

# Course 2 Section 2.8 - Your turn

Jiaying Wu

05/10/2020

## Step 1: Preparing the data

```
# Load tidyverse
library(tidyverse)

# Load rwalkr
library(rwalkr)

# Read in pedestrian data from 2016 to 2018 for 3 locations
ped_melbcnt <- melb_walk_fast(year = 2016:2018, sensor = "Melbourne Central")
invisible(gc())
ped_southbank <- melb_walk_fast(year = 2016:2018, sensor = "Southbank")
ped_bourkest.mallnorth <- melb_walk_fast(year = 2016:2018, sensor = "Bourke Street Mall (North)")

# Row bind 3 data frames
ped_melb.south.bourke <- bind_rows(ped_melbcnt, ped_southbank, ped_bourkest.mallnorth)

# Look at the data
glimpse(ped_melb.south.bourke)

## Rows: 78,912
## Columns: 5
## $ Sensor    <chr> "Melbourne Central", "Melbourne Central", "Melbourne Cent...
## $ Date_Time <dttm> 2016-01-01 00:00:00, 2016-01-01 01:00:00, 2016-01-01 02:...
## $ Date      <date> 2016-01-01, 2016-01-01, 2016-01-01, 2016-01-01, 2016-01-...
## $ Time      <int> 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, ...
## $ Count     <int> 2829, 2905, 1733, 1214, 357, 179, 160, 166, 190, 398, 601...
```

The variables in the data include:

- Sensor - location of the sensor device
- Date\_Time - date and time of the reading
- Date - date (yyyy-mm-dd)
- Time - hour of the day
- Count - total sensor count of pedestrians

```
# Load lubridate
library(lubridate)

# Create 'time' variables
ped_melb.south.bourke <- ped_melb.south.bourke %>%
  mutate(year = year(Date),
        month = month(Date, label = TRUE, abbr = TRUE),
        wday = wday(Date, label = TRUE, abbr = TRUE, week_start = 1),
        day = day(Date))
```

