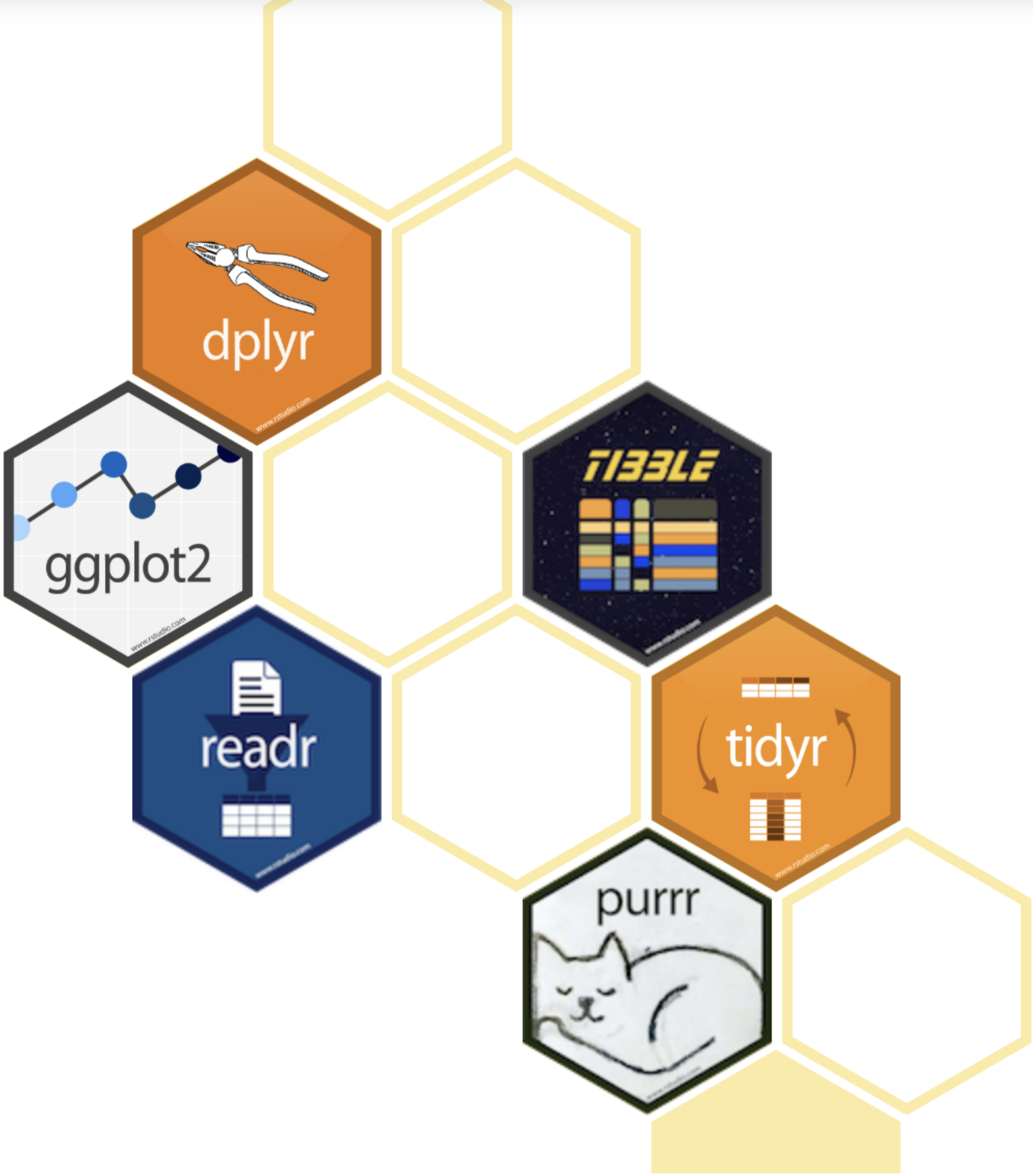
Introduction to the Tidyverse

Import, wrangle, model, and communicate data

2023-08-02


dplyr
ggplot2
TIBBLE
readr
tidyr
purrr
www.rstudio.com

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.

