Visualization of Crime Statistics

Final Finding: DUIs are proportionally more common in all districts in San Francisco on weekends than during weekdays.

Plan is to take crime statistics from San Francisco from the summer of 2015 and look at which types of crime are committed during the weekend (Sat-Sun) in the different districts.

```
library(data.table)
library(ggplot2)
library(reshape)
```

```
##
## Attaching package: 'reshape'
##
## The following object is masked from 'package:data.table':
##
##     melt
```

Loading in the San Francisco Data Set.

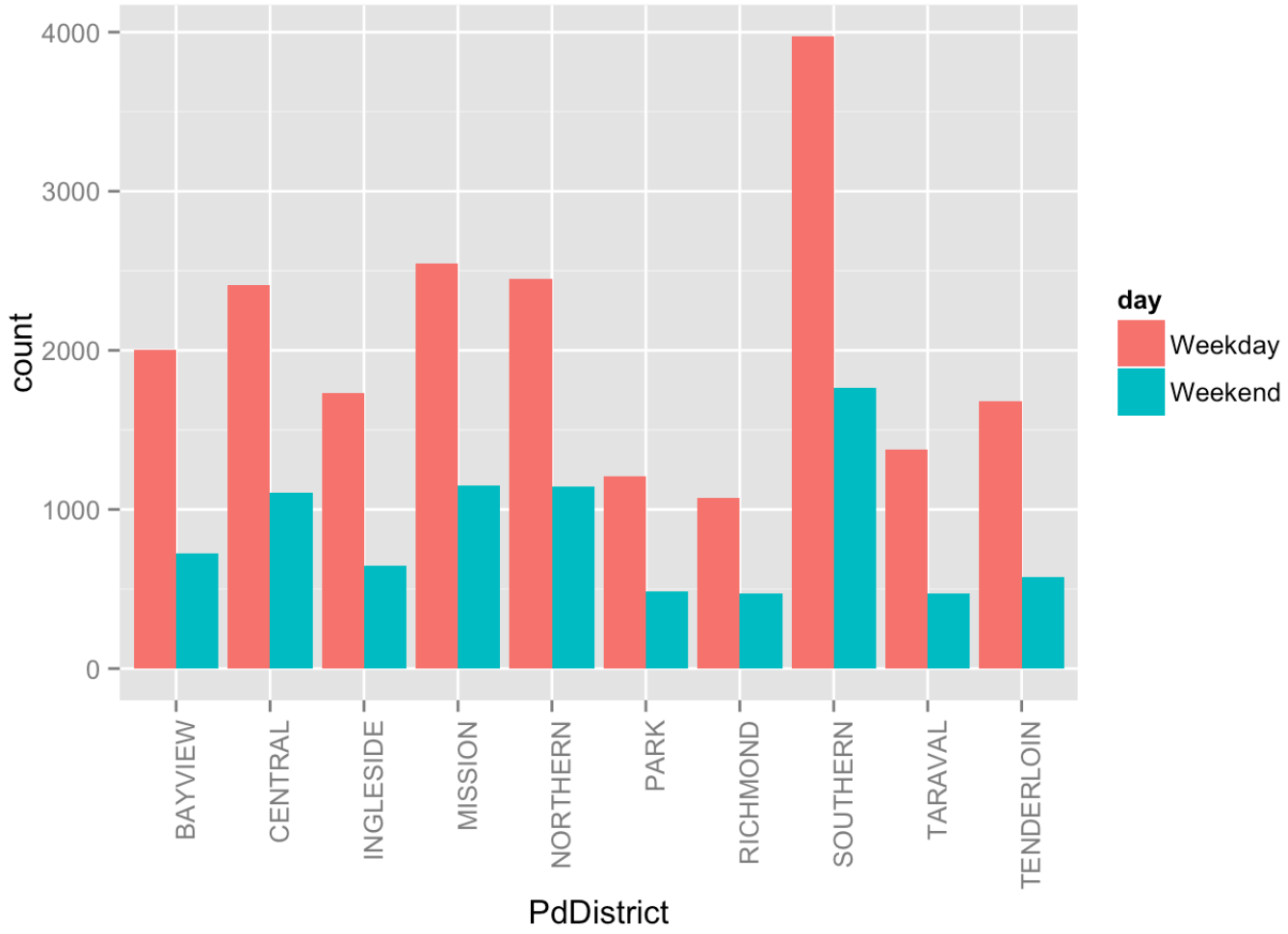
```
sf <- as.data.table(read.csv("sanfrancisco_incidents_summer_2014.csv"))
```

