## Daniel Sanchez

## BAN 502

## Dr. Stephen Hill

## Course Project - Phase 1

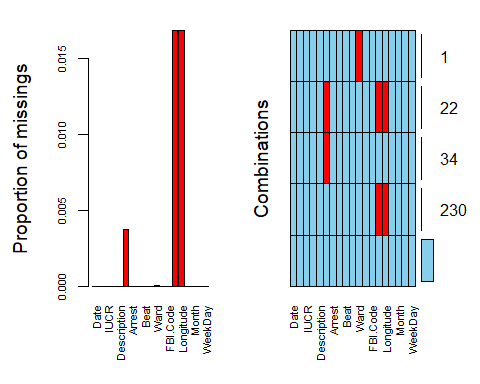
## June 15th, 2020

library(tidyverse)  
library(lubridate)  
library(mice)  
library(VIM)  
library(caret)

#data import and preparation  
chicago=read.csv("chicago2.csv")  
chicago=chicago%>%select(-ID,-X,-Case.Number,-Updated.On,-X.Coordinate,-Y.Coordinate,-Location,-Year)  
  
chicago = chicago %>% mutate(Date = mdy\_hms(Date))  
chicago = chicago %>% mutate(Hour = hour(Date))  
chicago = chicago %>% mutate(Month = month(Date))  
chicago = chicago %>% mutate(Day = day(Date))  
chicago = chicago %>% mutate(WeekDay = wday(Date))  
chicago = chicago %>% mutate(WeekDay = as\_factor(as.character(WeekDay)))%>%  
mutate(WeekDay = fct\_recode(WeekDay,  
"Sunday" = "1",  
"Monday" = "2",  
"Tuesday" = "3",  
"Wednesday" = "4",  
"Thursday" = "5",  
"Friday" = "6",  
"Saturday" = "7"))  
chicago = chicago %>% mutate(Beat = as\_factor(as.character(Beat)))  
chicago = chicago %>% mutate(Primary.Type = as\_factor(as.character(Primary.Type)))  
chicago = chicago %>% mutate(District = as\_factor(as.character(District)))  
chicago = chicago %>% mutate(Ward = as\_factor(as.character(Ward)))  
chicago = chicago %>% mutate(Block = as\_factor(as.character(Block)))  
chicago = chicago %>% mutate(Description = as\_factor(as.character(Description)))  
chicago = chicago %>% mutate(Location.Description = as\_factor(as.character(Location.Description)))  
chicago = chicago %>% mutate(IUCR = as\_factor(as.character(IUCR)))  
chicago = chicago %>% mutate(FBI.Code = as\_factor(as.character(FBI.Code)))  
chicago = chicago %>% mutate(Community.Area = as\_factor(as.character(Community.Area)))  
chicago = chicago %>% mutate(Arrest = as\_factor(as.character(Arrest)))%>%  
mutate(Arrest = fct\_recode(Arrest,  
"No Arrest" = "FALSE",  
"Arrest" = "TRUE"))  
chicago = chicago %>% mutate(Domestic = as\_factor(as.character(Domestic)))%>%  
mutate(Domestic = fct\_recode(Domestic,  
"Non Domestic" = "FALSE",  
"Domestic Violence" = "TRUE"))  
str(chicago)

