# Crime in Chicago, IL.

## Emily Watson

library(tidyverse)

## -- Attaching packages ------------------------------------------ tidyverse 1.3.0 --

## v ggplot2 3.3.0 v purrr 0.3.3  
## v tibble 2.1.3 v dplyr 0.8.5  
## v tidyr 1.0.2 v stringr 1.4.0  
## v readr 1.3.1 v forcats 0.5.0

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

library(caret)

## Loading required package: lattice

##   
## Attaching package: 'caret'

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

library(ranger)  
library(rpart)  
library(rattle)

## Loading required package: bitops

## Rattle: A free graphical interface for data science with R.  
## Version 5.4.0 Copyright (c) 2006-2020 Togaware Pty Ltd.  
## Type 'rattle()' to shake, rattle, and roll your data.

##   
## Attaching package: 'rattle'

## The following object is masked from 'package:ranger':  
##   
## importance

library(RColorBrewer)  
library(VIM)

## Loading required package: colorspace

## Loading required package: grid

## Loading required package: data.table

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

## 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:rattle':  
##   
## wine

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

library(mice)

##   
## Attaching package: 'mice'

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

library(MASS)

##   
## Attaching package: 'MASS'

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

library(ROCR)  
library(GGally)

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

##   
## Attaching package: 'GGally'

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

library(car)

## Loading required package: carData

##   
## Attaching package: 'car'

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

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

library(lmtest)

## Loading required package: zoo

##   
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':  
##   
## as.Date, as.Date.numeric

chicago <- 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.

View(chicago)

chicago <- dplyr::select(chicago, "X1", "Date", "Block", "IUCR", "Primary Type", "Description", "Location Description", "Arrest", "Domestic", "Beat", "District", "Ward", "Community Area", "FBI Code", "Year", "Latitude", "Longitude")

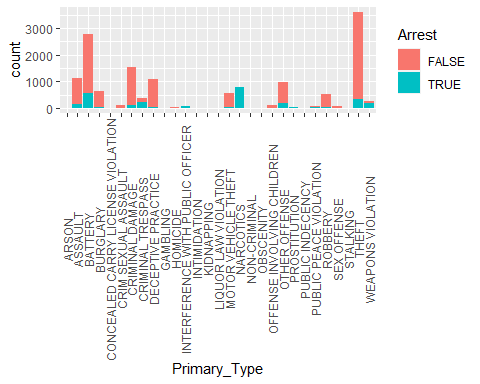
chicago = chicago %>% mutate(Date = lubridate::mdy\_hms(Date))  
chicago = chicago %>% mutate(Hour = lubridate::hour(Date))  
chicago = rename(chicago, "Primary\_Type" = "Primary Type")  
chicago = rename(chicago, "Location\_Description" = "Location Description")  
chicago = rename(chicago, "FBI\_Code" = "FBI Code")

chicago = chicago %>% mutate(Beat = as.character(as.factor(Beat)))  
chicago = chicago %>% mutate(District = as.character(as.factor(District)))  
chicago = chicago %>% mutate(Hour = as.character(as.factor(Hour)))  
chicago = chicago %>% mutate(Ward = as.character(as.factor(Ward)))  
chicago = chicago %>% mutate(`Community Area` = as.character(as.factor(`Community Area`)))  
chicago = chicago %>% mutate(FBI\_Code = as.character(as.factor(FBI\_Code)))  
chicago = chicago %>% mutate(IUCR = as.character(as.factor(IUCR)))  
chicago = chicago %>% mutate(IUCR = as.character(as.factor(IUCR)))  
chicago = chicago %>% mutate(Location\_Description = as.character(as.factor(Location\_Description)))

Shows primary crime type and what proportion resulted in arrest. High amount of theft, low proportion of arrests.

ggplot(chicago, aes(x = Primary\_Type, fill = Arrest)) + geom\_histogram(stat = "count") + theme(axis.text.x = element\_text(angle = 90, vjust=0.5))

