Course 1 Section 2.24 - Pedestrian activity around the City of Melbourne 2

Jiaying Wu

15/09/2020

```
library(tidyverse)
library(here)
```

Step 1: Create a new R Markdown and read in the data

```
ped <- read_csv(here("data", "melb_walk.csv"))
ped
```

```
## # A tibble: 31,992 x 5
##    Sensor                        Date_Time           Date        Time Count
##    <chr>                         <dttm>              <date>     <dbl> <dbl>
##  1 Bourke Street Mall (North)    2018-12-31 13:00:00 2019-01-01     0   918
##  2 Bourke Street Mall (South)    2018-12-31 13:00:00 2019-01-01     0   770
##  3 Melbourne Central             2018-12-31 13:00:00 2019-01-01     0    NA
##  4 Town Hall (West)              2018-12-31 13:00:00 2019-01-01     0  3025
##  5 Princes Bridge                2018-12-31 13:00:00 2019-01-01     0   531
##  6 Flinders Street Station Underpass 2018-12-31 13:00:00 2019-01-01     0  3284
##  7 Birrarung Marr                2018-12-31 13:00:00 2019-01-01     0  2733
##  8 Webb Bridge                   2018-12-31 13:00:00 2019-01-01     0   762
##  9 Southern Cross Station        2018-12-31 13:00:00 2019-01-01     0  1830
## 10 Victoria Point                2018-12-31 13:00:00 2019-01-01     0  1217
## # ... with 31,982 more rows
```

Step 2: Count the sensors, and more

Q1: Use a wrangling verb, to count the number of sensors. Do all the sensors have the same number of measurements?

```
num_sensor <- ped %>%
  group_by(Sensor) %>%
  summarise(num_sensor = n())

num_sensor
```

```
## # A tibble: 43 x 2
##    Sensor                         num_sensor
##    <chr>                              <int>
##  1 Alfred Place                         744
##  2 Birrarung Marr                       744
##  3 Bourke St-Russell St (West)          744
##  4 Bourke Street Mall (North)           744
##  5 Bourke Street Mall (South)           744
##  6 Chinatown-Lt Bourke St (South)       744
##  7 Chinatown-Swanston St (North)        744
##  8 City Square                          744
##  9 Collins Place (North)                744
## 10 Collins Place (South)                744
## # ... with 33 more rows
```

```
unique(num_sensor$num_sensor)
```

```
## [1] 744
```

All the sensors have the same number of measurements.

Q2: For each sensor, compute the total count for January. Which sensor had the largest count? Which sensor had the smallest count?

```
count_by_sensor <- ped %>%
  group_by(Sensor) %>%
  summarise(sum = sum(Count, na.rm = TRUE)) %>%
  arrange(sum)
```

```
head(count_by_sensor)
```

```
## # A tibble: 6 x 2
##   Sensor                             sum
##   <chr>                            <dbl>
## 1 City Square                          0
## 2 Flagstaff Station                    0
## 3 Flinders St-Elizabeth St (East)      0
## 4 Tin Alley-Swanston St (West)     38773
## 5 Waterfront City                  61481
## 6 Monash Rd-Swanston St (West)     66420
```

```
tail(count_by_sensor)
```

```
## # A tibble: 6 x 2
##   Sensor                                sum
##   <chr>                               <dbl>
## 1 The Arts Centre                    884885
## 2 Bourke Street Mall (North)         895483
## 3 Spencer St-Collins St (North)      910109
## 4 Flinders Street Station Underpass 1015331
## 5 Town Hall (West)                  1035715
## 6 Southbank                         1395117
```

Q3: For each sensor, compute the total number of missing counts. Which sensor had the most missing counts? Why might this be?

```
ped %>%
  group_by(Sensor) %>%
  select(Count) %>%
  summarise_all(funs(sum(is.na(.)))) %>%
  rename(na_count = "Count") %>%
  arrange(desc(na_count))
```

```
## Warning: `funs()` is 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 every 8 hours.
## Call `lifecycle::last_warnings()` to see where this warning was generated.
```

```
## # A tibble: 43 x 2
##    Sensor                           na_count
##    <chr>                               <int>
##  1 City Square                           744
##  2 Flagstaff Station                     744
##  3 Flinders St-Elizabeth St (East)       744
##  4 Birrarung Marr                        416
##  5 Melbourne Central                     127
##  6 Monash Rd-Swanston St (West)           50
##  7 Grattan St-Swanston St (West)          38
##  8 Tin Alley-Swanston St (West)           25
##  9 St Kilda Rd-Alexandra Gardens          24
## 10 Waterfront City                        21
## # ... with 33 more rows
```

Q4: Create a new table that contains the counts for the Melbourne Central and State Library sensors, then use a tidying verb to create two new columns that contain their counts.