Assign Weekend versus Weekday based on DayOfWeek

```
wend <- function(x){
  if(x=="Saturday") return("Weekend")
  if(x=="Sunday") return("Weekend")
  return("Weekday")
}

assignment <- as.factor(unlist(lapply(sf$DayOfWeek, wend)))
nopr <- sf[,day:=assignment]
```

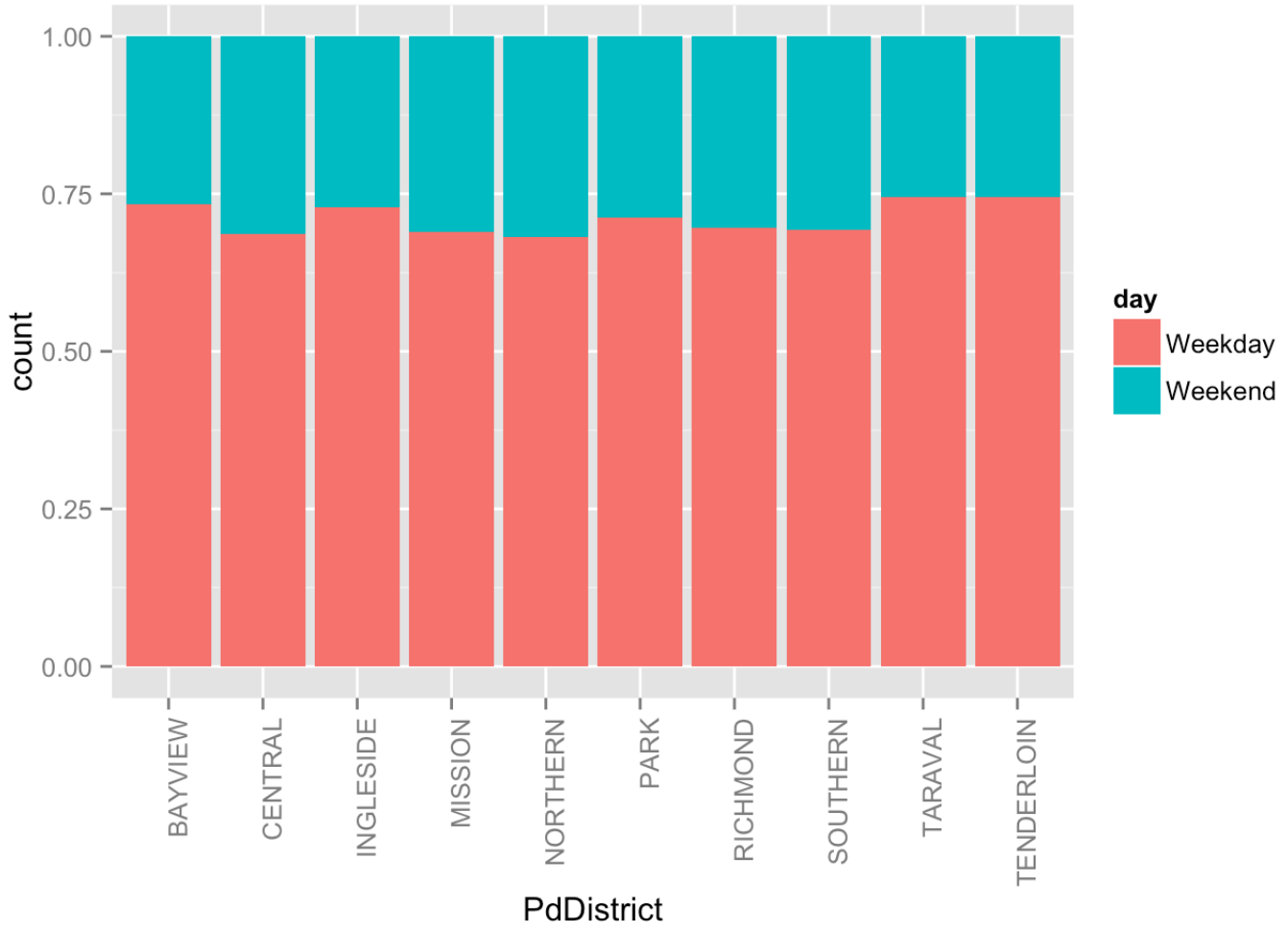
Look at breakdown of total crimes in each district by Weekday vs Weekend:

```
ggplot(sf, aes(x=PdDistrict, fill=day)) + geom_bar(position="dodge") + theme(axis.tex
t.x = element_text(angle = 90, hjust = 1))
```



