

# GETTING STARTED



**WE WANT TO  
UNDERSTAND  
AND USE DATA  
INTELLIGENTLY**

**HOW TO SEE**

**WHERE TO LOOK**

**WHAT IS VISIBLE**

**SEEING IS NOT  
AS SIMPLE AS  
IT LOOKS**

**WE WANT TO  
DRAW GOOD  
DATA GRAPHICS  
REPRODUCIBLY**

# Abstraction in Software

Less —————→ More

Easy things are awkward

Hard things are straightforward

Really hard things are doable

Easy things are trivial

Hard things are really awkward

Really hard things are impossible

D3

Grid

ggplot

Stata

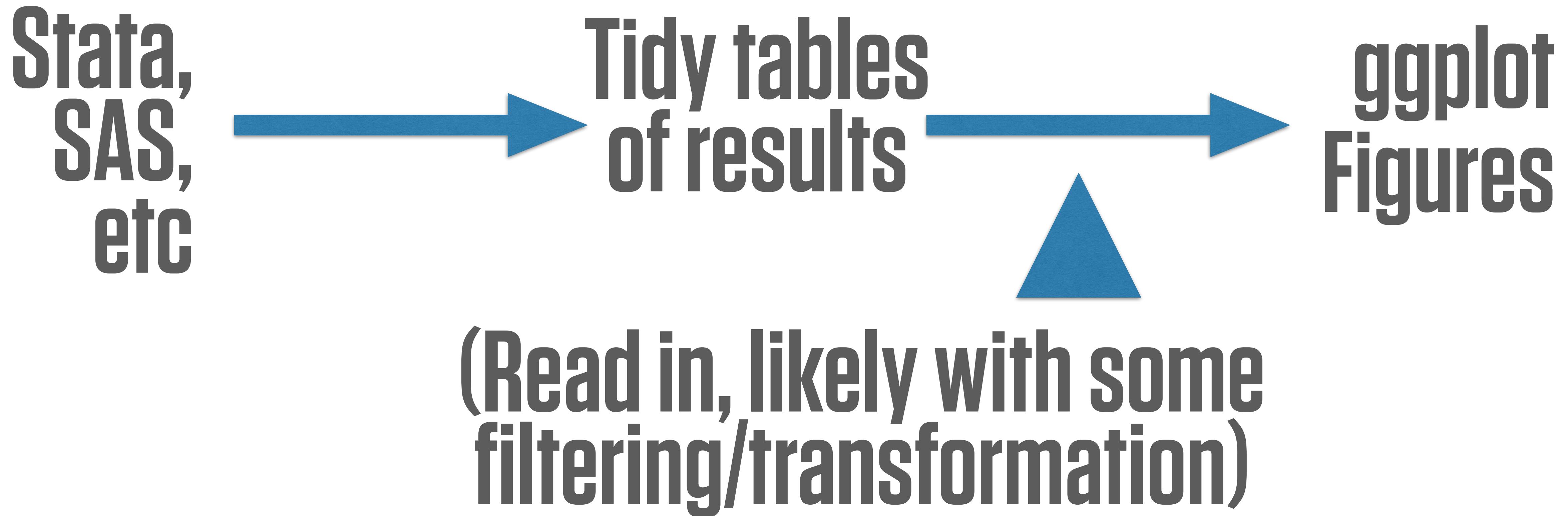
Excel

# Two ways to use R and ggplot

# 1. Do Everything in R



# 2. Just use ggplot

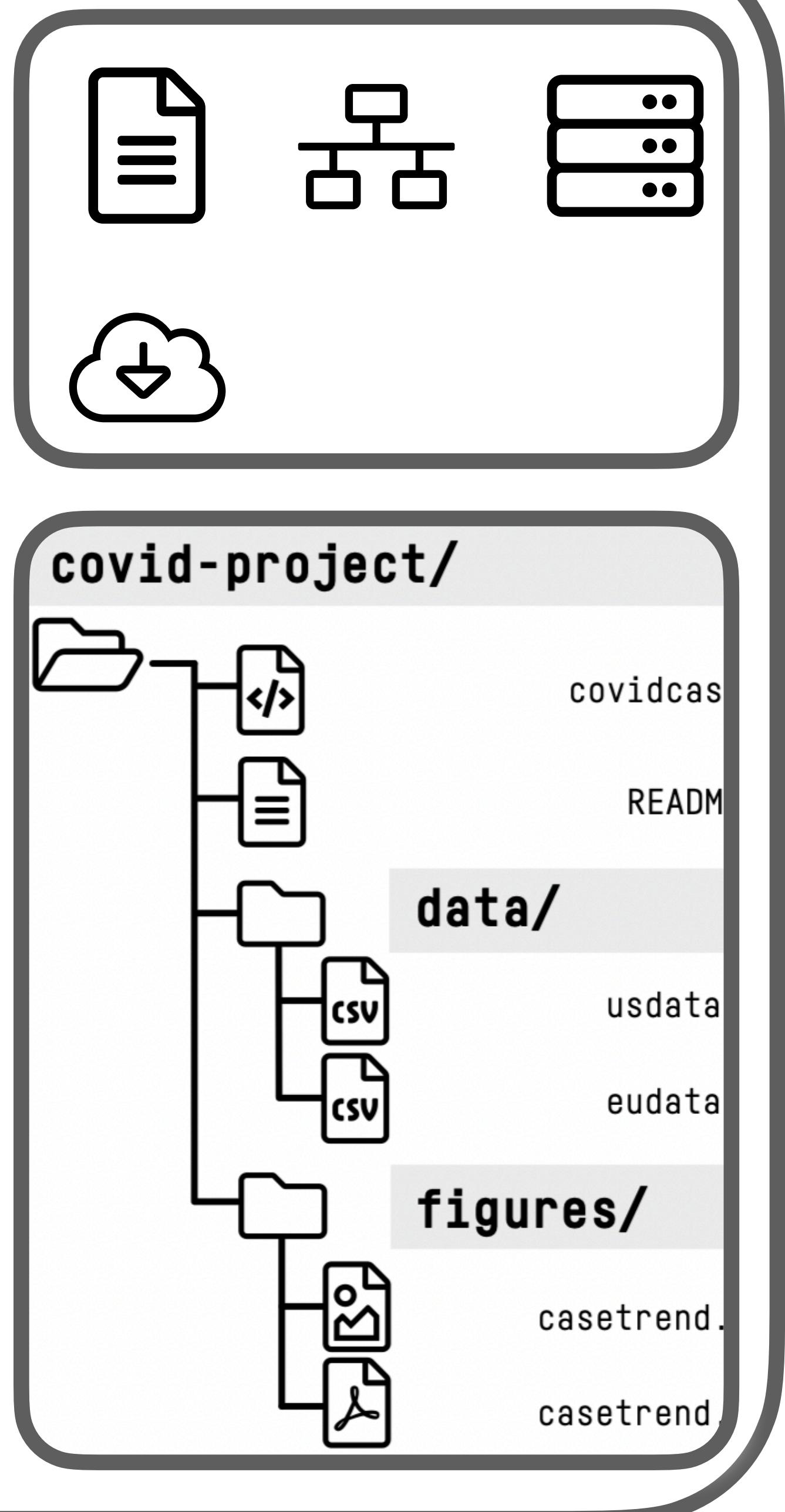


**THE RIGHT  
FRAME OF MIND**

**TYPE OUT YOUR  
CODE BY HAND**

**RSTUDIO**

```
# COVID      covidcases.Rmd  
  
## Get data from ECDC  
```{r get-data}  
covid_raw <- get_ecdc[url]  
``  
  
## Get data from the US  
```{r get-data}  
us_raw <- get_us[url]  
``
```

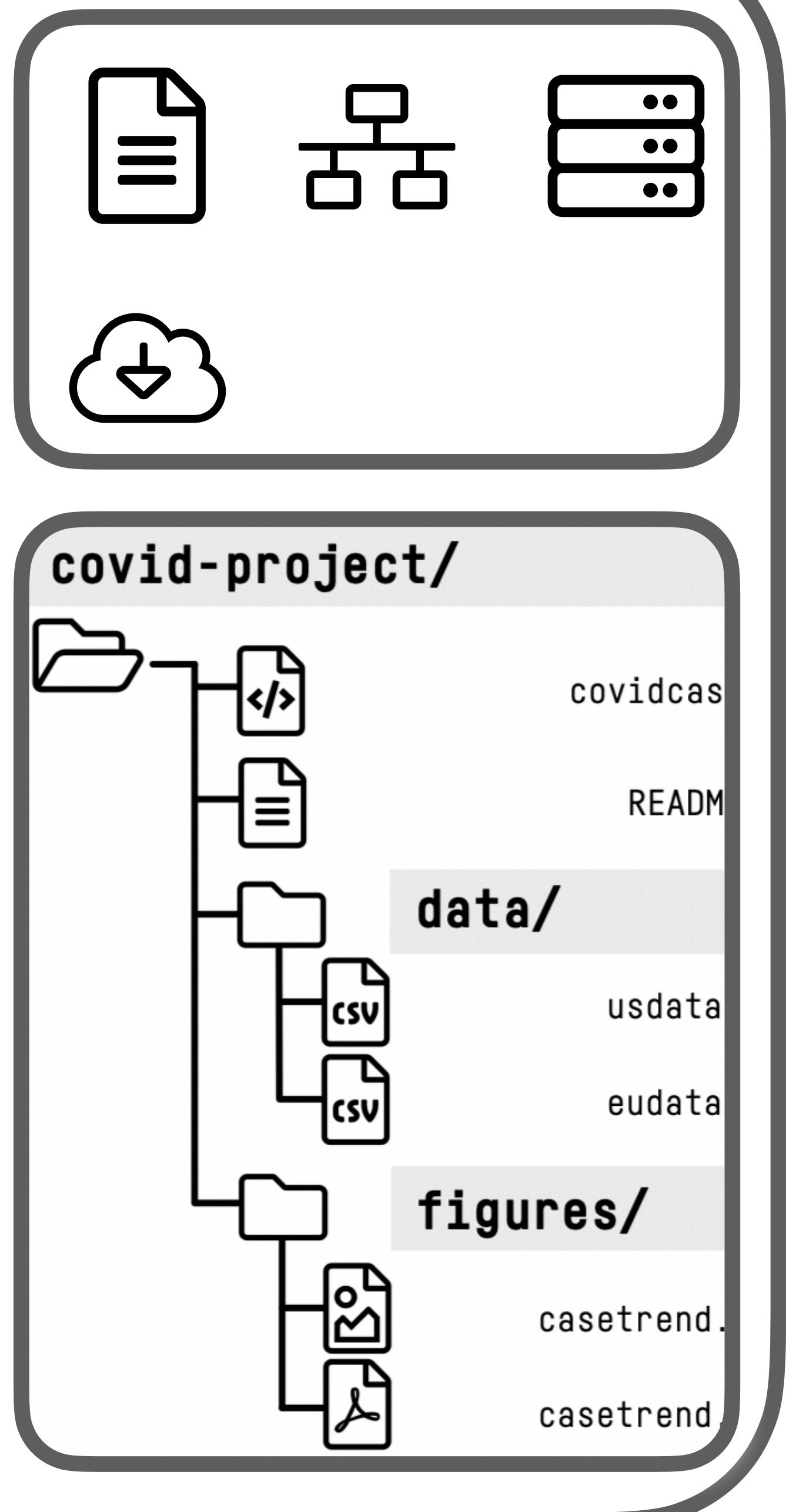


We write out instructions or recipes, rather than doing a series of point-and-click steps

```
# COVID          covidcases.Rmd

## Get data from ECDC
```{r get-data}
covid_raw <- get_ecdc[url]
```

## Get data from the US
```{r get-data}
us_raw <- get_us[url]
```
> -
```



