Lesson 1

David John Baker

04/05/2020

ggplot2

The thing you often hear a lot about with R is how it's great at making graphics. This is true. If you combine the functionality from a library like `ggplot2` with some of the tools we will look at next session, you'll be a plot pro.

The library that we're going to focus on today is called `ggplot2`. Continuing on from what we talked about in Lesson 0, the code below will look somewhat a familiar.

```
library(readr) # for getting data
library(ggplot2) # for plotting data

tips <- read_csv("tips.csv")
```

```
## Warning: Missing column names filled in: 'X1' [1]
```

```
## Parsed with column specification:
## cols(
##   X1 = col_double(),
##   total_bill = col_double(),
##   tip = col_double(),
##   sex = col_character(),
##   smoker = col_character(),
##   day = col_character(),
##   time = col_character(),
##   size = col_double()
## )
```

```
# You can change where this is output above in "Settings (by knit) > Chunk Output in Console"
tips
```

```
## # A tibble: 244 x 8
##       X1 total_bill   tip sex    smoker day   time    size
##    <dbl>      <dbl> <dbl> <chr>  <chr>  <chr> <chr>  <dbl>
##  1     0      17.0   1.01 Female No     Sun   Dinner     2
##  2     1      10.3   1.66 Male   No     Sun   Dinner     3
##  3     2      21.0   3.5  Male   No     Sun   Dinner     3
##  4     3      23.7   3.31 Male   No     Sun   Dinner     2
##  5     4      24.6   3.61 Female No     Sun   Dinner     4
##  6     5      25.3   4.71 Male   No     Sun   Dinner     4
```

```
##  7     6       8.77  2    Male   No     Sun   Dinner     2
##  8     7      26.9   3.12 Male   No     Sun   Dinner     4
##  9     8      15.0   1.96 Male   No     Sun   Dinner     2
## 10     9      14.8   3.23 Male   No     Sun   Dinner     2
## # ... with 234 more rows
```

Let's make a scatterplot of our total bill by tips with ggplot.

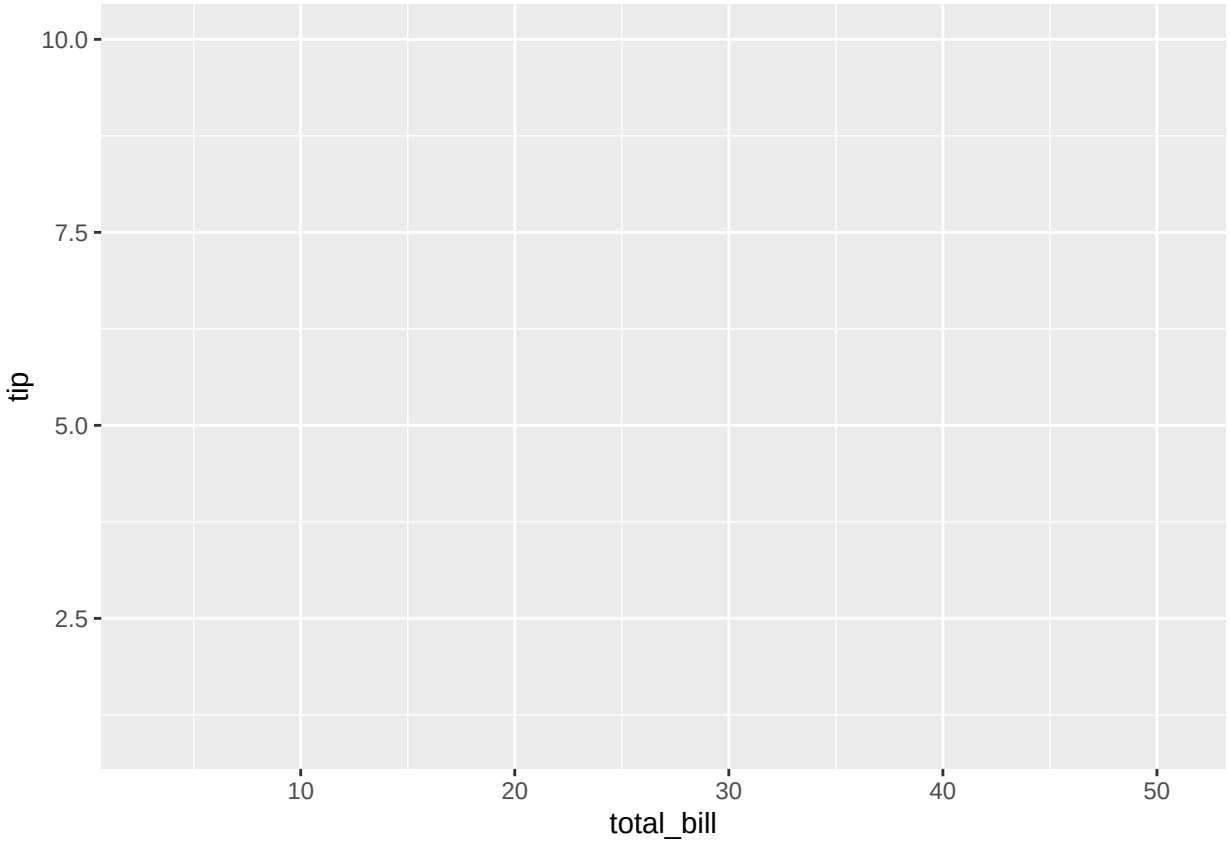
What do we need to tell ggplot if we want to make a plot of this?

