Day 2 Cheatsheet

Reproducibility

Major concepts

- **Reproducibility** - A different analyst re-performs the analysis with the same code and the same data and obtains the same result.
- **Repeatable** - keeping everything the same but repeating the analysis - do we get the same results
- **Reproducible** - using the same data and analysis but in the hands of another researcher - do we get the same results?
- **Replicable** - with new data do we obtain the same inferences?

Functions

Library/Package	Piece of code	Example of usage	What it does
Base `R`	`sessionInfo()`	`sessionInfo()`	Returns the R version information, the OS, and the attached packages in the current R session.

More resources

- The RMarkdown book
- Jenny Bryan's organizational strategies.
- Write efficient R code for science.
- Reproducibility in Cancer Informatics course

Data Input/Output

Major concepts

- **Delimited file** - columns within the file have boundaries created with some type of punctuation - for example, a csv file stands for comma separated values, thus the columns are delimited or separated by commas. txt files often use tabs.
- **tibble** – a rectangular data frame, where data are split into rows and columns.
- **File path** - where a file lives on your computer (or on the internet) - File paths can be relative or absolute.
- **Getting help** - For any function, you can write `?FUNCTION_NAME`, or `help("FUNCTION_NAME")` to look at the help file for that function.
- **R Projects** set the working directory where the .Rproj file is.
- **Reading in data (manual)** - *Download the data*, *Put data in the project folder*. Go to File, Import Dataset, From Text (`readr`), browse for the file, and finally click "Update" and "Import".

Functions

Library/Package	Piece of code	Example of usage	What it does
`tidyverse` (`readr`)	`read_csv()`	`df <- read_csv("<url>")`	Reads in a CSV file from a specified file path.
`tidyverse` (`readr`)	`read_tsv()`	`df <- read_tsv("<url>")`	Reads in a TSV file from a specified file path.
`tidyverse` (`readr`)	`read_delim()`	`df <- read_delim("<url>", delim = ":")`	Reads in a delimited file from a specified file path.
Base `R`	`head()` / `tail()`	`head(df)` / `tail(df)`	Returns the first or last part of a dataset (or other object).
Base `R`	`getwd()`	`getwd()`	Finds the current working directory.
Base `R`	`setwd()`	`setwd("Desktop")`	Changes the current working directory.
Base `R`	`View()`	`View(df)`	View a spreadsheet-style data viewer on a matrix-like R object.
Base `R`	`str()`	`str(df)`	Display the internal structure of an R object (dimensions, class, etc).

* This format was adapted from the cheatsheet format from AlexsLemonade.