Lab 03 - Nobel laureates

Status200OK: Kamila Palys, Manav Bhagat, Sai Harshitha Dalli, Shreehar Joshi

9/24/20

Load packages and data

```
library(tidyverse)
```

```
nobel <- read_csv("data/nobel.csv")
```

Exercise 1

The number of observations is 935 and the number of variables is 26.

Exercise 2

First off, creating a new dataframe by removing laureates for whom country is missing, who are not living anymore, and laureates who are organizations.

```
nobel_living <- filter(nobel, !is.na(country),
                       gender != "org", is.na(died_date))
```

The resulting number of observations is 228.

Exercise 3

Adding a new variable called country_us to identify whether the laureate was in the US when they won their prize and only considering Physics, Medicine, Chemistry, and Economics as the categories of prize.

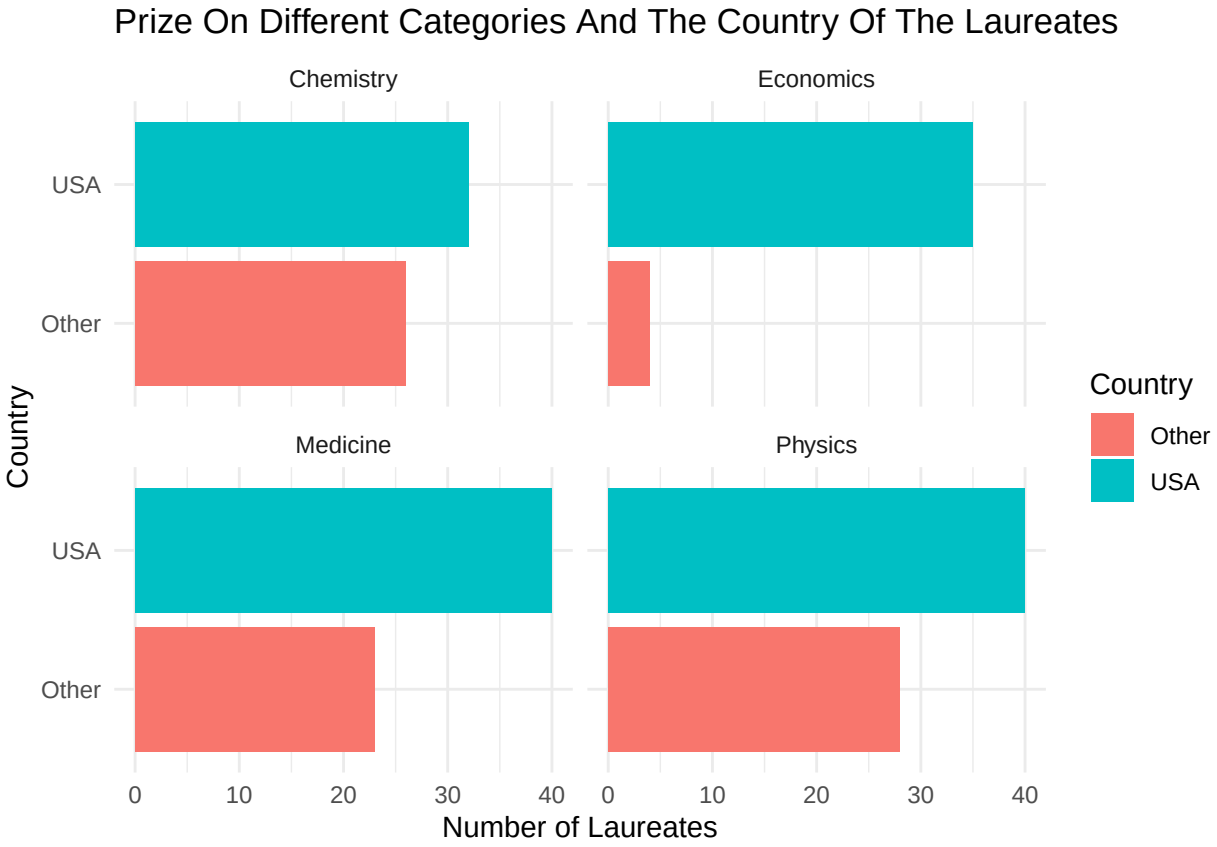
```
nobel_living <- nobel_living %>%
  mutate(
    country_us = if_else(country == "USA", "USA", "Other")
  )

nobel_living_science <- nobel_living %>%
  filter(category %in% c("Physics", "Medicine", "Chemistry", "Economics"))
```

Now, creating a faceted bar plot visualizing the relationship between the category of prize and whether the laureate was in the US when they won the nobel prize.

```
ggplot(data = nobel_living_science,
       mapping = aes(x = country_us, fill= country_us)) +
  geom_bar() +
  facet_wrap(~category) +
  labs(title= "Prize On Different Categories And The Country Of The Laureates ",
    x= "Country",
```

```
    y= "Number of Laureates",
    fill="Country") +
  coord_flip()+
  theme_minimal()
```



In all the above four categories, the number of laureates from the US is higher than the total number of laureates from the rest of the world. While the differences in the number of laureates from the US and the rest of the world varies in all the four categories (Economics having the highest difference and Chemistry having the lowest difference), in all four of the fields US exceeds the rest of the world put together in having the most number of laureates. Hence it fully supports the Buzzfeed's headline which states, "Most living Nobel laureates were based in the US when they won their prizes"

Exercise 4

Creating a new variable that distinguishes laureates on whether they were born in the US or not.

