Muley, Tushar Assignment 1-2 Week 1

June 13, 2021

# 1. Import, Plot, Summarize, and Save Data

# Using the Bureau of Labor Statistics data, choose a dataset that interest you.

# Then generate summary statistics for 2 variables, plot some of the features

# (e.g., histograms, box plots, density plots, etc.) of several variables, and save the

# data locally as CSV files.

#Imported Financial\_activities\_all\_states\_us.csv #variable to shorting name

fin\_act <- read.csv('C:/Users/Tushar/RStudioProjects/DSC630\_SUMMER\_2021/financial\_activities\_all\_states\_us.csv')

#Install pastecs package

install.packages("pastecs",repos = "http://cran.us.r-project.org")

## Installing package into 'C:/Users/Tushar/Documents/R/win-library/3.6'  
## (as 'lib' is unspecified)

## package 'pastecs' successfully unpacked and MD5 sums checked  
##   
## The downloaded binary packages are in  
## C:\Users\Tushar\AppData\Local\Temp\Rtmpg3gQFj\downloaded\_packages

library(pastecs,)

#See the column headers  
head(fin\_act,)

## area\_fips own\_code industry\_code agglvl\_code size\_code year qtr  
## 1 1000 1 1023 53 0 2020 4  
## 2 1000 2 1023 53 0 2020 4  
## 3 1000 3 1023 53 0 2020 4  
## 4 1000 5 1023 53 0 2020 4  
## 5 1001 3 1023 73 0 2020 4  
## 6 1001 5 1023 73 0 2020 4  
## disclosure\_code qtrly\_estabs month1\_emplvl month2\_emplvl month3\_emplvl  
## 1 18 136 142 146  
## 2 N 2 0 0 0  
## 3 133 1677 1675 1679  
## 4 14092 96165 95753 95065  
## 5 N 1 0 0 0  
## 6 120 399 393 401  
## total\_qtrly\_wages taxable\_qtrly\_wages qtrly\_contributions avg\_wkly\_wage  
## 1 3255664 0 0 1772  
## 2 0 0 0 0  
## 3 22571083 466753 4858 1035  
## 4 1970245660 62202613 966314 1584  
## 5 0 0 0 0  
## 6 5675563 290647 3720 1098  
## lq\_disclosure\_code lq\_qtrly\_estabs lq\_month1\_emplvl lq\_month2\_emplvl  
## 1 1.85 0.80 0.83  
## 2 N 1.37 0.00 0.00  
## 3 5.50 2.79 2.74  
## 4 1.16 0.85 0.84  
## 5 N 5.99 0.00 0.00  
## 6 1.43 0.62 0.61  
## lq\_month3\_emplvl lq\_total\_qtrly\_wages lq\_taxable\_qtrly\_wages  
## 1 0.85 0.71 0.00  
## 2 0.00 0.00 0.00  
## 3 2.76 2.87 1.11  
## 4 0.83 0.75 0.75  
## 5 0.00 0.00 0.00  
## 6 0.63 0.48 0.71  
## lq\_qtrly\_contributions lq\_avg\_wkly\_wage oty\_disclosure\_code  
## 1 0.00 0.85   
## 2 0.00 0.00 N  
## 3 1.81 1.04   
## 4 0.92 0.89   
## 5 0.00 0.00 N  
## 6 0.69 0.77   
## oty\_qtrly\_estabs\_chg oty\_qtrly\_estabs\_pct\_chg oty\_month1\_emplvl\_chg  
## 1 0 0.0 -1  
## 2 0 0.0 0  
## 3 -1 -0.7 -46  
## 4 364 2.7 297  
## 5 0 0.0 0  
## 6 1 0.8 -5  
## oty\_month1\_emplvl\_pct\_chg oty\_month2\_emplvl\_chg oty\_month2\_emplvl\_pct\_chg  
## 1 -0.7 2 1.4  
## 2 0.0 0 0.0  
## 3 -2.7 -52 -3.0  
## 4 0.3 -169 -0.2  
## 5 0.0 0 0.0  
## 6 -1.2 -11 -2.7  
## oty\_month3\_emplvl\_chg oty\_month3\_emplvl\_pct\_chg oty\_total\_qtrly\_wages\_chg  
## 1 11 8.1 41752  
## 2 0 0.0 0  
## 3 -52 -3.0 2554451  
## 4 -958 -1.0 213222335  
## 5 0 0.0 0  
## 6 -3 -0.7 474375  
## oty\_total\_qtrly\_wages\_pct\_chg oty\_taxable\_qtrly\_wages\_chg  
## 1 1.3 0  
## 2 0.0 0  
## 3 12.8 100770  
## 4 12.1 1215500  
## 5 0.0 0  
## 6 9.1 997  
## oty\_taxable\_qtrly\_wages\_pct\_chg oty\_qtrly\_contributions\_chg  
## 1 0.0 0  
## 2 0.0 0  
## 3 27.5 1064  
## 4 2.0 154025  
## 5 0.0 0  
## 6 0.3 -287  
## oty\_qtrly\_contributions\_pct\_chg oty\_avg\_wkly\_wage\_chg  
## 1 0.0 -28  
## 2 0.0 0  
## 3 28.0 143  
## 4 19.0 175  
## 5 0.0 0  
## 6 -7.2 108  
## oty\_avg\_wkly\_wage\_pct\_chg  
## 1 -1.6  
## 2 0.0  
## 3 16.0  
## 4 12.4  
## 5 0.0  
## 6 10.9

