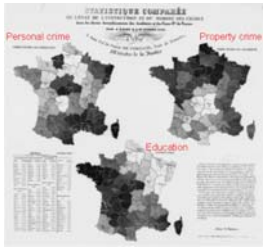
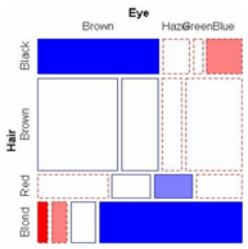
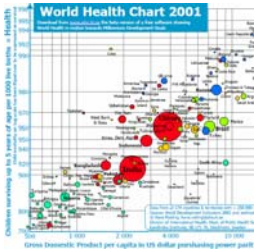






Visions of the Past, Present and Future of Statistical Graphics

Swiss Statistics Meeting, 2007

Michael Friendly

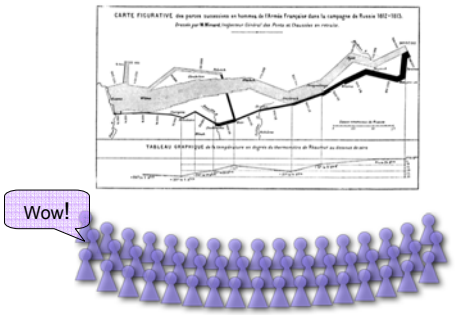
York University, Toronto, CA

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?

Different graphs for different purposes



Presentation

Exploration

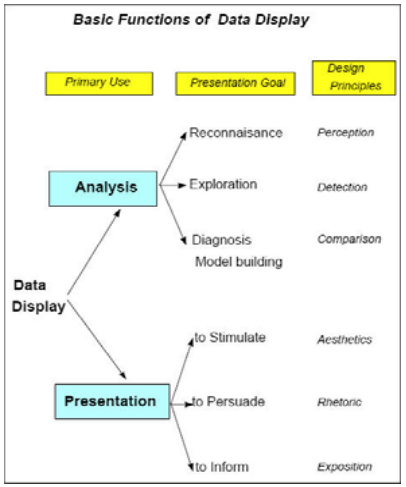
Goal: the Wow! experience
Single image for a large audience

Goal: the Ah ha! Experience

Many images, for a narrow audience (you!), linked to analysis

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

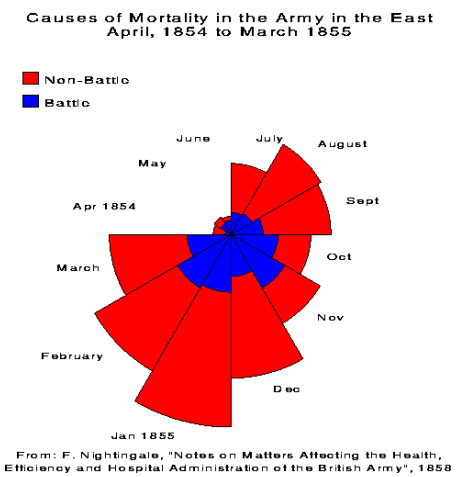


Presentation graph: Nightingale's coxcomb

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

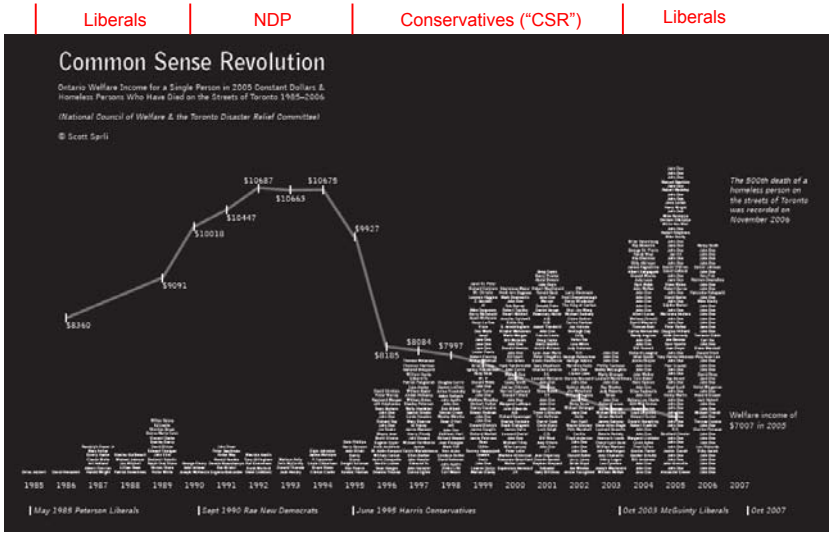
She used this to argue for better field hospitals (MASH units)

