Lab3

Priya (priku577) and Andreas(andch552)

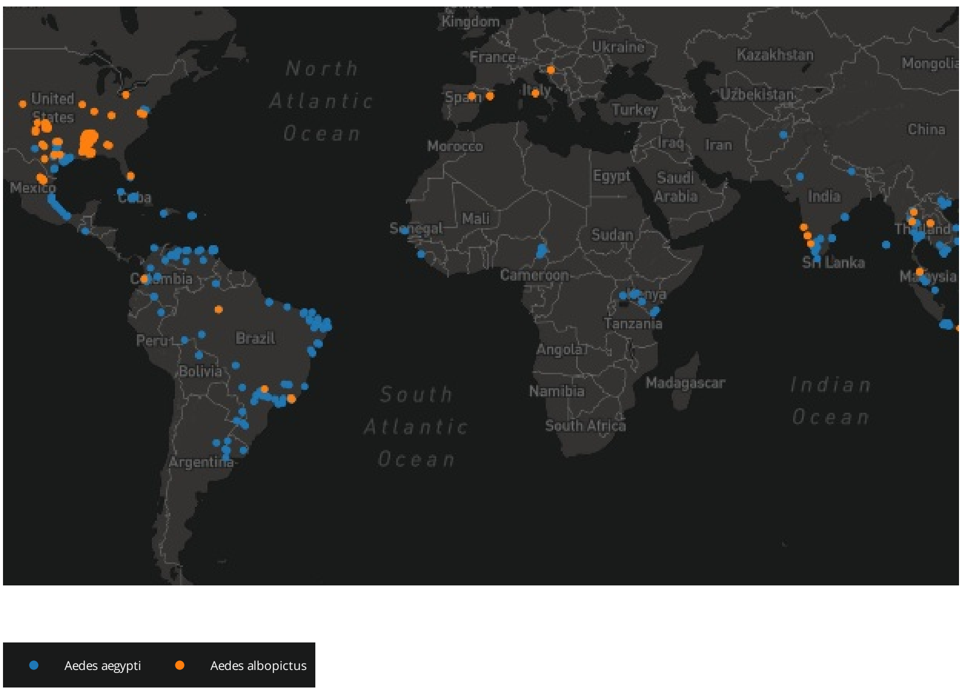
26 September 2018

Assignment 1

1.Mapbox plotly

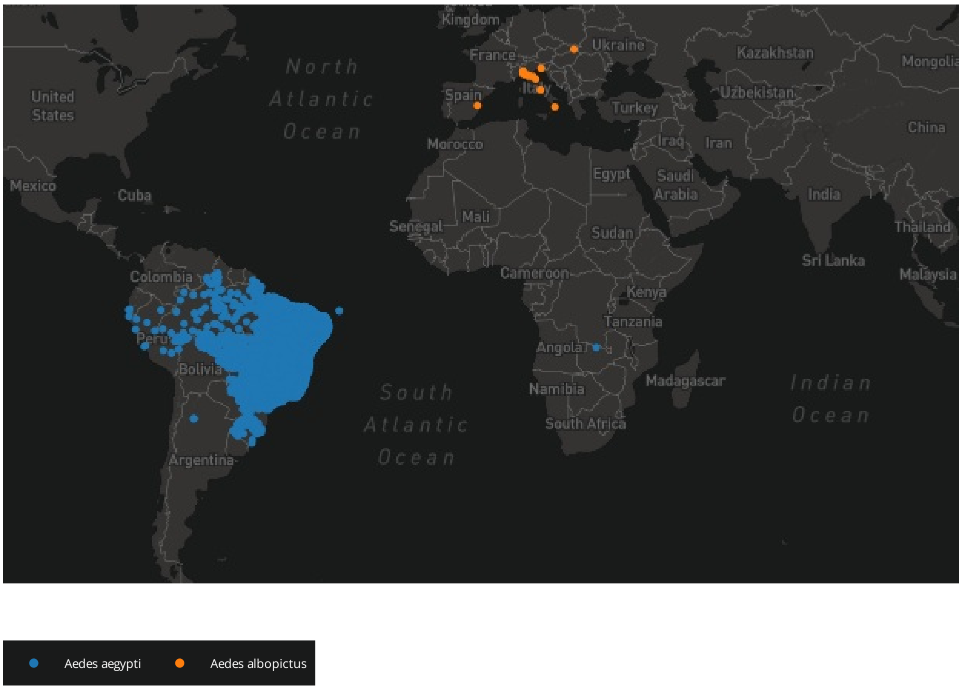
MAPBOX PLOT FOR YEAR 2004

```
library(tidyverse)
library(plotly)
library(RCurl)
library(gridExtra)
library(RColorBrewer)


Sys.setenv('MAPBOX_TOKEN'='pk.eyJ1IjoicXVhcnRlcm1haW5lIiwiYSI6ImNqbWJucjh4MjA2dm0zd25xMmp4ejZzMnQifQ.-FXHcA1t8b_YZ
kdRSUXuGw')

mosquitos<-read.csv('aegypti_albopictus.csv',header=TRUE)
p_2004<-mosquitos%>%filter(YEAR=='2004')%>%plot_mapbox(lat = ~Y, lon = ~X,
                                                        split=~VECTOR,colors = 'Set3',
                                                        mode = 'scattermapbox', hoverinfo='name')%>%
  layout(title = 'MOSQUITOS PLOTS FOR YEAR 2004',
         font = list(color='white'),
         plot_bgcolor = '#191A1A', paper_bgcolor = '#191A1A',
         mapbox = list(style = 'dark'),
         legend = list(orientation = 'h',font = list(size = 8)),
         margin = list(l = 25, r = 25,
                       b = 25, t = 25,
                       pad = 2))
p_2004
```



