DPSS 2022: International Policy Capstone Project

Hieu Nguyen

2022-07-25

Note: All the necessary libraries and packages used in this project can be found in the first part of the corresponding attached R script. They were not included in this R markdown knitted PDF document because I have hidden the chunk containing them (for ease of viewing). Thank you.

Quesion 1: Exploratory Data Analysis

Part (a)

```
# import data from library "unvotes"
un_votes <- un_votes
```

Part (b)

```
# rcid/issue count
no_issues <- n_distinct(un_votes$rcid) # 6202 distinct issues
# yes, no, abstain count
no_votes <- un_votes %>%
  group_by(vote) %>%
  count()
no_votes
```

```
## # A tibble: 3 x 2
## # Groups:   vote [3]
##   vote          n
##   <fct>     <int>
## 1 yes      693544
## 2 abstain  110893
## 3 no        65500
```

Part (c)

```
# create two columns from groups of countries (and filter)
un_votes_over100 <- un_votes %>%
  group_by(country) %>%
  summarise(percent_yes=round(mean(vote=="yes")*100, digits=2), n_votes=n()) %>%
  filter(n_votes>100)

# new dataset: countries that have voted more than 100 times
un_votes_over100
```

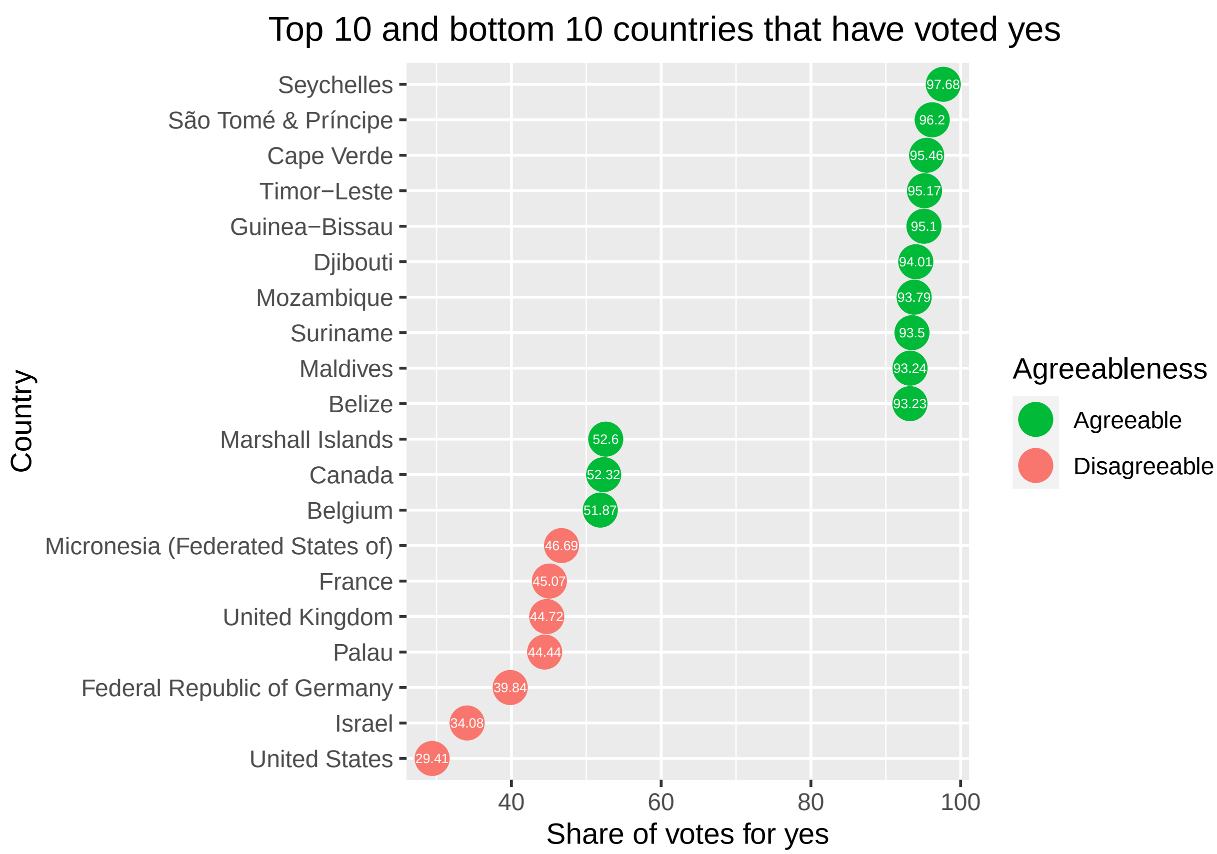
```
## # A tibble: 199 x 3
##    country           percent_yes n_votes
##    <chr>                   <dbl>   <int>
##  1 Afghanistan              85.3    5604
##  2 Albania                  71.0    4237
##  3 Algeria                  88.2    5289
##  4 Andorra                  66.2    2323
##  5 Angola                   91.9    3739
##  6 Antigua & Barbuda        91.6    3344
##  7 Argentina                79.1    6132
##  8 Armenia                  75.8    2361
##  9 Australia                56.2    6166
## 10 Austria                  64.5    5709
## # ... with 189 more rows
```

```
# top 10 and bottom 10 for percent_yes
un_votes_10 <- un_votes_over100 %>%
  arrange(desc(percent_yes)) %>%
  slice(c(1:10), c(n()-9):n())

# new column "agreeableness"
un_votes_10$agreeableness <- ifelse(un_votes_10$percent_yes>50,"Agreeable","Disagreeable")

# final dataset used for plotting
un_votes_10
```

```
## # A tibble: 20 x 4
##    country                          percent_yes n_votes agreeableness
##    <chr>                                  <dbl>   <int> <chr>
##  1 Seychelles                              97.7    2109 Agreeable
##  2 São Tomé & Príncipe                     96.2    2686 Agreeable
##  3 Cape Verde                              95.5    3941 Agreeable
##  4 Timor-Leste                             95.2    1387 Agreeable
##  5 Guinea-Bissau                           95.1    3595 Agreeable
##  6 Djibouti                                94.0    4073 Agreeable
##  7 Mozambique                              93.8    4152 Agreeable
##  8 Suriname                                93.5    4049 Agreeable
##  9 Maldives                                93.2    4644 Agreeable
## 10 Belize                                  93.2    3115 Agreeable
## 11 Marshall Islands                        52.6    2154 Agreeable
## 12 Canada                                  52.3    6176 Agreeable
## 13 Belgium                                 51.9    6161 Agreeable
## 14 Micronesia (Federated States of)        46.7    1962 Disagreeable
## 15 France                                  45.1    6084 Disagreeable
## 16 United Kingdom                          44.7    6140 Disagreeable
## 17 Palau                                   44.4    1366 Disagreeable
## 18 Federal Republic of Germany             39.8    2151 Disagreeable
## 19 Israel                                  34.1    5681 Disagreeable
## 20 United States                           29.4    6155 Disagreeable
```

```
# theme_set(theme_bw())
# plotting
ggplot(un_votes_10, aes(percent_yes, reorder(country, percent_yes), label = percent_yes)) +
```

```
geom_point(stat="identity", aes(col = agreeableness), size=5.5) +
geom_text(aes(label=percent_yes), size=1.75, color="white") +
scale_color_manual(name="Agreeableness",
                   labels=c("Agreeable", "Disagreeable"),
                values=c("Agreeable"="#00ba38","Disagreeable"= "#f8766d"))+
labs(title="Top 10 and bottom 10 countries that have voted yes") +
ylab("Country")+
xlab("Share of votes for yes")+
theme(plot.title = element_text(hjust=0.6))
```



