|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **X Variable** | **Y Variable** | **Purpose of analysis** | **Type of chart** | **Example** |
| Continuous (numerical) | Continuous (numerical) | How Y changes with X | Scatter plot | How cholesterol varies with Age? |
| Continuous (numerical) | Categorical | How the range of X varies for various category levels | Box plot | Cholesterol variation with Men and Women |
| Categorical | Categorical | What is the number or % of records of X which falls under each category | Stacked bar | How many men have heart disease compared to women? |
| Continuous | - | Look at the distribution of the values of the X variable | Histogram, boxplot | Distribution of cholesterol ranges |
| Impact of 2 X variables on a Y variable | | |  | Facet\_grid() | Distribution of chol across mean and women – compared for people who have and don’t have heart disease |

Continuous

Scatter plot = jointplot (kind=hex, kind = kde)

Histogram = distplot (distribution plot) – KDE – RUG

Pairplot(takes in dataframe) -> makes scatter plot for each pair

Categorical (x = numerical, y = categorical)

Stripplot(auto[‘a’], auto[‘b’], jitter=false)

Swarmplot – spreads the items for the same value

Boxplot(x=categorical, y=numerical) (no.of doors, horsepower) – hue = fuelpower

Q1-1.5IQR (INTER QUARTILE RANGE – Q3-Q1)

Whiskers, Q1, median, Q3, outliers

Barplot – pass two values + hue

Goes from 0. Bar goes all the way down to zero.

If you don’t want to go till zero use **point plots**

Countplot – categorical -> gets the count of each category. Can have hue

Catplot -> categorical plot

Catplot(x, y, data, hue, col, kind)

For kind=count there will be no y

Lmplot -> linear model plot

Line going through scatter plot. Line of best fit.

PointPlot

. - Any Character Except New Line

\d - Digit (0-9)

\D - Not a Digit (0-9)

\w - Word Character (a-z, A-Z, 0-9, \_)

\W - Not a Word Character

\s - Whitespace (space, tab, newline)

\S - Not Whitespace (space, tab, newline)

\b - Word Boundary

\B - Not a Word Boundary

^ - Beginning of a String

$ - End of a String

[] - Matches Characters in brackets

[^ ] - Matches Characters NOT in brackets

| - Either Or

( ) - Group

Quantifiers:

\* - 0 or More

+ - 1 or More

? - 0 or One

{3} - Exact Number

{3,4} - Range of Numbers (Minimum, Maximum)