Course Project - Phase 1

Jordan Whitaker

6/15/2020

chi <- read\_csv("chicago2.csv")

## Warning: Missing column names filled in: 'X1' [1]

## Parsed with column specification:  
## cols(  
## .default = col\_character(),  
## X1 = col\_double(),  
## ID = col\_double(),  
## Arrest = col\_logical(),  
## Domestic = col\_logical(),  
## Ward = col\_double(),  
## `Community Area` = col\_double(),  
## `X Coordinate` = col\_double(),  
## `Y Coordinate` = col\_double(),  
## Year = col\_double(),  
## Latitude = col\_double(),  
## Longitude = col\_double()  
## )

## See spec(...) for full column specifications.

#head(chi)  
  
chi = chi[, -c(2,3, 17, 18, 20, 23)]  
  
chi = chi %>% mutate(Date = mdy\_hms(Date))  
#head(chi)  
  
chi = chi %>% mutate(Hour = hour(Date))

# Creating chi\_new df  
chi\_new = chi %>% group\_by(chi$`Primary Type`) %>% filter(n() > 20)  
chi\_new$Quarter = format(chi\_new$Date, "%m")  
chi\_new = chi\_new[, -c(19)]

## Creating Factors and Prepping Data

chi\_new$Hour = as.factor(case\_when(  
 chi\_new$Hour <=5 ~ 'Early\_Morning',  
 chi\_new$Hour >=6 & chi\_new$Hour <=11 ~ 'Morning',  
 chi\_new$Hour >=12 & chi\_new$Hour <=17 ~ 'Afternoon',  
 chi\_new$Hour >=18 & chi\_new$Hour <=24 ~ 'Night'  
))  
  
  
chi\_new$Hour <- factor(chi\_new$Hour, levels=c("Early\_Morning", "Morning", "Afternoon", "Night"))  
  
chi\_new$Quarter = as.numeric(chi\_new$Quarter)  
  
chi\_new$Quarter = as.factor(case\_when(  
 chi\_new$Quarter <= 3 ~ 'Q1',  
 chi\_new$Quarter > 3 & chi\_new$Quarter <=6 ~ 'Q2',  
 chi\_new$Quarter >= 7 & chi\_new$Quarter <= 10 ~ 'Q3',  
 chi\_new$Quarter >= 10 & chi\_new$Quarter <= 12 ~ 'Q4'  
))  
  
  
  
  
chi\_new$Arrest = as.factor(case\_when(  
 chi\_new$Arrest == 'TRUE' ~ 'Arrest',  
 chi\_new$Arrest == 'FALSE' ~ 'NO\_Arrest'  
))  
  
  
  
chi\_new$Offense <- paste(chi\_new$`Primary Type`, chi\_new$Description)  
  
  
chi\_new$Hour = as.factor(chi\_new$Hour)  
chi\_new$`Primary Type` = as.factor(chi\_new$`Primary Type`)  
chi\_new$`Location Description` = as.factor(chi\_new$`Location Description`)  
  
  
chi\_new = chi\_new[, -c(5, 6)]  
  
  
chi\_new <-   
 chi\_new %>%   
 mutate(  
 Offense\_new = case\_when(  
 str\_detect(Offense, 'ASSAULT AGG') ~ 'AGGREVATED ASSAULT',  
 str\_detect(Offense, 'ASSAULT SIMPLE|ASSAULT PRO') ~ 'SIMPLE ASSAULT',  
 str\_detect(Offense, 'BATTERY') ~ 'BATTERY',  
 str\_detect(Offense, 'BURGLARY') ~ 'BURGLARY',  
 str\_detect(Offense, 'CRIM SEXUAL') ~ 'CRIMINAL SEX OFFENSE',  
 str\_detect(Offense, 'CRIMINAL DAMAGE') ~ 'CRIMINAL DAMAGE',  
 str\_detect(Offense, 'CRIMINAL TRESPASS') ~ 'CRIMINAL TRESPASS',  
 str\_detect(Offense, 'DECEPTIVE PRACTICE') ~ 'DECEPTIVE PRACTICE',  
 str\_detect(Offense, 'HOMICIDE') ~ 'HOMICIDE',  
 str\_detect(Offense, 'INTERFERENCE') ~ 'INTERFERENCE',  
 str\_detect(Offense, 'MOTOR VEHICLE THEFT') ~ 'MOTOR VEHICLE THEFT',  
 str\_detect(Offense, 'NARCOTICS') ~ 'NARCOTICS',  
 str\_detect(Offense, 'OFFENSE INVOLVING CHILDREN') ~ 'OFFENSE INVOLVING CHILDREN',  
 str\_detect(Offense, 'OTHER OFFENSE') ~ 'OTHER OFFENSE',  
 str\_detect(Offense, 'PROSTITUTION') ~ 'PROSTITUTION',  
 str\_detect(Offense, 'PUBLIC PEACE') ~ 'PUBLIC PEACE',  
 str\_detect(Offense, 'ROBBERY') ~ 'ROBBERY',  
 str\_detect(Offense, 'SEX OFFENSE') ~ 'SEX OFFENSE',  
 str\_detect(Offense, 'THEFT') ~ 'THEFT',  
 str\_detect(Offense, 'WEAPONS VIOLATION') ~ 'WEAPONS VIOLATION'  
 ))  
  
