

2018110493_정정룡_0526

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5      v purrr   0.3.4
## v tibble  3.1.6      v dplyr   1.0.8
## v tidyr   1.2.0      v stringr 1.4.0
## v readr   2.1.2      v forcats 0.5.1
```

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

```
library(modelr)
library(splines)
library(MASS)
```

```
##
## 다음의 패키지를 부착합니다: 'MASS'
```

```
## The following object is masked from 'package:dplyr':
##
##      select
```

```
library(lubridate)
```

```
##
## 다음의 패키지를 부착합니다: 'lubridate'
```

```
## The following objects are masked from 'package:base':
##
##      date, intersect, setdiff, union
```

```
bikes <- ISLR2::Bikeshare %>%
  tibble()

bikes <- bikes %>%
  mutate(
    season = factor(season, labels = c("winter", "spring", "summer", "fall"))
  ) %>%
  mutate(
    date = as.Date(day, origin = "2010-12-31"),
    datetime = make_datetime(year(date), month(date), mday(date), hr),
    weekday=as.factor(weekday),
    holiday=as.factor(holiday)
  )

glimpse(bikes)
```

```
## Rows: 8,645
## Columns: 17
## $ season      <fct> winter, winter, winter, winter, winter, winter, winter, win~
## $ mnth        <fct> Jan, Jan, Jan, Jan, Jan, Jan, Jan, Jan, Jan, Jan, Jan, Jan,~
## $ day         <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,~
## $ hr          <fct> 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 1~
## $ holiday     <fct> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
## $ weekday     <fct> 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6,~
## $ workingday  <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
## $ weathersit   <fct> clear, clear, clear, clear, clear, cloudy/misty, clear, cle~
## $ temp        <dbl> 0.24, 0.22, 0.22, 0.24, 0.24, 0.24, 0.22, 0.20, 0.24, 0.32,~
## $ atemp       <dbl> 0.2879, 0.2727, 0.2727, 0.2879, 0.2879, 0.2576, 0.2727, 0.2~
## $ hum         <dbl> 0.81, 0.80, 0.80, 0.75, 0.75, 0.75, 0.80, 0.86, 0.75, 0.76,~
## $ windspeed   <dbl> 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0896, 0.0000, 0.0~
## $ casual      <dbl> 3, 8, 5, 3, 0, 0, 2, 1, 1, 8, 12, 26, 29, 47, 35, 40, 41, 1~
## $ registered  <dbl> 13, 32, 27, 10, 1, 1, 0, 2, 7, 6, 24, 30, 55, 47, 71, 70, 5~
## $ bikers      <dbl> 16, 40, 32, 13, 1, 1, 2, 3, 8, 14, 36, 56, 84, 94, 106, 110~
## $ date        <date> 2011-01-01, 2011-01-01, 2011-01-01, 2011-01-01, 2011-01-01~
## $ datetime    <dtm> 2011-01-01 01:00:00, 2011-01-01 02:00:00, 2011-01-01 03:00~
```

```
colSums(is.na(bikes)) #결측치 없음
```

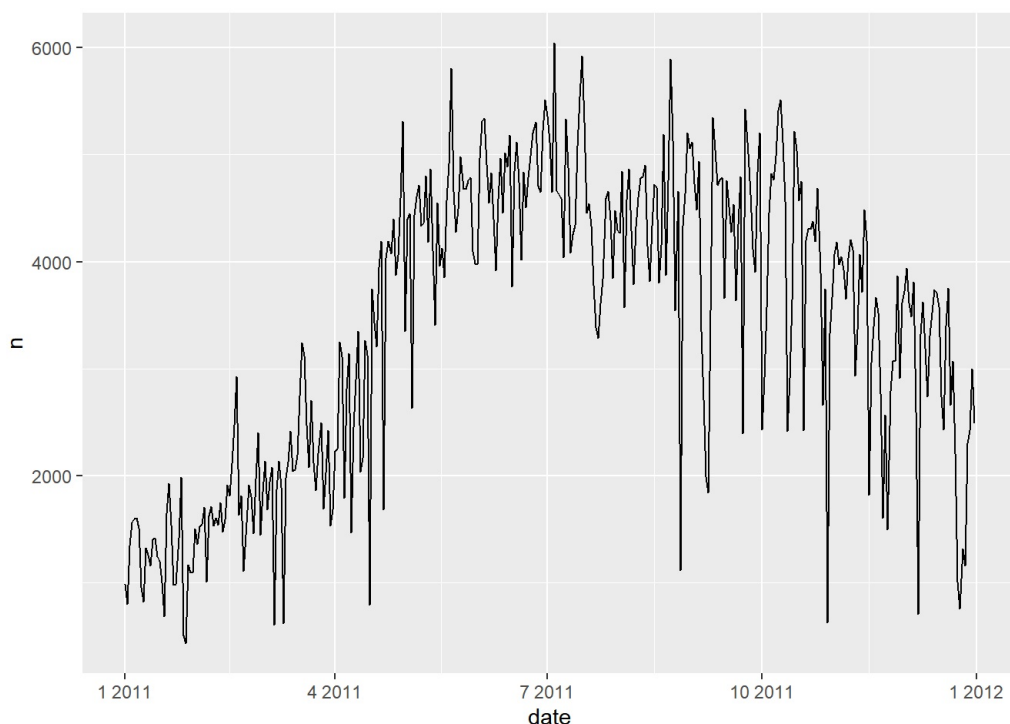
```
##      season      mnth      day      hr      holiday      weekday workingday
##      0          0          0          0          0          0          0
## weathersit      temp      atemp      hum      windspeed      casual registered
##      0          0          0          0          0          0          0
##      bikers      date      datetime
##      0          0          0
```

