

Hw7

Katherine Penney

8/22/2021

```
require(ggplot2)

## Loading required package: ggplot2
require(ggmap)

## Loading required package: 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.
require(readxl)

## Loading required package: readxl
require(gdata)

## Loading required package: gdata
## Warning in system(cmd, intern = intern, wait = wait | intern,
## show.output.on.console = wait, : running command 'C:\WINDOWS\system32\cmd.exe /c
## ftype perl' had status 2

## Warning in system(cmd, intern = intern, wait = wait | intern,
## show.output.on.console = wait, : running command 'C:\WINDOWS\system32\cmd.exe /c
## ftype perl' had status 2

## gdata: read.xls support for 'XLS' (Excel 97-2004) files ENABLED.

##
## gdata: Unable to load perl libaries needed by read.xls()
## gdata: to support 'XLSX' (Excel 2007+) files.

##
## gdata: Run the function 'installXLSXsupport()'
## gdata: to automatically download and install the perl
## gdata: libaries needed to support Excel XLS and XLSX formats.

##
## Attaching package: 'gdata'

## The following object is masked from 'package:stats':
##
##      nobs

## The following object is masked from 'package:utils':
##
```

```

##      object.size
## The following object is masked from 'package:base':
##
##      startsWith
require(dplyr)

## Loading required package: dplyr
##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:gdata':
##
##      combine, first, last
## The following objects are masked from 'package:stats':
##
##      filter, lag
## The following objects are masked from 'package:base':
##
##      intersect, setdiff, setequal, union
library(mapproj)

## Loading required package: maps
library(usdata)

## Warning: package 'usdata' was built under R version 4.1.1
library(usmap)

## Warning: package 'usmap' was built under R version 4.1.1
library(maptools)

## Loading required package: sp
## Checking rgeos availability: FALSE
##      Note: when rgeos is not available, polygon geometry      computations in maptools depend on gpclib
##      which has a restricted licence. It is disabled by default;
##      to enable gpclib, type gpclibPermit()
library(rgdal)

## rgdal: version: 1.5-23, (SVN revision 1121)
## Geospatial Data Abstraction Library extensions to R successfully loaded
## Loaded GDAL runtime: GDAL 3.2.1, released 2020/12/29
## Path to GDAL shared files: C:/Users/klein/Documents/R/win-library/4.1/rgdal/gdal
## GDAL binary built with GEOS: TRUE
## Loaded PROJ runtime: Rel. 7.2.1, January 1st, 2021, [PJ_VERSION: 721]
## Path to PROJ shared files: C:/Users/klein/Documents/R/win-library/4.1/rgdal/proj
## PROJ CDN enabled: FALSE
## Linking to sp version:1.4-5
## To mute warnings of possible GDAL/OSR exportToProj4() degradation,
## use options("rgdal_show_exportToProj4_warnings"="none") before loading rgdal.
## Overwritten PROJ_LIB was C:/Users/klein/Documents/R/win-library/4.1/rgdal/proj

```

```

install.packages("C:/Users/klein/Downloads/zipcode_1.0.tar.gz", repos=NULL, type="source")

## Installing package into 'C:/Users/klein/Documents/R/win-library/4.1'
## (as 'lib' is unspecified)

#Step 1: Load the Data ## Step 1.1 Read the data
data<- read_excel("C:/Users/klein/Downloads/MedianZIP_2_2_2.xls")

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

head(data)

## # A tibble: 6 x 4
##   `Data from: http://www.psc.isr.umich.edu/dis/c` ...2      ...3      ...4
##   <chr>          <chr>      <chr>      <chr>
## 1 Zip            Median     Mean       Pop
## 2 1001           56662.57349~ 66687.7508~ 16445
## 3 1002           49853.41769~ 75062.6343~ 28069
## 4 1003           28462        35121      8491
## 5 1005           75423        82442      4798
## 6 1007           79076.35400~ 85801.9750~ 12962

```

Step 1.2: Update column names

```
colnames(data) <- c('zip', 'median', 'mean', 'population')
```

