Homework 1.2 - Case 2

Statistical Data Analysis - 29th of October, 2019

Group 5

Antoñanzas Martínez, Guillermo
Pueblas Núñez, Rodrigo
Martínez Capella, Ignacio
Burrell, David

Introduction:

For this assignment, we have decided to choose two examples of different types of composite graphs and we have analyzed them and found their flaws.

First Composite Visualization:

Source:https://www.forbes.com/sites/ywang/2017/08/08/xiaomi-is-once-again-worlds-biggest-smartphone-brand-but-could-it-go-further/#2fa425431691

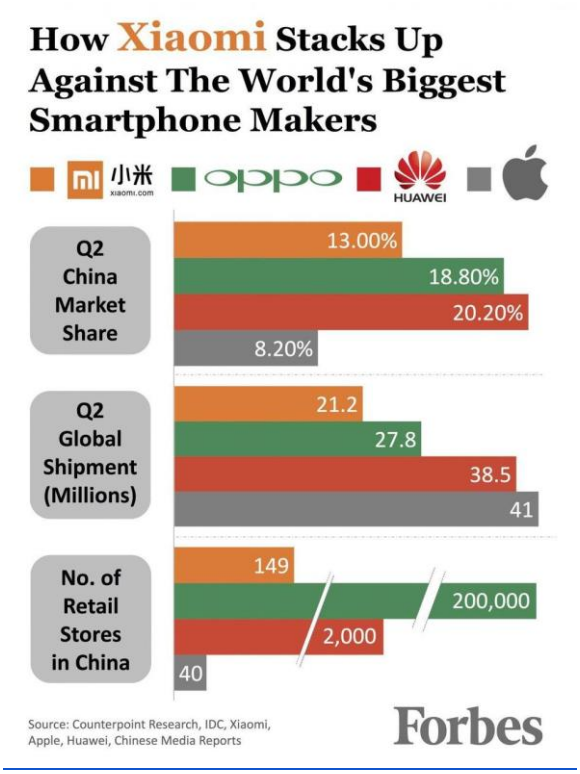

Figure 1: First composite graph to analyse

Figure 1 shows the first graph, which is used to describe how the telecom company Xiaomi is performing against its competitors in the smartphone market. Although at first glance this graph seems to be informative upon further inspection of the layout it can be seen there are many faults with the way the information is presented. Firstly, as the title includes the word “World’s” it is indicated that the statistics being shown are global. But in fact, two of the charts (the first and the third) are for China, Xiaomi’s domestic market, while the second is of the global market. By using the company’s domestic market, it can be argued that true comparison between Xiaomi and other companies is not being represented.

Issues can also be found in the layout of the legend. It requires the reader to know the company logos for the shown competitors. Although this is sufficient for the first three companies, as they include their names in the logo, the fourth (Apple) does not. Although it is fair to say that the Apple logo is globally recognised, it is a poor design decision to presume the reader knows what company is being shown here. Note that even in the original article there is no mention which companies are being used in the graph. An additional issue with the legend is it is not initially clear that it indicates each company's representative colour in the graphs. Due to the positioning of “oppo” as the largest logo in the centre the readers eyes are initially drawn there, but the green box looks like part of the logo. Combined with the Xiaomi logo looking similar to the legend indicator, an orange box, it takes the reader some time to deduce how the legend is comprised and related to the graphs.

There are also improvements that can be made to the graphs themselves. Firstly the “Q2 China Market Share” graph, although well laid out, is missing information. The values given total 60.2%, this leaves a percentage far larger than any other displayed. It should be indicated (not necessarily as a bar) what makes up the missing value. Presumably it is made up by lots of smaller companies, but this cannot be confirmed by the graph. The third bar graph, “No. of Retail Stores in China”, uses break on the bars to indicate a jump in the scale being used. But due to lack of X-scale it is not known how large the jumps are and may be confusing to a reader who is not aware of the concept. Although minor, this is another issue that may be confusing for a reader and lead to a delay in them understanding the information. The reason for this break usage is most likely to have all three graphs have the same limit. It would have been more appropriate to move this information to a different graph types such as a tree map as shown in Figure 2 (image not to scale).

