INFO 5100 Lucy He (lh486), Yeon Soo Park (yp246), Richard Quan (rq32)

P3: Mobile Data Visualization May 24, 2016

How & When Will You Die?

A. Data

Source

We gathered data from the Kaggle Database[1] on the Center For Disease Control and Prevention's detailed report on every death in the United States in 2014. Each row in the DeathRecords.csv is an individual death record. The original DeathRecords spreadsheet includes information on the cause (ICD-10 code) and time of death and various demographic background of the deceased like age, sex, race, education, resident status, and marital status. All data came from the CDC's National Vital Statistics Systems[2], with the exception of the Icd10Code[3], which are sourced from the World Health Organization.

Variables

We decided to focus on the following seven variables: age, sex, race, education, ICD-10 code (death cause), month of death, and day of the week of death.

For the death cause, we decided to focus specifically on the top 8 causes obtained from the 2014 Leading Causes of Death provided by the National Center for Health Statistics [4]. This is because the top 8 deaths accounted for a majority of all deaths and any other causes were not as significant in number. We further cleaned up our data by removing the Unknown category from education to keep the dataset coherent as well as the Other (Puerto Rico Only) category from race since there was no death record within this category for these causes.

---

[1] https://www.kaggle.com/cdc/mortality
[2] http://www.cdc.gov/nchs/data_access/vitalstatsonline.htm#Mortality_Multiple
[3] http://www.who.int/classifications/icd/en/
[4] http://www.cdc.gov/nchs/fastats/leading-causes-of-death.htm

This is a list of all the variables that were used in our visualizations:

- **V1: Age**
  - 1 - 114
- **V2: Sex**
  - M
  - F
- **V3: RaceRecode5**
  - 1: White
  - 2: Black
  - 3: American Indian
  - 4: Asian or Pacific Islander
- **V4: Education2003Revision**
  - 1: 8th Grade or Less
  - 2: 9th-12th Grade (No Diploma)
  - 3: High School Graduate or GED Completed
  - 4: Some College Credit, No Degree
  - 5: Associate's Degree
  - 6: Bachelor's Degree
  - 7: Master’s degree
  - 8: Doctorate or Professional Degree
- **V5: Icd10Code**
  - I01-I52: Heart Disease
  - C00-C97: Cancer (Malignant Neoplasms):
  - J40-J47: Chronic Lower Respiratory Disease
  - V01-X59: Accidents (Unintentional Injuries)
  - I60-I69: Stroke (Cerebrovascular Diseases)
  - G30: Alzheimer's Disease
  - E10-E14: Diabetes
  - J09-J18: Influenza and Pneumonia
- **V6: MonthOfDeath**
  - 1: January - 12: December
- **V7: DayOfWeekOfDeath**
  - 1: Sunday - 7: Saturday

Initial Data Load

We loaded the data from the DeathRecordsFinal csv file into different arrays pertaining to each cause of death: `heartDis`, `cancer`, `respDis`, `accident`, `stroke`, `alzheimer`, `diabetes`, and `flu`. This was done by checking if their ICD-10 codes fit in the ranges described above.

Image File

We found a tombstone image that was used in the page header from Cliparts.co[5] and removed the background with Adobe Photoshop.

B. Visual Elements

Demographics Input

We created four responsive button dropdown menus for each of the four demographic factors: age, sex, race, and education. For each variable, we added the No Preference option to allow users to select and filter the data even with a subset of those demographic factors instead of forcing users to use each and every factor. Users use their mouse to press on each dropdown menu or finger to touch and select choices.

Spinning Wheel Visualization

We created a “wheel of misfortune“ to show an interactive donut chart-like perspective of the distribution of the top 8 death causes as different personal attributes such as age, sex, race, and education level changed. Although the death causes were parsed down from their ICD-10 code format and organized into arrays, we found that it would be too data intensive to directly map the arrays onto our wheel. Instead, we opted to create a data object that held the label of the cause of death along with the count of the number of deaths in that category. This was much faster to update when the personal attribute inputs changed. The arcs are sized according to the count of each cause of death, very similar to how one would create a d3 pie chart. The text on each arc is rotated to the center of each arc, and comprises of the label of the cause of death along with a percentage of that cause compared to the total number of deaths currently being displayed. Text labels of the causes of death that comprise less than 1% of the total displayed deaths are hidden, although the small arcs themselves are still shown.

The demographic inputs are mapped to variables in the code. When an attribute is

[5] http://cliparts.co/tombstone-vector

changed on a dropdown, a change function is called which takes all the variables (set to null by default) and recalculates the count of each top 8 death cause. The recalculation is a simple filter of the full death cause arrays, checking if each data point i.e. victim's personal attributes match the currently selected personal attributes. The filtered arrays are placed in: `heartDisSel`, `cancerSel`, `respDisSel`, `accidentSel`, `strokeSel`, `alzheimerSel`, `diabetesSel`, and `fluSel`. When the inputs change, and the data is updated, the arcs and arc text labels dynamically move and resize to their new value to show differences between specific personal attributes. For example, changing the age range to "1 to 20" shows that young people die of accidents a lot more often than old people do (because cancer, heart disease, etc. are less likely). If there is no data for a certain personal attribute combination (victim aged 1-20 with a PhD), the words "No Data Found" are shown to the user.

