



Women in Parliament – data.table

Saghir Bashir

This version was compiled on September 25, 2020

We will use the World Bank's indicator data for "Women in Parliament" as a case study when working with the `data.table` R package. We will guide you through the geographical and time trends for the percentage of women in national parliaments. We will start by learning about and understanding the raw data, which we will then process ("wrangle") in preparation for some exploratory analysis.

Women in Parliament | World Bank Indicator | `data.table` | Tinyverse

1. Preface

We present a real-life case study for the `data.table`¹ package using the World Bank's "Women in Parliament" indicator data. To get the most out of this case-study guide, repeat the examples and do the exercises whilst reading it.

Guide materials. You can download materials for this guide from this link:

- <https://ilustat.com/shared/WiP-rdatatable.zip>

Unzip the file, which contains the data, this guide and an R script exercise file. We advise you to work with "WiP-Exercise.R" file to follow the examples and do the exercises. If you are using RStudio, you can double click on "WiP-dt.Rproj" to get started.

2. Objectives

Explore the geographical and time trends for the percentage of women² in national parliaments.

3. Understanding the Data

The World Bank Data. The raw data for "Proportion of seats held by women in national parliaments" includes the percentage of women in parliament ("single or lower parliamentary chambers only") by country (region) and year. It can be downloaded from:³

- <https://data.worldbank.org/indicator/SG.GEN.PARL.ZS>

As part of its "open data" mission the World Bank offers "free and open access to global development data" kindly licensed under the "Creative Commons Attribution 4.0 (CC-BY 4.0)".⁴

Source Data. The data originates from the "Inter-Parliamentary Union" (IPU)⁵ which provides an "Archive of statistical data on the percentage of women in national parliaments" going back to 1997 on a monthly basis:

- <http://archive.ipu.org/wmn-e/classif-arc.htm>

The World Bank data is for "single or lower parliamentary chambers only", while the IPU also presents data for "Upper Houses or Senates". Moreover, the IPU provides the actual numbers used to calculate the percentages (which the World Bank does not).

Data limitations. Take caution when interpreting these data, as parliamentary systems vary from country to country, and in some cases over time. Some of the issues to consider include:

- Who has, and who does not have, the right to become a Member of Parliament (MP)?
- How does someone become an MP? Through democratic elections? How is "democratic election" defined?
- What is the real power of MPs and their parliament? Can MPs make a difference?

Data definitions & assumptions.

"Women". The definition for "women" is not given, so we will assume that it refers to a binary classification for gender (sex).

"Country (Region)". The definition of countries and regions can change over time. (e.g. formation of new countries after conflicts, new member states joining a pre-existing collective). How are these changes reflected in the data? How do they affect the interpretation?

Pro tip. Understand the limitations of your data before anybody else points them out to you.

4. About the data file

The data is stored in a file called:

- `API_SG.GEN.PARL.ZS_DS2_en_csv_v2_1345211.csv`

To simplify things we have copied it to `WB-WiP.csv` (which also allows us to maintain the original file in case something goes wrong).

Pro tip. Always keep a backup copy of the data. Alternatively, set the data file(s) to "read-only" to protect it from being overwritten or modified.

Exercise. It is important to look at and understand the contents of the file before you start using it. Using a text editor or a spreadsheet software, open the `WB-WiP.csv` file (in the data directory). What do you observe in the contents of this file?

Content and Structure. The first four lines of `WB-WiP.csv` can be ignored, since they contain two lines of meta-information and two blank lines, as follows:

```
1 "Data Source","World Development Indicators",
2
3 "Last Updated Date","2020-09-08",
4
```

The fifth line contains the column (variable) names and the body of data starts in the sixth line. It is important to note that there was no collection of data for a majority of the years, which means that it is "missing".

5. Importing the data

Based on our findings above, we can "skip" the first four lines and treat the fifth line as column (variable) names. Also note the use of the `check.names` argument to ensure that the column names are compliant in R.

¹For more information on the `data.table` package see <http://r-datatable.com/>.

²The objective could be termed neutrally as "gender trends" but we will keep it per the World Bank data.

