Severeinjury\_Rscript.R

new

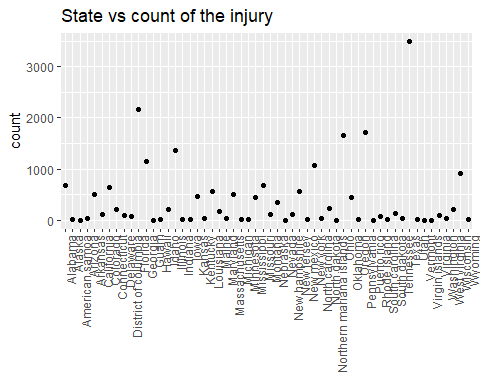
Wed Sep 13 09:37:49 2017

cat("\014") #clear the console

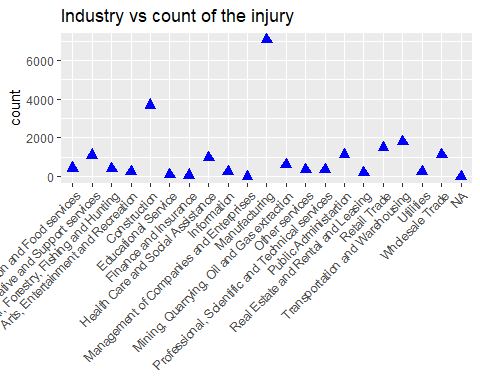
rm(list=ls()) #dclearing the environment variables  
  
library(data.table)  
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.4.1

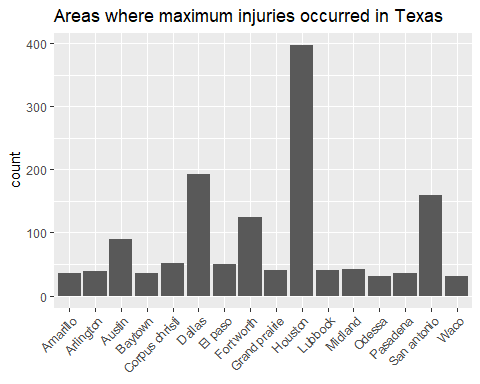
# loading raw data  
dataset = read.csv("severeinjury1.csv")  
dataset$Address = paste(dataset$Address1, dataset$Address2)  
dataset$Address1 = NULL  
dataset$Address2 = NULL  
dataset = as.data.frame(dataset)  
dataset = dataset[c(1,2,3,4,25,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24)]  
  
#loadingPre-processed data  
dataset = read.csv("processed\_data.csv")  
  
#function to convert first letter as a character  
first\_character\_uppercase = function(x) {  
 substr(x,1,1) = toupper(substr(x,1,1))  
 x  
}  
  
#dataset$Employer = paste(toupper(substr(dataset$Employer, 1, 1)),   
#substr(dataset$Employer, 2, nchar(dataset$Employer)), sep="")  
dataset$EventDate.1 = NULL  
  
dataset$Employer = tolower(dataset$Employer)  
dataset$City = tolower(dataset$City)  
dataset$State = tolower(dataset$State)  
dataset$Employer = first\_character\_uppercase(dataset$Employer)  
dataset$City = first\_character\_uppercase(dataset$City)  
dataset$State = first\_character\_uppercase(dataset$State)  
  
dataset$count = rep(1, nrow(dataset))  
  
dt = data.table(dataset)  
injurycount\_state = dt[,sum(count),by=State]  
colnames(injurycount\_state)[2] = "count"  
#rm(dt)  
  
state\_count = ggplot(injurycount\_state, aes(x=State, y=count))+geom\_point()+ ggtitle("State vs count of the injury")+xlab("State")+theme(axis.title.x = element\_blank(), axis.text.x = element\_text(angle = 90, hjust = 1))  
colnames(dataset)[27] = "Industry\_codes"  
  
dataset$industrynames = ifelse(dataset$Industry\_codes == "11", "Agricultural, Forestry, Fishing and Hunting",   
 ifelse(dataset$Industry\_codes == "21", "Mining, Quarrying, Oil and Gas extraction",  
 ifelse(dataset$Industry\_codes == "22", "Utilities",   
 ifelse(dataset$Industry\_codes == "23", "Construction",   
 ifelse(dataset$Industry\_codes == "31", "Manufacturing",   
 ifelse(dataset$Industry\_codes == "32", "Manufacturing",   
 ifelse(dataset$Industry\_codes == "33", "Manufacturing",   
 ifelse(dataset$Industry\_codes == "42", "Wholesale Trade",   
 ifelse(dataset$Industry\_codes == "44", "Retail Trade",   
 ifelse(dataset$Industry\_codes == "45", "Retail Trade",   
 ifelse(dataset$Industry\_codes == "48", "Transportation and Warehousing",   
 ifelse(dataset$Industry\_codes == "49", "Transportation and Warehousing",   
 ifelse(dataset$Industry\_codes == "51", "Information",   
 ifelse(dataset$Industry\_codes == "52", "Finance and Insurance",   
 ifelse(dataset$Industry\_codes == "53", "Real Estate and Rental and Leasing",   
 ifelse(dataset$Industry\_codes == "54", "Professional, Scientific and Technical services",  
 ifelse(dataset$Industry\_codes == "55", "Management of Companies and Enterprises",  
 ifelse(dataset$Industry\_codes == "56", "Administrative and Support services",   
 ifelse(dataset$Industry\_codes == "61", "Educational Service",   
 ifelse(dataset$Industry\_codes == "62", "Health Care and Social Assistance",  
 ifelse(dataset$Industry\_codes == "71", "Arts, Entertainment and Recreation",  
 ifelse(dataset$Industry\_codes == "72", "Accomodation and Food services",  
 ifelse(dataset$Industry\_codes == "81", "Other services", "Public Administartion")))))))))))))))))))))))  
  
dt1 = data.table(dataset)  
injurycount\_industry = dt1[,sum(count),by=industrynames]  
colnames(injurycount\_industry)[2] = "count"  
#rm(dt)  
  
industry\_count = ggplot(injurycount\_industry, aes(x=industrynames, y=count))+geom\_point(color = "blue", size = 3, shape= 17)+ ggtitle("Industry vs count of the injury")+xlab("Industry")+theme(axis.title.x = element\_blank(), axis.text.x = element\_text(angle = 45, hjust = 1))  
colnames(dataset)[27] = "Industrycode"  
  
#plot = ggplot()+geom\_point(dataset, aes(x=dataset$))  
  
#rmarkdown::render("dataset\_script.R", "html\_document")  
# multiplot(state\_count, industry\_count)  
print(state\_count)



print(industry\_count)



Injuryprone\_area = dataset[dataset$State == "Texas",]  
  
dt2 = data.table(Injuryprone\_area)  
injurycount\_city = dt2[,sum(count),by=City]  
colnames(injurycount\_city)[2] = "count"  
#rm(dt)  
  
Maxinjuries\_city = injurycount\_city[injurycount\_city$count > 30]  
city\_count = ggplot(Maxinjuries\_city, aes(x=City, y=count))+geom\_bar(stat = "identity")+ ggtitle("Areas where maximum injuries occurred in Texas")+xlab("City")+theme(axis.title.x = element\_blank(), axis.text.x = element\_text(angle = 45, hjust = 1))  
# colnames(dataset)[27] = "Industrycode"  
  
print(city\_count)



#rmarkdown::render("Severeinjury\_Rscript.R", "html\_document")