



FUNDAMENTALS OF DATA VISUALIZATION USING TABLEAU

LIVE SESSION ON 10TH JANUARY 2021



www.linkedin.com/in/aarthikumar

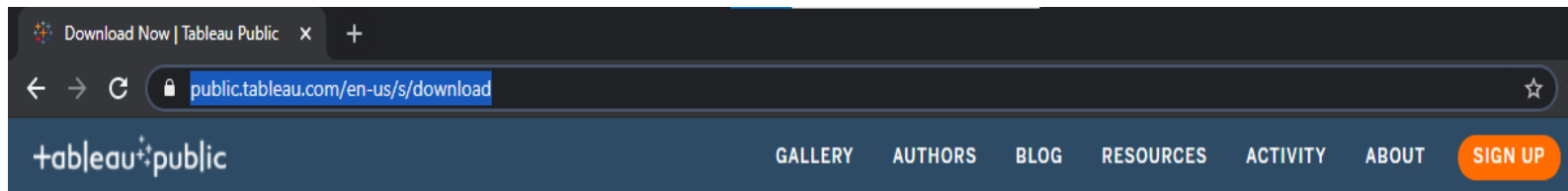
- Accountant + Data Scientist
- 10+ years of experience in DS
- Lead Data Scientist, Allianz Benelux
- Ambassador for Women in Data Science, a Stanford initiative

SESSION AGENDA

- Tableau public installation – ensure everyone has installed
- Why data visualization is important?
- Data visualization – the process
- A quick recap on the basics – Data types and Scale of measurement
- Tableau
 - Introduction to the tool (GUI)
 - Various chart options
 - Dashboard
- Design principles

TABLEAU PUBLIC INSTALLATION

- Go to <https://public.tableau.com/en-us/s/download>
- Enter your email address and start downloading the file (Windows / Mac)



You'll be exploring in minutes

Create interactive graphs, stunning maps, and live dashboards in minutes. Save your viz to your Tableau Public profile, and share it anywhere on the web. Anyone can do it, it's that easy—and it's free.

[DOWNLOAD THE APP](#)

2020.4 Available for Windows and Mac | [System Requirements](#)

WHY DATA VISUALIZATION

The greatest value of a picture is when it forces us to notice what we never expected to see.

