Midterm Project - Kiva.org

Jordan Lardieri - Boston University

December 12, 2018

##

Abstract

My interest is in the social sciences, and the use of data science for social good. From improving global health to city infrastructure, we can use data to help solve major societal issues. In my career, I'd like to contribute to these efforts.

I chose my project because it aligns with this goal. I'm interested in *Kiva.org's* cause because they empower so many individuals and communities around the world through crowdsourced loans, not donations. Anyone can go to the website, get familiar with someone's story and make a loan contribution.

Their website has a number of focus areas; women, single parents, conflict zones, water, and education to name a few. As a nonprofit, *Kiva's* mission is to connect people through lending to help alleviate poverty. Kiva supports 3 million borrowers in more than 80 countries, creating opportunities for individuals, their families, and their communities.

My project aims to take a deeper look at the Kiva loan data, to see if there are any underlying themes and behaviors that differ between regions and countries. Specifically, is there variation in category and amount of funding and to whom.

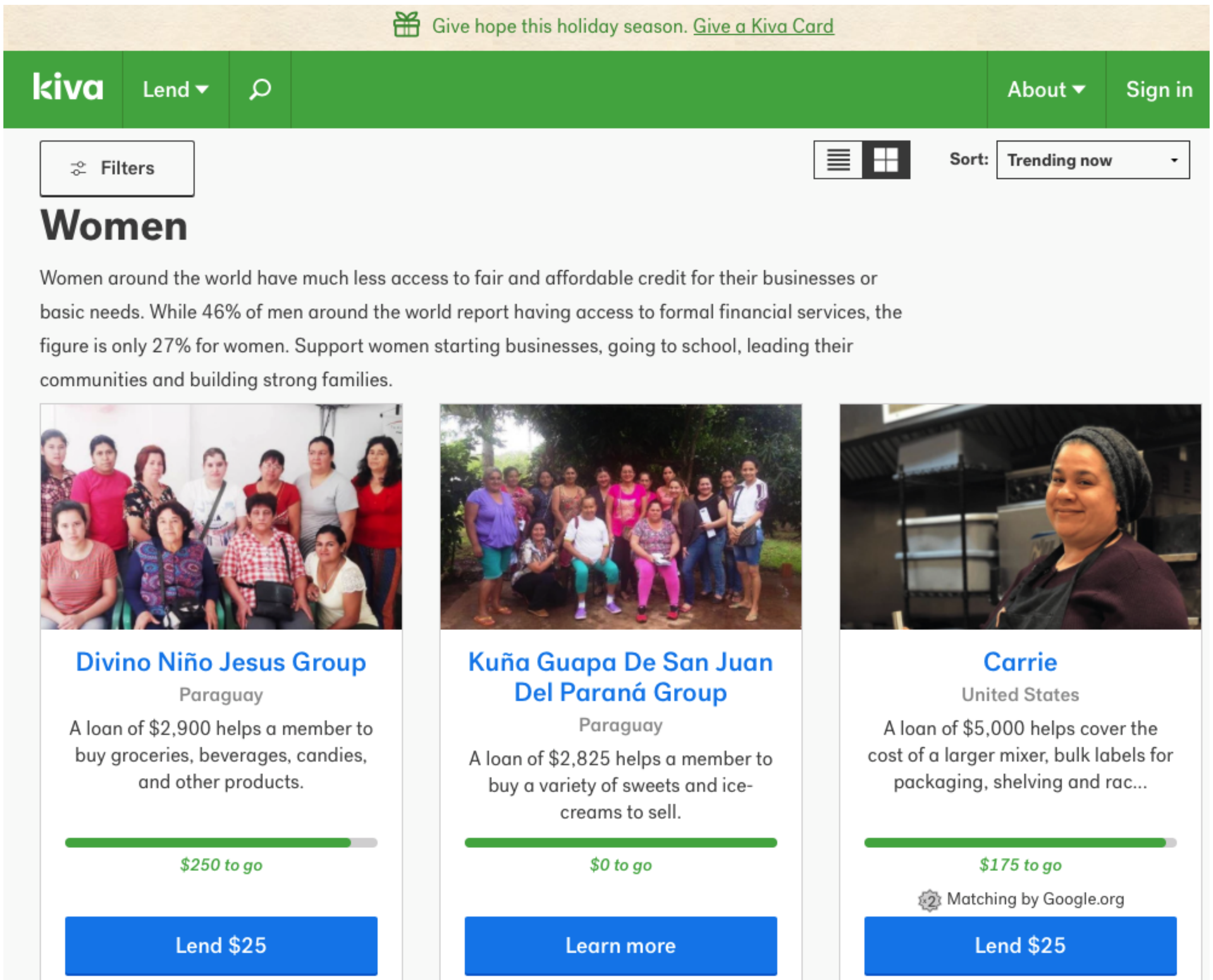


Figure 1: Kiva.org.

Method

The dataset is sourced and merged from 2 Kaggle datasets; loan detail from *Kiva.org* and regional multidimenisonal poverty index (MPI) detail. It was reshaped to have tag information per row (tags are additional information provided for borrowers, for example, "Elderly","Woman Owned Biz").

Variables $Partner.ID, Country, Region, WorldRegion, Sector, Activity, Use, Tags, Gender, Tags, Date, FundedLoanAr$

Note that there is more data to be merged that will be relevant to analysis, loan theme by region, human development index, and population below the poverty line in the future.

Dataset 1: Kiva Loans

The dataset contains the majority of the loan detail provided by *Kiva*.

```
##   partner_id funded_amount loan_amount         sector
## 1        247          2225        2225         Retail
## 2        334           250         250       Services
## 3        334           200         200    Agriculture
## 4        334           150         150 Transportation
## 5        334           250         250   Construction
## 6        334           250         250    Agriculture
##                  activity  country    region
## 1 Personal Products Sales Pakistan    Lahore
## 2                  Sewing    India Maynaguri
## 3                   Dairy    India Maynaguri
## 4          Transportation    India Maynaguri
## 5   Construction Supplies    India Maynaguri
## 6                 Farming    India  Dhupguri
##                                         borrower_genders
## 1 female,female,female,female,female,female,female,female
## 2                                                 female
## 3                                                 female
## 4                                                 female
## 5                                                 female
## 6                                                 female
##                                                                                     use
## 1                                                              to buy hair oils to sell.
## 2                                                         to purchase a sewing machine.
## 3                          To purchase a dairy cow and start a milk products business .
## 4  To repair their old cycle-van and buy another one to rent out as a source of income
## 5 to purchase stones for starting a business supplying stones to building contractors.
## 6                                            to cultivate broad beans on her own land.
##   term_in_months       date month year
## 1             11 2014-01-01    01 2014
## 2             43 2014-01-01    01 2014
```

```
## 3             43 2014-01-01    01 2014
## 4             43 2014-01-01    01 2014
## 5             43 2014-01-01    01 2014
## 6             43 2014-01-01    01 2014
##                                       tags Female.count Male.count
## 1 #Parent, #Repeat Borrower, user_favorite            8          0
## 2             user_favorite, user_favorite            1          0
## 3             user_favorite, user_favorite            1          0
## 4             user_favorite, user_favorite            1          0
## 5             user_favorite, user_favorite            1          0
## 6             user_favorite, user_favorite            1          0
##   Woman.Biz M.F.loan F.loan M.loan F.only.loan M.only.loan fully.funded
## 1         0        0      1      0           1           0            1
## 2         0        0      1      0           1           0            1
## 3         0        0      1      0           1           0            1
## 4         0        0      1      0           1           0            1
## 5         0        0      1      0           1           0            1
## 6         0        0      1      0           1           0            1
##   unfunded partial.funded Gender Gender.Var Water
## 1        0              0 female          1     0
## 2        0              0 female          1     0
## 3        0              0 female          1     0
## 4        0              0 female          1     0
## 5        0              0 female          1     0
## 6        0              0 female          1     0
```

Dataset 2: Multidimensional Poverty Index (MPI) and World Region Detail

This dataset contains World Region and MPI variables.

```
##       country     region world_region   MPI
## 1 Afghanistan Badakhshan   South Asia 0.387
## 2 Afghanistan    Badghis   South Asia 0.466
## 3 Afghanistan    Baghlan   South Asia 0.300
## 4 Afghanistan      Balkh   South Asia 0.301
## 5 Afghanistan     Bamyan   South Asia 0.325
## 6 Afghanistan   Daykundi   South Asia 0.313
```

Datasets Merged

This is one iteration of the dataset. I summarized loans and MPI separately by world region, region, and country for analysis.

```
## # A tibble: 6 x 4
## # Groups:   country [6]
```

```
##   country      MPI_country sumloan_amount sumfunded_amount
##   <chr>              <dbl>          <dbl>            <dbl>
## 1 Afghanistan        0.310          0.014            0.014
## 2 Burundi            0.412          2.28             2.17
## 3 Benin              0.320          0.05             0.05
## 4 Burkina Faso       0.548          2.7              2.64
## 5 Belize             0.0201         0.078            0.078
## 6 Brazil             0.0273         0.596            0.595
```

