Death by 3300 – An Interactive Data Visualization

Rob Aga: rra46
Eric Mei: em674
Ankur Biswas: ab2249

**Note: Please open this project in full-screen mode and on higher resolution for optimal viewing experience.
For full-viewing experience (aka to include music!) uncomment lines 200-214. The play/pause button will appear on the upper right-hand side.
Thank you and Enjoy.

Overview of the Project:

Death is inevitable. It is the one thing that is for certain that will happen to everyone in their lifetimes. Even though the consequence of death is the same for us all, people encounter death in different ways; some more painful than others. This happens because of a myriad of circumstances. One bad decision. One mistake. The lottery of birth. The causes of death also affect our perceptions of certain phenomena and experiences. Our aim in this project is to present a statistical overview of the reasons behind death in the modern era. Are sharks really as deadly as Jaws make them out to be? Is gun violence in the United States really that big of a deal? Is there any merit behind a law that criminalizes parents for not getting their children vaccines? These are some of the questions we hope to answer by examining both through a historic and present perspective, what events kill the most number of people.

Description of the Dataset:

The final dataset is as follows:

Col Name	Use	Data Type	Example
Causal Entity	An identification of the cause of death.	General String	Shark Attacks, Chicken Pox etc.
Number of Deaths	Number of Deaths Caused	Integer	1, 5, 10 etc.
Category	Broad Categorization of Data –later used for filtering	General String	Vaccine Effect, Animal, War etc.
Year of Accuracy	To identify whether the deaths caused were over a range of time or specific to a particular year	Integer	1960, 2015 etc.
Description	Text description of the data – to be presented to viewers of the application	General String	
Source	To maintain and ensure accuracy of the data and also to make sure that the data was being aggregated accurately	Hyperlink String	

Image	An image for the SVG display	PNG	

Total number of Data Points: 85

Filtering of the dataset:

The following table shows the main problems we identified in filtering the data and the steps we then took to maintain accuracy:

Problem	Description	Probable Cause	Filter Solution
Too many estimations	Death Statistics often come in ranges. A large number of these ranges greatly over estimate the number of deaths caused by a particular phenomenon.	Organizations that have these statistics are not objective in their nature. They are often pushing for some political or social agenda. Therefore they have an incentive to inflate statistics.	We tried to verify the data through credible and verified sources such as the UNHCR, WHO etc.
Lack of Specialization	Aggregate websites don't have single fields of specializations and the data can be misrepresentative.	Nature of how aggregation across multiple fields works.	We tried to get data from every category from a corresponding website in its field of expertise. Ex- CDC and WHO for disease statistics.
Too many data points	There are hundreds of diseases and catastrophic events that kill people every day.	Too much happening in the world.	We filtered on two parameters: 1. Frequency: We thought the frequent deadly diseases deserved places in the visualization. 2. Renown: We thought that phenomenon that have the reputation for being deadly deserved a place in the visualization (Ex Shark Attacks) especially to see how they compared to other data points in other categories.
Individual Outliers	There are certain events that skew the data of one category in a	Events like 9/11 are likely to increase deaths by terrorism, the Holocaust is likely	Events of a specific catastrophic nature were given their own points on the plot. Categorization helps to relate

	particular direction.	to increase deaths by murder.	them with other data points in a similar category.
Lack of readability due to too much variance	An interactive bar plot was unlikely to work because there is so much variance across the data we have: from 0 deaths attributed to marijuana usage to tens of millions due to smallpox.	Given the size and scale of the data (the sheer number of deaths taking place and number of ways to die) too much variance and a bad visualization as a result was bound to be a challenge.	To increase interactivity and readability of the data we created a custom elliptical plot which: 1. Sorts events by order of magnitude and arranges them in a respective radii. 2. Zooms outward so that newer data doesn't overlap with older data – but at the same time the older data isn't eliminated until it becomes insignificant.

After three rounds of scraping and filtering, we aggregated the data onto a CSV file and used it for our data visualization.

Data Mapping and Scaling Used:

- We mapped the number of deaths of a cause to the size of an image of that cause. In other words, if cause A has twice as many deaths as cause B, cause A's image will have twice the area of cause B's image.

- We then organized the causes into groups based on order of magnitude, and placed each group of causes onto its own ellipse. The end effect is a collection of concentric ellipses consisting of images whose sizes is proportional to the cause's number of deaths.

- To figure out how large to make each ellipse, we created our own method of rough estimation. Each ellipse has the same ratio of minor semiaxis to major semiaxis, but each axis is scaled according to the sum of all the widths and heights of the images we wish to place on the ellipse as well as the square root of the number of images. This method was developed mostly by trial and error, until we found a way to make visually appealing ellipses.

- It was tempting to just scale each ellipse by a factor of ten, but, although the items are grouped by factors of 10, the number of items in each group is different. When we tried this simple scaling, ellipses with many items would look too crowded, while ellipses with few items would look too sparse.

- However, the scaling for the zoom function did look very good when done by factors of ten. One unit of increase on the slider increases the width and the height of the current view box by one factor of ten.

Observations:

Our visualization tells us some obvious and some non-obvious things:

1. Cultural Norms dictate our understandings of certain issues even though data points in another directions: Movies like Jaws and a misunderstanding of sharks in general have created this perception of sharks as deadly creatures that will destroy humans upon contact. Yet, dogs were responsible for orders of magnitudes of more deaths than sharks in 2015.
2. Doctors sometimes get it wrong: It is interesting to note that the number of people killed by Hepatitis-B got higher in 2002 even though a vaccine has been established. Our research leads us to believe that the virus mutated to increase the spread of the infection, something that the medical community didn't account for happening in the way that it did. This probably explains the rise.
3. Political Power can be misplaced: Combating international terrorism has been on top of the American foreign policy agenda for a couple of decades now. The American government famously launched the 'War on Terror' in the Middle East as well and has spent billions of dollars and thousands of man hours on the same. It is interesting to note then that gun violence in the last year in the United States alone has claimed more American lives than terrorism has in the last ten. Our data suggests that the government should rethink its priorities and focus their interests at home rather than abroad.
4. Pre-Natal Care needs massive improvement, even in the 21st Century: It is sad to see that the lack of access to proper infrastructure led to the deaths of 1.1 million women in 2012. These are largely preventable deaths and steps need to be taken to increase access to pre-birth healthcare, especially in the developing world.
5. Marijuana does not kill people: That's it. Marijuana overdoses have not claimed a single life legally or illegally since it was legalized in the states of Colorado and Alaska.

Bibliography:

CSV File contains the Data-Set with all data sources attached.

P.S. We spent a long time on this project in order to create our own custom zoom and image generation functions without using d3. However, if you would like to see a 700,000 line prototype svg, please let us know.