INFO 7374 Assignment 1 : Kiva, WDI and WGI Datasets

Leela Gangadhar Vallabhaneni, Anudeep Chimakurthi, Anudeep Dasari, Anusha Modala

8 June 2016

Contents

1. Introduction

Using Kiva, WDI and WGI datasets we are going to explore the different aspects of how loans funded through Kiva are affecting people in countries with large loan count. This report is going to answer the following questions and each question will be mentioned prior to the doing the analysis required to answer them.

Q1. Which 5 countries have the largest number of loans funded?
Q2. Which gender has a high number of loans in those 5 countries?
Q3. What is Female unemployment rate (%) from 2005 to 2011 in these 5 countries ?
Q4. How is Female unemployment rate (%), Mortality rate (per 1,000 female adults), Ratio of female to male labor force participation rate (%), the Female labor force with secondary education(%) affected due to the loans funded to the female population in the Philippines?
Q5. Are the loans being paid back to lenders?
Q6. Is the voice and accountability of the people boosted due to loans?

To answer the above questions we are going to leverage various indicators from WDI and WGI datasets. Questions will be answered using visual plots and their explanation. Summaries for each question and information found due to the investigation will be provided.

2. Dataset description

We used three datasets - KIVA, WDI, WGI to answer the questions.

- Kiva is an international non-profit organization with a mission to connect people through lending to alleviate property. Kiva works with the microfinance institutions on five continents to provide loans to people without internet access.They are primarily funded through the support of lenders making optional donations. It contains 826,609 observations where each observation represents loan information like status, amount, date, location, borrower etc.

- World Development Indicators (WDI) provides recent global development available incorporating national, regional and global estimates. The World Bank's Open Data site provides access to the WDI database free of charge to all users. It contains 1,736 observations where each observation represents a country, year and a specific indicator.
- Worldwide Governance Indicators (WGI) contains aggregate and individual governance indicators for 213 economies over the period 1996-2009, for six dimensions of governance: Voice and Accountability; Political Stability and Absence of Violence/Terrorism; Government Effectiveness; Regulatory Quality; Rule of Law; Control of Corruption. It contains 1,736 observations where each observation represents a country, year and a specific indicator.

The following gives a brief overview describing each variable.

2.1 Description of variables from Kiva DataSet

Variable	Type	Description
`id`	numeric (no units)	The unique identifier(number) of the loan payment
`status`	character	A categorical value representing the status of the loan with various levels; "inactive_expired", "paid", "defaulted", "refunded", "deleted", "expired", "reviewed"
`sector`	character	The sector name of the requested loan
`funded_amount`	numeric (in $)	The amount that has been purchased by Kiva
`basket_amount`	numeric (in $)	Amount of the loan which lenders have saved in shopping baskets, but has not been confirmed as purchased
`paid_amount`	numeric (in $)	The amount of the loan that has been paid off
`delinquent`	logical	Whether or not the loan has turned delinquent
`country`	character	The name of the country
`country.code`	character	Unique identification code of the country
`location.town`	character	The name of the town
`loan_amount`	numeric (in $)	The amount of money disbursed to the borrower
`lender_count`	numeric (no units)	Number of lenders for this loan
`borrowers.male_count`	numeric (no units)	The number of male borrowers
`borrowers.female_count`	numeric (no units)	The number of female borrowers
`funded_date`	character	The date when loan was fully funded on Kiva
`paid_date`	character	The date when loan was full paid by the borrower

2.2 Description of variables from WDI and WGI Dataset

Variable	Type	Description
`female.unemploy.rate`	double (percentage)	Female Unemployment Rate derived from the female population
`female.labor.percent`	double (percentage)	Percentage of female labour force with secondary education
`female.mortality.rate`	double (no units)	Female mortality rate; average from 1000 females
`female.male.rate`	double (percentage)	Female to Male labour particiption rate (in %)
`voice.accountability.rank`	numeric (no units)	percentile rank, w.r.t all countries; higher the better
`country`	character	The name of the country
`country.code`	character	Unique identification code of the country
`region`	character	The name of the region
`longitude`	factor(with 211 levels)	The longitude position of the region

Variable	Type	Description
`latitude`	factor(with 211 levels)	The latitude position of the region
`year`	character	Year in which the respective indicator is taken

3. Dataset Preparation

Preparing a dataset with data from all three datasets requires a lot of data wrangling packages which will be described in the coming sections. Visualization of results, Parallel processing of data requires additional R packages which are explained below.

3.1.1 `magrittr` package

The `magrittr` package offers a set of operations which promote semantics that improves the code by structuring sequences of data operations left-to-right and avoiding nested function calls. We used pipe operator (%>%) that helped to pipe a value forward into an expression or function call.

```
library(magrittr)
```

3.1.2 `dplyr` package

The `dplyr` package is one of the popular data manipulation tools needed for data analysis in R.
It provides a uniform interface to work with any kind of data i.e., data in a data frame, a data table or a database.

```
library(dplyr)
```

3.1.3 `ggplot2` package

The `ggplot2` package is data visualization package for the statistical programming language R.
It is based on the grammar of graphics, the idea that you can build every graph from the same few components : a data set, a set of geoms - visual marks that represent data points, and a coordinate system.

```
library(ggplot2)
```

3.1.4 `rlist` package

The `rlist` package makes it easier to work with the lists by providing a wide range of functions that operate on non-tabular data stored in them. Most functions in the package are designed to be pipeline friendly so that data processing with lists can be chained.

```
library(rlist)
```

3.1.5 `parallel` package

The `parallel` package helps in creating clusters to do parallel processing. Since the data is large and takes more time to process sequentially, this package helps in creating a cluster with multiple cores which can be used in reading data from JSON files to dataframe.

```
library(parallel)
```

3.1.6 arules and arulesViz packages

`arules` - Provides the infrastructure for representing,manipulating and analyzing transaction data and patterns required to create association rules.

```
library(arules)
```

`arulesViz` - Extends package arules with various visualization techniques for association rules and itemsets.

```
library(arulesViz)
```

