Call Volume Data Exploration

```
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
```

```
## The following object is masked from 'package:base':
##
##     date
```

```
library(tidyverse)
```

```
## ── Attaching packages ─────── tidyverse 1.2.1 ──
```

```
## ✓ ggplot2 3.1.0     ✓ purrr   0.2.5
## ✓ tibble  1.4.2     ✓ dplyr   0.7.7
## ✓ tidyr   0.8.2     ✓ stringr 1.3.1
## ✓ readr   1.1.1     ✓ forcats 0.3.0
```

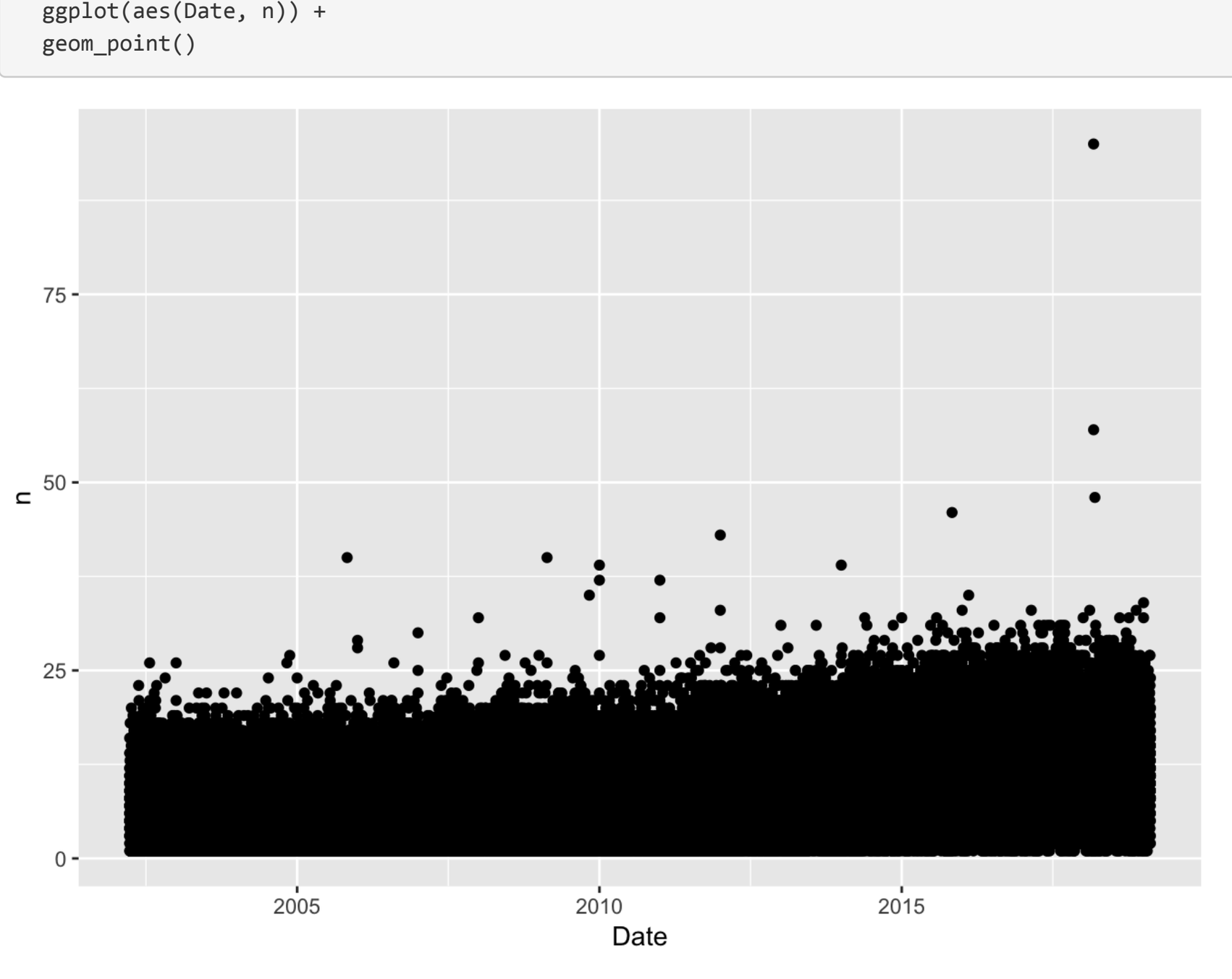
```
## ── Conflicts ────────── tidyverse_conflicts() ──
## ✗ lubridate::as.difftime() masks base::as.difftime()
## ✗ lubridate::date()        masks base::date()
## ✗ dplyr::filter()          masks stats::filter()
## ✗ lubridate::intersect()   masks base::intersect()
## ✗ dplyr::lag()             masks stats::lag()
## ✗ lubridate::setdiff()     masks base::setdiff()
## ✗ lubridate::union()       masks base::union()
```

```
vol <- read_csv('hourly_incidents_assigned_volume_feb_11.csv') %>%
  filter(year(Date) > 2000) #Since there were a couple of erroneous values in the year 1900
```

```
## Parsed with column specification:
## cols(
##   Date = col_datetime(format = ""),
##   n = col_integer()
## )
```

