DATA 607 Project 2 Part 1

Subhalaxmi Rout

03/08/2020

United Nations' Migration Data

Discussion thread created by : Subhalaxmi Rout

1.Introduction

People who have migrated across the countries all over the world and it was prepared and published by United Nation. Each of the origin countries, where the migrants are coming from is presented in each column and each of the destination countries, where the migrants are going to is represented in each row.The file contains a bunch of worksheets to include different years and data broken down by total / male / female. But I'm going to import 'Table 16' which contains the total migrants data for 2015 for this post.

Link: Dataset link

2.Load library

```
#install.packages("dplyr")
#install.packages("tidyr")
#install.packages("ggplot2")
#install.packages("DT")
library(dplyr)
```

```
##
## 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(tidyr)
library(ggplot2)
library(DT)
```

3.Data load and cleaning

Data is stored in the **Github** and loaded data from Github to Rstudio using **read.csv()** method.

```
# read csv file data in a variable using  read.csv and skip first 15 rows
data <- read.csv('https://raw.githubusercontent.com/SubhalaxmiRout002/Data-607-Project-2-Dataset-1/mast

# convert data to data frame
data <- data.frame(data)

# display data using datatable
datatable(data,options = list(scrollX = TRUE, paging=TRUE,fixedHeader=TRUE))
```

Show 10 entries Search:

	Var.1	Var.2	Var.3	Var.4	Var.5	Total	Other.North	Other.South	Afghanistan	Alban
1	1	WORLD		900		243700236	2139539	7644005	4843117	11229
2	2	Developed regions	(b)	901		140481955	539780	3520214	462239	11182
3	3	Developing regions	(c)	902		103218281	1599759	4123791	4380878	46
4	4	Least developed countries	(d)	941		11951316	241805	1005567	0	
5	5	Less developed regions excluding least developed countries		934		91266965	1357954	3118224	4380878	46
6	6	Sub-Saharan Africa	(e)	947		18993986	328171	1551645	122	1
7	7	Africa		903		20649557	350543	1586186	677	5
8	8	Eastern Africa		910		6129113	102542	457137	0	
9	9	Burundi		108	B R	286810	3242	20599		
10	10	Comoros		174	B	12555	597	754		

Showing 1 to 10 of 265 entries Previous 1 2 3 4 5 … 27 Next

3.1 Remove region and keep only contries

```
# remove blanks from Var5
data <- data %>% filter(data$Var.5 != "")

# display data using datatable
datatable(data,options = list(scrollX = TRUE, paging=TRUE,fixedHeader=TRUE))
```

Show 10 entries Search:

	Var.1	Var.2	Var.3	Var.4	Var.5	Total	Other.North	Other.South	Afghanistan	Albania
1	9	Burundi		108	B R	286810	3242	20599		
2	10	Comoros		174	B	12555	597	754		
3	11	Djibouti		262	B R	112351	1679	5042		
4	12	Eritrea		232	I	15941	693	1533		
5	13	Ethiopia		231	B R	1072949	17477	52431		
6	14	Kenya		404	B R	1084357	21112	39318		
7	15	Madagascar		450	C	32075	4785	3827		
8	16	Malawi		454	B R	215158	11708	35136		
9	17	Mauritius	(1)	480	C	28585	191	1361		
10	18	Mayotte		175	B	76992	796	800		

Showing 1 to 10 of 232 entries Previous 1 2 3 4 5 … 24 Next

3.2 Remove unnecessary columns

```
# using "-" sign to delete the column
data <- data %>% select(-Var.1, -Var.3, -Var.5, -Total, -starts_with("Other"))

# display data using datatable
datatable(data,options = list(scrollX = TRUE, paging=TRUE,fixedHeader=TRUE))
```

Show 10 entries Search:

	Var.2	Var.4	Afghanistan	Albania	Algeria	American.Samoa	Andorra	Angola	Anguilla
1	Burundi	108							
2	Comoros	174							
3	Djibouti	262							
4	Eritrea	232						346	
5	Ethiopia	231							
6	Kenya	404							
7	Madagascar	450							
8	Malawi	454							
9	Mauritius	480							
10	Mayotte	175							

