

Lesson 3: Aesthetics Solutions

Aesthetics and visualizations activity solutions

This document contains the solutions for the aesthetics and visualizations activity. You can use these solutions to check your work and ensure that your code is correct or troubleshoot your code if it is returning errors. If you haven't completed the activity yet, we suggest you go back and finish it before reading the solutions.

If you experience errors, remember that you can search the internet and the RStudio community for help:

<https://community.rstudio.com/#>

Step 1: Import your data

If you haven't exited out of RStudio since importing this data last time, you can skip these steps. Rerunning these code chunks won't affect your console if you want to run them just in case, though.

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:

Step 2: Refresh Your Memory

By now, you are pretty familiar with this data set. But you can refresh your memory with the `head()` and `colnames()` functions. Run two code chunks below to get at a sample of the data and also preview all the column names:

```
head(hotel_bookings)
```

```
##          hotel is_canceled lead_time arrival_date_year
## 1 Resort Hotel         0        342            2015
## 2 Resort Hotel         0        737            2015
## 3 Resort Hotel         0         7            2015
## 4 Resort Hotel         0        13            2015
## 5 Resort Hotel         0        14            2015
## 6 Resort Hotel         0        14            2015
##   arrival_date_week_number arrival_date_day_of_month status
## 1                      27                     1      1
## 2                      27                     1      1
## 3                      27                     1      1
## 4                      27                     1      1
```

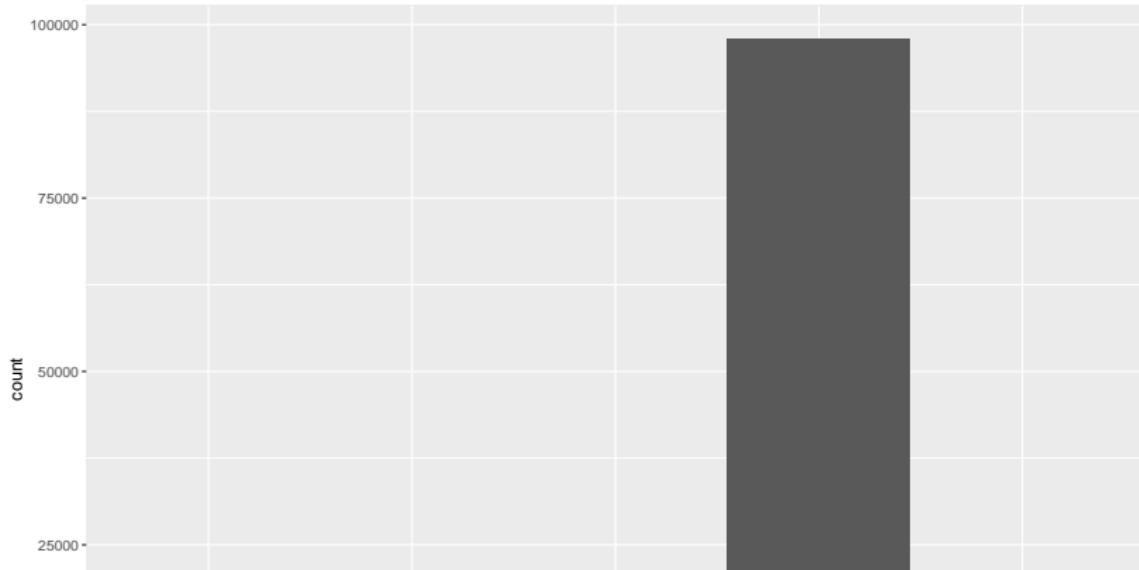
Step 3: Install and load the 'ggplot2' package (optional)

Run the code chunk below to install and load ggplot2 if you don't have it installed and loaded already. This may take a few minutes!

Step 4: Making a Bar Chart

Previously, you used `geom_point` to make a scatter plot comparing lead time and number of children. Now, you will use `geom_bar` to make a bar chart in this code chunk:

```
ggplot(data = hotel_bookings) +  
  geom_bar(mapping = aes(x = distribution_channel))
```



Practice quiz

Use the bar chart you created to answer this question: what distribution type has the most number of bookings? Note your answer and respond in the Coursera platform.

A: TA/TO B: Direct C: GDS D: Corporate Answer: A. The TA/TO distribution type has the most number of bookings.

Step 5: Diving deeper into bar charts

After exploring your bar chart, your stakeholder has more questions. Now they want to know if the number of bookings for each distribution type is different depending on whether or not there was a deposit or what market segment they represent.

Try running the code below to answer the question about deposits. You will use 'fill=deposit_type' to accomplish this.

