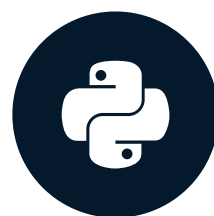*Part - 01*

# Count plots and bar plots
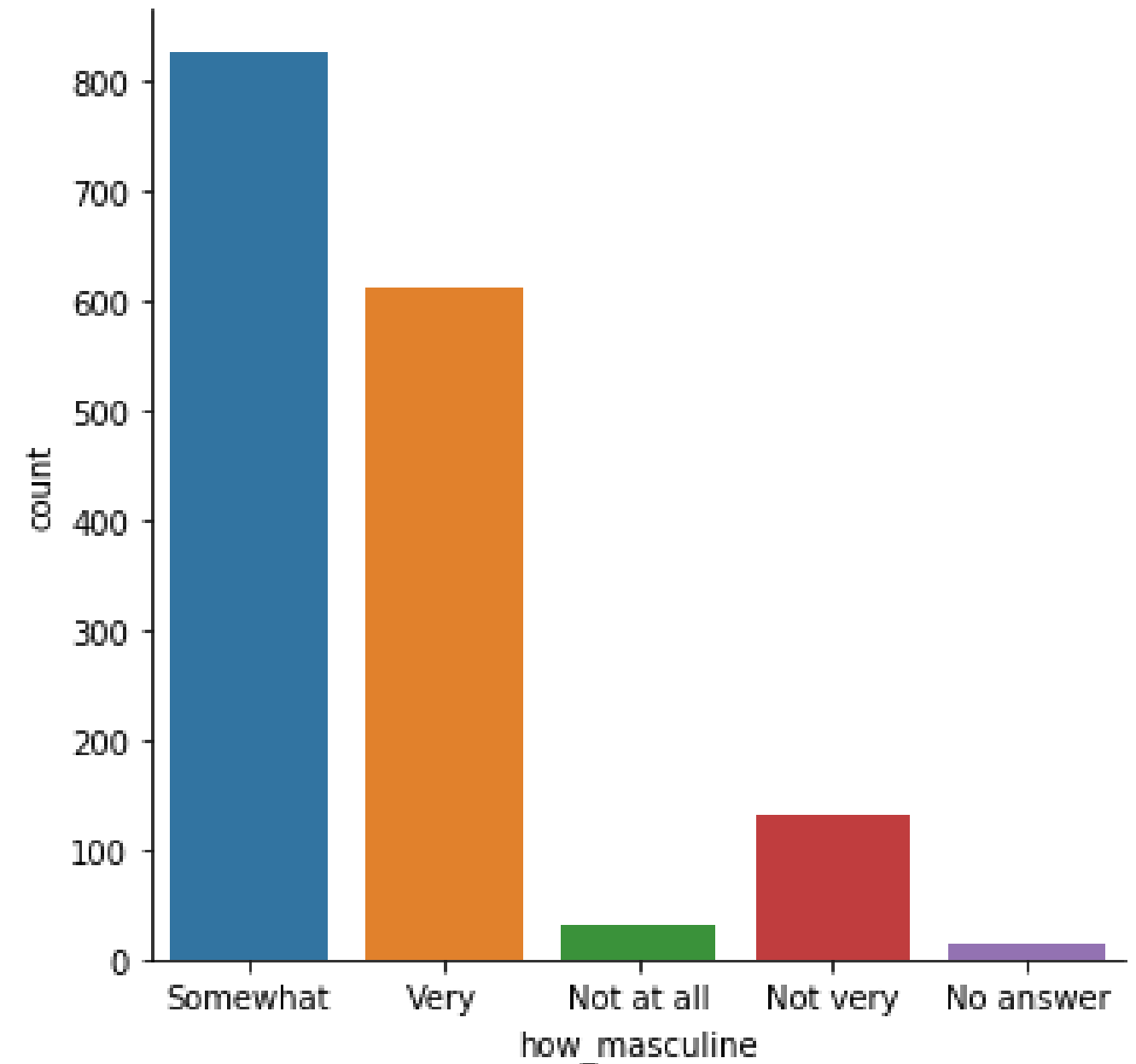
## INTRODUCTION TO DATA VISUALIZATION WITH SEABORN

**Erin Case**
Data Scientist

datacamp

# Categorical plots

- Examples: count plots, bar plots

- Involve a categorical variable

- Comparisons between groups
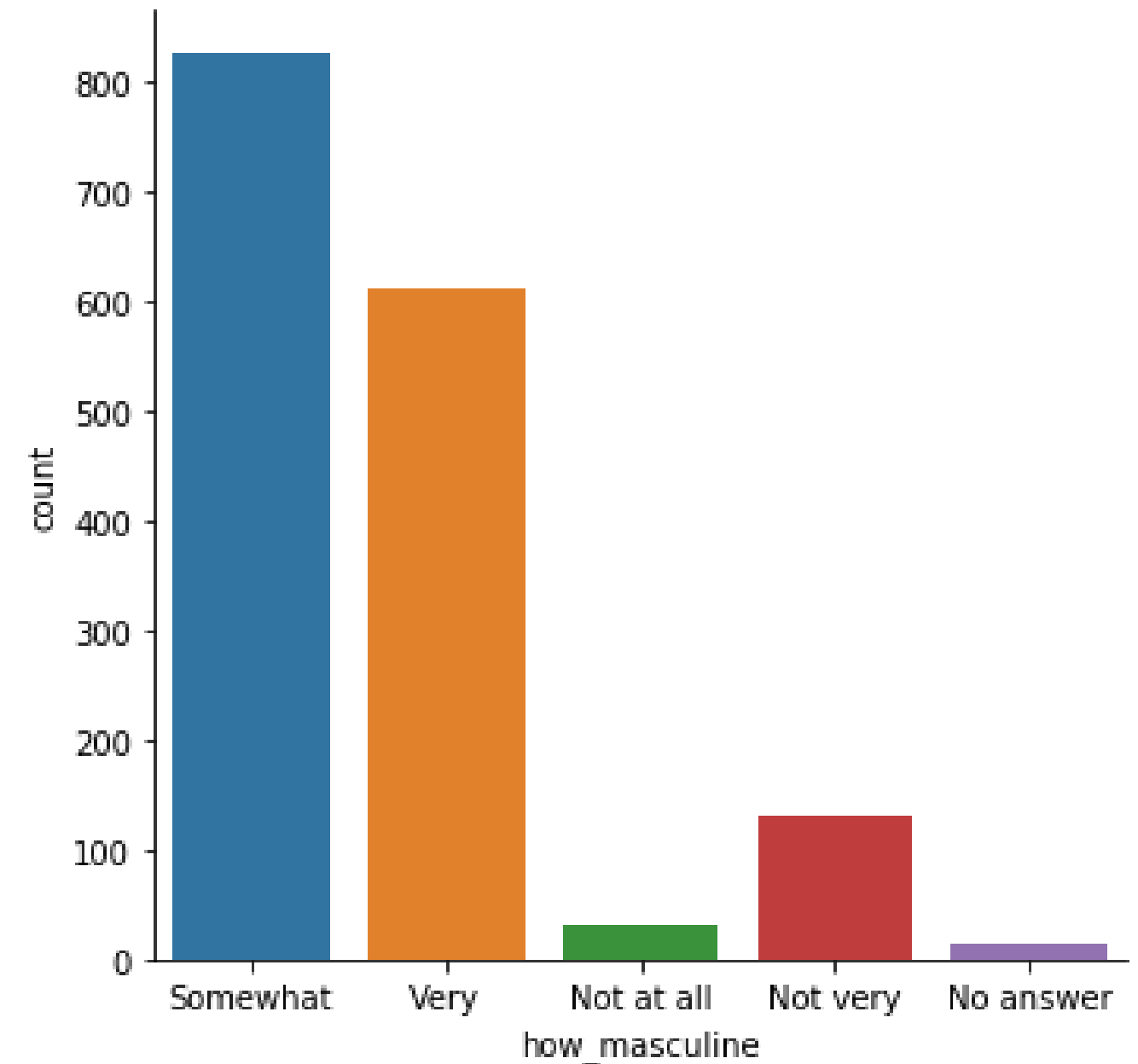
# catplot()

- Used to create categorical plots

- Same advantages of `relplot()`

- Easily create subplots with `col=` and `row=`

# countplot() vs. catplot()

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


sns.countplot(x="how_masculine",
              data=masculinity_data)


plt.show()
```
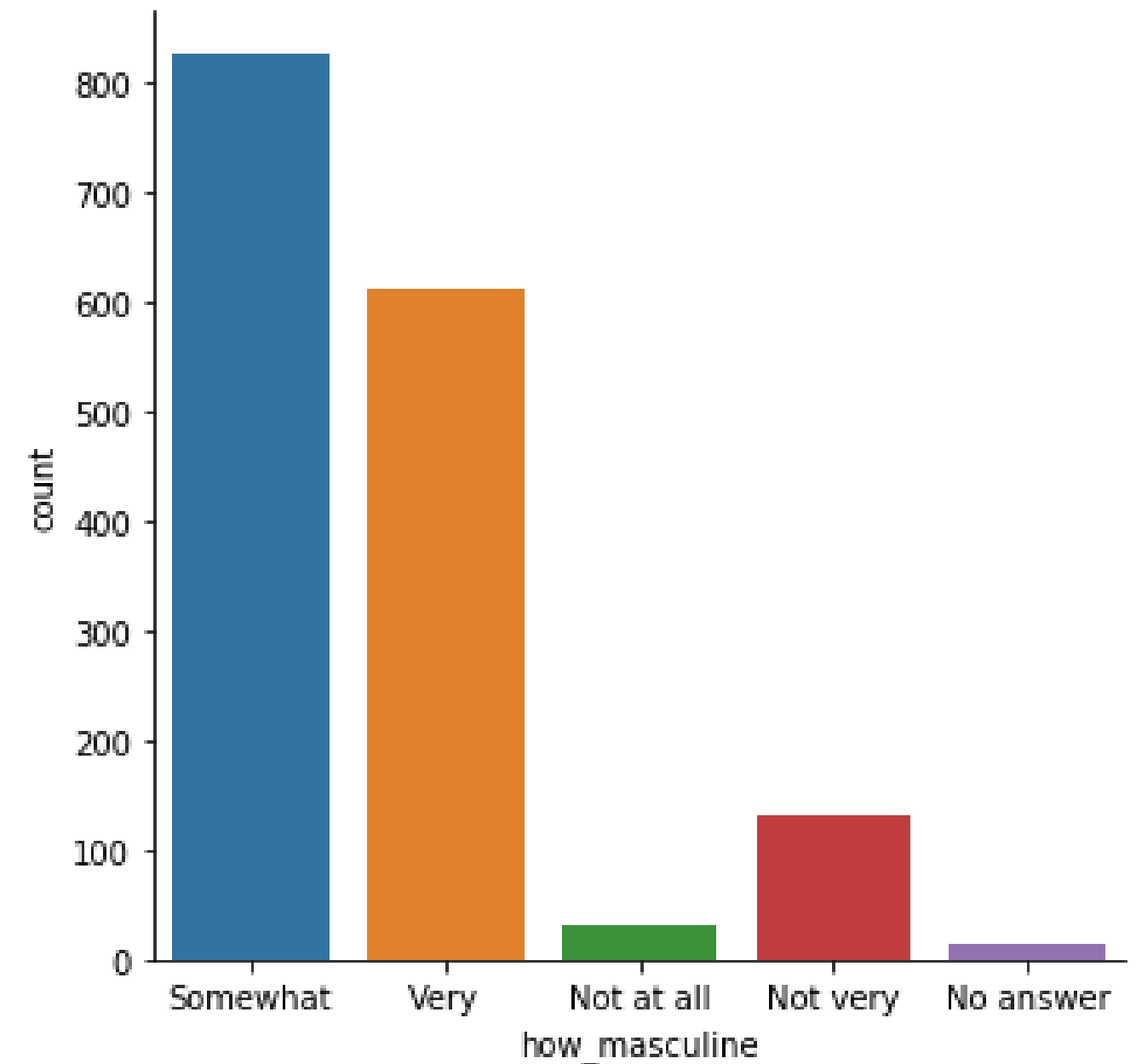
# countplot() vs. catplot()