```
df=bikes %>%
  group_by(date) %>%
  summarise(n=n())
table(df[,2])      #유의해야할 행 존재
```

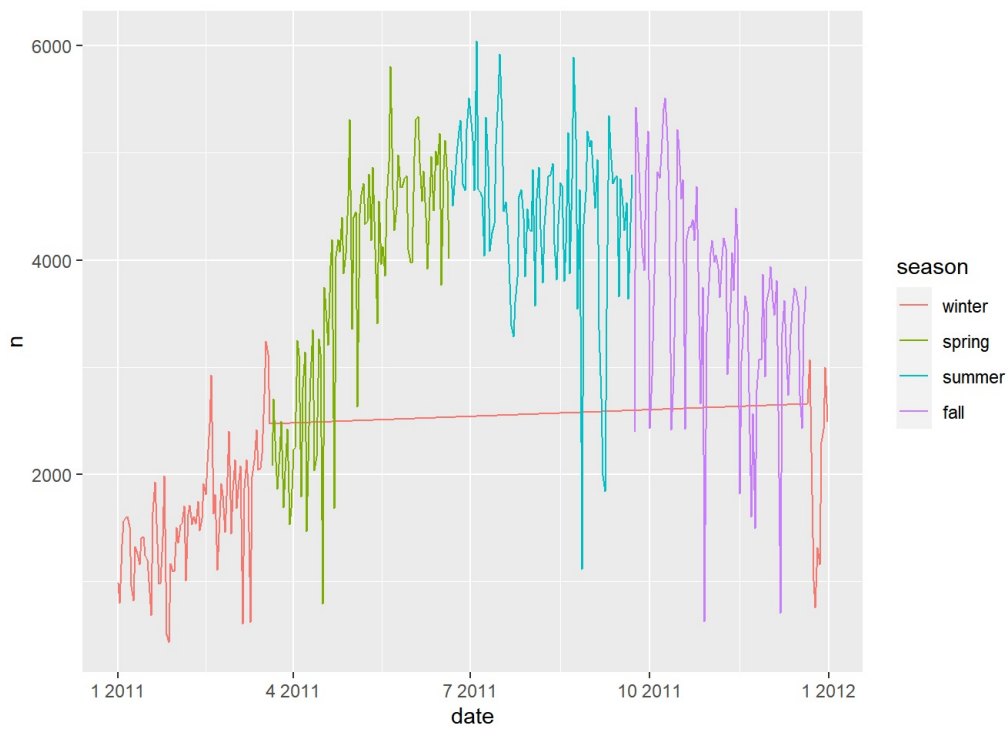
```
##
##  8 12 16 17 18 22 23 24
##  1  1  1  1  2  6 48 305
```

```
#daily
daily=bikes %>%
  group_by(date) %>%
  mutate(n=sum(bikers))