#Summary Statistics for 2 variables  
attach(fin\_act)  
sum\_fin\_act<-cbind(month1\_emplvl, total\_qtrly\_wages)  
stat.desc(sum\_fin\_act)

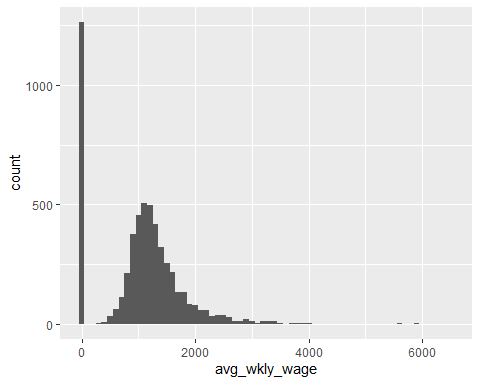
## month1\_emplvl total\_qtrly\_wages  
## nbr.val 5.548000e+03 5.548000e+03  
## nbr.null 1.265000e+03 1.263000e+03  
## nbr.na 0.000000e+00 0.000000e+00  
## min 0.000000e+00 0.000000e+00  
## max 8.178370e+05 3.846273e+10  
## range 8.178370e+05 3.846273e+10  
## sum 2.343910e+07 6.665326e+11  
## median 1.130000e+02 1.592396e+06  
## mean 4.224784e+03 1.201393e+08  
## SE.mean 3.926926e+02 1.458553e+07  
## CI.mean.0.95 7.698313e+02 2.859335e+07  
## var 8.555429e+08 1.180268e+18  
## std.dev 2.924967e+04 1.086402e+09  
## coef.var 6.923352e+00 9.042852e+00

#Plot some of the features # (e.g., histograms, box plots, density plots, etc.) of several variables

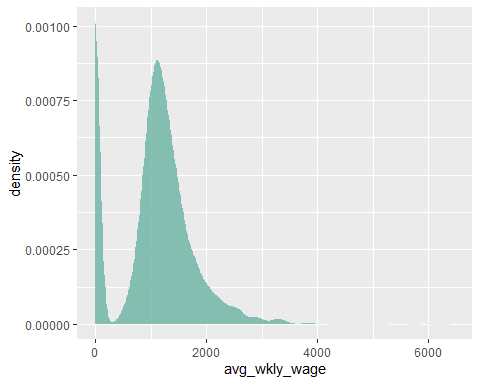
#Move data into a dataframe  
fin\_act\_df<-data.frame(fin\_act)  
  
#Preview the dataframe  
head(fin\_act\_df)

## area\_fips own\_code industry\_code agglvl\_code size\_code year qtr  
## 1 1000 1 1023 53 0 2020 4  
## 2 1000 2 1023 53 0 2020 4  
## 3 1000 3 1023 53 0 2020 4  
## 4 1000 5 1023 53 0 2020 4  
## 5 1001 3 1023 73 0 2020 4  
## 6 1001 5 1023 73 0 2020 4  
## disclosure\_code qtrly\_estabs month1\_emplvl month2\_emplvl month3\_emplvl  
## 1 18 136 142 146  
## 2 N 2 0 0 0  
## 3 133 1677 1675 1679  
## 4 14092 96165 95753 95065  
## 5 N 1 0 0 0  
## 6 120 399 393 401  
## total\_qtrly\_wages taxable\_qtrly\_wages qtrly\_contributions avg\_wkly\_wage  
## 1 3255664 0 0 1772  
## 2 0 0 0 0  
## 3 22571083 466753 4858 1035  
## 4 1970245660 62202613 966314 1584  
## 5 0 0 0 0  
## 6 5675563 290647 3720 1098  
## lq\_disclosure\_code lq\_qtrly\_estabs lq\_month1\_emplvl lq\_month2\_emplvl  
## 1 1.85 0.80 0.83  
## 2 N 1.37 0.00 0.00  
## 3 5.50 2.79 2.74  
## 4 1.16 0.85 0.84  
## 5 N 5.99 0.00 0.00  
## 6 1.43 0.62 0.61  
## lq\_month3\_emplvl lq\_total\_qtrly\_wages lq\_taxable\_qtrly\_wages  
## 1 0.85 0.71 0.00  
## 2 0.00 0.00 0.00  
## 3 2.76 2.87 1.11  
## 4 0.83 0.75 0.75  
## 5 0.00 0.00 0.00  
## 6 0.63 0.48 0.71  
## lq\_qtrly\_contributions lq\_avg\_wkly\_wage oty\_disclosure\_code  
## 1 0.00 0.85   
## 2 0.00 0.00 N  
## 3 1.81 1.04   
## 4 0.92 0.89   
## 5 0.00 0.00 N  
## 6 0.69 0.77   
## oty\_qtrly\_estabs\_chg oty\_qtrly\_estabs\_pct\_chg oty\_month1\_emplvl\_chg  
## 1 0 0.0 -1  
## 2 0 0.0 0  
## 3 -1 -0.7 -46  
## 4 364 2.7 297  
## 5 0 0.0 0  
## 6 1 0.8 -5  
## oty\_month1\_emplvl\_pct\_chg oty\_month2\_emplvl\_chg oty\_month2\_emplvl\_pct\_chg  
## 1 -0.7 2 1.4  
## 2 0.0 0 0.0  
## 3 -2.7 -52 -3.0  
## 4 0.3 -169 -0.2  
## 5 0.0 0 0.0  
## 6 -1.2 -11 -2.7  
## oty\_month3\_emplvl\_chg oty\_month3\_emplvl\_pct\_chg oty\_total\_qtrly\_wages\_chg  
## 1 11 8.1 41752  
## 2 0 0.0 0  
## 3 -52 -3.0 2554451  
## 4 -958 -1.0 213222335  
## 5 0 0.0 0  
## 6 -3 -0.7 474375  
## oty\_total\_qtrly\_wages\_pct\_chg oty\_taxable\_qtrly\_wages\_chg  
## 1 1.3 0  
## 2 0.0 0  
## 3 12.8 100770  
## 4 12.1 1215500  
## 5 0.0 0  
## 6 9.1 997  
## oty\_taxable\_qtrly\_wages\_pct\_chg oty\_qtrly\_contributions\_chg  
## 1 0.0 0  
## 2 0.0 0  
## 3 27.5 1064  
## 4 2.0 154025  
## 5 0.0 0  
## 6 0.3 -287  
## oty\_qtrly\_contributions\_pct\_chg oty\_avg\_wkly\_wage\_chg  
## 1 0.0 -28  
## 2 0.0 0  
## 3 28.0 143  
## 4 19.0 175  
## 5 0.0 0  
## 6 -7.2 108  
## oty\_avg\_wkly\_wage\_pct\_chg  
## 1 -1.6  
## 2 0.0  
## 3 16.0  
## 4 12.4  
## 5 0.0  
## 6 10.9