Step 1.3 and 1.4

```

library(zipcode)
data$zip <- clean.zipcodes(data$zip)
data$median<-as.numeric(data$median)

## Warning: NAs introduced by coercion
data$mean<- as.numeric(data$mean)

## Warning: NAs introduced by coercion
data$population <- as.numeric(data$population)

## Warning: NAs introduced by coercion
data("zipcode")
mergedata <- merge(data, zipcode, by="zip")
mergedata$median <- as.numeric(gsub(", ", "", mergedata$median))
mergedata$population <- as.numeric(gsub(", ", "", mergedata$population))
head(mergedata)

##      zip median    mean population      city state latitude longitude
## 1 01001 56662.57 66687.75     16445 Agawam    MA 42.07061 -72.62029
## 2 01002 49853.42 75062.63     28069 Amherst   MA 42.37765 -72.50323
## 3 01003 28462.00 35121.00     8491 Amherst   MA 42.36956 -72.63599
## 4 01005 75423.00 82442.00     4798 Barre     MA 42.41209 -72.10443

```

```
## 5 01007 79076.35 85801.98      12962 Belchertown    MA 42.27842 -72.41100
## 6 01008 63980.00 78391.00      1244   Blandford    MA 42.17431 -72.94828
```

Step 1.5: Remove Hawaii and Alaska

```
mergedata = mergedata %>%
  filter(state != "HI" & state != "AK")
```

Step 2: Show the income & population per state

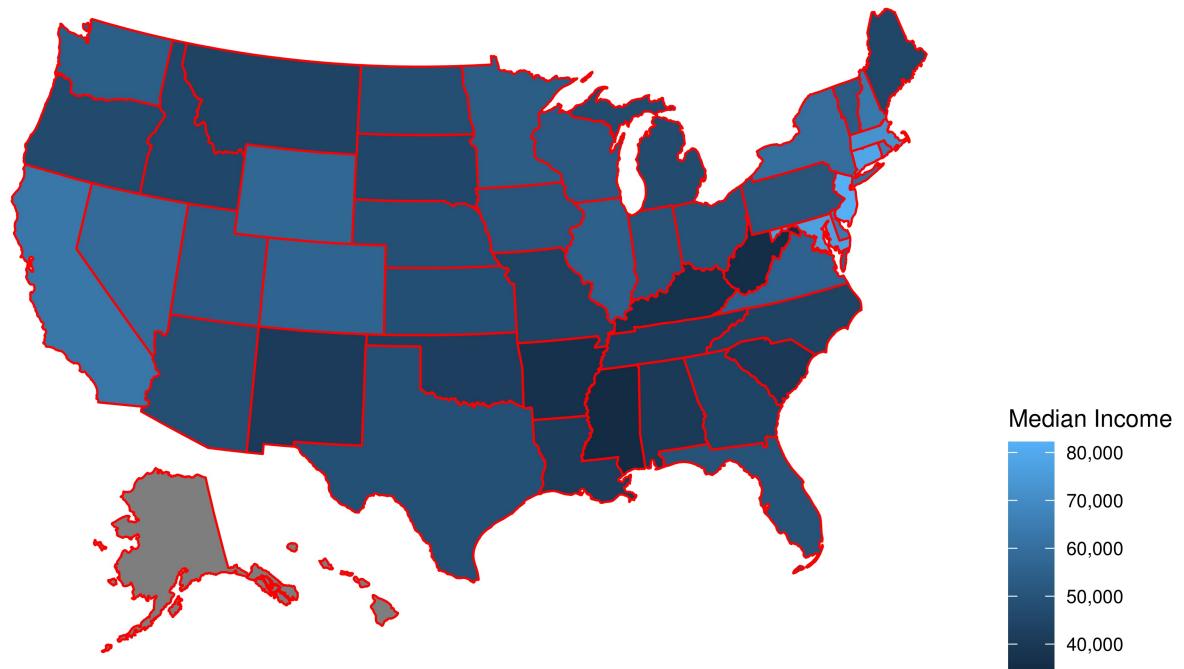
Step 2.1:

```
simpleDf = mergedata %>%
  group_by(state) %>%
  summarise(
    average_median = mean(median, na.rm = TRUE),
    average_population = mean(population, na.rm = TRUE)
  )
```

Step 2.3: Show map representing the color with the average median income of state

```
#GRAPH1
plot_usmap(data = simpleDf, values = "average_median", color = "red") +
  scale_fill_continuous(name = "Median Income", label = scales::comma) +
  ggtitle("Average Median Income per US states (Except Hawaii & Alaska)")+
  theme(legend.position = "right")
```

