

Visions of the Past, Present and Future of Statistical Graphics

Swiss Statistics Meeting, 2007

Michael Friendly

York University, Toronto, CA



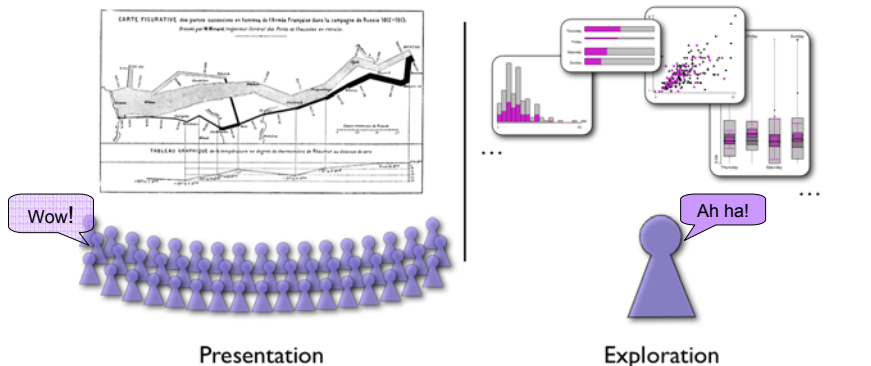
Slides: www.math.yorku.ca/SCS/Papers/swiss/

Outline

- Graphs as communication
- Success stories from the Golden Age
 - A.-M. Guerry & the invention of social science
 - The graphical vision of Ch. Jos. Minard
 - Galton's greatest graphical discovery
- Where are we today?
 - Graphs for data exploration & model fitting
 - Cognitive interfaces for models & graphics
- Where should we go in future?

2

Different graphs for different purposes



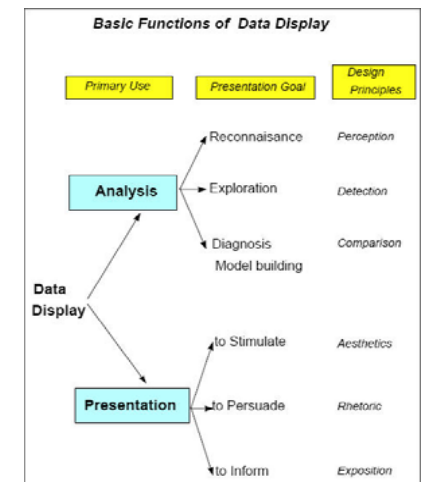
Goal: the Wow! experience
Single image for a large audience

Goal: the Ah ha! Experience
Many images, for a narrow audience (you!), linked to analysis

3

Different graphs for different purposes

- Graphs (& tables) as communication:
 - What is the audience?
 - What is the message?
- **Analysis graphs:** design to see patterns, trends, aid the process of data description, interpretation
- **Presentation graphs:** design to make a point, illustrate a conclusion



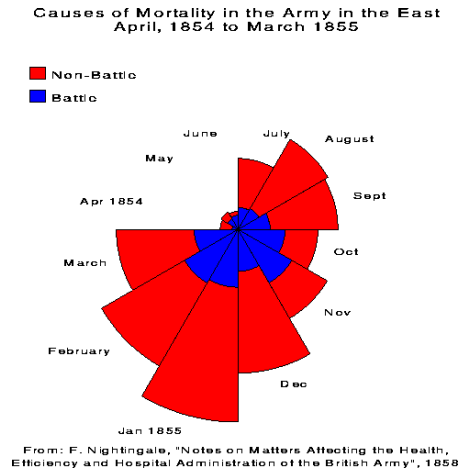
4

Presentation graph: Nightingale's coxcomb

Florence Nightingale: Deaths in the Crimean war from battle vs. other causes (disease, wounds)

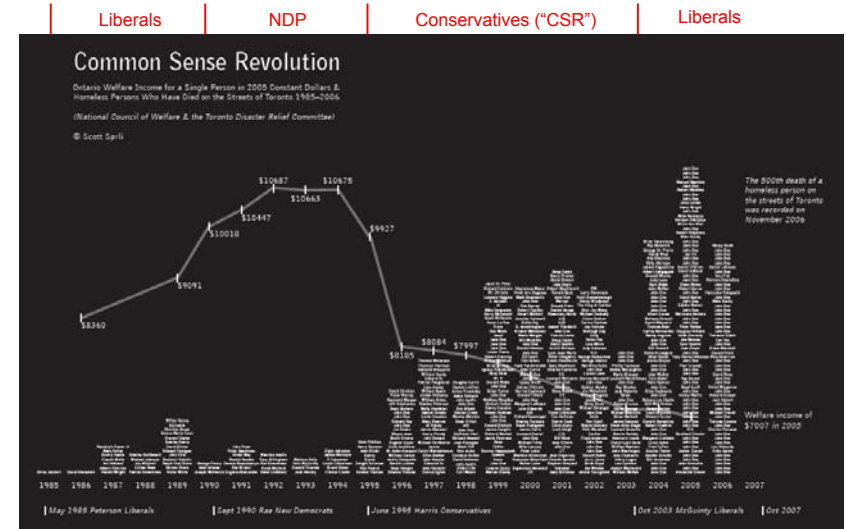
She used this to argue for better field hospitals (MASH units)

The best presentation graphs pass the **Intercocular Traumatic Test**: The message hits you between the eyes!



5

Rhetorical graph: Welfare income and Homeless deaths after the "Common Sense Revolution"



Analysis graph: Deaths vs. Income

Scatterplot of deaths vs. income

- Loess smooth + CI band
- Labels: year
- Color: party in power

The message here is interesting, but it lacks the power and eloquence of the original graph

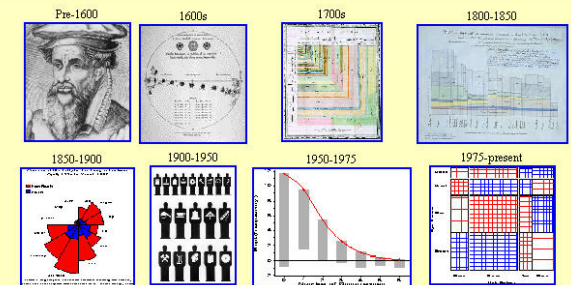


Context: Milestones Project

www.math.yorku.ca/SCS/Gallery/milestone

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

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



Up: Gallery | Introduction | Related | References | Term Index | Category XRef | Search

Pre-1600 | 1600s | 1700s | 1800+ | 1850+ | 1900+ | 1950+ | 1975+

Project goals:

- Comprehensive catalog of developments in history of data visualization
- Tool to study themes, antecedents, influences, patterns, trends, etc.

8

Milestones: Content Overview

Every picture has a story – Rod Stewart



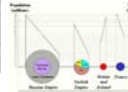
c. 550 BC: The first world map? (Anaximander of Miletus)



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



1701: First contour map- Edmund Halley



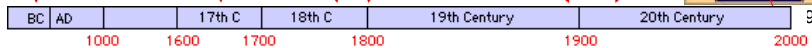
1801: Pie chart, circle graph - William Playfair

1896: Bivariate map- Jacques Bertillon

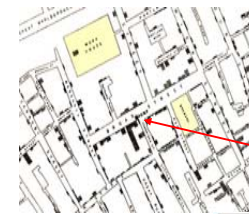


1924: Pictograms- Otto Neurath

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

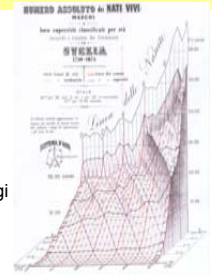


1850-1900: Golden Age



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

Broad St. pump



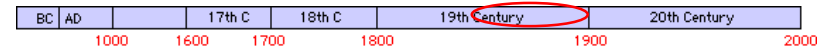
1879: Stereogram (3D population pyramid)- Luigi Perozzo



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



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



Stories from the Golden Age (1850-1900)

Stories:

- Guerry & the invention of social science
- Graphic vision of C. J. Minard
- Galton's graphical discoveries
- Statistical albums

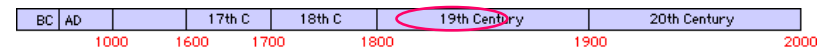
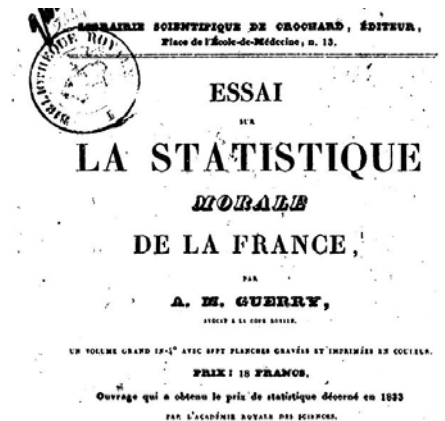
Themes:

- Statistics: numbers of the state
- Rise of visual thinking
- Visualization → Theory (graphic discovery)
- Theory → Practice
- Graphical excellence

A.-M. Guerry and the invention of social science

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".



