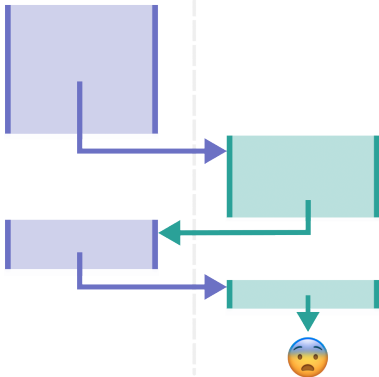


## Dynamic reporting with Quarto

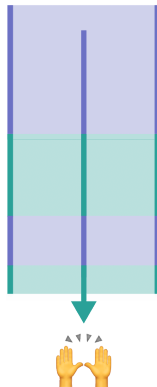


Analysis

Writing



Analysis + Writing



# Literate programming

*...a programming paradigm introduced by Donald Knuth in which a program is given as an explanation of the program logic in a natural language, such as English, interspersed with snippets of source code.*

— Wikipedia

The key point here is the **mixing** of text and code in a single document.

# Literate programming

This brings many benefits:

- Reproducibility
- Efficiency
- Reduces human error (e.g. copy and paste).
- Combines code with documentation

# Quarto is the successor to RMarkdown



Learn more @ <https://quarto.org>.



Artwork by @allison\_horst

```
---  
title: "My Quarto Report"  
author: "Ewan Carr"  
format: html  
---
```

## Quarto

Quarto enables you to weave together content and executable code into a finished document. To learn more about Quarto see <https://quarto.org>.

This is a code chunk:

```
```{r}  
library(tidyverse)  
library(lme4)  
```
```

You can add options to code chunks like this

```
```{r}  
#| echo: false  
df <- read_csv("raw_data.csv")  
```
```

# Markdown

- Markdown is an **easy-to-read, easy-to-write** plain text format.

<https://daringfireball.net/projects/markdown/>.

```
# Heading 1
```

```
## Heading 2
```

```
This is some bold text. This is italics.
```

- Quarto uses **Pandoc**, a piece of software that converts one markup format into another.

<http://pandoc.org/>

- With this, we can convert RMarkdown to Microsoft Word, PDF (via  $\text{\LaTeX}$ ), HTML, PowerPoint, ePub...



# It's not just for documents.

- Papers
- Books
- Websites
- Presentations
- Documentation

<https://quarto.org/docs/guide/>  
<https://quarto.org/docs/gallery>

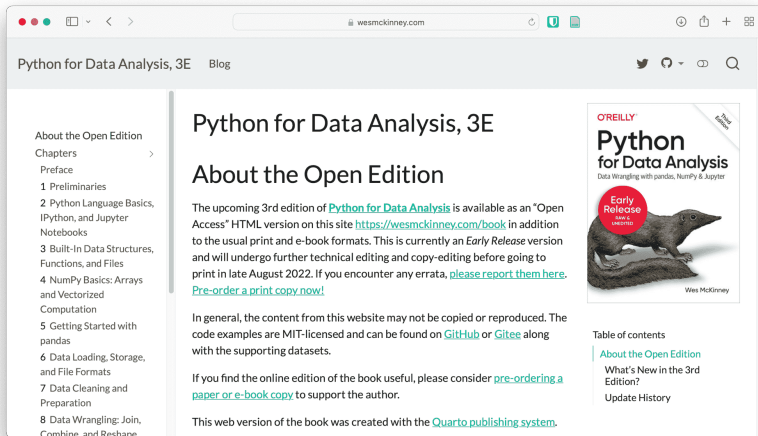
# It's not just for documents.

- Papers
- Books
- Websites
- Presentations
- Documentation

<https://quarto.org/docs/guide/>  
<https://quarto.org/docs/gallery>

Tip: Write your data cleaning scripts in Quarto.

# And it's not just for R...



The screenshot shows a web browser displaying the website for "Python for Data Analysis, 3E" by Wes McKinney. The browser's address bar shows the URL <https://wesmckinney.com>. The page has a light gray header with the title "Python for Data Analysis, 3E" and the word "Blog" on the left, and social media icons on the right. A left sidebar contains a table of contents with links to "About the Open Edition", "Chapters", "Preface", and numbered sections from 1 to 8. The main content area features the title "Python for Data Analysis, 3E" and "About the Open Edition". The text describes the upcoming 3rd edition as an "Open Access" HTML version available on the site, with a link to <https://wesmckinney.com/book>. It mentions that this is an "Early Release" version and will undergo further editing before print in late August 2022. A red circular badge with the text "Early Release READ & UNEDITED" is overlaid on the book cover image. The book cover itself shows the title "Python for Data Analysis" with the subtitle "Data Wrangling with pandas, NumPy & Jupyter" and the author's name "Wes McKinney". Below the main text, there are links to "Table of contents", "About the Open Edition", "What's New in the 3rd Edition?", and "Update History". A footer note states that the web version was created with the "Quarto publishing system".

