Billionaires

Fabian & Robin

31.12.2023

This is the R project of Fabian and Robin. We decided to analyse data from billionaires.
For that, we found a dataset on kaggle.com: https://www.kaggle.com/datasets/nelgiriyewithana/billionaires-statistics-dataset
This dataset contains data from the 4th of april 2023 and is used for 'exploring the global landscape of success'.

```
# We used these libraries:
library(ggplot2) # to plot graphs and images
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggmap)
```

```
## i Google's Terms of Service: <https://mapsplatform.google.com>
##   Stadia Maps' Terms of Service: <https://stadiamaps.com/terms-of-service/>
##   OpenStreetMap's Tile Usage Policy: <https://operations.osmfoundation.org/policies/tiles/>
## i Please cite ggmap if you use it! Use `citation("ggmap")` for details.
```

```
library(ggcorrplot)
library(ggthemes)
```

Read the CSV file into a dataframe called df_billionaires

```
df_billionaires <- read.csv("BillionairesStatisticsDataset.csv")
# display data
head(df_billionaires)
```

```
##   rank finalWorth              category               personName age
## 1    1     211000      Fashion & Retail Bernard Arnault & family  74
## 2    2     180000            Automotive                Elon Musk  51
## 3    3     114000            Technology               Jeff Bezos  59
## 4    4     107000            Technology            Larry Ellison  78
## 5    5     106000 Finance & Investments           Warren Buffett  92
## 6    6     104000            Technology               Bill Gates  67
```

```
##         country   city             source            industries
## 1        France  Paris               LVMH      Fashion & Retail
## 2 United States Austin      Tesla, SpaceX            Automotive
## 3 United States Medina             Amazon            Technology
## 4 United States  Lanai             Oracle            Technology
## 5 United States  Omaha Berkshire Hathaway Finance & Investments
## 6 United States Medina          Microsoft            Technology
##   countryOfCitizenship                      organization selfMade status gender
## 1               France LVMH Moët Hennessy Louis Vuitton    FALSE      U      M
## 2        United States                             Tesla     TRUE      D      M
## 3        United States                            Amazon     TRUE      D      M
## 4        United States                            Oracle     TRUE      U      M
## 5        United States   Berkshire Hathaway Inc. (Cl A)     TRUE      D      M
## 6        United States  Bill & Melinda Gates Foundation     TRUE      D      M
##         birthDate lastName firstName               title          date
## 1   3/5/1949 0:00  Arnault   Bernard     Chairman and CEO 4/4/2023 5:01
## 2  6/28/1971 0:00     Musk      Elon                 CEO 4/4/2023 5:01
## 3  1/12/1964 0:00    Bezos      Jeff Chairman and Founder 4/4/2023 5:01
## 4  8/17/1944 0:00  Ellison     Larry      CTO and Founder 4/4/2023 5:01
## 5  8/30/1930 0:00  Buffett    Warren                 CEO 4/4/2023 5:01
## 6 10/28/1955 0:00    Gates      Bill              Cochair 4/4/2023 5:01
##        state residenceStateRegion birthYear birthMonth birthDay cpi_country
## 1                                      1949          3        5      110.05
## 2      Texas                South      1971          6       28      117.24
## 3 Washington                 West      1964          1       12      117.24
## 4     Hawaii                 West      1944          8       17      117.24
## 5   Nebraska              Midwest      1930          8       30      117.24
## 6 Washington                 West      1955         10       28      117.24
##   cpi_change_country          gdp_country gross_tertiary_education_enrollment
## 1                1.1  $2,715,518,274,227                                 65.6
## 2                7.5 $21,427,700,000,000                                 88.2
## 3                7.5 $21,427,700,000,000                                 88.2
## 4                7.5 $21,427,700,000,000                                 88.2
## 5                7.5 $21,427,700,000,000                                 88.2
## 6                7.5 $21,427,700,000,000                                 88.2
##   gross_primary_education_enrollment_country life_expectancy_country
## 1                                      102.5                    82.5
## 2                                      101.8                    78.5
## 3                                      101.8                    78.5
## 4                                      101.8                    78.5
## 5                                      101.8                    78.5
## 6                                      101.8                    78.5
##   tax_revenue_country_country total_tax_rate_country population_country
## 1                        24.2                   60.7            67059887
## 2                         9.6                   36.6           328239523
## 3                         9.6                   36.6           328239523
## 4                         9.6                   36.6           328239523
## 5                         9.6                   36.6           328239523
## 6                         9.6                   36.6           328239523
##   latitude_country longitude_country
## 1         46.22764           2.213749
## 2         37.09024         -95.712891
## 3         37.09024         -95.712891
## 4         37.09024         -95.712891
```

```
## 5         37.09024        -95.712891
## 6         37.09024        -95.712891
```

```
# str(df_billionaires)
```

