Lesson 1:
Data Visualization in Data Analysis

SEARCH

RESOURCES

CONCEPTS

1. Introduction to Data Visualization
2. Motivation for Data Visualization
3. Further Motivation
4. Exploratory vs. Explanatory Analys...
5. Quiz: Exploratory vs. Explanatory
6. Visualization in Python
7. Course Structure
8. Lesson Summary

Mentor Help
Ask a mentor on our Q&A platform

Peer Chat 2
Chat with peers and alumni

Motivation for Data Visualization

Summary Statistics vs. Visualizations

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

Anscombe's Quartet Example

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

DOWNLOAD DATA

I		II		III		IV	
x	y	x	y	x	y	x	
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6

QUIZ QUESTION

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

They are different. They are different. They are different.

They are different.

QUESTION	ANSW
What is true for the means associated with any of the **X** columns?	They are the same.
What is true for the means associated with any of the **Y** columns?	They are the same.
What is true for the standard deviation associated with any of the **X** columns?	They are the same.