```python
import matplotlib.pyplot as plt

import seaborn as sns


sns.catplot(x="how_masculine",
            data=masculinity_data,
            kind="count")


plt.show()
```

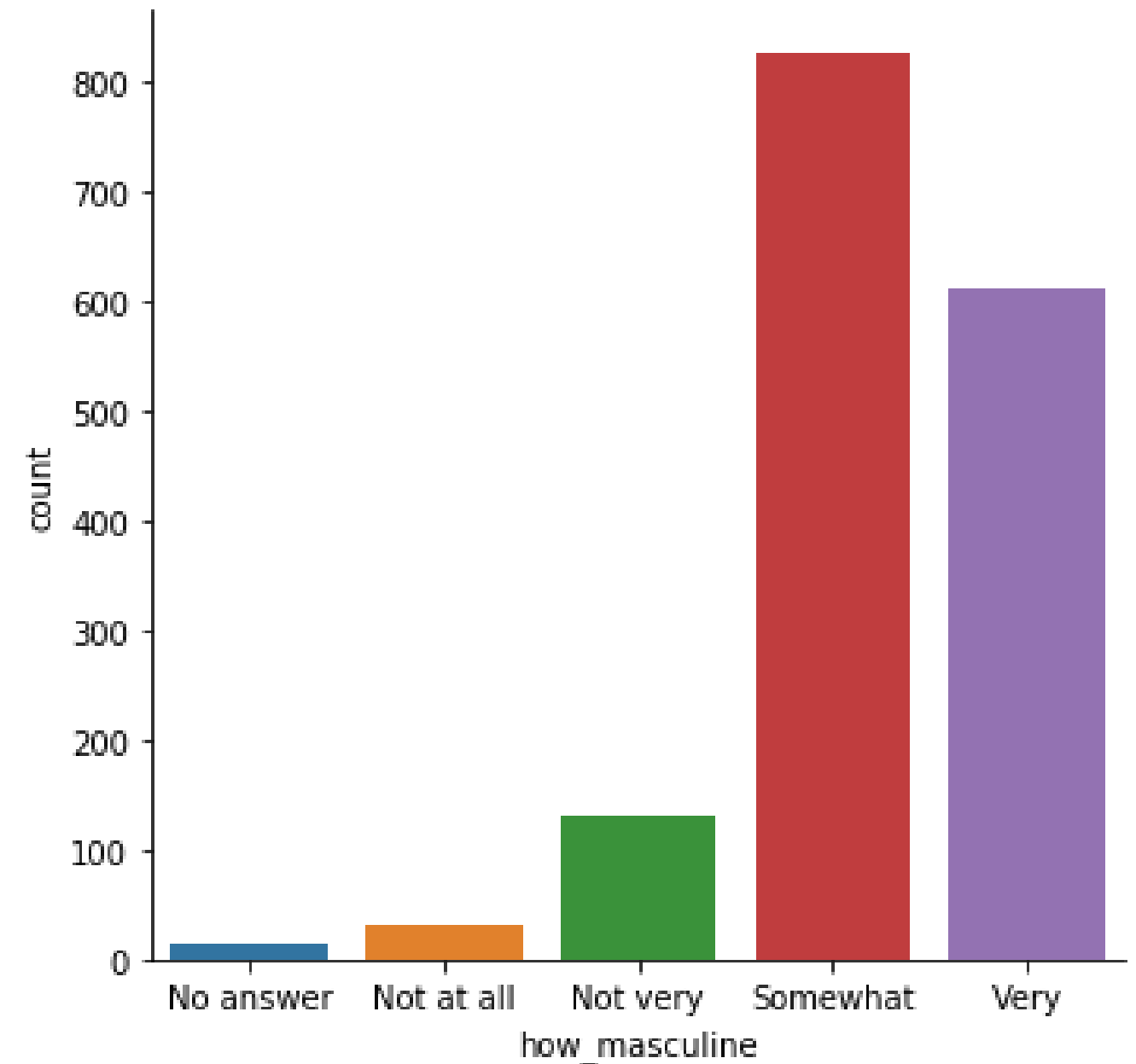y="Internet usage" , makes the countplot horizontal

# Changing the order

```python
import matplotlib.pyplot as plt
import seaborn as sns
category_order = ["No answer",
                  "Not at all",
                  "Not very",
                  "Somewhat",
                  "Very"]

sns.catplot(x="how_masculine",
            data=masculinity_data,
            kind="count",
            order=category_order)

plt.show()
```

works for all catplots()

# Bar plots

Displays mean of quantitative variable per category

one (day) to many (total_bills) relationship
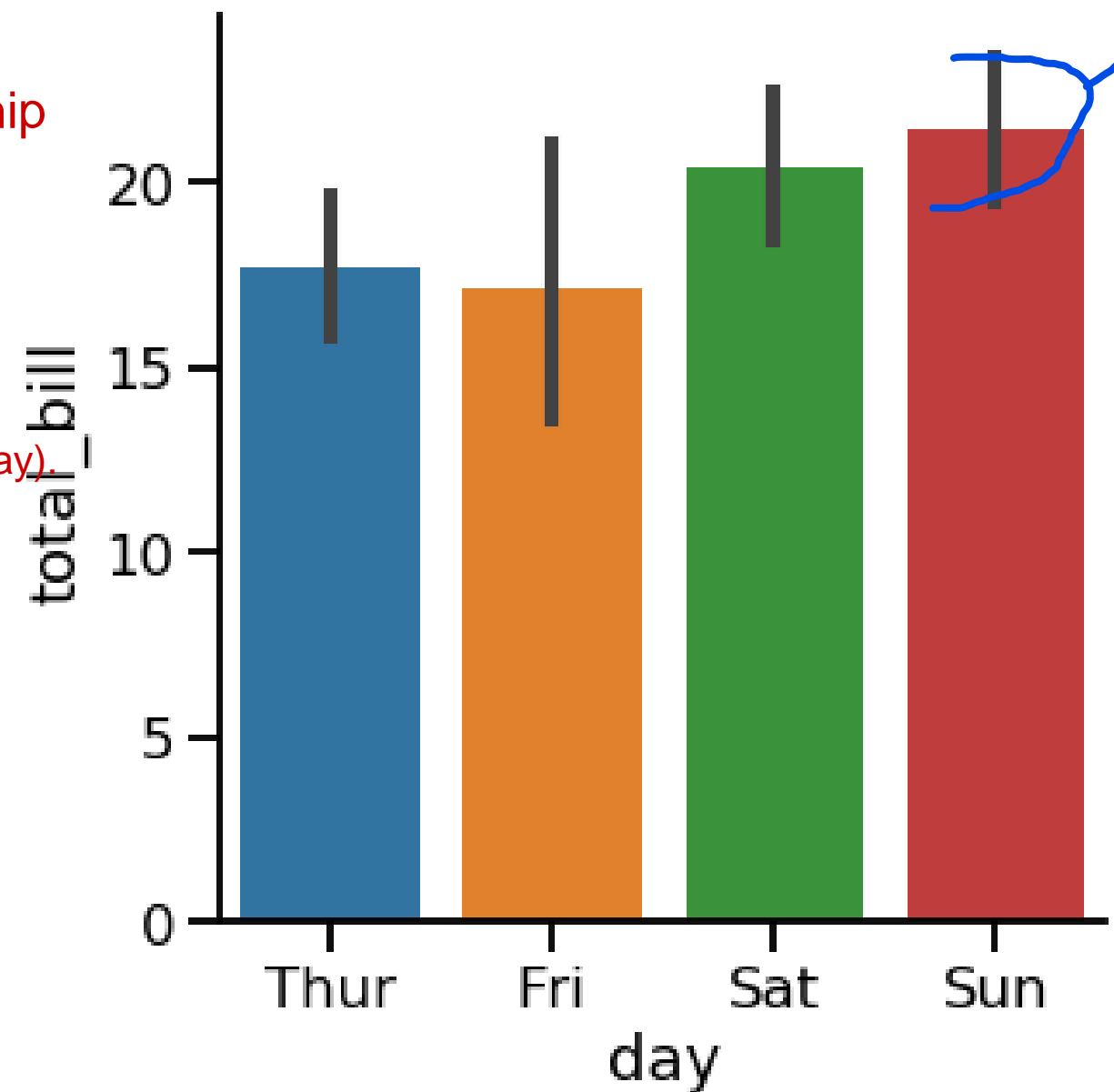
```python
import matplotlib.pyplot as plt

import seaborn as sns
```
it shows the mean (average) of the y variable (total_bill) for each category in x (day).

```python
sns.catplot(x="day",
            y="total_bill",
            data=tips,
            kind="bar")

plt.show()
```
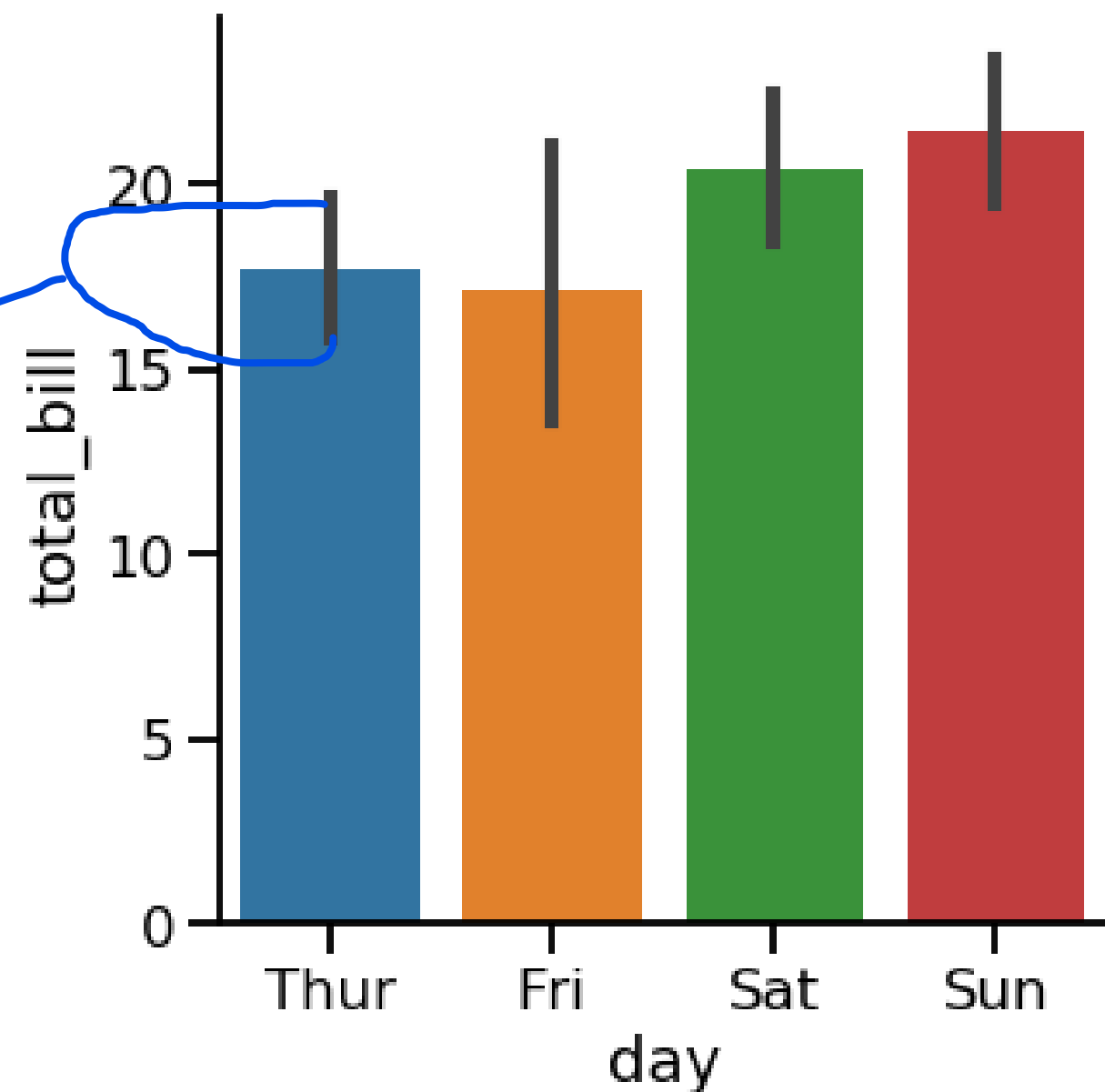
95% confidence interval as an error bar (by default)



[1] Waskom, M. L. (2021). seaborn: statistical data visualization. https://seaborn.pydata.org/

# Confidence intervals

- Lines show 95% confidence intervals for the mean

  by default ci=95

- Shows uncertainty about our estimate

- Assumes our data is a random sample

95% chance the mean lies in the range , ci=95



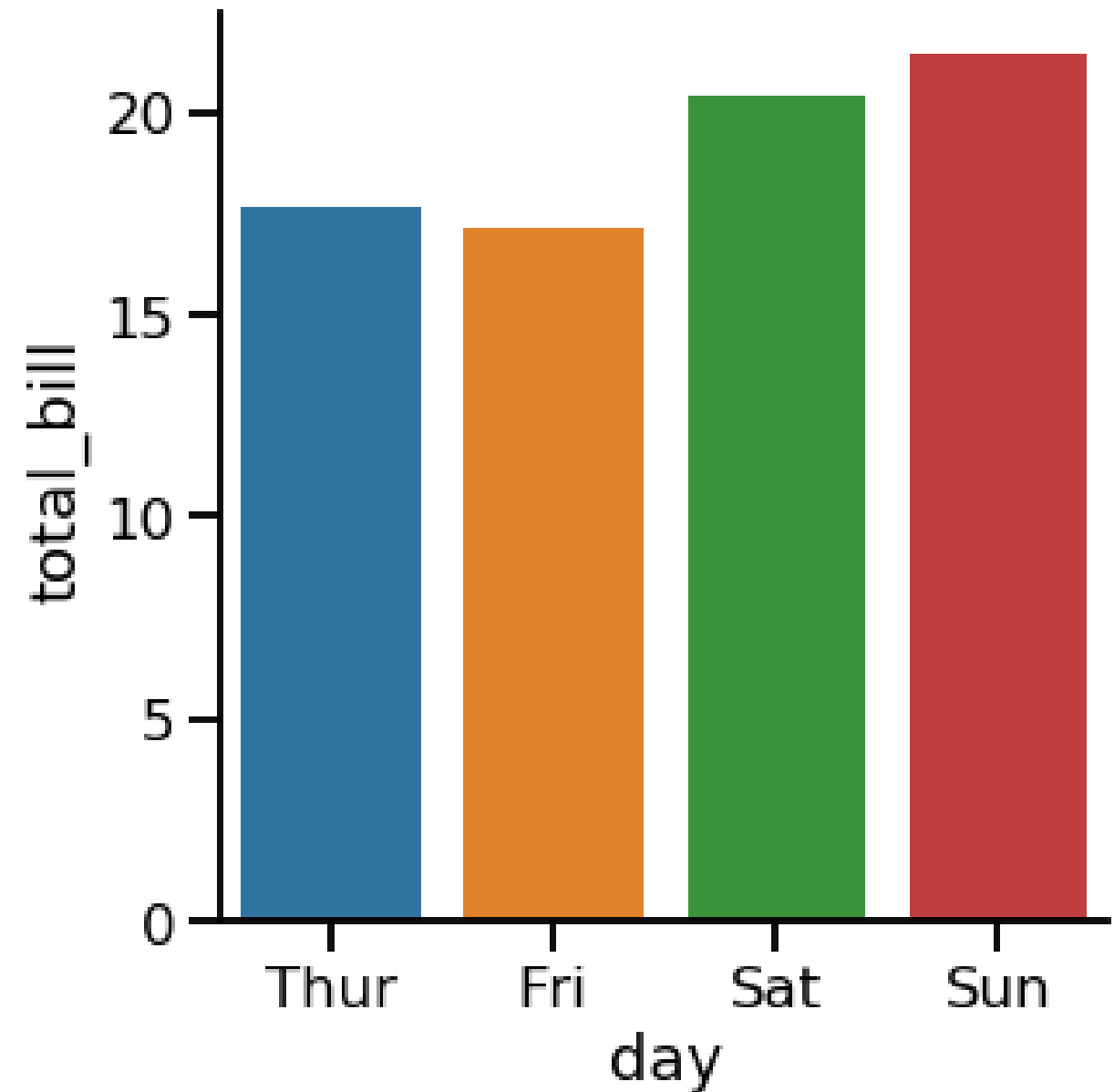[1] Waskom, M. L. (2021). seaborn: statistical data visualization. https://seaborn.pydata.org/

# Turning off confidence intervals

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

sns.catplot(x="day",
            y="total_bill",
            data=tips,
            kind="bar",
            ci=None)
plt.show()
```



