MIDTERM REDESIGN PROJECT

Group 15: Sarang Tirmanwar, Satyam Singh, Jaswanth Erusu

1.INTRODUCTION

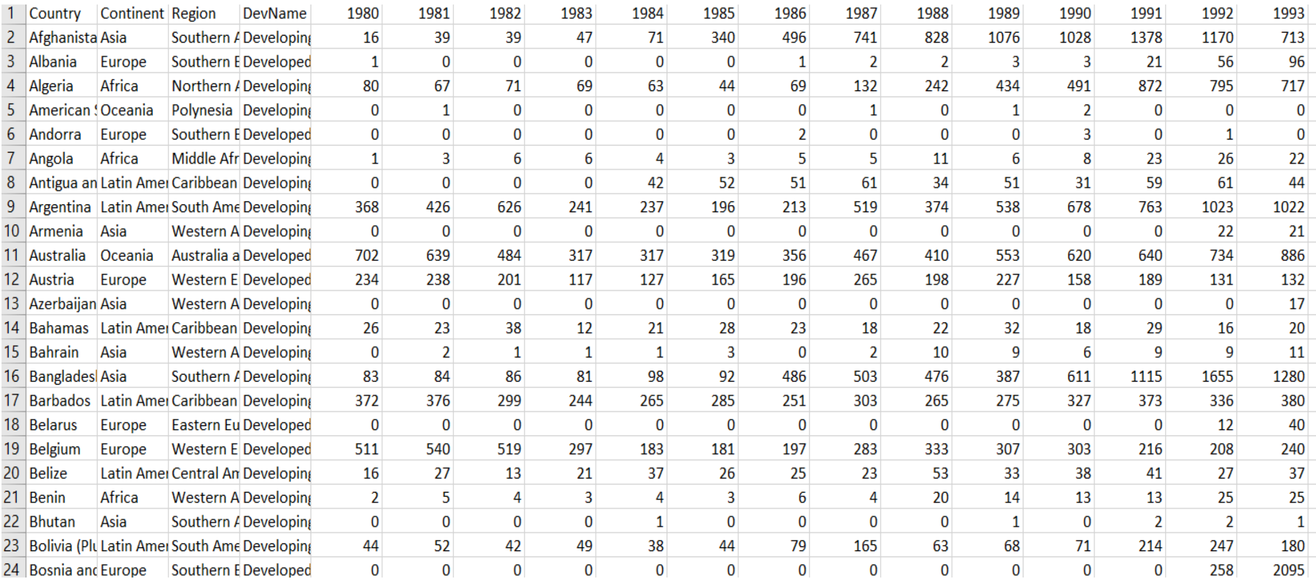
In this report, we will redesign two distinct sorts of bad graphs using all of the tools we learnt in stat 515. We applied the rules presented in class to clearly describe the graphs with correct comparison, and we added support by creating new graphs that emphasize the hidden truths that are not well understood.

We employed many forms of comparisons to better comprehend the plots and data collection related with immigration to Canada. The first graph we designed is a donut plot, and the second is a bar plot. We utilized a Kaggle data set with 195 rows and 39 columns of details regarding immigration to Canada between 1980 and 2013 for our research. We have included information on foreign visitors to Canada in this data set since it is now vital to learn about tourists' plans for the future. There are 33 years' worth of travel-related data from different countries. We have used a number of comparisons to help us understand the narrative and data set associated to immigration to Canada.

The dataset we have enables us to determine the annual immigration population by continent, region, and nation. Canada is one of the nations that receives the most immigrants each year.

We have improved our help by creating more graphs that draw attention to the hidden facts that are not readily comprehended, using the guidelines presented in class to clearly describe the graphs with precise comparison.We also utilized Rshiny to improve user comprehension and to make our project more user-friendly.

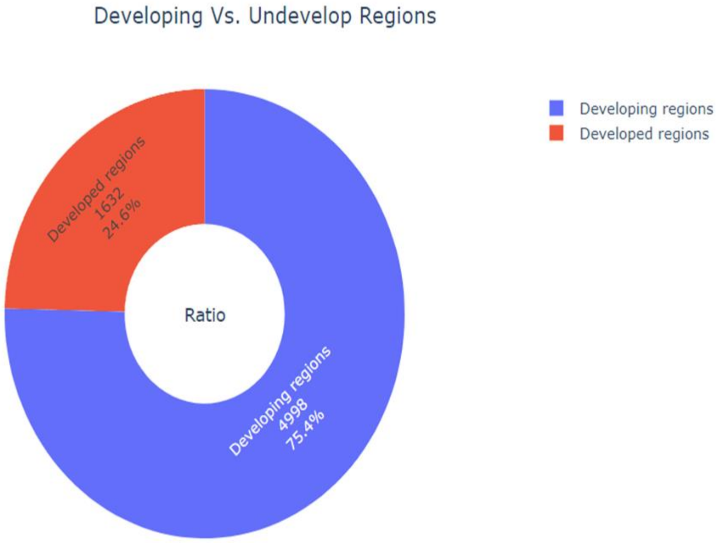
DATA SET SAMPLE:



1	Country	Continent	Region	DevName	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993
2	Afghanista	Asia	Southern A	Developing	16	39	39	47	71	340	496	741	828	1076	1028	1378	1170	713
3	Albania	Europe	Southern E	Developed	1	0	0	0	0	0	1	2	2	3	3	21	56	96
4	Algeria	Africa	Northern A	Developing	80	67	71	69	63	44	69	132	242	434	491	872	795	717
5	American S	Oceania	Polynesia	Developing	0	1	0	0	0	0	0	1	0	1	2	0	0	0
6	Andorra	Europe	Southern E	Developed	0	0	0	0	0	0	2	0	0	0	3	0	1	0
7	Angola	Africa	Middle Afr	Developing	1	3	6	6	4	3	5	5	11	6	8	23	26	22
8	Antigua an	Latin Amer	Caribbean	Developing	0	0	0	0	42	52	51	61	34	51	31	59	61	44
9	Argentina	Latin Amer	South Ame	Developing	368	426	626	241	237	196	213	519	374	538	678	763	1023	1022
10	Armenia	Asia	Western A	Developing	0	0	0	0	0	0	0	0	0	0	0	0	22	21
11	Australia	Oceania	Australia a	Developed	702	639	484	317	317	319	356	467	410	553	620	640	734	886
12	Austria	Europe	Western E	Developed	234	238	201	117	127	165	196	265	198	227	158	189	131	132
13	Azerbaijan	Asia	Western A	Developing	0	0	0	0	0	0	0	0	0	0	0	0	0	17
14	Bahamas	Latin Amer	Caribbean	Developing	26	23	38	12	21	28	23	18	22	32	18	29	16	20
15	Bahrain	Asia	Western A	Developing	0	2	1	1	1	3	0	2	10	9	6	9	9	11
16	Banglades	Asia	Southern A	Developing	83	84	86	81	98	92	486	503	476	387	611	1115	1655	1280
17	Barbados	Latin Amer	Caribbean	Developing	372	376	299	244	265	285	251	303	265	275	327	373	336	380
18	Belarus	Europe	Eastern Eu	Developed	0	0	0	0	0	0	0	0	0	0	0	0	12	40
19	Belgium	Europe	Western E	Developed	511	540	519	297	183	181	197	283	333	307	303	216	208	240
20	Belize	Latin Amer	Central Am	Developing	16	27	13	21	37	26	25	23	53	33	38	41	27	37
21	Benin	Africa	Western A	Developing	2	5	4	3	4	3	6	4	20	14	13	13	25	25
22	Bhutan	Asia	Southern A	Developing	0	0	0	0	1	0	0	0	0	1	0	2	2	1
23	Bolivia (Plu	Latin Amer	South Ame	Developing	44	52	42	49	38	44	79	165	63	68	71	214	247	180
24	Bosnia and	Europe	Southern E	Developed	0	0	0	0	0	0	0	0	0	0	0	0	258	2095

2 REDESIGN 1

2.1 INTRODUCTION:



This bad graph was obtained from Kaggle, a data science and artificial intelligence website. Large firms and groups post contests with monetary awards on this site. In addition to the contests, users may contribute their datasets and study those of others. This bad graph consists of two regions, one developed and one developing, and it attempts to convey the number of immigrants who have travelled to Canada from these regions. However, this graph is so confusing that we are unable to extract any useful information.

WHY THIS GRAPH IS A BAD GRAPH?

1. We can see that this graph is difficult to grasp since the axis isn't properly described, we can't see the years (1980–2013), and there isn't a lot of information there, which makes comparisons challenging.

2. This graph is missing some crucial components for communicating information, such as the year in which the immigration population in Canada peaked. Merely adding colors won't make this graph more interesting either.The theme of presentation will not be covered by simply representing the graph.

