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.

Use of this service is restricted to members of The Carpentries community; this is not for general purpose use (for that, try etherpad.wikimedia.org).

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

Logistics:

Lunch: West bank restaurants that are walkable for lunch (feel free to add your name if you'd like to coordinate lunch with other students):
Republic - pub at 221 Cedar Ave, Minneapolis
Acadia - pub at 329 Cedar Ave, Minneapolis
Dilla - Ethiopian at 1813 Riverside Ave
Red Sea - Ethiopian at 320 Cedar Ave South
Hard Times - cash only, vegan-friendly at 1821 Riverside Ave,
Chipotle - fast food Mexican, 229 Cedar Ave S
Campus Cafe - Turkish at 1810 Riverside Ave

Gender-neutral restrooms: https://egis.umn.edu/twincities/index.html?config=config_gender_neutral_restrooms.json

- Closest ones to Andersen are on Social Sciences (Floors 1,5, 6, 8, 9, 12, 14), Heller Hall (Floors 7 and 10) and Mondale Hall (floor 4)

Link to materials: https://chaochihl.github.io/2019-10-24-umntc/

BASH

GUI = Graphical User Interface

pwd = present working directory
cd = change directory (cd entered without pathway brings back to root computer directory)
/ = delineates pathway from root working directory

tab completion: if you type part of a command or file followed by the tab key, the shell will look for a file that might match what you've typed and complete it for you. If it can't match anything the shell will beep at you. Being lazy is good when programming :)

Up and down arrow keys will allow you to navigate past commands you've executed.

Commands can be concatenated for modified output

q = quit help file

If you are using gitbash it automically color codes things like zipped files, executables and directories. This is to help you pick certain file types out quickly. This can be modifed in both Unix or gitbash, but is a bit more advanced and requires editing your .profile file typically.

touch whean applied to an existing file modifies the time
cat shows contents of a file

Can create a file in nano, and it still can be used in vim.

GIT & GITHUB

Name/Position or Domain or Research Interests/Why you are interested in learning programming skills

Elena/PhD Student in Industrial-Organizational Psychology/Enable research related to working with large datasets & using techniques like machine learning, natural language processing, etc.; improve reproducibility of my analyses

Penny - Researcher focused in corn meiosis - would like to learn to be more effecient with my research in genomics. Can mine my own data for interesting questions.

Kayla/Masters student/Research Assistant-microbiome analyses and research-i need to know how to do this for my thesis projects incorporating very large data sets

David/ PhD Student in Wildlife Ecology/ Work with large GPS telemetry datasets to model animal movement; interested in version control for collaboration and reproducibility
Annika/ Research Technician in GCD, interested in filopodia formation in amoebas/ interested in programming to become better at quantitative data analysis

Annie/PhD student in Comparative molecular biosciences/ interested to learn the basics of R and continue to use it for data analysis and statistical anlysis for my reserach project which includes illumina/nanopore sequencing data, working with very large dataset for describing molecular epidemiology of bacterial foodbrone pathogens

Jonathan Dregni Entomology, work with lots of ecological data and need to better understand version control and data management

Cody - Librarian who helps researchers with text mining projects. I like how software carpentry brings together researchers across disciplinary boundaries!

Ky/PhD Student/Research assistant - microbiome research analysis

Virdi/Postdoc in Agronomy and plant genetics-Research ares in soybean genetics/genomics/biostatistics

Andy - Working in Marketing / Marketing Research - Interested in using R for Data Visualization, etc.

Brian - Postdoctoral Researcher CBS/EEB - I like to use programming to study all the omics. Also, I enjoy introducing people to the world of programming.

Wanda - Librarian - Data Curation for Data Repository for University of Minnesota (DRUM) and Data Curation Network (DCN). Looking to be able to curate R and other languages or at least have a better understanding of code

Jambay- VisitingScholar, Earth Science- Hydrological Research Data Analysis - I am interested to learn programming skills for for learning Hydrological modelling and teaching students

Rebecca - Research Scientist Plant Pathology. Gain skills to use R to anaylze field and greenhbouse data of plant disease assays. Also engaged in molecular work and need basic skills to do bioinformatics and optimize data visualization for manuscripts and presentations.