The Loans: Distribution of Fully Funded, Partially Funded, and Unfunded Loans

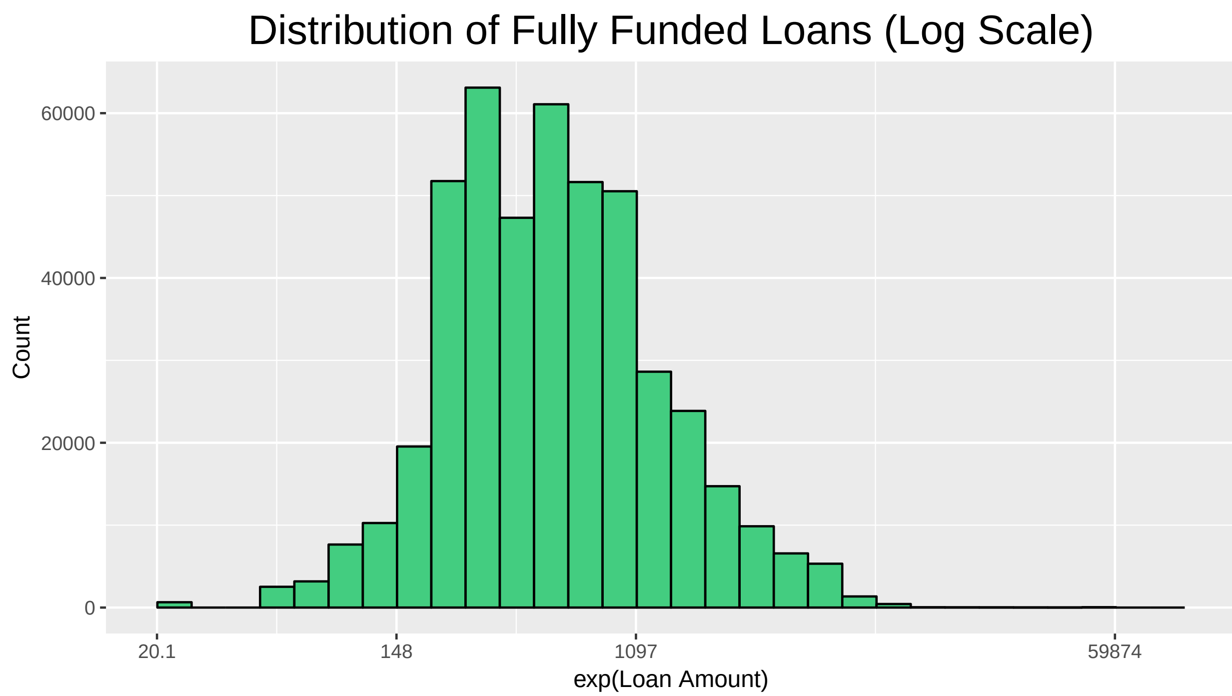
Not all loans receive full funding.

```
## [1] "No. of Fully Funded Loans = 423089"
```

```
## [1] "No. of Unfunded Loans = 2054"
```

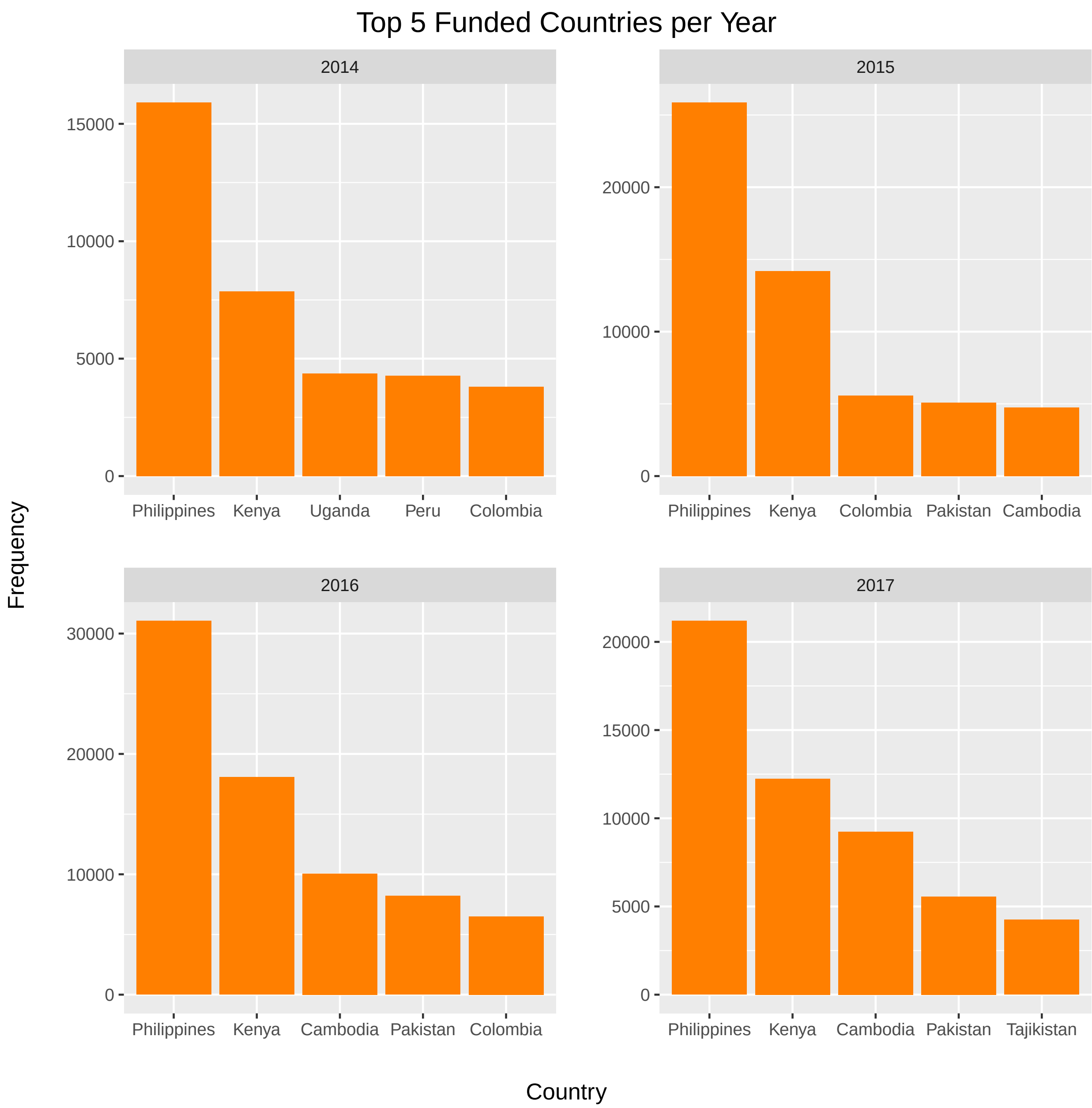
```
## [1] "No. of Partially Funded Loans = 37022"
```

```
## [1] "No. of Over Funded Loans = 2"
```

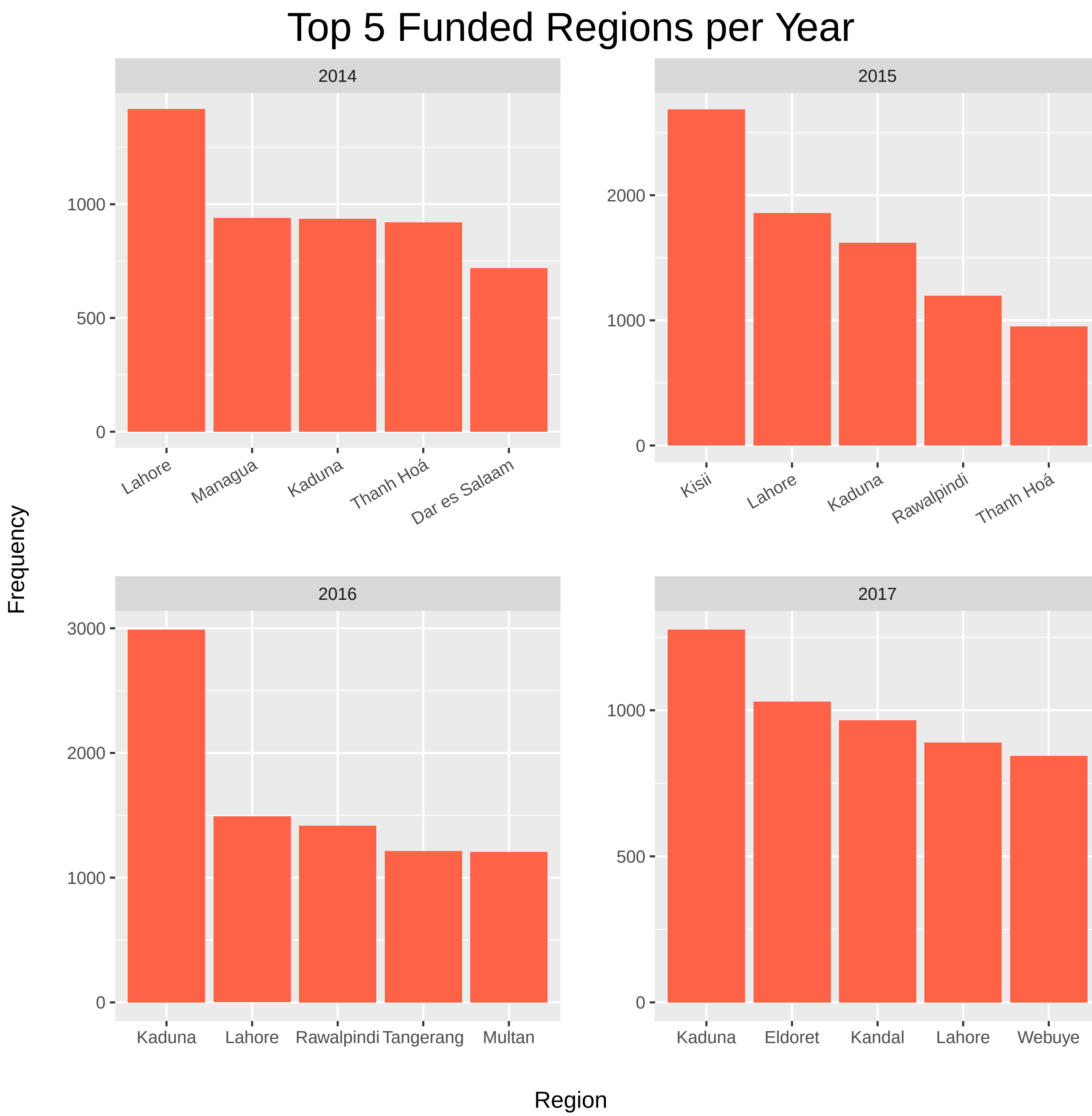


In the dataset, an unfunded loan is *FundedAmount*=$0, a partially funded loan is funded amount is less than loan amount and fully funded loan is funded amount=loan amount. Over 91% of Kiva loans are fully funded! From the histogram, the fully funded amounts in the dataset range from about 20 USD to about 5,9875 USD with an average of about 1,100 USD. How does this breakdown by *Country* and *Region*? Is funding going to the most impoverished areas? What about *Gender*? And *Sector*?

Frequently Funded Countries and Regions



From the bar plots, the top funded countries are consistently the Philippines and Kenya each year, Cambodia is also frequntly funded.



From the bar plot, there is a mix of top regions that are top funded per year.

Is there a relationship between frequently funded regions/countries and *MPI*?