We can see that, on the Country-axis, for the top 10 countries from Seychelles to Belize, the percentages for "yes" votes were very large, above 90 percent. And But until about the threshold of 55 percent can we see other countries that were still considered likely to be agreeable in the voting sessions. This shows the likelihood of having countries with large share of votes for "yes" is high. Meanwhile, the bottom 10 countries were consisted of a mixture of agreeable and disagreeable countries. Again, this implies that there were not many countries that posed an opposition to a voting session. Note that, most of the disagreeable countries were powerhouse (France, United Kingdom, Federal Republic of Germany, and United States).

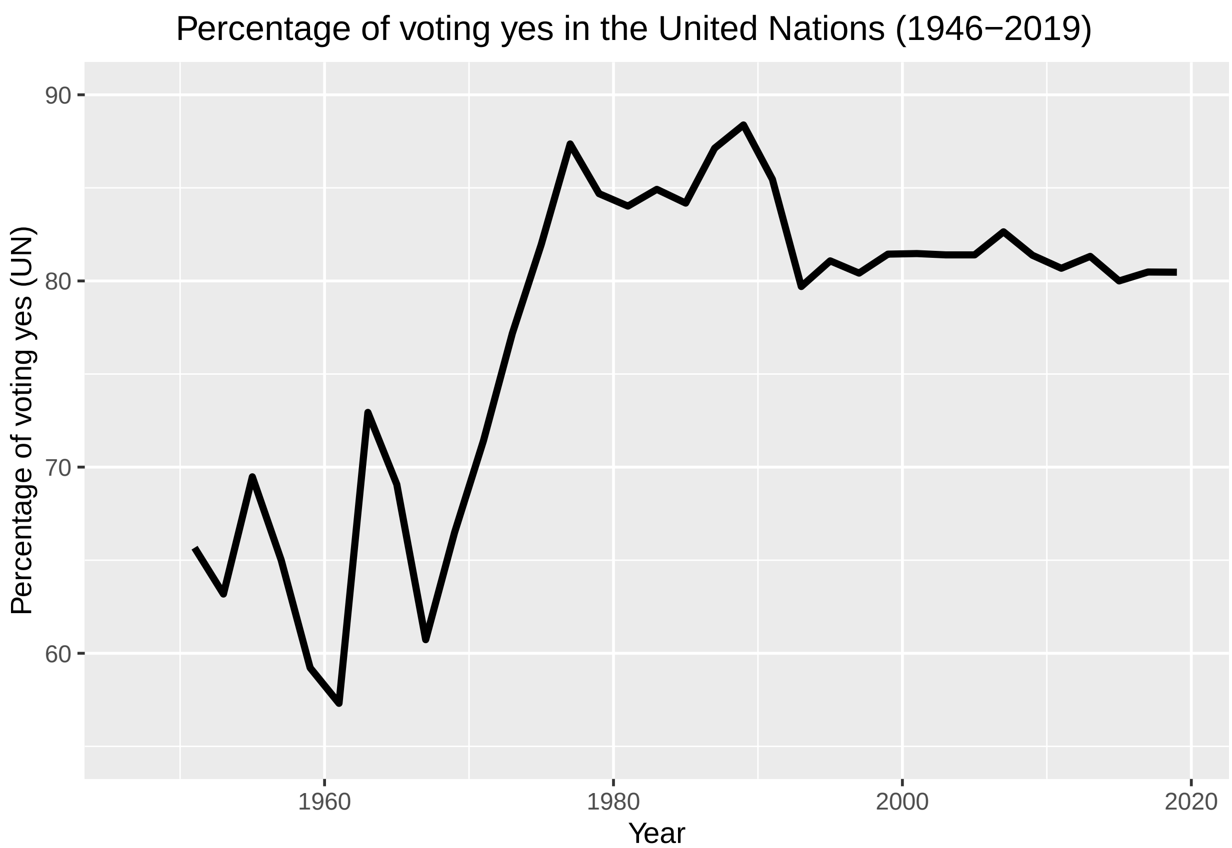
Question 2: Percentage of "yes" votes in the United Nations (1946-2019)

```r
# importing "un_roll_calls" from library "unvotes"
un_roll_calls <- un_roll_calls

# join "un_roll_call" with "un_votes"
q2a_joined <- merge(un_roll_calls, un_votes, by="rcid")
```

```r
# generate year column and percent_yes column
q2a_joined_year <- q2a_joined

q2a_joined_year$year <- format(as.Date(q2a_joined_year$date, format="%Y/%m/%d"), "%Y")
q2a_joined_year$year <- as.integer(q2a_joined_year$year)

q2a_joined_year <- q2a_joined_year %>%
  group_by(year) %>%
  filter(year%%2==1) %>%   # filtering odd years (for the sake of visualization)
  summarise(percent_yes=round(mean(vote=="yes")*100, digits=2), n_votes=n()) %>%
  filter(n_votes>100)

# new dataset used for plotting
q2a_joined_year
```

```
## # A tibble: 37 x 3
##     year percent_yes n_votes
##    <int>       <dbl>   <int>
##  1  1947        56.9    2039
##  2  1949        42.5    5700
##  3  1951        65.7     402
##  4  1953        63.2    1537
##  5  1955        69.5    2169
##  6  1957        65      4092
##  7  1959        59.2    4485
##  8  1961        57.3   10178
##  9  1963        72.9    3308
## 10  1965        69.1    4494
## # ... with 27 more rows
```

```r
# plotting the time-series (1946-2019)
ggplot(q2a_joined_year, aes(year, percent_yes, group=1)) +
  geom_line(size=1.25) +
  scale_y_continuous(limits=c(55,90)) +
  scale_x_continuous(limits=c(1947,2019)) +
  labs(title="Percentage of voting yes in the United Nations (1946-2019)") +
  ylab("Percentage of voting yes (UN)")+
  xlab("Year")+
  theme(plot.title = element_text(hjust=0.4))
```


Percentage of voting yes in the United Nations (1946−2019)
Percentage of voting yes (UN)
90
80
70
60
1960
1980
2000
2020
Year