MAPBOX PLOT FOR YEAR 2013

```
p_2013<-mosquitos %>%filter(YEAR=='2013')%>%plot_mapbox(lat = ~Y, lon = ~X,
                                                        split=~VECTOR,colors = 'Set3',mode = 'scattermapbox', hove
rinfo='name')%>%
  layout(title = 'MOSQUITOS PLOTS FOR YEAR 2013',
         font = list(color='white'),
         plot_bgcolor = '#191A1A', paper_bgcolor = '#191A1A',
         mapbox = list(style = 'dark'),
         legend = list(orientation = 'h',
                       font = list(size = 8)),
         margin = list(l = 25, r = 25,
                       b = 25, t = 25,
                       pad = 2))
p_2013
```



Comparing the two plots we can come to conclusions like

1.The amount of Aedes aegypti in Brazil has increased from 2004 to 2013.Brazil had one of the highest number of Aedes aegypti in 2013 as wellThe regions of rio di Janeiro had more cases.

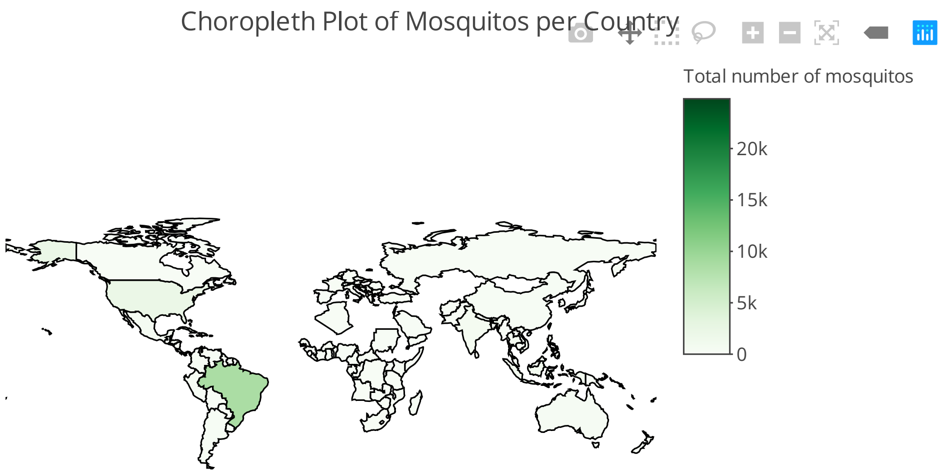
2.The amount of aedes albopictus was high in the US and Taiwan during the yer 2004.The regions most affected were mississippi in US and kaohsiung,Tainam in vietnam.Seems like US could eradicate the aedes albopictus from the missippi area completey but taiwan on other hand in taiwan their number increased.

The perception problems are overplotting,comparison on the poplulation of mosquitoes in each region is difficult.