Showing 1 to 10 of 232 entries Previous 1 2 3 4 5 … 24 Next

3.3 Rename column names

```
# rename Var.2 and Var.4 i.e "destination_country" and "country_code"
data <- data %>% rename(destination_country = Var.2, country_code = Var.4)
```

```
# display data using datatable
datatable(data,options = list(scrollX = TRUE, paging=TRUE,fixedHeader=TRUE))
```

Show 10 entries Search:

	destination_country	country_code	Afghanistan	Albania	Algeria	American.Samoa	Andorra	Angola
1	Burundi	108						
2	Comoros	174						
3	Djibouti	262						
4	Eritrea	232						34
5	Ethiopia	231						
6	Kenya	404						
7	Madagascar	450						
8	Malawi	454						
9	Mauritius	480						
10	Mayotte	175						

Showing 1 to 10 of 232 entries Previous 1 2 3 4 5 … 24 Next

3.4 Gather 232 columns to make it tidy

There are many pair of contries people did not migratate, so remove those contries.

```
# convert country column to row and remove "NA"
data <- data %>% gather(origin_country, migrants, Afghanistan:Zimbabwe, na.rm = TRUE)

# display data using datatable
datatable(data,options = list(scrollX = TRUE, paging=TRUE,fixedHeader=TRUE))
```

Show 10 entries Search:

	destination_country	country_code	origin_country	migrants
31	Egypt	818	Afghanistan	235
32	Libya	434	Afghanistan	320
39	Namibia	516	Afghanistan	39
40	South Africa	710	Afghanistan	83
61	Tajikistan	762	Afghanistan	7587
75	Malaysia	458	Afghanistan	498
77	Philippines	608	Afghanistan	1220
79	Thailand	764	Afghanistan	1
85	India	356	Afghanistan	8086
86	Iran (Islamic Republic of)	364	Afghanistan	2348369

Showing 1 to 10 of 11,228 entries Previous 1 2 3 4 5 … 1123 Next

3.5 Pair of contries with greater than 1 million people

```
# apply filter condition on migrants
data1 <- data.frame(data) %>% filter(data$migrants > 1000000)

# display data using datatable
datatable(data1,options = list(scrollX = TRUE, paging=TRUE,fixedHeader=TRUE))
```

Show 10 entries Search:

	destination_country	country_code	origin_country	migrants
1	Iran (Islamic Republic of)	364	Afghanistan	2348369
2	Pakistan	586	Afghanistan	1618687
3	France	250	Algeria	1430656
4	India	356	Bangladesh	3171022
5	Cte d'Ivoire	384	Burkina.Faso	1294323
6	China, Hong Kong Special Administrative Region	344	China	2307783
7	United States of America	840	China	2103551
8	United States of America	840	Cuba	1131284
9	United States of America	840	El.Salvador	1276489
10	Pakistan	586	India	2000908

Showing 1 to 10 of 39 entries Previous 1 2 3 4 Next

4. Analysis

This part shows relation between origin country and desitnation country. The **darkred** color country has low imigrants and the **blue** color country has more number of immirants.

