Statistics II

Week 1

Content for Today

1. How will the labs and assignments work
2. Introduction to `R Markdown`
3. Introduction to `dplyr`

Corona measures

- Make sure you sign the attendance sheet!
- 1.5 meter distance at all times (we cannot approach your screens)
- Masks can be taken off when once seated at your desk. Keep it on if possible.
- We will air the room every 30 minutes.
- Leave the room at ten to the hour.

Labs

The optional drop-in lab sessions exist primarily to help with the coding application of the materials presented in the lecture.

Each session will begin with a brief overview of the week's content and a few minutes for questions, followed by coding exercises that will help you to prepare for the assignments and develop your coding skills.

Please attend the lab you're assigned to.

Slides and scripts will be available on Moodle after the labs.

Assignments

- The professors will upload an assignment every 2 weeks on Moodle, with the corresponding data set to work with. You have 2 weeks to work on it.
- Your tasks:
  - Work on the assignment on an R Markdown document (in R)
  - Knit the document as an HTML file
  - Upload **both** the R Markdown and the HTML file on Moodle
- You can work collaboratively with peers but each one must write their own code and answers.

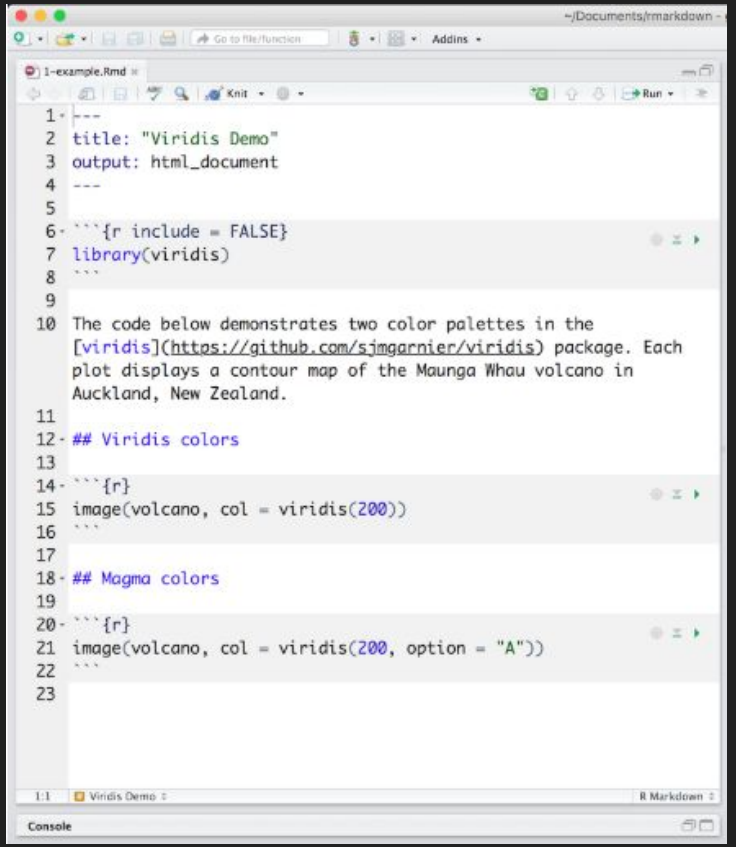
RMarkdown

Introduction to `Rmarkdown`

`RMarkdown` is an authoring framework for data science. A single `RMarkdown` file can be used to:

- Save and execute code
- Generate high quality reports that can be shared with an audience

