

# How to get started with R/RStudio

William Christiansen, Ph.D. - Mount St. Mary's University

March 4, 2024

## Contents

<b>1</b>	<b>Getting Started with R and RStudio</b>	<b>2</b>
1.1	Downloading and Installing R . . . . .	2
1.2	Downloading and Installing RStudio . . . . .	2
1.3	Opening RStudio and Setting Up Your Workspace . . . . .	2
1.4	Creating Your First Script . . . . .	2
1.5	Writing and Executing Code . . . . .	2
1.6	Basic R Operations . . . . .	3
1.7	Getting Help . . . . .	3
<b>2</b>	<b>Importing Data</b>	<b>3</b>
<b>3</b>	<b>Handling Missing Data</b>	<b>4</b>
3.1	Standardizing Missing Values . . . . .	4
3.2	Listwise Deletion . . . . .	4
3.3	Mean Imputation . . . . .	4
<b>4</b>	<b>Detecting and Handling Outliers</b>	<b>4</b>
4.1	Using Z-Scores for Outlier Detection . . . . .	4
<b>5</b>	<b>Data Transformation with tidy</b>	<b>5</b>
5.1	Pivot Longer . . . . .	5
5.2	Pivot Wider . . . . .	5
<b>6</b>	<b>Basic Data Exploration</b>	<b>5</b>
6.1	Summary Statistics . . . . .	5
6.2	Visualizing Data with ggplot2 . . . . .	5
<b>7</b>	<b>Conclusion</b>	<b>6</b>

# 1 Getting Started with R and RStudio

Before diving into data cleaning and analysis, it's essential to set up your R environment. R is a free software environment for statistical computing and graphics, while RStudio provides a powerful and user-friendly interface for R.

## 1.1 Downloading and Installing R

1. Visit CRAN (The Comprehensive R Archive Network).
2. Select your operating system (Linux, macOS, or Windows).
3. Follow the instructions to download and install R for your system.

## 1.2 Downloading and Installing RStudio

1. Visit the RStudio Download Page.
2. Download the RStudio Desktop (Free Version) suitable for your operating system.
3. Install RStudio on your computer.

## 1.3 Opening RStudio and Setting Up Your Workspace

Upon launching RStudio, you'll be greeted with three main panels (or four, if you open a script or RMarkdown file). Familiarize yourself with the Console, Environment/History, and Files/Plots/Packages/Help tabs.

## 1.4 Creating Your First Script

1. Go to File  $\hookrightarrow$  New File  $\hookrightarrow$  R Script to open a new script tab.
2. Write or paste your R code into this script.
3. Save your script for future reference (File  $\hookrightarrow$  Save or Save As).

## 1.5 Writing and Executing Code

- To run code, type it into the Console panel or in a script.
- Execute code from the script by highlighting the line(s) and pressing Ctrl + Enter (Cmd + Enter on macOS).
- Use the script to organize, annotate, and save your work.

## 1.6 Basic R Operations

---

```
1 # Assigning values to variables
2 x <- 10
3 y <- 20
4
5 # Basic arithmetic operations
6 sum <- x + y
7 difference <- y - x
8 product <- x * y
9 quotient <- y / x
10
11 # Displaying values
12 print(sum)
13 print(difference)
14 print(product)
15 print(quotient)
```

---

## 1.7 Getting Help

When in doubt, use the ‘help()’ function or ‘?’ operator in R.

---

```
1 ?mean
2 help("mean")
```

---

For more comprehensive resources, visit:

- R Project Help
- Stack Overflow
- RStudio Community

## 2 Importing Data

The foundation of any data analysis project lies in efficiently importing your dataset. R, equipped with the **rio** package, simplifies this process by supporting various file formats seamlessly.

---

```
1 # Install and load the rio package
2 install.packages("rio")
3 library(rio)
4
5 # Import data from a CSV file
6 data <- import("path/to/your/file.csv")
```

---

For more information on data import, visit the **rio** package documentation.

## 3 Handling Missing Data

Properly addressing missing data is crucial to ensure the accuracy of your analysis. R offers versatile strategies to manage missing values effectively.

### 3.1 Standardizing Missing Values

Convert unconventional missing value indicators to `NA` to maintain consistency.

---

```
1 # Standardize missing values
2 data[data == "" | data == "-9" | data == "-99"] <- NA
```

---

### 3.2 Listwise Deletion

Removing observations with any missing values to clean your dataset.

---

```
1 # Apply listwise deletion
2 clean_data <- na.omit(data)
```

---

### 3.3 Mean Imputation

Fill missing numeric values with the mean of their respective columns.

---

```
1 # Implement mean imputation
2 library(dplyr)
3 data <- data %>% mutate(across(where(is.numeric), ~ifelse
  ↪ (is.na(.), mean(., na.rm = TRUE), .)))
```

---

Discover more about handling missing data in the R documentation.

## 4 Detecting and Handling Outliers

Identifying and managing outliers is essential to prevent skewed analyses.

### 4.1 Using Z-Scores for Outlier Detection

Detect outliers by calculating z-scores, replacing extreme values with `NA`.

---

```
1 # Outlier detection with z-scores
2 z_scores <- (data - mean(data, na.rm = TRUE)) / sd(data,
  ↪ na.rm = TRUE)
3 data[abs(z_scores) > 3] <- NA # Adjust threshold as
  ↪ needed
```

---

## 5 Data Transformation with tidyr

Transforming your dataset's structure can unveil new insights and facilitate analysis.

### 5.1 Pivot Longer

Convert wide data into a long format for a more detailed examination.

---

```
1 library(tidyr)
2 data_long <- pivot_longer(data, cols = all_of(column_
  ↳ names), names_to = "name", values_to = "value")
```

---

### 5.2 Pivot Wider

Transition from a detailed long format back to a comprehensive wide format.

---

```
1 data_wide <- pivot_wider(data, names_from = name, values_
  ↳ from = value)
```

---

Learn more about tidyr in the tidyr package documentation.

## 6 Basic Data Exploration

Begin your data analysis journey with summary statistics and visualization to understand your dataset's distribution and main characteristics.

### 6.1 Summary Statistics

Quickly review your data's central tendency, dispersion, and shape.

---

```
1 summary(data)
```

---

### 6.2 Visualizing Data with ggplot2

Create a histogram to explore the distribution of a numeric variable.

---

```
1 library(ggplot2)
2 ggplot(data, aes(x = numeric_variable)) + geom_histogram
  ↳ ()
```

---

For more on ggplot2, check out the ggplot2 documentation.

## 7 Conclusion

I hope this is helpful! Please don't hesitate if you have any questions or suggestions for new applications/guides/resources. For further questions or feedback, please reach out to [w.t.christiansen@msmary.edu](mailto:w.t.christiansen@msmary.edu).