lab3_group25

Andreas

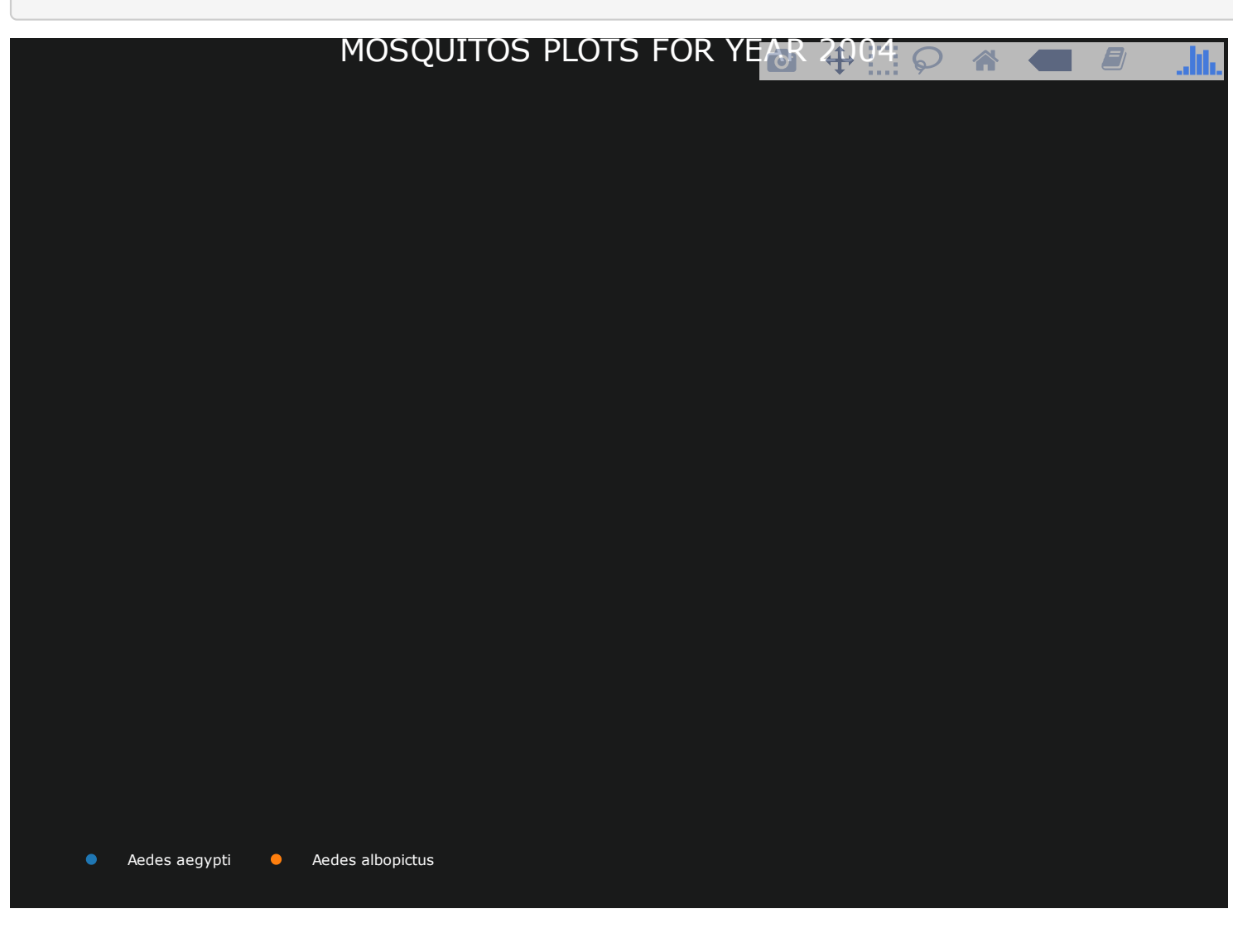
29 Oct 2018

Assignment 1

1. Mapbox plotly

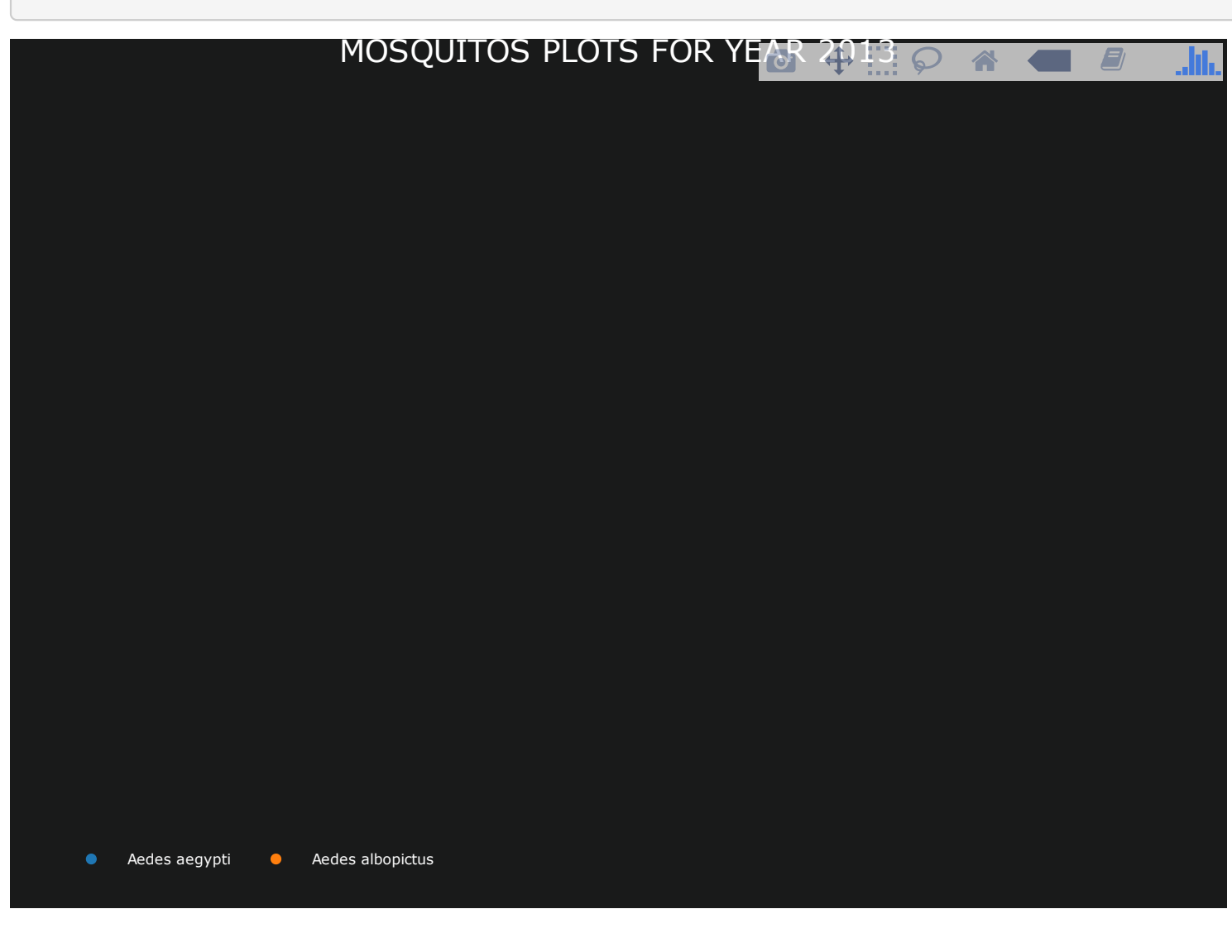
MAPBOX PLOT FOR YEAR 2004

```
library(tidyverse)
library(plotly)
library(RCurl)
library(gridExtra)
library(RColorBrewer)
#setwd("C:/Users/quartermaine/Documents/Visualization/lab_3")

Sys.setenv('MAPBOX_TOKEN'='pk.eyJ1IjoicXVhcnRlcm1haW5lIiwiYSI6ImNqbWJucjh4MjA2dm0zd25xMmp4ejZzMnQifQ.-FXHcA1t8b
_YZkdRSUXuGw')
