

Recreate variability plot

Alireza Mostafizi

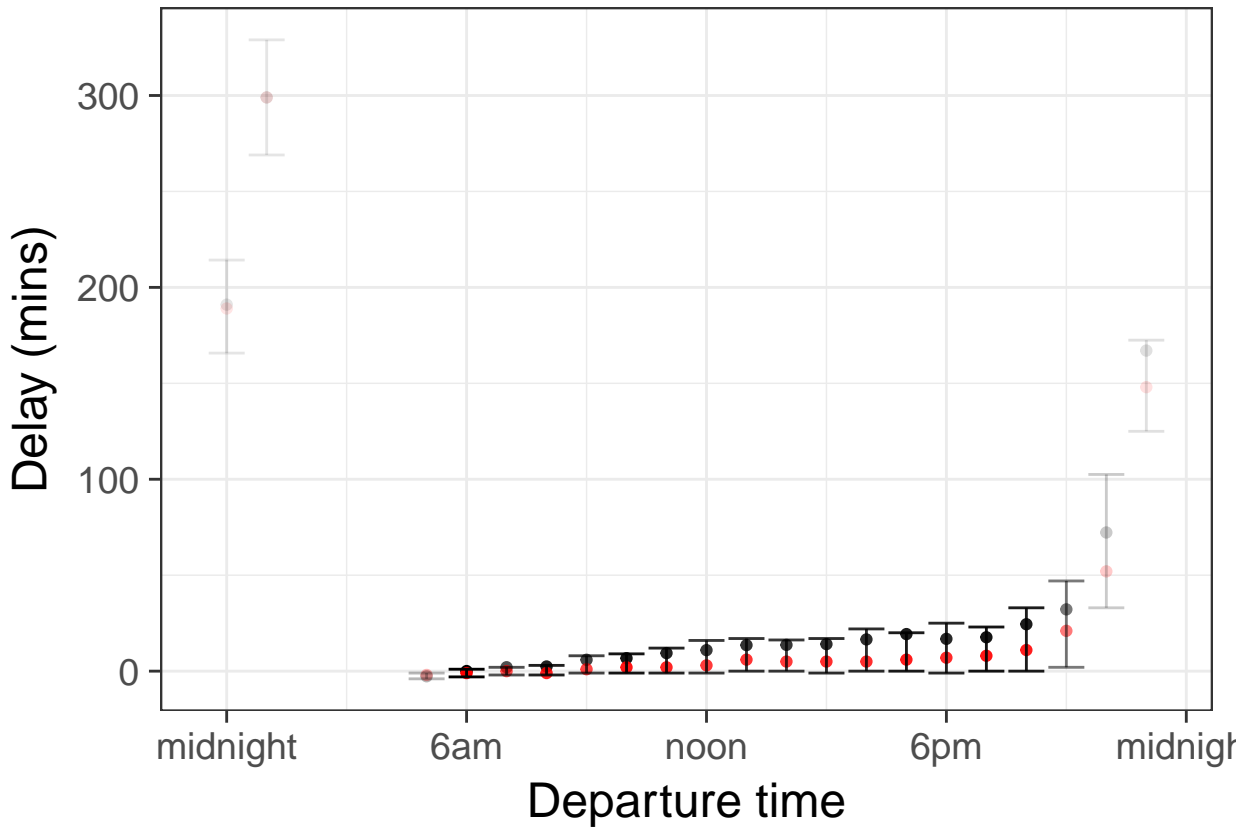
May 24, 2018

The code to generate the summaries, as mentioned in the assignment,

```
#install.packages("hflights")
library(tidyverse)
library(hflights)
hflights_df <- as_tibble(hflights)
hflights_df <- mutate(hflights_df,
                      DepHour = floor(DepTime/100),
                      DayOfWeek = factor(DayOfWeek,
                                          labels = c("Mon", "Tue", "Wed", "Thu",
                                                    "Fri", "Sat", "Sun")),
                      Date = ISOdate(Year, Month, DayofMonth)
                      )
hou <- filter(hflights_df, Origin == "HOU")
hou_mon <- filter(hou, DayOfWeek == "Mon")
# over all mondays in 2011, avg delay of flights departing by hour
hou_mon_avg <- hou_mon %>%
  group_by(DepHour) %>%
  summarise(avg_delay = mean(DepDelay))
# for each monday in 2011, avg delay of flights departing by hour
hou_mon_day <- filter(hou, DayOfWeek == "Mon") %>%
  group_by(Date, DepHour) %>%
  summarise(avg_delay = mean(DepDelay))
# quantiles for delay by time
hou_mon_q <- hou_mon %>% group_by(DepHour) %>%
  summarise(n = n(),
            q25 = quantile(DepDelay, probs = 0.25, na.rm = TRUE),
            q50 = quantile(DepDelay, probs = 0.5, na.rm = TRUE),
            q75 = quantile(DepDelay, probs = 0.75, na.rm = TRUE))
```

This code joins two tibbles, *hou_mon_avg* which contains the average delays and *hou_mon_q* that contains 25-, 50-, and 75-percentile values, and finally it plots the graph with error bars.

```
library(ggplot2)
library(dplyr)
hou_mon <- inner_join(hou_mon_avg, hou_mon_q)
ggplot(data = hou_mon, aes(alpha = n)) +
  geom_point(aes(DepHour, avg_delay)) +
  geom_point(aes(DepHour, q50), color = 'red') +
  geom_errorbar(data = hou_mon_q, mapping = aes(x = DepHour, ymin = q25, ymax = q75)) +
  ylab("Delay (mins)") +
  xlab("Departure time") +
  scale_x_continuous(breaks = seq(0, 24, 6),
                     labels = c("midnight", "6am", "noon", "6pm", "midnight")) +
  theme_bw(18)+
  guides(alpha = FALSE)
```



The error bars show the 25- and 75-percentile values. The red dot shows the median and the black dot represents the average delay. Also, transparency is correlated with the number of observations. The more transparent, the lower the number of observations.