ThulasiRam_RuppaKrishnan_HW7

```
# Step 1
# Load libraries

library(readxl)
library(zipcode)
library(ggplot2)
library(ggmap)
```

```
## Google's Terms of Service: https://cloud.google.com/maps-platform/terms/.
```

```
## Please cite ggmap if you use it! See citation("ggmap") for details.
```

```
library(stringi)

#download.file('https://www.dropbox.com/s/epqabzao74wq2hu/MedianZIP_2_2.xlsx?dl=0',
              #destfile="C:/Users/rkrishnan/Documents/01 Personal/MS/IST 687/MedianZIP_2_2.xl
sx",
              #method="auto")
# Clear objects
rm(list=ls())

# Load Data
MedianZIP_2_2 <- read_excel("C:/Users/rkrishnan/Documents/01 Personal/MS/IST 687/MedianZIP_2_
2.xlsx")
```

```
## New names:
## * `` -> `..2`
## * `` -> `..3`
## * `` -> `..4`
```

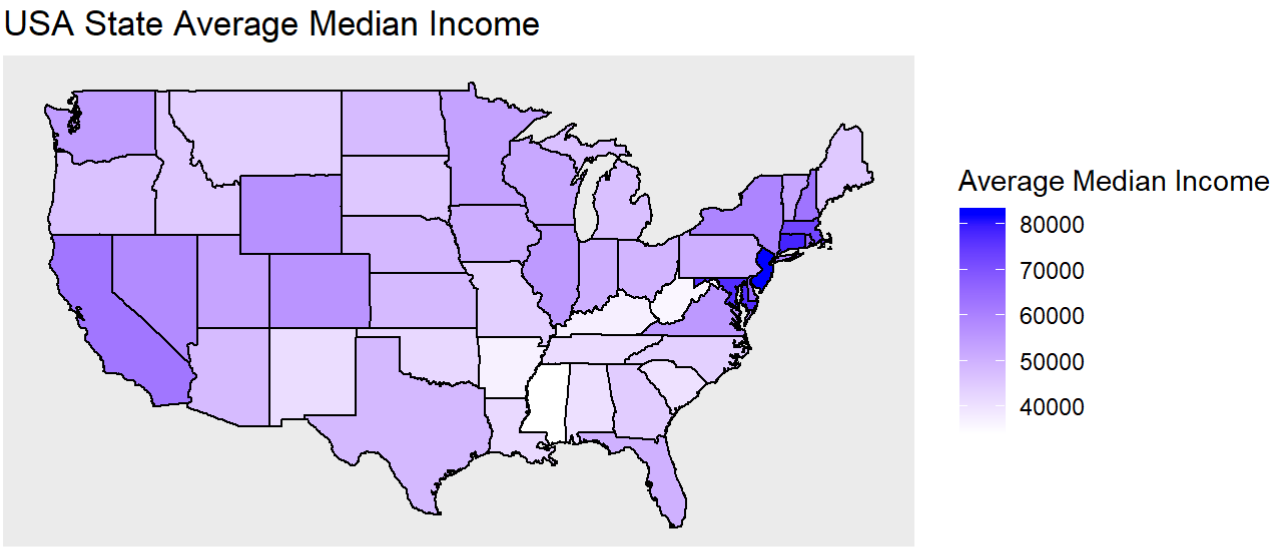
```
MedianZIP_2_2 <-MedianZIP_2_2[-1,]
colnames(MedianZIP_2_2)<-c("zip","median","mean","population")
#MedianZIP_2_2$zip <- as.numeric(MedianZIP_2_2$zip)
MedianZIP_2_2$zip <- stri_pad(MedianZIP_2_2$zip,width = 5,side = "left",pad = "0")
# MedianZIP_2_2 <-na.omit(MedianZIP_2_2)

data("zipcode")
colnames(zipcode) <- c("zip","city","state_cd","latitude","longitude")
df.state <- setNames(state.abb, state.name)
df.state <-data.frame(state_nm=names(df.state),state_cd=df.state)
rownames(df.state) <- NULL
df.zip <- merge(zipcode,df.state,by.x = "state_cd",by.y="state_cd",all.x = TRUE)
us <- map_data("state")

zip.ip <- merge(MedianZIP_2_2,df.zip)
zip.ip$state_nm <-tolower(zip.ip$state_nm)
zip.ip$city <-tolower(zip.ip$city)
zip.ip <- zip.ip[(zip.ip$state_cd != "AK" & zip.ip$state_cd !="DC" & zip.ip$state_cd !="HI"
),]
zip.ip$median <- as.numeric(zip.ip$median)
zip.ip$mean <- as.numeric(zip.ip$mean)
```

```
## Warning: NAs introduced by coercion
```

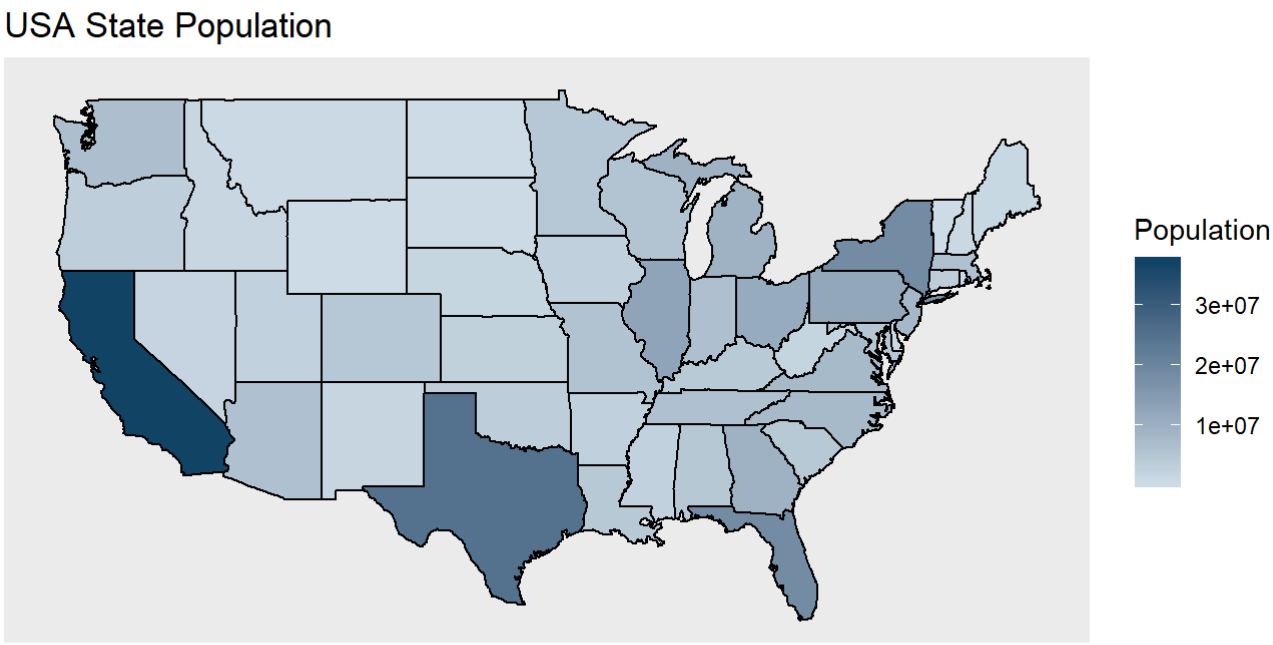
```
zip.ip$population <- as.numeric(zip.ip$population)
zip.ip <-na.omit(zip.ip)
length(unique(zip.ip$state_cd))
```

```
## [1] 48
```