3. The limitations mentioned above make it challenging to entice readers to study the information this graph may provide. In addition, we believe that the comparison is not

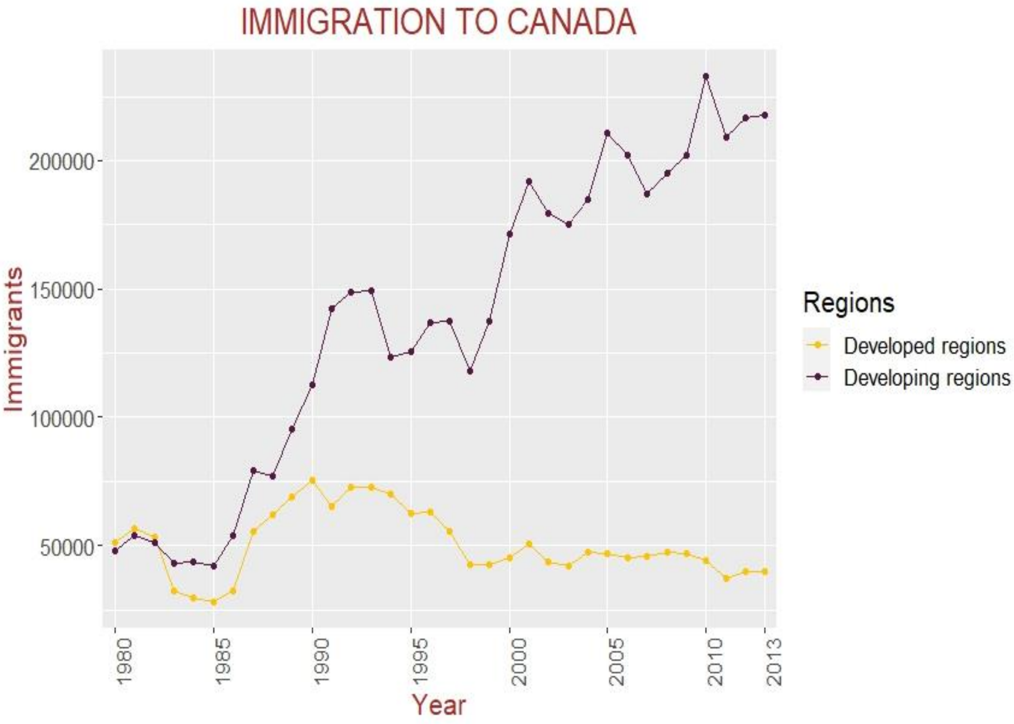
accurate since we are unsure of the meaning of the numbers and what they actually represent.

2.2 GOALS

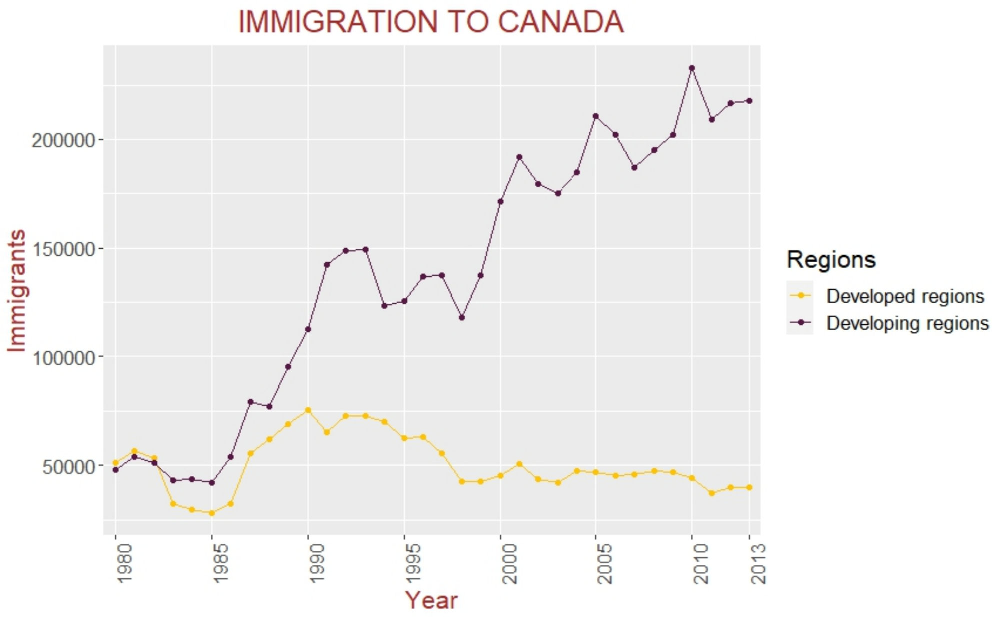
This poor redesign example seeks to address the above drawbacks by categorizing them according to distinct visual categories using a scientific standard, providing relevant context, and modifying visually appealing presentations.

