**ASSIGNMENT: 5 (09-09-25)**

**WEB SCRAPPING**

* **The web is an interface where we can store data.**
* **Scrapping means extracting.**
* **Web scraping means extracting data from the web.**
* **The Internet is compulsory.**

**Collecting data from the web**

* **The data should be in a tabular format.**
* **Power BI supports only tabular data.**
* **Then, only you can upload a web page to Power BI**
* **Static means constant**
* **Dynamic means that it will change with respect to time.**
* **The advantage of doing web scraping is not necessary to upload the data again and again. Automatically, the data will get updated.**
* **Disadvantage: the internet is compulsory.**

**WORLD HAPPINESS REPORT DOWNLOAD FROM GOOGLE.**

**[WIKIPEDIA PAGE]**

* **Copy the URL**
* **Go to GET DATA [HOME PAGE]**
* **CLICK ON THE WEB**
* **From two options: basic and advanced**
* **Go with basic**
* **Paste the URL.**
* **CLICK ON OK**
* **Select table 17**
* **Click on LOAD**
* **Double tap on table 17 and Change to WHR in the DATA pane**
* **Check the datatypes in the Table view.**
* **If any column represents geographical data, compulsory data category should be mentioned.**
* **Select the column, and go to the data category**
* **Select country/region.**
* **Whenever more numerical data is there, go with correlation**
* **Correlation is the relationship between two numerical values.**
* **A scatter plot is used to find the correlation between the values.**
* **Positive correlation:**

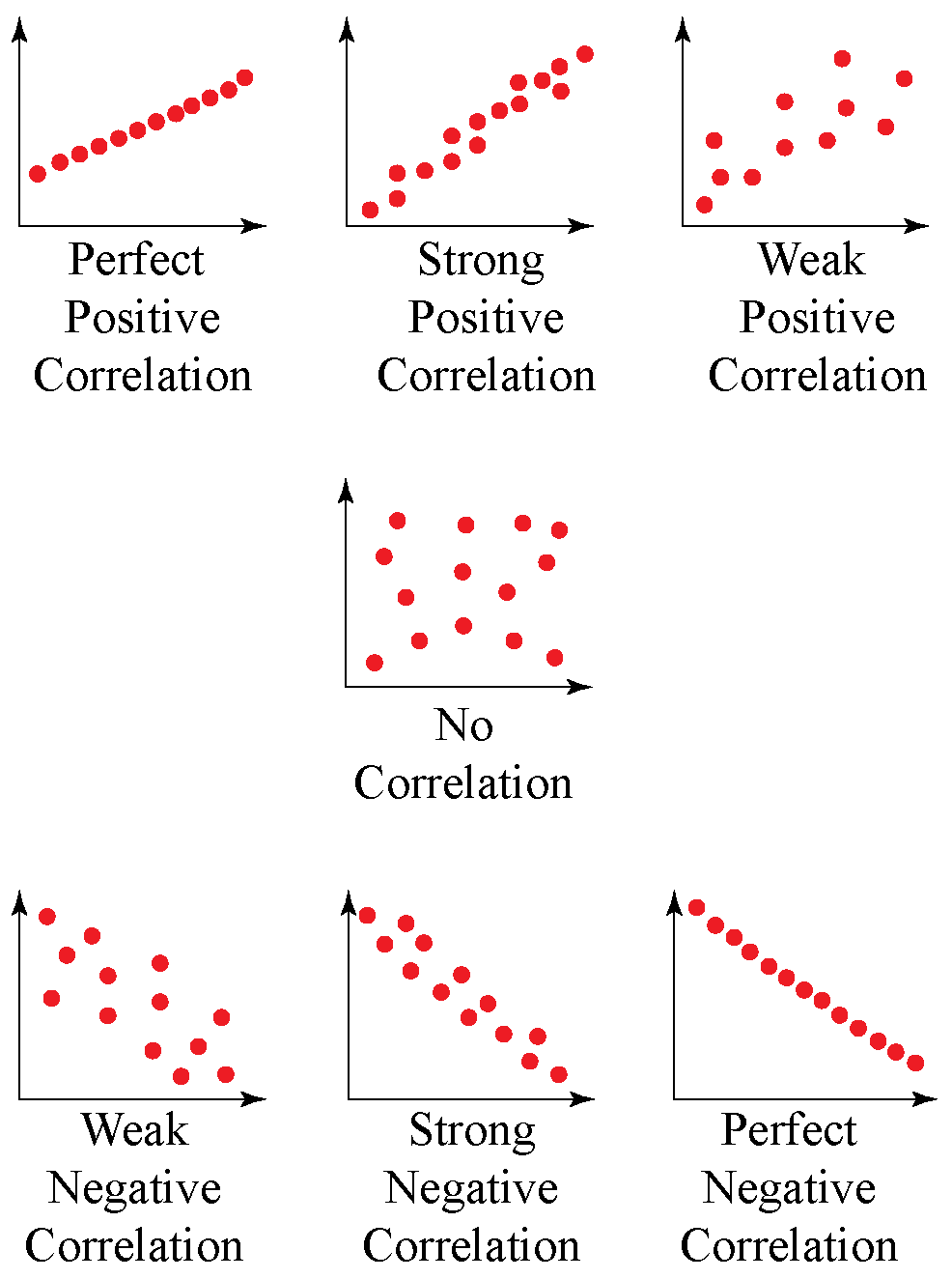
**As the temperature increases, ice cream sales tend to increase.**