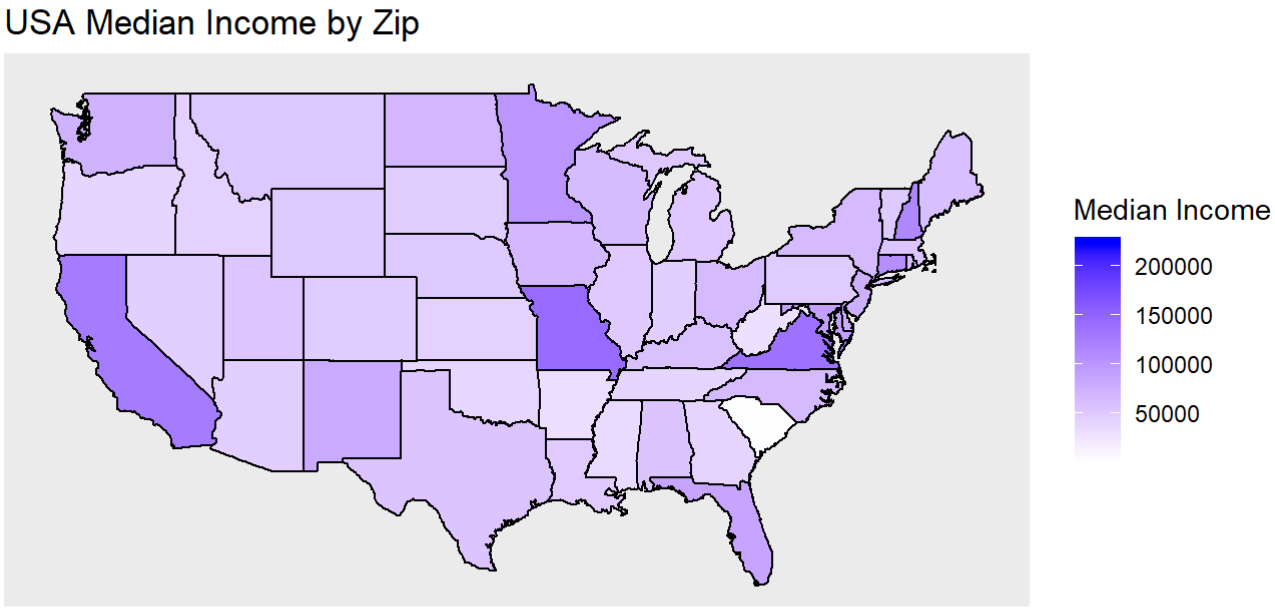
```
# Step 2
# Create a Simple data frame with Average Median Income and Population for Each State

state.i <- tapply(zip.ip$median, zip.ip$state_cd, mean)
state.i <-data.frame(state_cd=names(state.i),avg_median_income=state.i)
rownames(state.i) <- NULL

state.p <- tapply(zip.ip$population, zip.ip$state_cd, sum)
state.p <-data.frame(state_cd=names(state.p),state_population=state.p)
rownames(state.p) <- NULL

# Add State abbreviations and state names
state.ip <- merge(x=state.i,y=state.p,by.x = "state_cd",by.y = "state_cd",all = TRUE)
state.ip <- merge(x=state.ip,y=df.state,by.x="state_cd", by.y="state_cd",all.x = TRUE)
state.ip$state_nm <- tolower(state.ip$state_nm)

# U.S Map representing the color with the average median income of that state

ggplot(data=state.ip,aes(map_id=state_nm)) +geom_map(map=us, aes(fill=state.ip$avg_median_inc
ome),color="black") + expand_limits(x=us$long,y=us$lat) + coord_map() + ggtitle("USA State Av
erage Median Income") + labs(x = "", y = "")+ scale_x_continuous(breaks = NULL) + scale_y_con
tinuous(breaks = NULL) + scale_fill_continuous(low = "white", high= "blue", guide = guide_col
orbar(title = "Average Median Income"))
```



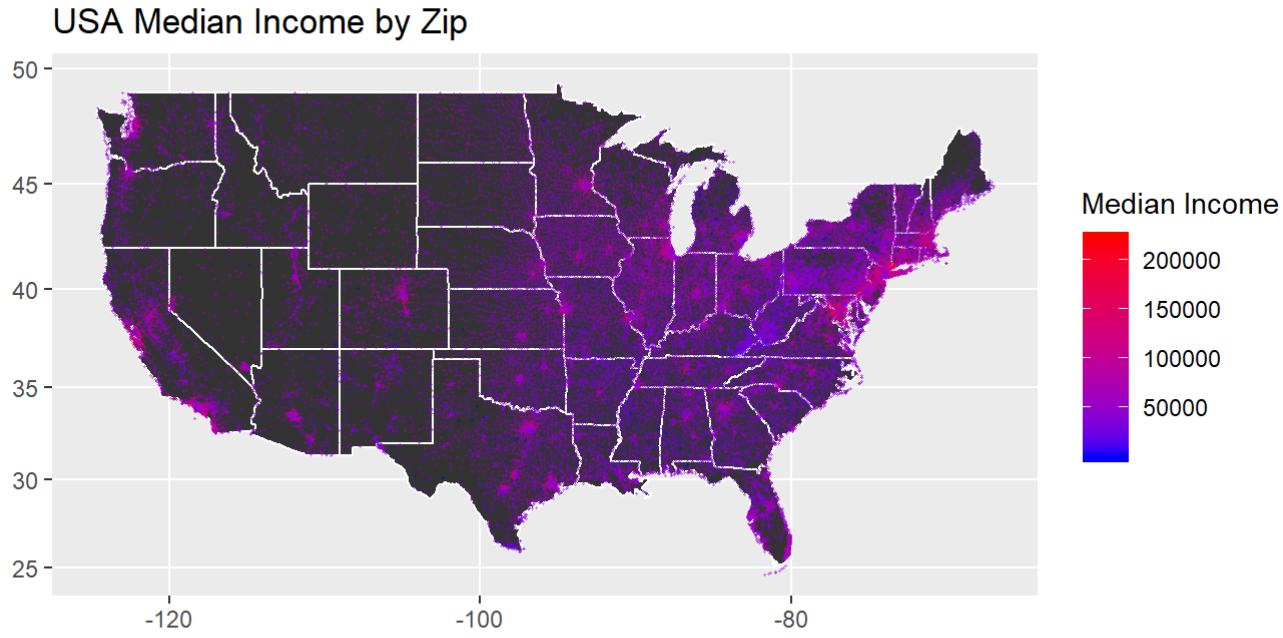

```
# U.S Map representing the color with the population of that state

ggplot(data=state.ip,aes(map_id=state_nm)) +geom_map(map=us, aes(fill=state.ip$state_populati
on),color="black") + expand_limits(x=us$long,y=us$lat) + coord_map() + ggtitle("USA State Pop
ulation") + labs(x = "", y = "")+ scale_x_continuous(breaks = NULL) + scale_y_continuous(brea
ks = NULL) + scale_fill_continuous(low = "#ccdbe5", high= "#114365",guide = guide_colorbar(ti
tle = "Population"))
```



