Alonso\_Week 7 Homework Assignment

Martin Alonso

8/30/2018

IST687 Introduction to Data Science: Week 7 Homework.

Map (S)mashing

### Part 1: The Code

# We're loading all the packages that will be needed for this exercise.  
require(ggplot2)  
require(ggmap)  
require(readxl)  
require(gdata)  
require(dplyr)  
require(zipcode)  
  
# Step 1: Load the Data.   
# Load the MedianZIP\_2\_2\_2.xlsx file.  
dat <- read\_excel('MedianZIP\_2\_2\_2.xlsx')  
  
# Let's take a quick survey of the data.   
str(dat)  
head(dat)  
summary(dat)  
  
# Looks like the column names will need to be updated and the second to fourth column will need to be converted to numerical type.  
names <- dat[1, ]  
colnames(dat) <- names  
dat <- dat[-1, ]  
  
numdat <- colnames(dat[, 2:4])  
for(i in numdat){  
 dat[[i]] <- as.numeric(dat[[i]])  
 dat[[i]][is.na(dat[[i]])] <- median(dat[[i]], na.rm = T)  
}  
  
# Let's check once again that the data is correct.  
str(dat)  
  
# Now let's upload the zipcode data, merge the data, and remove Alaska and Hawaii.  
data(zipcode)  
dat$zip <- clean.zipcodes(dat$Zip)  
  
zip\_comp <- merge(dat, zipcode, by = 'zip')  
zip\_comp <- zip\_comp %>%  
 filter(!state %in% c('AK', 'HI'))  
  
  
#Step 2: Show income & population per state.  
med\_dat <- zip\_comp %>%  
 group\_by(state) %>%   
 summarize(mean\_med\_income = mean(Median),   
 tot\_pop = sum(Pop))   
  
# Add the state names and convert to lower case.   
med\_dat$stateName <- state.name[match(med\_dat$state, state.abb)]  
med\_dat$stateName <- tolower(med\_dat$stateName)  
  
us <- map\_data("state")  
  
# Time to map average median income by state. We'll color code the map according to the median income.   
ggplot(med\_dat, aes(map\_id = stateName)) +   
 geom\_map(map = us, aes(fill = med\_dat$mean\_med\_income)) +   
 expand\_limits(x = us$long, y = us$lat) +  
 coord\_map() +  
 labs(x = 'Long', y = 'Lat', fill = 'Income') +   
 ggtitle('Avg median income by state')  
  
# And now we'll do the same by population.   
ggplot(med\_dat, aes(map\_id = stateName)) +   
 geom\_map(map = us, aes(fill = med\_dat$tot\_pop)) +   
 expand\_limits(x = us$long, y = us$lat) +  
 coord\_map() +  
 labs(x = 'Long', y = 'Lat', fill = 'Population') +   
 ggtitle('Population by state')  
  
  
# Step 3: Show income per zipcode.  
zip\_comp$stateName <- state.name[match(zip\_comp$state, state.abb)]  
zip\_comp$stateName <- tolower(zip\_comp$stateName)  
  
ggplot(zip\_comp, aes(map\_id = stateName)) +  
 geom\_map(map = us, fill = 'black', colour = 'white') +   
 expand\_limits(x = us$long, y = us$lat) +  
 geom\_point(data = zip\_comp, aes(zip\_comp$longitude, zip\_comp$latitude, color = zip\_comp$Median)) +  
 coord\_map() +  
 labs(x = 'Long', y = 'Lat', fill = 'Income') +   
 ggtitle('Median Income per Zipcode')  
  
# Step 4: Show zipcode density.  
ggplot(zip\_comp, aes(map\_id = stateName)) +  
 geom\_map(map = us, fill = 'black', colour = 'white') +   
 expand\_limits(x = us$long, y = us$lat) +  
 stat\_density2d(data = zip\_comp, aes(zip\_comp$longitude, zip\_comp$latitude)) +  
 coord\_map() +  
 labs(x = 'Long', y = 'Lat') +   
 ggtitle('Zipcode Density')  
  
# Step 5: Zoom in to the region around NYC.

### Part 2: Running the Code