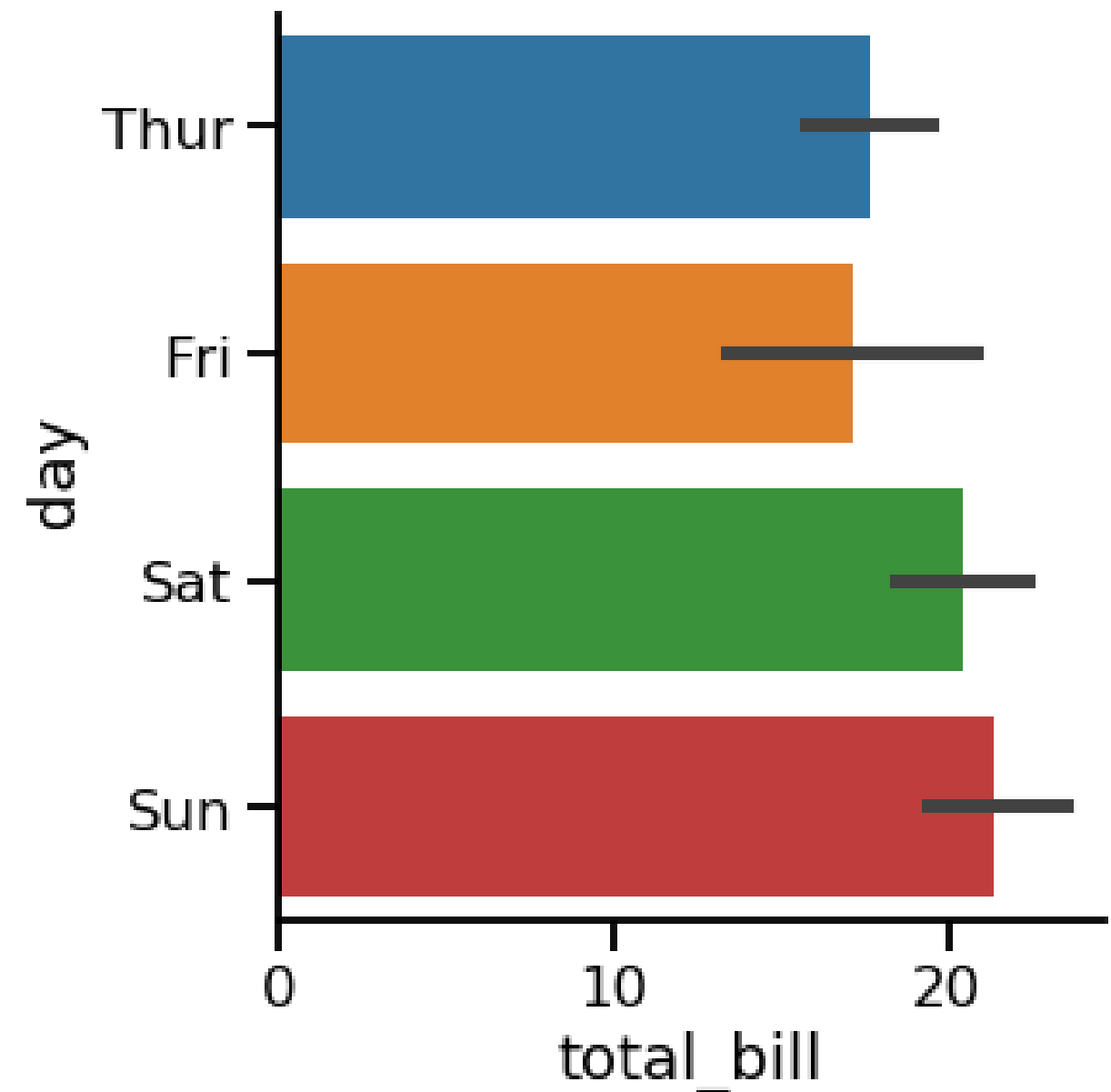[1] Waskom, M. L. (2021). seaborn: statistical data visualization. https://seaborn.pydata.org/

