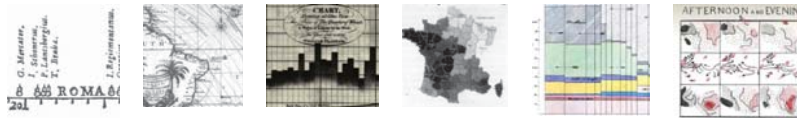## The *Milestones Project*: 7±2 Stories from the History of Data Visualization
[Old Wind into New Bottles]

Michael Friendly

York University

Geographic Visualization Across the Social Sciences
June 12, 2006

(Slides at: www.math.yorku.ca/SCS/Papers/7Stories-2x3.pdf)

Plan & Outline

- Introduction
  - Inspirational orienting ideas & widgets
  - Milestones Project
  - Whirlwind Milestones Tour
- Stories

```
Logo> TO TellStories :STORIES
        if empty? :STORIES [STOP]
        if TimeRunsOut? [STOP]
        #
        Tell first :STORIES
        TellStories BUTfirst :STORIES
        END

Logo> STORIES <- load("./gv2006.txt")
Logo> TellStories list( Langren Galton Playfair
                        Minard ... )
Logo> quit
```

2

Themes

- History of data visualization
- Graphical communication & excellence
  - Tasks: Lookup, comparison, detection
  - Effective visual comparisons
  - Effect ordering
- Statistical historiography
  - History as data
  - Understanding through reproduction
  - How to visualize a history?

3

Dimensions of Info Presentation

Where do we want to be?

NONE	0: R^0	• moment in time ($\triangle t$)	Something happens
ONE	1: R^1	• time (but, on what scale?) • oral presentation • avg(PPT) presentation	
TWO	2: R^2	• static graph/map • 2D timelines • written paper? (interactive skipping)	
MANY	2+: R^p	• Dynamic statistical graphics (ViSta, ggobi, R+extensions) • Web: java, Flash, blogs, wikis • PDAs, phones, GPS devices	
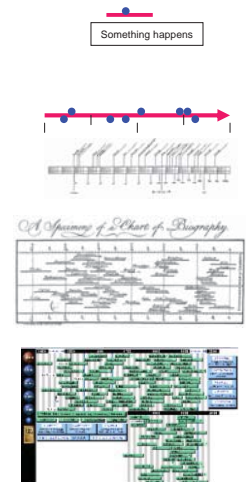
4

Milestones: Project Goals

- Comprehensive catalog of historical developments in *all fields* related to data visualization.
- → Collect representative bibliography, images, cross-references, web links, etc.
- → Enable researchers to find/study themes, antecedents, influences, patterns, trends, etc.
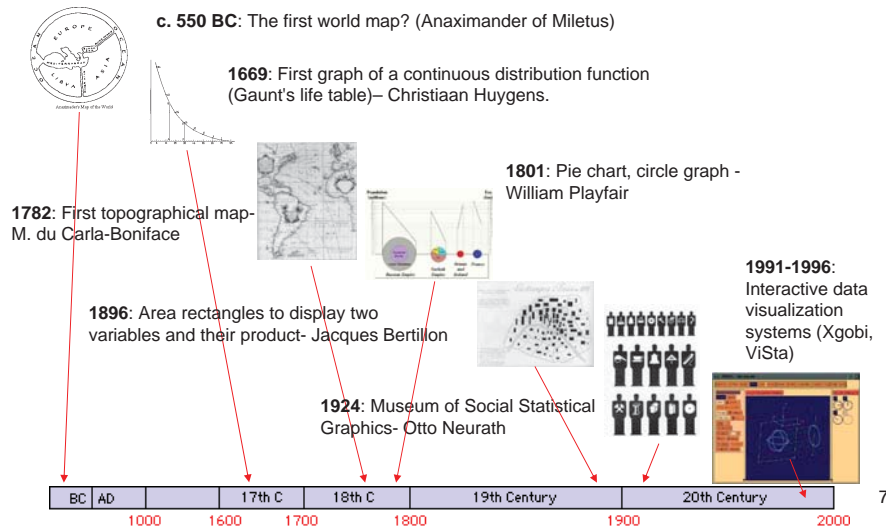- Web: http://www.math.yorku.ca/SCS/Gallery/milestone/

5

Milestones: Conceptual Overview

- Roots of Data Visualization
  - **Cartography**: map-making, geo-measurement, thematic cartography, GIS, geo-visualization
  - **Statistics**: probability theory, distributions, estimation, models, stat-graphics, stat-vis
  - **Data**: population, economic, social, moral, medical, …
  - **Visual thinking**: geometry, functions, mechanical diagrams, EDA, …
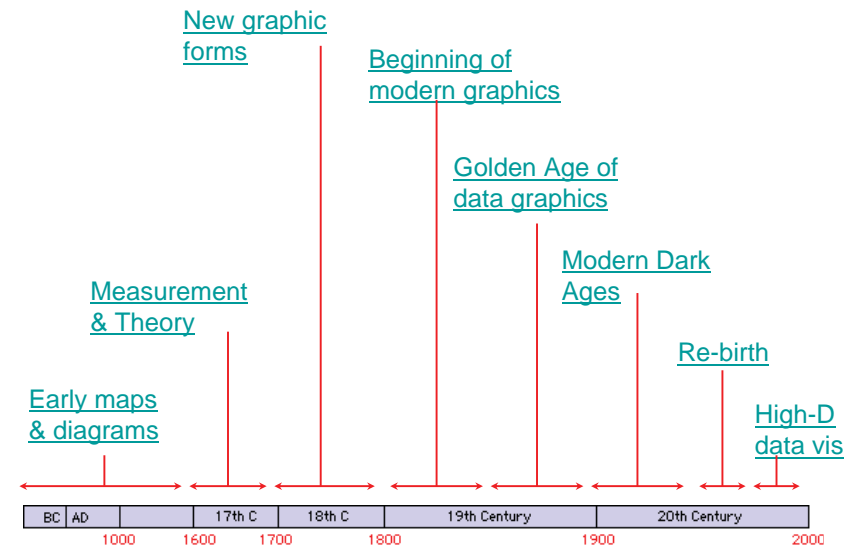 - **Technology**: printing, lithography, computing…

6

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.

1801: Pie chart, circle graph - William Playfair

1782: First topographical map- M. du Carla-Boniface

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

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

1924: Museum of Social Statistical Graphics- Otto Neurath

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

7

Milestones Tour

New graphic forms

Beginning of modern graphics

Golden Age of data graphics

Modern Dark Ages

Re-birth

Measurement & Theory

Early maps & diagrams

High-D data vis

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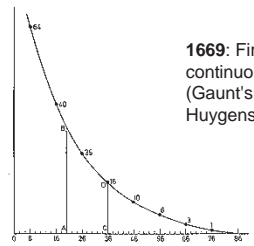1000 | 1600 | 1700 | 1800 | 1900 | 2000

8

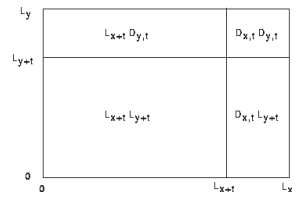## 1600-1699: Measurement and Theory

1626: Visual representations used to chart the changes in sunspots over time- Christopher Scheiner
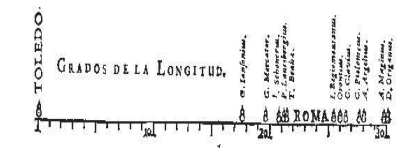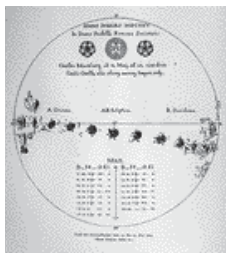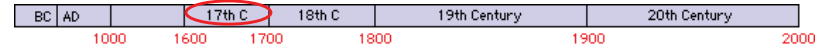
1644: First visual representation of statistical data- M.F. van Langren, Spain

1669: First graph of a continuous distribution function (Gaunt's life table)– Christiaan Huygens.
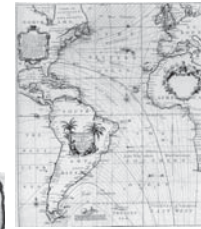
**1693**: First use of areas of rectangles to display probabilities of independent binary events- Edmund Halley, England
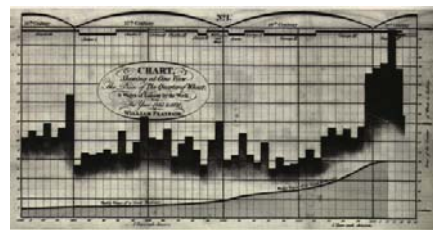
L_y			
	$L_{x+t} D_{y,t}$	$D_{x,t} D_{y,t}$	
L_{y+t}			
	$L_{x+t} L_{y+t}$	$D_{x,t} L_{y+t}$	
0		L_{x+t}	L_x

BC	AD		17th C	18th C	19th Century	20th Century	
	1000	1600	1700	1800	1900	2000	9

1700-1799: New graphic forms

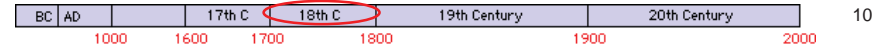**1701**: Isogon map, lines of equal magnetic declination – Edmund Halley

1765: Historical time line (life spans of famous people) Joseph Priestley
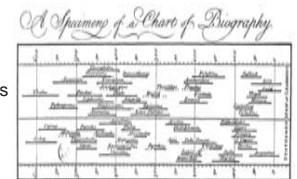
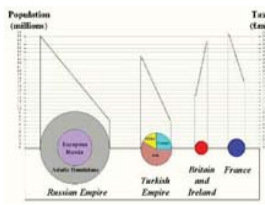**1786**: Bar chart, line graphs of economic data- William Playfair