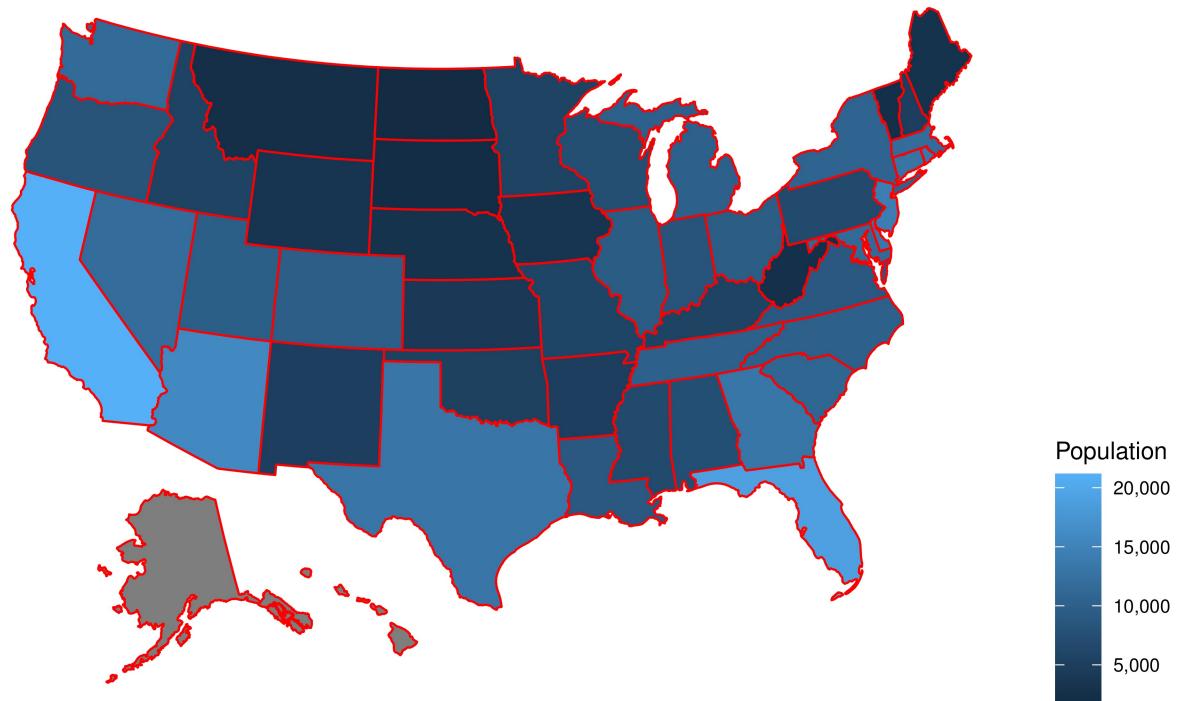
Average Median Income per US states (Except Hawaii & Alaska)



Step 2.4: Create map woth color representing the population of state

```
#GRAPH2
plot_usmap(data = simpleDf, values = "average_population", color = "red") +
  scale_fill_continuous(name = "Population", label = scales::comma) +
  ggtitle("Average Population per US states (Except Hawaii & Alaska)") +
  theme(legend.position = "right")
```

Average Population per US states (Except Hawaii & Alaska)



