Introduction to relational plots and subplots

INTRODUCTION TO DATA VISUALIZATION WITH SEABORN

Erin Case
Data Scientist

datacamp

scatterplot() vs. relplot()

Using `scatterplot()`

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

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

plt.show()
```

Using `relplot()`

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

sns.relplot(x="total_bill",
            y="tip",
            data=tips,
            kind="scatter")

plt.show()
```
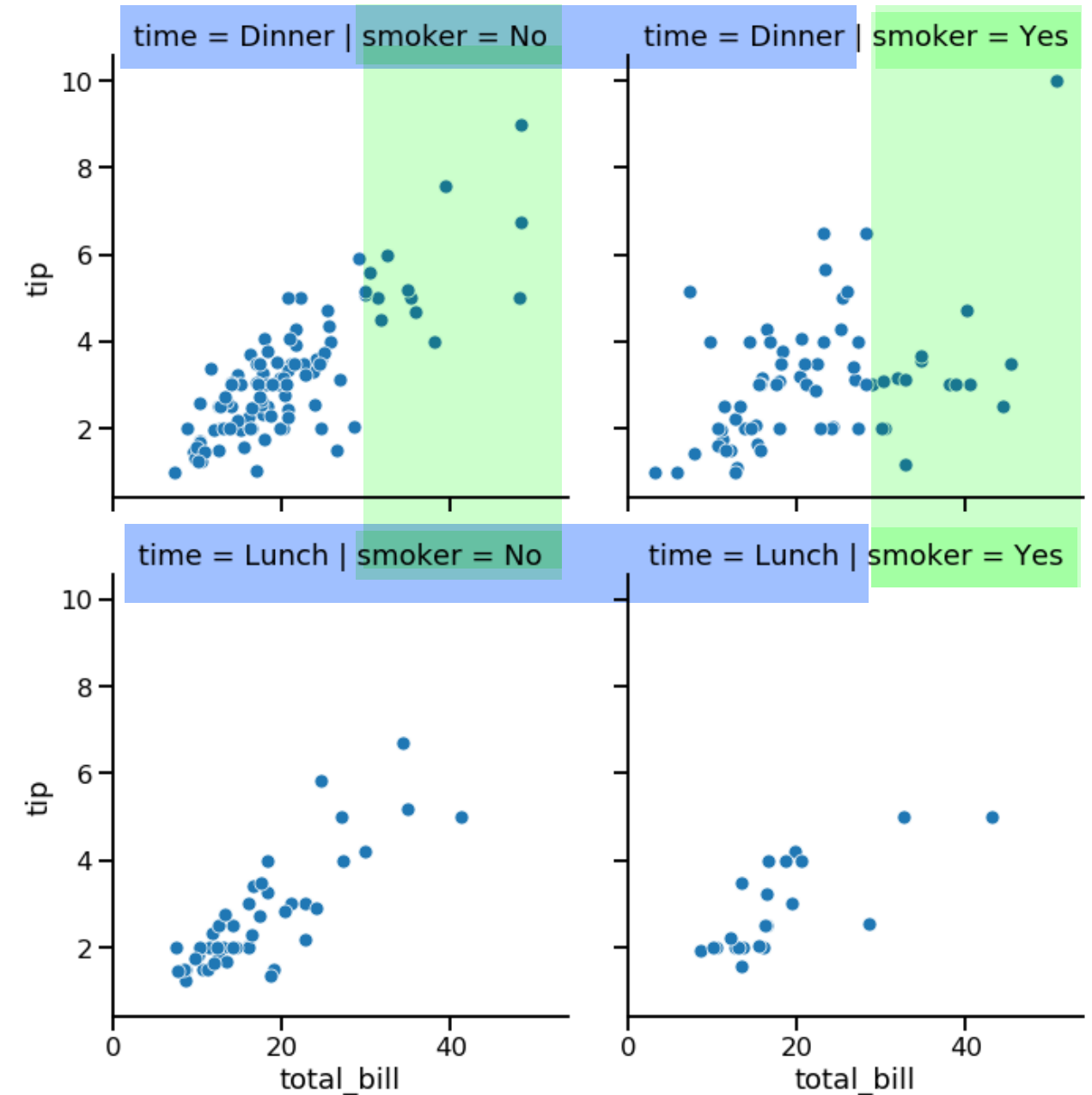
Subplots in rows and columns

```python
import seaborn as sns

import matplotlib.pyplot as plt


sns.relplot(x="total_bill",
            y="tip",
            data=tips,
            kind="scatter",
            col="smoker",
            row="time")

plt.show()
```
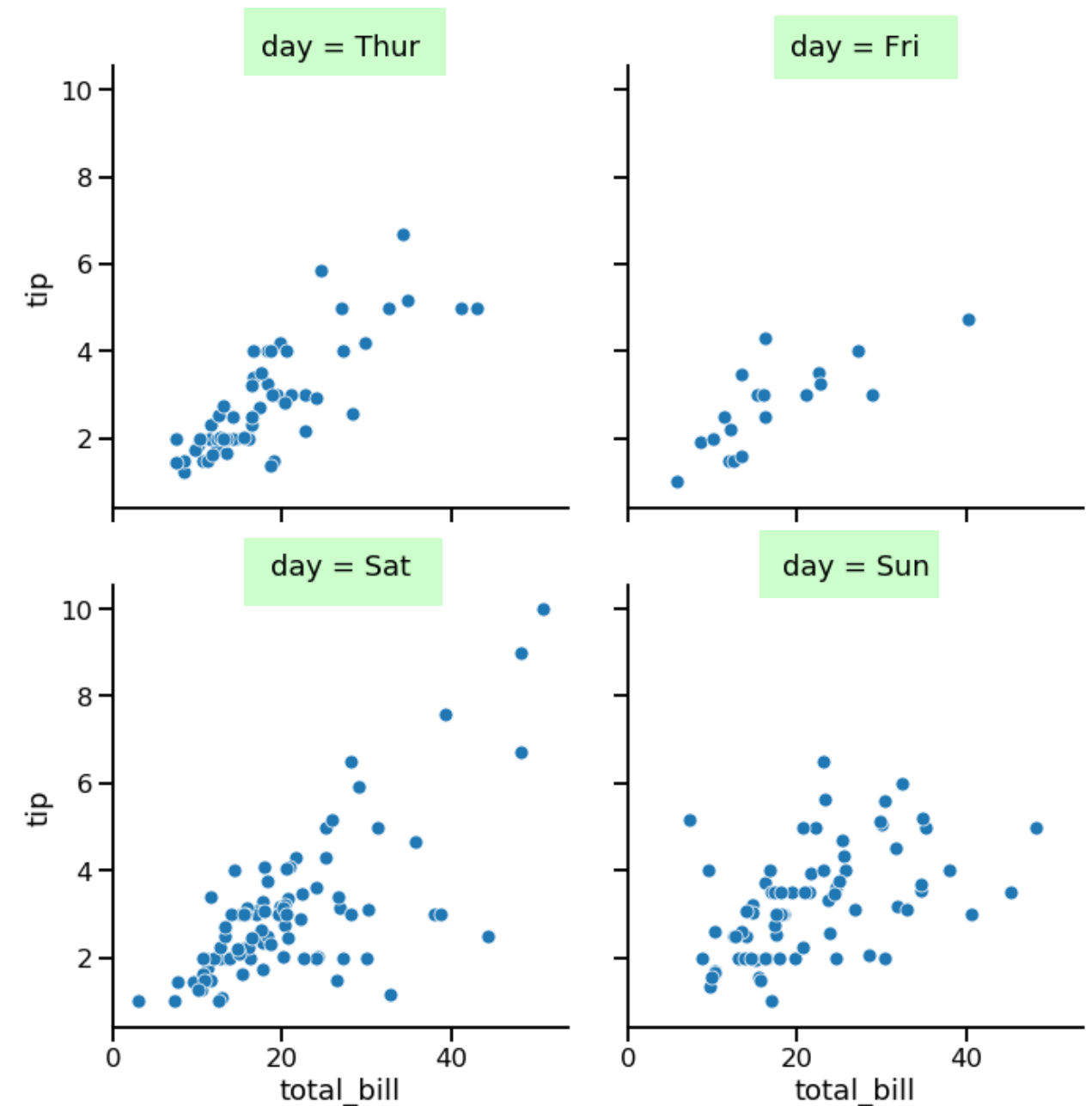
Ordering columns