1782: First topographical map- Marcellin du Carla-Boniface

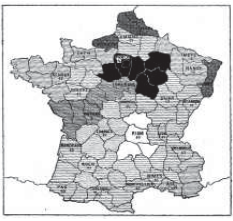| BC	AD		17th C	18th C	19th Century	20th Century	
	1000	1600	1700	1800	1900	2000	10

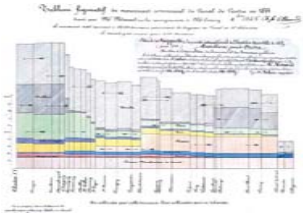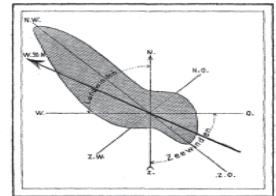## 1800-1849: Beginning of modern data graphics

1801: Pie chart, circle graph invented- William Playfair

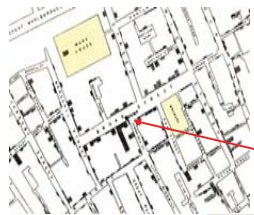**1819**: First modern statistical map (illiteracy in France)- Charles Dupin

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

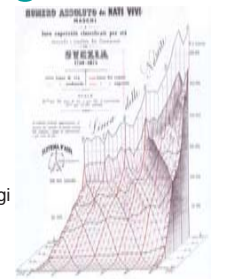**1844**: variable-width, divided bars, area ~ cost of transport- C. J. Minard

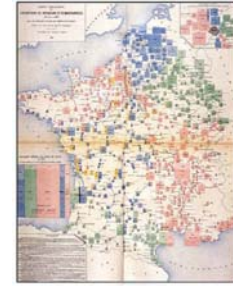| BC	AD		17th C	18th C	19th Century	20th Century	
	1000	1600	1700	1800	1900	2000	11

1850-1899: Golden Age

1855: Dot map of disease data (cholera)- John Snow
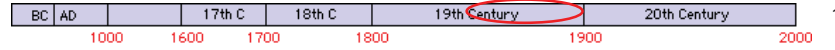
Broad St. pump

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

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

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

BC	AD		17th C	18th C	19th Century	20th Century	
	1000	1600	1700	1800	1900	2000	12

Data and Stories

- Statistical analysis: Data + Story
- Conf. presentation: .4*Data + .6*Story ?
- Historical analysis: .2*Data + .8*Story ?
    - Meta-Q: What is the "data" of history?
- A one-parameter family?
    - Wischenschaftliche-Forschung: *WF = p * Data + (1-p)*Story*
- Where are the pictures?
- Statistical graphics: Data + Picture + Story
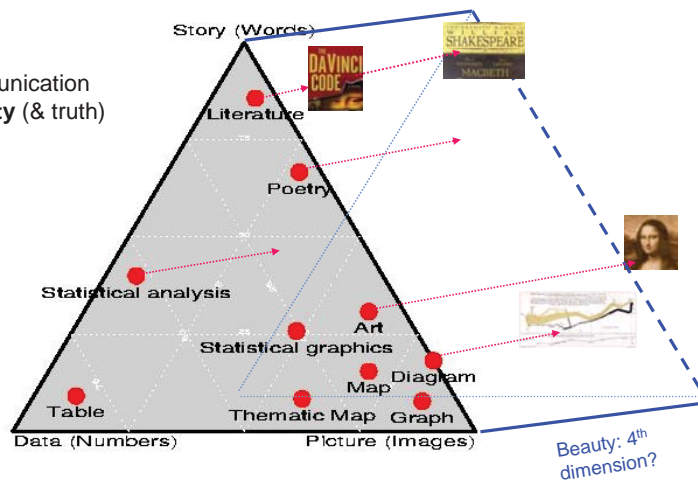
Data, Stories and Pictures

Modes of communication, as composed of data, stories and pictures in different proportions
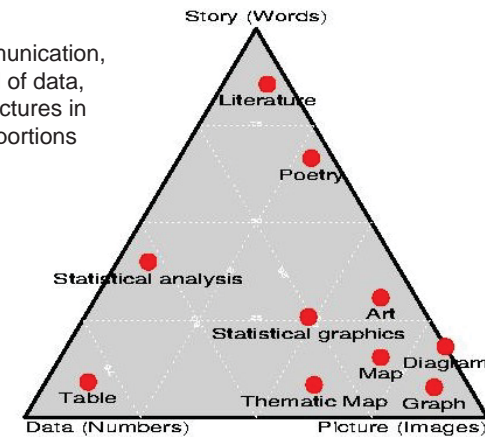
Data, Stories and Pictures

Modes of communication also vary in **beauty** (& truth)

Beauty: 4th dimension?

Logo> show STORIES

- Why the 1st statistical graph got it right
- Escaping flatland: ISOs from Halley to Galton
- Playfair's graphic sins and blunders
- The day social science was invented
- Why Minard?
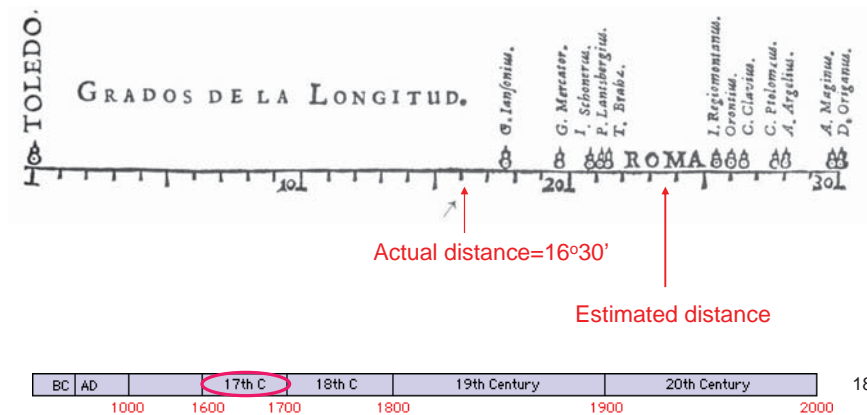- Graphic discoveries
- Visions of time & history

Slide 17

Logo> `plot STORIES`

Milestone Story Timeline

Slide 18
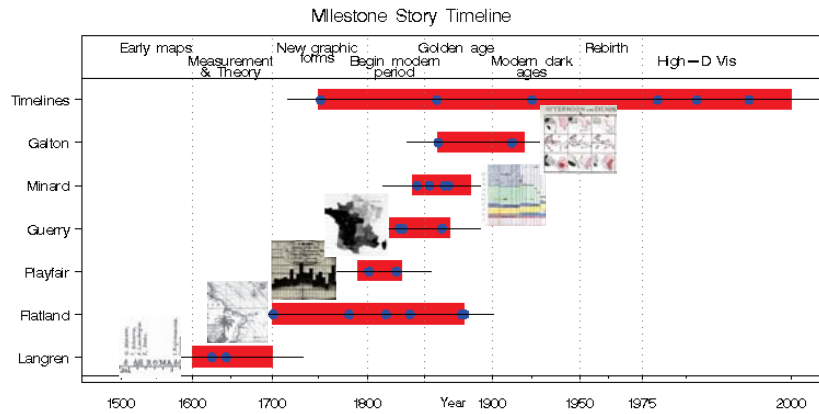
Why the 1st statistical graph got it right

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

Actual distance=16°30'

Estimated distance

Slide 19

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	
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	
1567	Clavius, C.	26.5	Germany
1578	Brahe, T.	21.5	
1582	Maginus, A.	29.8	Italy
1601	Organus, D.	30.1	

Slide 20

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

Sorted by Authority

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

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

Sorted by Longitude

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

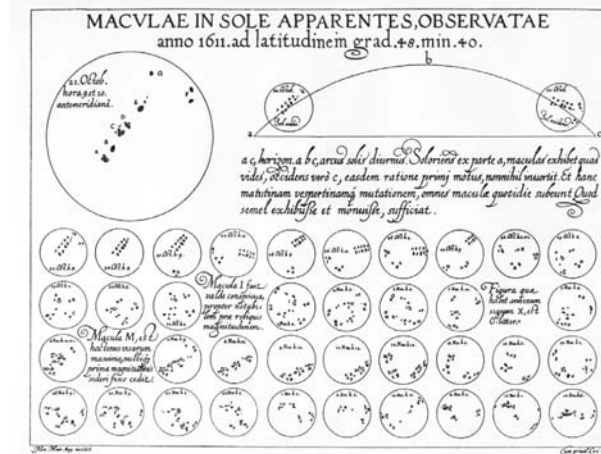## Only a graph shows…

- central location
- bias
- name labels

- wide variability
- clustering, detached observations
- (Did Tukey do this?)

range

bias

Actual distance=16º30'

Estimated distance
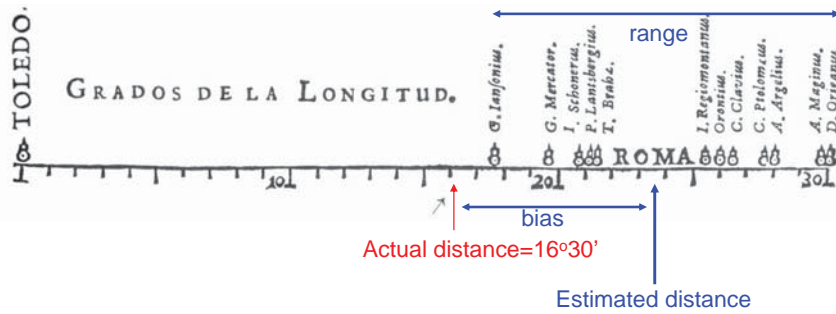
See: Friendly, M., & Kwan, E. (2003). Effect Ordering for Data Displays. *Computational Statistics and Data Analysis, 43*(4), 509--539.

