Introduction to Seaborn

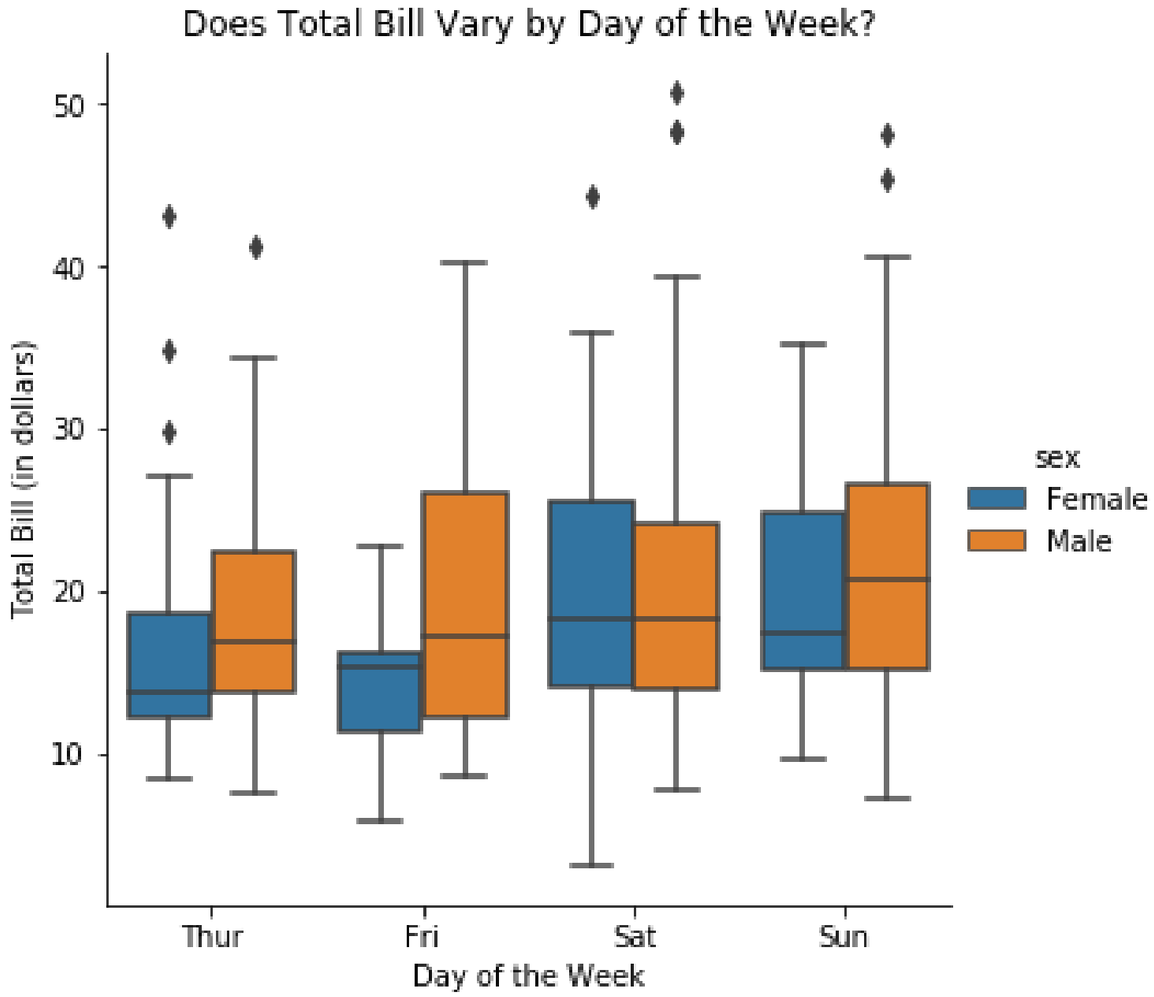
INTRODUCTION TO DATA VISUALIZATION WITH SEABORN

Erin Case
Data Scientist

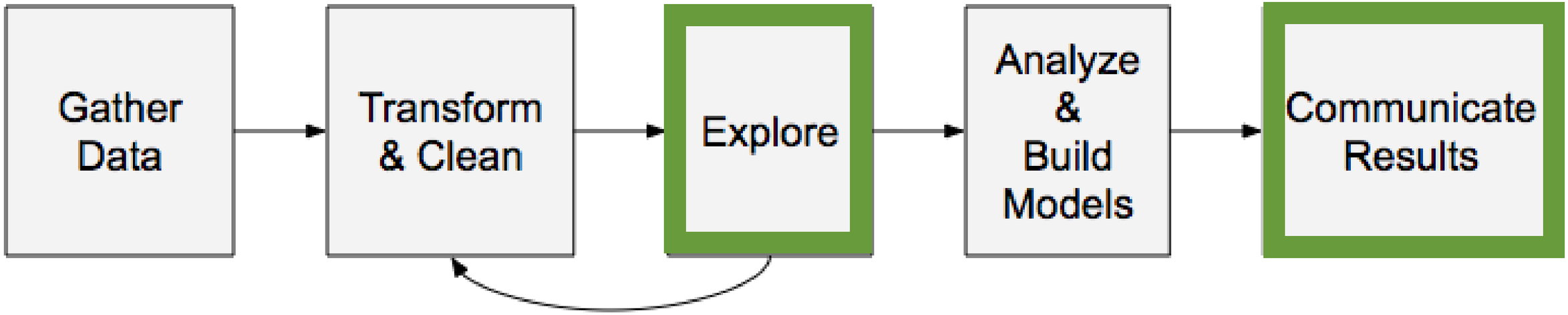
datacamp

What is Seaborn?

- Python data visualization library
- Easily create the most common types of plots



Why is Seaborn useful?



Advantages of Seaborn

- Easy to use
- Works well with `pandas` data structures
- Built on top of `matplotlib`

Getting started

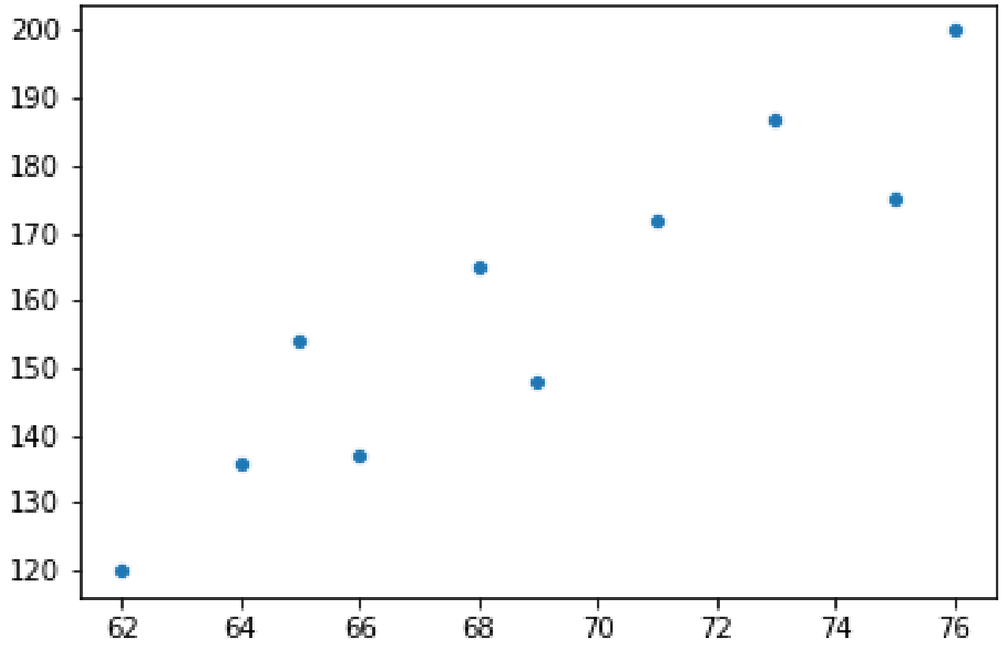
```
import seaborn as sns
import matplotlib.pyplot as plt
```

Samuel **N**orman **S**eaborn (`sns`)

