Unit-3

Data Visualization : What?

- Data visualization is the practice of translating information into a visual context, such as a **map** or **graph**, to make data easier for the human brain to understand and pull insights from.

- The main **goal** of data visualization is **to make it easier to identify patterns, trends and outliers** in **large data sets**.

- Data visualization is the process of creating a visual representation of the information within a dataset.

Data Visualization : Why?

- Visually depicting data often makes it easier to understand and draw insights from.
- Data visualization is an effective means of making data more accessible across an organization.
- It empowers stakeholders to back their actions using concrete information instead of relying on assumptions—resulting in more data-driven organizational processes.
- Data visualization can also help play an important role in communication with parties outside the institution (such as the media, investors, regulatory agents, and other stakeholders etc.).

Data Visualization : How?

- While there are hundreds of ways to visualize data, some of the most common data visualization techniques include:
  - Pie charts
  - Bar charts
  - Histograms
  - Gantt charts
  - Heat maps
  - Box-and-whisker plots
  - Waterfall charts
  - Area charts
  - Scatter plots
  - Infographics
  - Maps

Data Visualization Tools : How?

- Data scientists have to analyze, interpret, and visualize large datasets on a daily basis. Consequently, it is important for them to have the right **data visualization tools** at their disposal.

- Data visualization tools allow data scientists to **communicate their findings more effectively,** which is important because it allows them to share their insights with other people who may not be familiar with data science concepts.

Some

Data Visualization Tools

Tableau
(https://www.tableau.com/)

Tableau is a data visualization tool that can be used to create interactive graphs, charts, and maps.

It allows us to connect to different data sources and **create visualizations in minutes.**

We can also share our work with others and collaborate on projects.

Tableau Desktop is the original product. It's made for creating static visualizations that can be published on one or more web pages, but it doesn't create interactive maps.

Tableau Public is the free distribution of the Desktop product with some limitations.

Learning Tableau takes time and practice, but there are plenty of resources out there to help us learn how to use it.

As a data scientist, Tableau has to be the number one tool for you to learn and use in your everyday tasks.

QlikView

(https://www.qlik.com/us/products/qlikview)

- QlikView is a **data discovery platform** that empowers the users to make faster, more informed decisions by accelerating analytics, revealing new business insights, and increasing the accuracy of results.

- It has been an intuitive software development kit that has been used in organizations around the world for many years. It can combine various kinds of data sources with visualizations in color-coded tables, bar charts, line graphs, pie charts, and sliders.

- It has been developed on a **"drag & drops" visualization interface,** allowing users to easily add data from many different sources, such as databases or spreadsheets, without having to write any code.

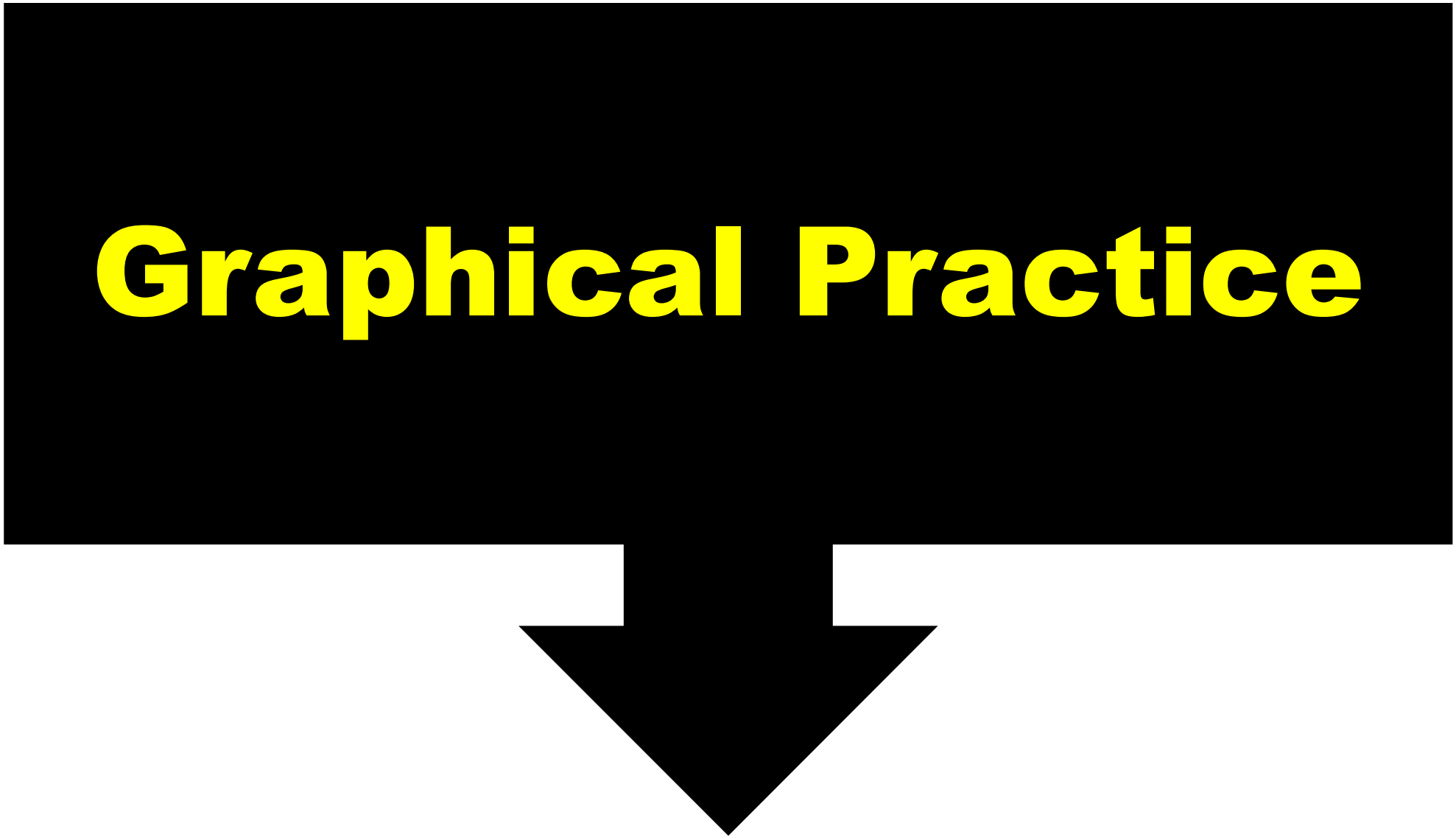
- Aforesaid characteristics also make it a relatively easier tool to learn and grasp.

Data Wrapper

(https://www.datawrapper.de/)

- It is an online data visualization tool that can be used in various contexts. It is very easy to use, and it has a **clean and intuitive user interface**.
- It allows users to create charts and maps directly in the browser by uploading their data files.
- The charts and maps created in Datawrapper are responsive and designed for all kinds of devices, so readers will be able to view them on any device that they are using.
- It's free for usage; but, there are certain limitations in the free version.

- More tools ….


Graphical Practice

Graphical Excellence

- **Graphical excellence** is well designed presentation of interesting data :
  - a matter of **substance**
  - a matter of **statistics**
  - a matter of **design**

