



DATA Visualization

Past – Present – Future

Sathish Sundaram
Jul 2019



Key Takeaways

- Data Visualization
 - Does it matter ?
 - Is this new ?
 - Where are we headed ?

Data Visualization

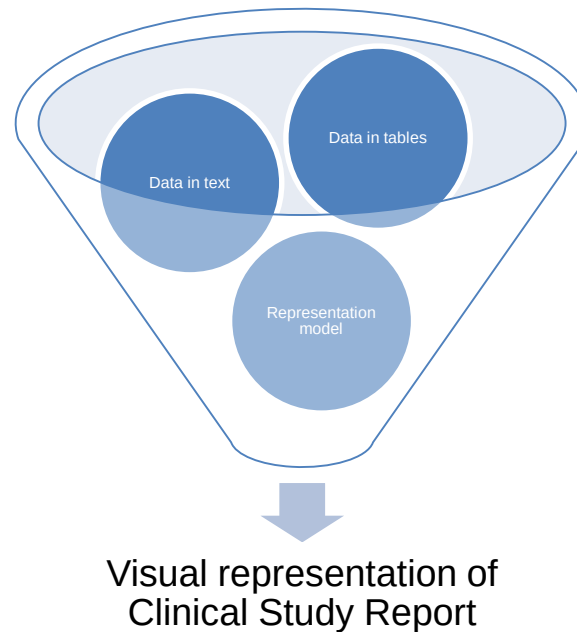


Does it Really Matter ?

Data on Visualization

- Usability Study
- Text → Visual representation of text
- Numbers → Visual representation
- Controlled study

Data Visualization Usability Study



<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5333215/>

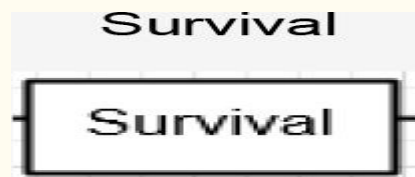
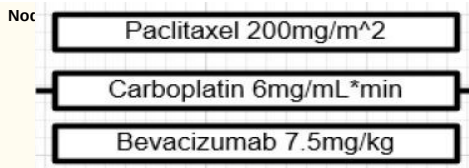
Data Visualization Usability Study

Excerpts from report

"Patients received up to six cycles of carboplatin/paclitaxel. Paclitaxel (200 mg/m²) was administered over 3 hours every 3 weeks. Carboplatin dosing was based on the Calvert formula with a target area under the curve of 6 mg/mL x min and glomerular filtration rate (GFR) estimated for males as $GFR = (140 - \text{age}) \times \text{weight} / 72 \times (\text{serum creatinine})$."

"Eleven patients discontinued treatment as a result of a nonfatal AE. Discontinuations occurred as a result of: hemorrhagic event (three patients) in the low-dose bevacizumab arm; a hemorrhagic event (one patient); Aspergillus lung abscess (one patient); aspiration pneumonia (one patient); thrombotic stroke (one patient); vertebral fracture (one patient); and peripheral neuropathy (paclitaxel-related; one patient) in the high-dose arm. In two cases, bevacizumab was discontinued following initiation of anticoagulant therapy. Bevacizumab was withheld from one patient with subclavian vein thrombosis."

"The overall response rate showed a trend toward improved response for patients receiving bevacizumab, with the highest response noted in the high-dose group (31.5%) and the lowest in the control group (18.8%; Table 2)...Median TTP was longer in the high-dose bevacizumab arm compared with the control arm (7.4 v 4.2 months; $P = .023$; Fig 1)...Survival for the high-dose bevacizumab arm was modestly longer than the control arm (17.7 v 14.9 months; $P = .63$; Fig 2)."