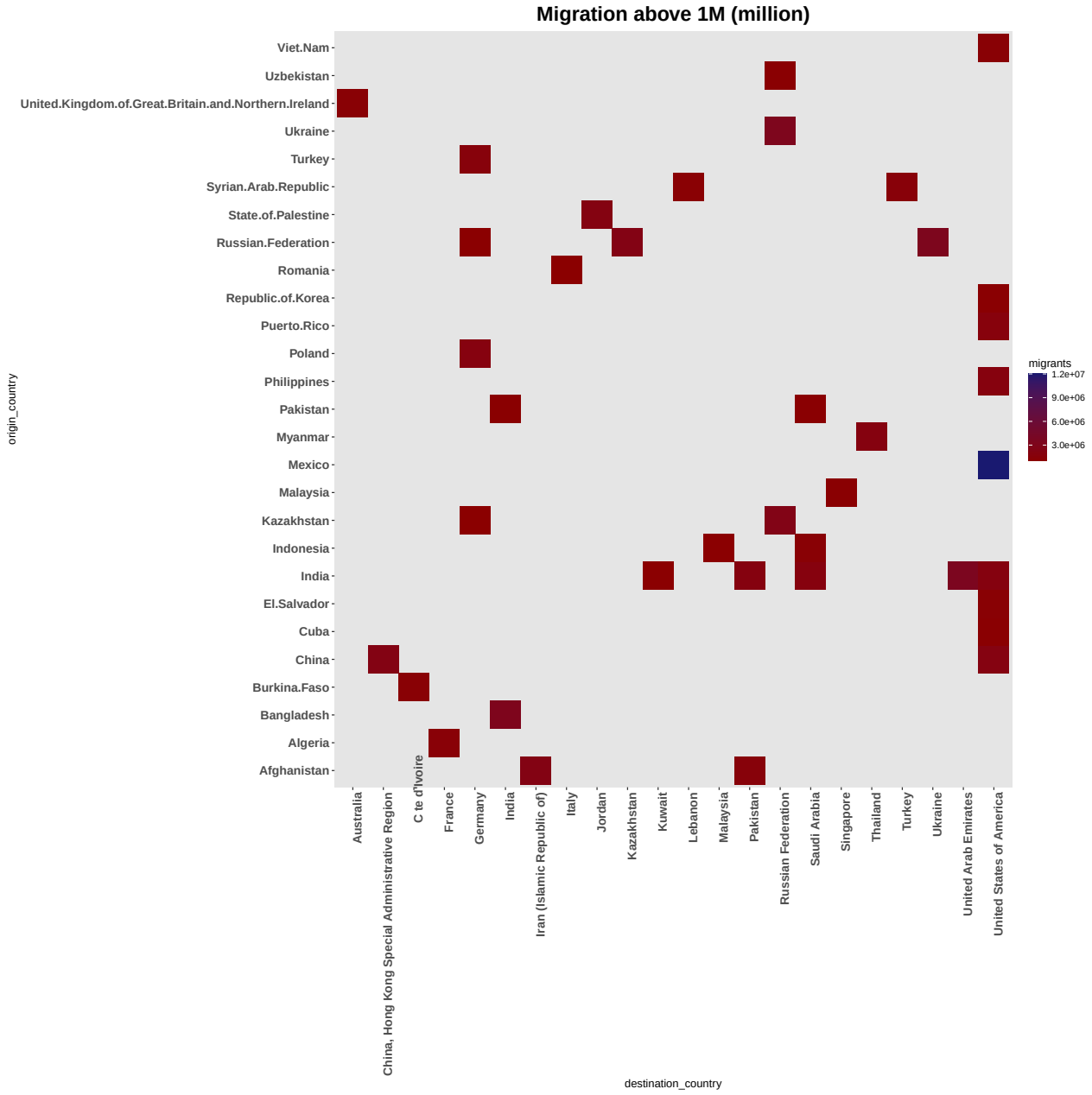
Note: Due to filter immigrants count shows in million.

4.1 Plot Destination country and Origin country mostly migrated using heat map

```
# stored data in data1
data1 <- data.frame(data1)

# plot heatmap using ggplot()
ggplot(data = data1, mapping = aes(x = destination_country, y = origin_country)) +
geom_tile(aes(fill = migrants)) + scale_fill_gradient(low = "darkred", high = "midnightblue") +
theme(panel.grid.major.x=element_blank(),
        panel.grid.minor.x=element_blank(),
        panel.grid.major.y=element_blank(),
        panel.grid.minor.y=element_blank(),
        panel.background=element_rect(fill="gray90"),
        axis.text.x = element_text(angle=90, hjust = 1,vjust=1,size = 12,face = "bold"),
        plot.title = element_text(size=20,face="bold"),
        axis.text.y = element_text(size = 12,face = "bold")) +
ggtitle("Migration above 1M (million)") +
theme(plot.title = element_text(hjust = 0.5))
```

```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## width unknown for character 0x87

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## width unknown for character 0x7f

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## width unknown for character 0x82

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## width unknown for character 0x7f

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## width unknown for character 0x47

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## width unknown for character 0x7f

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## width unknown for character 0x2

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## width unknown for character 0x7f

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## width unknown for character 0xfd

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## width unknown for character 0x7f

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## width unknown for character 0xc3

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## width unknown for character 0x7f

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## width unknown for character 0xf9

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## width unknown for character 0x7f

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## width unknown for character 0xf4

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## width unknown for character 0x7f

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## width unknown for character 0xf6
```

```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## width unknown for character 0x7f

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## width unknown for character 0xf1

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## width unknown for character 0x7f

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## width unknown for character 0xb6

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## width unknown for character 0x7f

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## width unknown for character 0x36

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## width unknown for character 0x7f

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## width unknown for character 0x31

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, : font
## width unknown for character 0x7f

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## font width unknown for character 0xfe

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## font width unknown for character 0x7f
```



