

Getting started with , and R Markdown

Professor Andy Field

 @profandyfield

 www.youtube.com/user/ProfAndyField/

 www.discoveringstatistics.com

 www.milton-the-cat.rocks

 www.discovr.rocks



ANDY FIELD



Key information

Teaching

- 1 × weekly lecture (2 in weeks 1 and 2) on theory
- 1 × weekly practical class (2 in weeks 1 and 2) about  and  Studio

Assessment

- 2 × 24-hour Take Away Papers (TAPs)
 - TAP 1: 30%
 - TAP 2: 30%
- 1 × Report (40%)



ANDY FIELD



Part 1: Introducing and Studio



ANDY FIELD





```
R Console
/Volumes/Alpha Lacertae/Documents/Academic/data/
[History restored from /Users/andyfield/.Rapp/History]

> library(tidyverse)
— Attaching packages —

tidyverse 1.2.1 —
✓ ggplot2 3.2.1   ✓ purrr 0.3.2
✓ tibble 2.1.3   ✓ dplyr 0.8.3
✓ tidyr 0.8.3    ✓ stringr 1.4.0
✓ readr 1.3.1    ✓ forcats 0.4.0
— Conflicts —

tidyverse_conflicts() —
* dplyr::filter() masks stats::filter()
* dplyr::lag()     masks stats::lag()
> library(here)
here() starts at /Users/andyfield
> getwd()
[1] "/Volumes/Alpha Lacertae/Documents/Academic/data/teaching_data/ais_data"
> zombie_tib <- readr::read_csv("ais_14_zombie_taser.csv")
Parsed with column specification:
cols(
  id = col_double(),
  immobility = col_double(),
  r_tms = col_double(),
  taser = col_double()
)
> |
```



ANDY FIELD







ANDY FIELD







- Free software environment for statistical analysis and graphics

R Studio®

- A free integrated development environment (IDE) for 
- You use  Studio as a way to interact with 
 - Install  and forget about it until such time that you need to update it
 - Install  Studio
 - Set up  Studio to suit your personality and enjoy your status as rad data scientist hipster



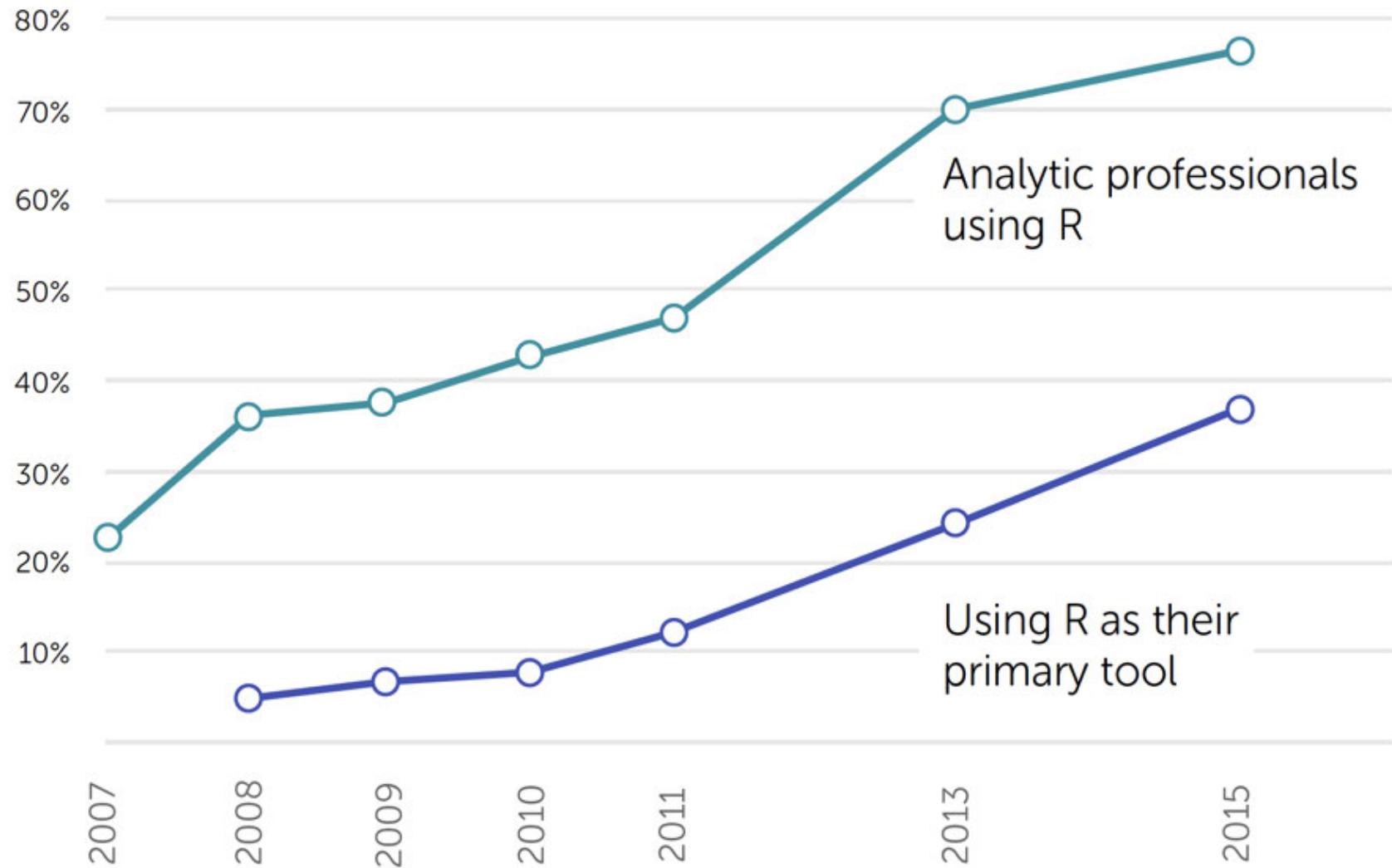
ANDY FIELD



Why R?

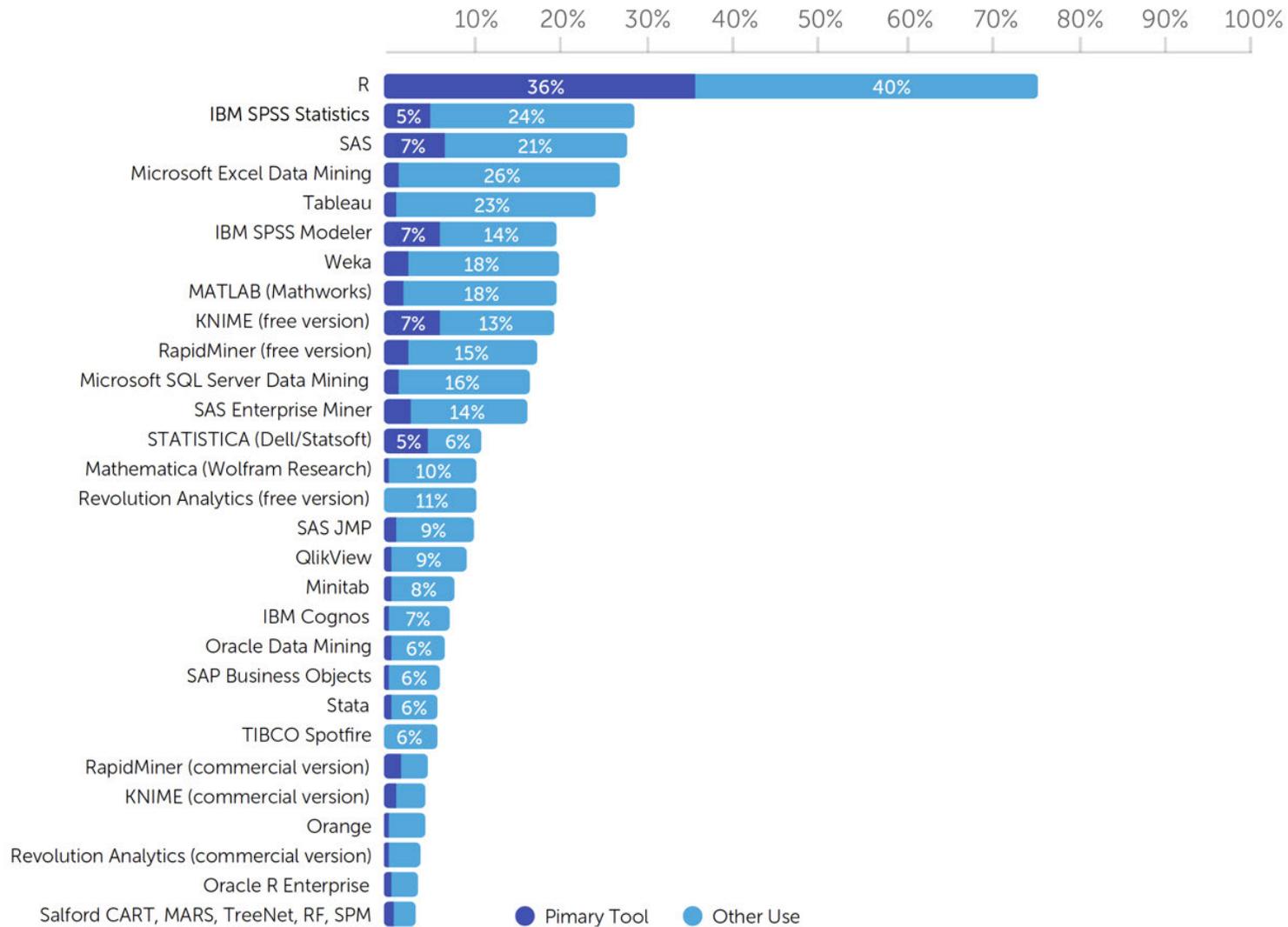
- Transferable skills
- R is the most widely used data analysis software
- Reproducible science
- Cutting edge
- One stop shop
- R Studio is amazeballs





