Data Science for WASH

An opportunity for the sector

Lars Schöbitz

2021-03-10

Welcome! 

Lars Schöbitz

Environmental Engineer
WASH Consultant
Instructor for Data Science with R

Georges Mikhael

Senior Urban Sanitation Specialist
Consultant at Aguaconsult, UK
Novice R user

Do you sometimes wonder:

- Where people defecate in the open within a city? And if there are water bodies nearby?
- Who lives downstream of contaminated water bodies? And what the prevalence of diarrhea is in those communities?
- Whether access to safe drinking water decreases the rate of diarrheal disease?

If so, then you might also wonder:

- How you to use your data, in combination with public open data, to answer these questions
- How you can read, transform and analyse data from different sources, and of different structures

Artwork from @juliesquid for @openscapes (illustrated by @allison_horst).

Relevance of the topic

- Deriving actions from data plays a key role in every organisation
- Demand on WASH professionals to analyse data and share knowledge is increasing
- Little attention given to the resources and competencies needed to satisfy this demand

Data Science - FAQ

Q: What is data science?
A: In data science you turn raw data into understanding, insight and knowledge

Q: What is R?
A: It's a computing language used for data science

Q: Is Data Science = Statistics
A: No, but they are closely related

Q: Is Data Science = Computer Science
A: No, but many themes are shared

Data Science for WASH - FAQ

Q: Will I learn how to do machine learning?
Nope, not at all. We explore core concepts such as descriptive statistics, data visualisation, and (some) basic modeling

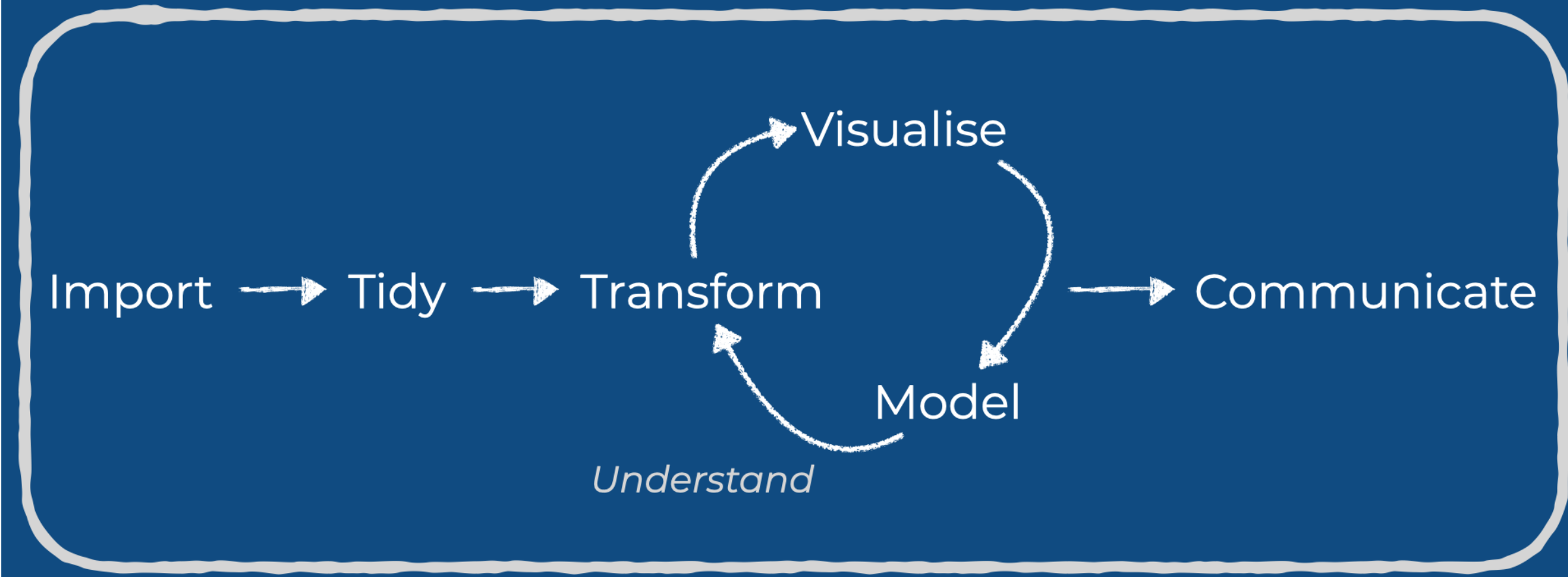
Q: Do I need any prior experience?
A: Nope, my courses are targeted at novice users

Q: What do I need to learn data science?
A: A mindset for openness to change; A good portion of vulnerability; A friendly and open community

Q: And how long does it take to learn data science?
A: A minimum of 3 months with tutoring support and time effort of 12 hours per week