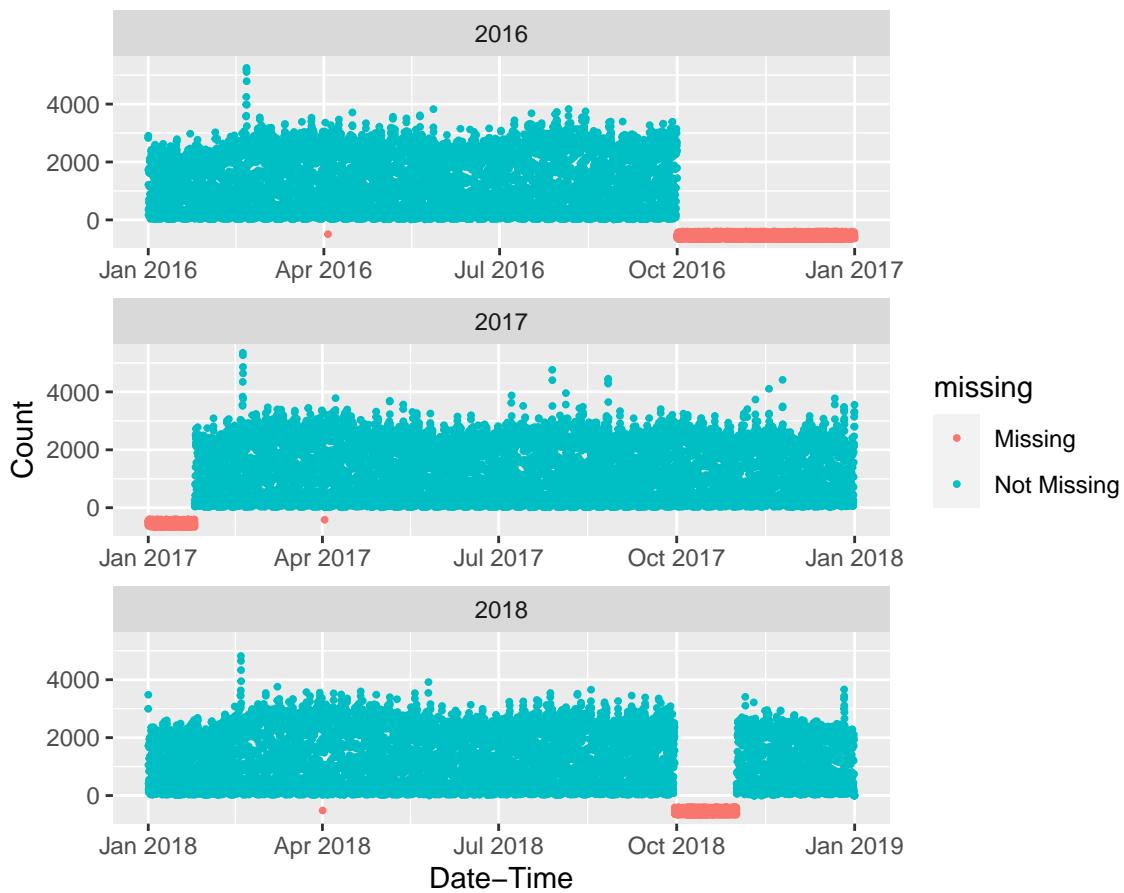
## Step 2: Exploring time gaps

Using what you have learned from Visualising Missing data, you will check for missing values/time gaps in the data.

```
library(visdat)
library(naniar)

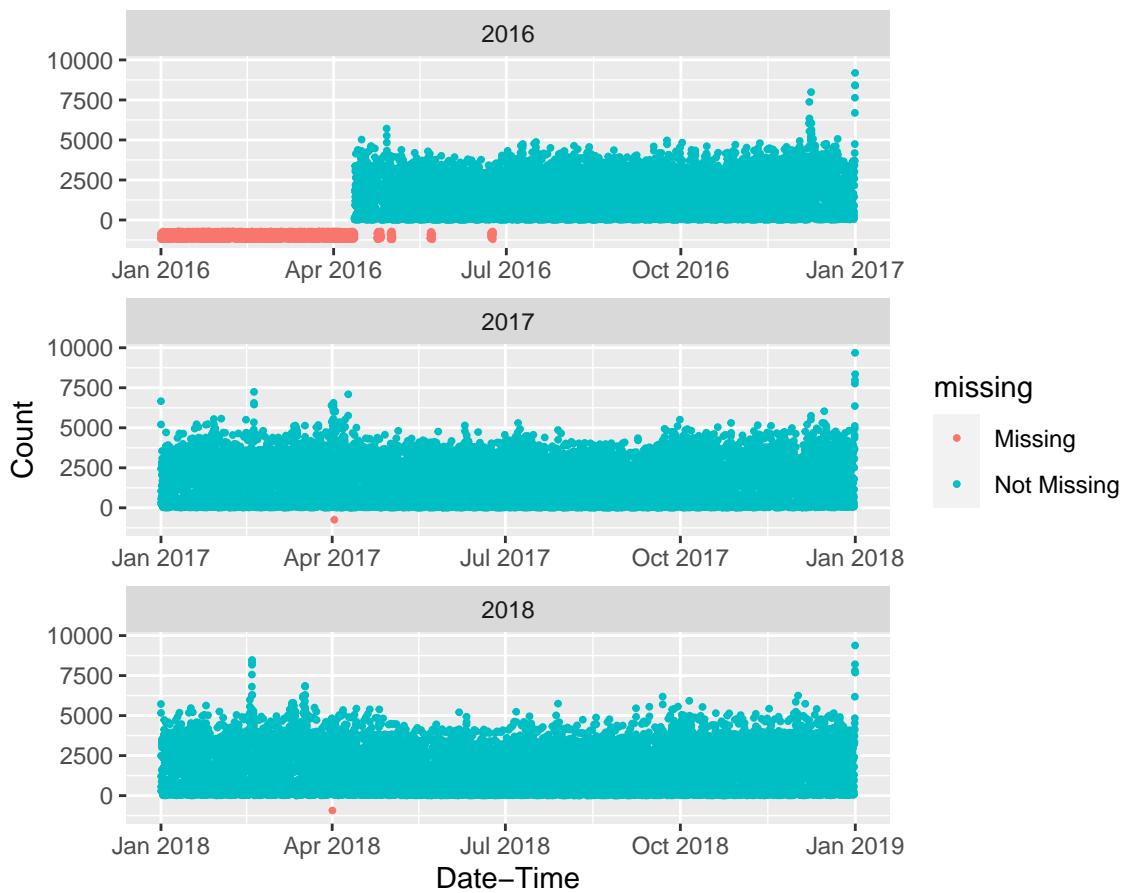
# Melbourne Central time gaps
ped_melb.south.bourke %>%
  filter(Sensor == "Melbourne Central") %>%
  ggplot(aes(x=Date_Time, y=Count)) +
  geom_miss_point(size = 0.7) +
  facet_wrap(year ~., scales = "free_x", nrow = 3) +
  labs(title = "Melbourne Central", y = "Count", x = "Date-Time")
```

