DATA607_w7 - Extra Credit

David Simbandumwe

2021-10-20

Introduction

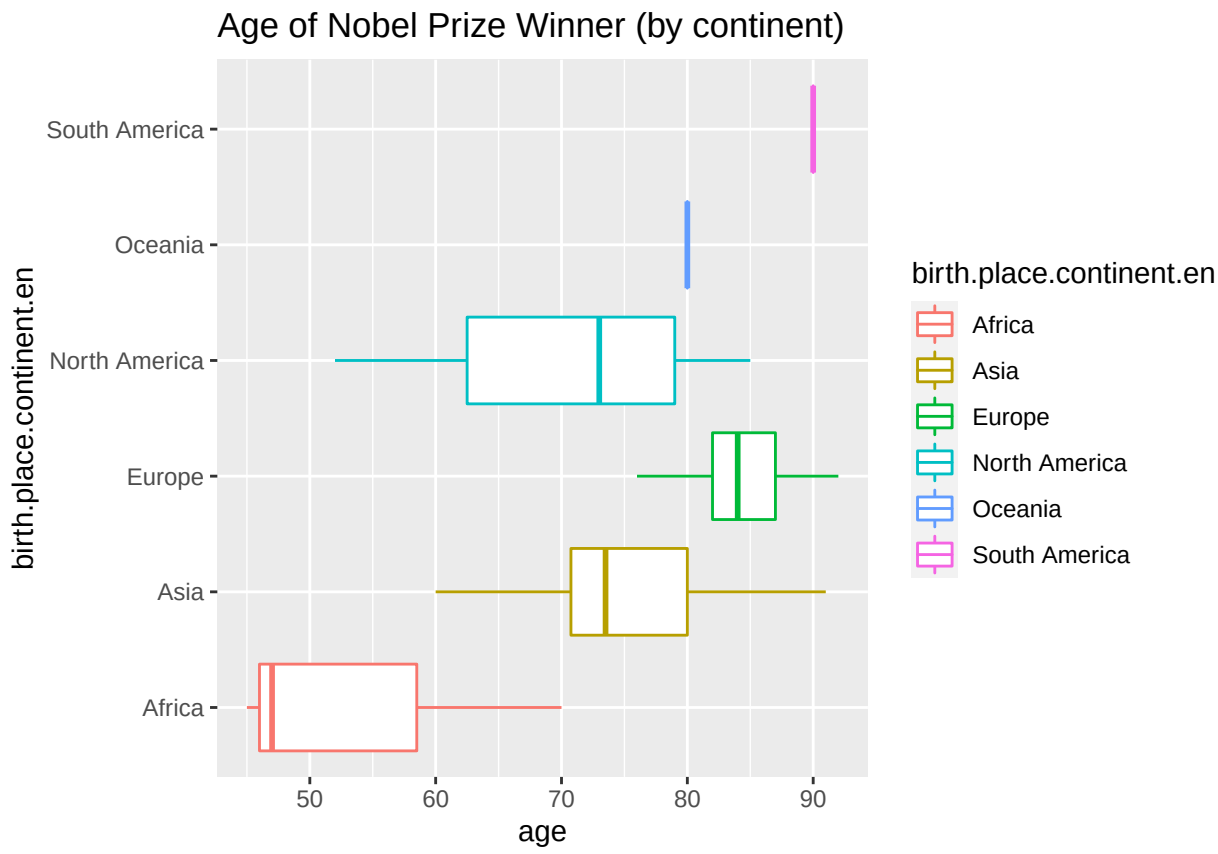
The objective of this effort was to use the Nobel Prize API to answer 3 questions about the data set. I started with the following questions. 1. How does life spans compair across regions and countries for Nobel Laureates? 2. How has the prize money for the Nobel Prize changed overtime? 3. Which country has recevied the most prize money from the Nobel committee?