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 systematically** 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)

Social context of crime in 1820s

- What to do about crime?
 - Crime a serious concern: Explosive growth in Paris, widespread unemployment, emergence of “dangerous classes”
 - **Liberal** (“philanthrope”) view: increase education, better prison conditions, religious instruction, better diet (bread and soup)
 - **Conservative** view: build more prisons, harsher treatment of recidivists
- But: there was very little data!
 - That would soon change

14

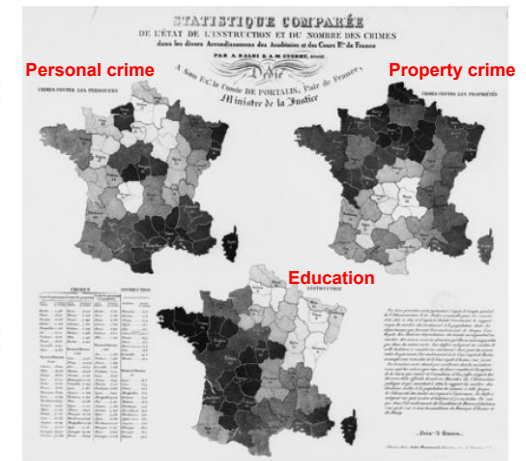
An avalanche of social numbers

- J.B.J. Fourier: *Recherches statistique sur la ville de Paris* (1821-9)
 - Massive tabulations: births, deaths (by cause), admission to insane asylums (age, sex, affliction)
- Ministry of Justice: *Compte generale* (1825--)
 - First **national** compilation of criminal justice data
 - **All** charges & dispositions, quarterly, 86 departments
- Other sources:
 - Bureau de Longitudes (illegitimate births)
 - Parent-Duchatelet (prostitution); Min. of War (desertions)
- Social issues could now be addressed with **DATA**

15

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!



16

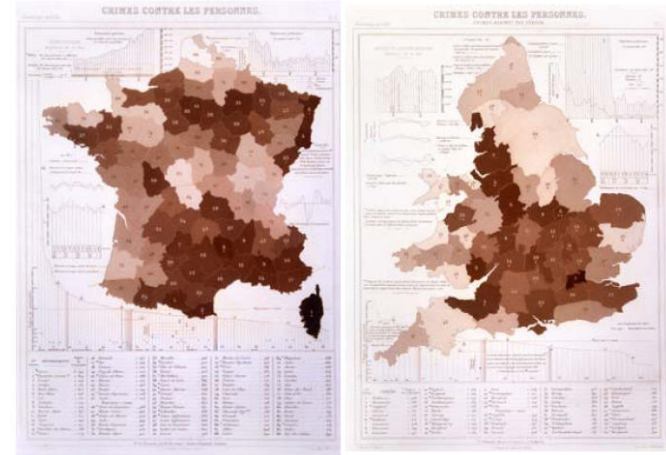
1864: *Statistique morale de l'Angleterre comparée...* Dayenu!

- ▶ 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!



17

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



18

Statistique analytique: General causes of crime

Plate XVII: M. Guerry's magnum opus

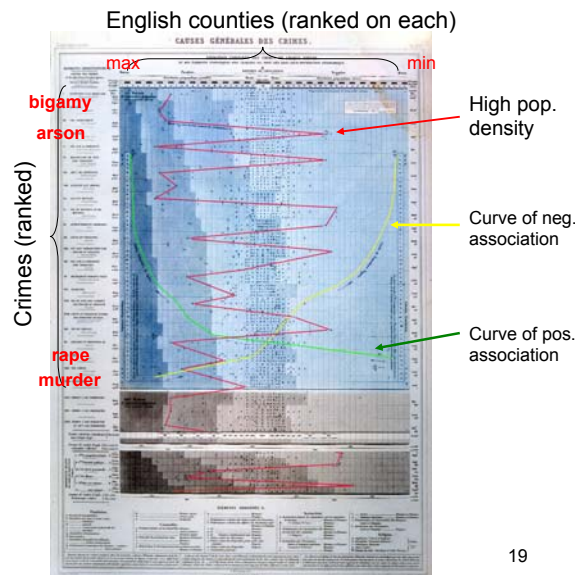
Goal:

- Show multivariate factors associated with distribution of crime
- Before invention of correlation

Entries: Codes for factors

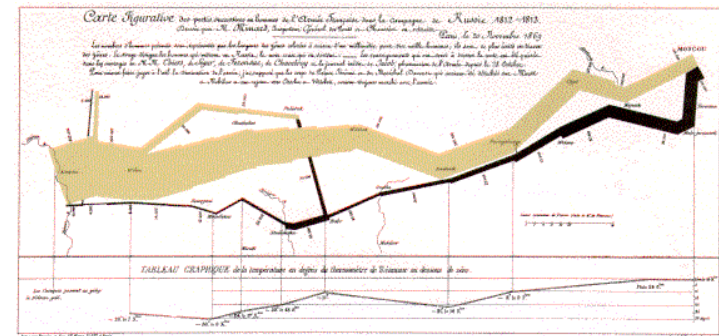
- Pop: (% Irish, domestics, ...)
- Criminality: (male, young, ...)
- Religion (Anglicans, dissenters, ...)

	g	h	v	z	λ	δ
x	f	e	β	a	n	c
	α	e	j	k	q	p
	r	γ	d	l	η	ε
					h	ν
					o	m



19

The graphic vision of C. J. Minard



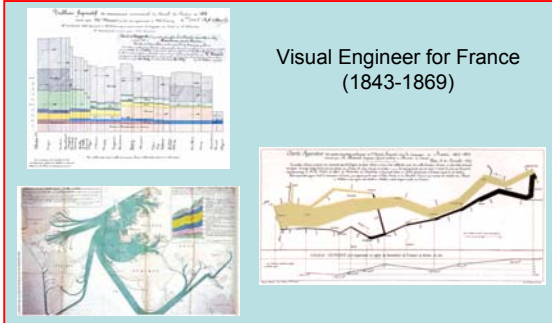
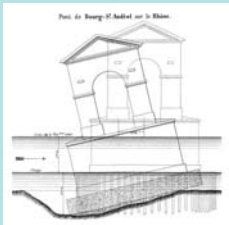
- Marey (1878): “defies the pen of the historian in its brutal eloquence”
- Tufte (1983): “the best statistical graphic ever produced”

20

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)



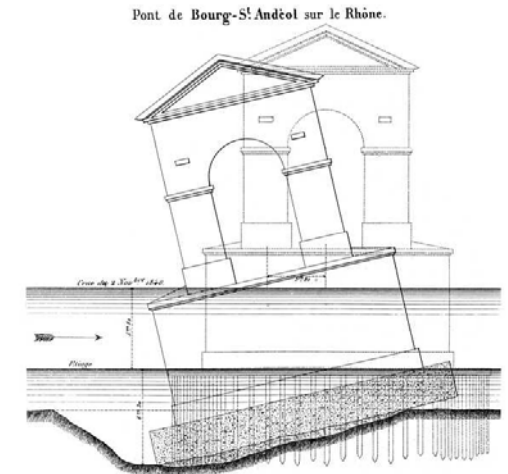
Visual Engineer for France
(1843-1869)

21

Visual thinking, visual explanation

1840: Why did the bridge at Bourg-St. Andéol collapse?

Minard's report consisted essentially of this self-explaining diagram.



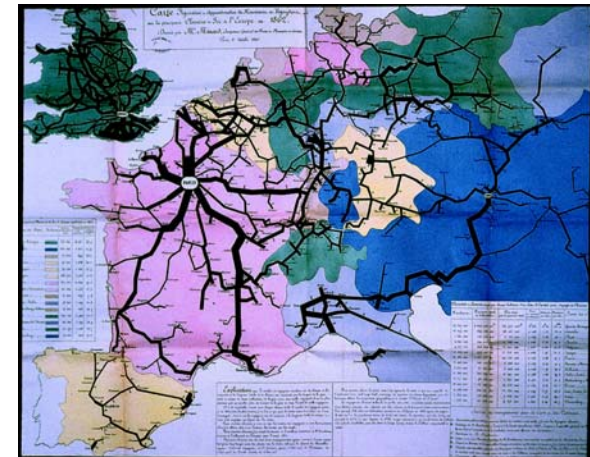
Visual tools for state planning

- 1830—1860: emergence of modern French state, dawn of globalization
- Trade, commerce, transportation:
 - Where to build railroads, canals?
 - Visualizing changes over time, differences over space
 - → Flow maps and other graphical innovations

23

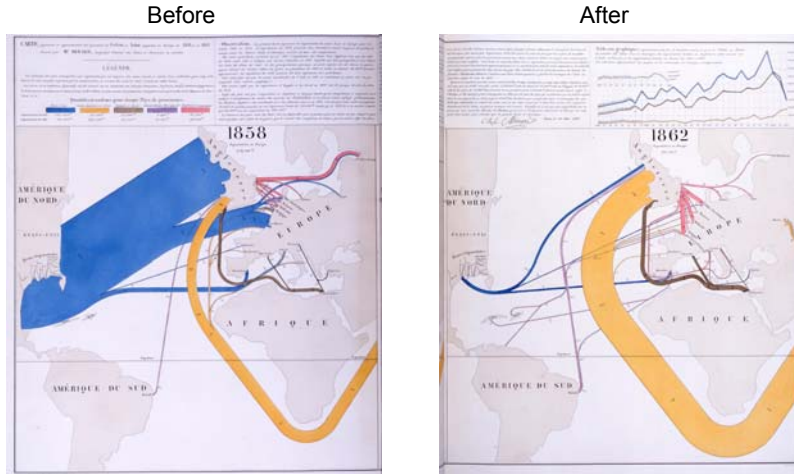
Flow maps as visual tools

Transport of passengers on the principal railroads in Europe in 1862



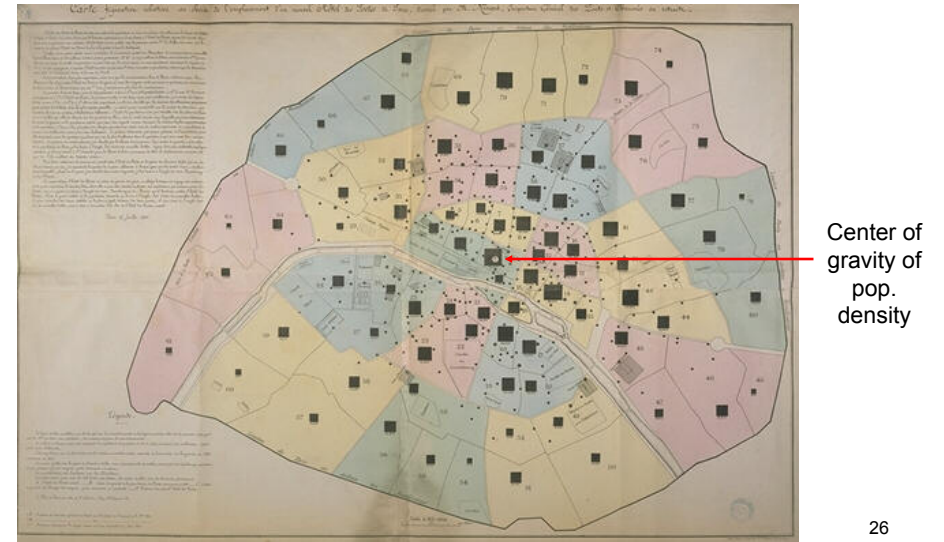
24

Effect of US civil war on cotton trade



25

Where to build a new post office?(1867)

