Data Visualization

Visualization tools

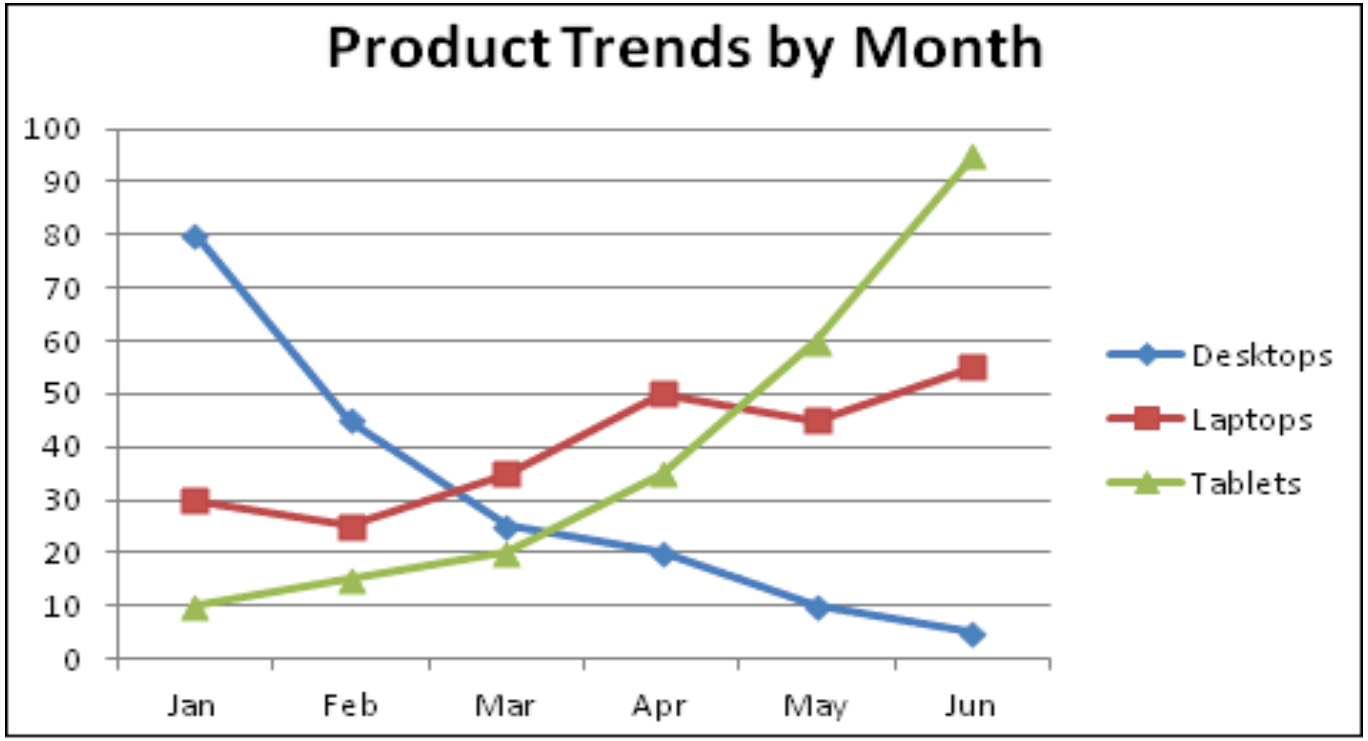
- Data visualization is the activity of accepting data and placing it in the visual context such as a map or a graph.
- Data visualizations help for better understanding of patterns , trends and outliers in groups of data.
- Since the purpose of data analysis is to gain insights, data is much more valuable when visualized.
- Without the visual representation of the insights, it can be difficult for the audience to grasp the insights

Different Visualization Tools

- Line Chart
- Area Chart
- Bar Chart
- Histogram
- Scatter Plot
- Bubble Plot
- Pie Chart
- Heat Map
- Box Plot
- Andrews Curve
- Chernoff’s Faces

