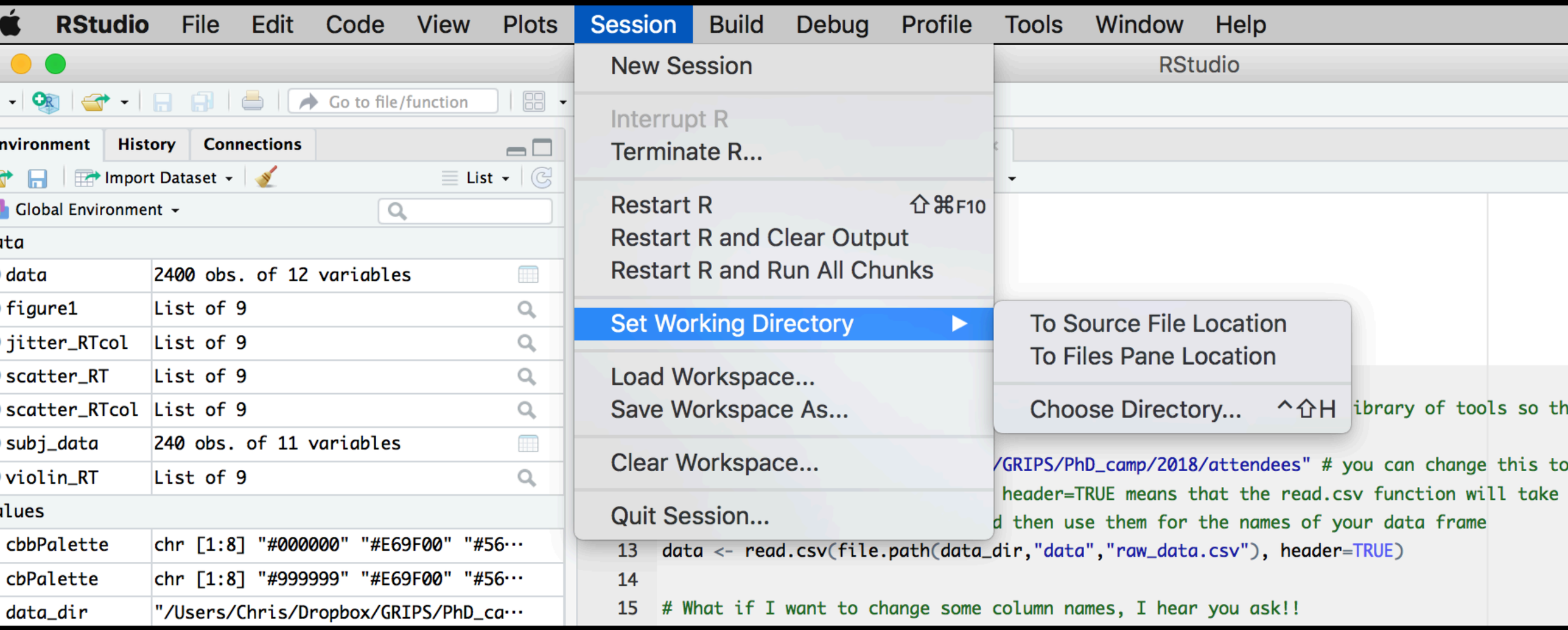
HESTER LAB 2018 - WORKSHOP

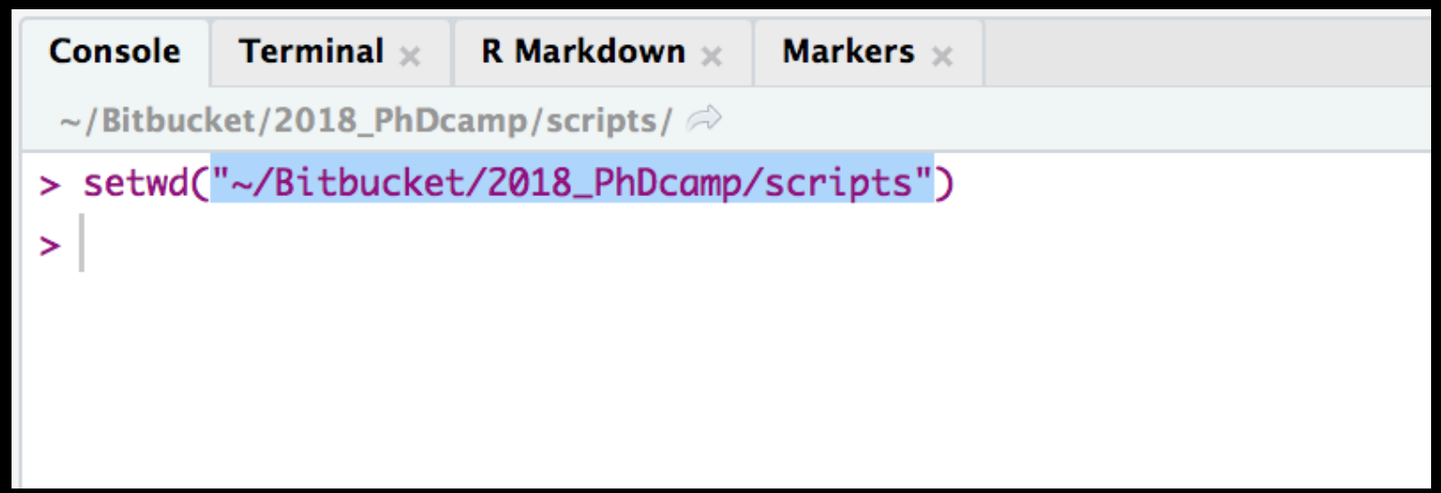
R FOR NEWBS

BY A NEWB (ME)

Select "to source file location" and R will workout where this file you opened is stored



Now this will appear in your console because R has figured out where the file is



Copy the stuff in YOUR SCRIPT in between the brackets and change this in all of the other scripts

Now this will appear in your console because R has figured out where the file is

Console | Terminal | R Markdown | Markers

~/Bitbucket/2018_PhDcamp/scripts/

```
> setwd("~/Bitbucket/2018_PhDcamp/scripts")
>
```

make_data.Rmd | ggplot_workshop.Rmd | PhDcamp2018_Rworkshop.Rmd

Knit

```
113 ```

    [1] 7
    [1] FALSE FALSE FALSE FALSE

114
115 Let's try and setup where the workshop files are on your computer. You should have all of this in a folder called 'PhDCamp_RWorkshop_20
116 First things first: let's figure out what your directory format looks like. Sometimes the way we specify directories in R and
117
118 ```{r setup_computer}
119
120 # ---------------- Setup base folder -------------------------
121 exp.basedir <- "/Users/Chris/Bitbucket/2018_PhDcamp/scripts" # base folder (you can change this, everything else is relative)
122 exp.rawdir <- file.path(exp.basedir, "data") # raw data files only
123 #exp.procdir <- file.path(exp.basedir, "processed_data")  # processed raw files
124 ```
```

Also do this for the ggplot script

Console | Terminal | R Markdown | Markers

~/Bitbucket/2018_PhDcamp/scripts/

```
> setwd("~/Bitbucket/2018_PhDcamp/scripts")
>
```

make_data.Rmd | ggplot_workshop.Rmd | PhDcamp2018_Rworkshop.Rmd

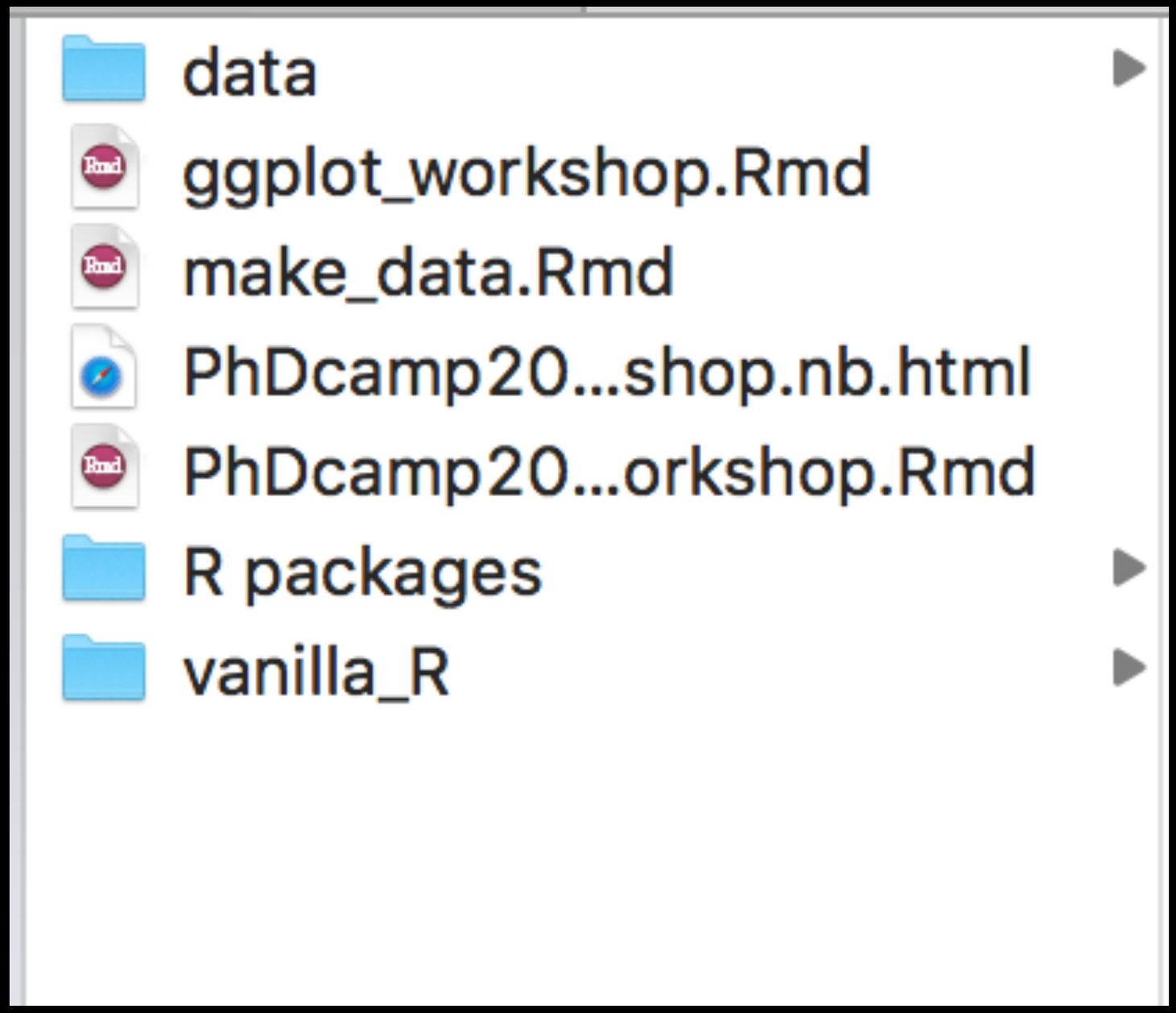
Knit | Insert

```
1  ---
2  title: "ggplot for beginners"
3  output: pdf_document
4  ---
5  Last edit: 14.05.2018 - Chris edited some errors she found post PhD workshop
6
7  ```{r load_data}
8  library(ggplot2) # we need to load the ggplot2 functions from our library of tools so that R/Rstudio can "see" them
9
10 data_dir = "/Users/Chris/Bitbucket/2018_PhDcamp/scripts" # you can change this to the place where you have stored this script
11 # Now we can read the csv file - header=TRUE means that the read.csv function will take the column
12 # headings from your csv file and then use them for the names of your data frame
13 data <- read.csv(file.path(data_dir,"data","raw_data.csv"), header=TRUE)
14
15 # What if I want to change some column names, I hear you ask!!
```

getwd()

```
> getwd()
[1] "/Users/Chris"
>
```

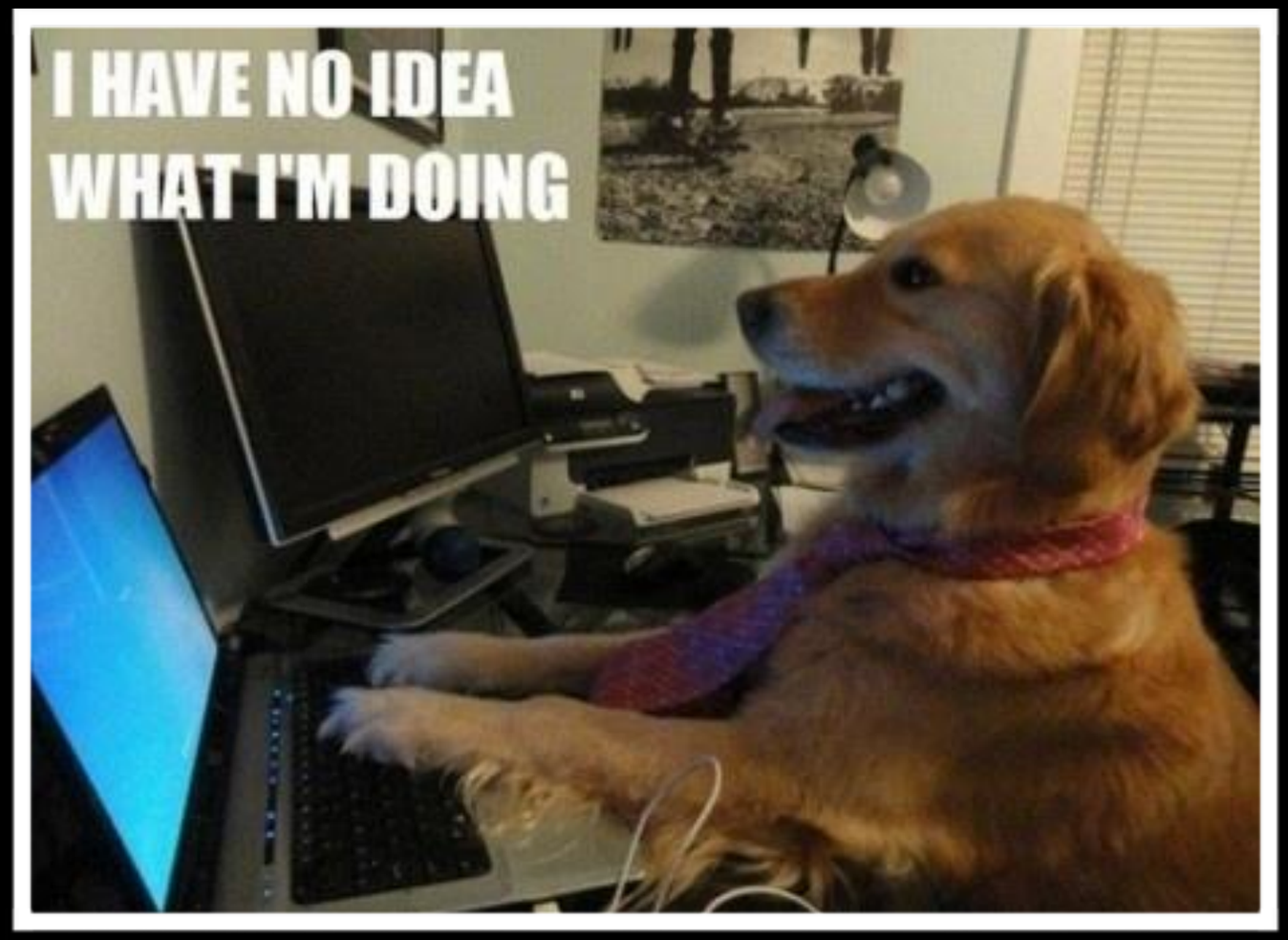
Make a folder called "attendees" somewhere you can easily find it