## Melbourne Central

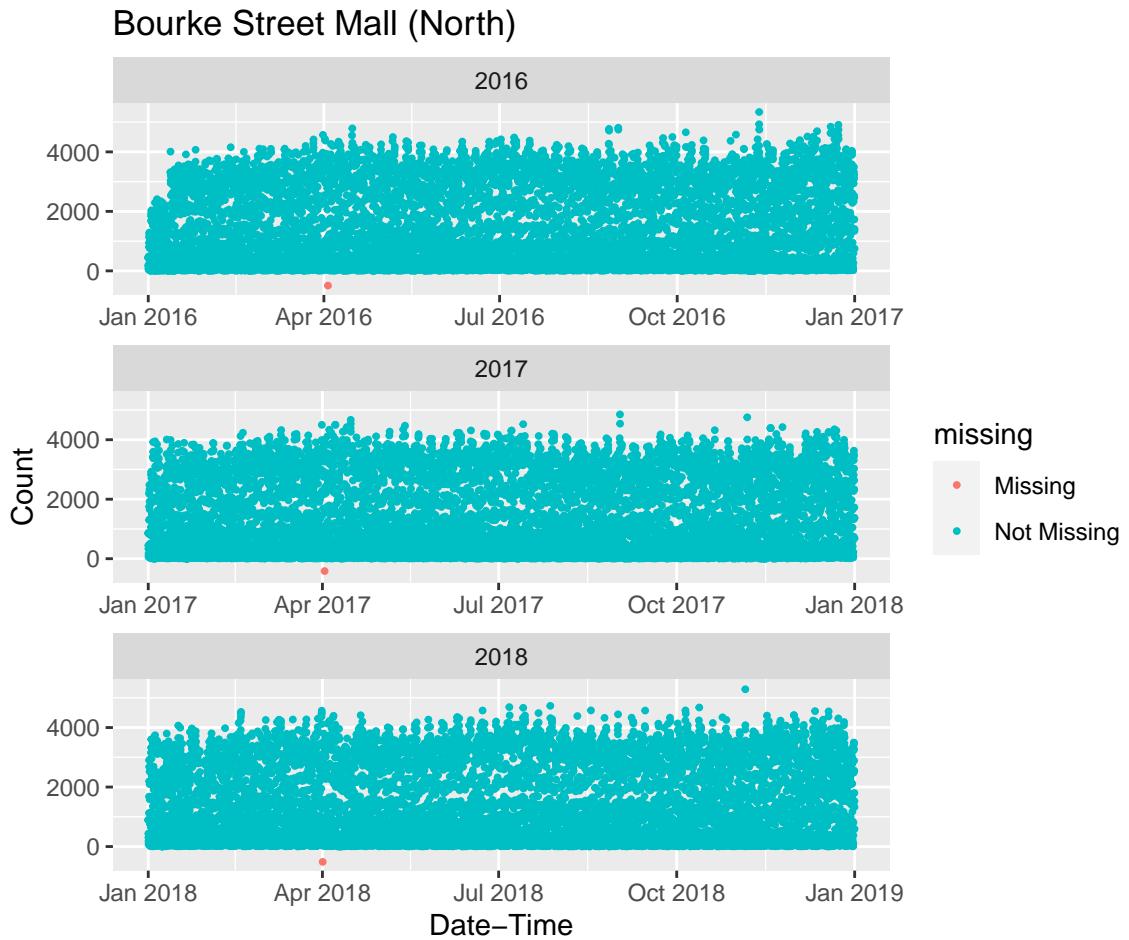


```
# Southbank time gaps
ped_melb.south.bourke %>%
  filter(Sensor == "Southbank") %>%
  ggplot(aes(x=Date_Time, y=Count)) +
  geom_miss_point(size = 0.7) +
  facet_wrap(year ~., scales = "free_x", nrow = 3) +
  labs(title = "Southbank", y = "Count", x = "Date-Time")
```

