Marci B Copeland BAN502 Project 1

library (tidyverse)

## -- Attaching packages --------------------------------------- tidyverse 1.2.1 --

## v ggplot2 3.2.1 v purrr 0.3.3  
## v tibble 2.1.3 v dplyr 0.8.3  
## v tidyr 1.0.0 v stringr 1.4.0  
## v readr 1.3.1 v forcats 0.4.0

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

library (lubridate)

##   
## Attaching package: 'lubridate'

## The following object is masked from 'package:base':  
##   
## date

library (MASS)

##   
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':  
##   
## select

library (ROCR)

## Warning: package 'ROCR' was built under R version 3.6.2

## Loading required package: gplots

## Warning: package 'gplots' was built under R version 3.6.2

##   
## Attaching package: 'gplots'

## The following object is masked from 'package:stats':  
##   
## lowess

library (caret)

## Warning: package 'caret' was built under R version 3.6.2

## Loading required package: lattice

##   
## Attaching package: 'caret'

## The following object is masked from 'package:purrr':  
##   
## lift

library (GGally)

## Warning: package 'GGally' was built under R version 3.6.2

## Registered S3 method overwritten by 'GGally':  
## method from   
## +.gg ggplot2

##   
## Attaching package: 'GGally'

## The following object is masked from 'package:dplyr':  
##   
## nasa

library (VIM)

## Warning: package 'VIM' was built under R version 3.6.2

## Loading required package: colorspace

## Loading required package: grid

## Loading required package: data.table

## Warning: package 'data.table' was built under R version 3.6.2

##   
## Attaching package: 'data.table'

## The following objects are masked from 'package:lubridate':  
##   
## hour, isoweek, mday, minute, month, quarter, second, wday,  
## week, yday, year

## The following objects are masked from 'package:dplyr':  
##   
## between, first, last

## The following object is masked from 'package:purrr':  
##   
## transpose

## VIM is ready to use.   
## Since version 4.0.0 the GUI is in its own package VIMGUI.  
##   
## Please use the package to use the new (and old) GUI.

## Suggestions and bug-reports can be submitted at: https://github.com/alexkowa/VIM/issues

##   
## Attaching package: 'VIM'

## The following object is masked from 'package:datasets':  
##   
## sleep

library (mice)

## Warning: package 'mice' was built under R version 3.6.2

## Registered S3 methods overwritten by 'lme4':  
## method from  
## cooks.distance.influence.merMod car   
## influence.merMod car   
## dfbeta.influence.merMod car   
## dfbetas.influence.merMod car

##   
## Attaching package: 'mice'

## The following object is masked from 'package:tidyr':  
##   
## complete

## The following objects are masked from 'package:base':  
##   
## cbind, rbind

library (rpart)

## Warning: package 'rpart' was built under R version 3.6.2

library (rattle)

## Warning: package 'rattle' was built under R version 3.6.2

## Rattle: A free graphical interface for data science with R.  
## Version 5.3.0 Copyright (c) 2006-2018 Togaware Pty Ltd.  
## Type 'rattle()' to shake, rattle, and roll your data.

##   
## Attaching package: 'rattle'

## The following object is masked from 'package:VIM':  
##   
## wine

library (RColorBrewer)

chicago<-read\_csv("chicago.csv")

## Parsed with column specification:  
## cols(  
## .default = col\_character(),  
## ID = col\_double(),  
## Arrest = col\_logical(),  
## Domestic = col\_logical(),  
## Beat = col\_double(),  
## District = col\_double(),  
## 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.

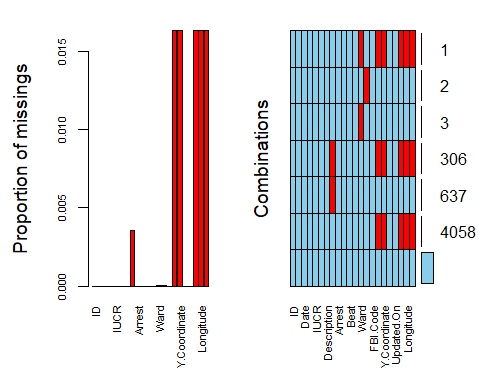
str(chicago)