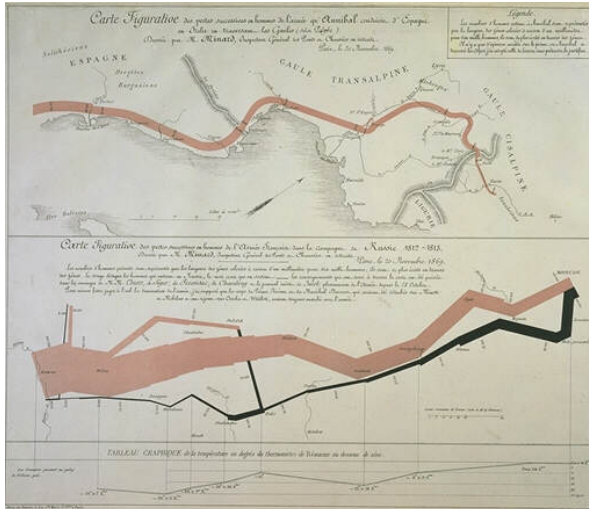


26

The March Re-Visited (1869)

Hannibal's retreat

Napoleon's 1812 campaign



27

Galton's visual discoveries- Bivariate normal correlation surface (1886)

Table 9.1 One of Galton's correlation tables

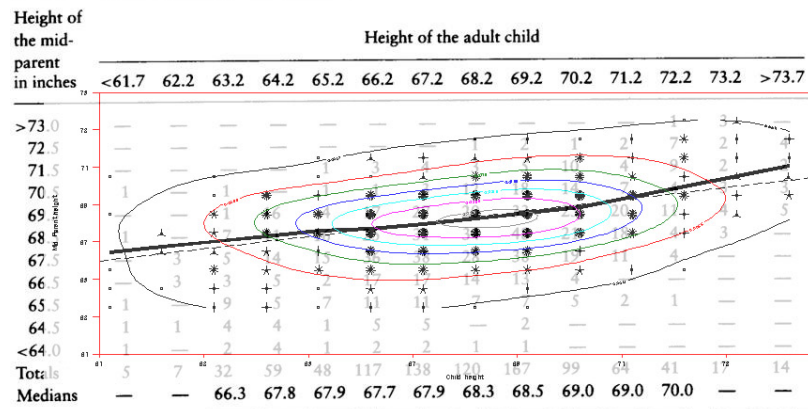
Height of the mid-parent in inches	Height of the adult child													
	<61.7	62.2	63.2	64.2	65.2	66.2	67.2	68.2	69.2	70.2	71.2	72.2	73.2	>73.7
>73.0	—	—	—	—	—	—	—	—	—	—	—	1	3	—
72.5	—	—	—	—	—	—	—	1	2	1	2	7	2	4
71.5	—	—	—	—	1	3	4	3	5	10	4	9	2	2
70.5	1	—	1	—	1	1	3	12	18	14	7	4	3	3
69.5	—	—	1	16	4	17	27	20	33	25	20	11	4	5
68.5	1	—	7	11	16	25	31	34	48	21	18	4	3	—
67.5	—	3	5	14	15	36	38	28	38	19	11	4	—	—
66.5	—	3	3	5	2	17	17	14	13	4	—	—	—	—
65.5	1	—	9	5	7	11	11	7	7	5	2	1	—	—
64.5	1	1	4	4	1	5	5	—	2	—	—	—	—	—
<64.0	1	—	2	4	1	2	2	1	1	—	—	—	—	—
Totals	5	7	32	59	48	117	138	120	167	99	64	41	17	14
Medians	—	—	66.3	67.8	67.9	67.7	67.9	68.3	68.5	69.0	69.0	70.0	—	—

Source: Galton (1886), p. 68.

28

Visual smoothing → Insight

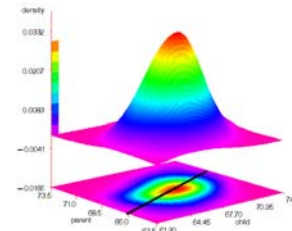
Table 9.1 One of Galton's correlation tables



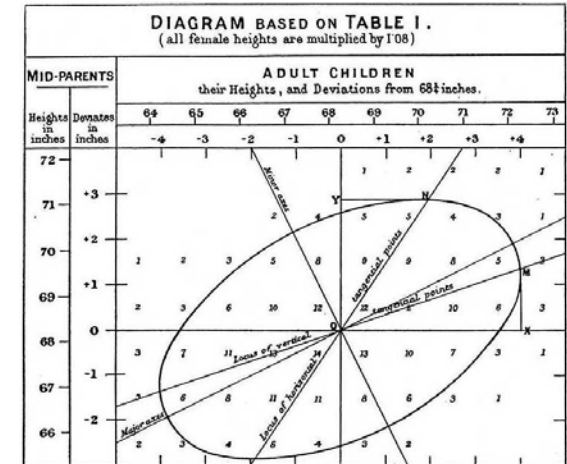
Source: Galton (1886), p. 68.

Visual insight → Theory

- 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 noteworthy scientific discoveries arising from analysis of pure observation (Pearson 1920, p37)



Galton (1886, Pl X): Smoothed contours of heights of parents and children 30

Galton's discovery of weather patterns- Perhaps the most notable *purely graphic* discovery ever!

METEOROGRAPHICA,

METHODS OF MAPPING THE WEATHER;

ILLUSTRATED BY UPWARDS OF 600 PRINTED AND LITHOGRAPHED DIAGRAMS

REFERRING TO

THE WEATHER OF A LARGE PART OF EUROPE,

During the Month of December 1861.

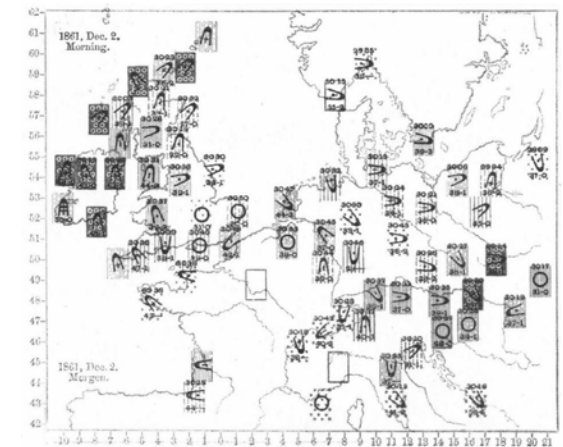
By FRANCIS GALTON, F.R.S.

(Galton, 1863)

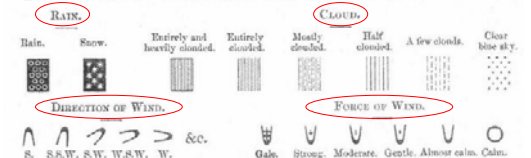
Method: All weather stations across Europe asked to record data 3x/day for all of Dec., 1861

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

Graphic analysis: 3x31=93 maps, each with multivariate glyphs showing all variables



EXPLANATION OF THE SYMBOLS USED IN THE WEATHER CHARTS.



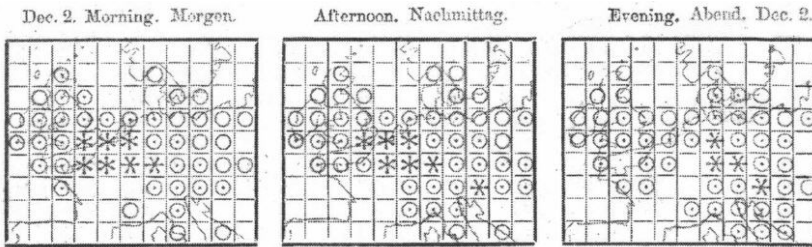
Visual ideas:

- Iconic symbols
- Multivariate glyphs (stamps!)

Visual abstraction → Patterns

How to see patterns of geographical variation over time?

