

CDC_Viz Final Report

Zanqing Feng ¹

University of Colorado, Boulder

Chase Kregor ²

University of Colorado, Boulder

Rachel Lewis ³

University of Colorado, Boulder

ABSTRACT

Visualizations can be used to provide insight and knowledge into the deaths from 1999-2015 within the United States. This information can be used to give insight to people's lives right now and can educate them on what health issues they may be at more risk of because of their race and gender. These visualizations will help people understand the relationship between race, age, and gender.

One visualization was created to allow the user to explore the number of deaths of each race or gender. This could then be narrowed down by selecting which state to take data from.

Keywords

ICD Chapter, Crude Rate, D3JS, Matplotlib,

Index Terms

Visualization, Interaction, Cause of Death, Race, Gender, Age,

1. Introduction

Understanding what people have died of and the relationship of race, age, and gender to the ways people die. Exploring this data allows insight into the nation's health on a large scale. Knowing the possible diseases of which you might die is very interesting and seeing how it changes depending on the different factors can lead to interesting questions that can be explored further by medical professionals.

2. Related Work

Inspiration for this project came from several other articles that use the same dataset.

2.1 How You Will Die

One of these articles is *How You Will Die* by Nathan Yau. Nathan utilizes the data to create a visualization using gender, race, and

age to show the percentage of people matching those parameters die of which diseases. With age included in the calculation, there is a way to control which causes may or may not be included. For example, if the age is 0 years old, perinatal deaths are a possible cause, but when the age is past that part of life, there will be no perinatal deaths factored in the percentages. [1]

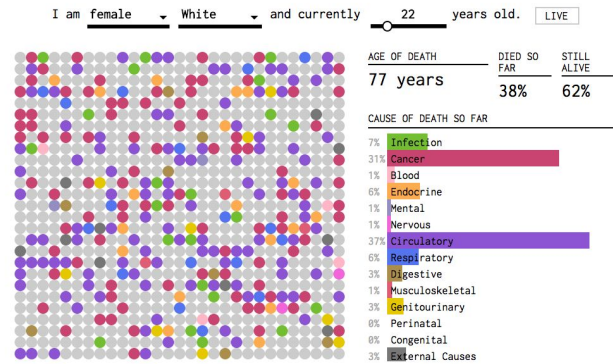


Figure1: Death visualization by Nathan Yau.

2.2 Causes of Death

A second article using this information is *Causes of Death*, also by Nathan Yau. The data was used within a flow chart to show the percentage of deaths each cause is responsible for at each age. This gives a much better high level view of how people die throughout the span of a lifetime. [2]

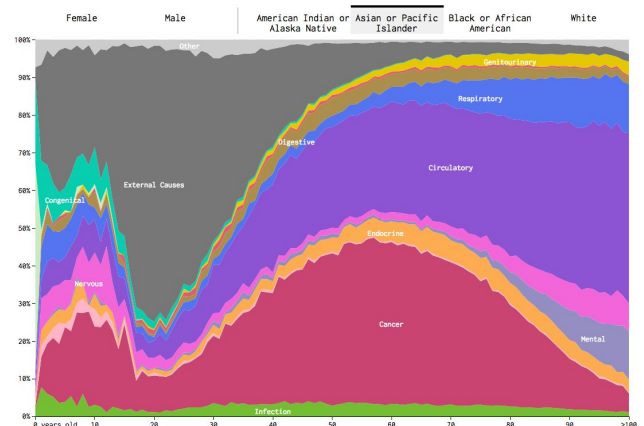


Figure2: Death visualization by Nathan Yau.

2.3 Dataset Documentation

Understanding the various intricacies of the dataset is always a good thing to understand. In reading the dataset's documentation we were given a greater understanding of how the numbers were curated and reported. [4]

2.4 Institute for Health Metrics and Evaluation

The Institute for Health Metrics and Evaluation is doing a very large scale visualization in this area. They allow the changing of variables such as country, disease, gender, and years to be

¹ Zanqing.Feng@colorado.edu

² chkr5069@colorado.edu

³ rachel.a.lewis@colorado.edu

accounted for. The method of representation can also be changed between time series, age patterns, and maps. [3]

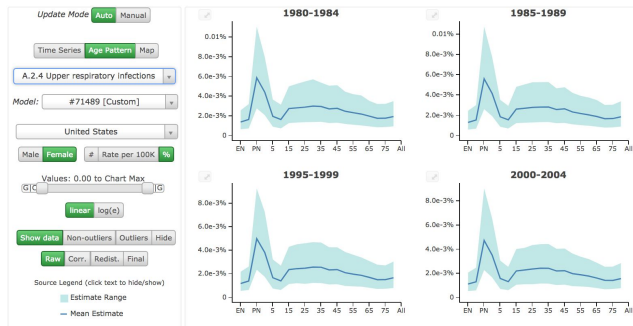


Figure3: United States female deaths from upper respiratory infection.