ANDY FIELD

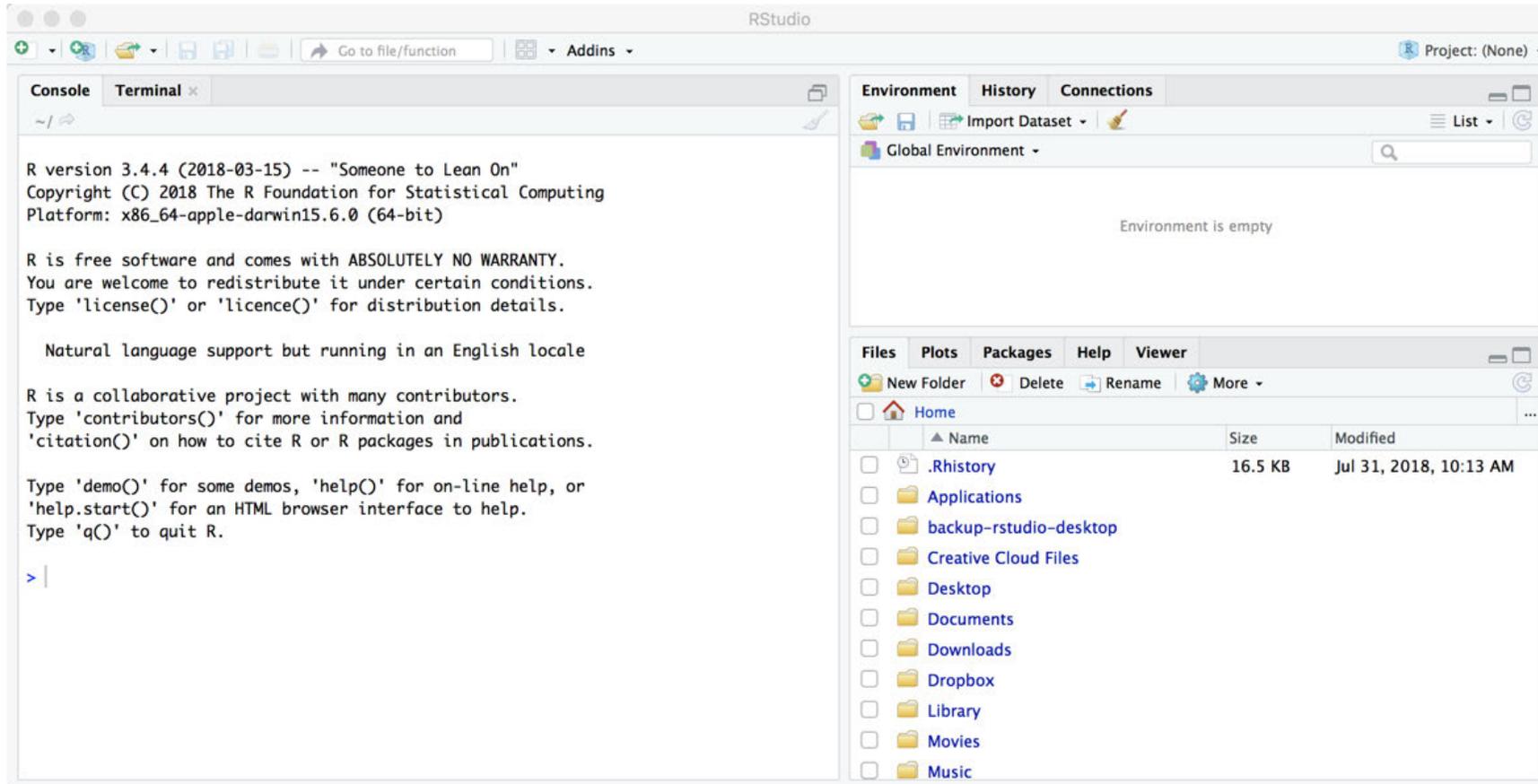




ANDY FIELD



Starting up R Studio



ANDY FIELD



Try it! Colour scheme

- Windows
 - Tools > Options
- MacOS
 - Tools > Global Options
 - RStudio > Preferences



The screenshot shows the RStudio Options dialog box with the 'Appearance' tab selected. The 'Editor theme' dropdown menu is open, showing a list of themes with 'Solarized Dark' highlighted. The 'RStudio theme' is set to 'Modern', the 'Editor font' is 'Monaco', and the 'Editor Font size' is '10'. A preview window on the right shows R code with the Solarized Dark color scheme applied.

```
# plotting of R objects
plot <- function(x, y, ...)
{
  if (is.function(x) &&
      is.null(attr(x, "class")))
  {
    if (missing(y))
      y <- NULL

    # check for ylab argument
    hasylab <- function(...)
      !all(is.na(
        pmatch(names(list(...)),
                 "ylab")))

    if (hasylab(...))
      plot.function(x, y, ...)

    else
      plot.function(
        x, y,
        ylab = paste(
          deparse(substitute(x)),
          "(x)",
          ...)
      )
  }
  else
    UseMethod("plot")
}
```



Try it! Pane locations



Options

Choose the layout of the panes in RStudio by selecting from the controls in each quadrant.

Source Console

Source
Console
Environment, History, Connections, Build, VCS, Presentation
Files, Plots, Packages, Help, Viewer

Environment, History, Connecti Files, Plots, Packages, Help, Vir

- Environment
- History
- Files
- Plots
- Connections
- Packages
- Help
- Build
- VCS
- Viewer

- Environment
- History
- Files
- Plots
- Connections
- Packages
- Help
- Build
- VCS
- Viewer

OK Cancel Apply



Each pane contains tabs

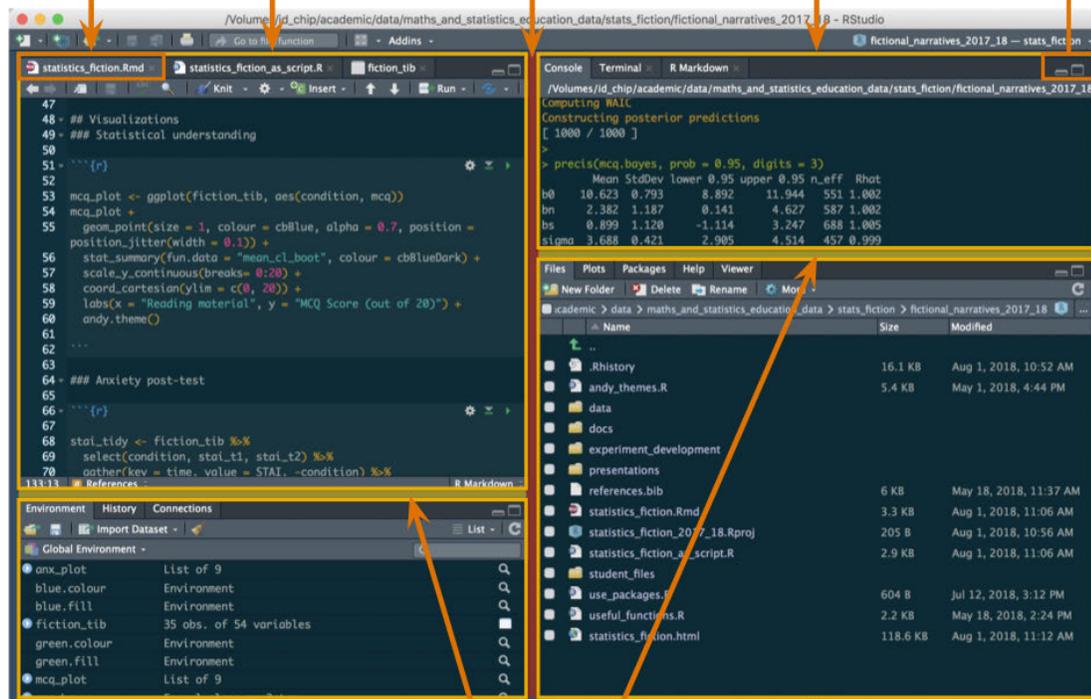
Click in the red region and drag left or right to adjust the relative widths of the left- and right-hand panes