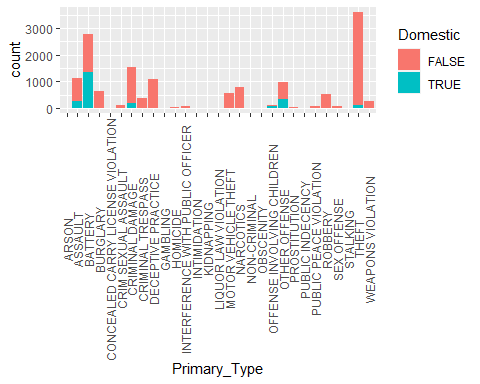
## Warning: Ignoring unknown parameters: binwidth, bins, pad



Shows primary crime type and what proportion were domestic.Very low proporation of domestic crimes spread out among type. Battery has highest proportion in domestic related issues.

ggplot(chicago, aes(x = Primary\_Type, fill = Domestic)) + geom\_histogram(stat = "count") + theme(axis.text.x = element\_text(angle = 90, vjust=0.5))

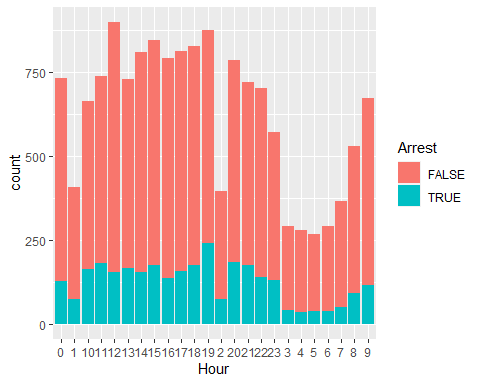
## Warning: Ignoring unknown parameters: binwidth, bins, pad



Shows what hours the crimes were committed and what proportion resulted in Arrest.12PM -7PM are hours with most reported crime, all hours have lower arrests than 7 PM.

ggplot(chicago, aes(x = Hour, fill = Arrest)) + geom\_histogram(stat = "count")

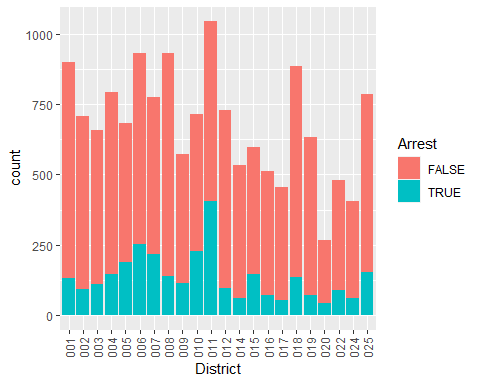
## Warning: Ignoring unknown parameters: binwidth, bins, pad



Shows what crime count by district and what proportion resulted in Arrest. Distrct 11 has highest amount of crime and by far highest amounts of arrests.

ggplot(chicago, aes(x = District, fill = Arrest)) + geom\_histogram(stat = "count") + theme(axis.text.x = element\_text(angle = 90, vjust=0.5))

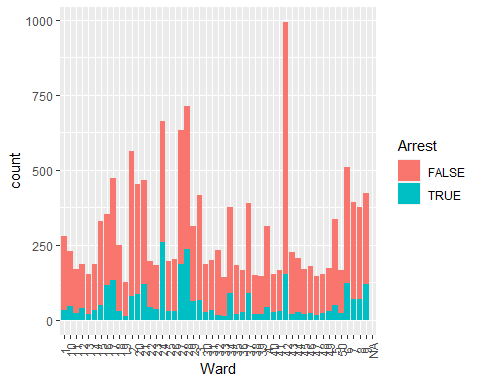
## Warning: Ignoring unknown parameters: binwidth, bins, pad



Shows the multiple places-location descriptions where crime was committed and of those what proportion resulted in arrrest. Ward 42 has highest amount of crime, but not nearly the highest amount of arrests.

ggplot(chicago, aes(x = Ward, fill = Arrest)) + geom\_histogram(stat = "count") + theme(axis.text.x = element\_text(angle = 90, vjust=0.5))

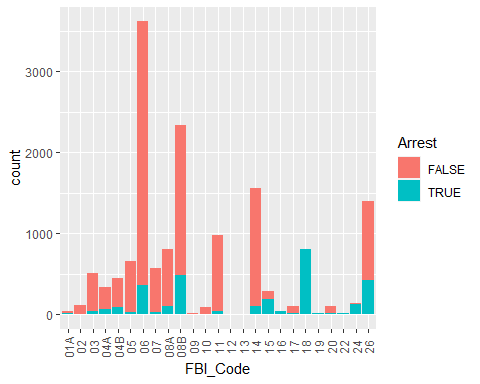
## Warning: Ignoring unknown parameters: binwidth, bins, pad