—John Tukey



The human brain can process an entire image for as little as 13 milliseconds

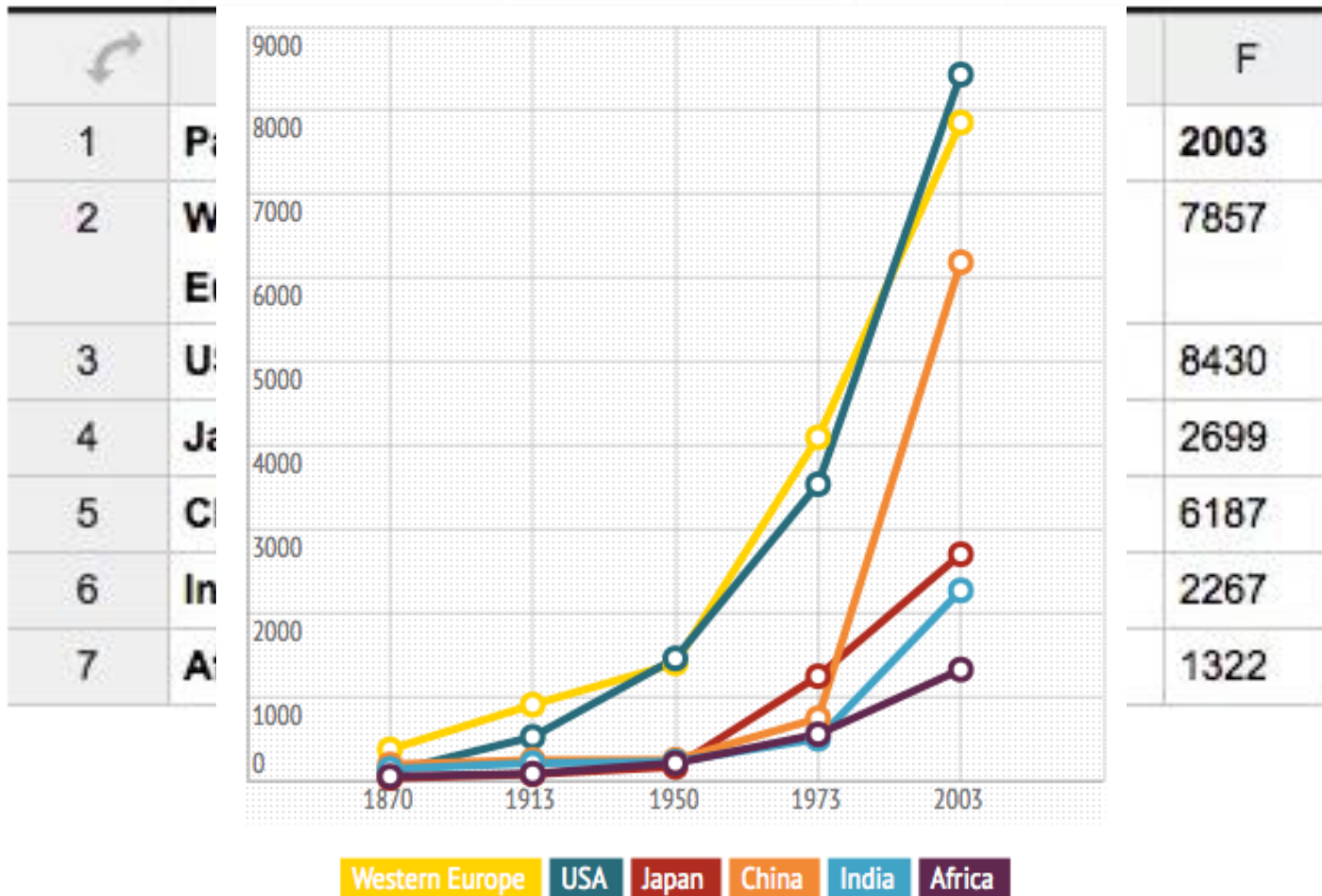


About 80% of the information we take in, is by eye

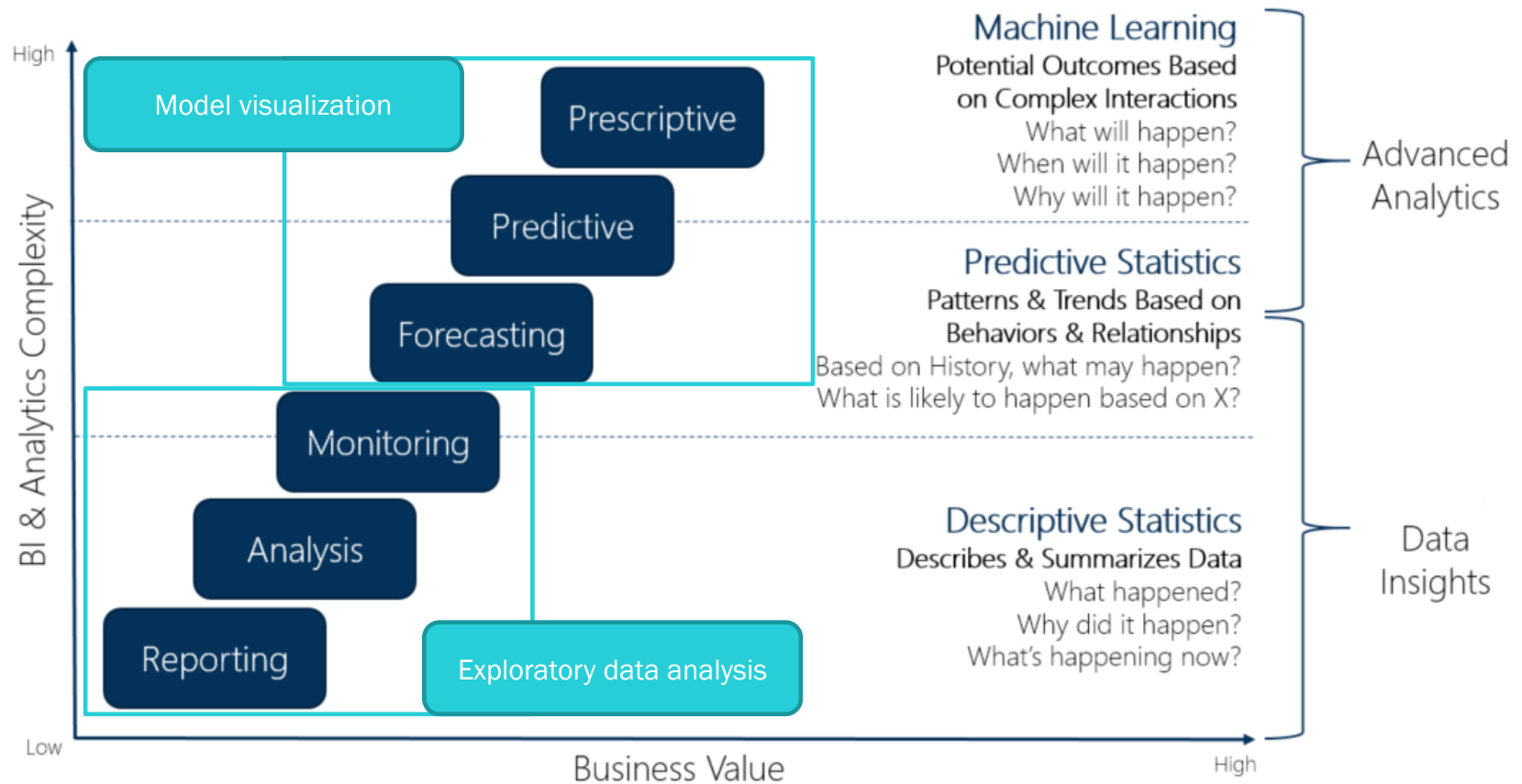


The human brain processes visuals 60,000 times faster than words

WHY DATA VISUALIZATION?



DATA VISUALIZATION IS USED THROUGHOUT THE ANALYTICS MATURITY CURVE



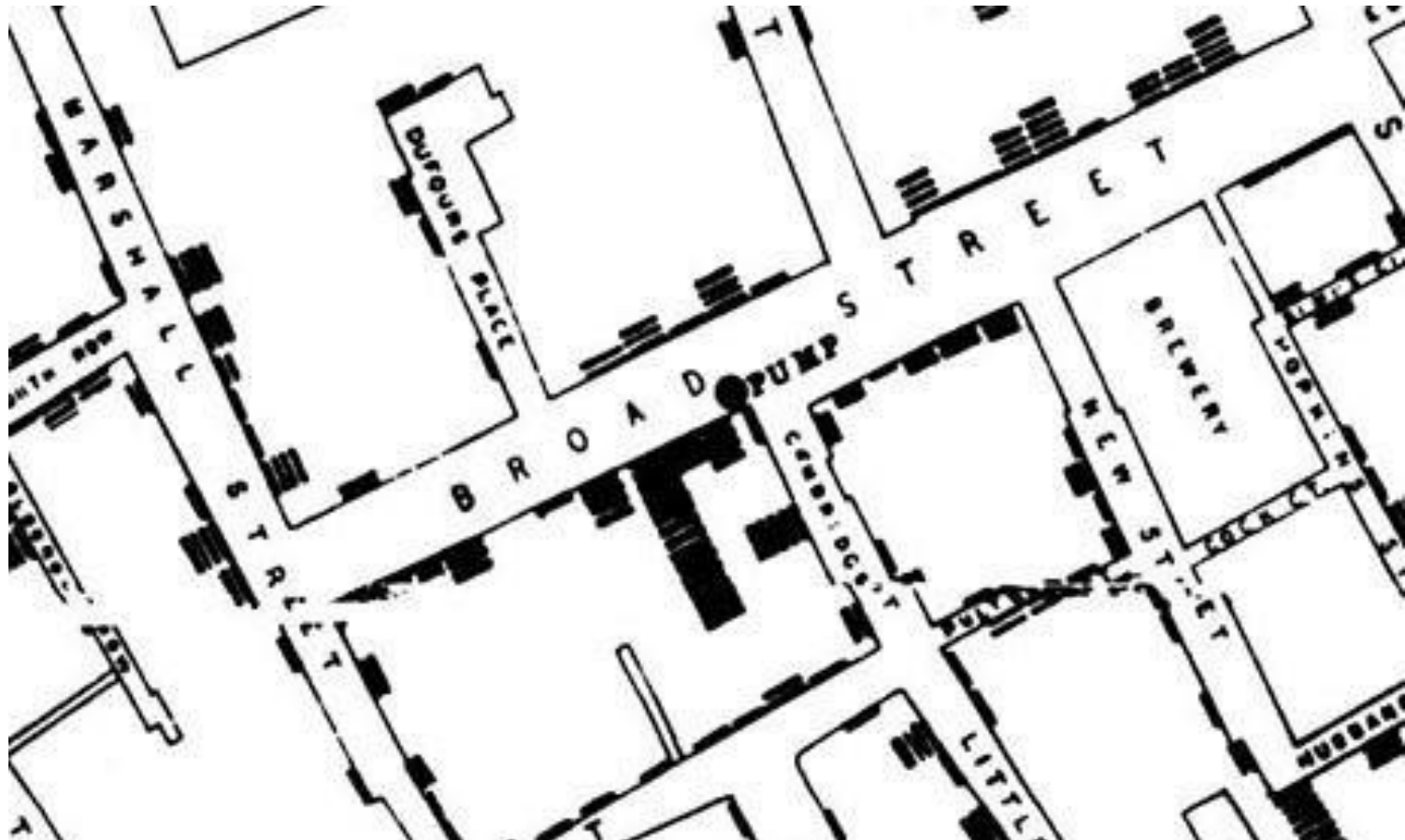


HISTORY OF DATA VISUALIZATION – IT IS NOT A NEW FIELD

Some of the path breaking visualization in the history are made by -

- William Playfair (1759–1823)
- Florence Nightingale (1820–1910)
- John Snow (1813–1858)
- Charles Joseph Minard (1781–1870)
- Edmond Halley (1656–1742)
- Charles de Fourcroy (1766–1824)
- Luigi Perozzo (1856–1916)

JOHN SNOW (1813–1858)