Data Science Lifecycle

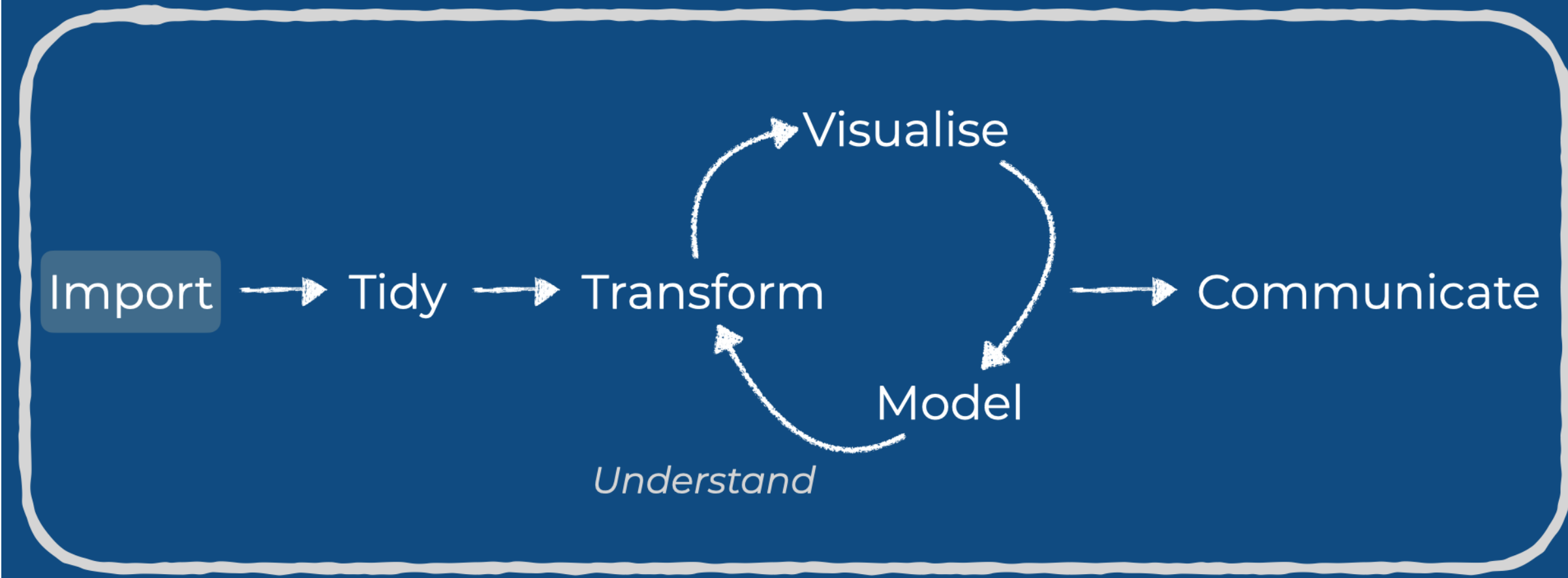
Data Science Lifecycle



From: https://r4ds.had.co.nz/

Data Science Lifecycle

Get your data into R

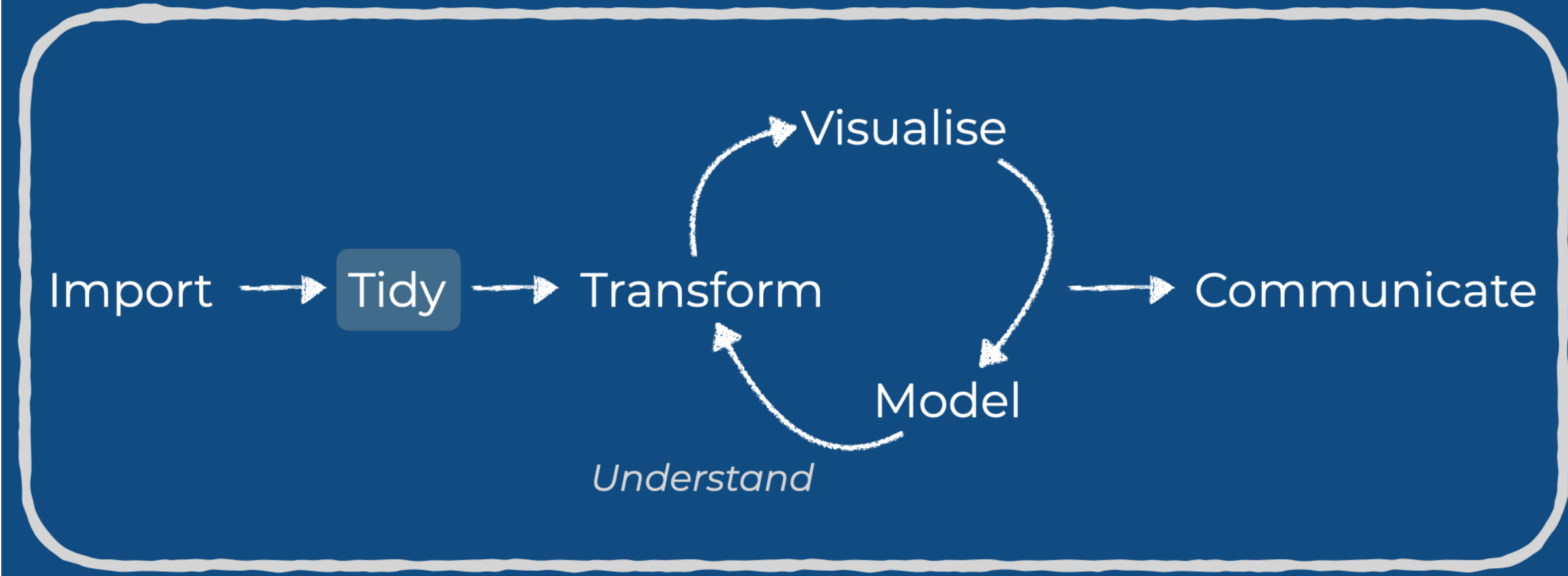


Program

From: https://r4ds.had.co.nz/

Data Science Lifecycle

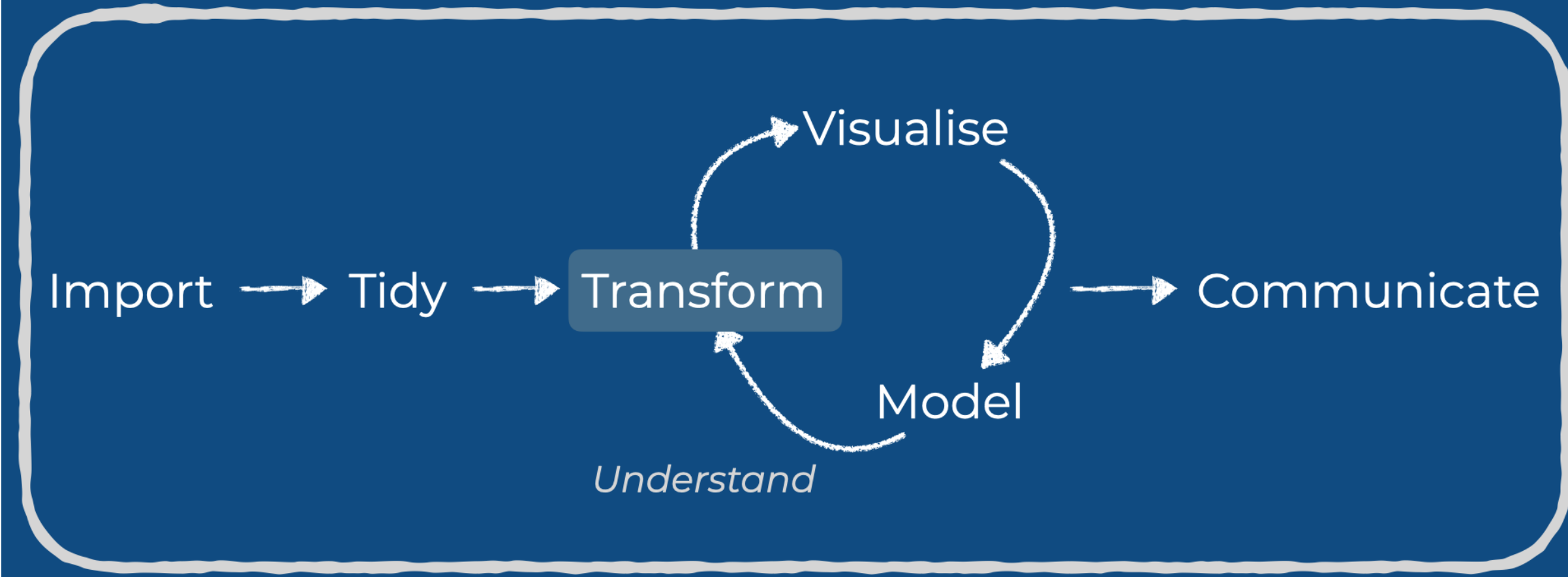
Store your data in a consistent form



From: https://r4ds.had.co.nz/

Data Science Lifecycle

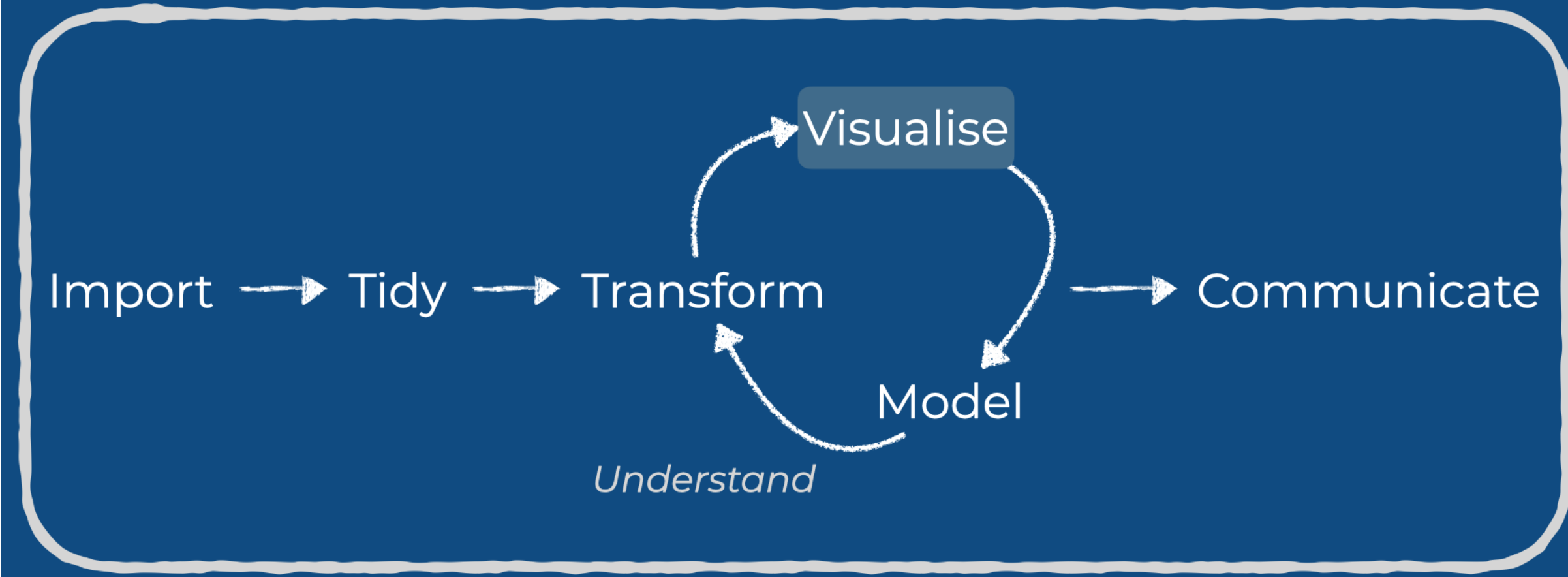
Narrow down + Create new variables + Summary stats



From: https://r4ds.had.co.nz/

Data Science Lifecycle

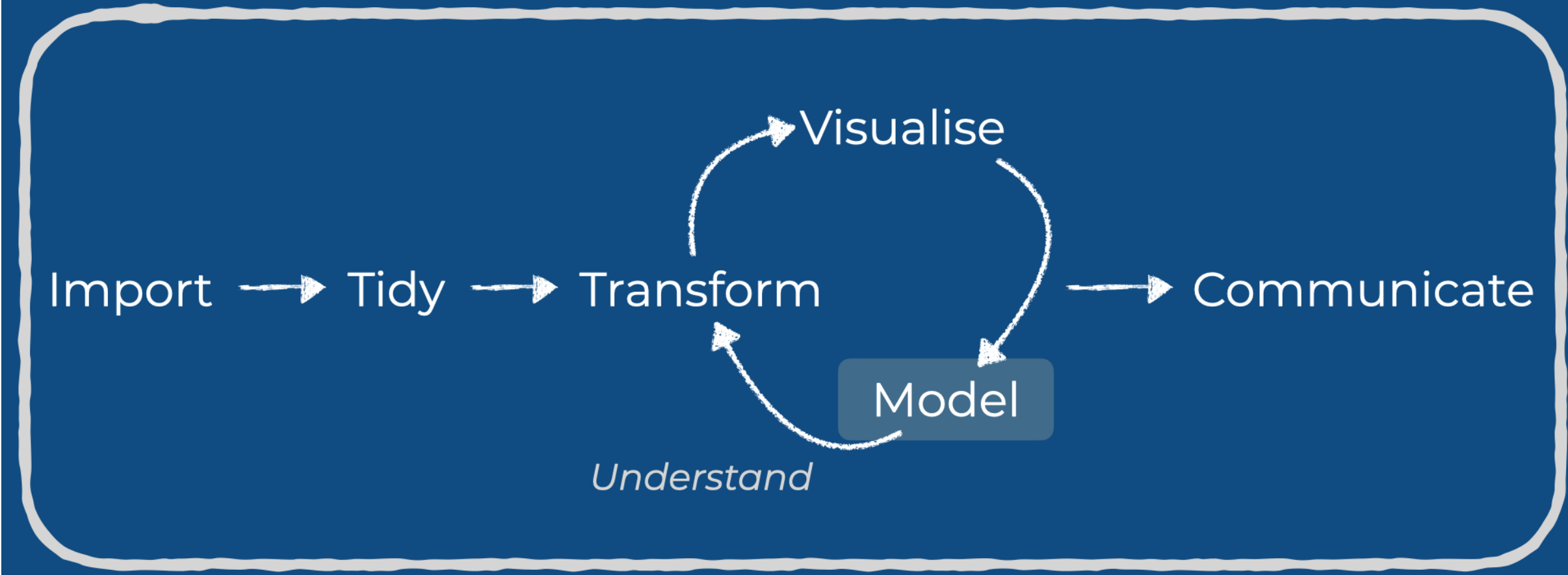
Explore your with visual representations



