# Module 5 - Assignment 1

## Broach, Stuart

### Data Transformation

library("tidyverse")

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

## v ggplot2 3.1.0 v purrr 0.2.5  
## v tibble 1.4.2 v dplyr 0.7.7  
## v tidyr 0.8.2 v stringr 1.3.1  
## v readr 1.1.1 v forcats 0.3.0

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

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

#### State Incomes

For this document, I will be using a subset of data from the state\_income.csv dataset. This will use the variables State\_Name, State\_ab, County, City, Type, ALand, Mean, MEdian, and Stdev.

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

## # A tibble: 10 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

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

## # A tibble: 10 x 9  
## State\_ab State\_Name County City Type SquareArea IncomeMean  
## <chr> <chr> <chr> <chr> <chr> <dbl> <int>  
## 1 AL Alabama Mobil~ Chic~ City 10894952 38773  
## 2 AL Alabama Barbo~ Loui~ City 26070325 37725  
## 3 AL Alabama Shelb~ Colu~ City 44835274 54606  
## 4 AL Alabama Mobil~ Sats~ City 36878729 63919  
## 5 AL Alabama Mobil~ Daup~ Town 16204185 77948  
## 6 AL Alabama Cullm~ Cull~ Town 8913021 50715  
## 7 AL Alabama Escam~ East~ City 8826252 33737  
## 8 AL Alabama Elmor~ Coos~ Town 10222339 46319  
## 9 AL Alabama Morga~ Eva Town 10544874 57994  
## 10 AL Alabama Talla~ Syla~ CDP 45178321 54807  
## # ... with 2 more variables: IncomeMedian <int>, IncomeStdev <int>

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

## # A tibble: 10 x 9  
## State\_ab State\_Name County City Type SquareArea IncomeMean  
## <chr> <chr> <chr> <chr> <chr> <dbl> <int>  
## 1 NC North Car~ Alama~ Elon CDP 3515396 89973  
## 2 NC North Car~ Johns~ Wend~ Town 23956770 67438  
## 3 NC North Car~ Samps~ Sted~ Town 1353212 43538  
## 4 NC North Car~ Hende~ Hend~ CDP 2625120 38120  
## 5 NC North Car~ Beauf~ Pine~ Town 4121722 30468  
## 6 NC North Car~ Davie~ Clem~ Town 5903422 97561  
## 7 NC North Car~ Blade~ Blad~ Town 5737410 38588  
## 8 NC North Car~ Samps~ Clin~ CDP 8562785 34778  
## 9 NC North Car~ Lee C~ Broa~ Town 3350431 60384  
## 10 NC North Car~ Guilf~ Burl~ City 75533002 54337  
## # ... with 2 more variables: IncomeMedian <int>, IncomeStdev <int>

#### NC Incomes

or our next analysis, we will be using a subset of the state\_income2 data set entitled “NC\_income”. This dataset only looks at the incomes within North Carolina including summaries by county, city and type.

arrange(NC\_income, County)

## # A tibble: 915 x 9  
## State\_ab State\_Name County City Type SquareArea IncomeMean  
## <chr> <chr> <chr> <chr> <chr> <dbl> <int>  
## 1 NC North Car~ Alama~ Elon CDP 3515396 89973  
## 2 NC North Car~ Alama~ Meba~ City 23213152 67397  
## 3 NC North Car~ Alama~ Hend~ Track 12734435 57073  
## 4 NC North Car~ Alama~ Ahos~ Track 199246026 54071  
## 5 NC North Car~ Alama~ Red ~ Track 93319263 30673  
## 6 NC North Car~ Alama~ Stat~ Track 10829691 40174  
## 7 NC North Car~ Alama~ Supp~ Track 29875162 45625  
## 8 NC North Car~ Alama~ Stat~ Track 37718022 55177  
## 9 NC North Car~ Alama~ Moor~ Track 13853696 106274  
## 10 NC North Car~ Alama~ Moor~ Track 7037037 93463  
## # ... with 905 more rows, and 2 more variables: IncomeMedian <int>,  
## # IncomeStdev <int>

head(NC\_income,10)

## # A tibble: 10 x 9  
## State\_ab State\_Name County City Type SquareArea IncomeMean  
## <chr> <chr> <chr> <chr> <chr> <dbl> <int>  
## 1 NC North Car~ Alama~ Elon CDP 3515396 89973  
## 2 NC North Car~ Johns~ Wend~ Town 23956770 67438  
## 3 NC North Car~ Samps~ Sted~ Town 1353212 43538  
## 4 NC North Car~ Hende~ Hend~ CDP 2625120 38120  
## 5 NC North Car~ Beauf~ Pine~ Town 4121722 30468  
## 6 NC North Car~ Davie~ Clem~ Town 5903422 97561  
## 7 NC North Car~ Blade~ Blad~ Town 5737410 38588  
## 8 NC North Car~ Samps~ Clin~ CDP 8562785 34778  
## 9 NC North Car~ Lee C~ Broa~ Town 3350431 60384  
## 10 NC North Car~ Guilf~ Burl~ City 75533002 54337  
## # ... with 2 more variables: IncomeMedian <int>, IncomeStdev <int>

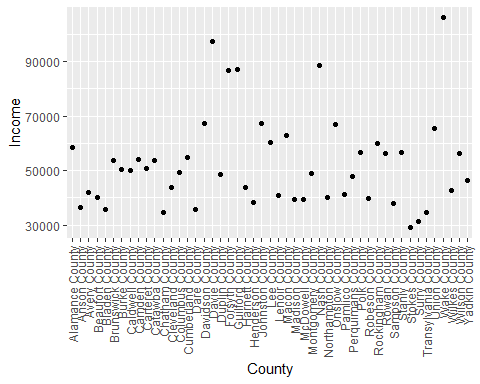
summary1 <- group\_by(NC\_income,County)  
summary1 <- summarise (summary1, mean= mean(IncomeMean))  
summary2 <- NC\_income %>%  
 group\_by(City) %>%  
 summarize (mean = mean(IncomeMean))

Summary1 list the mean of each county in alpabetical order. Summary2 does the exact same thing exact by City instead of County.

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

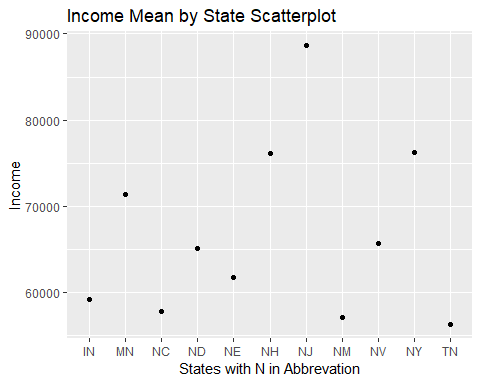
#### Income Visualizatrion

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))



According to the graph, Wake County has the largest average income. Stokes has the lowest average income, with Surry having the second lowest.

AvgStateIncome <- group\_by(state\_income2, State\_ab)  
AvgStateIncome <- summarise (AvgStateIncome, mean= mean(IncomeMean))  
AvgStateIncome <- filter(AvgStateIncome, grepl("N",State\_ab))  
ggplot(AvgStateIncome, aes(x=State\_ab,y=mean))+  
 geom\_point()+  
 labs(title = "Income Mean by State Scatterplot",  
 x = "States with N in Abbrevation",  
 y = "Income")



According to this graph, NJ had the largest average income while Tennesse had the lowest.