2.Choropleth Map

CHOROPLETH MAP Equirectangular Projection

```
d<-mosquitos%>%select(c("VECTOR","COUNTRY_ID"))
x<-group_by(d,COUNTRY_ID)%>%count()
s<-as.data.frame(x)

g1 <- list(
  showframe = FALSE,
  showcoastlines = FALSE,
  projection = list(type = 'equirectangular')
)


p1 <- s%>% plot_geo()  %>%
  add_trace(
    z = ~n, color=~n,colors = 'Greens',
    text = ~COUNTRY_ID,locations=~COUNTRY_ID
  ) %>% colorbar(title = 'Total number of mosquitos')%>%
  layout(title = 'Choropleth Plot of Mosquitos per Country',
    geo = g1
  )

p1
```

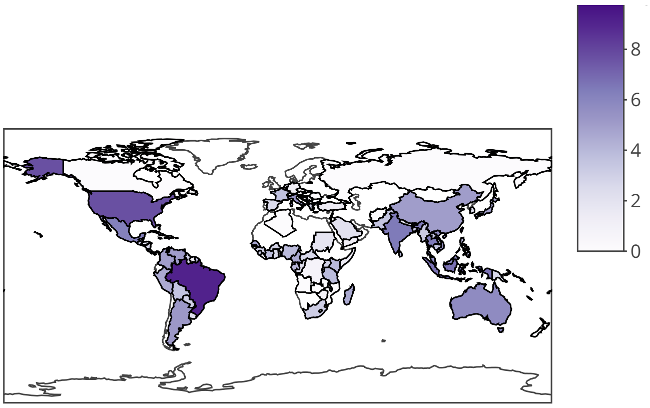


Here the colour scale is from 0-20k .It is very difficult to percieve the data based on this colous scale as instead of different hues intensity varies.

3.Choropleth Map with LOG(Z) transformation

a.Equidistant projection log(z)

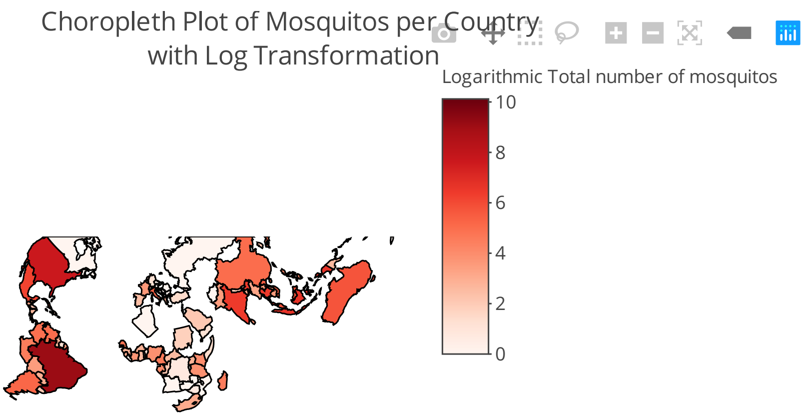
```
s$log_n<-log(s$n)

g2 <- list(
  projection = list(type = "equirectangular")
)


p2 <- s%>% plot_geo()  %>%
  add_trace(
    z = ~log_n, color=~log_n,colors = 'Purples',
    text = ~COUNTRY_ID,locations=~COUNTRY_ID
  ) %>% colorbar(title = 'Logarithmic Total number of mosquitos')%>%
  layout(title = 'Choropleth Plot of Mosquitos per Country \n with Log Transformation',
         geo = g2
  )

p2
```



Here we can see the colour scale is from 0-10 instead of 0-20K.The changes can be noted in a better way.A point to be noted here is Russia had a mosquito count of 1 earlier and now it log(1) and hence 0.Countries like India,USA can be diffetniated better now.

b.Conic Equal Projection log(Z)

```
s$log_n<-log(s$n)

g3 <- list(
  showframe = FALSE,
  showcoastlines = FALSE,
  projection = list(type = "conic equal area")
)


p3 <- s%>% plot_geo()  %>%
  add_trace(
    z = ~log_n, color=~log_n,colors = 'Reds',
    text = ~COUNTRY_ID,locations=~COUNTRY_ID
  ) %>% colorbar(title = 'Logarithmic Total number of mosquitos')%>%
  layout(title = 'Choropleth Plot of Mosquitos per Country \n with Log Transformation',
         geo = g3
  )

p3
```



comparing both the map it is clear that the map 3.a is better.The map 3b leads to misinterpretations.In conical projections as we move away from regions where cone intersects the globe the distortions become more.

