

Introduction to Seaborn

INTRODUCTION TO DATA VISUALIZATION WITH SEABORN



Erin Case
Data Scientist

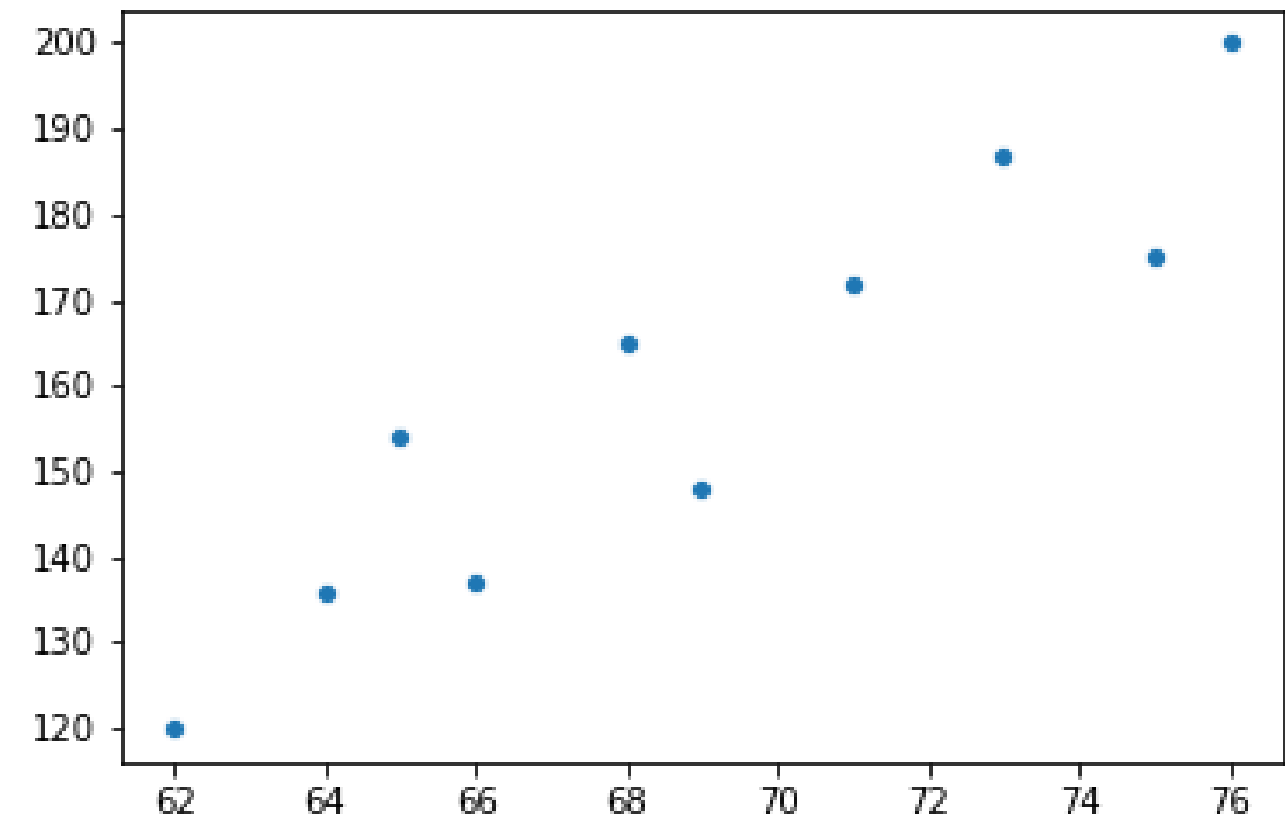
Scatter plot

```
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]

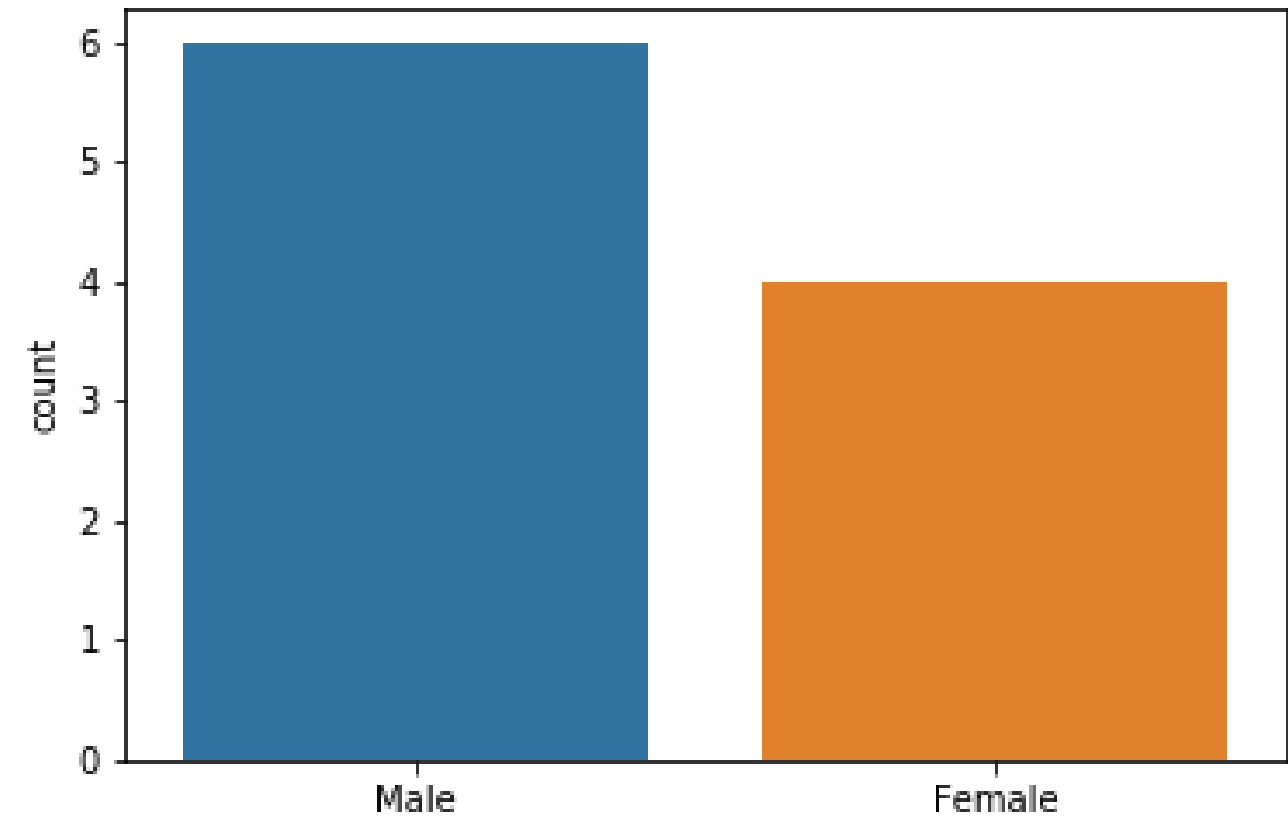
sns.scatterplot(x=height, y=weight)
plt.show()
```

