Final Project

Shubham Sharma - 22201541

2022-12-02

Part 1: Analysis

SIGI database

For this part, we are using the SIGI (Social Institutions and Gender Index) which has been provided by UNICEF and OECD. The OECD Development Centre's Social Institutions and Gender Index (SIGI) is a cross-country measure of discrimination against women in social institutions (formal and informal laws, social norms, and practices) across 180 countries.

Loading the data and converting the tibble to a dataframe

```
GIDDB2019_02122022151854504=read.csv(file='./GIDDB2019_02122022151854504.csv', header=TRUE)
sigi <- as.data.frame(GIDDB2019_02122022151854504)
```

```
head(sigi)
```

```
##   REGION Region LOCATION   Country INC      Income          VAR  Variable TIME
## 1    ASI   Asia      AUS Australia HIN High income   DF_HR_LAW       Law 2019
## 2    ASI   Asia      AUS Australia HIN High income   DF_HR_ATT Attitudes 2019
## 3    ASI   Asia      AUS Australia HIN High income DF_HR_PRACT  Practice 2019
## 4    ASI   Asia      AUS Australia HIN High income   DF_DV_LAW       Law 2019
## 5    ASI   Asia      AUS Australia HIN High income   DF_IN_LAW       Law 2019
## 6    ASI   Asia      AUS Australia HIN High income RPI_VAW_LAW       Law 2019
##   Year Value Flag.Codes Flags
## 1 2019  0.50         NA    NA
## 2 2019 21.10         NA    NA
## 3 2019  1.82         NA    NA
## 4 2019  0.00         NA    NA
## 5 2019  0.00         NA    NA
## 6 2019  0.75         NA    NA
```

```
summary(sigi)
```

```
##     REGION             Region            LOCATION           Country
##  Length:19676       Length:19676       Length:19676       Length:19676
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##      INC                Income              VAR              Variable
```

```
##  Length:19676       Length:19676       Length:19676       Length:19676
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##       TIME           Year          Value        Flag.Codes      Flags
##  Min.   :2019   Min.   :2019   Min.   :  0.00   Mode:logical   Mode:logical
##  1st Qu.:2019   1st Qu.:2019   1st Qu.:  0.25   NA's:19676     NA's:19676
##  Median :2019   Median :2019   Median :  1.00
##  Mean   :2019   Mean   :2019   Mean   : 16.02
##  3rd Qu.:2019   3rd Qu.:2019   3rd Qu.: 23.20
##  Max.   :2019   Max.   :2019   Max.   :115.80
```

```
dim(sigi)
```

```
## [1] 19676    13
```

We have 19676 values spread across 13 columns

We need to rename Flag Codes to Flag.Codes to process it further, since it has a space on it and might throw an error.

We can see that Flag.Codes and Flags have non-unique nan values. We can look to drop these columns.

```
colnames(sigi)[colnames(sigi) == "Flag Codes"] ="Flag.Codes"
```

Furthermore INC, LOCATION and REGION gives the shortform for Income, Country and Region respectively, and TIME and Year gives the same value, ie 2019. We can look to drop these columns as well

```
unique(sigi$INC)
```

```
## [1] "HIN" "UMI" "LWI" "LOI" "AIC"
```

```
unique(sigi$TIME)
```

```
## [1] 2019
```

```
unique(sigi$REGION)
```

```
## [1] "ASI" "ALL" "EUR" "AME" "AFR"
```

```
unique(sigi$Flags)
```

```
## [1] NA
```

```
unique(sigi$Flag.Codes)
```

```
## [1] NA
```

```
sigi2 <- subset(sigi, select = -c(Flags, Flag.Codes, INC, TIME, REGION, LOCATION))

head(sigi2)
```

```
##   Region   Country      Income         VAR  Variable Year Value
## 1   Asia Australia High income   DF_HR_LAW       Law 2019  0.50
## 2   Asia Australia High income   DF_HR_ATT Attitudes 2019 21.10
## 3   Asia Australia High income DF_HR_PRACT  Practice 2019  1.82
## 4   Asia Australia High income   DF_DV_LAW       Law 2019  0.00
## 5   Asia Australia High income   DF_IN_LAW       Law 2019  0.00
## 6   Asia Australia High income RPI_VAW_LAW       Law 2019  0.75
```

sigi2 contains the following columns:

Region: The continent that the country is part of

Country: The country for which we have the data

Income: The income group type that the country is a part of. It can be one of the following: unique(sigi$Income)

Year: We are looking at the data for which it was last available, ie 2019

VAR: This gives us the indicator that we are estimating.

```
unique(sigi2$VAR)
```

```
##  [1] "DF_HR_LAW"        "DF_HR_ATT"        "DF_HR_PRACT"      "DF_DV_LAW"
##  [5] "DF_IN_LAW"        "RPI_VAW_LAW"      "RPI_VAW_ATT"      "RPI_VAW_PRACT"
##  [9] "RPI_FGM_LAW"      "RPI_MW_PRACT"     "RPI_RA_LAW"       "RPI_RA_PRACT"
## [13] "RAPFR_SAL_LAW"    "RAPFR_SANL_LAW"   "RAPFR_SAFS_LAW"   "RAPFR_SAFS_PRACT"
## [17] "RAPFR_WR_LAW"     "RAPFR_WR_ATT"     "RAPFR_WR_PRACT"   "RCL_CR_LAW"
## [21] "RCL_PV_LAW"       "RCL_PV_ATT"       "RCL_PV_PRACT"     "RCL_FM_LAW"
## [25] "RCL_FM_PRACT"     "RCL_AJ_LAW"       "RCL_AJ_PRACT"     "RAPFR_SAL_PRACT"
## [29] "RAPFR_SANL_PRACT" "RPI_FGM_ATT"      "RPI_FGM_PRACT"    "DF_CM_LAW"
## [33] "DF_CM_PRACT"
```

- Discrimination in the family (DF) including information on marriage customs (age, inheritance, and polygamy) and decisionmaking power within a household (parental authority, repudiation).
  - Child marriage (CM) Whether women and men have the same legal minimum age of marriage
  - Household responsibilities (HR) Whether women and men have the same legal rights, decision-making abilities and responsibilities within the household
  - Inheritance (IN) Whether women and men have the same legal rights to inheritance of land and non-land assets.
  - Divorce (DV) Whether women and men have the same legal rights to initiate divorce and have the same requirements for divorce or annulment.
- Restricted physical integrity (RPI) capturing violence against women through traditional practices such as female genital mutilation or other attacks (e.g. rape, assault, harassment).
  - Violence against women (VAW) Whether the legal framework protects women from violence including intimate partner violence, rape and sexual harassment, without legal exceptions and in a comprehensive approach.
  - Female genital mutilation (FGM)
  - Missing women (MW)

