Module 1: Big Data Community Vote

# Goal

Throughout this course we will apply software designed to work with “big data” to actual data sets. So that we get the most out of this semester I want the data we choose to be of interest to you. This activity will help decide what data we use and how we will go about exploring and investigating questions.

# Instructions

1. Read about the proposed data sets below.
2. Look at the source data and metadata to identify any potential challenges to data access or quality issues.
3. Find research using the data set or similar data from other sources.
4. Write a post on Slack #dan606 outlining your findings (cover any issues you find with these and at least one interesting research angle).
5. Respond to the posts others make. Ask questions so that all can understand the data and possible questions.
6. VOTE for your top pick (A poll will be circulated at the end of the week).

# Data Sources

***US Census American Community Survey (ACS)***

Description: The Public Use Microdata (PUM) contains response data from the ACS and is the primary data for many other derived features.

<https://docs.data.world/uscensus/#10---the-american-community-survey>

<http://www.census.gov/acs>

Data:

<https://www2.census.gov/programs-surveys/acs/data/pums/>

***NYC Taxi Ride Data***

Description:

<https://www1.nyc.gov/site/tlc/about/data-and-research.page>

*Data:*

<https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page>

***Bureau of Transportation Statistics***

Description: Airport flight delay tracking

<https://www.transtats.bts.gov/OT_Delay/ot_delaycause1.asp?type=5&pn=1>

Data:

<https://www.transtats.bts.gov/DL_SelectFields.asp>

***AWS COVID Data Lake***

Description: There’s a lot in here. Don’t swim alone.

<https://aws.amazon.com/blogs/big-data/a-public-data-lake-for-analysis-of-covid-19-data/>

Data:

<https://registry.opendata.aws/aws-covid19-lake/>

***US Stock Market Data***

Description:

Data: <https://docs.intrinio.com/api-explorer?method=api_v2#%3Fmethod=download> (\*API key required)

***GapMinder***

Description: Gapminder is an independent educational non-proﬁt ﬁghting global misconceptions with focus on global demographic and economic data.

<https://www.gapminder.org/>

Data: <https://www.gapminder.org/data/>

***Suggest your own***

Description: Do you know a big data source in your field that may be of broader interest? Write a post about it and we will discuss.