³The `wbstats` R package (<https://cran.r-project.org/web/packages/wbstats/>) gives access to a "tidier" version of the World Bank indicator data.

⁴<https://datacatalog.worldbank.org/public-licenses/cc-by>.

⁵Inter-Parliamentary Union: <https://www.ipu.org/>.

```
library(data.table)
library(here)
wip <- fread(here("data", "WB-WiP.csv"),
             skip = 4, header = TRUE,
             check.names = TRUE)
```

Exercise. Check what you have read by typing “wip” in the console window. What do you observe? Type “class(wip)” and “str(wip)” to confirm that “wip” is of class “data.table”.

6. Data Wrangling Aims

We can simplify the production of summaries and plots by restructuring the current wip dataset (which has 65 columns) to the following format:

Country	Year	pctWiP
Country AAA	1997	##.##
Country AAA	1998	##.##
Country AAA	1999	##.##
...		

pctWiP refers to the percentage of women in parliament.

Key information retained. These three columns will contain the same information as the wip dataset but in a more usable format. We will also add a variable for the ratio of male to female MPs.

Superfluous columns. We will start by removing columns V65, Indicator.Name and Indicator.Code. There are years without any data but they will be removed automatically later (when restructuring from “wide” to “long” format).

Column V65 is created automatically due to an extra comma at the end of the column names (fifth) line of WB-WiP.csv:

```
... , "2016", "2017", "2018", "2019",
```

Check. Before removing it check that all values are NA.

```
wip[, .N, by=.(V65)]
#      V65      N
# 1:  NA  264
```

Column Indicator.Name has the unique value “Proportion of seats held by women in national parliaments (%)” and in Indicator.Code it is “SG.GEN.PARL.ZS”. As there is only one indicator in this dataset we will remove these two columns.

Exercise. Confirm that both Indicator.Name and Indicator.Code have the same values for all observations. Hint: Use the approach above for variable V65

Removing columns. The indicator and V65 columns can be removed. We will also rename “Country.Name” as “Country” and “Country.Code” as “Code”.

```
wip[, c("Indicator.Name", "Indicator.Code",
        "V65"):=NULL]
setnames(wip, c("Country.Name", "Country.Code"),
        c("Country", "Code"))
head(names(wip))
# [1] "Country" "Code"    "X1960"   "X1961"
# [5] "X1962"   "X1963"
tail(names(wip))
# [1] "X2014" "X2015" "X2016" "X2017" "X2018"
# [6] "X2019"
```

Reshape to long format. We want to transform the data so that for each country the year (column) data becomes a row. At the same time we will remove the missing data (with the na.rm option).

```
WP <- melt(wip,
           id.vars = c("Country", "Code"),
           measure = patterns("^X"),
           variable.name = "YearC",
           value.name = c("pctWiP"),
           na.rm = TRUE)
```

```
WP
#      Country Code YearC pctWiP
# 1:      Angola AGO X1997  9.545
# 2:      Andorra AND X1997  7.143
# 3:    Arab World ARB X1997  3.747
# 4: United Arab Emirates ARE X1997  0.000
# ---
# 5152:      Yemen, Rep. YEM X2019  0.332
# 5153:    South Africa ZAF X2019 46.348
# 5154:      Zambia ZMB X2019 17.964
# 5155:     Zimbabwe ZWE X2019 31.852
```

Final tweaks to WP. Create a numeric Year variable and a Ratio of men to women in parliament.

```
WP[, `:=`(Year=as.numeric(gsub("X", "", YearC)),
        Ratio = (100-pctWiP)/pctWiP)][
  , YearC:=NULL]
setcolorder(WP, c("Country", "Code", "Year",
                  "pctWiP", "Ratio"))
# Look at the contents of WP
# Note: Dropped Country name due to space
WP[, .(Code, Year, pctWiP, Ratio)]
#      Code Year pctWiP Ratio
# 1:      AGO 1997  9.545  9.48
# 2:      AND 1997  7.143 13.00
# 3:      ARB 1997  3.747 25.69
# 4:      ARE 1997  0.000  Inf
# ---
# 5152:      YEM 2019  0.332 300.00
# 5153:      ZAF 2019 46.348  1.16
# 5154:      ZMB 2019 17.964  4.57
# 5155:      ZWE 2019 31.852  2.14
```

7. Questions

The objective is to look at the geographical and time trends in the data. We will answer the following questions.

