**Name: Prashant Prakash (pxp141730)**

**MiniProject#2**

**Overview:**

2.a:

In this three plots are being drawn using map object . Map object is joined with data object based on state key. The state column is converted to lowercase as mentioned in the guidelines. For assigning the states different colors based on their values the breaks are decided seeing the min, max and mean values.

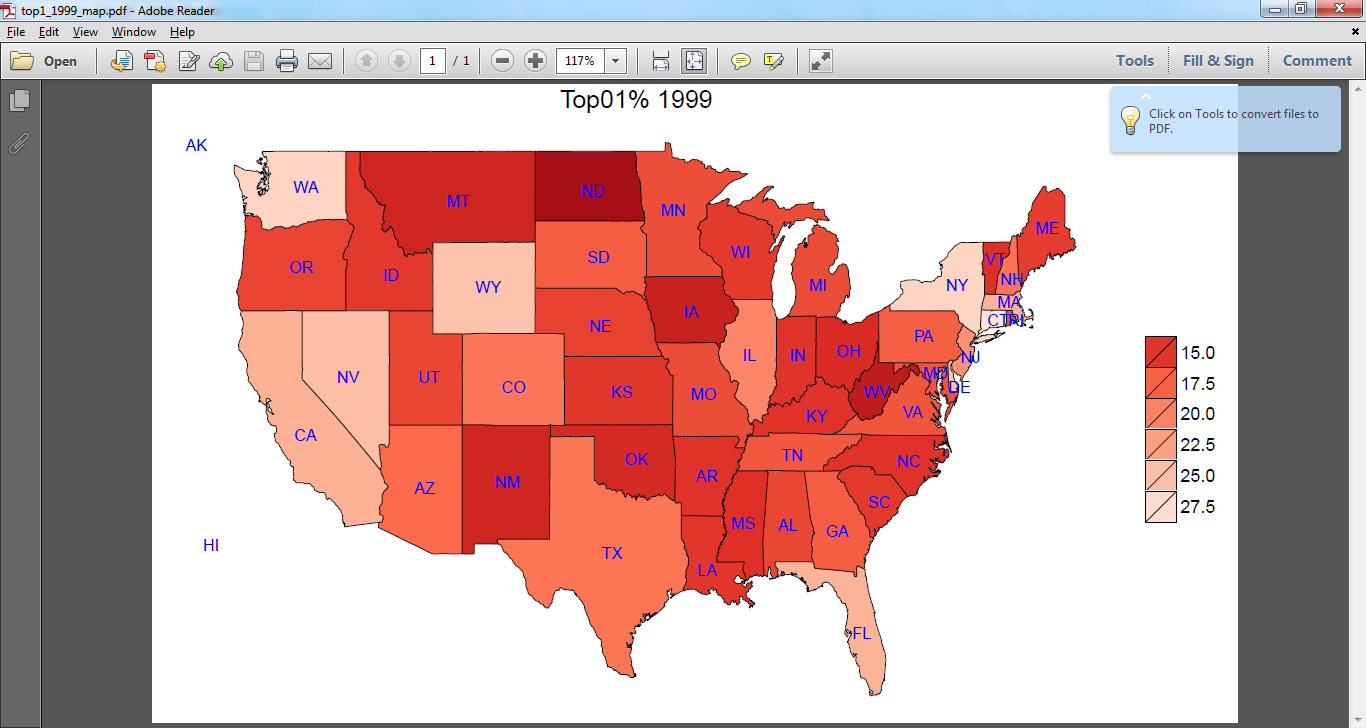
2.b:

Various plots like histogram and Box plot is drawn to understand the distribution. Also for understanding of Happy Planet index values on three different variables scatter plot is drawn to understand the correlation.

