MXB362 Advanced Visualisation and Data Science
Visualising the effects of bombings conducted on Germany and its effects on the production of military goods over the course of WW2

Tyler Hunter N10008241

November 1, 2019

Contents

Chapter 1

Executive Summary

The primary purpose of this visualisation is to identify whether the bombing campaigns conducted on Germany had a noticeable impact on the overall production of military goods or if instead led to an increased number of civilian deaths for little to no effect on production other than the moral effect it would have had.This purpose leads the visualisation to be more catered towards a casual audience to showcase 1. The intensity of the bombings throughout the war and 2. The production in spite of these bombings

This correspondence will be accomplished by using data from the Theater History of Operations Reports (THOR) which gave a plethora of information regarding bombings conducted by the allies during WW2. The primary aspect of this data that is going to be used is bombings located in and around the immediate geographical area of Germany and the information corresponding to that (such as targets and tons of HE used).

A secondary data set of German production during WW2 as well as the GDP of the Germain Reich(excluding occupied territories for accuracies sake) will be plotted to serve as a secondary visualisation to demonstrate how the production moved over time as the intensity/location of the bombings changed.

A novel aspect of this visualisation is the use of latitude/longitude data as well as having the potential to view the concentration of bombings over certain areas each month of the war to see if a trend can be determined rather than just viewing all the bombings at the end of the war and not being able to see if certain areas had increased bombing after a certain point in time or if it was a consistent location to be bombed over the course of the war.

During the project a major problem with data collection was encountered which changed the scope of the assignment from Britain vs Germany to just Germany alone, another difficulty that was faced was formatting the dates of the assignment such that they were all of the same format with manual alteration required for dates that needed ot be converted from the US format to the UK format. Additionally location mapping for Longitude and Latitude proved problematic on a 2D map but this was rectified, other minor problems were also encountered relating to the quality of the visualisation.

Chapter 2

Description of the Project

2.1 Data Sources

2.1.1 Primary Data Source

The primary data that is going to be visualised in this project is bombings conducted over Germany during World War 2 sourced from the Theater History of Operations Reports (THOR). This properties that are of greatest interest is the spatial coordinates as well as the number of bombings conducted. Spatial coordinates are of interest because it allows the visualisation of the bombings in a way to show the intensity of the bombings other than a flat number which is second property of great interest. Additional attributes such as target industy is also included which would be beneficial in expanding the scope of the visualisation. Included below is a segment of this data set.

A	B	C	D	E	F	G	H
MSNDATE	TGT_LOCA	TGT_TYPE	TGT_ID	TGT_INDUSTRY_COD	TGT_INDUSTRY	LATITUDE	LONGITUD
10/01/1940	BERLIN	THERMAL	3735	3	"PUBLIC UTILITIES -	52.53	13.42
26/08/1940	AACHEN	AERODRO	10	78	AIR FIELDS AND AIF	50.78	6.08
8/02/1940	HAMBURG	RAILROAD	17721	63	"RR INSTALLATION	53.55	10

2.1.2 Secondary Data Source/s

The secondary data sources to this visualisation is the GDP/Production data for Germany over the course of WW2. In the case of the GDP data this was only for the German Reich and doesn't incorporate its occupied countries. This isn't the case for the production of Rifles,Tanks and Bomber/Fighter Planes though given another point where the visualisation could be expanded. Below is an example entry for the rifle production.

