# Module 4- Assignment 1

## Bradley, Zackery

### Data Transformation

library(tidyverse)

## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5 v purrr 0.3.4  
## v tibble 3.1.2 v dplyr 1.0.7  
## v tidyr 1.1.3 v stringr 1.4.0  
## v readr 1.4.0 v forcats 0.5.1

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

library(readr)  
state\_income <- read\_csv("C:/Users/bradl/OneDrive/Desktop/MODULE 2/Module 4/state\_income.csv")

##   
## -- Column specification --------------------------------------------------------  
## cols(  
## id = col\_double(),  
## State\_Code = col\_double(),  
## 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\_double(),  
## Area\_Code = col\_double(),  
## ALand = col\_double(),  
## AWater = col\_double(),  
## Lat = col\_double(),  
## Lon = col\_double(),  
## Mean = col\_double(),  
## Median = col\_double(),  
## Stdev = col\_double()  
## )

## Warning: 1 parsing failure.  
## row col expected actual file  
## 27548 Area\_Code a double M 'C:/Users/bradl/OneDrive/Desktop/MODULE 2/Module 4/state\_income.csv'

#==Part 1

####**State Incomes** #I will be creating a subset from the state\_income dataset. The columns that will be included are state\_name, state\_ab, county, city, type, Aland, mean, median, and stdev.

#==Question 5-9

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[,1:9], 10)

## # A tibble: 10 x 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 Chickasaw City 1.09e7 38773 30506 33101  
## 2 AL Alabama Barbour Coun~ Louisville City 2.61e7 37725 19528 43789  
## 3 AL Alabama Shelby County Columbiana City 4.48e7 54606 31930 57348  
## 4 AL Alabama Mobile County Satsuma City 3.69e7 63919 52814 47707  
## 5 AL Alabama Mobile County Dauphin I~ Town 1.62e7 77948 67225 54270  
## 6 AL Alabama Cullman Coun~ Cullman Town 8.91e6 50715 42643 35886  
## 7 AL Alabama Escambia Cou~ East Brew~ 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 Co~ Sylacauga CDP 4.52e7 54807 41712 51359

#==Question 10-11

State\_income2 <- rename(State\_income2,SquareArea=ALand,Incomemean=Mean,IncomeMedian=Median,IncomeStDev=Stdev)  
head(State\_income2[,1:9], 10)

## # A tibble: 10 x 9  
## State\_ab State\_Name County City Type SquareArea Incomemean IncomeMedian  
## <chr> <chr> <chr> <chr> <chr> <dbl> <dbl> <dbl>  
## 1 AL Alabama Mobile C~ Chick~ City 10894952 38773 30506  
## 2 AL Alabama Barbour ~ Louis~ City 26070325 37725 19528  
## 3 AL Alabama Shelby C~ Colum~ City 44835274 54606 31930  
## 4 AL Alabama Mobile C~ Satsu~ City 36878729 63919 52814  
## 5 AL Alabama Mobile C~ Dauph~ Town 16204185 77948 67225  
## 6 AL Alabama Cullman ~ Cullm~ Town 8913021 50715 42643  
## 7 AL Alabama Escambia~ East ~ City 8826252 33737 23610  
## 8 AL Alabama Elmore C~ Coosa~ Town 10222339 46319 40242  
## 9 AL Alabama Morgan C~ Eva Town 10544874 57994 39591  
## 10 AL Alabama Talladeg~ Sylac~ CDP 45178321 54807 41712  
## # ... with 1 more variable: IncomeStDev <dbl>

#==Question 12-13

NC\_income<- filter(State\_income2,State\_Name=="North Carolina")  
head(NC\_income[,1:9], 10)

## # A tibble: 10 x 9  
## State\_ab State\_Name County City Type SquareArea Incomemean IncomeMedian  
## <chr> <chr> <chr> <chr> <chr> <dbl> <dbl> <dbl>  
## 1 NC North Caro~ Alamanc~ Elon CDP 3515396 89973 300000  
## 2 NC North Caro~ Johnsto~ Wende~ Town 23956770 67438 300000  
## 3 NC North Caro~ Sampson~ Stedm~ Town 1353212 43538 25196  
## 4 NC North Caro~ Henders~ Hende~ CDP 2625120 38120 31430  
## 5 NC North Caro~ Beaufor~ Pinet~ Town 4121722 30468 17951  
## 6 NC North Caro~ Davie C~ Clemm~ Town 5903422 97561 80720  
## 7 NC North Caro~ Bladen ~ Blade~ Town 5737410 38588 20838  
## 8 NC North Caro~ Sampson~ Clint~ CDP 8562785 34778 23603  
## 9 NC North Caro~ Lee Cou~ Broad~ Town 3350431 60384 52298  
## 10 NC North Caro~ Guilfor~ Burli~ City 75533002 54337 300000  
## # ... with 1 more variable: IncomeStDev <dbl>

#==Question14 #==Part 2 #### **NC Incomes**

#==Question15 #I will be using the NC\_income dataset to create summaries of the incomes within North Carolina including summaries by county, city and type.