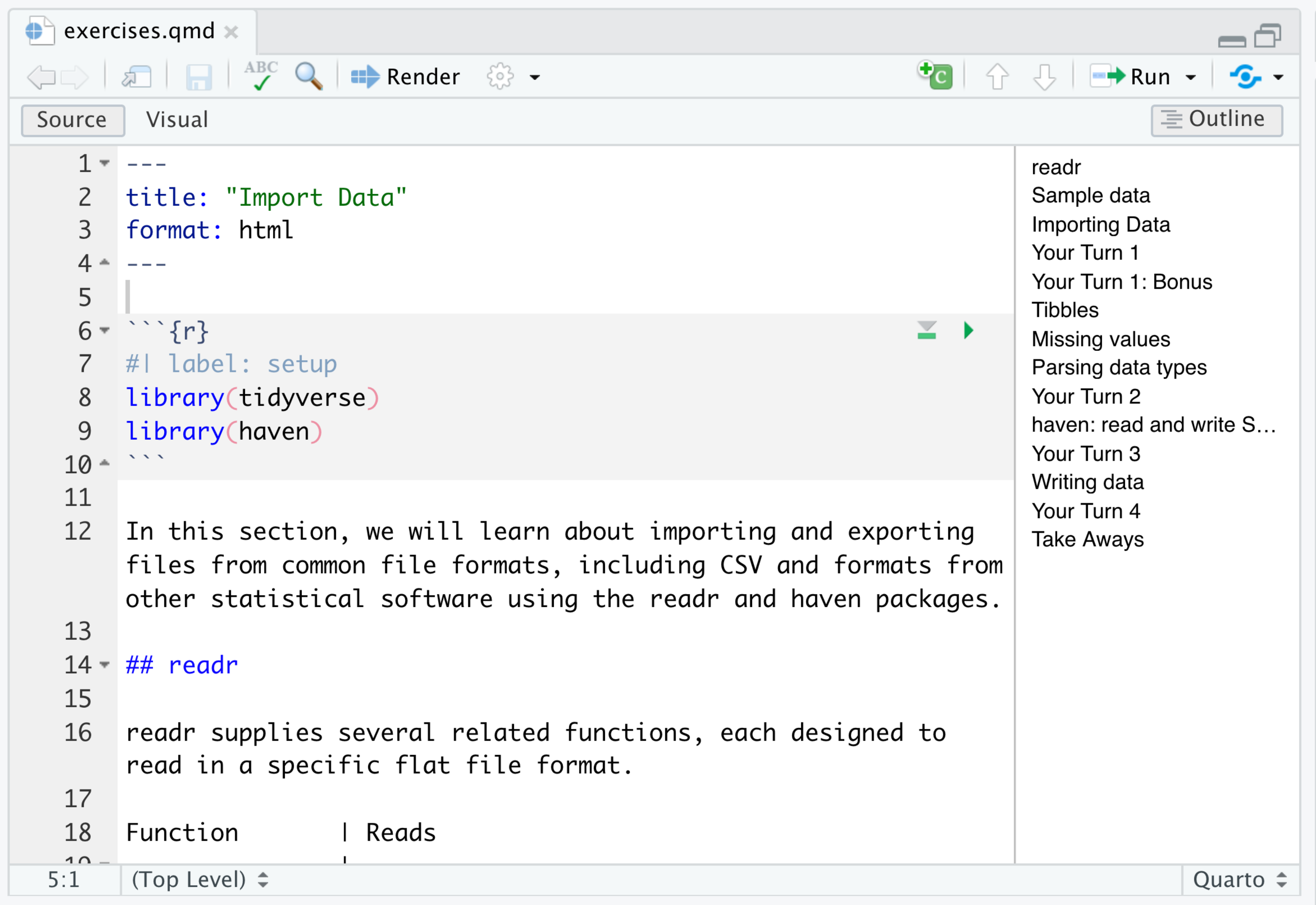
forcats, for factors.

lubridate, for dates and times.

This course is hands on!