First inspection of the data

Each row displays one person that is a billionaire. The columns contain the following details about each person:

- rank
  - The ranking of the billionaire in terms of wealth (from 1 to 2640).
- finalWorth
  - The final net worth of the billionaire in U.S. dollars.
- category
  - The category or industry in which the billionaire's business operates.
- personName
  - The full name of the billionaire.
- age
  - The age of the billionaire in years.
- country
  - The country in which the billionaire resides.
- city
  - The city in which the billionaire resides
- source
  - The source of the billionaire's wealth (eg. company name).
- industries
  - The industries associated with the billionaire's business interests.
- countryOfCitizenship
  - The country of citizenship of the billionaire.
- organization
  - The name of the organization or company associated with the billionaire.
- selfMade
  - Indicates whether the billionaire is self-made (True/False).
- status
  - D=Entrepreneur and U = Inherited
- gender
  - The gender of the billionaire.
- birthDate
  - The birthdate of the billionaire.

- lastName
  - The last name of the billionaire.
- firstName
  - The first name of the billionaire.
- title
  - The status or position of the billionaire (e.g., CEO, Founder).
- date
  - The date of data collection.
- state
  - The state in which the billionaire resides.
- redidenceStateRegion
  - The region or state of residence of the billionaire.
- birthYear
  - The birth year of the billionaire.
- birthMonth
  - The birth month of the billionaire.
- birthDay
  - The birth day of the billionaire.
- cpi_country
  - Consumer Price Index (CPI) for the billionaire's country.
- cpi_change_country
  - CPI change for the billionaire's country.
- gdp_country
  - Gross Domestic Product (GDP) for the billionaire's country.
- gross_tertiary_education_enrollment
  - Enrollment in tertiary education in the billionaire's country.
- gross_primary_education_enrollment_country
  - Enrollment in primary education in the billionaire's country.
- life_expectancy_country
  - Life expectancy in the billionaire's country.
- tax_revenue_country_country
  - Tax revenue in the billionaire's country.
- total_tax_rate_country
  - Total tax rate in the billionaire's country.
- population_country
  - Population of the billionaire's country.

- latitude_country
  - Latitude coordinate of the billionaire's country.

