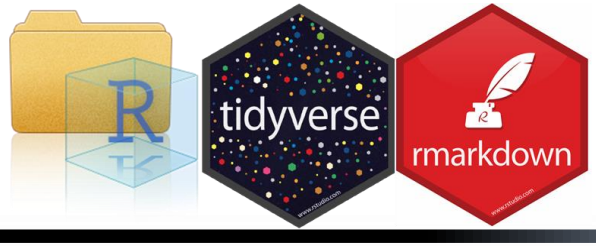
R: A Hitchhikers Guide to Reproducible Research

- Take control



Brendan Palmer,
Clinical Research Facility - Cork &
School of Public Health
@B_A_Palmer

CRF-C
HRB Clinical Research Facility Cork

UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh
School of Public Health

R Markdown

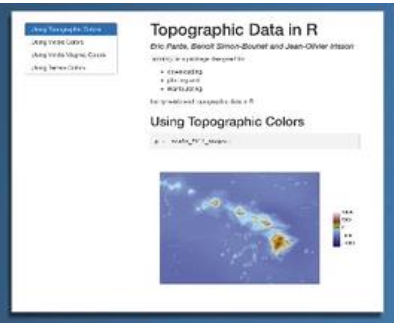
- R Markdown combines the code you wrote, the output produced and you own comments

- You can view it as a digital lab notebook, where you are both recording what you're doing, and what you were thinking while you were doing it!

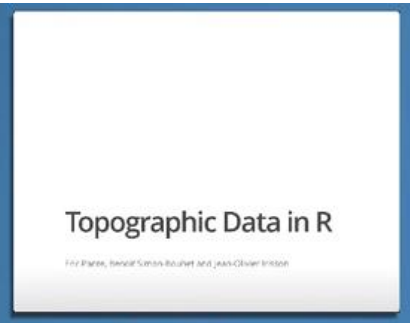
- R Markdown outputs can take many forms
    - Word documents, PDFs, slideshows etc.

- Once created the .Rmd file get sent to knitr, which executes the chunks of code and creates a new markdown document
    - this is then processed by pandoc which creates the finished file
        - knitr and pandoc are external websites

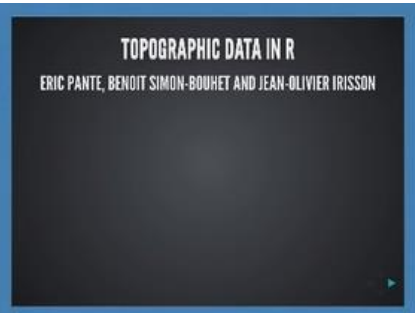
What has R Markdown ever done for us?



