Lab3

Group 22 - tejma768, balra340

September 11, 2018

Assignment1

Question 1

```r
library(plotly)

Sys.setenv('MAPBOX_TOKEN' = 'pk.eyJ1IjoidGVqYXNocmVlcm0iLCJhIjoiY2pta21ud212MHBxdTNsbmoxeGp1OGp4ZCJ9.mn_aZzsqDMNS9cXtkAhfIw')

df1 <- read.csv("aegypti_albopictus.csv")
df1$hover <- with(df1, paste(VECTOR, '<br>', "Year", YEAR))


df2004 <- subset(df1, YEAR == 2004)
p <- plot_mapbox(df2004, lat = ~Y, lon = ~X, color = ~factor(VECTOR), mode = 'scattermapbox')
p
```

- Aedes aegypti
- Aedes albopictus

```r
df2013 <- subset(df1, YEAR == 2013)
p1 <- plot_mapbox(df2013, lat = ~Y, lon = ~X, color = ~factor(VECTOR), mode = 'scattermapbox')
p1
```

- Aedes aegypti
- Aedes albopictus

There seems to be preattentive perception problem as a person faces difficulty in interpreting the data at first glance. At first glance we noticed that in the year 2004, there was higher occurence of Aedes Albopictus in southern parts of USA and lesser occurence of the same in Spain, Italy and Croatia. And in Taiwan at first when we see, we feel the population of mosquitos there is not that high, but only upon zooming we get to know that the occurence of Aedes Albopictus is more in Taiwan. We can see clear differences on the basis of mosquito population between years 2004 and 2013. In the year 2013, higher occurences of Aedes Aegypti was found to be in most parts of Brazil, and a lower occurence in Surabaya and Semarang of South Asia.

Mosquitos population seems to have reduced between the two time points of the mosquito type Aedes Albopictus, and the population of Aedes aegypti has increased over time specially in parts of Brazil.

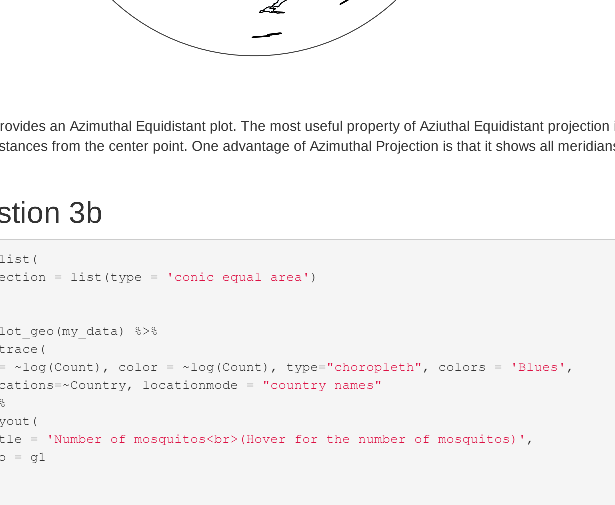
Question 2

```r
library(plotly)
library(dplyr)
library(ggplot2)

Z <- df1 %>% count(df1$COUNTRY)

my_data <- as.data.frame(Z)

colnames(my_data) <- c('Country', 'Count')

g <- list(
  projection = list(type = 'Equirectangular')
)
p <- plot_geo(my_data) %>%
  add_trace(
    z = ~Count ,color=~Count, type="choropleth", colors = 'Blues',
    locations = ~Country,locationmode ='country names'
  ) %>%
  layout(title = 'Number of Mosquitos per country detected
         during all study period ',
      geo = g
  )
p
```

There seems to be less information that we can depict from the plot as we are plotting the Number of mosquitos per country detected during all the study period, but we are not plotting the population of types of mosquito separately.

Question 3a

```r
Z <- df1 %>% count(df1$COUNTRY)

my_data <- as.data.frame(Z)

colnames(my_data) <- c('Country', 'Count')

g1 <- list(
  projection = list(type = 'azimuthal equidistant')
)

p <- plot_geo(my_data) %>%
  add_trace(
    z = ~log(Count), color = ~log(Count), type="choropleth",colors = 'Blues',
     locations= ~Country ,locationmode = "country names"
  ) %>%
  layout(
    title = 'Number of mosquitos<br>(Hover for the number of mosquitos)',
    geo = g1
  )
p
```



Map 3a provides an Azimuthal Equidistant plot. The most useful property of Aziuthal Equidistant projection is that all points are proportionally at correct distances from the center point. One advantage of Azimuthal Projection is that it shows all meridians with correct distances from the pole.

