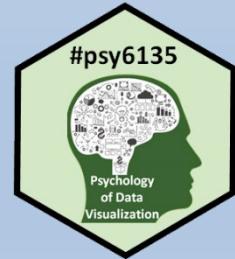
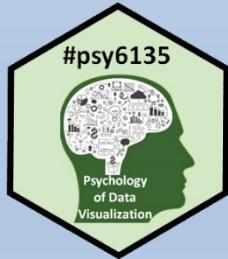


# History of Data Visualization

Michael Friendly  
Psych 6135



<https://friendly.github.io/6135>



# Outline



- Overview:
  - Roles of graphics in scientific discovery
  - Visualizing history: The *Milestones Project*
- Milestones tour of the history of data vis
  - Pre-history of visualization
  - The first statistical graph
  - The Big Bang: William Playfair
  - Influence of data, technology & visual thinking
- Other topics (later):
  - Moral statistics: the birth of social science
  - Graphs in the public interest: Nightingale, Farr and Snow
  - The Golden Age of statistical graphics

# Orienting questions

History in  
context



What  
motivated  
graphical  
inventions?

What was the  
communication  
goal?

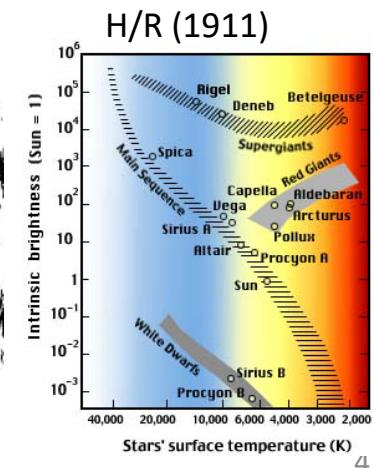
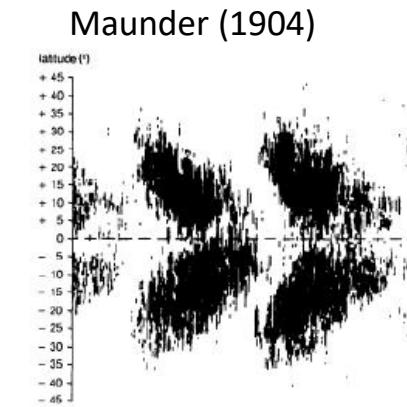
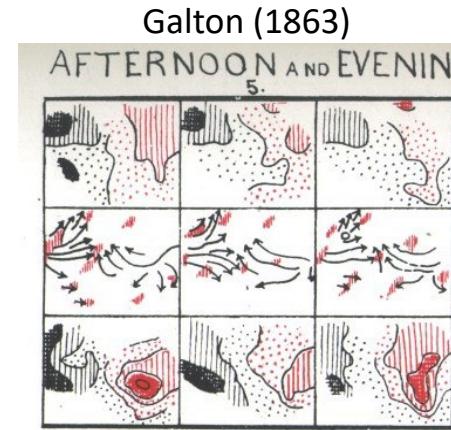
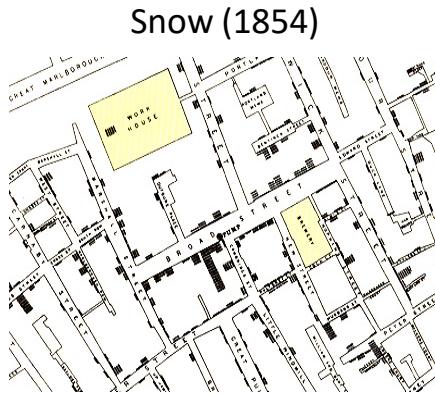
How does it  
relate to other  
developments?

- What were the  
pre-cursors?

How has this  
idea been  
used or re-  
invented  
today?

# Orienting Q: Visualization-based discoveries ??

- When have graphics led to discoveries that might not have been achieved otherwise?
  - Snow (1854): cholera as a water-borne disease
  - Galton (1883): anti-cyclonic weather patterns
  - E.W. Maunder (1904): 11-year sunspot cycle
  - Hertzsprung/Russell (1911): spectral classes of stars

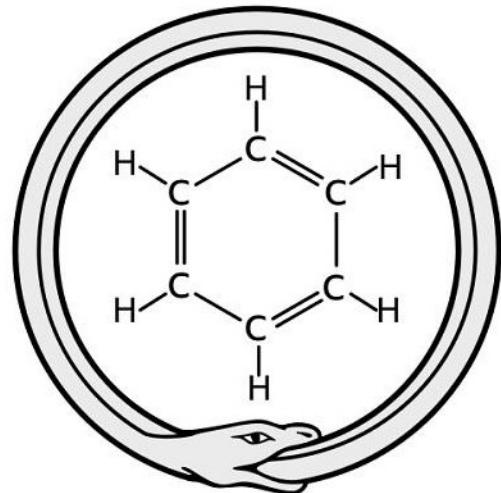


# Orienting Q: Visualization-based discoveries ??

- In the history of graphs, what features, and data led to such discoveries?
  - What were they thinking??
  - What visual ideas/representations were available?
  - What was needed to see/understand something new?
- As we go forward, are there any lessons?
  - What are the Big Questions for today?
  - How can data visualization help?

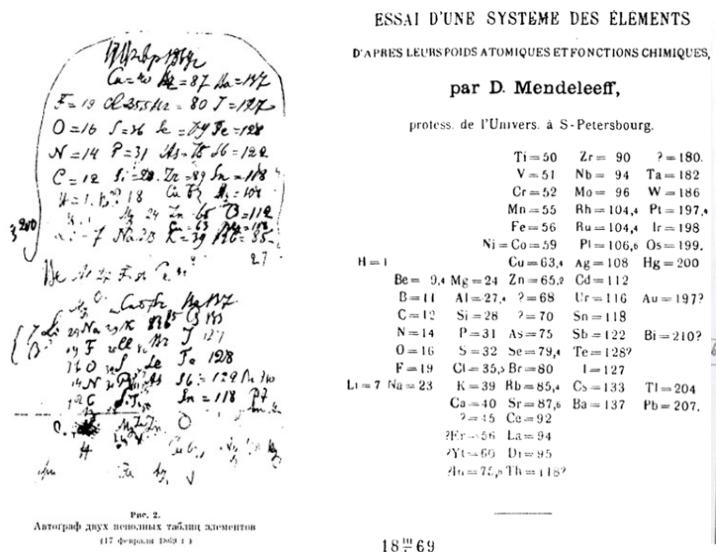
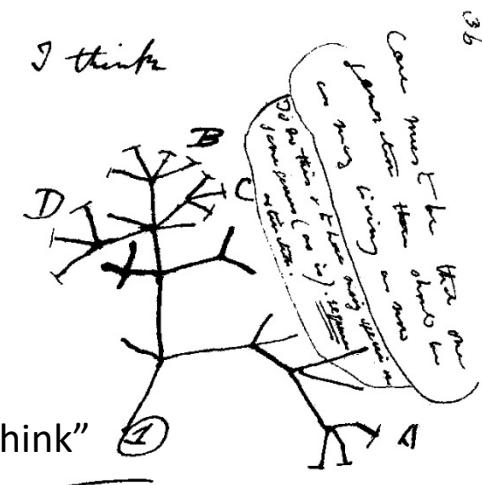


# Visual thinking & scientific discovery



**Dreams and snakes**  
August Kekulé (1862)  
discovers the structure of  
benzene in a **dream**

**Tree of evolution**  
Darwin (1859) **imagines**  
generations of species – “I think”



**Solitaire and the periodic table**  
Mendeleev (1869) organized chemical  
elements after a **mental image** of cards on a  
table.

See: <https://medium.com/@michael.friendly/visual-thinking-graphic-discoveries-128468677592>

# How to visualize travel? A route map!

In 1675, chartmaker John Ogliby told a graphic story of what you would see on a travel from London to Land's End

What was he thinking?

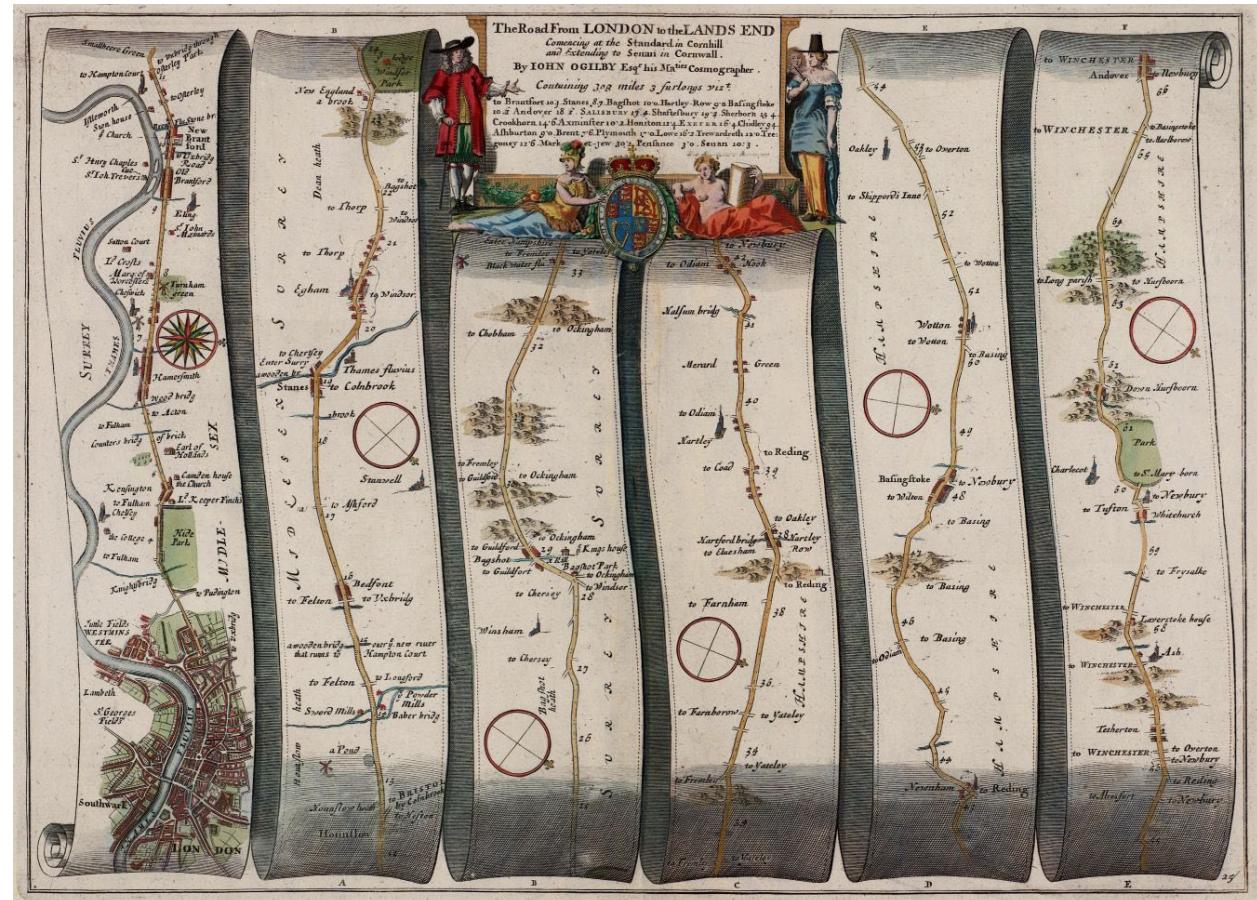
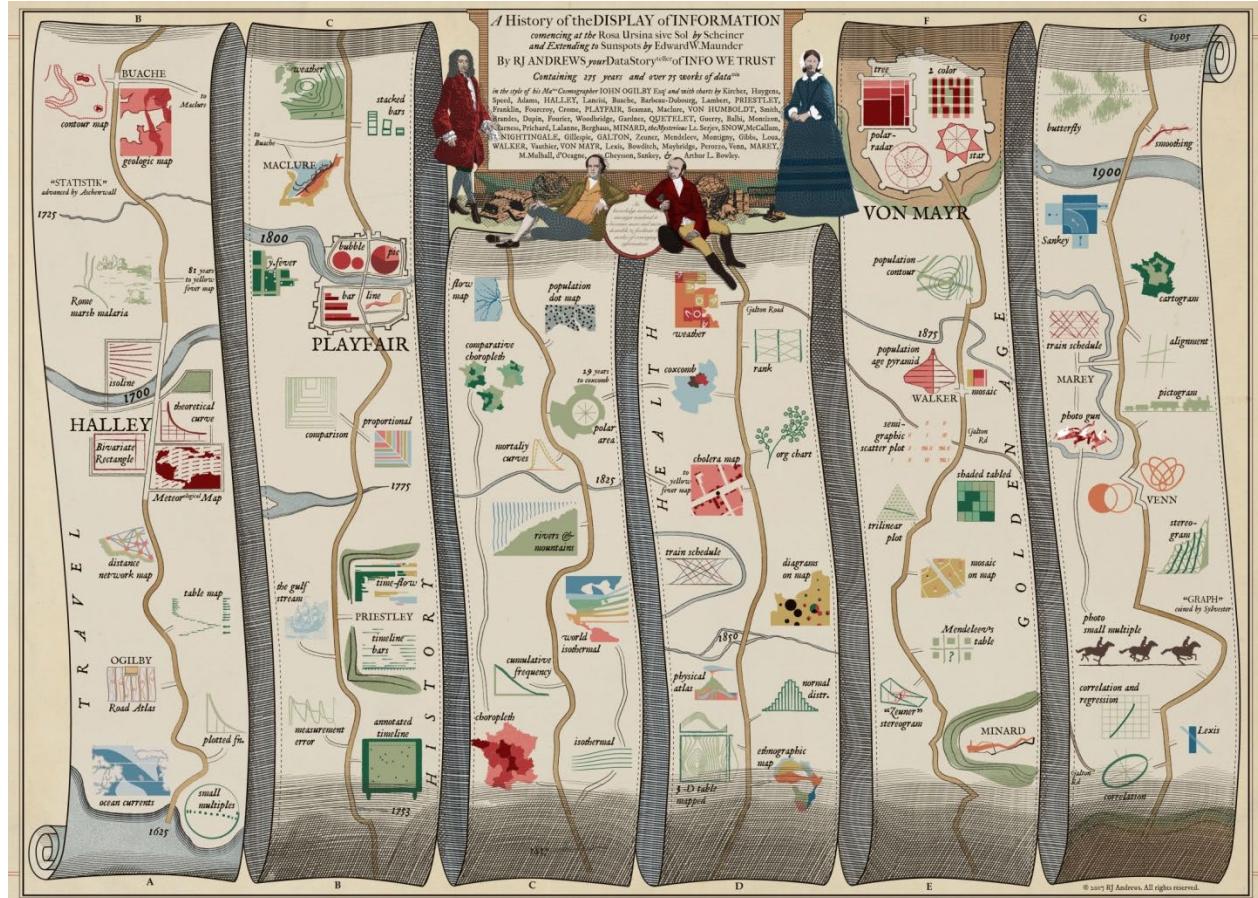


Image: [https://commons.wikimedia.org/wiki/File:Ogilby - The\\_Road\\_From\\_LONDON\\_to\\_the\\_LANDS-END\\_\(1675\).jpg](https://commons.wikimedia.org/wiki/File:Ogilby - The_Road_From_LONDON_to_the_LANDS-END_(1675).jpg)

# How to visualize history? A route map!

In 2017, RJ Andrews adapts Ogilvie's form show the history of data visualization.

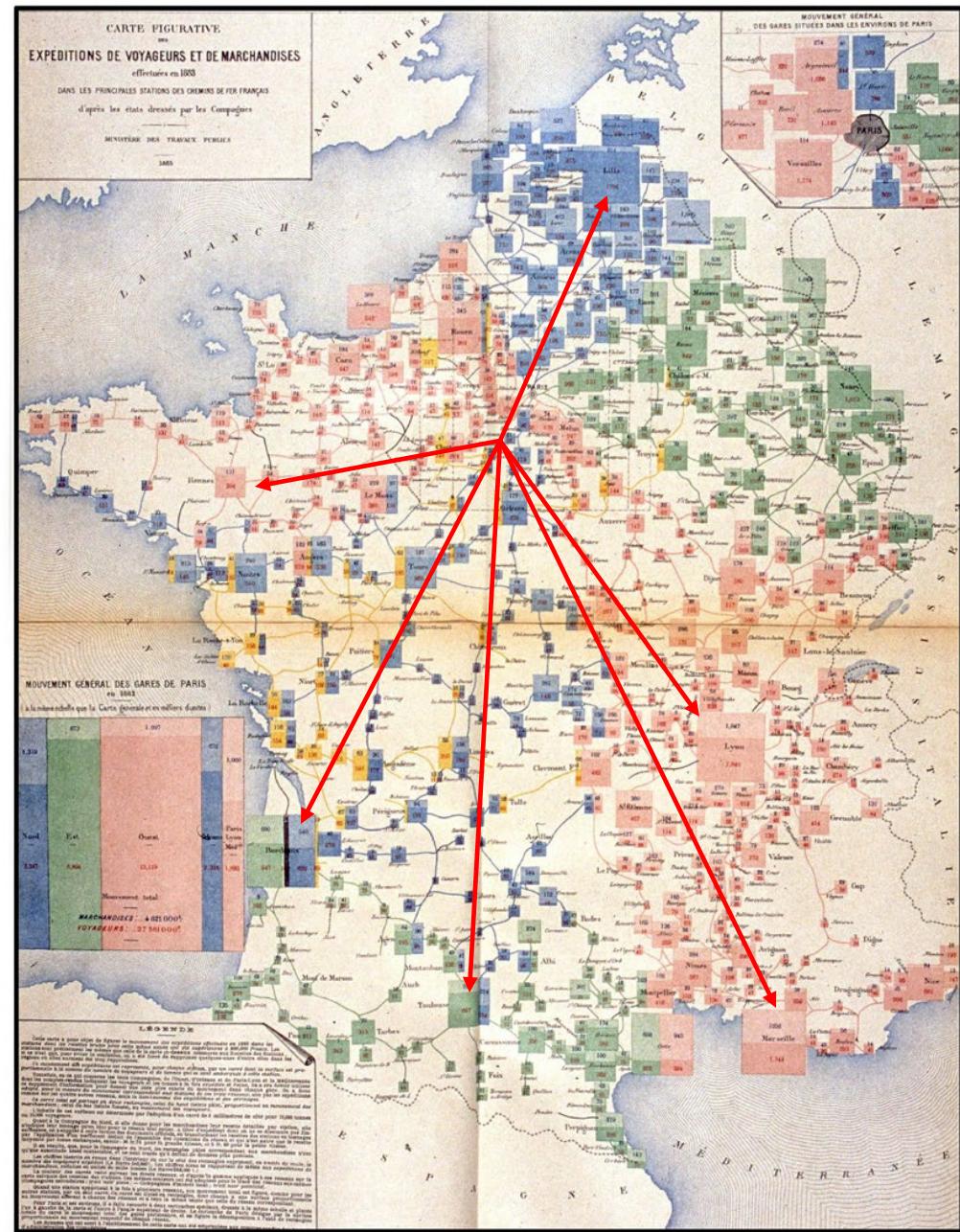
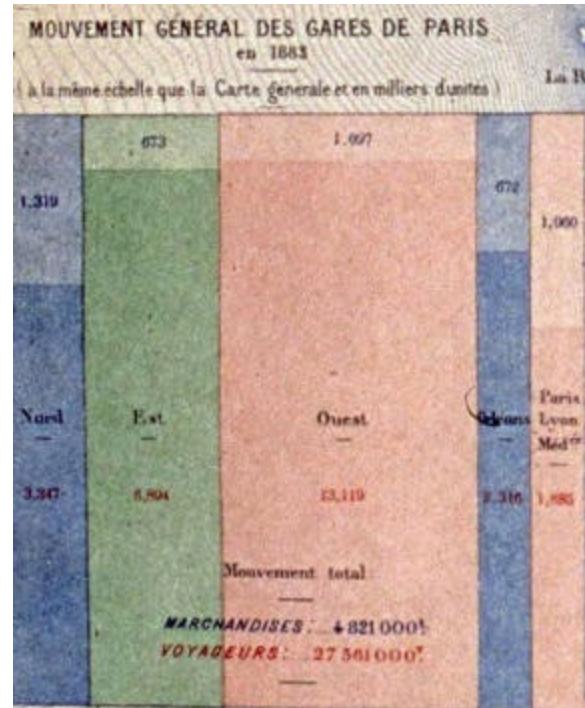


The online version, <https://history.infowetrust.com/> is fully interactive, with details about the images on this journey.

