

# Joining data & communicating results with Quarto

CVEN 5837 - Summer 2022

Lars Schöbitz

<https://cven5837-ss22.github.io/website/>

<https://cven5837-ss22.github.io/website/>

# Learning Objectives (for this week)

1. Learners can use Quarto and Netlify to publish an HTML file
2. Learners can add literature references to Quarto files using the navigation menu of RStudio visual editor
3. Learners can cross-reference figures and tables within an Quarto file
4. Learners can apply functions from the dplyr R Package to join multiple data sets

# Part 1: Joining data

# We...

...have multiple data frames

...want to bring them together

```
1 professions <- read_csv(here::here("data/scientists/professions.csv"))
2 dates <- read_csv(here::here("data/scientists/dates.csv"))
3 works <- read_csv(here::here("scientists/works.csv"))
```

# Data: Women in science

Information on 10 women in science who changed the world

name
Ada Lovelace
Marie Curie
Janaki Ammal
Chien-Shiung Wu
Katherine Johnson
Rosalind Franklin
Vera Rubin
Gladys West
Flossie Wong-Staal
Jennifer Doudna

# Inputs

professions

dates

works

# A tibble: 10 × 2

	name <chr>	profession <chr>
1	Ada Lovelace	Mathematician
2	Marie Curie	Physicist and Chemist
3	Janaki Ammal	Botanist
4	Chien-Shiung Wu	Physicist
5	Katherine Johnson	Mathematician
6	Rosalind Franklin	Chemist
7	Vera Rubin	Astronomer
8	Gladys West	Mathematician
9	Flossie Wong-Staal	Virologist and Molecular Biologist
10	Jennifer Doudna	Biochemist

# Desired output

```
# A tibble: 10 × 5
```

	name	profession	birth_year	death_year	
	<chr>	<chr>	<dbl>	<dbl>	<chr>
1	Ada Lovelace	Mathematician	NA	NA	first
2	Marie Curie	Physicist and Chemist	NA	NA	
3	Janaki Ammal	Botanist	1897	1984	
4	Chien-Shiung Wu	Physicist	1912	1997	
5	Katherine Johnson	Mathematician	1918	2020	
6	Rosalind Franklin	Chemist	1920	1958	<NA>



# Inputs, reminder

```
1 names(professions)
```

```
[1] "name"      "profession"
```

```
1 names(dates)
```

```
[1] "name"      "birth_year"  
"death_year"
```

```
1 names(works)
```

```
[1] "name"      "known_for"
```

```
1 nrow(professions)
```

```
[1] 10
```

```
1 nrow(dates)
```

```
[1] 8
```

```
1 nrow(works)
```

```
[1] 9
```

# Joining data frames

# Joining data frames

```
1 something_join(x, y)
```

- `left_join()`: all rows from x
- `right_join()`: all rows from y
- `full_join()`: all rows from both x and y
- ...

# Setup

For the next few slides...

```
1 x <- tibble(
2   id = c(1, 2, 3),
3   value_x = c("x1", "x2", "x3")
4 )
```

```
1 x
```

```
# A tibble: 3 × 2
```

```
      id value_x
<dbl> <chr>
1     1 x1
2     2 x2
3     3 x3
```

```
1 y <- tibble(
2   id = c(1, 2, 4),
3   value_y = c("y1", "y2", "y4")
4 )
```

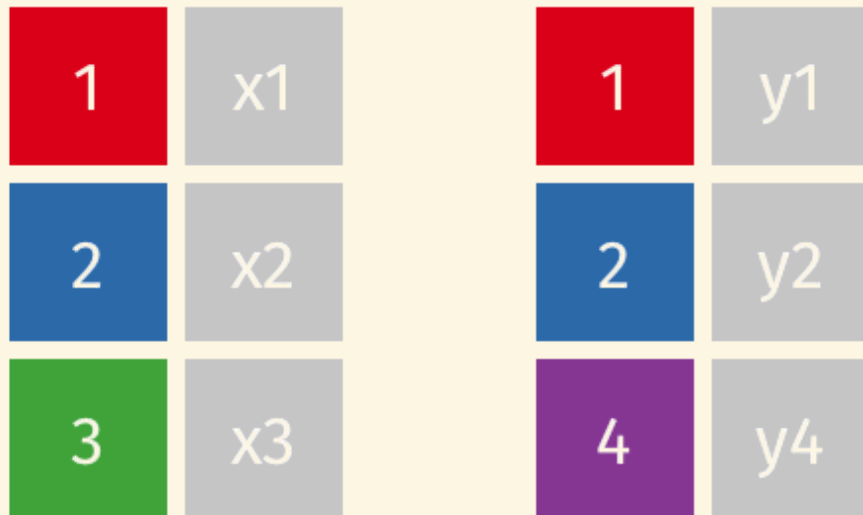
```
1 y
```

```
# A tibble: 3 × 2
```

```
      id value_y
<dbl> <chr>
1     1 y1
2     2 y2
3     4 y4
```

# left\_join()

left\_join(x, y)



```
1 left_join(x, y)
```

```
# A tibble: 3 × 3
```

```
      id value_x value_y
  <dbl> <chr>   <chr>
1     1 x1      y1
2     2 x2      y2
3     3 x3      <NA>
```

# left\_join()

```
1 professions %>%
2   left_join(dates)
```

# A tibble: 10 × 4

	name	profession	birth_year
	death_year		
	<chr>	<chr>	<dbl>
	<dbl>		
1	Ada Lovelace	Mathematician	NA
	NA		
2	Marie Curie	Physicist and Chemist	NA
	NA		
3	Janaki Ammal	Botanist	1897
	1984		
4	Chien-Shiung Wu	Physicist	1912
	1997		
5	Katherine Johnson	Mathematician	1918
	2020		

# right\_join()

right\_join(x, y)

1	x1	1	y1
2	x2	2	y2
3	x3	4	y4

```
1 right_join(x, y)
```

```
# A tibble: 3 × 3
```

```
      id value_x value_y
  <dbl> <chr>    <chr>
1     1 x1      y1
2     2 x2      y2
3     4 <NA>    y4
```

# right\_join()

```
1 professions %>%
2   right_join(dates)
```

```
# A tibble: 8 × 4
```

	name	profession	birth_year
	death_year		
	<chr>	<chr>	<dbl>
	<dbl>		
1	Janaki Ammal	Botanist	1897
	1984		
2	Chien-Shiung Wu	Physicist	1912
	1997		
3	Katherine Johnson	Mathematician	1918
	2020		
4	Rosalind Franklin	Chemist	1920
	1958		
5	Vera Rubin	Astronomer	1928
	2016		



# full\_join()

full\_join(x, y)

1	x1	1	y1
2	x2	2	y2
3	x3		
		4	y4

```
1 full_join(x, y)
```

# A tibble: 4 × 3

	id	value_x	value_y
	<dbl>	<chr>	<chr>
1	1	x1	y1
2	2	x2	y2
3	3	x3	<NA>
4	4	<NA>	y4

# full\_join()

```
1 dates %>%
2   full_join(works)
```

# A tibble: 10 × 4

	name <chr>	birth_year <dbl>	death_year <dbl>	known_for <chr>
1	Janaki Ammal	1897	1984	hybrid species, biodiversity protec...
2	Chien-Shiung Wu	1912	1997	confim and refine theory of radioac...
3	Katherine Johnson	1918	2020	calculations of orbital mechanics c...
4	Rosalind Franklin	1920	1958	<NA>
5	Vera Rubin	1928	2016	existence of dark matter
6	Gladys West	1930	NA	mathematical modeling of the shape ...
7	Flossie Wong-Staal	1947	NA	first scientist to clone HIV and cr...

# Putting it altogether

```
1 professions %>%
2   left_join(dates) %>%
3   left_join(works)
```

# A tibble: 10 × 5

name	profession	birth_year	death_year	
known_for				
<chr>	<chr>	<dbl>	<dbl>	<chr>
1 Ada Lovelace	Mathematician	NA	NA	first
co...				
2 Marie Curie	Physicist and Chemist	NA	NA	
theory o...				
3 Janaki Ammal	Botanist	1897	1984	
hybrid s...				
4 Chien-Shiung Wu	Physicist	1912	1997	
confim a...				
5 Katherine Johnson	Mathematician	1918	2020	
calculat...				
6 Rosalind Franklin	Chemist	1920	1958	<NA>

# Part 2: Communicate results with Quarto

# What is Quarto?

- Next generation version of R Markdown from RStudio
- Multi-language (Python, R, Julia, Observable)
- Authoring in plain text markdown or Jupyter notebooks

# Editing documents

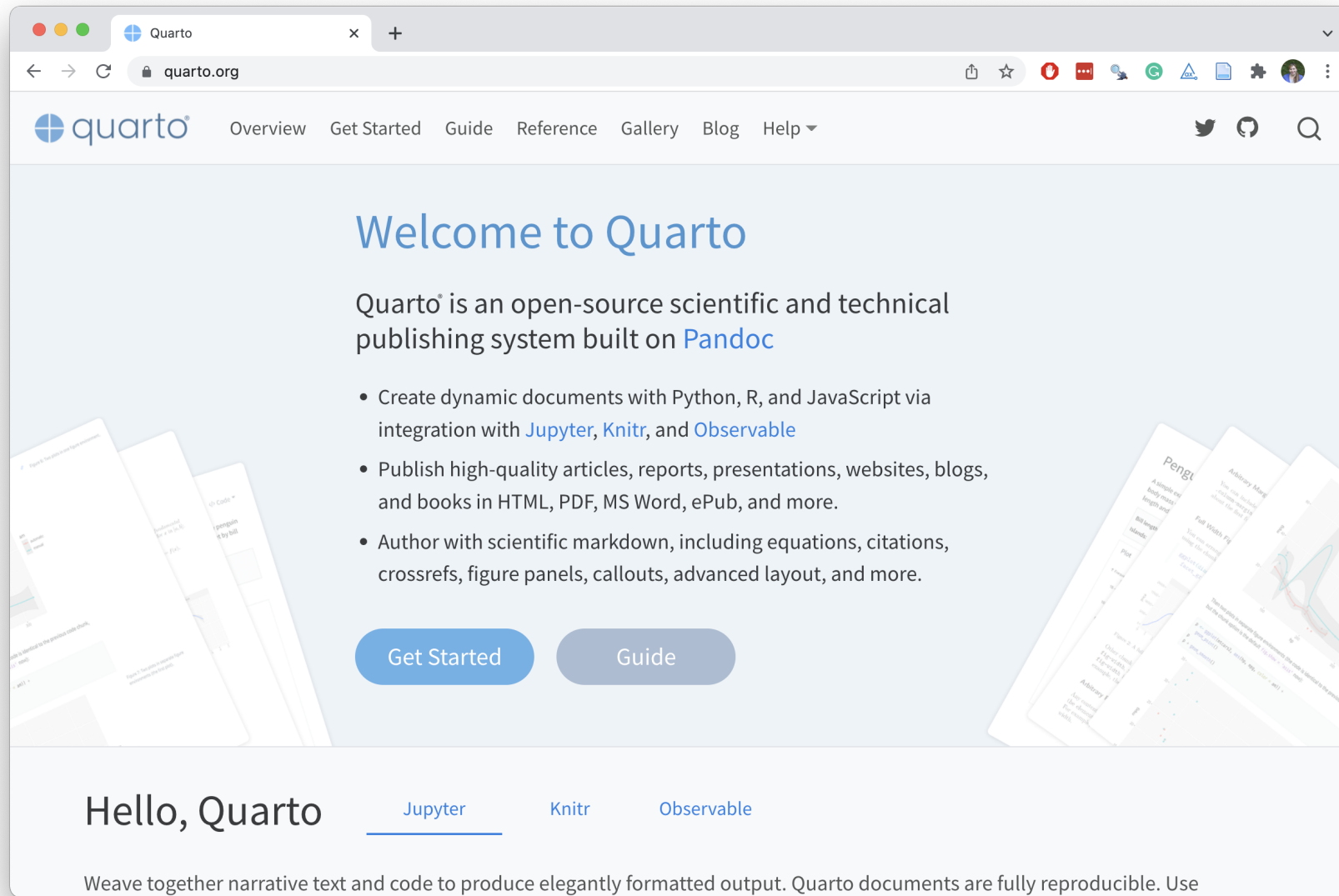
You're not limited to RStudio for editing Quarto documents...

- JupyterLab
- VS Code
- Text Editors

# Authoring

- Technical articles, reports, presentations, websites, blogs, and books in HTML, PDF, MS Word, ePub, and more.
- Author with scientific markdown, including equations, citations, crossrefs, figure panels, callouts, advanced layout, and more.

# Documentation



<https://cven5837-ss22.github.io/website/>



# Live Coding Exercise: Build a website

# live-05a-communicate

1. Head over to [rstudio.cloud](https://rstudio.cloud)
2. Open the workspace for the course (cven5837-ss22)
3. Open “Projects”
4. Open the “course-materials” project
5. Follow along with me

# Break



10:00

Photo by [Blake Wisz](#)

<https://cven5837-ss22.github.io/website/>

# Cross-references

- no space between `{r}` and `#|` `tbl-cap: "A table"`
- spelling `tbl` not `tab`
- no spaces (use dashes in `label`)

See [Table 1...](#)

```

1  ```{r}
2  #| tbl-cap: "A table"
3  #| label: tbl-simple-table
4
5  tibble(
6    id = c(1, 2, 3),
7    name = c("X", "Y", "Z")
8  ) %>%
9    knitr::kable()
10  ```

```

Table 1: A  
table

**id name**

<https://cven5837-ss22.github.io/website/>

id	name
1	X
2	Y
3	Z

# Homework week 5

# Homework due dates

- All material on [course website](#)
- Homework assignment due: Friday, 5th August
- Learning reflection due: Monday, 8th August

# What comes next?



# Get out the cloud

Install software on your computer:

- R: <https://cran.r-project.org/>
- RStudio IDE:  
<https://www.rstudio.com/products/rstudio/download/>
- Quarto CLI: <https://quarto.org/docs/get-started/>

# Start a project

- Guidance: <https://r4ds.had.co.nz/workflow-projects.html>
- Learn about project oriented workflows:  
<https://www.tidyverse.org/blog/2017/12/workflow-vs-script/>

# Stay in touch

- Email: [lars@lse.de](mailto:lars@lse.de)
- Data Science for WASH Slack Community: [Slack invite link](#)

Thanks! 

Slides created via revealjs and Quarto:

<https://quarto.org/docs/presentations/revealjs/> Access slides as [PDF on GitHub](#)

All material is licensed under [Creative Commons Attribution Share Alike 4.0 International](#).