You  eady?

Joe  oith, PhD

Slides available at https://joeroith.github.io/RUserGroup/

First Meeting of St.Kate's R Users Group

- Thank you for taking time to come
- The response has been fantastic from a diverse range of R users
- Introductions in a few minutes

Why use R? Some advantages to coding

- 1. Code is Text
  - Copy & Paste
  - Code is read-able
    - It has its own language and syntax
- 2. Code is open
  - Forever free
  - Forever customizable
    - Endless packages and graphical options
  - It is shareable
    - Better suited for collaboration

https://joeroith.github.io/RUserGroup/

Your Role

- Who do you want to be?
  - The Tourist
    - Survive and get by with minimal knowledge of language
    - Miss out on a lot of interesting things
  - The Fluent Speaker
    - Understand how to read/write/speak
    - Hold your own in conversation
  - The Native Speaker
    - Knowledge of sentence structures, recognize dialects
    - Takes times to reach
- I am a translator/guide

https://joeroith.github.io/RUserGroup/

Icebreaker Activity

- First open this Google Sheets link
- Catch the globe and state your name/department
- Note where your right index finger is and record in the document:
  - Land (Yes/No)
  - Name of Continent or Ocean
  - Name of Country
  - What year are you starting at St. Kate's?

https://joeroith.github.io/RUserGroup/

Get Started

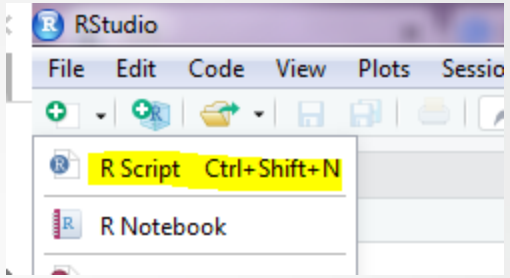
- Installation – R
  - https://www.r-project.org/
- Installation – RStudio
  - I use RStudio as an environment for R. This means in order to use RStudio, R must first (and always) be installed.
  - https://www.rstudio.com/

RStudio Layout

- Console
  - Where code gets compiled and tasks are executed
  - As long as you use the correct language, R will understand
- Environment
  - Object-oriented Programming
  - Manage objects (data sets, variables, functions, etc.)
  - Always consider saving the environment if you want to keep objects from a session
- Files/Plots/Packages/Help/Viewer
  - View graphical output
  - Browse help documentation
  - Install/load customized packages

https://joeroith.github.io/RUserGroup/

Bonus tip!



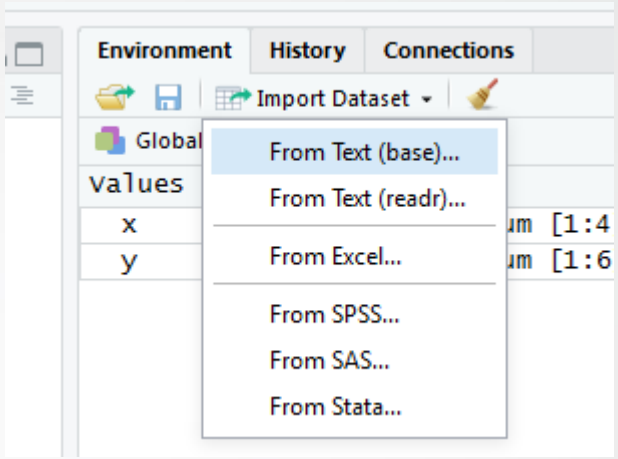
- Console can only handle one line of command at a time.
- R programmers like to write multiple lines of code to edit, check for bugs, or be more efficient.
- Use the green plus symbol in the upper left hand corner to open an R Script. This allows you to write code and run it when you want.
- A new window opens to type in. To run code through the console, highlight as much as you want and click the "Run" button at the top of the script.
- The biggest advantage of a script is the ability to save your work.

https://joeroith.github.io/RUserGroup/

Basic Idea

- R is an object oriented programming language
  - Save numbers, variables, data sets as named objects to refer to and use later
  - Use the *assign* notation "<-" to name and save objects
    - Ex. Type: `x <- 5` in the console and hit enter
- Objects can be used by simply referring to their name
  - See some of the tutorials at the end of the slides to dive deeper into using objects