DISCLAIMERS

- I am not an expert in coding - I don't have a background in computer science, non analytic mathematics, or engineering where coding is routine

- The advantage (?) of this is that I'm pretty good at knowing what it's like to be like this guy



INTRODUCTION TO R

FINDING YOUR WAY AROUND

RStudio

Go to file/function | Addins | Project: (None)

Environment | History

Import Dataset | List

Global Environment

Data	
data	412 obs. of 8 variables
raw_data	412 obs. of 381 variables
Values	
exp.basedir	"~/Dropbox/HesterLab_RWorkshop_2018"
exp.procdir	"~/Dropbox/HesterLab_RWorkshop_2018/processed_…
exp.rawdir	"~/Dropbox/HesterLab_RWorkshop_2018/raw_data"

Files | Plots | Packages | Help | Viewer

New Folder | Delete | Rename | More

Home

Name	Size	Modified
#*Org Agenda(a)*#	1.5 KB	Feb 20, 2018, 1:35 AM
#client#	0 B	Mar 13, 2018, 10:42 AM
mu4e-view	1 KB	Feb 20, 2018, 3:56 PM
Org Agenda(a)	1.5 KB	Feb 20, 2018, 1:29 AM
.RData	342.3 KB	Feb 6, 2018, 6:22 PM
.Rhistory	29.9 KB	Feb 6, 2018, 6:22 PM
anaconda2		
Desktop		
Documents		
Downloads		
Dropbox		
emacs_backup.txt	43 KB	Jan 8, 2018, 7:45 PM
emacs_backup.txt~	41.5 KB	Jan 5, 2018, 5:08 PM
films.txt	32 B	Apr 3, 2018, 4:50 PM
github		
Google Drive		
howto_python_samplecode.py	3.7 KB	Mar 13, 2018, 4:07 PM
Library		
MBI_code		
Movies		
Music		
News		
PhD		
Pictures		
Public		
Zotero		

run_bayes_ALM_SDL-CVH.Rmd | tutorial1.Rmd | data.exp1 | raw_data | run_bayes_ALM.Rmd | run_jags_model.Rmd | Untitled.Rmd | logit_model.Rmd | fit_log_func.Rmd | RL_learning.Rmd

Preview | Insert | Run

```
41  # todo explain how libraries work
42  library("plyr", "ggplot2")
43
44  # ------------------------------------------------------------------------
45  #                          Setup directories
46  # ------------------------------------------------------------------------
47  # ---------------- Setup base folder -------------------------
48  exp.basedir <- "~/Dropbox/HesterLab_RWorkshop_2018" # base folder (you can change this, everything else is relative)
49  exp.rawdir <- file.path(exp.basedir, "raw_data") # raw data files only
50  exp.procdir <- file.path(exp.basedir, "processed_data")  # processed raw files
51
52  # Read Oliver's data
53  raw_data <- read.csv(file.path(exp.rawdir,'oliver_honoursdata2.csv'), header = TRUE, quote = "", sep = ",", na.strings = "NaN")
54
55  #
56  head(raw_data)
57
58  colnames(raw_data)
59
60  # Create a new dataset with nicely named variables and only what we need
61  data <- data.frame(
62    cbind(
63      CLbinge = raw_data$C_L.Dich.B.NB,
64      age = raw_data$Age,
65      dob = raw_data$dob,
66      ageBand = raw_data$ageBand,
67      gender = raw_data$gender,
68      height = raw_data$height,
69      weight = raw_data$weight,
70      firstLang = raw_data$First.Language,
71      hand = raw_data$Handedness,
72      RT.go = raw_data$Av.Go.RT,
73      RT.go.sd = raw_data$Go.RT.StDev,
74      Av.SSRT = raw_data$Av.SSRT,
75      stop.acc = raw_data$Stop.Accuracy
76    )
77  )
78
```

Title
(chunk 1)
(chunk 2)
(Untitled)
(Untitled)
(Untitled)
(Untitled)
(Untitled)
(Untitled)
Setup base folder
check_distrib
filter_data
group_by
basic_plotting
run_tests

65:24 | Setup base folder | R Markdown

Console | Markers

```
+     dob = raw_data$dob,
+     ageBand = raw_data$ageBand,
+     gender = raw_data$gender,
+     height = raw_data$height,
+     weight = raw_data$weight,
+     firstLang = raw_data$First.Language,
+     hand = raw_data$Handedness,
+     RT.go = raw_data$Av.Go.RT,
+     RT.go.sd = raw_data$Go.RT.StDev,
+     Av.SSRT = raw_data$Av.SSRT,
+     stop.acc = raw_data$Stop.Accuracy
+   )
+ )
>
>
```

FINDING YOUR WAY AROUND

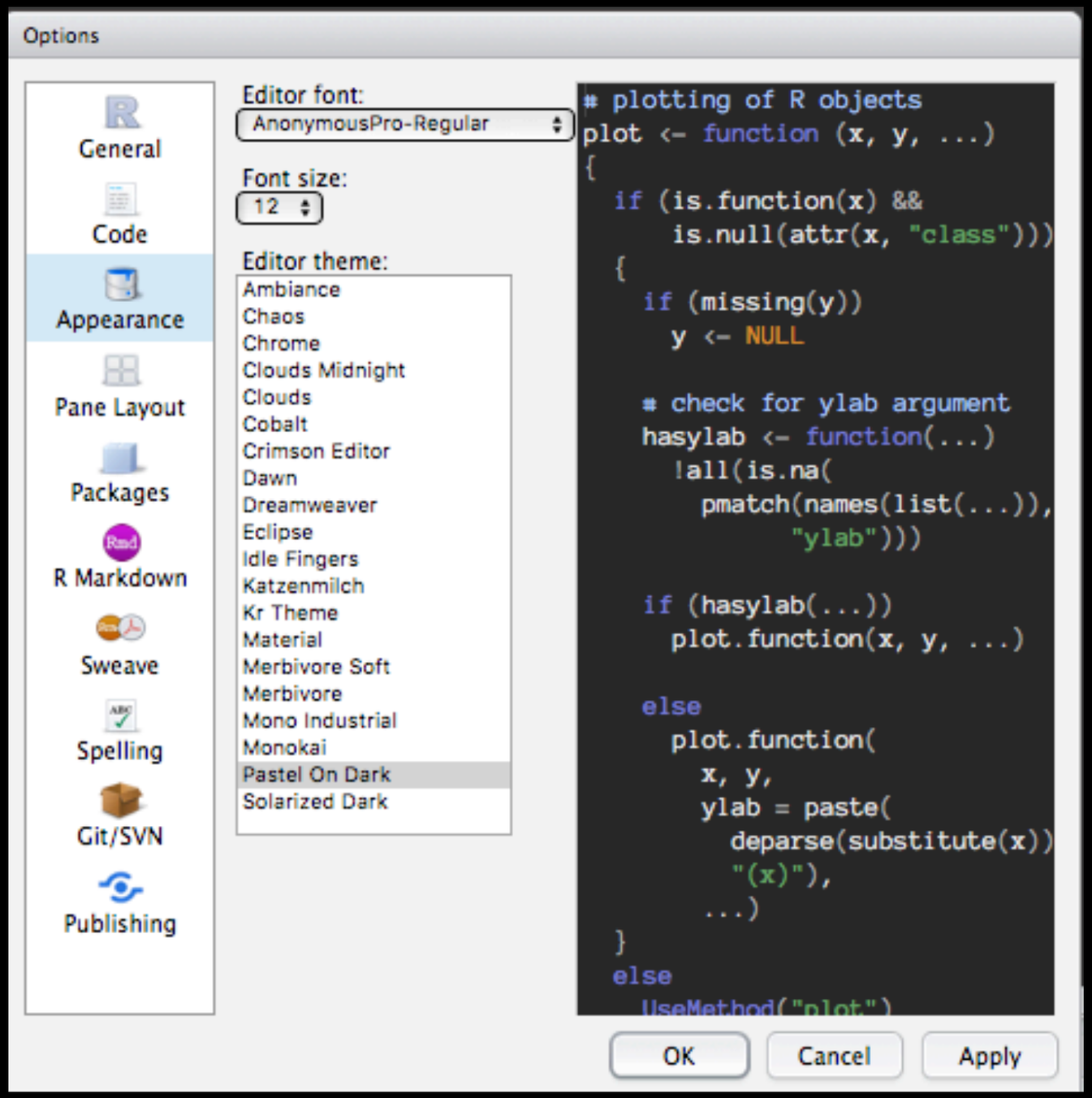
Changing settings

OSX (Mac)

- Studio > Preferences

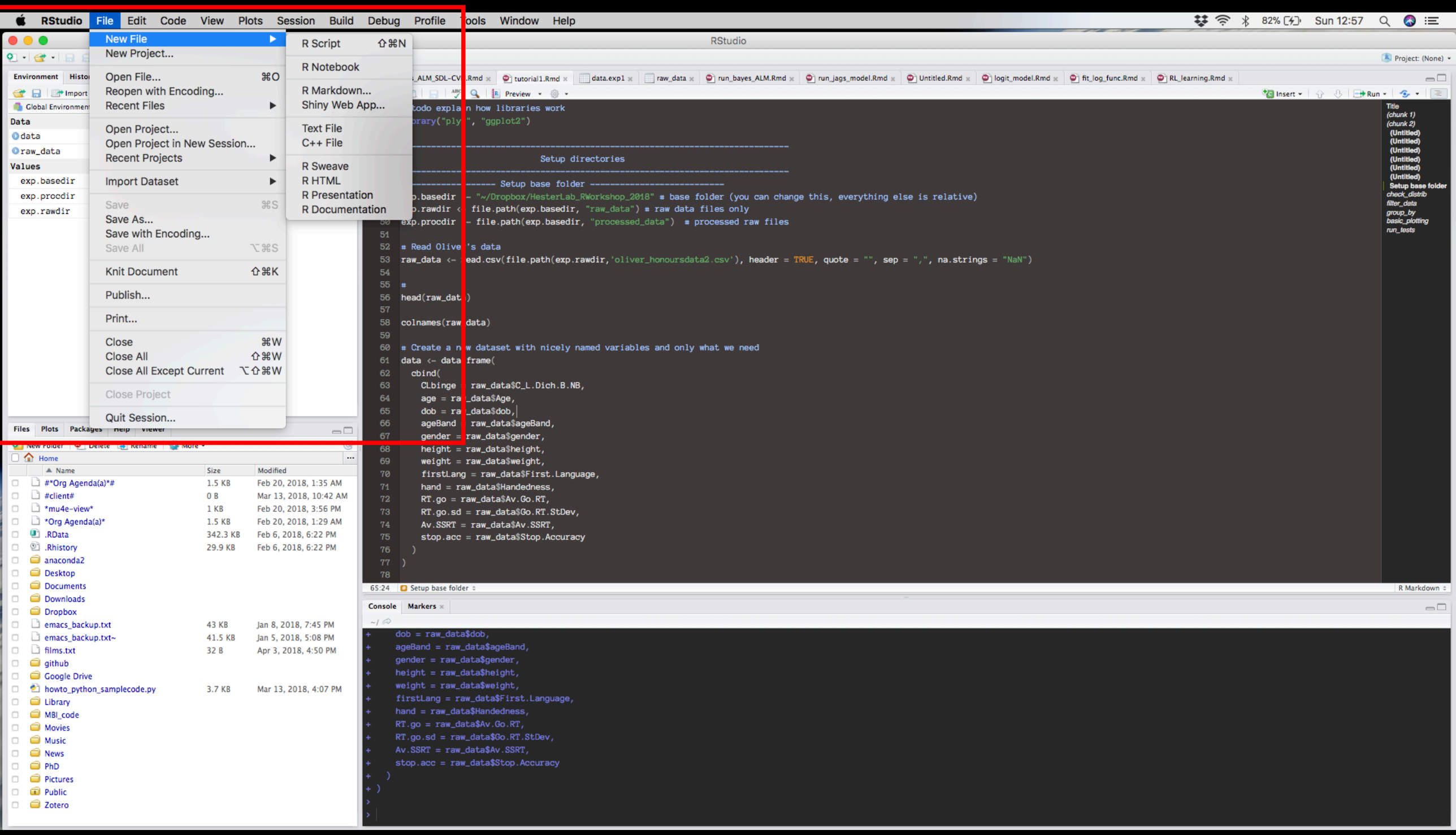
Windows

- Tools (?) > Options



FINDING YOUR WAY AROUND

Opening a file



FINDING YOUR WAY AROUND

The editor

RStudio

Environment | History

Import Dataset | List

Global Environment

Data	
data	412 obs. of 8 variables
raw_data	412 obs. of 381 variables
Values	
exp.basedir	"~/Dropbox/HesterLab_RWorkshop_2018"
exp.procdir	"~/Dropbox/HesterLab_RWorkshop_2018/processed_…
exp.rawdir	"~/Dropbox/HesterLab_RWorkshop_2018/raw_data"

