



toolkit for the modern statistician

mine çetinkaya-rundel

 bit.ly/modern-toolkit



data
transformation
and tidying with
tidyverse



tidyverse

opinionated collection of R
packages designed for data science

```
library(tidyverse)
```

- **ggplot2**: data visualization
- **dplyr**: data wrangling
- **tidyr**: data tidying
- **readr**: data reading/writing
- **forcats**: working with factors
- **stringr**: working with strings
- **tibble**: modern data frames
- **purrr**: functional programming

```
install.packages(tidyverse)    above + a few more
```



tidyverse

all packages share an underlying design philosophy, grammar, and data structures

- **tidy** data
- data pipelines with `%>%`

tidy data

country	year	cases	population
Afghanistan	1999	1845	19957071
Afghanistan	2000	2666	20395360
Brazil	1999	31737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	212766	128028583

variables

each variable must
have its own column

tidy data

country	year	cases	population
Afghanistan	1999	745	19557071
Afghanistan	2000	2666	20395360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	210766	1280428583

variables

each variable must
have its own column

country	year	cases	population
Afghanistan	1999	745	19557071
Afghanistan	2000	2666	20395360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	210766	1280428583

observations

each observation must
have its own row

tidy data

country	year	cases	population
Afghanistan	1999	745	19987071
Afghanistan	2000	2666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	213766	1280425583

variables

each variable must
have its own column

country	year	cases	population
Afghanistan	1999	745	19987071
Afghanistan	2000	2666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	213766	1280425583

observations

each observation must
have its own row

country	year	cases	population
Afghanistan	1999	745	19987071
Afghanistan	2000	2666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	213766	1280425583

values

each value must have
its own cell

task

I want to find my keys, then start my car, then drive to work, then park my car.

nested

```
find("keys")
```

nested

```
start_car(find("keys"))
```

nested

```
drive(start_car(find("keys")), to = "work")
```

nested

```
park(drive(start_car(find("keys")), to = "work"))
```

pip

```
find("keys")
```

piped

```
find("keys") %>%  
  start_car()
```

piped

```
find("keys") %>%  
  start_car() %>%  
  drive(to = "work")
```

piped

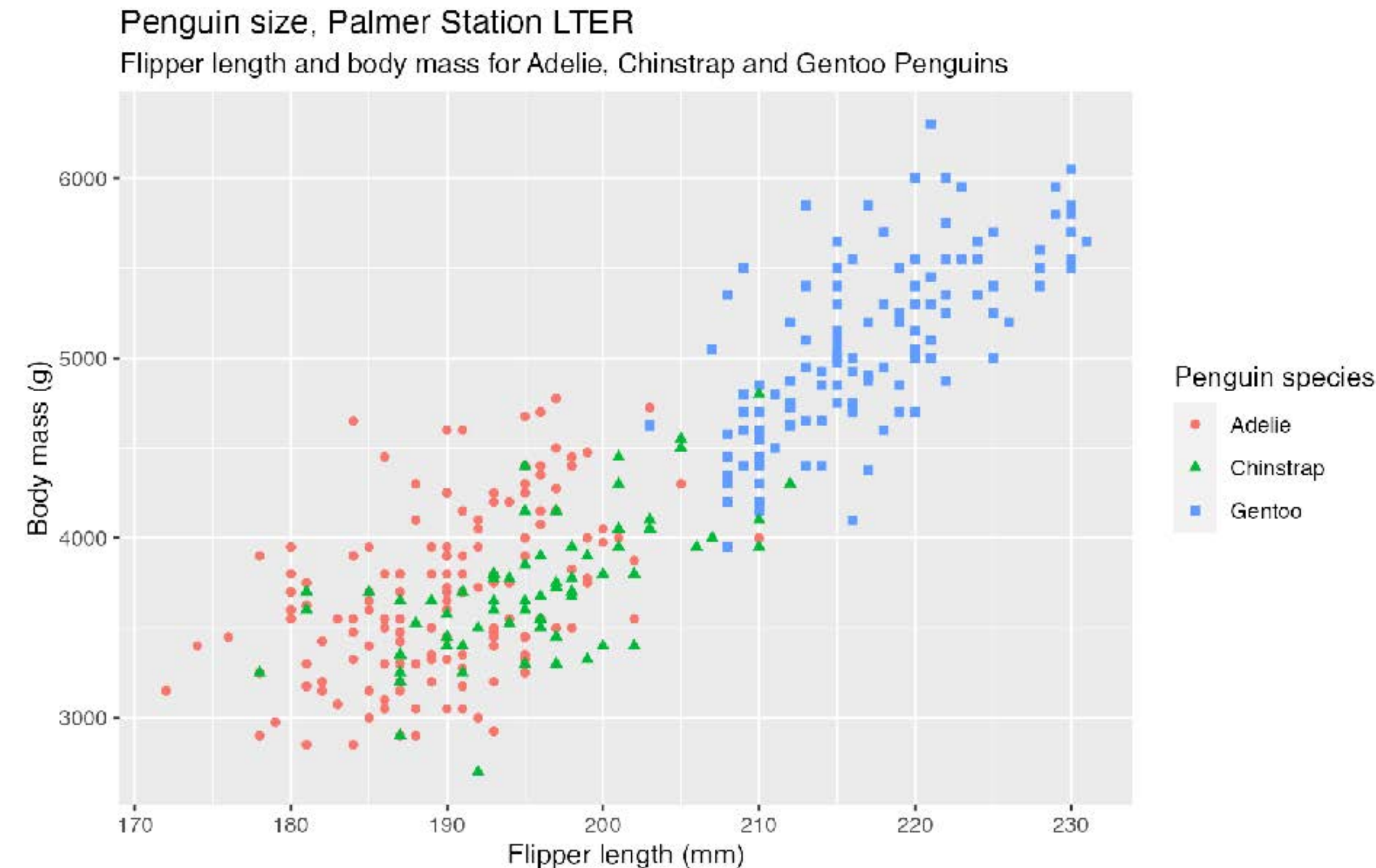
```
find("keys") %>%  
  start_car() %>%  
  drive(to = "work") %>%  
  park()
```

ex: ggplot2

visually pleasing defaults!

