*Michael Lengel, Benjamin Gonzalez, Frank Intrieri Jr. - Python Final Project - Backrow Team 11/28/2018*

***About the data that we used for our project:***

*We used a 1994 Census Bureau dataset downloaded from* <https://www.kaggle.com/uciml/adult-census-income>  This dataset contains just over 32,000 rows each containing 15 columns of descriptive variables.

***Scope of project and research questions addressed***

**Scope:**

* Using the census dataset we will attempt to analyze gender and various demographic characteristics and then show how they relate to variables like income level and education levels. We will attempt to identify these variable relations and hopefully uncover any existing trends and correlations.

***Research Questions:***

We set out to perform the following operations with the data:

Intelliscience for autocomplete in jupyter

Mike

1. Is there a relationship between a person’s level of education and their income level?

* Get\_education\_correlation

(returns a dictionary with keys of level\_of\_education & values of % making over 50K)

* Education\_correlation\_graph

(returns graph of info above)

Ben

1. Which gender has a higher income level?

* Get\_gender\_correlation

(returns dictionary of gender & % making over 50K)

* Gender\_correlation\_graph

(returns graph of above)

Frank

1. Does a person’s marital status have any correlation to their income level?

* Get\_marital correlation

(returns dictionary of marital & % making over 50K)

* Get\_marital\_graph

(returns graph)

Frank

1. What is the age range and average age of the population?
   * Get\_avg\_age. Returns avg age
   * Get\_range. Returns range

Mike

1. What is the relationship between the hours worked per week and education level? What is the correlation?

(Convert education level to years of school.)

* + Get hours\_education\_correlation

(returns dictionary with keys of level\_of\_education & values of avg years of education

* + Hours\_education\_graph
    1. Can return either continuous scatter plot of hours vs years of education or bar graph of hour groups & avg. years education

Ben

1. What is the relationship between hours worked per week and income level? What is the correlation?

(Group hours worked per week into increments of 10.)

* + Get\_hours\_income\_correlation

Can returns dictionary with keys of hour\_groups & values of % of ppl making over 50K)

* + Hours\_income\_graph

Return graph of information above

Mike

1. Which occupations work more hours per week?

* Occupation\_vs\_hours (avg. hours per week per occupation)

Ben

8. What is the relationship between a person’s age, hours worked per week and their income level?

* + Age\_vs\_hours (both variables are continuous. This can return a correlation)
  + Age\_vs\_hours\_graph (both variables are continuous. This can return a scatter plot).
  + Age\_vs\_income (group ages by 10 years vs % making over 50K)
  + Age\_vs\_income\_graph

1. Does a person’s occupation predict their income level?

(maybe we can train a binary model using occupation,age,race,sex,education etc to predict above or below 50K)

***Summary of the data***:

Looking at the statistics of the variables, and answering more general statistic questions:

* What is the average age?
* What percentage work in the private sector?
* Do most of the people in the dataset are identify as "White"?
* What is the average work week in hours?
* Do most individuals identify as male or female? What is the gender % breakdown?
* What percentage of people have income levels of >50K and <=50K?