Problem Set 7

Vivian Yeh

In this problem set, the data comes from lab 4 “flights”.

```r
library(stat20data)
library(tidyverse)
```

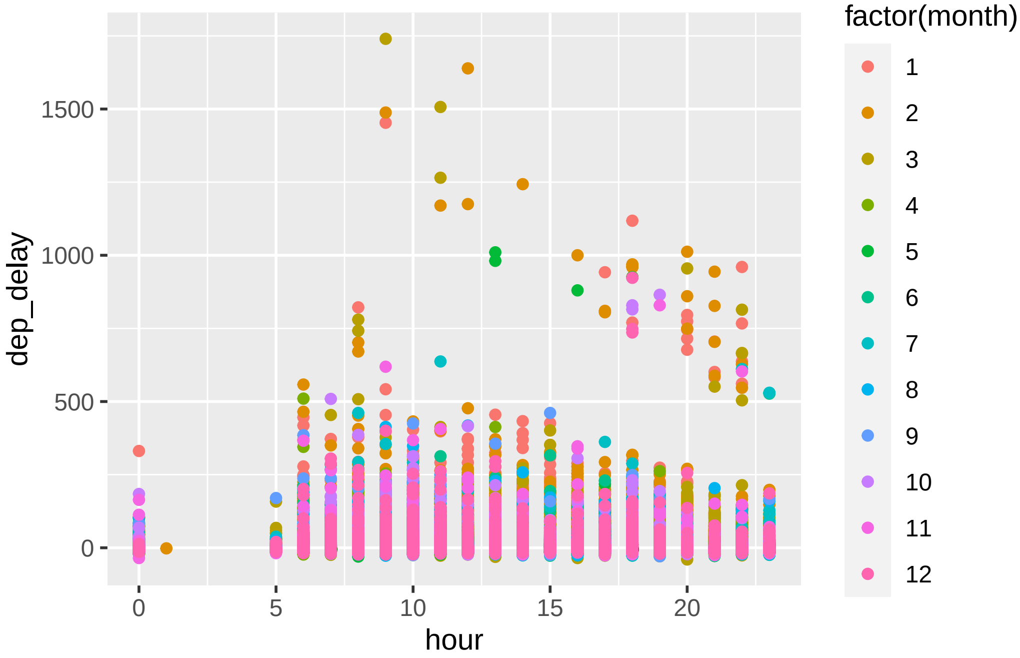
```
-- Attaching packages --------------------------------------- tidyverse 1.3.2 --
v ggplot2 3.3.6      v purrr   0.3.4
v tibble  3.1.8      v dplyr   1.0.10
v tidyr   1.2.1      v stringr 1.4.1
v readr   2.1.2      v forcats 0.5.2
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
```

```r
library(ggplot2)
data(flights)
```

Initial exploratory version:

```r
ggplot(flights) +
    geom_point(aes(x=hour, y=dep_delay, color=factor(month)))
```

```
Warning: Removed 7369 rows containing missing values (geom_point).
```



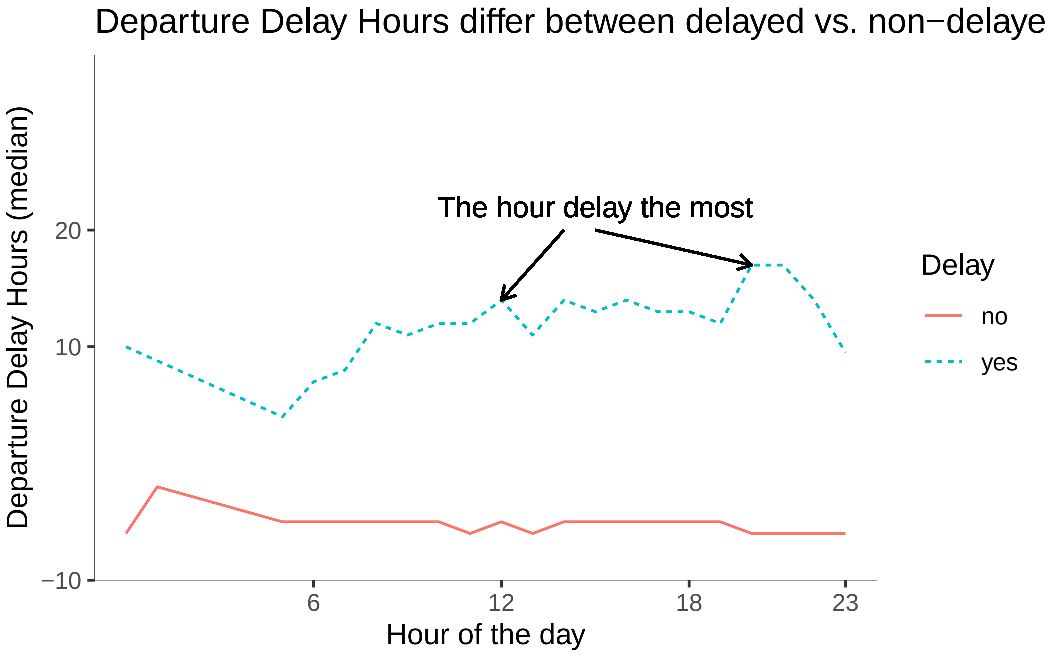
Communicatory plot:

```
# reduce to hourly data
by_delay <- flights %>%
  drop_na() %>%
  mutate(delay = ifelse(dep_delay > 0, 'yes', 'no')) %>%
  group_by(delay, hour) %>%
  summarise(dep_delay = median(dep_delay))
```

```
`summarise()` has grouped output by 'delay'. You can override using the
`.groups` argument.
```

```
morning_peak <- filter(by_delay, hour == 12, delay=='yes')$dep_delay
afternoon_peak <- filter(by_delay, hour == 20, delay=='yes')$dep_delay

ggplot(by_delay) +
  geom_line(aes(x = hour, y = dep_delay,
                color = delay, linetype = delay)) +
  theme(panel.background = element_blank(), # backgound/axis
        axis.line = element_line(colour = "black", size = .1),
        legend.key = element_rect(fill = "white")) +
  scale_x_continuous(breaks=c(6, 12, 18, 23), # x-axis
                     limits=c(0,23),
                     expand=c(0, 1)) +
  scale_y_continuous(breaks=c(-10, 10, 20), # y-axis
                     limits=c(-10, 35),
                     expand=c(0,0)) +
  labs(x = 'Hour of the day',
       y = 'Departure Delay Hours (median)',
       title = 'Departure Delay Hours differ between delayed vs. non-delayed') +
  guides(color = guide_legend(title = "Delay"),
         shape = guide_legend(title = "Delay"),
         linetype = guide_legend(title = "Delay")) +
  geom_text(aes(x = 15, y = 22, label = 'The hour delay the most')) +
  geom_segment(aes(x = 14, y = 20, xend = 12, yend = morning_peak),
               arrow = arrow(length = unit(0.03, 'npc'))) +
  geom_segment(aes(x = 15, y = 20, xend = 20, yend = afternoon_peak),
               arrow = arrow(length=unit(0.03, 'npc')))
```


Departure Delay Hours differ between delayed vs. non−delaye
The hour delay the most
Departure Delay Hours (median)
Hour of the day
Delay
no
yes