Introduction to R and R Studio

Grinnell College

January 23, 2026

Lab Today

Two parts:

1. Intro to R
   - Elements of R
   - Data frames
   - Data basics
2. R Markdown
   - Knit to PDF
   - Markdown formatting (headers, bold/italics, etc)
   - Code chunks

Why R?

First and foremost, this is *not* an R course

R provides several significant advantages over basic calculators or hand calculations:

- Handle very large quantities of data
- Able to read in data from a variety of different sources and formats
- Easily create sophisticated data summaries and visuals
- Large repositories of pre-built functions to perform statistical tests quickly

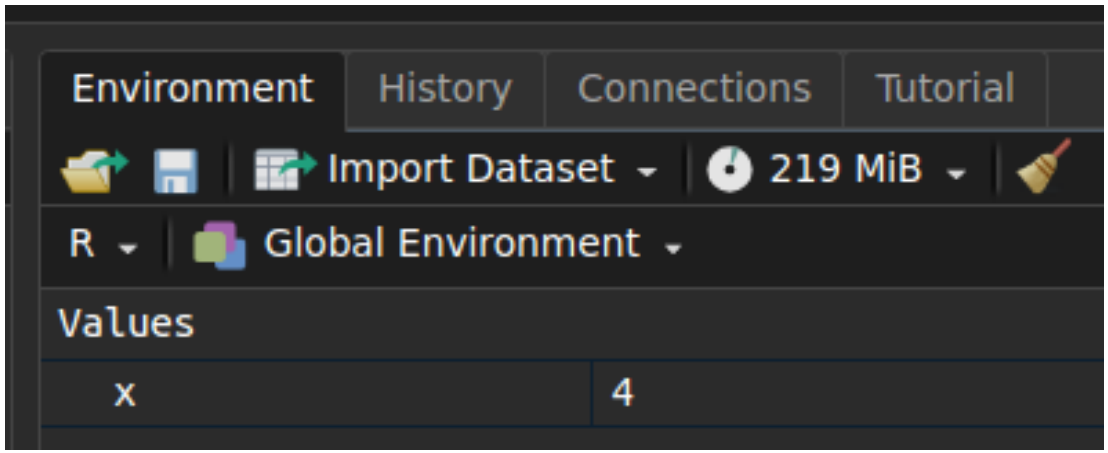
Also widely used across a number of disciplines

Basic Elements of R

Data in R is stored by assigning it to a name using <−. This relationship between a name and a value describes a *variable*

```
> x <- 4
> x
[1] 4
```

We can see all of the names we have assigned in the *environment* tab in the top right of RStudio



How is data stored in R?

Once names have been assigned, we can use just as we would their assigned values

```
> x <- 4
> y <- 3
> sqrt(x^2 + y^2)
[1] 5
```

Basic Elements of R

1. Vectors
   - All of one "type"
   - Can be short or long
2. Data frames
   - Shaped like a rectangular table
   - Rows are observations, columns are variables (vectors)
3. Functions
   - Prewritten pieces of code
   - Useful for performing common tasks
   - Things like `mean()`, `sqrt()` or `plot()`

What is data?

Data represents the raw, unprocessed facts or observations that become meaningful information once collected, organized, and analyzed

Generally speaking for this class, data will have four attributes:

1. **Variables** represent a quantity, quality, or attribute that can be measured
2. **Values** are the state of the variable when observed
3. **Observations** are a set of measurements taken on a single entity, often made up of several variables
4. **Tabular arrangement** describes the way observations and variables are traditionally stored

Data in Practice

We often uses a tabular form to store observations (rows) and variables (columns). This makes it simple to add or remove observations and variables with relative ease

Total Bill	Tip	Sex	Smoker	Day	Time	Size
13.42	1.58	Male	Yes	Fri	Lunch	2
16.27	2.50	Female	Yes	Fri	Lunch	2
10.09	2.00	Female	Yes	Fri	Lunch	2
20.45	3.00	Male	No	Sat	Dinner	4
13.28	2.72	Male	No	Sat	Dinner	2
22.12	2.88	Female	Yes	Sat	Dinner	2
24.01	2.00	Male	Yes	Sat	Dinner	4
15.69	3.00	Male	Yes	Sat	Dinner	3
⋮	⋮	⋮	⋮	⋮	⋮	⋮

Data in Practice

In R, tabular data is typically stored as a `data.frame`

```
> tips
     total_bill  tip    sex smoker  day   time size
  1:      16.99 1.01 Female     No  Sun Dinner    2
  2:      10.34 1.66   Male     No  Sun Dinner    3
  3:      21.01 3.50   Male     No  Sun Dinner    3
  4:      23.68 3.31   Male     No  Sun Dinner    2
  5:      24.59 3.61 Female     No  Sun Dinner    4
 ---
240:      29.03 5.92   Male     No  Sat Dinner    3
241:      27.18 2.00 Female    Yes  Sat Dinner    2
242:      22.67 2.00   Male    Yes  Sat Dinner    2
243:      17.82 1.75   Male     No  Sat Dinner    2
244:      18.78 3.00 Female     No Thur Dinner    2
```

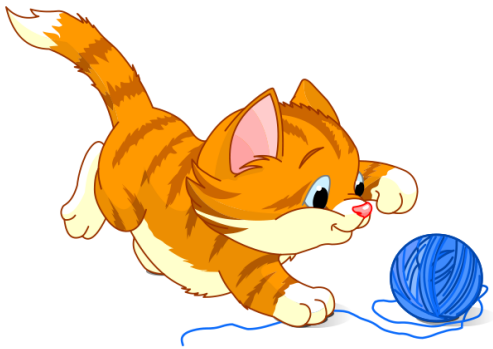
Using R Markdown

R Markdown describes a specific type of file that is used in R (.Rmd)

Uses *markdown* language to easily add headers, or write things in **bold** or *italics*

Alongside written text allows us to write and compute R code

Can be knit into pdf and submitted to gradescope



Go forth and conquer

1. Find lab on course website
2. Do it