Well, the first thing we should tell ggplot is our data. If we do this we get a nice blank screen, why? ggplot has no idea what data we want to plot!

```
ggplot(tips)
```

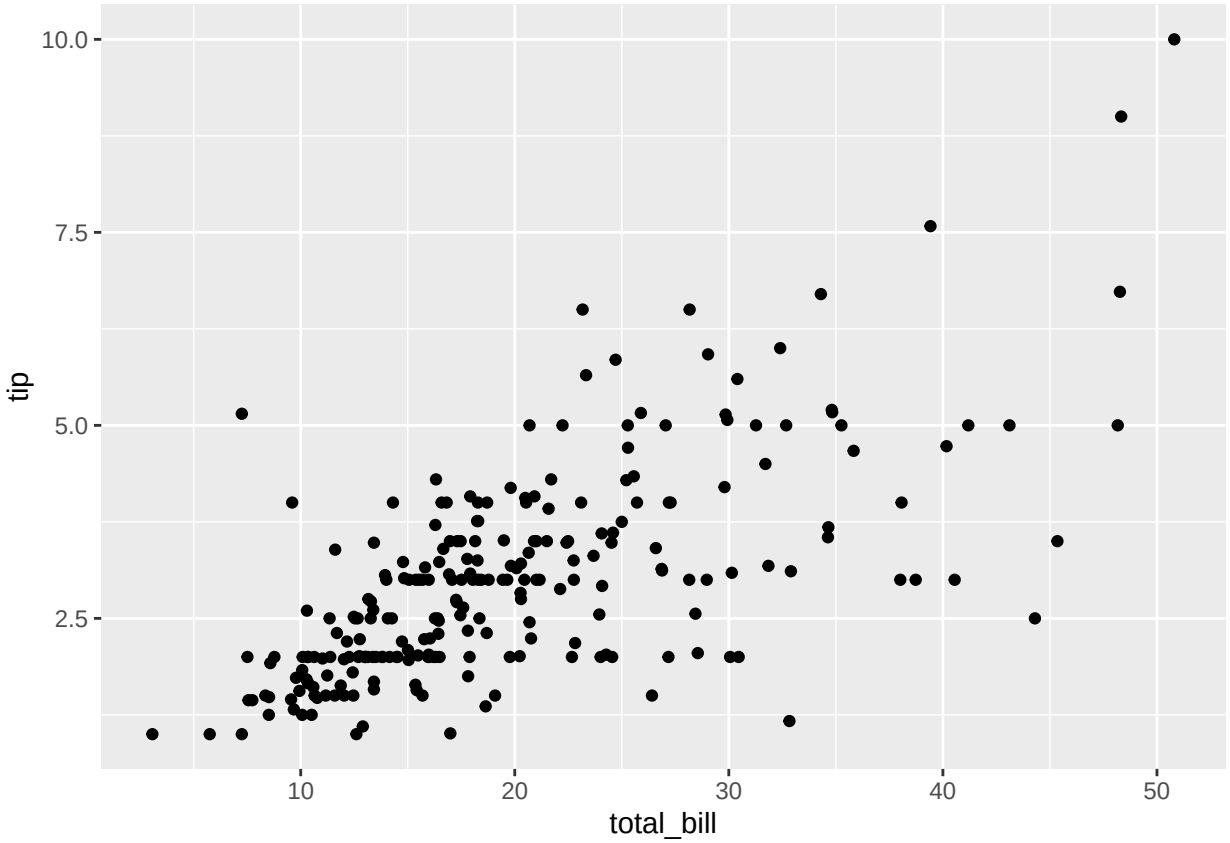

To change this, we need to tell it! How? We use the `aes()` argument which states for aesthetics.

```
ggplot(tips, aes(x = total_bill, y = tip))
```



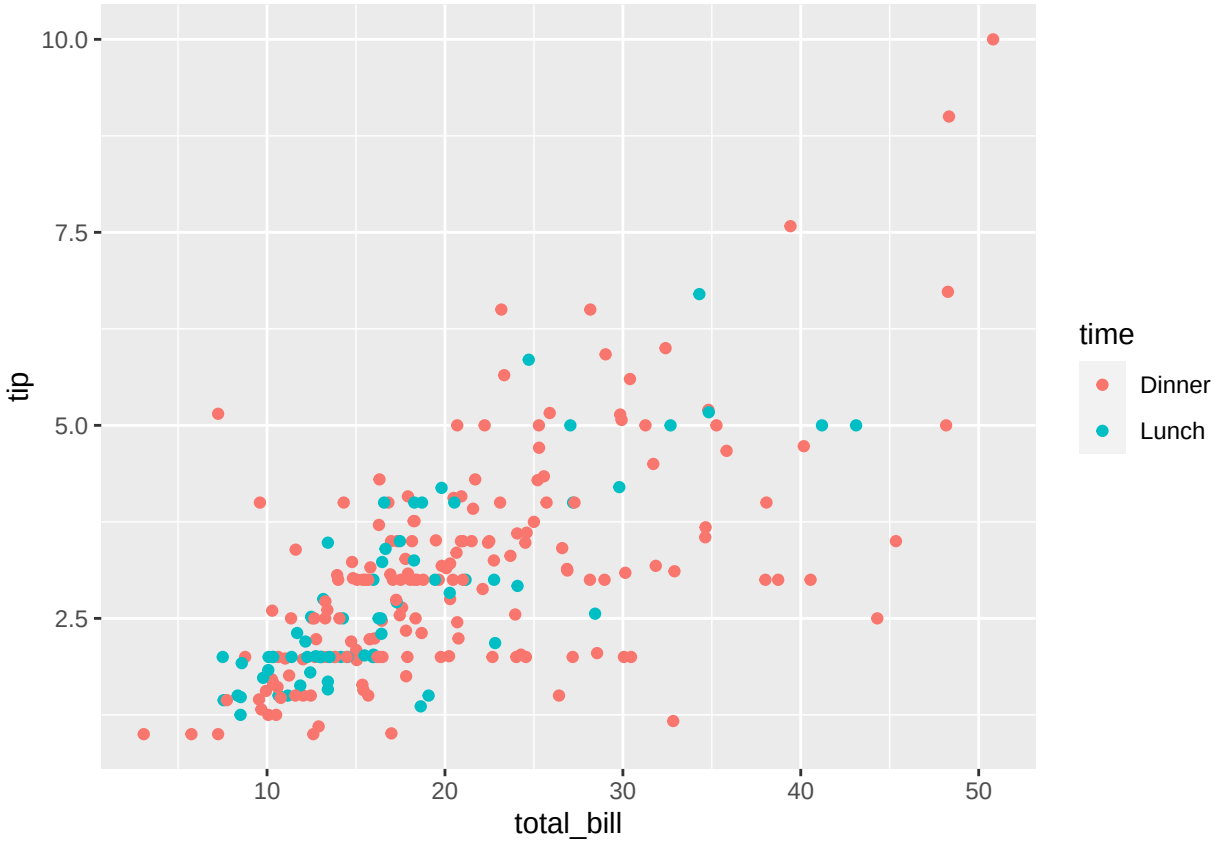
What do you now see?

```
ggplot(tips, aes(x = total_bill, y = tip)) +
  geom_point()
```