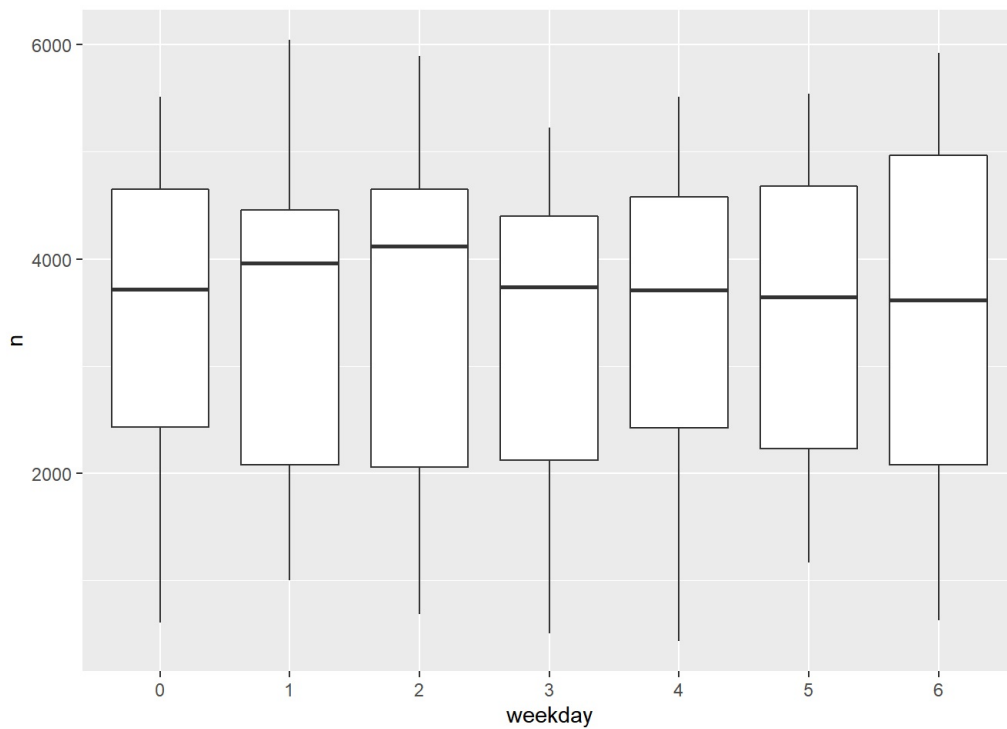
#일별
daily %>%
  ggplot()+
  geom_line(aes(date,n))
```



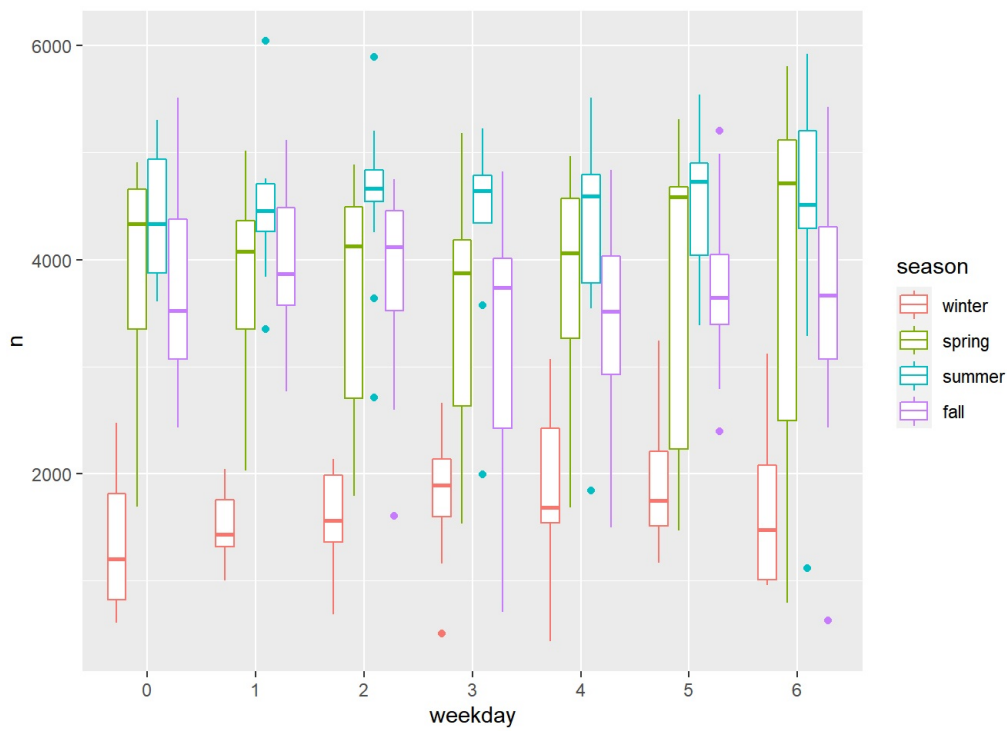
```
daily %>%
  ggplot()+
  geom_line(aes(date,n, color=season))
```