## Southbank



```
# Bourke Street Mall (North) time gaps
ped_melb.south.bourke %>%
  filter(Sensor == "Bourke Street Mall (North)") %>%
  ggplot(aes(x=Date_Time, y=Count)) +
  geom_miss_point(size = 0.7) +
  facet_wrap(year ~ ., scales = "free_x", nrow = 3) +
  labs(title = "Bourke Street Mall (North)", y = "Count", x = "Date-Time")
```



### Step 3: Visualising the data

return the mean and median hourly pedestrian counts in Melbourne Central, Southbank and Bourke Street Mall (North).

```
# Table of mean and median pedestrian count
ped_melb.south.bourke %>%
  group_by(Sensor) %>%
  summarise(meanCount = mean(Count, na.rm=TRUE),
            medianCount = median(Count, na.rm=TRUE)) %>%
ungroup()
```

| Sensor                     | meanCount | medianCount |
|----------------------------|-----------|-------------|
| Bourke Street Mall (North) | 1357.     | 872         |
| Melbourne Central          | 1294.     | 1141        |
| Southbank                  | 1719.     | 1654        |

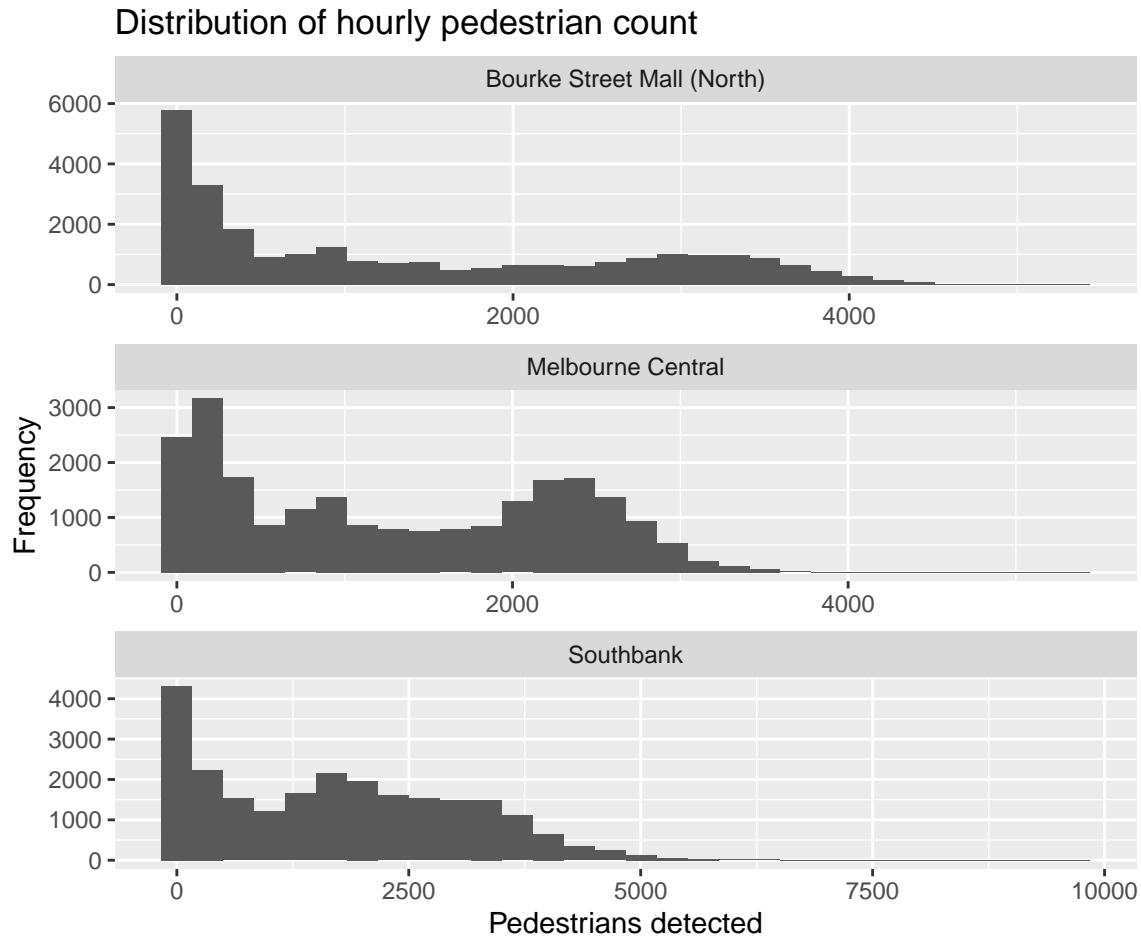
```
# Histogram of pedestrian count
ped_melb.south.bourke %>%
  ggplot(aes(x = Count)) +
  geom_histogram() +
```

```

  labs(title = "Distribution of hourly pedestrian count",
       x = "Pedestrians detected",
       y = "Frequency") +
  facet_wrap(~ Sensor, scales = "free", nrow = 3)

## Warning: Removed 6127 rows containing non-finite values (stat_bin).