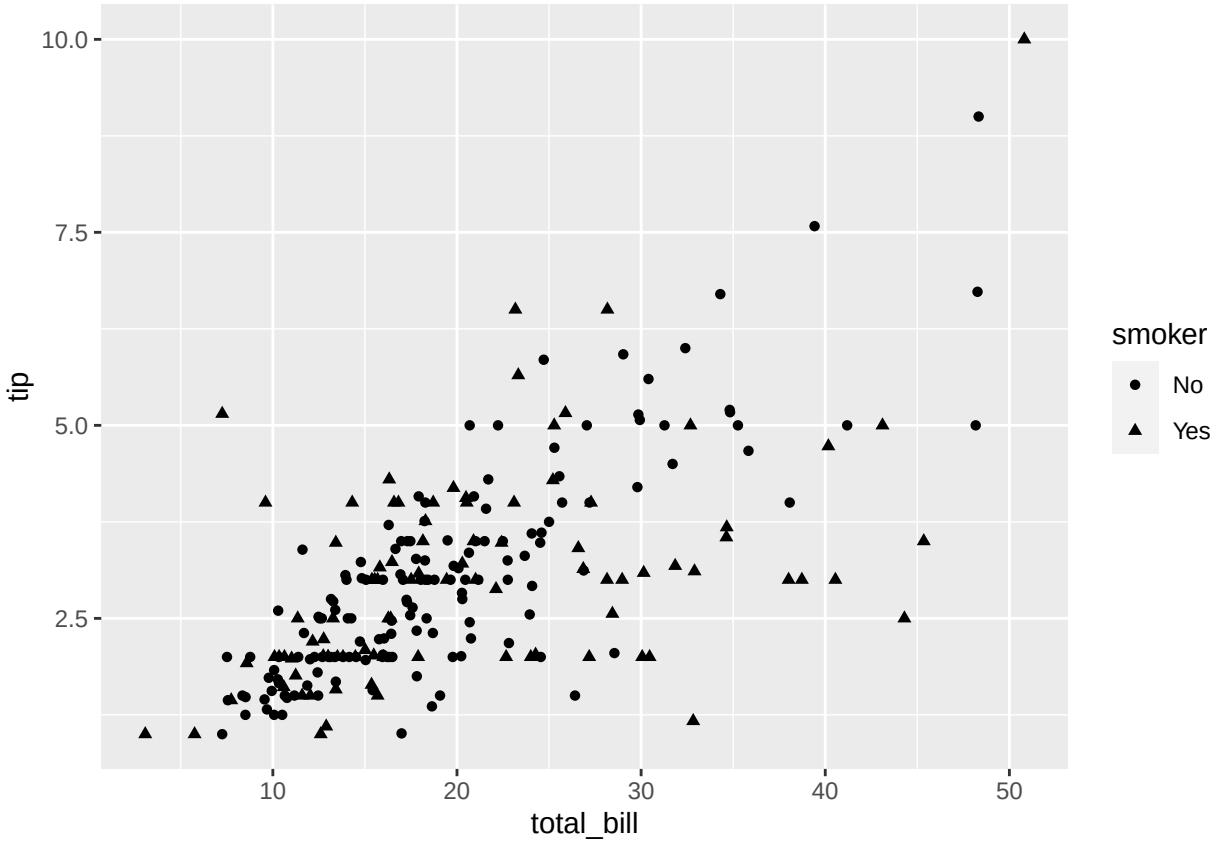


Now what if we want to keep modifying it? Let's dissect the code below.