21

Another early graphical success

1626: Scheiner's representation of **changes** in sunspots over time.

- "small multiples"
- allows comparison
- multiple legends
- A+ for info design!

22

Escaping flatland: ISOs from Halley to Galton

23

1701: Halley's contour maps showing curves of equal value (an isogonic map: lines of equal magnetic declination for the world) -- possibly the first thematic contour map of a data-based variable.

24
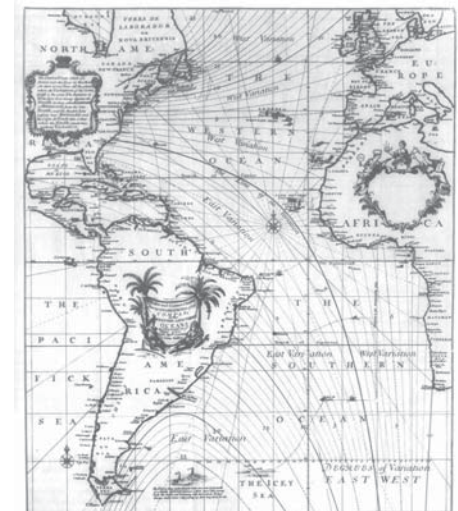
1782: **Marcellin du Carla, France**
The first topographical map, showing contours of equal elevation.
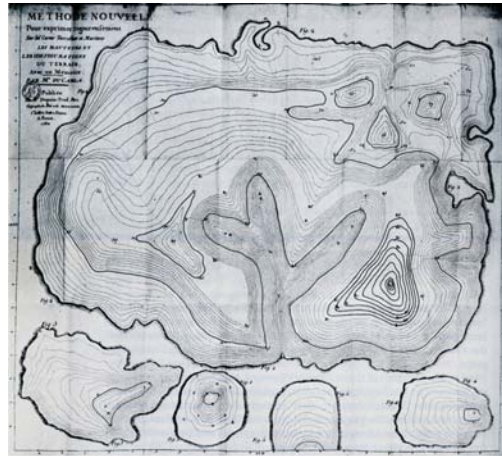
25

1817: **Alexander von Humboldt (1769-1859)** Germany
First **graph** of isotherms, showing mean temperature around the world by latitude and longitude.
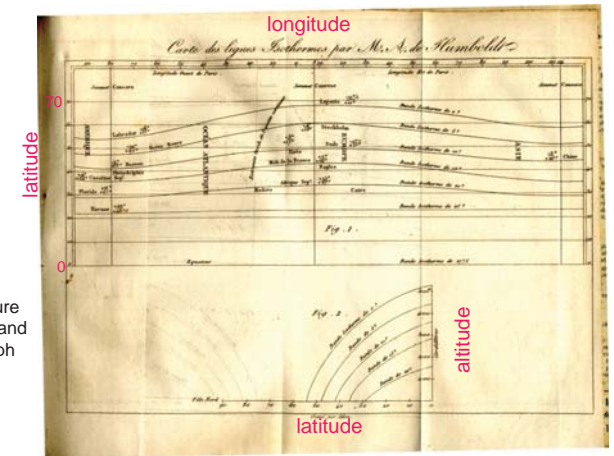
Recognizing that temperature depends more on **latitude** and **altitude**, a subscripted graph shows the direct relation of temperature on these two variables
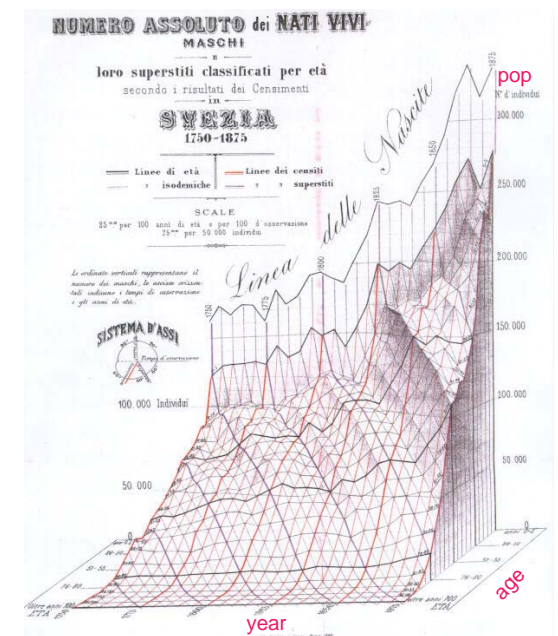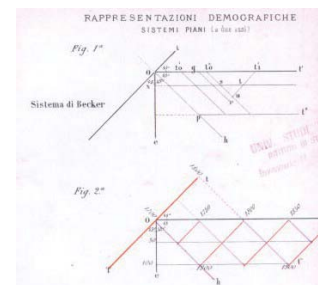
26

1843: Léon Lalanne (1811-1892) France
Contour diagram of a 3D table: temperature ~ hour x month
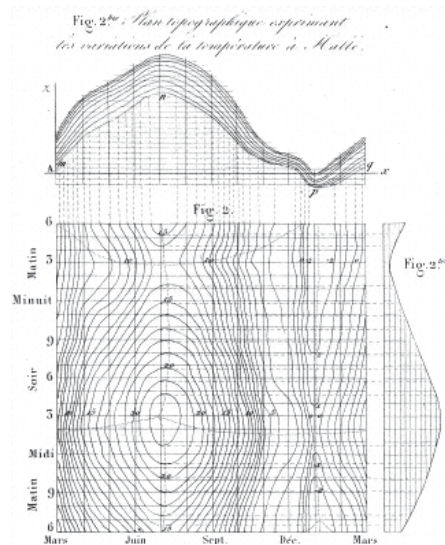
27

1879: Luigi Perozzo Italy
Stereogram (3D population pyramid) modeled on actual data (Swedish census, 1750--1875)
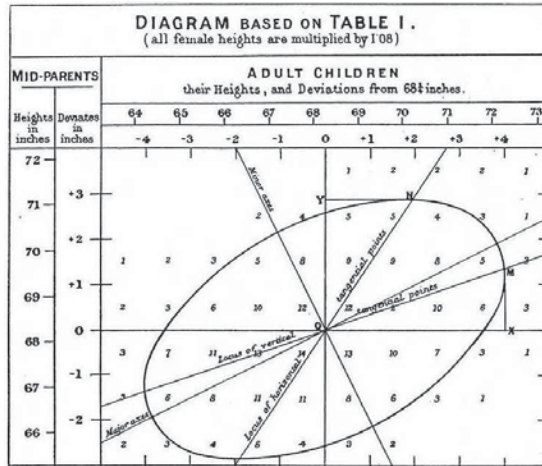
Construction diagrams for 3D surfaces

Galton's discovery of the bivariate normal correlation surface (1886)

• Level curves are **ellipses**
• Regression lines are loci of conjugate **tangents**
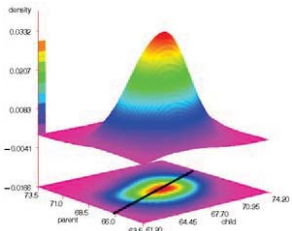
… that Galton should have evolved all this … is to my mind one of the most note-worthy scientific discoveries arising from analysis of pure observation (Pearson 1920, p37)

Galton (1886, Pl X): Smoothed contours of heights of parents and children
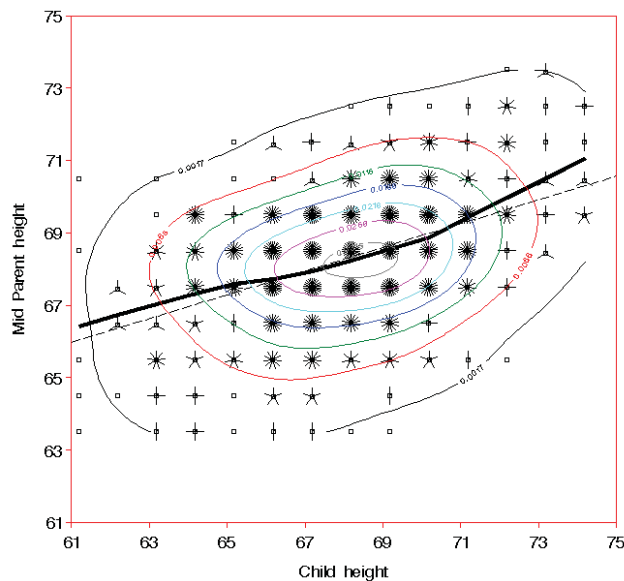
29

How did Galton reach this conclusion?

Literal application of Galton's smoothing method does not suggest "concentric ellipses" or lines of means as conjugate axes
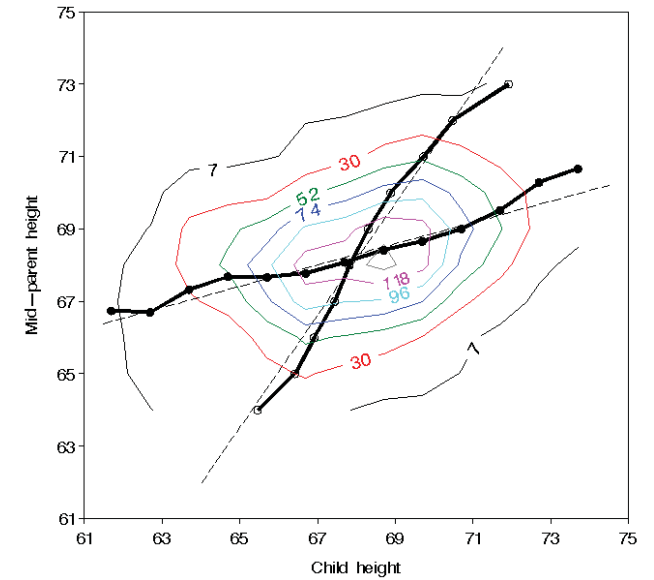
How did Galton reach this conclusion?