```



obtain the median hourly pedestrian counts by location and month.

```

# Median hourly pedestrian count for each month and location
ped_melb.south.bourke %>%
  group_by(Sensor, month) %>%
  summarise(medianCount = median(Count, na.rm=TRUE)) %>%
  ungroup() %>%
  spread(Sensor, medianCount)

```

```

## # A tibble: 12 x 4
##   month `Bourke Street Mall (North)` `Melbourne Central` `Southbank`
##   <ord>                <dbl>            <dbl>           <dbl>
## 1 Jan                  730.             1017            1931
## 2 Feb                  884.             1152            1889

```

```

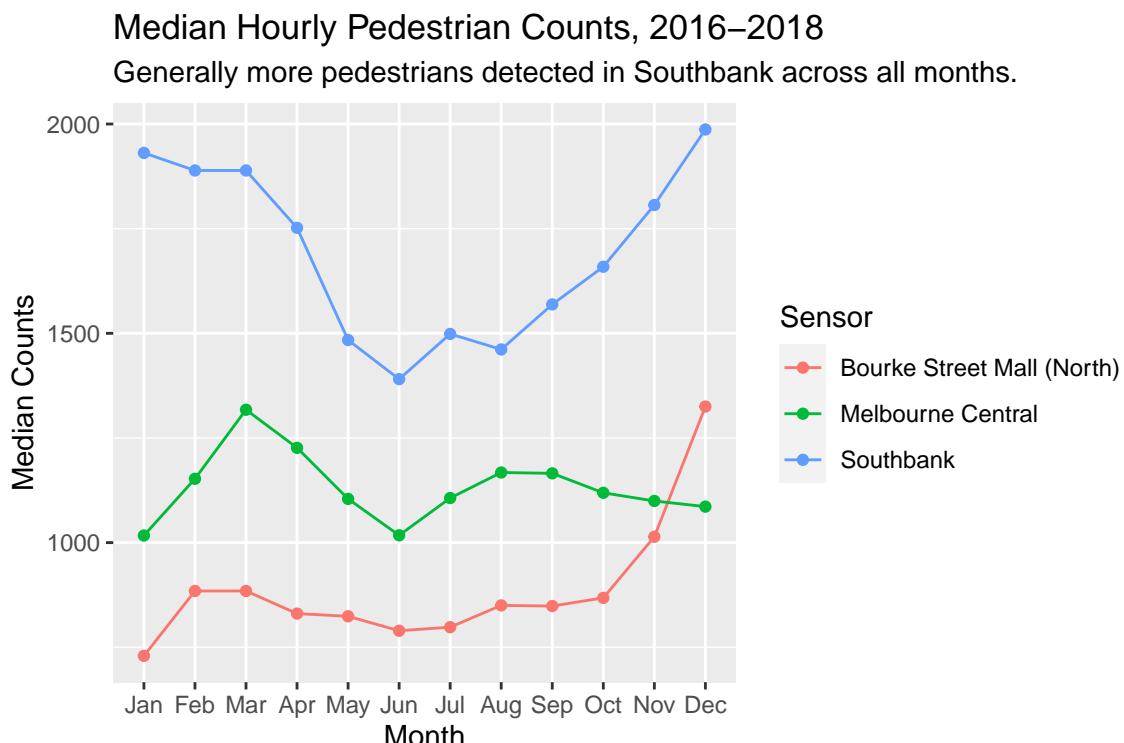
## 3 Mar 884. 1318. 1889
## 4 Apr 830. 1226. 1752
## 5 May 824. 1104. 1484
## 6 Jun 790. 1018. 1390.
## 7 Jul 798. 1106. 1498.
## 8 Aug 850. 1168. 1462.
## 9 Sep 848. 1166. 1569
## 10 Oct 868. 1119. 1659
## 11 Nov 1014. 1100. 1806.
## 12 Dec 1325. 1086. 1987

