ISA 401: Business Intelligence & Data Visualization

17: Charts Used for Comparisons, Relationships, Distributions and Correlations

Fadel M. Megahed, PhD

Associate Professor
Department of Information Systems and Analytics
Farmer School of Business
Miami University

Twitter: FadelMegahed
GitHub: fmegahed
Email: fmegahed@miamioh.edu
Office Hours: Automated Scheduler for Virtual Office Hours

Refresher: Important Data Viz Concepts

Activity | Solution Notes

> *Over the next several minutes, you will assess/demonstrate your knowledge of the covered material pertaining to the fundamentals of data visualization through 7-questions on Kahoot. To answer the questions, please do the following:*

- Go to www.kahoot.it
- Insert the game pin shared on the screen
- Add your first and last name when asked for a name
- To get the most points per question, answer each question **correctly** and **quickly**
- Winner is the person with the most points after seven questions

Refresher: Important Data Viz Concepts

Activity | Solution Notes

As we go over each question, I will highlight the correct answer and the logic associated with why such an answer was chosen. Please feel free to use the area below to take notes.

- Q1: _
- Q2: _
- Q3: _
- Q4: _
- Q5: _
- Q6: _
- Q7: _

Learning Objectives for Today's Class

- Identify strengths & weaknesses of basic charts
- Use appropriate charts based on objective
- Avoid using pie charts (never use pie charts)
- Avoid 3D graphs (unless VR changes their utility)

A Catalog of Commonly Used Graph Types

PolicyViz Data Visualization Catalog

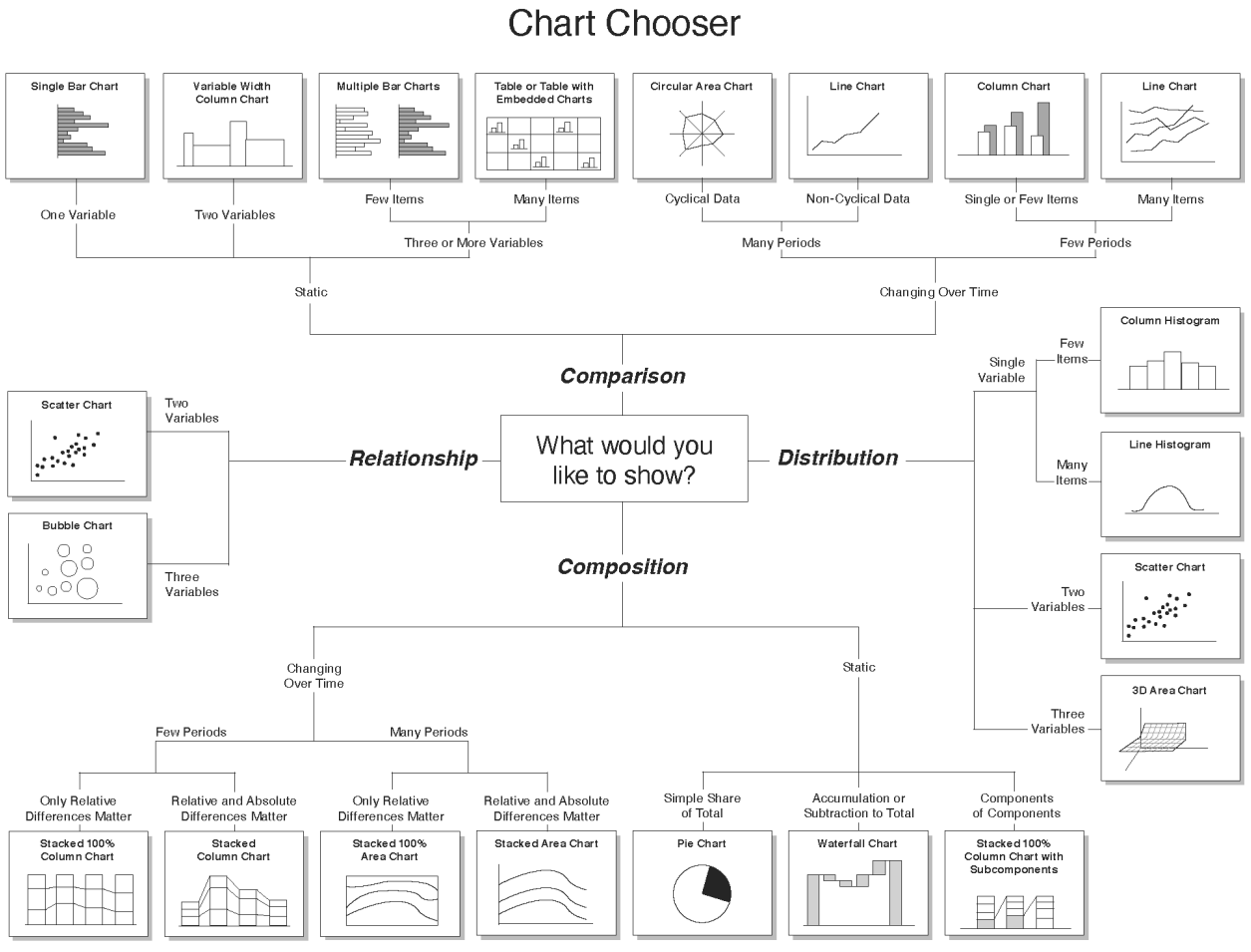
Graph Type | Organization | Author

Graph Type ▲	Description /Notes	Organization	Author	Small Multiples	Year	URL	Image

1 -

Google Data Studio

Chart Suggestions



Source: Chart Chooser, created by Andrew Abela in 2006 and last updated on Sept 06, 2020.

Charts Used for Comparing Data

(Unit of Analysis is Based on a Nominal Categorical Variable)

A Literal Bar Chart

05:00

Activity | Your Solution

Answer the following questions:

(1) How many variables do we have in this graph?

