

The Human Mortality Database

01405740 Florian Hauser a01405740@unet.univie.ac.at
01147464 Riccardo Pfeiler a01147464@unet.univie.ac.at

Fakultät für Informatik, Universität Wien

1 Motivation

1.1 Problem

Due to the unavailability of a comparable product, we wanted to produce a tool, that helps our users to gain insights to the Human Mortality Database and provide proper findings to fulfill defined explorational tasks such as comparing birth and death rates within countries and between different countries like Switzerland and Australia.

1.2 Tasks

Death Rate:

- Over the Years, how did death rate develop in countries where war took place?
- How did death rate develop in countries where no war took place?
- How do season correlate to the death rate?
- Comparing young and old people: How did death rate develop?
- Comparing males and females: How did death rate develop?

Birth Rate:

- While in some countries war took place, how did birth rate develop in those countries and how did it elsewhere?
- Are there any significant demographic changes in connection to a fall in birth rate?
- Is birth rate increasing with the increase of population?

Life Expectancy:

- How did life expectancy develop in different countries and cultures?

1.3 Users

Persona: Robert



- Persona: Primary
- Name: Robert Langdon
- Age: 45
- Job: Accountant

- About Robert:
Even though Robert has a very reputed job at his company, his passion is to write his own reports or essays on demographic circumstances. He sees himself more as a journalist than an accountant. He would like to have a tool that provides him the information he needs for exploring without spending too much time on configuration.

- Motto:
"Three things cannot be long hidden: the sun, the moon, and the truth."

User Goals:

- Have a reliable and pre-edited source of information about the development of death and birth rates in different countries.
- Being able to draw connection between these findings.
- Have a user oriented visualization
- Being able to communicate the gained information well and understandably to others.

1.4 Data

The Human Mortality Database (HMD) [1] was created to provide detailed mortality and population data to researchers, students, journalists, policy analysts, and others interested in the history of human longevity. Our process for computing mortality rates and life tables can be described in terms of six steps, corresponding to six data types that are available from the HMD. Here is an overview of the process:

- Births. Annual counts of live births by sex are collected for each population over the longest possible time period. These counts are used mainly for making population estimates at younger ages.
- Deaths. Death counts are collected at the finest level of detail available. If raw data are aggregated, uniform methods are used to estimate death counts by completed age (i.e., age-last-birthday at time of death), calendar year of death, and calendar year of birth.
- Population size. Annual estimates of population size on January 1st are either obtained from another source or are derived from census data plus birth and death counts.
- Exposure-to-risk. Estimates of the population exposed to the risk of death during some age-time interval are based on annual (January 1st) population estimates, with a small correction that reflects the timing of deaths within the interval.
- Death rates. Death rates are always a ratio of the death count for a given age-time interval divided by an estimate of the exposure-to-risk in the same interval.
- Life tables. To build a life table, probabilities of death are computed from death rates. These probabilities are used to construct life tables, which include life expectancies and other useful indicators of mortality and longevity.

2 Related Work

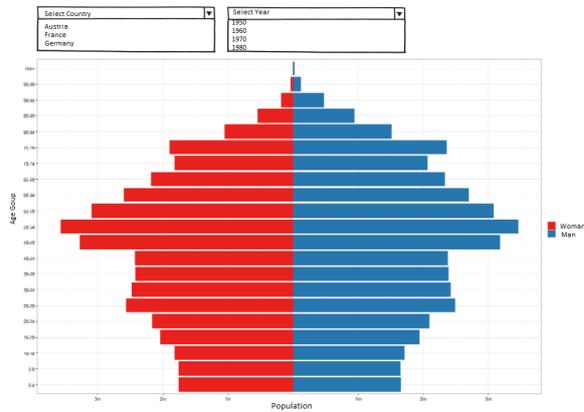
In general, there are two types of visualization. The visualization for presentation purposes and the visualization for exploration. The first one is designed to present your finding to a wider audience and the second one is for researchers who want to explore data to figure out what is going on. For our tool we used the second approach. [3]

Another paper states that "The basic objective of data visualization is to provide an efficient graphical display for summarizing and reasoning about quantitative information." [4]

2.1 Previous Visualization Ideas

To realize our project we had to reason and design various ideas but only a few made it to the final product. Here we provide a list of some of our design ideas:

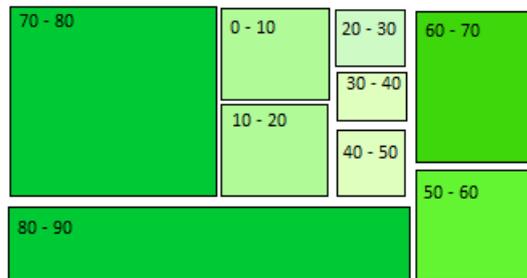
- Population Pyramid



This view shows the distribution of the various age groups in a population. The size of the population is depicted on the x-axis, which is separated in man and female, and age-groups on the y-axis. It is also possible to select different countries and years.

