Assignment I

Data Visualization

Benjámin Olivér Maday

2023-10-02

Task 1

In the following exercise I would like to show you a little fun fact about our favourite galaxy far far away. If we take a look at Star Wars characters

```
starwars_tatooine_summary <- starwars %>%
  filter(homeworld == "Tatooine") %>%
  drop_na(mass) %>%
  group_by(eye_color) %>%
  mutate('avg_mass' = mean(mass))
```

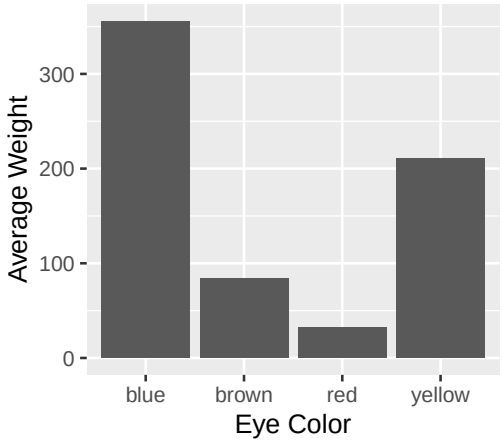

Figure 1: The average mass of Star Wars characters in relation to their eye colors.

Task 2

In the following task I would like to create a tidy data table from table4a. If we take a look at the original table (Table 1) we can see that 1999 and 2000 are represented as columns.

```
## # A tibble: 3 x 3
```

```
##   country     `1999` `2000`
##   <chr>        <dbl>  <dbl>
## 1 Afghanistan    745   2666
## 2 Brazil       37737  80488
## 3 China       212258 213766
```

This is not optimal for our purposes, since the variables we are interested in are not the years, but the occurence of tubercolosis and the years. Thus we pivot table4a to create a table where each observation has its own row, and the columns are country, year and tubercolosis_cases (Table 2).

```
table4a_tidy <- table4a %>%
  pivot_longer(
    c("1999", "2000"),
    names_to = "year",
    values_to = "tubercolosis_cases"
  )
```

```
## # A tibble: 6 x 3
##   country     year  tubercolosis_cases
##   <chr>       <chr>              <dbl>
## 1 Afghanistan 1999                 745
## 2 Afghanistan 2000                2666
## 3 Brazil      1999               37737
## 4 Brazil      2000               80488
## 5 China       1999              212258
## 6 China       2000              213766
```

Task 3

In task three, I created a new table4, by combining table4a and table4b. First I did the same thing with table4b as I did with 4a in the previous exercise. Then I joined the two columns based on the country and year columns. Finally I created a new column, which shows the rate of tubercolosis cases within the population.

```
## # A tibble: 6 x 5
##   country     year  tubercolosis_cases population      rate
##   <chr>       <chr>              <dbl>      <dbl>     <dbl>
## 1 Afghanistan 1999                 745   19987071 0.0000373
## 2 Afghanistan 2000                2666   20595360 0.000129
## 3 Brazil      1999               37737  172006362 0.000219
## 4 Brazil      2000               80488  174504898 0.000461
## 5 China       1999              212258 1272915272 0.000167
## 6 China       2000              213766 1280428583 0.000167
```