Show FBI code counts and proportion of arrests. Codes correspond with crime type as listed above. Larceny is the highest reported, proportionally few arrests. Narcotic related crimes have highest arrests (all arrested).

ggplot(chicago, aes(x = FBI\_Code, fill = Arrest)) + geom\_histogram(stat = "count") + theme(axis.text.x = element\_text(angle = 90, vjust=0.5))

## Warning: Ignoring unknown parameters: binwidth, bins, pad

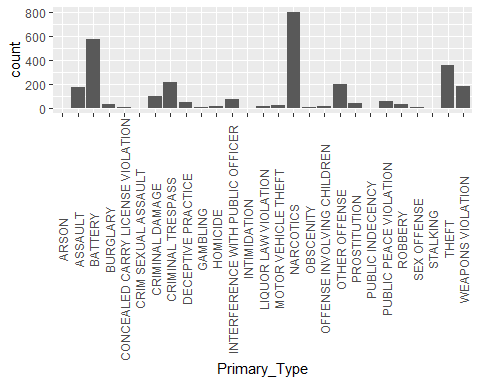


chicago\_arrests <- chicago %>% filter(Arrest %in% (TRUE))

Shows of what arrests made, what types of crimes there were.Battery and narcotics have highest arrests (proof on victim or perpertrator).

ggplot(chicago\_arrests, aes(x = Primary\_Type)) + geom\_histogram(stat = "count") + theme(axis.text.x = element\_text(angle = 90, vjust=0.5))

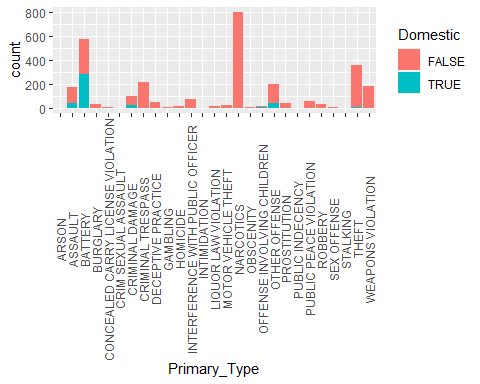
## Warning: Ignoring unknown parameters: binwidth, bins, pad



Shows the count of arrests and what proportion were domestic crimes.

ggplot(chicago\_arrests, aes(x = Primary\_Type, fill = Domestic)) + geom\_histogram(stat = "count") + theme(axis.text.x = element\_text(angle = 90, vjust=0.5))

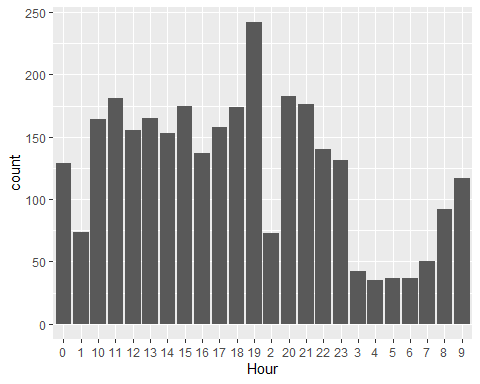
## Warning: Ignoring unknown parameters: binwidth, bins, pad



Shows of arrests made which hour the most arrests were made. Most arrests made at 7PM. Almost 100 more than next highest hour.

ggplot(chicago\_arrests, aes(x = Hour)) + geom\_histogram(stat = "count")

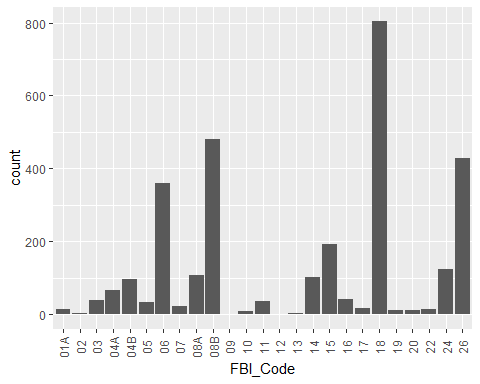
## Warning: Ignoring unknown parameters: binwidth, bins, pad