Funded Loan Amount and Poverty Index (MPI)

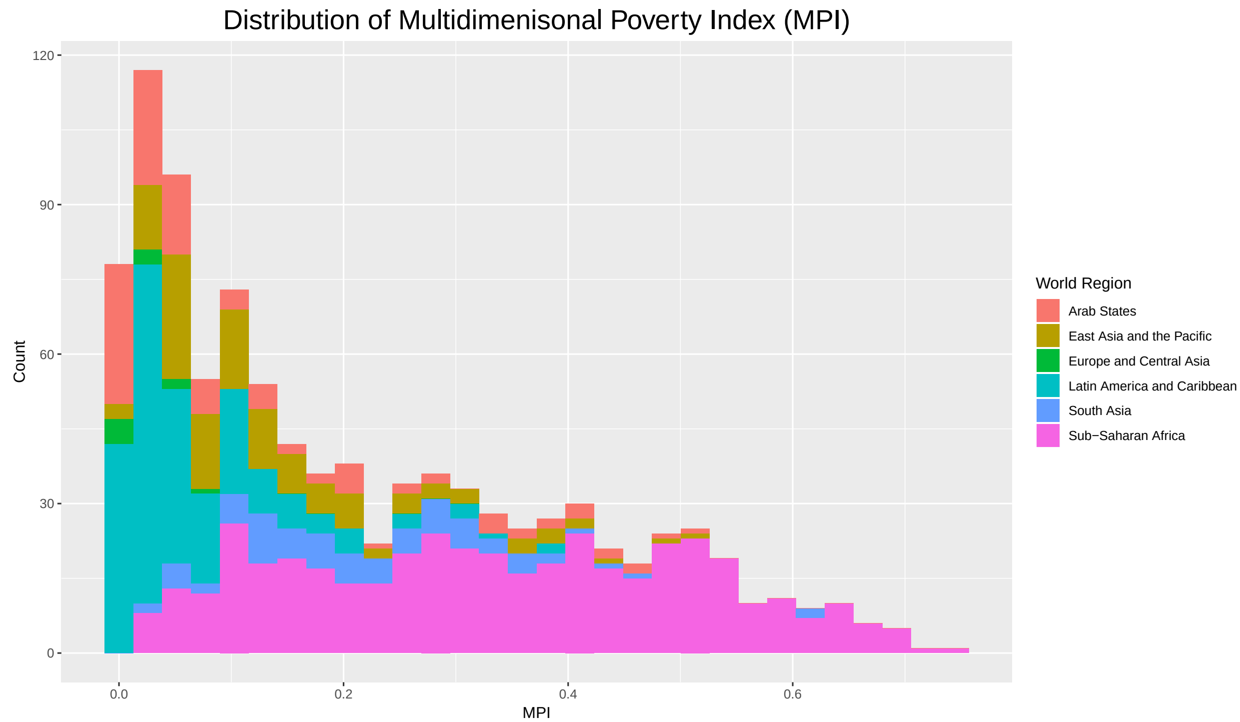
Here we introduce MPI and the 6 World Regions. This includes only regions and countries with an MPI.

```
## [1] "Max MPI = 0.74"
```

```
## [1] "Min MPI = 0.00"
```

```
## [1] "Med MPI = 0.15"
```

```
## [1] "Mean MPI = 0.21"
```

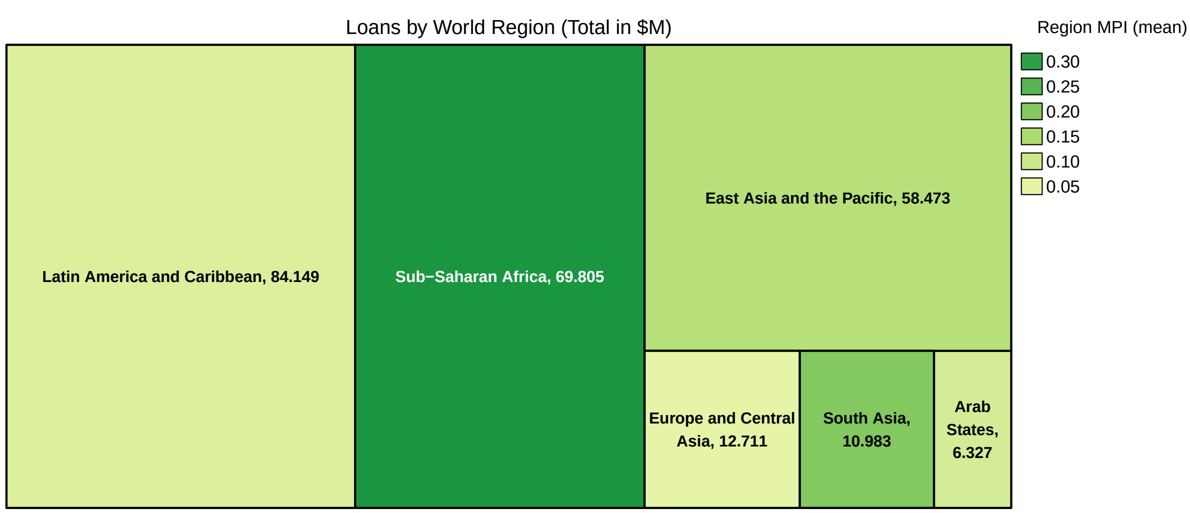


From distribution, the poorest world regions are Sub-Saharan Africa and South Asia. What proportion of loans are these *WorldRegions* receiving?

```
## [1] "No. of Total Regions = 928"
```

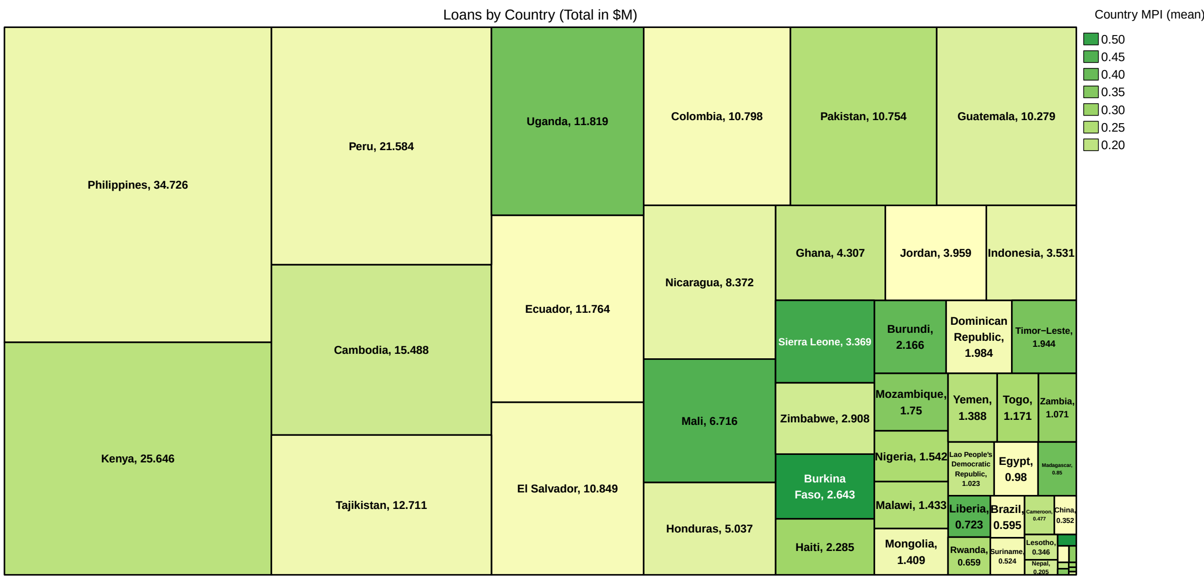
```
## [1] "No. of Total Countries = 102"
```

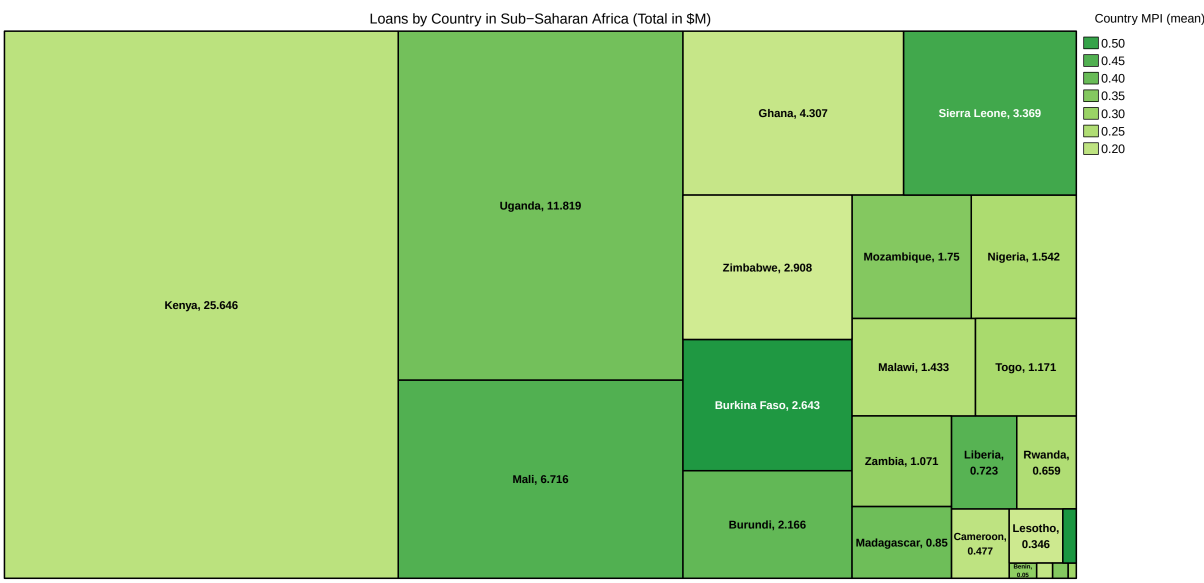
```
## [1] "No. of Total World Regions = 6"
```