library(ggplot2)  
# Basic histogram of average week wages  
ggplot(fin\_act\_df, aes(x=avg\_wkly\_wage)) + geom\_histogram(binwidth = 100)

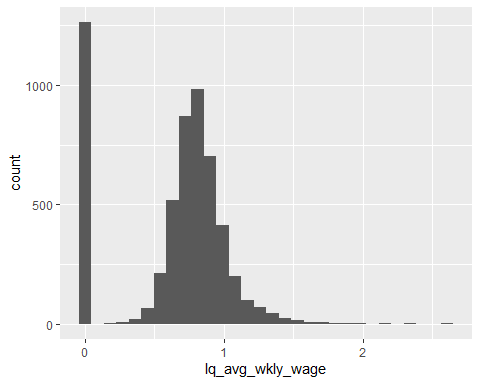


#Denisty Plot of average weekly wages  
ggplot(fin\_act\_df, aes(x=avg\_wkly\_wage)) +  
 geom\_density(fill="#69b3a2", color="#e9ecef", alpha=0.8)

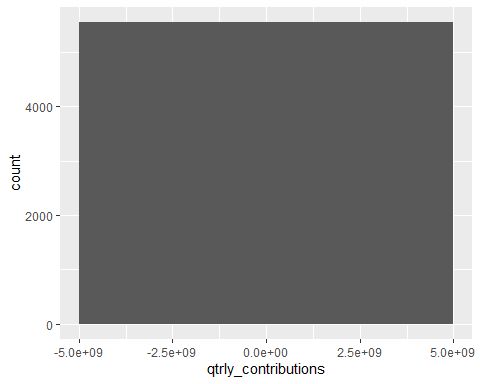


#Historam of average   
ggplot(fin\_act\_df, aes(x=lq\_avg\_wkly\_wage)) + geom\_histogram()

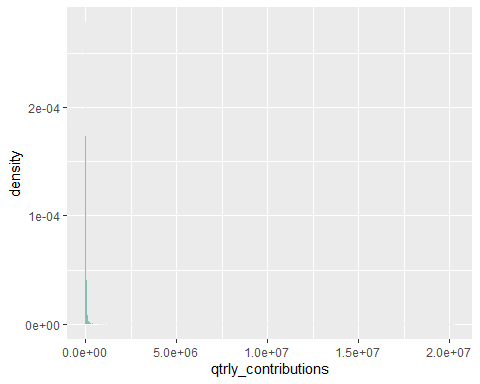
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



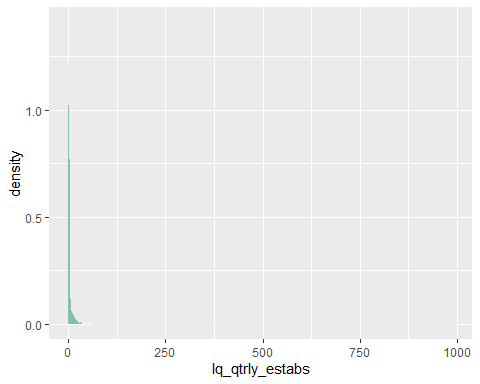
#Historam of quarterly contributions  
ggplot(fin\_act\_df, aes(x=qtrly\_contributions)) + geom\_histogram(binwidth = 10000000000)