3.2 Loading json data into a dataframe

Read all the JSON files with selected columns and saved into a data frame. Assuming we are in the directory that contains the loans folder, we are going to get the list of JSON file names from the directory.

```
data.folder <- "loans/"

loans.file <- list.files(data.folder, pattern="*.json")
```

`dfrow.from.list` is the function that converts a list to a dataframe.

```
dfrow.from.list = function(aList) {
  data.frame(rbind(unlist(aList)),
             stringsAsFactors=FALSE)
}
```

`readJSONFileIntoDataFrame` is the function that will read read the JSON file from a given path, converts to list by selecting required parameters and then uses the `dfrow.from.list` function to parse the list to dataframe.

```
readJSONFileIntoDataFrame <-
  function (filename) {
    # filename="1.json"
    if(file.size(paste(data.folder,
                       filename,
                       sep="")) != 0){
    paste(data.folder,
          filename,
          sep="") %>%
      fromJSON() %>%
{ .$loans } %>%
  list.select(status, sector, funded_amount,
              paid_amount,
              pending_amount = funded_amount - paid_amount,
              location.country_code = location$country_code,
              location.country = location$country,
              loan_amount, lender_count, funded_date, paid_date,
```

```
               gender = borrowers[[1]]$gender
  ) %>%
  lapply(dfrow.from.list) %>%
  bind_rows() %>%
  na.omit() %>%
    {
      x = .
      x$funded_year = format(
      as.Date.character(
        x$funded_date), "%Y"
      )
      x
    }
  }
}
```

Since the size of data is too large to read it sequentially, create a cluster for parallel processing. Create a cluster with 4 cores.

```
cl <- makeCluster(4)
```

Export all the variables and functions needed to the cluster.

```
clusterExport(cl,
              c('dfrow.from.list','data.folder','%>%',
                'list.select','bind_rows','fromJSON',
                'readJSONFileIntoDataFrame'))
```

Create a function called `create.loans.df.cl` to take the cluster, the list of all file paths and generate the loans dataframe.

```
create.loans.df.cl = function(cl, loans.file.in) {
  loans.file.in %>%
{ parLapply(cl, ., readJSONFileIntoDataFrame) } %>%
  bind_rows()
}
```

Now, run it ! Once completed stop the cluster.

```
system.time({
  loans.df = try(create.loans.df.cl(cl, loans.file))
})

stopCluster(cl)
```

`loans.df` is the dataframe that contains the KIVA loans data.

4.3 Setting the indicator codes for WDI

Set required indicator codes for analysis.

```r
library(WDI)

indicator.codes = c("SL.UEM.TOTL.FE.ZS",
                    "SP.DYN.AMRT.FE",
                    "SL.TLF.CACT.FM.ZS",
                    "SL.TLF.SECO.FE.ZS")
```

4.4 Extracting the dataset

Extracted the dataset pertaining to the above indicators.

```r
wdi_df <- WDI(indicator=indicator.codes,
              extra=TRUE
              )
wdi.df <- na.omit(wdi_df)
```

`wdi.df` is the dataframe that contains WDI indicators data along with country and year for each observation.

4.5 Merging of two datasets. KIVA and WDI.

Set the necessary column names which should be used to join the 2 dataframes.

```r
names(loans.df)[names(loans.df) == 'funded_year'] <- 'year'

names(loans.df)[names(loans.df) == 'location.country_code'] <- 'country.code'
names(wdi.df)[names(wdi.df) == 'iso2c'] <- 'country.code'
```

Change the indicator names to make them meaningful.

```r
names(wdi.df)[names(wdi.df) == 'SL.UEM.TOTL.FE.ZS'] <- "female.unemploy.rate"
names(wdi.df)[names(wdi.df) == "SL.TLF.CACT.FM.ZS"] <- 'female.male.rate'
names(wdi.df)[names(wdi.df) == 'SP.DYN.AMRT.FE'] <- 'female.mortality.rate'
names(wdi.df)[names(wdi.df) == 'SL.TLF.SECO.FE.ZS'] <- "female.labor.percent"
```

Merge the `loans.df` and `wdi.df` dataframes by common column names i.e, `country.code` and `year`. Clean the whole dataframe by removing the rows having NA fields.

```r
merged <- merge(loans.df, wdi.df,
                all.x=TRUE)

merged.df <- na.omit(merged)
```

`merged.df` is the dataframe that contains both KIVA and WDI data.