- "The West Wing" television show

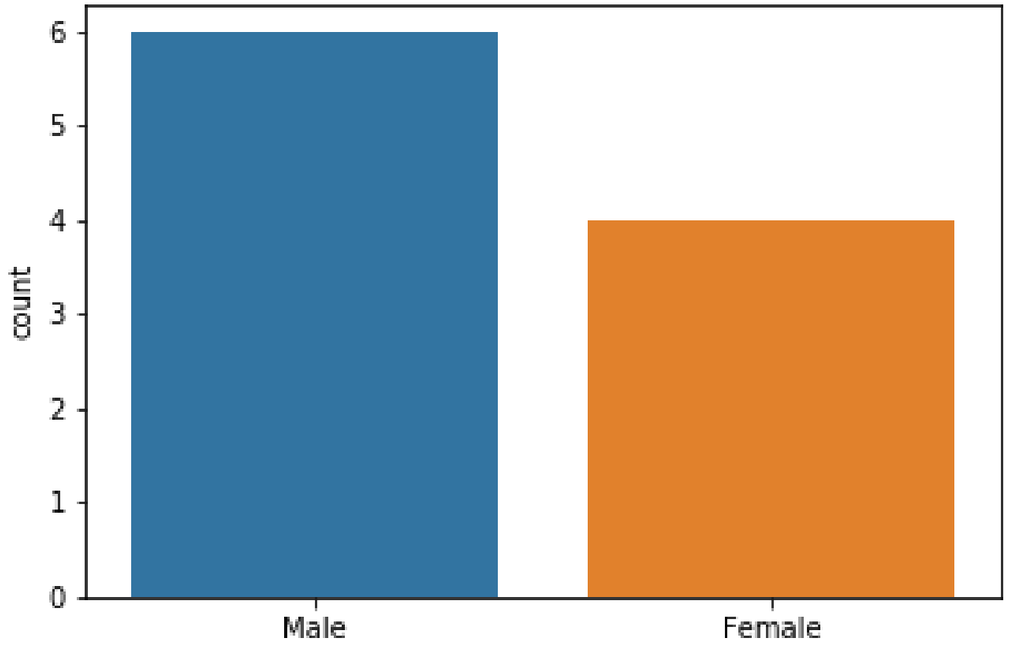
Example 1: Scatter plot

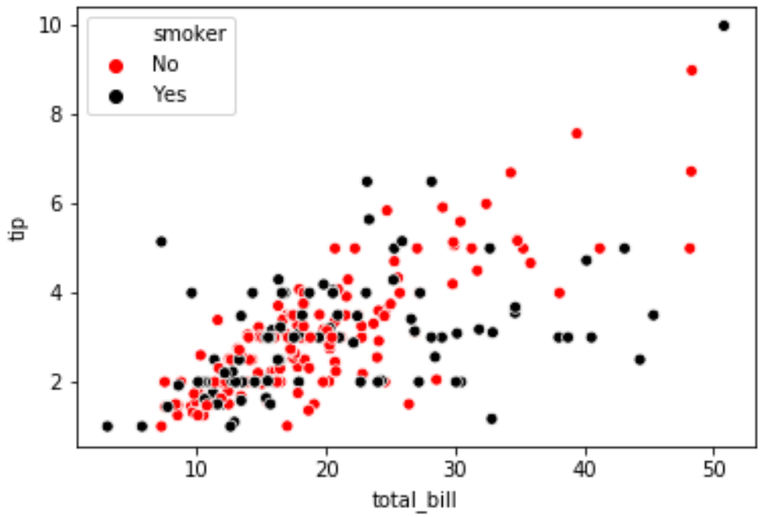
```
import seaborn as sns
import matplotlib.pyplot as plt
height = [62, 64, 69, 75, 66,
          68, 65, 71, 76, 73]
weight = [120, 136, 148, 175, 137,
          165, 154, 172, 200, 187]
sns.scatterplot(x=height, y=weight)
plt.show()
```

