# Explore Officer Traffic Stops Dataset

Chen Wang

This dataset contains 79884 records traffic stops by police officers in Charlotte, North Carolina in 2016. Firstly, the table below summarizes the composition of different reasons for the stops. It shows that most drivers were stoped because of Vehicle Regulatory (40.57%) and speeding(27.82%)

```
df %>%
  filter(!is.na(Reason_for_Stop)) %>%
  group_by(Reason_for_Stop) %>%
  summarise(count=n()) %>%
  mutate(prop = round(count/sum(count),4)*100) %>%
  arrange(desc(prop)) %>%
  knitr::kable(digits=2, col.names =  c("Reason for Stop","Number of stops",
                    "Proportion of the total(%) "), align = c('l','r','r'))
```

Reason for Stop	Number of stops	Proportion of the total(%)
Vehicle Regulatory	32405	40.57
Speeding	22222	27.82
Stop Light/Sign	7946	9.95
Vehicle Movement	7535	9.43
Safe Movement	4827	6.04
Investigation	1992	2.49
Other	1926	2.41
SeatBelt	631	0.79
CheckPoint	286	0.36
Driving While Impaired	114	0.14

Secondly, let's look at the age, gender and race composition of drivers who got stopped, which is shown in the diagram below. Apparently, male drivers were stopped more frequently than