

Graphical Excellence

(Professor Dave)²

Outline

The Visual Display of Quantitative Information provides a rich story of graph types and excellence including:

- What is excellence?
- Data maps
- Time series
- Space-time narrative designs
- Relational graphics
- Some history

How is excellence achieved?

When complex ideas are expressed with:

- Clarity
- Precision
- Efficiency

How is excellence achieved?

When complex ideas are expressed with:

- Clarity
- Precision
- Efficiency

How do we get there?

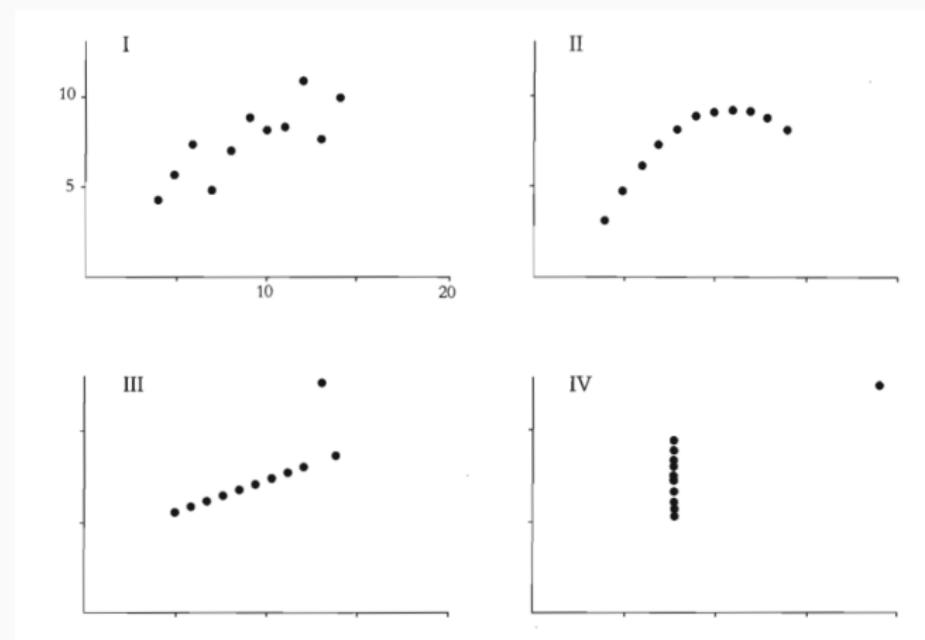
- **Do:** Show the data clearly, aim to have a quick understanding of the data trend, facilitate comparisons.
- **Don't:** Distort the data, compress data in small spaces, show data at different scales.

Why visualize? Anscombe's Quartet

I		II		III		IV	
X	Y	X	Y	X	Y	X	Y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

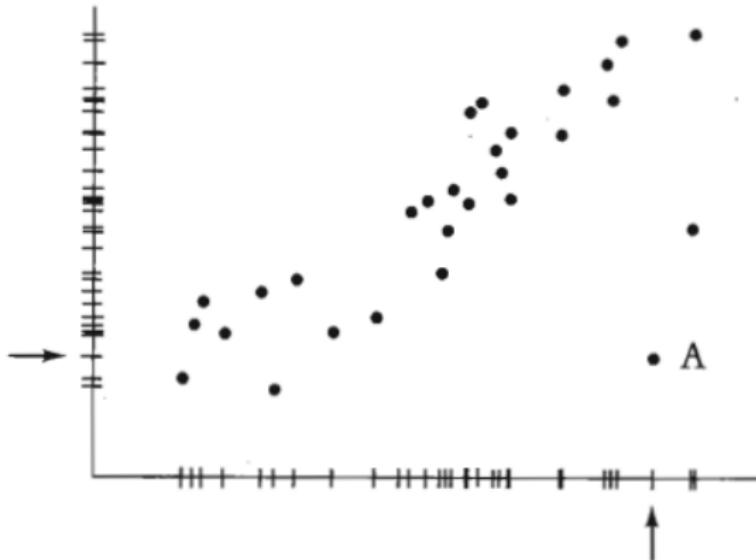
N = 11
mean of X's = 9.0
mean of Y's = 7.5
equation of regression line: $Y = 3 + 0.5X$
standard error of estimate of slope = 0.118
 $t = 4.24$
sum of squares $X - \bar{X} = 110.0$
regression sum of squares = 27.50
residual sum of squares of Y = 13.75
correlation coefficient = .82
 $r^2 = .67$

Why visualize? Anscombe's Quartet



Simple statistics fail to exhibit the data trend! Remember to always visualize!

Why visualize? Anomaly detection



Showing the data clearly reveals anomalies.