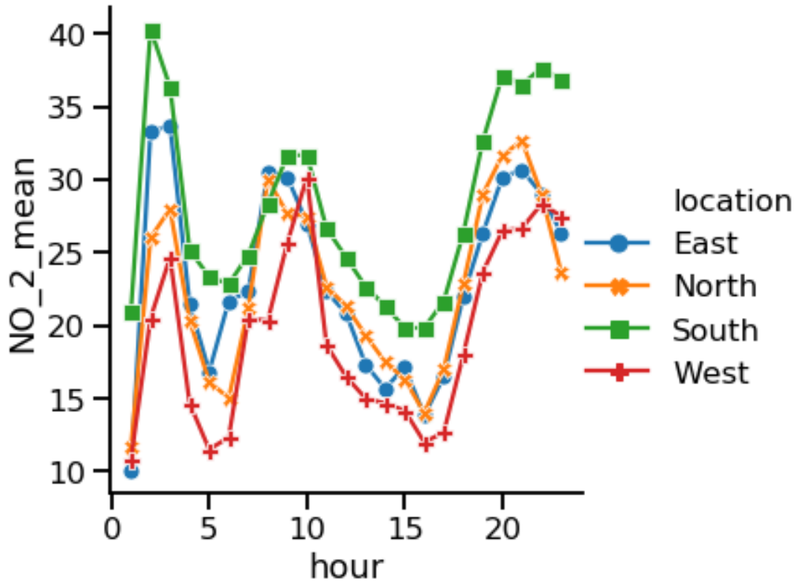


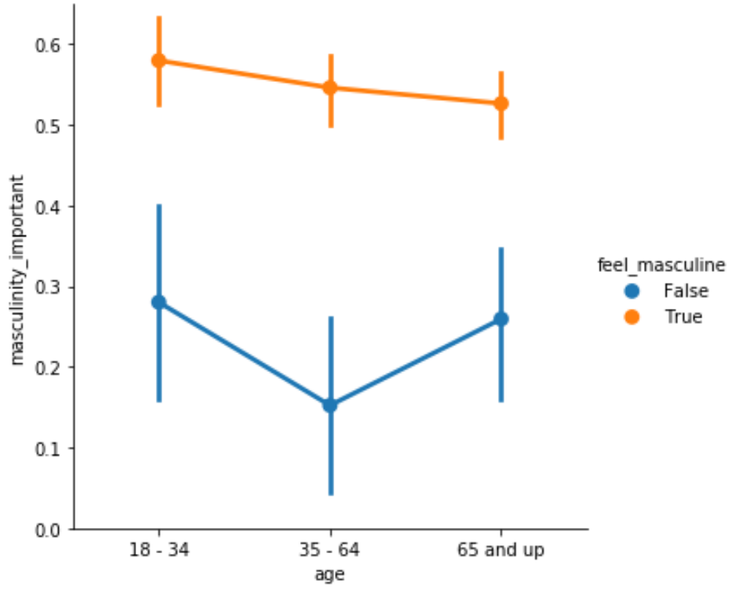
Example 2: Create a count plot

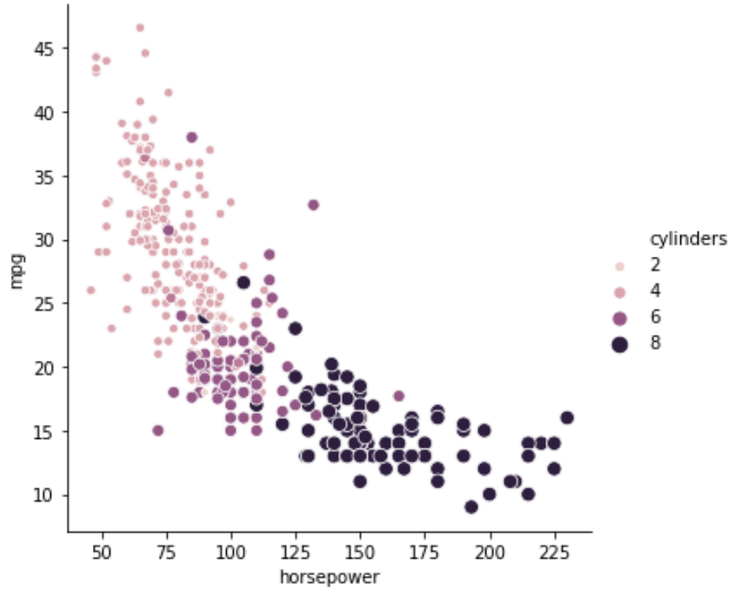
```
import seaborn as sns
import matplotlib.pyplot as plt
gender = ["Female", "Female",
          "Female", "Female",
          "Male", "Male", "Male",
          "Male", "Male", "Male"]
sns.countplot(x=gender)
plt.show()
```




smoker
No
Yes
tip
total_bill
10
8
6
4
2
10
20
30
40
50


location
East
North
South
West
NO_2_mean
hour
40
35
30
25
20
15
10
0
5
10
15
20


feel_masculine
False
True
masculinity_important
age
18 - 34
35 - 64
65 and up
0.6
0.5
0.4
0.3
0.2
0.1
0.0


cylinders
2
4
6
8
mpg
horsepower
45
40
35
30
25
20
15
10
50
75
100
125
150
175
200
225

Let's practice!

INTRODUCTION TO DATA VISUALIZATION WITH SEABORN

