Chicago Crime Analysis

library(dplyr)

## Warning: package 'dplyr' was built under R version 3.5.3

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.5.3

library(forecast)

## Warning: package 'forecast' was built under R version 3.5.3

library(stringr)

## Warning: package 'stringr' was built under R version 3.5.3

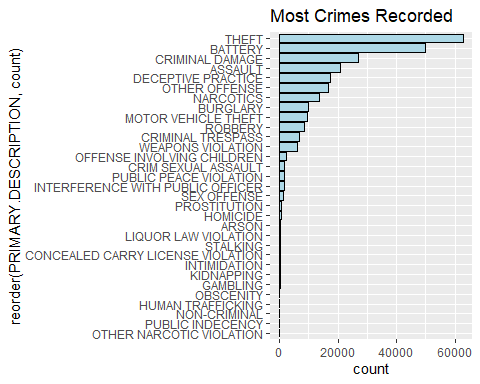
library(lubridate)

## Warning: package 'lubridate' was built under R version 3.5.2

##   
## Attaching package: 'lubridate'

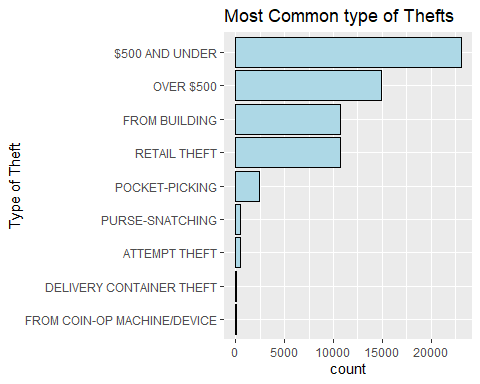
## The following object is masked from 'package:base':  
##   
## date

crime\_data <- read.csv("Crimes\_-\_One\_year\_prior\_to\_present.csv", sep= ",", stringsAsFactors = FALSE)  
crime\_data <- crime\_data[-c(4,10,12:14,17:19,21:23)]  
crime\_data$DATE..OF.OCCURRENCE = as.Date(crime\_data$DATE..OF.OCCURRENCE, format = "%m/%d/%Y %I:%M:%S %p")  
type\_of\_crime <- summarise(group\_by(crime\_data,PRIMARY.DESCRIPTION), count= n())  
ggplot(type\_of\_crime, aes(x= reorder(PRIMARY.DESCRIPTION,count), y= count))+ geom\_bar(stat= "identity",color= "black", fill= "light blue")+ coord\_flip()+ ggtitle("Most Crimes Recorded")



#write.table(crime\_data, file="crime\_data.csv",sep=",",row.names=F)

theft <- filter(crime\_data, PRIMARY.DESCRIPTION== 'THEFT')  
type\_of\_theft <- summarise(group\_by(theft, SECONDARY.DESCRIPTION), count= n())  
  
ggplot(type\_of\_theft, aes(x= reorder(SECONDARY.DESCRIPTION, count), y= count))+ geom\_bar(stat='identity', color= "black", fill= "light blue")+ coord\_flip()+ xlab('Type of Theft')+ ggtitle("Most Common type of Thefts")

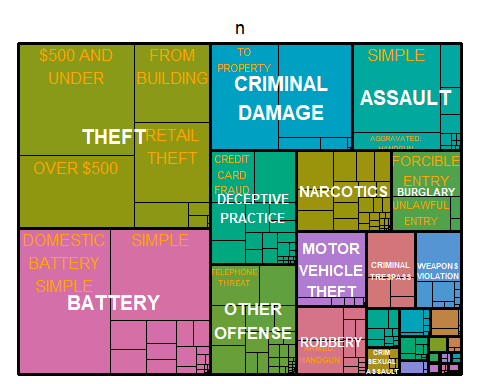


weapons\_crime <- filter(crime\_data, PRIMARY.DESCRIPTION== 'WEAPONS VIOLATION')  
weapons\_crime\_by\_area <- summarise(group\_by(weapons\_crime, Community.Areas), count= n())

library(dplyr)  
library(ggplot2)  
library(treemap)

## Warning: package 'treemap' was built under R version 3.5.3

treemap\_df <-  
 crime\_data %>%  
 group\_by(PRIMARY.DESCRIPTION, SECONDARY.DESCRIPTION) %>%  
 summarize(n = n())  
  
treemap(treemap\_df,   
 index=c("PRIMARY.DESCRIPTION","SECONDARY.DESCRIPTION"),   
 vSize="n",   
 type="index",  
 fontsize.labels=c(15,12),  
 fontcolor.labels=c("white","orange"),  
 fontface.labels=c(2,1),   
 bg.labels=c("transparent"),  
 align.labels=list(  
 c("center", "center"),   
 c("center", "top")  
 ),   
 overlap.labels=0.2,   
 inflate.labels=F  
 )