Click these icons to minimize (left) or expand (right) a pane



Source pane

Console pane



Environment pane

Files pane

Click in the yellow regions and drag up or down to adjust the relative heights of the top and bottom panes on each side

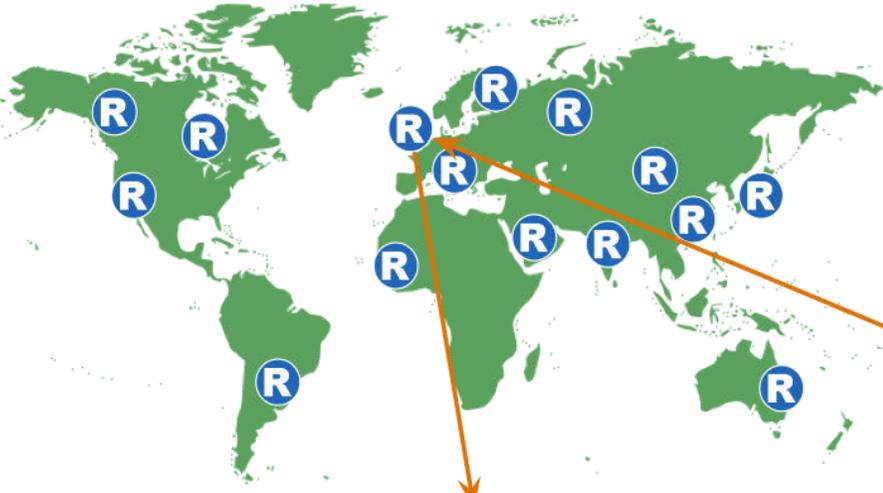


ANDY FIELD



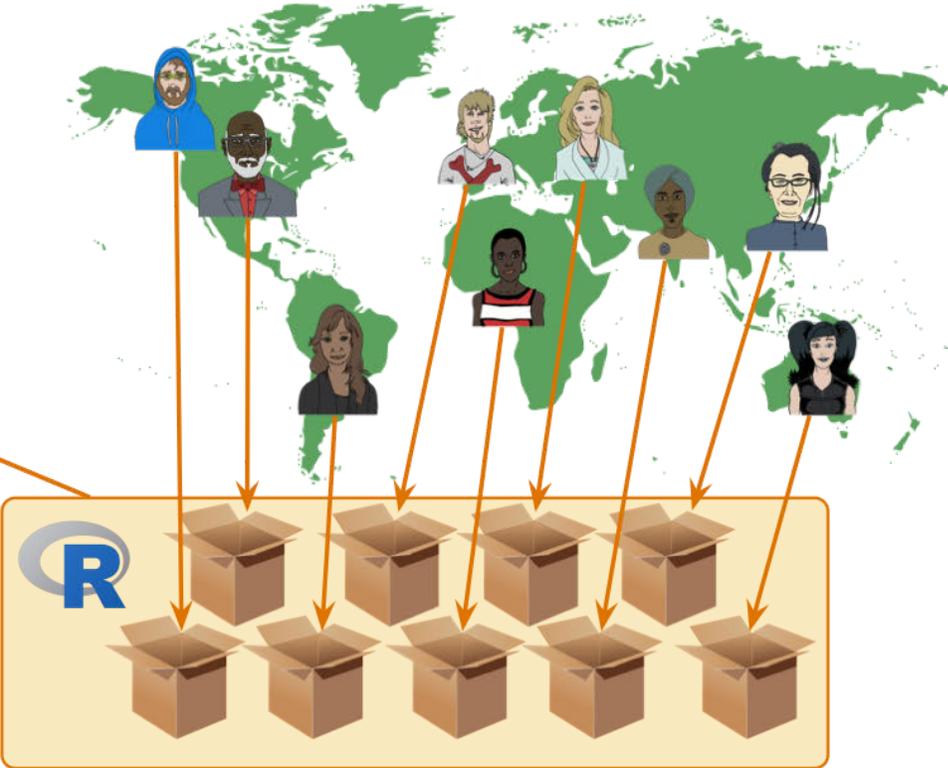
Packages

Mirrors



Your computer

Big brained contributors



CRAN

Installing and loading packages

Install the package from CRAN

- You need to install the package into R's repository of packages on your computer.
- Every time you update or re-install R you need to re-install packages to use them.
- `install.packages("package_name")`

Load the package

- To use a particular package in a current session load it from the repository.
- `library(package_name)`



Try it!

- Execute `install.packages("readr")` in the console.
- Execute `install.packages("remotes")` in the console.



Installing the `discover` package

```
remotes::install_github("profandyfield/discover")
```



Part 2: Interacting with and Studio



ANDY FIELD



Interacting with

- The console (boo!)
 - Type commands at the console prompt
 - Bad for reproducibility/your sanity
 - Great for getting help, trying things out
- Script files (meh!)
 - Great for reproducibility
 - Bad for integrating analysis into documents
-  markdown document (hooray!)
 - A document that combines marked-up text and code
 - **'Knits'** to a nicely-formatted .html, .docx or .pdf (LaTeX required) file.
 - Great for reproducibility
 - Great for integrating analysis into documents
 - **All coursework submitted as html file**



Open the file `my_first_markdown.Rmd` and click  to 'knit' the document.





R markdown document

Rendered or 'knitted' document

```

1 ---
2 title: "About me"
3 author: "Andy Field"
4 output: html_document
5 ---
6
7 ## [r setup, include=FALSE]
8 knitr::opts_chunk$set(echo = FALSE, message = FALSE)
9 ## Pets
10
11 # My favourite things
12 ## Pets
13 * I have a cocker spaniel called Milton
14 * His favourite equation is
15
16 SS $\text{text}[food]_i = \text{text}[\text{whistle}]_i + \text{text}[\text{run to owner}]_i + e_i$  SS
17
18 food_i = whistle_i + run to owner_i + e_i
19
20 ## Music
21 My favourite band is \[Iron Maiden\]\(www.ironmaiden.com\). I have seen Iron Maiden quite a few times and in several countries including England, France, Spain, Switzerland and the Netherlands. My favourite album by them is "Piece of Mind". I even like the cheesy dinosaur song that has these lyrics:
22
23 > In the time that dinosaurs walked the earth, When the land was swamp and caves were home, In an age when prize possession was fire, To search for landscapes men would roam.
24
25
26 My top 5 albums are:
27
28 1. Piece of Mind
29 2. Powerslave
30 3. The number of the beast
31 4. Seventh son of a seventh son
32 5. The final frontier (maybe)
33
34 Here's a summary of how people rate their studio albums on \[Amazon UK\]\(www.amazon.co.uk\) arranged by the mean rating. Four of my top 5 overlaps with the general consensus, the main difference being that I don't rate their debut album as highly as everyone else seems to and I seem to be a bit weird in liking *The final frontier* as much as I do.
35
36 ## [r, results = "asis"]
37 library(tidyverse)
38 amazon <- read_csv("maiden_ratings.csv")
39
40 amazon %>%
41   group_by(album) %>%
42   summarize(
43     Mean = mean(ratings) %>% round(, 3),
44     SD = sd(ratings) %>% round(, 3),
45     Median = median(ratings),
46     Range = max(ratings) - min(ratings)
47   ) %>%
48   arrange(desc(Mean)) %>%
49   knitr::kable(, format = "html") %>%
50   kableExtra::kable_styling(bootstrap_options = "striped")
51
52
53
54

```

This is a code 'chunk'. It is R code that is executed when the document is knitted. It creates a table of descriptive statistics

About me
Andy Field

My favourite things

Pets

- I have a cocker spaniel called Milton
- His favourite equation is $food_i = whistle_i + run\ to\ owner_i + e_i$

Music

My favourite band is [Iron Maiden](http://www.ironmaiden.com). I have seen **Iron Maiden** quite a few times and in several countries including England, France, Spain, Switzerland and the Netherlands. My favourite album by them is *"Piece of Mind"*. I even like the cheesy dinosaur song that has these lyrics:

In the time that dinosaurs walked the earth, When the land was swamp and caves were home, In an age when prize possession was fire, To search for landscapes men would roam.