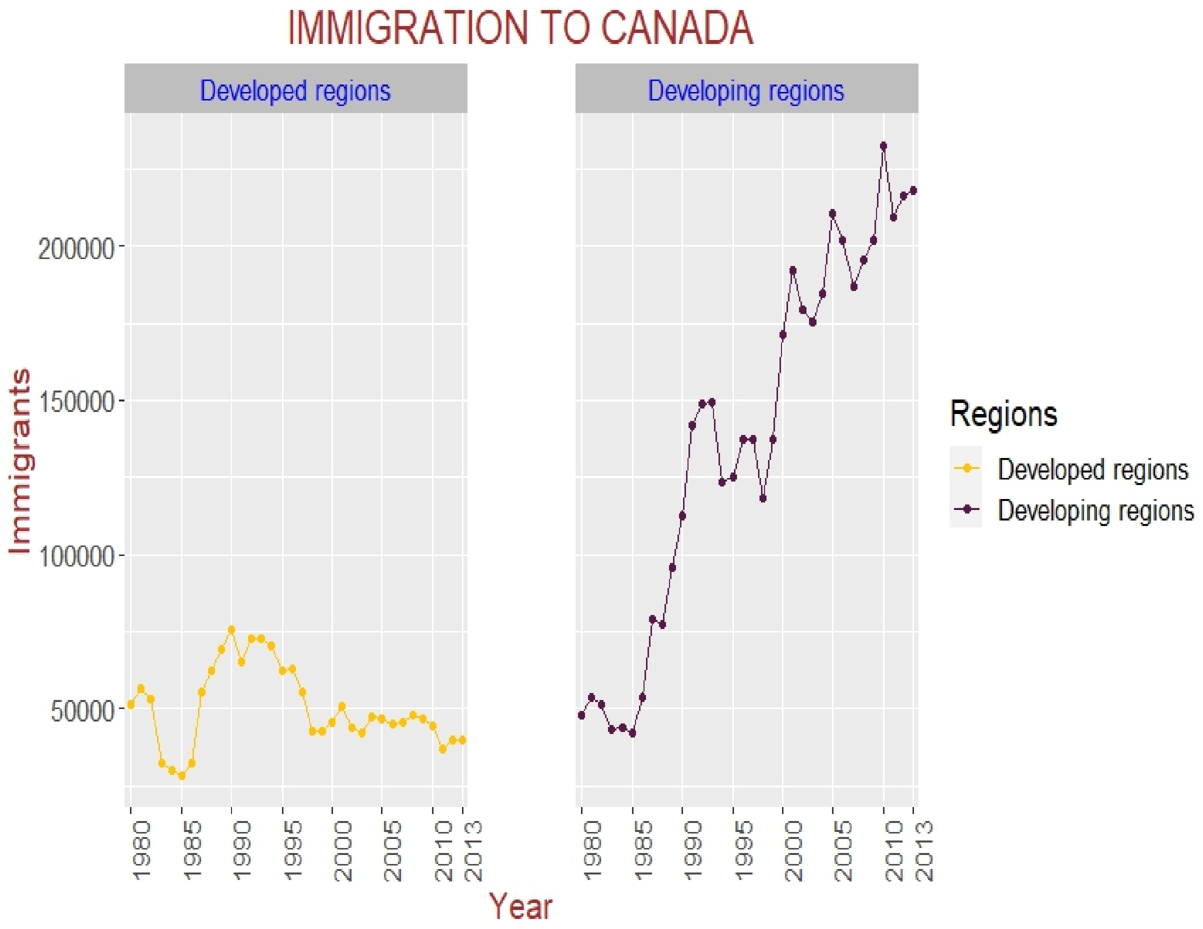
2.3 PROCESS

The primary tools used in this project are R Studio, the tidyr and dplyr packages for data preprocessing, the ggplot2 package for graphing, and the facet wrap () function for easier user comprehension. With the help of facet wrap(), we can easily see the graph separated into developed and developing regions. To make the graph more comprehensible and appealing, we removed the x and y axes. where a developing area is shown by the purple line and a developed area by the yellow line. The x and y axes, respectively, show the years and the total number of immigrants.



