# Data Visualisation Assignment 1-18230229

Sai Krishna Lakshminarayanan

1 February 2019

## INTRODUCTION

Data Visualisation helps humans in understanding the inferences effectively. It is much easier to look into the visual representation of data than reading through several paragraphs to understand the end result. It is also important to provide this visualisation in an eye pleasing way such that it is easily consumable. In this assignment, this aspect is going to be dealt in a detailed manner based on human perception and subjective analysis.

## PART 1

The given data set name is week1 and it consists of 44 rows and 3 columns. The column names are x_value , y_value and Day which consist of 4 days name day 1, day 2 , day 3 and day 4. The library that is needed for performing the task is ggplot2 and it is loaded followed by the csv file.

```
library(ggplot2)#loading the package
wk1<-read.csv("week1.csv")#loading the dataset
head(wk1)#to show the initial values present
```

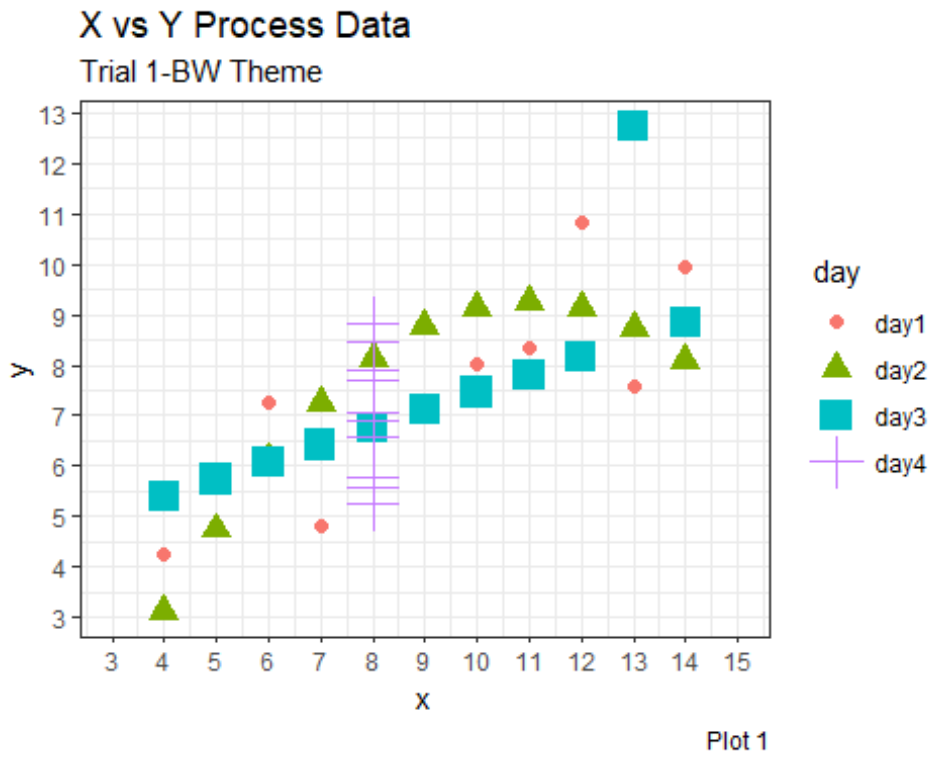
```
##   x_values y_values  day
## 1       10     8.04 day1
## 2        8     6.95 day1
## 3       13     7.58 day1
## 4        9     8.81 day1
## 5       11     8.33 day1
## 6       14     9.96 day1
```

## TRIAL 1

In trail 1, the predefined functions and their values are used directly in order to check whether any changes are needed to be done.

```
plot1<-ggplot(wk1, aes(x=x_values, y=y_values,col=day)) +#providing the data
for ggplot and giving x,y and colour values
  geom_point(aes(size=day,fill=day,shape=day))+ #giving the size,fill and
shape values for the point
  coord_cartesian(xlim=c(3,15))+#giving the coordinate limit
labs(title="X vs Y Process Data", subtitle="Trial 1-BW Theme", y="y", x="x",
caption="Plot 1")#providing the title and description for the plot
plot1 <-plot1 +
  scale_x_continuous(breaks=seq(2, 15, 1)) + #giving the breakdown for the x
```

```
and y axis
  scale_y_continuous(breaks=seq(3, 13, 1)) +
  theme_bw()
plot(plot1)#output for first plot

## Warning: Using size for a discrete variable is not advised.
```




**Results**

1) As the colours are pre-defined , it is seen that day 4 colour seems to be light and merging onto the bw theme lines. This means that it is difficult to distinguish the points as a human. So, it is better to remove the bw theme and keep the conventional classic theme as it wont have extra lines to see giving the visual separation of days a better chance.