From: https://r4ds.had.co.nz/

Data Science Lifecycle

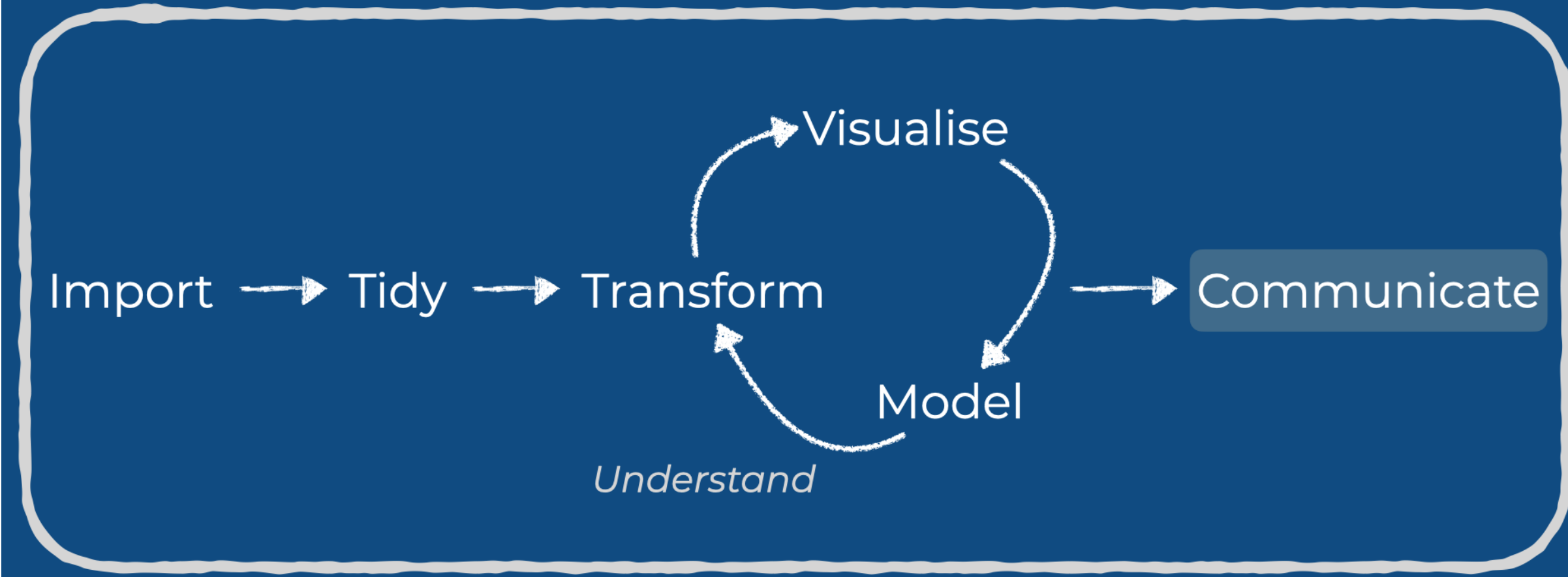
Explore your with visual representations