#Denisty Plot of quarterly contributions  
ggplot(fin\_act\_df, aes(x=qtrly\_contributions)) +  
 geom\_density(fill="#69b3a2", color="#e9ecef", alpha=0.8)



#Denisty Plot of average weekly wages  
ggplot(fin\_act\_df, aes(x=lq\_qtrly\_estabs)) +  
 geom\_density(fill="#69b3a2", color="#e9ecef", alpha=0.8)



#Need to filter the data by a level code. #Level code 53 Statewide, by Supersector – by ownership sector

head(fin\_act\_df)

## area\_fips own\_code industry\_code agglvl\_code size\_code year qtr  
## 1 1000 1 1023 53 0 2020 4  
## 2 1000 2 1023 53 0 2020 4  
## 3 1000 3 1023 53 0 2020 4  
## 4 1000 5 1023 53 0 2020 4  
## 5 1001 3 1023 73 0 2020 4  
## 6 1001 5 1023 73 0 2020 4  
## disclosure\_code qtrly\_estabs month1\_emplvl month2\_emplvl month3\_emplvl  
## 1 18 136 142 146  
## 2 N 2 0 0 0  
## 3 133 1677 1675 1679  
## 4 14092 96165 95753 95065  
## 5 N 1 0 0 0  
## 6 120 399 393 401  
## total\_qtrly\_wages taxable\_qtrly\_wages qtrly\_contributions avg\_wkly\_wage  
## 1 3255664 0 0 1772  
## 2 0 0 0 0  
## 3 22571083 466753 4858 1035  
## 4 1970245660 62202613 966314 1584  
## 5 0 0 0 0  
## 6 5675563 290647 3720 1098  
## lq\_disclosure\_code lq\_qtrly\_estabs lq\_month1\_emplvl lq\_month2\_emplvl  
## 1 1.85 0.80 0.83  
## 2 N 1.37 0.00 0.00  
## 3 5.50 2.79 2.74  
## 4 1.16 0.85 0.84  
## 5 N 5.99 0.00 0.00  
## 6 1.43 0.62 0.61  
## lq\_month3\_emplvl lq\_total\_qtrly\_wages lq\_taxable\_qtrly\_wages  
## 1 0.85 0.71 0.00  
## 2 0.00 0.00 0.00  
## 3 2.76 2.87 1.11  
## 4 0.83 0.75 0.75  
## 5 0.00 0.00 0.00  
## 6 0.63 0.48 0.71  
## lq\_qtrly\_contributions lq\_avg\_wkly\_wage oty\_disclosure\_code  
## 1 0.00 0.85   
## 2 0.00 0.00 N  
## 3 1.81 1.04   
## 4 0.92 0.89   
## 5 0.00 0.00 N  
## 6 0.69 0.77   
## oty\_qtrly\_estabs\_chg oty\_qtrly\_estabs\_pct\_chg oty\_month1\_emplvl\_chg  
## 1 0 0.0 -1  
## 2 0 0.0 0  
## 3 -1 -0.7 -46  
## 4 364 2.7 297  
## 5 0 0.0 0  
## 6 1 0.8 -5  
## oty\_month1\_emplvl\_pct\_chg oty\_month2\_emplvl\_chg oty\_month2\_emplvl\_pct\_chg  
## 1 -0.7 2 1.4  
## 2 0.0 0 0.0  
## 3 -2.7 -52 -3.0  
## 4 0.3 -169 -0.2  
## 5 0.0 0 0.0  
## 6 -1.2 -11 -2.7  
## oty\_month3\_emplvl\_chg oty\_month3\_emplvl\_pct\_chg oty\_total\_qtrly\_wages\_chg  
## 1 11 8.1 41752  
## 2 0 0.0 0  
## 3 -52 -3.0 2554451  
## 4 -958 -1.0 213222335  
## 5 0 0.0 0  
## 6 -3 -0.7 474375  
## oty\_total\_qtrly\_wages\_pct\_chg oty\_taxable\_qtrly\_wages\_chg  
## 1 1.3 0  
## 2 0.0 0  
## 3 12.8 100770  
## 4 12.1 1215500  
## 5 0.0 0  
## 6 9.1 997  
## oty\_taxable\_qtrly\_wages\_pct\_chg oty\_qtrly\_contributions\_chg  
## 1 0.0 0  
## 2 0.0 0  
## 3 27.5 1064  
## 4 2.0 154025  
## 5 0.0 0  
## 6 0.3 -287  
## oty\_qtrly\_contributions\_pct\_chg oty\_avg\_wkly\_wage\_chg  
## 1 0.0 -28  
## 2 0.0 0  
## 3 28.0 143  
## 4 19.0 175  
## 5 0.0 0  
## 6 -7.2 108  
## oty\_avg\_wkly\_wage\_pct\_chg  
## 1 -1.6  
## 2 0.0  
## 3 16.0  
## 4 12.4  
## 5 0.0  
## 6 10.9

fin\_act\_C53\_df <- fin\_act\_df[fin\_act\_df$agglvl\_code == '53',]

head(fin\_act\_C53\_df)

