STAT 515: Redesign Project
(Faie Almofeez and Navya Mote)

Very complex graphs or graphs that are poorly constructed can obscure its purpose and negatively impact the viewer. Usually graphs are used to provide a visual representation of the data and show an easier way to present the information. Therefore, graphs have to be easy to read and understand. Most graphs are open to interpretation, so it is important to represent the data in a clear and concise way. Problems like missing text or inconsistent scale can cause the viewer to spend a lot of time trying to extract the information and figure out what the author is trying to say. In this report, a graph from World Development Report 2018 team was redesigned to better represent the data and to fix some of the issues associated with it.

A very brief indication of bad graph or table context, content and intent

Title : School completion is always higher for children from wealthier families and urban settings, whereas gender gaps are more mixed
Subtitle : Gaps in grade 6 completion rates (percent) for 15-19-year-olds, by wealth, location, and gender

Figure O.6 School completion is higher for richer and urban families, but gender gaps are more context-dependent

Gaps in grade 6 completion rates (percent) for 15- to 19-year-olds, by wealth, location, and gender
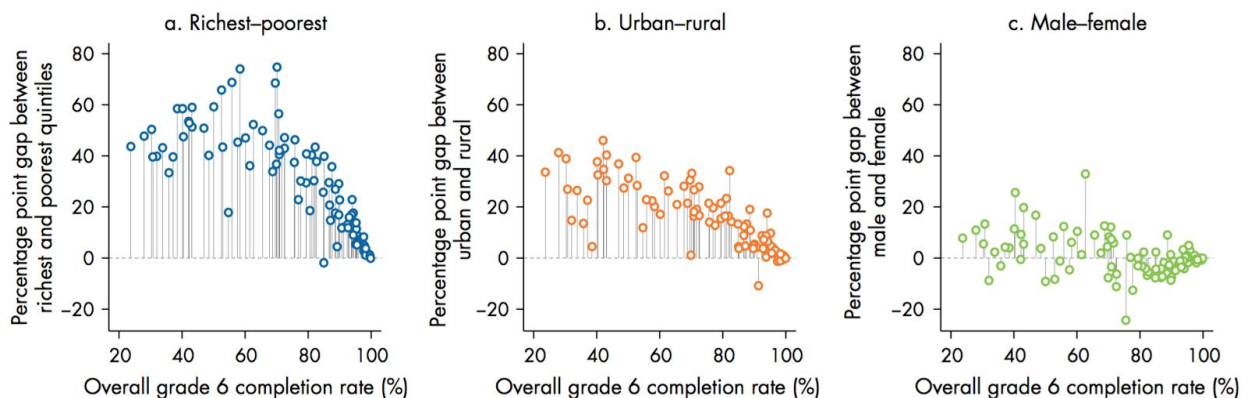
Source: WDR 2018 team, using data from Filmer (2016). Data at http://bit.do/WDR2018-Fig_O-6.

Note: The data presented are the latest available by country, 2005–14. Each vertical line indicates the size and direction of the gap for a country.

Figure O.6 displays grade 6 completion rates in developing countries from WDR (World Development Report). Each graph shows a factor that affects school completion in these countries. Graph A shows the gaps between rich and poor, Graph B shows the gaps between urban and rural and Graph C to the right shows the gap between male and female children.The x-axis is the overall percentage of completion and the y-axis shows the percentage point gap. The countries are shown in the colored circles and the grey vertical line symbolizes the direction of gap for each country. The data presented is the latest available by country, 2005-14.

The source of the bad graph :
Source : WDR 2018 team, using data from Filmer (2016).
Report URL : http://www.worldbank.org/wdr2018
Data URL : http://bit.do/WDR2018-Fig_O-6

A brief indication visual communication flaws or failure to show patterns

      The reason this graph was chosen is because of the poor representation of the information which made it more confusing for the viewer to understand and we believe that this graph can be better constructed to represent the information in much simpler and clearer way. This graph is showing three different pieces of information in one plot which are the overall completion rate, the percentage point gap and the countries. However, one of these information, which is the name of the countries, was not represented well. As shown in figure 1, each one of the countries represented by a circle but you can't actually tell which country is which and also there is no legend to clarify that. This issue causes the comparison between the factors and the countries impossible. Another problem in this graphs is that the circles are overlapping which make it harder for the viewer to distinguish between the different countries. Last, the missing grid could also make the graphs more confusing for the viewer.