From: https://r4ds.had.co.nz/

Data Science Lifecycle

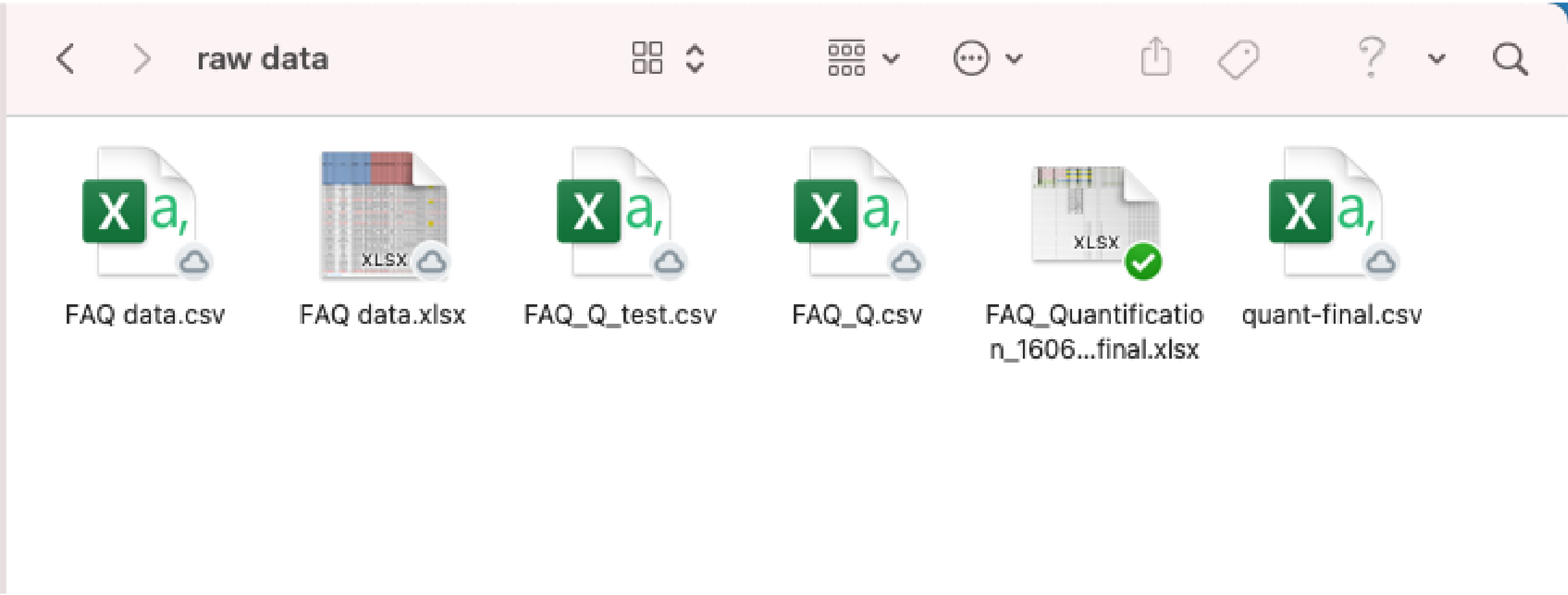
Share your findings with others

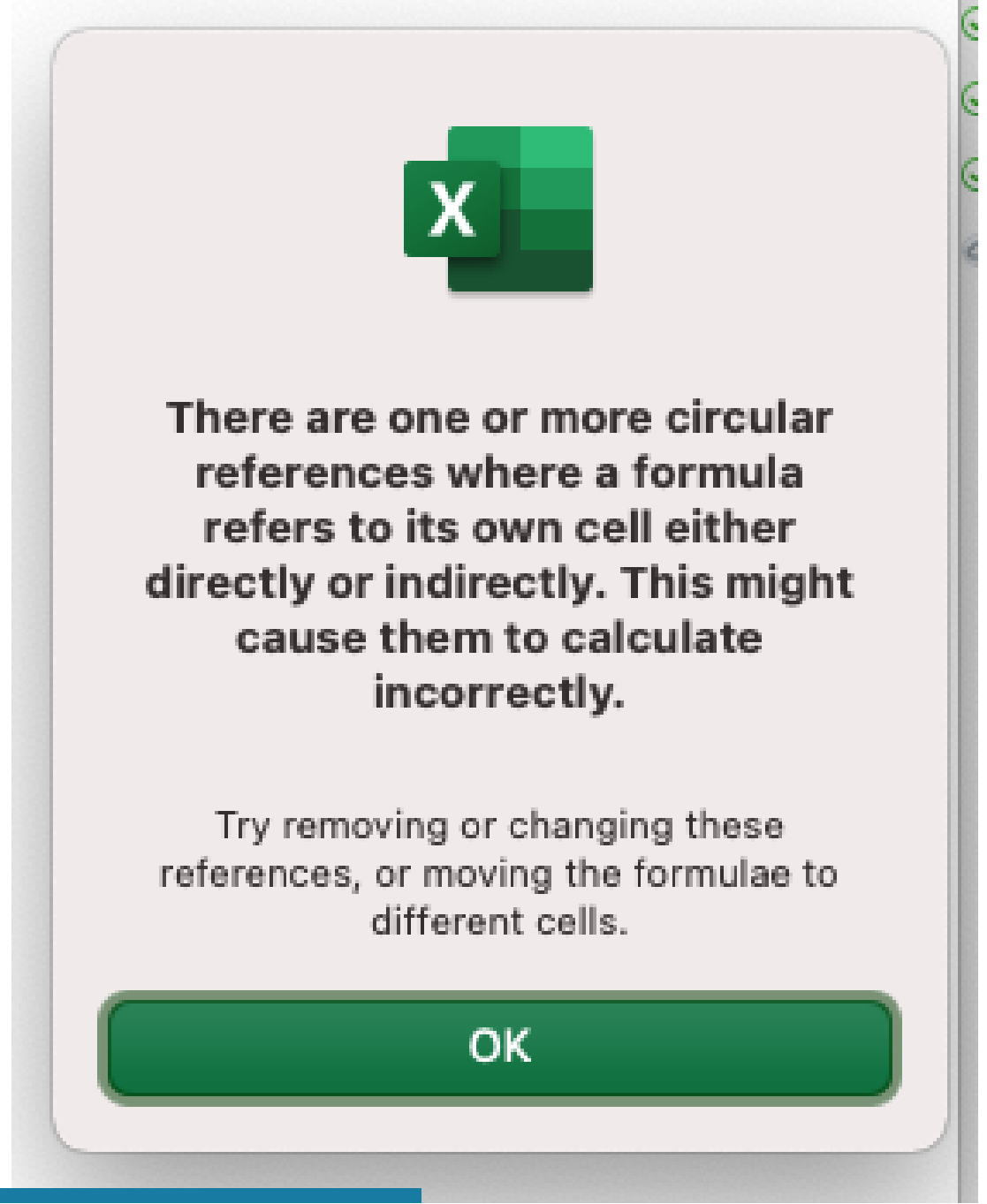


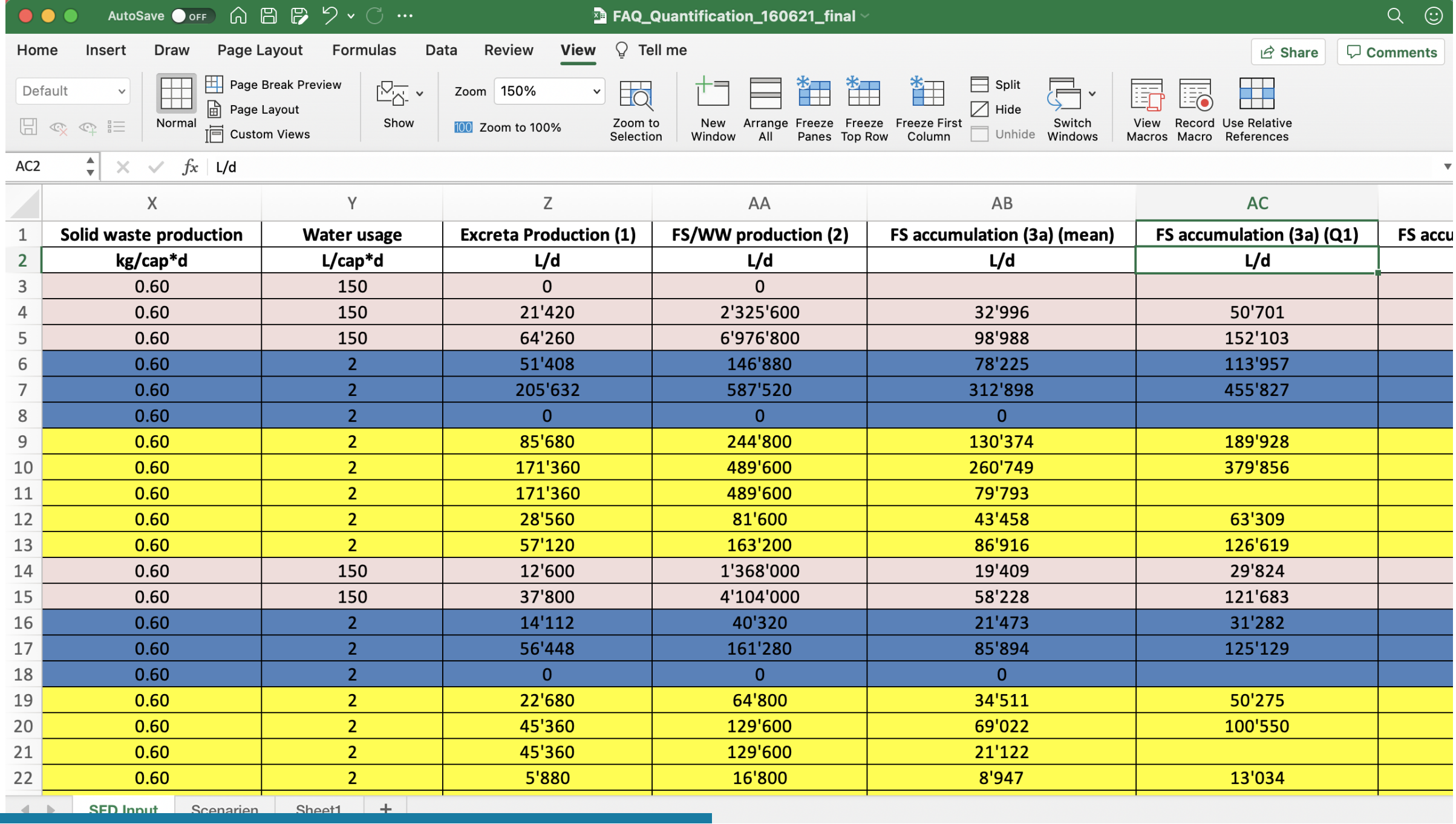
From: https://r4ds.had.co.nz/

My data analysis projects

what it used to look like


raw data
FAQ data.csv
XLSX
FAQ data.xlsx
FAQ_Q_test.csv
FAQ_Q.csv
XLSX
FAQ_Quantification_1606...final.xlsx
quant-final.csv

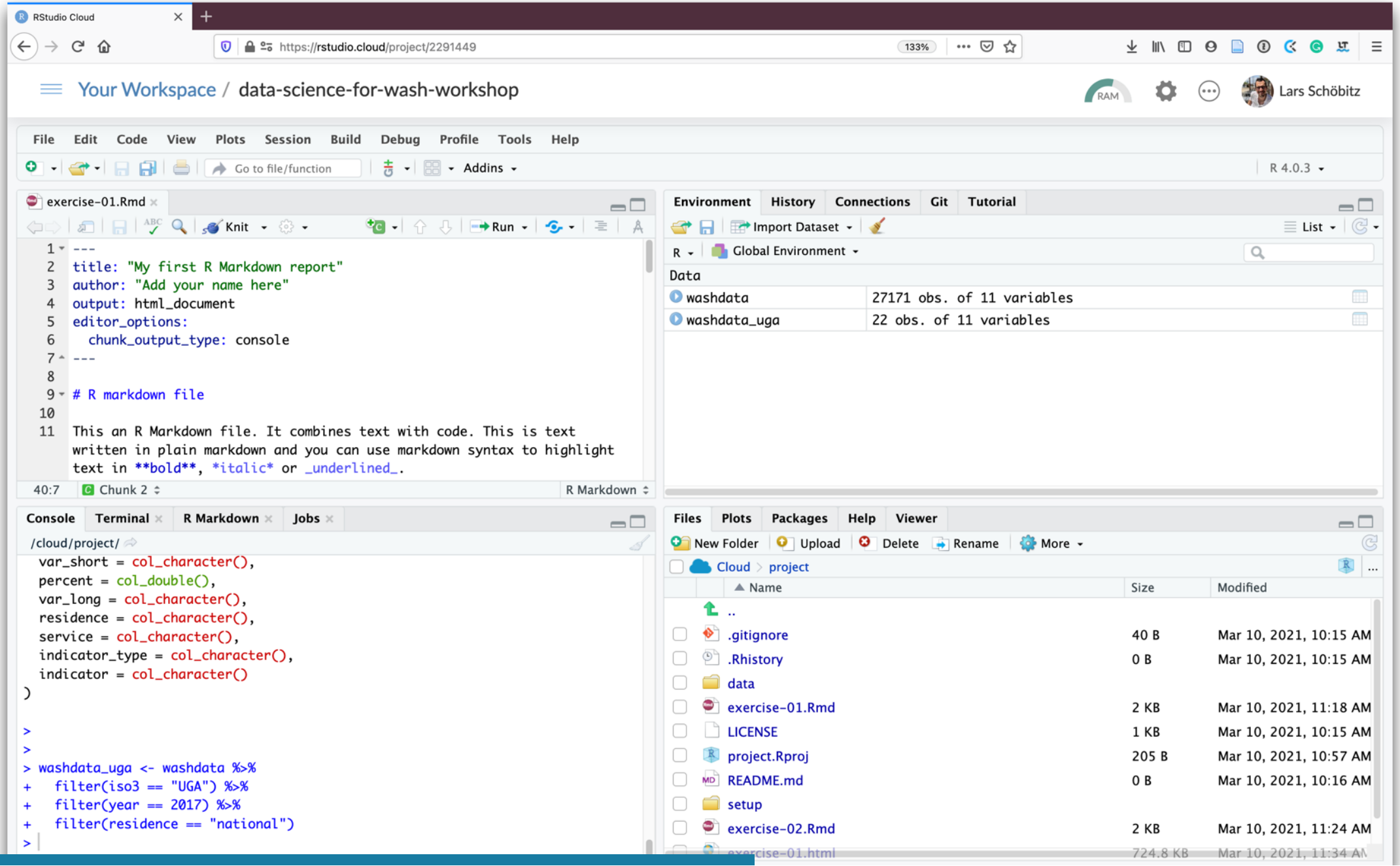

There are one or more circular references where a formula refers to its own cell either directly or indirectly. This might cause them to calculate incorrectly.
Try removing or changing these references, or moving the formulae to different cells.
OK



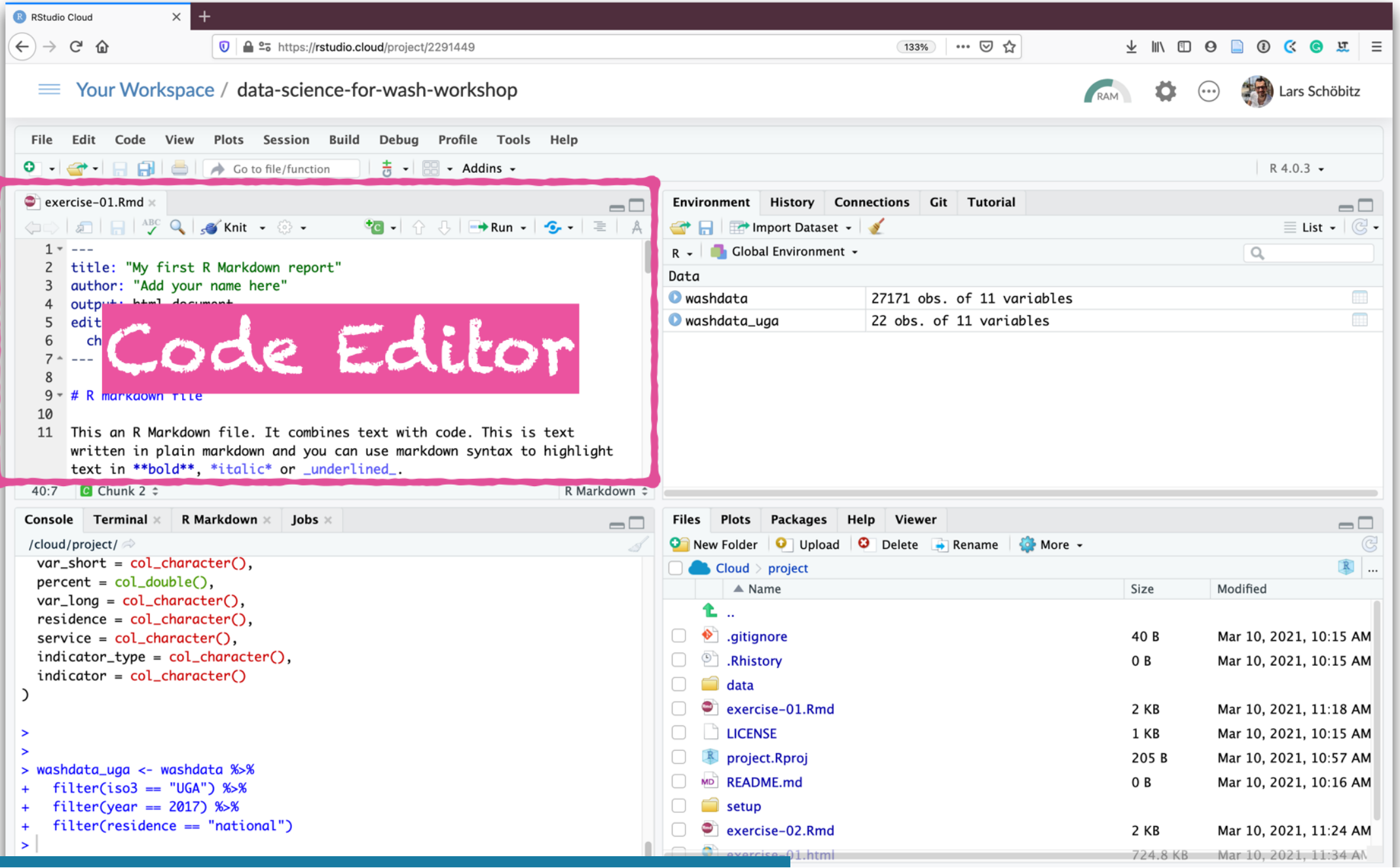
	X	Y	Z	AA	AB	AC