Our raw data  
and the recipes  
for doing things  
with it are what  
is “real” in our  
data analysis

RStudio File Edit Code View Plots Session Build Debug Profile Tools Window Help

stathorizons\_0820 - master - RStudio

01\_introduction.Rmd x Addins Environment History Connections Git Tutorial Import Dataset Global Environment List

Environment is empty

Paper, Report, Analysis, Notes, etc, in RMarkdown

1 ---  
2 title: "Data Visualization"  
3 author: "Kieran Healy"  
4 date: "10-January-2020"  
5 output: html\_document  
6 ---  
7  
8 ## Data Visualization Notes  
9  
10 This is a starter RMarkdown project template to accompany courses taught with [\\*Data Visualization\\*](#). You can use it to take notes, write your code, and produce a good-looking, reproducible document that records the work you have done. At the very top of the file is a section of [\\*metadata\\*](#), or information about what the file is and what it does. The metadata is delimited by three dashes at the start and another three at the end. You should change the title, author, and date to the values that suit you. Keep the 'output' line as it is for now, however. Each line in the metadata has a structure. First the [\\*key\\*](#) ("title", "author", etc), then a colon, and then the [\\*value\\*](#) associated with the key.  
11  
12 ## This Document is an RMarkdown File  
13  
14 Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <<http://rmarkdown.rstudio.com>>.  
15  
16 When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. A [\\*code chunk\\*](#) is

1:1 # Data Visualization

Console Jobs

~/Documents/courses/stathorizons\_0820/ Type 'demo()' for some demos, 'help()' for on-line help, or 'help.start()' for an HTML browser interface to help. Type 'q()' to quit R.

knitr hook "anchor" is now available Loading required package: testthat

Attaching package: 'testthat'

The following object is masked from 'package:devtools':

test\_file

>

Files Plots Packages Help Viewer

New Folder Delete Rename More

Home Documents courses stathorizons\_0820

| Name                          | Size    | Modified               |
|-------------------------------|---------|------------------------|
| ..                            |         |                        |
| .gitignore                    | 40 B    | Jul 21, 2020, 11:16 AM |
| 01_introduction.Rmd           | 4 KB    | Jul 21, 2020, 11:16 AM |
| 02_get_started.Rmd            | 3.3 KB  | Jul 21, 2020, 11:16 AM |
| 03_make_a_plot.Rmd            | 5.8 KB  | Jul 21, 2020, 11:16 AM |
| 04_show_the_right_numbers.Rmd | 4.8 KB  | Jul 21, 2020, 11:16 AM |
| 05_tables_and_labels.Rmd      | 9.9 KB  | Jul 21, 2020, 11:16 AM |
| 06_models.Rmd                 | 15 KB   | Jul 21, 2020, 11:16 AM |
| 07_maps.Rmd                   | 12.5 KB | Jul 21, 2020, 11:16 AM |
| 08_refine_plots.Rmd           | 21.7 KB | Jul 21, 2020, 11:16 AM |
| 09_supplementary_material.Rmd | 16.9 KB | Jul 21, 2020, 11:16 AM |
| assets                        |         |                        |
| data                          |         |                        |
| figures                       |         |                        |
| keynote                       |         |                        |
| LICENSE.md                    | 18.1 KB | Jul 21, 2020, 11:16 AM |
| materials                     |         |                        |
| README.md                     | 5.7 KB  | Jul 21, 2020, 11:16 AM |
| slides                        |         |                        |
| stathorizons_0820.Rproj       |         |                        |

205 R Jul 22, 2020, 8:50 AM

RStudio File Edit Code View Plots Session Build Debug Profile Tools Window Help

stathorizons\_0820 - master - RStudio

01\_introduction.Rmd x

1 ---  
2 title: "Data Visualization"  
3 author: "Kieran Healy"  
4 date: "10-January-2020"  
5 output: html\_document  
6 ---  
7  
8 ## Data Visualization Notes  
9  
10 This is a starter RMarkdown project template to accompany courses taught with [\\*Data Visualization\\*](#). You can use it to take notes, write your code, and produce a good-looking, reproducible document that records the work you have done. At the very top of the file is a section of [\\*metadata\\*](#), or information about what the file is and what it does. The metadata is delimited by three dashes at the start and another three at the end. You should change the title, author, and date to the values that suit you. Keep the 'output' line as it is for now, however. Each line in the metadata has a structure. First the [\\*key\\*](#) ("title", "author", etc), then a colon, and then the [\\*value\\*](#) associated with the key.  
11  
12 ## This Document is an RMarkdown File  
13  
14 Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <<http://rmarkdown.rstudio.com>>.  
15  
16 When you click the [\\*\\*Knit\\*\\*](#) button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. A [\\*code chunk\\*](#) is

1:1 # Data Visualization ▾ R Markdown ▾

Console Jobs x

~/Documents/courses/stathorizons\_0820/ ↵

Type 'demo()' for some demos, 'help()' for on-line help, or 'help.start()' for an HTML browser interface to help.

Type 'q()' to quit R.

knitr hook "anchor" is now available

Loading required package: testthat

Attaching package: 'testthat'

The following object is masked from 'package:devtools':

  test\_file

>

Environment History Connections Git Tutorial

Import Dataset

Global Environment

Environment is empty

Files Plots Packages Help Viewer

New Folder Delete Rename More

Home Documents courses stathorizons\_0820

|    | Name                          | Size    | Modified               |
|----|-------------------------------|---------|------------------------|
| .. |                               |         |                        |
|    | .gitignore                    | 40 B    | Jul 21, 2020, 11:16 AM |
|    | 01_introduction.Rmd           | 4 KB    | Jul 21, 2020, 11:16 AM |
|    | 02_get_started.Rmd            | 3.3 KB  | Jul 21, 2020, 11:16 AM |
|    | 03_make_a_plot.Rmd            | 5.8 KB  | Jul 21, 2020, 11:16 AM |
|    | 04_show_the_right_numbers.Rmd | 4.8 KB  | Jul 21, 2020, 11:16 AM |
|    | 05_tables_and_labels.Rmd      | 9.9 KB  | Jul 21, 2020, 11:16 AM |
|    | 06_models.Rmd                 | 15 KB   | Jul 21, 2020, 11:16 AM |
|    | 07_maps.Rmd                   | 12.5 KB | Jul 21, 2020, 11:16 AM |
|    | 08_refine_plots.Rmd           | 21.7 KB | Jul 21, 2020, 11:16 AM |
|    | 09_supplementary_material.Rmd | 16.9 KB | Jul 21, 2020, 11:16 AM |
|    | assets                        |         |                        |
|    | data                          |         |                        |
|    | figures                       |         |                        |
|    | keynote                       |         |                        |
|    | LICENSE.md                    | 18.1 KB | Jul 21, 2020, 11:16 AM |
|    | materials                     |         |                        |
|    | README.md                     | 5.7 KB  | Jul 21, 2020, 11:16 AM |
|    | slides                        |         |                        |
|    | stathorizons_0820.Rproj       | 205 B   | Jul 22, 2020, 9:50 AM  |

Console: Type or send code here, see results

RStudio File Edit Code View Plots Session Build Debug Profile Tools Window Help

stathorizons\_0820 - master - RStudio

01\_introduction.Rmd x

1 ---  
2 title: "Data Visualization"  
3 author: "Kieran Healy"  
4 date: "10-January-2020"  
5 output: html\_document  
6 ---  
7  
8 ## Data Visualization Notes  
9  
10 This is a starter RMarkdown project template to accompany courses taught with [\\*Data Visualization\\*](#). You can use it to take notes, write your code, and produce a good-looking, reproducible document that records the work you have done. At the very top of the file is a section of [\\*metadata\\*](#), or information about what the file is and what it does. The metadata is delimited by three dashes at the start and another three at the end. You should change the title, author, and date to the values that suit you. Keep the 'output' line as it is for now, however. Each line in the metadata has a structure. First the [\\*key\\*](#) ("title", "author", etc), then a colon, and then the [\\*value\\*](#) associated with the key.  
11  
12 ## This Document is an RMarkdown File  
13  
14 Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <<http://rmarkdown.rstudio.com>>.  
15  
16 When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. A [code chunk](#) is

1:1 # Data Visualization

Console Jobs x

~/Documents/courses/stathorizons\_0820/ ↵  
Type 'demo()' for some demos, 'help()' for on-line help, or  
'help.start()' for an HTML browser interface to help.  
Type 'q()' to quit R.

knitr hook "anchor" is now available  
Loading required package: testthat

Attaching package: 'testthat'

The following object is masked from 'package:devtools':

test\_file

>

Environment History Connections Git Tutorial

Import Dataset

Global Environment

Environment is empty

Files Plots Packages Help Viewer

New Folder Delete Rename More

Home Documents courses stathorizons\_0820

| Name                          | Size    | Modified               |
|-------------------------------|---------|------------------------|
| ..                            | 40 B    | Jul 21, 2020, 11:16 AM |
| .gitignore                    | 4 KB    | Jul 21, 2020, 11:16 AM |
| 01_introduction.Rmd           | 3.3 KB  | Jul 21, 2020, 11:16 AM |
| 02_get_started.Rmd            | 5.8 KB  | Jul 21, 2020, 11:16 AM |
| 03_make_a_plot.Rmd            | 4.8 KB  | Jul 21, 2020, 11:16 AM |
| 04_show_the_right_numbers.Rmd | 9.9 KB  | Jul 21, 2020, 11:16 AM |
| 05_tables_and_labels.Rmd      | 15 KB   | Jul 21, 2020, 11:16 AM |
| 06_models.Rmd                 | 12.5 KB | Jul 21, 2020, 11:16 AM |
| 07_maps.Rmd                   | 21.7 KB | Jul 21, 2020, 11:16 AM |
| 08_refine_plots.Rmd           | 16.9 KB | Jul 21, 2020, 11:16 AM |
| 09_supplementary_material.Rmd | 18.1 KB | Jul 21, 2020, 11:16 AM |
| assets                        | 5.7 KB  | Jul 21, 2020, 11:16 AM |
| data                          |         |                        |
| figures                       |         |                        |
| keynote                       |         |                        |
| LICENSE.md                    |         |                        |
| materials                     |         |                        |
| README.md                     |         |                        |
| slides                        |         |                        |
| stathorizons_0820.Rproj       | 205 B   | Jul 22, 2020, 9:50 AM  |

Project files, Plots, Help

RStudio File Edit Code View Plots Session Build Debug Profile Tools Window Help

stathorizons\_0820 master RStudio

01\_introduction.Rmd x

1 ---  
2 title: "Data Visualization"  
3 author: "Kieran Healy"  
4 date: "10-January-2020"  
5 output: html\_document  
6 ---  
7  
8 ## Data Visualization Notes  
9  
10 This is a starter RMarkdown project template to accompany courses taught with \*Data Visualization\*. You can use it to take notes, write your code, and produce a good-looking, reproducible document that records the work you have done. At the very top of the file is a section of \*metadata\*, or information about what the file is and what it does. The metadata is delimited by three dashes at the start and another three at the end. You should change the title, author, and date to the values that suit you. Keep the 'output' line as it is for now, however. Each line in the metadata has a structure. First the \*key\* ("title", "author", etc), then a colon, and then the \*value\* associated with the key.  
11  
12 ## This Document is an RMarkdown File  
13  
14 Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <<http://rmarkdown.rstudio.com>>.  
15  
16 When you click the \*\*Knit\*\* button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. A \*code chunk\* is

1:1 # Data Visualization ▾ R Markdown ▾

Console Jobs x

~/Documents/courses/stathorizons\_0820/ ↵

Type 'demo()' for some demos, 'help()' for on-line help, or 'help.start()' for an HTML browser interface to help.