html

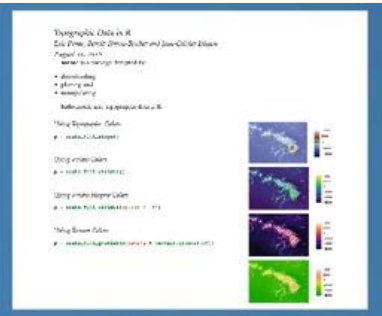


ioslides

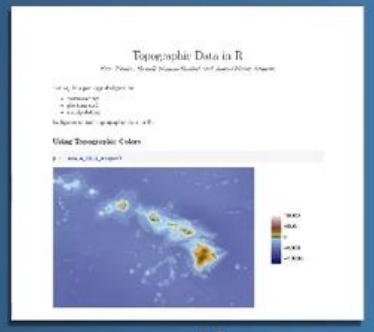


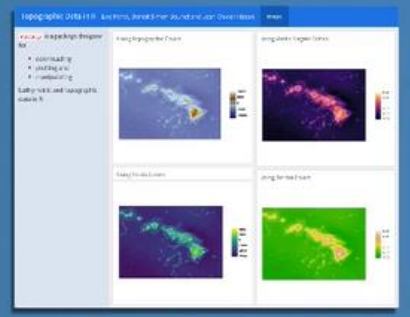
reveal.js


rtf

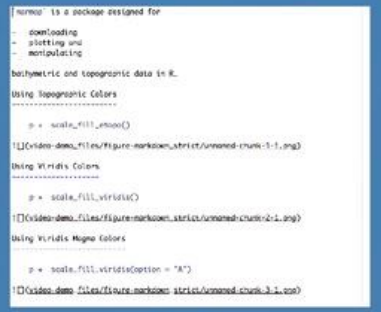

tufte handout

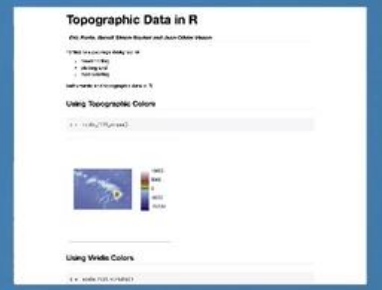
book


pdf

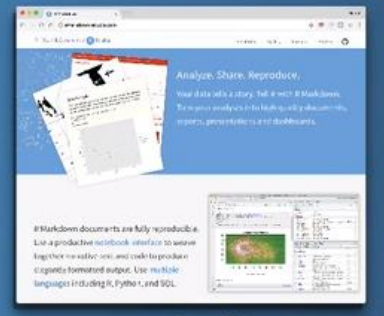

dashboard

slidy

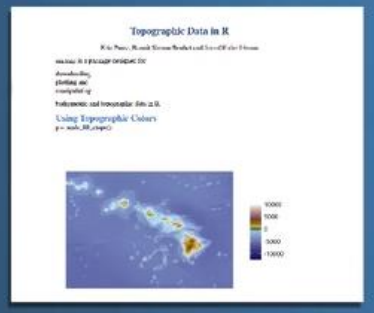

markdown

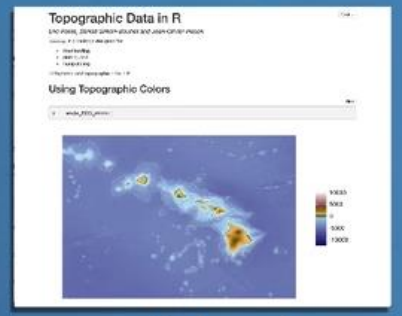


package vignette

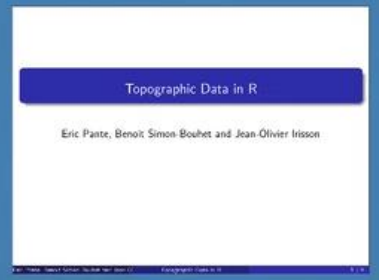


website


Word



notebook



beamer


latex

custom template


shiny app

R Markdown

YAML header

```
---
title: "Diamond sizes"
date: 2016-08-25
output: html_document
---
```

Chunks of code

````
```{r setup, include = FALSE}
library(ggplot2)
library(dplyr)
smaller <- diamonds %>%
filter(carat <= 2.5)
```
````