– Death by Age Groups

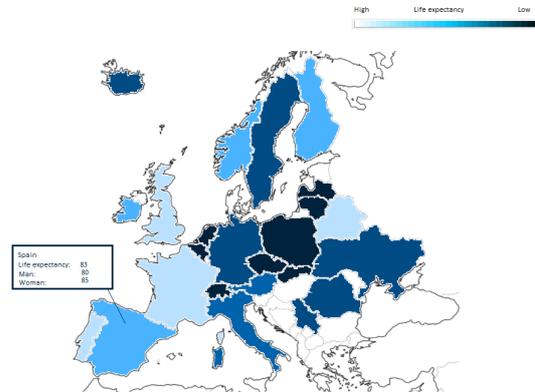
Deaths by age Groups



This visualization shows the deaths by age groups using a tree map. The color shades are used to give a clearer separation between the groups of age.

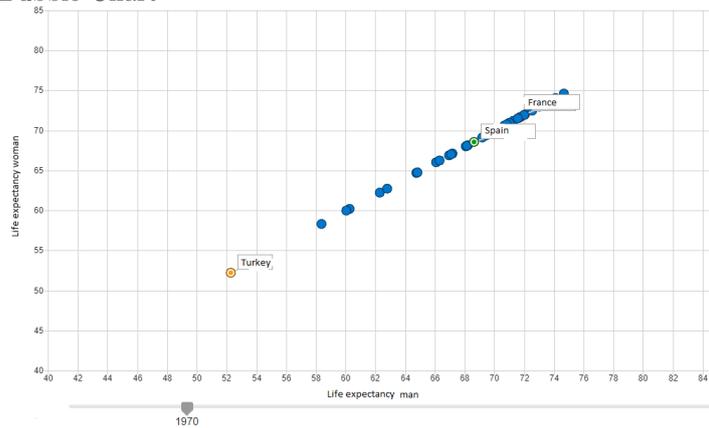
– Life Expectancy

- Map



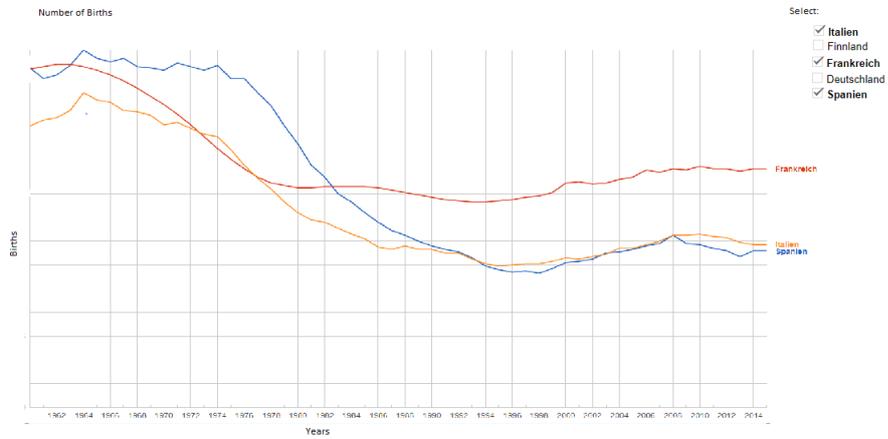
The map displays the life expectancy in different European countries. When a user hovers over a country a tool-tip occurs. The tool-tip shows the overall life expectancy and the life expectancy of woman and man in the chosen country.

- Bubble Chart

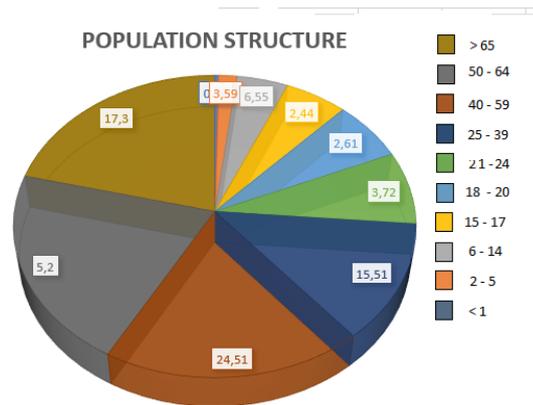


This bubble chart shows the correlation of male and female life expectancy per country. By using the slider on the bottom, a brushing feature is available to select a timespan as a subset.