(2) How many observations?

(3) Please discuss the type of variables in the graph? (i.e. nominal, ordinal, etc.)

(4) How is the data encoded in the graph?

(5) Any other comments/observations?

Source: The image is from How Much Does Beer Consumption Vary by Country?, and the data seems to be based on a 2004 report from the Kirin Holdings Company.

A Literal Bar Chart

05:00

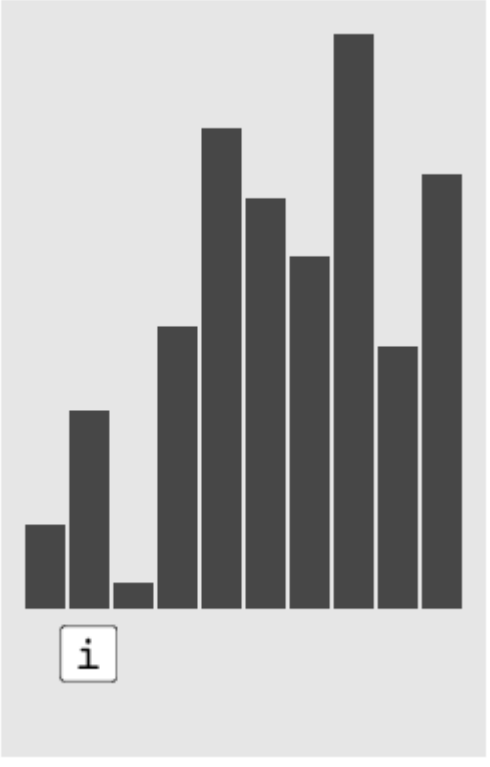
Activity | Your Solution

- Q1: __
- Q2: __
- Q3: __
- Q4: __
- Q5: __

Source: The image is from How Much Does Beer Consumption Vary by Country?, and the data seems to be based on a 2004 report from the Kirin Holdings Company.

Using a Bar Chart to Visualize R Code

```
insert_sort <- function(x) {
  i <- 2
  while(i <= length(x)) {
    j <- i
    while(j > 1 && x[j - 1] > x[j]) {
      j <- j - 1
      x[j + 0:1] <- x[j + 1:0]
    }
    i <- i + 1
  }
  x
}
```



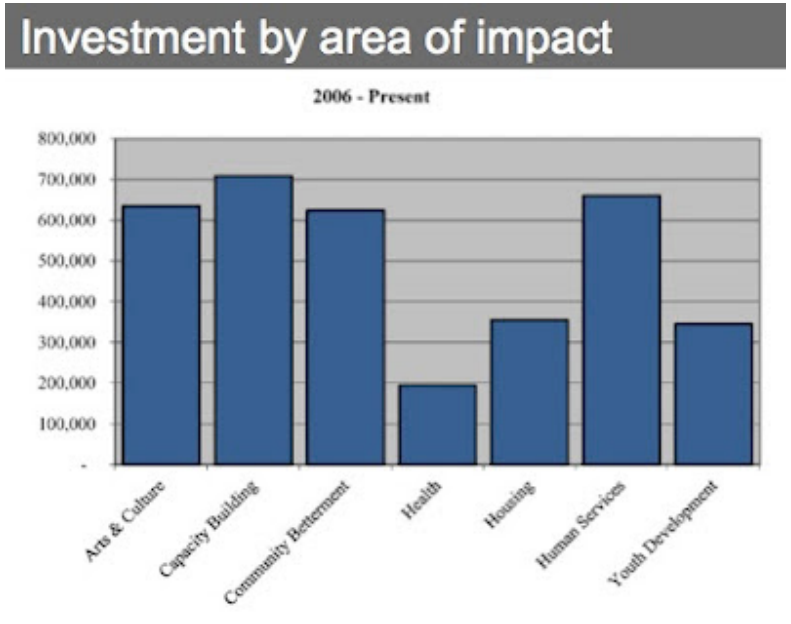
Source: See the amazing watcher package, which allows you to record the state of R function during evaluation. For more details, please click on the link.

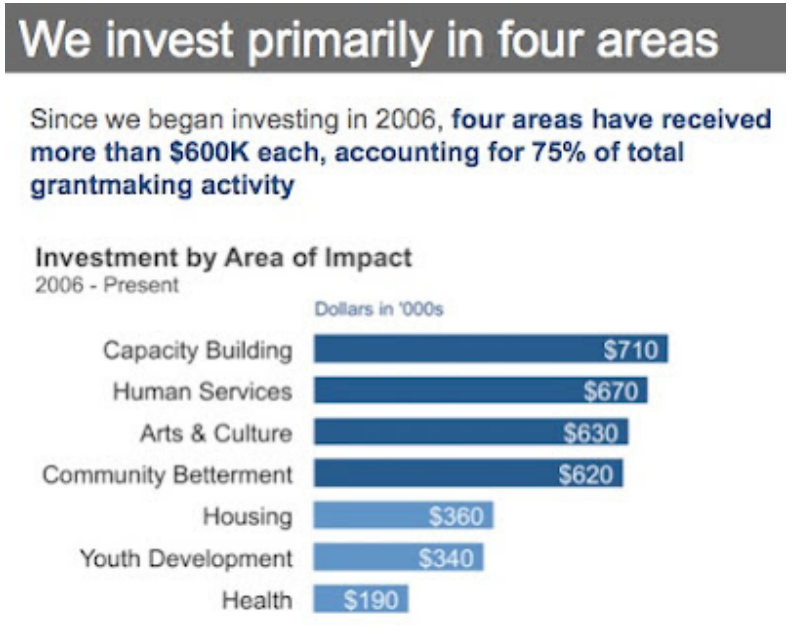
Non-graded activity: Two Bar Charts

05:00

Activity | Your Solution

> *Over the next five minutes, identify **3-4 differences that make the graph on the right better**, and suggest **how you can further improve the graph on the right***