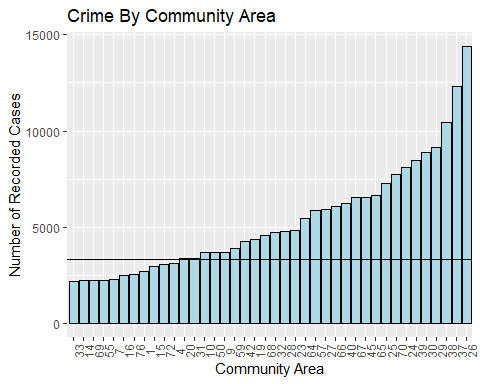


library(stringr)  
crime\_data$new\_block <- substr(crime\_data$BLOCK, start = 8, stop = str\_count(crime\_data$BLOCK))  
  
crime\_by\_area <- summarise(group\_by(crime\_data, new\_block), count= n())

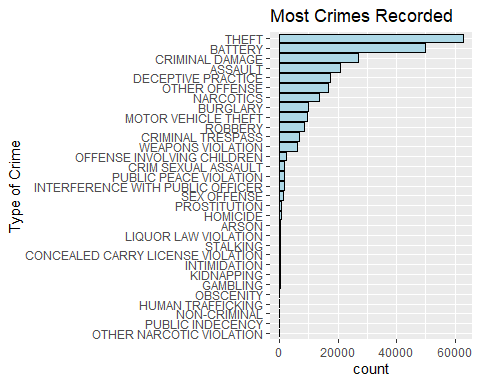
crime\_for\_community <- na.omit(crime\_data, cols=crime\_data$Community.Areas)  
crime\_by\_community\_area<- summarise(group\_by(crime\_for\_community, Community.Areas), count= n())  
crime\_by\_community\_area <- top\_n(crime\_by\_community\_area,40)

## Selecting by count

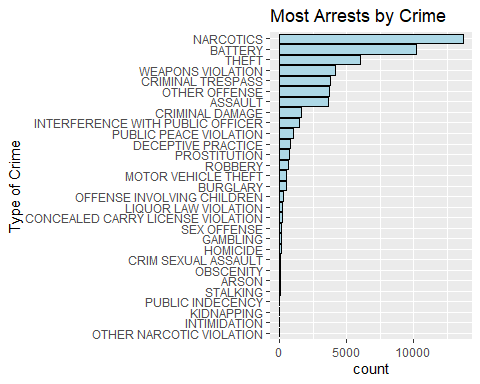
avg <- 3334.00  
ggplot(crime\_by\_community\_area, aes(x= reorder(Community.Areas, count), y=count))+ geom\_bar(stat= 'identity', color= "black", fill= "light blue") +geom\_hline(aes(yintercept=avg)) + theme(axis.text.x = element\_text(angle = 90))+ xlab('Community Area')+ ylab('Number of Recorded Cases')+ ggtitle("Crime By Community Area")



arrests <- filter(crime\_data, ARREST== 'Y')  
arrests\_by\_crime <- summarise(group\_by(arrests,PRIMARY.DESCRIPTION), count=n())  
  
ggplot(type\_of\_crime, aes(x=reorder(PRIMARY.DESCRIPTION, count), y= count))+ geom\_bar(stat= 'identity', color= "black", fill= "light blue")+ coord\_flip()+ xlab('Type of Crime')+ ggtitle("Most Crimes Recorded")



ggplot(arrests\_by\_crime, aes(x=reorder(PRIMARY.DESCRIPTION, count), y= count))+ geom\_bar(stat= 'identity', color= "black", fill= "light blue")+ coord\_flip()+ xlab('Type of Crime')+ ggtitle("Most Arrests by Crime")



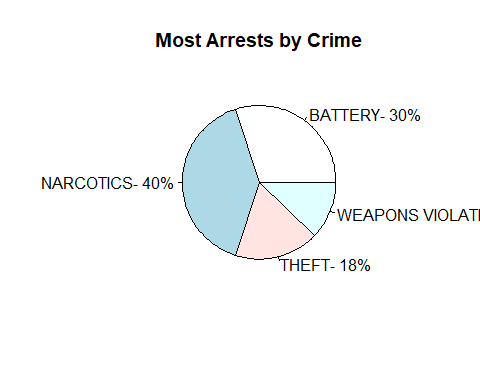
top\_crime\_recorded <- top\_n(type\_of\_crime,4)

