

Datavisualisering i historieforskning; metode til at finde nye oplysninger om Holocaust og dens ofre.

Data visualization in historical research; a method to find new information about Holocaust and its victims.

Navn: Lone Maor Guldborg Mikkelsen (KMC144)
Uddannelse: Kandidat i informationsvidenskab og kulturformidling
Dato for aflevering: **3. september 2018 kl. 13**
Vejleders navn: **Professor Niels Ole Finnemann**
Antal anslag inkl. mellemrum: 168.882

Summary

This project argues, that data visualization as a research method, can be used for finding geographical areas where it can be productive to use resources on qualitative methods to find not yet registered holocaust victims and other new knowledge about holocaust.

The visualizations show a clear pattern in the difference between the data sets relative to the visualizations from Poland. A rather big surprise is that generally more people are registered in YadVashem's name database (NDB) than in the Polish census of 1931. Therefore, additional visualizations have been made where NDB is held against a German data set from 1939, where data is more accurate, newer and closer to the period of NDB. These show the same trends, though to a lesser degree. This may argue, that there may be some redundancies in NDB, more than previously estimated.

Data visualization as a method cannot answer questions about why - but only show that there is a tendency. The method is reliable, but as all data driven research, there has to be critical analyses and preferably the method can be used to support other more qualitative methods.

Indhold

| | |
|---|----|
| 1. Historieforskning og digitalisering | 5 |
| 2. Nye værktøjer – nye muligheder | 7 |
| 2.1 Problemformulering | 10 |
| 3. Metode | 11 |
| 3.1 Afgrænsning | 14 |
| 3.2 Motivation | 15 |
| 3.3 Begrebsafklaring | 16 |
| 3.3.1 Holocaust | 16 |
| 3.3.2 Deep Mapping (Herefter DM) | 16 |
| 3.3.3 Organiserede datasæt – database (herefter kaldet DB) | 17 |
| 3.3.4 Big Data (Herefter kaldet BD) | 18 |
| 3.3.5 Historical Geographical Information Science (Herefter kaldet HGIS)..... | 19 |
| 3.3.6 Choropleth mapping og heat mapping | 19 |
| 3.3.7 Application Programming Interfaces (Herefter kaldet API)..... | 20 |
| 3.4 Proces | 20 |
| 3.4.1 Udvælgelsesprocessen af Datasæt | 22 |
| 3.4.2 Software | 30 |
| 3.4.3 Setup..... | 32 |
| 4.0 Visualiseringer | 34 |
| 4.1 Klassisk Graf i Excel..... | 36 |
| 4.2 Excel til produktion af choropleth kort | 37 |
| 4.3 Tableau til produktion af choropleth maps | 38 |
| 5.0 Hvad kan udledes af visualiseringerne..... | 42 |
| 5.1 Samlet billede - hvad springer først i øjnene?..... | 46 |
| 5.2 Einsatzgruppernes betydning | 48 |
| 5.3 Redundans? | 51 |
| 5.4 De sovjetiske områder..... | 55 |
| 5.5 De tyske områder | 58 |

| | |
|--|----|
| 5.6 Alfabeternes betydning | 61 |
| 6.0 Diskussion..... | 63 |
| 6.1 Digital Humanities (DH)..... | 63 |
| 6.2 Lidt om data | 66 |
| 6.3 Visualiseringer i tid og rum | 67 |
| 6.4 Visualiseringer kan ikke stå alene..... | 68 |
| 6.5 Gode visualiseringer kræver gode tværfaglige teams | 75 |
| 6.6 Hvordan kan brugen af DVIZ rammesættes? | 76 |
| 6.7 Digitale Research Infrastrukturers betydning for DVIZ metode | 79 |
| 6.7.1 EHRI – en digital forsker infrastruktur. | 81 |
| 6.8 Serendipitet og datavisualisering | 83 |
| 7. Konklusion..... | 84 |
| 8. Perspektivering..... | 86 |

1. Historieforskning og digitalisering

Der er på alle områder i samfundet i dag stort fokus på de nye muligheder, som digitaliseringen åbner for – ikke mindst indenfor forskning generelt men også indenfor historieforskningen (Hughes 2016, Gregory & Geddes, Kahn 2011, Fisher 2012, Miller 2015, Sagi 2017, Szewrań 2017, Yau 2013, White 2010)

Arkæologien har allerede i en del år (Gregory & Geddes 2014, Kitchin 2013, Earley-Spandoni 2017) brugt Geo Information Systems og andre nye værktøjer som Artificial Intelligence, herunder machine learning, Internet of Things, Big Data og Datavisualisering er bare nogle af de begreber, der bliver undersøgt som værktøjer til at høste ny viden indenfor historieforskningen.

Specielt datavisualisering er blevet et yndet værktøj til at prøve at gennemskue de enorme datamængder, der produceres - ikke bare over de sociale medier fra vores private enheder; smartphones, pc, smart TV, hvor der bliver uploadet 510.000 Facebook kommentarer, 350.000 tweets og 300 timers You-Tube videoer hvert minut¹. Men også fra alle de andre digitale enheder, der er i stand til at kommunikere med hinanden over internettet², og som efterlader enorme mængder af digitale spor, som opsamles og kan gemmes som data-exhaust (Mayer-Schönberger, 2014) for muligvis at kunne udnyttes senere. De producerede datamængder vokser eksponentielt og vil derfor fremadrettet fylde mere og mere indenfor alle dele af samfundet (Finnemann, 2015), - ikke mindst indenfor forskningen. Med de komplekse og omfattende datatyper, der skal visualiseres, er der i dag mere tale om et medie end et værktøj (Yau, 2013); et medie både kræver kritisk analyse og at undergå kildekritik som andre medier (Drucker 2014, Gray et.al 2016)

¹ <https://www.facebook.com/universitet/videos/10155699640166138/> set d. 21/5/18 kl.10:40

² IoT: Reale objekter ("ting"), (som modsætning til digitale repræsentationer) med IP-adresse. Dvs. en betegnelse for, hvordan alting med tiden kan få et unikt ID og kan knyttes sammen via internettet. For eksempel ved hjælp af RFID-chips, der kan aflæses på afstand. <http://www.informationsordbogen.dk/search.php?text=iot&type=all> set 2/9/18 15,30

Også indenfor Holocaustforskningen³ har man i flere år forsøgt at bruge forskellige digitale tiltag for at skaffe ny viden om emnet. I bogen *Geographies of the Holocaust*, skrevet af Anne Kelly Knowles, Tim Cole og Alberto Giordano i 2014, bliver der for alvor fokuseret på de digitale muligheder indenfor Holocaustforskning ved at afprøve flere forskellige digitale værktøjer på forskellige typer data. Dels for at se Holocaust fra nye vinkler, men også for at forstå Holocaust ved at gøre begrebet visuelt og derved mere synligt (Knowles, Cole & Giordano, 2014).

Der har siden slutningen Anden Verdenskrig (herefter kaldet krigen) været fokus på at registrere så mange Holocaust-ofre som muligt (Miller 2015, Sagi et.al. 2017, Aderet 2017), dels for at vise respekt for ofrene, dels for at kunne give fred til efterladte og dels for at kunne håndtere de store erstatningssager der er fulgt i kølvandet på krigen⁴.

Mange store internationale organisationer som Yad Vashem⁵, samarbejder på tværs af grænser og har for nuværende registreret mere end 4.2 millioner navne i deres digitale navnedatabase (NDB). Også i USA har United States Holocaust Memorial Museum⁶ (USHM) store databaser med navne på ofre i deres omfattende arkiver. I Europa er et af de største arkiver for holocaustforskning i Bad Arolsen⁷, hos International Tracing Service (ITS) som vedligeholder og opdaterer store databaser, for de mange mennesker der forsvandt under krigen.

Men selv om nazisterne med deres systematiske tilgang til udryddelsen af jøderne, blandt andet ved at føre lister over deporterede - i Auschwitz som eksempel, nummererede man fanger med tatoveringer og markerede de indsatte efter politisk ideologi, religion, kriminelle fortid, sociale status og lignende -, blev store dele af de nazistiske arkiver destrueret af nazisterne selv i krigens sidste tid. Det har stillet holocaust-forskere overfor uhyre store udfordringer i forhold til originale kilder. Det materiale, der har overlevet, og som man har til rådighed, er meget heterogent; spredte stumper fordelt over mange lande og skrevet på mange forskellige sprog og med forskellige alfabeter – både maskinskrevet og håndskrevet (Sagi 2017, Kahn 2011, Miller 2015). De få lister,

³ Argumentation for at skrive Holocaust med stort H følger under afsnittet begrebsafklaring.3.3.1 Holocaust

⁴ *The International* Commission on Holocaust Era Insurance Claims (ICHEIC) <https://icheic.ushmm.org/> set 2/9/18 kl. 20,50

⁵ [3.4.1 Udvælgelsesprocessen af Datasæt](#)

⁶ <https://www.ushmm.org/learn/mapping-initiatives/geographies-of-the-holocaust/landscapes-of-experience-evacuations-auschwitz/> set 23/8/2018 kl.11,30

⁷ <https://www.its-arolsen.org/en/about-its/> set 23/8/2018 kl.11,30

der er dukket op, er for længst ført ind i de store databaser⁸. De mange lister, man troede, der ville dukke op, har man nu til dels opgivet at finde, mens der bliver brugt flere og flere resurser på at finde nye metoder til at danne overblik over Holocaust og ikke mindst at finde data på de ofre, der stadig ikke er kendt og registrerede (Aderet 2017, Knowles, Cole & Giordano 2014)

2. Nye værktøjer – nye muligheder

En af de store udfordringer, som det jødiske folk har stået overfor siden krigens afslutning, er som nævnt at registrere navne på alle ofrene for Holocaust (Aderet 2017, Gilbert 1982, Sagi et.al. 2017, Kahn 2011). De heterogene data – og stumper af data - som er spredt over hele verden, de mange forskellige sprog, forskellige lovmæssige udfordringer på kryds og tværs af grænser og specielt omfanget af uhyrlighederne med det kolossalt store tal af Holocaustofre gør forskning i Holocaust og eftersøgning af ikke-registrerede ofre kompliceret, svært og uoverskueligt.

Der er derfor sideløbende med computerens udvikling, og ikke mindst gennem Internettets udvikling, kommet fokus på at bruge computerkraft, database-teknologi og andre digitale værktøjer til at registrere de data, man allerede har samlet og på at bearbejde disse enorme datamængder for at finde nye måder og veje til at finde frem til de ikke-registrerede ofre.

Et af de nyeste tiltag er et stort projekt med det mål at finde og registrere så mange ungarske ofre som muligt. Det blev sat i gang af Yad VaShems forskningsafdeling i Jerusalem i 2007. På det tidspunkt havde Yad VaShem i deres navnedatabase registreret ca. 43 % af de ca. 600.000 ungarske jøder, der blev myrdet under Holocaust. Ved hjælp af digitale værktøjer som blandt andet Machine Learning (ML) lykkedes det færdige projekt - afsluttet i 2017 - med at tilføje ca. 225.000 navne, så der nu er identificeret over 80 % af de ungarske holocaustofre (Aderet 2017, Sagi o.a. 2017), hvilket må siges, at være et overvældende godt resultat.

Ud over at der selvfølgelig har været investeret umådeligt menneskelige resurser i projektet, er der samtidig blevet udviklet nye API'er og andre digitale værktøjer, som har udgjort fundamentet for projektet. Disse vil fremadrettet blive videreudviklet og brugt i nye projekter.

⁸ <https://www.yadvashem.org/archive/about.html> set d. 2/7/2018 kl.8,30

Dette projekt viser, hvor vigtigt det er, at der også i holocaust forskning bliver tænkt digitalt, så der kan blive fundet nye måder og veje til at få så mange ofre som muligt registreret. Metoden kan fremlægge informationer på nye måder – og derved lægge op til nye undersøgelser, - ny debat.

Jeg vil her afprøve datavisualisering som en ny metode i efterforskning og argumentere for, at datavisualisering kan bruges til at udvide forskningsområdet indenfor historieforskning og finde ny viden.

Datavisualisering er som sådan ikke et nyt begreb, som er kommet med digitaliseringen (Friendly 2008, Gregory & Geddes 2014). Dels da der er grafiske visualiseringer helt tilbage til hulemalerierne, og dels har statistisk grafik og datavisualisering rødder helt tilbage til de første landkort, produceret i år 200 BC af Claudius Ptolemy i Alexandria, som havde en form for længde- og breddegrader. Dette kort blev frem til det 14. århundrede brugt som reference indenfor kortlægning (Friendly, 2008).

Op gennem oplysningstiden, som var drevet af den rationelle tænkning, udviklede også statistisk tænkning sig. Charles Joseph Minard, en fransk civilingeniør, producerede tilbage i år 1869 en visualisering der viser Napoleons hærs march- og ofre under den Russiske kampagne i 1812.

Ud af denne visualisering kan læses uhyre mange detaljer, og selvom den er fremstillet, længe før den moderne computer blev opfundet, er den et rigtigt godt eksempel på, hvor stærkt et værktøj visualisering generelt er; hvor meget information der kan formidles uden brug af tekst og tal.

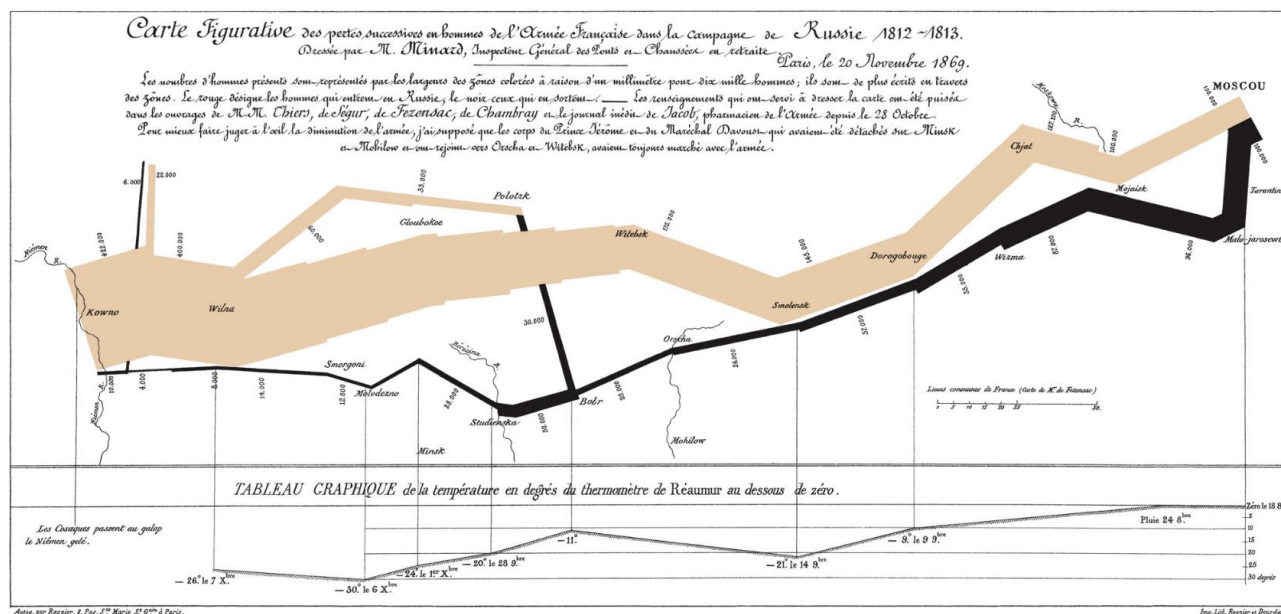


Illustration 1. Charles Minards kort over Napoleons hærs bevægelser og faldne i perioden 1812-1813. Billedet er sakset fra den engelske Wikipedia⁹.

De fleste, hvis ikke alle, typer visualiseringer som bliver brugt i dag, som for eksempel GIS-mapping, grafer og andre, er født og udviklet i andre discipliner (Drucker, 2014). Frem til midten af det 20. århundrede har nye teorier og nye grafiske former været en af årsagerne til, at statistisk grafik er blevet brugt oftere og oftere (Drucker, 2014).

Den populære statistiske grafik har sideløbende bevæget sig fra udelukkende at være formidlende til at være mere kompleks og forfinet. Da der i dag produceres kolossalt store datamængder - tendensen er, at den producerede datamængde vokser eksponentielt (Finnemann, 2014) - er det nu ikke bare forholdsvis nemt, men også nødvendigt at analysere disse store mængder med data-mining algoritmer¹⁰. Her er datavisualiseringer blevet et uundværligt værktøj til at danne overblik over disse store datamængder (Robinson 2017, Mikkelsen 2016).

Datavisualisering af store mængder data har i flere årtier, specielt i kølvandet på dataficeringen¹¹, været brugt indenfor forretningsverdenen og på de finansielle markeder. Også indenfor naturvidenskaberne har visualiserings-software længe været anvendt, både når de mindste partikler undersøges under mikroskop og når verdensrummet bliver undersøgt gennem teleskoper.

Dataanalyse af de store mængder data og visualisering af disse, bliver brugt både i søgen efter ny viden men også til understøttelse af allerede etableret viden (Finnemann, 2014). Der er her tale om både strukturerede datasæt og store ustrukturerede data-corpora – primært dog data, der er født digitalt. Fremtidens ”bog”, og hele måden vi deler information på vil være en være en kompleks kombination af mange forskellige elementer som tekst, digitale referencer med links, dataanalyse, søgning og henvisning ved hjælp af links, indeksering og databaseopbygning med tilhørende let tilgængelige værktøjer til visualisering af data og maskinlæsning. Alt dette formidlet over multimodale digitale platforme (Drucker, 2014) Et godt eksempel er de internationale

⁹ https://en.wikipedia.org/wiki/Charles_Joseph_Minard#/media/File:Minard.png 26/8/2018 10,30

¹⁰ En algoritme er en beskrivelse af en proces som en skal udføre en given opgave, som for eksempel en maskine (computer) skal udføre. <http://www.informationsordbogen.dk/concept.php?cid=2832> set 5/4/2018 kl. 10,30

¹¹ ”to datafy a phenomenon is to put it in a quantified format so it can be tabulated and analyzed” Meyer-Schönberger & Cukier, 2013. S. 78

forskningsplatforme, der allerede nu hurtigt vinder frem indenfor forskerverdenen og som kun bliver flere og flere og større og større. Disse vil jeg komme ind på senere i opgaven.

Visualiseringer fremstillet ved hjælp af digitale værktøjer kan forholdsvis let fremstille grafiske, overskuelige illustrationer af store mængder data, genereret ud fra givne forespørgsler. Det er som med mikroskopet, hvor en ny verden viser sig, når vores hjerne får mulighed for at se kendte objekter med et helt nyt perspektiv i den mikroskopiske verden.

Specielt med de store mængder meget heterogene data som i Holocaustforskning giver datavisualisering mulighed for at se data ”oppefra” i en slags fugleperspektiv, som ikke før har været muligt (Yau, 2013). Fugleperspektivet er en af Big Datas stærkeste fordele, men også i brugen af mindre datasæt, giver det mening at se data fra disse nye perspektiver.

Datavisualisering og dataanalyse kan ikke give et svar på *hvorfor* og *hvordan*, Holocaust kunne finde sted. Det vil derfor altid være nødvendigt når der bruges datadreven forskning, at bruge klassisk historisk metode, med fokus på nærlæsning og omhyggelig kildeanalyse. (Kristensen, 2007). Datavisualiseringer kan pege på nogle tendenser, korrelationer og mønstre, hvorefter der vil kunne stilles nye spørgsmål og der vil kunne bygges nye algoritmer, der kan høste nye data, som kan føre til ny viden indenfor historieforskningen.

2.1 Problemformulering

Hvordan kan vi bruge datavisualisering til at finde frem til de geografiske områder, som med stor sandsynlighed kan tilføje navne på endnu ikke-registrerede holocaustofre?

- *Hvordan kan datavisualisering som forskningsmetode bruges som metode til at pege på geografiske områder, hvor der stadig kan findes ikke-registrerede holocaustofre?*
- *Hvilken indflydelse på resultatet vil det have, at der bliver brugt grupperede data allerede i opbygningen af databaser?*
- *Kan metoden pege på andre interessante oplysninger om perioden?*
- *Er denne type datavisualisering egnet som metode indenfor holocaustforskning generelt?*

3. Metode

Visualiseringerne giver mulighed for ved hjælp af kvantitativ metode - at anskue et givent fænomen på makroniveau for derefter at kunne analysere eventuelt fremkomne mønstre og sammenhænge. Det er op til fagspecialister indenfor det respektive domæne at bruge analyserne som grundlag for yderligere analyse af eksemplerne, nu på mikroniveau hvor analyse, modsat makro niveau, kvalitativ metode, baseret på dybdegående analyse og undersøgelse af et givent fænomen, er grundlæggende (Yau 2013, Mikkelsen 2016).

Data kan ved hjælp af visualiseringer præsenteres på flere forskellige måder: grafisk som billeder, hvor specifikke situationer ses i et to- eller tredimensionelt billede, som 3D visualiseringer, hvor de rummelige tendenser giver yderligere muligheder for analyse og som Geo-spatial visualiseringer, hvor man kan observere ændringer over tid og sted (Yau 2013, Gregory & Geddes 2014). En af de mest effektive visualiseringer er de interaktive Geo- spacielle visualiseringer, som ved hjælp af avanceret software, giver brugeren mulighed for selv at vælge og styre mange forskellige parametre.

Metoden datavisualisering vil i dette projekt blive betragtet ud fra en post-positivistisk videnskabsteori, som en nyere, kritisk videreudvikling af den klassiske positivisme, hvor videnskabelig erkendelse kan opnås gennem systematisk indsamling af empiri¹² - men som, - modsat den klassiske positivisme - tager højde for og har fokus på, at alle fænomener kan fortolkes og derfor vil være påvirket subjektivt (Mikkelsen, 2016).

Ved at visualisere og kombinere en navnedatabase, der indeholder allerede registrerede holocaustofre, med en folkeoptælling fra 1931 og ved hjælp af visualiseringer se på de geografiske områder fra begge datasæt i samme geografiske kort, er tesen, at der vil fremkomme overlap eller manglende overlap, som derved kan pege på områder, hvor der i NDB endnu ikke er registreret det antal indbyggere, der befandt sig i det respektive geografiske område i 1931. Disse områder vil så være grundlaget for efterforskning på mikroniveau, hvor mere traditionelle historiske metoder vil

¹²Klassisk positivistisk teori kræver dog også, at data kan testes ved induktion; Induktion som metode, tager udgangspunkt i konklusion ud fra sammenfatning af mange enkeltstående observationer (Beck-Holm 2011)

blive brugt. Visualiseringerne vil derfor dels lægge op til fund af ny viden i forhold til den geografiske fordeling af registrerede ofre, og dels mindske den høstak, hvori nålene ligger gemt. Metoden lægger absolut også op til serendipitet og derved fund af ny viden.

Ved hjælp af forskellige lag i visualiseringerne, som i en form for heatmapping, Deep Mapping eller anden type visualisering, kan der måske ses mønstre, korrelationer eller andre tendenser, der kan give ny viden om området i perioden eller pege på geografiske områder, hvor det ville være realistisk, på mikroniveau¹³, at koncentrere ressourcer på at søge efter navne på ofre, der endnu ikke er registrerede, og derved bruge datavisualisering i eftersøgning af manglende navne på ofrene. Der er ligeledes fokus på processen i projektet og ved minutiøst at redegøre for denne, vil jeg undersøge, hvilke underliggende dataorganiseringsstrukturer, der egner sig til metoden. Opgaven er en metodeopgave, der har til hensigt at pege på nye metoder og værktøjer indenfor informationsvidenskab, formidling og informationsorganisering samt at diskutere fordele og ulemper, metoden kunne have.

At finde frem til noget, der IKKE er fundet, er selvsagt altid en udfordring, og derfor vil jeg her argumentere for, at netop datavisualisering, hvor det visuelle er det afgørende element, kan være med til at pege på geografiske områder, hvor det vil være hensigtsmæssigt at bruge ressourcer på blandt andet at finde ikke-registrerede ofre ved hjælp af metoder på mikroniveau.

Som eksempel på metodens anvendelse indenfor historieforskning er her valgt, ved hjælp af digitale værktøjer, at finde frem til geografiske områder, som viser tendens til, at der ved undersøgelser på mikroniveau kan være mulighed for at finde endnu ikke-registrerede navne på Holocaust ofre. Rammerne for projektet er derfor Digital Humanities, med fokus på datavisualisering indenfor Holocaustforskning.

Da rammerne for projektet er Digital Humanities, vil jeg i diskussionsafsnittet kort redegøre for hvor feltet står i dag.

¹³ Her overfor makroniveau, som datavisualisering af store datasæt hører ind under (Mikkelsen, 2016).

Det understreges, at der er fokus på *metoden* i opgaven, og det vil derfor være muligt at relatere opgavens metodiske resultater til datavisualiseringens potentiale som metode i andre sammenhænge og på andre fagområder end holocaustforskning. Der er lagt vægt på at afprøve datavisualisering som kvantitativ metode på makroniveau indenfor et forskningsområde, som historisk set er bygget på dybdegående, kvalitativ forskning på mikroniveau (Kristensen, 2007). Samtidig er der fokus på diskussionen om datadrevet forskning – herunder fordele og ulemper.

Da der (endnu) ikke findes ret mange projekter om visualiseringer af data fra Holocaust, har jeg, for ikke at sprede visualiseringerne ud over for mange domæner, benyttet en del materiale/visualiseringer fra bogen *Geographies of the Holocaust* af Knowles, A. K., Cole, T., Giordano, A - Copyright Date: 2014 - Published by: Indiana University Press.

Det skal dog her nævnes, at der indenfor de sidste par år, er kommet langt større fokus på brug af data indenfor holocaustforskning – her har EHRI været frontløberen med et stort udbud af seminarer, hvor der diskuteres Holocaust og Digital Humanities. Disse seminarer leder til forskellige pilotprojekter, hvor de digitale værktøjer bliver afprøvet. Et sådant er for eksempel projektet med at bruge Wikidata¹⁴ til at visualisere ghettoområder under krigen.

Opgaven her kan ses som et mindre projekt, hvor det undersøges om datavisualisering som metode er relevant for at finde frem til nye informationer om Holocaust generelt og specielt navne på holocaustofre, som endnu ikke er registreret i de store internationale databaser.

Der vil i denne opgave afprøves, om metoden med at arbejde med data hvor grupperinger af registreringer er grundlæggende er effektiv, eller om et langt større projekt hvor individuelle navne og/eller adresser på ofre/indbyggere vil give større muligheder for at finde frem til ikke-registrerede ofre.

Udvælgelse af data er et væsentligt parameter, ikke bare her i projektet men ved brug af alle datadrevne metoder, som jeg senere vil komme ind på, og der vil derfor i opgaven blive redegjort for, hvorfor de respektive data er valgt. Det er dog vigtigt at forstå, at fordi der er fokus på metoden, er datas konsistens og homogenitet ikke kritisk afgørende for opgavens resultat, da datavisualisering

¹⁴ <https://blog.ehri-project.eu/2018/02/12/using-wikidata/> set 1/8/2018 10,00

som metode generelt egner sig bedst til store data-corpora og til at pege på korrelationer, clusters, mønstre og afvigelser i disse, og vil derfor vil argumentere ud fra disse parametre. Hvis det viser sig, at metoden med at bruge data der er grupperet, vil det være en langt mindre ressourcekrævende metode at finde frem til nye oplysninger om holocaustofre på, end hvis der skal arbejdes med store datasæt.

Der vil blive redegjort for valg af software og valget vil blive begrundet.

Der vil også blive redegjort for de underliggende dataorganiseringsstrukturer og brugen af de valgte visualiseringsprogrammer, brugerflader samt for valg af præsentationsudtrykket og de grafiske udtryk. I analyseafsnittet vil jeg komme ind på, hvordan og hvorfor de valgte visualiseringer kan læses ud fra forskellige parametre og hvilken indflydelse, de forskellige valg af software har på resultaterne.

Da mine egne visualiseringer er interaktive, og brugeren blandt andet selv kan vælge at ændre på de forskellige parametre, vil der være mulighed for at logge på min offentlige profil på [Tableau Public](#), hvor disse kan afprøves.

De færdige visualiseringer analyseres i analyseafsnittet, hvor visualiseringerne studeres indgående og bliver kommenteret i forhold til deres effektivitet set ud fra den opstillede problemformulering. Hvor det vurderes relevant, vil egne visualiseringer bliver sammenlignet med andre visualiseringer indenfor emnet.

Resultaterne af analysen fremlægges og diskuteres i diskussionsafsnittet ud fra de nyeste forskningsartikler på området. Det vurderes, om metoden er brugbar og effektiv med fokus på mulige problematikker i forhold til kildekritik ved brug af digital visualisering i historieforsknings sammenhæng.

3.1 Afgrænsning

Da der i denne opgave er fokus på at finde jødiske holocaustofre, der er blevet myrdet under Holocaust, vil begrebet udelukkende gælde de jødiske ofre, der døde i perioden. Det vil ikke inkludere data fra alle de mange andre ofre; millioner af mennesker med forskellig nationalitet, sprog, kulturel baggrund og religiøs overbevisning, som også var ofre for nazisternes grusomheder, men som overlevede. Deres historie er ikke mindre vigtig end de myrdede jøders, men den er en anden, og disse ofre er derfor ikke en del af, hvad der i denne opgave er betegnet som holocaustofre.

Der er i denne opgave ikke taget stilling til etiske problemstillinger, som der derfor heller ikke vil blive redegjort for og diskuteret. Hele problematikken omkring etikken i holocaustforskning ses som et selvstændigt emne, og er derfor ikke her medtaget.

3.2 Motivation

Der er flere årsager til, at opgaven her netop har fokus på Holocaust: Først og fremmest fordi Holocaust er et afsnit i verdenshistorien, der har haft og stadig har stor betydning for hele verdenssamfundet. En direkte effekt af Holocaust er staten Israels tilblivelse i 1948, og netop her er Holocaust diskursen konstant under debat og vigtig for selvforståelse for det jødiske samfund både i Israel og diasporaen.

Det er vigtigt, at vi lærer mere om begivenheden, for at kortlægge perioden, og dermed prøve bedre at forstå, hvad der skete og hvorfor. Der må hele tiden insisteres på at bruge de nyeste teknikker og metoder indenfor forskningen til at lære nyt om Holocaust, - som genererer nye spørgsmål, hvis svar hjælper os til at prøve at forstå.

Da der i al forskning ved hjælp af digitale data og analyse af visualiseringer skal bruges klassisk funderet domænerelateret viden (Boyd & Crawford 2012, Ethington 2007, Gregory & Geddes 2014, Knowles, Cole & Giordano 2014), har det for mig også været oplagt at vælge Holocaust som "case" i mit speciale. Historie har altid haft min store interesse, og specielt Holocaust har haft min interesse i mange år. Senest har jeg tilføjet en del faglig viden også.

I løbet af mine informationsstudier har jeg ofte brugt emnet til opgaveskrivning, da jeg mener, at historiefaget har mulighed for at udvide forskningsområder ved at inddrage digitale værktøjer mere i forskningen.

I løbet af mine kandidatstudier har jeg blandt andet taget faget *Holocaust i Forskning og Debat* ved lektor og Ph.d. Lene Sofie Bak fra Københavns Universitet. Derudover har jeg deltaget i et EHRI-seminar i Trieste i Italien, hvor de nyeste informationer og værktøjer indenfor Holocaustforskning blev fremlagt af specialister fra hele verden. Jeg følger mange netværk for Holocaustforskning på forskellige digitale platforme og blogs, såvel som jeg løbende holder mig opdateret om nye tiltag og informationer indenfor faget.

Ovenstående har givet mig en del viden indenfor domænet, og vil derfor give mig mulighed for også at trække på egen viden indenfor holocausthistorien i mit arbejde med specialet, specielt i analysen af data.

3.3 Begrebsafklaring

Da der i opgaven bruges forskellige begreber, der ikke har klare definitioner, vil jeg her redegøre for, hvordan begreberne er brugt i netop denne opgave:

3.3.1 Holocaust

Begrebet Holocaust er i tiden efter krigen brugt som betegnelse for nazisternes forbrydelser mod jøderne. Ordet Holocaust blev brugt som engelsk oversættelse for det hebræiske ord ”Shoa”, som betyder ”Brændoffer”¹⁵.

I dag bliver begrebet dog brugt om folkedrab generelt. I nyere holocaustforskning og debat, bliver selve ordet nu mange steder skrevet med lille h, netop for at understrege, at det er ét folkedrab ud af flere, som nødvendigvis må sidestilles for netop at holde fokus på at undersøge, hvad der udløser disse begivenheder.¹⁶ Selvom jeg hører til denne sidste skole, har jeg i opgaven valgt at skrive Holocaust med stort H, for at pointere at der her KUN er tale om jøder.

