**ASSIGNMENT**

**Data visualization techniques**

**Submitted to: Submitted by:**

**Ms. Jetty Benjamin ABHINAND K S**

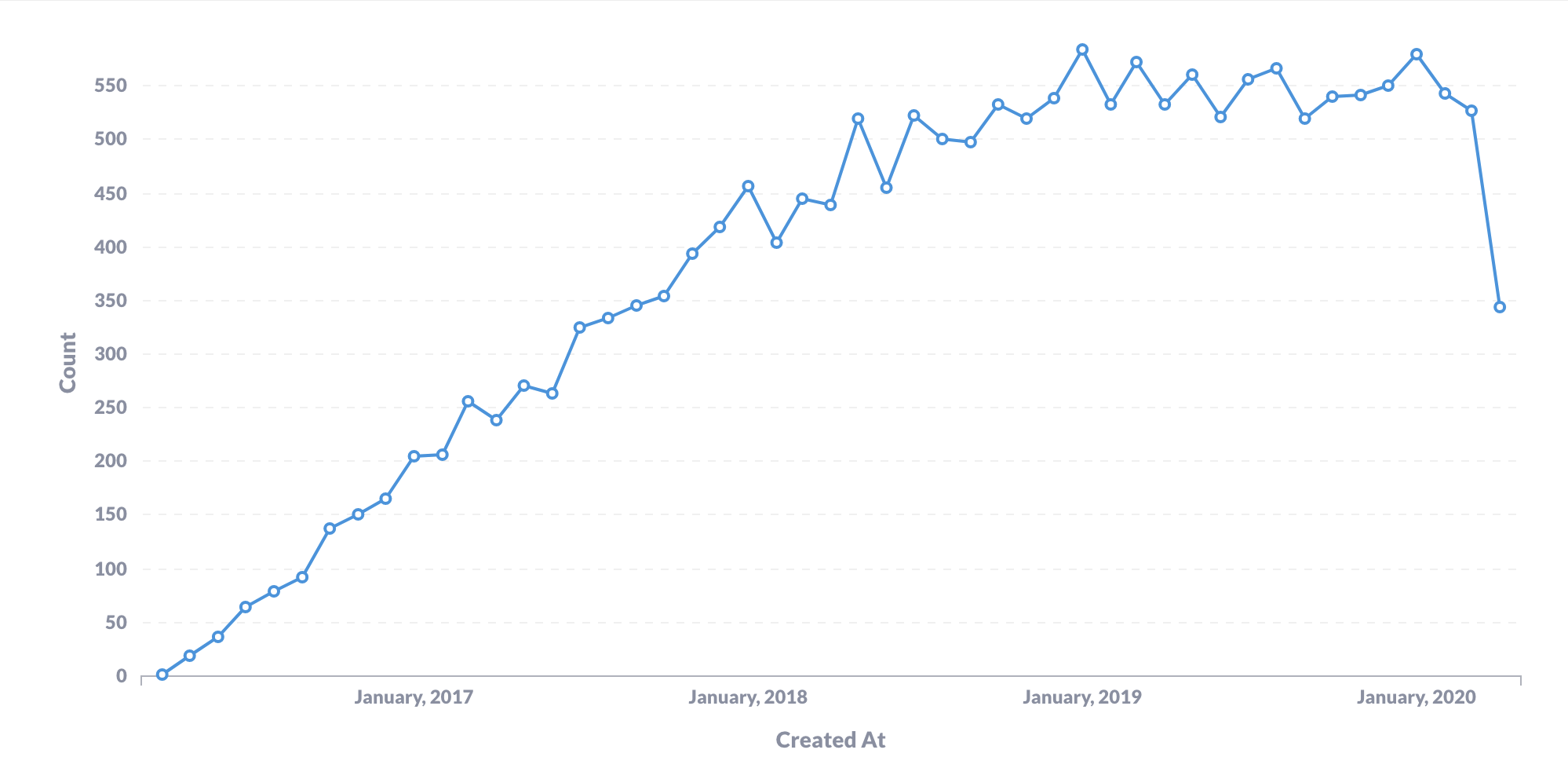
**S3-RMCA – A**

**ROLL NO - 03**

**Line plot**

Line plot is a common data visualization technique in machine learning used to display data points along a continuous interval, often representing trends, patterns, or relationships in the data. It provides valuable for visualizing various aspects:

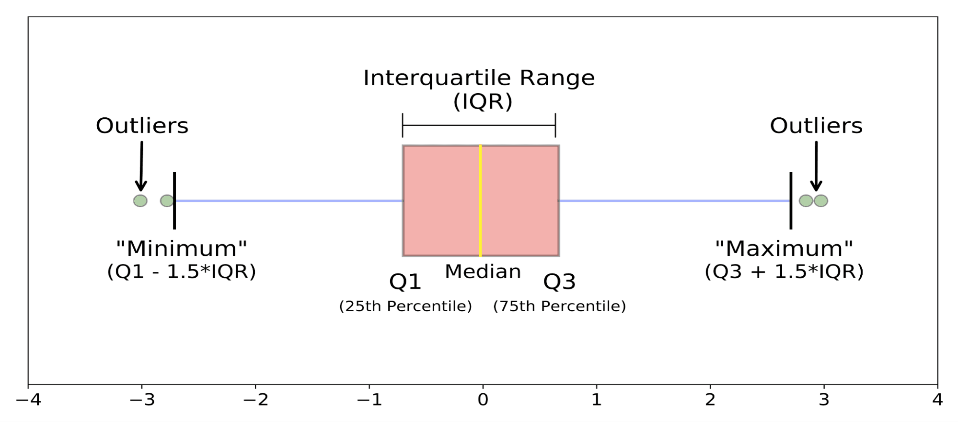
* Learning Curves: Line plots display how a model's performance metric (e.g., accuracy or loss) changes over epochs or iterations during training. This helps in understanding the learning progress and identifying issues like overfitting or underfitting.
* Feature Analysis: Line plots show the relationship between a specific feature and the target variable, helping analyze how the target variable changes concerning variations in that feature.
* Model Evaluation: Line plots compare performance metrics (such as accuracy, precision, or F1-score) of different machine learning models on the same dataset, aiding in model selection.
* Time Series Analysis: Line plots are used in time series forecasting to visualize historical data points and predicted values over time, facilitating trend analysis and forecasting accuracy assessment.



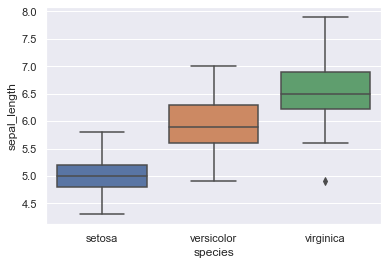
**Box and whisker plots**

Another great plot to summarize the distribution of a variable is boxplots. Boxplots provide an intuitive and compelling way to spot the following elements:

* **Median**. The middle value of a dataset where 50% of the data is less than the median and 50% of the data is higher than the median.
* **The upper quartile**. The 75th percentile of a dataset where 75% of the data is less than the upper quartile, and 25% of the data is higher than the upper quartile.
* **The lower quartile**. The 25th percentile of a dataset where 25% of the data is less than the lower quartile and 75% is higher than the lower quartile.
* **The interquartile range**. The upper quartile minus the lower quartile
* **The upper adjacent value**. Or colloquially, the “maximum.” It represents the upper quartile plus 1.5 times the interquartile range.
* **The lower adjacent value**. Or colloquially, the “minimum." It represents the lower quartile minus 1.5 times the interquartile range.
* **Outliers**. Any values above the “maximum” or below the “minimum.”



For example, the following boxplot shows the distribution of sepal length in three varieties of iris plants, drawing on the popular iris dataset.