- Iconic symbols on a geographical grid
- “Small multiples:” separate graphs laid out for direct comparison



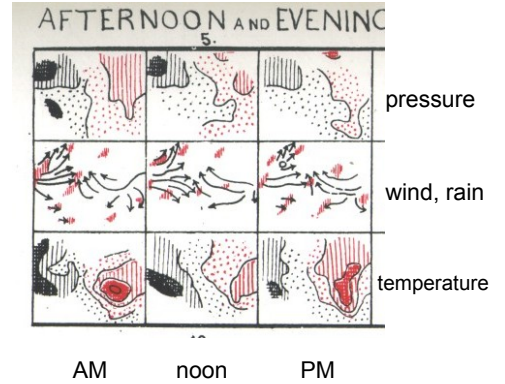
Symbols in Barometrical Charts.

Black	Inches. 29.95 to 29.71	Inches. 29.70 to 29.46	Inches. 29.45 to 29.21	Inches. 29.20 and below.
Red	○ 29.96 to 30.20	⊙ 30.21 to 30.45	✱ 30.46 to 30.70	● 30.71 and above.

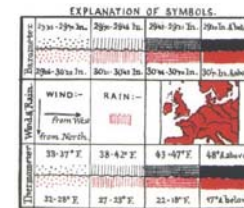
Visual abstraction → Patterns

What varies with what, over time and space?

- mini, abstract maps: vars x TOD
- iso-contours, shading to show equivalence
- arrows to show wind direction



Data for Dec 5, 1861

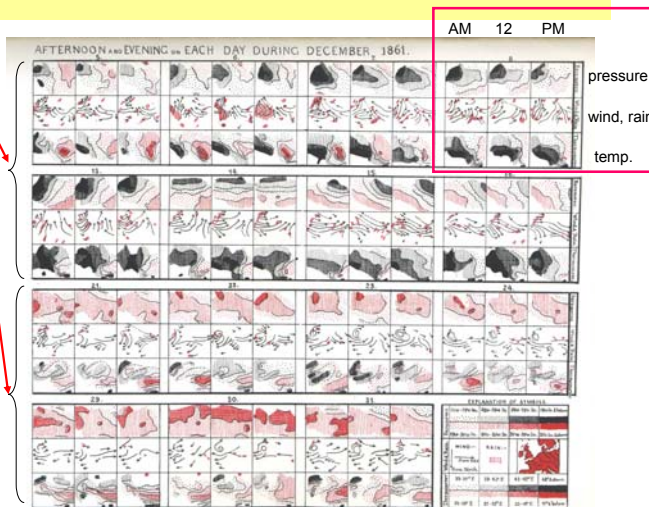


The large picture → Insight

Pattern:
 Low pressure (black) in early Dec. → CCW wind
 High pressure (red) in late Dec. → CW wind

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

(try this with your software!)



A series of weather maps from the *Metéorographien*.

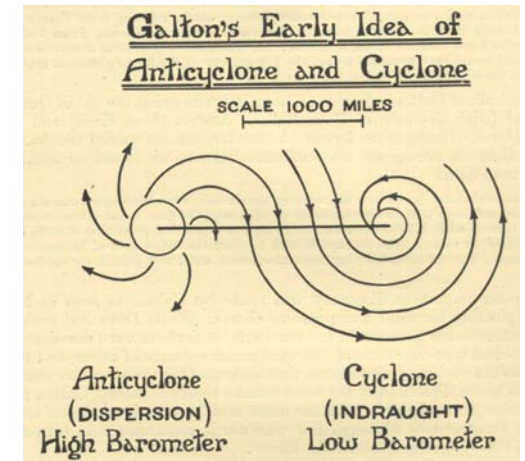
Visual insight → Theory

Visual insight from 93 (3x31) high-D graphs:

- Changes in wind dir w/ pressure over time
- → Winds revolve inwardly (CCW) in low pressure areas— as in a cyclone;
- → revolve outwardly (CW) in high pressure areas— “anti-cyclone”

Theory:

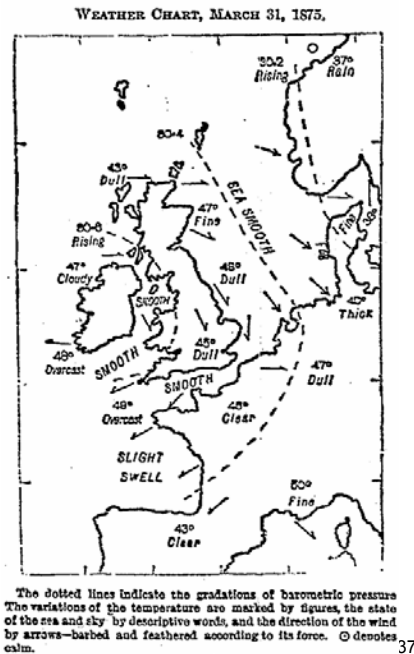
- Explained by Dove’s ‘Law of Gyration’
- Prediction: reversed pattern (CW/CCW) in southern hemisphere – confirmed!



Theory → Practice

The first modern weather map, *London Times*, Apr. 1, 1875

Galton did for weathermen what Kepler did for Tycho Brahe. This is no small accomplishment. (Wainer 2005)



37

Statistical albums: Data → practice & Graphical excellence

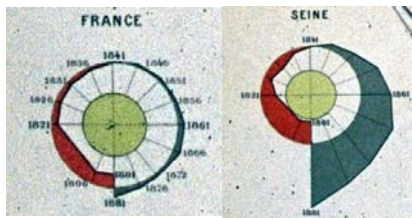
- Collection of gov't statistics on pop., trade, moral & political issues widespread in Europe & US, starting ~ 1820
- Statistical albums ~ 1870—1910
 - France: *Albums de Statistique Graphique*
 - Germany: local albums (Berlin, Frankfurt, etc.)
 - Switzerland: *Atlas graphique de la Suisse*
 - USA: Census atlases: 1870/80/90

38

Album de statistique graphique

- The pinnacle of the Golden Age
- 18 volumes published 1879-1899

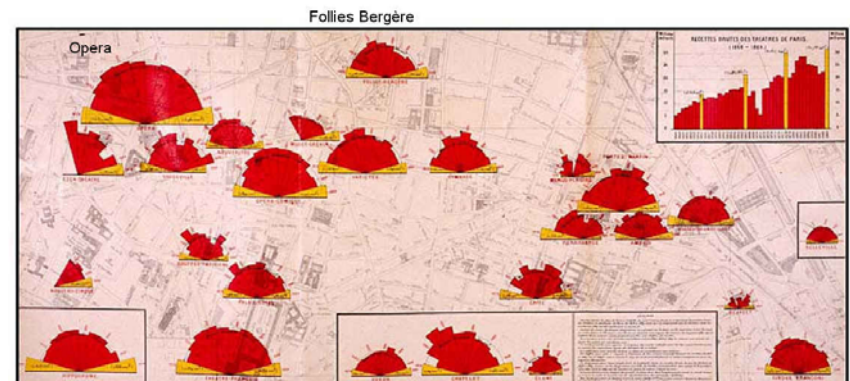
Changes in the population of France from 1801—1881, by department (Album, 1881)



39

Album de statistique graphique

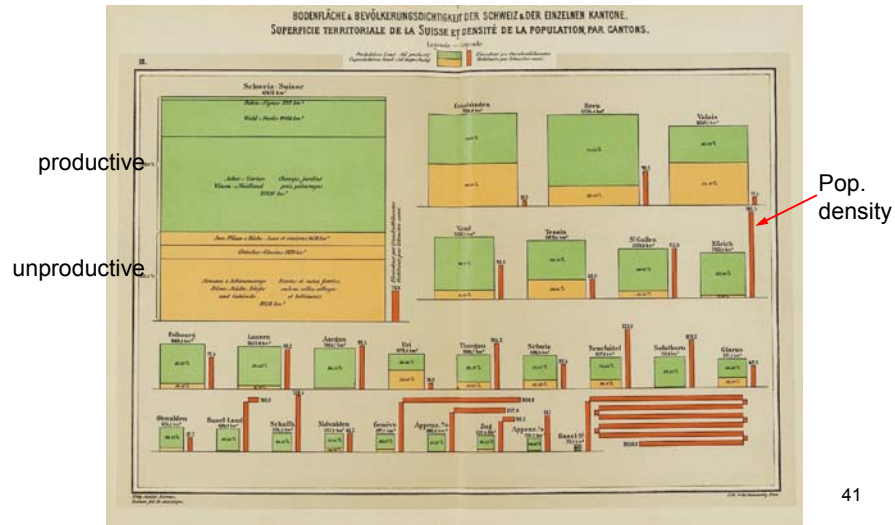
Gross receipts in theaters in Paris, 1848—1889, related to cultural events (Album, 1889)



40

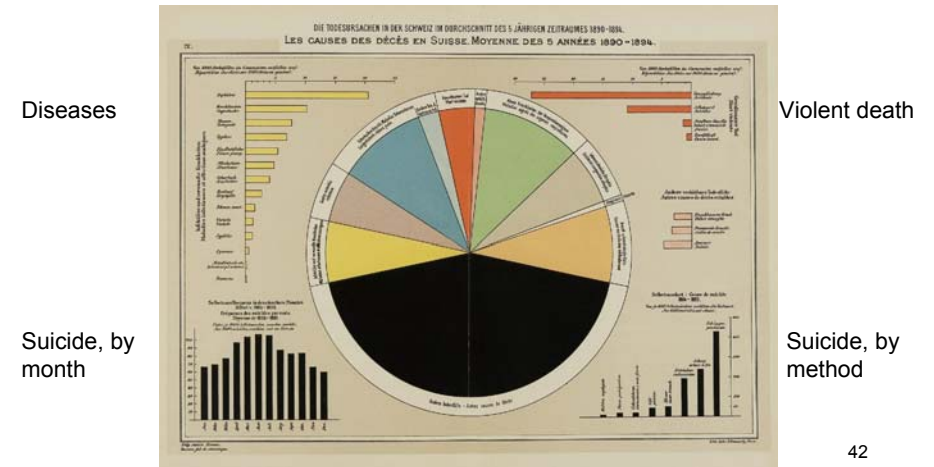
Atlas graphique de la Suisse (1897, 1914)

