Data Visualization and Analysis Potter – IB CS

Assignment I Due Dec 16. 18, 20

Each group of 4 will download the dataset and analyze it based on the questions posed.

Each group will need to verify the data. The data could have missing segments, extreme/odd outliers, or something similar

Each group will come up with 6 questions that could use the data to help illustrate. All questions will then be vetted by me, and then paired down to 4 questions.

Each group member will answer one of the questions posed.

Each group must use 3 different visualizations – Scatterplot, Pie, and Bar Graphs, with the 4th being any of the three.

You will present your findings via a Power Point presentation which would be comprised of:

* Description of the Dataset including given headers. (Person A)
* Any hinderances or problems of the dataset and how they were handled (Person B)
* The 4 questions, and how they are interesting (Person C)
* The analysis of the dataset based on the question (All members individually Persons A, B, C, and D)
* What more information they would have liked to have had and Conclusion (Person D)

Assignment Part II Due Jan 13/15/17

You and a partner from within your group will pair up and find a dataset of your choosing pending approval. \*Due prior to WinterBreak

You will create 2 questions from the dataset as well as use 2 distinct visualizations to help illustrate your analysis. I will create a 3rd question after you get the questions approved. The 3rd visualization is at your discretion.

You will present your findings via a Power Point presentation which comprised of:

* Description of the Dataset including given headers. (Person A)
* Any hinderances or problems of the dataset and how they were handled (Person B)
* The 3 questions and why they may be of interest (Person A)
* The analysis of the dataset based on the question (Person B x 2, Person A)
* What more information they would have liked to have had and Conclusion (Person A and B)

<https://bigdata-madesimple.com/70-amazing-and-free-data-sources-for-data-visualization/>

Topics

1. [Google Trends](https://trends.google.com/trends/explore)

**Curated by:** Google  
**Example data set:** ["Cupcake" search results](https://trends.google.com/trends/explore?q=%2Fm%2F03p1r4&date=all)There are a variety of filters to narrow down trends according to location (worldwide or by country), various time ranges, categories, or even specific search types (web vs image vs YouTube search results). You can easily see what topics are popular at the moment and what is currently trending on the Trends homepage. Google also highlights several interesting examples of trends with data visuals on that homepage.

2. [National Climatic Data Center](https://www.ncdc.noaa.gov/data-access/quick-links)

**Curated by:** National Centers for Environmental Information (formerly NOAA)  
**Example data set:** [Local Climatological Data (LCD)](https://www.ncdc.noaa.gov/cdo-web/datatools/lcd)

Here you can find an archive of climate and weather data sets across the US, the largest archive of environmental data in the world. It is a huge resource for all kinds of weather data, including meteorological, oceanic, climate, atmospheric, and geophysical data.

3. [Global Health Observatory data](http://www.who.int/gho/en/)

**Curated by:** World Health Organization (WHO)  
**Example data set:** [Universal access to reproductive health](http://www.who.int/gho/maternal_health/reproductive_health/en/)

The various data sets are organized according to themes, such as mortality, health systems, communicable and non-communicable diseases, medicines and vaccines, health risks, and so on. The WHO’s health statistics are to go-to source for global health information and is also used in the work of the US Centers for Disease Control and Prevention.

4. [Data.gov.sg](https://data.gov.sg/)

**Curated by:** Singaporean government  
**Example data set:** [Singapore Residents By Age Group, Ethnic Group And Gender, End June, Annual (2017)](https://data.gov.sg/dataset/resident-population-by-ethnicity-gender-and-age-group)

Singapore’s government data website is just so visually accessible. The homepage is full of small visualizations telling stories about each data set. Part of data visualization is making sure that not only does it display information in an accurate and relevant format, but also that it’s appealing catch interest. Most of the government data sites are utilitarian and simple, enough to get the data across in an easy to understand way. Singapore, however, brightens it up with colorful visualizations, splashes of color in the graphs, and a “Similar Datasets” section at the bottom of every data set to encourage readers to explore.

5. [Earthdata](https://earthdata.nasa.gov/)

**Curated by:** NASA  
**Example data set:** [Atmospheric Electricity (Lightning)](https://search.earthdata.nasa.gov/search?m=0!0!2!1!0!0%2C2&fst0=Atmosphere&fsm0=Atmospheric%20Electricity)

Via Earthdata, the public can access NASA’s data, news, and event information. It covers data from Earth’s atmosphere, solar radiance, the cryosphere (arctic/frozen areas), the ocean, land surface (gravity, geomagnetism, tectonics), and human environments.

6. [Amazon Web Services Open Data Registry](https://registry.opendata.aws/)

**Curated by:** Amazon  
**Example data set:** [1000 Genomes Project](https://docs.opendata.aws/1000genomes/readme.html)

Amazon has created a registry to find and share those various data sets. There are over 50 public data sets supported through Amazon’s registry, ranging from IRS filings to NASA satellite imagery to DNA sequencing to web crawling. The data sets also include usage examples, showing what other organizations and groups have done with the data.

7. [Pew Internet](http://www.pewinternet.org/datasets/)

**Curated by:** Pew Research Center  
**Example data set:** [Teens, Social Media & Technology 2018](http://www.pewinternet.org/2018/05/31/teens-social-media-technology-2018/)

The Pew Research Center’s mission is to collect and analyze data from all over the world. They cover all sorts of topics like politics, social media, journalism, the economy, online privacy, religion, and demographic trends. Access simply requires a brief registration on the site and credit to Pew Research Center as the source of the data.