

EDA Time Series Plots

Weijia Bao

10/03/2019

```
library(RColorBrewer)
library(ggplot2)
library(tidyverse)
library(lubridate)
```

```
Shooting_data <- read.csv("~/Downloads/NYPD_Shooting_Incident_Data_Historic.csv")
Shooting_data$Numeric_time = as.numeric(hms(Shooting_data$OCCUR_TIME))/60
```

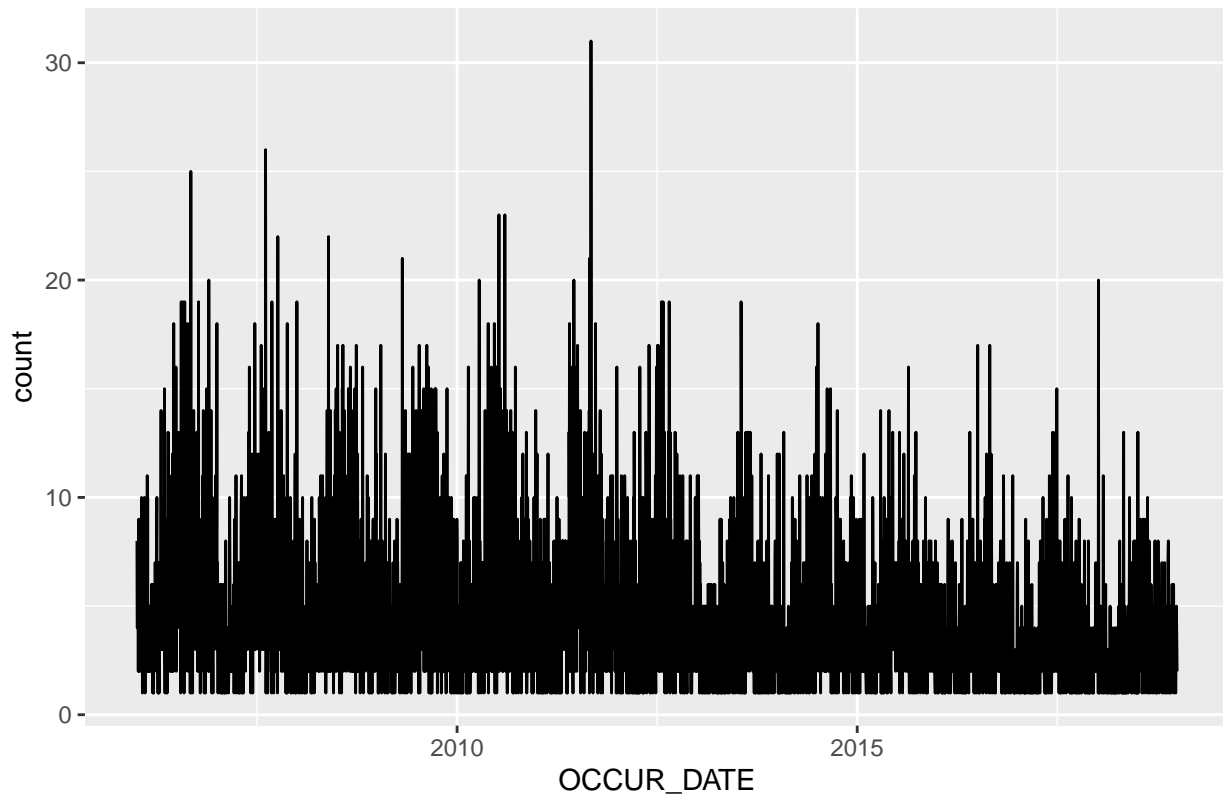
Time Series

- stat plot

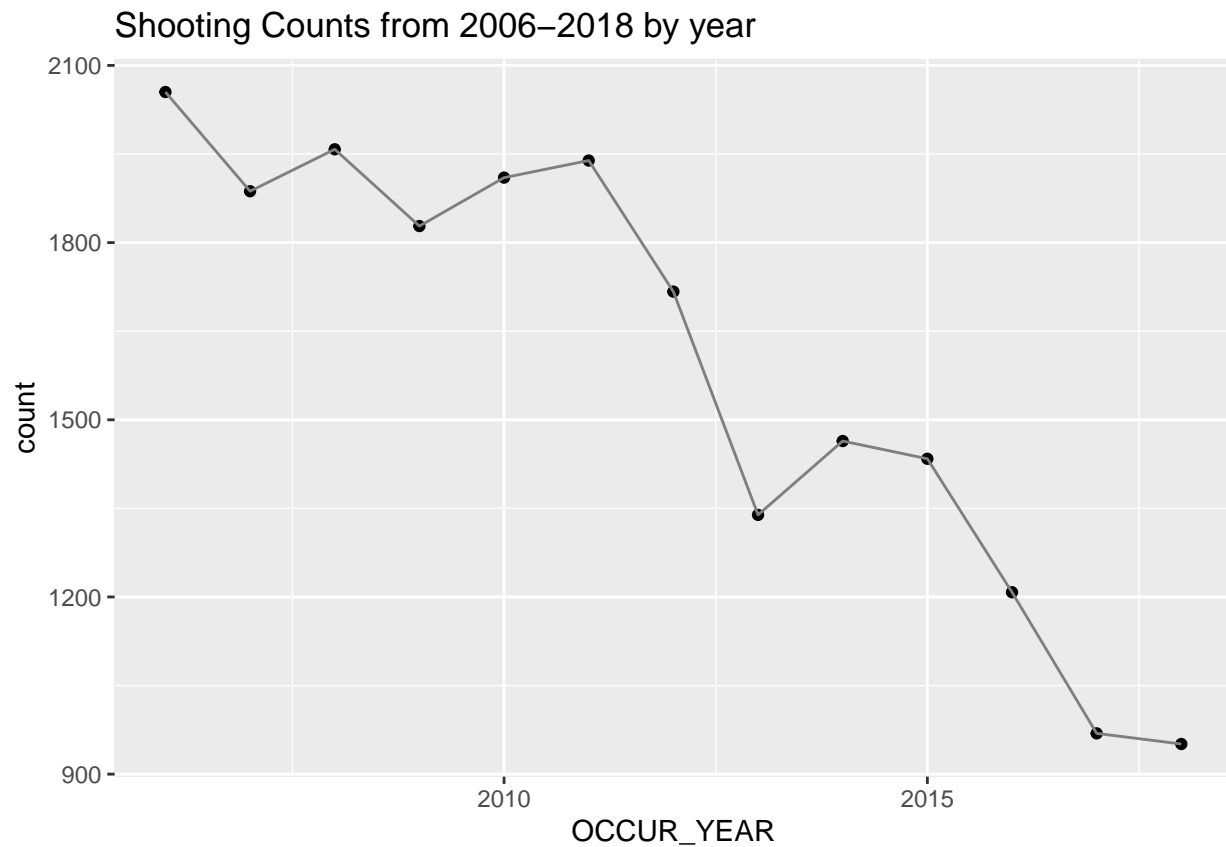
```
# Change the time format
Shooting_data$OCCUR_DATE<- format(as.Date(Shooting_data$OCCUR_DATE, format = "%m/%d/%Y"), "%Y-%m-%d")
Shooting_data$OCCUR_DATE <- as.Date(Shooting_data$OCCUR_DATE)

