

Playing with Visualization

```
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.2.1 --
## v ggplot2 3.2.1      v purrr  0.3.2
## v tibble  2.1.3      v dplyr  0.8.3
## v tidyr   0.8.3      v stringr 1.4.0
## v readr   1.3.1      v forcats 0.4.0

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()

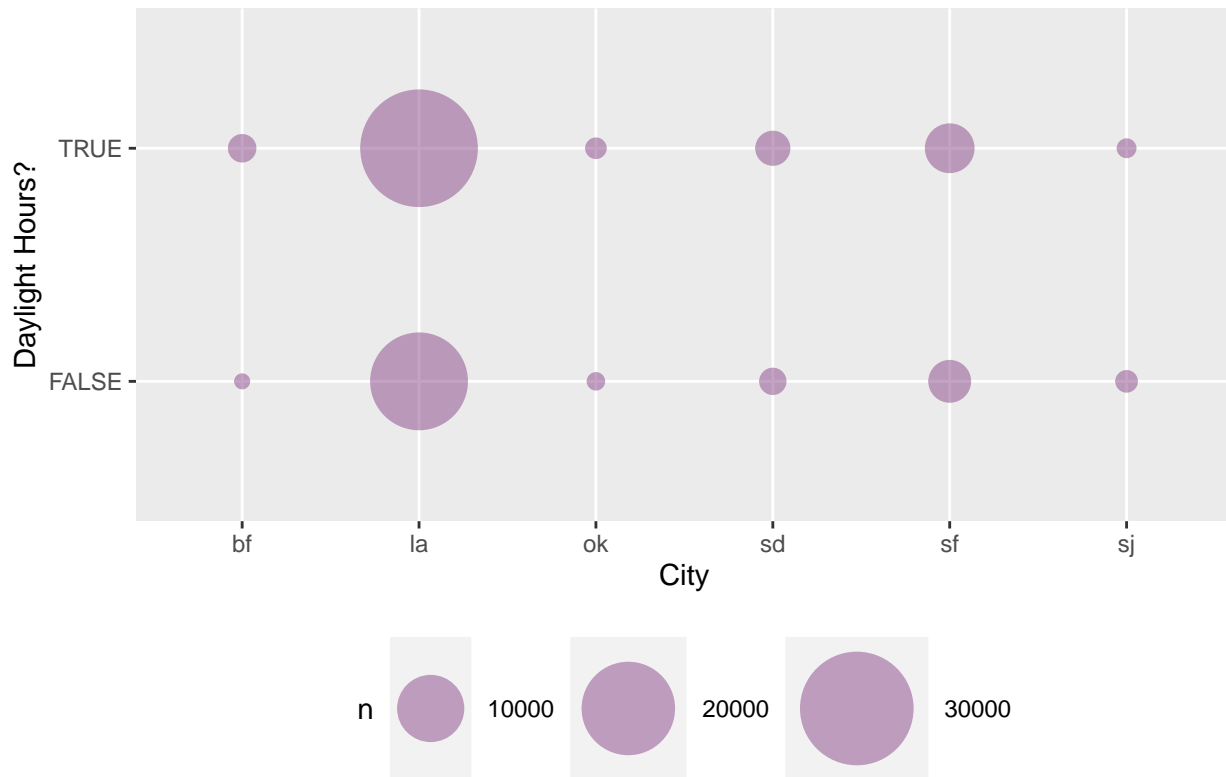
library(dplyr)
library(ggplot2)

load("cadata.RDA")
data.daytime <- filter(ca.df, daytime == "TRUE")
sum.dayvsnight.arrests <- ca.df %>%
  group_by(city, daytime)%>%
  tally()
```

The first visualization I want to investigate is a comparison of the number of arrests that happen across various cities in California and observe any noticeable differences in the number across cities and from daylight to nighttime hours. The size of the count dots indicate the number of arrests. As observed below, Los Angeles has a significantly larger number of arrests across the time period December 2009 to June 2018.

```
ggplot(ca.df)+
  geom_count(aes(x=city, y= daytime), alpha = 0.5, color= "orchid4") + scale_size_area(max_size = 20)+
  labs(x= "City",
       y = "Daylight Hours?",
       title = "Number of daytime and nighttime arrests in major cities in California")+
  theme(plot.title = element_text(hjust = 0.5))+ #puts title in the center
  theme(legend.position = "bottom") +
  scale_fill_discrete(name = "Number of Arrests") #supposed to change title of the legend
```

Number of daytime and nighttime arrests in major cities in California



The next area that would be interesting for visualization would be what proportion of arrests for the daytime were non-white as compared to the proportion at night.