chi\_new$Offense\_new = as.factor(chi\_new$Offense\_new)  
  
#View(chi)  
  
chi\_new$`Location Description` = case\_when(  
 str\_detect(chi\_new$`Location Description`, 'SIDEWALK|STREET') ~ 'STREET',  
 str\_detect(chi\_new$`Location Description`, 'AUTO|VEHICLE') ~ 'VEHICLE',  
 str\_detect(chi\_new$`Location Description`, 'BANK') ~ 'BANK',  
 str\_detect(chi\_new$`Location Description`, 'BOAT|LAKE') ~ 'WATER\_LOCATION',  
 str\_detect(chi\_new$`Location Description`, 'BRIDGE') ~ 'BRIDGE',  
 str\_detect(chi\_new$`Location Description`, 'CHURCH') ~ 'PLACE OF WORSHIP',  
 str\_detect(chi\_new$`Location Description`, 'COLLEGE') ~ 'COLLEGE\_UNIVERSITY',  
 str\_detect(chi\_new$`Location Description`, 'CTA') ~ 'CTA\_LOCATION',  
 str\_detect(chi\_new$`Location Description`, 'FEDERAL|GOVERNMENT|POLICE|FIRE') ~ 'GOVT BUILDING',  
 str\_detect(chi\_new$`Location Description`, 'HIGHWAY') ~ 'HIGHWAY',  
 str\_detect(chi\_new$`Location Description`, 'JAIL') ~ 'JAIL',  
 str\_detect(chi\_new$`Location Description`, 'FOREST|PARK PROPERTY') ~ 'PARK',  
 str\_detect(chi\_new$`Location Description`, 'PARKING LOT') ~ 'PARKING LOT',  
 str\_detect(chi\_new$`Location Description`, 'SCHOOL') ~ 'SCHOOL',  
 str\_detect(chi\_new$`Location Description`, 'SPORTS') ~ 'SPORTING ARENA',  
 str\_detect(chi\_new$`Location Description`, 'TAXI') ~ 'TAXI',  
 str\_detect(chi\_new$`Location Description`, 'STORE|CLUB|BAR|BOWLING|CAR WASH|COMMERCIAL|CONSTRUCTION SITE|DAY CARE CENTER|FACTORY|GAS STATION|LIBRARY|THEATER|SHOP|RESTAURANT|SAVINGS AND LOAN|ATM|CURRENCY EXCHANGE') ~ 'STORE',  
 str\_detect(chi\_new$`Location Description`, 'APARTMENT|DRIVEWAY|HALLWAY|HOTEL|HOUSE|NURSING HOME|PORCH|POOL ROOM|RESIDENCE|RESIDENTIAL|YARD') ~ 'RESIDENTIAL',  
 str\_detect(chi\_new$`Location Description`, 'AIR') ~ 'AIRPORT',  
 str\_detect(chi\_new$`Location Description`, 'HOSPITAL|MEDICAL') ~ 'HOSPITAL',  
 str\_detect(chi\_new$`Location Description`, 'OTHER|ABANDONED BUILDING|ALLEY|VACANT|WAREHOUSE') ~ 'OTHER'  
 )  
  
colnames(chi\_new)[5]= 'location\_description'  
   
chi\_new$location\_description = as.factor(chi\_new$location\_description)  
  
  
chi\_new$District = case\_when(  
 chi\_new$District == '001' ~ 1,  
 chi\_new$District == '002' ~ 2,  
 chi\_new$District == '003' ~ 3,  
 chi\_new$District == '004' ~ 4,  
 chi\_new$District == '005' ~ 5,  
 chi\_new$District == '006' ~ 6,  
 chi\_new$District == '007' ~ 7,  
 chi\_new$District == '008' ~ 8,  
 chi\_new$District == '009' ~ 9,  
 chi\_new$District == '010' ~ 10,  
 chi\_new$District == '011' ~ 11,  
 chi\_new$District == '012' ~ 12,  
 chi\_new$District == '013' ~ 13,  
 chi\_new$District == '014' ~ 14,  
 chi\_new$District == '015' ~ 15,  
 chi\_new$District == '016' ~ 16,  
 chi\_new$District == '017' ~ 17,  
 chi\_new$District == '018' ~ 18,  
 chi\_new$District == '019' ~ 19,  
 chi\_new$District == '020' ~ 20,  
 chi\_new$District == '021' ~ 21,  
 chi\_new$District == '022' ~ 22,  
 chi\_new$District == '023' ~ 23,  
 chi\_new$District == '024' ~ 24,  
 chi\_new$District == '025' ~ 25,  
)  
  
