VA - Roman Empire

Luigi Russo, Matteo Salvino

Sapienza University of Rome

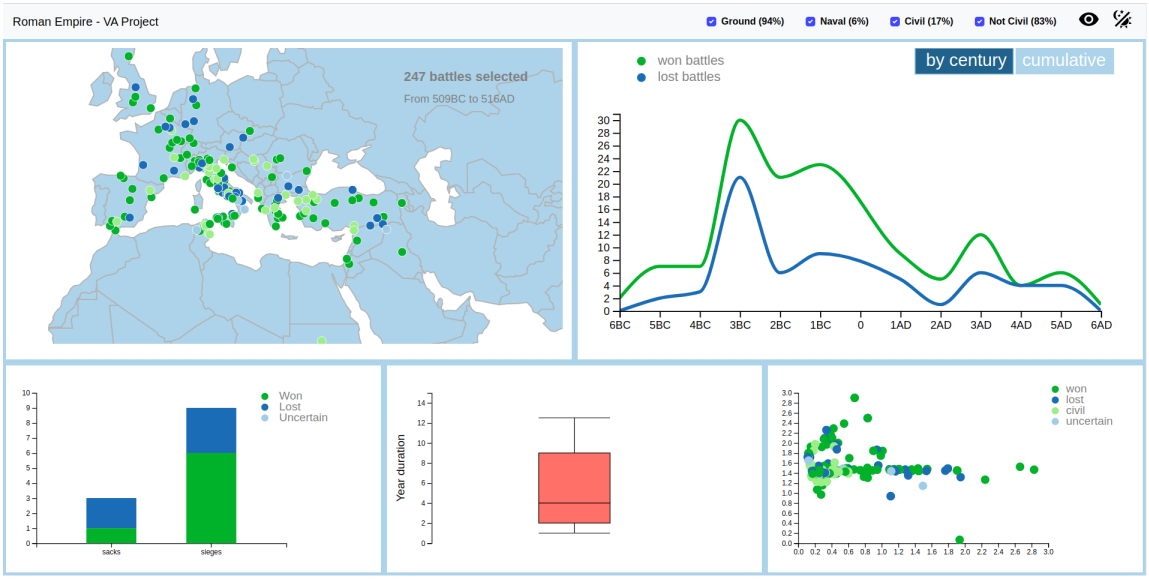

Figure 1: *Visualization with blind safe mode enabled.*

Abstract

The military of ancient Rome is unanimously considered a key element in the rise of Rome. We developed a visual analytics tool for educational purposes, mainly to give a visual representation of ancient Roman battles and wars. The multiple interactive views of our project present high level features, but the user also has the possibility to explore further information, e.g. via an info-box popup associated to each battle.

1 Introduction

> "If any people ought to be allowed to consecrate their origins and refer them to a divine source, so great is the military glory of the Roman People that when they profess that their Father and the Father of their Founder was none other than Mars, the nations of the earth may well submit to this also with as good a grace as they submit to Rome's dominion." ~T. Livius

The military of ancient Rome is unanimously considered a key element in the rise of Rome over above seven hundred years, from a small settlement in Latium to the capital of a wide empire around the shores of the Mediterranean sea. The role and structure of the military were modified many times during this long period: we decided to develop a tool to analyze the main battles and wars involving the Roman army.

We started with building our own dataset (see 2) from a Wikipedia page which contains useful information about the ancient Roman history. In particular, our dataset collects most of the Roman battles, from 6th century BC to

6th century AD. We later built a set of views (see 3): each of them provides a set of interactions. We focused on the type of features to provide to the users: in fact is possible to filter the data based on the given time period, analyze how many battles (and their type) the Romans won and lost, read further information on a battle thanks to the info-box popup that is shown when a battle is selected on the map. The primary goal of our project is to provide an educational tool, helping users to interactively learn the very detailed ancient Roman history through a multi coordinate visualization built with D3.js framework.

2 Dataset

The dataset used in this project has been generated starting from the (english) Wikipedia article List of Roman wars and battles. As all the Wikipedia pages, this article is written using the wikitext, a special hypertext markup used by the MediaWiki software as default formatter. In order to acquire the necessary information, we have used some script to scrap the content of the page and produce ER-tables: this is a much more flexible way to represent data [1], with respect to the wikitext, and in particular makes it possible to export as comma-separated values (**csv**) file the entire dataset.