4. Single variable summaries and visualization

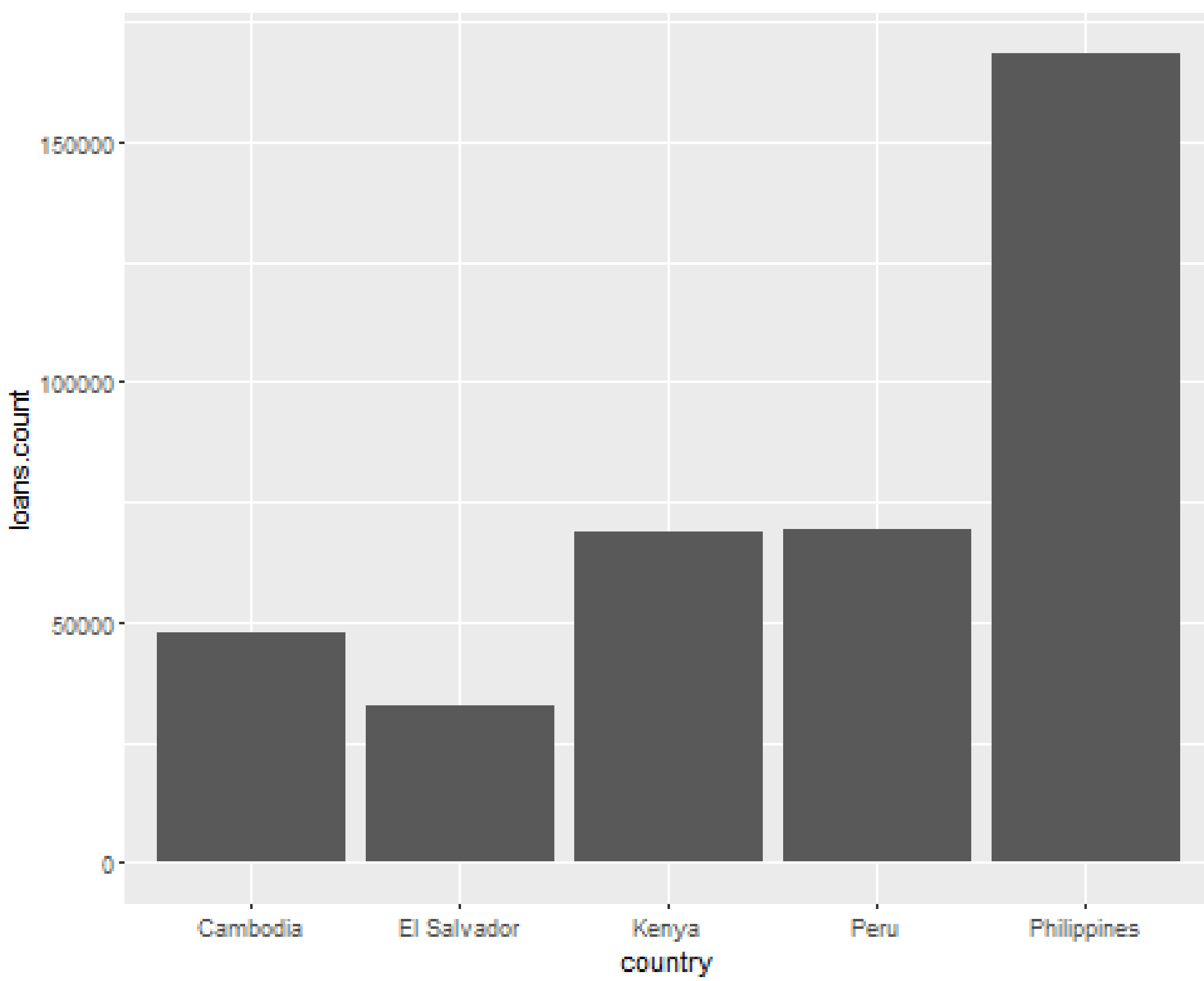
We are going to look into a single variable `country` and how many loans each country is funded.

4.1 Country vs Loans Count

The bar plot depicts the five countries having largest loan count.

Q1. Which 5 countries have the largest number of loans funded?

```
loans.df %>%
  group_by(country = country) %>%
  summarise(loans.count = n()) %>%
  .[order(-.$loans.count),] %>%
  .[1:5,] %>%
  ggplot  (aes(x    = country, y = loans.count)) +
  geom_bar(stat="identity")
```



As we see from the above graph, the top 5 countries with a large number of loans funded are Philippines, Peru, Kenya, Cambodia, El Salvador. The Philippines is top among all with 170,000 loans. The next country is Peru which has 70000 loans, which isn't even half of the loans funded to the Philippines. It shows how important the lenders feel the need for support to the people of Philippines. We are going to look into the data about these countries in the preceding sections.

5. Multiple variable summaries and visualization

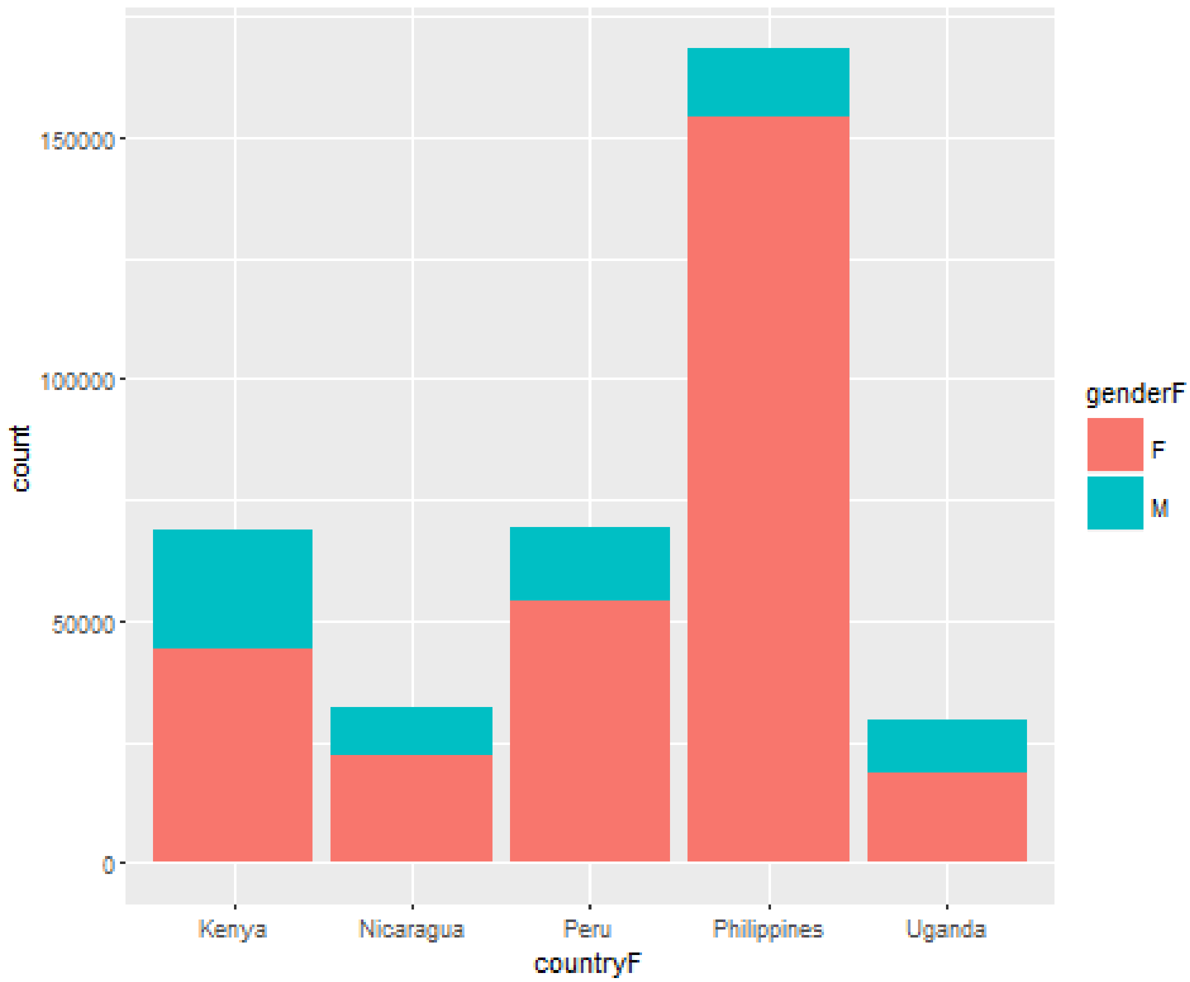
Let us look into how the loans funded to these countries are distributed according to gender.