female drivers. Also more people from age 20-30 are stopped than other age groups. Regarding the race of drivers, black drivers were involved in over 50% of the stops, ranking the first among all the races, while white drivers came second.

```
df <- df %>% mutate(
  Driver_Agegroup = cut(Driver_Age,breaks=seq(0,100,by=10),
                    labels=c("under 10", paste(seq(10,99,by=10),"s",sep=""))))
ggplot(df)+
  geom_bar(aes(x=Driver_Agegroup,fill=Driver_Race))+
  facet_grid( ~ Driver_Gender)+
  scale_fill_brewer(palette = "Accent")+
  xlab("Age Group")+
  ylab("Count")+
  labs(fill="Driver Race")+
  theme(axis.text.x=element_text(size=8.5, angle=35,vjust=0.8,hjust=0.5))+
  scale_y_continuous(labels=comma)
```
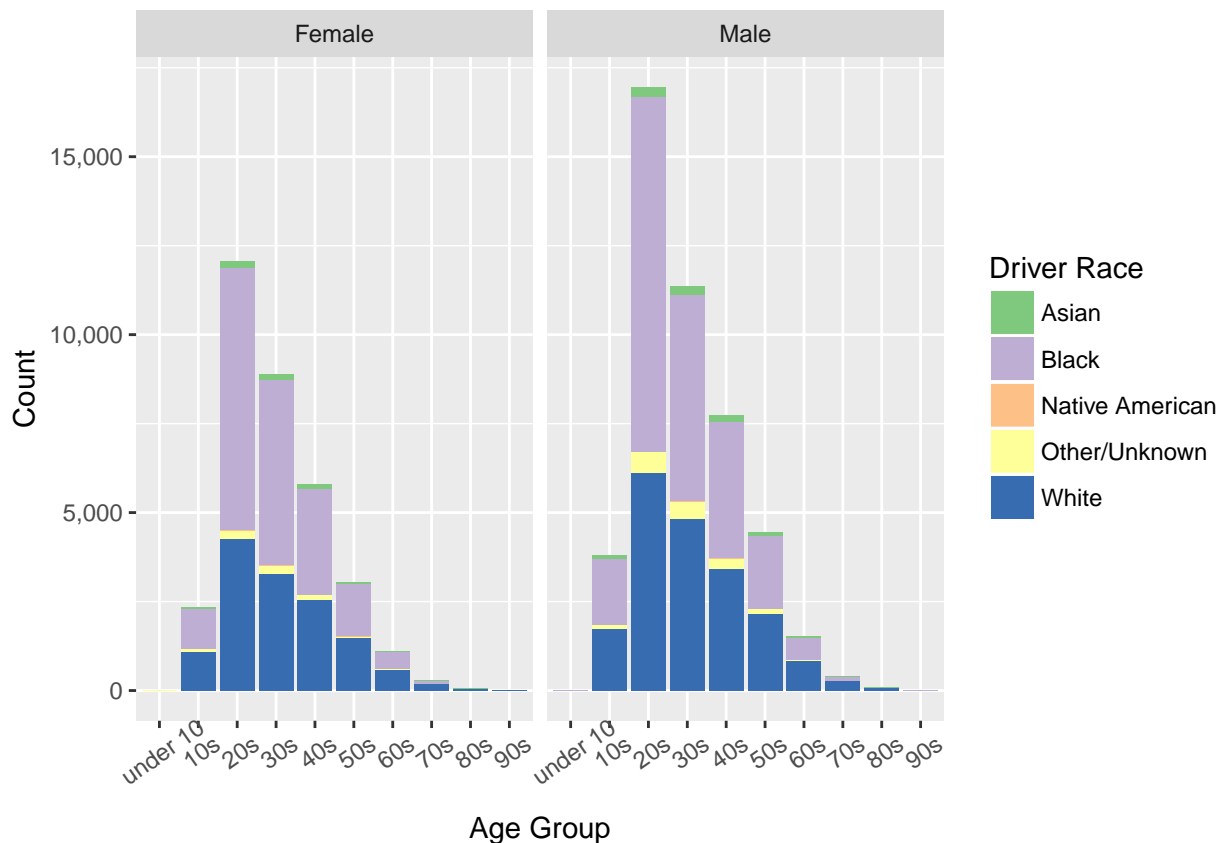
I did some research online, according to the 2010 United States Census, the racial composition of Charlotte was: White or Caucasian: 45.1%, Black or African American: 35.0%, Hispanic: 13.1%, Asian: 5.0%, Native American: 0.5%, Other: 9.6%. This indicates there should be more white drivers, but why are more black driver being stopped?

