
Homework 1

Data Summarization and Visualization

Purpose

This assignment asks you to inspect several types of data, choose summaries that fit the variables, and explain what the results show. Use R for every calculation and figure.

Data

Use `Summarizing Data Examples.xlsx`. The workbook contains four worksheets. Import each worksheet separately and confirm that the variables have the expected data type before beginning the analysis.

Tasks

Worksheet 1: Categorical Frequencies

The worksheet contains one categorical variable named `Var`.

1. Create a frequency table for X, Y, and Z. Include counts and percentages.
2. Create a labeled bar chart of the counts.
3. In two or three sentences, describe which category is most common and whether the distribution is balanced.

Worksheet 2: Degree Counts

The worksheet contains `Degree` and `Count`.

1. Create a labeled bar chart of count by degree.
2. Create a pie chart of the same data.
3. State which graph communicates the comparison more clearly and explain why.

Worksheet 3: Car Wax Effectiveness

The worksheet contains effectiveness measurements for two car wax brands.

1. Create side-by-side boxplots of effectiveness by brand.
2. Create comparable histograms for the two brands. Use the same axis limits and bin widths.
3. Compare the brands in terms of center, spread, shape, and unusual observations. Do not claim that one brand causes better results.

Worksheet 4: Numerical Summaries

The worksheet contains `x1`, `x2`, and `x3`.

1. Report the minimum, first quartile, median, mean, third quartile, maximum, and standard deviation for each variable in one table.
2. Identify the variable with the greatest spread and support your answer with the summaries.
3. Note any variable for which the mean and median differ noticeably. Explain what that may suggest about the distribution.

Submission Checklist

Submit through WebCampus (Canvas) by the deadline posted there. Submit both of the following files:

- a rendered PDF report; and
- the complete `.Rmd` source file used to produce the report.

The PDF must include your name and assignment title, answer every question in the order presented, show readable and labeled tables and figures, and explain results in complete sentences. The `.Rmd` file must contain all code needed to reproduce the report. Both files must run from the same data supplied for the assignment. Do not submit unexplained software output. You are responsible for checking all results and explaining every choice you made.

A Note About AI

AI can be used as a resource, but it cannot do the assignment for you. I do not want to read AI slop. Do not paste pages of generated writing, code, or interpretations into your report and expect me to sort through it. In the real world, you cannot send AI garbage to a boss, client, or customer. You should not submit it in this class either.

If you use AI, use it to help you learn. Check what it gives you. Make sure the code works. Internalize the material, interpret it, and then explain it in your own words and voice. You should understand every part of your submission and be able to defend the choices you made. Include a short note at the end of the report that names the tool and tells me how you used it. If the work reads like unreviewed AI output, it will receive a 0%, and you will have the opportunity to redo it under the policy in the syllabus.

How Your Work Will Be Evaluated

Credit is based on the statistical reasoning, correct use of R, appropriate checks of assumptions, quality of visual and numerical evidence, and accuracy of the interpretation. A correct number without an explanation is not a complete answer.