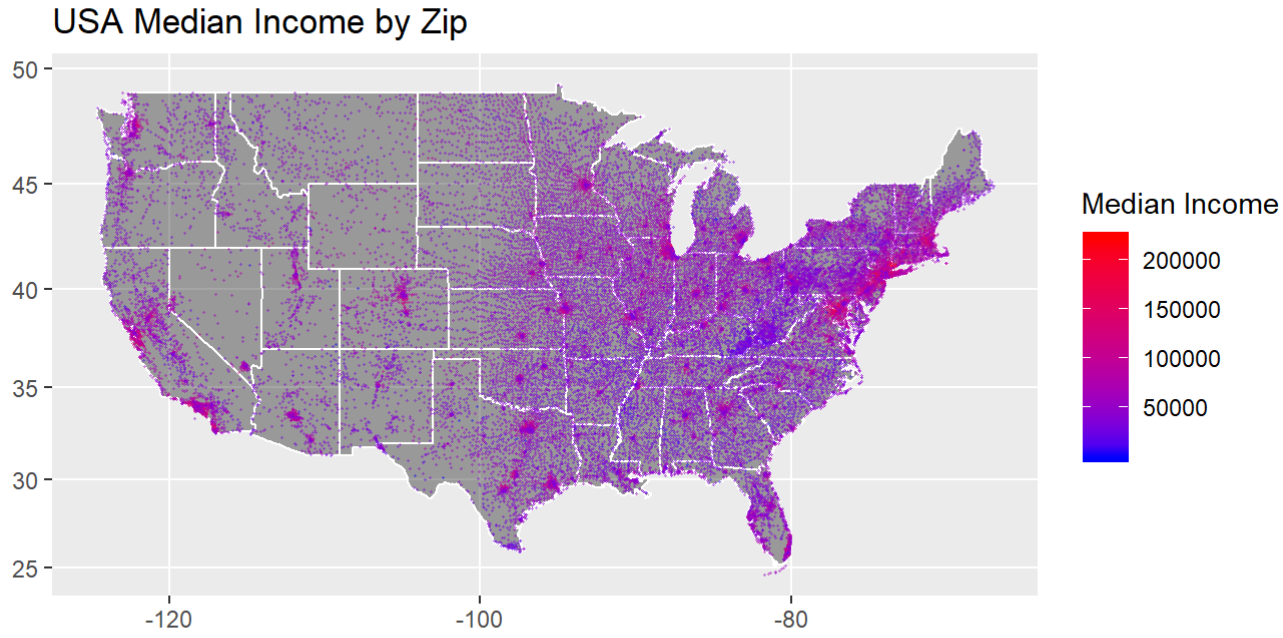

```
# Step 3
# Median Income by Zip using geom_point and fill
ggplot(data=zip.ip,aes(map_id=state_nm)) +geom_map(map=us,aes(fill=zip.ip$median),colour="bla
ck") + expand_limits(x=us$long,y=us$lat) + coord_map() + ggtitle("USA Median Income by Zip")
 + labs(x = "", y = "")+ scale_x_continuous(breaks = NULL) + scale_y_continuous(breaks = NULL
) + scale_fill_continuous(low = "white", high= "blue", guide = guide_colorbar(title = "Median
Income"))
```



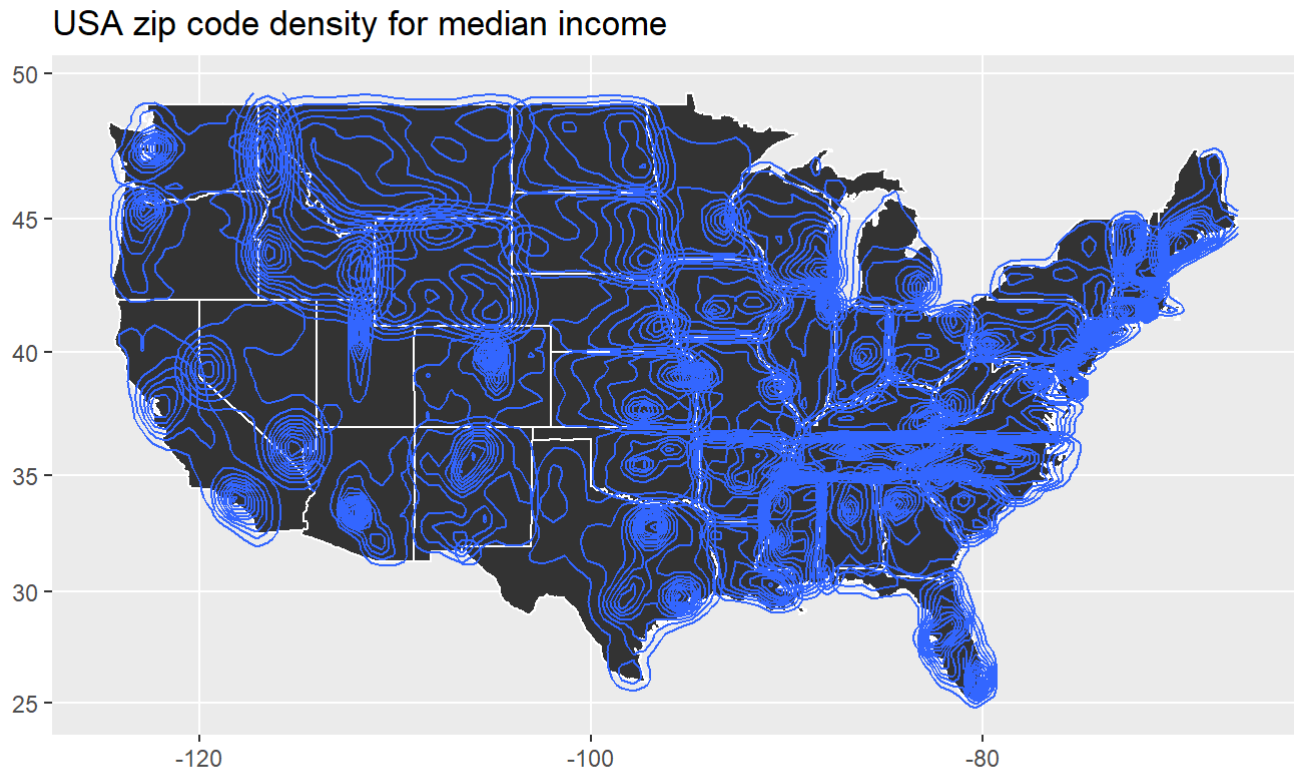

```
# Median Income by Zip geom_point and geom_map
ggplot(data=zip.ip,aes(map_id=state_nm)) +geom_map(map=us,colour="white") + expand_limits(x=us$long,y=us$lat) + coord_map()+ geom_point(aes(x=longitude,y=latitude,color=median),size=0.15,alpha=0.35) + ggtitle("USA Median Income by Zip") + labs(x = "", y = "",color ="Median Income") + scale_colour_gradient(low = "blue", high = "red")
```