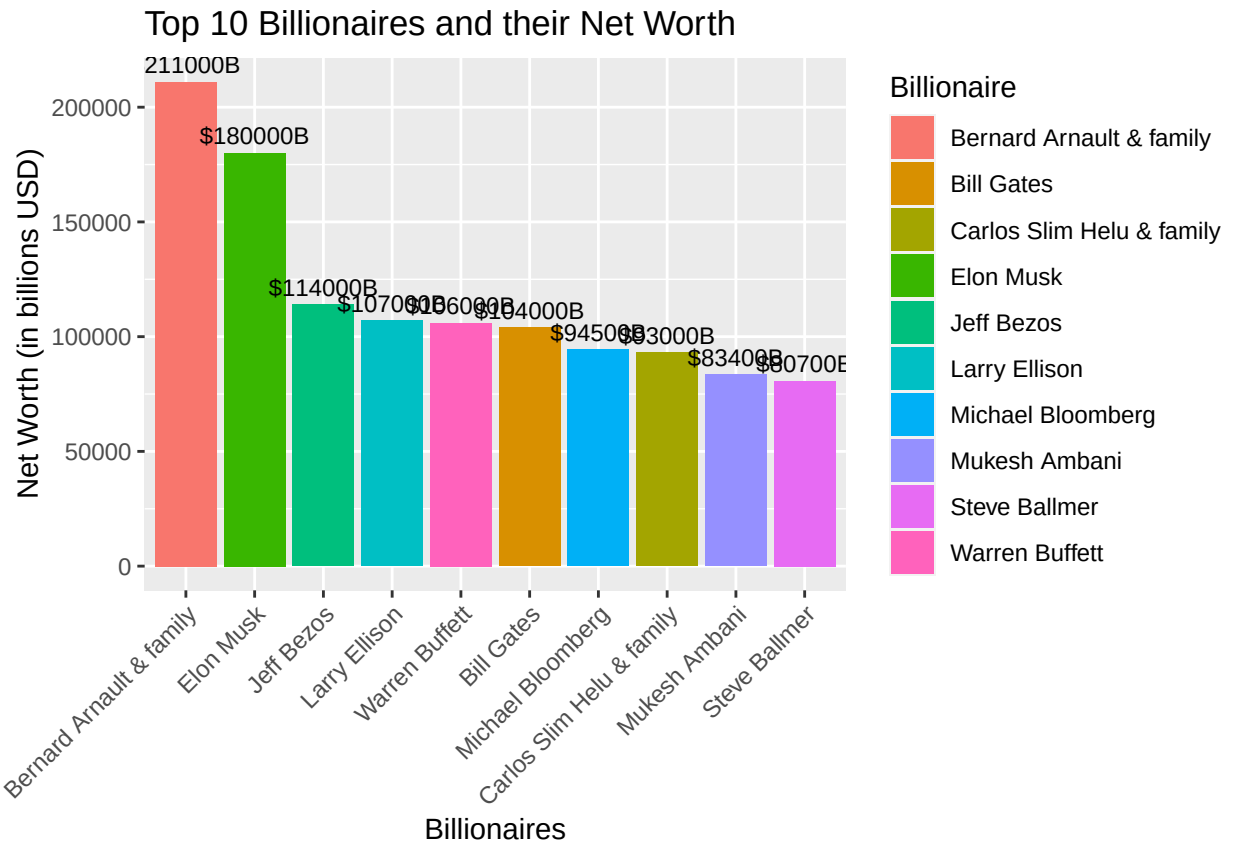
- longitude_country
  - Longitude coordinate of the billionaire's country.

We first started by creating a list of the top 10 richest billionaires:

```
# Create a data frame with the top 10 billionaires
top_10_billionaires <- df_billionaires[1:10, ]
```

Plot the top 10 richest billionaires

```
# Create a bar plot with text labels
top_10_billionaires_plot <- ggplot(top_10_billionaires, aes(x = reorder(personName, -finalWorth), y = fi
  geom_bar(stat = "identity") +
  geom_text(aes(label = paste0("$", finalWorth, "B")), vjust = -0.5, color = "black", size = 3) + # Add
  labs(
    title = "Top 10 Billionaires and their Net Worth",
    x = "Billionaires",
    y = "Net Worth (in billions USD)",
    fill = "Billionaire"
  ) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

# Print the plot
plot(top_10_billionaires_plot)
```



The plot displays ten bars in a barchart. Each bar represents the wealth of one billionaire. The plot only displays the top 10 richest billionaires. The height of the bar represents the wealth of the billionaire in US-dollars.

Next Steps

Next, we decided to formulate a thesis. Based on the dataset, we decided that we want to formulate our thesis about the correlation and causation between certain data columns and the wealth of the billionaires. Maybe this lets us speculate on certain circumstances that make it easier or harder for someone to become a billionaire.