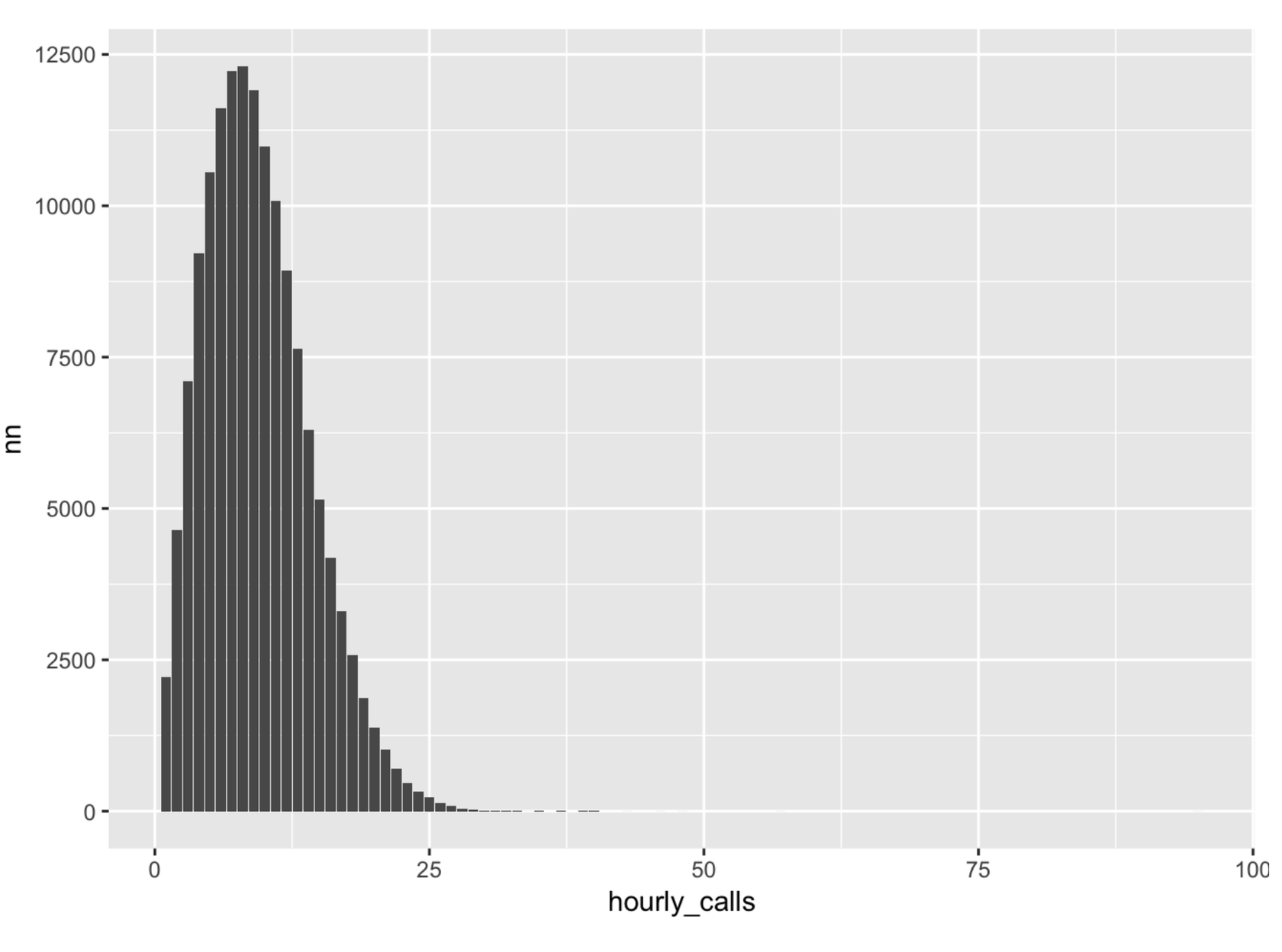
Purely by date

```
vol %>%
  ggplot(aes(Date, n)) +
  geom_point()
```