2.4 OBSERVATION AND RESULTS OF THE REDESIGN:

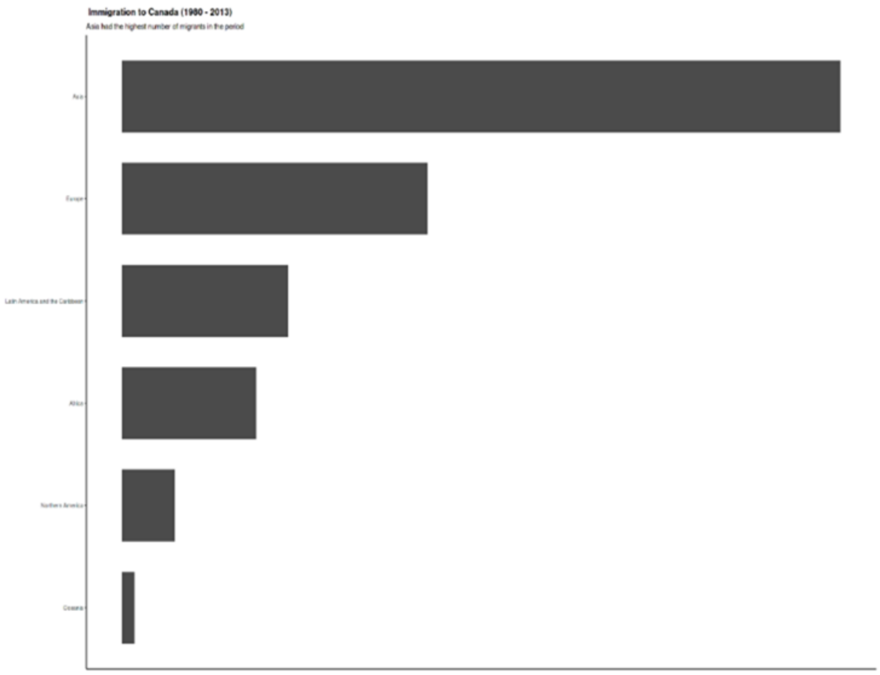




- First of all, when compared to the original graph, we can quickly see how many people immigrated to Canada between 1980 and 2013.
- The Y axis in the following revised graph displays the amount of immigrants, making it simple to compare developed and developing regions.
- According to our analysis, the number of immigrants from developing regions increased starting in the 2000s, whereas the number of migrants from developed regions decreased significantly between 2005 and 2013 and increased significantly between 1980 and 1990.
- The graph seems lot cleaner and is simpler to read after redesign.

3 REDESIGN 2

3.1 INTRODUCTION:



The graph above, which is also from Kaggle, shows how many individuals have immigrated to Canada per continent, but it is obvious that the information is so unclear that we are unable to make any inferences from it. The continents are shown on the Y axis, while there is nothing on the X axis.

WHY THIS GRAPH IS A BAD GRAPH?

1. The viewer may still be confused by this graph because the X-axis does not show the total number of immigrants. Therefore, it does not provide a clear comparison.
2. In the second poor graph, readers are unable to compare any values, and the total number of immigrants is not represented, which may have made the graph easier to grasp. However, sometimes readers want to know how the comparisons were made in order to grasp.
3. Only comparing which continent has more in relation to others defies the logic of bar charts. We can see that the graph has no colors for any of the continents. The reader will have a tough time making the connection as a result. No colour scheme might also lessen the reader's attraction.

3.2 GOALS

The goal of providing this second example of a poor redesign is to eliminate the flaws by introducing an x axis, the number of immigrants who came to Canada, and a color scheme to make the graph more appealing to the eye.

The second redesign graph should have improved in a number areas after being revised, including ease of comparison, easier access to more information than previously, enhanced reader interpretation, and increased reader attraction.

Also, we aim at making it more precise so that the audience can easily interpret and understand the figures related to each continent.

3.3 PROCESS

R Studio, the programs tidyr and dplyr for data preparation, and the ggplot2 package for graphing are the main tools used in this project.

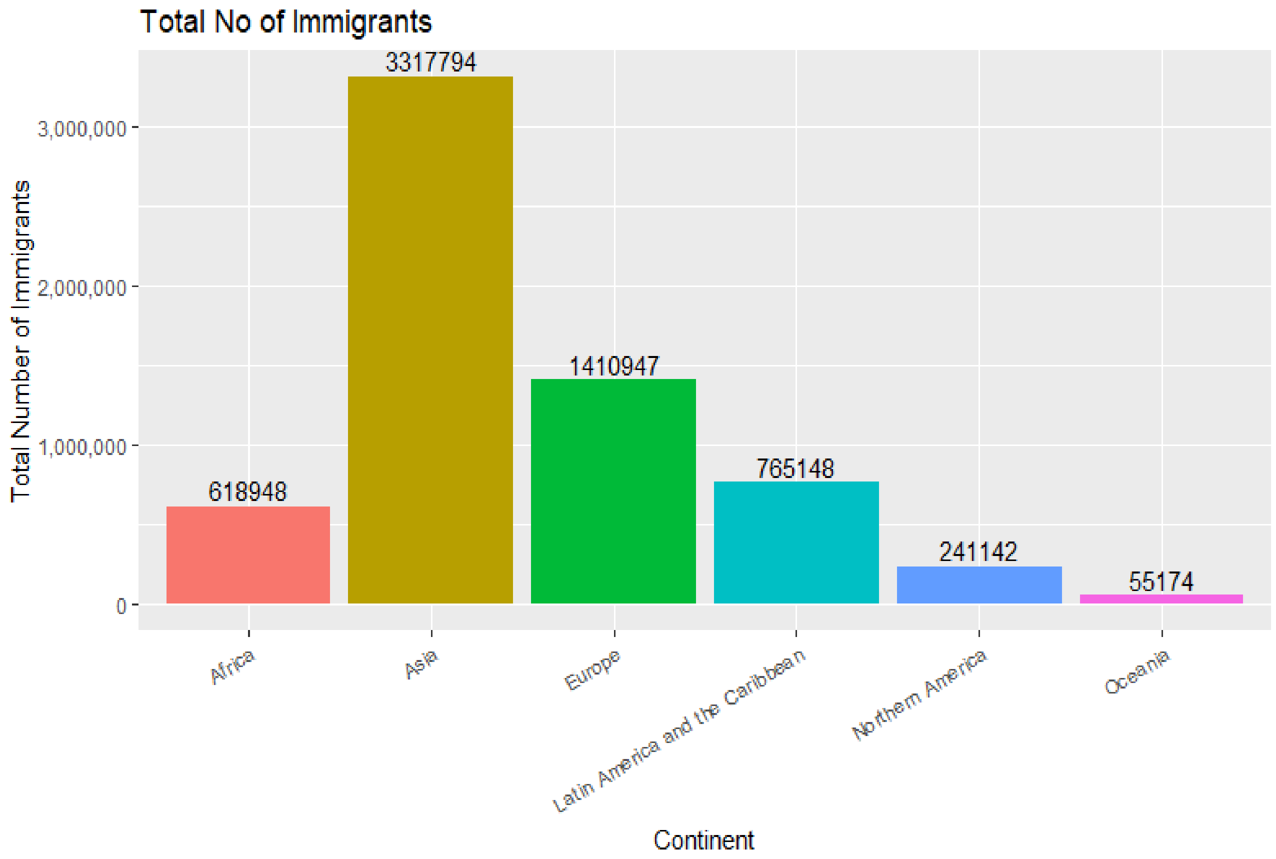
As you can see after redacting the graph, the y axis now indicates the overall number of immigrants, while the x axis now shows the continents. This was done to improve understanding.

