

**An Introduction to R**  
R Workshop 6: World Bank Data  
QS311 Section E: Baker University  
November 18th, 2019

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Lab Procedures</b>	<b>2</b>
2.1	Obtaining the Variables of Interest . . . . .	2
2.2	Obtaining a Panel with All Variables . . . . .	3
2.3	Variables . . . . .	3
<b>3</b>	<b>Analyzing the Data</b>	<b>4</b>
3.1	Questions . . . . .	4

# 1 Introduction

## Mapping One Variable

R Workshop in World Bank Global Gender Justice Data  
QS311 Section E: Baker University

*Goal: To explore global gender data from the World Bank in R.*

In this workshop we will focus on creating and loading World Bank data sets into R as well as some of the basic mapping capabilities in R.

1. To access World Bank data we need to use the API (Application Programming Interface). R has a package for the World Bank API called *wbstats*. Recall: if you use a computer without the API, you must install the package with `install.packages("wbstats")`. Once you have installed the package, remember to run the library command to call it into your session on R with the `library(wbstats)` command.

## 2 Lab Procedures

### 2.1 Obtaining the Variables of Interest

Steps:

1. Remember to run the library for `wbstats`: `library("wbstats")`.
2. Now, read in the data for each indicator code. For example, to get the domestic violence legislation indicator and call the resulting data frame `dv_leg`, you would use the following code (notice that the date is different from before):

```
dv_leg <- wb(country = "countries_only", indicator =  
  "SG.LEG.DVAW", startdate = 1960, enddate = 2017,  
  removeNA = FALSE)
```

The following are the indicator codes that you need. Call each data frame a different name. In parentheses are the names that I used for each data frame in my code.

(a) Human Capital

- SL.TLF.CACT.FE.ZS (*lf\_fempar*)
- SL.TLF.TOTL.FE.ZS (*lf\_per\_fem*)
- SE.TER.ENRR.FE (*tertiary\_female*)

(b) Well-being

- SG.LEG.DVAW (*dv\_leg*)
- SG.GEN.PARL.ZS (*np\_per\_fem*)
- SG.DMK.ALLD.FN.ZS (*three\_decisions*)

## 2.2 Obtaining a Panel with All Variables

1. Create a panel comprising each of these indicators. Remember that we can combine individual columns from different data frames into a single data frame with the variables we want. The following code enables this:

```
panel <- data.frame(dv_leg$iso3c, dv_leg$date,  
dv_leg$value, np_perfem$value, three_decisions$value,  
lf_fempar$value, lf_perfem$value, tertiary_female$value)
```

2. Rename the variables in the panel with something similar to the following code:

```
names(panel) <- c("country_code", "year", "dv_leg", "np_perfem",  
"three_decisions", "lf_fempar", "lf_perfem", "tertiary_female")
```

3. Recall that it is often useful to have a numeric year variable instead of a factor year variable. Create this with code similar to the following:

```
panel$year_num <- as.numeric(panel$year)
```

4. Finally, load in the the WB classifications data by importing an excel file. Then, utilize the merge command to combine the panel and WB classifications data frames.

```
final_panel <- merge(panel, wb_classifications, by.x = "country_code",  
by.y = "Code")
```

## 2.3 Variables

The full dataset has many variables based on the two merges. However, there are six primary variables about gender that we want to examine. They are listed below:

Table 1. Summary of Variables

Variable	Description
<b>Human Capital</b>	
<i>lf_per_fem</i>	Labor force participation rate, female (% of female population ages 15+) (modeled ILO estimate)
<i>tertiary_female</i>	School enrollment, tertiary, female (% gross)
<i>lf_fempar</i>	Labor force, female (% of total labor force)
<b>Well-being</b>	
<i>dv_leg</i>	Legislation exists on domestic violence (1=yes; 0=no)
<i>np_per_fem</i>	Proportion of seats held by women in national parliaments (%)
<i>three_decisions</i>	Women participating in the three decisions (own health care, major household purchases, and visiting family) (% of women age 15-49))

## 3 Analyzing the Data

### 3.1 Questions

1. Look at the overall labor force participation rate for women over time (i.e. create a fitted line). Now, create a fitted line by region. Do the same for the fraction of the labor force that is female. What is a human rights perspective on the heterogeneous participation rates.
2. Create a line graph (*geom\_line*) by country code of female enrollment in tertiary education for Sub-Saharan African countries. What does this say about the ability of women in this set of countries to participate in the workforce?
3. Create a boxplot for percent of women participating three of the major decisions around autonomy. What does an ethics of care perspective say about the geographic variation.
4. Generate a heat map (remember that this will require you to load in the coordinates that we learned about in the environmental data) of the world to look at the percent of women participating in national parliaments. How does the concept of free will apply to the ability to run for office? How does this go beyond the legal right to be an elected official?
5. Use the column chart command in R to see the percentage of countries by region and income group that have domestic violence legislation.