  - Reproductive autonomy (RA)
- Restricted access to productive and financial resources (RAPFR) indicating the quality of women's most basic economic right – to hold property, either in the form of bank loans, land, or other material assets.
  - Secure access to land (SAL) Whether women and men have the same legal rights and secure access to land assets
  - Secure access to non-land assets (SANL) Whether women and men have the same legal rights and secure access to non-land assets
  - Secure access to formal financial services (SAFS) Whether women and men have the same legal rights to open a bank account and obtain credit in a formal financial institution
  - Workplace rights (WR) Whether women and men have the same legal rights and opportunities in the workplace
- Restricted civil liberties (RCL) measuring the extent to which women can participate in social life (e.g. moving freely in public without the obligation to wear a veil or be escorted by male relatives).
  - Citizenship rights (CR) Whether women and men have the same citizenship rights and ability to exercise their rights
  - Freedom of movement (FM) Whether women and men have the same rights to apply for national identity cards (if applicable) and passports and travel outside the country
  - Political voice (PV) Whether the legal framework promotes women's equal political representation as men
  - Access to justice (AJ) Whether women and men have the same rights to provide testimony in court, hold public or political office in the judiciary and sue

Variable: Shows the level of discrimination in laws (Law / LAW), social norms (Attitudes / ATT) and practises (Practise / PRACT) Formal and informal laws, attitudes and practices that restrict women's and girls' access to rights, justice and empowerment opportunities. These are captured in a multi-faceted approach by SIGI's variables that combine qualitative and quantitative data, taking into account both the de jure and de facto discrimination of social institutions, through information on laws, attitudes and practices. The variables span all stages of a woman's life in order to show how discriminatory social institutions can interlock and bind them into cycles of poverty and disempowerment.

Value: 0 for no discrimination to 1 for very high discrimination. The qualitative information detailed in the SIGI country profiles are quantified using the following coding manual:

- 0: The legal framework provides women with the same rights as men, without legal exceptions regarding some groups of women. There is no customary, traditional or religious laws or practices that discriminate against women's rights.
- 0.25: The legal framework provides women with the same rights as men, without legal exceptions regarding some groups of women. However, some customary, traditional or religious laws or practices discriminate against women's rights.
- 0.5: The legal framework provides women with the same rights as men. However, it does not apply to all group of women.
- 0.75: The legal framework restricts some women's rights.
- 1: The legal framework fully discriminates against women's rights. In cases where no or insufficient information exists, variables are not assigned a value. The legal indicators are assessed based on all applicable legal frameworks, including civil law, religious law, customary law and traditional law.

```
head(sigi2[
  with(sigi2, order(Country, VAR)),
])
```

```
##            Region     Country            Income         VAR Variable Year Value
## 18825        Asia Afghanistan        Low income   DF_CM_LAW      Law 2019     1
## 18827 All regions Afghanistan        Low income   DF_CM_LAW      Law 2019     1
## 19077        Asia Afghanistan All income groups   DF_CM_LAW      Law 2019     1
## 19079 All regions Afghanistan All income groups   DF_CM_LAW      Law 2019     1
## 18826        Asia Afghanistan        Low income DF_CM_PRACT Practice 2019    35
## 18828 All regions Afghanistan        Low income DF_CM_PRACT Practice 2019    35
```

Since the values are noted twice, once under all regions/ income groups and once under the individual values, we can discard those values to avoid duplicates.

```
sigi3<-subset(sigi2, Region!="All regions")
sigi3<-subset(sigi3, Income!="All income groups")
```

Taking the look at the cleaned up data

```
head(sigi3)
```

```
##   Region   Country      Income          VAR  Variable Year Value
## 1   Asia Australia High income   DF_HR_LAW       Law 2019  0.50
## 2   Asia Australia High income   DF_HR_ATT Attitudes 2019 21.10
## 3   Asia Australia High income DF_HR_PRACT  Practice 2019  1.82
## 4   Asia Australia High income   DF_DV_LAW       Law 2019  0.00
## 5   Asia Australia High income   DF_IN_LAW       Law 2019  0.00
## 6   Asia Australia High income RPI_VAW_LAW       Law 2019  0.75
```

```
dim(sigi3)
```

```
## [1] 4919    7
```

The cleaned data has been reduced to 4919 rows spread across 7 columns

Including additional libraries

```
library("dplyr")
library(tidyverse)
```

```
#Filtering the dataset by checking gender inequality by law (which varies from 0 to 1), attitude and pr
df_law_vals = sigi3 %>% filter(str_detect(VAR, 'LAW'))
df_pract_vals = sigi3 %>% filter(str_detect(VAR, 'PRACT'))
df_att_vals = sigi3 %>% filter(str_detect(VAR, 'ATT'))

unique(df_pract_vals$VAR)
```

```
##  [1] "DF_HR_PRACT"      "RPI_VAW_PRACT"    "RPI_MW_PRACT"     "RPI_RA_PRACT"
##  [5] "RAPFR_SAFS_PRACT" "RAPFR_WR_PRACT"   "RCL_PV_PRACT"     "RCL_FM_PRACT"
##  [9] "RCL_AJ_PRACT"     "RAPFR_SAL_PRACT"  "RAPFR_SANL_PRACT" "RPI_FGM_PRACT"
## [13] "DF_CM_PRACT"
```

```
unique(df_law_vals$VAR)
```

```
##  [1] "DF_HR_LAW"      "DF_DV_LAW"      "DF_IN_LAW"      "RPI_VAW_LAW"
##  [5] "RPI_FGM_LAW"    "RPI_RA_LAW"     "RAPFR_SAL_LAW"  "RAPFR_SANL_LAW"
##  [9] "RAPFR_SAFS_LAW" "RAPFR_WR_LAW"   "RCL_CR_LAW"     "RCL_PV_LAW"
## [13] "RCL_FM_LAW"     "RCL_AJ_LAW"     "DF_CM_LAW"
```

```
unique(df_att_vals$VAR)
```

```
## [1] "DF_HR_ATT"    "RPI_VAW_ATT"  "RAPFR_WR_ATT" "RCL_PV_ATT"   "RPI_FGM_ATT"
```

```
# shows the following values can be compared across law, attitude and practises:
# df_hr ; rpi_vaw ; rapfr_wr ; rcl_pv ; rpi_fgm

# Filtering missing women around the world
df_mw_vals = sigi3 %>% filter(str_detect(VAR, 'MW'))
unique(df_mw_vals$VAR)
```

```
## [1] "RPI_MW_PRACT"
```