Type 'q()' to quit R.

knitr hook "anchor" is now available

Loading required package: testthat

Attaching package: 'testthat'

The following object is masked from 'package:devtools':

test\_file

>

Environment History Connections Git Tutorial

Import Dataset

Global Environment

Environment is empty

Inspect objects you create

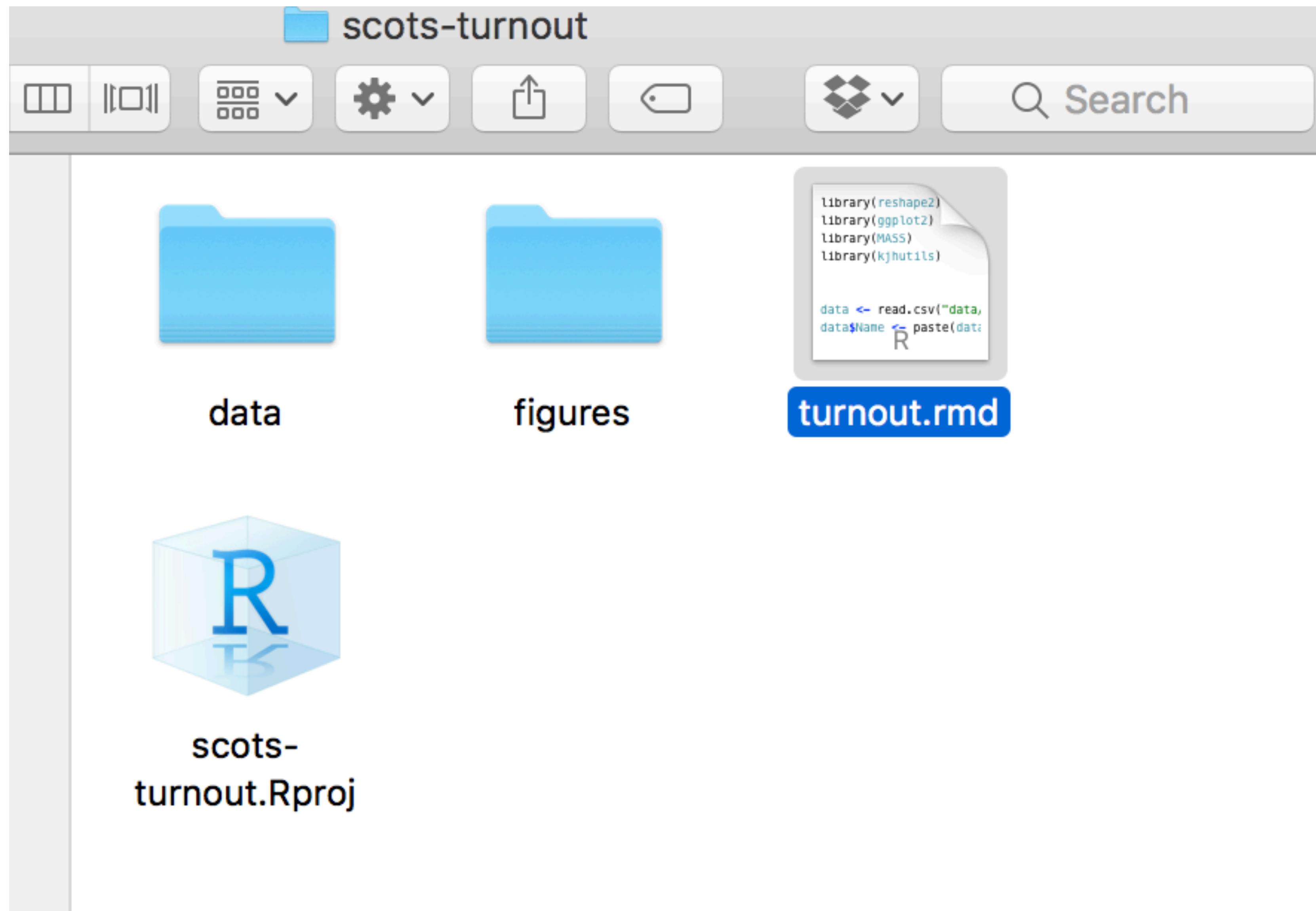
Files Plots Packages Help Viewer

New Folder Delete Rename More

Home > Documents > courses > stathorizons\_0820

| Name                          | Size    | Modified               |
|-------------------------------|---------|------------------------|
| ..                            |         |                        |
| .gitignore                    | 40 B    | Jul 21, 2020, 11:16 AM |
| 01_introduction.Rmd           | 4 KB    | Jul 21, 2020, 11:16 AM |
| 02_get_started.Rmd            | 3.3 KB  | Jul 21, 2020, 11:16 AM |
| 03_make_a_plot.Rmd            | 5.8 KB  | Jul 21, 2020, 11:16 AM |
| 04_show_the_right_numbers.Rmd | 4.8 KB  | Jul 21, 2020, 11:16 AM |
| 05_tables_and_labels.Rmd      | 9.9 KB  | Jul 21, 2020, 11:16 AM |
| 06_models.Rmd                 | 15 KB   | Jul 21, 2020, 11:16 AM |
| 07_maps.Rmd                   | 12.5 KB | Jul 21, 2020, 11:16 AM |
| 08_refine_plots.Rmd           | 21.7 KB | Jul 21, 2020, 11:16 AM |
| 09_supplementary_material.Rmd | 16.9 KB | Jul 21, 2020, 11:16 AM |
| assets                        |         |                        |
| data                          |         |                        |
| figures                       |         |                        |
| keynote                       |         |                        |
| LICENSE.md                    | 18.1 KB | Jul 21, 2020, 11:16 AM |
| materials                     |         |                        |
| README.md                     | 5.7 KB  | Jul 21, 2020, 11:16 AM |
| slides                        |         |                        |
| stathorizons_0820.Rproj       |         |                        |

205 R Jul 22, 2020, 9:50 AM



## Name

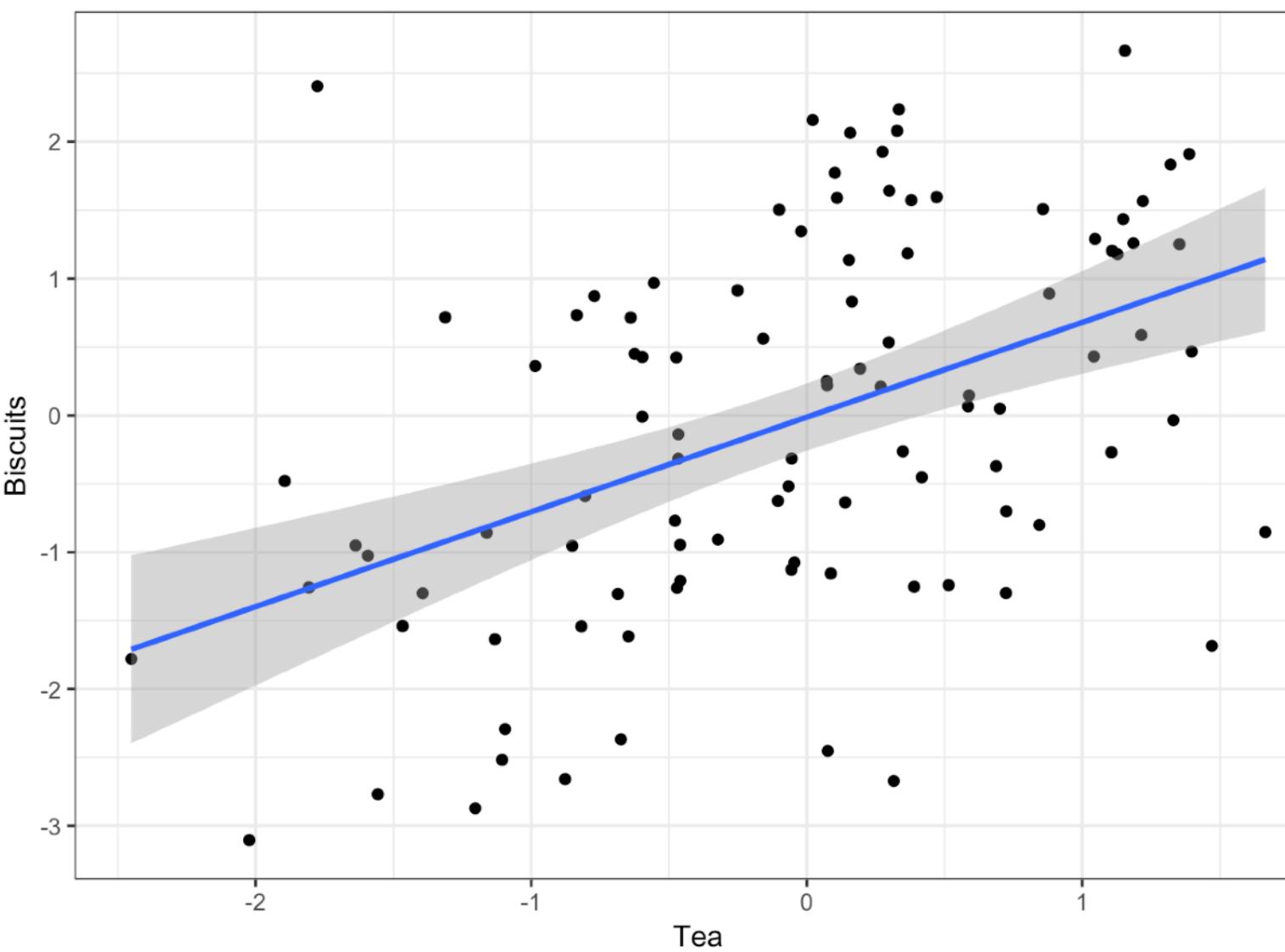
- ▶  **analysis**
- ▶  **cache**
- ▶  **data**
- ▶  **doc**
- ▶  **figures**
- ▶  **paper**
- ▶  **setup**
- ▶  **svyglm**
- ▶  **fin-capability.Rproj**

| Name  | ^   D |
|---|-------|
| ►  <b>data</b>               | 1     |
| ►  <b>data-raw</b>           | 1     |
| ►  <b>docs</b>               | 1     |
| ►  <b>inst</b>               | 1     |
| ►  <b>man</b>                | 1     |
| ►  <b>misc</b>               | 1     |
| ►  <b>R</b>                 | 1     |
| ►  <b>raw</b>              | 1     |
| ►  <b>rdoc</b>             | 1     |
| ►  <b>vignettes</b>        | 1     |
| ►  <b>vignettes-source</b> | 1     |
|  _pkgdown.yml              | 1     |
|  DESCRIPTION               | 1     |
|  gss_prep.Rmd              | 1     |
|  gssr.Rproj                | 1     |
|  LICENSE                   | 1     |
|  LICENSE.md                | 1     |
|  NAMESPACE                 | 1     |
|  NEWS.md                   | 1     |
|  README.md                 | 1     |
|  README.Rmd                | 1     |

**USE RMarkdown  
TO REPRODUCE  
YOUR OWN WORK**

# 1. Lorem Ipsum

  Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.



