Improving Graphs

Amit Sethi, faculty, IIT Bombay
Sudhakar Kumar, student, IIT Bombay

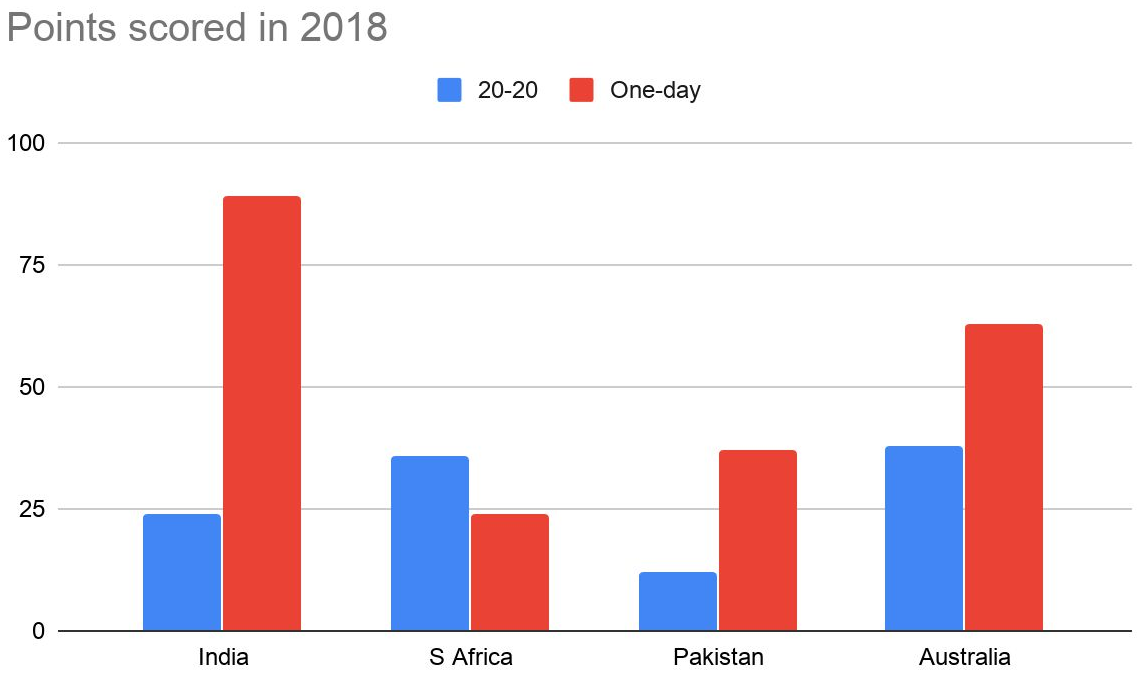
SHALA2020.GITHUB.IO
2020.04.20 :: 21:00 UTC+5:30

Learning Objectives

- List and draw basic graph types
- Add information to graphs
- Beautify graphs
- Highlight your message in graphs
- List and draw advanced graph types
- Decide between different types of graph

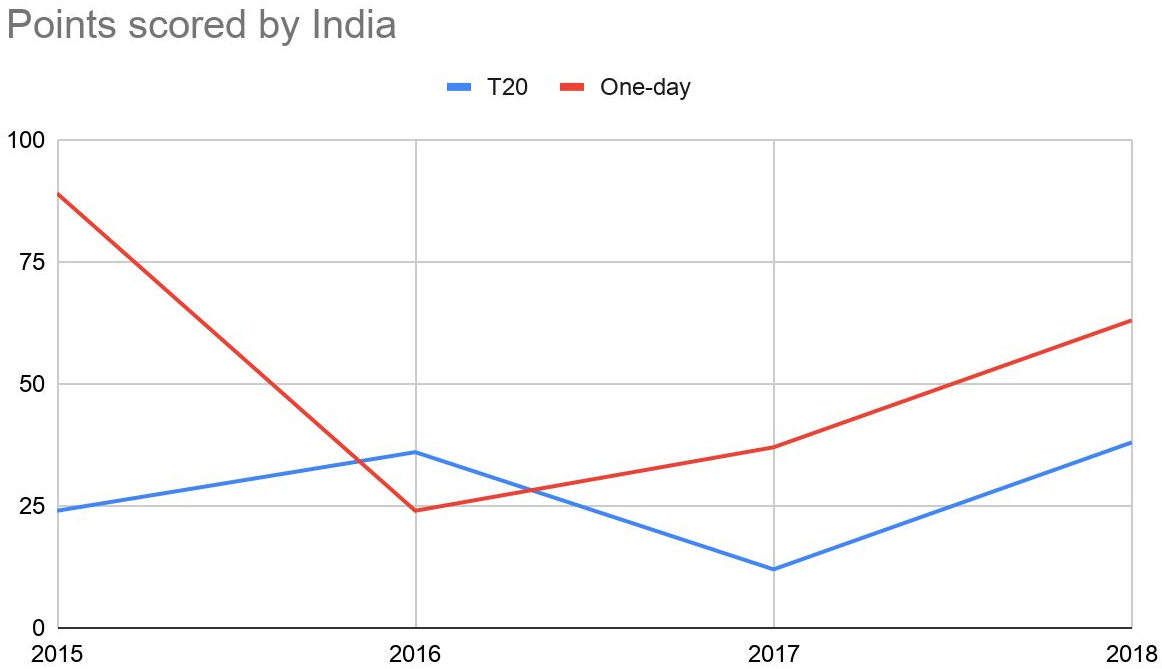
Bar graph

Used for comparing nominal entities



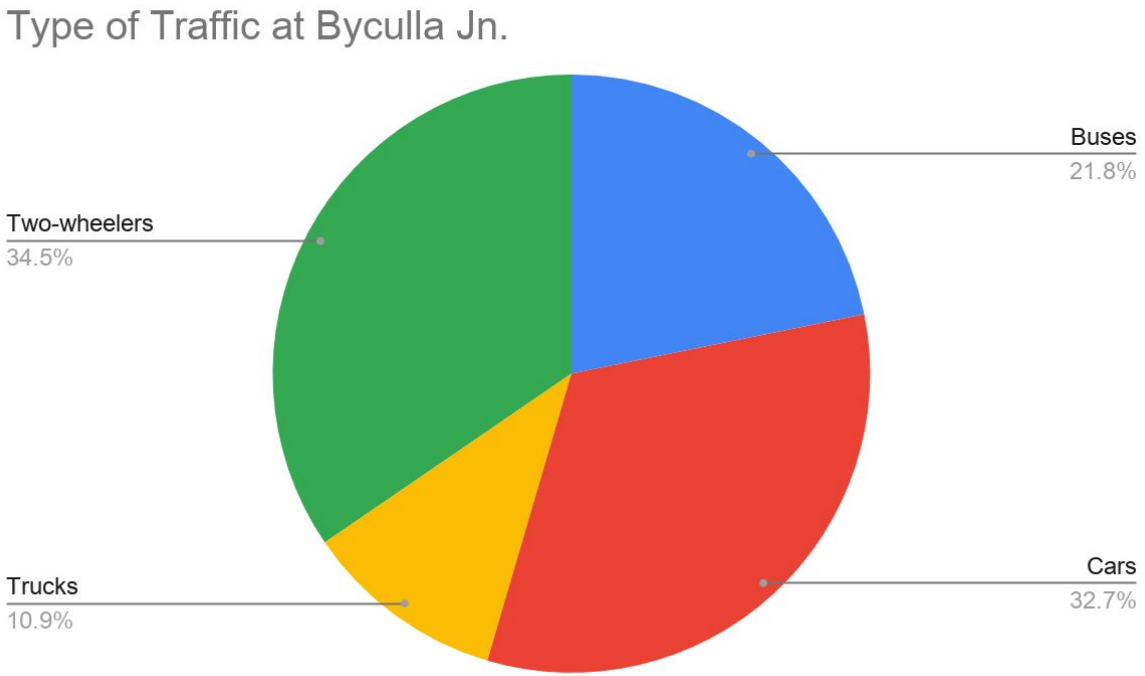
Line chart

Used for temporal order



Pie chart

Used to represent proportions



Scatter plot

Used to compare two metrics among several samples

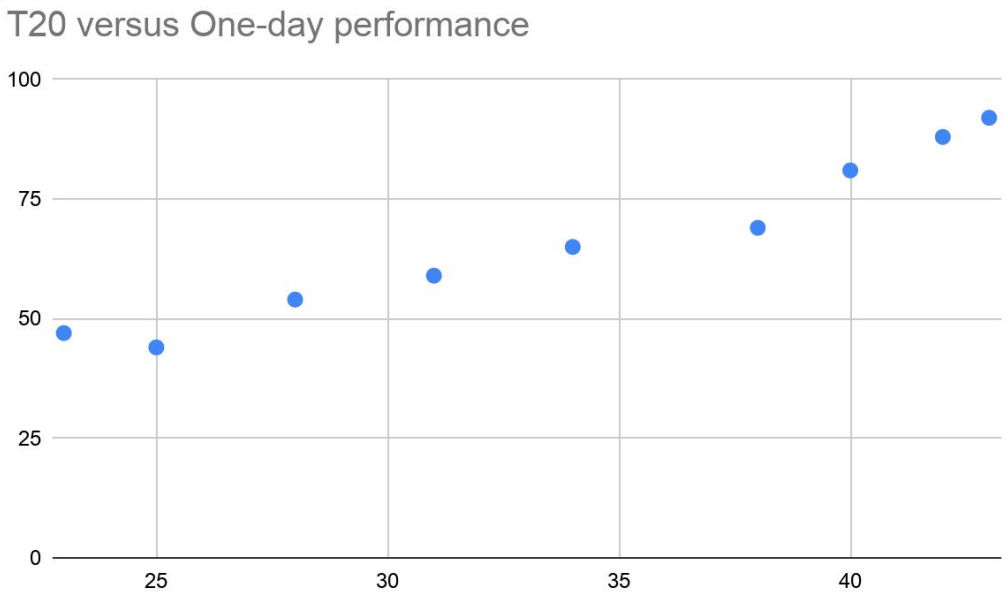


Chart Title

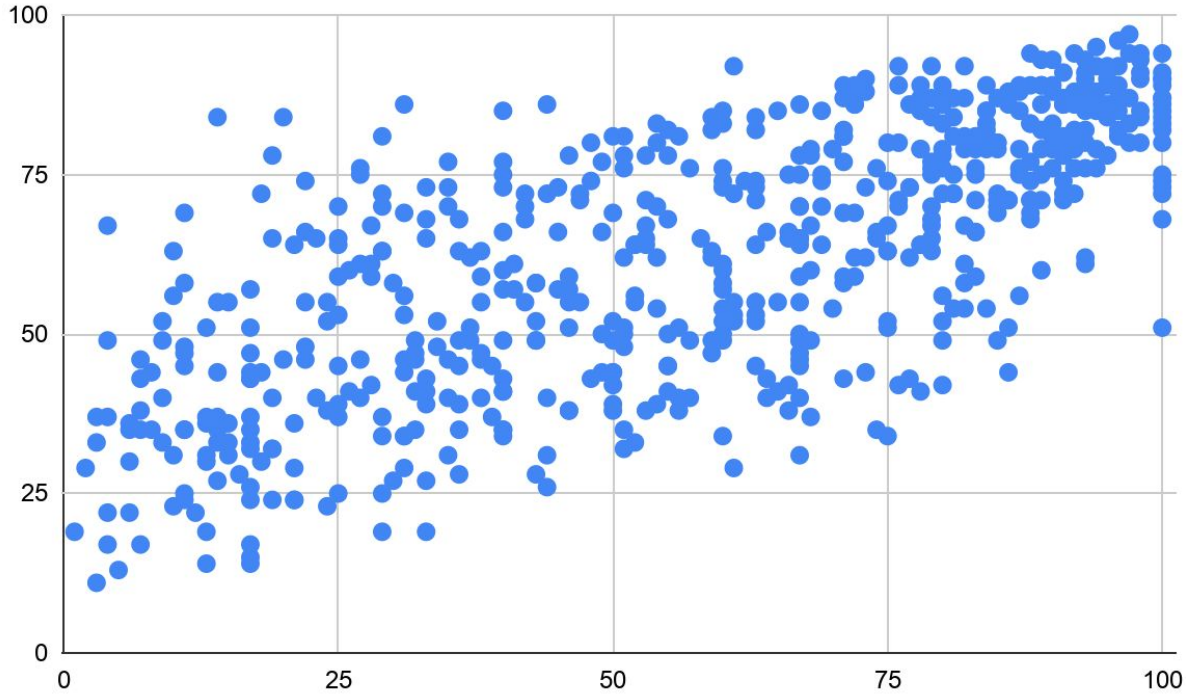
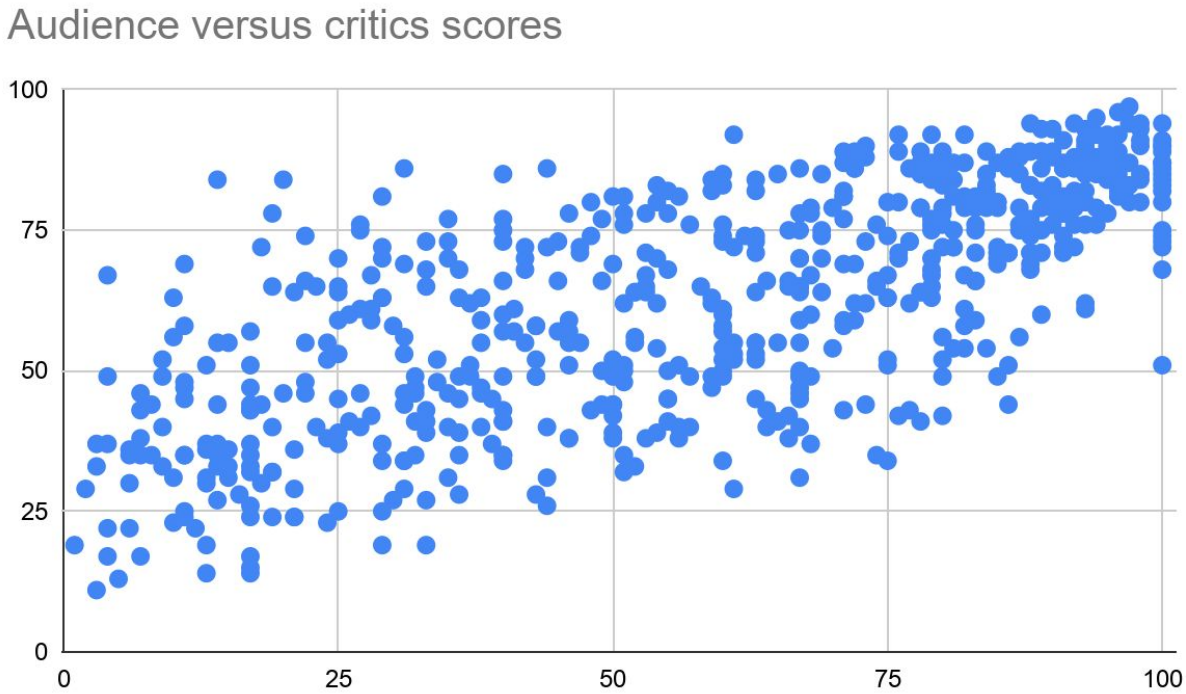
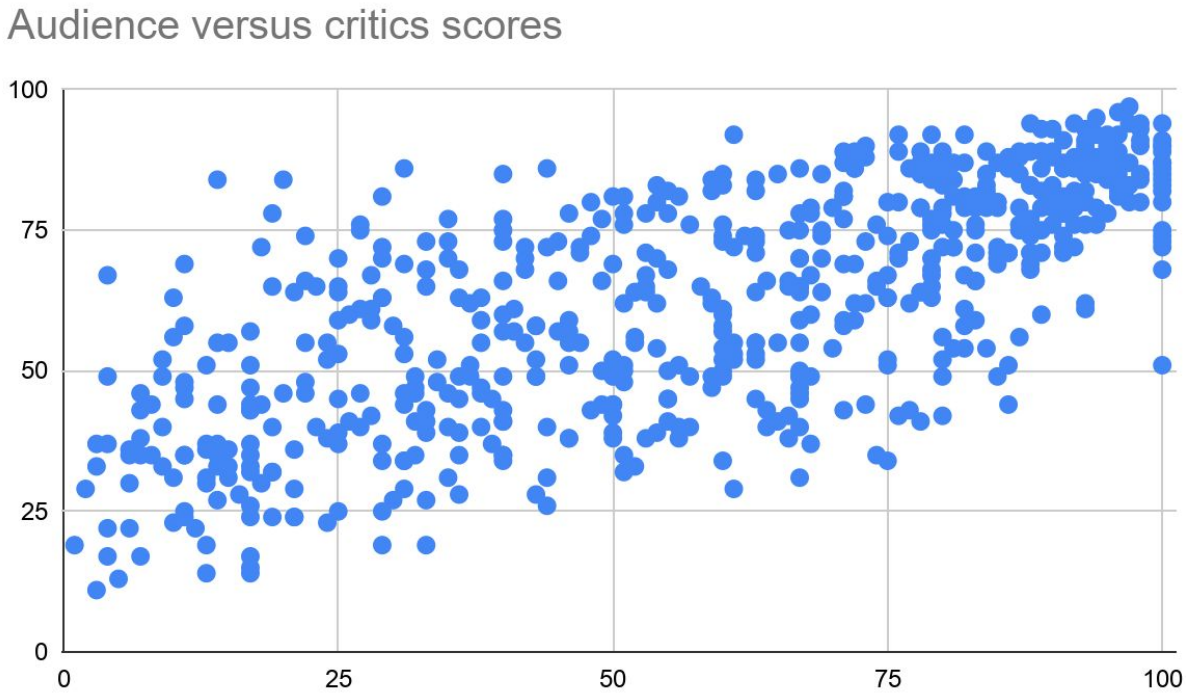


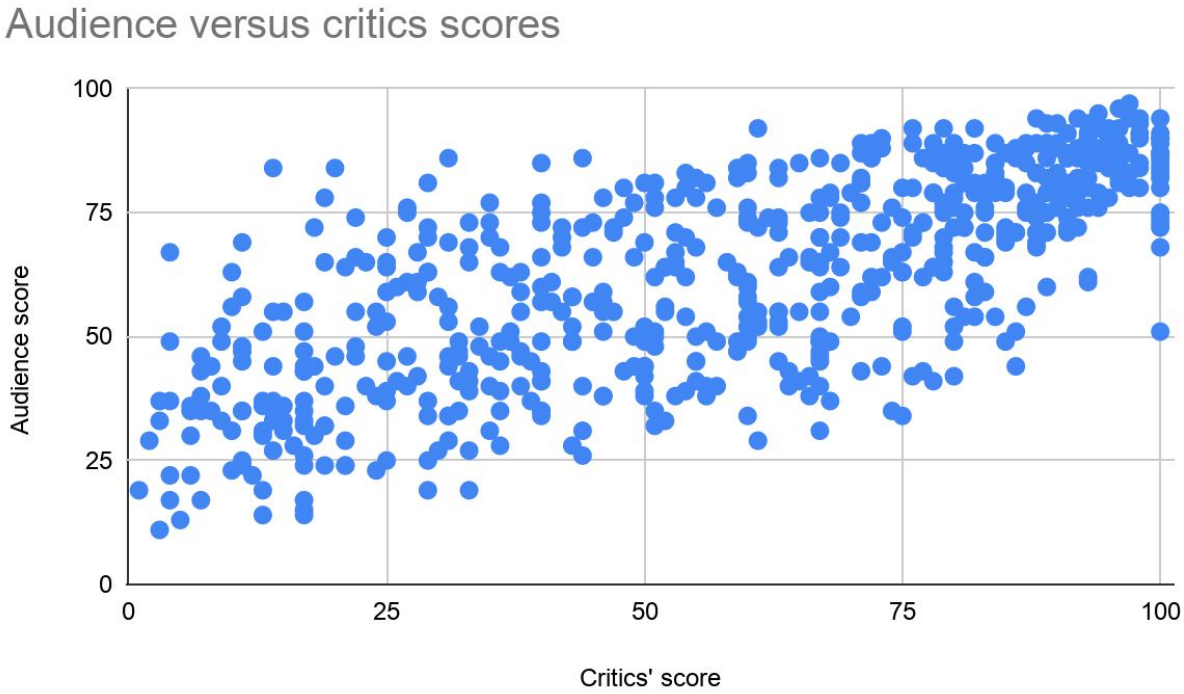
Chart Title



Axis titles

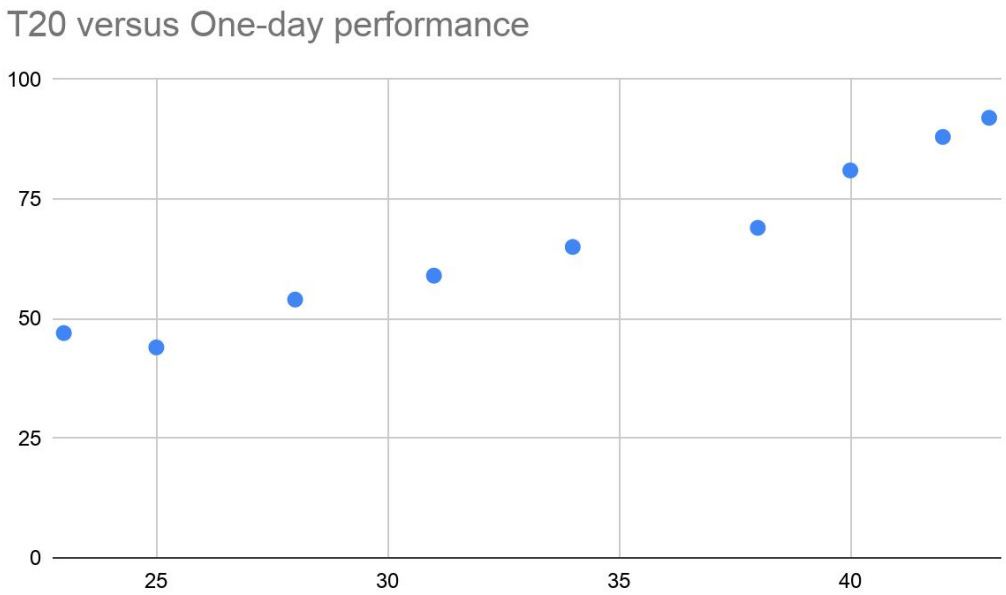


Axis titles


Audience versus critics scores
Audience score
Critics' score
0
25
50
75
100
0
25
50
75
100