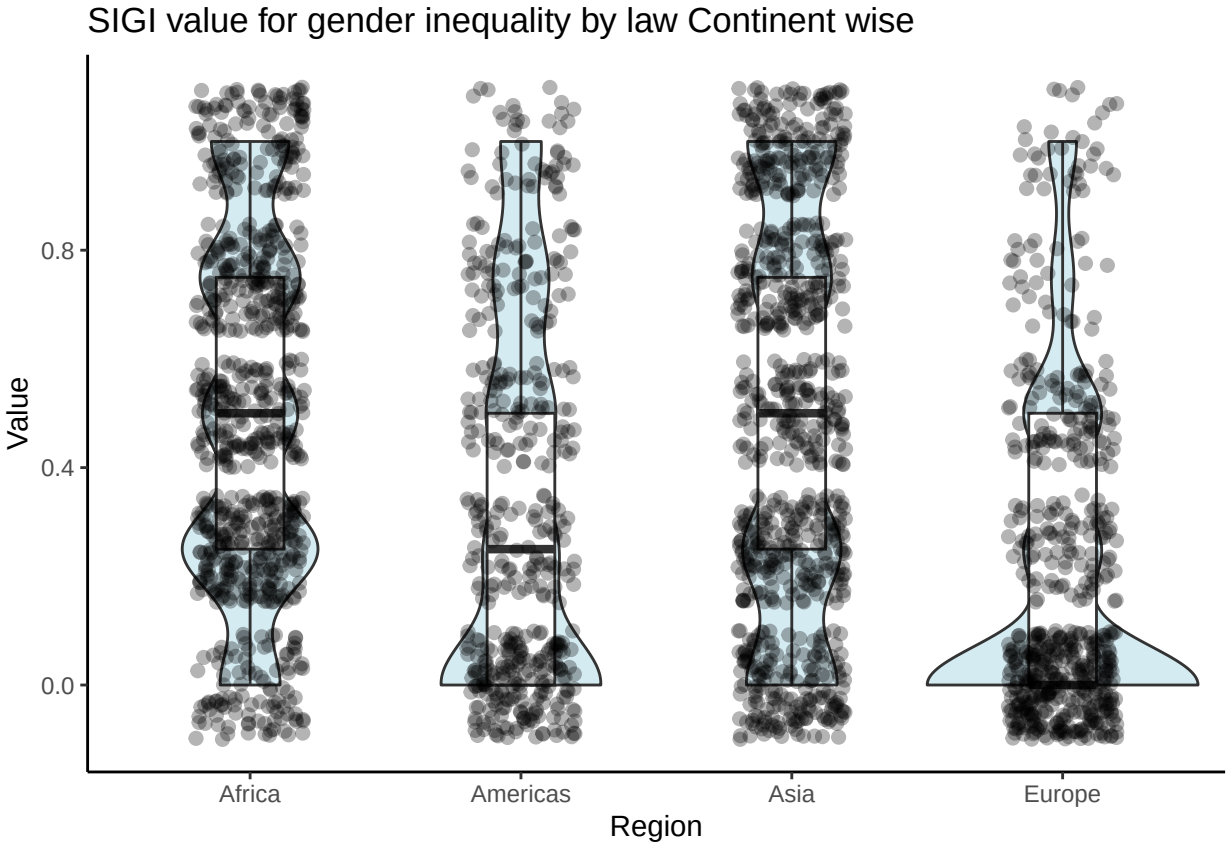
```
# this shows values by practise alone

# Filtering for violence against women across law, practise and attitude
df_rpi_law_vals = sigi3 %>% filter(str_detect(VAR, 'RPI_VAW_LAW'))
df_rpi_pract_vals = sigi3 %>% filter(str_detect(VAR, 'RPI_VAW_PRACT'))
df_rpi_att_vals = sigi3 %>% filter(str_detect(VAR, 'RPI_VAW_ATT'))
```

```
# Taking mean by values, grouping by continent and country
df_law_continent=df_law_vals %>%
  group_by(Region) %>%
  summarise_at(vars(Value), list(name = mean))

df_law_country=df_law_vals %>%
  group_by(Country) %>%
  summarise_at(vars(Value), list(name = mean))

# Taking mean by values, grouping by income groups
df_income=df_law_vals %>%
  group_by(Income) %>%
  summarise_at(vars(Value), list(Value = mean))

# Taking mean by values, grouping by discrimination of women in household responsibilites
df_hr_vals = sigi3 %>% filter(str_detect(VAR, 'DF_HR'))
df_hr=df_hr_vals %>%
  group_by(VAR) %>%
  summarise_at(vars(Value), list(Value = mean))

# Taking mean by values, grouping by gender inequality pertaining to a political voice
df_rcl_vals = sigi3 %>% filter(str_detect(VAR, 'RCL_PV'))
df_rcl=df_rcl_vals %>%
  group_by(VAR) %>%
```

```
  summarise_at(vars(Value), list(Value = mean))

# Taking mean by values, grouping by gender inqeuality in the workplace
df_wr_vals = sigi3 %>% filter(str_detect(VAR, 'RAPFR_WR'))
df_wr=df_wr_vals %>%
  group_by(VAR) %>%
  summarise_at(vars(Value), list(Value = mean))
```

Plotting a boxplot of Region vs Values

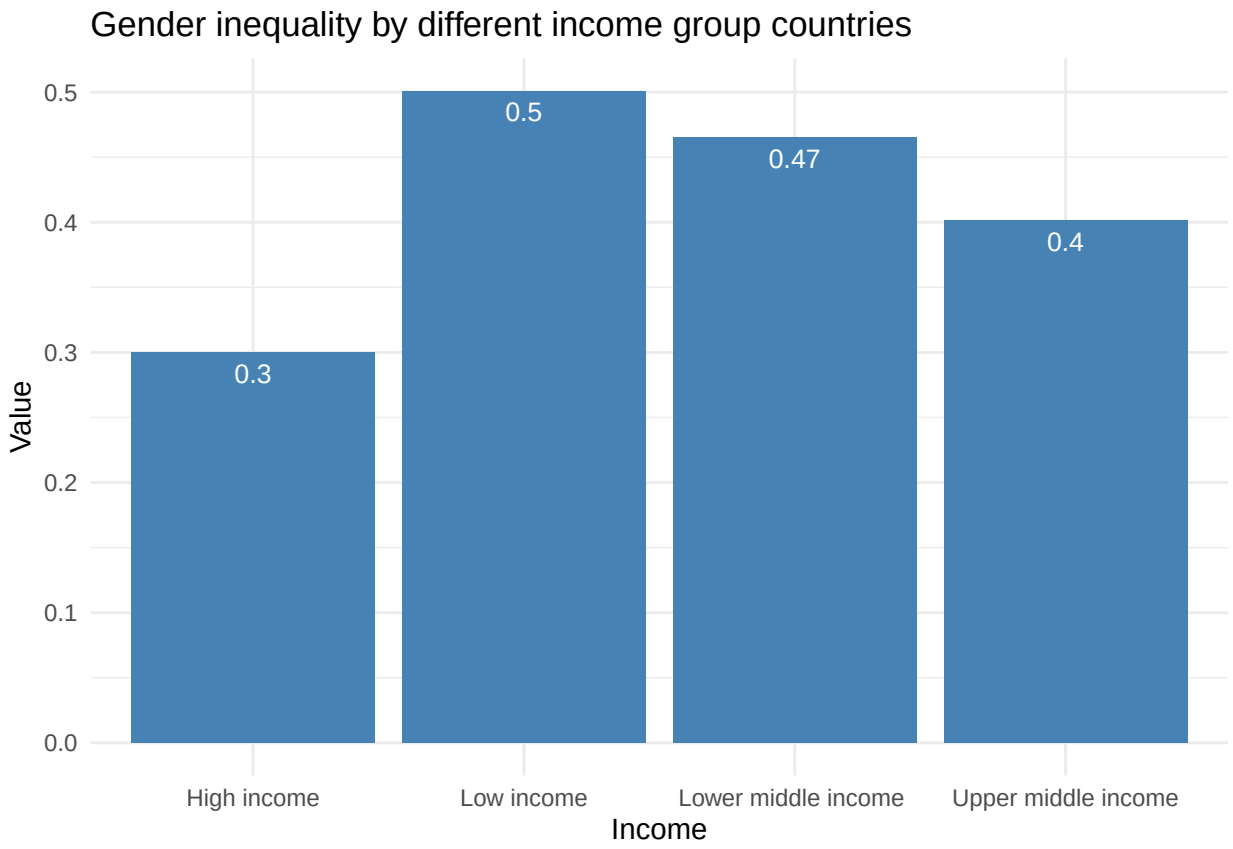
```
ggplot(df_law_vals, aes(x = Region, y = Value)) +
  geom_violin(alpha = 0.5, width = 1, fill = 'lightblue')+
    geom_boxplot(width = 0.25, fatten = 3, width = 0.3)+
      geom_jitter(color="black", size=2, alpha = 0.3, width = 0.2) +
    theme(axis.text.x = element_text(angle = 30, hjust = 1)) +
    ggtitle("SIGI value for gender inequality by law Continent wise") + theme_classic()
```