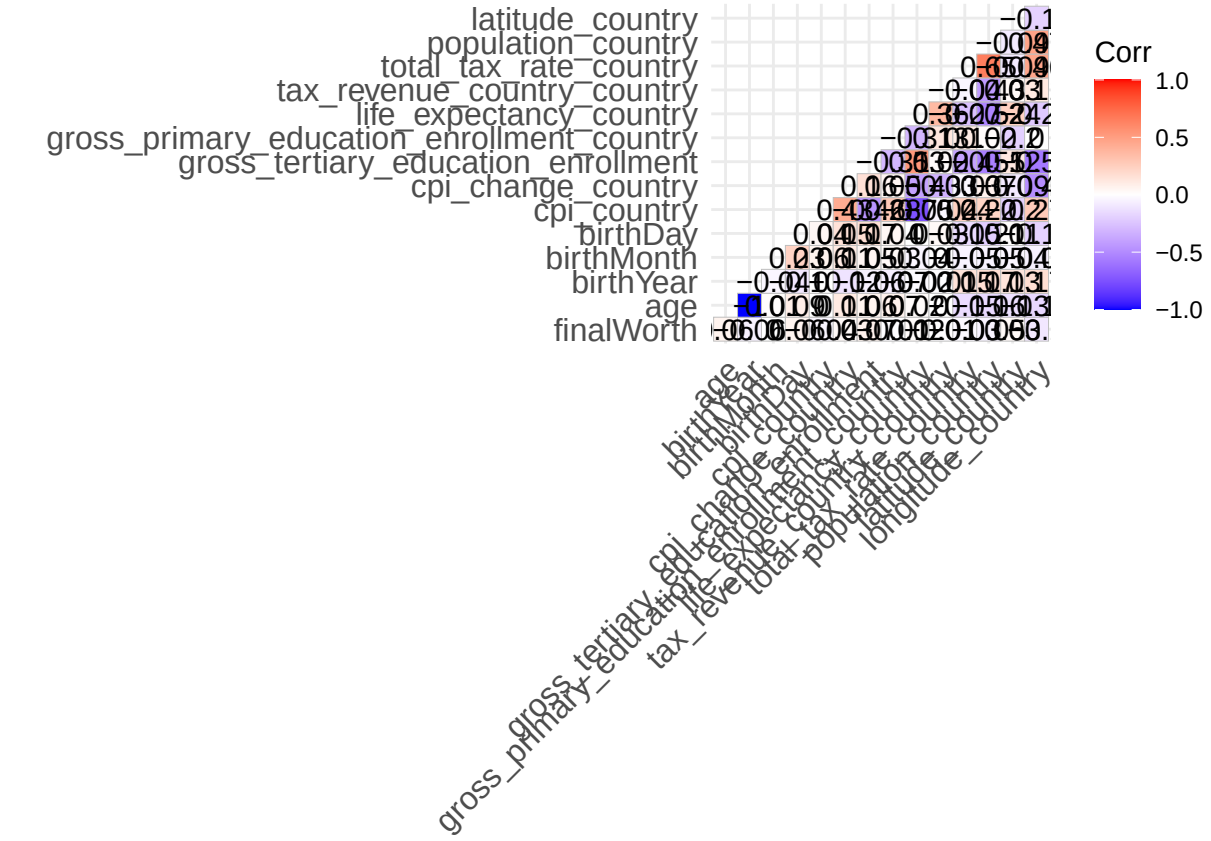
In the following we will create multiple plots to compare the correlation between certain data columns and the wealth. Because this dataset only contains billionaires, we won't focus on the difference between their wealth. Being part of this dataset means, that you're rich enough.

Correlation matrix between numeric values

The code calculates and visualizes the correlation matrix of numeric variables in the dataset, indicating the strength and direction of linear relationships between measures such as wealth, age, Consumer Price Index, Gross Domestic Product, life expectancy, and other relevant numeric columns.

```
df <- df_billionaires
# Convert 'birthDate' to Date type
df$birthDate <- as.Date(df$birthDate, format="%m/%d/%Y")

# Filter only numeric columns for correlation analysis
```

```r
numeric_cols <- sapply(df, is.numeric)
df_numeric <- df[, numeric_cols]

# Remove column rank
df_numeric <- df_numeric[, !colnames(df_numeric) %in% c('rank')]

# Remove NA and NaN values
df_numeric <- na.omit(df_numeric)

# Plot correlation matrix
cor_matrix <- cor(df_numeric)
ggcorrplot(cor_matrix, type = "lower", lab = TRUE)
```



As you can see... TODO: CHANGE SIZE OF PLOT TO MAKE IT READABLE?