Files | Plots | Packages | Help | Viewer

New Folder | Delete | Rename | More

Home

Name	Size	Modified
#*Org Agenda(a)*#	1.5 KB	Feb 20, 2018, 1:35 AM
#client#	0 B	Mar 13, 2018, 10:42 AM
mu4e-view	1 KB	Feb 20, 2018, 3:56 PM
Org Agenda(a)	1.5 KB	Feb 20, 2018, 1:29 AM
.RData	342.3 KB	Feb 6, 2018, 6:22 PM
.Rhistory	29.9 KB	Feb 6, 2018, 6:22 PM
anaconda2		
Desktop		
Documents		
Downloads		
Dropbox		
emacs_backup.txt	43 KB	Jan 8, 2018, 7:45 PM
emacs_backup.txt~	41.5 KB	Jan 5, 2018, 5:08 PM
films.txt	32 B	Apr 3, 2018, 4:50 PM
github		
Google Drive		
howto_python_samplecode.py	3.7 KB	Mar 13, 2018, 4:07 PM
Library		
MBI_code		
Movies		
Music		
News		
PhD		
Pictures		
Public		
Zotero		

run_bayes_ALM_SDL-CVH.Rmd | tutorial1.Rmd | data.exp1 | raw_data | run_bayes_ALM.Rmd | run_jags_model.Rmd | Untitled.Rmd | logit_model.Rmd | fit_log_func.Rmd | RL_learning.Rmd

Preview | Insert | Run

```
41  # todo explain how libraries work
42  library("plyr", "ggplot2")
43
44  # ------------------------------------------------------------------------
45  #                          Setup directories
46  # ------------------------------------------------------------------------
47  # ---------------- Setup base folder -------------------------
48  exp.basedir <- "~/Dropbox/HesterLab_RWorkshop_2018" # base folder (you can change this, everything else is relative)
49  exp.rawdir <- file.path(exp.basedir, "raw_data") # raw data files only
50  exp.procdir <- file.path(exp.basedir, "processed_data")  # processed raw files
51
52  # Read Oliver's data
53  raw_data <- read.csv(file.path(exp.rawdir,'oliver_honoursdata2.csv'), header = TRUE, quote = "", sep = ",", na.strings = "NaN")
54
55  #
56  head(raw_data)
57
58  colnames(raw_data)
59
60  # Create a new dataset with nicely named variables and only what we need
61  data <- data.frame(
62    cbind(
63      CLbinge = raw_data$C_L.Dich.B.NB,
64      age = raw_data$Age,
65      dob = raw_data$dob,
66      ageBand = raw_data$ageBand,
67      gender = raw_data$gender,
68      height = raw_data$height,
69      weight = raw_data$weight,
70      firstLang = raw_data$First.Language,
71      hand = raw_data$Handedness,
72      RT.go = raw_data$Av.Go.RT,
73      RT.go.sd = raw_data$Go.RT.StDev,
74      Av.SSRT = raw_data$Av.SSRT,
75      stop.acc = raw_data$Stop.Accuracy
76    )
77  )
78
```

65:24 | Setup base folder | R Markdown

Title
(chunk 1)
(chunk 2)
(Untitled)
(Untitled)
(Untitled)
(Untitled)
(Untitled)
(Untitled)
Setup base folder
check_distrib
filter_data
group_by
basic_plotting
run_tests

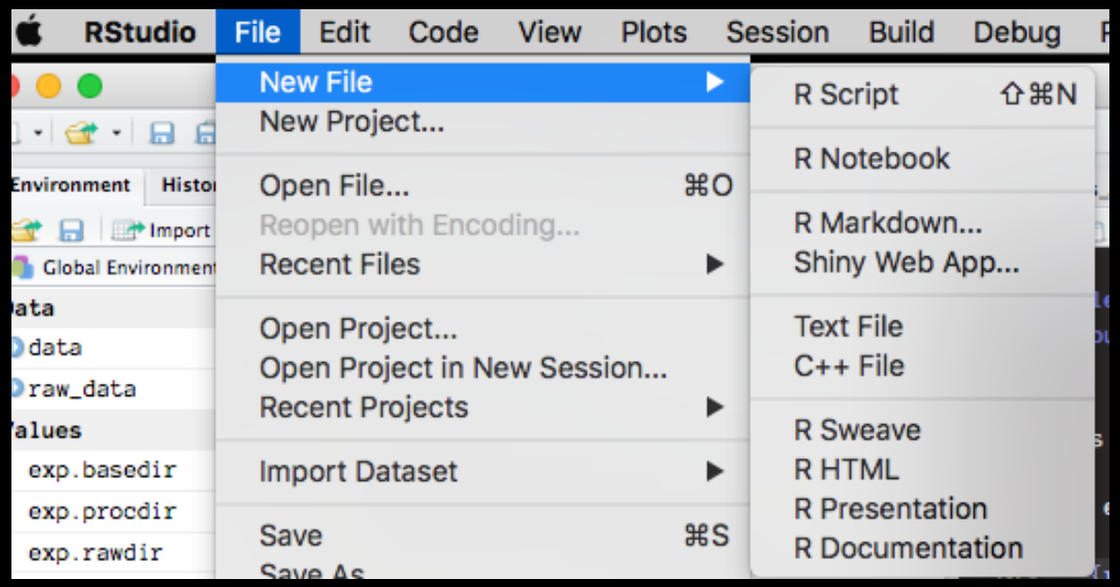
We can store instructions on how to process the data in documents called "scripts" - this stuff on the left is saved as a file called a [dot]Rmd file (stands for R Markdown) script

Console | Markers

```
+     dob = raw_data$dob,
+     ageBand = raw_data$ageBand,
+     gender = raw_data$gender,
+     height = raw_data$height,
+     weight = raw_data$weight,
+     firstLang = raw_data$First.Language,
+     hand = raw_data$Handedness,
+     RT.go = raw_data$Av.Go.RT,
+     RT.go.sd = raw_data$Go.RT.StDev,
+     Av.SSRT = raw_data$Av.SSRT,
+     stop.acc = raw_data$Stop.Accuracy
+   )
+ )
>
>
```

FINDING YOUR WAY AROUND

Types of scripts you can use



R script

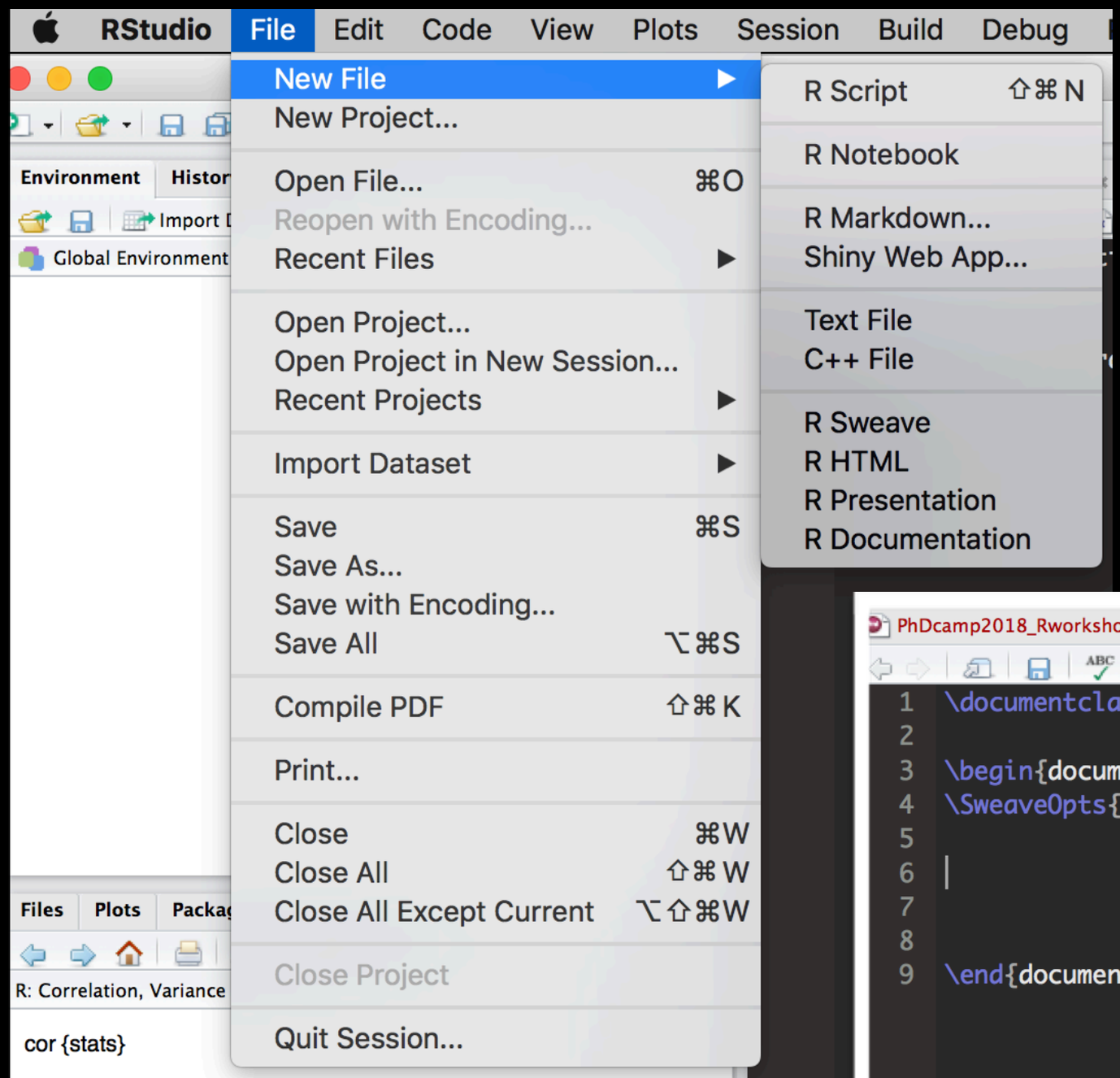
- ✓ Basic (depending on whether you think being "basic" = good)
- ✗ Does not display plots or the code result in a script
- ✗ Cannot run code in sections unless you highlight and select sections of code
- ✓ Not great for beginners
- ✓ Future proof

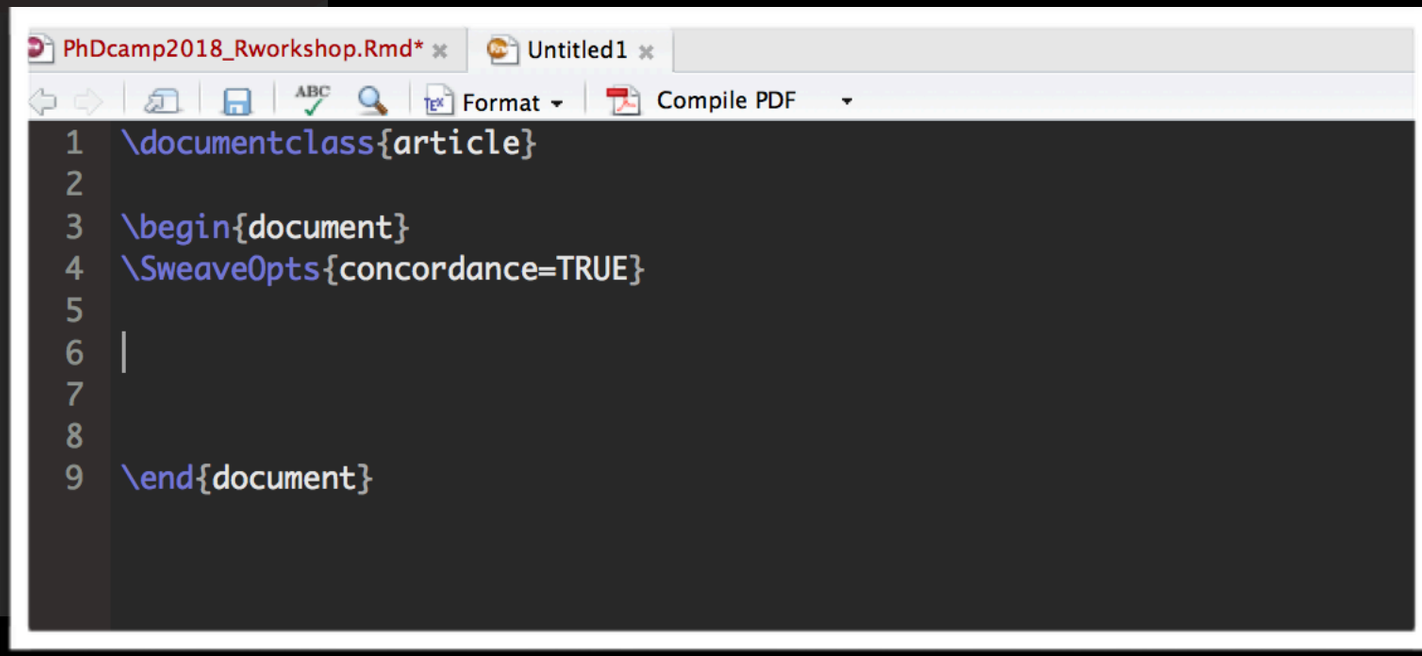
R Markdown

- ✓ Allows you to run sections of code, formatted in different ways
- ✓ Displays plots with code
- ✓ Can run in sections easily
- ✓ Can convert to pdf / doc / html *easily*
- ✓ Good for newbies to coding
- ✓ Write papers with analysis directly in it *easily*
- ✗ Not necessarily future proof

WRITING PAPERS IN R

Sweave allows you to write papers in RStudio using LaTeX (a coding alternative to word)





R Markdown Reference Guide

Learn more about R Markdown at rmarkdown.rstudio.com

Learn more about Interactive Docs at shiny.rstudio.com/articles

Contents:

1. **Markdown Syntax**
2. Knitr chunk options
3. Pandoc options

Syntax

```
Plain text

End a line with two spaces
to start a new paragraph.

*italics* and _italics_

**bold** and __bold__

superscript^2^

~~strikethrough~~

[link](www.rstudio.com)

# Header 1

## Header 2

### Header 3

#### Header 4

##### Header 5

###### Header 6
```

Becomes

Plain text

End a line with two spaces to start a new paragraph.

italics and *italics*

bold and **bold**

superscript^2

~~strikethrough~~

link

Header 1

Header 2

Header 3

Header 4

Header 5

Header 6

FINDING YOUR WAY AROUND

The console

When you run a script, you can see the output here. You can also use it like a calculator, try it!