Land distribution & population density by canton (Atlas, 1897)



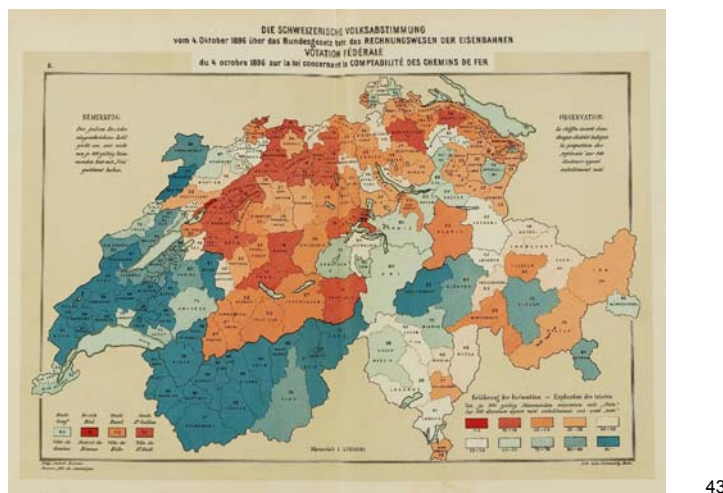
Atlas graphique de la Suisse (1897, 1914)

Causes of death, 1890—1894, color linking (Atlas, 1897)



Atlas graphique de la Suisse (1897, 1914)

Referendum on development of railroads in 1896 – bipolar color scale



Where are we today?

- Analysis graphs for exploration and model diagnosis (linear models)
- Graphical methods for categorical data & frequency tables (generalized linear models)
- Ease of use? Interactive data graphics?
- Cognitive interfaces for models and graphics?

Graphical methods for linear models

# of response variables	Classical linear models	Generalized linear models
1	LM family: $E(y)=X\beta$, $V(y X)=\sigma^2I$ ANOVA, regression, ... Many graphical methods: effect plots, spread-leverage, influence, ...	GLM: $E(y)=g^{-1}(X\beta)$, $V=V[g^{-1}(X\beta)]$ poisson, logistic, loglinear, ... Some graphical methods: mosaic plots, 4fold plots, diagnostic plots, ...
2+	MLM: $E(Y)=X\beta$, $V(Y X)=I\otimes\Sigma$ MANOVA, MMRReg, ... Graphical methods: ???	MGLM: ??? Graphical methods: ???

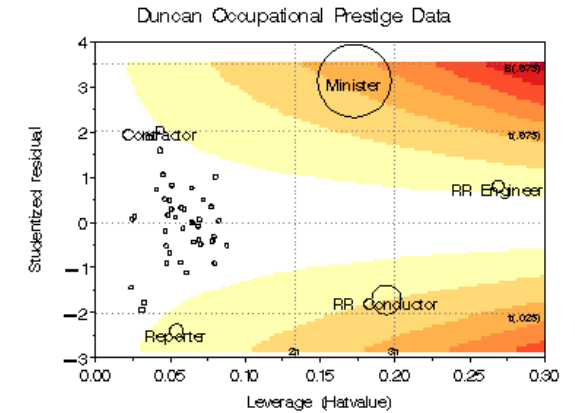
45

Model diagnosis: Influence in regression

Multiple regression model: prestige ~ income + education

Influence plots can show:

- model residual
- leverage (potential impact)
- influence ~ residual x leverage (Cook D statistic)
- contour map of influence



46

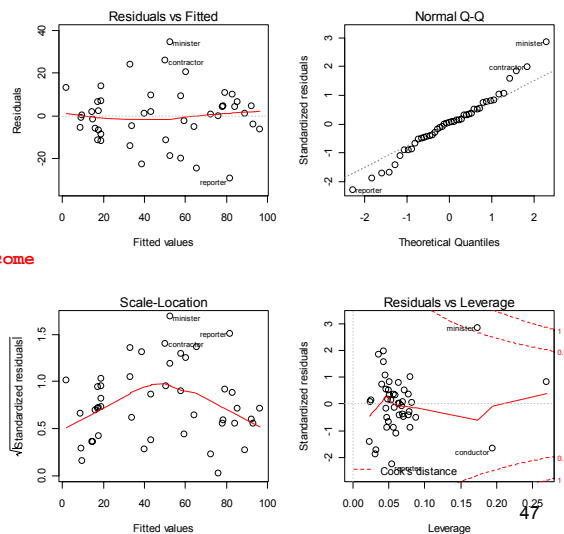
Model diagnosis: regression quartet

Statistical software should make it easy to get informative diagnostic plots

In R, plotting a `lm` model object → the “regression quartet” of plots

```
> model <- lm(prestige ~ income + education)
> plot(model)
```

(SAS has similar, using ODS graphics)



47

Graphical methods for GLMs

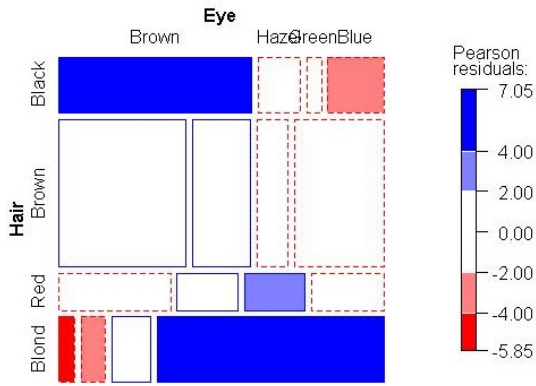
# of response variables	Classical linear models	Generalized linear models
1	LM family: $E(y)=X\beta$, $V(y X)=\sigma^2I$ ANOVA, regression, ... Many graphical methods: effect plots, spread-leverage, influence, ...	GLM: $E(y)=g^{-1}(X\beta)$, $V=V[g^{-1}(X\beta)]$ poisson, logistic, loglinear, ... Some graphical methods: mosaic plots, 4fold plots, diagnostic plots, ...
2+	MLM: $E(Y)=X\beta$, $V(Y X)=I\otimes\Sigma$ MANOVA, MMRReg, ... Graphical methods: ???	MGLM: ??? Graphical methods: ???

48

Visualizing categorical data: mosaic plots

Model: loglin(~ Hair + Eye)

Two-way table: [Hair][Eye]



A contingency table can be visualized by tiles whose area ~ cell frequency.

Shading: ~ Pearson residual,

$$d_{ij} = (O_{ij} - E_{ij}) / \sqrt{E_{ij}}$$

Color:

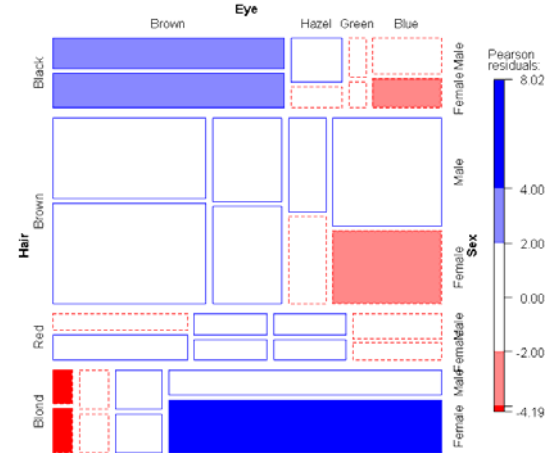
- blue: $O_{ij} > E_{ij}$; red: $O_{ij} < E_{ij}$

Interp: + association (dark hair, dark eyes), (light hair, light eyes)

N-way tables

Model: loglin(~ Hair + Eye + Sex)

Independence model: [Hair][Eye][Sex]



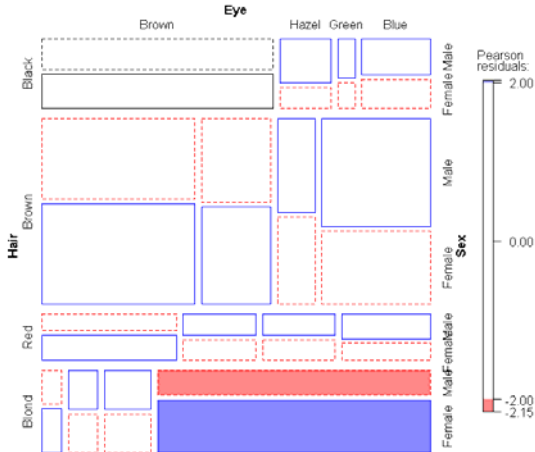
3+ way tables: split each tile ~ conditional proportions of the next variable.

Now, there are several different models that can be fit.

- Mutual independence: [H][E][S] → all vars unassociated
- Residuals: show associations not acct'd for by the model

N-way tables

Joint independence: [Hair, Eye][Sex]



The model of joint independence, [HE][S] allows Hair, Eye color association, but → [HE] assoc. is independent of Sex.

This model obviously fits much better, except for blue-eyed blonds, where females are more prevalent than the model allows.

