LO 1. Classify variables as numeric or categorical. Further distinguish between continuous and discrete numeric variables and ordinal/unordered/free-form categorical variables.

LO 2. Compute (by hand and with R) and interpret measurements of central values of a numeric variable.

- **mean**: If we observe n observations $x_1, \ldots, x_n$ of a variable, the mean is defined as:

$$\bar{x} = \frac{1}{n} \cdot \sum_i x_i.$$

- **median**: If we observe an odd number of observations of a numeric variable the median is the *middle value* that splits the data into two halves. That is, the median M is defined such that half of the data is no bigger than M and half of the data are no smaller than M. For an observation with an even number of points, the median is the average of the two middle values.

LO 3. Compute (by hand and with R) and interpret measurements of variability for a numeric variable.

- **variance**: If we observe n observations $x_1, \ldots, x_n$ of a variable, the variance is the average squared distance from the mean:

$$Var(x) = \frac{1}{n-1} \cdot \sum_i (x_i - \bar{x})^2.$$

- **standard deviation**: The standard deviation is the square-root of the variance.

$$sd(x) = \sqrt{Var(x)} = \sqrt{\frac{1}{n-1} \cdot \sum_i (x_i - \bar{x})^2}.$$

LO 4. Visually identify the **normal distribution** and understand the **central limit theorem (CLT)** through simulation studies. For us, the CTL gives that for large same sizes, the sampling distribution of the mean from an independent sample is approximately normal regardless of the original distribution.

LO 5. Define the **standard error**—the standard deviation of its sampling distribution—and use it to quantify the uncertainty in a test statistic.

LO 6. Understand the meaning of the Pearson correlation coefficient and be able to visually approximate the correlation from a scatter plot of data.

LO 7. Run, using R, two types of correlation tests between two numeric variables and interpret the results.

- **Pearson correlation coefficient** (r): relies on the central limit theorem approximation
- **Spearman's rank correlation coefficient** (ρ): a non-parametric test that attempts to describe a relationship by a monotonic function.
- **Kendall rank correlation coefficient** (τ): a non-parametric test that attempts to measure how similar the *rank* order of two datasets are. Conceptually similar to Spearman's coefficent.

LO 8. Describe the different assumptions that distinguish the Pearson correlation test from the Spearman/Kendall correlation tests.

LO 9. Understand the concept of **statistical power** and how this relates to the choice between parametric and non-parametric inference tests.

LO 10. Describe the definition of a **confidence interval** for a given **confidence level** and relate the concept back to hypothesis testing.

LO 11. Use the grammar of graphics to describe a data visualizations in R. Specifically, produce a scatter plot (with an optional best-fit line) for two variables from a given dataset or a histogram for a single numeric variable.

LO 12. Demonstrate the idea behind Simpson's paradox and explain how it can be alleviated through visualizations and good experimental design.