: 가 가 가 .



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()
```



Exercise 1: scatterplot ,countplot

Percent_literate:

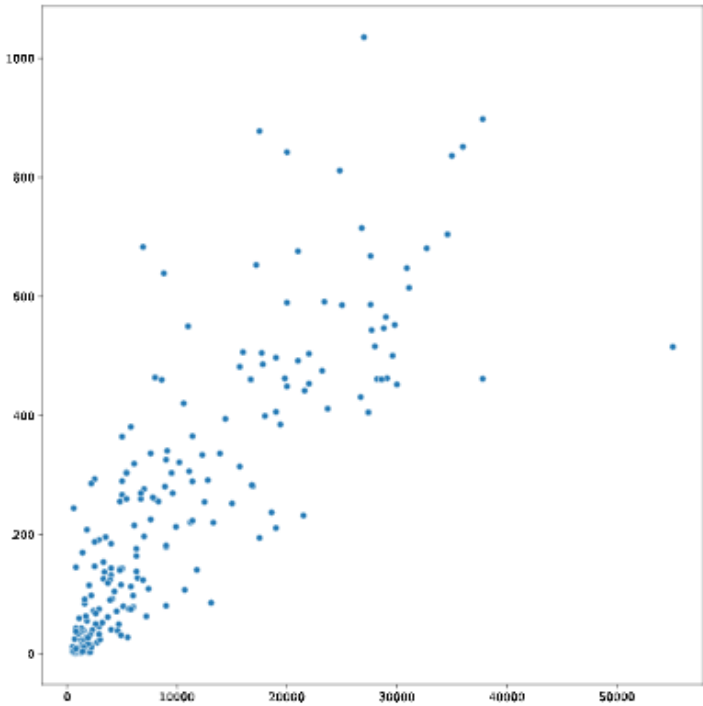
GDP: dollars per person

phones: number of mobile phones per 1,000 people in that country.

script.py

Light

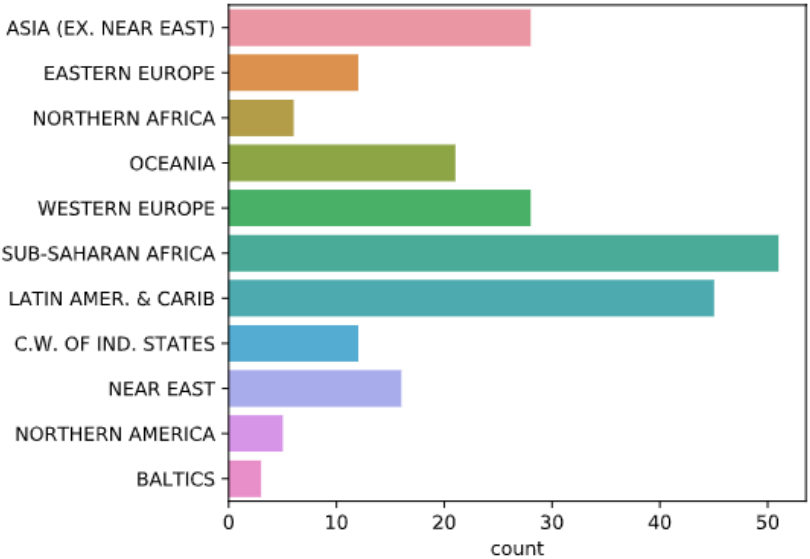
```
1 # Import Matplotlib and Seaborn
2 import matplotlib.pyplot as plt
3 import seaborn as sns
4
5 # Change this scatter plot to have percent literate on the y-axis
6 sns.scatterplot(x=gdp, y=percent_literate)
7
8 # Show plot
9 plt.show()
```



script.py

Light Mode