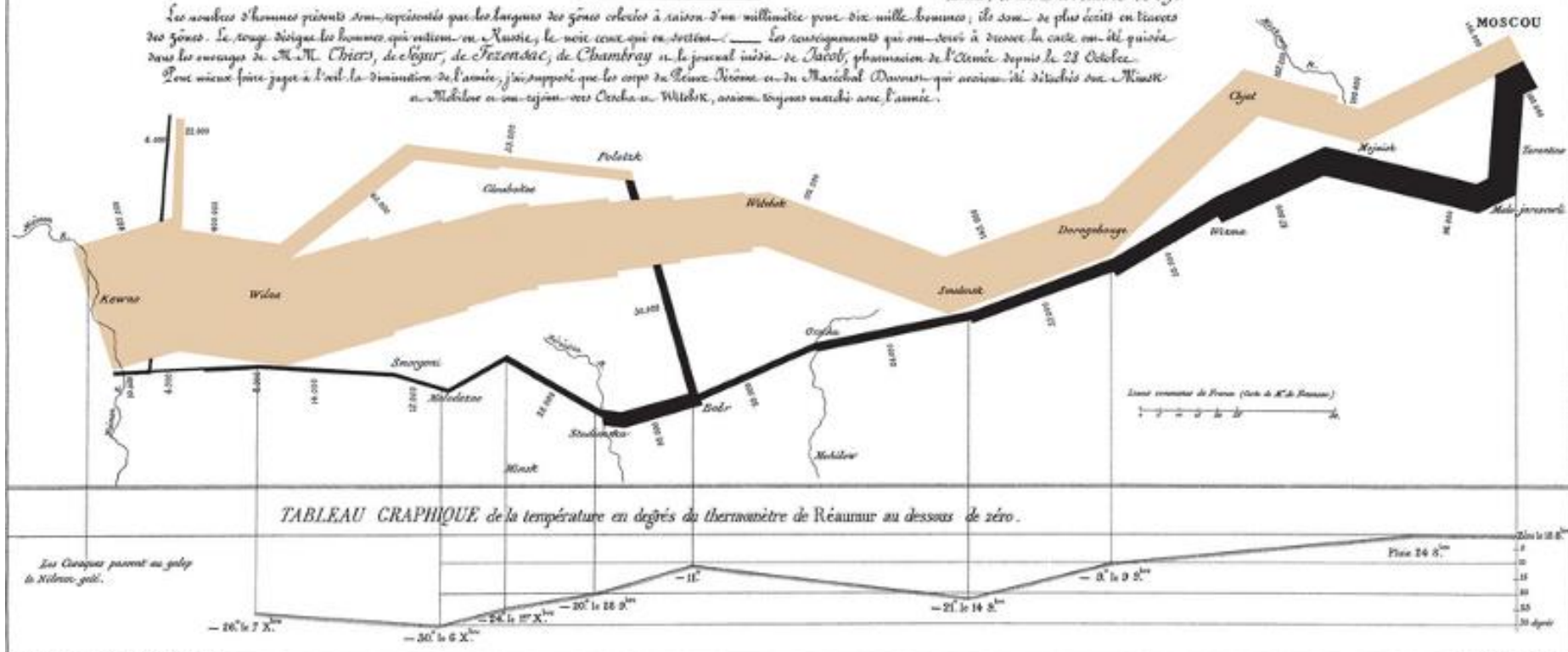


CHARLES JOSEPH MINARD (1781-1870)

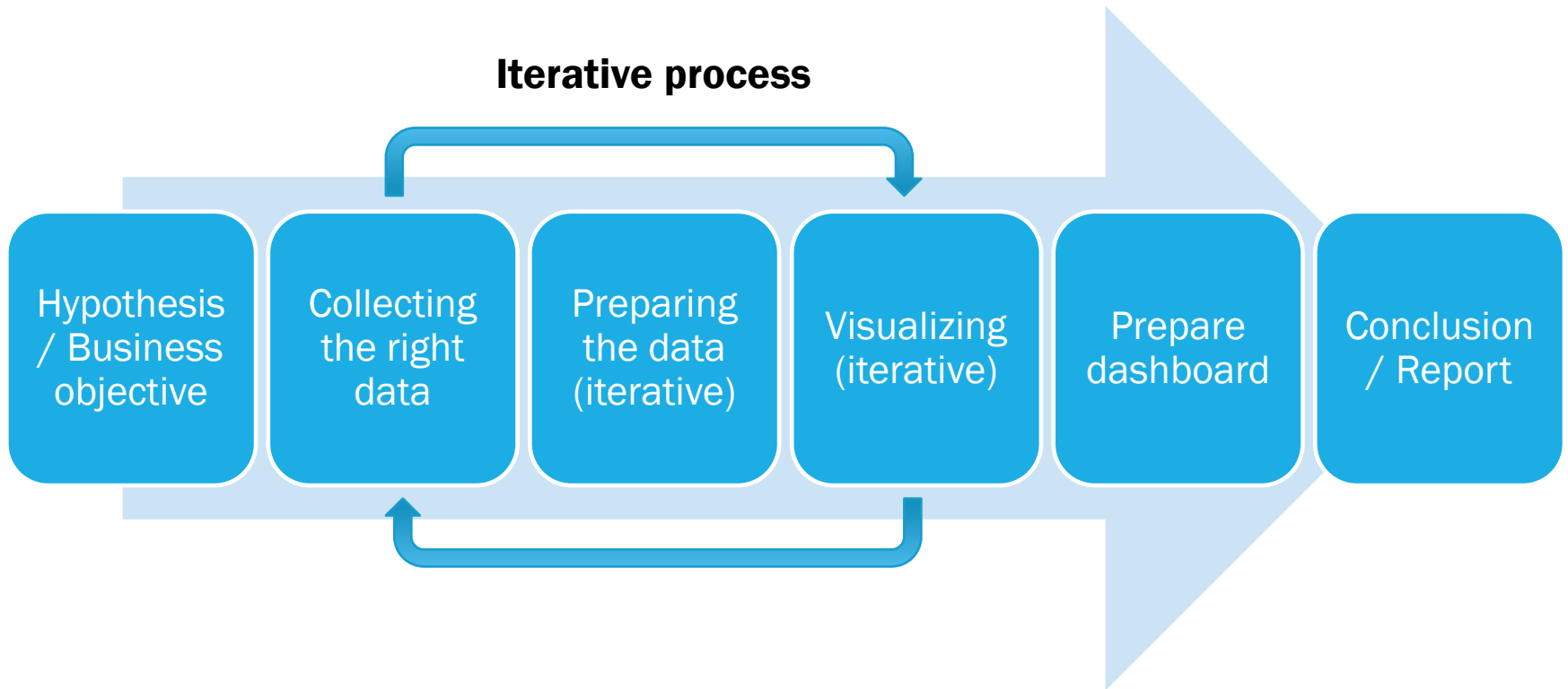
Carte Figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813.


Dessiné par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite. Paris, le 20 Novembre 1869.

Les nombres d'hommes présents sont représentés par les largueurs des zones colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en travers des zones. Le rouge désigne les hommes qui entrent en Russie, le noir ceux qui en sortent. Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de M. M. Chiers, de Légar, de Fozzardac, de Chambray et le journal russe de Jacob, pharmacien de l'Armée depuis le 23 Octobre. Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps de Reine Néron et du Maréchal Davoust qui avaient été détachés sur la Rousse en Moldavie et en ce jour vers Cernobila, avaient toujours marché avec l'armée.



PROCESS OF DATA VISUALIZATION





Your data is only as good as your ability to understand and communicate it, which is why choosing the right visualization is essential.

