Lab 02 - Nobel laureates

Ronit Singh 10/11/2020

Load packages and data

```
library(tidyverse)

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

Exercise 1

There are 935 rows or observations and 26 variables or columns in the dataset.

```
glimpse(nobel)

## Rows: 935
## Columns: 26
## $ id                    <dbl> 1, 2, 3, 4, 5, 6, 6, 8, 9, 10, 11, 12, 13, 14, …
## $ firstname             <chr> "Wilhelm Conrad", "Hendrik A.", "Pieter", "Henr…
## $ surname               <chr> "Röntgen", "Lorentz", "Zeeman", "Becquerel", "C…
## $ year                  <dbl> 1901, 1902, 1902, 1903, 1903, 1903, 1911, 1904,…
## $ category              <chr> "Physics", "Physics", "Physics", "Physics", "Ph…
## $ affiliation           <chr> "Munich University", "Leiden University", "Amst…
## $ city                  <chr> "Munich", "Leiden", "Amsterdam", "Paris", "Pari…
## $ country               <chr> "Germany", "Netherlands", "Netherlands", "Franc…
## $ born_date             <date> 1845-03-27, 1853-07-18, 1865-05-25, 1852-12-15…
## $ died_date             <date> 1923-02-10, 1928-02-04, 1943-10-09, 1908-08-25…
## $ gender                <chr> "male", "male", "male", "male", "male", "female…
## $ born_city             <chr> "Remscheid", "Arnhem", "Zonnemaire", "Paris", "…
## $ born_country          <chr> "Germany", "Netherlands", "Netherlands", "Franc…
## $ born_country_code     <chr> "DE", "NL", "NL", "FR", "FR", "PL", "PL", "GB",…
## $ died_city             <chr> "Munich", NA, "Amsterdam", NA, "Paris", "Sallan…
## $ died_country          <chr> "Germany", "Netherlands", "Netherlands", "Franc…
## $ died_country_code     <chr> "DE", "NL", "NL", "FR", "FR", "FR", "GB", "GB",…
## $ overall_motivation    <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,…
## $ share                 <dbl> 1, 2, 2, 2, 4, 4, 1, 1, 1, 1, 1, 2, 2, 1, 1, …
## $ motivation            <chr> "\"in recognition of the extraordinary services…
## $ born_country_original <chr> "Prussia (now Germany)", "the Netherlands", "th…
## $ born_city_original    <chr> "Lennep (now Remscheid)", "Arnhem", "Zonnemaire…
## $ died_country_original <chr> "Germany", "the Netherlands", "the Netherlands"…
## $ died_city_original    <chr> "Munich", NA, "Amsterdam", NA, "Paris", "Sallan…
## $ city_original         <chr> "Munich", "Leiden", "Amsterdam", "Paris", "Pari…
## $ country_original      <chr> "Germany", "the Netherlands", "the Netherlands"…
```

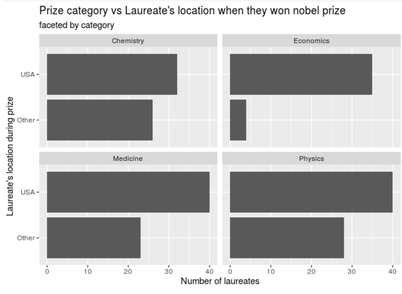
Exercise 2

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

## # A tibble: 228 x 26
##       id firstname surname  year category affiliation city  country born_date
##    <dbl> <chr>     <chr>   <dbl> <chr>    <chr>       <chr> <chr>   <date>
##  1    68 Chen Ning Yang     1957 Physics  Institute … Prin… USA     1922-09-22
##  2    69 Tsung-Dao Lee      1957 Physics  Columbia U… New … USA     1926-11-24
##  3    95 Leon N.   Cooper    1972 Physics  Brown Univ… Prov… USA     1930-02-28
##  4    97 Leo       Esaki    1973 Physics  IBM Thomas… York… USA     1925-03-12
##  5    98 Ivar      Giaever  1973 Physics  General El… Sche… USA     1929-04-05
##  6    99 Brian D.  Joseph…  1973 Physics  University… Camb… United… 1940-01-04
##  7   101 Antony    Hewish   1974 Physics  University… Camb… United… 1924-05-11
##  8   103 Ben R.    Mottel…  1975 Physics  Nordita     Cope… Denmark 1926-07-09
##  9   106 Samuel C… Ting     1976 Physics  Massachuse… Camb… USA     1936-01-27
## 10   107 Philip W. Anders…  1977 Physics  Bell Telep… Murr… USA     1923-12-13
## # … with 218 more rows, and 17 more variables: died_date <date>, gender <chr>,
## #   born_city <chr>, born_country <chr>, born_country_code <chr>,
## #   died_city <chr>, died_country <chr>, died_country_code <chr>,
## #   overall_motivation <chr>, share <dbl>, motivation <chr>,
## #   born_country_original <chr>, born_city_original <chr>,
## #   died_country_original <chr>, died_city_original <chr>, city_original <chr>,
## #   country_original <chr>
```

Exercise 3

```
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"))
ggplot(data = nobel_living_science, mapping = aes(y = country_us)) +
  facet_wrap(~category) +
  geom_bar() +
    labs(title = "Prize category vs Laureate's location when they won nobel prize",
         subtitle = "faceted by category",
       x = "Number of laureates", y = "Laureate's location during prize")
```



The graph clearly shows that in each category, the number of living nobel laureates were in the US (when getting the prize) is greater than that of other countries. Thus, the statement by Buzzfeed, "Most living Nobel laureates were based in the US when they won their prizes" is supported by the data.

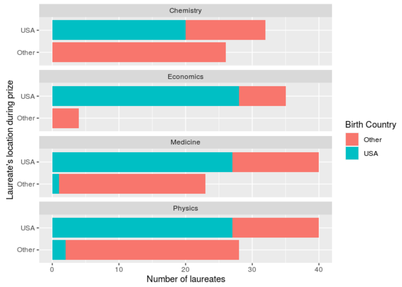
Exercise 4

Creating a new variable called born_country_us that has the value "USA" if the laureate is born in the US, and "Other" otherwise.

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

Exercise 5

```
nobel_living <- nobel_living %>%
  mutate(
    country_us = if_else(country == "USA", "USA", "Other"),
    born_country_us = if_else(born_country == "USA", "USA", "Other")
  )
nobel_living_science <- nobel_living %>%
  filter(category %in% c("Physics", "Medicine", "Chemistry", "Economics"))
ggplot(data = nobel_living_science, mapping = aes(y = country_us, fill = born_country_us)) +
  facet_wrap(~category, ncol = 1) +
  geom_bar() +
  labs(x = "Number of laureates", y = "Laureate's location during prize") + labs(fill = "Birth Country")
```



The data shows that laureates who were US citizens and won the nobel prize in the US (indicated by blue), many of them were born in different countries (other than the US), indicated by orange. Thus, the statement, "But of those US-based Nobel laureates, many were born in other countries" is supported by the data.

Exercise 6

```
nobel_living %>% filter(country == "USA",
                        born_country_us == "Other") %>% 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
```