My top 5 albums are:

- Piece of Mind
- Powerslave
- The number of the beast
- Seventh son of a seventh son
- The final frontier (maybe)

Here's a summary of how people rate their studio albums on [Amazon UK](http://www.amazon.co.uk) arranged by the mean rating. Four of my top 5 overlaps with the general consensus, the main difference being that I don't rate their debut album as highly as everyone else seems to and I seem to be a bit weird in liking *The final frontier* as much as I do.

album	Mean	SD	Median	Range
Powerslave	4.752	0.596	5.0	4
Iron Maiden	4.750	0.500	5.0	2
Piece of mind	4.708	0.622	5.0	4
The number of the beast	4.696	0.855	5.0	4
Seventh son of a seventh son	4.685	0.818	5.0	4
The book of souls	4.674	0.820	5.0	4
Brave new world	4.671	0.799	5.0	4
Killers	4.667	0.762	5.0	4
Somewhere in time	4.622	0.957	5.0	4
Dance of death	4.485	0.962	5.0	4
A matter of life and death	4.457	1.070	5.0	4
Fear of the dark	4.309	0.918	5.0	4
The X Factor	4.224	1.091	5.0	4
The final frontier	4.167	1.176	5.0	4
No prayer for the dying	3.913	1.304	4.5	4
Virtual XI	3.671	1.366	4.0	4

Inserting code chunks

Windows

- *ctrl + alt + i*
- Code > Insert Chunk

Mac

- *cmd + opt + i* (⌘ + ⌥ + i)
- Code > Insert Chunk



ANDY FIELD



Creating objects

```
metallica <- c("Lars", "James", "Jason", "Kirk")  
metallica <- metallica[metallica != "Jason"]  
metallica <- c(metallica, "Rob")
```

object

instructions

↑
Assignment
operator



ANDY FIELD



Try it

- In `my_first_markdown.Rmd` create this code chunk

```
energy_tib <- eddie_tib[c("energy", "valence")]
```

- Click  to execute the code in the code chunk
- Click  to execute all previous code chunks
- Add the second line to the code chunk and click 

```
energy_tib <- eddie_tib[c("energy", "valence")]  
energy_tib
```



Functions

```
mean(variable_name, ... arguments/options ...)  
mean(variable_name, na.rm = TRUE, trim = 0)
```

Mean

Variable list

- name
- instrument
- date_of_birth

Variable:

songs_written

Remove missing values Trim mean

OK Cancel

`mean(songs_written, na.rm = FALSE, trim = 0)`

Mean

Variable list

- name
- instrument
- date_of_birth

Variable:

songs_written

Remove missing values Trim mean

5

OK Cancel

`mean(songs_written, na.rm = TRUE, trim = 5)`

Try it

In your `my_first_markdown.Rmd` file

- Insert a code chunk, type the first command into it and execute.

```
mean_energy <- mean(energy_tib$energy)
mean_energy
```

- Click  to execute the code in the code chunk
- Insert a second code chunk, type the second command into it and execute.

```
plot(energy_tib)
```

- Click  to execute the code in the code chunk
- Click  to knit the document.



Try it

- Move one of your newly-created code chunks to before the code chunk called `subset_data`.
- Click  to knit the document.



Knitting Tips

- Code chunks are knitted **in the order they appear in the markdown document**.
 - Make sure you create objects BEFORE you try to use them.
- **Do NOT include `install.packages()` in markdown files** or the package will be installed every time you knit the document!
 - Execute `install.packages()` commands at the command line





Part 3: More Markdown



ANDY FIELD





R markdown document

Rendered or 'knitted' document

```

1 ---
2 title: "About me"
3 author: "Andy Field"
4 output: html_document
5 ---
6
7 ## [r setup, include=FALSE]
8 knitr::opts_chunk$set(echo = FALSE, message = FALSE)
9 ## Pets
10
11 # My favourite things
12
13 * I have a cocker spaniel called Milton
14 * His favourite equation is
15
16 SS\text{food}_i = \text{whistle}_i + \text{run to owner}_{i-1} + e_i SS
17
18 food_i = whistle_i + run to owner_i + e_i
19
20 ## Music
21 My favourite band is [Iron Maiden](www.ironmaiden.com). I have seen [Iron Maiden] quite a few times and in several countries including England, France, Spain, Switzerland and the Netherlands. My favourite album by them is "Piece of Mind". I even like the cheesy dinosaur song that has these lyrics:
22 > In the time that dinosaurs walked the earth, When the land was swamp and caves were home, In an age when prize possession was fire, To search for landscapes men would roam.
23
24 My top 5 albums are:
25
26 1. Piece of Mind
27 2. Powerslave
28 3. The number of the beast
29 4. Seventh son of a seventh son
30 5. The final frontier (maybe)
31
32 Here's a summary of how people rate their studio albums on [Amazon UK](www.amazon.co.uk) arranged by the mean rating. Four of my top 5 overlaps with the general consensus, the main difference being that I don't rate their debut album as highly as everyone else seems to and I seem to be a bit weird in liking [The final frontier] as much as I do.
33
34
35 ## [r, results = "asis"]
36 library(tidyverse)
37 amazon <- read_csv("maiden_ratings.csv")
38
39 amazon %>%
40   group_by(album) %>%
41   summarize(
42     Mean = mean(ratings) %>% round(, 3),
43     SD = sd(ratings) %>% round(, 3),
44     Median = median(ratings),
45     Range = max(ratings) - min(ratings)
46   ) %>%
47   arrange(desc(Mean)) %>%
48   knitr::kable(, format = "html") %>%
49   kableExtra::kable_styling(bootstrap_options = "striped")
50
51
52
53
54

```

- Use # to create a heading
- Use ## to create a subheading
- Start a line with * to create a bullet list
- You can include equations
- You can include hyperlinks
- Use ** to create bold text
- Use > to style text as a quotation
- Use numbers at the start of lines to create a numbered list
- Use * to create italic text
- This is a code 'chunk'. It is R code that is executed when the document is knitted. It creates a table of descriptive statistics

About me

Andy Field

My favourite things

Pets

- I have a cocker spaniel called Milton
- His favourite equation is $food_i = whistle_i + run\ to\ owner_{i-1} + e_i$

Music

My favourite band is [Iron Maiden](http://www.ironmaiden.com). I have seen **Iron Maiden** quite a few times and in several countries including England, France, Spain, Switzerland and the Netherlands. My favourite album by them is *Piece of Mind*. I even like the cheesy dinosaur song that has these lyrics:

In the time that dinosaurs walked the earth, When the land was swamp and caves were home, In an age when prize possession was fire, To search for landscapes men would roam.