3.3.2 Deep Mapping (Herefter DM)

Der er ingen akademisk konsensus omkring hvad begrebet ” Deep Mapping” er, men kan dog forklares med at være en mange-lags, digital kartografisk repræsentation, der kan visualiseres, redigeres og ændres i forhold til hvilke behov visualiseringen skal opfylde (Early-Spandoni, 2017). En DM er ikke bare et digitalt kort, men en kompleks konstruktion som er sammensat af meningsfulde lag af informationer, hvor brugeren selv kan være med til at skabe sin egen mening og historie (Early-Spandoni, 2017). DM kan ses som det redskab som ofte bliver brugt i interaktive visualiseringer, som egner sig bedst til at vise tendenser over tid og rum, da brugeren i interaktive visualiseringer selv kan påvirke visningen ved for eksempel at zoome ind og ud, se hvordan situationer ændrer sig over tid, sted og andet. Et godt eksempel er fra Mapping the Holocaust¹⁷ (Knowles, Cole & Giordano 2014) hvor brugeren kan se en afspilning af elementernes forandring og bevægelse over tid og rum. Min egen digitale udgave af visualiseringer i denne

¹⁵ <http://folkedrab.dk/sw52685.a>

¹⁶ <https://www.folkedrab.dk/artikl.er/begrebet-holocaust>

¹⁷ (https://web.stanford.edu/group/spatialhistory/cgi-bin/site/viz.php?id=379&project_id=0 23/6/2018 11.00)

opgave, ses som en forsimplet udgave af DM, da der ikke her er taget en tidsfaktor med i visualiseringerne.

DM er et effektivt redskab indenfor ikke bare historieforskning, men også arkæologi, da det blandt andet kan vise udviklinger over tid og rum. (Friendly 2008, Cole & Giordano 2014, Early-Spandoni 2017). Med de store mængder data der produceres og de store sociale netværks vokseværk, er det blevet nødvendigt at finde nye måder for at kunne gennemskue kompleksiteten af disse netværk og for at kunne visualisere data mere nuanceret. Indenfor arkæologi er begrebet *digital storytelling* oppe i tiden, hvor DVIZ bruges til at undersøge og forstå essensen af store mængder data (Early-Spandoni 2017, Finnemann 2014).

3.3.3 Organiserede datasæt – database (herefter kaldet DB)

En database er her brugt som udtryk for en udvalgt sæt data, organiseret under struktureret og specifikt udvalgt Meta Data.

DB består af forskellig størrelse datasæt og er sorteret i et Knowledge Organization system (KOS) ved brug af klassisk databaseopbygning og teknologi. Mange af disse datasæt er så store, at det ikke er muligt at analysere data uden hjælp fra informationsteknologi. Data er allerede organiseret under forudbestemt Meta Data ved hjælp af computersoftware og manuel sortering, og ligger i en struktureret database, klar til at blive analyseret og visualiseret. Den store fordel ved en organiseret DB er, at det nemt, ved hjælp af forskellige API'er, hvoraf mange er let tilgængelige, er muligt at køre forskellige mindre datasæt sammen ved hjælp af automatisk indeksering.

Data bliver hentet ind i en Master KOS, hvorefter indekseringen for eksempel kan foretages ved hjælp af Machine Learning hvor data sammenlignes med store mængder allerede eksisterende data og derefter indekseres, for igen at blive brugt til nye data og indeksering (Sagi, 2017) eller ved en kombination af Probabilistic og Rule-based indeksering. Metadata som indekseres i Master KOS skal, for at være konsistent, være holdt i et indekseringssprog som for eksempel XML og godkendt i henhold til gældende ISO-dokumentation for eksempel ISO 5964 (Mikkelsen 2015).

DB ses derfor i denne opgave som data der er indsamlet, manipuleret, rensset og eventuelt sammenkørt fra flere datasæt og lagt ind i strukturerede databaser, hvorfra der kan visualiseres ved hjælp af visualiseringssoftware. En DB vil derfor ikke være neutralt forankret, da den har gennemgået en lang proces hvor der har været truffet menneskelige valg under hele processen.

Opbygningen af DB specielt i forhold til HGIS-databaser, er som regel den del af et projekt, der tager længst tid. Det kræver fagviden at kunne udvælge de bedste parametre ud fra de kriterier projektet stiller. Selvom det nok er den fase, der kræver mest fagviden, er det en del af processen, der desværre ikke er nok fokus på (Gregory & Geddes, 2014).

3.3.4 Big Data (Herefter kaldet BD)

Der er ingen entydig og klar definition på begrebet BD. Det kan spores tilbage til 1997, NASA blandt andet har brugt det i forbindelse med datahåndtering NASA (Finnemann, 2014).

BD høster data fra Internettet ved hjælp af Crawlere¹⁸ og matematiske algoritmer¹⁹. De indsamlede data bliver lagret som enorme mængder usorteret data, som efterfølgende kan bruges til forskellige formål. BD har været brugt og bruges flittigt på de kommercielle markeder, blandt andet til at målrette reklamer.

Google Books er et godt eksempel på BD; selvom der kan rejses spørgsmål omkring Googles metode at samle data til Google Books, tilbyder den dog nye søge- og analysemetoder, som humanister med glæde burde tage imod (Finnemann, 2014).

Der er dog en vis konsensus (Mayer-Schönberger 2014, Finnemann 2014) omkring en definition på BD som mindst skal bestå af et datasæt der opfylder forskellige antal "V'er". Det synes, at et mindstekrav for at kunne definere et data-corpora som værende DB er at det indeholder data der kan beskrives ud fra mindst disse tre V'er:

Extremely large Volume of data – størrelsen på data korpus - som er de usorterede data der samlet ses som et data-corpora.

Extremely large Velocity of data – Velocity er en fysisk mængde-vektor; både størrelse og retning er nødvendige for at definere den. Hastigheden af den skalære absolutte værdi (størrelse) kaldes Velocity

¹⁸ En søgerobot til internet, der afgrænser sin søgning til sider, der er højrelevante for et givent domæne <http://www.informationsordbogen.dk/search.php?text=crawler&type=all> 27/8/2018 20,30

¹⁹ En endelig liste over veldefinerede instruktioner for hvordan en given opgave skal løses. En algoritme har en given initial tilstand og vil afslutte når en defineret sluttetilstand er opnået. Algoritmer ligger bl.a. til grund for konstruktionen af software. <http://www.informationsordbogen.dk/concept.php?cid=2832> 27/8/2018 20,35

Extremely large Variety of data – stor variation i data. Forskellige typer usorterede, heterogene data modsat sorterede data, der ligger under metadata i bearbejdede kategorier. Da BD er et meget stort og stadig uklart begreb, vil BD i denne opgave bruges i den brede definition som nævnt her ovenfor.

3.3.5 Historical Geographical Information Science (Herefter kaldet HGIS)

HGIS er her brugt som betegnelse for en DB hvortil der er tilføjet geografisk refererende data som blandt andet er nødvendigt for at kunne visualisere data på geografiske kort, som denne opgave er baseret på.

Mange datasæt indeholder allerede spatial attributter (Kitchin 2013, Edelstein 2017), da Historical Geographical Information Science (HGIS) i de sidste mange år har været praktiseret indenfor historieforskning (Gregory & Geddes 2014, Kitchin 2013, Early-Spandoni 2017). Tidligere indeholdt HGIS altid *attribute data*²⁰ i struktureret form, men med den hastige udvikling indenfor Informations Teknologi (IT), kan nu næsten alle typer af *attribute data* også ustruktureret, og så længe at der kan findes en location for disse, blive indlæst i et HGIS. (Gregory & Geddes 2014, Mikkelsen 2016). *Spatial data*²¹, er af kvantitativ type baseret på tal og matematiske formler. HGIS bliver i dag brugt indenfor mange forskellige faglige videnskabelige discipliner da den hastige teknologiske udvikling gør potentialet for at bruge HGIS både kvalitativ og kvantitativt større, - og spreder sig fra de historiskfaglige områder til andre områder af de humanistiske discipliner; det der i dag, som en forlængelse af Digital Humanities, bliver kaldt *Spatial Humanities* (Knowles, Cole & Giordano 2014. Gregory & Geddes 2014)

3.3.6 Choropleth mapping og heat mapping

Choropleth mapping er betegnelsen for en tematisk visualisering af data hvor forskellige områder bliver farvet eller gradueret i farver/skygger, i forhold til vægten af de data der bliver visualiseret. Heatmapping er meget lig choropleth mapping, men er ikke kombineret med geografiske data. Denne type visualisering er generelt god til at give et overblik over intensitet af store mængder data.

²⁰ Attribute Fysisk data siger *hvad* hvorimod...

²¹ ..spatial data siger *hvor* (rummelig)

Excels software i sin brug er meget rigtigt; bruger kan for eksempel ikke selv at vælge farver og grad af gennemsigtighed og der kan ikke arbejdes med Dual Axis (hvor forskellige datasæt sættes ind i samme kort – uden at blive blandet).

Choropleth mapping er, med alle sine fordele ikke egnet til denne opgave, men vil være interessant at arbejde med i et større Big Data projekt som arbejder på navne/adresseniveau.

Heat Mapping derimod kan være interessant at bruge for evt. at underbygge eventuelle opdagelser der kan dukke op i visualiseringerne.

Heat mapping er en type visualisering hvor tabeller bliver oversat til et sæt farver, og hvor intensiteten i summerede antal punkter derved træder tydeligt frem (se Illustration 5). Det vil eksempelvis sige at mørkere områder indikerer flere samlede datapunkter. I denne opgave vil en HM sammen med Geodata, kunne vise på et kort, hvor der er stor koncentration af dels jødiske indbyggere og dels ofre for Holocaust.

3.3.7 Application Programming Interfaces (Herefter kaldet API)

API'er er et stykke software²², der kan få de mange forskellige digitale sprog²³ der blandt andet bliver brugt til at bygge databaser med, til at samle data fra mange forskellige datakilder. De bruges til forskellige formål – også til geo-spatielle analyser og visualiseringer. Måden hvorpå data høstes, bearbejdes og ikke mindst visualiseres på, er medvirkende til at bestemme hvad der kan udledes af viden og dermed også til hvilke formål de respektive visualiseringer er egnede. Databaser og data-infrastrukturer, ændrer derved fundamentalt på måden og metoden for organisering af forskning (Kitchin, 2014).

3.4 Proces

For at starte processen, er jeg startet med at producere en manuel mockup²⁴, der visuelt skal vise, hvordan det endelige produkt kunne se ud. Min intention er at visualisere antallet af borgere fra et

²² En API kan være til et webbaseret system, operativsystem, database system, computer hardware eller software bibliotek

²³ Berners-Lee's semantiske Web er bygget på, at de objekter og ressourcer der figurerer på nettet, bruger et Mark-up sprog for at en iboende struktur af nettet skal kunne forbinde dokumenter og information, og at en information om et navngivet objekt eller ressource nemt skulle kunne findes (Kitchin, 2014).

²⁴ En illustration af hvordan et færdigt produkt vil kunne se ud.

datasæt fra før krigen overfor antal ofre registreret i *Yad Vashem's "The Central Database of Holocaust Victims' Names"* (NDB)

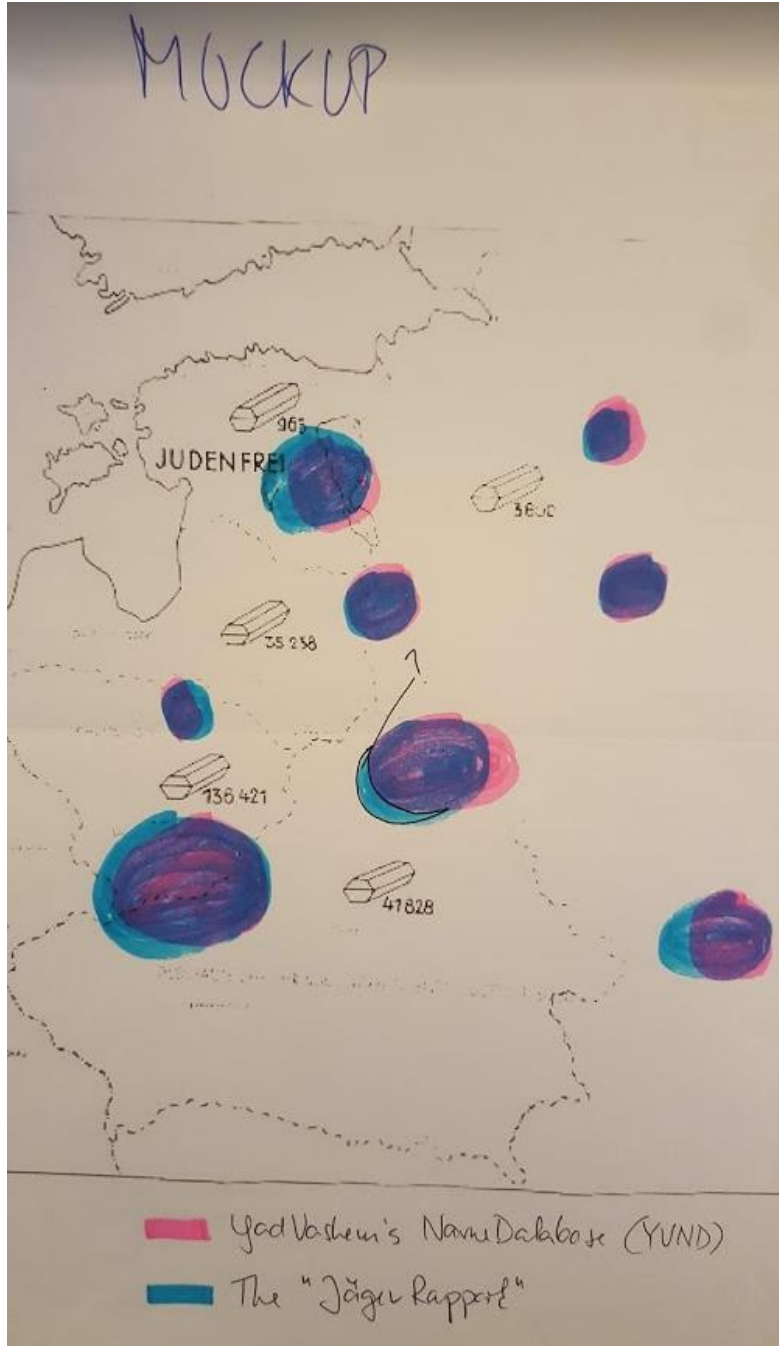


Illustration 1. Mockup af visualisering. Lavet på Jägers egen visualisering. Egen modificering.

På dette tidspunkt i processen, var der ikke endeligt taget beslutning om hvilke data der skulle bruges eller hvilket software der skulle bruges.

3.4.1 Udvælgelsesprocessen af Datasæt

Det næste step i processen er udvælgelse af data.

Det er besluttet at foretage visualiseringerne på gruppe-niveau og ikke på individ-niveau. På individ-niveau indeholder datasættene også data per indbygger/offer; det vil sige mindst fulde navne og hvor for eksempelvis specifikke adresser også kan fremgå. Et projekt af denne type vil generere meget store datasæt og vil kræve meget computerkraft. Visualiseringer på individ-niveau kræver blandt andet langt større databehandlings kapacitet og IT-faglig ekspertise i et eller flere programmeringssprog for at tilrette de store datamængder, samt en del menneskelige resurser. For generelt at teste visualiseringsmetoden for effektivitet, når der visualiseres på gruppe-niveau frem for på individ-niveau – som er et af formålene med opgaven, er det niveauet der er valgt i denne opgave. Opgaven, med sin begrænsning i tid og resurser, egner sig i øvrigt også bedst til at visualisere og analysere på gruppe-niveau.

Det må allerede her formodes, at et Big Data projekt, hvor data er på individ-niveau vil kunne levere et bedre og klarere resultat, idet jo større datamængder jo bedre muligheder for at vise mønstre, afvigelser, korrelationer og serendipitet.

Da opgaven er en metodeopgave, er der for at få så gode resultater som muligt, afprøvet forskellige datasæt fra forskellige kilder. Her nævnes de bedst egnede kilder fundet og afprøvet:

- *Jäger Rapporten*²⁵ - Den såkaldte Jäger Rapporten er en af de bedst bevarede originale kilder der findes indenfor registrering af holocaustofre. Rapportens fulde navn er *Complete tabulation of executions carried out in the Einsatzkommando 3 zone up to November 1941*. Rapporten er et originalt dokument med optegnelse af myrdede jøder og andre uønskede minoriteter, skrevet af Karl Jäger i 1941. Karl Jäger var øverstkommanderende for SD og Security Police i Litauen. Han var kommandør af Einsatzgruppe A, en likvideringsgruppe der primært opererede i Litauen, og var dermed ansvarlig for udslættelsen af Litauens jøder. I december 1941 rapporterede han, at der ikke var nogen jøder tilbage i Litauen bortset fra

²⁵ Yad Vashems kopi af [The Jaeger Rapport](#) set d.4/2/18 kl. 10,10

dem i tre ghettoer: Siauliai, Kovno og Vilna. Tre måneder senere udarbejdede Jæger en rapport, der omfattede antallet af mennesker dræbt af hans enhed: Dette er den såkaldte Jäger rapport.²⁶

- *Deutsches Bundesarchiv*, - De tyske forbunds arkiver har lovligt mandat til at sikre den føderale regerings arkiver på lang sigt og gøre dem tilgængelige til brug. Disse er dokumenter (herunder filer, kort, billeder, plakater, film og lydoptagelser i analog og digital form), der er produceret af det hellige romerske imperiums centrale kontorer (1495-1806), det tyske forbund (1815-1866) af det tyske rige (1867 / 71-1945), beskæftigelseszonen (1945-1949), den tyske demokratiske republik (1949-1990) og Forbundsrepublikken Tyskland (siden 1949)²⁷

Arkivet opbevarer en meget omfattende samling af dokumenter fra Holocaust-perioden, og er en oplagt datakilde til visualiseringer. Det er også her, at den original Jäger Rapport bliver opbevaret.

- *Zentralarchiv zur Erforschung der Geschichte der Juden in Deutschland*²⁸ - Er en institution under det Jødiske Centralråd i Tyskland. Det har arkiver i Heidelberg. Arkivet sørger for at opbevare og indsamle dokumentation for jødernes historie i Tyskland, og beskæftiger sig primært med forskning.
- *The 1939 German "Minority Census" Database*²⁹ - er blevet digitaliseret og organiseret i en online database af organisationen Tracing the Past. Tracing the Past's hovedformål er at fremme mindet om forfølgelsen. Tracing the Past blev grundlagt i 2014 som en registreret nonprofitorganisation i Berlin, Tyskland, af en ligesindet gruppe af berlinere, der primært består af historikere. Nuværende bidragsydere omfatter arkivarer og informationsteknologiske eksperter, der arbejder i mange aspekter af holocaust historie.³⁰

²⁶ YadVashems arkiv - <http://www.yadvashem.org/search.html?val=J%C3%A4ger%20Report&lang=en>. set 4/2/18 9,00

²⁷ <https://www.bundesarchiv.de/DE/Navigation/Home/home.html> set d.4/4/18 kl. 10,30

²⁸ [Zentralarchiv zur Erforschung der Geschichte der Juden in Deutschland](https://www.zentralarchiv.de/) set d.4/4/18 kl. 10,25

²⁹ [German Minority Census 1939](https://www.tracingthepast.org/) set d.4/4/18 kl. 10,20

³⁰ <https://www.tracingthepast.org/about> set 4/4/18 kl. 10,20

- *Atlas of the Holocaust* – som er baseret på data fra *Den Polske census 1931*³¹. Bogen er skrevet af Sir Martin Gilbert, University of Oxford (Gilbert, 2009) *Atlas of the Holocaust* udkom første gang i 1982. Ud over førsteudgaven i 1982 er bogen udkommet i tre efterfølgende opdaterede udgaver i henholdsvis 1993, 2002 og 2009. Bogen er den hidtil eneste bog som viser en kronologisk gennemgang af den *geografiske* side af Holocaust. De mange kort er håndtegnede, og de må derfor formodes, ikke at være så nøjagtige som eventuelle digitale kort fra samme områder, hvor HGIS ligger til grund for visualiseringerne.

På baggrund af analyser af kvaliteten af de forskellige data, og tests af samme, er det blevet besluttet at bruge *Atlas of the Holocaust* – som er baseret på data fra *Den Polske census 1939*, som grundlag for visualiseringerne af indbyggere før krigen

og

Yad Vashems "The Central Database of Holocaust Victims' Names" (NDB) er selvvalgt, da den er grundlæggende for det ene lag af visualiseringer, netop de ofre der er blevet myrdet, og som er registreret.

I de endelige visualiseringer er der udvalgt og høstet data om et antal registrerede jøder fra **tre forskellige kilder**:

Den første og til dels også vigtigste, da den er databasen hvor ofre er registreret og samtidig den mest nøjagtige og den største, er Yad Vashems "The Central Database of Holocaust Victims' Names" (Herefter kaldet NDB) hvor alle de registrerede er ofre, der er myrdet under Holocaust. Yad Vashem - "The World Holocaust Remembrance Center"- er etableret i 1953 i Jerusalem, Israel. Centret beskæftiger sig med alle former for Holocaustforskning, undervisning i Holocaust, fungerer som museum og fokuserer på innovative teknologier til registrering af den store samling. Databasen er kompleks, da der bliver registreret data både af forskere, medarbejdere ved Yad Vashem og af personer verden over med forbindelse til Holocaust.

³¹ Den Polske folketælling 1931 <https://www.archiwa.gov.pl/pl/578-archeion> set d.6/4/18 kl. 9,25

NDB bliver brugt til dels at registrere så mange ofre for holocaust som muligt, men også i stor udstrækning brugt til efterladte af holocaustofre til at finde frem til familiemedlemmer (mulige Holocaust ofre) der forsvandt under krigen. Derfor skal Databasen kunne håndtere mange forskellige sprog, og helt forskellige alfabeter, da ofrene kommer fra hele Europa, men skal samtidig fungere som en form for tesaurus, som kan håndtere at navne kan skrives forskelligt, læses forskelligt, udtales forskelligt eller være kælenavne inden for de forskellige sprog.

Allerede i 1954 begyndte det store projekt med at samle og registrere alle ofre for Holocaust. Først manuelt ved at udfylde "Pages of Testimony" med ofrenes biografiske data. Disse testimonies er blevet indekseret og arkiveret op gennem tiden, og i 1991 begyndte man at digitalisere de allerede samlede testimonies. I 2004 gik projektet The Central Database of Holocaust Victims' Names (NDB) online og har siden været under udvikling med mulighed for blandt andet at udfylde testimonies online og søge på flere sprog. Kilderne er skrevet på flere forskellige alfabeter: Hebraisk, latin, Kyrillisk og Græsk og på over 30 forskellige sprog, hvoraf mange af kilderne er håndskrevet. (Sage, 2017).

Efterhånden som der er dukket nyt materiale op, som lister der vurderes at være over ofre (for eksempel deportationslister, transportlister og lignende) er disse blevet tilføjet NDB. Både i processen med "Pages of Testimony" og andre data, kan der være mulighed for redundans. Denne mulighed er dog vurderet til at være forholdsvis ubetydelige (Sagi 2017, Kahn 2011).

Yad Vashems hjemmeside er opbygget som et facetteret Knowledge Organisation System (herefter kaldet KOS). I en hovedmenu er der adgang til online udstillinger, links til Yad VaShems egne databaser samt links til databaser over hele verden, der beskæftiger sig med holocaustforskning.

Yad Vashems digitale samlinger består i øjeblikket (august 2018) af følgende databaser, som findes under fanebladet "Digital Collections": Navnedatabasen (NDB), Fotodatabasen, Dokumentdatabasen, De Retfærdiges database, Biblioteksdatabasen, Filmdatabasen og Shoa Relateredes database.

Selve navnedatabasen er en del af informationssystemet, som kombinerer et klassifikationssystem med et navigationssystem og et søgesystem.

Der indsamles data fra mange forskelligartede kilder. De kan groft inddeles i tre typer datasæt, hvoraf det kun er det første, som den private bruger er involveret i. De to sidste er primært med fokus på forskningsinstitutioner:

I. **Pages of Testimony (PoT)** som er ensidede formularer, der er udfyldt og indleveret til Yad Vashem helt tilbage fra lige efter krigens afslutning. De PoT's som blev indsamlet inden den digitale databases opstart, er nu scannede, digitaliserede og metadata er lagt i databasen. Det er i dag muligt at udfylde en PoT online på Yad Vashems hjemmeside. Formularen indeholder 37 felter som udfyldes med offerets biografiske data. PoT kan også udskrives, udfyldes manuelt og mailes til Yad Vashem. Der er, med PoT som kilde ca. 2.600.000 navne i databasen i ca. 30 forskellige sprog, skrevet i 4 forskellige alfabeter. Formularens tekst til beskrivelse af felter, både i papirudgave og digitalt, fremstår på det sprog, der på forhånd er valgt at få præsenteret på hjemmesiden. Alle PoT's bliver manuelt gennemlæst, og hvis muligt holdt op mod anden dokumentation, af personale på Yad Vashem inden de endeligt bliver uploadede til databasen.

Denne type samling af data, falder ind under begrebet crowdsourcing³², som i de senere år er blevet flittigt brugt til andre projekter indenfor historieforskning. For eksempel hos Det Kongelige Bibliotek i København, hvor man har flere forskellige projekter i gang sideløbende, hvor et stort antal frivillige løbende digitaliserer historisk materiale, ved at taste data ind i databaser. Et andet stort projekt der med fordel har benyttet sig af crowdsourcing, er *The Holocaust Atlas of Lithuania*³³, som er et stort anlagt projekt, der netop ved hjælp af crowdsourcing søger at kortlægge Holocaust relaterede begivenheder ved hjælp af HGIS og visualiseringer. Projektet er sat i værk i 2010 af Vilna Gaon State Jewish Museum og Austrian Verein Gedenkdiens, med fokus på geografiske områder, hvor masse mordene fandt sted i Litauen i perioden 1933 – 1945. Der er online adgang til Atlasset, som er et godt eksempel på crowdsourcing, men også på hvad datavisualisering kan gøre for holocaustforskning. Projektet registrerer også en DB med navne på holocaustofre i Litauen (pr. 2/9/2018 er der 20.231 registreringer). Denne database vil eventuelt kunne bruges i et fremtidigt større projekt for at finde frem til ikke-registrerede ofre i NDB.

³² Anvendelse af internettet til at effektivt at outsource en aktivitet til en stor gruppe mennesker og trække på deres viden og kreativitet. <http://www.informationsordbogen.dk/search.php?text=crowdsourcing&type=all> set kl.2/7/18 9,30

³³ <http://www.holocaustatlas.lt/EN/#project/> set 2/9/18 kl. 10

- II. **Historiske informationer og dokumenter fra eksisterende arkiver**, som inkluderer korrespondance, lister, optegnelser og dokumentation fra både Nazisternes og de allierede landes myndigheder samt fra diverse organisationers. Dokumentationen håndteres, tjekkes af personale ved Yad Vashem og data indtastes i databasen. Sammenkørsler af registre og allerede opbyggede databaser foretages ved hjælp af API, udviklet til samme formål.
- III. **Lokale mindeprojekter**, - som for eksempel *Holocaust Atlas of Lithuania*, der finder sted over hele verden inden for Jødiske grupperinger og menigheder. Også disse eventuelle digitale databaser, køres sammen med YadVashems databaser.

I dag kører databaserne på YadVashem med komplicerede algoritmer, så man kan søge på navn og samtidig hente andre data frem om offeret, på tværs af de forskellige databaserne (Sagi, 2017). Yad VaShem er i dag verdenscenter for dokumentation for-, forskning i-, uddannelse omkring- og til at mindes Holocaust.³⁴

Yad VaShem har store forskerteams til rådighed, i samarbejde med teams i hele verden forsker intenst i Holocaust. Institutionen har adskillige digitale samlinger at arbejde med. Centeret vedligeholder blandt andet de syv forskellige online databaser, som alle kan tilgås online fra Yad VaShems hjemmeside, der fungerer som indgangsportal, blandt dem NDB.

Da Yad VaShem netop fokuserer på innovative teknologier og da NDB er en meget stor database, bliver NDB brugt i flere videnskabelige undersøgelser af, hvordan de forskellige digitale IT-teknologier arbejder, og hvordan de kan forbedres (Sagi, 2017, Aderet 2017, Brazzo 2018). NDB har pt. ca. 4,3 millioner separate individuelle ofres navne registrerede. Databasen er under konstant udvikling, både i omfang, men også i forhold til brugervenlighed og optimering af søgeværktøjer. NDB kører pt. online med en Beta version 109.2. Da denne Beta netop er udkommet, er det den forrige Beta (version 109.1) der er brugt i opgaven til at hente antal ofre for de specifikke byer i Polen til sammenligning med Census 1931.

³⁴ <http://www.yadvashem.org/yv/en/about/> set 2/7/18 kl. 9,15

Atlas of the Holocaust (Herefter kaldt **Census 1931**), skrevet af Sir Martin Gilbert, honorary Fellow of Merton College, University of Oxford og er baseret på data fra *Den Polske census 1931* (Gilbert, 2009).

Atlas of the Holocaust udkom første gang i 1982 og er en ældre bog der er udkommet i mange oplag i fire nyere udgaver. Bogen indeholder en kronologisk gennemgang af den geografiske side af Holocaust, rigt illustreret med geografiske kort. De fleste kort er håndtegnede, i en kvalitet som er god nok til at kunne bruge i denne sammenhæng, hvor de endelige visualiseringer bliver vist på meget nøjagtige digitale kort.

Jeg har valgt Gilberts kort som data til at stille overfor NDB, blandt andet fordi, de er baseret på en statslig Polsk folkeoptælling fra 1931, jeg har haft mulighed for at verificeret ved at se på tallene fra de originale optællinger, som kan tilgås fra Internettet i Pdf-filer³⁵. For at verificere data høstet fra Atlas of the Holocaust er der ved stikprøver, og sammen med en polsktalende tolk³⁶, tjekket op på om tallene stemmer overens med den originale Polske folkeoptælling fra 1931³⁷. Ved gennemgang af og sammenligning med de originale optællinger viser det sig, at tallene stemmer overens med de tal som Gilbert har brugt, bortset fra i nogle få tilfælde (bilag 1) Disse er rettet til data fra Census1931, da jeg vurderer disse for at være korrekte. Jeg har herved verificeret Gilberts Data. Der er her brugt den 4. udgave af bogen. Bogen er specielt valgt som kilde, da den giver et godt overblik over ikke bare tal som i den originale folkeoptælling, men også de geografiske placeringer. I øvrigt et godt eksempel på, at visuel præsentation kan gøre et kompliceret emne mere overskueligt og nemmere at forstå. Bogen indeholder primært mange håndtegnede kort med tal og data på ofre gennem hele perioden. Bogen er (desværre) ikke digitaliseret, men kan bruges i denne opgave, da der arbejdes med gruppeniveau og ikke individniveau.

Dette uddybes nedenfor i metodeafsnittet.

Der er i opgaven brugt tal fra kort nr. 28 side 19 der viser en opsummering af antal jødiske indbyggere på byplan, i de byer der på tidspunktet i 1939 lige inden krigens udbrud, havde 12.000

³⁵ file:///C:/Users/Imgmi/Dropbox/IVA/10%20semester%20Speciale/Cencus_Woj.kieleckie-Polska_spis_powszechny_1931.pdf

³⁶ Ania Czarnecka, - tidligere studerende på Uniwersytet im. A. Mickiewicza w Poznaniu.

³⁷

https://web.archive.org/web/20140317212240/http://statlibr.stat.gov.pl/exlibris/aleph/a18_1/apache_media/VUNVGMLANSCOQF_GYHCN3VDLK12A9U5.pdf set 2/7/16 kl. 14,15

eller flere jødiske indbyggere. (Gilbert, 2009 side 19). Dette datasæt vil fremadrettet blive kaldt for Census 1931.

Efter at have arbejdet med Census 1931 og NDB, er det blevet besluttet også at bruge visualiseringer fra *The 1939 German “Minority Census” Database*, da data er organiseret under samme metadata som de to andre databaser, og fordi det er interessant, at se om der vil være forskelle på, hvor mange ofre der er registreret i de to lande.

Det tredje datasæt der derfor bliver visualiseret og analyseret er, som tidligere nævnt, *The 1939 German “Minority Census” Database*³⁸ (herefter kaldet GMC).

Eftersom jeg fandt det interessant og relevant at se forskellen på antal registrerede ofre mellem Polen og Tyskland senere i forløbet med visualiseringerne, har jeg tilføjet *The 1939 German “Minority Census” Database*. Databasen har været brugt i forbindelse med et projekt - *The International Commission on Holocaust Era Insurance Claims*³⁹ - hvor man i perioden 1998 til 2006 har haft fokus på at finde ofre for Holocaust i Tyskland – for at kunne udbetale erstatning til overlevende holocaustofre.