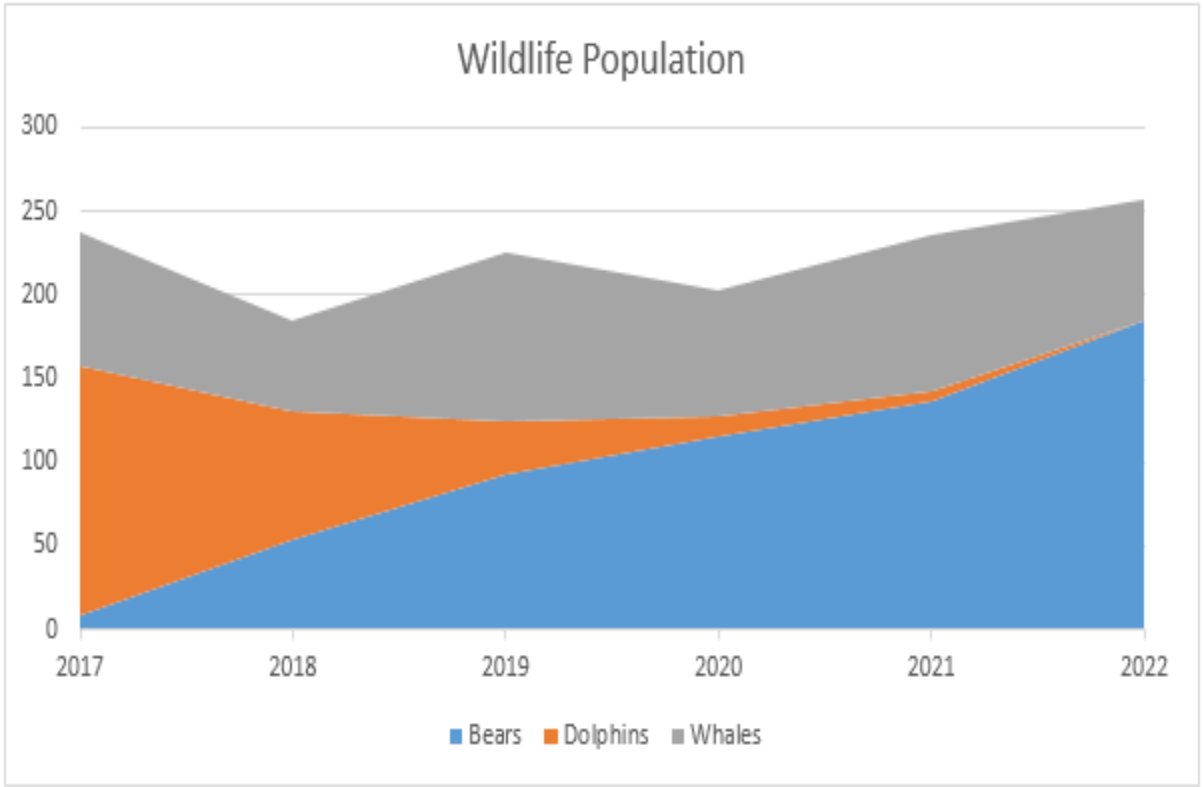
Line Chart

- Line chart illustrates changes over time. The x-axis is usually a period of time, while the y-axis is the quantity.



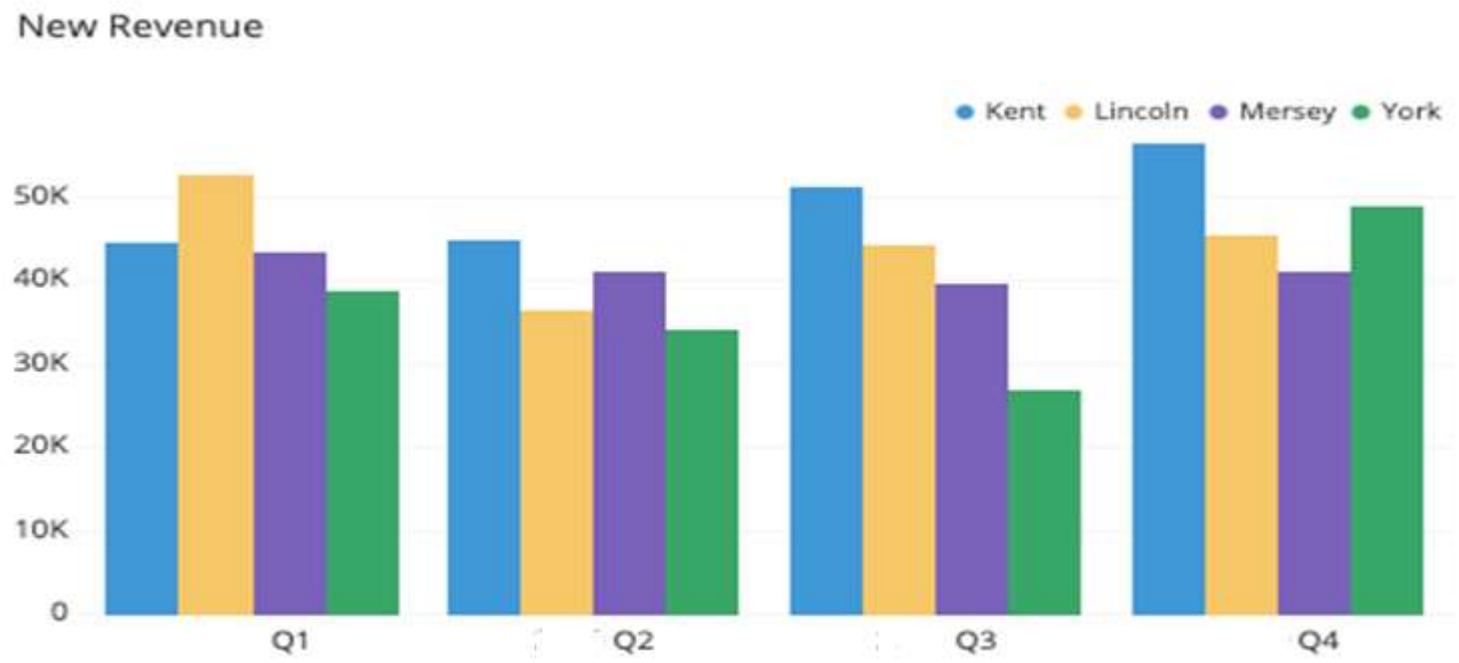
Area Chart

- It is an adaptation of line chart where the area under the line is filled to emphasize its significance.