- What are the time trends for Portugal?
- How does Portugal compare to other countries?
- Which countries have the highest percentage of women in parliament by year?
- How do continents compare?
- What are the global trends over time?

Exercise - Without Programming.

- Which country do you think has the highest percentage of women in parliament?
- In each continent (i.e. Africa, Americas, Asia, Europe and Oceania), which country has the highest percentage of women in parliament?
- What is the world percentage of women in parliament in 2019?

8. Exploratory Analysis

Select a country. This guide explores how Portugal performs over time and compared to other countries. Before continuing, select another country for yourself to repeat the examples and do the exercises.

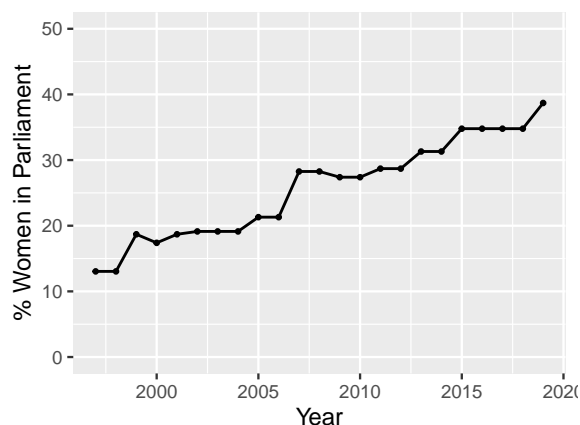
Time trends for Portugal. First look at the raw data.

```
WP[Country %in% "Portugal"]
```

#	Country	Code	Year	pctWiP	Ratio
# 1:	Portugal	PRT	1997	13.0	6.67
# 2:	Portugal	PRT	1998	13.0	6.67
# 3:	Portugal	PRT	1999	18.7	4.35
# 4:	Portugal	PRT	2000	17.4	4.75
# 5:	Portugal	PRT	2001	18.7	4.35
# 6:	Portugal	PRT	2002	19.1	4.23
# 7:	Portugal	PRT	2003	19.1	4.23
# 8:	Portugal	PRT	2004	19.1	4.23
# 9:	Portugal	PRT	2005	21.3	3.69
# 10:	Portugal	PRT	2006	21.3	3.69
# 11:	Portugal	PRT	2007	28.3	2.54
# 12:	Portugal	PRT	2008	28.3	2.54
# 13:	Portugal	PRT	2009	27.4	2.65
# 14:	Portugal	PRT	2010	27.4	2.65
# 15:	Portugal	PRT	2011	28.7	2.48
# 16:	Portugal	PRT	2012	28.7	2.48
# 17:	Portugal	PRT	2013	31.3	2.19
# 18:	Portugal	PRT	2014	31.3	2.19
# 19:	Portugal	PRT	2015	34.8	1.87
# 20:	Portugal	PRT	2016	34.8	1.87
# 21:	Portugal	PRT	2017	34.8	1.87
# 22:	Portugal	PRT	2018	34.8	1.87
# 23:	Portugal	PRT	2019	38.7	1.58
#	Country	Code	Year	pctWiP	Ratio

Visualisation. It is easier to find trends within a plot.

```
library(ggplot2)
library(magrittr)
WP[Country %in% "Portugal"] %>%
  ggplot(aes(Year, pctWiP)) +
    geom_line() + geom_point() +
    scale_y_continuous(limits=c(0, 50)) +
    ylab("% Women in Parliament")
```

