**Objective:**

Make maps of the states in U.S. & plot state-level income shares for different years to statistically analyze and visualize the differential wealth distribution using R. This helped to identify regions of stagnant growth.

The first map should plot state level income share of the top 1% of income earners in 2012. The second map should plot the same variable but for 1999.The third map should plot the 2012--‐1999 difference of the variable.

Visit the website:

<http://www.shsu.edu/eco_mwf/inequality.html>

To download the data and learn about the variable of interest.

**Implementation Details:**

Making a map essentially involves three steps --- getting a shape file for map, getting the data, and plotting.

Much of what you would see below has been learnt from:

(a) http://www.kevjohnson.org/making-maps-in-r/ for a general introduction map making in R

(b) https://www.nceas.ucsb.edu/~frazier/RSpatialGuides/colorPaletteCheatsheet.pdf for an introduction to colors in R

See also:

https://www.nceas.ucsb.edu/~frazier/RSpatialGuides/ggmap/ggmapCheatsheet.pdf

for an introduction to using Google maps in making maps using using the R package ggmap

**Note:**

Use ?functionname to see help on any unfamiliar functions.

If not already installed, use install.packages("packagename") to first install a package before loading with library command.

1. **Build 1st map to plot state level income share of the top 1% of income earners in 2012**

Used data from file “usa\_top1\_2012.csv”

**R Code:**

#####################################################################################

**# Making a map using R #**

**# Making a map essentially involves three steps --- getting a shape file for map,**

**# getting the data, and plotting.**

**# Make a map to plot state level income share of top 1% of income earners in 2012**

**# Data downloaded from website http://www.shsu.edu/eco\_mwf/inequality.html**

#####################################################################################

library(raster) # to get map shape file

library(ggplot2) # for plotting and miscellaneous things

library(ggmap) # for plotting

library(plyr) # for merging datasets

library(scales) # to get nice looking legends

library(maps) # draw geographical maps

**# Get a shape file of states of USA**

usa.shape <- getData("GADM", country = "USA", level = 1)

**# Plot to see how the map looks (may take a while**)

# plot(usa.shape)

**# Look at the names of states of USA**

# usa.shape$NAME\_1

#[1] "alabama" "alaska" "arizona"

#[4] "arkansas" "california" "colorado"

#[7] "connecticut" "delaware" "district of columbia"

#[10] "florida" "georgia" "hawaii"

#[13] "idaho" "illinois" "indiana"

#[16] "iowa" "kansas" "kentucky"

#[19] "louisiana" "maine" "maryland"

#[22] "massachusetts" "michigan" "minnesota"

#[25] "mississippi" "missouri" "montana"

#[28] "nebraska" "nevada" "new hampshire"

#[31] "new jersey" "new mexico" "new york"

#[34] "north carolina" "north dakota" "ohio"

#[37] "oklahoma" "oregon" "pennsylvania"

#[40] "rhode island" "south carolina" "south dakota"

#[43] "tennessee" "texas" "utah"

#[46] "vermont" "virginia" "washington"

#[49] "west virginia" "wisconsin" "wyoming"

# >

**# To merge population data to the shape file, convert the shape file into a dataframe**

usa.df <- map\_data("state")

**# Check the contents of the data frame**

# str(usa.df)

#'data.frame': 15537 obs. of 6 variables:

# $ long : num -87.5 -87.5 -87.5 -87.5 -87.6 ...

#$ lat : num 30.4 30.4 30.4 30.3 30.3 ...

#$ group : num 1 1 1 1 1 1 1 1 1 1 ...

#$ order : int 1 2 3 4 5 6 7 8 9 10 ...

#$ region : chr "alabama" "alabama" "alabama" "alabama" ...

#$ subregion: chr NA NA NA NA ...

#>

colnames(usa.df) [5] <- "state"

usa.df$state <- as.factor(usa.df$state)

#str(usa.df)

#'data.frame': 15537 obs. of 6 variables:

# $ long : num -87.5 -87.5 -87.5 -87.5 -87.6 ...

#$ lat : num 30.4 30.4 30.4 30.3 30.3 ...

#$ group : num 1 1 1 1 1 1 1 1 1 1 ...

#$ order : int 1 2 3 4 5 6 7 8 9 10 ...

#$ state : Factor w/ 49 levels "alabama","arizona",..: 1 1 1 1 1 1 1 1 1 1 ...

#$ subregion: chr NA NA NA NA ...

