



# SNS COLLEGE OF TECHNOLOGY

Coimbatore-37.

An Autonomous Institution



**COURSE NAME : 19CSE311- Data Visualization**

**Topic:** INTRODUCTION TO VISUAL ANALYTICS - CONNECTIONS AND CORRELATIONS

Ms.S.Vasuki

Assistant Professor

Department of Computer Science and Engineering



# Connections and correlations

- A **correlation** is how much two sets of data change together.
- For example, take height and weight for a set of adults.
- The taller an adult is, the heavier they are likely to be.
- As such, the two sets of data will correlate highly.



# Examples of time series analysis in action include:

- The first is a **Pearson Correlation**, which measures the **strength** of the linear relationship between variables.
- The second is a **Kendall Tau Correlation**, which measures the strength of agreement between two rankings in a dataset.



time series analy... X connections and... X Identify patterns, rel... X 5. Connections... X Learn about com... X Data Visualization... X How to Explain... X

https://medium.com/upskilling/data-visualization-part-7-correlations-differentiation-and-linear-regression-d147abd2...

Search Write

### Differences Between Correlation and Regression

learn.g2.com

| Correlation                        | Regression                       |
|------------------------------------|----------------------------------|
| 1 Relationship                     | 1 One affects the other          |
| 2 Variables move together          | 2 Cause and effect               |
| 3 x and y can be interchanged      | 3 x and y cannot be interchanged |
| 4 Data represented in single point | 4 Data represented by line       |

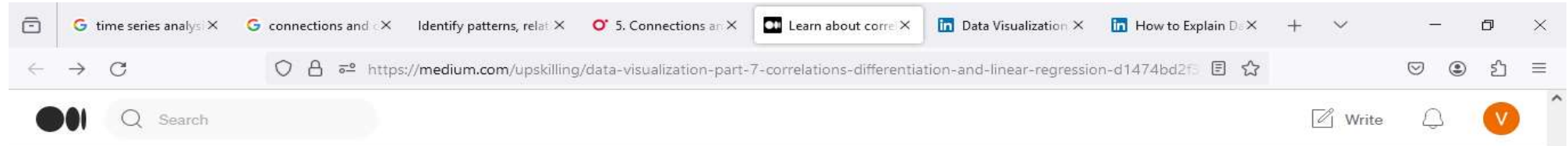
g2.com

Type here to search

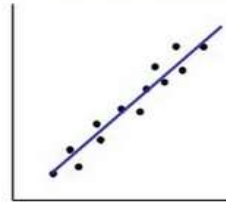
ENG 2:54 PM IN 2/19/2024



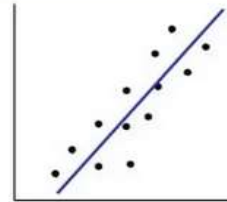
# Types:



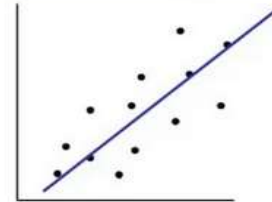
A **positive or negative** correlation is characterised by a **straight line** with a **positive /negative gradient**. The **strength** of the correlation depends on the **spread of points** around the imagined line.