**Experiment Results and Observation:**

**2.a**

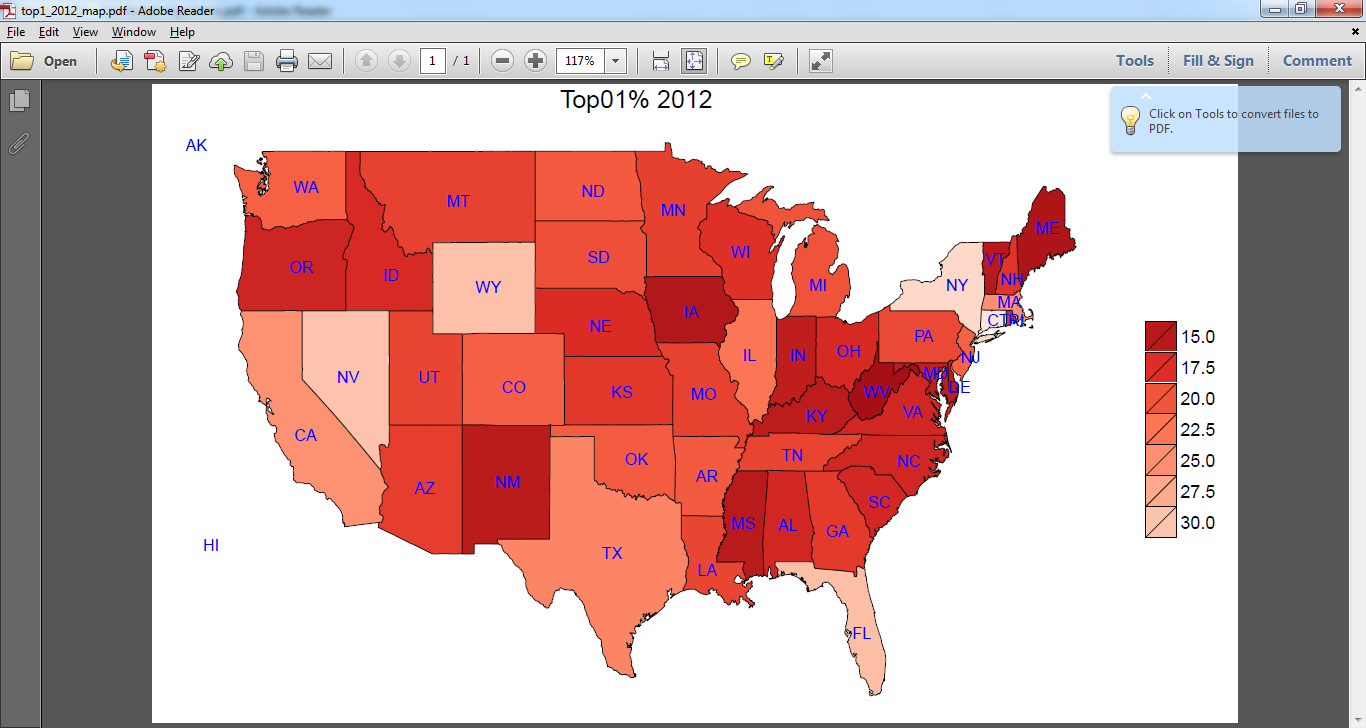
For year 1999 top 1% Income Share



Observation:

The best is among WA, WY, NV, NY. Most of the states have value between 15 and 17.5. So the mean must lie between 15 and 17.5. After calculation we get mean as 17.49. The lowest is among NM, ND, OK.