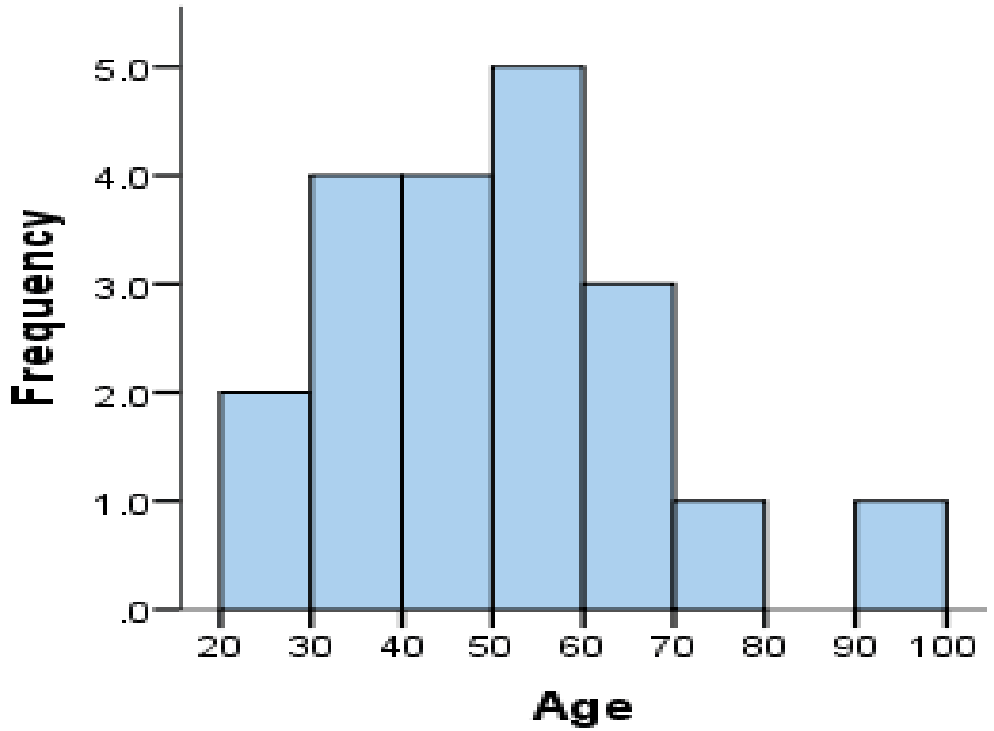
Bar chart

- It also illustrates changes w.r.t. time. But if there are more than one variable, a bar chart can make it easier to compare the data for each variable at each moment of time.



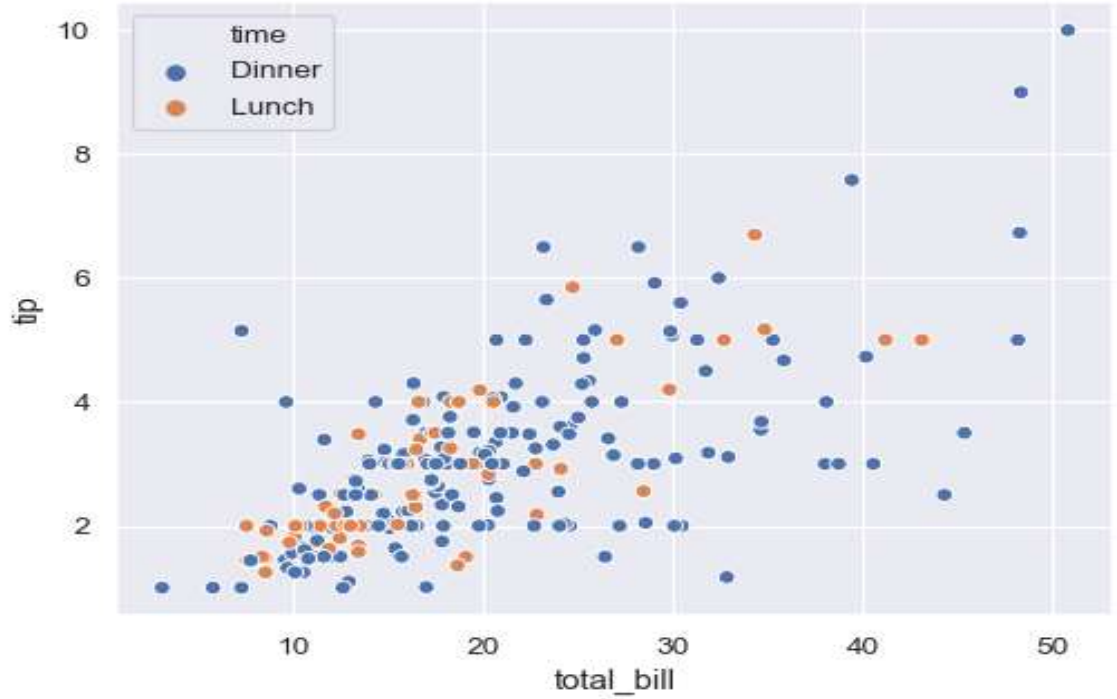
Histogram

- It measures frequency rather than trends in time. The x-axis is typically the intervals and y-axis is the frequency