# Changing the orientation

one to many relationship : data (1) ->(m)total_bill

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


sns.catplot(x="total_bill",
            y="day",
            data=tips,
            kind="bar")
plt.show()
```



[1] Waskom, M. L. (2021). seaborn: statistical data visualization. https://seaborn.pydata.org/
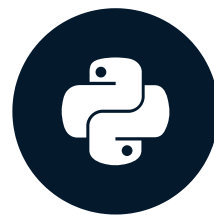
# Let's practice!

INTRODUCTION TO DATA VISUALIZATION WITH SEABORN
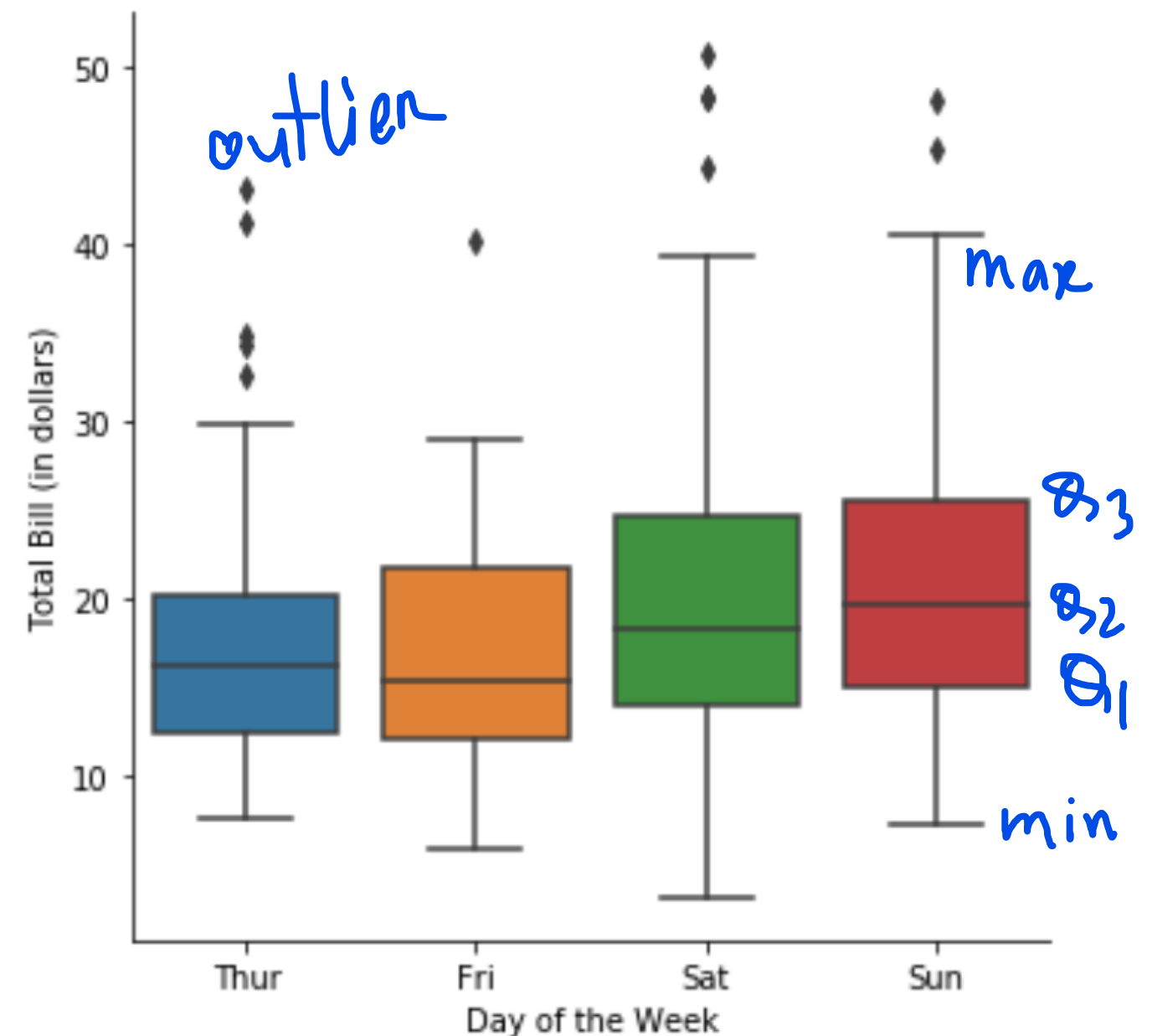
*Part - 01*

# Creating a box plot

## INTRODUCTION TO DATA VISUALIZATION WITH SEABORN

**Erin Case**
Data Scientist

# What is a box plot?

- Shows the distribution of quantitative data

- See median, spread, skewness, and outliers
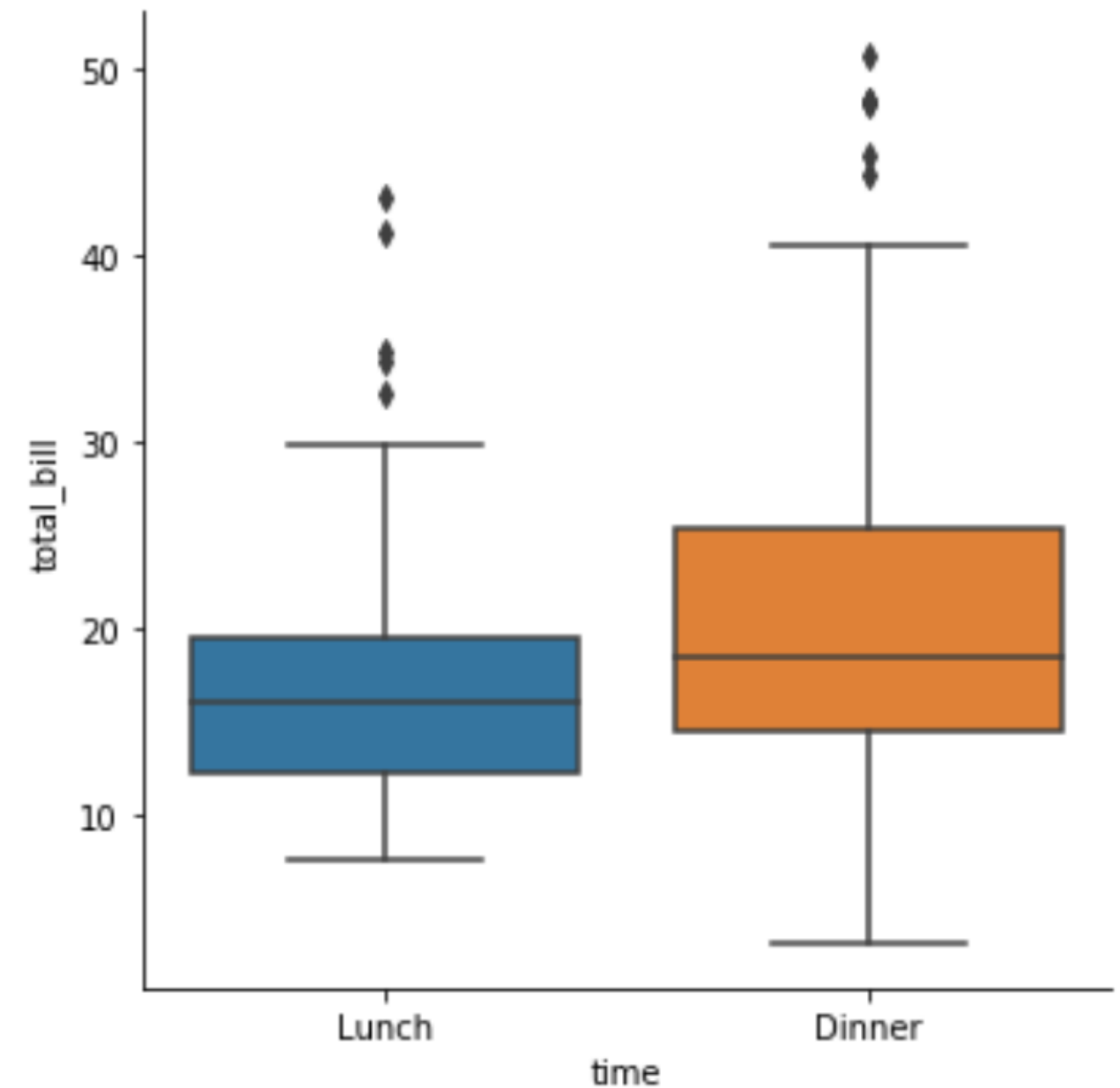
- Facilitates comparisons between groups



[1] Waskom, M. L. (2021). seaborn: statistical data visualization. https://seaborn.pydata.org/

# How to create a box plot

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

g = sns.catplot(x="time",
                y="total_bill",
                data=tips,
                kind="box")
plt.show()
```
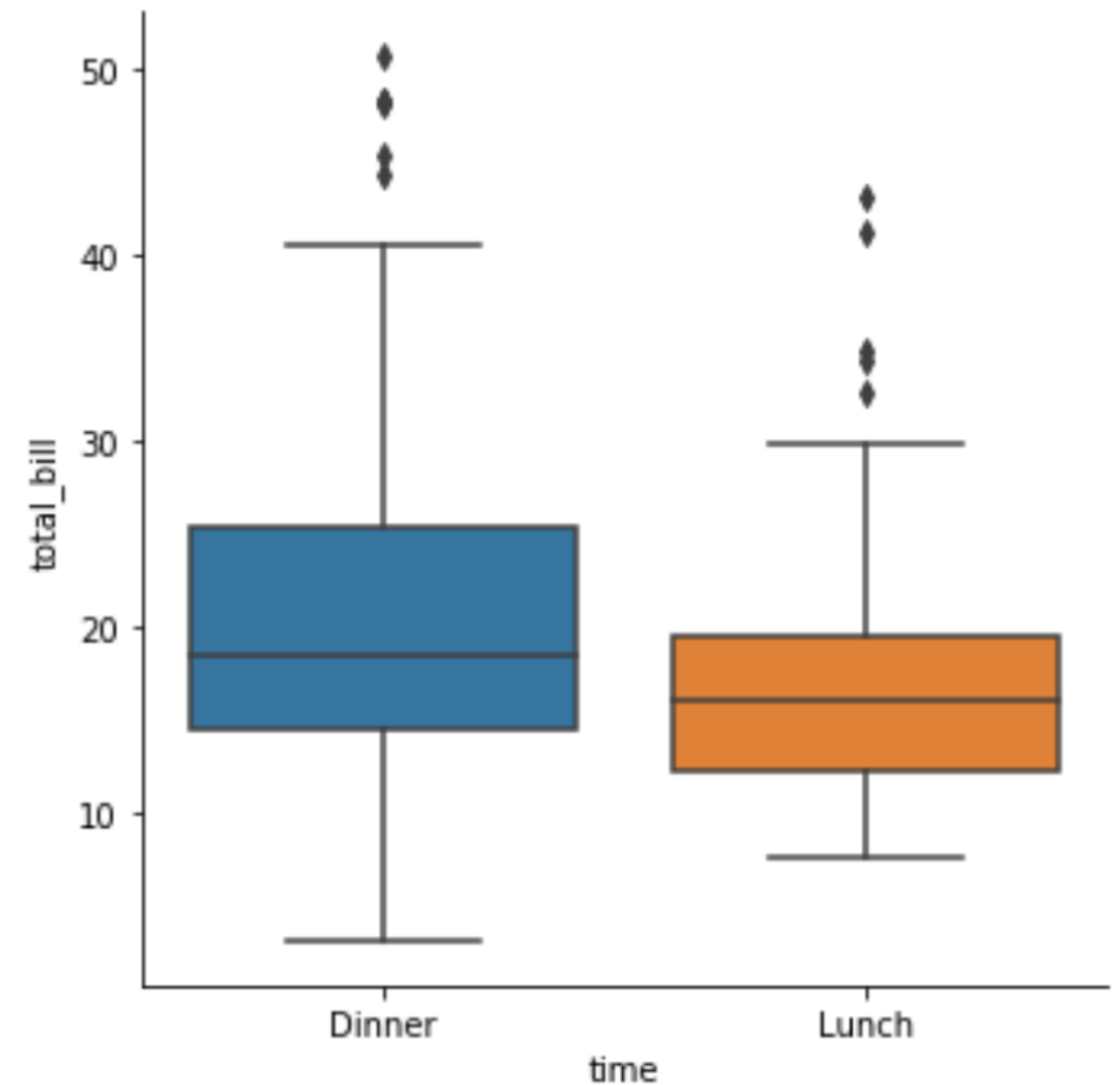


[1] Waskom, M. L. (2021). seaborn: statistical data visualization. https://seaborn.pydata.org/

# Change the order of categories

```python
import matplotlib.pyplot as plt

import seaborn as sns


g = sns.catplot(x="time",
                y="total_bill",
                data=tips,
                kind="box",
                order=["Dinner",
                       "Lunch"])

plt.show()
```



[1] Waskom, M. L. (2021). seaborn: statistical data visualization. https://seaborn.pydata.org/
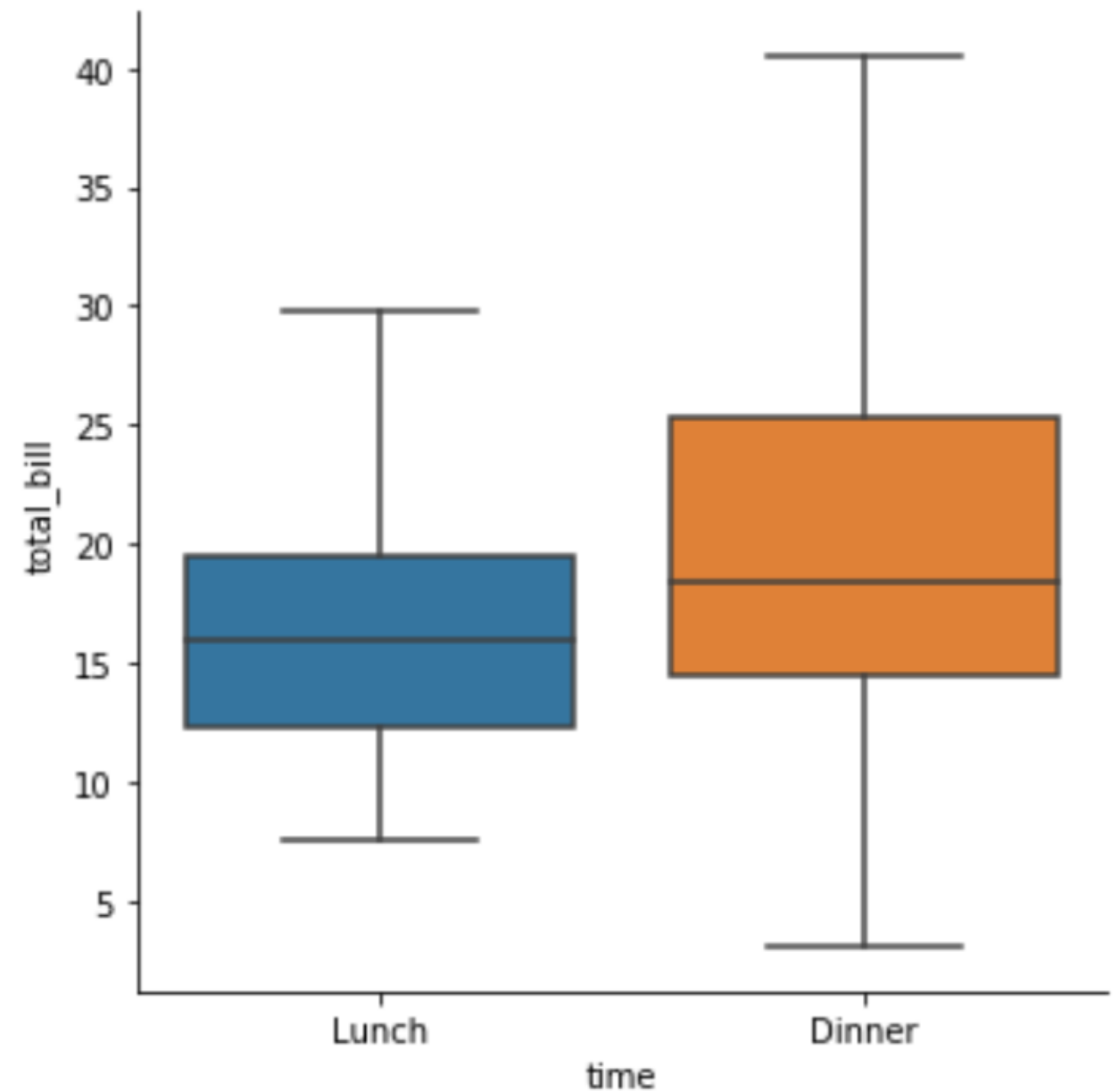
# Omitting the outliers using `sym`

