Houston Crime Report 2010

Aymen Rumi

Houston Crime Dataset

```
writeLines(names(crime))
```

```
time
date
hour
premise
offense
beat
block
street
type
suffix
number
month
day
location
address
lon
lat
```

Given our dataset with the above attributes, we will asses certain patterns & information about the crimes reported in Houston, TX.

Questions:

1. The distribution of crimes & patterns throughout a day
2. The distribution of crimes & patterns throughout a week
3. The distribution of crimes & patterns throughout a month
4. Distribution of offenses in relation to street
5. Crime densities by street

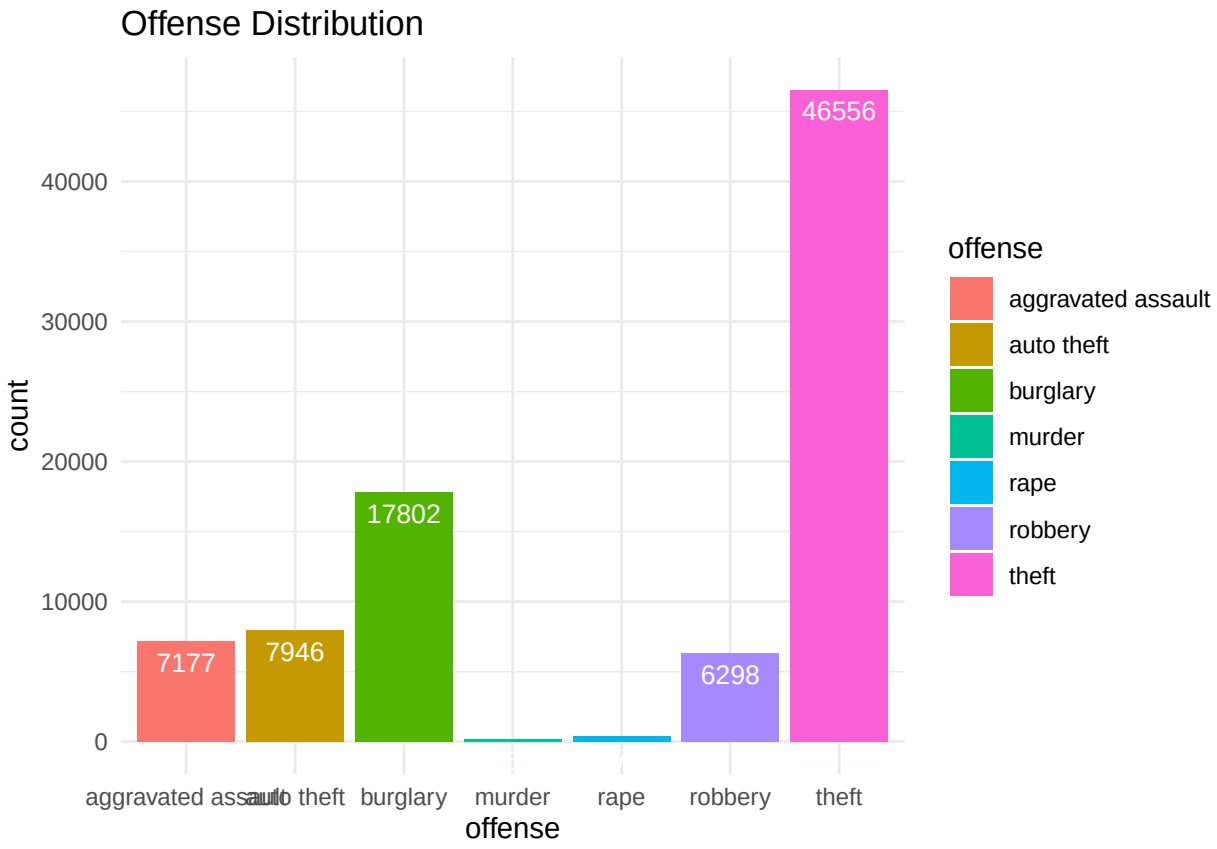
```
crime_by_hour<-crime%>%group_by(hour)%>%summarise(count=n())%>%mutate(prop=count/sum(count))

crime_by_week<-crime%>%group_by(day)%>%summarise(count=n())%>%mutate(prop=count/sum(count))

crime_by_month<-crime%>%group_by(month)%>%summarise(count=n())%>%mutate(prop=count/sum(count))
```

```
crime_by_offense<-crime%>%group_by(offense)%>%summarise(count=n())%>%mutate(prop=count/sum(count))
```

```
ggplot(crime_by_offense, aes(x=offense,y=count,fill=offense)) +   geom_bar(stat="identity")+geom_text(a
```