My top 5 albums are:

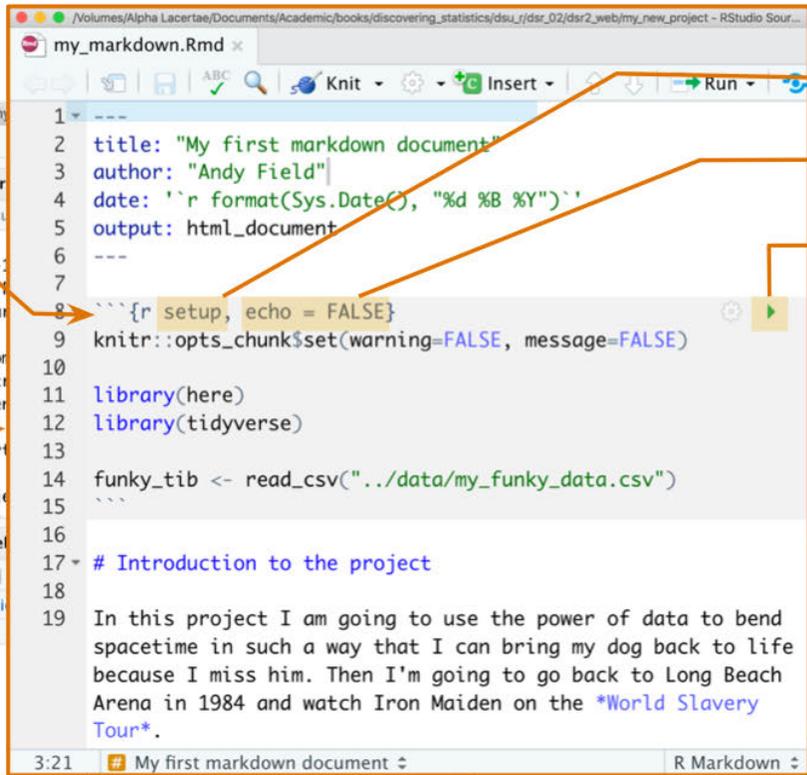
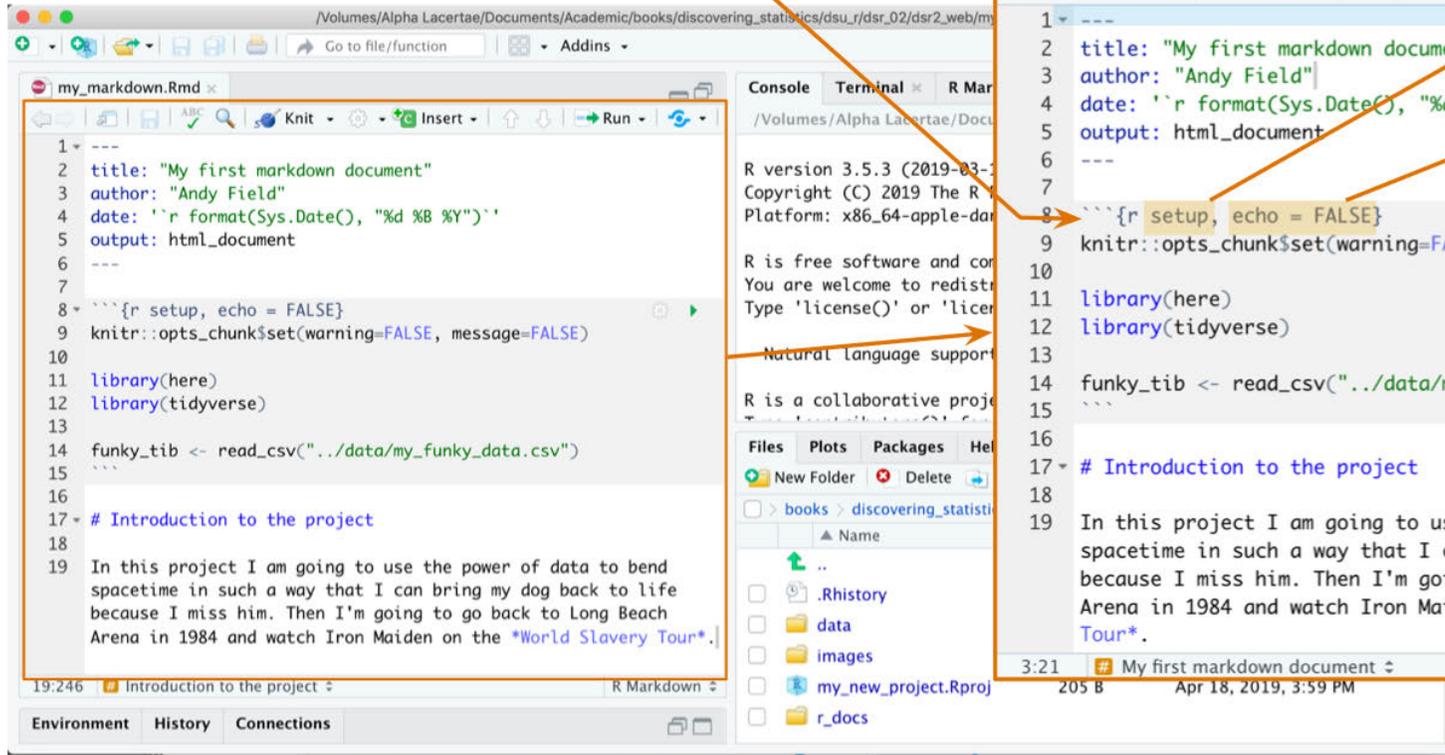
- Piece of Mind
- Powerslave
- The number of the beast
- Seventh son of a seventh son
- The final frontier (maybe)

Here's a summary of how people rate their studio albums on [Amazon UK](http://www.amazon.co.uk) arranged by the mean rating. Four of my top 5 overlaps with the general consensus, the main difference being that I don't rate their debut album as highly as everyone else seems to and I seem to be a bit weird in liking *The final frontier* as much as I do.

album	Mean	SD	Median	Range
Powerslave	4.752	0.596	5.0	4
Iron Maiden	4.750	0.500	5.0	2
Piece of mind	4.708	0.622	5.0	4
The number of the beast	4.696	0.855	5.0	4
Seventh son of a seventh son	4.685	0.818	5.0	4
The book of souls	4.674	0.820	5.0	4
Brave new world	4.671	0.799	5.0	4
Killers	4.667	0.762	5.0	4
Somewhere in time	4.622	0.957	5.0	4
Dance of death	4.485	0.962	5.0	4
A matter of life and death	4.457	1.070	5.0	4
Fear of the dark	4.309	0.918	5.0	4
The X Factor	4.224	1.091	5.0	4
The final frontier	4.167	1.176	5.0	4
No prayer for the dying	3.913	1.304	4.5	4
Virtual XI	3.671	1.366	4.0	4

This is a markdown file

This line is the chunk header



I've named this code chunk 'setup'

Options for the chunk

Click here to execute the code within the chunk

This is a code chunk. All R code goes in these chunks

This is R markdown



ANDY FIELD



Code chunk options

- **echo=FALSE**: the code will be evaluated but not reproduced in the knitted document.
- **eval=FALSE**: the code will not be evaluated.
- **include=FALSE**:  Studio will evaluate the code but neither the code nor output are displayed in the knitted document.
- **results="hide"**: the code will be evaluated and displayed in the knitted document but the output will be omitted.
- Include **fig.width=9** and **fig.height=5** to set the dimensions of an image
 - Obviously replace 9 and 5 with the values you want
- Use **results = 'asis'** when creating tables from code/output

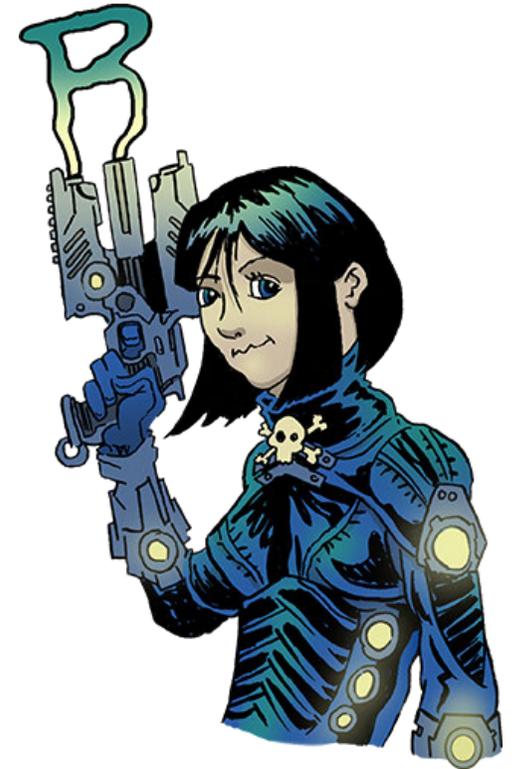


ANDY FIELD



Try it!