Modern smoothing methods (kernel density estimate) suggests that Galton:
• smoothed by 'eye & brain'
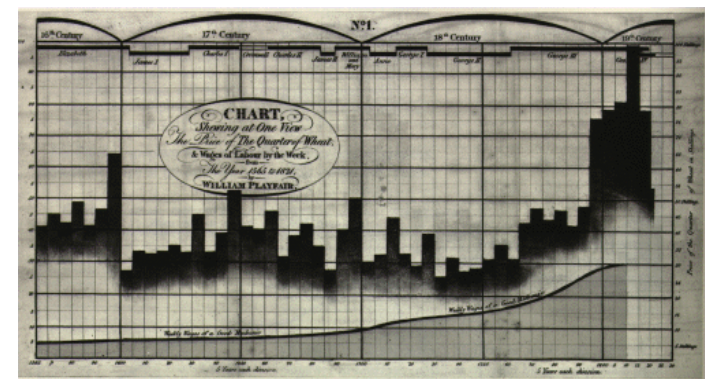• was probably looking for ellipses

Playfair's graphic sins and blunders

Playfair designed this graph to show that workers were best off in recent years, comparing prices to wages. Is this what you see?
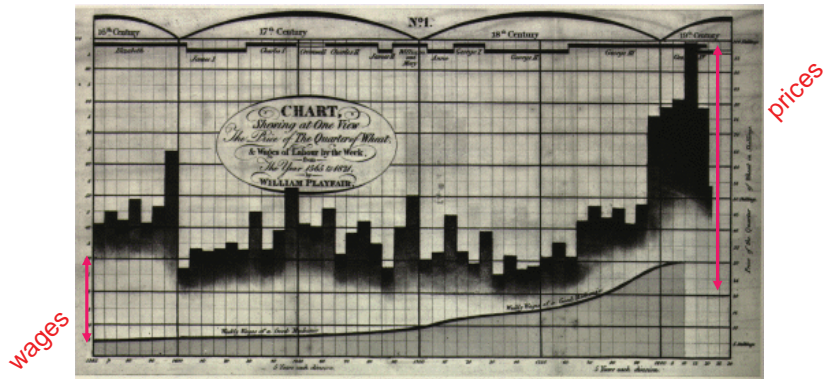

prices

wages

32

Slide 33

Graphic sins:
• The "double Y scale" trick
• Discrete bars for prices, smooth curve for wages

What do you *really* see?
• Volatility of prices
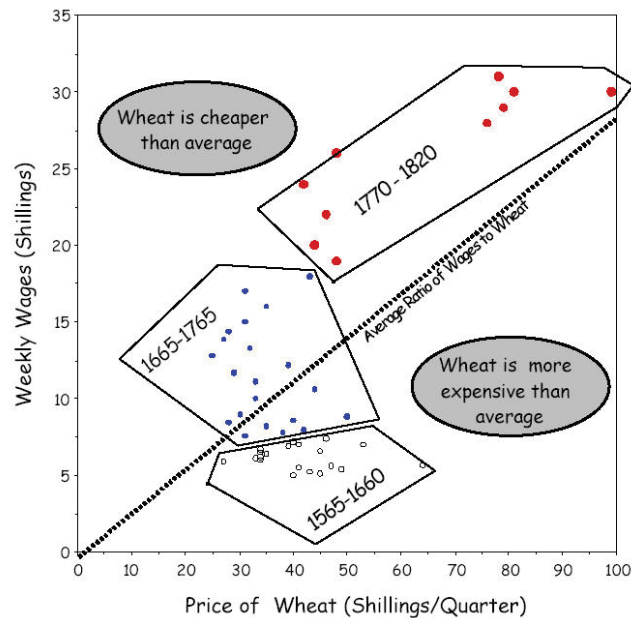• Small range for wages
• How to compare?

prices

wages

33

Slide 34

Playfair re-designed

CHART
Shewing in One View the Work
Required to Purchase
One Quarter of Wheat

Labour cost of wheat (Weeks/Quarter)

Year

34

Slide 35
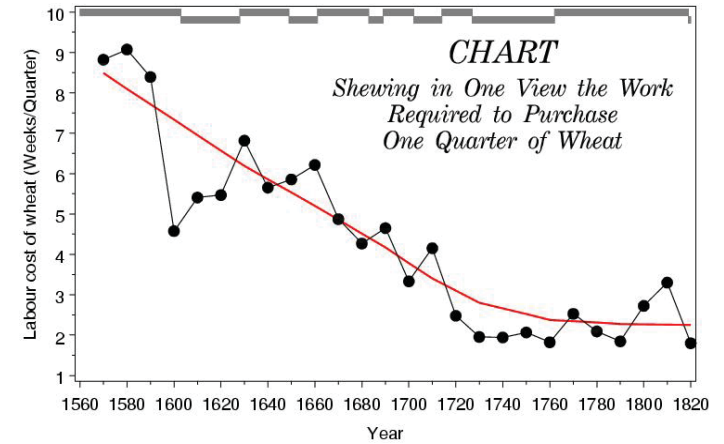
… Wainerized

Making Playfair's message explicit

Wheat is cheaper than average

1770 - 1820

1665-1765

Average Ratio of Wages to Wheat

Wheat is more expensive than average

1565-1660

Weekly Wages (Shillings)
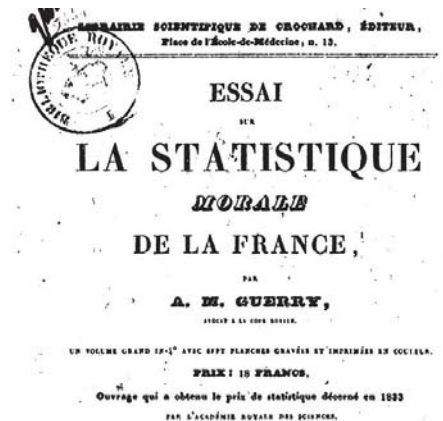
Price of Wheat (Shillings/Quarter)

35

Slide 36

The day social science was invented

André-Michel Guerry (1802-1866)

• *Essai* presented to Academie des Sciences Français: July 2, 1832

• First analysis of comprehensive data on crime, suicide, other 'moral' variables.

•Along with Quetelet, established the study of "moral statistics" -> criminology, sociology, "social science".

ESSAI
SUR
LA STATISTIQUE
MORALE
DE LA FRANCE,
PAR
A. M. GUERRY,

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

36

The discovery of "social facts"
Stability and Variation

Guerry's results were both compelling and startling:

- Rates of crime and suicide remained remarkably invariant over time, yet varied sytematically by region, sex of accused, type of crime, etc.
- In any given French city or department, almost the same number committed suicide, stole, gave birth out of wedlock, etc.

Year	1826	1827	1828	1829	1830	Avg
Sex	All accused (%)					
Male	79	79	78	77	78	78
Female	21	21	22	23	22	22
Age	Accused of Theft (%)					
16–25	37	35	38	37	37	37
25–25	31	32	30	31	32	31
Crime	Committed in summer (%)					
Indecent assault	.	36	36	35	38	36
Assault & battery	.	28	27	27	27	28

"We are forced to conclude that the facts of the moral order are subject, like those of the physical order to invariable laws." (Guerry, 1833, p14)


1829: *Statistique comparée de l'état de l'instruction...*

- Done with Adriano Balbi
- Single-sheet set of three shaded maps (darker = worse)
- Crime against persons, property (pop per crime)
- Instruction (# male school children)
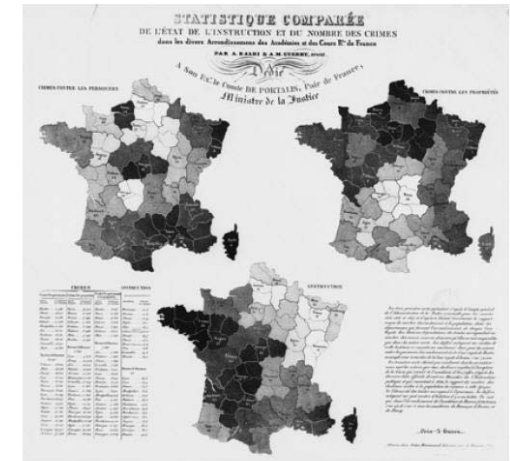
1829: *Statistique comparée de l'état de l'instruction...*

- First shaded thematic maps of crime data
- First comparative maps of social data
- ↦ crime against persons seemed inversely related to crime against property!
- Instruction: ↦ *France obscure* and *France éclairée* (Dupin, 1826)
- North of France highest in education, but also in property crime!
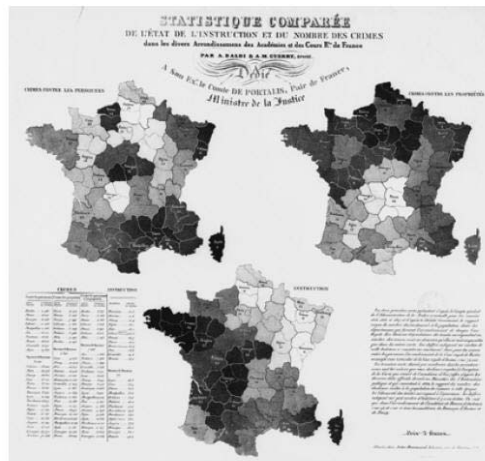
## 1864: *Statistique morale de l'Angleterre comparée...*
Dayenu!