## Classes 'spec\_tbl\_df', 'tbl\_df', 'tbl' and 'data.frame': 267185 obs. of 22 variables:  
## $ ID : num 11948745 11945748 11944351 11933173 11924359 ...  
## $ Case.Number : chr "JD112418" "JD108838" "JD107176" "JC561174" ...  
## $ Date : chr "1/1/2018 0:00" "1/1/2018 0:00" "1/1/2018 0:00" "1/1/2018 0:00" ...  
## $ Block : chr "069XX N CLARK ST" "070XX N KEDZIE AVE" "072XX W BALMORAL AVE" "047XX N ARTESIAN AVE" ...  
## $ IUCR : chr "1753" "1130" "1153" "1752" ...  
## $ Primary.Type : chr "OFFENSE INVOLVING CHILDREN" "DECEPTIVE PRACTICE" "DECEPTIVE PRACTICE" "OFFENSE INVOLVING CHILDREN" ...  
## $ Description : chr "SEX ASSLT OF CHILD BY FAM MBR" "FRAUD OR CONFIDENCE GAME" "FINANCIAL IDENTITY THEFT OVER $ 300" "AGG CRIM SEX ABUSE FAM MEMBER" ...  
## $ Location.Description: chr "RESIDENCE-GARAGE" "APARTMENT" "RESIDENCE" "RESIDENCE" ...  
## $ Arrest : logi FALSE FALSE FALSE FALSE FALSE FALSE ...  
## $ Domestic : logi FALSE FALSE FALSE TRUE FALSE FALSE ...  
## $ Beat : num 2431 2411 1613 1911 932 ...  
## $ District : num 24 24 16 19 9 22 12 7 9 16 ...  
## $ Ward : num 49 50 41 40 16 21 27 15 20 41 ...  
## $ Community.Area : num 1 2 10 4 61 73 28 67 61 76 ...  
## $ FBI.Code : chr "2" "11" "11" "17" ...  
## $ X.Coordinate : num NA NA NA NA NA ...  
## $ Y.Coordinate : num NA NA NA NA NA ...  
## $ Year : num 2018 2018 2018 2018 2018 ...  
## $ Updated.On : chr "1/12/2020 15:48" "1/9/2020 15:57" "1/8/2020 15:49" "1/4/2020 15:44" ...  
## $ Latitude : num NA NA NA NA NA ...  
## $ Longitude : num NA NA NA NA NA ...  
## $ Location : chr NA NA NA NA ...  
## - attr(\*, "spec")=  
## .. cols(  
## .. ID = col\_double(),  
## .. Case.Number = col\_character(),  
## .. Date = col\_character(),  
## .. Block = col\_character(),  
## .. IUCR = col\_character(),  
## .. Primary.Type = col\_character(),  
## .. Description = col\_character(),  
## .. Location.Description = col\_character(),  
## .. Arrest = col\_logical(),  
## .. Domestic = col\_logical(),  
## .. Beat = col\_double(),  
## .. District = col\_double(),  
## .. Ward = col\_double(),  
## .. Community.Area = col\_double(),  
## .. FBI.Code = col\_character(),  
## .. X.Coordinate = col\_double(),  
## .. Y.Coordinate = col\_double(),  
## .. Year = col\_double(),  
## .. Updated.On = col\_character(),  
## .. Latitude = col\_double(),  
## .. Longitude = col\_double(),  
## .. Location = col\_character()  
## .. )

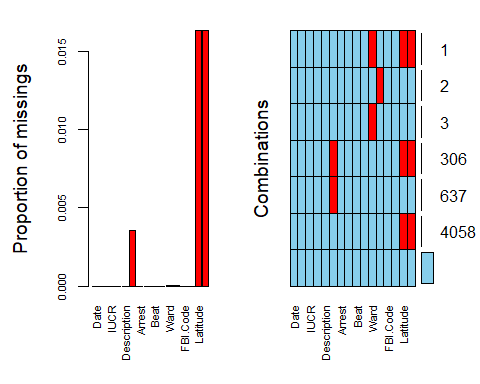
summary(chicago)

## ID Case.Number Date   
## Min. : 23757 Length:267185 Length:267185   
## 1st Qu.:11286550 Class :character Class :character   
## Median :11375979 Mode :character Mode :character   
## Mean :11352463   
## 3rd Qu.:11465374   
## Max. :11968996   
##   
## Block IUCR Primary.Type   
## Length:267185 Length:267185 Length:267185   
## Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character   
##   
##   
##   
##   
## Description Location.Description Arrest Domestic   
## Length:267185 Length:267185 Mode :logical Mode :logical   
## Class :character Class :character FALSE:213769 FALSE:223427   
## Mode :character Mode :character TRUE :53416 TRUE :43758   
##   
##   
##   
##   
## Beat District Ward Community.Area   
## Min. : 111 Min. : 1.0 Min. : 1.00 Min. : 0.00   
## 1st Qu.: 611 1st Qu.: 6.0 1st Qu.:10.00 1st Qu.:23.00   
## Median :1031 Median :10.0 Median :24.00 Median :32.00   
## Mean :1143 Mean :11.2 Mean :23.45 Mean :36.47   
## 3rd Qu.:1723 3rd Qu.:17.0 3rd Qu.:35.00 3rd Qu.:53.00   
## Max. :2535 Max. :31.0 Max. :50.00 Max. :77.00   
## NA's :4 NA's :2   
## FBI.Code X.Coordinate Y.Coordinate Year   
## Length:267185 Min. :1092706 Min. :1813897 Min. :2018   
## Class :character 1st Qu.:1153330 1st Qu.:1859425 1st Qu.:2018   
## Mode :character Median :1166911 Median :1894254 Median :2018   
## Mean :1165006 Mean :1886600 Mean :2018   
## 3rd Qu.:1176455 3rd Qu.:1908739 3rd Qu.:2018   
## Max. :1205119 Max. :1951535 Max. :2018   
## NA's :4365 NA's :4365   
## Updated.On Latitude Longitude Location   
## Length:267185 Min. :41.65 Min. :-87.93 Length:267185   
## Class :character 1st Qu.:41.77 1st Qu.:-87.71 Class :character   
## Mode :character Median :41.87 Median :-87.66 Mode :character   
## Mean :41.84 Mean :-87.67   
## 3rd Qu.:41.91 3rd Qu.:-87.63   
## Max. :42.02 Max. :-87.53   
## NA's :4365 NA's :4365