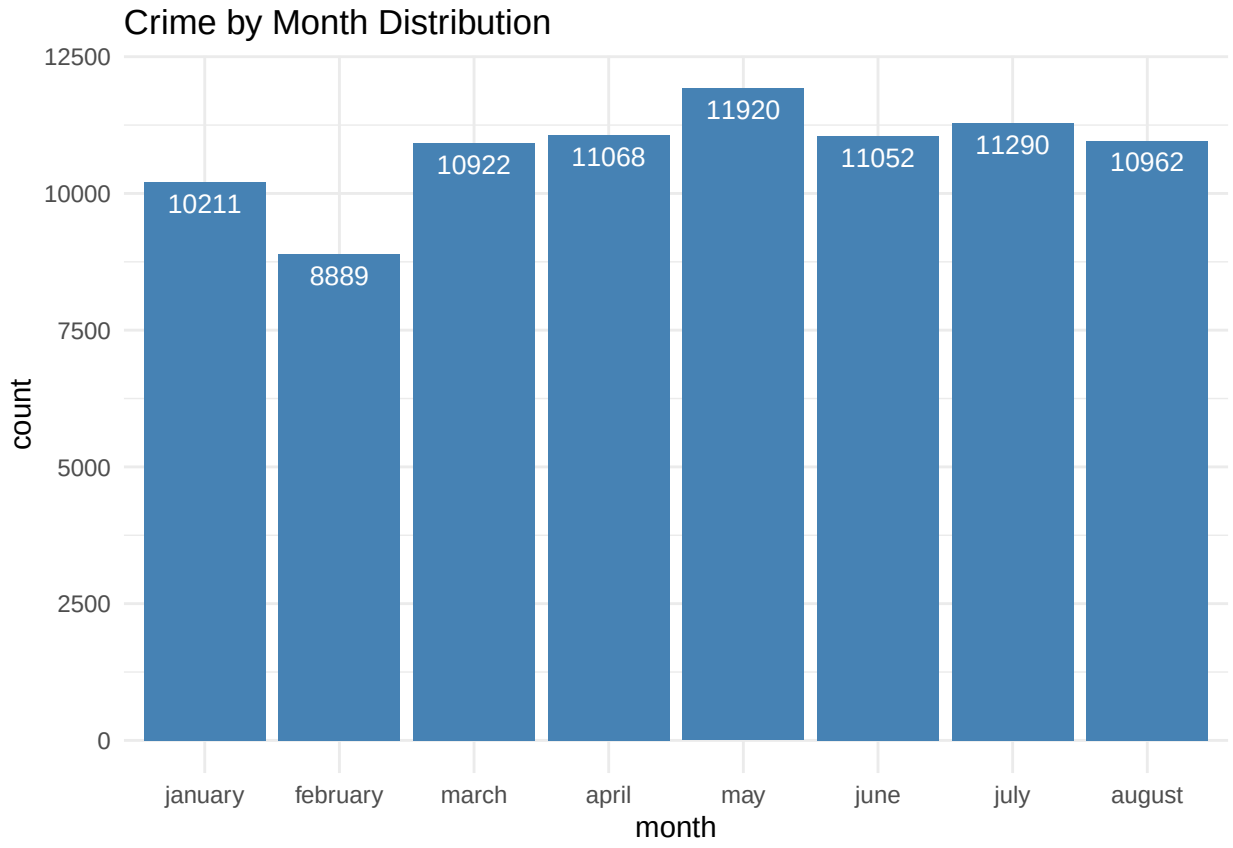

```
crime_by_offense%>%kable()
```

offense	count	prop
aggravated assault	7177	0.0831499
auto theft	7946	0.0920592
burglary	17802	0.2062470
murder	157	0.0018189
rape	378	0.0043794
robbery	6298	0.0729661
theft	46556	0.5393795

Analysis:

We cans see observe that crimes in the houston area are predominantly theft, accounting for more than 53% of total crimes, burglary 2nd at 20% both accounting for a total of 73% of crimes. Proportion of muder and rape account for 0.5%, while it it relatively low it is far from 0 accoutning for a total of 500 during 2010. We will later explore details of these specific crimes