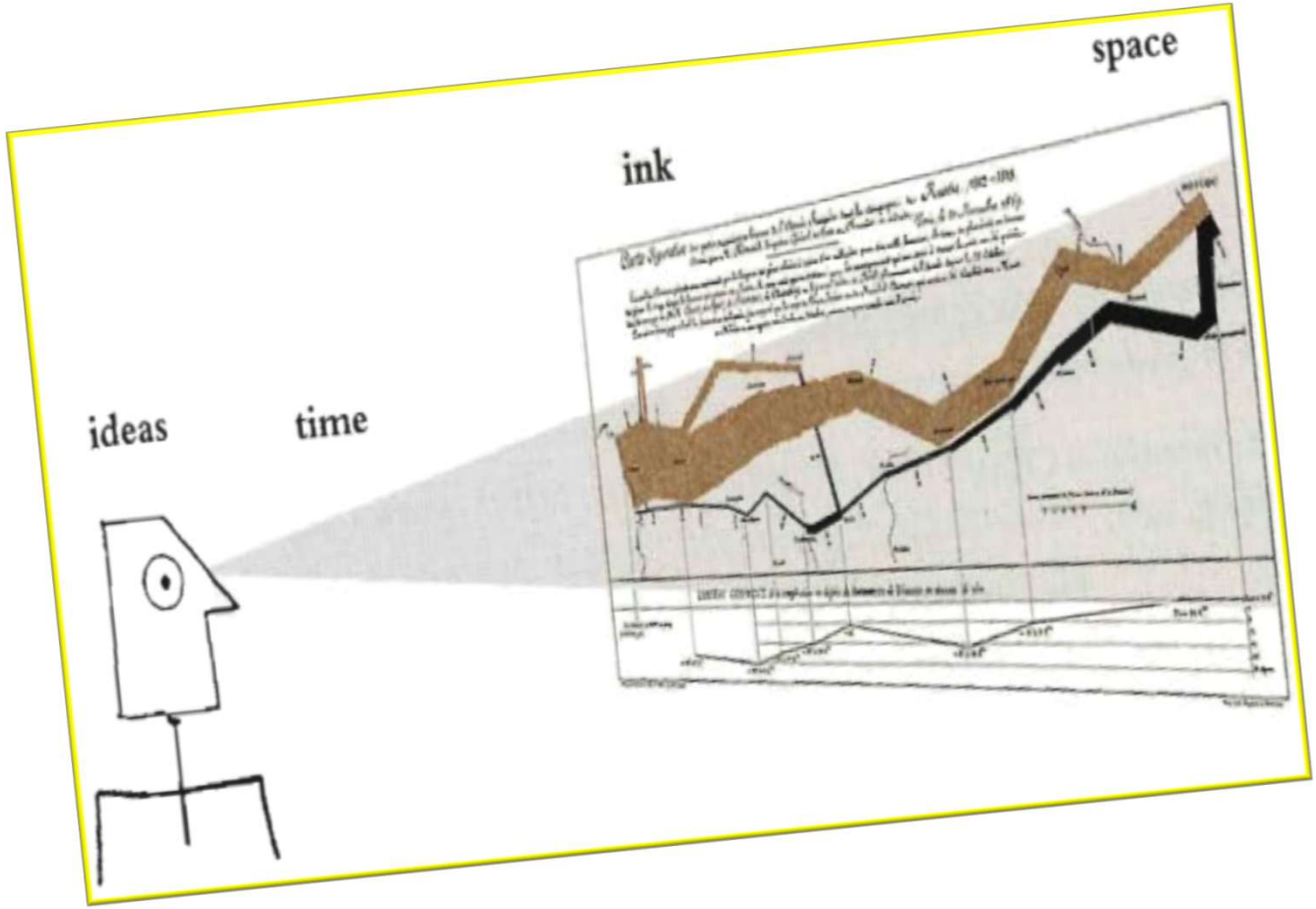
- Graphical Excellence consists of **complex ideas** **communicated** with **clarity**, **precision** & **efficiency**.

- Graphical excellence is something that delivers to it's viewers **greatest number of ideas** in the **shortest span of time** with the **least ink** in the **smallest space**.

Principles of Graphical Excellence

- Tufte's Principles of Graphical Excellence
  - Graphical excellence is
    - ✓the well-designed presentation of interesting data – a matter of substance, of statistics, and of design
    - ✓consists of complex ideas communicated with clarity, precision and efficiency
    - ✓is that which gives to the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space
    - ✓requires telling the truth about the data.

Graphical Excellence - Example



- **Graphical Excellence is almost always multivariate.**
- **Graphical Excellence requires telling the truth about data.**

Graphical Integrity

- Graphical Integrity – refers to how accurately visual elements represent data.

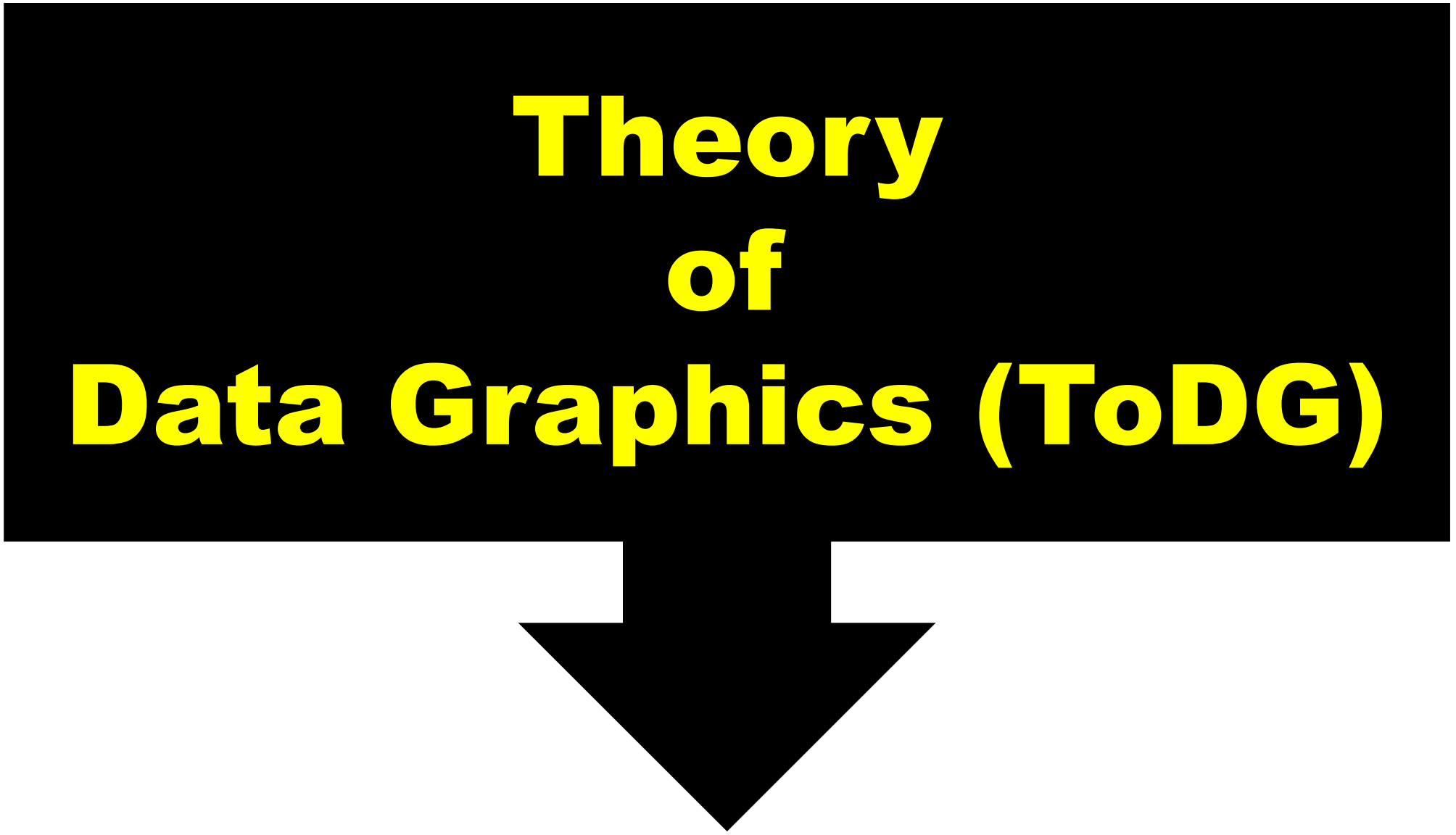
- Any graphical design – that gives false impression of the data and leads to incorrect conclusions, is forbidden by principles of Graphical Integrity.

Tufte's Graphical Integrity

- Some lapses intentional, some not
- Lie Factor = $\frac{\text{size of effect in graph}}{\text{size of effect in data}}$
- Misleading uses of area
- Misleading uses of perspective
- Leaving out important context
- Lack of taste and aesthetics

Graphical Integrity - Practices

- Numerical representations that match the true proportions.
- Clear and detailed labeling.
- Designs that do not vary for some ulterior motive, but show only data variation.
- The use of well-known units when representing money.
- Uniform number of dimensions represented and number of dimensions in the data.


Theory
of
Data Graphics (ToDG)

Data-ink & Graphical Re-design

- While choosing options to represent data within **charts** or **graphs**, it's important to do so **correctly** & **aesthetically** to produce the desired effect … which is **representing the data** above all else.