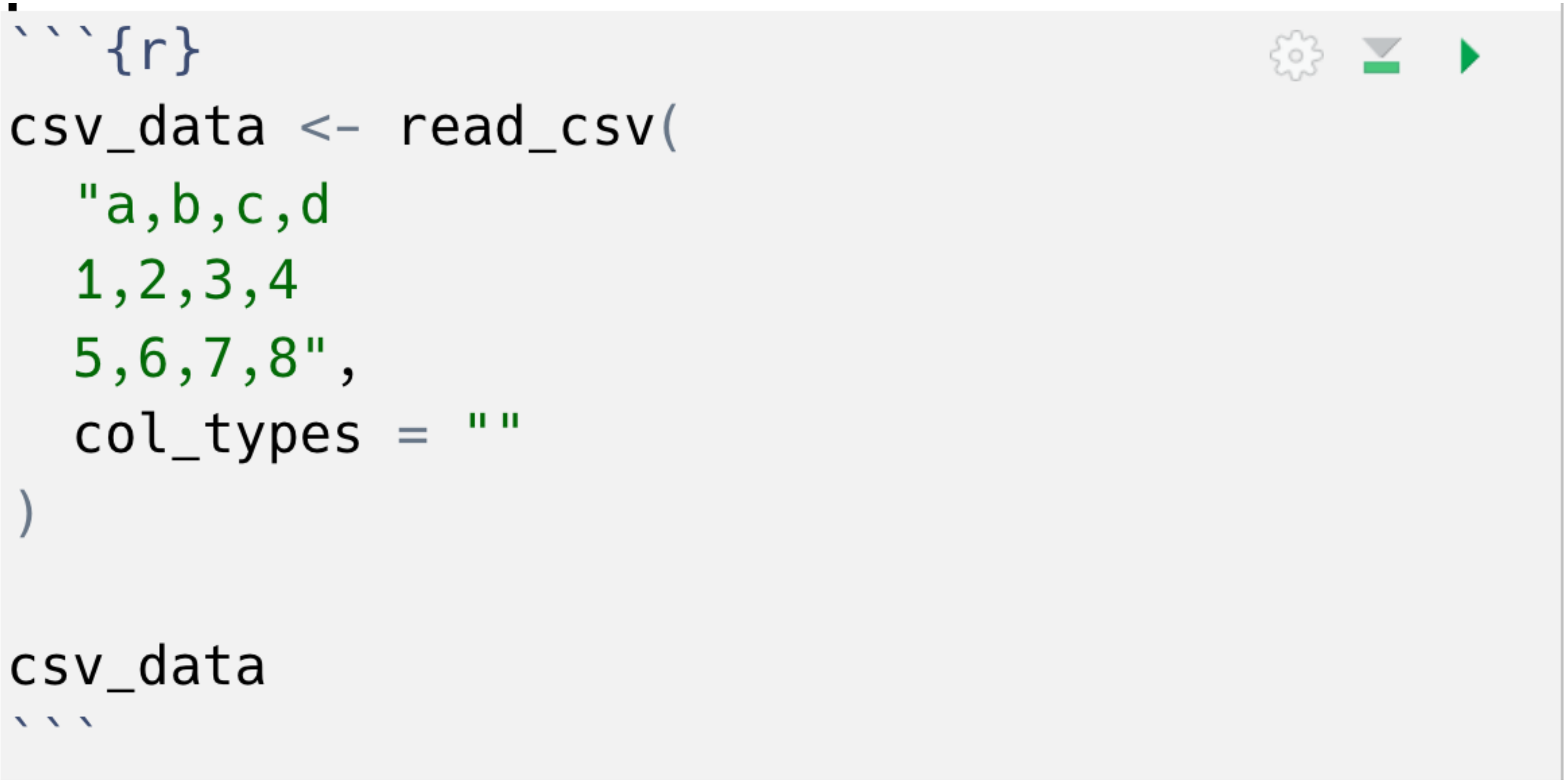
Each section has an exercises file: ***`exercises.qmd`***