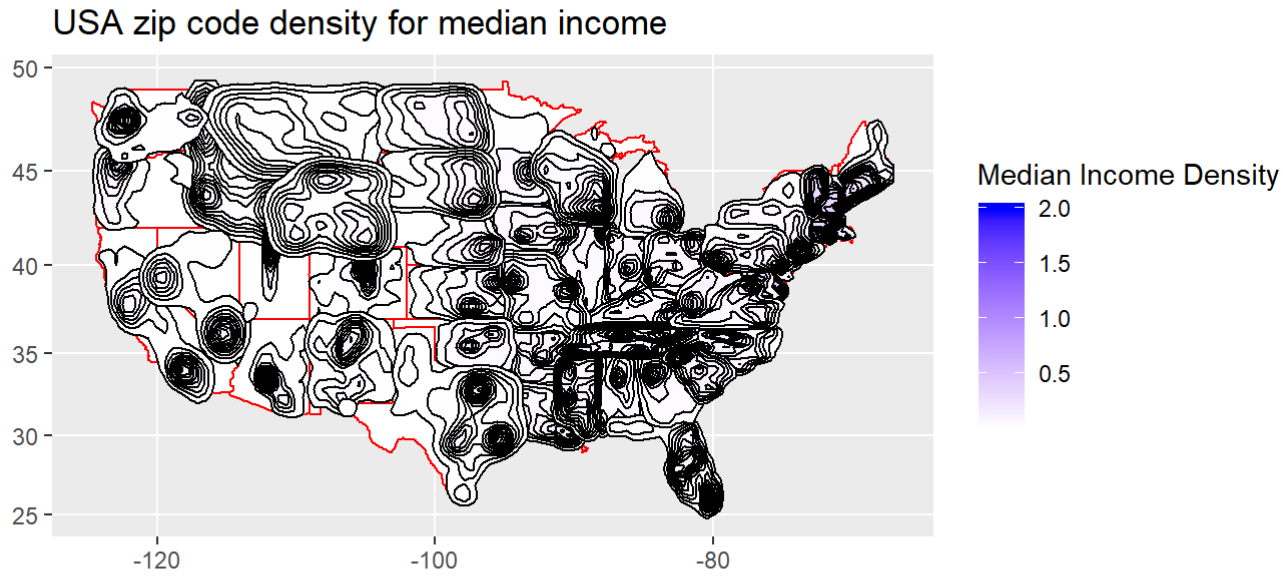

```
# Median Income by Zip using geom_point and geom_polygon
ggplot(data=zip.ip,aes(x=longitude, y=latitude)) +geom_polygon(data=us,aes(long,lat,group=gro
up),colour="white",fill="black",alpha=0.35) + geom_point(aes(color=median),size=0.15,alpha=0.
35) + ggtitle("USA Median Income by Zip") + labs(x = "", y = "",color ="Median Income") + sca
le_colour_gradient(low = "blue", high = "red") + coord_map()
```



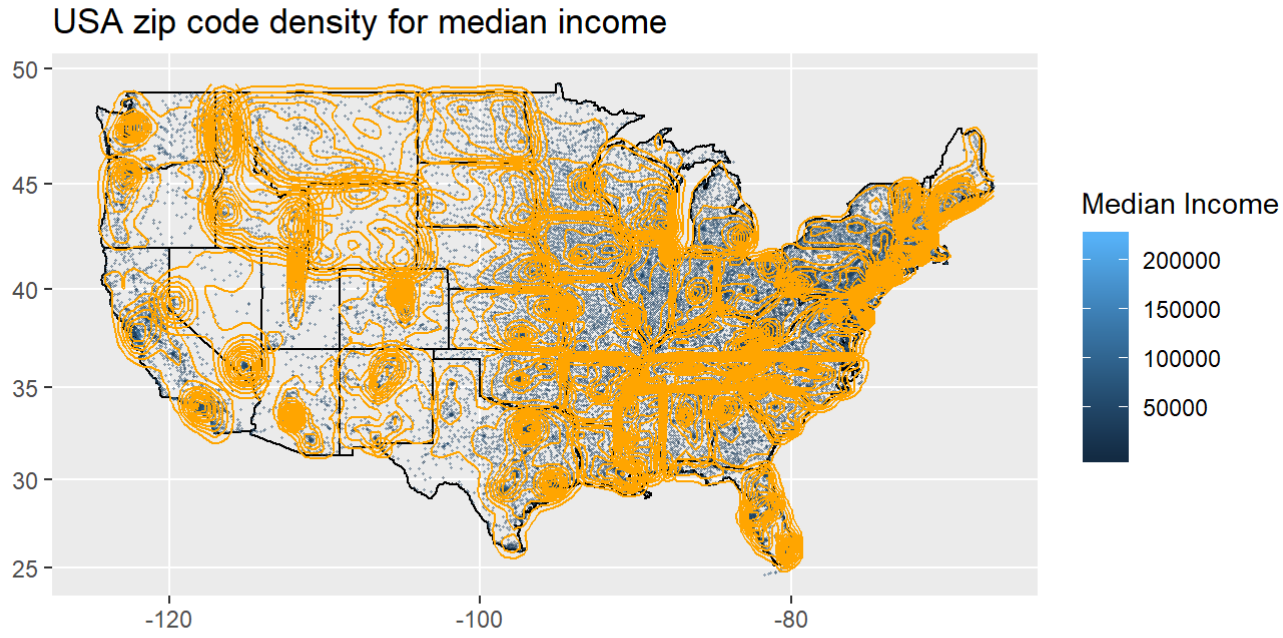

```
# Step 4
# Zip code density using geom_map without geom_point()

ggplot(data=zip.ip,aes(map_id=state_nm)) +geom_map(map=us,colour="white") + expand_limits(x=u
s$long,y=us$lat)  + coord_map() +ggtitle("USA zip code density for median income") + labs(x =
"", y = "",color ="Median Income") + geom_density_2d(aes(x=longitude,y=latitude,color=media
n))
```