A brief comment about special efforts. Obtaining, processing or creating data or the graphics side of production.

      The process of finding a graph to redesign was kind of challenging. We had to look through many graphs and analyze them and see if there is a way we can redesign them to look better. Through our search, we wanted to choose a graph that was challenging and had some meaningful data. The first few choices were not successful because we struggled finding the raw data of these graphs. Once we decided on the gap in grade 6 completion rate graph and we obtained the raw data for that graph, we had to rearrange these data and format the values in CSV to fit better in Rstudio. Then, we had to read the CSV file into our R script. The first attempt in redesigning this graph is using dot plots to redo the graph entirely as shown in figure 2 but then felt the challenge was in representing the graph the way it is (which is complicated) and to make it better than it already is.

Describe suggested improvements and their correspondence to class (or other) design and reasoning guidelines.
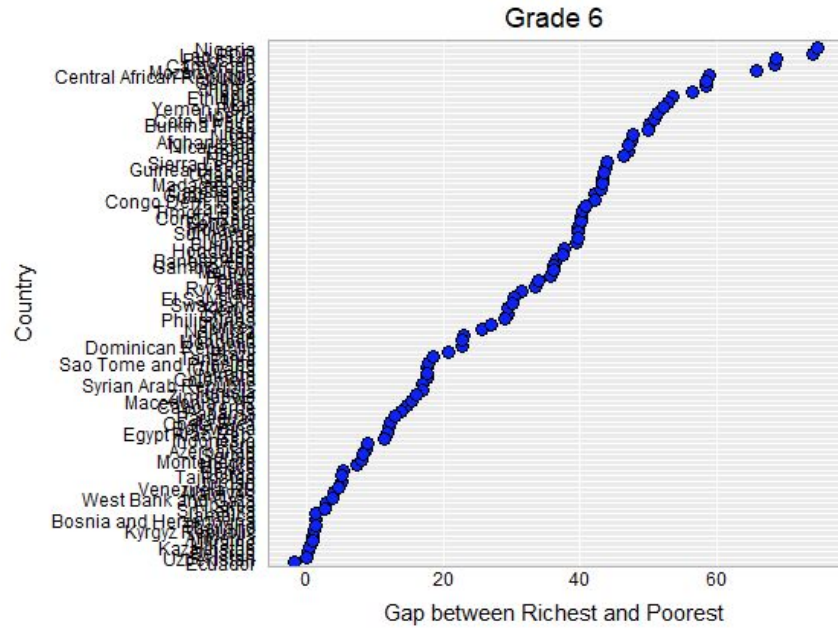
Attempt 1 - Graph 1:

Figure 1

- In order to redesign the graphs we started writing the scripts initially using R file. However, the y-axis which represents the countries was messed up and all the countries names were overlapping (Fig 1) and it just didn't work.
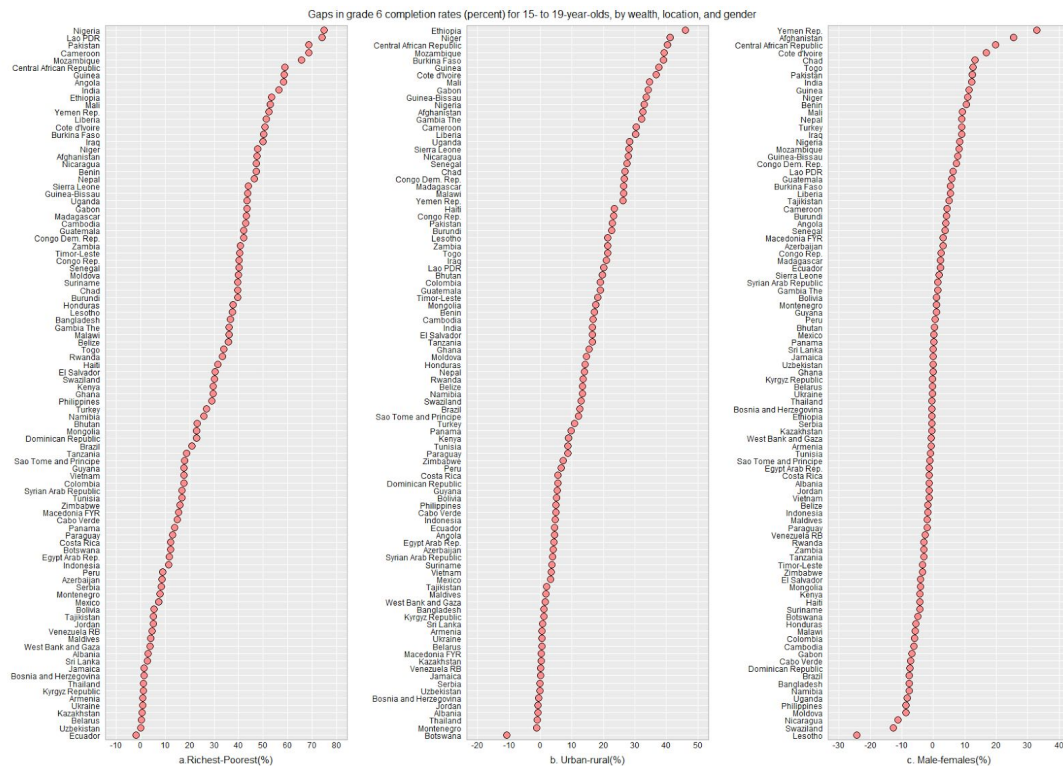
Attempt 2 - Graph 2:

Figure 2

- Using R markdown worked best with the size of the window and the data corresponding to 96 countries fit in perfectly.
- We arranged the countries in increasing order of the gap by using reorder along y axis and descend = true because sorting the countries would make it easier for the viewer to read and compare gaps of different countries
- Using scale_x_continuous helped to change the axis sequence
- We could put 3 graphs in one window using grid.arrange (Fig 2)
- As our graph looked very different from the bad graph, we decided to try again to make our graph look more like the bad graph while making the data interpretation better.
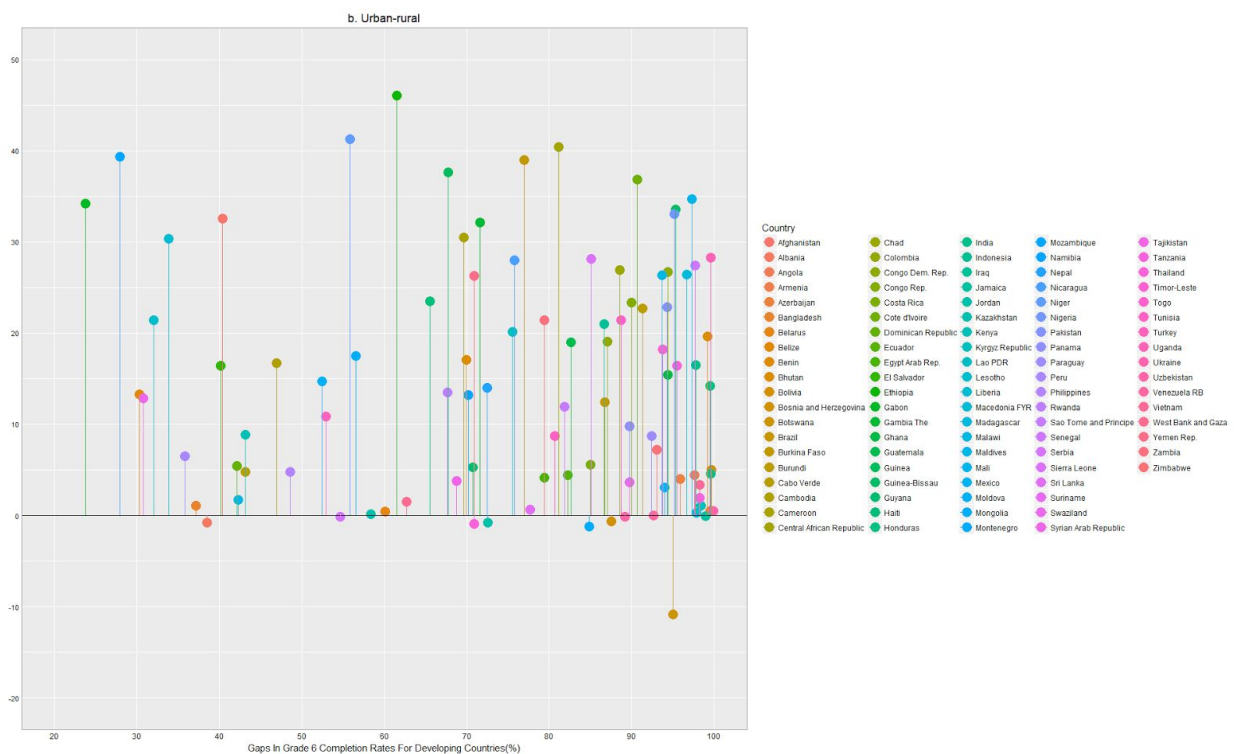
Attempt 3 - Graph 3 :

