02. Motivation for Data Visualization

Summary Statistics vs. Visualizations

Summary statistics like the mean and standard deviation can be great for attempting to quickly understand aspects of a dataset, but they can also be misleading if you make too many assumptions about how the data distribution looks.

Anscombe's Quartet Example

Consider we have the following four datasets of x, y pairs. You can download the data using the button below. A link to a Google Sheet with the data is also available here.

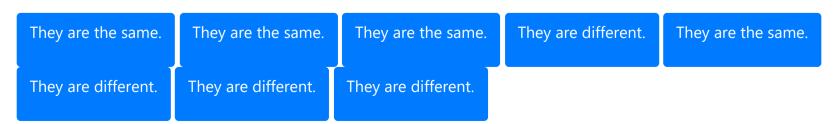
DOWNLOAD DATA

| ı | | II | | III | |
|------|-------|------|------|------|-------|
| х | у | х | У | х | у |
| 10.0 | 8.04 | 10.0 | 9.14 | 10.0 | 7.46 |
| 8.0 | 6.95 | 8.0 | 8.14 | 8.0 | 6.77 |
| 13.0 | 7.58 | 13.0 | 8.74 | 13.0 | 12.74 |
| 9.0 | 8.81 | 9.0 | 8.77 | 9.0 | 7.11 |
| 11.0 | 8.33 | 11.0 | 9.26 | 11.0 | 7.81 |
| 14.0 | 9.96 | 14.0 | 8.10 | 14.0 | 8.84 |
| 6.0 | 7.24 | 6.0 | 6.13 | 6.0 | 6.08 |
| 4.0 | 4.26 | 4.0 | 3.10 | 4.0 | 5.39 |
| 12.0 | 10.84 | 12.0 | 9.13 | 12.0 | 8.15 |
| 7.0 | 4.82 | 7.0 | 7.26 | 7.0 | 6.42 |
| 5.0 | 5.68 | 5.0 | 4.74 | 5.0 | 5.73 |

QUIZ QUESTION:

Use the data above to match an answer to each of the following questions. (Assume rounding to 2 digits)

ANSWER CHOICES:



Question

What is true for the means associated with any of the **X** columns?

they are same 9

| What is true for the means associated with any of the $f Y$ columns? | they are same 7.5 | | | |
|---|---------------------|--|--|--|
| What is true for the standard deviation associated with any of the ${\bf X}$ columns? | they are same 3.162 | | | |
| What is true for the standard deviation associated with any of the $f Y$ columns? | they are same 1.936 | | | |
| | | | | |
| | | | | |

Next Concept