```
1 # Import Matplotlib and Seaborn
2 import matplotlib.pyplot as plt
3 import seaborn as sns
4
5 # Create count plot with region on the y-axis
6 sns.countplot(y=region)
7
8 # Show plot
9 plt.show()
```



Using pandas with Seaborn

INTRODUCTION TO DATA VISUALIZATION WITH SEABORN



Erin Case
Data Scientist

Working with DataFrames

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

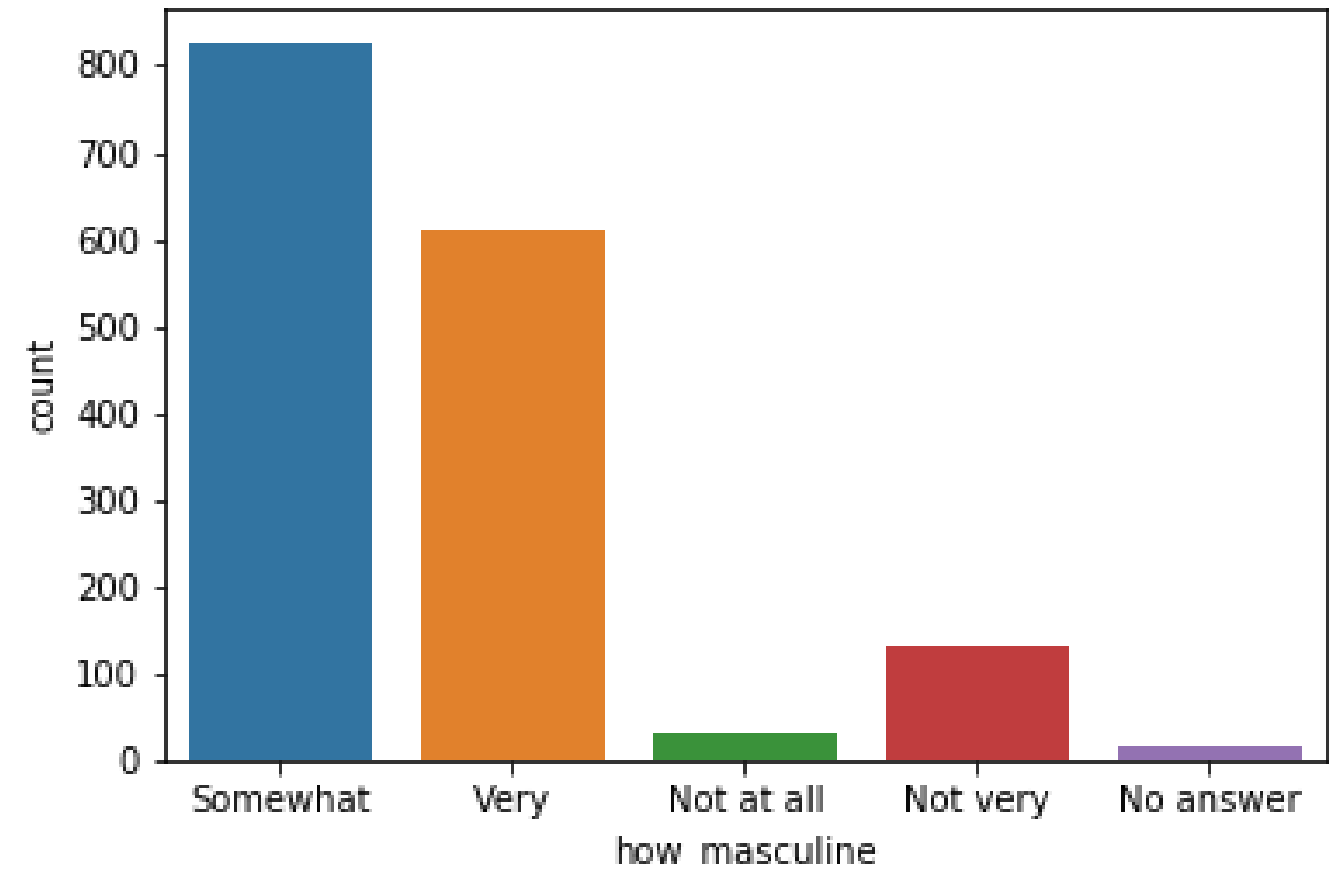
	participant_id	age	how_masculine	how_important
0	1	18 - 34	Somewhat	Somewhat
1	2	18 - 34	Somewhat	Somewhat
2	3	18 - 34	Very	Not very
3	4	18 - 34	Very	Not very
4	5	18 - 34	Very	Very

3	:	?	
4	:		?

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()
```



: Seaborn dataframe data
tidy (,)

Example 2 : 가

script.py

```
1 # Import Pandas
2 import pandas as pd
3
4 # Create a DataFrame from csv file
5 df = pd.read_csv(csv_filepath)
6
7 # Print the head of df
8 print(df.head())
```

<script.py> output:

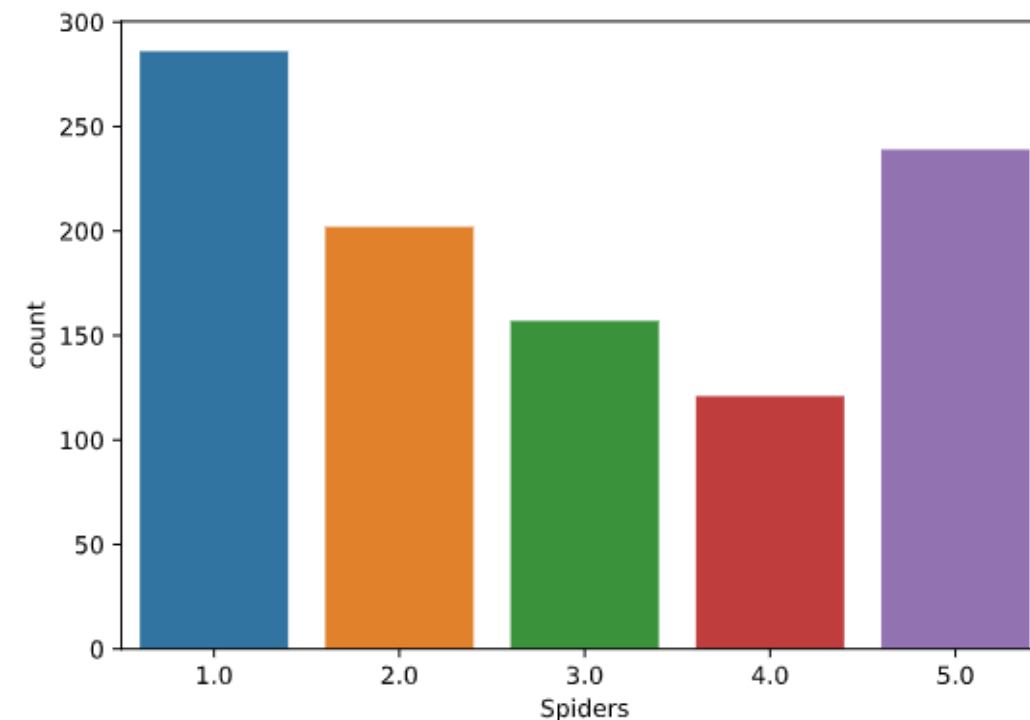
	Unnamed: 0	How old are you?
0	Marion	12
1	Elroy	16
2	NaN	What is your favorite animal?
3	Marion	dog
4	Elroy	cat

: untidy . [2]

Example 3 : How many young people repot being scared of spiders? (1~5 , 가 .)

script.py

```
1 # Import Matplotlib, Pandas, and Seaborn
2 import matplotlib.pyplot as plt
3 import seaborn as sns
4 import pandas as pd
5
6 # Create a DataFrame from csv file
7 df = pd.read_csv(csv_filepath)
8
9 # Create a count plot with "Spiders" on the x-axis
10 sns.countplot(x="Spiders", data=df)
11
12 # Display the plot
13 plt.show()
```



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()
```

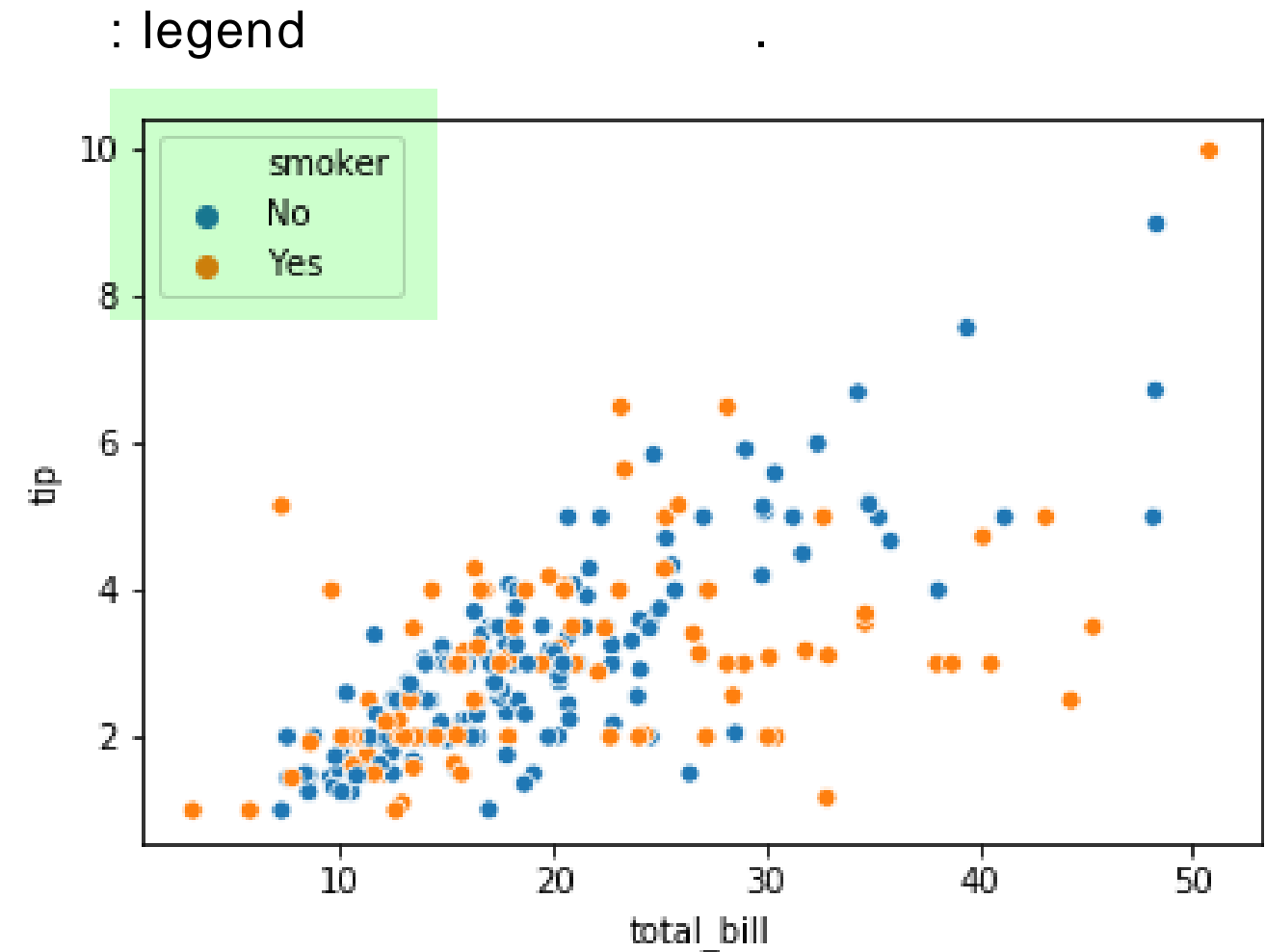
	total_bill	tip	sex	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3
2	21.01	3.50	Male	No	Sun	Dinner	3
3	23.68	3.31	Male	No	Sun	Dinner	2
4	24.59	3.61	Female	No	Sun	Dinner	4

A 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()
```