I selected the data for only white and black drivers, and would like to compare the reasons for stop for both group.The result is displayed in the graph below. One noticable point is that actually more white drivers were stopped for speeading, but the number of black drivers stopped for vehicle regulatory is almost double of that of white driver. Common understanding of vehicle regulatory includes expired registration, missing plates, or anything related to vehicle regulatory issues. So the question remains is that does black driver tends to have more vehicle regulatory issues or they were just simply stopped more often?

```
df %>%
  filter(Driver_Race%in%c("White","Black")) %>%
  group_by(Driver_Race) %>%
  count(Reason_for_Stop) %>%
  ggplot()+
  geom_bar(aes(x=Driver_Race, y=n, fill=Reason_for_Stop),
           position="dodge",stat="identity")+
  labs(fill="Reason for Stop")+
  xlab("Race")+
  ylab("Count")+
  scale_y_continuous(labels=comma)
```
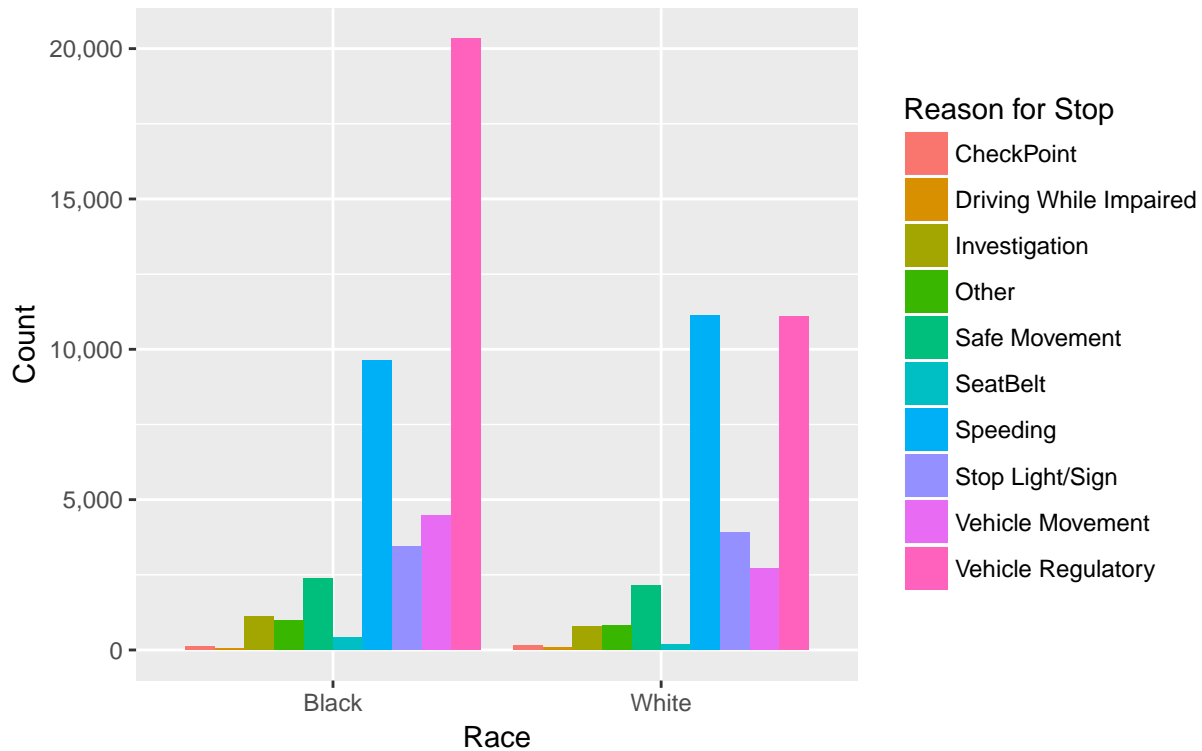
Now I would like to explore whether the race of driver who got stopped is related to the officer's gender. For example, officer of a specific race may be more likely to stop a driver from a different race or vise versa. The graph below summarize