Life Span

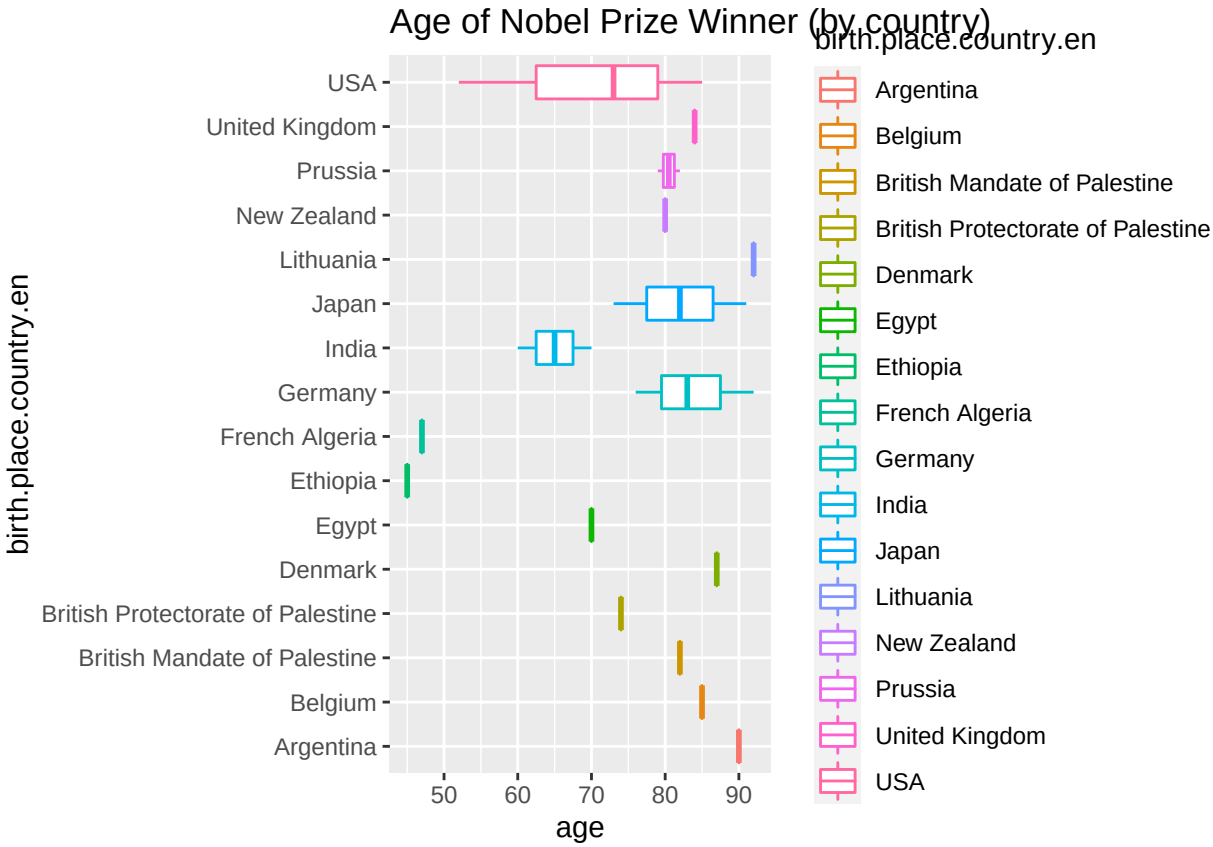
```
raw_json = fromJSON("https://api.nobelprize.org/2.1/laureates",
                    simplifyDataFrame = TRUE
                    )


life_df  <- pluck(raw_json, "laureates") %>%
  as_tibble() %>%
  flatten() %>%
  select(id, familyName.en, givenName.en, gender, birth.date, death.date, birth.place.continent.en, bir
  mutate (
    death.date = ifelse(is.na(death.date), as.character.Date(Sys.Date()), death.date),
    death.date = as_date(death.date, tz = NULL, format = NULL),
    birth.date = as_date(birth.date, tz = NULL, format = NULL),
    age = year(death.date) - year(birth.date)
  ) %>%
  drop_na()
```

```
## Warning: 2 failed to parse.
```

```
life_df %>%
  ggplot() +
  geom_boxplot(mapping = aes(y=birth.place.continent.en, x=age, color=birth.place.continent.en)) +
  labs (title = "Age of Nobel Prize Winner (by continent)" )
```



