# Module 4 - Assignment 1

## Angelo, Elizabeth

### Data Transformation

library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.1 ──

## ✔ ggplot2 3.3.5 ✔ purrr 0.3.4  
## ✔ tibble 3.1.6 ✔ dplyr 1.0.7  
## ✔ tidyr 1.1.4 ✔ stringr 1.4.0  
## ✔ readr 2.1.1 ✔ forcats 0.5.1

## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()

state\_income <- read.csv("~/Desktop/MIS 503/Module 4/state\_income.csv")

#### State Incomes

I will be using the state\_income dataset to create a second dataset “state\_income2”, using the following variables: State\_Name (state name), State\_ab (state abbreviation), County, City, Type (city, town, CDP), ALand (sq area of land), Mean (mean household income), Median (median household income), and Stdev (standard deviation of household income).

library(dplyr)  
state\_income2 <- state\_income %>%  
 select(State\_Name, State\_ab, County, City, Type, ALand, Mean, Median, Stdev) %>%  
 select(State\_ab, everything())  
head(state\_income2, 10)

## State\_ab State\_Name County City Type ALand Mean  
## 1 AL Alabama Mobile County Chickasaw City 10894952 38773  
## 2 AL Alabama Barbour County Louisville City 26070325 37725  
## 3 AL Alabama Shelby County Columbiana City 44835274 54606  
## 4 AL Alabama Mobile County Satsuma City 36878729 63919  
## 5 AL Alabama Mobile County Dauphin Island Town 16204185 77948  
## 6 AL Alabama Cullman County Cullman Town 8913021 50715  
## 7 AL Alabama Escambia County East Brewton City 8826252 33737  
## 8 AL Alabama Elmore County Coosada Town 10222339 46319  
## 9 AL Alabama Morgan County Eva Town 10544874 57994  
## 10 AL Alabama Talladega County Sylacauga CDP 45178321 54807  
## Median Stdev  
## 1 30506 33101  
## 2 19528 43789  
## 3 31930 57348  
## 4 52814 47707  
## 5 67225 54270  
## 6 42643 35886  
## 7 23610 28256  
## 8 40242 38941  
## 9 39591 47235  
## 10 41712 51359

state\_income2 <- state\_income2 %>%  
 rename(SquareArea = ALand,  
 IncomeMean = Mean,  
 IncomeMedian = Median,  
 IncomeStDev = Stdev)  
head(state\_income2, 10)

## State\_ab State\_Name County City Type SquareArea  
## 1 AL Alabama Mobile County Chickasaw City 10894952  
## 2 AL Alabama Barbour County Louisville City 26070325  
## 3 AL Alabama Shelby County Columbiana City 44835274  
## 4 AL Alabama Mobile County Satsuma City 36878729  
## 5 AL Alabama Mobile County Dauphin Island Town 16204185  
## 6 AL Alabama Cullman County Cullman Town 8913021  
## 7 AL Alabama Escambia County East Brewton City 8826252  
## 8 AL Alabama Elmore County Coosada Town 10222339  
## 9 AL Alabama Morgan County Eva Town 10544874  
## 10 AL Alabama Talladega County Sylacauga CDP 45178321  
## IncomeMean IncomeMedian IncomeStDev  
## 1 38773 30506 33101  
## 2 37725 19528 43789  
## 3 54606 31930 57348  
## 4 63919 52814 47707  
## 5 77948 67225 54270  
## 6 50715 42643 35886  
## 7 33737 23610 28256  
## 8 46319 40242 38941  
## 9 57994 39591 47235  
## 10 54807 41712 51359

NC\_income <- state\_income2 %>%  
 filter(State\_Name == "North Carolina")  
head(NC\_income, 10)

## State\_ab State\_Name County City Type SquareArea  
## 1 NC North Carolina Alamance County Elon CDP 3515396  
## 2 NC North Carolina Johnston County Wendell Town 23956770  
## 3 NC North Carolina Sampson County Stedman Town 1353212  
## 4 NC North Carolina Henderson County Hendersonville CDP 2625120  
## 5 NC North Carolina Beaufort County Pinetown Town 4121722  
## 6 NC North Carolina Davie County Clemmons Town 5903422  
## 7 NC North Carolina Bladen County Bladenboro Town 5737410  
## 8 NC North Carolina Sampson County Clinton CDP 8562785  
## 9 NC North Carolina Lee County Broadway Town 3350431  
## 10 NC North Carolina Guilford County Burlington City 75533002  
## IncomeMean IncomeMedian IncomeStDev  
## 1 89973 300000 44245  
## 2 67438 300000 38524  
## 3 43538 25196 48097  
## 4 38120 31430 24810  
## 5 30468 17951 32450  
## 6 97561 80720 72495  
## 7 38588 20838 40643  
## 8 34778 23603 23077  
## 9 60384 52298 46462  
## 10 54337 300000 29951

#### NC Incomes

I 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 <- NC\_income %>%  
 arrange(County)  
head(NC\_income, 10)

