

Homework assignment 2

Mikayla Lalli

Script setup

Part A

1. Read in data file

```
data <- read_csv("data/lotr-trilogy.csv")
```

```
Rows: 3 Columns: 7
```

```
-- Column specification -----
```

```
Delimiter: ","
```

```
chr (1): movie
```

```
dbl (6): elf_female, elf_male, Hobbit_female, hobbit_Male, man_Female, Man_male
```

```
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.
```

2. List all the ways the dataset is not tidy.

1. Data is in wide format; each variable does not have its own column and each observation does not have its own row.
2. Some variable names are capitalized inconsistently (Hobbit_female, hobbit_Male, Man_female, Man_male)

3. How many columns and rows would you have if this dataset was tidy?

There would be four columns and 18 rows.

4. What would the column names be in tidy format?

The column names would be: movie, race, sex, and words_spoken

Part B

1. Tidy the dataset from A1.

```
data <- data |>
  rename(hobbit_female = "Hobbit_female",
         hobbit_male = "hobbit_Male",
         man_female = "man_Female",
         man_male = "Man_male")
```

```
data <- data |>
  pivot_longer(elf_female:man_male,
               names_to = c("race", "sex"),
               names_pattern = "(.*)_(.*)",
               values_to = "words_spoken")
```

2. What's the total number of words spoken by:

a) male hobbits

```
male_hobbit_tib <- data |>
  filter(race == "hobbit", sex == "male")

male_hobbit <- sum(male_hobbit_tib$words_spoken)
male_hobbit
```

```
[1] 8780
```

8780 words were spoken by male hobbits.

b) female elves

```
female_elf_tib <- data |>
  filter(race == "elf", sex == "female")

female_elf <- sum(female_elf_tib$words_spoken)
female_elf
```

```
[1] 1743
```

1743 words were spoken by female elves.

c) male elves?

```
male_elf_tib <- data |>
  filter(race == "elf", sex == "male")

male_elf <- sum(male_elf_tib$words_spoken)
male_elf
```

```
[1] 1994
```

1994 words were spoken by male elves.

3. Is the number of spoken words in a movie dominated by a single race?

```
movie_tib <- data |>
  group_by(movie, race) |>
  summarise(words_spoken = sum(words_spoken), .groups = "keep") |>
  ungroup() |>
  group_by(movie) |>
  arrange(desc(words_spoken), .by_group = TRUE)
```

Yes, but only in The Fellowship of the Ring and The Return of the King. In The Two Towers, the number of words spoken by the races, Hobbit and Man were very close.

4. Does the dominant race depend on the movie?

Yes. Man is dominant in The Return of the King (and technically has more words in The Two Towers but not by much), while Hobbit is dominant in The Fellowship of the Ring.