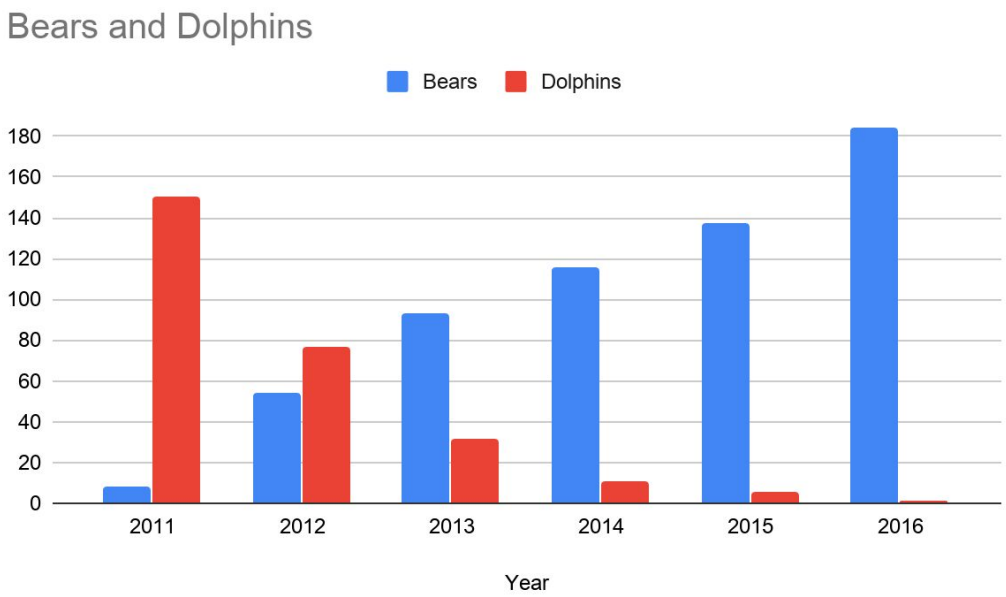
Axis units

Used to compare two metrics among several samples

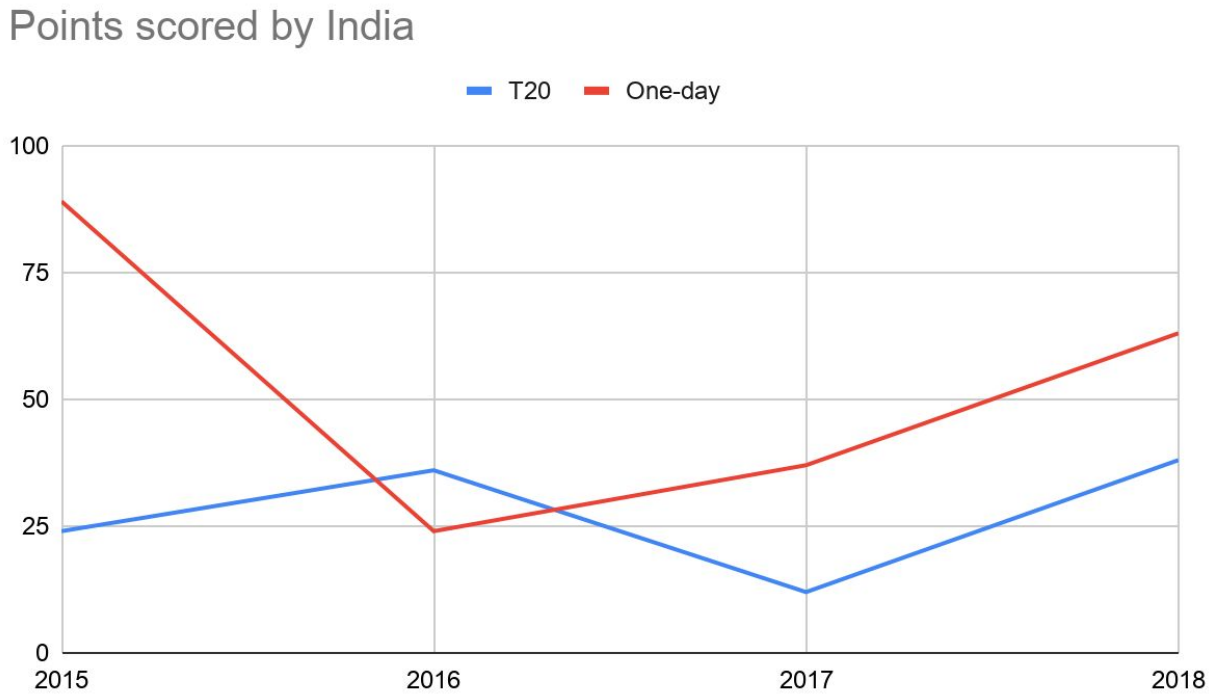


Grid lines

Used to compare two metrics among several samples

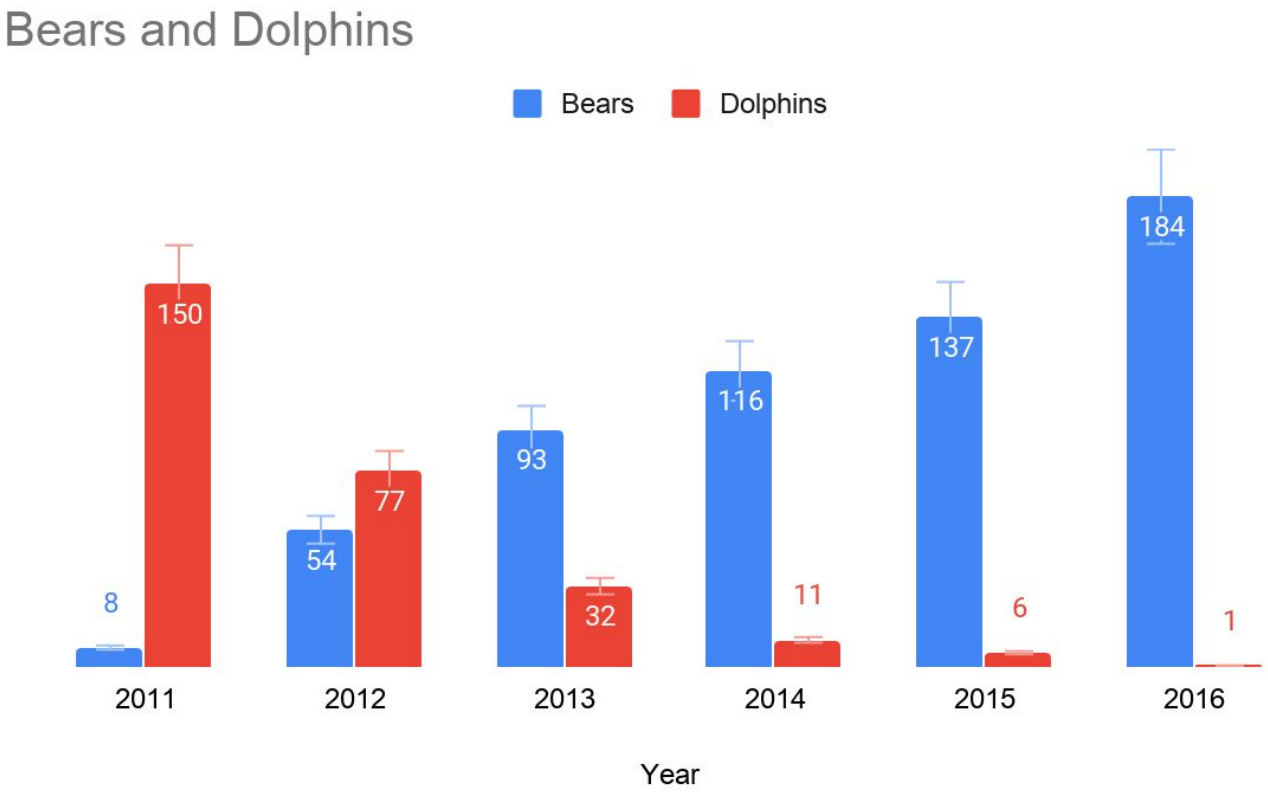


Legend


Points scored by India
T20
One-day
100
75
50
25
0
2015
2016
2017
2018

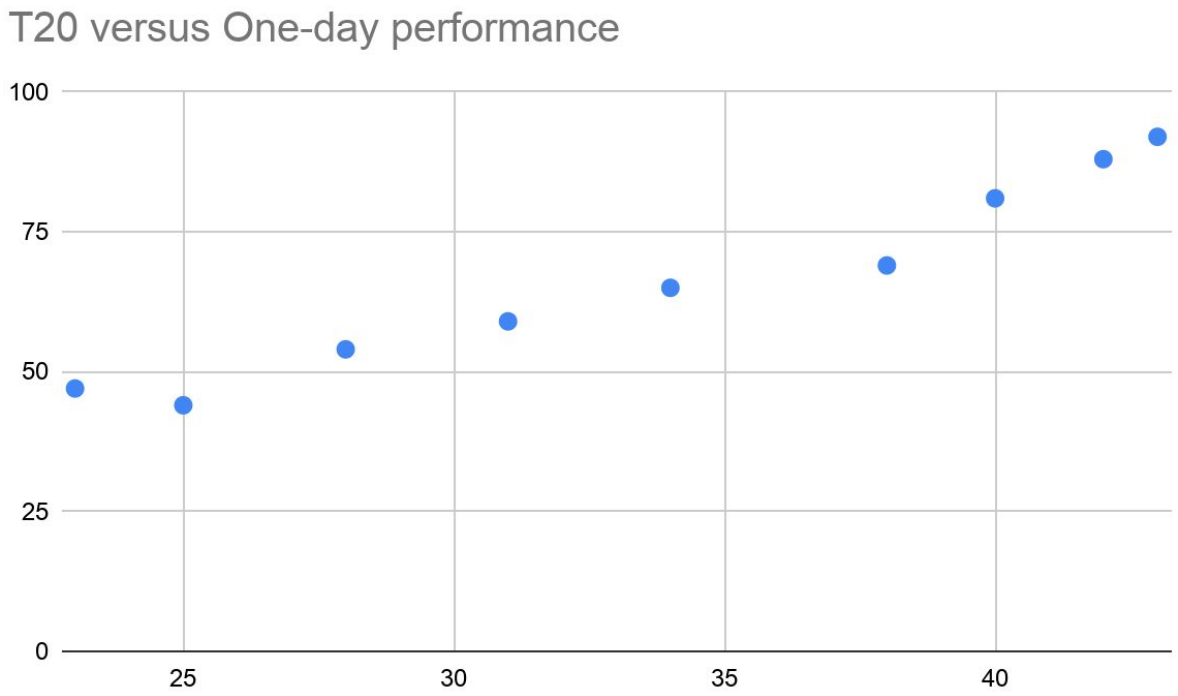
Error bars

Used to compare two metrics among several samples

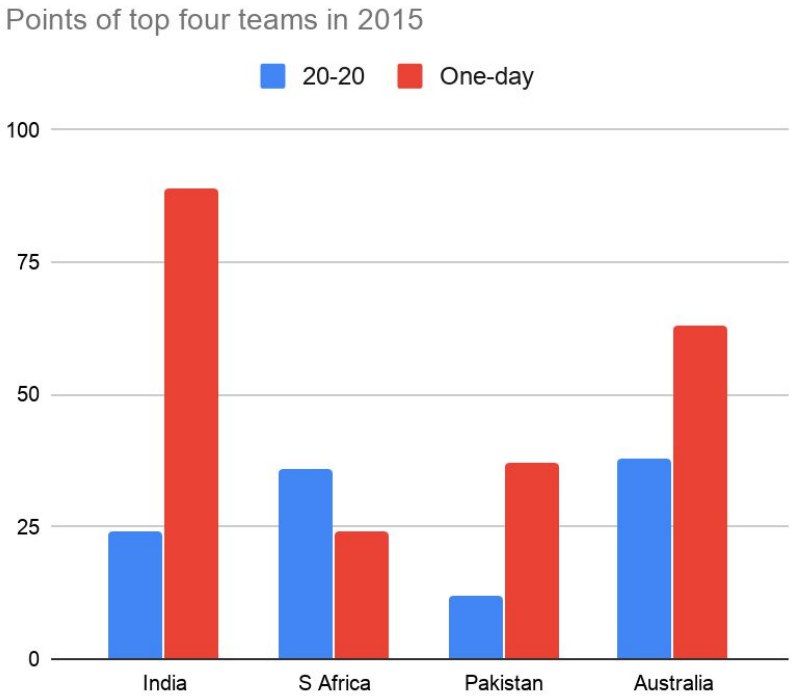


Confidence intervals

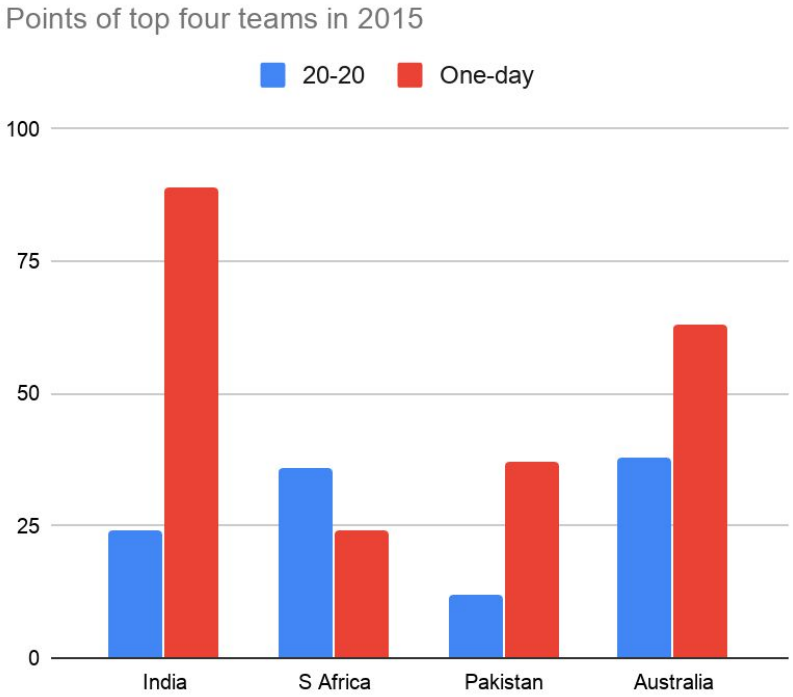
Used to compare two metrics among several samples