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', h
overinfo='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

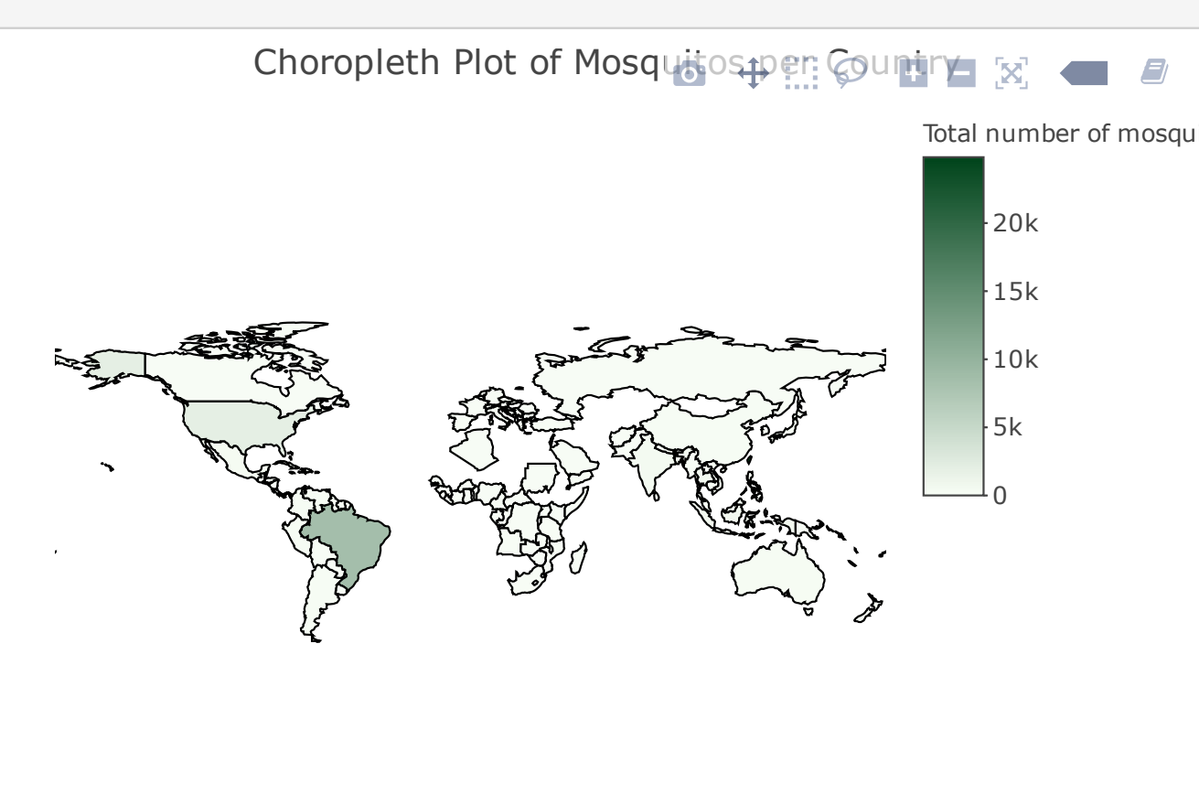
1.The amount of Aedes aegypti in Brazil has increased from 2004 to 2013.Brazil had one of the highest number of Aedes aegypti in 2013 as wellThe regions of rio di Janeiro had more cases.

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.

2Choropleth Map

CHOROPLETH MAP Equirectangular Projection #3Choropleth Map