- **The data** must be the **most important item** on the page and **be portrayed accurately** with all relevant information available.

- When data is misrepresented in any way, whether through **inaccuracy** or **lack of qualifying information**, it becomes a useless tool.

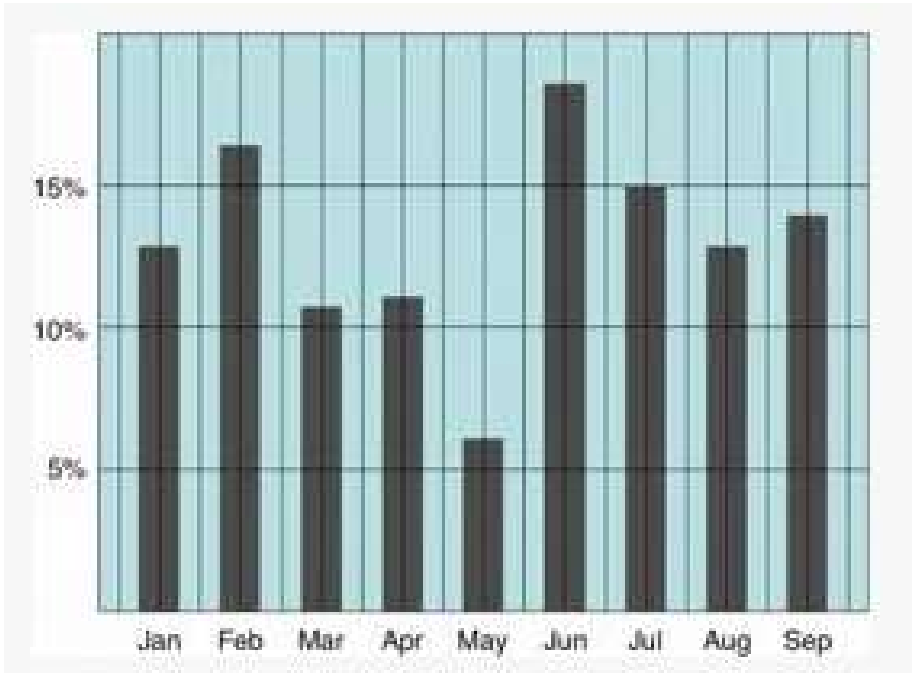
- Visual representations must be designed in such a way that **meaning of data** is abundantly clear and easily recognized.

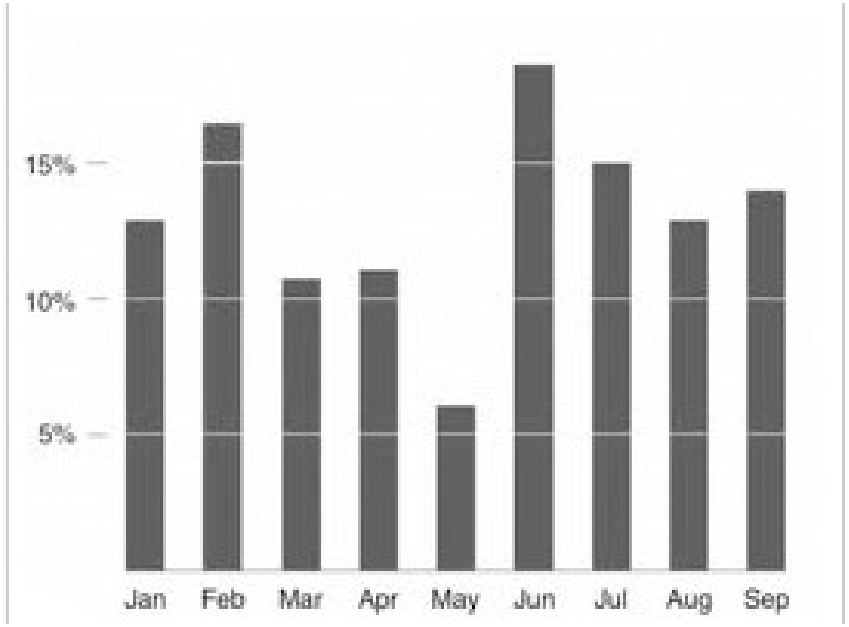
Data-ink Maximization & Graphical Design

- Importance of the **amount of ink** and **how it is used** to design charts & graphs for visual display of quantitative information in statistical representations is well documented.

- Using too much ink for information other than what's pertinent and relevant can diminish the data and make it less effective.

- It's important when viewing and analysing statistical information in a visual form to remember that the data is what's most important and should be the focal point.





Both graphics represent the same data, though one maximizes the ink used to convey the information while the other has vast amounts of wasted ink.

Source:
https://jasonhelmer.wordpress.com/

ToDG: Data-ink & Graphical Re-design

- Large share of ink on a graphic should present data-information, the ink changing as the data change (Tufte 2001, 2:93)
- **Data-ink**: **Non-erasable core** of a graphic, the **non-redundant ink** arranged in response to variation in the numbers represented (Tufte 2001, 2:93)

- $\text{Data-ink ratio} = \frac{\text{data-ink}}{\text{total ink used to print the graphic}}$ = proportion of a graphic's ink devoted to the non-redundant display of data-information = 1.0 - proportion of a graphic that can be erased without loss of data-information (Tufte 2001, 2:93)

- **Rule**: Maximize the data-ink ratio, within reason (Tufte 2001, 2:96)
- Two **Erasing Principles** (Tufte 2001, 2:96)
  - Erase non-data-ink, within reason
  - Erase redundant data-ink, within reason

Source: Tufte, Edward R. 2001. *The Visual Display of Quantitative Information*. Vol. 2. Graphics press Cheshire, CT.

Tufte's Principle on Data-ink Ratio

- **Maximize** the **data-ink ratio**:

$$\text{Data-ink ratio} = \frac{\text{data ink}}{\text{total ink used in graphic}}$$

- **Avoid "chart junk"**

Chart Junk (as per Edward Tufte)

- According to Tufte, chartjunk consists of non-data and redundant data elements in a graph.
- It comes in various types: sometimes artistic decoration, but more **often in the form of conventional graphical elements that are unnecessary in that they add no value**.
- Chartjunk is not informative and is often harmful.

Chart Junk (contd..)

- "**Chart junk**" refers to all visual elements in charts and graphs that are not necessary to comprehend the information represented on the graph, or that distract the viewer from this information.
- Examples of chart junk include:
  - heavy or dark grid lines
  - unnecessary text
  - inappropriately complex or gimmicky font faces
  - ornamented chart axes, and display frames
  - pictures, backgrounds or icons within data graphs
  - ornamental shading and unnecessary dimensions

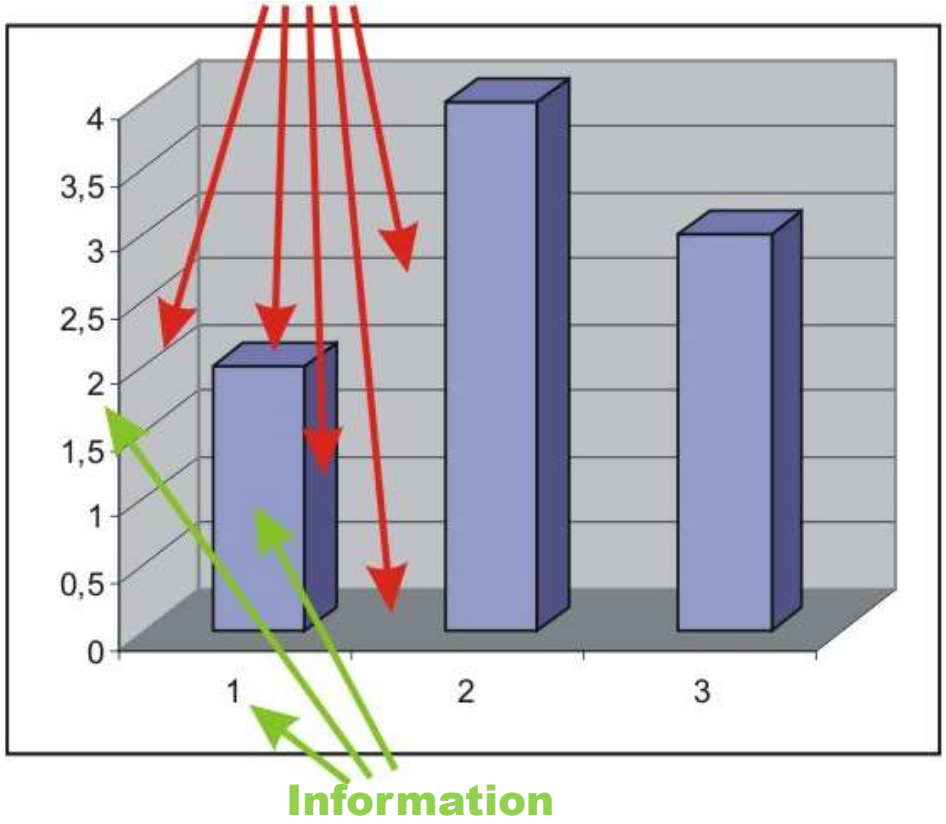
Alternative Definition of Chart Junk

- ***Chart junk** is any element of a chart that **does not contribute to clarifying the intended message.***