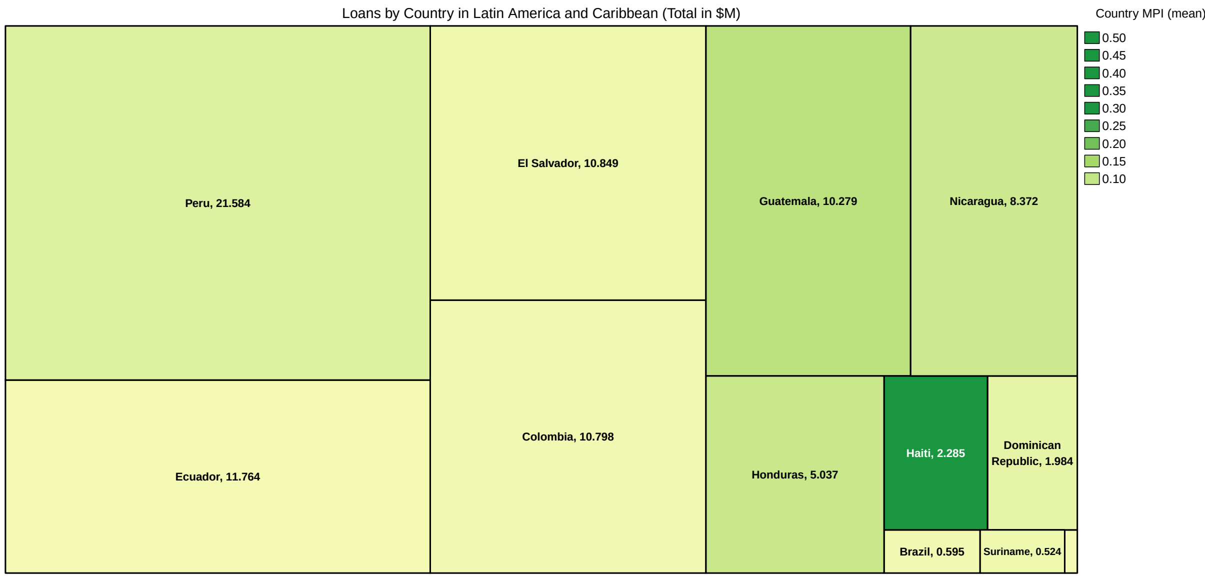


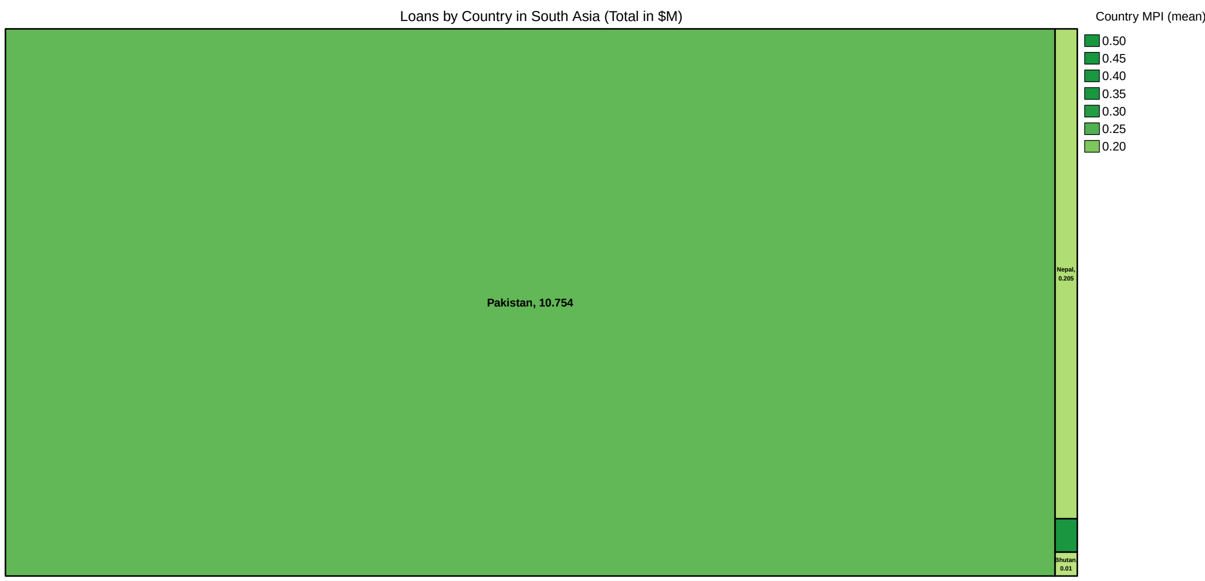
As we saw from the *MPI* distribution, Sub-Saharan Africa is the poorest *WorldRegion*. From the treemap, Sub-Saharan Africa has received a large portion of the total funded loans. While South

Asia, high on the poverty index, receives the second smallest portion of funded loans. South Asia might be an area to focus on to identify loan trends.







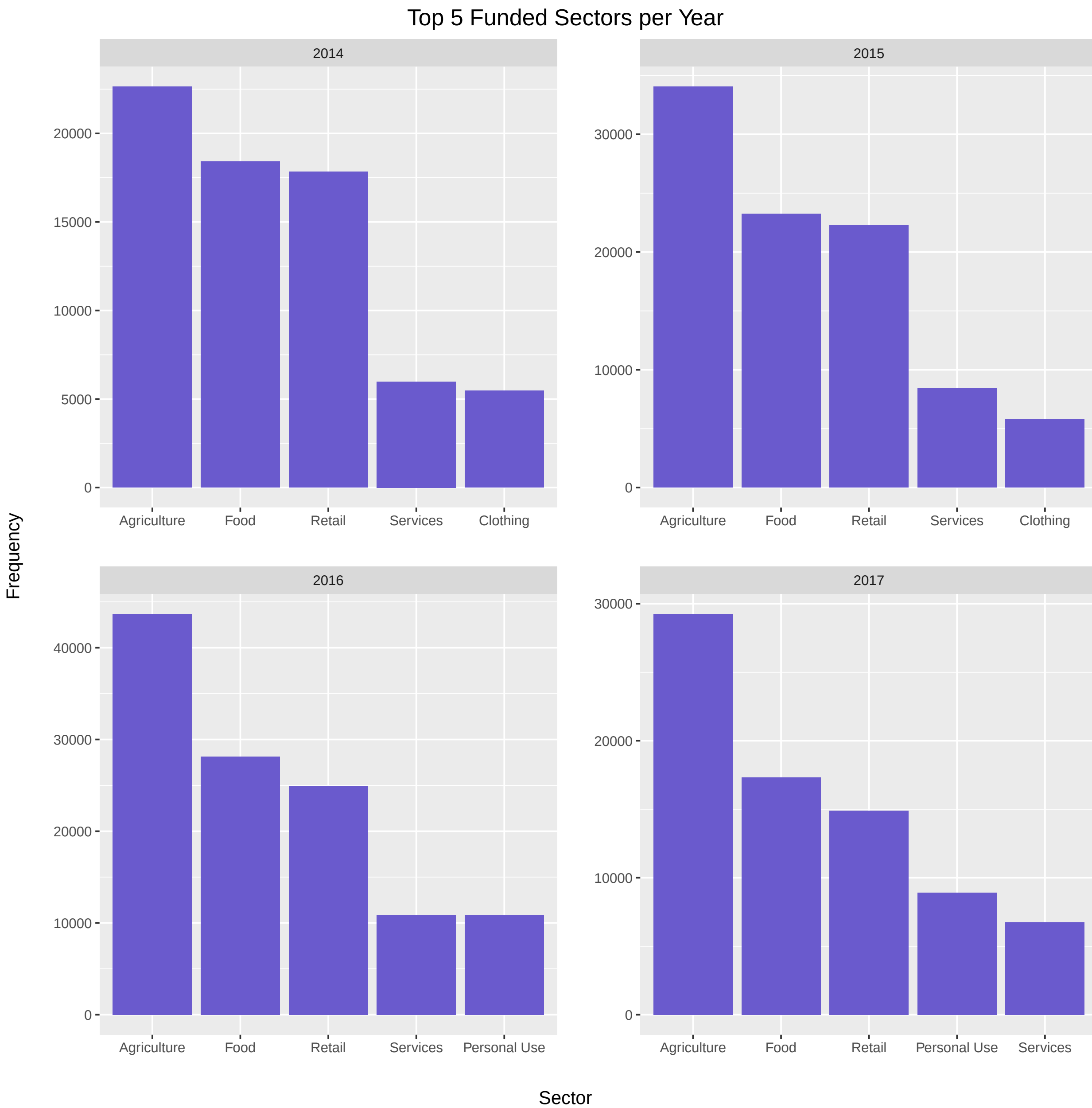


From these treemaps, some of the poorest countries are receiving a small portion of the total Kiva loans. Can see the darker green more prominent in the lower most corner.

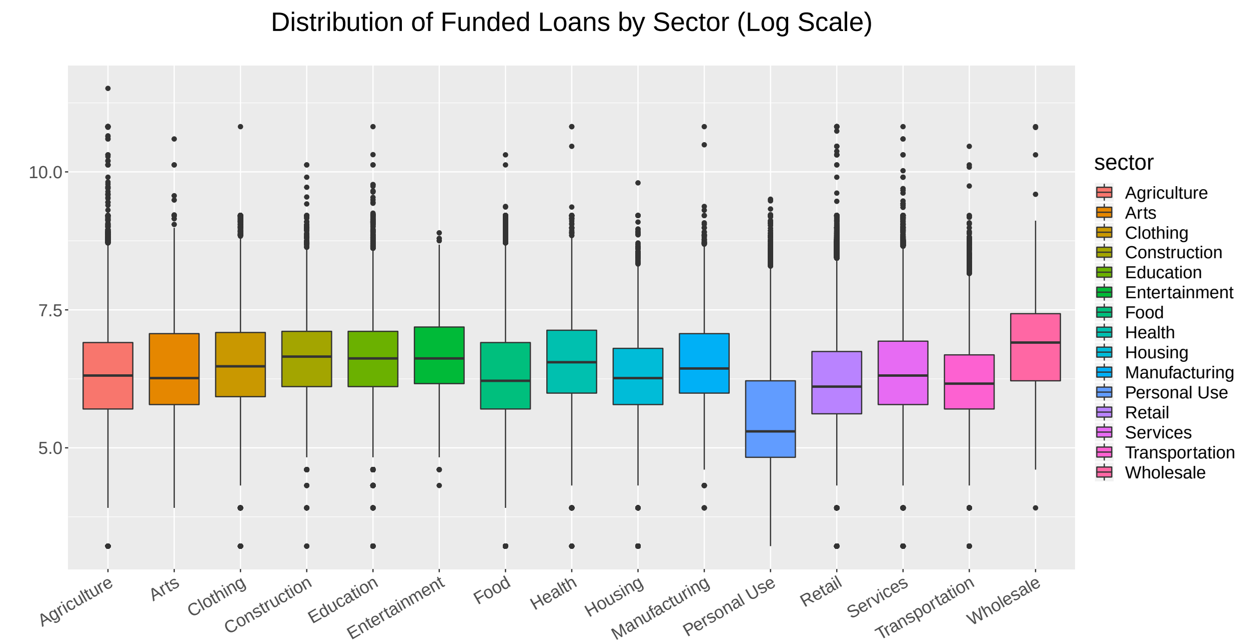
Burkina and South Sudan within Sub-Saharan Africa, Haiti within Latin America and Caribbean, and Afghanistan within South Asia are receiving a small portion of the funded amounts.