This is what  
we want to end up  
with: nicely  
formatted text,  
plots, and tables.

Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

# # Lorem Ipsum

  Lorem ipsum dolor sit amet, consectetur adipisicing elit,  
  sed do \*eiusmod tempor\* incididunt ut labore et dolore magna  
  aliqua. Ut enimad minim veniam, quis nostrud exercitation  
  ullamco laboris nisi ut aliquip ex ea commodo consequat.

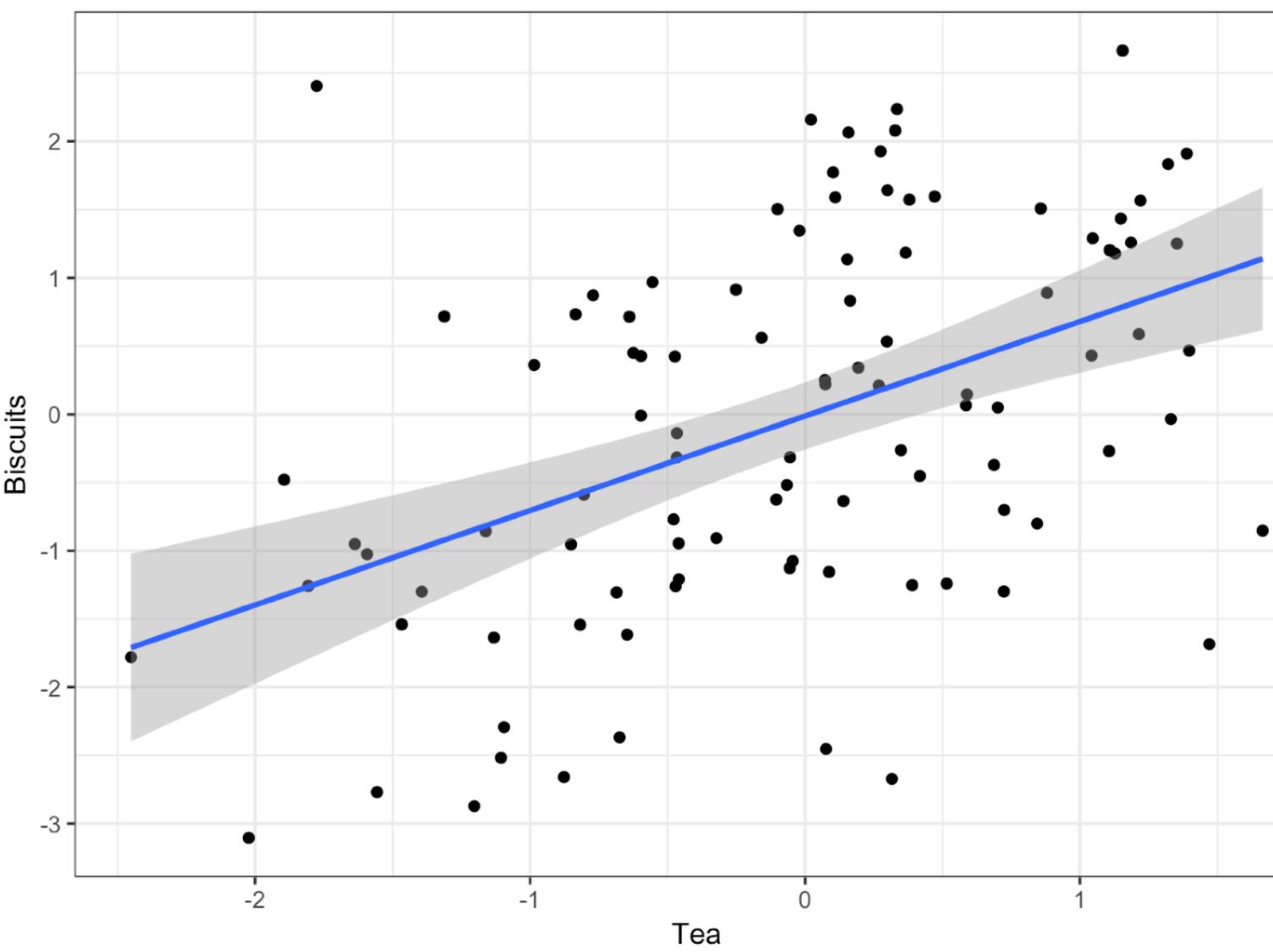
```
library(ggplot2)
tea <- rnorm(100)
biscuits <- tea + rnorm(100, 0, 1.3)
data <- data.frame(tea, biscuits)
p <- ggplot(data, aes(x = tea, y = biscuits)) +
  geom_point() +
  geom_smooth(method = "lm") +
  labs(x = "Tea", y = "Biscuits") + theme_bw()
print(p)
```

Duis aute irure dolor in reprehenderit in voluptate velit esse  
cillum dolore eu fugiat nulla pariatur. Excepteur sint  
occaecat cupidatat non proident, sunt in culpa qui officia  
deserunt mollit anim id est laborum.

In a **Literate Programming** approach to documents, chunks of code are processed and replaced with their output

# 1. Lorem Ipsum

  Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.



Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

In a Literate Programming approach to documents, chunks of code are processed and replaced with their output

```
# Report      notes.Rmd  
We can see this *relationship*  
in a scatterplot.
```

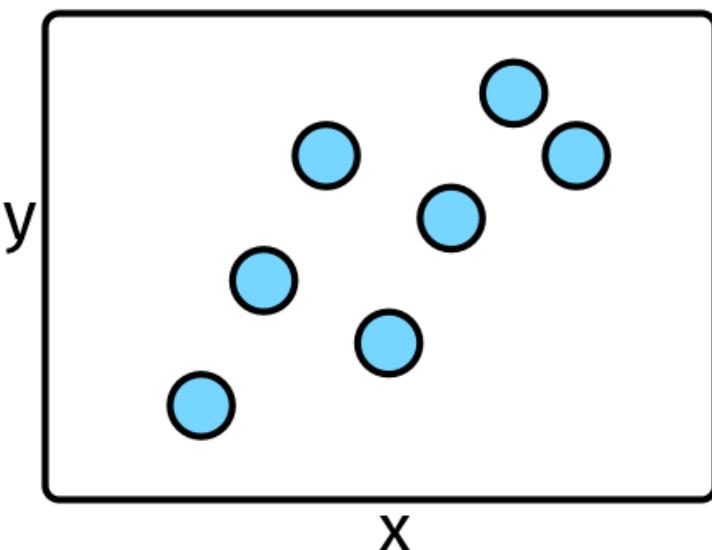
```
```{r my-code}  
p <- ggplot(data, mapping)  
p + geom_point()  
```
```

As you can see, this plot  
looks pretty nice.

knit in R

**Report** notes.pdf

We can see this *relationship*  
in a scatterplot.



As you can see, this plot  
looks pretty nice.

An Rmd document lets you  
keep your code and notes  
together in plain text

And produce good-looking  
output in a range of formats

```
# Report      notes.Rmd  
We can see this *relationship*  
in a scatterplot.
```

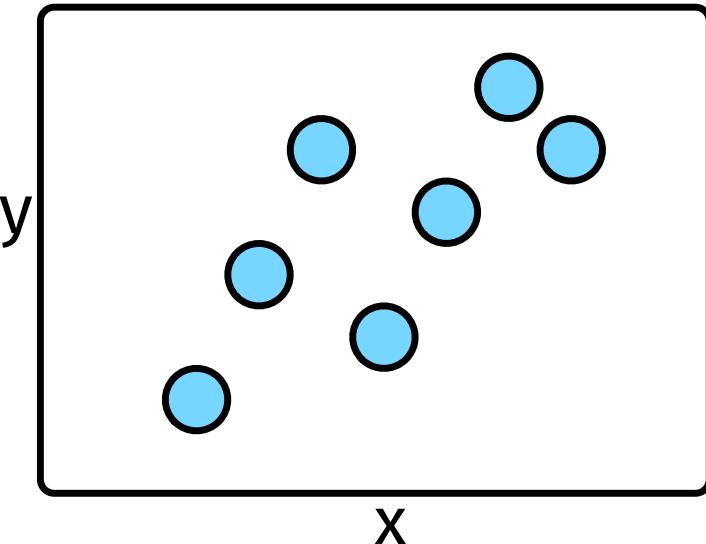
```
```{r my-code}  
p <- ggplot(data, mapping)  
p + geom_point()  
```
```

As you can see, this plot  
looks pretty nice.

knit in R

**Report** notes.html

We can see this *relationship*  
in a scatterplot.



As you can see, this plot  
looks pretty nice.

An Rmd document lets you  
keep your code and notes  
together in plain text

And produce good-looking  
output in a range of formats

```
# Report      notes.Rmd  
We can see this *relationship*  
in a scatterplot.
```

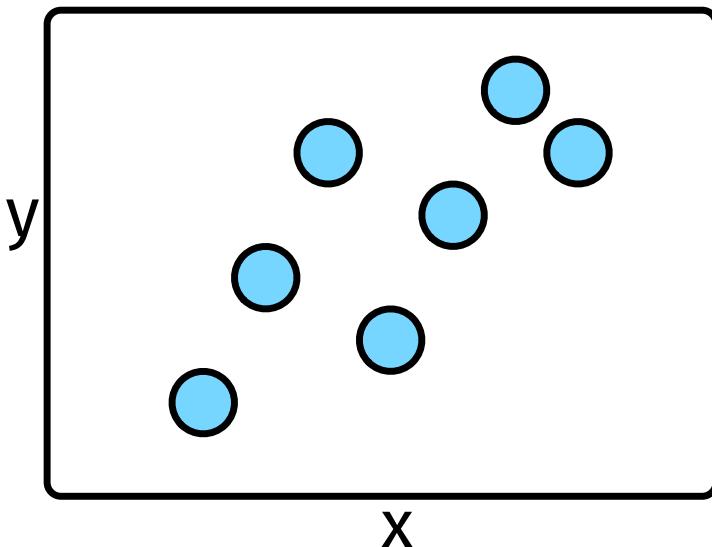
```
```{r my-code}  
p <- ggplot(data, mapping)  
p + geom_point()  
```
```

As you can see, this plot  
looks pretty nice.

knit in R

**Report** notes.docx

We can see this *relationship*  
in a scatterplot.



As you can see, this plot  
looks pretty nice.

An Rmd document lets you  
keep your code and notes  
together in plain text

And produce good-looking  
output in a range of formats

## Markdown

```
# Header  
## Subhead  
Plain text  
*italics*  
**bold**  
'verbatim'  
1. List  
2. List  
- Bullet 1  
- Bullet 2  
Footnote.[^1]  
[^1]: The footnote.
```

## Output

```
Header  
Subhead  
Plain text  
italics  
bold  
verbatim  
1. List  
2. List  
◦ Bullet 1  
◦ Bullet 2  
Footnote1
```



A **Markdown Processor** turns the marked-up plain text into actually formatted output in **HTML, PDF, DOCX or other file types**.