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



From: F. Nightingale, "Notes on Matters Affecting the Health, Efficiency and Hospital Administration of the British Army", 1858

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

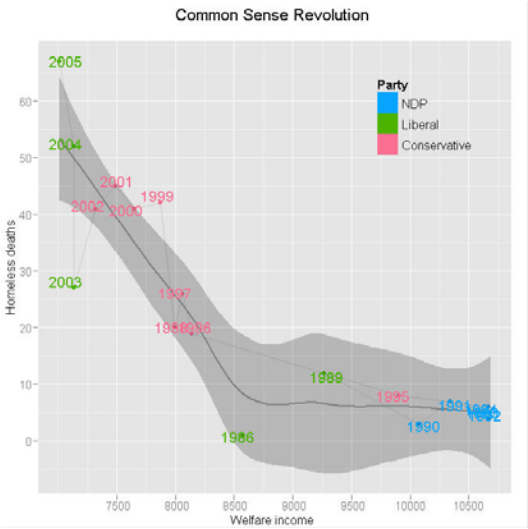


Analysis graph: Deaths vs. Income

Scatterplot of deaths vs. income

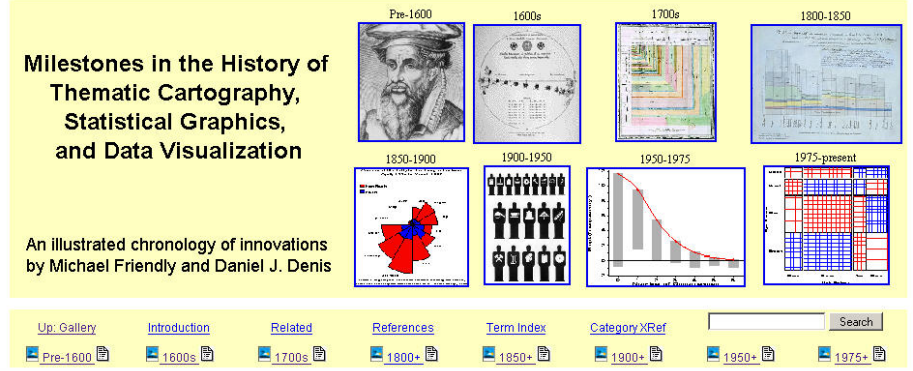
- Loess smooth + CI band
- Labels: year
- Color: party in power

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



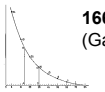
Project goals:

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

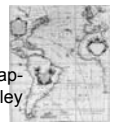
Milestones: Content Overview

Every picture has a story – Rod Stewart

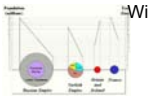
c. 550 BC: The first world map? (Anaximander of Miletus)