Importing Data



- Easiest way…
  - Put data in an Excel File (or CSV file)
  - Environment window > Import Dataset > Choose file type
  - Typically any data collection/management tool (Qualtrics, REDCap, etc.) will let you export data as a .csv file
- Big Data
  - 10,000+ rows you may want to consider importing a different way (I can help)
  - I'd still try it the easy way first, no harm will be done

https://joeroith.github.io/RUserGroup/

Using Data

- There are basic statistical functions loaded into R to use with data
  - Need to know the proper syntax to use it
  - Luckily, it is usually intuitive
  - Cheat sheets are available on the R User Group website
- Functions have the following format:

  function_name(arguments)
- Let's try some!

https://joeroith.github.io/RUserGroup/

Globe Example

- Export the shared Google Sheet with our results to your computer. (Make sure to note the file location so you can find it)
- Try to import the data into RStudio
  - Notice the object name in the environment is the name of the downloaded file. This is your dataset name.
- In order to use the variables in `Globe_Activity`, we need to attach the data
  - `attach(Globe_Activity)`

- We can now use `Land` or `Years` for any function that will take the appropriate categorical or quantitative variable.
  - Ex. `table(Land)`
  - Ex. `mean(Years)`
- Sometimes a function can have another function inside it.
  - Ex. `barplot(table(Continent_Ocean_Name))`
  - In the previous example, to make a bar chart we first need the category counts from the table. Everything inside of parentheses happens first

- More complex functions have multiple arguments and parameters.
  - Check the help documentation to see what is required for a function
  - Use '?' before the function name
  - Ex. `?prop.test`
- Let's test whether two-thirds of the planet is covered with water.
  - First make a new object for the number of land and water observations.

    ```
    land_counts <- table(Land)
    ```
  - Now use `land_counts` as an argument in the proportion test function, along with our null hypothesis value of p = 0.67

    ```
    prop.test(land_counts, p = 0.67)
    ```

Motivation to Learn more…

- Tourist

```
boxplot(Years ~ Land)
```

- Fluent

```
boxplot(Years ~ Land, main = "Side by Side Boxplots", xlab =
"Land", ylab = "Years at St. Kates", col=c(“blue", “red"))
```

- Native

```
ggplot(Globe_Activity, aes(Land, Years)) +
    geom_violin(scale = "area", aes(fill = Land)) +
    ggtitle("Violin Plot of Land vs. Years at St. Kate's")
```

https://joeroith.github.io/RUserGroup/

A Glimpse

- A short presentation of examples won't make you an expert
  - I hope it helps by making you a little more familiar and a little more curious to dive into the tutorials
- You will have errors in your code
  - I have errors in my code (rarely do I write it correctly the first time)
  - Don't be afraid to search the web for examples of what you are trying to do.

https://joeroith.github.io/RUserGroup/

Helpful resources

- https://stackoverflow.com/
- Swirl, http://swirlstats.com/students.html
- DataCamp, https://www.datacamp.com
- *R for data science*, Wickham & Grolemund
  - Free here: http://r4ds.had.co.nz/
- Youtube, Google

iNZight

- Consider this software as a point and click option to use with R

  https://www.stat.auckland.ac.nz/~wild/iNZight/

- Warning: this program is relatively new, I haven't had much time to test it out, but my first impression is that it is a nice way to do some exploratory data analysis for basic datasets.

R Tutoring

- We will have a student hold weekly office hours (TBD) in the Statistical Collaborative Center, Library 108.
- Drop in for help troubleshooting R
- Also considering "office calls" for those who can't make the hours held in the Library
- I am available to answer questions and help.

https://joeroith.github.io/RUserGroup/

Next Meeting

- Do you want a next meeting?
- When should it be?

https://joeroith.github.io/RUserGroup/

Questions and Help

- I am happy to help troubleshoot and answer questions individually right now.
- Anyone who has experience in R is welcome to help their neighbor out as well.

https://joeroith.github.io/RUserGroup/

Thank you for coming!!!

- Spread the word to your colleagues!
- I will be in touch shortly with follow-up information

https://joeroith.github.io/RUserGroup/