Dorchester and Boston Crime Data

1. Scrape Data from online
2. Exploratory Analysis

About the data - Universal Hub, started in 2005, is a community news and information site for the Boston area based on the idea that residents can help keep themselves informed. Combining reports from a pool of hundreds of local bloggers, photographers and tweeters, Universal Hub posts daily news and links, often before the mainstream media, and gives local folks a place to discuss the day's events - and to post news on their own.

Clean Data

```
# Clean crime
crime_data <- crime_data %>% mutate(crime = tolower(crime)) # make crimes all lowercase

shooting_vec <- c("shooting", "shootin", "shoting", "shotting") #all the shooting spellings
crime_data$crime <- ifelse(crime_data$crime %in% shooting_vec, "shooting", crime_data$crime)
```

The Top 5 Most Common Crimes in Dorchester, MA

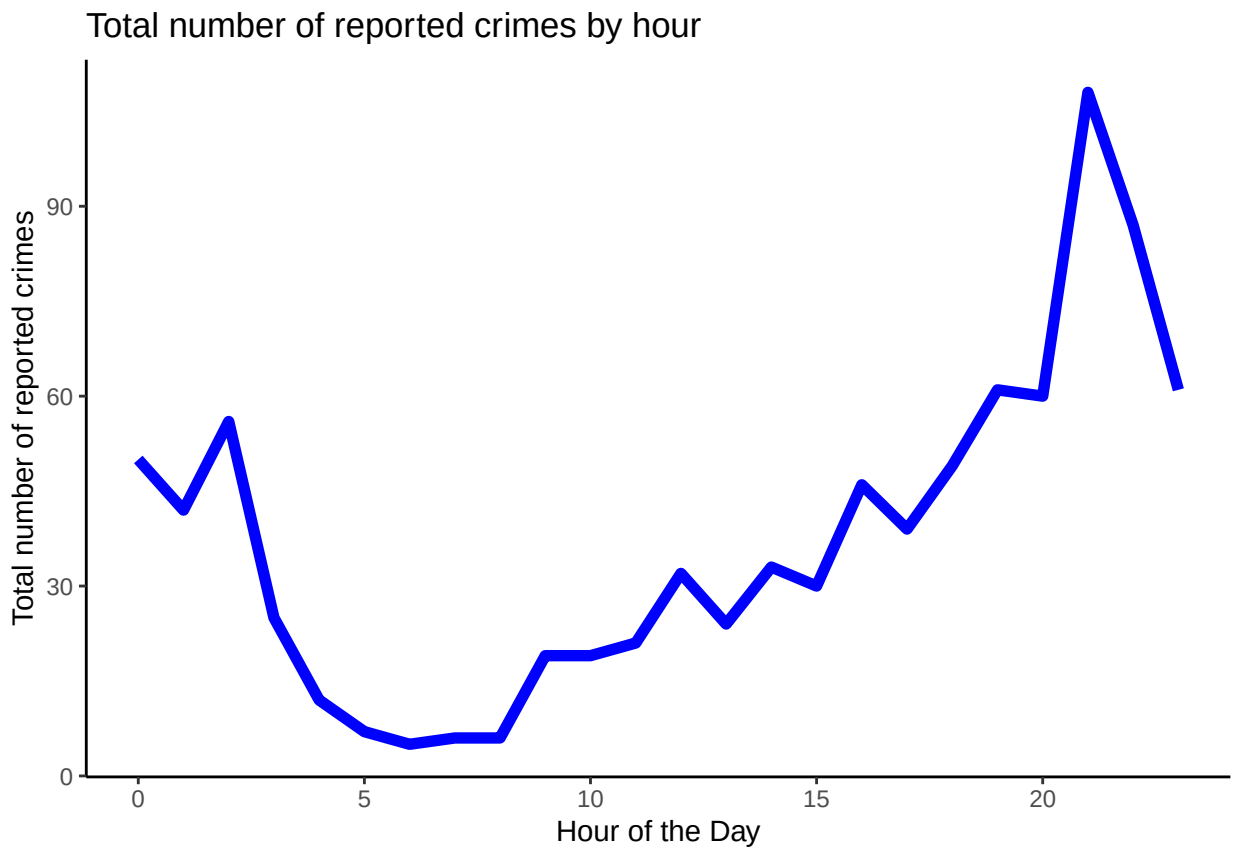
```
crime_count <- crime_data %>%
  group_by(crime) %>%
  summarize (count = n()) %>%
  arrange(desc(count))
top_5_crimes = crime_count %>% slice(1:5)
top_5_crimes
```

```
## # A tibble: 5 x 2
##   crime                   count
##   <chr>                   <int>
## 1 shooting                  170
## 2 illegal gun possession    130
## 3 gunfire                   127
## 4 murder                    113
## 5 stabbing                   74
```