Figure 3

- In order to represent the graph as it is, we had to implement functionalities that we had not worked on before.This took some time and effort to try out the possibilities.
- The original graph has three pieces of data that needed to be represented in one graph. However, in order to implement this method, we learnt to rearrange the data in an order that could be read in R, to change the axis values, to represent each dot as a country and to give separate colours to them(using aes(color))
- To add vertical lines in order to better indicate the direction of the gap from (using geom_segment)
- To highlight the 0 line(using geom_hline (y-intercept = 0))
- To change the font of the titles(using axis.title.x/axis.title.y)

- But the problem with this graph was that , we were not able to distinguish between the 96 countries clearly as some colours look similar.
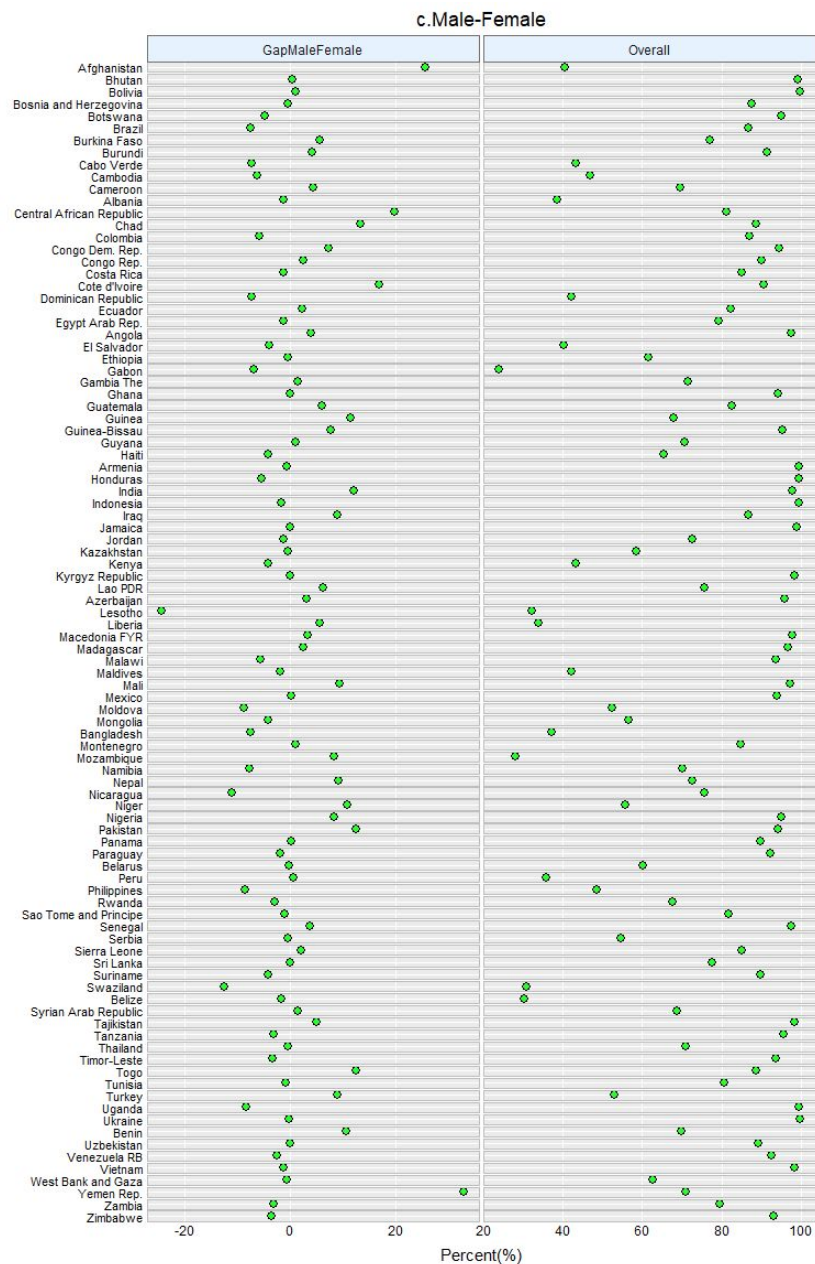
Attempt 4 - Graph 4:

Figure 4

- We tried using facet_grid to group the data so that we can represent the gap and the overall completion rate next to each other.
- This way we can see the gap % and the overall % in one straight line against the country name.

- We felt that this way of representing the graph was better to understand the respective gaps for the countries individually.

Discuss redesigned graph and the patterns that its shows :

The end result of our redesigned project is shown in figure 4. Since the original graph has three set of data, our best option was to split the overall completion rate and the percentage completion gap into two panels. Each panel represent a set of data against the respective country. The original graphs were missing the country names but by redesigning the graphs this way, we were able to present the country names along with the other two sets of data. We decided to arrange the countries alphabetically so that the difference between one factor (eg: Richest-poorest/Urban-Rural/Male-Female) to another can be easily spotted (Appendix B).

Make one more concluding remarks

We feel that there is more than one way of making the graph better. If we could try more ways, it would be to put all the three graphs in one window so that the comparison is easier. We could sort the gap percent in descending order so that we can be able to clearly spot the countries that have a positive gap and the ones that have a negative gap. We could also group the countries in sets of five each which would make the graph easier to focus on.

After trying different ways of making the bad graph better, we would like to conclude by saying that our graph shown in Fig. 4 has been able to better represent the gaps for all the countries individually. With the help of which we can clearly identify the names of the countries and their corresponding gaps. Tomorrow if an analyst wants to derive insightful information related to a specific developing country, they can do so with the help of our graph.

Appendix

A. Script

```
---
title: "Male-Female Gap"
author: "Faie Almofeez, Navya Mote"
date: "October 29, 2017"
input: "CSV File"
---
```{r echo=FALSE}
library(tidyverse)
library(lattice)
source('hw.r')
library(gridExtra)

colNames <- function(x){
  colNam = colnames(x)
  colInt = 1:length(colNam)
  names(colInt) = colNam
  return(colInt)
}
```


```{r}
studentMaleFemale <- read.csv(file="GapMaleFemale - Copy.csv")
studentMaleFemale
```


