Data Carpentry Workshop Group Notes

Workshop webpage: https://bigelowlab.github.io/2024-07-08-wph-online/

July 8 Attendance

Name, favorite food

1. Julia Brown, strawberry ice cream
2. Nick Record, curry
3. Ben Tupper, cheese and crackers
4. Ralph Cammack, 'za
5. Frendli Napo, Cheesecake
6. Johnathan Evanilla, fish
7. Aramish Barker, Biryani
8. Bentley Davis, ice cream
9. Nikki Iozzia, pasta
10.
11.
12. Alison Green-Parsons, lobster
13. Michaela Pinette
14. Nameera Fatima, Sugar candy
15. Nicolle Hincks, Peanut Butter Fudge
16. Rachel Sipler, peaches
17. Charlene Huerth, lobster
18. Glendon Zernicke, pizza
19. Aislyn Keyes, pizza
20.
21.
22.

"Matt Parker Comedy" – spreadsheet humor

Getting data ready before going into R:

- Keep an untouched copy of your original data sheet before playing with it in a spreadsheet program
- Have a text file to explain what was done if making changes to raw data file.
- Data formatting: Columns are variables and rows are observations, Avoid having multiple pieces of information in the same column (Try limiting to one piece), Avoid highlighting questionable data (adding a quality control column may be helpful), Avoid having multiple tabs of informing (can another column be added or do we need to start another workbook)?

FYI, you can follow along with this lesson at the lesson link (also on the website): https://datacarpentry.org/spreadsheet-ecology-lesson/

"Datasheets are best when designed with intentionality." or something like that, Julia Brown 2024-07-08

If entering data into a spreadsheet, use column validation to help minimize keystroke and/or data entry errors.

Spreadsheet resources from Nick:

Hadley Wickham, *Tidy Data*, Vol. 59, Issue 10, Sep 2014, Journal of Statistical Software. http://www.jstatsoft.org/v59/i10.

Karl W. Broman & Kara H. Woo (2018) Data Organization in Spreadsheets, The American Statistician, 72:1, 2-10, DOI: 10.1080/00031305.2017.1375989.

wd = “working directory”

getwd() = prints path to working directory
    (also found next to R version at the top of console pane)

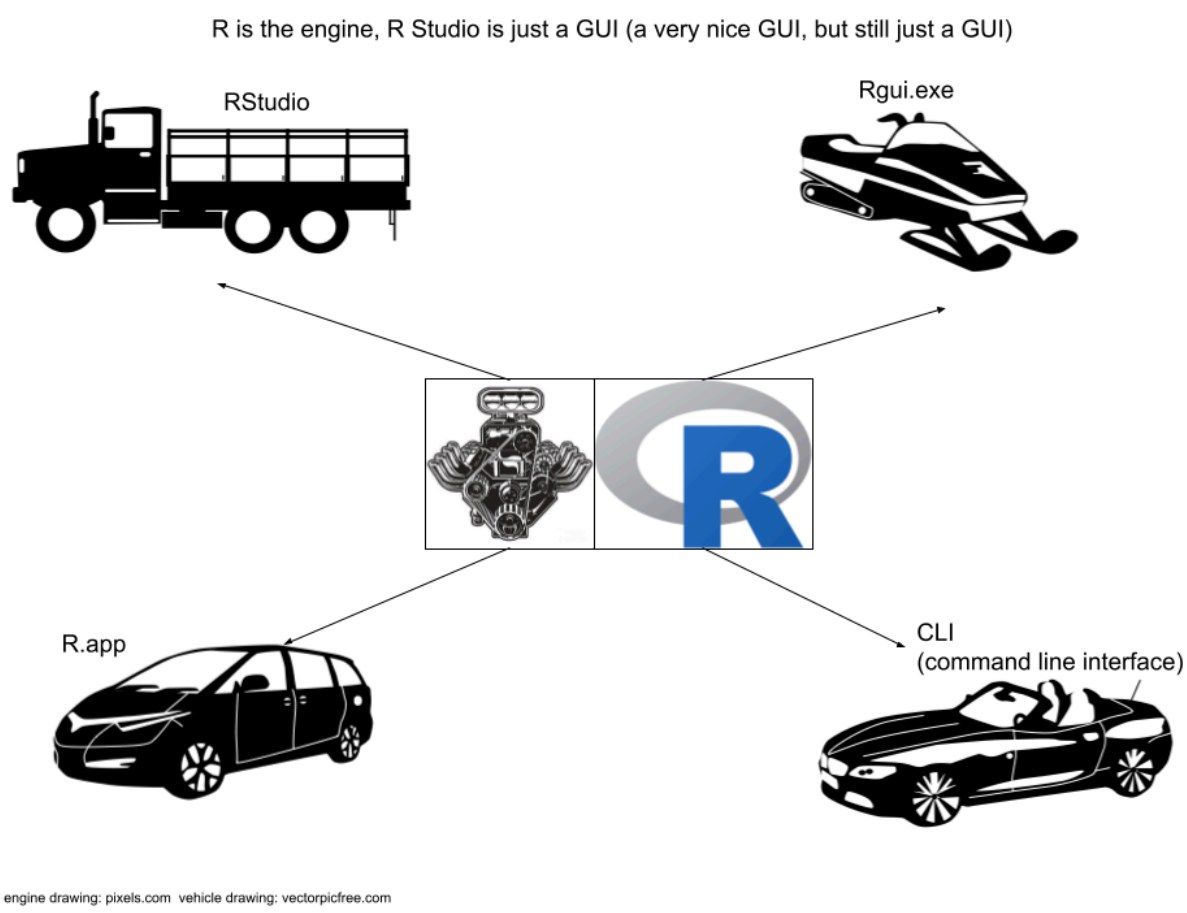
Helpful settings in RStudio:

Tools -> Global Options -> (under) Workspace
    Uncheck “restore .Rdata into workspace at startup”
    Set “Save workspace to .Rdata on exit” to “Never”
Click “Apply” then “OK”

Use library(“name”) to load an R package by putting its “name” in quotes

Save your R code in a script by saving it in a file with the “.R” extension
    In RStudio click: File -> New File -> R Script (don’t forget to save it!)

Steps to start a new R project in RStudio
    Click the menu in the upper right corner (“Project - (None)”)
    Choose R project -> New project
    Provide a project name and select a directory for the project to live in



THis is my mental model for the relationship between R (the computing engine) and the way we interact with it. I think of R an an engine that can be dropped into different models of cars/trucks/etc.

R is R

R.app is the simple wrapper of R on Apple computers.

Rgui.exe is the simple wrapper of R on WIndows computers

CLI is the unix/linux boring interface to R where it's all just typing (nothing pretty).

RStudio is a swiss-army-knife of wrappers that works on all computers. It has all of the whizz-bang doo-dads to make the experience easy and efficient. But RStudio is not R itself.

End of day survey link:
https://forms.gle/NBPLS5w4KWWbGqVYA

Day 2

Attendance

1. Frendli Napo
2. Nick Record
3. Nikki Iozzia
4. Aramish Barker
5. Bentley Davis
6. Nicolle Hincks
7. Charlene Huerth
8. Johnathan Evanilla
9. Michaela Pinette
10. Julia Brown
11. Alison Green-Parsons
12. Lisa Hogan
13. Glendon Zernicke
14.
15.
16.
17.
18.
19.
20.

Today's tutorial: https://datacarpentry.org/R-ecology-lesson-alternative/visualizing-ggplot.html

Colorblind Friendly Plotting

- https://r-graphics.org/recipe-colors-palette-discrete-colorblind
- https://medialab.github.io/iwanthue/
- https://cran.r-project.org/web/packages/viridis/vignettes/intro-to-viridis.html
- https://colorbrewer2.org/#type=diverging&scheme=PuOr&n=9

R help search bar

- https://rseek.org/

ggplot "cheat sheet"

- https://rstudio.github.io/cheatsheets/html/data-visualization.html
- https://rstudio.github.io/cheatsheets/data-visualization.pdf

Challenge on plot modifications

Modify the previous plot by adding a descriptive subtitle. Increase the font size of the plot title and make it bold.