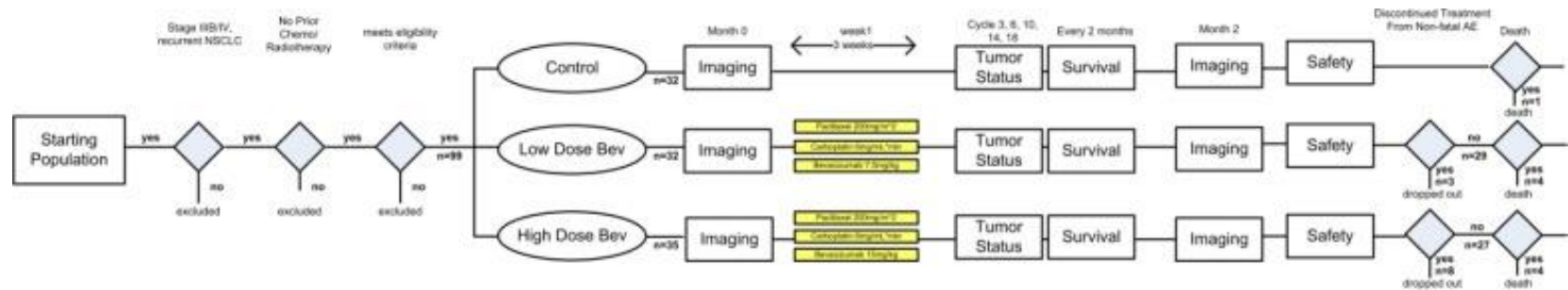
Variables within Node

- drug name(s)
- dosage(s)
- duration
- frequency
- Hemorrhagic event
- Aspergillus lung abscess
- aspiration pneumonia
- thrombotic stroke
- vertebral fracture
- peripheral neuropathy
- following initiation of anticoagulant therapy
- subclavian vein thrombosis

- median survival (month)
- median survival (range)
- % PFS vs. time
- % survival vs. time

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5333215/>

Data Visualization Usability Study



<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5333215/>



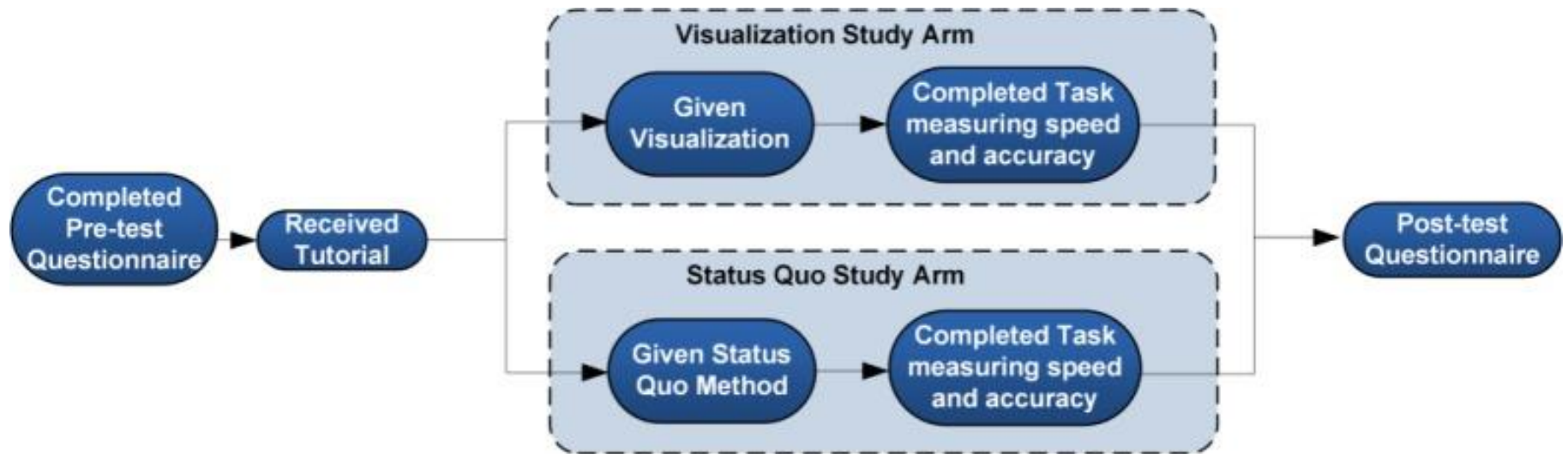
Data Visualization Usability Study

- Test the ease of interpretation and time saved against the status quo (i.e., the published report)
- Test the intuitiveness of the visualization through reproducibility of the visualization,
- Elicit preferences from a targeted user group