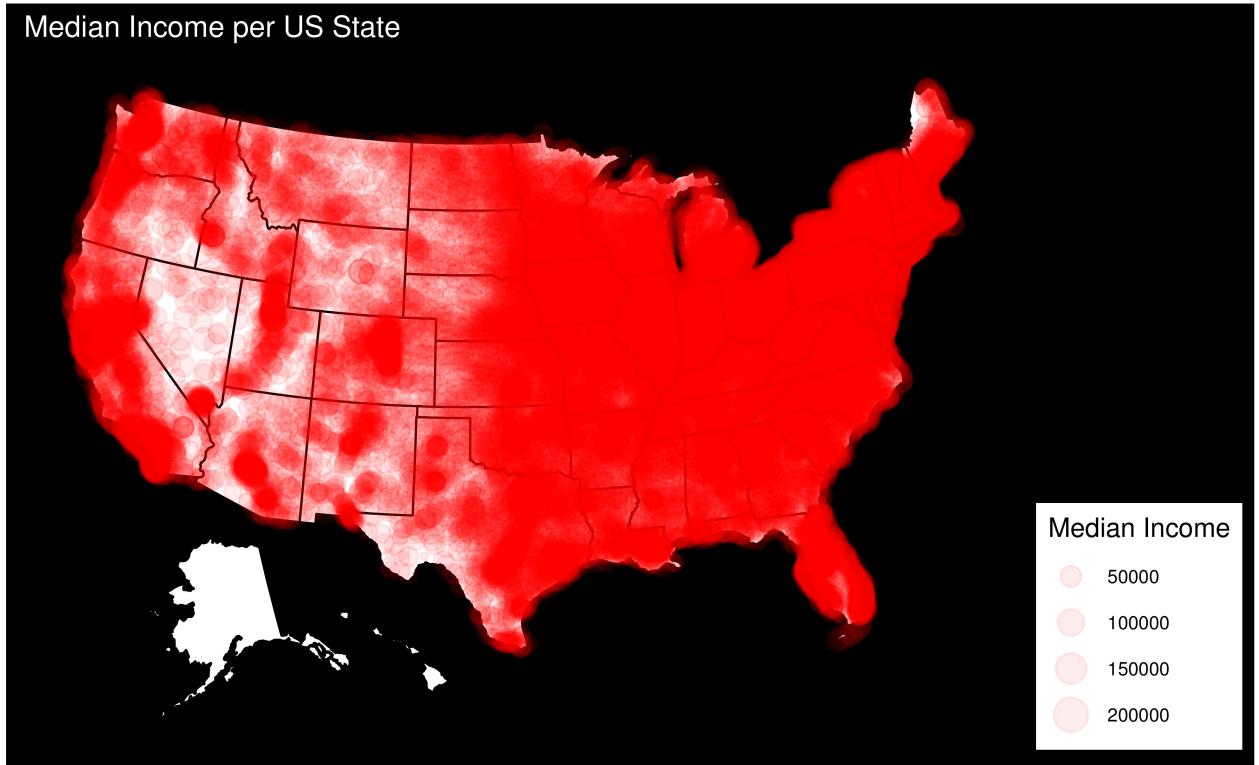
Step 3: Show the income per zip code

Step 3.1

```
#GRAPH3
mergedata <- mergedata %>% relocate(latitude, .before = zip)
mergedata <- mergedata %>% relocate(longitude, .before = latitude)
mergedata_transformed <- usmap_transform(mergedata)

## Warning in showSRID(uprojargs, format = "PROJ", multiline = "NO", prefer_proj =
## prefer_proj): Discarded datum unknown in Proj4 definition

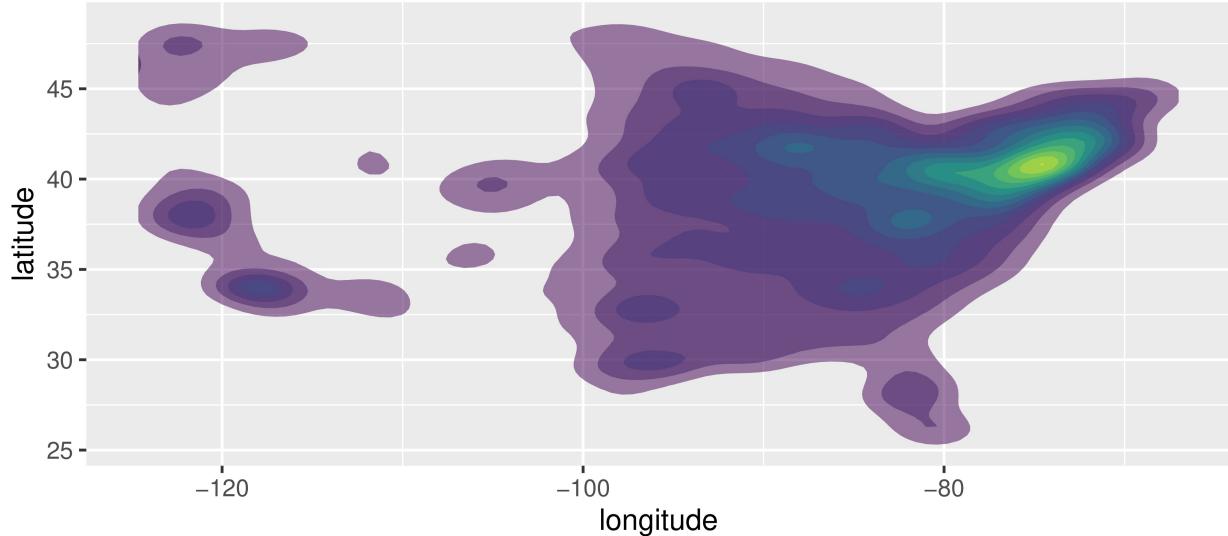
plot_usmap() +
  geom_point(data = mergedata_transformed, aes(x = longitude.1, y = latitude.1, size = median),
  color = "red", alpha = 0.08) +
  ggtitle("Median Income per US State")+
  labs(size = "Median Income")+
  theme(legend.position = "right",
  legend.title = element_text(color = "black", size = 10),
  title = element_text(colour = "white"),
  plot.background = element_rect(fill = "black"))
```