Question 3: Examining major countries

Part(a): process data

```
# importing new dataset
un_roll_call_issues <- un_roll_call_issues

# joining un_roll_call_issues and q2a_joined
q3a_joined <- merge(un_roll_call_issues, q2a_joined, by="rcid")

# list the issues (for my own reference)
un_roll_call_issues %>% group_by(issue) %>% count()
```

```
## # A tibble: 6 x 2
## # Groups:   issue [6]
##   issue                                    n
##   <fct>                                <int>
## 1 Colonialism                            957
## 2 Arms control and disarmament          1092
## 3 Economic development                   765
## 4 Human rights                          1015
## 5 Palestinian conflict                  1061
## 6 Nuclear weapons and nuclear material   855
```

```
# filtering countries that voted  more than 10 times
filtering_countries_10 <- q3a_joined %>%
  group_by(country) %>%
  summarise(total_vote = n()) %>%
  filter(total_vote>10)

# final data for part (a)
q3a_final <- q3a_joined %>%
  filter(issue!="NA") %>%
  filter(country %in% filtering_countries_10$country)

# we can also see that the "issue" column does not have any "NA" values if we use:
# sapply(q3a_joined, function(x) sum(is.na(x)))
```

Part (b): calculate percent_yes for 5 countries by 6 issues, then plot

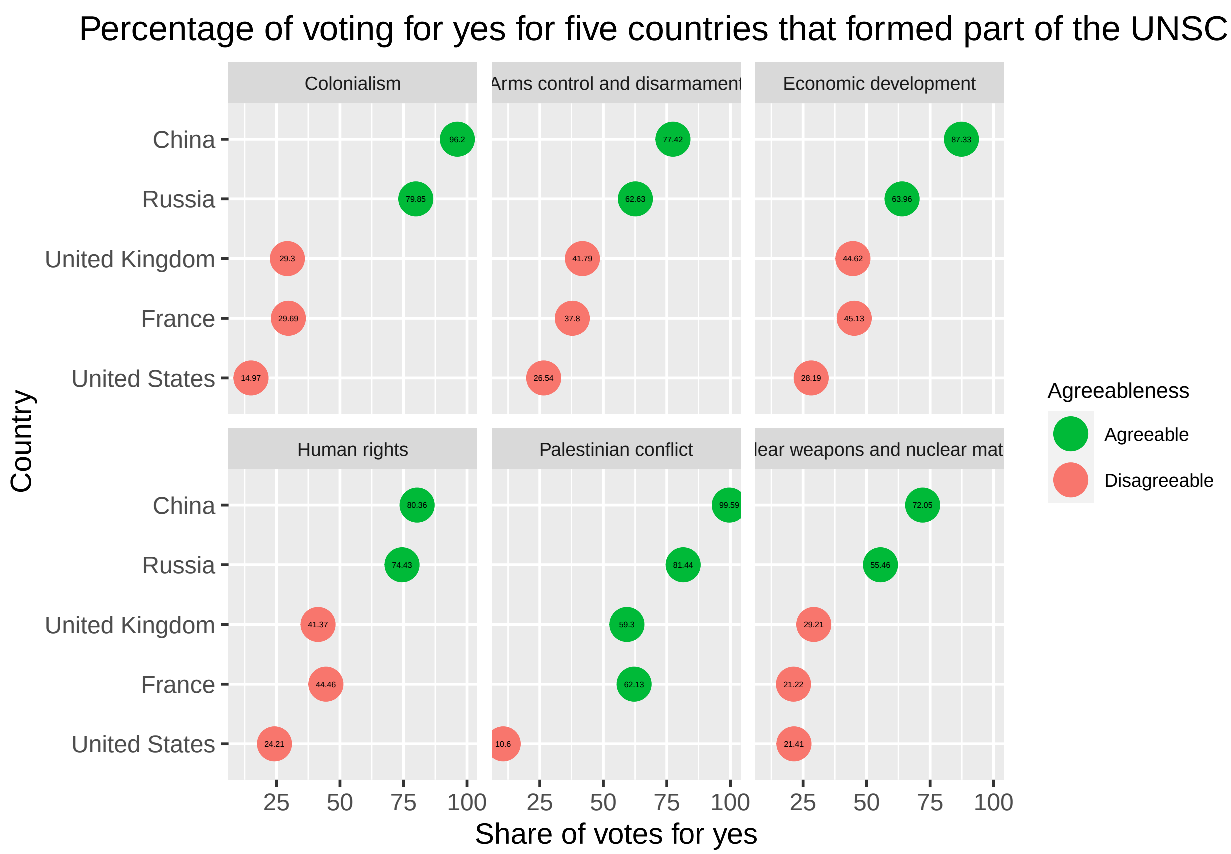
```
# calculate percent_yes by 5 countries, for each of 6 issues
q3b_plot <- q3a_final %>%
  group_by(country, issue) %>%
  summarise(percent_yes=round(mean(vote=="yes")*100, digits=2), n_votes=n()) %>%
  filter(country %in% c("United States", "United Kingdom", "France", "China", "Russia"))
```

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

```
# add agreeableness column
q3b_plot$agreeableness <- ifelse(q3b_plot$percent_yes>50,"Agreeable","Disagreeable")

# final dataset for plotting
q3b_plot
```

```
## # A tibble: 30 x 5
## # Groups:   country [5]
##    country issue                                percent_yes n_votes agreeableness
##    <chr>   <fct>                                      <dbl>   <int> <chr>
##  1 China   Colonialism                                 96.2     658 Agreeable
##  2 China   Arms control and disarmament                77.4     899 Agreeable
##  3 China   Economic development                        87.3     576 Agreeable
##  4 China   Human rights                                80.4     881 Agreeable
##  5 China   Palestinian conflict                        99.6     967 Agreeable
##  6 China   Nuclear weapons and nuclear materi~         72.0     712 Agreeable
##  7 France  Colonialism                                 29.7     933 Disagreeable
##  8 France  Arms control and disarmament                37.8    1082 Disagreeable
##  9 France  Economic development                        45.1     760 Disagreeable
## 10 France  Human rights                                44.5    1001 Disagreeable
## # ... with 20 more rows
```

```r
# plotting
ggplot(q3b_plot, aes(percent_yes, reorder(country, percent_yes))) +
  geom_point(stat="identity", aes(col = agreeableness), size=5.5) +
  geom_text(aes(label=percent_yes), size=1) +
  scale_color_manual(name="Agreeableness",
                     labels=c("Agreeable", "Disagreeable"),
                  values=c("Agreeable"="#00ba38","Disagreeable"= "#f8766d"))+
  facet_wrap(vars(issue)) +
  labs(title="Percentage of voting for yes for five countries that formed part of the UNSC") +
  ylab("Country")+
  xlab("Share of votes for yes")+
  theme(plot.title = element_text(hjust=0.4))+
  theme(legend.title = element_text(size=8)) +
  theme(legend.text = element_text(size=7)) +
  theme(strip.text.x = element_text(size = 6.5))
```



Part (c): including World

```
# create a new dataframe for the non-members of UNSC

# get World statistics
q3c_non_member <- q3a_final %>%
  filter(!(country %in% c("United States", "United Kingdom", "France", "China", "Russia"))) %>%
  group_by(issue) %>%
  summarise(percent_yes=round(mean(vote=="yes")*100, digits=2), n_votes=n())

# create a country column for World
q3c_non_member$country <- c("World")

# add agreeableness column
q3c_non_member$agreeableness <- ifelse(q3c_non_member$percent_yes>50,"Agreeable","Disagreeable")

# dataframe for the World
q3c_non_member
```