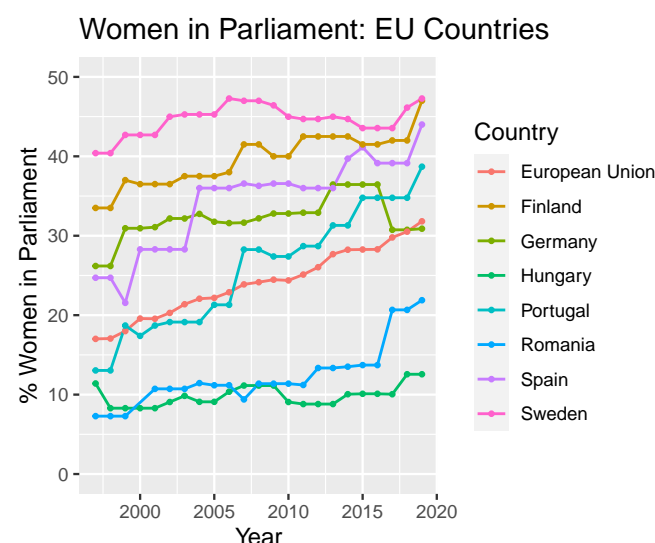


Interpretation. In 1997 Portugal had 13.0% women in parliament (i.e. 12.2 men for each woman), which increased to 38.7% (i.e. 1.58 men for each woman) in 2019. This still falls short of 50% (i.e. point of gender parity in parliament).

Exercise. For your chosen country look at the time trend data and the plot. What is your interpretation? How does it compare to Portugal?

Portugal versus European Union (EU) countries. We selected six EU countries (due to space limitations) for comparison. It would be better to compare all EU and/or all European countries.

```
WP[Country %in% c("Portugal", "Sweden", "Spain",
                  "Hungary", "Romania", "Finland", "Germany",
                  "European Union")] %>%
  ggplot(aes(Year, pctWiP, colour=Country)) +
    geom_line() +
    geom_point() +
    scale_x_continuous(breaks=seq(1995, 2020, 5)) +
    scale_y_continuous(limits=c(0, 50),
                      breaks=seq(0, 50, by=10)) +
    ggtitle("Women in Parliament: EU Countries") +
    ylab("% Women in Parliament")
```



Interpretation. Since 2007 Portugal has had more women in parliament than the European Union average. The key point to note is that none of these countries reaches equality between males and females in parliament, although Sweden and Finland come closest.

A couple of points to note.

“Germany”. In October 1990, the process of “German reunification” led to the creation of Germany, which united the former “German Democratic Republic” (East Germany) and the “Federal Republic of Germany” (West Germany). Therefore, since reunification, the data is presented for the reunified “Germany” only. Careful thought should be given to handling, analysing and interpreting any pre-reunification data (if available).

“European Union”. The “European Union” has changed over time (unlike the “continent of Europe”). It started in the 1950s as a block of six European countries (known as the “European Community”) and has expanded over the years to 28 countries (with the United Kingdom about to depart). This raises the question of how the European Union average is calculated. For a given year, is it calculated based on the actual member states in that year or on all of the current member states?

Exercises. Compare the country of your choice to four or five other countries by plotting a line graph similar to the one above.

Countries with the highest percentage of women in parliament.

A quick answer can be obtained by looking at the highest percentages.

```
WP[order(-pctWiP), head(.SD, 10)]
#      Country Code Year pctWiP Ratio
# 1: Rwanda RWA 2013  63.8 0.569
# 2: Rwanda RWA 2014  63.8 0.569
# 3: Rwanda RWA 2015  63.8 0.569
# 4: Rwanda RWA 2016  63.8 0.569
# 5: Rwanda RWA 2017  61.2 0.633
# 6: Rwanda RWA 2018  61.2 0.633
# 7: Rwanda RWA 2019  61.2 0.633
# 8: Rwanda RWA 2008  56.2 0.778
# 9: Rwanda RWA 2009  56.2 0.778
# 10: Rwanda RWA 2010  56.2 0.778
```

Data speaks. Are you surprised? Data can be very enlightening.