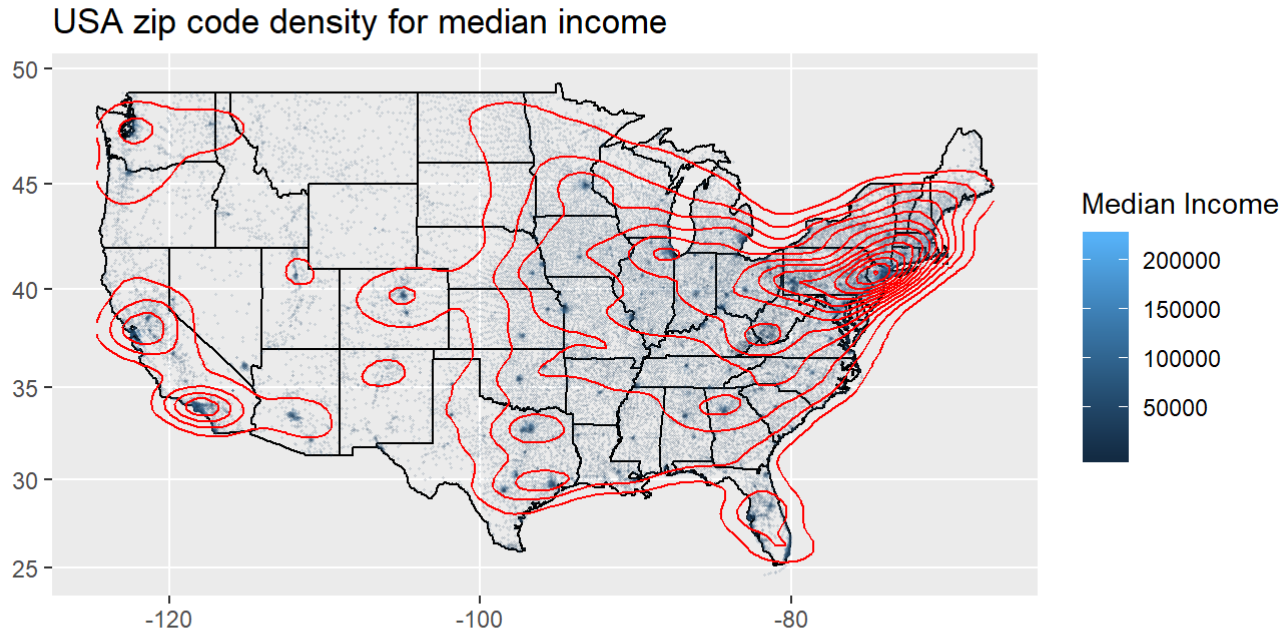

```
ggplot(data=zip.ip,aes(map_id=state_nm)) +geom_map(map=us,colour="red",fill="white") + expand
_limits(x=us$long,y=us$lat)  + coord_map() +ggtitle("USA zip code density for median income")
+ labs(x = "", y = "",color ="Median Income") + geom_density_2d(aes(x=longitude,y=latitude),c
olor="white") + stat_density_2d(aes(x=longitude,y=latitude,fill = stat(level)),color="black",
geom = "polygon")+ scale_fill_continuous(low = "white", high= "blue", guide = guide_colorbar
(title = "Median Income Density"))
```



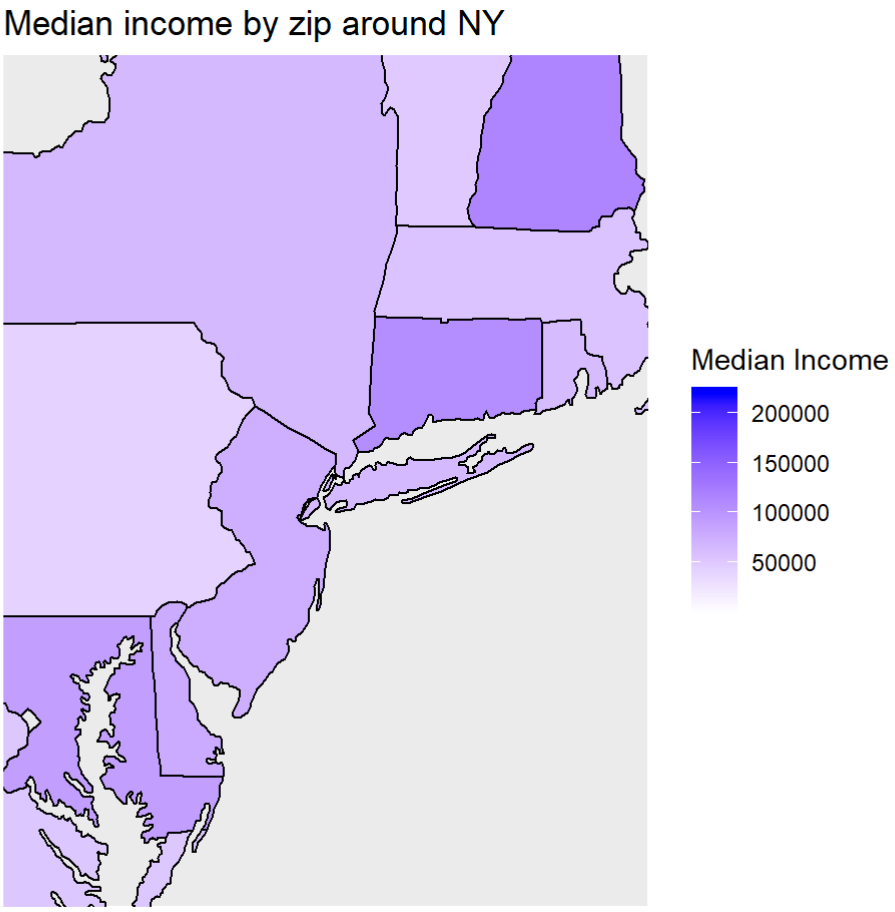

```
# Zip code density using geom_map with geom_point

ggplot(data=zip.ip,aes(map_id=state_nm)) +geom_map(map=us,colour="black",fill=NA) + expand_li
mits(x=us$long,y=us$lat) + coord_map()+ geom_point(aes(x=longitude,y=latitude,color=median),s
ize=0.15,alpha=0.35) + geom_density_2d(aes(x=longitude,y=latitude),color="orange") + ggtitle(
"USA zip code density for median income") + labs(x = "", y = "",color ="Median Income")
```