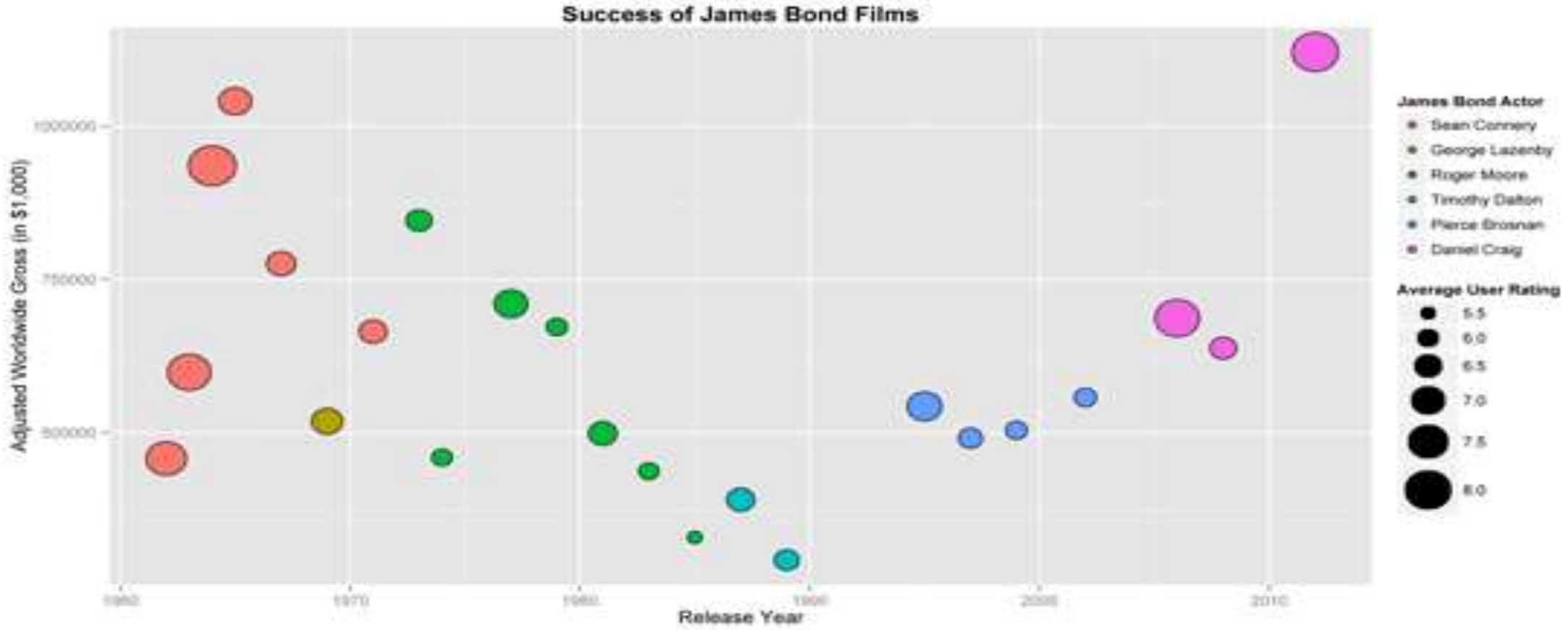
Scatter Plot

- These are used to find correlations. The points trend in a certain manner to give an idea of correlation. If the points are observed to be truly scattered, the variables can be concluded to be uncorrelated.

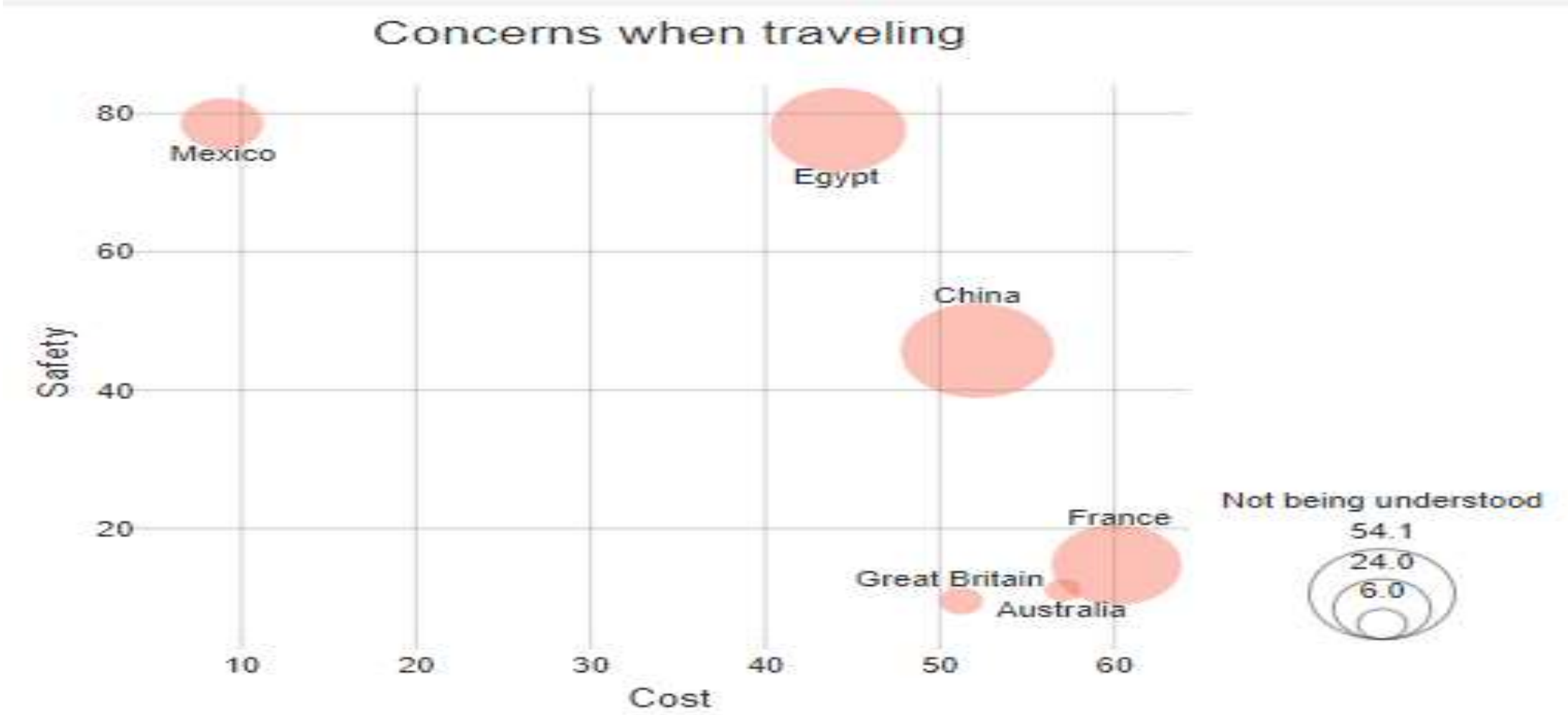


Bubble Plot

- Each point is illustrated as a bubble whose area has a meaning in addition to its placement on the axis. The negative point of this chart is the limited space within the axes to fit bubbles at times.