# >

**# Get the state level income share of top 1% data based on 2012 census**

usa.dat <- read.table("usa\_top1\_2012.csv", header = T, sep = ",")

**# Merge the shape data with the population data by state name**

usa.df <- join(usa.df, usa.dat, by = "state", type = "inner")

# str(usa.df)

# 'data.frame': 15537 obs. of 8 variables:

# $ long : num -87.5 -87.5 -87.5 -87.5 -87.6 ...

# $ lat : num 30.4 30.4 30.4 30.3 30.3 ...

# $ group : num 1 1 1 1 1 1 1 1 1 1 ...

# $ order : int 1 2 3 4 5 6 7 8 9 10 ...

# $ state : Factor w/ 49 levels "alabama","arizona",..: 1 1 1 1 1 1 1 1 1 1 ...

# $ subregion: chr NA NA NA NA ...

# $ income : num 16.7 16.7 16.7 16.7 16.7 ...

# $ abb : Factor w/ 51 levels "AK","AL","AR",..: 2 2 2 2 2 2 2 2 2 2 ...

# >

#range(usa.df$income)

# [1] 13.68477 33.00785

# >

**# Divide income into class intervals --- there will be one color for each interval**

**# provide only the upper limits of the intervals (the break points)**

brks <- c(-10,-5,0,5,10,15,20,25,30,35)

**#Create the binding of the abbreviation to the population data**

cnames <- aggregate(cbind(long, lat) ~ abb, data= usa.df,

FUN=function(x)mean(range(x)))

**#Plot map with the abbreviations for each state**

p <- ggplot(usa.df,aes(long,lat)) +

# with borders (slower)

geom\_polygon(data = usa.df, aes(x = long, y = lat, group = group, fill = income),

color = "black", size = 0.15) +

scale\_fill\_distiller(palette = "Reds", breaks = brks, trans = "reverse") +

theme\_nothing(legend = TRUE) +

labs(title = "Top 1% income earners in 2012 in USA", fill = "") +

geom\_text(data=cnames, aes(long, lat, label = abb),size = 2)

**# Note: we are using shades of red for plotting; trans = "reverse" option**

**# Makes the shades go from dark to light as the income percent increases, thus**

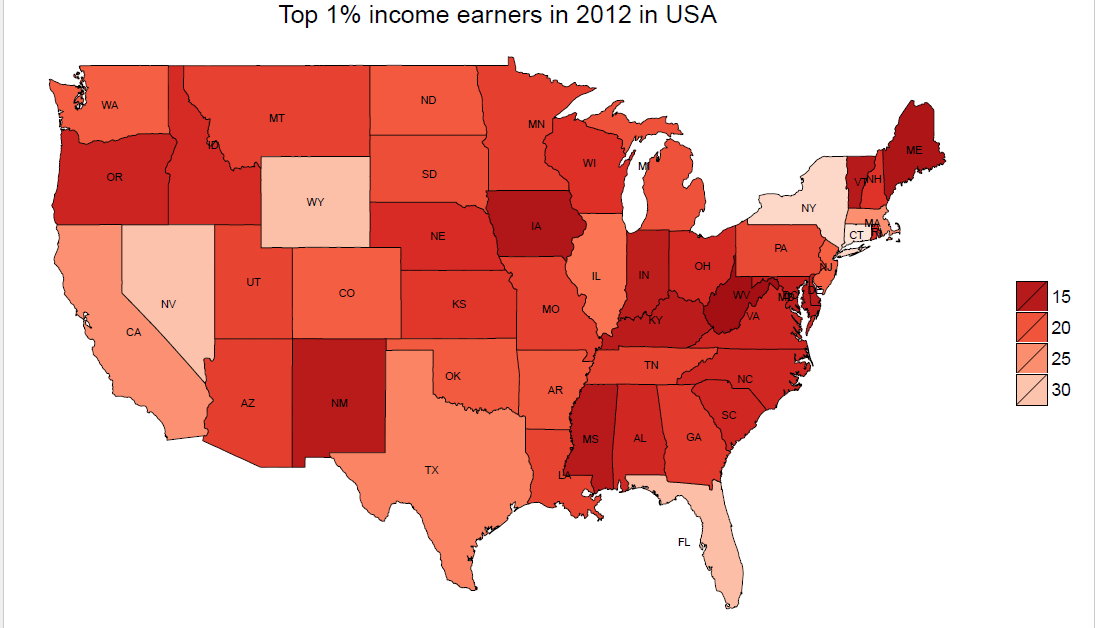
**# ensuring that darkest red = worst case scenario.**

**# Save the map to a file to viewing**

ggsave(p, file = "usa\_income\_map\_2012.pdf")

##################################################################################