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


g = sns.catplot(x="time",
                y="total_bill",
                data=tips,
                kind="box",
                sym="")
plt.show()
```

sym -> how outliers are shown :"" mean dont show



[1] Waskom, M. L. (2021). seaborn: statistical data visualization. https://seaborn.pydata.org/

# Changing the whiskers using `whis`

whis = 1.5 (default) Extend to 1.5 × IQR from Q1/Q3

- By default, the whiskers extend to 1.5 * the interquartile range

- Make them extend to 2.0 * IQR: `whis=2.0`

- Show the 5th and 95th percentiles: `whis=[5, 95]`

- Show min and max values: `whis=[0, 100]`

Whiskers extend from the minimum to the maximum value. This means: All data points are included in the whiskers — no outliers shown!

# Changing the whiskers using `whis`

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

g = sns.catplot(x="time",
                y="total_bill",
                data=tips,
                kind="box",
                whis=[0, 100])
plt.show()
```
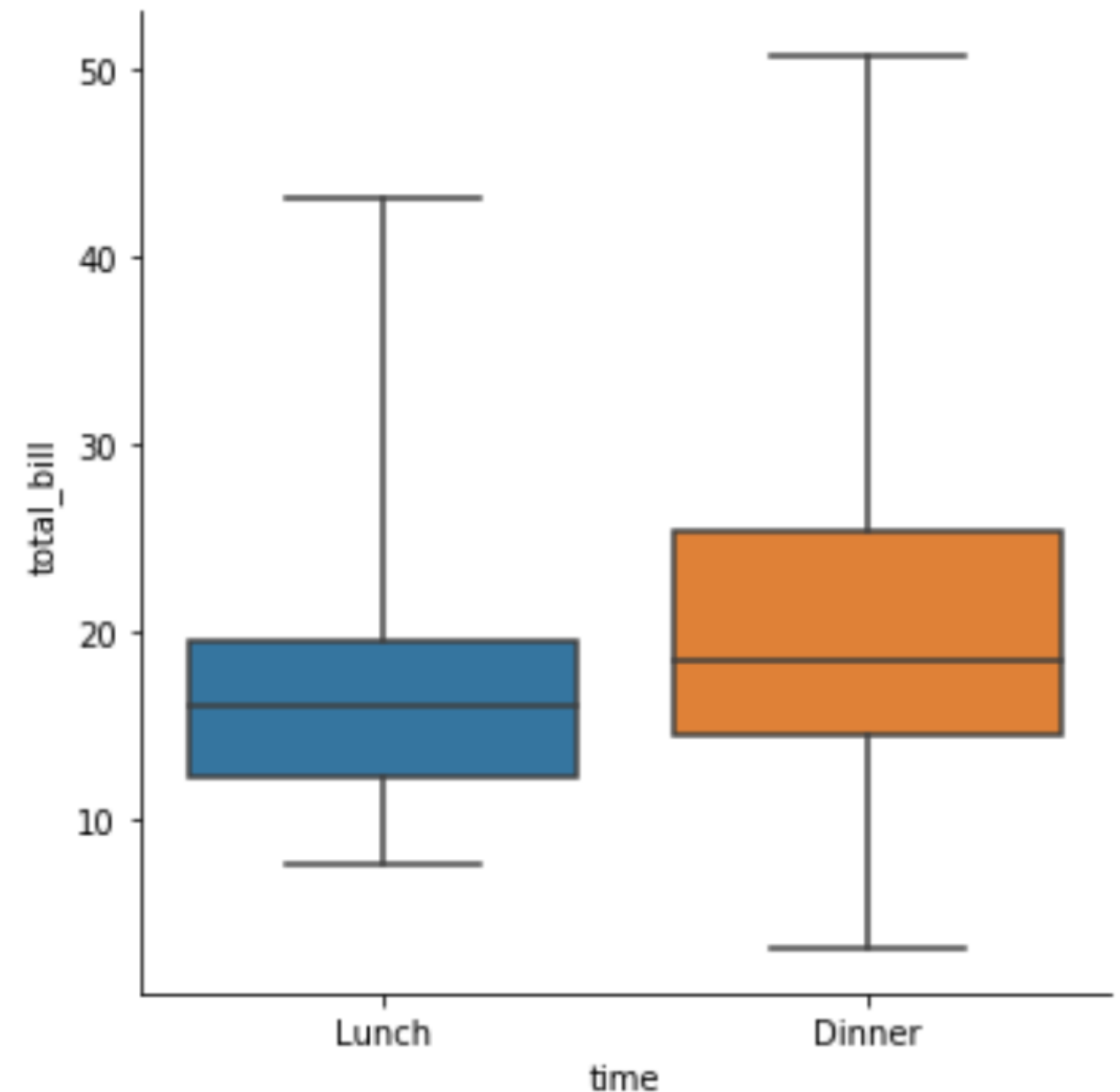
IQR=Q3–Q1

Q1 – 1.5 × IQR (if there's a data point <= that)

Q3 + 1.5 × IQR (if there's a data point >= that)



[1] Waskom, M. L. (2021). seaborn: statistical data visualization. https://seaborn.pydata.org/
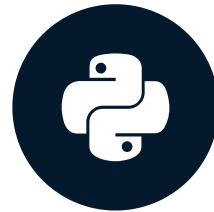
# Let's practice!