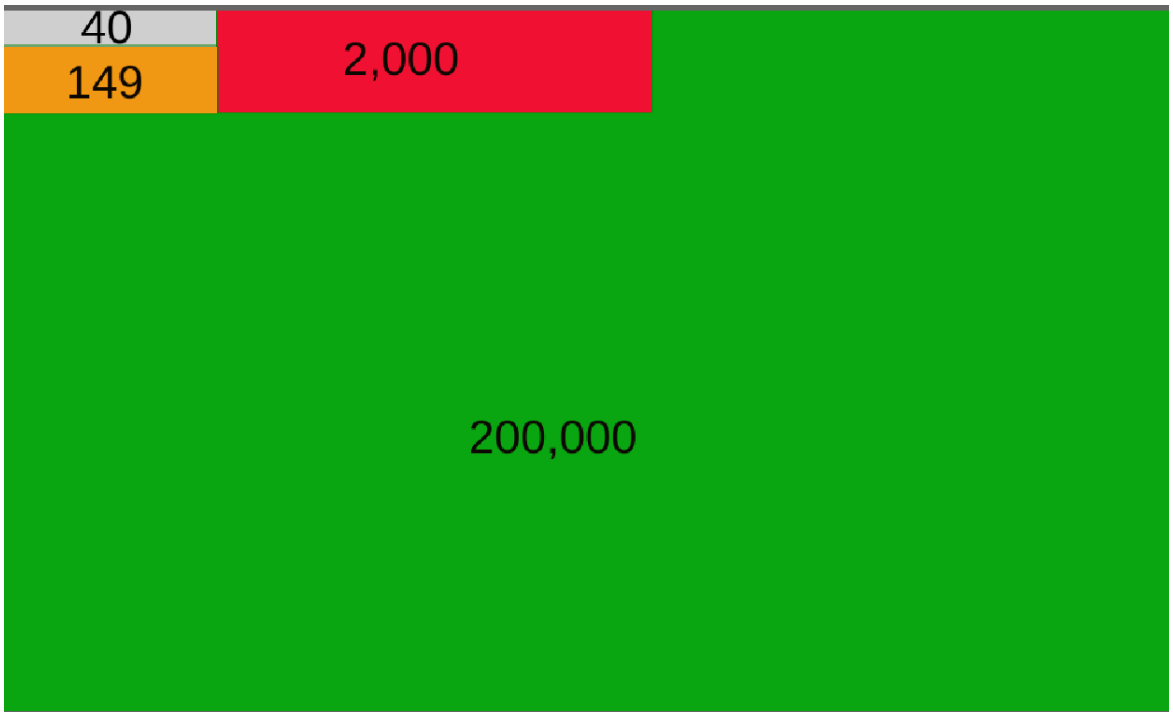

Figure 2: Graph better suited for the third barplot

This attempt though was probably to fit the 3 graphs to the same scale, allowing the image to fit better with the article web page.

Second Composite Visualization:

Source: http://www.yangon.climatemps.com/

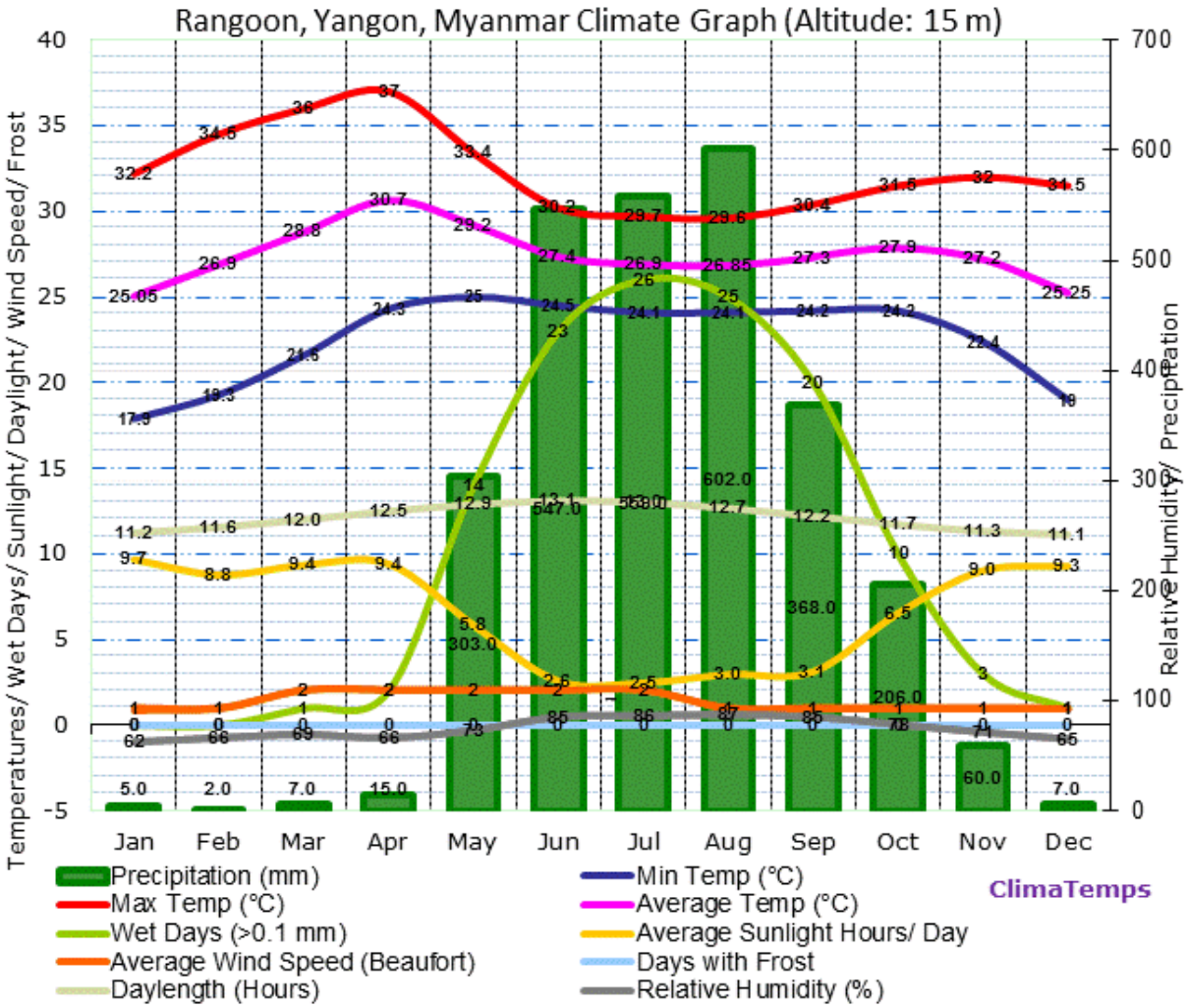

Figure 3: Second composite graph to analyse

Figure 3 displays climate information of the city of Rangoon (Myanmar). It displays the evolution of several variables across a whole year in monthly steps. The accuracy in the representation offered by the graph is good, since each variable is represented using an adequate figure for the information it holds. However, there are several problems with this composite graph.

1. The first and main one is that it is trying to display too much information at the same time. If we look at the legend, we can see that the graph shows ten variables, overloading the reader with information instead of providing a clear representation of the data. For example, climate graphs usually limit themselves to the precipitation and the temperature, like the one shown in the following image.

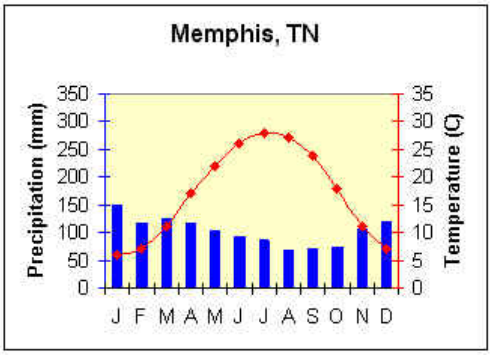

Figure 4: Example of simple climate graph

https://www.earthonlinemedia.com/ebooks/tpe_3e/images/atmosphere/climate/climographs/memphis.jpg

2. The second problem is that it only uses two different types of plots for the ten variables. The amount of precipitation is represented using a barplot, whereas the other nine variables are displayed using lines connecting the different values of the variable for each month. This makes it difficult to distinguish between these nine variables without constantly checking the legend, since they all are represented in the same way.

3. Also, the authors of the graph decided to add the numeric values to the figure. Although this is not a bad practice in itself, because of the high number of variables, the numbers are everywhere in the graph, making it more confusing. Moreover, in some cases, the numbers are placed one above another, with the information being illegible. And, speaking of illegibility, the legend of the right y-axis also overlaps with its numbers.

One way of correcting this composite graph is to split the information in different ones, in order to achieve a clear representation. The first one would be a standard climograph relating the amount of precipitation per month with its average temperature.