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

sns.relplot(x="total_bill",
            y="tip",
            data=tips,
            kind="scatter",
            col="day",
            col_wrap=2,
            col_order=["Thur",
                       "Fri",
                       "Sat",
                       "Sun"])
plt.show()
```
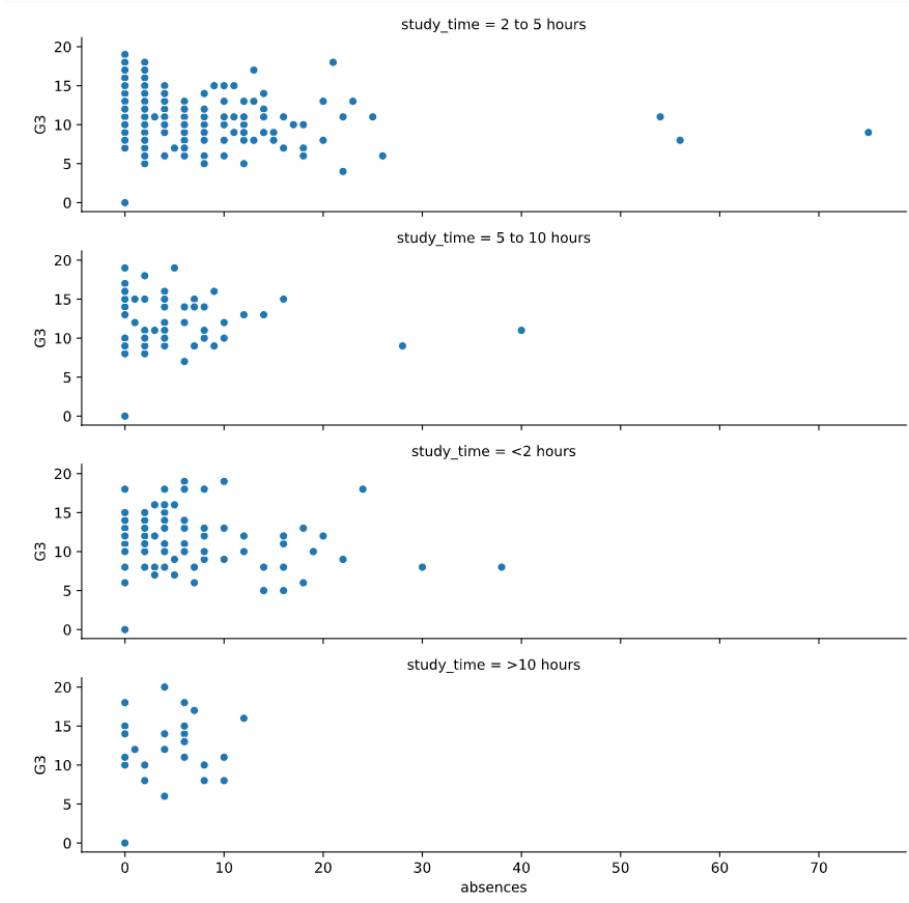
EXERCISE 1:

· EXERCISE 2: first grade   last grade   ?

?   famsup, schoolsup   grade

```
script.py                                    ☀ Li
1    # Change this scatter plot to arrange the plots in rows
     instead of columns
2    sns.relplot(x="absences", y="G3",
3                data=student_data,
4                kind="scatter",
5                row="study_time")
6
7    # Show plot
8    plt.show()
```