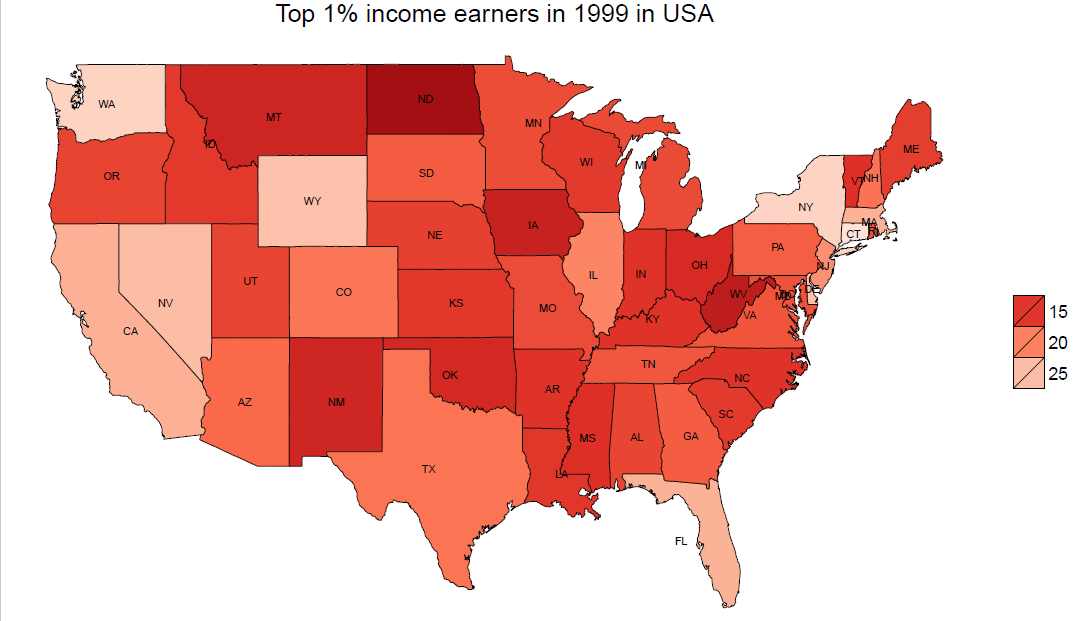
**Result & Analysis:**



The map incorporates the World Top Income Database the top 1% income shares for each US state in 2012. The income percent ranges from 13.68477% for West Virginia to 33.00785% for Connecticut. We are using shades of red for plotting. The shades go from dark to light as the income percent increases, thus ensuring that darkest red as the worst case scenario. Income percent below 15% have the darkest shade and income percent above 30% has the lightest shade in the map.

1. **Second map to plot state level income share of top 1% income earners in 1999**

Used data from file “usa\_top1\_1999.csv”

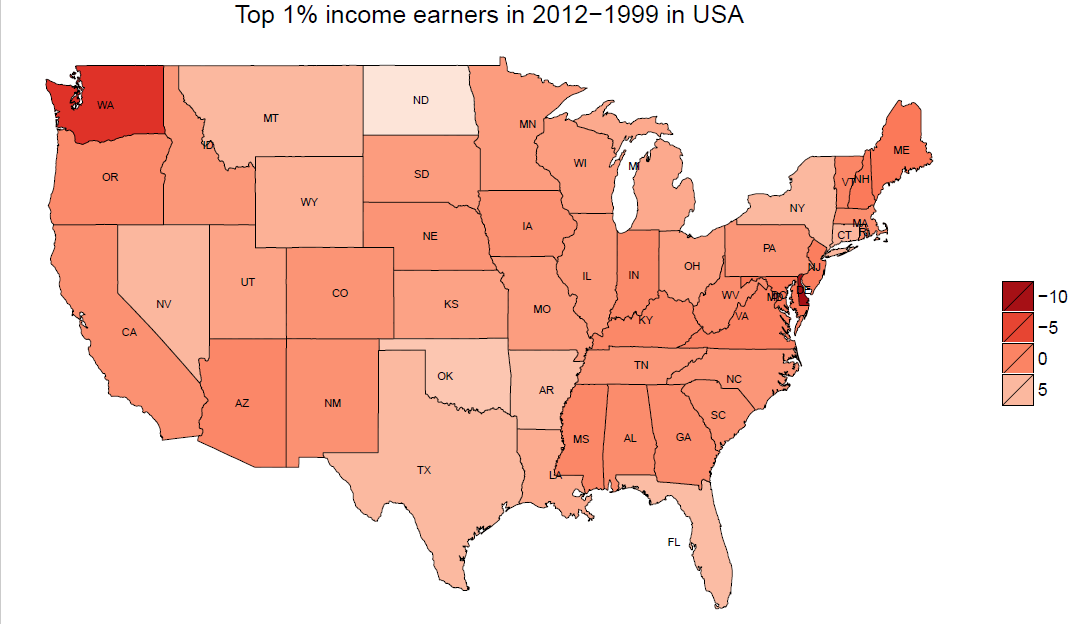


The map incorporates the World Top Income Database the top 1% income shares for each US state in 1999. The income percent ranges from 11.24333% for North Dakota to 28.15289% for Connecticut. We are using shades of red for plotting. The shades go from dark to light as the income percent increases, thus ensuring that darkest red as the worst case scenario as income percent is lowest. All income percent less than 15% have the darkest shade.

Incomes between 15% to 20% have the medium shade and incomes 20% and higher have the lightest shade in the map.

1. **Third map to plot the difference in 2012 and 1999 of state level income share of top 1% income earners**

Used data from file “usa\_top1\_2012-1999.csv”



This map is used to plot the difference in 2012 and 1999 of state level income share of top 1% income earners. We can see the difference ranges from -10.1696% for Delaware to 9.027267% for North Dakota. All income percent between -10% and -5% have the darkest shade and all income percent above 5% has the lightest shade.