Markdown puts formatting instructions in plain-text documents

```
---
```

```
title: "My Notes"
```

```
author: "Kieran healy"
```

```
date: "12/7/2016"
```

```
output: html_document
```

```
--
```

# Header section provides metadata and sets options

```
```{r setup, include=FALSE}
```

```
knitr::opts_chunk$set(echo = TRUE)
```

```
```
```

```
## R Markdown
```

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <<http://rmarkdown.rstudio.com>>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
```{r cars}
```

```
summary(cars)
```

```
```
```

```
## Including Plots
```

You can also embed plots, for example:

```
```{r pressure, echo=FALSE}
```

```
plot(pressure)
```

```
```
```

## Code chunks can have their own names and options

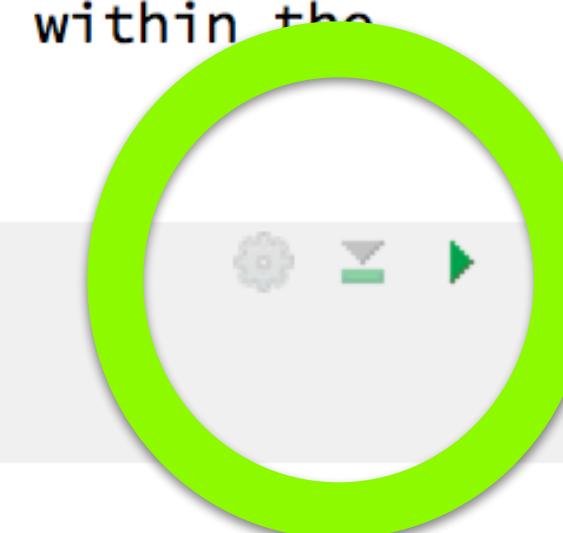
Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.



Code chunk



Text with  
Markdown  
formatting



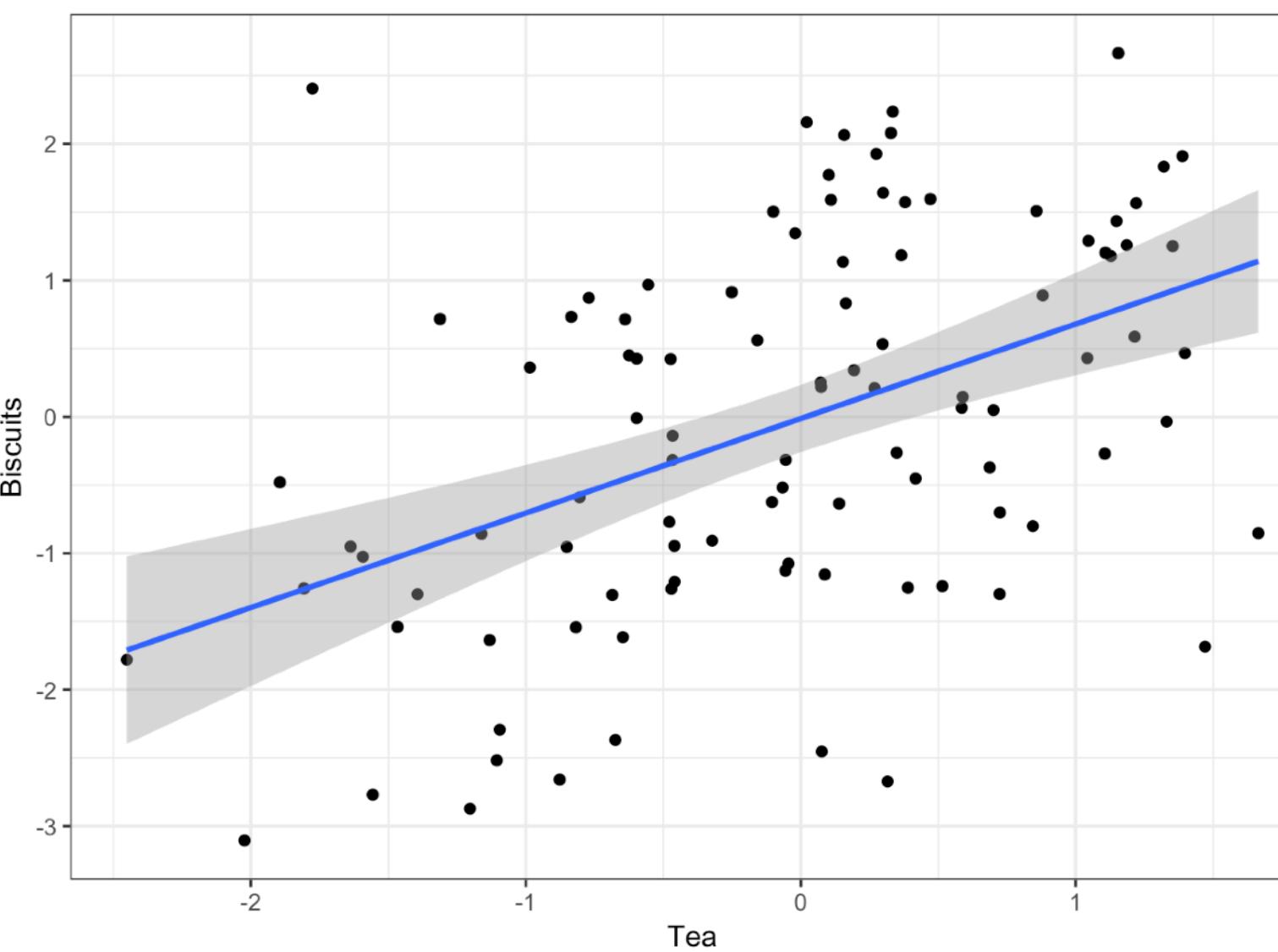
In RStudio, code  
chunks can be  
"played" one at a  
time



Chunks are  
replaced by their  
output when the  
document is  
made

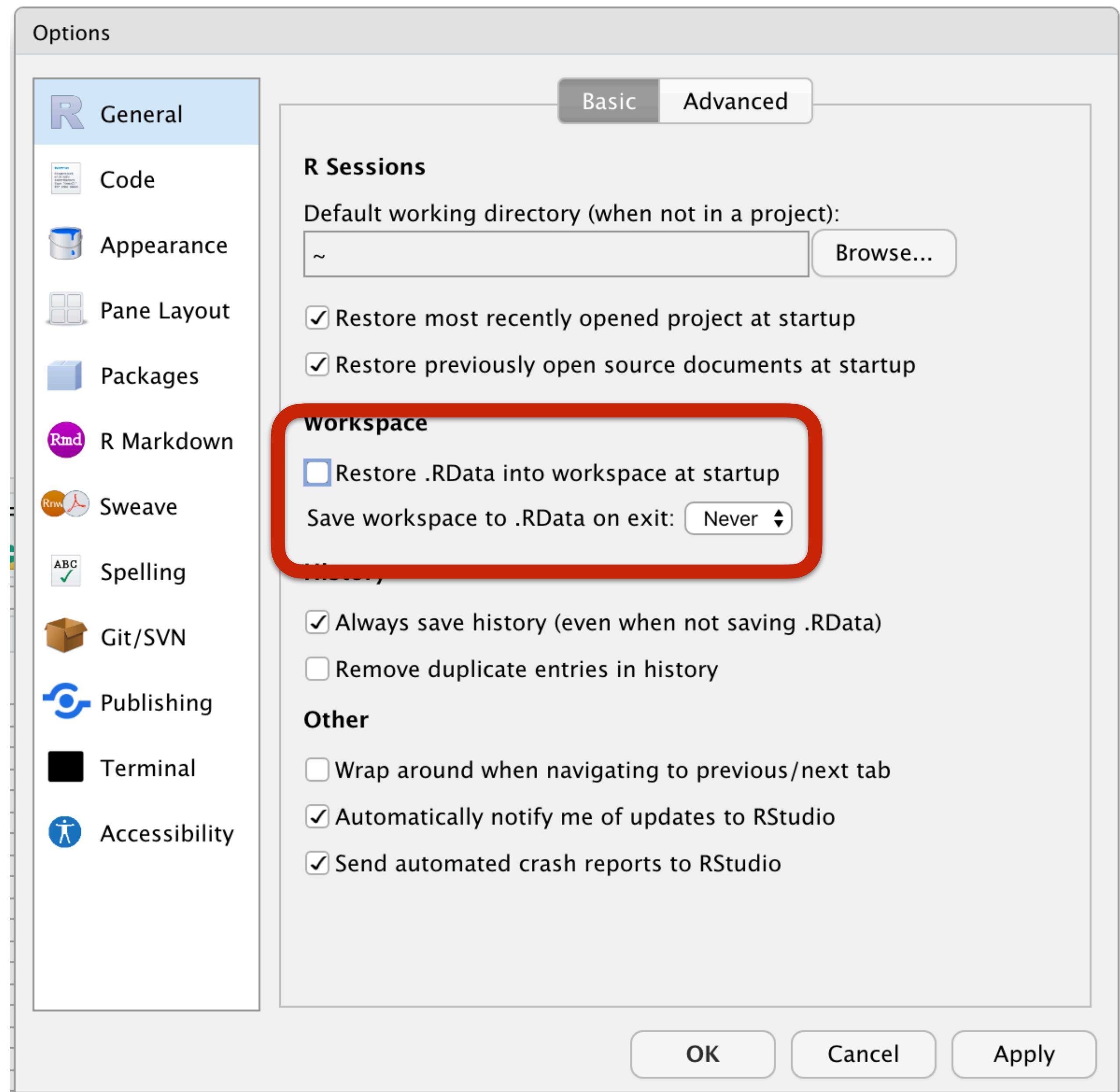
# 1. Lorem Ipsum

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.

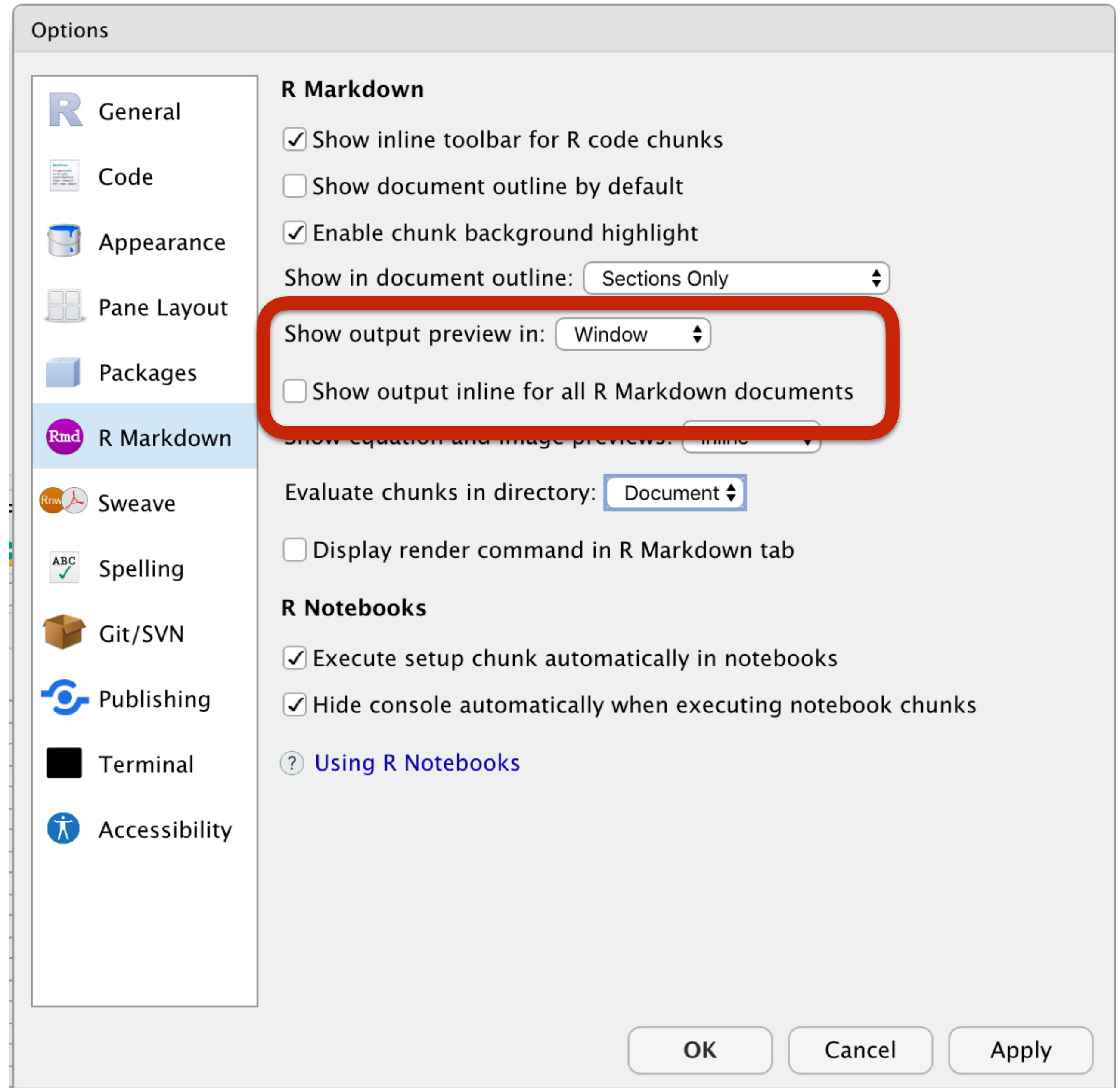


RStudio will do all the work for you when it comes to processing your document—i.e., getting it from plain-text Rmd to HTML, Word, or PDF.

Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.



In general,  
your code is  
what's "real"  
in your project,  
not the objects  
you create.



Consider not  
showing your  
output inline

**GETTING  
ORIENTED**

# The Tidyverse