the composition of driver's race against officer race. However, it does not seem like it is making a big difference. Except Asian officers, officers from other races tends to stop more black drivers than white drivers.

```
df %>%
  filter(Officer_Race!=" ") %>%
  group_by(Officer_Race) %>%
  count(Driver_Race) %>%
  mutate(feq=round(n/sum(n),4)*100) %>%
  ggplot(aes(x=Officer_Race,y=feq,fill=Driver_Race))+
  geom_bar(stat="Identity", width=0.4)+
  theme(axis.text.x=element_text(size=10, angle=30,vjust=0.8,hjust=0.5))+
  coord_flip()+
  scale_fill_brewer(palette = "Accent")+
  ylab("Percentge(%)")+
  xlab("Officer Race")+
  labs(fill="Driver Race")
```
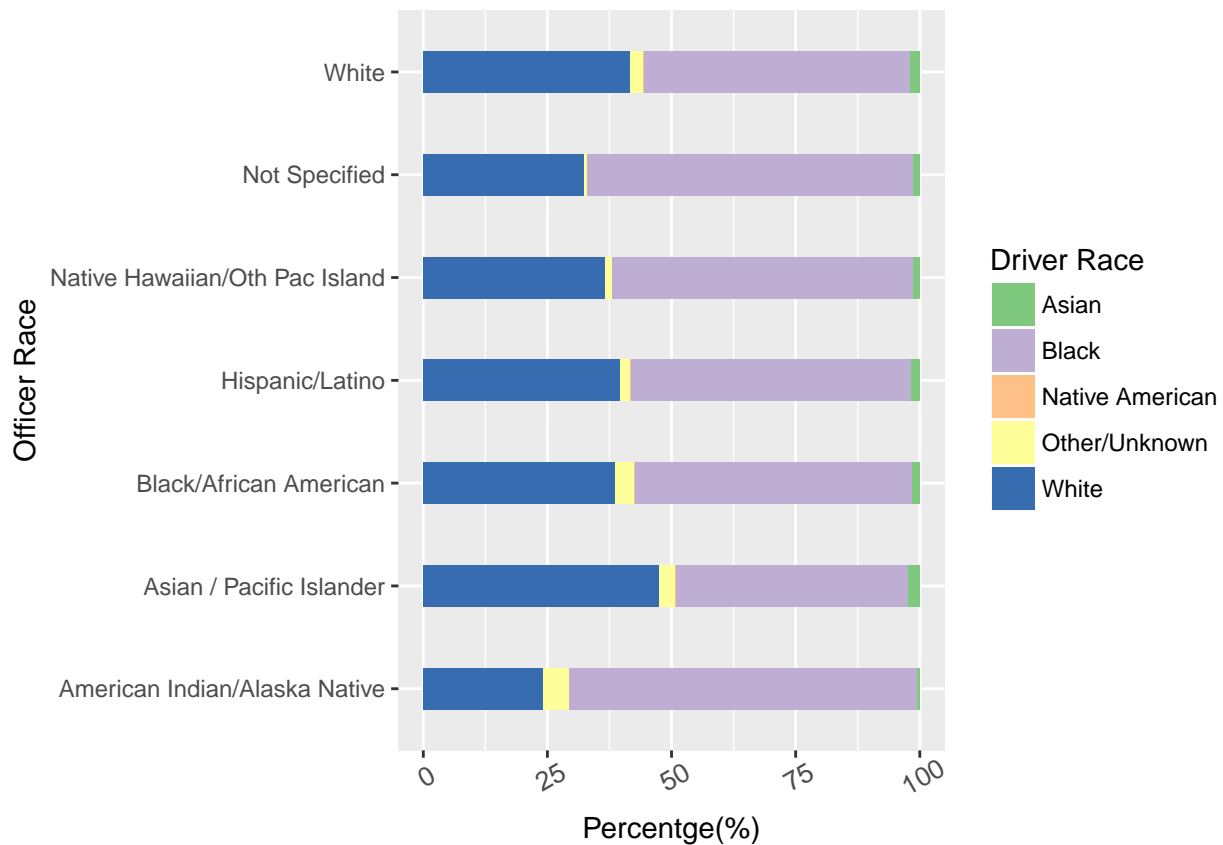
Next, I would like to look at if drivers of different age group or gender will be stopped for different reasons. For example, we may think more percentage of male drivers between age 20 and 30 be stopped due to speeding than other age groups because young male drivers are thought to be more aggresive. However, the diagram des not seem to indicate such a trend. The proportion of different reasons for stopping tends to be pretty similar in each age group and across gender.