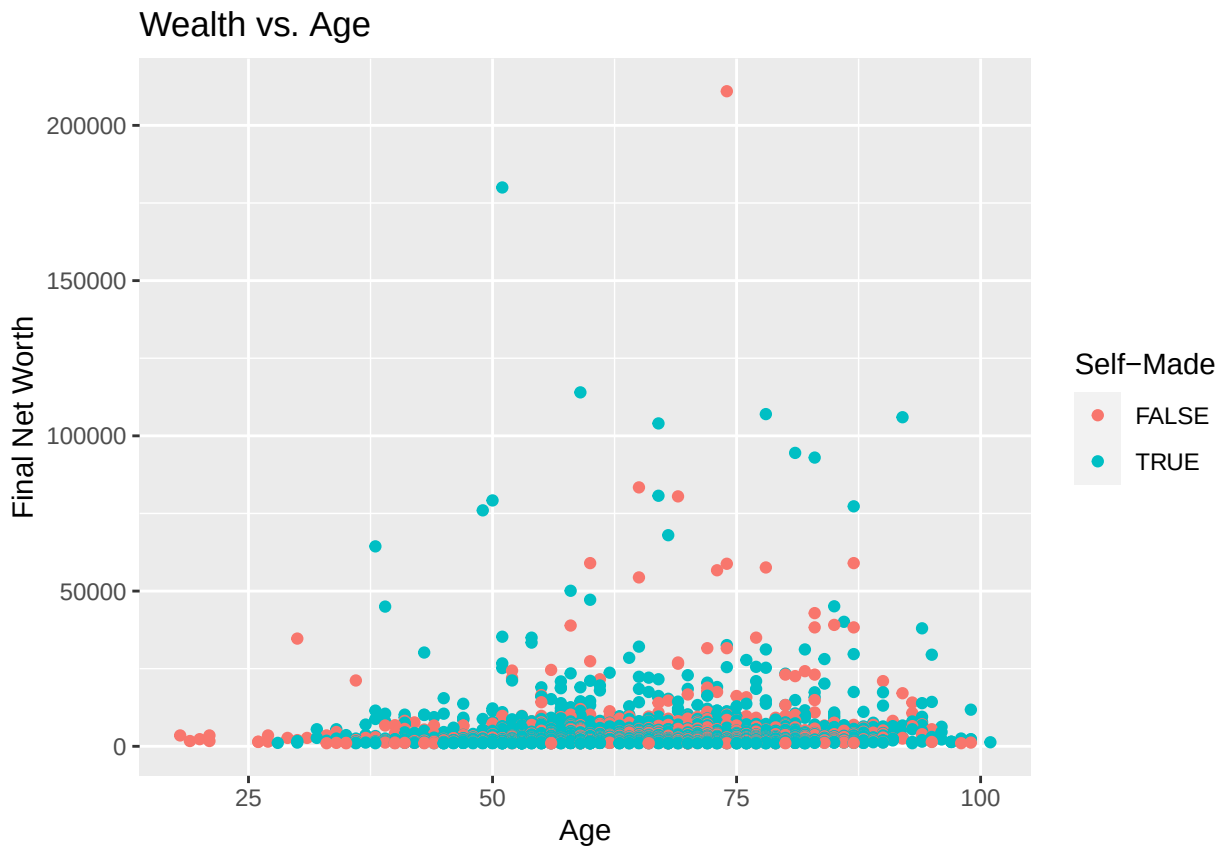
Scatter plot between age and wealth

This plot gives on overview on the generated or inherited wealth by each billionaire indexed by their age. The x-axis demonstrates the age from 1 to 100+ and the y-axis demonstrates the wealth. Blue dots indicate a self-made wealth, while red dots indicate inherited wealth.

```r
# Scatter plot: Wealth vs. Age
ggplot(df, aes(x = age, y = finalWorth, color = selfMade)) +
  geom_point() +
  labs(title = "Wealth vs. Age",
       x = "Age",
```

```
       y = "Final Net Worth",
       color = "Self-Made")
```

```
## Warning: Removed 65 rows containing missing values (`geom_point()`).
```




```
# Remove NA values
df <- na.omit(df)
# Filter for self-made billionaires and find the minimum age
youngest_age <- min(df[df$selfMade == TRUE, ]$age)
# Print the result
cat("The age of the youngest self-made billionaire is", youngest_age, "years.")
```

```
## The age of the youngest self-made billionaire is 28 years.
```