- Using the code chunks we have already created
- Change the code chunk headers to include `echo=FALSE`, re-knit the document and compare the result.
- Repeat this process but including `eval=FALSE`



Markdown: headers

Markdown

- # Level 1 heading
- ## Level 2 heading
- ### Level 3 heading
- #### Level 4 heading

Knitted text

- **Level 1 heading**
- **Level 2 heading**
- **Level 3 heading**
- Level 4 heading

Markdown: lists

Markdown

```
* This is the first bullet point
+ this is a sub-bullet
+ so is this
* This is the second bullet
+ This is a sub-bullet
  - A third layer of bullet madness
  - It had to be done
* This is the third main bullet
```

Knitted text

- This is the first bullet point
 - this is a sub-bullet
 - so is this
- This is the second bullet
 - This is a sub-bullet
 - A third layer of bullet madness
 - It had to be done
- and this is the third bullet



Markdown: emphasizing text

Markdown

```
*italic text*  
**bold text**  
Text^superscript^  
Text~subscript~
```

Knitted text

italic text

bold text

Text^{superscript}

Text_{subscript}



Markdown: images

```
![Caption for image](filepath_for_image)
```

```
![Figure 1: my lovely spaniel]  
(media/milton_circle_2019.png)
```



Figure 1: my lovely spaniel

Markdown: hyperlinks

```
[text_describing_link](url)
```

Markdown

The `**discover**` package is a collection of interactive tutorials for learning R. Get it from `www.discovr.rocks/discover/`

Knitted text

The `discover` package is a collection of interactive tutorials for learning R. Get it from www.discovr.rocks/discover/



Markdown: equations

Markdown

We can include the linear model in its own paragraph like this:

```
$$ Y_i = b_0 + b_1X_i + \epsilon_i $$
```

Knitted text

We can include the linear model in its own paragraph like this:

$$Y_i = b_0 + b_1X_i + \epsilon_i$$



Tables

Raw markdown

```
tap_tib
```

```
## # A tibble: 1,740 × 9
##   id      intervention efficacy did_bhv id_school time_cat intention time_num positivity
##   <chr> <chr>          <dbl>  <dbl> <chr>      <chr>      <dbl>    <dbl>      <dbl>
## 1 zy56g Group           3        0 school_1 baseline      35         0          5
## 2 xt30h Leaflet         3        1 school_1 baseline      36         0          3
## 3 st59q Group           4        0 school_1 baseline      34         0          4
## 4 vh02j Group           7        1 school_1 baseline      33         0          2
## 5 vn97y Leaflet         3        1 school_1 baseline      31         0          3
## 6 av87d Leaflet         5        1 school_1 baseline      36         0          4
## 7 nu84x Leaflet         3        0 school_1 baseline      33         0          6
## 8 zr06y Group           7        1 school_1 baseline      34         0          5
## 9 ls25l Leaflet         2        0 school_1 baseline      32         0          3
## 10 zp67x Group           5        0 school_1 baseline      35         0          2
## # ... with 1,730 more rows
```



Tables

Using `results = 'asis'` and `kable()`

```
tap_tib %>%  
  knitr::kable(caption = "Table 1: My nice table", digits = 2)
```

Table 1: My nice table

id	intervention	efficacy	did_bhv	id_school	time_cat	intention	time_num	positivity
zy56g	Group	3	0	school_1	baseline	35	0	5
xt30h	Leaflet	3	1	school_1	baseline	36	0	3
st59q	Group	4	0	school_1	baseline	34	0	4
vh02j	Group	7	1	school_1	baseline	33	0	2
vn97y	Leaflet	3	1	school_1	baseline	31	0	3
av87d	Leaflet	5	1	school_1	baseline	36	0	4

Try it!

- In your markdown file create a level 1 header that reads 'About me'
- Write a sentence or two about yourself using bold and italic somewhere
- Create a level two heading 'My favourite things'
- Within that section write a list of five of your favourite things. Include a URL to one of them.
 - (It could be books, movies, bands, songs, statisticians).



Part 4: Workflow in Studio



ANDY FIELD



R Studio project files

A file created by R Studio with the extension `.Rproj`

- Stores information about the containing folder
- Restores the previous state of the project (i.e. what documents/tabs were open)

Opening a project file sets the working directory to the folder containing the project file

- You can use relative file paths
- The project will work on any machine you care to use
- You can share your project folder with others and it'll work for them

Use project files!



Filepaths made easy (well, easier ...)

Relative paths

If you use an  Studio project, you can use relative paths

- `tap_tib <- readr::read_csv("../data/tap_parenting.csv")`

The **here** package

If you use an  Studio project, `here::here()` returns the project folder

- Text within `here()` returns the folder or file within the project folder that matches the text
- `tap_tib <- here::here("data/tap_parenting.csv") %>% readr::read_csv()`



```
library(here)  
here::here()
```

```
## [1]  
"/Volumes/alpha_lacertae/documents/academic/teaching/disc_stats_module/disc_stat_presentations"
```

```
here::here("data")
```

```
## [1]  
"/Volumes/alpha_lacertae/documents/academic/teaching/disc_stats_module/disc_stat_presentations/data"
```

```
here::here("data/tap_parenting.csv")
```

```
## [1]  
"/Volumes/alpha_lacertae/documents/academic/teaching/disc_stats_module/disc_stat_presentations/data/tap_pa
```

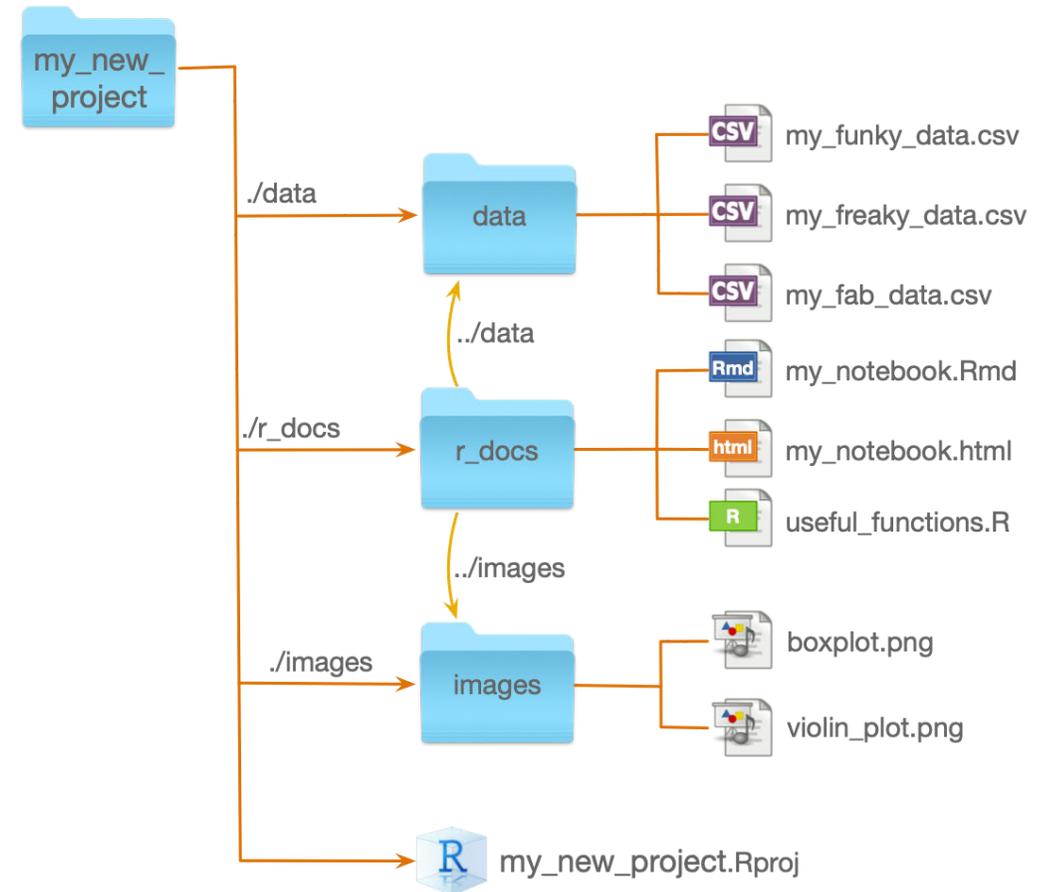


Get organized!