The names of the continents are easier to see on the X-Axis since they have been slanted.

To make our graph more appealing visually, we also added various colors to various types of continents.

In order to provide precise information, each bar is labelled with the specific number of immigrants.

3.4 OBSERVATION AND RESULTS OF THE REDESIGN:



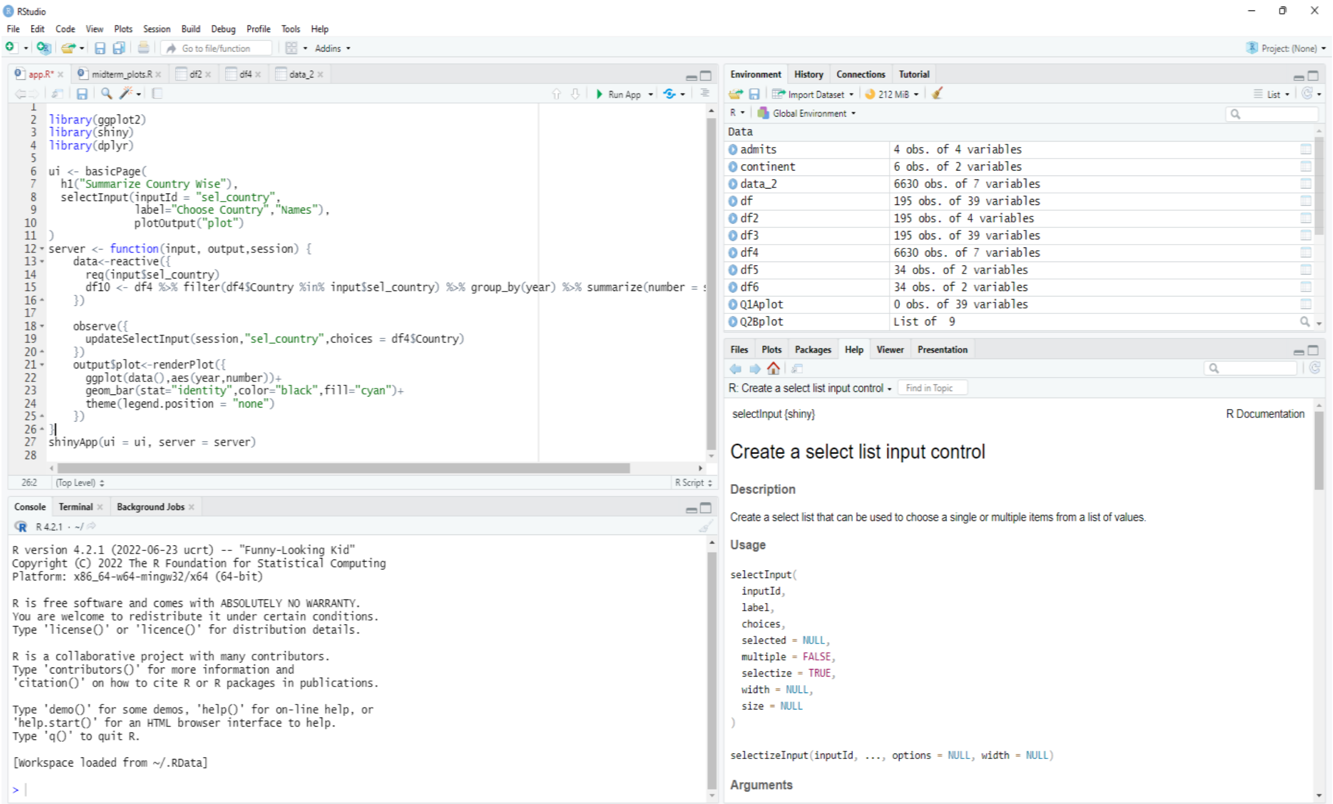
- The revised graph addresses all issues with the original, including the lack of an axis in the earlier, less-than-ideal graph and the addition of colours, which make it simpler for readers to understand how many individuals travelled to Canada from the continents shown above.
- We can state with ease that Asia is the continent from which the greatest number of people moved to Canada, followed by Europe and then Oceania, where the least number of people went.