2) Alot of overlapping occurs in the place where x=8 and some of it isn't clearly visible to the eyes. So, it means that shape and size should be manually changed inorder to support it and fill and stroke can also be tried to increase the intensity of the borders. This will ensure for quickly grasping the different points as it will be significantly different for each day.

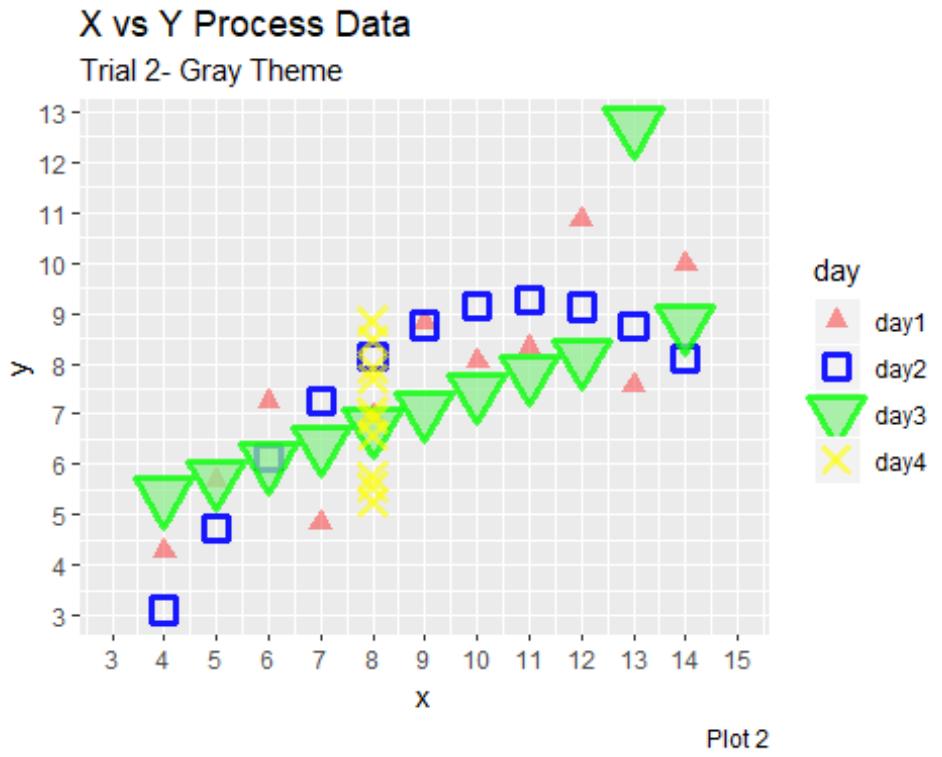
3) The 4 shapes are visually separable and clearly distinguishable. So it can be kept as it is or can be enhanced by adding borders and changing alpha values to support the better visual separation of the 4 days.

4) The x limits and the y limits seem to be in proper way as it removes the extra space which will be unused. It helps in focusing onto the required part directly and so the coordinates can be kept as it is for now.

## TRIAL 2

In this, let's consider fill and stroke to create some visual separation and improve from the faults present in trial 1. This includes providing manual size,colour and shape values as needed and using the classic theme.

```
plot2 <- ggplot(wk1, aes(x=x_values, y=y_values,shape=day)) +
 geom_point(aes(col=day, size=day,alpha=day),stroke=2,fill="lightgreen") +
#giving stroke values
    coord_cartesian(xlim=c(3,15))  +
  labs(title="X vs Y Process Data", subtitle="Trial 2- Gray Theme", y="y",
x="x", caption="Plot 2")+

  scale_shape_manual(values=c(17, 0, 25, 4)) + # giving the shapes manually

  scale_color_manual(values=c("red", "blue", "green","yellow")) + #giving the
colours manually

  scale_size_manual(values=c(2,3,6,3))+ # giving the size manually

  scale_alpha_manual(values = c(0.4,0.9,0.75,0.6)) #providing the alpha
values manually
plot2 <-plot2 +
  scale_x_continuous(breaks=seq(2, 15, 1)) +
  scale_y_continuous(breaks=seq(3, 13, 1)) +
  theme_gray()
plot(plot2)#output for second plot
```




**Results**

1) The colours seem to be inseparable especially in the clustered part in x=8. This is happening inspite of changing the values in alpha and size manually.So, it visually suggests day 4. This means that it is extremely difficult to differentiate between the days and it will consume alot of time.