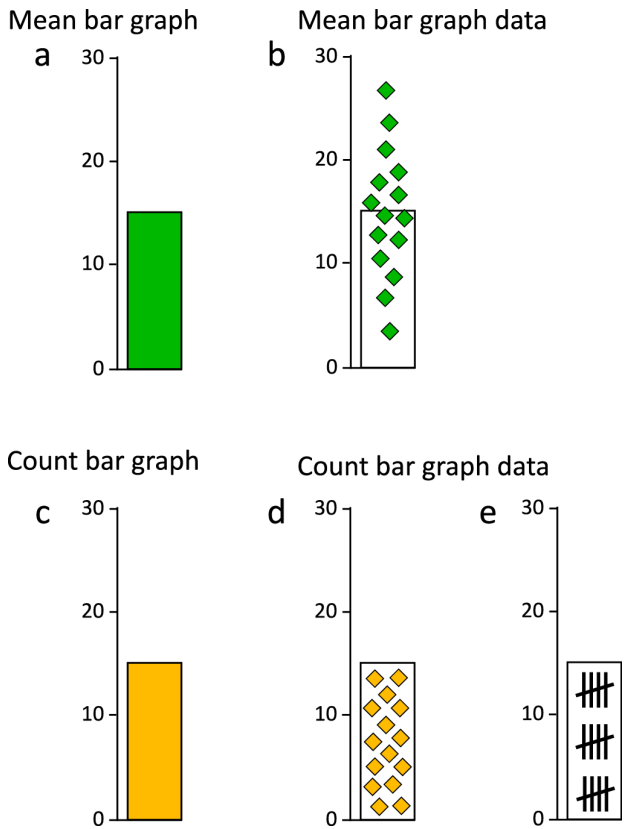
Non-graded activity: Two Bar Charts

05:00

Activity | Your Solution

Insert your differences and suggestions for improvement below.

Issues with the Interpretation of Bar Charts



Data distribution differs categorically between mean and count graphs. (a) Mean bar graphs and (c) count bar graphs do not differ in basic appearance, but they do depict categorically different data distributions.

Source: Kerns and Wilmer (2021). *Journal of Vision*. DOI: https://doi.org/10.1167/jov.21.12.17

Key Takeaway 1

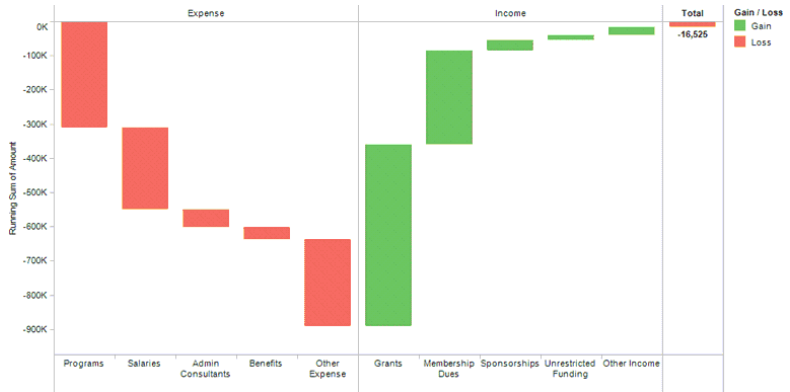
> The typically used **bar** chart should not be to depict means of categorical variables.

Waterfall Charts

05:00

Activity | Your Solution

> *What are the advantages and disadvantages of these two charts? They are using the same exact data. Please try to list 2-4 in each category for each chart.*



