Math Camp: Lesson 2.5

Introduction to R

UW–Madison Political Science

August 18, 2020

What is R?

R is...

R is...

- a programming language and software environment for statistical computing and graphics
- open source (free!)
- based on the popular C programming language

R in the wild

R in the wild

- R graphics are often featured in research in the top political science journals
- If you know R, you will be able to create cool graphics too!
- Check out Kastellec and Leoni's 2007 *Perspectives on Politics* article, Using Graphs Instead of Tables in Political Science
-

Getting Started

Getting Started

- Download R (engine): pick the CRAN mirror closest to you!
- Download RStudio (car): download the free version!
- Open RStudio
- Create New Folder called `sample-folder`
- Go to File -> New File -> R Script
- File -> Save As -> save in `sample-folder`

Tips

Tips

- Google is your friend!
- That's it

Set Working Directory

- Copy folder path

Set Working Directory

- Copy folder path

- In MacOS

```
setwd("/Users/marcyshieh/Desktop/sample-folder")
```

Set Working Directory

- Copy folder path

- In MacOS

```
setwd("/Users/marcyshieh/Desktop/sample-folder")
```

- In Windows

```
setwd("C:/Users/marcyshieh/Desktop/sample-folder")
```

Use R as a Calculator

Use R as a Calculator

- `1+3`
- `5-3`
- `2*2`
- `9/9`
- `2^2`
- `sqrt(4)`

Creating Objects in R

What does it mean to create an object in R?

Creating Objects in R

What does it mean to create an object in R?

You are basically putting stuff in a box, then labeling the box.

Creating Objects in R

What does it mean to create an object in R?

You are basically putting stuff in a box, then labeling the box.

- You want to put 1 + 3 into an object called "four"

```
four <- 1+3
```

Creating Objects in R

What does it mean to create an object in R?

You are basically putting stuff in a box, then labeling the box.

- You want to put 1 + 3 into an object called "four"

```
four <- 1+3
```

- Now when you want to perform a function on "four," you will be referring to 1 + 3

```
sqrt(four)
```

Creating Objects in R

What does it mean to create an object in R?

You are basically putting stuff in a box, then labeling the box.

- You want to put 1 + 3 into an object called "four"

```
four <- 1+3
```

- Now when you want to perform a function on "four," you will be referring to 1 + 3

```
sqrt(four)
```

- You want to put the word "hi" into an object called "hello"

```
hello <- "hi"
```

Creating Objects in R

What does it mean to create an object in R?

You are basically putting stuff in a box, then labeling the box.

- You want to put 1 + 3 into an object called "four"

```
four <- 1+3
```

- Now when you want to perform a function on "four," you will be referring to 1 + 3

```
sqrt(four)
```

- You want to put the word "hi" into an object called "hello"

```
hello <- "hi"
```

- You can overwrite objects

- Here, you are taking out the word "hi" from the "hello" object and replacing it with "hi, nice to meet you"

Creating Objects in R

What does it mean to create an object in R?

You are basically putting stuff in a box, then labeling the box.

- You want to put 1 + 3 into an object called "four"

```
four <- 1+3
```

- Now when you want to perform a function on "four," you will be referring to 1 + 3

```
sqrt(four)
```

- You want to put the word "hi" into an object called "hello"

```
hello <- "hi"
```

- You can overwrite objects

- Here, you are taking out the word "hi" from the "hello" object and replacing it with "hi, nice to meet you"

```
hello <- "hi, nice to meet you"
```

Create Data Frame

Create Data Frame

A data frame is essentially a table.

Create Data Frame

A data frame is essentially a table.

- Create a column called "streaming"

```
streaming <- c("Netflix", "Hulu", "Amazon Prime")
```

Create Data Frame

A data frame is essentially a table.

- Create a column called "streaming"

```
streaming <- c("Netflix", "Hulu", "Amazon Prime")
```

- Create a column called "approve"

```
approve <- c(50, 30, 15)
```

Create Data Frame

A data frame is essentially a table.

- Create a column called "streaming"

```
streaming <- c("Netflix", "Hulu", "Amazon Prime")
```

- Create a column called "approve"

```
approve <- c(50, 30, 15)
```

- Create a columnn named "disapprove"

```
disapprove <- c(50, 70, 85)
```

Create Data Frame

A data frame is essentially a table.

- Create a column called "streaming"

```
streaming <- c("Netflix", "Hulu", "Amazon Prime")
```

- Create a column called "approve"

```
approve <- c(50, 30, 15)
```

- Create a columnn named "disapprove"

```
disapprove <- c(50, 70, 85)
```

- Put all the columns together to create one table

```
streaming_dataframe <- data.frame(streaming, approve, disapprove)
```

Create Data Frame

A data frame is essentially a table.