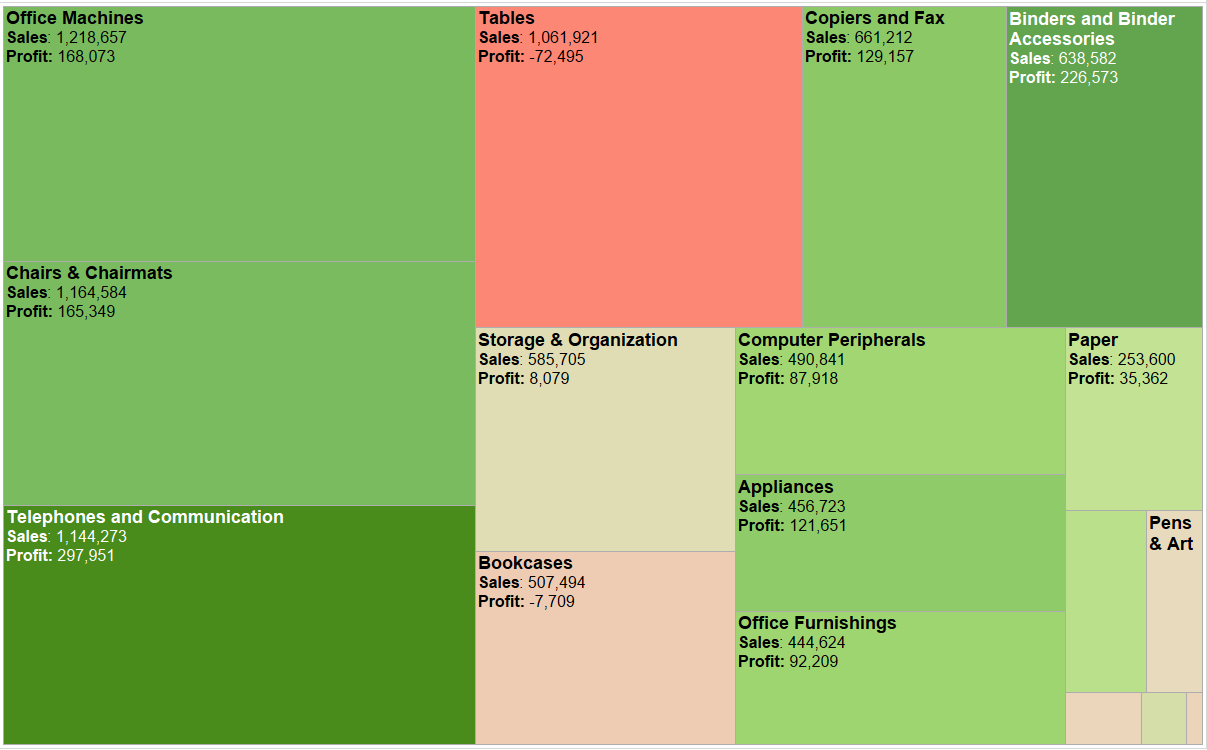
### Treemaps

Treemaps are suitable to show part-to-whole relationships in data. They display hierarchical data as a set of rectangles. Each rectangle is a category within a given variable, whereas the area of the rectangle is proportional to the size of that category. Compared to similar visualizations, like pie charts, tree maps are considered more intuitive and preferable.

**Hierarchy Representation**: The treemap starts with a single, large rectangle representing the entire dataset or category. Within this rectangle, smaller rectangles are nested to represent sub-categories or sub-divisions of the data.

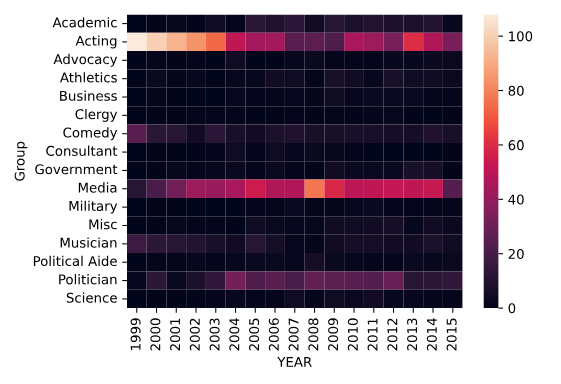
**Rectangle Size and Color**: The size of each rectangle represents a quantitative value. Larger rectangles indicate larger values. Additionally, the color of the rectangles can be used to convey additional information, such as different categories or data ranges.

**Hierarchy Navigation**: Users can navigate through the hierarchy by interacting with the rectangles. Clicking on a larger rectangle can zoom in to reveal its sub-categories, allowing for a detailed exploration of the data's hierarchical structure.



**Heat maps**

A heatmap is a graphical representation of data in which values in a matrix are represented as colors. It is a two-dimensional representation where individual values are represented as colors within a grid. Heatmaps are commonly used to visualize data in various fields such as statistics, biology, and data analysis. In a heatmap, different shades or colors represent different values, allowing viewers to quickly grasp patterns, trends, and variations in the data. Heatmaps are particularly useful for identifying correlations, clustering, and analyzing large datasets, providing a visual summary of complex information.



### Word clouds

Word clouds are useful for visualizing common words in a text or data set. They're similar to bar plots but are often more visually appealing. However, at times word clouds can be harder to interpret. World clouds are useful in the following scenarios:

* Quickly identify the most important themes or topics in a large body of text.
* Understand the overall sentiment or tone of a piece of writing.
* Explore patterns or trends in data that contain textual information.
* Communicate the key ideas or concepts in a visually engaging way.