```
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
```

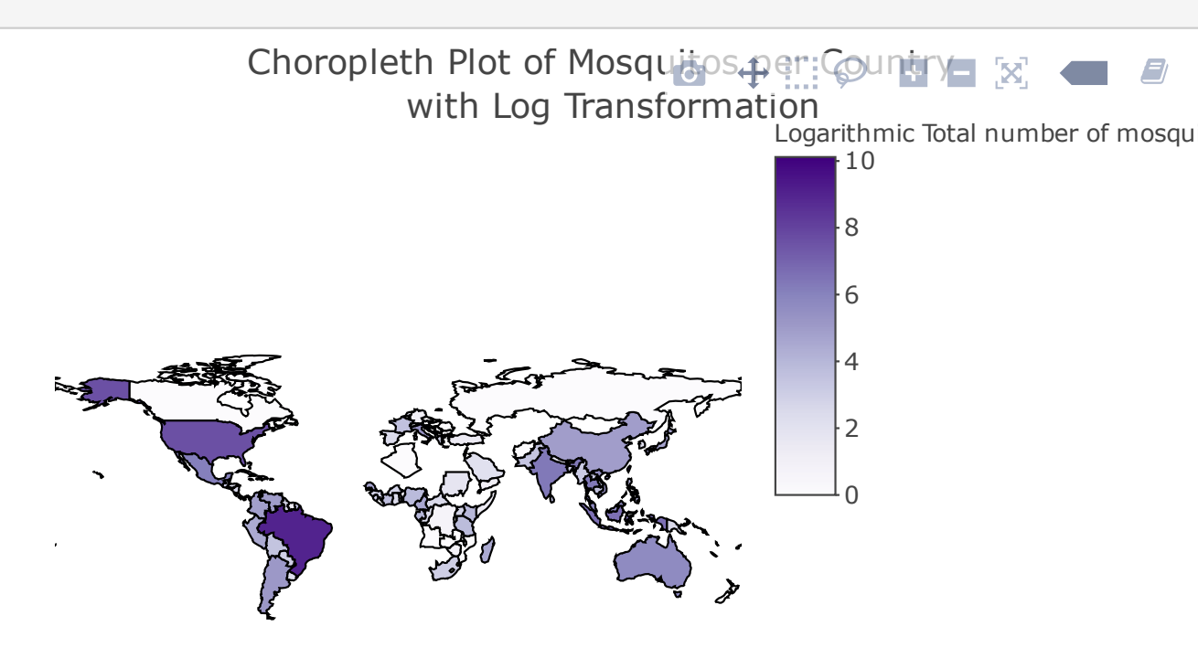


Since we have not scaled the data,most countries are not visible.Brazil has very high number of mosquitoes and hence that is visible.since the range varies from 0-20k countries with small number of mosquitoes are not visible.

3Choropleth Map with LOG(Z) transformation

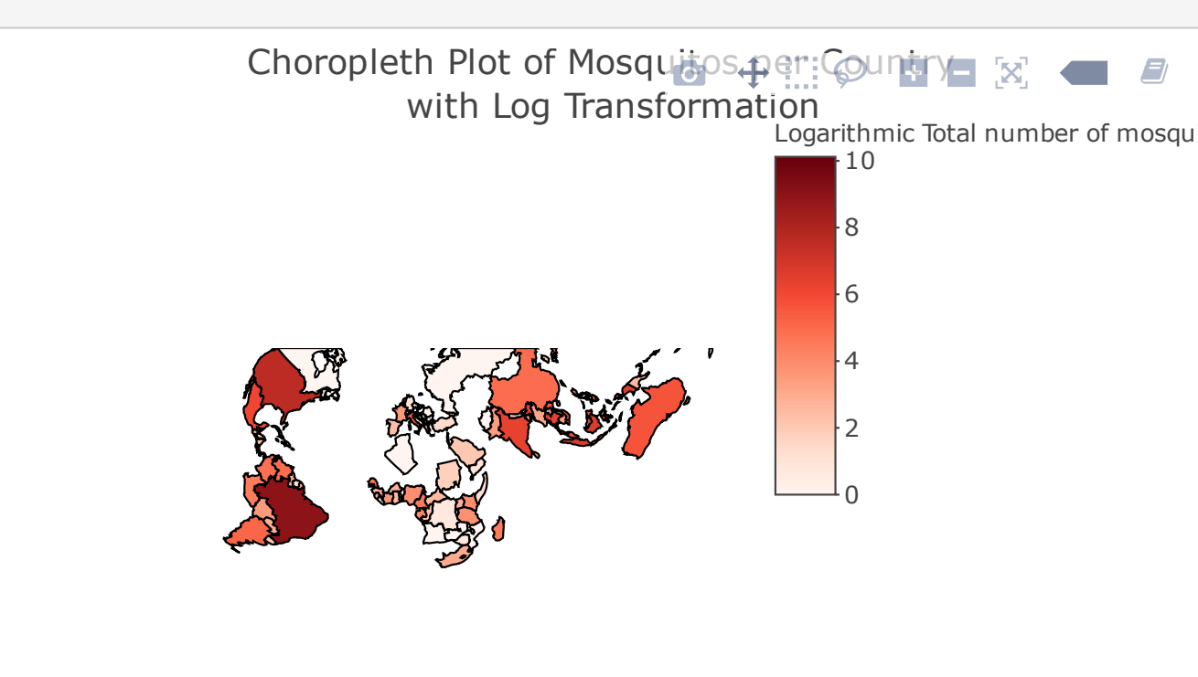
a.Equidistant projection log(z)

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

g2 <- list(
  showframe = FALSE,
  showcoastlines = FALSE,
  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
```