2.5 Top 10 Leading Causes of Death: United States, 1999–2013

The National Center for Health Statistics, a branch of the Center for Disease Control is doing a Data Visualization Pilot. This pilot includes various projects on a variety of health topics. The projects include tableau dashboards that provided some insight into the type of subfield and interactions we wanted to represent in our visualizations. [5]

2.6 USA Life Expectancy

Total deaths by cause based on gender and age. This information goes into more detail about what the causes of death are. For example, lung disease is distinguished from lung cancer. The rank of the cause of death and how many deaths were caused by it are included to give information to the reader. It is easy to follow the trend of the different causes throughout the age groups. [6]

2.7 Chronic Diseases and the Emphasis on Prevention

NYC Data Science Academy, a data science bootcamp published a blog post with some visualizations about chronic diseases. These visualizations helped lead us to the visualizations we created in our tableau workbook. We used this tableau workbook as a quick exploratory and design tool to better understand our dataset and the visualizations we wanted to make from it. [7]

2.8 Top 10 Causes of Death: World Health Organization

This article from the World Health Organization divides up the causes of death by economic income group. More than half of all deaths in low-income countries were caused by conditions such as communicable diseases, maternal causes, and nutritional deficiencies. This number is less than 7% in high-income countries. This can lead to research in what aid can be offered to these lower-income countries to reduce the number of preventable deaths. [8]

3. Project Implementation

3.1 Data Wrangling

We used the Center for Disease Control's CDC Wonder database to perform a request on the correct subfields for the Cause of Death, 1999-2015 database. We grouped by ICD Chapter, State, Gender, and Race. Unfortunately the CDC Wonder Database only lets you pull a certain amount so we were not able to include age as well for our D3 visualization.

3.2 Data Analysis

Some sections of the crude rate are deemed "unreliable" due to low numbers of reporting. You mainly see this in the Native American population demographics. Here we were presented with the problem of the inclusion or exclusion of these numbers. We ultimately chose to include these numbers creating our own version of the crude rate, called crude rate revised in the dataset with the acknowledgement that in very little instances the numbers may not be absolutely perfect but still provide adequate numbers to provide insightful and useful visualizations.

3.3 Data Visualization

3.3.1 D3 Visualization

Race and gender are two very interesting categories that reveal a lot of information in the dataset. For this reason they were chosen as options to organize the death information. Selecting which state you live in can also give interesting insight. For example, the amount of deaths from external causes of morbidity and mortality might be larger in a state with many large cities or factors that increase danger.

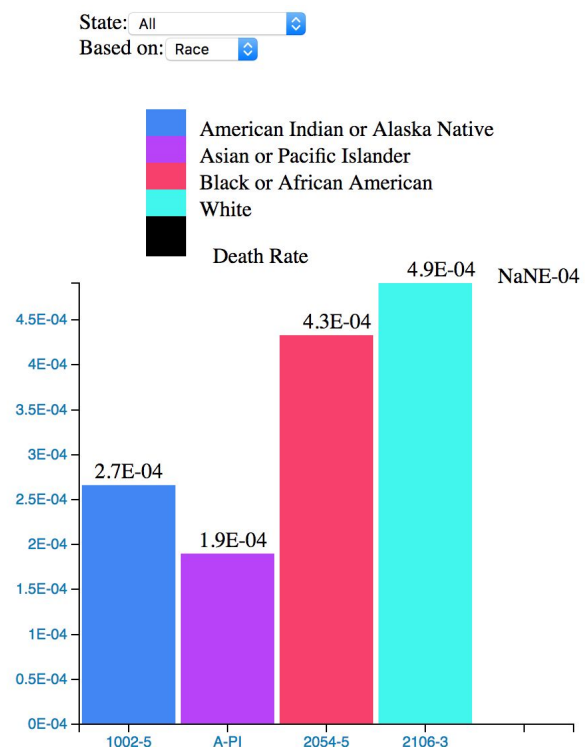


Figure4: Death based on race in the state of Colorado.

