Hotel Bookings

2022-04-30

Hotels Booking Dataset

Step 1: Importing Data

The data in this example is originally from the article Hotel Booking Demand Datasets (https://www.sciencedirect.com/science/article/pii/S2352340918315191), written by Nuno Antonio, Ana Almeida, and Luis Nunes for Data in Brief, Volume 22, February 2019.

The data was downloaded and cleaned by Thomas Mock and Antoine Bichat for #TidyTuesday during the week of February 11th, 2020 (https://github.com/rfordatascience/tidytuesday/blob/master/data/2020/2020-02-11/readme.md).

You can learn more about the dataset here: https://www.kaggle.com/jessemostipak/hotel-booking-demand

Run the code below to read in the file 'hotel_bookings.csv' into a data frame:

```
hotel_bookings <- read.csv("hotel_bookings.csv")
```

Step 2: Refresh Your Memory

```
head(hotel_bookings)
```

```
##          hotel is_canceled lead_time arrival_date_year arrival_date_month
## 1 Resort Hotel           0       342              2015               July
## 2 Resort Hotel           0       737              2015               July
## 3 Resort Hotel           0         7              2015               July
## 4 Resort Hotel           0        13              2015               July
## 5 Resort Hotel           0        14              2015               July
## 6 Resort Hotel           0        14              2015               July
##   arrival_date_week_number arrival_date_day_of_month stays_in_weekend_nights
## 1                       27                         1                       0
## 2                       27                         1                       0
## 3                       27                         1                       0
## 4                       27                         1                       0
## 5                       27                         1                       0
## 6                       27                         1                       0
##   stays_in_week_nights adults children babies meal country market_segment
## 1                    0      2        0      0   BB     PRT         Direct
## 2                    0      2        0      0   BB     PRT         Direct
## 3                    1      1        0      0   BB     GBR         Direct
## 4                    1      1        0      0   BB     GBR      Corporate
## 5                    2      2        0      0   BB     GBR      Online TA
## 6                    2      2        0      0   BB     GBR      Online TA
##   distribution_channel is_repeated_guest previous_cancellations
## 1               Direct                 0                      0
## 2               Direct                 0                      0
```

```
## 3               Direct                0                    0
## 4            Corporate                0                    0
## 5                TA/TO                0                    0
## 6                TA/TO                0                    0
##   previous_bookings_not_canceled reserved_room_type assigned_room_type
## 1                              0                  C                  C
## 2                              0                  C                  C
## 3                              0                  A                  C
## 4                              0                  A                  A
## 5                              0                  A                  A
## 6                              0                  A                  A
##   booking_changes deposit_type agent company days_in_waiting_list customer_type
## 1               3   No Deposit  NULL    NULL                    0     Transient
## 2               4   No Deposit  NULL    NULL                    0     Transient
## 3               0   No Deposit  NULL    NULL                    0     Transient
## 4               0   No Deposit   304    NULL                    0     Transient
## 5               0   No Deposit   240    NULL                    0     Transient
## 6               0   No Deposit   240    NULL                    0     Transient
##   adr required_car_parking_spaces total_of_special_requests reservation_status
## 1   0                            0                         0          Check-Out
## 2   0                            0                         0          Check-Out
## 3  75                            0                         0          Check-Out
## 4  75                            0                         0          Check-Out
## 5  98                            0                         1          Check-Out
## 6  98                            0                         1          Check-Out
##   reservation_status_date
## 1              2015-07-01
## 2              2015-07-01
## 3              2015-07-02
## 4              2015-07-02
## 5              2015-07-03
## 6              2015-07-03
```

```
colnames(hotel_bookings)
```

```
##  [1] "hotel"                          "is_canceled"
##  [3] "lead_time"                      "arrival_date_year"
##  [5] "arrival_date_month"             "arrival_date_week_number"
##  [7] "arrival_date_day_of_month"      "stays_in_weekend_nights"
##  [9] "stays_in_week_nights"           "adults"
## [11] "children"                       "babies"
## [13] "meal"                           "country"
## [15] "market_segment"                 "distribution_channel"
## [17] "is_repeated_guest"              "previous_cancellations"
## [19] "previous_bookings_not_canceled" "reserved_room_type"
## [21] "assigned_room_type"             "booking_changes"
## [23] "deposit_type"                   "agent"
## [25] "company"                        "days_in_waiting_list"
## [27] "customer_type"                  "adr"
## [29] "required_car_parking_spaces"    "total_of_special_requests"
## [31] "reservation_status"             "reservation_status_date"
```

Step 3: Install and load the 'ggplot2' package

Install and load the `ggplot2` package

Step 4: Annotating your chart

The first step will be adding a title; that is often the first thing people will pay attention to when they encounter a data visualization for the first time. To add a title, you will add `labs()` at the end of your `ggplot()` command and then input a title there:

```
ggplot(data = hotel_bookings) +
  geom_bar(mapping = aes(x = market_segment,fill=market_segment)) +
  facet_wrap(~hotel) +
  labs(title="Comparison of market segments by hotel type for hotel bookings")
```

I also want to add another detail about what time period this data covers. To do this, I need to find out when the data is from.

