Python for Visualization

Agenda

- Visualization Quiz
- Visualization - One Variable
- Visualization - Two Variables
- Visualization - Multiple Variables

Let's begin the discussion by answering a few questions on visualization.

Visualization Quiz

What are the three statistics presented in box of a boxplot?

A. First Quartile (Q1), Mode, Third Quartile (Q3)

B. Lower whisker, Median, Upper whisker

C. First Quartile (Q1), Median, Third Quartile (Q3)

D. First Quartile (Q1), Mean, Third Quartile (Q3)

Visualization Quiz

What are the three statistics presented in box of a boxplot?

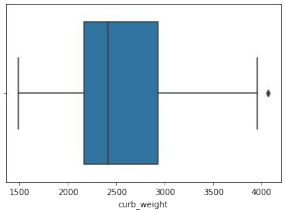
A. First Quartile (Q1), Mode, Third Quartile (Q3)

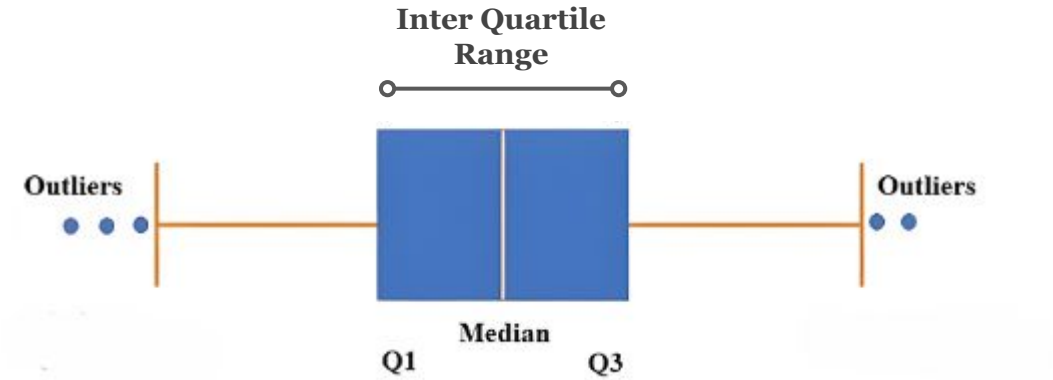
B. Lower whisker, Median, Upper whisker

C. First Quartile (Q1), Median, Third Quartile (Q3)

D. First Quartile (Q1), Mean, Third Quartile (Q3)

Visualization - One Variable

Plot	Type of Data	Usage	Example
Boxplot	Numerical	Helps us understand data distribution and skewness by displaying the data in the form of a box divided by quartiles	



Visualization Quiz

Which of the following represents the general formula for computing the lower whisker (fence) of a boxplot?

A. Q1 - 1.5 * IQR

B. Q1 - 2 * IQR

C. Q2 - 1.5 * IQR

D. Q2 - 2 * IQR

Visualization Quiz

Which of the following represents the general formula for computing the lower whisker (fence) of a boxplot?