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



1701: First contour map- Edmund Halley

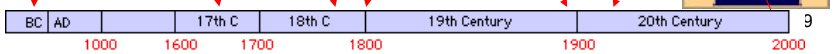


1801: Pie chart, circle graph - William Playfair

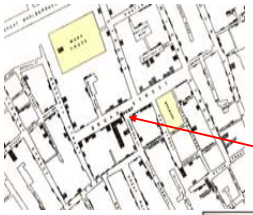
1896: Bivariate map- Jacques Bertillon

1924: Pictograms- Otto Neurath

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



1850-1900: Golden Age

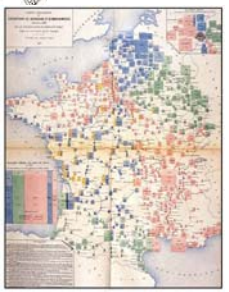


1855: Dot map of disease data (cholera)- John Snow

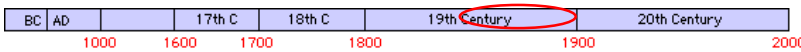
Broad St. pump

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

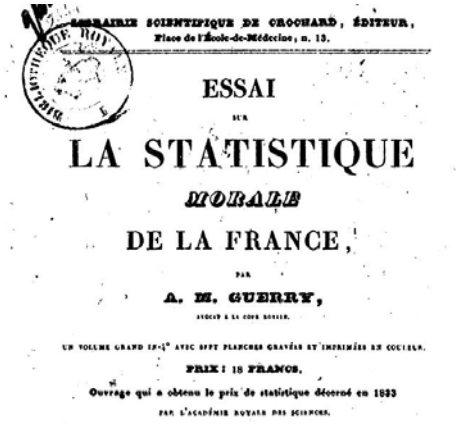
A.-M. Guerry and the invention of social science

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".



LIBRAIRIE SCIENTIFIQUE DE CROCHARD, ÉDITEUR,
Place de l'École-de-Médecine, n. 13.

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

UN VOLUME GRAND IN-4° AVEC SEPT PLANCHES GRAVÉES ET IMPRIMÉES EN COULEUR.
PRIX : 15 FRANCS.
Ouvrage qui a obtenu le prix de statistique décerné en 1833
PAR L'ACADÉMIE ROYALE DES SCIENCES.

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)

Social context of crime in 1820s

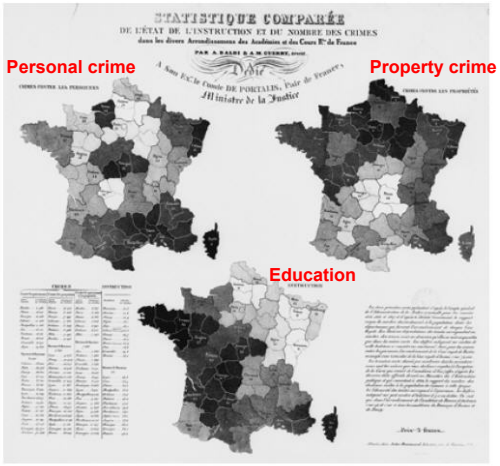
- What to do about crime?
  - Crime a serious concern: Explosive growth in Paris, widespread unemployment, emergence of "dangerous classes"
  - Liberal ("philanthrope") view: increase education, better prison conditions, religious instruction, better diet (bread and soup)
  - Conservative view: build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

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**

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...*

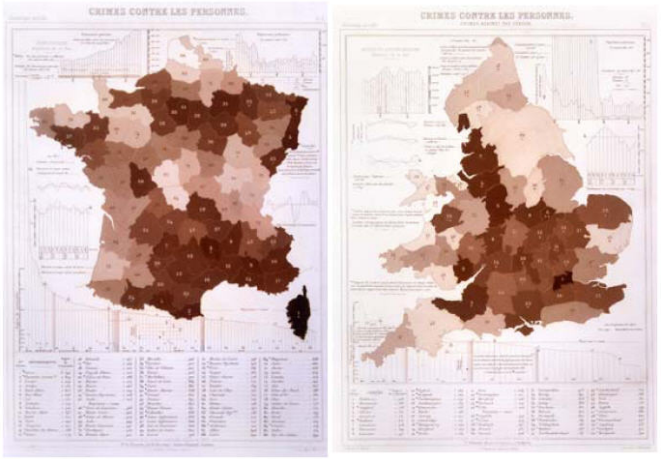
Dayenu!

- 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!