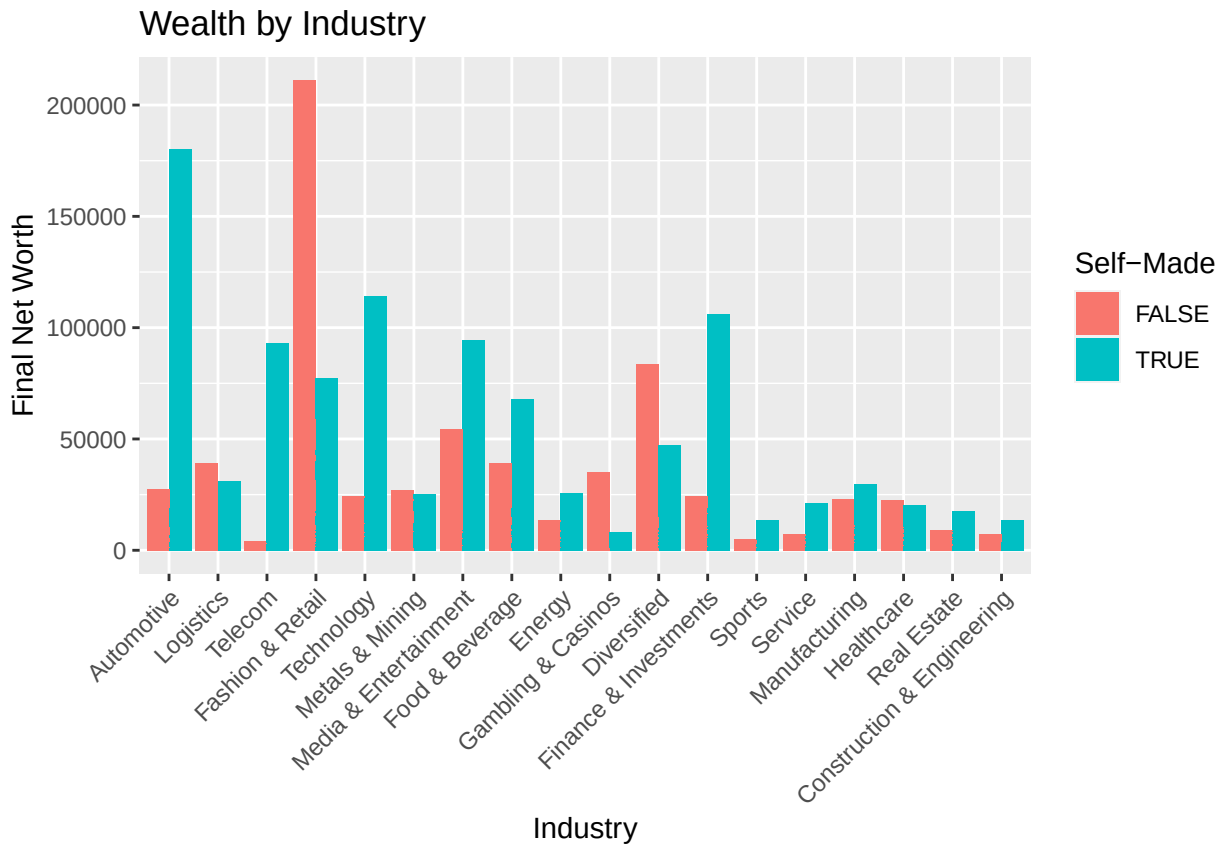
The plot indicates that young billionaires before the age of 28 most definitely inherited their wealth. The youngest self-made billionaire is 28 years old. As the age increases, there is no more pattern. The only thing worth mentioning is that the richest billionaires are mostly self-made. This probably results out of the fact of inflation and the growing gap between rich and poor.

Bar chart wealth by industry

This bar chart shows the wealth distribution over all industries given in the dataset. It also differentiates between self-made wealth and inherited wealth.

```
# Bar chart: Wealth by Industry
ggplot(df, aes(x = reorder(category, -finalWorth), y = finalWorth, fill = selfMade)) +
  geom_bar(stat = "identity", position = "dodge") +
```

```
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  labs(title = "Wealth by Industry",
       x = "Industry",
       y = "Final Net Worth",
       fill = "Self-Made")
```



The billionaires working in Automotive are mainly self-made billionaires while the Fashion & Retail industry is dominated by billionaires that inherited their wealth. Modern industries like Telecom, Automotive, Technology and Media & Entertainment are domiated by self-made billionaires. Most of the wealth by these billionaires comes from the industries Automotive, Telecom, Fashion & Retail, Technology, Diversified, Media & Entertainment, Food & Beverages and Finance & Investments. We should also mention that this data is probably not 100% accurate due to the rough classification of the industry of each billionaire.