What's the distribution of volumes?

```
vol %>%
  count(hourly_calls = n) %>%
  ggplot(aes(hourly_calls, nn)) +
  geom_col()
```



Looks like the CAD export didn't include any 0-call hours. I'll have to re-export to make sure we get those.

Or just interpolate

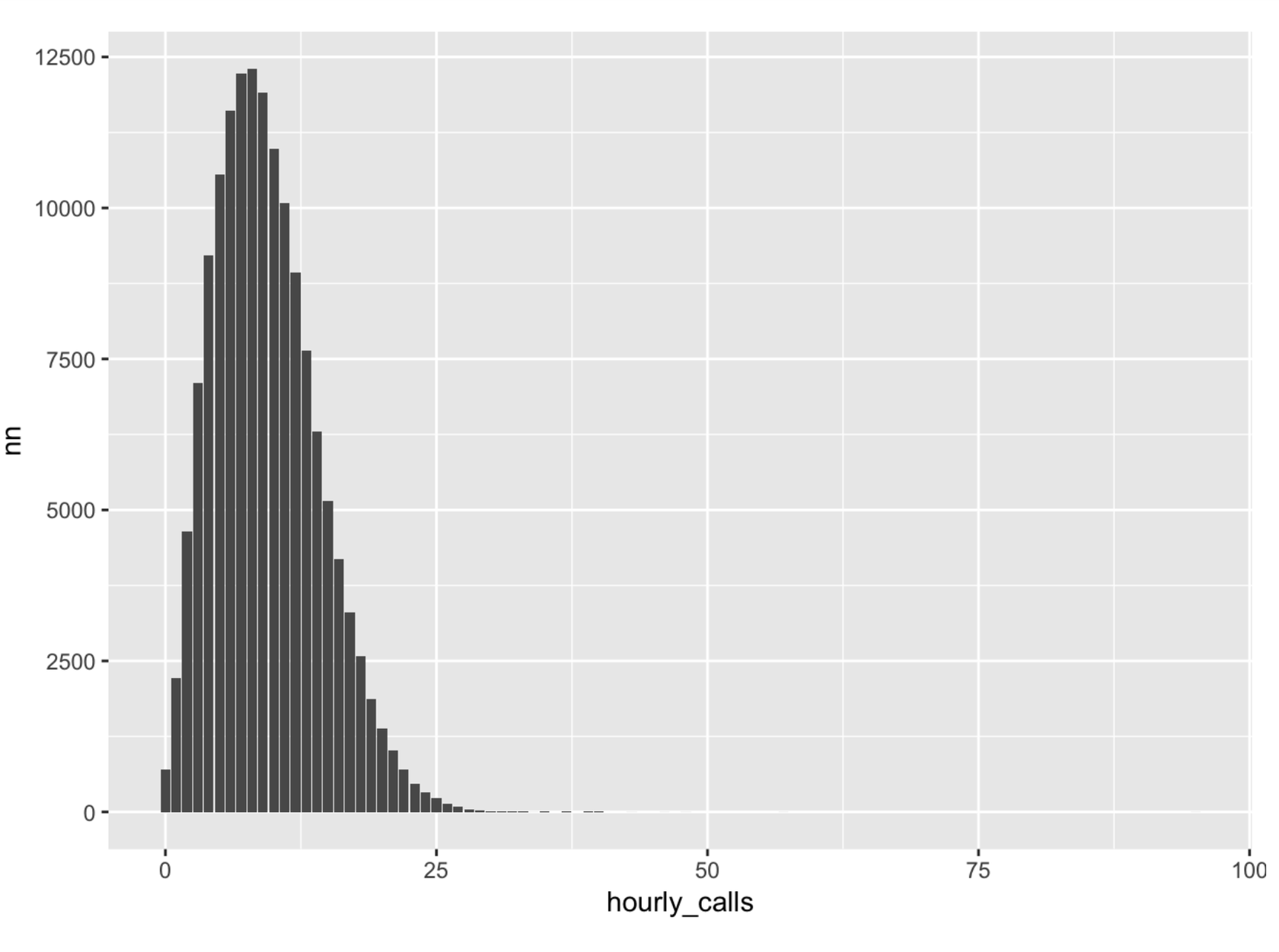
```
all_hours <- tibble(Date = seq.POSIXt(from = min(vol$Date),
                                      to = max(vol$Date),
                                      by = 'hour')
                    ) #Makes a vector of all date/hours in the dataset

vol <- vol %>%
  full_join(all_hours) %>% #Joins with existing dataset
  replace_na(replace = list(n = 0))#A result of the full join is NA where there isn't a match, so this replaces NA with 0
```

```
## Joining, by = "Date"
```