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5333215/>



Data Visualization Usability Study



<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5333215/>

Data Visualization Usability Study

Trial	Visualization				Status Quo			
	Accuracy (%)	SD	Time (min)	SD	Accuracy (%)	SD	Time (min)	SD
1	77.3%	5.34%	27.3	10.73	68.9%	10.01%	34.5	10.63
2	58.5%	11.67%	25.0	8.86	53.8%	14.80%	34.3	9.63
3	84.5%	6.10%	24.6	12.42	78.8%	10.23%	38.2	11.20

Measures of performance as a function of errors and time

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5333215/>



Data Visualization Usability Study

CONCLUSIONS

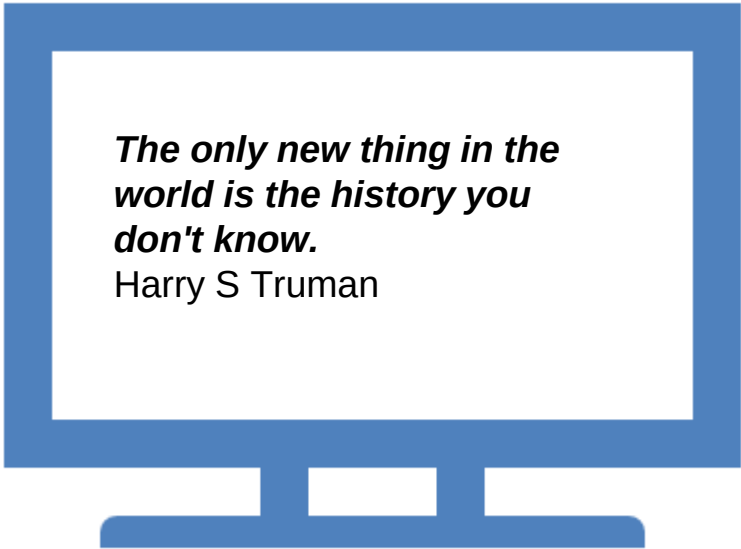
- *VISUALIZATIONS FOR TRIAL REPORTS COULD PROVIDE TIME SAVINGS AND ACHIEVE SIMILAR ACCURACY*
- *UNDERSTANDING THE STRENGTH AND QUALITY OF CLINICAL TRIALS CAN BE ALLEVIATED WITH A VISUALIZATION*



Data Visualization



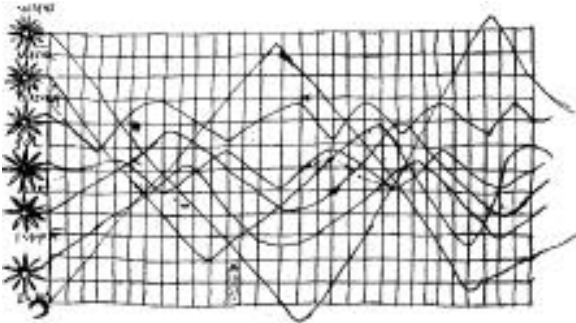
Data Visualization



***The only new thing in the
world is the history you
don't know.***

Harry S Truman

Data Visualization – Journey Through History - ~10th Century



- Top: Planetary movements shown as cyclic inclinations over time, by an unknown astronomer, appearing in a 10th century appendix to commentaries by A. T. Macrobius on Cicero's *In Somnium Scipionis*. Source: Funkhouser (1936, p. 261).

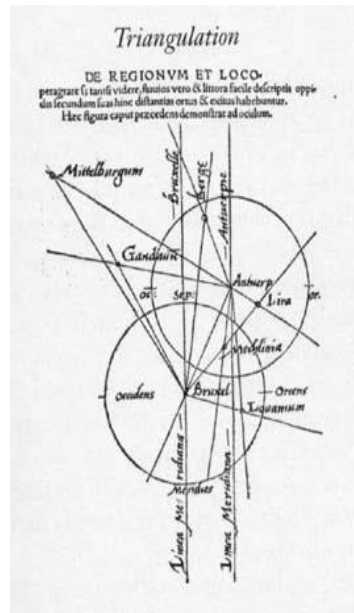
Data Visualization – Journey Through History - ~14th Century