Code 18 has highest arrests=Narcotics.

ggplot(chicago\_arrests, aes(x = FBI\_Code)) + geom\_histogram(stat = "count") + theme(axis.text.x = element\_text(angle = 90, vjust=0.5))

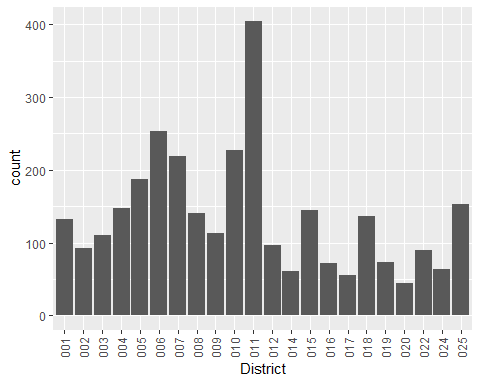
## Warning: Ignoring unknown parameters: binwidth, bins, pad



Shows of count of arrests, which district had the highest count.District 11.

ggplot(chicago\_arrests, aes(x = District)) + geom\_histogram(stat = "count") + theme(axis.text.x = element\_text(angle = 90, vjust=0.5))

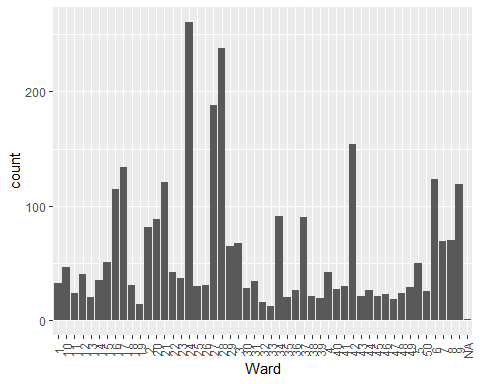
## Warning: Ignoring unknown parameters: binwidth, bins, pad



Shows of which wards, which had the highest arrests, 24 & 28.

ggplot(chicago\_arrests, aes(x = Ward)) + geom\_histogram(stat = "count") + theme(axis.text.x = element\_text(angle = 90, vjust=0.5))

## Warning: Ignoring unknown parameters: binwidth, bins, pad

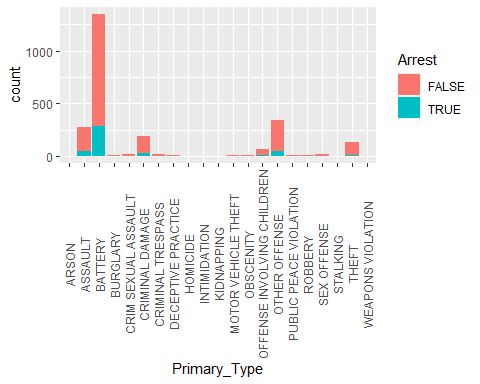


chicago\_domestic <- chicago %>% filter(Domestic %in% (TRUE))

Of domestic related crimes, which resulted in arrest. Battery has the highest count and arrests.

ggplot(chicago\_domestic, aes(x = Primary\_Type, fill = Arrest)) + geom\_histogram(stat = "count") + theme(axis.text.x = element\_text(angle = 90, vjust=0.5))

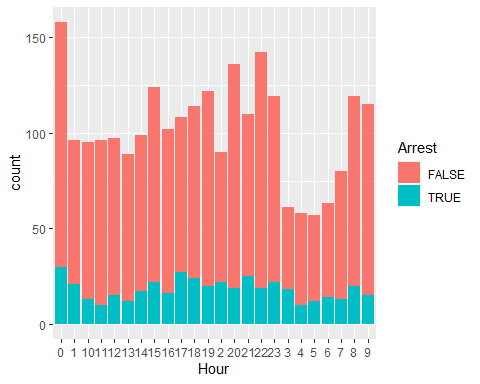
## Warning: Ignoring unknown parameters: binwidth, bins, pad



At what hour do most domestic crimes occur? From about 8 PM to 12 AM.

ggplot(chicago\_domestic, aes(x = Hour, fill = Arrest)) + geom\_histogram(stat = "count")