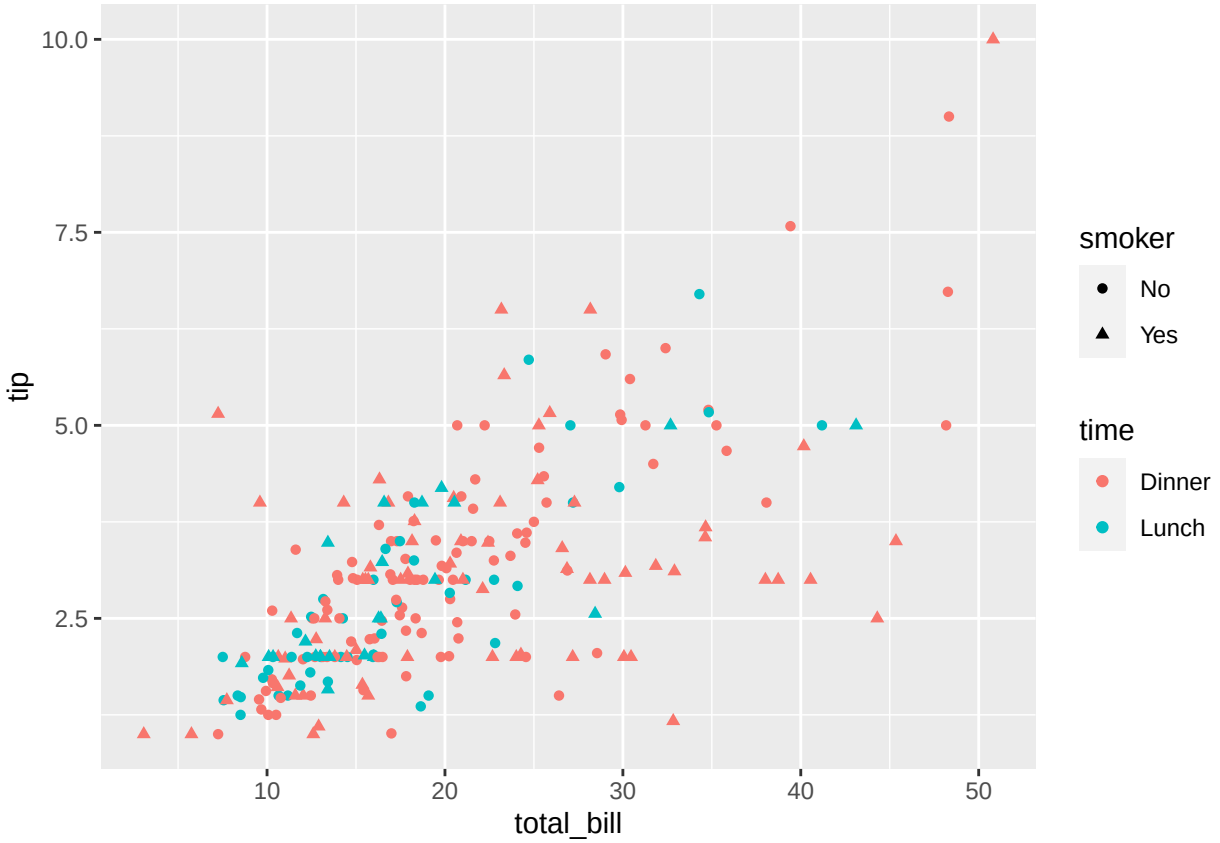
```
ggplot(tips, aes(x = total_bill, y = tip, color = time)) +
  geom_point()
```



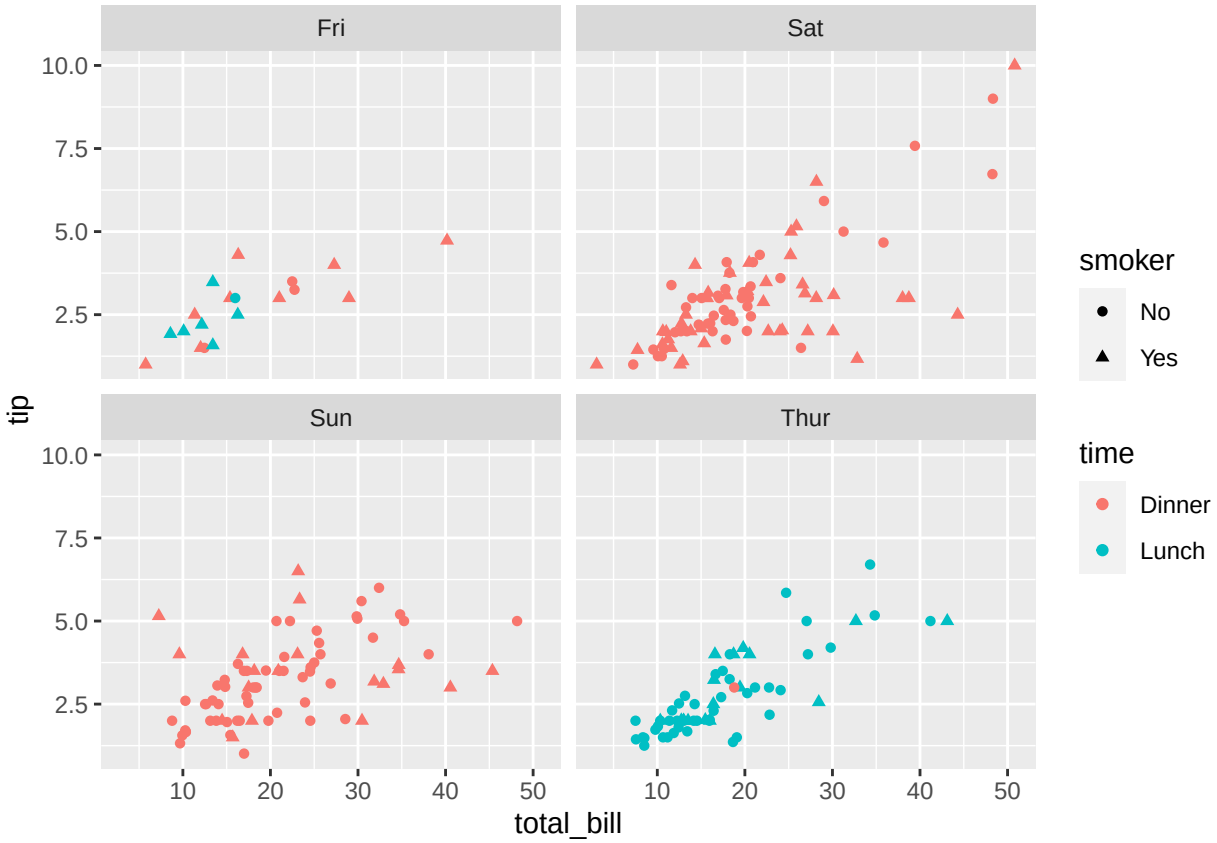

```
ggplot(tips, aes(x = total_bill, y = tip, shape = smoker)) +
  geom_point()
```




```
ggplot(tips, aes(x = total_bill, y = tip, color = time, shape = smoker)) +
  geom_point()
```

