

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:

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

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 a sample of the data and also preview all the column names:

```
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 (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!

```
install.packages('ggplot2')
```

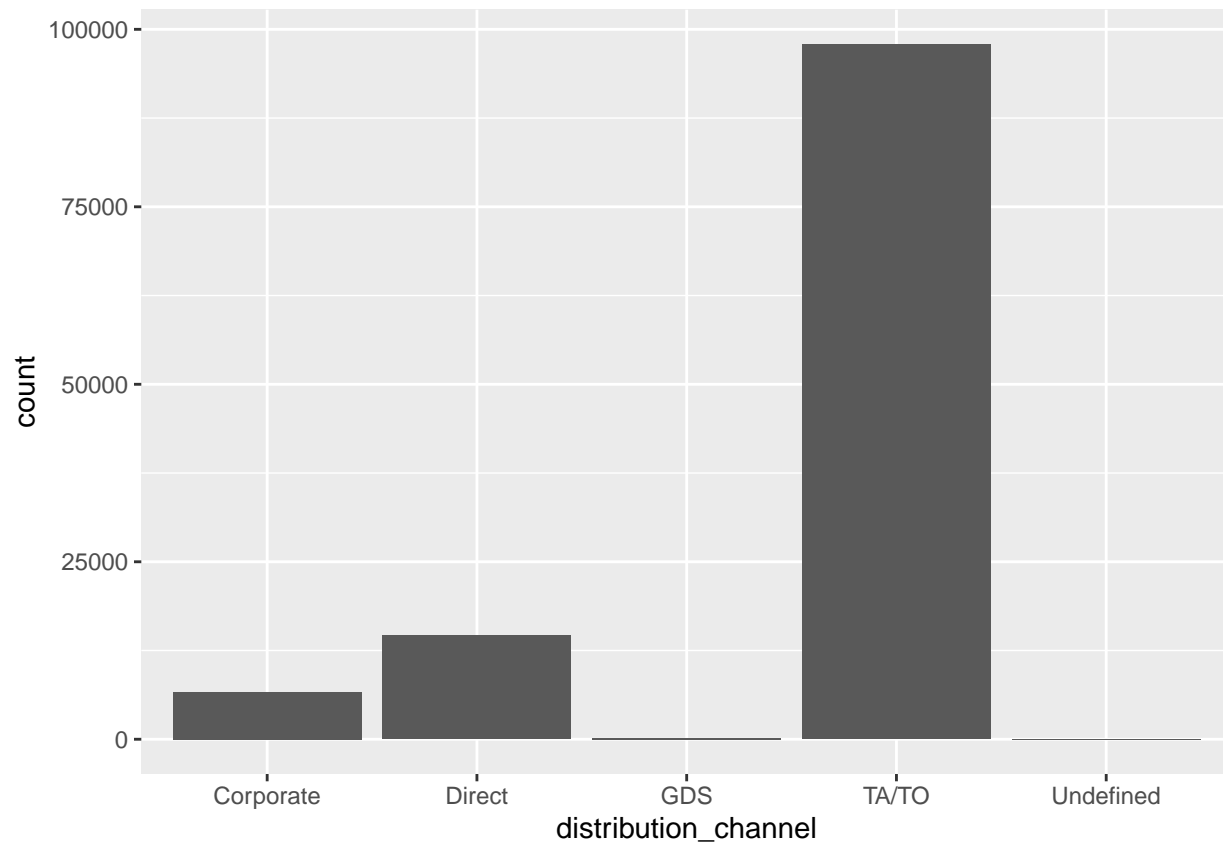
```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.3'
## (as 'lib' is unspecified)
```

```
library(ggplot2)
```

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.