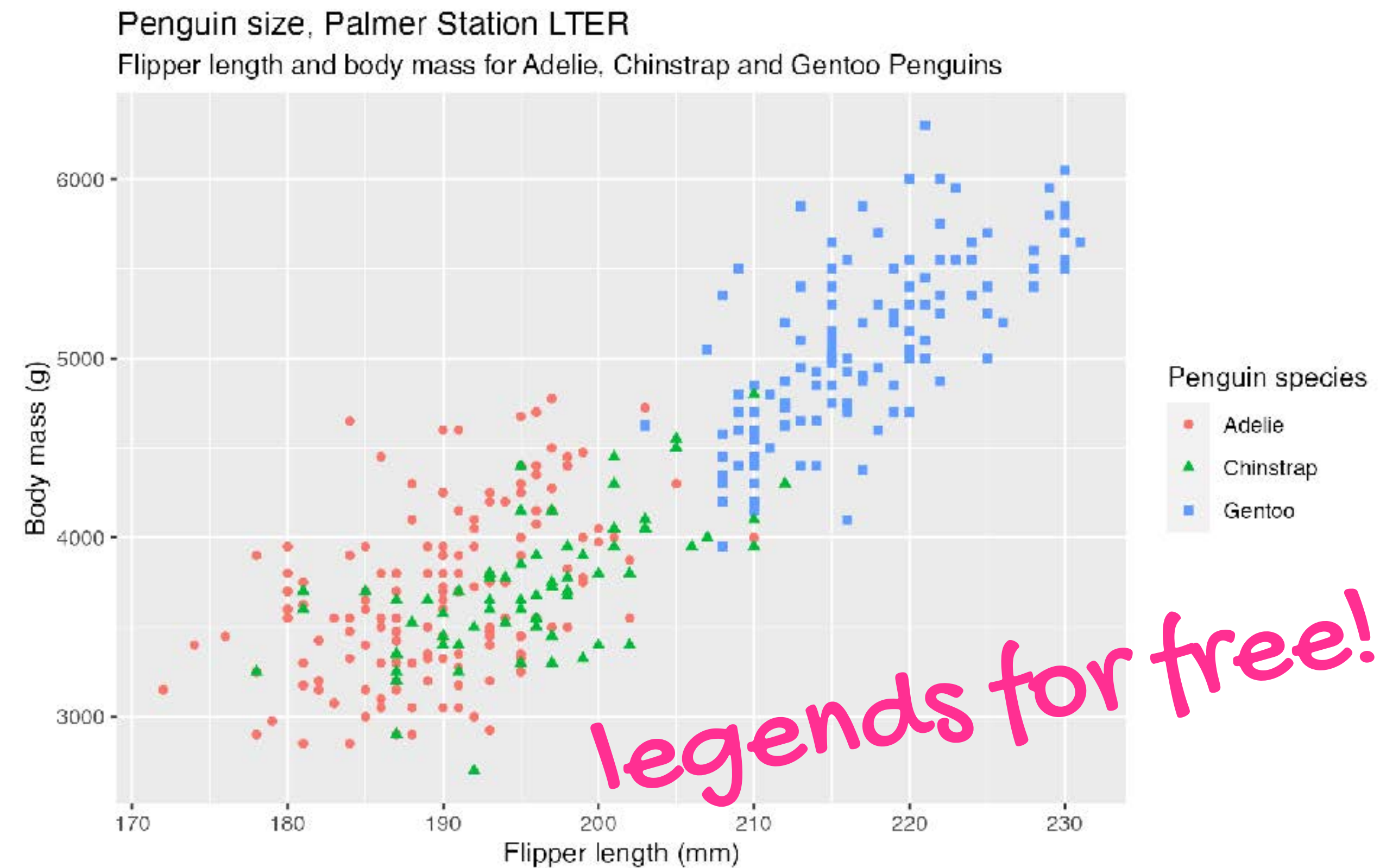
```
library(palmerpenguins)
library(tidyverse)
```

```
ggplot(data = penguins, aes(x = flipper_length_mm, y = body_mass_g)) +
  geom_point(aes(color = species, shape = species)) +
  labs(
    title = "Penguin size, Palmer Station LTER",
    subtitle = "Flipper length and body mass
for Adelie, Chinstrap and Gentoo
Penguins",
    x = "Flipper length (mm)",
    y = "Body mass (g)",
    color = "Penguin species",
    shape = "Penguin species"
  )
```



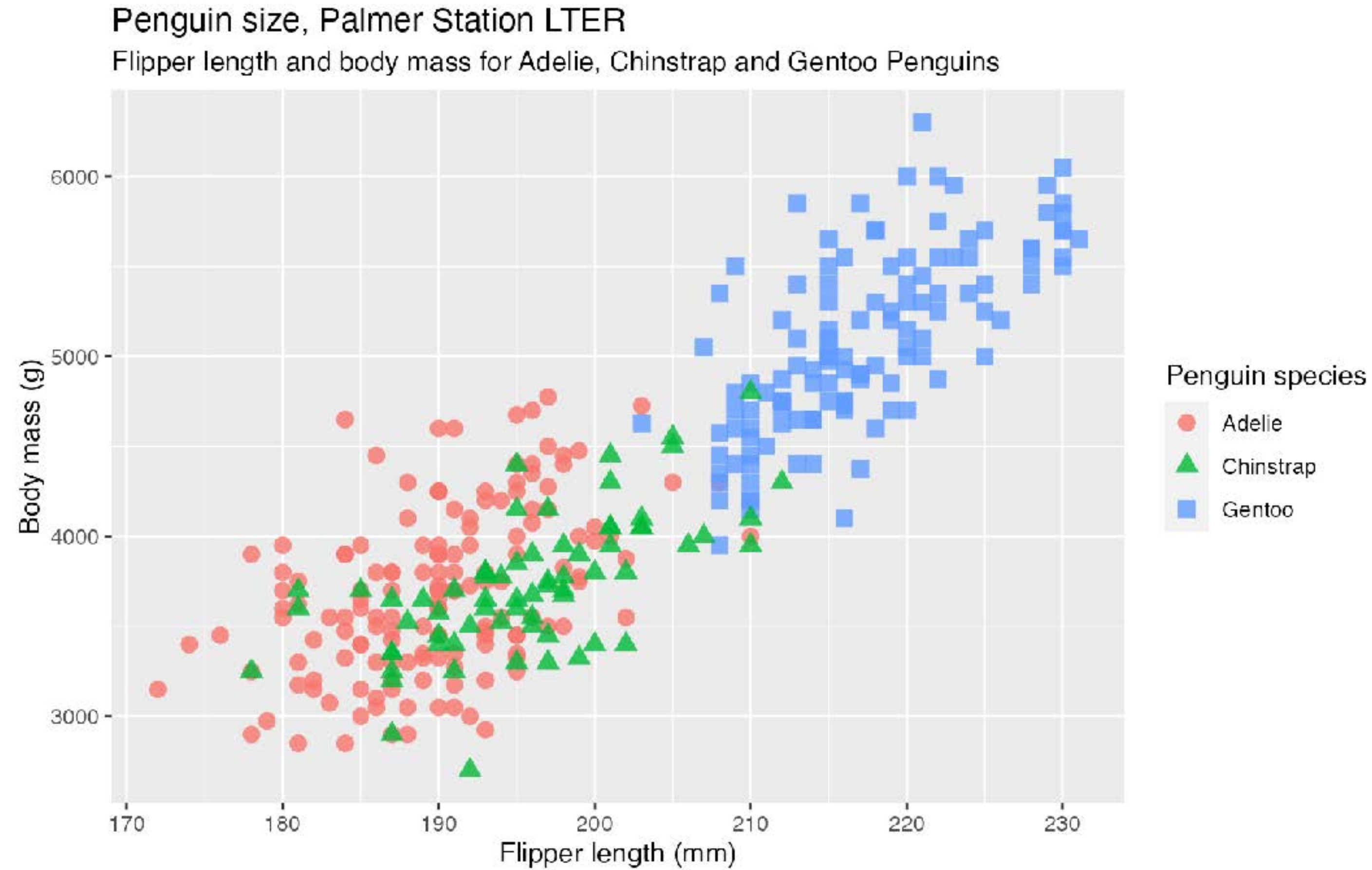
```
library(palmerpenguins)
library(tidyverse)
```

```
ggplot(data = penguins, aes(x = flipper_length_mm, y = body_mass_g)) +
  geom_point(aes(color = species, shape = species)) +
  labs(
    title = "Penguin size, Palmer Station LTER",
    subtitle = "Flipper length and body mass
for Adelie, Chinstrap and Gentoo
Penguins",
    x = "Flipper length (mm)",
    y = "Body mass (g)",
    color = "Penguin species",
    shape = "Penguin species"
  )
```