Data maps

In highest decile,
statistically significant



Significantly high, but
not in highest decile



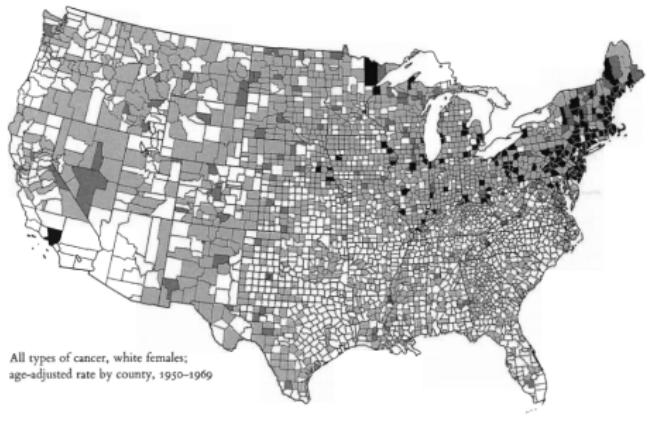
In highest decile, but not
statistically significant



Not significantly different
from U.S. as a whole



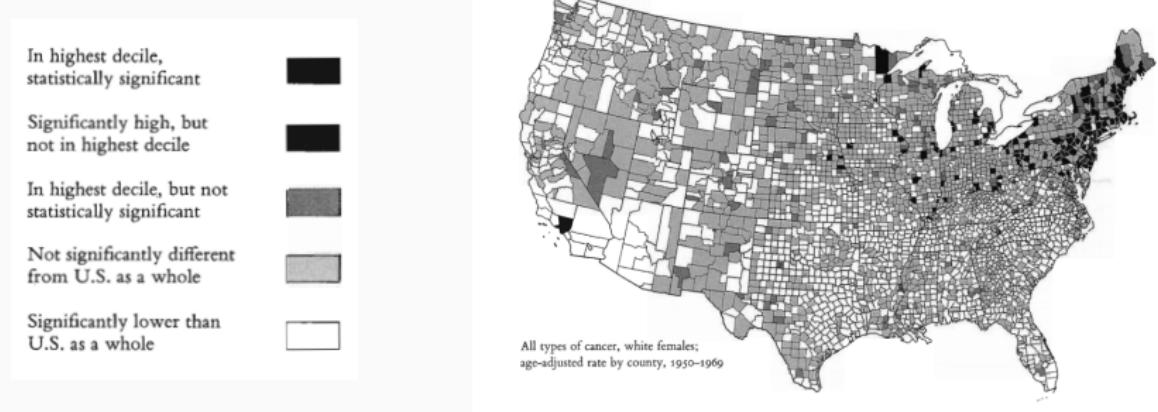
Significantly lower than
U.S. as a whole



Each map conveys a large amount of information in a small space.

They invite us to search for patterns and to explain their nature.

Data maps



High death rates in the Northeast region and Great Lakes.

Low death rates across central and south bands.

Data maps

In highest decile,
statistically significant



Significantly high, but
not in highest decile



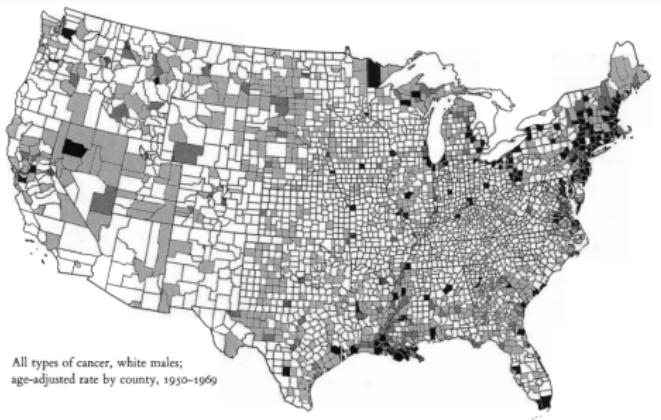
In highest decile, but not
statistically significant