- Proposes to replace simple "moral statistics" (tables) with "analytical statistics"
  - calculation, graphic display
  - ↦ general, abstract results
- 17 large color plates (56 × 39 cm):
  - data for France (1825–1855), England (1834–1855)
  - crimes against persons and property decomposed in various ways
  - first attempt to delineate multivariate relations among moral variables
- Voluminous data:
  - 85,564 suicide records (1836–1860), classified by motive
  - 226,224 accused of personal crime
  - numbers, in a line → 1170 meters!

1864: *Statistique morale de l'Angleterre comparée...*
### Comparing France and England
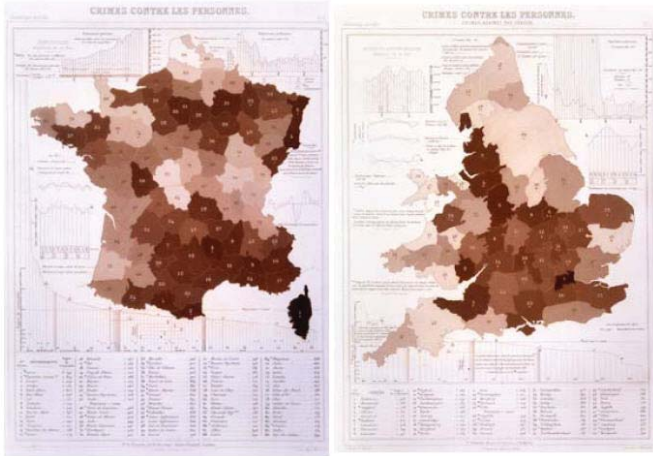
41

Statistique Analytique: General Causes of Crime
Plate XVII: M. Guerry's Magnum Opus

- ► Analysis of the factors associated with crimes and their geographic distribution
- ► Rows: 23 crimes, ordered by frequency and seriousness
  - ► keeping baudy house, bigamy, cattle stealing, ...
  - ► ... fraud, rape, murder
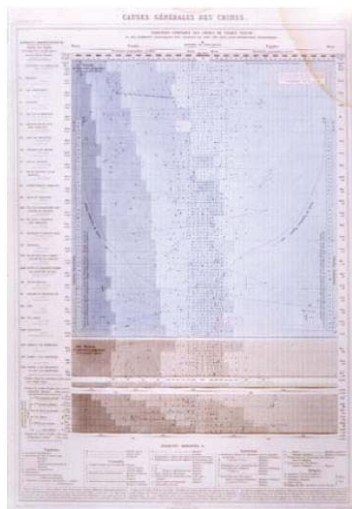- ► Cols: Rank order of degree of criminality of English counties
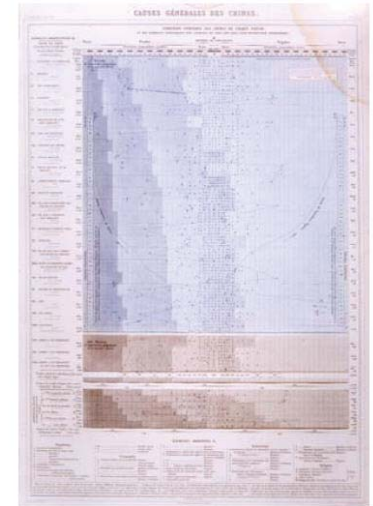
42

Statistique Analytique: General Causes of Crime
Plate XVII: M. Guerry's Magnum Opus

- ► Entries: Symbols for associated moral aspects
  - ► Population (% Irish, agricultural, domestics, ...)
  - ► Criminality (Male, young, ...)
  - ► Religion (Anglicans, "dissenters", ...)
- ► Curves: positive and negative coincidences

43

Why Minard?

- • *"The best statistical graphic ever produced… defies the pen of the historian"*

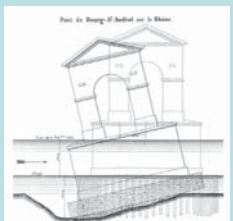| BC | AD | | 17th C | 18th C | 19th Century | 20th Century |

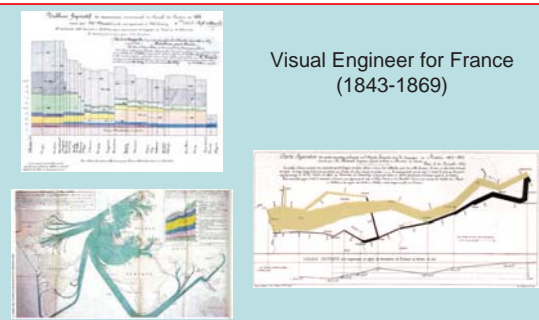1000   1600   1700   1800   1900   2000
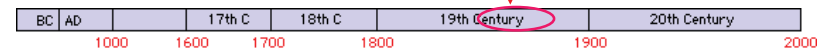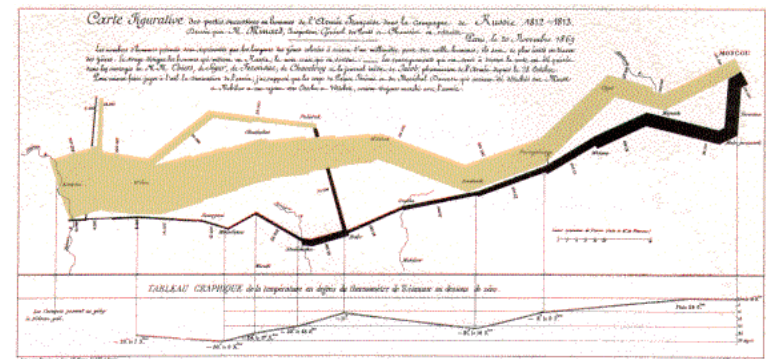
44

Why Minard?

- Study breadth and depth of his work
  - How related to work in his time?
  - How related to modern statistical graphics?
  - How related to his personal history?
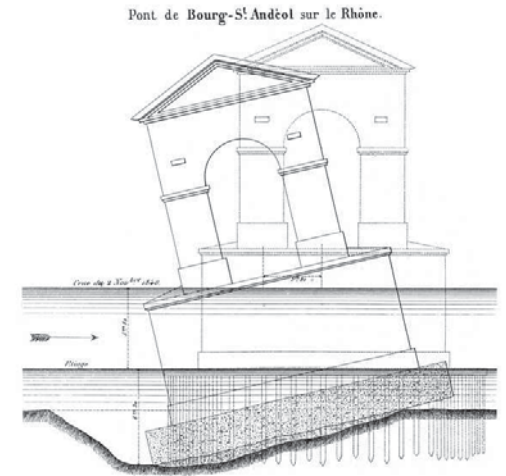
Civil Engineer for ENPC (1810-1842)
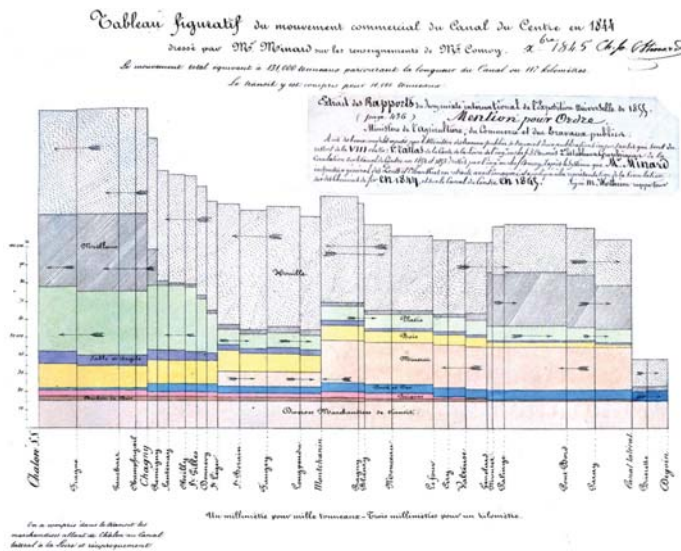
Visual Engineer for France (1843-1869)
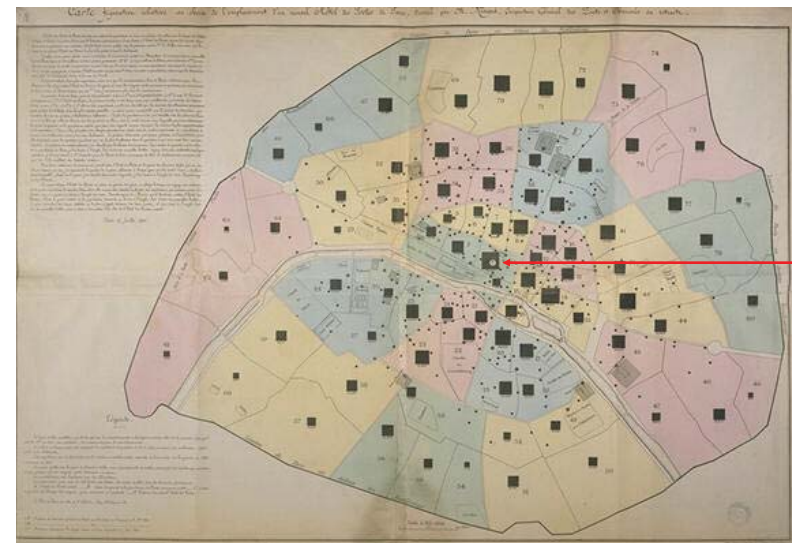
45

Visual thinking

Why did the bridge at Bourg-St. Andèol collapse?