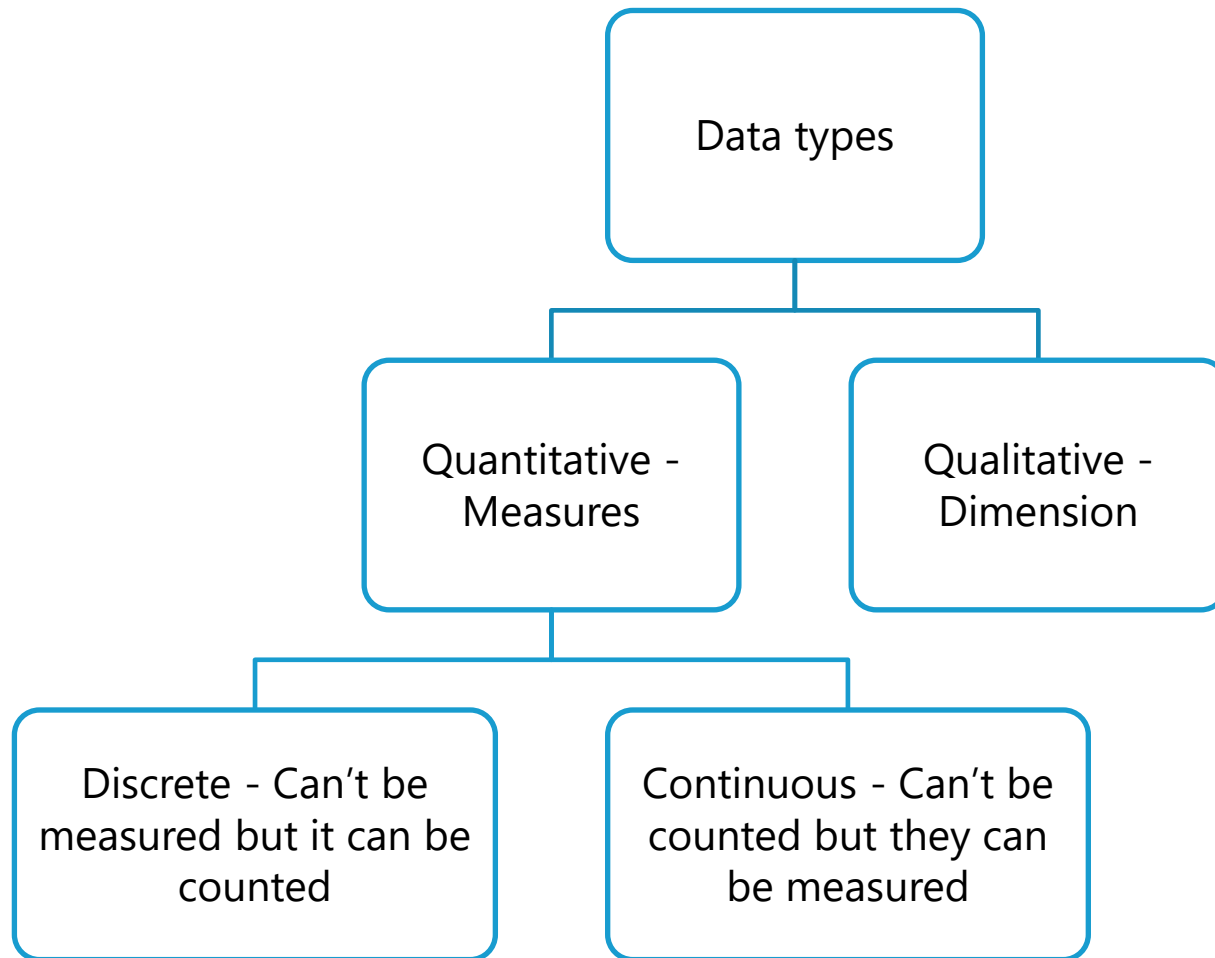


KNOW YOUR DATA

DATA TYPES & SCALE OF MEASUREMENT



DATA TYPES



SCALES OF MEASUREMENT

Nominal

- Nominal values represent discrete units and are used to label variables, that have no quantitative value

Ordinal

- Ordinal values represent discrete and ordered units.
- It is therefore nearly the same as nominal data, and that it is the ordering that matters

Interval

- Ordered units that have the same difference (interval). They don't have a true zero

Ratio

- The same as interval values, with the difference that they do have an absolute zero

VISUALIZATION TOOLS - LANDSCAPE

Power BI and
Tableau are the
leaders in the
Gartner magic
quadrant on BI
platforms

Figure 1. Magic Quadrant for Analytics and Business Intelligence Platforms



Source: Gartner (February 2020)

CHART TYPES (IN TABLEAU)

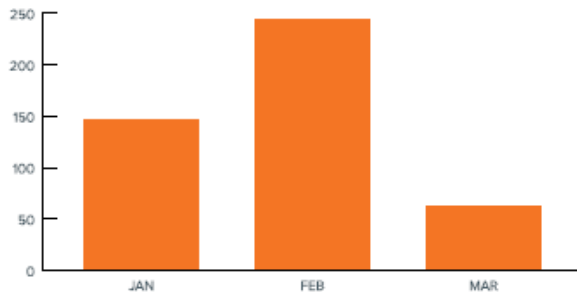
1. Text Table (Crosstab)
2. Heat Map
3. Highlight Table
4. Symbol Map
5. Filled Map (Area)
6. Pie Chart
7. Horizontal Bar Chart
8. Stacked Bar Chart
9. Side-by-Side Bar Chart
10. Treemap
11. Circle View
12. Side-by-Side Circle View
13. Line Charts (Continuous & Discrete)
14. Dual-Line Chart (Non-Synchronized)
15. Area Charts (Continuous & Discrete)
16. Scatter Plot
17. Histogram
18. Box-and-Whisker Plot
19. Gantt Chart
20. Bullet Graph
21. Packed Bubbles

BAR

- Bar charts are very versatile. They are best used to show change over time, compare different categories, or compare parts of a whole.

VARIATIONS OF BAR CHARTS

PAGE VIEWS, BY MONTH



VERTICAL (COLUMN CHART)

Best used for chronological data (time-series should always run left to right), or when visualizing negative values below the x-axis.

CONTENT PUBLISHED, BY CATEGORY



HORIZONTAL

Best used for data with long category labels.

STACKED BAR

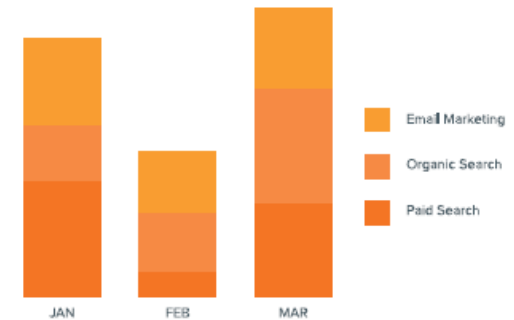
Stacked

- Best used when there is a need to compare multiple part-to-whole relationships. These can use discrete or continuous data, oriented either vertically or horizontally.

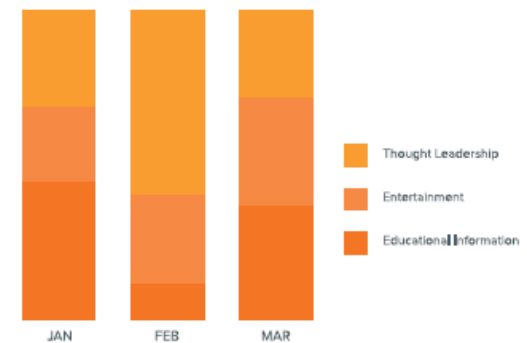
100% Stacked

- Best used when the total value of each category is unimportant and percentage distribution of subcategories is the primary message.

MONTHLY TRAFFIC, BY SOURCE



PERCENTAGE OF CONTENT PUBLISHED, BY MONTH





BAR CHART BEST PRACTICES

USE HORIZONTAL LABELS

- Avoid steep diagonal or vertical type, as it can be difficult to read.

SPACE BARS APPROPRIATELY

- Space between bars should be $\frac{1}{2}$ bar width.

START THE Y-AXIS VALUE AT 0

- Starting at a value above zero truncates the bars and doesn't accurately reflect the full value.

USE CONSISTENT COLORS

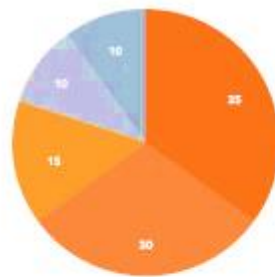
- Use one colour for bar charts. You may use an accent colour to highlight a significant data point. (unless colour is a dimension to display the data)

ORDER DATA APPROPRIATELY

- Order categories alphabetically, sequentially, or by value.