- Here, the plot depicts the level of discrimination as per the laws laid down, continent wise.
- We can see that Europe and the Americas have relatively lower levels of discrimination due to the cluster of points close to 0.
- Asia has it's values spread across, dictating gender inequality in certain countries more than others.
- Africa has a lots of points clustered around 0.2-0.4, which indicated relatively lower gender inequality, but this is countered by considering the number of points close to 1.

Plotting a bargraph of Income groups vs values

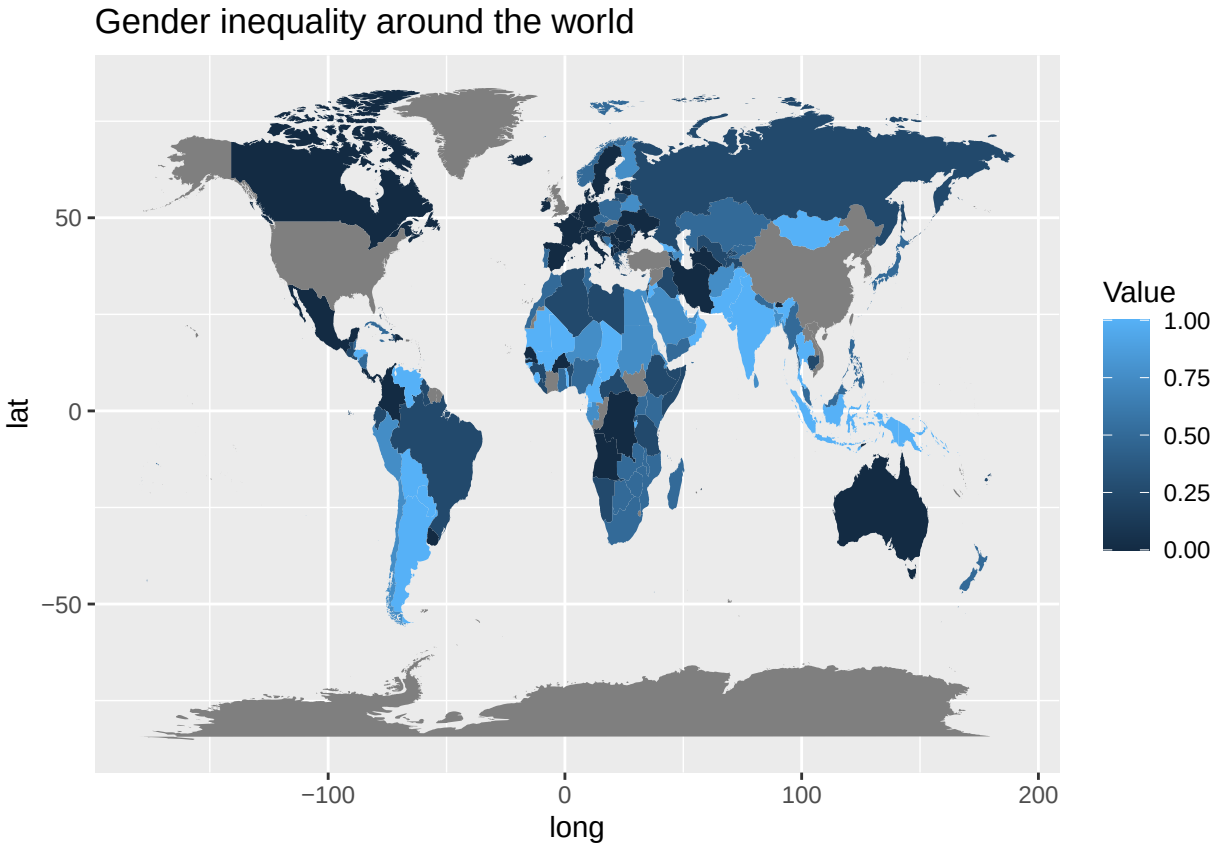
```
ggplot(data=df_income, aes(x=Income, y=Value)) +
  geom_bar(stat="identity", fill="steelblue")+
  geom_text(aes(label=round(Value, digits=2)), vjust=1.6, color="white", size=3.5) +
  ggtitle("Gender inequality by different income group countries") +
  theme_minimal()
```



- The following plot shows the gender inequality as laid down by the law for different strata of countries.
- We can see that high income countries have a relatively lower value of gender inequality than low income countries.
- We can assume a causal relationship, but more insights can be drawn by the F-test to affirm this.
- Furthermore, for high income countries the legal framework provides women with the same rights as men, without legal exceptions regarding some groups of women. However, some customary, traditional or religious laws or practices discriminate against women's rights.
- The lower middle countries, also have a higher level of inequality than the upper middle income countries.
- According to the key provided above, we can confirm that the legal framework provides women with the same rights as men for low income countries. However, it does not apply to all group of women.