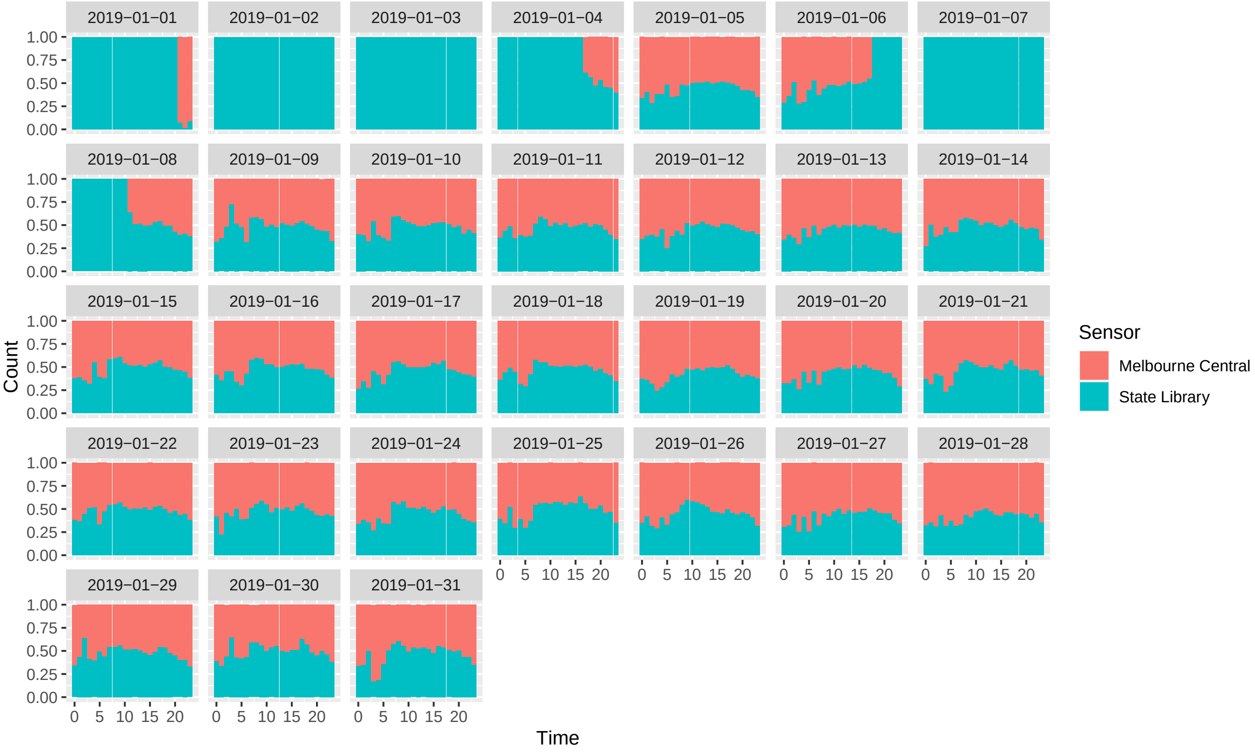
```
ped %>%
  filter(Sensor %in% c("Melbourne Central", "State Library")) %>%
  spread(key = Sensor, value = Count)
```

```
## # A tibble: 744 x 5
##    Date_Time           Date        Time `Melbourne Central` `State Library`
##    <dttm>              <date>     <dbl>               <dbl>           <dbl>
##  1 2018-12-31 13:00:00 2019-01-01     0                  NA            1548
##  2 2018-12-31 14:00:00 2019-01-01     1                  NA            1494
##   3 2018-12-31 15:00:00 2019-01-01     2                  NA              878
##   4 2018-12-31 16:00:00 2019-01-01     3                  NA              309
##   5 2018-12-31 17:00:00 2019-01-01     4                  NA              133
##   6 2018-12-31 18:00:00 2019-01-01     5                  NA              110
##   7 2018-12-31 19:00:00 2019-01-01     6                  NA               42
##   8 2018-12-31 20:00:00 2019-01-01     7                  NA               50
##   9 2018-12-31 21:00:00 2019-01-01     8                  NA               83
##  10 2018-12-31 22:00:00 2019-01-01     9                  NA              128
## # ... with 734 more rows
```

Q5: Create the following 100 percent chart to compare the foot traffic at Melbourne Central and the State Library during different hours of the day. Explain why the first 8 days of January appear this way.

```r
ped %>%
  filter(Sensor %in% c("Melbourne Central", "State Library")) %>%
  ggplot(aes(x = Time , y = Count, fill = Sensor)) +
  geom_bar(stat = "identity", position = "fill") +
  facet_wrap(~Date, ncol = 7)
```

```
## Warning: Removed 127 rows containing missing values (position_stack).
```


Count
Time
Sensor
Melbourne Central
State Library