

## LA-6: Data Visualization (15 points)

### Learning Outcomes

In this assignment, you will:

- Learn how to use `ggplot2`, which is a package included in the `tidyverse` suite of packages, to create visualizations using data.
- Practice using functions from LA-5, including `mutate()`, `freq()`, and `descr()`.

#### Tip

Read all the instructions carefully before starting the assignment.

### Instructions

- 1) Load the packages below and download the `Utilities.csv` data file from Canvas. Read the dataset into R. Use the codebook on Canvas to familiarize yourself with the data in this file.
  - `tidyverse`
  - `summarytools`
- 2) Read and follow along with Chapter 5 of [this online book](#). Use the data provided in the chapter to follow the steps in Ch. 5.

#### Tip

Do not skip this step. It will help you understand `ggplot2`.

- 3) Make a histogram to display the distribution of customers' total monthly bill.
  - a) Use the histogram you created and describe the distribution. Is the distribution skewed or symmetric? If skewed, is the distribution skewed positively or negatively?
  - b) At what cost do most of the total monthly bills fall? Provide a range between which the total monthly bills fall.
- 4) Make a scatterplot to display gas bill by month. Use your graph to answer the following questions.
  - a) Which month do you think has the highest average gas bill?
  - b) Use R to calculate the average gas bill for the months of January and December. Which is higher?
- 5) Make a scatterplot to display electric bill by month. Use your graph to answer the following questions.
  - a) Which month do you think has the highest average electric bill?
  - b) Calculate the average electric bill for the months of September and December. Which is higher?

- 6) Make a scatter plot to display the relationship between gas usage and gas bill. Place gas usage on the x-axis and monthly gas bill on the y-axis. Does there appear to be a relationship between gas usage and gas bill? If so, describe the relationship (e.g., does a relationship exist? Is it linear? Is there a positive or negative relationship between the variables?).
- 7) Make a scatterplot to show the relationship between electricity usage and electric bill. Does there appear to be a relationship between electricity usage and electric bill? If so, describe the relationship.
- 8) Create a new categorical variable called **season**. Have the variable equal **winter** if the bill was from December, January, or February. Have the variable equal **spring** if the bill was from March, April, or May. Have the variable equal **summer** if the bill was from June, July, or August. Have the variable equal **fall** if the bill was from September, October, or November.

Then, examine a frequency table of **season**.

- a) How many bills in the data set are from fall?
  - b) How many bills in the data set are from summer?
- 9) **BONUS:** Make a bar chart using the `geom_col()` to display the count or proportion of donors by **season**.

First, make a new variable, **DonorStatus**. Set **DonorStatus** equal to 1 if the billee donated money to *Operation Fuel* and 0 otherwise. Check that you have created the **DonorStatus** variable correctly.

Then, use `ggplot()` to create your graph. In your own words, describe the graph. Does donation vary by season?

---

## Submission

Submit your R script (named `LA-#_FirstName-LastName.R`) to Canvas.

Your R script should:

- 1) Include commands and functions that are necessary to address all the questions in the assignment.
- 2) Contain comments that answer the questions in the assignment.
- 3) Run in its entirety without errors.

To ensure that your R script runs without errors, you should:

- Save your script.
- Navigate back to Your Workspace on Posit Cloud.
- Reopen your project.
- Run the entire script line-by-line without editing it to ensure there are no errors.

### ! Important

These standards apply to all submissions in this course that require R scripts. You should follow these instructions for preparation, naming, and saving of your R script for all of your individual lab assignments.