Using pandas with Seaborn

INTRODUCTION TO DATA VISUALIZATION WITH SEABORN

Erin Case
Data Scientist

datacamp

What is pandas?

- Python library for data analysis
- Easily read datasets from csv, txt, and other types of files
- Datasets take the form of `DataFrame` objects

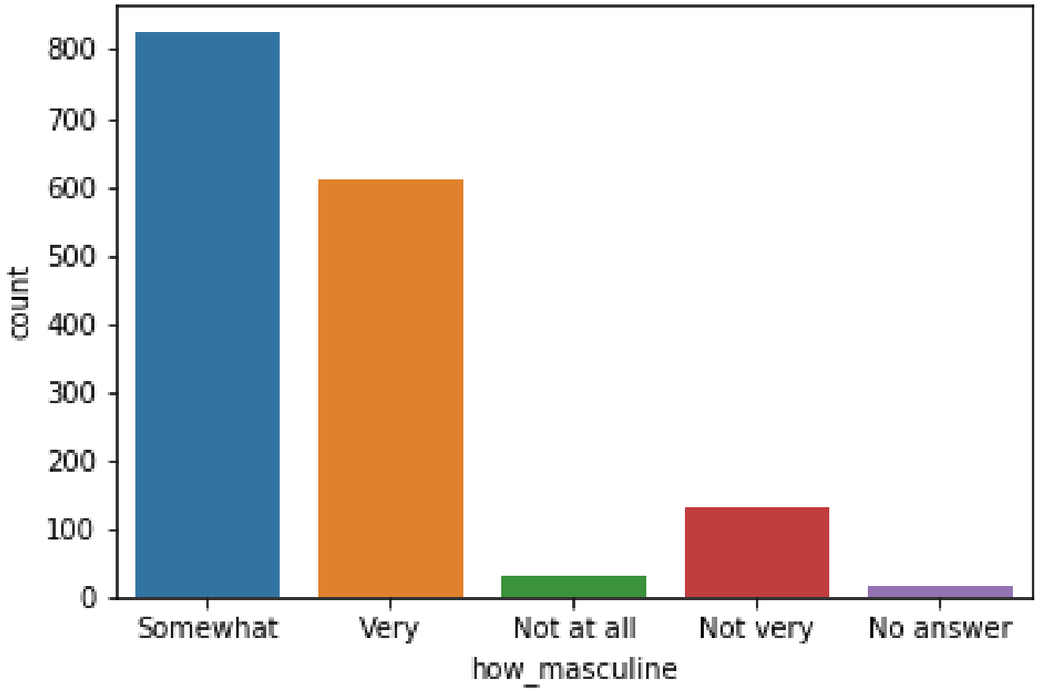
Working with DataFrames

```
import pandas as pd
df = pd.read_csv("masculinity.csv")
df.head()
```

```
   participant_id       age   how_masculine   how_important
0               1  18 - 34        Somewhat         Somewhat
1               2  18 - 34        Somewhat         Somewhat
2               3  18 - 34            Very         Not very
3               4  18 - 34            Very         Not very
4               5  18 - 34            Very             Very
```