Type of Weapon	1939	1940	1941	1942	1943	1944	1945 (2-3 months)
Battle tanks	962	1,573	3,399	4,386	5,813	7,983	956

Another data source from simplemaps gave the lat/lon of major German cities which was used in creating the base map for the bombing plot. An example of this data is as follows

city	lat	lng	country	iso2	admin	capital	population	population_proper
Berlin	52.516667	13.4	Germany	DE	Berlin	primary	3406000	3094014
Stuttgart	48.782343	9.180819	Germany	DE	Baden-Württemberg	admin	2944700	606588
Frankfurt	50.11552	8.684167	Germany	DE	Hesse	minor	2895000	679664
Mannheim	49.496706	8.479547	Germany	DE	Baden-Württemberg	minor	2362000	313174
Hamburg	53.575323	10.01534	Germany	DE	Hamburg	admin	1757000	1739117
Essen	51.45657	7.012282	Germany	DE	North Rhine-Westphalia	minor	1742135	573468

2.2 Visualisation Techniques

The visualisation technique that were given the data above is to have the bombing location superimposed over a map of Germany.The spatial data of the bombings will be used to make a type of heat map of points showing the most commonly bombed locations if they existed or if the bombings were more widespread. This plot will be coupled with a line graph that details the levels of military production of Germany, this graph will exist as a subplot to the main data set.

The last plot that will be used is a table of sorts detailing the number of attacks against specific industries as well as the total tons of HE used (this feature is made possible by the extensive amount of information offered in the primary data set.)

2.3 Visualisation Environments and Tools

The primary tool that will be needed for this visualisation is Matlab.In particular Matlab's ability to make simple animations will be used which when coupled with the relative ease of making plots/subplots with additional features such as a heat map will ensured a relatively smooth process in creating the visualisation.

The mapping toolbox in matlab was also used in generating the primary visualisation, featuring functions such as worldmap and geoshow which did the heavy lifting of converting lat/lon to usable coordinates and placing them accordingly.

Chapter 3

Results and Outputs

3.1 Background Research

There wasn't a great deal of background research required for this assignment since WW2 is common knowledge, this restricted the background research to investigating possible data sources to determine if enough information existed for the purpose of the visualisation. This research wasn't flawlwess however which will be covered later in the report.

3.2 Data Massaging/Manipulation

3.2.1 Bombing Data

The bombing data had to be manipulated in such a way that all of the chronological values were ordered in the same format from the beginning to the end of the war, this was essential in plotting a timelapse of the bombings. The data was further massaged by target country which was set to be Germany, this wasn't foolproof however since the THOR data set counted bombings on occupied countries as being bombings on Germany which had further impacts on the visualisation. For the most part the data was manually formatted so only entries in the Lat/Lon boundaries of the visualisation were used but this still wasn't flawless.

3.2.2 Production Data

The production data only had data entries avaliable at year points, since the visualisation operated on a month by month basis this data needed to be extrapolated to fit in the gaps. This raised another issue where the data has to be carefully shown to be production levels (i.e For month X if the production data during this month was the

same over the whole year Y amount of Tanks would be produced rather than Y amount of tanks was produced this month) This could have been rectified by dividing the results by 12 but this risks misrepresenting the data since this then assumes the production is equal for all months.

3.3   Description of the Code

3.3.1   Main Function: Visualisation

The main focus of the main function is initializing and formatting all of the data for the visualisation as well as setting up the required subplots. The process of this initialization is as follows.

The table of THOR data is taken in as an input and is sorted into Two Arrays of BombingDates and BombingLocations. After this the production data is taken in and data is interpolated to create data points for each month. The bombing data from earlier is then sorted into Year and Month groups (i.e One dataset is Bombings1939September). Now the main plot gets created and the bombing visualisation plot is populated with city locations and names. Further plots are made for the GDP and Production.

All of this information is then passed into the Bombplot function which has inputs of Date,Location,Map to be placed on, Time between months in the visualisation, the handles of all the production plots and lastly the production data.

3.3.2   Secondary Function:BombPlot

This function is largely responsible for updating the various elements of the visualisation for each month in the timespan.

This is accomplished by taken in the data that has been sorted by date and seperating it into it's Longitude and Latitude components (this data also has other attributes but they are beyond the scope of the visualisation). The current Year and Month is then found by reading the appropriate data from these tables which is used for indexing purposes as well as updating the title of the masterplot. After doing some arithmetic to the Array Position the X position can be calculated and then all of the production plots are updated with their respective production values for the current month.

