

GGPlot

The Scenario

In this scenario, you are a junior data analyst working for a hotel booking company. You have cleaned and manipulated your data, and gotten some initial insights you would like to share. Now, you are going to create some simple data visualizations with the `ggplot2` package. You will use basic `ggplot2` syntax and troubleshoot some common errors you might encounter

Step 1: Import data

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

Step 2: Look at the data

```
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)
```

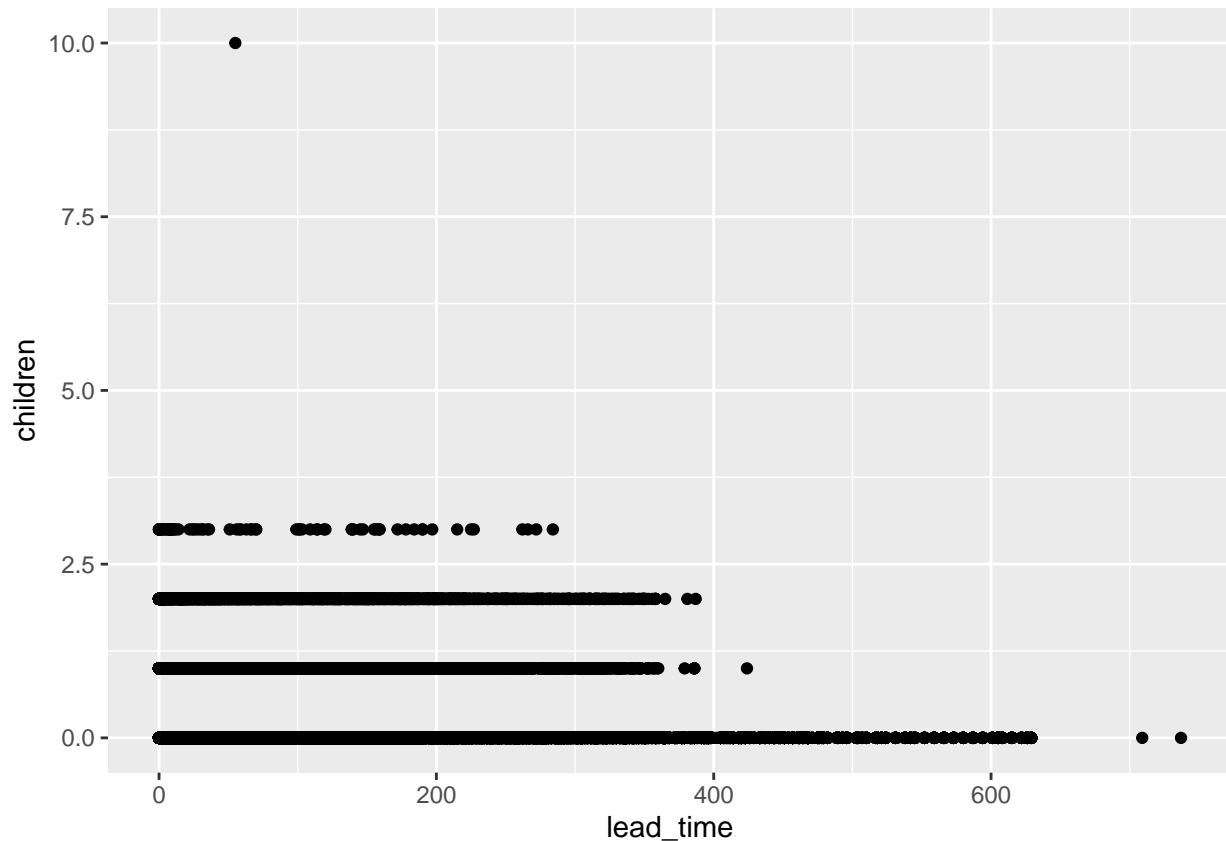
```
## [1] "hotel" "is_canceled"
## [3] "lead_time" "arrival_date_year"
## [5] "arrival_date_month" "arrival_date_week_number"
## [7] "arrival_date_day_of_month" "stays_in_weekend_nights"
## [9] "stays_in_week_nights" "adults"
## [11] "children" "babies"
## [13] "meal" "country"
## [15] "market_segment" "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

Step 4: Creating a plot

```
ggplot(data = hotel_bookings) +
  geom_point(mapping = aes(x = lead_time, y = children))
```

```
## Warning: Removed 4 rows containing missing values (`geom_point()`).
```



Next, your stakeholder says that she wants to increase weekend bookings, an important source of revenue for the hotel. Your stakeholder wants to know what group of guests book the most weekend nights in order to target that group in a new marketing campaign. She suggests that guests without children book the most weekend nights. Is this true?

Step 5: Solve

```
ggplot(data = hotel_bookings) +
  geom_point(mapping = aes(x = stays_in_weekend_nights, y = children))

## Warning: Removed 4 rows containing missing values (`geom_point()`).
```