4.2 People migrated to United States

```
# filter with destination country
data2 <- data.frame(data) %>% filter(data$destination_country == "United States of America")

# order by desc with immigrants
data2 <- data2 %>% arrange(desc(migrants))

# display data using datatable
datatable(data2,options = list(scrollX = TRUE, paging=TRUE,fixedHeader=TRUE))
```

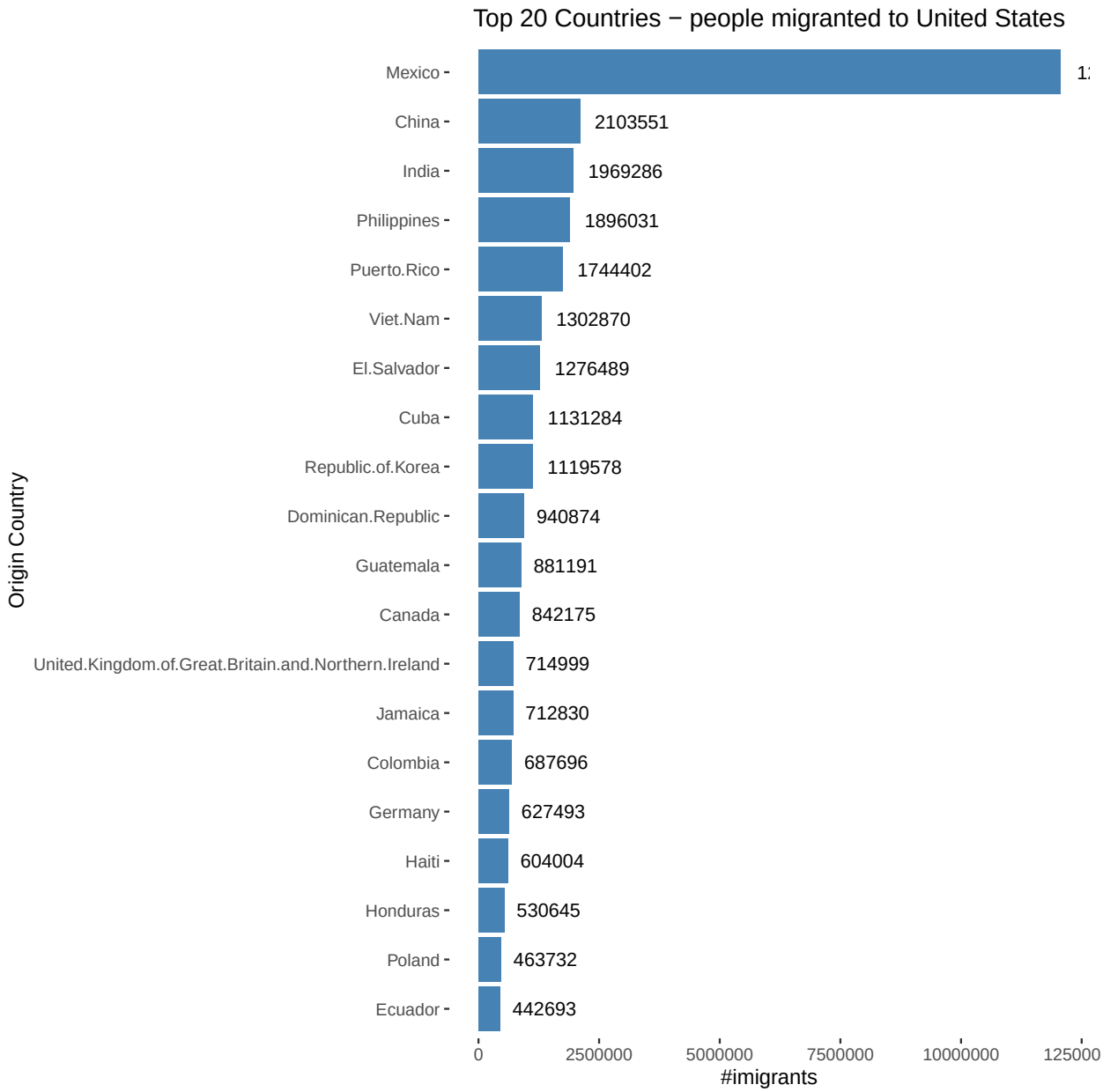
Show 10 entries Search:

	destination_country	country_code	origin_country	migrants
1	United States of America	840	Mexico	12050031
2	United States of America	840	China	2103551
3	United States of America	840	India	1969286
4	United States of America	840	Philippines	1896031
5	United States of America	840	Puerto.Rico	1744402
6	United States of America	840	Viet.Nam	1302870
7	United States of America	840	El.Salvador	1276489
8	United States of America	840	Cuba	1131284
9	United States of America	840	Republic.of.Korea	1119578
10	United States of America	840	Dominican.Republic	940874

Showing 1 to 10 of 150 entries Previous 1 2 3 4 5 … 15 Next

4.3 Plot top 20 country people migrated to United States

```
# stored top 20 top 20 country, people mostly migrated to United States, stored data in var (data2)
data2 <- head(data2, 20)

# plot bar graph using ggplot()
ggplot(data = data2, mapping = aes(x = origin_country, y = migrants)) +
  geom_bar(aes(reorder(origin_country,migrants),migrants),stat = "identity",fill = "steelblue") +
  coord_flip() +  xlab("Origin Country") + ylab("#imigrants") +
  ggtitle("Top 20 Countries - people migranted to United States") +
  theme(plot.title = element_text(hjust = 0.5),panel.background = element_rect(fill = "white", color = 
  geom_text(aes( y = migrants,label=migrants), hjust = -0.20, color="black", size=3.5)
```



4.4 Americans migrated country

```
# filter with origin country
data3 <- data.frame(data) %>% filter(data$origin_country == "United.States.of.America")

# order by desc with immigrants
data3 <- data3 %>% arrange(desc(migrants))

# display data using datatable
datatable(data3,options = list(scrollX = TRUE, paging=TRUE,fixedHeader=TRUE))
```