## area\_fips own\_code industry\_code agglvl\_code size\_code year qtr  
## 1 1000 1 1023 53 0 2020 4  
## 2 1000 2 1023 53 0 2020 4  
## 3 1000 3 1023 53 0 2020 4  
## 4 1000 5 1023 53 0 2020 4  
## 145 2000 1 1023 53 0 2020 4  
## 146 2000 2 1023 53 0 2020 4  
## disclosure\_code qtrly\_estabs month1\_emplvl month2\_emplvl month3\_emplvl  
## 1 18 136 142 146  
## 2 N 2 0 0 0  
## 3 133 1677 1675 1679  
## 4 14092 96165 95753 95065  
## 145 5 25 26 26  
## 146 N 15 0 0 0  
## total\_qtrly\_wages taxable\_qtrly\_wages qtrly\_contributions avg\_wkly\_wage  
## 1 3255664 0 0 1772  
## 2 0 0 0 0  
## 3 22571083 466753 4858 1035  
## 4 1970245660 62202613 966314 1584  
## 145 656187 0 0 1967  
## 146 0 0 0 0  
## lq\_disclosure\_code lq\_qtrly\_estabs lq\_month1\_emplvl lq\_month2\_emplvl  
## 1 1.85 0.80 0.83  
## 2 N 1.37 0.00 0.00  
## 3 5.50 2.79 2.74  
## 4 1.16 0.85 0.84  
## 145 2.99 0.96 1.01  
## 146 N 59.64 0.00 0.00  
## lq\_month3\_emplvl lq\_total\_qtrly\_wages lq\_taxable\_qtrly\_wages  
## 1 0.85 0.71 0.00  
## 2 0.00 0.00 0.00  
## 3 2.76 2.87 1.11  
## 4 0.83 0.75 0.75  
## 145 1.02 0.82 0.00  
## 146 0.00 0.00 0.00  
## lq\_qtrly\_contributions lq\_avg\_wkly\_wage oty\_disclosure\_code  
## 1 0.00 0.85   
## 2 0.00 0.00 N  
## 3 1.81 1.04   
## 4 0.92 0.89   
## 145 0.00 0.82   
## 146 0.00 0.00 N  
## oty\_qtrly\_estabs\_chg oty\_qtrly\_estabs\_pct\_chg oty\_month1\_emplvl\_chg  
## 1 0 0.0 -1  
## 2 0 0.0 0  
## 3 -1 -0.7 -46  
## 4 364 2.7 297  
## 145 0 0.0 0  
## 146 0 0.0 0  
## oty\_month1\_emplvl\_pct\_chg oty\_month2\_emplvl\_chg oty\_month2\_emplvl\_pct\_chg  
## 1 -0.7 2 1.4  
## 2 0.0 0 0.0  
## 3 -2.7 -52 -3.0  
## 4 0.3 -169 -0.2  
## 145 0.0 1 4.0  
## 146 0.0 0 0.0  
## oty\_month3\_emplvl\_chg oty\_month3\_emplvl\_pct\_chg oty\_total\_qtrly\_wages\_chg  
## 1 11 8.1 41752  
## 2 0 0.0 0  
## 3 -52 -3.0 2554451  
## 4 -958 -1.0 213222335  
## 145 3 13.0 8021  
## 146 0 0.0 0  
## oty\_total\_qtrly\_wages\_pct\_chg oty\_taxable\_qtrly\_wages\_chg  
## 1 1.3 0  
## 2 0.0 0  
## 3 12.8 100770  
## 4 12.1 1215500  
## 145 1.2 0  
## 146 0.0 0  
## oty\_taxable\_qtrly\_wages\_pct\_chg oty\_qtrly\_contributions\_chg  
## 1 0.0 0  
## 2 0.0 0  
## 3 27.5 1064  
## 4 2.0 154025  
## 145 0.0 0  
## 146 0.0 0  
## oty\_qtrly\_contributions\_pct\_chg oty\_avg\_wkly\_wage\_chg  
## 1 0 -28  
## 2 0 0  
## 3 28 143  
## 4 19 175  
## 145 0 -82  
## 146 0 0  
## oty\_avg\_wkly\_wage\_pct\_chg  
## 1 -1.6  
## 2 0.0  
## 3 16.0  
## 4 12.4  
## 145 -4.0  
## 146 0.0

#Min  
min(fin\_act\_C53\_df$total\_qtrly\_wages)

## [1] 0

#Max  
max(fin\_act\_C53\_df$total\_qtrly\_wages)

## [1] 35504756396

#Median  
median(fin\_act\_C53\_df$total\_qtrly\_wages)

## [1] 5740782

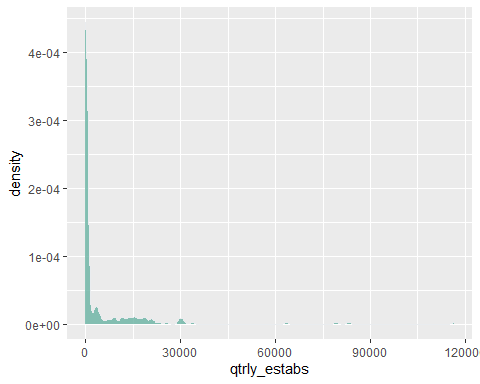
#Mean  
mean(fin\_act\_C53\_df$total\_qtrly\_wages)