Using DataFrames with countplot()

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
df = pd.read_csv("masculinity.csv")
sns.countplot(x="how_masculine",
              data=df)
plt.show()
```



	participant_id	age	how_masculine	how_important
0	1	18 - 34	Somewhat	Somewhat
1	2	18 - 34	Somewhat	Somewhat
2	3	18 - 34	Very	Not very
3	4	18 - 34	Very	Not very
4	5	18 - 34	Very	Very
5	6	18 - 34	Very	Somewhat
6	7	18 - 34	Somewhat	Not very
7	8	18 - 34	Somewhat	Somewhat
8	9	18 - 34	Very	Not at all
9	10	18 - 34	Somewhat	Somewhat

	AMONG ADULT MEN	Unnamed: 1	Adult Men	Age	Unnamed: 4	Unnamed: 5
0				18 - 34	35 - 64	65 and up
1	In general, how masculine or "manly" do you feel?					
2		Very masculine	37%	29%	42%	37%
3		Somewhat masculine	46%	47%	46%	47%
4		Not very masculine	11%	13%	9%	13%
5		Not at all masculine	5%	10%	2%	3%
6		No answer	1%	0%	1%	1%
7	How important is it to you that others see you as masculine?					
8		Very important	16%	18%	17%	13%
9		Somewhat important	37%	38%	37%	32%
10		Not too important	28%	18%	31%	37%
11		Not at all important	18%	26%	15%	18%
12		No answer	0%	0%	1%	0%

Let's practice!

INTRODUCTION TO DATA VISUALIZATION WITH SEABORN

Adding a third variable with hue

INTRODUCTION TO DATA VISUALIZATION WITH SEABORN

Erin Case
Data Scientist

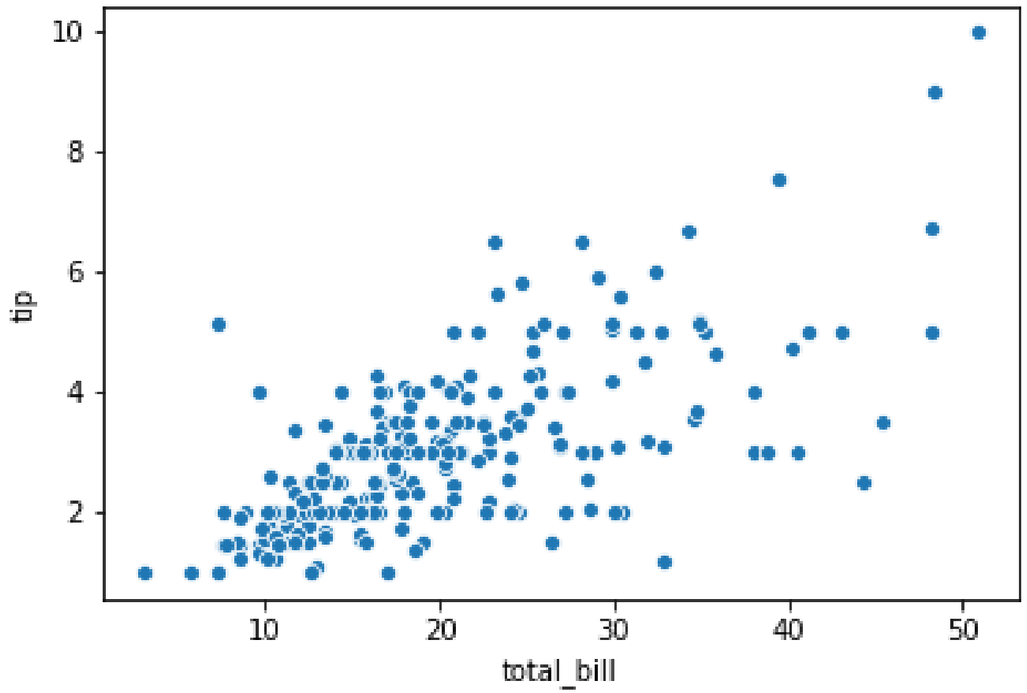
Tips dataset