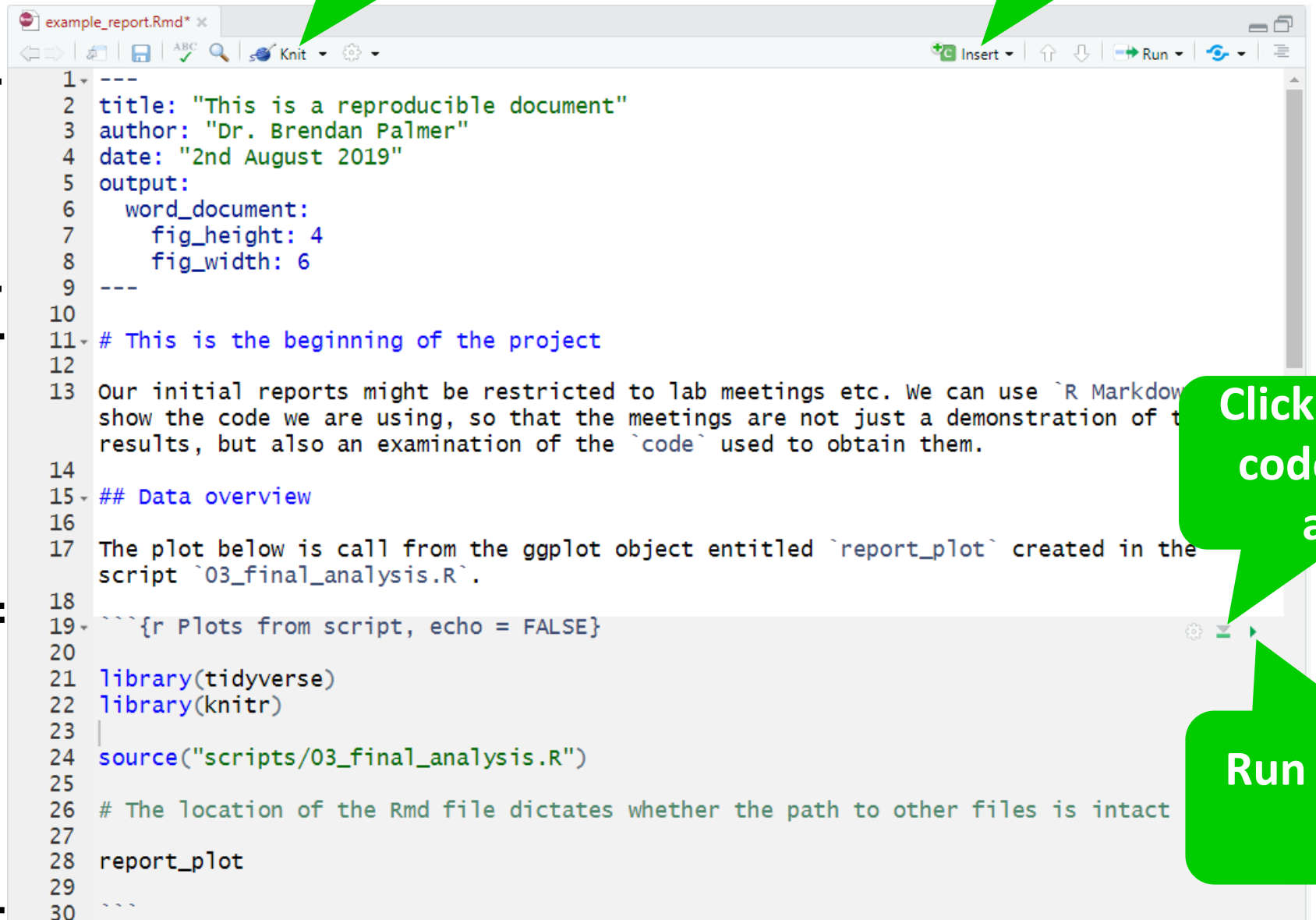
Plain text with data outputs from R code

````
We have data about `r nrow(diamonds)`
diamonds. Only
`r nrow(diamonds) - nrow(smaller)` are
larger than
2.5 carats. The distribution of the
remainder is shown below:
````

Chunks of code

````
```{r, echo = FALSE}
smaller %>%
ggplot(aes(carat)) +
geom_freqpoly(binwidth = 0.01)
```
````