Look again for hourly call distributions

```
vol %>%
  count(hourly_calls = n) %>%
  ggplot(aes(hourly_calls, nn)) +
  geom_col()
```



Poisson distribution?

What's with the super busy hours?

```
vol %>%
  filter(n > 40)
```

```
## # A tibble: 5 x 2
##   Date                    n
##   <dttm>              <dbl>
## 1 2012-01-01 01:00:00    43
## 2 2015-11-01 01:00:00    46
## 3 2018-03-05 00:00:00    95
## 4 2018-03-05 05:00:00    57
## 5 2018-03-13 00:00:00    48
```

2012-01-01 01:00 I suspect this may be legit (new years' eve)

2015-11-01 01:00 is a suprise

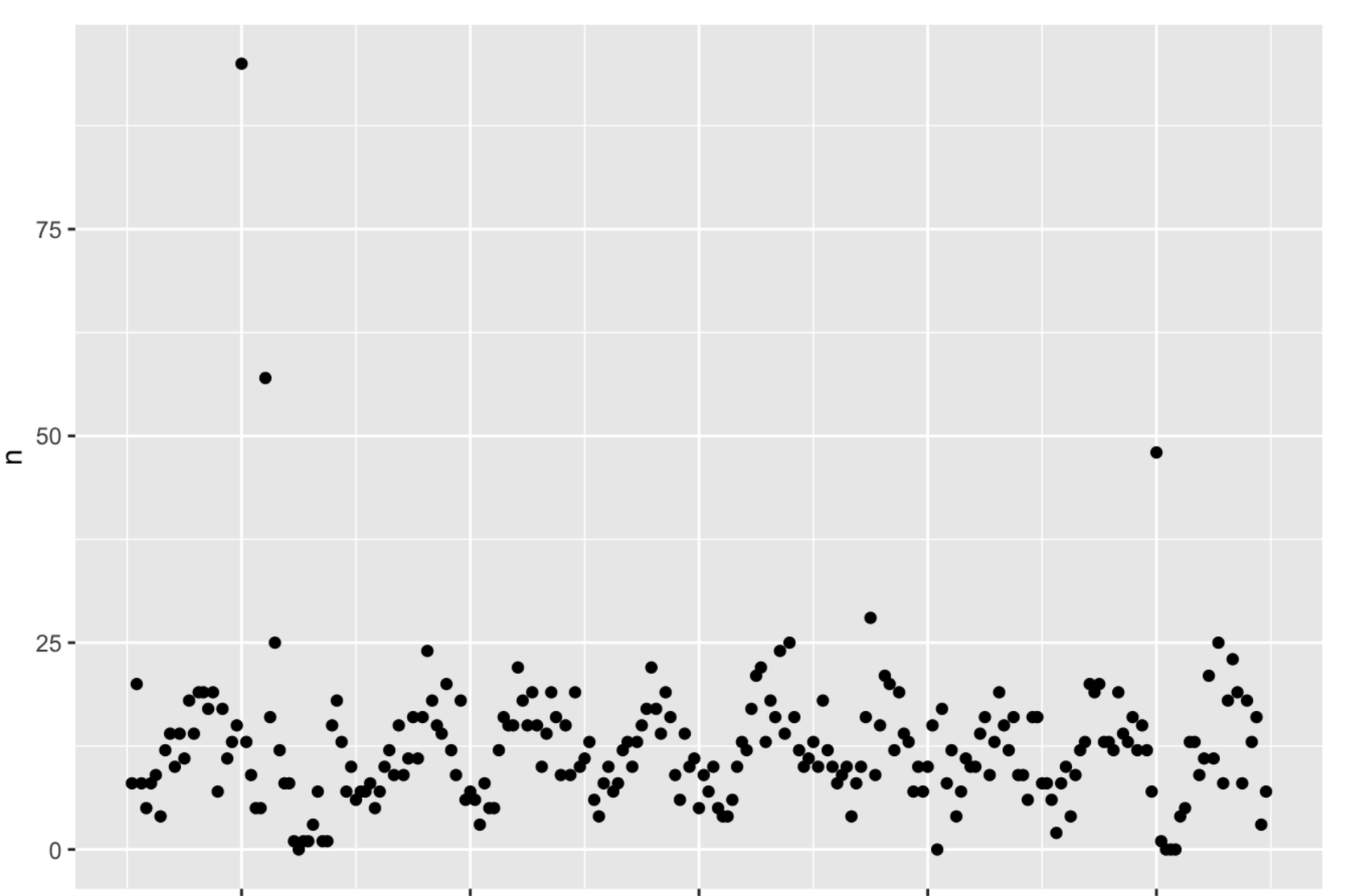
2018-03-05 00:00 Near a CAD downtime, likely catching up from previous hours.

2018-03-05 05:00 Near a CAD downtime, likely catching up from previous hours.

2018-03-13 00:00 Near a CAD downtime, likely catching up from previous hours.

For the last three or four we may have to figure out a way to deal with these.

```
vol %>%
  filter(Date > ymd('2018-03-04')) %>%
  filter(Date < ymd('2018-03-14')) %>%
  ggplot(aes(Date, n)) +
  geom_point()
```



