Assignment 1

AUTHOR
Junhao Fan

Overview: *For this assignment, you are going to download some variables using the* `wbstats` *and* `vdemdata` *packages. Then you are going to wrangle these data and merge the two data sets into one and analyze how they relate to women's representation in national parliaments. Do your work for each step in the code chunks provided. Be sure to label your code chunks.*

Step 1: Download data from the V-Dem Dataset

Look at the V-Dem [codebook](). Identify two measures of democracy that are not the polyarchy score that we used in Module 1.2.

Try to pick variables that will relate to women's representation and read the description of the variable to make sure that it includes data for a sufficient number of years, e.g. that the data will be somewhat recent. Select the most recent 20 years of data for your analysis.

Make sure to load the packages that you need and `glimpse()` *the data or* `View()` *it to make sure that it downloaded properly.*

```
library(wbstats) # upload world bank package
library(vdemdata) # Upload the V-dem package
library(dplyr) # Use dplyr to view and analysis the dataset
```

```
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(janitor)
```

```
Attaching package: 'janitor'

The following objects are masked from 'package:stats':

    chisq.test, fisher.test
```

```r
# Download the data
dem <- vdem |>
  filter(year >= 2003) |> # most recent data in 20 years
  select(
    country = country_name,
    vdem_ctry_id = country_id,
    year,
    partidem = v2x_partipdem, # participatory democracy (1st democracy indicator)
    liberal_dem = v2x_libdem, # Liberal democracy (2nd democracy indicator)
    equal_access = v2xeg_eqaccess, # Equal access to the resources (might relate with wom
    equal_dist = v2xeg_eqdr, # Equal distribution of resources (women_rep)
    region = e_regionpol_6C
  ) |>
  mutate(
    region = case_match(region,
                        1 ~ "Eastern Europe",
                        2 ~ "Latin America",
                        3 ~ "Middle East",
                        4 ~ "Africa",
                        5 ~ "Western",
                        6 ~ "Asia")
  )
# View the data
glimpse(dem)
```

```
Rows: 3,568
Columns: 8
$ country      <chr> "Mexico", "Mexico", "Mexico", "Mexico", "Mexico", "Mexico…
$ vdem_ctry_id <dbl> 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, …
$ year         <dbl> 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 201…
$ partidem     <dbl> 0.440, 0.441, 0.441, 0.417, 0.397, 0.394, 0.395, 0.405, 0…
$ liberal_dem  <dbl> 0.503, 0.510, 0.510, 0.482, 0.454, 0.456, 0.455, 0.459, 0…
$ equal_access <dbl> 0.550, 0.550, 0.550, 0.550, 0.569, 0.569, 0.569, 0.592, 0…
$ equal_dist   <dbl> 0.312, 0.312, 0.312, 0.312, 0.312, 0.312, 0.312, 0.312, 0…
$ region       <chr> "Latin America", "Latin America", "Latin America", "Latin…
```

Step 2: Download data from the World Bank

*Next, download the variable on women's represenation that we used in Module 1.2 ("SG.GEN.PARL.ZS")
and at least one additional measure related to women's empowerment. Go to the [WDI](#) site the
wb_search() function to identify relevant variables. Download the most recent 20 years of data for your
analysis.*

*Make sure that the indicator has enough data to conduct your analysis, i.e. that it has data for most
countries and years.*

```r
library(wbstats)
library(dplyr)
```

```r
library(janitor)

fsemp_indicators <- wb_search("female share of employment")

# Store a list of object
indicators <- c("fsemp" = "SL.EMP.SMGT.FE.ZS", "women_rep" = "SG.GEN.PARL.ZS")

# fesmp means the female share of employment in senior and middle management

# Download the data:
women_emp <- wb_data(indicators, mrv = 20) |> # Most recent 20 years value
  select(!iso2c) |>
  rename(year = date) |>
  mutate(
    fsemp = round_to_fraction(fsemp, denominator = 100),
    women_rep = round_to_fraction(women_rep, denominator = 100)
  )

# Glimpse of data
glimpse(women_emp)
```

```
Rows: 4,340
Columns: 5
$ iso3c     <chr> "ABW", "ABW", "ABW", "ABW", "ABW", "ABW", "ABW", "ABW", "ABW…
$ country   <chr> "Aruba", "Aruba", "Aruba", "Aruba", "Aruba", "Aruba", "Aruba…
$ year      <dbl> 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, …
$ women_rep <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, …
$ fsemp     <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, …
```

Step 3: Merge the data

Now add country codes using the `countrycode` package and merge the data using `left_join()`.

```r
library(dplyr)
library(remotes)
install_github('vincentarelbundock/countrycode') # Dowmload the package
```

```
Skipping install of 'countrycode' from a github remote, the SHA1 (016e3970) has not
changed since last install.
  Use `force = TRUE` to force installation
```

```r
library(countrycode) # Import the country code and add it in the dataset
dem <- dem |>
  mutate(iso3c = countrycode(
    sourcevar = vdem_ctry_id,
    origin = "vdem",
    destination = "wb",
```

```
  )) |>
  relocate(iso3c, .after = vdem_ctry_id)
```

Warning: There was 1 warning in `mutate()`.
ℹ In argument: `iso3c = countrycode(...)`.
Caused by warning:
! Some values were not matched unambiguously: 128, 139, 236

```
glimpse(dem)
```