A. Q1 - 1.5 * IQR

B. Q1 - 2 * IQR

C. Q2 - 1.5 * IQR

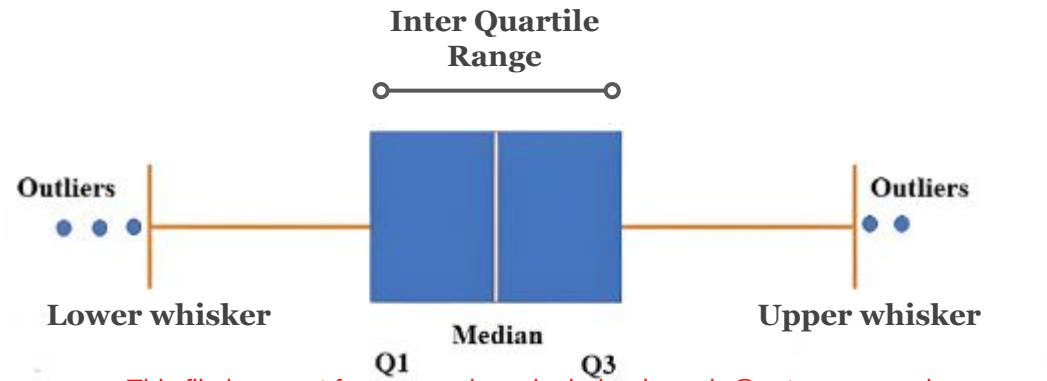
D. Q2 - 2 * IQR

Visualization - One Variable

The whiskers of a boxplot show the range of the data, excluding outliers.

```
Upper whisker: Q3 + 1.5 * IQR
Lower whisker: Q1 - 1.5 * IQR
```

Data points to the left of the lower whisker and to the right of the upper whisker are generally considered to be outliers



Visualization Quiz

How does a KDE plot differ from a histogram?

A. A KDE plot displays the frequency of data points, while a histogram shows the probability density.

B. A KDE plot provides a smoother representation of the data distribution compared to a histogram.

C. A KDE plot is suitable for categorical data, while a histogram is designed for numerical data.

D. A KDE plot cannot handle large datasets in general, whereas a histogram can.

Visualization Quiz

How does a KDE plot differ from a histogram?

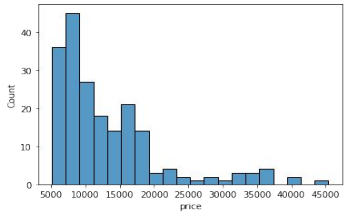
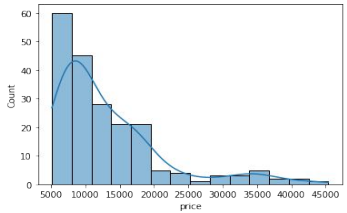
A. A KDE plot displays the frequency of data points, while a histogram shows the probability density.

B. A KDE plot provides a smoother representation of the data distribution compared to a histogram.

C. A KDE plot is suitable for categorical data, while a histogram is designed for numerical data.

D. A KDE plot cannot handle large datasets in general, whereas a histogram can.

Visualization - One Variable

Plot	Type of Data	Usage	Example
Histogram	Numerical	Helps us understand data distribution by dividing it into bins and showing the number of observations in each bin via bars	
Kernel Density Estimation	Numerical	Helps us understand data distribution by displaying a distribution curve on top of the histogram bars	

Visualization Quiz

Which statements accurately describe scatterplot?

A. Each point on a scatterplot represents a single observation or data point

B. Primarily visualize the relationship between two continuous variables

C. Explores both positive and negative correlations

D. Scatterplots are only useful for handling categorical variables

Visualization Quiz

Which statements accurately describe scatterplot?

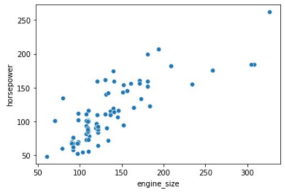
A. Each point on a scatterplot represents a single observation or data point

B. Primarily visualize the relationship between two continuous variables

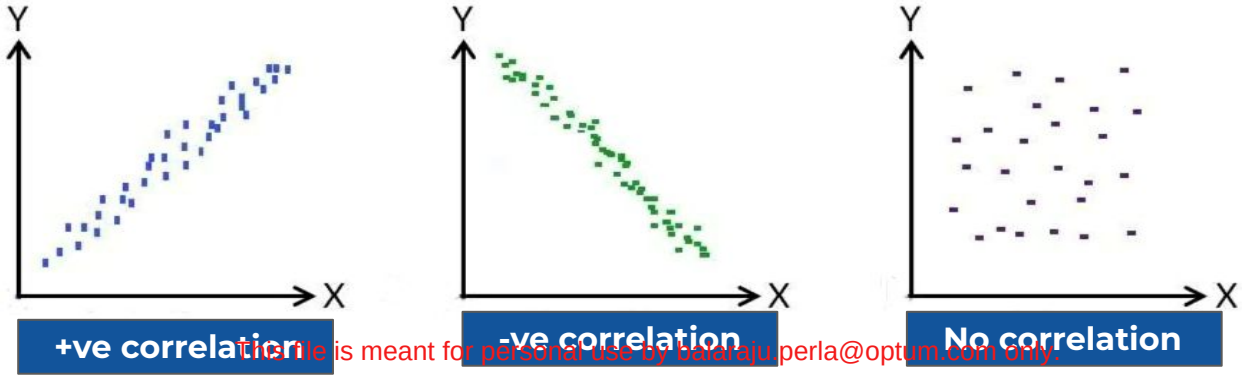
C. Explores both positive and negative correlations