```
script.py
1    # Adjust further to add subplots based on family support
2    sns.relplot(x="G1", y="G3",
3                data=student_data,
4                kind="scatter",
5                col="schoolsup",
6                col_order=["yes", "no"],
7                row="famsup",
8                row_order=["yes", "no"])
9
10   # Show plot
11   plt.show()
```
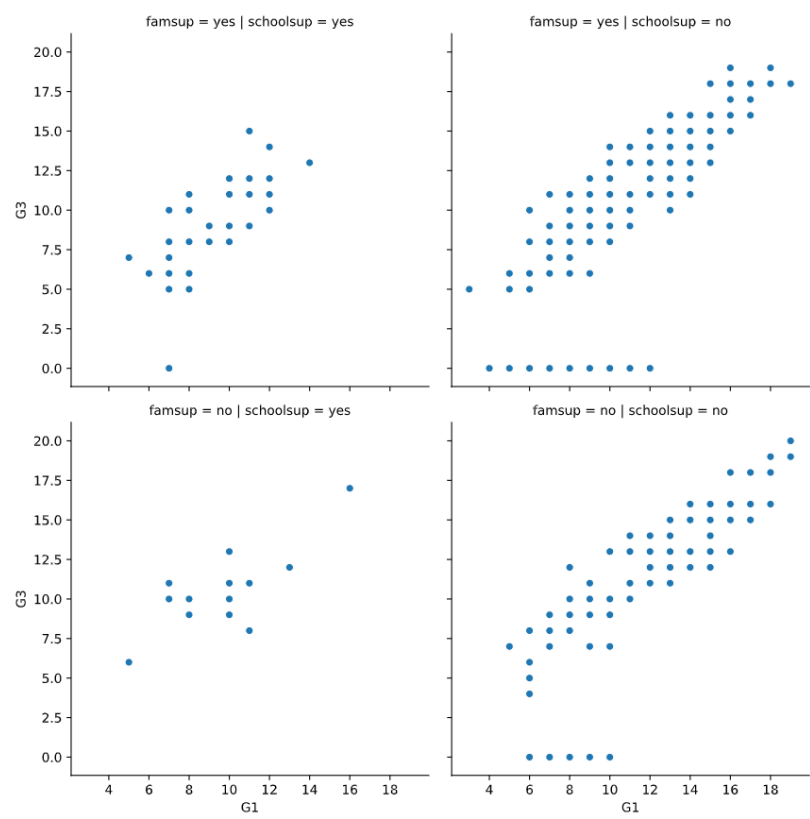
Customizing scatter plots

INTRODUCTION TO DATA VISUALIZATION WITH SEABORN

Erin Case
Data Scientist
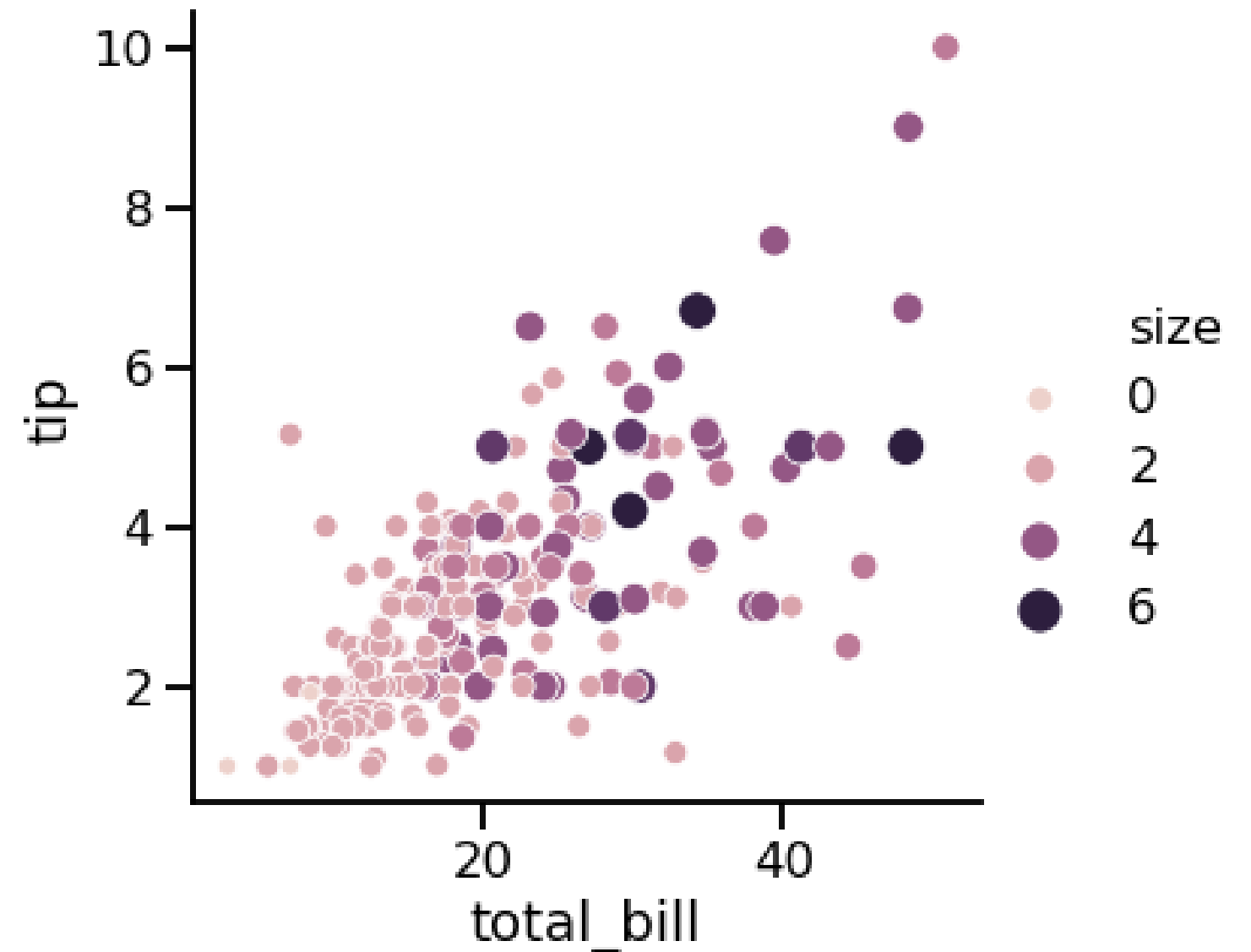
datacamp

Point size and hue

```
import seaborn as sns

import matplotlib.pyplot as plt


sns.relplot(x="total_bill",
            y="tip",
            data=tips,
            kind="scatter",
            size="size",
            hue="size")

plt.show()
```
: size
  seaborn   !
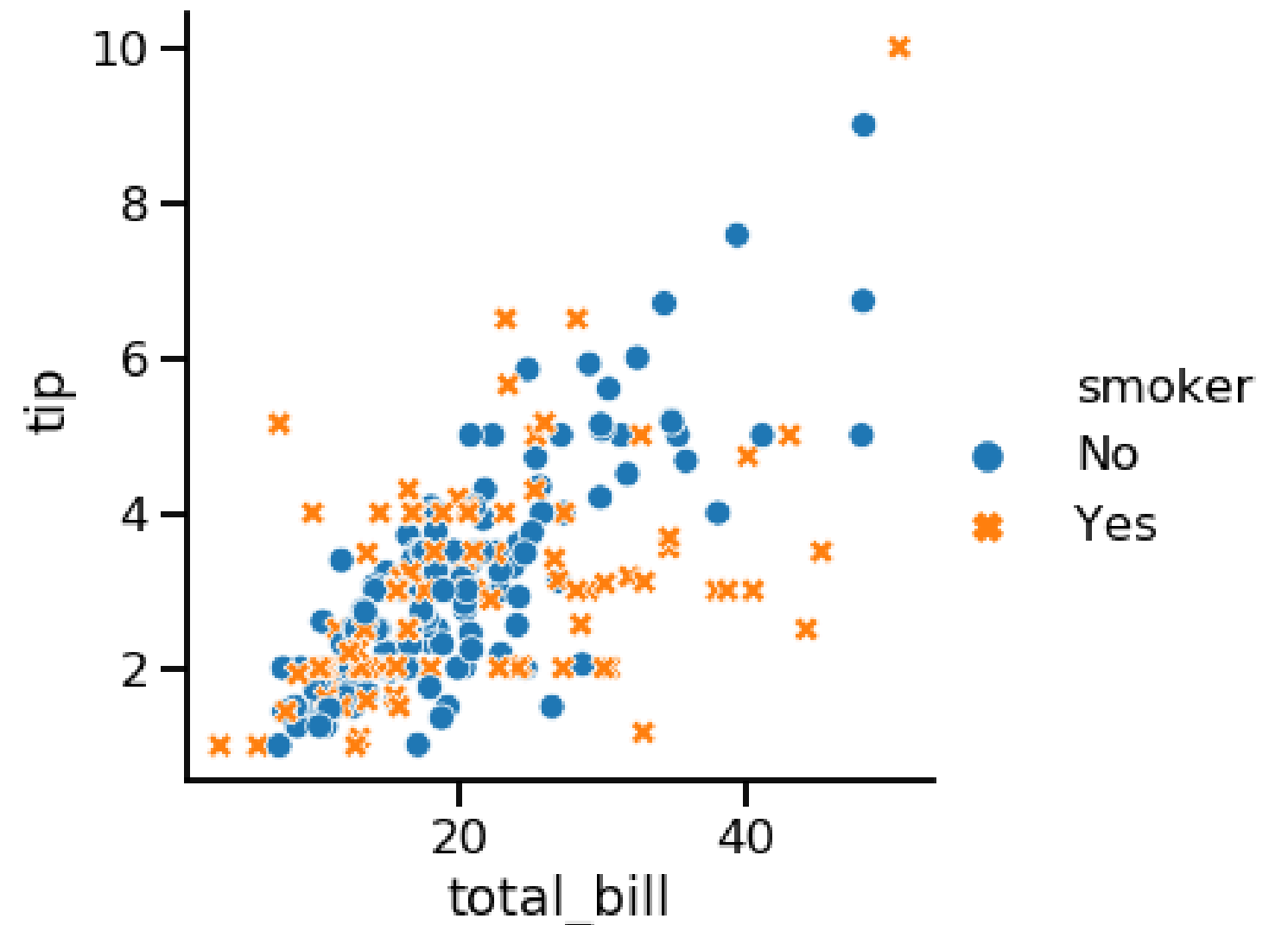

:

Subgroups with point style