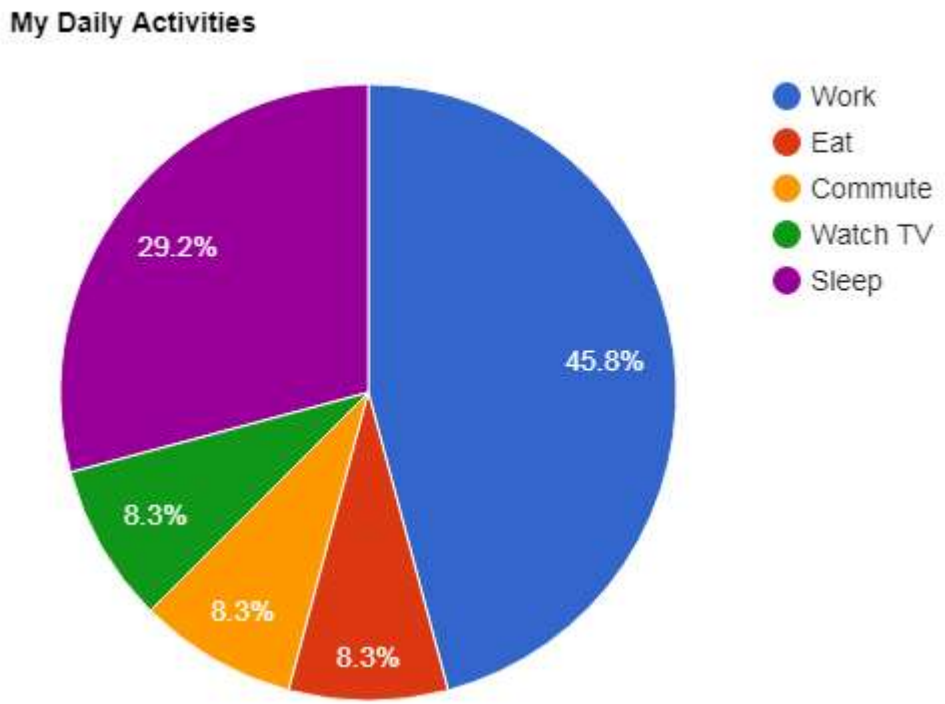


Bubble Plot


Concerns when traveling
Safety
80
60
40
20
10
20
30
40
50
60
Cost
Mexico
Egypt
China
France
Great Britain
Australia
Not being understood
54.1
24.0
6.0

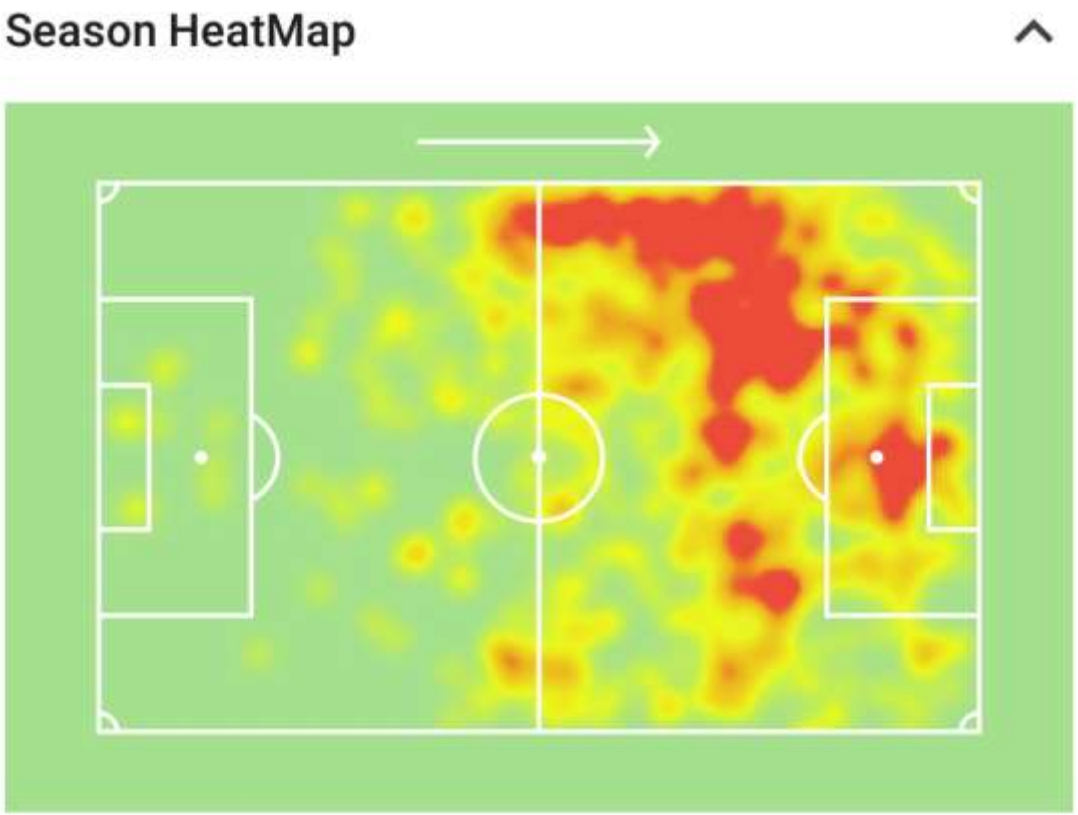
Pie Chart

- It is used for illustrating percentages. It shows elements as part of a whole.