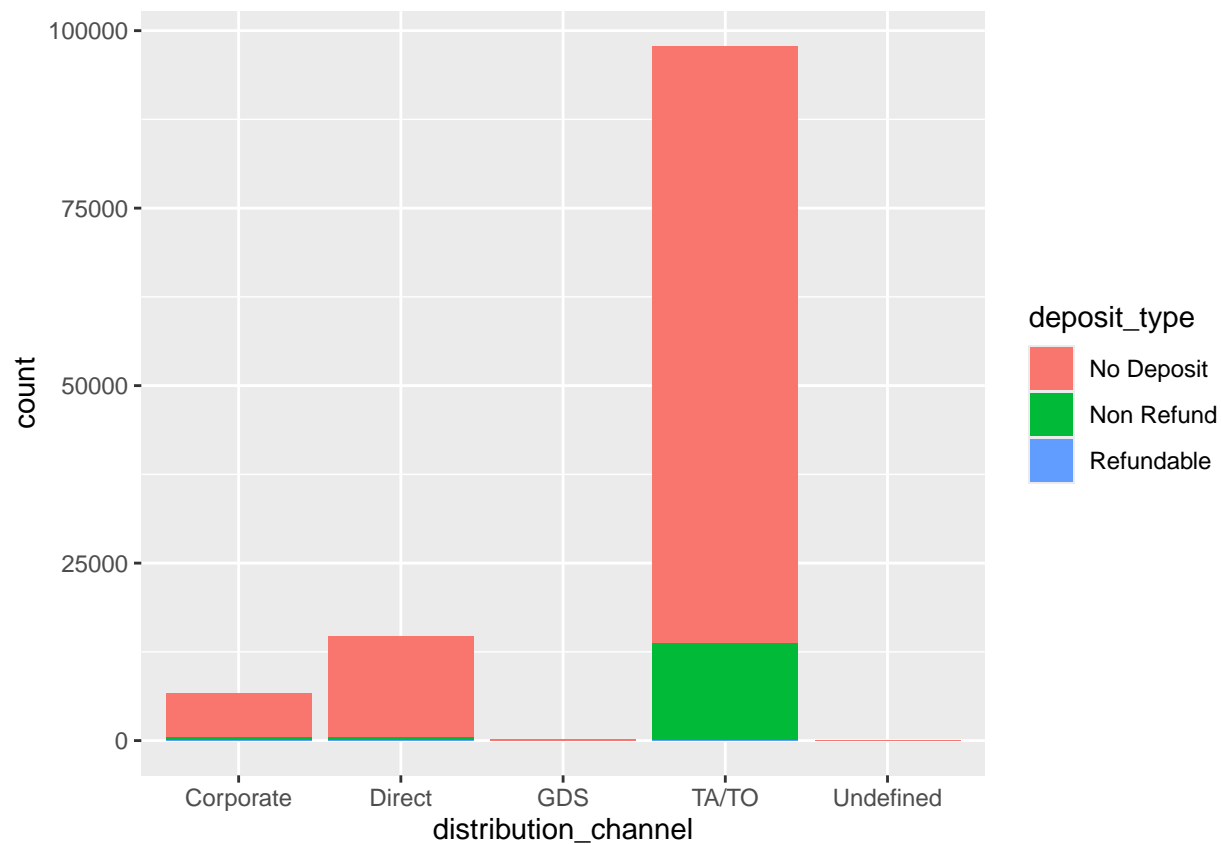
A: TA/TO B: Direct C: GDS D: Corporate Answer: A. The TA/TO distribution type has the most number of bookings.

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.

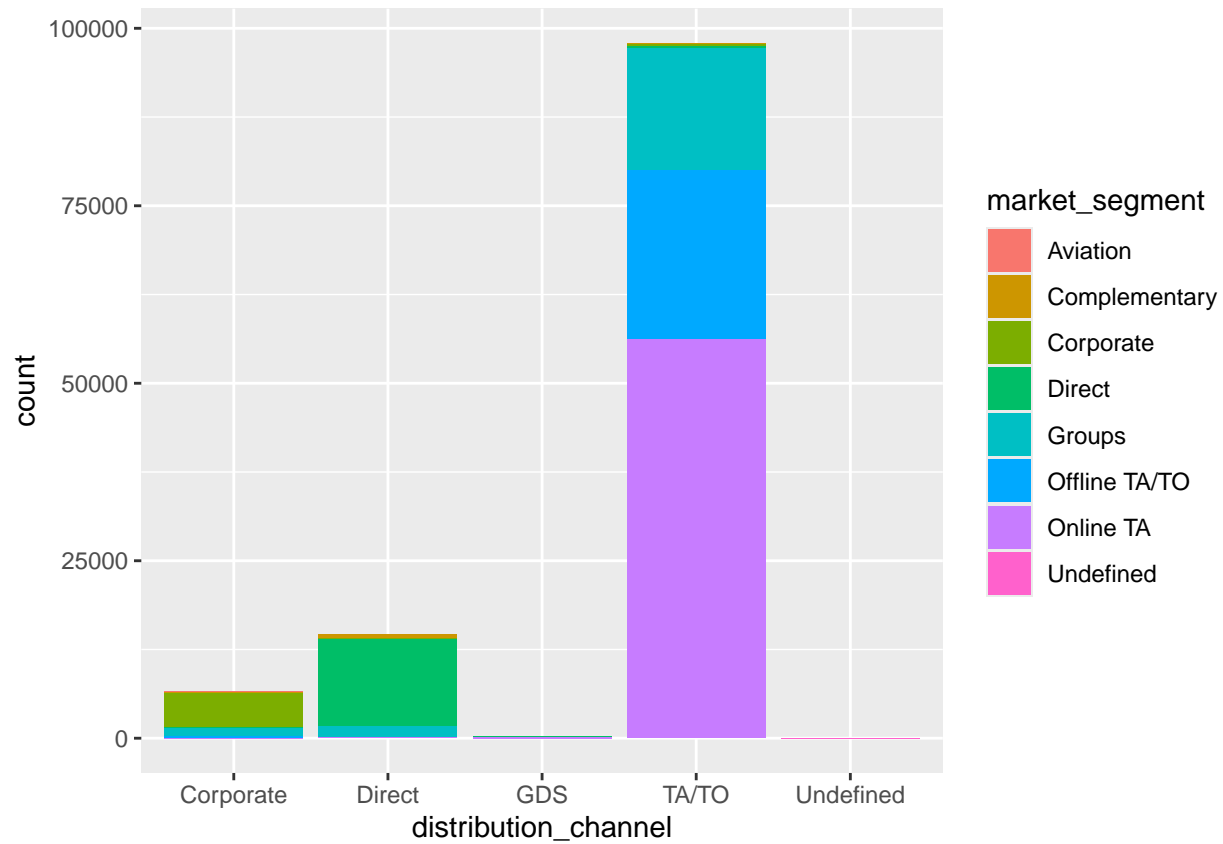
Try running the code below to answer the question about deposits. You will use 'fill=deposit_type' to accomplish this.

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