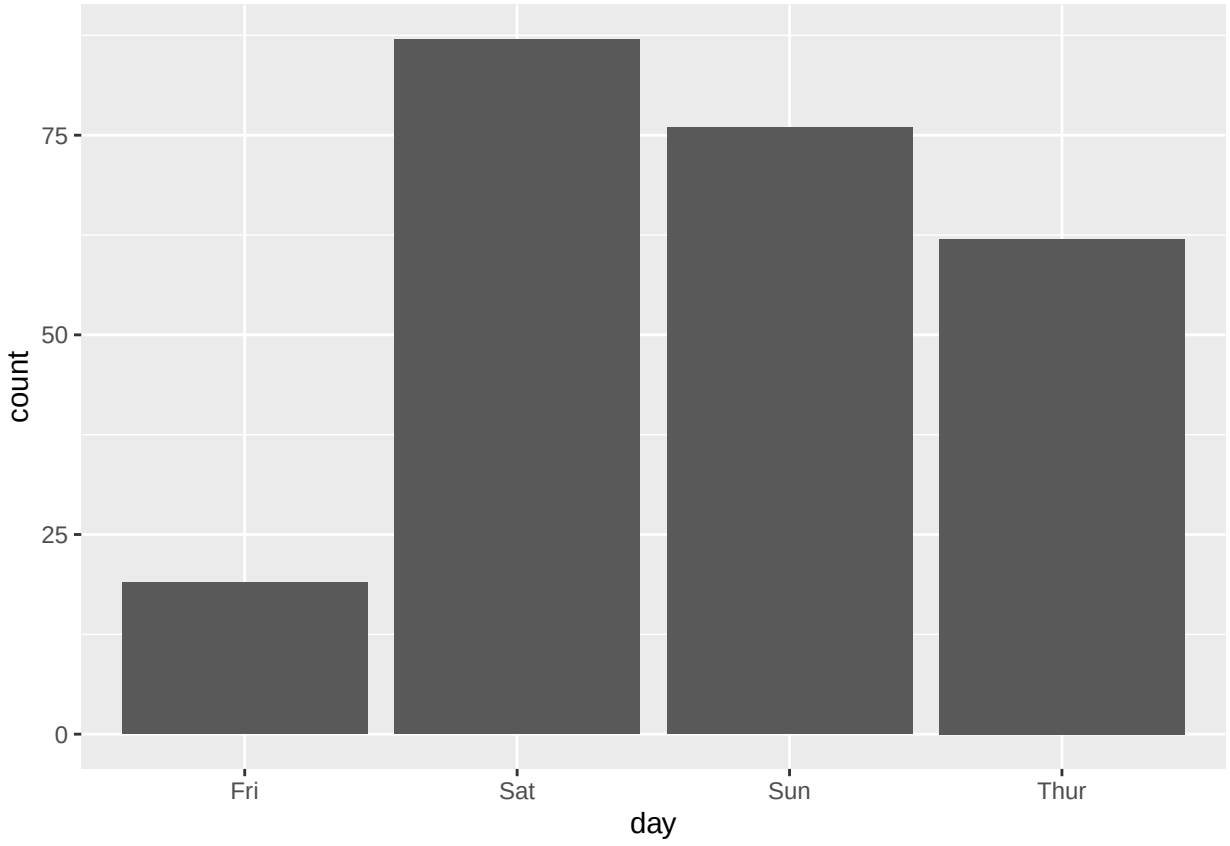



```
ggplot(tips, aes(x = total_bill, y = tip, color = time, shape = smoker)) +
  geom_point() +
  facet_wrap(~day)
```