b.Conic Equal Projection log(Z)

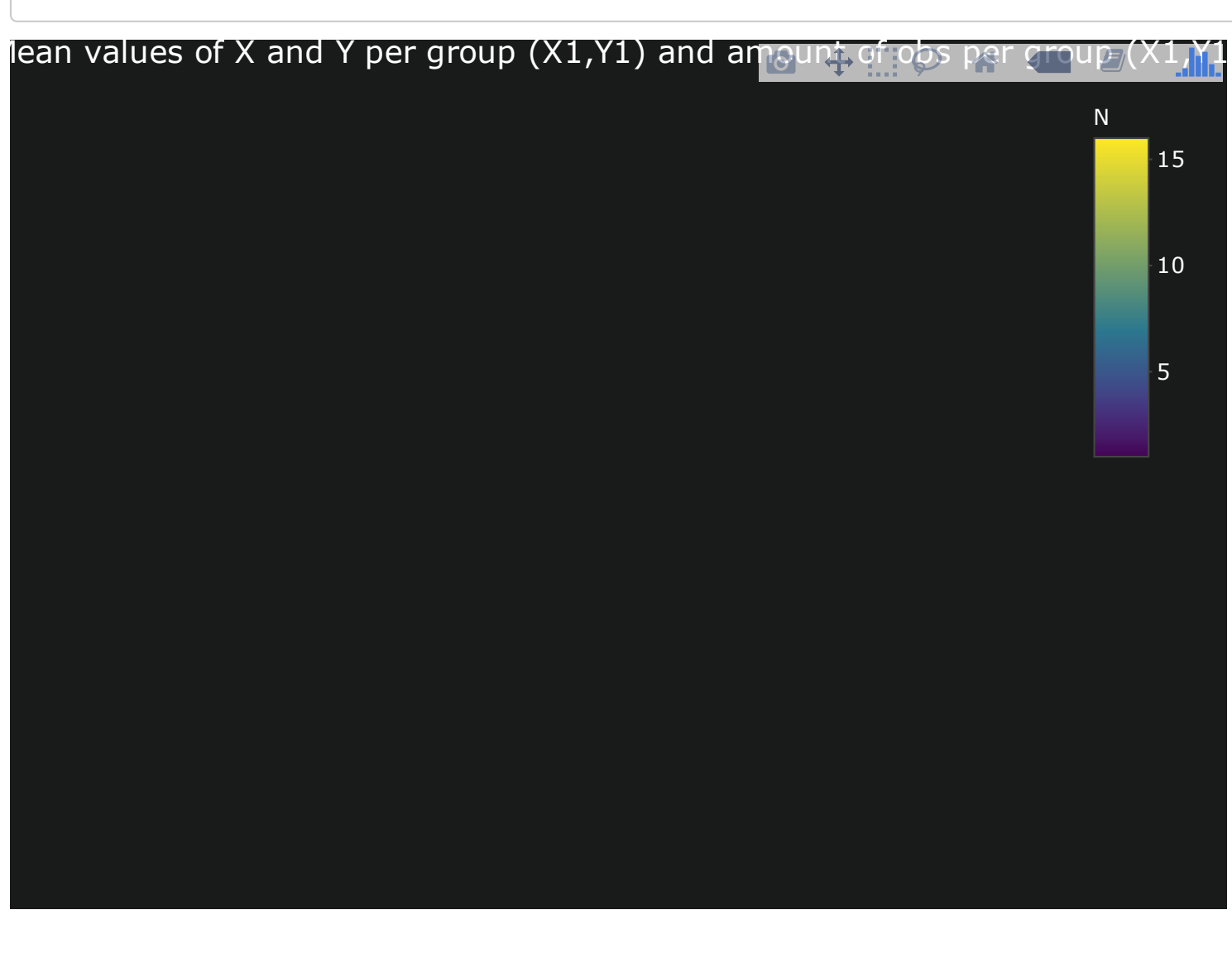
```
s$log_n<-log(s$n)

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
```



