

How Much Time You Have Left and How You Spend It

INFO 5100 Project 2

Min Zheng, Kexin Jiang, Yicheng Zhu

Overview

Through this project, we would like to expose the viewers to the topic of death, and make them realize how precious the time is in order to cherish their lives. We started by showing the top 5 death causes of U.S. residents at different age groups. Then, we designed an individualized experience to let viewers input their demographic data to see how much time they have spent in their whole life and how much time is left for them to spend in the future. By interacting with the time dots, users can also obtain information of their daily time usage in different life stages, and get a general sense of how they would be spending the rest of their lives.

Data

Three datasets have been combined in this project. The first is used to summarize the number of deaths and major death causes for each age group. The second is used to predict the time expectation length, and the last one is for break-down of the time spending. The links of those three datasets are provided at the end of this report.

The first dataset is from CDC. It is the annual report on death in the US under the National Vital Statistics Systems. It includes information of education level, death age, sex, autopsy, death cause, race, and etc. from the year of 2005 to the year of 2015. Python pandas library is used to subtract and group the information we need including gender, death cause, and death age. And only 2015's data is used because it contained enough amount of data points. For specific ages, the top five death causes are chosen for keeping the data valuation clean and free of distraction. The age of data has several metric codes. We chose the one that binned by [<1], [1,4], [5-14], and then increases by step size 10 years, since it best captures the overview of the age of interest. Despite that the data contains information about different demographic background and residency status, we decided to go with US residents only.

In the report of both US government and WHO, this dataset is also used to determine life expectancy of male and female in the US. We applied this data in the second plot too. We used the prediction up to 2015, which is consistent with the data we chose from the first plot. Based on the report, it showed a difference between men and women for the life expectancy; thus, gender is taken into consideration here as an input.

The third data set contains the information of total time distribution for different activities under specific age groups. We aim to provide a personalized information but easy user experience. Considering the combination with the second plot, where we have the user to choose gender and age, we decided to split the data into the same manner: specified by age group and gender as well. Since the selection of age is continuous, but the data for time use is in age range, we decided to introduce a mouse hover effect on the second plot as an selection button to view the third plot (time use). A normalization is applied to get the percentage of time spent on specific

activities. Similarly, the dataset showed a difference between gender about how they spend their times; so, both male and female data are chosen.

Table 1 Variable selection, data source, data manipulation and mapping

Index	Variable	Data Source	Data Manipulation	Location in Project
1	age	CDC*		Session 1
2	death cause	CDC*		Session 1
3	death cause percentage	CDC*	$= 2.a / \sum (2.1, 2.2 \dots 2.n)$	Session 1
4	gender	CDC*		Session 2
5	life length left	WHO*		Session 2
6	time spent in specific activity	BLS*		Session 3
7	time spent in specific activity percentage	BLS*	$= 6.a / \sum (6.1, 6.2 \dots 6.n)$	Session 3

*Link:

[CDC/WHO/BLS](#)

Mapping

Overall, we use black color as the background to show how serious the death is. And we use the bright color to present data to show how precious the time we can live.

There are two layers of session one. The outer circle showed the top 5 death causes. The ratio of perimeter showed the percentage of each death causes. The rest of death causes will be counted into others to keep the color theme from being messy. The inner circle represents death, while a number is showing the specific age. The size of the small circles flying in is also proportional to the percentage of each death cause, and the color is consistent. The animation that outliner is blending into the inner center is for presenting the process of death. For all of the circles data, we used `scaleSqrt()` to avoid the confusions.

For session two, after two inputs - age and gender, the circle matrix showed your total life length. Each circle represents each year, in which grey presents the life you have already lived and the green shows the rest of life you could possibly live. The animation is inspired by hourglass to show the eclipse of time.

After clicking on each age group, a bubble chart will show the time spending breakdown, while different colors representing different activities and the radius is mapped by `scaleSqrt()` to the percentage of activity spending time.

Story

Time is an interesting topic. A few fun facts about time spending could be like: we only spent 115 days laugh for whole life but spent 5 months complaining. Or we spent 10.5 years working while spent half year longer on TV. So have you ever wondered how long could you live, what causes you to die and how could you spend the rest of your life better? Our project aims to give you some insights by presenting the cause of death based on age, and allowing the user to explore how much time left in his/her life and how he/she spends their time in each stage of their life.

Dataset

<https://www.kaggle.com/robotgames/cause-of-death-word-cloud/data>
<http://apps.who.int/gho/data/view.main.SDG2016LEXREGv?lang=en>
<https://www.bls.gov/charts/american-time-use/activity-by-agem.htm>