# Why study history of #datavis?

*Those who don't know history are doomed to plagiarize it.*

**Recursive mosaic:** Distribution of passengers and goods from the Paris railways to the rest of France [Album, 1884, pl. 11]





# Quest for the Albums

- British library, BNF, Library of Congress: just a few copies
- Richard Langdon, U of T Fisher Rare Book Library: check out this bookshop, 3 rue des Beaux Arts, Paris
  - A complete set: all albums 1879 – 1899!
- Les Chevaliers
  - Collective purchase, owned by all, each held “in trust” by one member
  - “chevaliers”: Foster a spirit of collegial study of history of data visualization & thematic cartography
  - Conference sessions: RC33 (Cologne, 2000), GFKL (Dortmund, 2004), JSM (Toronto, 2004), ...
  - Regular “Chevalier Lunch”

# *Les Chevaliers des Albums*

Antoine de Falguerolles



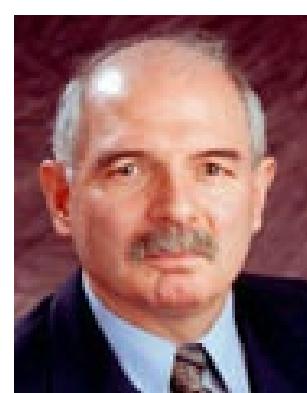
Gilles Palsky



Antony Unwin



Ian Spence



Howard Wainer



Michael Greenacre



RJ Andrews



Sandra Rendgen



Stephen Stigler



David Rumsey



# The Milestones Project

## Milestones in the History of Thematic Cartography, Statistical Graphics, and Data Visualization

An illustrated chronology of innovations by Michael Friendly and Daniel J. Denis

[Home](#) | [Introduction](#) | [Milestones Project](#) | [Varieties of Data Visualization](#) | [Related](#) | [References](#) | [Keyword Index](#)

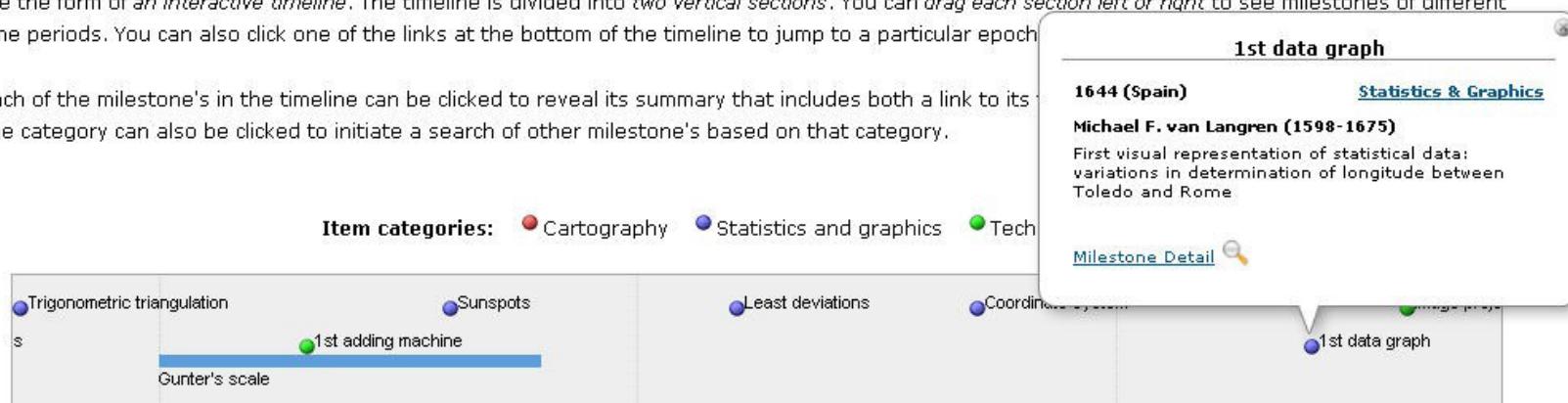
[Search](#)

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

### Timeline

This page provides a graphic overview of the events in the history of data visualization that we call "**milestones**." These milestones are shown below in the form of an *interactive timeline*. The timeline is divided into two vertical sections. You can drag each section left or right to see milestones of different time periods. You can also click one of the links at the bottom of the timeline to jump to a particular epoch.

Each of the milestone's in the timeline can be clicked to reveal its summary that includes both a link to its category and a search of other milestone's based on that category.



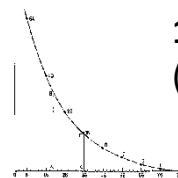
The web site: <http://datavis.ca/milestones> has an interactive timeline, allowing different kinds of search

# Milestones: Content Overview

*Every picture has a story* – Rod Stewart

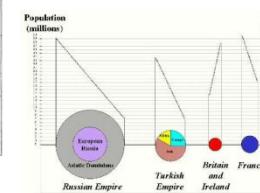
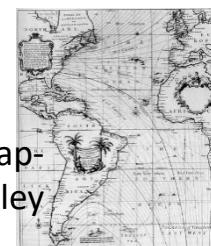


c. 550 BC: The first world map? (Anaximander of Miletus)



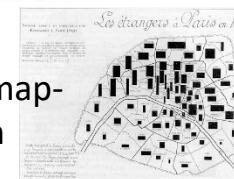
1669: First graph of a continuous distribution function  
(Gaunt's life table)– Christiaan Huygens.

1701: First contour map-  
Edmund Halley



1801: Pie chart, circle  
graph - William Playfair

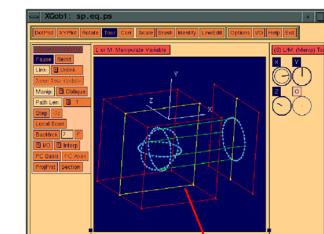
1896: Bivariate map-  
Jacques Bertillon



1924: ISOTYPE-  
Otto Neurath

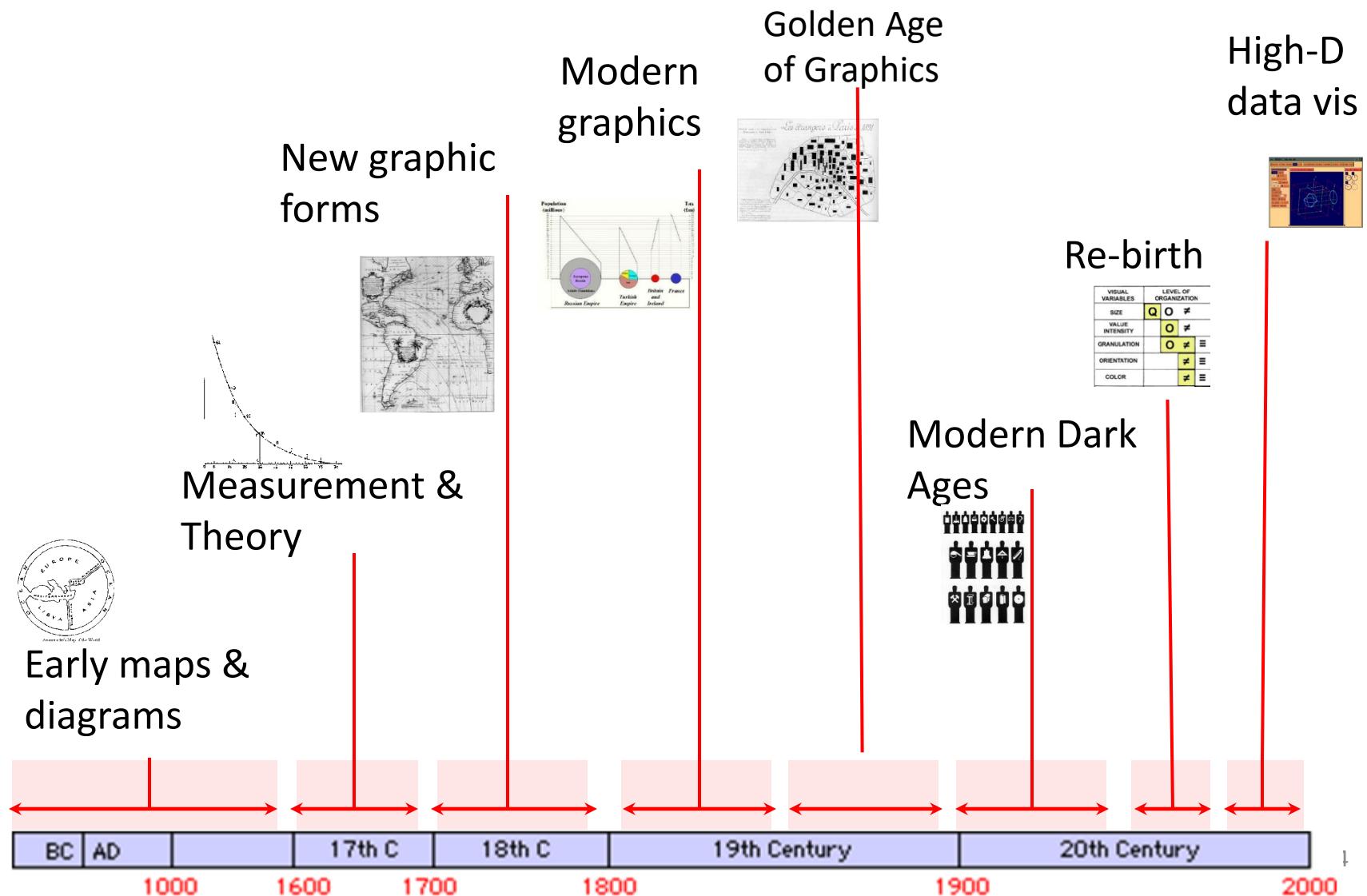


1991-1996:  
Interactive data  
visualization  
systems (Xgobi,  
ViSta)



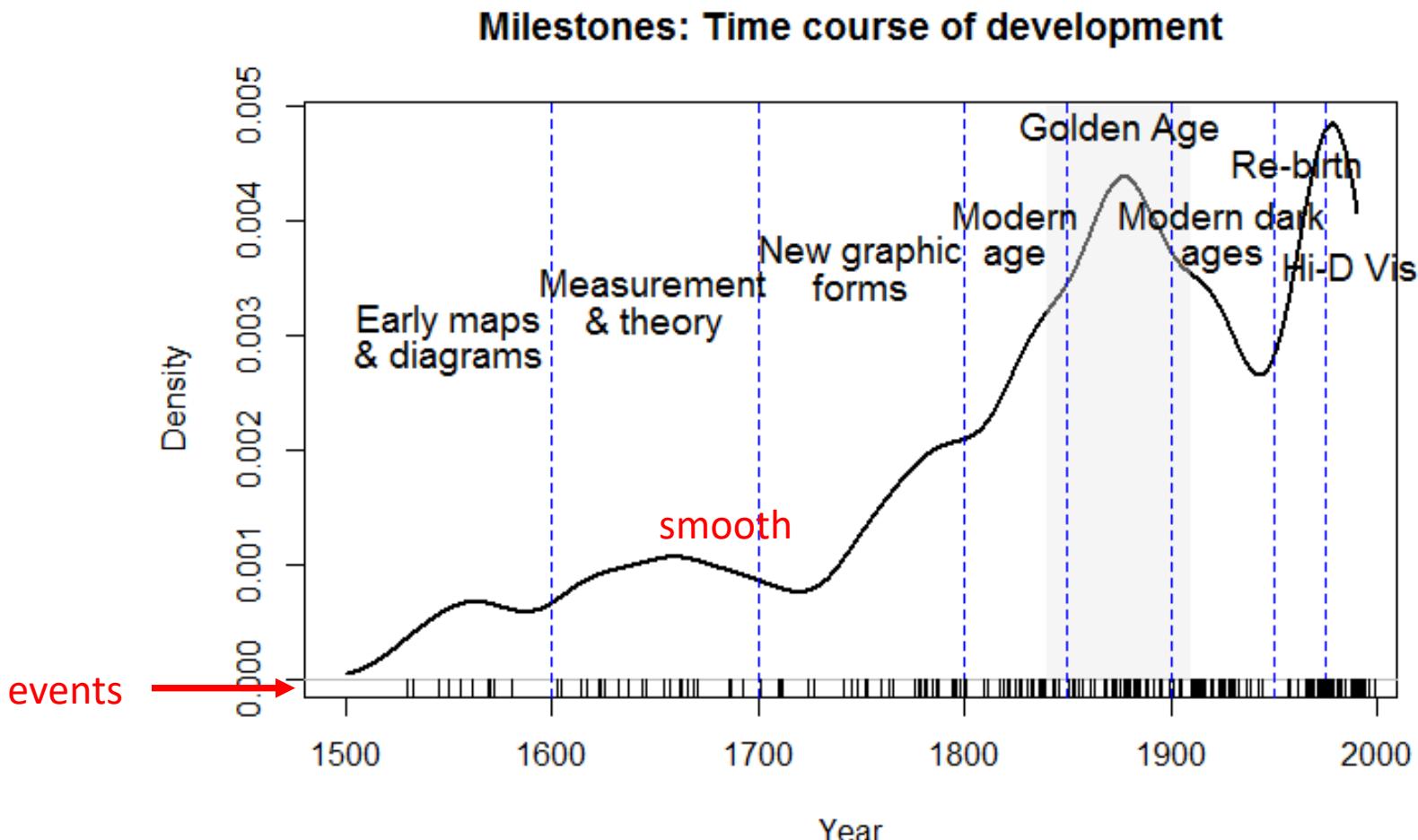
BC	AD	1000	1600	1700	1800	19th Century	20th Century	2000
----	----	------	------	------	------	--------------	--------------	------

# Milestones Tour: Epochs



# Statistical historiography

Historical information, suitably organized can be treated as data, and analyzed. This plot shows a smoothed frequency distribution of 248 milestones items over time, in relation to the named time periods.



# Prehistory of visualization

Lascaux Cave, ~ 15000 BCE, the “Sistine Chapel of pre-historic art”



Lascaux II, Main chamber

# Lascaux: What were they thinking?



Lascaux II, Chamber of the Bulls

- Visual features:
  - show perspective, a sense of motion, rich use of color & texture
- What was the purpose?
  - Hunting success? NO (they hunted reindeer)
  - mostly symbolic – visual language, story of communal myths
- How to understand them?
  - A **cognitive revolution**: evidence for the modern human mind in Cro Magnon man
  - inner vision, visual thinking, mental imagery – a gleam in the mind's eye
- Other cave art [20000BC – 10000BC]: Altamira (Spain); Chauvet (France), Cueva de las Manos (Argentina), ...

# Prehistory: Diagrams, graphic stories

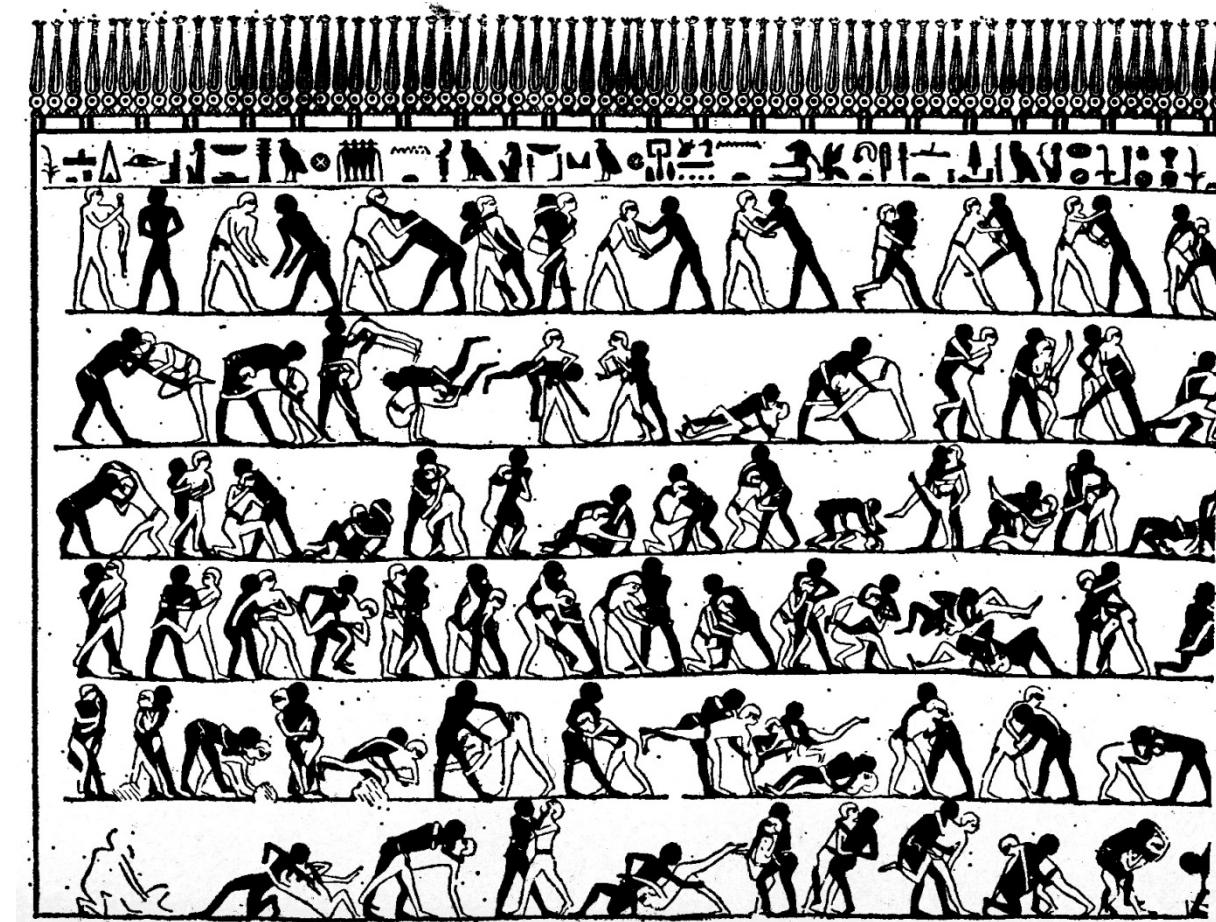
Early Egyptian animated graphic diagram

Wrestling scene on east wall, tomb of Baqt at Beni Hasan (ca. 2000 BCE).