```{r echo=TRUE, fig.width=8, fig.height=12}
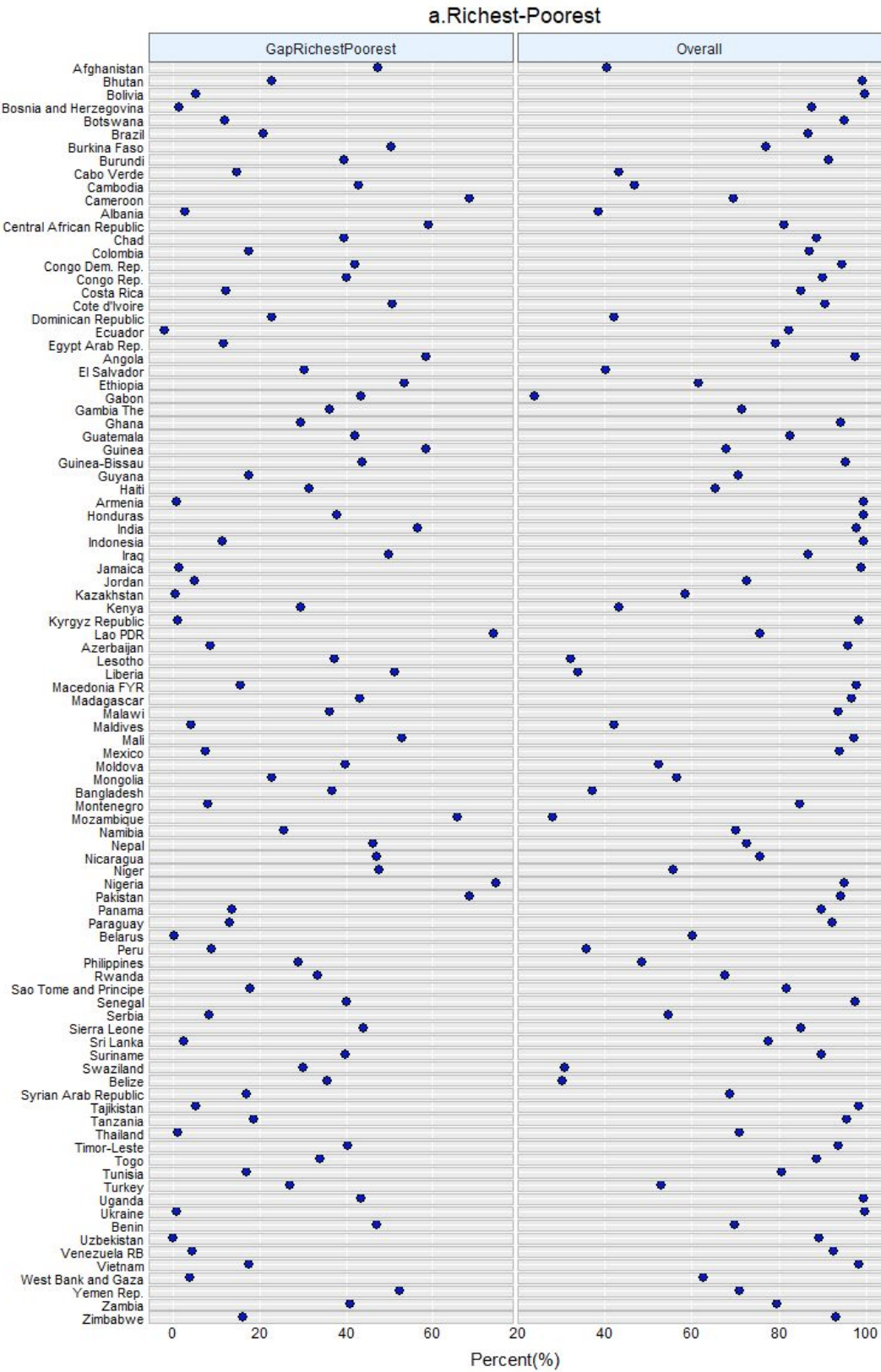
ggplot(studentMaleFemale,
  aes(x = Values, y = reorder(Country,Values), descend = TRUE)) +
geom_point(shape = 21, fill = "#00FF00",
  col = "black",size = 2.5) +
  facet_grid(Group~Gap, scale = "free",
  labeller = labeller(varName = stripLabs)) +
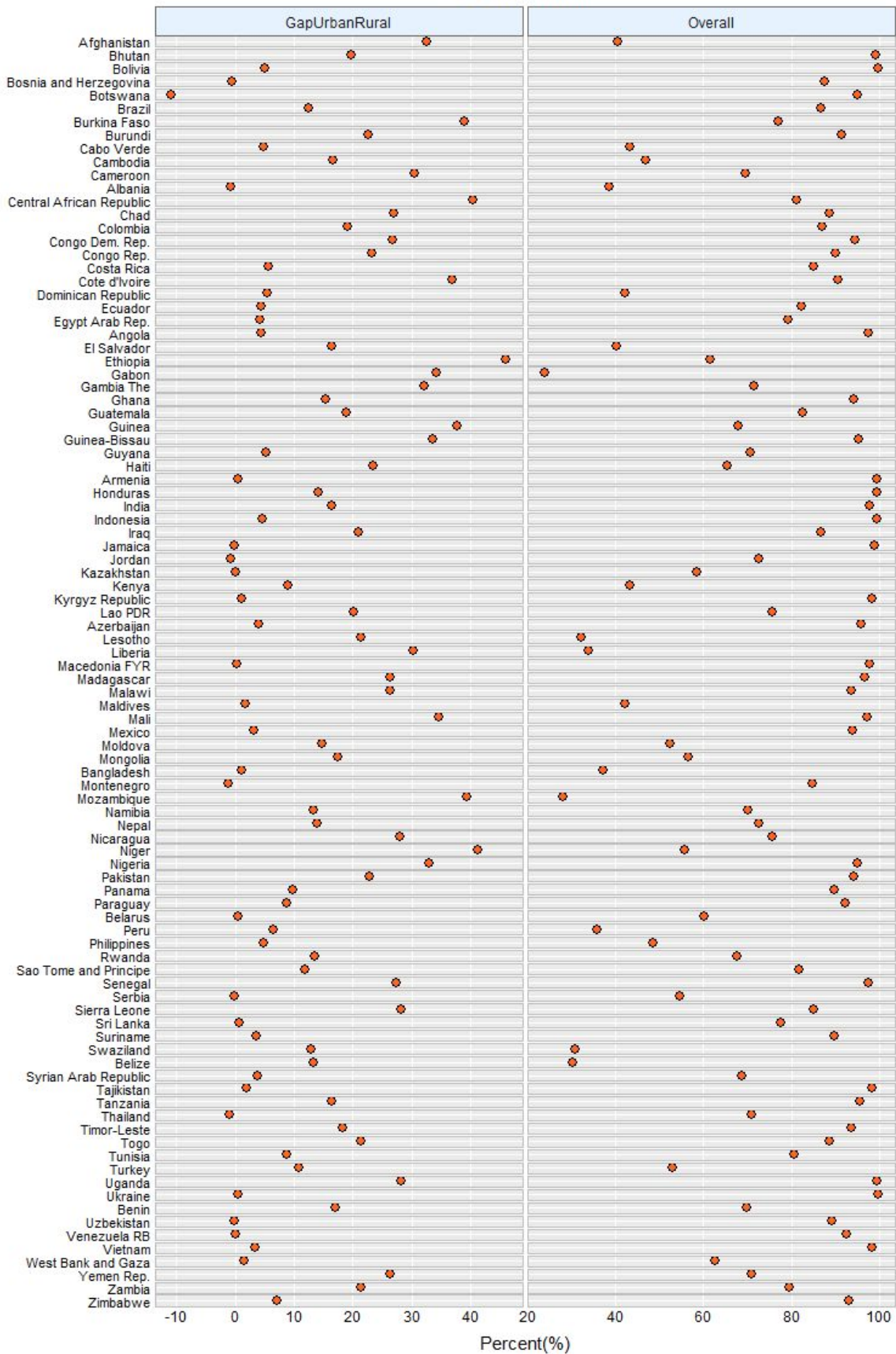labs(x = "Percent(%)",y = '',
  title = paste0("c.Male-Female")) + hw
```
```

