Introduction to R for Data Management and Analysis

Marcel Ramos

Review Session

- Questions from previous class
- Typical workflow using R

Topics to cover

- Reporting tools
- RMarkdown syntax basics
- Analysis example
  - OpenCaseStudies - Health Expenditure

Reporting Tools

- Reproducibility is a major issue across disciplines[1]
- Being able to reproduce your *own* analysis is essential
- Reviewers may ask to redo your analysis with slight modifications
- "Future" you will *thank you* for creating an analysis workflow that is easy to follow and well annotated
- R Markdown is an essential tool for effective communication and dissemination of reproducible results

[1] {https://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibility-1.19970}

R Markdown Syntax

1. Visit the RStudio Tutorial page
2. Go through the first 9 lessons up to and including Output Formats
3. Download the RStudio Cheatsheet here[2]

[2] https://www.rstudio.com/resources/cheatsheets/

Rmd header

- yaml 'front matter'
- yaml stands for 'YAML ain't markup language'
- A special type of syntax similar to JSON for configurations
- Driven by a key value pairs

```
---
title: "Session 8"
output: word_document
---
```

- Tells R what kind of output to generate and how

Rmd vs Md

- `Rmd` Supports R code
- `md` is an intermediate between `Rmd` and `HTML`
- `Rmd` $>$ `md` $>$ `HTML`
- When the output is `HTML`
- `md` is a popular format for GitHub READMEs

Bioconductor 2020 Workshops (real-world example)

- Bioc2020 Workshops
- Use an Rmd to generate an md file
- Run the R code in the chunks
- Use the md file to publish on the website

RMarkdown basics

- Opening an Rmd file in RStudio
- Simplifies HTML
  - # means <h1> or level 1 header
- Documents need to be generated using `knit`
- create lists using the asterisk or dash
- Nesting lists

OpenCaseStudies Example

- Health Expenditures link