```python
import seaborn as sns

import matplotlib.pyplot as plt


sns.relplot(x="total_bill",
            y="tip",
            data=tips,
            kind="scatter",
            hue="smoker",
            style="smoker")

plt.show()
```
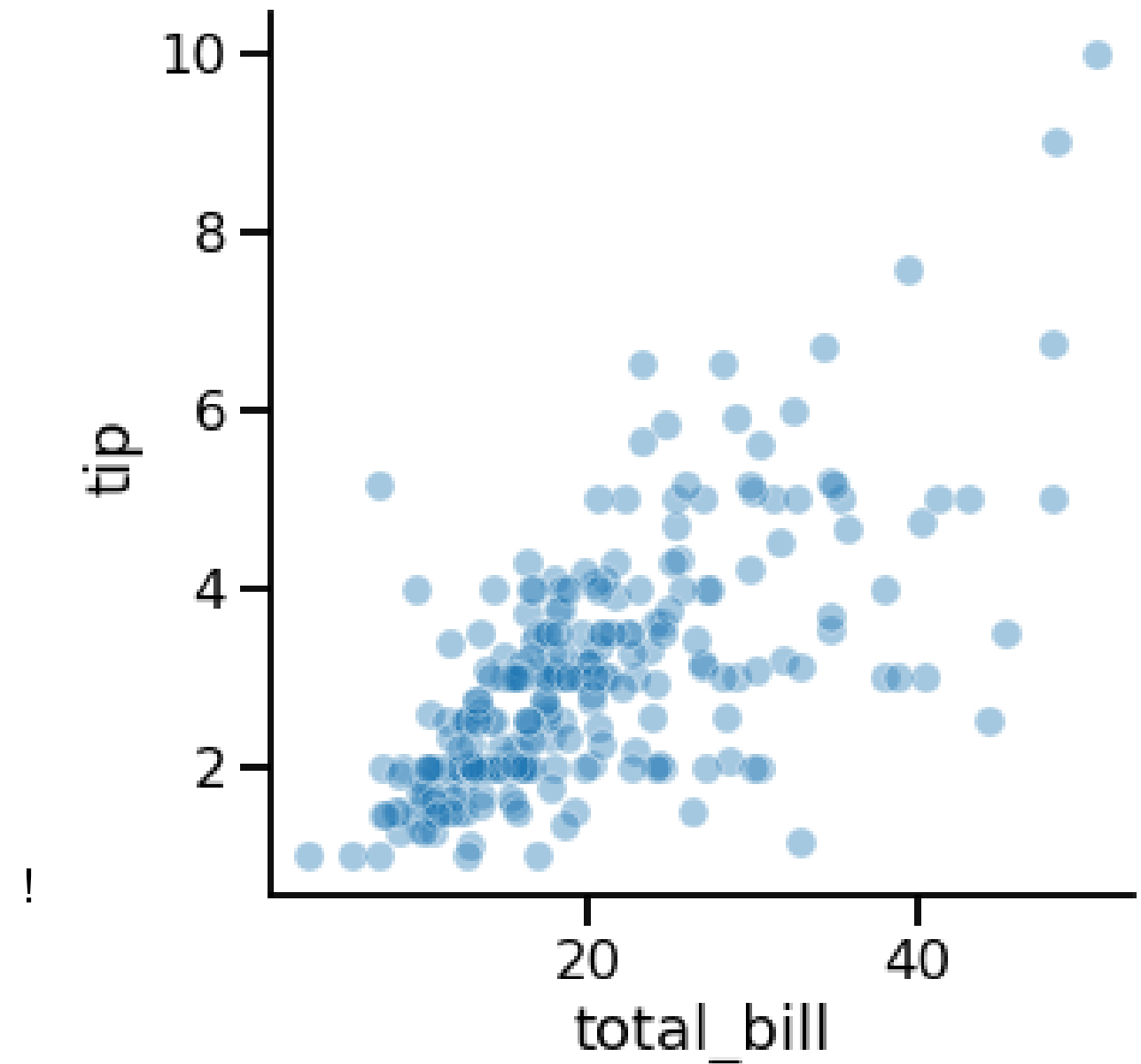
Changing point transparency

```python
import seaborn as sns

import matplotlib.pyplot as plt

# Set alpha to be between 0 and 1
sns.relplot(x="total_bill",
            y="tip",
            data=tips,
            kind="scatter",
            alpha=0.4)

