# Welcome & get ready for the course

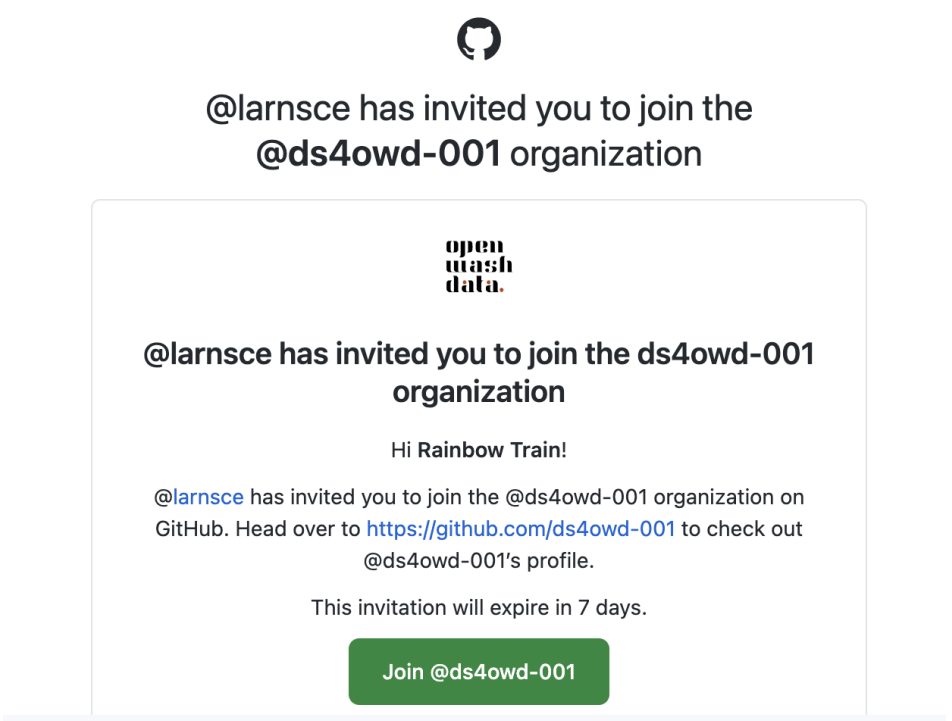
ds4owd - data science for the course

Lars Schöbitz

2023-10-31

# Email from GitHub?

While we are getting ready, please check for this email from GitHub and accept the invitation to join the GitHub organisation for the course. Used Gmail to sign up? Check the folders that aren't your primary inbox (e.g Updates).




ds4owd-001.github.io/website/

# Welcome! 👋

ds4owd-001.github.io/website/

# Meet the team

## Lars Schöbitz

- Env. Engineer
- retired WASH Researcher
- RStudio certified instructor

## Mian Zhong

- Data Scientist
- Advocate for data for social good

## Sophia Skorik

- Computer Scientist
- Technical support for the course

ds4owd-001.github.io/website/

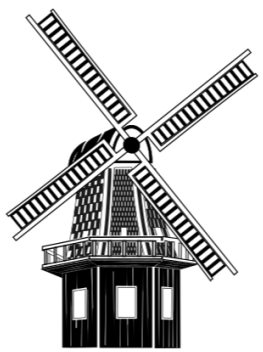
open wash data.

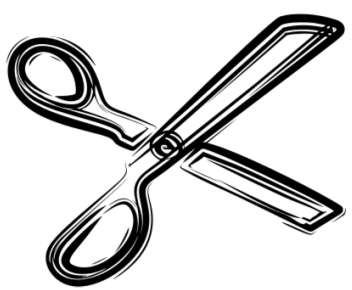
# Learning Goals (for the course)

1. Be able to use a common set of data science tools (R, RStudio IDE, Git, GitHub, tidyverse, Quarto) to illustrate and communicate the results of data analysis projects.
2. Learn to use the Quarto file format and the RStudio IDE visual editing mode to produce documents with citations, footnotes, cross-references, figures, and tables.

# Your turn: About you

> Pick an item and take notes for 1 minute:
>
> What does the item you have picked have to do with the reason for you being here?





Images from: https://openclipart.org/

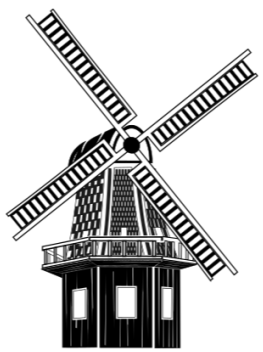
ds4owd-001.github.io/website/

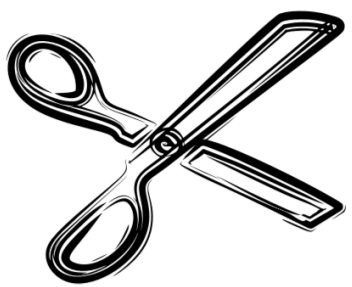
01:00

# In break-out rooms

**Take 2 minutes each to share with your room partner:**

What does the item you have picked have to do with the reason for you being here?





Images from: https://openclipart.org/



05:00

# Course Calendar

| date | week | topic | module |
|---|---|---|---|
| 31 October 2023 | 1 | Welcome & get ready for the course | module 1 |
| 07 November 2023 | 2 | Data science lifecycle & Quarto | module 2 |
| 14 November 2023 | 3 | Exploratory data analysis using visualization | module 3 |
| 21 November 2023 | 4 | Data organization in spreadsheets | module 4 |
| 28 November 2023 | 5 | Data transformation, descriptive statistics and gt | module 5 |
| 05 December 2023 | 6 | Concept of tidy data & Vectors in R | module 6 |
| 12 December 2023 | 7 | Joining data & wrting functions | module 7 |
| 19 December 2023 | 8 | Using AI for software development in R | module 8 |
| 26 December 2023 | 9 | Break | NA |
| 02 January 2024 | 10 | Break | NA |
| 09 January 2024 | 11 | Break | NA |
| 16 January 2024 | 12 | Personal website development with Quarto and publication of capstone project | module 9 |
| 23 January 2024 | 13 | Work on Capstone project | NA |
| 30 January 2024 | 14 | Final submission date of Capstone project | module 10 |
| 06 February 2024 | 15 | Graduation party of openwashdata academy | NA |

# Course structure

- My turn: Lecture segments + live coding
- Our turn: Live coding + follow along
- Your turn: Exercises in break-out rooms

# My turn: Lecture segments + live coding

- Instructor writes and narrates code out loud
- Instructor explains concepts and principles that are relevant

# Our turn: Live coding + follow along

- Instructor writes and narrates code out loud
- Instructor explains concepts and principles that are relevant
- Code is displayed on second screen / split screen
- Learners join by writing and executing the same code

# Your turn: Exercises in break-out rooms

- Two learners work together in a break out session
- One person (the driver) shares the screen and does the typing
- The other person (the navigator) offers comments and suggestions
- Roles get switched

# Getting help

- During my turn and our turn segments: Please keep your microphone on mute. Send message to the Zoom chat Mian and Sophia will support you.
- During your turn segments: Due to the large number of participants, it will not be feasible to join individual break-out rooms, but you will always be working in pairs.

# Platforms and Tools

- R
- tidyverse R Packages
- Posit Cloud
- RStudio IDE
- Quarto publishing system
- Element

# Bookmark

ds4owd-001.github.io/website/

# Learning Objectives (for this week)

1. Learners can access the Posit Cloud workspace for the course.
2. Learners can use the Element chat to introduce themselves.
3. Learners can open an issue on GitHub and tag the course instructor.
4. Learners can clone a repository from GitHub and use the GitHub PAT to push a commit from their local repository to GitHub.


ds4owd-001.github.io/website/

# Posit Cloud

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins | R 4.3.0

live-01a-setup.qmd | live-01b-data-science-lifecycle-so...

Render | Run

Source | Visual | Normal | Format | Insert | Table

3. calculating summary statistics (like counts or the mean)

## Gapminder data

**Goal:** Calculate the median life expectancy at birth by continent for 2007.

```{r}

# before loading library, write code

library(dplyr)
```

(Top Level) | Quarto

Data import
Gapminder data
Data tidying
Gapminder data
Data transformation
Gapminder data
Data visualisation...
Gapminder data
Data communicat...

Environment | History | Connections | Git | Tutorial

Import Dataset | 172 MiB | List

R | Global Environment

Environment is empty

Console | Terminal | Background Jobs

R 4.3.0 · /cloud/project/

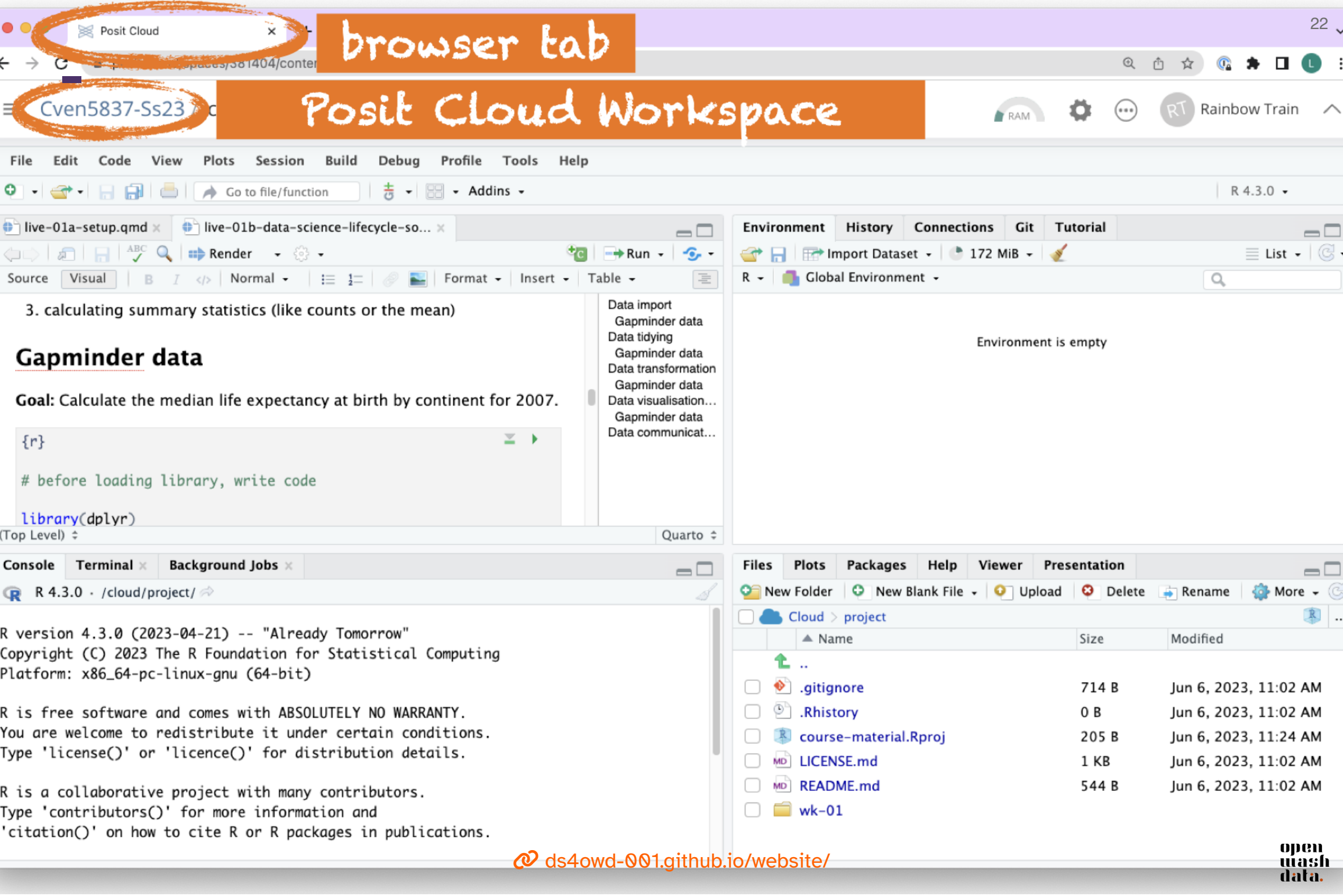
```
R version 4.3.0 (2023-04-21) -- "Already Tomorrow"
Copyright (C) 2023 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
```

Files | Plots | Packages | Help | Viewer | Presentation

New Folder | New Blank File | Upload | Delete | Rename | More

Cloud > project

| Name | Size | Modified |
|---|---|---|
| .. | | |
| .gitignore | 714 B | Jun 6, 2023, 11:02 AM |
| .Rhistory | 0 B | Jun 6, 2023, 11:02 AM |
| course-material.Rproj | 205 B | Jun 6, 2023, 11:24 AM |
| LICENSE.md | 1 KB | Jun 6, 2023, 11:02 AM |
| README.md | 544 B | Jun 6, 2023, 11:02 AM |
| wk-01 | | |


Posit Cloud
browser tab
Cven5837-Ss23
Posit Cloud Workspace
Rainbow Train
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
R 4.3.0
live-01a-setup.qmd
live-01b-data-science-lifecycle-so...
Render
Run
Source Visual
Normal
Format Insert Table
3. calculating summary statistics (like counts or the mean)
Gapminder data
Goal: Calculate the median life expectancy at birth by continent for 2007.
{r}
# before loading library, write code
library(dplyr)
Data import
Gapminder data
Data tidying
Gapminder data
Data transformation
Gapminder data
Data visualisation...
Gapminder data
Data communicat...
(Top Level)
Quarto
Environment History Connections Git Tutorial
Import Dataset
172 MiB
List
R
Global Environment
Environment is empty
Console Terminal Background Jobs
R 4.3.0 · /cloud/project/
R version 4.3.0 (2023-04-21) -- "Already Tomorrow"
Copyright (C) 2023 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Files Plots Packages Help Viewer Presentation
New Folder
New Blank File
Upload
Delete
Rename
More
Cloud > project
Name
Size
Modified
..
.gitignore 714 B Jun 6, 2023, 11:02 AM
.Rhistory 0 B Jun 6, 2023, 11:02 AM
course-material.Rproj 205 B Jun 6, 2023, 11:24 AM
LICENSE.md 1 KB Jun 6, 2023, 11:02 AM
README.md 544 B Jun 6, 2023, 11:02 AM
wk-01
ds4owd-001.github.io/website/
open wash data.

Posit Cloud

posit.cloud/spaces/381404/content/6066891

Cven5837-Ss23 / course-material-rainbow-train

Rainbow Train

RStudio IDE Menu

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins | R 4.3.0

live-01a-setup.qmd | live-01b-data-science-lifecycle-so...

Render | Run | Source | Visual | Normal | Format | Insert | Table

3. calculating summary statistics (like counts or the mean)

## Gapminder data

**Goal:** Calculate the median life expectancy at birth by continent for 2007.

```{r}

# before loading library, write code

library(dplyr)
```

(Top Level) | Quarto

Data import
Gapminder data
Data tidying
Gapminder data
Data transformation
Gapminder data
Data visualisation...
Gapminder data
Data communicat...

Environment | History | Connections | Git | Tutorial

Import Dataset | 172 MiB | List

R | Global Environment

Environment is empty

Console | Terminal | Background Jobs

R 4.3.0 · /cloud/project/

```
R version 4.3.0 (2023-04-21) -- "Already Tomorrow"
Copyright (C) 2023 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
```

Files | Plots | Packages | Help | Viewer | Presentation

New Folder | New Blank File | Upload | Delete | Rename | More

Cloud > project

| Name | Size | Modified |
|---|---|---|
| .. | | |
| .gitignore | 714 B | Jun 6, 2023, 11:02 AM |
| .Rhistory | 0 B | Jun 6, 2023, 11:02 AM |
| course-material.Rproj | 205 B | Jun 6, 2023, 11:24 AM |
| LICENSE.md | 1 KB | Jun 6, 2023, 11:02 AM |
| README.md | 544 B | Jun 6, 2023, 11:02 AM |
| wk-01 | | |

Posit Cloud

posit.cloud/spaces/381404/content/6066891

Cven5837-Ss23 / course-material-rainbow-train

RAM

Rainbow Train

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function

Addins

RStudio IDE Menu

R 4.3.0

live-01a-setup.qmd

live-01b-data-science-lifecycle-so...

Render

Run

Source Visual

Normal

Format Insert Table

3. calculating summary statistics (like counts or the mean)

## Gapminder dat

Code Editor

**Goal:** Calculate the me for 2007.

```{r}

# before loading library, write code

library(dplyr)
```

Data import
Gapminder data
Data tidying
Gapminder data
Data transformation
Gapminder data
Data visualisation...
Gapminder data
Data communicat...

Environment History Connections Git Tutorial

Import Dataset

172 MiB

List

R

Global Environment

Environment is empty

Console Terminal Background Jobs

R 4.3.0 · /cloud/project/

```
R version 4.3.0 (2023-04-21) -- "Already Tomorrow"
Copyright (C) 2023 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
```

Files Plots Packages Help Viewer Presentation

New Folder New Blank File Upload Delete Rename More

Cloud > project

| Name | Size | Modified |
|---|---|---|
| .. | | |
| .gitignore | 714 B | Jun 6, 2023, 11:02 AM |
| .Rhistory | 0 B | Jun 6, 2023, 11:02 AM |
| course-material.Rproj | 205 B | Jun 6, 2023, 11:24 AM |
| LICENSE.md | 1 KB | Jun 6, 2023, 11:02 AM |
| README.md | 544 B | Jun 6, 2023, 11:02 AM |
| wk-01 | | |

Posit Cloud

posit.cloud/spaces/381404/content/6066891

Cven5837-Ss23 / course-material-rainbow-train

RAM

Rainbow Train

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function

Addins

R 4.3.0

RStudio IDE Menu

live-01a-setup.qmd

live-01b-data-science-lifecycle-so...

Render

Run

Source Visual

Normal

Format Insert Table

3. calculating summary statistics (like counts or the mean)

## Gapminder dat

Code Editor

**Goal:** Calculate the me for 2007.

```{r}

# before loading library, write code

library(dplyr)
```

Data import
Gapminder data
Data tidying
Gapminder data
Data transformation
Gapminder data
Data visualisation...
Gapminder data
Data communicat...

Environment History Connections Git Tutorial

Import Dataset

172 MiB

List

R

Global Environment

Environment is empty

Environment

Git

Console Terminal Background Jobs

R 4.3.0 · /cloud/project/

```
R version 4.3.0 (2023-04-21) -- "Already Tomorrow"
Copyright (C) 2023 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
```

Files Plots Packages Help Viewer Presentation

New Folder New Blank File Upload Delete Rename More

Cloud > project

| Name | Size | Modified |
|---|---|---|
| .. | | |
| .gitignore | 714 B | Jun 6, 2023, 11:02 AM |
| .Rhistory | 0 B | Jun 6, 2023, 11:02 AM |
| course-material.Rproj | 205 B | Jun 6, 2023, 11:24 AM |
| LICENSE.md | 1 KB | Jun 6, 2023, 11:02 AM |
| README.md | 544 B | Jun 6, 2023, 11:02 AM |
| wk-01 | | |

Posit Cloud

posit.cloud/spaces/381404/content/6066891

Cven5837-Ss23 / course-material-rainbow-train

RAM

Rainbow Train

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function

Addins

R 4.3.0

# RStudio IDE Menu

live-01a-setup.qmd

live-01b-data-science-lifecycle-so...

Render

Run

Source Visual

Normal

Format Insert Table

3. calculating summary statistics (like counts or the mean)

## Gapminder dat

# Code Editor

**Goal:** Calculate the me for 2007.

```{r}

# before loading library, write code

library(dplyr)
```

Data import
Gapminder data
Data tidying
Gapminder data
Data transformation
Gapminder data
Data visualisation...
Gapminder data
Data communicat...

Environment History Connections Git Tutorial

Import Dataset

172 MiB

List

R

Global Environment

Environment is empty

# Environment

# Git

Console Terminal Background Jobs

R 4.3.0 · /cloud/project/

```
R version 4.3.0 (2023-04-21) -- "Already Tomorrow"
Copyright (C) 2023 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gn

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
```

# Console

Files Plots Packages Help Viewer Presentation

New Folder New Blank File Upload Delete Rename More

Cloud > project

| Name | Size | Modified |
|---|---|---|
| .. | | |
| .gitignore | 714 B | Jun 6, 2023, 11:02 AM |
| .Rhistory | 0 B | Jun 6, 2023, 11:02 AM |
| course-material.Rproj | 205 B | Jun 6, 2023, 11:24 AM |
| LICENSE.md | 1 KB | Jun 6, 2023, 11:02 AM |
| README.md | 544 B | Jun 6, 2023, 11:02 AM |
| wk-01 | | |

ds4owd-001.github.io/website/

Posit Cloud

posit.cloud/spaces/381404/content/6066891

Cven5837-Ss23 / course-material-rainbow-train

RAM

Rainbow Train

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function

Addins

R 4.3.0

RStudio IDE Menu

live-01a-setup.qmd

live-01b-data-science-lifecycle-so...

Render

Run

Source Visual

Normal

Format Insert Table

3. calculating summary statistics (like counts or the mean)

## Gapminder dat

Code Editor

**Goal:** Calculate the me for 2007.

```{r}

# before loading library, write code

library(dplyr)
```

Data import
Gapminder data
Data tidying
Gapminder data
Data transformation
Gapminder data
Data visualisation...
Gapminder data
Data communicat...

Environment History Connections Git Tutorial

Import Dataset

172 MiB

List

R

Global Environment

Environment is empty

Environment

Git

Console Terminal Background Jobs

R 4.3.0 · /cloud/project/

```
R version 4.3.0 (2023-04-21) -- "Already Tomorrow"
Copyright (C) 2023 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gn

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
```

Console

Files Plots Packages Help Viewer Presentation

New Folder New Blank File Upload Delete Rename More

Cloud > project

| Name | Size | Modified |
|---|---|---|
| README.md | 544 B | Jun 6, 2023, 11:02 AM |
| wk-01 | | |

File Manager

Viewer

ds4owd-001.github.io/website/

open wash data.

# Screen setup - Poll

# One screen

# Two screens or more

# Hello Quarto

# Meeting you where you are

I'll **assume** you

- do not have R or git experience
- have not worked in an IDE before (e.g. RStudio IDE)
- want to learn about R
- want to learn about Quarto and publishing
- want to learn about project management with GitHub

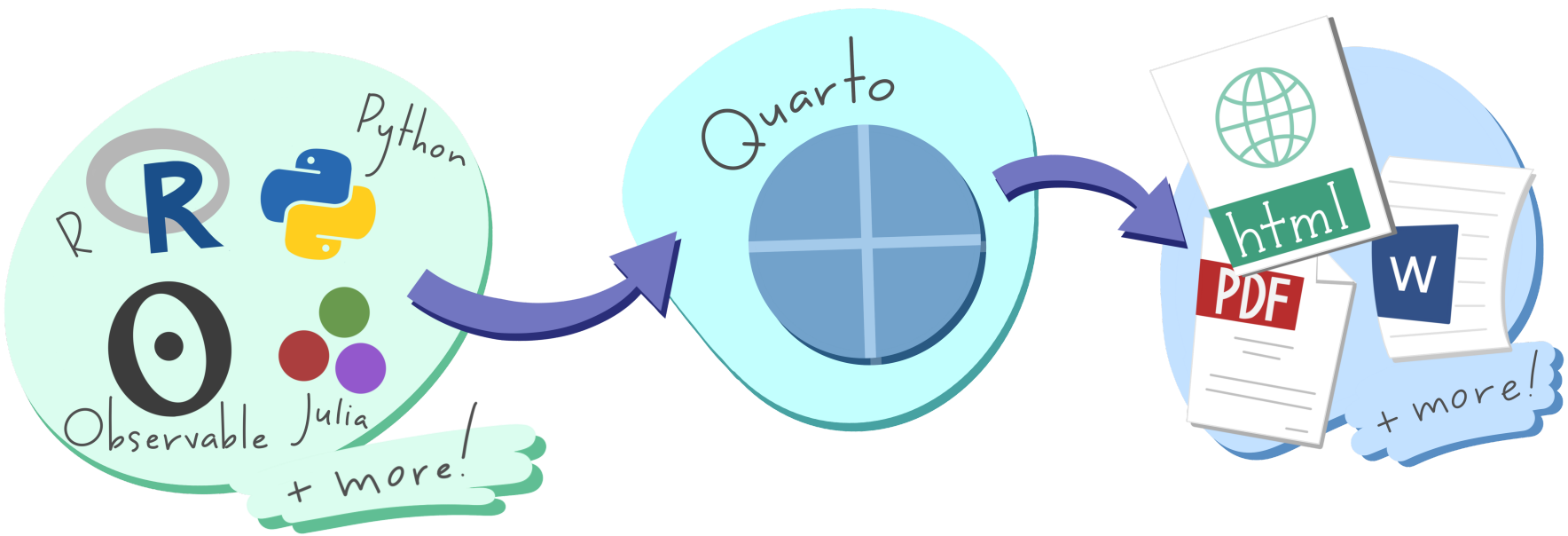
I'll **teach** you

- R
- Quarto syntax and formats
- Markdown
- Git via RStudio GUI
- GitHub issues, project management, and publishing

# What is Quarto?

# Quarto ...

- is a new, open-source, scientific, and technical publishing system
- aims to make the process of creating and collaborating dramatically better



# My turn: A tour of Quarto

Sit back and enjoy!

# Your turn: Log into Posit Cloud with GitHub account

- Go to the Posit Cloud Sign Up page: login.posit.cloud/register
- Click on the Sign Up with GitHub button.
- Enter your GitHub username and password when prompted.
- Open and accept the workspace invitation (Link is in the Zoom chat now).
- Bookmark the address of the open tab in your browser.

**GitHub Authorisation**

- If this is your first time logging in to Posit Cloud with your GitHub account, you will be prompted to authorize Posit Cloud to access your GitHub account information.
- Once you have authorized access, you will be redirected back to the Posit Cloud website and logged in to your account.

# Take a break

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

Photo by Blake Wisz

ds4owd-001.github.io/website/

10:00

open wash data.

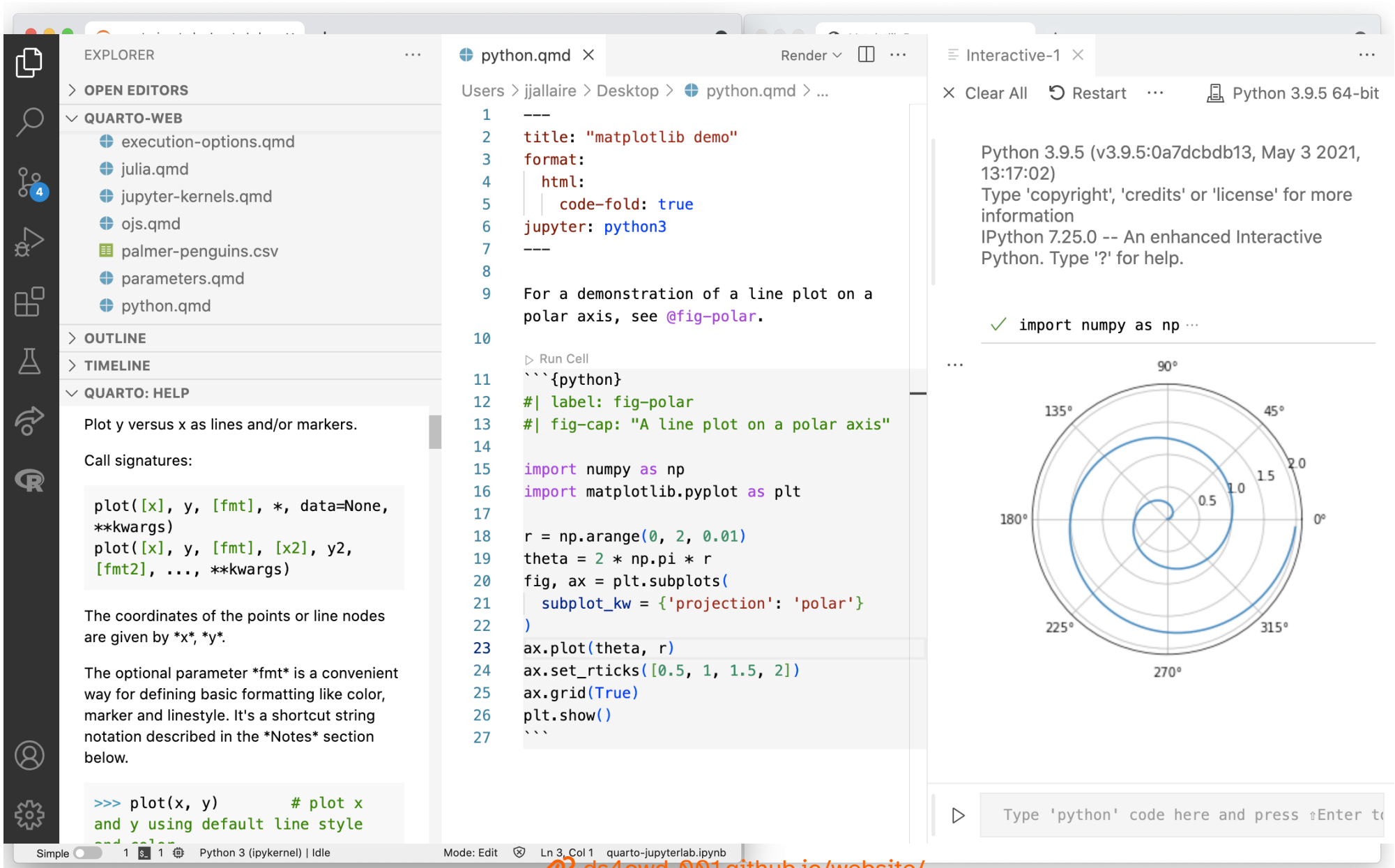
# Your turn: md-01-exercises

1. Open posit.cloud in your browser (use your bookmark).
2. Open the ds4owd workspace for the course.
3. Click `Start` next to `md-01-exercises`.
4. In the File Manager in the bottom right window, locate the `hello-quarto.qmd` file and click on it to open it in the top left window.
5. Render the document.
6. Add `author:` to the YAML header and add your name
7. Re-render the document
8. Inspect components of the document and make one more update and re-render.
9. Discuss notes about updates you've made with your neighbor. Note any aspects of the document that are not clear after the tour and your first interaction with it.



10:00

# From the comfort of your own workspace




ds4owd-001.github.io/website/

open wash data.

# Quarto formats

# One install, “Batteries included”

- RMarkdown grew into a large ecosystem, with **varying syntax**.
- Quarto comes **“batteries included”** straight out of the box
  - HTML reports and websites
  - PDF reports
  - MS Office (Word, Powerpoint)
  - Presentations (Powerpoint, Beamer, `revealjs`)
  - Books
- Any language, *exact* same approach and syntax

# Many Quarto formats

| Feature | R Markdown | Quarto |
|---|---|---|
| Basic Formats | html_document | html |
| | pdf_document | pdf |
| | word_document | docx |
| Beamer | beamer_presentation | beamer |
| PowerPoint | powerpoint_presentation | pptx |
| HTML Slides | xaringan | revealjs |
| | ioslides | |
| | revealjs | |
| Advanced Layout | tufte | Quarto Article Layout |
| | distill | |

# Many Quarto formats

| Feature | R Markdown | Quarto |
|---|---|---|
| Cross References | html_document2<br>pdf_document2<br>word_document2 | Quarto Crossrefs |
| Websites & Blogs | blogdown<br>distill | Quarto Websites<br>Quarto Blogs |
| Books | bookdown | Quarto Books |
| Interactivity | Shiny Documents | Quarto Interactive Documents |
| Journal Articles | rticles | Journal Articles |
| Dashboards | flexdashboard | Quarto Dashboards |

# Your turn: Create a new Quarto document

In your exercises project in RStudio on Posit Cloud, go to File > New File > Quarto document to create a Quarto document with HTML output.

- Render the document, which will ask you to give it a name – you can use `my-first-document.qmd`.

Use the visual editor for the next steps.

- Add a title and your name as the author.
- Create four sections with headings of level 2 (Introduction, Methods, Results, Conclusions).
- **Stretch goal**: Add a table of contents. Note: Watch out for the indentation.
- **Stretch goal**: Change the html theme to `sketchy`. Tipp: Check quarto.org and use search function 🔍 with "HTML theming"

# Version Control

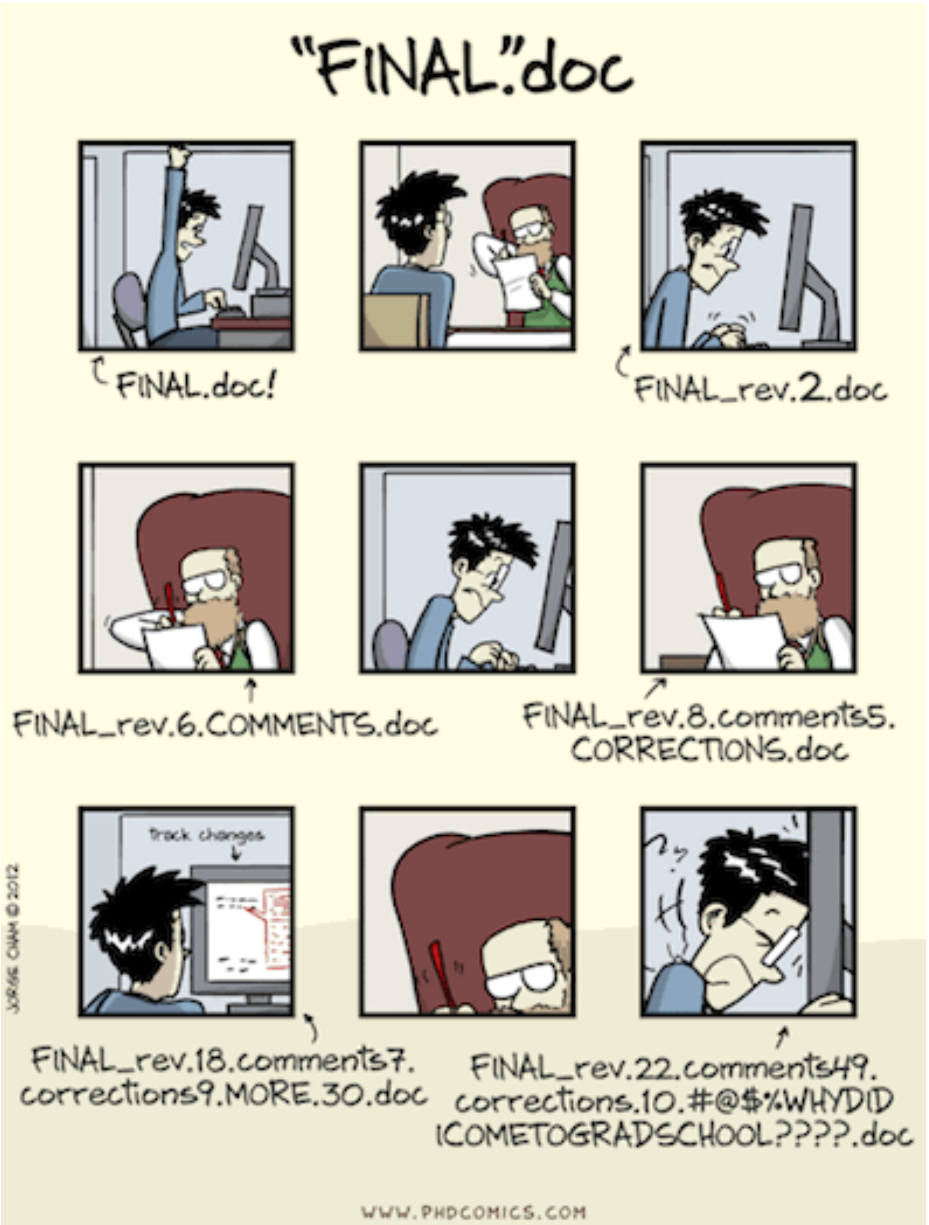
# Version Control with Git and GitHub

A way to share files with others, so they can:

- download
- re-use
- contribute

You can view the history of files, and jump back in time to any point.

# Why is it useful?




ds4owd-001.github.io/website/

# Git and GitHub

- Git is a software for version control
- Created in 2005
- Popular among programmers collaboratively developing code
- Tracks changes in a set of files (directory/folder/repository)

- GitHub is a hosting platform for version control using Git
- Launched in 2008, aquired by Microsoft in in 2018, Microsoft for US$ 7.5 billion
- > 100 million Users (20.5 in 2022 alone) (October, 2023)
- Social media for software developers

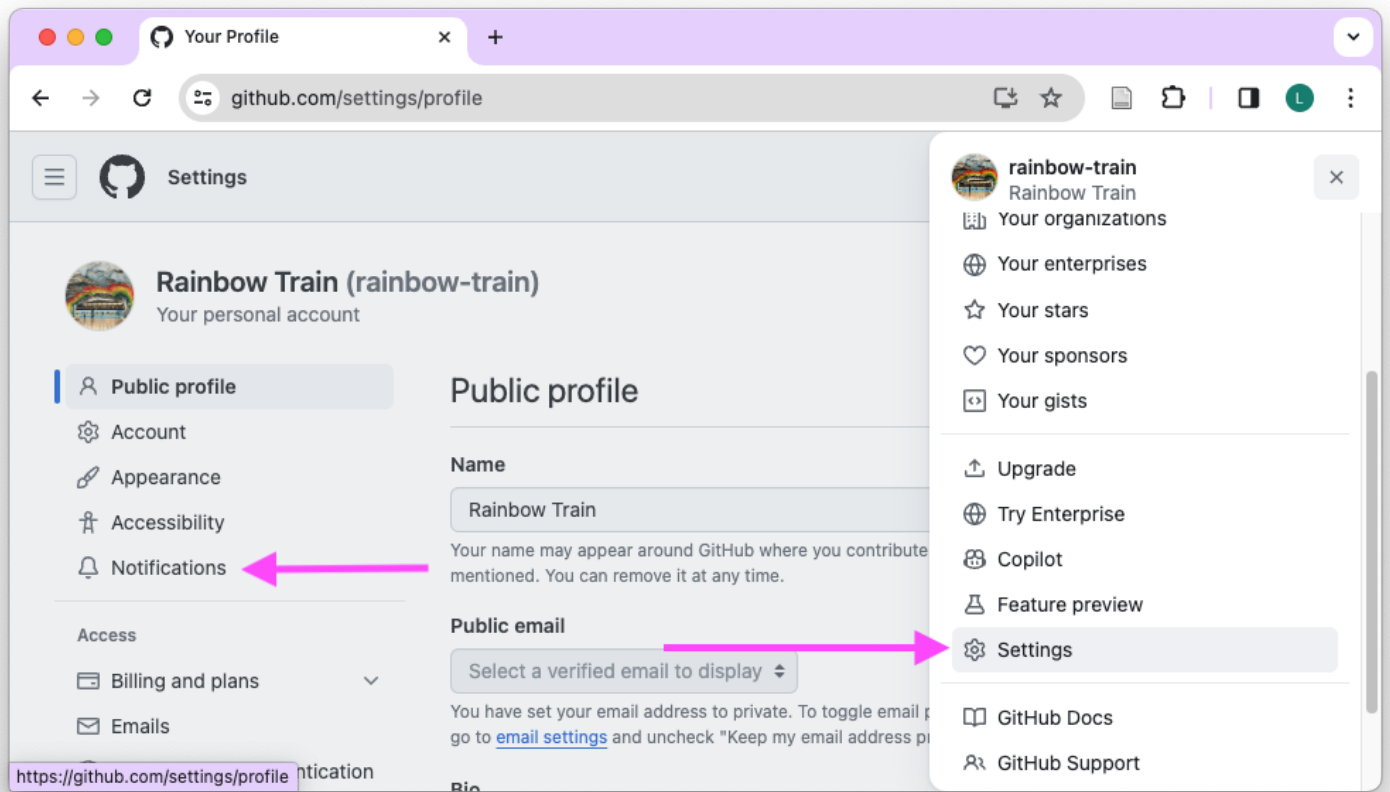
# My turn: A tour of GitHub

Sit back and enjoy!

ds4owd-001.github.io/website/

# Our turn: Configure Notifications settings

> Currently, you receive emails when someone mentions you in a comment on GitHub. Let's change the settings to receive notifications On GitHub.

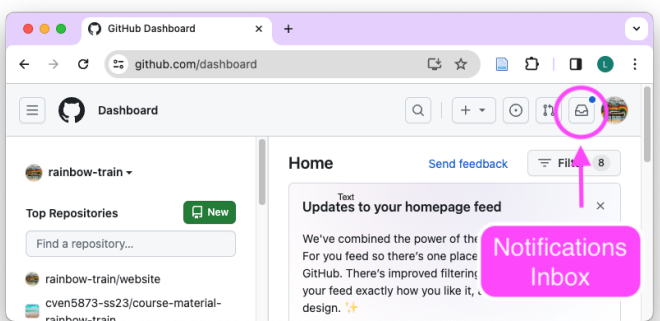



ds4owd-001.github.io/website/

05:00

# Your turn: Create an issue on GitHub

1. Open github.com in your browser and login with your credentials
2. Exchange your GitHub username with your room partner
3. Find and open the md-001-assignments-USERNAME repository that ends with your GitHub username
4. Find the issue tracker
5. Open an issue with the title “Support for module 1 homework”
6. Add your room partner to the list of Assignees on the right panel
7. Add a comment to the issue and tag Mian with `@mianzg`, Sophia with `@sskorik001`, and your `room partner` to ask for support during the homework assignments
8. Click submit new issue
9. Check if you have received a notification on GitHub, and open the inbox.
10. Open the issue and respond to the comment of your room partner.




ds4owd-001.github.io/website/

10:00

# Take a break

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

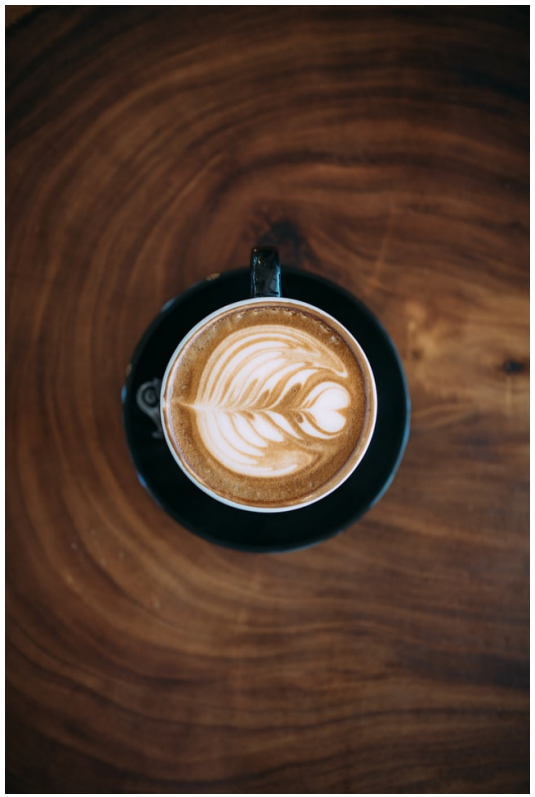


Photo by Blake Wisz

ds4owd-001.github.io/website/

10:00

open wash data.

# 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.

````
---
title: "ggplot2 demo"
date: "5/23/2023"
format: html
---

## MPG

There is a relationship between city and highway mileage.

```{r}
#| label: fig-mpg

library(ggplot2)

ggplot(mpg, aes(x = cty, y = hwy)) +
  geom_point() +
  geom_smooth(method = "loess")
```
````



Source: https://en.wikipedia.org/wiki/Literate_programming

# Metadata

# YAML

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

```yaml
---
key: value
---
```

# Output options

```yaml
---
format: something
---
```

```yaml
---
format: html
---
```

```yaml
---
format: pdf
---
```

```yaml
---
format: revealjs
---
```

# Output option arguments

Indentation matters!

```yaml
---
format:
  html:
    toc: true
    code-fold: true
---
```

# 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 `:`

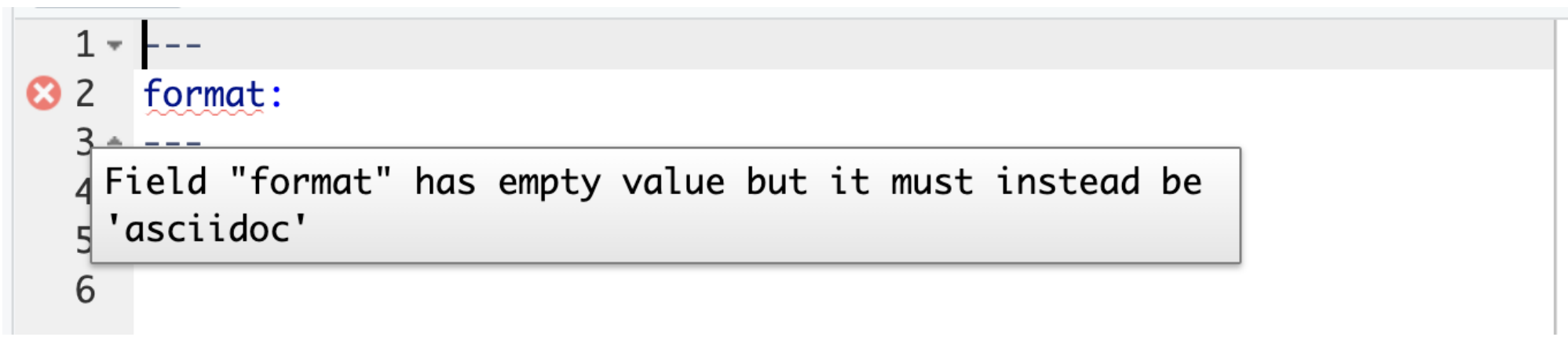
```yaml
format: html
```

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

```yaml
format:
  html:
    toc: true
```

# Quarto linting

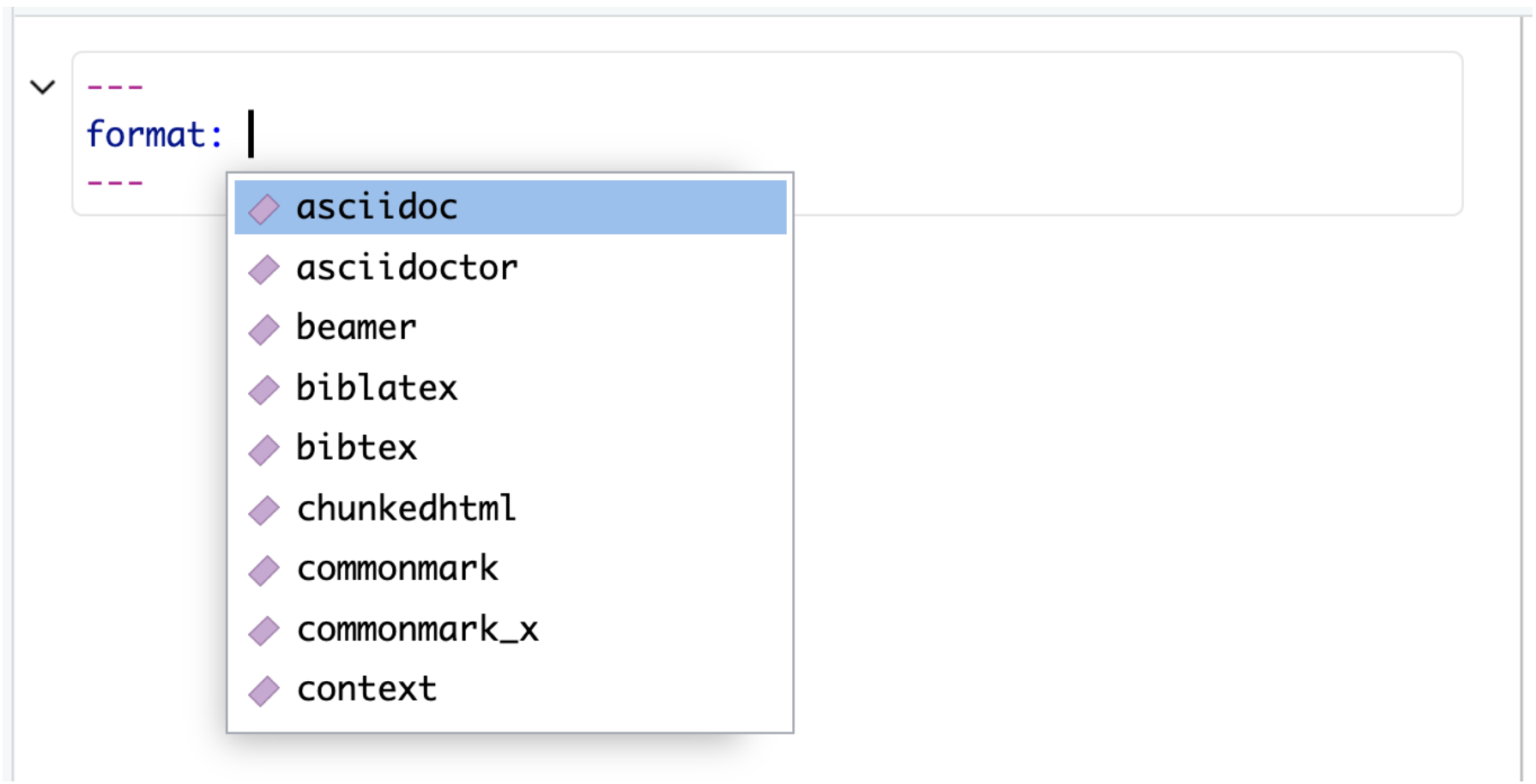
Lint, or a linter, is a static code analysis tool used to flag programming errors, bugs, stylistic errors and suspicious constructs.




Source: https://en.wikipedia.org/wiki/Lint_(software)

# Quarto YAML Intelligence

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



# R fundamentals

# Packages

## base R

```
1 sqrt(49)
2 sum(1, 2)
```

- Functions come with R

## R Packages

```
1 library(dplyr)
```

- Installed once in the Console: `install.packages("dplyr")`
- Loaded per script

# Functions & Arguments

```
1 library(dplyr)
2
3 filter(.data = gapminder,
4        year == 2007)
```

- Function: `filter()`
- Argument: `.data =`
- Arguments following: `year == 2007` **What do do with the data**

# Objects

```
1 library(dplyr)
2
3 gapminder_yr_2007 <- filter(.data = gapminder,
4                             year == 2007)
```

- Function: `filter()`
- Argument: `.data =`
- Arguments following: `year == 2007` **What do do with the data**
- Object: `gapminder_yr_2007`

# Operators

```
1 library(dplyr)
2
3 gapminder_yr_2007 <- gapminder |>
4   filter(year == 2007)
```

- Function: `filter()`
- Argument: `.data =`
- Arguments following: `year == 2007` **What do do with the data**
- Object: `gapminder_yr_2007`
- Assignment operator: `<-`
- Pipe operator: `|>`

# Rules

Rules of `dplyr` functions:

- First argument is always a data frame
- Subsequent arguments say what to do with that data frame
- Always return a data frame
- Don’t modify in place

# Course information

ds4owd-001.github.io/website/

# Weekly Structure

| | |
|---|---|
| **Monday** | |
| **Tuesday** | Module from 2 pm to 4:30 pm CET |
| **Wednesday** | |
| **Thursday** | Office hours on Zoom (2 pm to 3:30 pm CET) |
| **Friday** | |

# Homework assignments

- Weekly assignments (module 1 homework is required for participation)
- Submitted as rendered Quarto documents on GitHub
- Reviewed by course instructors for errors
- Management and support through GitHub issue tracker

# Capstone Project

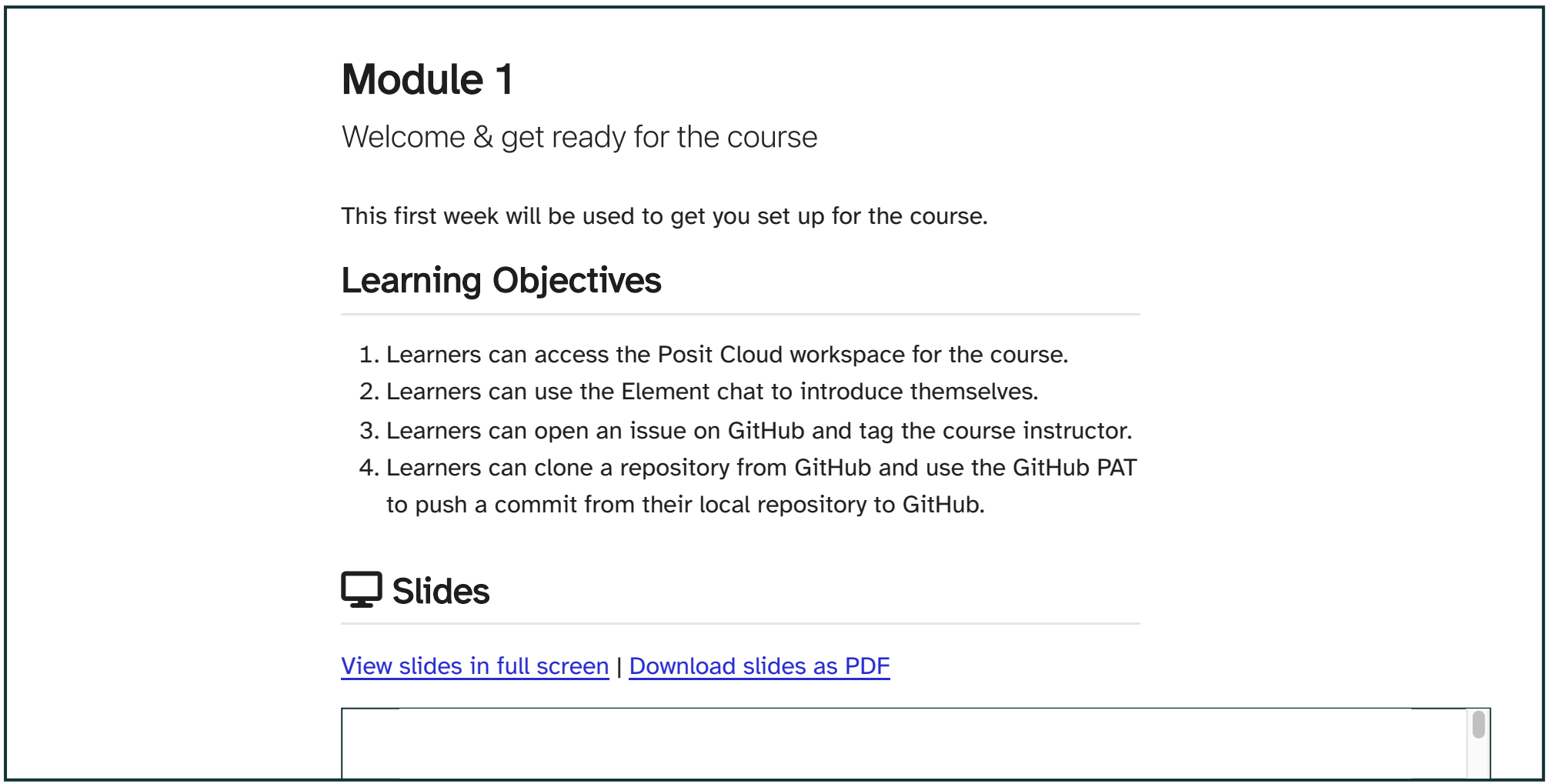
- Data analysis project report with a dataset of your choice
- Submitted as rendered Quarto document on GitHub
- Submission required for successful completion of the course

# Homework assignments module 1

ds4owd-001.github.io/website/

# Module 1 documentation

ds4owd-001.github.io/website/modules/md-01.html



## Module 1

Welcome & get ready for the course

This first week will be used to get you set up for the course.

### Learning Objectives

1. Learners can access the Posit Cloud workspace for the course.
2. Learners can use the Element chat to introduce themselves.
3. Learners can open an issue on GitHub and tag the course instructor.
4. Learners can clone a repository from GitHub and use the GitHub PAT to push a commit from their local repository to GitHub.

### Slides

View slides in full screen | Download slides as PDF

# Homework due date

- Homework assignment due: Monday, November 6th

# Wrap-up

ds4owd-001.github.io/website/

# Thanks! 🌻

Slides created via revealjs and Quarto: https://quarto.org/docs/presentations/revealjs/ Access slides as PDF on GitHub