INTRODUCTION TO DATA VISUALIZATION WITH SEABORN

Part-03

# Point plots
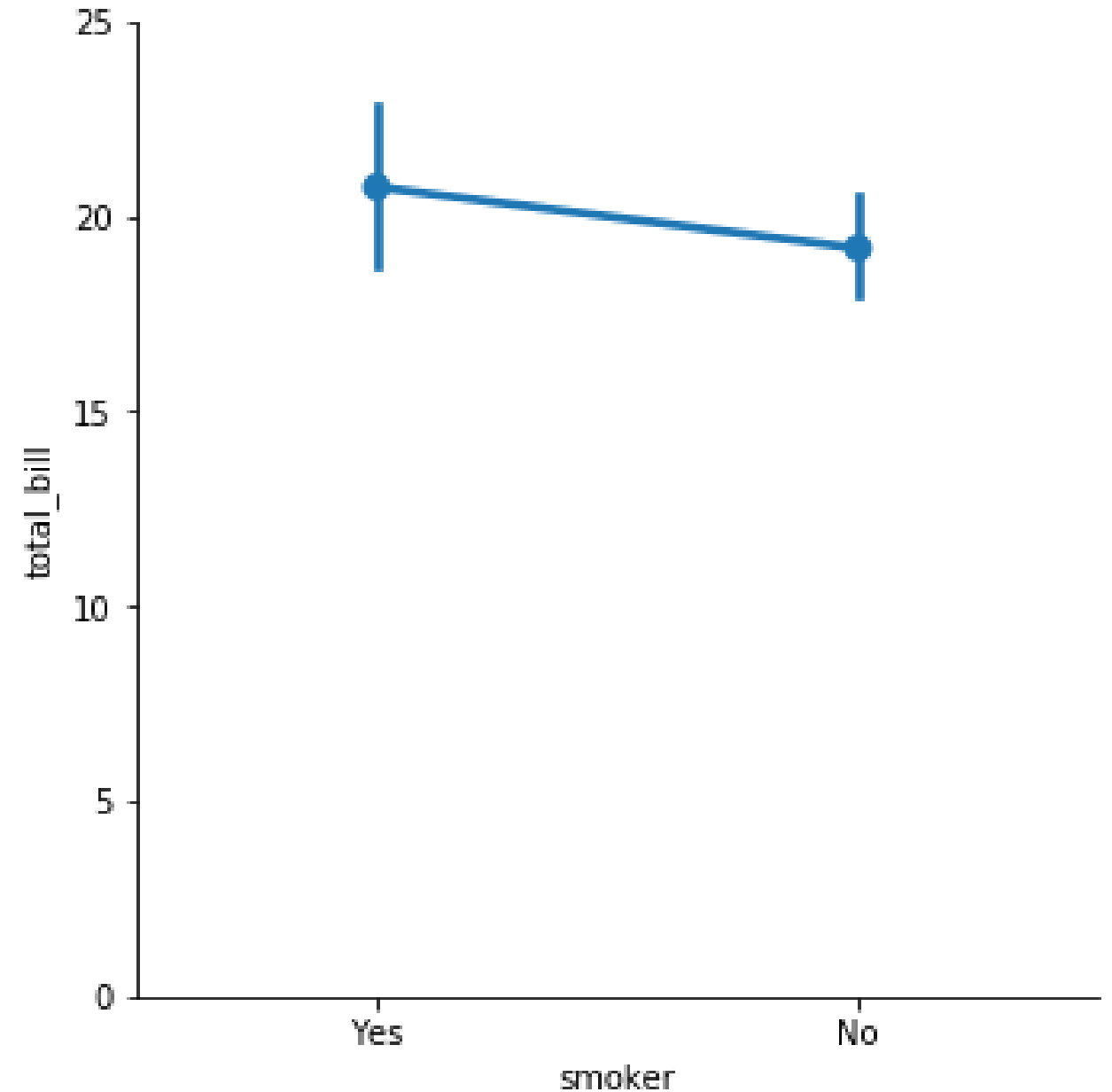
INTRODUCTION TO DATA VISUALIZATION WITH SEABORN

**Erin Case**
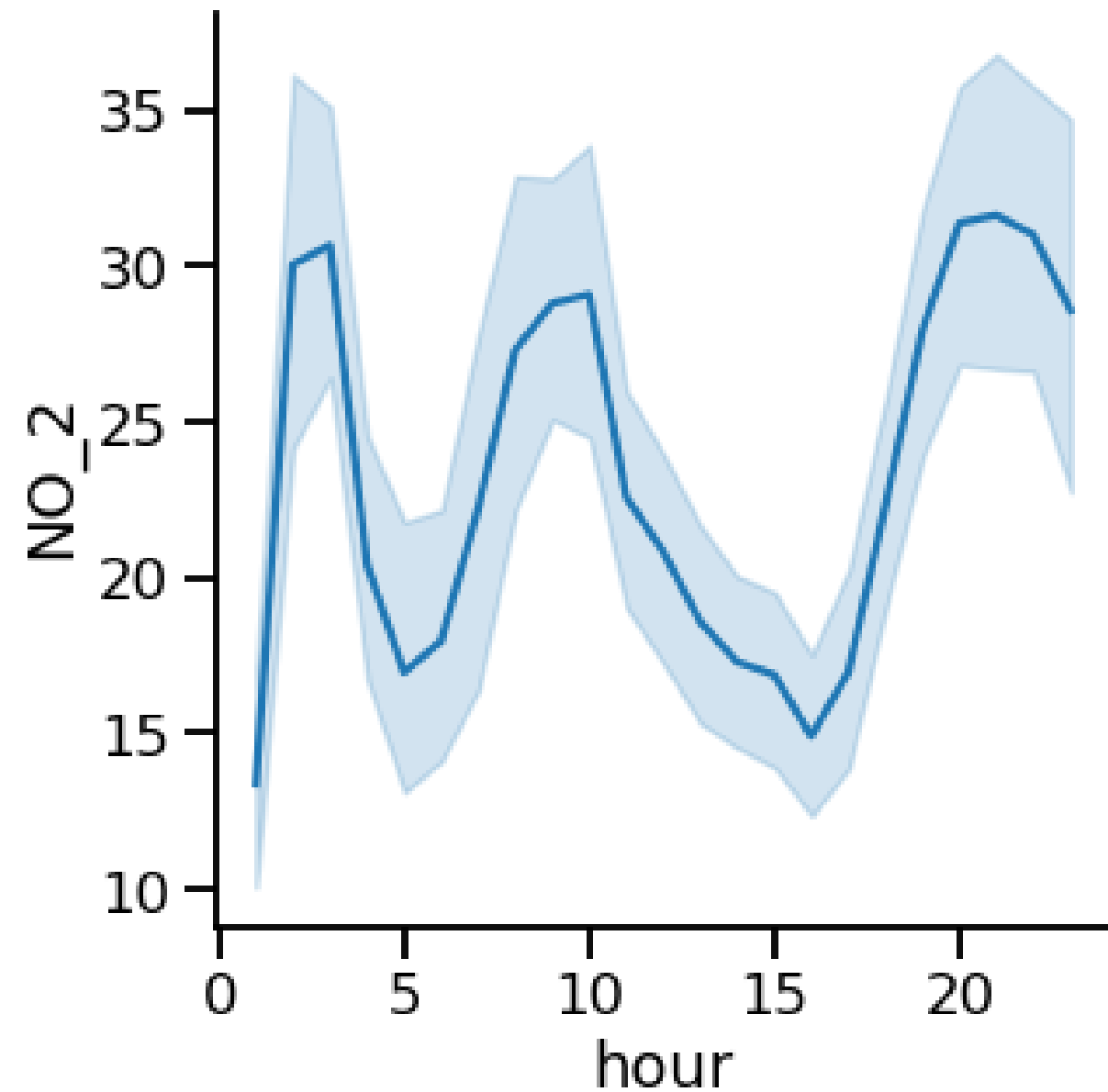Data Scientist

# What are point plots?

- Points show mean of quantitative variable

- Vertical lines show 95% confidence intervals     that mean lies in the range
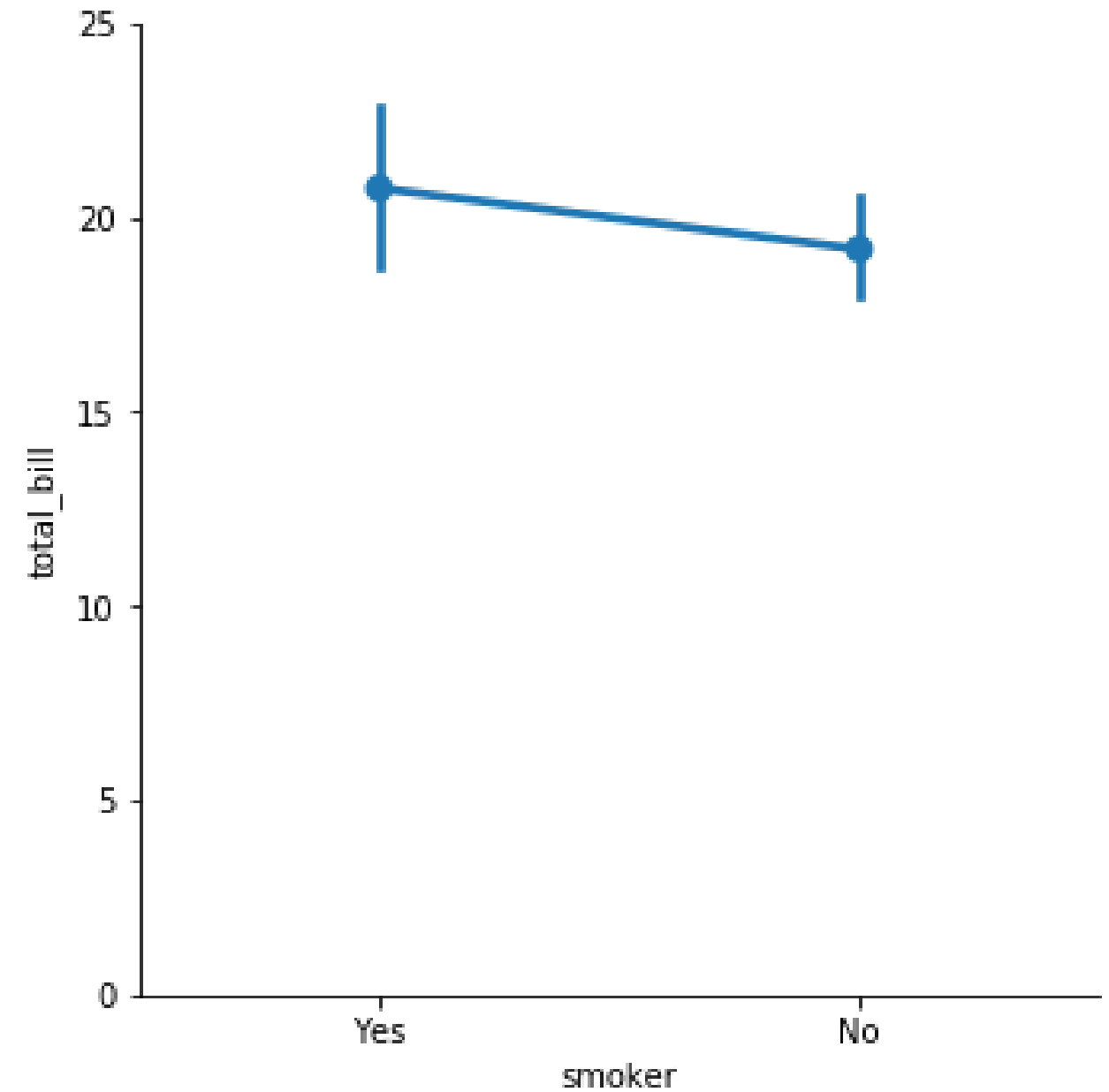
default ci=95



[1] Waskom, M. L. (2021). seaborn: statistical data visualization. https://seaborn.pydata.org/

## Line plot: average level of nitrogen dioxide over time



## Point plot: average restaurant bill, smokers vs. non-smokers



[1] Waskom, M. L. (2021). seaborn: statistical data visualization. https://seaborn.pydata.org/

# Point plots vs. line plots

Both show:

- Mean of quantitative variable

- 95% confidence intervals for the mean

Differences:

- Line plot has **quantitative** variable (usually time) on x-axis

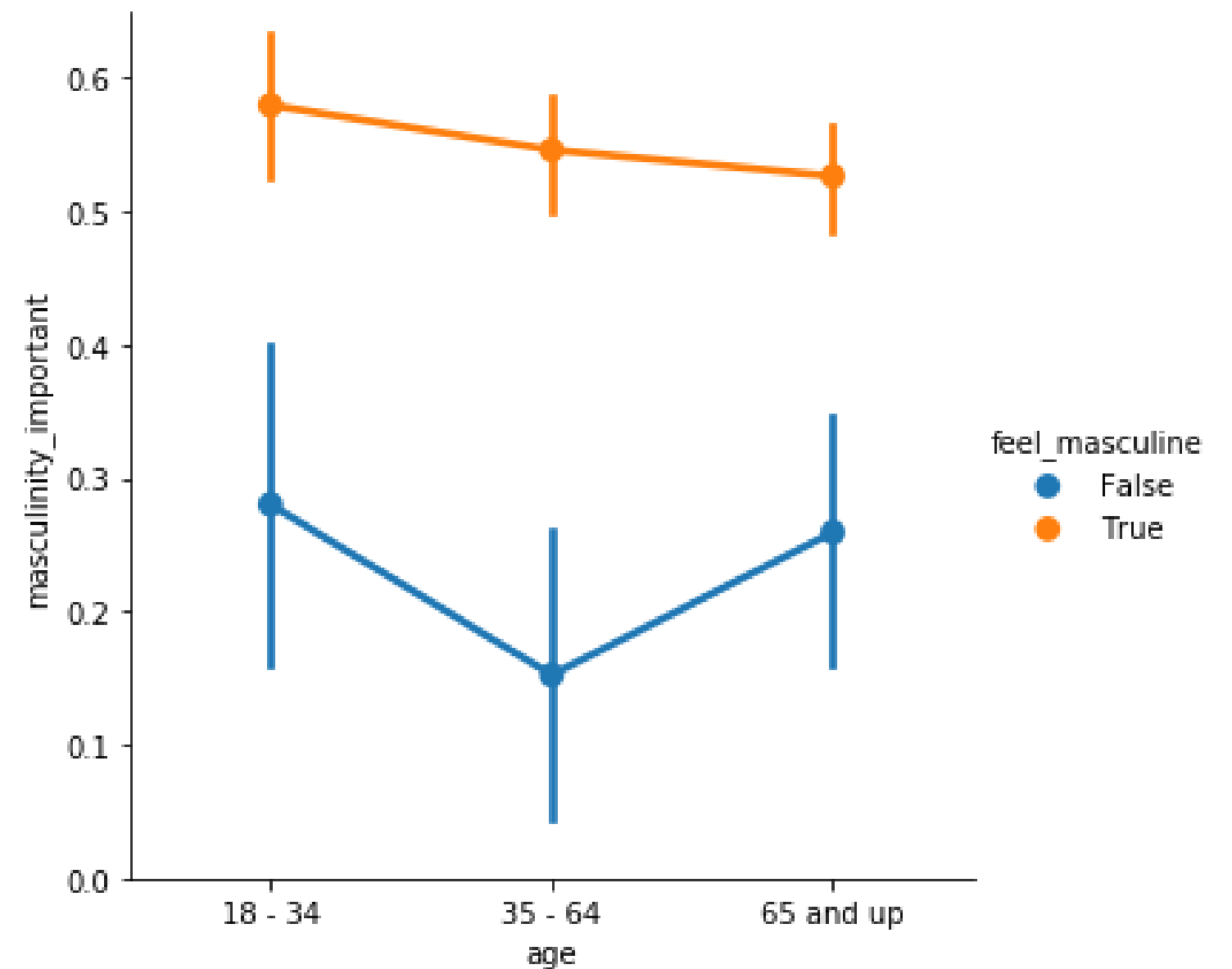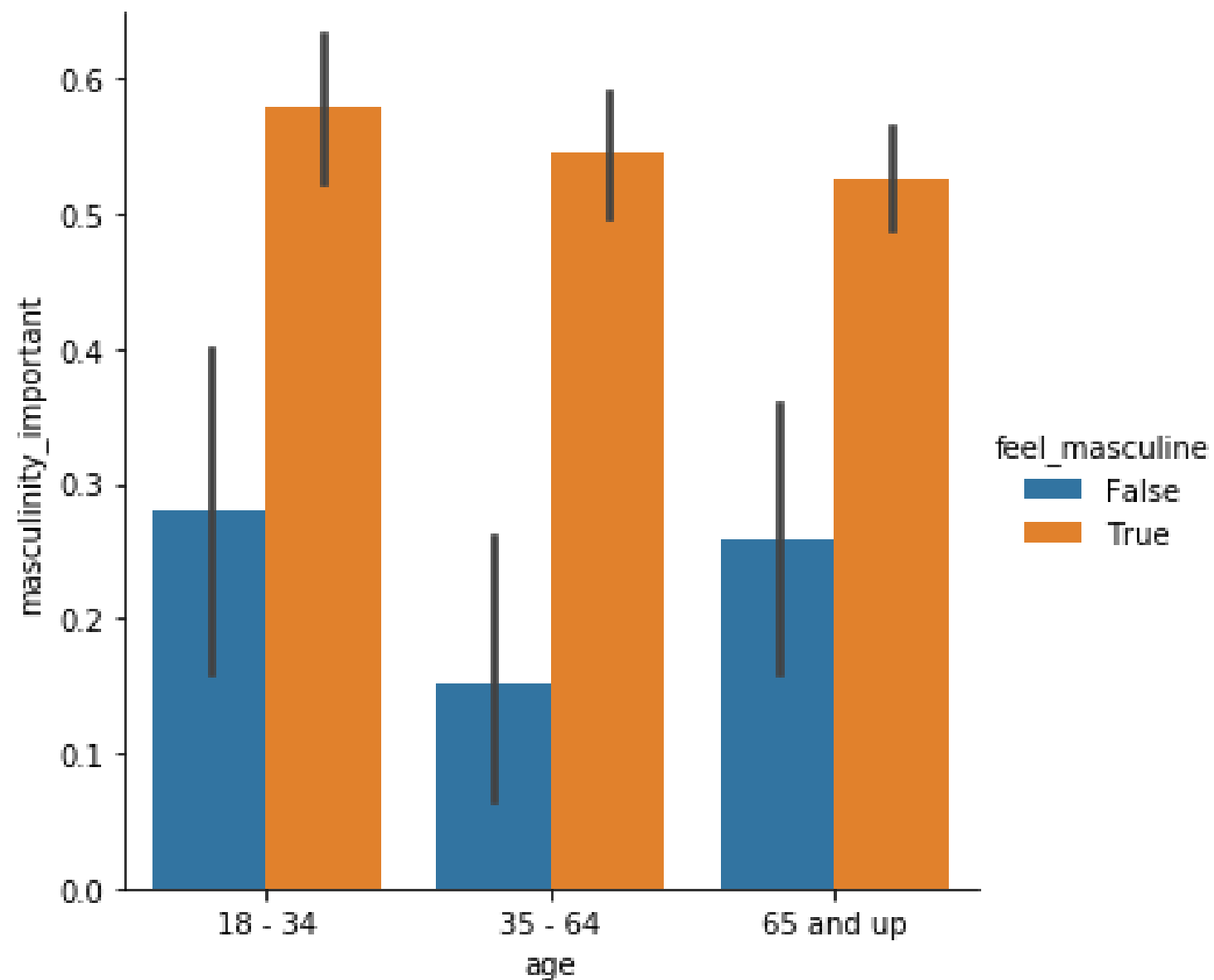- Point plot has **categorical** variable on x-axis