46

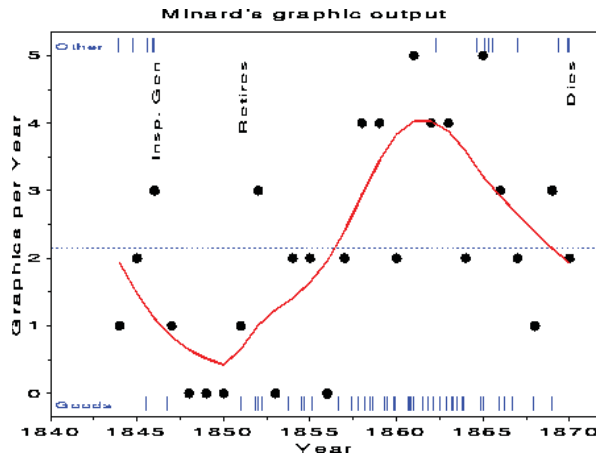**1844**: *Tableau-graphique*: variable-width, divided bars, area ~ cost of transport- Charles Joseph Minard

47

Where to build a new post office?(1867)

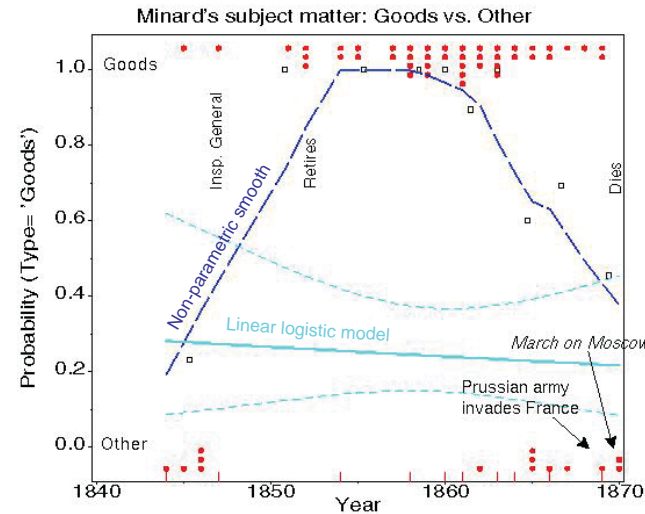Center of gravity of pop. density

48

The Ebb and Flow of Minard's Graphic Output

Graphical insight: smoothing helps the mind's eye

Life skills insight: retirement may not be a bad thing.
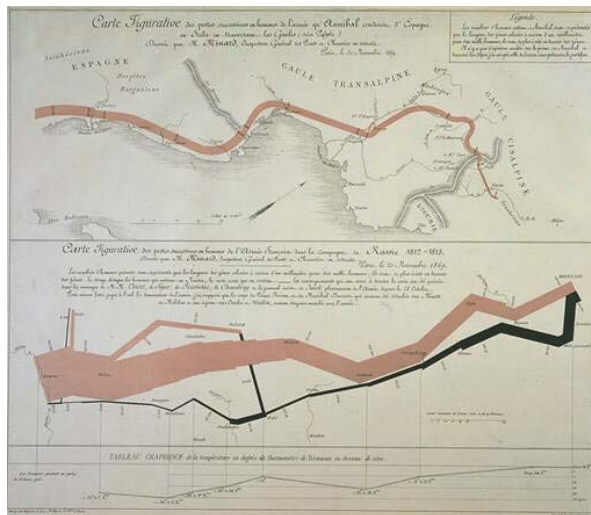
49

Minard's themes: Goods vs. Other

Graphical insight: discrete data are hard to show effectively

Statistical insight: models are often crude approximations

50

The March Re-Visited (1869)

Hannibal's retreat

Napoleon's 1812 campaign

51

Graphic discoveries

- When have graphics led to scientific discoveries that might not have been achieved otherwise?
  - Galton (1883): anti-cyclonic weather pattern
  - Moseley (1913): concept of atomic number
  - Hertzsprung/Russell (1911): temperature classes of stars (spectral type)
 - Phillips (1958): "Phillips curve" of inverse relation between inflation and unemployment

52

Galton's discovery of anti-cyclonic weather patterns

Data: recordings of barometric pressure, wind, rain, temp. 3x/day, 50 weather stations in Europe.

Graphic: 3x3x31 grid, mapping {pressure, wind/rain, temperature} x {AM, 12, PM} x day {1:31}

(try this with your software!)

Galton did for weathermen what Kepler did for Tycho Brahe (Wainer 2005, p56)
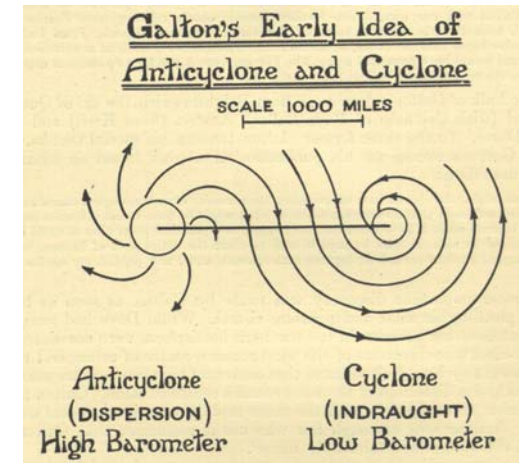

53
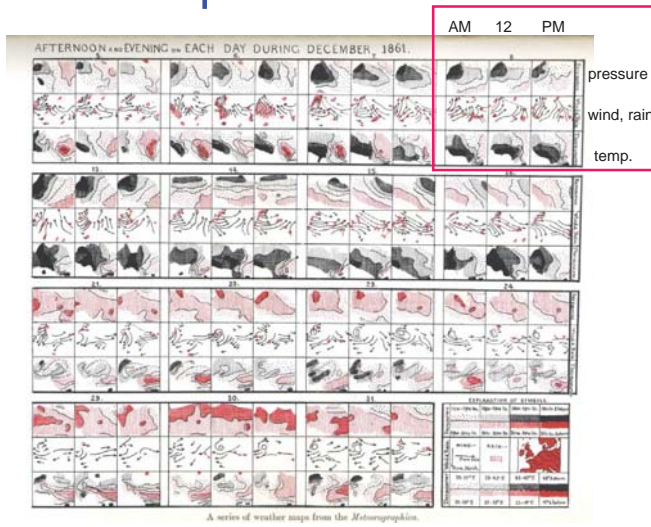
Visual comparison → Insight

The visual insight from 93 (3x31) high-D graphs:
• Changes in wind dir w/ pressure over time
• → Winds revolve CCW in low pressure areas; CW in high pressure

54

Moseley's discovery of atomic number

Plot of serial numbers of elements vs. square root of frequencies from X-ray spectra:

• linear relations → periodic table better explained by atomic number than weight (serial # must have a physical basis).

• gaps in series → predicted existence of several undiscovered elements!

• multiple lines later explained with discovery of the spin of electrons.

The hallmark of good science is the discovery of laws that unify & simplify findings, and allow prediction of yet unobserved phenomena.
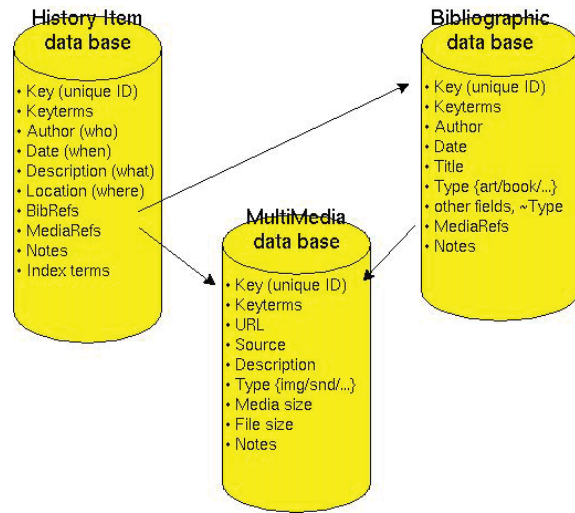

55

Visions of time & history

- How to export advances in data visualization to an historical realm?
  - How might a graphically-minded statistician look at history of data visualization?
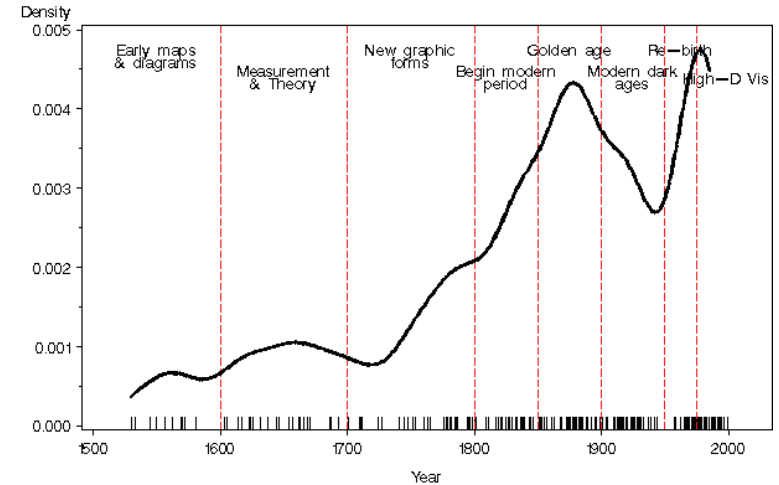- History as data
- How to visualize a history?