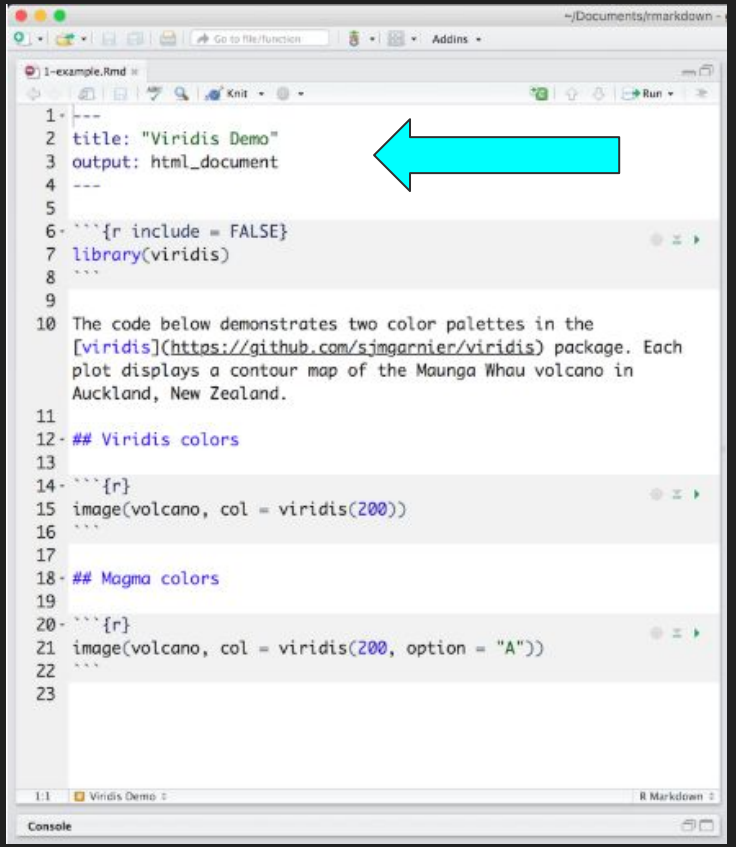
We will use `RMarkdown` to submit our weekly assignments.



Contents in `Rmarkdown`

`RMarkdown` files support three types of content:

- YAML headers surrounded by ---
  Meta-data that guides the file build-up process.
  (not used in assignments, since submissions are anonymous)

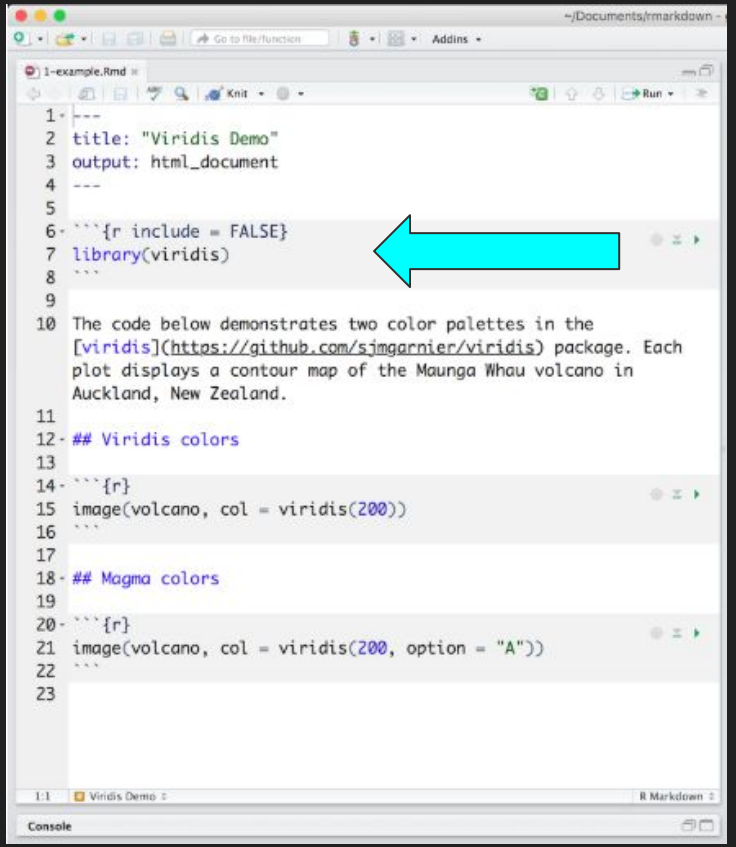


Contents in `Rmarkdown`

`RMarkdown` files support three types of content:

- R code chunks surrounded by ```
  Chunks take code as an input. It works in the same way your .R scripts did for Stats I.

start a chunk: ```` ```{r} ````
end a chunk: ```` ``` ````

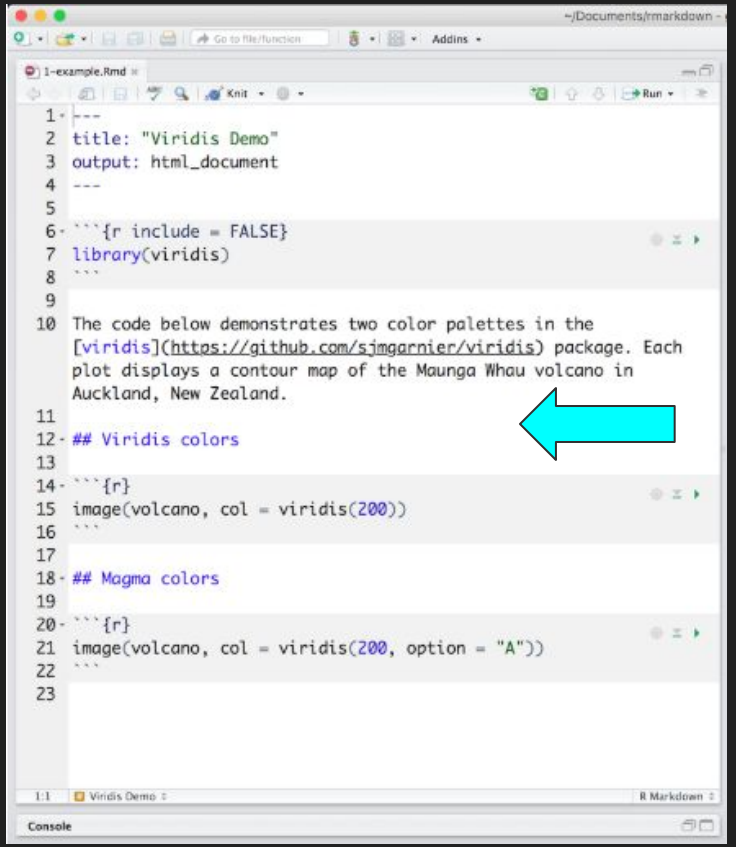


Contents in `Rmarkdown`

`RMarkdown` files support three types of content:

- Text mixed with simple text formatting
  Takes text as input.

[RMarkdown Cheatsheet]



Example

Let’s see what an assignment can look like in R.

Coding with `dplyr`

Working with R: `dplyr`

Throughout this lab, we will be using the `dplyr` package for most data-wrangling (rather than base R functionality):

We'll often use the pipe operator (`%>%`) to string together commands, and rely on the `dplyr` "verbs". For example:

`select:` subset columns
`filter:` subset rows
`arrange:` reorder rows
`mutate:` add columns to existing data
`summarize:` summarize values in the dataset
`group_by:` defines groups within dataset

Helpful Hints

What to do when you get stuck on coding problems

First, don’t panic. Then:

1. Check your code (missing parentheses, packages, stray commas, etc.)
2. Google the error message
3. Search on https://stackoverflow.com/ or look on YouTube
4. Ask for help (from stackoverflow, friends, or your TA)

Further resources

- Reminder of the basics: https://tinyurl.com/vkebh2f
- A comprehensive guide to R: http://qpolr.com/
- RMarkdown: The definitive guide https://tinyurl.com/y4tyfqmg
- RMarkdown cheatsheet
- Data wrangling with `dplyr`: https://tinyurl.com/vyrv596
- `dplyr` video tutorial: https://www.youtube.com/watch?v=jWjqLW-u3hc
- Data wrangling cheatsheet