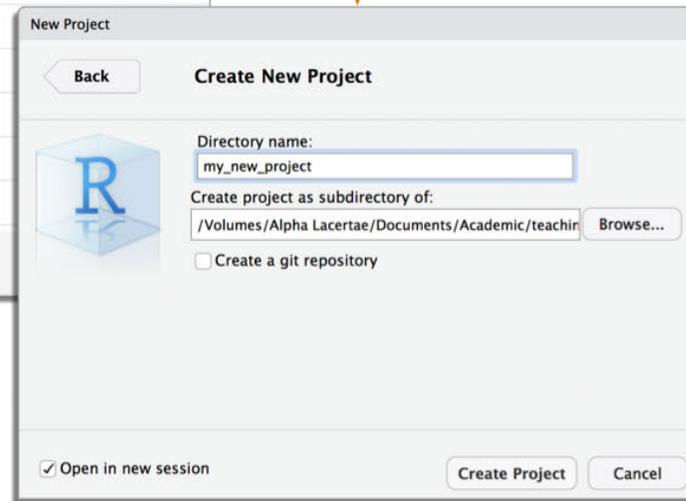
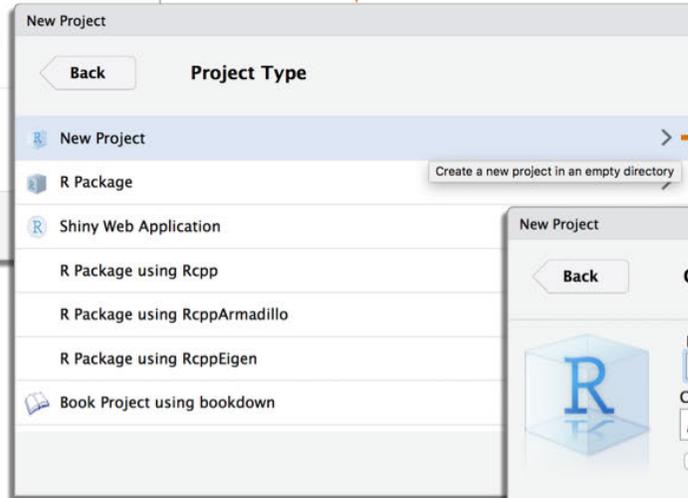
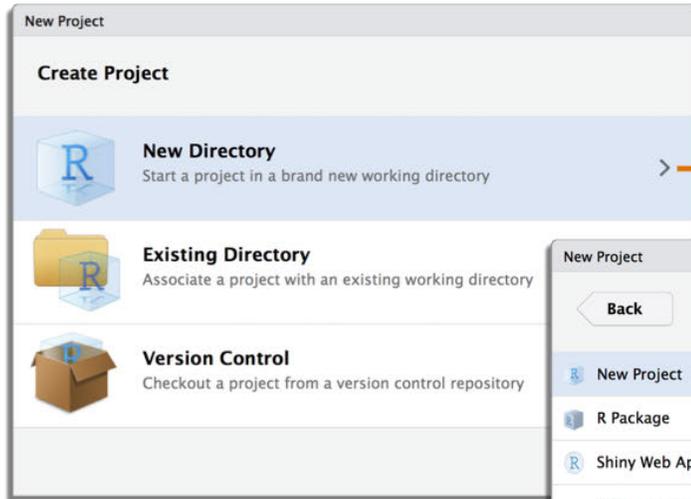
C:/Users/andyfield/Documents/my_new_project/data/my_funky_data.csv

./data/my_funky_data.csv

../data/my_funky_data.csv



Creating a project: **File > New Project**



Try it!

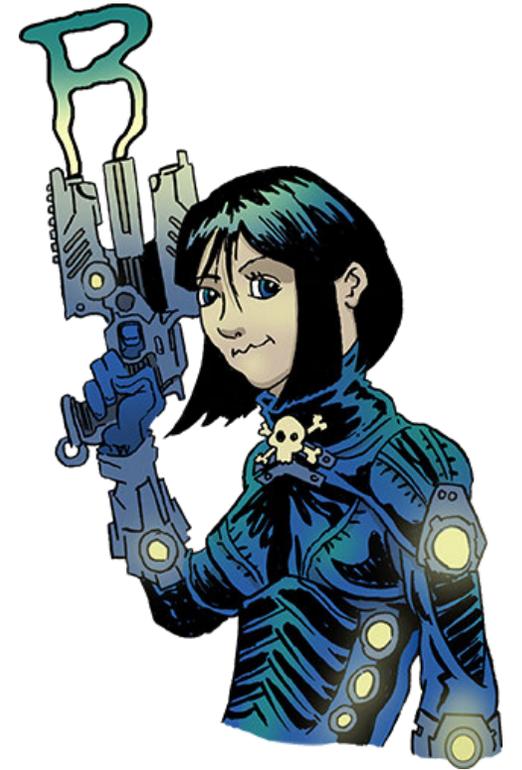
Create an  Studio project called `my_adventr` on `Sussex One Drive`

Within it create folders called

- `data`
- `r_docs`

Create a markdown file

- Use the `File > New File > R Markdown ...` menu
- Save it to the `r_docs` folder with the name `sample_tap.Rmd`



Now, let's save some data

On CANVAS go to [Files > data_files](#)

Download the sample TAP data ([tap_parenting.csv](#))

Copy the file into the data folder of your project



ANDY FIELD





Part 5: Tidyverse

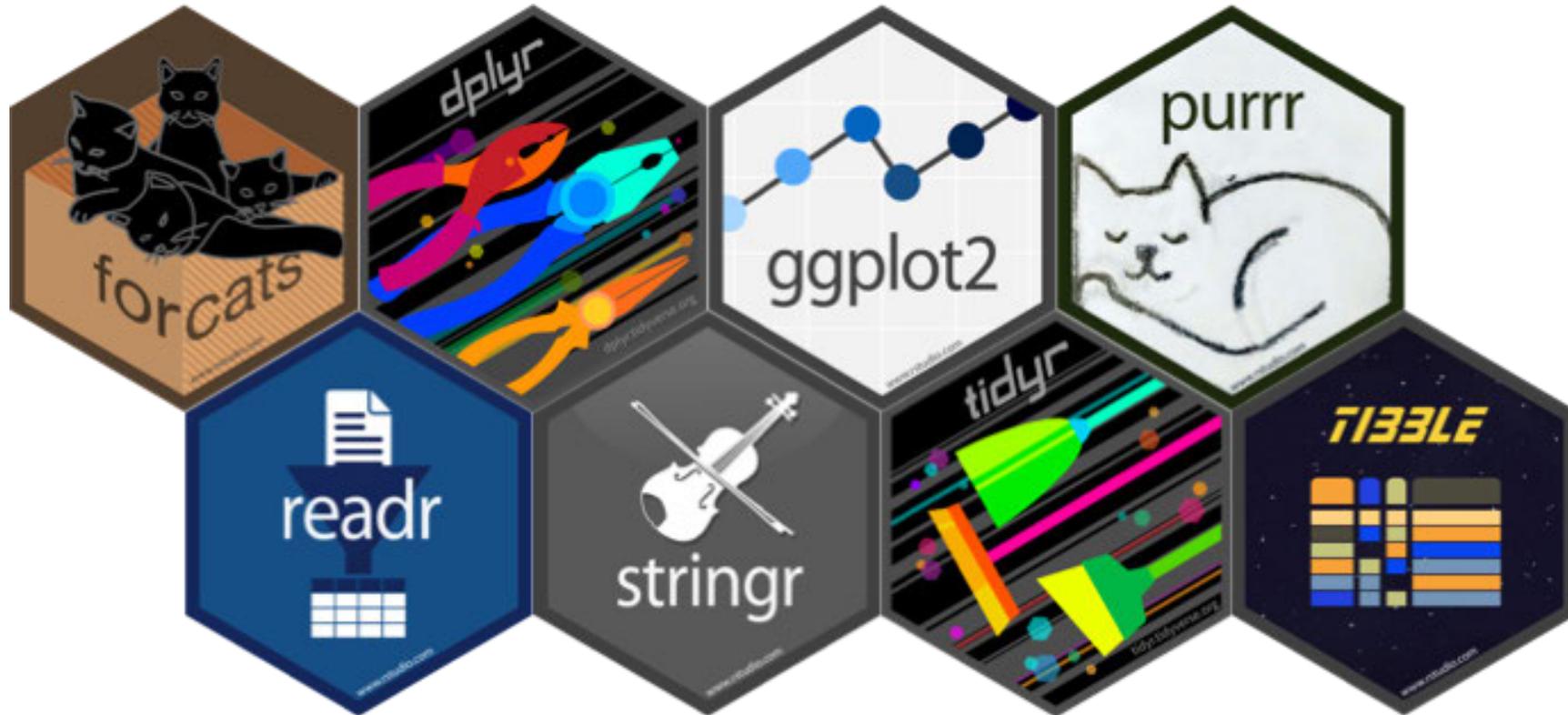


ANDY FIELD



The tidyverse

A set of packages built upon a common philosophy of data science



ANDY FIELD



The pipe operator (%>%)

```
metallica
```

```
## [1] "Lars" "James" "Kirk" "Rob"
```

Multiple commands:

```
metalli_core <- subset(metallica, metallica != "Rob")  
metalli_core <- sort(metalli_core)
```