Again, these areas of focus for loan trends and what is driving these differences. One thing to consider are potential sector and activity differences between the regions/countries. Do the loan needs of the poorer countries cost less than the others? Can we use this to estimate poverty levels and needs for those countries?

Frequently Funded Sectors



From the bar plots, the most frequently funded *sector*s are consistently Agriculture, Food, and Retail. What is the overall loan distribution among the sectors?

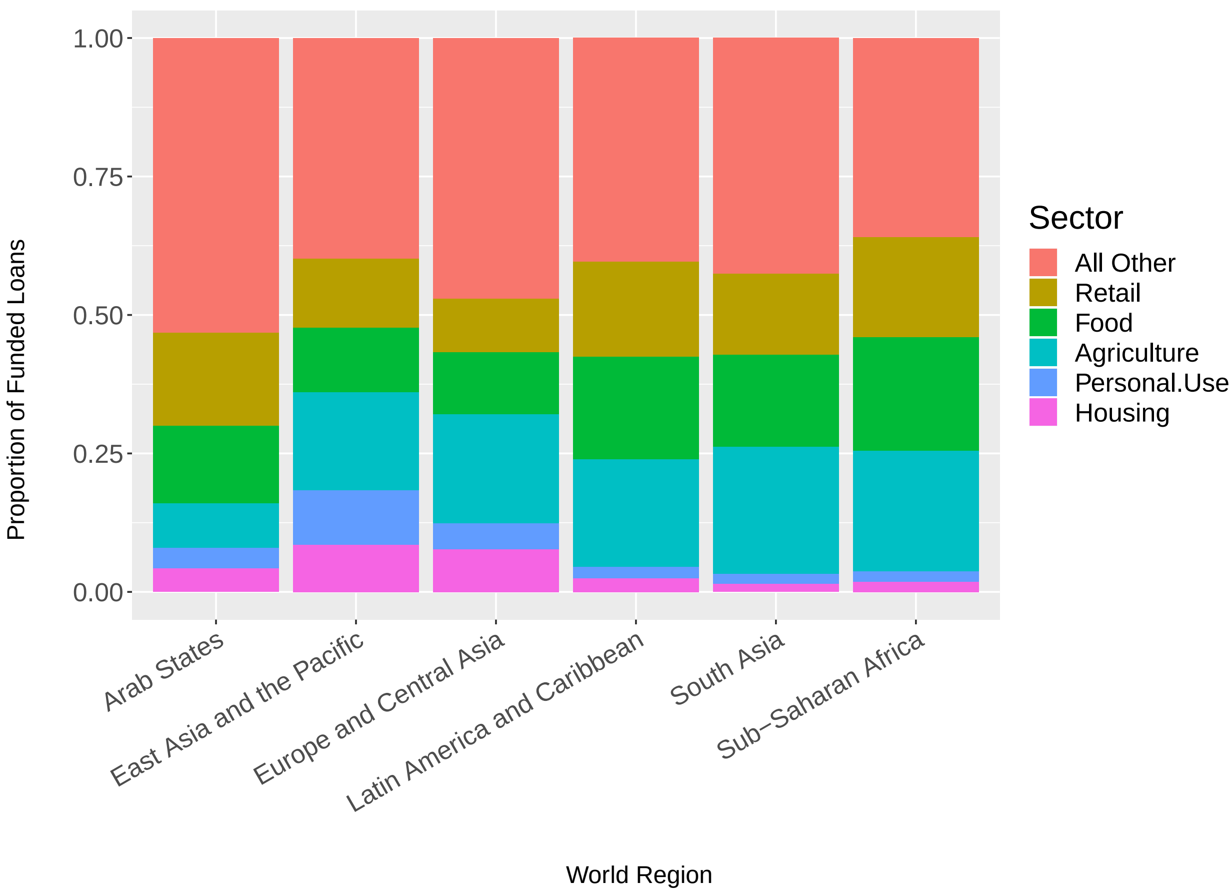


From the box plot, there is some variation in the loan amounts among the *sector*s. The Food, Housing, and Personal Use sectors have the lowest medians. This could correspond to what we saw from the treemaps. So, what is the funded loan breakdown for these top sectors for World Region and Country?

Proportion of Funded Loans for Sectors by World Region

```
## Warning: Column `country` joining factors with different levels, coercing
## to character vector
```

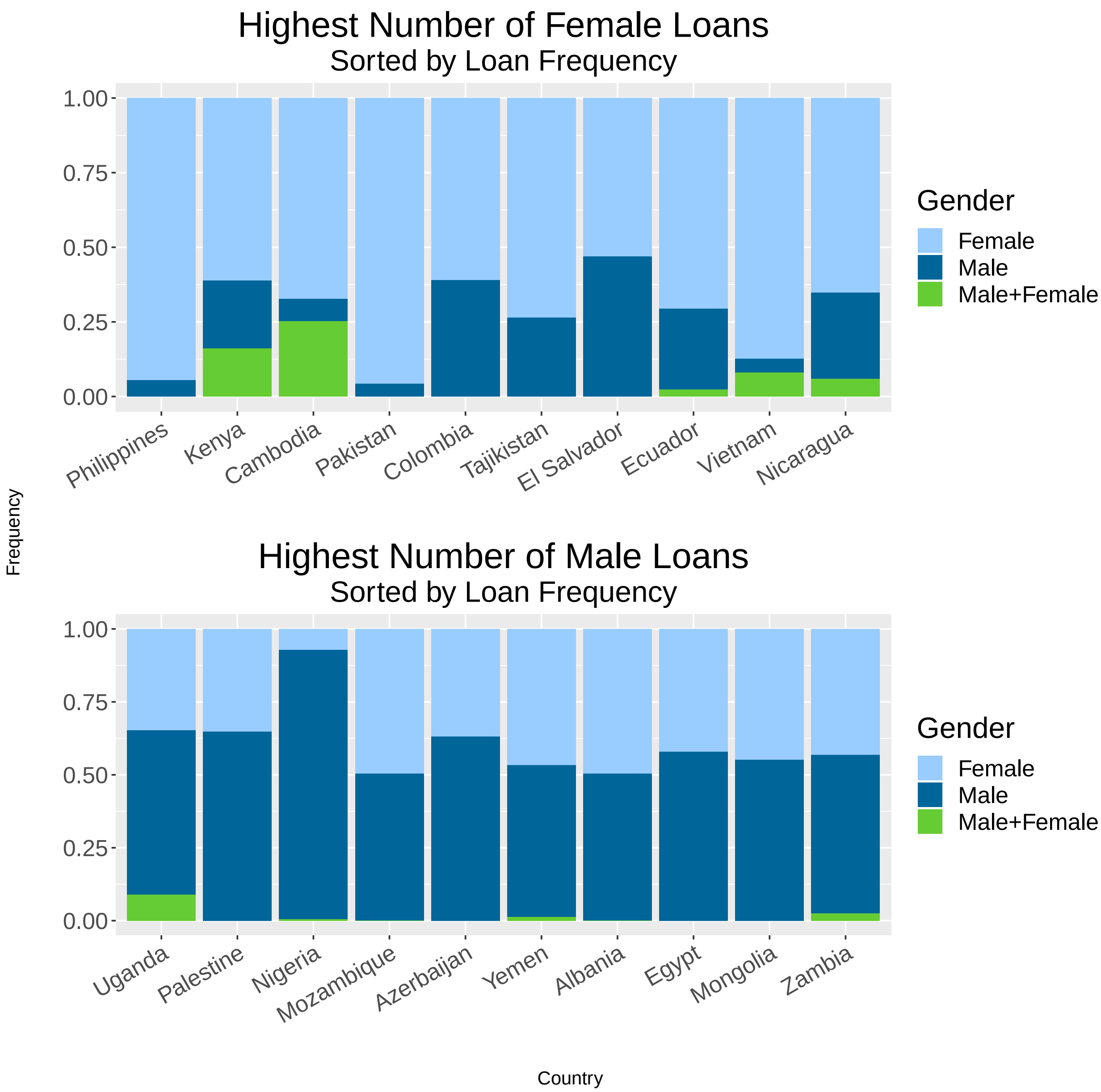
Relationship Between World Regions and Top Sectors