Highest percentage by year. Which countries have the highest percentage of women in parliament by year?

```
WP[order(Year, -pctWiP), head(.SD, 1), by = Year]
#      Year Country Code pctWiP Ratio
# 1: 1997 Sweden SWE  40.4 1.475
# 2: 1998 Sweden SWE  40.4 1.475
# 3: 1999 Sweden SWE  42.7 1.342
# 4: 2000 Sweden SWE  42.7 1.342
# 5: 2001 Sweden SWE  42.7 1.342
# 6: 2002 Sweden SWE  45.0 1.223
# 7: 2003 Rwanda RWA  48.8 1.051
# 8: 2004 Rwanda RWA  48.8 1.051
# 9: 2005 Rwanda RWA  48.8 1.051
# 10: 2006 Rwanda RWA  48.8 1.051
# 11: 2007 Rwanda RWA  48.8 1.051
# 12: 2008 Rwanda RWA  56.2 0.778
# 13: 2009 Rwanda RWA  56.2 0.778
# 14: 2010 Rwanda RWA  56.2 0.778
# 15: 2011 Rwanda RWA  56.2 0.778
# 16: 2012 Rwanda RWA  56.2 0.778
# 17: 2013 Rwanda RWA  63.8 0.569
# 18: 2014 Rwanda RWA  63.8 0.569
# 19: 2015 Rwanda RWA  63.8 0.569
# 20: 2016 Rwanda RWA  63.8 0.569
# 21: 2017 Rwanda RWA  61.2 0.633
# 22: 2018 Rwanda RWA  61.2 0.633
# 23: 2019 Rwanda RWA  61.2 0.633
#      Year Country Code pctWiP Ratio
```

Merging continent. The variable Country in the WP dataset is a mix of countries and regions (e.g. “European Union”, “South Asia” and “World”). To present the highest percentages grouped by continent we need to add it. Luckily, given the large number of R packages available, we can merge the “continent” from the “codelist” dataset in the “countrycode” package.

```
# Ensure that 'countrycode' package is installed.
# install.packages("countrycode")
library(countrycode)
cl <- as.data.table(codelist)[, .(continent, wb)]
setnames(cl, c("continent"), c("Continent"))
cWP <- merge(WP, cl, by.x = "Code", by.y = "wb",
             all.x = TRUE)
```

Highest percentages by year and continent. Which countries have the highest percentages in 1997 and 2019?

```
cWP[Year %in% c(1997, 2019) & !is.na(Continent)][
  order(Year, -pctWiP), head(.SD, 1),
  by = .(Year, Continent)][
  order(Continent, Year),
  .(Continent, Year, Country, pctWiP)]
#      Continent Year      Country pctWiP
# Africa 1997      Seychelles  27.3
# Africa 2019      Rwanda      61.2
# Americas 1997     Argentina  27.6
# Americas 2019      Cuba      53.2
# Asia 1997      Vietnam      26.2
# Asia 2019 United Arab Emirates  50.0
# Europe 1997      Sweden      40.4
# Europe 2019      Sweden      47.3
# Oceania 1997     New Zealand  29.2
# Oceania 2019     New Zealand  40.8
```

Decline in percentage. Which countries have had a decline in percentage of women in parliament since their first and last available measurement?

```
dWP <- cWP[order(Country, Year), .SD[c(1, .N)],
  by=Country][,
  pctDiff := pctWiP - shift(pctWiP, by=Country)[
  pctDiff<0][
  order(pctDiff)]
dWP[!is.na(Continent),
  .(Country, pctWiP, pctDiff)]
#      Country pctWiP pctDiff
# 1:      Seychelles 21.212 -6.0606
# 2:      Mali      9.524 -2.7211
# 3: Korea, Dem. Peoples Rep. 17.613 -2.4745
# 4:      Bahamas, The 12.821 -2.1795
# 5:      Tuvalu      6.250 -2.0833
# 6:      Congo, Dem. Rep. 10.000 -2.0000
# 7:      Papua New Guinea 0.000 -1.8349
# 8:      Maldives      4.598 -1.6523
# 9:      Haiti      2.542 -1.0721
# 10:      Congo, Rep. 11.258 -0.7417
# 11:      Yemen, Rep.  0.332 -0.3322
# 12:      Oman      2.326 -0.0841
# 13:      Nigeria      3.380 -0.0385
```