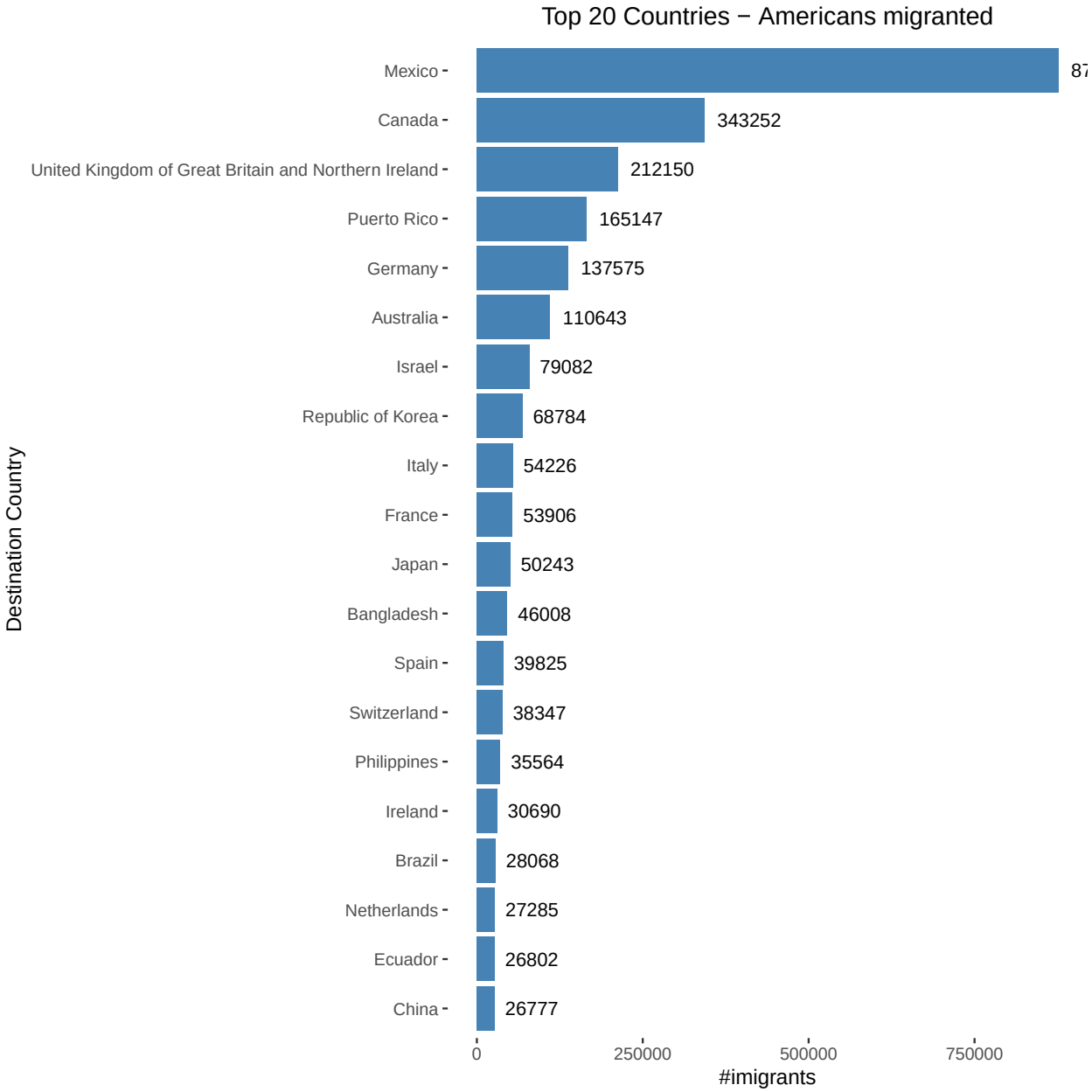
Show 10 entries Search:

	destination_country	country_code	origin_country	migrants
1	Mexico	484	United.States.of.America	876528
2	Canada	124	United.States.of.America	343252
3	United Kingdom of Great Britain and Northern Ireland	826	United.States.of.America	212150
4	Puerto Rico	630	United.States.of.America	165147
5	Germany	276	United.States.of.America	137575
6	Australia	36	United.States.of.America	110643
7	Israel	376	United.States.of.America	79082
8	Republic of Korea	410	United.States.of.America	68784
9	Italy	380	United.States.of.America	54226
10	France	250	United.States.of.America	53906

Showing 1 to 10 of 161 entries Previous 1 2 3 4 5 … 17 Next

4.5 Plot top 20 contries where Americans migrated

```
# top 20 contries where Americans mostly migrated, stored data in var (data3)
data3 <- head(data3, 20)

ggplot(data = data3, mapping = aes(x = destination_country, y = migrants)) +
  geom_bar(aes(reorder(destination_country,migrants),migrants),stat = "identity",fill = "steelblue") +
  coord_flip() +  xlab("Destination Country") + ylab("#imigrants") +
  ggtitle("Top 20 Countries - Americans migranted") +
  theme(plot.title = element_text(hjust = 0.5),panel.background = element_rect(fill = "white", color = 
  geom_text(aes( y = migrants,label=migrants), hjust = -0.20, color="black", size=3.5)
```



5. Conclusion

The data set contains 232 columns, using differnt method of **tydr, dplyr** converted those colums to row. Applied filter condition to get below analysis.

- People migrated from original country to migrated country
- Top 20 original country - where people migrated to United states
- Top 20 desination country- where americans migrated to destination contry