**Research Questions:**

1. Search to find 2 additional data visualization techniques. These can be related to any kinds of data, including social networking, GPS, public health, finance, etc. (include references or web links)

One type of data visualization technique is generating a time series curve in which a given metric from a dataset is analyzed over time. This metric is normally a numeric column that is either summed, averaged or counted. On a time series graph, the x-axis is usually the date field and the y-axis contains the actual value that changes. In time series analysis, we look for patterns and seasonal trends that impact the data. An example of this technique being used in the finance industry is to help predict stock prices. This can include looking at data from several previous years and help us make a prediction about the future value. (<https://www.aptech.com/blog/introduction-to-the-fundamentals-of-time-series-data-and-analysis/>)

Another type of data visualization technique is the LDA topic modeling that is used in NLP datasets. LDA topic modeling creates a graph that illustrates how closely topics are related to each other. This is mainly used in recommendation systems. An example in which LDA graph is needed can be to determine the types of articles a user reads on a given website. If the circles on the graph on relatively close to each other or if there is an overlap in circles, this usually mean that the topics are somewhat related. If they are far apart on the coordinate plane, then they are not related. (<https://www.tdktech.com/tech-talks/topic-modeling-explained-lda-to-bayesian-inference>)

2. Find an example that uses the Hamming Distance (a.k.a. “Manhattan distance”) or Jaccard Coefficient, and give your explanation / summary of how it is used. (include references or web links)

Jaccard Coefficient is a metric that is used for determining the relationships between sample sets of data. The Jaccard Coefficient is calculated by dividing the intersection of the datasets by the union of the datasets. The coefficient is usually a percentage from 0 to 100 where the higher the percentage, the more similarity. One of the drawbacks in the Jaccard Coefficient is that it does not produce accurate results on small datasets. This algorithm produces its best results when there is a large enough dataset. The Jaccard Distance is a measure of how different the sets of data are and it is measured by subtracting 100 from the Jaccard Coefficient. (<https://www.statisticshowto.com/jaccard-index/>)