Not significantly different
from U.S. as a whole



Significantly lower than
U.S. as a whole

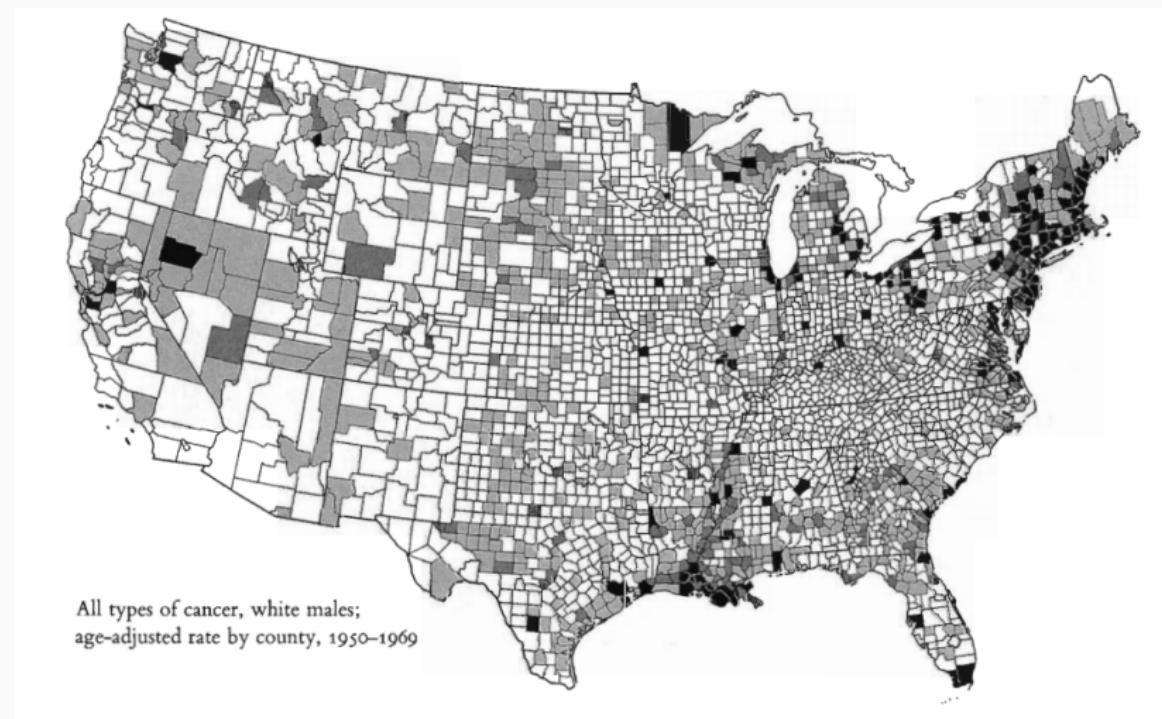


We are now looking at **white males** ...

Observations?

Explanations?

Data maps



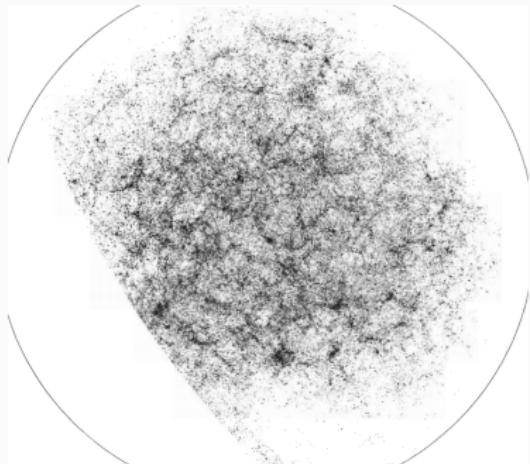
High death rates in the Louisiana area. **Asbestos exposure**

Data maps

Shortcomings:

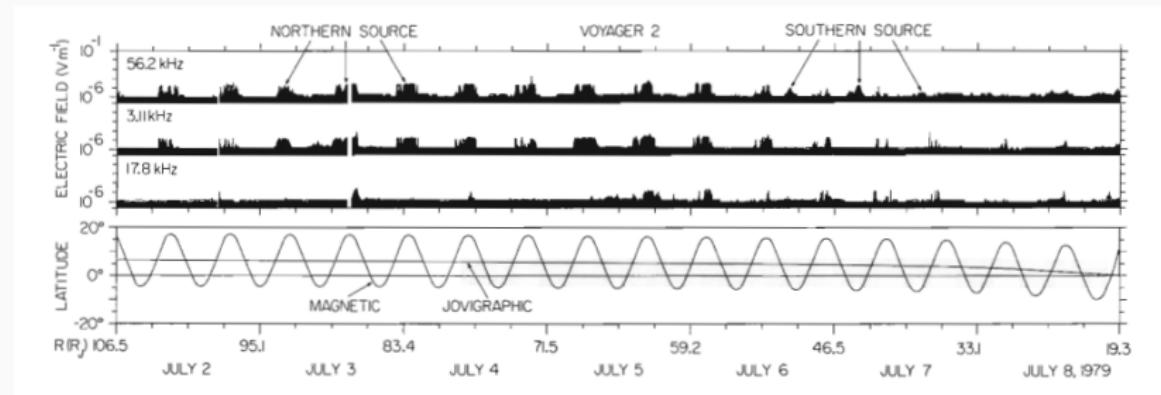
- The population of each county is not considered (only area).
- Changes across counties are abrupt (not smooth).
- How reliable is the data? (the diagnosis may be biased).