A **visual explanation**  
of a wrestling match

Anticipates modern  
graphic novels

Why? Perhaps Baqt's  
last lesson as a  
wrestler in his youth  
and later as a coach



# Pre 17<sup>th</sup> C.: Early maps & diagrams

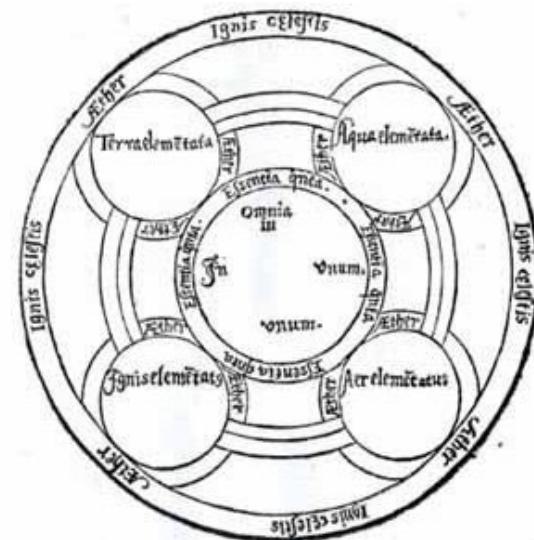
c. 550 BC: The first world map? (Anaximander of Miletus)



**1350:** Bar graph of theoretical function N.  
Oresme, France



## **1305:** Mechanical diagram of knowledge- Ramon Llull, Spain



**1375:** Catalan Atlas, an exquisitely beautiful visual cosmography, perpetual calendar, and thematic representation of the known world- Abraham Cresques, Spain



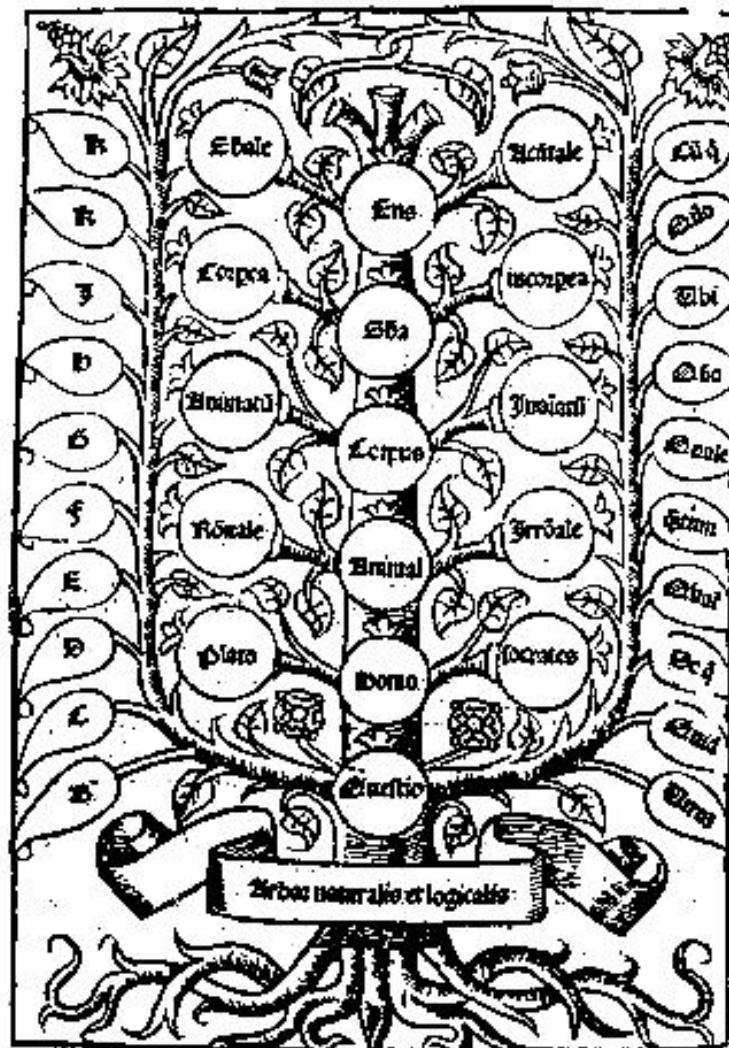
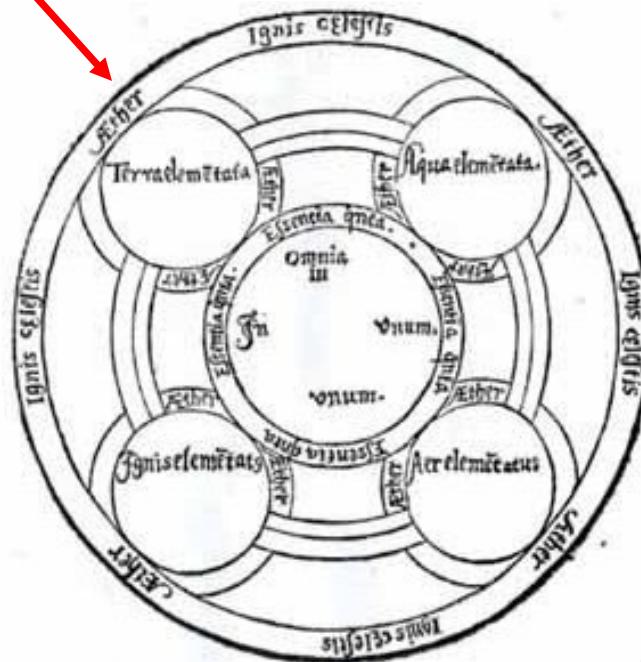
## **1305: Mechanical diagram of knowledge- Ramon Llull, Spain**

*Ars Magna* → inspiration for symbolic logic (Leibnitz)



## *Tree of porphyry*: Aristotle's categories of knowledge (center)

- Left: questions
  - Right: rotating disks → answers



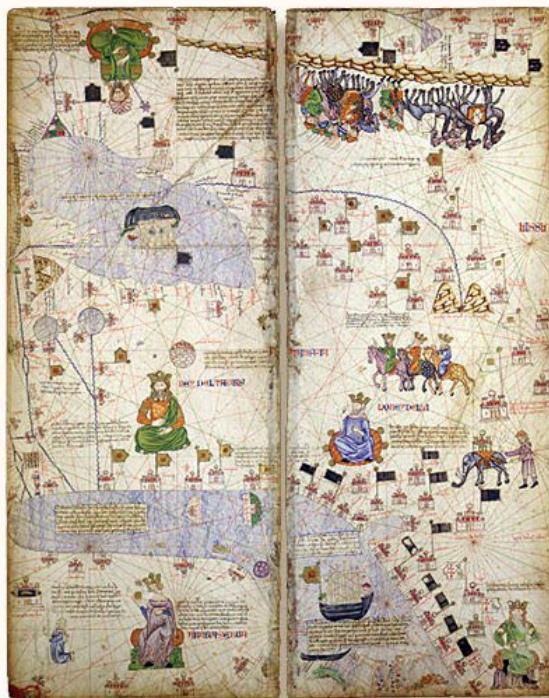
BC	AD		17th C	18th C	19th Century	20th Century		
			1000	1600	1700	1800	1900	2000



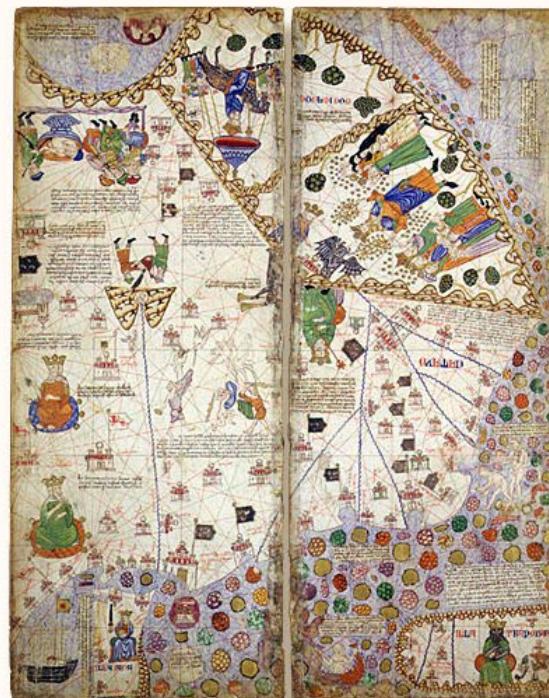
1375: Catalan Atlas, an exquisitely beautiful visual cosmography, perpetual calendar, and thematic representation of the known world- Abraham Cresques, Majorca, Spain [BNF: ESP 30]



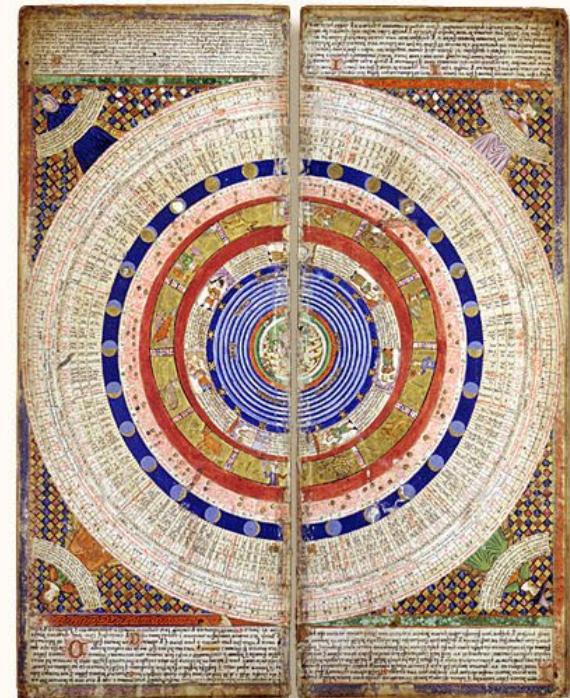
Western world



Eastern world (Marco Polo)



Perpetual calendar

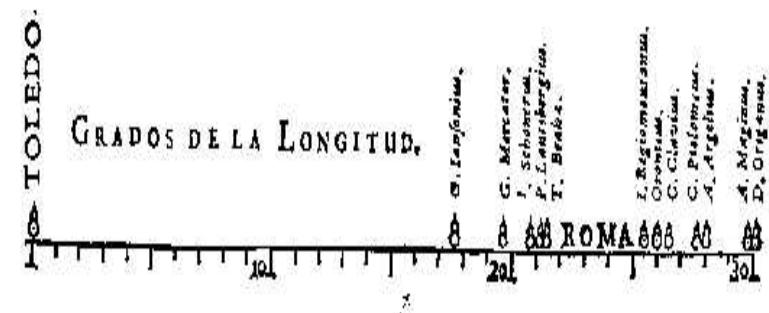
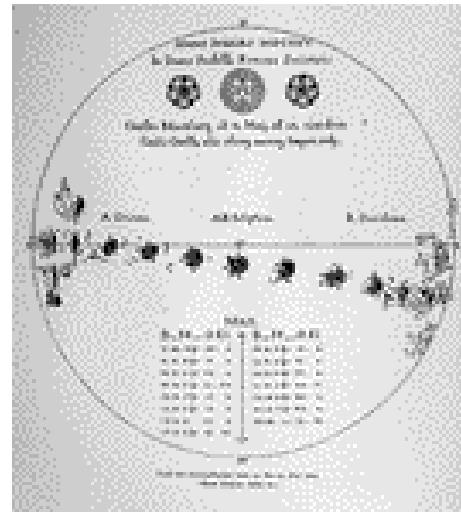


# 1600-1699: Measurement and Theory

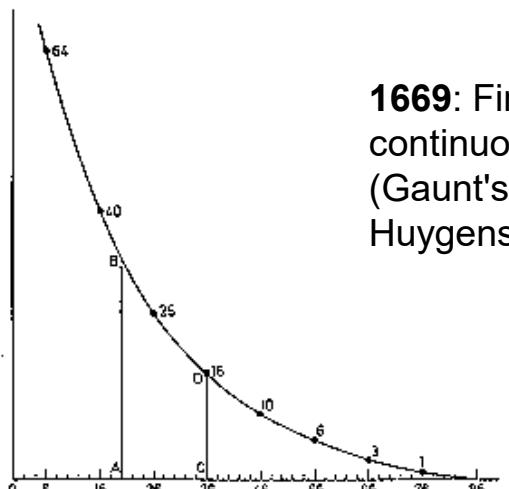
- The 17<sup>th</sup> century saw growth in theory and the dawn of attempts at visualization.
- Featured in this were:
  - the rise of **analytic geometry**: (x, y) coordinates (Descartes),
  - theories of **errors of measurement**: astronomical observations (Laplace)
  - the birth of **probability theory**-- games of chance, annuities (Fermat, DeMoivre, ...),
  - automatic **graphic recording** (Scheiner)
  - the first graphical representations of **statistical data** (van Langren)

# 1600-1699: Measurement and Theory

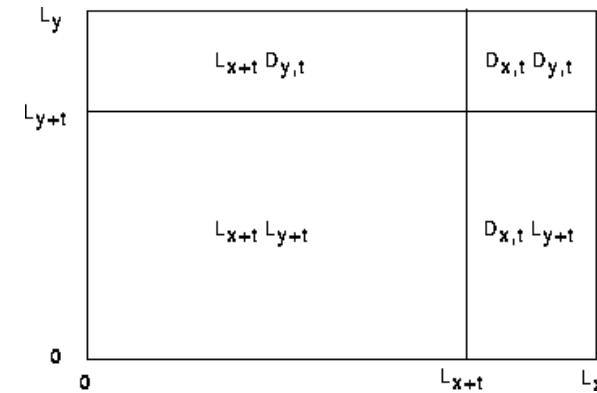
1626: Visual representations used to chart the changes in sunspots over time—Christopher Scheiner



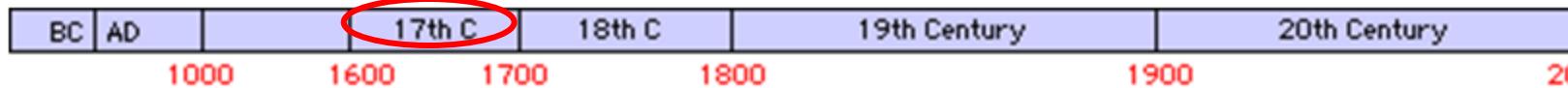
1644: First visual representation of statistical data—M.F. van Langren, Spain



1669: First graph of a continuous distribution function (Gaunt's life table)—Christiaan Huygens.



1693: First use of areas of rectangles to display probabilities of independent binary events—Edmund Halley, England





# Sunspots: Galileo



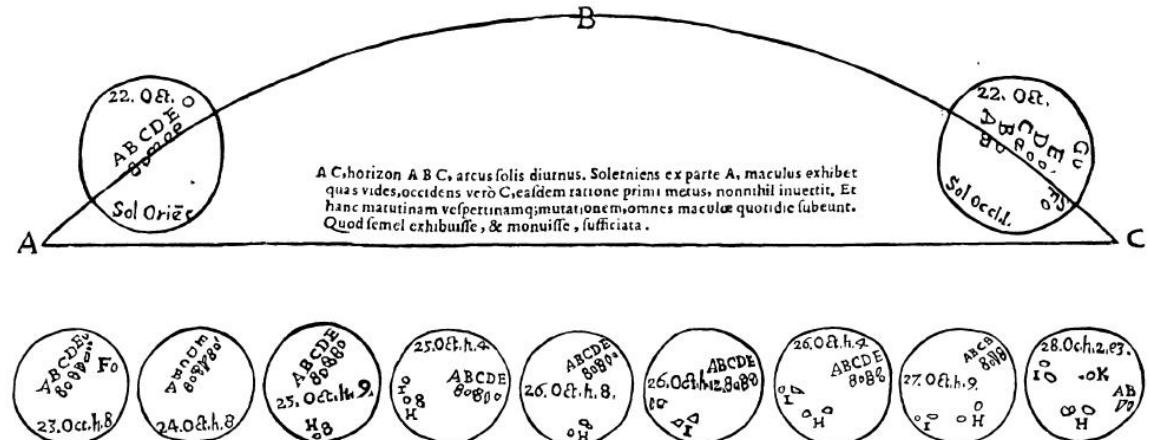
1608: telescope (Hans Lippershey, NL)

↓  
1610: Galileo (*Sidereus Nuncius*)

1611: Galileo records **movement** of sunspots over time (*Three letters on sunspots*, 1613)

## Visual ideas:

- Animated graphic
- “Small multiples”
- Allows comparison
- Self-explaining diagram

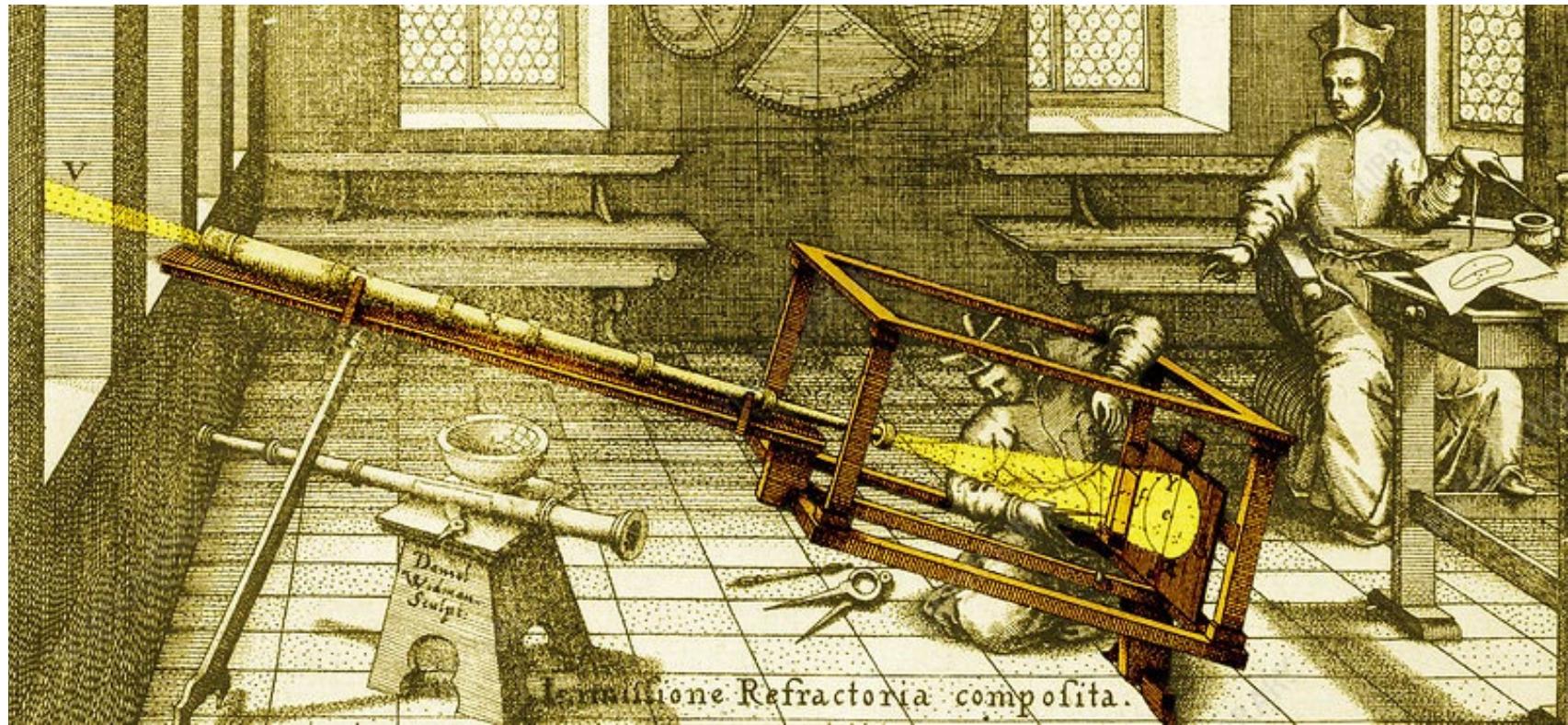


A+ for info design!

