Welcome to The Carpentries Etherpad!

This pad is synchronized as you type, so that everyone viewing this page sees the same text. This allows you to collaborate seamlessly on documents.

Users are expected to follow our code of conduct: https://docs.carpentries.org/topic_folders/policies/code-of-conduct.html

All content is publicly available under the Creative Commons Attribution License: https://creativecommons.org/licenses/by/4.0/

Workshop links:
Workshop site: https://ucsbcarpentry.github.io/2021-05-06-IntroR/

- *contains setup instructions, curriculum, daily schedule, link to this document, workshop survey links

Setup instructions: https://ucsbcarpentry.github.io/2021-05-06-IntroR/ (See "Setup" section at the bottom of the page)
Setup files: https://github.com/UCSBCarpentry/2021-05-06-IntroR/raw/gh-pages/data/IntroR_gapminder.zip

- *Note: Please use the latest version of R (4.2 +)

Thursday feedback form: https://forms.gle/wdtdPv7qzUCVw9pb6

- *please give feedback at the end of day 1

Post-workshop survey: https://forms.gle/64vnFxnihmCekWwp9

- *please complete after the workshop ends

Scripts:
- Day 1 Intro to Rstudio: https://drive.google.com/file/d/1LA6UGs2_sUPPorY8dO90wGAvyo8w2k0-/view?usp=sharing
- Day 1 Review and Subsetting: https://drive.google.com/file/d/1kXTpMg1cl2oRonz6m8TcX8HbWNIHyS1Y/view?usp=sharing

-

General Links:
UCSB Carpentry Home Page: https://ucsbcarpentry.github.io/

- *contains list of all upcoming and past workshops

UCSB Carpentry List-serve: https://groups.google.com/a/library.ucsb.edu/forum/#!forum/carpentry

- *sign-in with a ucsb.edu email address and click "join" to be updated about all future carpentry workshops

UCSB Carpentry Slack: https://join.slack.com/t/ucsbcarpentry/shared_invite/zt-qgv0qdeo-kglHcR2O9ctKiRARdsnRzA

-

R Resources:

https://www.r-graph-gallery.com/
http://www.cookbook-r.com/
https://www.youtube.com/watch?v=h29g21z0a68 "plotting anything with ggplot"
https://ggplot2.tidyverse.org/reference/ggtheme.html

Questions from Feedback day 1:
what is a workspace?/"Don't save your workspace" https://www.r-bloggers.com/2017/04/using-r-dont-save-your-workspace/
"numeric vs. integer" https://stackoverflow.com/questions/23660094/whats-the-difference-between-integer-class-and-numeric-class-in-r
dataframes vs. tibbles : https://cran.r-project.org/web/packages/tibble/vignettes/tibble.html
factors : https://swcarpentry.github.io/r-novice-inflammation/12-supp-factors/
*note factors are R's way to work with categorical data. It *used* to be that any character values were converted to a factor data type (which has different 'behavior' compared to character) automatically when you created a dataframe. With R version 4 + , R is moving away from that default. We didn't get into them much in this workshop because this default is changing. You may find that in certain situations factors may be useful for you, and we encourage you to do some more reading if that's the case!

Workshop Notes:
*answered questions, helpful comments and links from the chat, and exercise questions will be posted here
----

Q: Is there a way to run scripts separately or in chunks?
A: Rmarkdown allows you to chunk things separately, but this is another type of R format and we won't be going over Rmarkdown in this workshop. Another way is to select/highlight the script you want to run and then run the selected code. You can run R script line by line using Shift + Enter

Q: Do you have to install the same package every time you open R, even when working on the same project?
A: No, but you may have to load it everytime (if its not checked off in the packages list panel).

Q:when the warning message states "conflicts" after installing tidy verse, what does that mean?
A: It means a different package has it open or has loaded one or more packages as a dependency. Rstudio will not load it twice so it is just letting you know you already have it elsewhere.

Q: If you only received a "The downloaded binary packages are in..." does this mean you installed the packages?
A: Yes. This message is just telling you where those packages were saved or downloaded into on your computer.

Q: Whats the difference between the System Library and the User Library under the Packages tab?
A: The system library = packages that come with your base rstudio, The user library = stuff you downloaded from CRAN or a different repository (like a package from github).

Q: are commands with _ different than with . ? read_csv vs read.csv
A: Yes. read_csv is a command for reading things in Tidyverse. Read.csv is the read command from a different package.
A2: They both read in data, there are just slight differences. We prefer using the tidyverse read_csv

Q: What is a data frame?
A: A data frame is basically R's version of a data table with rows (observations of data) and columns (types of data)

Q: Is 'c' defining the data frame? Challenge 1 (I think)
A: c creates a vector (which is misleading). A vector is a set of values (all of the same type i.e. numeric) that can be assigned to a variable.

Q: Why cant we do character_vector <- as.numeric(character_vector)
A: You can, but you will overwrite the character_vector variable. Which is ok if that's what you want to do, but we're just being safe by giving it a new name (i.e. variable name).

Q: Is length() and ncol() the same?
A: Not really. length will give you the length of thet list which can be the number of columns, but not always. ncol() will definitively give you the number of columns.

================================= Day 2 Notes
===

FYI: Each time you run gapminder[sample(nrow(gapminder), 5),] you'll get differnent results as it selects a random set in the sample

Q: Does cbind automatically know which rows to match the logical variable to? or are all the values for each variable present?
A: short answer is no, we ensure it matches correctly by using a vector the same length as the data frame but R will match the index # of the vector with the same index # of the dataframe.

Q: Hi. Tidyverse loaded properly but I'm not set to my data folder so I receive the following message: 'data/nordic-data.txt' does not exist in current working directory. What was the script to set this folder as work directory?
A: Before setting the working directory, find where the file actually is, and then confirm your own working directory by either typing in getwd() in the console or click on the options gear in your files panel and click on 'get working directory. After you figure out where things are, move the files to the correct directory or set the working directory using the command setwd() or 'set working directory' in the files panel.

Tip: You can soft wrap the text so its easily readable. Tools > Global Options > Code > Soft-wrap R source files

Q: Why do we use single quote versus double quotes when using a list?

Q: How do you drop a vector? Without deleting everything from the global environment?
A: If you want to delete one item of you global environment you can use the function rm(name_of_the_object_you_want_to_delete)

Q: I'm confused on how TRUE and FALSE are connected to the vector.

Q: So the piping operator links functions for a dataframe?

A: Yes

Q: Where can we find these teaching materials (for the ggplot2 lesson)?
A: https://ucsbcarpentry.github.io/2021-05-06-IntroR/07-plot-ggplot2/index.html

Q: Does this also work for filtering for multiple values? %in%in%
A: Yes, but the %in% operator starts making more sense if you have a lot of possible values