Mosaic displays: visual assessment of model fit:

Better model → “cleans the mosaic”



Cognitive interfaces for data exploration & model-fitting

- “Dumb” vs. “Smart” data tables: variables assigned default **roles** (factor, covariate, response, label)
- Model syntax, e.g., loglin(~ Hair+Eye+Sex) vs. model-building widgets & GUIs
- Begging for graphs vs. getting them (semi-)automatically
- Model history: easily compare, modify or explore multiple models for the same data
- Dynamic, interactive graphics
 - Multiple, linked views: data table, 2D graphs, 3D plots
 - Selection: subset or highlight all views for a given selection
 - Projection & low-D views of hi-D data

JMP: Model summary = graphs + numbers

Variable roles

Default graphs

53

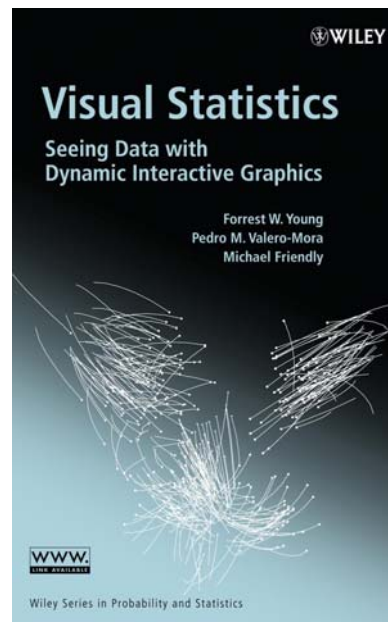
JMP: Model summary = graphs + numbers

54

- Summarizes 15 years of research
- Visual interfaces for statistical analysis & dynamic graphics
- Testbed: ViSta
<http://forrest.psych.unc.edu/research/>



- Spreadplots: graphic analog of spreadsheet
- Work maps: visual GUI for path(s) of analysis
- Guide maps: visually guide novices thru typical analysis steps



ViSta: regression spreadplots

Model: mpg ~ PC(,), data=auto

3D spin biplot

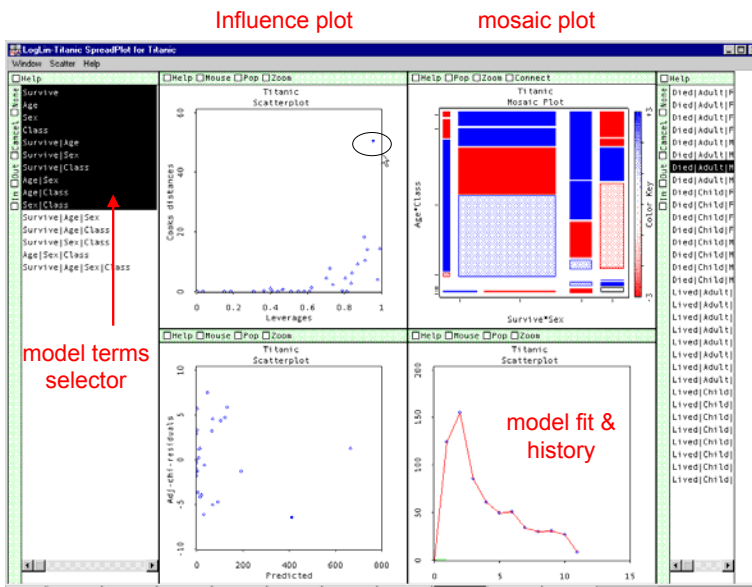
Influence plot

predictor overview

var/obs selector

56

ViSta: loglin spreadplots



The future of statistical graphics?

- New visualizations for statistical methods
 - Closer integration of analysis (summarization) and graphics (exposure)
 - E.g., HE plots for MLMs
- Public data, public graphics
 - Worldmapper – the world in cartograms
 - Facebook for data & graphs?
 - Gapminder – analyzing trends and relations
- Graphical excellence

Graphical methods for MLMs

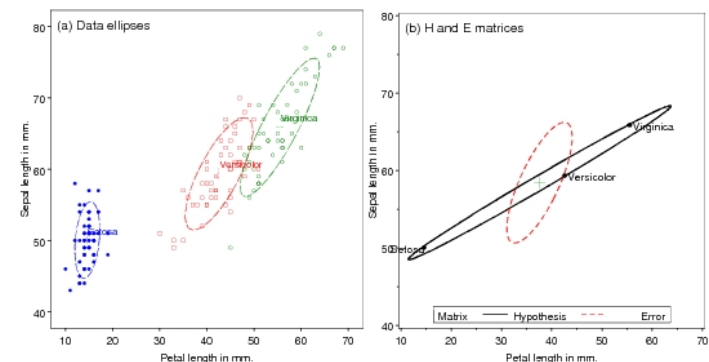
	Classical linear models	Generalized linear models
1	LM family: $E(y) = X\beta$, $V(y X) = \sigma^2 I$ ANOVA, regression, ...	GLM: $E(y) = g^{-1}(X\beta)$, $V = V[g^{-1}(X\beta)]$ poisson, logistic, loglinear, ...
	Many graphical methods: effect plots, spread-leverage, influence, ...	Some graphical methods: mosaic plots, 4fold plots, diagnostic plots, ...
2+	MLM: $E(Y) = X\beta$, $V(Y X) = I\otimes\Sigma$ MANOVA, MMRreg, ...	MGLM: ???
	Graphical methods: ???	Graphical methods: ???

of response variables

HE plots for MANOVA, MMRreg

HE plots provide a way to visualize hypothesis tests in MANOVA and multivariate multiple regression, using data ellipses for fitted (H) and residual (E) co-variances.

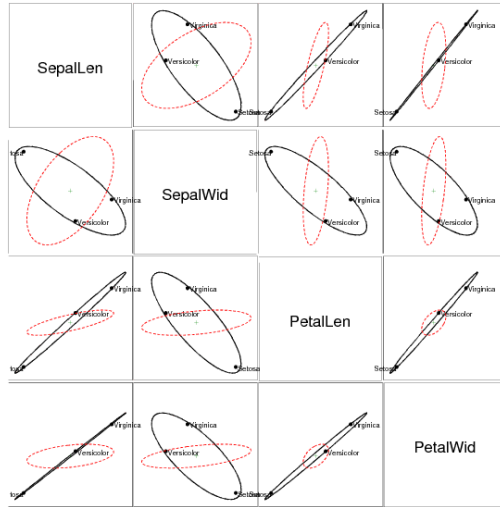
Graphic ideas: (a) Data ellipses summarize H & E (co)variation; (b) Scale H ellipse so it projects outside E ellipse *iff* effect is significant (Roy test)



HE plot matrices

HE plots in a scatterplot matrix show effects for all pairs of responses.

For the iris data, the Species means are highly correlated on all variables except Sepal length.



61

HE plots: 2-way MANOVA

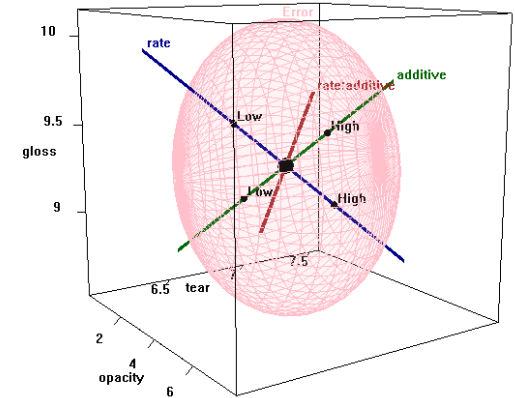
Plastic film data: 2x2 MANOVA
(gloss, opacity, tear) ~ rate*additive

MANOVA tests show that both main effects are significant:

```
Type II MANOVA Tests: Roy test statistic
                        Df approx F Pr(>F)
rate                    1 7.5543 0.003034 **
additive                1 4.2556 0.024745 *
rate:additive           1 1.3385 0.301782
```

HE plot shows the nature of these effects, e.g.,

high rate: ↑tear, ↑opacity, ↓gloss



1 df tests: H ellipsoid collapses to a line

62

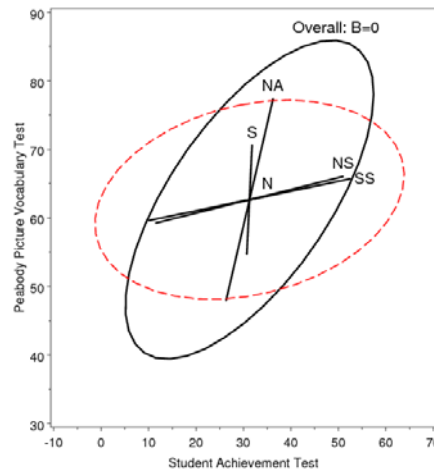
HE plots: Multivariate regression

Rohwer data: Cognitive ability and PA tests: n=37, Low SES group

(SAT, PPVT, Raven) ~ n + s + ns + na + ss

- Only one predictor, NA, is (barely) significant

- Yet, overall multivariate test: $H_0: \mathbf{B} = \mathbf{0}$ is highly so!



63

Public data, public graphics

- Worldmapper – the world in cartograms
- ManyEyes, Swivel – Facebook for data & graphs
- Gapminder – analyzing trends and relations
- Online communities?
- Visualization tools: availability & accessibility

64

Worldmapper: The world in cartograms

How to visualize social, economic, disease, ... data for geographic units?
 worldmapper.com : cartograms: area ~ variable of interest (350+ maps)

WORLDMAPPER *The world as you've never seen it before*

Home | Map Categories | Thumbnail Index | A-Z Map Index | About Worldmapper | Help

Worldmapper is a collection of world maps, where territories are re-sized on each map according to the subject of interest. There are 366 maps, also available as PDF posters. Use the menu above or click on a thumbnail image below to view a map.

Reference maps ...

- Total Population
- Land Area
- Labelled Map
- Appendix A (Areas included)

Newest maps ...