The pie chart

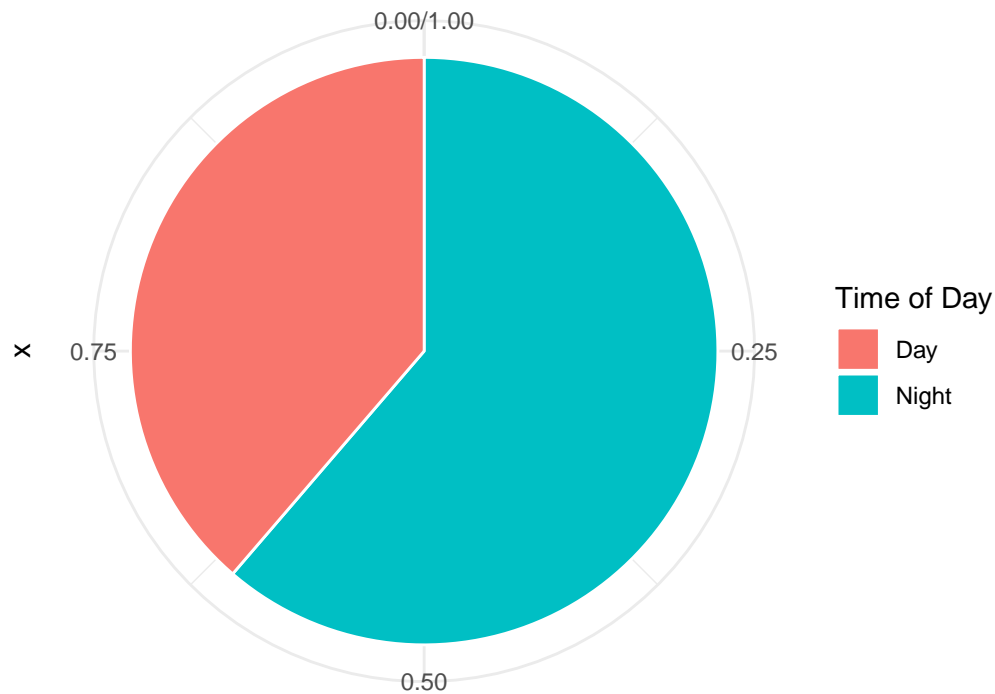
```
arrests.white.sf <- filter(ca.df, city == "sf", subject_race == "white") %>%
  group_by(daytime) %>%
  tally() %>%
  mutate(proportion = n / sum(n)) %>%
  mutate_at(vars(proportion), funs(round(., 3)))
```

```
## Warning: funs() is soft deprecated as of dplyr 0.8.0
## Please use a list of either functions or lambdas:
##
##   # Simple named list:
##   list(mean = mean, median = median)
##
##   # Auto named with `tibble::lst()` :
##   tibble::lst(mean, median)
##
##   # Using lambdas
##   list(~ mean(., trim = .2), ~ median(., na.rm = TRUE))
## This warning is displayed once per session.
```

```
ggplot(arrests.white.sf, aes(x = "", y = proportion, fill = daytime)) +
  geom_bar(width = 1, stat = "identity", color = "white") +
  coord_polar("y", start = 0) +
  theme_minimal() + labs(title = "Stops of white people in San Francisco") +
  labs(fill = "Time of Day") +
```

```
scale_fill_discrete(labels = c("Day", "Night"))
```