- Nicole Orseme's Proto Bar Graph

<http://datavis.ca/papers/hbook.pdf>

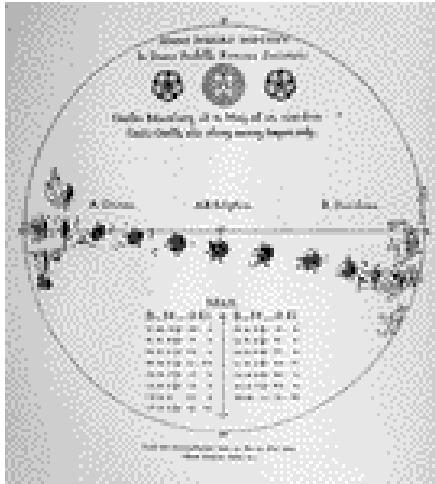
Data Visualization – Journey Through History - ~16th Century



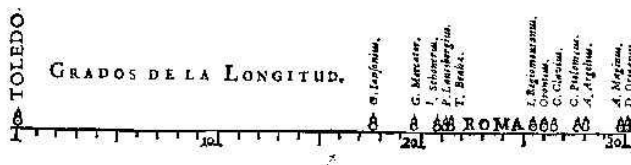
- Gemma-Frisius Diagram of triangulation

<http://datavis.ca/papers/hbook.pdf>

Data Visualization – Journey Through History - 17th Century

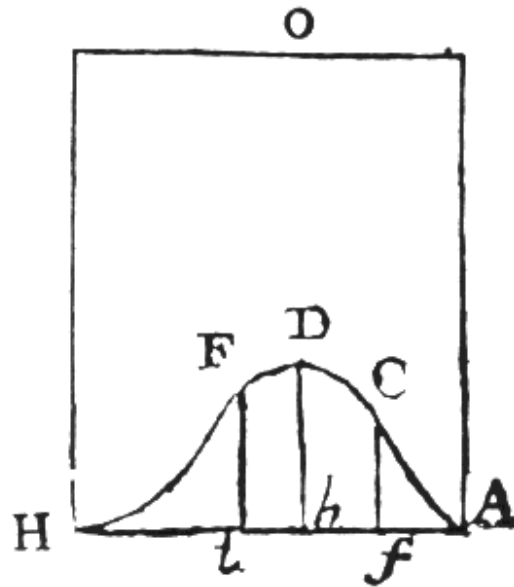


- Top: Scheiner's 1626 representation of the changes in sunspots over time
- Bottom: Langren's 1644 graph of determinations of the distance, in longitude, from Toledo to Rome. The correct distance is $16^{\circ}30'$.