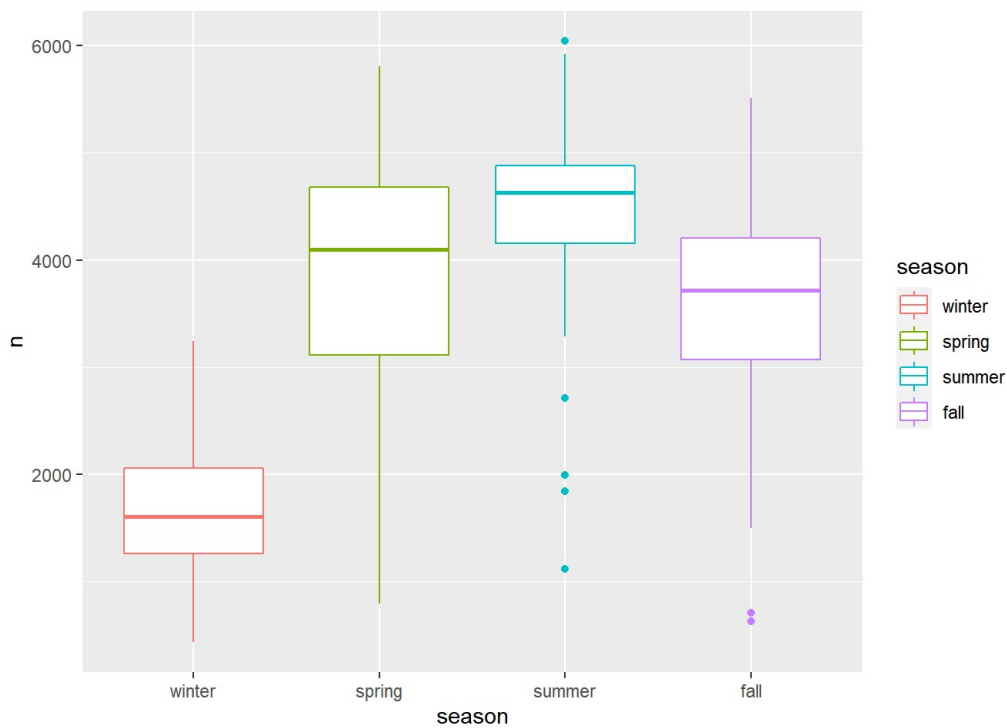
```
#요일별
ggplot(daily, aes(weekday,n))+
  geom_boxplot()
```



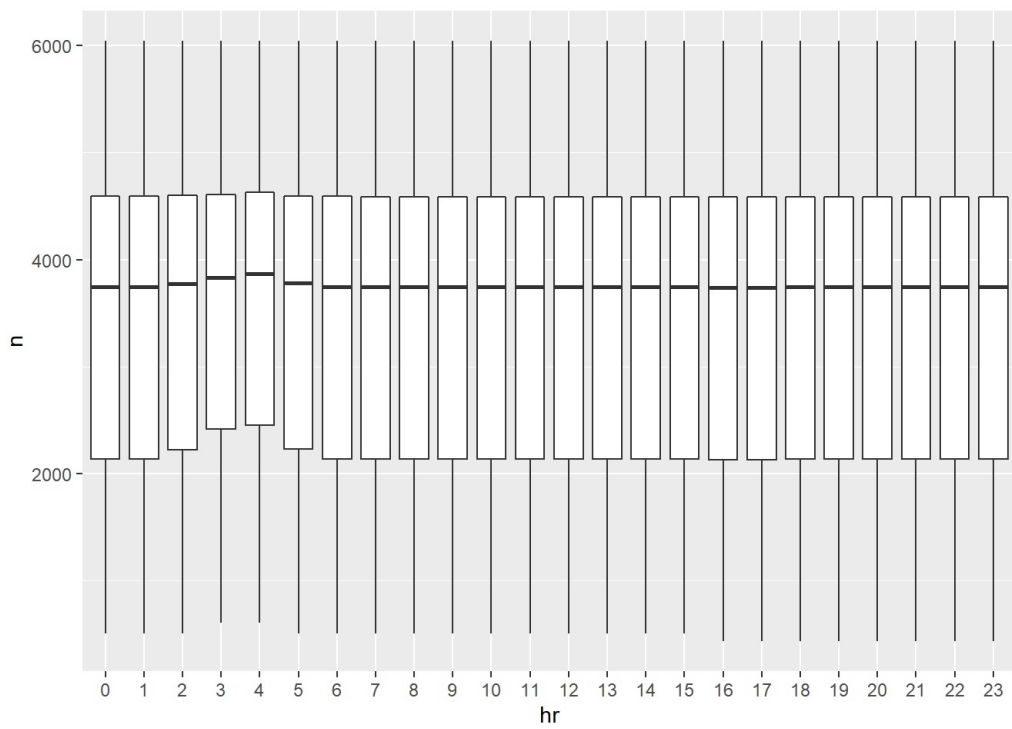
```
ggplot(daily, aes(weekday,n, color=season))+
  geom_boxplot()
```