Stops of white people in San Francisco



proportion

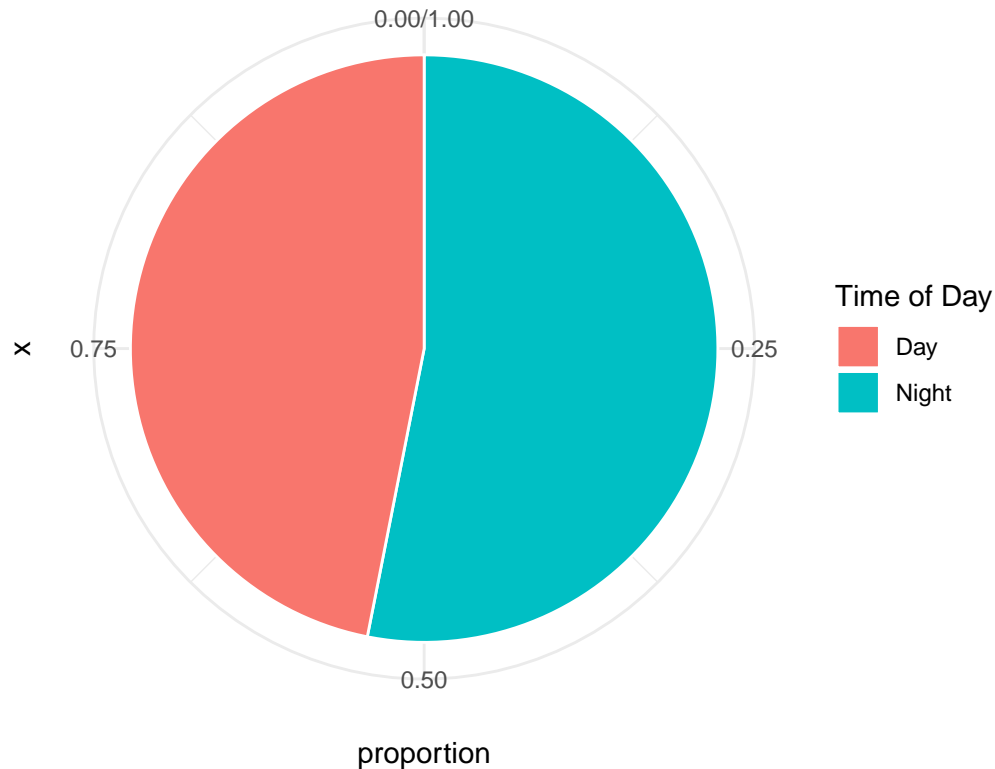
The proportion should

be the same for black and hispanic drivers.

```
arrests.black.sf <- filter(ca.df, city == "sf", subject_race == "black") %>%
  group_by(daytime) %>%
  tally() %>%
  mutate(proportion = n / sum(n) ) %>%
  mutate_at(vars(proportion), funs(round(., 3)))

ggplot(arrests.black.sf, aes(x = "", y = proportion, fill = daytime))+
  geom_bar(width = 1, stat = "identity", color = "white") +
  coord_polar("y", start = 0)+
  theme_minimal() +
  labs(title = "Stops of black people in San Francisco")+
  labs(fill = "Time of Day")+
  scale_fill_discrete(labels = c("Day", "Night"))
```