2.1 ER Tables

We obtained ten different relationships, listed in table 1.

For sake of simplicity and in order to avoid to compute joins on tables at runtime, we organized the final (i.e. the one passed in input to the visualization tool) dataset into 5 csv files, hosted on the project repository and loaded only once, when the tool is started:

- battles.csv: the join of many different tables is precomputed
- wars.csv: this is the join of wars info and period
- allies.csv, commanders.csv, images.csv

Main	Relationship
Battles	info
	type
	outcome
	places
	allies
	commanders
	period
	images
Wars	info
	period
	bibliographical sources

Table 1: *The underlying dataset as set of tables.*

In the future, our tool will support other tables as well (see 5): in particular, we remark that the only table not yet integrated is the one containing the bibliographical sources.

2.1.1 Battles

This table comes with different attributes:

- id - unique identifier of the battle
- warId - identifier of the corresponding war (if any)
- year - year of the event
- yearAU - year of the event, ab Urbe condita
- relativeYear - year of the event, relative to roman phase
- label - the name of the battle
- locationLabel - place of the battle
- latitude - north–south position
- longitude - east-west position
- outcome - win, loss, ...
- civil - marks the battle as civil
- naval - distinguishes between naval and ground battles
- siege - indicates that the battle is actually a blockade
- sack - indicates the event as a sack
- final - marks the battle as decisive fight
- period - Roman phase

2.1.2 Allies

This table comes with different attributes:

- id - unique identifier of the battle
- RomanAllies EnemyAllies - list of allies
- RomanStrength EnemyStrength -
- RomanLosses EnemyLosses -

2.1.3 Commanders

This table comes with two main attributes:

- RomanCommanders - list of Roman commanders
- EnemyCommanders - list of adversarial commanders

2.1.4 Wars

This table comes with different attributes:

- id - unique identifier of the war
- label - name of the war
- startYear
- endYear

2.2 Preprocessing

The dataset already offers many attributes; however, we decided to perform some preprocessing in order to enhance it. The preprocessing phase consisted of adding four extra attributes for battles

- currentCountry - the current country in which the battle took place
- stoaId - Pleiades identifier
- y1, y2 - the result of a static MCA

and one extra attribute for wars

- wikidata - Wikidata item identifier

2.2.1 Pleiades Identifier

Pleiades is a community-built gazetteer and graph of ancient places. It publishes authoritative information about ancient places and spaces, providing unique services for finding, displaying, and reusing that information under open license: it publishes for the widening array of computational research and visualization tools that support humanities teaching and research.

We downloaded one of their dumps in order to statically retrieve the place identifier and to adjust some latitude and longitude values. This allowed the visualization tool to properly render the position on the geographic map of an event and to provide an hyperlink to Pleiades website in the battle info-box. Moreover, we adopted this strategy since it is more flexible. Indeed, binding an ancient place to its corresponding Pleiades entry allows many benefits, among them we remark:

- we can fully exploit the Pleiades graph and its daily-basis update system
- it is simpler to detect errors and eventually fix them

2.2.2 Reverse Geocoding

Through reverse geocoding we were able to add an attribute for each tuple in the battles dataset. We used the reverse_geocoder Python 3 library to do that. The script is available as a Jupyter Notebook project, named rg.ipynb.

2.2.3 MCA

Finally, we run a Multiple Correspondence Analysis (**MCA**) [2] on the tuples of battles dataset, and we stored the two principal components by appending them to the input data: we named these two coordinates y_1 and y_2 and we have used them in the scatterplot view to give a 2D representation of the battles.

Roughly speaking, MCA is an extension of the well-known Correspondence Analysis [3]: the idea behind it is to simply compute the one-hot encoded version of a dataset and apply CA on it; we decided to run this analysis because most of the variables are categorical.

The MCA was executed on a subset of the attributes, using the popular Prince Python 3 library: we have excluded attributes like "id", since they are just used to properly represent the ER schema, but carry no meaningful information for the analysis.

Note that MCA has been executed only once: the coordinates associated to each battle are static and bound to each tuple before running the visualization and analytics tool.

2.2.4 Wikidata binding

We also used the MediaWiki API in order to retrieve the Wikidata unique id associated to each war: this identifier is appended to each tuple in the column "wikidata".