Waterfall Charts

05:00

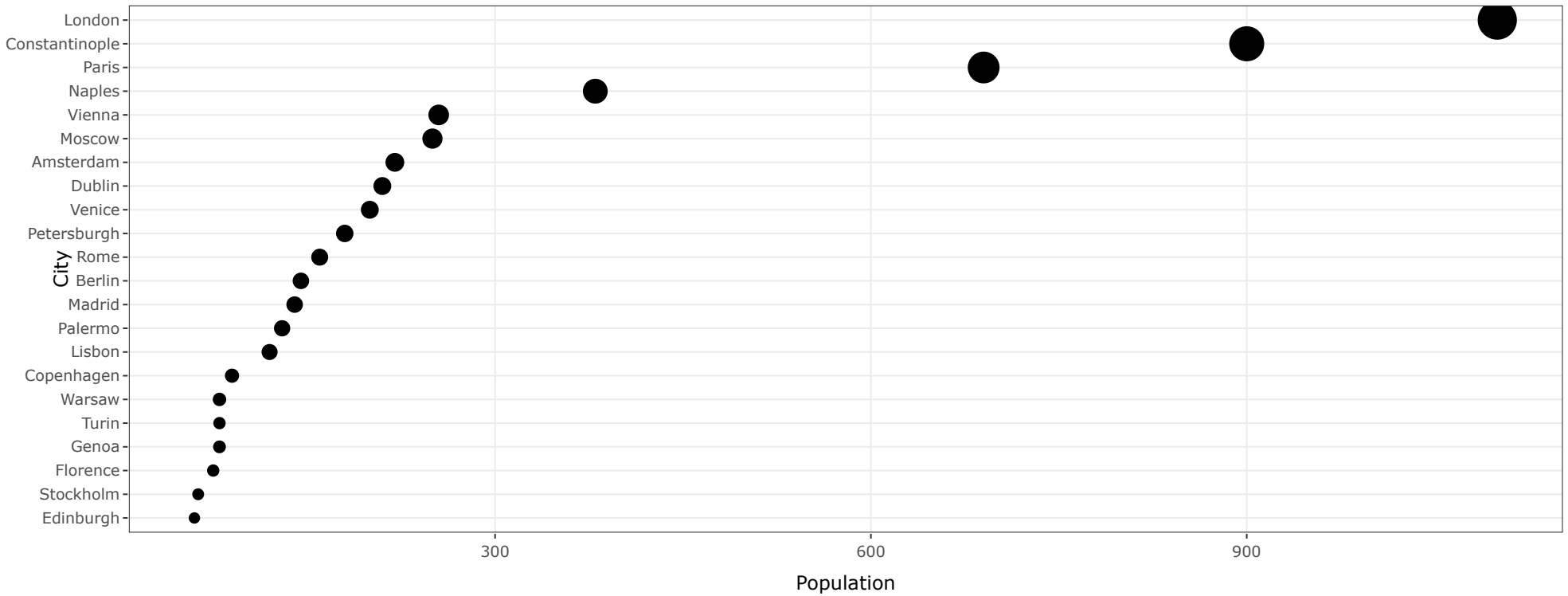
Activity | Your Solution

Insert your advantages and disadvantages below

3D Charts are Awful: Possible Exception

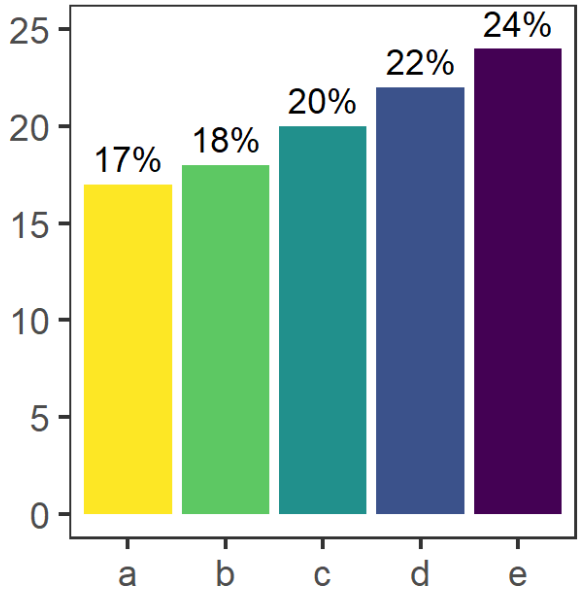
Source: See the interactive version of the chart by clicking here