- Often Preventable Deaths
- Morphing animation
- Deaths from Non-Communicable Illnesses
- All Injury Deaths

Worldmapper: The world in cartograms

Carbon Emissions 2000

Carbon dioxide causes roughly 60% of the 'enhanced greenhouse effect' or global warming resulting from certain gases emitted by human activities. In 2000 there were almost 23 billion tonnes of carbon dioxide emitted worldwide. Of this, 28% came from North American territories; 0.09% came from Central African territories.

Emissions of carbon dioxide vary hugely between places, due to differences in lifestyle and ways of producing energy. Whilst people living in 66 territories emitted less than 1 tonne per person in 2000, more than 10 tonnes per person were emitted by people living in the highest polluting 21 territories that year.

Territory size shows the proportion of carbon dioxide emissions in 2000 that were directly from those territories.

MOST AND LEAST CARBON DIOXIDE EMISSIONS IN 2000

Rank	Territory	Value	Rank	Territory	Value
1	Qatar	64	181	Democratic Republic Congo	0.005
2	Bahrain	27	192	United Republic of Tanzania	0.004
3	Brunei Darussalam	21	193	Malawi	0.004
4	Kuwait	21	194	Liberia	0.004
5	Trinidad & Tobago	20	195	Comoros	0.004
6	Luxembourg	19	196	Niger	0.004
7	United States	19	197	Burundi	0.004
8	United Arab Emirates	18	198	Cambodia	0.045
9	Australia	18	199	Chad	0.047
10	Saudi Arabia	17	200	Afghanistan	0.040

tonnes of carbon dioxide emitted in 2000 per person living in that territory*

"If the world does not learn now to show respect to nature, what kind of future will the new generations have?"
 Rigobera Menchú Tum, 1992

© Copyright 2006, 2010 Group University of Bath and Bath Spa University of Bath

Worldmapper: The world in cartograms

Richest Fifth

Central Africa is the region with the lowest earnings per person amongst the richest fifth of the population of each territory. The anomaly in this region is Equatorial Guinea, where the world's fourth highest earnings per person in the richest fifth of the population are found. Equatorial Guinea is quite small on the map due to its tiny total population of 500 thousand. The Democratic Republic of Congo and Burundi, also in Central Africa, are home to some of the poorest 'richest fifths of population'. The regions where the richest fifth make the most money are North America, Western Europe and Japan.

Territory size shows the earnings of the richest fifth of the population living there, as a proportion of the earnings of the poorest tenth living in all territories.

HIGHEST AND LOWEST ANNUAL EARNINGS OF THE RICHEST FIFTH OF POPULATION

Rank	Territory	Value	Rank	Territory	Value
1	Luxembourg	132280	191	Tajikistan	168
2	United States	81188	192	Madagascar	158
3	Ireland	79106	193	Guinea-Bissau	157
4	Equatorial Guinea	70025	194	Sierra Leone	152
5	Norway	68825	195	Yemen	1720
6	Costa Rica	66115	196	Democratic Republic Congo	1648
7	Hong Kong (China)	66127	197	Burundi	1636
8	Canada	65778	198	Ethiopia	1502
9	Australia	65442	199	Mali	1481
10	Switzerland	65442	200	United Republic of Tanzania	1502

earnings by the richest fifth of the population in USD purchasing power parity (PPP) per rich person

"The 51 year old former rickshaw driver and millionaire businessman was apprehended in his pajamas by South African police, who ... accused him of helping to finance a botched coup ... in the oil-rich state of Equatorial Guinea." *Vivienne Walt, 2004*

© Copyright 2006, 2010 Group University of Bath and Bath Spa University of Bath

Worldmapper: The world in cartograms

HIV Prevalence

HIV, or Human Immunodeficiency Virus Infection, attacks the immune system. It eventually causes AIDS, which stands for Acquired Immune Deficiency Syndrome. With cases first recognised in the United States in 1981, AIDS increases the risk of many infections and tumours.

In 2003, the highest HIV prevalence was Swaziland, where 38%, or almost 4 in every 10 people aged 15 to 49 years, were HIV positive. All ten territories with the highest prevalence of HIV are in Central and Southern Africa.

Transmission of HIV is through sex, using infected needles, and in the womb. Infected children are not shown here. HIV/AIDS often has an acquired social stigma.

Territory size shows the proportion of all people aged 15-49 with HIV (Human Immunodeficiency Virus) worldwide, living there.

MOST PEOPLE WITH HIV

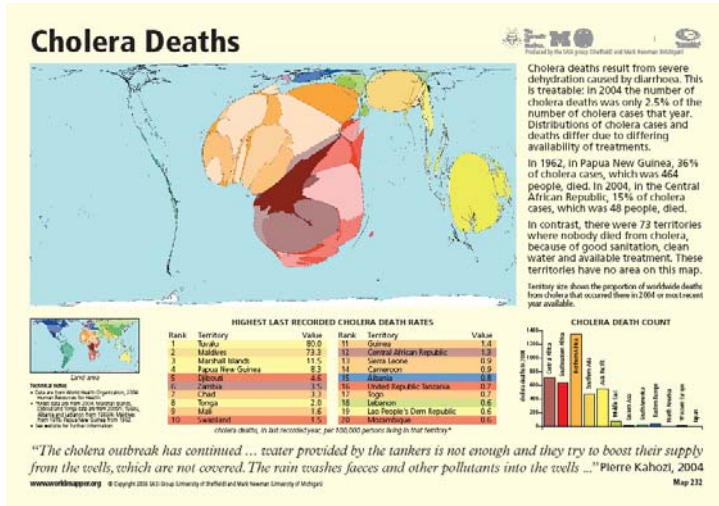
Rank	Territory	Value	Rank	Territory	Value
1	Swaziland	38	195	United Republic of Tanzania	2.0
2	Botswana	27	196	Gabon	2.1
3	Lesotho	25	197	Cote d'Ivoire	2.0
4	Senegal	24	198	Cameroon	2.2
5	South Africa	21	199	Kenya	2.0
6	Zambia	16	200	Rwanda	2.0
7	Malawi	16	201	Malawi	2.0
8	Kenya	16	202	Malawi	2.0
9	Central African Republic	13	203	Nigeria	2.0
10	Microronesia	12	204	Madagascar	2.0

percentage of people aged 15 to 49 living with HIV*

"I have come to the conclusion that HIV/AIDS is not entirely about death. People die and will continue to die for one reason or the other. AIDS is also about the living."
 Kiliza Ngonzi, 2004

© Copyright 2006, 2010 Group University of Bath and Bath Spa University of Bath

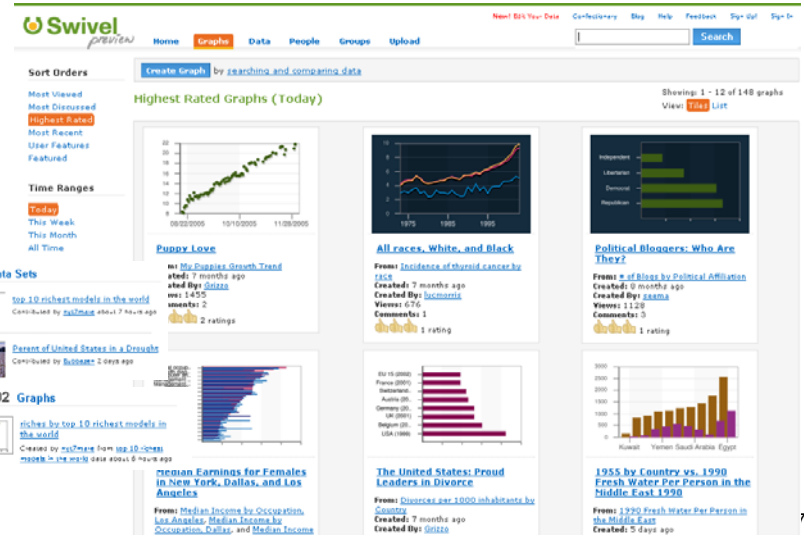
Worldmapper: The world in cartograms



69

Swivel: Facebook for data & graphs

<http://www.swivel.com/> : upload & explore your data



70



ManyEyes: Social data analysis & visualization

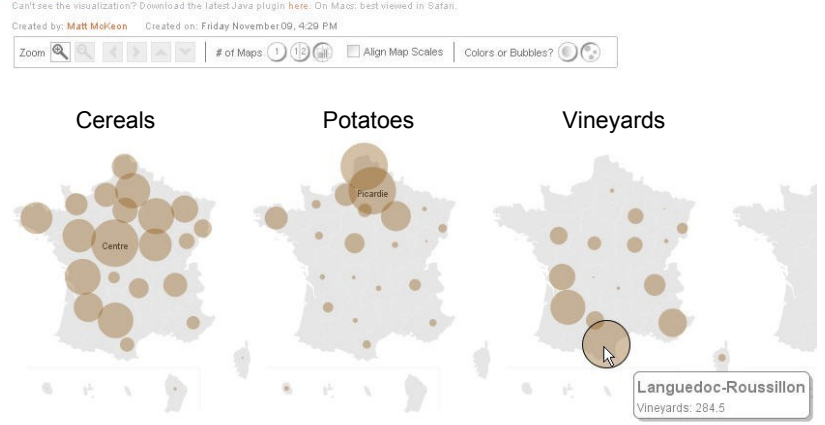
<http://services.alphaworks.ibm.com/manyeyes/app>

The ManyEyes application interface includes a navigation menu with options like 'explore visualizations', 'participate register', and 'learn more'. It features a 'Try Our Featured Visualizations' section with examples such as 'U.S. Teen Births in 2003', 'National Parks Bubbles', and 'Brazilian Constitution'. There is also a 'Featured Topic Hubs' section with categories like 'Transportation', 'Decision 2008', and 'OECD Factbook 2007'. The interface is branded with the ManyEyes logo and the IBM logo at the bottom.