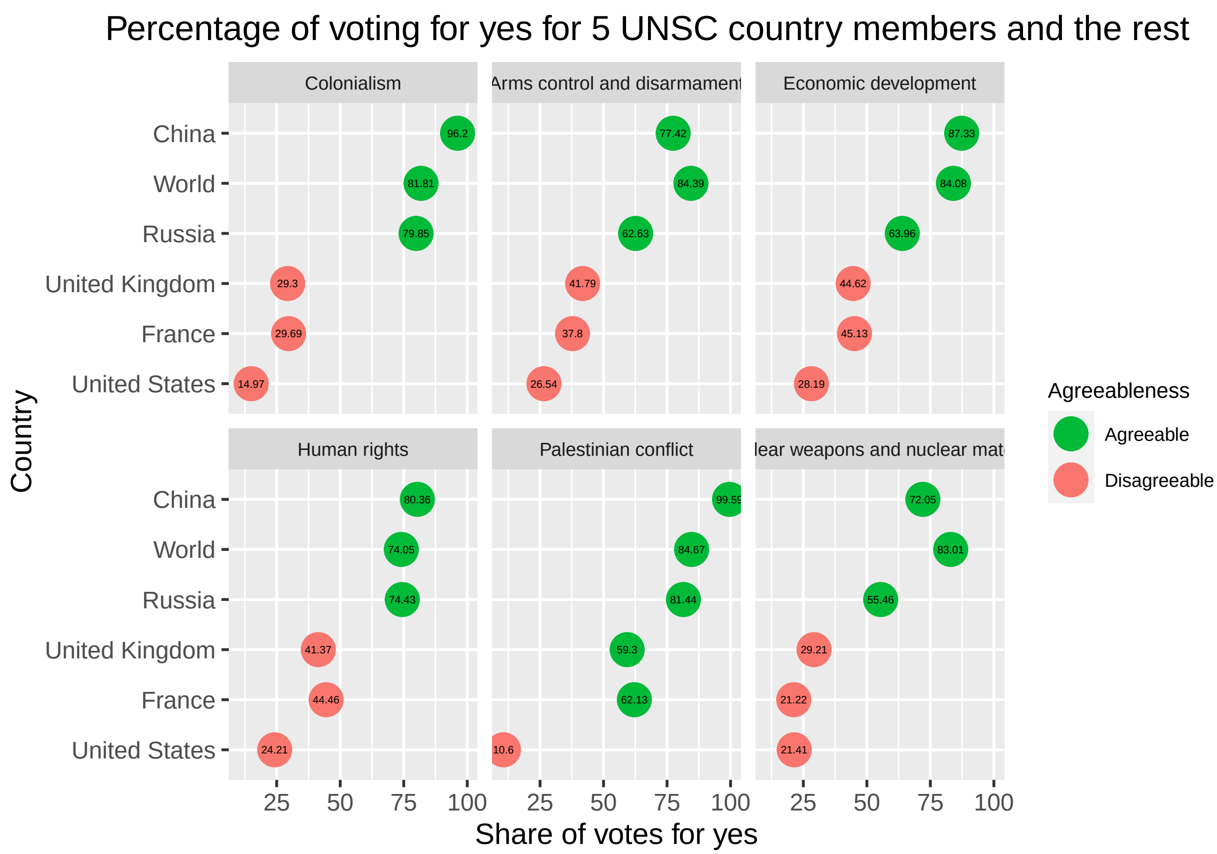
```
## # A tibble: 6 x 5
##   issue                              percent_yes n_votes country agreeableness
##   <fct>                                    <dbl>   <int> <chr>   <chr>
## 1 Colonialism                               81.8  125267 World   Agreeable
## 2 Arms control and disarmament              84.4  165254 World   Agreeable
## 3 Economic development                      84.1  105138 World   Agreeable
## 4 Human rights                              74.0  151716 World   Agreeable
```

```
## 5 Palestinian conflict                       84.7  153458 World   Agreeable
## 6 Nuclear weapons and nuclear material        83.0  129534 World   Agreeable
```

```
# bind rows of the two dataframes: the member and the non-member dataframes
q3c_plot <- bind_rows(q3b_plot, q3c_non_member, id=NULL)

# final data used for plotting
q3c_plot
```

```
## # A tibble: 36 x 5
## # Groups:   country [6]
##    country issue                                percent_yes n_votes agreeableness
##    <chr>   <fct>                                      <dbl>   <int> <chr>
##  1 China   Colonialism                                 96.2     658 Agreeable
##  2 China   Arms control and disarmament                77.4     899 Agreeable
##  3 China   Economic development                        87.3     576 Agreeable
##  4 China   Human rights                                80.4     881 Agreeable
##  5 China   Palestinian conflict                        99.6     967 Agreeable
##  6 China   Nuclear weapons and nuclear materi~         72.0     712 Agreeable
##  7 France  Colonialism                                 29.7     933 Disagreeable
##  8 France  Arms control and disarmament                37.8    1082 Disagreeable
##  9 France  Economic development                        45.1     760 Disagreeable
## 10 France  Human rights                                44.5    1001 Disagreeable
## # ... with 26 more rows
```

```
# add the World statistics to the 5-member plot
ggplot(q3c_plot, aes(percent_yes, reorder(country, percent_yes))) +
  geom_point(stat="identity", aes(col = agreeableness), size=5.5) +
  geom_text(aes(label=percent_yes), size=1.5) +
  scale_color_manual(name="Agreeableness",
                    labels=c("Agreeable", "Disagreeable"),
                 values=c("Agreeable"="#00ba38","Disagreeable"= "#f8766d"))+
  facet_wrap(vars(issue)) +
  labs(title="Percentage of voting for yes for 5 UNSC country members and the rest") +
  ylab("Country")+
  xlab("Share of votes for yes")+
  theme(plot.title = element_text(hjust=0.4)) +
  theme(legend.title = element_text(size=8)) +
  theme(legend.text = element_text(size=7)) +
  theme(strip.text.x = element_text(size = 6.5))
```



Overall, the World tended to be agreeable on all issues, which might imply that there could be a wide census. In more detail, China and Russia were the two countries that consistently voted "yes", possibly expressing mutual (agreeable) understanding and shared international relation ideologies with each other. Meanwhile, the United Kindom, France, and the United States were giving oppositions to almost all issues. Especially, the United States was the least likely to be agreeable on these issues compared to other powerhouse countries and the World. There were some issues that clearly showed a voting division among these five countries and the World such as Colonialism, Human rights, and Palestinian conflict.