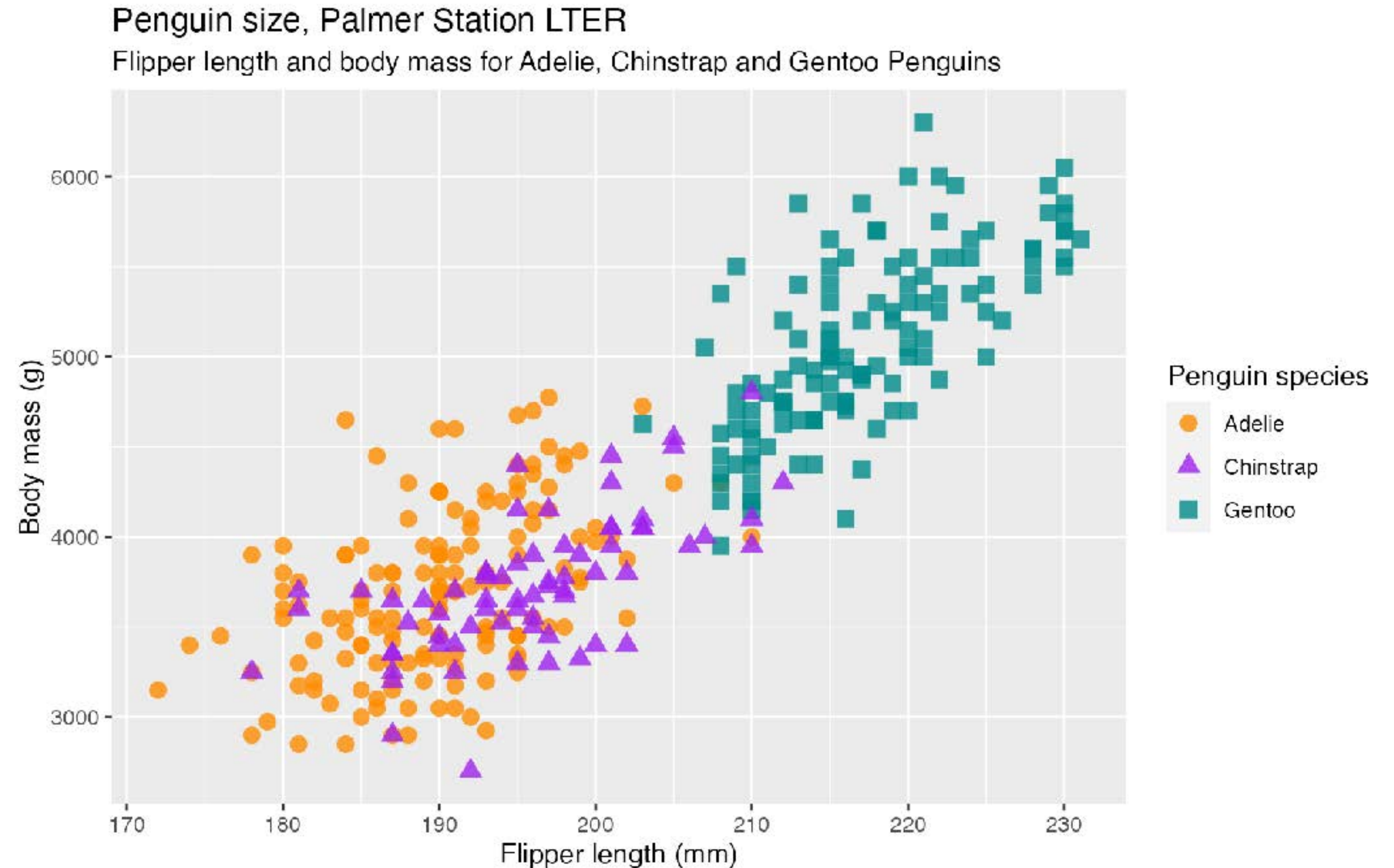


customize to your heart's desire!

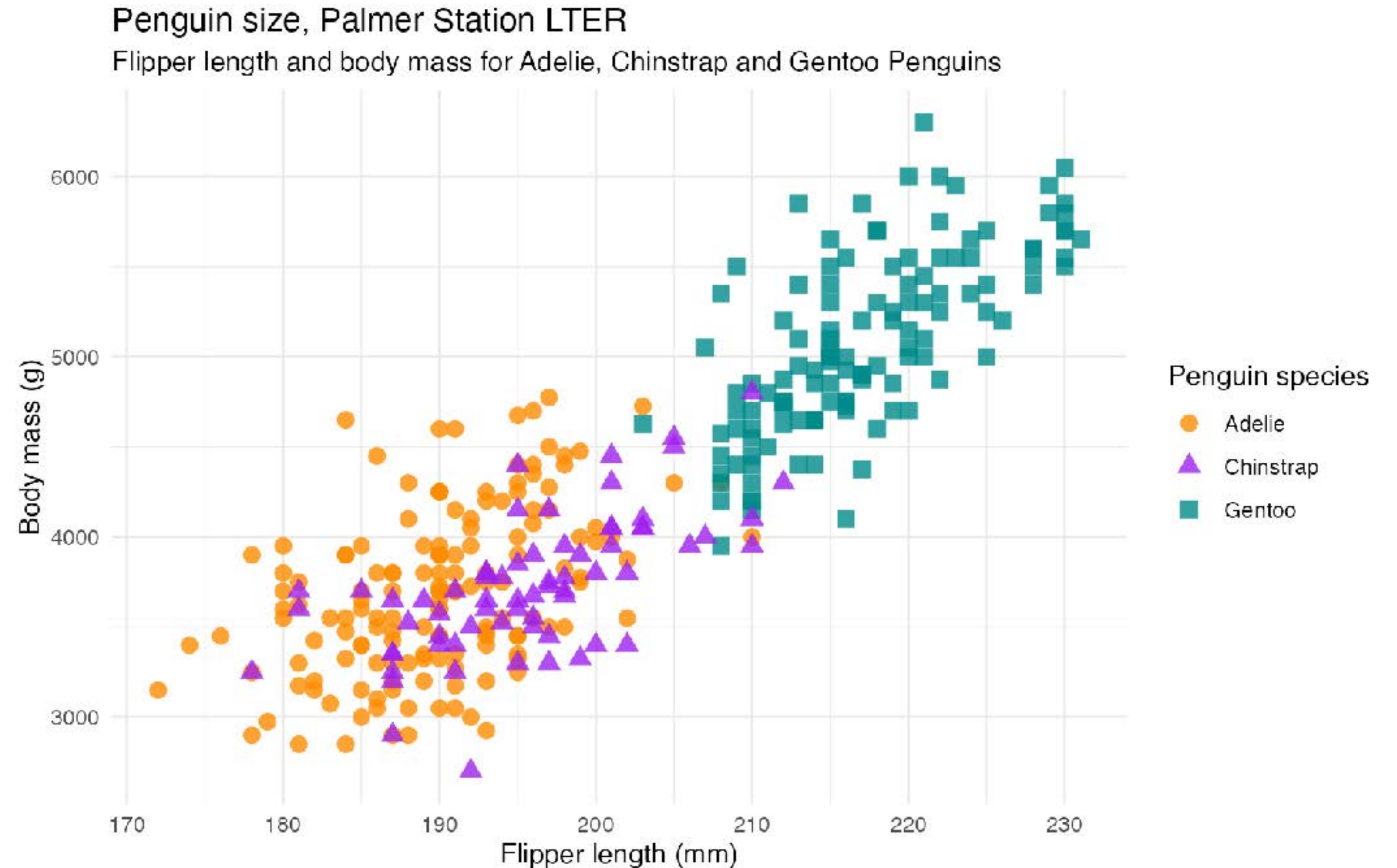
```
ggplot(data = penguins, aes(x = flipper_length_mm, y = body_mass_g)) +
  geom_point(aes(color = species, shape = species), size = 3, alpha = 0.8) +
  labs(
    title = "Penguin size, Palmer Station LTER",
    subtitle = "Flipper length and body mass
for Adelie, Chinstrap and Gentoo
Penguins",
    x = "Flipper length (mm)",
    y = "Body mass (g)",
    color = "Penguin species",
    shape = "Penguin species"
  )
```



```
ggplot(data = penguins, aes(x = flipper_length_mm, y = body_mass_g)) +
  geom_point(aes(color = species, shape = species), size = 3, alpha = 0.8) +
  labs(
    title = "Penguin size, Palmer Station LTER",
    subtitle = "Flipper length and body mass
for Adelie, Chinstrap and Gentoo
Penguins",
    x = "Flipper length (mm)",
    y = "Body mass (g)",
    color = "Penguin species",
    shape = "Penguin species"
  ) +
  scale_color_manual(
    values = c("darkorange",
              "purple",
              "cyan4"))
```