This would be more clearly seen if instead of comapring bars, there was a percentage number.

```
ggplot(sf, aes(x=PdDistrict, fill=day)) + geom_bar(position="fill")+ theme(axis.text.
x = element_text(angle = 90, hjust = 1))
```



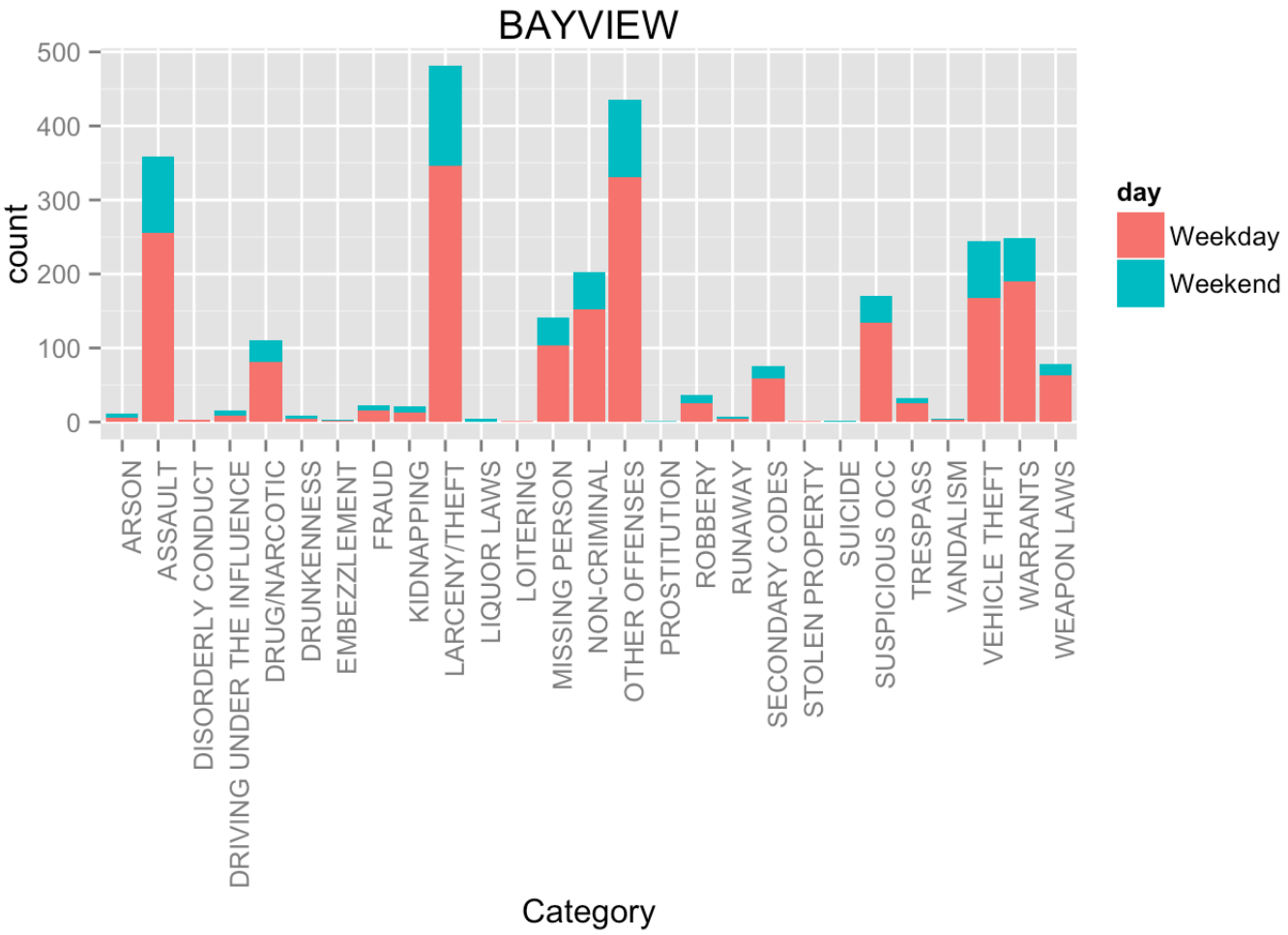
For all districts, there are more crimes during Monday-Fri, which is 5 days, than Sat-Sunday, which is 2 days.