Question 4: what issue is dividing opinions the most in the UN

Part (a): process data

```
# restate the dataset (in question 2: un_votes + un_roll_calls)
q4a_joined <- q2a_joined

# vector from un_votes that shows the issues (rcid) that had more than 5 votes at a UN session
rcid_5_votes <- q4a_joined %>%
  group_by(rcid) %>%
  summarise(total_vote=n()) %>%
  filter(total_vote>5)

# using the above filter (sessions with more than 5 votes)
q4a_final <- q4a_joined %>%
  filter(rcid %in% rcid_5_votes$rcid)
```

```
# create two columns: year and vote_value

# column year
q4a_final$year <- format(as.Date(q4a_final$date, format="%Y-%m-%d"), "%Y")
q4a_final$year <- as.integer(q4a_final$year)

# column vote_value
q4a_final$vote_value <- ifelse(q4a_final$vote=="yes", 1, ifelse(q4a_final$vote=="no", -1, 0))

# filter the year > 2000 and get the final dataset for part (a)
q4a_final <- filter(q4a_final, year>2000)
```

Part (b): 50 most disputed rcids and filter

```
# find 50 disputed rcids from "var()" function (higher variance = higher dispute)
q4b_disputed <- q4a_final %>%
  group_by(rcid) %>%
  summarise(variance = var(vote_value)) %>% # var function applied to the "vote_value" column
  arrange(desc(variance)) %>%
  slice_head(n=50)    # pick the 50 most disputed issues

# filter un_roll_calls (for more details) that only contained the 50 disputed rcids above
q4b_disputed_topics <- un_roll_calls %>%
  filter(rcid %in% q4b_disputed$rcid) %>%
  dplyr::select(rcid, date, short, descr)

# top 50 disputed issues in the 21st century
q4b_disputed_topics
```

```
## # A tibble: 50 x 4
##     rcid date       short                                                        descr
##    <int> <date>     <chr>                                                        <chr>
##  1  4352 2001-12-05 HUMAN RIGHTS, COERCIVE MEASURES                              Huma~
##  2  4354 2001-12-05 DEMOCRATIC ORDER                                             Prom~
##  3  4355 2001-12-05 HUMAN RIGHTS, CHARTER PRINCIPLES                             Resp~
##  4  4430 2002-12-05 DEMOCRATIC ORDER                                             Prom~
```

```
## 5  4434 2002-12-05 HUMAN RIGHTS, COERCIVE MEASURES                        Huma~
## 6  4470 2003-12-22 Human rights and unilateral coercive measures : resol~ Huma~
## 7  4538 2004-12-20 Promotion of a democratic and equitable international~ Prom~
## 8  4592 2006-05-08 Investing in the United Nations: for a stronger Organ~ Inve~
## 9  4609 2005-12-16 Promotion of peace as a vital requirement for the ful~ Prom~
## 10 4612 2005-12-16 Human rights and unilateral coercive measures : resol~ Huma~
## # ... with 40 more rows
```

(Note that I included the "desc" column because some issues did not have any meaningful "short" value. I think excluding the "NA" values in the "short" column would not be very reasonable, especially when we also had a "descr" column.)

Part (c): get correlation between the voting pattern of Russia and other countries

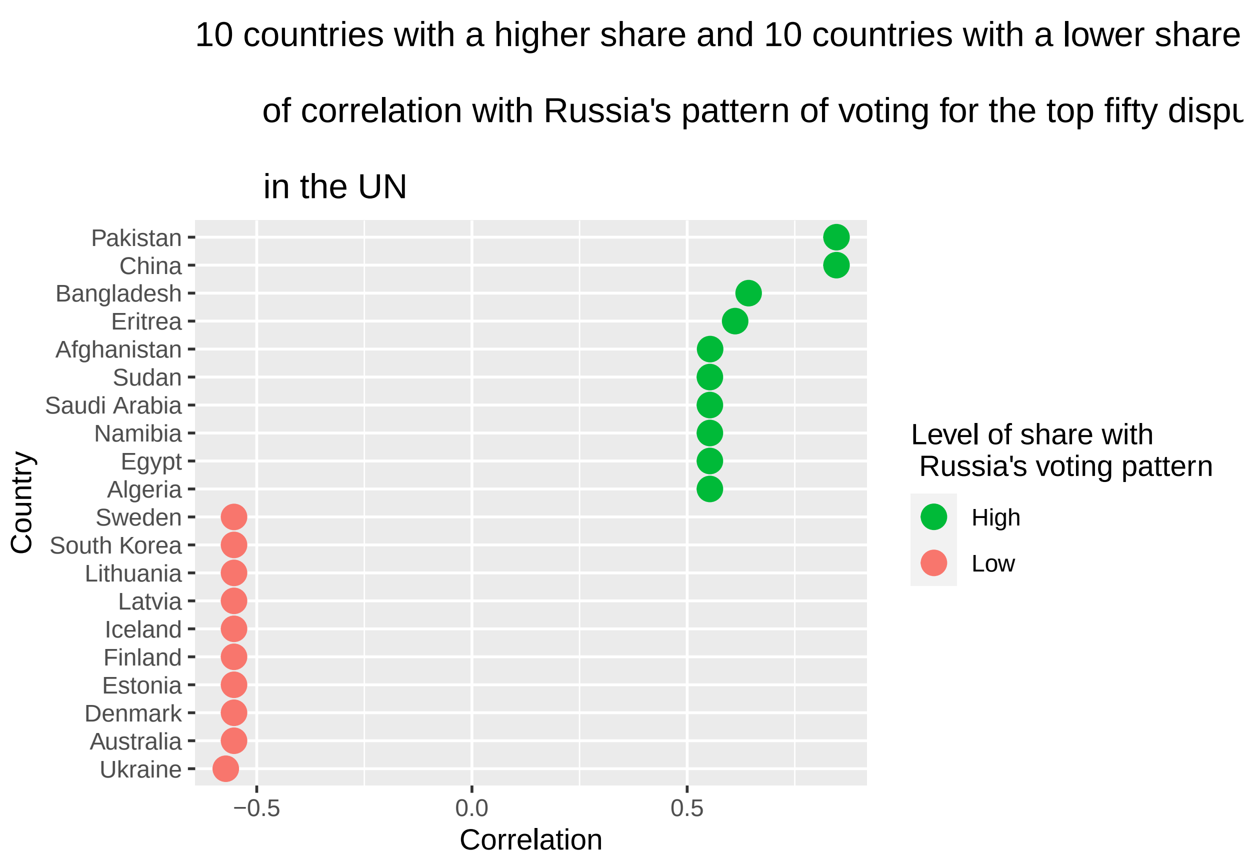
```r
# find what country had a voting pattern opposite to Russia

# process un_votes
q4c_disputed <- un_votes %>%
  filter(rcid %in% q4b_disputed_topics$rcid)

# creata a new column "vote_value"
q4c_disputed$vote_value <- ifelse(q4c_disputed$vote=="yes", 1,
                                  ifelse(q4c_disputed$vote=="no", -1, 0))

# pair_wise: correlations of pairs of items and filter to get Russia
q4c_corr_russia <- q4c_disputed %>%
  pairwise_cor(country, rcid, vote_value) %>%
  filter(item2=="Russia")

# other countries' voting pattern correlation with Russia
q4c_corr_russia
```

```
## # A tibble: 192 x 3
##    item1              item2  correlation
##    <chr>              <chr>        <dbl>
##  1 United States      Russia     -0.553
##  2 Canada             Russia     -0.553
##  3 Bahamas            Russia      0.0867
##  4 Cuba               Russia      0.371
##  5 Haiti              Russia      0.185
##  6 Dominican Republic Russia      0.104
##  7 Jamaica            Russia      0.225
##  8 Trinidad & Tobago  Russia      0.406
##  9 Barbados           Russia      0.352
## 10 Dominica           Russia      0.142
## # ... with 182 more rows
```

(Note: To my understanding of correlation, I think that correlation can be negative. A negative correlation indicates a relationship in which the voting patterns of Russia and other countries were conflicting or going in an opposite direction.)

