

LA-5: Examining and Managing Data (15 points)

Learning Outcomes

In this assignment, you will learn how to:

- Run and examine frequency distributions using the `freq()` function in the `summarytools` package.
- View descriptive statistics using the `descr()` function in the `summarytools` package.
- Recoding variables using the `case_when()` function.
- Create new variables from existing variables in a data frame using the `mutate()` function.

Tip

Read all the instructions carefully before starting the assignment.

Instructions

- 1) Install (`install.packages()`) and load (`library()`) the packages below. Then download `wind-turbines.csv` dataset from Canvas and read it in to R. You will also want to download the codebook for the data from Canvas.
 - `tidyverse`
 - `summarytools`
 - `rstatix`
- 2) The dataset includes data about wind turbines in the US. Use the `freq()` function to examine the frequency distributions of the variables `Site.State` and `Year`. Use the frequency distributions to answer the following questions.
 - a) How many wind turbines are located in Utah?
 - b) How many wind turbines became operational in 2021?
 - c) What percentage of turbines became operational between 1983 and 1999 (include these years in the calculation)?
- 3) Now, subset the data to answer the following question: How many turbines in Utah have been operational since 2010?
- 4) Examine the frequency distribution of `Turbine.Capacity` based on the subset in #3.
 - a) How many turbines in this subset have a capacity of 2,300 kW?
 - b) What percentage of turbines in this subset have a capacity of 1,500 kW?
- 5) Often, we want to look at descriptive statistics of a quantitative variable. We can do this using the `descr()` function that is in the package `summarytools`. Examine the descriptive statistics for `Turbine.Total_Height` and answer the questions below. Include the units of measurement in your answers.
 - a) What is the mean?

- b) What is the median?
 - c) What is the maximum?
- 6) Next, we will learn how to use the function, `mutate()`. We use this function to create new variables (or “mutate” them) from existing ones. Let’s create a new variable called `Turbine_Space` that is the mean of two existing variables, `Turbine.Hub_Height` and `Turbine.Rotor_Diameter`. Then, examine the descriptive statistics of the new variable, `Turbine_Space`, and answer the following questions:
- a) What is the mean of `Turbine_Space`?
 - b) What is the minimum value of `Turbine_Space`?
 - c) What is the maximum value of `Turbine_Space`?
- 7) Sometimes, we may want to make continuous variables into categorical variables. For example, let’s say we are interested in the number of turbines in a project, `Project.Number_Turbines`, but we want to categorize them into low vs. high numbers of turbines in projects.

We refer to this process as **recoding variables**. For this exercise, let’s recode projects with 100 turbines or less into the **low** category and those that have more than 100 turbines into the **high** category.

Use the `mutate()` and `case_when()` functions to recode `Project.Number_Turbines` into a new variable, `Project.Number`, with only two categories, low (`<= 100`) and high (`> 100`). Then, answer the following questions.

- a) How many projects are in the **low** category?
 - b) How many projects are in the **high** category?
- 8) **BONUS:** In the previous step, you created a new variable called `Project.Number`. What is the mean `Project.Capacity` among projects in the (a) low and (b) high `Project.Number` categories?
-

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.