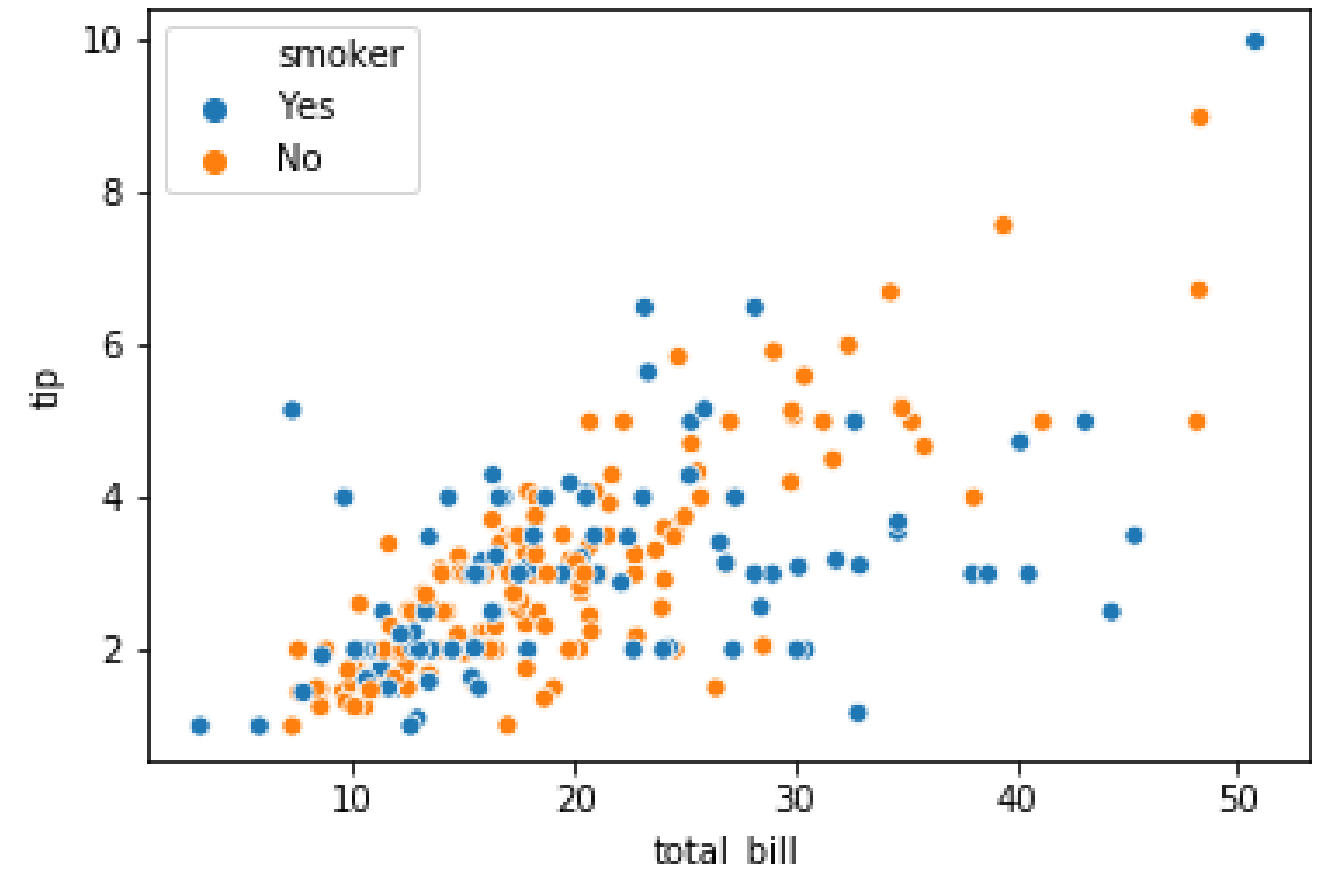


Setting hue order

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

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