# overview
overall_data<- Shooting_data %>% group_by(OCCUR_DATE) %>% summarize(count = n())
ggplot(overall_data, aes(OCCUR_DATE, count))+
  geom_line()+
  ggtitle("Overview Shooting Counts from 2006-2018")
```

Overview Shooting Counts from 2006–2018



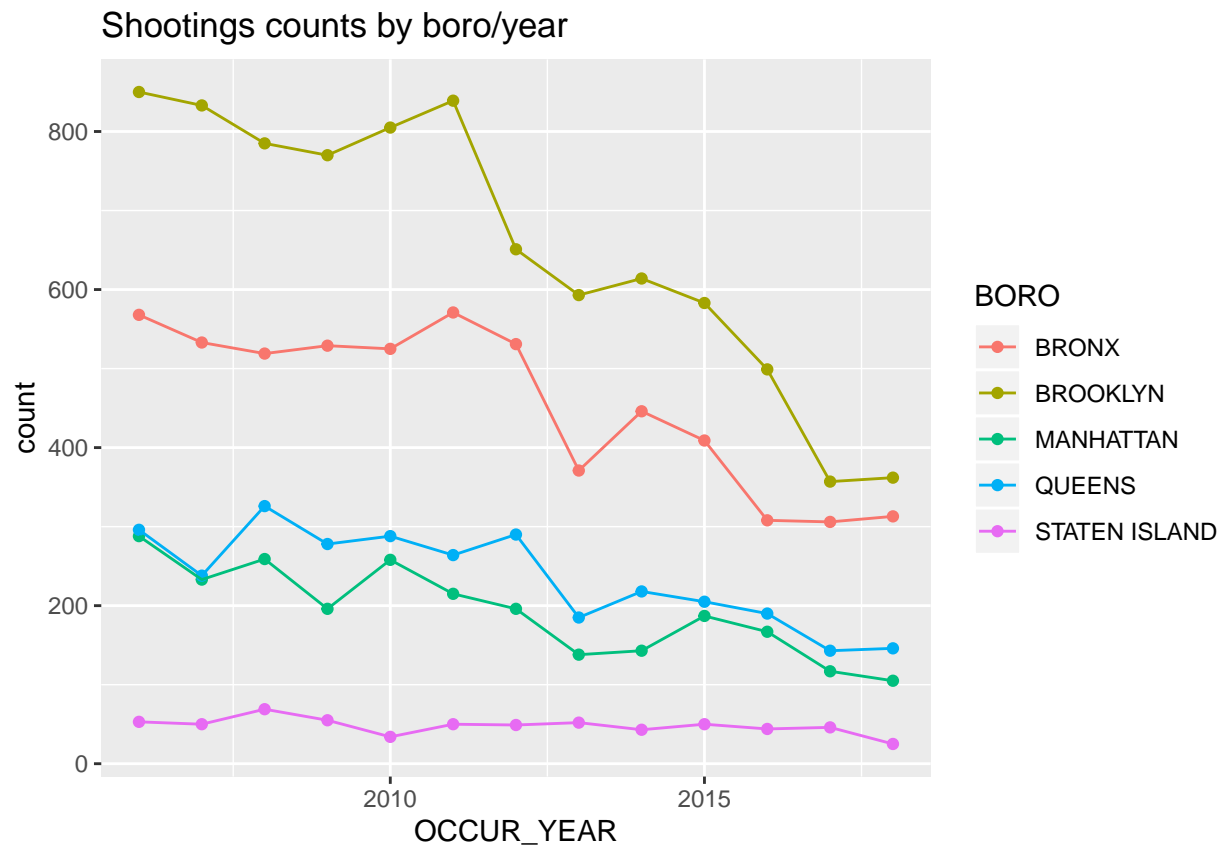
```
##### Year #####  
# Shooting counts by year  
Shooting_data$OCCUR_YEAR <- year(Shooting_data$OCCUR_DATE)  
  
year_data <- Shooting_data %>% group_by(OCCUR_YEAR) %>% summarize(count = n())  
ggplot(year_data, aes(OCCUR_YEAR, count))+  
  geom_point()+  