Plotting a world heat map of countries vs values

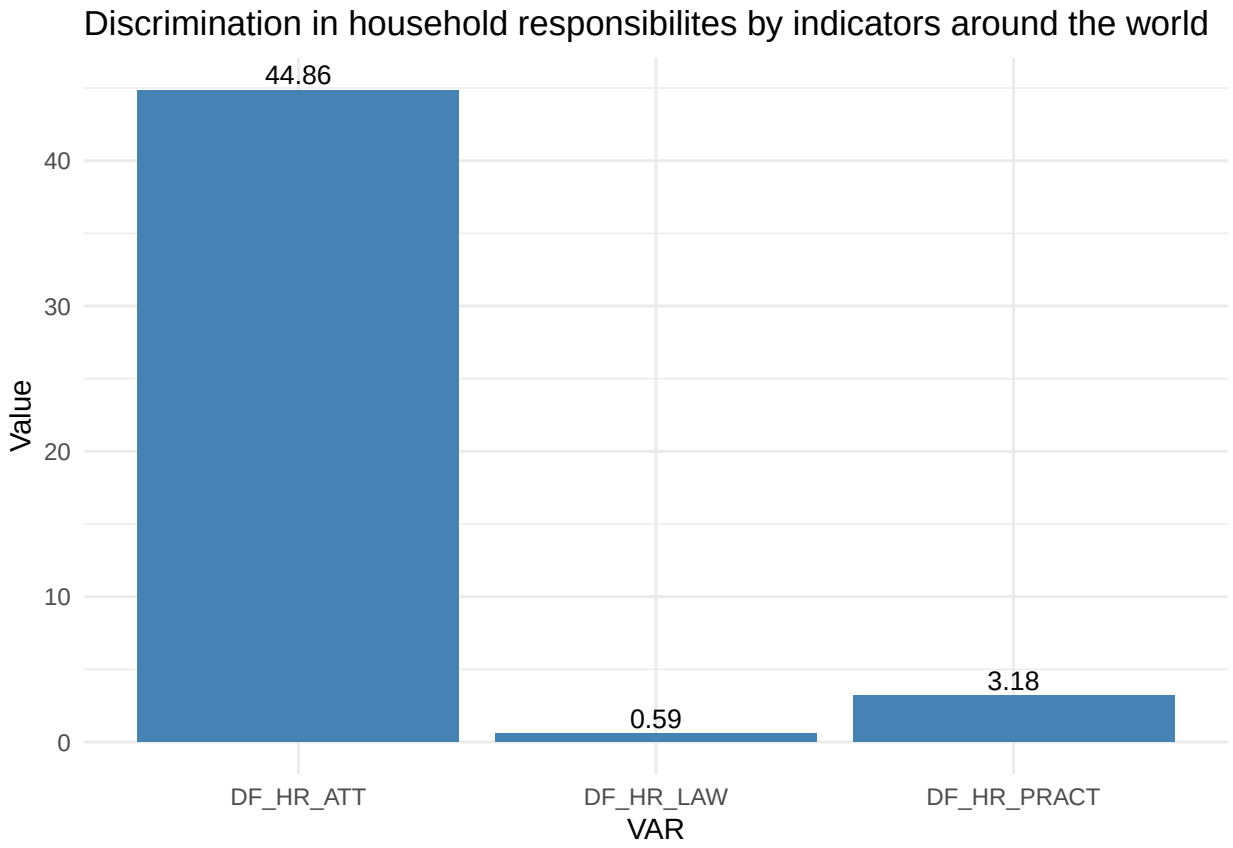
```
world_coordinates <- map_data("world")
world_coordinates %>%
  merge(df_law_vals, by.x = "region", by.y = "Country", all.x = T) %>%
  arrange(group, order) %>%
  ggplot(aes(x = long, y = lat, group = group, fill = Value)) + labs(title="Gender inequality around th
```



- The following plot shows the discrimination against women around the world as laid down by the law across all 4 factors (stated above).
- Lighter blue values indicate greater gender inequality.
- We can see that countries like Australia, Canada, and several European countries have fared well in this regard.
- Certain South American, African and Asian countries (like India, Argentina, Mongolia, Mali etc) haven't performed as well, and have scope for improvement.

Plotting a barplot of discrimination in household responsibilites by indicators around the world

```
  ggplot(data=df_hr, aes(x=VAR, y=Value)) +
geom_bar(stat="identity", fill="steelblue")+
geom_text(aes(label=round(Value, digits=2)), vjust=-0.3, size=3.5) +
ggtitle("Discrimination in household responsibilites by indicators around the world") +
theme_minimal()
```



- We can see that although the law dictates low (0.59) gender inequality, this isn't practised nor are the social norms followed around the world.
- There is still discrimination in the household - whether women and men have the same legal rights, decision-making abilities and responsibilities within the household.
- Although, the value in practise is fairly low, the attitude is higher than expected, and should be mended.

The following countries have fairly large and small values:

```
head(df_hr_vals[order(df_hr_vals$Value),]$Country) # low values
```

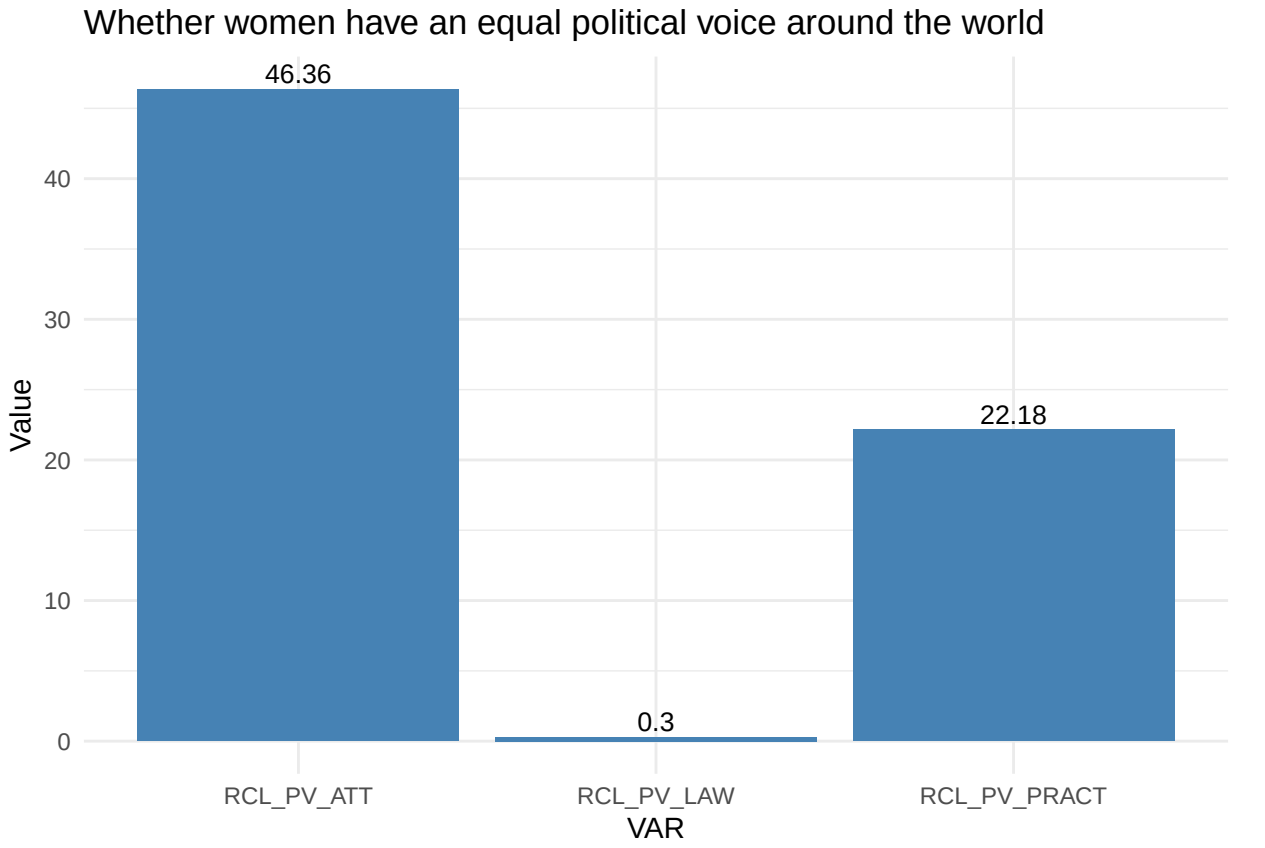
```
## [1] "Switzerland"           "Colombia"              "Japan"
## [4] "Malta"                 "Cabo Verde"            "Sao Tome and Principe"
```

```
tail(df_hr_vals[order(df_hr_vals$Value),]$Country) # high values
```

```
## [1] "India"
## [2] "Tunisia"
## [3] "Qatar"
## [4] "Palestinian Authority or West Bank and Gaza Strip"
## [5] "Yemen"
## [6] "Jordan"
```