First, let's make some data

- *Experiment:*
  - Stimuli – faces and houses
  - Trials – 280
  - Difficulty (noise) – 1 (easy), 2 (moderate), 3 (hard)
  - Decision – did I see face (1) or house (2)
  - Confidence – rate how sure you are that you saw the face or house from 1 to 5
  - Reaction time – RT_day1, RT_day 2

RESOURCES

BASICS

- https://software-carpentry.org/lessons/
  - The RezBaz at UniMelb run regular workshops: http://melbourne.resbaz.edu.au

- An introduction to statistics in R http://compcogscisydney.org/psyr/
- Associated textbook: http://compcogscisydney.org/learning-statistics-with-r/

- Sean Murphy - https://github.com/seanchrismurphy/A-Psychologists-Guide-to-R

ADVANCED (e.g. hierarchical models, bayesian linear and hierarchical models)

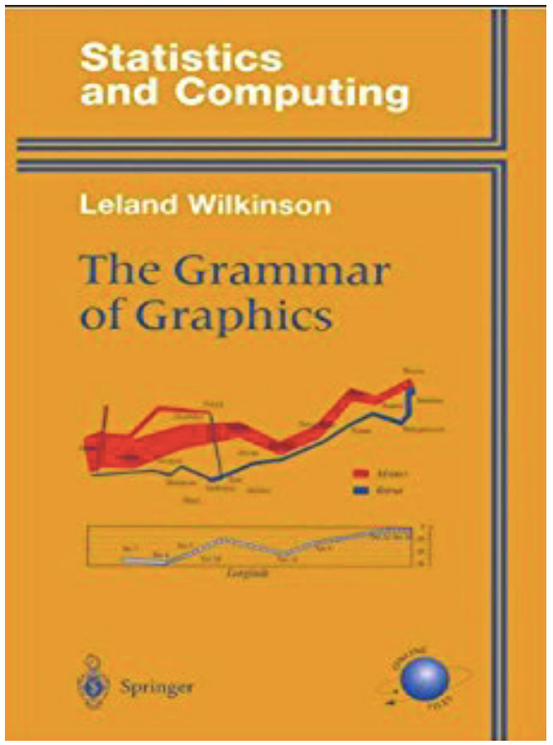
- Statistical rethinking: http://xcelab.net/rm/statistical-rethinking/
- This book has youtube videos: https://www.youtube.com/channel/UCNJK6_DZvcMqNSzQdEkzvzA/videos

Part 2
ggplot2 workshop

Aims of this workshop

- Understanding how to load data from a csv or excel file into R studio
- Understand what a data frame looks like and how it can be used for plotting
- Understand how to make basic plots
- Learn how to save your plots

Some history



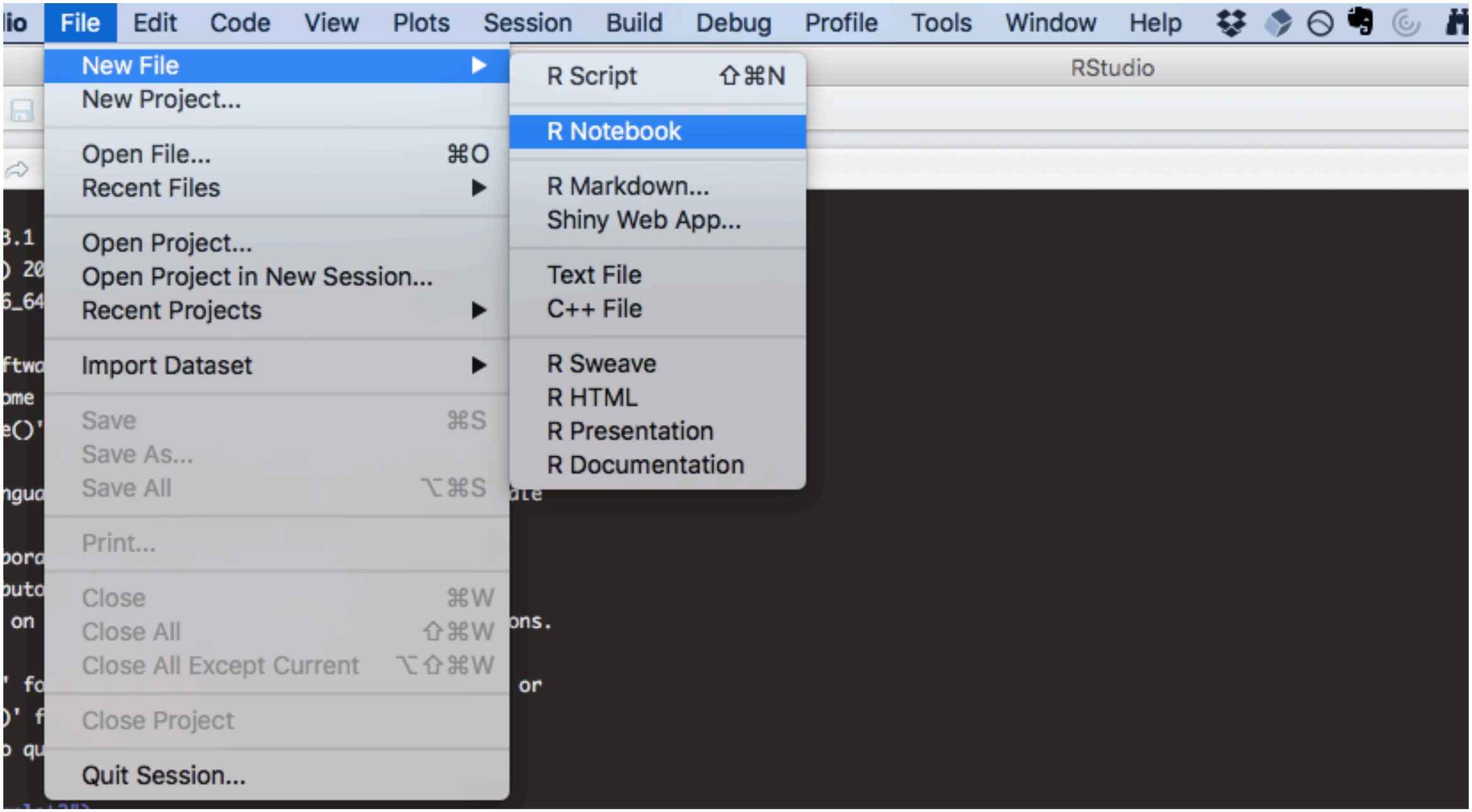
- ggplot = grammar of graphics plot
  - Bertin (1983)
  - Based on a book called *The Grammar of Graphics* by Lenland Wilkinson (SYSTAT, SPSS, Tableau, H_2O)
  - "Good grammar" – what are the ingredients I need for a plot ?
    - Data
    - Axes
    - Stuff I do with the data (dots, lines, bars)
  - Inspired Hadley Wickham to develop 'ggplot2' package for R

Good grammar = how to build a plot in a logical way

Why use ggplot?

- Data set
- A set of "geometries" – i.e. how we represent our data – scatter, bar, histogram etc
- A set of "aesthetics" – i.e. things we do to make our data more informative – change size, shape etc. of dots, bars etc.

Open a notebook file



We can change the type of output

```
---
title: "R Notebook"
output: pdf_notebook
---

This is an [R Markdown](http://rmarkdown.rstudio.com) Notebook. When you execute code within the notebook, the results appear beneath the code.

Try executing this chunk by clicking the *Run* button within the chunk or by placing your cursor inside it and pressing *Cmd+Shift+Enter*.

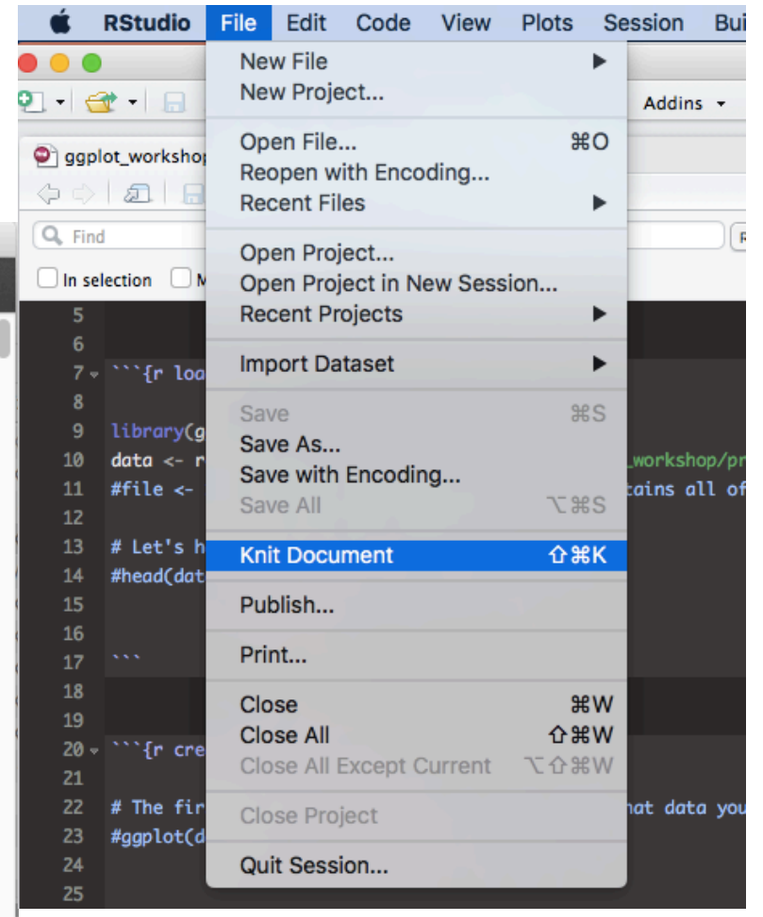
```{r}
plot(cars)
```

Add a new chunk by clicking the *Insert Chunk* button on the toolbar or by pressing *Cmd+Option+I*.

When you save the notebook, an HTML file containing the code and output will be saved alongside it (click the *Preview* button or press *Cmd+Shift+K* to preview the HTML file).
```

Knit it good

RStudio: View PDF
Page: 1 of 25
Automatic Zoom
ggplot for beginners
library(ggplot2)
Warning: package 'ggplot2' was built under R version 3.3.2
data <- read.csv("~/Dropbox/GRIPS/PhD_camp/ggplot_workshop/processed_data.csv", header=TRUE)
#file <- Sys.glob("*data.csv")[1] # this file contains all of raw data for all subjects in one file
Let's have a quick look at our data
#head(data)
The first thing you need to do is tell ggplot what data you want to plot
#ggplot(data)
Step1 - Tell ggplot what bits of the data you want to plot ggplot data x = trial, y = RT_daty1
Step 2 - Tell ggplot what you want to display it as points
Step 3 - Let's try adding some colour
data[,'conf'] <- as.factor(data[,'conf'])
# geom_point inherits everything from ggplot(data) part
# What if I would like a bar plot?
# Imagine I want to know what the distributions of responses according to confidence ratings look like
# ggplot(data, aes(x = resp)) + geom_bar(aes(conf))
# We can also change how thin or thick the bars are - the thinness or thickness
# tells us how many data points we count (get frequencies) over. These are called "bins"
# Below we can change the bin size to different values using 'binwidth' inside the geom_histogram argum
ggplot(data, aes(x = RT_day1)) + geom_histogram(binwidth = 0.2)


RStudio
File
Edit
Code
View
Plots
Session
New File
New Project...
Open File... ⌘O
Reopen with Encoding...
Recent Files
Open Project...
Open Project in New Session...
Recent Projects
Import Dataset
Save ⌘S
Save As...
Save with Encoding...
Save All ⌥⌘S
Knit Document ⇧⌘K
Publish...
Print...
Close ⌘W
Close All ⇧⌘W
Close All Except Current ⌥⇧⌘W
Close Project
Quit Session...
Addins
In selection

Things to be aware of (before you get your data)

- Short vs long format
- ggplot likes data frames

```
# Let's have a quick look at our data
head(data)
```

```
```

	trial <int>	stim <int>	trials <int>	difficulty <int>	resp <int>	corr <int>	conf <int>	RT <dbl>
1	1	2	1	1	2	1	4	0.2017097
2	2	2	2	2	2	1	3	0.9322453
3	3	1	3	3	0	0	1	1.0279934
4	4	1	4	2	1	1	4	0.7351871
5	5	2	5	1	2	1	4	0.2370090
6	6	1	6	1	1	1	4	0.2102735

6 rows

ggplot uses layers

1. The basics - layers

ggplot uses layers – keep adding until you like plot

statistics - regression line, binning, confidence intervals (garnish)