Mapbox of Brazil

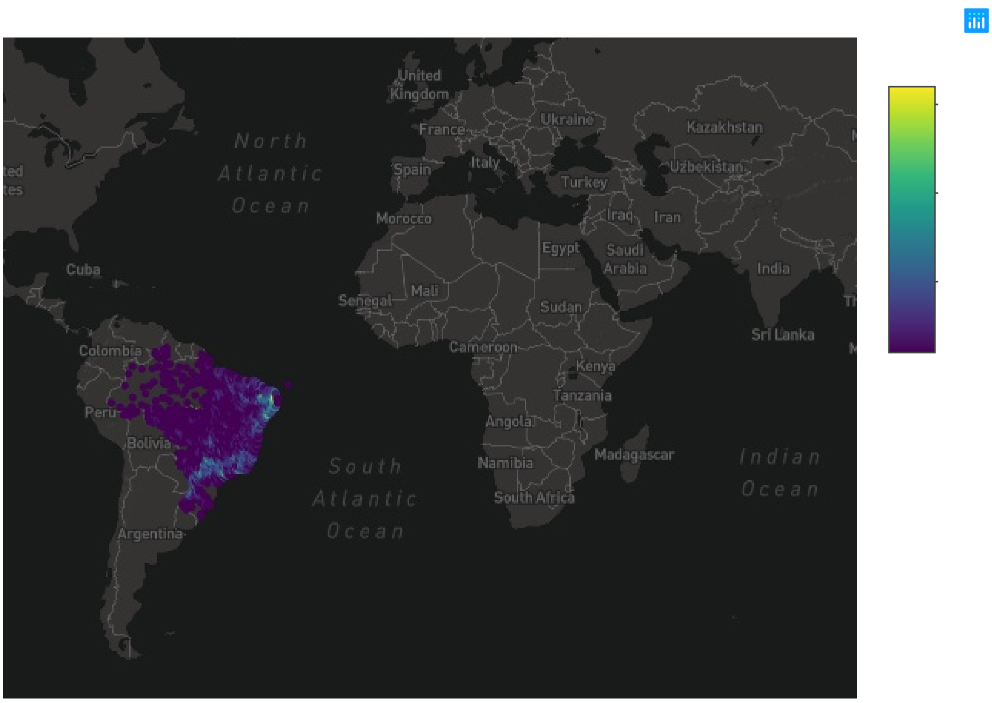
```
mosquitos_brazil<-mosquitos[(mosquitos$COUNTRY == "Brazil" & mosquitos$YEAR =="2013" ), ]

mosquitos_brazil$X1<-cut_interval(mosquitos_brazil$X,100)
mosquitos_brazil$Y1<-cut_interval(mosquitos_brazil$Y,100)


#mos_group_X1<-as.data.frame(mosquitos_brazil%>%group_by(X1)%>%summarise(mean_group_X=mean(X),n_group_X=n()))
#mos_group_Y1<-as.data.frame(mosquitos_brazil%>%group_by(Y1)%>%summarise(mean_group_Y=mean(Y),n_group_Y=n()))


mos<-as.data.frame(mosquitos_brazil)%>%group_by(X1,Y1)%>%summarise(m1=mean(X),m2=mean(Y),N=n())
```

```
## Warning: Factor `X1` contains implicit NA, consider using
## `forcats::fct_explicit_na`
```

```
## Warning: Factor `Y1` contains implicit NA, consider using
## `forcats::fct_explicit_na`
```

```
mos<-as.data.frame(mos)

br <-mos %>%plot_mapbox(lat = ~m2, lon = ~m1,color = ~N ,mode = 'scattermapbox', hoverinfo='name')%>%
  layout(title = 'Mean values of X and Y per group (X1,Y1) and amount of obs per group (X1,Y1) ',
         font = list(color='white'),
         plot_bgcolor = '#191A1A', paper_bgcolor = '#191A1A',
         mapbox = list(style = 'dark'),
         legend = list(orientation = 'h',
                       font = list(size = 8)),
         margin = list(l = 25, r = 25,
                       b = 25, t = 25,
                       pad = 2))

br
```

```
## Warning: Ignoring 1 observations
```



As we can observe from the plot Nova Cruz,Guarabira and Sao Paulo are some cities with high levels of Mosquitos. This discretization defenetly help in analyzing the distribution of mosquitoes accross Brazil.