Databasen – GMC - ligger online, og er organiseret med samme metadata, som jeg bruger i de andre datasæt. Data er indlæst i Excel og holdt op mod Yad VaShems database, med præcis samme fremgangsmåde som data fra Census 1931.

Visualiseringer fra GMC vil indgå i analysen og blive sammenlignet med de to ovenstående datasæt.

Udvælgelse af data som skal ligge til grund for visualiseringer, er helt afgørende for kvaliteten af analyseresultaterne (Gray 2016, Friendly 2008, Drucker 2014, Kudym 2018). I et større projekt end dette, baseret på data med individuelle navne og adresser, vil der være behov for at finde nye datakilder, hvor disse metadata er inkluderet. Her kan blandt andet henvises til projektet *Mapping the Lives* og andre.

³⁸ The 1939 German “Minority Census” Database: <https://www.census.tracingthepast.org/index.php/en/minority-census/census-database> set 23/6/2018 kl. 10,30

³⁹ <https://icheic.ushmm.org/> set d. 5/7/18 kl. 12,30

Selve opbygning af databaserne er også en vigtig del af processen, specielt vigtigt er det at dokumentere, hvordan og med hvilket software data er rensset.

Dataudrensning eller ”Data Wash” er processen med at finde og korrigere eller fjerne unøjagtige optegnelser fra en tabel eller en database, og refererer til identifikation af ufuldstændige, ukorrekte, unøjagtige eller irrelevante dele af dataene. Dataudrensning kan udføres manuelt eller ved hjælp af software.

I denne opgave, hvor der er tale om forholdsvis små datasæt, er dataudrensningen primært foretaget manuelt, men også programmet Open Refine⁴⁰ er brugt.

Gennem hele denne proces, tages der konstant subjektiv stilling til, hvad visualiseringen skal indeholde, og ikke indeholde. Mere om dette i diskussionen senere.

I arbejdet med Census 1931, er jeg startet med at oprette en database i Excel ved manuelt at taste tallene fra kort nr. 28 (Gilbert 2009 s. 19) som er tal fra Polsk folketælling 1931.

Metadata er:

- Byens navn med Latinske bogstaver
- Antal indbyggere (mænd, kvinder og børn er lagt sammen) Tallene er opsummeringer på byplan, i de byer der på tidspunktet havde 12.000 eller flere jødiske indbyggere. (Gilbert, 2009 side 19).
- Geografiske data (længde og breddegrad – hentet med et online Google Maps værktøj⁴¹)
- Ud fra disse data vil jeg visualisere på gruppeniveau; dvs. ikke på individuelt person niveau, men i grupper af ofre; mænd, kvinder og børn fra samme by/område.

Alle tre datasæt vil herefter blive organiseret under samme metadata.

3.4.2 Software

Der er i opgaven testet forskelligt software:

⁴⁰ <http://openrefine.org/> set d. 5/7/18 kl. 11,35

⁴¹ <https://www.latlong.net/> set d. 5/7/18 kl. 11,30

Til opbygning af database er SQL light, Microsoft Access og Excel blevet testet. Det er valgt her at bruge Excel, som passer bedst til de forholdsvis små datasæt. Til mindre databaser er Excel et meget brugervenligt værktøj, som samtidig er fleksibelt og har rigtig mange funktioner – også til visualisering⁴². Da denne opgave bygger på antal personer der allerede er samlet i totaler pr. by, er der her tale om en ganske lille database og derfor er Excel netop valgt. Excel har også den fordel, at mange andre programmer har mulighed for at importere data fra Excel, uden at der skal skrives nye API'er.

Til visualiseringerne er programmet Tableau⁴³ blevet brugt flittigt. Det har vist sig, efter at jeg blandt andre har testet Q-Gis, Excels eget geo-visualiseringsværktøj – som i øvrigt er godt til at lave heatmapping og andre visualiseringsværktøjer på data - at netop Tableau har været et meget effektivt værktøj til visualisering af Geodata. Tableau har et brugervenligt interface, er gearet til at snakke med mange forskellige databaser, har mange online vejledninger og visualiseringerne kan tilgås online fra Tableau Public.

Data er blevet administreret og organiseret i Excel og læst ind i Tableau. Tableau arbejder med live data og alle ændringer der evt. bliver foretaget i databasen, bliver automatisk læst ind i Tableau og ændrer visualiseringerne således, at de hele tiden er opdaterede.

Tableau har den store fordel, at programmet kan håndtere enorme datamængder og kan med fordel bruges til visualisering af Big Data også. Det vil dog kræve, at der er store og hurtige mængder computerkraft til rådighed, og at data for eksempel håndteres i Hadoop⁴⁴ – et Open Source-program som er skalerbart og som kan håndtere Big Data fra mange forskellige datakilder. Denne kombination vil være et oplagt valg, i et Geo-spatial projekt (Szewranski, 2017), der i stedet for at operere med samlede grupper af holocaustofre, arbejder med data på navne- og adresseniveau – et projekt hvor datamængden vil være enorm. Der vil i et projekt af denne art, som skal ses som et Big Data projekt, skulle bruges helt andre værktøjer og software end der er brugt her.


⁴² Der er blandt andet testet en heapmapping (bilag 5) på datasættene. Denne type visualisering viser sig dog ikke at være egnet til netop dette projekt, da der opereres på gruppeniveau. Det vil evt. være en bedre type for arbejde med individ niveau

⁴³ <https://www.tableau.com/> 2/6/2018 9,35

⁴⁴ <http://hadoop.apache.org/> set 2/6/2018 9,30

3.4.3 Setup

Projektet er baseret på grupperinger af borgere/ofre på byplan. -Der opereres med 3 forskellige farver af de respektive data:

Blå runde bobler/ringe er data fra Atlas of the Holocaust kaldet Census 1931 

Røde gennemsigtige cirkler/ringe eller bobler er data fra NDB 

Grønne runde bobler/ringe er data fra German Minority Census 1939 GMC 

Farverne er valgt, da det er farver der ofte bruges i forbindelse med Holocaust og jeg har før brugt disse i visualiseringer. Der findes en løbende debat i forskerkredse, om hvordan man af etiske årsager skal forholde til film, kunst og geografiske kort der beskæftiger sig med Holocaust formidling. Også hvilke farver der kan regnes for passende, hvor kraftige må de være og lignende (Knowles, Cole & Giordano, 2014) debatteres fortsat.

De henholdsvis røde, blå og grønne markeringer på det geografiske kort illustrerer **summen** af antal henholdsvis ofre (NDB) og indbyggere (Census 1931 og GMC). Det understreges her, at der ikke er tale om samme tidsinterval. I denne undersøgelse, giver det ingen mening at tilføje tid, da det er givet, at der er tidsforskel på Census 1931, GMC og NDB. Der er fokus på antal og de geografiske områder - så ikke tid og sted, men sted og antal. Vigtigheden for tid generelt i DVIZ, vil blive diskuteret senere.

” One of the most unusual aspects of our approach to the Holocaust is our conviction that spatial analysis and geovisualization can complement and help specify the humanistic understanding of space and place by exploring and quantifying relationships among things and people to discover and visualize spatial patterns of activity”

Knowles, A. K., Cole, T., Giordano, A. (2014) *Geographies of the Holocaust - Series: The Spatial Humanities* - Copyright Date: 2014 - Published by: Indiana University Press. Side 4.

De 29 byer som er visualiseret i opgaven er:

- Bedzin
- Bialystok
- Kolomyja
- Kowel
- Rowne
- Rzeszow

- Brest
- Chelm
- Cracow
- Czestochowa
- Drohobycz
- Grodno
- Kalisz
- Kielce
- Lodz
- Lublin
- Luck
- Lvov
- Miedzyrzec
- Pinsk
- Przemysl
- Radom
- Sidlce
- Sosnowiec
- Stanislawow
- Tarnopol
- Tarnow
- Vilna
- Warszawa

Ved at søge i NDB over registrerede ofre fra Holocaust fra samme byer repræsenteret i Census 1931 og GMC, har jeg høstet antal ofre fra NDB. Der er indhentet accept fra YadVashems afdeling for forskning til brug af data fra YadVashem databasen⁴⁵.

Der er i den online database foretaget udvidet søgning. I søgefeltet for **Place**, er der fire muligheder:

- *Place of birth*
- *Before the war*
- *During the war*
- *Place of death*

NDBs database er sat op til, i søgefeltet *Place* at kunne søge på synonymer, bogstavering og fonetisk søgning.

Her er de ovennævnte bynavne tastet ind i søgefeltet **Place - before the war**, for at få en tidsramme så tæt på Census 1931 og GMC som muligt. Bynavnet er stavet med samme latinske bogstaver som er oversat fra Census 1931 og GMC. Alle andre søgeparametre er = 0. Jeg er herved kommet frem til samlet antal indbyggere før krigen og som pr. d. 27/3/2018⁴⁶ er registreret i NDB.

Søgningen er gentaget for alle de ovennævnte byer.

Jeg har testet mange forskellige typer for visualiseringer; heat-maps, bobler og andre, for at finde de der virker som bedst egnede til at vise overlap i antal for byerne.

⁴⁵ Eliot Nidam Orvieto, Head, Academic Affairs Section
International Institute for Holocaust Research, Yad Vashem The World's Holocaust Remembrance Center
POB 3477, Jerusalem 9103401, Israel, Tel: 972-2-6443480

⁴⁶ Den dato, hvor søgningerne til det datasæt der er brugt i visualiseringerne, er foretaget.

De første visualiseringer med de forskellige datasæt blev lavet i Google Maps (Illustration 1), da det er et meget let tilgængeligt værktøj, blandt andet til at afprøve data. Værktøjet er dog sat op til at visualisere specifikke udvalgte parametre, og er for rigtigt et software til at kunne bruges til andet end validering i denne opgave.

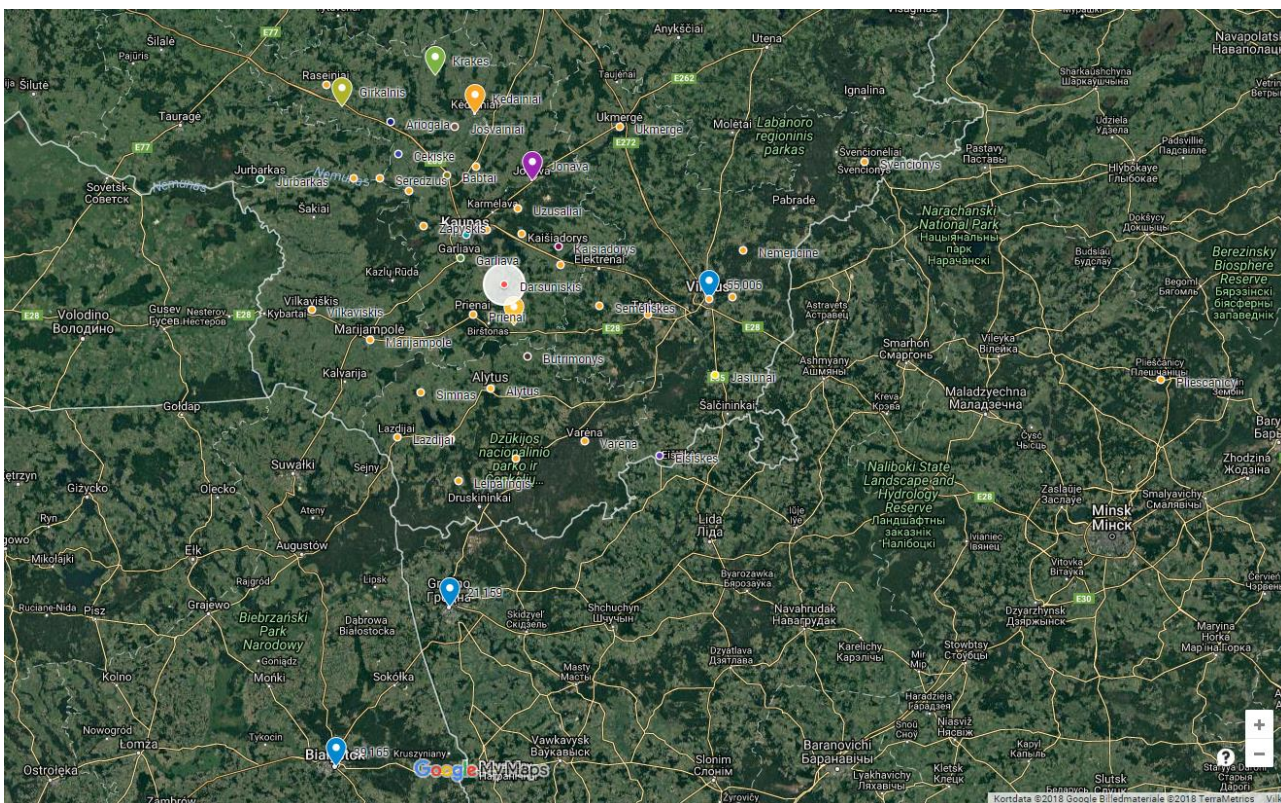


Illustration 2. Google Maps visualisering af henholdsvis Jäger Rapporten og NDB. Egen illustration

Ud fra de udvalgte visualiseringer, vil jeg argumentere for, at visualiseringer kan være en pegepind i forhold til at finde frem til ny viden om Holocaust og eventuelt endnu ikke-registrerede ofre i disse specifikke områder.

Der vil løbende blive diskuteret hvilke fordele og ulemper vi skal være opmærksomme på i arbejdet med datavisualiseringer som historisk metode. Der vil samtidig blive afprøvet forskellige værktøjer og metoder for visualisering, og det vil blive diskuteret løbende, hvilke er bedst egnede til netop dette projekt – hvorfor eller hvorfor ikke.

4.0 Visualiseringer

Der er her vist forskellige eksempler og typer af visualiseringer. Da det er vigtigt at forstå analysen, er der også vist få eksempler på visualiseringer, der ikke har med emnet Holocaust at gøre, og som

ikke er egne visualiseringer. De viser dog visualiseringstyper, som er omtalt i og som er vigtige for forståelse af datavisualisering. Egne visualiseringer er alle udvalgt mellem mange udførte, til at bruge som grundlag for analyse.

*“Visualizations and diagrams depend on the same basic graphic principles as other visual sign systems: the rationalization of a surface (setting an area or space apart so that it can sustain signification), the distinction of figure and ground (as elements of a co-dependent relation of forces and tensions in a graphical field), and the delimitation of the domain of visual elements so that they function as a relational system (framing or putting them in relation to a shared reference). Without these basic principles, no graphical system can work.” - Drucker. J. (2014) *Graphesis – Visual Forms of Knowledge Production*. Cambridge Mass. Harvard University Press.*

Dette citat af Johanna Drucker (2014) forklarer kort, hvilke grundlæggende principper visualiseringer bygger på. Citatet siger dog ikke noget om hvad DVIZ kan. Visualiseringer er en måde at formidle noget på. En måde, hvorpå man formidler data ved at præsentere disse i et udtryk, hvor beskuere har fælles referencerammer og derved får mulighed for at ”læse” tal eller andre data, selvom det er et stort datasæt.

Illustrationen fra Excel der, selvom den er simpel, lever op til definitionen på en visualisering. Excel er et kendt og meget anvendt regneark, som også kan bruges til hurtige og ikke komplicerede visualiseringer (Yau, 2014).

4.1 Klassisk Graf i Excel

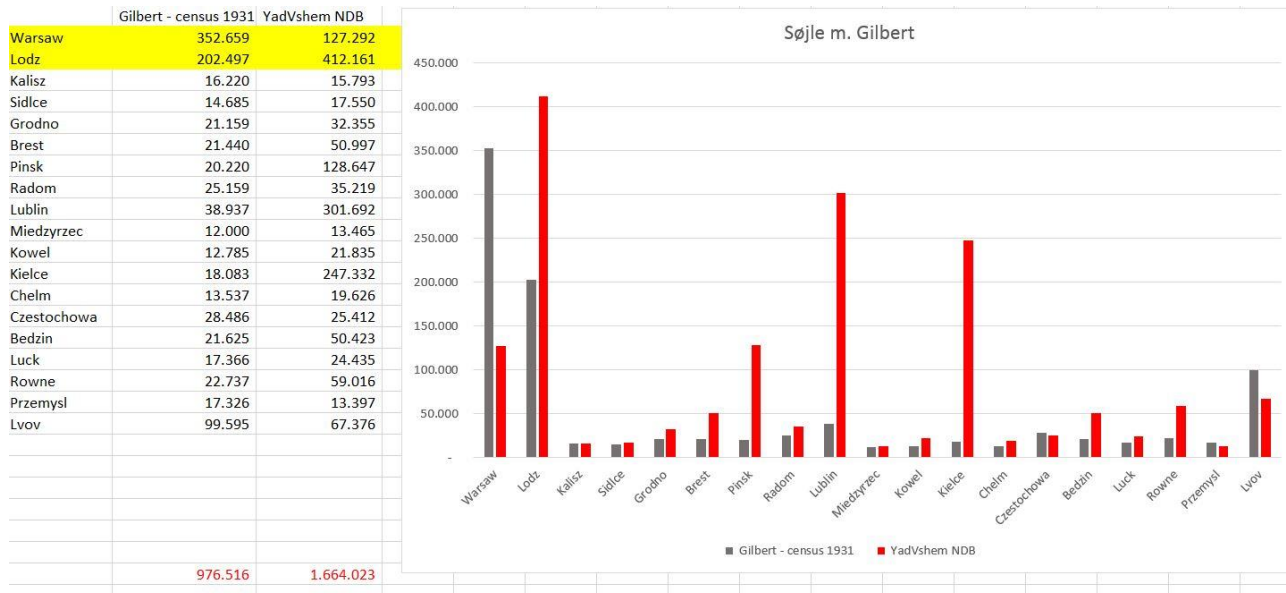


Illustration 2 – standard grafvisning af data. (egen visualisering i Excel)

Visualisering ovenfor (Ill2.) er en klassisk graf-visualisering i Excel.

Excelgrafer er gode visualiseringer der har været brugt i mange år, og som bliver brugt indenfor mange fagområder. Excel giver let adgang til at sammenligne data, som er grundlæggende i al visualisering (Yau, 2013). Her repræsenterer de røde farver NDB og de grå farver Census 1931. Netop denne type visualisering er udmærket til klart at vise forskellige antal *n*. Denne visualisering i opgaven er ikke fyldestgørende, da der her er fokus på geografiske lokationer.

4.2 Excel til produktion af choropleth kort

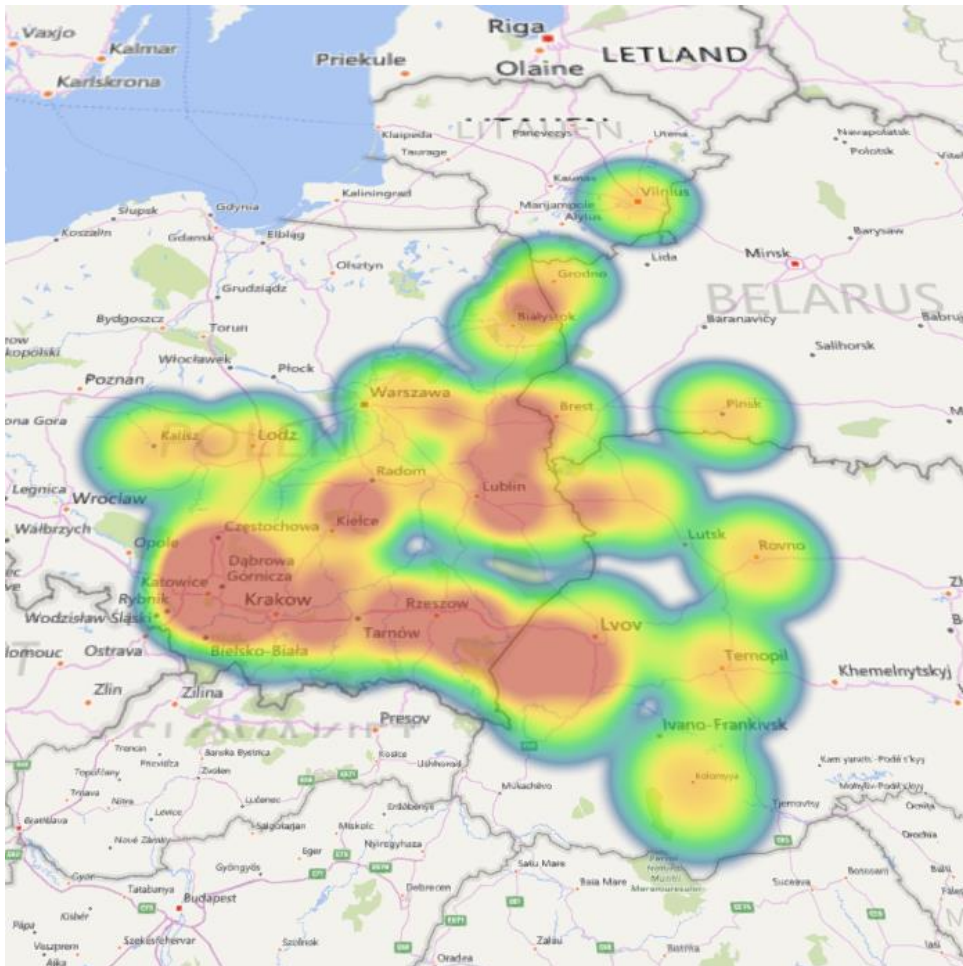
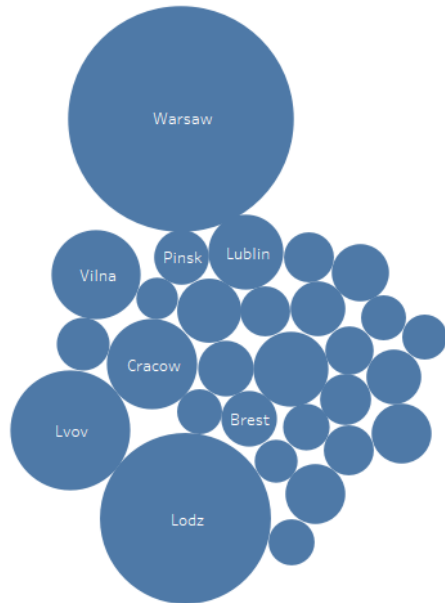


Illustration 3. De to databaser vist som choropleth maps (egen visualisering i Excel maps)

Ovenfor er vist en choropleth kort af de to databaser NDB og Census1931, foretaget med Excels eget værktøj til heatmapping og choropleth kort, (Illustration 3) kombineret med Excel World. Denne type visualisering hvor en heatmapping er kombineret med Geo-data, har, mod forventning, vist sig ikke at være egnet til opgaven her. Det skyldes primært, at der bliver arbejdet med data på gruppe-niveau. I et større projekt med store datasæt på individ-niveau, vil heatmapping være en spændende type at arbejde med. Det kræver dog, at det software som bruges, tilbyder muligheder for selv at stille på væsentligt flere parametre, end Excel Maps tilbyder. Heatmaps bruges ofte i forbindelse med visualiseringer af BD, hvor farvernes tæthed gør det lettere at vise koncentrationer af mange forskellige variabler og derved analysere store datamængder. Det er ikke nødvendigvis på

baggrund af geografiske kort, men bliver også brugt til at stille data overfor hinanden, som jeg har gjort i illustration 5.

Gilbert



NDB

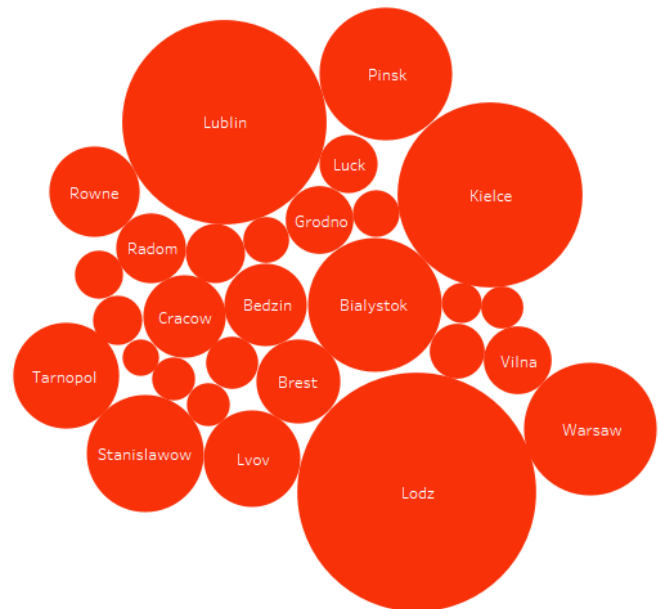


Illustration 4. De to databaser vist som heatmapping - bobler ved siden af hinanden (egen visualisering i Tableau)

4.3 Tableau til produktion af choropleth maps

Tableau er et ganske avanceret visualiseringsprogram, der bruges ofte til Business Intelligence (BI), indenfor den finansielle verden (Yau 2013, Szewrań 2017). Programmet er godt til at lave interaktive visualiseringer, og selvom det er et dyrt software (som man i øvrigt som studerende eller underviser har gratis adgang til), kan man vælge at publicere sine visualiseringer i Tableau Public, hvor alle så har adgang til disse.

Jeg har valgt primært at bruge Tableau til mine visualiseringer.

I Illustration 4 er data vist i en simpelt heatmapping som bobler, hvor hver boble repræsenterer en by og størrelsen angiver antal indbyggere/ofre - blå for Census 1931 og rød for NDB. Der er her ikke brugt graduering i farverne. Ved at se de to visualiseringer ved siden af hinanden, opnår man ikke det der umiddelbart var meningen; at se i hvilke geografiske områder der har fundet frem til flest holocaustofre.

Denne type visualiseringer bruges ofte til at vise forskel på størrelser uafhængigt af fysiske rammer. Visualiseringen kan, som Excelgrafene give en indikation på forskellene mellem NDB og Census 1931. Denne visualisering er dog ikke umiddelbart egnet til formålet med opgaven her, da forskellen i antal registrerede ofre og ikke-registrerede ofre ikke ses i geografiske rammer.

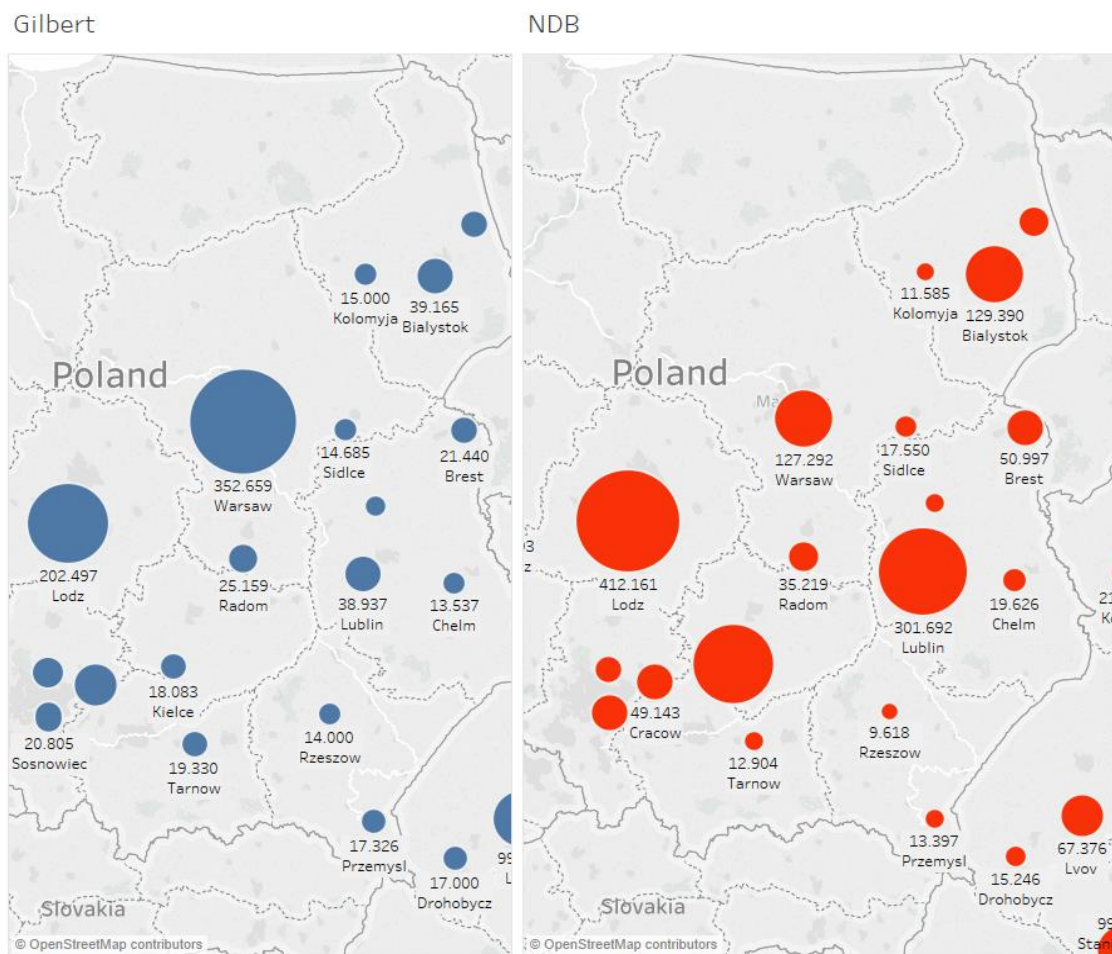


Illustration 5. De to databaser vist i choropleth mapping med geografisk placering ved hjælp af bobler (egen visualisering i tableau)

I Illustration 5 er data læst ind i Tableau og der er tilføjet geografiske data. Der er brugt Tableaus egne landekoder og der er valgt samme gennemsigtighed og størrelsesforhold til begge databaser. Her kan man se eventuelle mønstre i forhold til geografien og størrelsesforholdet mellem registrerede og ikke-registrerede ofre. Denne visualisering, er ikke egnet til at vise forskelle, mønstre og korrelationer, da de to kort er visualiseret ved siden af hinanden, og jeg har derfor fravalgt denne visualisering som grundlag for analysen.

Da der er tale om visualiseringer på gruppe-niveau, har jeg valgt at medtage tallene for ofrene/indbyggerne. Da kortene ikke er lagt sammen, er dette en mulighed, og kan evt. bruges til at verificere visuelle tendenser i det følgende kort.

Denne visualisering er grundlæggende for næste visualisering, hvor de to kort bliver lagt sammen med fælles akse (Dual Axis), så forskellene mellem NDB og Census 1931 fremtræder tydeligere.

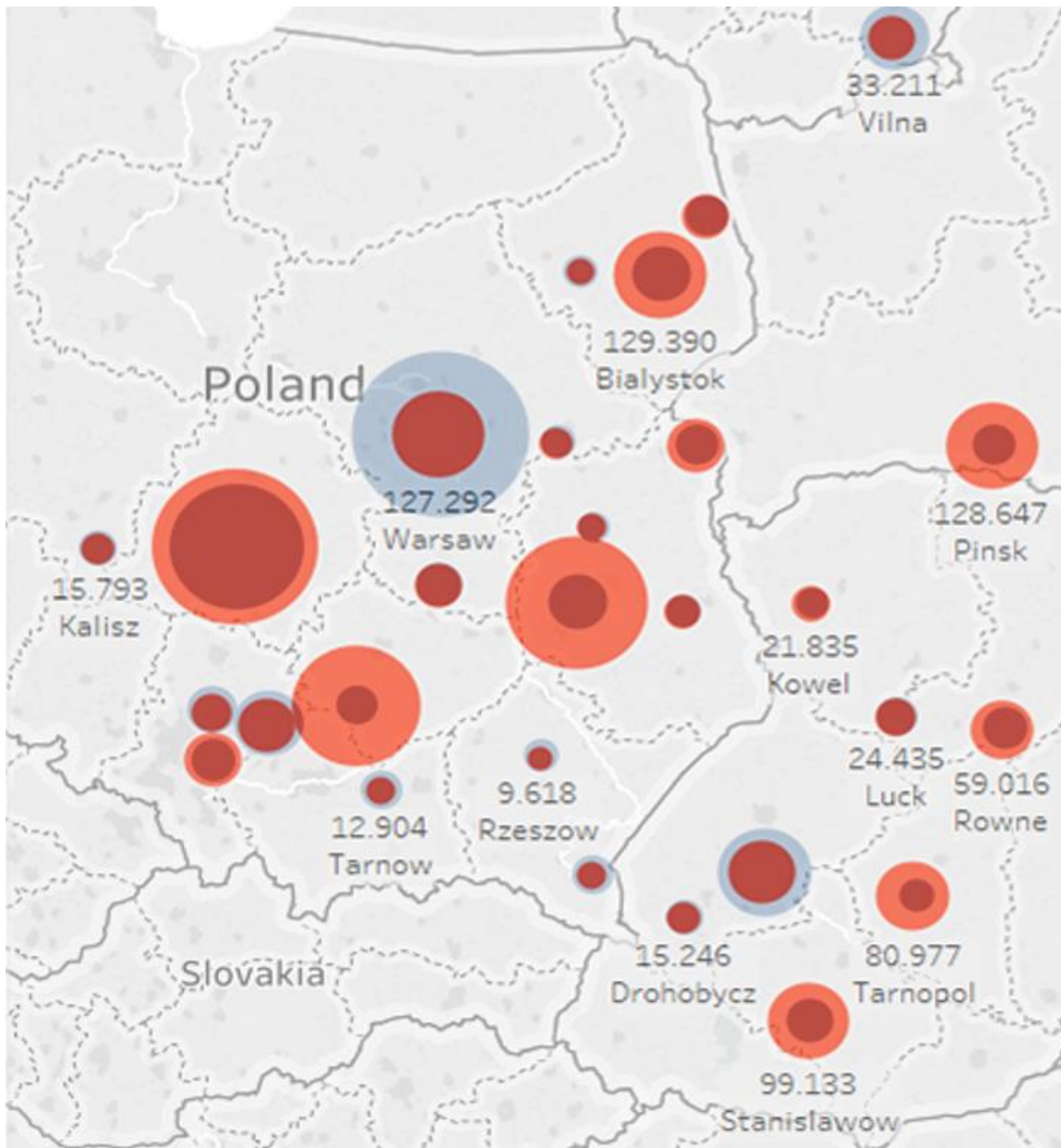


Illustration 6. De to databaser lagt sammen med Dual Axis. (egen visualisering i Tableau)

Ser vi på Illustration 6, nærmer vi os det kort, som mit mockup (Illustration 1) visualiserede i starten af opgaven. Her er boblerne lagt ovenpå hinanden, så man visuelt kan se forskellen på

boblernes størrelse. Visualiseringen er dog ikke helt tydelig, da farverne blandes i denne visualisering og derfor ikke giver et klart indtryk af forskellene på data.