The idea of diagrams for visualizing phenomena had arrived.

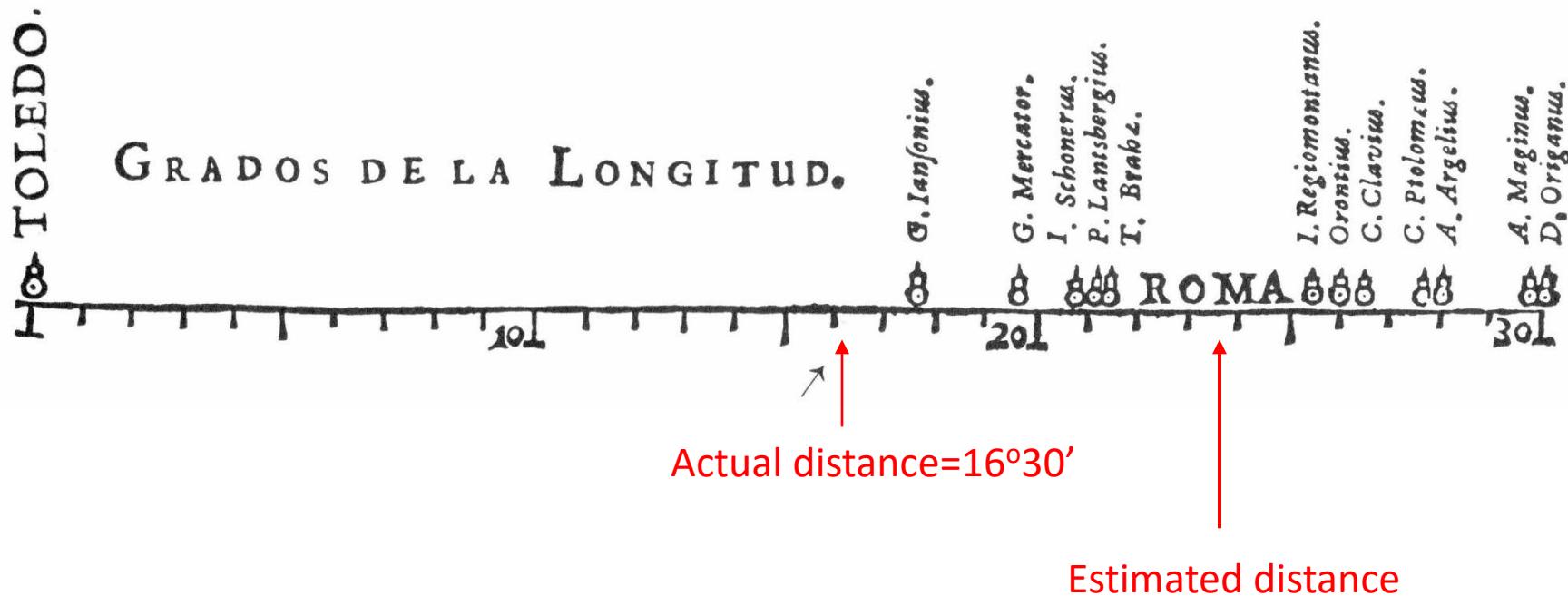
# Scheiner: systematic recording

1626: Christoph Scheiner invents helioscope & camera obscura to record sunspots (*Rosa Ursina sive Sol*, 1626-1630)



# Why the 1<sup>st</sup> statistical graph got it right

1644: First visual representation of statistical data: determination of longitude between Toledo and Rome- Michael Florent van Langren, Spain



BC	AD	1000	1600	1700	1800	19th Century	20th Century	2000
				17th C				

# What else could he have done?

- What would occur to men of his time to convey a message to the king?
- ... he could used a *table* have sorted by *year* to establish *priority* (or show change).

Sorted by Priority

Year	Name	Longitude	Where
150	Ptolomeus, C.	27.7	Egypt
1471	Regiomontanus,	25.4	Germany
1501	Ianfonius, G.	17.7	
1530	Lantsbergius, P.	21.1	
1536	Schonerus, I.	20.8	Germany
1541	Argelius, A.	28.0	
1542	Ortonius	26.0	France
1567	Mercator, G.	19.6	Flanders
1567	Clavius, C.	26.5	Germany
1578	Brahe, T.	21.5	Denmark
1582	Maginus, A.	29.8	Italy
1601	Organus, D.	30.1	

Sorted by Authority

Name	Longitude	Year	Where
Argelius, A.	28.0	1541	
Brahe, T.	21.5	1578	Denmark
Clavius, C.	26.5	1567	Germany
Ianfonius, G.	17.7	1501	
Lantsbergius, P.	21.1	1530	
Maginus, A.	29.8	1582	Italy
Mercator, G.	19.6	1567	Flanders
Organus, D.	30.1	1601	
Orontius	26.0	1542	France
Ptolomeus, C.	27.7	150	Alexandria
Regiomontanus, I.	25.4	1471	Germany
Schonerus, I.	20.8	1536	Germany

Sorted by Longitude

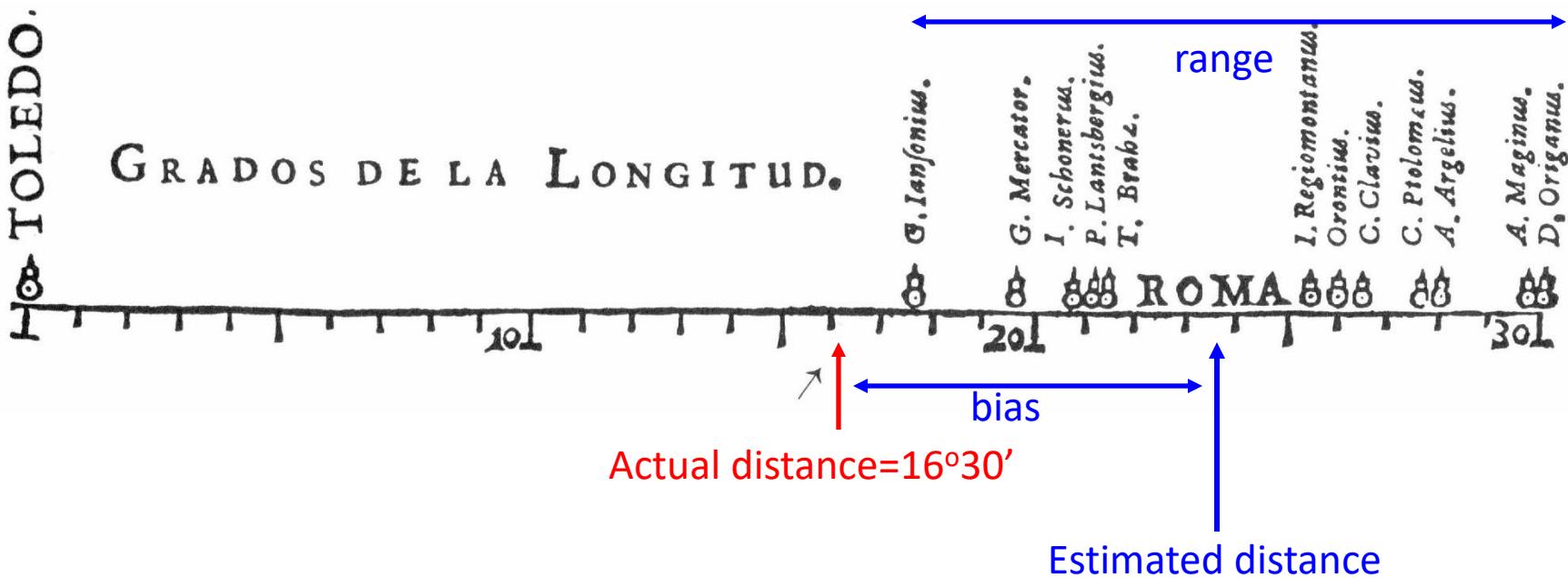
Longitude	Name	Year	Where
17.7	G. Ianfonius	1501	
19.6	G. Mercator	1567	Flanders
20.8	I. Schonerus	1536	Germany
21.1	P. Lantsbergius	1530	
21.5	T. Brahe	1578	Denmark
25.4	I. Regiomontanus	1471	Germany
26.0	Orontius	1542	France
26.5	C. Clavius	1567	Germany
27.7	C. Ptolomeus	150	Egypt
28.0	A. Argelius	1541	
29.8	A. Maginus	1582	
30.1	D. Organus	1601	

- ... he could have sorted by *name*, to show *authority*.

- ... he could have sorted by *longitude* to show the *range*.

# Only a graph shows...

- central location
- bias
- name labels– avoiding overplotting
- wide variability
- clustering, detached observations



See: Friendly, M., & Kwan, E. (2003). Effect Ordering for Data Displays. *Computational Statistics and Data Analysis*, 43(4), 509–539;  
Friendly et al (2010), The First (Known) Statistical Graph: Michael Florent van Langren and the "Secret" of Longitude  
*The American Statistician*, 64, 185-191

# 1700-1799: New graphic forms

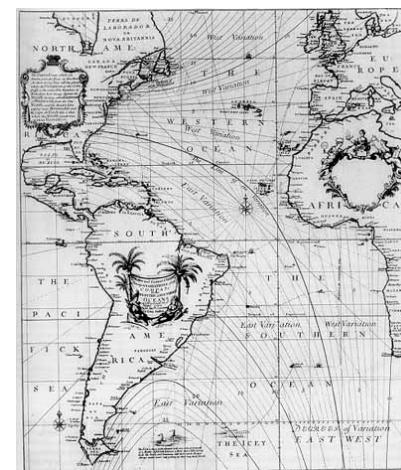
- The 18<sup>th</sup> century witnessed the germination of the seeds of visualization & visual thinking, planted earlier.
  - Map-makers began to try to show more than just geographical position-- the beginnings of **thematic mapping** of physical quantities
    - topographical maps
    - iso-contour maps
  - New graphic forms were invented:
    - bar chart,
    - line chart,
    - timelines
- } The Big Bang



BC	AD	1000	1600	1700	1800	19th Century	20th Century	2000

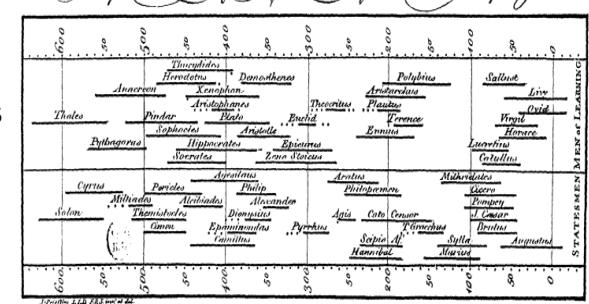
# 1700-1799: New graphic forms

**1701:** Isobar map, lines of equal magnetic declination – Edmund Halley

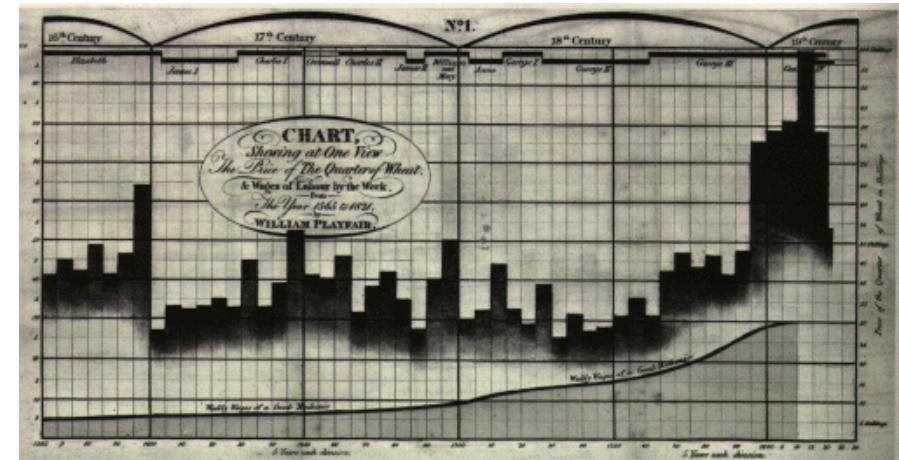


**1782:** First topographical map- Marcellin du Carla-Boniface

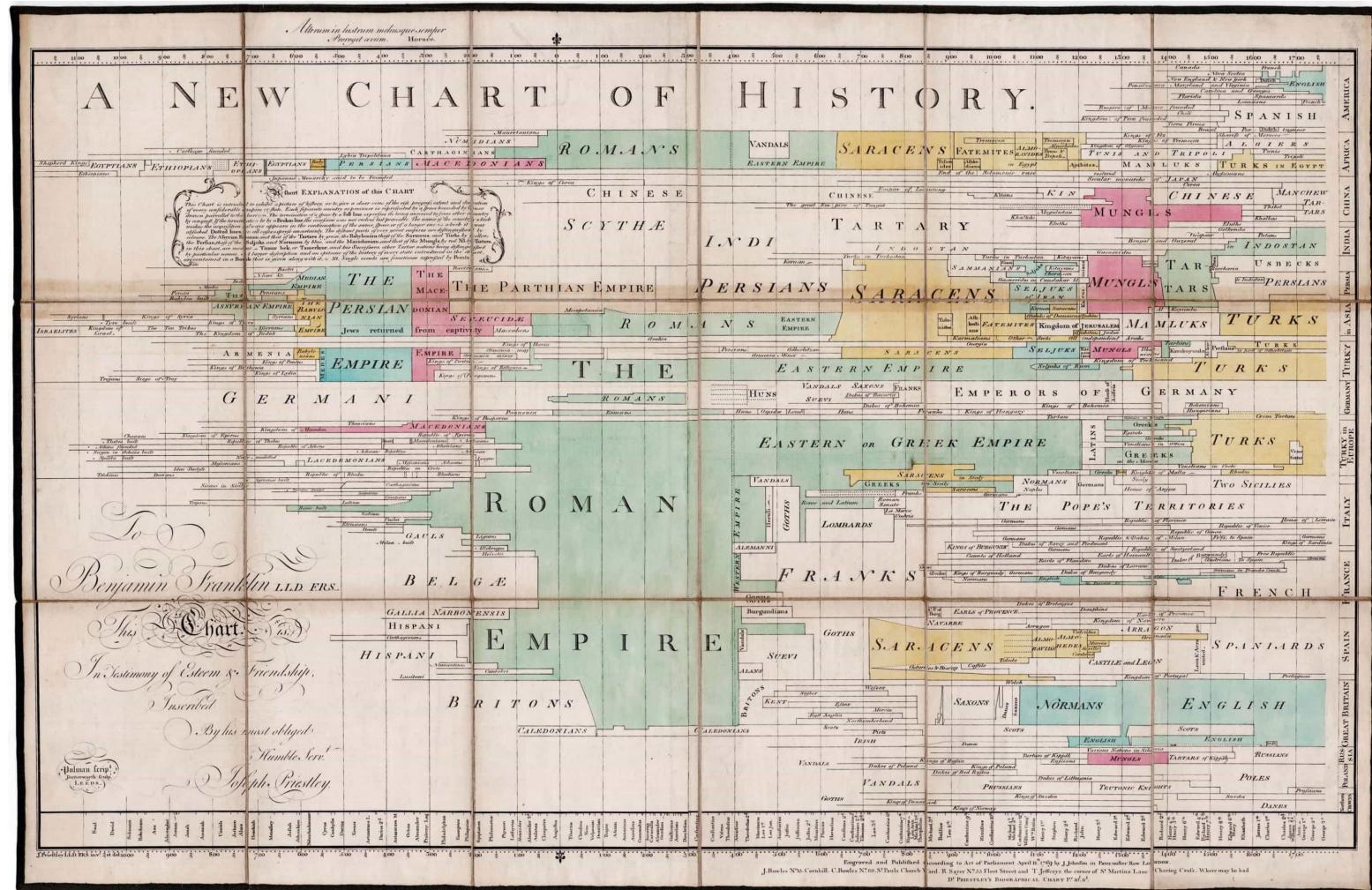
**1765:** Historical time line (life spans of famous people) Joseph Priestley



**1786:** Bar chart, line graphs of economic data- William Playfair



**1769:** Visualization of the history of civilizations & empires over ~3000 years --Joseph Priestley



1000 BC

0

1000

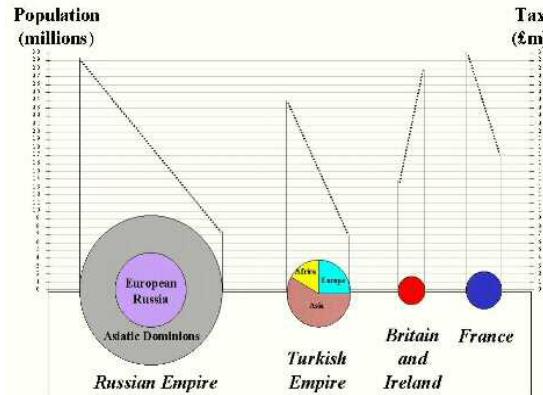
# 1800-1849: Beginning of modern data graphics

- The first half of the 19<sup>th</sup> century witnessed an explosive growth in statistical graphics and thematic mapping
  - Polar coordinates, log axes
  - Shaded (choropleth) maps of social data (literacy, crime)
- The **birth of data**: widespread national collection of data on social and medical issues
  - France: data on crime, literacy, prostitution, ... collected centrally
  - England: Births, deaths, disease mortality collected by Registrar General (William Farr)

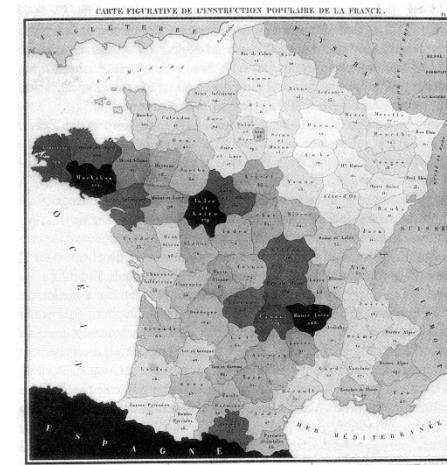


# 1800-1849: Beginning of modern data graphics

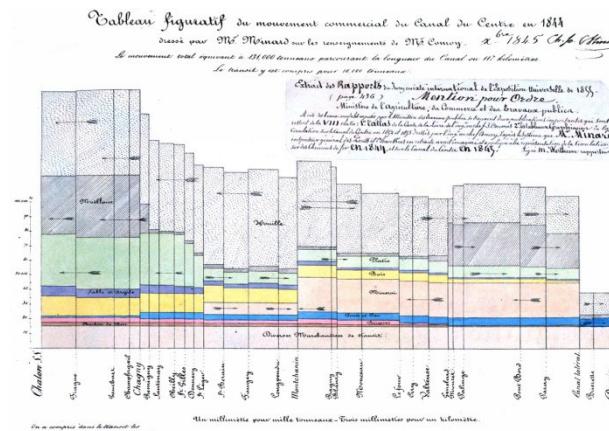
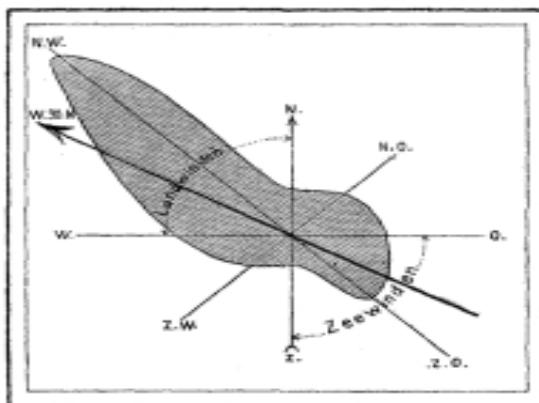
**1801:** Pie chart, circle graph invented- William Playfair