Now try running the code below to answer the question about different market segments. You will use 'fill=market_segment' to accomplish this.

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

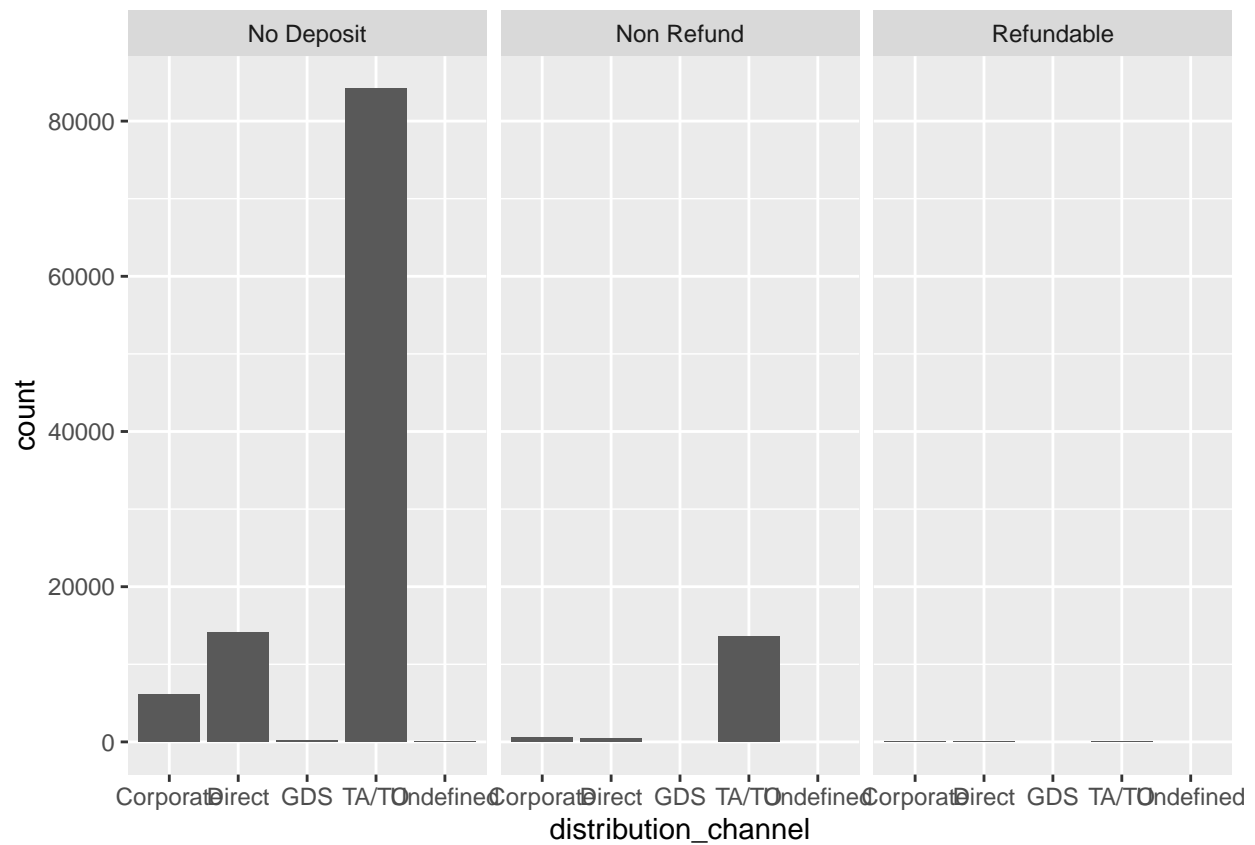


Step 6: Facets galore

After reviewing the new charts, your stakeholder asks you to create separate charts for each deposit type and market segment to help them understand the differences more clearly.