Assignment 2

1.Young,Adult and senior of the swedish counties are grouped.

```
library(RCurl)
library(tidyverse)

swedish<-read.csv("kd.csv", header=TRUE,encoding = "latin1")

swedish_wide <- spread(swedish, age,X2016)
swedish_wide =swedish_wide[,-2]

names(swedish_wide)<-c("Region","Young","Adult","Senior")
```

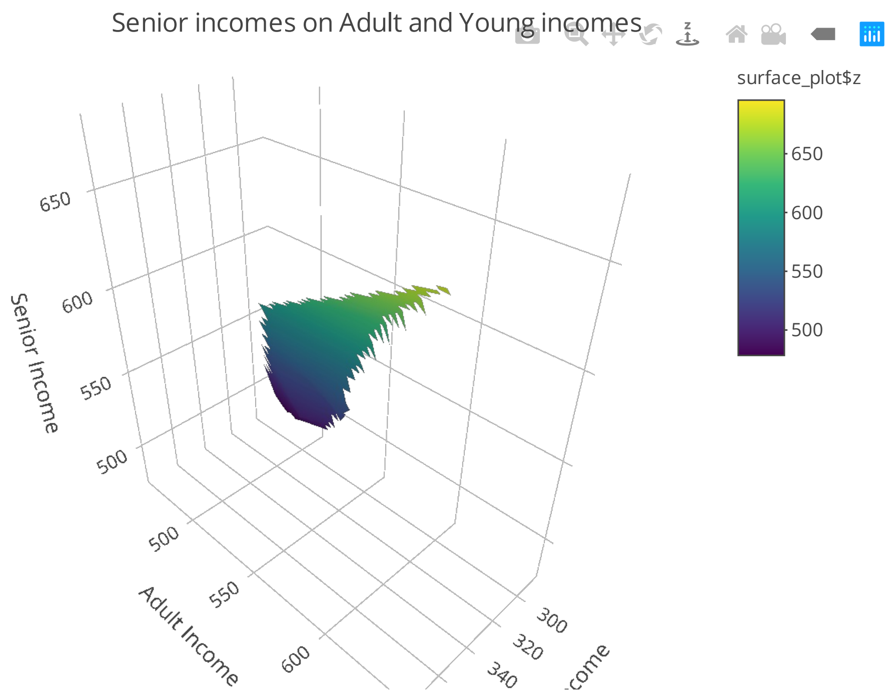
2.Violin plot

```
ee <- swedish_wide %>%
  plot_ly(type = 'violin') %>%
  add_trace(y = ~Young,name = 'Young',box = list(visible = T),
    meanline = list(visible = T),line = list(color = 'pink')
  )%>%
  add_trace(y = ~Adult,name = 'Adults',box = list(visible = T),
            meanline = list(visible = T),line = list(color = 'blue')) %>%
  add_trace(
    y = ~Senior,name = 'Senior',box = list(visible = T),
    meanline = list(visible = T),line = list(color = 'green')
  )%>%
  layout(yaxis = list(title = "income ",
                      zeroline = F), xaxis = list(title = "Age Group"), title = "Income vs  Age Group"
  )


ee
```

The mean income is lowest for age 18-24.This is true may be coz they just start working. All three groups have a small group of people having very high income range (outliers). The age groups of 30-49 and 50-64 have almost similar mean and distribution, with majority having income in range 500 K SEK.

3.Surface plot

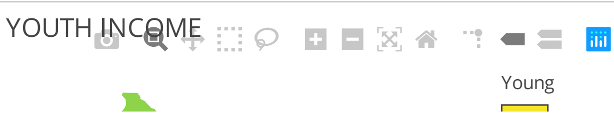
```
library(akima)

surface_plot=interp(swedish_wide$Young, swedish_wide$Adult, swedish_wide$Senior, duplicate = "mean")

plot_ly(x=~surface_plot$x, y=~surface_plot$y, z=~surface_plot$z, type="surface") %>%  layout(
  title = "Senior incomes on Adult and Young incomes",
  scene = list(
    xaxis = list(title = "Young Income"),
    yaxis = list(title = "Adult Income"),
    zaxis = list(title = "Senior Income")
  ))
```