**1826:** First modern statistical map (illiteracy in France)- Charles Dupin



**1843:** Wind-rose (polar coordinates)- L. Lalanne

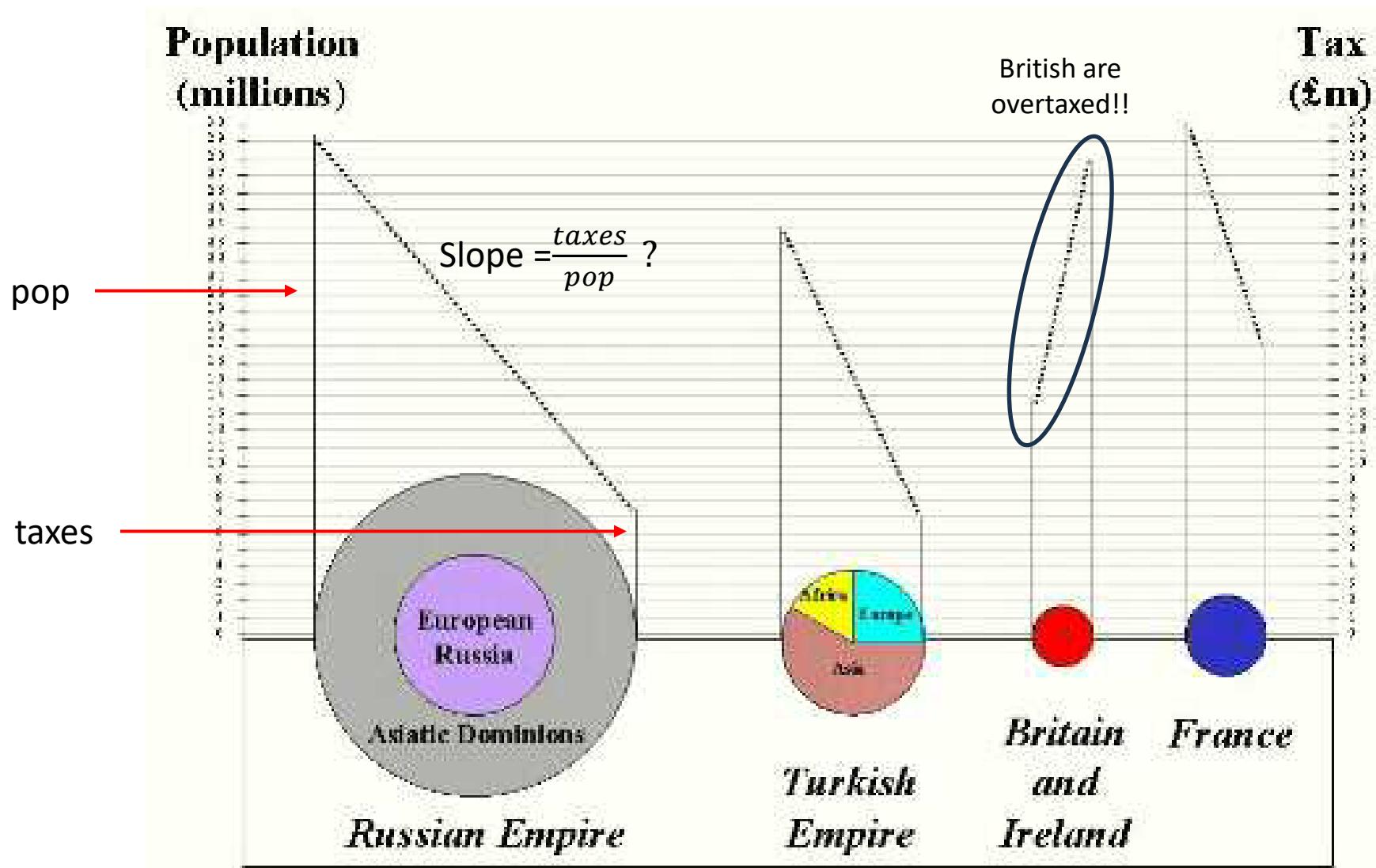


**1844:** variable-width, divided bars, area ~ cost of transport- C. J. Minard

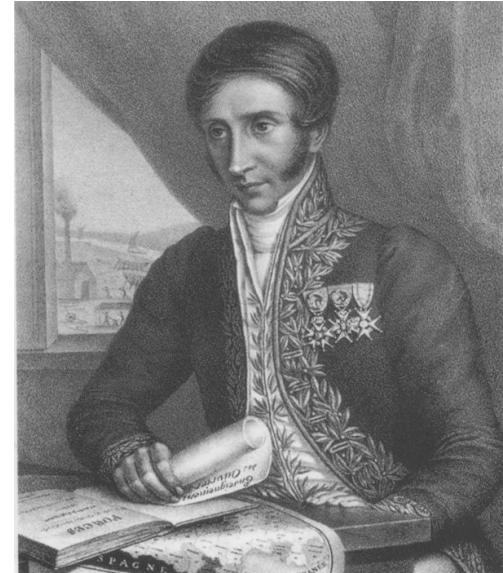
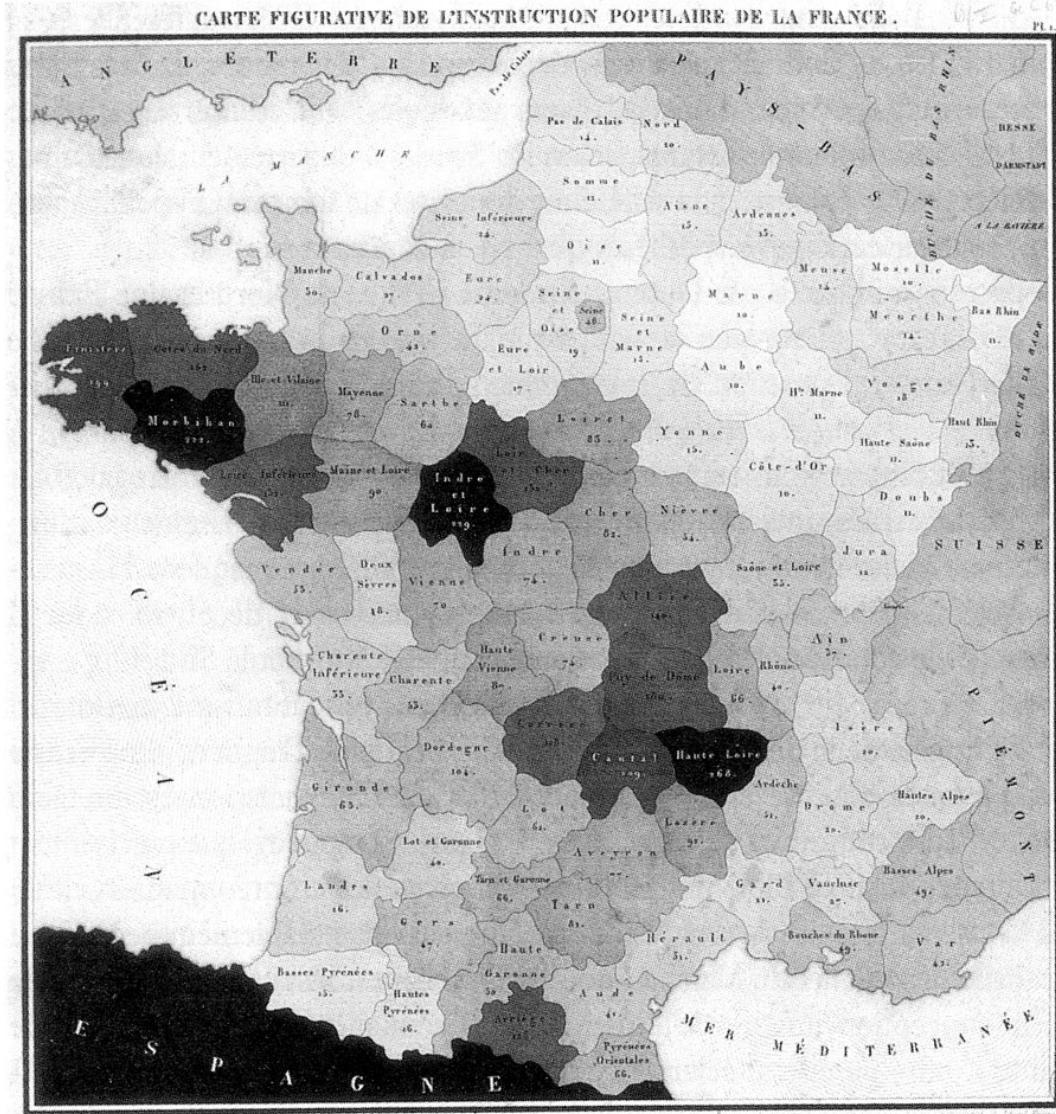


# 1801: Pie chart, circle graph invented- William Playfair

(But with a graphic sin & fallacy – What are they?)



# 1826: The 1<sup>st</sup> choropleth map, showing the distribution of literacy in France – Baron Charles Dupin



Social variables became:

- visual
- subject to scientific discussion

# 1844: Tableau-graphique: variable-width, divided bars, area ~ cost of transport- Charles Joseph Minard

## Tableau figuratif du mouvement commercial du Canal du Centre en 1844

dressé par M<sup>e</sup> Minard sur les renseignements de M<sup>e</sup> Comoy. <sup>6e</sup> - 1845 Ch. J. Minard

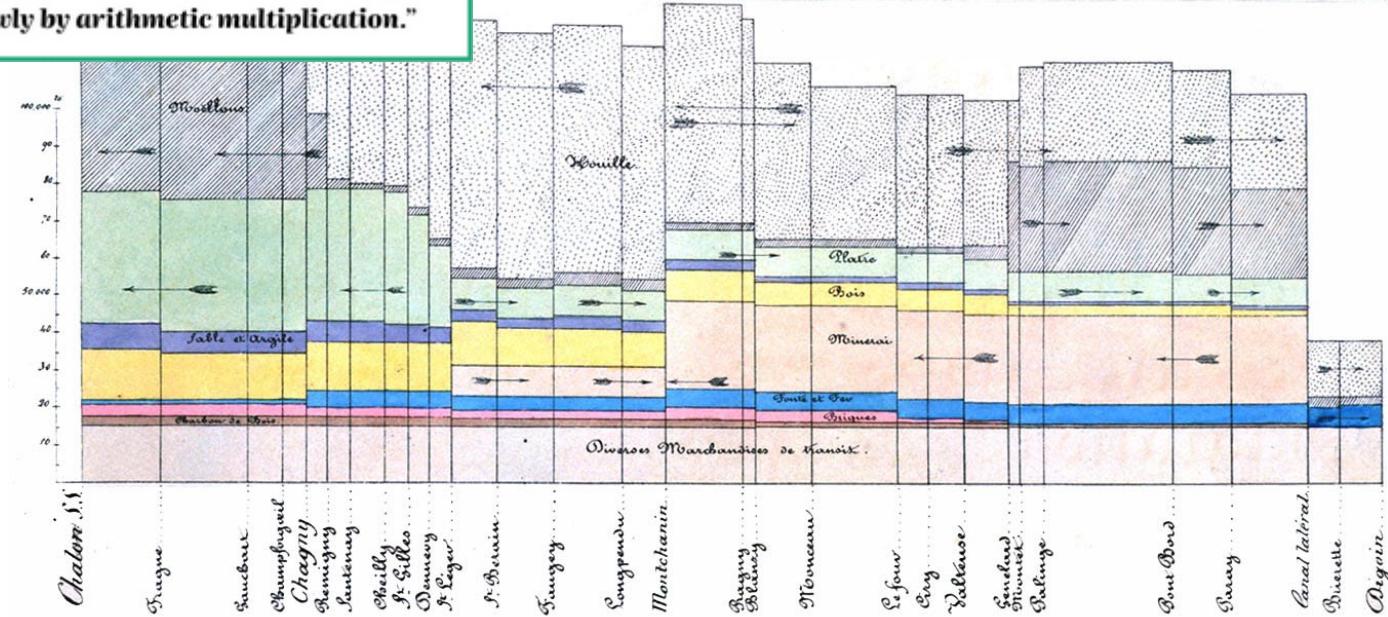
Le mouvement total équivaut à 131,000 tonnes au parcourir la longueur du Canal en 117 kilomètres.

Le transit y est compris pour 16,000 tonnes.

"the total areas of each rectangle being proportional to the number of passengers carried at 1 kilometer [...], the comparison of these surfaces gives promptly, by the image, what written numbers only give slowly by arithmetic multiplication."

Extrait des Rappports à l'Exposition internationale de l'Exposition Universelle de 1855.  
(page 436) Mention pour Ordre.

Ministère de l'Agriculture, du Commerce et des Travaux publics.  
A côté des bons modèles exposés par le Ministère des Travaux publics se trouvent diverses publications importantes qui sont du résultat de la VIII<sup>e</sup> classe. L'Atlas de la carte de la route de l'ingénieur Comte 2. Tableau graphique de la circulation sur le canal du Centre en 1842 et 1843. Il démontre l'exactitude du système que M. Minard introduit généralement dans ses travaux en ce qu'il applique à la représentation de la circulation sur les chemins de fer en 1844, et sur le canal du Centre en 1845. Signé M. Wertheim rapporteur.

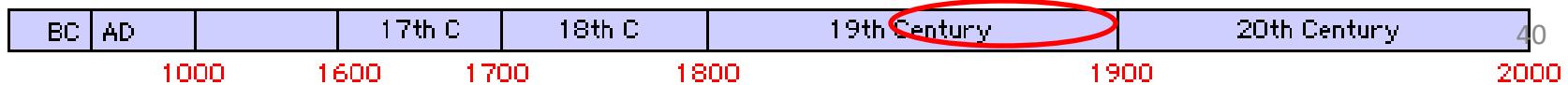


Un millimètre pour mille tonnes - Trois millimètres pour un kilomètre.

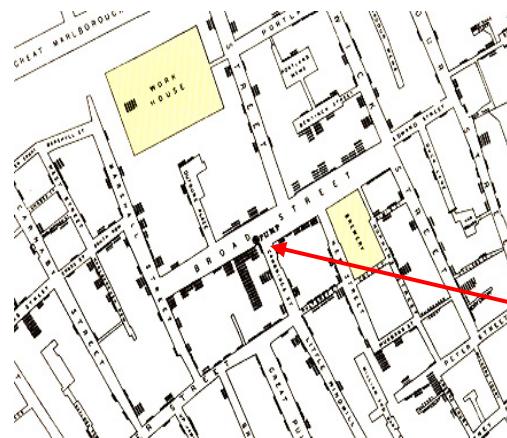
On a compris dans le transit les marchandises allant de Chalon au Canal latéral à la Saône et vice versa.

# 1850-1900: Golden Age

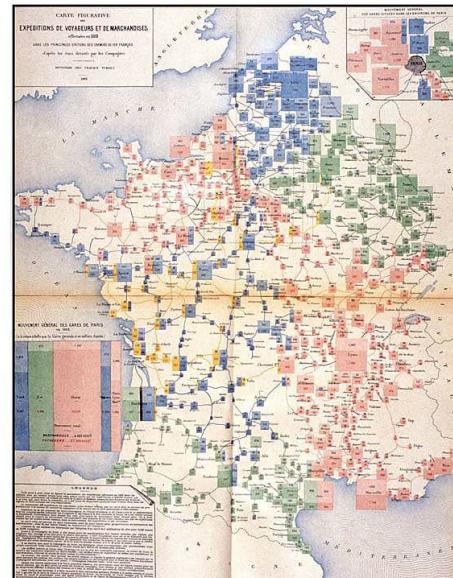
- By the last half of the 19<sup>th</sup> century the conditions for rapid growth of visualization had been established:
  - widespread data collection for planning, commerce, social theory
  - the beginnings of statistical theory and visual thinking
  - a wide range of graphic forms, reasonably well understood
  - technology:
    - lithography and color printing
    - automatic recording devices
    - calculation: machines & graphical calculators
- The result was a perfect storm-- among the most exquisite graphics ever produced.



# 1850-1900: Golden Age

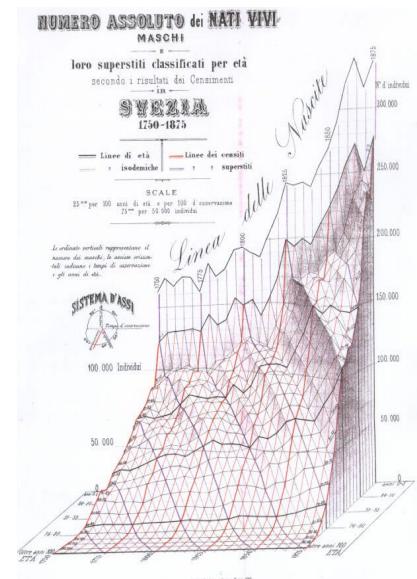


1855: Dot map of disease data (cholera)- John Snow

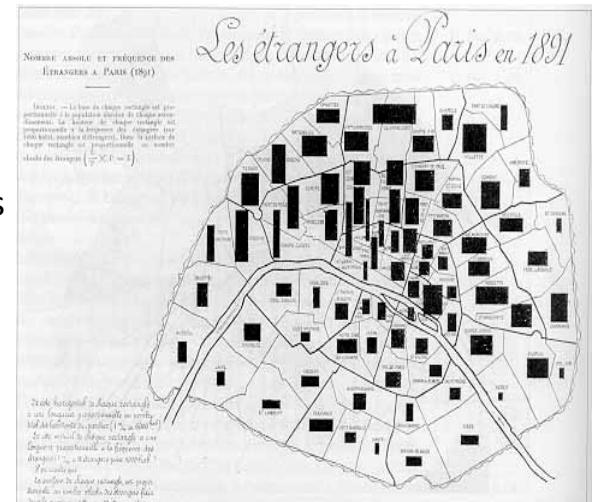


1884: Recursive multi-mosaic on a map- Emile Cheysson

1879: Stereogram (3D population pyramid)- Luigi Perozzo



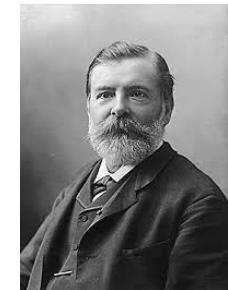
1896: Area rectangles on a map to display two variables and their product- Jacques Bertillon



BC	AD	1000	1600	1700	17th C	18th C	1800	19th Century	20th Century	2000
----	----	------	------	------	--------	--------	------	--------------	--------------	------

# E.-J. Marey: *La Méthode Graphique*

- How to make human and animal motion subject to precise scientific study?
- e.g., aerial locomotion of flying insects & birds
  - What is the frequency of wings of different species?
  - What are the mechanisms of wings to produce lift and forward motion?