* **Negative correlation:**

**As the number of hours spent studying increases, the likelihood of failing an exam decreases.**

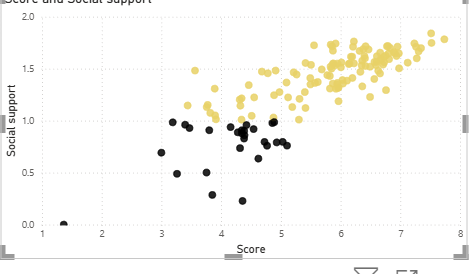
* **No correlation:**

**There is no clear linear relationship between a person’s shoe size and their height. The following correlation graphs show the examples of different ranges of values for a correlation coefficient:**

****

**A scatter plot is completely for numerical data, not for categorical data.**

* **Take a scatter plot**
* **Check the relationship between social support and score**
* **Pass score to the x-axis**
* **Pass social support to the y-axis**
* **Sum is an aggregated function [whatever value you will pass, it will combine and give one value.]**
* **A scatter plot should not be summarized by anything**
* **Click on the x-axis and y-axis, make sure they don’t summarize.**
* **Click on don’t summarize.**

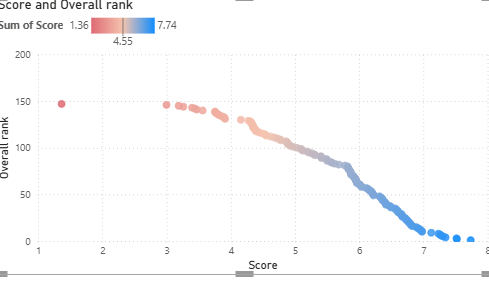
****

**INSIGHT: IT IS A WEAK POSITIVE CORRELATION.**

**Take another scatter plot**

* **Pass score to the x-axis**
* **Pass rank to the y-axis.**
* **Don’t summarize**

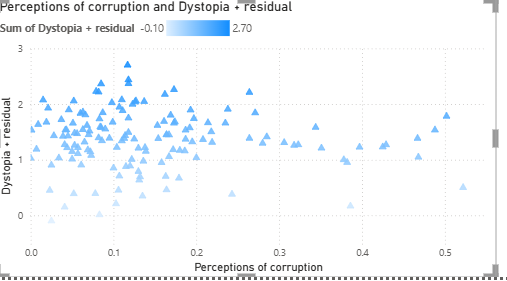
**INTERPRETATION: STRONG NEGATIVE CORRELATION**

****

**Take another scatter plot**

* **Check with the perception of corruption and dystopia+residual**
* **X axis – perception of corruption**
* **Y axis – dystopia+residual**
* **Go to format your visual**
* **Select Marker option**

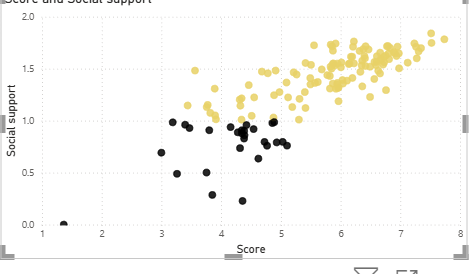
**INSIGHT: NO CORRELATION**

****

**Perform conditional formatting based on social support**

**0 to 1 and 1 to 2 for chart 1**

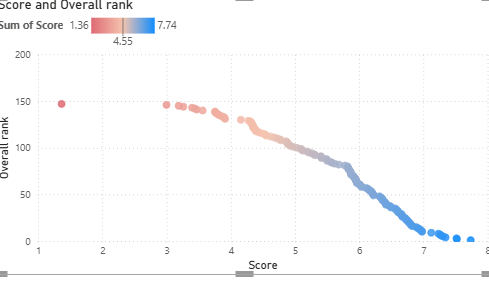
* **Click on fx**
* **Select rules based on social support.**

****

**For chart 2**

**Gradient applied for score.**

**Format Empty values – don’t format**

****