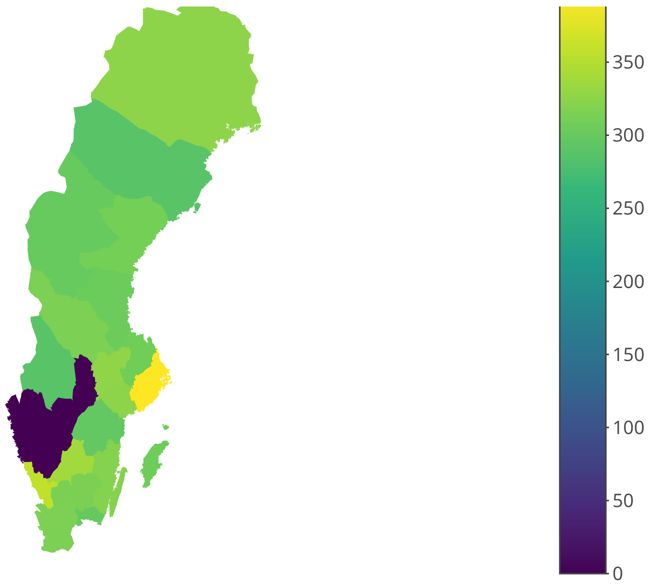


There is linear relationship and hence linear model fits well.

4.Choropleth Plots of Adults and Youngs

```
library(RCurl)
library(tidyverse)
#ibrary(sf)
library(plotly)
library(RCurl)
library(tidyverse)
library(sf)
library(plotly)

swedish<-read.csv("kd.csv", header=TRUE,encoding = "latin1")

swedish_wide <- spread(swedish, age,X2016)
swedish_wide =swedish_wide[,-2]

names(swedish_wide)<-c("Region","Young","Adult","Senior")
map<-readRDS("gadm36_SWE_1_sf.rds")


swedish <- swedish_wide %>% separate(Region,c("Num","Region","County"),extra = 'drop')%>%select("Region","Young","
Adult","Senior")

rownames(swedish)=swedish$Region
map$Young=swedish[map$NAME_1, "Young"]
map$Young[is.na(map$Young)]=0
map$Adult=swedish[map$NAME_1, "Adult"]
map$Adult[is.na(map$Adult)]=0
Young_value <- plot_ly() %>% add_sf(data=map, split=~NAME_1, color=~Young, showlegend=F, alpha=1) %>%
  layout(title="YOUTH INCOME")
Young_value
```

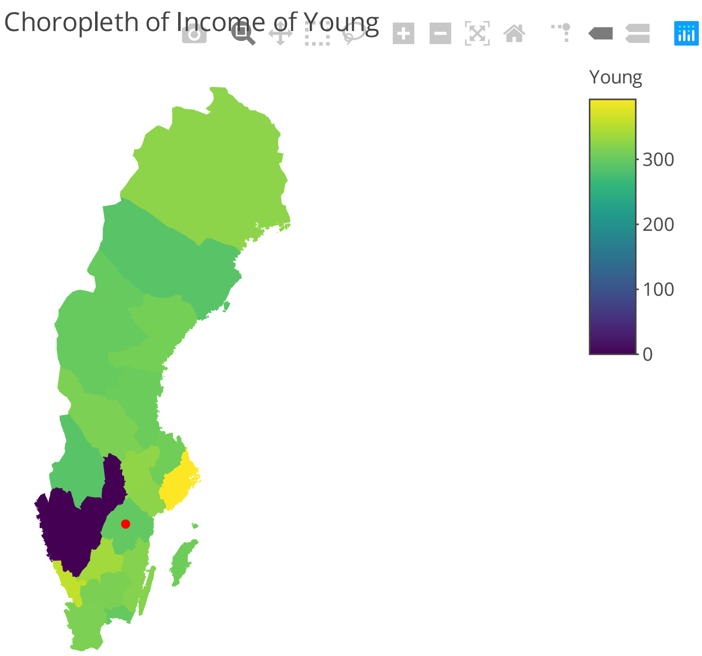




As we can observe from both Choropleth Plots both Young and Adults in Stockhom have higher income regarding rest of Sweden.This seems to be reasonable as stockholm is the capital city and there would be lot of job oppertunities there.It is easier to compare between regions.BUt when we look the country map it is difficult to say about the inome distribution because of the continous colour scale of colour blue.