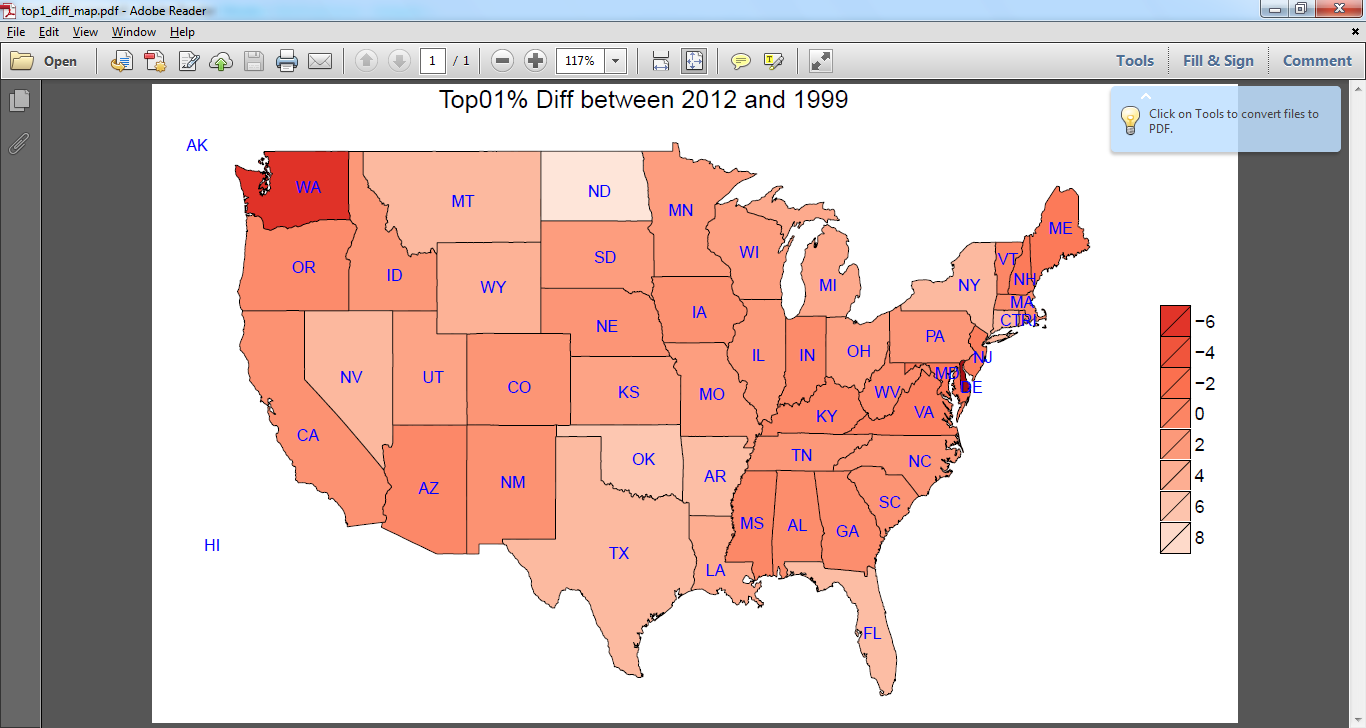
For year 2012 Top 1% Income share



Observation:

In this case maximum is among WY, NV,NY and WA is not among in best as compared to 1999, and in this case most of the states have value between 17.5 and 20 so mean is between 17.5 and 20. After mean calculation we get value as 19.39. So the mean has been increased as compared to year 1999.

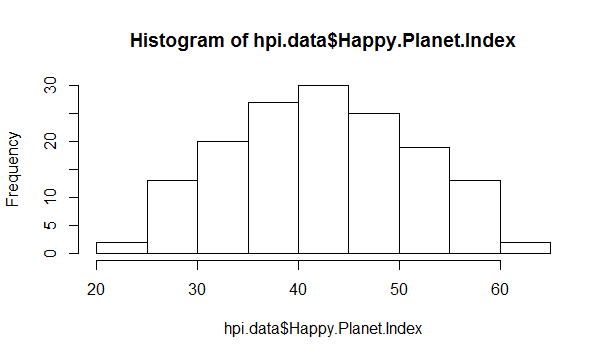
For diff Top 1% (2012 -1999)



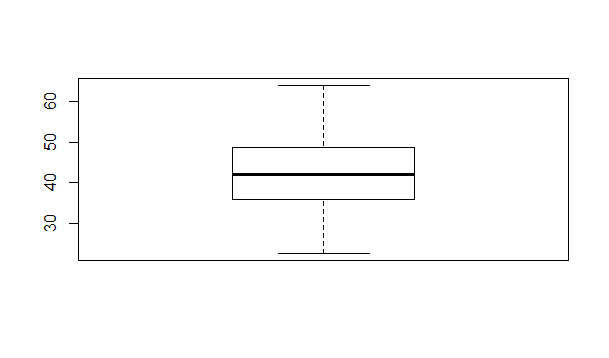
Observation: clearly the state WA has a significant reduction in Income share from 1999 and 2012. States like NV and ND has a significant increase in Income share from 1999 and 2012 which is close to maximum increase in Income share between 1999 and 2012. The mean increase from year 1999 to 2012 is :1.90 which is somewhat obvious from the map as most of the values are between 0 and 2.