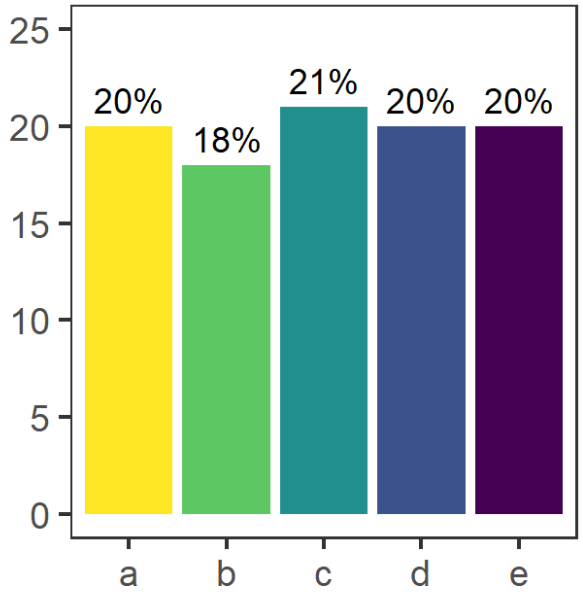
Dot Charts: Recall the Playfair Example

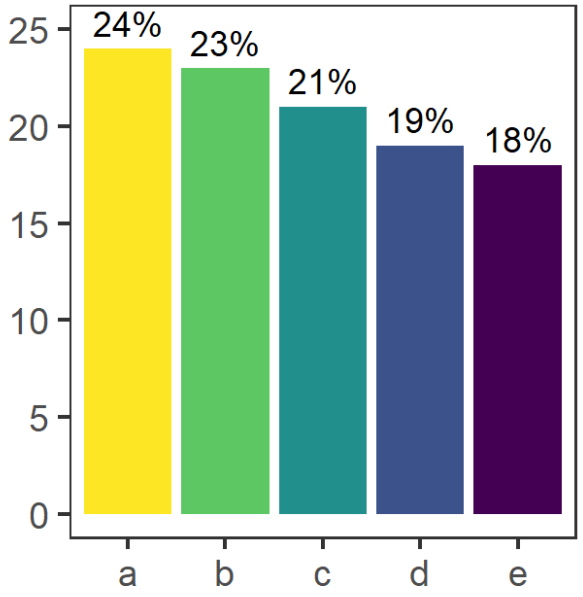


Proportions

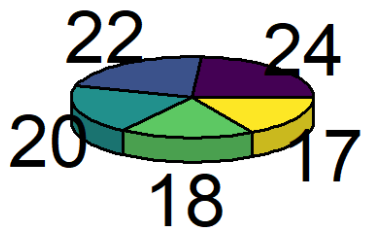
Pie Charts are Awful By Design



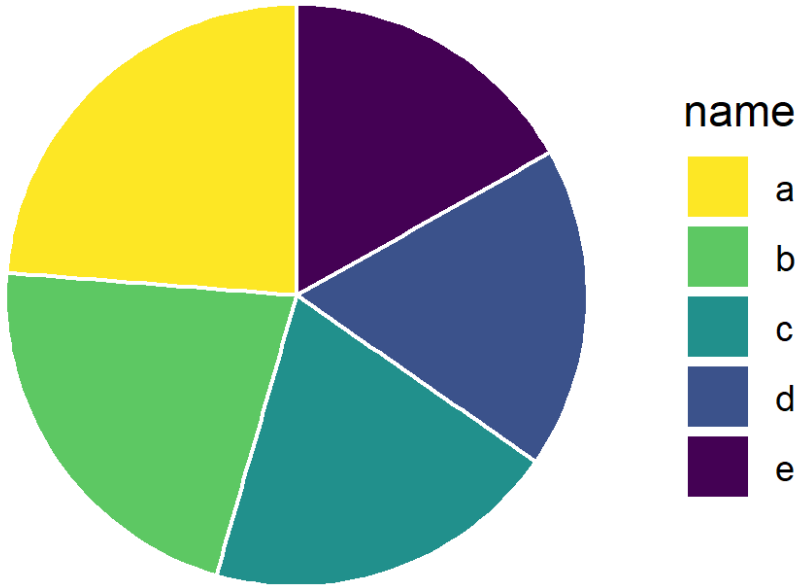




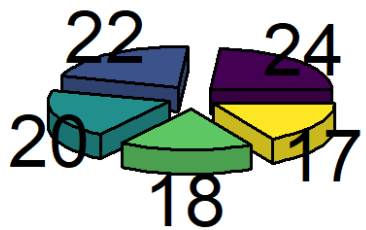
And often made even worse: 3D



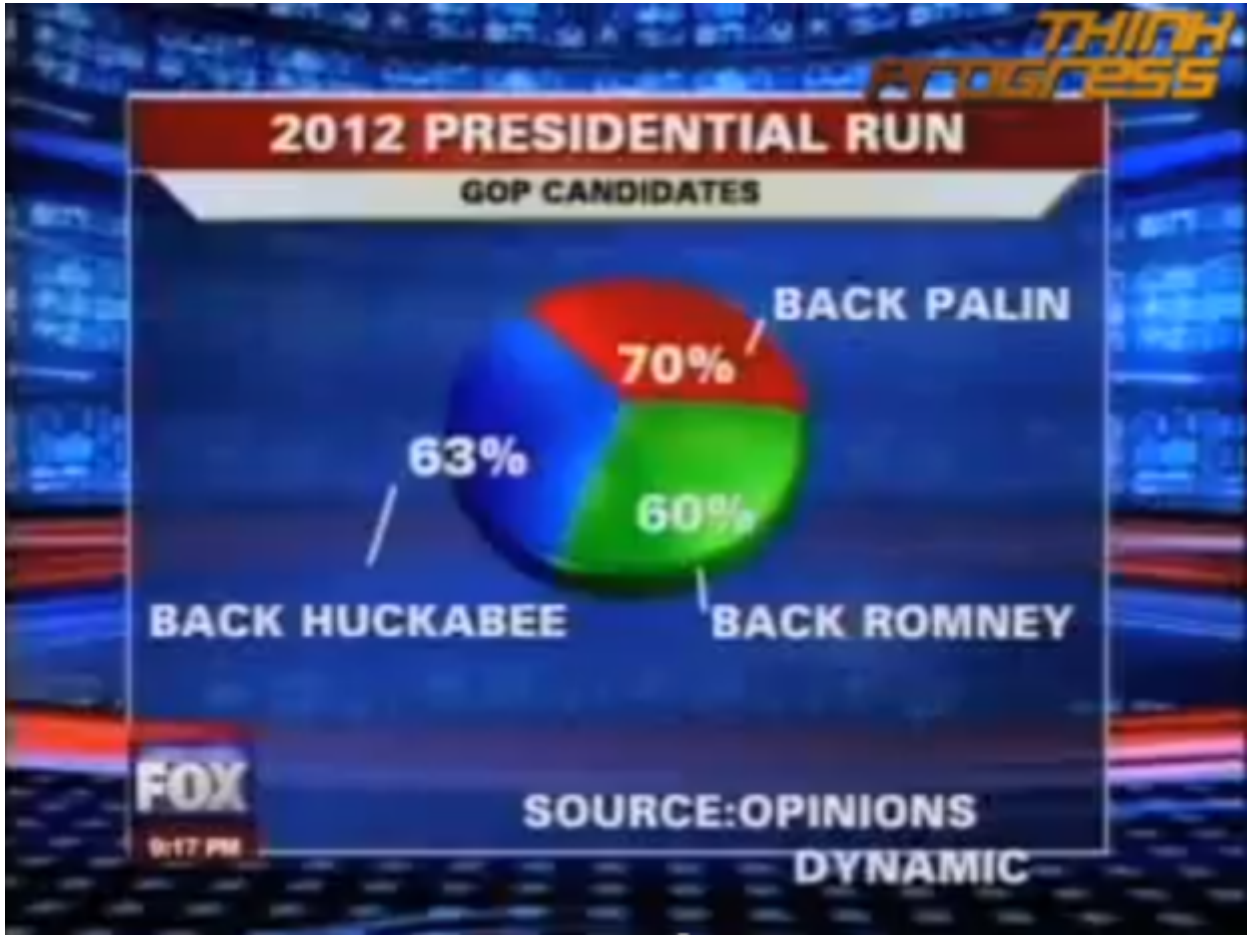
And often made even worse: Side Legend



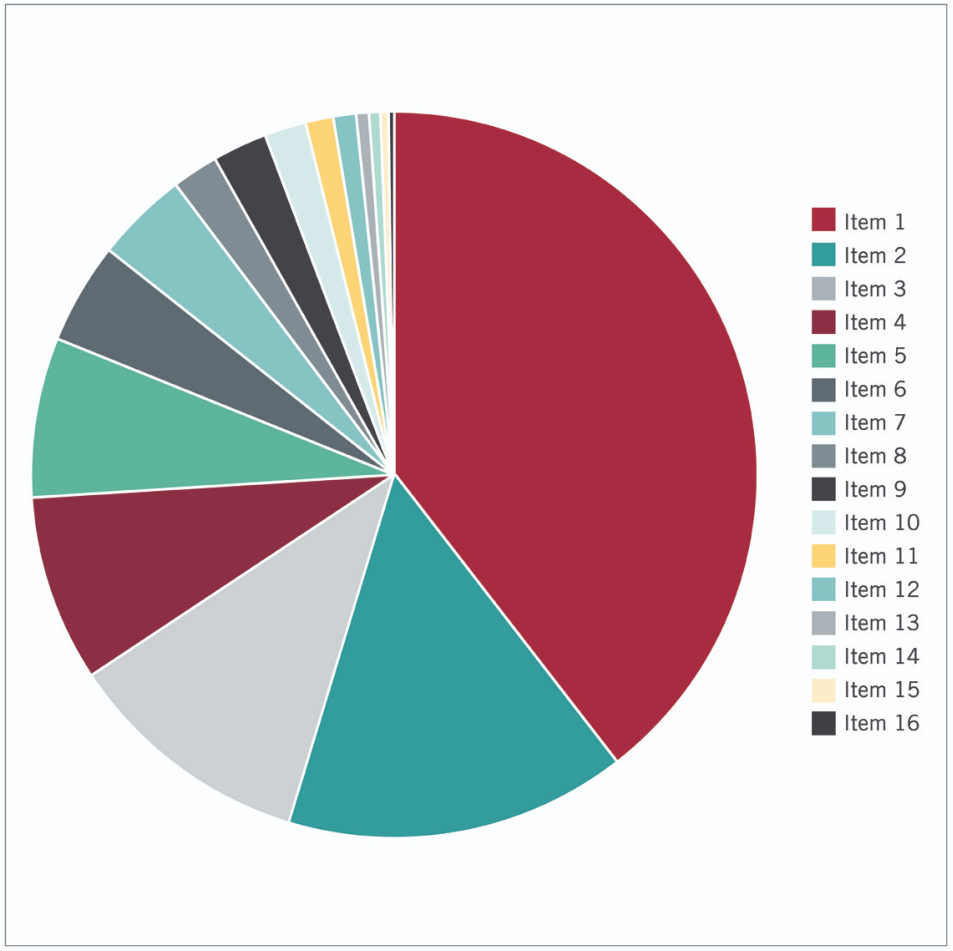
And often made even worse: Exploded Pie



And often made even worse: SUM(%) != 100%