STATISTIQUE MORALE
DE
L'ANGLETERRE
COMPARÉE AVEC
LA STATISTIQUE MORALE
DE
LA FRANCE
PAR A.-M. GUERRY
ATLAS
PROSPECTUS
PARIS
J.-B. BAILLIÈRE ET FILS,
LONDRES. NEW-YORK.
MADRID.

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

Comparing France and England



Statistique analytique: General causes of crime

Plate XVII: M. Guerry's magnum opus

Goal:

• Show multivariate factors associated with distribution of crime

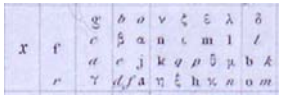
• Before invention of correlation

Entries: Codes for factors

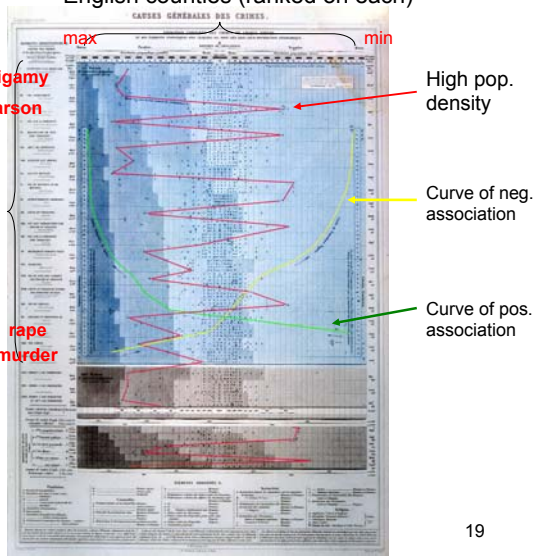
• Pop: (% Irish, domestics, …)

• Criminality: (male, young, …)

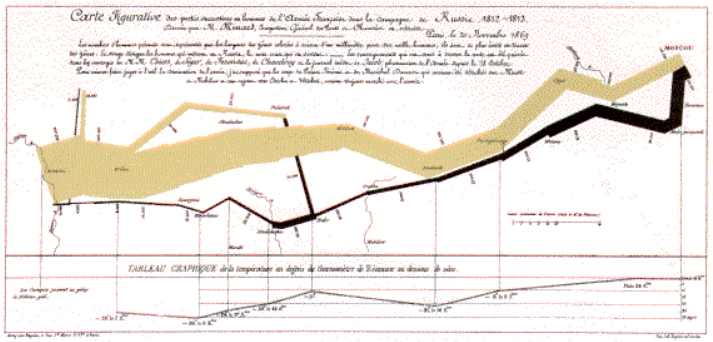
• Religion (Anglicans, dissenters, …)



English counties (ranked on each)



The graphic vision of C. J. Minard

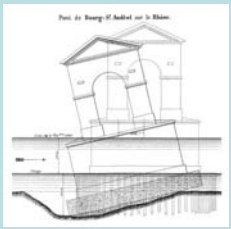


- Marey (1878): *"defies the pen of the historian in its brutal eloquence"*
- Tufte (1983): *"the best statistical graphic ever produced"*

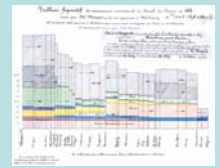
Why Minard?

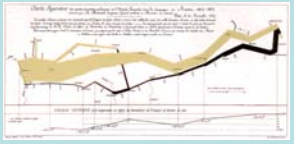
- Study breadth and depth of his work
  - How related to work in his time?
  - How related to modern statistical graphics?
  - How related to his personal history?