A correlated visualization gives further insight into the data. In addition to the number of deaths, this shows what the causes are for that demographic selected. Easily seeing patterns in this data can lead to different questions for further research.

3.3.2: Jupyter Notebook Visualization

For this part of the project we wanted to create a more accessible version of our visualizations. In this notebook the user can input the race, gender, and disease they want to study. The output shows the crude on the y axis over the people's life span on the x axis. We displayed this visualization both in seaborn and matplotlib. We believe providing a jupyter notebook with the code and implementation of the visualizations provides a lower barrier to entry in understanding how to make interesting visualizations from these types of projects.

Visualizing your selected disease in Matplotlib

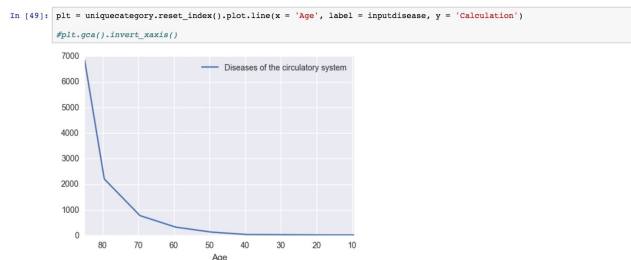


Figure5: Matplotlib visualization within the Jupyter Notebook

Visualizing your selected disease in Seaborn

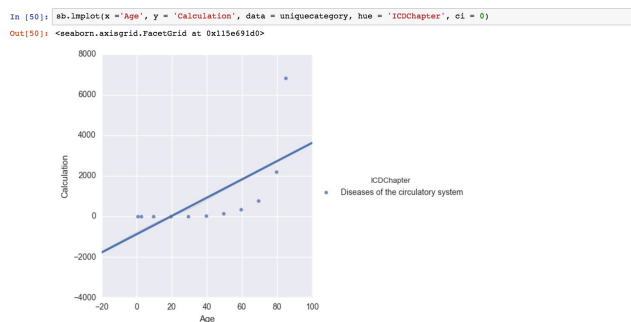


Figure6: Seaborn visualization within the Jupyter Notebook

4. Findings

There is a lot of research into the area of causes of death on large scales. Visualizations are being used to make this information more attainable to the common person and shed light on what may cause their death.

Social patterns can also be seen using visualizations. Using attributes in other data sets that include factors such as suicide, homicide, and car accidents can be used to see what social conditions contribute to the dominance of each of these causes of death.

Visualizing this dataset along with other causes of death datasets proves useful in enabling users to quickly understand public health at a macro level.

The rate at which people start to die goes up exponentially starting at the 60 year mark regardless of race or gender.

While we understand that these snapshots are not necessarily absolute, rather we acknowledge that these are undeniable trends and forecasts into the ways in which institutions and various group of people live out their lives.

ACKNOWLEDGEMENTS

Dr. Danielle Szafir for feedback and direction.

Dr. Brian Keegan for teaching jupyter notebooks, matplotlib, and seaborn.

REFERENCES

- [1] Yau, N. (2016, August 30). How You Will Die. Retrieved May 08, 2017, from <https://flowingdata.com/2016/01/19/how-you-will-die/>
- [2] Yau, N. (2016, September 12). Causes of Death. Retrieved May 08, 2017, from <https://flowingdata.com/2016/01/05/causes-of-death/>
- [3] Institute for Health Metrics and Evaluation (IHME), Causes of Death (COD) Visualization. (n.d.). Retrieved May 08, 2017, from <http://www.healthdata.org/data-visualization/causes-death-cod-visualization>
- [4] CDC WONDER FAQ Help Contact Us WONDER Search. (n.d.). Retrieved May 08, 2017, from <https://wonder.cdc.gov/wonder/help/ucd.html#Top15>
- [5] NCHS Data Visualization Pilot. (2017, April 21). Retrieved May 08, 2017, from <https://blogs.cdc.gov/nchs-data-visualization/>
- [6] USA CAUSES OF DEATH BY AGE AND GENDER. (n.d.). Retrieved May 08, 2017, from <http://www.worldlifeexpectancy.com/usa-cause-of-death-by-age-and-gender>
- [7] Academy, N. D. (2016, October 27). NYC Data Science Academy. Retrieved May 08, 2017, from <http://blog.nycdatascience.com/student-works/r-visualization/chronic-diseases/>
- [8] The top 10 causes of death. (n.d.). Retrieved May 08, 2017, from <http://www.who.int/mediacentre/factsheets/fs310/en/index1.html>