# Point plots vs. bar plots

Both show:

- Mean of quantitative variable

- 95% confidence intervals for the mean
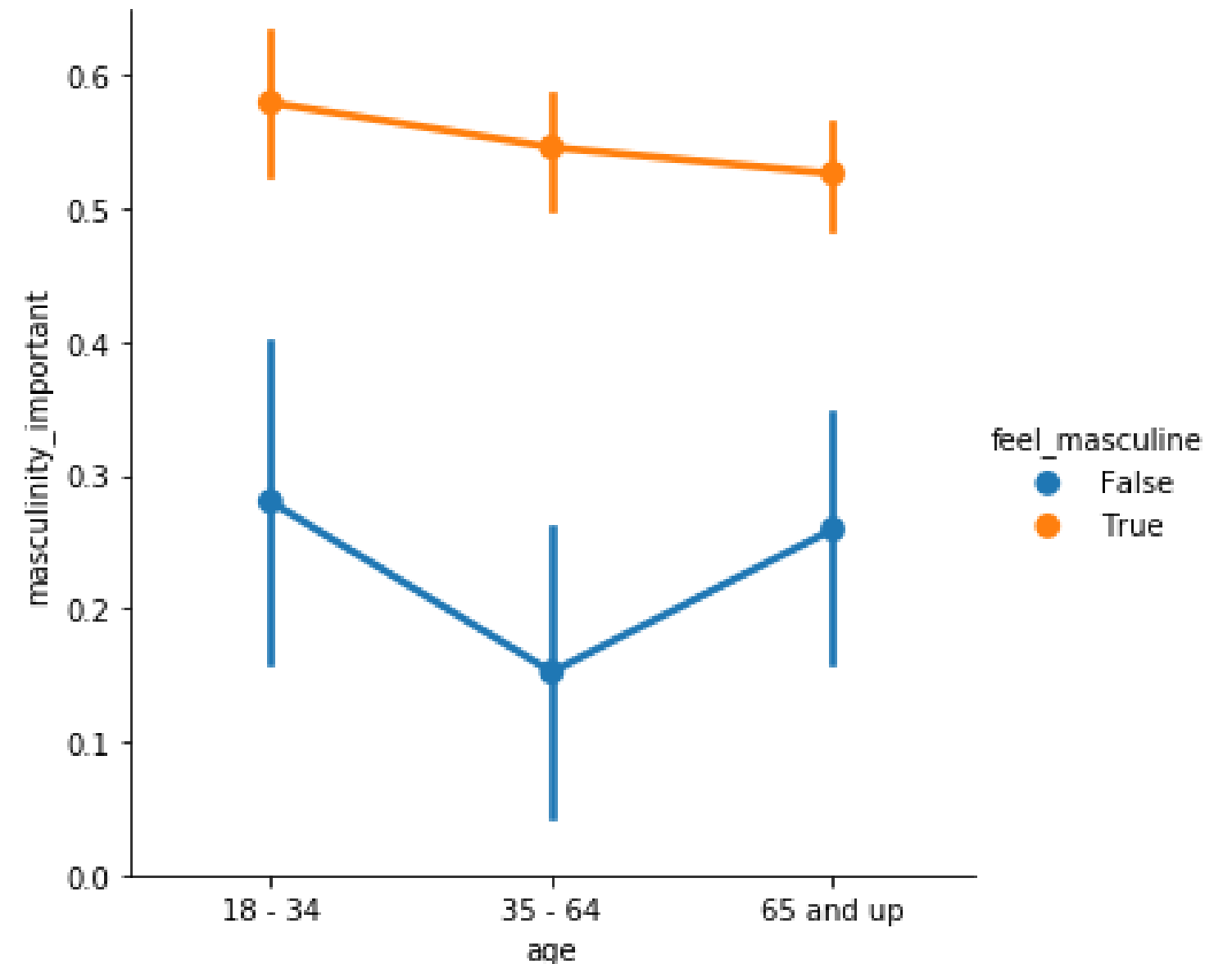
# Point plots vs. bar plots

# Creating a point plot

```python
import matplotlib.pyplot as plt

import seaborn as sns


sns.catplot(x="age",
            y="masculinity_important",
            data=masculinity_data,
            hue="feel_masculine",
            kind="point")


plt.show()
```
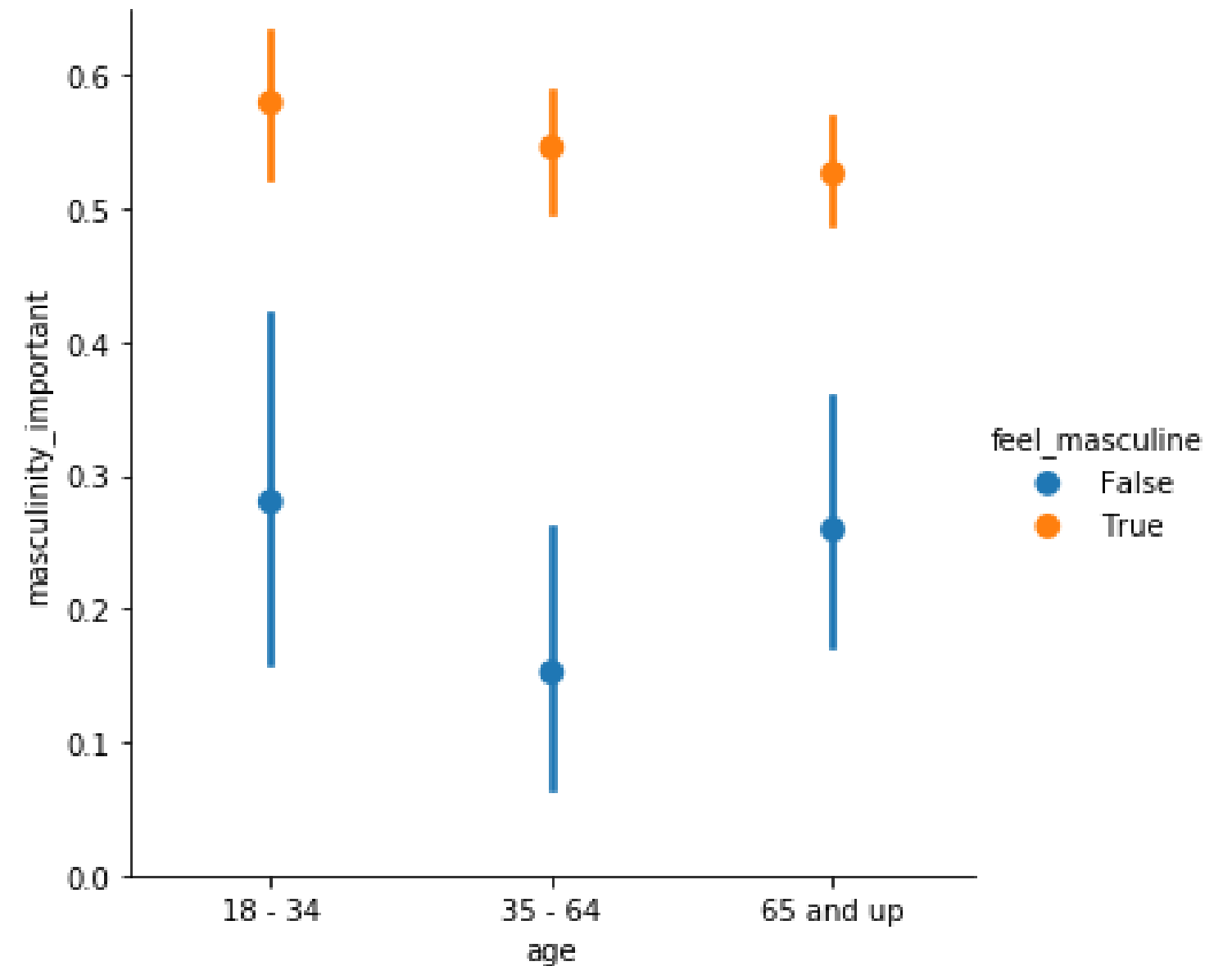
# Disconnecting the points

```python
import matplotlib.pyplot as plt

import seaborn as sns

sns.catplot(x="age",
            y="masculinity_important",
            data=masculinity_data,
            hue="feel_masculine",
            kind="point",
            join=False)

plt.show()
```
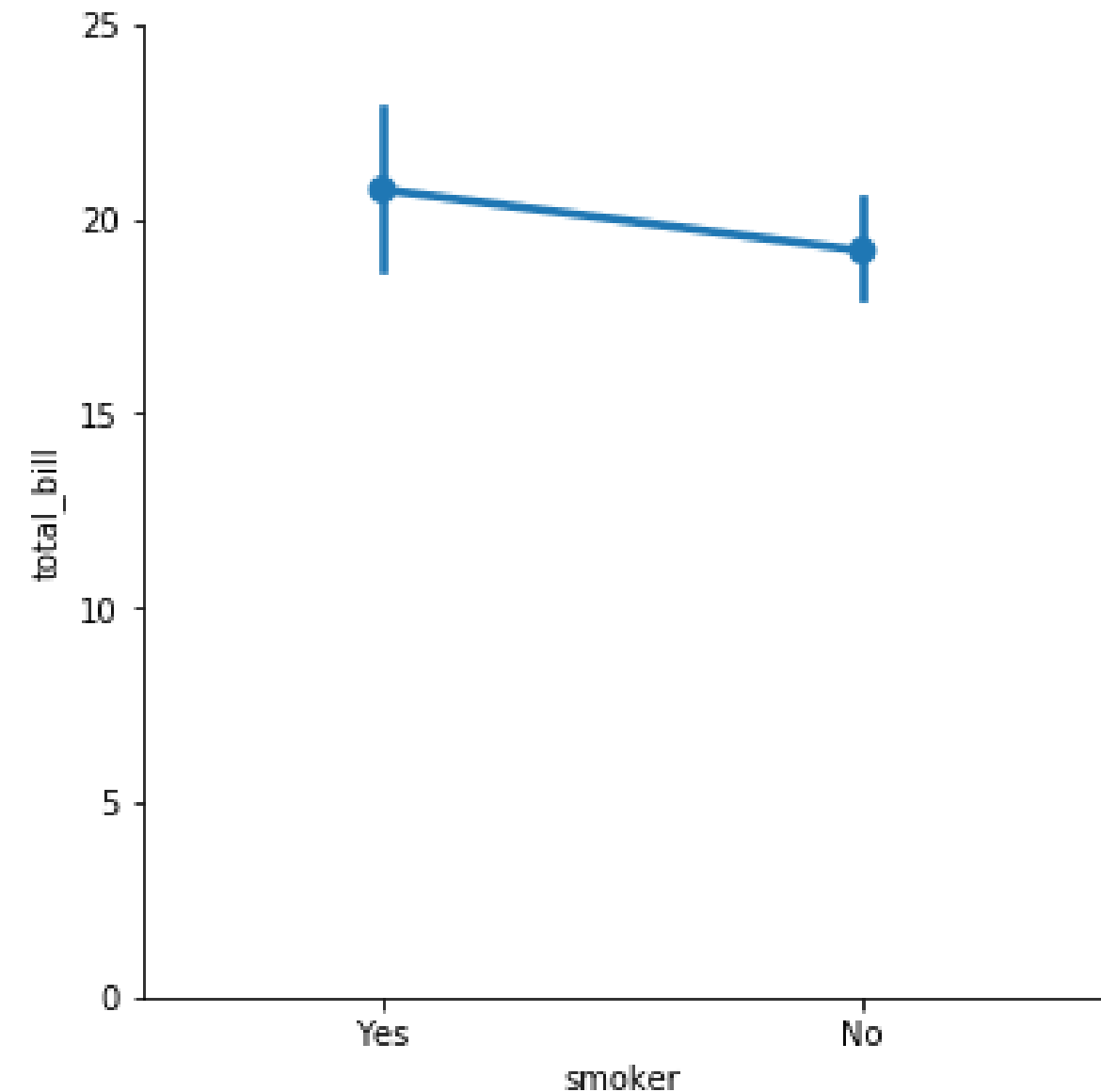
# Displaying the median