Civil Engineer for ENPC (1810-1842)



Visual Engineer for France (1843-1869)

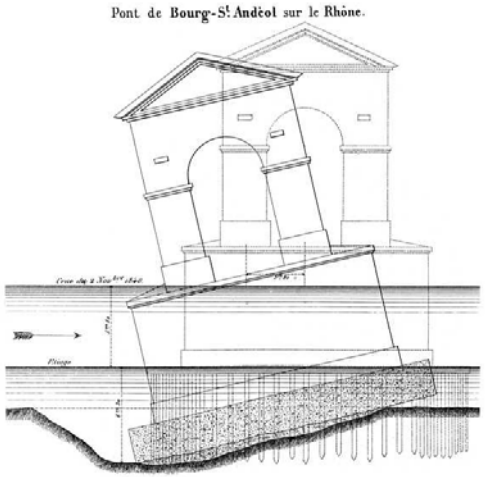




Visual thinking, visual explanation

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

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

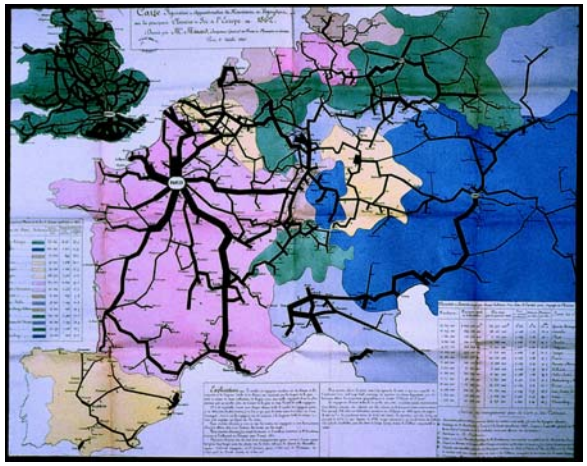


Visual tools for state planning

- 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

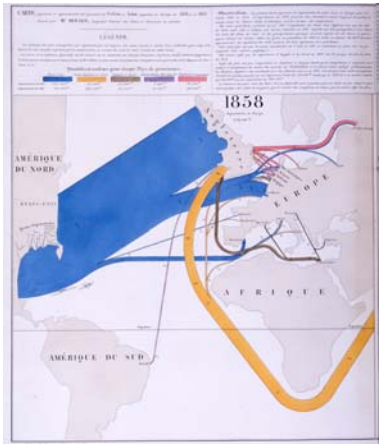
Flow maps as visual tools

Transport of passengers on the principal railroads in Europe in 1862



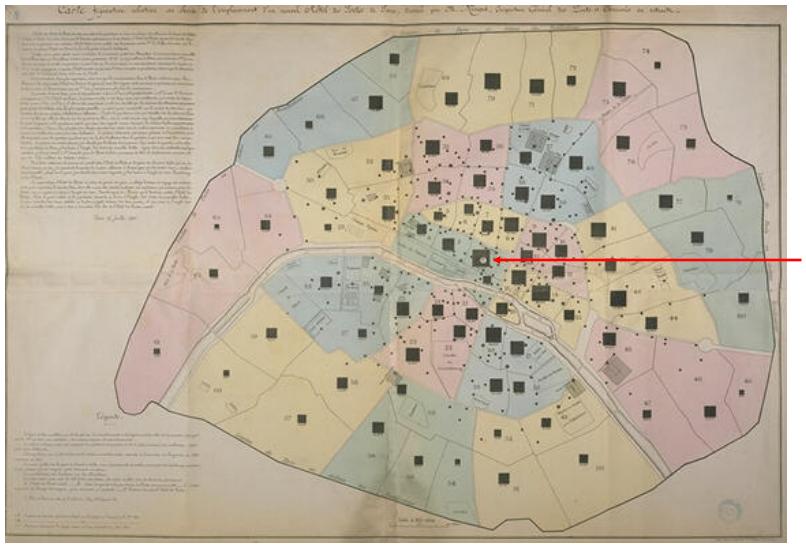
Effect of US civil war on cotton trade

Before



After

Where to build a new post office?(1867)