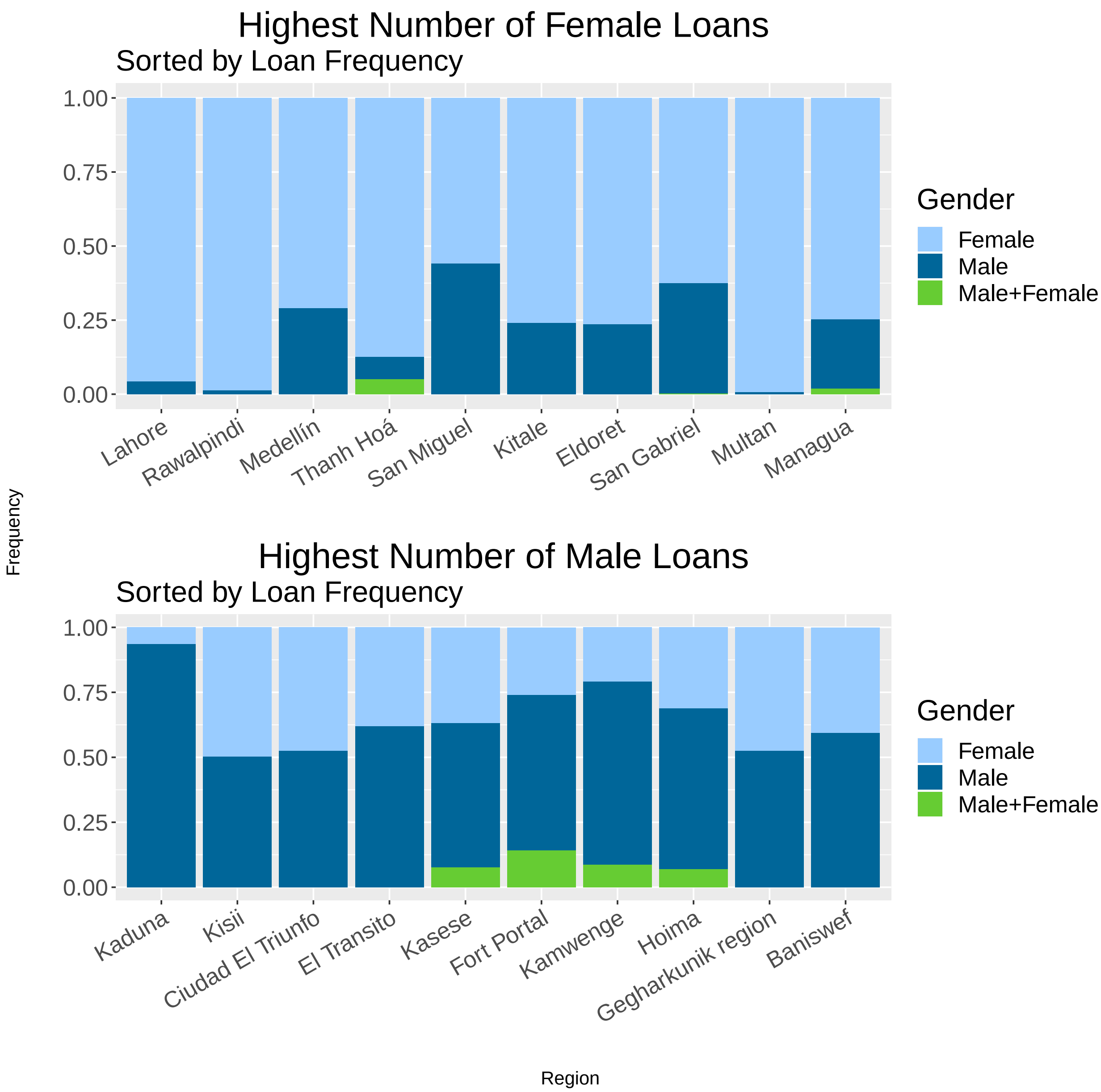
The chart indicates that Agricultural loans make up a good portion of the funded loan amounts. Recalling both the treemap by World Region and the sector boxplots, I would have expected a larger proportion of Personal Use and Housing within Sub-Saharan Africa and South Asia. (I will revisit this). Now we will take a look how *Gender* plays a role in the dataset.

Loan Breakdown by Gender

```
## [1] "No. of Total Loans = 462167"
## [1] "Female Only Loans = 70%"
## [1] "Male Only Loans = 23%"
## [1] "Female+Male Loans = 7%"
## [1] "Total Loans - Funded = 92%"
## [1] "Female Only Loans - Funded = 94%"
## [1] "Male Only Loans - Funded = 83%"
## [1] "Female+Male Loans - Funded = 92%"
```

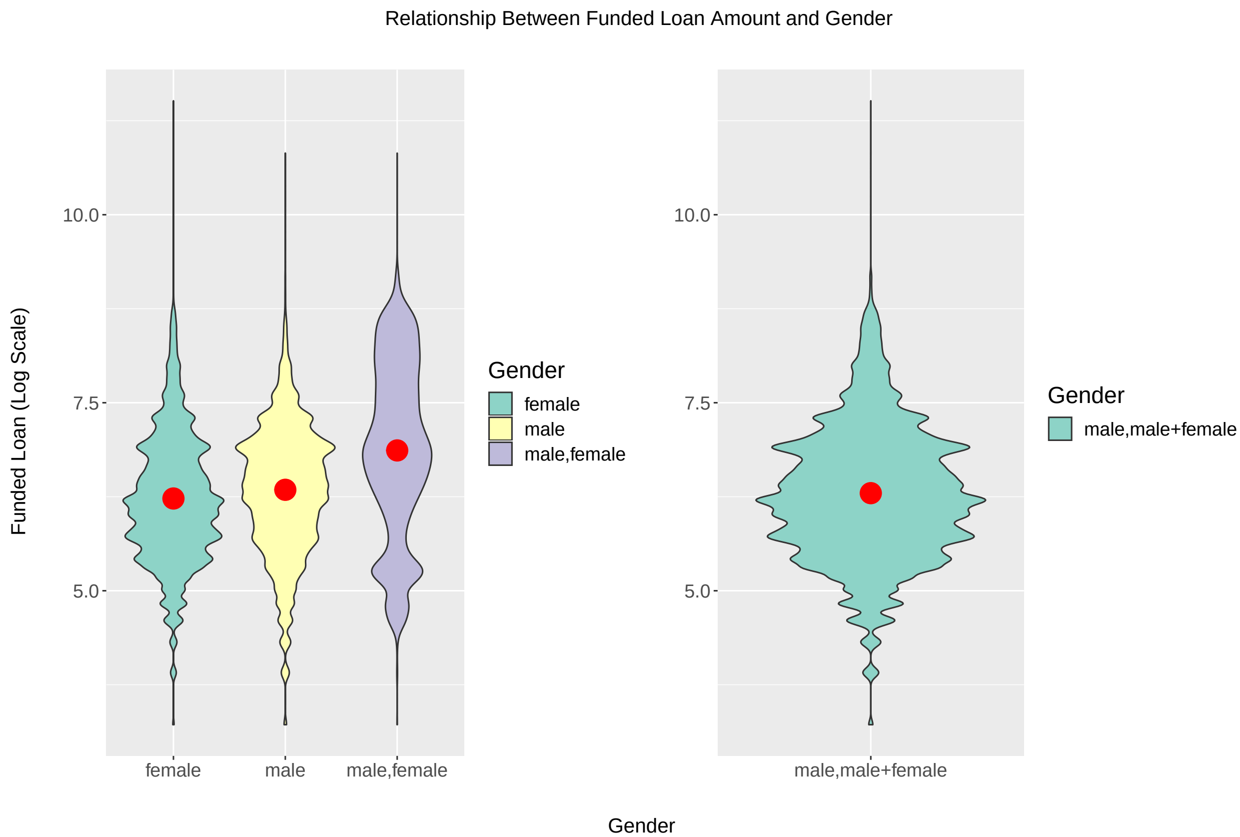


Highest Number of Female Loans
Sorted by Loan Frequency
1.00
0.75
0.50
0.25
0.00
Philippines
Kenya
Cambodia
Pakistan
Colombia
Tajikistan
El Salvador
Ecuador
Vietnam
Nicaragua
Gender
Female
Male
Male+Female
Frequency
Highest Number of Male Loans
Sorted by Loan Frequency
1.00
0.75
0.50
0.25
0.00
Uganda
Palestine
Nigeria
Mozambique
Azerbaijan
Yemen
Albania
Egypt
Mongolia
Zambia
Gender
Female
Male
Male+Female
Country



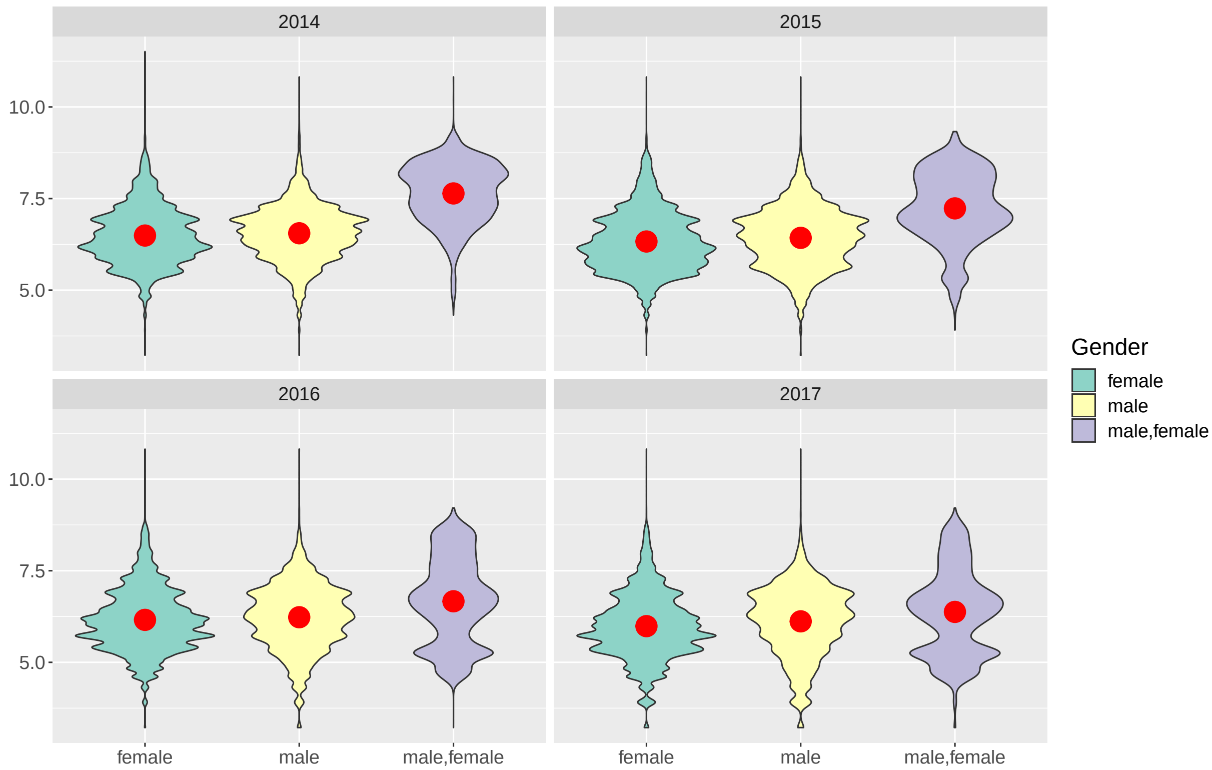
These charts provide good summaries for the *Gender* differences across countries and regions. There are clear differences among the countries and regions for who is taking out the loan.

We will also want to consider if the average loan amount differs between *Gender* across *Country*.