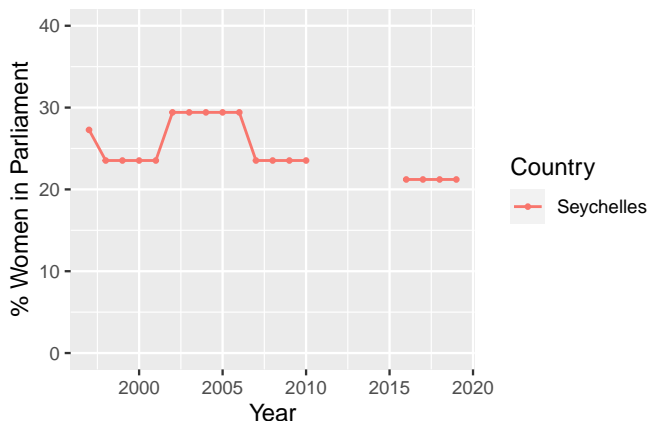
Visualisation. We will plot the trend lines for countries with at least a 5% decline. Note that the “5%” is arbitrarily selected.

```
# Select the countries to plot
dclpct <- unique(dWP[!is.na(Continent) &
  pctDiff <= -5]$Country)

WP[Country %in% dclpct] %>%
  ggplot(aes(Year, pctWiP, colour=Country)) +
  geom_line() +
  geom_point() +
  scale_x_continuous(breaks=seq(1995, 2020, 5)) +
  scale_y_continuous(limits=c(0, 40),
    breaks=seq(0, 40, by=10)) +
  ggtitle("Women in Parliament: Decline >=5%") +
  ylab("% Women in Parliament")
```

```
# Warning: Removed 5 rows containing missing values
# (geom_point).
```


Women in Parliament: Decline >=5%



Interpretation. Only Seychelles had a decline to greater than 5% which could be due a low number of seats in the parliament.

Ranked status. Another way to look at the data is to look at the ranking of countries, which could be done at a global level or by continent. Nonetheless, the results should be interpreted with caution and an understanding of the actual percentages. For example, if most countries were around the 50% mark, rankings could be misleading and subject to random fluctuations.

Global ranks by year. We will rank the countries by year based on the percentage of women in parliaments. The countries with the highest percentage will be ranked first and the lowest last. A total for the number of countries with data is included as it varies by year.

```
cWP[!is.na(Continent),
      `:=`(RankG = rank(-pctWiP), TotalG = .N),
      by = .(Year)]
```

Global ranking – Portugal.

```
cWP[Country=="Portugal",
      .(Country, Year, pctWiP, Ratio, RankG, TotalG)][
      order(Year)]
```

#	Country	Year	pctWiP	Ratio	RankG	TotalG
# 1:	Portugal	1997	13.0	6.67	41.5	160
# 2:	Portugal	1998	13.0	6.67	49.5	162
# 3:	Portugal	1999	18.7	4.35	24.0	153
# 4:	Portugal	2000	17.4	4.75	34.0	157
# 5:	Portugal	2001	18.7	4.35	33.0	167
# 6:	Portugal	2002	19.1	4.23	42.0	161
# 7:	Portugal	2003	19.1	4.23	46.0	174
# 8:	Portugal	2004	19.1	4.23	54.0	181
# 9:	Portugal	2005	21.3	3.69	46.0	185
# 10:	Portugal	2006	21.3	3.69	50.0	189
# 11:	Portugal	2007	28.3	2.54	28.0	188
# 12:	Portugal	2008	28.3	2.54	28.0	187
# 13:	Portugal	2009	27.4	2.65	33.0	187
# 14:	Portugal	2010	27.4	2.65	34.0	187
# 15:	Portugal	2011	28.7	2.48	31.0	188
# 16:	Portugal	2012	28.7	2.48	35.0	189
# 17:	Portugal	2013	31.3	2.19	36.0	184
# 18:	Portugal	2014	31.3	2.19	36.0	187
# 19:	Portugal	2015	34.8	1.87	29.0	189
# 20:	Portugal	2016	34.8	1.87	27.0	192
# 21:	Portugal	2017	34.8	1.87	28.0	190
# 22:	Portugal	2018	34.8	1.87	29.0	191

```
# 23: Portugal 2019 38.7 1.58 28.0 192
#          Country Year pctWiP Ratio RankG TotalG
```

Interpretation. Portugal has generally been ranked in the first quartile (25%) of countries in the world, with the fluctuations of its ranking most likely due to random variation.