1	Solid waste production	Water usage	Excreta Production (1)	FS/WW production (2)	FS accumulation (3a) (mean)	FS accumulation (3a) (Q1)
2	kg/cap*d	L/cap*d	L/d	L/d	L/d	L/d
3	0.60	150	0	0		
4	0.60	150	21'420	2'325'600	32'996	50'701
5	0.60	150	64'260	6'976'800	98'988	152'103
6	0.60	2	51'408	146'880	78'225	113'957
7	0.60	2	205'632	587'520	312'898	455'827
8	0.60	2	0	0	0	
9	0.60	2	85'680	244'800	130'374	189'928
10	0.60	2	171'360	489'600	260'749	379'856
11	0.60	2	171'360	489'600	79'793	
12	0.60	2	28'560	81'600	43'458	63'309
13	0.60	2	57'120	163'200	86'916	126'619
14	0.60	150	12'600	1'368'000	19'409	29'824
15	0.60	150	37'800	4'104'000	58'228	121'683
16	0.60	2	14'112	40'320	21'473	31'282
17	0.60	2	56'448	161'280	85'894	125'129
18	0.60	2	0	0	0	
19	0.60	2	22'680	64'800	34'511	50'275
20	0.60	2	45'360	129'600	69'022	100'550
21	0.60	2	45'360	129'600	21'122	
22	0.60	2	5'880	16'800	8'947	13'034

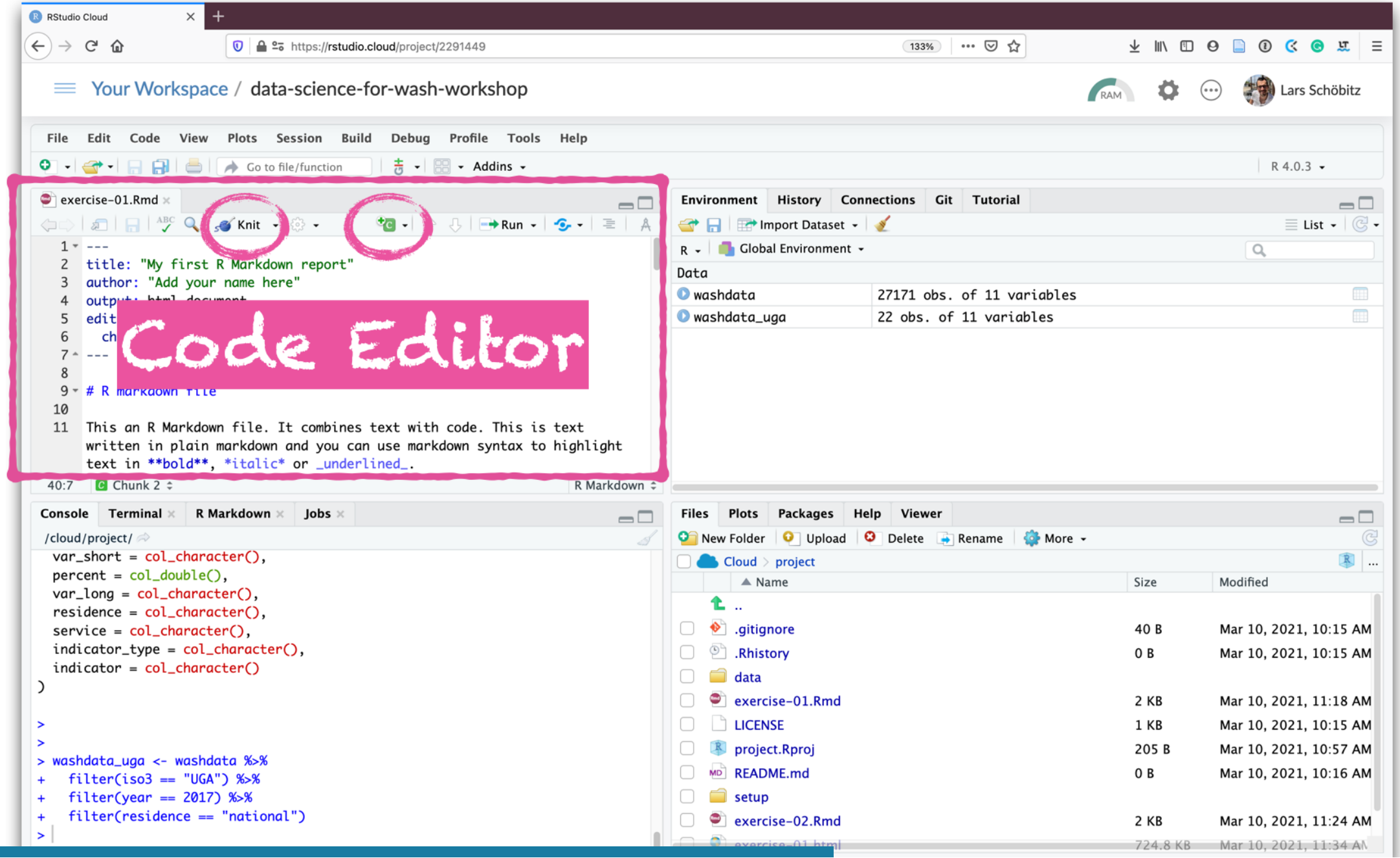
My data analysis projects

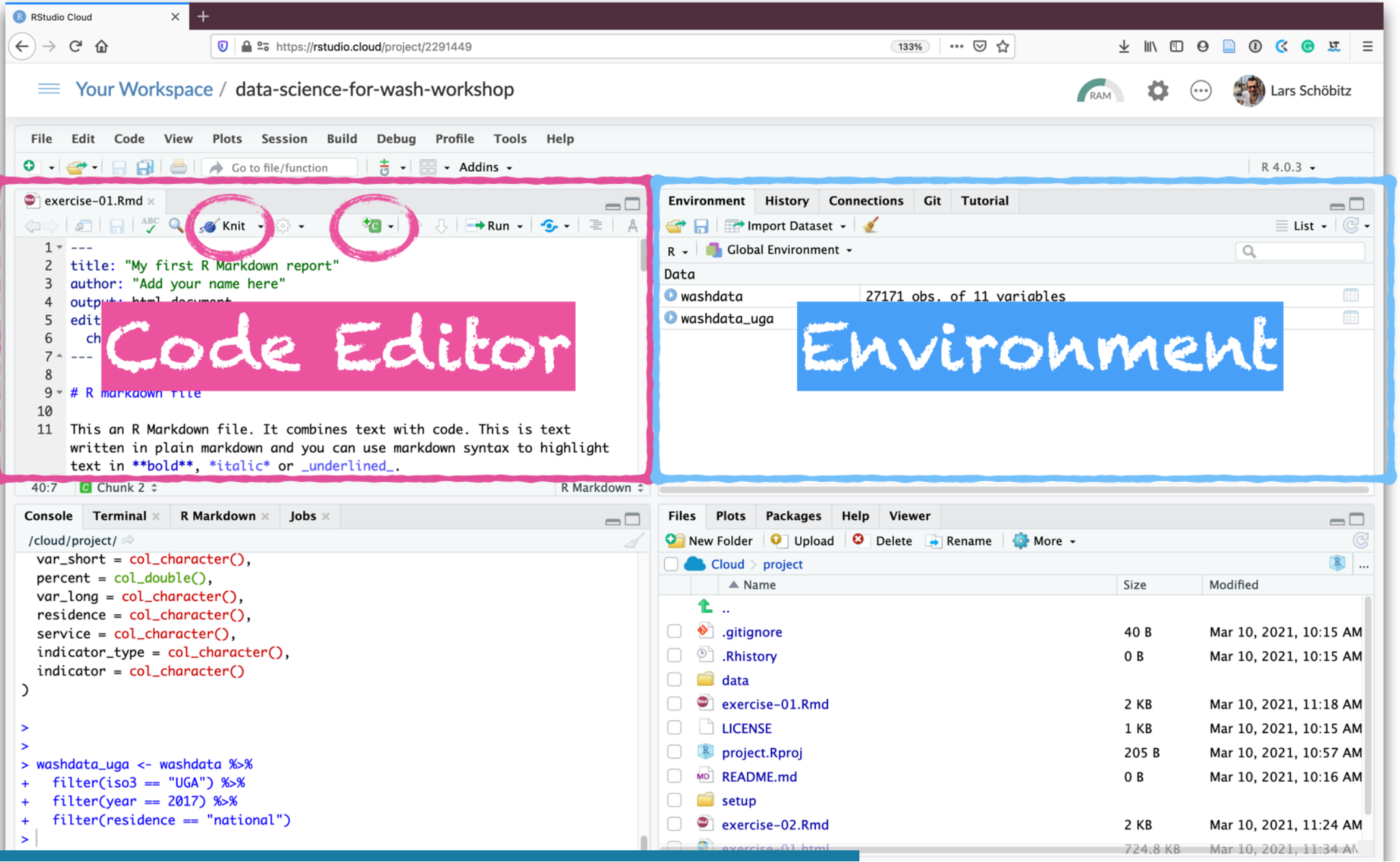
what it looks like now

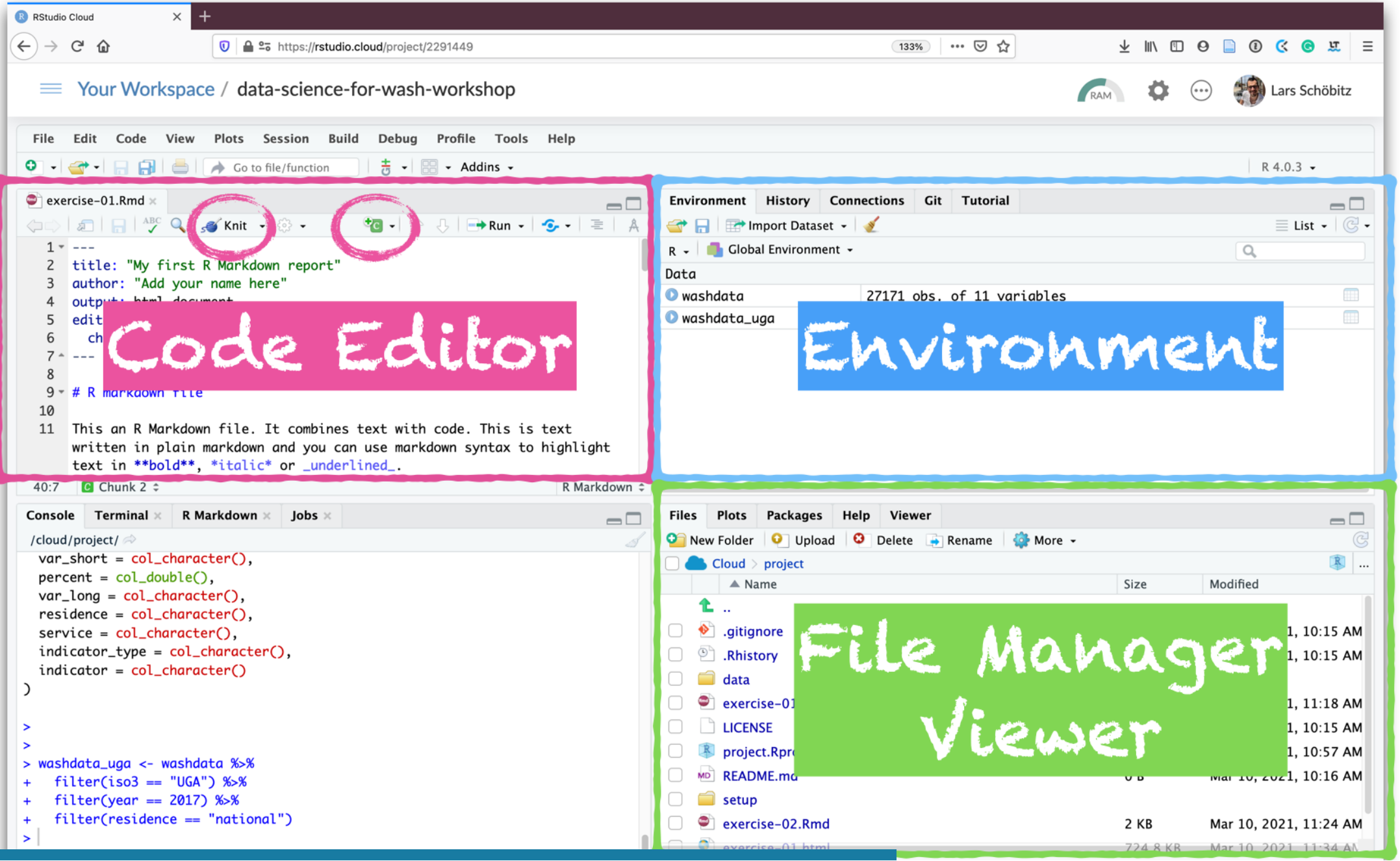

RStudio Cloud
https://rstudio.cloud/project/2291449
133%
Your Workspace / data-science-for-wash-workshop
RAM
Lars Schöbitz
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
R 4.0.3
exercise-01.Rmd
Knit
Run
---
title: "My first R Markdown report"
author: "Add your name here"
output: html_document
editor_options:
chunk_output_type: console

R markdown file
This an R Markdown file. It combines text with code. This is text
written in plain markdown and you can use markdown syntax to highlight
text in **bold**, *italic* or _underlined_.
40:7
Chunk 2
R Markdown
Environment History Connections Git Tutorial
Import Dataset
List
R
Global Environment
Data
washdata 27171 obs. of 11 variables
washdata_uga 22 obs. of 11 variables
Console Terminal R Markdown Jobs
/cloud/project/
var_short = col_character(),
percent = col_double(),
var_long = col_character(),
residence = col_character(),
service = col_character(),
indicator_type = col_character(),
indicator = col_character()
)
>
>
> washdata_uga <- washdata %>%
+ filter(iso3 == "UGA") %>%
+ filter(year == 2017) %>%
+ filter(residence == "national")
>
Files Plots Packages Help Viewer
New Folder Upload Delete Rename More
Cloud project
Name Size Modified
.. .gitignore 40 B Mar 10, 2021, 10:15 AM
.Rhistory 0 B Mar 10, 2021, 10:15 AM
data
exercise-01.Rmd 2 KB Mar 10, 2021, 11:18 AM
LICENSE 1 KB Mar 10, 2021, 10:15 AM
project.Rproj 205 B Mar 10, 2021, 10:57 AM
README.md 0 B Mar 10, 2021, 10:16 AM
setup
exercise-02.Rmd 2 KB Mar 10, 2021, 11:24 AM
exercise-01.html 724.8 KB Mar 10, 2021, 11:34 AM

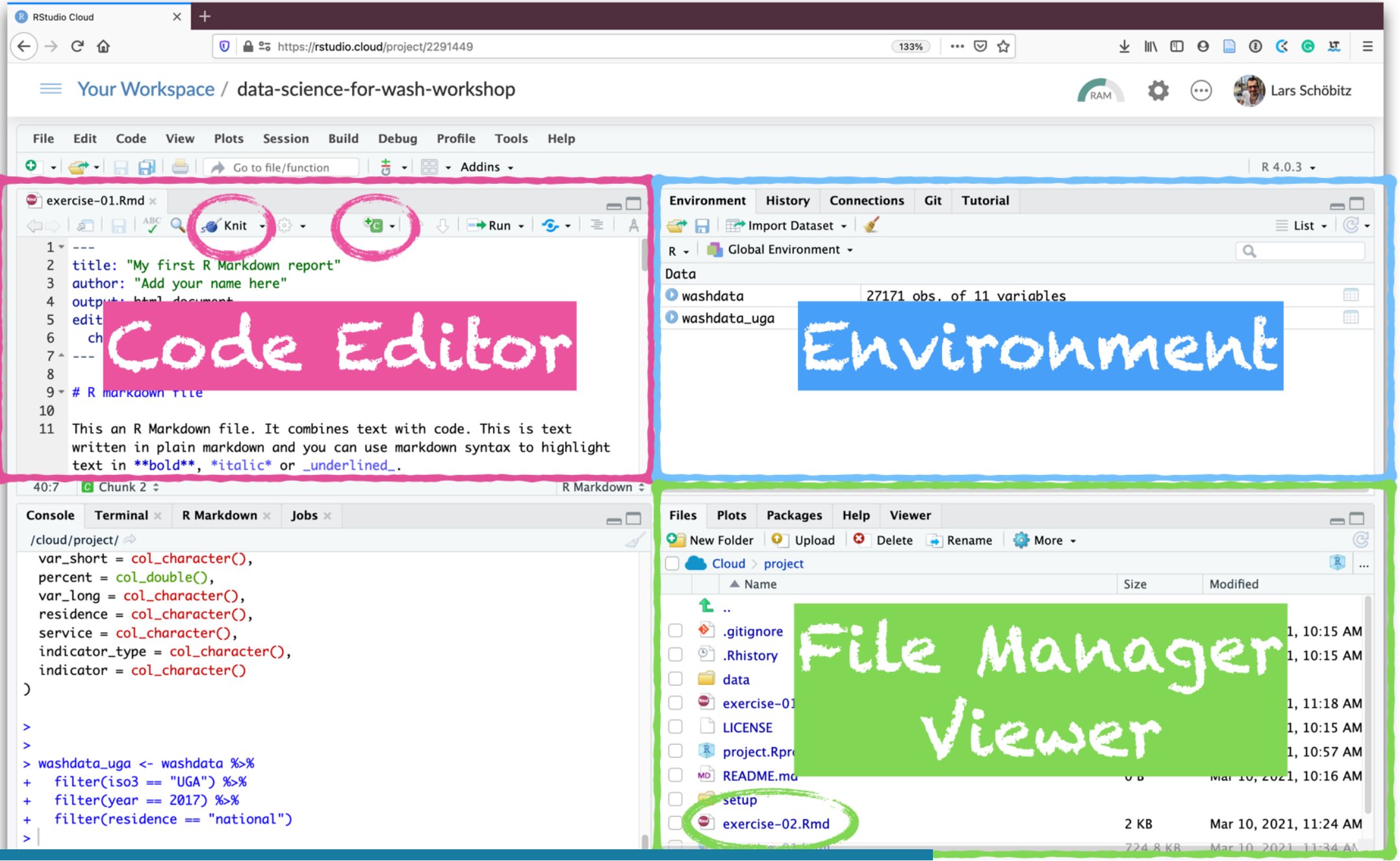

RStudio Cloud
https://rstudio.cloud/project/2291449
Your Workspace / data-science-for-wash-workshop
Lars Schöbitz
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
R 4.0.3
exercise-01.Rmd
Knit
Run
1 ---
2 title: "My first R Markdown report"
3 author: "Add your name here"
4 output: html_document
5 edit
6 ch
7 ---
8
9 # R markdown file
10
11 This an R Markdown file. It combines text with code. This is text
written in plain markdown and you can use markdown syntax to highlight
text in **bold**, *italic* or _underlined_.
Code Editor
40:7
Chunk 2
R Markdown
Environment History Connections Git Tutorial
Import Dataset
List
R
Global Environment
Data
washdata 27171 obs. of 11 variables
washdata_uga 22 obs. of 11 variables
Console Terminal R Markdown Jobs
/cloud/project/
var_short = col_character(),
percent = col_double(),
var_long = col_character(),
residence = col_character(),
service = col_character(),
indicator_type = col_character(),
indicator = col_character()
)
>
>
> washdata_uga <- washdata %>%
+ filter(iso3 == "UGA") %>%
+ filter(year == 2017) %>%
+ filter(residence == "national")
>
Files Plots Packages Help Viewer
New Folder Upload Delete Rename More
Cloud project
Name Size Modified
.. .gitignore 40 B Mar 10, 2021, 10:15 AM
.Rhistory 0 B Mar 10, 2021, 10:15 AM
data
exercise-01.Rmd 2 KB Mar 10, 2021, 11:18 AM
LICENSE 1 KB Mar 10, 2021, 10:15 AM
project.Rproj 205 B Mar 10, 2021, 10:57 AM
README.md 0 B Mar 10, 2021, 10:16 AM
setup
exercise-02.Rmd 2 KB Mar 10, 2021, 11:24 AM
exercise-01.html 724.8 KB Mar 10, 2021, 11:34 AM