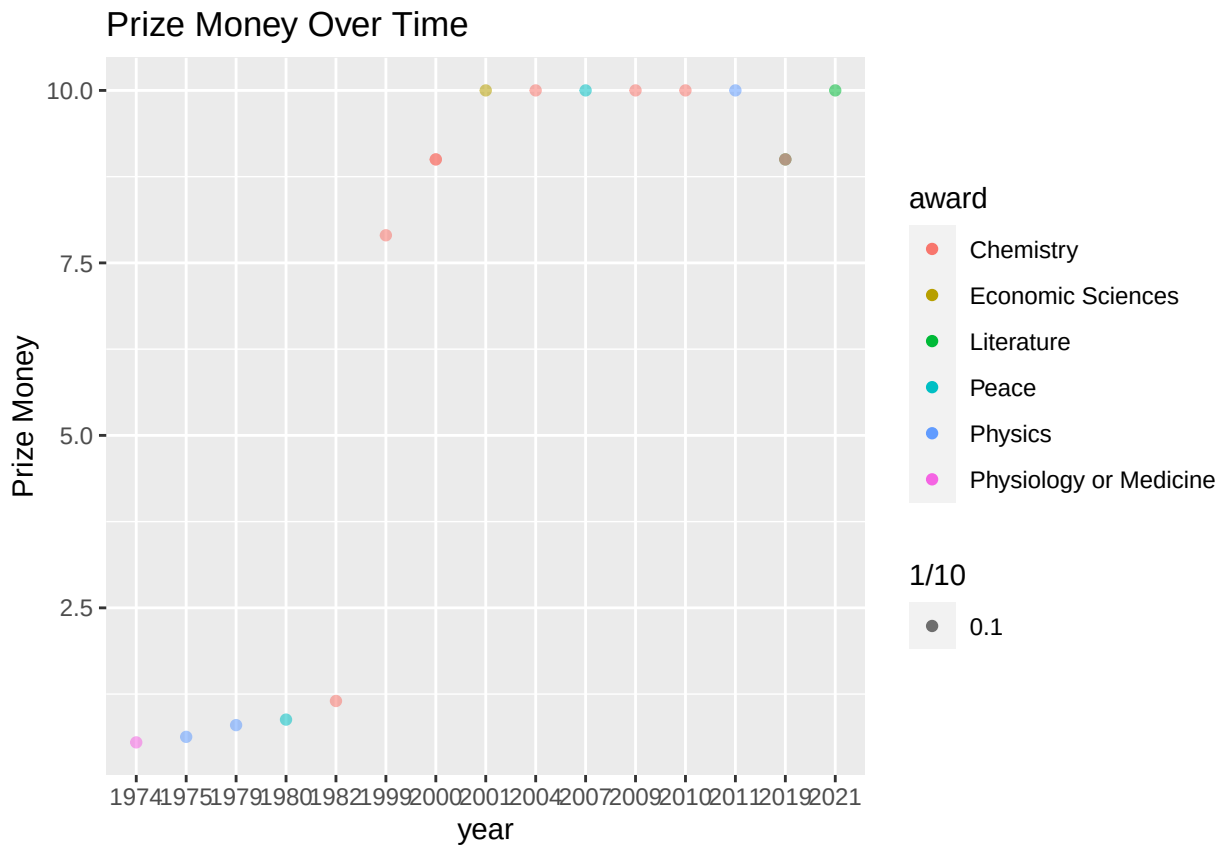

```
life_df %>%
  ggplot() +
  geom_boxplot(mapping = aes(y=birth.place.country.en, x=age, color=birth.place.country.en)) +
  labs (title = "Age of Nobel Prize Winner (by country)" )
```



- **Lithuania has set the mark for the country with the Nobel Laureates that lived the longest.**
- **South america sets the mark for the region with the Nobel Laureates that lived the longest.**

```
raw_json = fromJSON("https://api.nobelprize.org/2.1/laureates")

prize_df  <- pluck(raw_json, "laureates") %>%
  as_tibble() %>%
  flatten()


prize_df <- prize_df %>%
  select(id, familyName.en, givenName.en, gender, birth.date, birth.place.countryNow.en, nobelPrizes)


prize_df <- prize_df %>%
  unnest_wider(nobelPrizes, simplify = TRUE) %>%
  unnest_wider(category, simplify = TRUE) %>%
  select(id, familyName.en, givenName.en, gender, birth.date, birth.place.countryNow.en, awardYear, en
  mutate (
    prizeAmount = prizeAmount/1000000
  ) %>%
  rename (
    award = en
  )
```

```
prize_df %>%
  filter(awardYear > 1970) %>%
  ggplot() +
  geom_point(mapping = aes(x=awardYear, y=prizeAmount, color=award, alpha=1/10)) +
  labs(
    title = "Prize Money Over Time",
    x = "year",
    y = "Prize Money"
  )
```



- **Prize money was relativily flat until 1982 then it spiked upwards**

```
prize_df %>%
  group_by(birth.place.countryNow.en) %>%
  mutate (
    sum_prize = sum(prizeAmount)
  ) %>%
  select(birth.place.countryNow.en, sum_prize) %>%
  distinct() %>%
  ggplot(aes( y=birth.place.countryNow.en, x=sum_prize)) +
  geom_bar(position="dodge", stat="identity") +
  labs(
    title = "Prize Money by Country",
    y = "Country",
```

```
    x = "Prize Money"
)
```

- **The united states has won the lions share of the Nobel Awards over the years**