# Module 4 - Assignment 1

## Agyabeng Prince

### Data Transformation

library(tidyverse)

## Warning: package 'tidyverse' was built under R version 3.4.4

## -- Attaching packages ------------------------------------------- tidyverse 1.2.1 --

## v ggplot2 2.2.1 v purrr 0.2.4  
## v tibble 1.4.1 v dplyr 0.7.5  
## v tidyr 0.8.0 v stringr 1.3.1  
## v readr 1.1.1 v forcats 0.2.0

## Warning: package 'dplyr' was built under R version 3.4.4

## Warning: package 'stringr' was built under R version 3.4.4

## -- Conflicts ---------------------------------------------- tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library(readr)  
state\_income <- read\_csv("state\_income.csv")

## Parsed with column specification:  
## cols(  
## id = col\_integer(),  
## State\_Code = col\_integer(),  
## State\_Name = col\_character(),  
## State\_ab = col\_character(),  
## County = col\_character(),  
## City = col\_character(),  
## Place = col\_character(),  
## Type = col\_character(),  
## Primary = col\_character(),  
## Zip\_Code = col\_integer(),  
## Area\_Code = col\_integer(),  
## ALand = col\_double(),  
## AWater = col\_double(),  
## Lat = col\_double(),  
## Lon = col\_double(),  
## Mean = col\_integer(),  
## Median = col\_integer(),  
## Stdev = col\_integer()  
## )

## Warning in rbind(names(probs), probs\_f): number of columns of result is not  
## a multiple of vector length (arg 2)

## Warning: 1 parsing failure.  
## row # A tibble: 1 x 5 col row col expected actual file expected <int> <chr> <chr> <chr> <chr> actual 1 27548 Area\_Code an integer M 'state\_income.csv' file # A tibble: 1 x 5

View(state\_income)

#### State Incomes

I will be creating a subset of data from the file

State\_income2 <- select(state\_income,State\_Name,State\_ab,County,City,Type, ALand,Mean,Median, Stdev)  
  
State\_income2%>%select(State\_ab, everything())

## # A tibble: 32,526 x 9  
## State\_ab State\_Name County City Type ALand Mean Median Stdev  
## <chr> <chr> <chr> <chr> <chr> <dbl> <int> <int> <int>  
## 1 AL Alabama Mobile Co~ Chickas~ City 1.09e7 38773 30506 33101  
## 2 AL Alabama Barbour C~ Louisvi~ City 2.61e7 37725 19528 43789  
## 3 AL Alabama Shelby Co~ Columbi~ City 4.48e7 54606 31930 57348  
## 4 AL Alabama Mobile Co~ Satsuma City 3.69e7 63919 52814 47707  
## 5 AL Alabama Mobile Co~ Dauphin~ Town 1.62e7 77948 67225 54270  
## 6 AL Alabama Cullman C~ Cullman Town 8.91e6 50715 42643 35886  
## 7 AL Alabama Escambia ~ East Br~ City 8.83e6 33737 23610 28256  
## 8 AL Alabama Elmore Co~ Coosada Town 1.02e7 46319 40242 38941  
## 9 AL Alabama Morgan Co~ Eva Town 1.05e7 57994 39591 47235  
## 10 AL Alabama Talladega~ Sylacau~ CDP 4.52e7 54807 41712 51359  
## # ... with 32,516 more rows

head(State\_income2, n = 10)

## # A tibble: 10 x 9  
## State\_Name State\_ab County City Type ALand Mean Median Stdev  
## <chr> <chr> <chr> <chr> <chr> <dbl> <int> <int> <int>  
## 1 Alabama AL Mobile Co~ Chickas~ City 1.09e7 38773 30506 33101  
## 2 Alabama AL Barbour C~ Louisvi~ City 2.61e7 37725 19528 43789  
## 3 Alabama AL Shelby Co~ Columbi~ City 4.48e7 54606 31930 57348  
## 4 Alabama AL Mobile Co~ Satsuma City 3.69e7 63919 52814 47707  
## 5 Alabama AL Mobile Co~ Dauphin~ Town 1.62e7 77948 67225 54270  
## 6 Alabama AL Cullman C~ Cullman Town 8.91e6 50715 42643 35886  
## 7 Alabama AL Escambia ~ East Br~ City 8.83e6 33737 23610 28256  
## 8 Alabama AL Elmore Co~ Coosada Town 1.02e7 46319 40242 38941  
## 9 Alabama AL Morgan Co~ Eva Town 1.05e7 57994 39591 47235  
## 10 Alabama AL Talladega~ Sylacau~ CDP 4.52e7 54807 41712 51359