```
library(tidyverse)
```

```
Loading tidyverse: ggplot2
```

```
Loading tidyverse: tibble
```

```
Loading tidyverse: tidyr
```

```
Loading tidyverse: readr
```

```
Loading tidyverse: purrr
```

```
Loading tidyverse: dplyr
```

- ◀ Draw graphs
- ◀ Nicer data tables
- ◀ Tidy your data
- ◀ Get data into R
- ◀ Cool functional programming stuff
- ◀ Action verbs for manipulating data

# Course-Specific Library

```
library(socviz)
```

# CODE YOU CAN TYPE AND RUN

```
## Inside chunks of code, lines beginning with  
## the hash character are comments  
my_numbers <- c(1, 1, 4, 1, 1, 4, 1)
```

## OUTPUT

```
my_numbers
```

```
## [1] 1 1 4 1 1 4 1
```

What R Looks Like

**FOUR  
THINGS  
TO KNOW  
ABOUT R**

# 1: Everything has a Name

my\_numbers

data

p

Some names are forbidden

FALSE TRUE Inf

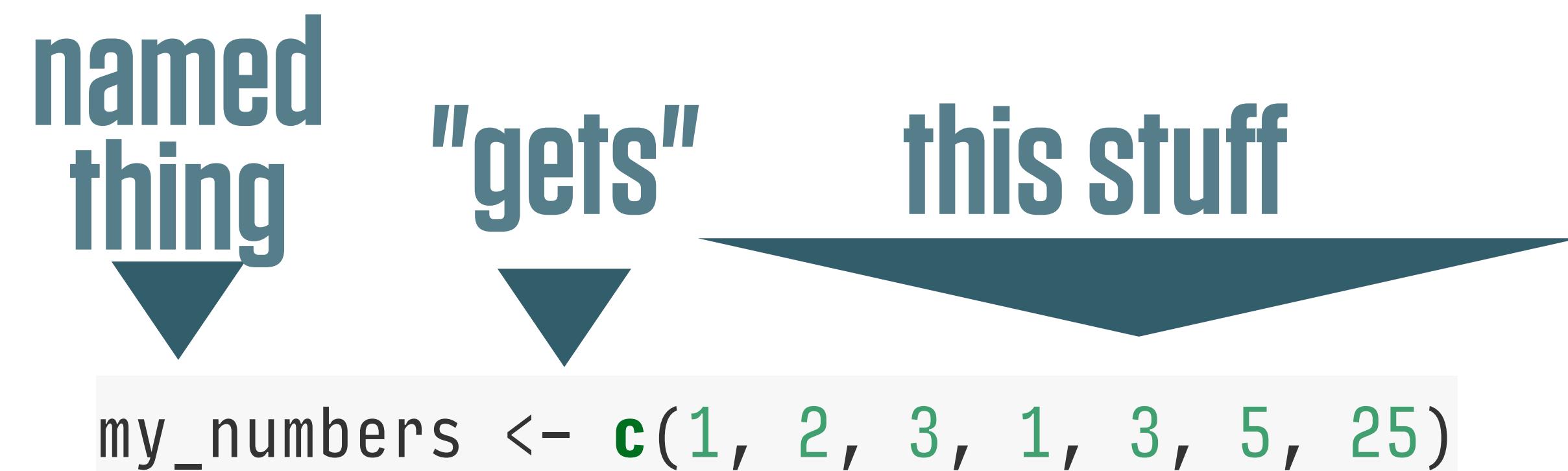
for if break

function

# 2. Everything is an Object

```
> letters  
## [1] "a" "b" "c" "d" "e" "f" "g" "h" "i" "j" "k" "l" "m" "n" "o" "p" "q" "r" "s"  
[20] "t" "u" "v" "w" "x" "y" "z"
```

You **create** objects by  
**assigning** a thing to a name



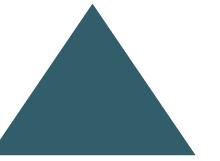
The diagram illustrates the process of creating an R object. Three teal-colored text labels are positioned above a line of R code. Arrows point from each label to a specific part of the code:

- A downward arrow points from the label "named thing" to the variable name "my\_numbers".
- A downward arrow points from the label "'gets'" to the assignment operator "<-".
- A large, wide downward arrow points from the label "this stuff" to the vector elements "c(1, 2, 3, 1, 3, 5, 25)".

```
my_numbers <- c(1, 2, 3, 1, 3, 5, 25)
```

# You create objects by assigning a thing to a name

```
my_numbers <- c(1, 2, 3, 1, 3, 5, 25)
```



The assignment operator performs the action of creating objects. Use a keyboard shortcut to type it:

option - Mac

alt - Windows

# 3. You do things with named objects using functions and operators



`c()` is a function that takes comma-separated numbers or strings and joins them together into a vector

# Functions

take arguments,  
perform actions,  
produce outputs

Functions have parentheses  
at the end of their name.  
This is where the inputs,  
or **arguments** go.

`mean()`

“Take this object ...”

`mean(x = my_numbers)`

Named argument. “Calculate the mean of what, please?”  
These names are internal to functions.

# Functions

take arguments,  
perform actions,  
produce outputs

```
mean(my_numbers)
```



If you just write the name of the input,  
R assigns it to the function's arguments  
in order. Look at the function's help page  
to see the order it expects its arguments.

# You can assign a function's output to a named object

```
my_summary <- summary(my_numbers)
```

```
my_sd <- sd(my_numbers)
```

```
my_summary
```

```
my_sd
```

# Objects you create exist until you overwrite or delete them

```
rm(my_numbers)
```

```
my_numbers
```

```
my_numbers <- c(1, 2, 3, 1, 3, 5, 25)
```

# Objects come in **types** and **classes**

```
class(my_numbers)
```

Vectors

numeric

character

factor

Arrays

matrix

data.frame

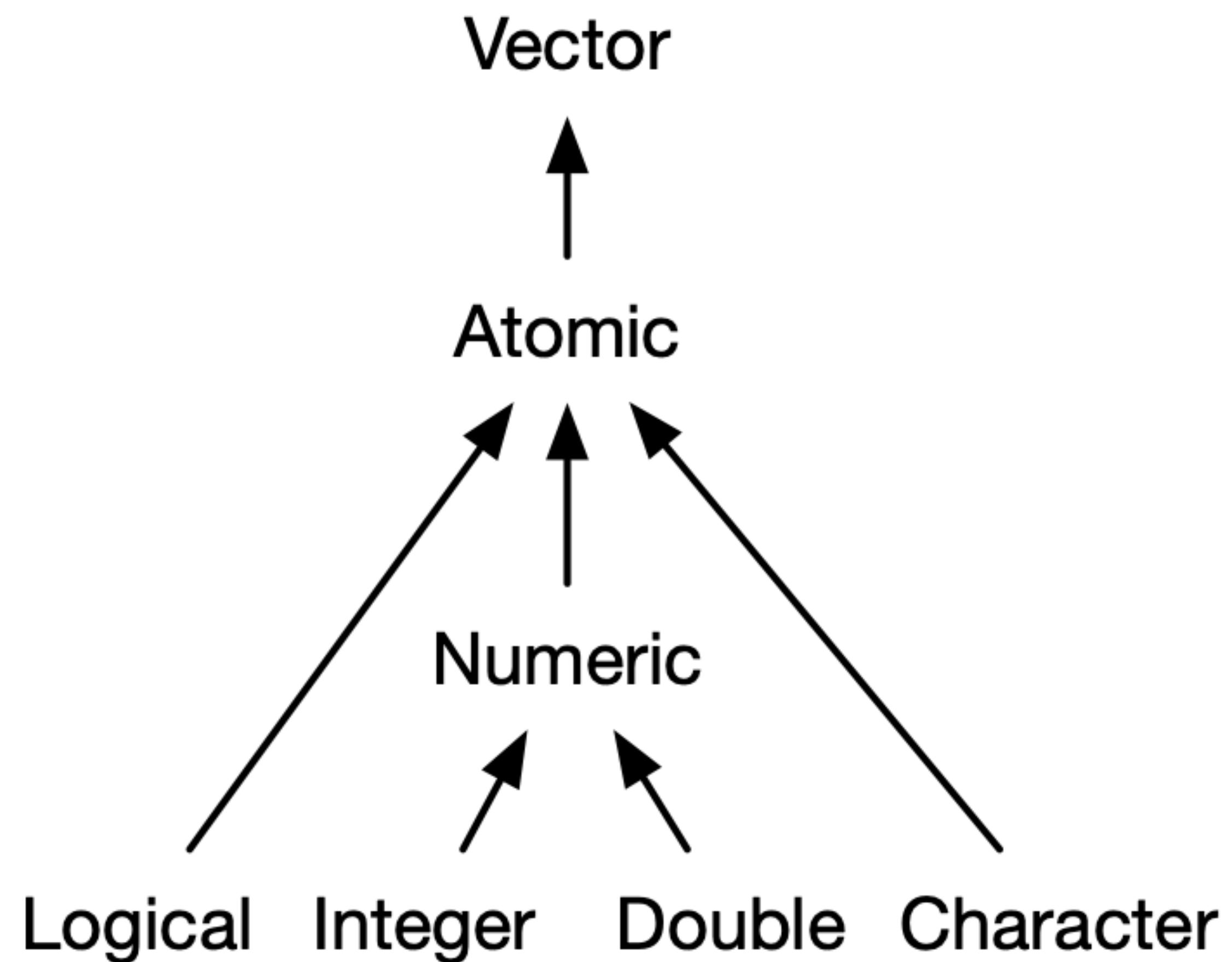
tibble

Models

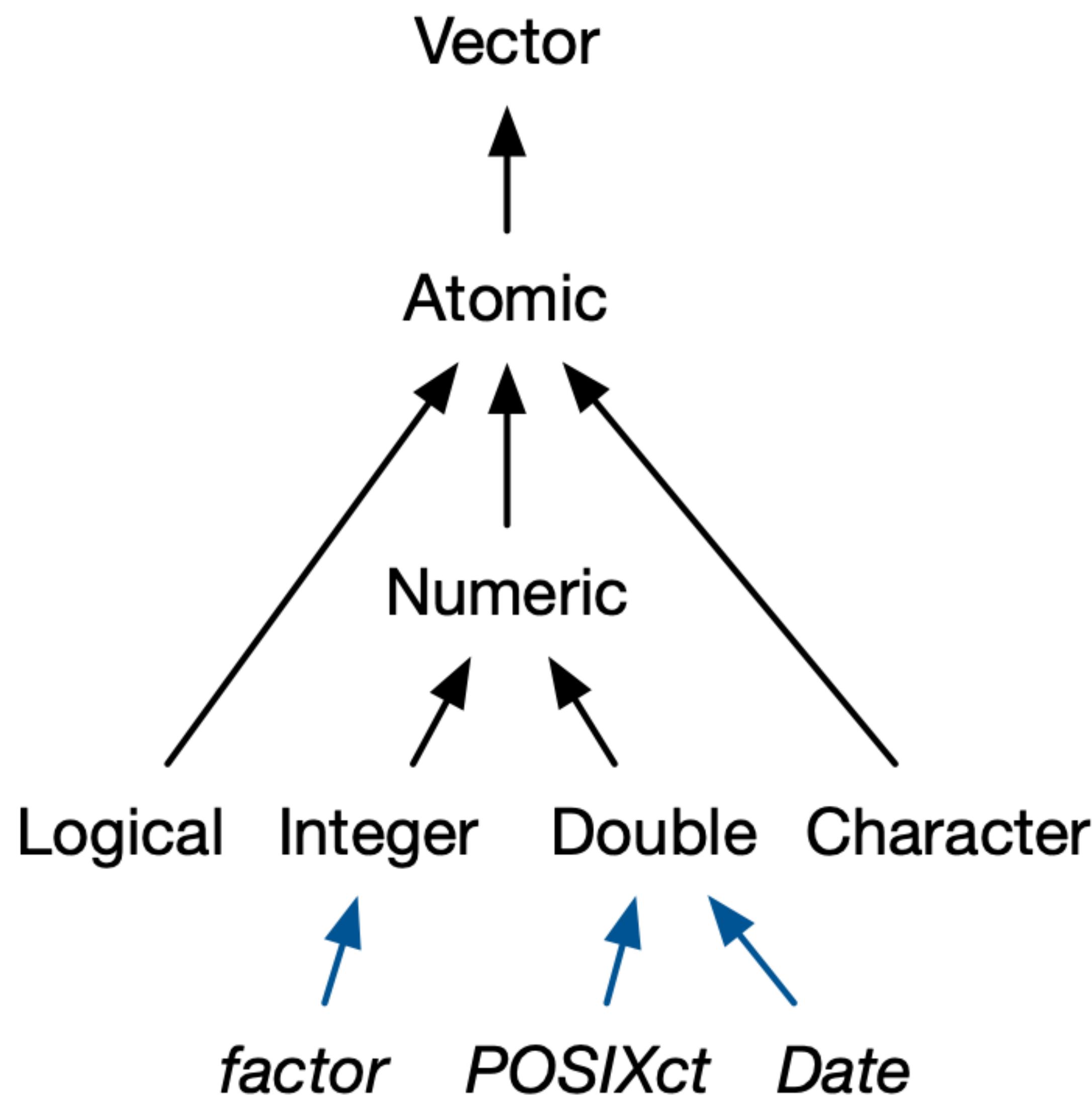
lm

glm

# Objects come in **types** and **classes**



# Objects come in **types** and **classes**



# Things to try on Objects