exercises.qmd



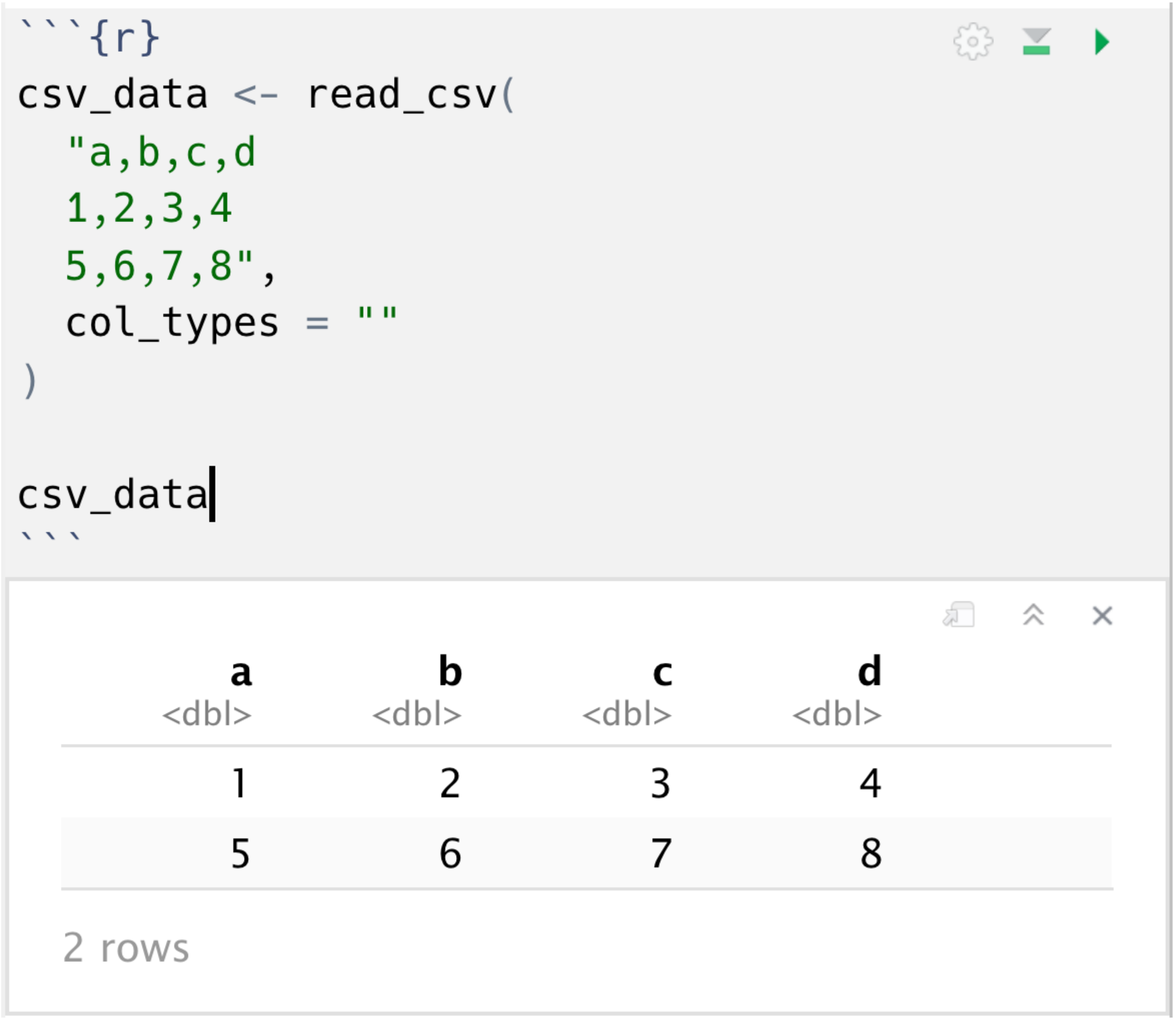
~~~
---
title: "Import Data"
format: html
---

```{r}
#| label: setup
library(tidyverse)
library(haven)
```

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.

## readr

readr supplies several related functions, each designed to read in a specific flat file format.

Function        | Reads
~~~