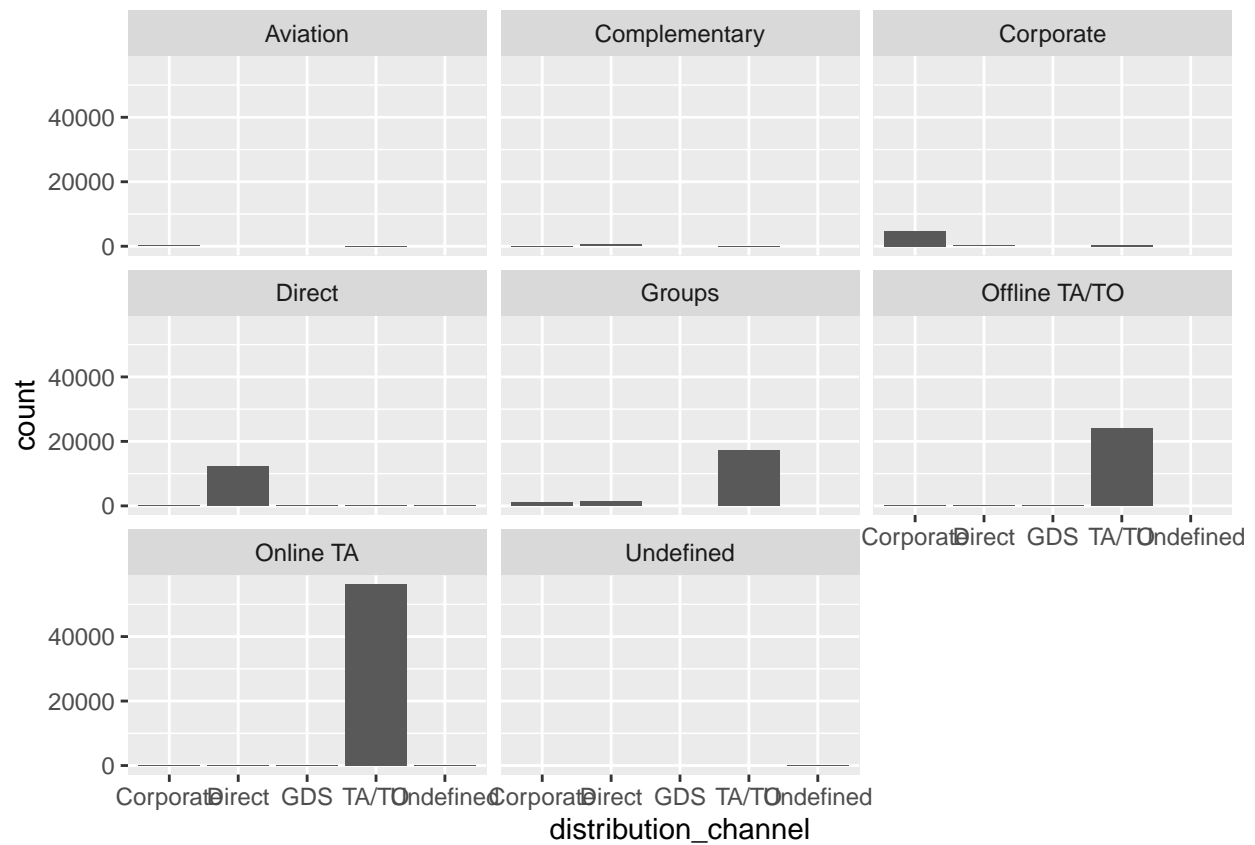
Run the code chunk below to create a different chart for each deposit type:

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



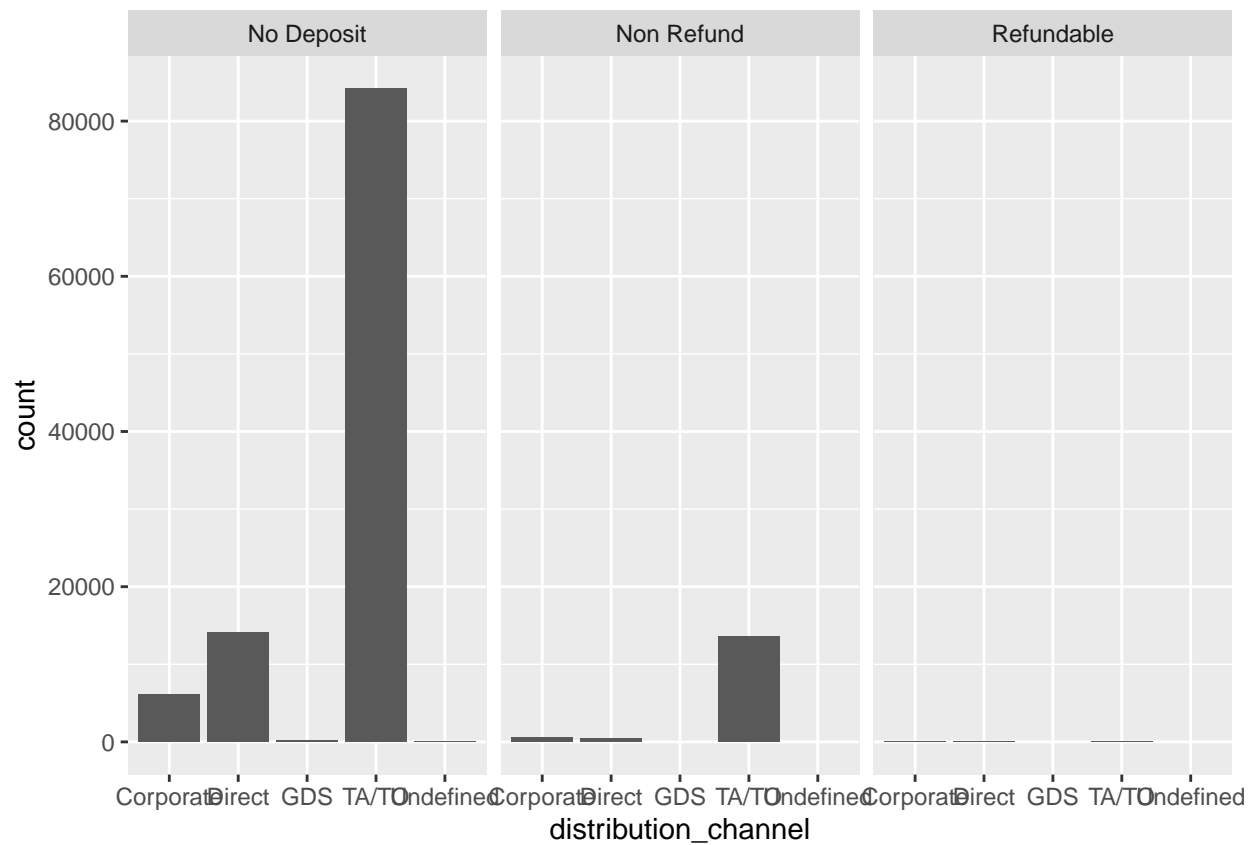
Run the code chunk below to create a different chart for each market segment:

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



The `facet_grid` function does something similar. The main difference is that `facet_grid` will include plots even if they are empty. Run the code chunk below to check it out:

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



Now, you could put all of this in one chart and explore the differences by deposit type and market segment. Run the code chunk below to find out; notice how the ~ character is being used before the variables that the chart is being split by:

```
ggplot(data = hotel_bookings) +  
  geom_bar(mapping = aes(x = distribution_channel)) +  
  facet_wrap(~deposit_type~market_segment)
```