Week 6 Programming Assignment

Instructor: Chien-Hao Fu

3/28/2019

Due: 4/03/2019

Source Data

- 縣市重要統計指標　http://statdb.dgbas.gov.tw/pxweb/Dialog/statfile9.asp

Preparation

1. Use the file `counties.xlsx` from last week

2. Download, modify, and run the code from the course website (`W4_Solution.R`)

    - Use `source("file_name")` to import R codes from other files

3. Make sure you have the following data frame in your environment:

    - `mydata`
        - This data frame should contain all the observations imported from `counties.xlsx`. It should also include all the modifications that we have done in Week 4 (changing variable names, transforming string variable to factor variables, etc.)
    - `sub_data2015`
        - This data frame should contain all observations in 2015, excluding those from `Taiwan` and `Total`

4. Load the `ggplot2` package

5. Modify the data frame `mydata`

- Name the new data frame `sub_Taiwan`
- This new data frame should include only observations from the region `Taiwan`
- Hint: Use the `filter` function to create this data frame

Programming Assignment

1. Show the distribution of city's population densities in 2015

- Data set: `sub_data2015`
- Function: `geom_histogram()`
- Variable: `pop_density`

Q: Is there an outlier in the dataset? Check your dataset and see which city the outlier is.

2. Show the distribution for youth unemployment rate. Set the number of bins to 15.

- Dataset: `sub_data2015`
- Function: `geom_histogram()`
- Variable: `youth_un`

Q: Is there an outlier in the dataset? Check your dataset and see which city the outlier is.

3. Show Taiwan's population change over the years

- Dataset: `sub_Taiwan`
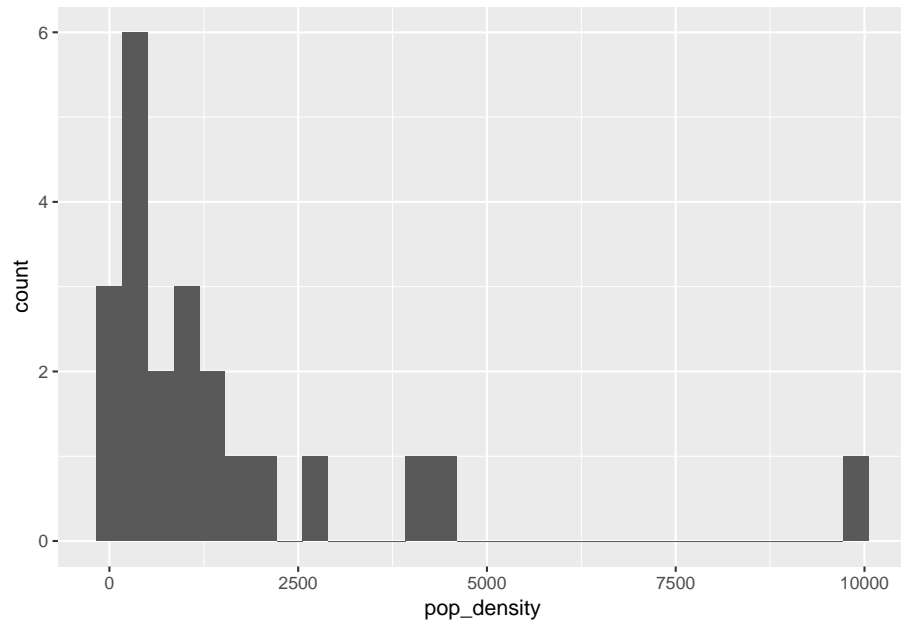- Function `geom_col()`
- Variables: `pop_change`, `year`

- Submit your answers, plots, and source codes using the file: "W6 Student Submission.docx."
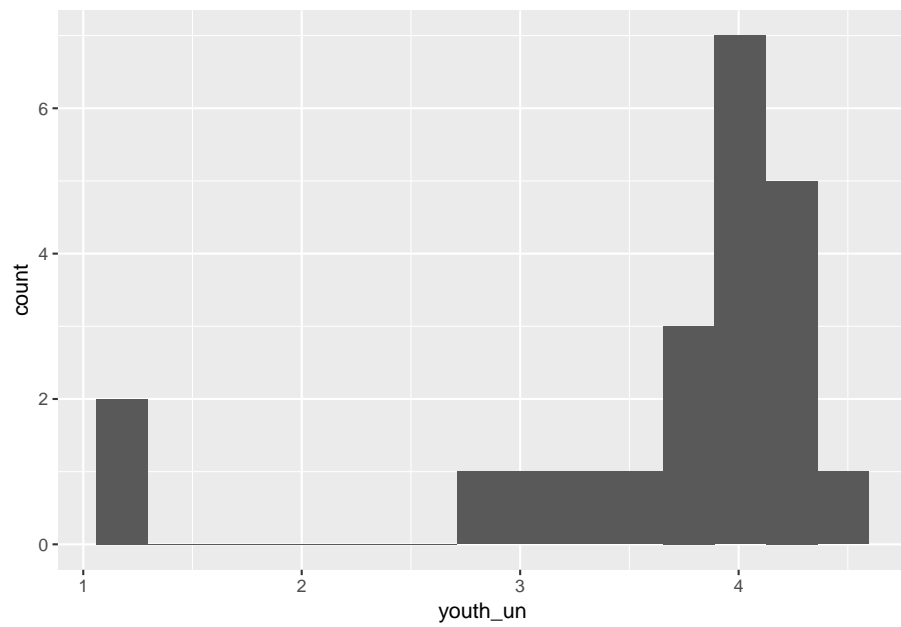
- Bonus Tasks

  - Try to change the color of these charts!

Your plots may look similar to the following:

For Question 1:

For Question 2:

For Question 3: