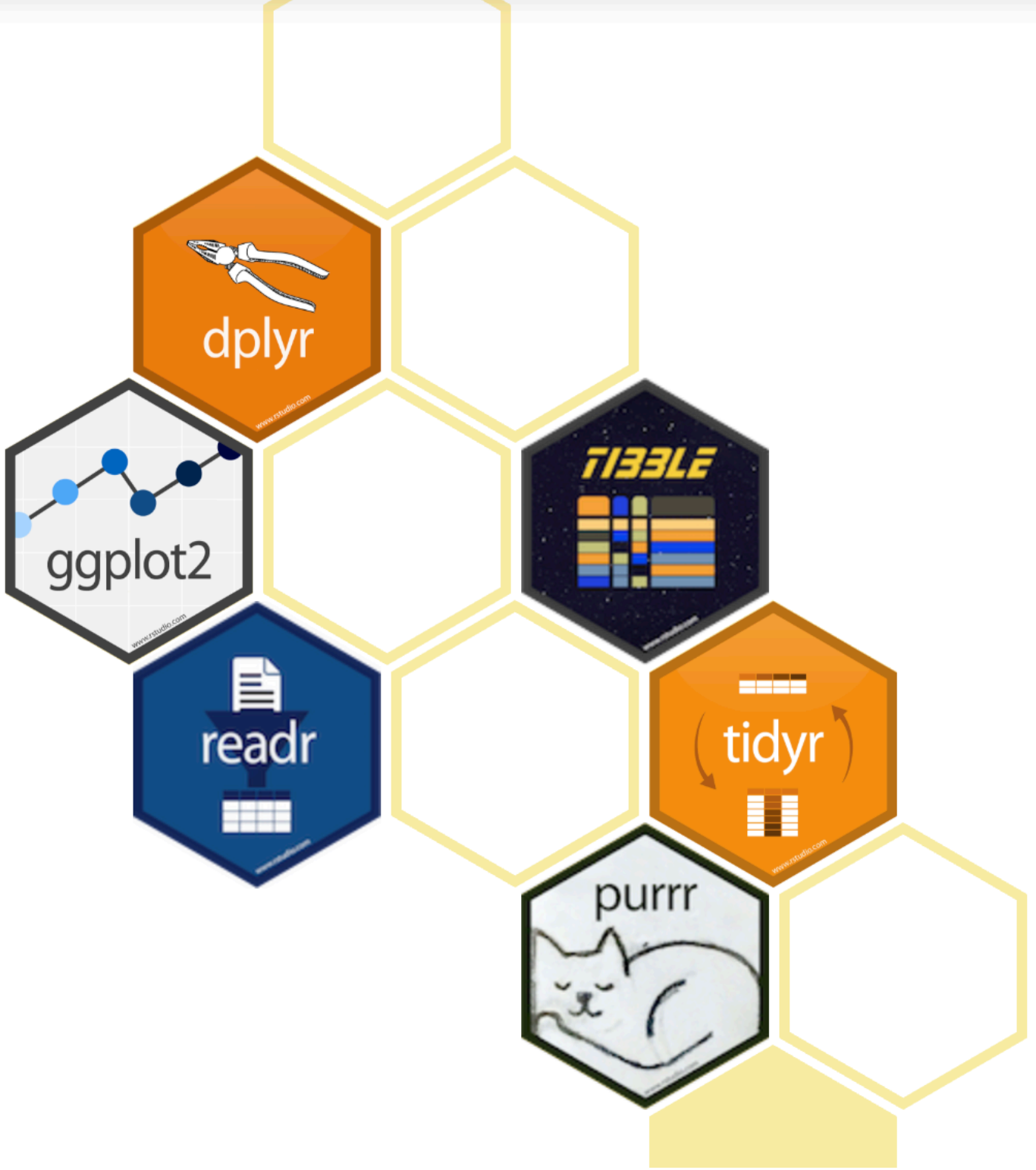
Introduction to the Tidyverse

Import, wrangle, model, and communicate data

2024-08-13


dplyr
ggplot2
tibble
readr
tidyr
purrr

Working with data in R

the tidyverse is a collection of *friendly and consistent* tools for data analysis and visualization.

They live as R packages, each of which does one thing well.

`library(tidyverse)` will load the core packages:

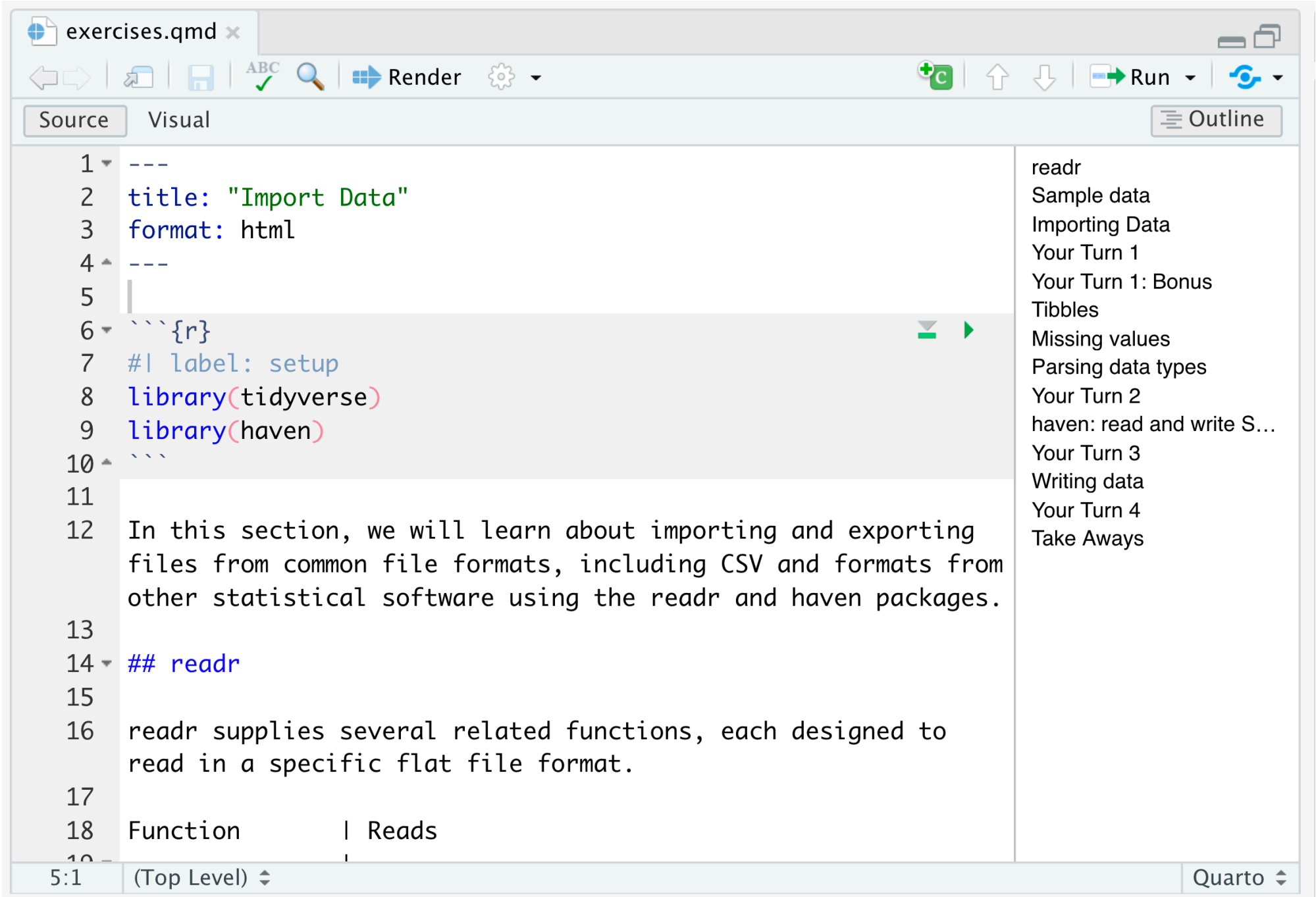
ggplot2, for data visualisation.
dplyr, for data manipulation.
tidyr, for data tidying.
readr, for data import.
purrr, for functional programming.
tibble, for tibbles, a modern re-imagining of data frames.
stringr, for strings.
forcats, for factors.
lubridate, for dates and times.

This course is hands on!

Each section has an exercises file:

exercises.qmd

exercises.qmd


exercises.qmd
Render
Run
Source
Visual
Outline
1 ---
2 title: "Import Data"
3 format: html
4 ---
5
6 ```{r}
7 #| label: setup
8 library(tidyverse)
9 library(haven)
10 ```
11
12 In this section, we will learn about importing and exporting files from common file formats, including CSV and formats from other statistical software using the readr and haven packages.
13
14 ## readr
15
16 readr supplies several related functions, each designed to read in a specific flat file format.
17
18 Function | Reads
readr
Sample data
Importing Data
Your Turn 1
Your Turn 1: Bonus
Tibbles
Missing values
Parsing data types
Your Turn 2
haven: read and write S...
Your Turn 3
Writing data
Your Turn 4
Take Aways
5:1
(Top Level)
Quarto

Code chunks

```{r}
csv_data <- read_csv(
  "a,b,c,d
  1,2,3,4
  5,6,7,8",
  col_types = ""
)

csv_data
```

Running code chunks

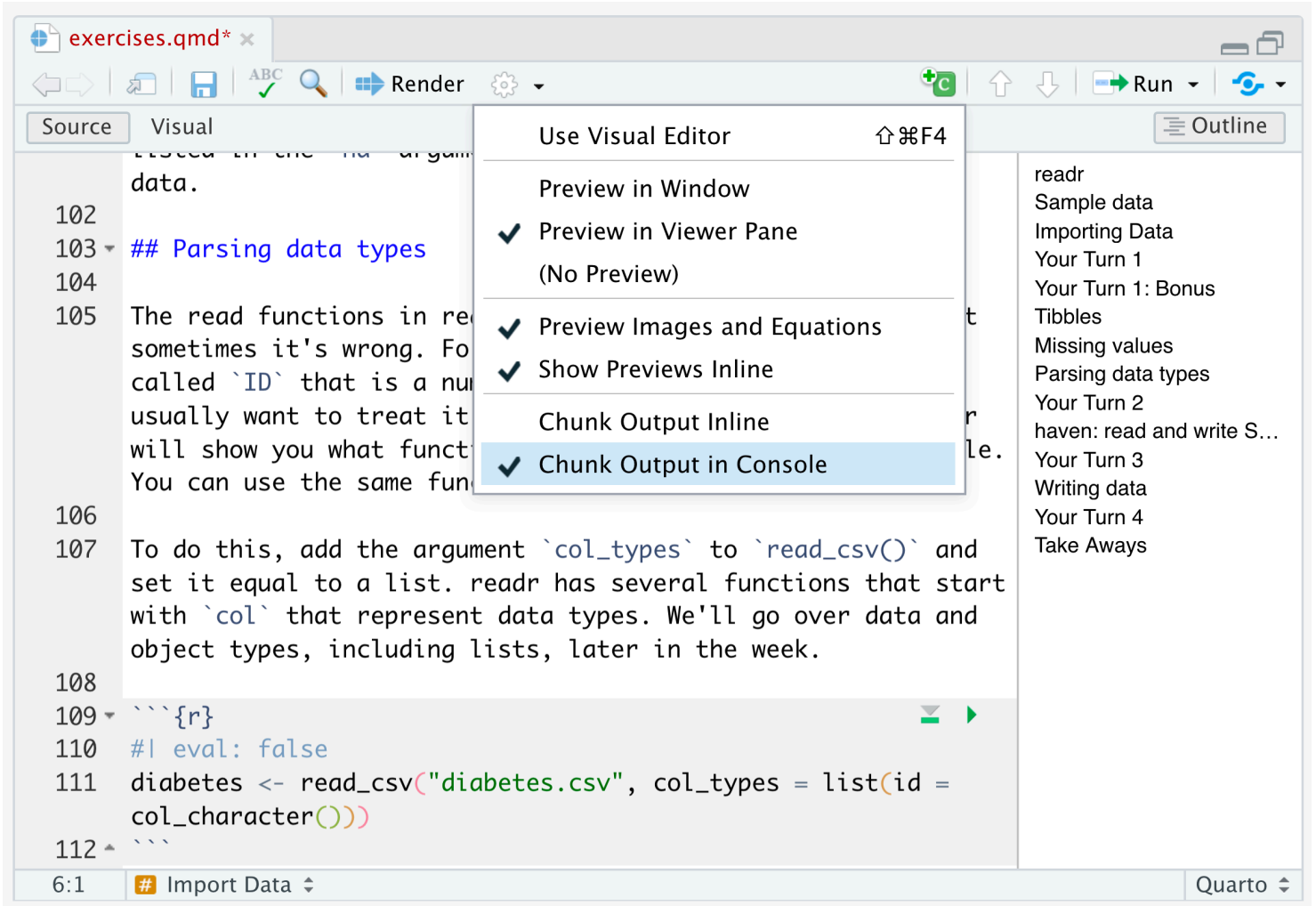
```{r}
csv_data <- read_csv(
  "a,b,c,d
  1,2,3,4
  5,6,7,8",
  col_types = ""
)

csv_data
```

a <dbl>	b <dbl>	c <dbl>	d <dbl>
1	2	3	4
5	6	7	8

2 rows

Outputting to the console



Project contents

```
├── 01-dplyr_5verbs
│   ├── cheatsheet_dplyr_5verbs.pdf
│   ├── diabetes.csv
│   ├── exercises.qmd
│   ├── slides.pdf
```