PIE CHART

- Pie charts are best used for making part-to-whole comparisons with discrete or continuous data. They are most impactful with a small data set.



STANDARD

Used to show part-to-whole relationships.



DONUT

Stylistic variation that enables the inclusion of a total value or design element in the center.

PIE CHART BEST PRACTICES

VISUALIZE NO MORE THAN 5 CATEGORIES PER CHART

- It is difficult to differentiate between small values; depicting too many slices decreases the impact of the visualization. If needed, you can group smaller values into an “other” or “miscellaneous” category, but make sure it does not hide interesting or significant information.

DON'T USE MULTIPLE PIE CHARTS FOR COMPARISON

- Slice sizes are very difficult to compare side-by-side. Use a stacked bar chart instead.

MAKE SURE ALL DATA ADDS UP TO 100%

- Verify that values total 100% and that pie slices are sized proportionate to their corresponding value.

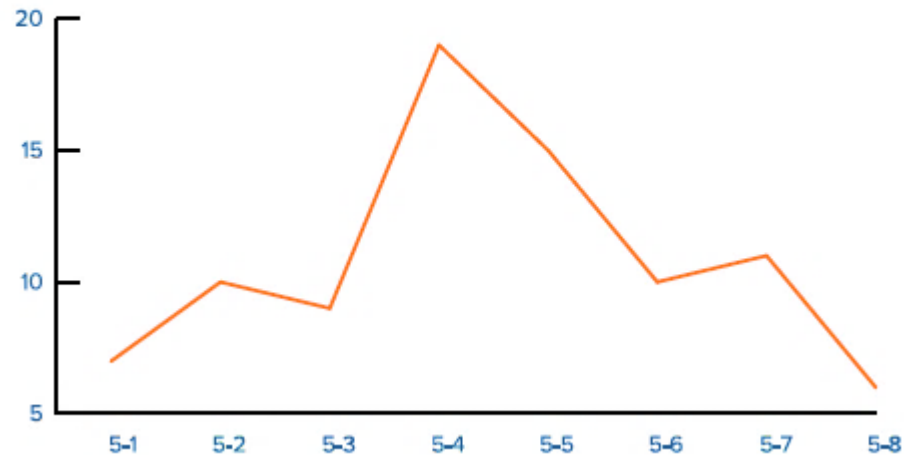
ORDER SLICES CORRECTLY

- There are two ways to order sections, both of which are meant to aid comprehension:
 - OPTION 1 – CLOCKWISE
 - OPTION 2 – COUNTER CLOCKWISE

LINE CHART

- Line charts are used to show time-series relationships with continuous data. They help show trend, acceleration, deceleration, and volatility.

DIRECT MARKETING VIEWS, BY DATE





LINE CHART BEST PRACTICES

INCLUDE A ZERO BASELINE IF POSSIBLE

- Although a line chart does not have to start at a zero baseline, it should be included if possible. If relatively small fluctuations in data are meaningful (e.g., in stock market data), you may truncate the scale to showcase these variances.

DON'T PLOT MORE THAN 4 LINES

- If you need to display more, break them out into separate charts for better comparison.

USE SOLID LINES ONLY

- Dashed and dotted lines can be distracting.

LABEL THE LINES DIRECTLY

- This lets readers quickly identify lines and corresponding labels instead of referencing a legend.

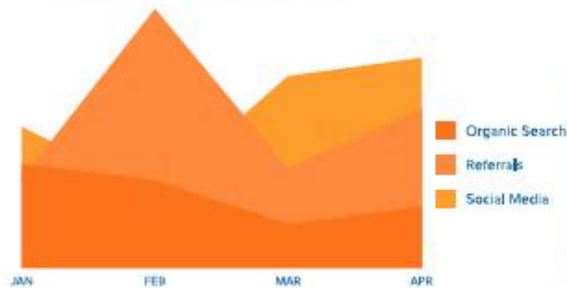
USE THE RIGHT HEIGHT

- Plot all data points so that the line chart takes up approximately two-thirds of the y-axis' total scale.

AREA CHART

- Area charts depict a time-series relationship, but they are different than line charts in that they can represent volume.

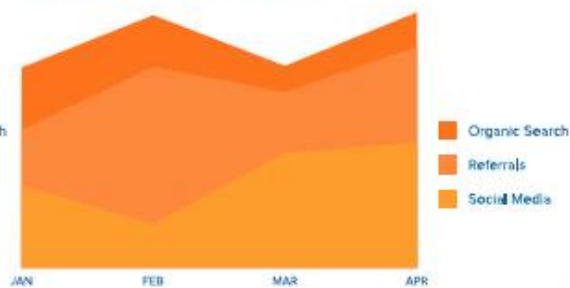
NEW CONTACTS, BY SOURCE



AREA CHART

Best used to show or compare a quantitative progression over time.

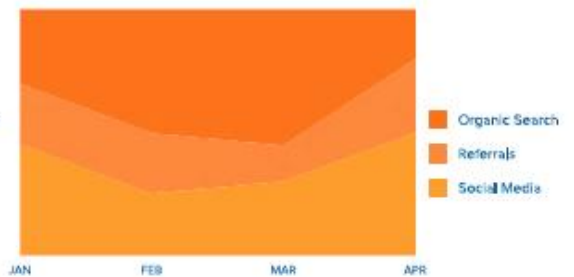
NEW CONTACTS, BY SOURCE



STACKED AREA

Best used to visualize part-to-whole relationships, helping show how each category contributes to the cumulative total.

NEW CONTACTS, BY SOURCE



100% STACKED AREA

Best used to show distribution of categories as part of a whole, where the cumulative total is unimportant.



AREA CHART BEST PRACTICES

MAKE IT EASY TO READ

- In stacked area charts, arrange data to position categories with highly variable data on the top of the chart and low variability on the bottom.

START Y-AXIS VALUE AT 0

- Starting the axis above zero truncates the visualization of values.

DON'T DISPLAY MORE THAN 4 DATA CATEGORIES

- Too many will result in a cluttered visual that is difficult to decipher

USE TRANSPARENT COLORS

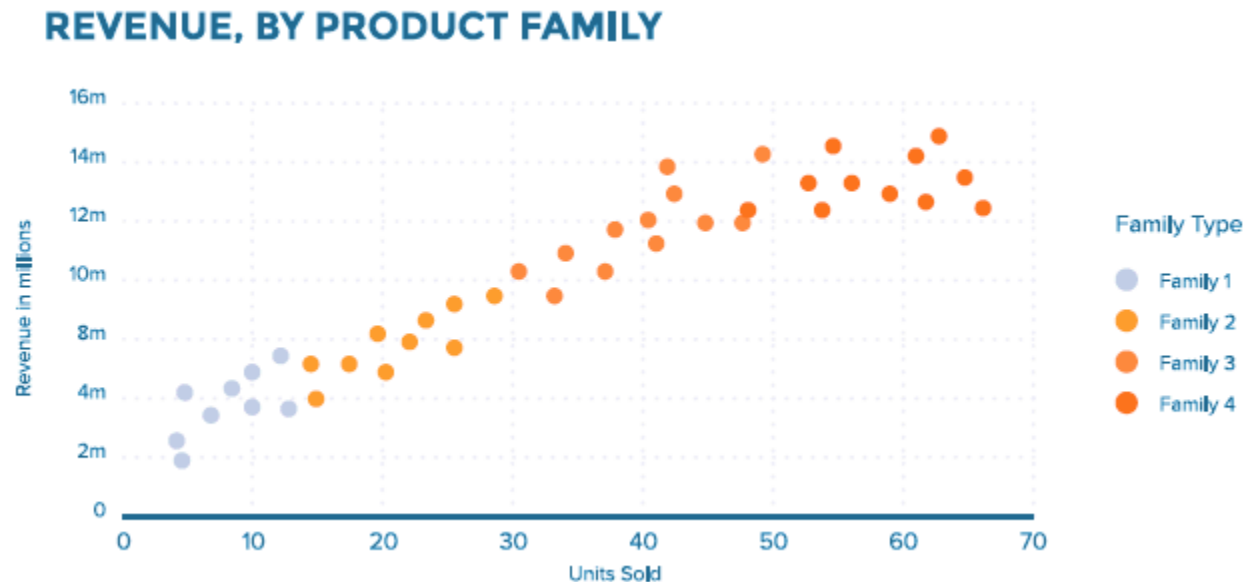
- In standard area charts, ensure data isn't obscured in the background by ordering thoughtfully and using transparency.