Looking at the number of crimes on the weekend as a filled histogram, shows that there is a skew towards the weekdays for Tenderloin, Traval, Park, Ingleside, and Bayview.

Some districts have proportionally more crime on the weekend, including Central, Mission, Northern, and Southern.

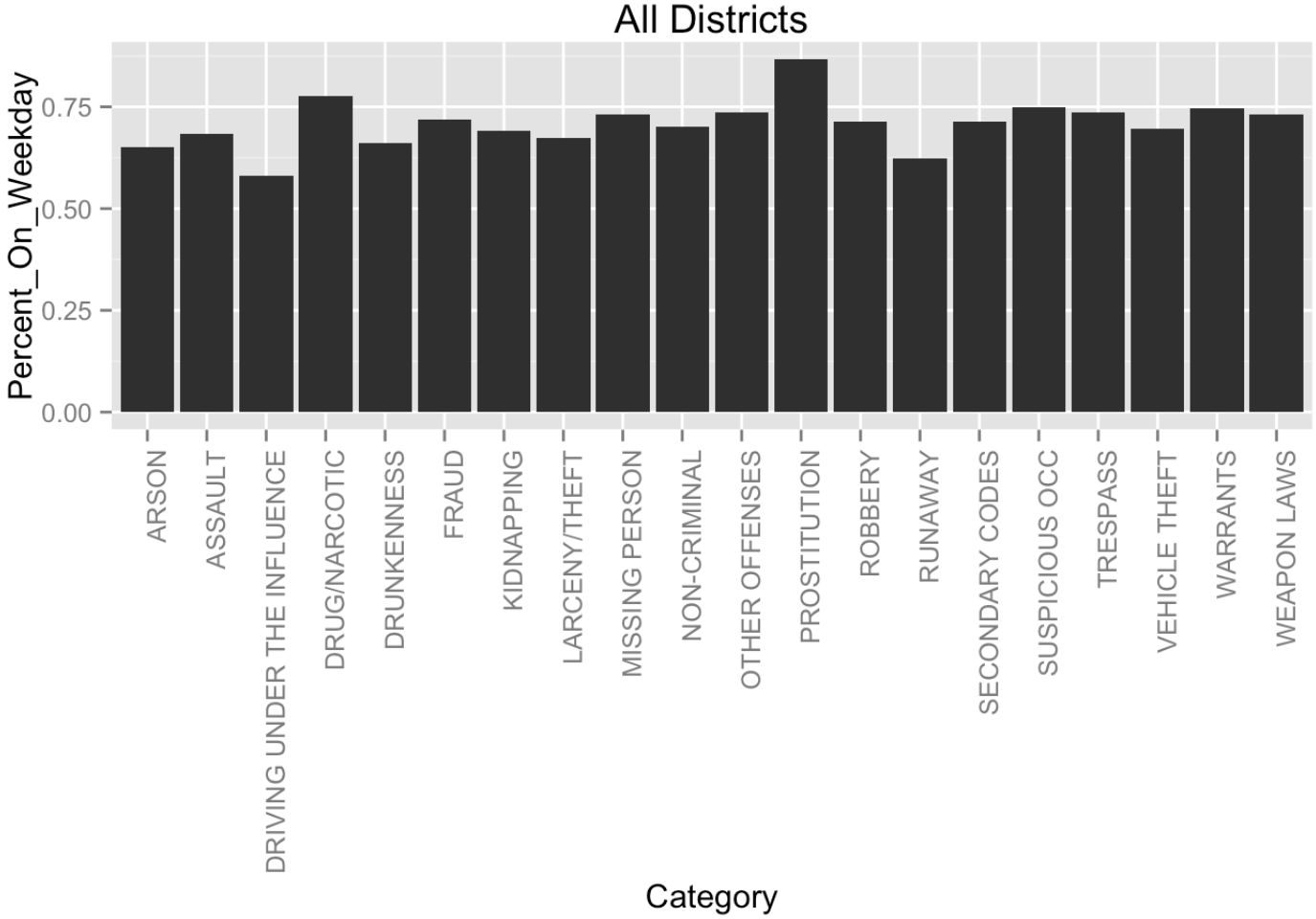
Now, we'll investigate relative proportions of different types of crimes in a district during the week versus the weekend. Example for 1 district

```
districts <- levels(sf$PdDistrict)

i <- 1
plot1 <- ggplot(sf[PdDistrict==districts[i]], aes(x=Category, fill=day)) + geom_histo
gram()
plot1 + theme(axis.text.x = element_text(angle = 90, hjust = 1)) + ggtitle(districts[
i])
```



