Lab 02 - Nobel laureates

Aleksandra Medina October 15, 2020

Load packages and data

```
library(tidyverse)
```

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

Exercise 1

There are 26 columns which equates to 26 variables in the dataset (26 things measured for each laureate), and there are 935 rows equating to 935 observations (935 laureates).

```
glimpse(nobel)
```

```
## Rows: 935
## Columns: 26
## $ id                  <dbl> 1, 2, 3, 4, 5, 6, 6, 8, 9, 10, 11, 12, 1
## $ firstname           <chr> "Wilhelm Conrad", "Hendrik A.", "Pieter"
## $ surname             <chr> "Röntgen", "Lorentz", "Zeeman", "Becquer
## $ year                <dbl> 1901, 1902, 1902, 1903, 1903, 1903, 1911
## $ category            <chr> "Physics", "Physics", "Physics", "Physic
## $ affiliation         <chr> "Munich University", "Leiden University"
## $ city                <chr> "Munich", "Leiden", "Amsterdam", "Paris"
## $ country             <chr> "Germany", "Netherlands", "Netherlands",
## $ born_date           <date> 1845-03-27, 1853-07-18, 1865-05-25, 185
## $ died_date           <date> 1923-02-10, 1928-02-04, 1943-10-09, 190
## $ gender              <chr> "male", "male", "male", "male", "male",
## $ born_city           <chr> "Remscheid", "Arnhem", "Zonnemaire", "Pa
## $ born_country        <chr> "Germany", "Netherlands", "Netherlands",
## $ born_country_code   <chr> "DE", "NL", "NL", "FR", "FR", "PL", "PL"
## $ died_city           <chr> "Munich", NA, "Amsterdam", NA, "Paris",
## $ died_country        <chr> "Germany", "Netherlands", "Netherlands",
```

```
## $ died_country_code     <chr> "DE", "NL", "NL", "FR", "FR", "FR", "FR"
## $ overall_motivation    <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,
## $ share                 <dbl> 1, 2, 2, 2, 4, 4, 1, 1, 1, 1, 1, 1, 2, 2
## $ motivation            <chr> "\"in recognition of the extraordinary s
## $ born_country_original <chr> "Prussia (now Germany)", "the Netherland
## $ born_city_original    <chr> "Lennep (now Remscheid)", "Arnhem", "Zon
## $ died_country_original <chr> "Germany", "the Netherlands", "the Nethe
## $ died_city_original    <chr> "Munich", NA, "Amsterdam", NA, "Paris",
## $ city_original         <chr> "Munich", "Leiden", "Amsterdam", "Paris"
## $ country_original      <chr> "Germany", "the Netherlands", "the Nethe
```