A harness, designed to register the trajectory, force and speed of a bird's wing in flight

Marey (1870) *Animal Mechanism*

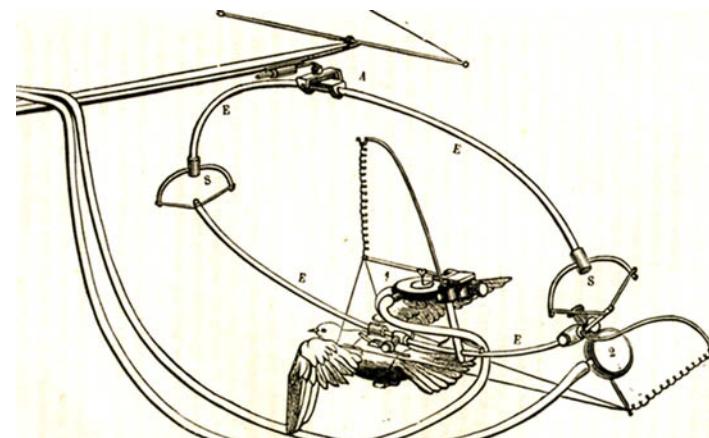


FIG. 104.—Suspension of the bird in the instrument. E E E E, an ellipse of metal capable of oscillating freely in every direction, by means of the double suspension A. S S, india-rubber supports allowing the lower part of the ellipse to oscillate in the vertical direction. The suspensory apparatus is fixed on the back of the pigeon. The lever-drum (1) receives the movements executed by the wing in a vertical direction. The lever-drum (2) receives those of the horizontal movements.

# E.-J. Marey: Chronophotography



Fig. 1. Voie d'emploi du fusil photographique.

Rather than separate frames, Marey's "fusil photographique" allowed one to see motion continuously in a single static image.

This provides a visual analysis of a sprint:

- The runner takes about  $\frac{1}{2}$  second (7 frames) to make it to an upright position
- Successive frames alternate between power push from the hind leg to landing on the opposite leg

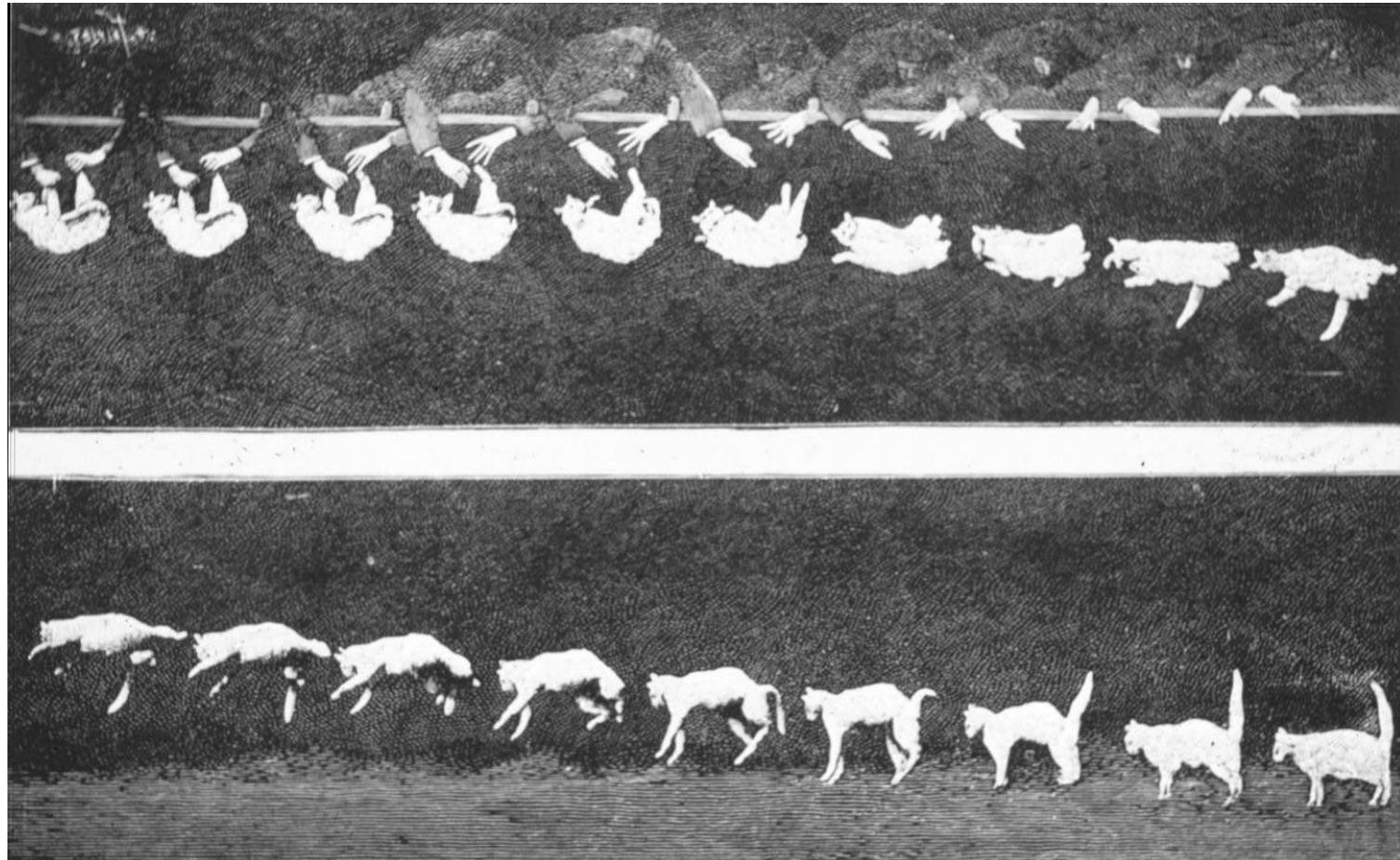


Source: <https://lightsmellscloud.wordpress.com/tag/etienne-jules-marey/>

# The Falling Cat Problem

Another fundamental problem answered by chronophotography:

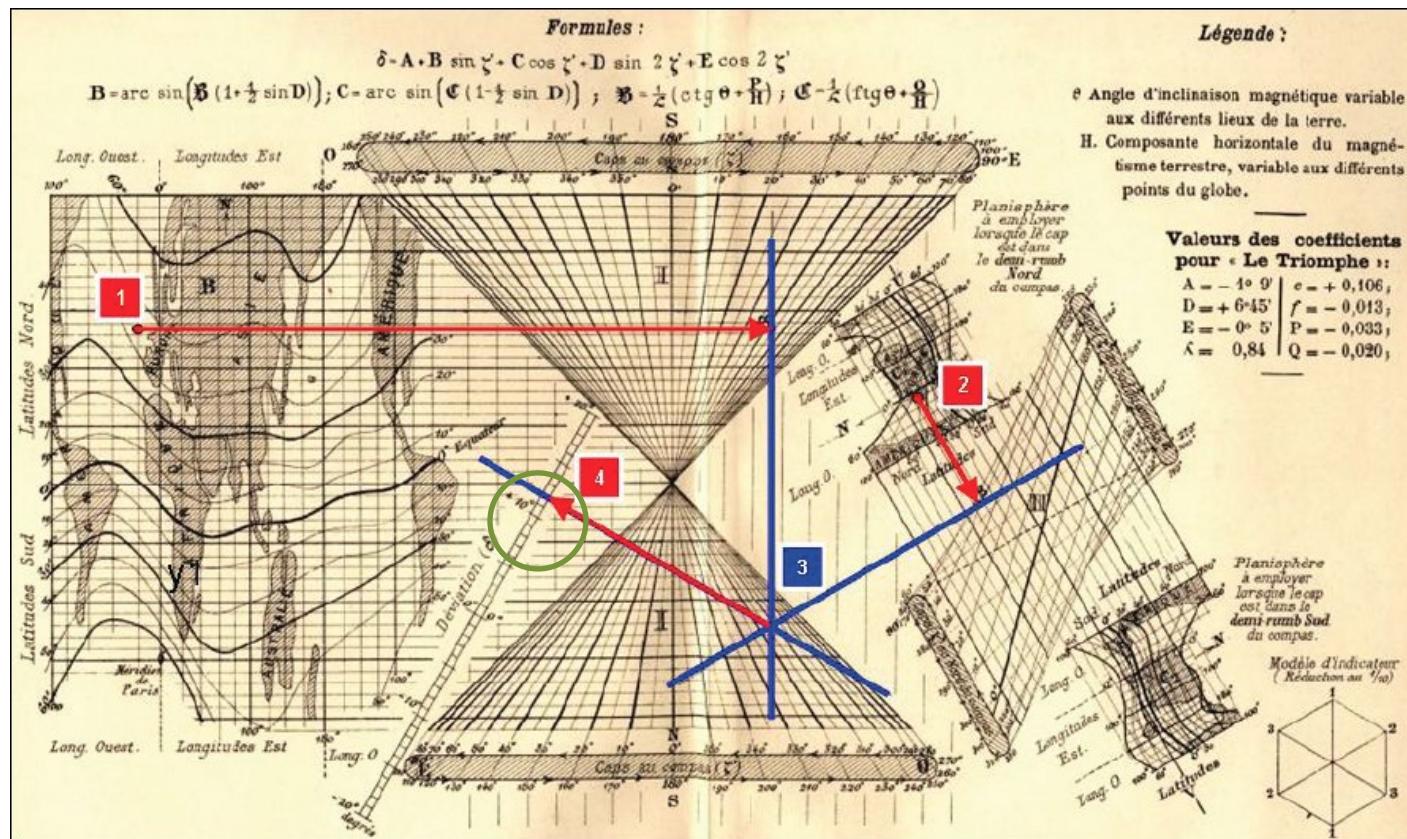
- How does a falling cat usually land on her feet? An OMG moment!



# Nomography

1885: Charles Lallemand, graphical calculator for compass course corrections of a ship at sea

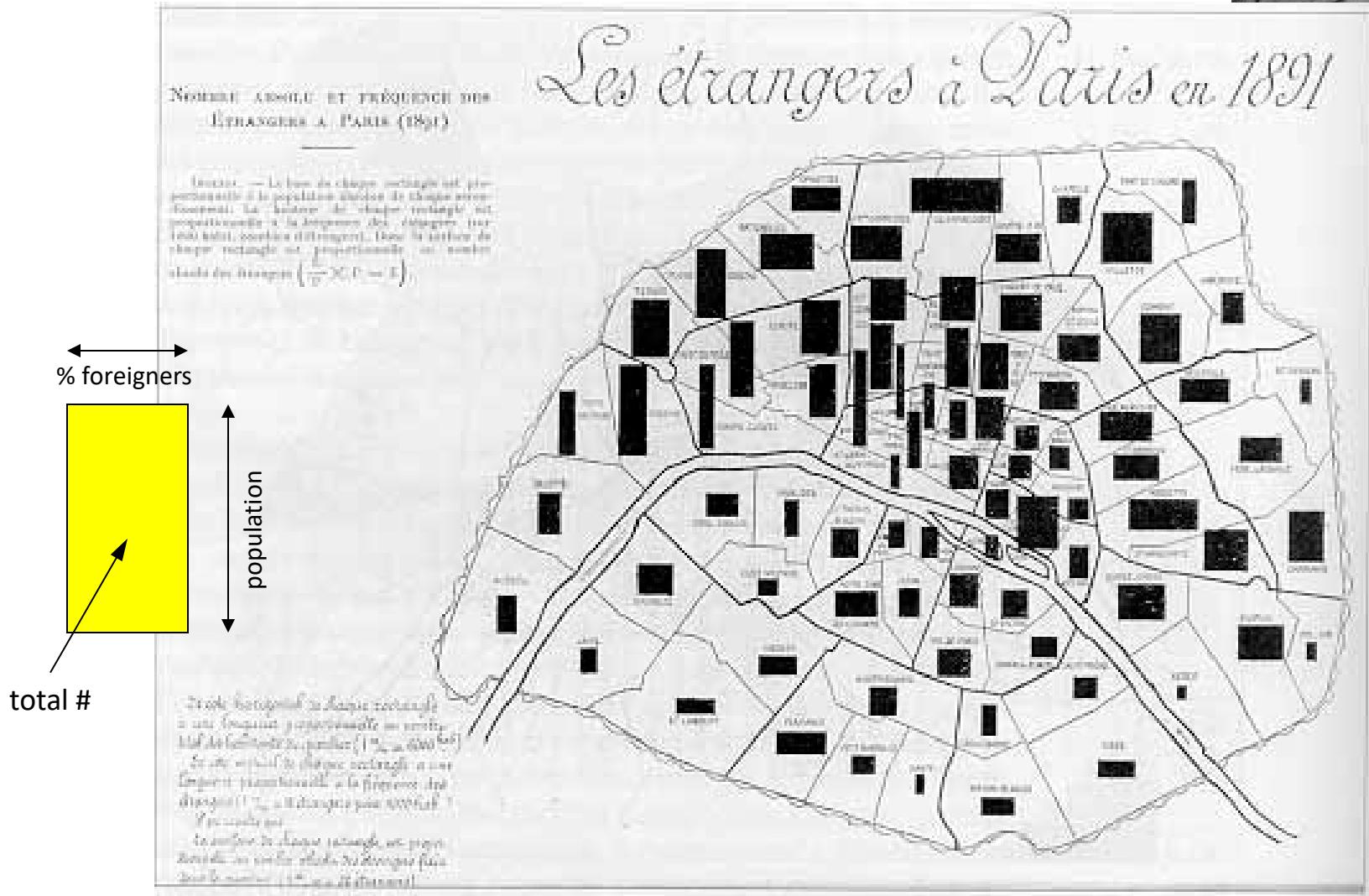
Combines: anamorphic maps, hexagonal coordinates, trigonometric scales (5 eqns)



Lallemand

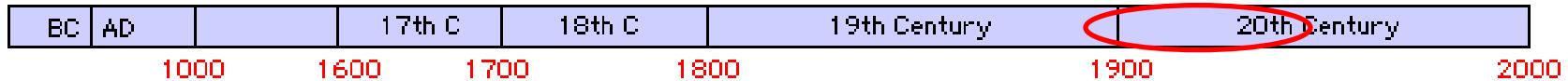


# 1896: Area rectangles on a map to display two variables and their product- Jacques Bertillon



# 1900-1949: The Modern Dark Ages

- By the 1930s, the growth of statistical methods supplanted enthusiasm for graphics
  - There were few graphic innovations
  - In statistics: numbers were precise; graphs were just “pretty pictures”
- But graphical methods had entered the mainstream & were popularized
  - Text books, college courses
- There were several graphic-based scientific discoveries
- Electronic computers were born



# 1900-1949: The Modern Dark Ages

**1914:** Brinton: *Graphic Methods for Presenting Facts*

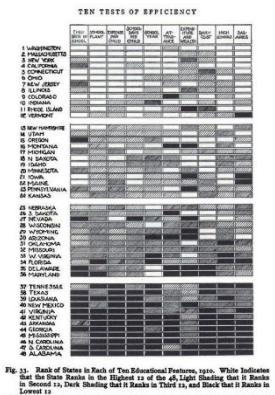
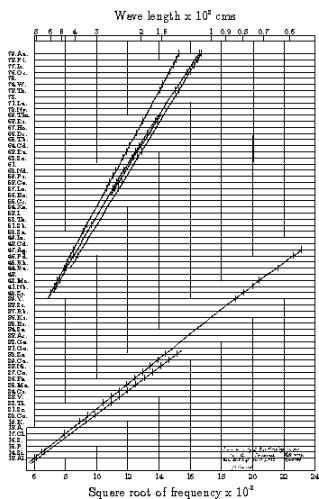
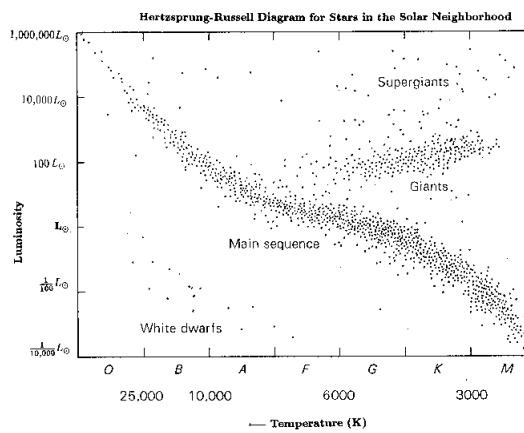


Fig. 21. Rank of States in Each of Ten Educational Features, 1920. White Indicates that the State Ranks in the Highest 12 of the 48; Light Shading that It Ranks in Second 12; Dark Shading that It Ranks in Third 12; and Black that It Ranks in Lowest 12.