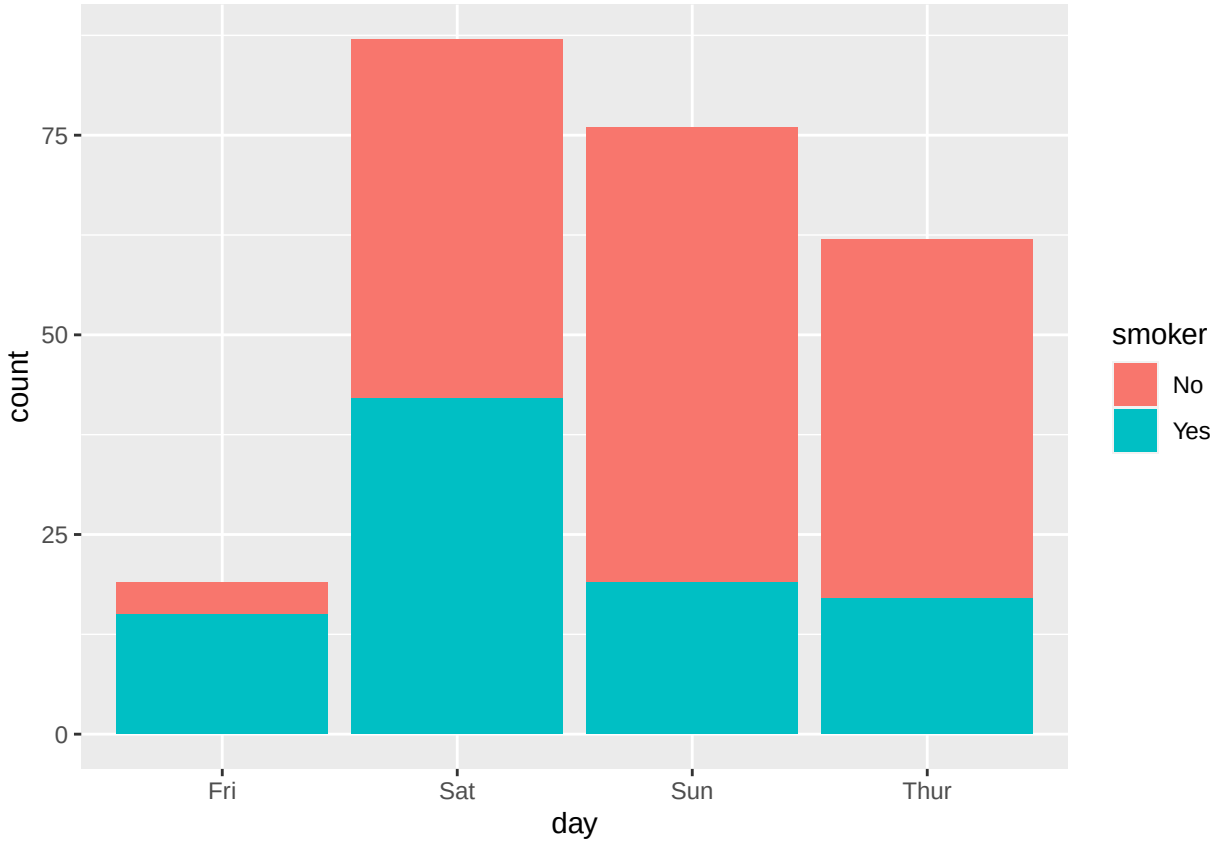


Let's dive into the data a bit more and make a barplot comparing how many counts we have of each day! What is similar about this plot compared to the ones before? What is different?

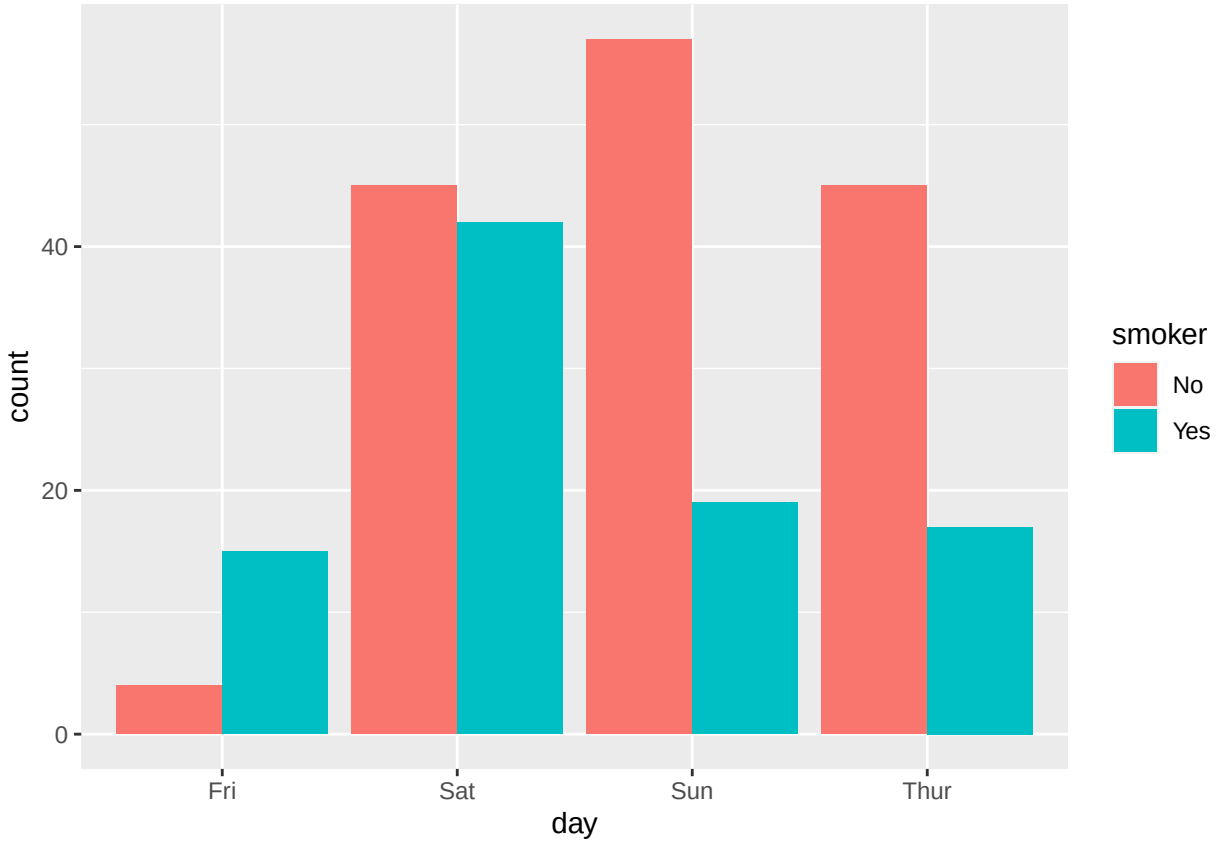
```
ggplot(tips, aes(x = day)) +
  geom_bar()
```