WHen we scale the data more information fits in.3a and 3b provide similar information .Only display is different.

```
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
```

```
## 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)
myfile <- getURL("http://www.statistikdatabasen.scb.se/sq/56793", ssl.verifyhost=FALSE, ssl.verifypeer=FALSE)
swedish<-read.csv(textConnection(myfile), header=TRUE)

swedish_wide <- spread(swedish, age,X2016)

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

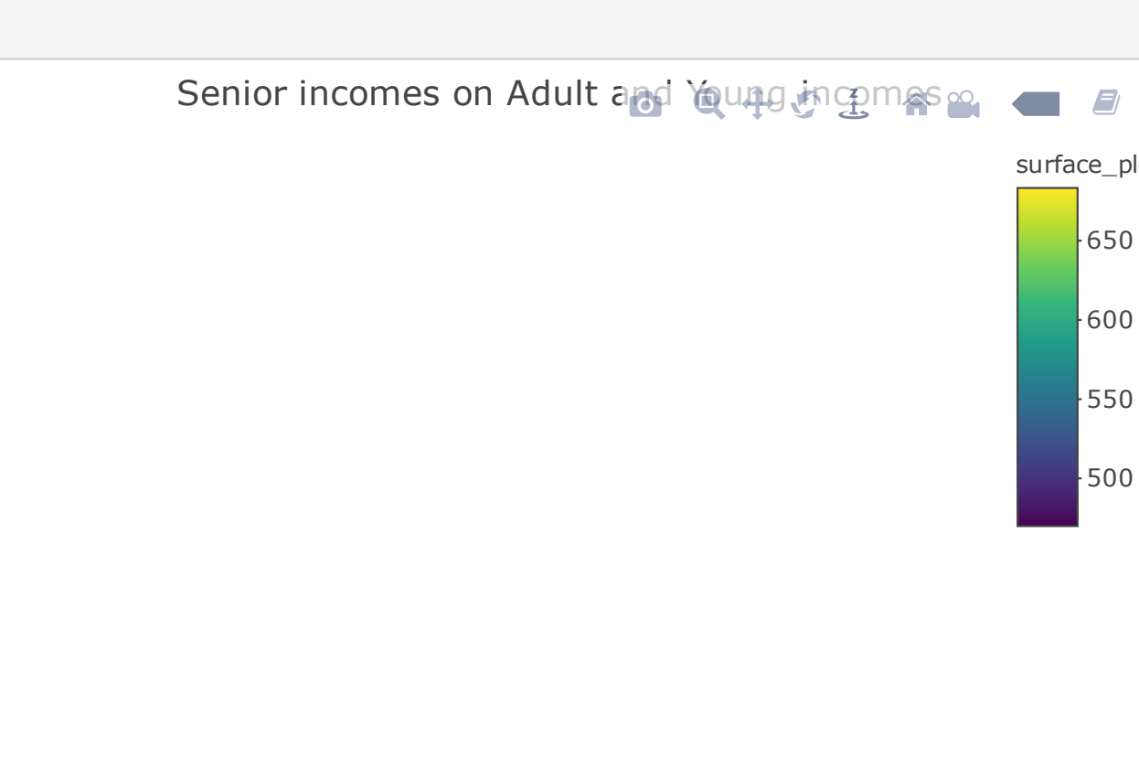
2.Violin plot