– Number of Births



This line chart shows the development of births in different countries. By selecting different countries, a comparison of the birth rates per year is applicable.



– Population Structure

The pie chart shows the distribution of the various age groups in a population. The size of the population is depicted on the size of the slices and age-groups separated in colors. It is also possible to select different countries and years. This chart was just a design study. We wanted to try new ways of showing the population structure of a country.

2.2 Tool References

3 Approach

3.1 Visualization Design

For this project we choose to operate a design study where we want to gain insights to the Human Mortality Database and provide proper findings.

3.2 Reasons for Design Choices

During our prototyping phase we used a tree-map to display our data for "Deaths by age group", but eventually we decided to change the visualization to a bar chart because area-encoding of a number is suboptimal for perceptual reasons.

We used the bubble chart/scatter plot to display "Life Expectancy" in our final product and dismissed the map visualization. We think a bubble chart has more impact on perception than a map with color encoding due to perceptual reasons like changes in distribution (of e.g. curves) are easier to detect than a (slight) change in color.

By interpreting the pie chart for "Population Structure" we were able to discover drawbacks:

- High data ink ratio
- Old fashioned 3D structure of the pie These negative aspects made us to prefer "Population Pyramid" over the pie chart - which both basically display the same but in a different encoding.

high data ink ratio and the old fashioned 3D structure of the pie

4 Implementation

- Brief Description:

Our solution will implement an interactive visualization where the users can adjust the desired timespan. To do so, we will have usage of the visualization tool 'Tableau'. [2] The main reasons we will use Tableau are:

- The dashboard can easily be embedded into our webpage
- Interactive reports can be built

Because the data is not properly formatted for the usage with Tableau, we had to re-format the dataset of use.

- Difficulties:

We had major problems on processing the datasets we wanted to visualize. The data was split up per country, which was again split up into births, deaths and population. Each individual data source had a different formatting. Also the tool Tableau did change our data as we imported them and tried to connect each source - this made us to spend much more time on preparing the data than on our actual implementation of a Prototype/Dashboard with the desired functions. Therefore we had to delay our missing features like comparing countries to Milestone 4.

5 Results

5.1 Scenario of Use

– Use Case

Roberts first project deals with the effects of wars on the population structure in different countries for example Switzerland and Australia. The fact that the number of 20-30-year-old males decreased during wars drives him to dive deep into the dataset of Human Mortality Database and retrieve answers to his questions. He wants to show the development of population despite the huge reduction of swiss males. Also, he wants to compare these findings with countries that were also involved in martial activities, for example Australia.

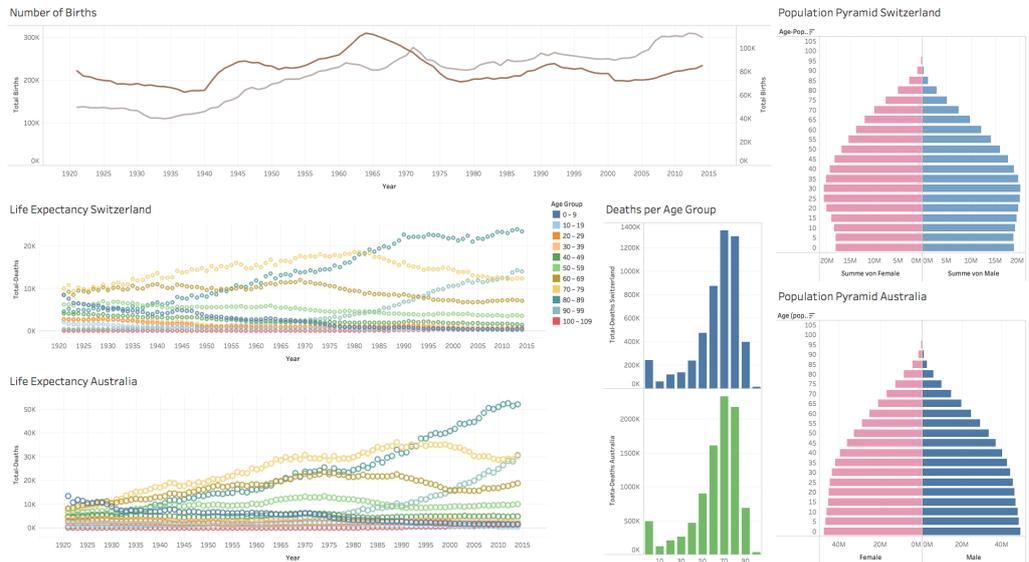
Upon opening the tool, just by looking at the dashboard, especially the chart displaying the number of births Robert was able to discover the following right away: Until 1914 the number of births per year was relatively stable. His findings show that during the first world war the number of births decreased hugely. After the first world war a short peak appeared until the great depression in the 1920's. He was staggered as he saw that the birth rate increased drastically with the rise of the second world war. After this time, the rates went almost through the roof until the introduction of the oral contraceptive-pill, which provoked a massive fall in birth rates.