plt.show()
```
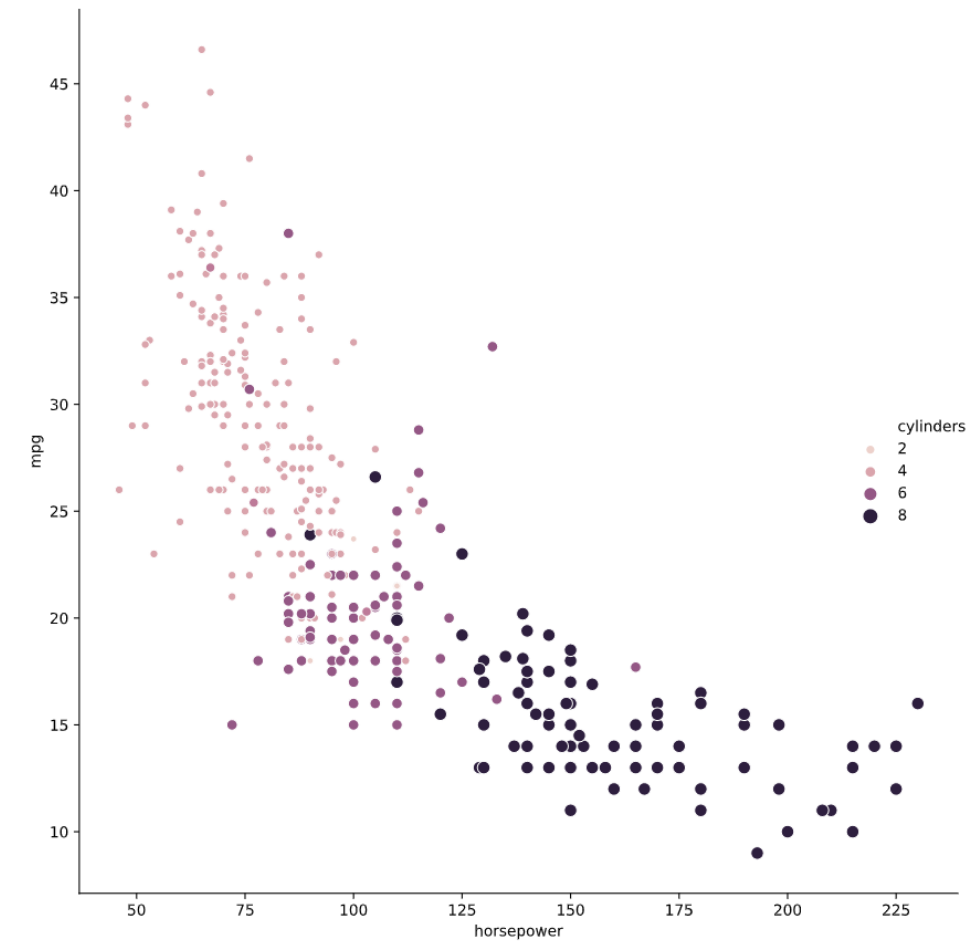
EXERCISE 3: mpg data   horsepower   mpg   ?
                             cylinder   ?


```python
# Import Matplotlib and Seaborn
import matplotlib.pyplot as plt
import seaborn as sns

# Create scatter plot of horsepower vs. mpg
sns.relplot(x="horsepower", y="mpg",
            data=mpg, kind="scatter",
            size="cylinders",
            hue="cylinders")