```
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
```

From the graph the can see that Seniors have the highest income from the other 2 groups the Adults follow and as epected the Youngs have the smallest income from the other 2 categories.

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")
  ))
```



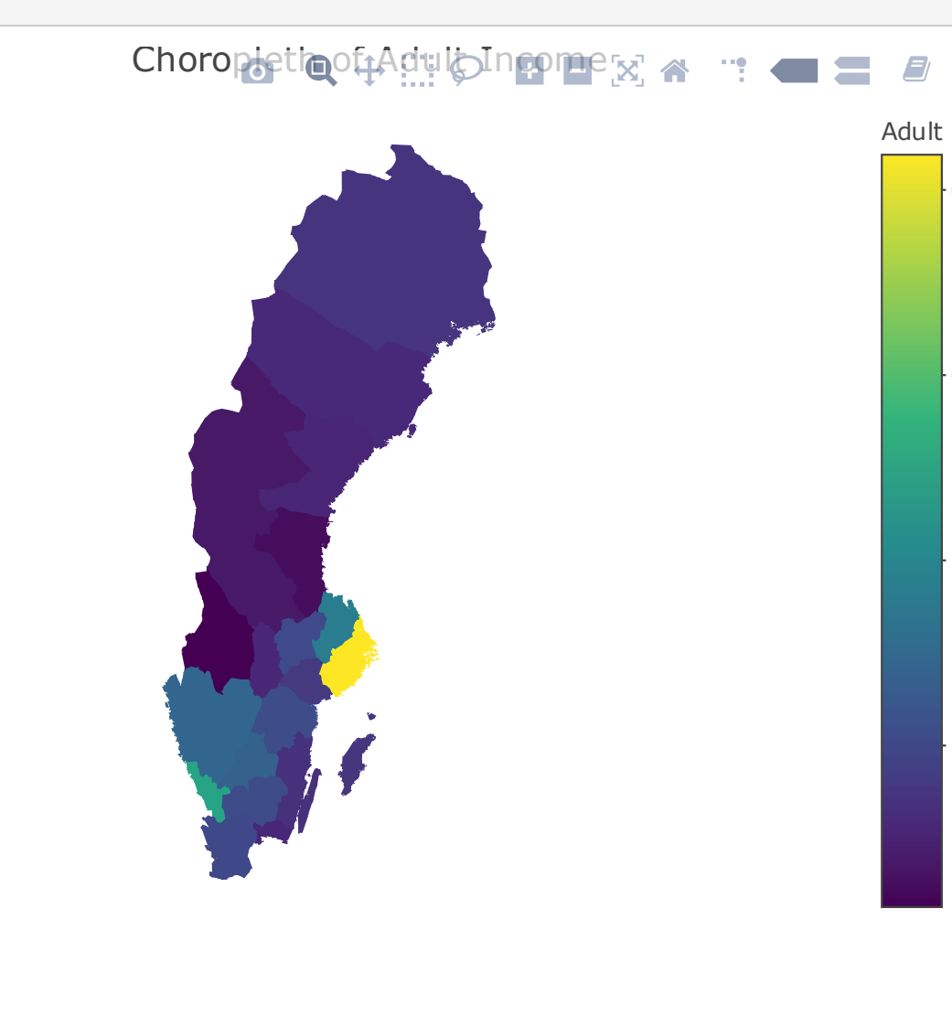
There is linear relationship and hence linear model fits well.