And often made even worse: Many Levels



Key Takeaway 2:

> Please do not use pie charts.

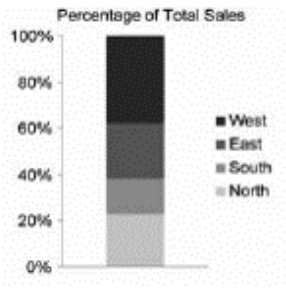
- If you need any further evidence, please check *?pie()* in R. Even statistical software are recommending against using pie charts!!

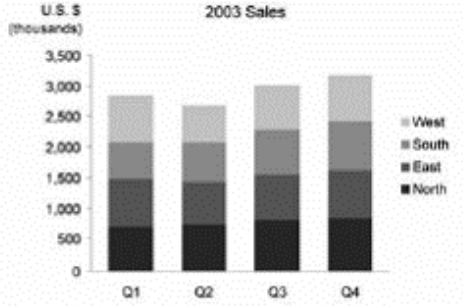
Stacked Bar Charts

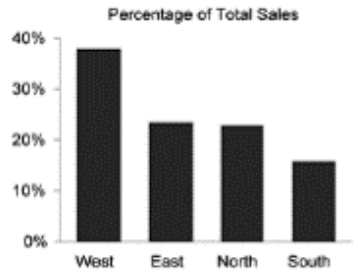
05:00

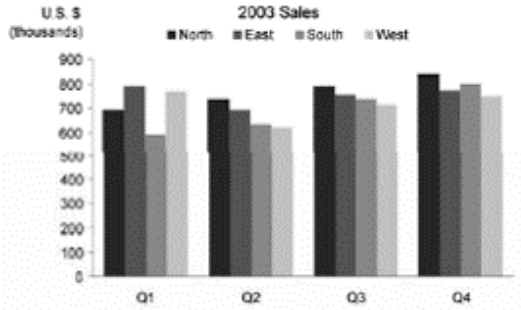
Activity | Your Solution

When it is best to use the four charts below? They are using the same exact data.