I realize I can use the `min()` function on the year column in the data:

```
min(hotel_bookings$arrival_date_year)
```

```
## [1] 2015
```

And the `max()` function:

```
max(hotel_bookings$arrival_date_year)
```

```
## [1] 2017
```

But I will need to save them as variables in order to easily use them in my labeling; the following code chunk creates two of those variables:

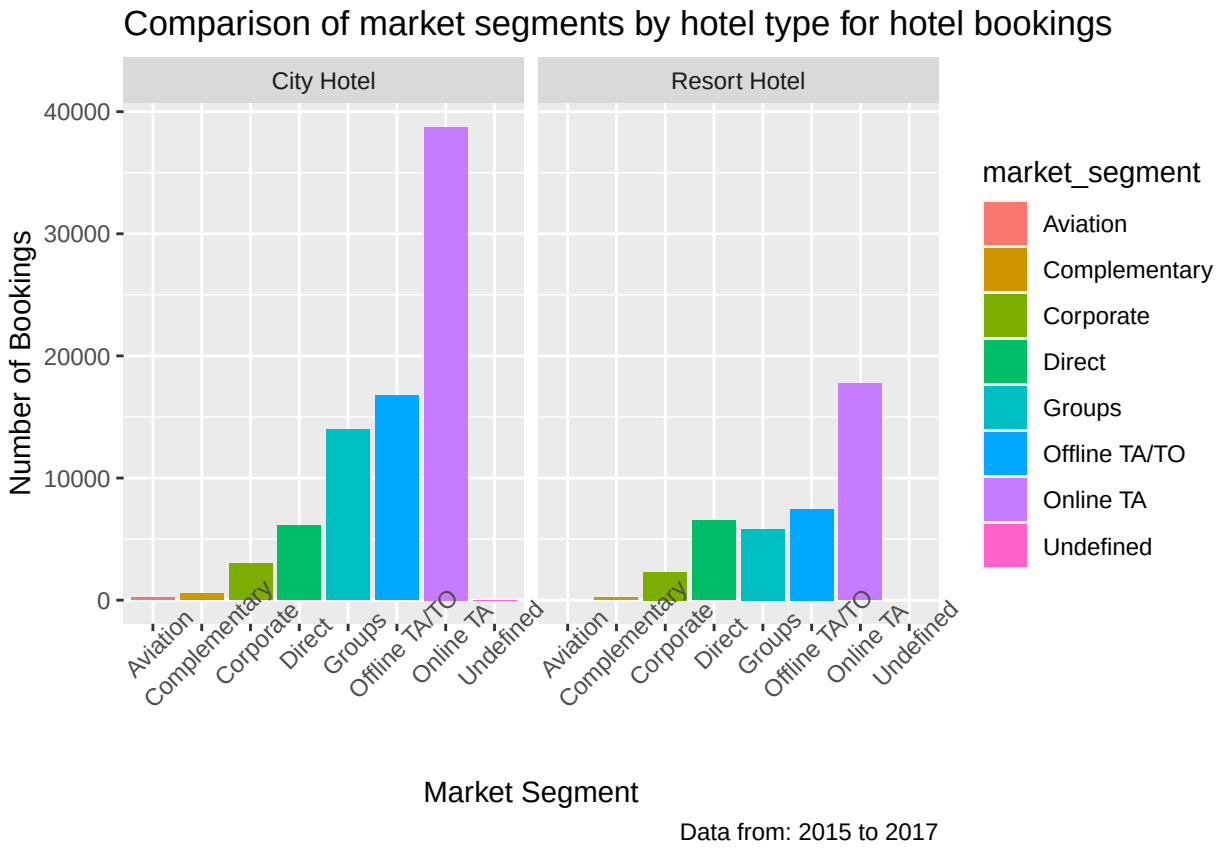
```
mindate <- min(hotel_bookings$arrival_date_year)
maxdate <- max(hotel_bookings$arrival_date_year)
```

I decide to switch the `subtitle` to a `caption` which will appear in the bottom right corner instead.

```
ggplot(data = hotel_bookings) +
  geom_bar(mapping = aes(x = market_segment,fill=market_segment)) +
  facet_wrap(~hotel) +
  labs(title="Comparison of market segments by hotel type for hotel bookings",
       caption=paste0("Data from: ", mindate, " to ", maxdate))
```

Now I want to clean up the x and y axis labels to make sure they are really clear. To do that, I add to the `labs()` function and use `x=` and `y=`.

```
ggplot(data = hotel_bookings) +
  geom_bar(mapping = aes(x = market_segment,fill=market_segment)) +
  facet_wrap(~hotel) +
  theme(axis.text.x = element_text(angle = 45))+
  labs(title="Comparison of market segments by hotel type for hotel bookings",
       caption=paste0("Data from: ", mindate, " to ", maxdate),
       x="Market Segment",
       y="Number of Bookings",
       )
```



Step 5: Saving Chart

The `ggsave()` function was used to save the last plot that was generated, so if I have run something after running the code chunk above, then run that code chunk again.

```
ggsave('hotel_booking_chart.png')
```

```
## Saving 6.5 x 4.5 in image
```