These histograms are difficult to interpret on a district by district basis. I think graphing as proportions of weekday to weekend is better. I'm also going to focus on more common crimes to get more signal versus noise. I'm going to use an arbirary cutoff of 50 incidents to include a category.

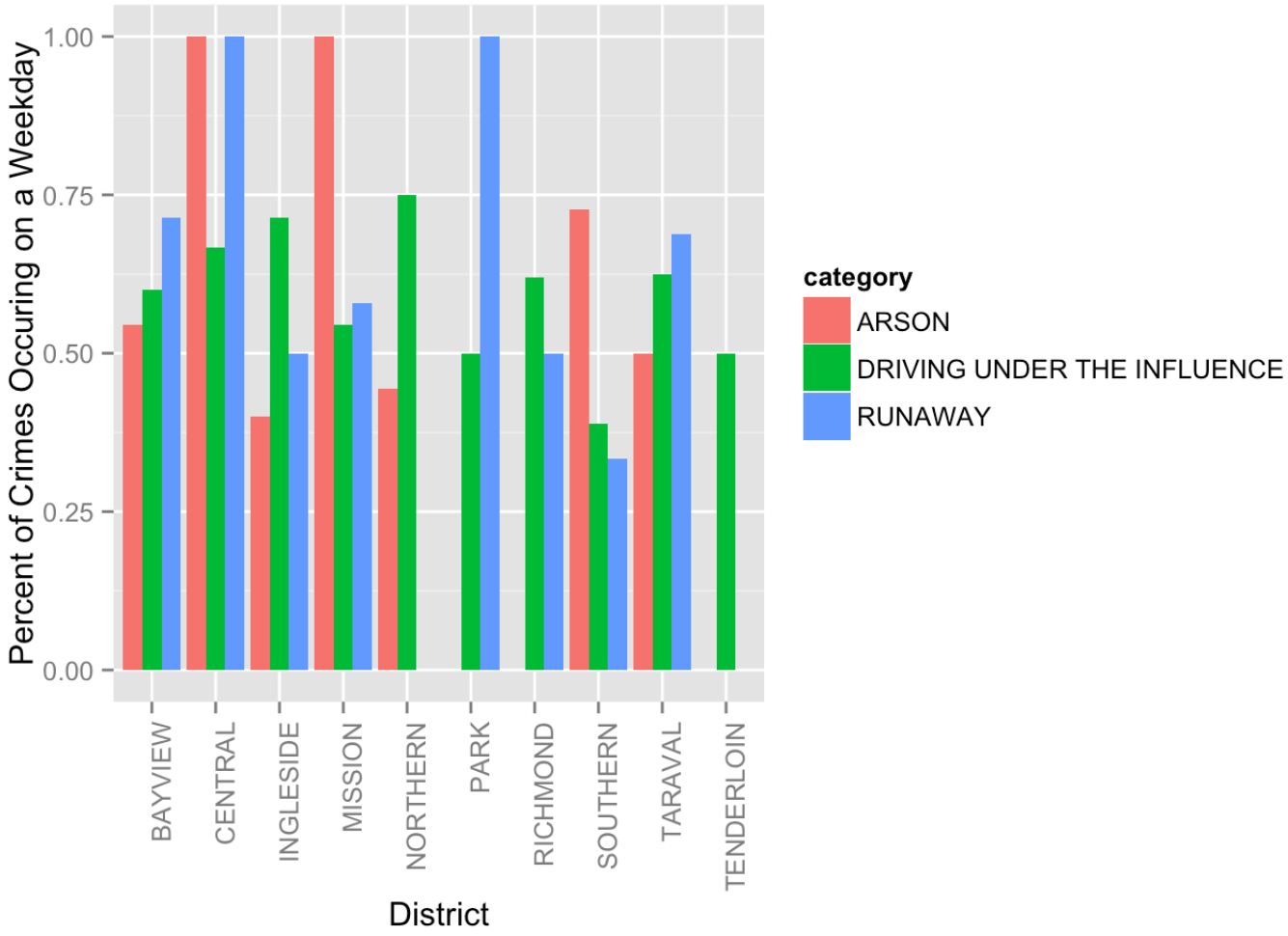
```
sfFocused <- sf[,.(Category, PdDistrict, DayOfWeek, day)]

catCrime <- unique(sf$Category)
percent <- rep(0, length(catCrime))
totalInCat <- rep(0, length(catCrime))

#Number of each category during week and weekend:
for(i in 1:length(catCrime)){
  weekday <- nrow(sfFocused[day=="Weekday" & Category==catCrime[i]])
  weekend <- nrow(sfFocused[day=="Weekend" & Category==catCrime[i]])
  percent[i] <- weekday/(weekday+weekend)
  totalInCat[i] <- weekday+weekend
}

perPlot <- as.data.table(catCrime)
perPlot[,Percent_On_Weekday:=percent]
perPlot[,total:=totalInCat]
setnames(perPlot, c("catCrime"), c("Category"))

ggplot(perPlot[totalInCat>=50], aes(x=Category, y=Percent_On_Weekday))+ geom_bar(stat
="identity") + theme(axis.text.x = element_text(angle = 90, hjust = 1)) + ggtitle("Al
l Districts")
```