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



chicago <- subset(chicago, select = -c(ID,Case.Number,X.Coordinate,Y.Coordinate,Updated.On,Location))  
vim\_plot = aggr(chicago, numbers = TRUE, prop = c(TRUE, FALSE),cex.axis=.7)



chicago = chicago %>% mutate(Date = mdy\_hms(Date))

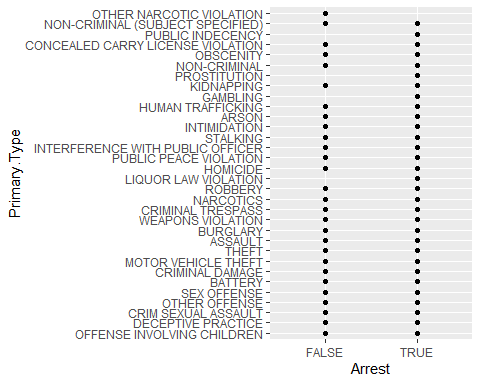
chicago = chicago %>% mutate(Hour = hour(Date)) #creates new variable in dataset with the  
#hour extracted from each date/time.  
#See the lubridate package reference above for the other time objects that can be extracted  
summary(chicago$Hour) #shows summary of this new variable

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 18 18 18 18 18 18

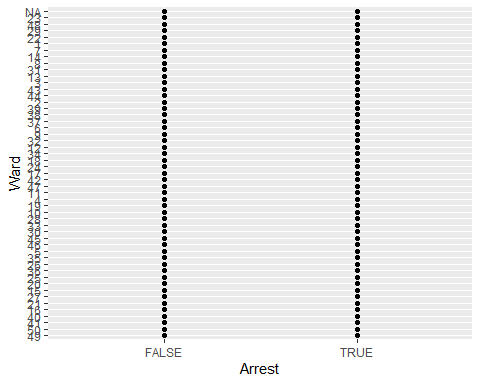
chicago = chicago %>%  
 mutate(`Primary.Type` = as\_factor(as.character(`Primary.Type`))) %>%  
 mutate(`Description` = as\_factor(as.character(`Description`))) %>%  
 mutate(`Location.Description` = as\_factor(as.character(`Location.Description`)))%>%  
 mutate(`Arrest` = as\_factor(as.character(`Arrest`)))%>%  
 mutate(`Domestic` = as\_factor(as.character(`Domestic`)))%>%  
 mutate(`Beat` = as\_factor(as.character(`Ward`)))%>%  
 mutate(`District` = as\_factor(as.character(`Ward`)))%>%  
 mutate(`Ward` = as\_factor(as.character(`Ward`)))%>%  
 mutate(`Community.Area` = as\_factor(as.character(`Community.Area`)))%>%  
 mutate(`FBI.Code` = as\_factor(as.character(`FBI.Code`)))%>%  
 mutate(`Hour` = as\_factor(as.character(`Hour`)))  
str(chicago)

