Notes - Data Analysis with R Programming

Wonnetz Phanthavong

email: wonnetz2@gmail.com.com
linkedin: https://www.linkedin.com/in/wonnetz/

Abstract

- Understanding the Basics of R
- Programming using RStudio
- Working with data in R
- Visualizations, aesthetics, and annotations
- Documentation and reports

Week 1: Programming and Data Analytics

Why should you use RStudio?

- Free and open source.
- It makes it convenient to view and interact with environmental objects.
- Graphics are more accessible for casual users.

What was covered?

- This week was basically an introduction to Programming Languages and RStudio.

Week 2: Programming using RStudio

How do you assign variables?

- Variables are assigned with '<-'
- These can be strings, floats, booleans, etc

What are vectors?

- They're basically arrays.
- Can be created using 'c()'.
- Can only store elements of the same type.

What are pipes?

- Pipes are similar to an assembly line.
- The output of a pipe is an input to another segment of code.
- Pipes are represented by '%>%'
- You can think of pipes as an 'and then' statement

What are lists?

- Lists are similar to vectors, but they can store different types of elements.
  - These elements can be dates, numbers, boolean values, matrices, etc.
- Lists can be made using the 'list()' function.

What are dates?

- As you would think they are. You can work with dates in R using the lubridate package.

Logical Operators

- AND '&' or '&&'
- OR '|' or '||'
- NOT '!'

Week 3: Data in R

What are Tibbles?

- Tibbles are a streamlined data frames that make working with data easier.
  - They never change data types of the inputs.
  - They never change the name of your variables or create new row names.
  - They make printing in R easier.

The Bias Function

- Numerically calculates the bias between two variables.

Week 4: Visualizations in R

Other Visualization Packages in R

- ggplot2
- Plotly
- Lattice
- RGL
- Dygraphs
- Leaflet
- Highcharter
- Patchwork
- gganimate
- ggridges

Benefits of ggplot2

- Create different types of plots
- Customize the look and feel of plots
- Create high quality visuals
- Combine data manipulation and visualization

Cheat Sheet for ggplot2

Geom Functions

- geom_point()
- geom_bar()
- geom_line()
- geom_smooth()
- geom_jitter()

Faccet Functions

- facet_wrap()
  - creates separate plots based on one variable
- facet_grid()
  - creates separate plots based on multiple variables

Week 5: Documentation and Reports

This section covered...

- What is R Markdown?
- How to use R Markdown in RStudio to create .rmd files
- Structure of these files and how to format them to make reports
- Code chunks and how to include them in your documents
- How to take all of your analyses and transform it into a report