Modern data maps



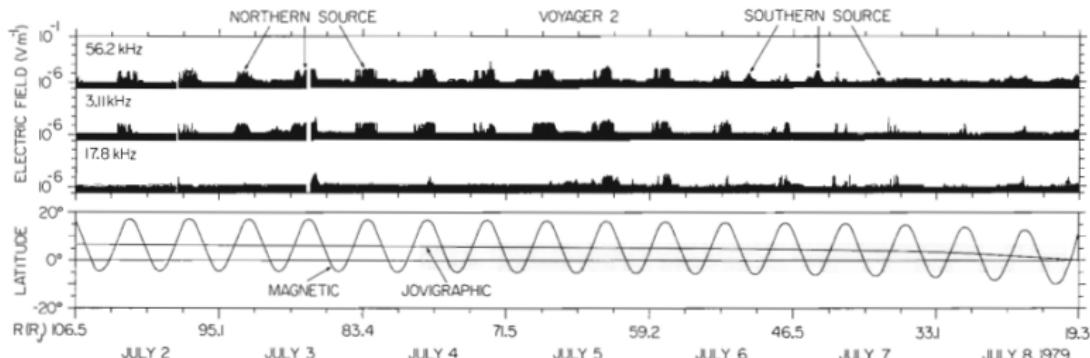
- Computerized cartography of the universe
- Distribution of 1.3 million galaxies in the Northern Galactic hemisphere.
- The darker the gray tone, the more galaxies in that region.

Time series



- This is the most frequent type of graph design.
- The x-axis contains time in one of many possible units (seconds, minutes, hours, days, months, years, etc.).

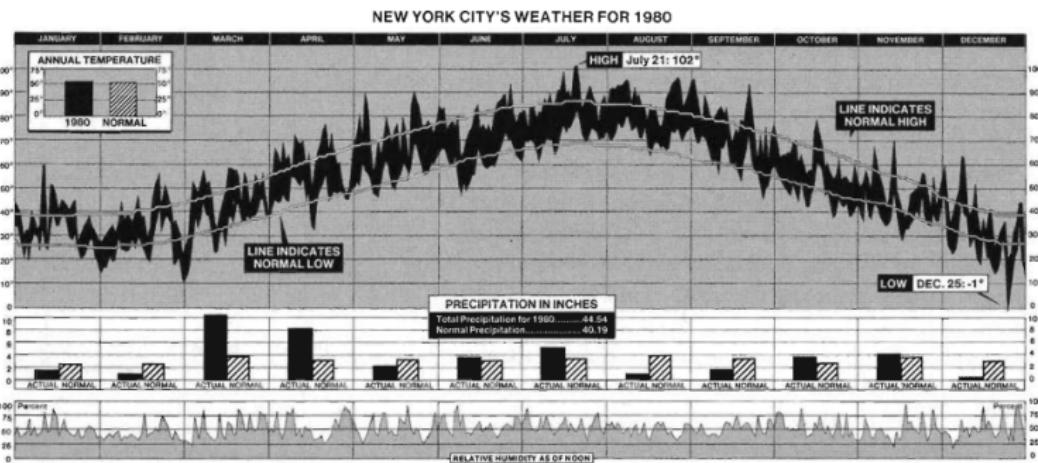
Time series



- Radio emissions from Jupiter captured by Voyager 2 (1979)
- x-axis is both time and distance from Jupiter
- Top three panels are different radio bands
- Bottom panel is orientation of spacecraft

Time series

New York City weather summary (1980)

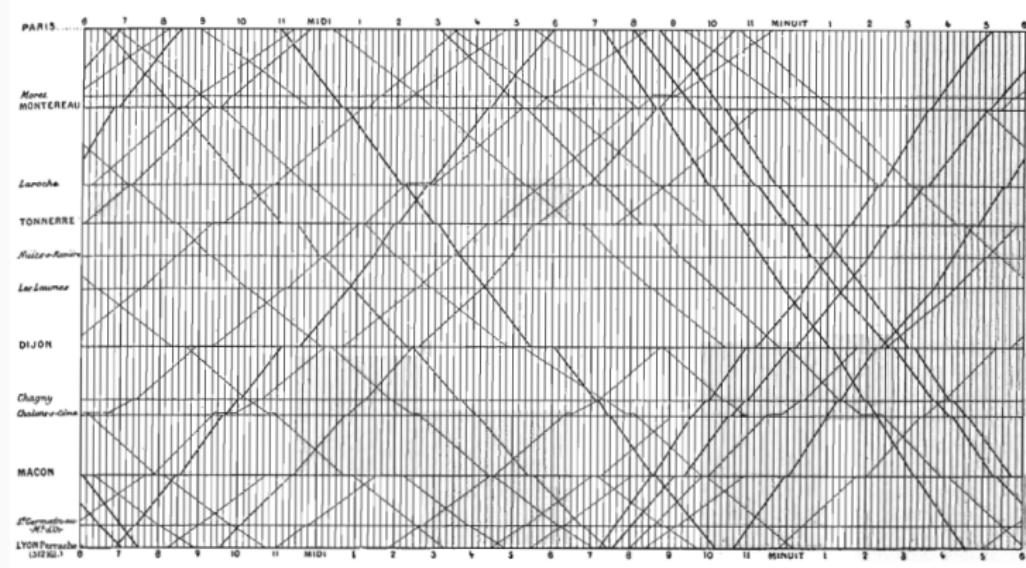


How much can you learn from the graph?