RStudio Cloud
https://rstudio.cloud/project/2291449
133%
Your Workspace / data-science-for-wash-workshop
RAM
Lars Schöbitz
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
R 4.0.3
exercise-01.Rmd
Knit
Run
Code Editor
1 ---
2 title: "My first R Markdown report"
3 author: "Add your name here"
4 output: html_document
5 edit
6 ch
7 ---
8
9 # R markdown file
10
11 This an R Markdown file. It combines text with code. This is text
written in plain markdown and you can use markdown syntax to highlight
text in **bold**, *italic* or _underlined_.
40:7
Chunk 2
R Markdown
Environment History Connections Git Tutorial
Import Dataset
List
R
Global Environment
Data
washdata 27171 obs. of 11 variables
washdata_uga 22 obs. of 11 variables
Console Terminal R Markdown Jobs
/cloud/project/
var_short = col_character(),
percent = col_double(),
var_long = col_character(),
residence = col_character(),
service = col_character(),
indicator_type = col_character(),
indicator = col_character()
)
>
>
> washdata_uga <- washdata %>%
+ filter(iso3 == "UGA") %>%
+ filter(year == 2017) %>%
+ filter(residence == "national")
>
Files Plots Packages Help Viewer
New Folder Upload Delete Rename More
Cloud project
Name Size Modified
.. .gitignore 40 B Mar 10, 2021, 10:15 AM
.Rhistory 0 B Mar 10, 2021, 10:15 AM
data
exercise-01.Rmd 2 KB Mar 10, 2021, 11:18 AM
LICENSE 1 KB Mar 10, 2021, 10:15 AM
project.Rproj 205 B Mar 10, 2021, 10:57 AM
README.md 0 B Mar 10, 2021, 10:16 AM
setup
exercise-02.Rmd 2 KB Mar 10, 2021, 11:24 AM
724.8 KB Mar 10, 2021, 11:34 AM

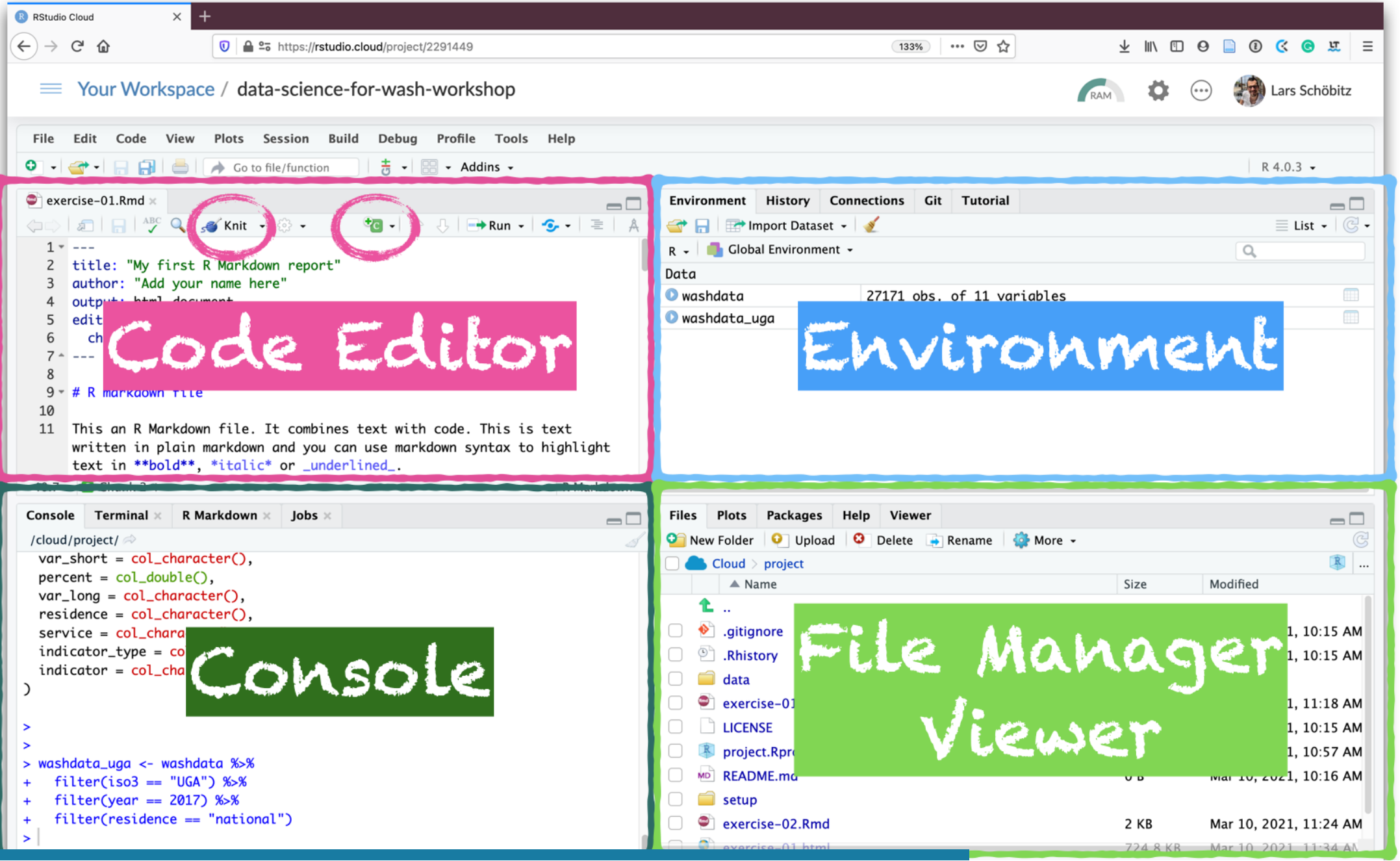

RStudio Cloud
https://rstudio.cloud/project/2291449
Your Workspace / data-science-for-wash-workshop
Lars Schöbitz
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
R 4.0.3
exercise-01.Rmd
Knit
Run
1 ---
2 title: "My first R Markdown report"
3 author: "Add your name here"
4 output: html_document
5 edit
6 ch
7 ---
8
9 # R markdown file
10
11 This an R Markdown file. It combines text with code. This is text
written in plain markdown and you can use markdown syntax to highlight
text in **bold**, *italic* or _underlined_.
Code Editor
40:7
Chunk 2
R Markdown
Environment
History
Connections
Git
Tutorial
Import Dataset
List
R
Global Environment
Data
washdata
27171 obs. of 11 variables
washdata_uga
Environment
Console
Terminal
R Markdown
Jobs
/cloud/project/
var_short = col_character(),
percent = col_double(),
var_long = col_character(),
residence = col_character(),
service = col_character(),
indicator_type = col_character(),
indicator = col_character()
)
>
>
> washdata_uga <- washdata %>%
+ filter(iso3 == "UGA") %>%
+ filter(year == 2017) %>%
+ filter(residence == "national")
>
Files
Plots
Packages
Help
Viewer
New Folder
Upload
Delete
Rename
More
Cloud
project
Name
Size
Modified
..
.gitignore 40 B Mar 10, 2021, 10:15 AM
.Rhistory 0 B Mar 10, 2021, 10:15 AM
data
exercise-01.Rmd 2 KB Mar 10, 2021, 11:18 AM
LICENSE 1 KB Mar 10, 2021, 10:15 AM
project.Rproj 205 B Mar 10, 2021, 10:57 AM
README.md 0 B Mar 10, 2021, 10:16 AM
setup
exercise-02.Rmd 2 KB Mar 10, 2021, 11:24 AM
724.8 KB Mar 10, 2021, 11:34 AM


RStudio Cloud
https://rstudio.cloud/project/2291449
Your Workspace / data-science-for-wash-workshop
Lars Schöbitz
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
R 4.0.3
exercise-01.Rmd
Knit
Run
---
title: "My first R Markdown report"
author: "Add your name here"
output: html_document
edit
ch

R markdown file
This an R Markdown file. It combines text with code. This is text written in plain markdown and you can use markdown syntax to highlight text in **bold**, *italic* or _underlined_.
Code Editor
40:7
Chunk 2
R Markdown
Environment
History
Connections
Git
Tutorial
Import Dataset
List
R
Global Environment
Data
washdata
27171 obs. of 11 variables
washdata_uga
Environment
Console
Terminal
R Markdown
Jobs
/cloud/project/
var_short = col_character(),
percent = col_double(),
var_long = col_character(),
residence = col_character(),
service = col_character(),
indicator_type = col_character(),
indicator = col_character()
)
> washdata_uga <- washdata %>%
+ filter(iso3 == "UGA") %>%
+ filter(year == 2017) %>%
+ filter(residence == "national")
Files
Plots
Packages
Help
Viewer
New Folder
Upload
Delete
Rename
More
Cloud
project
Name
Size
Modified
..
.gitignore
.Rhistory
data
exercise-01
LICENSE
project.Rpr
README.md
setup
exercise-02.Rmd
2 KB
Mar 10, 2021, 11:24 AM
File Manager
Viewer


