Scatter Plots

October 5, 2020 12:46 AM

Scatter plots → allow comparison of [2 variables] for a set of data.

→ Can interpret correlation (depends on trends of scatter points)

+ve

−ve

→ go.Scatter → defining Scatter plot → x, y, mode → type of data point. → dict/{}

go.Layout → title, xaxis, yaxis, hovering mode → dict/{}

go.Figure → to intialize data & layout on figure → dict/{}

→ ALWAYS A LIST [ ]

Line Charts

October 5, 2020 1:53 AM

→ series of data points (markers) → connected by line segments

~ Similar to scatter plot = but measure points are ordered + joined by straight line

~ to visualize trend data over intervals of time = time series

STYLES IN PLOTLY → lines / lines + markers / markers

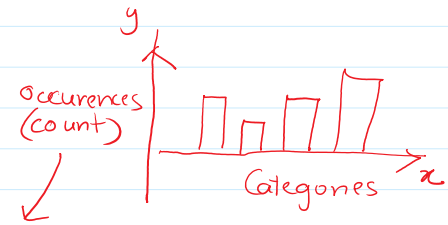
Bar Charts

October 5, 2020 2:20 PM

Categorical data → Rectangular bars (heights / lengths)

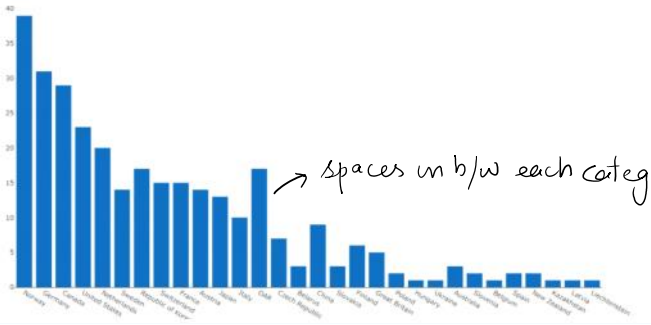
Categorical / Continuous

- Information (discrete) → ex Gender, status
- measurements (values) → ex weight, height etc



Count, Sum, average etc
↳ default

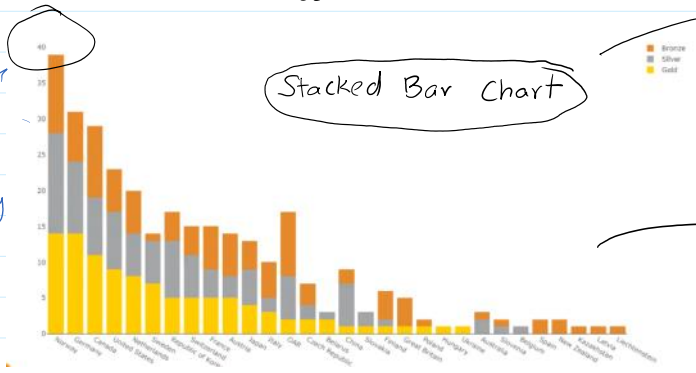
ex



Count of medals

→ spaces in b/w each category.

↓ countries



Stacked Bar Chart

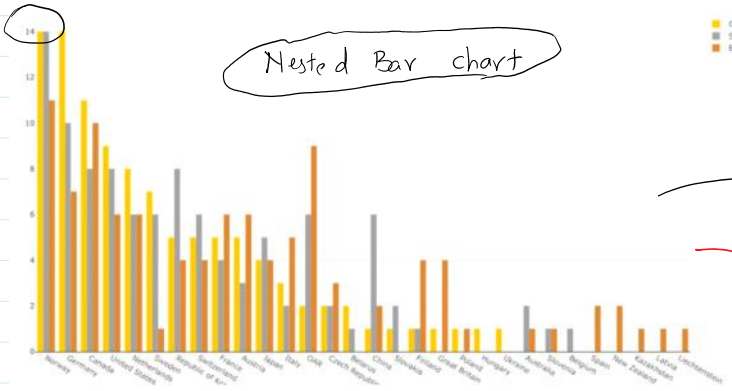
→ With in single bar, it is difficult to identify classes

Total is easy to identify

→ representing medals in further classification.

↓
Bronze, Gold, Silver

This is reduced as for each class, separate bar chart is created.