Da det desuden ikke kan aflæses hvilke database tallene tilhører under boblerne, er illustrationen/visualiseringen af den grund stadig ikke fyldestgørende nok.

Denne visualisering viser både forskelle mellem grupperingerne og kan samtidig vise mønstre og relationer/sammenfald i data. Korrelationer, som er forholdet mellem de viste variabler, træder her tydeligt frem som for eksempel ses i Illustration 6, hvor der er ganske få blå bobler, som er fremtrædende. Visualiseringen her er dog en smule misvisende, da farverne ændrer sig, når de bliver blandet sammen. Det fremstår ikke klart, hvor stor forskel der er på de to datasæt.

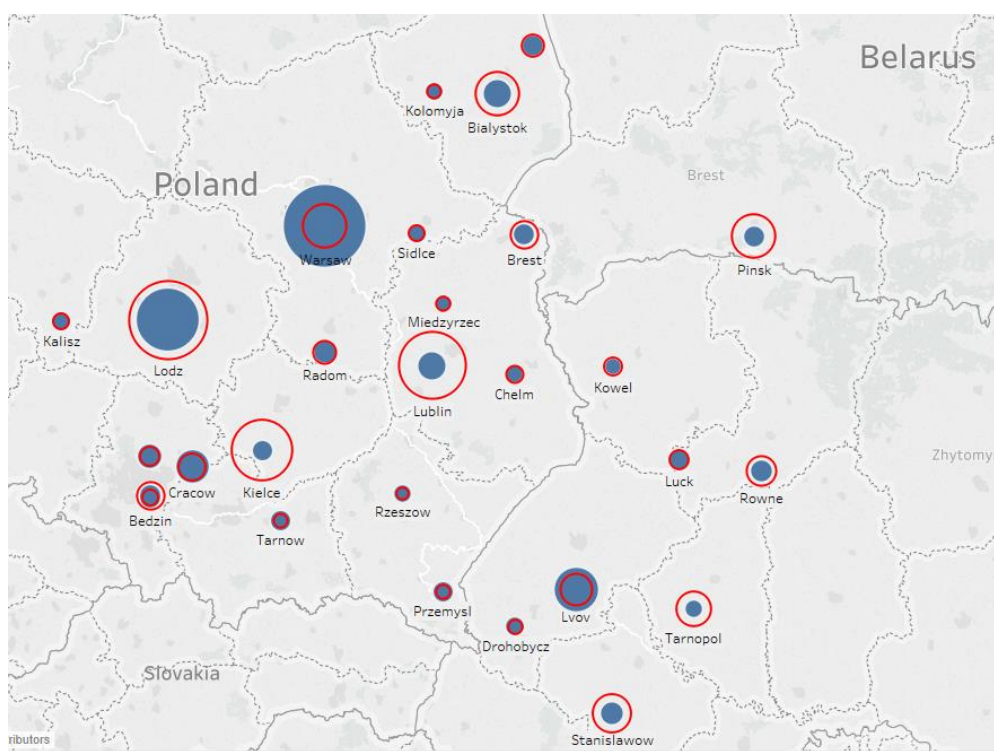


Illustration 7. Endelig visualisering: NDB og PC 1931 kombineret geografisk og vist med blå bobler for PC 1931 og røde ringe for NDB (egen visualisering i Tableau)

Før det endelige resultat på den næste visualisering (Illustration 7), er der afprøvet forskellige muligheder for at opnå det bedst mulige resultat. I den endelig udgave af visualiseringen er fyldet fra de røde bobler fjernet, så der kun er en rød ring tilbage. Hele feltet indenfor denne røde ring viser det antal ofre der er registreret i NDB og derved fremtræder forskellighederne tydeligere.

Det er denne visualisering, jeg har valgt som værende den endelige, og som er grundlæggende for analysen.

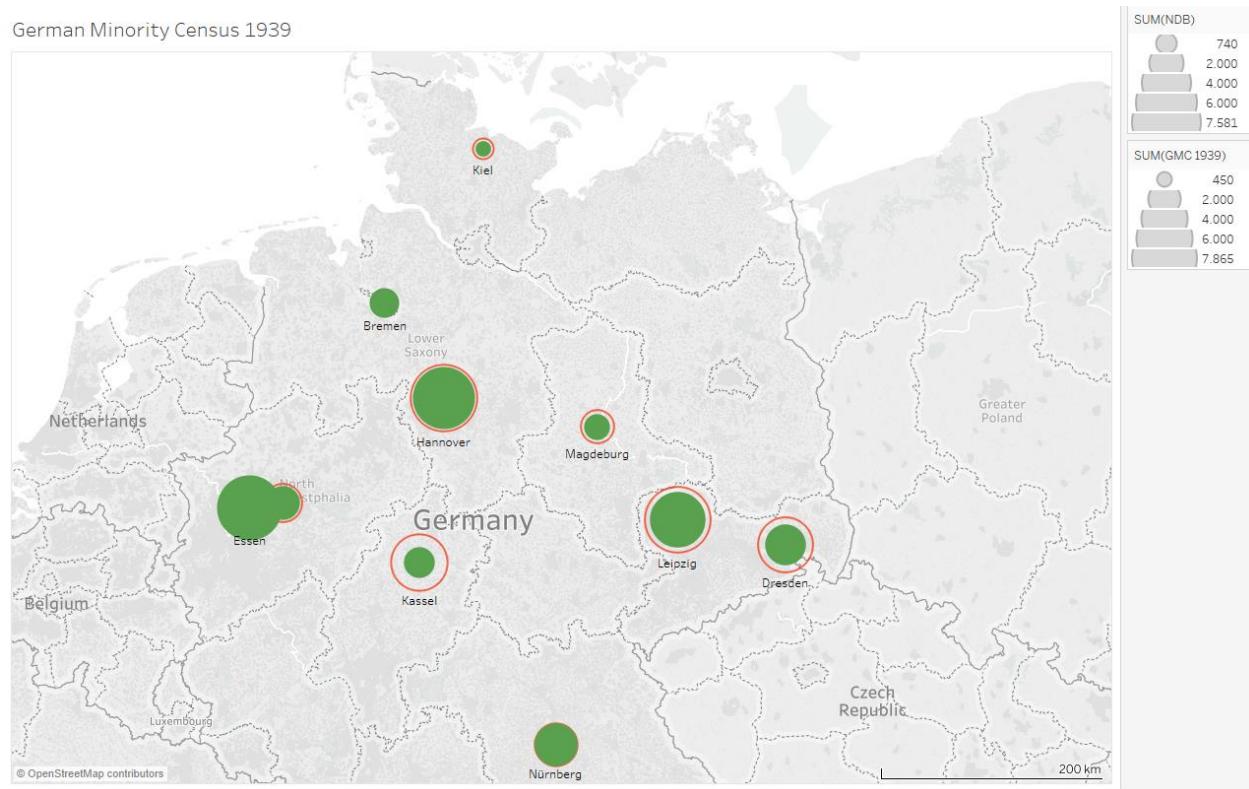


Illustration 8. NDB og GMC 1939 kombineret geografisk og vist med grønne bobler for GMC 1939 og røde ringe for NDB (egen visualisering i Tableau)

I Illustration 8, hvor databasen fra GMC er visualiseret overfor NDB, ses, at bobler og cirkler er mere ens end i visualiseringerne fra Polen. Den viser knapt så store forskelle i antal registrerede og ikke-registrerede ofre.

Det er interessant, at mønsteret hvor NDB har flere registrerede end GMC, så her, er der også, behov for at tage forbehold for eventuel redundans i data, men også at have fokus på, at datasættene ikke kan visualiseret 1:1. Størrelsen generelt på de to datasæt GMC og Census 1931 er meget forskellige, og kan derfor være misvisende. Her må man igen have i mente, at DVIZ ikke kan stå alene, som diskuteres nedenfor.

5.0 Hvad kan udledes af visualiseringerne

Indledningsvis vil jeg gøre opmærksom på, at jeg har arbejdet med de interaktive visualiseringer, som giver mulighed for at zoome ind og ud, hvilket er en stor fordel, når visualiseringerne skal

analyseres. De kan tilgås [her](#) via min profil i Tableau Public. At bruge de interaktive visualiseringer, giver blandt andet større mulighed for at bruger selv kan zoome ind og ud for at undersøge, om der evt. kan findes mønstre i den måde boblerne henholdsvis ringene lægger sig på – og som muligvis har betydning for årsagen til at netop nogle byer eller områder har flere eller færre registrerede ofre end andre. I de interaktive visualiseringer, er der også mulighed for at brugeren selv kan vælge mellem parametre, der er lagt ind i databasen.

En god visualisering skal kunne fortælle en historie og ikke blot vise en grafik over en masse tal. (Yau, 2015). Dette gælder også i de viste visualiseringer. Det vi ser, er bobler, som hver især kan fortælle hvor mange ofre, der er registreret i området. Man kan ikke se, hvem ofrene var, men visualiseringerne fortæller alligevel en bagvedliggende historie om, det menneskelige behov der er, for at finde frem til de ofre der endnu ikke er registreret, og bag hver registrering ligge en historie om en menneskelig skæbne. Dette er måden visualiseringer kan fortælle en historie på.

*” Data visualizations are not only bright adornments to our 21st-century information environments. They embody and engender not only particular ways of **seeing**, but also ways of **knowing** and ways of **organizing** collective life in our digital age. In other words, data visualizations reflect and articulate their own particular modes of rationality, epistemology, politics, culture, and experience.”* Gray, J., Bounegru, L., Milan, S., & Ciuccarelli, P. (2016) Ways of Seeing Data- Toward a Critical Literacy for Data Visualizations as Research Objects and Research Devices. *Forthcoming in Sebastian Kubitschko and Anne Kaun (eds.) Innovative Methods in Media and Communication Research. London: Palgrave Macmillan S. side 228*

I denne analyse, som er baseret på gruppevisualiseringer, giver det ingen mening at zoome ind på vej/adresse niveau, men i et større projekt, hvor det er individuelle adresser og navne, der danner datagrundlag, vil det være en oplagt måde at arbejde på – og samtidig bruge choropleth mapping. Der er dog en udfordring ved denne type cartogrammer, da de geografiske visualiseringer ikke altid viser de helt korrekte forhold i størrelse på landsdele (Yau, 2013). Dette vil kun have mindre betydning i et sådant projekt, hvor det er forskelle, relationer, korrelationer og mønstre, som er interessante, og ikke specifikke oplysninger om antal borgere i specifikke geografiske områder.

For at illustrere hvordan en datavisualisering ser ud, når der bruges individuel-baserede data-corpora, har jeg valgt at bruge en visualisering fra *Geographies of the Holocaust*, (illustration 9) hvor der er foretaget clusteranalyser af arrestationer på jødiske borgere i Italien i 1943. Man har her valgt på det store kort, at sammenligne antal af arrestationer, med en Census der viser antal jødiske borgere i Italien i 1938 på det lille kort i højre hjørne – at vise korrelationer og sammenhold på to forskellige kort, er efter min mening ikke det bedste valg, og jeg vil mene at analyseresultaterne fra Italien ville blive bedre, hvis man havde valgt at anvende dual-Axis.

Visualiseringen viser klart, at arrestationerne foregik i alle dele af Nord- og Mellem Italien, bortset fra nogle få undtagelser som eksempelvis nord for Rom. Umiddelbart følger billedet af arrestationerne de områder der også ses med flest jødiske indbyggere på det lille kort, som er Italy Minority Census 1938. Der ligger flere clusters i visualiseringen: det fremgår, at der er flere arrestationer i de store byer, men også i byerne der ligger i umiddelbar nærhed af disse (Knowles, Cole & Giordano, 2014).

Det kunne her være interessant at kombinere disse data med data fra NDB, for at se om der eventuelt kunne findes sammenhænge med ofre og arrestationer.

En clusteranalyse er et godt redskab til at vise mønstre og sammenhænge. I visualiseringen fra Italien, ser man mønstre hvor der tættere forekomster af arrestationer. Disse er her markeret med ellipseformede afgrænsninger, så man kan se hvor områderne med mange arrestationer forekommer.

Der er ikke lavet specifikke clusteranalyser i denne opgave, da det ikke er hovedformålet, som er at fokusere på forskelle mellem registrerede og ikke-registrerede i NDB. Derudover giver clusteranalyse af data, som på forhånd er grupperet i databasen, ikke helt mening når der er udvalgt specifikke byer. Man kan i mine visualiseringer se, hvordan antallet af registrerede generelt forholder sig til hinanden geografisk, som clusters, som det er observeret, at større byer har generelt flere ikke-registrerede end registrerede ofre. Generelt mener jeg dog, at en god clusteranalyse skal baseres på store mængder, helst usortet data.

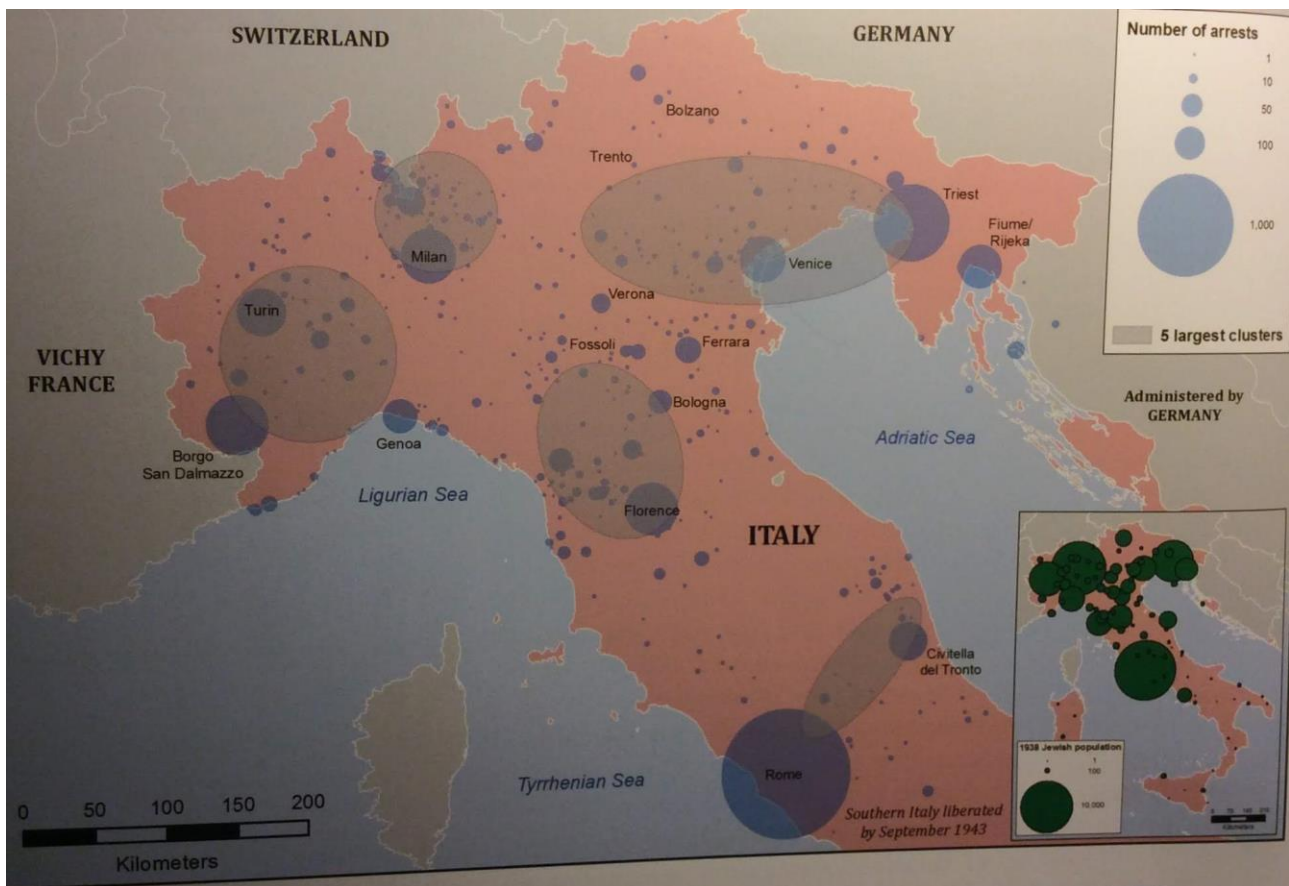


Illustration 9. Arrestations clusters sammenlignet med Italiens indbyggerantal 1938. Knowles, A. K., Cole, T., Giordano, A. (2014) *Geographies of the Holocaust - Series: The Spatial Humanities* - Copyright Date: 2014 - Published by: Indiana University Press. Side 68

Sammenligner man denne type visualisering med den gruppevisualisering jeg har foretaget med NDB og Census 1931 (illustration 7), vil man se, at clustervisualiseringen kan give langt mere information end gruppevisualiseringen. Clusteranalyser hvor store data-corpora ligger til grund for visualiseringer som i illustration 9, har vist sig, at vise mønstre om mobilitet og immobilitet, koncentration og sårbarhed, sammenfald og spredning i måden arrestationerne på de italienske jøder foregik på (Knowles, Cole & Giordano, 2014). Derudover har visualiseringerne kunne pege på at identiteten af gerningsmændene var en stor faktor i den historiske begivenhed; der var større risiko for at blive sendt til Auschwitz hvis man blev arresteret af tyskerne, end hvis man blev arresteret af italiensk politi (Knowles, Cole & Giordano, 2014). Endnu et argument for, at et større projekt end det foreliggende, med langt større datasæt, vil kunne være yderst interessant.

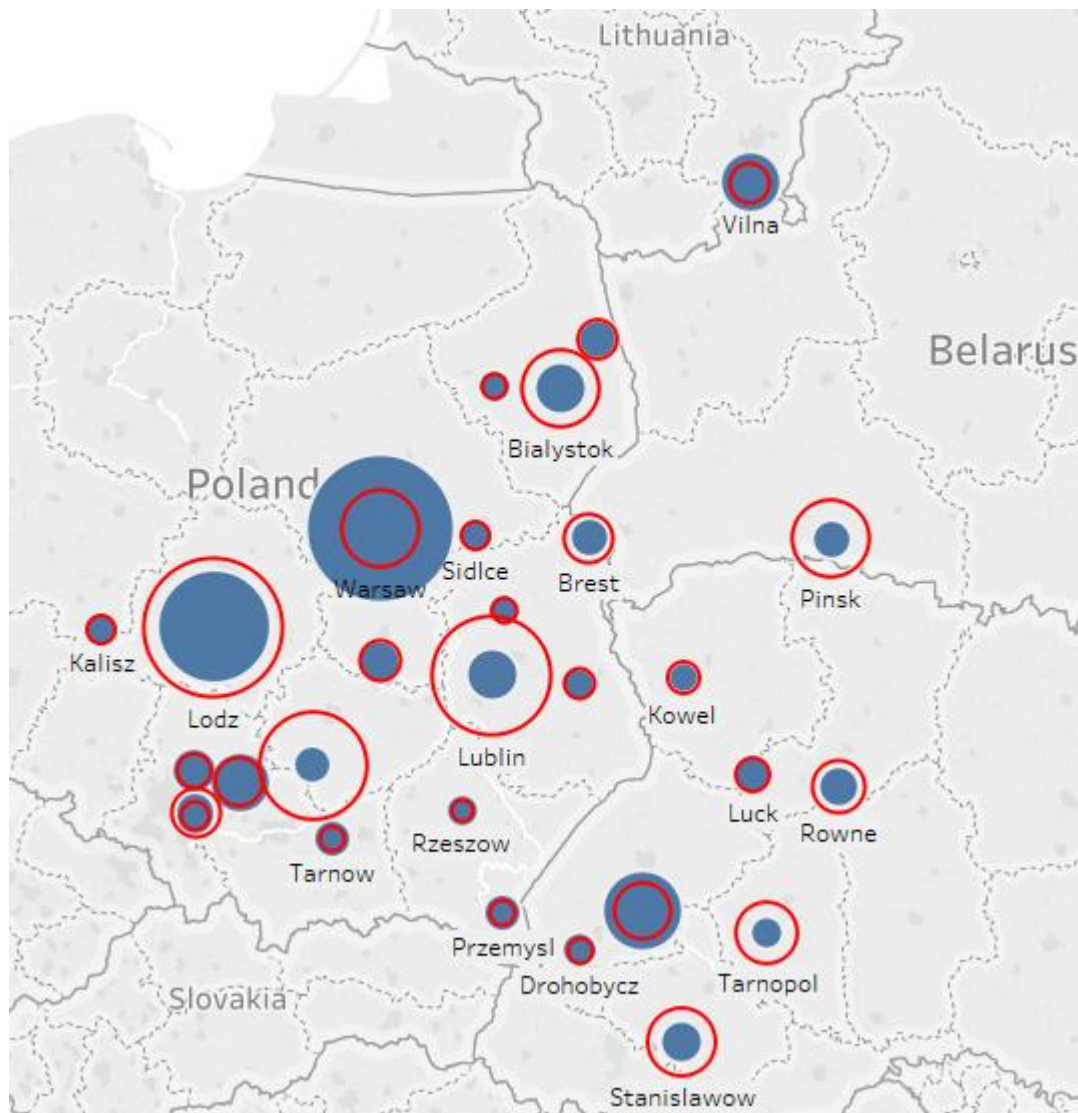


Illustration 10. NDB og Census 1931 kombineret geografisk og vist med blå bobler for Census 1931 og røde ringe for NDB (egen visualisering i Tableau)

5.1 Samlet billede - hvad springer først i øjnene?

Jeg starter med at se på det overordnede billede, ved at zoom ud på kortet til alle 29 visualiserede byer fra Census 1931 og NDB kan ses. (Illustration 10)

Det første der falder i øjnene, som det klart fremgår af den første choropleth-visualisering ovenfor, er, at der er flere byer, hvor de røde ringe (17) er større end de blå bobler (12). Det betyder, at mange byer ifølge visualiseringerne har flere registrerede ofre fra NDB end indbyggere før krigen (i 1931) - eksempelvis i byerne Lublin, Kielce, Lodz, Bedzin, Stanislawow, Tarnopol og Rowne. Et mønster der skal vise sig at være gentaget, i alle de her foretagne visualiseringer.

At der umiddelbart lader til at være flere registrerede ofre i NDB end i Census 1931 kan have mange forskellige årsager; jeg vil her kort komme ind på et par af de årsager, der kan ligge til grund i selve datasættene og dernæst stille spørgsmålstegn ved, om der kan ligge specifikke historiske begivenheder til grund for dette. Det er netop her, at man ser og – og derved kan argumentere for, at datavisualiseringer er stærkest, når der er tale om meget store datasæt og ikke som her, hvor der er visualiseret på gruppe-niveau. Jo større data-corpora der ligger til grund for visualiseringerne, jo større chancer er der for at se korrelationer og mønstre – og derved pege på områder, hvor der med fordel kan findes ikke-registrerede ofre.

Dette betyder også for denne opgave, at det må konstateres, at metoden her som bruger data på gruppe-niveau ikke egner sig til at finde frem til områder med ikke-registrerede holocaustofre.

Der kan som tidligere nævnt være flere grunde til at der ser ud til at være flere registrerede ofre i NDB end i Census 1931.

Et par af de oplagte årsager kunne være: Grundlaget for data er forskellige i de to datasæt – de er meget generelle i Census 1931, da det i Census 1931 kun er de større byer, med over 12.000 indbygger som er registrerede.

Det kan også delvis skyldes, at der opereres på gruppeniveau og at NDB's metadata sorterer de store byers omkringliggende småbyer ind under storbyer. Her må en analyse af datasættene kræves. Et tæt samarbejde med de forskere der administrerer NDB må være et krav, hvis metoden eventuelt skal bruges i et større projekt hvor data er på individ-niveau, netop til det formål at finde frem til ikke-registrerede ofre.

En tredje oplagt årsag, som ofte ses i forbindelse med både sammenkørsel af registre og indsamling af data ved hjælp af crowdsourcing, kan være redundans; at der i NDB er mange, dobbelte registreringer, hvor samme offer er registreret flere gange og tit oftere end to gange. Dette har vist sig efter at have udført nedenstående undersøgelse.

For at afgøre om redundans kunne være et problem i NDB, har jeg udført en kort undersøgelse.

Som eksempel har jeg brugt navnet på et kendt Holocaustoffer, som jeg før har set registreret i NDB. Navnet Anne Frank er blevet søgt på i NDB og resultatet viser, at mindst fire forskellige indførsler/registreringer omhandler samme offer. (se bilag 1) Denne dobbeltregistrering er et problem, man *er* opmærksomme på forekomster af redundans på Yad VaShem. Man mener dog, at denne redundans er af mindre grad og ikke betydelig (Sagi, 2017).

Man kan med disse visualiseringerne argumentere for, at problemet muligvis er større, end det hidtil har været formodet.

Denne udfordring med redundans, vil typisk være til aktuel, når der bliver brugt navne som data og specielt, når en del af data er baseret on Crowdsourcing⁴⁷, som NDB kan betegnes, da en meget stor del af registreringerne kommer fra Pages of Testimony (bilag 2), som overlevende familiemedlemmer eller bekendte har sendt ind til registrering.⁴⁸ Det er en typisk udfordring indenfor netop Holocaustforskning, hvor data er samlet på mange forskellige sprog og kørt sammen fra forskellige lister, der er fundet hele eller i fragmenter.

Det fremgår også tydeligt, at det primært er i de store og mellemstore byer, at der er flere registrerede ofre. Dette er specielt interessant, da det stiller spørgsmål ved metodens holdbarhed. Det lægger op til kritisk kvalitativ analyse og spørgsmål til historikere, der har ekspertviden om Holocaust, om hvad der kan være årsagen. Er der historiske fakta, som kan forklare hvorfor, eller skal der kigges på data?

Mange årsager til eventuelle mønstre, korrelationer, sammenfald, clusters og lignende, kunne være interessante at se på i samarbejde med en domænespecialist. Med min viden om Holocaust historiografien, vil jeg ikke kunne få øje på og se ind i disse detaljer.

En analyse af visualiseringerne peger på nogle tendenser i forhold til om metoden kan bruge til at finde frem til ikke-registrerede ofre; det er hvad jeg derfor ser efter. Desuden vil jeg argumentere for, at hvis datavisualisering skal kunne vise klare, gode og interessante tendenser af anden slags, - som for eksempel clusters og lignende, må der langt større datasæt til og mere detaljerede data.

5.2 Einsatzgruppernes betydning

Der er visse tendenser der kan pege på informationer om registreringer af ofre: På Illustration 10 ser vi netop, at specielt de små byer i området omkring Lvov, har de røde ringe indenfor den blå boble, og derfor har færre registrerede ofre i NDB end Census 1931: **Przemysl** med 17.326 registreringer i Census 1931 og 13.397 med registreringer i NDB, **Drohobycz** med 17.000 registreringer i Census

⁴⁷ Anvendelse af internettet til at effektivt at outsource en aktivitet til en stor gruppe mennesker og trække på deres viden og kreativitet <http://www.informationsordbogen.dk/search.php?text=crowdsourcing&type=all> set 19/8/2018 12,00

⁴⁸ <https://www.yadvashem.org/archive/hall-of-names/database.html> set 19/8/2018 11,52

1931 og 15.246 med registreringer i NDB og **Rzeszow** med 14.000 registreringer i Census 1931 og med 9.618 i NDB. Umiddelbart er det ikke store forskelle, men da det er forholdsvis små byer, er her tale om relativt mange mennesker.

En af de historisk betingede årsager til de store udsving i forskelle på registrerede og ikke-registrerede ofre, kan være, at mange områder der er taget med i denne visualisering (se Illustration 11), er områder hvor de nazistiske Einsatzgrupper⁴⁹ efter ordre fra Heydrich specielt fra forsommeren 1941 henrettede hele landsbyer, og hvor der kun eksisterer meget fragmenterede kilder om navne på ofrene (Stone, 2004).

Det er i netop dette område at Einsatzgruppe C (og D) opererede under invasionen af Rusland i juni 1941 (Stone 2004, Gilbert 1982, Dickmann 2001). Det tyder på, at netop den måde Einsatzgrupperne opererede på, har gjort det svært at finde frem til navne på ofre. Det var ganske kaotiske masse mord, hvorfra kun få overlevendes vidneudsagn kan kaste lys over begivenhederne. Måske netop her kunne en visualiseringsmetode gøre en forskel.

Kunne det tænkes, at Nazisterne ikke lykkedes med at holde fast i deres ellers så systematiske udryddelser af jøderne og ikke formåede at registrere ofrene, eller har de strategisk undladt beviser, af disse forfærdelige gerninger? Man var i de sidste år af krigen meget omhyggelige med at grave de tidligere fyldte massegrave op, for at brænde ligene i stedet, netop for ikke at efterlade beviser for ugerningerne.

⁴⁹ Specielle SS-grupper, som blev sendt ind i de sovjetiske områder i kølvandet på operation Barbarossa i 1941, for at likvidere jøder (Stone, 2004).



Illustration 11. Invasionen af Rusland og de forskellige Einsatzgruppers fremrykning juni 1941 - Martin Gilbert 1982. s. 52

Det er tydeligt at se, i hvilke af de små byer det vil være udbytterigt at fokusere på en nærmere, kvalitativ undersøgelse af eksempelvis de lokale arkiver og museer. Dette er dog en opgave for historikere, da sådanne undersøgelser lægger op til omhyggelig kildeanalyse og dybdegående undersøgelser på mikroniveau.

Det vil være mindre ressourcekrævende, hvis der forud for en undersøgelse, kunne foretages visualiseringer af områder, - helst på individuelt niveau – før man tog beslutning om, hvor man skal sætte ind med mikro-undersøgelser. Det er netop dette visualiseringerne kan!

I dette projekt bliver der brugt grupperinger fra byer med over 12.000 indbyggere fra Census 1931; med bedre data og større datasæt, vil der sandsynligvis kunne findes frem til nogle af de steder, hvor der stadig er mange ikke-registrerede ofre.

Der findes helt sikkert lokale Minority Censuses, som GMC i andre lande i Europa (Knowles, Cole & Giordano, 2014). Selv har jeg ved selvsyn erfaret, at et sådant findes på *Regional Institute for the History of the Liberation Movement of Friuli-Venezia Giulia* i Trieste i Italien. Her findes i arkiverne en meget omfattende census fra 1943 hvor alle jødiske borgere er registreret. Også denne

har været brugt til økonomisk at kompensere de overlevende og Holocaustofrenes efterkommere. Materialet er håndskrevet, og selvom man fra arkivets side prøver at digitalisere så meget som muligt, er det et stort arbejde, der kræver langt flere ressourcer end det lokale arkiv har til rådighed. Der kunne med fordel overvejes at bruge Crowdsourcing for at få materialet digitaliseret.