## [1] 1216410953

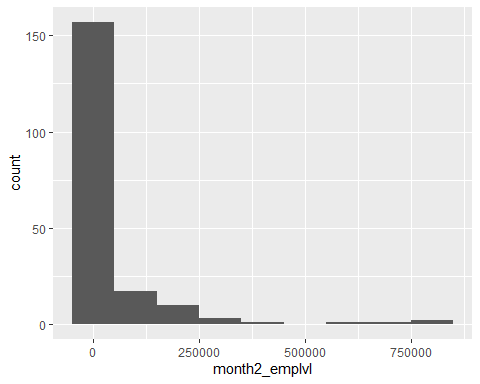
#Median  
median(fin\_act\_C53\_df$avg\_wkly\_wage)

## [1] 1536

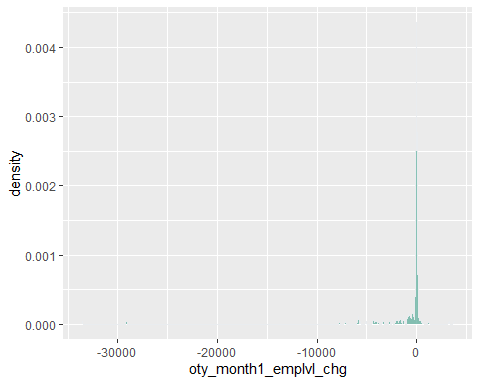
#Denisty Plot of qtrly\_establishment  
ggplot(fin\_act\_C53\_df, aes(x=qtrly\_estabs)) +  
 geom\_density(fill="#69b3a2", color="#e9ecef", alpha=0.8)



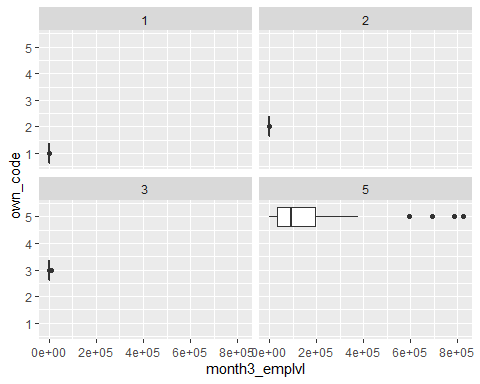
#Histogram of month2\_emplvl  
ggplot(fin\_act\_C53\_df, aes(x=month2\_emplvl)) + geom\_histogram(binwidth = 100000)



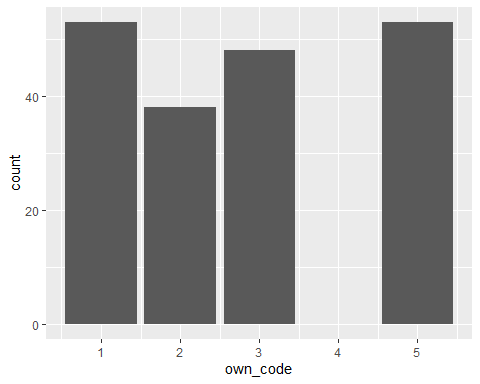
#Density plot of oty\_month1\_emplvl\_chg  
ggplot(fin\_act\_C53\_df, aes(x=oty\_month1\_emplvl\_chg)) +  
 geom\_density(fill="#69b3a2", color="#e9ecef", alpha=0.8)



#Boxplot of month 3 Employment and own code  
ggplot(fin\_act\_C53\_df, aes(x=month3\_emplvl, y=own\_code, group=own\_code)) +   
 geom\_boxplot() + facet\_wrap(~own\_code)



#Bar graph of own codes  
ggplot(fin\_act\_C53\_df) +  
 geom\_bar(mapping = aes(x = own\_code))



#2. Explore Some Bivariate Relations #Use the same dataset within the same website to explore some bivariate relations #(e.g. bivariate plot, correlation, table cross table etc.)

install.packages("ggpubr", repos = "http://cran.us.r-project.org")

## Installing package into 'C:/Users/Tushar/Documents/R/win-library/3.6'  
## (as 'lib' is unspecified)

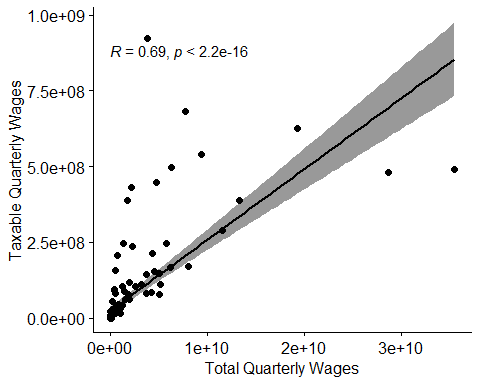
## package 'ggpubr' successfully unpacked and MD5 sums checked  
##   
## The downloaded binary packages are in  
## C:\Users\Tushar\AppData\Local\Temp\Rtmpg3gQFj\downloaded\_packages

library("ggpubr")