Stops of black people in San Francisco

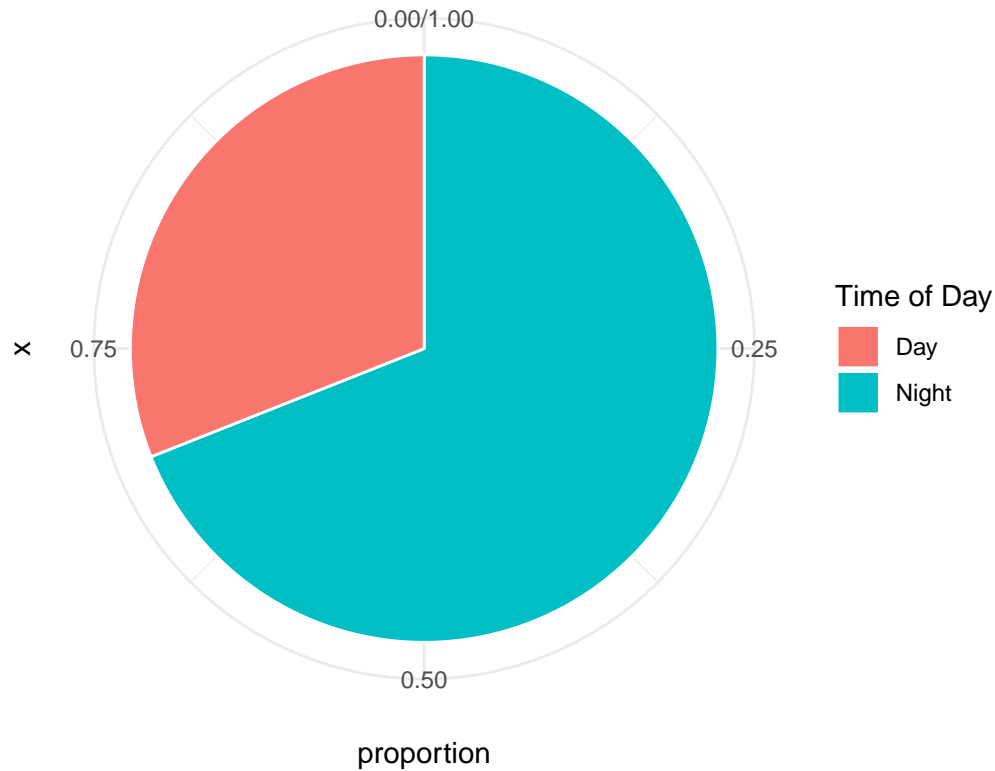


The above visualization can be created for Los Angeles and San Jose data.

```
arrests.white.la <- filter(ca.df, city == "la", subject_race == "white") %>%
  group_by(daytime) %>%
  tally() %>%
  mutate(proportion = n / sum(n) ) %>%
  mutate_at(vars(proportion), funs(round(., 3)))

ggplot(arrests.white.la, aes(x = "", y = proportion, fill = daytime))+
  geom_bar(width = 1, stat = "identity", color = "white") +
  coord_polar("y", start = 0)+
  theme_minimal() +
  labs(title = "Stops of white people in LA")+
  labs(fill = "Time of Day")+
  scale_fill_discrete(labels = c("Day", "Night"))
```

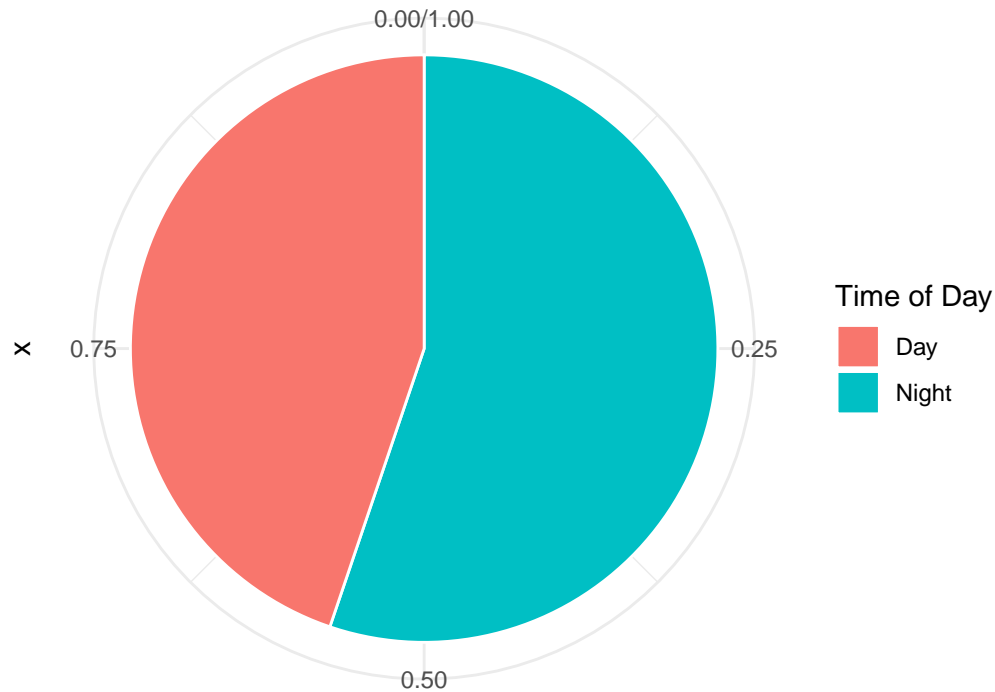
Stops of white people in LA



```
arrests.black.la <- filter(ca.df, city == "la", subject_race == "black") %>%
  group_by(daytime) %>%
  tally() %>%
  mutate(proportion = n / sum(n) ) %>%
  mutate_at(vars(proportion), funs(round(., 3)))

ggplot(arrests.black.la, aes(x = "", y = proportion, fill = daytime))+
  geom_bar(width = 1, stat = "identity", color = "white") +
  coord_polar("y", start = 0)+
  theme_minimal() +
  labs(title = "Stops of black people in LA")+
  labs(fill = "Time of Day")+
  scale_fill_discrete(labels = c("Day", "Night"))
```

Stops of black people in LA



proportion

From the pie charts above, we can observe that in even just these two cities, something interesting is going on. Out of the total number of black people stopped, there is a fairly equal distribution from night to day. This means that approximately half the stops were during the day and half at night. However, for both cities, a significantly larger proportion of the total number of white people stopped was at night as compared to the day. If this trend exists across all cities then this means that white people were stopped more often at night than during the day when it was easier to see the driver in the car.

Further Investigation

I am now inclined to