Part (d): plot the correlation with Russia's pattern of voting

```r
# slice top 10 and bottom 10 regarding share of correlation with Russia's voting pattern
q4d_corr_russia_plot <- q4c_corr_russia %>%
  arrange(desc(correlation)) %>%
  slice(c(1:10), c(n()-9):n()) %>%  # top 10 and bottom 10 correlations with Russia
  mutate(share_level = ifelse(correlation > 0, "High", "Low"))
  # add a new column to color the high/low share of correlation

# plot
ggplot(q4d_corr_russia_plot, aes(correlation, reorder(item1, correlation))) +
  geom_point(stat="identity", aes(col = share_level), size=4) +
  scale_color_manual(name="Level of share with\n Russia's voting pattern",
                     labels=c("High", "Low"),
                     values=c("High"="#00ba38","Low"= "#f8766d")) +
  labs(title="10 countries with a higher share and 10 countries with a lower share\n
       of correlation with Russia's pattern of voting for the top fifty disputed issues\n
       in the UN") +
  ylab("Country")+
  xlab("Correlation") +
  theme(plot.title = element_text(hjust=0.001))
```



(Note: You can see that I made an assumption that if the correlation is higher than 0, it is considered a high share, and vice versa.)

Question 5: The effect of natural disasters of different magnitudes influence how countries - where those disasters took place - vote in UN voting sessions for resolutions concerning environmental issues

Part (a): import climate-vote.csv

```
# import and familiarize "climate-vote.csv" file
climate_vote <- read.csv("C:\\dpss-capstone\\climate-vote.csv")
```

Part (b): import natual-disaster.csv

```
# import and familiarize "natural-disaster.csv" file
natural_disaster <- read.csv("C:\\dpss-capstone\\natural-disaster.csv")
```

Part (c): process data and analyze regressions

```
# join the two datasets
joined <- merge(climate_vote, natural_disaster, by=c("rcid", "country_code"))

# Note that there were 5010 observations in this joined dataframe, rather than 5019 like in dataframes
# climate_vote or natural_disaster. This is because there were 9 observations in dataframe
# natural_disaster that had value "0" for country_code, which matched nothing with the dataframe
# climate_vote. Try these two codes to see:
# (1) `%!in%` <- Negate(`%in%`)
# (2) natural_disaster %>% filter(country_code %!in% joined$country_code)
```

Now, I will conduct some regression models. In general, there are two regressions to explore:

(1) pro-climate vote (*pro_climate_vote*) versus the number of disasters (*number_disasters*)
(2) pro-climate vote (*pro_climate_vote*) versus the existence of a disaster one year before a voting session (*disaster_before_vote*)

For each regression, I chose LPM and PM models because the dependent variable holds binary results. That is, the *pro_climate_vote* variable had the values of either 0 or 1. I decided to include both models in this regression analysis with the view to comparing the results and identifying any noticeable issues. Also, for probit models, I wanted to show a comprehensive comparison of having and not having the average partial effect. Lastly, for the comparison purpose as well, I decided to include two-way fixed-effects models to check the potential heterogeneity bias. Six models might be numerous for a solely regression analysis, but I think it is worthwhile to view the regression results from multiple perspectives. Note that we assume the significance level is 5 percent.

Regression (1):

- Linear probability model (LPM)
- Probit model (PM)
- Probit model with Average Partial Effect (PM with APE)
- Linear probability model with Fixed Effects (LPM with FE)
- Probit model with Fixed Effects (PM with FE)
- Probit model with Average Partial Effect and Fixed Effects (PM with APE and FE)

Linear probability model (LPM)

```r
# Regression 1 - linear probability model (LMP)
lpm_model_1 <- lm(pro_climate_vote ~ number_disasters, data=joined)
tidy(lpm_model_1)
```

```
## # A tibble: 2 x 5
##   term             estimate std.error statistic p.value
##   <chr>               <dbl>     <dbl>     <dbl>   <dbl>
## 1 (Intercept)       0.880     0.00525    168.    0
## 2 number_disasters -0.00688   0.00277     -2.48  0.0131
```

On average, holding other factors constant, an increase in the number of natural disasters of a country (within a year before a voting session over a UN resolution regarding environmental issues) by one unit is associated with a decrease in the probability of that country's vote being a pro-climate type by 0.69 percent. The estimated coefficient is statistically significant given the small p-value (0.0131).

Probit model (PM)

```r
# regression 1 - Probit model (PM) without Average Partial Effect (APE)
probit_model_1 <- glm(pro_climate_vote ~ number_disasters, data=joined,
                      binomial(link="probit"))
tidy(probit_model_1)
```

```
## # A tibble: 2 x 5
##   term             estimate std.error statistic p.value
##   <chr>               <dbl>     <dbl>     <dbl>   <dbl>
## 1 (Intercept)        1.17      0.0254     46.2   0
## 2 number_disasters  -0.0303    0.0124     -2.45  0.0144
```

On average, holding other factors constant, an increase in the number of natural disasters in a country by one unit is associated with a decrease in the probability of that country's vote being a pro-climate type by 3.03 percent. The estimated coefficient is statistically significant given the small p-value (0.0144).

Probit model with Average Partial Effect (PM with APE)

```r
# regression 1 - Probit model (PM) with Average Partial Effect (APE)
probit_ape_model_1 <- probitmfx(pro_climate_vote ~ number_disasters,
                                data=joined, atmean=FALSE)
tidy(probit_ape_model_1)
```

```
## # A tibble: 1 x 6
##   term             atmean estimate std.error statistic p.value
##   <chr>            <lgl>     <dbl>     <dbl>     <dbl>   <dbl>
## 1 number_disasters FALSE  -0.00626   0.00256     -2.45  0.0144
```

On average, holding other factors constant, an increase in the number of natural disasters in a country by one unit is associated with a decrease in the probability of that country's vote being a pro-climate type by 0.63 percent. The estimated coefficient is statistically significant given the small p-value (0.0144).