## Warning: Ignoring unknown parameters: binwidth, bins, pad

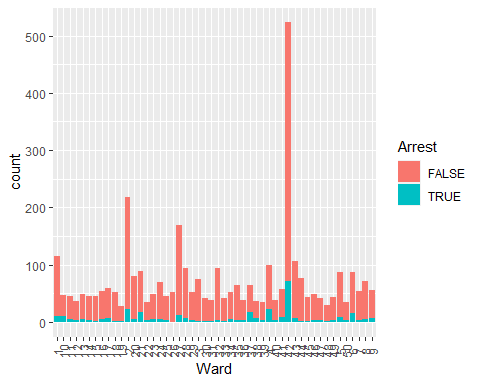


chicago\_theft <- chicago %>% filter(Primary\_Type %in% ("THEFT"))

Theft was a notable crime in tracking the reports. This graph looks at which ward faces the most theft and how many arrests occur. Ward 42 has the most reports of theft. They do have the most arrests, but is low proportionally to the reports.

ggplot(chicago\_theft, aes(x = Ward, fill = Arrest)) + geom\_histogram(stat = "count") + theme(axis.text.x = element\_text(angle = 90, vjust=0.5))

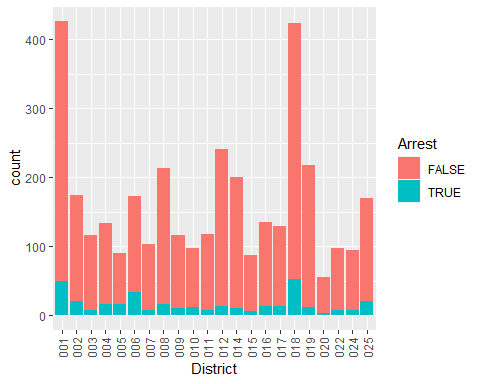
## Warning: Ignoring unknown parameters: binwidth, bins, pad



Which districts have the most reports of theft? Districts 1 and 18. They do have higher arrest rates for that theft, but proportionally low compared to rates of commission.

ggplot(chicago\_theft, aes(x = District, fill = Arrest)) + geom\_histogram(stat = "count") + theme(axis.text.x = element\_text(angle = 90, vjust=0.5))

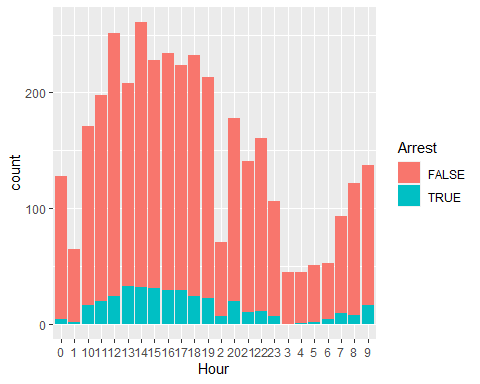
## Warning: Ignoring unknown parameters: binwidth, bins, pad



Of the theft reports, which hours had the most crime reported, and which were arrests. 12PM to 7PM had the most reports of theft. Arrests are higher at that time, but still proportionally low.

ggplot(chicago\_theft, aes(x = Hour, fill = Arrest)) + geom\_histogram(stat = "count")

## Warning: Ignoring unknown parameters: binwidth, bins, pad

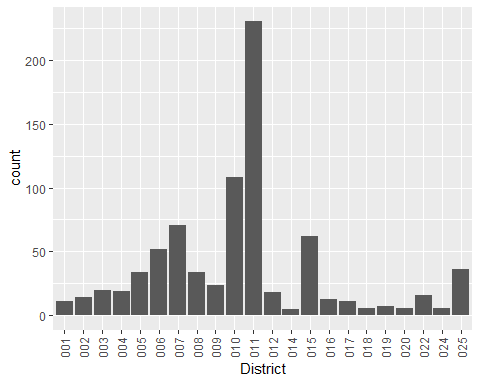


chicago\_narcotics <- chicago %>% filter(Primary\_Type %in% ("NARCOTICS"))

Of the districts, which has the most narcotic crimes? District 11. Did not need to do a proportion for arrests as all resulted in arrests.

ggplot(chicago\_narcotics, aes(x = District)) + geom\_histogram(stat = "count") + theme(axis.text.x = element\_text(angle = 90, vjust=0.5))