For all of the districts combined, there are three crimes that occur proportionally more often on the weekend. These three crimes are Arson, DUIs, and Runaway.

How do these crimes breakdown for the percent occuring on a weekday for each district?

```
focus <- c("DRIVING UNDER THE INFLUENCE", "RUNAWAY", "ARSON")

weekendCrimes <- sfFocused[Category %in% focus]

district <- unique(weekendCrimes$PdDistrict)
catCrime <- focus

dist30 <- rep(district, 3)
focus30 <- rep(focus, 10)
percent30 <- rep(0, 30)

table <- as.data.table(dist30)
table[,category:=focus30]
table[,percentWeekday:=percent30]


#Number of each category during week and weekend:
for(i in 1:length(catCrime)){
  for(j in 1:length(district)){
    weekday <- nrow(weekendCrimes[day=="Weekday" & Category==catCrime[i] & PdDistrict
==district[j]])
    weekend <- nrow(weekendCrimes[day=="Weekend" & Category==catCrime[i] & PdDistrict
==district[j]])
    perc <- weekday/(weekday+weekend)
    if(is.na(perc)){perc <- 0}
    table[category==catCrime[i] & dist30==district[j],percentWeekday:=perc]
  }}



ggplot(table, aes(x=dist30, fill=category, y=percentWeekday)) + geom_bar(position="do
dge", stat="identity") + ylab("Percent of Crimes Occuring on a Weekday") + xlab("Dist
rict") + theme(axis.text.x = element_text(angle = 90, hjust = 1))
```