5.1 Country vs Total Loan Count - Two factor variables - Male, Female loan count

The bar plot depicts the top 5 countries having the largest number of loans borrowed, distributed between males and females.

> Q2. Which gender has a high number of loans in those 5 countries?

```
loans.df %>%
  subset(country %in%
           c("Kenya","Nicaragua","Peru","Philippines","Uganda")) %>%
  {
    .$genderF = factor(.$gender)
    .$countryF = factor(.$country)
    .$sectorF = factor(.$sector)
    .$yearF = factor(.$year)
    .$statusF = factor(.$status)
    .
  } %>%
  {
    ggplot(data=.,aes(x=countryF)) +
      geom_bar(aes(fill=genderF))
  }
```



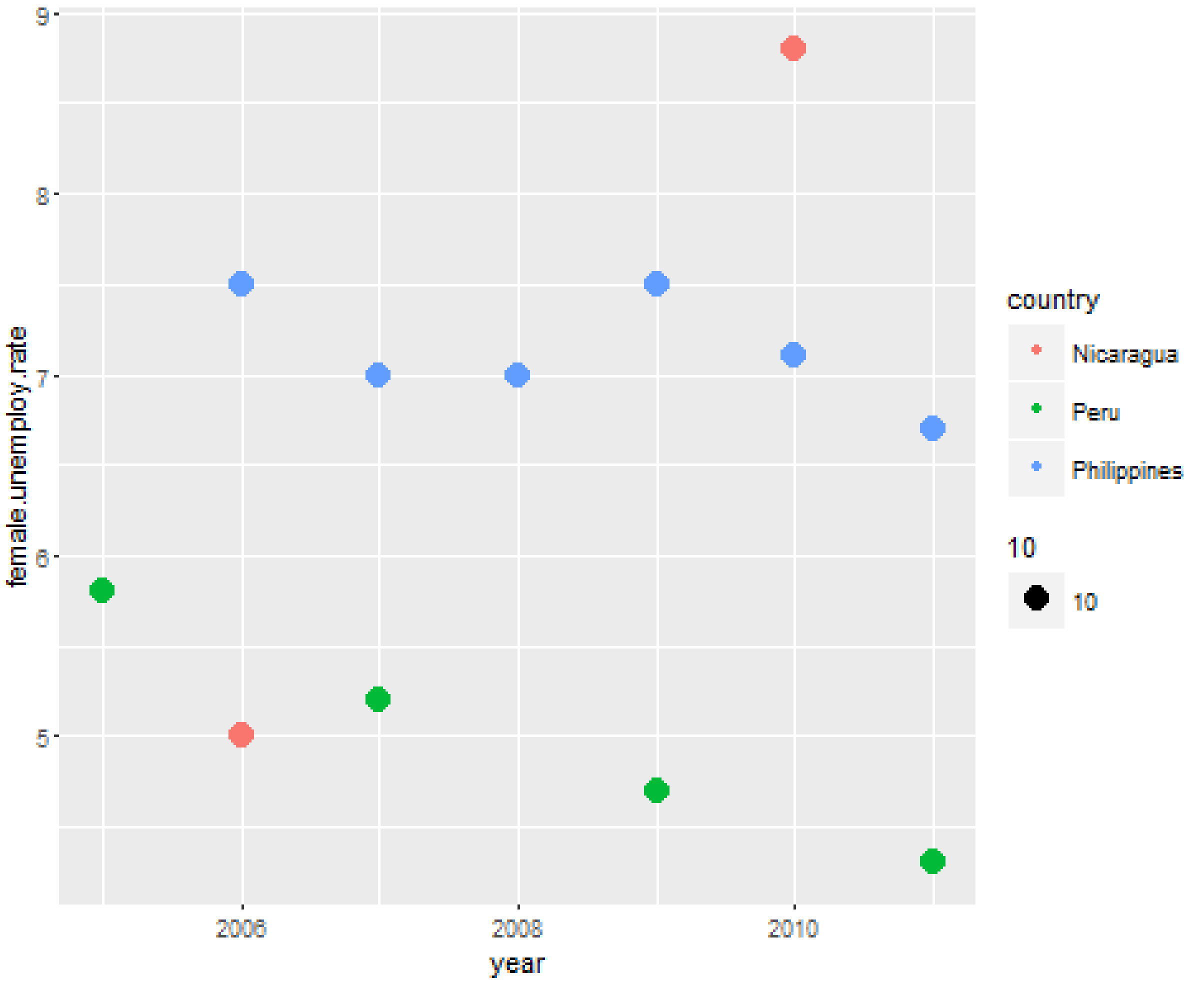
From the above it is very clear that females are funded the most. Let us concentrate our investigation more into the female population. Particularly in the Philippines the ratio of female to male loan count is very high when compared to other countries.

5.2 Year Vs Unemployment Rate in Women - Five-factor variables - Countries

The scatter plot depicts the female unemployment rates of countries with high loan count over the years, 2005 to 2011.

Q3. What is Female unemployment rate(%) from 2005 to 2011 in these 5 countries ?

```
wdi.df %>%
  subset(country %in%
           c("Kenya","Nicaragua","Peru","Philippines","Uganda")) %>%
  ggplot(aes(x=year, y=female.unemploy.rate)) +
  geom_point(aes(color=country, size=10))
```



The unemployment rate of females for different countries with large loan count is plotted from years 2005 to 2011. Philippines is the country with continuos data available. Further investigation needs continuos data over the years and more loans with respect to females. Since the Philippines satisfies both these conditions we are going to investigate the lives of females in the Philippines from the years 2008 to 2011. This should give us a better insight to whether the loans funded are actually making a difference in the lives of female population in the Philippines.