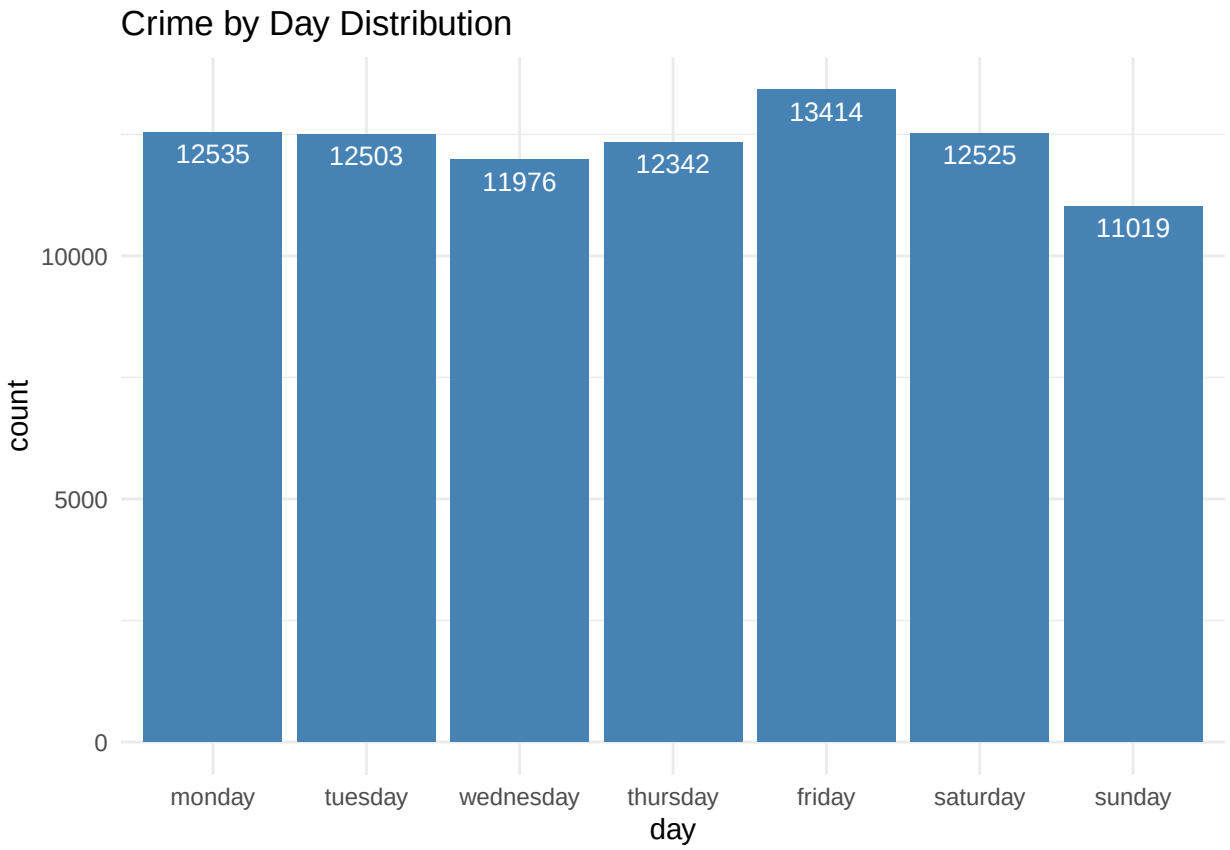
```
ggplot(data=crime_by_month, aes(x=month, y=count)) +
  geom_bar(stat="identity", fill="steelblue")+
  geom_text(aes(label=count), vjust=1.6, color="white", size=3.5)+
  theme_minimal()+ggtitle("Crime by Month Distribution")
```



Analysis:

Rate of crimes stay relatively uniform throughout the year with a slighy dip in february and sligh upward trend in may, there will be further investigates of the offense distribution by month to see if they follow the same uniform distribution as well