#Checking to see Quarterly wage correlate to Taxable Quarter Wages

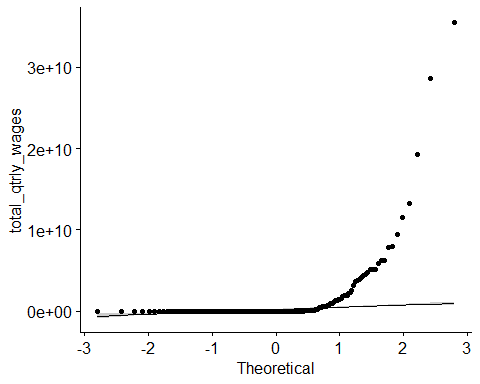
ggscatter(fin\_act\_C53\_df, x = "total\_qtrly\_wages", y = "taxable\_qtrly\_wages",   
 add = "reg.line", conf.int = TRUE,   
 cor.coef = TRUE, cor.method = "pearson",  
 xlab = "Total Quarterly Wages", ylab = "Taxable Quarterly Wages")

## `geom\_smooth()` using formula 'y ~ x'

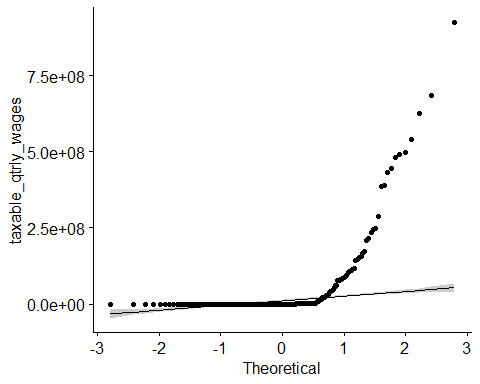


#QQPlot

# total\_qtrly\_wages  
ggqqplot(fin\_act\_C53\_df$total\_qtrly\_wages, ylab = "total\_qtrly\_wages")



# Taxable Quarterly Wages  
ggqqplot(fin\_act\_C53\_df$taxable\_qtrly\_wages, ylab = "taxable\_qtrly\_wages")



#Pearson Correlation

person\_cor <- cor.test(fin\_act\_C53\_df$total\_qtrly\_wages,   
 fin\_act\_C53\_df$taxable\_qtrly\_wages, method = "pearson")  
person\_cor

##   
## Pearson's product-moment correlation  
##   
## data: fin\_act\_C53\_df$total\_qtrly\_wages and fin\_act\_C53\_df$taxable\_qtrly\_wages  
## t = 13.168, df = 190, p-value < 2.2e-16  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.6087066 0.7582125  
## sample estimates:  
## cor   
## 0.6907718

#Pearson Correlation:

#t is the t-test statistic value (t = 13.168), #df is the degrees of freedom (df= 190), #p-value is p-value = 2.2e^-16}). conf.int is the confidence interval of the correlation coefficient at 95% (conf.int = [0.6087, 0.7582]);

sample estimates is the correlation coefficient (Cor.coeff = 0.6907).

#Kendall Rank Correlation

kendall\_cor <- cor.test(fin\_act\_C53\_df$total\_qtrly\_wages,   
 fin\_act\_C53\_df$taxable\_qtrly\_wages, method="kendall")  
kendall\_cor

##   
## Kendall's rank correlation tau  
##   
## data: fin\_act\_C53\_df$total\_qtrly\_wages and fin\_act\_C53\_df$taxable\_qtrly\_wages  
## z = 12.692, p-value < 2.2e-16  
## alternative hypothesis: true tau is not equal to 0  
## sample estimates:  
## tau   
## 0.6706145

#Kendall Correlation: #Correlation coeffient between total quarterly wages and taxable quarterly wages #0.6706 and p-value = 2.2e^-16

#Spearman Correlation

spreaman\_cor <-cor.test(fin\_act\_C53\_df$total\_qtrly\_wages,   
 fin\_act\_C53\_df$taxable\_qtrly\_wages, method = "spearman")

## Warning in cor.test.default(fin\_act\_C53\_df$total\_qtrly\_wages,  
## fin\_act\_C53\_df$taxable\_qtrly\_wages, : Cannot compute exact p-value with ties

spreaman\_cor

##   
## Spearman's rank correlation rho  
##   
## data: fin\_act\_C53\_df$total\_qtrly\_wages and fin\_act\_C53\_df$taxable\_qtrly\_wages  
## S = 230102, p-value < 2.2e-16  
## alternative hypothesis: true rho is not equal to 0  
## sample estimates:  
## rho   
## 0.8049345

#Spearman Correlation: #Correlation coefficient between quarterly wages and taxable quarterly wages is #0.8049 rho and p-value is 2.2e^-16

#Based on this these correlation I would say we have a positive correlation between #quarterly wages and taxable quarterly wages.

#It is on a perfect correlation but strong positive correlation: #Pearson: 0.6907 #Kendall: 0.6706 #Spearan: 0.8049

#If quarter wages go up so does taxable quarterly wages.