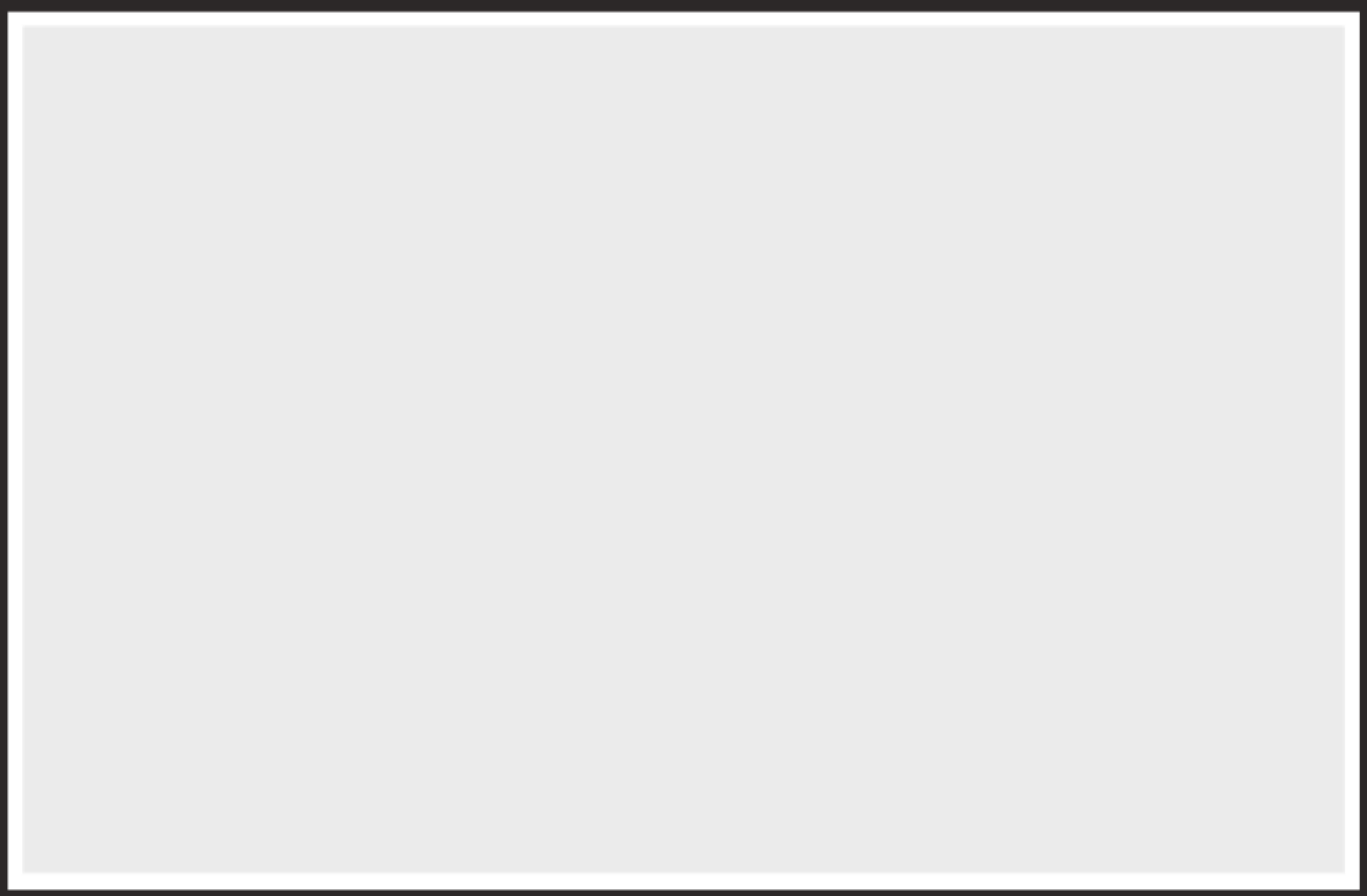
asethetics - x, y, size, shape, colour (colours of icing, shape of cake)

geometries points, line, bar, text (layers of cake) – *these also have asthetics*

data (cake)

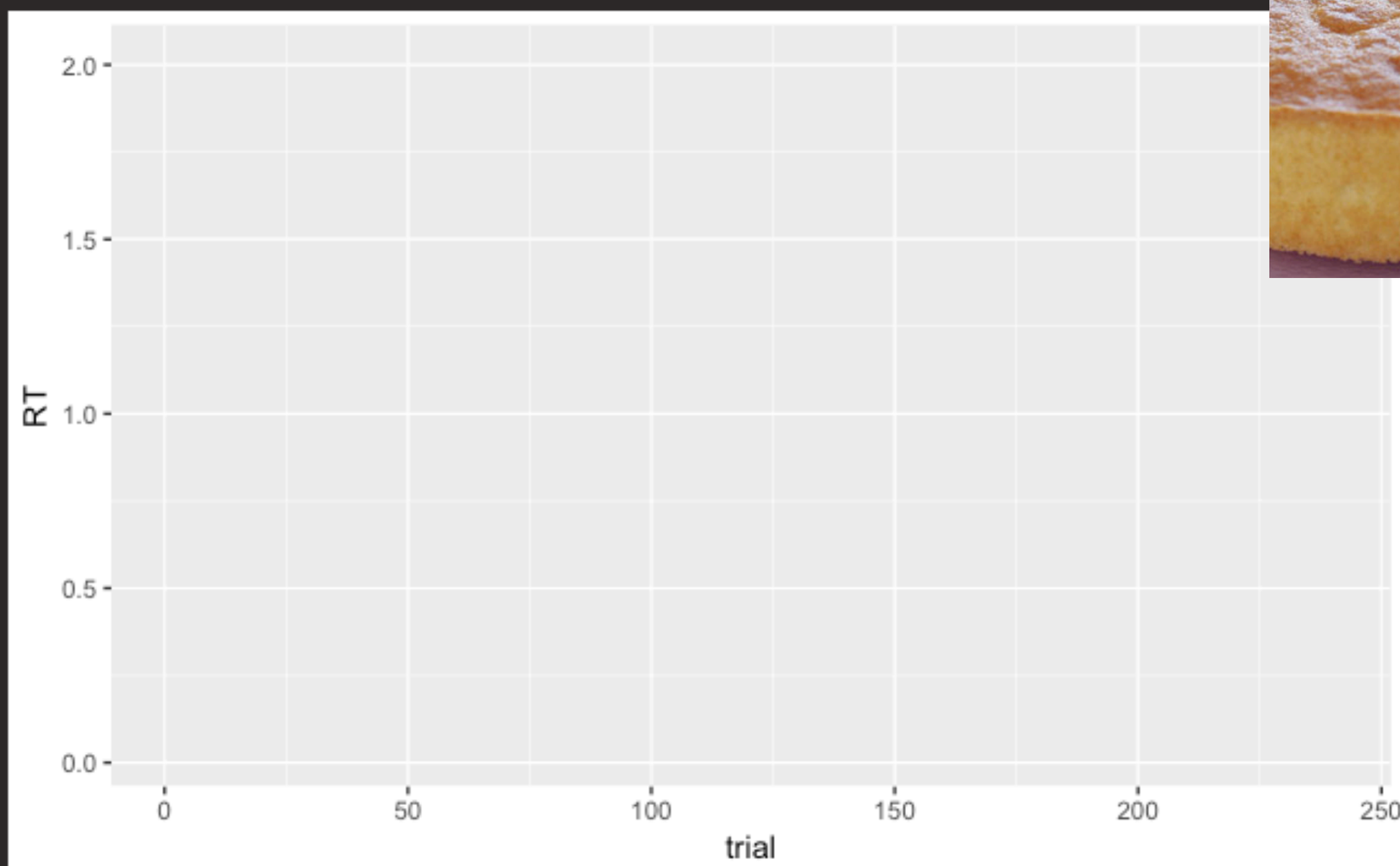
Let’s start with the ingredients for the cake (load data)

```{r create_plot_space}

# The first thing you need to do is tell ggplot what data you want to plot
ggplot(data)

```



Let’s start with the data/cake (what kind of cake?)

```{r}

# Step1 - Tell ggplot what bits of the data you want to plot
ggplot(data, aes(x = trial, y = RT))

```



Aesthetic to data mapping (aes)

statistics - regression line, binning, confidence intervals (garnish)

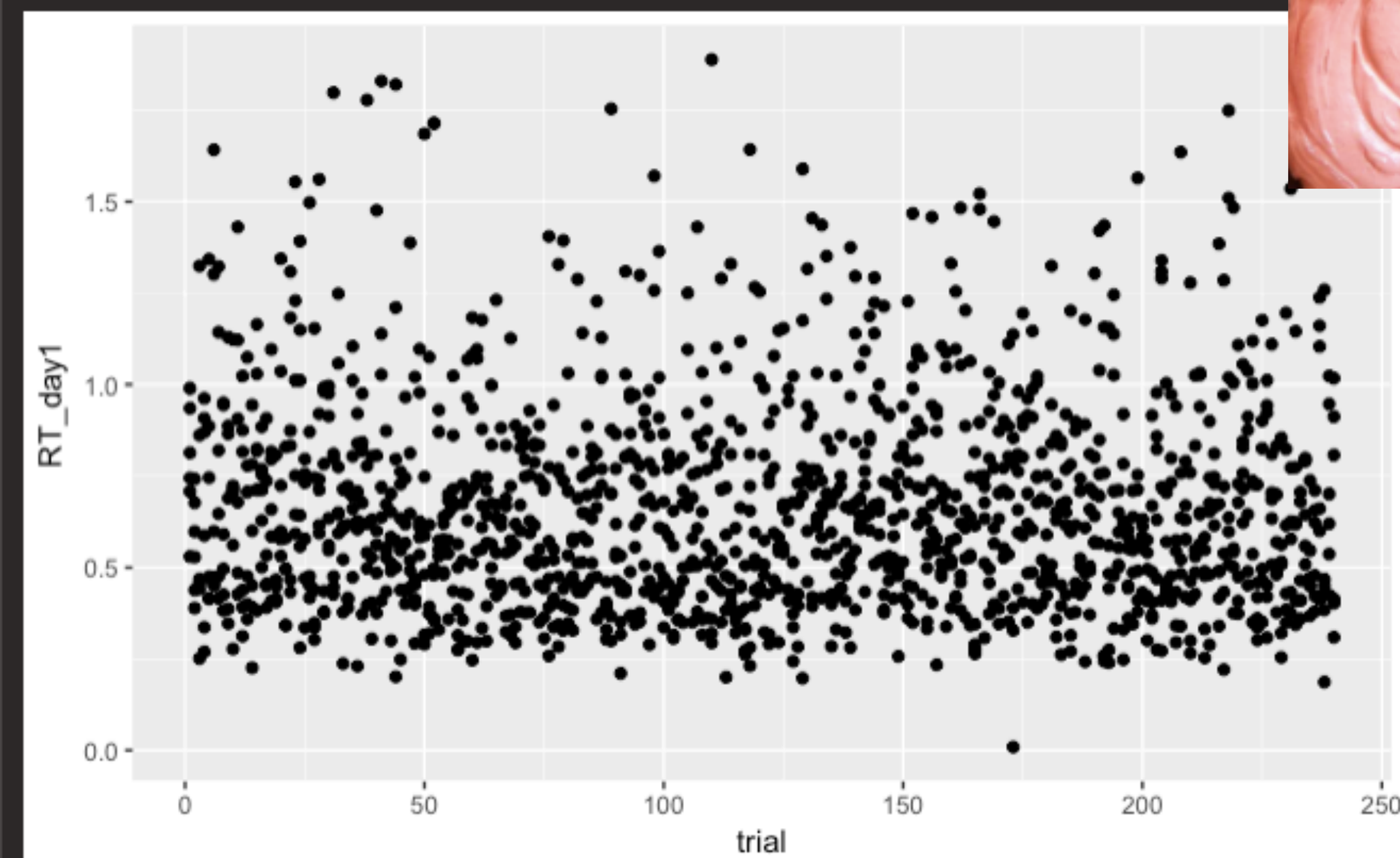
asethetics - x, y, size, shape, colour (colours of icing, shape of cake)

coordinate system(base)

data (cake)

Aesthetic to data mapping (aes)

```{r geom_points_RT}

# Step 2 - Tell ggplot what you want to display it as
ggplot(data, aes(x = trial, y = RT_day1)) + geom_point()

```



Aesthetic to data mapping (aes)

geometries points, line, bar, text (layers of cake) – *these also have asthetics*