Nested commands:

```
metalli_core <- sort(subset(metallica, metallica != "Rob"))
```

Piped commands:

```
metalli_core <- metallica %>%  
  subset(., metallica != "Rob") %>%  
  sort(.)
```

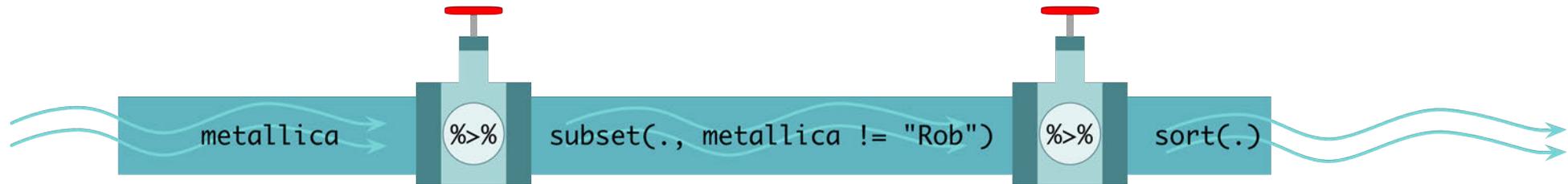


Piped commands

```
metalli_core <- metallica %>%  
  subset(., metallica != "Rob") %>%  
  sort(.)
```

```
metalli_core
```

```
## [1] "James" "Kirk" "Lars"
```



R Style (Wickham, 2014)

R is case sensitive. In code chunks use lower case wherever possible

Variable and function names should be lowercase with an underscore (_) to separate words

- Names should be concise and meaningful
- A variable representing children's anxiety levels might be named **child_anxiety**
- **scores_on_the_child_manifest_anxiety_scale** is meaningful but too long, and **ca** is concise but not meaningful

Place spaces around all operators (=, +, -, <-) to make code easier on the eye!

My practice: use suffixes that identify types of objects. Some examples

- **_tib** to denote a tibble (more them later), e.g. **anx_tib** for a tibble of data relating to anxiety)
- **_mod** to denote a model (e.g., **anx_mod** for a model in which child anxiety is predicted)
- **_lm** to denote a model built using **lm()**
- **_out** to denote output (e.g., **anx_out** contains the summary output from the above model)



Messy vs. Tidy data

Messy data (aka 'wide' data)

What IBM SPSS Statistics uses

- Each row represents a unique case/entity

Tidy data (aka 'long' data)

What many (but not all)  functions require

- Each row represents an instance of the outcome measure
- Rows code information about that 'instance'



Tidy data example

Are invisible people mischievous?

- Placed participants in an enclosed community riddled with hidden cameras
- Measured how many mischievous acts participants performed in a week
- Manipulated whether or not there was access to an invisibility cloak



Independent design

- 12 participants given an invisibility cloak
- 12 participants not given an invisibility cloak
- A random sample of $N = 7$ from each group is shown

id	cloak	mischief
Anupama	No cloak	1
Collin	No cloak	5
Darrell	No cloak	4
Kinaana	No cloak	5
Steven	No cloak	2
Tyler	No cloak	6
Vanessa	No cloak	6
Conan	Cloak	6
Devante	Cloak	7
Jerry	Cloak	5
Kathryn	Cloak	2
Sage	Cloak	3
Shajee'a	Cloak	4
Tamara	Cloak	8

Repeated measures design

- 12 participants given an invisibility cloak for one week
- During a different week they did not have an invisibility cloak
- A random subsample of $N = 7$ participants is displayed

id	cloak	mischief
Anupama	No cloak	1
Anupama	Cloak	3
Collin	No cloak	5
Collin	Cloak	6
Darrell	No cloak	4
Darrell	Cloak	5
Kinaana	No cloak	5
Kinaana	Cloak	5
Steven	No cloak	2
Steven	Cloak	4
Tyler	No cloak	6
Tyler	Cloak	5
Vanessa	No cloak	6
Vanessa	Cloak	8

```
tap_tib <- here::here("data/tap_parenting.csv") %>%
  readr::read_csv()
```

```
tap_tib
```

tap_parenting.csv

id	intervention	efficacy	did_bhv	id_school	time_cat	intention	time_num	positivity
aa54j	Group	7	1	school_7	baseline	28	0	1
aa54j	Group	7	1	school_7	1 month	46	1	5
aa54j	Group	7	1	school_7	6 months	29	6	2
ac09v	Group	4	1	school_6	baseline	54	0	5
ac09v	Group	4	1	school_6	1 month	48	1	5
ac09v	Group	4	1	school_6	6 months	59	6	5
ad17o	Leaflet	6	1	school_7	baseline	26	0	2
ad17o	Leaflet	6	1	school_7	1 month	28	1	5
ad17o	Leaflet	6	1	school_7	6 months	26	6	2
ad43a	Group	3	1	school_6	baseline	57	0	4

Part 6: Tips and good habits



ANDY FIELD



Getting help

To get help, use the `help()` function or ?

- `help(thing_you_want_help_with)`
- `?thing_you_want_help_with`

The  help files are frequently incomprehensible to mere mortals, but they are getting better!

Tip: Execute help commands at the command line

- **Do NOT include `help()` or `?` in R markdown files**

Try it!

Access the help files for the `mean()` function, by executing:

- `?mean`



Using a 'setup' code chunk

Use a setup code chunk that

- Sets global options for all of your code chunks
- Loads all of the packages you plan to use (in alphabetic order)
- Loads any data that you plan to use

```
knitr::opts_chunk$set(echo = TRUE, warning=FALSE, message=FALSE)
```

```
library(here)  
library(kableExtra)  
library(tidyverse)
```

```
tap_tib <- here::here("data/tap_parenting.csv") %>%  
  readr::read_csv()
```



Part 7: Getting the most from practical classes



ANDY FIELD



Practical classes

The practical classes are based on a package of interactive tutorials called **discover** that I wrote

- You work at your own pace
- You can work with friends/peers to support each other
- Tutors will wander around giving you one-to-one help when you need it



ANDY FIELD



The `discover` package

Installing `discover`

```
if(!require(remotes)){  
  install.packages('remotes')  
}  
remotes::install_github("profandyfield/discover")
```



ANDY FIELD



Running a tutorial

```
library(discover)
learnr::run_tutorial("name_of_tutorial", package = "discover")
learnr::run_tutorial("discover_03", package = "discover")
```

The screenshot shows the RStudio interface with the following components:

- Editor:** A file named `ais_workshop.Rmd` containing a title "An adventure in statistics: introductory workshop", author "Andy Field", and code for setting up the environment (including `knitr::opts_chunk$set(echo = TRUE)` and `use_packages("kableExtra", "tidyverse")`).
- Environment:** A panel on the right titled "Tutorial" showing three available tutorials:
 - discover: Confidence intervals** (package: `discover_03`): Description: "Confidence intervals: interactive app demonstrating what a confidence interval is, computing normal and bootstrap confidence intervals using R, adding confidence intervals to data summaries." A red arrow points to the "Start Tutorial" button.
 - discover: Visualizing data** (package: `discover_05`): Description: "Visualizing data. The ggplot2 package, boxplots, plotting means, violin plots, scatterplots, grouping by colour, grouping using facets, adjusting scales, adjusting positions." A "Start Tutorial" button is visible.
 - discover: the beast of bias** (package: `discover_06`): Description: "the beast of bias". A "Start Tutorial" button is visible.
- Console:** Shows the R version (4.1.1) and copyright information for The R Foundation for Statistical Computing.
- Files:** A file browser showing the current directory structure: `pha_lacertae > documents > academic > teaching > an_adventure_in_statistics > my_adventr`.



ANDY FIELD



Suggested workflow

Create an  Studio project called `my_adventr`

- Within it create folders called `data` and `r_docs`
- Save all of the data files for the tutorials (on Canvas) into the `data` folder
- Save the file `tutorial_template.Rmd` into your `r_docs` folder

Run a tutorial and open it in a separate window

Create a learning journal for the tutorial

- Open the file `tutorial_template.Rmd` and save it with a name related to the tutorial
- As you work through the tutorial, copy the code you've written in the tutorial into code chunks in the R markdown file
- Make notes (for example, anything you didn't understand at first, or things to help you remember what you did and why you did it). This will help with your reflective statements
- Save the markdown file for future reference, and/or knit it into an html document

Watch the video at <https://youtu.be/FhoYCsZttGc>