  geom_line(color = "grey50")+  
  ggtitle("Shooting Counts from 2006-2018 by year")
```



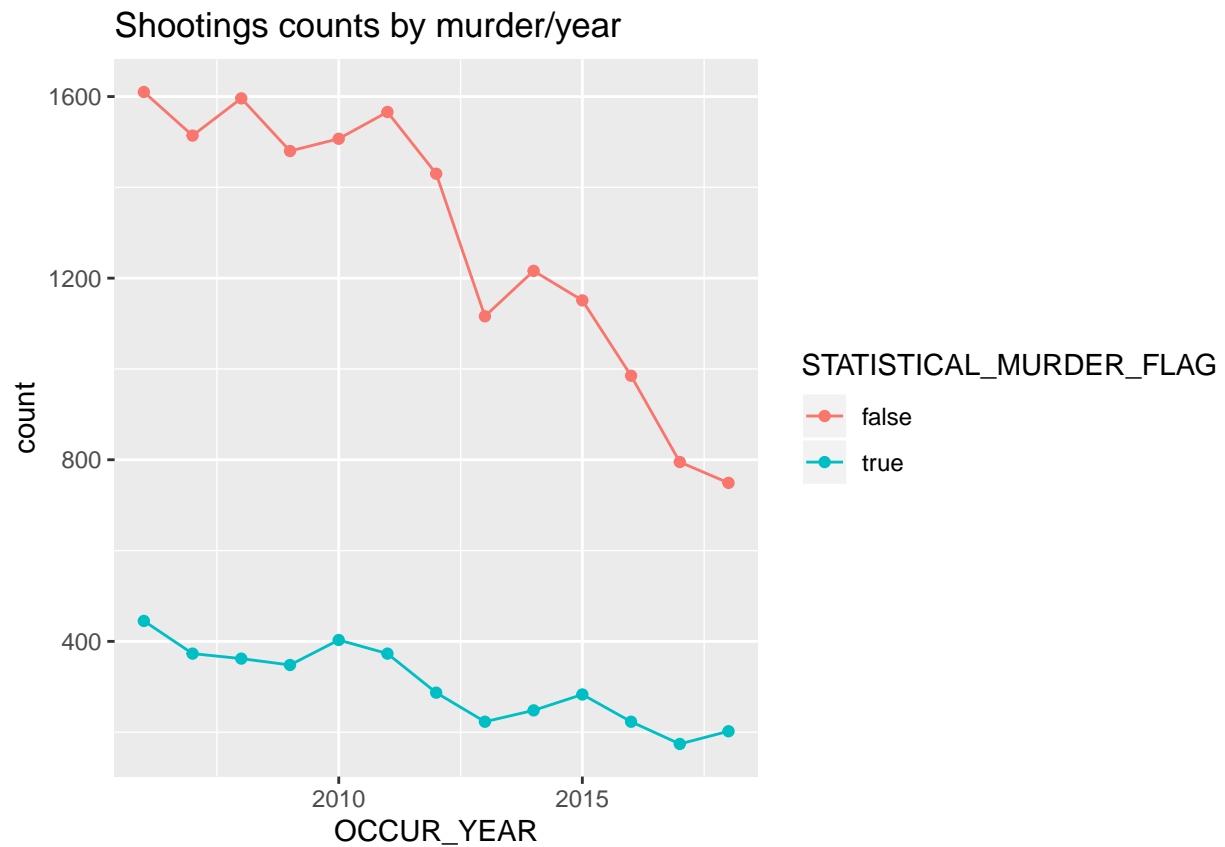
```
# year shooting by boro
year_boro <- Shooting_data %>% group_by(OCCUR_YEAR, BORO) %>%
  summarize(count = n())

#year_boro <- Shooting_data %>% group_by(Month = floor_date(OCCUR_DATE, "month"), BORO) %>%
  #summarize(count = n())

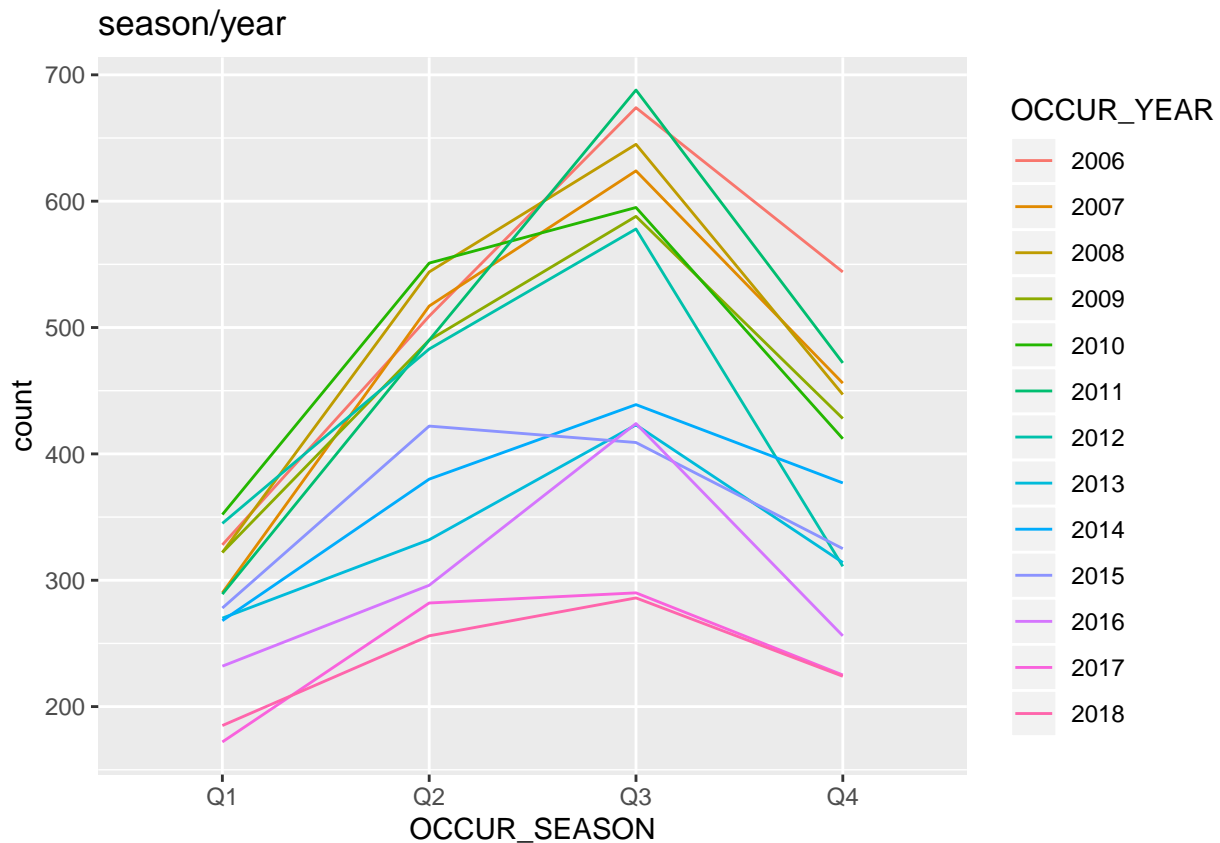
ggplot(year_boro, aes(OCCUR_YEAR, count, color = BORO))+
  geom_point()+
  geom_line(aes(group= BORO))+
  ggtitle("Shootings counts by boro/year")
```