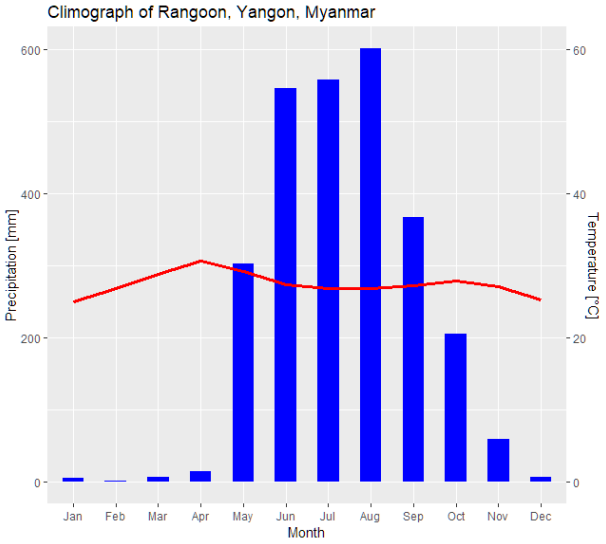

Figure 5: Temperature and precipitation by month

As for the rest of the variables, they can be grouped according to the information they offer. We have divided in:

- Temperature information: The variables Max Temp, Average Temp and Min Temp are displayed in a heatmap plot. They could have been represented using a line plot, but the heatmap offers an alternative representation that achieves similar results. We can see the evolution of each variable across time and we can compare the different variables among themselves, just as with a line plot. Since we are going to use line plots for the two remaining plots, we have decided that the heatmap would be a nice variation.

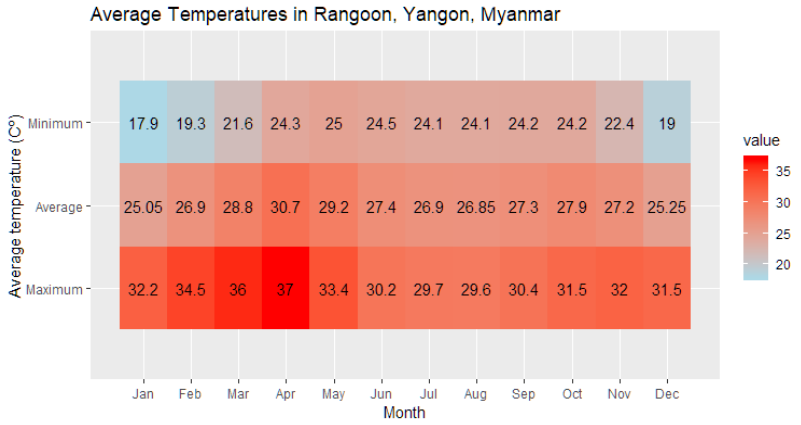

Figure 6: Minimum, maximum and average temperatures by month

- Precipitation information: The variables Precipitation, Relative humidity and Wet days are portrayed using a line plot.

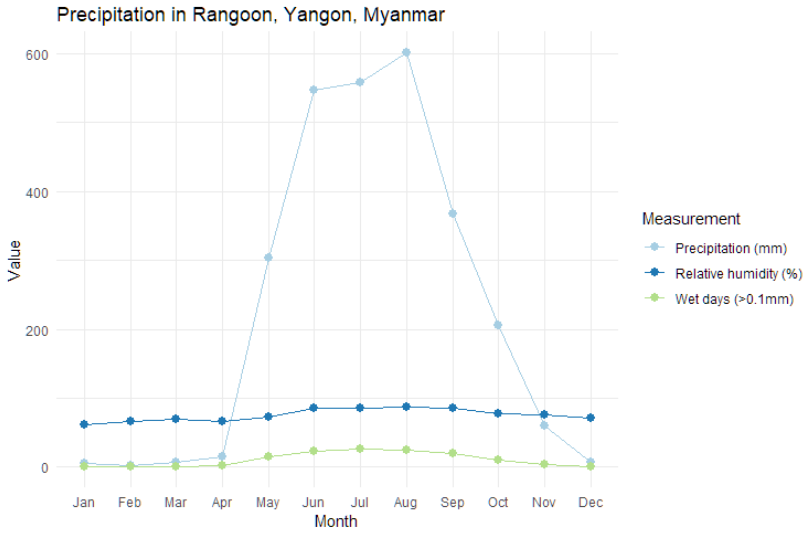

Figure 7: Precipitation, relative humidity and wet days by month

- Sun and wind information: The rest of the variables, which are concerned with the sun information, the days with frost and the wind, are also plotted together in a line plot.

