

Information Visualization Project Proposal and Dataset

G3-A

75180 – Adriana
Domingos

75541 – Tomás Alves

75551 – Miguel Santos



01

INITIAL DATASET

Initial Dataset

Description

Composed by static tables of the Wikipedia related to the WW2. The tables regard data from the Eastern and Western fronts.

Data sample

Germany: 157,621 total casualties (c. 49,000 dead)
1,236 aircraft lost[5][9]
795 tanks destroyed[10]

Italy: 6,029

Total: 163,650 casualties

02

**SELECTED / DERIVED
DATA**

Selected data

Data description

The selected data were the battles' names, its location, its winner and the side of each country, its casualties, its dates and the commanders.

Derived data

Derived data description

There were no derived measures since the group had to gather all the data by hand and create the final dataset.

The data that the group took from the tables were the only ones who had interest to answer the proposed questions, since they are about the number of casualties and the commanders for each side of the war.

03

DATA ABSTRACTION

Data abstraction

Description:

Commander_Allies.csv and

Commander_Axis.csv: table containing the name and country of an Ally or Axis commander, respectively, that participated in a certain battle.

Dataset type:

Static table

Data abstraction

```
Battle;CommandersAlliesCountry;CommandersAlliesName;  
Battle of Britain;United Kingdom;Winston Churchill;
```

- Items: battles' commanders.
- Attributes:
 - Battle, CommandersXCountry and CommandersXName: nominal, where X is his side.
- Meaning:
 - Battle: battle's name;
 - CommanderXCountry: commander's country and name, respectively, where X is his side.

Data abstraction

Description:

Date_Casualties_Victory.csv: table containing the battles, the location, the dates, casualties for each side and the winner.

Dataset type:

Static table

Data abstraction

MainBattle;Battle;StartDate;EndDate;TotalDays;CasualtiesAllies;CasualtiesAxis;Location;Victory

Battle of Britain;Battle of Britain;10/07/40;01/11/40;111;91964;4303;London;Allies

- Items: battles.

- Attributes:

- MainBattle, Battle, Location and Victory: nominal;
- StartDate and EndDate: quantitative, sequential, hierarchical;
- TotalDays, CasualtiesAllies and CasualtiesAxis: quantitative, ratio.

MainBattle;Battle;StartDate;EndDate;TotalDays;CasualtiesAllies;CasualtiesAxis;Location;Victory

Battle of Britain;Battle of Britain;10/07/40;01/11/40;111;91964;4303;London;Allies

- Meaning:

- MainBattle, Battle and Location: main battle's name, battle's name and location, respectively;
- StartDate and EndDate: battle's starting and ending date;
- CasualtiesX: casualties' numbers where X is the side;
- Victory: the side who won the battle.

04

DATASET PROCESSING

Dataset processing

Dataset cleaning description

Since the group was collecting the data by hand (e.g. copy/paste the data from the Wikipedia tables), the cleaning process was done in the selection step.

There wasn't the need for any further cleaning, since the data was all specifically picked.

Dataset processing

Problems found:

No data regarding casualties

- Delete the data from the dataset

Insufficient data regarding the casualties

- Input a value that made sense according to the data found (e.g. “unknown, but higher”, “unknown, but significant”).

05

MAPPING

Mapping

- **How did the casualties vary throughout the war?**

Use the dates from each battle and the casualties of each side.

- **Which was the biggest battle?**

Use the battles' names and casualties of each side.

- **Who was the worst commander?**

Use the commander's name and the battles' names where he participated, along its casualties.