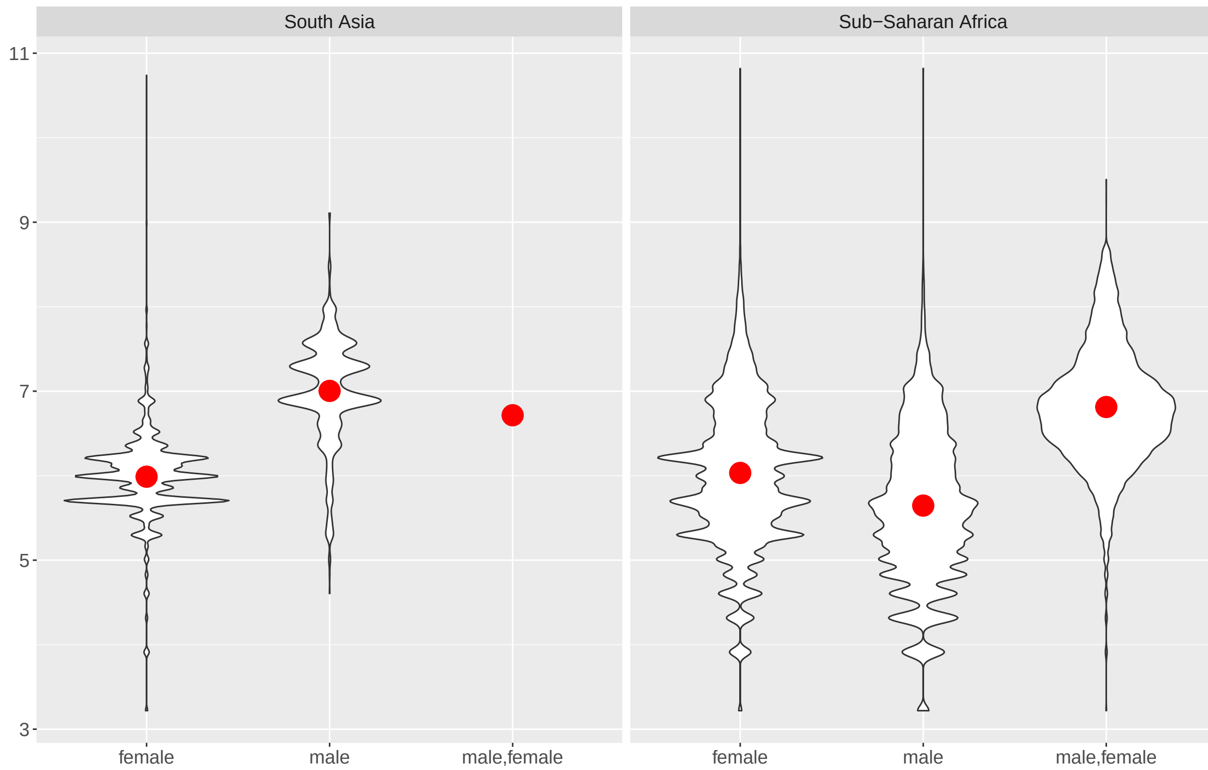
From the violin plots, there is a difference in the average funded loan amount (red dot) between females and males on an overall basis.

We may see even bigger differences by looking at the gender differences across World Regions and Countries and over time.

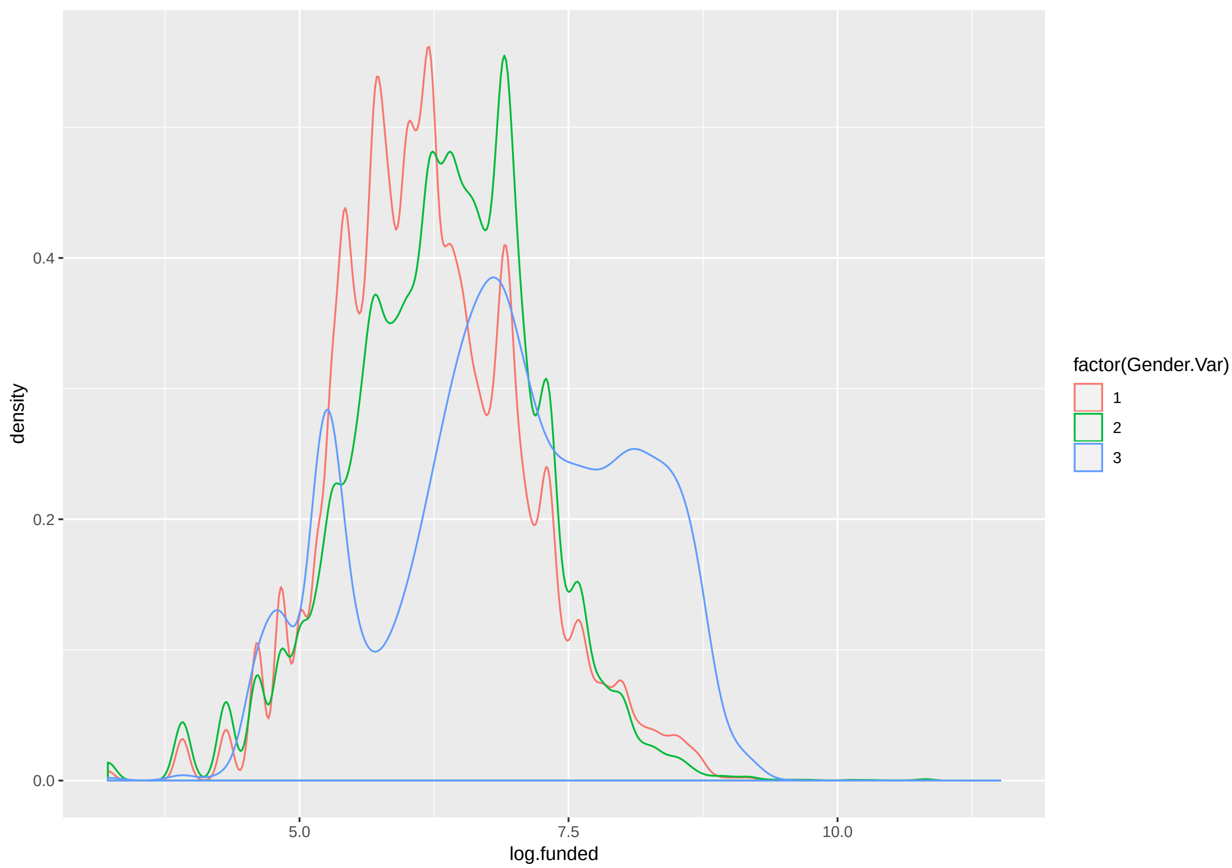


From the violin plots by year, the average loans overall seem to be decreasing, but also leveling between female, male, and male+female.

We may see something interesting across World Regions and Countries.



There was just one loan that had male,female variable for South Asia (in 2014). (We will revisit this in more detail in our modeling)



Logistic Regression

Will take a look at logistic regression for specific *sector*s of interest. For example, Agriculture or not Agriculture outcome on *MPI*.

Multinomial Model with Random Effect

Will take a look at outcome of all *sector* types and how *MPI* is asssociated with them. *Sector* as outcome and *MPI* as predictor and a random effect.

Multilevel Model

Here we will look at funded loan amount (log scale) as the outcome, sector as predictor, and gender+country as random effects.

```
##                                         (Intercept) factor(Gender.Var)2 factor(Gender.Var)3
## Paraguay                                0.364667826        -0.936526384         -0.65888504
## The Democratic Republic of the Congo    0.753548684        -0.905664348         -0.17798350
## Ghana                                  -0.437209431        -0.708252809          0.51643843
## Somalia                                 0.371876986        -0.629399900          1.05128409
```

```
## Guatemala                          -0.008162036     -0.599304841      -0.20332047
## Cote D'Ivoire                       2.420548190     -0.547021922      -1.88047743
## Mexico                             -0.089219642     -0.463233366       0.94311437
## Azerbaijan                          0.214096857     -0.429670667      -0.05704167
## Congo                               1.326077330     -0.396818129       0.16846824
## Costa Rica                          0.032531180     -0.392649709       0.08516294
## Vietnam                            -0.182509741     -0.337403095       0.53935218
## Armenia                            -0.031675768     -0.319074347       0.11811456
## Bolivia                            -0.017021214     -0.313468703       1.10864043
## Bhutan                              1.318971172     -0.298075514      -1.02468339
## Afghanistan                         1.293419229     -0.292301007      -1.00483258
## United States                       1.235913064     -0.227997397      -0.03129426
## Brazil                              0.348903458     -0.192851728       0.39020329
## Peru                               -0.414384781     -0.190357144       1.29908144
## Yemen                              -0.290616512     -0.185756800      -0.03264880
## Vanuatu                             0.376787656     -0.168295199       0.73329011
## Kyrgyzstan                         -0.220574848     -0.165551624       0.23309912
## Ecuador                            -0.373996248     -0.157875654       1.78369718
## Burkina Faso                       -0.268566389     -0.144074350      -0.16889510
## Jordan                             -0.222429489     -0.138286362      -0.14945540
## Palestine                           0.219779299     -0.135257757      -0.14621025
## Thailand                            0.452557286     -0.128489128      -0.34406913
## Mozambique                         -0.873880318     -0.110329616       0.46531772
## Zambia                             -0.094943779     -0.106433144       1.07935843
## Lebanon                             0.075410320     -0.097111462      -0.38573147
## Madagascar                         -1.594957983     -0.087816760       1.36757529
## El Salvador                        -0.817546655     -0.086278114       1.09045358
## Lesotho                             0.231537093     -0.078609596       0.14447322
## Colombia                           -1.031438546     -0.068536588       0.88775999
## Burundi                             0.651477494     -0.065174393       0.05831551
## Nicaragua                          -0.559743846     -0.062730673       0.56653007
## Mongolia                            0.307983824     -0.053265908      -0.24394892
## Albania                            -0.040799322     -0.038378114       0.04533907
## Tajikistan                         -0.746986826     -0.031397183       0.40827789
## Israel                              0.975975865     -0.022216232      -0.81506756
## Georgia                            -0.100283986     -0.021447555       0.09055194
## Tanzania                           -0.578538870     -0.021045548       0.67018318
## Nigeria                            -1.814005837     -0.007562967       1.88937849
## Mauritania                          1.463864989     -0.005945271      -1.23036569
## Saint Vincent and the Grenadines    0.809996935     -0.003662807      -0.68068838
## Benin                               0.805200121     -0.003270201      -0.67676364
## Cameroon                           -1.306073403      0.001733996       1.09876662
## Moldova                             0.284662592      0.008929988      -0.24214734
## Malawi                             -0.354210186      0.021111422       0.29207184
## Kenya                              -1.360363009      0.025126904       0.76816285
## Namibia                             0.720524726      0.029070895      -0.61476589
## Indonesia                          -0.790648958      0.039176286       1.41415559
## Zimbabwe                           -0.608138617      0.069896353       0.67504268
```

```
## China                                   0.645308707      0.070675668     -0.56338487
## Belize                                 -0.432099400      0.090587851      0.33771407
## Solomon Islands                        -0.487877594      0.110255908      0.37902274
## Honduras                               -0.618599798      0.110650482      0.82804632
## Timor-Leste                            -0.493401529      0.120631240      0.38069817
## Ukraine                                 0.067425776      0.145778564     -0.09853295
## Turkey                                 -0.677513416      0.153111883      0.52634717
## Suriname                                0.431901172      0.172125129     -0.41284717
## South Africa                            0.032274546      0.172619249     -0.07664093
## Chile                                   1.525671763      0.180875787     -0.78680104
## Uganda                                 -0.999424985      0.189547765      1.20425605
## Philippines                            -1.427162776      0.191419989      1.14631350
## Rwanda                                 -1.432767269      0.196768909      0.32771364
## Cambodia                               -0.778684182      0.202620605     -0.29918275
## India                                  -1.026900603      0.203806086      1.71035845
## Mali                                   -0.283948495      0.236560987      0.29826082
## Panama                                 -0.149211769      0.268199302      0.04871242
## Dominican Republic                      1.035929584      0.369330019     -0.27696104
## Egypt                                  -1.138762836      0.393846163      0.84555988
## Senegal                                -0.829534175      0.428028350      0.57549849
## Lao People's Democratic Republic       -1.297367512      0.444595065      1.55810284
## Samoa                                  -0.715517943      0.475115157      0.46603989
## Haiti                                  -1.218625674      0.555317797      2.24375905
## Sierra Leone                           -0.934407009      0.608389223      0.71098946
## South Sudan                            -1.373672985      0.630008968      0.97558324
## Togo                                   -1.833077402      0.731575009      2.28590654
## Myanmar (Burma)                        -0.149204262      0.775330021     -0.41269307
## Liberia                                -1.458452104      0.967003120      0.95034730
## Pakistan                               -1.209517851      1.022900634      0.73135869
## Nepal                                  -1.285177835      1.468123931      0.66087665
```