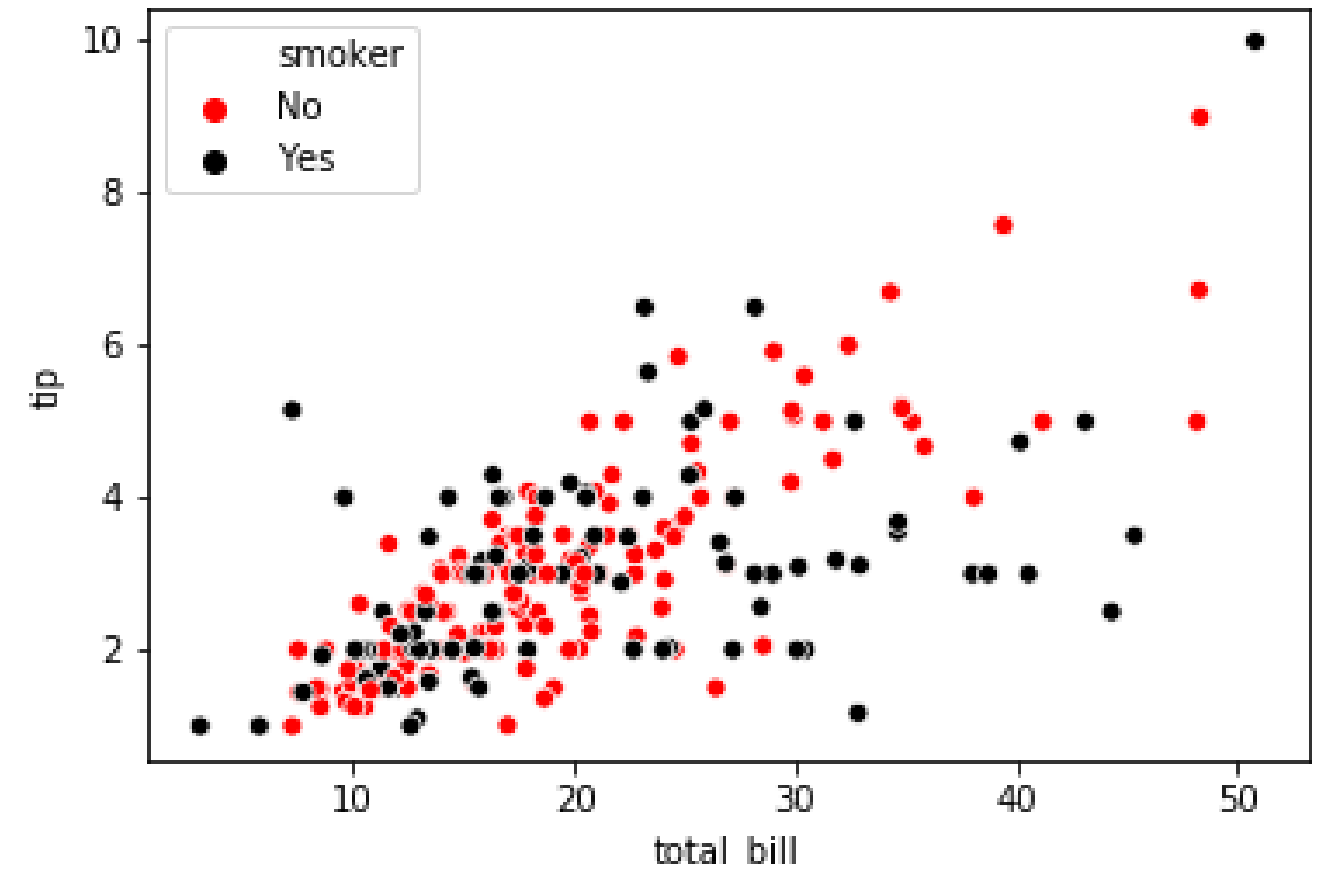
```
plt.show()
```




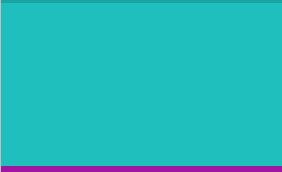
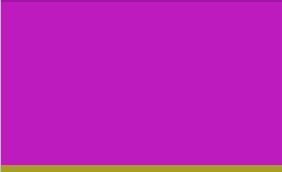