The wheel itself has two interactive properties. One is the random spin, which is triggered by pressing the spin button in the middle of the wheel. This rotates the wheel at least two full rotations in addition to a random degree from 1 to 360. The rotation has a tweened animation to allow a small amount of suspense. This is meant to serve as a lighthearted random chancing of the user's cause of death based on their personal attributes. Where the wheel stops is the random cause of death output, and is used to later decide when a person may die based on previous deaths of that category. The second interactive property is the user's ability to drag the wheel around. This allows the user to see data for a different cause of death manually. The random spin and drag motion functions are triggered by both touch and mouse events, and thus will work on both desktop computers and mobile devices. The x and y positions of the touched or clicked position on the wheel is used to calculate an arctan which calculates the angle of the position that is being clicked on the wheel. The wheel rotates accordingly to follow the mouse or touch.

The angle of the wheel's rotation is also how we know where the user has randomly spun or dragged the wheel to. When the wheel is populated or updated with cause of death data, an array called sliceArr is created which holds the labels of each cause of death, and a different number of each of them based on the count of each cause. The current angle of the wheel can be taken as a fraction of 360 degrees, and that fraction can be used a fraction of sliceArr to index, in turn showing the currently picked cause of death. The currently selected/spun cause of death is shown in a larger text headline under the wheel, and is updated during spins, drags, and attribute selection.

```
angleDeg = Math.atan2(y, x);
picked = Math.floor(((angleDeg % 360) / 360) * sliceArr.length);
```

Calendar Heat Map Visualization

Our second major visualization is the calendar heat map, which is a calendar-like visualization that takes into account the four personal attributes and the cause of death selected to display the most probable month of death and the probability of dying in each day of the week for that particular month.

As soon as the cause of death was selected, we reorganized the data for that cause (already filtered by the four personal attributes) so that the data can be easily accessed by using `deathMonthData[month][day]`. We then calculated the month with the highest death months and then calculated the death counts for each day of the week. Using those counts for each day, we created the opacity scale for the heat map display:

```
var opacityScale = d3.scale.linear().domain([dayCountMin,
dayCountMax]).range([0.1, 1]);
```

We used this scale to change the opacity of bars for the seven days so that the day with the least death counts would have a very light maroon color and the day with the least death counts would have a very dark, striking maroon color. We chose the opacity of 0.1 so that the lightest bar is very light but is still visible. We also displayed the color legend at the bottom of the calendar with dynamically changing minimum and maximum count labels to provide users with a clearer idea of how many people have died in a highlighted weekday of a month.

Responsive Design

We used a responsive Twitter Bootstrap framework consisting of a fixed navigation bar on the top, header jumbotron section, and a container for the selection drop-down menus and visualizations. Moreover, we used media queries to hide and display hidden classes like smallsize-only and largersize-only that moved parts of the header to the center and removed large chunks of text based on the screen width for mobile, tablet, and desktop versions. Moreover, we also made each list item on the selection dropdown-menus for age, sex, race, and education expand to full width by applying a btn-responsive class that adjusts for smaller devices. We set the width of both the wheel and the calendar to be 300px, which is smaller than the width of the narrowest smartphone screen (320px for iPhone5), and made them float horizontally so that they are always centered within each fluid column of Bootstrap's grid system. And with the help of Bootstrap's grid system, we were able to display the wheel and the calendar side by side on desktops and vertically stacked on smaller screens.

C. Story

Over the last five years, the main causes of death in the US have remained fairly consistent, although accidents became the fourth leading cause of death in 2014, while stroke became the fifth. Just two, heart disease and cancer, account for over half of all deaths in the country. This trend has persisted for many years.

It's been said that "statistics are human beings with the tears wiped off." This is especially true with this dataset. Each death record represents somebody's loved one, often connected with a lifetime of memories and sometimes tragically too short. With this in mind, we decided to display this information in a casual, lighthearted manner with a color scheme centered on shades of maroon and grays to reflect death.

Our two visualizations allow users to create a demographic profile for themselves by inputting variables for age, sex, race, and education and watching how slices of the wheel change based on how likely they are to die from the top eight leading causes of death in the United States. As users input one or more of these demographic variables, they not only see the wheel change but also see a calendar heat map depicting what month and day of the week they are most likely to die on for that particular cause of death. The “wheel of misfortune” interaction was meant to serve as a sarcastic perspective of the chance of each cause of death.

Interestingly, we found that younger ages are more susceptible to dying from accidents and cancer compared to older ages who were likely to die from heart disease. We also discovered a slightly higher percentage of females dying from old age diseases like alzheimer's compared to men most likely due to the fact that women live longer than men. Moreover, we discovered that American Indians are the two times more susceptible to dying from accidents compared to other races. We discovered little difference in causes of death for individuals with different educational backgrounds. Across all demographics, individuals die most from Heart Disease, Diabetes, Flu/Pneumonia, Cancer, and Respiratory Disease in January and died from Stroke and Alzheimer’s in December. This is most likely due to the fact that low temperatures cause more problems for the body's cardiovascular and respiratory system. Interestingly, most people died of these causes on Wednesday compared to any other day of the week.

References:

https://d3js.org/