## 'data.frame': 15000 obs. of 19 variables:  
## $ Date : POSIXct, format: "2018-09-13 14:00:00" "2018-07-06 23:00:00" ...  
## $ Block : Factor w/ 8940 levels "050XX S LAWNDALE AVE",..: 1 2 3 4 5 6 7 8 9 10 ...  
## $ IUCR : Factor w/ 236 levels "1320","0460",..: 1 2 3 4 5 3 6 7 8 4 ...  
## $ Primary.Type : Factor w/ 29 levels "CRIMINAL DAMAGE",..: 1 2 1 3 4 1 5 2 6 3 ...  
## $ Description : Factor w/ 221 levels "TO VEHICLE","SIMPLE",..: 1 2 3 4 5 3 6 7 8 4 ...  
## $ Location.Description: Factor w/ 105 levels "STREET","APARTMENT",..: 1 1 2 1 1 3 4 2 5 3 ...  
## $ Arrest : Factor w/ 2 levels "No Arrest","Arrest": 1 1 1 1 1 1 1 1 2 1 ...  
## $ Domestic : Factor w/ 2 levels "Non Domestic",..: 1 1 1 1 1 1 1 2 1 1 ...  
## $ Beat : Factor w/ 274 levels "821","1023","823",..: 1 2 3 4 5 6 1 7 8 9 ...  
## $ District : Factor w/ 22 levels "8","10","2","5",..: 1 2 1 2 3 4 1 5 6 4 ...  
## $ Ward : Factor w/ 50 levels "14","12","15",..: 1 2 3 4 5 6 1 7 8 6 ...  
## $ Community.Area : Factor w/ 78 levels "62","30","66",..: 1 2 3 4 5 6 7 8 9 10 ...  
## $ FBI.Code : Factor w/ 25 levels "14","08B","06",..: 1 2 1 3 4 1 5 2 6 3 ...  
## $ Latitude : num 41.8 41.9 41.8 41.9 41.8 ...  
## $ Longitude : num -87.7 -87.7 -87.7 -87.7 -87.6 ...  
## $ Hour : int 14 23 2 5 7 17 13 10 12 8 ...  
## $ Month : num 9 7 6 8 11 6 9 3 4 4 ...  
## $ Day : int 13 6 25 9 2 25 19 12 10 1 ...  
## $ WeekDay : Factor w/ 7 levels "Thursday","Friday",..: 1 2 3 1 2 3 4 3 5 6 ...

#check for missingness  
vim\_plot = aggr(chicago, numbers = TRUE, prop = c(TRUE, FALSE),cex.axis=.7)

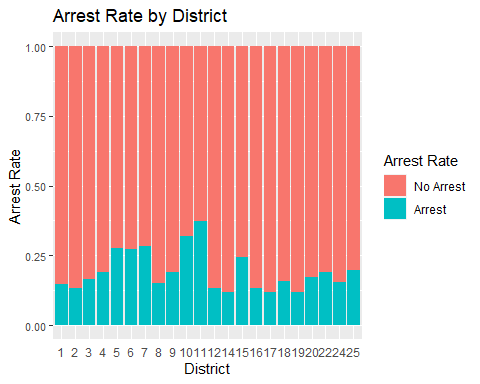


#row-wise deletion of na values  
chicago = chicago%>%drop\_na()  
  
#remove crime types with low numbers  
chicago= chicago%>%filter(Primary.Type!="PUBLIC INDECENCY",Primary.Type!="NON-CRIMINAL",Primary.Type!="STALKING",Primary.Type!="OBSCENITY",Primary.Type!="KIDNAPPING",Primary.Type!="CONCEALED CARRY LICENSE VIOLATION",Primary.Type!="INTIMIDATION",Primary.Type!="GAMBLING",Primary.Type!="ARSON",Primary.Type!="LIQUOR LAW VIOLATION")

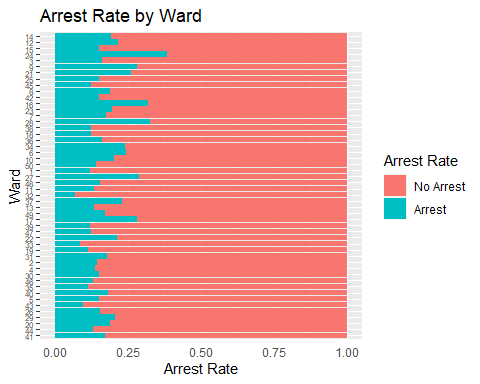
#testing importance of variables  
chicago\_rf\_fit=chicago%>%select(-Date,-Location.Description,-Block,-IUCR,-Community.Area,-Latitude,-Longitude,-Description,-Day,-Beat,-WeekDay)  
fit\_control = trainControl(method = "cv",   
 number = 10)  
set.seed(123)   
rf\_fit = train(x=chicago\_rf\_fit[,-3], y=chicago\_rf\_fit$Arrest,  
 method = "ranger",   
 importance = "permutation",  
 trControl = fit\_control,  
 num.trees = 100)  
  
varImp(rf\_fit)