- Create a column called "streaming"

```
streaming <- c("Netflix", "Hulu", "Amazon Prime")
```

- Create a column called "approve"

```
approve <- c(50, 30, 15)
```

- Create a columnn named "disapprove"

```
disapprove <- c(50, 70, 85)
```

- Put all the columns together to create one table

```
streaming_dataframe <- data.frame(streaming, approve, disapprove)
```

- Then you can take a look at the table that you created!

```
streaming_dataframe
```

Summarize the Data Frame

Summarize the Data Frame

- Calculate the mean for the "approve" column

```
mean(streaming_dataframe$approve)
```

Summarize the Data Frame

- Calculate the mean for the "approve" column

```
mean(streaming_dataframe$approve)
```

- Calculate the mean for the "disapprove" column

```
mean(streaming_dataframe$disapprove)
```

Summarize the Data Frame

- Calculate the mean for the "approve" column

```
mean(streaming_dataframe$approve)
```

- Calculate the mean for the "disapprove" column

```
mean(streaming_dataframe$disapprove)
```

- Subtract the means in the approve and disapprove columns

```
mean(streaming_dataframe$approve) - mean(streaming_dataframe$disapprove)
```

Summarize the Data Frame (cont)

Summarize the Data Frame (cont)

- Find the absolute value from subtracting the approve and disapprove columns

```
abs(mean(streaming_dataframe$approve) - mean(streaming_dataframe$disapprove))
```

Summarize the Data Frame (cont)

- Find the absolute value from subtracting the approve and disapprove columns

```
abs(mean(streaming_dataframe$approve) - mean(streaming_dataframe$disapprove))
```

- Find the median of the approve column

```
median(streaming_dataframe$approve)
```

Summarize the Data Frame (cont)

- Find the absolute value from subtracting the approve and disapprove columns

```
abs(mean(streaming_dataframe$approve) - mean(streaming_dataframe$disapprove))
```

- Find the median of the approve column

```
median(streaming_dataframe$approve)
```

- Look at the summary statistics for the entire table

```
summary(streaming_dataframe)
```

Count rows and columns

- Count the number of rows

```
nrow(streaming_dataframe)
```

Count rows and columns

- Count the number of rows

```
nrow(streaming_dataframe)
```

- Count the number of variables

```
ncol(streaming_dataframe)
```

Count rows and columns

- Count the number of rows

```
nrow(streaming_dataframe)
```

- Count the number of variables

```
ncol(streaming_dataframe)
```

- Count rows and columns

```
dim(streaming_dataframe)
```

Any questions?

Write & Read CSV Files

Write & Read CSV Files

- Write csv file into folder

```
write.csv(streaming_dataframe, "streaming_dataframe.csv")
```

Write & Read CSV Files

- Write csv file into folder

```
write.csv(streaming_dataframe, "streaming_dataframe.csv")
```

- Open up your streaming_dataframe.csv file and rename the "approve" cell into "approve2"

```
streaming_dataframe_v2 <- read.csv("streaming_dataframe.csv")
```

Write & Read CSV Files

- Write csv file into folder

```
write.csv(streaming_dataframe, "streaming_dataframe.csv")
```

- Open up your streaming_dataframe.csv file and rename the "approve" cell into "approve2"

```
streaming_dataframe_v2 <- read.csv("streaming_dataframe.csv")
```

- Find the mean for each column one by one

```
mean(streaming_dataframe_v2$approve2)   mean(streaming_dataframe_v2$disapprove)
```

Write & Read CSV Files

- Write csv file into folder

```
write.csv(streaming_dataframe, "streaming_dataframe.csv")
```

- Open up your streaming_dataframe.csv file and rename the "approve" cell into "approve2"

```
streaming_dataframe_v2 <- read.csv("streaming_dataframe.csv")
```

- Find the mean for each column one by one

```
mean(streaming_dataframe_v2$approve2)   mean(streaming_dataframe_v2$disapprove)
```

- Or find "summary" statistics info for all the columns

```
summary(streaming_dataframe_v2)
```

Graphs

- Create a graph for approval

```
barplot(streaming_dataframe_v2$approve, xlab="Streaming", ylab="Approval
Rating", names.arg=c("Netflix", "Hulu", "Amazon Prime"))
```

- Create a graph for disapproval

```
barplot(streaming_dataframe_v2$disapprove, xlab="Streaming",
ylab="Disapproval Rating", names.arg=c("Netflix", "Hulu", "Amazon Prime"))
```

Install Packages

- Install the package once

```r
install.packages("ggplot2")
```

- Access the package in your current environment

```r
library("ggplot2")
```

Create graph in ggplot2

- Create graph

```
streaming_graph<-ggplot(data=streaming_dataframe_v2, aes(x=streaming,
y=approve)) + geom_bar(stat="identity")
```

- Load graph

```
streaming_graph
```

- Add color to the graph

```
streaming_graph_blue<-ggplot(data=streaming_dataframe_v2, aes(x=streaming,
y=approve)) + geom_bar(stat="identity", fill="steelblue")
```

- Load graph with color

```
streaming_graph_blue
```

Knit a practice R Markdown file!

Knit a practice R Markdown file!

1. Download LaTeX.
2. Go to RStudio.
3. Go to File > R Markdown
4. The "New to R Markdown" window will come up.
5. Enter a title for your file (e.g., Practice R Markdown Document).
6. Select PDF as the output format.
7. Save As "practice.Rmd".
8. Click "Knit".
9. Compare the PDF to the Rmd file.
10. Check out the R Markdown Cheatsheet.

Final Takeaways

Final Takeaways

- Coding is all about building confidence
- Don't be discouraged
- Don't be afraid to ask for help
- Rely on Google! Use it like it's a friend!
- swirl is a good resource for learning R

Let's call it a day

Homework is online

https://github.com/shirikov/math-camp-2020