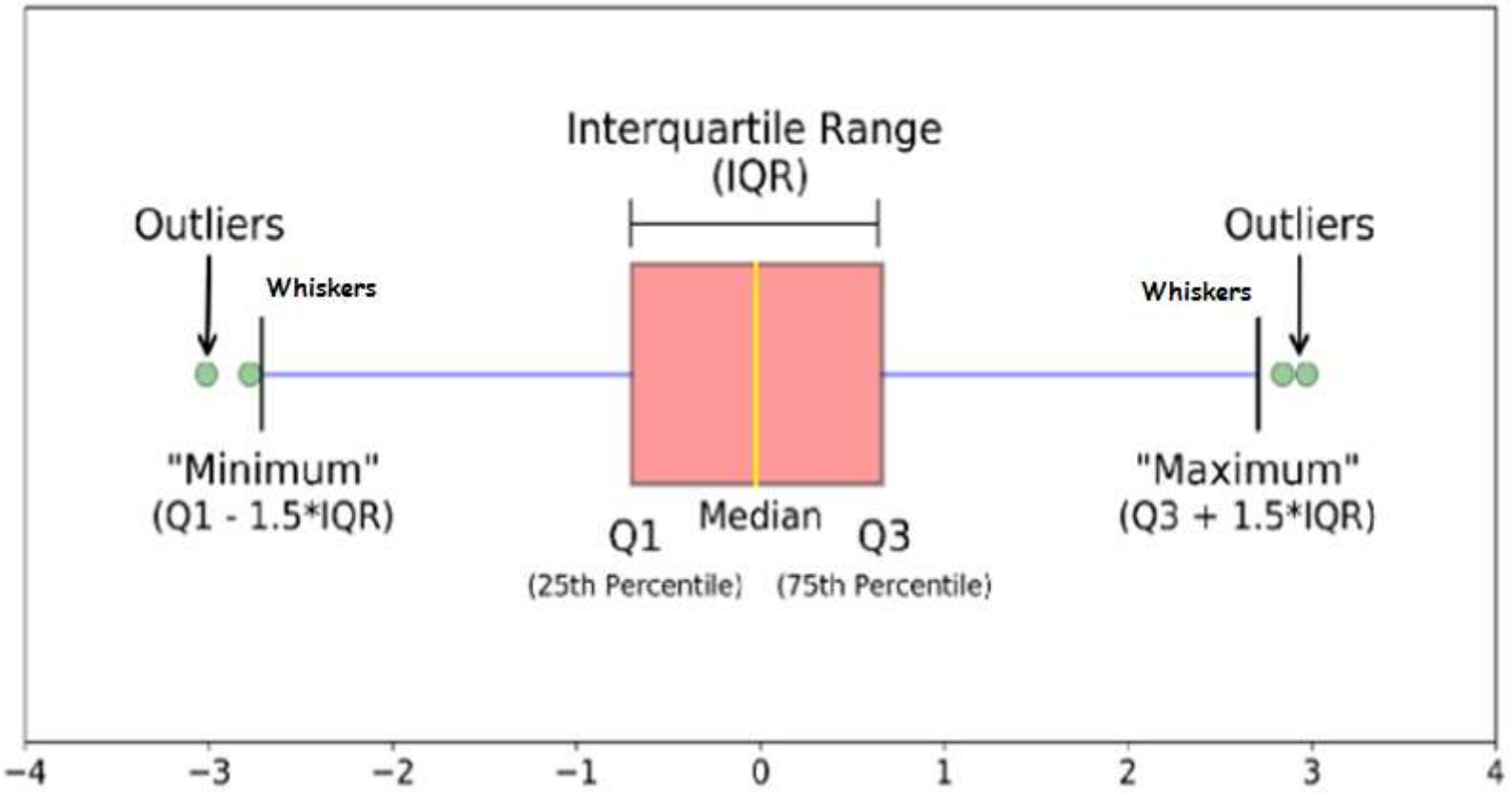
Heat Map

- It is a colour coded matrix. This type of visualization is helpful because colours are quicker to interpret than numbers.



Cristiano Ronaldo movement on the football field in the Serie-A season for Juventus club as a left winger

Box(Box and Whisker) Plot



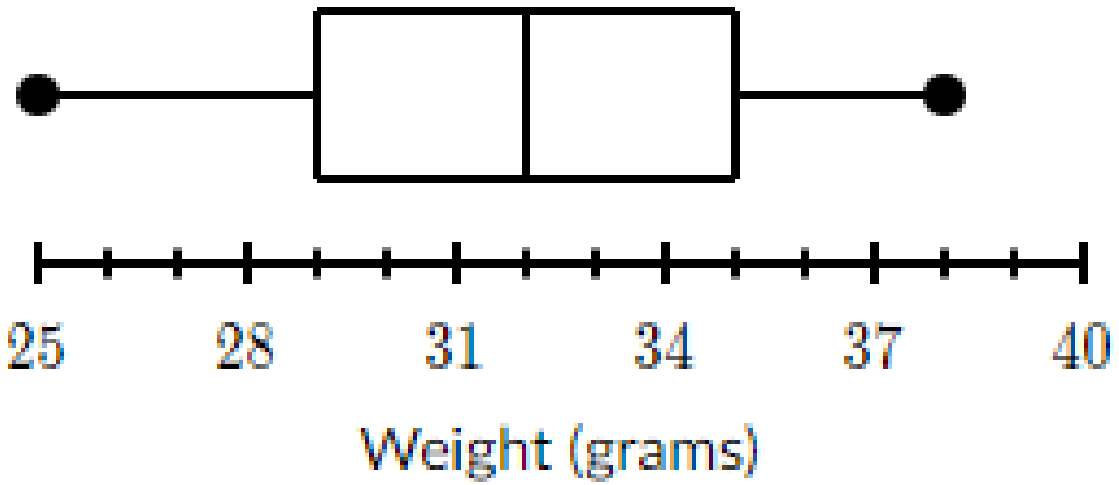
https://towardsdatascience.com/understanding-boxplots-5e2df7bcbd51

Box Plot Contd..

- Arrange the given data in ascending order
- Find Median, first quartile(Q1), third quartile (Q3).
- Find IQR: Inter quartile range= Q3-Q1
- Find Minimum: Q1-(1.5xIQR)
- Find Maximum: Q3-(1.5xIQR)
- Identify Outlier. Outliers are those numbers which are less than minimum or greater than maximum

Questions on Box Plot

Q. 1:Box plot shows the representation of weights of raisins. About what percentage of raisins weighed more than 29gms?



Ans: 75%

Questions on Box Plot

Rank ordered data is - 5,7,10,15,19,21,21,22,22,23,23,23,23,23,24,24,24,24,25	
Median	**?**
Quartile 1	**?**
Quartile 3	**?**
Inter Quartile Range	**?**
Minimum	**?**
Maximum	**?**
Outliers	**?**

Questions on Box Plot

Rank ordered data is - 5,7,10,15,19,21,21,22,22,23,23,23,23,23,24,24,24,24,25	
Median	**23**
Quartile 1	**19**
Quartile 3	**24**
Inter Quartile Range	**5**
Minimum	**11.5**
Maximum	**31.5**
Outliers	**5,7,10**