Specifying hue colors

```
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"	#00bfff
	purple	"magenta"	"m"	#bf00bf
	yellow	"yellow"	"y"	#ffff00
	black	"black"	"k"	#000000
	white	"white"	"w"	#ffffff

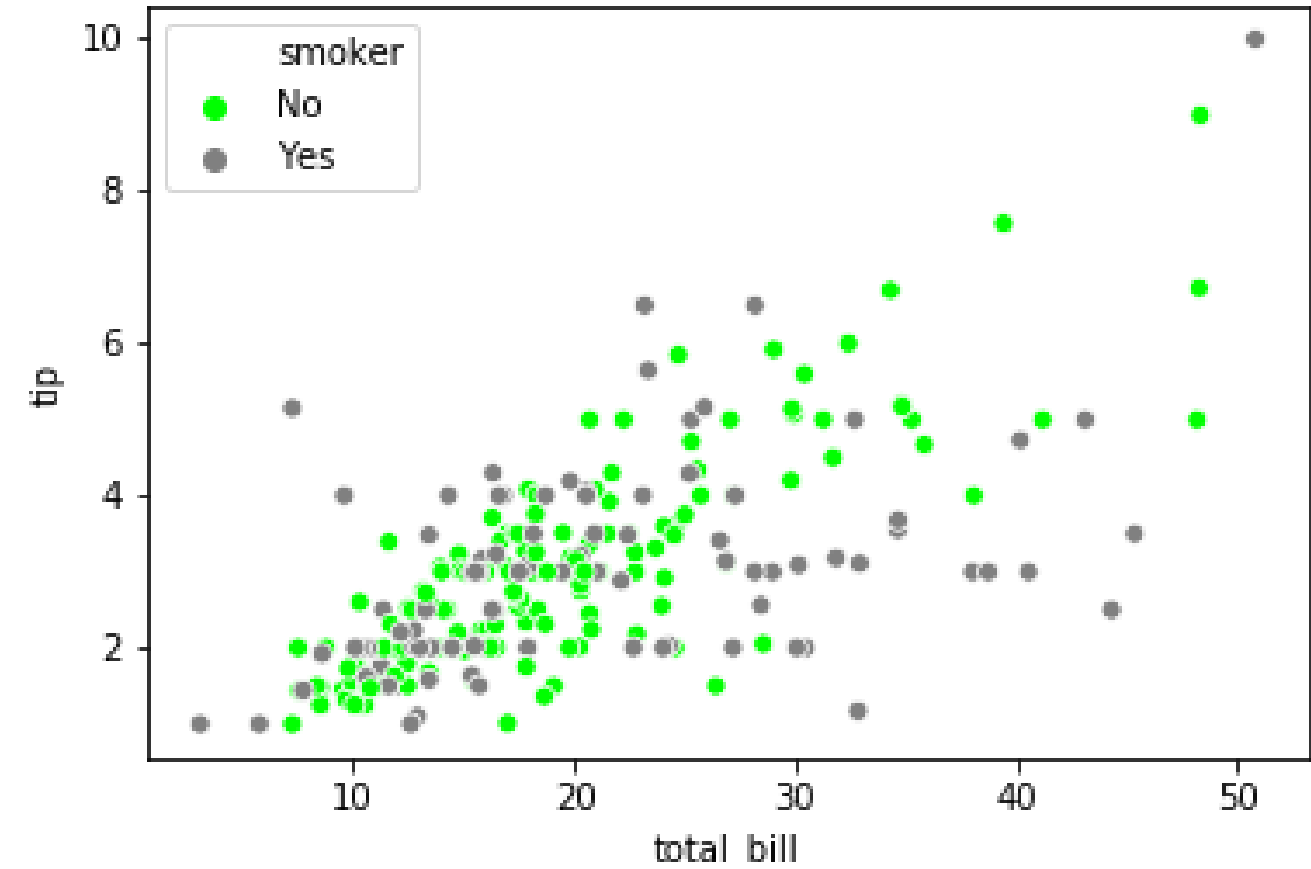
Using HTML hex color codes with hue