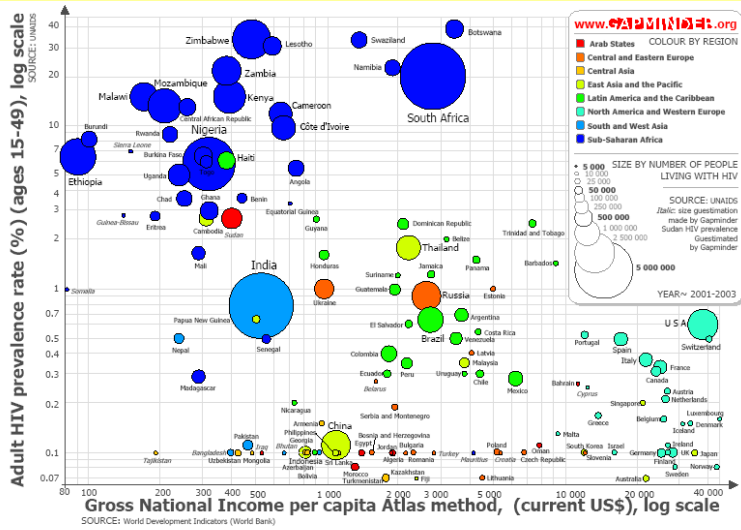
ManyEyes: Comparing maps

Visualizations : Vineyards in France

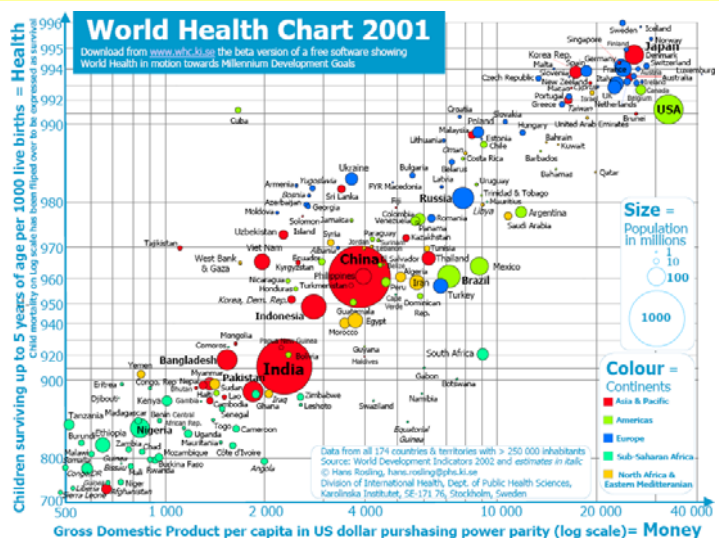


72

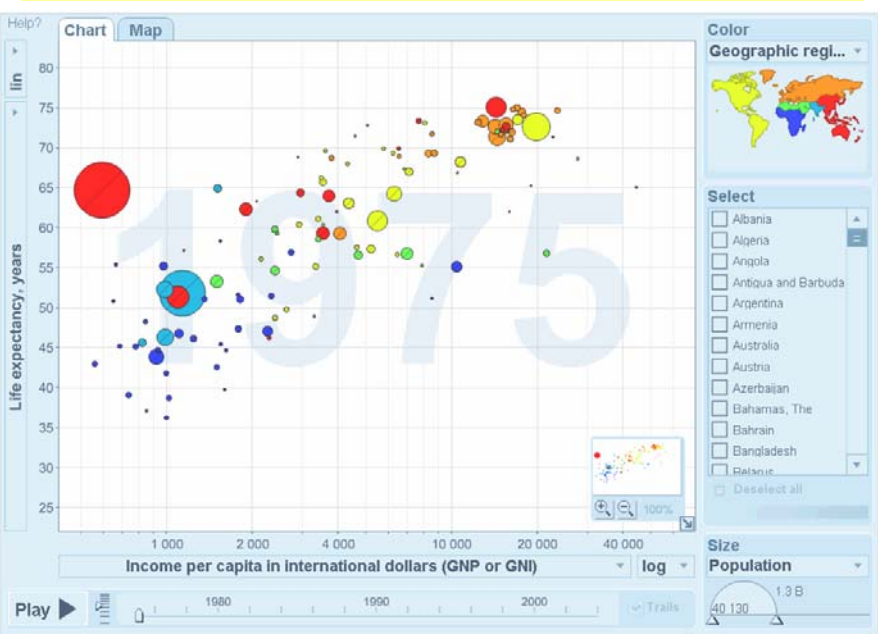
Gapminder: Exploring trends in world health



Gapminder: Exploring trends in world health

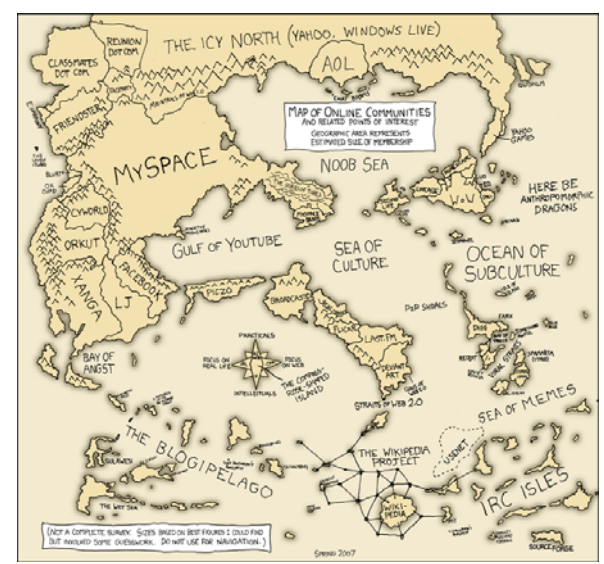


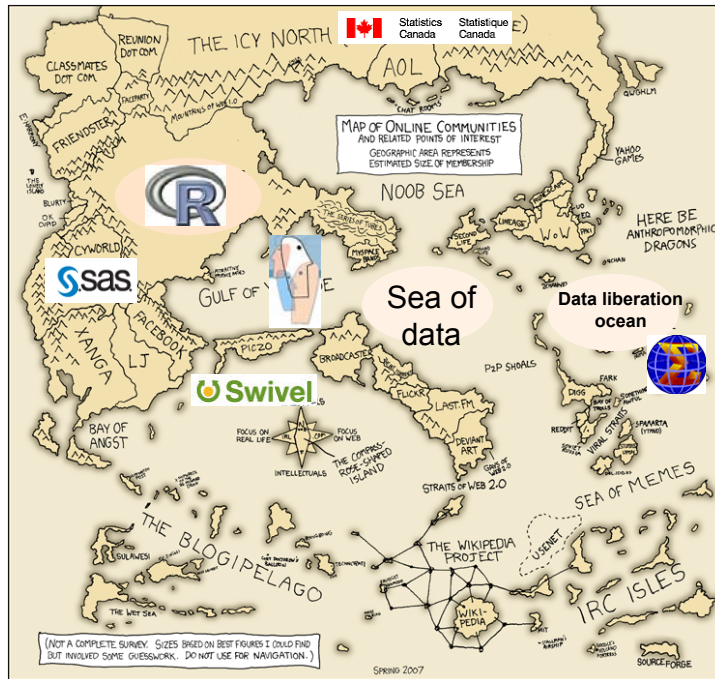
Gapminder: visualizing trends and relations



Online communities

Map by Randall Munroe of online communities. Area ~ # users (Berners-Lee projection)

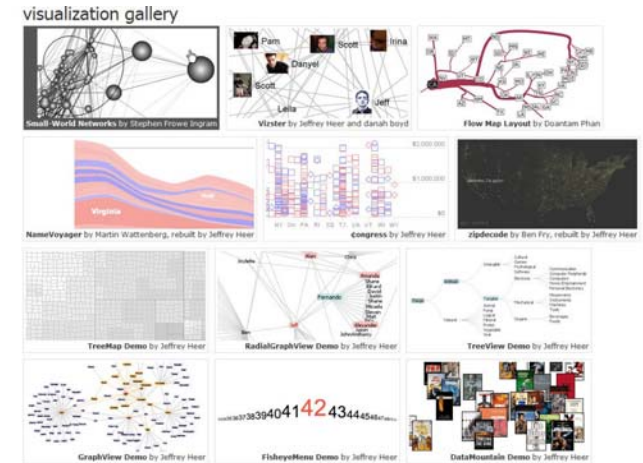




What else do we need? Visualization Tools: Availability & Accessibility

Practical power = Statistical power x Probability of use (J.W. Tukey)

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



3

Summary & conclusions

- Past history teaches us that—
 - Statistical graphics always has a purpose: tell a story, inform a decision
 - Statistical graphics is hard work, but can have both beauty & truth
 - Advances require: data, technology & visual thinking

Summary & conclusions

- Present history suggests that—
 - Generalized models → generalized graphics
 - Consumers & producers: different strokes for different folks
 - Most want graphical toasters: data in, picture out (but: what picture(s)?)
 - Some want complete control of graphic details
 - Graphic developers want it all: freedom to invent

Summary & conclusions

- The future of statistical graphics?
 - New visualizations for statistical methods
 - Open source, open data, online communities
 - Leveraging new technologies for visualization, analysis & data → {insight, practice}