Through this identifier it is possible to access the pages of MediaWiki projects (e.g. en.wiki, itwiki, etc.): this allows the analyst to explore the details of the events. Again, as already stated above, we are sure that this approach empowers the flexibility because in case of errors it is simpler to fix them.

3 Visualization

Table 2: *Relationship between charts and ER-tables.*

ER-Table	Chart
Battles	Geographic Map
	Line Chart
	Stacked Bar Chart
	Scatter plot
Wars	Box Plot

3.1 Geographic Map

The first component of our visualization tool is the geographic map, whose goal is to show to the end customer the precise projected point (using longitude and latitude attributes) where a particular battle occurred. In particular, we took an European continent picture with its respective countries boundaries information, and projected the real battles coordinates over it, in order to properly fit them to our figure. We represented each battle with a clickable circle, whose color depends on the battle outcome (i.e. won, lost, etc.); in order to improve the quality of the analysis, we offer the possibility to know some details about the currently selected battle: indeed, we have provided a modal popup which summarizes some useful information such as battle name, the year, the coordinates and its outcome. Furthermore, it is possible to select multiple battles via the brushing approach. In fact, when a number $n > 2$ of battles is selected, a small legend will show that n battles have been selected together with the range (of years) in which they have occurred. Note that, when brushing, the unselected battles are hidden.

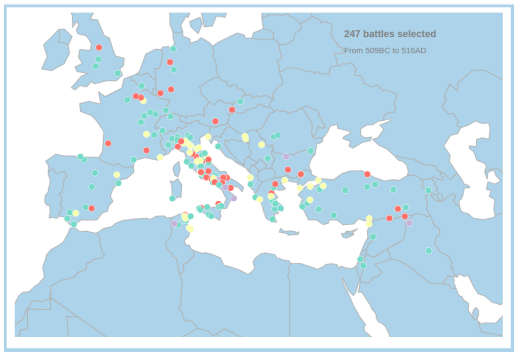

Figure 2: *Geographic map*

In this scenario, the other charts of the visualization will be properly updated. Conversely, also the geographic map will be updated when a particular time period is selected in the Line chart or a group of battles is selected in the Scatter plot. In the first case, the battles outside the selected time period are hidden, whereas in the second scenario the selected battles are highlighted with a black border.

3.2 Infobox

In order to allow the end customer to explore further information about a specific battle, we have implemented an information container, using Bootstrap modal feature, that is shown when he clicks on the marker of a battle. This container displays the image of the battle, the date, the location and its outcome. When clicking the location label, it is possible to explore more information about that ancient place on Pleiades. Furthermore, we can find the com-

manders of the Roman and enemies empire, as well as their allies. At the end of the container, we can find some statistical measures about the strength and losses of the two counterparts.

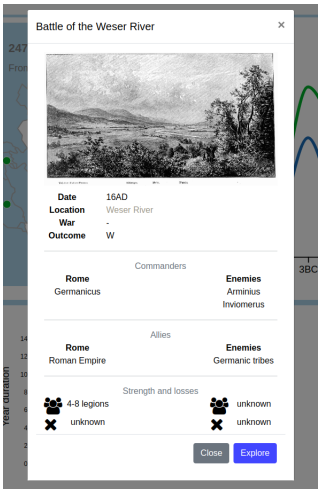

Figure 3: *Infobox example*

The last thing we want to underline is the possibility to explore further war details of the chosen battle through the button Explore, if and only if the corresponding war identifier is present.

3.3 Line Chart

The next component is the Line chart, which is responsible for communicating to the analyst the trends of won and lost battles in a given time period. We have provided two analysis approaches:

- **by centuries** in order to analyze the previous trends from a general point of view in century form
- **cumulative**, which allow us to perform a more specific analysis with respect to the previous one, taking into account the trends in year form.

The Line chart has two main axes: the x axis represents centuries or years, and the y axis represents the number of battles. The first interaction on this chart is reached through brushing plus zooming approach. In fact, selecting a specific time period via brushing, the zoom function will update the chart axis's ranges and data. To remove a selection, the user has just to perform a double click. In this scenario, the selected data will update also the geographic map, the stacked bar chart and the scatter plot. On the contrary, this chart will be updated if only if any change occurs in the geographic map.