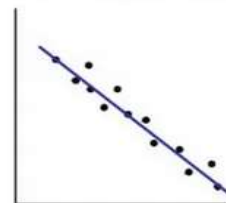
Strong Positive



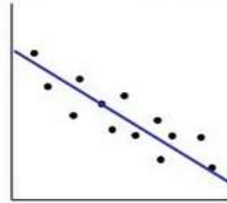
Moderate Positive



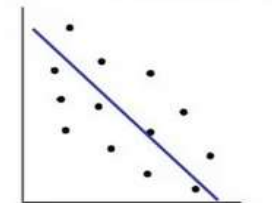
Weak Positive



Strong negative



Moderate Negative

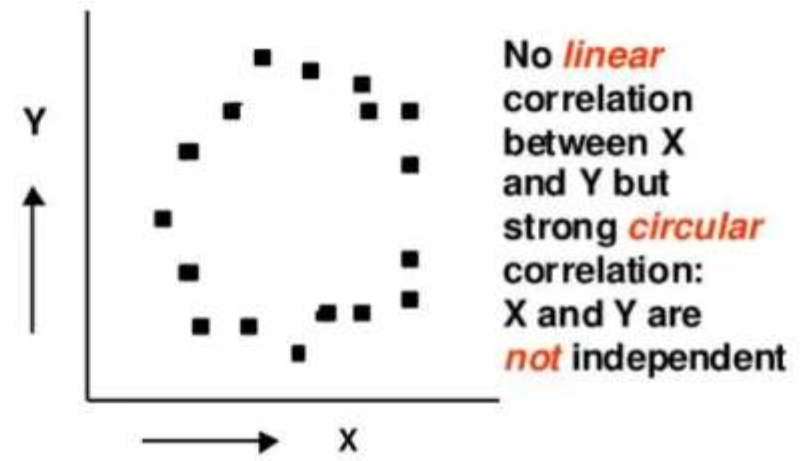
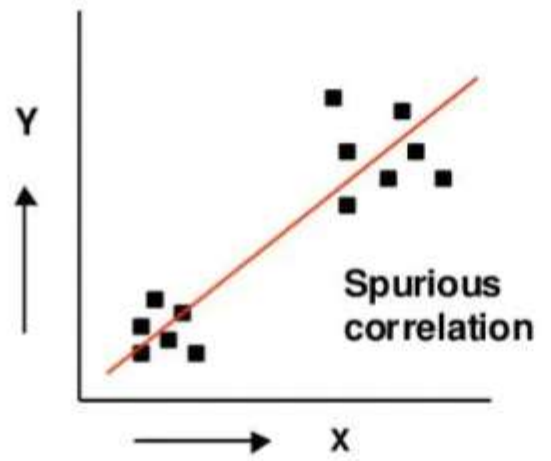
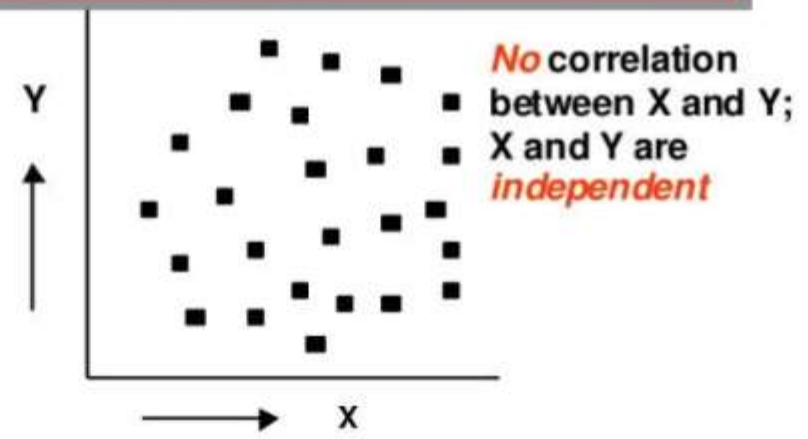
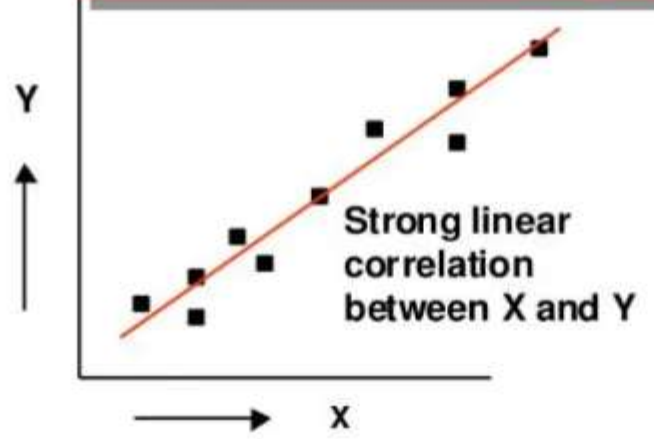


Weak negative

Correlations



## Scatter plots strong linear, no, spurious, and circular correlation





# Discrete/Continuous variables

- **Discrete variables** are meant to group things. For example, imagine you are in a doctor's office that doesn't consider more than two genders.
- In that case, genders are discrete values — assuming there are only two boxes you can tick.
- The patients in a hospital are also discrete values — they can only be one person or another.
- **Categorical variables** are usually discrete. For example, I could have a categorical variable of skin colors.
- There are several ways to measure skin color, so this will be a continuous variable by excel standards.



# Discrete/Continuous variables

- If I'm looking at a population census, though, I'd find either one race or another marked because the choices presented limit how you must identify yourself.
- Race on a census forms are discrete values, even if they have "continuous" properties you can measure.
- **Continuous variables** are essentially numbers. The age of a patient or the length of an object are examples of continuous variables.





THANK YOU