Python for Data Analysis, 3E Blog

## Python for Data Analysis, 3E

### About the Open Edition

The upcoming 3rd edition of [Python for Data Analysis](#) is available as an "Open Access" HTML version on this site <https://wesmckinney.com/book> in addition to the usual print and e-book formats. This is currently an *Early Release* version and will undergo further technical editing and copy-editing before going to print in late August 2022. If you encounter any errata, [please report them here](#). [Pre-order a print copy now!](#)

In general, the content from this website may not be copied or reproduced. The code examples are MIT-licensed and can be found on [GitHub](#) or [Gitee](#) along with the supporting datasets.

If you find the online edition of the book useful, please consider [pre-ordering a paper or e-book copy](#) to support the author.

This web version of the book was created with the [Quarto publishing system](#).

**O'REILLY**  
**Python**  
for Data Analysis  
Data Wrangling with pandas, NumPy & Jupyter  
Early Release  
READ & UNEDITED  
Wes McKinney

Table of contents  
[About the Open Edition](#)  
What's New in the 3rd Edition?  
Update History

<https://wesmckinney.com/book/>

# Demonstration!

1. Create a new document
2. Compiling ('knitting')
3. Markdown syntax
4. Code chunks and inline code.
5. The visual editor
6. Figures
7. Chunk options
  - `eval`
  - `echo`
  - `tbl-cap`
8. Layout options
9. Citations

Publication-ready tables

# Packages for creating tables

There are many options:

1. `tinytable`

<https://vincentarelbundock.github.io/tinytable>

2. `tabyl` from the `janitor` package

<https://sfirke.github.io/janitor/index.html>

3. `gt` and `gtsummary`

<https://gt.rstudio.com>

<https://www.danielsjoberg.com/gtsummary>

4. `table1`

<https://github.com/benjaminrich/table1>

# You could do this manually...

```
> library(tidyverse)
> library(palmerpenguins)
> penguins |>
>   group_by(species) |>
>   summarise(
>     across(where(is.numeric),
>       .fns = list(
>         mean = \(x) mean(x, na.rm = TRUE),
>         sd = \(x) sd(x, na.rm = TRUE)
>       )
>     )
>   )
# A tibble: 3 × 11
  species    bill_length_mm_mean bill_length_mm_sd bill_depth_mm_mean
  <fct>          <dbl>          <dbl>          <dbl>
1 Adelie         38.8           2.66           18.3
2 Chinstrap      48.8           3.34           18.4
3 Gentoo         47.5           3.08           15.0
```

...but it's a lot of work.

```

> penguins |>
>   group_by(species) |>
>   summarise(
>     across(where(is.numeric),
>       .fns = list(
>         mean = \(x) mean(x, na.rm = TRUE),
>         sd = \(x) sd(x, na.rm = TRUE)
>       )
>     )
>   ) |>
>   pivot_longer(-species) |>
>   mutate(var = str_replace(name, "_mean$|_sd$", ""),
>     measure = if_else(str_detect(name, "_mean$"), "mean", "sd")
>   )
>   select(-name) |>
>   pivot_wider(names_from = measure,
>     values_from = value) |>
>   mutate(cell = make_cell(mean, sd)) |>
>   select(species, var, cell) |>
>   pivot_wider(names_from = species,
>     values_from = cell)

```

|   | var               | Adelie         | Chinstrap      | Gentoo         |
|---|-------------------|----------------|----------------|----------------|
|   | <chr>             | <glue>         | <glue>         | <glue>         |
| 1 | bill_length_mm    | 38.8 [2.7]     | 48.8 [3.3]     | 47.5 [3.1]     |
| 2 | bill_depth_mm     | 18.3 [1.2]     | 18.4 [1.1]     | 15.0 [1.0]     |
| 3 | flipper_length_mm | 190.0 [6.5]    | 195.8 [7.1]    | 217.2 [6.5]    |
| 4 | body_mass_g       | 3700.7 [458.6] | 3733.1 [384.3] | 5076.0 [504.1] |
| 5 | year              | 2008.0 [0.8]   | 2008.0 [0.9]   | 2008.1 [0.8]   |



A flexible package for creating tables of summary statistics and regression models, building on the powerful **gt** package.

gtsummary 2.0.4

🏠 Reference

Articles ▾


News

Search for

🔍

## tbl\_summary() tutorial

Source: `vignettes/articles/tbl_summary.Rmd`



### Introduction

The `tbl_summary()` function calculates **descriptive statistics** for continuous, categorical, and dichotomous variables in **R**, and presents the results in a **beautiful, customizable summary table** ready for publication (for example, Table 1 or demographic tables).

This vignette will walk a reader through the `tbl_summary()` function, and the various functions available to modify and make additions to an existing table summary object.