Then after having processed the previous findings, Robert inspected the deaths per age group chart and found that males represent a higher value than females within the age group of 70-79. Females are overrepresented within the age group of 80 to 89 year olds. Therefore he hypothesize that the general life expectancy of females is greater than the one of males.

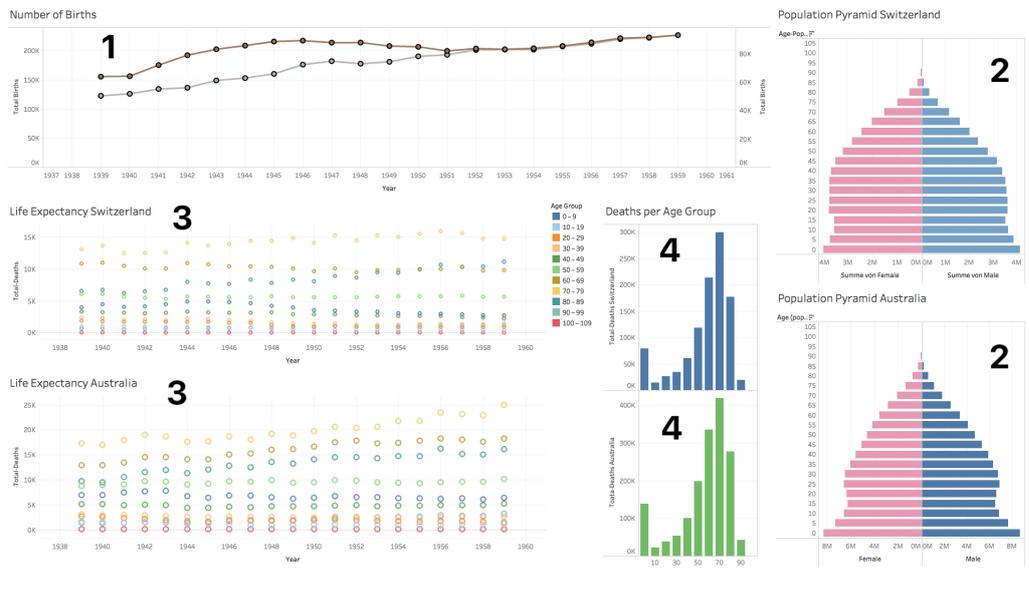
Before heading over to the population pyramid, Robert examined the scatter plot displaying life expectancy. He discovered that child mortality decreased over time due to the improvement of medical health care. A drastic increase of deaths occurred in 1918, due to the spread of the spanish flu during the first world war.

Finally when Robert reached the last chart - the population pyramid - he was able to compare the two pyramids and show the distribution of the population of two different countries: Switzerland and Australia at the time when e.g. the spanish flu spreaded in 1918.

