Lab No. 12

Elizabeth Walter

29 November, 2021

```
knitr::opts_chunk$set(echo = TRUE,
                      fig.align = 'center')
```

```
library(tidyverse)
library(lubridate)
```

```
booking_raw <- read.csv('https://ssc442.netlify.app/data/Lab11_booking.csv')

roomrates_raw <-read.csv("https://ssc442.netlify.app/data/Lab11_roomrates.csv", header = FALSE)
#preserve dates when reading in, takes couple extra steps
colnames(roomrates_raw) <- roomrates_raw[1,] #make first row col names
roomrates_raw <- roomrates_raw[-1,] # remove first row (col names row as obs)
# columns after A are dates in d-m-yy format

parking_raw <- read.csv("https://ssc442.netlify.app/data/Lab11_parking.csv")
# corp C & D - add parking to cost - $60, all other corps, $0
```

1. Total up spending by corporation and report the biggest and smallest spenders inclusive of rooms and parking.

Smallest spending corporation, rooms + parking: E ($87,661.18). Largest spending corporation, rooms + parking: D ($101,098.70).

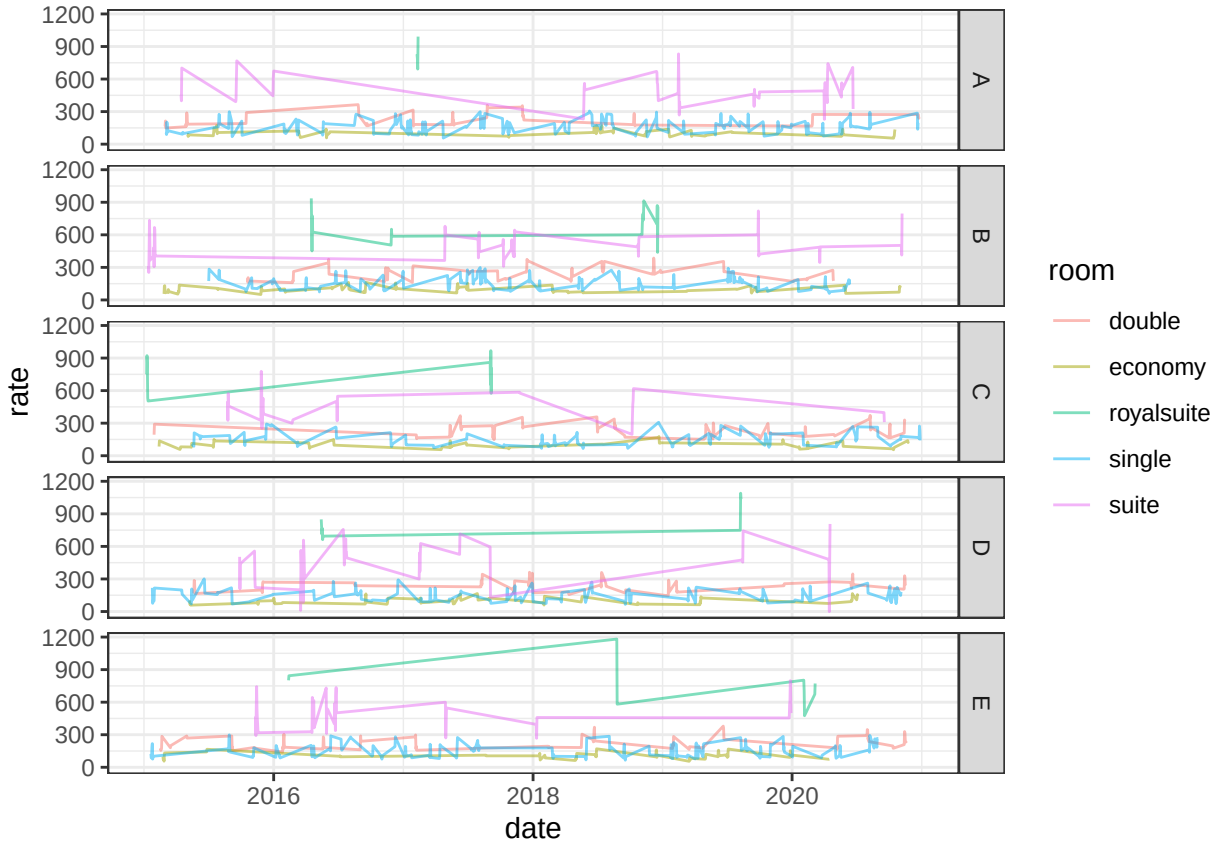
```
booking_cl <- booking_raw %>%
  mutate(date_occupied = ymd_hms(date_occupied),
  date = round_date(date_occupied, "day")) %>% # remove h:m:s from dates
  separate(room_use, into = c("room", NA)) %>%
  select(!date_occupied)

roomrates_cl <- roomrates_raw %>%