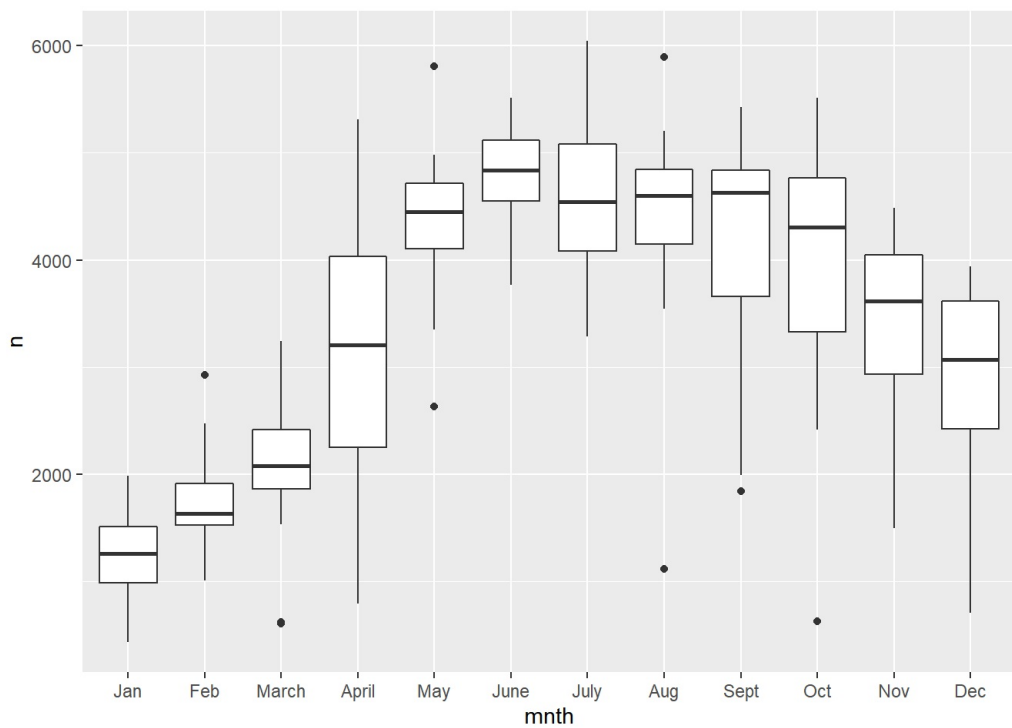
```
#계절별
ggplot(daily, aes(season,n, color=season))+
  geom_boxplot()
```



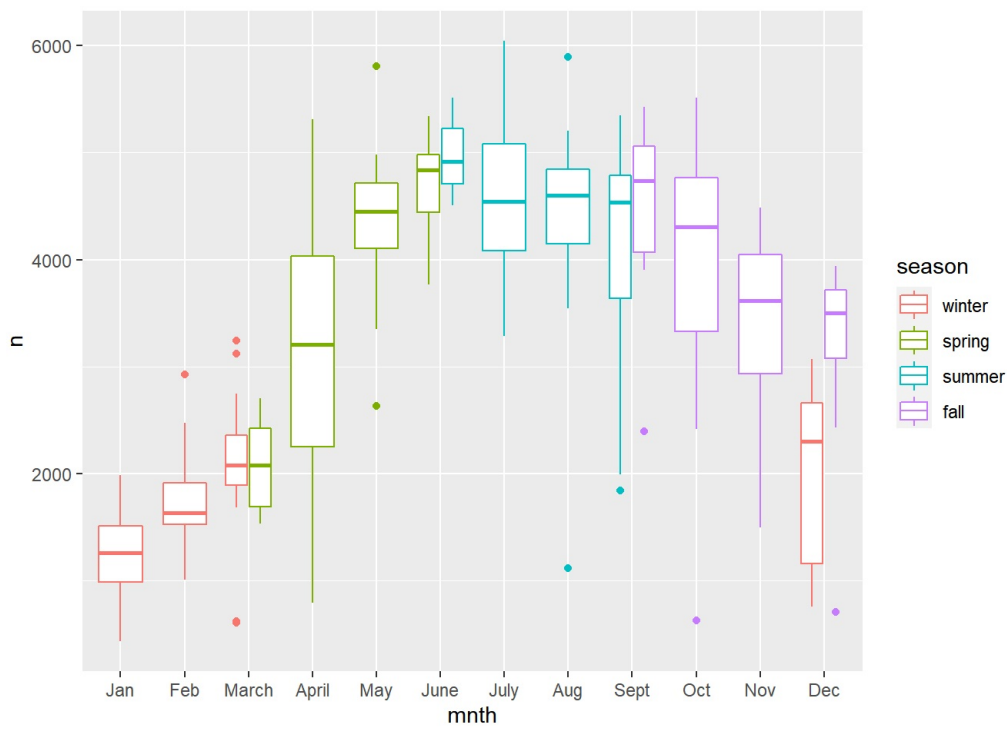
```
#시간별 box
ggplot(daily, aes(hr,n))+
  geom_boxplot()
```



