**Project 2 Proposal**

Khrystyne Clifton, Benjamin Brown, Donald Ricumstrict, Andrew Doran

Dataset:

World Happiness Report- https://www.kaggle.com/datasets/unsdsn/world-happiness

Description/Outline:

A brief analysis of happiness by country

Providing a breakdown of satisfaction of various country’s populaces

Questions to answer:

Which country has the highest happiness score?

* The lowest? (Dystopia [imaginary country made as a baseline])

Breakdown of tasks:

Cleaning the data

* Merge CSVs into one file/database
* Omit Statistical Analysis due to inconsistent inclusion (see data quirks below).

GitHub Merge authority

Visualizations

Requirements.txt- “Library==version of library”

Push to github

Visualizations:

Map plot

* Layers for metrics
* Chloropleth
* .bindPopup

Ipyleaflet- Ben found. ***Interactivity approved by Carson- no mobile visualizations, no 3D.***

* Incorporate widgets
  + Graph Happiness Score by year
  + If there’s enough time, include other columns (GDP, health, trust, etc…)

Ethical concerns:

Taiwan is a country? Not according to China 2015 has just Taiwan, 2017 credits it to China

* Geographical area rather than geopolitical- Megan (TA)

Gallup world poll was entirely voluntary

* Limited access to internet/small response rate in certain countries could skew

Ethical concerns arise primarily in how we present the data

* We don’t want to compare countries to countries.
  + We want to compare countries to the world average
  + No personal identifying info- Per Megan (TA)

Data quirks:

2015 is the only spreadsheet with Standard Deviation (0.0479 average).

2016 and 2017 have whisker bounds- If we had the Z-score we could calculate the Standard Deviation.

2018 and 2019 have no Deviation or Whisker data available.

* Hence, we are omitting the Whisker and standard deviation columns from our cleaned dataset rather than having null values included.

Not every country is included in every year. INCLUDE NULL DATA YEARS

2017 doesn’t have regions- Region column invalid. Carson recommended Right merge 2015 and 2017 to copy the regions column over, and manually filling in the NA lines.

DF.merge- ,suffix=[] defaults to \_x, \_y if it didn’t merge (\_x in left, \_y in right) google pandas merge and look at suffix argument.

|  |  |  |
| --- | --- | --- |
| Task | Assigned-1 | Assigned-2 |
| Cleaning the data and merging into one database | Andrew |  |
| GitHub Merge authority | Ben |  |
| Chloropleth | Khrystyne | Ben |
| Popups and zoom to bounds | Khrystyne | Ben |
| Layers- Capitol marker | Don | Andrew |
| Push HTML, JS, etc… to Github Pages | Don |  |
| Analysis | ALL |  |

Manish’s feedback:

Take opportunity to use (for the experience):

Beautiful soup, flask