Exercise 1:

Imagine you drafted an excellent paragraph for a paper you are writing, but later ruin it. How would you retrieve the *excellent* version of your conclusion? Is it even possible?
I would got back to a previous version
i save a copy to cloud drives but it is stil scary to loose the nice version

time machine on my Mac

I just always manually backup my cloud files by downloading that dates version
look to a previous version in Google Docs
Only with a very meticulous backup system that tracks the datetime and recent changes to the working file.

I use google docs so I am always able to retrieve an earlier version

Imagine you have 5 co-authors. How would you manage the changes and comments they make to your paper? If you use LibreOffice Writer or Microsoft Word, what happens if you accept changes made using the Track Changes option? Do you have a history of those changes?

i can download the older version and keep track of changes, but thats a lot of work and would like something easier to keep tracks
we would use google docs to collaborate and then convert before publication

I would go back a to a previous saved version. If I remember correctly some items are tracked as they are changed multiple times.

R for Reproducible Scientific Analysis

Using git in R studio: https://support.rstudio.com/hc/en-us/articles/200532077-Version-Control-with-Git-and-SVN

Data links:
https://raw.githubusercontent.com/swcarpentry/r-novice-gapminder/gh-pages/_episodes_rmd/data/gapminder_data.csv

https://raw.githubusercontent.com/swcarpentry/r-novice-gapminder/gh-pages/_episodes_rmd/data/gapminder_wide.csv

```
cats_2 <-data.frame(coat = c("calico","black","tabby","tabby"),
                    weight = c(2.1,5.0,3.2, "2.3 or 2.4"),
                    likes_string = c(1,0,1,1))
write.csv(x = cats_2, file = "data/feline-data_v2.csv", row.names=F)
```

Best practices for variable naming:

Some people use underscores (my_variable), some use periods (my.variable), or camel case (MyVariable)

Spaces will be problematic (my variable), and you can't start with a period, underscore, or number

Variables are case sensitive

Vectors need to all be the same "type" (numbers, characters)

You can print variables by typing print(x) or just typing x where x is the variable

To run a command, you can either click "run" in RStudio, or Ctrl+Enter (Command+Enter in Mac)

Packrat is a package management system that will record which version of a package you're using. R packages are available through CRAN. There are some "off label" packages through GitHub, but those are less consistently updated

You can organize your script to be more clear using ctrl + shift + a (command instead of ctrl in MAC)
"getwd()" is the equivalent in R of "pwd" in the terminal

code:

```
cats=data.frame(coat=c("calico","black","tabby"),weight=c(2.1,5.0,3.2),likes_string=c(1,0,1))
write.csv(cats,file="Data/cat-data.csv")#export csv file
cats=read.csv(file="Data/cat-data.csv")
cats
gapminder_data=read.csv(file="Data/gapminder_data.csv")

calcGDP <- function(dat, year=NULL, country=NULL) {
  if(!is.null(year)) {
    dat <- dat[dat$year %in% year, ]
  }
  if(!is.null(country)) {
```

```
  dat <- dat[dat$country %in% country, ]
 }
 gdp <- dat$pop * dat$gdpPercap
 new <- cbind(dat, gdp=gdp)
 return(new)
}
```

dplyr (data manipulation) lesson:
https://swcarpentry.github.io/r-novice-gapminder/13-dplyr/index.html

ggplot2 (visualization) lesson:
https://swcarpentry.github.io/r-novice-gapminder/08-plot-ggplot2/index.html

%>% is a pipe, it pipes data through a set of functions without having to specify and save the data each time

select() selects variables/columns from dataset

filter() selects rows from dataset

Helpful Resource!
data wrangling cheatsheet: https://resources.rstudio.com/rstudio-developed/data-transformation

google R colorbrewer palette
can set color="uniquecolorID"