The graph tells a story!

Time series

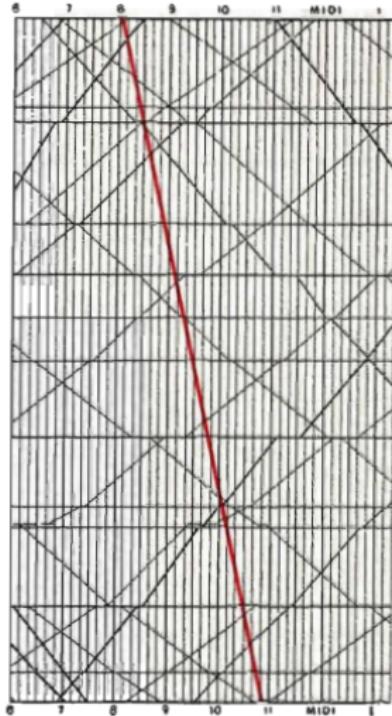
Graphical train schedule for Paris to Lyon (1880s).



y-axis: Arrivals/departures, top: Paris, bottom: Lyon.

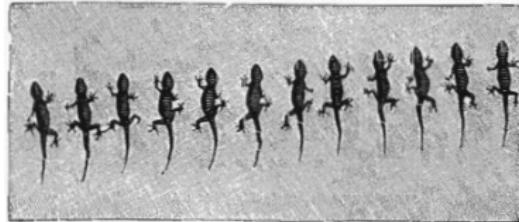
x-axis: time! Slope of line corresponds to the speed of the train.

Time series

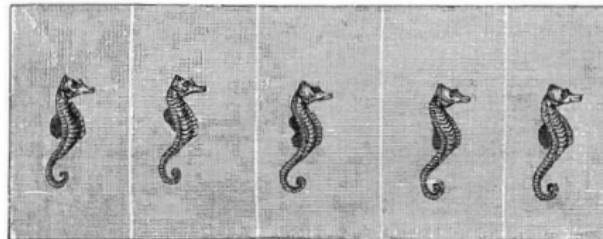


Extremely fast train came in 1981.

Time series



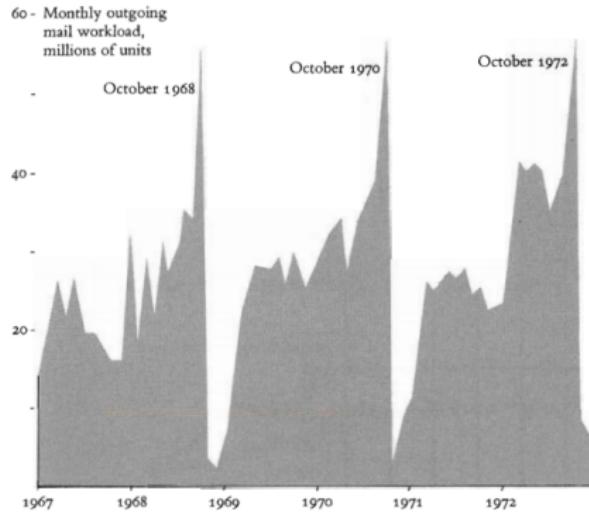
as well as the advance of the gecko.



the undulations of the dorsal fin of a descending sea-horse,

Time series can focus on the graphical evolution of body parts.

