

# **Introduction to R**

## **Experimental Research**

**Professor Yamil Velez (9/29/22)**

# What is R?

- Most popular programming environment in statistics
- Free and open source
- Diverse array of packages
  - Data cleaning
  - Data visualization
  - Analysis

# Alternatives to R

- STATA (proprietary): most popular statistical software in economics. Relatively expensive, programming is not straightforward.
- Matlab (proprietary): popular in the hard sciences. Programming is similar to R.
- Python (open source): general purpose programming language. Using Python for statistics is not as straightforward as R.

# RStudio Cloud

- R can be installed on your machine by visiting <https://cran.rstudio.com/>
- RStudio is a GUI for programming in R
  - <https://www.rstudio.com/products/rstudio/download/>
- In class, we will be using RStudio.cloud.
  - Version of RStudio that does not require installation
  - Sharing scripts and data is easy

# Calculation

```
> # Sample Session
> # This is a comment
>
> 2 # print a number
[1] 2
> 2+3 # perform a simple calculation
[1] 5
> log(2) # natural log
[1] 0.6931472
```

# Storing numeric or string objects

```
> x <- 2 # store an object
> x # print this object
[1] 2
> (x <- 3) # store and print an object
[1] 3
>
> x <- "Hello" # store a string object
> x
[1] "Hello"
```



# Storing and manipulate vectors

```
> Height <- c(168, 177, 177, 177, 178, 172, 165, 171, 178, 170) #store a vector
> Height # print the vector
[1] 168 177 177 177 178 172 165 171 178 170
>
> Height[2] # Print the second component
[1] 177
> Height[2:5] # Print the second, the 3rd, the 4th and 5th component
[1] 177 177 177 178
>
> (obs <- 1:10) # Define a vector as a sequence (1 to 10)
[1] 1 2 3 4 5 6 7 8 9 10
>
> Weight <- c(88, 72, 85, 52, 71, 69, 61, 61, 51, 75)
>
> BMI <- Weight/((Height/100)^2) # Performs a simple calculation using vectors
> BMI
[1] 31.17914 22.98190 27.13141 16.59804 22.40879 23.32342 22.40588 20.86112
[9] 16.09645 25.95156
```

# Summary statistics

```
> length(Height)
[1] 10
> mean(Height) # Compute the sample mean
[1] 173.3
> var(Height)
[1] 22.23333
```



# Packages

- R packages are essentially add-ons that can facilitate different tasks (e.g., cleaning, visualization).
- We'll be working with the “tidyverse,” which is a collection of packages that make programming in R a bit more intuitive
  - Consistent data input/output structure
  - Intuitive function names (e.g., group\_by, summarize)
  - Consistent style (verbs)
  - Piping functions
    - `dataset %>% select(variable)`
- To install the tidyverse: `install.packages(“tidyverse”)`

# Analyzing an experiment

[bit.ly/experiment example](https://bit.ly/experiment_example)