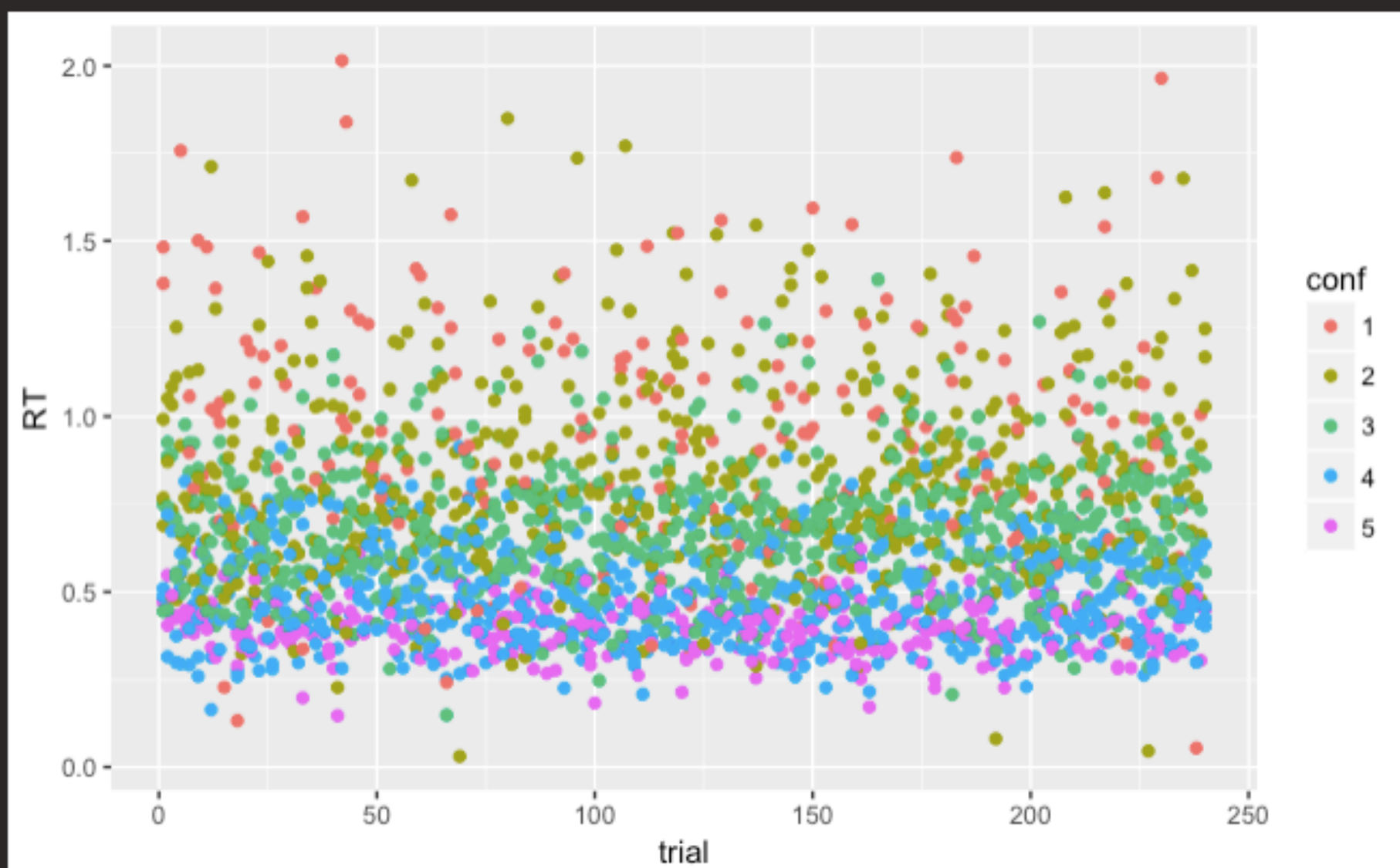
asethetics - x, y, size, shape, colour (colours of icing, shape of cake)

coordinate system(base)

data (cake)

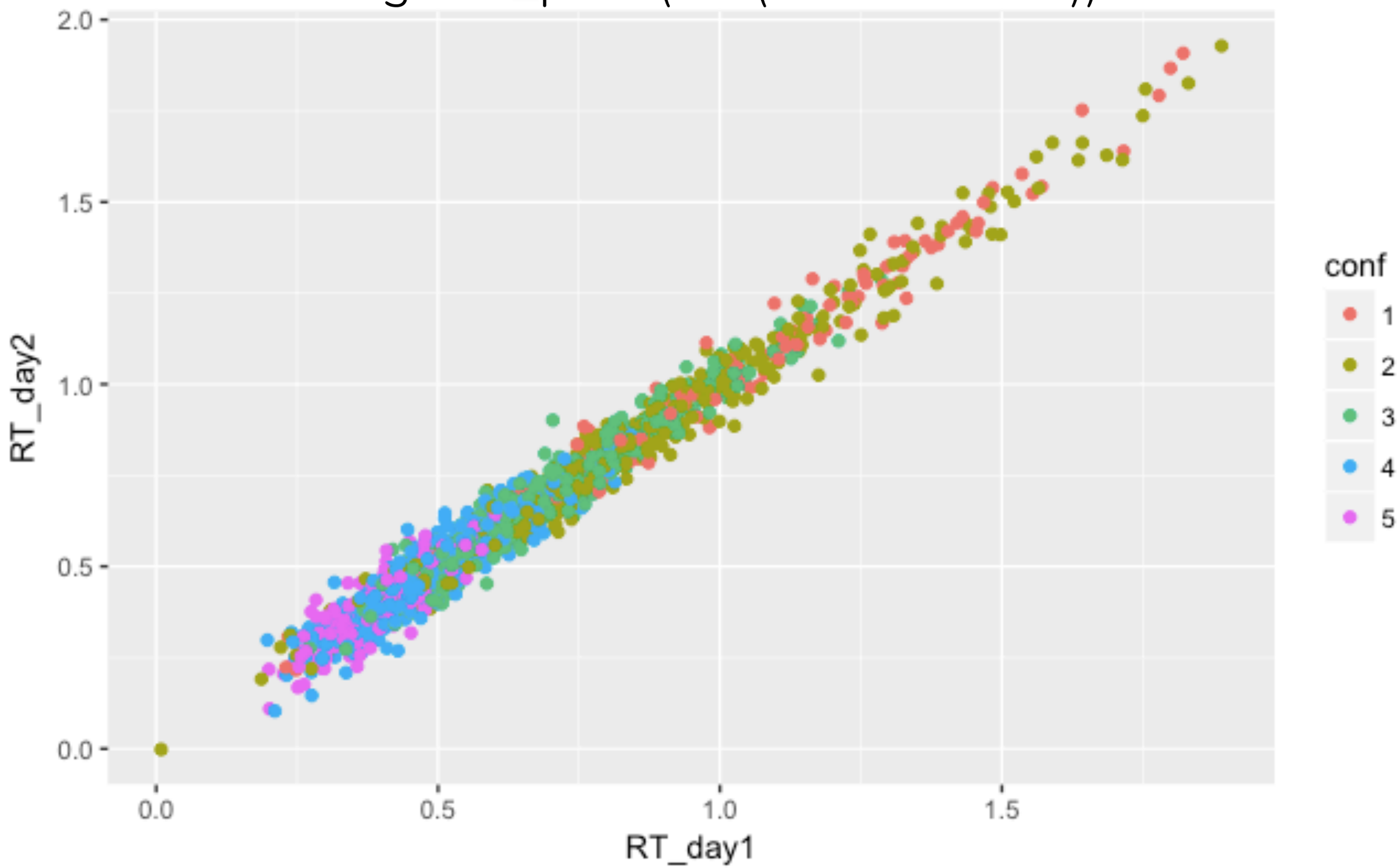
Aesthetic to data mapping (aes)

```{r add_colour}

# Step 3 - Let's try adding some colour
data[,'conf'] <- as.factor(data[,'conf'])
ggplot(data, aes(x = trial, y = RT)) + geom_point(aes(colour = conf))

```



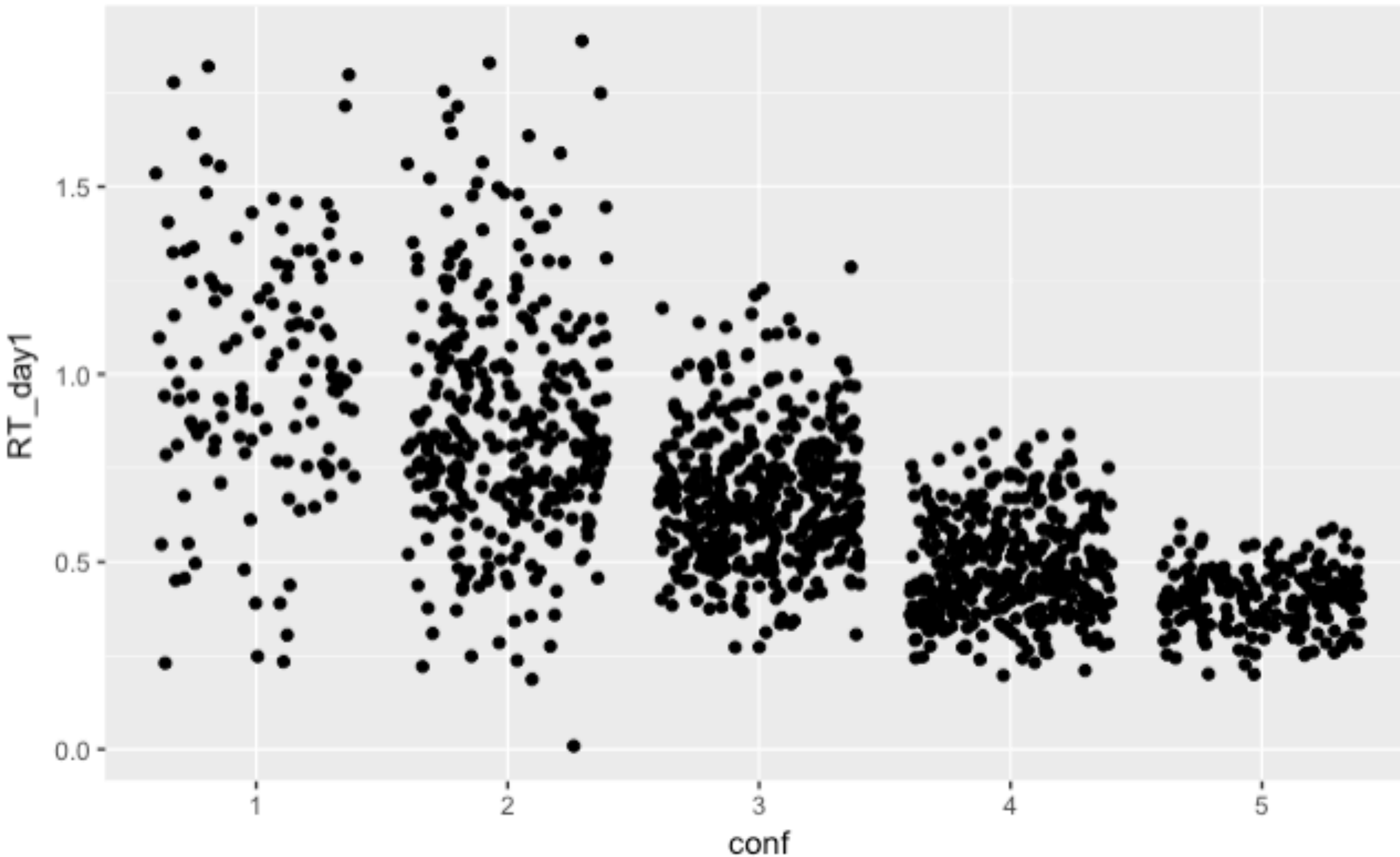
Scatterplots

```
ggplot(data, aes(x = RT_day1, y = RT_day2)) +
geom_point(aes(colour = conf))
```