The bombing plot is then updated by plotting all the values for the current month although code exists to instead have these values placed one by one to have a higher clarity, this introduced problems however when used which will be described later in the report.

3.3.3   Secondary Code: gif

This is a function created by Chad Greene ("gif - File Exchange - MATLAB Central", 2019) that was used to create the gif of the animation. It uses a simple filename deceleration and then a subsequent call of gif for each frame, the only issue encountered is potential slowdown of applications.

3.4   Visualisation Outputs

The output of this visualisation is a timelapse over the course of WW2 that shows the bombings conducted over Germany and the immediate surrounding area. This is contrasted with a production/GDP plot for Germany. A snapshot of this visualisation is given at its end point with the animation being included in the zip file

Bombings from September 1939 to September 1945
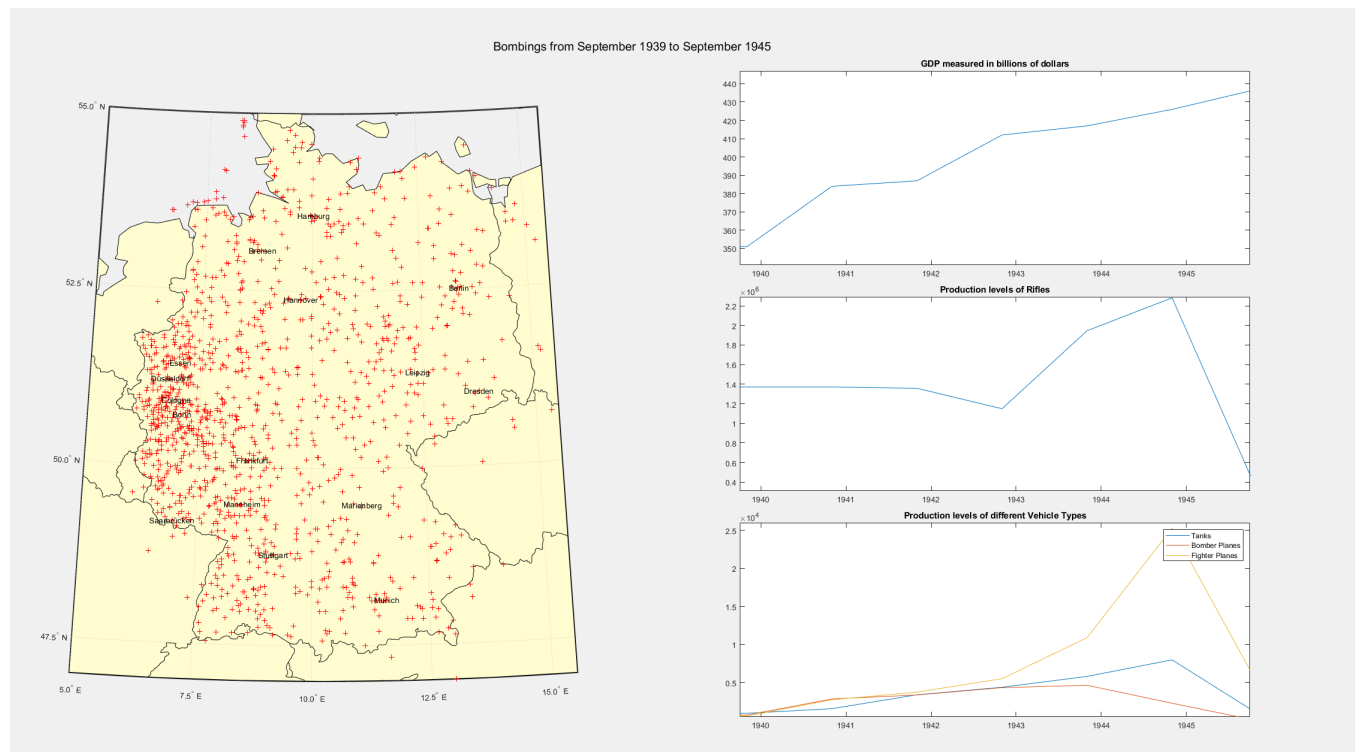
3.5 Effective Visualisation Issues