Plotting a barplot of gender inequality in having a political voice around the world

```
ggplot(data=df_rcl, aes(x=VAR, y=Value)) +
  geom_bar(stat="identity", fill="steelblue")+
  geom_text(aes(label=round(Value, digits=2)), vjust=-0.3, size=3.5) +
  ggtitle("Whether women have an equal political voice around the world") +
  theme_minimal()
```



- We can see that although the law dictates very low gender inequality across the world (0.29), this isn't practised nor are the social norms followed.
- Both in practise and in attitude, the legal framework doesn't promotes women's equal political representation as men.
- The attitude shown is twice as worse as the practises followed.

The following countries have fairly large and small values:

```
head(df_rcl_vals[order(df_rcl_vals$Value),]$Country) # low values
```

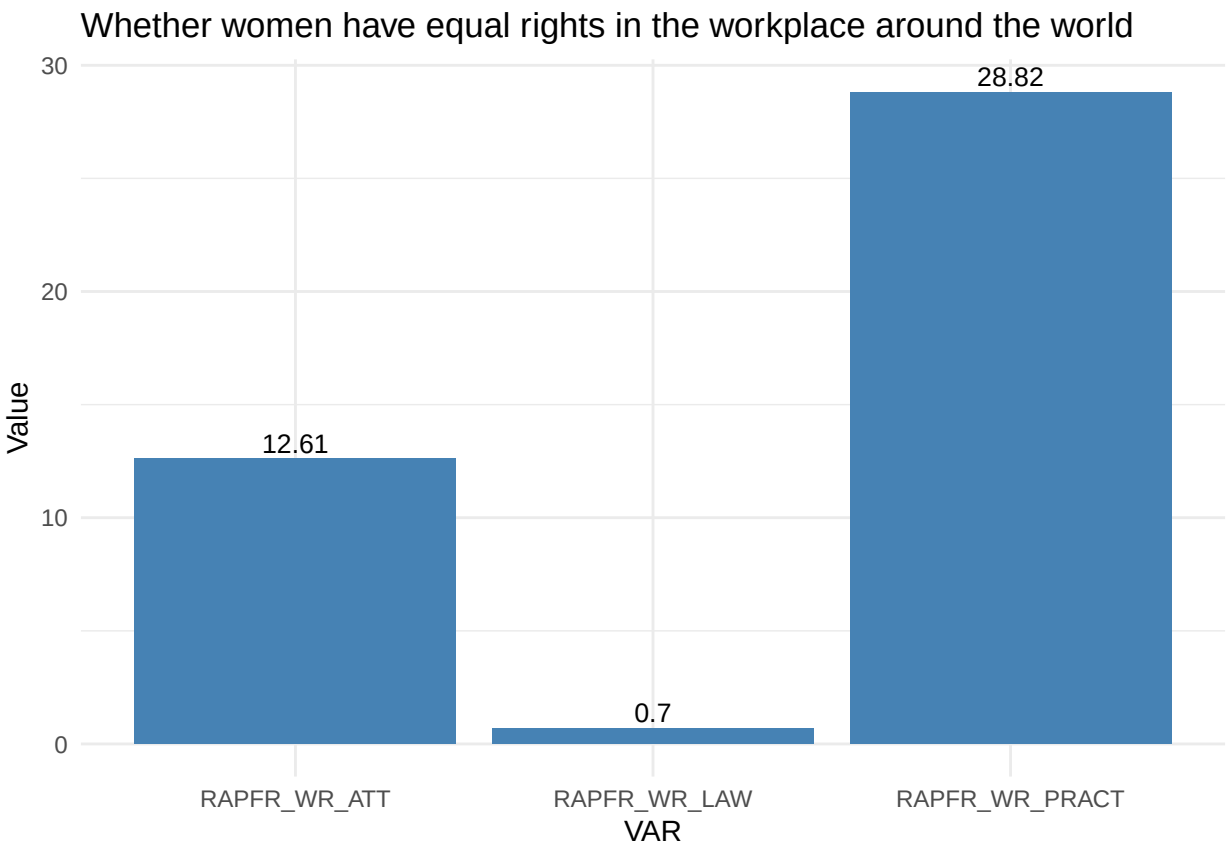
```
## [1] "Belgium" "Denmark" "Finland" "France"  "Iceland" "Ireland"
```

```
tail(df_rcl_vals[order(df_rcl_vals$Value),]$Country) # high values
```

```
## [1] "Palestinian Authority or West Bank and Gaza Strip"
## [2] "Jordan"
## [3] "Ghana"
## [4] "Yemen"
## [5] "Qatar"
## [6] "Egypt"
```

Plotting a barplot of gender inqueality in the workplace around the world

```
ggplot(data=df_wr, aes(x=VAR, y=Value)) +
  geom_bar(stat="identity", fill="steelblue")+
  geom_text(aes(label=round(Value, digits=2)), vjust=-0.3, size=3.5) +
  ggtitle("Whether women have equal rights in the workplace around the world") +
  theme_minimal()
```



According to the law (0.69), the legal framework restricts some women's rights when to comes to the workplace around the world. Herein, the practises followed are twice as worse as the attitude or the social norms dictates.