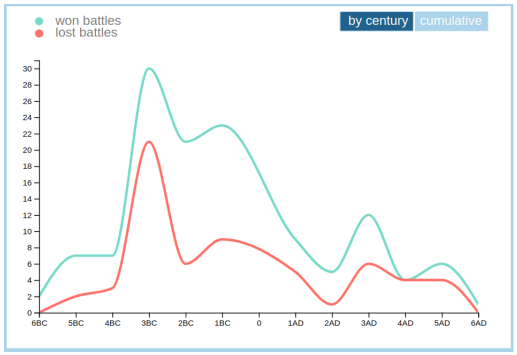

Figure 4: *Line chart showing century by century the number of battles.*

3.4 Stacked Bar Chart

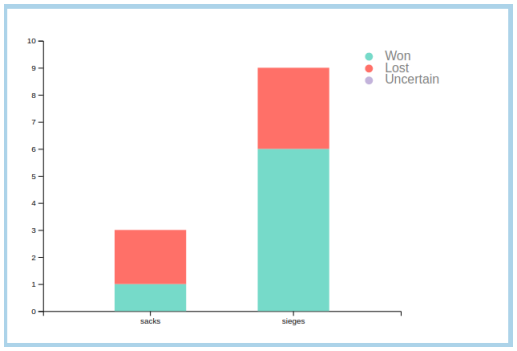

Figure 5: *Stacked Bar chart*

The stacked bar chart is another component of our visualization that offers a different type of analysis. It's constituted by two axes: the x axis represents two battle types: **sacks** and **sieges**; the y axis, instead, represents the number of the battles. This analysis is based on won, lost and uncertain battles (as usual we reported a legend for reasons of clarity).

This chart is not able to update the other ones, but can be updated with a selection performed in the geographic map, in the line chart or over both of them.

In case of doubts about the actual number of battles outcome with a particular type, the user can simply move the mouse pointer over that specific layer, thus producing a small label (tooltip) showing the exact number: this can be extremely useful because of the intrinsic nature of the stacked bar chart and its bars displayed one on top of the other.

3.5 Box Plot

The Box Plot is another component of our visualization tool: it aims at giving a visual representation of the Roman wars, rather than the battles, as the other charts.

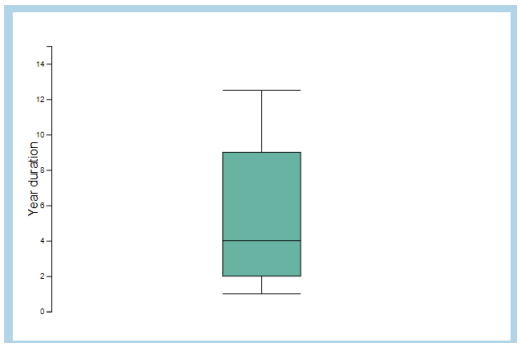

Figure 6: *Box plot*

A unique box plot, indeed, is drawn to graphically render the duration of the wars, expressed in years, through quantiles: these quantiles are dynamically computed, since the user interacting with the other views intrinsically filters some battles and their corresponding wars.

3.6 Scatter Plot

The scatter plot is one more component of our dashboard, used to represent the roman battles after a dimensionality reduction on two dimensions.

In particular, this chart represents the first two components, y_1 and y_2, obtained after the execution of a static MCA on the battles.

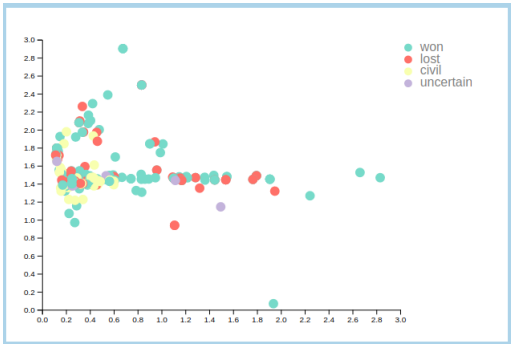

Figure 7: *Scatter plot*

This chart allows the user to:

- find similar events, taking into account the distance between the points
- individuating clusters
- discover outliers

We have decided to add a chromatic encoding of the points, associating to each color a different outcome: the colors are the same used in the geographic map. The user is also given the possibility to brush a zone of interest on the graph: the points within the zone are highlighted in the geographic map as well.

