Activity 1: Class Activity and Speaking; Locating Public Data

Submit deliverable(s) through your assignment GitHub repository.
Place your work in the `writing/` directory

Objectives

To learn how to use online sources to gather sets of data.To identify the traits of the dataset itself and to envision types of questions that it could be used to answer.

Overview

When working on research projects, It is very likely that some form of data will be necessary to complete the project. The data chosen for a research project must be able to help the researcher to investigate main the research questions of the project and therefore, must be carefully chosen for its scope.

In this work, we will spend some time to see what kinds of datasets are available using public sources. We will then choose a dataset and then study its contents determine some of questions that could be addressed by some type of analysis of the dataset. *We note, that we do not need to perform any form of analysis – we are only studying the dataset for its potential value to a research project.*

On the next class day, we will present the dataset and then discuss two main questions that could be answered by a study (analysis) of the dataset. *Again, we are not going to perform the analysis, only to suggest that one could be made to address two main questions.*

What To Do

This assignment has two components.

1. **Writing Your Ideas**: You are to use the below GitHub link to create your working repository where you will work by editing your report files directly in your browser. In your repository, you will find the file, `writing/report.md` in which you are to respond to questions about your dataset. You will be writing your work using Markdown.

2. **Presenting Your Ideas**: Prepare two or three slides using Google Docs or another software to use for a *lightning talk* in which you introduce your dataset to your colleagues. This talk is to take no more than three or four minutes and will consist of about three or four slides. you will be using your own laptop to make your presentation in class while using Zoom or Google Meet. In this talk, you are to respond to the below basic themes about your dataset.

    - What is the dataset?
    - Where did this dataset come from (reference)?
    - Who put this dataset together?
    - What two questions can it be used to answer?

Links to Data Sources

Below are some suggestions of sources of data. Whether you find another source of data or use one of the below sets, **please be sure to cite the source in your work**.

- `http://www.seanlahman.com/baseball-archive/statistics/`
- `https://Baseball-Almanac.com`
- `https://TheBaseballCube.com`
- `https://sabr.org/how-to/statistical-databases-and-websites`
- `http://m.mlb.com/statcast/leaderboard#exit-velo,r,2019`
- `https://www.kaggle.com/`
- `https://www.baseball-reference.com/`
- Awesome Public Datasets (listing of sites):
    - `https://github.com/awesomedata/awesome-public-datasets`

GitHub Starter Link

Clone Your Assignment Repository

`https://classroom.github.com/a/zRen0tB7`

To use this link, please follow the steps below.

- Click on the link and accept the assignment

- Find the necessary files to edit and then begin your edit (using the pen button) on the right side of the page.
- Be sure to often use the *Commit changes* part of the web page to save your work regularly. Note that before you can commit your changes, you will need to add a simple message to explain what the commit entails. Please add some meaningful text to complete this step.

Required Deliverables

1. A completed version of the file `writing/report.md` containing your responses.
2. A *lightning talk* presentation of three or four slides to introduce your dataset to your colleagues. The elements to address are discussed above. You do not need to submit these slides to the instructor for your presentation; they will only be used to guide your talk.

Please be sure to ask the instructor if you have any questions.