Hint: “bold” is referred to as a font “face”

Contacts

Nick Record nrecord@bigelow.org
Ben Tupper btupper@bigelow.org
Johnathan Evanilla jevanilla@bigelow.org
Julia Brown julia@bigelow.org

End of day survey: https://forms.gle/4ShthUmz42KfvATk7

Day 3 "Exploring and understanding data"

Attendance
1. Nick Record
2.Bentley Davis
3. Michaela Pinette
4. Nicolle Hincks
5. Charlene Huerth
6.Aramish Barker
7.Alison Green-Parsons
8. Frendli Napo
9. Glendon Zernicke
10. Nikki Iozzia
11.Lisa Hogan
12
13.

Notes

Changing RStudio theme:
https://docs.posit.co/ide/user/ide/guide/ui/appearance.html

Function - A set of instructions (method, subroutine, etc.) to execute based on inputs and returns a specific type of output. The output will change depending on the input. Can be used more than one time.

Arguments - Values that are required or optional as input to execute a function.

Package - A collection of functions that can be installed and loaded to use.

Challenge 1: Coercion

Since vectors can only hold one type of data, something has to be done when we try to combine different types of data into one vector.

What type will each of these vectors be? Try to guess without running any code at first, then run the code and use class() to verify your answers.

```
num_logi <- c(1, 4, 6, TRUE)
num_char <- c(1, 3, "10", 6)
char_logi <- c("a", "b", TRUE)
```

```
tricky <- c("a", "b", "1", FALSE)
```

How many values in combined_logical are "TRUE" (as a character)?

```
combined_logical <- c(num_logi, char_logi)
```

Now that you've seen a few examples of coercion, you might have started to see that there are some rules about how types get converted. There is a hierarchy to coercion. Can you draw a diagram that represents the hierarchy of what types get converted to other types?

Day 3 second session

CHALLENGE 2: CREATING SEQUENCES

Write some code to generate the following vector:

[1] -3 -2 -1 0 1 2 3 -3 -2 -1 0 1 2 3 -3 -2 -1 0 1 2 3

CHALLENGE 2: CREATING SEQUENCES (CONTINUED)

Calculate the quantiles for the `complete_old` hindfoot lengths at every 5% level (0%, 5%, 10%, 15%, etc.)

End of day survey: https://forms.gle/PF3CbffhGzkxZ8Pt9

Day 4 "Working with data"

Attendance
1. Nick Record
2. Julia Brown
3. Jonathan Evanilla

4. Ben Tupper
5. Frendli Napo
6. Nikki Iozzia
7. LIsa Hogan
8. Aramish Barker
9. Nicolle Hincks
10. Glendon Zernicke
11.Bentley Davis
12. Michaela Pinette
13.Charlene Huerth
14. Alison Green-Parsons

Challenge 1: selecting and filtering

1. Use the surveys data to make a data.frame that has only data with years from 1980 to 1985.
2. Use the surveys data to make a data.frame that has only the following columns, in order: `year`, `month`, `species_id`, `plot_id`.

Challenge 2: piping

Use the surveys data to make a data.frame that has the columns `record_id`, `month`, and `species_id`, with data from the year 1988. Use a pipe between the function calls.

Challenge 3: plotting dates

Because the `ggplot()` function takes the data as its first argument, you can actually pipe data straight into `ggplot()`. Try building a pipeline that creates the date column and plots weight across date.

Challenge 4: making a time series

1. Use the split-apply-combine approach to make a `data.frame` that counts the total number of animals of each sex caught on each day in the `surveys` data.
2. Now use the data.frame you just made to plot the daily number of animals of each sex caught over time. It's up to you what `geom` to use, but a `line` plot might be a good choice. You should also think about how to differentiate which data corresponds to which sex.

Post Workshop Survey:
https://carpentries.typeform.com/to/UgVdRQ?slug=2024-07-08-wph-online

Short answer survey:
https://docs.google.com/forms/d/e/1FAIpQLSdgcdsdqyB3qZF_hGtVEAGWFw5SPsIHQBGZ8jepz9KVdCnZPw/viewform?usp=sf_link