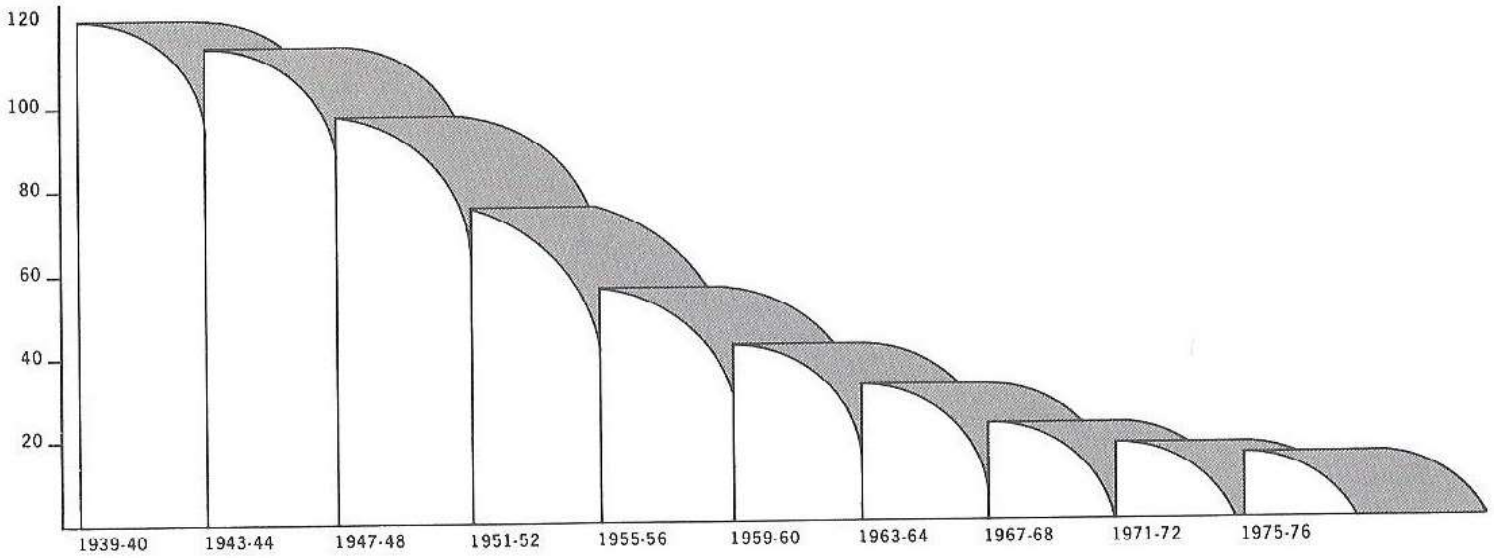
Chart Junk Examples

??

No extra information

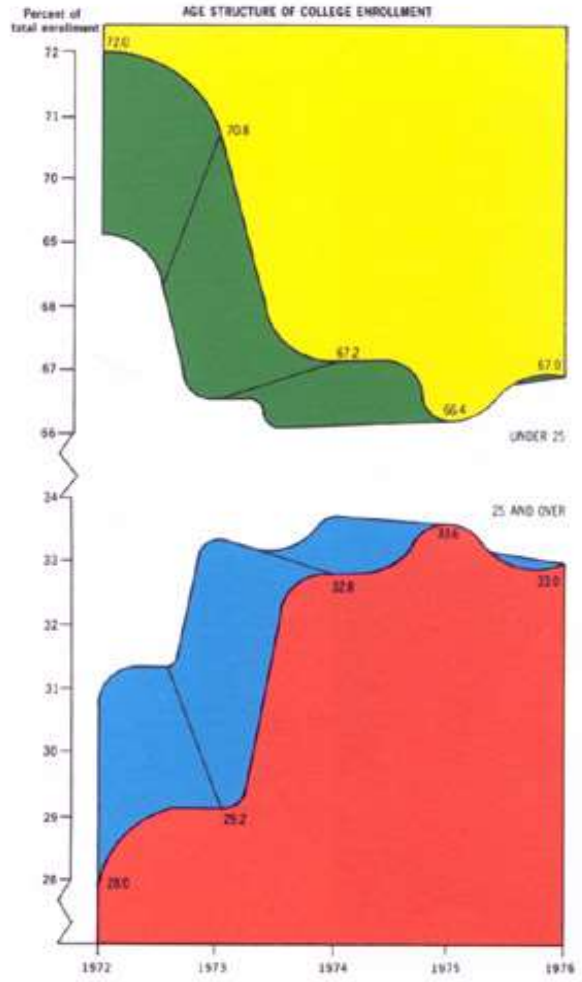


Information



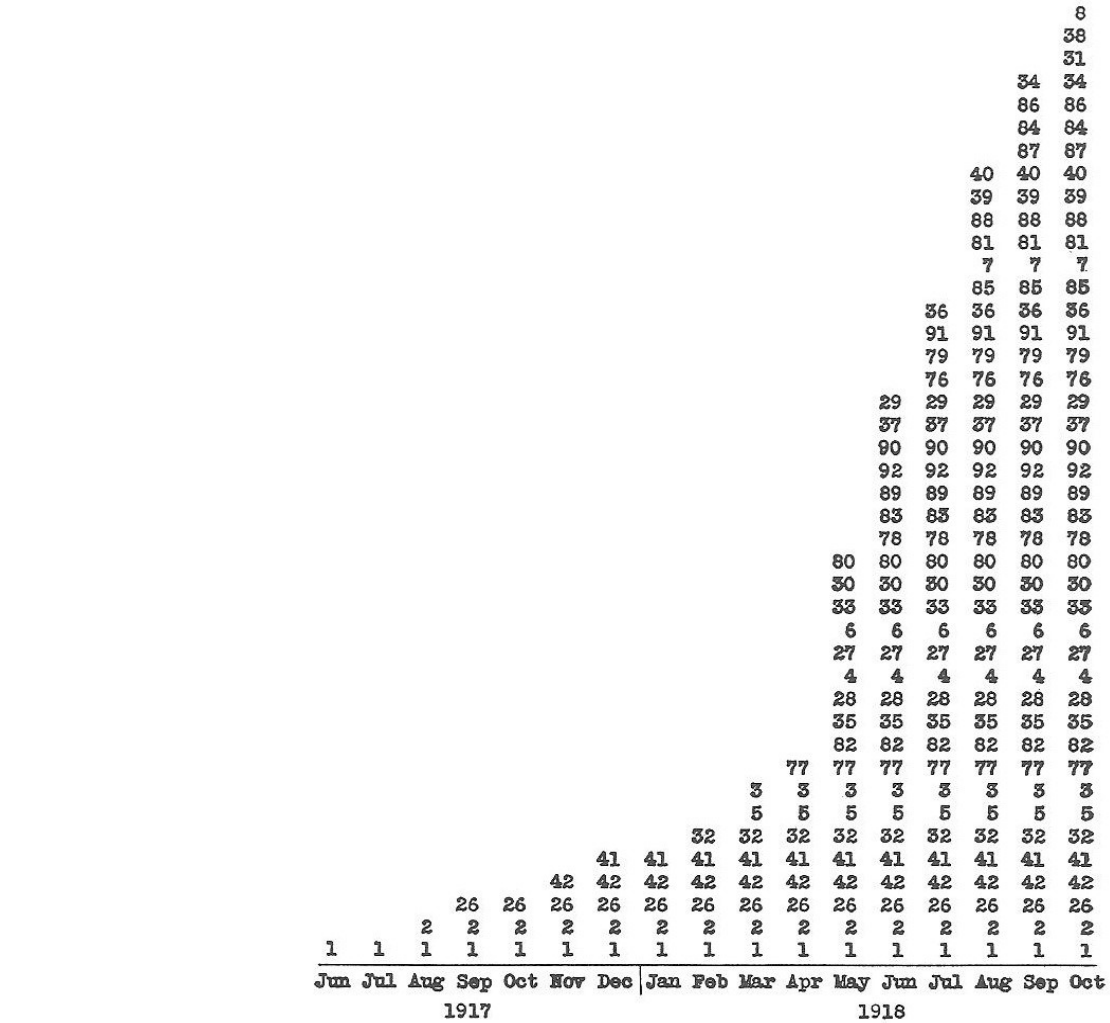
Source:
The Visual Display of Quantitative Information
by E. Tufte

Worst.Graph. Ever.