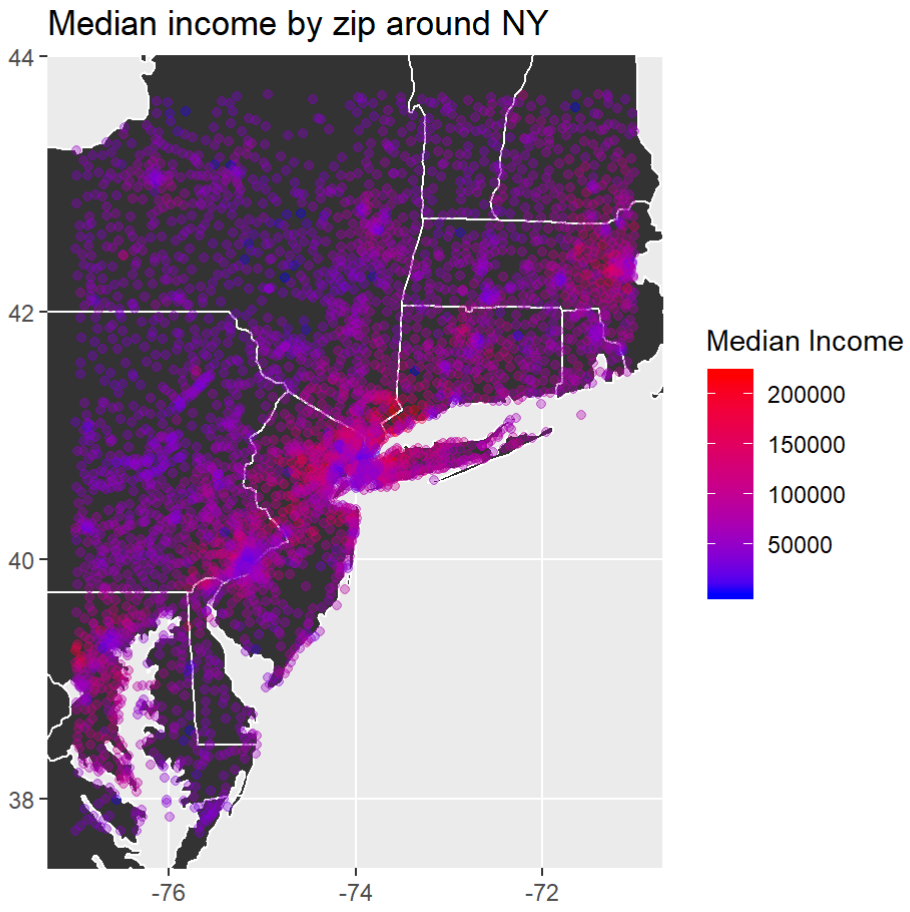

```
# Zip code density using geom_polygon

ggplot(data=zip.ip,aes(x=longitude, y=latitude)) +geom_polygon(data=us,aes(long,lat,group=gro
up),colour="black",fill=NA,alpha=0.35) + geom_point(aes(color=median),size=0.15,alpha=0.1) +
 geom_density_2d(color="red")  + labs(x="",y="",color ="Median Income") +  ggtitle("USA zip c
ode density for median income") + coord_map()
```



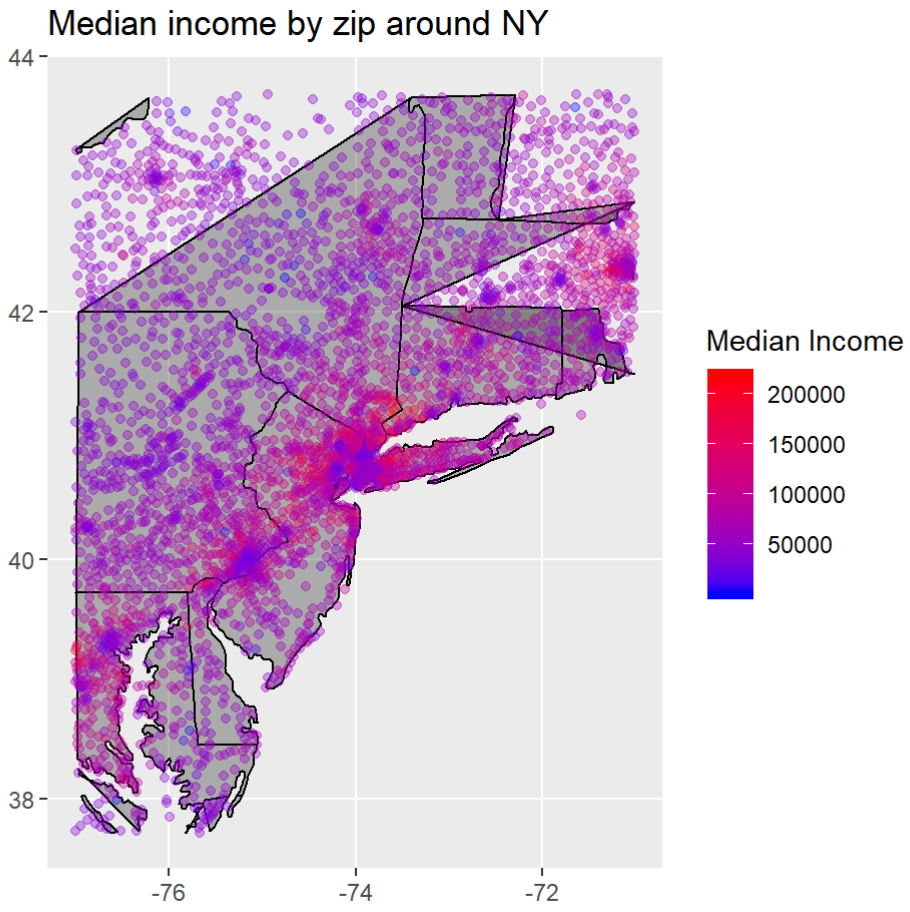

```
# Step 5
#Zoom into the region around NYC
# store corodintates for NYC and calculate x , y range for zoom_amt 3
ny_long <- -74.00594
ny_lat <- 40.71278
zoom_amt <- 3
xlimit_frm_ny <- c(ny_long - zoom_amt,ny_long + zoom_amt)
ylimit_frm_ny <- c(ny_lat - zoom_amt, ny_lat + zoom_amt)

#create a dataset for NYC with zip code x,y range for +- zoom_amt
zoom.ny.zip.ip <- zip.ip
zoom.ny.zip.ip <- zoom.ny.zip.ip[zoom.ny.zip.ip$longitude>xlimit_frm_ny[1],]
zoom.ny.zip.ip <- zoom.ny.zip.ip[zoom.ny.zip.ip$longitude<xlimit_frm_ny[2],]
zoom.ny.zip.ip <- zoom.ny.zip.ip[zoom.ny.zip.ip$latitude>ylimit_frm_ny[1],]
zoom.ny.zip.ip <- zoom.ny.zip.ip[zoom.ny.zip.ip$latitude<ylimit_frm_ny[2],]

# Repeat step 3 for NY data set

# Median Income by Zip using geom_point and fill
ggplot(data=zoom.ny.zip.ip,aes(map_id=state_nm)) +geom_map(map=us,aes(fill=zoom.ny.zip.ip$med
ian),colour="black") + expand_limits(x=xlimit_frm_ny,y=ylimit_frm_ny) + coord_map() + ggtitle
("Median income by zip around NY") + labs(x = "", y = "")+ scale_x_continuous(breaks = NULL)
 + scale_y_continuous(breaks = NULL) + scale_fill_continuous(low = "white", high= "blue", gui
de = guide_colorbar(title = "Median Income"))
```