#3. Organize a Data Report #Generate a summary report. Make sure to include: summary for every variable, #structure and type of data elements, discuss four results of your data.

str(fin\_act\_C53\_df)

## 'data.frame': 192 obs. of 42 variables:  
## $ area\_fips : Factor w/ 3707 levels "1000","10000",..: 1 1 1 1 659 659 659 659 1967 1967 ...  
## $ own\_code : int 1 2 3 5 1 2 3 5 1 3 ...  
## $ industry\_code : int 1023 1023 1023 1023 1023 1023 1023 1023 1023 1023 ...  
## $ agglvl\_code : int 53 53 53 53 53 53 53 53 53 53 ...  
## $ size\_code : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ year : int 2020 2020 2020 2020 2020 2020 2020 2020 2020 2020 ...  
## $ qtr : int 4 4 4 4 4 4 4 4 4 4 ...  
## $ disclosure\_code : Factor w/ 3 levels "","-","N": 1 3 1 1 1 3 3 1 1 1 ...  
## $ qtrly\_estabs : int 18 2 133 14092 5 15 3 1540 12 18 ...  
## $ month1\_emplvl : int 136 0 1677 96165 25 0 0 10395 125 761 ...  
## $ month2\_emplvl : int 142 0 1675 95753 26 0 0 10384 126 764 ...  
## $ month3\_emplvl : int 146 0 1679 95065 26 0 0 10309 127 784 ...  
## $ total\_qtrly\_wages : num 3.26e+06 0.00 2.26e+07 1.97e+09 6.56e+05 ...  
## $ taxable\_qtrly\_wages : int 0 0 466753 62202613 0 0 0 54053368 0 179839 ...  
## $ qtrly\_contributions : int 0 0 4858 966314 0 0 0 816004 0 7386 ...  
## $ avg\_wkly\_wage : int 1772 0 1035 1584 1967 0 0 1392 2220 1070 ...  
## $ lq\_disclosure\_code : Factor w/ 2 levels "","N": 1 2 1 1 1 2 2 1 1 1 ...  
## $ lq\_qtrly\_estabs : num 1.85 1.37 5.5 1.16 2.99 ...  
## $ lq\_month1\_emplvl : num 0.8 0 2.79 0.85 0.96 0 0 0.59 0.5 0.86 ...  
## $ lq\_month2\_emplvl : num 0.83 0 2.74 0.84 1.01 0 0 0.61 0.5 0.84 ...  
## $ lq\_month3\_emplvl : num 0.85 0 2.76 0.83 1.02 0 0 0.61 0.5 0.87 ...  
## $ lq\_total\_qtrly\_wages : num 0.71 0 2.87 0.75 0.82 0 0 0.41 0.48 0.83 ...  
## $ lq\_taxable\_qtrly\_wages : num 0 0 1.11 0.75 0 0 0 0.98 0 0.29 ...  
## $ lq\_qtrly\_contributions : num 0 0 1.81 0.92 0 0 0 0.96 0 1.41 ...  
## $ lq\_avg\_wkly\_wage : num 0.85 0 1.04 0.89 0.82 0 0 0.68 0.97 0.97 ...  
## $ oty\_disclosure\_code : Factor w/ 2 levels "","N": 1 2 1 1 1 2 2 1 1 2 ...  
## $ oty\_qtrly\_estabs\_chg : int 0 0 -1 364 0 0 0 118 0 0 ...  
## $ oty\_qtrly\_estabs\_pct\_chg : num 0 0 -0.7 2.7 0 0 0 8.3 0 0 ...  
## $ oty\_month1\_emplvl\_chg : int -1 0 -46 297 0 0 0 -852 0 0 ...  
## $ oty\_month1\_emplvl\_pct\_chg : num -0.7 0 -2.7 0.3 0 0 0 -7.6 0 0 ...  
## $ oty\_month2\_emplvl\_chg : int 2 0 -52 -169 1 0 0 -723 -2 0 ...  
## $ oty\_month2\_emplvl\_pct\_chg : num 1.4 0 -3 -0.2 4 0 0 -6.5 -1.6 0 ...  
## $ oty\_month3\_emplvl\_chg : int 11 0 -52 -958 3 0 0 -733 1 0 ...  
## $ oty\_month3\_emplvl\_pct\_chg : num 8.1 0 -3 -1 13 0 0 -6.6 0.8 0 ...  
## $ oty\_total\_qtrly\_wages\_chg : num 4.18e+04 0.00 2.55e+06 2.13e+08 8.02e+03 ...  
## $ oty\_total\_qtrly\_wages\_pct\_chg : num 1.3 0 12.8 12.1 1.2 0 0 5.2 1.2 0 ...  
## $ oty\_taxable\_qtrly\_wages\_chg : int 0 0 100770 1215500 0 0 0 -1470347 0 0 ...  
## $ oty\_taxable\_qtrly\_wages\_pct\_chg: num 0 0 27.5 2 0 0 0 -2.6 0 0 ...  
## $ oty\_qtrly\_contributions\_chg : int 0 0 1064 154025 0 0 0 -27202 0 0 ...  
## $ oty\_qtrly\_contributions\_pct\_chg: num 0 0 28 19 0 0 0 -3.2 0 0 ...  
## $ oty\_avg\_wkly\_wage\_chg : int -28 0 143 175 -82 0 0 161 32 0 ...  
## $ oty\_avg\_wkly\_wage\_pct\_chg : num -1.6 0 16 12.4 -4 0 0 13.1 1.5 0 ...