<http://datavis.ca/papers/hbook.pdf>

Data Visualization – Journey Through History - 18th Century



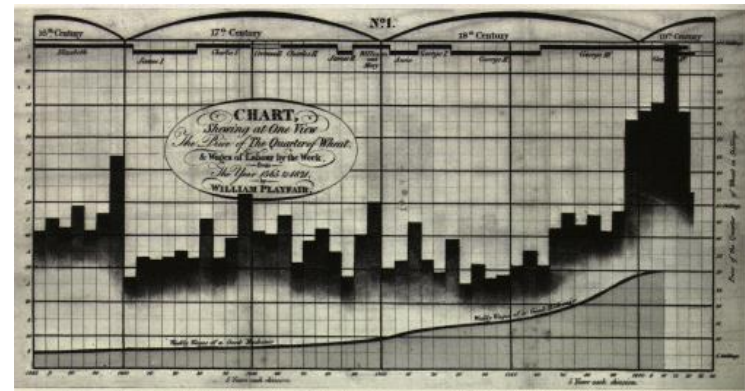
- Bayes' Graph of the beta density

<http://datavis.ca/papers/hbook.pdf>

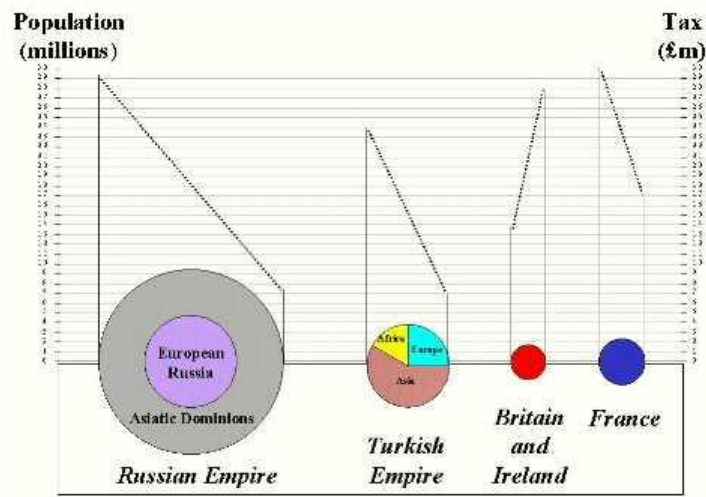
Data Visualization

– Journey Through History - 19th Century

- William Playfair's 1821 time series graph of prices, wages, and ruling monarch over a 250 year period.
Source:



<http://datavis.ca/papers/hbook.pdf>



Data Visualization – Journey Through History - 19th Century

- Re-drawn version of a portion of Playfair's 1801 pie-circle-line chart, comparing population and taxes in several nations.

<http://datavis.ca/papers/hbook.pdf>

- Gantt Chart 1917.

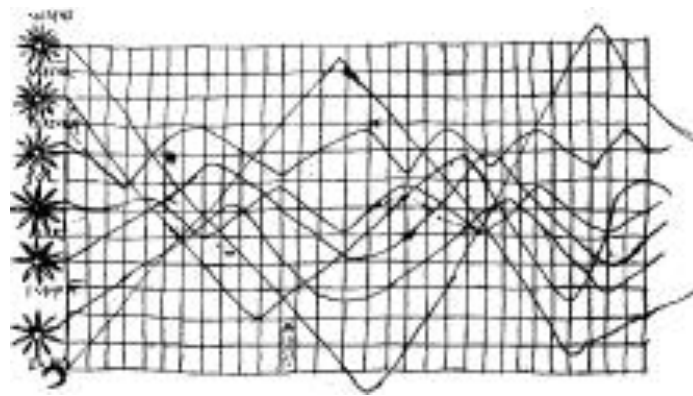
Created Using Milestones Professional
www.KIDASA.com



Data Visualization



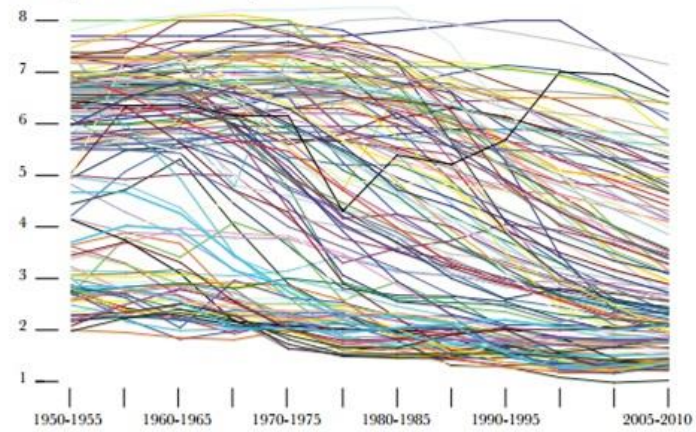
Where are we headed?



10th Century

Fertility Rate

Average number of children per woman over her lifetime
Showing all countries for which complete data is available



Today

Visualization Importance



- VISUALIZING LARGE AMOUNTS OF COMPLEX DATA IS EASIER THAN PORING OVER SPREADSHEETS OR REPORTS
- YOU CAN EXPERIMENT WITH DIFFERENT SCENARIOS
- DATA VISUALIZATION CAN ALSO:
 - IDENTIFY AREAS THAT NEED ATTENTION OR IMPROVEMENT.
 - CLARIFY WHICH FACTORS INFLUENCE OUTCOMES
 - SEE THE FUTURE

https://www.sas.com/en_in/insights/big-data/data-visualization.html#dmusers



How Complex Biopharma data is?