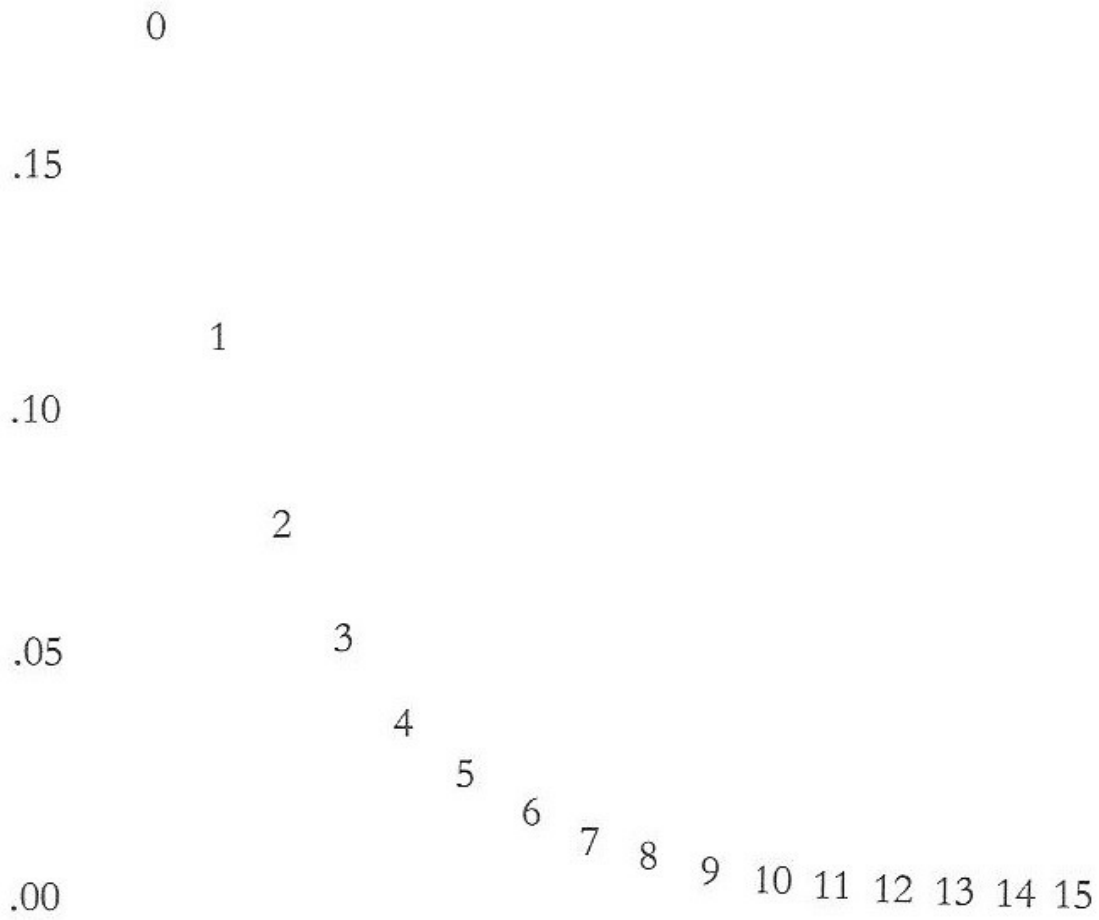


Source:
The Visual Display of Quantitative Information
by E. Tufte

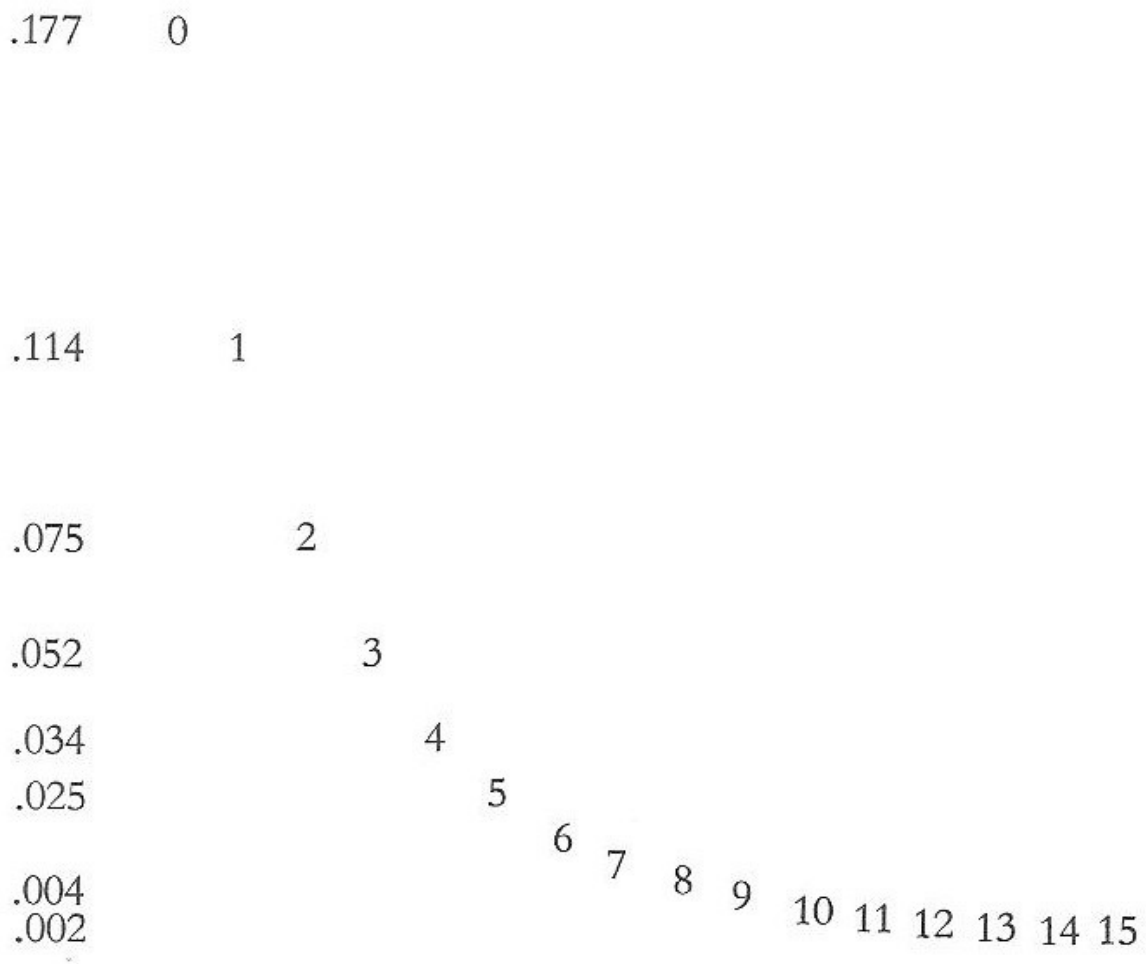
Multifunctioning
Graphical Elements



Source:
The Visual Display of Quantitative Information
by E. Tufte



Source:
The Visual Display of Quantitative Information
by E. Tufte



Source:
The Visual Display of Quantitative Information