5.3 Redundans?

Jeg vil argumentere for, at ved brug af visualiseringer som forskningsmetode som i denne opgave der er baseret på mindre, grupperede datasæt, bør man være opmærksom på udfordringen med redundans i data. I visualiseringerne af Big Data med store uorganiserede datasæt hvor der viser billeder fra fugleperspektiv, og som lægger op til at vise mønstre og korrelationer mere end detaljerede informationer, egner denne type metode sig netop til Big Data analyse, frem for til mere specifikke opgaver som i her. Da mindre datasæt, jo større indflydelse har redundans på visualiseringerne.

Det kunne have været super spændende, at få tid med ind over visualiseringerne, for at se hvordan forløbet af registreringerne har udviklet sig over tid og sted. En sådan visualisering er en god mulighed for at vise hvor meget bedre en interaktiv visualisering er til at vise udvikling over tid. Vil man kunne se perioder der afspejler tendenser i holocaust diskursen i samfundet generelt? Er der perioder, hvor det har været mere effektivt at finde frem til ofre end andre? Kan man for eksempel se, at der er sket en ændring efter 1989, hvor murens fald og Sovjetunionens sammenbrud åbnede for tilgang til et væld af gemte lister og originale kilder blandt andet fra arkiver i det tidligere Øst Tyskland (Stone, 2004), så som deporteringslister og lignende?

Denne form for visualisering vil dog kræve en del mere tid end der er til rådighed i denne opgave, samt langt større og detaljerede datasæt.

I storbyerne er der generelt flere registrerede end på landet (illustration 8), - dette kan have mange forskelligartede årsager: en af disse er nok, at de jødiske borgere, blev flyttet meget rundt i perioden 1933- 1945. Data brugt fra NDB må man huske, viser opholdssted før krigen – der er ikke angivet, hvilket årstal dette indebærer. Desuden, har Census 1931 indordnet en del af de mindre landsbyer i udkanten af storbyerne.

Her er det vigtigt at huske, at visualisering som metode, kun kan pege på tendenser. Metoden kan understøtte og vejlede, men vil altid kræve en nærmere, og mere dybdgående undersøgelse af de originale kilder.

Det er også vigtigt at huske på, at i perioden 1931 og frem til krigens slutning, specielt i Tyskland og Polen har der været massive folkevandringer, dels flugt fra det Nazistiske styre, mens også flytning af borgere fra hele landsbyer og byområder først mod øst, og senere flytning til ghettoer. Det ses blandt andet, ved det kaos der opstod efter krigen, hvor mange millioner flygtninge, tidligere krigsfanger og overlevende, stod foran at skulle starte livet forfra (Knowles, Cole & Giordano 2014, Stone 2004, Kudym 2018).

Projektet *Holocaust Mobilities: Journeys Through Minsk* (Kudym, 2018), er et forholdsvis nyt projekt, hvor der er brugt HGIS og visualiseringssoftware for at vise hvordan Holocaust overlevende har flyttet rundt under og efter krigen. (Illustration 11)

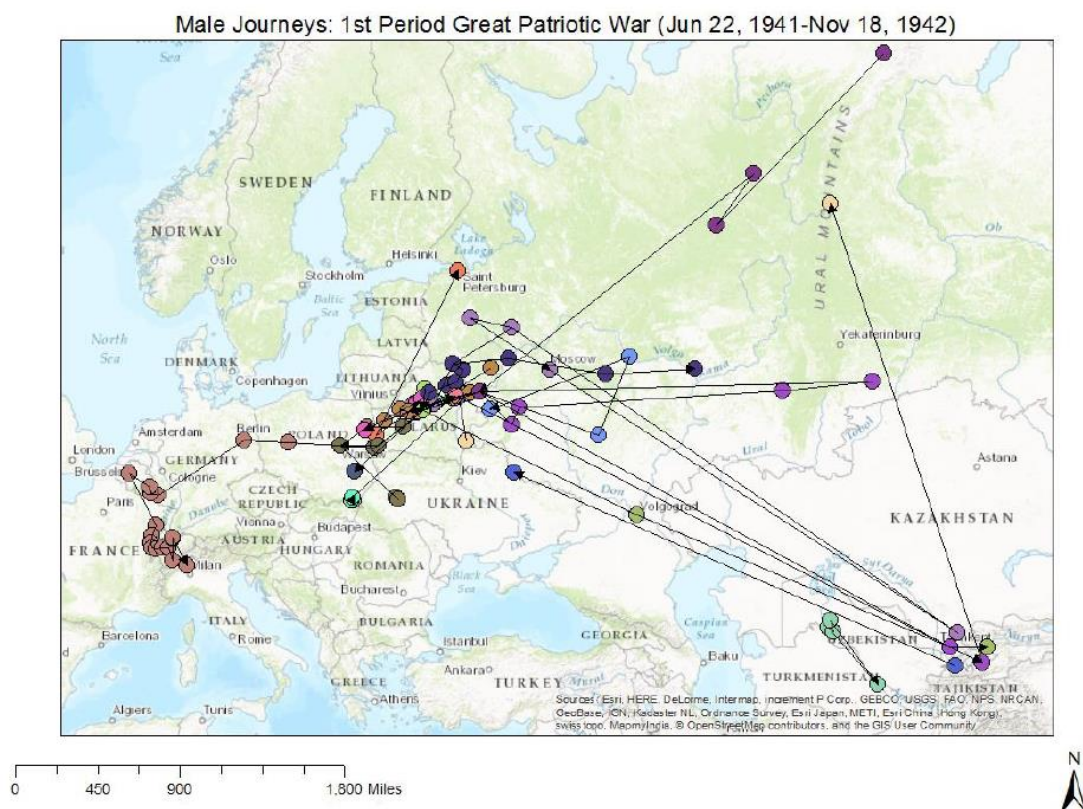


Illustration 12. Eksempel på visualisering fra migrerings/deporteringsruter fra Projektet *Holocaust Mobilities: Journeys Through Minsk*. Visualisering Kudym 2018 s. 99

Ved hjælp af ArchMap 10.5, har Kudym (2018) visualiseret 30 forskellige deporterings- flugt- og migrationsruter, vist over både tid og sted. Data er høstet fra transskriptioner af 30 audiovisuelle vidneudsagn fra 30 jødiske Holocaustoverlevende fra Østeuropa og Tyskland.

Et meget interessant projekt, som sætter fokus på, hvordan Holocaust har været årsagen til store, både tvungne og frivillige migreringer af de jødiske minoriteter, her specielt fra Rusland. Ikke

mindst interessant er det, at Kudym (2018) i sit speciale også bruger netop visualisering som en del af sin metode. Kudym (2018) har valgt at kombinere visualisering som kvantitativ metode, med narrativ analyse som kvalitativ metode. At der her er valgt at bruge begge metoder, er med til at understøtte min argumentation om, at visualiseringer som metode fungerer bedst, når det er som understøttende til andre metoder. Metoden kan ikke stå alene.

Kudym (2018) visualiseringer viser blandt andet, at rejserne der blev foretaget via ruterne sjældent gik glat og problemfrit, men snarere var en hyppigt afbrudt rejse, med adskillige ufrivillige stop undervejs, og ofte med lange perioder hvor man måtte forblive samme sted, for eksempel på grund af fængsling, perioder med tvunget arbejde eller i skjul (Kudym, 2018)

Ovennævnte projekt kan også ses som endnu et argument for, at visualiseringer på gruppeniveau som her i opgaven, ikke er tilstrækkelig nøjagtige og for uigennemsigtige til at kunne bruges reelt til at finde ikke-registrerede ofre.

Det betyder samtidig, at et større projekt, hvor databaser er baseret på individuelle adresser som eksempel, kunne give et langt mere nøjagtigt billede, selvom der i et sådant projekt, stadig må forventes en del unøjagtigheder, specielt i storbyerne, som her, dels fordi der i perioden blev flyttet meget rundt på befolkningsgrupper, og dels fordi som tidligere nævnt, at de kilder man har er meget heterogene.

Det er helt naturligt at der er tendens til redundans i NDB, da der er indsamlet data ind fra så mange forskellige kilder og hvor også metoden i høstningen er forskellig. Bare det faktum, at der i tidens løb er tilføjet ca. 500.000 navne fra deportationslister, er helt sikkert medvirkende til en del redundans (Sagi, 2017).

Jeg vil dog her, med resultaterne fra mine visualiseringer som belæg, argumenter for, at det vil være på sin plads at undersøge redundans oftere, når der bliver brugt DVIZ. Og ikke bare når det gælder denne metode, men måske generelt undersøge, hvor meget redundans der forekommer i NDB.

Da databasen bliver brugt flittigt til forsøg med forskellige datadrevne metoder og software (Sagi, 2017), vil det være oplagt, at ét af projekterne kunne være at lede efter redundans. Eliminering af redundans vil være med til at skabe bedre og mere akkurate data for fremtidige undersøgelser.

Hele problematikken omkring redundans i NDB er blevet undersøgt lidt nærmere i artiklen *Multi-source uncertain entity resolution: Transforming holocaust victim reports into people*, skrevet af Sagi, Gal, Barkol, Bergman og Avram og publiceret i 2017.

Det er dog ikke som problemstilling, redundans her bliver undersøgt. I denne artikel, beskrives, hvordan et team af forskere fra Israel, har undersøgt, - ikke hvor meget redundans der eventuelt er at finde i NDB, - hvordan de har udviklet en digital model – Uncertain Entity Resolution Model (UERM) - som ved hjælp af algoritmer og machine-learning, kan finde frem til samme personer, selv hvis deres navne er stavet på forskellige måder, er skrevet med forskellige alfabeter eller er baseret på billeder – kort sagt; en model der arbejder digitalt med heterogene kilder.

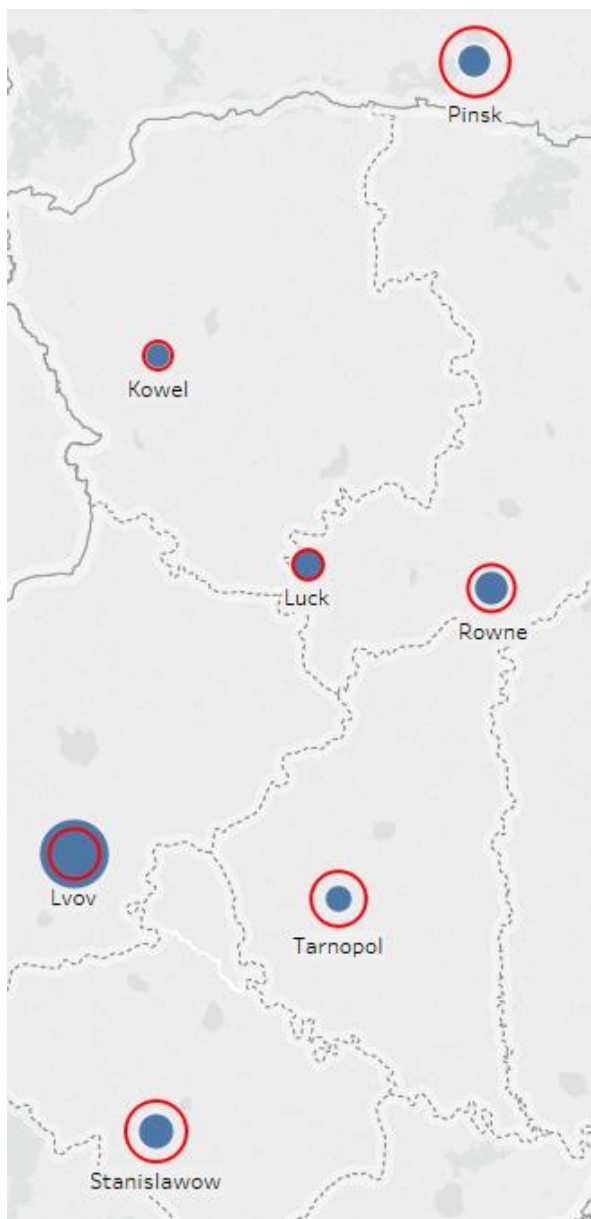
Man har i artiklen ønsket at finde frem til personlige fortællinger, ved at linke så mange informationer fra YadVashems mangeartede databaser sammen som muligt. De mange forskellige fragmenter og sproglige kilder, ligger digitaliseret i flere forskellige databaser hos Yad Vashem - som NDB, fotodatabasen og andre.

Artiklen beskriver indgående den tekniske side af projektet, og det er interessant at se, hvordan modellen/softwaren lykkes med at finde frem til, og sætte ansigter på mange holocaustofre (Sagi, 2017).

Projektet UERM, er endnu en datadrevet måde til at arbejde med de store mængder heterogene data som er vores kildemateriale til forståelse af Holocaust.

Man kunne forestille sig, at en videreudvikling på en model som denne kunne bruges til at finde frem til den meste redundans i NDB, og derved fremadrettet give mulighed for at arbejde med mere nøjagtig data i visualiserings metoder som bliver brugt her i opgaven.

Modellen kunne muligvis også med fordel bruges i samarbejde med visualiseringer, hvor disse peger på områder med ikke-registrerede ofre, og hvor det så forholdsvis let vil være muligt for UERM-softwaren at ”tage over” for at finde frem til nogle af disse.



5.4 De sovjetiske områder

Interessant er det også, at der tegner sig et billede af, at jo mere østpå man trækker, jo flere registreringer er der i NDB (Illustration 9A). Der er i Pinsk (128.647 registreringer i NDB, 20.220 registrerede i Census 1931), Rowne (59.016 registreringer i NDB, 22.737 i Census 1931), Tarnopol (80.977 registrerede i NDB, 14.000 i Census 1931) og Stanislawow (99.133 i NDB, 24.823 i Census 1931) langt flere registreringer i NDB end i Census 1931! Dette kan ikke umiddelbart forklares, men er interessant, og lægger derfor op til nærmere studier.

Det fremgår klart af visualiseringen (Illustration 10), at det primært er de større byer som Warszawa, Lodz, Lublin og andre, der har de største forskelle. De mindre byer som Radom, Chelm, Cestochowa Rzeszow og Rarnow ligger boble og ring næsten nøjagtigt ovenpå hinanden, og tyder på, at af det indbyggertal der figurerer i Census 1931, er en stor del registreret i NDB.

Illustration 10A. Udsnit af Illustration 10 der viser mønster af større røde ringe i de sydlige egne. Egen visualisering i Tableau

Meget iøjnefaldende er det også, at Warszawa har betydeligt flere ikke-registrerede ofre end alle andre byer på kortet. Det skal her bemærkes, at Warszawa var en by, hvor man samlede jøderne sammen fra hele oplandet i ghettoer, og at det derfor har været, og stadig er, svært at finde frem til ofrenes identitet. Hele familier blev flyttet flere gange og flere familier blev ”slået sammen” blandt andet på grund af pladsmangel i ghettoen. (Stone 2004, Kudym 2018) Selvom nazisterne førte detaljerede lister, over alle der blev deporteret, er der ikke mange af disse tilbage. De lister der indtil nu er dukket op, er i stor stil blevet digitaliseret og er inkluderet i NDB (Sagi, 2017).

En anden tendens, er at mange ofre blev flyttet til arbejdslejre, som kunne ligge langt fra deres hjemegn, og derfor kan være registreret på lister fra disse områder i NDB. Dette betyder, at der vil komme en skævhed i forhold til at bruge visualiseringerne til at pege på ikke-registrerede ofre ved sammenligning af databaser før og efter krigen. De områder, hvortil ofre var sendt for at udføre tvangsarbejde, er absolut ikke nødvendigvis tæt på de områder, hvor de opholdt sig før krigen. Dette fremgår i øvrigt tydeligt af de visualiseringer Kudym (2018) har foretaget. Dette kan ses som belæg for, at netop denne undersøgelse ikke er optimal til at finde frem til ikke-registrerede ofte. Det betyder ikke nødvendigvis, at visualisering generelt ikke er en holdbar forskningsmetode. Det peger på, at det er vigtigt, at data som lægges til grund for visualiseringer, passer til det formål de skal bruges. Det peger samtidig på, at der kræves fagekspertise og domænekendskab gennem hele processen – specielt i opbygning af database (Knowles, Cole & Giordano 2018, Drucker 2014).

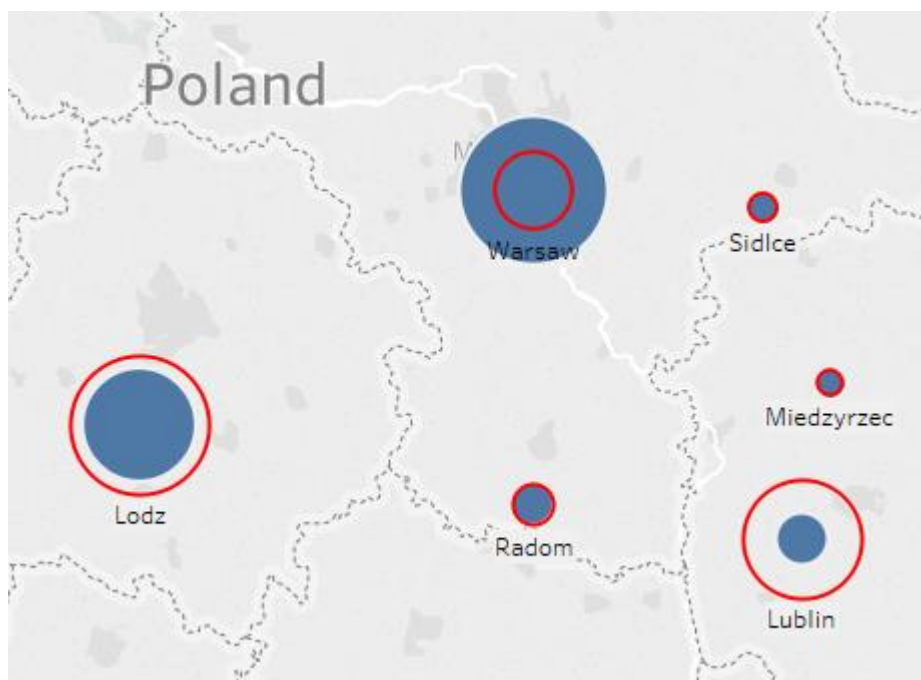


Illustration 12. Udsnit med zoom på 3 store og 3 små byer. Egen DVIZ med Tableau.

Hvis vi ser på illustration 12, hvor der er zoomet ind på 3 af de store og 3 af de små Polske byer: Warszawa, Lodz og Lublin, ses tydeligt store forskelle. Både mellem store og små, men også mellem antal registrerede i de to databaser.

Hvis vi starter med de store byer: I Lodz er der registreret 202.497 borgere i Census 1931 og flere end dobbelt så mange ofre - 412.161 - i NDB. I Lublin er forskellene langt større; her er der

registreret 38.937 borgere i Census 1931 og 301.692 ofre i NDB – altså næsten 8 gange flere end i Census 1931. I begge disse byer, er der flere registrerede i NDB end i Census 1931.

Her må der igen ses på data, og overvejes, om disse visualiseringer er tilstrækkelig detaljerede til at finde ikke-registrerede ofre.

Jeg argumenterer her for, at hvis der havde været mere nøjagtige og langt større data-corpora benyttet til visualiseringerne, ville også resultatet være langt mere nøjagtigt. I dette tilfælde, vil data som ligger tættere på hinanden, som for eksempel med NDB og GMC, give et bedre billede, og derfor kunne pege på geografiske områder, hvor der kan sættes ind med kvalitative undersøgelser på mikro-niveau. Umiddelbart, må man udlede af visualiseringerne for de to byer, at der ikke er mange ikke-registrerede ofre.

Warszawa er til gengæld én af de byer, hvor det fremgår af visualiseringen at der næsten er 2,5 gange så mange indbyggere i Census 1931, end der er registrerede i NDB. At der er så mange ofre, der ikke er registrerede i Warszawa, er hverken overraskende eller ny viden.

Warszawa er en af de byer under hele krigen, som blev mest ødelagt. Efter Warszawa opstanden i 1944 blev næsten hele byen jævnet med jorden af tyskerne. Ghettoområdet blev brændt af, og de overlevende indbygger blev deporteret direkte til udryddelseslejre (Stone 2004, Winstone 2015). På grund af byens historie under krigen, er det ganske oplagt, at der vil være mange ikke-registrerede. Der mangler grundlæggende data, og det data der findes, er som de fleste efterladte originale kilder fra perioden, ganske sparsomme og meget fragmenterede.

I tilfældet med Warszawa er der ikke brug for datavisualiseringer for at overveje, om der skal bruges resurser på at finde frem til ikke-registrerede ofre. Det er selvfølgelig oplagt, at der i områderne i og omkring Warszawa efterforskes løbende, hvilket der også bliver, dels af de lokale forskere på Det Jødiske Historiske Institut og ikke mindst på det nye Polske museum POLIN (Janika, 2015).

Når vi sammenligner de tre mindre byer Sidlce, Miedzyrzec og Radom med de tre store byer (Illustration 12), er det første der falder i øjnene, at der her er mindre forskelle mellem de to databaser. I Sidlce er der 14.685 registrerede borgere i Census 1931 og 17.550 ofre i NDB. I Miedzyrzec er der 12.000 registrerede borgere i Census 1931 og 13.465 ofre i NDB. I den største af de tre mindre byer, Radom, er der 25.159 borgere i Census 1931 og 35.219 ofre i NDB.

Her går mønsteret med, at der er flere registrerede i NDB end i Census 1931 igen. Der er dog en væsentlig forskel på differencen mellem antallet af de registrerede i byerne.

5.5 De tyske områder

Hvis vi ser tilbage på illustration 10, vil vi se, at der viser sig et mønster for næsten alle byer i visualiseringen fra Polen. De små byer har mindre differencer, mens der stadig er en tendens til, at der er flere registrerede i NDB end i Census 1931.

Efter at have set de store forskelle i visualiseringerne fra Polen, følte jeg det oplagt at sammenligne samme type visualiseringer fra Polen og Tyskland. Dette har fået mig til at lave samme type visualiseringer for områder i Tyskland, hvor datagrundlaget her er GMC overfor samme NDB.

Vil vi se samme tendenser i visualiseringer, der er baseret på GMC, hvor data er mere præcise, idet at GMC har data fra 1939, altså hele 8 år senere end Census 1931?

Det kunne måske give et delvist svar på, om det er data i Census 1931 der er hovedårsagen til, at der generelt ses flest registreringer i NDB. Hvis der ses samme mønster i visualiseringer med NDB overfor GMC, kan dette bruges som argument for, at de forholdsvis gamle data fra Census 1931 ikke udelukkende er årsagen til at der er flest registreringer i NDB. Dermed samtidig sagt, at det tyder på større redundans i NDB end som så.

I Illustration 13 hvor de Tyske byer (GMC), er visualiseret overfor NDB, ser man igen samme mønster som, i visualiseringerne fra Polen, hvor en større del af byerne har flere registrerede ofre i NDB end der er registreret i GMC. Det skal her gøres opmærksom på, at det er helt andre antal registreringer der er tale om – også i de større byer.

Her er det endda et endnu skarpere billede der viser sig: Der er her tale om hele 7 ud af 9 byer som har højere antal NDB-registreringer. Det er kun de to forholdsvis store byer Essen (3.394 registrerede i NDB, 7.865 registrerede i GMC) og Bremen (1.018 registreringer i NDB, 1.650 registreringer i GMC), hvor der umiddelbart mangler registreringer i NDB. Det er specielt interessant, fordi der her er tale om en Census, der er taget samme år som krigen brød ud og ikke som i Census 1931. Umiddelbart vil man formode, at det er væsentlig mere akkurate data, der ligger til grund for visualiseringen.

German Minority Census 1939

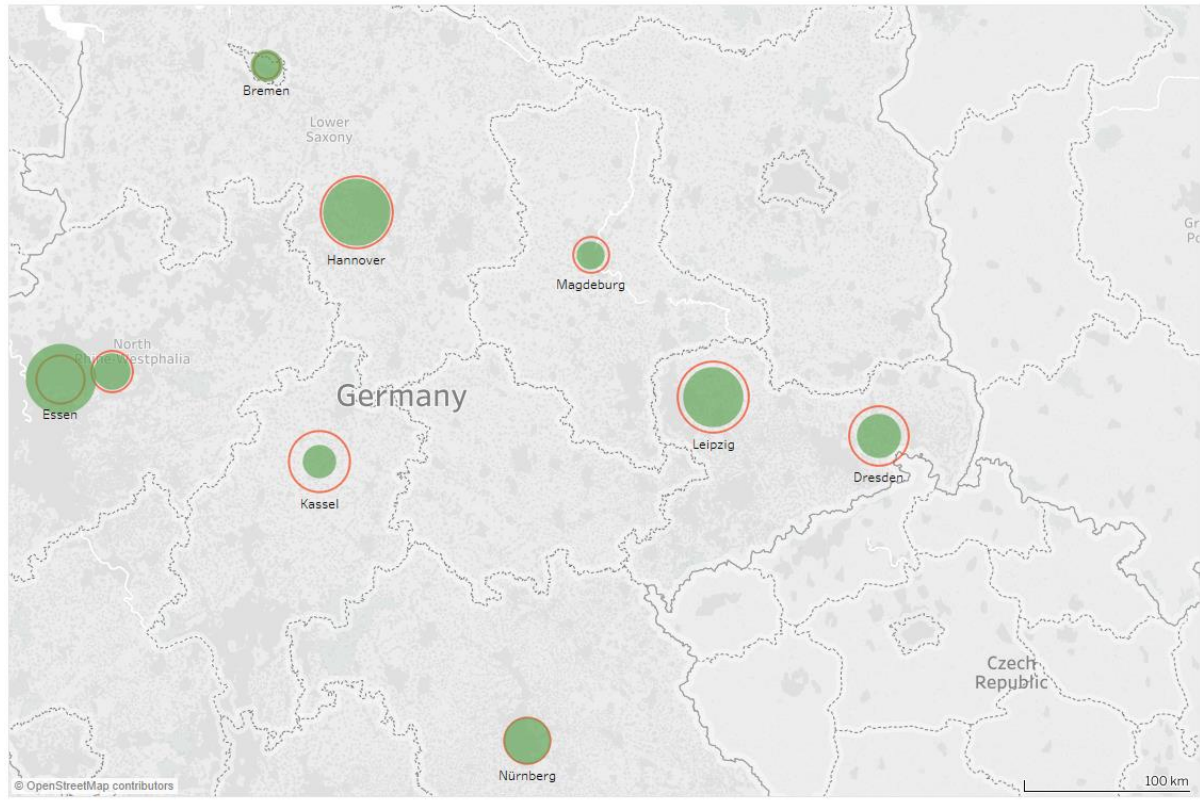


Illustration 13. GMC over for NDB. Egen visualisering i Tableau - (de grønne bobler er her gjort lidt mere utydelige, da det ellers er svært at se forskellene på farverne).

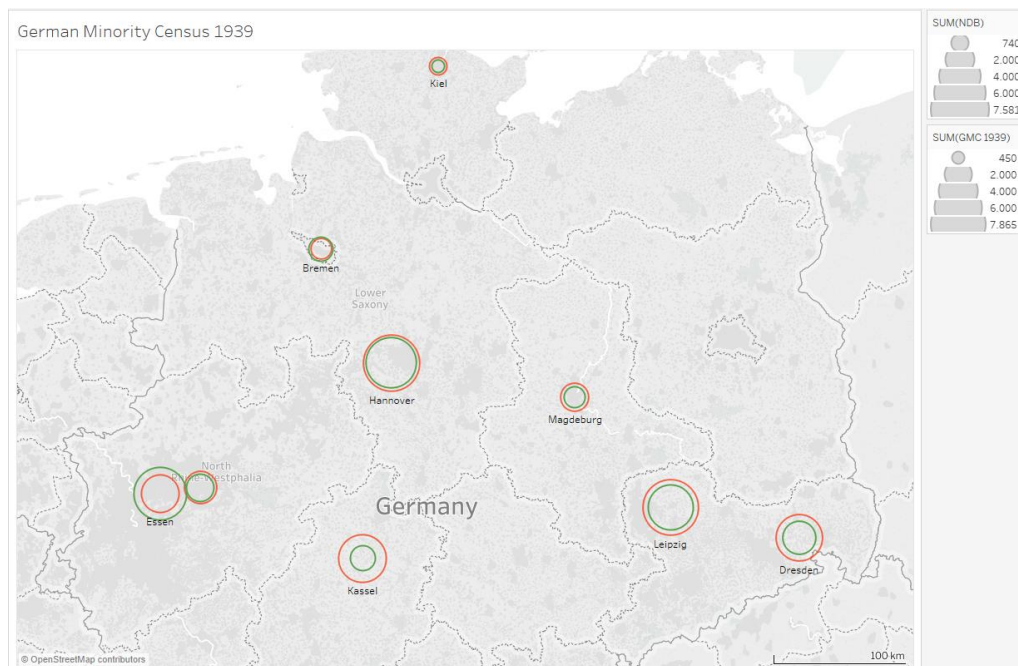


Illustration 14. Kort af GMC til sammenligning med illustrationerne fra Polen. Egen visualisering i Tableau.

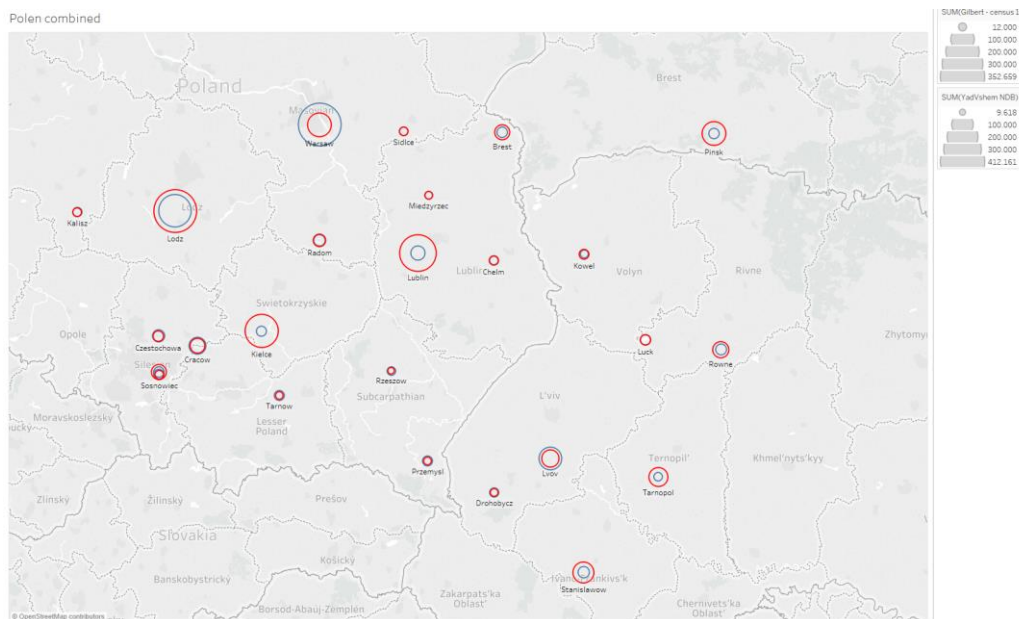


Illustration 15. Census 1931 og NDB vist med ringe til sammenligning med Tyske data. Egen visualisering i Tableau.

Illustration 14 stillet overfor illustration 15 viser helt klart samme tendenser; der er flere registreringer i NDB end i GMC – men forskellen mellem antal registrerede, er i de tyske data langt mindre.

Samme mønster som visualiseringerne fra Polen gentager sig her, også når der ses på store og små byer. Det er vigtigt at holde sig for øje, at der generelt er langt flere registreringer i Polen, selv med dette taget med i overvejelserne, er der stadig samme tendens til, at der er større forskelle i de store byer end i de mindre; at der er færre ofre fra NDB registreret end borgere fra GMC. Kun i den næststørste af de byer der er visualiseret i GMC, Essen, er billedet anderledes. Her er der 7.865 indbyggere registrerede i GMC og 3.394 ofre registreret i NDB.

Det ses samtidig tydeligt, at der er langt mindre difference mellem antal ofre i NDB og antal indbyggere fra GMC, selvom der generelt er flere registrerede i NDB.

Dette forhold kan som nævnt tidligere, tyde på, at data fra GMC er nyere og dermed mere akkurate end Census 1931 da der er kortere tid mellem det tidspunkt GMC er blevet udført på og slutningen af krigen. Det er dog også til dels skyldes, at der siden slutningen af krigen er blevet investeret mange ressourcer og meget tid på at finde frem til de tyske ofre for Holocaust, for at kunne udbetale erstatning.

Visualiseringerne kan ikke svare på hvorfor men kun vise, at der er en tendens. Der kan selvfølgelig ligge andre årsager til grund for denne difference, og det vil kræve undersøgelser på mikro-niveau, at klarlægge hvorfor. Det er dog interessant, da der viser sig et helt klart mønster i difference mellem de to datasæt i forhold til visualiseringerne fra Polen, hvor der ses store forskelle mellem datasættene.