- **Volume** - National Health Service (NHS) in the UK adds 19 million inpatient records, 90 million outpatient records, and 18 million accident and emergency records to its databases every year.
- **Cost** - Trials typically cost more than \$2 billion for every approved drug (including the cost of failures), and the average new drug has approximately a 10% chance of making it through trials successfully.



<https://www.bcg.com/publications/2015/big-data-advanced-analytics-biopharmaceuticals-making-big-data-work-biopharma.aspx>



ACCENT Principles of effective graphical display

A_{pp}rehension:

Ability to correctly perceive relations among variables.

Does the graph maximize apprehension of the relations among variables?

C_larity:

Ability to visually distinguish all the elements of a graph.

Are the most important elements or relations visually most prominent?

C_onsistency:

Ability to interpret a graph based on similarity to previous graphs.

Are the elements, symbol shapes and colors consistent with their use in previous graphs?

E_{ff}iciency:

Ability to portray a possibly complex relation in as simple a way as possible.

Are the elements of the graph economically used?

Is the graph easy to interpret?

N_ecessity:

The need for the graph, and the graphical elements.

Is the graph a more useful way to represent the data than alternatives (table, text)?

Are all the graph elements necessary to convey the relations?

T_ruthfulness:

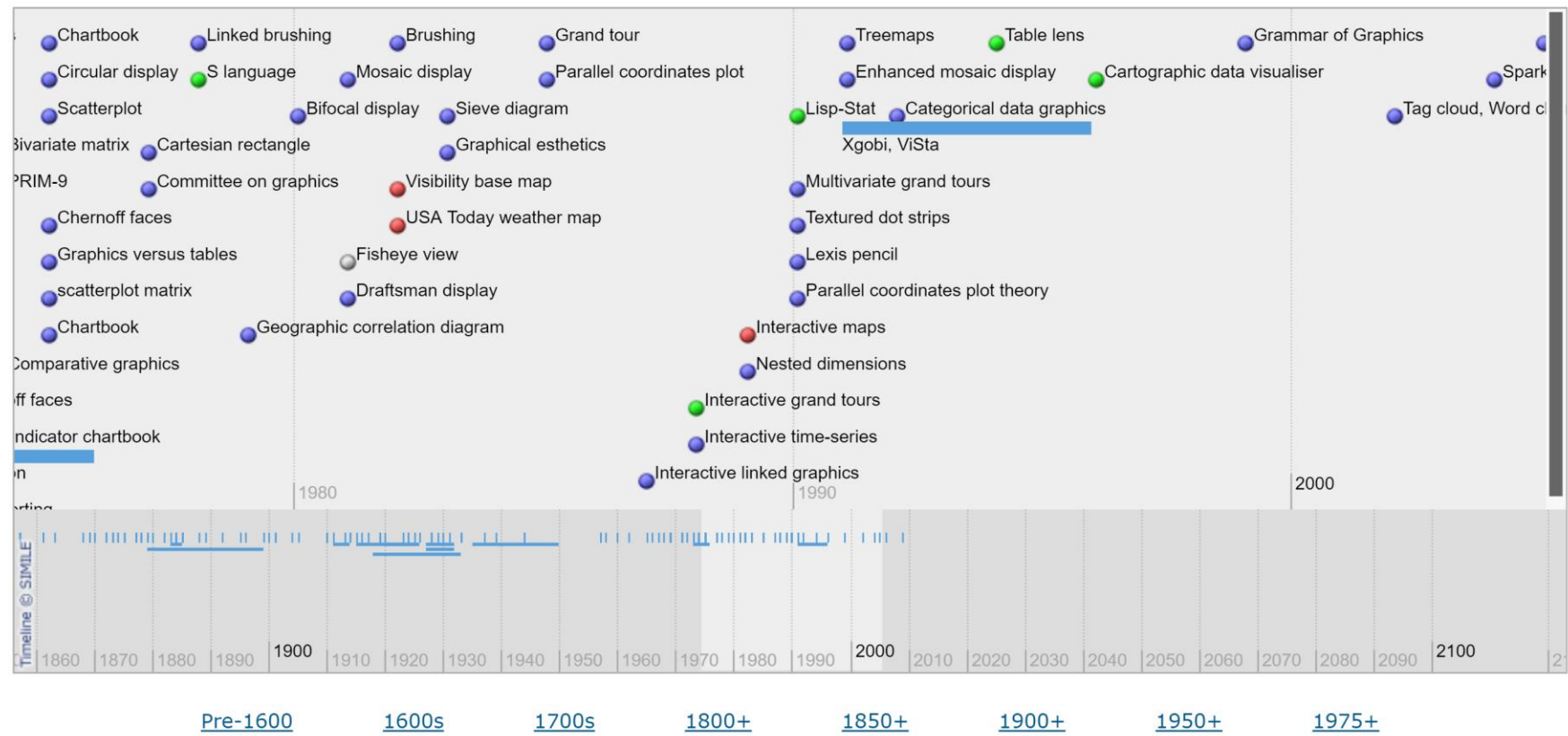
Ability to determine the true value represented by any graphical element by its magnitude relative to the implicit or explicit scale.