56

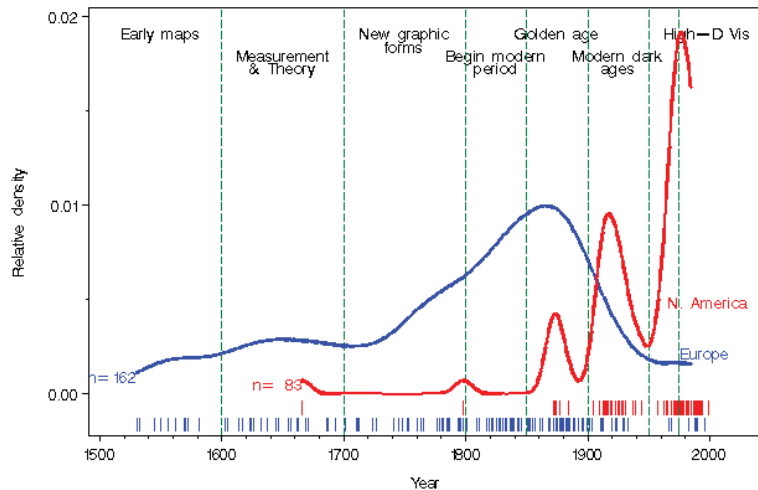## What is milestone "data"?

Milestones data as a relational data base

History Item data base
- Key (unique ID)
- Keyterms
- Author (who)
- Date (when)
- Description (what)
- Location (where)
- BibRefs
- MediaRefs
- Notes
- Index terms

Bibliographic data base
- Key (unique ID)
- Keyterms
- Author
- Date
- Title
- Type {art/book/...}
- other fields, ~Type
- MediaRefs
- Notes

MultiMedia data base
- Key (unique ID)
- Keyterms
- URL
- Source
- Description
- Type {img/snd/...}
- Media size
- File size
- Notes

57

Analyzing milestones data?

58

Analyzing milestones data?

59

How to visualize a history?

- Timeline: obvious, but:
  - 8000+ years, but most in last 300-400
  - Problems of display, resolution, access
  - Linear: no representation of content
- Lessons from the past?
  - Dubourg's Scroll of History
  - Priestly's Charts of Biography, History
  - Marey's life spans of British monarchs

60

Priestly's Chart of Biography

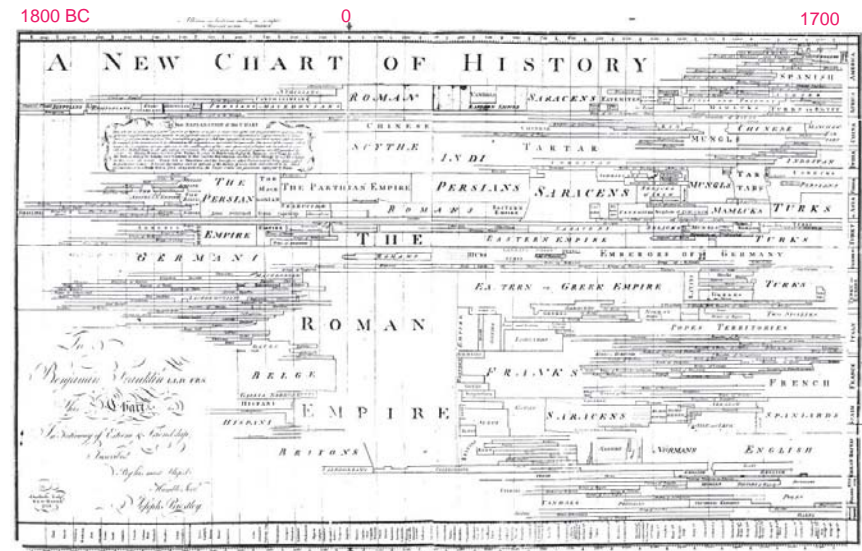- Life spans of famous people, 1200 BC to 1750
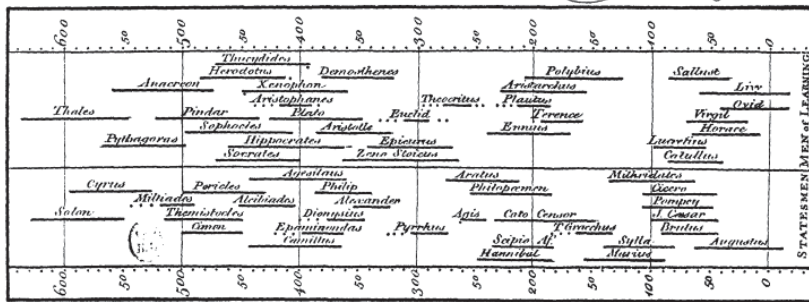
(Priestley, 1765)

61

Priestley's Chart of History

(Priestley, 1770)

62

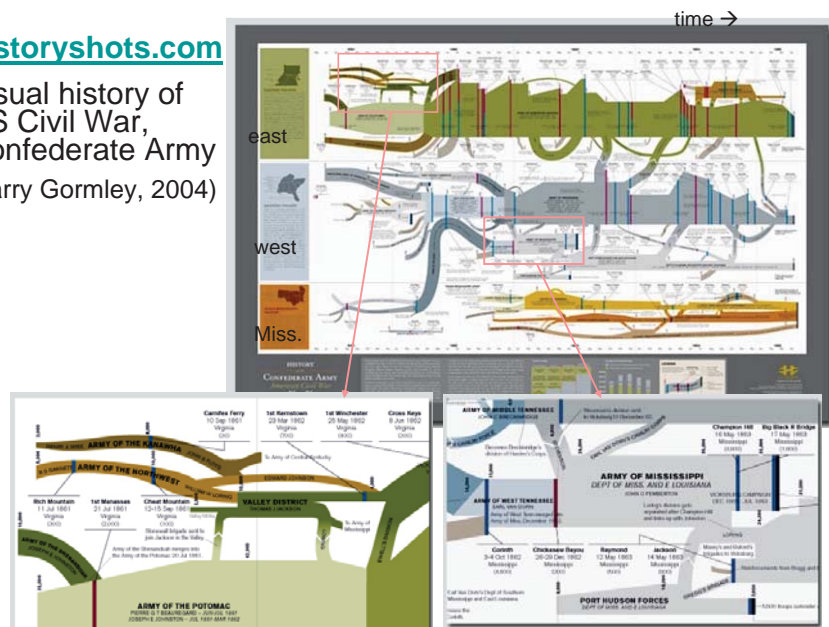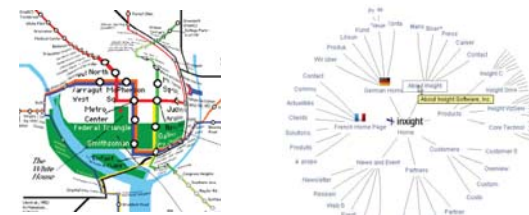**historyshots.com**

Visual history of
US Civil War,
Confederate Army
(Larry Gormley, 2004)
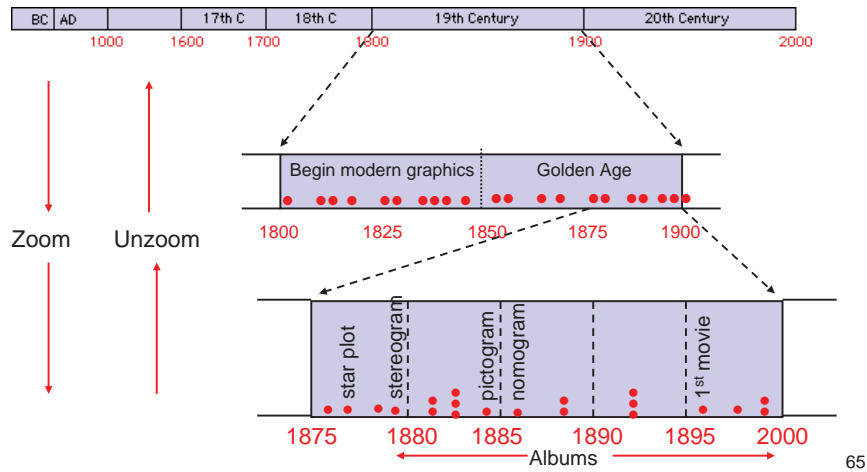

time →

east

west

Miss.

63

Lessons from data visualization

- Zoom, focus & resolution
  - Non-linear scales for space & time
  - Table lens
- Network representations
- Tree representations

4

Zoom, Focus & Resolution

Zoom   Unzoom
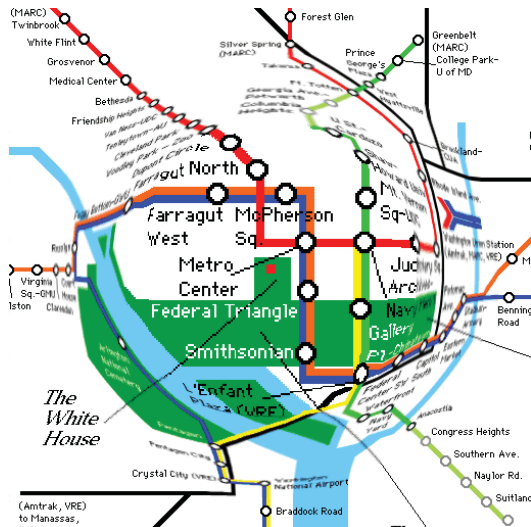
65

Non-linear Scales for Space and Time

Hand with Sphere,
M. E. Esher

66

Non-linear Scales for Space

Fisheye map of central
Washington D.C.