```
class(my_numbers)  
table(my_numbers)
```

◀ Notice that these  
are functions

```
x <- c(my_numbers, 5)  
y <- c(my_numbers, "hello")
```

◀ How do x and  
y differ?

```
mean(c(my_numbers, my_numbers))
```

◀ Functions can be  
nested, and will be  
evaluated from the  
inside out.

# Some operators

`<-` or `=`

Assignment ("gets")

`+, -, *, /, ^`

Arithmetic

`<, >, <=, >=, ==, !=`

Relational

`&, &&, |, ||, !`

Logical

`%*%, %in%, %>%`

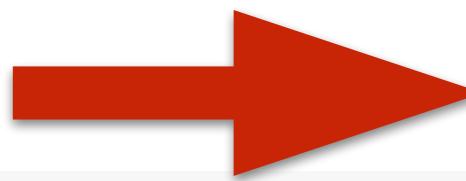
Special

# The pipe operator

%>%

"and then"

```
mean(my_numbers)
```



```
my_numbers %>% mean()
```

```
round(mean(my_numbers))
```



```
my_numbers %>% mean() %>% round()
```

This will be very convenient later on

# 4. R will be Frustrating

We're going to be joining a lot of objects and functions together

```
ggplot(data = mpg,  
       mapping = aes(x = displ, y = hwy)) +  
  geom_point()
```

"+"  
goes  
here

```
ggplot(data = mpg, mapping = aes(x = displ, y = hwy))  
+ geom_point()
```

not here

```
❸ 39 p <- ggplot(data = gapminder  
40                               mapping = aes(x = gdpPercap,  
41                                         y = lifeExp))
```

# 4. R will be Frustrating

RStudio tries its best to help.  
Learn to attend to what it's trying to tell you

```
38  ````{r 03-make-a-plot-4}
✖ 39 p <- ggplot(data = gapminder
  40           mapping = aes(x = gdpPercap,
  41                         y = lifeExp))
```

```
38  ````{r 03-make-a-plot-4}
✖ 39 p <- ggplot(data = gapminder
  expected ',' after expression
  40           mapping = aes(x = gdpPercap,
  41                         y = lifeExp))
```

```
39 p <- ggplot(data = gapminder,
  40               mapping = aes(x = gdpPercap,
  41                             y = lifeExp))|
```

```
39 p <- ggplot(data = gapminder,
  40               mapping = aes(x = gdpPercap,
  41                             y = lifeExp)))
```

```
39 p <- ggplot(data = gapminder,
  40               mapping = aes(x = gdpPercap,
  41                             y = lifeExp)))
✖ 41 unexpected token ')'
42 |
```

**LET'S GO**

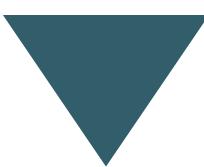
```
library(gapminder)
```

# gapminder

# A tibble: 1,704 x 6

|    | country     | continent | year  | lifeExp | pop      | gdpPerCap |
|----|-------------|-----------|-------|---------|----------|-----------|
|    | <fctr>      | <fctr>    | <int> | <dbl>   | <int>    | <dbl>     |
| 1  | Afghanistan | Asia      | 1952  | 28.801  | 8425333  | 779.4453  |
| 2  | Afghanistan | Asia      | 1957  | 30.332  | 9240934  | 820.8530  |
| 3  | Afghanistan | Asia      | 1962  | 31.997  | 10267083 | 853.1007  |
| 4  | Afghanistan | Asia      | 1967  | 34.020  | 11537966 | 836.1971  |
| 5  | Afghanistan | Asia      | 1972  | 36.088  | 13079460 | 739.9811  |
| 6  | Afghanistan | Asia      | 1977  | 38.438  | 14880372 | 786.1134  |
| 7  | Afghanistan | Asia      | 1982  | 39.854  | 12881816 | 978.0114  |
| 8  | Afghanistan | Asia      | 1987  | 40.822  | 13867957 | 852.3959  |
| 9  | Afghanistan | Asia      | 1992  | 41.674  | 16317921 | 649.3414  |
| 10 | Afghanistan | Asia      | 1997  | 41.763  | 22227415 | 635.3414  |

Named thing gets ...



... using these  
arguments

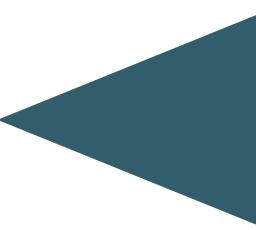
```
p <- ggplot(data = gapminder,  
             mapping = aes(x = gdpPerCap,  
                            y = lifeExp))
```

... the output of  
this function ...



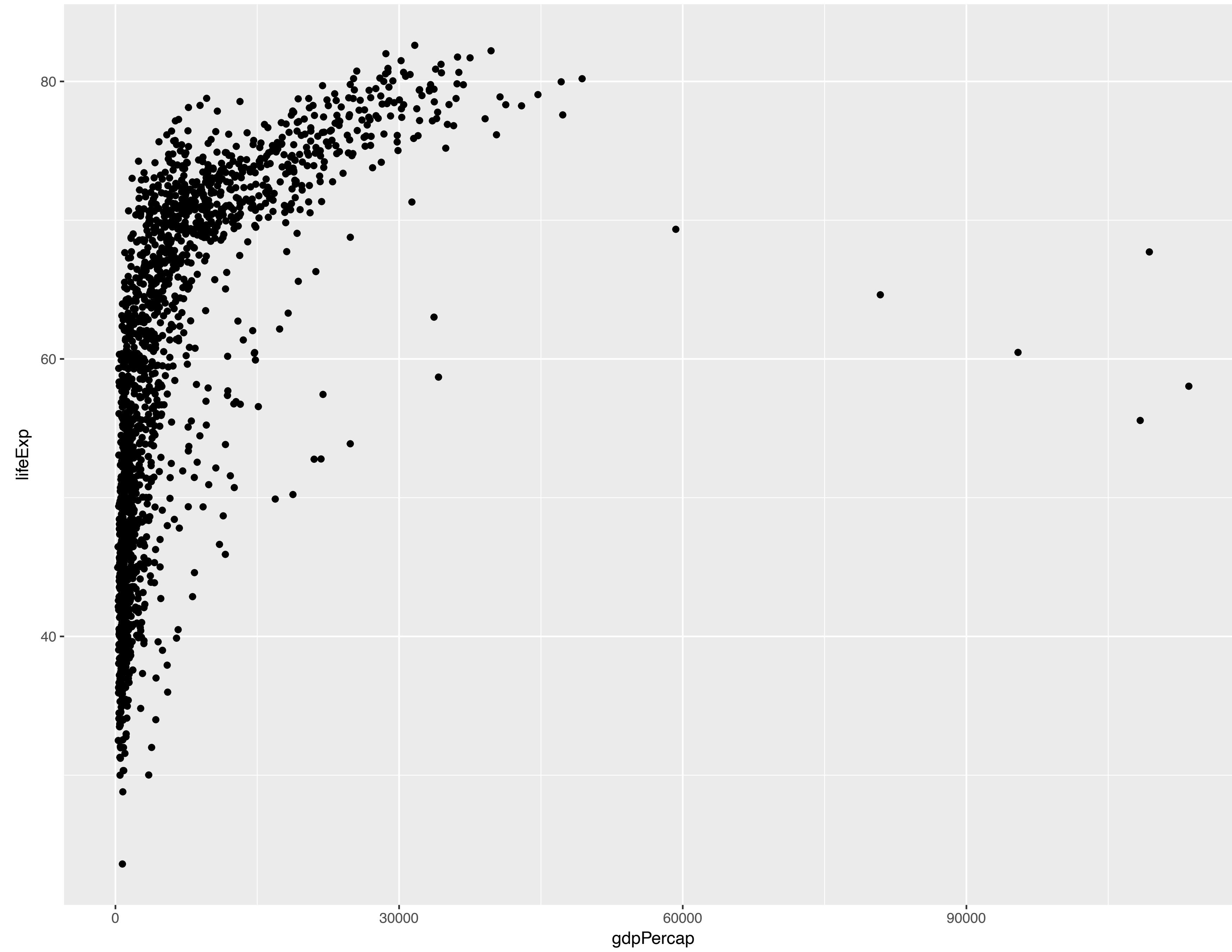
p