```
#월별 box
ggplot(daily, aes(mnth,n))+
  geom_boxplot()
```



```
ggplot(daily, aes(mnth,n, color=season))+
  geom_boxplot()
```

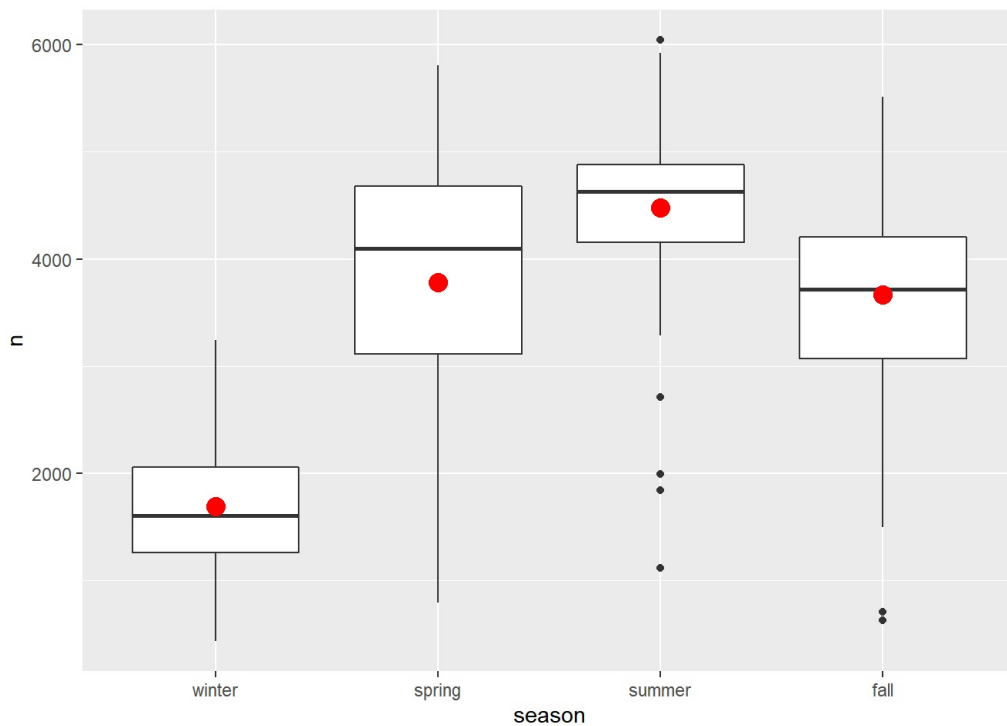


#####계절의 효과가 매우 크다는 것을 알 수 있다

```
#계절의 효과를 제거
mod=lm(n~season, data=daily)

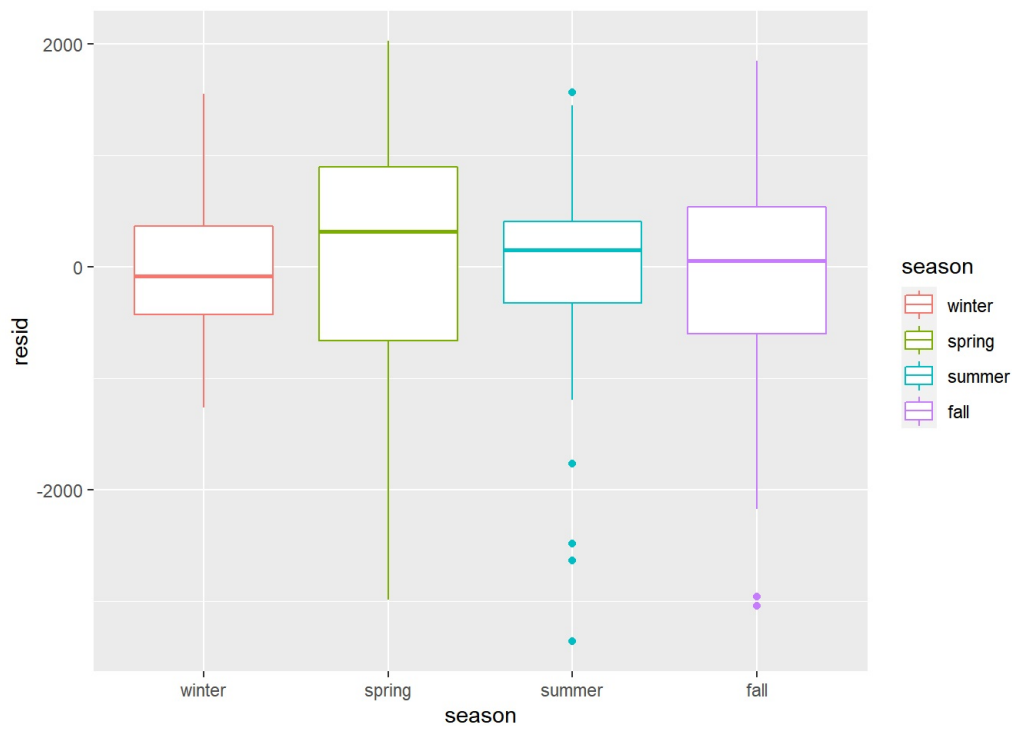
grid <- daily %>%
  data_grid(season) %>%
  add_predictions(mod, "n")

