7-2021

Data Analytics for Humanities and Historical Analysis

Henry Wu
Purdue University, wu1461@purdue.edu

Data Analytics for Humanities and Historical Analysis

Mentor: Dr. Sorin A. Matei
Henry Wu
Purdue University

Introduction & Motivation

Historical text and images feed the complex narrative we view history with today. Images and texts such as a series of battle maps or a list of casualties contain quantifiable data that can be visualized. In this project, I aim to use historical data from the World War II battle: 1941 Operation Barbarossa, to reconstruct the events that unfolded. Before visualizing the battle, the data has to be cleaned, crosschecked, related, and analyzed.

The project is currently in the stage of organizing and structuring the data to prepare for its use in visualization and analysis.

Methodology

The data used in the project came in various formats such as battle maps, casualty lists, engagement data, and unit movements. As a result, scripts written in Python are implemented to relate the different forms of data and consolidate them into a central data sheet.

First, the data is checked for completeness. Entries with missing data are checked and fixed accordingly

Secondly, the data is consolidated into a central database. Information such as unit coordinates and unit type were grouped together.

Thirdly, relations among the data were realized and added to the master data sheet.

Results & Discussions

Two main datasets became apparent after processing the data

The first dataset is a master table containing information relating to map data such as

- division unit number and type
- date
- coordinates
- country
- casualty count and type
- battle and posture

Figure 1.

ORIG_FID	Unit Type	UniqueUnitCode	DATE	POINT_X	POINT_Y	Battle	Posture	KIA	WIA	MIA	ILL	ACD
59	Mot ID	40014	1941-07-15	29.83195125	55.5278373	4000	1	8	21	0	0	0
60	Mot ID	40014	1941-07-16	30.05077249	55.53068799	4000	1	0	2	0	0	0
61	Mot ID	40014	1941-07-17	30.57098916	55.65427681	4000	1	0	0	0	0	0
62	Mot ID	40014	1941-07-19	29.81201162	56.02110513	4000	1	5	5	0	0	0
63	Mot ID	40014	1941-07-20	30.48258403	55.92589301	4000	1	5	3	3	0	0
64	Mot ID	40014	1941-07-22	29.92248038	56.05466001	4000	1	0	3	0	0	0
65	Mot ID	40014	1941-07-26	31.84299431	55.98540264	4000	2	1	1	0	0	0
66	Mot ID	40014	1941-07-27	32.04404916	55.83413011	4000	2	0	0	0	0	0

This dataset allows us to plot divisions onto a map and animate their movements throughout the battle.

The second dataset is a master matrix containing information of the battle and posture of each unit on each date during the battle

- For each unique unit code in the master table, there is a recorded posture and battle of that unit corresponding to the correct date

Figure 2.

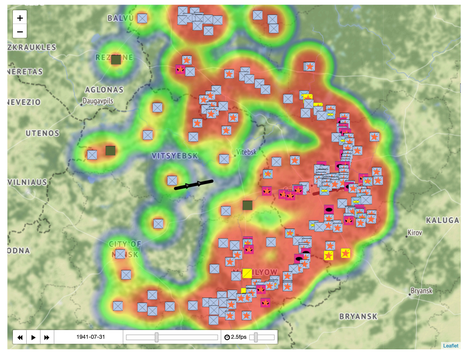
	1941-06-22	1941-06-23	1941-06-24	1941-06-25	1941-06-26	1941-06-27	1941-06-28	1941-06-29
20005	1	3	1	1	3	3	3	1
20006	1	1	1	1	3	3	3	3
20007	1	1	1	1	1	1	1	1
20008	1	1	1	1	1	1	1	1

Whenever we need to know the battle and posture of a specific unit on a specific date, it can be easily found in this dataset. This helps us interpret the state of the battle in terms of military postures (1: attack, 2: defense, 3: movement, or 4: reserve)

Conclusions

Effective mapping and scripting techniques have helped us consolidate data from many scattered sources to one or two centralized datasets, creating a complete and accurate source of data to use in future analysis and visualization. Only with the help of these techniques can we efficiently create visualizations such as this early version of our weather map to help with the analysis of the battle

Figure 3.



Future Work

The next phase of the project will involve improving visualizations and start the analysis of the events that unfolded during the battle. We will look at details such as the general movement of the battle through momentum vectors and concentration of units through dynamic heat maps.

References

Germany. Heer. Generalstab, C. (1942) *Der Feldzug gegen Sowjet-Russland: Band I. Operationen Sommer-Herbstvom 21. Juni-6. Dezember*. [Berlin?: Generalstab, ?] [Map] Retrieved from the Library of Congress, https://www.loc.gov/item/map51000141/.

Acknowledgments: The authors acknowledge support from the Lilly Endowment, Inc.