DON'T USE AREA CHARTS TO DISPLAY DISCRETE DATA

- The connected lines imply intermediate values, which only exist with continuous data.

SCATTER PLOT

- Scatter plots show the relationship between items based on two sets of variables. They are best used to show correlation in a large amount of data.





SCATTER PLOT BEST PRACTICES

START Y-AXIS VALUE AT 0

- Starting the axis above zero truncates the visualization of values.

INCLUDE MORE VARIABLES

- Use size and dot colour to encode additional data variables.

USE TREND LINES

- These help draw correlation between the variables to show trends.

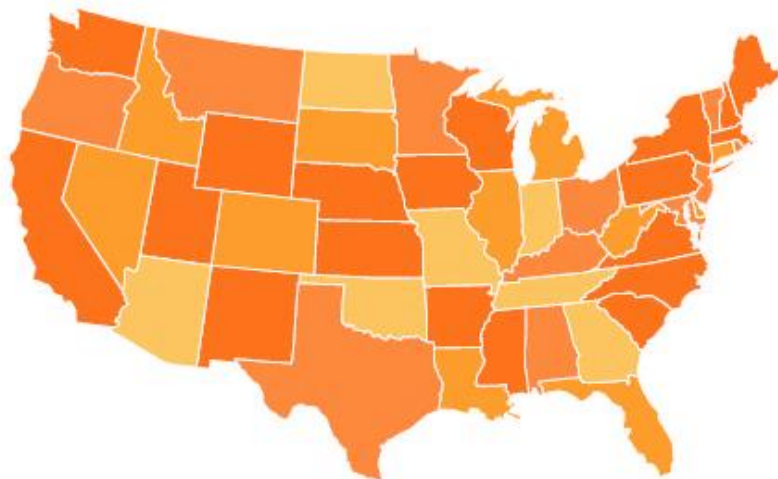
DON'T COMPARE MORE THAN 2 TREND LINES

- Too many lines make data difficult to interpret.

HEAT MAP

- Heat maps display categorical data, using intensity of colour to represent values of geographic areas or data tables.

STATES WITH NEW SERVICE CONTRACTS



75-76 77-78 79-80 81+

QUARTERLY SALES BY VEHICLE TYPE

	Sample Size	F19Q1	F19Q3	F20Q1	F19Q1	F19Q3	F20Q1
Bike		50% 591 0.5042	51% 379 0.5134	47% 432 0.4687	49% 7,272 0.4641	47% 4,304 0.3990	43% 5,026 0.3785
Car		22% 255 0.2177	19% 139 0.1891	25% 226 0.2450	27% 3,969 0.2533	29% 2,681 0.2485	30% 3,502 0.2637
Horse		28% 326 0.2781	30% 219 0.2975	29% 264 0.2862	24% 3,547 0.2264	24% 2,207 0.2046	27% 3,115 0.2346



HEAT MAP BEST PRACTICES

USE A SIMPLE MAP OUTLINE

- These lines are meant to frame the data, not distract

SELECT COLORS APPROPRIATELY

- Some colours stand out more than others, giving unnecessary weight to that data. Instead, use a single colour with varying shade or a spectrum between two analogous colours to show intensity. Also remember to intuitively code colour intensity according to values.

CHOOSE APPROPRIATE DATA RANGES

- Select 3-5 numerical ranges that enable fairly even distribution of data between them.



FINDING THE STORY IN YOUR DATA

FINDING THE STORY IN YOUR DATA



TRENDS, CORRELATION, OUTLIERS

TRENDS



Example:
Ice cream sales
over time

CORRELATIONS



Example:
Ice cream sales vs.
temperature

OUTLIERS



Example:
Ice cream sales in an
unusual region

DATA RELATIONSHIPS

Nominal Comparison	<ul style="list-style-type: none">• This is a simple comparison of the quantitative values of subcategories. Example: Number of visitors to various websites.
Time Series	<ul style="list-style-type: none">• This tracks changes in values of a consistent metric over time. Example: Monthly sales.
Correlation	<ul style="list-style-type: none">• This is data with two or more variables that may demonstrate a positive or negative correlation to each other. Example: Salaries according to education level
Distribution	<ul style="list-style-type: none">• This shows data distribution, often around a central value. Example: Heights of players on a basketball team.

DATA RELATIONSHIPS (CONTD.)

Part-to-whole Relationships

- This shows a subset of data compared to the larger whole. Example: Percentage of customers purchasing specific products.

Ranking

- This shows how two or more values compare to each other in relative magnitude. Example: Historic weather patterns, ranked from the hottest months to the coldest.

Deviation

- This examines how data points relate to each other, particularly how far any given data point differs from the mean. Example: Amusement park tickets sold on a rainy day vs. a regular day.



DASHBOARD

GATHER YOUR CHARTS AND CREATE STORIES





DASHBOARD DESIGN PRINCIPLES

- Include text to explain your dashboard
- Each dashboard should be constructed based on a story (e.g. Sales performance by multiple dimensions like month, author, genre, book, territory, etc.)
- Do not clutter the dashboard
- Add filters on top of the dashboard (applicable to all charts)
- Build interaction in the charts – add filters only if necessary
- Use colour legends next to the charts if there are more than one



BEST PRACTICES FOR BUILDING EFFECTIVE DASHBOARDS

In general

- Know your audience
- Limit the number of views & colours
- Add interactivity to encourage exploration
- Test your dashboard before deploying

Tableau specific

- Consider display size
- Plan for fast load times
- Leverage tooltips, the story within your story



CLOSING NOTES

POINTS TO REMEMBER WHILE CREATION VISUALS



10 DATA DESIGN DOS AND DON'TS

Dos

- Do use one colour to represent each category.
- Do order data sets using logical hierarchy.
- Do highlight important or interesting information.
- Do visualize data in a way that is easy for readers to compare values.
- Do use icons to enhance comprehension and reduce unnecessary labelling.