Rows: 3,568
Columns: 9
$ country   <chr> "Mexico", "Mexico", "Mexico", "Mexico", "Mexico", "Mexico…
$ vdem_ctry_id <dbl> 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, …
$ iso3c   <chr> "MEX", "MEX", "MEX", "MEX", "MEX", "MEX", "MEX", "MEX", "…
$ year   <dbl> 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 201…
$ partidem   <dbl> 0.440, 0.441, 0.441, 0.417, 0.397, 0.394, 0.395, 0.405, 0…
$ liberal_dem <dbl> 0.503, 0.510, 0.510, 0.482, 0.454, 0.456, 0.455, 0.459, 0…
$ equal_access <dbl> 0.550, 0.550, 0.550, 0.550, 0.569, 0.569, 0.569, 0.592, 0…
$ equal_dist   <dbl> 0.312, 0.312, 0.312, 0.312, 0.312, 0.312, 0.312, 0.312, 0…
$ region   <chr> "Latin America", "Latin America", "Latin America", "Latin…

```
library(readr)
library(dplyr)

# Perform left join using common iso3c and year
dem_women <- left_join(dem, women_emp, by = c("iso3c", "year")) |>
  rename(country = country.x) |> # set country column as country.x
  select(!country.y) # Delete the country.y column

# View data
glimpse(dem_women)
```

Rows: 3,568
Columns: 11
$ country   <chr> "Mexico", "Mexico", "Mexico", "Mexico", "Mexico", "Mexico…
$ vdem_ctry_id <dbl> 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, …
$ iso3c   <chr> "MEX", "MEX", "MEX", "MEX", "MEX", "MEX", "MEX", "MEX", "…
$ year   <dbl> 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 201…
$ partidem   <dbl> 0.440, 0.441, 0.441, 0.417, 0.397, 0.394, 0.395, 0.405, 0…
$ liberal_dem <dbl> 0.503, 0.510, 0.510, 0.482, 0.454, 0.456, 0.455, 0.459, 0…
$ equal_access <dbl> 0.550, 0.550, 0.550, 0.550, 0.569, 0.569, 0.569, 0.592, 0…
$ equal_dist   <dbl> 0.312, 0.312, 0.312, 0.312, 0.312, 0.312, 0.312, 0.312, 0…
$ region   <chr> "Latin America", "Latin America", "Latin America", "Latin…
$ women_rep   <dbl> 22.60, 22.60, 24.20, 22.60, 23.20, 23.20, 27.60, 26.20, 2…
$ fsemp   <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, 34.9, 35.3, 34.8,…

Step 4: Summarize your combined data set

Use `group_by()`, `summarize()` *and* `arrange()` *to glean insights about your data. For example, how do regions compare on mean values of women's representation and how do these values relate to the values of other variables in your data set? Which countries stand out in terms of women's representation? Etc.*

```
# group_by(), summarize() and arrange()
dem_summary <- dem_women |> # save result as new object
  group_by(region)  |> # group dem_women data by region
  summarize(              # summarize following vars (by region)
    partidem = mean(partidem, na.rm = TRUE), # calculate mean, remove NAs
    liberal_dem = mean(liberal_dem, na.rm = TRUE),
    fsemp = mean(fsemp, na.rm = TRUE),
    women_rep = mean(women_rep, na.rm = TRUE)
  ) |>
  arrange(desc(women_rep)) # arrange in descending order by women representation score
```

Now we can see that the western countries has the highest score of women's representation score around the world.

```
# group_by countries
dem_summary_country <- dem_women |> # save result as new object
  group_by(country)  |> # group dem_women data by country
  summarize(              # summarize following vars (by country)
    partidem = mean(partidem, na.rm = TRUE), # calculate mean, remove NAs
    liberal_dem = mean(liberal_dem, na.rm = TRUE),
    fsemp = mean(fsemp, na.rm = TRUE),
    women_rep = mean(women_rep, na.rm = TRUE)
  ) |>
  arrange(desc(women_rep)) # arrange in descending order by women representation score
```

From the information above, we can see that Rwanda, Sweden, Cuba, Finland, and south africa has the highest score of women's representation.

Step 5: Interpret your data

Write a paragraph or so relating your insights to the Norris reading about electoral reform and women's representation. Picking a handful of country cases from your data, how does your analysis relate to her arguments about the relevance of electoral systems, statutory quotas, reserved seats or voluntary quotas?

My answers to question 5:

In the conclusion of Norris' paper, the author states that with a PR electoral system, especially with a large district magnitude, parties may able to balance the lists of candidates, which can reduce the discrimination on either gender, race, or other elements. Further more, the statutory quotas also allows

women to have certain mandatory seats in the government, which also balanced the fairness of election. According to Norris' paper, in table 2, figure 2, we can see that Rwanda, Spain, and Finland are progressively using the PR system, and they all have a high score in women representation. As mentioned in the passage later, the statutory gender quota has been widely applied in European and Latin America countries. From the dem_summary results, we can see that the Western and Latin America region have the highest value of women representation (Western: 29.83, Latin America: 23.55); also, Finland and Norway applied this quota, they also have a relatively high score on women representation (For further information, please refer to the chart dem_summary_country). For the conclusion based on Norris' paper and the data analysis, countries with a PR electoral system and which applied the statutory quota have a higher proportion of women's representation in the government than the counties which haven't.