by E. Tufte

Data Density
&
Small Multiples

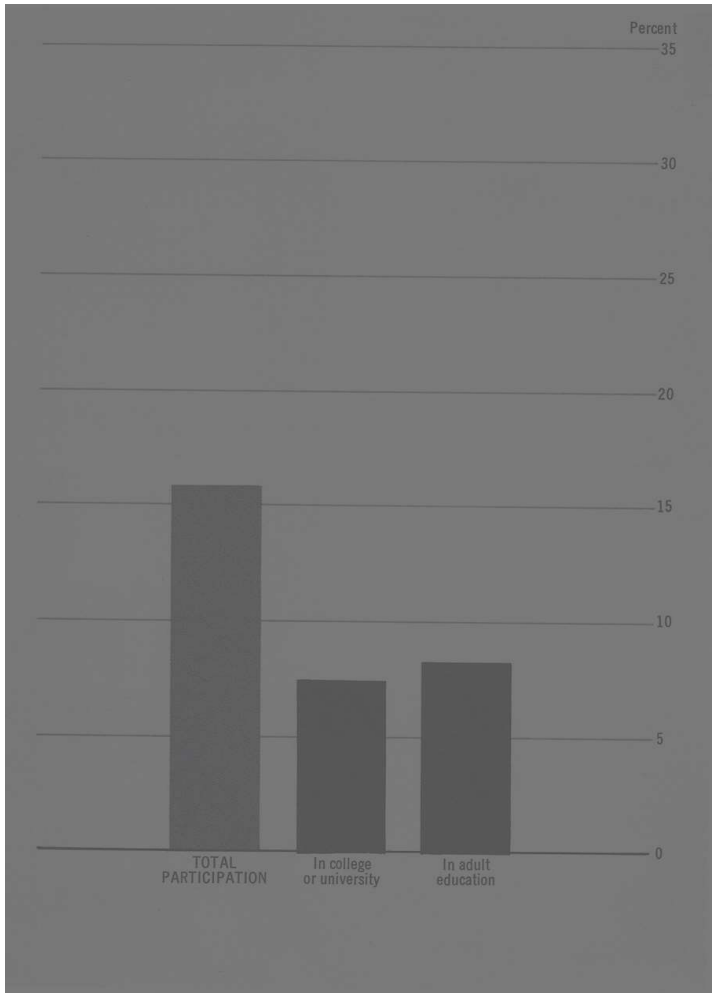
Data Density

Data Density

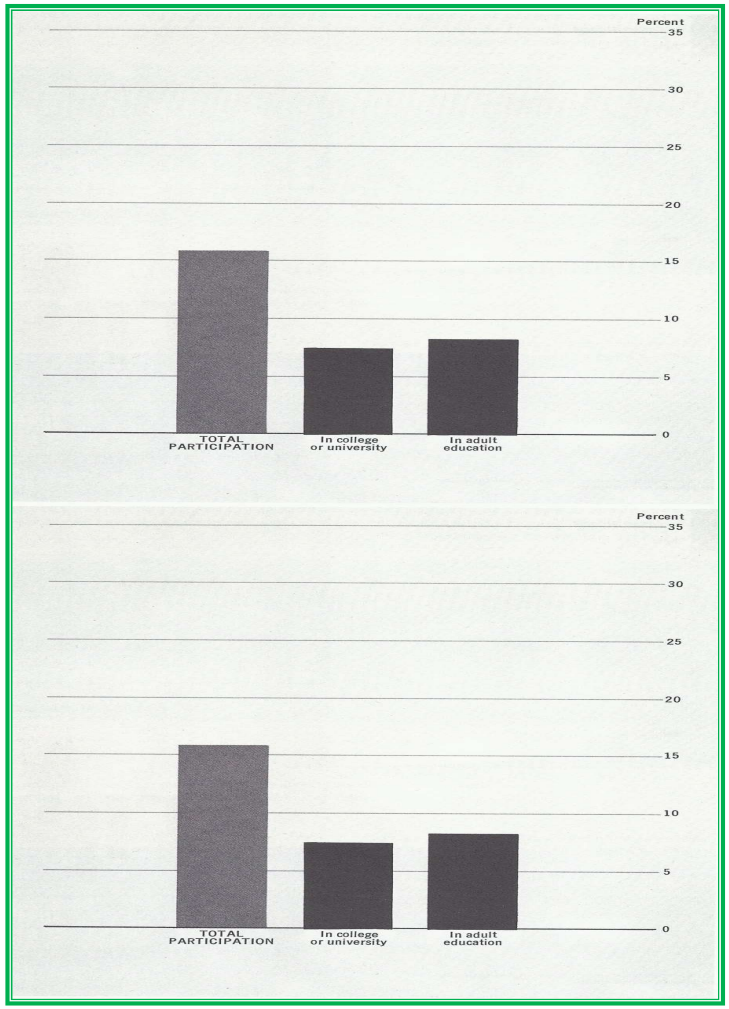
$$\frac{Number\ of\ entries\ in\ data\ matrix}{Area\ of\ Data\ Graphic}$$

Low Data Density

Number of entries = 4
Graph Area = 26.5 square inches
Data Density = $\frac{4\ data\ entries}{26.5\ sq.\ in.}$
=.15 data entries per sq. in.

