# Experiment No. 9

---

**Aim:** Use the inbuilt dataset ‘Titanic’ as used in previous experiment . Plot a box plot for distribution of age with respect to each gender along with the information anout wheather they survived or not.(Cloumn name ‘sex’ and ‘age’)

---

**Requirement:**

- Anaconda Installer
- Windows 10 OS
- Linux
- Jupyter Notebook

**Theory:**

## What is Data Visualization:

Data visualization is defined as a graphical representation that contains the information and the data. By using visual elements like charts, graphs, and maps, data visualization techniques provide an accessible way to see and understand trends, outliers, and patterns in data.

In modern days we have a lot of data in our hands i.e, in the world of Big Data, data visualization tools, and technologies are crucial to analyze massive amounts of information and make data-driven decisions.

It is used in many areas such as:

• To model complex events.
• Visualize phenomenons that cannot be observed directly, such as weather patterns, medical conditions, or mathematical relationships.

**Data Visualization Technique:**

### 1. Box and Whisker Plot

• This plot can be used to obtain more statistical details about the data.
• The straight lines at the maximum and minimum are also called whiskers.
• Points that lie outside the whiskers will be considered as an outlier.
• The box plot also gives us a description of the 25th, 50th,75th quartiles.
• With the help of a box plot, we can also determine the Interquartile range(IQR) where maximum details of the data will be present. Therefore, it can also give us a clear idea about the outliers in the dataset.

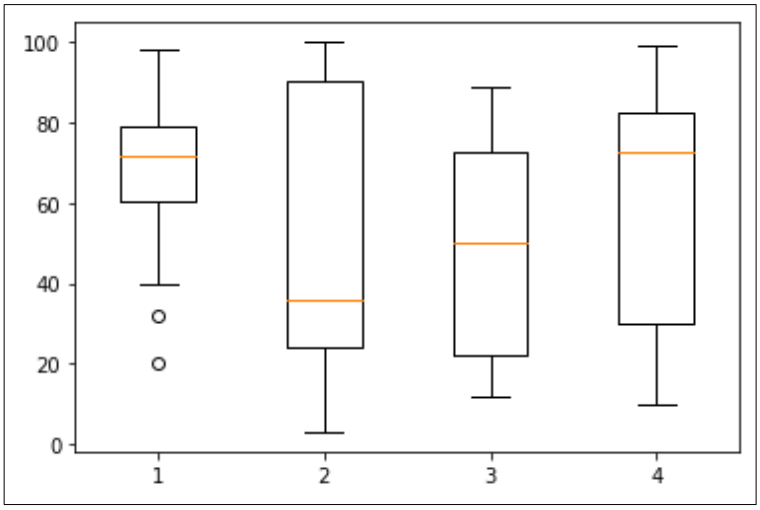


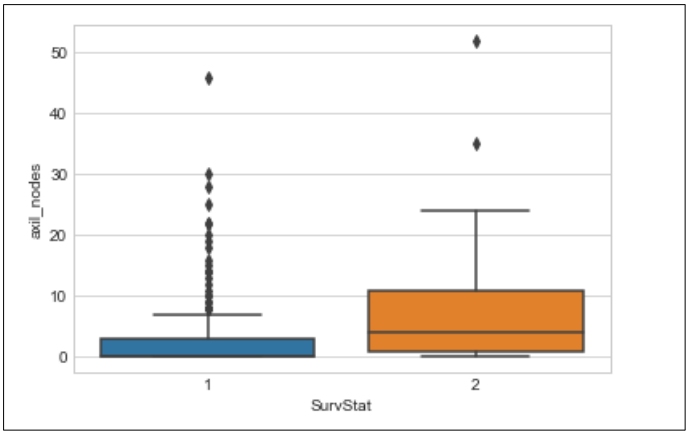

Fig. General Diagram for a Box-plot

**Implementation:**

- Boxplot is available in the Seaborn library.
- Here x is considered as the dependent variable and y is considered as the independent variable. These box plots come under univariate analysis, which means that we are exploring data only with one variable.
- Here we are trying to check the impact of a feature named “axil_nodes” on the class named “Survival status” and not between any two independent features.

The code snippet is as follows:

```
sns.boxplot(x='SurvStat',y='axil_nodes',data=hb)
```




**Libraries Used:**

Seaborn: Seaborn is a data visualization library built on top of matplotlib and closely integrated with pandas data structures in Python.

**Conclusion:** In this experiment we have studied about the data visualization technique and implemented data visualization on the built in data set in seaborn laibrary i.e., titanic dataset