CHALLENGES

The first difficulty in this project is choosing the examples because it might be difficult to locate datasets for the graphs we want to remodel. We spent a lot of effort on it since we had to locate correct dataset. Then, in the first redesign example, we found it difficult to simply plot the graph separately for the developed and developing regions, and even after we had done so, we thought it was not aesthetically appealing. So, we used facet wrap() to divide both regions so that we could get a clear sense individually.

After finding certain graphs our main focus was to make the reader easily grasp the information so that they will need to take minimum efforts to understand. Just by simply looking at the graphs the reader can understand many things unlike the previous one.

4.R- Shiny



```
library(ggplot2)
library(shiny)
library(dplyr)

ui <- basicPage(
  h1("Summarize Country Wise"),
  selectInput(inputId = "sel_country",
              label="Choose Country","Names"),
              plotOutput("plot")
)
server <- function(input, output,session) {
    data<-reactive({
      req(input$sel_country)
      df10 <- df4 %>% filter(df4$Country %in% input$sel_country) %>% group_by(year) %>% summarize(number = s
    })

    observe({
      updateSelectInput(session,"sel_country",choices = df4$Country)
    })
    output$plot<-renderPlot({
      ggplot(data(),aes(year,number))+
      geom_bar(stat="identity",color="black",fill="cyan")+
      theme(legend.position = "none")
    })
}
shinyApp(ui = ui, server = server)
```

```
R version 4.2.1 (2022-06-23 ucrt) -- "Funny-Looking Kid"
Copyright (C) 2022 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Workspace loaded from ~/.RData]

> 
```

```
Data
admits       4 obs. of 4 variables
continent    6 obs. of 2 variables
data_2       6630 obs. of 7 variables
df           195 obs. of 39 variables
df2          195 obs. of 4 variables
df3          195 obs. of 39 variables
df4          6630 obs. of 7 variables
df5          34 obs. of 2 variables
df6          34 obs. of 2 variables
Q1Aplot      0 obs. of 39 variables
Q2Bplot      List of  9
```

selectInput {shiny}   R Documentation

Create a select list input control

Description

Create a select list that can be used to choose a single or multiple items from a list of values.

Usage

```
selectInput(
  inputId,
  label,
  choices,
  selected = NULL,
  multiple = FALSE,
  selectize = TRUE,
  width = NULL,
  size = NULL
)

selectizeInput(inputId, ..., options = NULL, width = NULL)
```

Arguments

This is the code we used for Rshiny. It is a very helpful tool and extremely easy to create statistical Web Apps.

Firstly, we installed R-Shiny libraries and loaded them. R Shiny has basically 2 components, the R-Shiny user interface, and the R-shiny server function. Both functions are passed to the R-Shiny App function as arguments. The R-Shiny app function then creates a shiny app object from this UI/ server pair.

Our R-shiny code is quite simple to understand,

In the UI interface we make a basic page with a simple heading and using the select input function we give inpud Id and label it. Also, we name the Output Plot as plot.

In the main server function, we define the arguments and choose the reactive function which helps in plotting and setting the App. As according the country we need to keep changing the plots, we should use the reactive function which takes the input value as its argument.

After taking the input value, we use some data manipulation methods from dyplyr library which helps us summarize and filter a particular country based on the input value.

Lastly, we use the ggplot library to plot the manipulated data above to get our country specific plot regarding the population from 1980 to 2013.