**2.b**

Histogram for Happy planet Index:

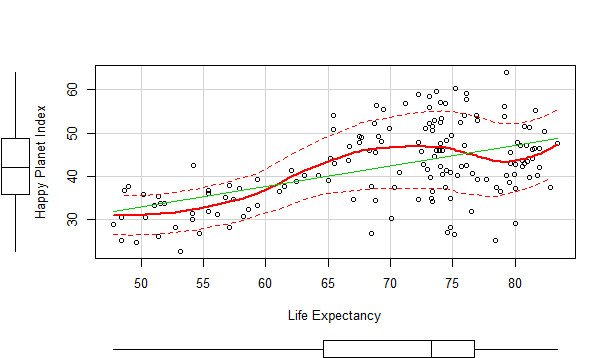


Box Plot for happy Planet Index:

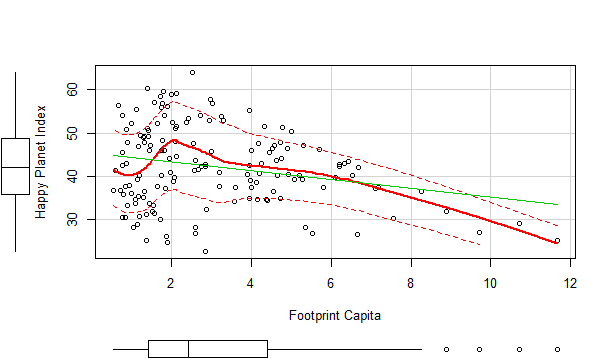


Observation: Clearly from Histogram the distribution is a Normal distribution and for Normal Distribution mean and standard deviation are the right parameters to represent.

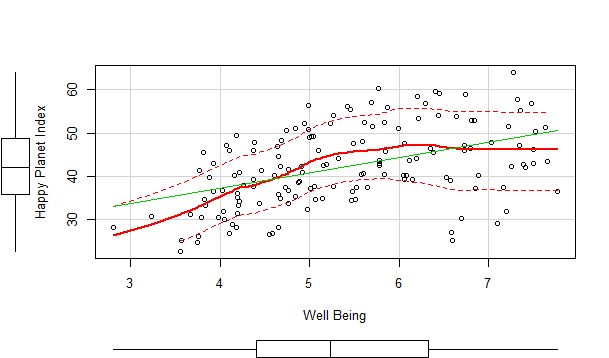
Drawing scatter plot of HPI against Life expectancy we get plot as given below:



HPI against Footprint Capita



HPI against Well Being



Observation:

Clearly from the scatter plot it is obvious that:

1. HPI increases as with the increase in Life expectancy.
2. HPI decreases as with the increase in FootPrint Capita.
3. HPI increase as with the increase in wellbeing.

Now seeing this scatter plot we can easily infer that there is relation of HPI with all the three variables. To understand the extent of relation we can understand the correlation of HPI with these 3 variables.

Correlation of HPI with Footprint Capita: -0.2382588

Correlation of HPI with wellbeing: 0.4530088

Correlation of HPI with Life Expectancy: 0.5109242

From results of correlation it is clear that HPI is highly related to life expectancy.

**Programs:**

**2.a**

# including all libraries

library(gdata) # to read from xls file

library(raster) # to get map shape file

library(ggplot2) # for plotting and miscellaneuous things

library(ggmap) # for plotting

library(plyr) # for merging datasets

library(scales) # to get nice looking legends

library(maps) # for plotting map

# getting usa shape object to plot the map , parameter USA is passed to get details about USA

usa.shape <- getData("GADM", country = "USA", level = 1)

# building data frame for all states

usa.df <- map\_data("state")

# assigning the column name as state which is same as from xls file loaded in next step

colnames(usa.df) [5] <- "state"

# loading data from the file , note that the column state has been changed to lowercase to join from map

usa.data <- read.xls("C:\\Users\\PRASHANT\\Downloads\\usstatesWTID.xls",header=TRUE);

# filtering only three columns Year,State and Top1\_adj which is required to build map

usa.filtercol.data <- usa.data[,c("Year","state","Top1\_adj")];

# filtering data from filtered set only for year 1999

usa.1999.data <- subset(usa.filtercol.data, Year == 1999)

# filtering data from filtered set only for year 2012

usa.2012.data <- subset(usa.filtercol.data, Year == 2012)

# merging data set obtained for year 1999 and 2012 on key state so that all the three plots can be handled from one data frame only

mergeddataset=merge(usa.1999.data,usa.2012.data,by="state")

mergeddataset$diff <- 0.00

mergeddataset$diff <- mergeddataset$Top1\_adj.y - mergeddataset$Top1\_adj.x

# joining the data set obtained in previous set with the map dataset based on state

usa.df <- join(usa.df, mergeddataset, by = "state", type = "inner")