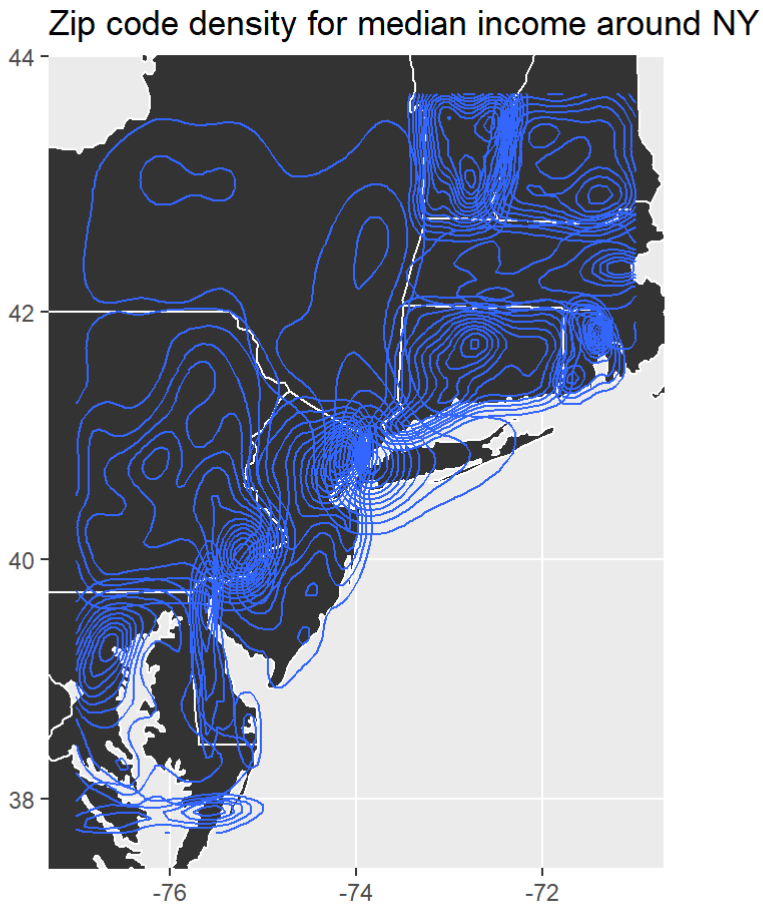

```
# Median Income by Zip geom_point and geom_map
ggplot(data=zoom.ny.zip.ip,aes(map_id=state_nm)) +geom_map(map=us,colour="white") + expand_li
mits(x=xlimit_frm_ny,y=ylimit_frm_ny) + coord_map()+ geom_point(aes(x=longitude,y=latitude,co
lor=median),alpha=0.35) + ggtitle("Median income by zip around NY") + labs(x = "", y = "",col
or ="Median Income") + scale_colour_gradient(low = "blue", high = "red")
```



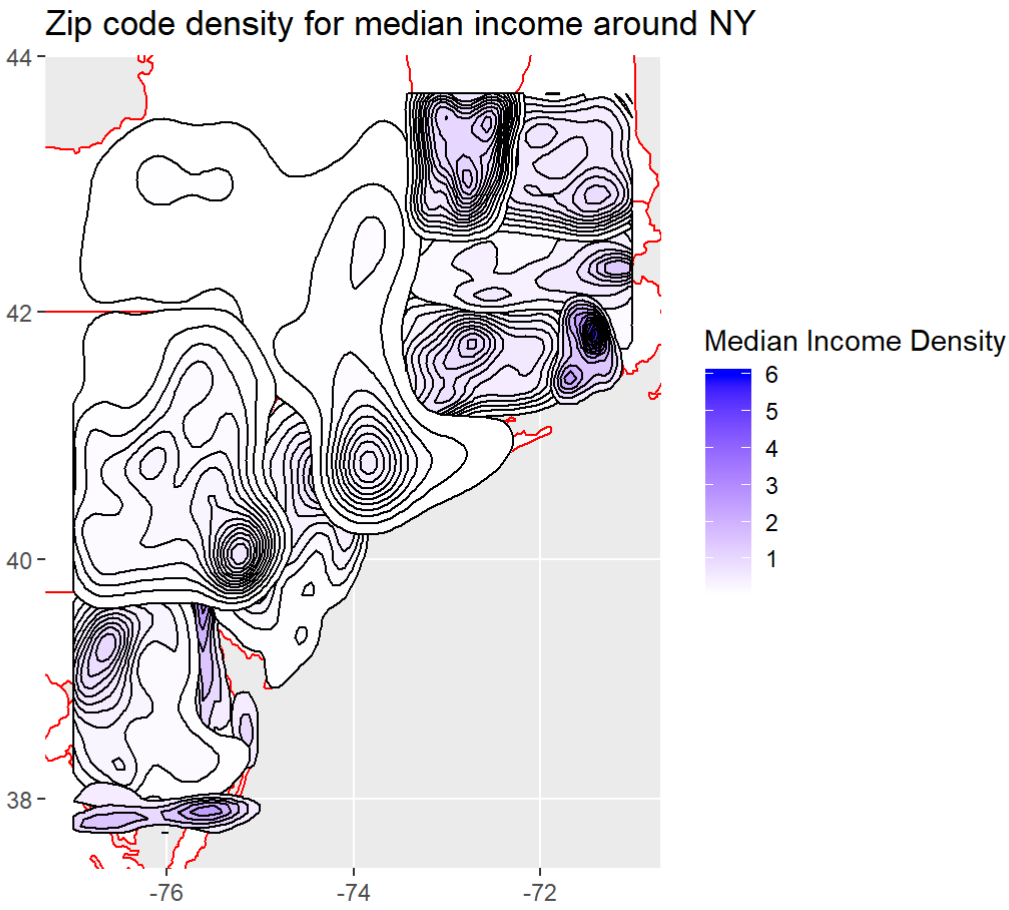

```
# Median Income by Zip using geom_point and geom_polygon
ggplot(data=zoom.ny.zip.ip,aes(x=longitude, y=latitude))  +geom_polygon(data=us,aes(long,lat,
group=group),colour="black",alpha=0.35) + xlim(xlimit_frm_ny) +ylim(ylimit_frm_ny) +geom_poin
t(aes(color=median),alpha=0.35) + ggtitle("Median income by zip around NY") + labs(x = "", y
 = "",color ="Median Income") + scale_colour_gradient(low = "blue", high = "red") + coord_map
()
```