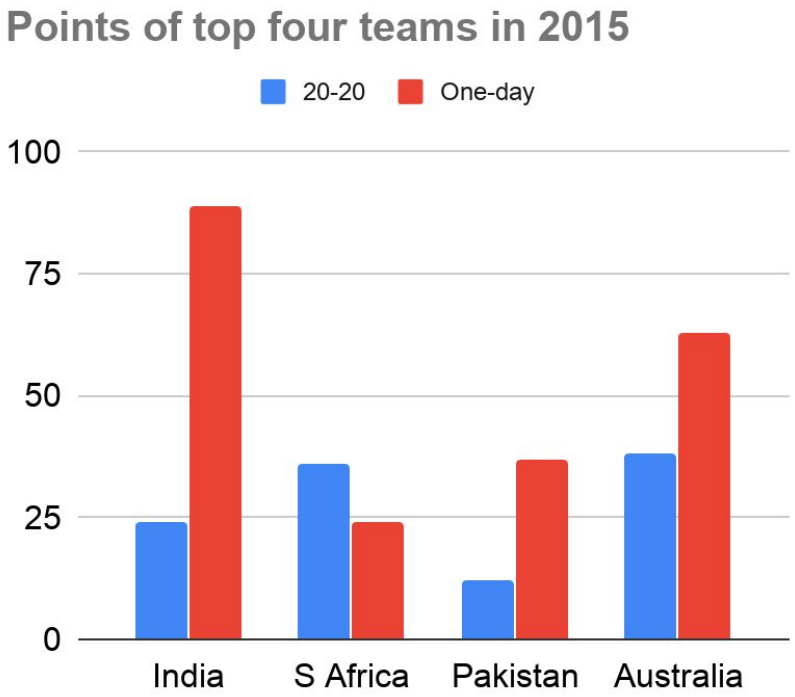


Legibility of fonts

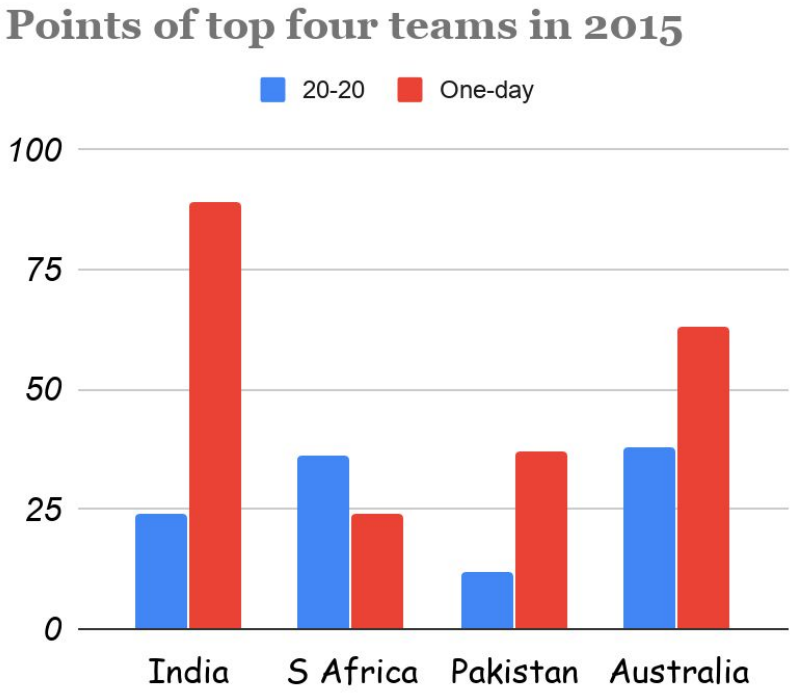


Legibility of fonts

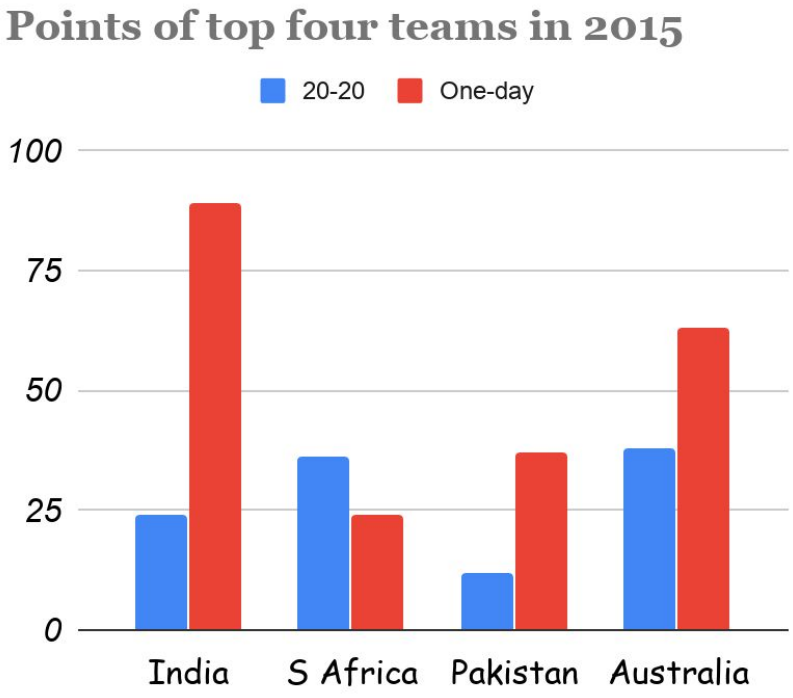


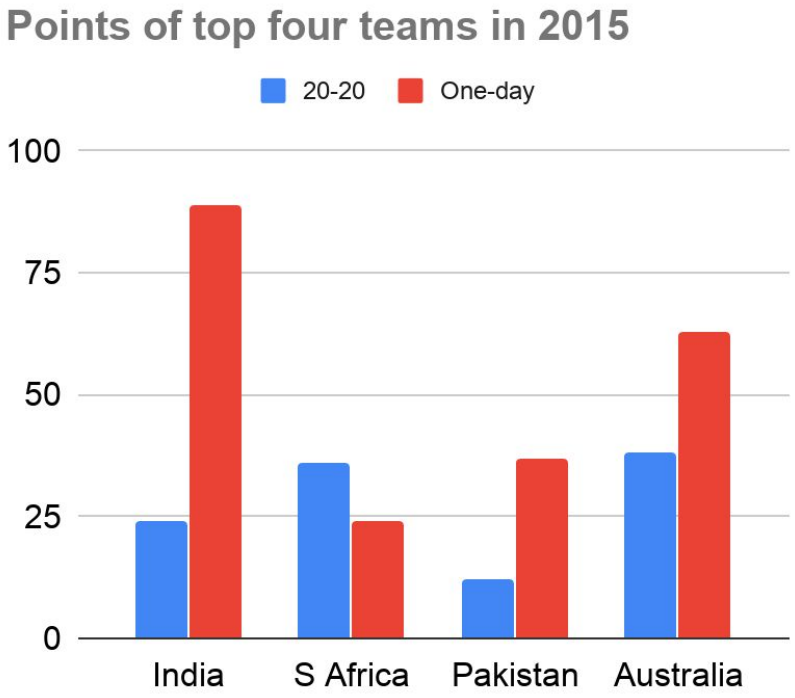


Consistency of fonts

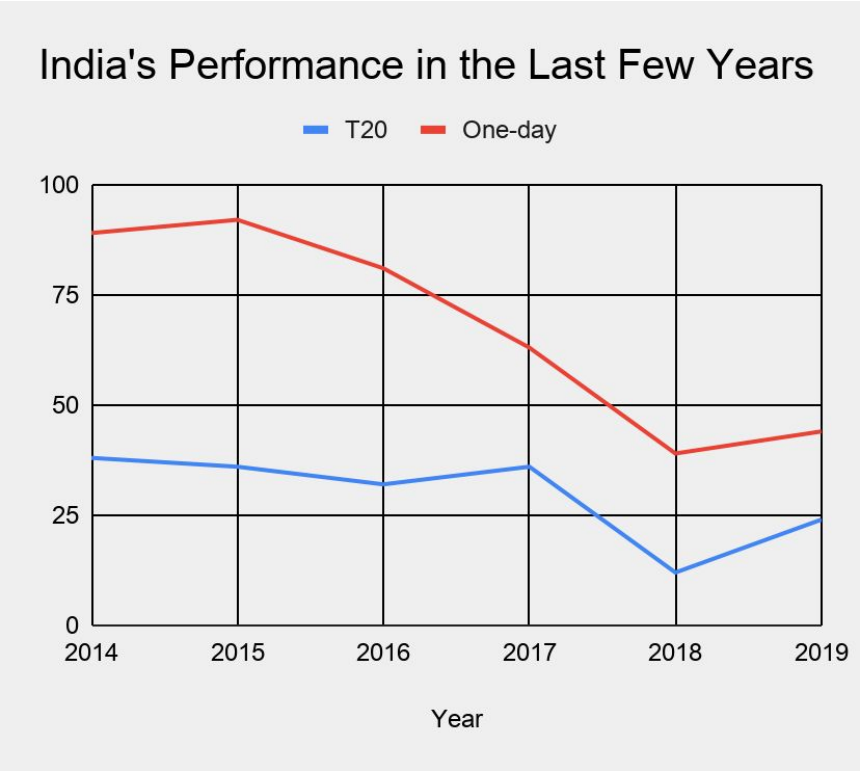


Consistency of fonts

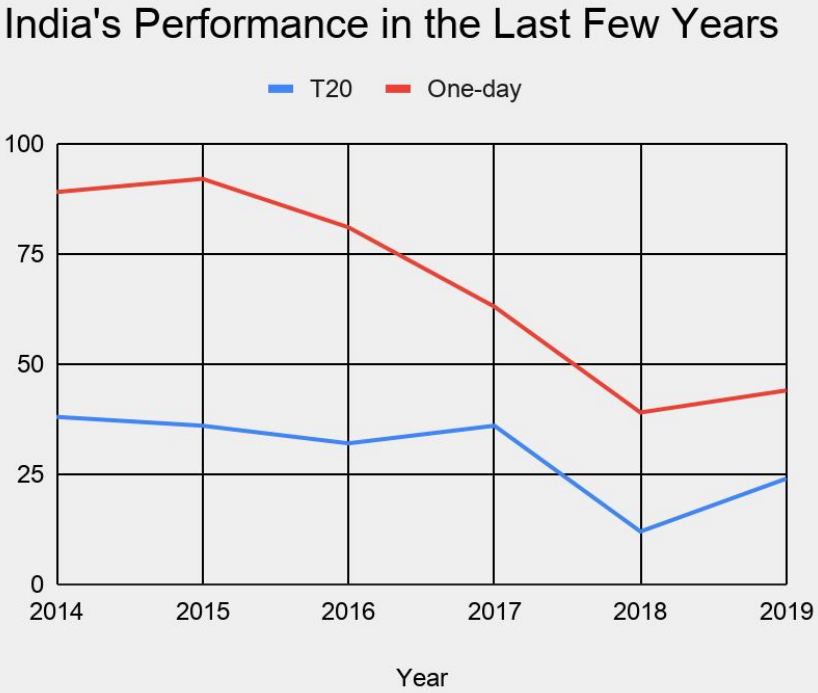


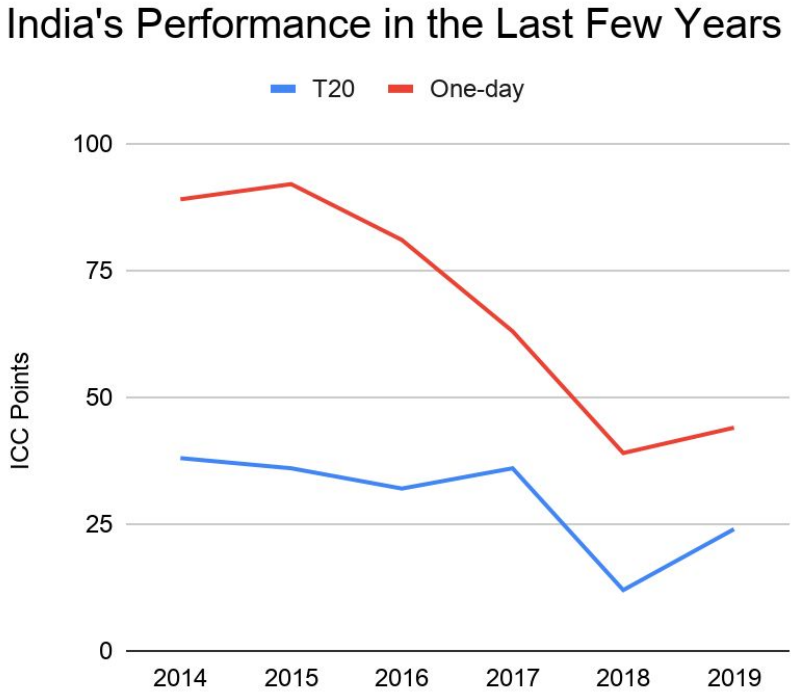


Information to ink ratio (adding whitespace)

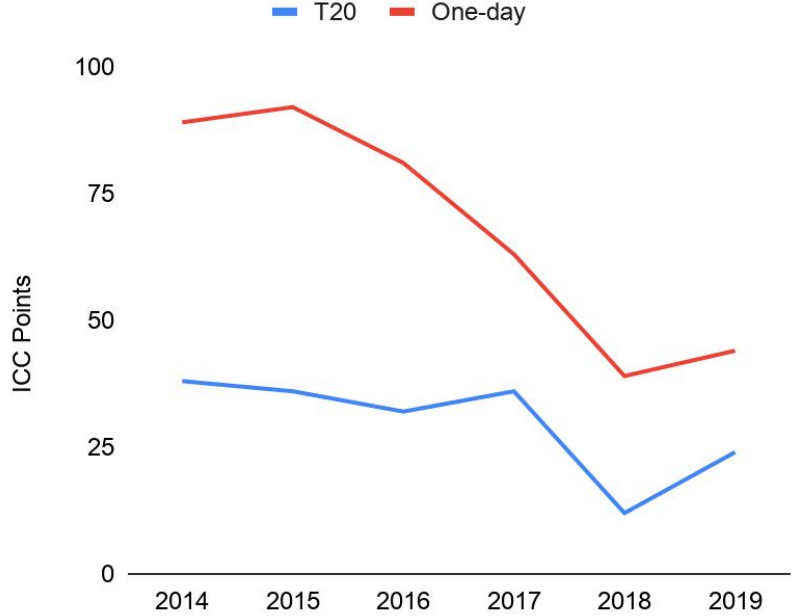


Information to ink ratio (adding whitespace)

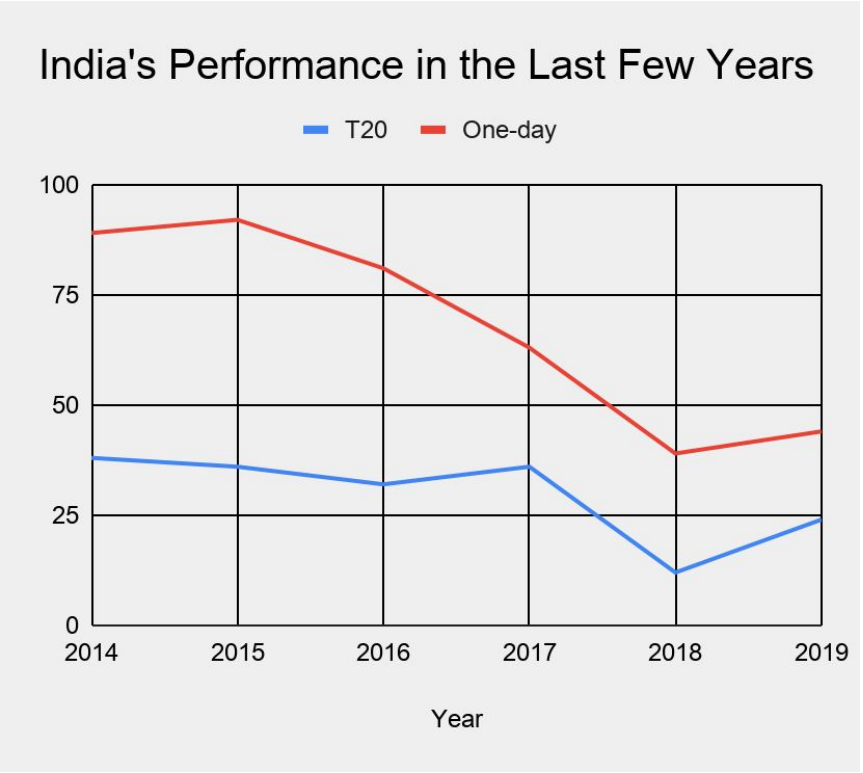


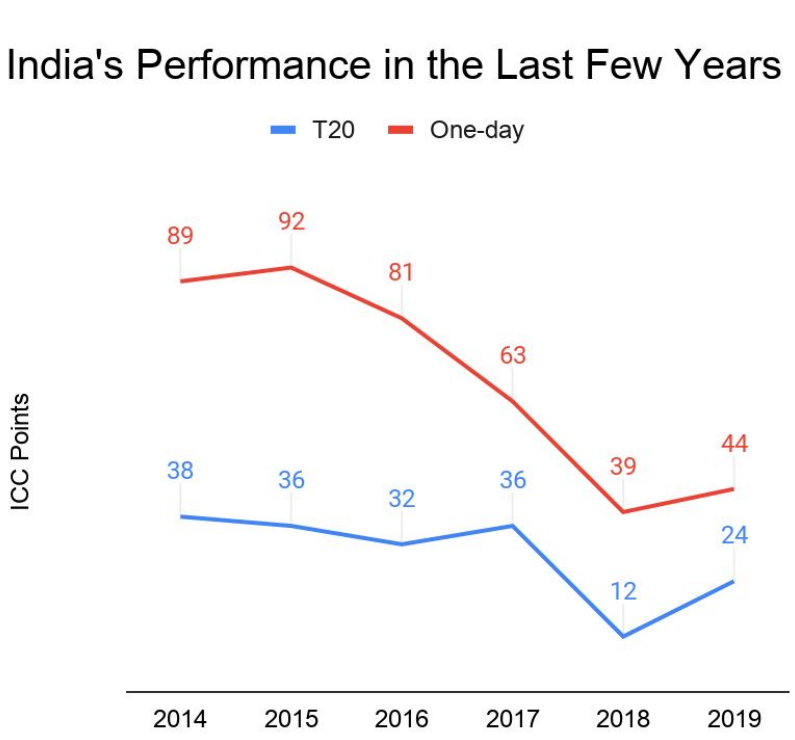


Information to ink ratio (adding whitespace)

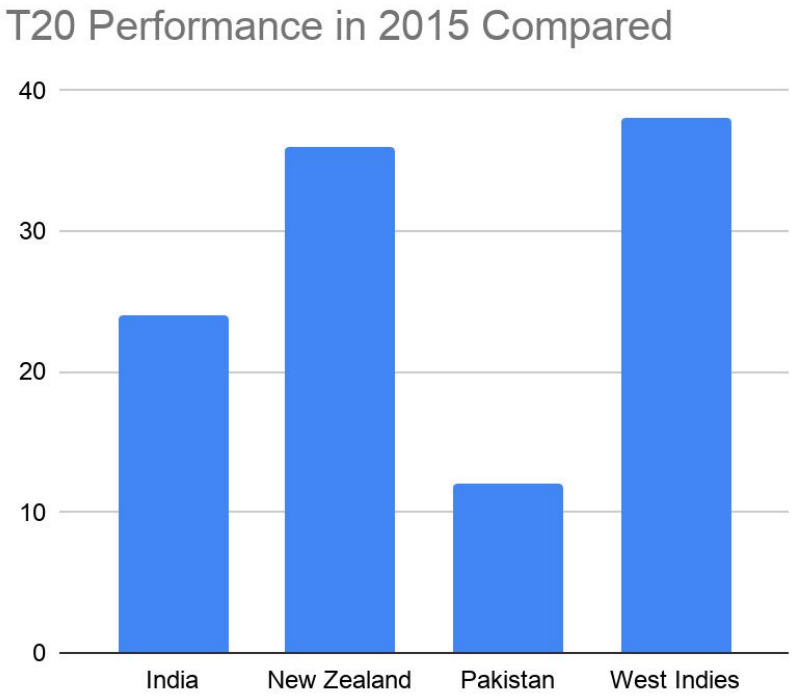


Information to ink ratio (adding whitespace)

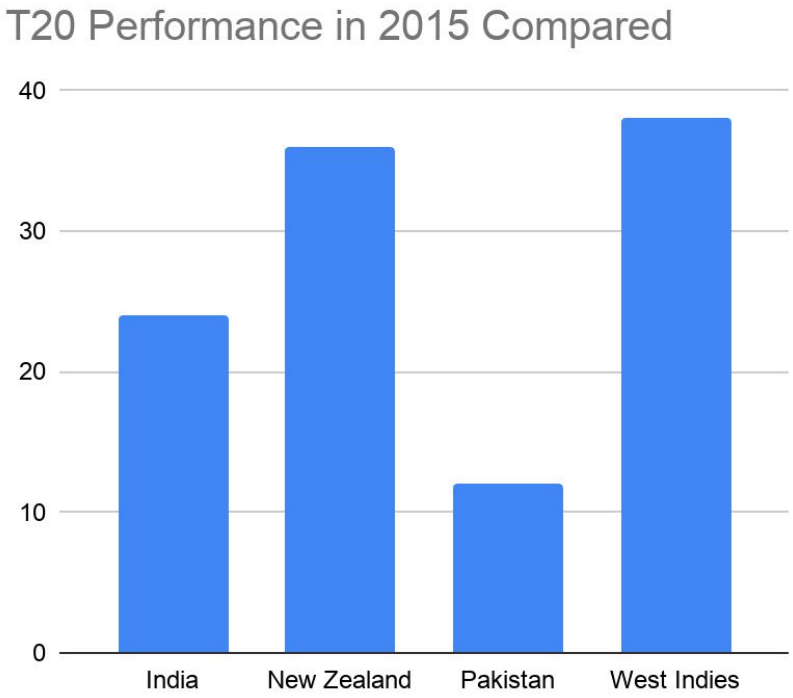


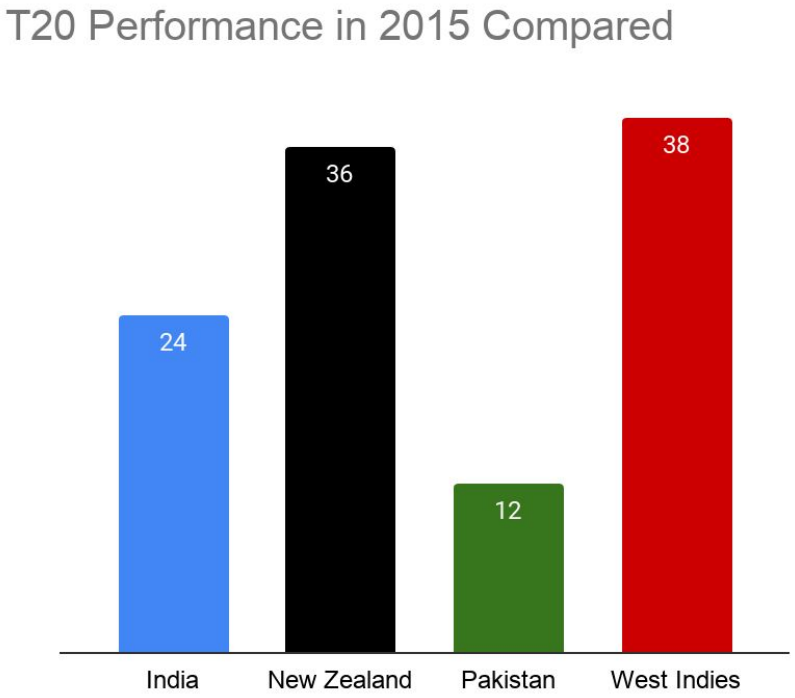


Use of color

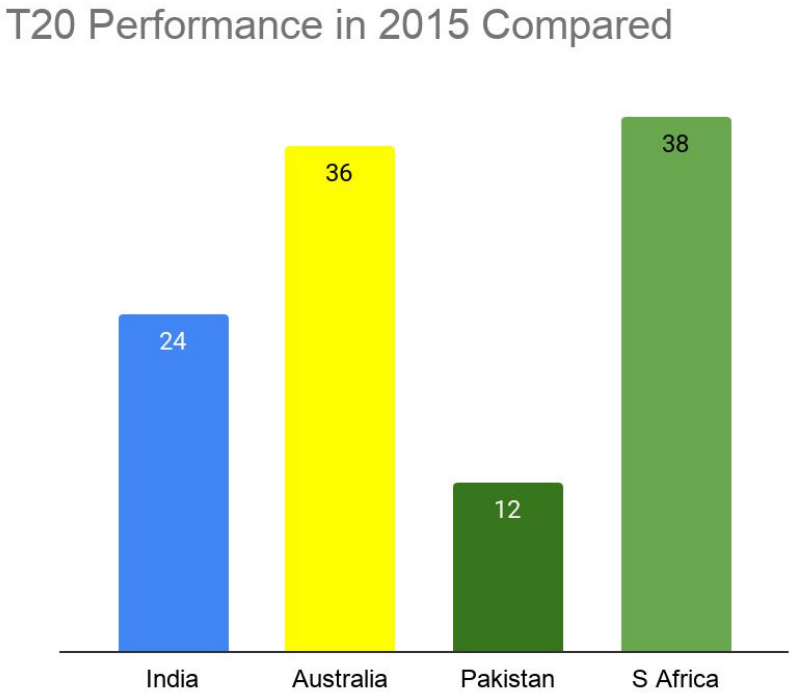

T20 Performance in 2015 Compared
40
30
20
10
0
India
New Zealand
Pakistan
West Indies

Use of color

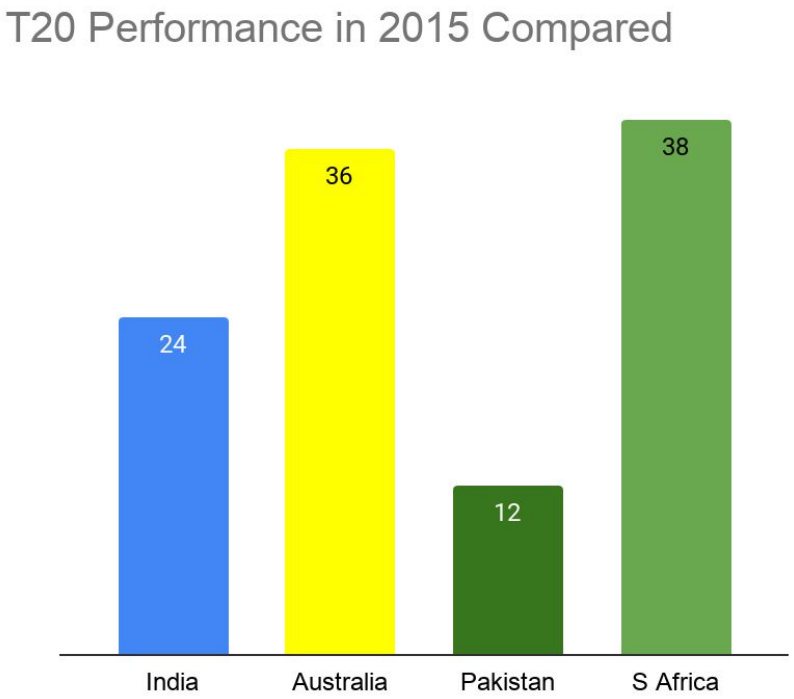

T20 Performance in 2015 Compared
40
30
20
10
0
India
New Zealand
Pakistan
West Indies