5.Choropleth Plots of Adults and Youngs with Linkoping coords

```
library(sf)
lat= 58.409814
longitude = 15.624525
name = "Linkoping"
desc = "Linköping, Östergötlands län, SE"

data_point <- data.frame(lat, longitude, name, desc)
plot_ly() %>% add_sf(data=map, split=~NAME_1, color=~Young, showlegend=F, alpha=1,
                     type = "scatter") %>% layout(title = "Choropleth of Income of Young") %>% add_markers(data =
data_point,
                                                                                                     y = ~la
t, x = ~longitude, color = I("red"), text='Linkoping')
```



APPENDIX

```
Appendix Code
GROUP_25
```

```
# Assignment 1 ------------------------------------------------------------


# 1 -------------
library(tidyverse)
library(plotly)
library(RCurl)
library(gridExtra)
library(RColorBrewer)


Sys.setenv('MAPBOX_TOKEN'='pk.eyJ1IjoicXVhcnRlcm1haW5lIiwiYSI6ImNqbWJucjh4MjA2dm0zd25xMmp4ejZzMnQifQ.-FXHcA1t8b_YZ
kdRSUXuGw')

mosquitos<-read.csv('aegypti_albopictus.csv',header=TRUE)
p_2004<-mosquitos%>%filter(YEAR=='2004')%>%plot_mapbox(lat = ~Y, lon = ~X,
                                                       split=~VECTOR,colors = 'Set3',
                                                       mode = 'scattermapbox', hoverinfo='name')%>%
  layout(title = 'MOSQUITOS PLOTS FOR YEAR 2004',
         font = list(color='white'),
         plot_bgcolor = '#191A1A', paper_bgcolor = '#191A1A',
         mapbox = list(style = 'dark'),
         legend = list(orientation = 'h',font = list(size = 8)),
         margin = list(l = 25, r = 25,
                       b = 25, t = 25,
                       pad = 2))
p_2004

p_2013<-mosquitos %>%filter(YEAR=='2013')%>%plot_mapbox(lat = ~Y, lon = ~X,
                                                        split=~VECTOR,colors = 'Set3',mode = 'scattermapbox', hove
rinfo='name')%>%
  layout(title = 'MOSQUITOS PLOTS FOR YEAR 2013',
         font = list(color='white'),
         plot_bgcolor = '#191A1A', paper_bgcolor = '#191A1A',
         mapbox = list(style = 'dark'),
         legend = list(orientation = 'h',
                       font = list(size = 8)),
         margin = list(l = 25, r = 25,
                       b = 25, t = 25,
                       pad = 2))
p_2013

# 2 -------------

d<-mosquitos%>%select(c("VECTOR","COUNTRY_ID"))
x<-group_by(d,COUNTRY_ID)%>%count()
s<-as.data.frame(x)

g1 <- list(
  showframe = FALSE,
  showcoastlines = FALSE,
  projection = list(type = 'equirectangular')
)


p1 <- s%>% plot_geo()  %>%
  add_trace(
    z = ~n, color=~n,colors = 'Greens',
    text = ~COUNTRY_ID,locations=~COUNTRY_ID
  ) %>% colorbar(title = 'Total number of mosquitos')%>%
  layout(title = 'Choropleth Plot of Mosquitos per Country',
         geo = g1
  )

p1

# 3 -------------

s$log_n<-log(s$n)

g2 <- list(
  projection = list(type = "equirectangular")
)


p2 <- s%>% plot_geo()  %>%
  add_trace(
```

```
    z = ~log_n, color=~log_n,colors = 'Purples',
    text = ~COUNTRY_ID,locations=~COUNTRY_ID
  ) %>% colorbar(title = 'Logarithmic Total number of mosquitos')%>%
  layout(title = 'Choropleth Plot of Mosquitos per Country \n with Log Transformation',
         geo = g2
  )

p2


s$log_n<-log(s$n)

g3 <- list(
  showframe = FALSE,
  showcoastlines = FALSE,
  projection = list(type = "conic equal area")
)


p3 <- s%>% plot_geo()  %>%
  add_trace(
    z = ~log_n, color=~log_n,colors = 'Reds',
    text = ~COUNTRY_ID,locations=~COUNTRY_ID
  ) %>% colorbar(title = 'Logarithmic Total number of mosquitos')%>%
  layout(title = 'Choropleth Plot of Mosquitos per Country \n with Log Transformation',
         geo = g3
  )

p3

# 4 --------------

mosquitos_brazil<-mosquitos[(mosquitos$COUNTRY == "Brazil" & mosquitos$YEAR =="2013" ), ]