  gather(date, rate, -name, convert = TRUE) %>% #gather all columns except name
  separate(name, into = c(NA, "room"), extra = "merge") %>%
  mutate(date = dmy(date),
         rate = as.numeric(rate)) # changes dates from d/m/yy to YYYY-MM-DD

# make royal_suite and royalsuite the same between 2 datasets so dont lose when joining
# from: https://stackoverflow.com/questions/43108310/how-to-rename-a-specific-cell-in-r
roomrates_cl$room <- gsub("royal_suite", "royalsuite", roomrates_cl$room)
```

```r
cost <- left_join(booking_cl, roomrates_cl, by = 'date') %>%
  # remove room obs from roomrates_cl that dont match those in booking_cl
  # use to get rates from roomrates_cl that match day/room combos for booking_cl
  filter(room.x == room.y)

# dplyr::setequal(cost$room.x, cost$room.y) # double check matches

# clean up - remove duplicate room row & rename
cost <- cost %>% select(!room.y) %>% rename(room = room.x) %>%
  # add parking on, get new daily totals
  mutate(parking = ifelse(cost$corp == "C" | cost$corp == "D", 60, 0),
         total = rate + parking)

# Find totals by corporation
cost_tot <- cost %>% group_by(corp) %>% summarize(sum_tot = sum(total)) %>%
  arrange(sum_tot)

head(cost_tot)
```

```
## # A tibble: 5 x 2
##   corp  sum_tot
##   <chr>   <dbl>
## 1 E      87661.
## 2 A      88516.
## 3 C      89954.
## 4 B      98318.
## 5 D     101099.
```

