hw-2_notebook.qmd

Matin Borhani

```
#Q1.
lord_data<-readr::read_csv("lord-of-the-rings-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.
```

```
#Q2. The occasional and inconsistent use of capital letters. Also in tidy data we
#usually have species, sex, and the number of observation (count),
#in seperate columns instead of giving them their own columns and
#having them as seperate variables.
#This also leads to a single column to have more than one type of data.
#for example "elf-female" tells us about both elf and female and this is not tidy.
#Each Column should contain only one variable.
```

```
#Q3.I would have 3 columns (race, sex, count) and 18 rows
#(6 rows for each movie becuase we have 3 species/races that each consist of
#male and female)
```

```
#Q1A. Column names: "race", "sex", "count"
```

```
lord_data_tidy_s2 <- lord_data |>
  tidyr::pivot_longer(c('elf_female', 'elf_male', 'Hobbit_female', 'hobbit_Male',
                        'man_Female',
                'Man_male'),
```

```
    names_to = "race_sex",
    values_to = "count",
    values_drop_na = TRUE)

print(lord_data_tidy_s2)
```

```
# A tibble: 18 x 3
   movie                     race_sex      count
   <chr>                     <chr>         <dbl>
 1 The Fellowship of the Ring elf_female     1229
 2 The Fellowship of the Ring elf_male        971
 3 The Fellowship of the Ring Hobbit_female    14
 4 The Fellowship of the Ring hobbit_Male    3644
 5 The Fellowship of the Ring man_Female        0
 6 The Fellowship of the Ring Man_male       1995
 7 The Two Towers             elf_female      183
 8 The Two Towers             elf_male        510
 9 The Two Towers             Hobbit_female     2
10 The Two Towers             hobbit_Male    2673
11 The Two Towers             man_Female      268
12 The Two Towers             Man_male       2459
13 The Return of the King     elf_female      331
14 The Return of the King     elf_male        513
15 The Return of the King     Hobbit_female     0
16 The Return of the King     hobbit_Male    2463
17 The Return of the King     man_Female      401
18 The Return of the King     Man_male       3589
```

```
lord_data_tidy_s2 |>
  dplyr::count(race_sex)
```

```
# A tibble: 6 x 2
  race_sex          n
  <chr>         <int>
1 Hobbit_female     3
2 Man_male          3
3 elf_female        3
4 elf_male          3
5 hobbit_Male       3
6 man_Female        3
```

```
#Here I can see that each race_sex data point repeats 3 times in each of the 3 movies
#so in total I should have 18 rows total.
```

```
lord_data_tidy_s4 <- lord_data_tidy_s2 |>
   tidyr::separate(race_sex, c("race", "sex"), sep = "_")

lord_data_tidy_s4
```

```
# A tibble: 18 x 4
   movie                      race   sex    count
   <chr>                      <chr>  <chr>  <dbl>
 1 The Fellowship of the Ring elf    female  1229
 2 The Fellowship of the Ring elf    male     971
 3 The Fellowship of the Ring Hobbit female    14
 4 The Fellowship of the Ring hobbit Male    3644
 5 The Fellowship of the Ring man    Female     0
 6 The Fellowship of the Ring Man    male    1995
 7 The Two Towers             elf    female   183
 8 The Two Towers             elf    male     510
 9 The Two Towers             Hobbit female     2
10 The Two Towers             hobbit Male    2673
11 The Two Towers             man    Female   268
12 The Two Towers             Man    male    2459
13 The Return of the King     elf    female   331
14 The Return of the King     elf    male     513
15 The Return of the King     Hobbit female     0
16 The Return of the King     hobbit Male    2463
17 The Return of the King     man    Female   401
18 The Return of the King     Man    male    3589
```

```
#Q2B
  male_hobbits_words_spoken <-dplyr::filter(lord_data_tidy_s4, race == "hobbit"
                                            & sex == "male")
  male_hobbits_words_spoken
```

```
# A tibble: 0 x 4
# i 4 variables: movie <chr>, race <chr>, sex <chr>, count <dbl>
```

```
  dplyr::summarise(male_hobbits_words_spoken, sum_words = sum(count))
```

```
# A tibble: 1 x 1
  sum_words
      <dbl>
1         0
```

```
#Q2BCONTINUED

  female_elf_words_spoken <-dplyr::filter(lord_data_tidy_s4, race == "elf" & sex
                                          == "female")

  female_elf_words_spoken
```

```
# A tibble: 3 x 4
  movie                      race  sex    count
  <chr>                      <chr> <chr>  <dbl>
1 The Fellowship of the Ring elf   female  1229
2 The Two Towers             elf   female   183
3 The Return of the King     elf   female   331
```

```
  dplyr::summarise(female_elf_words_spoken, sum_words = sum(count))
```

```
# A tibble: 1 x 1
  sum_words
      <dbl>
1      1743
```

```
  male_elf_words_spoken <-dplyr::filter(lord_data_tidy_s4, race == "elf"
                                        & sex == "male")

  male_elf_words_spoken
```

```
# A tibble: 3 x 4
  movie                      race  sex    count
  <chr>                      <chr> <chr>  <dbl>
1 The Fellowship of the Ring elf   male     971
2 The Two Towers             elf   male     510
3 The Return of the King     elf   male     513
```

```
  dplyr::summarise(male_elf_words_spoken, sum_words = sum(count))
```

```
# A tibble: 1 x 1
  sum_words
      <dbl>
1      1994
```

```
library(stringr)

  lord_data_tidy_s4 <- dplyr::mutate(lord_data_tidy_s4,
    race = str_to_lower(race),  # making race column to lowercase
    sex = str_to_lower(sex)     # making sex column to lowercase
  )
```

```
#Q3B and Q4B

lord_data_nosex <-lord_data_tidy_s4
lord_data_nosex <- dplyr::select(lord_data_tidy_s4, movie, race, count)

lord_data_nosex
```

```
# A tibble: 18 x 3
   movie                      race   count
   <chr>                      <chr>  <dbl>
 1 The Fellowship of the Ring elf     1229
 2 The Fellowship of the Ring elf      971
 3 The Fellowship of the Ring hobbit    14
 4 The Fellowship of the Ring hobbit  3644
 5 The Fellowship of the Ring man        0
 6 The Fellowship of the Ring man     1995
 7 The Two Towers             elf      183
 8 The Two Towers             elf      510
 9 The Two Towers             hobbit     2
10 The Two Towers             hobbit  2673
11 The Two Towers             man      268
12 The Two Towers             man     2459
13 The Return of the King     elf      331
14 The Return of the King     elf      513
15 The Return of the King     hobbit     0
16 The Return of the King     hobbit  2463
17 The Return of the King     man      401
18 The Return of the King     man     3589
```

```
#"Hobbit" in the movie "fellow ship of the king" and "man" in "return of the king"
#seem to be dominant race speaking words in those movies.

#Therefore the number of words spoken by the dominant race does depend on the race
```