Are the graph elements accurately positioned and scaled?

Ref: <http://www.datavis.ca/gallery/accent.php>



Item categories: ● Cartography ● Statistics and graphics ● Technology ● Other

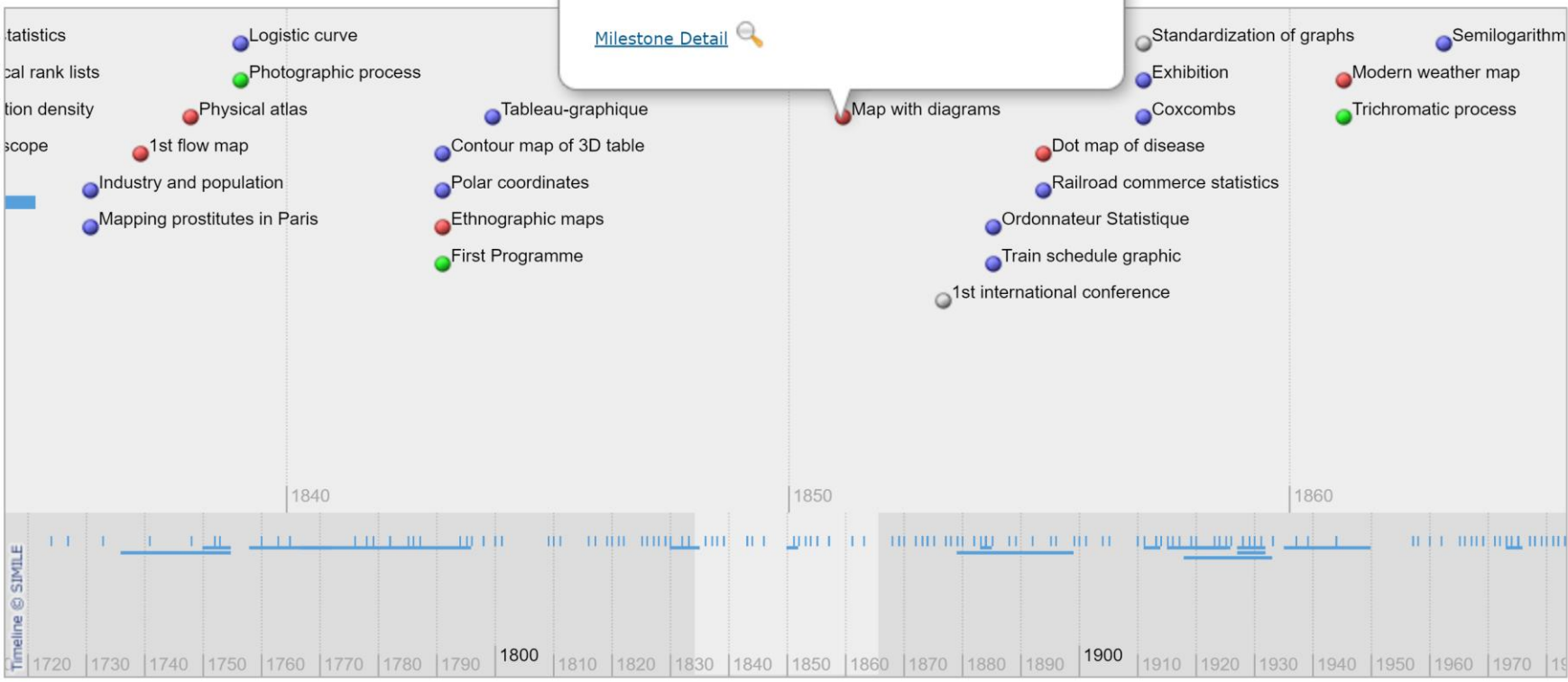


Item categories:

1851 (France) [Cartography](#)

Charles Joseph Minard (1781-1870)
Map incorporating statistical diagrams: circles proportional to coal production (published in 1861)

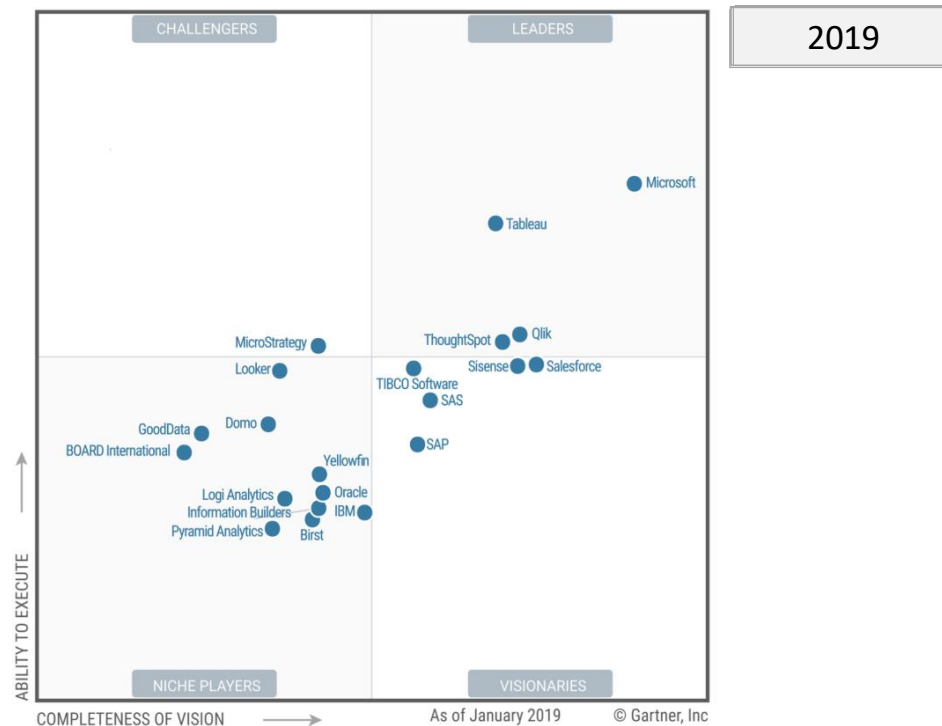
[Milestone Detail](#)



[Pre-1600](#) [1600s](#) [1700s](#) [1800+](#) [1850+](#) [1900+](#) [1950+](#) [1975+](#)

Data Visualization Tools Landscape

Gartner Magic Quadrant for Analytics and Business Intelligence Platforms
2014-2019



Visualization – More Reasons

- Enable real time analysis.
- Dynamic and Interactive presentation of reports.
- High level of end user interaction with the reports.
- Allow end users to share and collaborate securely.
- Interactive visualization, you can take the concept a step further by using technology to drill down into charts and graphs for more detail, interactively changing what data you see and how it's processed



References

- <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5333215/>
- [https://www.phusewiki.org/wiki/index.php?title=Data Visualisations for Clinical Data](https://www.phusewiki.org/wiki/index.php?title=Data_Visualisations_for_Clinical_Data)
- <http://datavis.ca/milestones/index.php>