The following regression models are like above but now with two-way fixed effects:

Linear probability model with Fixed Effects (LPM with FE)

```r
# regression 1 - LMP with Fixed Effects (FE)
lpm_model_1_fe <- lm(pro_climate_vote ~ number_disasters + factor(country_code)
                     + factor(date), data=joined)

# convert the results to dataframe
# change some columns to have 2 digit after decimals (for ease of viewing)
lpm_model_1_fe_results <- tidy(lpm_model_1_fe) %>%
  mutate(estimate = format(round(estimate, 4), nsmall=2)) %>%
  mutate(p.value = format(round(p.value, 4), nsmall=2))

# show results
lpm_model_1_fe_results
```

```
## # A tibble: 215 x 5
##    term                    estimate  std.error statistic p.value
##    <chr>                   <chr>         <dbl>     <dbl> <chr>
##  1 (Intercept)             " 0.7487"   0.0692    10.8    0.0000
##  2 number_disasters        "-0.0023"   0.00431   -0.524  0.6005
##  3 factor(country_code)AE "-0.2122"   0.0838    -2.53   0.0113
##  4 factor(country_code)AF "-0.1107"   0.0884    -1.25   0.2105
##  5 factor(country_code)AG " 0.0723"   0.0891     0.812  0.4169
##  6 factor(country_code)AL "-0.0015"   0.0873    -0.0170 0.9864
##  7 factor(country_code)AM " 0.0825"   0.0881     0.936  0.3495
##  8 factor(country_code)AO " 0.0785"   0.0862     0.911  0.3625
##  9 factor(country_code)AR " 0.0250"   0.0842     0.297  0.7668
## 10 factor(country_code)AT "-0.0649"   0.0873    -0.743  0.4575
## # ... with 205 more rows
```

On average, holding other factors constant, an increase in the number of natural disasters in a country by one unit is associated with a decrease in the probability of that country's vote being a pro-climate type by 0.23 percent. The estimated coefficient is not statistically significant given the large p-value (0.6005).

Probit model with Fixed Effects (PM with FE)

```r
# regression 1 - PM with FE
probit_model_1_fe <- glm(pro_climate_vote ~ number_disasters + factor(country_code)
                         + factor(date), data=joined, binomial(link="probit"))
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```r
# convert the results to dataframe and
# change some columns to have 2 digit after decimals (for ease of viewing)
probit_model_1_fe_results <- tidy(probit_model_1_fe) %>%
  mutate(estimate = format(round(estimate, 4), nsmall=2)) %>%
  mutate(p.value = format(round(p.value, 4), nsmall=2))

# show results
probit_model_1_fe_results
```

```
## # A tibble: 215 x 5
##    term                    estimate  std.error statistic p.value
##    <chr>                   <chr>         <dbl>     <dbl> <chr>
```

```
##   1 (Intercept)            " 0.7079"    0.421    1.68    0.0927
##   2 number_disasters       "-0.0068"    0.0268  -0.254   0.7993
##   3 factor(country_code)AE "-0.9584"    0.471   -2.03    0.0421
##   4 factor(country_code)AF "-0.6268"    0.507   -1.24    0.2166
##   5 factor(country_code)AG " 4.6218"  494.       0.00936 0.9925
##   6 factor(country_code)AL " 0.0043"    0.560    0.00761 0.9939
##   7 factor(country_code)AM " 4.6652"  488.       0.00957 0.9924
##   8 factor(country_code)AO " 4.6583"  453.       0.0103  0.9918
##   9 factor(country_code)AR " 0.1850"    0.556    0.333   0.7393
##  10 factor(country_code)AT "-0.3305"    0.514   -0.642   0.5206
## # ... with 205 more rows
```

On average, holding other factors constant, an increase in the number of natural disasters in a country by one unit is associated with a decrease in the probability of that country's vote being a pro-climate type by 0.68 percent. The estimated coefficient is not statistically significant given the large p-value (0.7993).

Probit model with Average Partial Effect and Fixed Effects (PM with APE and FE)

```r
# regression 1 - PM with APE and FE
probit_ape_model_1_fe <- probitmfx(pro_climate_vote ~ number_disasters
                                   + factor(country_code) + factor(date),
                                   data=joined, atmean=FALSE)
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```r
# convert the results to dataframe and
# change some columns to have 2 digit after decimals (for ease of viewing)
probit_ape_model_1_fe_results <- tidy(probit_ape_model_1_fe) %>%
  mutate(estimate = format(round(estimate, 4), nsmall=2)) %>%
  mutate(p.value = format(round(p.value, 4), nsmall=2))

# show results
probit_ape_model_1_fe_results
```

```
## # A tibble: 214 x 6
##    term                   atmean estimate  std.error statistic p.value
##    <chr>                  <lgl>  <chr>         <dbl>     <dbl> <chr>
##  1 number_disasters       FALSE  "-0.0010"   0.00392  -0.254   0.7992
##  2 factor(country_code)AE FALSE  "-0.1840"   0.107    -1.73    0.0843
##  3 factor(country_code)AF FALSE  "-0.1117"   0.105    -1.06    0.2869
##  4 factor(country_code)AG FALSE  " 0.1244"   0.00447  27.8     0.0000
##  5 factor(country_code)AL FALSE  " 0.0006"   0.0816    0.00762 0.9939
##  6 factor(country_code)AM FALSE  " 0.1244"   0.00436  28.5     0.0000
##  7 factor(country_code)AO FALSE  " 0.1244"   0.00434  28.7     0.0000
##  8 factor(country_code)AR FALSE  " 0.0252"   0.0702    0.359   0.7195
##  9 factor(country_code)AT FALSE  "-0.0541"   0.0930   -0.582   0.5605
## 10 factor(country_code)AU FALSE  "-0.3004"   0.107    -2.81    0.0050
## # ... with 204 more rows
```

On average, holding other factors constant, an increase in the number of natural disasters in a country by one unit is associated with a decrease in the probability of that country's vote being a pro-climate type by 0.10 percent. The estimated coefficient is not statistically significant given the large p-value (0.7992).

So, to conclude, we see that despite the variations in statistical significance across six models, the relationship between the number of disasters and the chance of having a pro-climate vote is likely to be negative and the difference is small.

Regression 2:

- Linear probability model (LPM)
- Probit model (PM)
- Probit model with Average Partial Effect (PM with APE)
- Linear probability model with Fixed Effects (LPM with FE)
- Probit model with Fixed Effects (PM with FE)
- Probit model with Average Partial Effect and Fixed Effects (PM with APE and FE)

Linear probability model (LPM)

```
# regression 2 - LPM
lpm_model_2 <- lm(pro_climate_vote ~ disaster_before_vote, data=joined)
tidy(lpm_model_2)
```

```
## # A tibble: 2 x 5
##   term                 estimate std.error statistic p.value
##   <chr>                   <dbl>     <dbl>     <dbl>   <dbl>
## 1 (Intercept)            0.881    0.00603    146.    0
## 2 disaster_before_vote  -0.0171   0.00958     -1.79  0.0739
```

On average, holding other factors constant, the existence of a disaster one year before a voting session in a country is associated with a decrease in the probability of that country's vote being a pro-climate type by 1.71 percent. The estimated coefficient is not statistically significant given the large p-value (0.0739).

Probit model (PM)

```
# regression 2 - PM without APE
probit_model_2 <- glm(pro_climate_vote ~ disaster_before_vote, data=joined,
                      binomial(link="probit"))
tidy(probit_model_2)
```

```
## # A tibble: 2 x 5
##   term                 estimate std.error statistic p.value
##   <chr>                   <dbl>     <dbl>     <dbl>   <dbl>
## 1 (Intercept)            1.18      0.0296     39.9   0
## 2 disaster_before_vote  -0.0821    0.0460     -1.78  0.0746
```

On average, holding other factors constant, the existence of a disaster one year before a voting session in a country is associated with a decrease in the probability of that country's vote being a pro-climate type by 8.21 percent. The estimated coefficient is not statistically significant given the large p-value (0.0746).