3.7 Filters - Header

Figure 8: *Header filters*

Another main component of our project is the header. Indeed, in order to allow the analyst to further customize its analysis, we decided to implement a set of simple filters such as **Ground**, **Naval**, **Civil** and **Not Civil**. By default, all filters are selected, i.e. the whole dataset is visualized. The activation (resp. deactivation) of a filter triggers a function that hides (resp. shows) all the data corresponding to that filter.

3.8 Dark and blind safe modes

Other additional features available in our visualization to allow the analyst to feel comfortable are the **dark theme** and **blind safe mode**. We decided to introduce the first one, since a human in different time periods of the day can be mentally tired, and couldn't pay so much attention to several visualization information. In order to reduce this issue, we decided to provided a more relaxing colors combination for the eyes, thanks to ColorBrewer web tool [4]. The default theme is the light one, which uses a colors combination brighter and lighter with respect to the one used for the dark theme. Another problem that could arise is referred to the colorblind people. In fact, these people see particular colors in different ways, for example a green color can be seen as brown one. For this reason, we decided to introduce also a blind safe mode, always supported by ColorBrewer, which uses a colors combination that have been scientifically chosen in order to reduce the confusion likelihood as much as possible.

We want to underline that we have used the same blind safe colors combination for both themes.

4 Analytics

As we have previously mentioned, in our visualization play a fundamental role the interactions between the available charts. In fact, the interactions transform a static visualization into a dynamic one, that is what we humans are used to see. For space reasons, we will represent with a $*$ all available charts in our project. In particular, the interactions provided are the following:

- **geographic map** $\rightarrow *$: there is a possibility to select one or multiple battles through brushing technique. In particular, when a window is drawn, we can leverage its boundary in order to trigger a run time computation to exclude (i.e. hide) the battles outside the brush area and show only the ones inside it.At the end of this task, another run time function is triggered to update all the remaining charts, constraining their visualization domain in order to show only the selected battles.
- **line chart** $\rightarrow$ **geographic map, stacked bar chart, scatter plot**: in this case we used brushing and zooming techniques for a possible interaction. As before, with the brushing we are able to understand the subset of battles that we want to show and with zooming we are able to update axis's ranges and data. To make things clear, the zoom area is exactly the same of the brushing one. Once that this update is performed, a run time function is called to update the data shown in the geographic map, in the stacked bar chart and in the scatter plot.
- **stacked bar chart** $\rightarrow$ null: this chart does not provide any interaction, it is only updated by others.
- **box plot** $\rightarrow$ null: this chart does not provide any interaction as well, but it is updated by others.
- **scatter plot** $\rightarrow$ **geographic map**: in this chart we have used the brushing approach to select a group of battles that will be highlighted with a dark border in case of light theme (resp. with a white border in case of dark theme). This action also triggers another run time function, which is responsible to highlight the selected battles also in the geographic map.

The analytics part is constituted by two main functions:

- **Percentage of ground / naval / civil / not civil battles**: at run time will be triggered a function that computes the exact amount of each type of battles reported above, that will be displayed nearby the corresponding battle type label.
- **Box plot quantiles**: they are the principal components of this chart, which are computed at run time.

5 Conclusions

We believe that this work already represents a good and innovative tool for an interactive analysis of the Roman military: it comes in handy because it offers

- the possibility to select and filter data
- an instant visual information of the selected events
- some basic analytics
- hyperlinks for further details

Several directions can be investigated in the future and many extensions may be taken into account. Among them we remark:

- Dataset extension: improve the dataset quality, adding more fields, e.g. external sources or bibliography for each battle or war
- Predefined filters: provide to the user a list of historical relevant filters, such as "Pax Romana"
- Dynamic boundaries visualization: dynamically change the borders, depending on the period

References

[1] Peter Pin-Shan Chen. The entity-relationship model, toward a unified view of data. *ACM Transactions on Database Systems*, 1(1):9–36, March 1976.

[2] A. Hervé and V. Dominique. Multiple correspondence analysis. *Encyclopedia of Measurement and Statistics*, January 2007.

[3] M. O. Hill. Correspondence analysis: A neglected multivariate method. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 23(3):340–354, 1974.

[4] Mark Harrower and Cynthia Brewer. Colorbrewer.org: An online tool for selecting colour schemes for maps. *Cartographic Journal The*, 40:27–37, 06 2003.