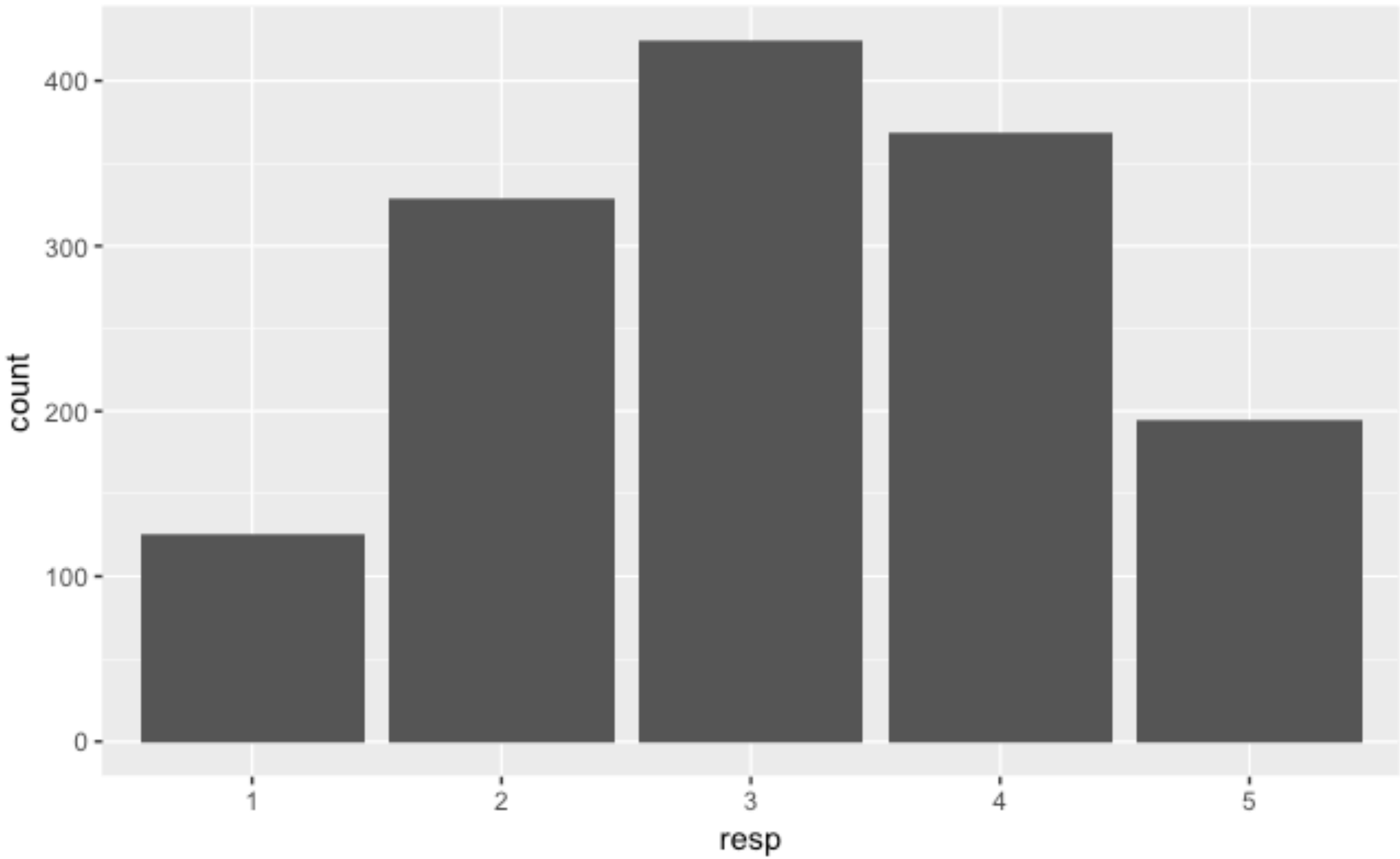


Scatterplots

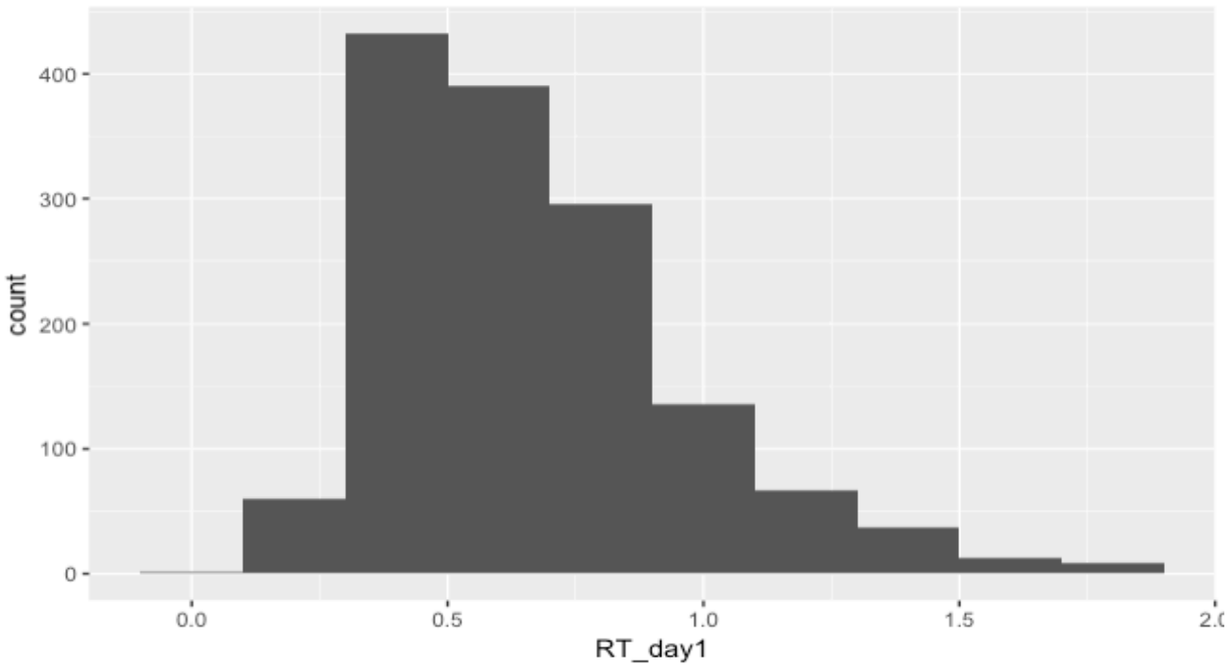
ggplot(data, aes(y = RT_day1, x = conf)) + geom_jitter()

