# Module 4 - Assignment 1

## Black, Tyler

### Data Transformation

library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.4 ✔ readr 2.1.5  
## ✔ forcats 1.0.0 ✔ stringr 1.5.1  
## ✔ ggplot2 3.5.1 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.3 ✔ tidyr 1.3.1  
## ✔ purrr 1.0.2   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

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

## Warning: One or more parsing issues, call `problems()` on your data frame for details,  
## e.g.:  
## dat <- vroom(...)  
## problems(dat)

## Rows: 32526 Columns: 18  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (7): State\_Name, State\_ab, County, City, Place, Type, Primary  
## dbl (11): id, State\_Code, Zip\_Code, Area\_Code, ALand, AWater, Lat, Lon, Mean...  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

#### State Incomes

Subsets will be created from the state\_income database. Different variables and conditions will be used. Variables will also be renamed. A subset for just North Carolina data will be created.

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

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

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

## # A tibble: 10 × 9  
## State\_Name State\_ab County City Type SquareArea IncomeMean IncomeMedian  
## <chr> <chr> <chr> <chr> <chr> <dbl> <dbl> <dbl>  
## 1 Alabama AL Mobile Co… Chic… City 10894952 38773 30506  
## 2 Alabama AL Barbour C… Loui… City 26070325 37725 19528  
## 3 Alabama AL Shelby Co… Colu… City 44835274 54606 31930  
## 4 Alabama AL Mobile Co… Sats… City 36878729 63919 52814  
## 5 Alabama AL Mobile Co… Daup… Town 16204185 77948 67225  
## 6 Alabama AL Cullman C… Cull… Town 8913021 50715 42643  
## 7 Alabama AL Escambia … East… City 8826252 33737 23610  
## 8 Alabama AL Elmore Co… Coos… Town 10222339 46319 40242  
## 9 Alabama AL Morgan Co… Eva Town 10544874 57994 39591  
## 10 Alabama AL Talladega… Syla… CDP 45178321 54807 41712  
## # ℹ 1 more variable: IncomeStdev <dbl>

NC\_Income <- filter(state\_income2,State\_Name == 'North Carolina')  
head(NC\_Income, n =10)

## # A tibble: 10 × 9  
## State\_Name State\_ab County City Type SquareArea IncomeMean IncomeMedian  
## <chr> <chr> <chr> <chr> <chr> <dbl> <dbl> <dbl>  
## 1 North Carolina NC Alama… Elon CDP 3515396 89973 300000  
## 2 North Carolina NC Johns… Wend… Town 23956770 67438 300000  
## 3 North Carolina NC Samps… Sted… Town 1353212 43538 25196  
## 4 North Carolina NC Hende… Hend… CDP 2625120 38120 31430  
## 5 North Carolina NC Beauf… Pine… Town 4121722 30468 17951  
## 6 North Carolina NC Davie… Clem… Town 5903422 97561 80720  
## 7 North Carolina NC Blade… Blad… Town 5737410 38588 20838  
## 8 North Carolina NC Samps… Clin… CDP 8562785 34778 23603  
## 9 North Carolina NC Lee C… Broa… Town 3350431 60384 52298  
## 10 North Carolina NC Guilf… Burl… City 75533002 54337 300000  
## # ℹ 1 more variable: IncomeStdev <dbl>

#### NC Income

The data set created earlier for just North Carolina data will be used to create income summaries by county, city, and type.

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

## # A tibble: 10 × 9  
## State\_Name State\_ab County City Type SquareArea IncomeMean IncomeMedian  
## <chr> <chr> <chr> <chr> <chr> <dbl> <dbl> <dbl>  
## 1 North Carolina NC Alama… Elon CDP 3515396 89973 300000  
## 2 North Carolina NC Alama… Meba… City 23213152 67397 55632  
## 3 North Carolina NC Alama… Hend… Track 12734435 57073 41022  
## 4 North Carolina NC Alama… Ahos… Track 199246026 54071 42038  
## 5 North Carolina NC Alama… Red … Track 93319263 30673 20786  
## 6 North Carolina NC Alama… Stat… Track 10829691 40174 27569  
## 7 North Carolina NC Alama… Supp… Track 29875162 45625 32324  
## 8 North Carolina NC Alama… Stat… Track 37718022 55177 48504  
## 9 North Carolina NC Alama… Moor… Track 13853696 106274 83085  
## 10 North Carolina NC Alama… Moor… Track 7037037 93463 79991  
## # ℹ 1 more variable: IncomeStdev <dbl>

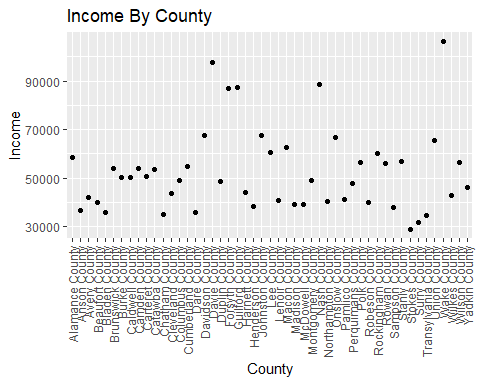
summary1 <- group\_by(NC\_Income,County)  
summary1 <- summarise (summary1, mean= mean(IncomeMean))  
  
summary2 <- NC\_Income %>%  
group\_by(City) %>%  
summarise (mean = mean(IncomeMean))

The summary1 data set groups the mean state incomes for North Carolina by county.  
Wake County has the highest mean state income.  
The summary2 data set groups the mean state incomes for North Carolina by city. Davidson has the highest mean state income.

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

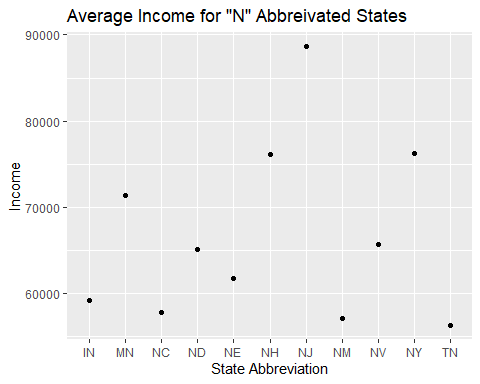
#### Income Visualization

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



Within the R Markdown document, create a new paragraph that states which county has the largest average income and which county has the lowest average income. Which has the second lowest income?  
**Wake County has the largest income while Stokes County has the lowest and Surry County had the second to lowest**

AvgStateIncome <- state\_income2 %>%  
group\_by(State\_ab) %>%  
summarise (mean = mean(IncomeMean))  
  
AvgStateIncome <- filter(AvgStateIncome, grepl('N',State\_ab))  
  
ggplot(data=AvgStateIncome,aes(x=State\_ab, y = mean)) +  
 geom\_point() +  
 labs(title = 'Average Income for "N" Abbreivated States', x= 'State Abbreviation',  
 y = 'Income')



**TN had the lowest average State Income while NJ had the largest State Income.**