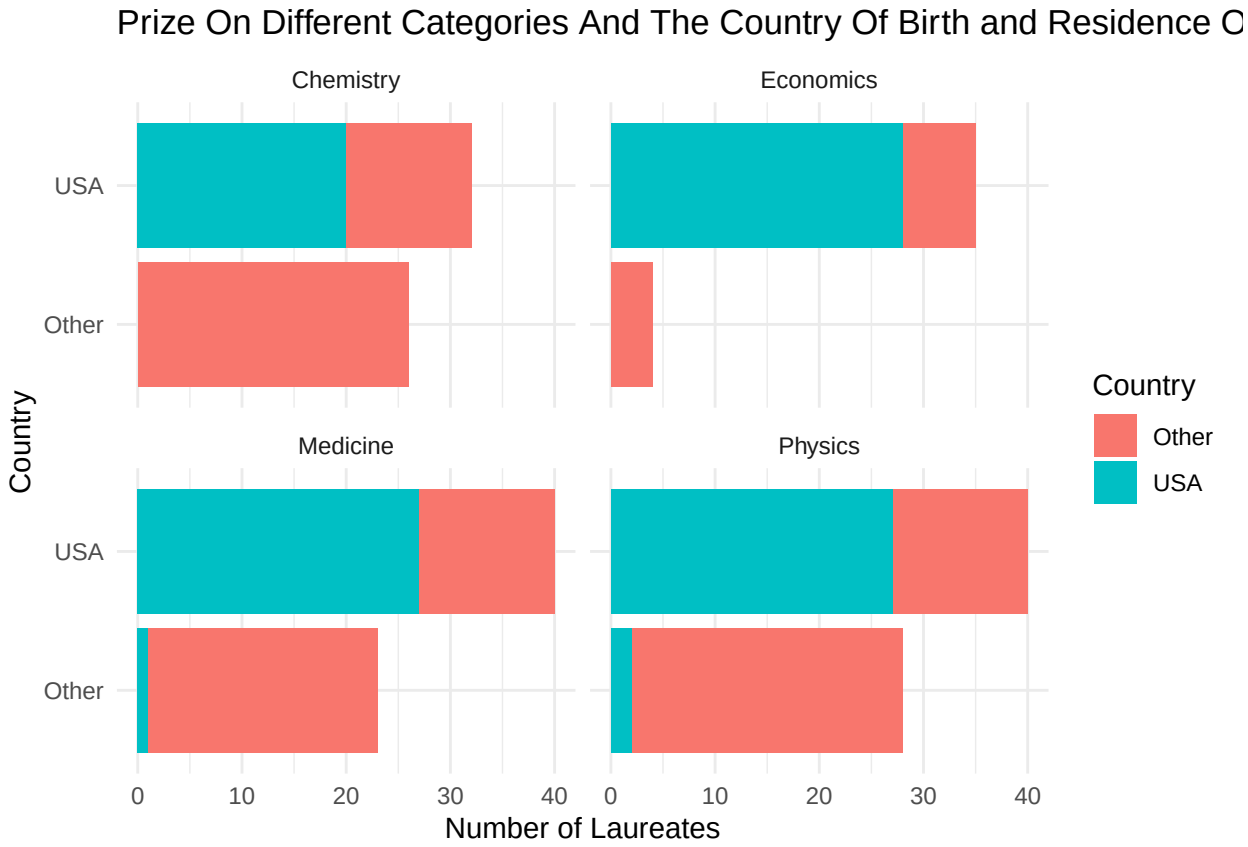
```
nobel_living_science <- nobel_living_science %>%
  mutate(born_country_us = if_else(born_country == "USA", "USA", "Other" ))
```

Exercise 5

Now, creating a faceted bar plot visualizing the relationship between the category of prize and whether the laureate was in the US when they won the nobel prize and if they were born in the US.

```
ggplot(data = nobel_living_science,
       mapping = aes(x = country_us, fill= born_country_us)) +
```

```
  geom_bar() +
  facet_wrap(~category) +
  labs(title= "Prize On Different Categories And The Country Of Birth and Residence Of The Laureates ",
    x= "Country ",
    y= "Number of Laureates",
    fill="Country") +
  coord_flip()+
  theme_minimal()
```



In all the bars representing the countries of Laureates as the USA, there appear a number of laureates who were not born in the US (which is represented by the red blocks beside the blue ones). Hence the claim made by the Buzzfeed ("But of those US-based Nobel laureates, many were born in other countries…") is supported by the bar chart above.

Exercise 6

Now, lets look onto the number of laureates from different countries who were not born in US but were living in US when they won the prize.

```
nobel_living_science %>%
  filter(country == "USA", born_country != "USA") %>%
  count(born_country, sort = TRUE)
```

```
## # A tibble: 21 x 2
##    born_country       n
##    <chr>          <int>
##  1 Germany            7
```

```
##  2 United Kingdom     7
##  3 China              5
##  4 Canada             4
##  5 Japan              3
##  6 Australia          2
##  7 Israel             2
##  8 Norway             2
##  9 Austria            1
## 10 Finland            1
## # ... with 11 more rows
```