## Selecting by count

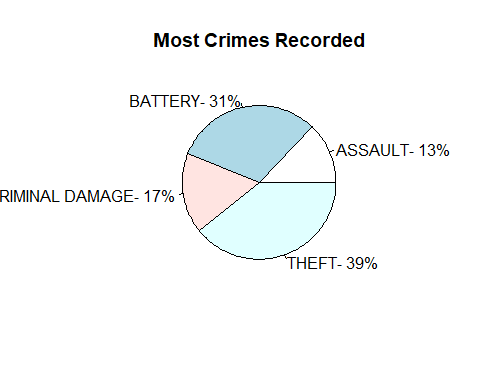
top\_arrests\_recorded <- top\_n(arrests\_by\_crime,4)

## Selecting by count

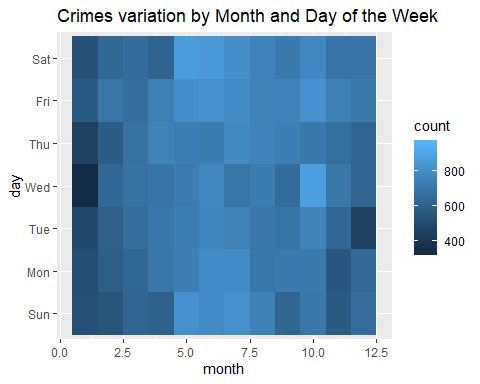
slices <- c(top\_arrests\_recorded$count)  
lbls <- c(top\_arrests\_recorded$PRIMARY.DESCRIPTION)  
pct <- round(slices/sum(slices)\*100)  
lbls1 <- paste(lbls,"- ", pct,"%", sep="")  
pie(slices, labels = lbls1, main="Most Arrests by Crime")



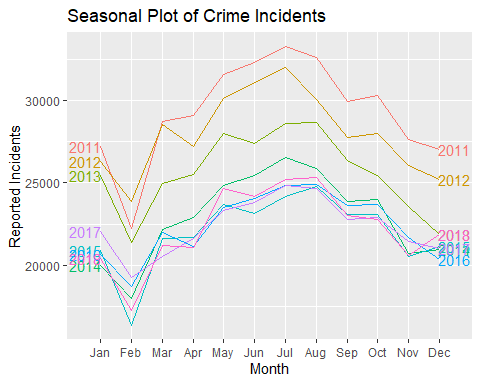
slices <- c(top\_crime\_recorded$count)  
lbls <- c(top\_crime\_recorded$PRIMARY.DESCRIPTION)  
pct <- round(slices/sum(slices)\*100)  
lbls1 <- paste(lbls,"- ", pct,"%", sep="")  
pie(slices, labels = lbls1, main="Most Crimes Recorded")



library(lubridate)  
crime\_by\_date <- summarise(group\_by(crime\_data, DATE..OF.OCCURRENCE), count= n())  
crime\_by\_date$month <- month(crime\_by\_date$DATE..OF.OCCURRENCE)  
crime\_by\_date$day <- as.Date(crime\_by\_date$DATE..OF.OCCURRENCE)  
crime\_by\_date$day <- wday(crime\_by\_date$day, label=TRUE)  
  
ggplot(crime\_by\_date, aes(x= month, y= day, fill= count))+ geom\_tile() + ggtitle("Crimes variation by Month and Day of the Week")



crime\_data\_2011\_2018 <- read.csv("Crimes\_2011\_2018.csv", sep= ",", stringsAsFactors = FALSE)  
crime\_data\_2011\_2018 <- crime\_data\_2011\_2018[-c(1,4:8,11:13,15:22)]  
crime\_data\_2011\_2018$Date = as.Date(crime\_data\_2011\_2018$Date, format = "%m/%d/%Y %I:%M:%S %p")  
crime\_data\_2011\_2018$Year <- lubridate::year(crime\_data\_2011\_2018$Date)  
crime\_data\_2011\_2018$Month <- lubridate::month(crime\_data\_2011\_2018$Date)  
crime\_by\_year <- summarise(group\_by(crime\_data\_2011\_2018, Year, Month), count=n())  
  
  
  
violent\_ts <- ts(crime\_by\_year$count, start=c(2011, 1), end=c(2018, 12), frequency=12)   
  
 ggseasonplot(violent\_ts, year.labels=TRUE, year.labels.left=TRUE) +  
 ylab("Reported Incidents") +  
 ggtitle("Seasonal Plot of Crime Incidents")



ggplot(crime\_by\_year)+ aes(x= Month, y= count)+ geom\_bar(stat= "identity", color= "black", fill= "light blue")+ ggtitle("Crime of Years")