```
ggplot(df)+
  geom_bar(aes(x=Driver_Agegroup, fill=Reason_for_Stop),color="white")+
  facet_grid( ~ Driver_Gender)+
  ylab("Count")+
  xlab("Driver Age Group")+
  labs(fill="Reason for Stop")+
  scale_y_continuous(labels=comma)+
  theme(axis.text.x=element_text(size=8.5, angle=45,vjust=0.8,hjust=0.5))
```
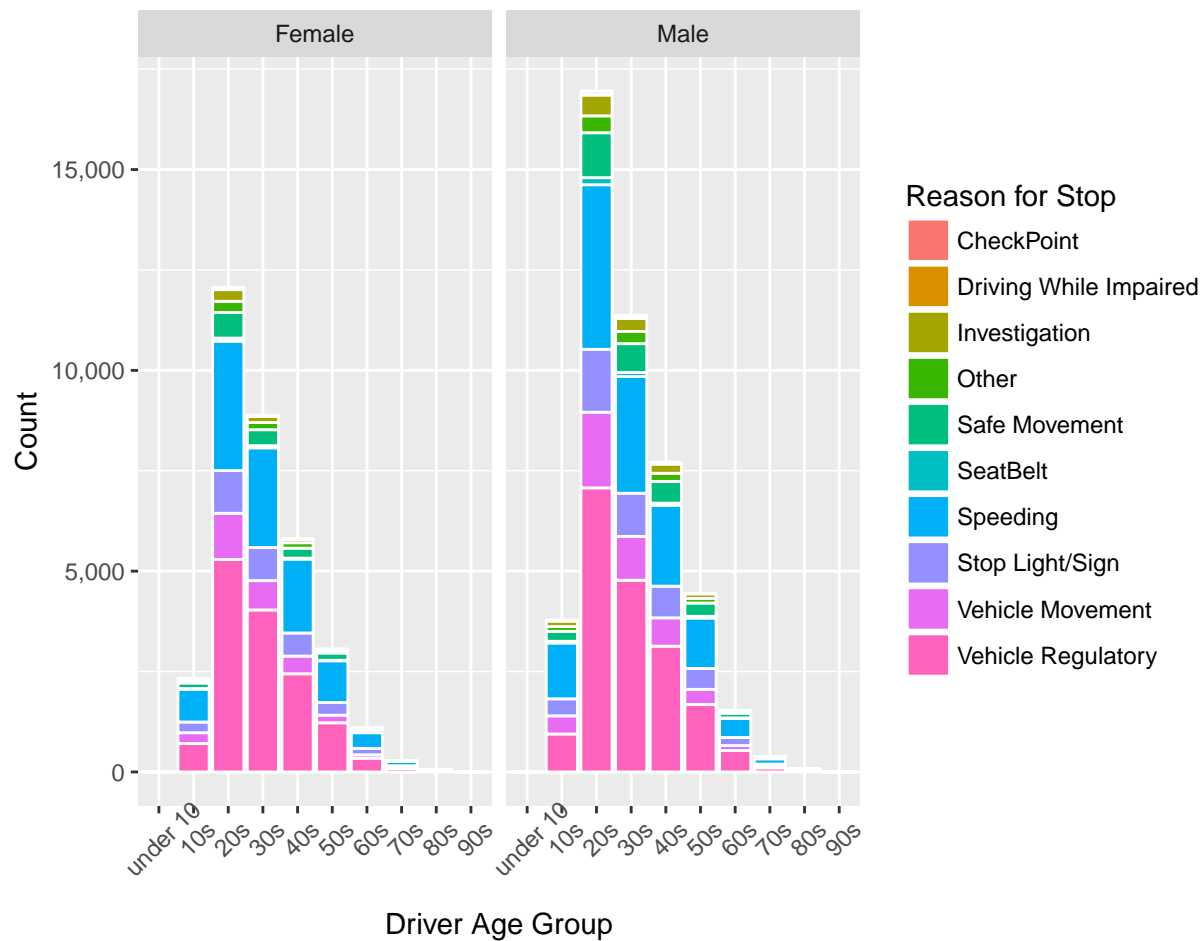
I also summarised the stops made in each months in 2016 in the graph below. Surprisingly, the stops in December is significantly fewer than the other months of the year. Probably it is because as holiday is approaching more officers are on vacation, or people tends to stay home during the holiday month so that are less traffic on the road.

