

CSC196V: Data Visualization

Interactive Design

Team: Dixita Bhanderi, Pallavi Yadikar, Ravi Trivedi

Dataset Description:

Tobacco smoking is one of the world's largest health problems. Over the course of the 20th century, it killed around 100 million people, most of them in today's rich countries[1]. The health burdens of smoking are now moving from high-income to low-to-middle income countries; some estimates have suggested that one billion people could die from tobacco over the 21st century[2]. According to the Global Burden of Disease study, more than 8 million people died prematurely as a result of smoking in 2017.

The data is published by the Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2017 (GBD 2017) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2018 with the time span from 1990 to 2017. It consists of data related to deaths due to smoking with different age categories, countries, gender, consumption per smoker per day, income range for the smoker. Link: <http://ghdx.healthdata.org/gbd-2017>

Design Decisions:

For this interactive design assignment, initially, we explored various D3 libraries for different kinds of charts and their designs. We examined a few datasets related to all of the team members' interests.

We all liked the health-related dataset and decided to work on any health-related problems the world is facing nowadays. After examining the variety of data values of this dataset we finalized this dataset and explored the questions to plot on the data visualizations.

We used the small subset from this dataset to consider for the question we are addressing in this interactive design. The class-in activity and readings done earlier helped us to gain/brush up the basic idea of javascript, D3 and how do we write the code and modify D3 libraries when necessary. We had four group meetings to finalize the dataset, design of the visualization, work division and collaboration.

Alternatively, we have also explored the world's data breaches dataset. We designed our visualization based on D3 Bubble Chart and dropdown. We also looked at some of the well known D3 charts such as Voronoi Diagram, Collapsible Force Layout, Sequence subset. This effort was an insightful experience for us, it gave us a formal intuition on how to consider and understand any dataset, its fields and the relation between them, how to raise questions from the dataset, how to build better data visualizations, how to work with different D3 libraries and functionality.

CSC196V: Data Visualization

Interactive Design

Team: Dixita Bhanderi, Pallavi Yadikar, Ravi Trivedi

Development Process:

Data wrangling:

The dataset holds five different age groups, such as "Under 5 years old", "5-14 years old", "15-49 years old", "All Ages", "50-69 years old", "70+ years old". While doing data wrangling, there were zero deaths from under 5 years and a 5-14-year-old age group. Hence, we removed those categories in our visualization design.

Questions:

For this interactive design, we are answering the question of “Death rates from smoking by age over years” (Visualization design 1) and “Country wise death rate change over years” (Visualization design 2).

Visualization Designs:

As our data has time series values (years) we decided to give selection for the years in the visualization design. Numerical values for deaths we decide to plot using a bar chart. We used two different libraries from D3, A slider for time and a horizontal bar chart and merged them in one visualization design and used the play button to quickly slide over all the years. Finally, we gave color encoding to the different age group bars and tooltip and hover effects. Our caption for the visualization is also dynamic, year changes according to the selected year.

From the first visualization page, we can navigate to the second visualization design to see another question. This design holds a selection drop-down for the countries. Dropdown helps to filter the data by each country. The bar chart shows the number of deaths from the age group greater than 70 years old.

Work and time Distribution:

Initial explorations and findings related to dataset and interactive visualization tools (All - 6 hours total)

Initial efforts of selecting visualization designs & charts (All - 5 hours total)

Alternative designs & explorations (Dixita - 7 hours)

Health data exploration, insights, & finalizing designs (All - 2 hours)

Dataset wrangling and d3 libraries, slider & various design charts exploration (Ravi - 3 hours)

Data visualization design & question findings (All - 2 hours)

Data visualization wrangling, D3 bar chart & slider merge code implementation (Dixita - 10 hours)

Data visualization wrangling and D3 bar chart & dropdown code implementation (Pallavi - 8 hours)

Report (All - 5 hours)

CSC196V: Data Visualization

Interactive Design

Team: Dixita Bhanderi, Pallavi Yadikar, Ravi Trivedi

Interactive designs:

Design 1: Death rates from smoking by age, over the years (from 1990 to 2016)

As we can see in our first visualization of the breakdown of the deaths, most deaths from smoking occurred in older populations. When we look at the breakdown of deaths from smoking by age we see that it is mainly older populations that are affected.

In the visualizations, we show the death rates from smoking by age category and the share of annual deaths that occur in each age group. Here we see that death rates from smoking are much higher in people older than 70 years old, followed by those aged 50 to 69. Death rates for younger adults and children are very low.

This has also reflected in the number of deaths by age: in 2016 just over half of the people who died prematurely from smoking were older than 70 years old, and around 93% were older than 50 years.

Design 2: Country-wise death rate change over the years (for Age>70 Category)

With the help of the above insights, we thought of creating another interactive visualization where we can see death rates over the years for a specific category of people (i.e Age> 70). In this design, we used a dropdown where we can select the country and then a bar chart for a selected country is plotted with new values.

For this design, we had to wrangle our dataset again because needed different combinations of rows and columns from the actual dataset. We used trifacta for wrangling the data.

Summary:

This assignment made us familiar with D3, JavaScript, CSS, and other client-side web technologies. This will be helpful for our final project where we need to create many more interactive designs.

References:

[1] Jha, P. (2009). Avoidable global cancer deaths and total deaths from smoking. Nature Reviews Cancer, 9(9), 655.

CSC196V: Data Visualization

Interactive Design

Team: Dixita Bhanderi, Pallavi Yadikar, Ravi Trivedi

[2] Mathers, C. D., & Loncar, D. (2006). Projections of global mortality and burden of disease from 2002 to 2030. PLoS medicine, 3(11), e442.