mosquitos_brazil$X1<-cut_interval(mosquitos_brazil$X,100)
mosquitos_brazil$Y1<-cut_interval(mosquitos_brazil$Y,100)


#mos_group_X1<-as.data.frame(mosquitos_brazil%>%group_by(X1)%>%summarise(mean_group_X=mean(X),n_group_X=n()))
#mos_group_Y1<-as.data.frame(mosquitos_brazil%>%group_by(Y1)%>%summarise(mean_group_Y=mean(Y),n_group_Y=n()))


mos<-as.data.frame(mosquitos_brazil)%>%group_by(X1,Y1)%>%summarise(m1=mean(X),m2=mean(Y),N=n())
mos<-as.data.frame(mos)

br <-mos %>%plot_mapbox(lat = ~m2, lon = ~m1,color = ~N ,mode = 'scattermapbox', hoverinfo='name')%>%
  layout(title = 'Mean values of X and Y per group (X1,Y1) and amount of obs per group (X1,Y1) ',
         font = list(color='white'),
         plot_bgcolor = '#191A1A', paper_bgcolor = '#191A1A',
         mapbox = list(style = 'dark'),
         legend = list(orientation = 'h',
                       font = list(size = 8)),
         margin = list(l = 25, r = 25,
                       b = 25, t = 25,
                       pad = 2))

br

# Assignment 2 ------------------------------------------------------------

# 1 --------------
library(RCurl)
library(tidyverse)

swedish<-read.csv("kd.csv", header=TRUE,encoding = "latin1")

swedish_wide <- spread(swedish, age,X2016)
swedish_wide =swedish_wide[,-2]

names(swedish_wide)<-c("Region","Young","Adult","Senior")

# 2 --------------
ee <- swedish_wide %>%
  plot_ly(type = 'violin') %>%
  add_trace(y = ~Young,name = 'Young',box = list(visible = T),
            meanline = list(visible = T),line = list(color = 'pink')
  )%>%
  add_trace(y = ~Adult,name = 'Adults',box = list(visible = T),
            meanline = list(visible = T),line = list(color = 'blue')) %>%
```

```
  add_trace(
    y = ~Senior,name = 'Senior',box = list(visible = T),
    meanline = list(visible = T),line = list(color = 'green')
  )%>%
  layout(yaxis = list(title = "income ",
                      zeroline = F), xaxis = list(title = "Age Group"), title = "Income vs  Age Group"
  )


ee

# 3 --------------

library(akima)

surface_plot=interp(swedish_wide$Young, swedish_wide$Adult, swedish_wide$Senior, duplicate = "mean")

plot_ly(x=~surface_plot$x, y=~surface_plot$y, z=~surface_plot$z, type="surface") %>%  layout(
  title = "Senior incomes on Adult and Young incomes",
  scene = list(
    xaxis = list(title = "Young Income"),
    yaxis = list(title = "Adult Income"),
    zaxis = list(title = "Senior Income")
  ))

# 4 --------------

library(RCurl)
library(tidyverse)
#ibrary(sf)
library(plotly)
library(RCurl)
library(tidyverse)
library(sf)
library(plotly)

swedish<-read.csv("kd.csv", header=TRUE,encoding = "latin1")

swedish_wide <- spread(swedish, age,X2016)
swedish_wide =swedish_wide[,-2]

names(swedish_wide)<-c("Region","Young","Adult","Senior")
map<-readRDS("gadm36_SWE_1_sf.rds")


swedish <- swedish_wide %>% separate(Region,c("Num","Region","County"),extra = 'drop')%>%select("Region","Young","
Adult","Senior")

rownames(swedish)=swedish$Region
map$Young=swedish[map$NAME_1, "Young"]
map$Young[is.na(map$Young)]=0
map$Adult=swedish[map$NAME_1, "Adult"]
map$Adult[is.na(map$Adult)]=0
Young_value <- plot_ly() %>% add_sf(data=map, split=~NAME_1, color=~Young, showlegend=F, alpha=1) %>%
  layout(title="YOUTH INCOME")
Young_value

# 4 --------------
library(sf)
lat= 58.409814
longitude = 15.624525
name = "Linkoping"
desc = "Linköping, Östergötlands län, SE"

data_point <- data.frame(lat, longitude, name, desc)
plot_ly() %>% add_sf(data=map, split=~NAME_1, color=~Young, showlegend=F, alpha=1,
                     type = "scatter") %>% layout(title = "Choropleth of Income of Young") %>% add_markers(data =
data_point,
                                                                                                           y = ~la
t, x = ~longitude, color = I("red"), text='Linkoping')
```