# **Data Location Assignment**

**Introduction**:

This assignment asks you to locate, clean and summarize a dataset that you will use for your final project for the course.

The final project will ask you to examine a dataset using the methods we have covered in the course (descriptive statistics, data visualization, multivariate regression analysis).

**Assignment:**

**Please read these directions carefully.**

Find a dataset that you will analyze for the final project and complete and submit the worksheet (attached).

**Guidelines:**

* Your dataset should come from a **reputable source**.
* You will most likely need to **drop a lot of variables** – you only need 5-10 variables for the final project.
* You may need to **drop observations**. For example, you may decide to analyze one wave from a multi-wave panel dataset.
* Keep the dataset **manageable in size** – the purpose of the project is to demonstrate your ability to use the statistical methods and data visualization skills ﻿﻿﻿﻿﻿covered in the course.
* In short, the dataset you use for the project can be a subset of a larger dataset.

# **Data Location Assignment Worksheet (10 points)**

1. **Name of dataset (1 points):** Merged Census (ACS) & IPEDS
2. **2-3 sentence description (2 points):** The dataset comes from selecting specific variables from the ACS program (years 2015-19) using the Census API. It also includes select Fall admission and enrollment IPEDS variables for 2016-20. The unit of analysis is at the state level.
3. **Number of observations (1 points):** 255 (1 per state (plus DC) per year from 2015-19)
4. **Potential dependent variable (1 points):** State-wide college enrollment rates
5. **Potential independent variable(s) of interest (1 points):** See all variables below in the chart, except for the enrollment rate dependent variable.
6. **Total number of variables in the dataset that you’ll be using (1 points):** 9
7. **Variable definitions (3 points):**

*Complete the table adding more rows as necessary.*

| **Variable** | **Source** | **Year(s)** | **Definition** |
| --- | --- | --- | --- |
| 1. Statewide college average enrollment rate | IPEDS | 2016-20 | Dependent variable.  Derived variable. Calculated the average enrollment rate for all institutions in each state for each fall. |
| 1. State Institutions' Average Admission Rate | IPEDS | 2016-20 | Derived variable. Calculated the average admission rate for all institutions in each state for each fall. |
| 1. Unemployment Rate | ACS | 2015-19 | Percentage of people unemployed who are still in the civilian labor force. |
| 1. Median Household Income | ACS | 2015-19 | Median household income, adjusted in 2019 dollars. |
| 1. State Population | ACS | 2015-19 | Total population, ages 1 year and over. |
| 1. Bachelor’s Degree or Higher | ACS | 2015-19 | Percentage of residents who earned a bachelor’s degree or higher. Population 25 years old and over. |
| 1. HS Grad Rate | ACS | 2015-19 | Percentage of residents who earned a high school diploma or equivalent. Population 25 years old and over. |
| 1. Total number of institutions per state | IPEDS | 2016-20 | Sum of institutions listed per state. |
| 1. Average in-state tuition rate at public institutions | IPEDS | 2016-20 | Derived variable. Calculate the average tuition rate for each year per state. I may include all public and private institutions for consistency. |

* Highlighted = Will likely use to create derived variable(s), which may include…
  + Median household income:cost of living (is this power purchasing parity?)
  + Median household income:state’s average tuition rate.
  + Number of state institutions:state population.

Parameters:

* Institutions (n = 5632) included are any Title IV within the US that offer any credential accessible without already having a postsecondary credential. In other words, institutions with graduate programs but no undergraduate programs are excluded, as one would already have a postsecondary credential to enroll.

As of 10/29, IPEDS has not released Fall 2020 data. So as of this date I cannot include admissions fall 2020 data. For now, I can use ‘16-19.