State\_income2 <- rename(State\_income2, SquareArea = ALand, IncomeMean = Mean, IncomeMedian = Median, IncomeStDev = Stdev)  
head(State\_income2)

## # A tibble: 6 x 9  
## State\_Name State\_ab County City Type SquareArea IncomeMean  
## <chr> <chr> <chr> <chr> <chr> <dbl> <int>  
## 1 Alabama AL Mobile Coun~ Chickasaw City 10894952 38773  
## 2 Alabama AL Barbour Cou~ Louisville City 26070325 37725  
## 3 Alabama AL Shelby Coun~ Columbiana City 44835274 54606  
## 4 Alabama AL Mobile Coun~ Satsuma City 36878729 63919  
## 5 Alabama AL Mobile Coun~ Dauphin Is~ Town 16204185 77948  
## 6 Alabama AL Cullman Cou~ Cullman Town 8913021 50715  
## # ... with 2 more variables: IncomeMedian <int>, IncomeStDev <int>

NC\_Income <- filter(State\_income2, State\_Name == "North Carolina")

## Warning: package 'bindrcpp' was built under R version 3.4.4

head(NC\_Income, n = 10)

## # A tibble: 10 x 9  
## State\_Name State\_ab County City Type SquareArea IncomeMean  
## <chr> <chr> <chr> <chr> <chr> <dbl> <int>  
## 1 North Carol~ NC Alamance C~ Elon CDP 3515396 89973  
## 2 North Carol~ NC Johnston C~ Wendell Town 23956770 67438  
## 3 North Carol~ NC Sampson Co~ Stedman Town 1353212 43538  
## 4 North Carol~ NC Henderson ~ Henderso~ CDP 2625120 38120  
## 5 North Carol~ NC Beaufort C~ Pinetown Town 4121722 30468  
## 6 North Carol~ NC Davie Coun~ Clemmons Town 5903422 97561  
## 7 North Carol~ NC Bladen Cou~ Bladenbo~ Town 5737410 38588  
## 8 North Carol~ NC Sampson Co~ Clinton CDP 8562785 34778  
## 9 North Carol~ NC Lee County Broadway Town 3350431 60384  
## 10 North Carol~ NC Guilford C~ Burlingt~ City 75533002 54337  
## # ... with 2 more variables: IncomeMedian <int>, IncomeStDev <int>

#### NC Incomes

will be using the NC\_income dataset to create summaries of the incomes within North Carolina including summaries by county, city and type.

NC\_Income\_County <- arrange(NC\_Income, County)  
head(NC\_Income\_County, n = 10)

## # A tibble: 10 x 9  
## State\_Name State\_ab County City Type SquareArea IncomeMean  
## <chr> <chr> <chr> <chr> <chr> <dbl> <int>  
## 1 North Carol~ NC Alamance C~ Elon CDP 3515396 89973  
## 2 North Carol~ NC Alamance C~ Mebane City 23213152 67397  
## 3 North Carol~ NC Alamance C~ Henderso~ Track 12734435 57073  
## 4 North Carol~ NC Alamance C~ Ahoskie Track 199246026 54071  
## 5 North Carol~ NC Alamance C~ Red Spri~ Track 93319263 30673  
## 6 North Carol~ NC Alamance C~ Statesvi~ Track 10829691 40174  
## 7 North Carol~ NC Alamance C~ Supply Track 29875162 45625  
## 8 North Carol~ NC Alamance C~ Statesvi~ Track 37718022 55177  
## 9 North Carol~ NC Alamance C~ Mooresvi~ Track 13853696 106274  
## 10 North Carol~ NC Alamance C~ Mooresvi~ Track 7037037 93463  
## # ... with 2 more variables: IncomeMedian <int>, IncomeStDev <int>

summary1 <- group\_by(NC\_Income,County)  
summary1 <- summarise (summary1, mean= mean(IncomeMean))  
summary2 <- NC\_Income %>%  
group\_by(City) %>%  
summarise (mean = mean(IncomeMean))

The new data contains mean data income for the counties and mean income for the cities for North Carolina

summary3 <- group\_by(NC\_Income,Type)  
summary3 <- summarise (summary3, mean= mean(IncomeMean))

#### Income Visualization

library("tidyverse")  
  
ggplot(summary1,aes(County, mean))+  
 geom\_point()+theme(axis.text.x = element\_text(angle = 90,vjust =0.5, hjust=1))