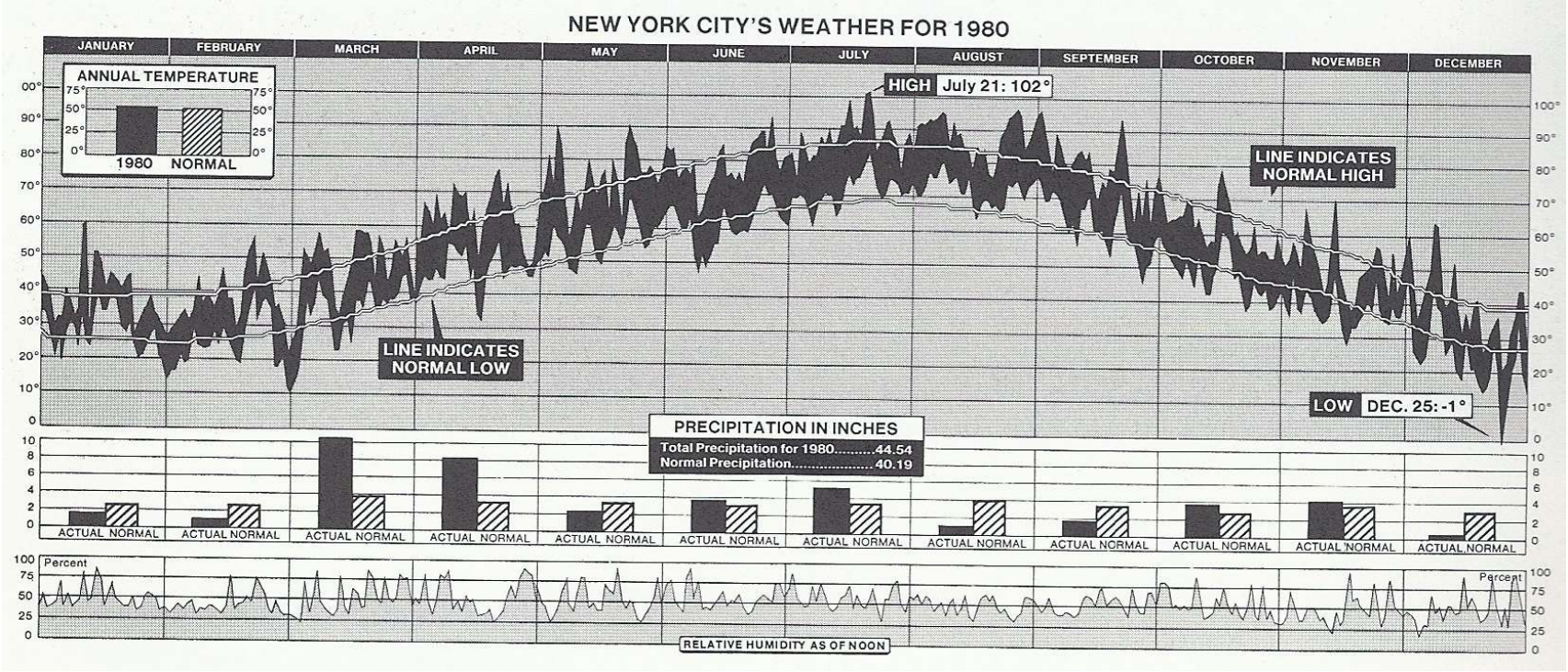


Low Data Density



Source:
The Visual Display of Quantitative Information
by E. Tufte

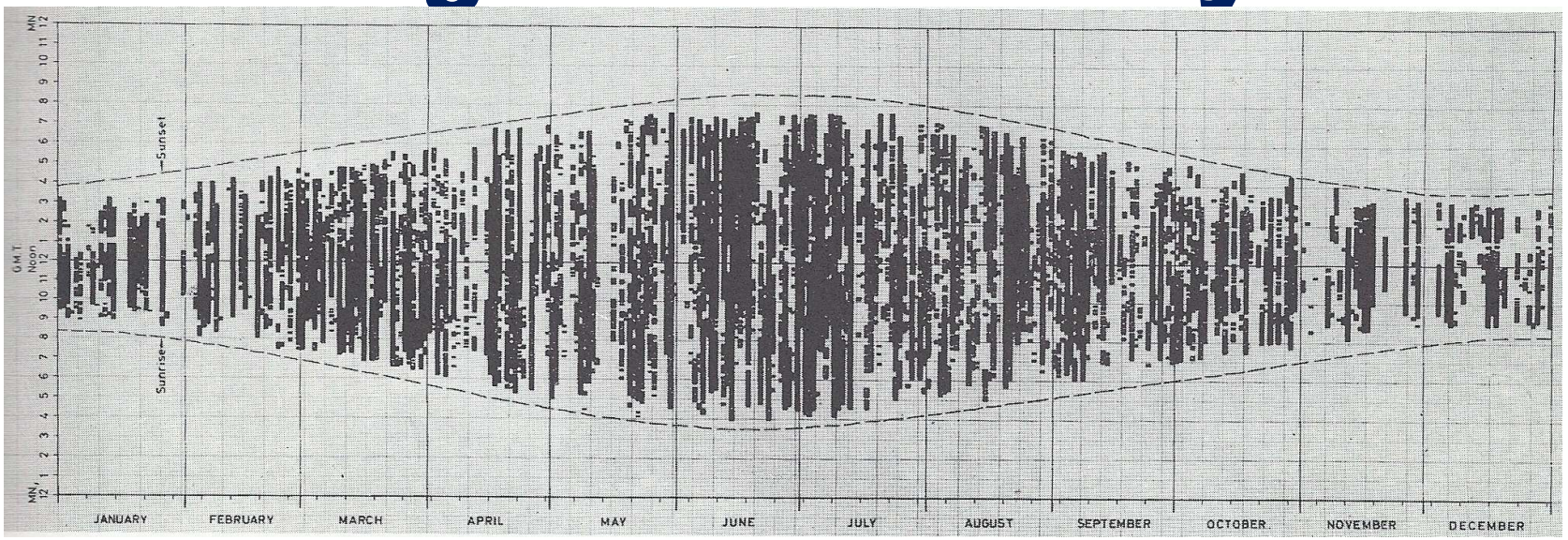
High Data Density



181 Numbers per square inch

Source:
The Visual Display of Quantitative Information
by E. Tufte

High Data Density

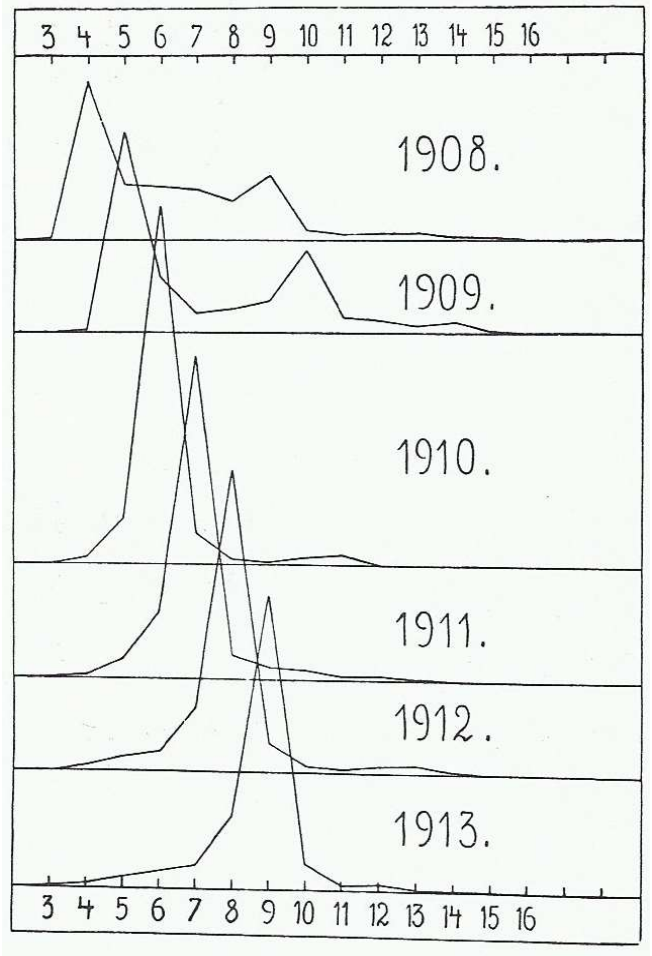


1,000 Numbers per square inch

Source:
The Visual Display of Quantitative Information
by E. Tufte

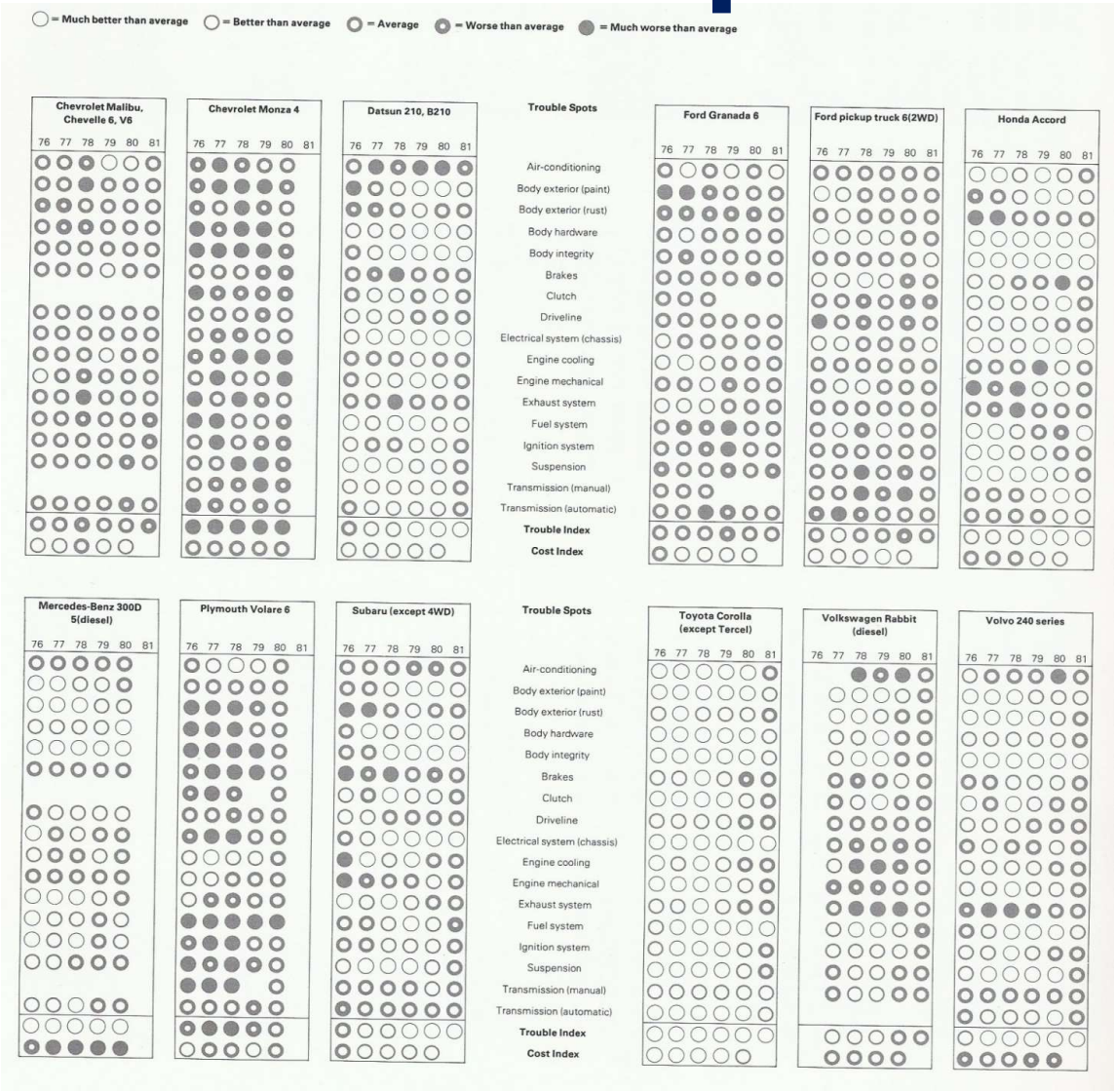
Small Multiples

Small Multiples



Source:
The Visual Display of Quantitative Information
by E. Tufte

Small Multiples

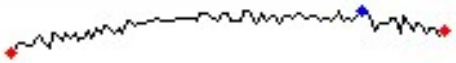
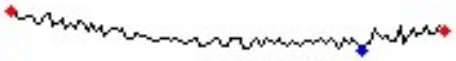
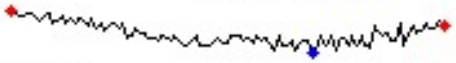
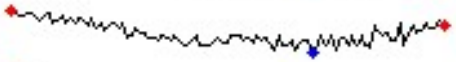
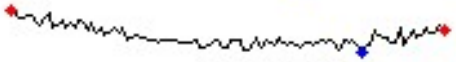
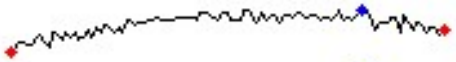
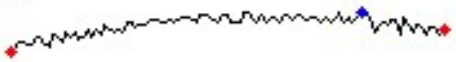


Source:
Consumer Reports

Tufte’s Graphs

Sparkline
Slope Graph

Sparklines

	0	GDP Per Capita	100	min	max
Arabia	-3.8097		-1.5757	-3.8097	0.6381
Brazil	3.8097		1.5757	-0.6381	3.8097