Exercise 2

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

nobel_living
```

```
## # A tibble: 228 x 26
##       id firstname surname  year category affiliation city  country bo
##    <dbl> <chr>     <chr>   <dbl> <chr>    <chr>       <chr> <chr>   <d
##  1    68 Chen Ning Yang     1957 Physics  Institute … Prin… USA     19
##  2    69 Tsung-Dao Lee      1957 Physics  Columbia U… New … USA     19
##  3    95 Leon N.   Cooper   1972 Physics  Brown Univ… Prov… USA     19
##  4    97 Leo       Esaki    1973 Physics  IBM Thomas… York… USA     19
##  5    98 Ivar      Giaever  1973 Physics  General El… Sche… USA     19
##  6    99 Brian D.  Joseph…  1973 Physics  University… Camb… United… 19
##  7   101 Antony    Hewish   1974 Physics  University… Camb… United… 19
##  8   103 Ben R.    Mottel…  1975 Physics  Nordita     Cope… Denmark 19
##  9   106 Samuel C… Ting     1976 Physics  Massachuse… Camb… USA     19
## 10   107 Philip W. Anders…  1977 Physics  Bell Telep… Murr… USA     19
## # … with 218 more rows, and 17 more variables: died_date <date>, gende
## #   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_origin
## #   country_original <chr>
```

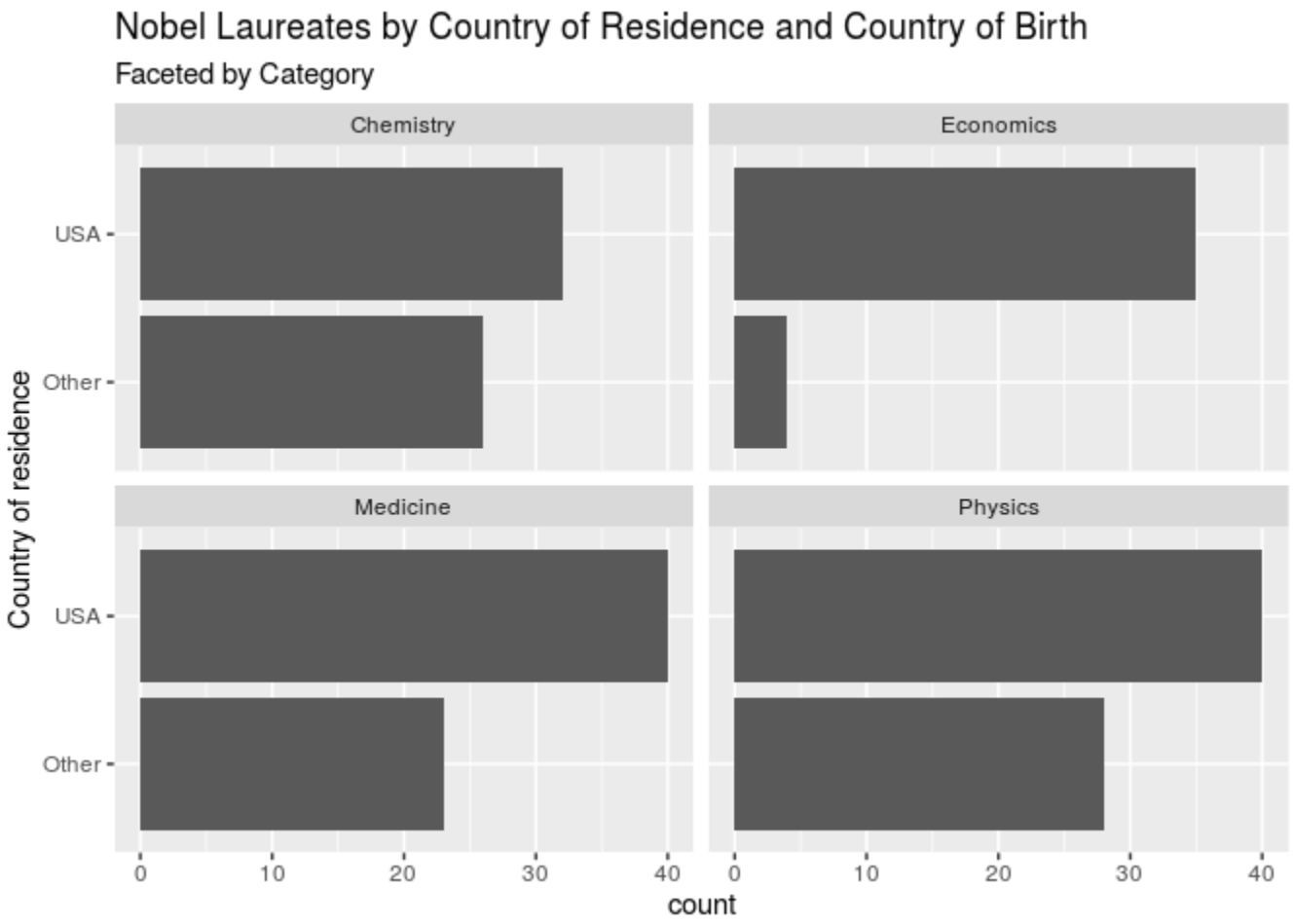
Exercise 3

Yes, technically, the Buzzfeed article was correct. However, one must emphasize that the laureates were in the US at the time of receiving the prize. That doesn't mean that they are US citizens.

```
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")