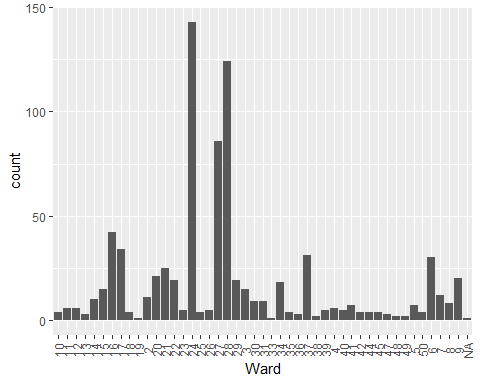
## Warning: Ignoring unknown parameters: binwidth, bins, pad



Which wards resulted in the highest reports of narcotic crimes? Districts 24 & 28. Since all the narcotic crimes resulted in arrest, this makes sense with the arrest counts in the graph above.

ggplot(chicago\_narcotics, aes(x = Ward)) + geom\_histogram(stat = "count") + theme(axis.text.x = element\_text(angle = 90, vjust=0.5))

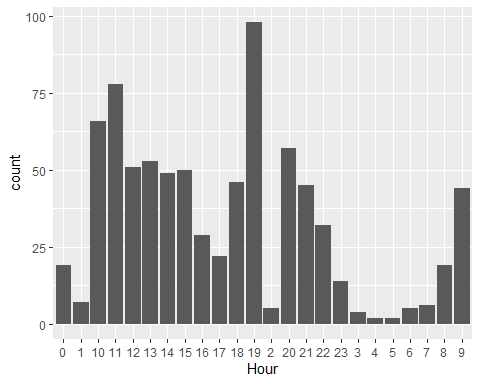
## Warning: Ignoring unknown parameters: binwidth, bins, pad



Which hour do most narcotics crimes get reported? 7 PM.

ggplot(chicago\_narcotics, aes(x = Hour, )) + geom\_histogram(stat = "count")

## Warning: Ignoring unknown parameters: binwidth, bins, pad

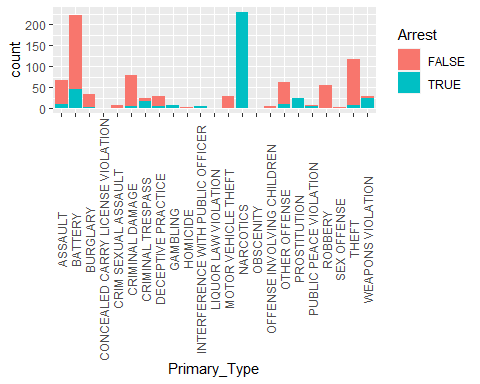


chicago\_11 <- chicago %>% filter(District %in% ("011"))

District 11 was pointed at showing a large amount of crime. Reviewing the types of crimes reported there, the two crimes that are the most notable are narcotic related crimes and battery.

ggplot(chicago\_11, aes(x = Primary\_Type, fill = Arrest)) + geom\_histogram(stat = "count") + theme(axis.text.x = element\_text(angle = 90, vjust=0.5))

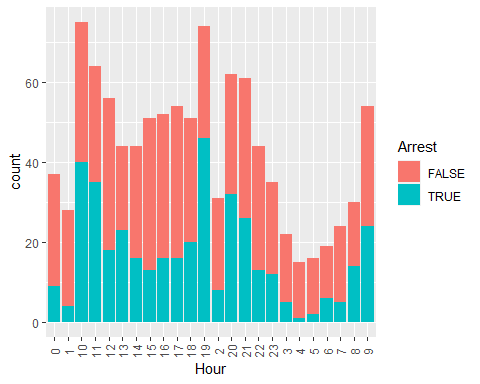
## Warning: Ignoring unknown parameters: binwidth, bins, pad



Which hour were most the crimes committed in District 11? Spiking at 10 AM and 7 PM.

ggplot(chicago\_11, aes(x = Hour, fill = Arrest)) + geom\_histogram(stat = "count") + theme(axis.text.x = element\_text(angle = 90, vjust=0.5))