# Show plot
plt.show()
```


: size   hue   !!

: cylinder   .
  horseposer   mpg   .

Introduction to line plots

INTRODUCTION TO DATA VISUALIZATION WITH SEABORN

Erin Case
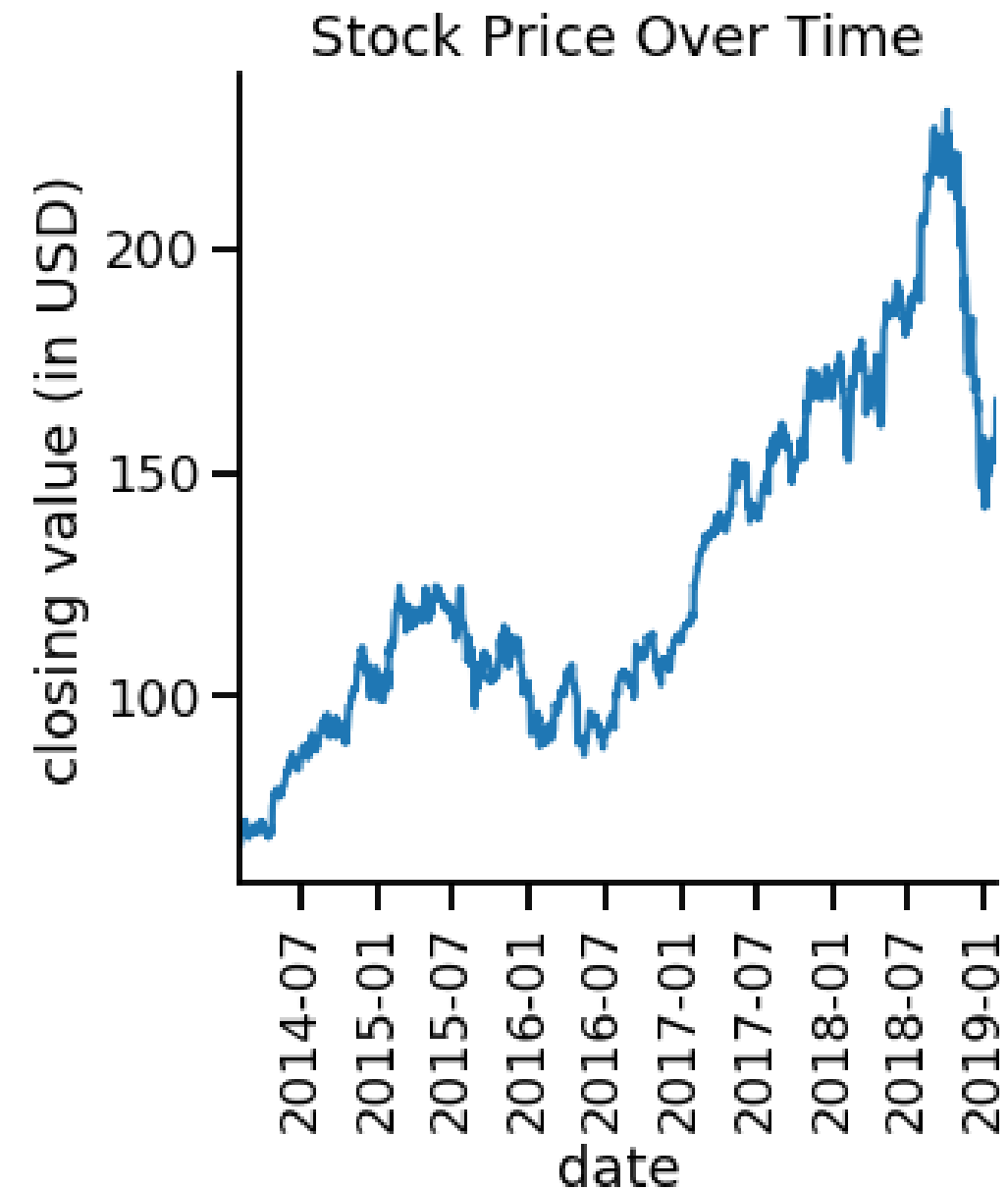Data Scientist

What are line plots?

Two types of relational plots: scatter plots and line plots

Scatter plots

- Each plot point is an independent observation

Line plots

- Each plot point represents the same "thing", typically tracked over time
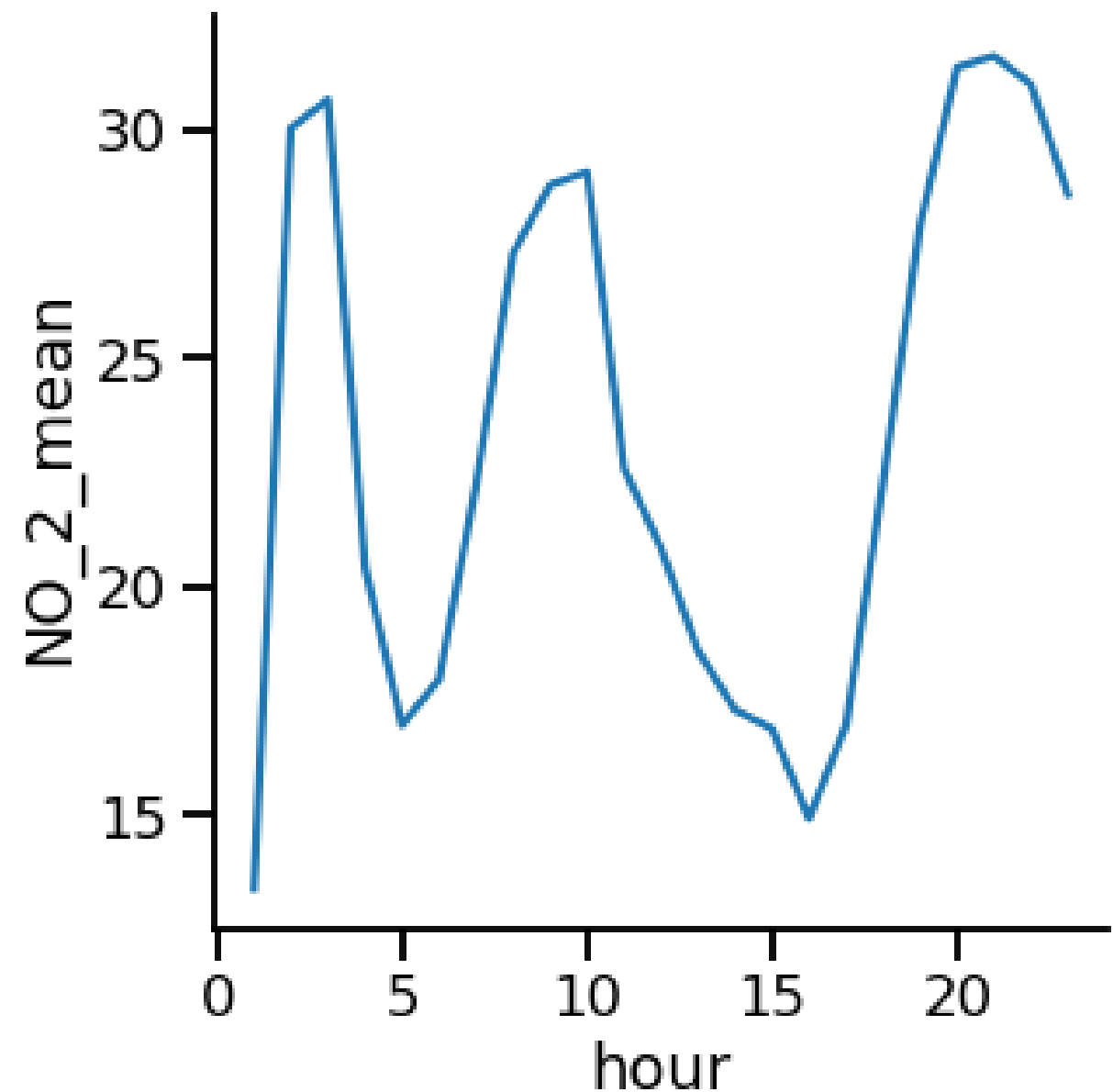
Line plot

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

sns.relplot(x="hour", y="NO_2_mean",
            data=air_df_mean,
            kind="line")

plt.show()
```
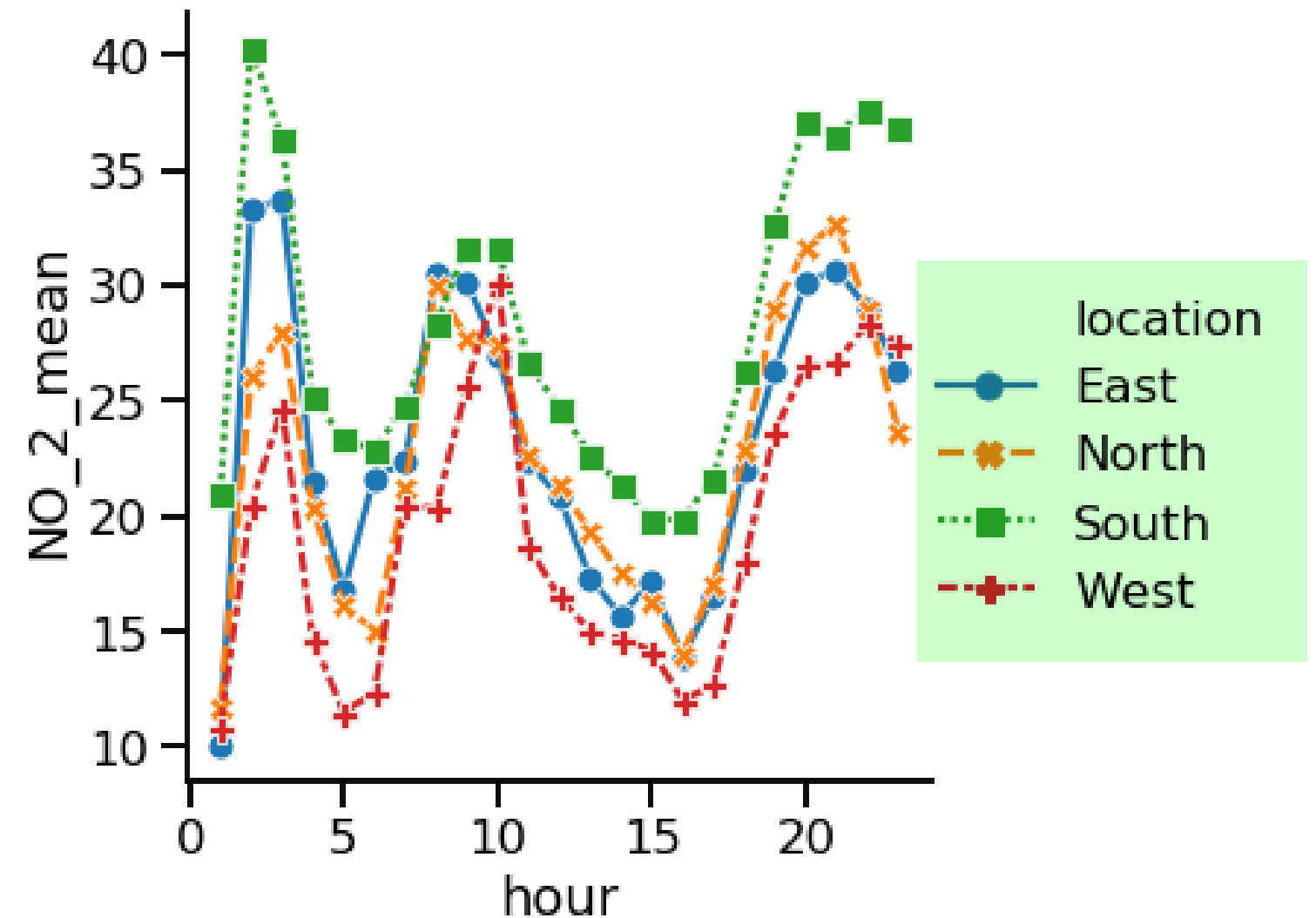
Adding markers

location : East, North, South, West .

```python
import matplotlib.pyplot as plt

import seaborn as sns


sns.relplot(x="hour", y="NO_2_mean",
            data=air_df_loc_mean,
            kind="line",
            style="location",
            hue="location",
            markers=True)

plt.show()
```