R Markdown



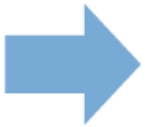
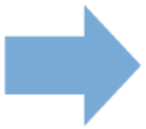

```
---
title: "This is a reproducible document"
author: "Dr. Brendan Palmer"
date: "2nd August 2019"
output:
  word_document:
    fig_height: 4
    fig_width: 6
---

# This is the beginning of the project

Our initial reports might be restricted to lab meetings etc. We can use `R Markdow
show the code we are using, so that the meetings are not just a demonstration of t
results, but also an examination of the `code` used to obtain them.

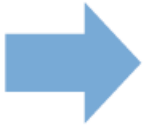
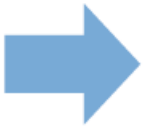
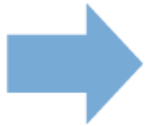
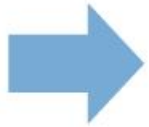
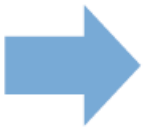
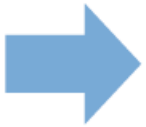
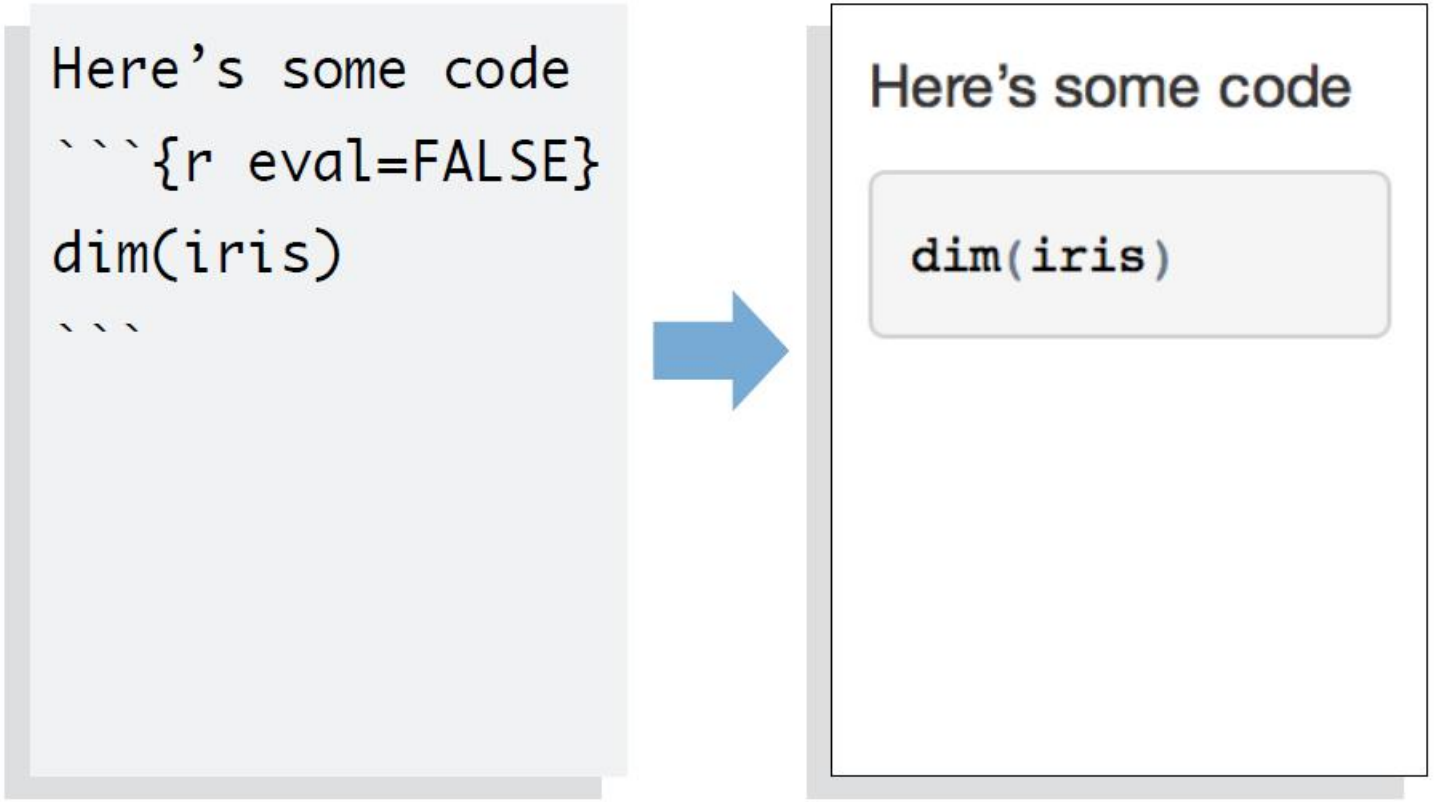
## Data overview

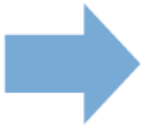
The plot below is call from the ggplot object entitled `report_plot` created in the
script `03_final_analysis.R`.

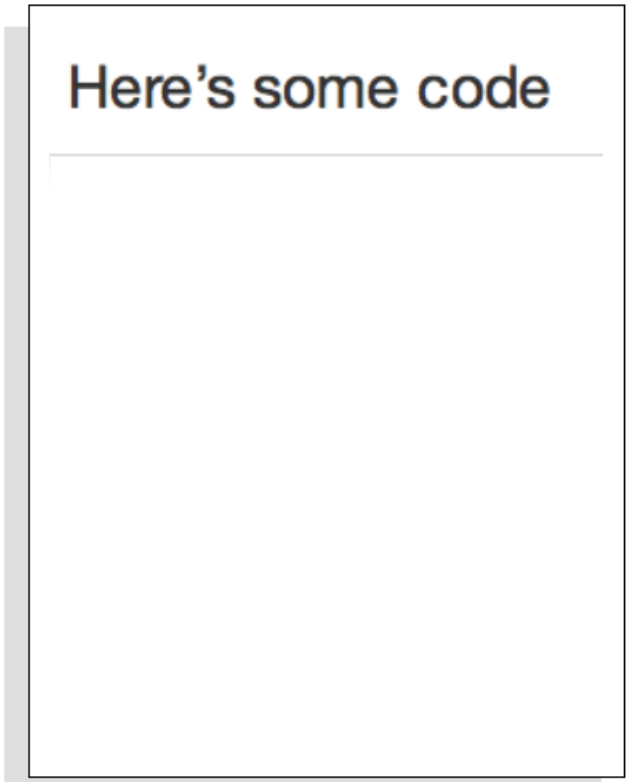
```{r Plots from script, echo = FALSE}