```
import pandas as pd
import seaborn as sns
tips = sns.load_dataset("tips")
tips.head()
```

```
   total_bill   tip     sex smoker  day    time  size
0       16.99  1.01  Female     No  Sun  Dinner     2
1       10.34  1.66    Male     No  Sun  Dinner     3
2       21.01  3.50    Male     No  Sun  Dinner     3
3       23.68  3.31    Male     No  Sun  Dinner     2
4       24.59  3.61  Female     No  Sun  Dinner     4
```

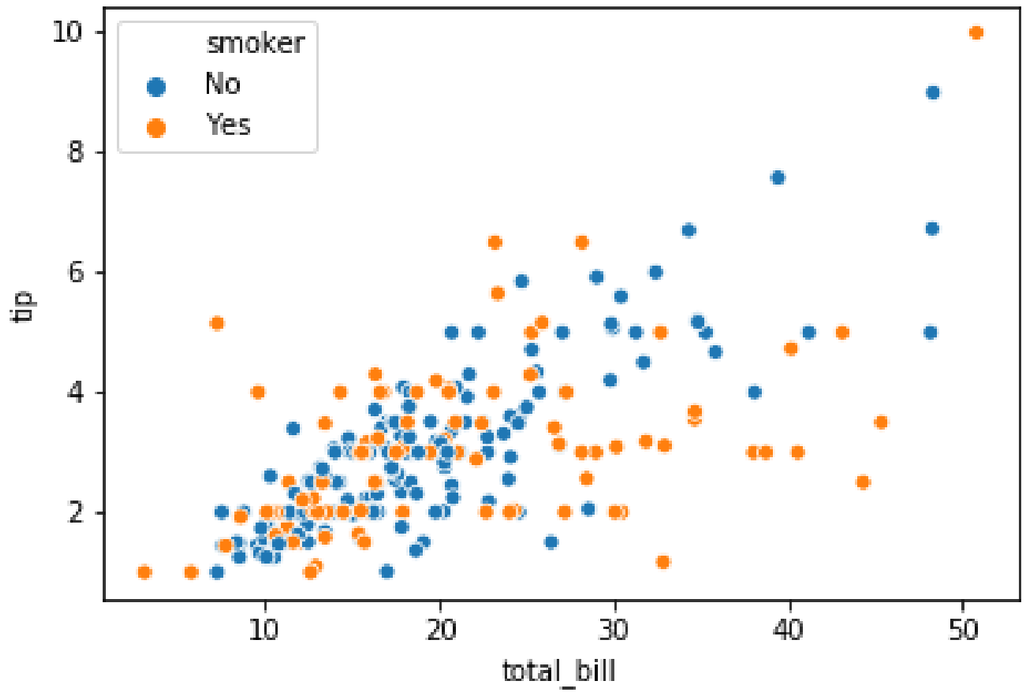
A basic scatter plot

```
import matplotlib.pyplot as plt
import seaborn as sns

sns.scatterplot(x="total_bill",
                y="tip",
                data=tips)

plt.show()
```



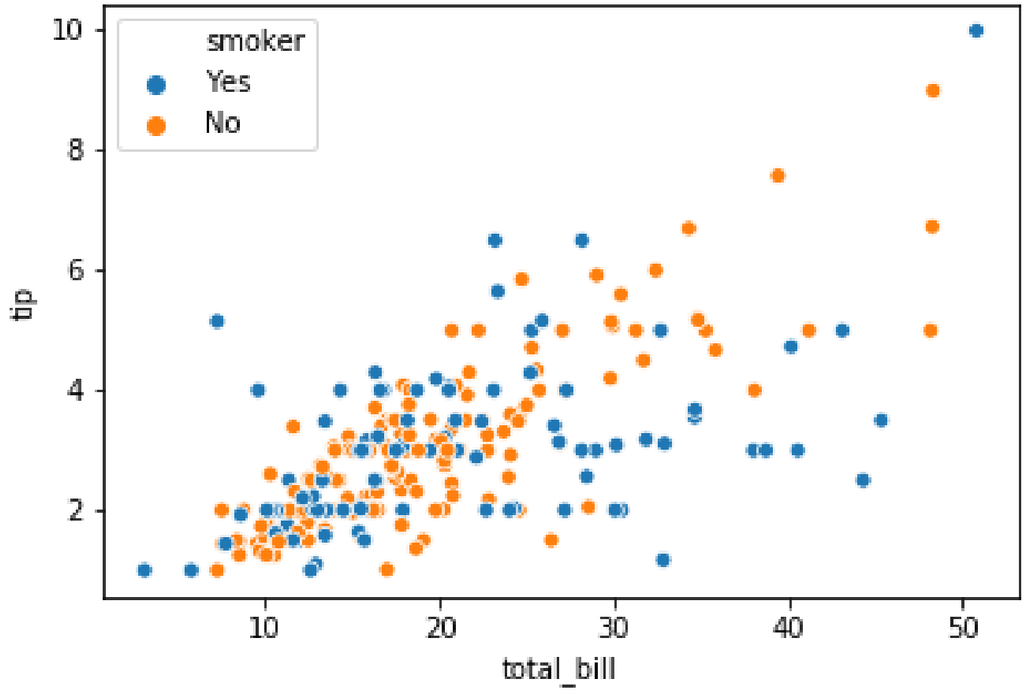
A scatter plot with hue

