Exercise 3: Income of Hawaii Residents

ECON 256 Data Analysis and Visualization

Objective

Generate some statistics and plots regarding the earnings of Hawaii residents.

1 Set up Your R Workspace

1. Start a new R Script
2. Assign a working directory
3. Initialize tidyverse using `library(tidyverse)`

2 Load a New Data Set

This assignment will use data from the American Community Survey, which is a survey of the American population. I have compiled a data set of 18-65 year old respondents from 2022 in Hawaii. Download the file (acs_HI_2022.csv) from Laulima. Load the data into R using the `read_csv()` command to make an object called `mydata`.

Type `View(mydata)` into the console and hit enter. You should be looking at some demographic data. Each row is an individual person from Hawaii. Take a second and think about what each variable means.

3 Summarize Income in Hawaii

Using the `summary()` function, calculate the median income for a person in Hawaii.

Create a histrogram that shows the income distribution for Hawaii.

Use this as a guide and fill in the blanks:

```
ggplot(______,aes(x=______))+
  geom_histogram()+
  ggtitle("______")+
  xlab("______")+
  ylab("Number of Survey Respondents")
```

4 Income Differences by Education and Age

Use the `filter()` function to create three new data objects. The first object should contain only those respondents who graduated college (COLLEGE==1). The second object should contain only people under 30 (AGE<30). The third object should contain only those who are college graduates AND under 30. Call the objects `grads`, `young` and `younggrads` respectively.

What is the median income of a college grad? a person under 30? a college grad who is under 30? Use the `summary()` function to find the answers. Record your answers in your script as comments (#).

It looks like young people earn less. Let's plot this relationship. Fill in the blanks of the following ggplot() function to show the average income earned across ages:

```
ggplot(______,aes(x=AGE,y=______))+
  geom_smooth()+
  theme_minimal()+
  ggtitle("______")+
  xlab("______")+
  ylab("______")
```

By looking at your graph, at about what age does income peak?

Redo the graph you just made, but use data on only college graduates. Now make the plot for only non-college graduates (you may need to make a new data object that contains only non-graduates).

According to your graph, at age 40, about how much more do college educated people make relative to non-college educated people?

Record your answers in your script as comments (#).

5 Send me Your Code

Save your R script. Name it with your last name, followed by the exercise number.

Submit it on Laulima.