**Team 15 Project Proposal: The health rankings system in the United States**

**Problem:**

Analyzing the overall life expectancy based on the gender, age and disease across 51 states in the United States

**Motivation:**

Mortality is one of healthcare outcome measurements. If we combine the mortality with causes of death, it could be used to evaluate how disease/non-disease affect our life expectancy.

World health organization (WHO) and the United States centers for disease control and prevention (CDC) make the statistics data public every year. Understanding these data remains a major challenge. To solve this problem, we decided to build up the mortality ranking system, integrate the mortality with causes of death, and visualize them. It helps us to choose the good life style in the future.

**Dataset:**

**World health rankings website:**

**(**[**https://www.worldlifeexpectancy.com/usa-cause-of-death-by-age-and-gender**](https://www.worldlifeexpectancy.com/usa-cause-of-death-by-age-and-gender)**)**

The entire data are from the public website (shown as above), which contain the death rate, the causes of disease, gender, and the age for different states in the USA.

The house-hould income for 51 states in the United States in 2017, from #####

**Methodology**

**Packages**

**1 Pandas**

* **Official website:** [**here**](https://pandas.pydata.org/)
* **Installation**

**The best way to get pandas is via conda:**

**conda install pandas**

**OR**

**pip install pandas**

**2 Sqlite**

* **Official website:** [**here**](https://docs.python.org/2/library/sqlite3.html)

**3 Numpy**

* **Official website:** [**here**](https://www.scipy.org/install.html)

**4 Bokeh**

* **Official website:** [**here**](https://bokeh.pydata.org/en/latest/)
* **Installation**

**The best way to get bokeh is via conda:**

**conda install bokeh**

**OR**

**pip install bokeh**

**5 Holoviews**

* **Official website:** [**here**](http://holoviews.org/)
* **Installation**

**pip install holoviews**

**OR**

**conda install -c pyviz/label/dev holoviews**

**6 Plotly**

* **Official website:** [**here**](https://plot.ly/python/)
* **Installation**

**pip install plotly**

**7 Matplotlib**

* **Official website:** [**here**](https://matplotlib.org/)
* **Installation**

**python -m pip install -U pip**

**python -m pip install -U matplotlib**

**8 Seaborn**

* **Official website:** [**here**](https://seaborn.pydata.org/)
* **Installation**

**The best way to get bokeh is via conda:**

**conda install seaborn**

**OR**

**pip install seaborn**

**The plan**

Our proposed solution is to use a python package (BeautifulSoup) to extract data from the websites (shown above) and build up the overall health ranking system and house-hould income across 51 states. Firstly, we organize the data according to gender, ages, and states. Then we are going to visualize the data by showing the number of deaths vs. different causes.

Our system can provide health advice to people who input their location, age, and gender.

**The timelines and steps for the plan**

|  |  |  |
| --- | --- | --- |
| Step | Estimated completion time | Person(s) in charge |
| **1. Extract the data, and organize/clean up the data by the states** | **One week** | **Shang-Wei Hung, Ye Gao** |
| **2. Build the health evaluation system ( ages, genders, causes of death, and states)** | **Two week** | **Ye Gao, Yuhan Zhang** |
| **3. Use bokeh to do data visualization ( interactively analyze the data and trend)** | **Two week** | **Shang-Wei Hung, Yuhan Zhang** |

**Visualization**

Please see file: ######## (the combined files).ipyn

Here (link for these files)

Please click here to see the interactive ranking system (show the link for our ranking system)

**The presentation slides Here (link)**