The following countries have fairly large and small values:

```
head(df_wr_vals[order(df_wr_vals$Value),]$Country) # low values
```

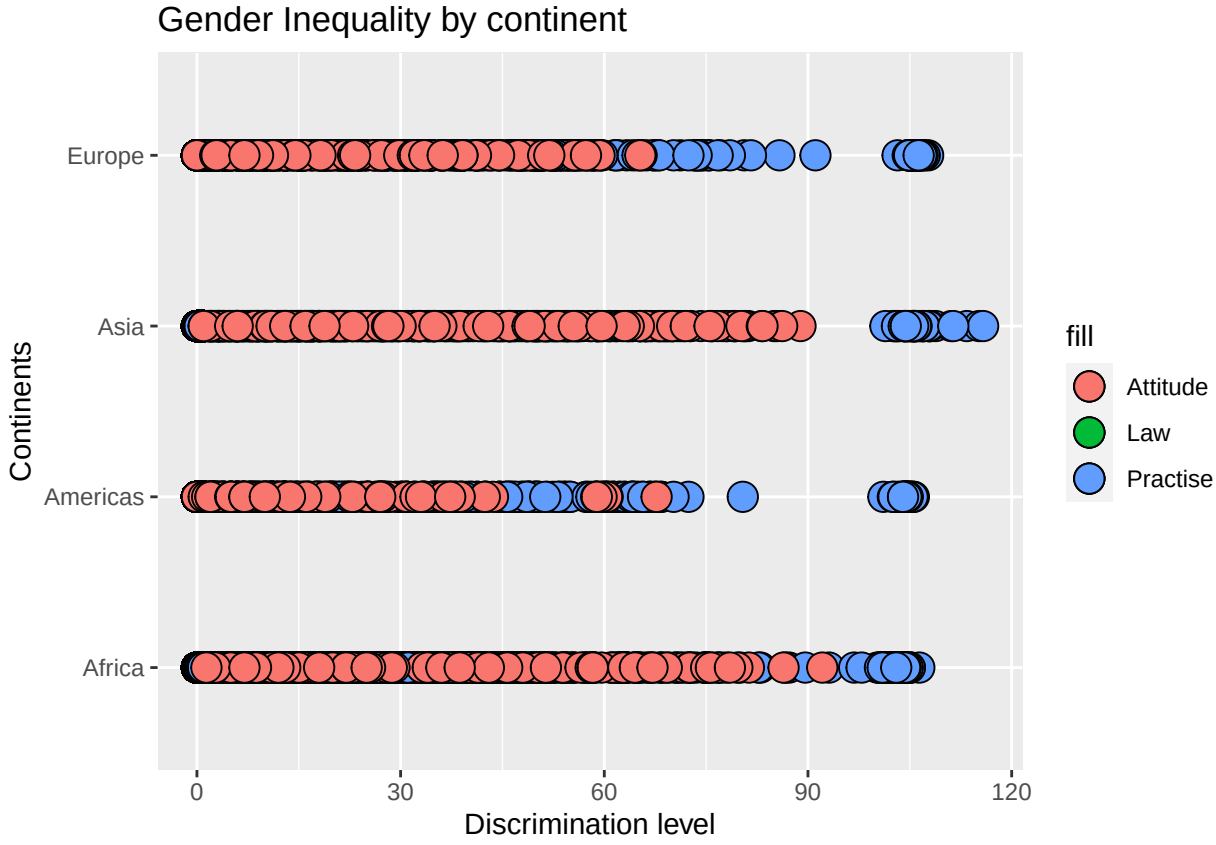
```
## [1] "Austria" "Belgium" "Canada"  "Canada"  "Denmark" "Finland"
```

```
tail(df_wr_vals[order(df_wr_vals$Value),]$Country) # high values
```

```
## [1] "Yemen"       "Philippines" "Bahamas"     "Colombia"    "Jamaica"
## [6] "Pakistan"
```

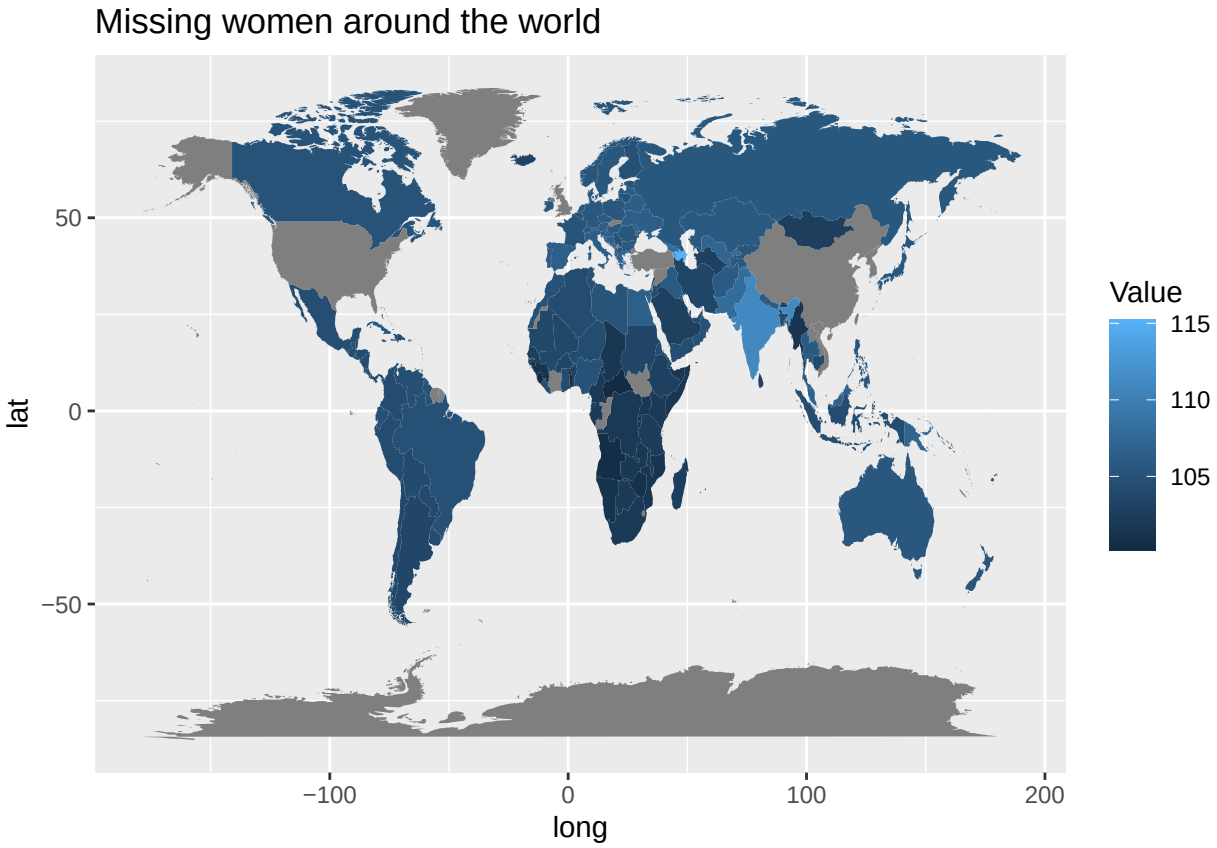
Scatter plot showing the gender inequality by continent and variables

```
ggplot() +
  geom_point(data = df_law_vals, aes(Value, Region, fill='Law'),
             size = 5, shape = 21)+
  geom_point(data = df_pract_vals, aes(Value, Region, fill='Practise'),
             size = 5, shape = 21)+
  geom_point(data = df_att_vals, aes(Value, Region, fill='Attitude'),
             size = 5, shape = 21)+
    scale_color_manual(
    values = c("red", "blue", "green")
  ) +
  labs(x = "Discrimination level", y = "Continents")+
  ggtitle("Gender Inequality by continent")
```