## Classes 'tbl\_df', 'tbl' and 'data.frame': 267185 obs. of 17 variables:  
## $ Date : POSIXct, format: "2020-01-01 18:00:00" "2020-01-01 18:00:00" ...  
## $ Block : chr "069XX N CLARK ST" "070XX N KEDZIE AVE" "072XX W BALMORAL AVE" "047XX N ARTESIAN AVE" ...  
## $ IUCR : chr "1753" "1130" "1153" "1752" ...  
## $ Primary.Type : Factor w/ 32 levels "OFFENSE INVOLVING CHILDREN",..: 1 2 2 1 1 3 3 1 1 4 ...  
## $ Description : Factor w/ 301 levels "SEX ASSLT OF CHILD BY FAM MBR",..: 1 2 3 4 4 5 6 4 4 7 ...  
## $ Location.Description: Factor w/ 132 levels "RESIDENCE-GARAGE",..: 1 2 3 3 3 4 4 3 2 3 ...  
## $ Arrest : Factor w/ 2 levels "FALSE","TRUE": 1 1 1 1 1 1 1 2 1 1 ...  
## $ Domestic : Factor w/ 2 levels "FALSE","TRUE": 1 1 1 2 1 1 1 1 1 1 ...  
## $ Beat : Factor w/ 50 levels "49","50","41",..: 1 2 3 4 5 6 7 8 9 3 ...  
## $ District : Factor w/ 50 levels "49","50","41",..: 1 2 3 4 5 6 7 8 9 3 ...  
## $ Ward : Factor w/ 50 levels "49","50","41",..: 1 2 3 4 5 6 7 8 9 3 ...  
## $ Community.Area : Factor w/ 78 levels "1","2","10","4",..: 1 2 3 4 5 6 7 8 5 9 ...  
## $ FBI.Code : Factor w/ 26 levels "2","11","17",..: 1 2 2 3 3 1 1 3 3 4 ...  
## $ Year : num 2018 2018 2018 2018 2018 ...  
## $ Latitude : num NA NA NA NA NA ...  
## $ Longitude : num NA NA NA NA NA ...  
## $ Hour : Factor w/ 1 level "18": 1 1 1 1 1 1 1 1 1 1 ...

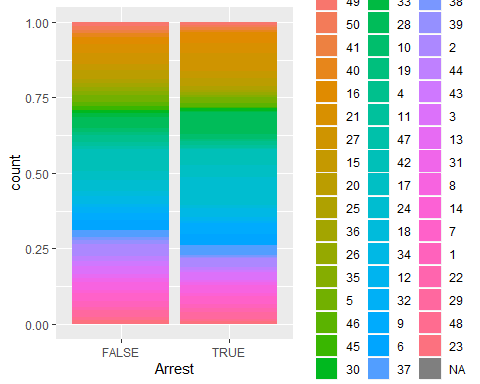
ggplot(chicago,aes(x=Arrest,y=Primary.Type))+  
 geom\_point()+  
 geom\_smooth(method=lm, se = TRUE)



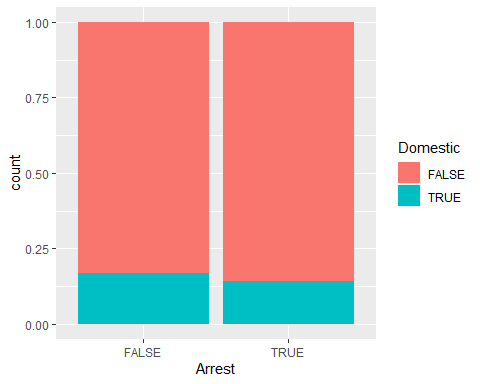
ggplot(chicago,aes(x=Arrest,y=Ward)) +  
 geom\_point()



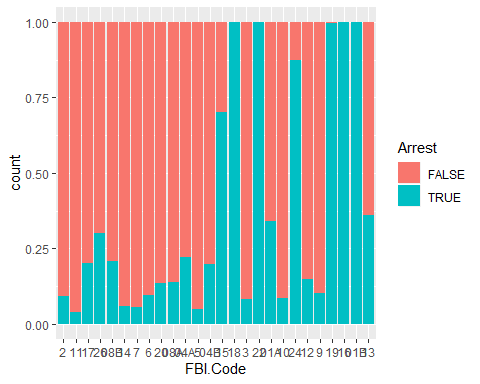
ggplot(chicago, aes(x=Arrest, fill=Ward)) + geom\_bar(position="fill")



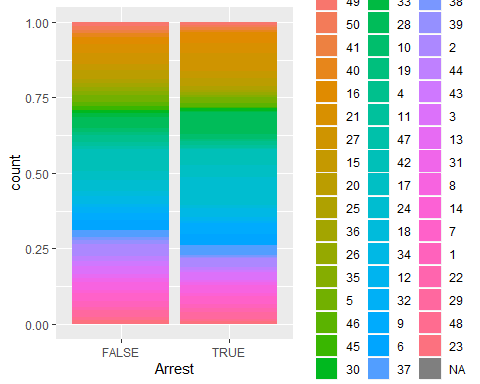
ggplot(chicago, aes(x=Arrest, fill=Domestic)) + geom\_bar(position="fill")



ggplot(chicago, aes(x=FBI.Code, fill=Arrest)) + geom\_bar(position="fill")



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



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