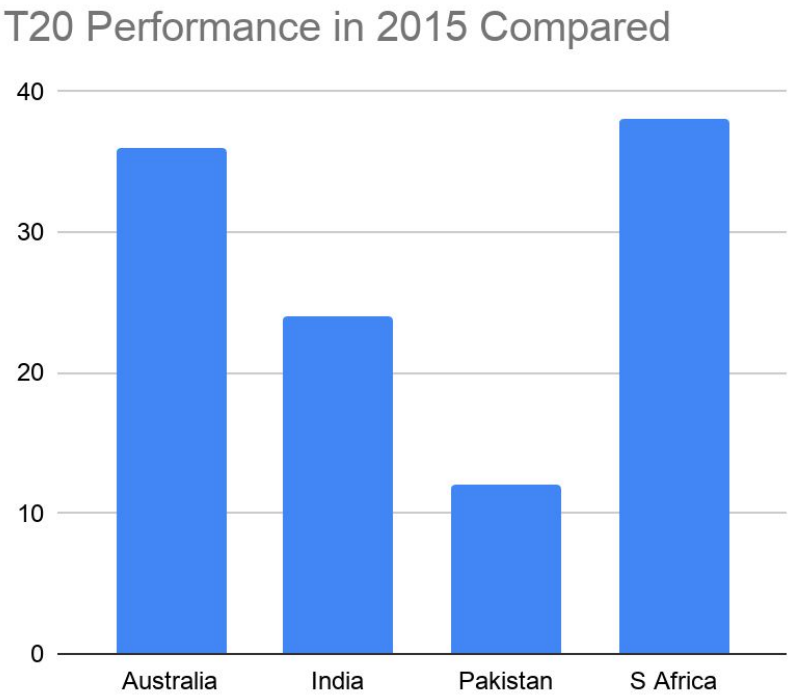

T20 Performance in 2015 Compared
24
36
12
38
India
New Zealand
Pakistan
West Indies

Use of color

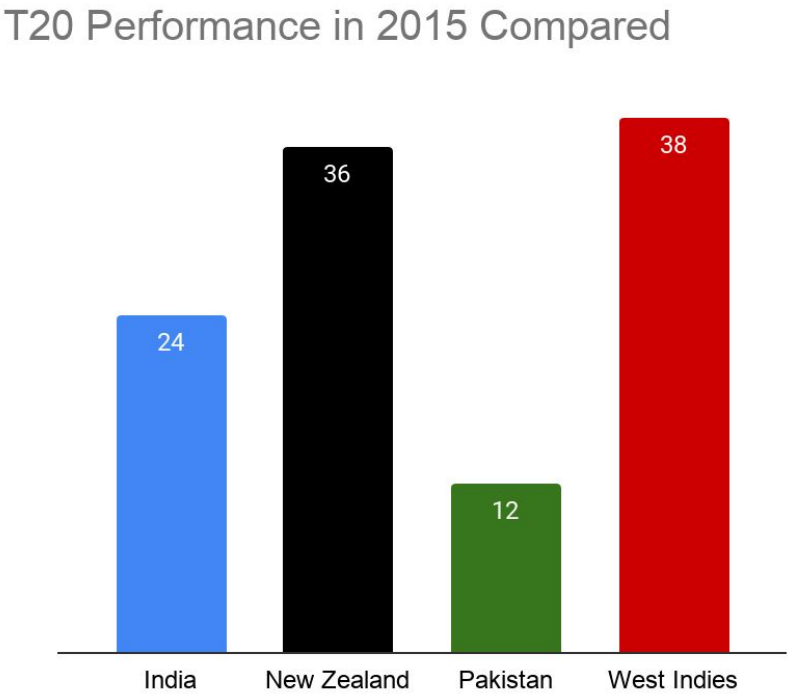

T20 Performance in 2015 Compared
24
36
12
38
India
Australia
Pakistan
S Africa

Use of color

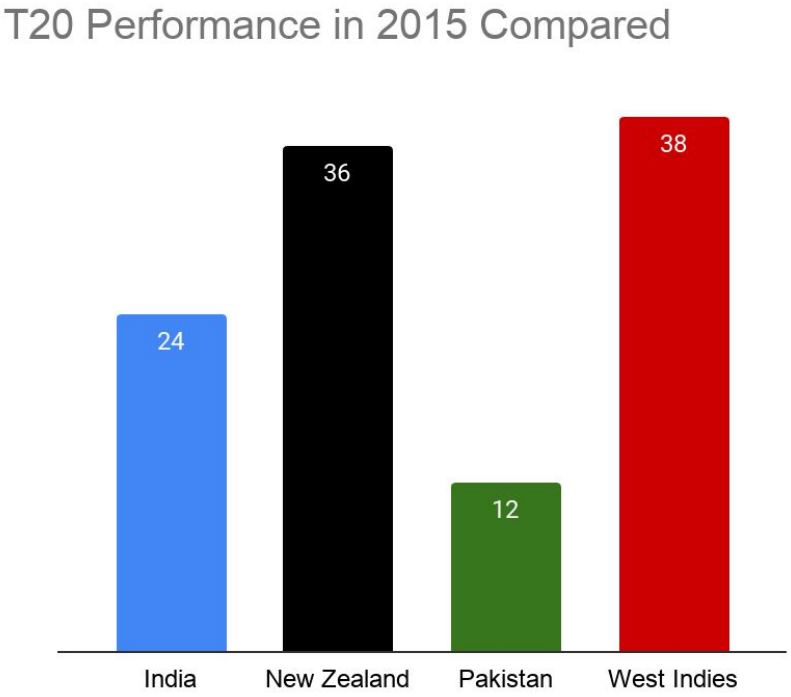

T20 Performance in 2015 Compared
24
36
12
38
India
Australia
Pakistan
S Africa

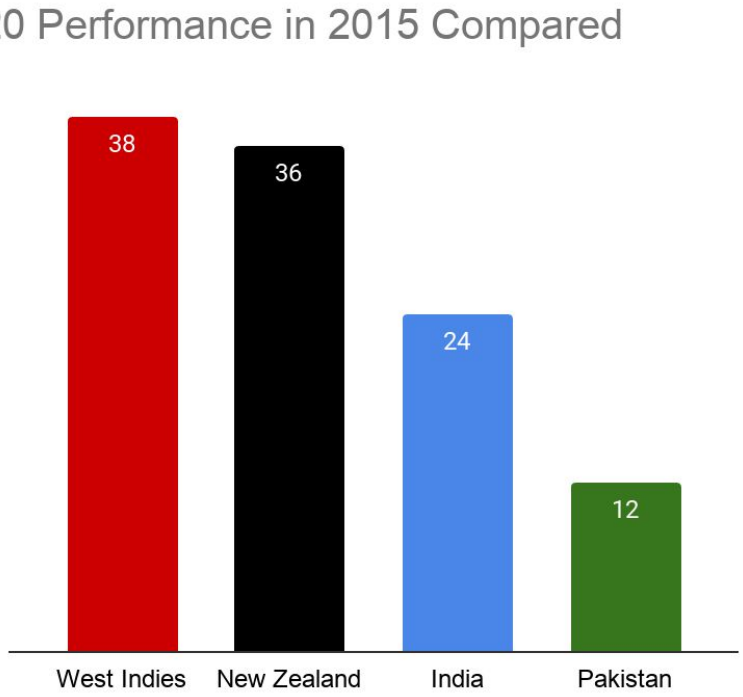

T20 Performance in 2015 Compared
40
30
20
10
0
Australia
India
Pakistan
S Africa

Sorting

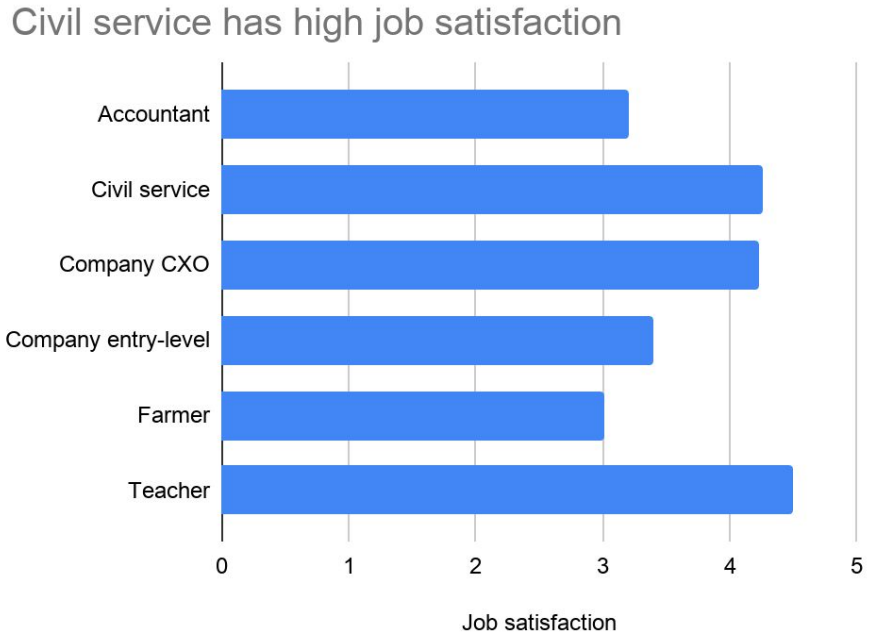

T20 Performance in 2015 Compared
24
36
12
38
India
New Zealand
Pakistan
West Indies

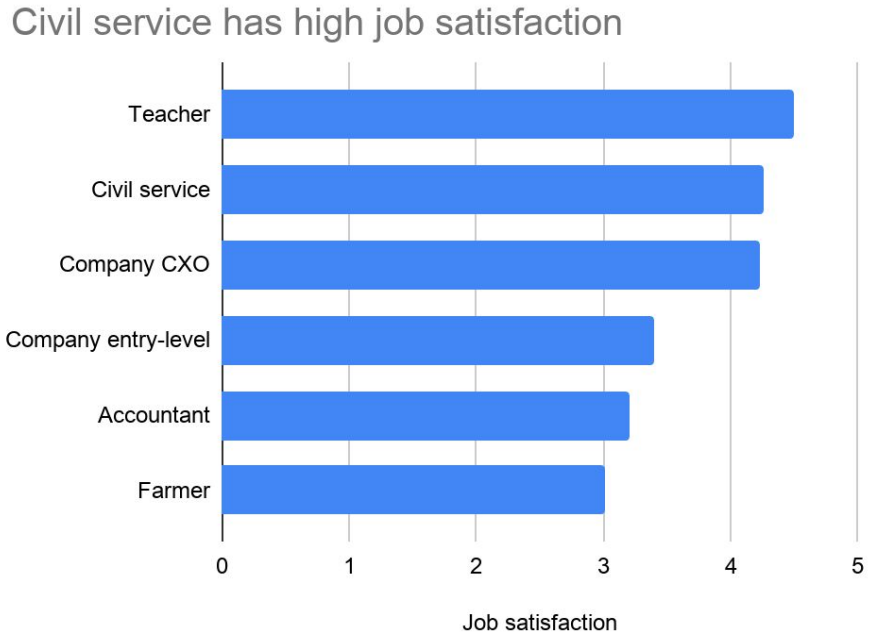
Sorting


T20 Performance in 2015 Compared
24
36
12
38
India
New Zealand
Pakistan
West Indies

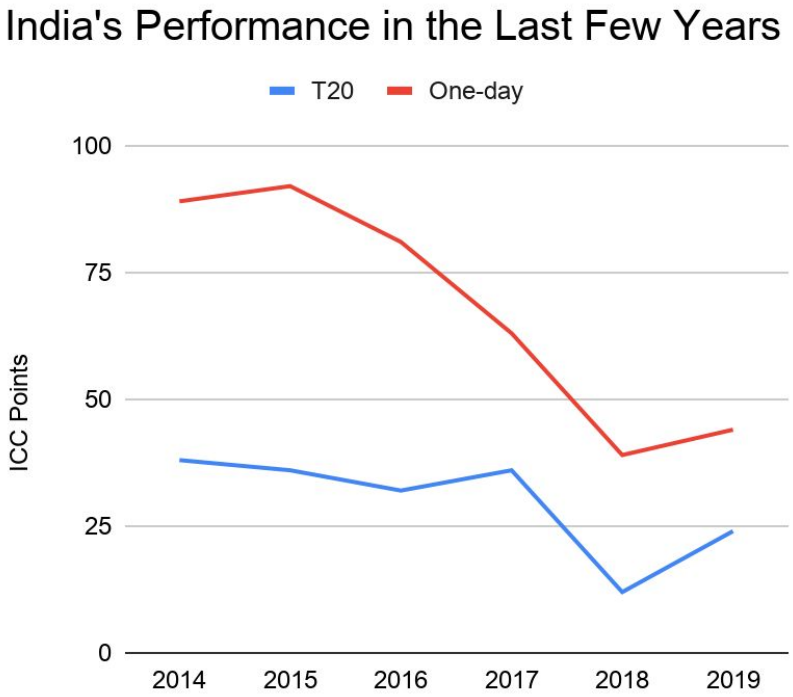

T20 Performance in 2015 Compared
38
36
24
12
West Indies
New Zealand
India
Pakistan

Sorting alphabetically or by value?

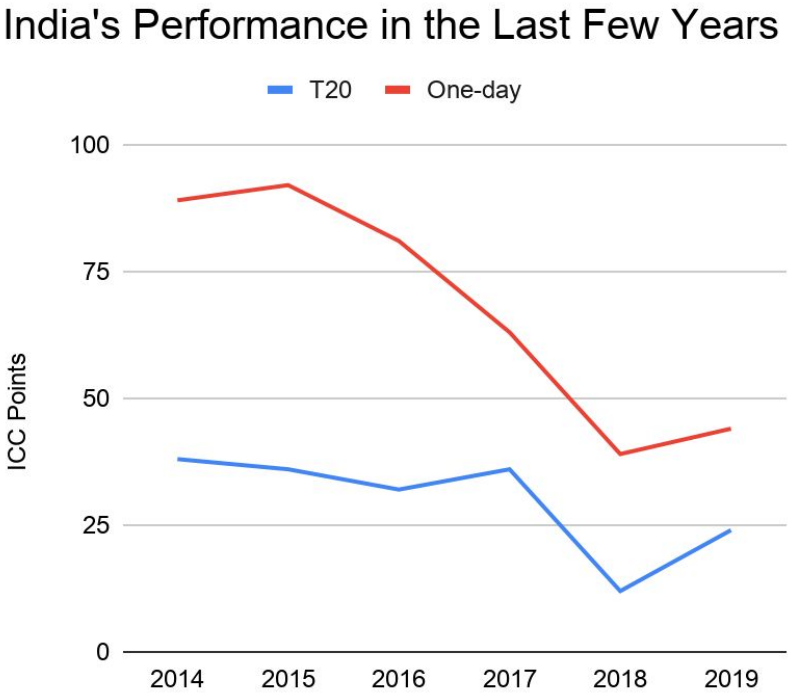


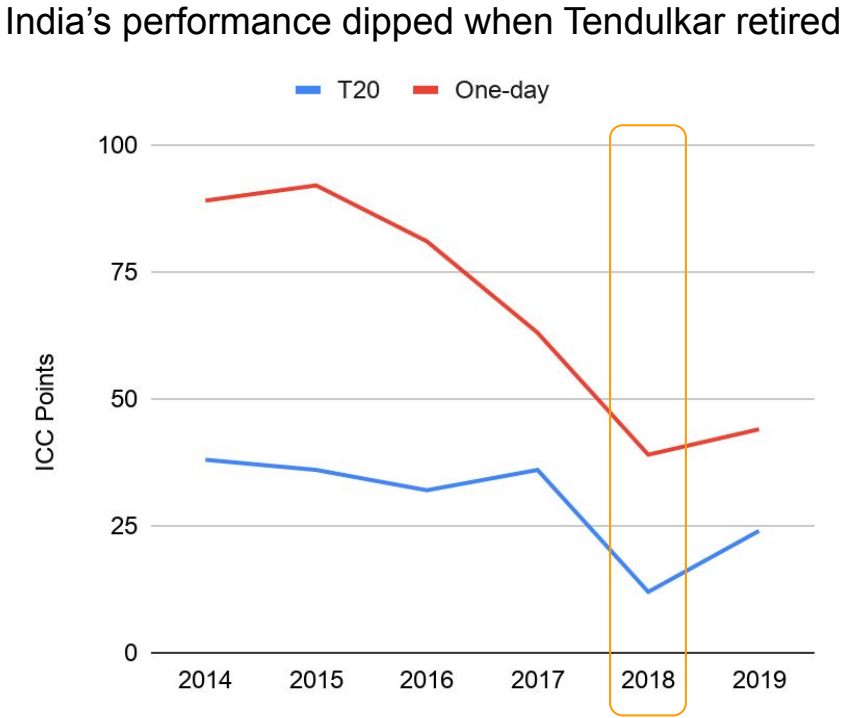


Boxes

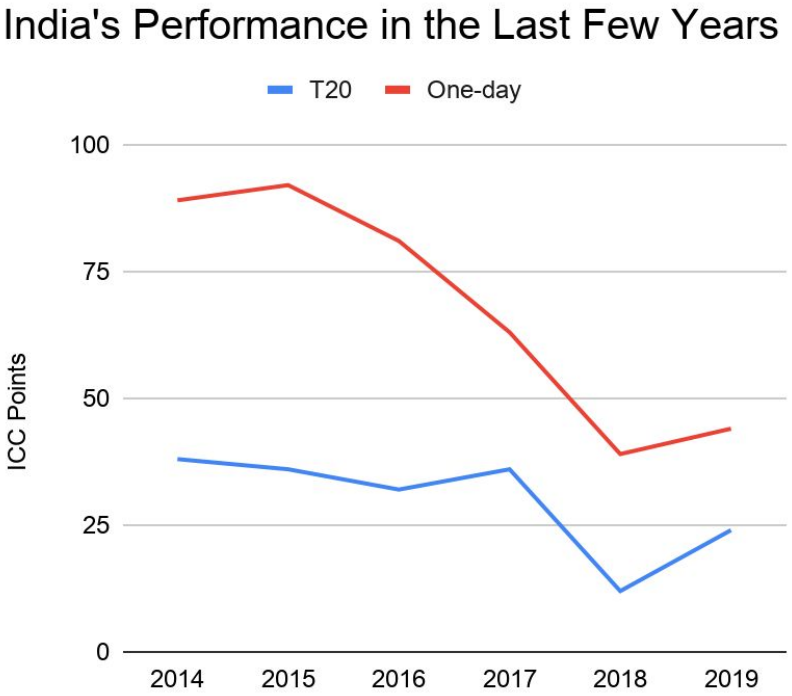

India's Performance in the Last Few Years
T20
One-day
100
75
50
25
0
ICC Points
2014
2015
2016
2017
2018
2019

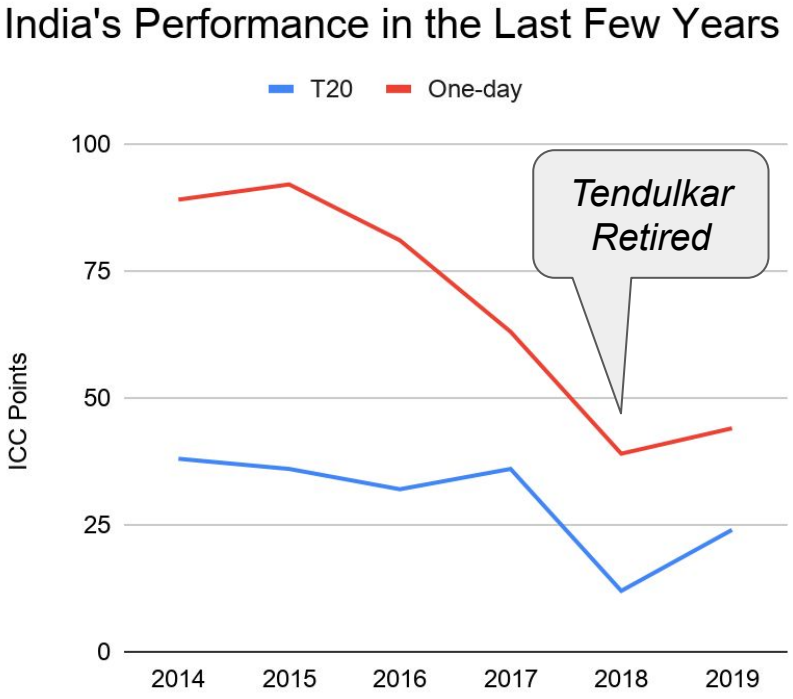
Boxes


India's Performance in the Last Few Years
T20
One-day
100
75
50
25
0
ICC Points
2014
2015
2016
2017
2018
2019