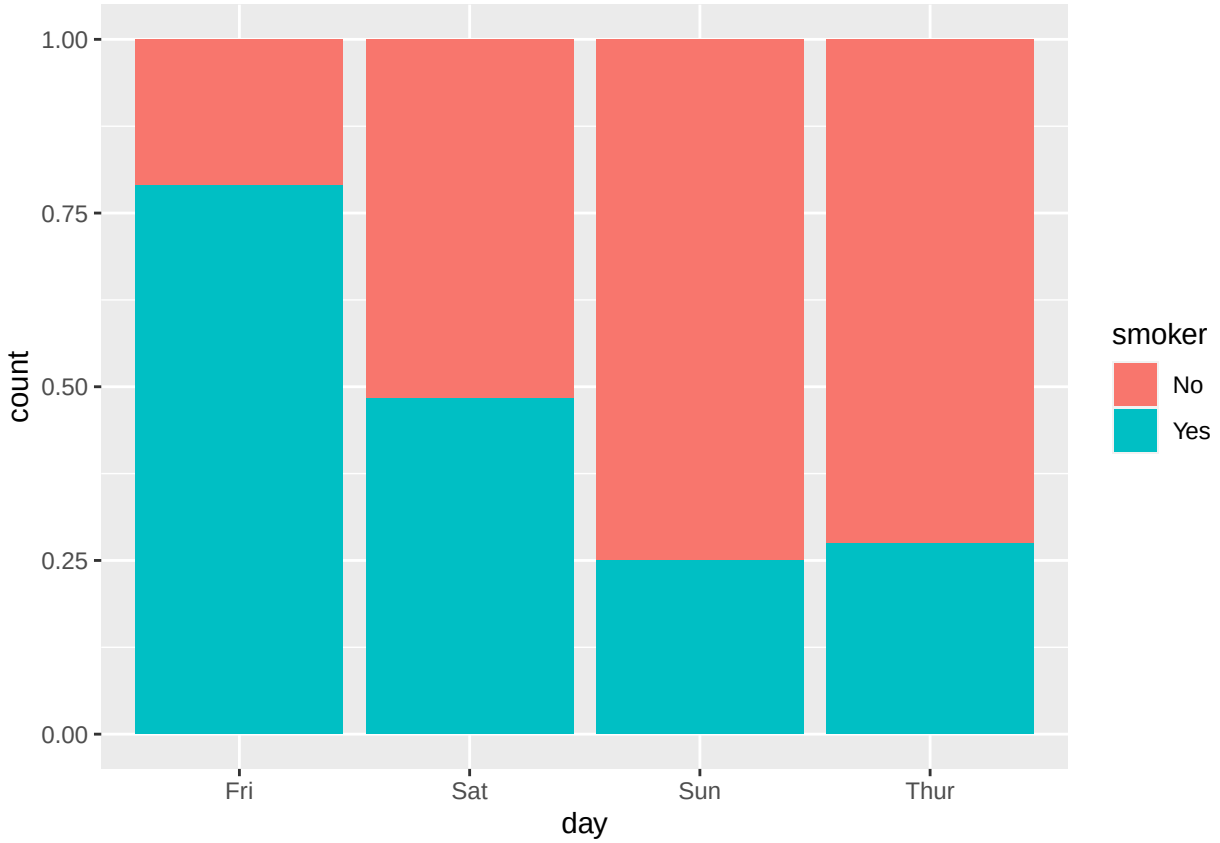

```
ggplot(tips, aes(x = day, fill= smoker)) +
  geom_bar()
```




```
ggplot(tips, aes(x = day, fill= smoker)) +
  geom_bar(position = "dodge")
```

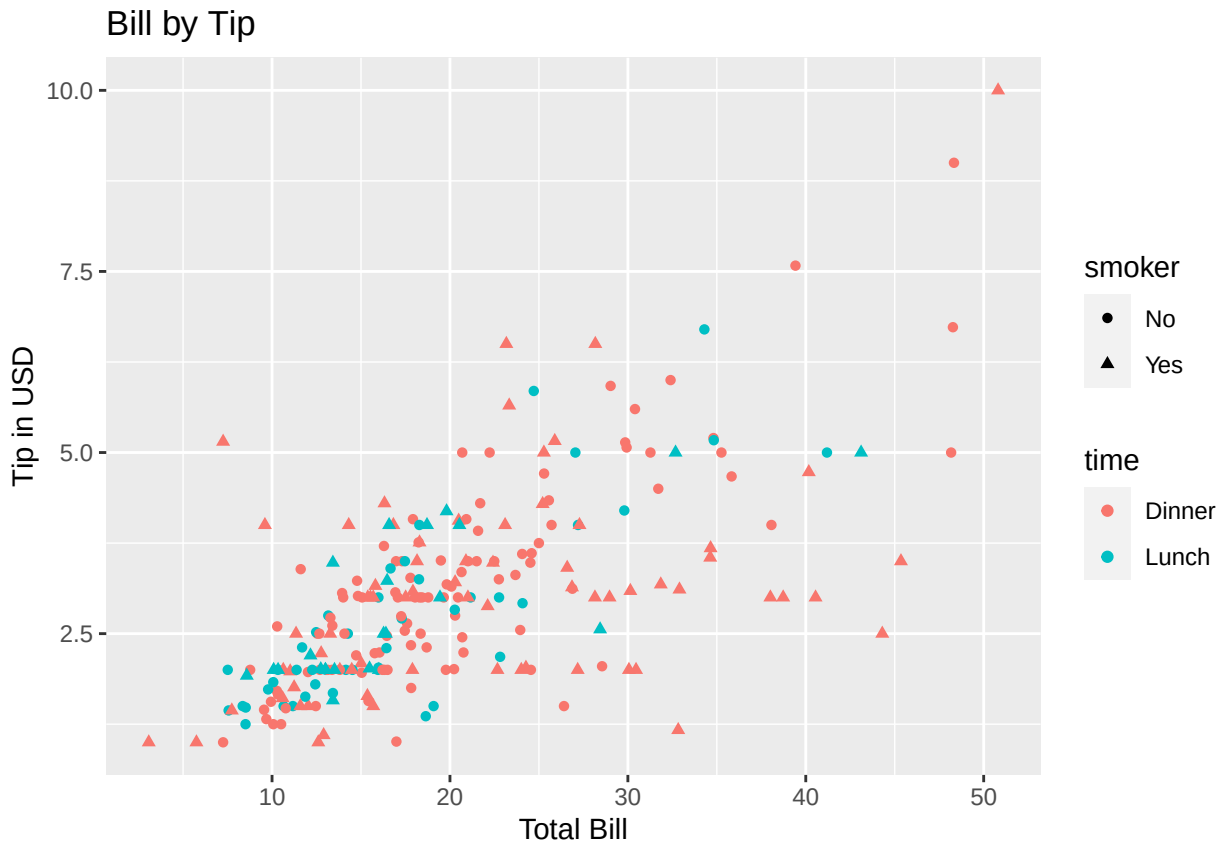



```
ggplot(tips, aes(x = day, fill= smoker)) +
  geom_bar(position = "fill")
```