```
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()
```

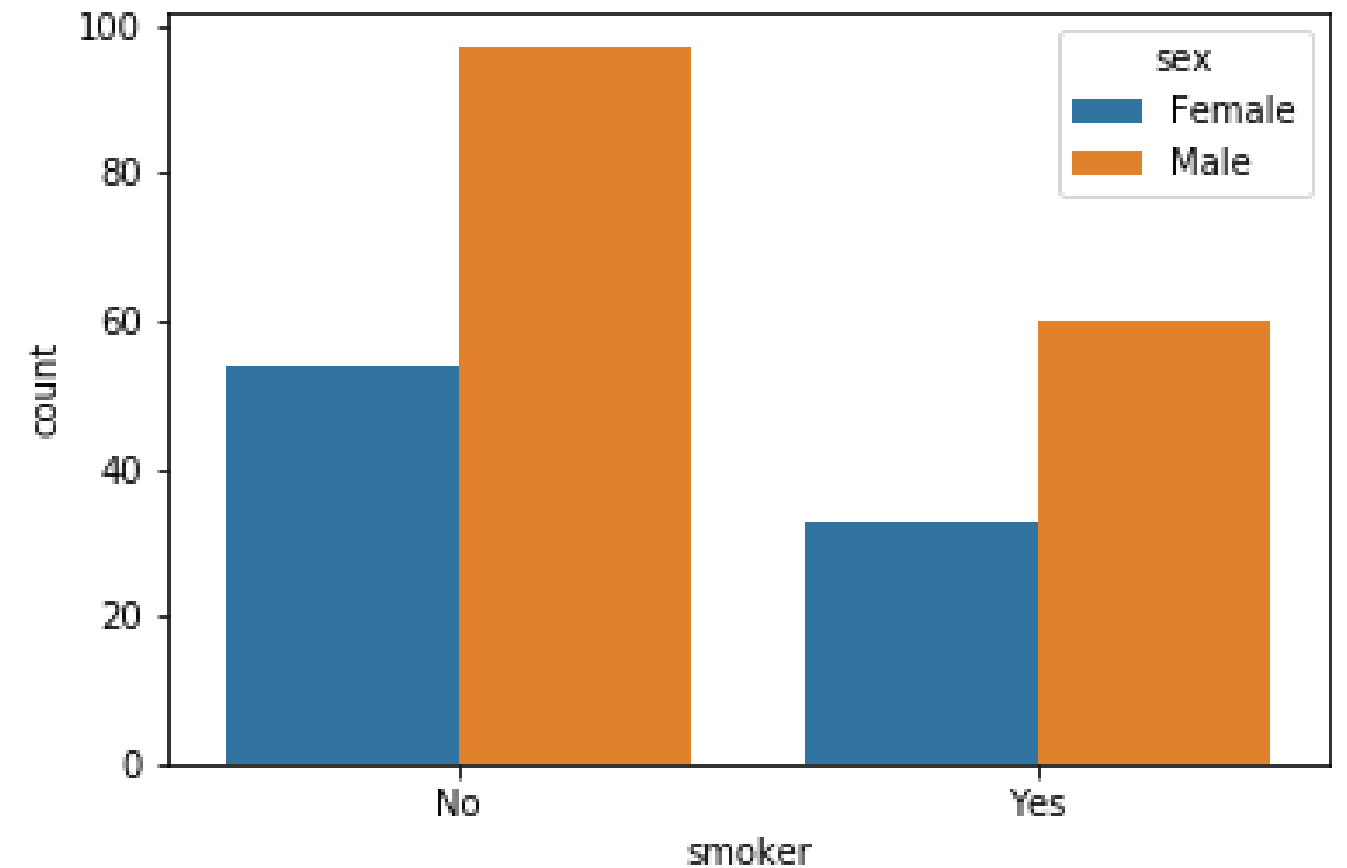


Using hue with count plots

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

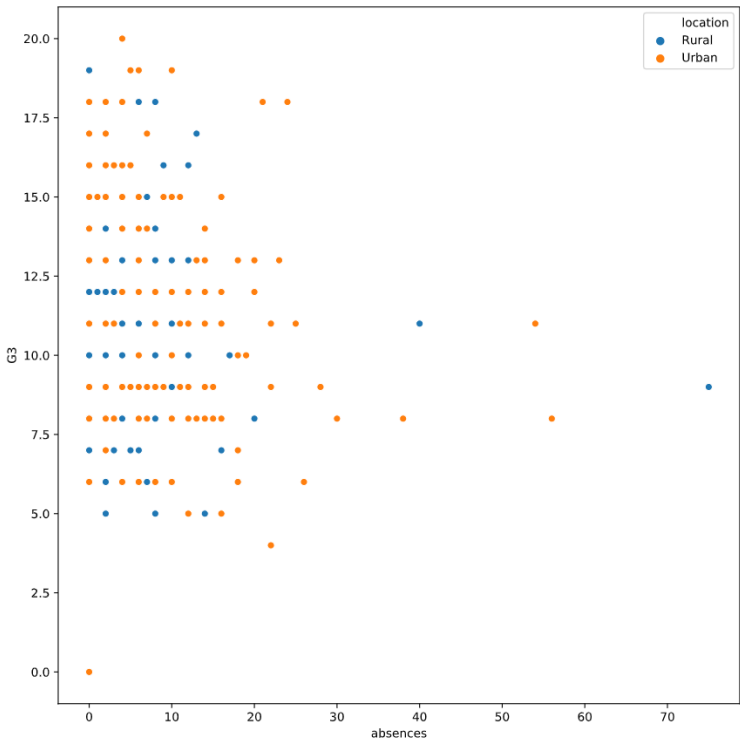
sns.countplot(x="smoker",
              data=tips,
              hue="sex")

plt.show()
```



Example 4 : . (segmented by where the student lives (rural, urban area))

```
script.py
1 # Import Matplotlib and Seaborn
2 import matplotlib.pyplot as plt
3 import seaborn as sns
4
5 # Change the legend order in the scatter plot
6 sns.scatterplot(x="absences", y="G3",
7                 data=student_data,
8                 hue="location", hue_order=["Rural", "Urban"])
9
10 # Show plot
11 plt.show()
```



Example 5 : school (GP/MS)
How many students live , 가

```
script.py
1 # Import Matplotlib and Seaborn
2 import matplotlib.pyplot as plt
3 import seaborn as sns
4
5 # Create a dictionary mapping subgroup values to colors
6 palette_colors = {"Rural": "green", "Urban": "blue"}
7
8 # Create a count plot of school with location subgroups
9 sns.countplot(x="school", data=student_data, hue="location", palette=palette_colors)
10
11 # Display plot
12 plt.show()
```