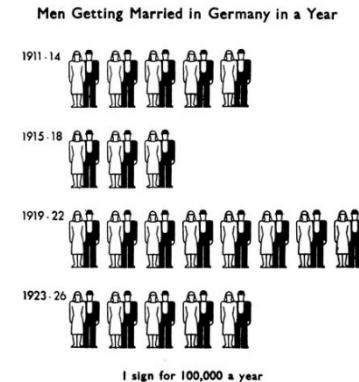
**1913:** Discovery of atomic number, based on graphical analysis- H. Moseley



**1911-1913:** The Hertzsprung-Russell diagram & evolution of stars



**1924:** ISOTYPE method of pictorial graphics—Otto Neurath



1 sign for 100,000 a year

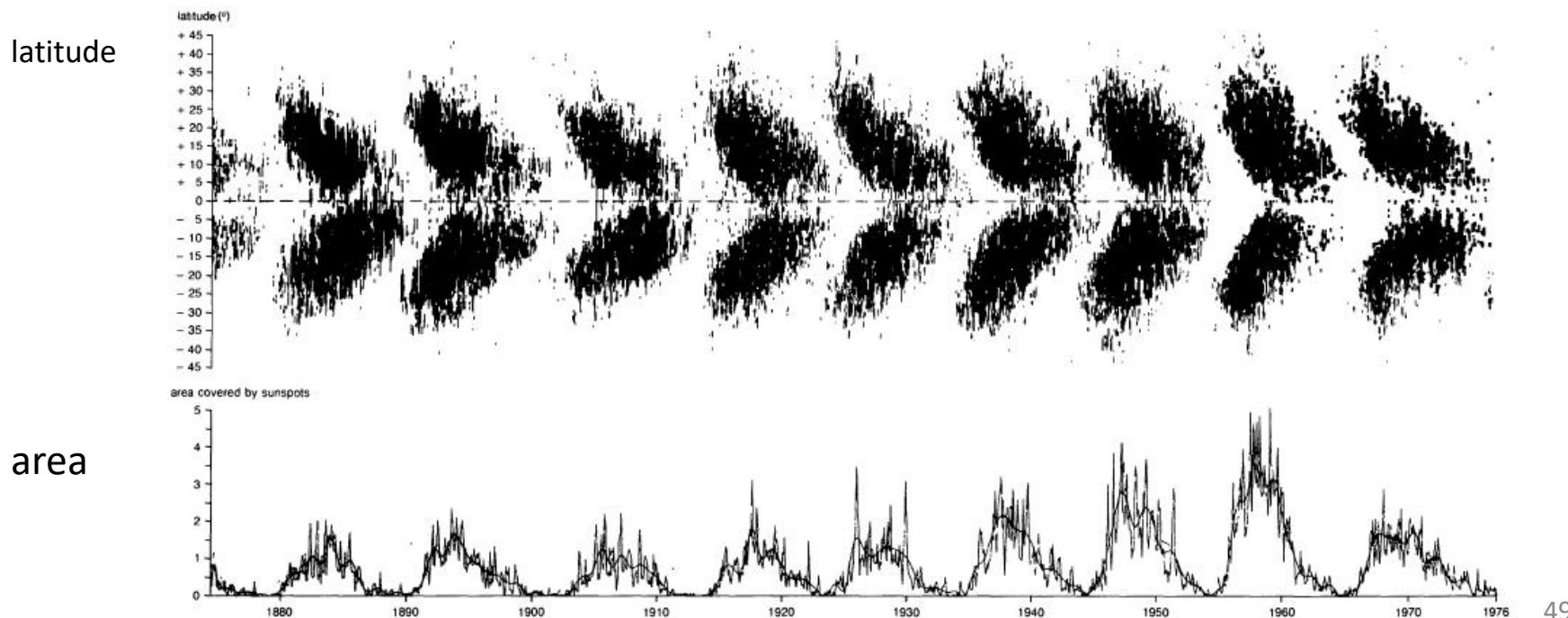
**1944:** Harvard's Mark I, the first digital computer- Howard Aiken, Grace Hopper



# Maunder: Butterfly diagram

1904: E.W. Maunder plots distribution of sunspots in sun's latitude by time

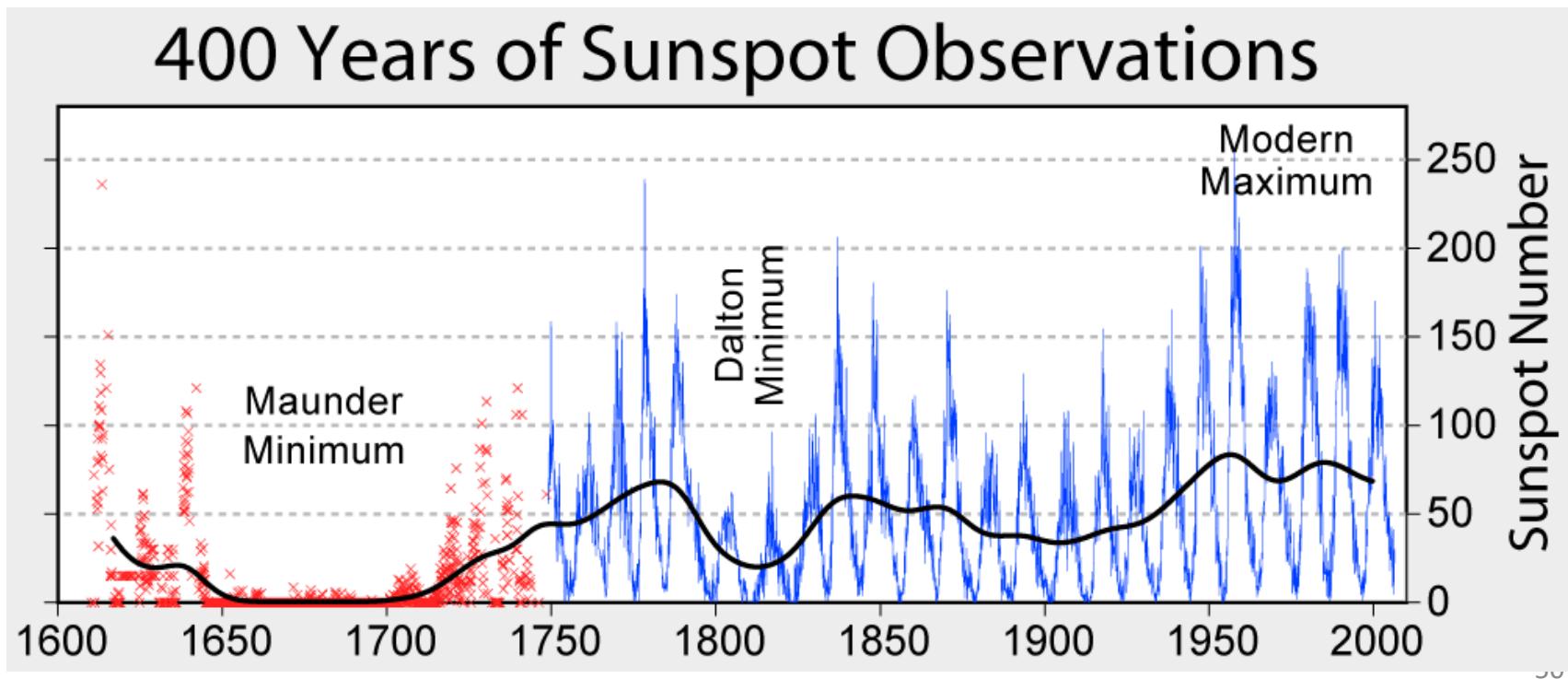
- Discovery of 11-year sunspot cycles (& 22-yr: reversal of sun's magnetic field)



# Maunder: Butterfly diagram

1904: E.W. Maunder plots distribution of sunspots in sun's latitude by time

- Discovery of “Maunder minimum” (1645-1715): “Little Ice Age”
- Smoothing reveals other extrema



**1914:** Willard C. Brinton publishes *Graphic Methods for Presenting Facts*, the 1<sup>st</sup> popular book on the topic



heatmap

TEN TESTS OF EFFICIENCY

	CHILDREN IN SCHOOL	SCHOOL-PLANT	EXPENSE PER CHILD	SCHOOL-DAYS PER CHILD	SCHOOL-YEAR	ATTENDANCE	EXPENDITURE AND WEALTH	DAILY COST	HIGH SCHOOLS	SALARIES
1 WASHINGTON										
2 MASSACHUSETTS										
3 NEW YORK										
4 CALIFORNIA										
5 CONNECTICUT										
6 OHIO										
7 NEW JERSEY										
8 ILLINOIS										
9 COLORADO										
10 INDIANA										
11 RHODE ISLAND										
12 VERMONT										
13 NEW HAMPSHIRE										
14 UTAH										
15 OREGON										
16 MONTANA										
17 MICHIGAN										
18 N DAKOTA										
19 IDAHO										
20 MINNESOTA										
21 IOWA										
22 MAINE										
23 PENNSYLVANIA										
24 KANSAS										
25 NEBRASKA										
26 S. DAKOTA										
27 NEVADA										
28 WISCONSIN										
29 WYOMING										
30 ARIZONA										
31 OKLAHOMA										
32 MISSOURI										
33 W. VIRGINIA										
34 FLORIDA										
35 DELAWARE										
36 MARYLAND										
37 TENNESSEE										
38 TEXAS										
39 LOUISIANA										
40 NEW MEXICO										
41 VIRGINIA										
42 KENTUCKY										
43 ARKANSAS										
44 GEORGIA										
45 MISSISSIPPI										
46 N. CAROLINA										
47 S. CAROLINA										
48 ALABAMA										

Fig. 33. Rank of States in Each of Ten Educational Features, 1910. White Indicates that the State Ranks in the Highest 12 of the 48, Light Shading that it Ranks in Second 12, Dark Shading that it Ranks in Third 12, and Black that it Ranks in Lowest 12.

pictogram

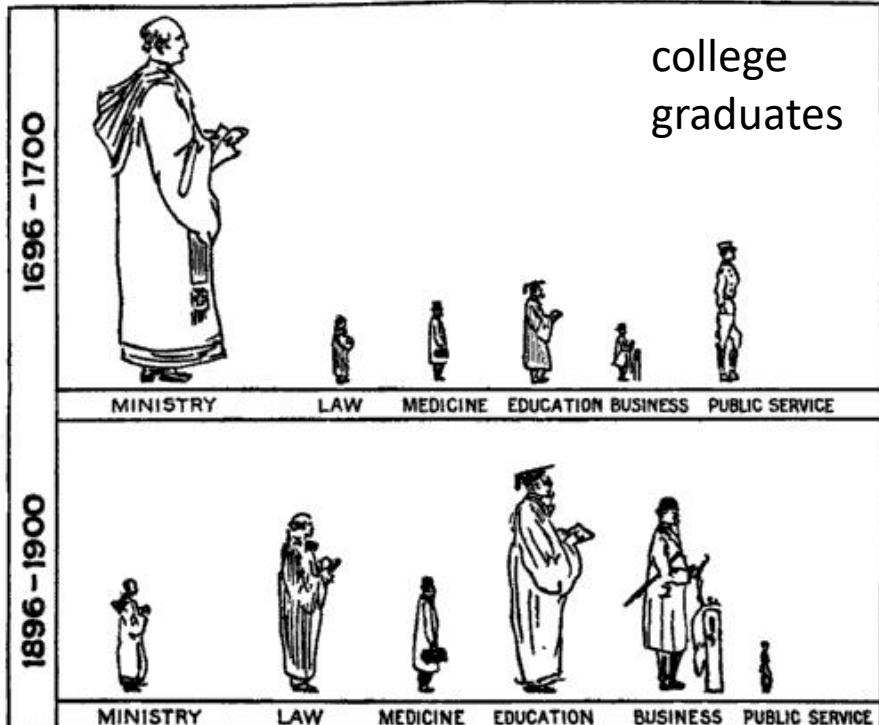


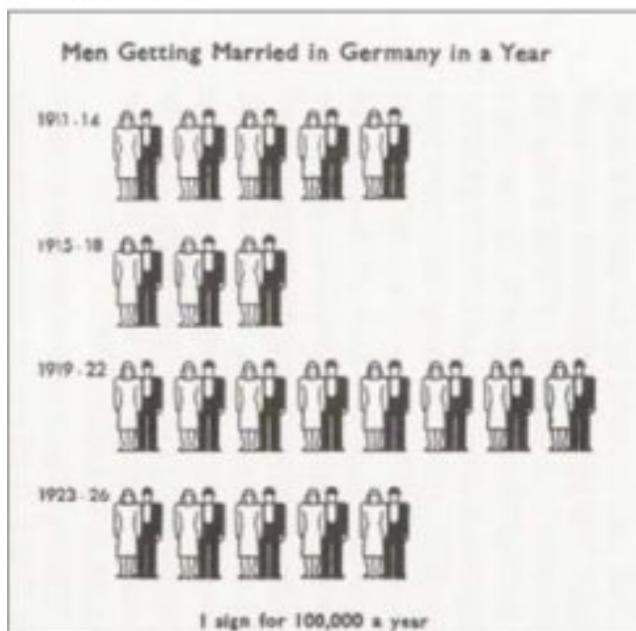
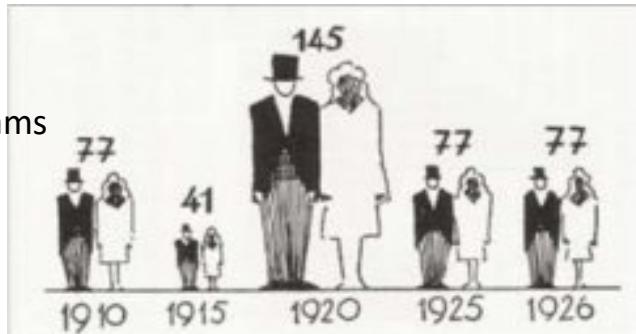
Fig. 39. Proportion of College Graduates in Different Professions in 1696-1700 and in 1896-1900

Charts of this kind with men represented in different sizes are usually so drawn that the data are represented by the height of the man. Such charts are misleading because the area of the pictured man increases more rapidly than his height. Considering the years 1696-1700, the pictured minister has about two and one-half times the height of the man representing public service. The minister looks over-important because he has an area of more than six times that of the man drawn to represent public service. This kind of graphic work has little real value.

**1924:** Otto Neurath developed the **Isotype** (International System of Typographic Picture Education) method to communicate statistical information to the broad public in an intuitive, pictorial way.

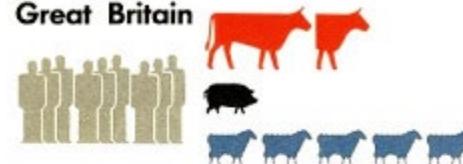


NOT  
pictograms



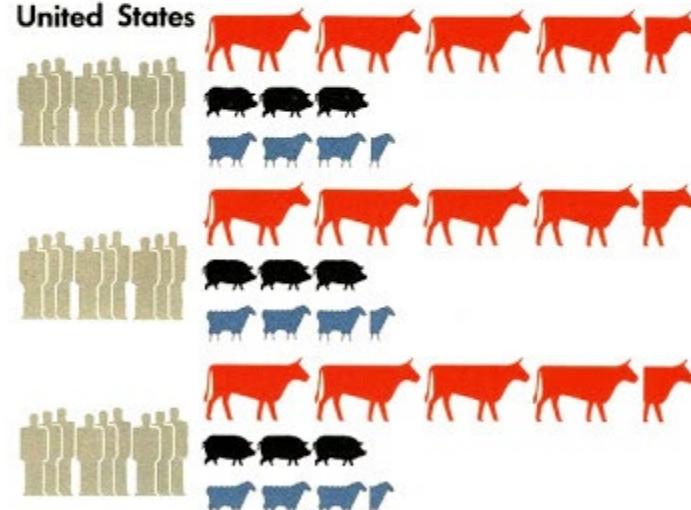
### Population and Live Stock

Great Britain



even complex,  
multivariate data

United States



Each grey figure represents 5 million population  
Each complete red symbol represents 5 million cattle  
Each complete black symbol represents 5 million pigs  
Each complete blue symbol represents 5 million sheep

Average for 1935 - 1939

## EXPLORATORY DATA ANALYSIS



# 1950-1974: Re-birth of graphics

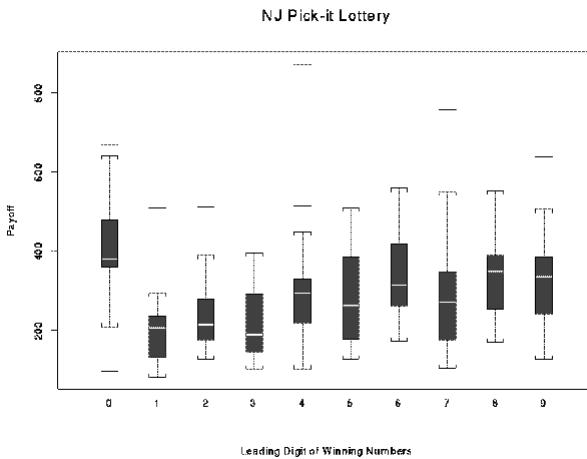
- Visualization began to rise from dormancy in the mid 1960s, spurred largely by:
  - J. W. Tukey's *Exploratory Data Analysis*: The power of graphics to show the unexpected in data analysis
  - Jacques Bertin's *Semiologie Graphique*: A general theory of composing graphs and maps
  - computer hardware for computation and display
  - the advent of statistical and graphics software

VISUAL VARIABLES	LEVEL OF ORGANIZATION		DEPLOYMENT MODE		
	PUNCTUAL	LINEAR	ZONAL		
SIZE	Q O ≠	• ● ●	— — —	■ ■ ■	■ ■ ■
VALUE INTENSITY	O ≠	○ ○ ○	— — —	■ ■ ■	■ ■ ■
GRANULATION	O ≠ ≡	■ ■ ■	— — —	■ ■ ■	■ ■ ■
ORIENTATION	≠ ≡	/   -	— — —	■ ■ ■	■ ■ ■
COLOR	≠ ≡	● ● ●	— — —	■ ■ ■	■ ■ ■
FORM	≠ ≡	▲ ● ■	— — —	■ ■ ■	■ ■ ■

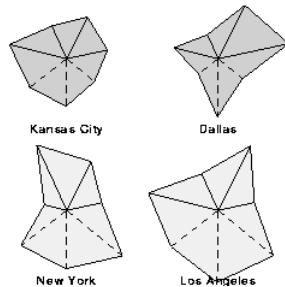


# 1950-1974: Re-birth of graphics

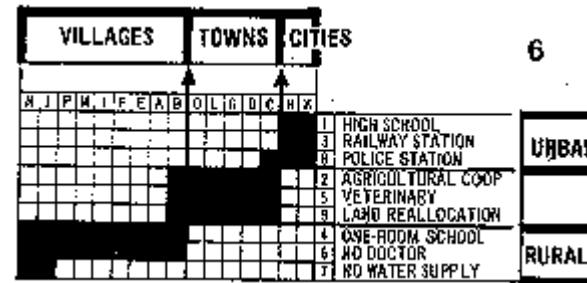
1969: Graphical innovations for EDA (stem-and-leaf, box-plots, etc.)- J.W. Tukey



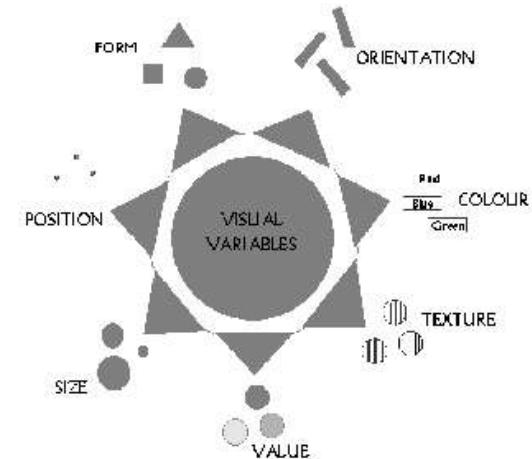
## Multivariate glyphs



1971: Star plots- J. H. Siegel et al

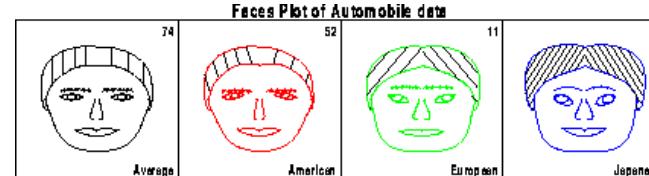


1967: Comprehensive theory of graphical symbols & graphics representation- Jacques Bertin



1967: Reorderable matrix- Jacques Bertin

1973: Face plots- Herman Chernoff



# Digital display devices

The biggest limitation in the early development of dynamic and interactive graphics was in graphics display devices.



Only B/W, but for the first time, **dynamic** displays became possible.

By the late 1950s, pen-like input devices allowed rudimentary direct **interaction**

# 1975-present

## Technology:

- Progressively more powerful computation & graphics
  - Mainframes → PCs → workstations → servers → cloud computing
  - pen plotters → CRTs → graphics hardware & firmware
  - stand-alone → client-server architecture
- Internet
  - email → bitnet -> file sharing (FTP) → www (HTML) → Java → javascript
  - data: open data initiatives (~1995) → APIs (census, health, ...)
  - eCommerce: Amazon, Netflix, ... → BIG data, recommender systems
- Software
  - Graphics packages: SYSTAT, Data Desk, XGobi, ViSta
  - Statistical packages: SAS, SPSS
  - Statistical programming environments: R, matlab, Stata
  - Contributed package archives: CTAN (latex), CPAN (perl), CRAN (R)
  - Collaborative development sites: github, bitbucket, ...