Question 3b

```r
g1 <- list(
  projection = list(type = 'conic equal area')
)

p <- plot_geo(my_data) %>%
  add_trace(
    z = ~log(Count), color = ~log(Count), type="choropleth", colors = 'Blues',
    locations=~Country, locationmode = "country names"
  ) %>%
    layout(
    title = 'Number of mosquitos<br>(Hover for the number of mosquitos)',
    geo = g1
  )
p
```

Map 3b provides a Conic Equal Area plot. Conic projections are suited best for mappig regional areas.The main disadvantage of Conic Projections is that the shapes will not be preserved.

When the plot is zoomed to center, the plots are visible in a much better way, we can see data with respect to countries which is advantageous. But its difficult to find the relative population of mosquitoes at first glance which seems to be disadvantageous.

Question 4

```r
Brazil <- filter(df2013 , COUNTRY == "Brazil")

#Question 4a
Brazil$X1 <- cut_interval(Brazil$X, 100)

#Question 4b
Brazil$Y1 <- cut_interval(Brazil$Y, 100)

#Question 4c
Calculate_mean <- Brazil %>%
    group_by(X1, Y1) %>%
    summarise(mean_X = mean(X), mean_Y = mean(Y), Observations = n())

#Question 4d
plot_Brazil <- plot_mapbox(data = Calculate_mean, x = ~mean_X, y = ~mean_Y, color = ~log(Observations),
                          mode ="scattermapbox")
plot_Brazil
```

By looking at the plot in Question 1, we see that population of mosquitos is more in Brazil. But as we look at the plot in question 4 it gets clear that the regions Guarabira, Caicara and Soa Paulo that are covered with Yellow colour depicts that the respective area is very much affected by mosquitos next least affected areas are covered by Green colour and they are, Ribeirao Preto, Lavras, Aracatuba, Valadares and most part of the regions that are covered in blue color like Brasilia, Corumba and Ji Parana are having traces of mosquitos.

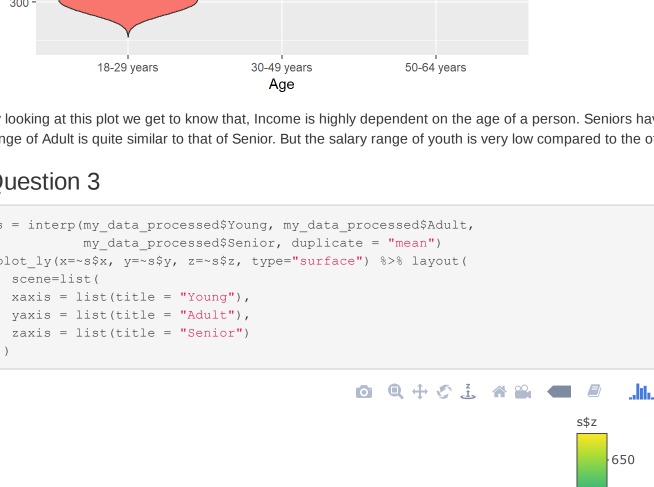
Assignment 2

Question 1

```r
my_data = read.csv("000000KD.csv")
rds = readRDS('gadm36_SWE_1_sf.rds')
my_data_processed = data.frame(region = unique(my_data$region))

my_data_split = split(my_data, my_data$age)
for (i in seq_along(my_data_split)) {
  my_data_processed[[names(my_data_split)[i]]] = merge(my_data_split[[i]],
                                                       my_data_processed$region,
                                                       by.x = 'region',
                                                       by.y = 1, all = T)$X2016
}

colnames(my_data_processed) = c("region", "Young", "Adult", "Senior")
my_data_processed$region = gsub(" county", "", my_data_processed$region)
my_data_processed$region = gsub("\\d{2} ", "", my_data_processed$region)
my_data_processed$region = gsub("Örebro", "Orebro", my_data_processed$region)
rownames(my_data_processed) = my_data_processed$region
```

Question 2

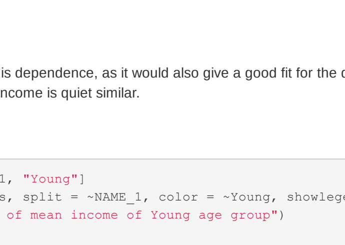
```r
p3 <- ggplot(my_data, aes(y=X2016, x=age, fill = age)) +
      xlab("Age") + ylab("Income") + geom_violin(trim = FALSE) +
      stat_summary(fun.y=mean, geom="point", size=2, color="black") +
  ggtitle("Distribution of income in different age groups")

p3
```



By looking at this plot we get to know that, Income is highly dependent on the age of a person. Seniors have a high salary range and the salary range of Adult is quite similar to that of Senior. But the salary range of youth is very low compared to the other age groups.