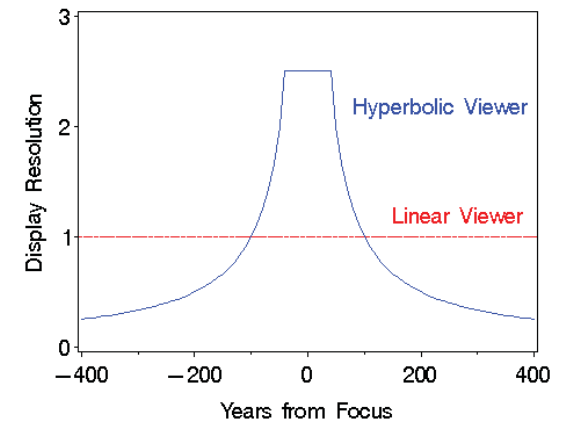Dynamic: move the
cursor to change the
focal point

67

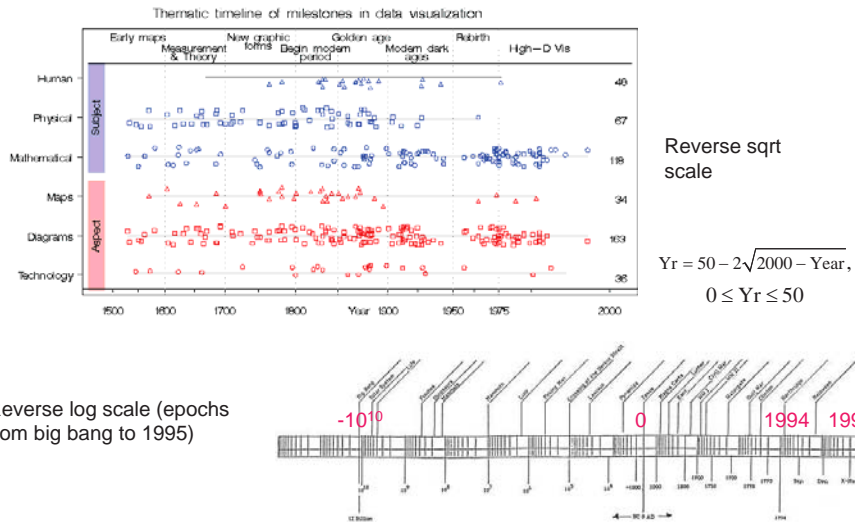# Non-linear Scales for Time

Dynamic views:

A hyperbolic viewer varies
resolution smoothly, trading
off **detail** against **span** of the
view:

*Resolution * Span = constant*

68

Non-linear Scales for Time

Static views:

Thematic timeline of milestones in data visualization

Reverse sqrt scale

$$Yr = 50 - 2\sqrt{2000 - Year},$$
$$0 \le Yr \le 50$$

Reverse log scale (epochs from big bang to 1995)

-10¹⁰ 0 1994 1995

Tree-based views:
Geschictesbaum Europa (2003)

GESCHICHTSBAUM EUROPA

• space-filling design: resolution ~ time²
• natural metaphors for roots, branches
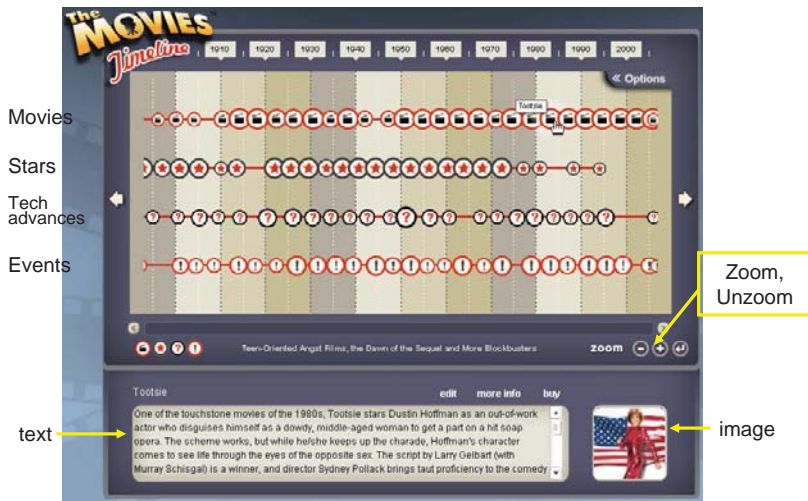

Tree-based views

• Branches for countries & domains of thought

• linear horizontal scale → area ~ time²

71

Lessons from the Web

• Dynamic timelines (flash, java, svg, …)
• Hyper histories
• Apps for your PDA, phone, …
• What do we need to escape flatland?

72

Dynamic timelines

Movies

Stars

Tech advances

Events

Zoom, Unzoom

text

image

Flash app: www.5etdemi.com (Patrick Mineault)
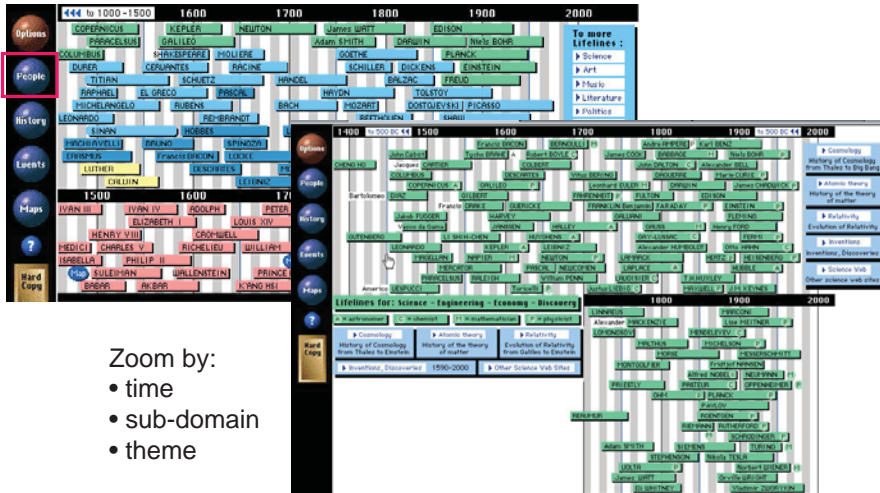
73

HyperHistory Online

http://www.hyperhistory.com/

74

HyperHistory Online

People view (a la Priestley): time x domain

Zoom by:
• time
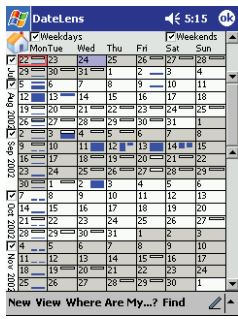• sub-domain
• theme

5

HyperHistory Online

Event view

History by location & time
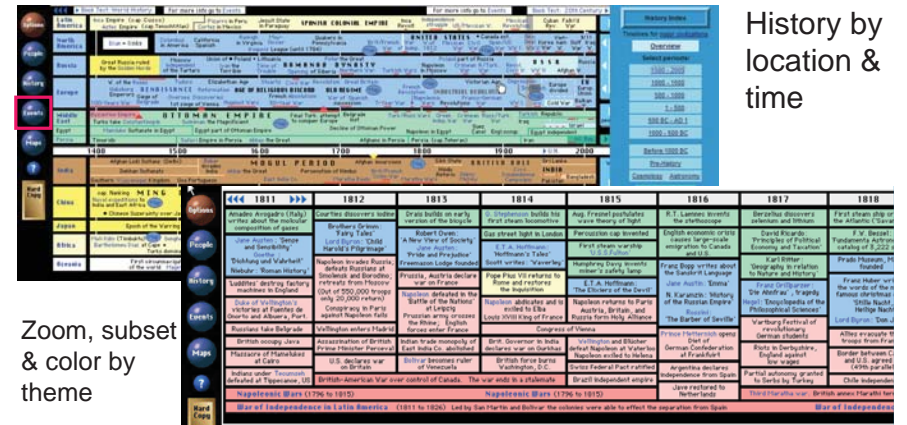
Zoom, subset & color by theme
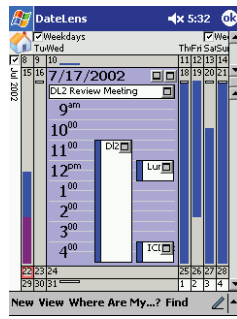
76

DateLens for your PDA

Useful visual interfaces support a variety of selectable **views** and a range of **scales**.

Big view

Graphic view, focused
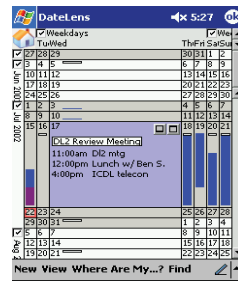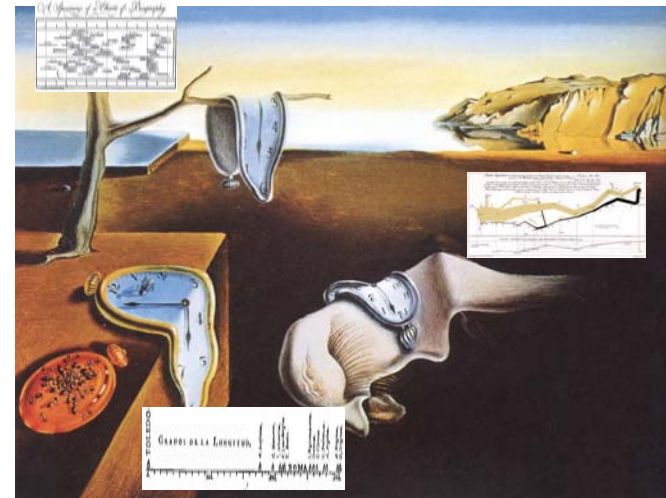
Table view, zoomed

Dali TimeScape

What do we need?
Vis Tools: Availability & Accessibility

www.prefuse.org: Java toolkit for building interactive visualization apps