2) Day 4 seems to require another shape as it appears to be inseparable when present close to each other and also requires a lighter colour as it seems to be too contrasting while viewing making it difficult to focus.

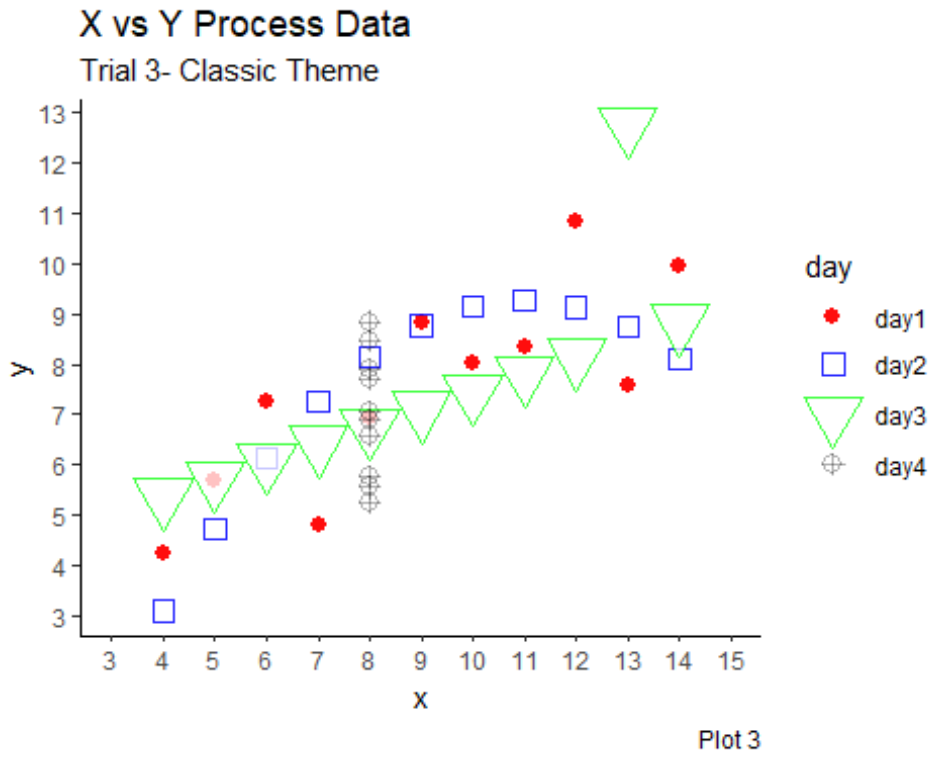
3) The fill effect provided in the day 3 has caused a scenario where the values present in it aren't visible clearly. Therefore, it is better to apply white fill so that there is transparency and helps in viewing the points inside in a better way.

## TRIAL 3

From the issues faced in trial 1 and trial 2, steps are to be taken in trial 3 inorder to bring better separation that is possible. This is possible by altering the sizes to ensure that overlap is minimum and changing alpha and colour values suitably to ensure transparency while visually seeing them.

```
plot3 <- ggplot(wk1, aes(x=x_values, y=y_values,shape=day)) +
   geom_point(aes(col=day, size=day,alpha=day),stroke=1,fill="white") +
    coord_cartesian(xlim=c(3,15))  +
  labs(title="X vs Y Process Data", subtitle="Trial 3- Classic Theme", y="y",
```

```
x="x", caption="Plot 3")+

  scale_shape_manual(values=c(16, 0, 25, 10)) +

  scale_color_manual(values=c("red", "blue", "green","black")) +

  scale_size_manual(values=c(2,3,7,3))+

  scale_alpha_manual(values=c(0.95,0.85,0.75,0.3))



plot3 <-plot3 +
  scale_x_continuous(breaks=seq(2, 15, 1)) +
  scale_y_continuous(breaks=seq(3, 13, 1))+
  theme_classic()
plot(plot3)
```




**Results**

1) In this, the colour combination appears to be different from each other which can be easily separated through first glimpse. Therefore this colour combination of red , blue, green and black for the days is kept.

2) The size is given to the days in such a way that day 3 is large and the ones which will be overlapping are small so that it could not be too confusing to pick them up. It proves to be supporting as it is easy to visually seaparate the 4 days.

3) The shapes are assigned in such a way that it allows transparency to ensure overlapping points are visible. Also,day 4 is taken as that particular shape of a circle with a cross inside so that it helps in locating the points even when they are clustered like in the case of x=8. This has helped in seeing how much points of day 3 and 4 are present successfully without much strain.