Metodeanvendelsen til at finde frem til områder hvor der stadig er mange ikke-registrerede ofre, vil også her kunne benyttes med fordel.

5.6 Alfabeternes betydning

Det kunne tænkes, at metoden vil være en anelse mere sikker i sin anvendelse på data fra Tyskland, dels fordi GMC som nævnt, er data som er samlet tæt på starten af registreringerne i NDB og derfor er mere relevant data, og dels fordi der i Tyskland og Nordeuropa arbejdes løbende med at digitalisere de originale kilder, og med at organisere dem i store databaser. Mange Internationale Organisationer som International Tracing service (ITS) i Bad Arolsen⁵⁰, JewishGen⁵¹, United States Holocaust memorial Museum⁵² og mange andre, arbejder løbende med at dokumentere de mange ofre for Holocaust, både overlevende og ikke overlevende.

Også andre store Holocaust organisationer som EHRI⁵³ bruger ressourcer på at henvise til databaser, som kan medvirke til nye registreringer. Det gør et eventuelt projekt, hvor der bruges HGIS til at visualisere registrerede ofre langt nemmere, da en stor del af data allerede er samlet i databaser. Det er lettere at sammenkøre data, end at skulle opbygge nye databaser (Knowles, Cole, Giordano, 2014).

En anden fordel ved at arbejde med visualiseringerne fra tyske data, er at sproget bruger samme latinske alfabet. Det er en kolossal udfordring at arbejde med det polske sprog som har rødder i det Kyrilliske alfabet, også selvom der kun er tale om navne, som i dette projekt, hvis man ikke taler og

⁵⁰ <https://www.its-arolsen.org/en/about-its/> set 28/8/2018 9,50

⁵¹ <https://www.jewishgen.org/databases/#Germany> 28/8/2018 9,55

⁵² <https://www.ushmm.org/> 28/8/2018 10,15

⁵³ <https://ehri-project.eu/> 28/8/2018 10,20

læser sproget. Det er bare én af de mange årsager til, at der stadig ikke er lavet tilnærmelsesvis så mange undersøgelser af Holocaust i Polen som i resten af Nordeuropa.

At det polske sprog har et andet alfabet end de Nordeuropæiske lande, har i det hele taget stor betydning for Holocaustforskningen: det har blandt andet betydet, at mange af de hjemmesider der tilbyder software til at bearbejde data, specielt i forbindelse med tekstanalyser, ikke kan anvendes. Et redskab som Google Ngram, eksempelvis, hvori der kan søges i tekst-corpora på adskillige forskellige sprog, inkluderer ikke på polske data⁵⁴.

Derudover findes der meget litteratur om Holocaust, som endnu ikke er oversat fra polsk til for eksempel engelsk eller tysk. Her ligger sandsynligvis mange informationer og venter på at bliver taget frem i dagens lys.

Det kan også være et argument for at bruge visualiseringer i historieforskning, at data bliver oversat til grafisk kommunikation, som er forståeligt, uanset sproglig baggrund. Det gør også, at visualiseringer kan ses som et universelt sprog. Data repræsenterer en simplificering – et abstrakt billede af virkeligheden, og visualiseringer er som et snapshot, et billede der har indfanget denne abstrakte verden i et øjeblik – eller over tid når der er tale om interaktive visualiseringer, der viser begivenheder over tid og rum (Yau, 2013).

At udlede viden og forståelse fra en visualisering er ikke så nemt som at kigge på et fotografi – det kræves at man forstår den underliggende, semantiske strukturs principper, for at forstå og analysere en visualisering (Yau 2013, Drucker 2014), men mange af de fordele fotografiet har, deles af visualiseringer. Et eksempel kunne være, at visualiseringer er nemme at dele med andre, på tværs af grænser og sprog – hvilket er en fordel indenfor forskningen, hvor videndeling er en vigtig parameter i forhold til at finde og dokumentere ny viden. Jo mere domænespecifik og sofistikerede visualiseringerne er, jo bedre er de til at formidle ny viden, skabt fra analyse af disse (Drucker, 2014)

⁵⁴ <https://books.google.com/ngrams> set 27/7/2018 10,30

6.0 Diskussion

Baseret på de ovenstående visualiseringer vil jeg argumentere for, at datavisualisering er et digitalt redskab, som holocaustforskere såvel som historieforskere må forstå vigtigheden af, og at datadrevne forskningsmetoder, ikke bare er kommet for at blive, men at de kan åbne døre, som før har været lukkede. Samtidig vil jeg argumentere for, at data aldrig kan stå alene, men kræver kritisk analyse.

6.1 Digital Humanities (DH)

Selvom der ikke findes en faglig definition omkring hvad digital Humanities, er der bred konsensus om DH, er når der bruges datadrevne forskningsmetoder ,og arbejdes med computerteknologier indenfor humaniora (McCarty 2003, Early-Spandoni 2017).

Klassiske, konventionelle forskningsmetoder indenfor humaniora har tidligere kun været kvalitative metoder, hvor forskere har været begrænset til et forholdsvis lille datamateriale.

Med computerens hastige indtog i forskermiljøerne, giver computerkraft nu forskere indenfor humaniora mulighed for at arbejde med langt større datasæt: tekstanalyse kan nu udføres ved hjælp af software og nye muligheder for at drive forskning med enormt store datasæt, giver helt nye muligheder, som for eksempel ved hjælp af ordstatistik i GoogleNgram⁵⁵, et stykke online software, udviklet af Google.

Digital Humanities har fået mange betegnelser hæftet på sig op gennem tiden, og er blandt andet kaldt en visionær diskurs med en utopisk kerne, en revolution og et fag, der åbner for den største historiske transformation i menneskets historie (Van Eijnatten, Pieters & Verheul 2013). Det kunne godt lyde overdrevent, men der er nok ingen tvivl hos nogen om, at de mange, digitale værktøjer der bliver brugt til datadrevet forskning, vil ændre den klassiske måde, hvorpå der bliver forsket - også indenfor Humaniora.

⁵⁵ <https://books.google.com/ngrams/info> set 28/8/2018 11.30

Der er op gennem de sidste 15 år (Hughes, 2016) lavet forskellige forslag til at strukturere og kategorisere ”indholdet” af DH i en taksonomi; hvad falder ind under feltet? Kan man bygge en taksonomi, der kan være rammesættende for DH?

Et af resultaterne kommer fra Oxford University Digital Humanities Programme, og er konstitueret med de følgende kategorier:

- kommunikation og samarbejde
- Dataanalyse
- Data-mining
- Dataudgivelse og formidling
- Datastrukturering og forbedring
- Praksis-ledet forskning
- Strategi og projektledelse

Derudover er der en inddeling i underniveauer:

- Audiovisuel analyse
- Søgning og linking
- Statistisk analyse
- Tekstanalyse
- Anden analyse
- Visualisering

Denne taksonomi ses som rammesættende, når der skal defineres projekter indenfor DH.

Visualisering som metode falder ifølge denne Taksonomi ind under Digital Humanities. Ved hjælp af computerkraft, kan man visualisere store mængder data, som måske i kraft af deres størrelse kan bibringe forskningen bedre gennemsigtighed, og dermed en dybere og samtidig mere sofistikeret forståelse af de historiske hændelser (Knowles, Cole & Giordano 2014, Gregory & Geddes & Geddes 2014, Kitchin 2013).

De digitale værktøjer som blandt andet bliver brugt her til visualiseringerne, er stærke værktøjer, blandt andet i kraft af at de kan bearbejde og behandle datamængder, det ikke er muligt at håndtere med traditionelle metoder. Men de digitale værktøjer kræver også nye, anderledes og specielle måder at arbejde med forskning på. Som tidligere nævnt, er det vigtigt, når der bruges datadrevne

analysemetoder, at man samler teams af specialister fra forskellige faggrupper – at et tværfagligt arbejde etableres i starten af projektet, så de rigtige kompetencer kan komme i spil.

Kitchin (2013) mener, at der i de sidste årtier ikke har været nok fokus på at uddanne nye, unge forskere der har de rette kompetencer til at analysere store datasæt, specielt indenfor geografiske visualiseringer. Det kan have stor indflydelse på, hvor godt visualisering virker som metode. Det lægger op til, at der bliver sat en dialog i gang, omkring hvilke kompetencer der kræves indenfor feltet.

Er det en mulighed, at Digital Humanities, tager fat i spørgsmål og udfordringer som dette? Det er klart, at der fremadrettet, kun bliver brug for flere mennesker, der kan analysere visualiseringer, da de eksponentielt voksende datamængder kræver digitale visualiseringer for overhovedet at kunne overskues og derfra blive analyseret og forstået.

Måske er der brug for langt mere avancerede programmer til at kunne fordøje fremtidige datamængder? Kan man forestille sig, at disse programmer kommer til at være vigtigere end den menneskelige ressource? Hvilke kompetencer skal der satses på? Hard-core IT-udvikling og/ eller visualisering/kommunikation og analyse?

I et større projekt end det jeg her præsenterer, med visualisering af holocaustofre på individ-niveau, vil der som minimum være brug for IT – udviklere til at binde de forskellige software sammen, dataanalytikere til at analysere data i samarbejde med historikere med speciale indenfor netop Holocaust og krigen, og ikke mindst vil der være brug for informationsspecialister, til blandt andet at organisere de store mængder data og dokumentere projektets arbejdsgang.

Det er vigtigt for forskningen, fremadrettet, at arbejde i tværfaglige teams som ovenfor beskrevet, for at få så stort udbytte af data som muligt. De mange store EU-støttede digitale forskerinfrastrukturer som EHRI⁵⁶ (som er beskrevet mere detaljeret senere i denne opgave), DARIAH⁵⁷, CLARIN⁵⁸ og andre, er med til at gøre dette samarbejde lettere. Det er ikke nødvendig at forskere for de forskellige fagområder absolut sidder sammen, - infrastrukturene giver mulighed

⁵⁶ <https://ehri-project.eu/> set 26/6/2018 9,50

⁵⁷ <https://www.dariah.eu/> set 26/6/2018 9,53

⁵⁸ <https://www.clarin.eu/> set 26/6/2018 9,55

for, at der kan arbejdes i teams uden at tage hensyn til tid og sted. (Hughes 2016, Van Eijnatten, Gregory & Geddes & Geddes 2014, Knowles, Cole & Giordano 2014).

Eksperter fra hvert deres område, kan i et bredt samarbejde ud fra data online producere ny viden om historiske begivenheder.

6.2 Lidt om data

Begreber som viden og data er komplekse størrelser, og kræver her en kort redegørelse:

Jeg bruger her Bernt Gustavssons (2000) definition af begrebet ”viden” som kan beskrives som:

”Viden er noget vi indoptager i os selv, og som forandrer vores opfattelse og øger vores forståelse af omverdenen” Gustavsson, B. (2000)
Videns filosofi. Forlaget Klim, Århus N. 1. udg. 3. oplag, 2011 s.15.

Kitchin (2014) forklarer i et forsøg på at definere data, at data er det der eksisterer før argumentet der konverterer disse til informationer og fakta. Ordet data stammer fra det latinske *dare*, som betyder *at give*. Det betyder, her, at data er de rå elementer der kan trækkes ud af fænomenet for derefter at blive manipuleret på forskellig måde. Begrebet data refererer generelt til elementer der allerede er *taget* ud af noget andet ved hjælp af observationer, manipuleringer og eksperimenter, fra arkiver. Derfor er ordet for data faktisk unøjagtigt, og burde i stedet mere korrekt være *Capta*, som stammer fra det latinske *capere*, der betyder *at tage*, - *capta* værende de enheder af data, som er blevet udvalgt og høstet ud af alle de potentielle data (Kitchin, 2014).

*“Data are capta, taken not given,
constructed as an interpretation of the
phenomenal world, int inherent in it.”*

Drucker, J. 20104. S. 128.

Ved at have *capta* i bevidstheden, accepteres hermed, at data altid er fortolkede idet de er valgt ud af et potentielt materiale for at blive brugt til et specifikt formål og med forventninger om at se specifikke resultater. Data er ikke stabile entiteter der eksisterer a

priori. De er entiteter der altid vil være subjektivt påvirket af den virkelighed de stammer fra (Drucker, 2014).

”Data are commonly understood to be the raw material produced by abstracting the world into categories, measures and other representational forms – numbers, characters, symbols, images, sounds, electromagnetic waves, bits – that constitute the building blocks from which information and knowledge are created.” Kitchin, R. (2014) *The Data Revolution, Big Data, Open Data, Data Infrastructures & their consequences*. Sage Publications, Los Angeles, USA. s.1

6.3 Visualiseringer i tid og rum

Når der tales om datavisualisering, og specielt når der tales om datavisualisering af historiske begivenheder, spiller tid og sted en stor rolle. Datavisualisering giver jo netop mulighed for at visualisere store mængder data over tid og rum, hvilket ikke har været muligt tidligere. I takt med den teknologiske og digitale udvikling og øget computerkraft, er mulighederne vokset for at kunne arbejde med større og større datasæt. Det har stor betydning for datavisualisering, at der nu er mulighed for at visualisere i både 3D og interaktive visualiseringer hvor brugeren selv kan sætte parametre op for visualiseringerne (Earley-Spadoni 2017, Edelstein o.a 2017, Gregory & Geddes 2014, Knowles, Cole & Giordano, 2014)

I artiklen “Placing the Past: Groundwork for spatial Theory of History” reflekterer Ethington (2007) over begrebet “tid”. Dette begreb, som er grundlæggende i diskursen omkring spatial HGIS, er blevet diskuteret ofte af filosoffer og historikere langt tilbage i tiden (Ethington, 2007).

Det er svært at finde en fast definition for tid. Tid som begreb er blevet til, for at vi som mennesker, kan forholde os til den virkelighed som den fysiske verden omkring os udgør. Tiden eksisterer ikke ”i sig selv”. Begrebet kan ses som et måleinstrument for bevægelse i rum. Ethington mener, at historie derfor kun kan repræsenteres ved steder, hvor menneskelig udfoldelse er forekommet og at (Ethington 2007,) at historien ikke er en redegørelse for ændringer over tid, men mere en ændring gennem rum. Artiklen udfordrer her den klassiske historieforskning ved at fokusere på begrebet ”tid” og ”rum” i historieforståelsen, og har stor relevans for betydningen af muligheden for at bruge digitale geo-spatiale-visualiseringer af historiske begivenheder eller perioder. Det er derfor også

vigtigt, at historikere forstår denne ”spatiale drejning” indenfor humaniora, hvor der er fokus på kombinationen af tid og rum, i stedet for at se udelukkende på begivenheder der har fundet sted i en kronologisk rækkefølge, som klassisk historieforskning lægger op til (Ethington 2007, Knowles Cole & Giordano 2014). Ethington fremlægger her en teori om at tid er *transcendent a priori*, meget lig flere af de teorier som flere af oplysningstidens store filosoffer som Kant og Heidegger, trods deres store forskelligheder, var tilhængere af.

Netop forståelse af denne teoretiske tilgang, vil have stor betydning for hvordan historikere fremadrettet vil kunne udnytte de digitale værktøjer til at analysere og fortolke visualiseringer, og derved se historien fra nye vinkler. Det er blandt andet, denne kombination af tid og rum der er disse værktøjers styrke.

I projektet her, er der ikke brugt oplysninger om tid som en del af datasættene. Det er dog vigtigt at understrege, at et muligt fremtidigt projekt, hvor netop denne faktor også bliver taget med i visualiseringerne, - for eksempel med data om hvor mange indbyggere der boede i de respektive byer i årene mellem Census 1931 og 1945 - sandsynligvis vil kunne vise langt mere interessante interaktive visualiseringer, hvor man vil kunne se variationerne af de forskellige byers størrelser og ændringer over tid. Den type visualiseringer lægger op til at se hele periodes udvikling over tid og kan derfor muligvis tilføje vigtig viden, som ikke før er blevet opdaget. I et større projekt, vil der samtidig være større data-corpora, og derfor også større muligheder for serendipitet (Knowles, Cole & Giordano 2014).

6.4 Visualiseringer kan ikke stå alene

Det er vigtigt, i al form for kvantitativ undersøgelse hvor man arbejder med sammenkørte databaser og specielt i arbejdet med store datasæt og specielt BD at understrege, at data og DVIZ ikke kan og må stå alene. At visualisere store datasæt kan ses som en forlængelse af de geo-spatielle undersøgelsesmetoder, der tilbage fra starten af 70'erne, har været brugt indenfor geografien (Gregory & Geddes, 2014) og da tal stammer fra specifikke institutioner, og organisationer med forskellige politiske agendaer og interesser, vil der altid være brug for at stille kritiske spørgsmål til data og databehandlingen (boyd & Crawford, 2011), når der formidles ved hjælp af visualiseringer. At DH gør brug af datadrevet forskning som er baseret på software, betyder ikke nødvendigvis, at DH har en positivistisk vinkel. Tværtimod, er der hos humanisterne stort fokus på, at metoderne brugt i DH må evalueres, og løbende kritisk analyseres (Hughes, boyd & Crawford 2012, Kitchin 2013, Gray 2016)

”Et af de områder der kræver stort fokus, er netop selve fortællingen ud fra data. Ikke kun tolkningen, men at bygge en historiefortælling ud fra visualiseringerne, som kan formidle den viden de kan tilføje området – og formidle den på en måde, så den viser så korrekt et billede af historien som muligt. Historieforskning er også historiefortælling. Specielt indenfor Holocaustforskningen, er det vigtigt at have micro-analysen i mente, når de store penselstrøg overtager. Der ligger i alle data og tal menneskeskæbner til grund. Visualiseringerne kan vise hvordan og så er det op til humaniora at svare på hvorfor!” - Mikkelsen, L. BA-opgave 2016, s. 32. INF, Det Humanistiske Fakultet, Københavns Universitet.

At tro, at korrelationer kan erstatte kausalitet, er at bedrage os selv og andre (Barnes 2013).

At forholde sig realistisk til datavisualisering, må være baseret på en ide om, at fænomenet afhænger af øjet der ser – at visualiseringen er afhængigt af personen der observerer. At forholde sig til en visualisering som værende det samme som det observerede fænomen i sig selv, øger den kritiske distance mellem verden som fænomen og måden denne bliver forstået på.

Her vil jeg argumentere for, at metoden med at visualisere store datasæt som i denne opgave og specielt måske, når der er tale om meget store datasæt – organiserede eller uorganiserede, med fordel kan bruges til at vise korrelationer og mønstre som *understøtter* kausalitet.

Metoden som et forskningsværktøj, kan være med til at hjælpe os til analysere, filtrere, gennemse og udforske komplekse informationer. Specielt datavisualiseringer hvor geografiske data er grundlæggende og også den tidslige dimension, er en del af data, vil have stor betydning i historieforskning (Kitchin 2013, Kitchin 2007, Gregory & Geddes 2014, Earley-Spandoni 2017).

Det er umuligt at forestille sig et fænomen eller en begivenhed, uden at forestille sig at den er placeret et sted (Ethington, 2007). Det er derfor vigtigt, at der i historieforskningen, også når der visualiseres, bliver brugt geografiske data og at visualiseringerne grundlæggende er baseret på spatial data, samtidig med at man holder for øje, at visualiseringerne ikke kan stå alene.

”... maps are never fully formed and their work I never complete. Maps are of-the-moment, beckoned into being through practices; they are always mapping. From their perspective maps are fleeting, contingent, relational and context-dependent, emerging through transductive processes to solve relational problems. This theoretical turn has led us to suggest that cartography I processual, not representational, in nature. Rather than cartography being narrowly understood as the scientific pursuit of how best to represent the spaces of the world (focused on issues such as form and accuracy, cartography becomes understood as the pursuit of representational solutions (not necessarily pictorial) to solve relational, spatial problems. In so doing, cartography shifts from being ontical in status, wherein the ontological assumptions about how the world can be known and measured are implicitly secure to an ontological project that questions more fully the work maps do in the world.”

Kitchin, R. & Dodge, M. (2007) Rethinking maps. *Progress in Human Geography* 31(3) 2007.pp. 331-344. Side 343

De store datamængder der produceres i dag, stiller forskerne overfor store udfordringer. Det er absolut nødvendigt, hvis forskningen skal kunne producere ny viden, at der er værktøjer, der kan bearbejde disse enorme mængder data. Selvom det er blevet nemmere og billigere at høste, opbevare og visualisere data, kræver de nye teknologier konstant vedligehold og opdateringer. At det er blevet muligt at tilgå og opbevare data online, gør det nemmere at holde software opdateret, da det kan foregå centralt og automatisk, men kræver alligevel, at den brugte software løbende bliver kritisk evalueret (Hughes 2016, Mikkelsen 2016). Netop fordi meget software og data bliver opbevaret er Sky-baserede, og derfor er en del af Internettet, som ikke er et statisk, rigtigt medie, men et evigt foranderligt netværk der altid er i bevægelse og ændres hvert sekund, er denne del af datadreven forskning en større udfordring end den umiddelbart virker som.

Data er altid udvalgt: det vil sige, at der har været truffet menneskelige beslutninger om, hvilke data der er tale om. Data er derfor grundlæggende aldrig neutrale, de vil altid være fabrikerede (Kitchin 2014, Boyd & Crawford 2011). Kunne man forestille sig, at dette blev ændret i fremtiden?

”Selv i Big Data exhaust, hvor data er fuldstændig tilfældige bits and bytes, er der i det øjeblik data bliver visualiserede, en

menneskelig tolkning ind over – og dermed er det ikke muligt at visualisere data 100% objektivt – data vil være filtrerede på et eller andet niveau, som den givne forsker på forhånd har lagt, bevidst eller ubevidst. Lidt ligesom i Heisenbergs ubestemthedsrelation, hvor i det øjeblik, der i en undersøgelse bliver målt på kvantefysisk niveau, vil ske ændringer i empirien.⁵⁹ ” Lone Mikkelsen, BA-opgave 2016. s 29

Mange forskellige parametre har haft indflydelse på visualiseringerne: hvilke data er valgt og ud fra hvilke præmisser? Hvis specifikke data udeladt og andre valgt til, må man stille spørgsmål til hvorfor.

Hvis datavisualiseringer bliver brugt til at belyse historiske perioder eller begivenheder, må det tages i betragtning, at mange af de originale kilder er født digitalt, og derfor har været gennem digitaliseringsproces af en slags. I denne proces har nogen taget stilling til problematikker i forbindelse med selve digitaliseringen, - disse kan have påvirket materialet subjektivt.

I alle former for arbejde med visualiseringer af data og specielt i arbejdet med store data corpora og BD kan der opstå muligheder for at mistolke eller bare misforstå data. Man synes måske at se sammenhænge, hvor disse reelt set ikke er til stede. Det gør det derfor vigtigt i arbejdet med digital datavisualisering, at man i analysen af data, arbejder tæt sammen med fagspecialiser, i dette tilfælde med historikere, som har indgående kendskab til Holocaust, så man kan få afstemt kvaliteten af data. Det er vigtigt i arbejdet med specielt opbygningen af database som grundlag for datavisualiseringer, at der før en analyse stilles klare spørgsmål til hvilke mønstre og korrelationer, der eventuelt KAN ses, og at det bliver holdt for øje, at disse ikke nødvendigvis viser et sandt billede, men måske snarere et billede man forventer at se.

Et eksempel herpå kunne være at visualiseringerne foretaget her, umiddelbart ville formodes at vise, at der var færre registrerede ofre i NDB end registreringer i henholdsvis Census 1931 og GMC, da

⁵⁹Heisenbergs ubestemthedsrelation: <http://fysikl.eksikon.nbi.ku.dk/u/ubestemthedsrelation/> set d. 28/8/18 kl. 9

der stadig formodes, at der ud af de ca. 6 millioner Holocaustofre stadig mangler at blive registreret ca. 1.8 millioner ofre i NDB.

Der kan argumenteres for, at datasæt kan blive så store, at de kommer tæt på at vise troværdig viden (Andersson, 2008), men set fra en post-positivistisk vinkel, vil der altid være en subjektivitet ind over ikke bare analysen, men også gennem hele processen i udvælgelsen af hvilke data, frem til beslutning om hvilket software der skal bruges til opbygning af database og frem til visualiseringen. Ikke mindst, bliver brugergrænsefladen altid fremstillet ud fra ønsker, der er forudbestemt (Knowles, Cole & Giordano, 2014). Dette betyder, at den viden, der i sidste ende vil blive produceret, er data som er fanget: trukket ud af fænomenet – de er ikke identiske med fænomenet (Kitchin 2014).

Der er en del kritikere der mener, at brug af Big Data og visualiseringer indenfor videnskab, er en tilbagevenden til de rent positivistiske metoder. Som kritikeren Martin Fricke i sin artikel fra 2015 – *Big data and its epistemology* siger: "*Datadriven science is a chimera*" (Fricke 2015 s. 651). Fricke argumenterer for, ved at referere til videnskabsfilosofiske ikoner som Hume og Popper at Big Data ikke kan leve op til præmisserne for disse grundlæggende videnskabsteorier, og derfor ikke kan bruges i videnskabelige sammenhænge. Fricke mener, at vores viden om data er fejlagtig og data er derfor er hypotetisk (Fricke 2015). Også Barnes (2013) i sin artikel "Big data, little history" kalder det at drive videnskab ved hjælp af Big Data og visualisering af disse for "gammel vin på nye flasker"; og at metoden grundlæggende peger på en del problematikker som tilbage i tiden da, den geografiske kvantitative revolution holdt sit indtog, allerede er blevet diskuteret. Fricke (2015) stiller spørgsmålstejn ved om metoden kan bidrage med ny viden, der ikke bare er ny data.

Eksemplerne i denne opgave, hvor der bliver brugt grupperinger i visualiseringerne, er ikke optimalt til at vise korrelationer, men ved at nærstudere visualiseringerne som her, kan der måske observeres geografiske områder, hvor det er iøjnefaldende, at der i nogle af områderne er en tendens til, at der er flere registreringer end andre, som for eksempel i Illustration 8 hvor det tydeligt træder frem, at der er flere byer, hvor de røde cirkler er større end de blå bobler, eller i illustration 7, som klart viser, at der i de tyske byer er langt flere registrerede ofre end i de polske byer.

Ved visualiseringer som disse må man dog passe på med at komme ind på en positivistiske glidebane. Selv de resultater en visualisering kan vise i form af mønstre eller korrelationer der træder frem, må omhyggeligt analyseres og kritisk diskuteres af domænespecialister.

Måske specielt i store projekter, hvor der arbejdes med meget store datasæt, da disse kan være helt umulige at gennemskue *uden* visualiseringerne. Det lægger her igen op til, at opbygningen af database er en vigtig del af hele processen og at der ligger veldokumenteret materiale til grund for denne proces. For at producere gode visualiseringer, er den struktur der ligger bag opbygning af en database - selve dataorganisering kritisk.

Johanne Drucker (2014) argumenterer for, at der i dag er for lidt fokus på denne del af processer, da nogle visualiseringsformater, som for eksempel tabeller, er så generelle i deres udtryk og lette at genbruge, at fokus er forsvundet fra selve strukturen i data – vi accepterer deres formål uden kritik – og glemmer ofte, de bagvedliggende strukturer, som er grundlaget for visualiseringerne. Drucker (2014) peger som eksempel på, at tabeller i sig selv, indeholder information, men når mennesker læser disse, kan tabellerne kombineres og skabe ny viden.

Rækker og linjer – som i tabeller - er helt grundlæggende i en DB. I mere sofistikerede visualiseringer som geografiske kort, er der tale om at tabellerne bliver lagt i et koordinatsystem, og giver dermed mulighed for at vise relationer mellem data afhængigt af et sæt grafiske parametre. Beslutninger om hvordan disse grafiske parametre skal bruges i en visualisering, er grundlæggende for, hvad der vil kunne blive observeret, og er derfor et meget magtfuldt retorisk instrument i brugen af visualiseringer (Drucker, 2014), da disse parametre er kritiske i forhold til hvordan de viste elementer udtrykker sig i forhold til hinanden, og derved for analyse af visualiseringen.

” The challenge is enormous, but essential, if the humanistic worldview, grounded in the recognition of the interpretative nature of knowledge, is to be part of the graphical expressions that come into play in the digital environment. If we do not engage with this challenge, we give the game away in advance, ceding the territory of interpretation to the ruling authority of certainty established on the false claims of observer-independent objectivity in the “visual display of quantitative information.”

Drucker, J. (2014) *Graphesis – Visual Forms of Knowledge Production*. Cambridge Mass. Harvard University Press. Side 136

Jeg argumenterer således for, at visualiseringer - som data - aldrig kan stå alene, men at de kan være med til at se historiske begivenheder fra nye og anderledes vinkler, end dem vi hidtil har brugt, og derved muligvis frembringe ny information – og ikke mindst stille nye spørgsmål.

Datavisualiseringer er med til at forstærke vores sanser og vores muligheder til at give verden omkring os mening. Men datavisualiseringer er ikke neutrale i deres forankring og derfor heller ikke et neutralt redskab. Visualiseringerne kommer med en specifik vinkel hvorfra der bliver ”set”, specifikke analyser, specifik mediering og dagsordner for specifikke historiefortællinger, som vi altid må være opmærksomme på, når datavisualisering bliver brugt i forbindelse med forskning (Gray 2016, Boyd & Crawford, 2011, Barnes 2013).

Dette taler mod at DVIZ og datadreven forskning kan kaldes rent positivistisk, som det end imellem er blevet argumenteret (Gregory & Geddes & Geddes 2014).

Det vil som regel være nødvendigt med kvalitative analyser af data i det grundlæggende arbejde med at indsamle og analysere data, før en database er bygget. Der bør i forbindelse med visualiseringer af store datasæt følge dokumentation på, hvorfra data er hentet og hvordan er blevet rensat, sorteret eller manipuleret (Mikkelsen, 2016). En historiedatabase er ikke bare en passiv databeholder som er fyldt op af data fra en åben hane; udvælgelsen og bearbejdningen af data inden den bliver organiseret i en database er et stykke historieforskning, godt gennemarbejdet og fortolket af de eksperter der har bygget databasen (Knowles, Cole & Giordano 2014) Dette er et godt argument for, at datadreven forskning ikke er en rent positivistisk metode, men lige så meget en heuristisk metode, baseret på subjektive skøn og beslutninger.

De digitale værktøjer, som visualiseringsmetoden er baseret på, kan forstås som værktøjer, der giver mulighed for at se historiske perioder og historiske fænomener fra en ny vinkel.

Selvom metoden med visualiseringer i holocaustforskning umiddelbart synes at være et rigtig godt værktøj til at formidle holocaustviden og foretage analyser med, er det vigtigt, at vi er omhyggelige med, ikke bare at lade os forføre af den magt og de store muligheder som store datasæt og visualiseringer umiddelbart tilbyder, men hele tiden holde os for øje, at metoden også har sine begrænsninger (Mayer-Schönberger & Cukier 2014, Knowles, Cole & Giordano 2014, Sagi 2017, Van Einatten 2013).

Visualiseringer er gode til at komplementere allerede etableret viden, og kan være et godt og nyttigt værktøj til brug af analyse af store mængder, komplicerede datasæt. I opgaven her er der omhyggeligt udvalgt datasæt, og i et større projekt, med større datasæt, vil det stadig kræve omhyggelig udvælgelse og kritik af data, inden en eventuel database kan visualiseres.

Visualiseringer som bliver brugt til analyse af historisk materiale, må altid, for at opnå sikker viden, trække på domæneviden, og kan derfor bruges som understøttende argumentation. Det er vigtigt, at man har den nødvendige domæneviden, så der kan analyseres fra skiftevis fra makro- og mikroniveau (Mikkelsen, 2016).

Da metoder drevet af data som her, og specielt til historiske analyser ikke er objektiv, som eksempelvis ved at data er udvalgt i forhold til bestemte dagsordner og ud fra forventninger om at få svar på forhånds stillede spørgsmål, vil metoden altid kræve domæneviden. Datavisualiseringer som metode, er ikke bare grafiske billeder af specifikke områder, som kan illustrere ny viden som er fundet ved hjælp af andre metoder. Datavisualisering i historieforsknings kontekst, kan meget mere end det. Datavisualisering er i sig selv metoden til at forske. Visualiseringer kan få os til at stille spørgsmål til kilder, som måske ikke var blevet stillet, hvis ikke visualiseringerne havde peget på disse og dermed gået glip af oplysninger om fortiden, som kunne være med til at belyse historiske begivenheder som Holocaust (White, 2010).

Det er meget vigtigt, at netop dette argument bliver forstået i forhold til datavisualisering i holocaustforskning. Jeg vil argumentere for, at man som i mine visualiseringer her i opgaven ser, tendenser til, at der er flere registrerede ofre i NDB, end der er registrerede indbyggere i de to andre databaser, Census 1931 og GMC, at metoden kan bruges til at udforske og analysere også ældre data, som ikke er digitalt født, såvel som nye datasæt der hele tiden kommer til.