China	1.9518		1.2553	0.11066	1.9518
Denmark	1.9518		1.2553	0.11066	1.9518
England	3.8097		1.5757	-0.6381	3.8097
France	-3.8097		-1.5757	-3.8097	0.6381
Gibraltar	-3.8097		-1.5757	-3.8097	0.6381

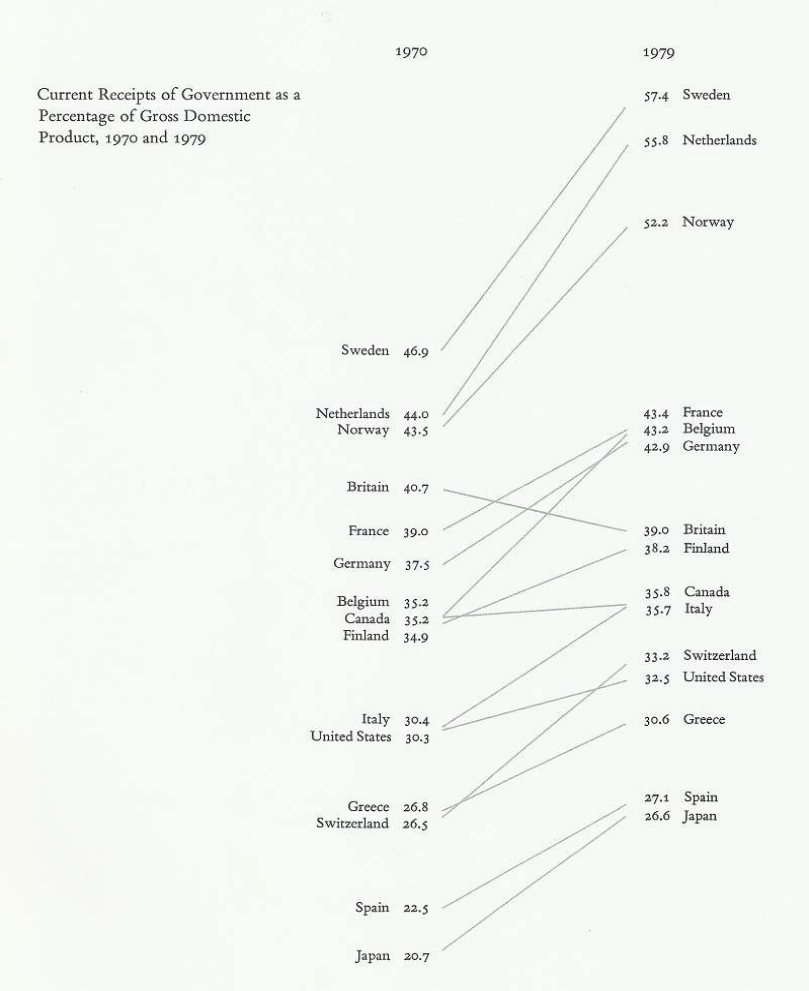
Source:
The Visual Display of Quantitative Information
by E. Tufte

Sparklines

Source:
The Visual Display of Quantitative Information
by E. Tufte

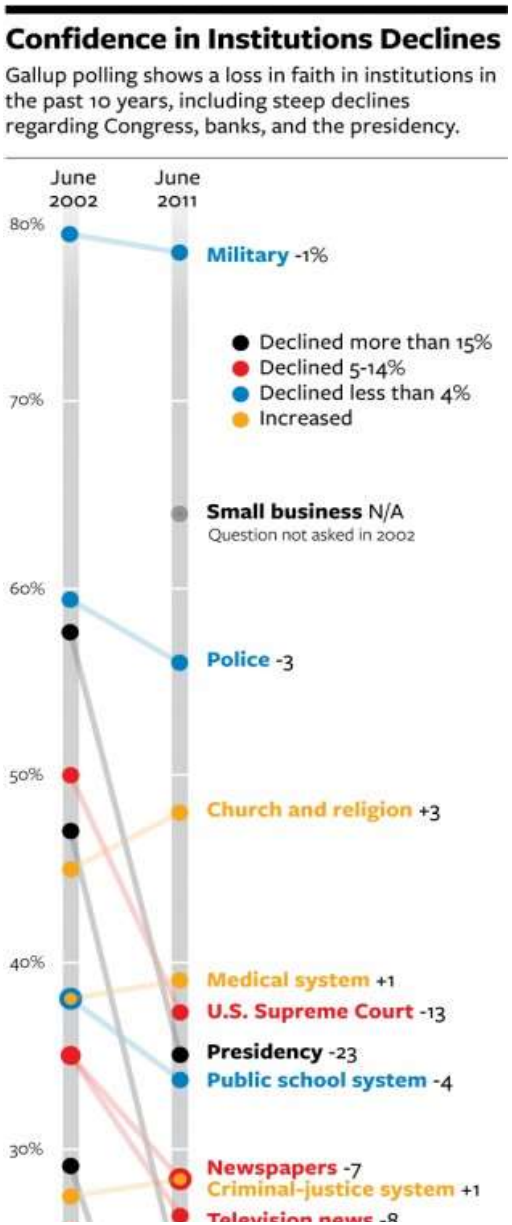
Source:
The Visual Display of Quantitative Information
by E. Tufte

Slope Graph



Source:
The Visual Display of Quantitative Information
by E. Tufte

Slope Graph



Source: The Atlantic, June 30, 2012

Takeaways

- Maintain Graphical Integrity
- Maximize Data-Ink Ratio, within reason
- Avoid Chartjunk and Ducks
- Use Multifunctioning Graphical Elements (if possible)
- Keep Labels with data
- Maximize Data Density

Aesthetics
&
Techniques in Graphical Design

References & Acknowledgements

- Edward R. Tufte (2001): The Visual Display of Quantitative Information
- https://infovis-wiki.net/
- https://jasonhelmer.wordpress.com/
- https://eagereyes.org