Probit model with Average Partial Effect (PM with APE)

```r
# regression 2 - PM with APE
probit_ape_model_2 <- probitmfx(pro_climate_vote ~ disaster_before_vote,
                                data=joined, atmean=FALSE)
tidy(probit_ape_model_2)
```

```
## # A tibble: 1 x 6
##   term                 atmean estimate std.error statistic p.value
##   <chr>                <lgl>     <dbl>     <dbl>     <dbl>   <dbl>
## 1 disaster_before_vote FALSE   -0.0171   0.00969     -1.77  0.0773
```

On average, holding other factors constant, the existence of a disaster one year before a voting session in a country is associated with a decrease in the probability of that country's vote being a pro-climate type by 1.71 percent. The estimated coefficient is not statistically significant given the large p-value (0.0773).

The same regression models like above but now with fixed effects:

Linear probability model with Fixed Effects (LPM with FE)

```r
# regression 2 - LMP with FE
lpm_model_2_fe <- lm(pro_climate_vote ~ disaster_before_vote +
                       factor(country_code) + factor(date), data=joined)

# convert the results to dataframe and
# change some columns to have 2 digit after decimals (for ease of viewing)
lpm_model_2_fe_results <- tidy(lpm_model_2_fe) %>%
  mutate(estimate = format(round(estimate, 4), nsmall=2)) %>%
  mutate(p.value = format(round(p.value, 4), nsmall=2))

# show results
lpm_model_2_fe_results
```

```
## # A tibble: 215 x 5
##    term                    estimate  std.error statistic p.value
##    <chr>                   <chr>         <dbl>     <dbl> <chr>
##  1 (Intercept)             " 0.7451"    0.0692   10.8    0.0000
##  2 disaster_before_vote    "-0.0152"    0.0120   -1.26   0.2067
##  3 factor(country_code)AE "-0.2099"    0.0838   -2.51   0.0123
##  4 factor(country_code)AF "-0.1036"    0.0880   -1.18   0.2393
##  5 factor(country_code)AG " 0.0725"    0.0891    0.814  0.4159
##  6 factor(country_code)AL " 0.0046"    0.0875    0.0528 0.9579
##  7 factor(country_code)AM " 0.0826"    0.0881    0.937  0.3489
##  8 factor(country_code)AO " 0.0875"    0.0864    1.01   0.3112
##  9 factor(country_code)AR " 0.0336"    0.0844    0.398  0.6909
## 10 factor(country_code)AT "-0.0620"    0.0873   -0.710  0.4778
## # ... with 205 more rows
```

On average, holding other factors constant, the existence of a disaster one year before a voting session in a country is associated with a decrease in the probability of that country's vote being a pro-climate type by 1.52 percent. The estimated coefficient is not statistically significant given the large p-value (0.2067).

Probit model with Fixed Effects (PM with FE)

```r
# regression 2 - PM with FE
probit_model_2_fe <- glm(pro_climate_vote ~ disaster_before_vote
                         + factor(country_code) + factor(date), data=joined,
                         binomial(link="probit"))
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```r
# convert the results to dataframe and
# change some columns to have 2 digit after decimals (for ease of viewing)
probit_model_2_fe_results <- tidy(probit_model_2_fe) %>%
  mutate(estimate = format(round(estimate, 4), nsmall=2)) %>%
  mutate(p.value = format(round(p.value, 4), nsmall=2))

# show results
probit_model_2_fe_results
```

```
## # A tibble: 215 x 5
##    term                  estimate  std.error statistic p.value
##    <chr>                 <chr>         <dbl>     <dbl> <chr>
##  1 (Intercept)           " 0.6788"    0.421    1.61    0.1072
##  2 disaster_before_vote  "-0.0975"    0.0780  -1.25    0.2113
##  3 factor(country_code)AE "-0.9429"    0.471   -2.00    0.0455
##  4 factor(country_code)AF "-0.5555"    0.504   -1.10    0.2700
##  5 factor(country_code)AG " 4.6214"  494.       0.00936 0.9925
##  6 factor(country_code)AL " 0.0400"    0.559    0.0716  0.9429
##  7 factor(country_code)AM " 4.6652"  488.       0.00957 0.9924
##  8 factor(country_code)AO " 4.7387"  452.       0.0105  0.9916
##  9 factor(country_code)AR " 0.2547"    0.558    0.457   0.6478
## 10 factor(country_code)AT "-0.3089"    0.515   -0.600   0.5486
## # ... with 205 more rows
```

On average, holding other factors constant, the existence of a disaster one year before a voting session in a country is associated with a decrease in the probability of that country's vote being a pro-climate type by 9.75 percent. The estimated coefficient is not statistically significant given the large p-value (0.2113).

Probit model with Average Partial Effect and Fixed Effects (PM with APE and FE)

```r
# regression 2 - PM with APE and FE
probit_ape_model_2_fe <- probitmfx(pro_climate_vote ~ disaster_before_vote
                                   + factor(country_code) + factor(date),
                                   data=joined, atmean=FALSE)
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```r
# convert the results to dataframe and
# change some columns to have 2 digit after decimals (for ease of viewing)
probit_ape_model_2_fe_results <- tidy(probit_ape_model_2_fe) %>%
  mutate(estimate = format(round(estimate, 4), nsmall=2)) %>%
  mutate(p.value = format(round(p.value, 4), nsmall=2))

# show results
probit_ape_model_2_fe_results
```

```
## # A tibble: 214 x 6
##    term                   atmean estimate  std.error statistic p.value
##    <chr>                  <lgl>  <chr>         <dbl>     <dbl> <chr>
##  1 disaster_before_vote   FALSE  "-0.0143"   0.0115    -1.24   0.2140
##  2 factor(country_code)AE FALSE  "-0.1803"   0.106     -1.70   0.0896
##  3 factor(country_code)AF FALSE  "-0.0971"   0.101     -0.959  0.3375
##  4 factor(country_code)AG FALSE  " 0.1243"   0.00449   27.7    0.0000
##  5 factor(country_code)AL FALSE  " 0.0058"   0.0793     0.0727 0.9421
##  6 factor(country_code)AM FALSE  " 0.1244"   0.00437   28.5    0.0000
##  7 factor(country_code)AO FALSE  " 0.1245"   0.00423   29.4    0.0000
##  8 factor(country_code)AR FALSE  " 0.0337"   0.0661     0.509  0.6104
##  9 factor(country_code)AT FALSE  "-0.0502"   0.0919    -0.546  0.5850
## 10 factor(country_code)AU FALSE  "-0.2815"   0.109     -2.59   0.0096
## # ... with 204 more rows
```

On average, holding other factors constant, the existence of a disaster one year before a voting session in a country is associated with a decrease in the probability of that country's vote being a pro-climate type by 1.43 percent. The estimated coefficient is not statistically significant given the large p-value (0.2140).

In conclusion, we see that the results were relatively consistent among LPM, PM with APE, LPM with FE, and PM with APE and FE. However, the estimated coefficients of all models were not statistically significant. This means we probably need more control variables.

Thank you very much for preparing this fun capstone project. I have learned a lot!

I will remember all of you. Wish you all the best!