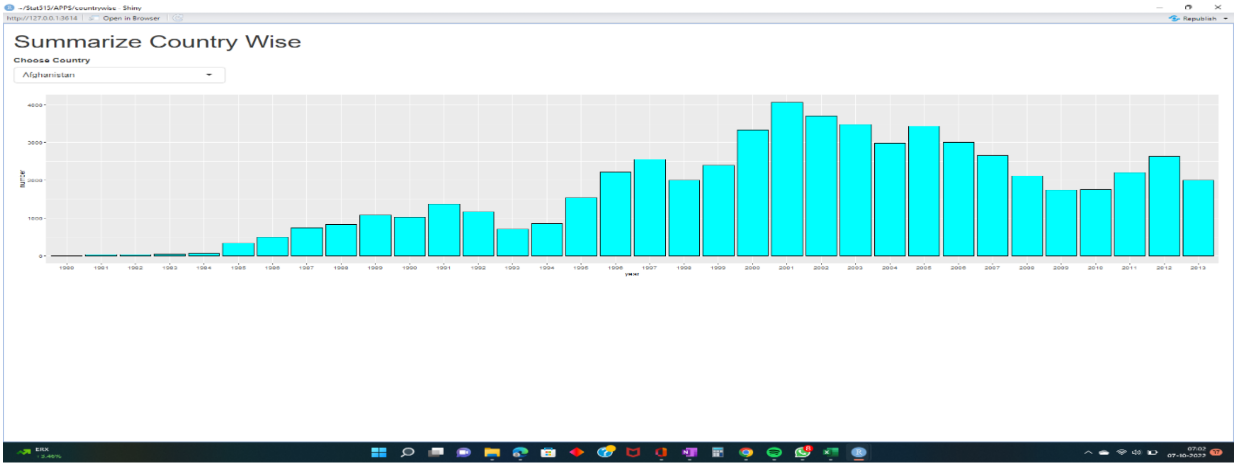
We should also note that the observe function takes in the value of the labels given before in the UI function which helps it relate to the part in the code. Also, to have many choices we select the data frame with the column name of country to get multiple choices.

Below is the snapshot of our code and working.

I have also uploaded on the internet using the Rshiny apps.

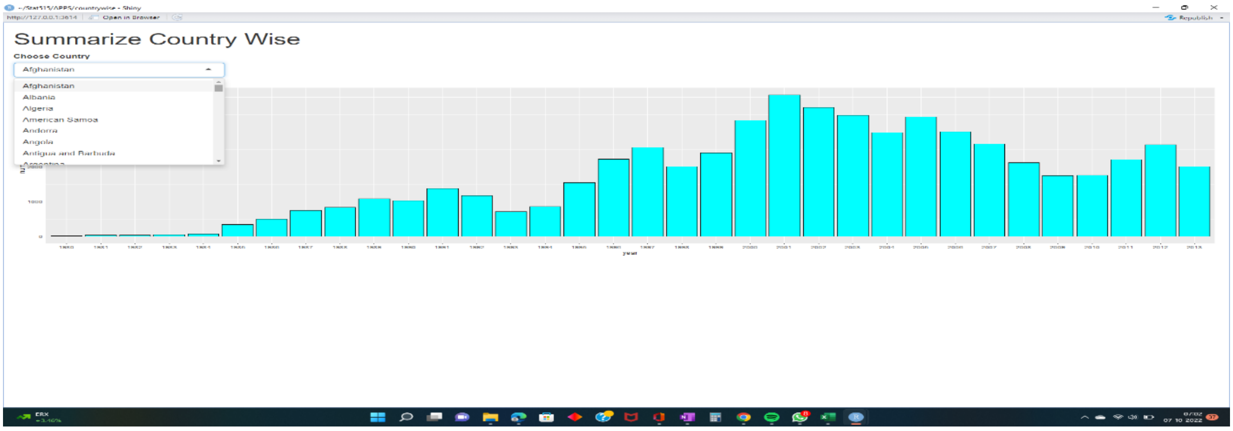
Here is the link.

https://satyam17.shinyapps.io/countrywise/?_ga=2.50314662.1688704609.1665192088-1452630271.1664999684



Summarize Country Wise

Choose Country: Afghanistan



Summarize Country Wise

Choose Country: Afghanistan

Conclusion

In this project, we updated two distinct kinds of poor graphs and enhanced them in four ways: accurate comparisons, simplified presentation, understanding, and reader engagement. We employ all STAT515-learned packages, such as "ggplot2," "tidyr," and "dplyr," during this approach. Along with learning how to plot graphs using the ggplot2 package, we also gained knowledge of the capabilities of tidyr and dplyr and learnt how to produce appealing graphs.These functions also helped us changing data sets as per the needs. As we know we might have appropriate data but to plot it we need to keep changing the datasets columns and rows as per the need for plotting which may not be the case all the time.

With the help of R-Shiny, we understood the ease of plotting statistical Web apps, these help in increased audience interaction which is one of the key motives. They help user in better understand and help in direct comparison.

References

Immigration to canada | Kaggle

Canada Immigration | Kaggle