ggplot(daily, aes(season, n)) +
  geom_boxplot() +
  geom_point(data = grid, colour = "red", size = 4)
```

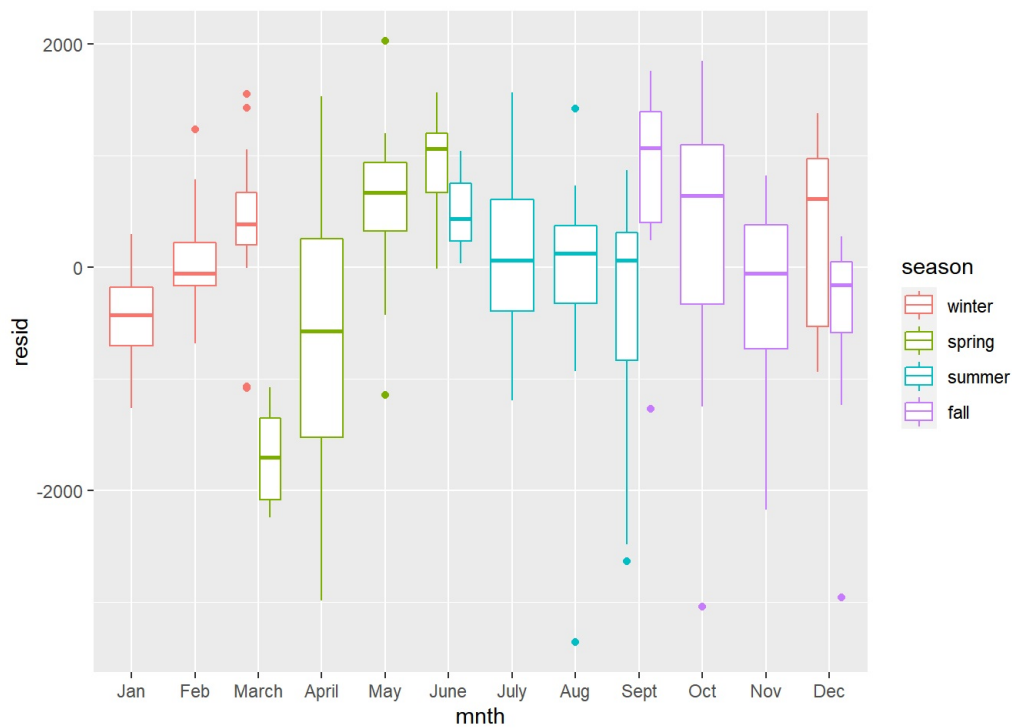


```
daily=daily %>%
  add_residuals(mod)

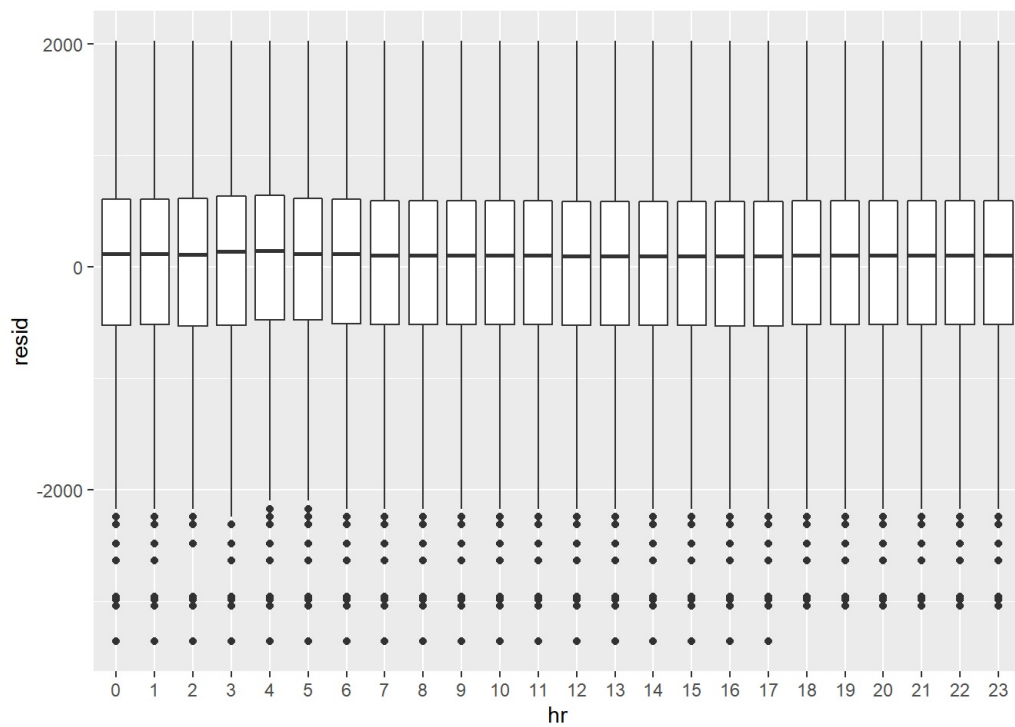
daily %>%
  ggplot()+
  geom_boxplot(aes(season, resid, color=season))
```