Step 4: Show Zip Code Density

```
ggplot(mergedata, aes(x = longitude, y = latitude)) +  
  coord_equal() +  
  ggtitle("Density Plot using Stat Density2D function") +  
  xlab('longitude') +  
  ylab('latitude') +  
  stat_density2d(aes(fill = ..level..), alpha = .5,  
  geom = "polygon", data = mergedata) +  
  scale_fill_viridis_c() +  
  theme(legend.position = 'none')
```

Density Plot using Stat Density2D function

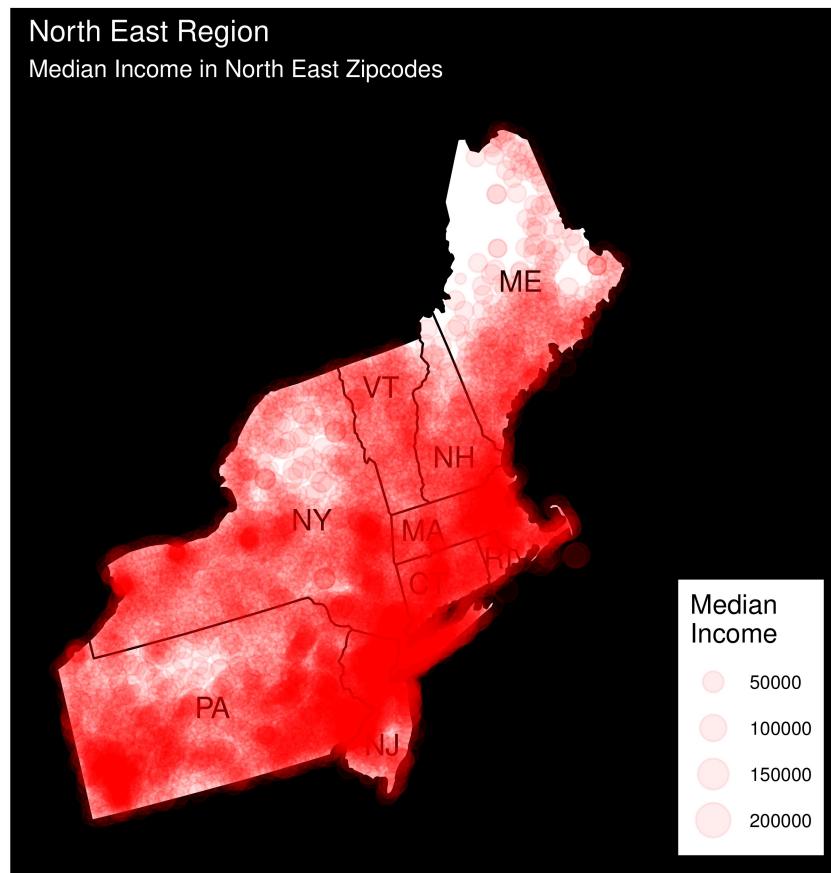


```
# Only Northeast region
```

Step 5: Zoom in to the region around NYC

```
#Graph 5
north_east <- mergedata %>%
  filter(state=="NY" | state=="RI" | state == "CT" | state=="NJ" |
  state=="PA" | state == "NH" | state=="VT" | state == "MA" |
  state=="ME")
north_east_transformed <- usmap_transform(north_east)

## Warning in showSRID(uprojargs, format = "PROJ", multiline = "NO", prefer_proj =
## prefer_proj): Discarded datum unknown in Proj4 definition
plot_usmap(include = .northeast_region, labels = TRUE) +
  geom_point(data = north_east_transformed, aes(x = longitude.1, y = latitude.1, size = median),
  color = "red", alpha = 0.08) +
  labs(title = "North East Region", subtitle = "Median Income in North East Zipcodes",size = "Median
Income") +
  theme(legend.position = "right",
  legend.title = element_text(color = "black", size = 10),
  title = element_text(colour = "white"),
  plot.background = element_rect(fill = "black"))
```



```
#Graph 6
plot_usmap(data = mergedata, values = "median", include = .northeast_region, color = "red", labels = TRUE) +
  scale_fill_continuous(low = "white", high = "red", name = "Population", label = scales::comma) +
  labs(title = "North East Region", subtitle = "Median Income in North East States", size = "Median Income") +
  theme(legend.position = "right",
        legend.title = element_text(color = "black", size = 10),
        title = element_text(colour = "white"),
        plot.background = element_rect(fill = "black"))
```