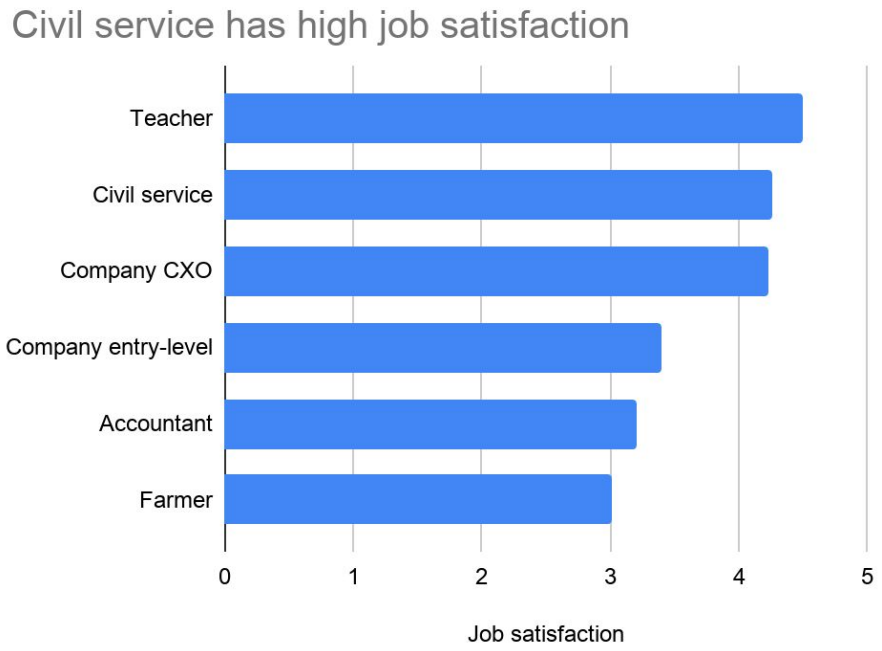

India’s performance dipped when Tendulkar retired
T20
One-day
100
75
50
25
0
ICC Points
2014
2015
2016
2017
2018
2019

Call-outs and pointers

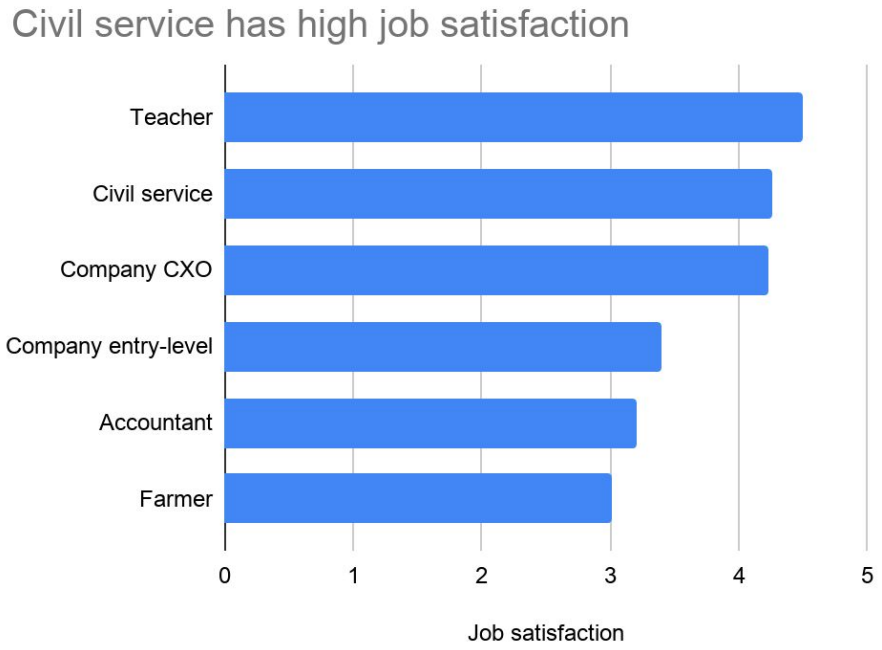


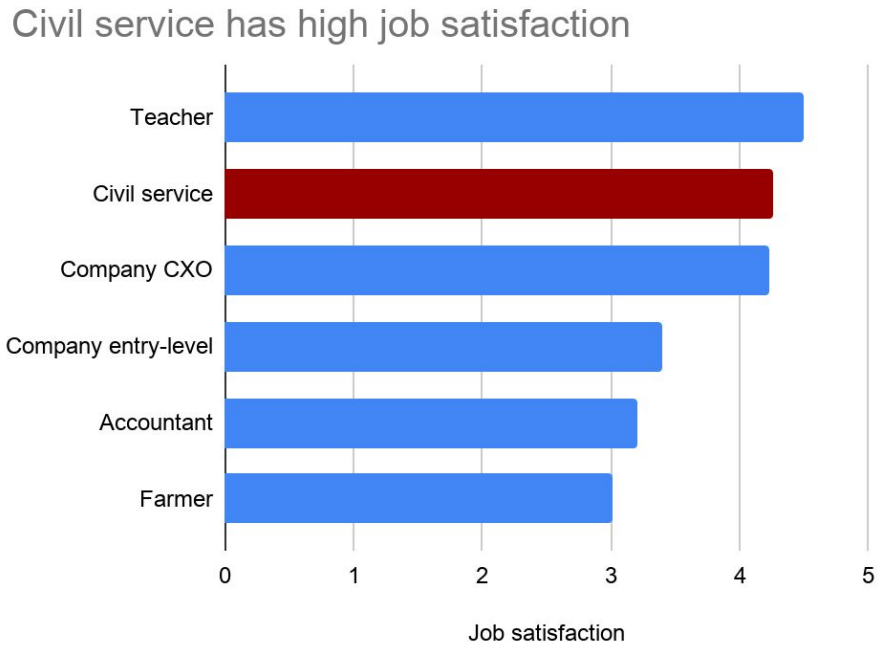


Use of color

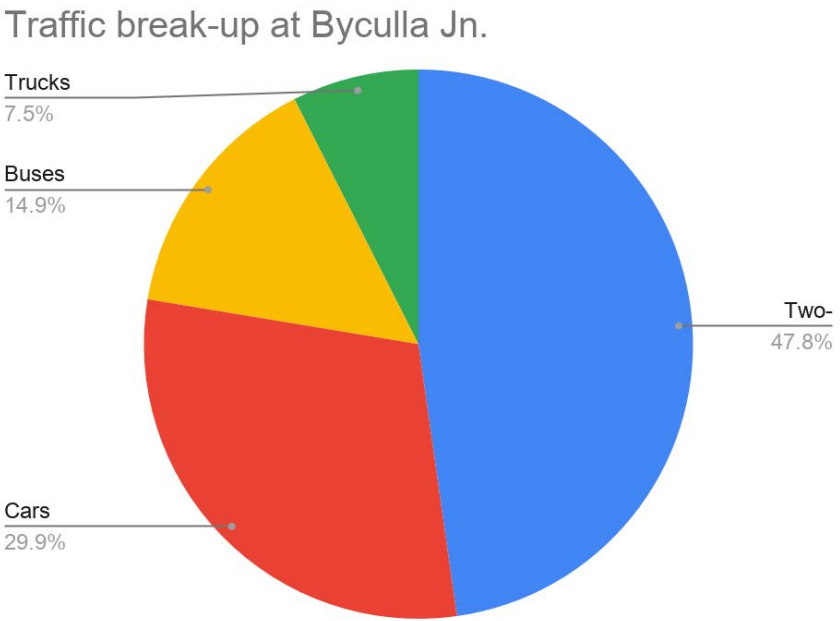

Civil service has high job satisfaction
Teacher
Civil service
Company CXO
Company entry-level
Accountant
Farmer
0
1
2
3
4
5
Job satisfaction

Use of color

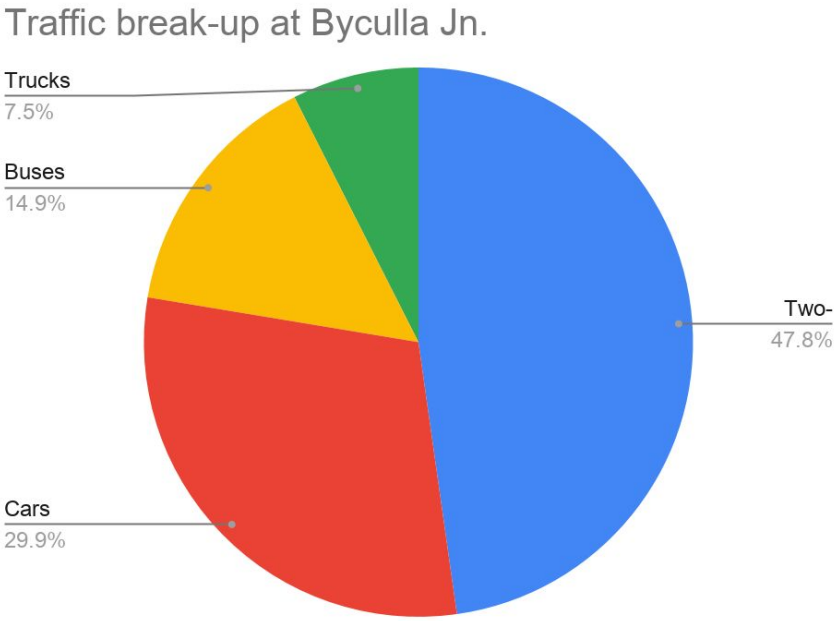

Civil service has high job satisfaction
Teacher
Civil service
Company CXO
Company entry-level
Accountant
Farmer
0
1
2
3
4
5
Job satisfaction

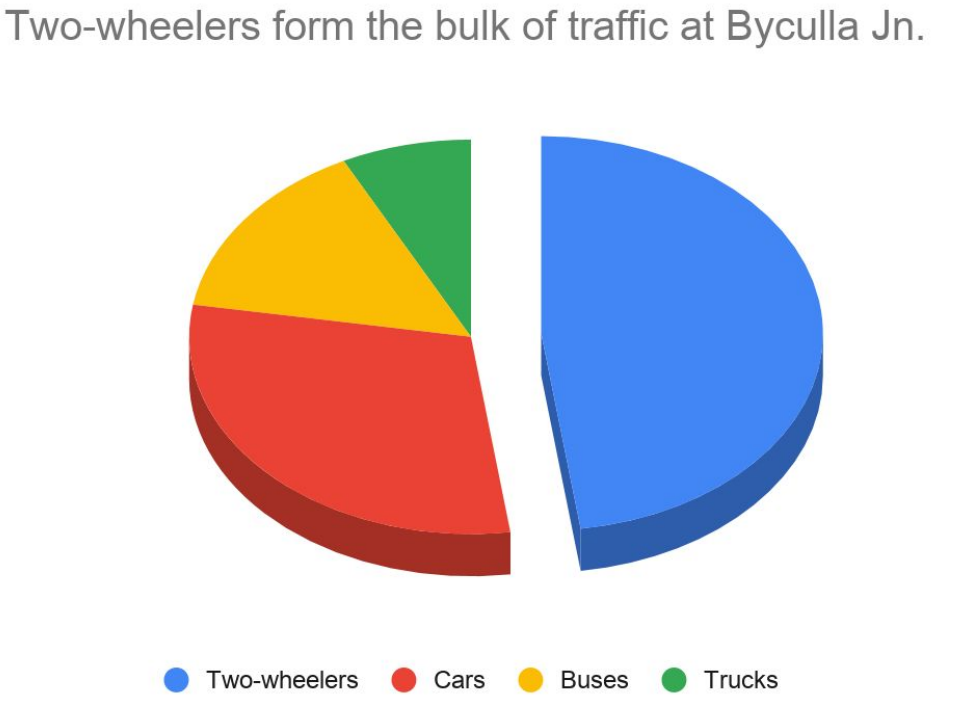

Civil service has high job satisfaction
Teacher
Civil service
Company CXO
Company entry-level
Accountant
Farmer
0
1
2
3
4
5
Job satisfaction

