

Creating and publishing scholarly articles with Quarto and GitHub pages

ds4owd - data science for openwashdata

Lars Schöbitz

2024-01-16



Q: How do I successfully complete the course?

You successfully complete the course and you will receive a certificate of completion if you:

hand in a complete capstone project report that uses a dataset of
your choice by 13th February 2024

Required items: <https://ds4owd-001.github.io/website/project/>

This is the only requirement to successfully complete the course,
independent of how many classes you attended or how many
homework assignments you completed.

Learning Objectives (for this week)

1. Learners can add literature references to Quarto files using the navigation menu of RStudio visual editor.
2. Learners can cross-reference figures and tables within an Quarto file.
3. Learners can use the GitHub pages service to publish a repository as a standalone website.

Anatomy of a Quarto document

Components

1. Metadata: YAML
2. Text: Markdown
3. Code: Executed via knitr or jupyter

Weave it all together, and you have beautiful, powerful, and useful outputs!

Literate programming

Literate programming is writing out the program logic in a human language with included (separated by a primitive markup) code snippets and macros.

```
1 ---  
2 title: "ggplot2 demo"  
3 date: "5/23/2023"  
4 format: html  
5 ---  
6  
7 ## MPG  
8  
9 There is a relationship between city and highway mileage.  
10  
11 ````{r}  
12 #| label: fig-mpg  
13  
14 library(ggplot2)  
15  
16 ggplot(mpg, aes(x = cty, y = hwy)) +  
17   geom_point() +  
18   geom_smooth(method = "loess")  
19 ````
```

Metadata

YAML

“Yet Another Markup Language” or “YAML Ain’t Markup Language” is used to provide document level metadata.

```
1  ---
2  key: value
3  ---
```

Output options

```
1  ---
2  format: something
3  ---
```

```
1  ---
2  format: html
3  ---
```

```
1  ---
2  format: pdf
3  ---
```

```
1  ---
2  format: revealjs
3  ---
```

Output option arguments

Indentation matters!

```
1  ---
2  format:
3    html:
4      toc: true
5      code-fold: true
6  ---
```

YAML validation

- Invalid: No space after :

```
1 ---  
2 format:html  
3 ---
```

- Invalid: Read as missing

```
1 ---  
2 format:  
3 html  
4 ---
```

YAML validation

There are multiple ways of formatting valid YAML:

- Valid: There's a space after :

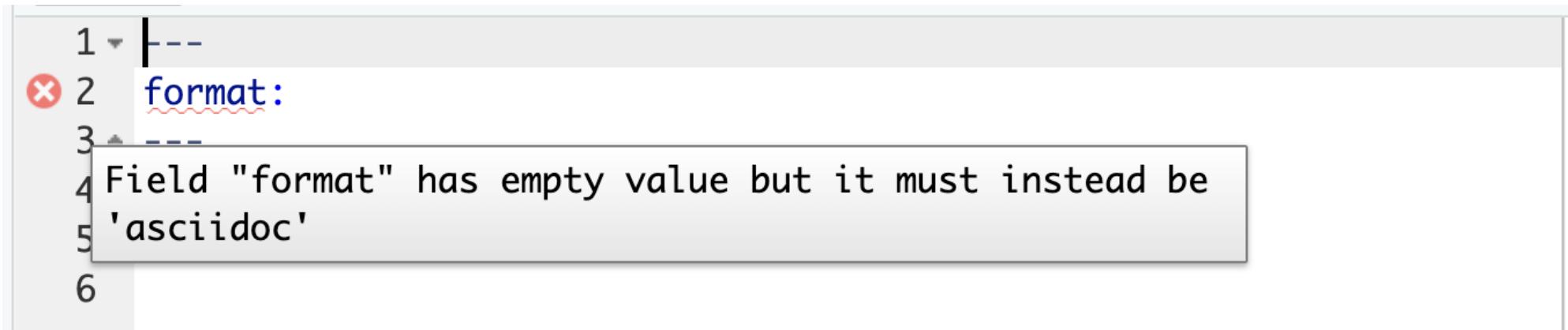
```
1 format: html
```

- Valid: `format: html` with selections made with proper indentation

```
1 format:  
2   html:  
3     toc: true
```

Quarto linting

Lint, or a linter, is a static code analysis tool used to flag programming errors, bugs, stylistic errors and suspicious constructs.



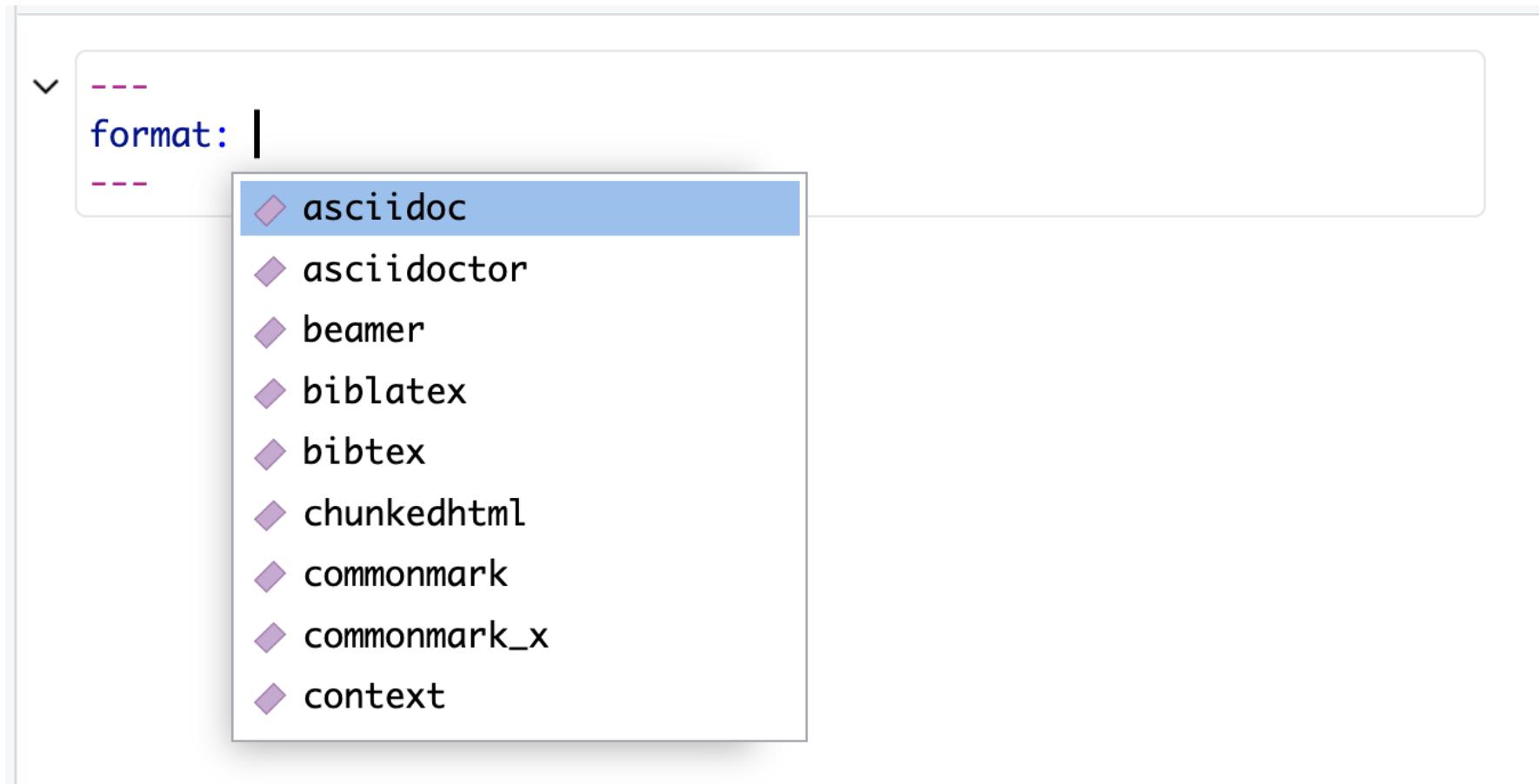
The screenshot shows a code editor with the following content:

```
1 ---  
2 format:  
3 ---  
4 Field "format" has empty value but it must instead be  
5 'asciidoc'  
6
```

A red circle with a white 'X' is positioned next to the line number 2, indicating an error. The word 'format:' is underlined in red, signifying it is the source of the error. A tooltip or callout box is displayed over the error line, containing the message: "Field 'format' has empty value but it must instead be 'asciidoc'".

Quarto YAML Intelligence

RStudio + VSCode provide rich tab-completion - start a word and tab to complete, or **Ctrl + space** to see all available options.



My turn: A tour of Quarto (once again)

Sit back and enjoy!

List of valid YAML fields

- Many YAML fields are common across various outputs
- But also each output type has its own set of valid YAML fields and options
- Definitive list: quarto.org/docs/reference/formats/html

Text (in Markdown)

Text Formatting

Markdown Syntax

Output

italics and **bold**

italics and **bold**

superscript^2^ / subscript~2~

superscript^2 / subscript_2

~~strikethrough~~

~~strikethrough~~

`verbatim code`

verbatim code

Headings

Markdown Syntax	Output
# Header 1	<h1>Header 1</h1>
## Header 2	<h2>Header 2</h2>
### Header 3	<h3>Header 3</h3>
#### Header 4	<h4>Header 4</h4>
##### Header 5	<h5>Header 5</h5>
###### Header 6	<h6>Header 6</h6>

Links

There are several types of “links” or hyperlinks.

Markdown

- 1 You can embed [named hyperlinks](<https://quarto.org/>),
- 2 direct urls like <<https://quarto.org/>>. The syntax is similar for
- 3 inline image:
- 4 ! [Penguins playing with ball](img/lec-08/penguins-quarto-ball.png)

Output

You can embed
named
hyperlinks, direct
urls like
<https://quarto.org/>.
The syntax is
similar for
embedding an
inline image:



Lists

Unordered list:

Markdown:

```

1 - unordered list
2   - sub-item 1
3   - sub-item 1
4     - sub-sub-item 1

```

Output

- unordered list
 - sub-item 1
 - sub-item 1
 - sub-sub-item 1

Ordered list:

Markdown:

```

1 1. ordered list
2 2. item 2
3   i. sub-item 1
4     A. sub-sub-item 1

```

Output

1. ordered list
2. item 2
 - i. sub-item 1
 - a. sub-sub-item 1

Quotes

Markdown:

```
1 > Let us change our traditional attitude to the construct  
2 > - Donald Knuth, Literate Programming
```

Output:

Let us change our traditional attitude to the construction of programs: Instead of imagining that our main task is to instruct a computer what to do, let us concentrate rather on explaining to human beings what we want a computer to do. - Donald Knuth, Literate Programming

Your turn: md-08-exercises

1. Open [posit.cloud](#) in your browser (use your bookmark).
2. Open the [ds4owd workspace](#) for the course.
3. Click [Start](#) next to [md-08-exercises](#).
4. In the File Manager in the bottom right window, locate the [md-08-markdown-syntax.qmd](#) file and click on it to open it in the top left window.
5. Use the [source editor](#) mode
6. Follow the instructions in the document, then exchange one new thing you've learned with your neighbor.

Take a break

Please get up and move! Let your emails rest in peace.



Anatomy of a Quarto scholarly article

Components

1. Metadata: YAML
2. Text: Markdown
3. Code: Executed via knitr or jupyter

Weave it all together, and you have a beautiful, reproducible journal article!

Scholarly writing - four terms

- Citation
- Reference
- Bibliography
- Citation Style Language (CSL)

What's a Citation?

- Inequality underpins waste management systems, structuring who can or cannot access services ([Kalina et al., 2023](#)).
- Many visitors still expect a personal pick-up, despite the availability of taxi services ([Tilley & Kalina, 2021](#)).
- In Tilley & Kalina ([2021](#)), the authors describe how visitors still expect a personal pick-up, despite the availability of taxi services.

What's a Citation?

- Inequality underpins waste management systems, structuring who can or cannot access services ([Kalina et al., 2023](#)).
- Many visitors still expect a personal pick-up, despite the availability of taxi services ([Tilley & Kalina, 2021](#)).
- In Tilley & Kalina ([2021](#)), the authors describe how visitors still expect a personal pick-up, despite the availability of taxi services.



Important: The period is after the citation.

What's a Reference?

- detailed description of the source of information
- author's name, title, year of publication, publisher, DOI, etc.

Tilley, E., & Kalina, M. (2021). “My flight arrives at 5 am, can you pick me up?”: The gatekeeping burden of the african academic. *Journal of African Cultural Studies*, 33(4), 538–548.

<https://doi.org/10.3929/ethz-b-000493677>

What's a Bibliography?

- list of references in a research paper or project
- includes all sources used, whether they were directly quoted or not
- listed alphabetically by the author's last name in the reference list

References

Kalina, M., Makwetu, N., & Tilley, E. (2023). "The rich will always be able to dispose of their waste": A view from the frontlines of municipal failure in Makhanda, South Africa.

Environment, Development and Sustainability. <https://doi.org/10.1007/s10668-023-03363-1>

Knuth, D. E. (1984). Literate Programming. *The Computer Journal*, 27(2), 97–111.
<https://doi.org/10.1093/comjnl/27.2.97>

Tilley, E., & Kalina, M. (2021). "My flight arrives at 5 am, can you pick me up?": The gatekeeping burden of the ~~african academic~~ <https://www.academia.edu/website>  *Journal of African Cultural Studies*, 33(4),

What's the Citation Style Language (CSL)?

- It's what your citation and generated bibliography look like
- APA (American Psychological Association) Style, Chicago Style, IEEE Style, Vancouver Style, etc. (over 10,000 styles in [Zotero Style Repository](#))

What's the Citation Style Language (CSL)?

author-date: Many visitors still expect a personal pick-up, despite the availability of taxi services ([Tilley & Kalina, 2021](#)).

numeric Many visitors still expect a personal pick-up, despite the availability of taxi services [1].

Why use a reference management tool?

Managing references
manually:

- is a lot of work
- is prone to mistakes
- makes you lose track



Why use Zotero?

- free
- open source: developed in public
- transparent about access to your own data
- cross-platform (Windows, Mac, Linux)
- collaboration in groups
- integration with word processors



Zotero setup guide

- Not teaching how to use Zotero, but encourage you to learn and use it.
- Prepared a setup guide: <https://ds4owd-001.github.io/website/guide/>

Scholarly Articles in Quarto

Quarto supports

- a standardized schema for authors and affiliations that can be expressed once in the source document,
- the use of Citation Style Language (CSL) to automate the formatting of citations and bibliographies, and
- outputting to `pdf`, `html`, and `docx` with custom formatting,

according to the styles required for various journals,
and creating the LaTeX required for submission to multiple
journals.

Front matter

Quarto provides a rich set of YAML metadata keys to describe the details required in the front matter of scholarly articles.

- title
- author
- affiliation
- abstract
- keywords
- citation
- licensing
- etc.

Give me a title

AUTHOR

First Last 

AFFILIATION

ETH Zurich

PUBLISHED

October 17, 2023

ABSTRACT

The abstract will be placed here. Breaks can be added by hitting return on the keyboard.

```
---
title: "Give me a title"
date: 2023-10-17
author:
  - name: First Last
    orcid: 0000-0003-2196-5015
    email: me@web.org
    affiliation:
      - name: ETH Zurich
        url: https://ethz.ch/de.html
abstract: >
  The abstract will be placed here. Breaks can be added by hitting return
  on the keyboard.
```

Our turn: md-08-exercises

1. Open [posit.cloud](#) in your browser (use your bookmark).
2. Open the [ds4owd workspace](#) for the course.
3. Open [md-08-exercises](#).
4. In the File Manager in the bottom right window, locate the [scholarly-writing.qmd](#) file and click on it to open it in the top left window.
5. Follow along on the screen using the instructions in the document.

Publishing

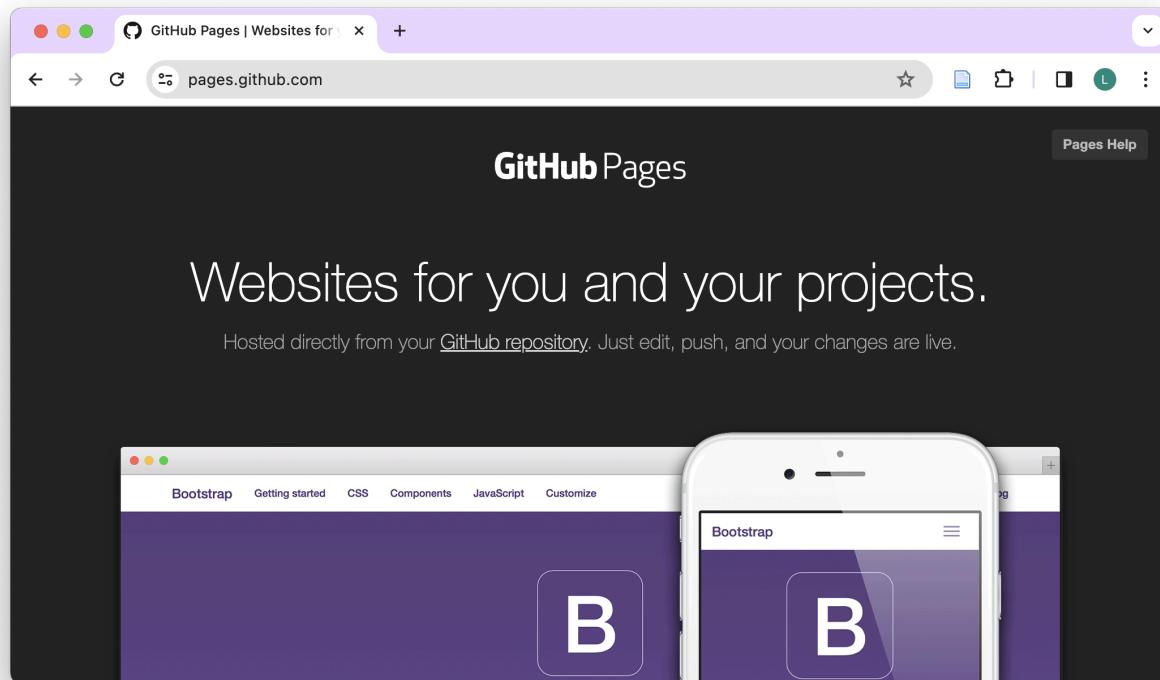
Our turn: md-08-publish-USERNAME

Clone the repository from GitHub

1. Open github.com in your browser and navigate to the [GitHub organisation for the course: https://github.com/ds4owd-001/](#).
2. Find the repository **md-08-publish-USERNAME** that ends with your GitHub username, and open it.
3. Click on the green “Code” button.
4. Copy the HTTPS URL to your clipboard.
5. Open the ds4owd workspace on [posit.cloud](#)
6. Click New Project > New Project from Git Repository
7. Paste the HTTPS URL from GitHub into the “URL of your Git Repository” field.
8. Wait until the project is deployed.
9. From the Files Manager in the bottom right window, open **docs** folder, then click on [index.qmd](#).
10. Indicate the open Poll with “ready to go” when you are ready.

GitHub Pages

- GitHub Pages is a free service for hosting static websites. It is ideal for blogs, course or project websites, books, presentations, and personal hobby sites.



✉️ ds4owd-001.github.io/website/

Minimal Example - Requirements

- Landing site needs to be stored as `index.qmd`
- The `index.qmd` needs to be stored in `docs` folder
- Example works well for a report/article as a stand-alone page
- Quarto provides a framework and examples for more complex websites: <https://quarto.org/docs/websites/>

Course Guide

- Steps for publishing the capstone project report are described in course guide
- <https://ds4owd-001.github.io/website/guide/>

Take a break

Please get up and move! Let your emails rest in peace.



Capstone project

Course certificate

- You will receive a course certificate if you complete the capstone project.
- The course certificate will be issued by the openwashdata academy.
- The certificate will highlight the time you have invested and the tools you learned to navigate.
- The certificate can include a link to your public capstone project report (voluntary).
- We would like to add a graduates section to <https://openwashdata.org/> and highlight course graduates (e.g. link to report, GitHub profile, LinkedIn profile, ORCID profile)

Submission

- The submission due date is: **Tuesday, 13th February.**
- You will need to work through Module 5 & Module 7 homework assignments to get started.
- We will use the GitHub issue tracker to communicate and ask questions about the capstone project.
- A list of required items for submission is covered on the course website: <https://ds4owd-001.github.io/website/project/>
- If you require an extension, please reach to us via email:
ghe@mavt.ethz.ch

Your turn: Capstone project - Read and take notes

1. Open: <https://ds4owd-001.github.io/website/project/>.
2. Read through the page.
3. For the list in “Required items” note down the numbers of those that are unclear to you and why.
4. After the time is up, you will join a break-out room and discuss the unclear items with your peers.

Your turn: Capstone project - Discuss unclear items

1. Join the break-out room.
2. Discuss with your peers the unclear items you noted down.

Your turn: Capstone project - Share remaining unclear items

1. Open this repository: <https://github.com/ds4owd-001/project>.
2. Add your questions for unclear items to the issue tracker as described.

Wrap-up

Homework assignment

- No more homework assignments
- Use the time to work on your capstone project

Student hours

- Every Thursday, 2:00 to 3:30 PM (CET)
- Final student hours: Thursday, February 08, 2:00 to 3:30 PM (CET)

Next week: openwashdata webinar

- Date: Tuesday, January 23, 2:00 - 3:00 PM (CET)
- Zoom Registration: https://ethz.zoom.us/meeting/register/u50Ice-vrD0iGN06ywPPr2gPU84Ljr_zmNAf
- More info: <https://openwashdata.org/pages/events/2023-01-24-webinar-data-publishing>

Time	Title	Speaker	Affiliation
02:00 - 02:05	Welcome	Elizabeth Tilley	Professor at ETH Zurich, Global Health Engineering
02:05 - 02:25	openwashdata community & workflow	Lars Schöbitz	Open Science Specialist at ETH Zurich, Global Health Engineering
02:25 - 02:45	Open data: A publishers perspective	Lauren Cadwallader	Open Research Manager at Public Library of Science (PLOS)
02:45 - 03:00	Discussion	Elizabeth Tilley	Professor at ETH Zurich, Global Health Engineering

Module 09: Using AI for software development in R

- Date: Tuesday, January 30, 2:00 - 4:30 PM (CET)

Module 10: Graduation party

- Date: Tuesday, February 20, 2:00 - 3:00 PM (CET)

Attribution

Content was re-used from a workshop hosted by [Mine Çetinkaya-Rundel](#) at the 2023 Symposium on Data Science and Statistics and stored at <https://github.com/mine-cetinkaya-rundel/quarto-sdss>. The original content is licensed under a [Creative Commons Attribution 4.0 International License](#).

Thanks! 🌻

Slides created via revealjs and Quarto:

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