```

```

# Challenge: Line plots median hourly pedestrian count
ped_melb.south.bourke %>%
  group_by(Sensor, month) %>%
  summarise(medianCount = median(Count, na.rm=TRUE)) %>%
  ungroup() %>%
  ggplot(aes(x= month, y = medianCount, color = Sensor, group = Sensor)) +
  geom_line() +
  geom_point() +
  labs(title = "Median Hourly Pedestrian Counts, 2016–2018",
       subtitle = "Generally more pedestrians detected in Southbank across all months.",
       x = "Month",
       y = "Median Counts")

```



### Box plots of pedestrian counts

```

# Box plot of pedestrian counts
ped_melb.south.bourke %>%

```

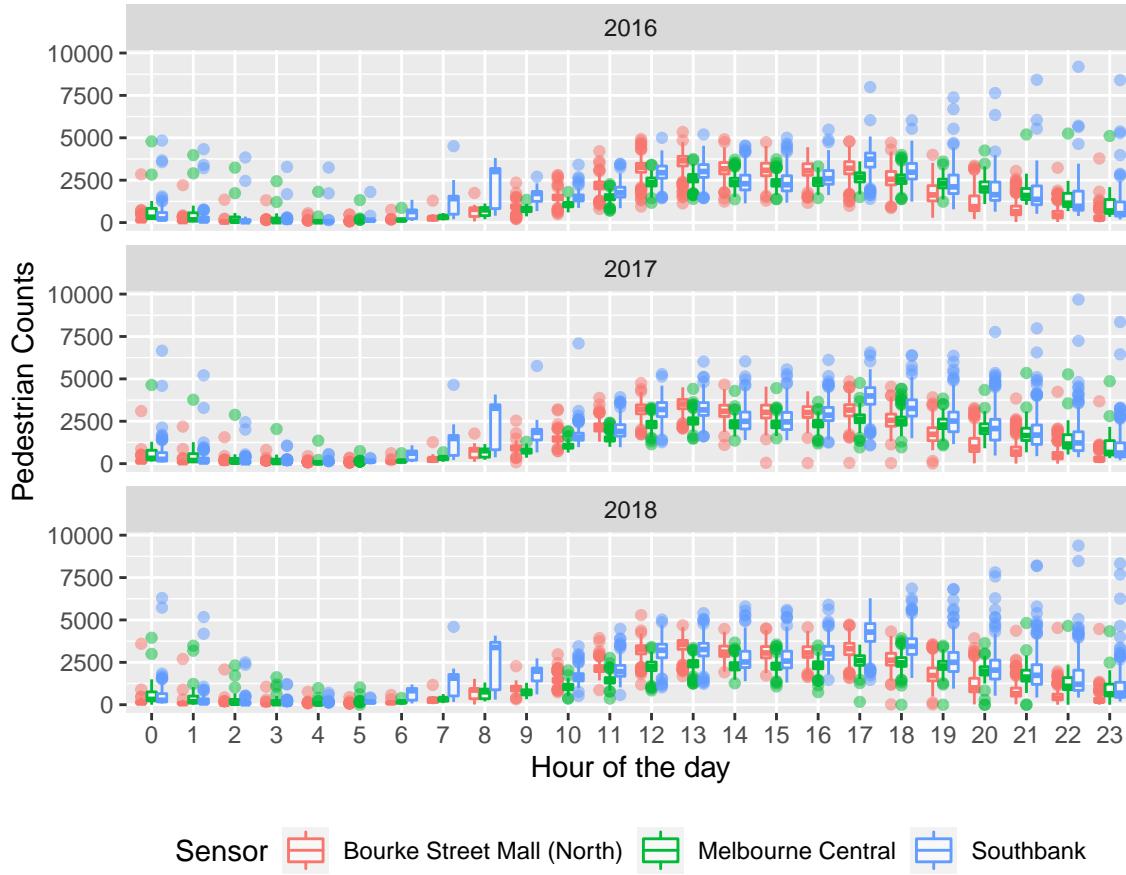
```

ggplot(aes(x = as.factor(Time), y = Count, colour = Sensor)) +
  geom_boxplot(alpha = 0.5) +
  facet_wrap(~ year, nrow = 3) +
  theme(legend.position = "bottom") + # change the legend position
  labs(title = "Distribution of pedestrian counts at each hour of the day", y = "Pedestrian Counts", x =