Andrews Curve

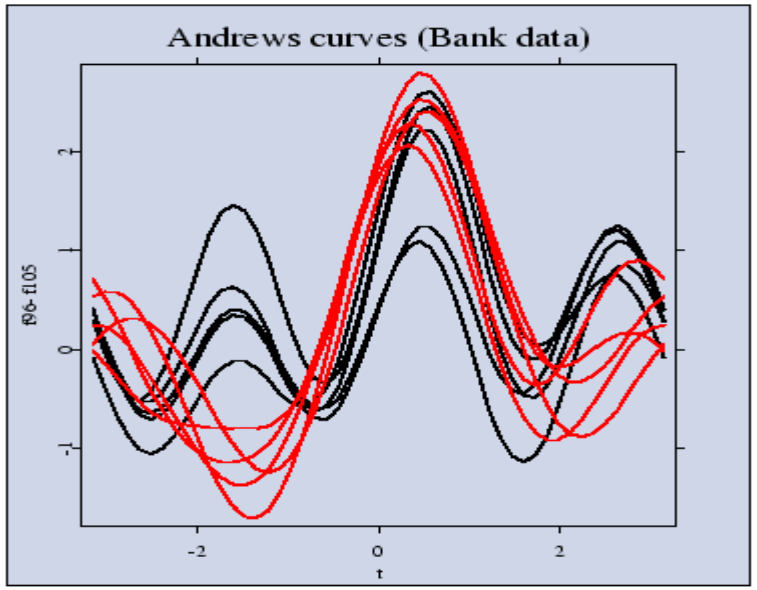
- The basic problem of graphical displays of multivariate data is the dimensionality.
- Scatter plots work well up to three dimensions More than three dimensions have to be coded into displayable 2D or 3D structures.
- The idea of coding and representing multivariate data by curves was suggested by Andrews in 1972.
- Each multivariate observation is transformed into a curve as follows:
- It defines a finite Fourier series.
- $fx(t)=x_1/\sqrt{2} + x_2\sin(t) + x_3\cos(t) + x_4\sin(2t) + x_5\cos(2t)+..$

- Let us consider swiss bank data set. http://www.statistics4u.com/fundstat_eng/data_fluriedw.html

Variables are – length of the note, width of left edge, width of right edge, bottom margin width, top margin width, diagonal all in mm.

- The 96th observation is – [215.6,129.9, 129.9,9.0,9.5,141.7]

$$f_{96}(t) = \frac{215.6}{\sqrt{2}} + 129.9\sin(t) + 129.9\cos(t) + 9.0\sin(2t) + 9.5\cos(2t) + 141.7\sin(3t).$$

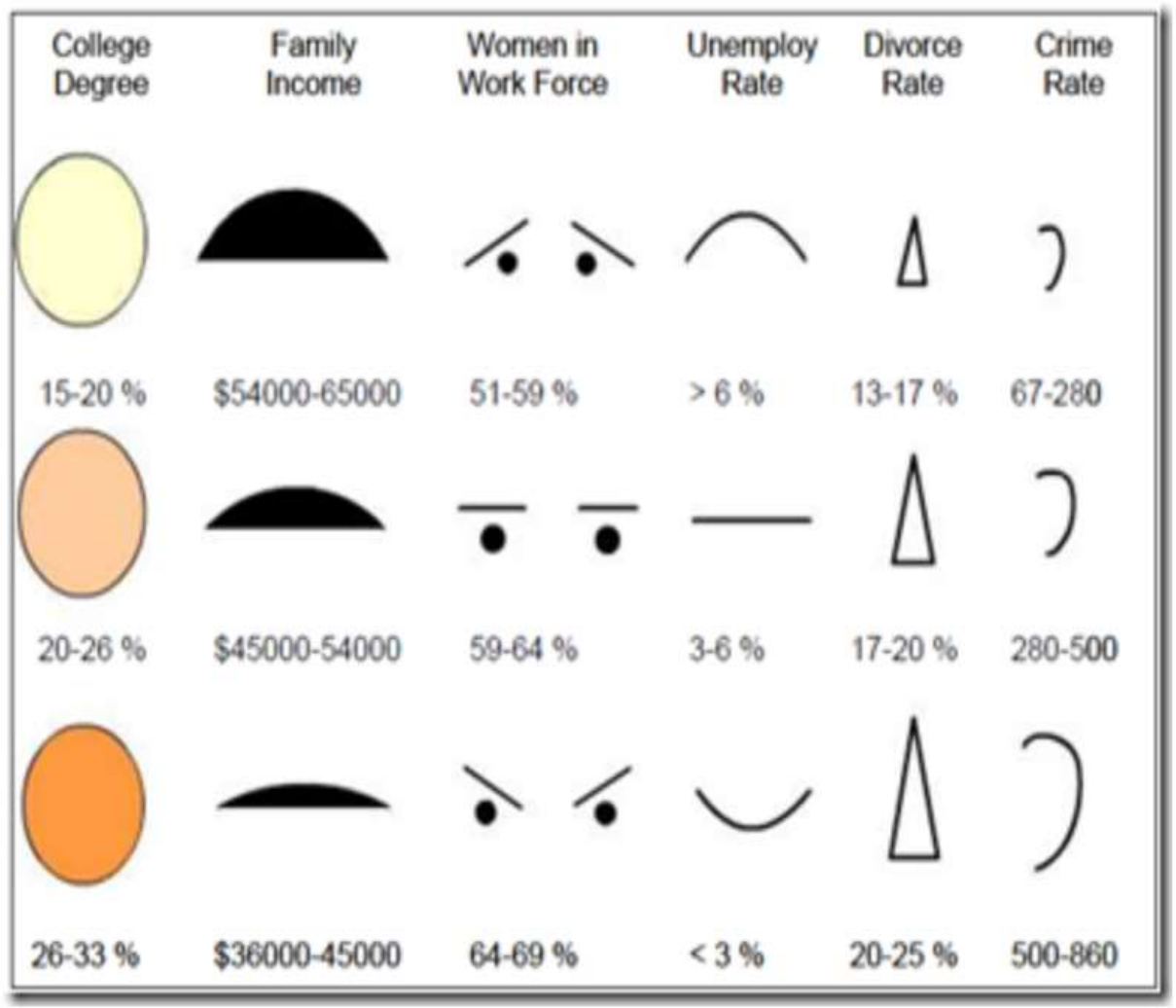


- Figure shows the Andrews' curves for observations 96-105 of the Swiss bank note data set.
- The observations 96-100 represent genuine bank notes, and that the observations 101-105 represent counterfeit bank notes.