Geographical distribution of billionaires on a world map

This does not work because we need a google api key... TODO: F*ck Google die ehrenlosen Hunde

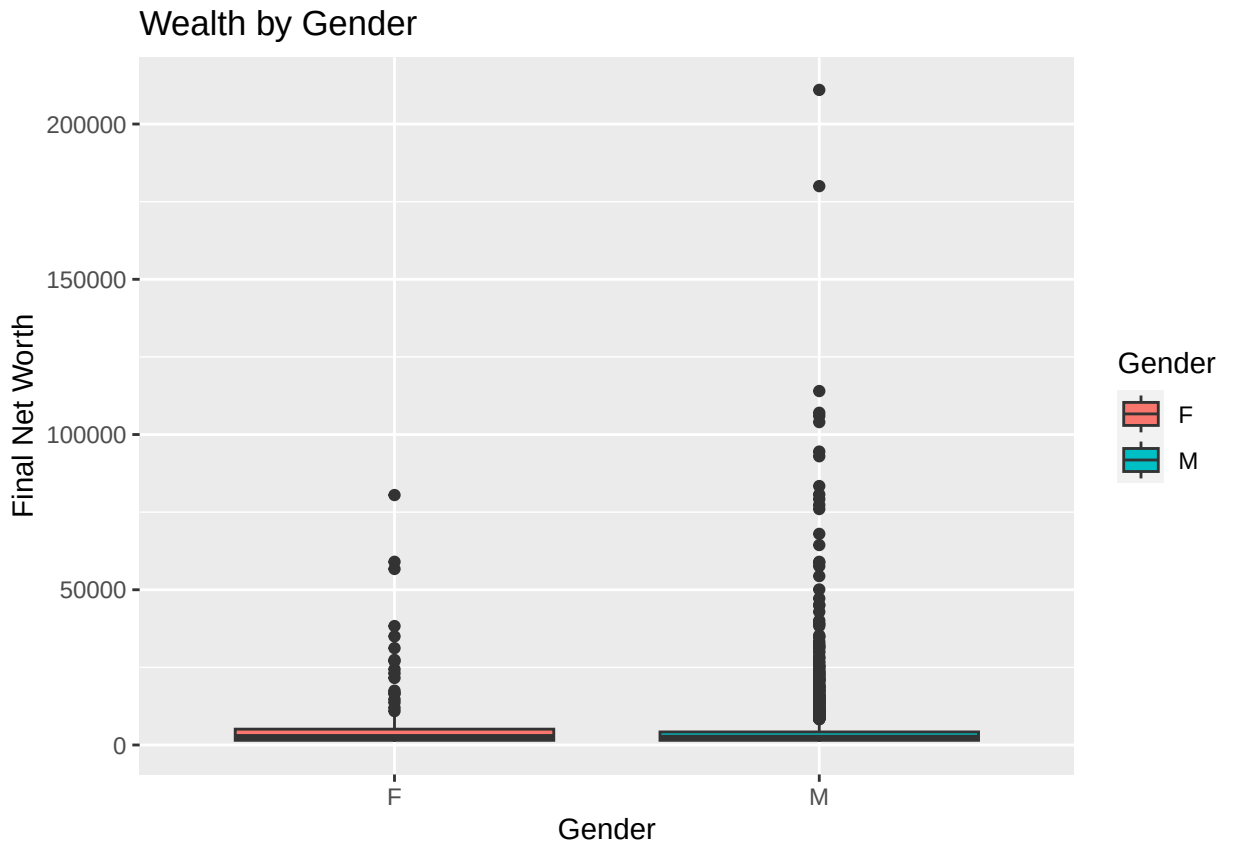
```
# Map: Geographical distribution of billionaires
```

TODO: Here we need to describe the plot that is to generate!

Wealth distribution by gender

This boxplot demonstrates the wealth distribution by biological gender.

```
# Boxplot: Wealth by Gender
ggplot(df, aes(x = gender, y = finalWorth, fill = gender)) +
  geom_boxplot() +
  labs(title = "Wealth by Gender",
       x = "Gender",
       y = "Final Net Worth",
       fill = "Gender")
```



Generally men are richer than women. TODO: ZOOM ON BOXPLOT AND FOCUS ON MAIN STATS INSTEAD OF RICHEST 25%! –> BOXPLOT LOOKS WEIRD