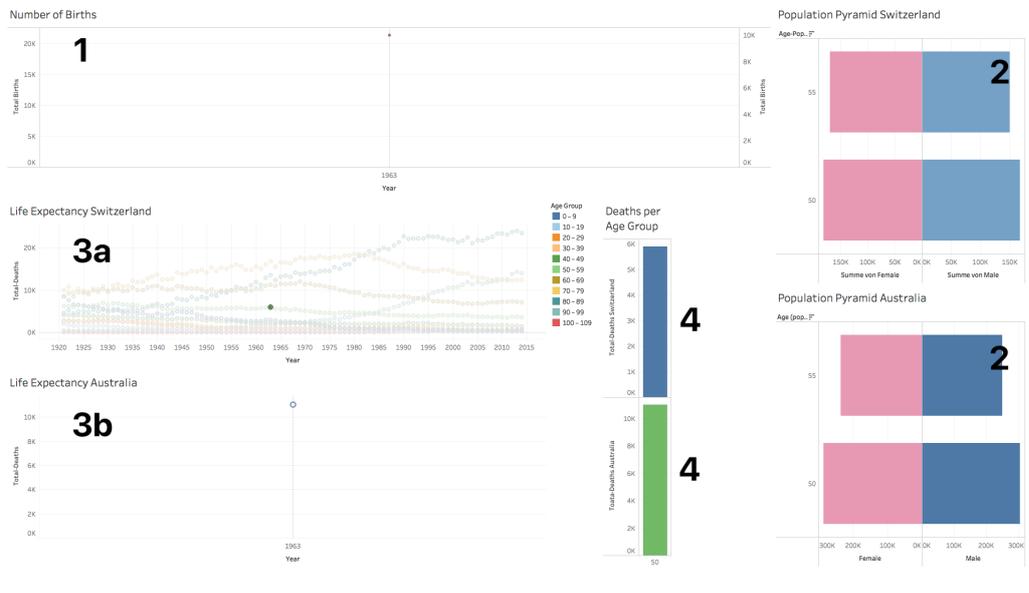
– Screen-shots



This dashboard provides insights into the demographic progress of two countries. Number of births, deaths per age groups, life expectancy and a population pyramid are embedded in our final view. Number of births provides information about the number of births over time. Life expectancy shows the progress of life expectancy over time per age group. Deaths by age group indicates the amount of deaths of a specific age group within a country. The population pyramid assembles the number of males and females per age group in a country. All views are linked; therefore, each view gives information within the selected parameters. When a user selects for example an age group or a gender at the population pyramid, the changes take effect to other six views.



Upon selecting a timespan at "Number of Births" [1], the effects take change to both "Population Pyramid" [2], "Life Expectancy" [3] and "Deaths per Age Group" [4] for the two countries.



By selecting an age group at "Life Expectancy" [3a] a filter is applied to "Life Expectancy" [3b], "Population Pyramid" [2] and "Deaths per Age Group" [3].

5.2 Feedback

The feedback we got from the evaluations helped us on building our tool. At an early stage of our project we received the information from a coach not to focus on a single task but to show a greater picture where the user can explore and discover his own findings rather than guiding him to a result.

We also got the information that the tree-map we are using is not a real tree-map because of the lack of a hierarchy. Also the fact that area encoding is bad for encoding a number like age groups has been delivered by the feedback from our coaches.

6 Discussion

Our visualization transforms a huge amount of data into an easy understandable and meaningful visualization. Our aim was not to guide the user through the data and point out interesting aspects of it. We want to give him a tool with whom he can experiment and discover his own findings.

At first glance the number and simplicity of our graphs might be considered as a weakness in our visualization. But it serves the given purpose, which is to provide a tool with whom you can analyze different aspects of the data and is also beginner friendly.

There were some major changes in comparison to the original idea. We underestimated the time it would take to implement a feature for example. Since it was a relatively small project it was not a big problem. In big company projects though it would be wise to invest more time into the beginning phase of a project. On the other hand, it is very hard to estimate the amount of time it will take when you don't have worked on similar projects.

One strength of our project is, that we realized it with Tableau. Tableau allows experimenting with data on a very high visualization level. The generated charts support numerous user interactivities such as filtering brushing etc. Furthermore, you can easily switch between different charts and look what visualization fits best for your purpose. This is a very cool feature in the implementation process, since you can try different visualization techniques.

6.1 Lessons learnt

One lesson we learnt is that data preparation is a time consuming but necessary task in data analytics. The problem with our data was that we had a lot of different data-files. This data files did not have the same format. In our opinion Tableau is not quite suitable for this use case because of the large number of variables you get when you use many different files. To solve this problem, we first implemented a basic concept of our tool and added more and more features with an iterative approach.

7 Work Distribution

7.1 Responsibilities

	Hauser	Pfeiler
Finalize Dashboard	x	x
Report	x	x

x = responsible for the milestone

7.2 Time Table

(might not be complete)

Datum	Anwesenheit	Tätigkeit(en)
January 15th	Hauser,Pfeiler	M4 Document Latex set-up, added two-countries dashboard
January 18th	Hauser,Pfeiler	M4 Document wrap-up and finalization, upload to homepage

References

1. Human mortality database. <http://www.mortality.org/>. Accessed: 2018-01-20.
2. Tableau. <https://www.tableau.com/>. Accessed: 2018-01-20.
3. James Moody Kieran Healy. Data visualization in sociology. *The Annual Review of Sociology*, 2014.
4. Andrei Zinovyev. Data visualization in political and social sciences. <https://arxiv.org/ftp/arxiv/papers/1008/1008.1188.pdf>. Accessed: 2018-01-20.