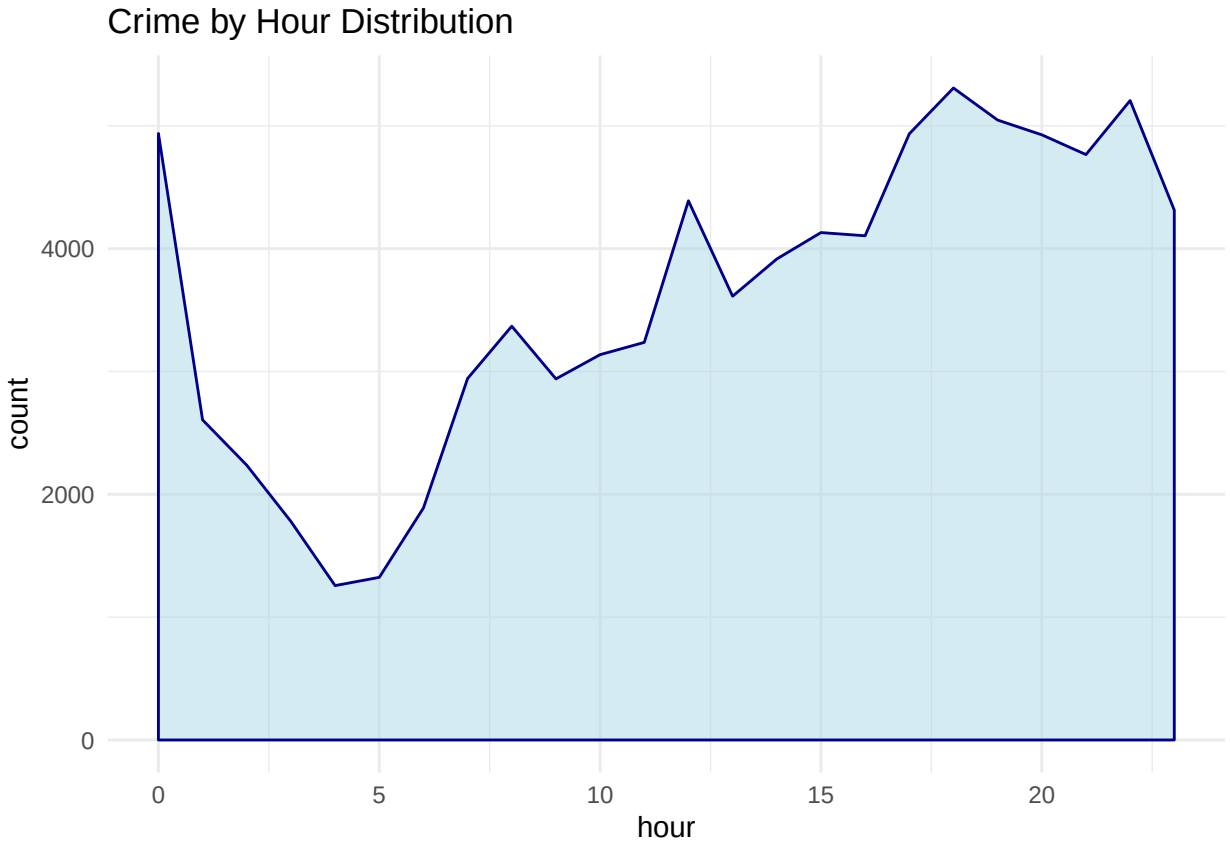
```
ggplot(data=crime_by_week, aes(x=day, y=count)) +
  geom_bar(stat="identity", fill="steelblue")+
  geom_text(aes(label=count), vjust=1.6, color="white", size=3.5)+
  theme_minimal()+ggtitle("Crime by Day Distribution")
```



Analysis:

Rate of crimes stay relatively uniform throughout the week as well with a dip on Sunday(I guess even criminals need a day off), and a little up on Friday

```
ggplot(crime_by_hour, aes(x=hour,y=count)) +  geom_density(stat="identity",color="darkblue", fill="ligh
```



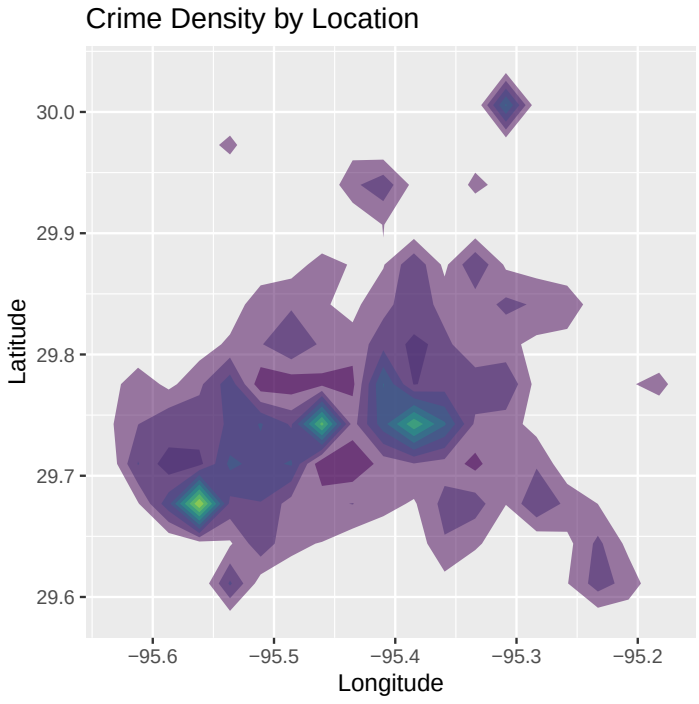

```
slice(crime_by_hour%>%arrange(desc(count)),1:10)%>%kable()
```

hour	count	prop
18	5308	0.0614964
22	5206	0.0603147
19	5047	0.0584726
0	4938	0.0572097
17	4935	0.0571750
20	4927	0.0570823
21	4766	0.0552170
12	4390	0.0508608
23	4314	0.0499803
15	4131	0.0478601

Analysis:

Rate of crimes throughout the day show an interesting pattern, there is an upward trend from 5 AM all the way to 12 AM and a big spike down, with some peaks


```
## Warning: Removed 5 rows containing non-finite values (stat_density2d).
```



Analysis:

We observe particular locations in the Houston Area with dense distribution of crimes