Billionaires wealth distribution by countries

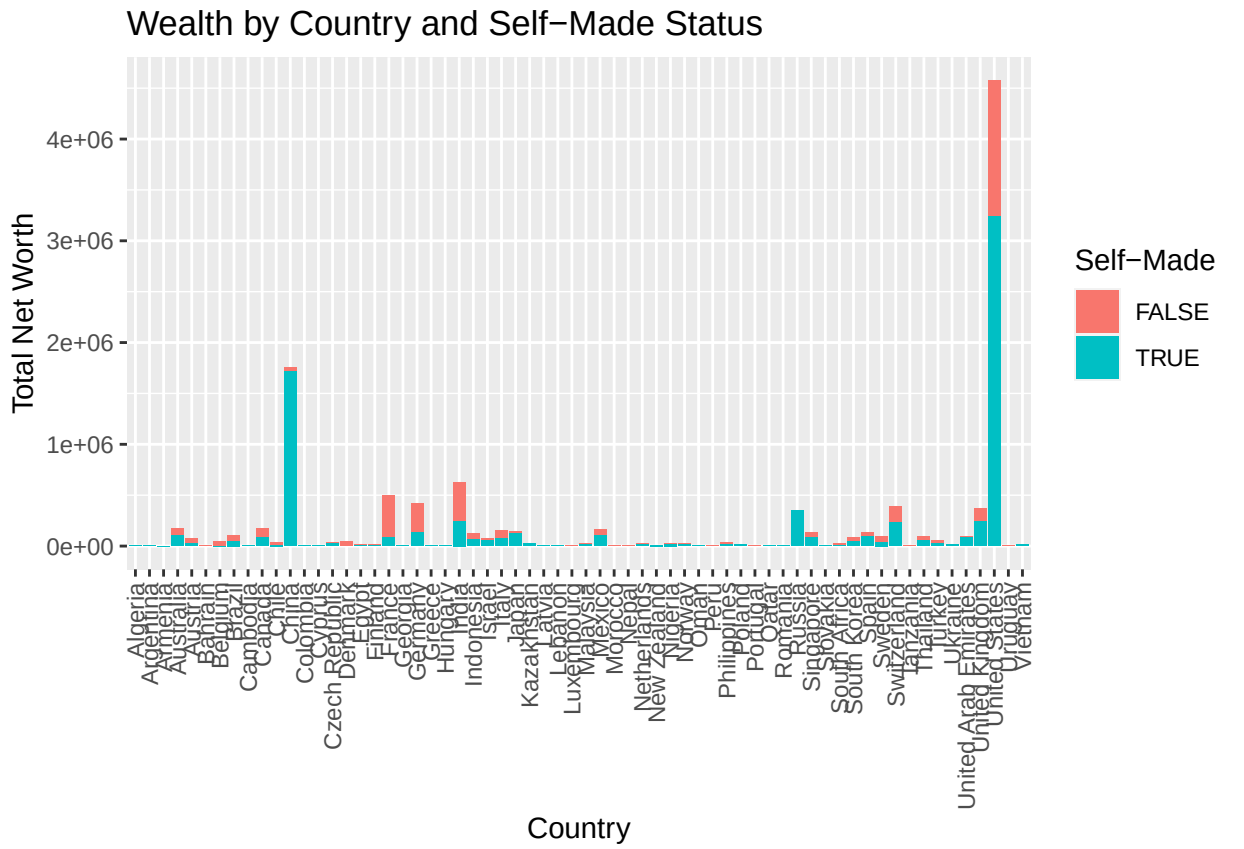
This plot shows the wealth of all billionaires belonging to the country they live in.

```
# Stacked bar chart: Wealth by Country and Self-Made Status
df_country <- df %>% group_by(country, selfMade) %>% summarise(totalWorth = sum(finalWorth))
```

```
## `summarise()` has grouped output by 'country'. You can override using the
## `.groups` argument.
```

```
ggplot(df_country, aes(x = country, y = totalWorth, fill = selfMade)) +
  geom_bar(stat = "identity") +
  labs(title = "Wealth by Country and Self-Made Status",
       x = "Country",
       y = "Total Net Worth",
       fill = "Self-Made") +
```

```
  theme(axis.text.x = element_text(angle = 90, hjust = 1, vjust = 0.5))
```



The most wealth lies in usa and china. Chinese billionaires are mostly self-made. This may be, because China's economy just recently transformed from a largely agrarian and impoverished nation to the world's second-largest economy, driven by their market reforms and globalization. In the U.S. roughly 1/3 of the billionaires inherited their wealth.