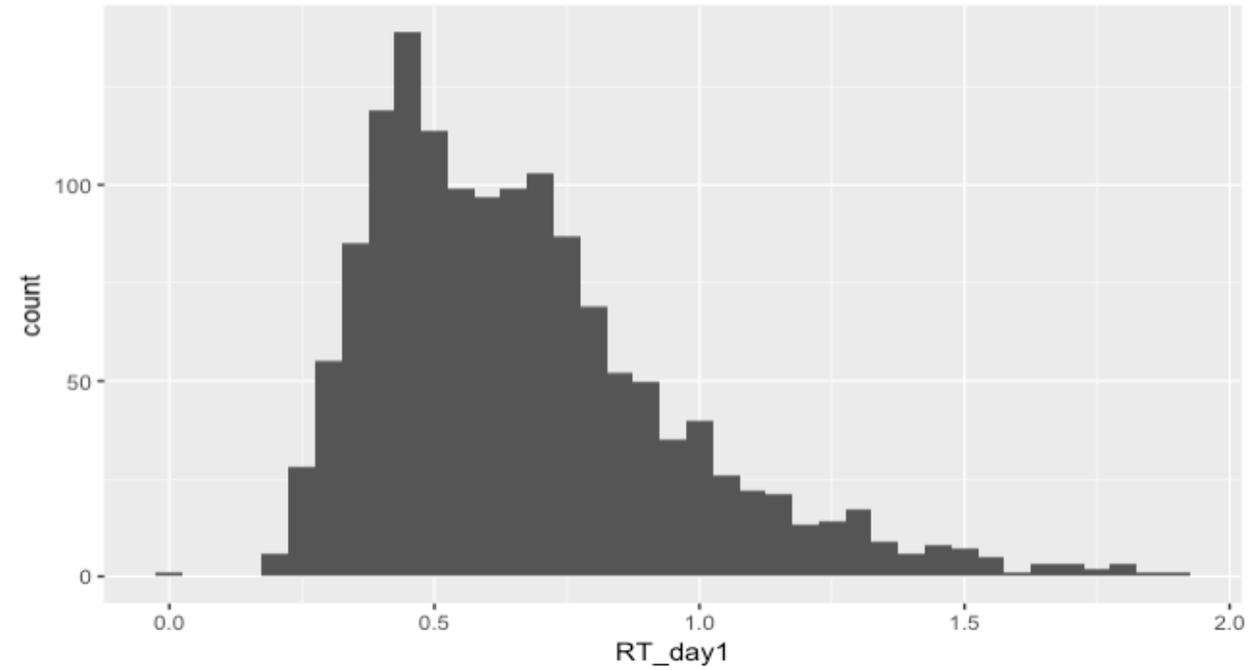


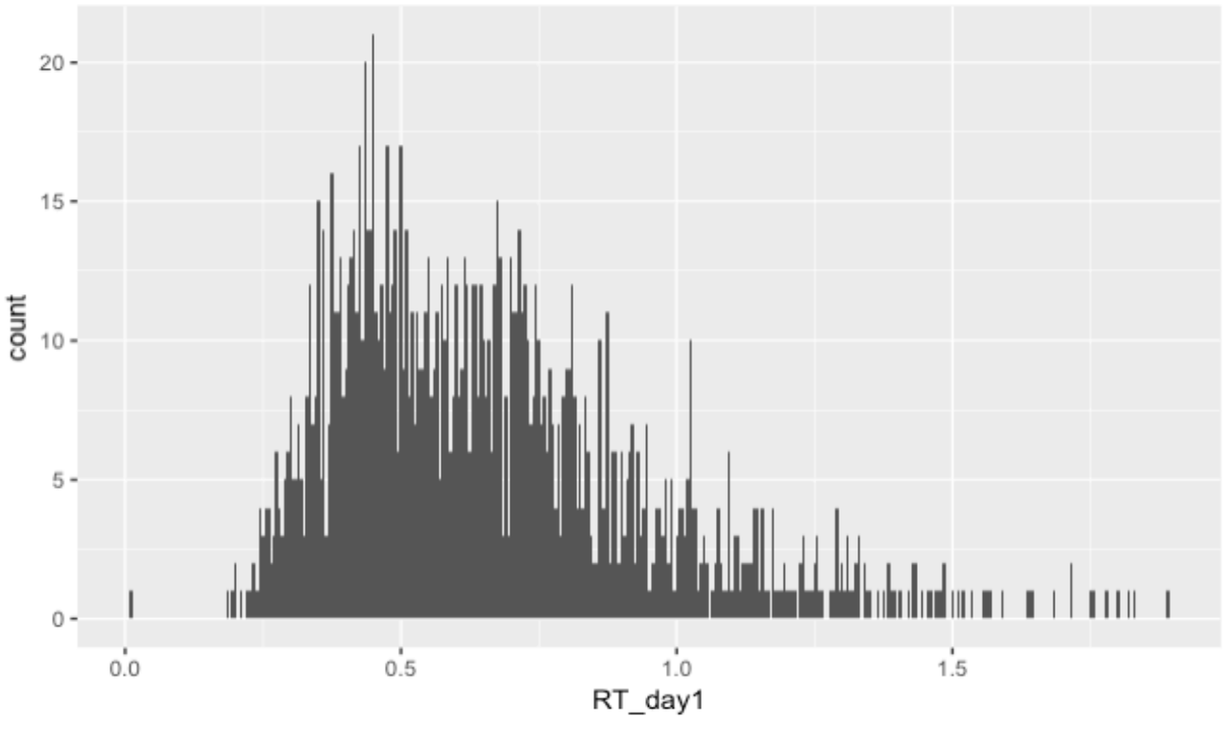
Bar graph

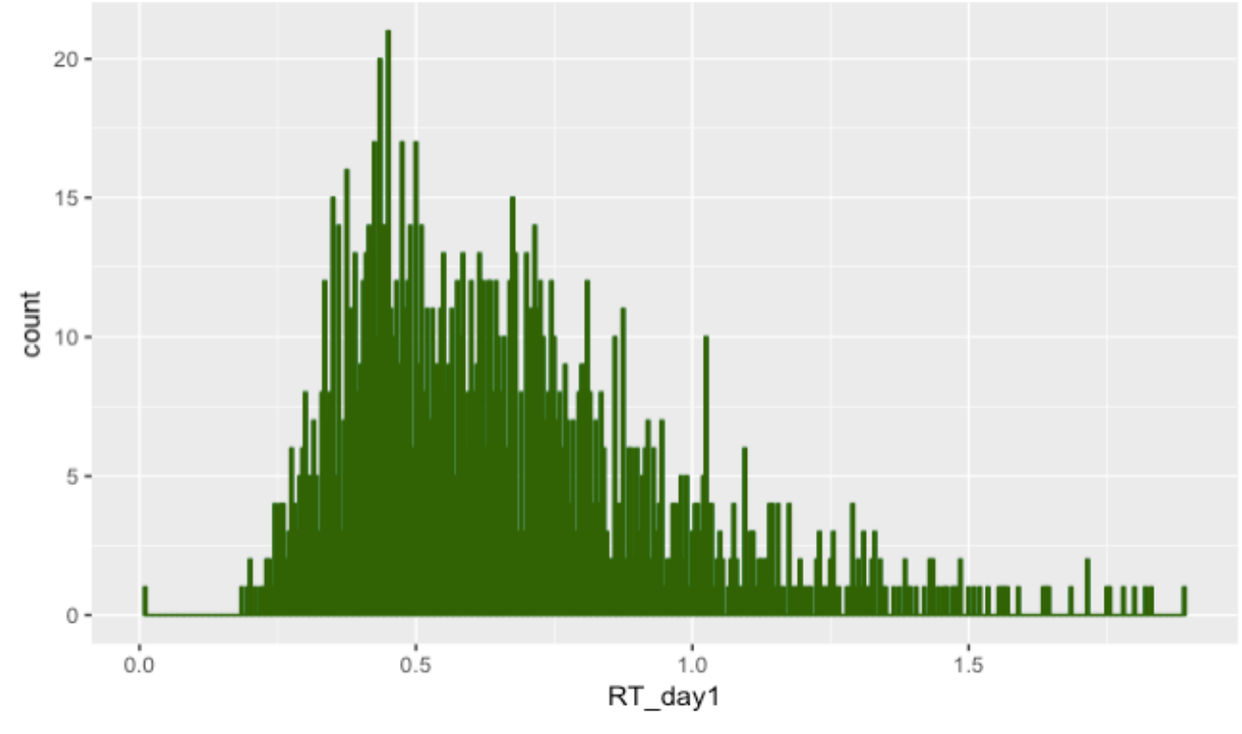


Histograms

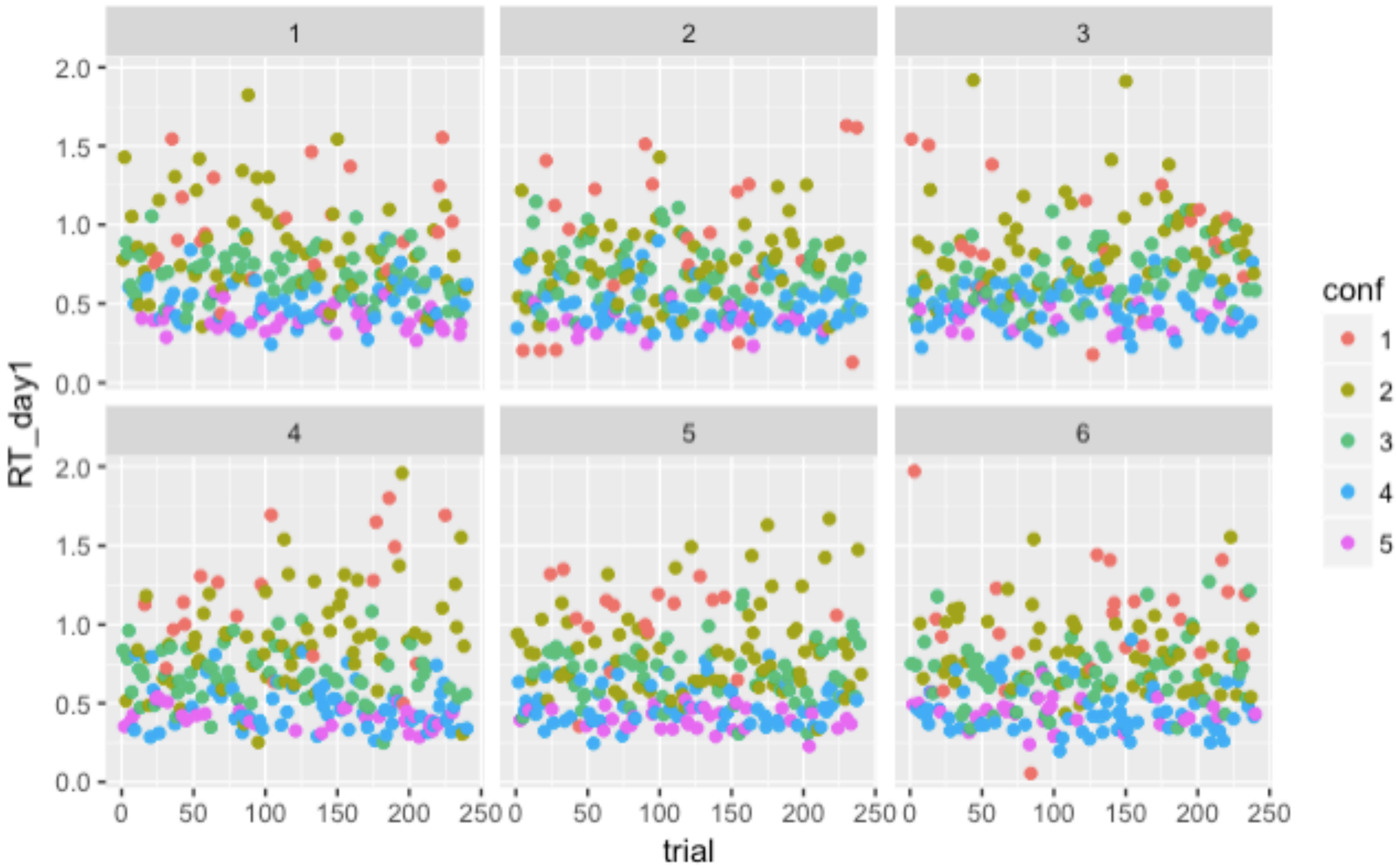

count
400
300
200
100
0
0.0
0.5
1.0
1.5
RT_day1


count
100
50
0
0.0
0.5
1.0
1.5
RT_day1

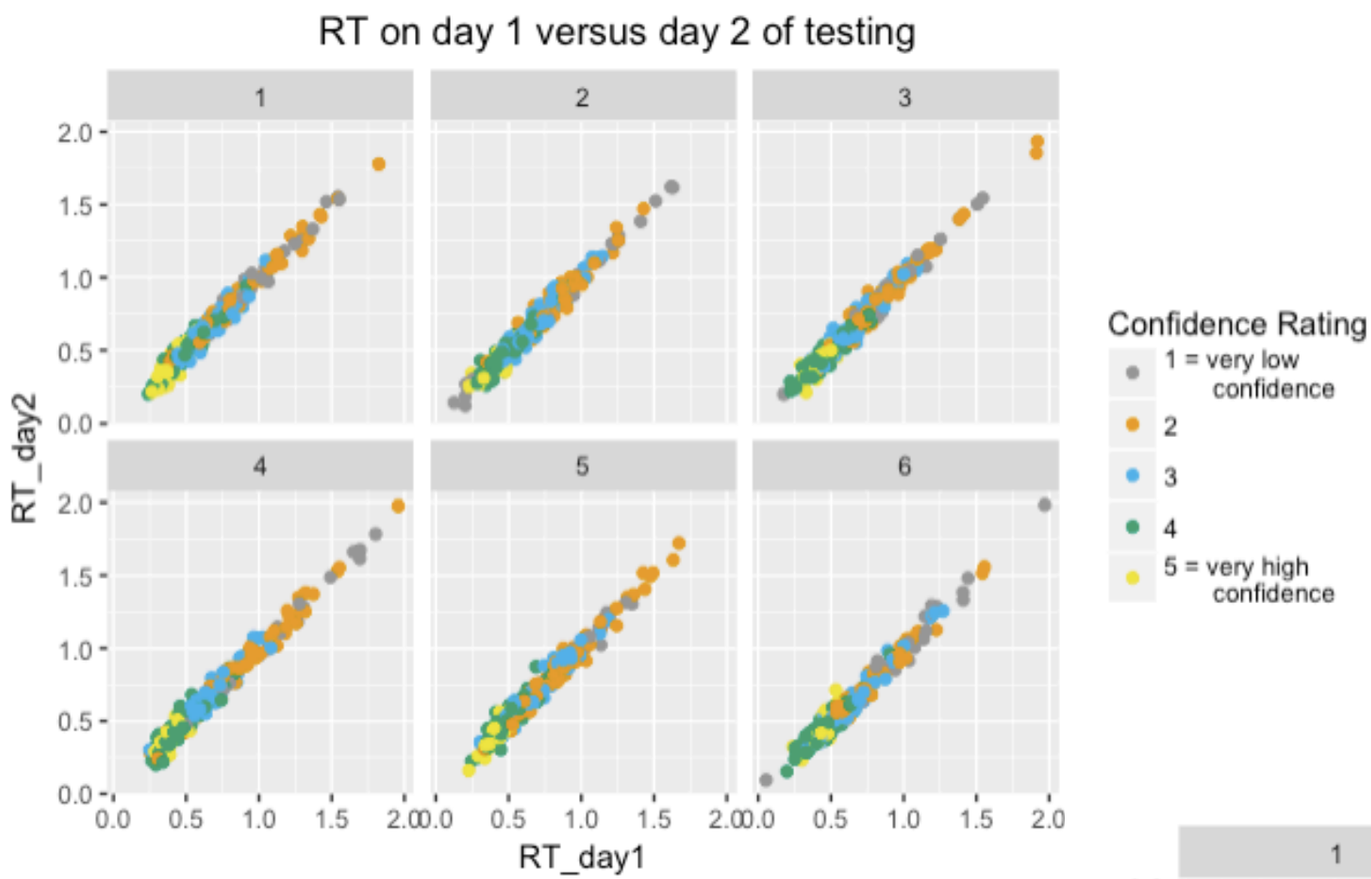

count
20
15
10
5
0
0.0
0.5
1.0
1.5
RT_day1

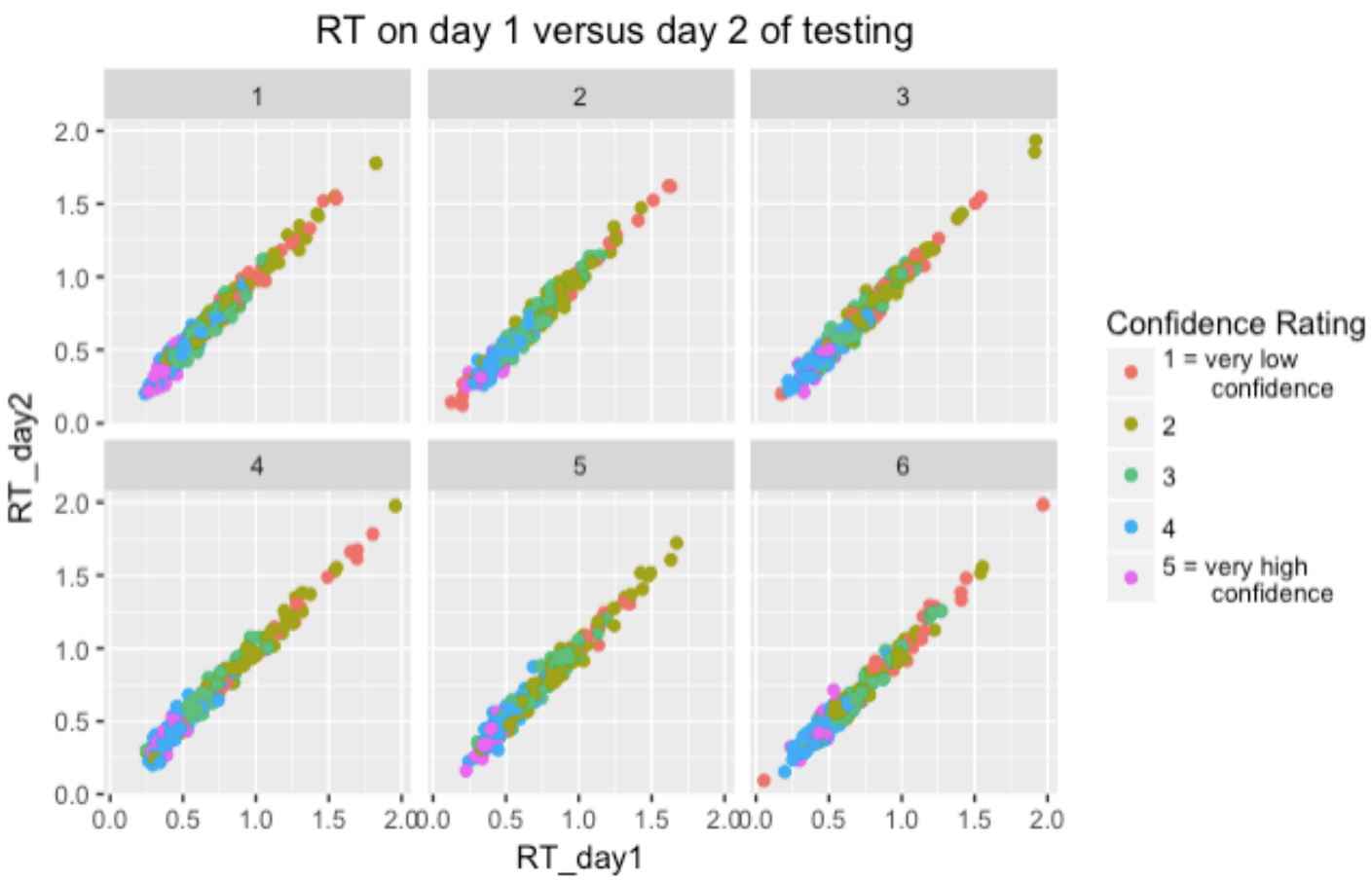

count
20
15
10
5
0
0.0
0.5
1.0
1.5
RT_day1

Plotting multiple subjects

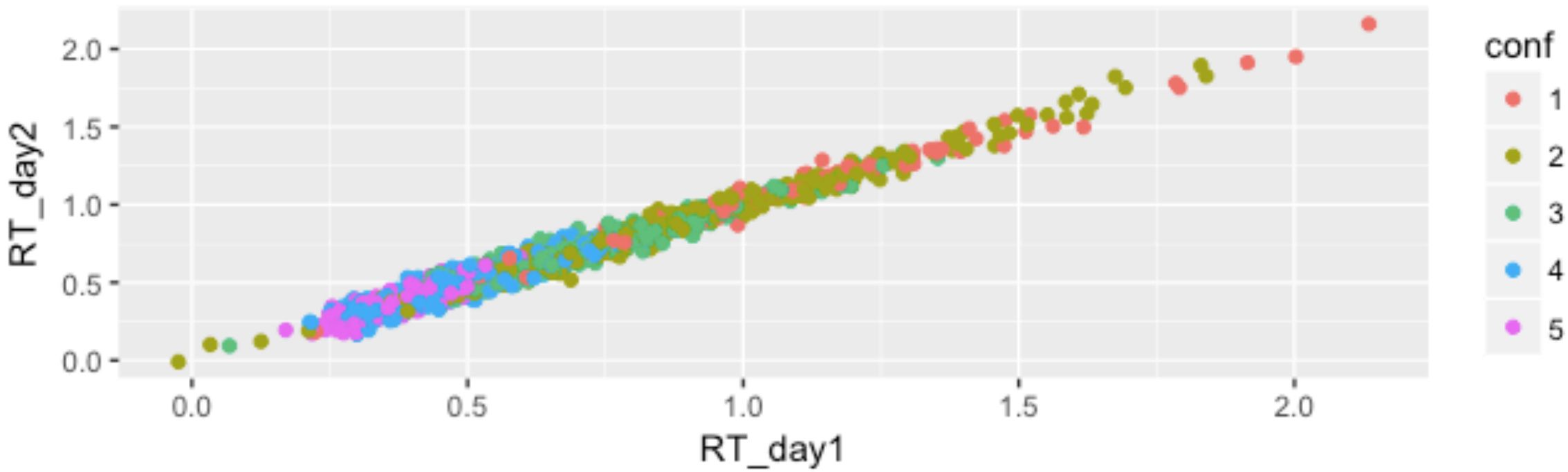


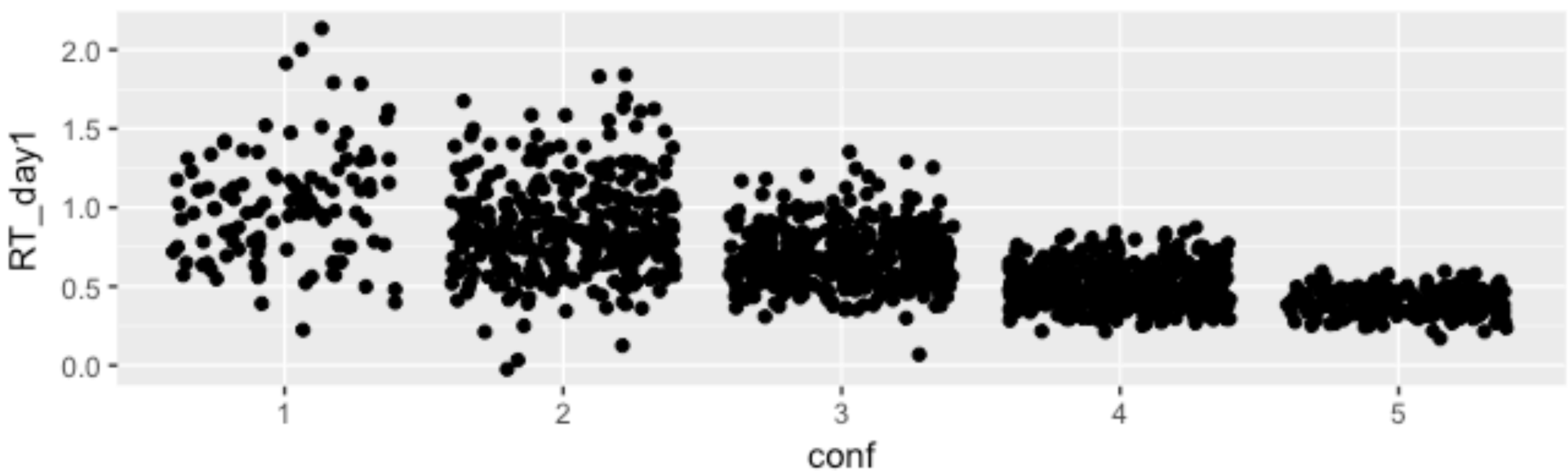
Colours and legends





Grid plotting





Thank you