# getting all statenames help to annotate the map

state.names <- data.frame(id = 1:length(usa.shape$NAME\_1), state = usa.shape$NAME\_1)

statenames<-data.frame(state.center,state.abb,tolower(state.names))

# these breaks are decided based on seeing the min , mean , max for the columns to plot

# for 1999 , min : 10.74 , max : 28.152 mean: 17.49

# for 2012 , min : 12.50 , max : 33.00 mean: 19.39

# for diff , min : -10.16 , max : 9.02 mean: 1.90

brks <- c(15, 17.5, 20, 22.5, 25, 27.5,30)

brksdiff <-c(-6, - 4 , -2 , 0 , 2 , 4 , 6, 8)

# plot for year 1999

p <- ggplot() +

# with borders (slower)

geom\_polygon(data = usa.df, aes(x = long, y = lat, group = group, fill = Top1\_adj.x),

color = "black", size = 0.15) +

scale\_fill\_distiller(palette = "Reds", breaks = brks, trans = "reverse") +

theme\_nothing(legend = TRUE) +

labs(title = "Top01% 1999", fill = "") +

geom\_text(data=statenames,aes(x,y,label = state.abb),color="blue",size=3)

# saving the plot as pdf file

ggsave(p, file = "top1\_1999\_map.pdf")

# plot for year 2012

q <- ggplot() +

# with borders (slower)

geom\_polygon(data = usa.df, aes(x = long, y = lat, group = group, fill = Top1\_adj.y),

color = "black", size = 0.15) +

scale\_fill\_distiller(palette = "Reds", breaks = brks, trans = "reverse") +

theme\_nothing(legend = TRUE) +

labs(title = "Top01% 2012", fill = "") +

geom\_text(data=statenames,aes(x,y,label = state.abb),color="blue",size=3)

# save plot as pdf file

ggsave(q, file = "top1\_2012\_map.pdf")

# plot for diff : 2012 - 1999

r <- ggplot() +

# with borders (slower)

geom\_polygon(data = usa.df, aes(x = long, y = lat, group = group, fill = diff),

color = "black", size = 0.15) +

scale\_fill\_distiller(palette = "Reds", breaks = brksdiff, trans = "reverse") +

theme\_nothing(legend = TRUE) +

labs(title = "Top01% Diff between 2012 and 1999", fill = "") +

geom\_text(data=statenames,aes(x,y,label = state.abb),color="blue",size=3)

# save plot as pdf file

ggsave(r, file = "top1\_diff\_map.pdf")

**2.b**

# get all the library required

library(xlsx) # for reading from xlsx file

library(car) # for using scatterplot

# get hpi data from the Rank Order Sheet

hpi.data <- read.xlsx("C:\\Users\\PRASHANT\\Downloads\\hpi-data.xlsx", sheetName ="Rank Order" , colIndex= 2:10, rowIndex =7:158)

# Draw Histogram for Happy Planet Index to understand the distribution

hist(hpi.data$Happy.Planet.Index)

# Box Plot for happy Planet Index

boxplot(hpi.data$Happy.Planet.Index)

hpi.data.lifeexp <- hpi.data[,c("Life..Expectancy","Happy.Planet.Index")]

scatterplot(hpi.data.lifeexp$Life..Expectancy,hpi.data.lifeexp$Happy.Planet.Index, xlab="Life Expectancy" , ylab="Happy Planet Index" )

hpi.data.footprint <-hpi.data[,c("Footprint..gha.capita.","Happy.Planet.Index")]

scatterplot(hpi.data.footprint$Footprint..gha.capita.,hpi.data.footprint$Happy.Planet.Index, xlab="Footprint Capita" , ylab="Happy Planet Index" )

hpi.data.wellbeing<- hpi.data[,c("Well.being..0.10.","Happy.Planet.Index")]

scatterplot(hpi.data.wellbeing$Well.being..0.10.,hpi.data.wellbeing$Happy.Planet.Index, xlab="Well Being" , ylab="Happy Planet Index" )

# capturing correlation

cor(hpi.data$Footprint..gha.capita.,hpi.data$Happy.Planet.Index ,use ="everything", method = c("pearson"))

cor(hpi.data$Well.being..0.10.,hpi.data$Happy.Planet.Index ,use ="everything", method = c("pearson"))

cor(hpi.data$Life..Expectancy,hpi.data$Happy.Planet.Index ,use ="everything", method = c("pearson"))