4Choropleth Plots of Adults and Youngs

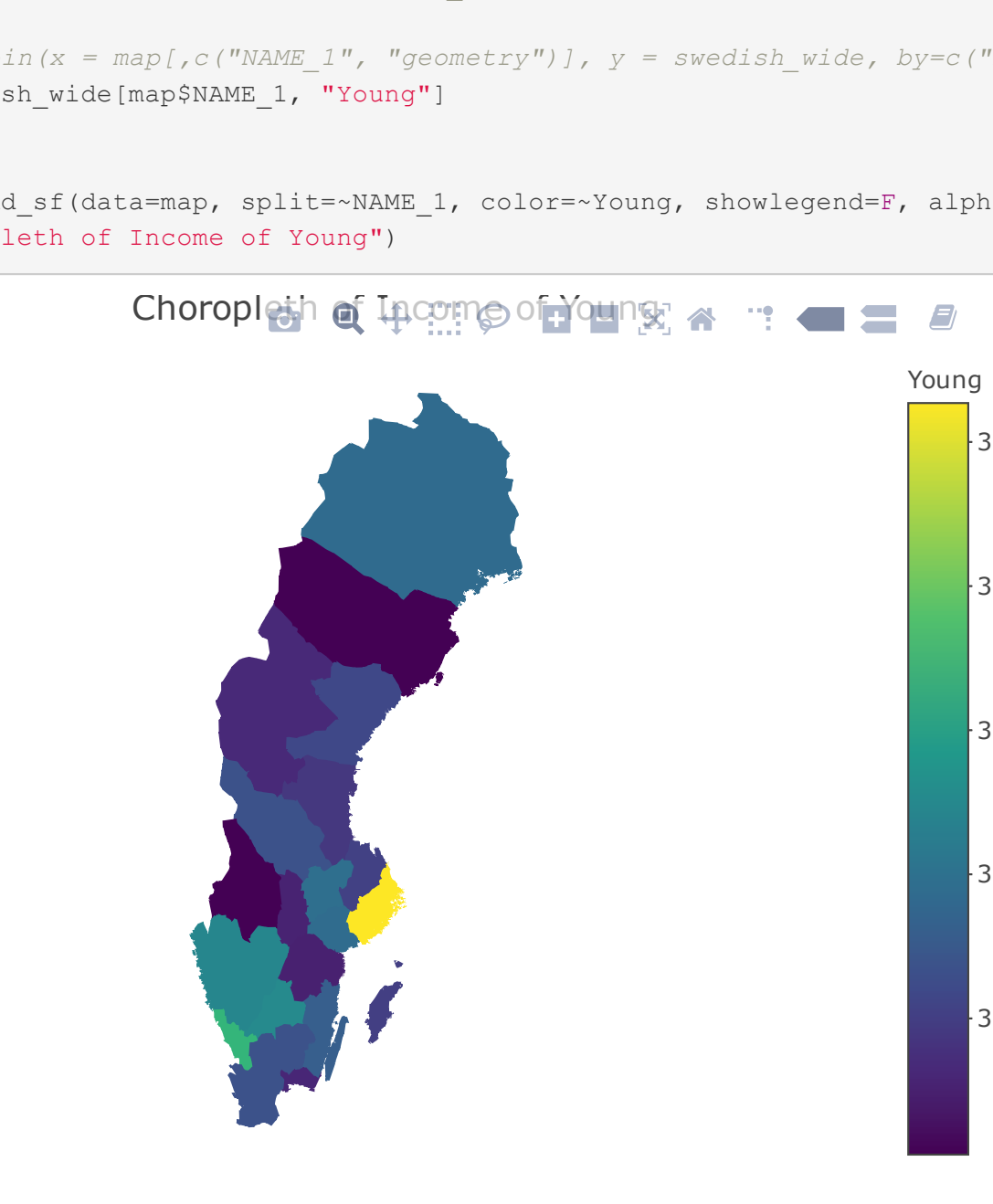
```
library(RCurl)
library(tidyverse)
#setwd("C:/Users/quartermaine/Documents/Visualization/lab_3")

myfile <- getURL("http://www.statistikdatabasen.scb.se/sq/56793", ssl.verifyhost=FALSE, ssl.verifypeer=FALSE)
swedish<-read.csv(textConnection(myfile), header=TRUE,encoding = "UTF-8")

swedish_wide <- spread(swedish, age,X2016)

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


map<-readRDS("gadm36_SWE_1_sf.rds")
map$NAME_1<-replace(map$NAME_1,c(3,6,7,12,13,14,17,18,19,20,21),
                    c("Gavleborg","Jamtland","Jonkoping","Ostergotland",
                      "Skane","Sodermanland","Varmland","Vasterbotten","Vasternorrland","Vastmanland","Vastra"
))
library(dplyr)


swedish_wide<- swedish_wide %>% separate(Region, c("region_no", "region", "type"), " ")
```

```
## Warning: Expected 3 pieces. Additional pieces discarded in 1 rows [12].
```

```
swedish_wide$region <- as.character(swedish_wide$region)
rownames(swedish_wide)=swedish_wide$region



#map_i <- inner_join(x = map[,c("NAME_1", "geometry")], y = swedish_wide, by=c("NAME_1" = "region"))
map$Adult<-swedish_wide[map$NAME_1, "Adult"]


plot_ly() %>% add_sf(data=map, split=~NAME_1, color=~Adult, showlegend=F, alpha=1, type = "scatter") %>% layout
(title = "Choropleth of Adult Income")
```



```
#library(RCurl)
#library(tidyverse)
#setwd("C:/Users/quartermaine/Documents/Visualization/lab_3")
#myfile <- getURL("http://www.statistikdatabasen.scb.se/sq/56793", ssl.verifyhost=FALSE, ssl.verifypeer=FALSE)
#swedish<-read.csv(textConnection(myfile), header=TRUE,encoding = "UTF-8")

#swedish_wide <- spread(swedish, age,X2016)

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


#map<-readRDS("gadm36_SWE_1_sf.rds")
#map<-readRDS("gadm36_SWE_1_sf.rds")
# map$NAME_1<-replace(map$NAME_1,c(3,6,7,12,13,14,17,18,19,20,21),
                    # c("Gavleborg","Jamtland","Jonkoping","Ostergotland",
                     #  "Skane","Sodermanland","Varmland","Vasterbotten","Vasternorrland","Vastmanland","Vastr
a"))
#library(dplyr)


#swedish_wide<- swedish_wide %>% separate(Region, c("region_no", "region", "type"), " ")
#swedish_wide$region <- as.character(swedish_wide$region)

#map <- inner_join(x = map[,c("NAME_1", "geometry")], y = swedish_wide, by=c("NAME_1" = "region"))
map$Young<-swedish_wide[map$NAME_1, "Young"]


plot_ly() %>% add_sf(data=map, split=~NAME_1, color=~Young, showlegend=F, alpha=1, type = "scatter") %>% layout
(title = "Choropleth of Income of Young")
```