Turning off line style

```python
import matplotlib.pyplot as plt

import seaborn as sns


sns.relplot(x="hour", y="NO_2_mean",
            data=air_df_loc_mean,
            kind="line",
            style="location",
            hue="location",
            markers=True,
            dashes=False)

plt.show()
```
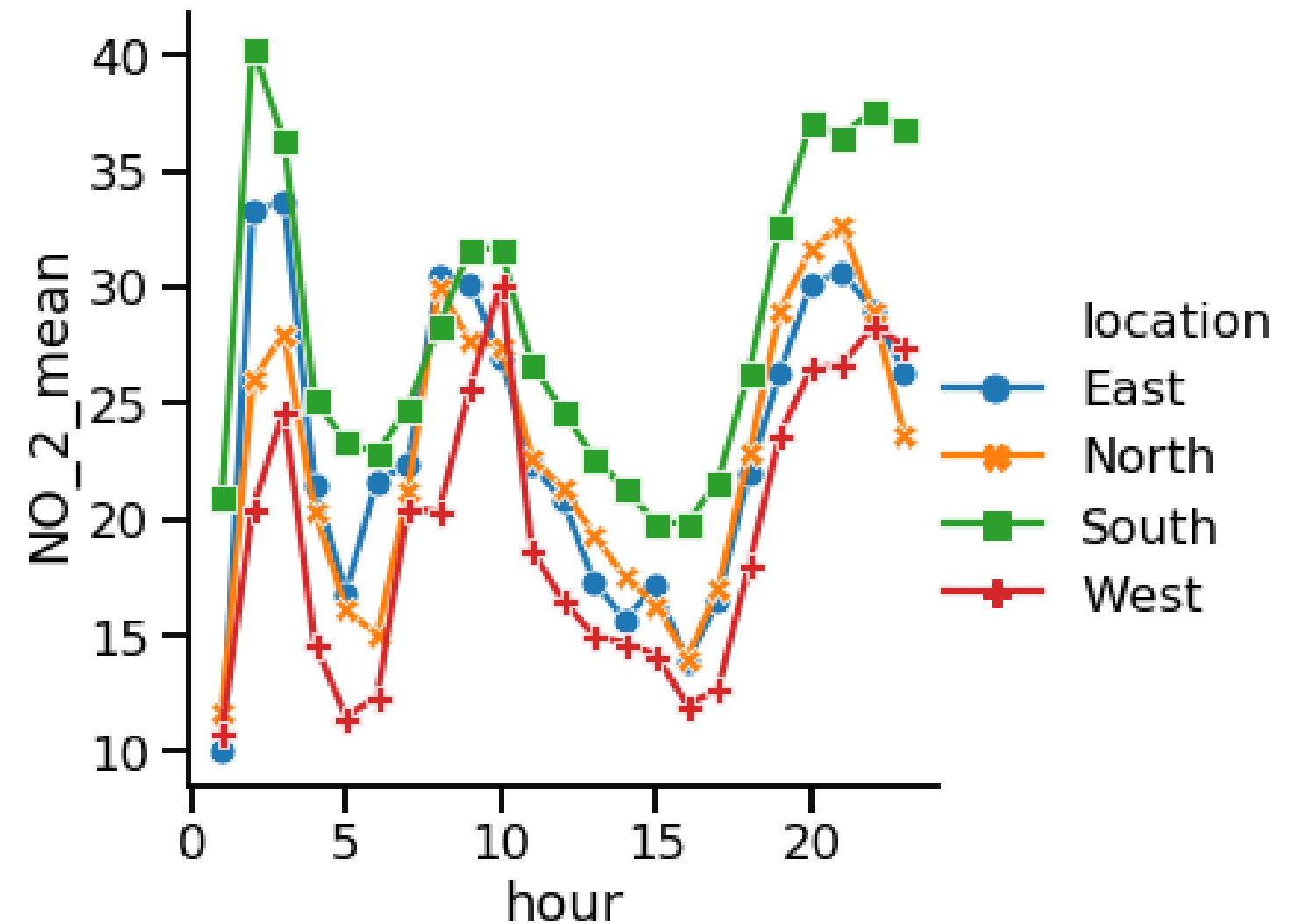
style   line