6.5 Gode visualiseringer kræver gode tværfaglige teams

Det er, specielt ved brug af disse digitale metoder, uhyre vigtigt, at der finder et tværfagligt samarbejde sted for at udnytte metoden fuldt ud, forstå data og visualiseringerne, og ikke mindst for at kunne verificere de nye oplysninger der eventuelt kunne komme ud af undersøgelse. Den tekniske side af metoden, kræver, at der er IT-specialister, der har kompetencer til at vælge netop de visualiseringsværktøjer, der egner sig bedst til opgaven, og eventuelt kan udvikle nye API'er som kan være med til at samle heterogent data.

Derudover bør der være informationsspecialister på teamet, som kan være med til at organisere og opbygge databaser i tæt samarbejde med historikere med speciale indenfor netop det område, som bliver undersøgt ved hjælp af DVIZ. Arbejdes der med strukturerede data – som her, er det vigtigt, at der er professionelle historikere der kan tage beslutninger om og klart definere, hvilke parametre der skal visualiseres.

Der kræves tværfaglige, brede samarbejder mellem eksperter for at forstå og analysere de billeder visualiseringer producerer.

En af årsagerne til jeg har valgt at arbejde med grupperinger af data i stedet for data på individuelt niveau i visualiseringerne i opgaven her, er netop, at jeg som enkeltperson ikke har ressourcerne til at arbejde med data på individ-niveau. Det kræver langt større IT-ekspertise, computerkraft og adgang til software end jeg har til rådighed i denne opgave. Det kræver langt flere kompetencer at håndtere et projekt på individ-niveau, hvor de grundlæggende datamængder ville være væsentligt større, end de er i grupperinger.

Uanset om der vil blive besluttet at se på enkelte lande, som her Polen og Tyskland eller om tilgangen vil være, at landegrænser ingen betydning har, da disse vil vise sig i en visualisering under alle omstændigheder, vil datagrundlaget være meget stort. Det kan i princippet være det, der vil være kernen i projektet; at bruge visualisering som metode til at visualisere så mange data som muligt, for derved muligvis at kunne pege på specifikke områder, hvor der er stort potentiale for at finde navne på endnu ikke-registrerede Holocaustofre.

Projektet her har dog med sine visualiseringer på gruppe-niveau vist, at metoden har potentiale blandt andet ved at pege på områder, der kunne være interessante for yderligere undersøgelser.

6.6 Hvordan kan brugen af DVIZ rammesættes?

I en artikel fra 2016 af Jonathan Gray o.a., bliver datavisualiseringer i forskerøjemed analyseret, og Gray opstiller en tredelt heuristisk ramme for hvad der må tages i betragtning, når der bliver læst om, arbejdet med og foretaget forskning med datavisualisering. Udspillet skal ses som et oplæg til hvordan kritisk og reflektiv mediering ved brug af datavisualisering kan rammesættes (Gray, 2016):

1. Medieringen fra verden til data - her er der fokus på kildemateriale og ”rå” data som er lagt til grund for den specifikke visualisering.
2. Medieringen fra data til selve billedet af den grafiske repræsentation af informationerne.

3. Medieringen fra billede til øje og relationer til den sociale, kulturelle og historiske specifikke ”vinkel” datavisualiseringerne er foretaget ud fra.

Denne rammesætning er et ganske godt redskab til at bruge når der er tale om formidling ved hjælp af datavisualisering og ligger i forlængelse af advarsler mod at bruge de datadrevne metoder og frygten for at lade integriteten i de humanistiske fag blive undermineret af en mulig opblomstring af empirisme og pseudo-positivismen (Gray 2016, Kitchin 2013).

Ved at bruge rammerne udstukket af Gray, kan datavisualisering netop leve op til at være en metode hvor der indenfor forskningen kan findes nye informationer.

Selve opbygningen af databasen der ligger til grund for enhver visualisering af den art som projektet her beskæftiger sig med, er den del af arbejdet med visualiseringerne, som har mest betydning for selve resultatet og kvaliteten af analysen. De kilder der bliver hentet information fra skal vælges omhyggeligt og organiseringen, rensning og verificering af data, danner tilsammen en database, som ikke er et passivt, neutralt værktøj for visualiseringerne, men mere et resultat af dybdegående og kritisk forskning. At sammensætte en effektiv og god database, og i dette tilfælde af ekstremt heterogene data, som kan levere gode, klare og gennemskuelige visualiseringer, er en af de største dele af hele arbejdet i og omkring visualisering. Ikke mindst kræves der dybdegående domænekendskab og kendskab til klassisk kildeanalyse (Knowles, Cole & Giordano, 2014).

Her vil jeg igen argumenter for, at det er nødvendigt, fremadrettet, at arbejde mere tværfagligt indenfor forskning, - ikke kun historieforskning, men forskning generelt.

Der er bred konsensus omkring, at al datadreven forskning fordrer domæneekspertise. Det vil sige, at det er nødvendigt at arbejde tværfagligt, så man har de fornødne kompetencer i et team der arbejder med metoder som datavisualisering (boyd & Crawford 2012, Ethington 2007, Gregory & Geddes 2014, Knowles, Cole & Giordano 2014, Brazzo 2018).

For selvom de datadrevne forskningsmetoder som datavisualisering er stærke værktøjer, er de dog ikke stærkere, end den software der bliver fremstillet til brugen. Det er derfor super vigtigt, at der løbende bliver evalueret på de forskellige værktøjer, og at der hele tiden bliver udviklet nye værktøjer som kan følge med de store og hastigt voksende mængder data. Brugergrænseflader som giver historikere mulighed for uden dybdegående IT-kompetencer og uden kendskab til kodning, bliver hele tiden forbedrede og forfinede. Dette kan være både godt og skidt. Har en historieforsker de nødvendige kompetencer til, kritisk at evaluere et stykke software, eller et databaseværktøj, så resultaterne bliver så nøjagtige som mulige?

Mange af de digitale forskningsværktøjer er open source⁶⁰, som for eksempel Datavisualiserings værktøjet Gephi, og ligger frit tilgængelige på internettet. Hvordan undgår man, at de bliver brugt til at fremstille ”fake history”⁶¹?

Videnskaben har velkendte, velafprøvede, traditionelle værktøjer til rådighed til at kontrollere at ny viden ER ny viden, og at denne evt. er tilstrækkelig veldokumenteret til at blive publiceret, så som Peer Reviewing som kan ses som et kvalitetsstempel.

Hele problematikken omkring verificering af forskning publiceret på nettet og grå litteratur, er et stort emne, som ikke har plads her, men det er nødvendigt, at nævne, at helt samme problematik gælder for udnyttelsen af software som ligger tilgængeligt på nettet, og de ”forskningsresultater” der eventuelt kan frembringes af amatører eller andre, der kan have interesse i at publicere ”fake history” indenfor forskellige domæner.

Muligheden for at store datalagre bliver Open Source og kan tilgås fra enhver computer, og i øvrigt fra mange forskellige platforme, er samtidig med til at fremme, at forskning bliver uafhængig af tid og rum. Det er umiddelbart en stor fordel, da det lægger op til langt bedre videndeling, end tidligere, men denne tendens betyder samtidig, at der fremadrettet vil komme udfordringer i forhold til, hvad viden overhovedet er, som nævnt ovenfor - hvem der skal tage beslutninger om dette, hvor denne viden skal opbevares, om der skal restriktioner omkring hvem der skal have tilgang til denne viden og hvem der kvalitetssikrer disse enorme mængder af producerede informationer. Dette er tidligere noget blandt andet universiteter, museer, arkiver og biblioteker har haft procedurer omkring og fagfolk til at håndtere – det er ikke en option mere, da datamængderne er for store, og ressourcerne slet ikke matcher udfordringen.

Det kan samtidig være langt sværere at finde fejl og problemer i dataanalyse visualiseret ud fra store data-corpora end at udføre klassiske statistiske tests, der har forudbestemte regler om data. Datadreven forskning med datavisualisering som metode, er en langt mindre sort/hvid metode hvor det som eksempel gælder om at finde elektroniske bugs i software. At finde frem til fejl i dataanalyse er mere komplekst og komplicerede; dette lægger op til en debat om, at de digitale

⁶⁰ Open Source software bliver udviklet indenfor alle domæner, og er frit tilgængeligt for alle.

⁶¹ ”Fake history” ses her som pendant til ”Fake News” som er bevidst manipulerede nyheder, som bliver publiceret som værende verificerede nyheder.

værktøjer fremadrettet måske bør udvikles indenfor nogle faste rammer, der nok giver en ikke-ekspert mulighed for at bruge disse metoder, men også gør det klart at data ikke kan stå alene men kræver indgående analyse.

6.7 Digitale Research Infrastrukturers betydning for DVIZ metode

Der er i løbet af de senere år kommet adskillige store internationale forskningsplatforme online. Disse kaldes Research Infrastructures (RI) og har stor vigtighed for fremtidens forskning generelt, RI, Virtual Research Environments (VRE) og specielt om de helt nye tiltag med tværfaglige understrukturer, som giver forskere mulighed for at sidde forskellige steder i verden og samarbejde online og i realtid via netværksteknologi – uden af skulle forholde sig til tid og rum.

Jeg vil argumentere for, at det er oplagt, at disse digitale infrastrukturer fremadrettet vil spille en stor rolle i forskning indenfor alle fag, - ikke mindst indenfor holocaustforskning, hvor der informationer og data er spredt over så mange forskellige lande, sprog og fragmenteringer. Gennem de store platforme, vil det fremadrettet også være muligt at trække på de tværfaglige ressourcer som datadrevet forskning kræver.

” An important condition for making a success of EHRI is having an interdisciplinary structure. The scholars involved in EHRI are not only historians but also archivists and digital research infrastructure specialists. The collaboration between these specialists is crucial to EHRI’s ambition. EHRI will be an example and forerunner for digital research in a wide range of new disciplines in historical research and archival practices.”

Kahn, R. (2011) The EHRI Project: building an online archive for European Holocaust research. SCOUNL Focus 52 2011. s. 22

Med mulighed for at netværke, opbevare og søge informationer i organiserede databaser, bruge fælles software – og blive undervist i brugen af dem via e-learning og dele viden og information indenfor ikke bare samme domæne, men også tværfagligt, åbner der sig helt nye muligheder for at samle store mængder data fra perioden til brug i DVIZ. Der kommer hele tiden nye data til, også indenfor holocaustforskning, da der er stort fokus på at digitalisere originale kilder. Når først materialerne er digitaliseret, vil det være oplagt, at forskere igen uanset hvor og hvornår, vil kunne tilgå data, og arbejde sammen om brugen disse.

Der er flere digitale tiltag i gang med at kortlægge og bringe nye perspektiver til forskellige områder af Holocaust.

Ud over det meget interessante projekt af Sagi fra 2017, som der i øvrigt arbejdes videre på dels gennem en afholdt Hackatron⁶² arrangeret af Yad Vashems forskerafdeling i Jerusalem, og dels ved at implementere flere dele af Sagi modellen i Yad Vashems Databaser er der flere forskellige tiltag i gang indenfor Holocaustforskning.

Et af mange eksempler er projektet *Mapping The Lives*; et digitalt initiativ, under udførelse af Tracing The Past⁶³, som ved hjælp af et avanceret HGIS vil prøve at samle personlige data i et enkelt system, som så vil blive lagt tilgængelig online og som ved hjælp af blandt andet metoder som crowdsourcing⁶⁴, hele tiden vil udvides og ændres. En form for Wiki metode, men med fokus på at validere oplysninger, før de bliver offentligt tilgængelige. Hvordan, hvem og hvilke metoder der skal bruges i denne sammenhæng, er på tidspunktet for denne opgave ikke oplyst nogen steder i projektets materiale.

Der er rigtig mange andre digitale projekter i gang og på tegnebrættet inden for holocaustforskning. Jeg vil ikke komme ind på dem, men gøre opmærksom på, at jo mere materiale der er digitalt, - jo bedre kan visualiseringsmetoden gøres.

De databaser som jeg her har arbejdet med, vil eventuelt kunne bruges til at danne fundament for yderligere forskning omkring registrering af ofre. De digitale muligheder for at arbejde uafhængigt af tid og rum, kan forebygge dobbeltarbejde. Som allerede nævnt, er hele opbygningen af databasen, den mest tidskrævende faktor i denne type projekt. Derfor vil det kun være en fordel, at andre vil kunne bygge videre på denne. Det vil selvfølgelig kræve, at de forskere der benytter databasen og arbejder på den, dokumenterer deres arbejde.

RI giver også god mulighed for at finde frem til netop den type projekt, man som forsker kunne være interesseret i – og lægger også op til serendipitet indenfor forskning, som igen vil kunne tilføre innovation og nytænkning.

Der har i mange år eksisteret online databaser for information af forskellig art, som for eksempel Web of Science, som forskere stadig bruger flittigt, til at finde relevant litteratur om deres specielle

⁶² En Hackatron er kort beskrevet en begivenhed, hvor man inviterer gæster med digitale/IT-kompetencer, som så får stillet specifik software til rådighed, med henblik på at vurdere hvordan softwaren er bygget, hvad det kan bruges til, samt kvaliteten af den.

⁶³ <https://www.tracingthepast.org/>

⁶⁴ Anvendelse af internettet til at effektivt at outsource en aktivitet til en stor gruppe mennesker og trække på deres viden og kreativitet. Et typisk fænomen i såkaldt. Informationsordbogen.dk – set d. 2/7/2018 kl. 9:00

emner. Disse databaser indeholder vigtige materialer og grundlæggende ressourcer for forskere fra alle domæner, men fremadrettet, må den type databaser være bare en del af større infrastrukturer. En oplagt mulighed for disse RI vil være at samle emner for forskning. Kunne man eventuelt forestille sig, at et fremtidigt projekt for at finde ikke-registrerede holocaustofre, kunne bruge en af disse platforme, og lægge op til samarbejde mellem forskellige interesserede forskere og opfordre til at dele data.

6.7.1 EHRI – en digital forsker infrastruktur.

En af disse RI er European Holocaust Research Infrastructure – EHRI⁶⁵, - et projekt under EU der startede op i 2010 og siden da har arbejdet på at oprette et digitalt – men også menneskeligt fagligt netværk for Holocaustforskning i Europa.

For at skabe en form for overblik over de mange lokale museers, bibliotekers og arkivers materialer, forsøger EHRI at skabe en platform, hvor forskere og andre interesserede, kan søge Holocaust relateret materiale og sparre med andre forskere på området. EHRI organiserer seminarer om forskellige temaer forskellige steder i Europa, og lægger vægt på, at unge forskere har mulighed for at træffe hinanden og diskutere deres projekter.

Projektet er finansieret af EU. Projektets første fase varede indtil 2015 og bestod da af over tyve kerneorganisationer fra 17 forskellige lande. Planen er pt. at der skal støde yderligere partnere fra både Syd- og Østeuropa samt USA i 2019.

Ideen med projektet, er at udvikle en struktur, der kan fungere som en digital platform; en form for portal, hvor interesserede forskere og andre, kan finde frem til lokale materialer og databaser. Infrastrukturen har ikke selv databasestrukturer, som kan tilgås online, men er mere et redskab til at finde frem til de mange fragmenterede materialer der er spredt for alle vinde (Kahn 2011, Hughes 2016, Adaret 2017). Den fungerer som en portal ved at samle navne og links til de arkiver, forskningsinstitutioner, museer og biblioteker der har materiale ét sted søgbare og som har en underliggende fælles semantik (Sagi 2017) som gør det lettere for forskere fra mange forskellige fag, at finde frem til ønsket materiale.

⁶⁵ <https://ehri-project.eu/> set d. 2/7/2018 kl.. 10,50

Der er også mulighed for at følge med i EHRI's "Latest News" kanal, hvor der løbende bliver udbudt seminarer, efterspurgt ny litteratur, og publiceret datoer for møder og konferencer. EHRI udbyder selv Fellow Ships løbende, til Ph.d. studerende og individuelle forskere. Det er også muligt at komme med indlæg og kommentarer på deres blog. Et ganske nyt tiltag, er online kurser indenfor Holocaustforskning.

EHRI har en online blog på sin hjemmesides forside, hvor alle emner indenfor holocaustforskning bliver taget op.

Et af de nyere blogindlæg⁶⁶, præsenterer et nyudviklet digitalt software til at formatere digitaliserede dokumenter til Optical Character Recognition (OCR) format, som kan lave søgninger i disse og som fremadrettet vil blive brugt som springbræt til at udnytte machine-learning. Det er stadig på ide-niveau, men lægger op til, at selve bloggen bliver brugt til forskning. Ikke bare til formidling af information, men til konkret forskning, hvor mange forskellige forskere inviteres til at deltage i diskussionen (Brazzo, 2018).

Platforme som EHRI er vigtige tiltag, som blandt andet gør det muligt at informere om og gøre brug af de mange nye digitale værktøjer der hele tiden bliver udviklet til høst og analyse af data.

Da de historiske materialer er fra en tid hvor intet er digitalt født, er selve digitaliseringen af materialerne et af de største bump på vejen for at bruge dataanalyse til at finde frem til ofre. Mange dokumenter er en kombination af håndskrevne og maskinskrevne oplysninger, og selv om de maskinskrevne oplysninger kan konverteres til digitalt format ved hjælp af forskelligt avanceret OCR software, er det stadig en stor udfordring, at mange af de originale kilder er håndskrevne dokumenter – og ikke bare håndskrevne, men formularer hvori der er en kaotisk sammenblanding af både trykt tekst, maskinskrevet tekst, håndskrevet tekst, stempler og billeder. Derudover er materialerne, som tidligere nævnt på flere forskellige sprog, idet Holocaust fandt sted over det meste af Europa. Her kunne det være en fordel at udlicitere til Crowdsourcing, som for eksempel Det Kongelige Bibliotek har gjort med succes, med flere forskellige projekter⁶⁷.

Selv har jeg i denne opgave kun brugt EHRI en smule, og specielt i den indledende fase, hvor der skulle findes passende data.. Da forskerinfrastrukturen fungerer som en form for portal, giver den

⁶⁶ <https://blog.ehri-project.eu/2018/03/29/quod/> set d. 28/8/2018 16,50

⁶⁷ <http://www.kb.dk/da/nb/materialer/luftfoto/about.html>

ikke direkte adgang til databaser, bortset fra de arkiver, museer og lignende der har valgt at linke til deres hjemmeside.

Jeg må tilføje, at selvom EHRI har relativt gode søgefunktioner og en let forståelig brugergrænseflade, var det ikke let at komme i gang med EHRI. Det tog tid at finde ud af hvordan informationer var strukturerede, og hvordan man i øvrigt fik mulighed for at tilgå disse.

Som jeg ser det, er de store forskerinfrastrukturer gode til videndeling ved hjælp af fælles kartoteker, blogs og lignende, og bestemt også til brug i mere klassisk, kvalitativ forskning, hvor der er brug for at finde data på mikro-niveau.

6.8 Serendipitet og datavisualisering

Der er flere gange i denne opgave nævnt begrebet serendipitet. Begrebet fortjener her at blive uddybet, da netop indenfor datavisualisering, er der store muligheder for at finde nye informationer og dermed ny viden, da man ser data fra nye vinkler.

Kort kan man forklare serendipitet med uplanlagte fund, hvilket vil sige, at man ved at lede efter noget, gå på opdagelse i, eller bare være nysgerrig generelt, falder over nogle informationer eller viden man ikke umiddelbart var på udkig efter (Björneborn, 2008).

Måske specielt indenfor Holocaustforskning, hvor størstedelen af data ikke er digitalt født, men stammer fra et stort heterogent kaos af materialer, er der muligheder for at finde nye informationer ved hjælp af datavisualisering. Meget af det ikke digitalt-fødte materiale er blevet digitaliseret med et specifikt formål for øje.

Netop her kan visualisering være et værktøj, der åbner for nye muligheder. Når data først er digitaliseret, er der mange muligheder for visualiseringer. Og her kommer serendipitet ind i billedet. Specielt når der tales om meget store data-corpora, som for eksempel sammenkørsel af disse ukurante og forskelligartede heterogene data, hvoraf mange har geografiske informationer tilknyttet, lægger i visualisering op til serendipitet (Mikkelsen, 2016).

I processen med bearbejdning af data, analyse og visualisering af store datasæt, vil der være store muligheder for at der kan vise sig mønstre, der ikke specifikt blev kigget efter, men som ”dukkede” op i den store mængde af data (Knowles, Cole & Giordano). Netop hermed argumenterer jeg for, at også serendipitet er et af de begreber, der knytter sig tæt til brugen af datavisualisering – specielt indenfor historieforskning.

I visualiseringerne foretaget i denne opgave, kan der argumenteres for, at der dukket mønstre op, der kan pege på tendenser, som ikke umiddelbart var information jeg havde forudset, eller havde regnet med at finde. At der er flere registrerede ofre i NDB end i de andre datasæt der ligger til grund for visualiseringerne, er ikke et øjeblik faldet mig ind, før jeg så visualiseringerne. Dette viser klart potentialet for datavisualisering, selv når der kun er brugt små datasæt der er grupperede på by-niveau. Denne form for serendipitet er specifikt, hvad jeg mener er datavisualiserings stærkeste side. Med store datasæt, er det absolut ikke muligt at skabe overblik, se sammenhænge, mønstre, korrelationer eller lignende på andre måder! Netop derfor, er det min overbevisning, at datavisualisering fremadrettet bliver en metode, der vil blive benyttet flittigt indenfor alle fag – ikke mindst indenfor historieforskning, - specielt i takt med, at flere og flere ikke digitalt-fødte materialer bliver digitaliseret. Når først materialet ligger som digitale data, er det kun et spørgsmål om tid og ressourcer, før det vil blive visualiseret.

7. Konklusion

Ved hjælp af forskellige visualiseringer foretaget med softwareprogrammet Tableau, er der i projektet vist mønstre, korrelationer og andre tendenser, der kan give ny viden om antallet af dels jødiske indbyggere registreret i den Polske Census fra 1931 og dels holocaustofre der er registreret i YadVashems navnedatabase. Visualiseringerne peger på geografiske områder, hvor det ville være realistisk, på mikro-niveau, at koncentrere ressourcer for at søge efter navne på ofre, der endnu ikke er registreret, og derved bruge datavisualisering i eftersøgning af de over en million manglende navne på holocaustofre.

Der er ligeledes fokus på processen, for at argumentere for metoden, som i projektet har vist sig at være en effektiv metode til brug i Holocaustforskning, specielt ved brug af store data-corpora, hvor mønstre, clusters, afvigelser, tendenser og korrelationer vil træde frem.

Opgaven viser, at datavisualisering er et stærkt digitalt redskab, som Holocaustforskerne såvel som historieforskere generelt ikke kan tillade sig at se bort fra. Metoden egner sig til at finde nye informationer, se historiske begivenheder fra nye vinkler og et nyt perspektiv, men også som værktøj til at verificere allerede kendt viden og som her være med til at generere ny viden, ved at pege på specifikke områder hvor der kan findes ikke-registrerede holocaustofre. Brug af metoden til at finde frem til områder hvor der stadig er mange ikke-registrerede ofre, vil derfor kunne benyttes

med fordel, som grundlæggende for flere undersøgelser på mikro-niveau, som kan finde frem til specifikke ofre.

Metoden giver samtidig mulighed for at anskue et givent fænomen på makro-niveau for derefter at kunne analysere eventuelt fremkomne mønstre og sammenhænge. Det er derefter vigtigt, at fagspecialister indenfor det respektive domæne bruger analyserne som grundlag for yderligere analyse af visualiseringerne, nu på mikro-niveau hvor analyse, baseret på dybdegående undersøgelse af et givent fænomen, er grundlæggende. Visualiseringer kan pege på tendenser, korrelationer og mønstre, hvorefter der vil kunne stilles nye spørgsmål og der vil kunne bygges nye algoritmer, der kan høste nye data, som kan føre til ny viden indenfor historieforskningen.

I processen med bearbejdning af data, analyse og visualisering af store data-corpora, vil der være store muligheder for, at der kan vise sig mønstre, der ikke specifikt blev kigget efter, men som ”dukker” op i den store mængde af data. Dette ses som argument for, at også serendipitet er et af de begreber, der knytter sig tæt til brugen af datavisualisering. Der ligger stort potentiale gemt i de enorme mængder data, der hele tiden produceres i det digitale samfund; og ikke bare det som bliver født digitalt, men også de store mængder historisk arkivmateriale som løbende bliver digitaliseret. Projektet her peger igen her på, at meget store data-corpora kan give mere information end visualisering af grupperet data, som her er anvendt.

Visualiseringerne viser et helt klart mønster i differencen mellem de to datasæt i forhold til visualiseringerne fra Polen, hvor der ses store forskelle mellem datasættene. En ganske stor overraskelse er det specielt, at der generelt ses langt flere registrerede i YadVashems navnedatabase (NDB) end i den Polske Census fra 1931. Der er derfor foretaget yderligere visualiseringer hvor NDB bliver holdt op mod et Tysk datasæt fra 1939, hvor data er mere akkurate, nyere og tættere på perioden for NDB. Disse viser samme tendenser, dog i noget mindre grad. Dette finder jeg som argument for, at der muligvis kan være en del redundans i NDB, mere end hidtil anslået.

Ved brug af visualiseringer som forskningsmetode, hvor visualiseringerne er baseret på mindre, grupperede datasæt, bør man derfor være opmærksom på udfordringen med redundans i data. Dette har vist sig at spille en ganske afgørende rolle i visualiseringer af forholdsvis små datasæt, som skal give svar på specifikke spørgsmål. Til gengæld er det i visualiseringerne af Big Data med store uorganiserede datasæt mindre kritisk, da de viser billeder fra fugleperspektiv, og lægger op til at vise mønstre og korrelationer – ikke specifik information.

Datavisualiseringer som metode, kan ikke svare på spørgsmål om hvorfor - men kun vise, at der er en tendens. Årsager vil kræve undersøgelser med andre mere kvalitative metoder. Til gengæld, kan metoden med fordel bruges til at understøtte andre, kvalitative metoder.

8. Perspektivering

I dag er der mange gode projekter og initiativer i gang hvor digitaliserede data ligger til grund indenfor Holocaustforskning. Som eksempler kan nævnes nogle af de større: Stanfords ”Spatial History Project”⁶⁸ som er grundlæggende for *Geographies of The Holocaust* (Knowles, Cole & Giordano), ”Holocaust Atlas of Lithuania”⁶⁹, ”Forced Labor 1939 – 1945” og andre.

Også på YadVashem og gennem EHRI, er der hele tiden nye initiativer, som bygger på datadreven forskning. På mange af de lokale arkiver bliver der forsket ved hjælp af datadrevne metoder. De store internationale platforme, ville kunne gøre en stor forskel for forskningen, ved at opfordre til at alle disse små og allerede strukturerede data kunne lægges i fælles arkiver. Det er grundlæggende i forlængelse af hvad EHRI allerede gør, - de henviser til de steder hvor data ligger, - kunne man ikke hente data i stedet? Et stort projekt, ja – men absolut et som kunne gøre en forskel.

Der vil i fremtiden givetvis komme flere digitale projekter til, og dermed komme mere digitaliseret materiale om Holocaust – datadreven forskning forstærker sig selv. Ikke mindst metoden Crowdsourcing, kommer helt sikkert til at spille en stor rolle i denne digitaliseringsproces. Med de store data-corpora der uomtvisteligt vil blive samlet, vil der være brug for metoder til at arbejde med, analysere og verificere disse data. Som metode, vil datavisualisering være bare en af mange metoder, som må bruges indenfor Digital Humanities.

Der er i denne opgave fokus på datavisualisering i Holocaust. Det er forholdsvis gængse datavisualiseringer, hvor der er brugt klassiske, kendte softwareprogrammer til at foretage visualiseringer af kvantitative data bestående af numeriske optegnelser, som generelt relaterer til de fysiske sider af fænomenet (Kitchin, 2014).

⁶⁸ <https://web.stanford.edu/group/spatialhistory/cgi-bin/site/project.php?id=1015> set 2/9/18 12,30

⁶⁹ <http://www.holocaustatlas.lt/EN/> set 2/9/18 12,35

Det er dog på sin plads her at nævne, at der indenfor mange områder udvikles nye metoder til at arbejde med de store mængder data der hober sig op - ikke bare visualiseringsmetoder, men også software som er baseret på genkendelse, som også Machine Learning er baseret på.

Indenfor holocaustforskning vil det, i takt med at denne type software bliver mere og mere nøjagtigt, være oplagt at tage disse typer metoder i brug. Man kunne for eksempel forestille sig, fremtidige projekter, hvor ansigtsgenkendelsessoftware⁷⁰ kunne bruges til at finde frem til ikke-registrerede ofre i NDB. Dette kræver selvfølgelig, at al originalt kildemateriale er digitaliseret og scannet ind, men også metoderne til denne del af processen, er under hastig udvikling og kan i fremtiden lette dette arbejde.

Et yderst interessant fremtidigt projekt kunne være at digitalisere de geografiske kort og data fra Martin Gilberts *Atlas of The Holocaust* (Gilbert, 1989). Med disse digitaliserede data, ville der kunne komme mere viden ud af data, end der kan ud fra de håndtegnede kort. En ide, som forhåbentlig bliver en realitet i fremtiden.

*“A spring day the smell of lilac
Between the ruins of your city
A beautiful day to fish in the river
Inside me my heart is broken
There it was and it wasn't
Your child is a small woman
People that no-one knows
There isn't even a house that you'll remember

And if you're going, where are you going
Forever is just ashes and dust
Where are you going, where are you going
Years and nothing is erased...”*

Yehuda Poliker- barn af Holocaust overlevende – fra albummet *Ash and Dust* 1988

⁷⁰ https://en.wikipedia.org/wiki/Facial_recognition_system set 2/9/18 kl. 9,00



Litteratur

- Aderet, O. (2017) An ambitious project ID'd 80% of Hungarian Jews murdered in the Holocaust. (2017) *Haaretz*, October 21st, 2017
- Barnes, T. J. (2013) *Big data, little history*. *Dialogues in Human Geography*. 3(3) s. 297-302
- Beck Holm, A. (2011): *Videnskab i virkeligheden – en grundbog i videnskabsteori*. Samfundslitteratur, Frederiksberg C.
- Bjørneborn, L. (2008). Serendipitetsfaktorer og brugeradfærd på det fysiske bibliotek. *Dansk Biblioteksforskning*, 4(2): 41-54
- boyd & Crawford (2012). Critical Questions for Big Data. *Information, Communication & Society* 15, no. 5 s. 662-679
- Brazzo, L. (2018) Reto Speck, "Introduction" in Holocaust Research and Archives in the Digital Age, ed. Laura Brazzo, Reto Speck, Quest. Issues in Contemporary Jewish History. Journal of Fondazione CDEC, n. 13 August 2018
- Dieckmann, C. (2001) Alytus 1941-1944: Massenmorde in einer Kleinstadt - Ein Fallbeispiel deutscher Besatzungspolitik in Litauen. *Lithuanian Foreign Policy Review*, 200. <http://lfpr.lt/wp-content/uploads/2015/07/LFPR-8-Dieckmann.pdf> (set 24/8/2018) s. 1-28
- Drucker. J. (2014) *Graphesis – Visual Forms of Knowledge Production*. Cambridge Mass. Harvard University Press.
- Edelstein, D., Findlen, P., Ceserani, G., Winterer, C., & Coleman, N. (2017) Historical Research in a Digital Age- Reflections from the Mapping the Republic of Letters Project. *AHR Forum*. Published by Oxford University Press on behalf of the American Historical Association.

Earley-Spadoni, T. (2017) Spatial History, deep mapping and digital storytelling- archaeology's future imagined through an engagement with the Digital Humanities. *Journal of Archaeological Science* 84 (2017)

Ethington, P. J., Placing the Past: 'Groundwork' for a Spatial Theory of History (2007) - *Rethinking History*. Vol. 11, No. 4, december 2007, pp. 465 – 493

Finnemann, N. O. (2014) Big Data – et nyt humanvidenskabeligt forskningsfelt. Tiltrædelsesforelæsning som professor i Internetforskning og Digital Humanities ved Det Informationsvidenskabelige Akademi, Københavns Universitets Humanistiske Fakultet, d. 21.11.2014.

Fisher, D. Popov, I., Drucker, S., Schraefel (2012) Trust me, I'm partially right: Incremental visualization lets analysts explore large dataset faster. *CHI 2012, May 5-10, 2012, Austin, Texas, USA*

Frické, Martin (2014). Big data and its epistemology. *Journal of the Association for Information Science and Technology*, 66(4), 651-661. DOI: 10.1002/asi.23212

Friendly M. (2008) A Brief History of Data Visualization. In: *Handbook of Data Visualization*. Springer Handbooks Comp.Statistics. Springer, Berlin, Heidelberg.