Question 3

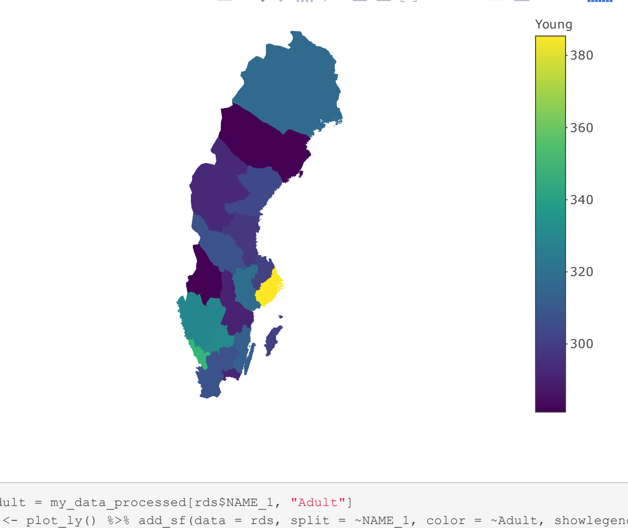
```r
s = interp(my_data_processed$Young, my_data_processed$Adult,
           my_data_processed$Senior, duplicate = "mean")
plot_ly(x=~s$x, y=~s$y, z=~s$z, type="surface") %>% layout(
  scene=list(
  xaxis = list(title = "Young"),
  yaxis = list(title = "Adult"),
  zaxis = list(title = "Senior")
))
```



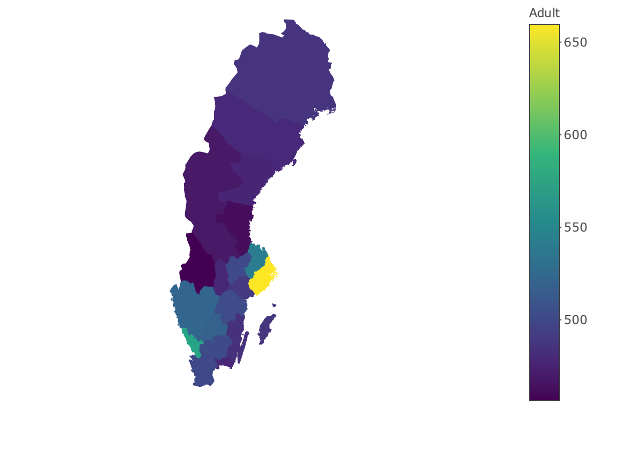
Yes, linear regression will be a suitable model for this dependence, as it would also give a good fit for the data. The range of income for Youth is lower, whereas for adults and seniors the range of income is quiet similar.

Question 4

```r
rds$Young = my_data_processed[rds$NAME_1, "Young"]
Young <- plot_ly() %>% add_sf(data = rds, split = ~NAME_1, color = ~Young, showlegend = F, alpha = 1) %>%
        layout(title = "Choropleth plot of mean income of Young age group")
Young
```



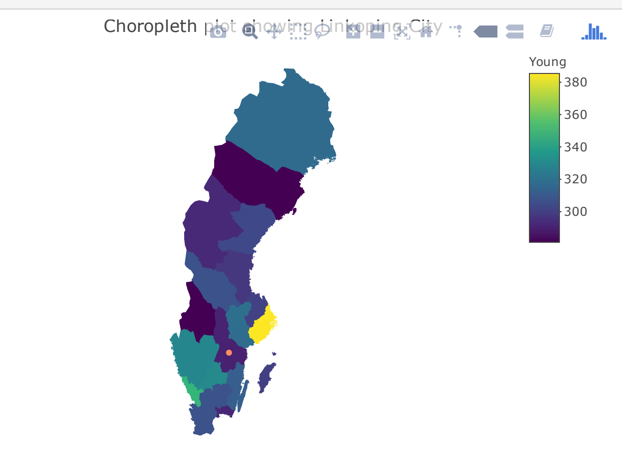

```r
rds$Adult = my_data_processed[rds$NAME_1, "Adult"]
Adult <- plot_ly() %>% add_sf(data = rds, split = ~NAME_1, color = ~Adult, showlegend = F, alpha = 1) %>%
         layout(title = "Choropleth plot of mean income of Adult age group")
Adult
```



This plot provides good information of how income is also dependent on the region you are in, which was not present in the previous plots.The income seems to be higher in the southern parts of Sweden when compared to the northern parts. And also Stockholm has the highest paying jobs.

Question 5

```r
rds$Young = my_data_processed[rds$NAME_1, "Young"]
Linkoping <- plot_ly() %>% add_sf(data = rds, split = ~NAME_1, color = ~Young, showlegend = F, alpha = 1) %>%
             add_markers(x = 15.621373, y = 58.410809, color = "red", hoverinfo = "text", text = "We are here a
t                       Linkoping!!") %>% layout(title = "Choropleth plot showing Linkoping City")

Linkoping
```



Choropleth map with Linkoping City marked in red.