Don'ts

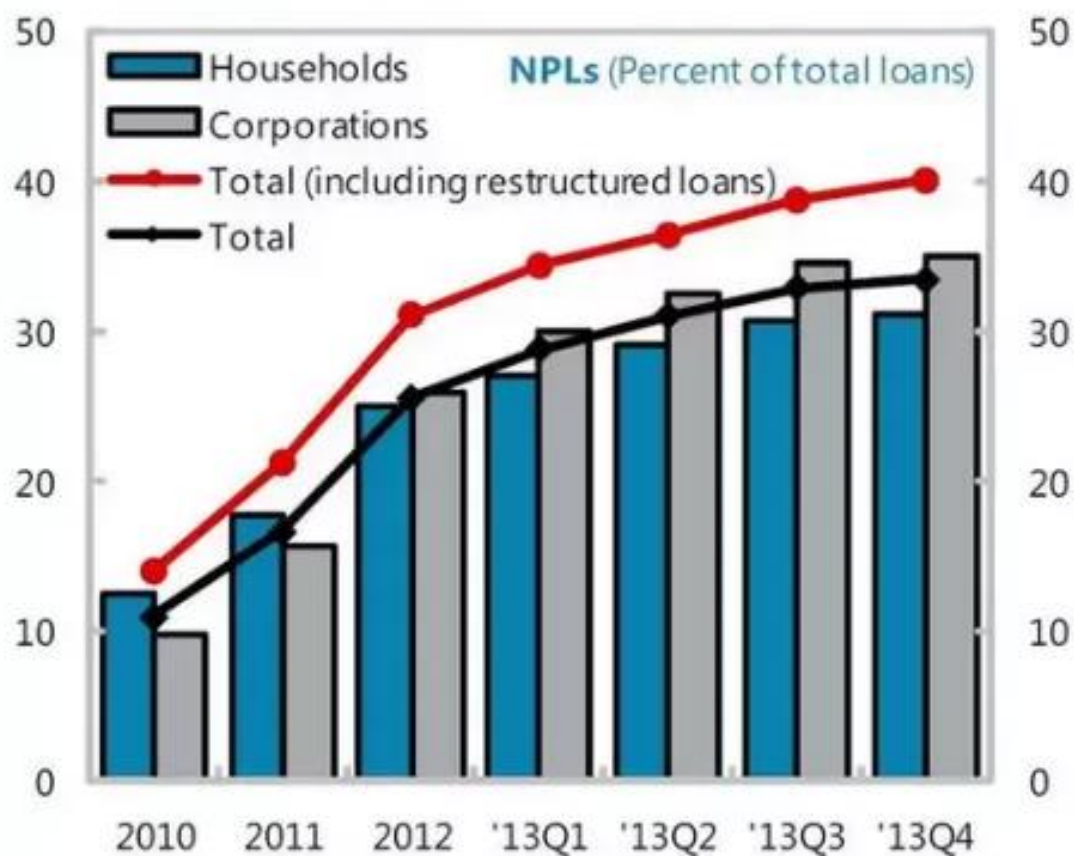
- Don't use high contrast colour combinations such as red/green or blue/yellow unless you want to show contrasting information (+ve / -ve)
- Don't use 3D charts. They can skew perception of the visualization.
- Don't add chart junk. Unnecessary illustrations, drop shadows, or ornamentations distract from the data.
- Don't use more than 6 colours in a single layout
- Don't use distracting fonts or elements (such as bold, italic, or underlined text).



COMMON MISTAKES TO AVOID

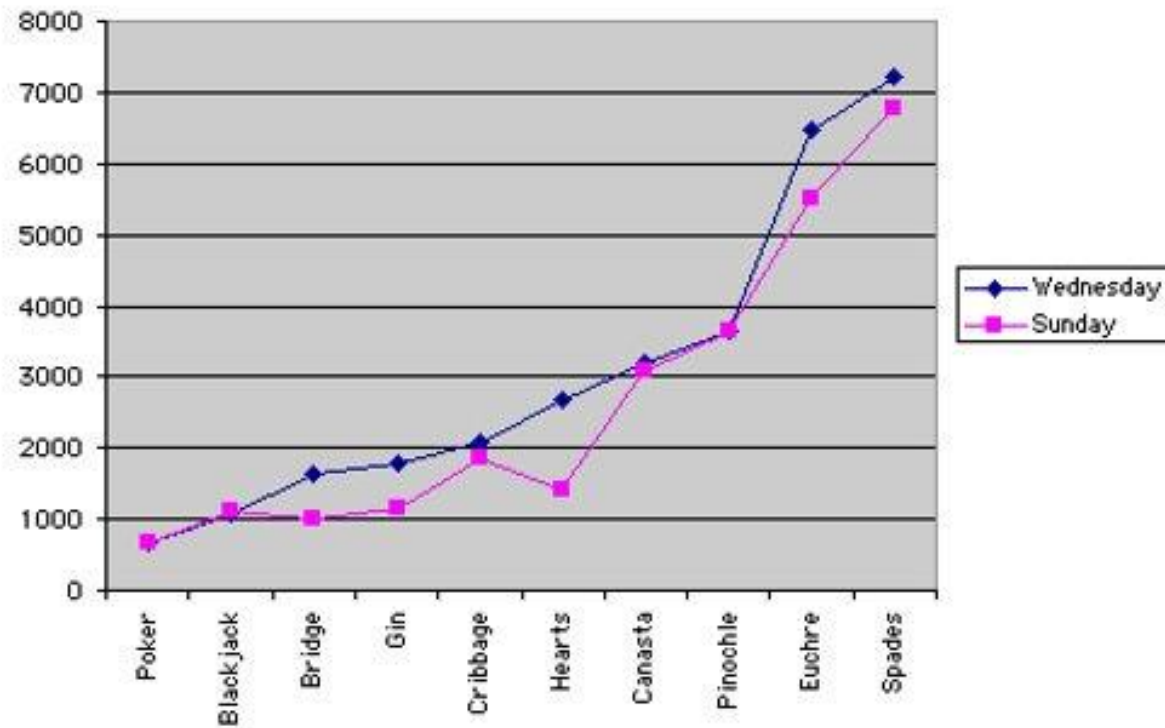
- Trying to answer too many questions with a single dashboard
- Using metrics no one understands
- Cluttering the dashboard with low-value graphics and widgets
- Not viewing your dashboard as your users will
- Forgetting to check in with your audience

WHAT IS WRONG IN THIS CHART?