In the case of the production since the dataset was so small to begin with it was relatively straight forward already to see the rise and the fall so visualising this in a line graph is only a minor boost to how easy it is to see the data. The real strength of the production plots is in comparing them to one another where it is easier to view where the main focus of production for Germany was (from the productions visualised).

In the case of the bombing locations though a great deal more information can be gathered by plotting the points according to the geographic location. As well as by using a timelapse it is possible to see how the intensity of bombings increased and to see which areas were most heavily bombed features that whilst possible in the original table version of the data were incredibly difficult and time consuming to identify.

3.6 Insights and Knowledge Gained

During this visualisation I learnt a lot about effectively mapping Lat/Lon coordinates to a mapping system as well as other types of coordinate systems. The most useful knowledge gained was in using matlabs GeoShow feature and it's mapping toolbox which would no doubt help in future endeavours involving geographically located data.

Chapter 4

Exegesis of the Effectiveness of the Outputs

Critically looking at the visualisations there are a few features that could be improved for a better visualisation, each of these features will be listed in detail below.

4.1 Bombing Plot

4.1.1 Contrast between Bombings and City Names

As the bombing map becomes more and more saturated with bombing points (due to the fact that the prior months points can't be cleared) it becomes difficult to see the city names. This could be rectified by using a high contrast text on the city names that would stand out with the red bombing point markers (which themselves are the best color to maintain clarity with the color of the map as well as the city names whilst making the visualisation relatively nice to look at)

4.1.2 Legend on the Bombing Plot Map

Without proper documentation it wouldn't be instantly clear what each of the points represent until partway through the visualisation which could be rectified with the use of a legend.

4.1.3 Map Detail

For the sake of clarity only the most populated cities were included but this means that some bombings could be misattributed to one city rather than their true target, this isnt a major point however since if the map was too busy then it would be difficult to see where the bombings occured. For this to be most effective a balance should be found between the two.

4.1.4 Bombing Data Range

Since the range of the bombings was an area much wider than germany the area of the visualisation was restricted to ensure individual entries could still be seen, this had the effect though of reducing the clarity and not showing the full scope of the data which imposed a degree of inaccuracy in the production data.

4.2 Production Data

4.2.1 Possibility of confusion with data points

As stated earlier the production data per month is interpolated in such a way that isn't intuitive meaning that its possible that this data could be taken in the wrong way, this could be circumvented by finding accurate data for each month but that poses the challenge of finding the data to begin with.

4.2.2 Accuracy regarding Production outside GDP

Since the production isnt stated to be strictly germany it's possible that occupied countries could be contributing to this production which isn't shown in the main display. This could be solved by increasing the visualisation range for the bombing but then clarity is lost, an alternative is trying to find data that represents only the German Reich.

4.3 Overall Plot

4.3.1 Visuals

Visually the master plot is very bare bones and has a significant amount of empty space that could be used to enhance the visualisations, this could be fixed by using the position property for the plots but this raises issues where the visualisation might not properly work on smaller resolution screens.

Chapter 5

Updated Project Timeline
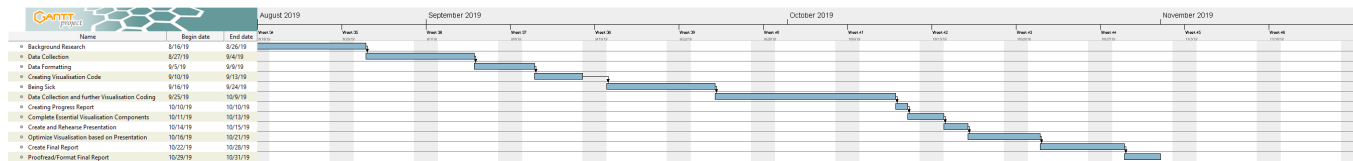
Since the progress report progress has gone mostly according to the timeline presented in the progress report with minute differences.