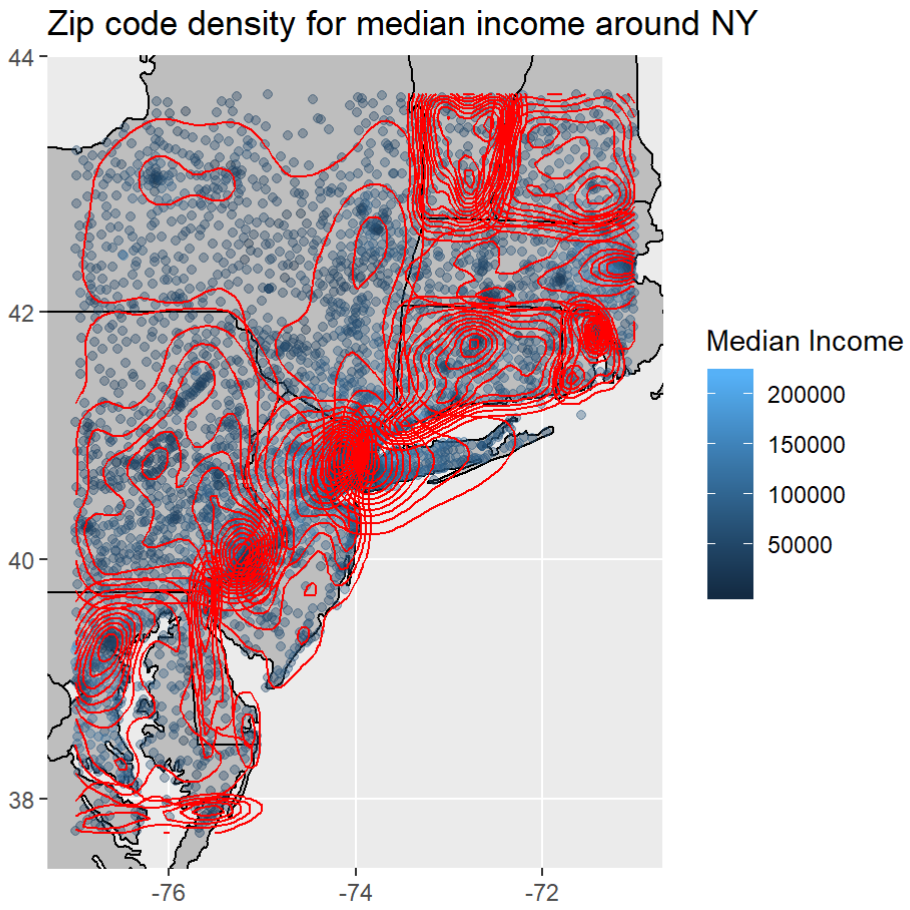

```
# Repeat step 4 for NY data set

# Zip code density using geom_map without geom_point()

ggplot(data=zoom.ny.zip.ip,aes(map_id=state_nm)) +geom_map(map=us,colour="white") + expand_li
mits(x=xlimit_frm_ny,y=ylimit_frm_ny)  + coord_map() +ggtitle("Zip code density for median in
come around NY") + labs(x = "", y = "",color ="Median Income") + geom_density_2d(aes(x=longit
ude,y=latitude,color=median))
```



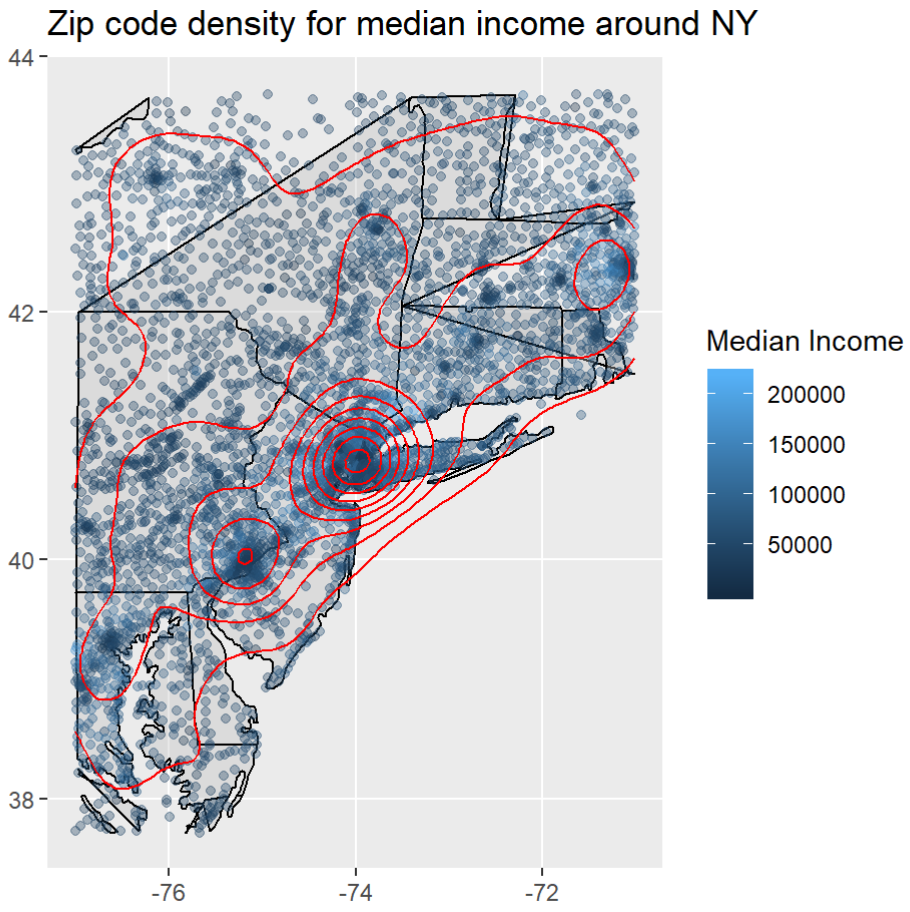

```
ggplot(data=zoom.ny.zip.ip,aes(map_id=state_nm)) +geom_map(map=us,colour="red",fill="white")
 + expand_limits(x=xlimit_frm_ny,y=ylimit_frm_ny)  + coord_map() +ggtitle("Zip code density f
or median income around NY") + labs(x = "", y = "",color ="Median Income") + geom_density_2d
(aes(x=longitude,y=latitude),color="white") + stat_density_2d(aes(x=longitude,y=latitude,fill
= stat(level)),color="black", geom = "polygon")+ scale_fill_continuous(low = "white", high=
"blue", guide = guide_colorbar(title = "Median Income Density"))
```




```
# Zip code density using geom_map with geom_point

ggplot(data=zoom.ny.zip.ip,aes(map_id=state_nm)) +geom_map(map=us,colour="black",fill="gray")
+ expand_limits(x=xlimit_frm_ny,y=ylimit_frm_ny) + coord_map()+ geom_point(aes(x=longitude,y=
latitude,color=median),alpha=0.35) + geom_density_2d(aes(x=longitude,y=latitude),color="red")
+ ggtitle("Zip code density for median income around NY") + labs(x = "", y = "",color ="Media
n Income")
```




```
# Zip code density using geom_polygon

ggplot(data=zoom.ny.zip.ip,aes(x=longitude, y=latitude)) +geom_polygon(data=us,aes(long,lat,g
roup=group),colour="black",fill="gray",alpha=0.35) + xlim(xlimit_frm_ny) +ylim(ylimit_frm_ny)
+ geom_point(aes(color=median),alpha=0.35) + geom_density_2d(color="red")  + labs(x="",y="",c
olor ="Median Income") +  ggtitle("Zip code density for median income around NY") + coord_map
()
```


Zip code density for median income around NY
44
42
40
38
-76
-74
-72
Median Income
200000
150000
100000
50000