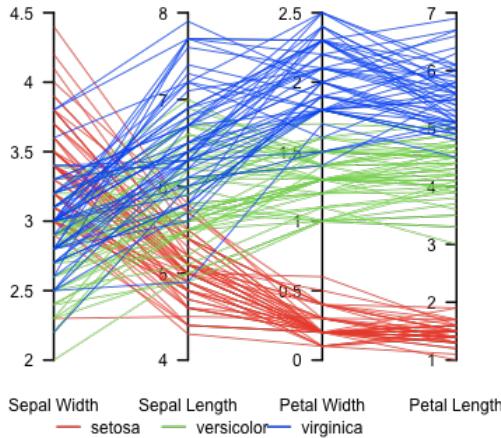
# 1975-present

Themes in data visualization:

- high-D problems of progressively higher dimensions
  - grand tour: n-D → 2D projections
  - Dimension reduction methods (PCA, MDS, biplots)
- graphics & methods for other data types:
  - categorical, frequency data,
  - networks, trees, ...
  - text (word clouds, ...)
  - spatial data & models
- interactive data vis
  - linked views
  - direct manipulation: select, zoom, filter
  - dynamic graphics & animation

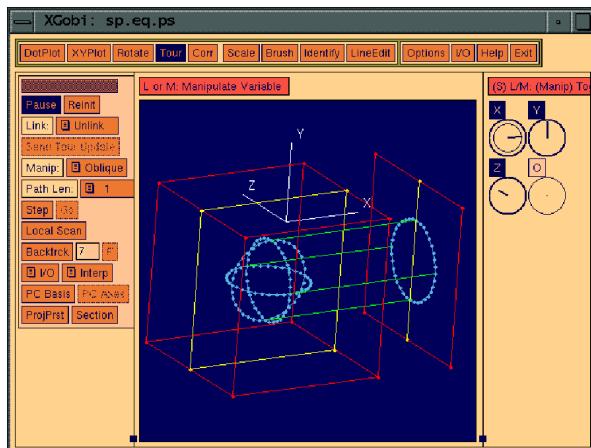
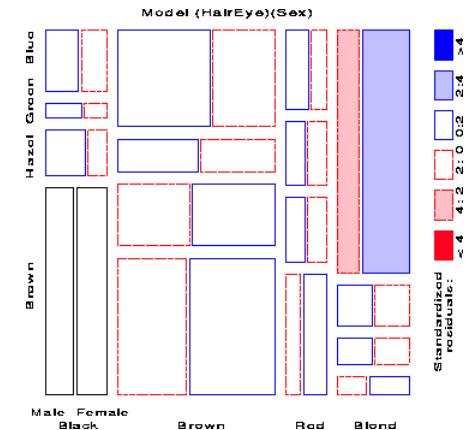
# 1975-present

Parallel coordinate plot, Fisher's Iris data



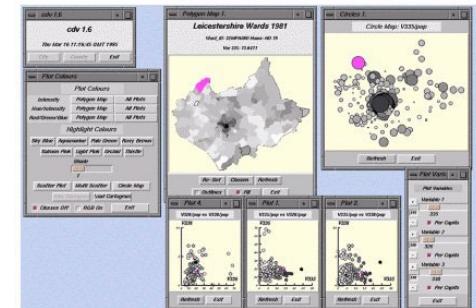
1985: Parallel coordinates plots for high-D data- Alfred Inselberg

1991: Mosaic display for visual analysis of log-linear models- Michael Friendly



1991-1996: High-interaction systems for data analysis and visualization, e.g., *XGobi*, *ViSta*

1996: Cartographic Data Visualiser – Jason Dykes



# Tukey: PRIM-9

1973: a group at the Stanford Linear Accelerator developed PRIM-9

- Picturing, Rotating, Isolation, Masking in up to 9 dimensions
- \$400K graphic display & keypad; computations on a mainframe, \$500/hr



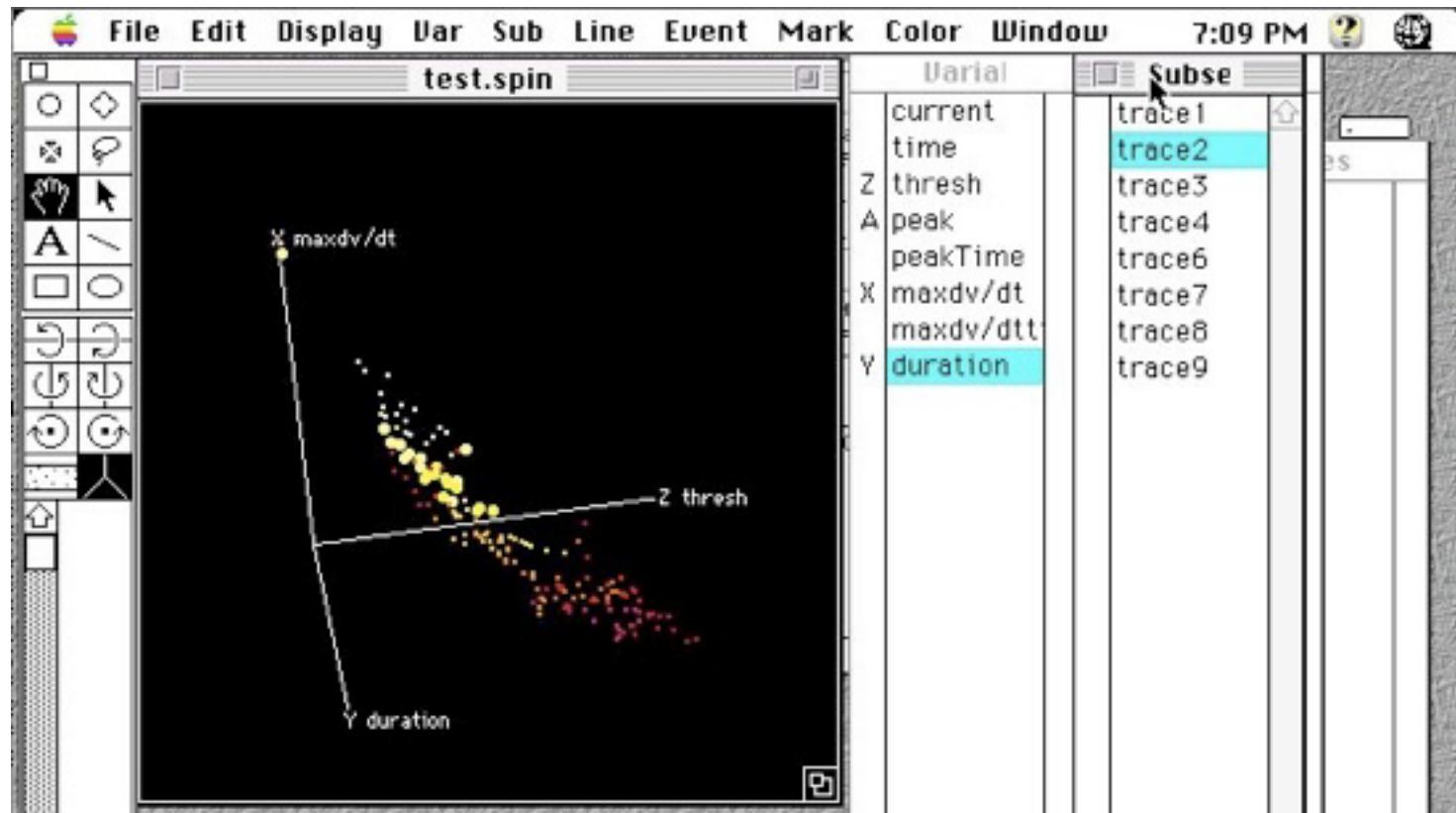
# Next steps: Hardware

- Dynamic 3D graphics was painfully slow for larger data sets.
- Specialized 3D graphics hardware:
  - Early 1970s: Simple LSI graphics chips for video games
  - 70s—80s: Graphics co-processors (GPUs) with increasing graphics capabilities
  - 80s—90s: Silicon Graphics develops high-performance 3D graphics workstations



# Software

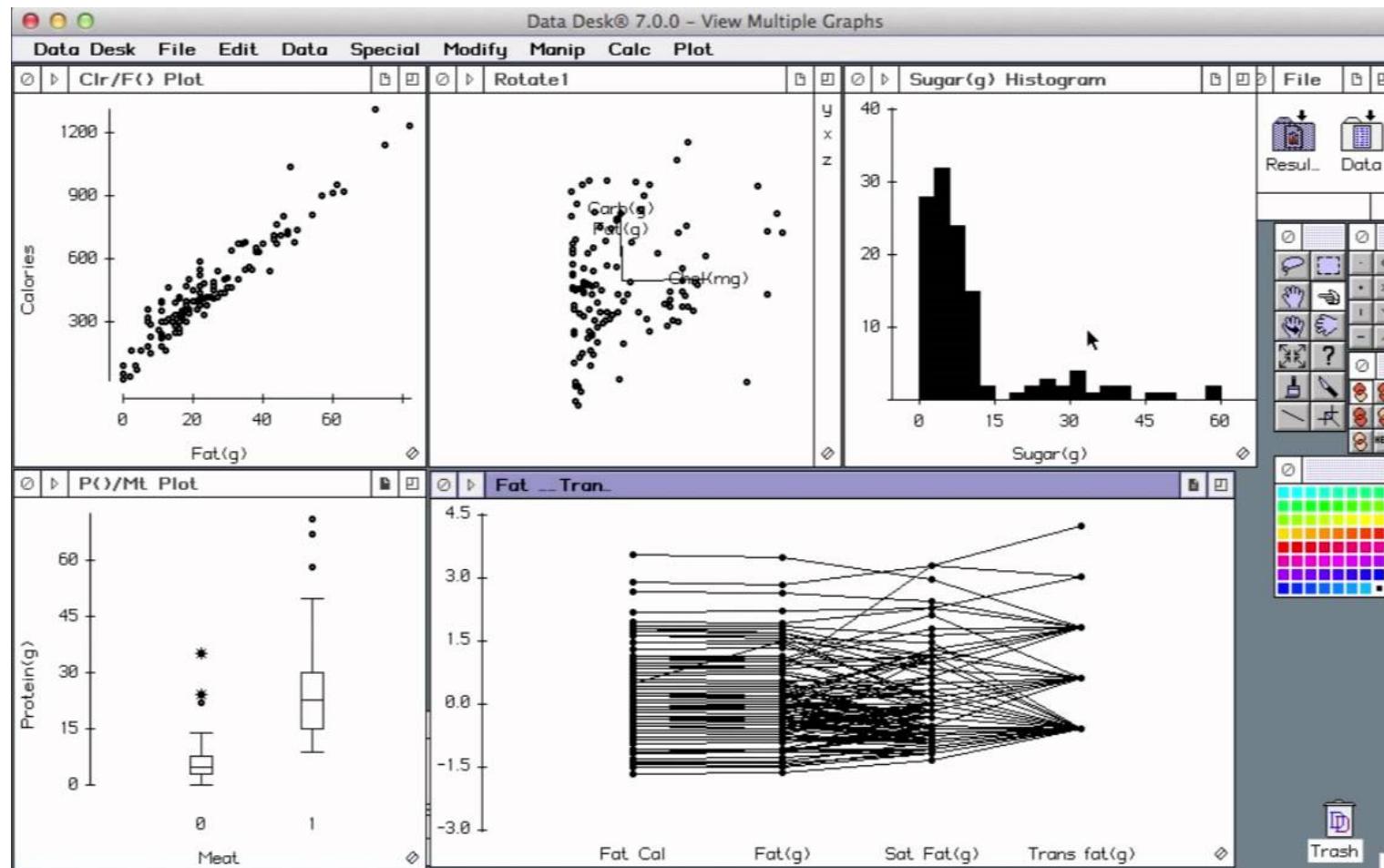
- MacSpin – Andrew & David Donoho (1984—85). At ASA meetings 1986, “dynamic graphics became as portable as a 25-lb Macintosh”



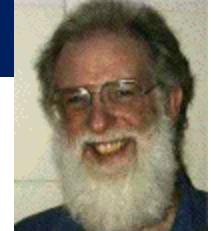
# Linking, brushing, 3D rotation

Paul Velleman (~1985): Data Desk provided multiple 1D, 2D, 3D views

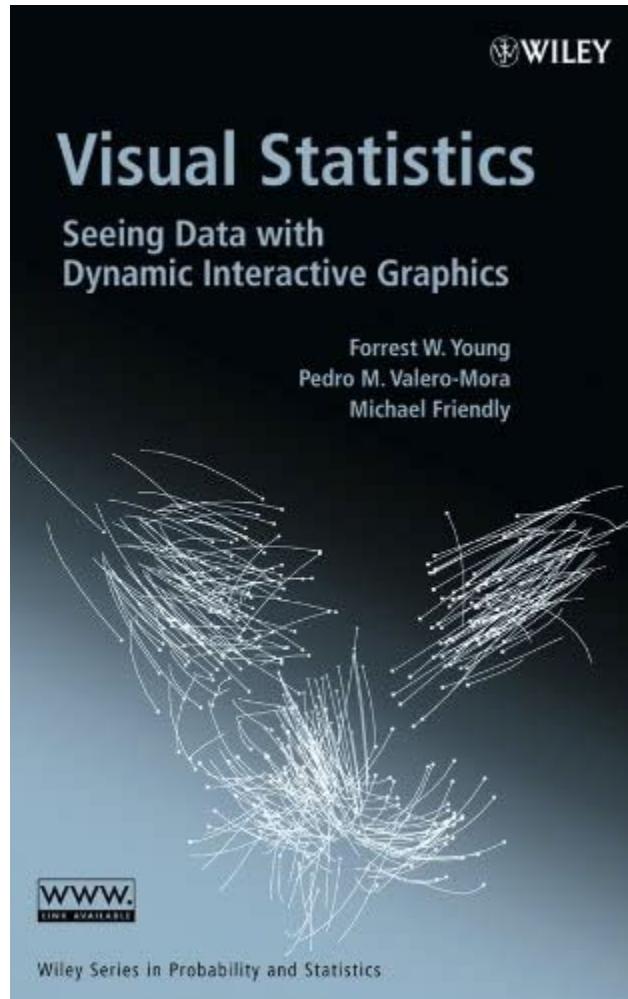
- **Brushing:** selection of points, regions, ... via mouse
- **Linking:** Any action in one plot reflected in all others



# Visual Statistics



Young, Valero-Mora & Friendly (2006)



A philosophy & pedagogy for statistics based on dynamic interactive graphics

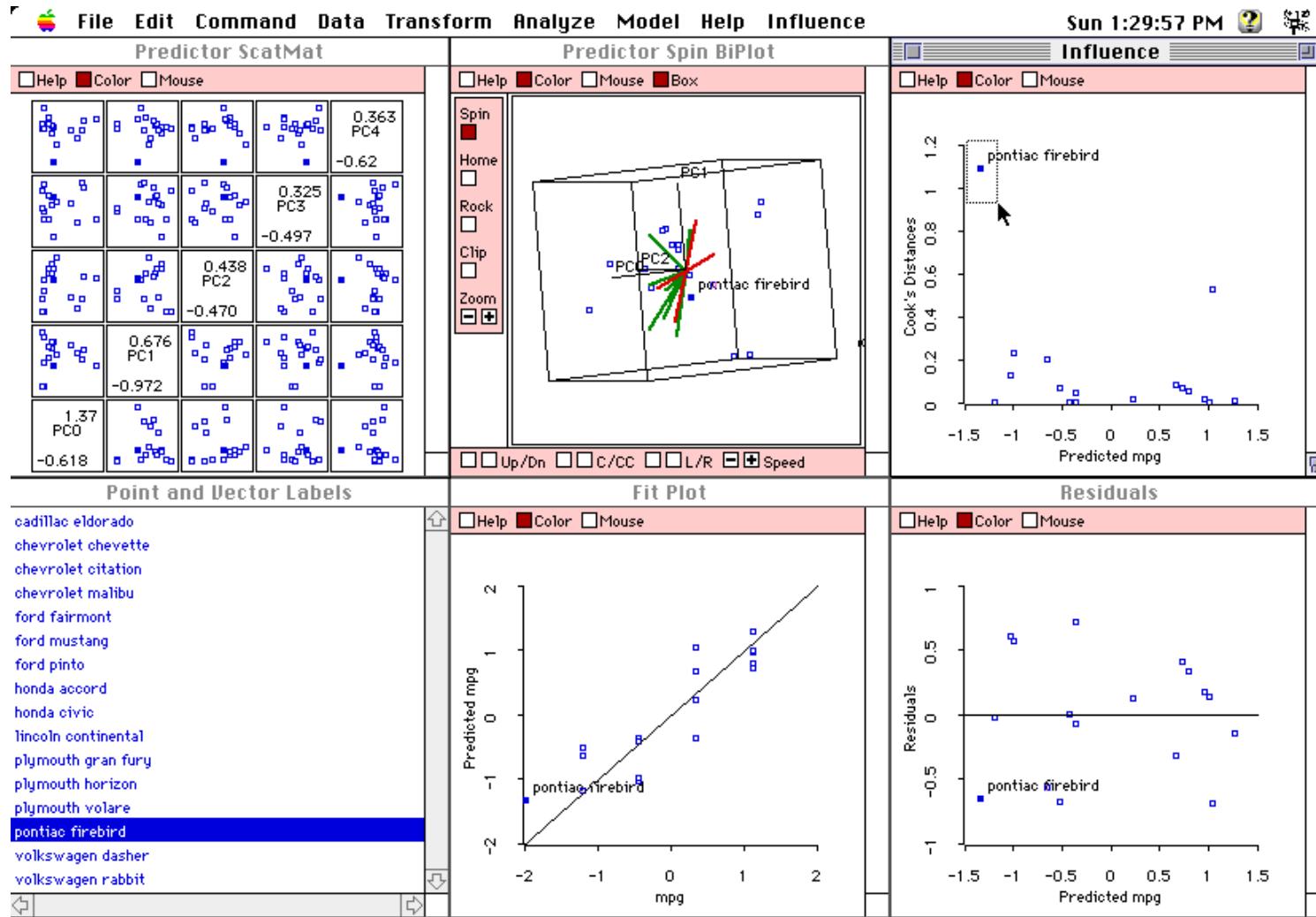
A theory of #datavis software:

- objects (data, model, ...)
- methods (print, plot, )
- manipulating plot objects & dimensions
- spin plots: rotating 3D plots
- spreadplots: dynamically linked views
- workmaps: visual record of analysis steps

Details: <https://www.uv.es/visualstats/>

See: The History of ViSta: The Visual Statistics System,  
<https://onlinelibrary.wiley.com/doi/full/10.1002/wics.1203>

# ViSta: Visual Statistics





# Summary

- Data Visualization has deep & wide roots:
  - **Cartography:** map-making, geo-measurement, thematic cartography, GIS, geo-visualization
  - **Statistics:** probability theory, distributions, estimation, models, stat-graphics, stat-visualization
  - **Data:** population, economic, social, moral, medical, ...
  - **Visual thinking:** geometry, functions, mechanical diagrams, EDA, ...
  - **Technology:** printing, lithography, computing...
- **Problem driven:** developments often driven by practical and theoretical problems of the day
- **Communication driven:** developments often arose from a desire to communicate better

# Conclusions

- Why study the history of data visualization?
  - “The only new thing in the world is the history you don’t know” – Harry S. Truman
  - “Those who cannot remember the past are condemned to repeat it.” – George Santayana (*The Life of Reason*, 1905)
  - “No scientific discovery is named after its original discoverer” – Stigler’s Law of Eponomy (1980). But: originally due to Merton!
- Today:
  - Narrow, specialized work in many fields
  - New methods “invented” and re-named w/o knowing history.
    - mosaic displays: Georg von Mayr (1877)
    - heatmaps: Loua (1873); Brinton (1914), Bertin (1967)
    - Nightingale (1859) rose diagram: polar diagrams by Guerry (1829), Lalanne (1843)