library(tidyverse)
library(knitr)

source("scripts/03_final_analysis.R")

# The location of the Rmd file dictates whether the path to other files is intact

report_plot

```
```

R Markdown - Headers

```
# Header 1
## Header 2
### Header 3
#### Header 4
##### Header 5
###### Header 6
```



Header 1

Header 2

Header 3

Header 4

Header 5

Header 6

R Markdown - Formatting

```
Text
_italics_
__bold__
`code`
```



Text

italics

bold

`code`

R Markdown - Lists

```
Bullets

* bullet 1
* bullet 2

Numbered list

1. item 1
2. item 2
```



Bullets

- bullet 1
- bullet 2

Numbered list

1. item 1
2. item 2

R Markdown - Hyperlinks

```
This is a
[link](www.git.com).
```



This is a link.

R Markdown - Equations

```
According to
Einstein,
$E=mc^{2}$
```



According to Einstein, $E = mc^2$

R Markdown - Images

```
![](images/1.png)
The RStudio logo.
```



R Markdown – Code chunks

````
Here's some code
```{r}
dim(iris)
```
````



Here's some code

```
dim(iris)
```

```
## [1] 150   5
```

R Markdown – Code chunks

````
Here's some code
```{r echo=FALSE}
dim(iris)
```
````



Here's some code

```
## [1] 150   5
```

- Displays the results but not the code

R Markdown – Code chunks



- Displays the code, but not the results (code is not run)

R Markdown – Code chunks

````
Here's some code
```{r include=FALSE}
dim(iris)
```
````





- Neither code nor results displayed (but the code is run)

R Markdown

Tips:

- Ensure each notebook has a descriptive title
- If you reach a research dead end, don't delete it
    - Write a note about it. It may be useful later