### Setup

Before going through the tutorial, install and load (gtsummary).

```
# install.packages("gtsummary")
library(gtsummary)
```

#### On this page

- Introduction
- Setup
- Example data set
- Basic Usage
- Customize Output
- Select Helpers
- Multi-line Continuous Summaries
- Advanced Customization
- Set Default Options with Themes
- Survey Data
- Cross Tables

I won't go into depth here; see documentation to learn more.

# tbl\_summary

```
penguins |>
  select(species,
         where(is.numeric)) |>
  tbl_summary()
```

| Characteristic                      | N = 344 <sup>†</sup> |
|-------------------------------------|----------------------|
| species                             |                      |
| Adelie                              | 152 (44%)            |
| Chinstrap                           | 68 (20%)             |
| Gentoo                              | 124 (36%)            |
| bill_length_mm                      |                      |
| Unknown                             | 2                    |
| bill_depth_mm                       | 17.30 (15.60, 18.70) |
| Unknown                             | 2                    |
| flipper_length_mm                   | 197 (190, 213)       |
| Unknown                             | 2                    |
| body_mass_g                         | 4,050 (3,550, 4,750) |
| Unknown                             | 2                    |
| year                                |                      |
| 2007                                | 110 (32%)            |
| 2008                                | 114 (33%)            |
| 2009                                | 120 (35%)            |
| <sup>†</sup> n (%); Median (Q1, Q3) |                      |

# tbl\_summary

```
penguins |>  
  select(species, where(is.numeric)) |>  
  tbl_summary(by = species)
```

| Characteristic    | Adelie<br>N = 152 <sup>1</sup> | Chinstrap<br>N = 68 <sup>1</sup> | Gentoo<br>N = 124 <sup>1</sup> |
|-------------------|--------------------------------|----------------------------------|--------------------------------|
| bill_length_mm    | 38.8 (36.7, 40.8)              | 49.6 (46.3, 51.2)                | 47.3 (45.3, 49.6)              |
| Unknown           | 1                              | 0                                | 1                              |
| bill_depth_mm     | 18.40 (17.50, 19.00)           | 18.45 (17.50, 19.40)             | 15.00 (14.20, 15.70)           |
| Unknown           | 1                              | 0                                | 1                              |
| flipper_length_mm | 190 (186, 195)                 | 196 (191, 201)                   | 216 (212, 221)                 |
| Unknown           | 1                              | 0                                | 1                              |
| body_mass_g       | 3,700 (3,350, 4,000)           | 3,700 (3,475, 3,950)             | 5,000 (4,700, 5,500)           |
| Unknown           | 1                              | 0                                | 1                              |
| year              |                                |                                  |                                |
| 2007              | 50 (33%)                       | 26 (38%)                         | 34 (27%)                       |
| 2008              | 50 (33%)                       | 18 (26%)                         | 46 (37%)                       |

# tbl\_summary

```
penguins |>
  select(species, where(is.numeric)) |>
  tbl_summary(by = species,
              label = list(bill_length_mm = "Bill length (mm)",
                           bill_depth_mm = "Bill depth (mm)",
                           missing_text = "Missing")
```

| Characteristic    | Adelie<br>N = 152 <sup>1</sup> | Chinstrap<br>N = 68 <sup>1</sup> | Gentoo<br>N = 124 <sup>1</sup> |
|-------------------|--------------------------------|----------------------------------|--------------------------------|
| Bill length (mm)  | 38.8 (36.7, 40.8)              | 49.6 (46.3, 51.2)                | 47.3 (45.3, 49.6)              |
| Missing           | 1                              | 0                                | 1                              |
| Bill depth (mm)   | 18.40 (17.50, 19.00)           | 18.45 (17.50, 19.40)             | 15.00 (14.20, 15.70)           |
| Missing           | 1                              | 0                                | 1                              |
| flipper_length_mm | 190 (186, 195)                 | 196 (191, 201)                   | 216 (212, 221)                 |
| Missing           | 1                              | 0                                | 1                              |
| body_mass_g       | 3,700 (3,350, 4,000)           | 3,700 (3,475, 3,950)             | 5,000 (4,700, 5,500)           |
| Missing           | 1                              | 0                                | 1                              |
| year              |                                |                                  |                                |

# tabyl

`tabyl` is a function from the `janitor` package.

It's great for quick tables that don't require the heavy lifting of `tbl_summary`.

```
> library(janitor)
> penguins |>
> tabyl(species)
```

| species   | n   | percent   |
|-----------|-----|-----------|
| Adelie    | 152 | 0.4418605 |
| Chinstrap | 68  | 0.1976744 |
| Gentoo    | 124 | 0.3604651 |

```
> library(janitor)
> penguins |>
> tabyl(species) |>
> adorn_totals() |>
> adorn_pct_formatting()
```

| species   | n   | percent |
|-----------|-----|---------|
| Adelie    | 152 | 44.2%   |
| Chinstrap | 68  | 19.8%   |
| Gentoo    | 124 | 36.0%   |
| Total     | 344 | 100.0%  |

See [this page](#) to learn more.

# Practical: Quarto and tables