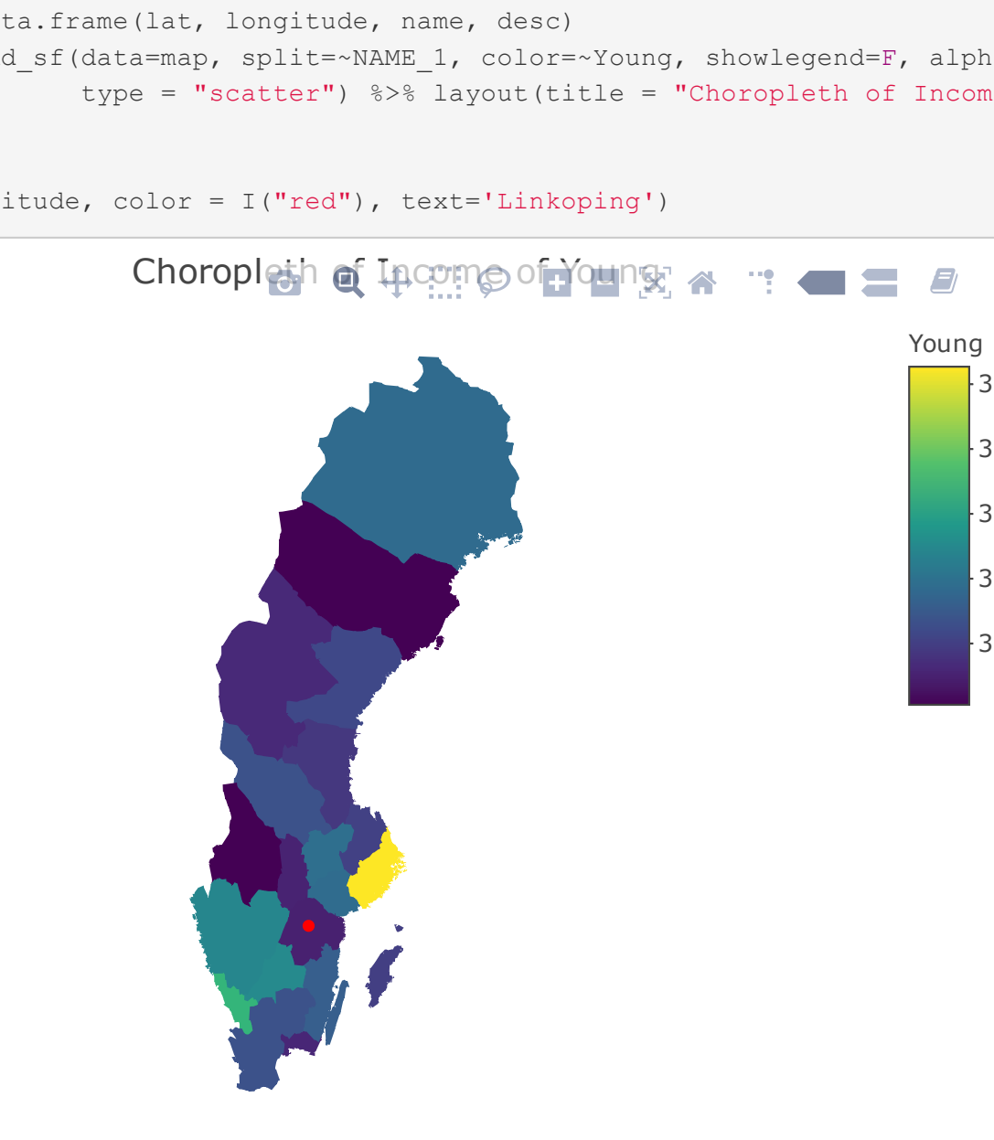
As we can observe from both Choropleth Plots both Young and Adults in Stockhom have higher income regarding rest of Sweden.

5

```
library(sf)
```

```
## Linking to GEOS 3.6.1, GDAL 2.2.3, proj.4 4.9.3
```

```
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 =
 ~lat, x = ~longitude, color = I("red"), text='Linkoping')
```



```
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=~Adult, showlegend=F, alpha=1,
                     type = "scatter") %>% layout(title = "Choropleth of Income of Adults") %>% add_markers(dat
a = data_point,
                                                                                                            y =
 ~lat, x = ~longitude, color = I("red"), text='Linkoping')
```