# Exploring data with Python

Tony Yao-Jen Kuo

# Project Overview



Assignment from Introduction to Data Science in Python

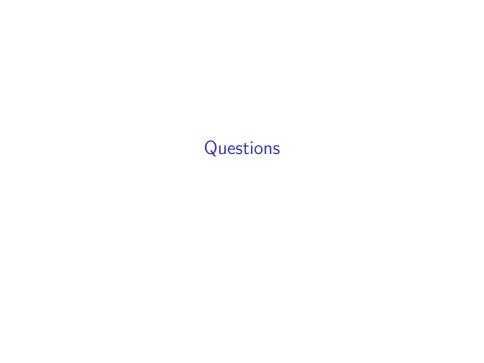
Data

# Downloading data

```
import pandas as pd
us_census = pd.read_csv('https://storage.googleapis.com/py
```

The top 5 rows of us\_census data

us\_census.head()



Which state has the most counties in it? (hint: consider the sumlevel key carefully! You'll need this for future questions too...)

Only looking at the three most populous counties for each state, what are the three most populous states (in order of highest population to lowest population)?

Which county has had the largest change in population within the 6 year period (hint: population values are stored in columns POPESTIMATE2010 through POPESTIMATE2015, you need to consider all 6 columns)?

In this datafile, the United States is broken up into four regions using the "REGION" column.

Create a query that finds the counties that belong to regions 1 or 2, whose name starts with 'Washington', and whose POPESTIMATE 2015 was greater than their POPESTIMATE 2014.

You should return a 5x2 DataFrame with the columns = ['STNAME', 'CTYNAME'] and the same index ID as the census\_df (sorted ascending by index).