

# 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	<code>sessionInfo()</code>	<code>sessionInfo()</code>	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.
- **R Markdown** files look for file paths relative to where the .Rmd file is located

## Functions

Library/Package	Piece of code	Example of usage	What it does
readr	read_csv()	df <- read_csv("<url>")	Reads in a TSV file from a specified file path. This function can be tailored to read in other common types of files. i.e. read_csv(), read_rds(), etc.
readxl	read_excel()	read_excel("dataset.xlsx")	Read in an excel file.
Base R	getwd()	getwd()	Finds the current working directory.
Base R	setwd()	setwd("Desktop")	Changes the current working directory.

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