**Title**: Relationship between college major and Income

**Description**: Answer the following-

1. Do students in more popular majors make more money
2. How many majors are predominantly male? Female?
3. Which category of majors have the most students?

* Rank - Rank by median earnings (the dataset is ordered by this column).
* Major\_code - Major code.
* Major - Major description.
* Major\_category - Category of major.
* Total - Total number of people with major.
* Sample\_size - Sample size (unweighted) of full-time.
* Men - Male graduates.
* Women - Female graduates.
* ShareWomen - Women as share of total.
* Employed - Number employed.
* Median - Median salary of full-time, year-round workers.
* Low\_wage\_jobs - Number in low-wage service jobs.
* Full\_time - Number employed 35 hours or more.

In this guided project, we work with a dataset on the job outcomes of students who graduated from college between 2010 and 2012. The original data was collected and aggregated by the [American Community Survey](https://www.census.gov/programs-surveys/acs/data/pums.html), though the set we will be using is the cleaned version from [FiveThirtyEight](https://github.com/fivethirtyeight/data/tree/master/college-majors). Each row represents a different college major and contains information on things such as gender diversity, employment rates, median salaries, and more. The columns in the dataset are as follows.

* Rank - Rank by median earnings. Note that the dataset is ordered by this column.
* Major\_code - Major code.
* Major - Major description.
* Major\_category - Category of major, as determined by the report [*What's It Worth?: THe Economic Value of College Majors* by Carnevale et al](http://cew.georgetown.edu/whatsitworth)
* Total - Total number of people with major.
* Sample\_size - Sample size (unweighted) of full-time, year-round workers **only**, who reported earnings of at least $1. This sample is used for computing the earnings columns Median, P25th, and P75th for each major.
* Men - Number of male graduates.
* Women - Number of female graduates.
* ShareWomen - Proportion of women as share of total.
* Employed - Number employed in civilian jobs.
* Full\_time - Number employed who work 35 hours or more per week.
* Part\_time - Number employed who work less than 35 hours per week
* Full\_time\_year\_round - Number employed at least 50 weeks of the year and work at least 35 hours per week.
* Unemployed - Number unemployed (not employed or in the armed forces and actively looking for work)
* Unemployment\_rate - (Number Unemployed) / (Number Unemployed + Number Employed).
* Median - Median earnings of full-time, year-round workers, in USD.
* P25th - 25th percentile of earnings of full-time, year-round workers, in USD.
* P75th - 75th percentile of earnings of full-time, year-round workers, in USD.
* College\_jobs - Number with job requiring a college degree.
* Non\_college\_jobs - Number with job not requiring a college degree.
* Low\_wage\_jobs - Number in low-wage service jobs.

This goal of this project is to use visualizations to explore questions from the dataset. Some of the questions we will explore include:

* Do students in more popular majors make more money? (Use scatterplots.)
* How many majors are predominantly male? Predominantly female? (Use histograms.)
* Which categories of majors have the most students? (Use bar plots.)

While we can use the [pandas interface with matplotlib](https://pandas.pydata.org/pandas-docs/stable/user_guide/visualization.html) for plotting data that is stored in [Series](https://pandas.pydata.org/pandas-docs/stable/reference/series.html#plotting) and [DataFrame](https://pandas.pydata.org/pandas-docs/stable/reference/frame.html" \l "plotting) objects, we will focus instead on using [seaborn](http://seaborn.pydata.org/index.html), which is a high-level library for making statistical graphics in Python that is built upon matplotlib and closely integrated with pandas.