Nested Bar chart

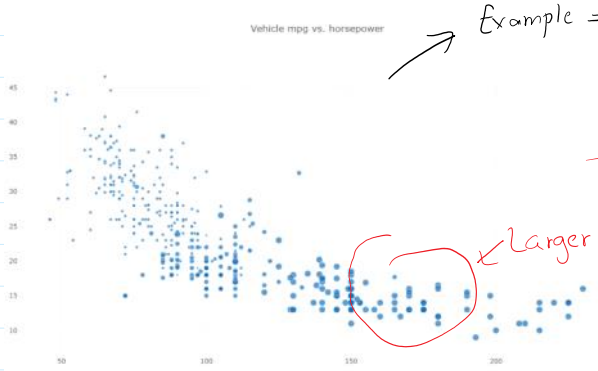
→ Clearly distinguish no. of medals in each type

→ But this doesn't tell the total.

Bubble Charts

October 5, 2020 9:46 PM

→ Similar to scatter plots → Now they have 'Third Variable' → through size of markers



Example = all data points are not same size (like in scatter plot)

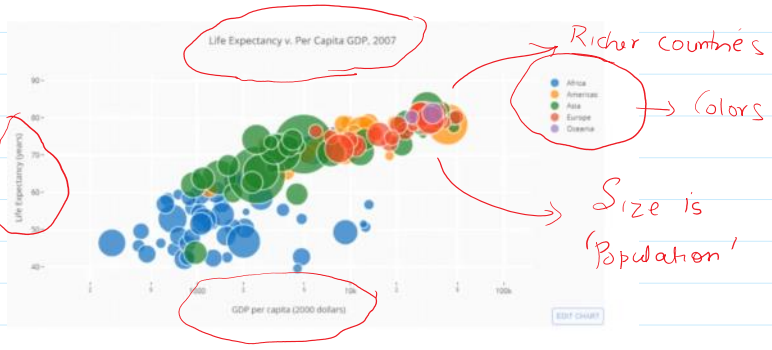
→ HP ↑ MPG ↓

Larger = size of each bubble → no. of cylinders (3rd feature)

Reveals multiple info.

In 2D

1. life expectancy
2. per capital GDP
3. Continent
4. Countries (Size bubble)



Richer countries

→ Colors

Size is 'Population'

Box Plots

October 12, 2020 9:59 AM

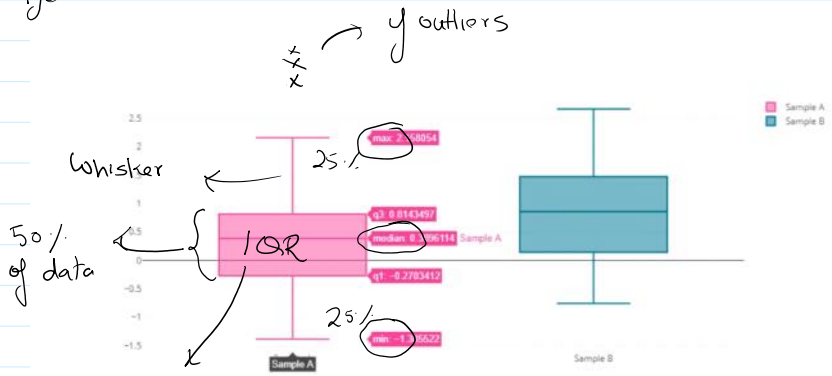
→ Visualize variation of a feature → depicting continuous numerical data } quartiles

\+ separate data based on categ. feature
↳ to compare cont. value

4 parts

25 %
50 %
25 %

-> Data distribution
-> Real analysis



y outliers

Whisker

25 %

50 % of data

IQR

25 %

Inter quartile region

Upper outliers

IQR >

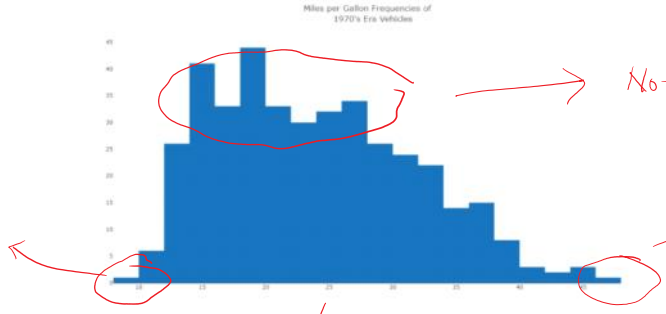
$(q_3 - q_1) \times 1.5$

Histogram

October 12, 2020 1:57 PM

accurate representation of overall distribution of a continuous features ⇒ divide entire range of values of cont. features into series of intervals.