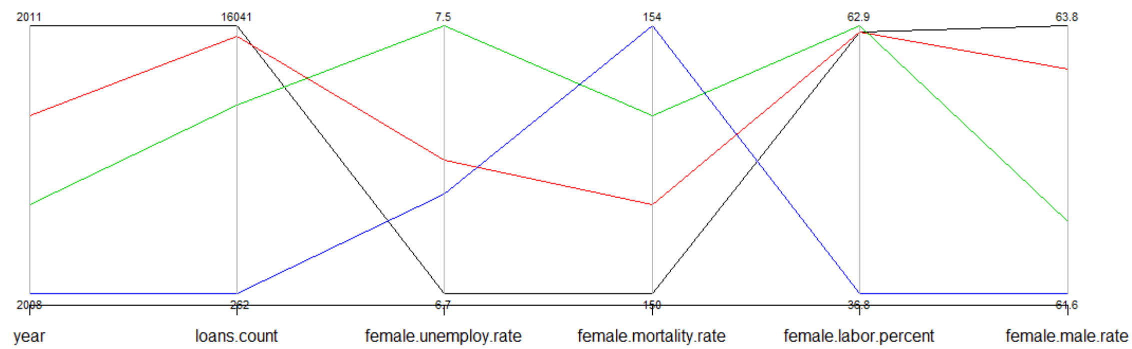
5.3 Analyzing multiple indicators which give insight into female lives in the Philippines

In particular, loans funded to the Philippines are mostly to women. So, we will dig more into how these loans are effecting female population of Philippines using the Parallel Coordinate Charts with the indicators taken from WDI data.

> Q4. How is Female unemployment rate(%), Mortality rate (per 1,000 female adults), Ratio of female to male labor force participation rate (%), Female labor force with secondary education(%) affected due to the loans funded to the female population in the Philippines?

```
merged.df %>%
  subset(country %in% c("Philippines") & gender %in% c("F")) %>%
  {
    .$yearF = factor(.$year)
```

```
  .
} %>%
group_by(year = year) %>%
summarise(loans.count = n(),
          female.unemploy.rate=mean(as.numeric(female.unemploy.rate)),
          female.mortality.rate=mean(as.numeric(female.mortality.rate)),
          female.labor.percent=mean(as.numeric(female.labor.percent)),
          female.male.rate=mean(as.numeric(female.male.rate))) %>%
.[order(-.$loans.count),] %>%
{
 .$year <- as.numeric(.$year)
  .
} %>%
parcoord(col = c(1,2,3,4), var.label=TRUE)
```



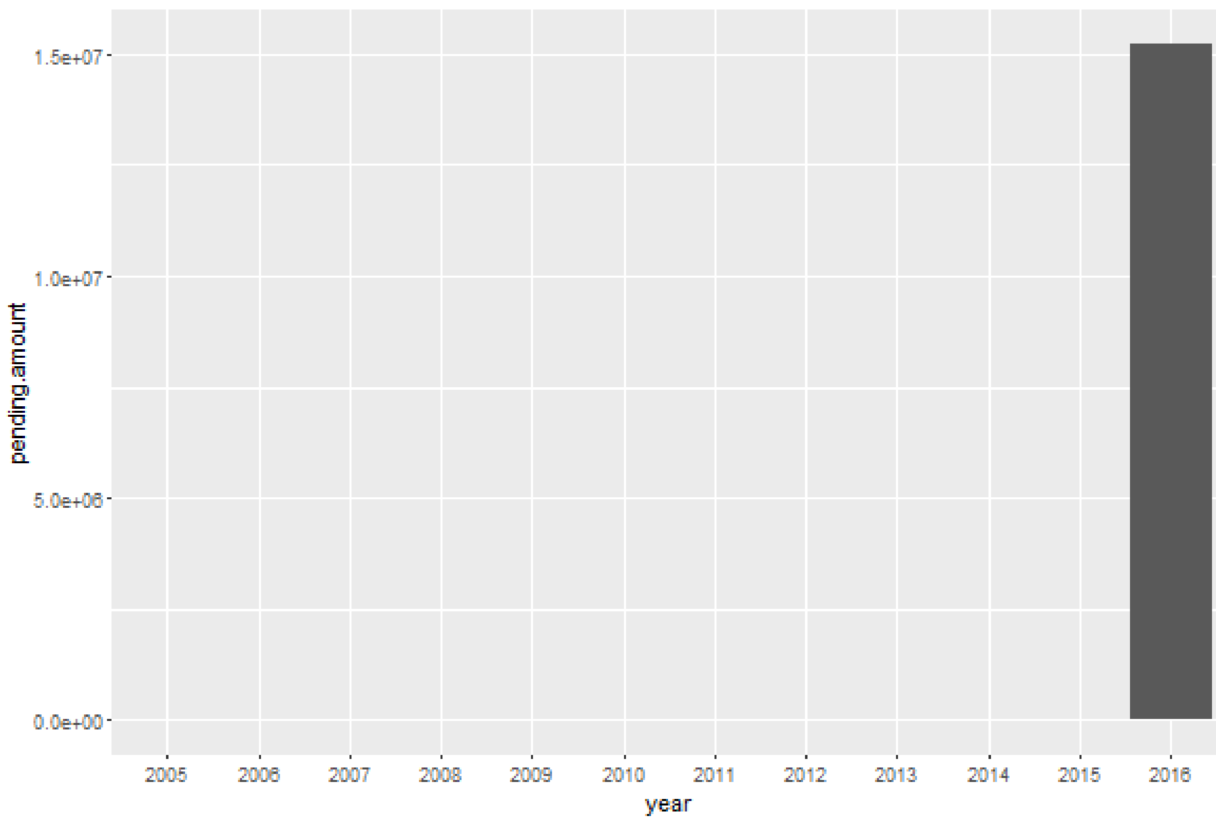
The parallel coordinate charts help us to see the change in values of multiple dimensions at the same time. As you can see in the above graph, the loans keep increasing through time.
The unemployment rate of female population although increases in the first two years to 7.5%, gradually goes to a low point, 6.7%. This shows that although females get a large number of loans, they require time to be employed. Why is that? The reason might be that the females are getting a secondary education due to the loans funded to other females in their households. The financial support is helping them to get proper education before they are employed.
The female mortality rate has been decreasing steadily over the years. This is a good sign of improved health conditions of females due to their financial support through the funded loans.
We can see that percentage of female labor force with secondary education increases from 2008 to 2009 and is stable from 2009 to 2011. The steady state from 2009 to 2011 shows the effect of the increase unemployment rate from 2008 to 2010, although there is a huge increase in the number of loans funded, is not increasing the value of this indicator. Since there has been an increase in unemployment rate the percentage of female labor force with secondary education is steady over the years without continuous increment.
The ratio of female to male work force is steadily increasing show the effect of loans on men when compared to women. Since women get more financial support it's encouraging more women to be part of this process and become employed.