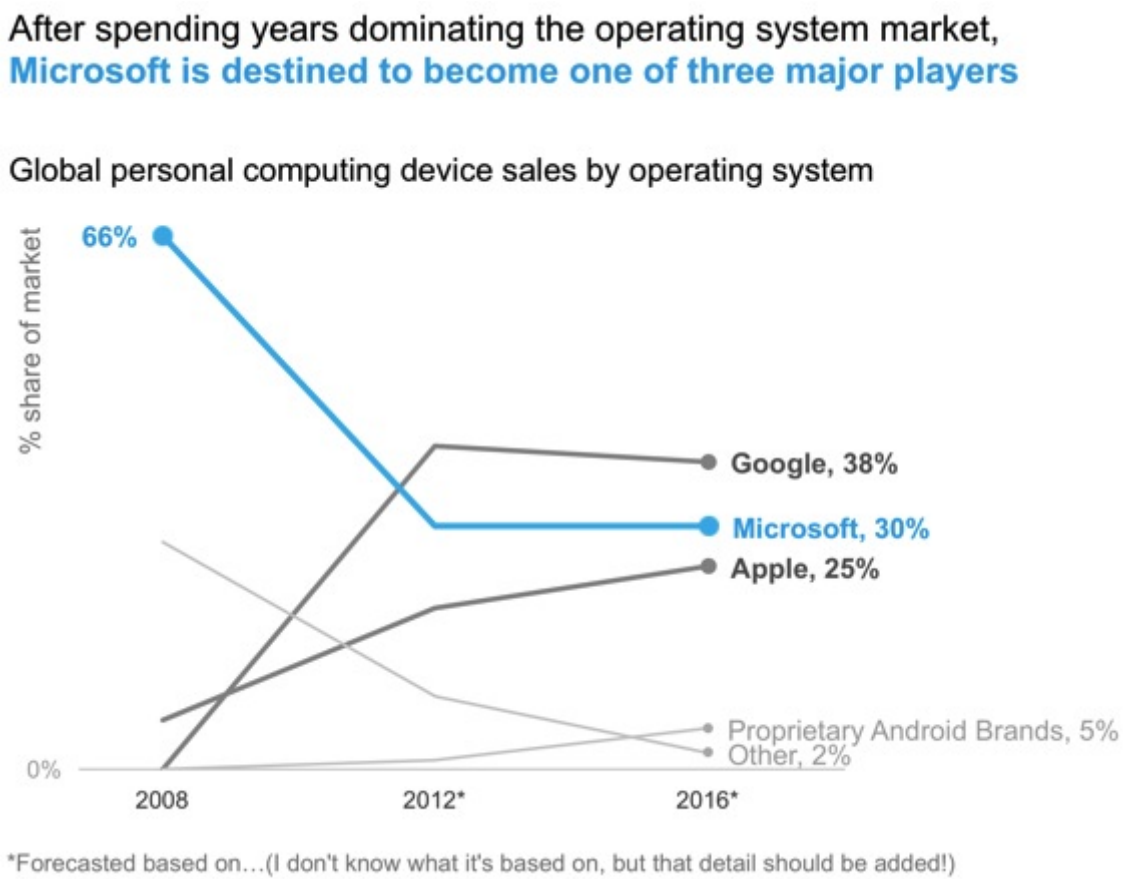
Stacked Bar Charts

05:00

Activity | Your Solution

Insert best usage scenario for each chart below

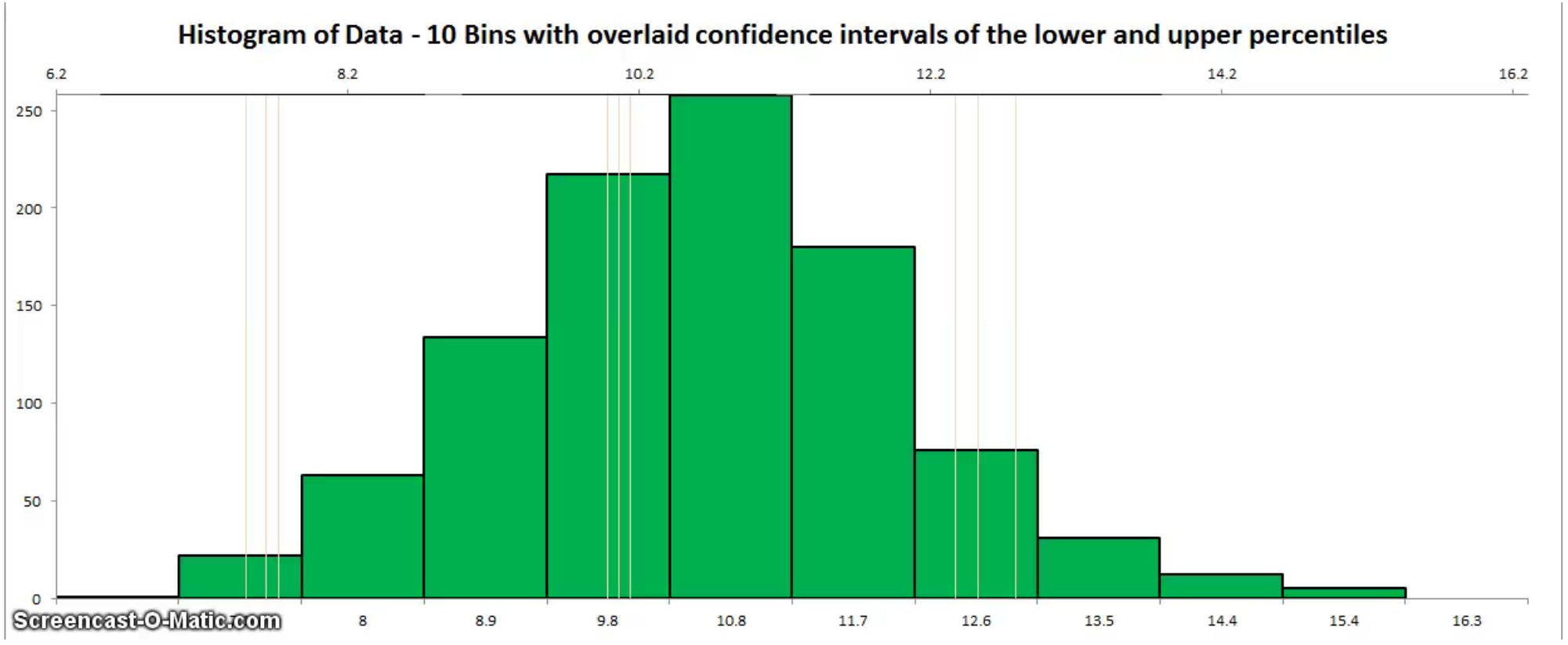
A Note on Stacked Bar Charts



Source: Storytelling with Data: To Stack or Not to Stack

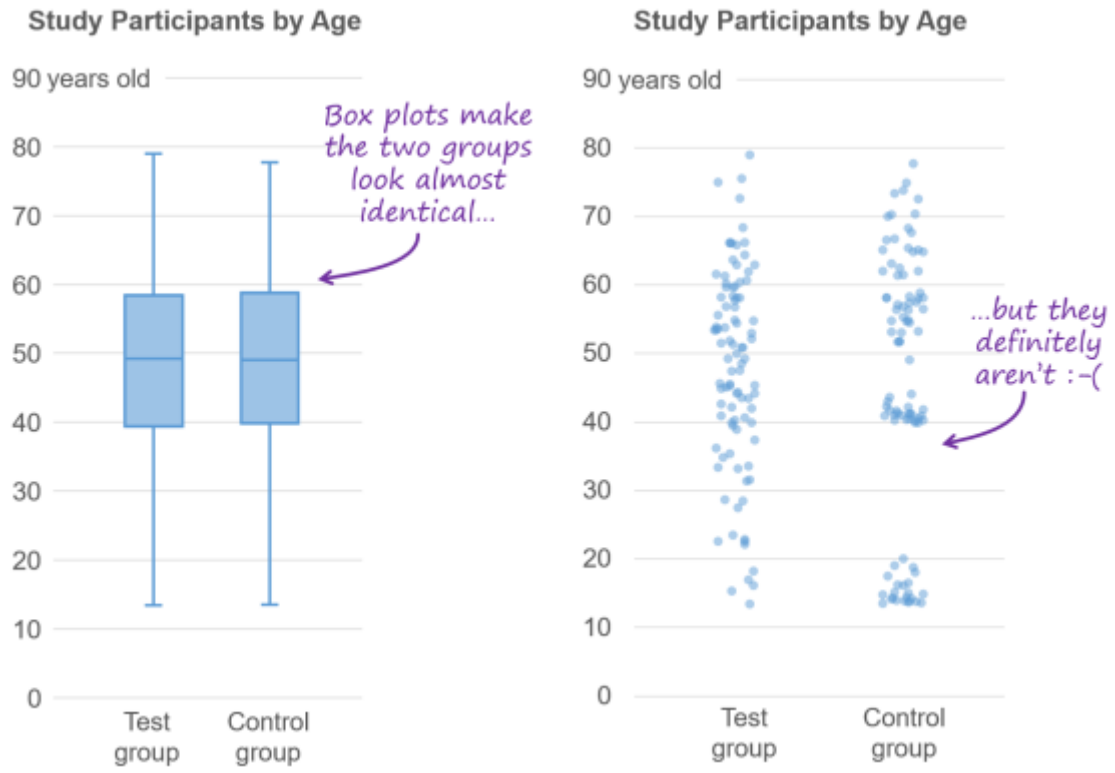
Distributions and Correlations

Issues with Histograms



Source: In a quality improvement project with a major manufacturer of electronic devices, we discovered an issue in the customer's seemingly 'normal' data.

Issues with Box Plots

