HW 04 - Design + details

INFO 3312/5312 - Spring 2025

Your name

2025-03-06

Setup

Load packages and data:

```
library(tidyverse)
```

```
── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0
──
✔ dplyr     1.1.4     ✔ readr     2.1.5
✔ forcats   1.0.0     ✔ stringr   1.5.1
✔ ggplot2   3.5.1     ✔ tibble    3.2.1
✔ lubridate 1.9.4     ✔ tidyr     1.3.1
✔ purrr     1.0.4
── Conflicts ────────────────────────────────────────── tidyverse_conflicts()
──
✖ dplyr::filter() masks stats::filter()
✖ dplyr::lag()    masks stats::lag()
ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all
conflicts to become errors
```

```
library(scales)
```

```

Attaching package: 'scales'

The following object is masked from 'package:purrr':

    discard

The following object is masked from 'package:readr':

    col_factor
```

```
library(palmerpenguins)
library(rvest)
```

```

Attaching package: 'rvest'

The following object is masked from 'package:readr':

    guess_encoding
```

```
library(colorspace)
library(ggtext)
```

Exercises

Exercise 1

```
p <- ggplot(
  data = penguins,
  mapping = aes(
    x = flipper_length_mm,
    y = body_mass_g,
    color = species
  )
) +
  geom_point(
    mapping = aes(
      shape = species
    ),
    size = 3,
    alpha = 0.8
  ) +
  geom_smooth(method = "lm", se = FALSE) +
  scale_x_continuous(labels = label_number(scale_cut = cut_si(unit = "mm"))) +
  scale_y_continuous(labels = label_number(scale_cut = cut_si(unit = "g"))) +
  labs(
    title = "Penguin size, Palmer Station LTER",
    subtitle = "Flipper length and body mass for Adelie, Chinstrap, and Gentoo
Penguins",
    x = "Flipper length",
    y = "Body mass",
    color = "Penguin species",
    shape = "Penguin species"
  )
```

```
# add code here
```

Exercise 2

````
```{r}
#| label: ex2

# add code here
```
````

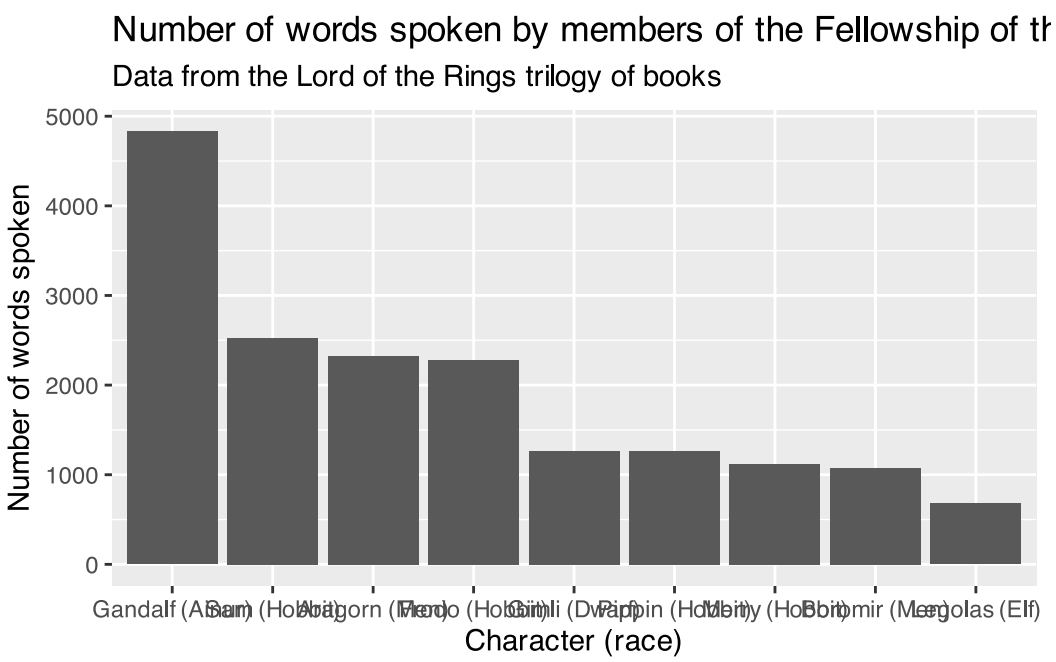
Exercise 3

```
# load LOTR data
# source: https://github.com/MokoSan/FSharpAdvent/blob/master/Data/WordsByCharacter.csv
lotr_words <- read_csv(file = "data/LOTRWordsByCharacter.csv")
```

```
Rows: 731 Columns: 5
── Column specification ────────────────────────────────────────────────────────
Delimiter: ","
chr (4): Film, Chapter, Character, Race
dbl (1): Words

ℹ Use `spec()` to retrieve the full column specification for this data.
ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
lotr_words |>
  summarize(
    n_words = sum(Words),
    .by = c(Character, Race)
  ) |>
  # filter and only keep members of the fellowship of the rings
    filter(Character %in% c("Frodo", "Sam", "Merry", "Pippin", "Gandalf", "Aragorn", "Legolas", "Gimli", "Boromir")) |>
  mutate(Character = str_glue("{Character} ({Race})") |>
           fct_reorder(.x = n_words, .desc = TRUE)) |>
  ggplot(mapping = aes(x = Character, y = n_words)) +
  geom_col() +
  labs(
    x = "Character (race)",
    y = "Number of words spoken",
    title = "Number of words spoken by members of the Fellowship of the Ring",
    subtitle = "Data from the Lord of the Rings trilogy of books"
  )
```



Exercise 4

```
# add code here
```