```
df %>%
  group_by(Month_of_Stop) %>%
  count() %>%
  ggplot(aes(x=Month_of_Stop, y=n))+
  geom_bar(stat="identity",fill="#6666CC", width=0.6)+
  xlab("Month")+
  ylab("Number of Stops")+
  theme(plot.title = element_text(hjust = 0.45))+
  theme(axis.text.x=element_text(size=9, angle=20,vjust=0.8,hjust=0.5))
```
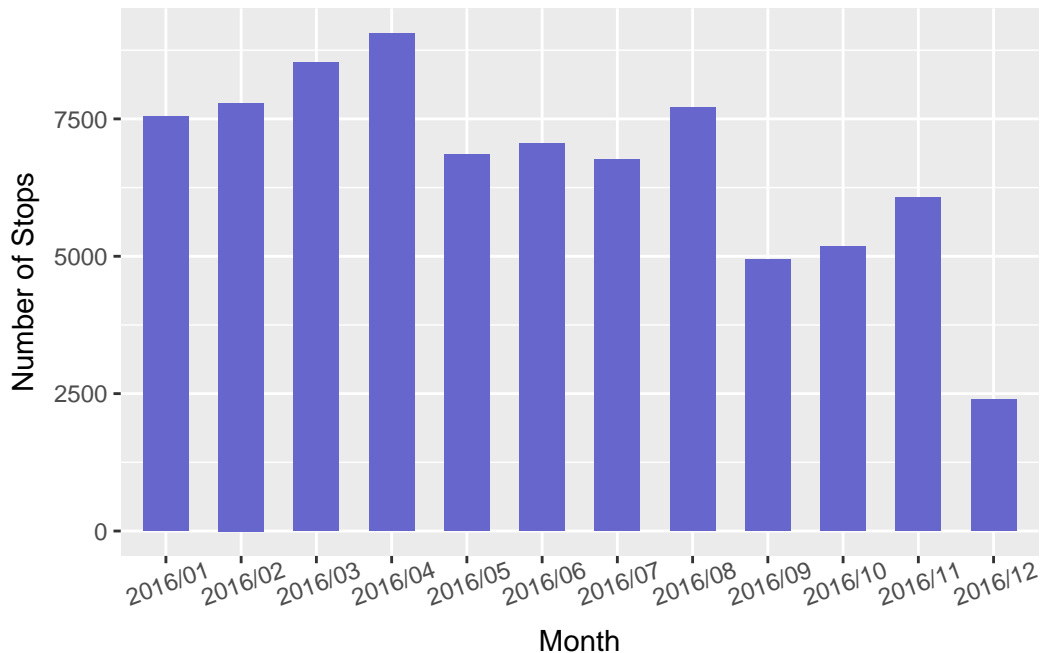
Looking at the result of traffic stops, it looks like different reasons for stop leads to different results. If drivers were stopped because they were drving while imparied, over 75% of time, they would got arrested, and over 70% of pull over at check point either led to arrest or citation. Officers were also quite strict about speeding, more than half of the stops due to speeding ended with citations. It seems that officers have a little higher tolerance when drivers were stopeed due to vehicle movement or safe movement reasons. over 50% of these cases ended without a citation.

```
df %>%
  group_by(Reason_for_Stop) %>%
  count(Result_of_Stop) %>%
  mutate(feq=round(n/sum(n),4)*100) %>%
  ggplot(aes(x=Reason_for_Stop,y= feq,fill=Result_of_Stop))+
  geom_bar(stat="identity", width=0.6)+