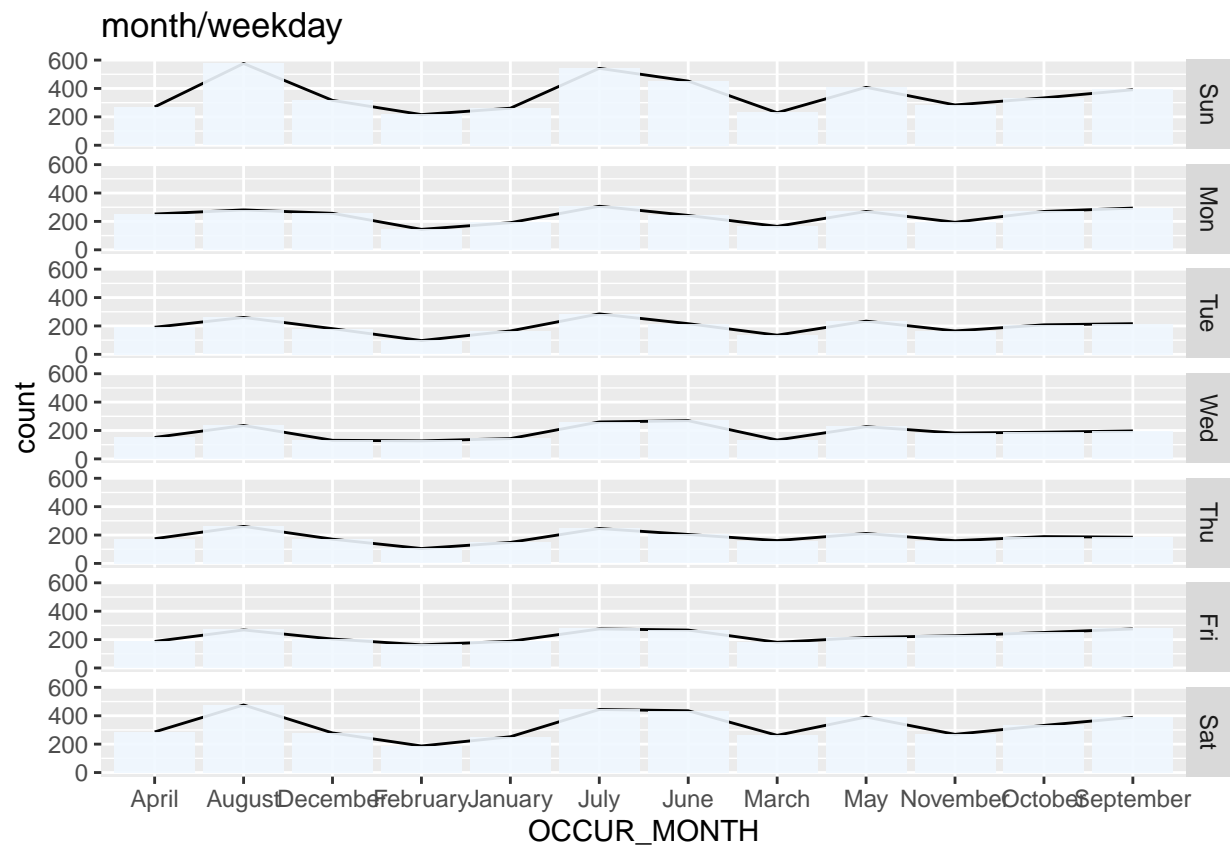
```
# year shooting by murder
year_murder <- Shooting_data %>% group_by(OCCUR_YEAR, STATISTICAL_MURDER_FLAG) %>%
  summarize(count = n())
ggplot(year_murder, aes(OCCUR_YEAR, count, color = STATISTICAL_MURDER_FLAG))+
  geom_point()+
  geom_line(aes(group= STATISTICAL_MURDER_FLAG))+
  ggtitle("Shootings counts by murder/year")
```



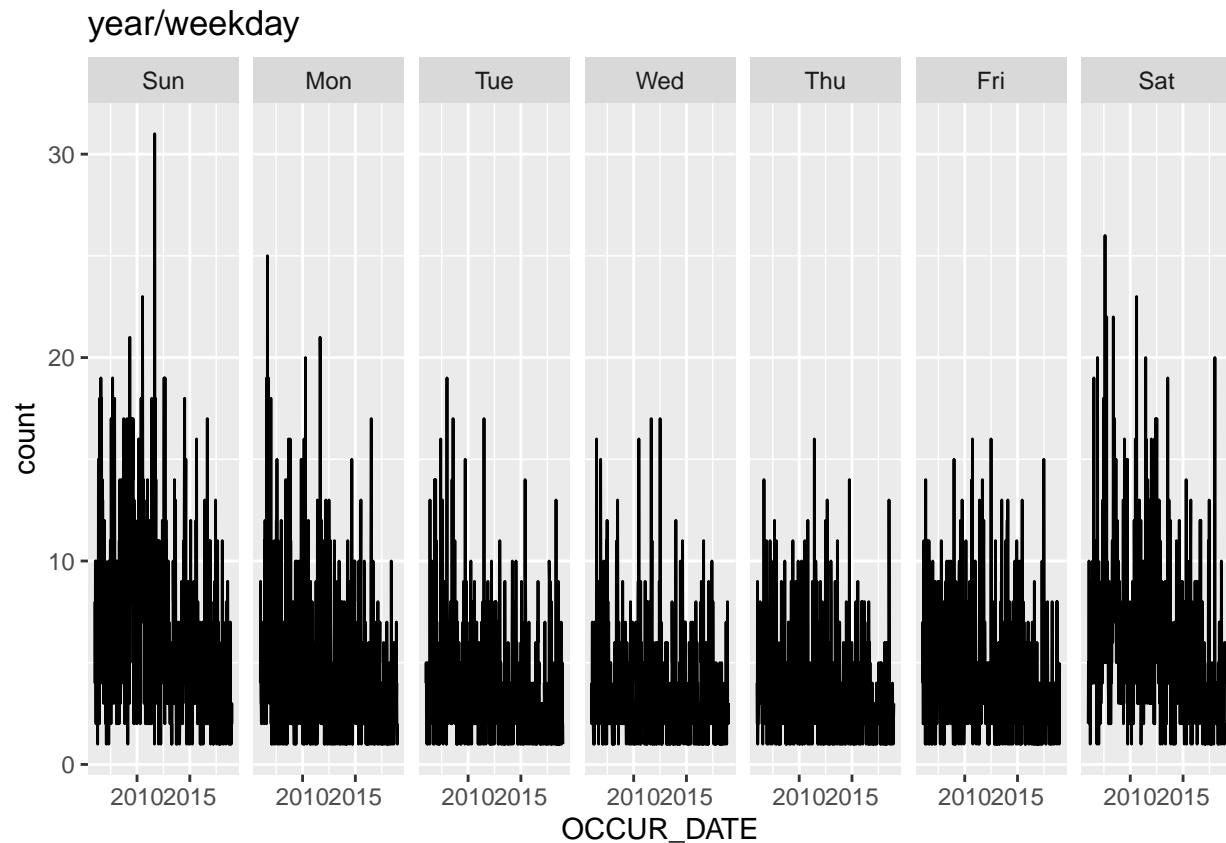
```
##### Season #####
Shooting_data$OCCUR_YEAR<-as.factor(Shooting_data$OCCUR_YEAR)
Shooting_data$OCCUR_SEASON<-quarters(Shooting_data$OCCUR_DATE)