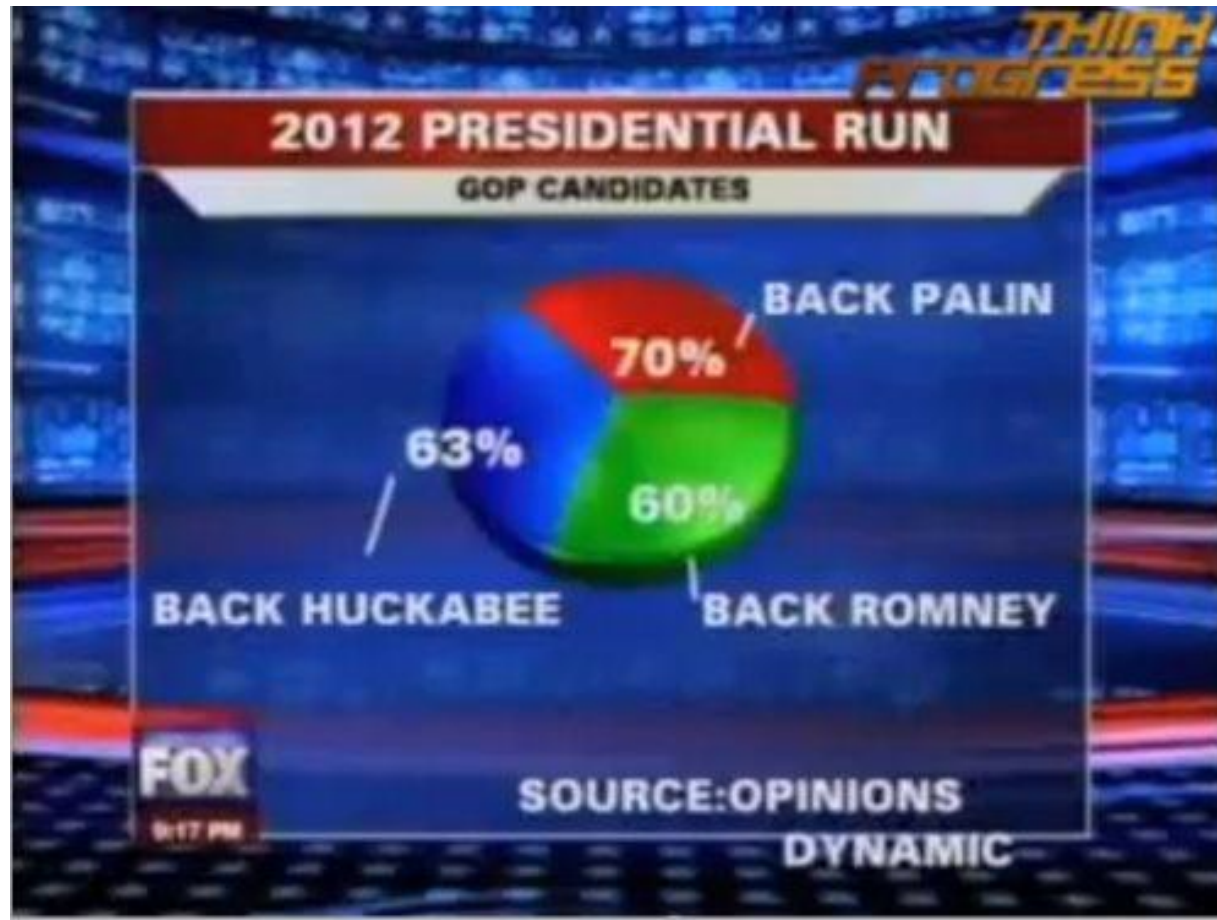


Sources: Bank of Greece; and IMF staff calculations.

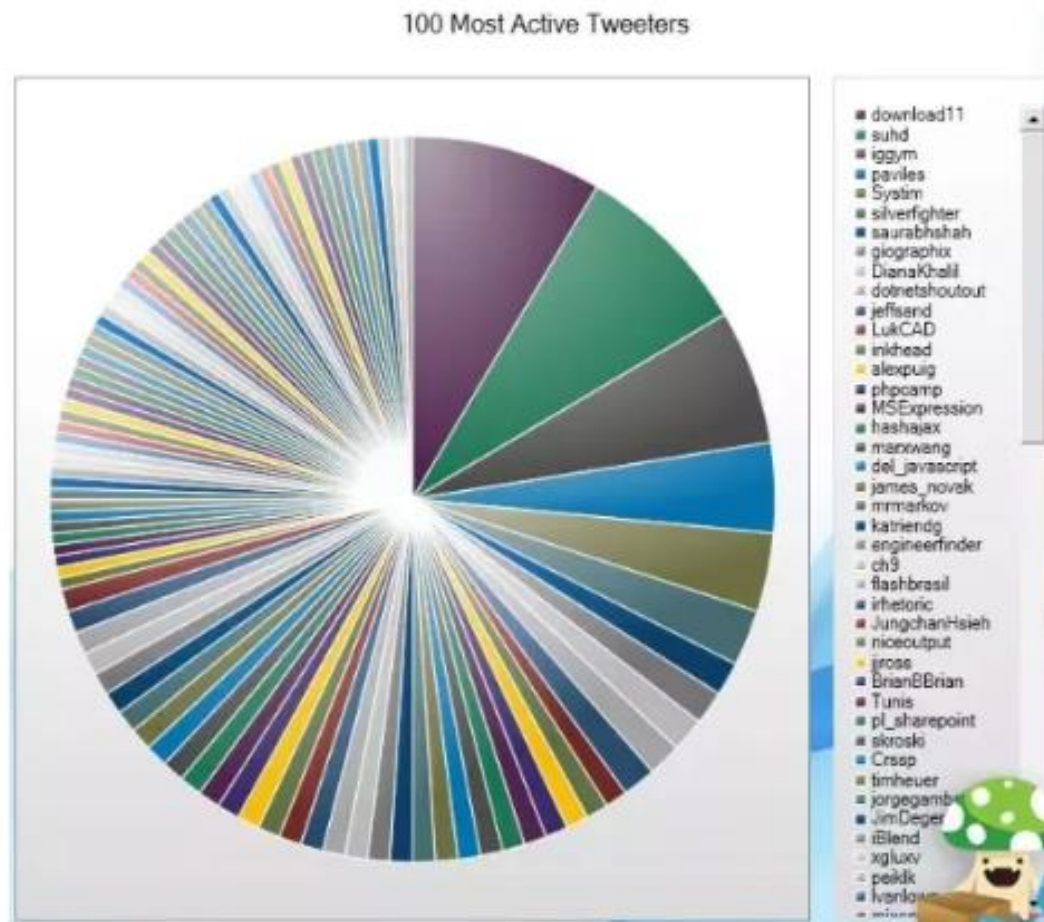
WHAT IS WRONG IN THIS CHART?



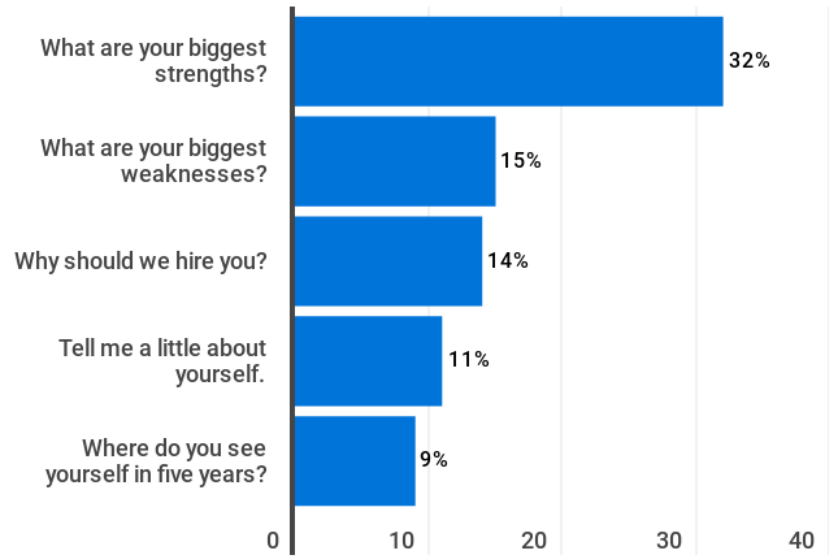
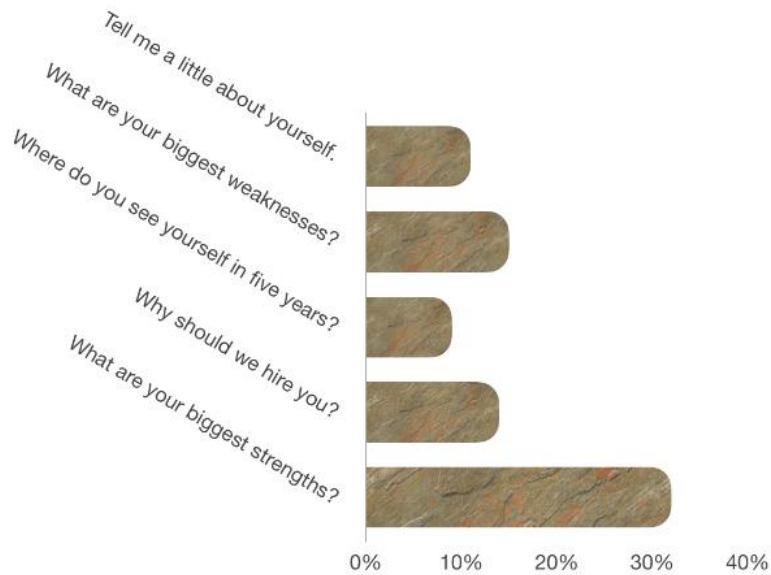
WHAT IS WRONG IN THIS CHART?



WHAT IS WRONG IN THIS CHART?



WHAT IS WRONG IN THIS CHART?





THANK YOU, QUESTIONS?