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


sns.catplot(x="smoker",
            y="total_bill",
            data=tips,
            kind="point")


plt.show()
```



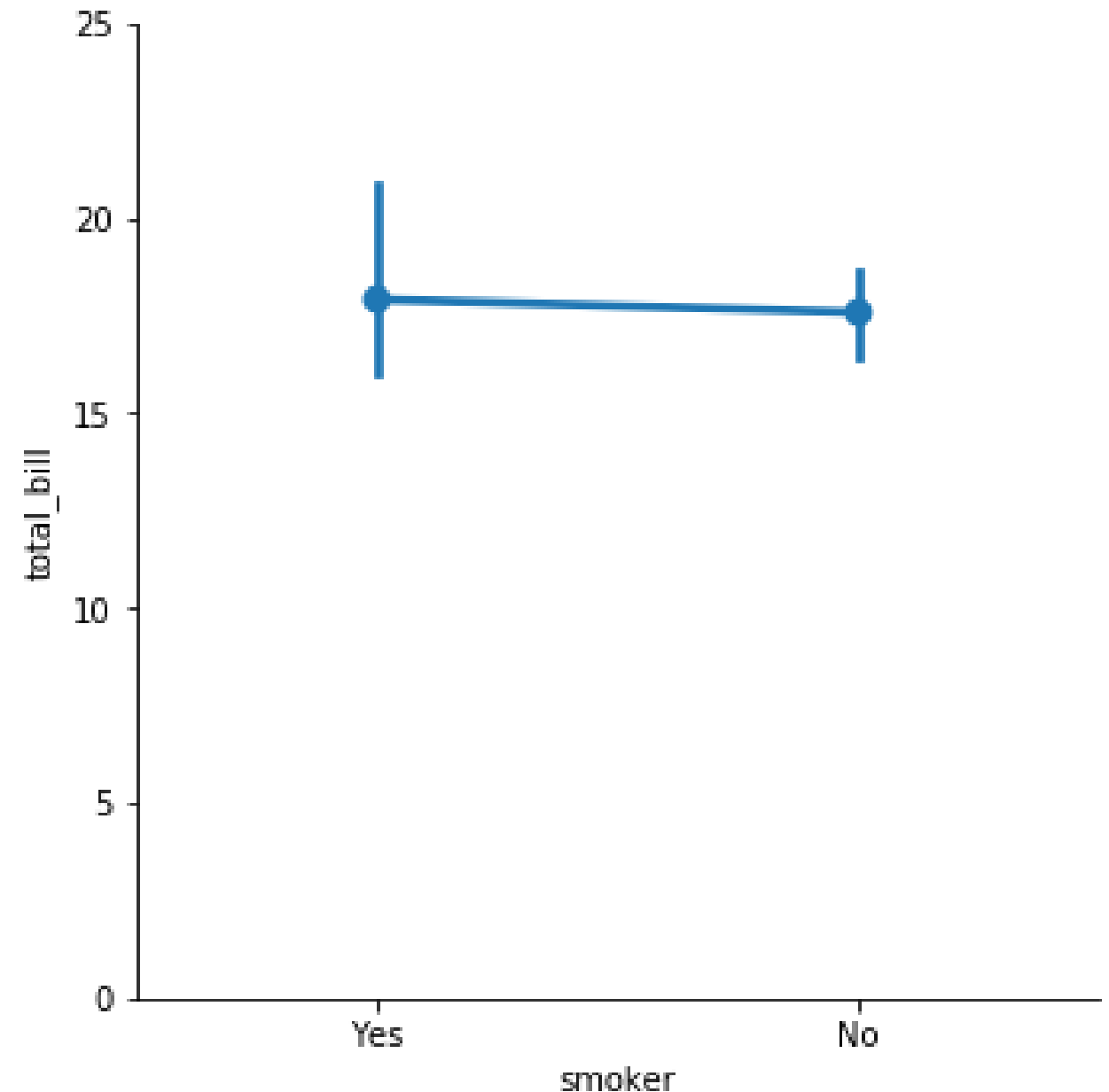[1] Waskom, M. L. (2021). seaborn: statistical data visualization. https://seaborn.pydata.org/

# Displaying the median

```python
import matplotlib.pyplot as plt
import seaborn as sns
from numpy import median


sns.catplot(x="smoker",
            y="total_bill",
            data=tips,
            kind="point",
            estimator=median)



plt.show()
```
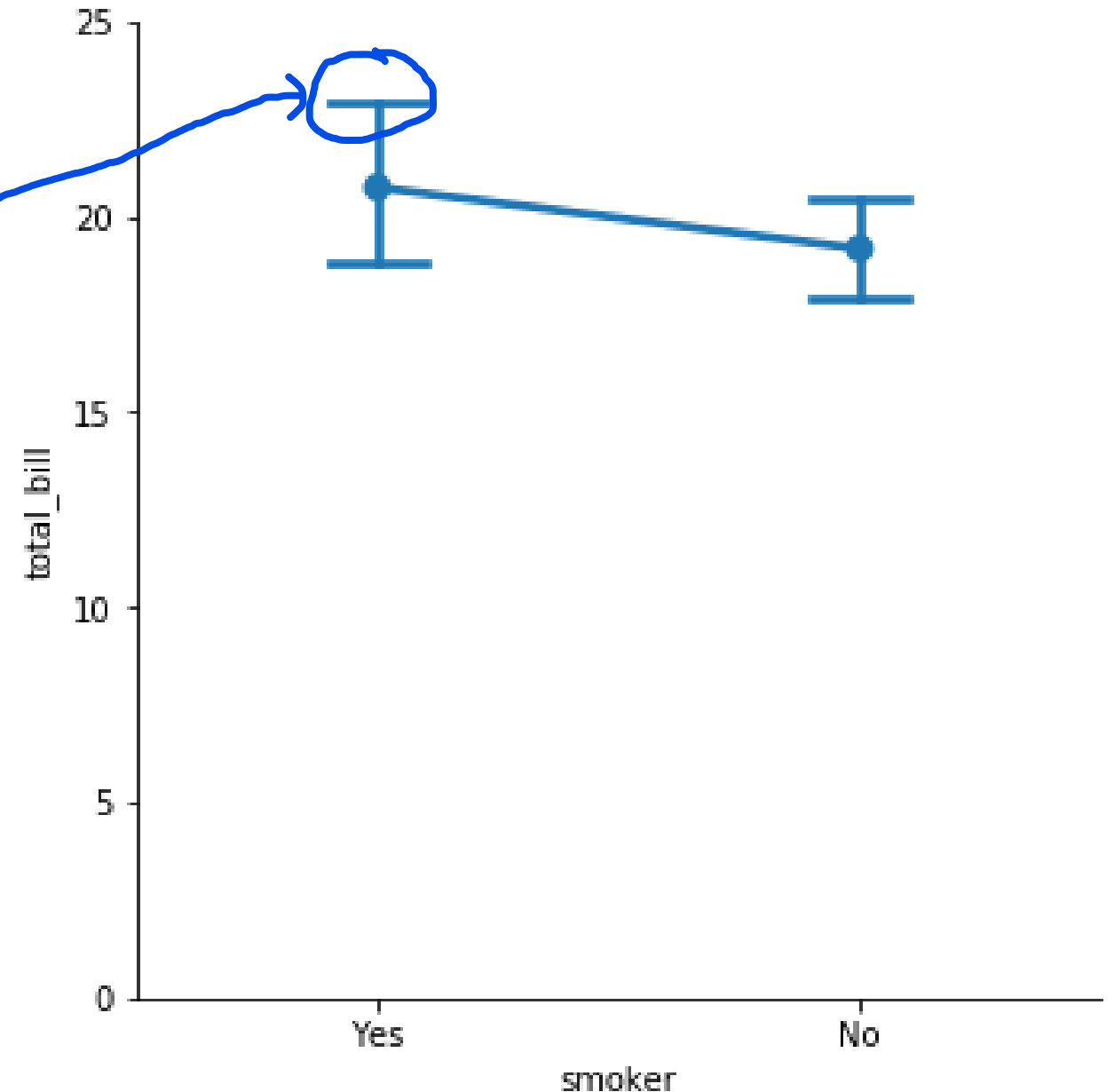


[1] Waskom, M. L. (2021). seaborn: statistical data visualization. https://seaborn.pydata.org/

# Customizing the confidence intervals

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

sns.catplot(x="smoker",
            y="total_bill",
            data=tips,
            kind="point",
            capsize=0.2)
plt.show()
```

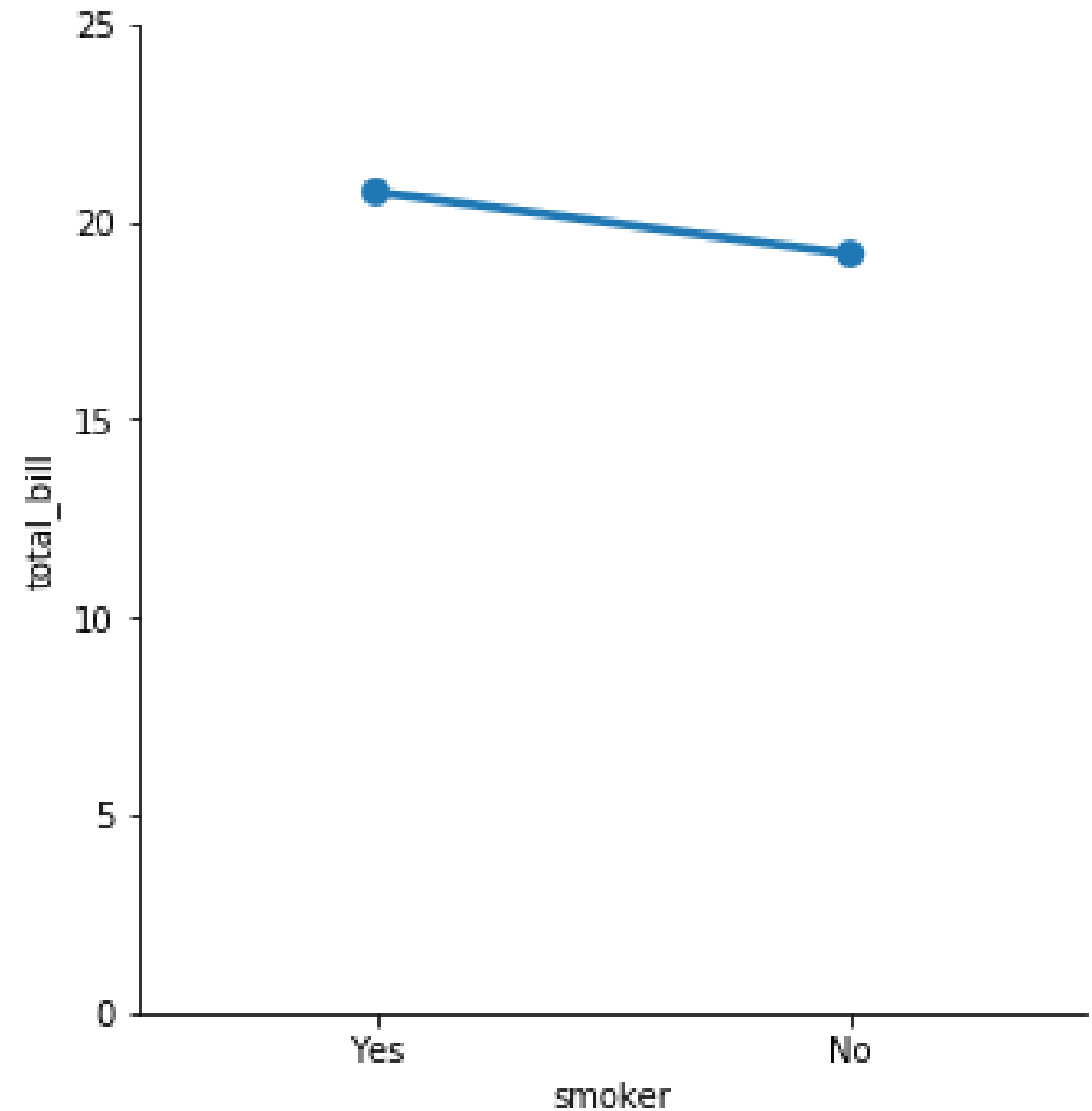Without capsize, the default is no caps (just plain vertical lines).



[1] Waskom, M. L. (2021). seaborn: statistical data visualization. https://seaborn.pydata.org/

# Turning off confidence intervals

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

sns.catplot(x="smoker",
            y="total_bill",
            data=tips,
            kind="point",
            ci=None)
plt.show()
```



[1] Waskom, M. L. (2021). seaborn: statistical data visualization. https://seaborn.pydata.org/

# Let's practice!

INTRODUCTION TO DATA VISUALIZATION WITH SEABORN

datacamp