```
import matplotlib.pyplot as plt
import seaborn as sns

sns.scatterplot(x="total_bill",
                y="tip",
                data=tips,
                hue="smoker")
plt.show()
```



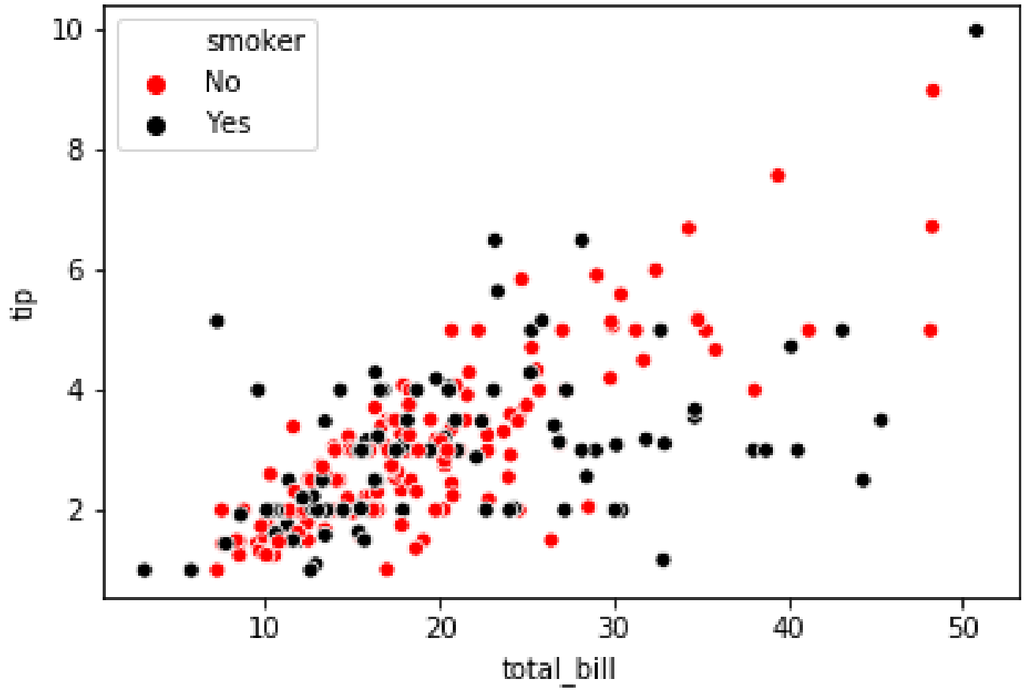
Setting hue order