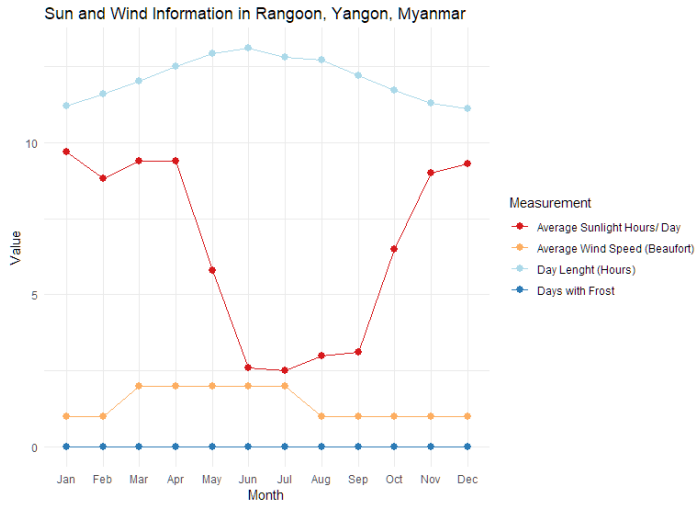

Figure 8: Average hours of different measures by month

The previous graphs can be presented individually or combined in a composite graph. If we compare the resulting composite graph with the original one, we can see that the information is clearer and more organized, which facilitates the understanding of the reader.