nobel_living_science %>%
  ggplot(
    mapping = aes(x = country_us)
  )+
  geom_bar()+
  facet_wrap(~ category, nrow = 2)+
  coord_flip()+
   labs (
    title = "Nobel Laureates by Country of Residence and Country of Birth
    subtitle = "Faceted by Category",
    x = "Country of residence"
    )
```



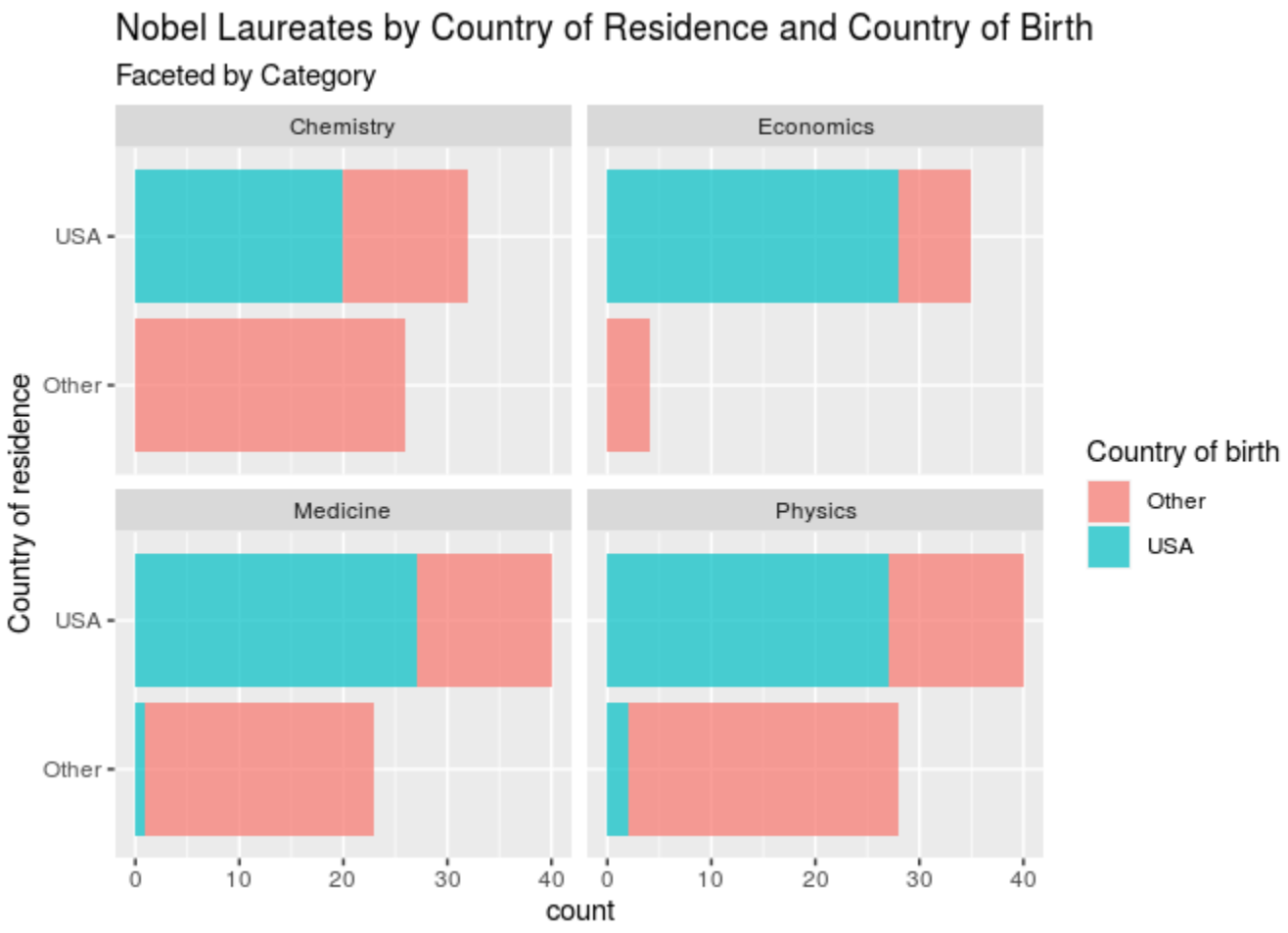
Exercise 4

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

Exercise 5

Technically, the Buzzfeed article was correct - most recipients of the Nobel prize were inside the US at the time of recieval. However, the article fails to emphasize that a considerable chunk of these recipients are immigrants to the US and were born elsewhere. While most recipients of the Nobel Prize in Economics were born in the US (maybe their parents weren't?), a significant amount of recipients of the Nobel Prize in Chemistry, Medicine and Physics were not born in the US.

```
nobel_living_science %>%
  ggplot(
    mapping = aes(x = country_us, fill = born_country_us)
  )+
  geom_bar(
    alpha = 0.7,
  )+
  facet_wrap(~ category, nrow = 2)+
  coord_flip()+
  labs (
    title = "Nobel Laureates by Country of Residence and Country of Birth
    subtitle = "Faceted by Category",
    x = "Country of residence",
    fill = "Country of birth"
    )
```



Exercise 6

```
nobel_living_science %>%
```

```
  filter (country_us == "USA",
          born_country != "USA"
          ) %>%
  count(born_country) %>%
  arrange(desc(n))
```

```
## # 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
```