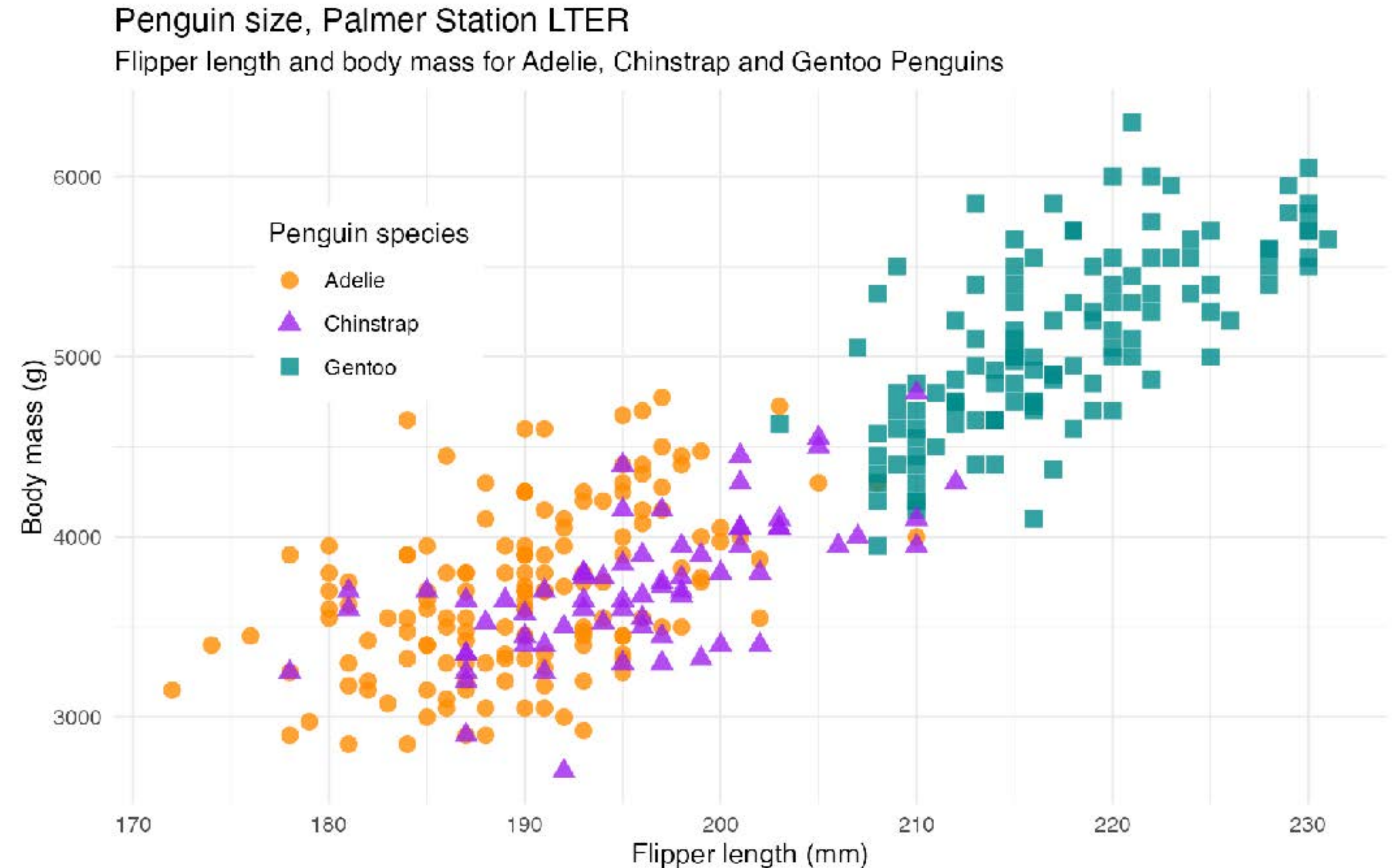
```
ggplot(data = penguins, aes(x = flipper_length_mm, y = body_mass_g)) +
  geom_point(aes(color = species, shape = species), size = 3, alpha = 0.8) +
  labs(
    title = "Penguin size, Palmer Station LTER",
    subtitle = "Flipper length and body mass
for Adelie, Chinstrap and Gentoo
Penguins",
    x = "Flipper length (mm)",
    y = "Body mass (g)",
    color = "Penguin species",
    shape = "Penguin species"
  ) +
  scale_color_manual(
    values = c("darkorange",
              "purple",
              "cyan4")) +
  theme_minimal()
```



```

ggplot(data = penguins, aes(x = flipper_length_mm, y = body_mass_g)) +
  geom_point(aes(color = species, shape = species), size = 3, alpha = 0.8) +
  labs(
    title = "Penguin size, Palmer Station LTER",
    subtitle = "Flipper length and body mass
for Adelie, Chinstrap and Gentoo
Penguins",
    x = "Flipper length (mm)",
    y = "Body mass (g)",
    color = "Penguin species",
    shape = "Penguin species"
  ) +
  scale_color_manual(
    values = c("darkorange",
              "purple",
              "cyan4")) +
  theme_minimal() +
  theme(
    legend.position = c(0.2, 0.7),
    legend.background =
      element_rect(
        fill = "white",
        color = NA
      )
  )

```



ex: tidyr

```
experiment_data
#> # A tibble: 6 x 5
#>   patient group    bp_1    bp_2    bp_3
#>   <dbl> <chr>    <chr>    <chr>    <chr>
#> 1     1 treatment 120/80  135/93  125/90
#> 2     2 control  172/105 171/82  161/117
#> 3     3 treatment 140/89  133/92  121/86
#> 4     4 control  151/92  112/109 150/83
#> 5     5 treatment 175/93  173/90  120/118
#> 6     6 control  180/85  173/94  174/106
```



```
#> # A tibble: 18 x 5
#>   patient group measurement systolic diastolic
#>   <dbl> <chr>    <chr>          <int>    <int>
#> 1     1 treatment 1             120      80
#> 2     1 treatment 2             135      93
#> 3     1 treatment 3             125      90
#> 4     2 control  1             172     105
#> 5     2 control  2             171      82
#> 6     2 control  3             161     117
#> # ... with 12 more rows
```

```
experiment_data %>%
```

```
  pivot_longer(  
    cols = contains("bp"),  
    names_to = "measurement",  
    names_prefix = "bp_",  
    values_to = "value"  
  )
```

```
#> # A tibble: 18 x 4
```

```
#>   patient group measurement value  
#>   <dbl> <chr>    <chr>    <chr>  
#> 1      1  treatment 1      120/80  
#> 2      1  treatment 2      135/93  
#> 3      1  treatment 3      125/90  
#> 4      2  control   1      172/105  
#> 5      2  control   2      171/82  
#> 6      2  control   3      161/117
```

```
#> # ... with 12 more rows
```

```
experiment_data
```

```
#> # A tibble: 6 x 5
```

```
#>   patient group bp_1 bp_2 bp_3  
#>   <dbl> <chr>    <chr> <chr> <chr>  
#> 1      1 treatment 120/80 135/93 125/90  
#> 2      2 control  172/105 171/82 161/117  
#> 3      3 treatment 140/89 133/92 121/86  
#> 4      4 control  151/92 112/109 150/83  
#> 5      5 treatment 175/93 173/90 120/118  
#> 6      6 control  180/85 173/94 174/106
```

```

experiment_data %>%
  pivot_longer(
    cols = contains("bp"),
    names_to = "measurement",
    names_prefix = "bp_",
    values_to = "value"
  ) %>%

```

```

  separate(value, into = c("systolic", "diastolic"), convert = TRUE)