```
import matplotlib.pyplot as plt
import seaborn as sns

sns.scatterplot(x="total_bill",
                y="tip",
                data=tips,
                hue="smoker",
                hue_order=["Yes",
                           "No"])
plt.show()
```



Specifying hue colors

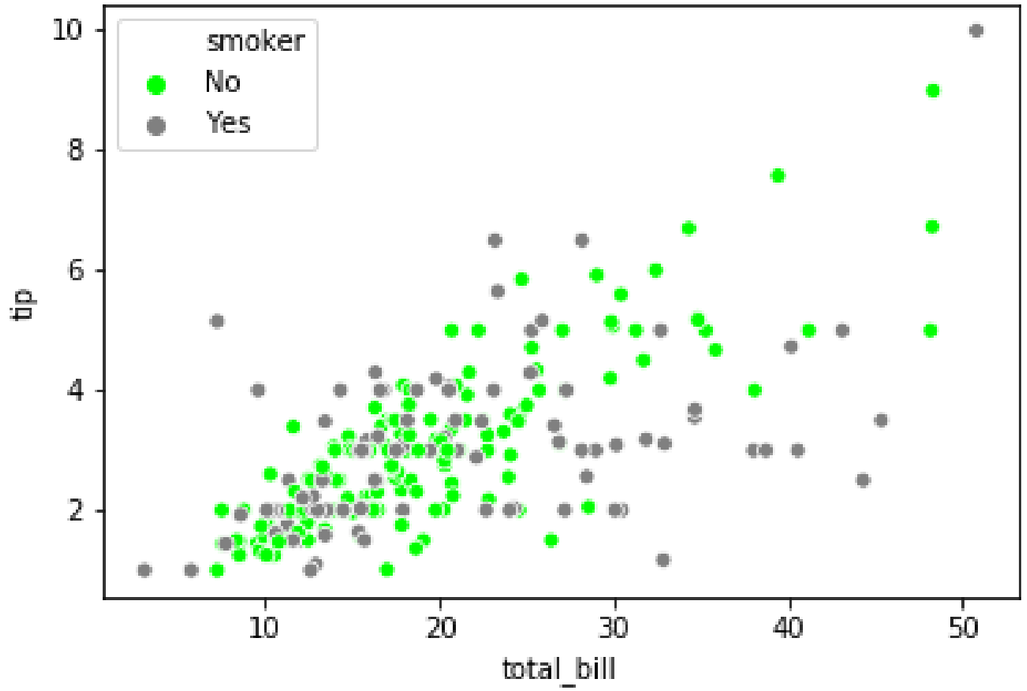
```
import matplotlib.pyplot as plt
import seaborn as sns
hue_colors = {"Yes": "black",
              "No": "red"}
sns.scatterplot(x="total_bill",
                y="tip",
                data=tips,
                hue="smoker",
                palette=hue_colors)

plt.show()
```



	Color	Matplotlib name	Matplotlib abbreviation	HTML color code (hex)
	blue	"blue"	"b"	#0000ff
	green	"green"	"g"	#008000
	red	"red"	"r"	#ff0000
	green/blue	"cyan"	"c"	#00bfbf
	purple	"magenta"	"m"	#bf00bf
	yellow	"yellow"	"y"	#bfbf00
	black	"black"	"k"	#000000
	white	"white"	"w"	#ffffff

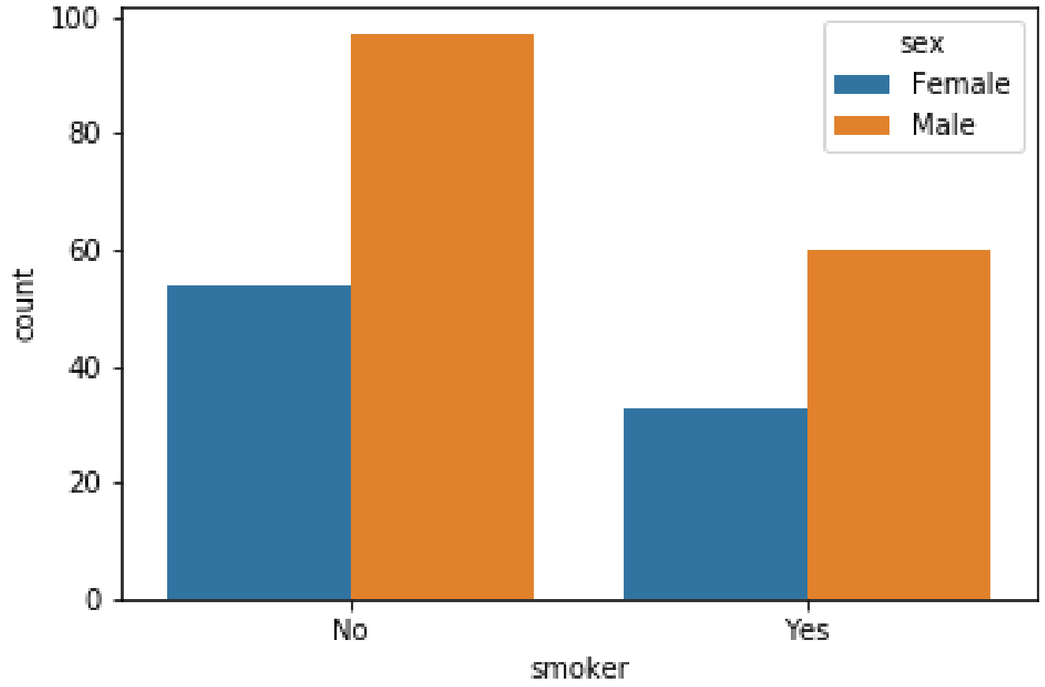
Using HTML hex color codes with hue

```
import matplotlib.pyplot as plt
import seaborn as sns

hue_colors = {"Yes": "#808080",
              "No": "#00FF00"}

sns.scatterplot(x="total_bill",
                y="tip",
                data=tips,
                hue="smoker",
                palette=hue_colors)
plt.show()
```



Using hue with count plots

```
import matplotlib.pyplot as plt
import seaborn as sns

sns.countplot(x="smoker",
              data=tips,
              hue="sex")
plt.show()
```



Let's practice!

INTRODUCTION TO DATA VISUALIZATION WITH SEABORN