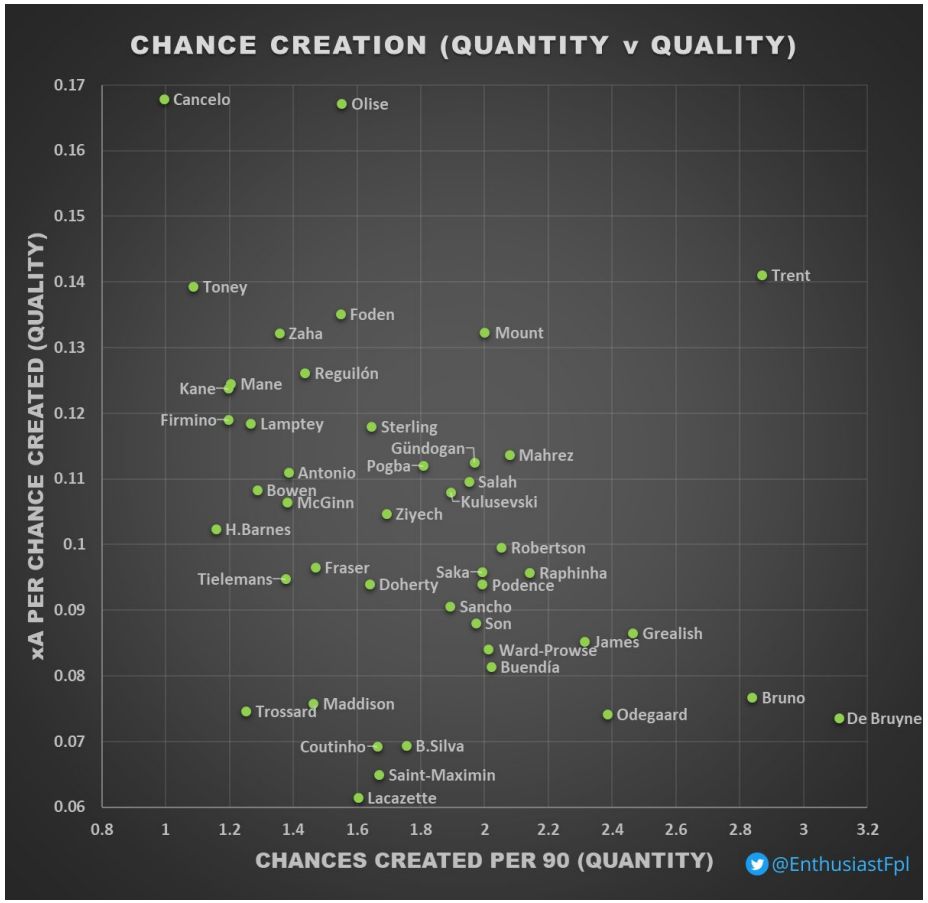


Source: Nick Desbarats (2021) I've Stopped Using Box Plots. Should You?

Key Takeaway 3:

> To capture the variability in a dataset, the use of box plots **may not be** appropriate!!!!

Scatter Plots



Source: Chart created by @EnthusiastFpl and shared on on March 25, 2022.

Recap

Summary of Main Points

- Identify strengths & weaknesses of basic charts
- Use appropriate charts based on objective
- Avoid using pie charts (never use pie charts)
- Avoid 3D graphs (unless VR changes their utility)