```

```

#> # A tibble: 18 x 5
#>   patient group      measurement systolic diastolic
#>   <dbl> <chr>      <chr>          <int>    <int>
#> 1     1     1 treatment      1      120     80
#> 2     2     1 treatment      2      135     93
#> 3     3     1 treatment      3      125     90
#> 4     4     2 control        1      172    105
#> 5     5     2 control        2      171     82
#> 6     6     2 control        3      161    117
#> # ... with 12 more rows

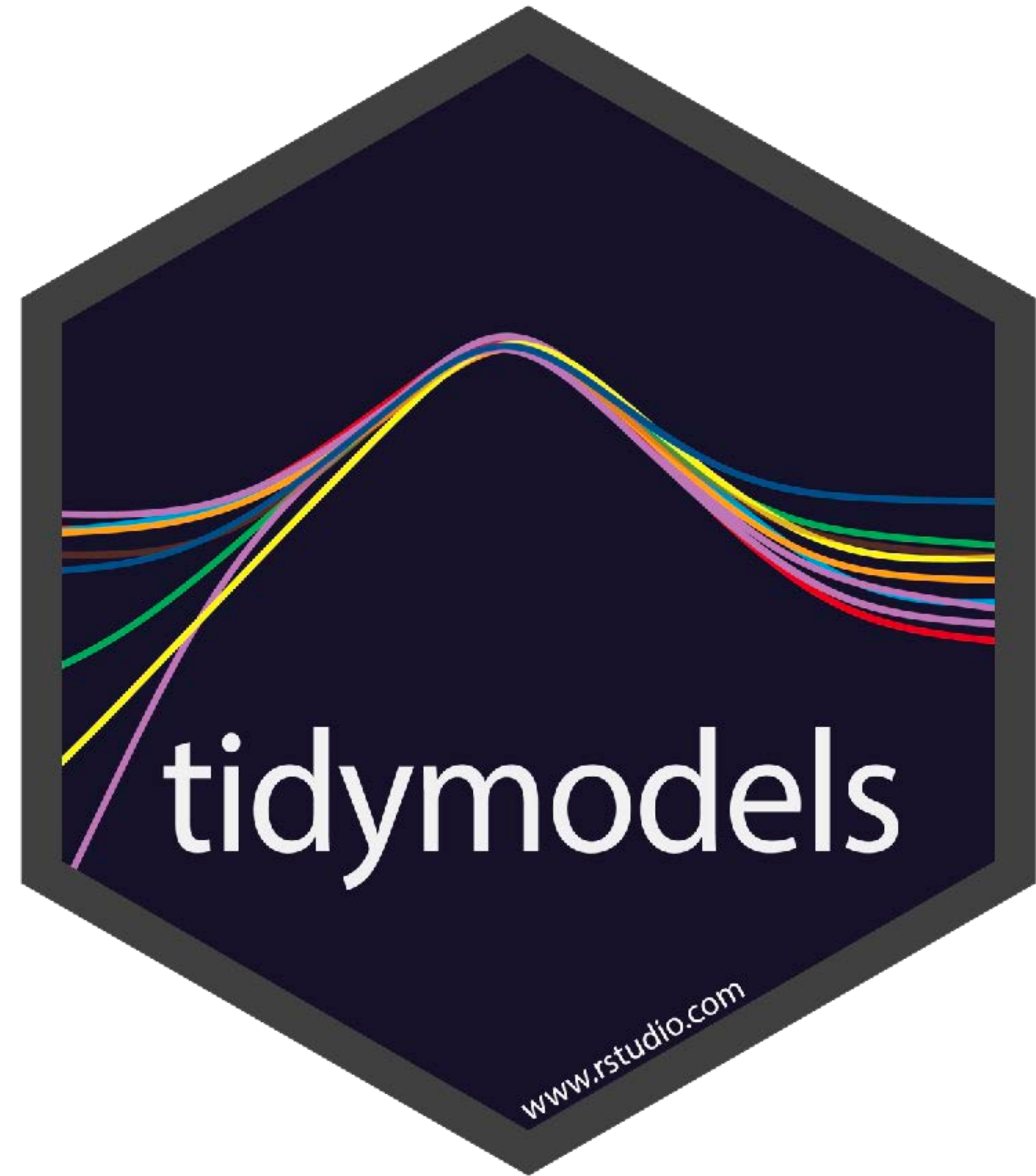
```

```

#> # A tibble: 18 x 4
#>   patient group      measurement value
#>   <dbl> <chr>      <chr>          <chr>
#> 1     1     1 treatment      1      120/80
#> 2     2     1 treatment      2      135/93
#> 3     3     1 treatment      3      125/90
#> 4     4     2 control        1      172/105
#> 5     5     2 control        2      171/82
#> 6     6     2 control        3      161/117
#> # ... with 12 more rows

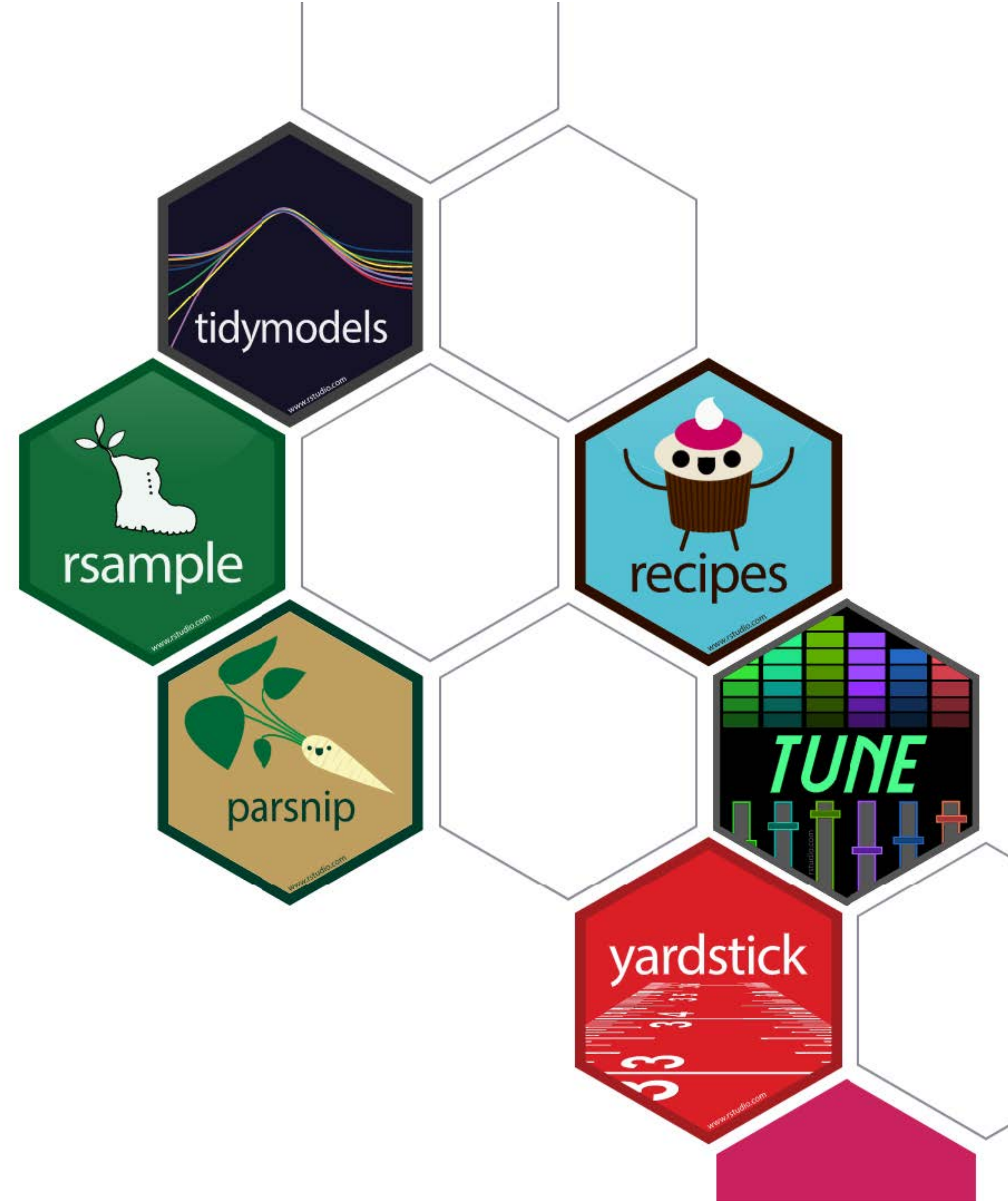
```

**modeling and
machine learning
with **tidymodels****



tidymodels

- collection of packages for modeling and machine learning using tidyverse principles
- **parsnip**: unified interface to models that can be used to try a range of models without getting bogged down in the syntactical minutiae of the underlying packages
- **recipes**: tidy interface to data pre-processing tools for feature engineering
- **rsample**: efficient resampling for estimation and model evaluation
- “many models” in a single data frame to avoid environment clutter and easy access with helper functions



A scenic landscape featuring a calm lake in the foreground, a dense forest of green and yellow trees on the opposite shore, and a vibrant rainbow arching across the sky. A large, dark evergreen tree is visible on the right side of the frame, partially obscuring the view. The text "a vast tidy ecosystem" is overlaid in the center-left area.