Time series

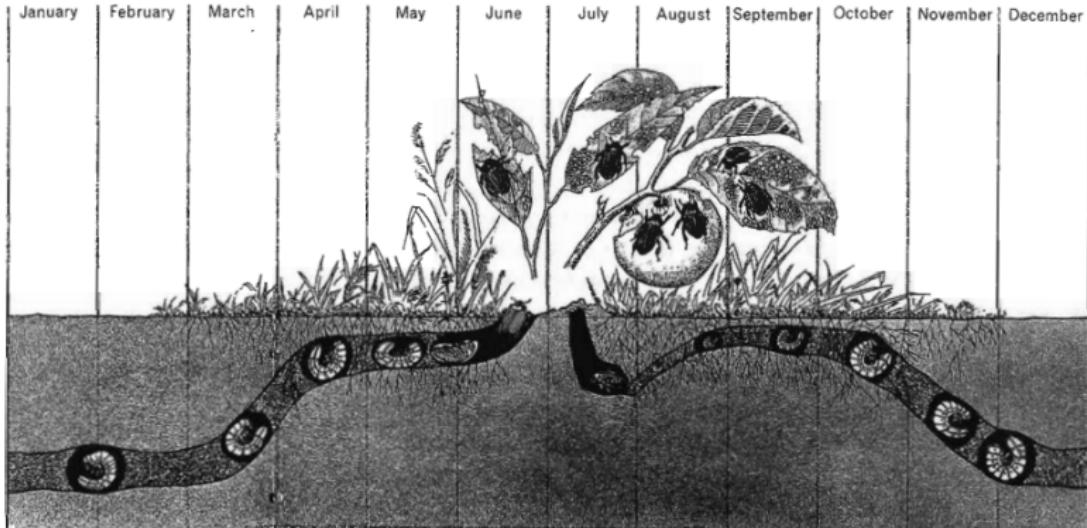


Causal information is encoded here ... how?

Outgoing mail (millions) by incumbent representatives.

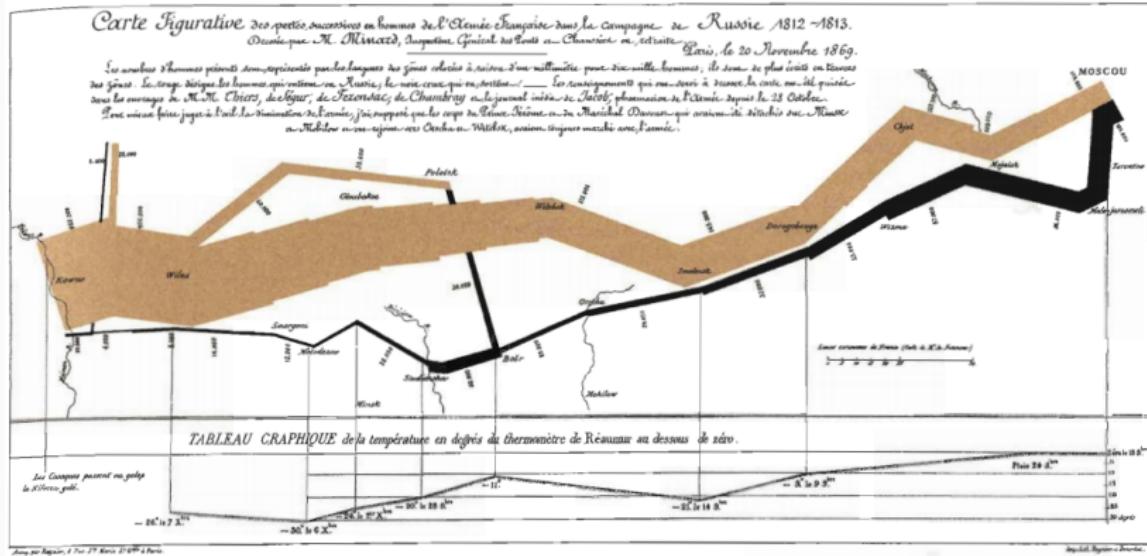
Representatives use the privilege of free mail to send many letters during re-election campaigns.

Space-time narrative designs

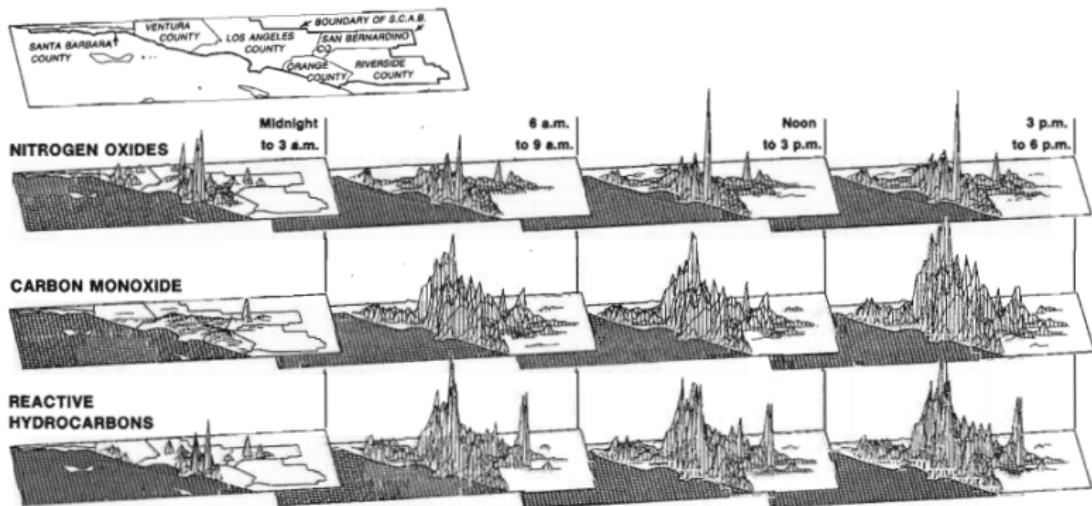


- Time-series graphs can be enhanced by adding spatial dimensions.
- This adds multivariate complexity that should be easy to visualize and interpret.
- Example: Life-cycle of the Japanese beetle.