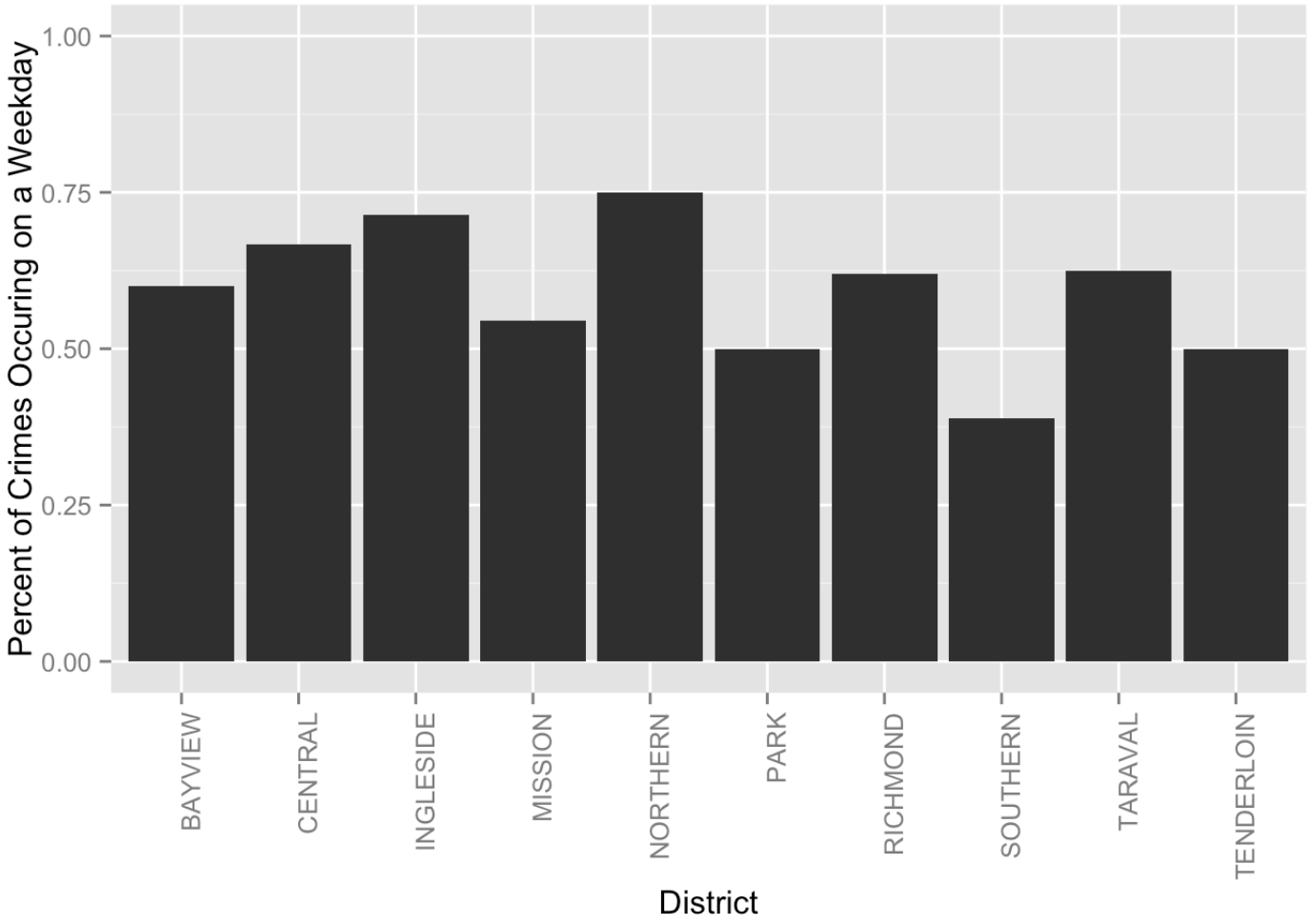


The only category that has some stability by each district is “Driving under the influence.” Arson and Runaway both flucuate in their weekend versus weekday predominance depending on the district.

Based on these visualizations, the only category of crime that appears to consistantly occur proportionally more often on the weekend, regardless of the part of town, is DUIs. People really do drink more on the weekend.

To drive this point home, here is a graph showing just the percentage of DUIs on weekdays. Rememeber that there are only 2 days here that count as weekends and 5 days that count as weekdays.

```
ggplot(table[category=="DRIVING UNDER THE INFLUENCE"], aes(x=dist30,  y=percentWeekda
y)) + geom_bar(position="dodge", stat="identity") + ylab("Percent of Crimes Occuring
on a Weekday") + xlab("District") + theme(axis.text.x = element_text(angle = 90, hjus
t = 1)) + ylim(c(0,1))
```


Percent of Crimes Occuring on a Weekday
1.00
0.75
0.50
0.25
0.00
BAYVIEW
CENTRAL
INGLESIDE
MISSION
NORTHERN
PARK
RICHMOND
SOUTHERN
TARAVAL
TENDERLOIN
District