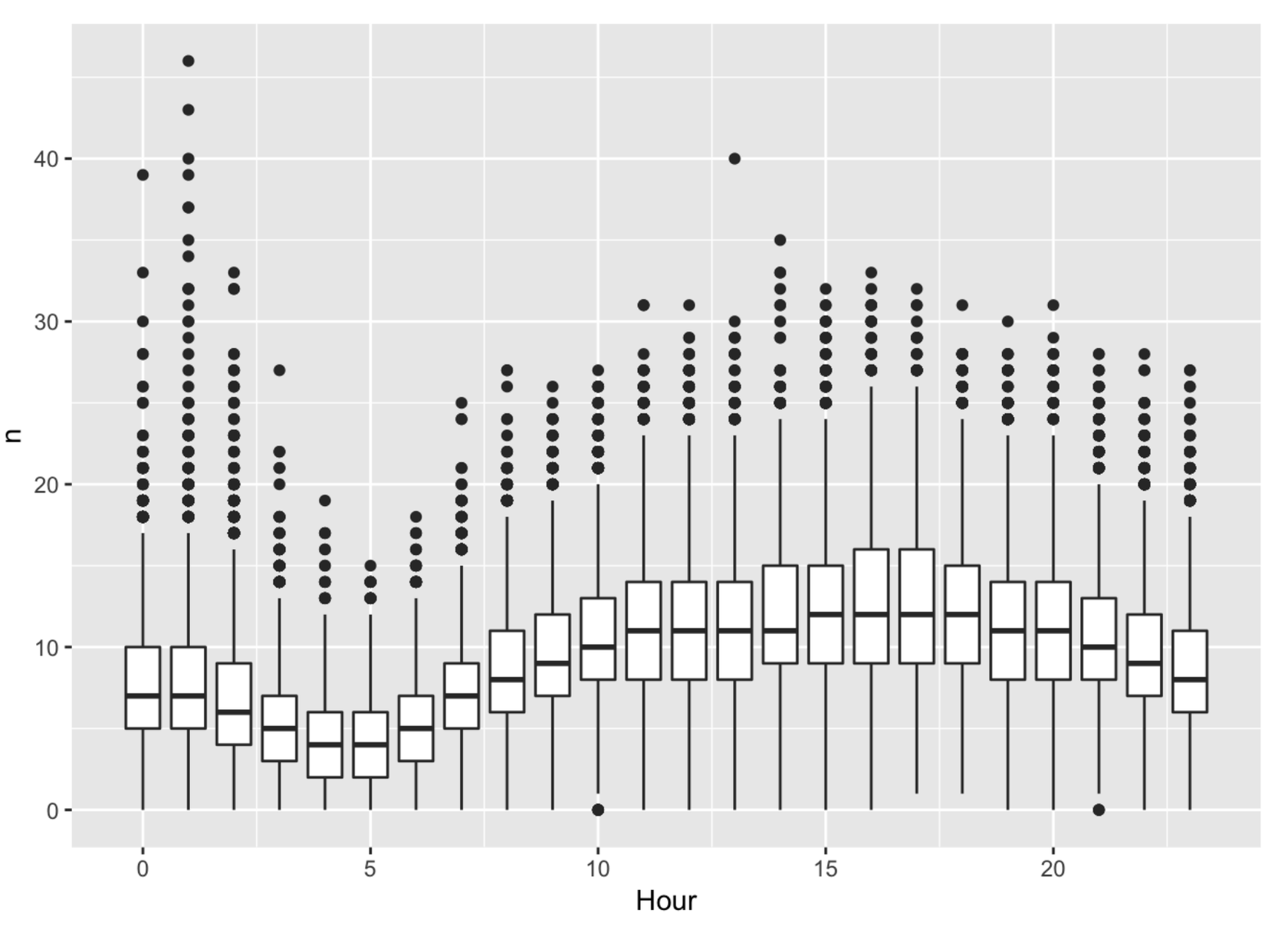
For now I'll just ignore those March 2018 outliers

```
vol <- vol %>%
  filter(n < 48)


vol <- vol %>%
  mutate(Hour = hour(Date)) %>%
  mutate(Month = month(Date)) %>%
  mutate(yday = yday(Date)) %>%
  mutate(wday = wday(Date))
```