## ranger variable importance  
##   
## Overall  
## Arrest 1.000e+02  
## FBI.Code 4.877e+00  
## Primary.Type 4.753e+00  
## Ward 7.906e-02  
## District 7.598e-02  
## Hour 6.589e-03  
## Month 0.000e+00

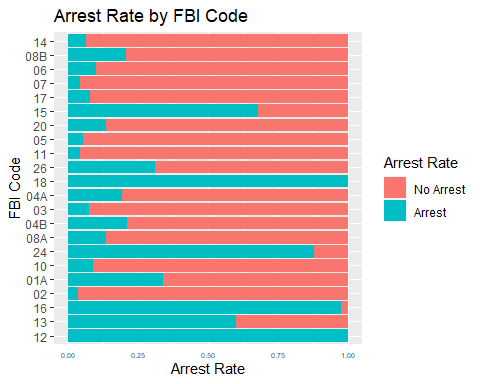
#Charts of important variables  
ggplot(chicago,aes(x=fct\_inseq(factor(District)), fill=Arrest))+  
 geom\_bar(position="fill")+  
 labs(x = "District", y = "Arrest Rate", title = "Arrest Rate by District", fill = "Arrest Rate")+  
 theme(axis.text.y = element\_text(size = 8), axis.ticks.x = element\_blank())



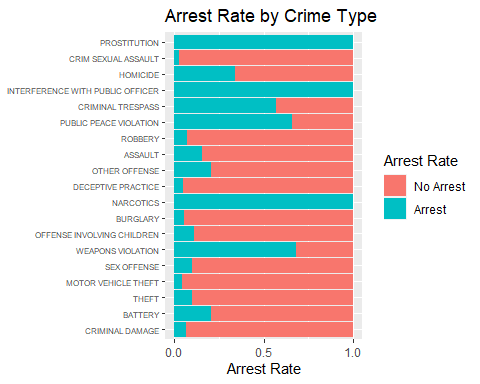
ggplot(chicago,aes(x=fct\_rev(fct\_inorder(factor(Ward))), fill=Arrest))+  
 geom\_bar(position="fill")+  
 coord\_flip()+  
 labs(x = "Ward", y = "Arrest Rate", title = "Arrest Rate by Ward", fill = "Arrest Rate")+  
 theme(axis.text.y = element\_text(size = 6), axis.ticks.x = element\_blank())



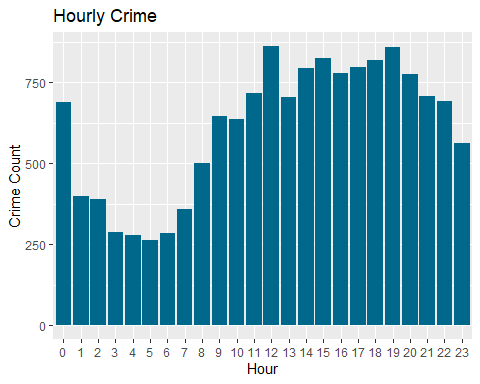
ggplot(chicago,aes(x=fct\_rev(fct\_inorder(factor(FBI.Code))), fill=Arrest))+  
 geom\_bar(position = "fill") +  
 coord\_flip()+  
 theme(axis.text.x = element\_text(size = 5))+  
 labs(x = "FBI Code", y=("Arrest Rate"), title = "Arrest Rate by FBI Code", fill = "Arrest Rate")+  
 theme(axis.ticks.x = element\_blank())