dasher=False

Multiple observations per x-value

Line plot

```
import matplotlib.pyplot as plt

import seaborn as sns


sns.relplot(x="hour", y="NO_2",
            data=air_df,
            kind="line")


plt.show()
```
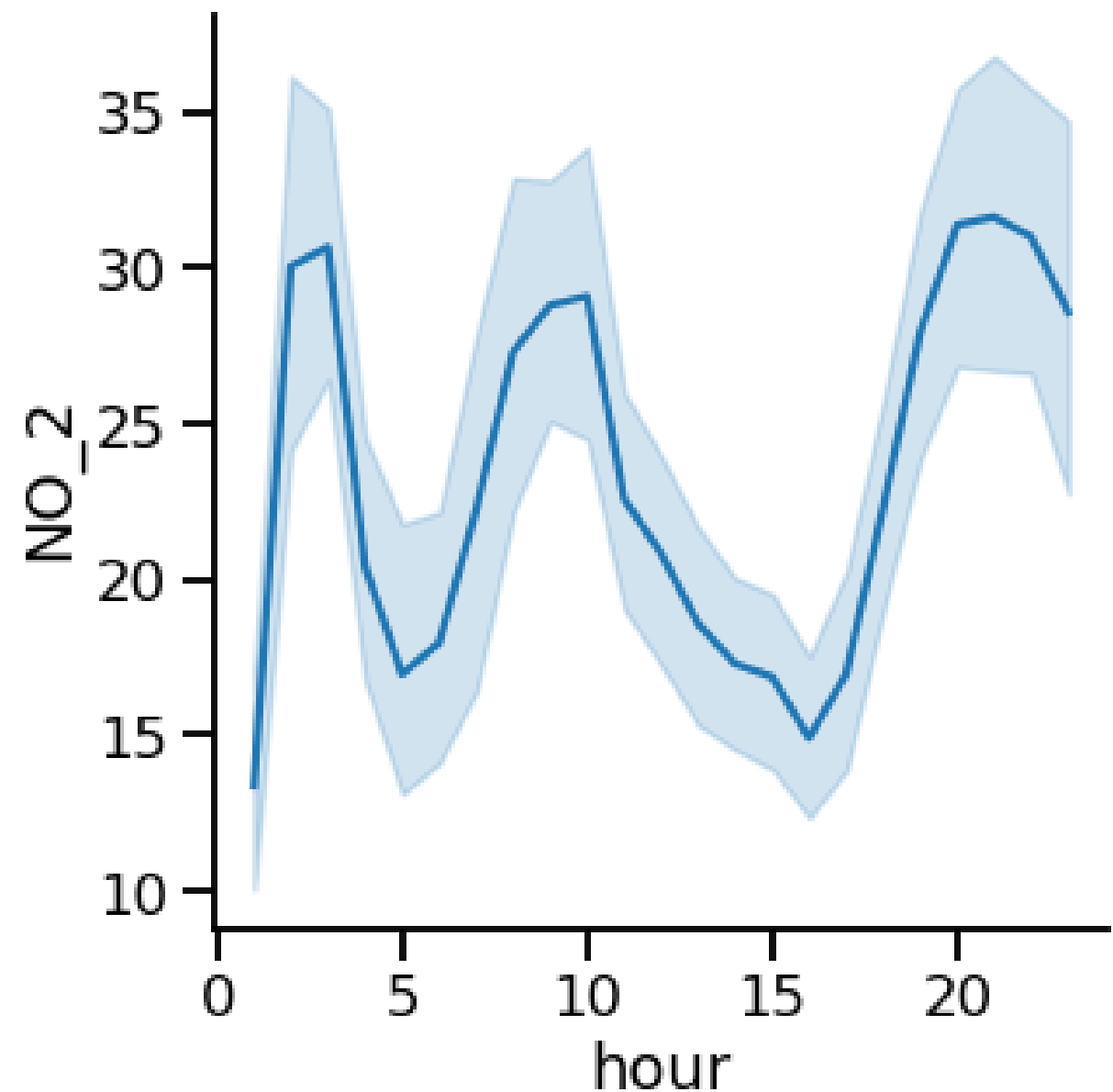

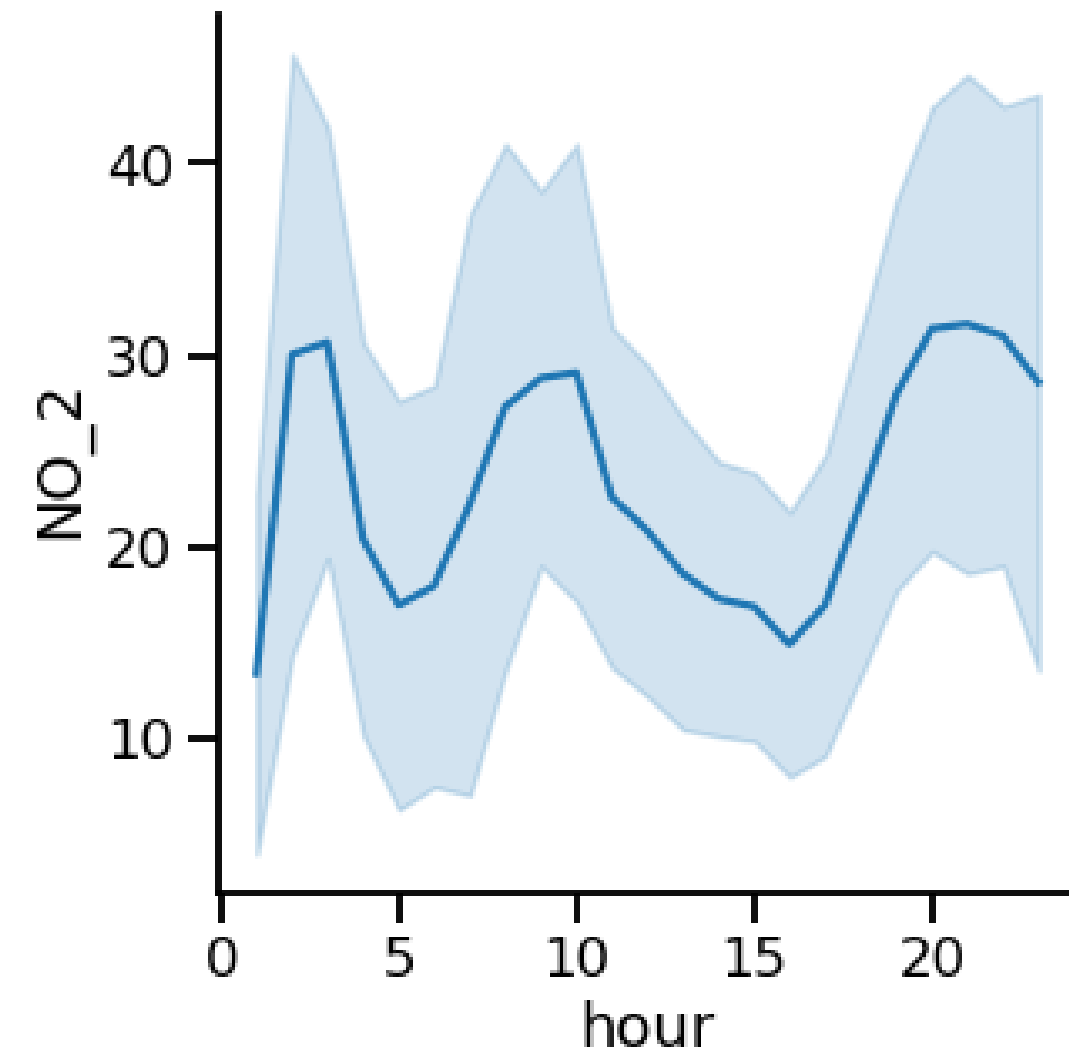Shaded region   random sample   95%

Replacing confidence interval with standard deviation

```python
import matplotlib.pyplot as plt

import seaborn as sns


sns.relplot(x="hour", y="NO_2",

            data=air_df,

            kind="line",

            ci="sd")

plt.show()  ci='None'              graph
```

EXERCISE 4: mpg data   horsepower   time   ?
                    country   ?

: Time   horsepower   .

country   .

```
script.py
1    # Import Matplotlib and Seaborn
2    import matplotlib.pyplot as plt
3    import seaborn as sns
4
5    # Add markers and make each line have the same style
6    sns.relplot(x="model_year", y="horsepower",
7                data=mpg, kind="line",
8                ci=None, style="origin",
9                hue="origin",
10               marker=True,
11               dashes=False)
12
13   # Show plot
14   plt.show()
```

origin(   )   style   hue
dash(   )   .

     marker=True   marker   ..?
           ..