Number of Reported Crimes by Hour

```
crimes_by_hour <- crime_data %>%
  group_by(hour) %>%
  summarize (count = n())

ggplot(crimes_by_hour, aes(x=hour, y=count)) +
  geom_line(color = "blue", linetype="solid", size =2) + theme_classic() +
  labs(title="Total number of reported crimes by hour", x="Hour of the Day",
       y = "Total number of reported crimes")
```

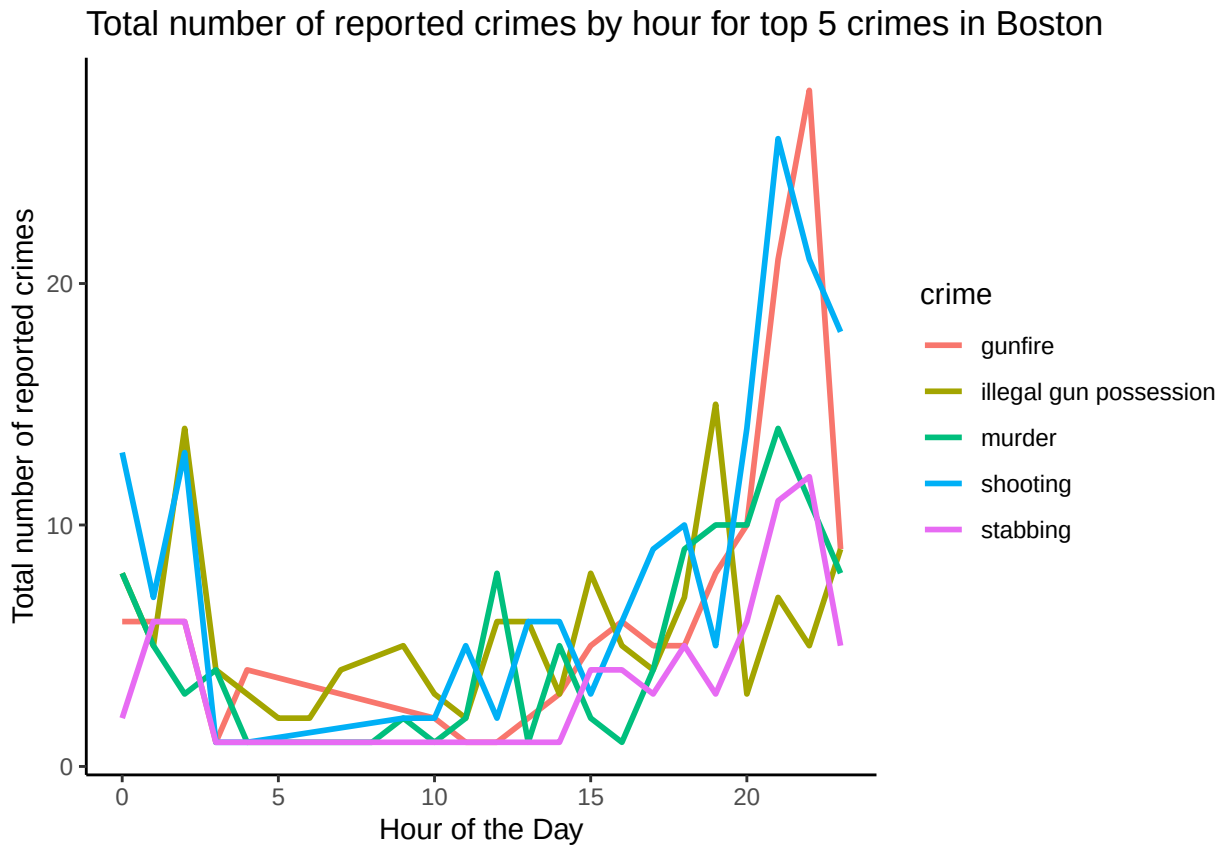


The total number of reported crimes by hour is interestingly U-shaped. There is a drop in the number of crimes reported in the early morning and then steadily increases throughout the day with the most amount of crimes occurring during the hours of 9pm to 11 pm.

```
top_5 <- top_5_crimes$crime
crimes_by_hour_type <- crime_data %>%
  group_by(hour, crime) %>%
  summarize (count = n()) %>%
  filter(crime %in% top_5)
```

```
## `summarise()` has grouped output by 'hour'. You can override using the
## `.groups` argument.
```

```
ggplot(crimes_by_hour_type, aes(x=hour, y=count, color = crime)) +
  geom_line(linetype="solid", size = 1) + theme_classic() +
  labs(title="Total number of reported crimes by hour for top 5 crimes in Boston",
       x="Hour of the Day", y = "Total number of reported crimes")
```