Gilbert, M. (1982) *The Atlas of the Holocaust*. 4. Edition. Routledge. 711 Third Avenue, New York NY 10017

Gray, J., Bounegru, L., Milan, S., & Ciuccarelli, P. (2016) Ways of Seeing Data- Toward a Critical Literacy for Data Visualizations as Research Objects and Research Devices. Forthcoming in Sebastian Kubitschko and Anne Kaun (eds.) *Innovative Methods in Media and Communication Research*. London: Palgrave Macmillan S. 227 – 251.

Gustavsson, B. (2000) *Vidensfilosofi*. Forlaget Klim, Århus N. 1. udg. 3. oplag, 2011

Gregory & Geddes, I. N. (2014) *Toward Spatial Humanities: Historical GIS and Spatial History*. Indiana University Press. Bloomington, Indiana USA

Hughes, L., Constantopoulos, P., Dallas, C. (2016) Digital Methods in the Humanities: Understanding and Describing their Use across the Disciplines. *A New Companion to Digital Humanities*, First Edition. Edited by Susan Schreibman, Ray Siemens, and John Unsworth 2016 John Wiley & Sons.

Janicka E. (2015) The Square of Polish Innocence: POLIN Museum of the History of Polish Jews in Warsaw and its symbolic topography, *East European Jewish Affairs*, 45:2-3, 200-214, DOI: 10.1080/13501674.2015.1059246

Kahn, R. (2011) The EHRI Project: building an online archive for European Holocaust research. *SCONUL Focus* 52 2011

- Kitchin, R. (2013) Big data and human geography: Opportunities, challenges and risks. *Dialogues in Human Geography* 3(3) 262–267
- Kitchin, R. (2014) *The Data Revolution, Big Data, Open Data, Data Infrastructures & their consequences*. Sage Publications, Los Angeles, USA
- Kitchin, R. Dodge, M. (2007) Rethinking maps. *Progress in Human Geography* 31(3) 2007.pp. 331-344
- Knorr Cetina, k. (2009) The Synthetic Situation: Interactionism for a Global World. *Symbolic Interaction* 32. 61-87 2009.
- Knowles, A. K., Cole, T., Giordano, A. (2014) *Geographies of the Holocaust - Series: The Spatial Humanities* - Copyright Date: 2014 - Published by: Indiana University Press
- Kristensen, B. E. (2007) *Historisk metode*. 1. udg. 1. oplag. Hans Reitzels Forlag. København s. 9-55.
- Kudym, S. R. (2018) Holocaust Mobilities: Journeys Through Minsk. *Masters Thesis - University of Nebraska at Omaha, ProQuest Dissertations Publishing*, 2018.
- Mayer-Schönberger, V., Cukier, K. (2014) *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Mariner Books, New York
- Mikkelsen, L. M. G. (2015) Eksamensopgave i Digitale Videnssystemer. IVA forår 2015
- Mikkelsen, L. M. G. (2015) Eksamensopgave i Videnskabsteori og Metode. IVA forår 2015
- Mikkelsen, L. M. G. (2016) Bacheloropgave i Informationsvidenskab og Kulturformidling. IVA forår 2016
- Miller, R. et.al. (2015) Project:” Mapping the Lives”. *Tracing the Past e.V. 2015*.
<https://www.tracingthepast.org/mappingthelives>
- Sagi, T. et.al. (2017) Multi-source uncertain entity resolution: Transforming Holocaust victim reports into people. *Information Systems* 65 (2017) 124–136
- Stone, D. (2004) *The Historiography of the Holocaust*. Pelgrave Macmillan, 2004.
- Szewrań S., Kazak, J. Sylla, M., Gorzata Ś M. (2017) *Spatial Data Analysis with the Use of ArcGIS and Tableau Systems*. Springer International Publishing AG 2017 I. Ivan et al. (eds.),
- Van Eijnatten, J., Pieters, T. & Verheul, J. (2013) Big Data for Global History. The Transformative Promise of Digital Humanities. *BMGM - Low Countries Historical Review*, Volume 128-4 (2013), pp. 55-77
- White, R. What is Spatial History? (2010). *Working paper; Submitted 1 February 2010*. Spatial History Lab at Stanford University

Winstone, M. (2015) *The Holocaust Sites of Europe*. I. B. Tauris & Co. Ltd, 6 Salem Road, London W2 4BU.

Yau N. 2013. *Data Points – Visualization that means something*. John Wiley & Sons, Inc. 10475 Crosspoint Boulevard, Indianapolis, IN46256.

Tableau Public – min profil:

<https://public.tableau.com/profile/lone.maor.guldborg.mikkelsen#!/vizhome/holocaustvictimsfinalversion/combined?publish=yes>

Bilag:

Bilag 1 – Udsnit af Polish Census1931

32 WOJEWÓDZTWO KIELECKIE

Tabl. 11. Ludność według płci i wyznania — Population d'après le sexe et la confession

Powiaty (Miasta i wieś, Miasta razem, Miasta o ludności powyżej 20 000. Wieś)

| Powiaty Districts | Ludność ogółem Population totale | Wyznanie — Confession | | | | | | | | | | | | |
|-------------------------|---|---|---|------------------------------------|------------------------------------|-----------------------------------|-----------------------------------|---|---|--|---|---|---|--|
| | | rym- skoka- tołockie i or- mia- skoka- tołockie cath. romane et cath. armé- nienne | grecko- katoli- ckie i ortry- dok- skań- skie cath. grecque et rita oriental de l'Église cathol. | prawo- sławne ortho- doxe | ewangeliczne — protestante | | | bez biskup- stwa catho- lickie non peccidae | inne chrze- ścijań- skie autres confes- sions chré- tiennes | inne nie- chrze- ścijań- skie autres confes- sions non chré- tiennes | nie okre- szone ouza bez wy- zna- nia non précisé et popom- mes non confes- sion | nie okre- szone ouza bez wy- zna- nia non précisé et popom- mes non confes- sion | nie okre- szone ouza bez wy- zna- nia non précisé et popom- mes non confes- sion | |
| | | | | | augu- stynie huthé- ranne | refo- rmowane réfor- més | inne autres unio- naires | | | | | | | |
| Będzin | r 231 310 | 198 211 | 101 | 365 | 180 | 33 | 4 | 95 | 107 | 31 875 | 1 | 21 | 317 | |
| | m 114 135 | 98 074 | 57 | 2031 | 92 | 19 | 3 | 40 | 48 | 15 496 | 1 | 14 | 148 | |
| | k 117 175 | 100 137 | 44 | 162 | 88 | 14 | 1 | 55 | 59 | 16 439 | — | 7 | 169 | |
| Miasta razem | r 105 542 | 77 114 | 17 | 183 | 108 | 24 | — | 63 | 82 | 27 852 | 1 | 16 | 62 | |
| Willes ensemble | m 51 481 | 37 733 | 10 | 96 | 55 | 14 | — | 27 | 37 | 13 465 | — | 10 | 25 | |
| | k 54 061 | 39 384 | 18 | 87 | 53 | 10 | — | 36 | 45 | 14 387 | — | 6 | 37 | |
| m. Będzin | r 47 597 | 35 790 | 21 | 45 | 26 | 11 | — | 10 | 16 | 21 625 | — | 10 | 35 | |
| | m 22 818 | 12 314 | 9 | 25 | 15 | 4 | — | 9 | 6 | 10 447 | — | 6 | 13 | |
| | k 24 779 | 13 476 | 12 | 20 | 11 | 7 | — | 10 | 9 | 11 198 | — | 4 | 22 | |
| m. Cieladź | r 21 003 | 19 820 | 8 | 28 | 21 | 5 | — | 14 | 5 | 1 077 | 1 | 1 | 8 | |
| | m 10 557 | 9 969 | 6 | 18 | 12 | 4 | — | 4 | 4 | 525 | — | — | — | |
| | k 10 446 | 9 851 | 2 | 10 | 9 | 1 | — | 10 | 1 | 552 | — | — | — | |
| m. Dąbrowa | r 36 942 | 31 502 | 8 | 110 | 61 | 8 | — | 30 | 63 | 5 180 | — | 5 | 6 | |
| Górnica | m 18 091 | 15 439 | 4 | 53 | 28 | 6 | — | 14 | 27 | 2 513 | — | 3 | 4 | |
| | k 18 851 | 16 063 | 4 | 57 | 33 | 2 | — | 16 | 35 | 2 637 | — | 2 | 2 | |
| Wieś | r 125 768 | 121 097 | 64 | 182 | 73 | 9 | 4 | 32 | 25 | 4 023 | — | 3 | 255 | |
| Communes rurales | m 62 654 | 60 342 | 38 | 107 | 37 | 5 | 3 | 13 | 11 | 1 973 | — | 4 | 123 | |
| | k 63 114 | 60 755 | 26 | 75 | 35 | 4 | 1 | 19 | 14 | 2 052 | — | 1 | 132 | |
| M. Częstochowa | r 117 179 | 90 843 | 74 | 835 | 443 | 29 | 3 | 260 | 37 | 25 588 | 3 | 19 | 45 | |
| | m 58 954 | 41 165 | 38 | 170 | 207 | 17 | 2 | 186 | 19 | 12 188 | 3 | 10 | 22 | |
| | k 63 225 | 49 178 | 36 | 165 | 236 | 12 | 1 | 174 | 18 | 13 423 | — | 9 | 23 | |
| Częstochowa | r 182 466 | 174 127 | 48 | 82 | 575 | 5 | 1 | 182 | 80 | 6 927 | — | 14 | 427 | |
| | m 89 639 | 85 515 | 27 | 53 | 282 | 2 | 1 | 96 | 36 | 3 425 | — | 10 | 189 | |
| | k 92 827 | 88 612 | 21 | 29 | 291 | 3 | — | 86 | 44 | 3 499 | — | 4 | 238 | |
| Miasta | r 14 461 | 11 260 | 2 | 18 | 2 | — | — | 1 | 4 | 1 161 | — | — | 13 | |
| Villes | m 7 101 | 5 499 | 2 | 11 | 1 | — | — | 1 | 3 | 1 577 | — | — | 7 | |
| | k 7 360 | 5 761 | — | 7 | 1 | — | — | — | 1 | 1 584 | — | — | 6 | |
| Wieś | r 168 005 | 162 867 | 46 | 64 | 571 | 5 | 1 | 181 | 76 | 3 766 | — | 14 | 414 | |
| | m 82 538 | 80 016 | 25 | 42 | 281 | 2 | 1 | 95 | 33 | 1 851 | — | 10 | 182 | |
| | k 85 467 | 82 851 | 21 | 22 | 290 | 3 | — | 86 | 43 | 1 915 | — | 4 | 232 | |
| Ilza | r 159 718 | 147 062 | 51 | 105 | 388 | 27 | 4 | 241 | 56 | 11 377 | 1 | 6 | 348 | |
| | m 79 046 | 72 901 | 39 | 60 | 179 | 13 | 3 | 120 | 57 | 6 486 | — | 2 | 175 | |
| | k 80 672 | 74 161 | 22 | 46 | 209 | 14 | 1 | 121 | 41 | 5 881 | — | 4 | 173 | |
| Miasta | r 12 432 | 8 150 | 15 | 20 | 1 | 1 | — | 6 | — | 4 330 | — | — | 9 | |
| | m 6 089 | 3 985 | 11 | 12 | — | — | — | 5 | — | 2 071 | — | — | 5 | |
| | k 6 343 | 4 165 | 4 | 8 | 1 | 1 | — | 1 | — | 2 259 | — | — | 4 | |
| Wieś | r 147 286 | 138 912 | 46 | 85 | 387 | 26 | 4 | 235 | 98 | 7 147 | 1 | 6 | 339 | |
| | m 72 957 | 68 916 | 28 | 48 | 179 | 13 | 3 | 115 | 57 | 3 485 | — | 3 | 175 | |
| | k 74 329 | 69 996 | 18 | 37 | 208 | 13 | 1 | 120 | 41 | 3 722 | — | 4 | 169 | |
| Jędrzejów | r 108 533 | 98 714 | 21 | 88 | 18 | 6 | — | 4 | 4 | 9 435 | — | 2 | 292 | |
| | m 53 543 | 48 746 | 14 | 23 | 7 | 3 | — | 4 | 2 | 4 587 | — | 1 | 160 | |
| | k 54 990 | 49 968 | 7 | 15 | 11 | 3 | — | — | 2 | 4 848 | — | 1 | 130 | |
| Miasto | r 12 857 | 8 136 | 4 | 23 | 15 | 3 | — | — | — | 4 440 | — | 1 | 13 | |
| | m 6 165 | 3 998 | 3 | 13 | 5 | 3 | — | — | — | 2 138 | — | — | 5 | |
| | k 6 692 | 4 138 | 1 | 11 | 10 | 0 | — | — | — | 2 302 | — | — | 8 | |
| Wieś | r 95 676 | 86 578 | 17 | 65 | 3 | — | — | 4 | 4 | 4 995 | — | 1 | 279 | |
| | m 47 378 | 44 748 | 11 | 11 | 2 | — | — | 4 | 3 | 2 449 | — | — | 151 | |
| | k 48 298 | 41 830 | 6 | 4 | 1 | — | — | — | 1 | 2 546 | — | 1 | 128 | |
| Kielce | r 219 297 | 190 249 | 164 | 881 | 942 | 30 | 8 | 79 | 30 | 26 896 | — | 20 | 580 | |
| | m 106 430 | 92 237 | 109 | 219 | 410 | 22 | 6 | 45 | 15 | 13 092 | — | 27 | 264 | |
| | k 112 867 | 98 012 | 55 | 162 | 423 | 10 | 2 | 47 | 15 | 13 804 | — | 9 | 321 | |

34 WOJEWÓDZTWO KIELECKIE

Tabl. 11. Ludność według płci i wyznania (dok.)

| Powiaty | r — razem m — mężczyźni k — kobiety | Ludność ogółem | Wyznania | | | | | | | | | | | |
|----------------|---|-------------------|------------------------|------------------------------|------------------|--------------|---------------|----------|--------------|--------|------------------------------|--------------------------------------|-----------------------|-----------------------|
| | | | rzymsko- katolickie | protestan- skie i inne | prawo- sławne | ewangelickie | | | | | inne chrześ- cijańskie | inne nie- chrześ- cijańskie | nie okre- szone | nie okre- szone |
| | | | | | | lutherskie | metodystyczne | unijskie | baptystyczne | innych | | | | |
| Opoczno (dok.) | | | | | | | | | | | | | | |
| Wieś | r | 121 618 | 113 887 | 57 | 34 | 1 144 | 5 | — | 124 | 3 | 5 005 | — | — | 459 |
| | m | 59 672 | 55 067 | 30 | 14 | 566 | 3 | — | 66 | 2 | 2 819 | — | — | 205 |
| | k | 61 946 | 57 920 | 27 | 20 | 578 | 2 | — | 58 | 1 | 2 086 | — | — | 254 |
| Pińczów | r | 126 088 | 112 747 | 54 | 33 | 5 | 2 | 1 | 4 | 4 | 12 900 | — | — | 147 |
| | m | 60 603 | 54 132 | 31 | 18 | 2 | 2 | — | 3 | 1 | 6 255 | — | — | 159 |
| | k | 65 485 | 58 615 | 23 | 15 | 3 | — | 1 | 1 | 3 | 6 645 | — | — | 188 |
| Miasta | r | 15 999 | 6 700 | 7 | 8 | 2 | 1 | — | 1 | 1 | 9 175 | — | — | 11 |
| | m | 7 671 | 3 231 | 5 | 3 | 1 | 1 | — | 2 | 1 | 4 423 | — | — | 4 |
| | k | 8 328 | 3 469 | 2 | 5 | 1 | — | — | — | — | 4 752 | — | — | 7 |
| Wieś | r | 110 099 | 105 957 | 47 | 25 | 3 | 1 | 1 | 1 | 3 | 3 724 | — | — | 136 |
| | m | 52 930 | 50 001 | 26 | 11 | 1 | 1 | — | — | 1 | 1 832 | — | — | 115 |
| | k | 57 169 | 55 956 | 21 | 14 | 2 | — | — | 1 | 2 | 1 892 | — | — | 181 |
| M. Radom | r | 77 902 | 51 811 | 64 | 264 | 304 | 44 | 1 | 139 | 2 | 25 159 | 1 | 24 | 70 |
| | m | 37 028 | 24 509 | 28 | 133 | 145 | 23 | — | 63 | 11 | 12 085 | 1 | 23 | 24 |
| | k | 40 874 | 27 302 | 36 | 131 | 159 | 21 | 1 | 76 | 10 | 13 074 | — | 1 | 46 |
| Radom | r | 150 447 | 140 896 | 44 | 43 | 807 | 50 | — | 332 | 6 | 7 774 | 1 | 1 | 523 |
| | m | 74 673 | 69 960 | 25 | 28 | 387 | 31 | — | 155 | 1 | 3 830 | — | — | 255 |
| | k | 75 774 | 70 936 | 19 | 15 | 420 | 19 | — | 177 | 5 | 3 944 | 1 | — | 268 |
| Miasta | r | 4 359 | 3 392 | 3 | — | — | — | — | — | — | 1 011 | — | — | 16 |
| | m | 2 130 | 1 618 | — | — | — | — | — | — | — | 499 | — | — | 10 |
| | k | 2 229 | 1 774 | — | — | — | — | — | — | — | 512 | — | — | 6 |
| Wieś | r | 146 088 | 137 557 | 41 | 43 | 807 | 50 | — | 332 | 6 | 6 743 | — | — | 507 |
| | m | 72 543 | 68 342 | 22 | 28 | 387 | 31 | — | 155 | 1 | 3 311 | — | — | 245 |
| | k | 73 545 | 69 215 | 19 | 15 | 420 | 19 | — | 177 | 5 | 3 432 | — | — | 262 |
| Sandomierz | r | 123 862 | 107 315 | 42 | 75 | 128 | 5 | — | 28 | 1 | 15 427 | — | 3 | 338 |
| | m | 60 045 | 52 379 | 20 | 35 | 66 | 5 | — | 11 | — | 7 855 | — | — | 153 |
| | k | 63 817 | 54 936 | 22 | 40 | 62 | — | — | 17 | 1 | 8 072 | — | 2 | 185 |
| Miasta | r | 20 188 | 11 181 | 18 | 59 | 7 | 4 | — | — | — | 8 681 | — | 3 | 34 |
| | m | 9 808 | 5 624 | 11 | 46 | 3 | 4 | — | — | — | 4 104 | — | 1 | 15 |
| | k | 10 380 | 5 557 | 7 | 13 | 4 | — | — | — | — | 4 578 | — | 2 | 19 |
| Wieś | r | 102 974 | 95 734 | 24 | 16 | 121 | 1 | — | 26 | 1 | 6 745 | — | — | 306 |
| | m | 50 217 | 46 755 | 9 | 9 | 63 | 1 | — | 11 | — | 3 251 | — | — | 138 |
| | k | 52 757 | 48 979 | 15 | 7 | 58 | — | — | 15 | 1 | 3 494 | — | — | 168 |
| M. Sosnowiec | r | 108 969 | 86 606 | 97 | 479 | 566 | 47 | 4 | 164 | 89 | 20 805 | — | 35 | 68 |
| | m | 52 346 | 41 350 | 60 | 220 | 275 | 27 | 1 | 80 | 42 | 10 233 | — | 25 | 33 |
| | k | 56 623 | 45 256 | 37 | 259 | 291 | 20 | 3 | 84 | 47 | 10 572 | — | 10 | 35 |
| Stopnica | r | 153 091 | 135 246 | 58 | 36 | 226 | 4 | — | 4 | 4 | 17 157 | — | 1 | 355 |
| | m | 74 147 | 65 481 | 32 | 16 | 116 | 2 | — | 2 | 2 | 8 520 | — | — | 176 |
| | k | 78 944 | 69 765 | 26 | 20 | 110 | 2 | — | 2 | 2 | 8 837 | — | — | 180 |
| Miasta | r | 12 878 | 5 518 | 18 | 10 | 5 | 1 | — | — | — | 7 312 | — | — | 13 |
| | m | 6 082 | 2 541 | 10 | 6 | 2 | 1 | — | — | — | 3 517 | — | — | 5 |
| | k | 6 796 | 2 977 | 8 | 4 | 3 | — | — | — | — | 3 795 | — | — | 8 |
| Wieś | r | 140 213 | 129 728 | 40 | 26 | 220 | 3 | — | 4 | 4 | 9 845 | — | — | 342 |
| | m | 68 065 | 62 940 | 22 | 10 | 114 | 1 | — | 2 | 2 | 4 803 | — | — | 179 |
| | k | 72 148 | 66 788 | 18 | 16 | 106 | 2 | — | 2 | 2 | 5 042 | — | — | 174 |
| Włoszczowa | r | 100 607 | 92 575 | 32 | 35 | 501 | 11 | 3 | 43 | — | 7 131 | — | 1 | 8 |
| | m | 50 002 | 46 047 | 17 | 20 | 241 | 7 | 1 | 16 | — | 3 526 | — | — | 267 |
| | k | 50 605 | 46 528 | 15 | 15 | 260 | 4 | 2 | 27 | — | 3 605 | — | — | 122 |
| Miasta | r | 13 052 | 7 443 | 4 | 11 | 5 | 9 | — | 12 | — | 5 557 | — | — | 11 |
| | m | 6 439 | 3 678 | 2 | 6 | 3 | 5 | — | 3 | — | 2 735 | — | — | 7 |
| | k | 6 613 | 3 765 | 2 | 5 | 2 | 4 | — | 9 | — | 2 822 | — | — | 4 |
| Wieś | r | 87 555 | 85 132 | 28 | 24 | 496 | 2 | 3 | 31 | — | 1 574 | — | — | 250 |
| | m | 43 563 | 42 369 | 15 | 14 | 238 | 2 | 1 | 13 | — | 791 | — | — | 115 |
| | k | 43 992 | 42 763 | 13 | 10 | 258 | — | 2 | 18 | — | 783 | — | — | 141 |
| Zawiercie | r | 181 073 | 119 491 | 29 | 80 | 177 | 11 | — | 42 | 559 | 10 535 | — | 8 | 147 |
| | m | 88 482 | 57 914 | 22 | 38 | 98 | 3 | — | 21 | 269 | 5 052 | — | 6 | 64 |
| | k | 92 591 | 61 577 | 7 | 42 | 79 | 8 | — | 21 | 290 | 5 483 | — | 2 | 83 |
| M. Zawiercie | r | 32 872 | 26 060 | 8 | 44 | 110 | 3 | — | 30 | 1 | 5 677 | — | 2 | 19 |
| | m | 15 490 | 12 681 | 7 | 21 | 59 | 1 | — | 16 | — | 2 697 | — | — | 5 |
| | k | 17 382 | 13 379 | 1 | 23 | 51 | 2 | — | 14 | — | 2 980 | — | — | 14 |
| Wieś | r | 98 201 | 92 522 | 21 | 36 | 58 | 8 | — | 12 | 552 | 4 858 | — | 6 | 128 |
| | m | 47 992 | 45 233 | 15 | 17 | 34 | 4 | — | 5 | 268 | 2 355 | — | 4 | 59 |
| | k | 50 209 | 47 289 | 6 | 19 | 24 | 6 | — | 7 | 284 | 2 503 | — | 2 | 69 |








32 WOJEWÓDZTWO KIELECKIE

Tabl. 11. Ludność według płci i wyznania — Population d'après le sexe et la confession

Powiaty (Miasta i wieś, Miasta razem, Miasta o ludności powyżej 20 000. Wieś)

| Powiaty Districts | Ludność ogółem Population totale | rzymsko- katolickie i or- miańsko- katolickie cath.- romaine et cath.- arménienne | Wyznanie — Confession | | | | | | | | | | | nie okre- szone ou non précisé et non indiqué |
|---|---|--|-------------------------------------|-----------------------------------|-----------------------------------|----------------|--|---|-----------------|----------------------------|-------------|---------------|-------------------|---|
| | | | prawo- sławne orthod- oxes | ewangelickie — protestante | | | inne chrześ- cijańskie autres confes- sions chré- tiennes | nie okre- szone ou non précisé et non indiqué | | | | | | |
| | | | | augu- stynie augu- stins | refo- rmowane réfor- més | inne autres | | | | | | | | |
| Bedzin | r 231 310 m 114 125 k 117 175 | 198 211 98 074 100 137 | 101 57 44 | 365 2031 162 | 180 92 88 | 33 19 14 | 4 3 1 | 95 40 50 | 107 48 59 | 31 875 15 496 16 439 | 1 1 — | 21 14 7 | 317 148 169 | |
| Miasta razem Villes ensemble | r 105 542 m 51 481 k 54 061 | 77 114 37 733 39 384 | 17 10 18 | 183 96 87 | 108 55 53 | 24 14 10 | — — — | 63 27 36 | 82 37 45 | 27 852 13 465 14 387 | 1 — — | 16 10 6 | 62 25 37 | |
| m. Bedzin | r 47 597 m 22 818 k 24 779 | 35 790 12 314 23 476 | 21 9 12 | 45 25 20 | 26 15 11 | 11 4 7 | — — — | 10 9 10 | 16 6 9 | 21 625 10 447 11 198 | — — — | 10 6 4 | 35 13 22 | |
| m. Cieladź | r 21 003 m 10 557 k 10 451 | 19 820 9 969 9 853 | 8 6 2 | 28 18 10 | 21 12 9 | 5 4 1 | — — — | 14 4 10 | 5 4 1 | 1 077 525 552 | 1 — — | 1 — — | 8 13 — | |
| m. Dąbrowa Górnica | r 36 942 m 18 091 k 18 851 | 31 502 15 439 16 063 | 8 4 4 | 110 53 57 | 61 28 33 | 8 6 2 | — — — | 30 14 16 | 63 27 35 | 5 180 2 513 2 637 | — — — | 5 3 2 | 6 2 2 | |
| Wieś Communes rurales | r 125 768 m 62 654 k 63 114 | 121 097 60 342 60 755 | 64 38 26 | 182 107 75 | 78 37 35 | 9 5 4 | 4 3 1 | 32 13 19 | 25 11 14 | 4 023 1 973 2 052 | — — — | 3 1 1 | 255 222 132 | |
| M. Częstochowa | r 117 179 m 58 954 k 63 225 | 90 843 41 165 49 178 | 74 38 36 | 835 170 165 | 443 207 236 | 29 17 12 | 3 2 1 | 260 186 124 | 37 19 18 | 25 588 12 188 13 423 | 3 3 — | 19 10 9 | 45 22 23 | |
| Częstochowa | r 182 466 m 89 639 k 92 827 | 174 127 85 515 88 612 | 48 27 21 | 82 53 29 | 575 282 291 | 5 2 3 | 1 1 — | 182 96 86 | 80 36 44 | 6 927 3 425 3 499 | — — — | 14 10 4 | 427 189 238 | |
| Miasta Villes | r 14 461 m 7 101 k 7 360 | 11 260 5 499 5 761 | 2 2 — | 18 11 7 | 2 1 1 | — — — | — — — | 1 1 — | 4 3 1 | 1 161 577 584 | — — — | — — — | 13 7 6 | |
| Wieś | r 168 005 m 82 538 k 85 467 | 162 867 80 016 82 851 | 46 25 21 | 64 42 22 | 571 281 290 | 5 2 3 | 1 1 — | 181 95 86 | 70 33 43 | 3 766 1 851 1 915 | — — — | 14 10 4 | 414 182 232 | |
| Itza | r 159 718 m 79 046 k 80 672 | 147 062 72 901 74 161 | 51 39 22 | 105 60 46 | 388 179 209 | 27 13 14 | 4 3 1 | 241 120 121 | 56 57 41 | 11 377 6 486 5 881 | 1 — — | 6 2 4 | 948 175 173 | |
| Miasta | r 12 432 m 6 089 k 6 343 | 8 150 3 985 4 165 | 15 11 4 | 20 12 8 | 1 — 1 | 1 — 1 | — — — | 6 5 1 | — — — | 4 330 2 071 2 259 | — — — | — — — | 9 5 4 | |
| Wieś | r 147 286 m 72 957 k 74 329 | 138 912 68 916 69 996 | 46 28 18 | 85 48 37 | 387 179 208 | 26 13 13 | 4 3 1 | 235 115 120 | 98 57 41 | 7 142 3 485 3 722 | 1 1 — | 6 2 4 | 320 175 169 | |
| Jędrzejów | r 108 533 m 53 543 k 54 990 | 98 714 48 746 49 968 | 21 14 7 | 88 23 15 | 18 7 11 | 6 3 2 | — — — | 4 4 2 | 4 2 2 | 9 435 4 587 4 848 | — — — | 2 1 1 | 292 160 130 | |
| Miasto | r 12 857 m 6 165 k 6 692 | 8 336 3 998 4 338 | 4 3 1 | 23 13 11 | 15 5 10 | 3 3 2 | — — — | — — — | — — — | 4 440 2 138 2 308 | — — — | — — — | 13 5 8 | |
| Wieś | r 95 676 m 47 378 k 48 298 | 90 378 44 748 45 630 | 17 11 6 | 65 31 34 | 73 42 31 | 3 2 1 | — — — | 4 4 — | 4 3 1 | 4 995 2 449 2 546 | — — — | — — — | 279 151 128 | |
| Kielce | r 219 297 m 106 430 k 112 867 | 190 249 92 237 98 026 | 164 109 56 | 881 219 162 | 942 419 423 | 30 22 10 | 8 6 2 | 79 45 47 | 30 15 16 | 26 896 13 092 13 808 | — — — | 20 27 9 | 580 264 321 | |

Bilag 2 - Test af NDB med navnet Anne Frank: fire forskellige indførsler.

| | | | | | | |
|---|--------------|-------------|-----------------------------------|---|-----------------|---|
| <p>▼ Anne Anneliese Marie</p> | <p>Frank</p> | <p>1929</p> | <p>Amsterdam, The Netherlands</p> | <p>Page of Testimony</p> | <p>murdered</p> |  |
| <div style="display: flex;"> <div style="flex: 1;">  </div> <div style="flex: 2;"> <p>Anne anneliese marie Frank was born in Frankfurt am Main, Germany in 1929 to Otto and Edith nee Hollaender. She was a child. Prior to WWII she lived in Amsterdam, The Netherlands. During the war she was in Amsterdam, The Netherlands.</p> <p>Anne anneliese marie was murdered in the Shoah.</p> <p>This information is based on a Page of Testimony (displayed here) submitted by her father's wife, Elfriede Frank</p> <p>More Details</p> </div> </div> | | | | | | |
| <p>▼ Anne</p> | <p>Frank</p> | <p>1929</p> | <p>Frankfurt am Main, Germany</p> | <p>Page of Testimony (digital)</p> | <p>murdered</p> |  |
| <div style="display: flex;"> <div style="flex: 1;">  </div> <div style="flex: 2;"> <p>Anne Frank nee Frank was born in Frankfurt am Main, Germany in 1929 to Otto and Edith. She was a child. Prior to WWII she lived in Frankfurt am Main, Germany. During the war she was in Amsterdam, The Netherlands.</p> <p>Anne was murdered in the Shoah.</p> <p>This information is based on a Page of Testimony (digital) (displayed here) submitted by her other, Mireille Sijm</p> <p>More Details</p> </div> </div> | | | | | | |
| <p>▼ Anneliese Anne</p> | <p>Frank</p> | <p>1929</p> | <p>Amsterdam, The Netherlands</p> | <p>Page of Testimony (digital)</p> | <p>murdered</p> |  |
| <div style="display: flex;"> <div style="flex: 1;">  </div> <div style="flex: 2;"> <p>Anneliese anne Frank was born in Frankfurt am Main, Germany in 1929 to Otto and Edith nee Holländer. Prior to WWII she lived in Amsterdam, The Netherlands. During the war she was in Amsterdam, The Netherlands.</p> <p>Anneliese anne was murdered in the Shoah.</p> <p>This information is based on a Page of Testimony (digital) (displayed here) submitted by a researcher, Kate Gaddis</p> <p>More Details</p> </div> </div> | | | | | | |
| <p>▼ Anne Annelies Marie</p> | <p>Frank</p> | <p>1929</p> | <p>Aachen, Germany</p> | <p>List of murdered Jews from Germany</p> | <p>murdered</p> |  |
| <div style="display: flex;"> <div style="flex: 1;">  </div> <div style="flex: 2;"> <p>Anne annelies marie Frank was born in Frankfurt am Main, Germany in 1929. Prior to WWII she lived in Aachen, Germany. During the war she was in Bergen Belsen, Germany.</p> <p>Anne annelies marie was murdered in the Shoah.</p> <p>This information is based on a List of murdered Jews from Germany, found in List of Jewish victims from the Memorial book "Victims of the Persecution of Jews under the National Socialist Tyranny in Germany 1933 - 1945" prepared by the German Federal Archives.</p> <p>More Details</p> </div> </div> | | | | | | |