5.4 Loans repayment

Let us explore if the loans being funded to the Philippines are actually being paid back to lenders.

Q5. Are the loans being paid back to lenders?

```r
loans.df %>%
  subset(country = "Philippines") %>%
  group_by(year) %>%
  summarise(pending.amount = sum(as.numeric(pending_amount)),
            funded.amount = sum(as.numeric(funded_amount)),
            paid.amount = sum(as.numeric(paid_amount))
            ) %>%
  ggplot(aes(x = year, y = pending.amount)) +
  geom_bar(stat="identity")
```



The graph shows a very high value for the year 2016 since its current and isn't yet complete. But the rest are paid back with very negligible unpaid loan amounts. We can conclude that the process of repaying the loans is being done smoothly. This is encouraging lenders to fund more loans over time.

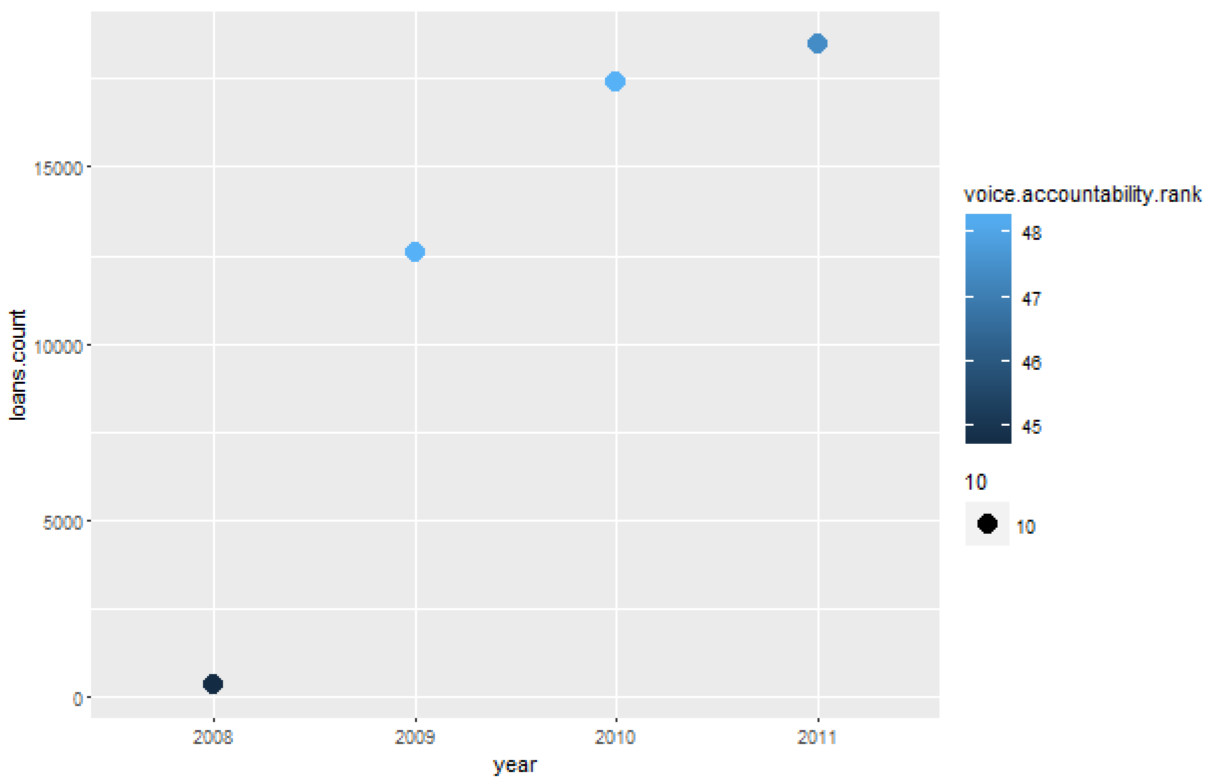
5.5 Voice and Accountability of citizens

From the information gathered till now, we can say that the funded loans are having an effect on the female population of Philippines. But how much of it is being used to actually increase their morale and be accountable citizens that will help them participate in the functioning of the nation. Let us look into the percentile ranks of Philippines from the year 2008 to 2011.

Q6. Is the voice and accountability of the people boosted due to loans?

```r
WDI(indicator="VA.PER.RNK",
    extra=TRUE) %>%
```

```
na.omit() %>%
{
  names(.)[names(.) == "VA.PER.RNK"] <- 'voice.accountability.rank'
  .
} %>%
merge(loans.df,
      .,
      all.x=TRUE) %>%
na.omit() %>%
subset(country %in% c("Philippines")) %>%
{
  .$yearF = factor(.$year)
  .
} %>%
group_by(year = year) %>%
summarise(loans.count = n(),
          voice.accountability.rank=mean(voice.accountability.rank)) %>%
.[order(-.$loans.count),] %>%
ggplot(aes(x=year, y=loans.count)) +
geom_point(aes(color=voice.accountability.rank, size=10))
```



The scatter plot clearly shows the increase in voice and accountability of citizens of Philippines is increased with the loans they are funded. We can conclude that not only the financial standards of the people have increased but the morality and responsibility of the people have increased over time. This shows us that the Philippines as a country is benefited from loans provided through Kiva.

