

# STAT 3011: Discussion 015

Week 2

Talha Hamza

University of Minnesota  
College of Science & Engineering


Spring 2026

# Office Hours & Contact

---

## Office Hours

- Wednesday & Thursdays: 2:00 – 3:00 PM (in-person, Ford Hall 495)

 **Email:** [hamza050@umn.edu](mailto:hamza050@umn.edu)

# Review: Creating and Viewing Objects

---

## Example Code:

```
# Assigning values to objects
x <- 10
y <- 5
z <- c(1, 2, 3) # this is a vector

# Displaying object values
z
```

## Environment Pane:

- Objects you create appear here.
- Useful for tracking your variables.

# RStudio Layout

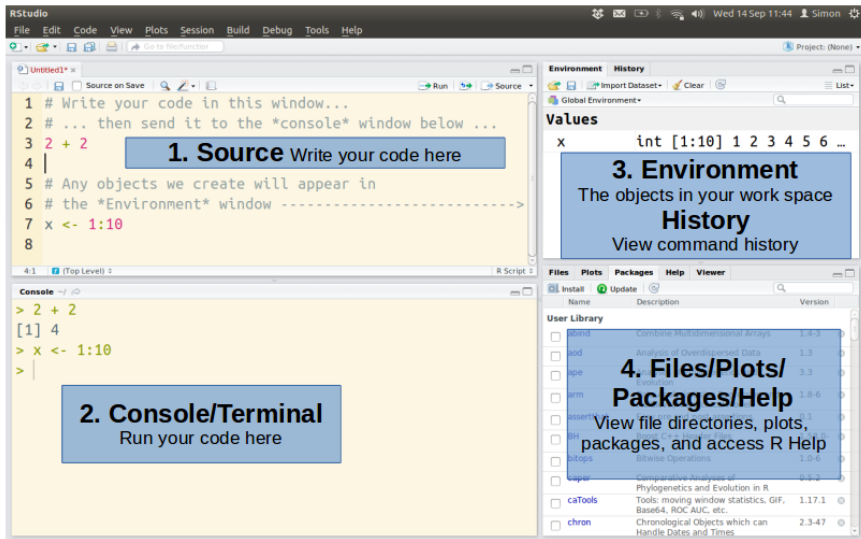


Image credit: <https://www.simonqueenborough.info/R/basic/first-look>

# Preliminary Definitions

---

## Variable Types:

- **Categorical:** Qualitative data representing categories
- **Quantitative:** Numerical data that can be measured or counted
  1. **Discrete:** Countable values (e.g., number of people)
  2. **Continuous:** Measurable values (e.g., weight, time)

**Explanatory Variable:** Variable thought to influence the response.

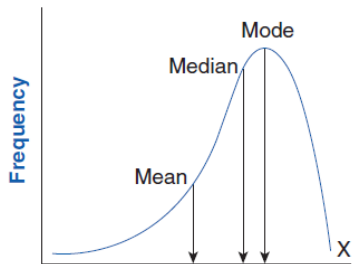
**Response Variable:** Outcome measured in a study.

## Distribution Shapes:

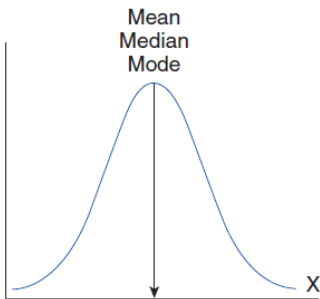
- **Symmetric:** Balanced on both sides (e.g., normal distribution)
- **Skewed:** Asymmetric (left or right skewed)

# Skewness

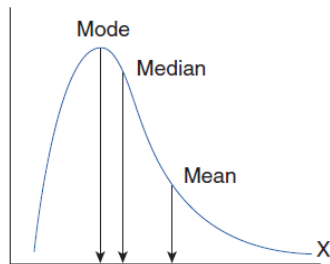
(a) Negatively skewed



(b) Normal (no skew)



(c) Positively skewed



# Research Study Concepts

---

## Study Design:

- **Population:** The entire group of interest.
- **Sample:** Subset of the population actually studied.
- **Subjects:** Individual participants in the study

## Statistical Measures:

- **Parameter:** A number that summarizes a population.
- **Statistic:** A number that summarizes a sample.

## Study Types:

- **Observational:** Researcher observes without intervention
- **Experimental:** Researcher assigns treatments

# Lab 1: R Cheat Sheet

---

## Basic Vector Operations:

- `vec0 <- c(1, 121, 100)` Create vector with specific values
- `vec1 = 3:100` Create sequence vector
- `length(vec1)` Get number of elements
- `mean(vec1)` Calculate mean
- `var(vec1)` Calculate variance

## Vectorized Operations:

- `vec2 <- vec0*2` Multiply all elements by 2
- `vec3 <- vec0 + 10` Add 10 to all elements
- `vec4 <- vec0^2` Square all elements
- `vec23 <- vec2*vec3` Element-wise multiplication



# Lab 1: R Cheat Sheet Conitnued

---

## Data Import & Manipulation:

- `hur0 <- read.csv("path/to/file.csv", header = TRUE)` Import CSV file
- `hur0 <- read.csv(file.choose(), header = TRUE)` Interactive file selection
- `dim(hur0), nrow(hur0), ncol(hur0)` Dataset dimensions
- `names(hur0)` View column names
- `head(hur0), tail(hur0)` View first 6/last 6 rows

## Subsetting Data:

- `hur3 <- hur0[,3:5]` Select columns 3-5
- `hur <- hur3[-c(193:235),]` Remove rows 193-235
- `hur$MaxWind` Access MaxWind column within hur dataframe

## Lab 2: R Cheat Sheet

---

### Visualization:

- `hist(hur$MaxWind)`    Create histogram
- `boxplot(hur$MaxWind)`    Create boxplot
- `boxplot(numeric_variable ~ categorical_variable, data = dataframe)`  
Side-by-side boxplots comparing a numeric variable across categories

Questions?