Space-time narrative designs

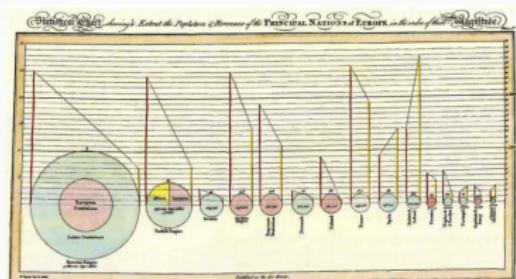


Space-time narrative designs



One more example is the levels of air pollutants in southern California during the day.

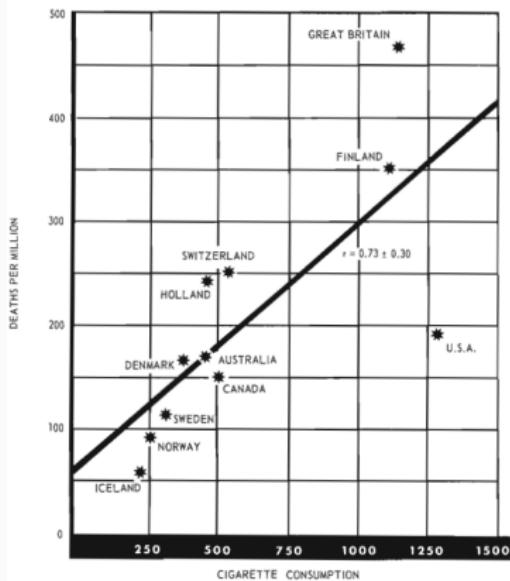
Slide 27



- Relational Graphics
- It is similar to previous graphs, but the dimensions can differ from spatial coordinates.
- The word "relational" refers to the dependence of two variables (any variables).
- Example: Each circle refers to a country.
- The size of the circle correlates with the area of the country.
 - The line on the left is the population in millions.
 - The line on the right is the taxes collected in millions of pounds.
- What is important is the sign of

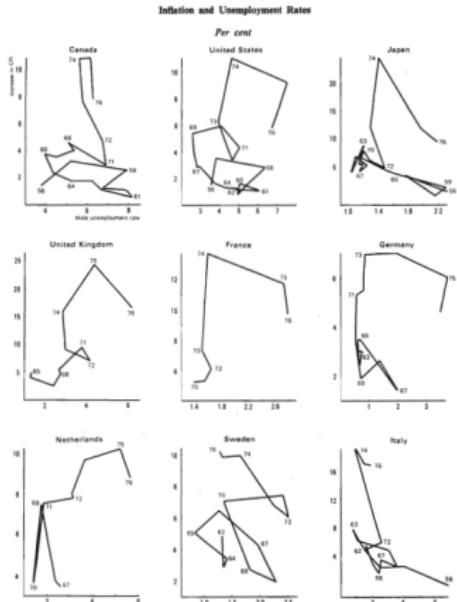
Slide 28

CRUDE MALE DEATH RATE FOR LUNG CANCER
IN 1950 AND PER CAPITA CONSUMPTION OF
CIGARETTES IN 1930 IN VARIOUS COUNTRIES.



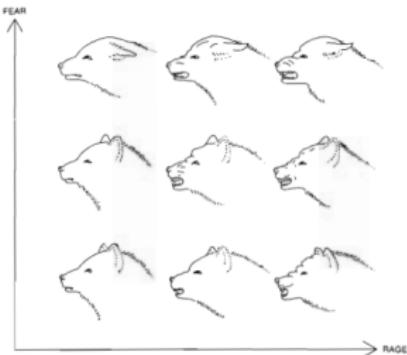
- Relational Graphics
- Relational graphs are the greatest of graphical designs.
- They help look for causal information.
- Example: The relationship of lung cancer and smoking.

Slide 29



- Relational Graphics
- Relational graphs are the greatest of graphical designs.
- They help look for causal information.
- Is there a connection between inflation and unemployment?
- Apparently NOT!

Slide 30



- Relational Graphics
- Graphs can be very informative.
- Data points can convey a lot of information.
- Rage (x-axis) vs Fear (y-axis).

Slide 31

Most of the material comes from Chapter 1 of *The Visual Display of Quantitative Information* by Edward R. Tufte, 2nd Edition, Graphics Press, 2007.