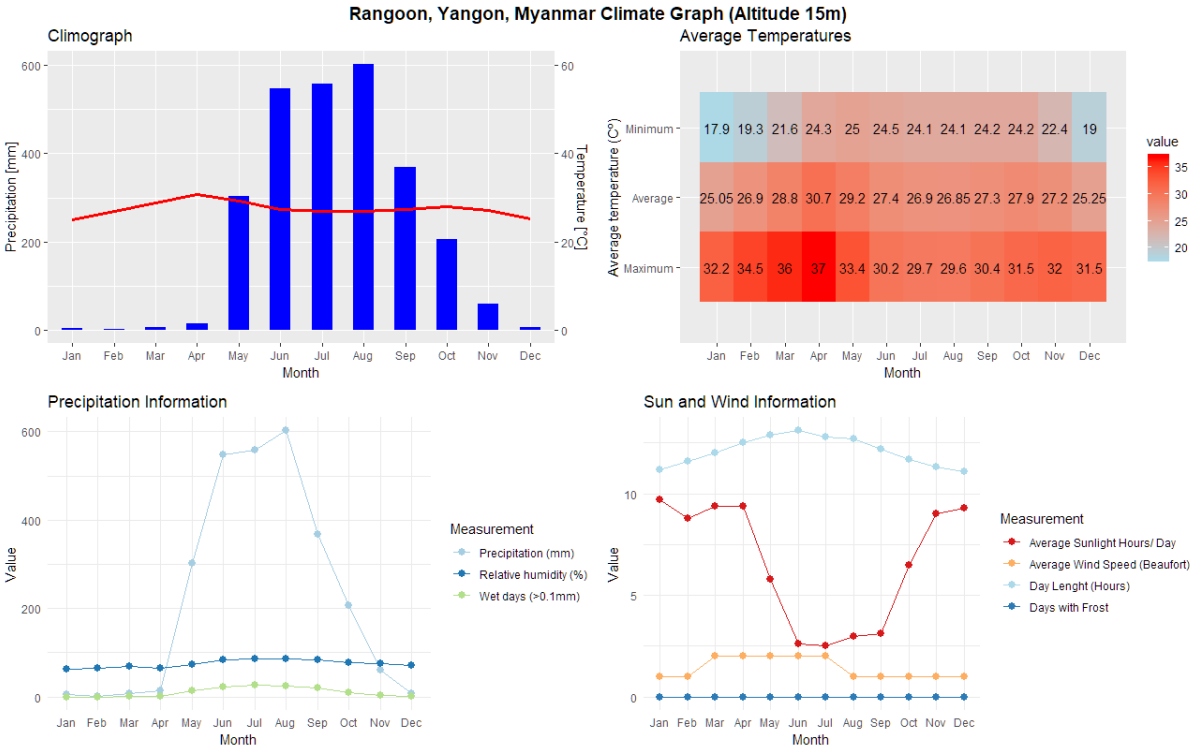

Figure 9: Second composite graph fixed

Appendix

Second visualization:

```
library(ggplot2)
library(RColorBrewer)
library(reshape2)
library(ggpubr)

#Climograph
climograph <- data.frame(
  Month=c("Jan",
"Feb","Mar","Apr","May","Jun","Jul","Aug","Sep","Oct","Nov","Dec"),
  precipitation=c(5,2,7,15,303,547,558,602,368,206,60,7),
  average_temp <-
c(25.05,26.9,28.8,30.7,29.2,27.4,26.9,26.85,27.3,27.9,27.2,25.25))

correct_order <- c("Jan","Feb","Mar","Apr","May","Jun",
"Jul","Aug","Sep","Oct","Nov","Dec")

plot_climograph <-  ggplot(data = climograph, mapping = aes(x = Month, y =
precipitation, group = 1)) +
  geom_bar(stat = "identity", fill = "blue", width = 0.5) +
  geom_line(mapping = aes(y = average_temp*10),
            color = "red",
            size = 1.2) +
  scale_y_continuous(
       "Precipitation [mm]",
       sec.axis = sec_axis ( ~ . *0.1, name = "Temperature [°C]")) +
  scale_x_discrete(limits = correct_order) +
  ggtitle("Climograph")
plot_climograph

#Temperature
temp_mat <-
matrix(c(32.2,34.5,36,37,33.4,30.2,29.7,29.6,30.4,31.5,32,31.5,25.05,26.9,28.8,30.7
,29.2,27.4,26.9,26.85,27.3,27.9,27.2,25.25,17.9,19.3,21.6,24.3,25,24.5,24.1,24.1,24
.2,24.2,22.4,19),
                    nrow = 12,ncol = 3)

mat.melted <- melt(temp_mat)
print(mat.melted)

plot_temp<-ggplot(mat.melted, aes(x = Var1, y = Var2, fill = value)) +
  geom_tile()+
  scale_fill_gradient(low = "lightblue", high = "red")+
  geom_text(aes(label = value))+
 scale_x_discrete(limit = c("Jan",
"Feb","Mar","Apr","May","Jun","Jul","Aug","Sep","Oct","Nov","Dec"))+
scale_y_discrete(limit = c("Maximum", "Average", "Minimum"))+
  labs(x = "Month", y = "Average temperature (Cº)", title = "Average Temperatures")
```

```
#Precipitation
  prec_df <- data.frame(
       month=c("Jan",
"Feb","Mar","Apr","May","Jun","Jul","Aug","Sep","Oct","Nov","Dec"),
       data=c(5,2,7,15,303,547,559,602,368,206,60,7,62,66,69,66,73,85,86,87,85,78,7
5,71,0,0,1,2,14,23,26,25,20,10,3,1),
       Measurement=rep(c("Precipitation (mm)","Relative humidity (%)","Wet days
(>0.1mm)"),each=12))

  plot_prec<-ggplot(prec_df, aes(x=month, y=data, group=Measurement)) +
       geom_line(aes(color=Measurement))+
       geom_point(aes(color=Measurement),size=2.5)+
       scale_color_brewer(palette="Paired")+
       theme_minimal()+
       labs(x = "Month", y = "Value", title = "Precipitation Information")+
       scale_x_discrete(limits = correct_order)

  #Other (Sun and Wind)
  other_df <- data.frame(
       month=c("Jan",
"Feb","Mar","Apr","May","Jun","Jul","Aug","Sep","Oct","Nov","Dec"),
       data =
c(9.7,8.8,9.4,9.4,5.8,2.6,2.5,3,3.1,6.5,9,9.3,11.2,11.6,12,12.5,12.9,13.1,12.8,12.7
,12.2,11.7,11.3,11.1,0,0,0,0,0,0,0,0,0,0,0,0,1,1,2,2,2,2,2,1,1,1,1,1),
       Measurement=rep(c("Average Sunlight Hours/ Day","Day Lenght (Hours)","Days
with Frost","Average Wind Speed (Beaufort)"),each=12))

  plot_other<-ggplot(other_df, aes(x=month, y=data, group=Measurement)) +
       geom_line(aes(color=Measurement))+
       geom_point(aes(color=Measurement),size=2.5)+
       scale_color_brewer(palette="RdYlBu")+
       theme_minimal()+
       labs(x = "Month", y = "Value", title = "Sun and Wind Information")+
       scale_x_discrete(limits = correct_order)

  plot_combined<-
ggarrange(plot_climograph,plot_temp,plot_prec,plot_other,nrow=2,ncol = 2)+
       labs(title = "Rangoon, Yangon, Myanmar Climate Graph (Altitude 15m)",x = "",
y="", fill="")+
       theme(plot.title = element_text(hjust = 0.5,face = "bold"))
  plot_combined
```