Exercise. For your chosen country, interpret its ranking over the years. How does it compare to Portugal?

Continent ranks by year. We will rank the countries by year within a continent based on the percentage of women in parliaments. The countries with the highest percentage will be ranked first and the lowest last. A total for the number of countries with data, within each continent, is included as it varies by year.

```
cWP[!is.na(Continent),
      `:=`(RankC = rank(-pctWiP), TotalC = .N),
      by = .(Continent, Year)]
```

Portugal's ranking in Europe.

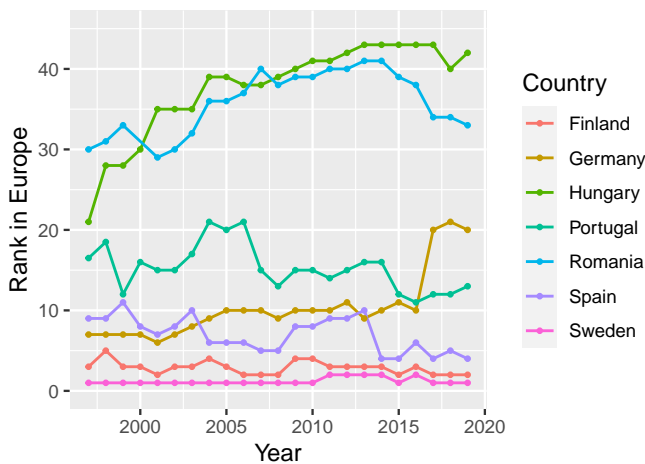
```
cWP[Country=="Portugal",
      .(Country, Year, pctWiP, Ratio, RankC, TotalC)][
      order(Year)]
```

#	Country	Year	pctWiP	Ratio	RankC	TotalC
# 1:	Portugal	1997	13.0	6.67	16.5	38
# 2:	Portugal	1998	13.0	6.67	18.5	36
# 3:	Portugal	1999	18.7	4.35	12.0	37
# 4:	Portugal	2000	17.4	4.75	16.0	37
# 5:	Portugal	2001	18.7	4.35	15.0	40
# 6:	Portugal	2002	19.1	4.23	15.0	38
# 7:	Portugal	2003	19.1	4.23	17.0	40
# 8:	Portugal	2004	19.1	4.23	21.0	41
# 9:	Portugal	2005	21.3	3.69	20.0	41
# 10:	Portugal	2006	21.3	3.69	21.0	43
# 11:	Portugal	2007	28.3	2.54	15.0	43
# 12:	Portugal	2008	28.3	2.54	13.0	43
# 13:	Portugal	2009	27.4	2.65	15.0	43
# 14:	Portugal	2010	27.4	2.65	15.0	43
# 15:	Portugal	2011	28.7	2.48	14.0	43
# 16:	Portugal	2012	28.7	2.48	15.0	43
# 17:	Portugal	2013	31.3	2.19	16.0	43
# 18:	Portugal	2014	31.3	2.19	16.0	43
# 19:	Portugal	2015	34.8	1.87	12.0	43
# 20:	Portugal	2016	34.8	1.87	11.0	43
# 21:	Portugal	2017	34.8	1.87	12.0	43
# 22:	Portugal	2018	34.8	1.87	12.0	43
# 23:	Portugal	2019	38.7	1.58	13.0	43