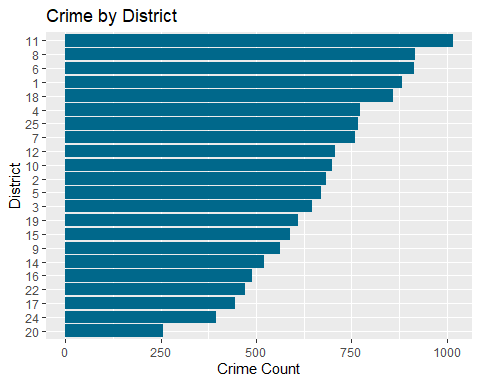
ggplot(chicago,aes(x=Primary.Type, fill=Arrest)) +  
 geom\_bar(position = "fill") +  
 coord\_flip()+  
 theme(axis.text.y = element\_text(size = 6))+  
 scale\_y\_continuous(breaks=seq(0,1,1/2))+  
 labs(x = NULL,y = "Arrest Rate", title = "Arrest Rate by Crime Type", fill = "Arrest Rate")



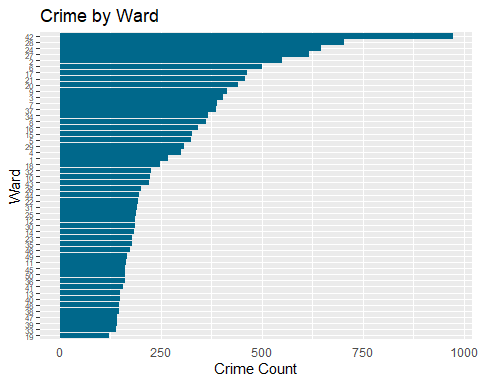
#Other descriptive factors  
ggplot(chicago,aes(x=fct\_inseq(factor(Hour))))+  
 geom\_bar(fill="deepskyblue4")+  
 labs(x = "Hour", y = "Crime Count", title = "Hourly Crime")



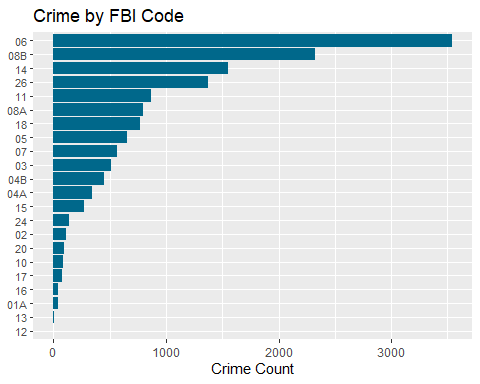
ggplot(chicago,aes(x=fct\_rev(fct\_infreq(factor(District)))))+  
 geom\_bar(fill="deepskyblue4")+  
 coord\_flip()+  
 labs(x = "District", y = "Crime Count", title = "Crime by District")



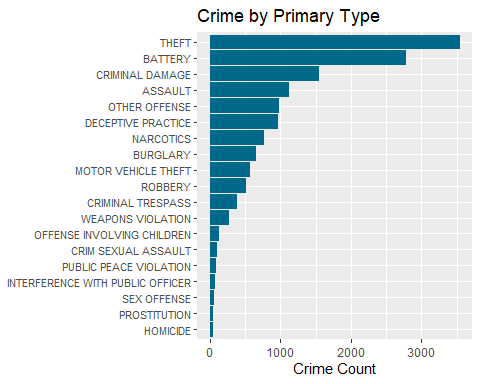
ggplot(chicago,aes(x=fct\_rev(fct\_infreq(factor(Ward)))))+  
 geom\_bar(fill="deepskyblue4")+  
 coord\_flip()+  
 labs(x = "Ward", y = "Crime Count", title = "Crime by Ward")+  
 theme(axis.text.y = element\_text(size = 6), axis.ticks.x = element\_blank())