Pie chart with a breakout



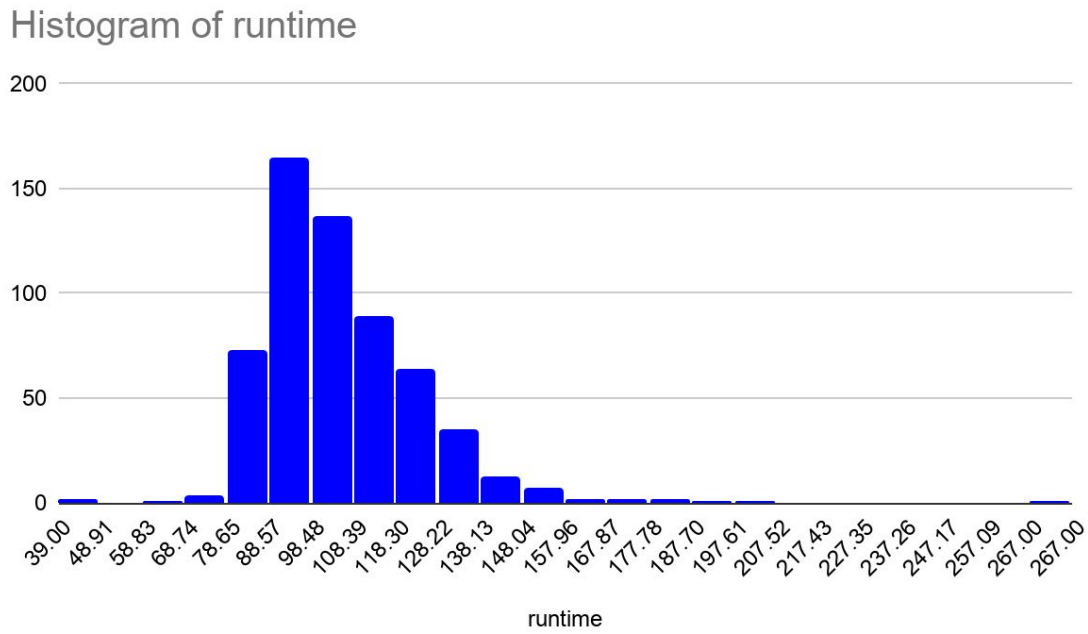
Pie chart with a breakout





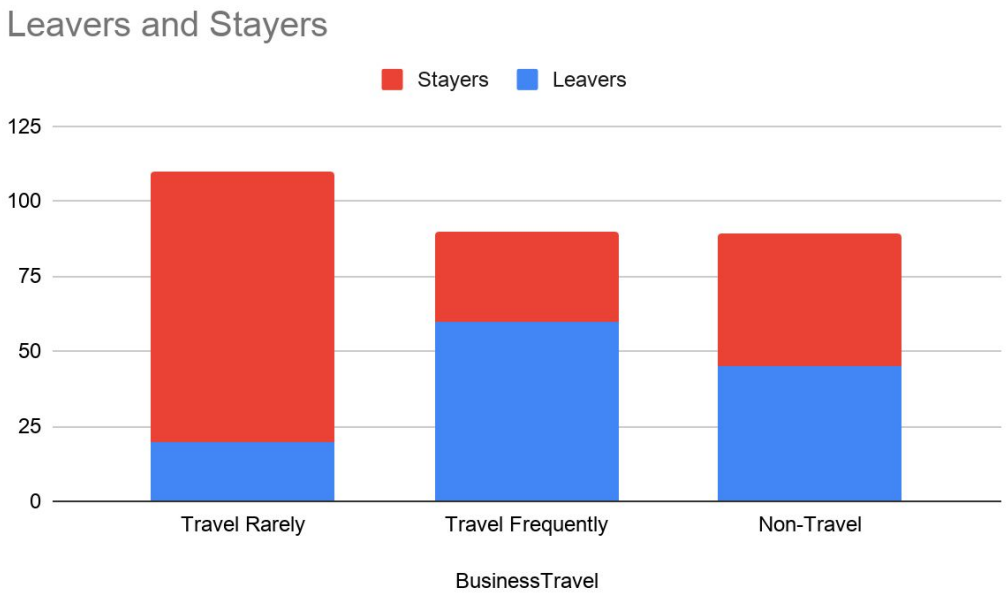
Histogram

Used to compare two metrics among several samples



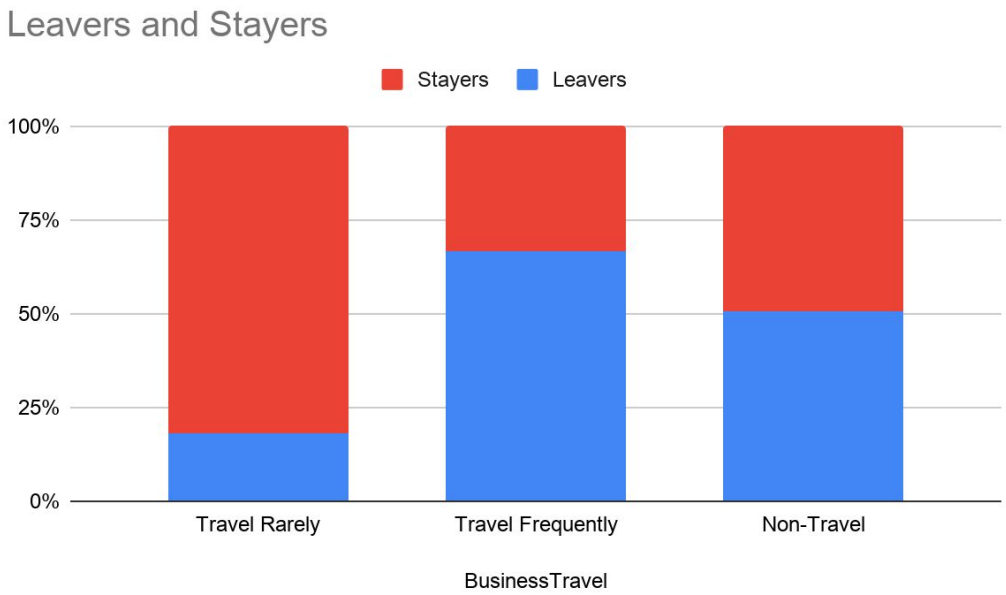
Stacked bar

Used to compare two metrics among several samples



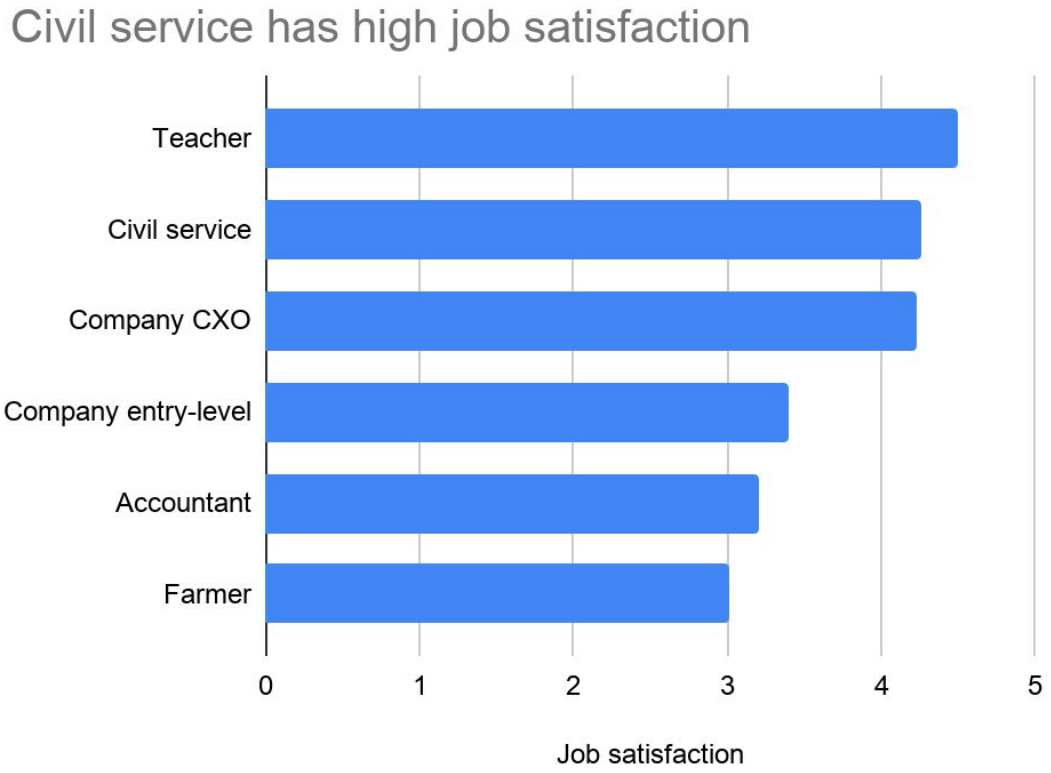
100% Stacked bar

Used to compare two metrics among several samples



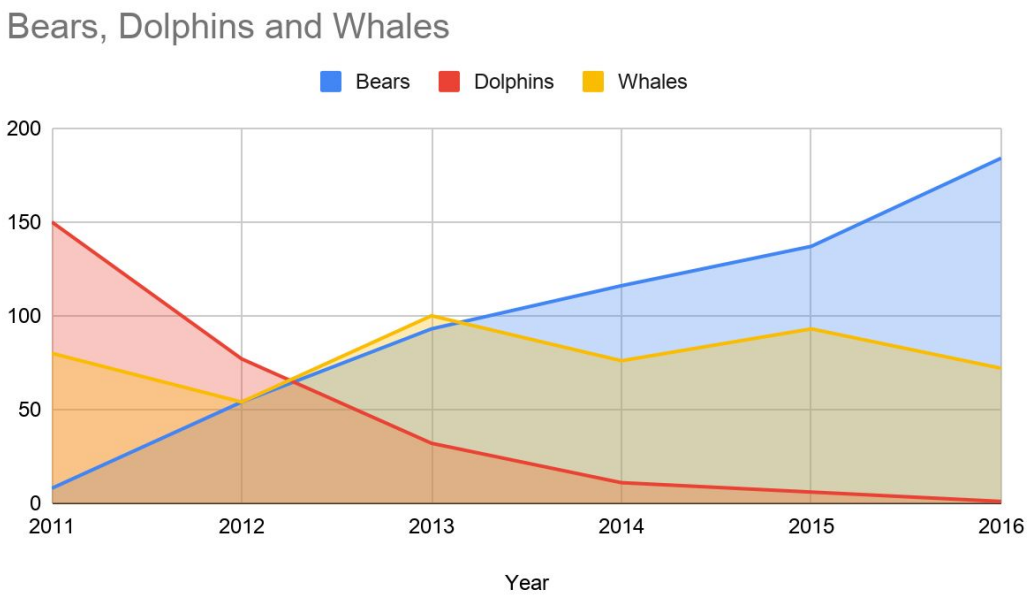
Column chart

Like bar chart, but rotated 90 degrees; always for nominal categories



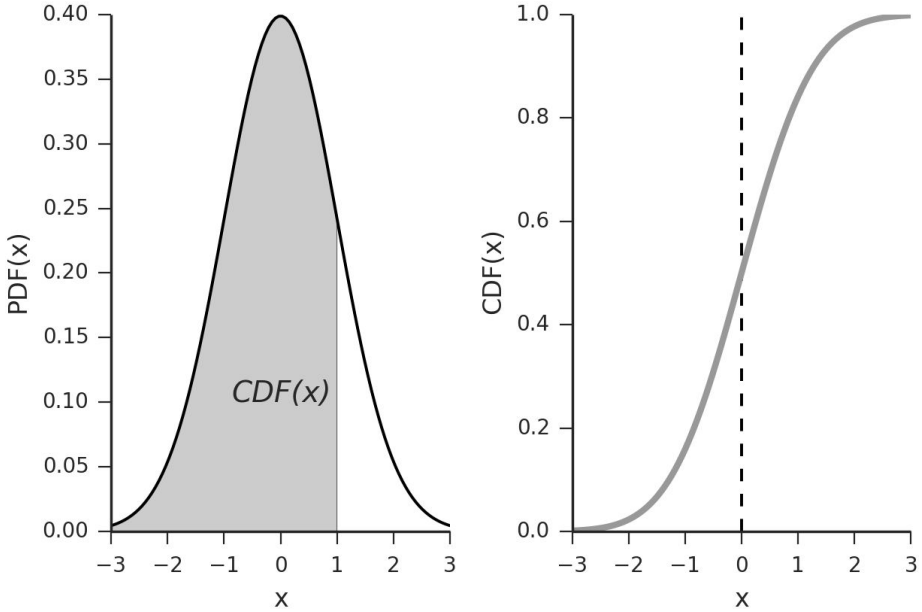
Area chart

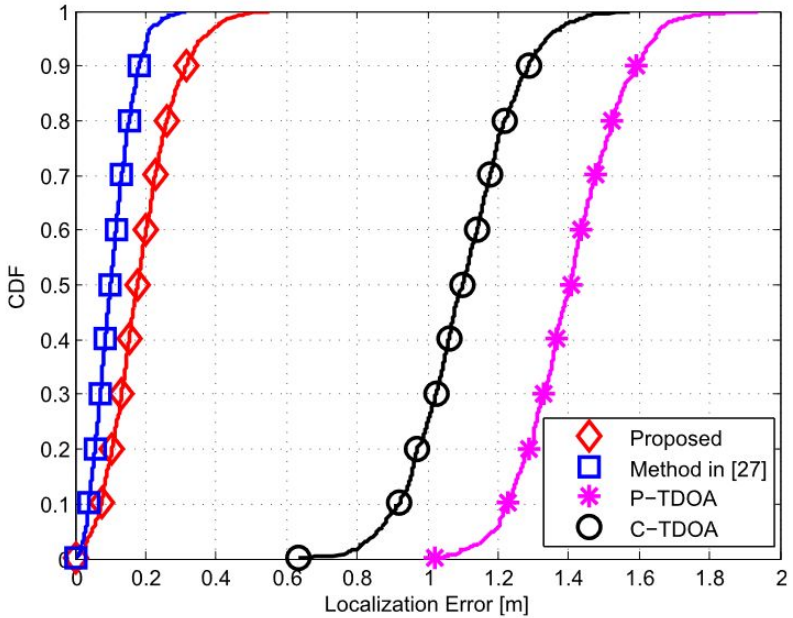
Can be used to compare two histograms



Cumulative distribution function

Can be used to compare two distributions; non-decreasing and goes from 0 to 1.

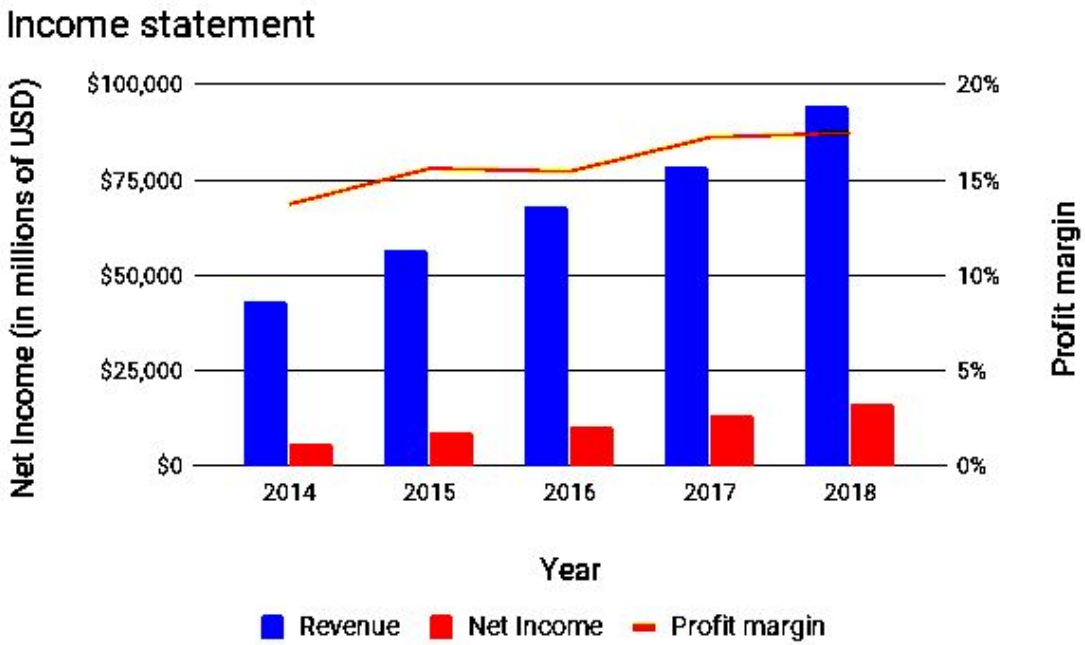




Sources: deepai.org; “A Hybrid Localization Algorithm Based on TOF and TDOA for Asynchronous Wireless Sensor Networks.” Wang et al. IEEE Access, 2019

Bar and line chart with dual y-axis

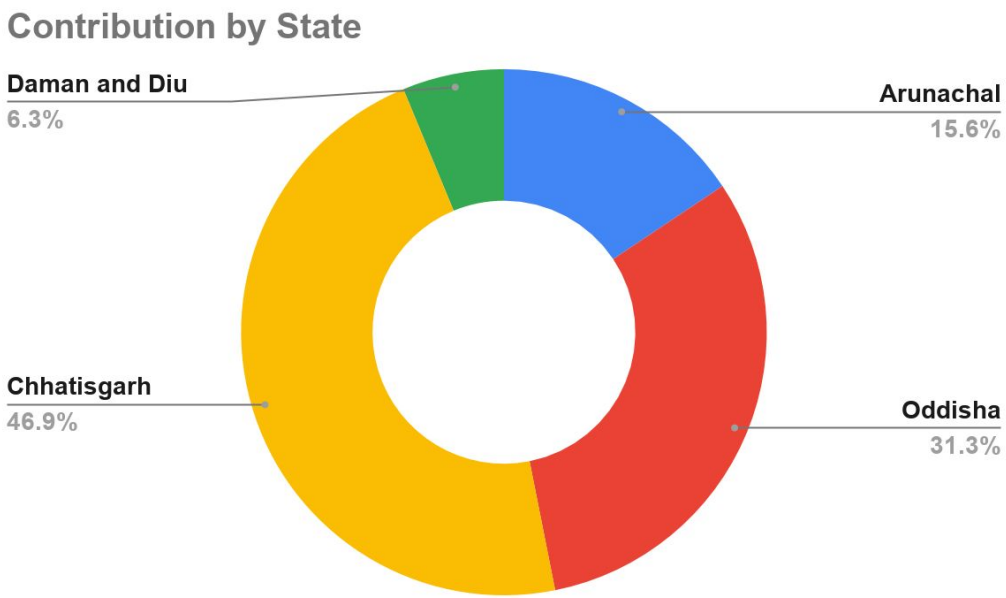
Used to compare two metrics among several samples



Source: https://support.google.com/docs/answer/9142593?hl=en

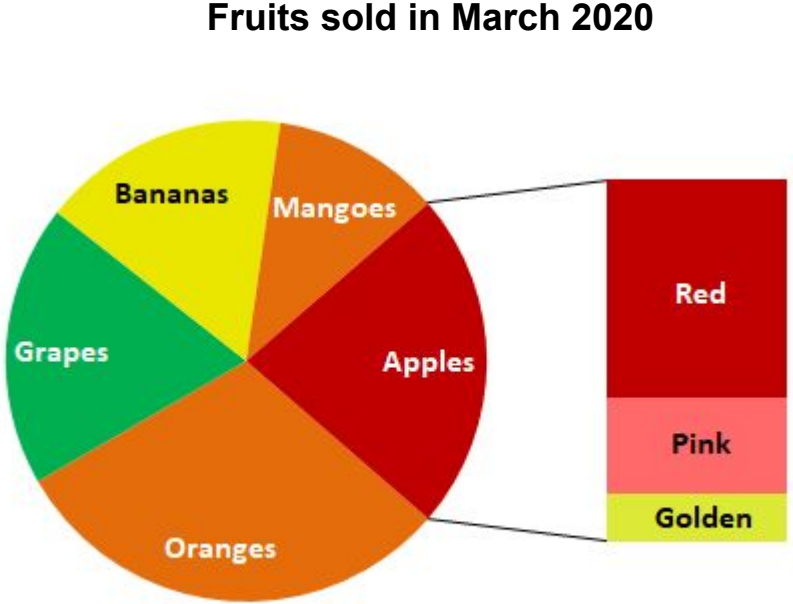
Doughnut chart

Like a pie chart, but with a hole in the middle



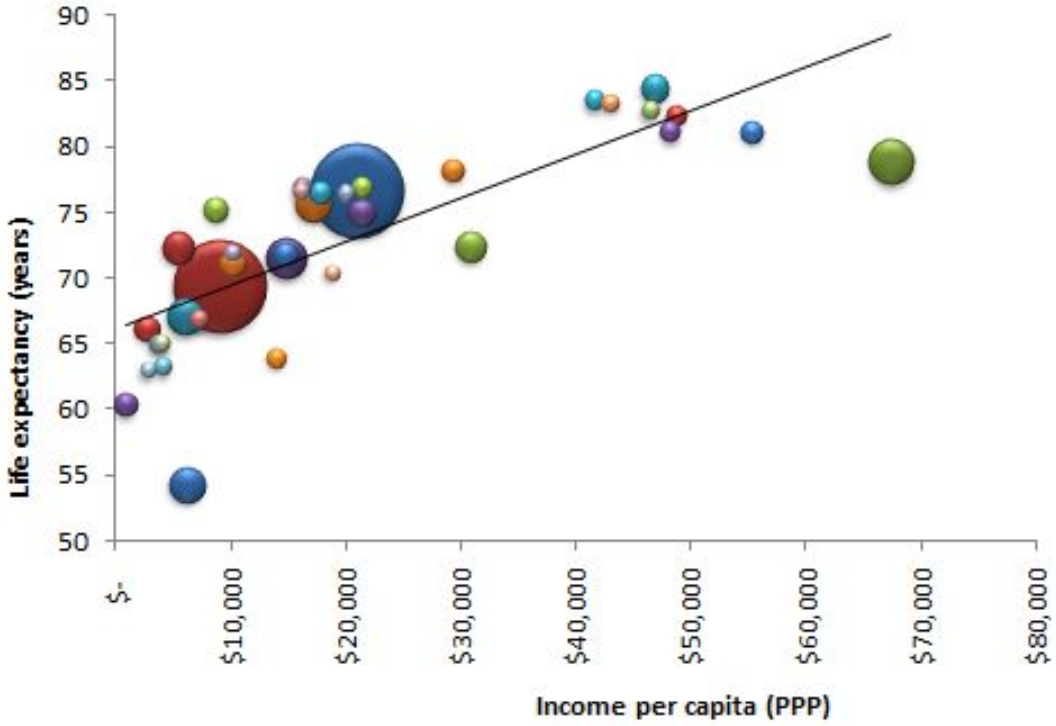
Mixed charts, e.g. pie and stacked chart

Show breakup of something using a pie, and then take one component and break it further using a bar



Bubble chart

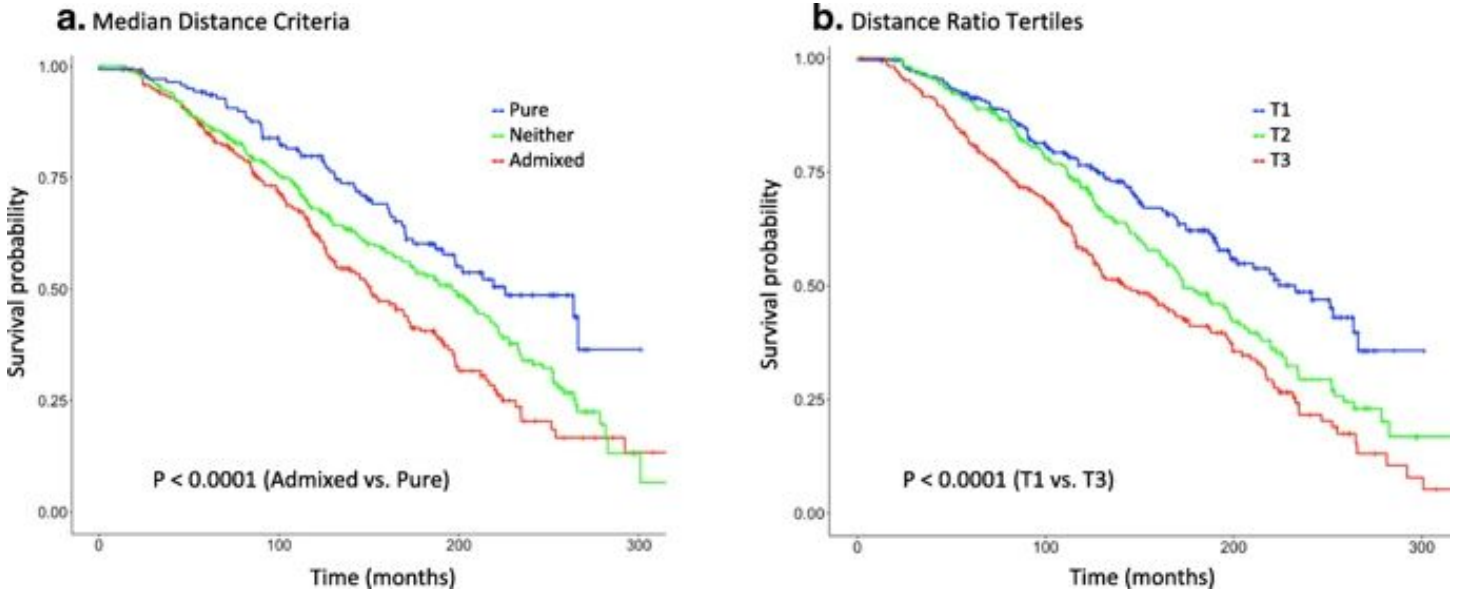
Scatter plot with a third variable for size



Data Source: Wikipedia.org

Survival (or Kaplan-Meier) curves

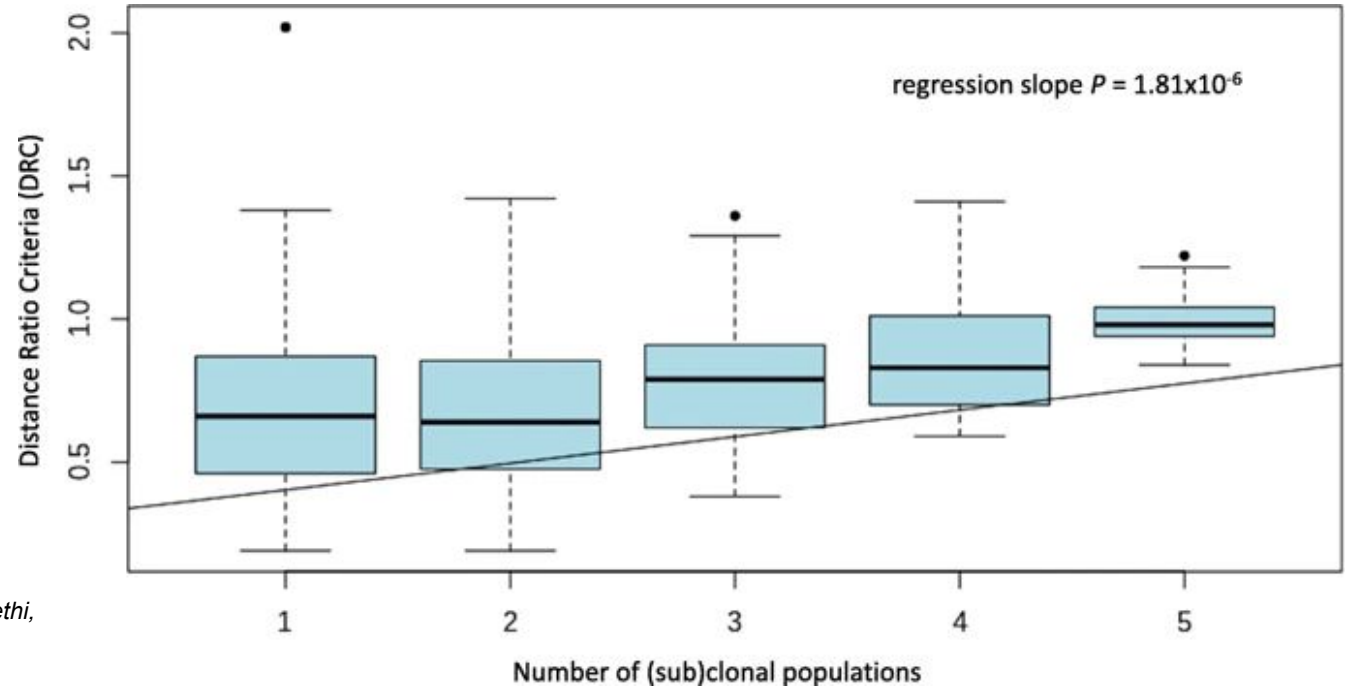
To show when failure occurs



Source: "Quantification of intrinsic subtype ambiguity in Luminal A breast cancer and its relationship to clinical outcomes" by Kumar, Zhao, Bhaumik, Sethi, Gann, in BMC Cancer 2019

Box and whiskers plot

To show slices of a distribution

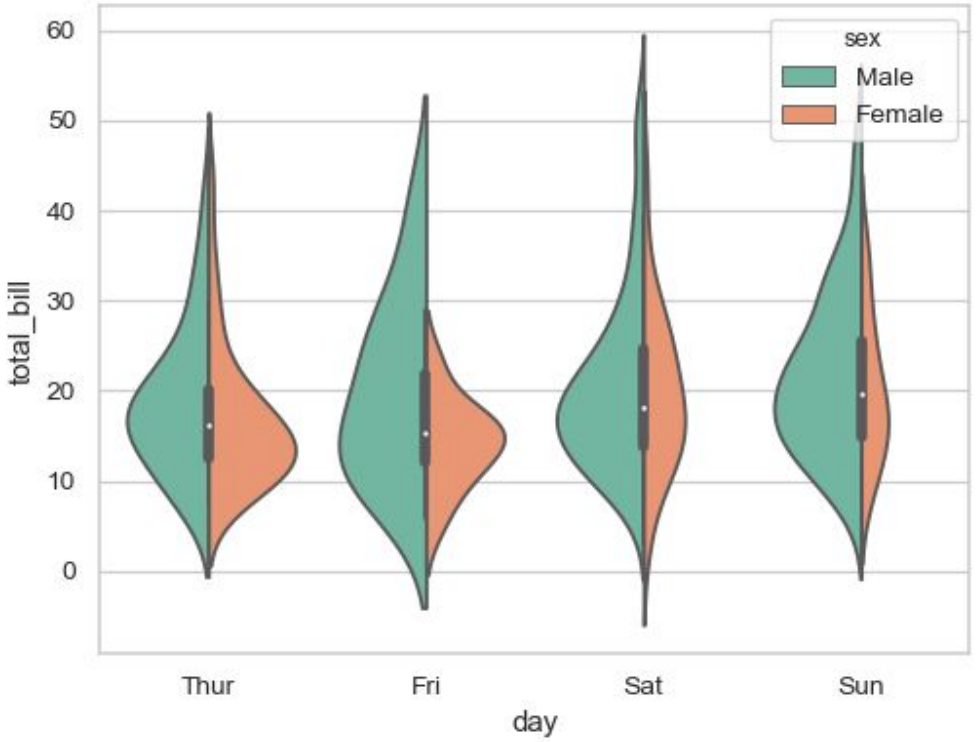


Sources:

"Quantification of intrinsic subtype ambiguity in Luminal A breast cancer and its relationship to clinical outcomes" by Kumar, Zhao, Bhaumik, Sethi, Gann, in BMC Cancer 2019

Violin plot

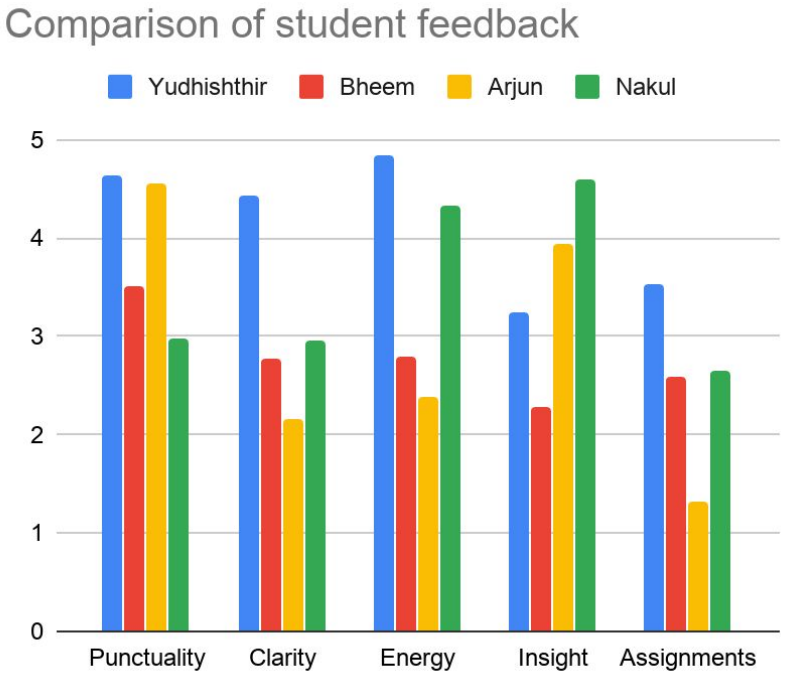
To show several distributions together

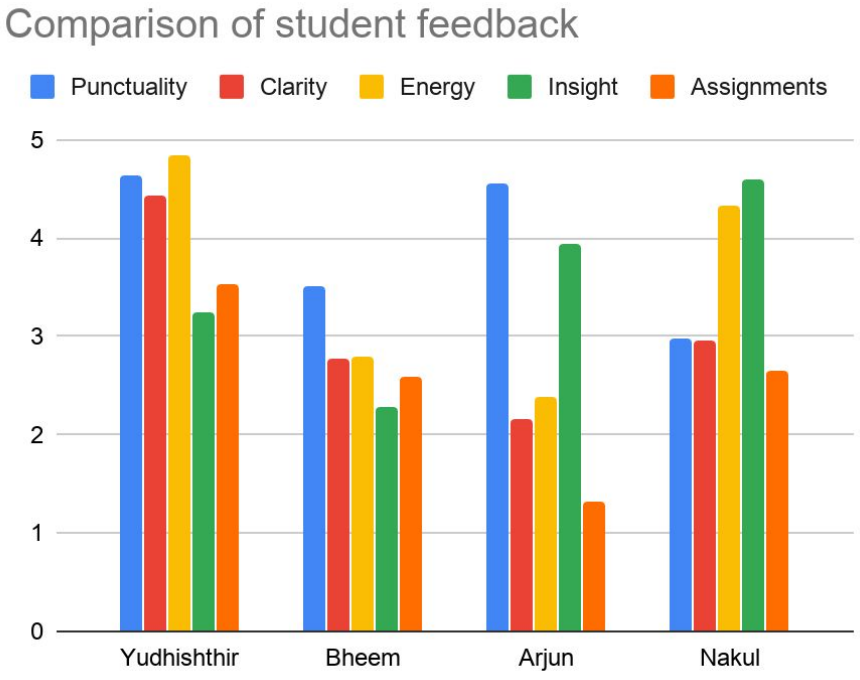


Source: https://seaborn.pydata.org/generated/seaborn.violinplot.html

Radar or spider chart

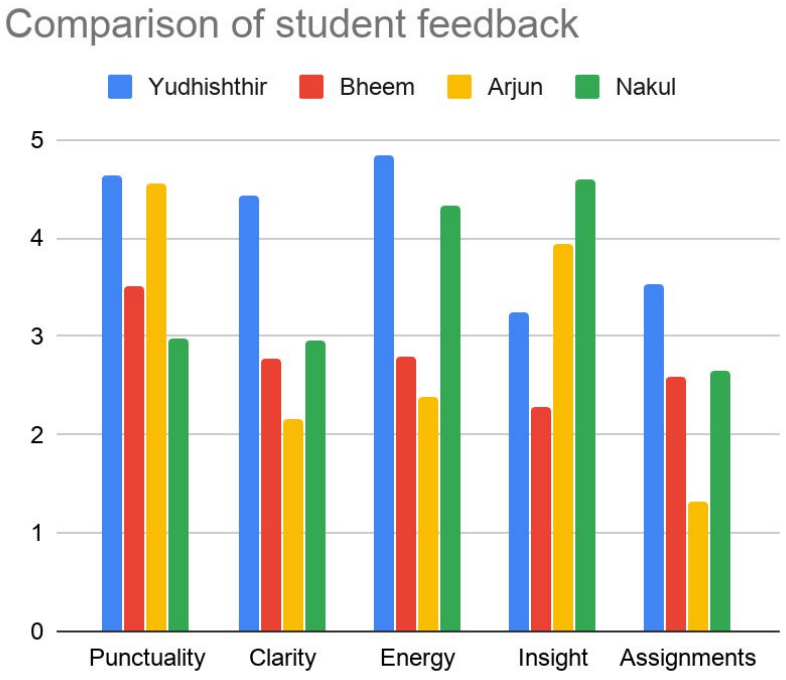
To compare several metrics together instead of having many bars

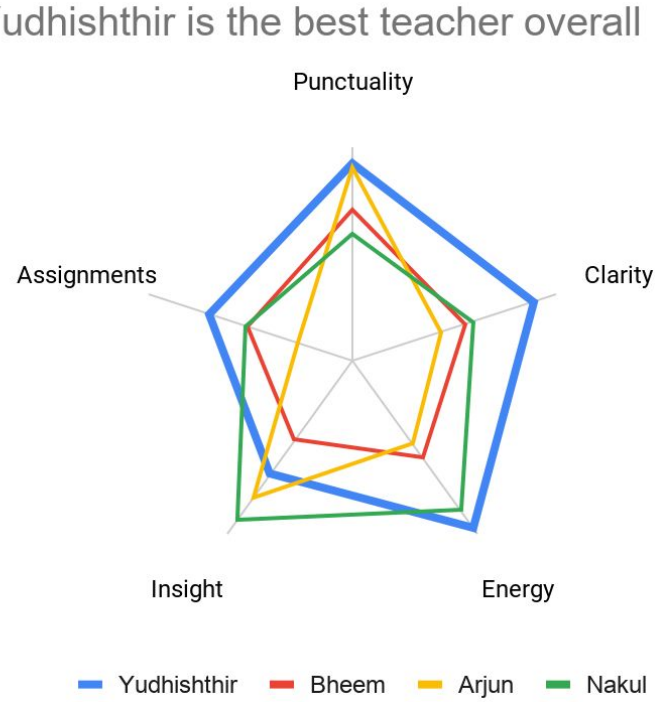




Radar or spider chart

To compare several metrics together instead of having many bars





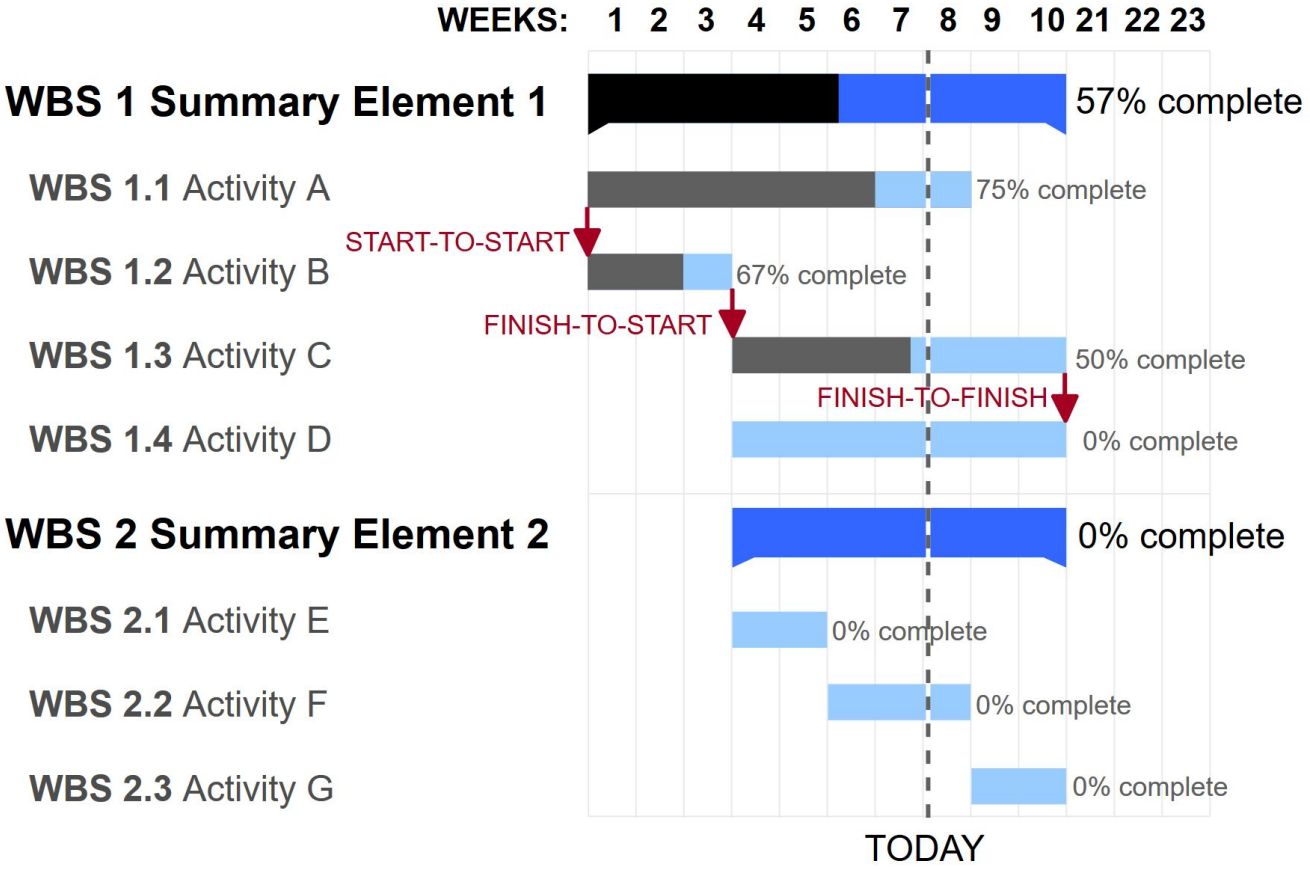
Waterfall chart

To show parts of a whole

Source: https://en.wikipedia.org/wiki/Waterfall_chart

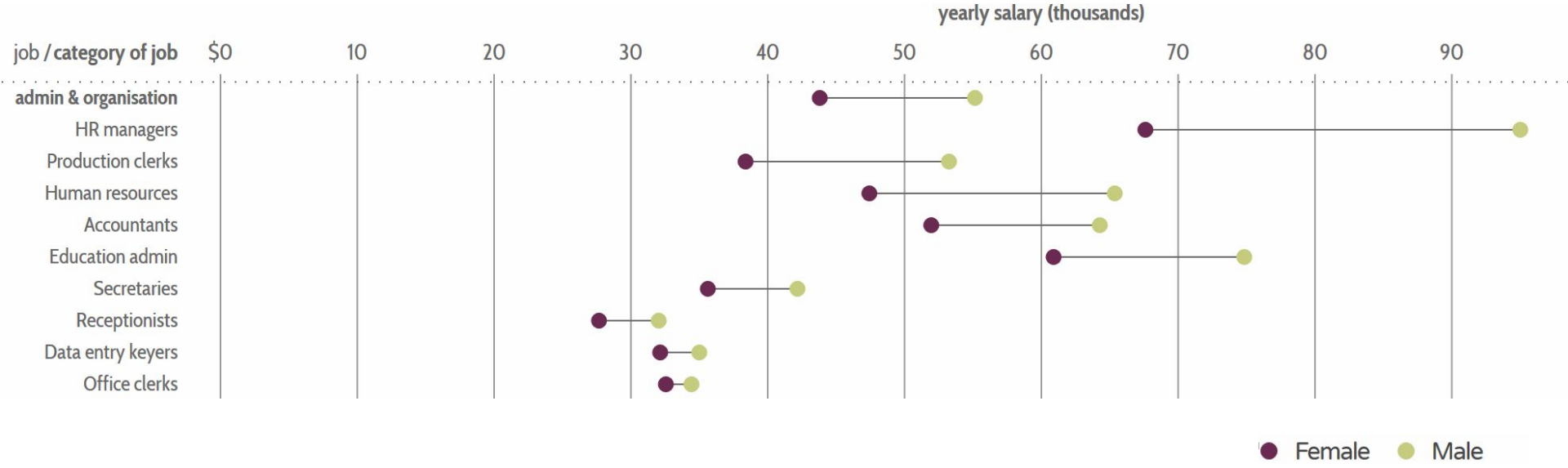
Gantt chart

To show timelines



Source: https://en.wikipedia.org/wiki/Gantt_chart

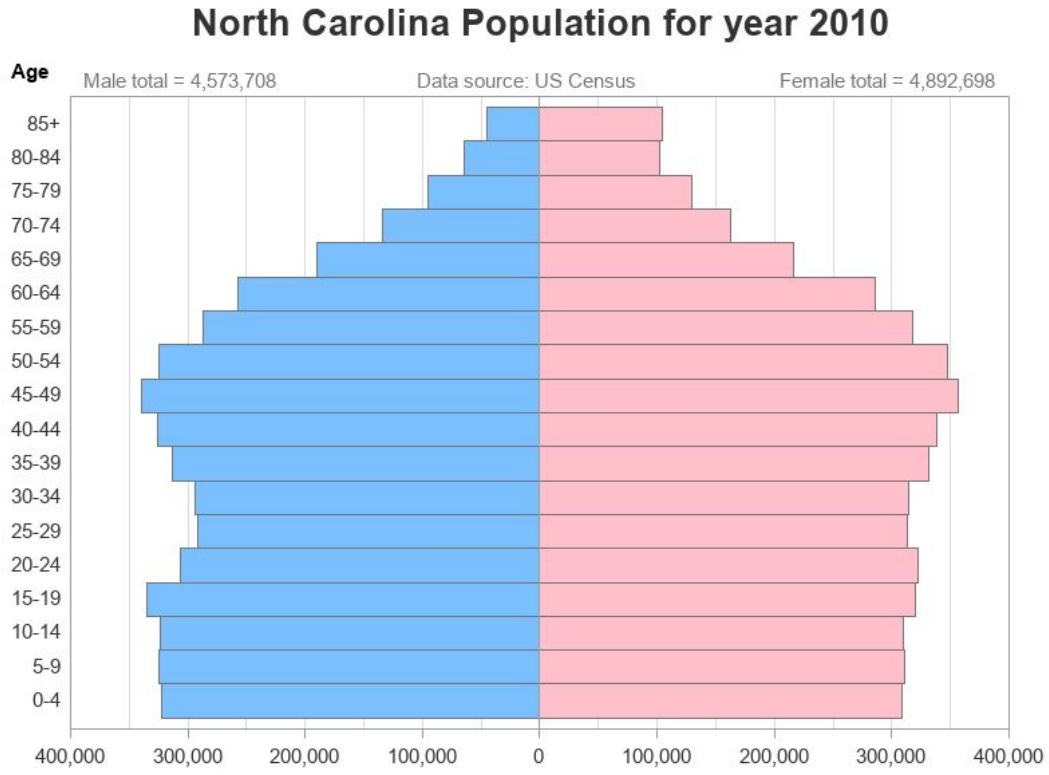
Pair-wise comparison



Source: https://informationisbeautiful.net/visualizations/gender-pay-gap/

Dual histogram

To compare male and female age histograms, for example

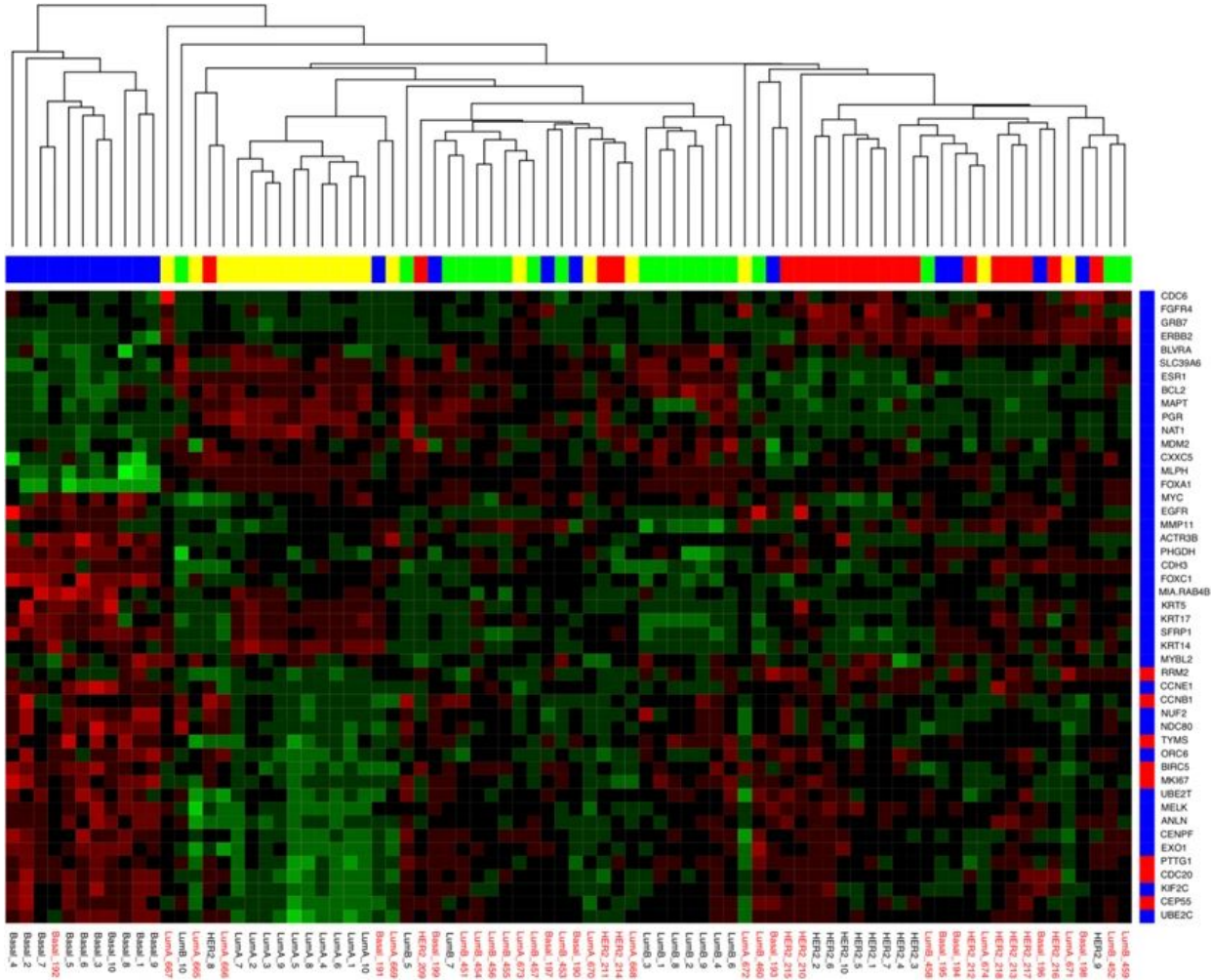


Source:

https://blogs.sas.com/content/graphicallyspeaking/files/2019/02/nc_population_age_gender_almost_final.png