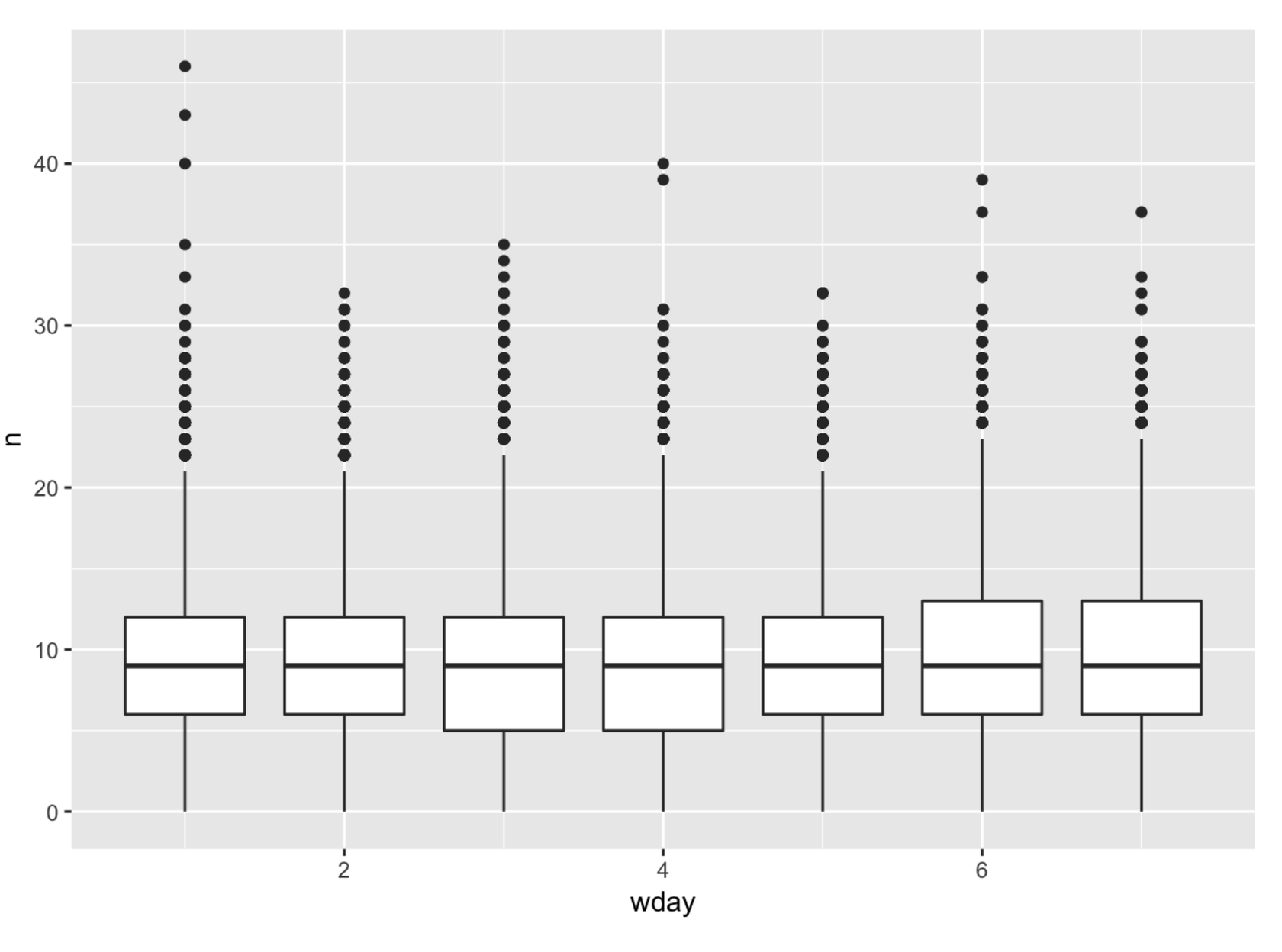
By hour of day

```
vol %>%
  ggplot(aes(Hour, n, group = Hour)) +
  geom_boxplot()
```