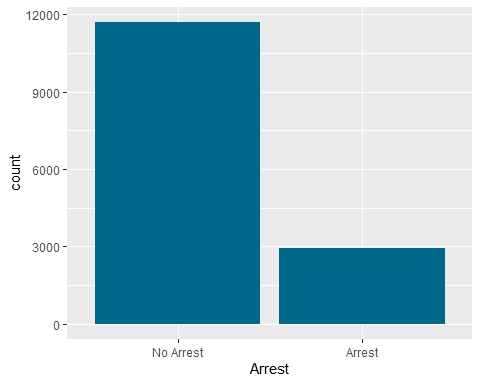
ggplot(chicago,aes(x=fct\_rev(fct\_infreq(factor(FBI.Code)))))+  
 geom\_bar(fill="deepskyblue4")+  
 coord\_flip()+  
 theme(axis.text.y = element\_text(size = 8))+  
 labs(x = NULL, y = "Crime Count", title = "Crime by FBI Code")



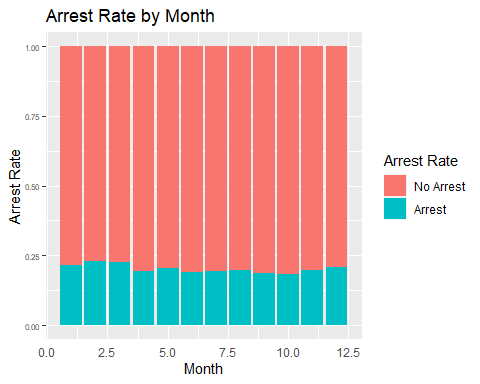
ggplot(chicago,aes(x=fct\_rev(fct\_infreq(factor(Primary.Type)))))+  
 geom\_bar(fill="deepskyblue4")+  
 coord\_flip()+  
 theme(axis.text.y = element\_text(size = 8))+  
 labs(x = NULL, y = "Crime Count", title = "Crime by Primary Type")



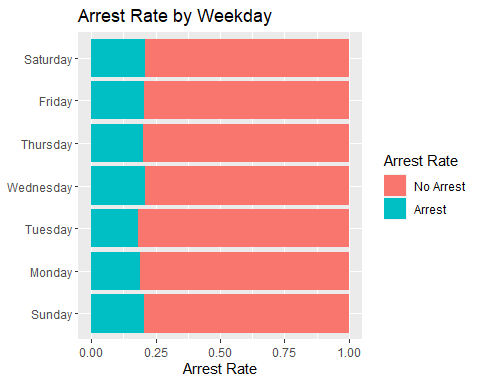
ggplot(chicago,aes(x=Arrest))+  
 geom\_bar(fill="deepskyblue4")



ggplot(chicago,aes(x=Month, fill=Arrest))+  
 geom\_bar(position="fill")+  
 labs(x = "Month", y = "Arrest Rate", title = "Arrest Rate by Month", fill = "Arrest Rate")+  
 theme(axis.text.y = element\_text(size = 6), axis.ticks.x = element\_blank())



chicago%>%  
 mutate(Weekday = fct\_relevel(WeekDay,   
 "Sunday", "Monday", "Tuesday",   
 "Wednesday", "Thursday", "Friday",   
 "Saturday"))%>%  
ggplot(aes(x=Weekday, fill=Arrest))+  
 geom\_bar(position="fill")+  
 coord\_flip()+  
 labs(x = NULL, y = "Arrest Rate", title = "Arrest Rate by Weekday", fill = "Arrest Rate")



ggplot(chicago,aes(x=Hour, fill=Arrest))+  
 geom\_bar(position="fill")+  
 labs(x = "Hour", y = "Arrest Rate", title = "Arrest Rate by Hour", fill = "Arrest Rate")