```
p + geom_point()
```

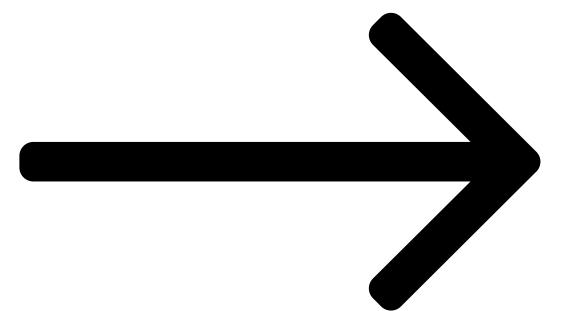
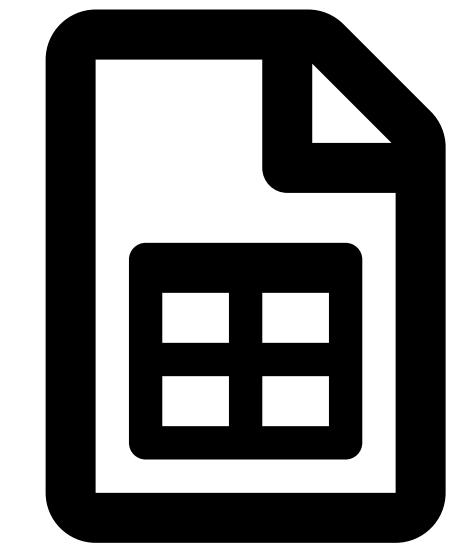


Objects created by  
ggplot() are unusual in  
that you can “add”  
things to them, and  
they will work as  
though you wrote all  
the code at once.

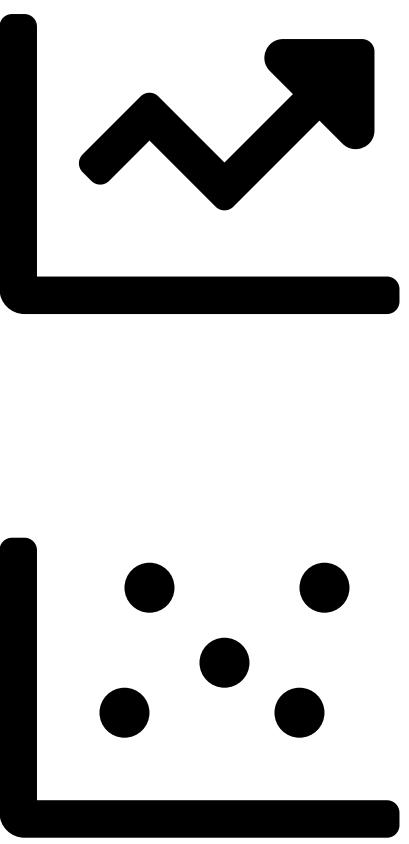
```
p <- ggplot(data = gapminder,  
             mapping = aes(x = gdpPercap,  
                            y = lifeExp))  
  
p + geom_point()
```



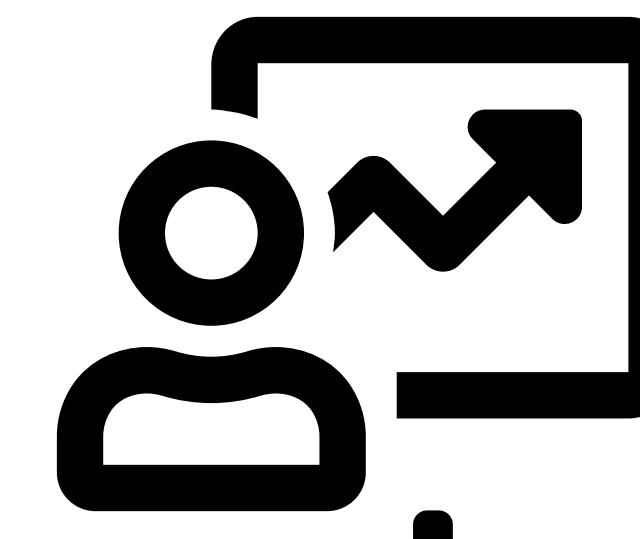
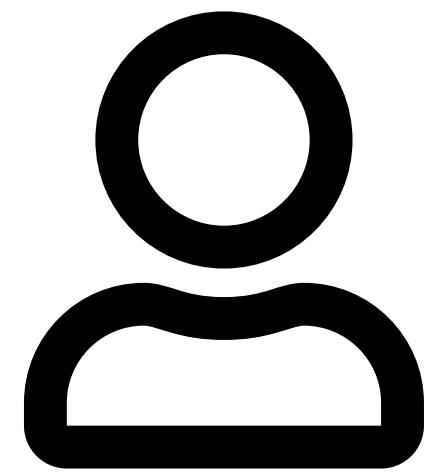
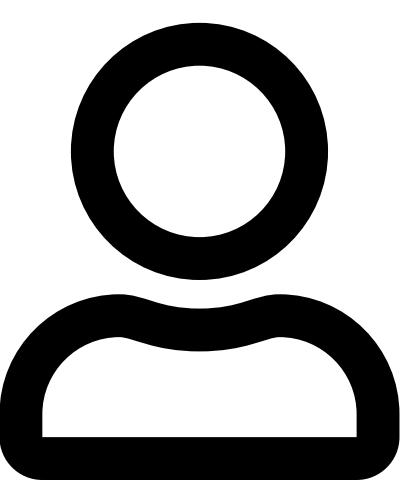
**What exactly is a  
graph, anyway?**



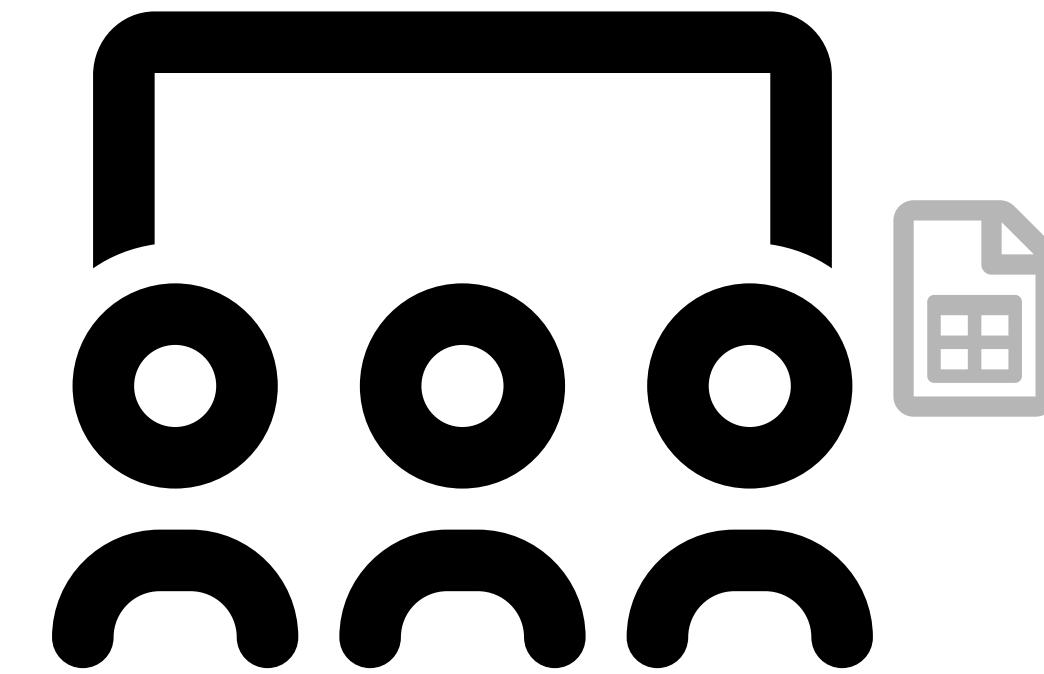
Encode



Represent



Decode



Interpret

**And what is R doing?**

R objects are just  
lists of stuff to use,  
or things to do

```
p <- ggplot(data = gapminder,  
             mapping = aes(x = gdpPercap,  
                            y = lifeExp))  
  
p + geom_point()
```



p

## Data

|   | <fctr>      | <fctr> | <int> |
|---|-------------|--------|-------|
| 1 | Afghanistan | Asia   | 1952  |
| 2 | Afghanistan | Asia   | 1957  |
| 3 | Afghanistan | Asia   | 1962  |
| 4 | Afghanistan | Asia   | 1967  |
| 5 | Afghanistan | Asia   | 1972  |
| 6 | Afghanistan | Asia   | 1977  |
| 7 | Afghanistan | Asia   | 1982  |
| 8 | Afghanistan | Asia   | 1987  |
| 9 | Afghanistan | Asia   | 1992  |

## Mappings

— Represent or Map

“gdpPercap” using the x axis

— Represent or Map

“gdpPercap” using the y axis

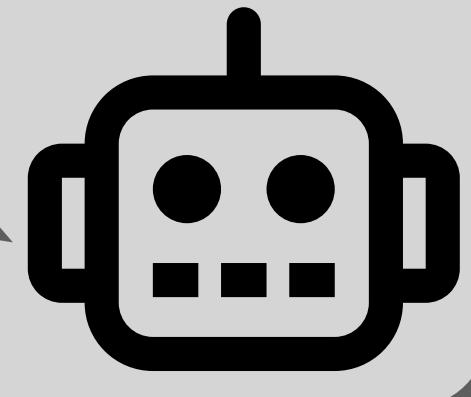
— Represent or Map

“continent” using colors

Just deal with these for me, robot

coordinates  
scales  
theme  
plot\_env

bleep bloop



## Labels

— Label the x axis “GDP per Capita”

— Label the y axis “Life Expectancy”

— Label the color key “Continent”

| Name          | Type                              | Value                                 |
|---------------|-----------------------------------|---------------------------------------|
| ▶ p           | list [9] (S3: gg, ggplot)         | List of length 9                      |
| ▶ data        | list [1704 x 6] (S3: tbl_df, tbl, | A tibble with 1704 rows and 6 columns |
| layers        | list [0]                          | List of length 0                      |
| ▶ scales      | environment [1] (S3: ScalesLis    | <environment: 0x7f8f08c1e010>         |
| ▶ mapping     | list [3] (S3: uneval)             | List of length 3                      |
| theme         | list [0]                          | List of length 0                      |
| ▶ coordinates | environment [5] (S3: CoordCa      | <environment: 0x7f8f08c27b40>         |
| ▶ facet       | environment [2] (S3: FacetNul     | <environment: 0x7f8f08c55210>         |
| ▶ plot_env    | environment [6]                   | <environment: R_GlobalEnv>            |
| ▶ labels      | list [3]                          | List of length 3                      |

# Peek in using the Object Inspector

```
p <- ggplot(data = gapminder,  
             mapping = aes(x = gdpPercap,  
                           y = lifeExp))  
  
p + geom_point()
```