## Warning: Ignoring unknown parameters: binwidth, bins, pad

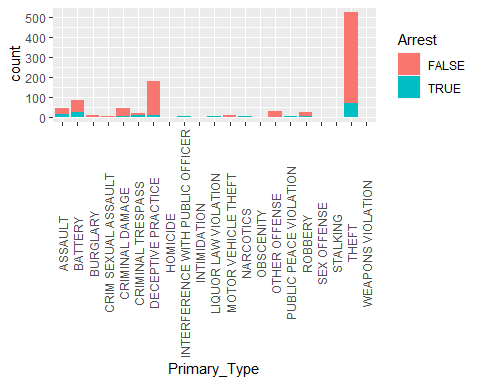


chicago\_42 <- chicago %>% filter(Ward %in% ("42"))

Ward 42 was also notable for the amount of crime committed. Most notable types is theft, most of which do not lead to arrests.

ggplot(chicago\_42, aes(x = Primary\_Type, fill = Arrest)) + geom\_histogram(stat = "count") + theme(axis.text.x = element\_text(angle = 90, vjust=0.5))

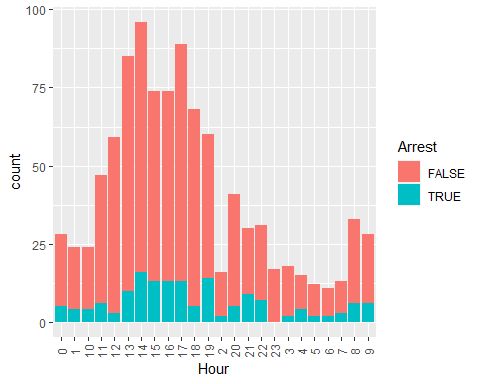
## Warning: Ignoring unknown parameters: binwidth, bins, pad



Which hours do crimes get committed in Ward 42? Spiked 1-2 PM and again at 5 PM.

ggplot(chicago\_42, aes(x = Hour, fill = Arrest)) + geom\_histogram(stat = "count") + theme(axis.text.x = element\_text(angle = 90, vjust=0.5))

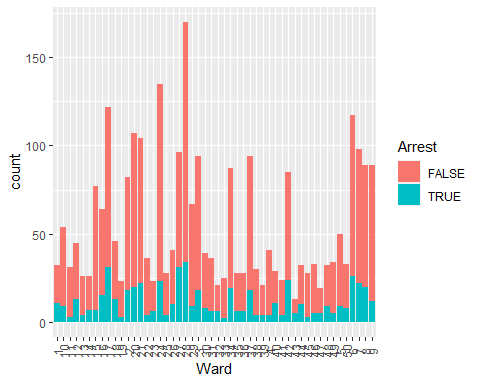
## Warning: Ignoring unknown parameters: binwidth, bins, pad



chicago\_battery <- chicago %>% filter(Primary\_Type %in% ("BATTERY"))

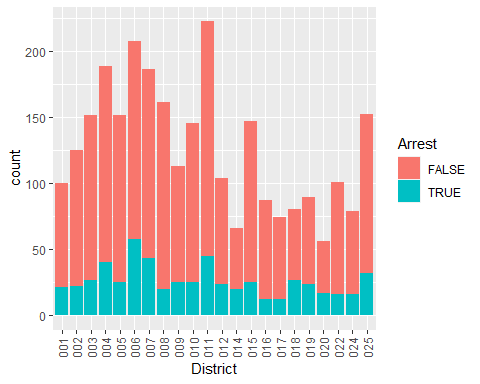
ggplot(chicago\_battery, aes(x = Ward, fill = Arrest)) + geom\_histogram(stat = "count") + theme(axis.text.x = element\_text(angle = 90, vjust=0.5))

## Warning: Ignoring unknown parameters: binwidth, bins, pad



ggplot(chicago\_battery, aes(x = District, fill = Arrest)) + geom\_histogram(stat = "count") + theme(axis.text.x = element\_text(angle = 90, vjust=0.5))

## Warning: Ignoring unknown parameters: binwidth, bins, pad



ggplot(chicago\_battery, aes(x = Hour, fill = Arrest)) + geom\_histogram(stat = "count") + theme(axis.text.x = element\_text(angle = 90, vjust=0.5))

## Warning: Ignoring unknown parameters: binwidth, bins, pad