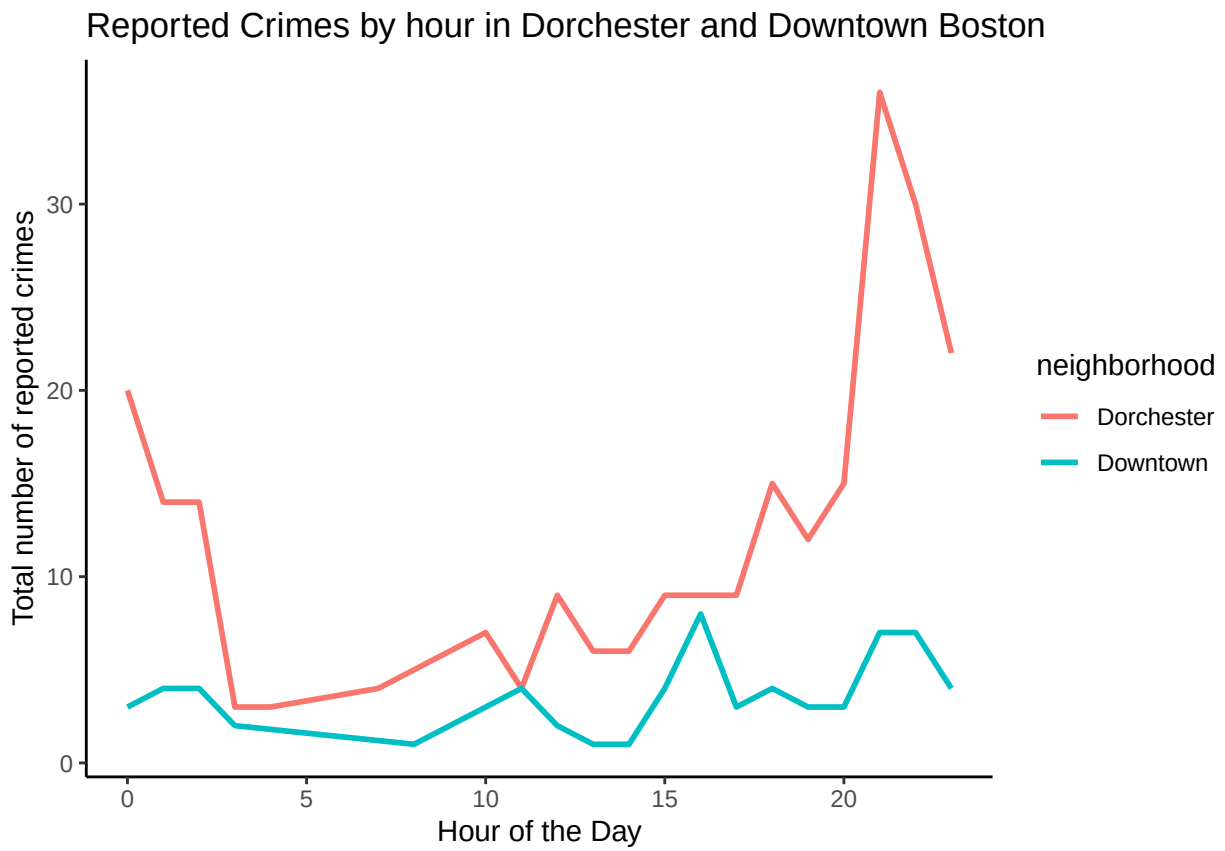


Although the five crimes types are nested together in the daytime, the volume of reported shootings increases after late afternoon and becomes the highest in the evening.

```
crime_by_hour_neighborhood <- crime_data %>%
  group_by(hour, neighborhood) %>%
  summarize (count = n()) %>%
  filter(neighborhood %in% c("Downtown", "Dorchester"))
```

```
## `summarise()` has grouped output by 'hour'. You can override using the
## `.groups` argument.
```

```
ggplot(crime_by_hour_neighborhood ,
                                  aes(x=hour, y=count, color = neighborhood)) +
  geom_line(linetype="solid", size = 1) + theme_classic() +
  labs(title="Reported Crimes by hour in Dorchester and Downtown Boston",
       x="Hour of the Day", y = "Total number of reported crimes")
```



There is less reported crime in Downtown Boston than in Dorcester. In the early morning there is a decrease in reported crime in Dorcester and it slowly increase throughout the day peaking at night. In Downtown Boston, crime is reported most in the late afternoon and then drops in evening.