6. Association rules

Association rules are used to define and analyze relationships between variables. They are useful for analyzing and predicting the behavior of the observations.

Definition of functions to create factor variables.

```
quantile.vector <- function(col){
  c(-Inf,
    quantile(col,
             c(0.25, 0.50, 0.75),
             na.rm=TRUE
    ),
    Inf
  )
}
```

We are going to use quantiles to help create itemsets.

```
# changes the range of quantiles into factors
quantile.factors <- function(col){
  cut(col,
      breaks = quantile.vector(col),
      labels=c('Q1','Q2','Q3','Q4')
      )
}
```

Implementing above functions on `merged.df` dataset for `female.unemploy.rate`, `female.male.rate`, `female.mortality.rate` for female population in Philippines.

```
rules.df <- merged.df %>%
  subset(country %in% c("Philippines") & gender %in% c("F")) %>%
{
  .$female.unemploy.rate.f = quantile.factors(.$female.unemploy.rate)
  .$female.mortality.rate.f = quantile.factors(.$female.mortality.rate)
  .$female.male.rate.f = quantile.factors(.$female.male.rate)
  .$year.f = factor(.$year)
  .
}
```

Using the `apriori` to mine frequent itemsets, association rules using the Apriori algorithm.

```
rules <- apriori(rules.df[,c("female.unemploy.rate.f","female.mortality.rate.f",
                             "female.male.rate.f",
                             "year.f")],
                 parameter = list("maxlen" = 3))
```

`rules` contains the lhs, rhs, support, confidence and lift. Sort `rules` by lift.

```
rules.sorted.by.lift = sort(rules, by="lift")
```

`rules.sorted.by.lift` contains all rules sorted by lift in decreasing order. Find Redundant Rules.

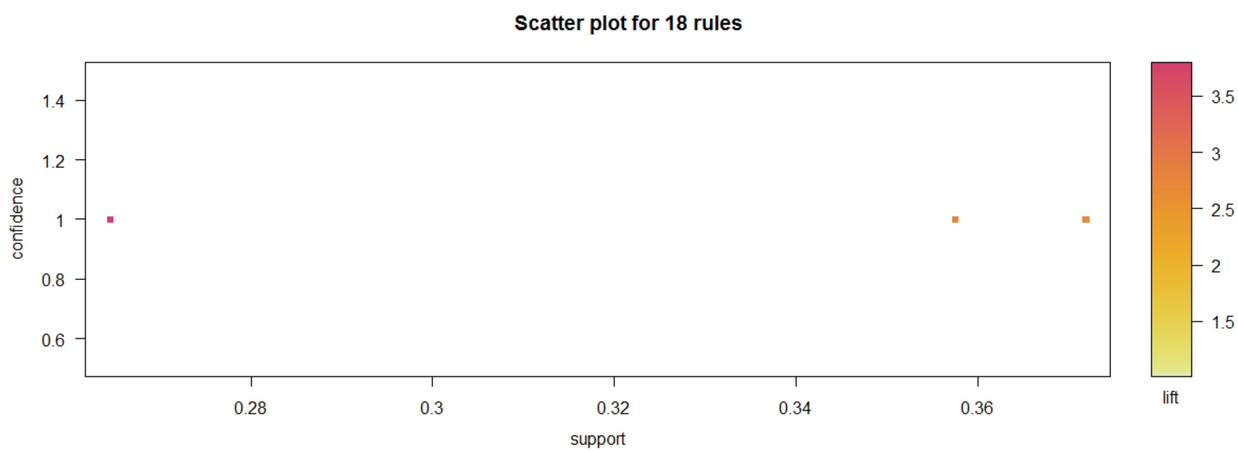
```
subset.matrix <- is.subset(rules.sorted.by.lift, rules.sorted.by.lift)
subset.matrix[lower.tri(subset.matrix, diag=T)] <- NA
redundant <- colSums(subset.matrix, na.rm=T) >= 1
```

`redundant` contains the indexes of the rules which don't give any new information. Remove of Redundant files.

```
rules.pruned <- rules.sorted.by.lift[!redundant]
inspect(rules.pruned)
```

`rules.pruned` is pruned set of rules. `inspect(rules.pruned)` gives all the rules derived after removing the redundant rules. Let us visualize a scatter plot of 18 rules w.r.t `support`, `confidence` and `lift`.

```
plot(rules.pruned)
```



The plot shows that the confidence is constant through all the rules. But lift decreases as support increases. This tells that more observations have LHS and RHS together are reducing the predictability of the rules.

Let us look into the quantile values for the three indicators in the female population of Philippines.

```
merged.df %>%
  subset(country %in% c("Philippines") & gender %in% c("F")) %>%
{
  list(female.unemploy.rate = quantile.vector(.$female.unemploy.rate),
       female.male.rate = quantile.vector(.$female.male.rate),
       female.mortality.rate = quantile.vector(.$female.mortality.rate)
       )
}
```

`female.unemploy.rate` - Q1 = (-Inf, 6.7] ; Q2 = [6.7, 7.1] ; Q3 = [7.1, 7.5] ; Q4 = [7.5, Inf)

`female.male.rate` - Q1 = (-Inf, 62.16898] ; Q2 = [62.16898, 63.43001] ; Q3 = [63.43001, 63.78446] ; Q4 = [63.78446, Inf)

`female.mortality.rate` - Q1 = (-Inf, 149.565] ; Q2 = [149.565, 150.894] ; Q3 = [150.894, 152.224] ; Q4 = [152.224, Inf)

The following are the insights drawn from the association rules

- Rule 1 -

```
        LHS                         RHS              support   confidence     lift
   {year.f=2009}  => {female.male.rate.f=Q1}   0.2644795       1          3.696118
```

This rule tells us the relationship between LHS, `year` at 2009, and RHS, `female.male.rate.f` at Q1. Support tells us that 26.44% of the total observations have LHS & RHS values, which is alright considering the amount of loans funded during the year 2009. The confidence of 1 tells us that all the transactions containing the LHS contain RHS, which means that `female.male.rate.f` in all observations containing `year.f` as 2009, is in Q1. Lift of 3.69 tells us that this rule occurs 3.69 times more often than expected. This rule tells us that in the year 2009, female to male ratio is lower since that was beginning of increased loans to the female population of Philippines.