chi\_new = chi\_new %>% group\_by(chi\_new$location\_description) %>% filter(n() > 20)  
  
chi\_new <- drop\_na(chi\_new)

## Plotting Data

offense\_arrest <- ggplot(chi\_new, aes(Offense\_new, fill = Arrest)) +  
 geom\_bar(aes(y=..count../sum(..count..))) +   
 theme\_gdocs() + theme(axis.text.x = element\_text(angle = 30)) +  
 scale\_y\_continuous(limits = c(0, .25), labels=percent\_format()) +  
 #scale\_y\_log10() +  
 scale\_fill\_manual(values = c("dark grey", "light blue")) +  
 labs(title= "Offense VS Arrest Status", x = "Offense", y ="Percent")   
#location\_arrest  
  
ggplotly(offense\_arrest, tooltip =FALSE)

hour\_arrest <- ggplot(chi\_new, aes(Arrest, fill = Hour)) +  
 geom\_bar(aes(y=..count../sum(..count..)), position = "dodge") +  
 #geom\_bar(position = "dodge") +  
 theme\_gdocs() +  
 scale\_fill\_brewer(type = "seq", palette = "YlOrRd") +   
 scale\_y\_continuous(limits = c(0, .30), labels=percent\_format()) +  
 #scale\_fill\_manual(values = c("orange", "light blue", "yellow", "dark grey")) +  
 labs(title = "Time of Offense VS Arrest Status", x="Arrest Status", y="Percent")   
  
hour\_arrest\_plotly <- ggplotly(hour\_arrest, tooltip = FALSE)   
hour\_arrest\_plotly

#   
# district\_arrest <- ggplot(chi\_new, aes(District, fill = Arrest)) +  
# geom\_bar(aes(y=..count../sum(..count))) + theme(axis.text.x = element\_text(vjust = 0.5,  
# angle = 90)) +  
# scale\_fill\_manual(values = c("dark grey", "light blue")) +  
# theme\_gdocs()+  
# scale\_x\_continuous(breaks = c(1:25, 0)) +  
# labs(title="Police District VS Arrest Status", y="Count", x="Police District")  
# district\_arrest  
  
quarter\_arrest <- ggplot(chi\_new, aes(Arrest, fill = Quarter)) +  
 geom\_bar(aes(y=..count../sum(..count..)), position = "dodge") +  
 #geom\_bar(position = "dodge") +  
 theme\_gdocs() +  
 scale\_fill\_brewer(type = "seq", palette = "YlOrRd") +   
 scale\_y\_continuous(limits = c(0, .30), labels=percent\_format()) +   
 #scale\_fill\_manual(values = c("orange", "light blue", "yellow", "dark grey")) +  
 labs(title = "Quarter VS Arrest Status", x= "Arrest Status", y ="Percent")  
quarter\_arrest\_plotly <- ggplotly(quarter\_arrest, tooltip = FALSE)   
quarter\_arrest\_plotly

location\_arrest <- ggplot(chi\_new, aes(location\_description, fill = Arrest)) +  
 geom\_bar(position = "dodge") +   
 theme\_gdocs() + theme(axis.text.x = element\_text(angle = 30)) +  
 scale\_fill\_manual(values = c("dark grey", "light blue")) +  
 scale\_y\_log10() +   
 labs(title= "blank title", x = "Offense Location", y ="Count", caption="Y-Axis on Log10 Scale")   
#location\_arrest  
# location\_arrest  
  
# test <- ggplot(chi\_new, aes(location\_description, fill = Arrest)) +  
# geom\_bar(aes(y=..count../sum(..count..))) +  
# scale\_y\_continuous(labels = percent\_format()) + theme(axis.text.x = element\_text(vjust = 0.5,   
# angle = 30))  
#   
# ggplotly(test)  
   
 location\_arrest\_plotly <- ggplotly(location\_arrest, tooltip = FALSE) %>%   
 layout(title = list(text = paste0('Offense Location VS Arrest Status',  
 '<br>',  
 '<sup>',  
 'Y-Axis on Log10 Scale',  
 '<br>',  
 '</sup>')))  
 location\_arrest\_plotly