D. Scatterplots are only useful for handling categorical variables

Visualization - Two Variable

Plot	Type of Data	Usage	Example
Scatterplot	Numerical	Helps us understand potential relationship between two numerical variables	

Enables identification of **correlation** and **patterns** between the variables.



Visualization Quiz

When should a jointplot be used instead of a scatterplot?

A. When there are more than two variables

B. When we only want to visualize one variable

C. When we want to visualize both the relationship between variables and the distribution of variables

D. When we want to visualize only the relationship between variables and not the distribution of variables

Visualization Quiz

When should a jointplot be used instead of a scatterplot?

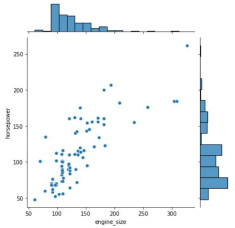
A. When there are more than two variables

B. When we only want to visualize one variable

C. When we want to visualize both the relationship between variables and the distribution of variables

D. When we want to visualize only the relationship between variables and not the distribution of variables

Visualization - Two Variables

Plot	Type of Data	Usage	Example
Jointplot	Numerical	Helps us understand the distribution and relationship between two numerical variables on the same plot.	

Integrates scatterplot with variable-specific histograms for comprehensive visualization

Visualization Quiz

Which of the following statements accurately describes a pairplot?

A. It visualizes the relationship between multiple variables

B. It is composed of boxplots and histograms

C. It displays pairwise relationships in a grid format

D. The diagonal line represents histograms of each variable

Visualization Quiz

Which of the following statements accurately describes a pairplot?

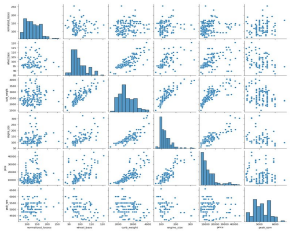
A. It visualizes the relationship between multiple variables

B. It is composed of boxplots and histograms

C. It displays pairwise relationships in a grid format

D. The diagonal line represents histograms of each variable

Visualization - Multiple Variables

Plot	Type of Data	Usage	Example
Pairplot	Numerical	Helps us understand the relationship between two or more pairs of numerical variables	

Offers simultaneous examination of multiple variables

Visualization Quiz

Which of the following statements are TRUE for heatmap?

A. Provide detailed information about outliers

B. Condense information into a single plot for easier pattern identification

C. Exclusively designed for visualizing relationships between categorical variables

D. Represent the relationship between two numerical variables through color gradients

Visualization Quiz

Which of the following statements are TRUE for heatmap?

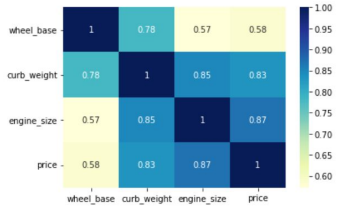
A. Provide detailed information about outliers

B. Condense information into a single plot for easier pattern identification

C. Exclusively designed for visualizing relationships between categorical variables

D. Represent the relationship between two numerical variables through color gradients

Visualization - Multiple Variables

Plot	Type of Data	Usage	Example
Heatmap	Numerical	Helps us understand the correlation between pairs of columns in the data by visualizing it as a matrix	

Provide quick insights into patterns

Great Learning

Happy Learning !