2. Visualize (using ggplot) each corporation's spending at the hotel over time and by roomtype. Make one plot with ggplot that shows this.

```
plot1 <- cost %>% arrange(date) %>%
  ggplot(aes(x = date, y = rate, color = room)) +
  geom_line(size =.5, alpha = .5) +
  facet_grid(corp ~ .) +
  theme_bw()
plot1
```



3. Visualize (using ggplot) the room rates over time by room type. Can you tell what factors determine room price? Make two total plots, the first showing the room rates over time by room type, and the second explaining some feature of one of the room rates (e.g. when is the double room price high? When is it low?).

Plot 1 - Room rates over time by room type

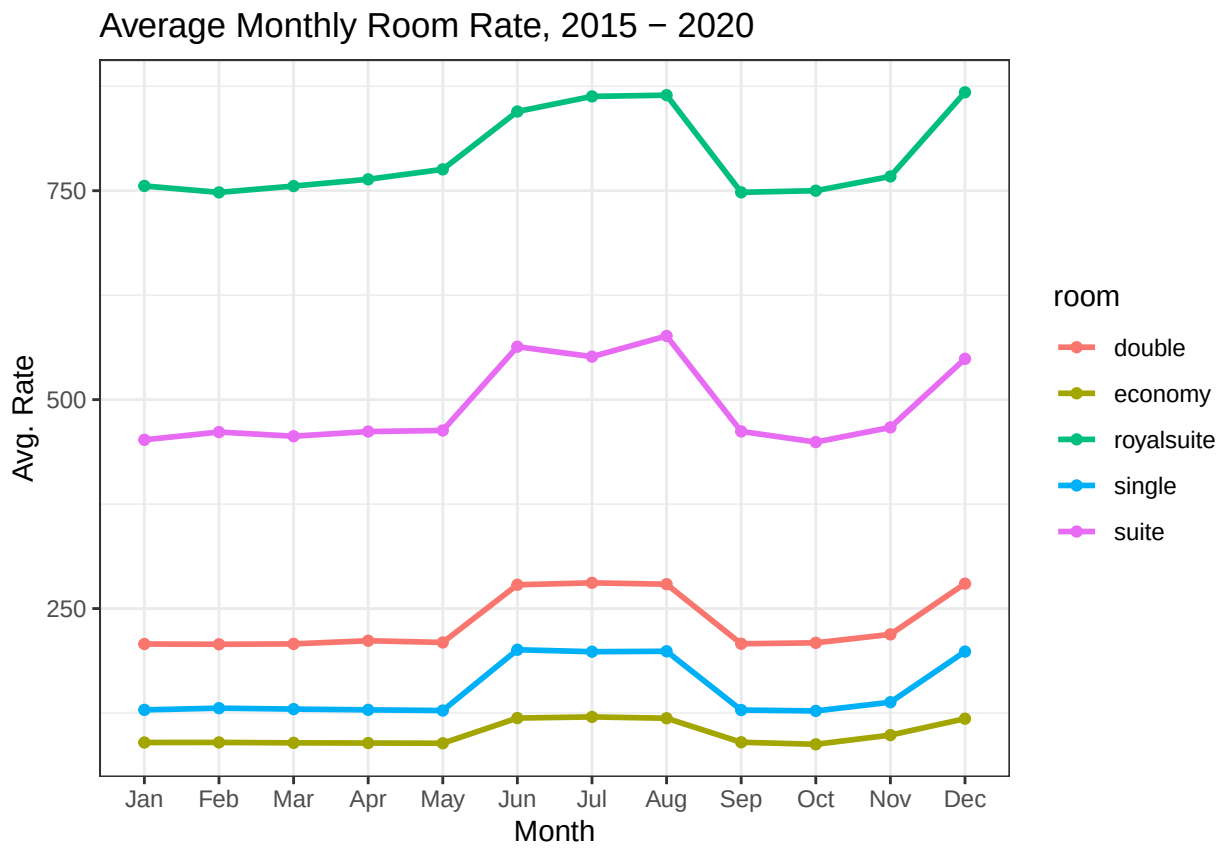
```
plot2 <- roomrates_cl %>%
  ggplot(aes(x = date, y = rate, color = room)) +
  geom_line(alpha = .4, size = 1) +
  theme_bw() +
  theme(legend.position = "bottom") +
  scale_x_date(date_breaks = "1 year", date_labels = "%Y") +
  labs(title = "Room Rates Over Time")
plot2
```

Plot 2 - Explaining a feature of room rates

Based on plot 1 of Q3 (plot2), it appears that the changes in room rate are fairly cyclical and that overall the pattern exists across room types. Therefore, I suspected that room rates were adjusted based on the month or season, as change in demand for rooms probably follows a similar (seasonal) pattern each year. To visualize this, I found the average rate for each month, averaged across the 5 years of data available, for each

room type. Once visualized, we see a clear pattern that generally holds across all room types: Room rates are fairly steady in the beginning of the year, then they increase for the summer months of June-August before declining back to pre-summer levels, and then heading into the holiday season, rates increase again to similar peak-summer rates. These trends align with the general public's travel plans: Most people take time off to travel in the summer when children are out of school for summer break or work is slower, and again at the end of the year when kids have winter break and holidays prompt people to travel to visit friends and family.

```
plot3 <- roomrates_cl %>%
  arrange(room, date) %>%
  mutate(month = month(date),
         month_chr = month(date, label = TRUE),
         day = day(date),
         year = year(date)) %>%
  group_by(room, month) %>%
  summarize(avg_rt = mean(rate)) %>%
  ggplot(aes(x = month, y = avg_rt, color = room)) +
  geom_point() +
  geom_line(size = 1) +
  theme_bw() +
  labs(title = "Average Monthly Room Rate, 2015 - 2020",
       x = "Month",
       y = "Avg. Rate") +
  scale_x_discrete(breaks = waiver(), limits = c("Jan", "Feb", "Mar", "Apr",
                                                 "May","Jun", "Jul", "Aug",
                                                 "Sep", "Oct", "Nov", "Dec"))
plot3
```


Average Monthly Room Rate, 2015 − 2020
Avg. Rate
750
500
250
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Month
room
double
economy
royalsuite
single
suite