**a vast tidy
ecosystem**

work with data pipelines

work with ggplot2 layers



data cleaning



pretty (complex)
tables for PDF
output



laying out
multiple plots



highlighting
data in ggplots

*these are just some
of my favourite packages!*

**share and
communicate with
rmarkdown**



rmarkdown

- ◆ create computational documents that knit together text, code, results, and figures into polished outputs that are easy to read and share
- ◆ reproducible *by default*
- ◆ **bookdown**: and make them into books...
- ◆ **xaringan**: and make them into slides...
- ◆ **blogdown** / **distill**: and make them into websites...
- ◆ **rticles**: and make them into manuscripts...
- ◆ ...

2021-modern-stat-toolbox-niss - RStudio

reproducible-penguins.Rmd

Normal

B I U </>

Format Insert Table

```
---
title: Reproducible penguins

# to produce blinded version set to 1
blinded: 1

authors:
- name: Mine \c{etinkaya-Rundel} ^[mcetinka@ed.ac.uk]
  affiliation: |
    | School of Mathematics, University of Edinburgh, RStudio,
    | and Department of Statistical Science, Duke University
  email: mcetinka@ed.ac.uk
keywords: |
  R language, penguins, statistics, data science, tidyverse

abstract: >
  This is the awesome abstract of my awesome paper about Palmer Penguins.

bibliography: references.bib
output:
  rticles::asa_article:
    keep_tex: yes
---
```

```
{r include = FALSE}
library(tidyverse)
library(tidymodels)
library(palmerpenguins)
library(kableExtra)
library(knitr)
knitr::opts_chunk$set(echo = TRUE, out.width = "80%", warning = FALSE)
options(knitr.kable.NA = "")
```

Introduction

Data analysis

Introduction

The data come from

Data analysis

Figure below is a scatterplot.

Environment History Connections Tutorial

2021-modern-stat-toolbox-niss

Console Terminal R Markdown Jobs

R 4.0.4 · ~/Desktop/Talks/2021-modern-stat-toolbox-niss/

R version 4.0.4 (2021-02-15) -- "Lost Library Book"
Copyright (C) 2021 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin17.0 (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.

Natural language support but running in an English locale

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.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> |

Files Plots Packages Help Viewer

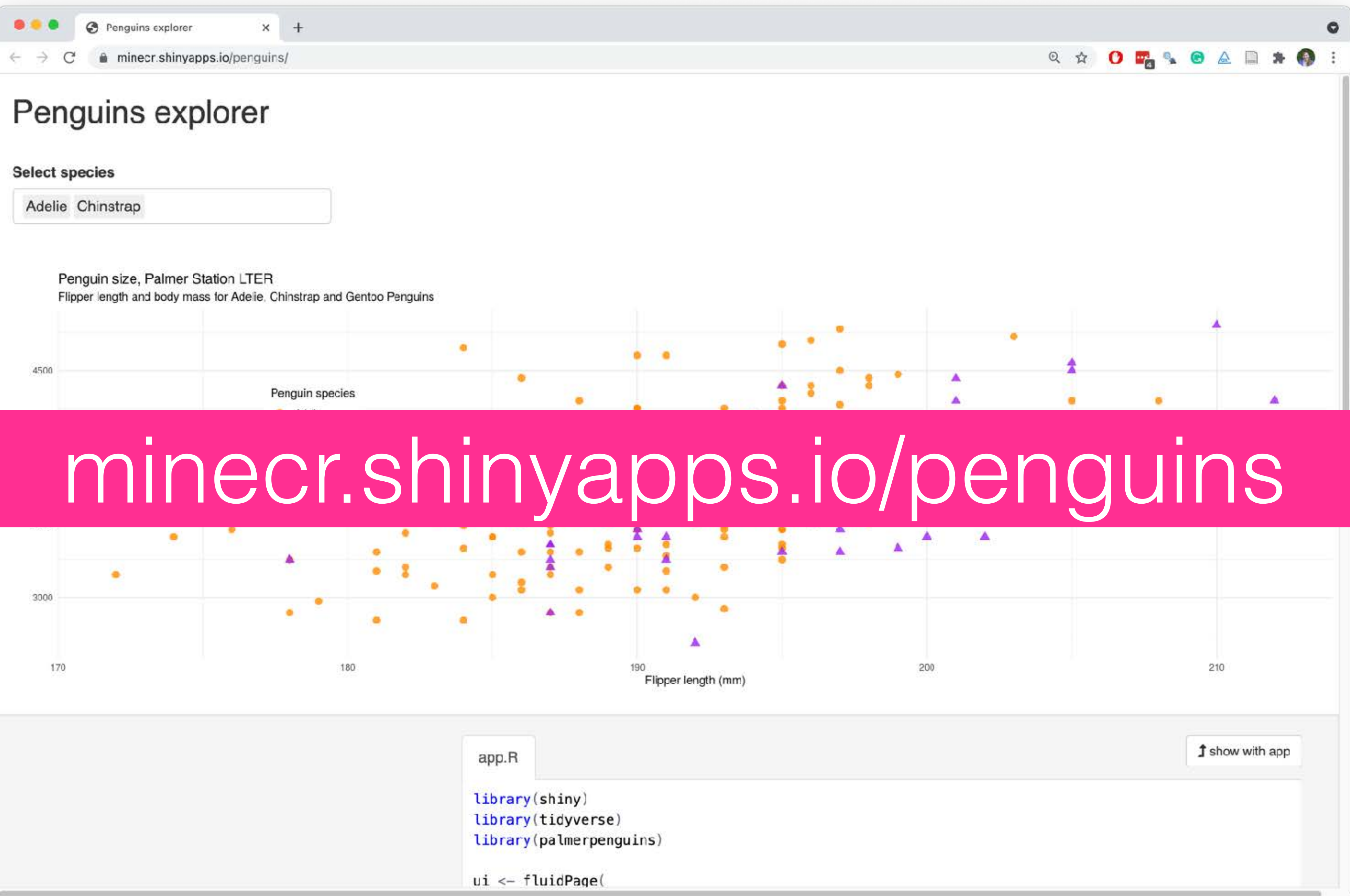
New Folder Delete Rename More

Home Desktop Talks 2021-modern-stat-toolbox-niss paper

Name	Size	Modified
..		
images		
references.bib	514 B	Mar 30, 2021, 2:44 PM
reproducible-penguins_files		
reproducible-penguins.pdf	242.3 KB	Mar 30, 2021, 3:20 PM
reproducible-penguins.Rmd	876 B	Mar 30, 2021, 3:21 PM
reproducible-penguins.tex	6.6 KB	Mar 30, 2021, 3:20 PM
reproducible-penguins.log	27.5 KB	Mar 30, 2021, 3:20 PM

interact with
shiny







calcat.covid19.ca.gov/cacovidmodels

**version control and
collaborate
with **git** and **github****



Git

THIS IS GIT. IT TRACKS COLLABORATIVE WORK
ON PROJECTS THROUGH A BEAUTIFUL
DISTRIBUTED GRAPH THEORY TREE MODEL.

COOL. HOW DO WE USE IT?

NO IDEA. JUST MEMORIZE THESE SHELL
COMMANDS AND TYPE THEM TO SYNC UP.
IF YOU GET ERRORS, SAVE YOUR WORK
ELSEWHERE, DELETE THE PROJECT,
AND DOWNLOAD A FRESH COPY.



ims - master - RStudio

00-preface.Rmd

1

Preface {.unnumbered}

2

3

We hope readers will take away three ideas from this book in addition to forming a foundation of statistical thinking and methods.

4

5

1. Statistics is an applied field with a wide range of practical applications.

6

2. You don't have to be a math guru to learn from interesting, real data.

7

3. Data are messy, and statistical tools are imperfect. However, when you understand the strengths and weaknesses of these tools, you can use them to learn interesting things about the world.

8

9

Textbook overview {.unnumbered}

10

11

1. **Introduction to data.** Data structures, variables, summaries, graphics, and basic data collection techniques.

12

2. **Exploratory data analysis.** Descriptive statistics, probability, and simulation and randomization techniques.