  scale_fill_brewer(palette ="Paired")+
  xlab("Reason for Stop")+
  ylab("Percentage of Result of Stop")+
  labs(fill="Result of Stop")+
  theme(axis.text.x=element_text(size=9, angle=20,vjust=0.8,hjust=0.5))
```
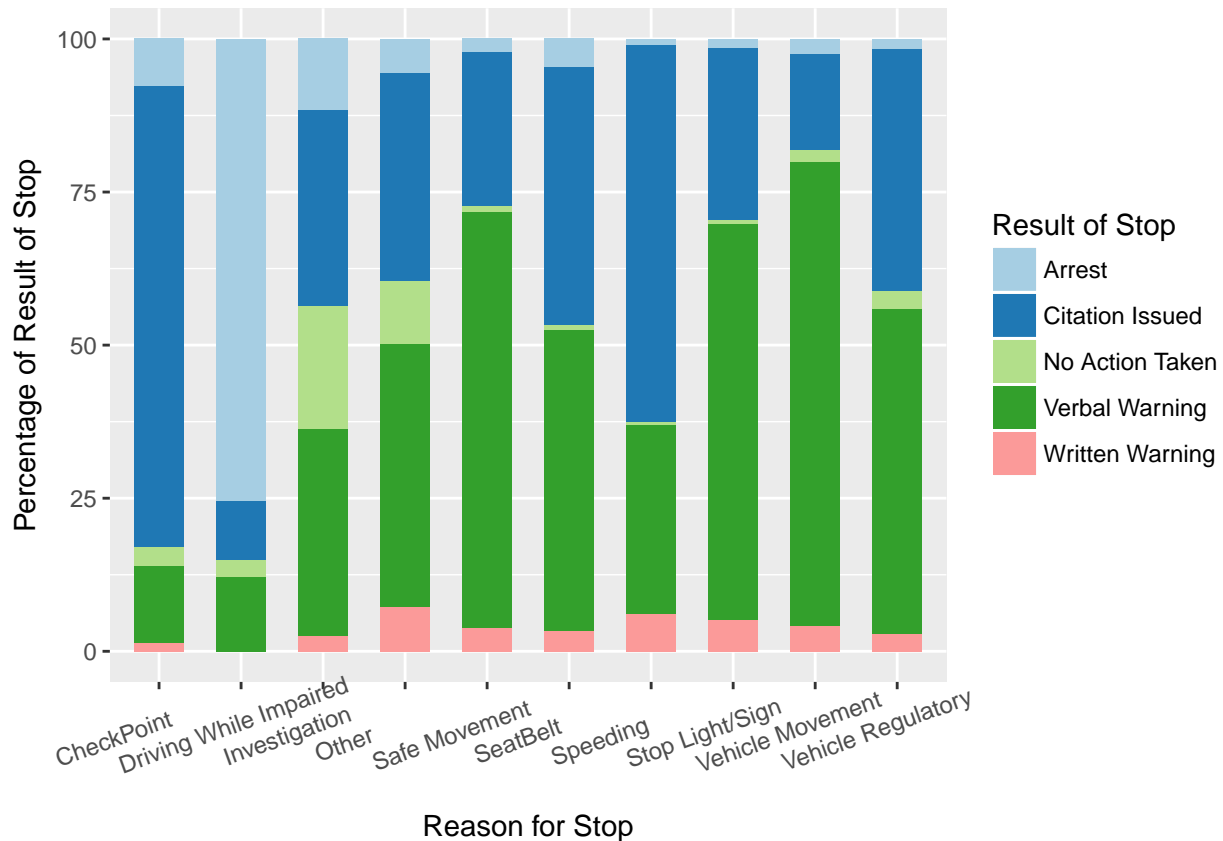
Lastly, I would like to expolore if an officer's gender will affect the result of stops. Would a male officer more strict than a female officer? The diagram below is created to show side by side the result of stop by officer genders. Female officers tends to give a little more verbal warnings and fewer citations, but the difference is not very distinguishable.

```
a <- df %>%
  select(Officer_Gender,Result_of_Stop) %>%
  group_by(Officer_Gender) %>%
  count(Result_of_Stop) %>%
  mutate(prop = n/sum(n))

ggplot(a, aes(x=Result_of_Stop,y=prop))+
  geom_bar(stat="identity",fill="#6666CC", width=0.6)+
  facet_grid( ~ Officer_Gender)+
  scale_y_continuous(labels=percent,limits = c(0,0.6))+
  xlab("Result of Stop")+
  ylab("Percentage")+
  theme(axis.text.x=element_text(size=10, angle=25,vjust=0.8,hjust=0.5))+
  geom_text(data=a,aes(label=paste(round(prop*100,2),"%"), y=prop+0.015),size=3.5)
```