- Rule 2 -

```
             LHS                            RHS                    support    confidence        lift
{female.mortality.rate.f=Q2} => {female.male.rate.f=Q2}         0.3575238         1           2.797017
```

This rule tells us the relationship between LHS, `female.mortality.rate.f` at Q2, and RHS, `female.male.rate.f` at Q2.
Support tells us that 35.75% of the total observations have LHS & RHS values, which makes sense because lower mortality rate of females has increased the female to male ratio in labor force. The confidence of 1 tells us that all the transactions containing the LHS contain RHS, which means that `female.male.rate.f` in all observations where `female.mortality.rate.f` is Q2, is in Q2. Lift of 2.79 tells us that this rule occurs 2.79 times more often than expected. This rule tells us that the decrease in `female.mortality.rate.f` correlates with increase in `female.male.rate.f`.

- Rule 3 -

```
     LHS                          RHS                 support    confidence     lift
{year.f=2010}  =>  {female.unemploy.rate.f=Q2}  0.3575238        1          2.750287
```

This rule tells us the relationship between LHS, `year.f` at 2010, and RHS, `female.unemploy.rate.f` at Q2. Support tells us that 35.75% of the total observations have LHS & RHS values, which is alright considering the amount of loans funded during the year 2010. The confidence of 1 tells us that all the transactions containing the LHS contain RHS, which means that `female.unemploy.rate.f` in all observations where `year.f` is 2010, is in Q2. Lift of 2.75 tells us that this rule occurs 2.75 times more often than expected. This rule tells us that as loan count increased with time the unemployment rate of females, `female.unemploy.rate.f`, gets to a lower level.

- Rule 4 -

```
             LHS                                RHS                    support    confidence      lift
{female.unemploy.rate.f=Q1}  => {female.mortality.rate.f=Q1}    0.3719221        1          2.688735
```

This rule tells us the relationship between LHS, `female.unemploy.rate.f` at Q1, and RHS, `female.unemploy.rate.f` at Q1.
Support tells us that 37.19% of the total observations have LHS & RHS values, which makes sense since we have seen female unemployment rate and female mortality rate moving together. The confidence of 1 tells us that all the transactions containing the LHS contain RHS, which means that `female.mortality.rate.f` in all observations where `female.unemploy.rate.f` is Q1, is in Q1. Lift of 2.68 tells us that this rule occurs 2.68 times more often than expected. This rule tells us that as the female unemployment rate decreased their mortality rate also decreased. This is a good indication that good financial support is helping them improve their living standards.

- Rule 5 -

```
             LHS                            RHS                support   confidence   lift
{female.male.rate.f=Q1} => {female.unemploy.rate.f=Q3}   0.2644795   0.9775473   3.696118
```

This rule tells us the relationship between LHS, `female.male.rate.f` at Q1, and RHS, `female.unemploy.rate.f` at Q3.
Support tells us that 26.44% of the total observations have LHS & RHS values. The confidence of 0.97 tells us that 97% the transactions containing the LHS contain RHS, which means that `female.unemploy.rate.f` in all observations where `female.male.rate.f` is Q1, is in Q3. This also indicates that female unemployment rate might be even more when female to male labor ratio is low. Lift of 3.69 tells us that this rule occurs 3.69 times more often than expected. This rule tells us that where ever the female to male labor ratio decreased the female unemployment rate increased.

7. Conclusion

Our work used the Kiva, WDI and WGI datasets to understand the impact of loans on the various countries, particularly the female population of Philippines. Here are the conclusions of our questions:

> Q1. Which 5 countries have the largest number of loans funded?

We have found that Philippines, Peru, Kenya, Cambodia and El Salvador are the countries with more funding than other countries.

> Q2. Which gender has a high number of loans in those 5 countries?

Females are funded the most. This holds true for all the countries mentioned in the previous question.

> Q3. What is Female unemployment rate(%) from 2005 to 2011 in these 5 countries ?

The Philippines is the country with continuos data available. Further investigation needs continuos data over the years and more loans with respect to females. Since the Philippines satisfies both these conditions we are going to investigate the lives of females in the Philippines from the years 2008 to 2011. This will give us a better insight to whether the loans funded are actually making a difference in the lives of female population of Philippines.

> Q4. How is Female unemployment rate(%), Mortality rate (per 1,000 female adults), Ratio of female to male labor force participation rate (%), Female labor force with secondary education(%) affected due to the loans funded to the female population in the Philippines?

The unemployment rate of female population although increases in the first two years to 7.5%, gradually goes to a low point, 6.7%. This shows that although females get a large number of loans, they require time to be employed. The financial support is helping them to get proper education before they are employed. The female mortality rate has been decreasing steadily over the years. This is a good sign of improved health conditions of females due to their financial support through the funded loans. The steady state of female labor force with secondary education from 2009 to 2011 shows the effect of the increase unemployment rate from 2008 to 2010, although there is a huge increase in the number of loans funded, is not increasing the value of this indicator. Since there has been an increase in unemployment rate the percentage of female labor force with secondary education is steady over the years without continuous increment. The ratio of female to male work force is steadily increasing show the effect of loans on men when compared to women. Since women get more financial support it's encouraging more women to be part of this process and become employed.

Q5. Are the loans being paid back to lenders?

We conclude that the process of repaying the loans is being done smoothly. This is encouraging lenders to fund more loans over time.

Q6. Is the voice and accountability of the people boosted due to loans?

We conclude that not only the financial standards of the people have increased but the morality and responsibility of the people have increased over time. This shows us that the Philippines as a country is benefited from loans provided through Kiva.

On an end note, we can confidently say that the loans being funded through Kiva to the Philippines are a success with respect to its female population and the general population.