#==Question 16

arrange(NC\_income,County)

## # A tibble: 915 x 9  
## State\_ab State\_Name County City Type SquareArea Incomemean IncomeMedian  
## <chr> <chr> <chr> <chr> <chr> <dbl> <dbl> <dbl>  
## 1 NC North Caro~ Alamanc~ Elon CDP 3515396 89973 300000  
## 2 NC North Caro~ Alamanc~ Mebane City 23213152 67397 55632  
## 3 NC North Caro~ Alamanc~ Hende~ Track 12734435 57073 41022  
## 4 NC North Caro~ Alamanc~ Ahosk~ Track 199246026 54071 42038  
## 5 NC North Caro~ Alamanc~ Red S~ Track 93319263 30673 20786  
## 6 NC North Caro~ Alamanc~ State~ Track 10829691 40174 27569  
## 7 NC North Caro~ Alamanc~ Supply Track 29875162 45625 32324  
## 8 NC North Caro~ Alamanc~ State~ Track 37718022 55177 48504  
## 9 NC North Caro~ Alamanc~ Moore~ Track 13853696 106274 83085  
## 10 NC North Caro~ Alamanc~ Moore~ Track 7037037 93463 79991  
## # ... with 905 more rows, and 1 more variable: IncomeStDev <dbl>

head(NC\_income[,1:9], 10)

## # A tibble: 10 x 9  
## State\_ab State\_Name County City Type SquareArea Incomemean IncomeMedian  
## <chr> <chr> <chr> <chr> <chr> <dbl> <dbl> <dbl>  
## 1 NC North Caro~ Alamanc~ Elon CDP 3515396 89973 300000  
## 2 NC North Caro~ Johnsto~ Wende~ Town 23956770 67438 300000  
## 3 NC North Caro~ Sampson~ Stedm~ Town 1353212 43538 25196  
## 4 NC North Caro~ Henders~ Hende~ CDP 2625120 38120 31430  
## 5 NC North Caro~ Beaufor~ Pinet~ Town 4121722 30468 17951  
## 6 NC North Caro~ Davie C~ Clemm~ Town 5903422 97561 80720  
## 7 NC North Caro~ Bladen ~ Blade~ Town 5737410 38588 20838  
## 8 NC North Caro~ Sampson~ Clint~ CDP 8562785 34778 23603  
## 9 NC North Caro~ Lee Cou~ Broad~ Town 3350431 60384 52298  
## 10 NC North Caro~ Guilfor~ Burli~ City 75533002 54337 300000  
## # ... with 1 more variable: IncomeStDev <dbl>

#==Question 18

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

#==Question 20

#both summary 1 and 2 group the NC\_income dataset by county and city and also summarize the both groups by Incomemean.

#==Question 21

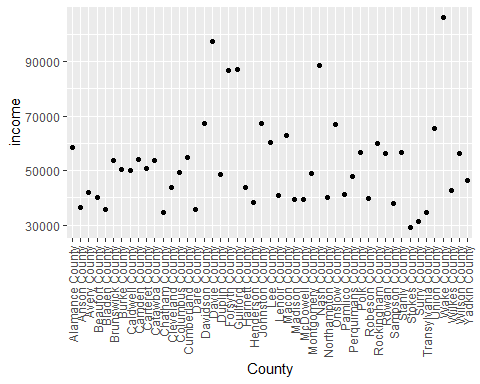
summary3 <- group\_by(NC\_income,Type)   
summary3 <- summarise(summary3, mean=mean(Incomemean))

#== Question 22 #==Part 3

#### **Income Visualization**

#==Question 23

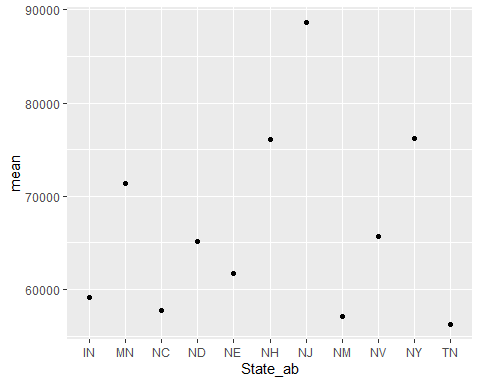
ggplot(data = summary1, aes(x=County, y=mean))+  
geom\_point()+  
ylab("income")+  
theme(axis.text.x = element\_text(angle = 90,vjust =0.5, hjust=1))

 #== Question 25

#The county with the highest average income is Wake County, with 106,139.The county with the lowest income is Stokes County with 28,921.50. The second lowest incomes is Surry County with 31,455.

#==Question 26

AvgStateIncome<- group\_by(State\_income2,State\_ab)   
AvgStateIncome<- summarise (AvgStateIncome, mean= mean(Incomemean))  
AvgStateIncome <- filter(AvgStateIncome,grepl("N",State\_ab))  
ggplot(data = AvgStateIncome, aes(x=State\_ab, y=mean))+  
geom\_point()



#==Question 29

#The smallest average income from our data is Tennessee (TN) with an avwerage income of 56,271.95. The largest average income from our data belongs to New Jersey (NJ) with an average total of 88,657.64.