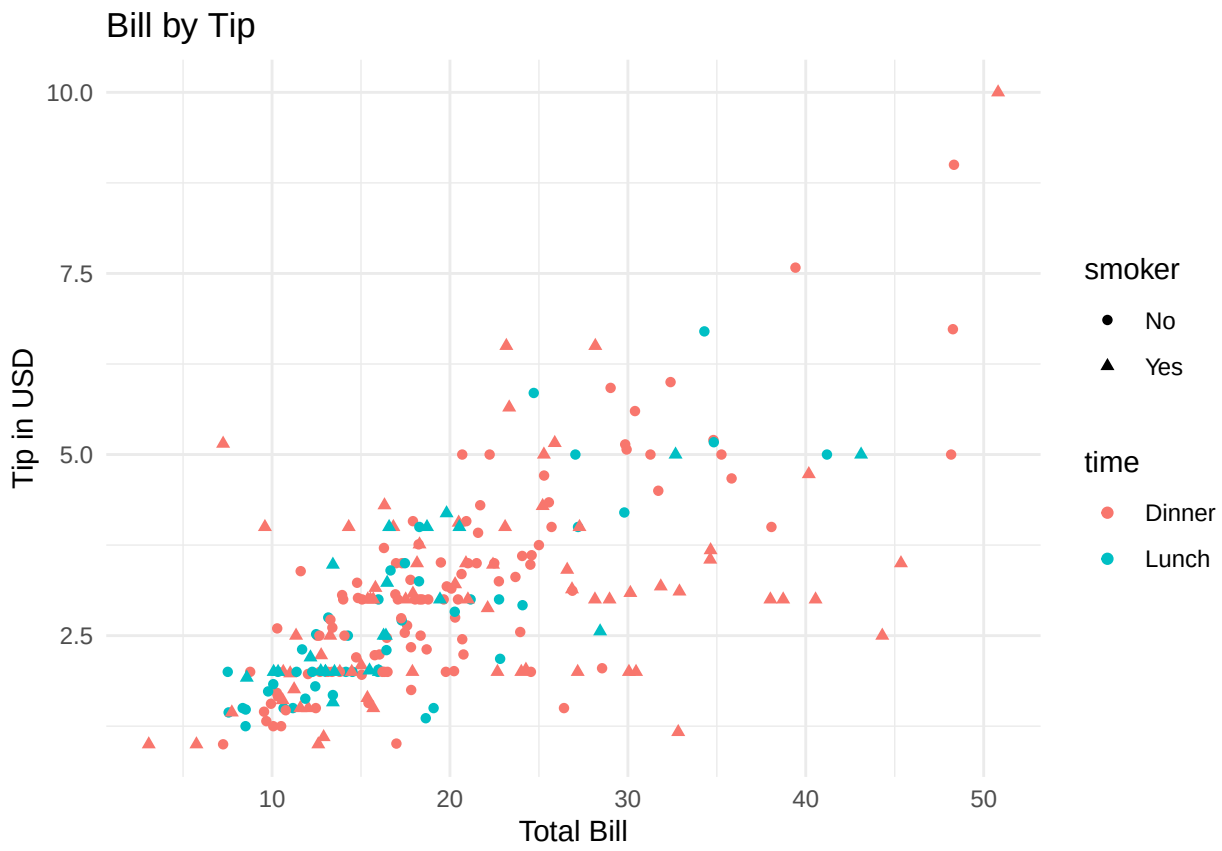


Let's now take a plot from above that we liked and begin to modify it!

```
ggplot(tips, aes(x = total_bill, y = tip, color = time, shape = smoker)) +
  geom_point() +
  labs(title = "Bill by Tip", x = "Total Bill", y = "Tip in USD")
```




```
ggplot(tips, aes(x = total_bill, y = tip, color = time, shape = smoker)) +
  geom_point() +
  labs(title = "Bill by Tip", x = "Total Bill", y = "Tip in USD") +
  theme_minimal()
```


Bill by Tip
10.0
7.5
5.0
2.5
Tip in USD
10
20
30
40
50
Total Bill
smoker
No
Yes
time
Dinner
Lunch