- Graphical Excellence
- Data Maps
- Time Series
- Space-Time Narrative Designs
- Relational Graphics
- History
- Visual Displays
- Graphical Excellence

Slide 32

History

First, data maps were created thousands of years ago.

A significant map of the eleventh century was found in China.

In the sixteenth century, maps began to show similarities to modern maps.

- 5,000 years back: Geographic maps in clay tablets
- 11th century: Detailed map in China
- 16th century: Maps look similar to modern graphs in Europe
- 1750-1800: Graphics, time series, scatter plots, multivariate displays by William Playfair
- 1686 Edmond Halley: Map of trade winds and monsoons on a world map.



History
Edmond Halley's 1686 chart.

Slide 34



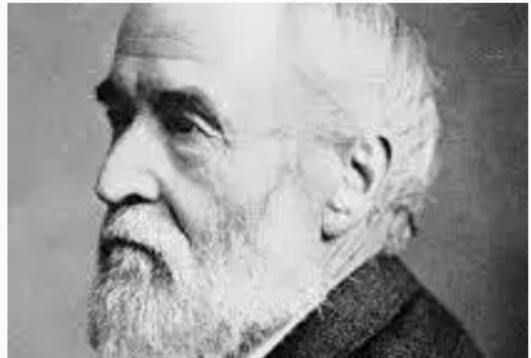
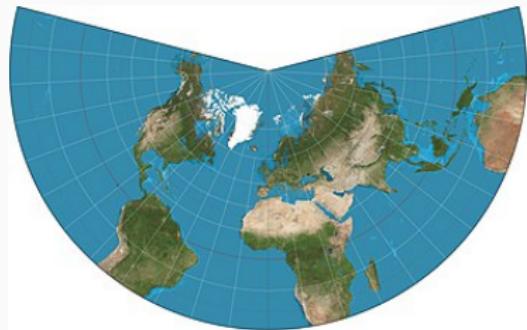
History

1854

Location of Cholera deaths in Central London by Dr. John Snow.

- Pump in Broad Street at the center of the outbreak.
- The handle of the pump was removed, and the disease did not spread any more.

Slide 35

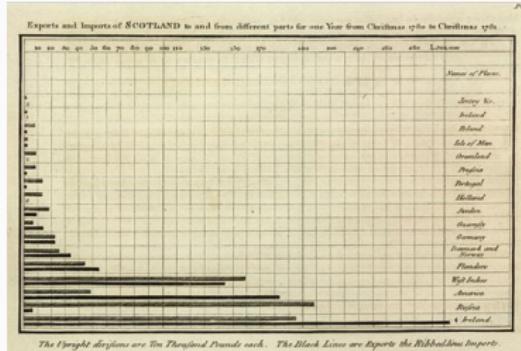


History

Great inventors of modern graphical design.

- Johan Heinrich Lambert:
Swiss-German scientist and
mathematician (1728-1777)
- Pioneered the idea of map
projections.

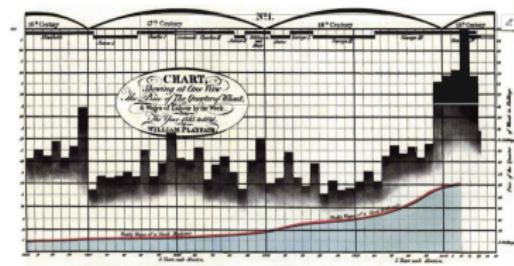
Slide 36



History

Great inventors of modern graphical design.

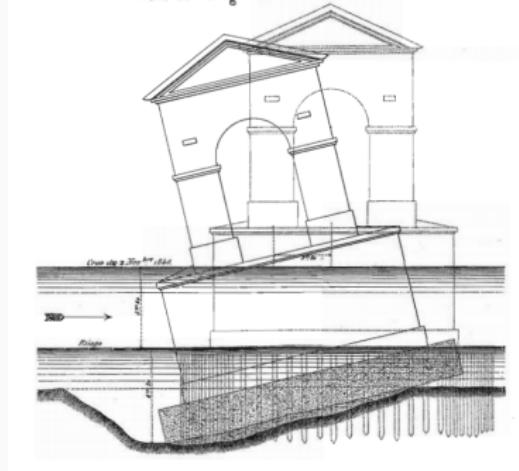
- William Playfair: Scottish political economist (1759-1823)
- Published the first time series economic data and created many types of charts, including the bar chart.
- Diverse professional



History

William Playfair showed in this graph prices, wages, and reigns of the British monarchy.

The novelty lies in combining all that information in one graph.



Summary - Conclusions

- Good graphs succinctly convey substantial information.
- Coming up with a good design is challenging.
- Ideas must be conveyed with clarity, precision, and efficiency.