13

3. **Correlation and regression.** Descriptive statistics, probability, and simulation and randomization techniques.

14

4. **Multiple regression.** Descriptive statistics, probability, and simulation and randomization techniques.

15

5. **Foundations for inference.** Descriptive statistics, probability, and simulation and randomization techniques.

16

6. **Inference for categorical data.** Descriptive statistics, probability, and simulation and randomization techniques.

4:1 Preface

Console

Terminal

Markers

Jobs

R 4.0.4 · ~/Desktop/Open Intro/ims/

R version 4.0.4 (2021-02-15) -- Lost Library Book

Copyright (C) 2021 The R Foundation for Statistical Computing

Changes

History

master

Stage

Revert

Ignore

Staged

Status

Path

00-preface.Rmd

Commit message

10 characters

Minor edit

Amend previous commit

Commit

Show

Staged

Unstaged

Context

5 lines

Ignore Whitespace

Stage All

Discard All

Stage chunk

Discard chunk

@@ -1,11 +1,11 @@

1 1 # Preface {.unnumbered}

2 2

3 3 We hope readers will take away three ideas from this book in addition to forming a foundation of statistical thinking and methods.

4 4

5 5 1. Statistics is an applied field with a wide range of practical applications.

6 2. You don't have to be a math guru to learn from interesting, real data.

6 2. You don't have to be a math wiz to learn from interesting, real data.

7 7 3. Data are messy, and statistical tools are imperfect. However, when you understand the strengths and weaknesses of these tools, you can use them to learn interesting things about the world.

8 8

9 9 #### Textbook overview {.unnumbered}

10 10

11 11 1. **Introduction to data.** Data structures, variables, summaries, graphics, and basic data collection techniques.

Environment

History

Connections

Build

Git

Tutorial

Diff

Commit

Pull

Push

History

Staged

Status

Path

00-preface.Rmd

01-data-hello.Rmd

31.3 KB

Mar 26, 2021, 1:39 AM

02-data-design.Rmd

36.5 KB

Mar 18, 2021, 4:50 PM

03-data-applications.Rmd

14.8 KB

Mar 26, 2021, 1:39 AM

04-explore-categorical.Rmd

31.3 KB

Mar 26, 2021, 1:39 AM

GitHub

- ◆ web hosting for projects version controlled with Git
- ◆ collaboration and project management
- ◆ discoverability and publishing (with ghpages)
- ◆ where the technical side of the R community lives:
 - ◆ look for code samples
 - ◆ make feature requests
 - ◆ contribute to packages

mine-cetinkaya-rundel/covid19-r

github.com/mine-cetinkaya-rundel/covid19-r

Search or jump to...

Pull requestsIssuesMarketplaceExplore

mine-cetinkaya-rundel / covid19-r

Unwatch14

Star148

Fork30

<> Code

Issues

Pull requests

Actions

Projects

Wiki

Security

Insights

Settings

master1 branch0 tags

Go to fileAdd fileCode

mine-cetinkaya-rundel Update README.md

36c0be5 on 13 Jul 202075 commits

README.mdUpdate README.md9 months ago

_config.ymlSet theme jekyll-theme-minimal9 months ago

README.md

Contributor Covenant v2.0 adopted

covid19-r

This repo links to a collection of analyses on and representations of COVID19 data in R. Inclusion on the list does not mean the analysis was verified by me or that I endorse the findings. It should also be noted that some of the analysis / data might be out of date, so these should not be viewed, by default, as current findings.

I'm collecting them as they get posted online so that I can come back to them later to develop educational examples based on theses data. I should also note that one should think about using these as class examples thoroughly -- the situation is affecting all of us (and our students) greatly and it's possible some students are personally affected by the pandemic.

Subheadings are in flux as of now, as more analyses get added I might recategorize them.

If you have examples to add, please open an issue or make a pull request.

Blog posts

- Simulating COVID-19 interventions with R

About

Collection of analyses, packages, visualisations of COVID-19 data in R

mine-cetinkaya-rundel.github.io/co...

Readme

Releases

No releases published

Create a new release

Packages

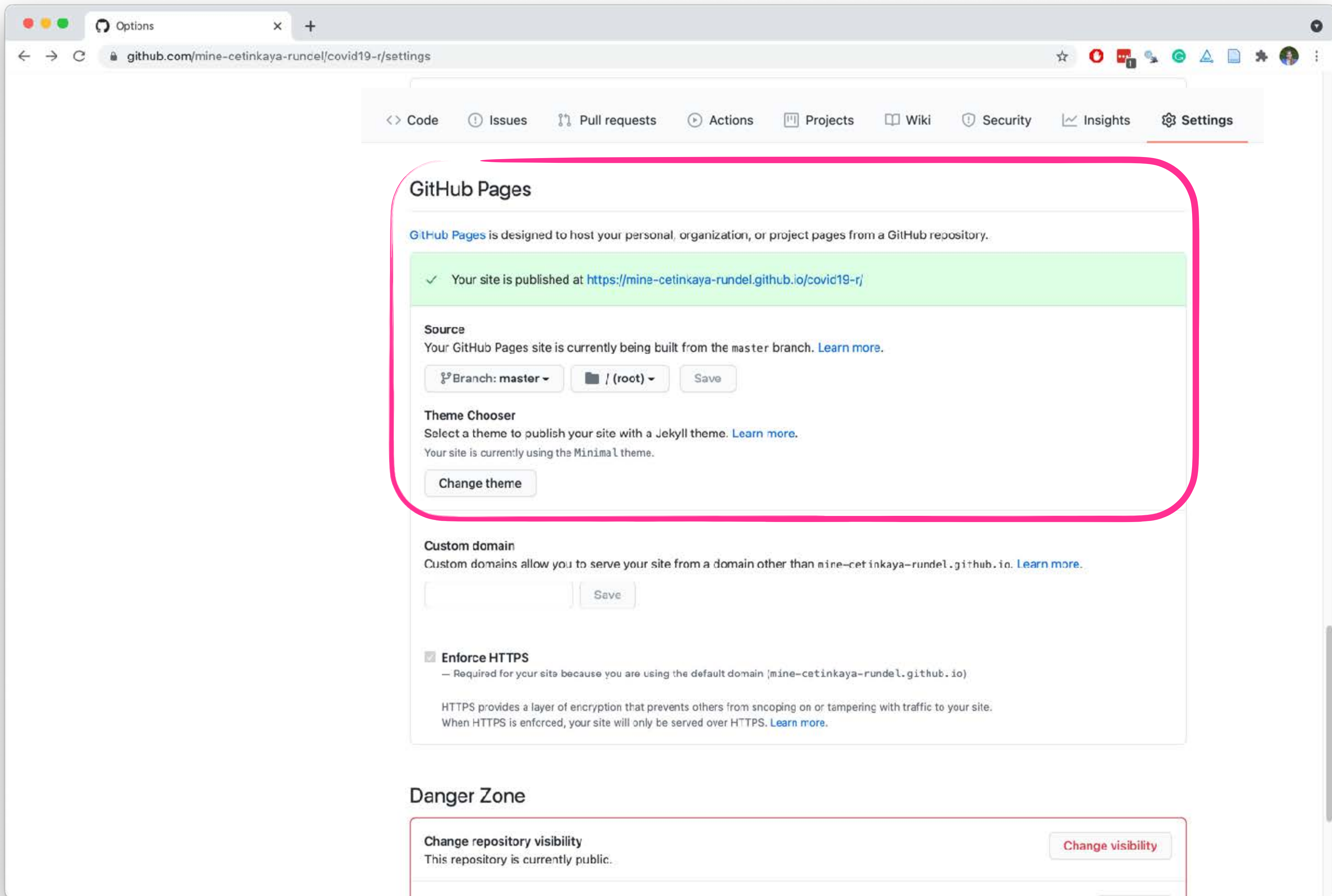
No packages published

Publish your first package

Contributors10

Environments1

github-pagesActive



covid19-r | Collection of analyses

mine-cetinkaya-rundel.github.io/covid19-r/

covid19-r

Collection of analyses, packages, visualisations of COVID-19 data in R

[View the Project on GitHub](#)
mine-cetinkaya-rundel/covid19-r

Contributor Covenant

v2.0 adopted

covid19-r

This repo links to a collection of analyses on and representations of COVID19 data in R. Inclusion on the list does not mean the analysis was verified by me or that I endorse the findings. It should also be noted that some of the analysis / data might be out of date, so these should not be viewed, by default, as current findings.