Heatmaps

To show a matrix of values

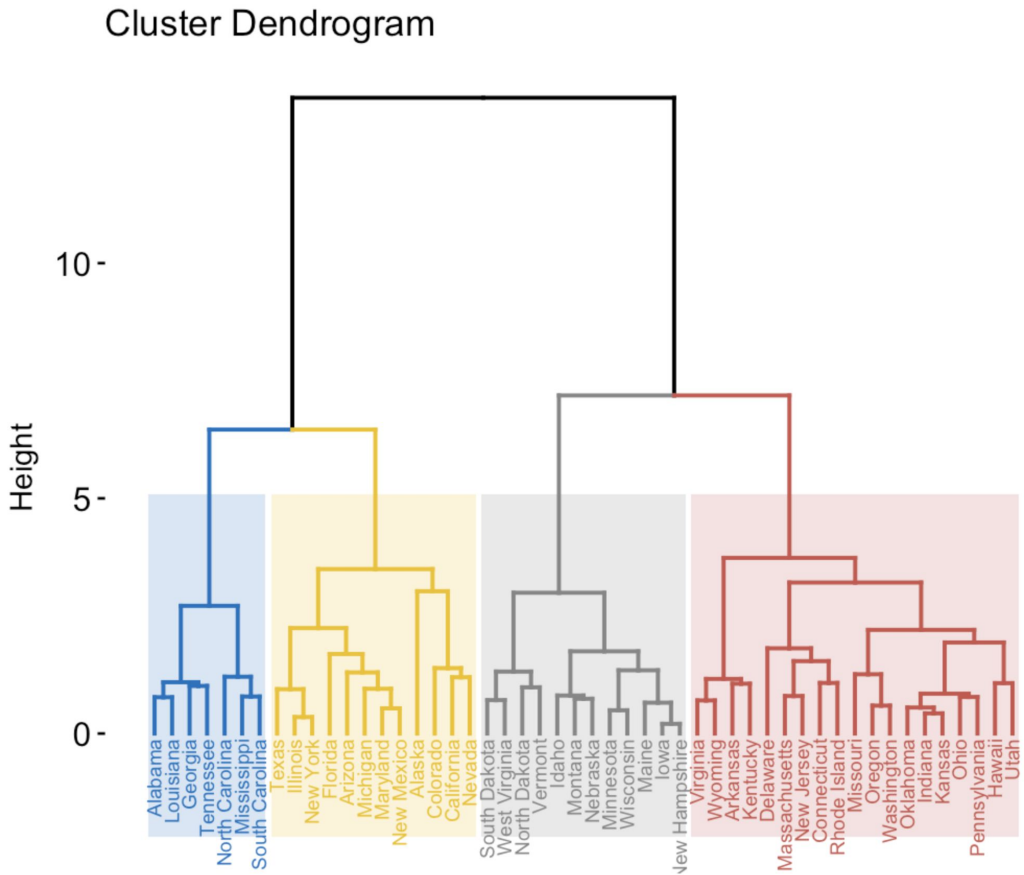


Sources:

"Quantification of intrinsic subtype ambiguity in Luminal A breast cancer and its relationship to clinical outcomes" by Kumar, Zhao, Bhaumik, Sethi, Gann, in BMC Cancer 2019

Dendrograms

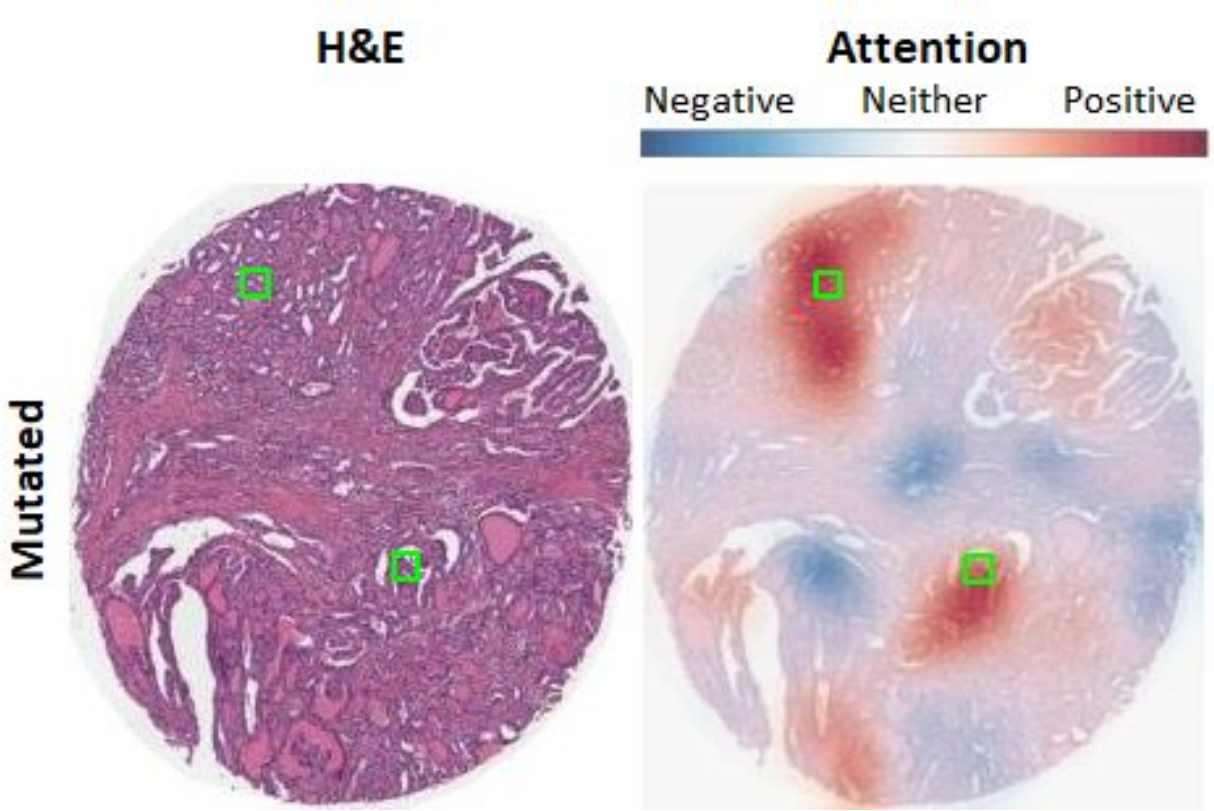
To show hierarchical clustering



Sources:
https://datavizproject.com/data-type/dendrogram/

Image overlays

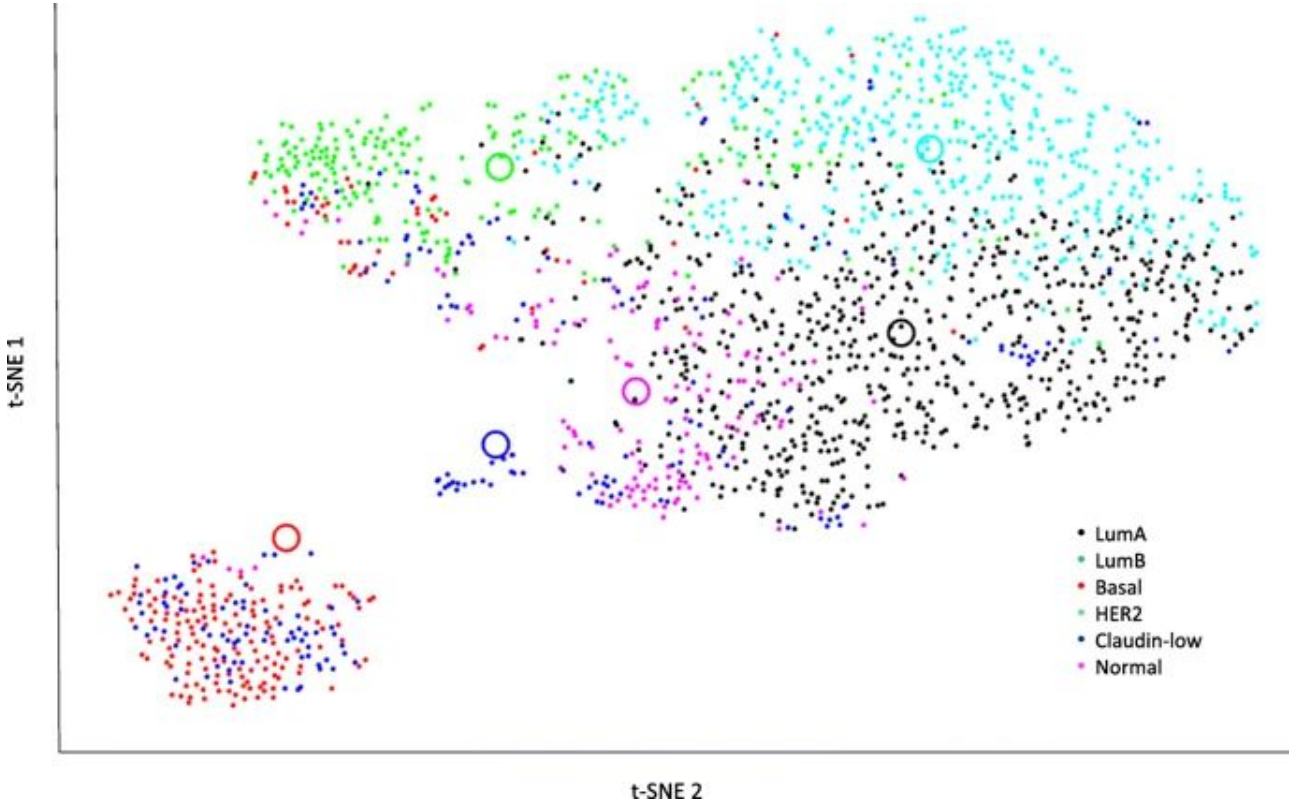
To highlight image analysis results



Source: Unpublished work by Kumar Yashashwi, Deepak Anand, Neeraj Kumar, Swapnil Rane, and Amit Sethi

Dimension reduction tools for high-dimensional data

Such as T-SNE



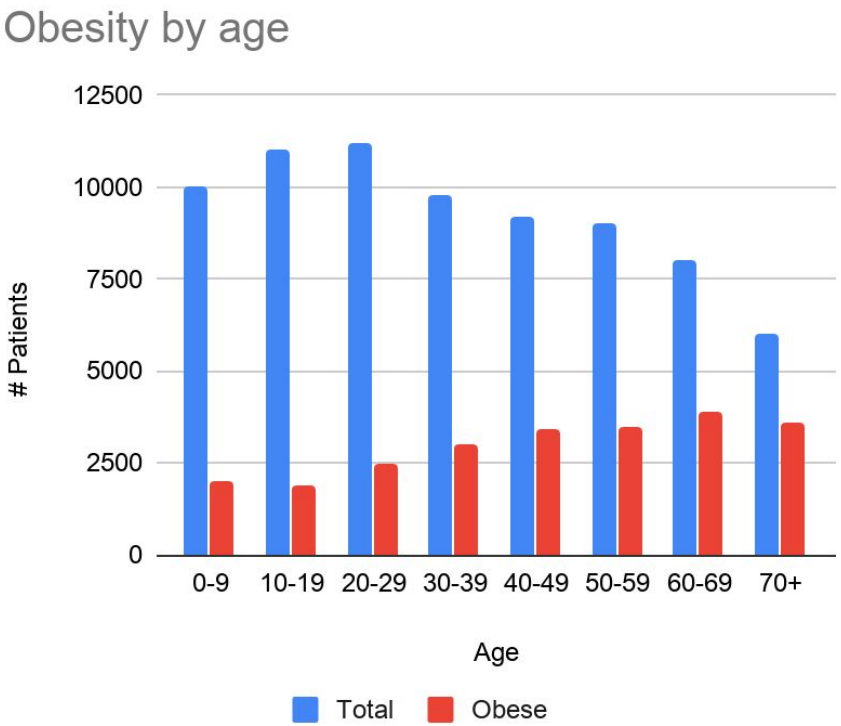
Sources:

Maarten and Hinton https://lvdmaaten.github.io/tsne
"Quantification of intrinsic subtype ambiguity in Luminal A breast cancer and its relationship to clinical outcomes" by Kumar, Zhao, Bhaumik, Sethi, Gann, in BMC Cancer 2019

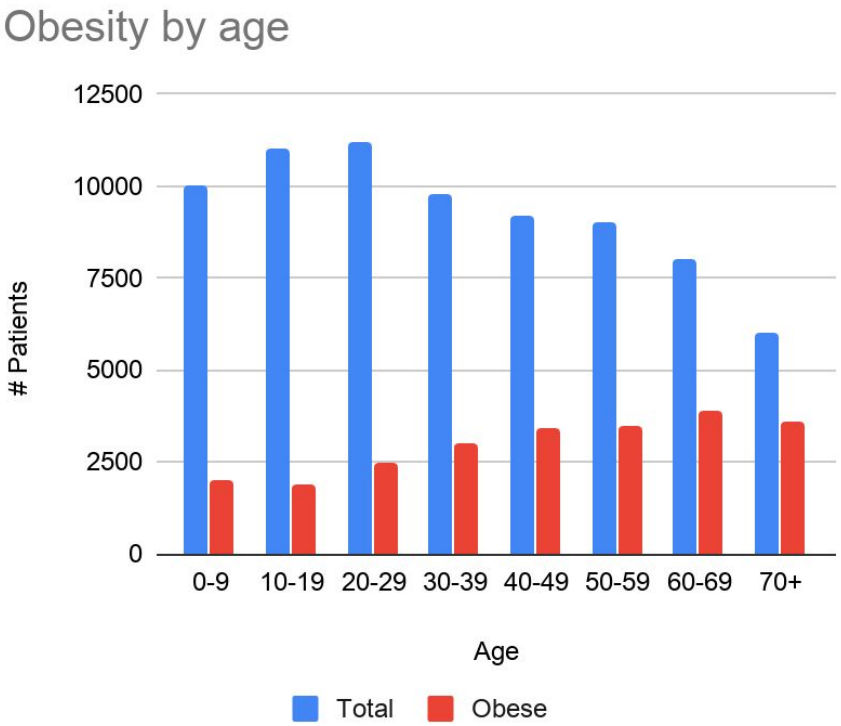
Data transformations

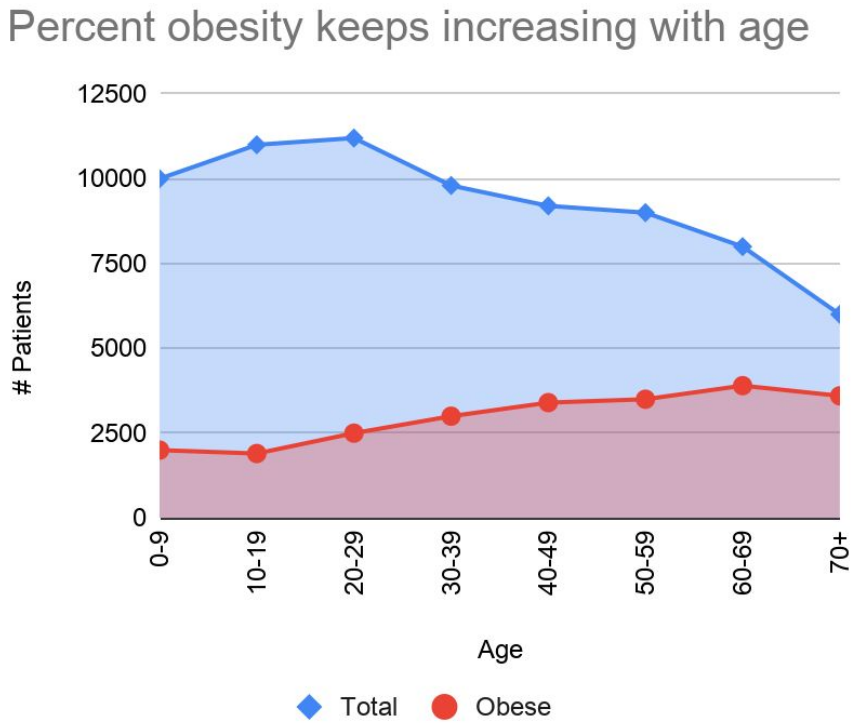
- When to use log-scale?
  - When the rise is exponential
  - When the order of magnitude is more important
- When to take ratio of to two variables?
  - When the variation within the variable is too great
- When to use polar form
  - Part of a whole (pie, doughnut)
  - Many variables (radar)
  - Cyclical variable (hour of day, day of week, month of year, seasons)

How will you improve this chart?

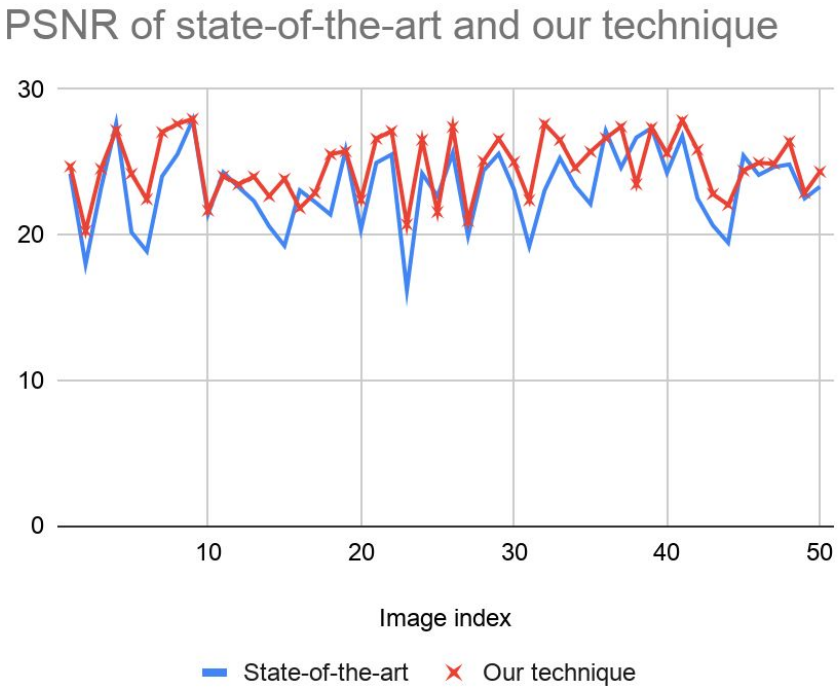


How will you improve this chart?

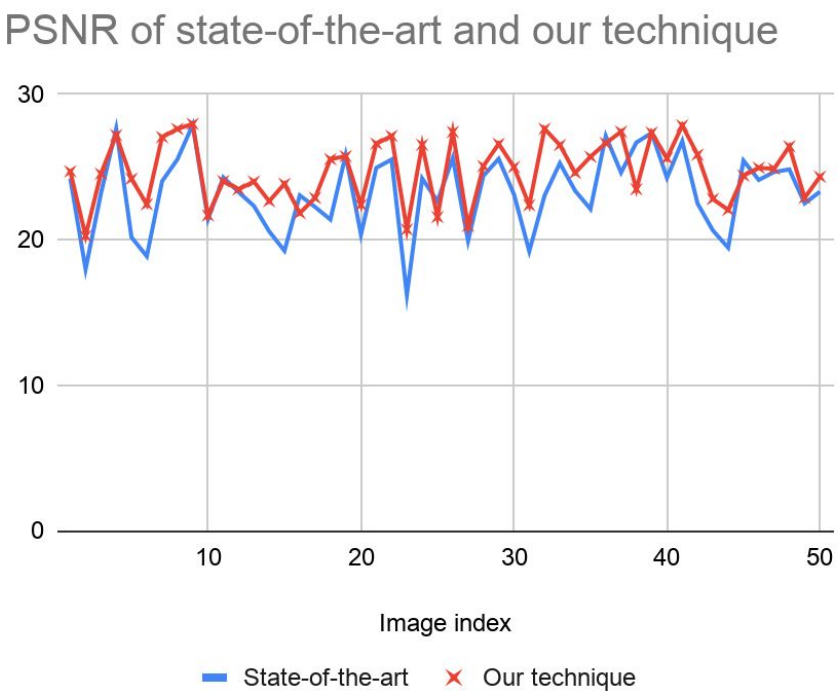


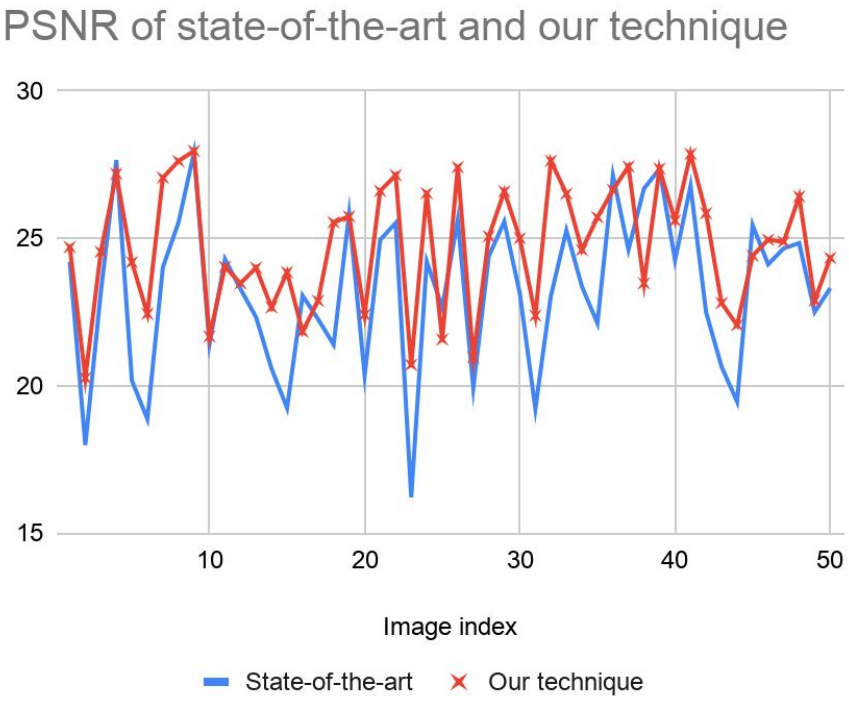


How will you improve this chart?



How will you improve this chart?





How will you improve this chart?