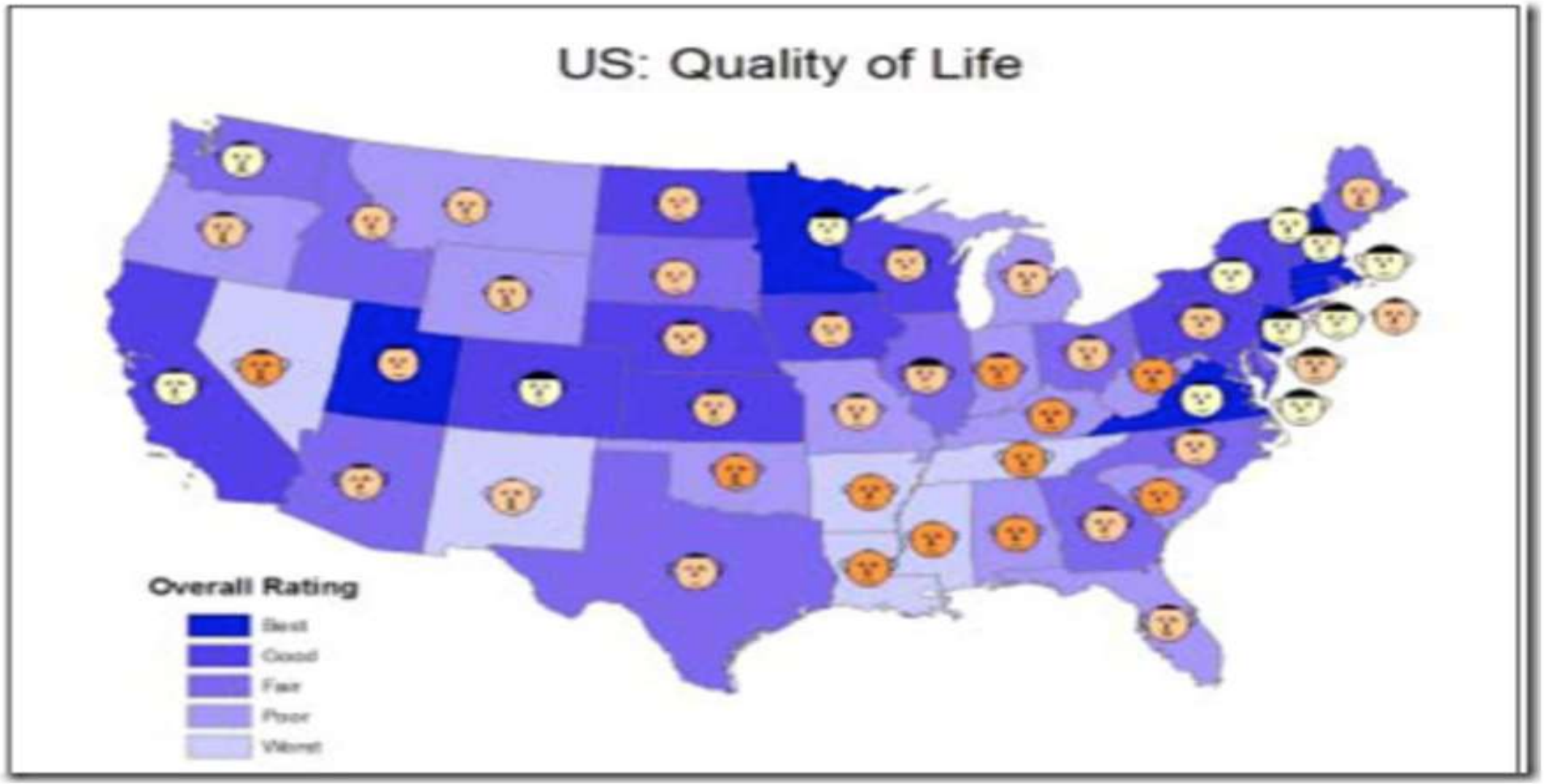
Chernoff faces

- Used to display multivariate data in the shape of human face. The individual parts, such as eyes, ears, nose, mouth, face length are used to represent values of the variables by their shape, size, placement and orientation. The humans easily recognize faces and notice small changes without difficulty.

Quality of Life – Chernoff faces composition



Quality of Life – Chernoff faces composition



Batting Records

Parameter	Tendulkar	Dhoni	Sehwag	Gambhir	Yuvraj
Batting Average	44.83	50.57	35.05	39.68	36.55
Strike Rate	86.23	87.56	104.33	85.25	87.67
No. of Fours / match	4.354	2.36	4.510	3.816	2.984
No. of Sixers / match	0.421	0.654	0.541	0.115	0.510
Ratio of innings to total matches played	0.976	0.848	0.976	0.972	0.914

Metrics and features association

Metrics	Features
Batting Average	Height of Face
Strike Rate	Curve of Smile
Number of fours per match	Width of Eyes
Number of sixes per match	Height of Eyes
Ratio of innings to total matches played	Width of Face

Chernoff Faces

Sachin

Sehwag

Dhoni

Yuvraj

Gambhir

Conclusion

- As can be seen, the happiest face seems to be of Sehwag, no surprise there since he has the highest strike rate, a variable mapped to curve of the smile. Also, notice Dhoni has a very long face, this is again due to the fact because the batting average is mapped to the height of face and Dhoni has a very good batting average.
- Another thing to notice is the width of eyes for both Dhoni and Yuvraj, its very small, testament to the fact that both these players, although very good stroke makers, made a lot of runs by running between the wickets.