This orders *Gender* by country from more female dominated to more male dominated based on funded amount (log scale). The top 10 average loan amounts for male versus female are: Nepal, Pakistan, Liberia, Myanmar (Burma), Togo, South Sudan, Sierra Leone, Haiti, Samoa, and Lao People's Democratic Republic. (Will look into this more)

This is a work in progress. There are many different levels to this dataset.

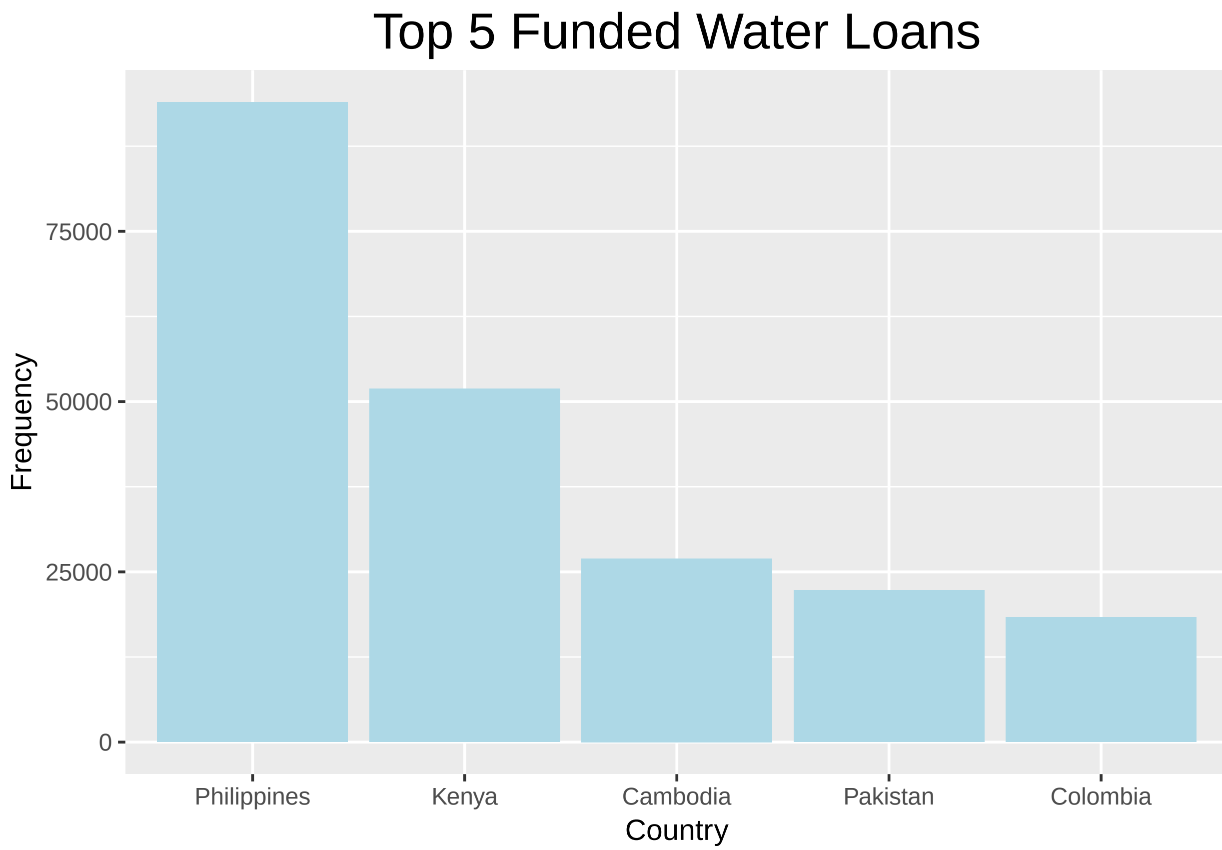
Conclusion/Next Steps

There is a lot to consider in this analysis. This project is ongoing and will be diving deeper into modeling next. It is important to understand the underlying themes and behaviors that differ between regions and countries. This data can help Kiva in supporting these areas.

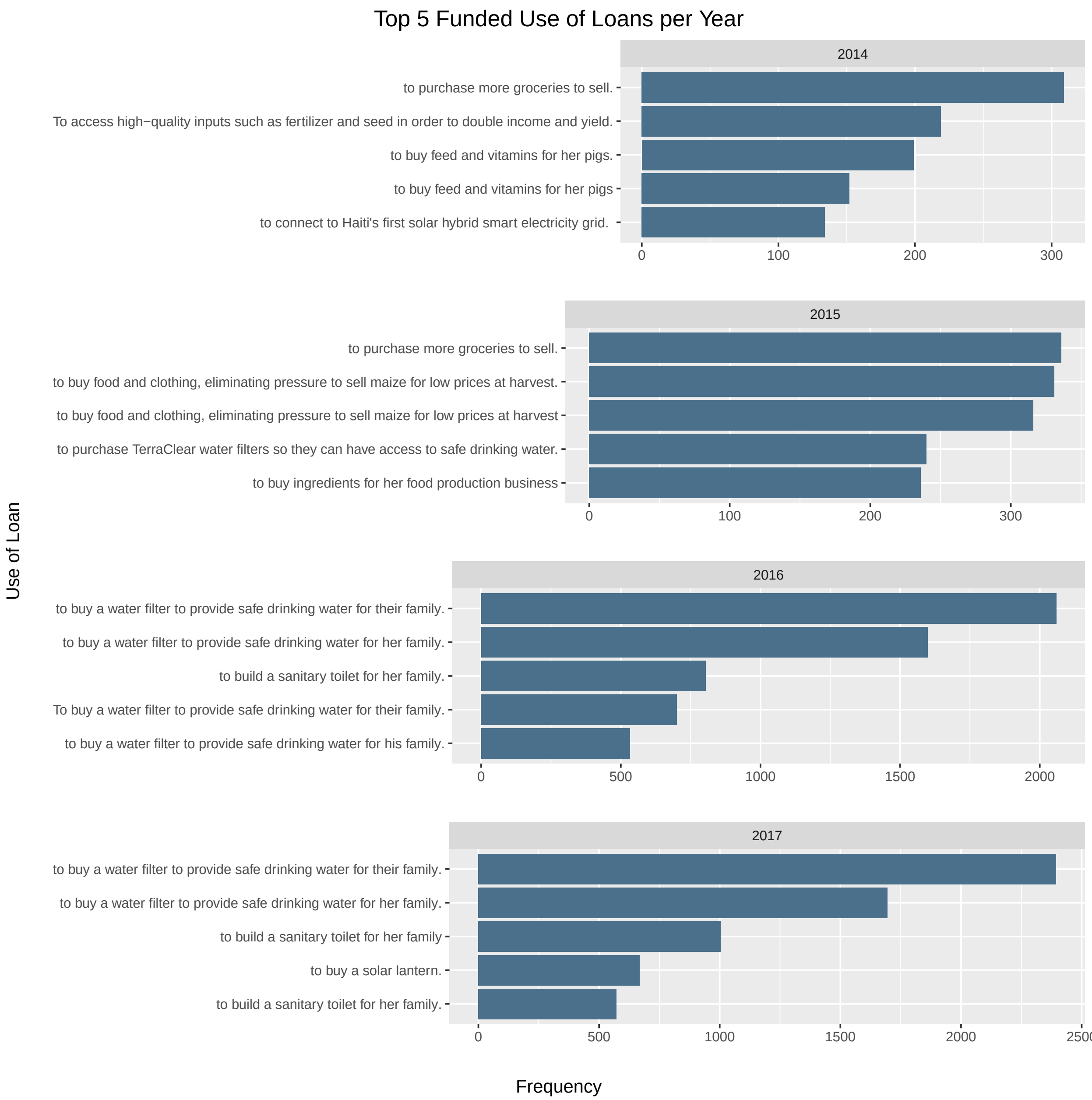
Appendix

Items to look at in the future:

Where is the use of water loan most prevelant?



How are loans being used?



From the bar plots, there is some variation in the use of the loan. For water uses, does this vary by region and become more prominent during dry seasons? If there is an expected dry season can we expect water loans to increase?

```
#- Most imporverished areas
#- What is being funded/partially funded/not funded and likelihood?
#- Female vs. male borrowers and whether having a male in the group affects loan behavior?
#- Repeat and type of borrowers
#EDA plot outcome on number of F and number of male and ratio of male count to female count to i
#Does a male impact on the loan.
```

```
#Number of females might not matter, but once adding in a male that could affect the loan amount
#Linear regression model per country on amount of loan for gender
```