```

## Warning: Removed 6127 rows containing non-finite values (stat\_boxplot).

Distribution of pedestrian counts at each hour of the day



Pedestrian count prior to NYE fireworks

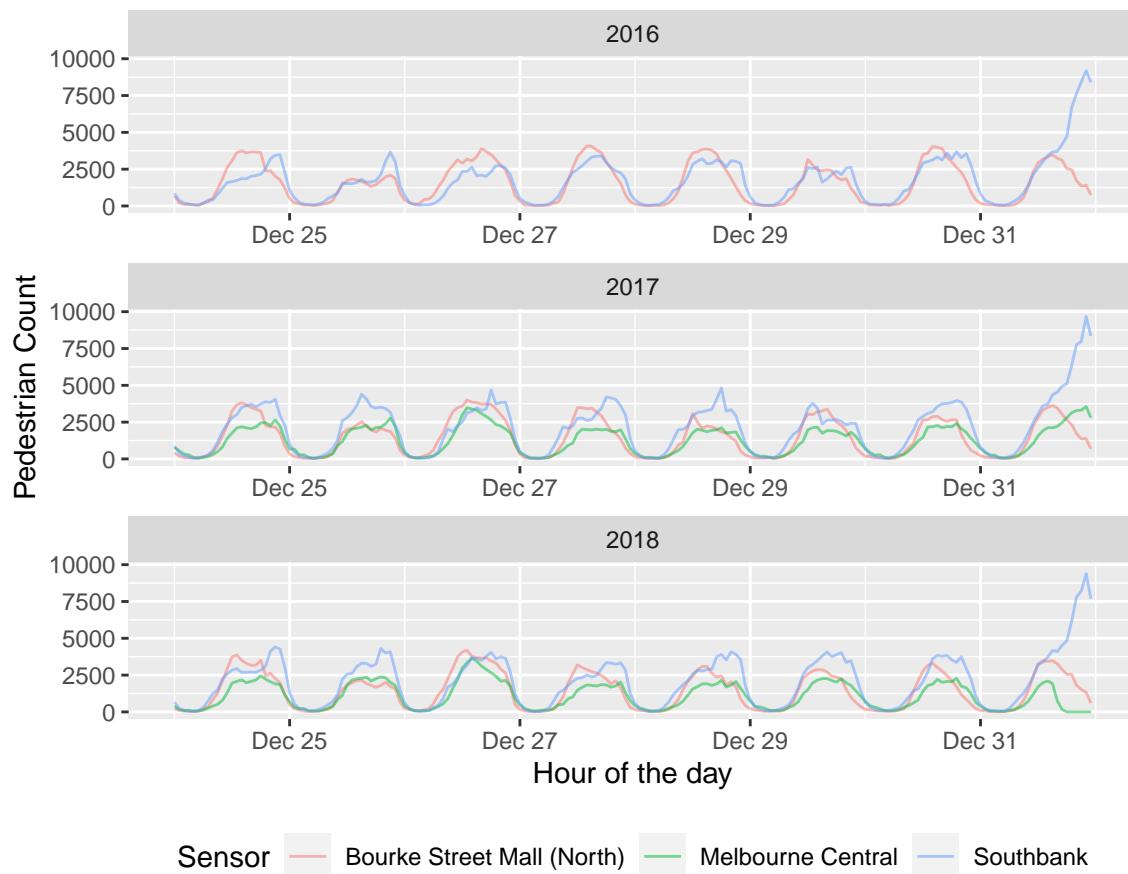
```

# Fill out ???
ped_melb.south.bourke %>%
  filter(month == "Dec", day %in% 24:31) %>%
  ggplot(aes(x = Date_Time, y = Count, colour = Sensor)) +
  geom_line(alpha = 0.5) +
  facet_wrap(year ~., scales = "free_x", nrow = 3) +
  theme(legend.position = "bottom") + # change the legend position
  labs(title = "Pedestrian count at each hour of the day leading up to NYE", y = "Pedestrian Count", x =

```

## Warning: Removed 192 row(s) containing missing values (geom\_path).

## Pedestrian count at each hour of the day leading up to NYE



Sensor Bourke Street Mall (North) Melbourne Central Southbank