# season by year
season_year <- Shooting_data %>% group_by(OCCUR_SEASON, OCCUR_YEAR) %>%
  summarize(count = n())
ggplot(season_year, aes(OCCUR_SEASON, count, color = OCCUR_YEAR))+
  geom_line(aes(group= OCCUR_YEAR))+
  ggtitle("season/year")
```



```
##### Month & Weekday #####
Shooting_data$OCCUR_MONTH<-months(Shooting_data$OCCUR_DATE)
Shooting_data$OCCUR_WEEKDAY <-wday(Shooting_data$OCCUR_DATE, label = TRUE)
# month by weekday
month_weekday <- Shooting_data %>% group_by(OCCUR_MONTH, OCCUR_WEEKDAY) %>%
  summarize(count = n())
# facets
ggplot(month_weekday, aes(OCCUR_MONTH, count))+
  geom_line(aes(group= OCCUR_WEEKDAY))+
  facet_grid(OCCUR_WEEKDAY~.)+
  geom_bar(stat = "identity", fill="aliceblue", alpha= 0.9)+
  ggtitle("month/weekday")
```

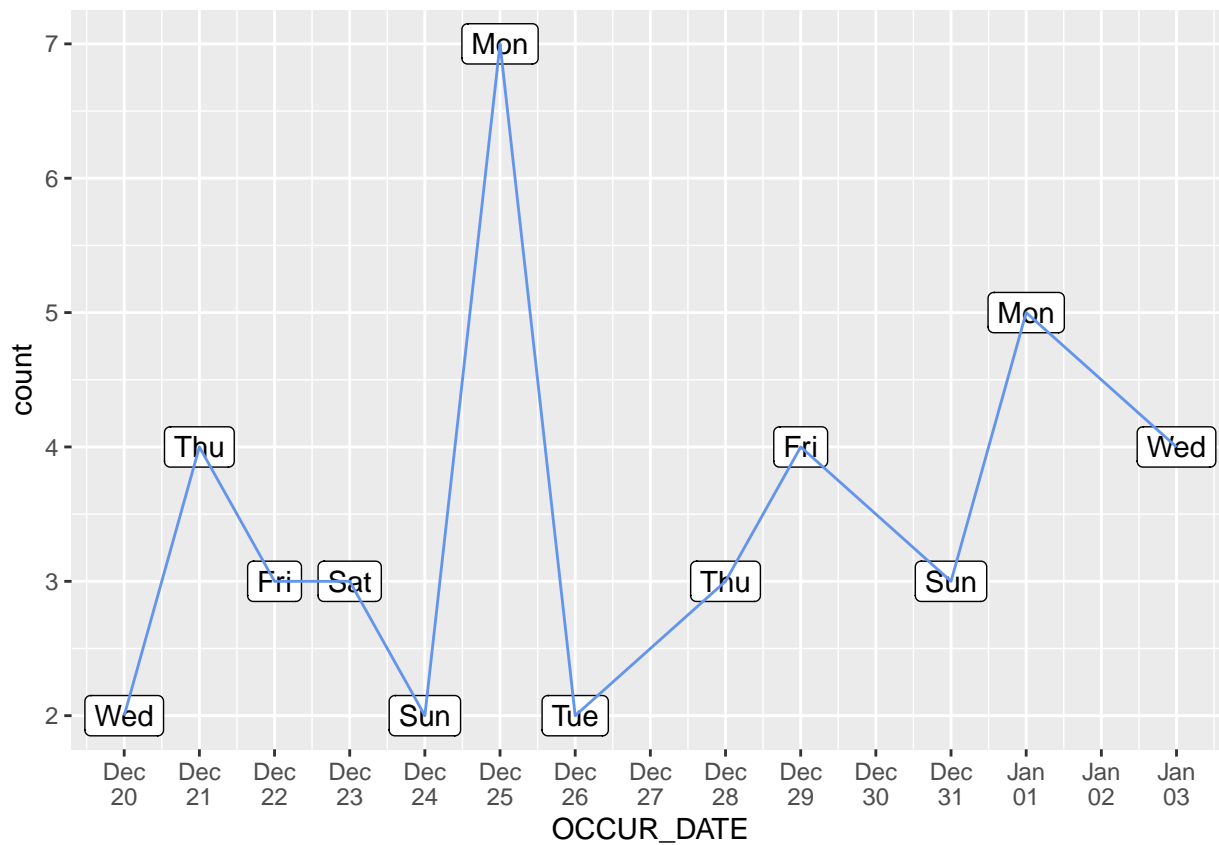


```
ggplot(overall_data, aes(OCCUR_DATE, count))+
  geom_line()+
  facet_grid(~wday(overall_data$OCCUR_DATE, label = TRUE))+
  ggtitle("year/weekday")
```

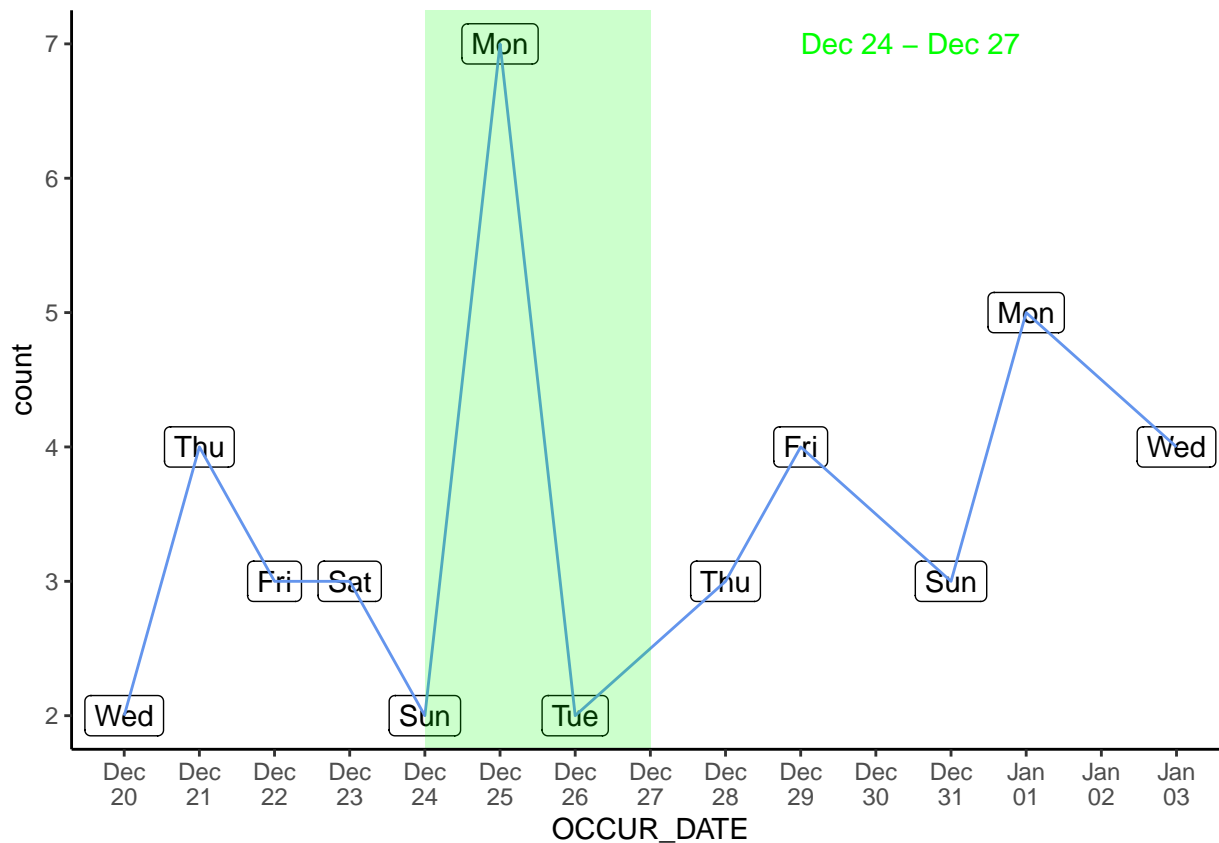


```
##### Different pattern for Christmas Week #####
# Take 2017 for example
# Label by day of month
christmas_2017 <- Shooting_data %>%
  filter(OCCUR_DATE >= as.Date("2017-12-20") &