## State\_ab State\_Name County City Type SquareArea  
## 1 NC North Carolina Alamance County Elon CDP 3515396  
## 2 NC North Carolina Alamance County Mebane City 23213152  
## 3 NC North Carolina Alamance County Hendersonville Track 12734435  
## 4 NC North Carolina Alamance County Ahoskie Track 199246026  
## 5 NC North Carolina Alamance County Red Springs Track 93319263  
## 6 NC North Carolina Alamance County Statesville Track 10829691  
## 7 NC North Carolina Alamance County Supply Track 29875162  
## 8 NC North Carolina Alamance County Statesville Track 37718022  
## 9 NC North Carolina Alamance County Mooresville Track 13853696  
## 10 NC North Carolina Alamance County Mooresville Track 7037037  
## IncomeMean IncomeMedian IncomeStDev  
## 1 89973 300000 44245  
## 2 67397 55632 52141  
## 3 57073 41022 47417  
## 4 54071 42038 44806  
## 5 30673 20786 33771  
## 6 40174 27569 33710  
## 7 45625 32324 42467  
## 8 55177 48504 33945  
## 9 106274 83085 75605  
## 10 93463 79991 61872

summary1 <- group\_by(NC\_income, County)  
 summary1 <- summarise(summary1,mean = mean(IncomeMean))  
print(summary1)

## # A tibble: 49 × 2  
## County mean  
## <chr> <dbl>  
## 1 Alamance County 58430.  
## 2 Anson County 36559.  
## 3 Avery County 41915   
## 4 Beaufort County 40029.  
## 5 Bladen County 35796   
## 6 Brunswick County 53794.  
## 7 Burke County 50283.  
## 8 Caldwell County 50050   
## 9 Camden County 53950   
## 10 Carteret County 50700   
## # … with 39 more rows

summary2 <- NC\_income %>%  
 group\_by(City) %>%  
 summarise(mean = mean(IncomeMean))  
print(summary2)

## # A tibble: 335 × 2  
## City mean  
## <chr> <dbl>  
## 1 Aberdeen 71839   
## 2 Advance 96650   
## 3 Ahoskie 51360   
## 4 Albemarle 61028.  
## 5 Andrews 39213   
## 6 Angier 58414.  
## 7 Apex 113786.  
## 8 Arapahoe 58334   
## 9 Arden 85641   
## 10 Ash 50636   
## # … with 325 more rows

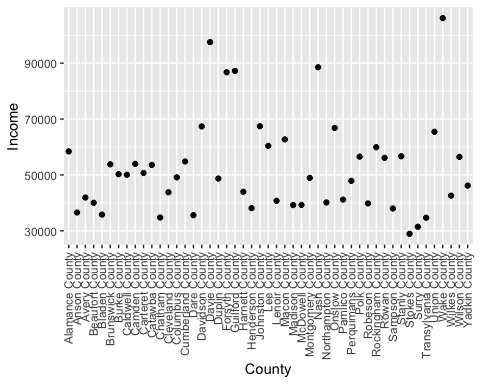
The first dateset shows North Carolina income data in ascending order, containing the 9 variables: state abbreviation, state name, county, city, type, square area, income mean, income median, and income standard deviation. The second dataset contains North Carolina Counties and their income mean. The final dataset contains North Carolina Cities and their income mean.

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

## # A tibble: 5 × 2  
## Type mean  
## <chr> <dbl>  
## 1 CDP 45853.  
## 2 City 55884.  
## 3 Town 53116.  
## 4 Track 58381.  
## 5 Village 0

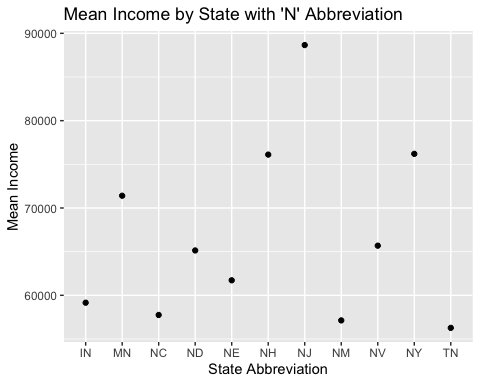
#### Income Visualization

library(ggplot2)  
ggplot(summary1, aes(x = County, y = mean)) +  
 geom\_point() +  
 labs(y = "Income") +  
 theme(axis.text.x = element\_text(angle = 90, vjust = 0.5, hjust = 1))



Wake county has the largest average income, while Stokes County has the lowest average income. Surry County has the second lowest average income.

AvgStateIncome <- state\_income2 %>%  
 group\_by(State\_ab) %>%  
 summarise(mean\_Income = mean(IncomeMean))  
AvgStateIncome\_N <- filter(AvgStateIncome, grepl("N", State\_ab))  
ggplot(AvgStateIncome\_N, aes(x = State\_ab, y = mean\_Income)) +  
 geom\_point() +  
 labs(x = "State Abbreviation", y = "Mean Income", title = "Mean Income by State with 'N' Abbreviation")



New Jersey, New Hampshire, and New York had the highest mean incomes, while Tennessee, New Mexico, and North Carolina, had the lowest mean incomes.