Plot of Portugal's ranking in Europe. Below we reproduce the percentage plot to show how Portugal ranks in relation to six other European countries. Note that the highest percentage is ranked first and the lowest last.

```
cWP[Country %in% c("Portugal", "Sweden", "Spain",
                    "Hungary", "Romania", "Finland", "Germany")] %>%
  ggplot(aes(Year, RankC, colour=Country)) +
  geom_line() +
  geom_point() +
  scale_x_continuous(breaks=seq(1995, 2020, 5)) +
  scale_y_continuous(limits=c(0, 45),
                      breaks=seq(0, 45, by=10)) +
  ggtitle("Women in Parliament: Ranked") +
  ylab("Rank in Europe")
```

Women in Parliament: Ranked



Interpretation. A total of 38 in 1997 and 43 in 2019. Within Europe, Portugal was typically ranked in the second quartile (25-50%) with the fluctuations of its ranking most likely due to random variation.

Exercise. How does your chosen country rank within its continent?

Highest rank by year and continent. Which countries have the highest rank in 1997 and 2019? The answer will coincide with the highest percentages (see above).

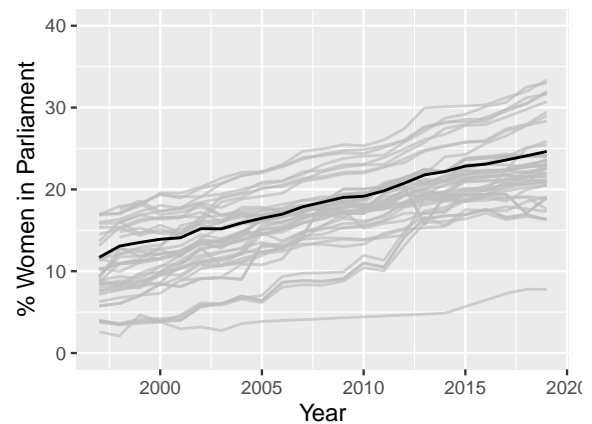
```
cWP[Year %in% c(1997, 2019) & RankC==1][
  order(Continent, Year),
  .(Continent, Year, Country, pctWiP)]
```

#	Continent	Year	Country	pctWiP
#	Africa	1997	Seychelles	27.3
#	Africa	2019	Rwanda	61.2
#	Americas	1997	Argentina	27.6
#	Americas	2019	Cuba	53.2
#	Asia	1997	Vietnam	26.2
#	Asia	2019	United Arab Emirates	50.0
#	Europe	1997	Sweden	40.4
#	Europe	2019	Sweden	47.3
#	Oceania	1997	New Zealand	29.2
#	Oceania	2019	New Zealand	40.8

Overall picture. What are the trends globally? There are various regions defined in the World Bank data. We can plot them and highlight the world “average”.

```
library(gghighlight)
cWP[is.na(Continent)] %>%
  ggplot(aes(Year, pctWiP, group=Country)) +
  geom_line() +
  gghighlight(Country=="World",
    use_direct_label = FALSE,
    use_group_by = FALSE) +
  scale_x_continuous(breaks=seq(1995, 2020, 5)) +
  scale_y_continuous(limits=c(0, 40),
    breaks=seq(0, 40, by=10)) +
  ggtitle("Women in Parliament: Global Trends") +
  ylab("% Women in Parliament")
```

Women in Parliament: Global Trends



Interpretation. The grey lines show that regardless of how we define region the general trends are upwards. The “World” percentage (black line) increased between 1997 and 2019. In 2018, women in parliament represented 24% (i.e. a ratio of 3.17 men to each woman), which is still less than half the level before gender parity can be claimed.

9. Conclusion

This guide presented an analysis of the percentage of women in parliament as a real-life case study for the data.table package. Although the format limited what could be presented, we can conclude that the percentage of women in parliament is increasing but that gender parity in parliaments is still far-off.

There is a lot more that can be said and discussed about the limitations, interpretation and potential impact of this data which the World Bank has nicely summarised.⁶ You are strongly encouraged to read their discussion for a more complete understanding.



⁶<https://databank.worldbank.org/data/reports.aspx?source=2&type=metadata&series=SG.GEN.PARL.ZS>.