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

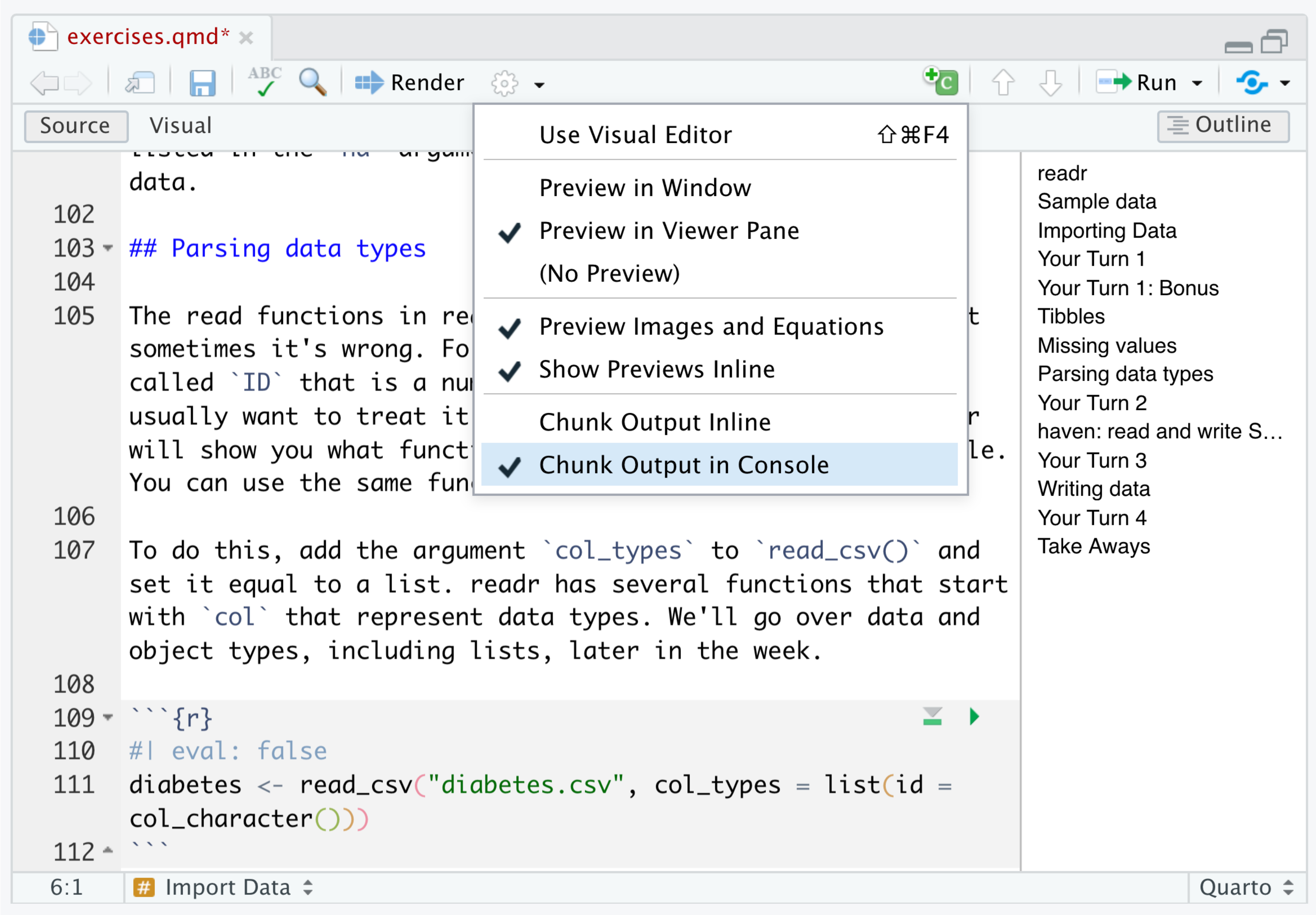


~~~
```{r}
csv_data <- read_csv(
  "a,b,c,d
  1,2,3,4
  5,6,7,8",
  col_types = ""
)

csv_data
```
~~~

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

2 rows

Outputting to the console



data.

Parsing data types

The read functions in re t
sometimes it's wrong. Fo
called `ID` that is a nu
usually want to treat it r
will show you what funct le.
You can use the same fun

To do this, add the argument `col_types` to `read_csv()` and set it equal to a list. readr has several functions that start with `col` that represent data types. We'll go over data and object types, including lists, later in the week.

~~~
```{r}
#| eval: false
diabetes <- read_csv("diabetes.csv", col_types = list(id =
col_character()))
```
~~~

Project contents

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