I'm collecting them as they get posted online so that I can come back to them later to develop educational examples based on these data. I should also note that one should think about using these as class examples thoroughly - the situation is affecting all of us (and our students) greatly and it's possible some students are personally affected by the pandemic.

Subheadings are in flux as of now, as more analyses get added I might recategorize them.

If you have examples to add, please open an issue or make a pull request.

Blog posts

- [Simulating COVID-19 interventions with R](#)
- [Analysing COVID-19 \(2019-nCoV\) outbreak data with R - part 1](#)
- [Analysing COVID-19 \(2019-nCoV\) outbreak data with R - part 2](#)
- [Modelling the effects of public health interventions on COVID-19 transmission using R - part 1](#)
- [Modelling the effects of public health interventions on COVID-19 transmission using R - part 2](#)
- [COVID-19: The Case of Germany](#)
- [Map Visualization of COVID-19 Across the World with R](#)
- [COVID-19 cumulative observed case fatality rate over time](#)
- [Impact of a country's age breakdown on COVID-19 case fatality rate](#)
- [COVID-19 Tracker: Days since N](#)
- [COVID-19 Has Become a Partisan Issue \(How to Make Dumbbell Plots in R\)](#)
- [Get Your Epidemiology from Epidemiologists](#)
- [Flatten the COVID-19 curve](#)
- [A Beginner's Guide to Creating a Corona "Dashboard"](#)
- [Infectious diseases and nonlinear differential equations](#)
- [COVID-19: The Case of Spain](#)
- [Epidemic modelling of COVID-19 in the UK using an SIR model](#)

This project is maintained by [mine-cetinkaya-rundel](#)

Hosted on: [GitHub Pages](#) — Theme by [orderedlist](#)

**stay current and
connected with
#rstats community**



ask (good) questions

- ◆ make **re**producible **e**xamples
- ◆ make them as minimal as you can
- ◆ If asking publicly (RStudio Community, Stack Overflow, etc.) try to use data available in a package
- ◆ let reprex take care of checking for reproducibility and formatting for you!



RStudio

Go to file/function Addins Project: (None)

Untitled1 x

Source on Save Run Source

```
1 library(tidyverse)
2 library(palmerpenguins)
3
4 penguins <- penguins %>%
5   mutate(body_mass_kg = body_mass_g / 1000)
6
7 penguins %>%
8   group_by(species) %>%
9   summarise(mean = mean(body_mass_kg))
10
```

Console Terminal Jobs

R 4.0.4 ~/

```
> library(reprex)
> penguins %>%
+   group_by(species) %>%
+   summarise(mean = mean(body_mass_kg))
# A tibble: 3 x 2
  species    mean
  <fct>    <dbl>
1 Adelie    NA
2 Chinstrap 3.73
3 Gentoo    NA
>
```

Files Plots Packages Help Viewer

10:3 (Top Level) R Script

Environment History Connections Tutorial

Import Dataset 198 MiB

R Global Environment

Data

penguins	344 obs. of 9 variables
----------	-------------------------

+ Create a new Topic

NAs when calculating means

Uncategorized

tags key packages, functions, & topics

🗨️ **B** *I* 🔗 ⌞ ⌵ ⬆️ ☰ ☷ 😊 📅 ⚙️

I know that mass is recorded for some Adelie and Gentoo penguins. Why am I getting NAs for my means?

```
library(tidyverse)
library(palmerpenguins)

penguins <- penguins %>%
  mutate(body_mass_kg = body_mass_g / 1000)
```

```
penguins %>%
  group_by(species) %>%
  summarise(mean = mean(body_mass_kg))
#> # A tibble: 3 x 2
#>   species mean
#>   <fct>   <dbl>
#> 1 Adelie NA
#> 2 Chinstrap 3.73
#> 3 Gentoo NA
````
```

<sup><sup></sup>Created on 2021-03-30 by the [reprex package](https://reprex.tidyverse.org) (v1.0.0)</sup>



I know that mass is recorded for some Adelie and Gentoo penguins. Why am I getting NAs for my means?

```
library(tidyverse)
library(palmerpenguins)

penguins <- penguins %>%
 mutate(body_mass_kg = body_mass_g / 1000)

penguins %>%
 group_by(species) %>%
 summarise(mean = mean(body_mass_kg))

#> # A tibble: 3 x 2
#> species mean
#> <fct> <dbl>
#> 1 Adelie NA
#> 2 Chinstrap 3.73
#> 3 Gentoo NA
```

Created on 2021-03-30 by the [reprex package](#) (v1.0.0)

+ Create Topic

cancel

« hide preview

RStudio

Project: (None)

```
1 library(tidyverse)
2 library(palmerpenguins)
3
4 penguins <- penguins %>%
5 mutate(body_mass_kg = body_mass_g / 1000)
6
7 penguins %>%
8 group_by(species) %>%
9 summarise(mean = mean(body_mass_kg))
10
```

Console

```
R 4.0.4 ~/
group_by(species) %>%
+ summarise(mean = mean(body_mass_kg))
A tibble: 3 x 2
 species mean
 <fct> <dbl>
1 Adelie NA
2 Chinstrap 3.73
3 Gentoo NA
> reprex()
i Rendering reпреx...
✓ Rendered reпреx is on the clipboard.
> reprex()
i Rendering reпреx...
✓ Rendered reпреx is on the clipboard.
> reprex()
i Rendering reпреx...
✓ Rendered reпреx is on the clipboard.
>
```

Files | Plots | Packages | Help | Viewer

```
library(tidyverse)
library(palmerpenguins)

penguins <- penguins %>%
 mutate(body_mass_kg = body_mass_g / 1000)

penguins %>%
 group_by(species) %>%
 summarise(mean = mean(body_mass_kg))
#> # A tibble: 3 x 2
#> species mean
#> <fct> <dbl>
#> 1 Adelie NA
#> 2 Chinstrap 3.73
#> 3 Gentoo NA
```

9:37 (Top Level) R Script

Environment | History | Connections | Tutorial

205 MiB

R Global Environment

Data

|          |                         |
|----------|-------------------------|
| penguins | 344 obs. of 9 variables |
|----------|-------------------------|

Created on 2021-03-30 by the reпреx package (v1.0.0)

# community

- ◆ #rstats on Twitter
- ◆ R Weekly newsletter: [rweekly.org](http://rweekly.org)
- ◆ TidyTuesday: [github.com/rfordatascience/tidytuesday](https://github.com/rfordatascience/tidytuesday)
- ◆ RLadies: [rladies.org](http://rladies.org) + community Slack
- ◆ useR groups: [r-consortium.org/blog/2019/09/09/r-community-explorer-r-user-groups](http://r-consortium.org/blog/2019/09/09/r-community-explorer-r-user-groups)
- ◆ talk to each other (including your students!) about computing

# resources

## ◆ learn

- ◆ tidyverse: [tidyverse.org/learn](https://tidyverse.org/learn)
- ◆ tidymodels: [tidymodels.org/start](https://tidymodels.org/start)
- ◆ rmarkdown: [rmarkdown.rstudio.com/lesson-1.html](https://rmarkdown.rstudio.com/lesson-1.html)
- ◆ RStudio visual editor: [rstudio.github.io/visual-markdown-editing/#](https://rstudio.github.io/visual-markdown-editing/#)
- ◆ shiny: [shiny.rstudio.com/tutorial](https://shiny.rstudio.com/tutorial)
- ◆ Git and GitHub: [happygitwithr.com](https://happygitwithr.com)
- ◆ **teach:** [datasciencebox.org](https://datasciencebox.org)

# toolkit for the modern statistician



[bit.ly/modern-toolkit](https://bit.ly/modern-toolkit)

@minebocek

mine-cetinkaya-rundel

cetinkaya.mine@gmail.com