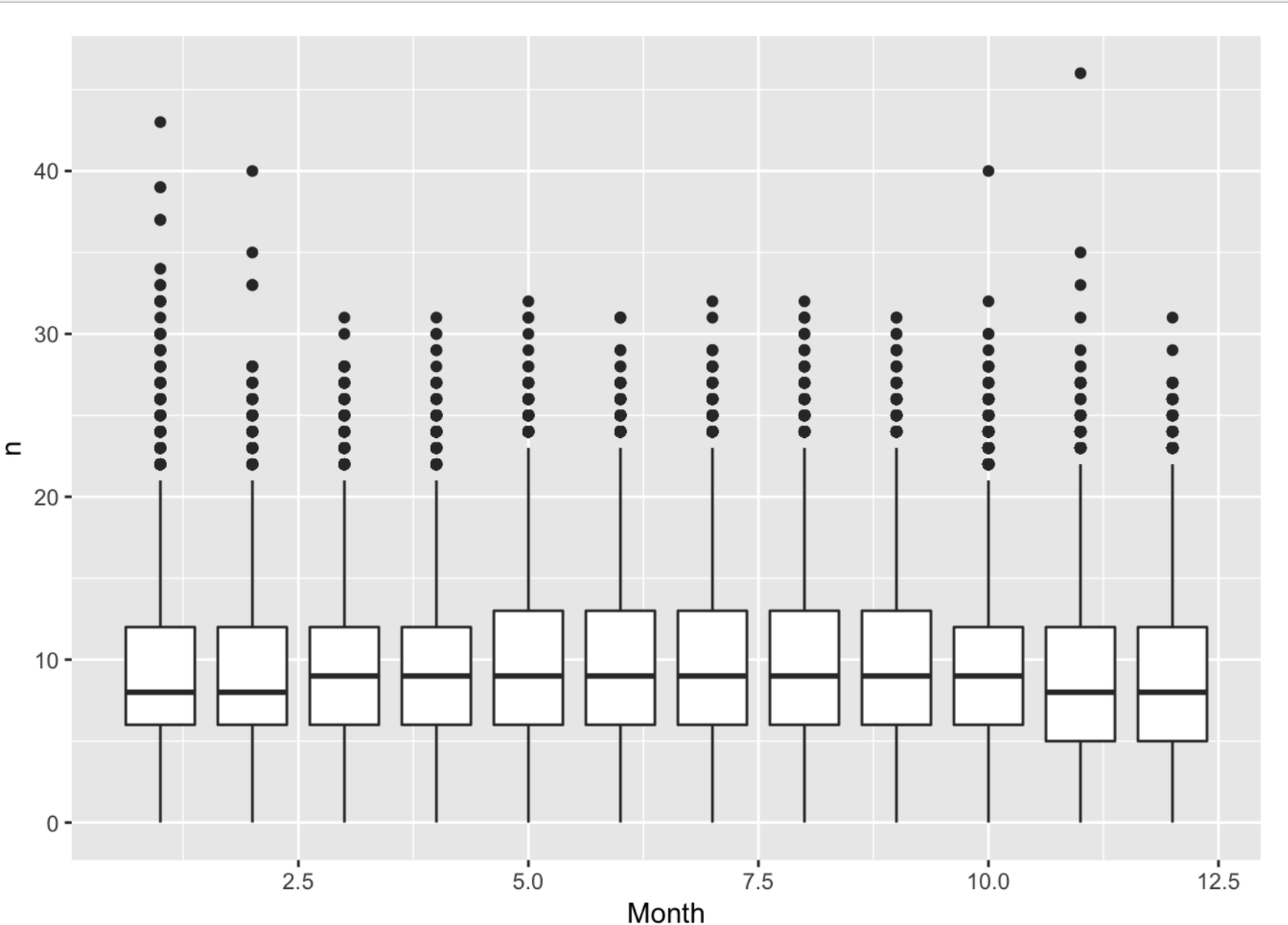
By day of week

```
vol %>%
  ggplot(aes(wday, n, group = wday)) +
  geom_boxplot()
```



By month

```
vol %>%
  ggplot(aes(Month, n, group = Month)) +
  geom_boxplot()
```



Temps

```
temps <- read_csv('Project/Weather/houry_DEN_weather.csv')
```

```
## Parsed with column specification:
## cols(
##   DATE = col_datetime(format = ""),
##   Temp = col_double(),
##   Code = col_integer()
## )
```

```
temps_rounded <- temps %>%
  group_by(Date = round_date(DATE, unit = 'hour')) %>%
  summarise(Temp = mean(Temp, na.rm = T))
```

```
vol %>%
  left_join(temps_rounded) %>%
  ggplot(aes(Temp, n)) +
  geom_point() +
  geom_smooth(method = 'lm') +
  facet_grid(Hour ~ wday)
```

```
## Joining, by = "Date"
```

```
## Warning: Removed 501 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 501 rows containing missing values (geom_point).
```