4) Alpha value for day 2 and day 3 are given high and for day 1 and 2 it is given low as they are present alot together. Only when the alpha value is given low for them, it helps in viewing them clearly when present together. Also, Fill value is given as white as only then it helps in maintaining the trasparency in viewing the points . The stroke value is kept at 1 so that it doesn't appear to be too thick to cause distraction while viewing. In these ways, the 4 days points are made visually separable as much as possible.

## PART 2

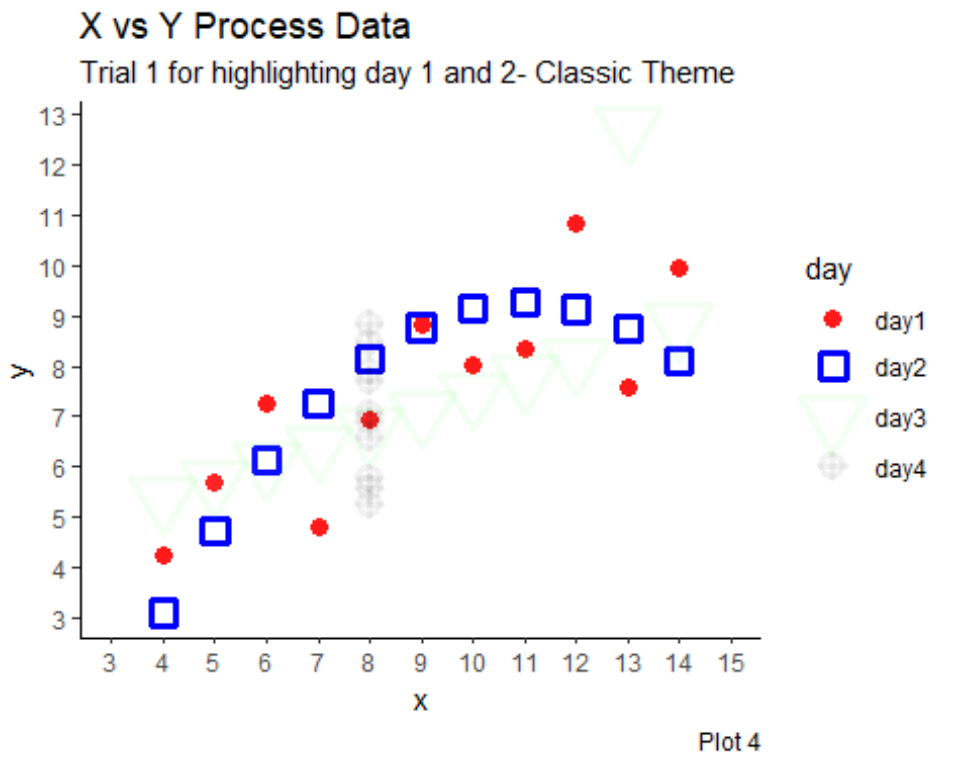
In this, the task is to highlight the points of day 1 and 2 in the plot. In order to achieve this, let's make use of alpha,size,fill and stroke values. When this is done manually, it makes sure that the required points are projected with more importance than the other ones.

## TRIAL 1

In this attempt, the alpha values are manually altered and fill is given as white and stroke value is given as 2 in order to ensure that day 1 and 2 points are highlighted.

```
plot4 <- ggplot(wk1, aes(x=x_values, y=y_values,shape=day)) +
   geom_point(aes(col=day, size=day,alpha=day),stroke=2,fill="white") +
    coord_cartesian(xlim=c(3,15))  +
  labs(title="X vs Y Process Data", subtitle="Trial 1 for highlighting day 1
and 2- Classic Theme", y="y", x="x", caption="Plot 4")+

  scale_shape_manual(values=c(16, 0, 25, 10)) +

  scale_color_manual(values=c("red", "blue", "green","black")) +

  scale_size_manual(values=c(2,3,7,3))+

  scale_alpha_manual(values=c(0.9,1,0.05,0.05))



plot4 <-plot4 +
  scale_x_continuous(breaks=seq(2, 15, 1)) +
  scale_y_continuous(breaks=seq(3, 13, 1))+
```

```
  theme_classic()
plot(plot4)
```




**Results**

1) From the initial look, the points of day 1 and day 2 are certainly highlighted when compared to day 3 and day 4.

2) This is done by lowering the values of alpha values of day 3 and day 4 and keeping the fill as white.

3) This ensures that the viewer experiences a plot in which day 1 and day 2 points are present brightly and easy to calculate when compared to the day 3 and 4 points. They are also present but are deliberately made illegible to highlight the days 1 and 2.