         OCCUR_DATE <= as.Date("2018-01-03")) %>%
  group_by(OCCUR_DATE) %>%
  summarize(count = n())

g <- ggplot(christmas_2017, aes(OCCUR_DATE, count)) +
  geom_label(aes(label = wday(OCCUR_DATE, label = TRUE))) +
  geom_line(color = "cornflowerblue") +
  scale_x_date(date_labels = "%b\n%d",
               date_breaks = "1 day")
g
```

```
# Highlight the abnormality
start <- as.Date("2017-12-24")
end <- as.Date("2017-12-27")
g+annotate("rect", xmin = start, xmax = end,
            ymin = -Inf, ymax = Inf, fill = "green",
            alpha = .2) +
  annotate("text", x = end + 2,
          y = 7, label = "Dec 24 - Dec 27",
          color = "green", hjust = 0) +
  theme_classic()
```



```
##### Independence Day #####
# Take 2017 for example
year2017 <- Shooting_data %>%
  filter(OCCUR_DATE >= as.Date("2017-01-01") &
         OCCUR_DATE <= as.Date("2017-12-31")) %>%
  group_by(OCCUR_DATE) %>%
  summarize(count = n())
ggplot(year2017, aes(OCCUR_DATE, count)) +
  geom_line() +
  geom_vline(xintercept = ymd("2017-07-04"), color = "blue", size = 1) +
  scale_x_date(date_labels = "%b %y") +
  theme_grey(16)
```