B. All Redesigned Graphs

a.Richest-Poorest

b.Urban-Rural

Figure comparing GapUrbanRural and Overall percentages across countries: Afghanistan, Bhutan, Bolivia, Bosnia and Herzegovina, Botswana, Brazil, Burkina Faso, Burundi, Cabo Verde, Cambodia, Cameroon, Albania, Central African Republic, Chad, Colombia, Congo Dem. Rep., Congo Rep., Costa Rica, Cote d'Ivoire, Dominican Republic, Ecuador, Egypt Arab Rep., Angola, El Salvador, Ethiopia, Gabon, Gambia The, Ghana, Guatemala, Guinea, Guinea-Bissau, Guyana, Haiti, Armenia, Honduras, India, Indonesia, Iraq, Jamaica, Jordan, Kazakhstan, Kenya, Kyrgyz Republic, Lao PDR, Azerbaijan, Lesotho, Liberia, Macedonia FYR, Madagascar, Malawi, Maldives, Mali, Mexico, Moldova, Mongolia, Bangladesh, Montenegro, Mozambique, Namibia, Nepal, Nicaragua, Niger, Nigeria, Pakistan, Panama, Paraguay, Belarus, Peru, Philippines, Rwanda, Sao Tome and Principe, Senegal, Serbia, Sierra Leone, Sri Lanka, Suriname, Swaziland, Belize, Syrian Arab Republic, Tajikistan, Tanzania, Thailand, Timor-Leste, Togo, Tunisia, Turkey, Uganda, Ukraine, Benin, Uzbekistan, Venezuela RB, Vietnam, West Bank and Gaza, Yemen Rep., Zambia, Zimbabwe.

Percent(%)

c.Male-Female