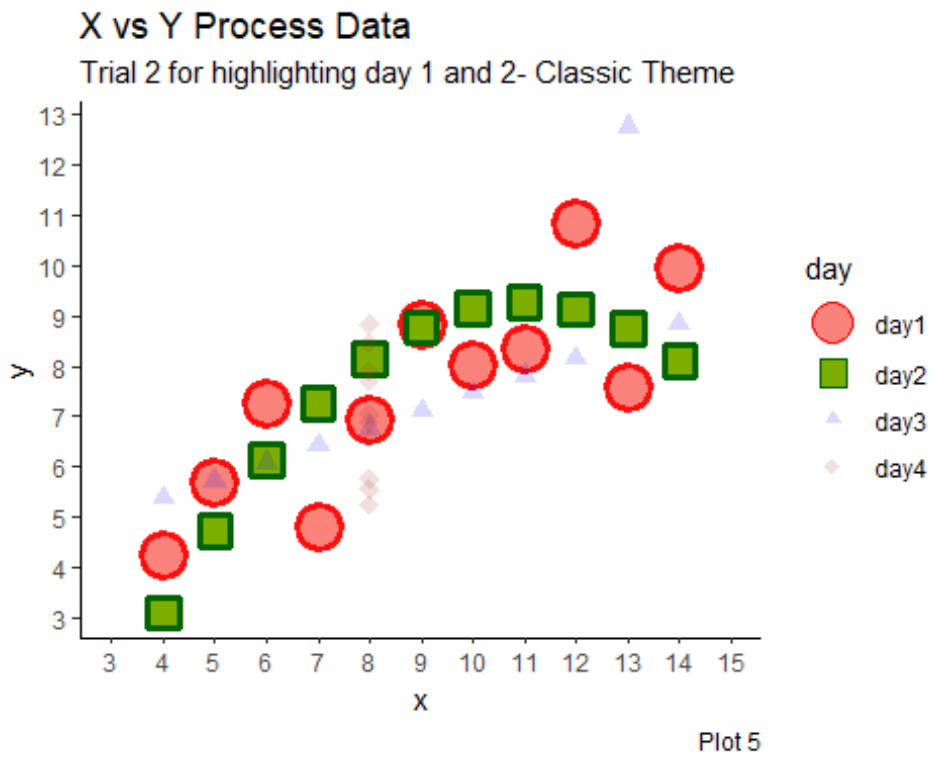
4) The issue in this plot is that with this shape and colour it is difficult to count the day 3 and 4 if needed. So, it is neccessary to make them countable and also ensure that day 1 and 2 are highlighted. This will create the best overall visual experience with respect to the plot.

## TRIAL 2

In this second trial, the issues with first trial are rectified. This is done by means of altering the values of alpha,size,colour,stroke,shape and fill functions for the given data.

```
plot5 <- ggplot(wk1, aes(x=x_values, y=y_values,shape=day)) +
   geom_point(aes(col=day, size=day,alpha=day,fill=day,stroke=1.75 ))+
```

```
    coord_cartesian(xlim=c(3,15))  +
  labs(title="X vs Y Process Data", subtitle="Trial 2 for highlighting day 1
and 2- Classic Theme", y="y", x="x", caption="Plot 5")+

  scale_shape_manual(values=c(21,22, 17, 18)) +

  scale_color_manual(values=c("red", "darkgreen", "blue","brown")) +

  scale_size_manual(values=c(7,5,2,2.5))+

  scale_alpha_manual(values=c(0.9,1,0.15,0.15 ))



plot5 <-plot5 +
  scale_x_continuous(breaks=seq(2, 15, 1)) +
  scale_y_continuous(breaks=seq(3, 13, 1))+
  theme_classic()
plot(plot5)
```




**Results**

1) In this, the points of day 1 and 2 are highlighted and also the points of day 3 and 4 are not completely illegible and can be counted if needed.This is done by means of changing the shapes to accept fill values and providing bright colours. This means that it will catch the immediate attention of the person viewing it.

2) The strokes are given are given a comparatively large value and also the alpha values are given a slight high value. This helped in providing some visibility for day 3 and day 4 while also maintaining to highlight the points 1 and 2.

3) The size is also increased alot for day 1 and 2. This helped in highlighting both of them even while overlapping like in the case where x=9. This is because it is visually easily to pick and differentiate the values with large size and different colour and shape. Therefore, in this way it is seen that trial 2 plot is much more effective.

## CONCLUSION

Through the two tasks, the various ways in which ggplot package can be used is seen . The usage and benefits of various functions in the ggplot like fill,shape,size,colour and alpha are also realised in a practical manner successfully. Through the help of these, plot in which the points of the days are visually seaparable and the plot in which the points of day 1 and 2 are highlighted is generated successfully.