- The following scatter plot shows the gender inequality around the world, grouped by continents and colored by the variable.
- We can see that in general, inequality fairs worse in practise than in attitude.
- This is more so in Asia where there is a clear gap between practise and attitude points.
- Even certain practises in the Americas fair worse than the attitude against women.
- The law points are spread across 0 and 1, and can't be visually inferred via this plot.

```
world_coordinates %>%
  merge(df_mw_vals, by.x = "region", by.y = "Country", all.x = T) %>%
  arrange(group, order) %>%
  ggplot(aes(x = long, y = lat, group = group, fill = Value)) + labs(title="Missing women around the wo
```



- The following world map shows the number of missing women around the world.
- We can see that there is relatively higher number of missing women in and around India as compared to the rest of the world.
- Since this is an absolute number, we can draw a causal relationship to the population of this region. Since this region is densily populated, we can expect higher cases of missing women here than the rest of the world.

Countries with high cases of missing women

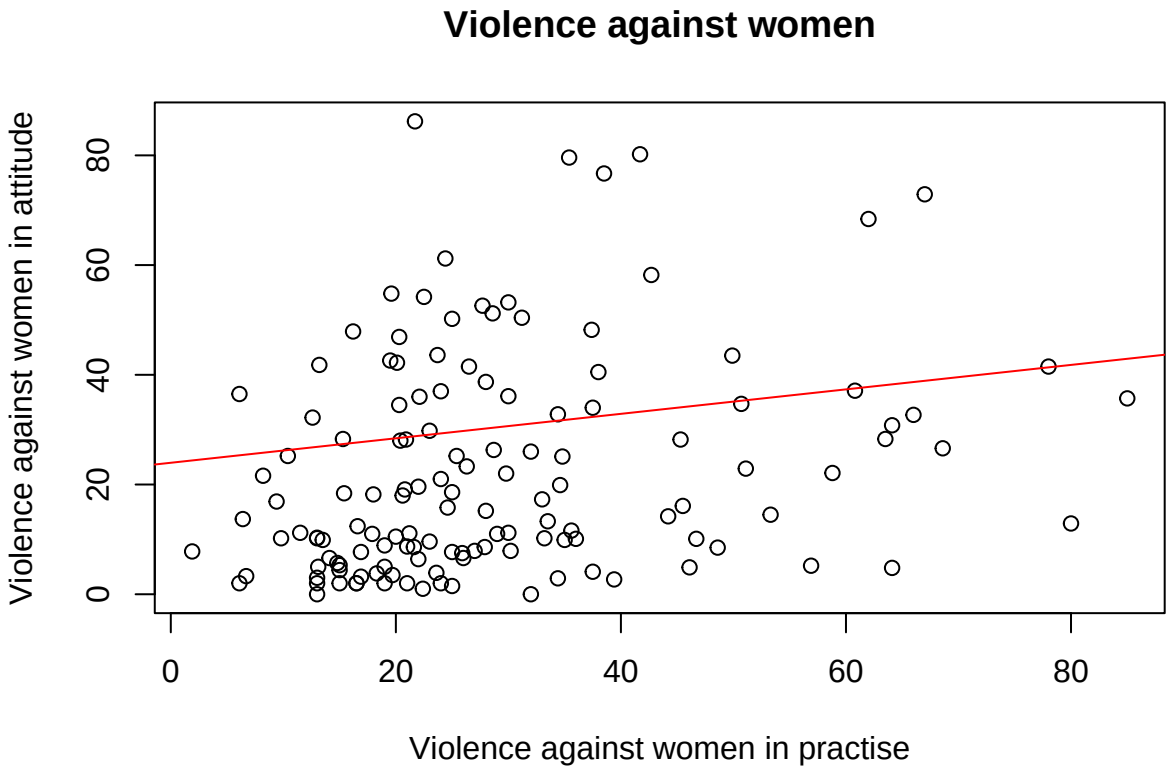
```
df_mw_vals_high = filter(df_mw_vals, Value>110)
df_mw_vals_high$Country
```

```
## [1] "Armenia"                      "Azerbaijan"
## [3] "China (People's Republic of)" "India"
## [5] "Viet Nam"
```

```
head(df_mw_vals[order(df_mw_vals$Value),]$Country) # low values
```

```
## [1] "Central African Republic" "Sierra Leone"
## [3] "Angola"                   "Rwanda"
## [5] "Togo"                     "Guinea-Bissau"
```

```
plot(df_rpi_pract_vals$Value, df_rpi_att_vals[1:nrow(df_rpi_pract_vals),]$Value,
xlab="Violence against women in practise", ylab="Violence against women in attitude",
   main="Violence against women",)
abline(lm(df_rpi_pract_vals$Value ~ df_rpi_att_vals[1:nrow(df_rpi_pract_vals),]$Value), col="red")
```



We can see that there is a mild linear correlation for violence against women in practise and in attitude. There is a cluster in the bottom left (below 20), however certain outliers lie in both practise and attitude.

In conclusion, from the SIGI dataset provided by OECD, we can infer that there is significant data to infer that gender inequality exists across several domains in law, attitude and practise around the world.

Part 2: R Package

OpenCV : CRAN link

OpenCV is a highly optimized library to experiment with computer vision algorithms and machine learning in R with focus on real-time applications. This package exposes some of the available OpenCV algorithms, such as edge, body or face detection. These can either be applied to analyze static images, or to filter live video footage from a camera device. OpenCV is open source and released under the Apache 2 License. It is free for commercial use. In RSudio IDE the image objects will automatically be displayed in the viewer pane. Real life application include but are not limited to perform tasks like face detection, objection tracking, landmark detection, and much more.

Installing and loading the package:

```
install.packages("opencv")
install.packages("jpeg")
```

OpenCV building blocks:

- ocv_read and ocv_write to load/save images on disk
- ocv_picture and ocv_video to use your webcam.

Reading an image of the winners of the world cup, Argentina group photo