```
daily %>%
  ggplot()+
  geom_boxplot(aes(mnth, resid, color=season))
```



```
daily %>%
  ggplot()+
  geom_boxplot(aes(hr, resid))
```

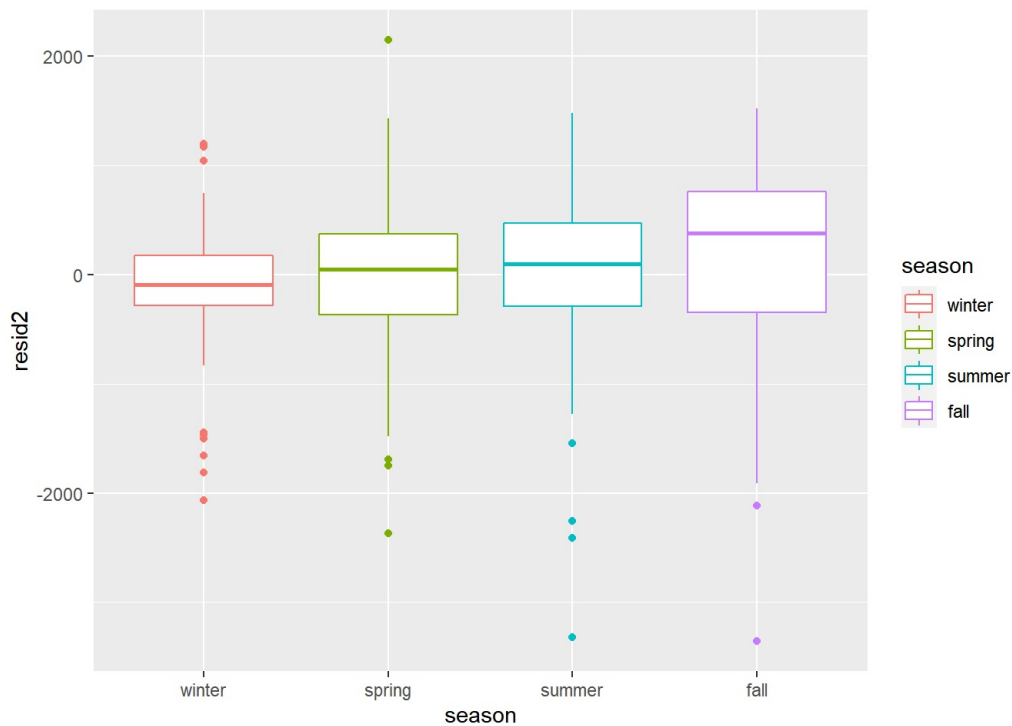


#계절의 효과가 매우 유의하다는 것을 알 수 있다

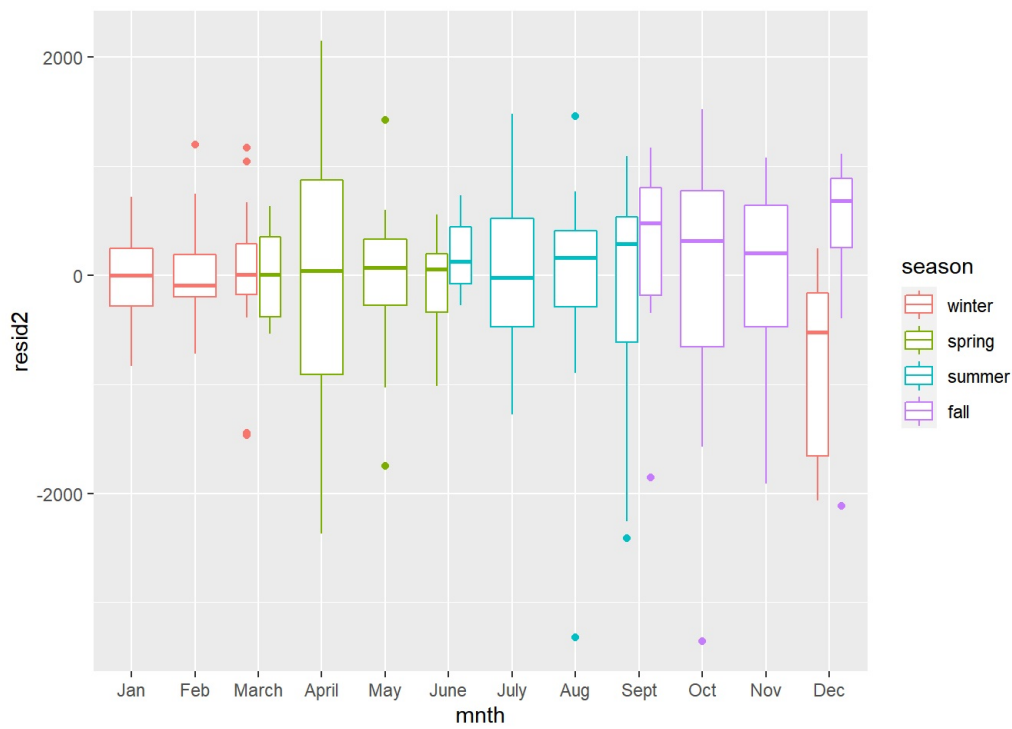
#월별에서 약간 차이가 난다는 것을 확인, 월별 효과 제거하고 진행
 mod2=lm(n~mnth, data=daily)

```
daily=daily %>%
  add_residuals(mod2, 'resid2')
```

```
daily %>%
  ggplot()+
  geom_boxplot(aes(season, resid2, color=season))
```



```
daily %>%
  ggplot()+
  geom_boxplot(aes(mnth, resid2, color=season))
```

```
daily %>%
  ggplot()+
  geom_boxplot(aes(hr, resid2))
```