MPG of vehicles histogram

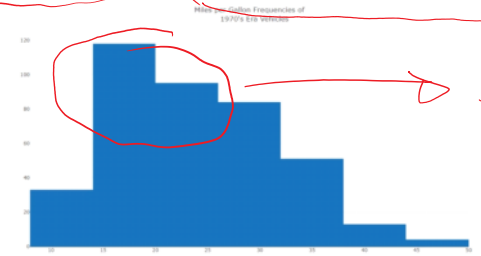


Not finer details (Need to inc bin size)

only 1 car with MPG (0, 10)

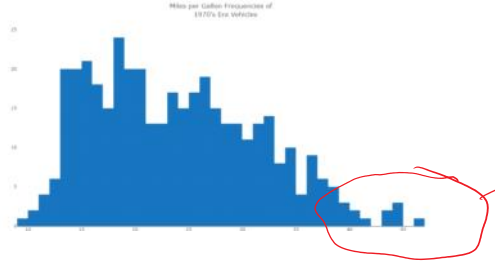
only 1 car with MPG (45 - 50)

Increased bin size (decreased number of bins)



Still doesn't give details

Decreased bin size (increased number of bins)



gaps with no counts (not useful)

To choose properly # bins (keep it default)

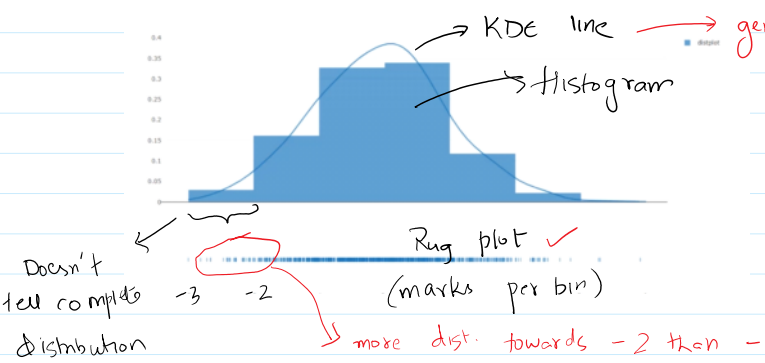
DistPlots

October 12, 2020 4:42 PM

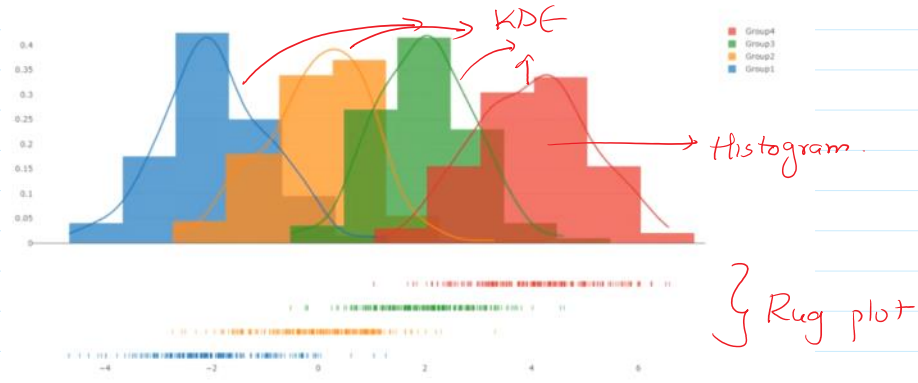
Distribution plots (3 plots on top of one another) → HISTOGRAM (each data point is placed inside a bin of similar values)

→ RUG PLOT
→ marks = x-axis for every data point
↓
LET YOU SEE DISTRIBUTION OF VALUES INSIDE EACH BIN.

→ KDE line (Kernel Density Estimate) – describes shape of distribution



Interesting plot to see distributions of diff categories



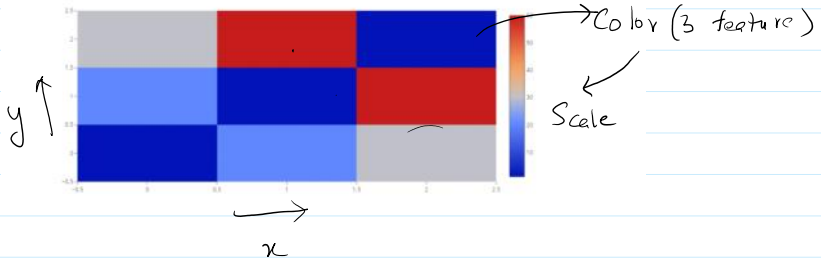
Heatmaps

October 12, 2020 7:37 PM

Visualization → 3 features → Categorical / Continuous } x-y + color (3 feature)

Example (continuous)

X and Y axis are separated into intervals to form a grid



Example (categorical)

Real data