- At the end of each day run a clean knit of the note book

- If there's an error message, correct it while its still fresh in your mind

- If you want your code to be reproducible in the long run, you'll need to keep a rigorous track of the package versions
    - Consider using the packrat package to help with this

- For an deeper dive into R Markdown visit https://bookdown.org/yihui/rmarkdown

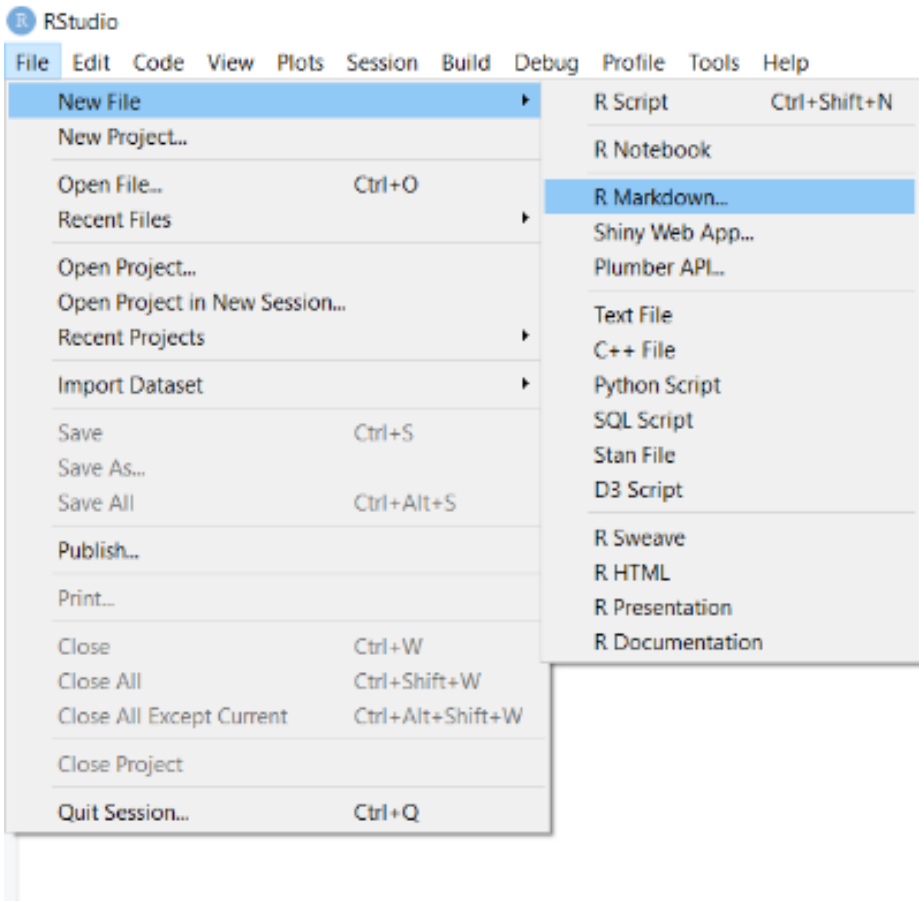
Introduction to R Markdown

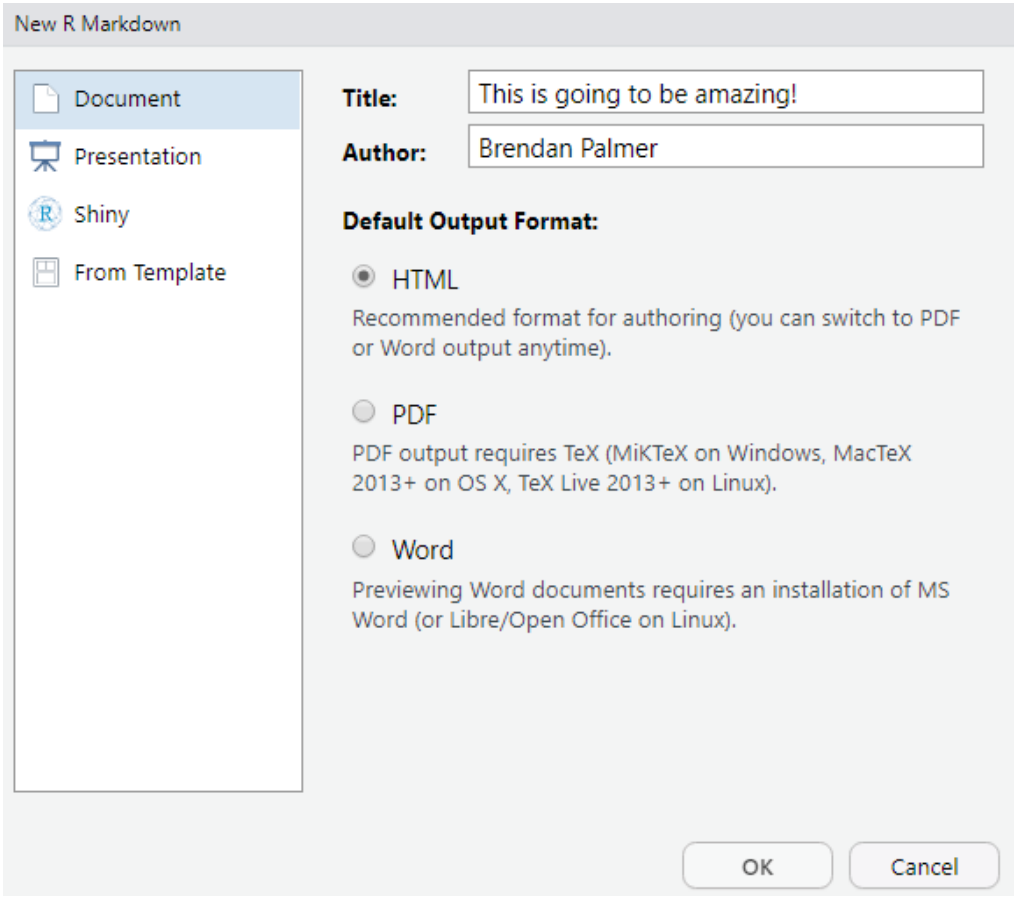
- We're now going to look at a R Markdown file that provides some of the tips and tricks you'll need yourselves
    - Code chunks
    - Formatting
    - Tables
    - Figures etc.

- Open the R Markdown file Day_3/intro_to_RMarkdown.rmd
- Open the R Markdown file Day_3/code_chunks.rmd
- Open the R Markdown file fancy_R-markdown_bits.rmd

Worksheet – R Markdown DIY

- Create a R Markdown document and begin compiling





- Save the file as Day_3/diy_r_markdown.rmd

Where to next?

- Understanding basic statistical concepts
www.khanacademy.org

- Collection of YouTube videos describing statistics through R
http://rafalab.github.io/pages/harvardx.html

- You know what you want to do, but don't know how to do it
https://stats.stackexchange.com/

Structured training……

coursera

Catalog | statistics with r

Fo

Course Languages

- [ ] English 557
- [ ] Spanish 12
- [ ] Chinese (Simplified) 7

Show More

Subtitle Languages

- [x] English 586
- [ ] Chinese (Simplified) 62
- [ ] Spanish 60

Show More

All Topics

- [ ] Data Science 212
- [ ] Business 207
- [ ] Computer Science 162

Show More

You searched for **statistics with r.** 586 matches

Active filters: English ×

Courses and Specializations

Statistics with R

5-course Specialization · Duke University

The R Programming Environment

Johns Hopkins University

Basic Statistics

University of Amsterdam

Advanced Linear Models for Data Science 2: Statistical Linear Models

Johns Hopkins University

Viewing 42 results matching

Search:

"statistics r" ✖ CLEAR ALL

Refine your search

Availability

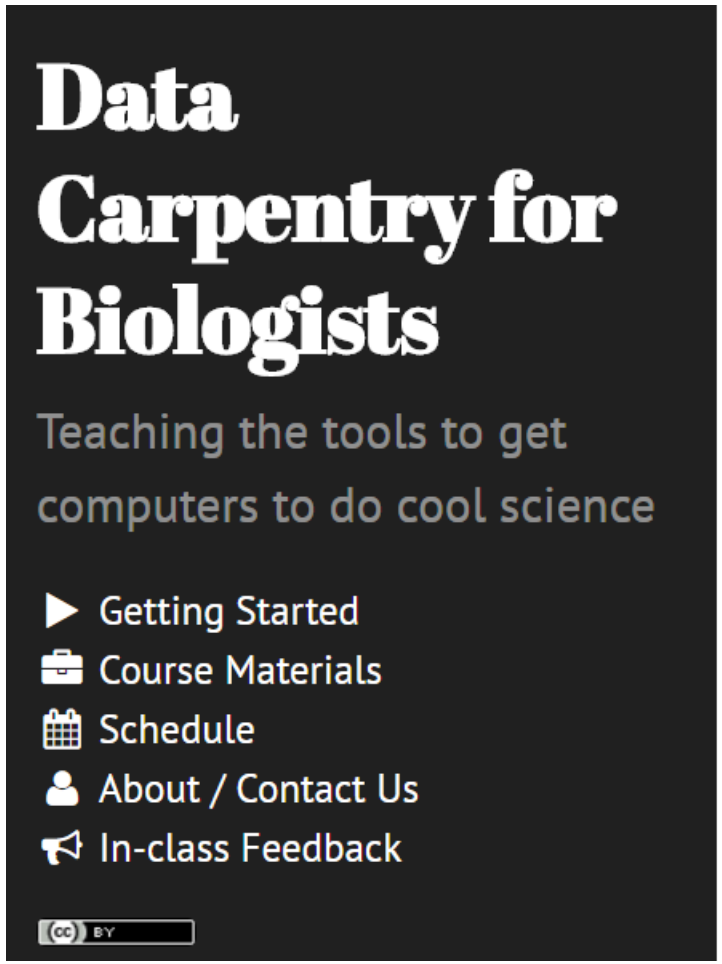
Current	20
Starting Soon	7
Upcoming	5
Self-Paced	24
Archived	8

Subjects

Biology & Life Sciences	9
Business & Management	3
Computer Science	12
Data Analysis & Statistics	

Statistics and R	HarvardX	Course	7/12/2017
Explore Statistics with R	KIx	Course	7/7/2015
Introduction to R for Data Science	Microsoft	Course	10/1/2017
Programming with R for Data Science	Microsoft	Course	10/1/2017
Analyzing Big Data with Microsoft R	Microsoft	Course	10/1/2017
Statistical Analysis in Bioinformatics	USMx	Course	10/23/2017

And finally



Googling for Help

Notes

> Check that top Google results haven't change and adjust as needed
> Current top 3 hits:
>
> 1. https://blog.exploratory.io/selecting-columns-809bdd1ef615
> 2. https://dplyr.tidyverse.org/reference/select.html
> 3. https://stackoverflow.com/questions/21502465/replacement-for-rename-in-dplyr

Link to the data carpentry webpage

Digital Badge
- Responsible Conduct of Research

Digital Badge in the Responsible Conduct of Research

The UCC Digital Badge in the Responsible Conduct of Research is offered to researchers who complete a tailored, blended learning course on responsible, reproducible research. The digital badge aims enhance Research Integrity awareness among the research community in UCC by promoting frank discussions about responsible research. Transparent data practices are a key skill of RI and participants will also learn about the FAIR principles, good data management practices and reproducibility. The badge is available to researchers at all levels and career stages from PhD to PI.

- Coming to a School of Applied Psychology near you!!

meetup

Cork (Ireland) R-Users Group