Chapter 6

Conclusion

In conclusion the project went relatively smooth and the end result is within the range of what I was after but there is definite room for improvement as noted in the exegesis. As a proof of concept this works well and could be expanded for multiple countries with the possibility of utilizing the full THOR dataset to look at bombings not just in Germany. The timelapse feature of the visualisation could be enhanced too with the use of a slider to allow a further look in depth at each month. If I had to do this again I would be much more indepth with my research into matlab functions before diving into my own haphazard way of approximating the work, this also applies into research into data sources where in my haste I chose a source that wasn't actually usable when I set the scope of the assignment.

Chapter 7

Problems Encountered

7.1 Data Collection of Britain Bombings

During the data collection phase an issue was encountered where the best/only place to collect the necesssary data for a core part of the visualisation was actually unable to be used because of the nature of the data. This was exacerbated by the fact that the data was purely physical and located in the UK meaning I couldn't attempt to do a basic hashing together of the data either. This problem arose because of not properly looking in depth at the data sources that I had chosen and could have been avoided with more time spent researching. The effects of this was the scope of the visualisation had to be change and as such it wasn't as effective of a visualisation as it could have been.

7.2 Formatting Date Data

The dates that was in the original THOR data set weren't of the same format (or excel's automatic date conversion misinterpreted the original data). This was problematic because it wasn't possible to chronologically order this data, this was ambended by using a KuTools extension for excel which automatically converted the data all to the same format of DD/MM/YY, the only problem that arose from this is some dates were in the MM/DD/YY format and this led to having bombings occur on months before the war had started and after the war had finished. This problem didn't have an automatic tool that could fix this and the only solution was to manually go through the dates by hand and shift them into the correct format.

7.3 Timelapse Animation

When creating the timelapse the optimal desing would have been to show 1 bombing point at a time so it was easy to see where each bombing was. This took a very long time though and wasn't feasible. In an ideal scenario a slider would be implemented so each month could be viewed but in the current stage the bombings by month are shown which isn't a very effective visualization for the purpose I want.

7.4 Clearing Datapoints for each month

In the process of having 100% accurate bombing locations using GeoShow created another problem where the bombings from each month cant be easily cleared leading to the visualisation becoming cluttered overtime but has the positive side of being able to view the density of the bombings on each area overtime.This problem mostly comes down to a matter of taste/convenience, for the sake of this report accuracy was chosen over clearing the bombings monthly.

7.5 Mapping Lon/Lat Data

This was eventually fixed by using geoshow and matlabs mapmaking toolbox but initially a function was created that haphazardly altered the points to X and Y coordinates and it wasn't awfully accurate and was a result solely born from poor researching (again).

Chapter 8

References

References

German arms production > WW2 Weapons. (2019). Retrieved 11 October 2019, from
https://ww2-weapons.com/german-arms-production/

Germany Cities Database | Simplemaps.com. (2019). Retrieved 1 November 2019, from
https://simplemaps.com/data/de-cities

gif - File Exchange - MATLAB Central. (2019). Retrieved 1 November 2019, from
https://au.mathworks.com/matlabcentral/fileexchange/63239-gif?
focused=7757452&tab=function

Mapping Toolbox. (2019). Retrieved 1 November 2019, from
https://au.mathworks.com/products/mapping.html?requestedDomain=

Military production during World War II. (2019). Retrieved 11 October 2019, from
https://en.wikipedia.org/wiki/Military_production_during_World_War_II

World War II THOR Data - dataset by datamil. (2019). Retrieved 11 October 2019, from
https://data.world/datamil/world-war-ii-thor-data