Exercise 2: Plotting Data

ECON 256   Data Analysis and Visualization

Objective

Plot data using the ggplot function

1 Set up Your R Workspace

Recall the setup steps from Exercise 1:

1. Create a folder on your desktop (or use the same one as last time)

2. Start a new R Script

3. Assign a working directory with the `setwd()` function at the top of your code

4. Initialize tidyverse by adding: `library("tidyverse")` to your code

5. We will use the state.csv data again (download from Laulima if you don't have it already)

6. Add states.csv as an object in RStudio by adding to your R Script: `mydata <- read_csv("states.csv")`

Type the following command into the console (not the script) and hit enter:

`View(mydata)`

This will show you a spreadsheet with the data set. Think about what each of the variables means.

2 Plot Something

Okay, lets try to plot a graph.

Do states with high house prices also have high rents? Let's check. Plot the following graph:

`ggplot(mydata,aes(x=median_home_value,y=median_rent))+`

`geom_point()`

Is there a correlation between these two variables? (answer all questions with comments in your code #)

3 Is Hawaii Relatively Expensive for Housing?

So where is Hawaii on this plot? You could look up Hawaii by using `View(mydata)` and then find it on the plot. What if we wanted to represent it on the plot?

Add this command:

`mydata2<-mutate(mydata,hawaii=as.factor(ifelse(state=="Hawaii",1,0)))`

This command adds a new variable to `mydata`, called `hawaii`, and saves a new object called `mydata2`. The `ifelse()` function has three arguments. The first is a logical condition, the second is the value the new variable will take if the condition is true, the third argument is the value the new variable will take if the condition is false. The `as.factor()` function tells R what type of variable you want `hawaii` to be.

Check your data using `View(mydata2)`. `mydata2` should have a new variable that takes a value of 1 if the state is Hawaii and 0 if it is not Hawaii.

Let's make Hawaii a unique color. When we specify x and y variables, we can also specify a "color variable".

Add this command to make a second chart in your code:

```
ggplot(mydata2,aes(x=median_home_value,y=median_rent,color=hawaii))+
  geom_point()
```

According to your graph, is Hawaii an expensive state for housing?

4   Graph a Different Relationship

Now create a third plot that shows the relationship between a state's population in millions and the number of housing units in the state in millions. You will need to create two new variables to convert the raw population and housing counts into millions (use the mutate() function). Then you should have everything you need to create a plot using a `ggplot()` function.

Does a state's population and number of housing units have a close relationship?

5   Send me Your Code

Save your R script. Name it with your last name, followed by the exercise number. eg) Tyndall2.R

Submit it on Laulima under Exercise 2.