RStudio Cloud
https://rstudio.cloud/project/2291449
Your Workspace / data-science-for-wash-workshop
Lars Schöbitz
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
R 4.0.3
exercise-01.Rmd
Knit
Run
---
title: "My first R Markdown report"
author: "Add your name here"
output: html_document
edit
ch

R markdown file
This an R Markdown file. It combines text with code. This is text written in plain markdown and you can use markdown syntax to highlight text in **bold**, *italic* or _underlined_.
Code Editor
40:7
Chunk 2
R Markdown
Environment
History
Connections
Git
Tutorial
Import Dataset
List
R
Global Environment
Data
washdata
27171 obs. of 11 variables
washdata_uga
Environment
Console
Terminal
R Markdown
Jobs
/cloud/project/
var_short = col_character(),
percent = col_double(),
var_long = col_character(),
residence = col_character(),
service = col_character(),
indicator_type = col_character(),
indicator = col_character()
)
> washdata_uga <- washdata %>%
+ filter(iso3 == "UGA") %>%
+ filter(year == 2017) %>%
+ filter(residence == "national")
Files
Plots
Packages
Help
Viewer
New Folder
Upload
Delete
Rename
More
Cloud
project
Name
Size
Modified
..
.gitignore
.Rhistory
data
exercise-01
LICENSE
project.Rpr
README.md
setup
exercise-02.Rmd
2 KB
Mar 10, 2021, 11:24 AM
File Manager
Viewer


RStudio Cloud
https://rstudio.cloud/project/2291449
Your Workspace / data-science-for-wash-workshop
Lars Schöbitz
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
R 4.0.3
exercise-01.Rmd
Knit
Run
1 ---
2 title: "My first R Markdown report"
3 author: "Add your name here"
4 output: html_document
5 edit
6 ch
7 ---
8
9 # R markdown file
10
11 This an R Markdown file. It combines text with code. This is text
written in plain markdown and you can use markdown syntax to highlight
text in **bold**, *italic* or _underlined_.
Code Editor
Environment History Connections Git Tutorial
Import Dataset
List
R
Global Environment
Data
washdata 27171 obs. of 11 variables
washdata_uga
Environment
Console Terminal R Markdown Jobs
/cloud/project/
var_short = col_character(),
percent = col_double(),
var_long = col_character(),
residence = col_character(),
service = col_chara
indicator_type = co
indicator = col_cha
)
Console
> washdata_uga <- washdata %>%
+ filter(iso3 == "UGA") %>%
+ filter(year == 2017) %>%
+ filter(residence == "national")
Files Plots Packages Help Viewer
New Folder Upload Delete Rename More
Cloud project
Name Size Modified
..
.gitignore
.Rhistory
data
exercise-01
LICENSE
project.Rpr
README.md
setup
exercise-02.Rmd 2 KB Mar 10, 2021, 11:24 AM
File Manager
Viewer

Let's put work into this workshop

Requirements

1. ~~Registration for Colorado WASH Symposium 2021~~
2. A free account on RStudio Cloud
    - https://rstudio.cloud/plans/free
3. One of Mozilla Firefox, Google Chrome, Microsoft Edge, Safari
    - just **not** the Internet Explorer
4. A laptop or desktop computer
    - it will be hard to do on a phone or tablet

Screen setup

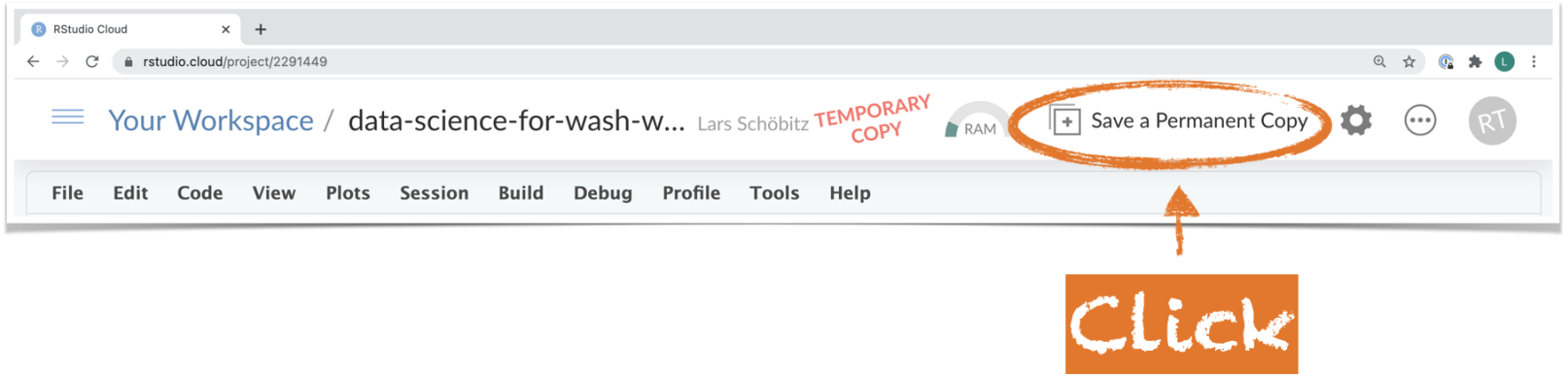
Add screenshot that shows setup on one small screen

Your Turn

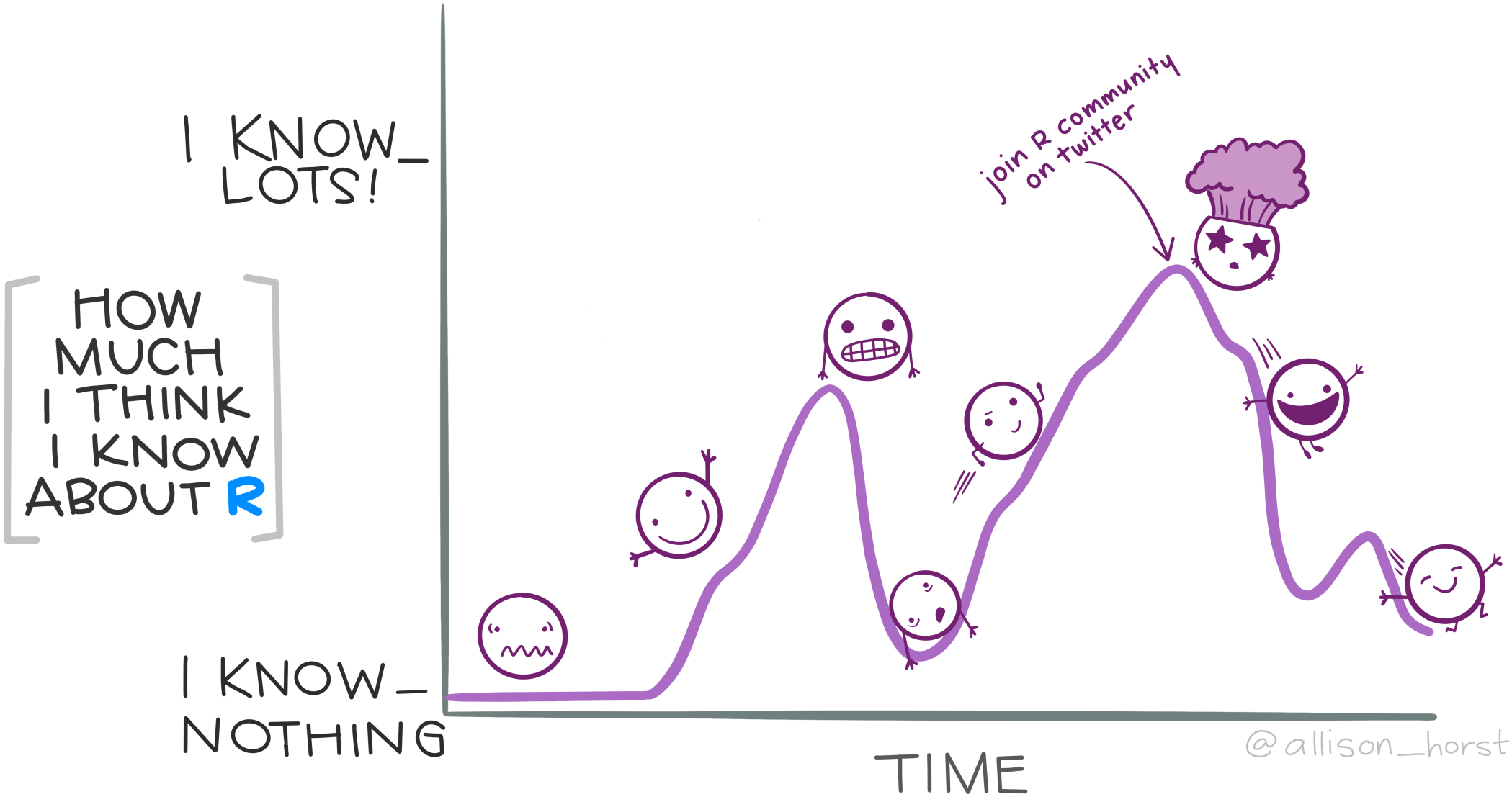
Step 1: Open this link in your browser

kutt.it/cowash-ws

Step 2: Create your own permanent project copy



What's next?



Artwork by @allison_horst

If you could not follow through the exercises for any reason

- Sign up for another round of this workshop

TODO: Schedule a Zoom Meeting that people can sign up for

If you want to continue learning online

- Twitter: #rstats
- Online learning community @R4DScommunity
- Minority R Users @miR_community
- We are @R-Ladies
- RStudio Community

Artwork by @allison_horst

If you want a complete introduction to R Markdown

Tutorial by Danielle Navarro

- Slides: https://slides.djnavarro.net/starting-rmarkdown
- YouTube videos: https://youtube.djnavarro.net/starting-rmarkdown

If you are interested in following along the development of Data Science for WASH

- Slack: kutt.it/washdata-slack
- Twitter: @washdata
- E-Mail: Fill out this form **Add link to form**

If you have unanswered questions or want to leave a comment (anonymously)

kutt.it/cowash-gd

Survey form

We want to learn a bit more about you

- Add Google Form Link

🌻 Thank you!

For joining!

For R packages {xaringan} and {xaringanthemer}, which where used to create these slides.