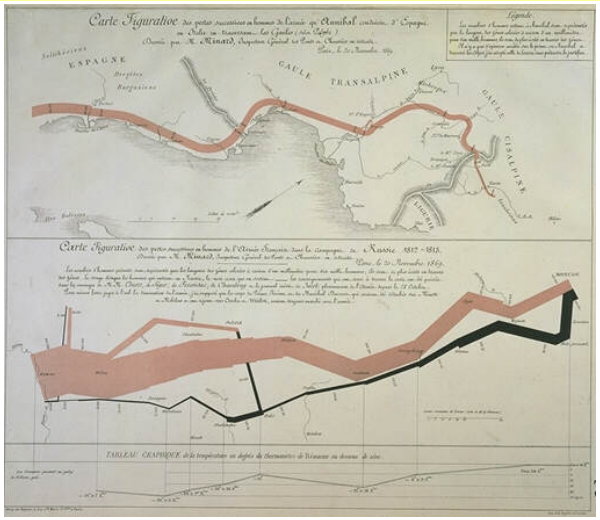


Center of gravity of pop. density

The March Re-Visited (1869)

Hannibal's retreat

Napoleon's 1812 campaign



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.

Visual smoothing → Insight

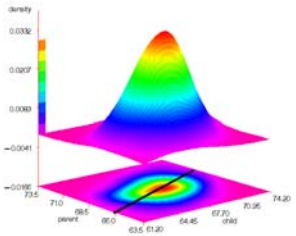
Table 9.1 One of Galton's correlation tables

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	7	2	2	—	—	—
71.5	—	—	—	1	—	1	3	4	10	4	2	—	—	—
70.5	—	—	1	—	1	6	5	10	14	12	3	3	—	—
69.5	—	1	—	1	3	3	9	18	14	7	4	3	1	—
68.5	1	—	1	3	5	10	26	28	20	11	4	—	—	—
67.5	—	3	3	5	10	9	15	20	12	4	2	1	—	—
66.5	3	3	5	2	7	11	14	7	5	2	1	—	—	—
65.5	1	9	5	7	11	11	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	187	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	—	—

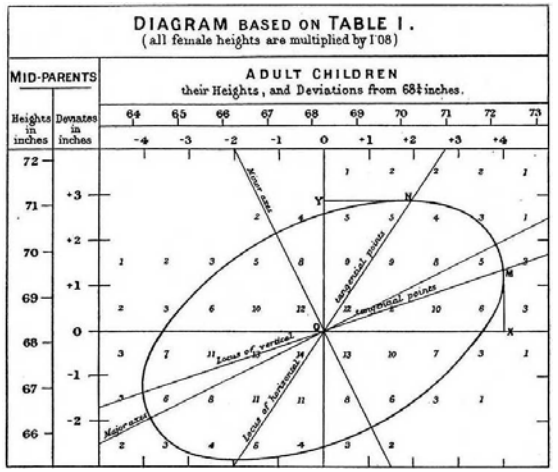
Source: Galton (1886), p. 68.

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 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

Galton's discovery of weather patterns-

Perhaps the most notable *purely graphic* discovery ever!

METEOROGRAPHICA,

OR

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.

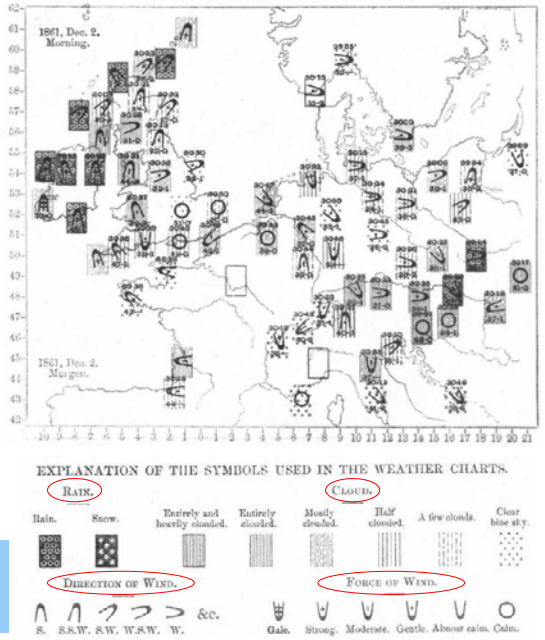
(Galton, 1863)

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.

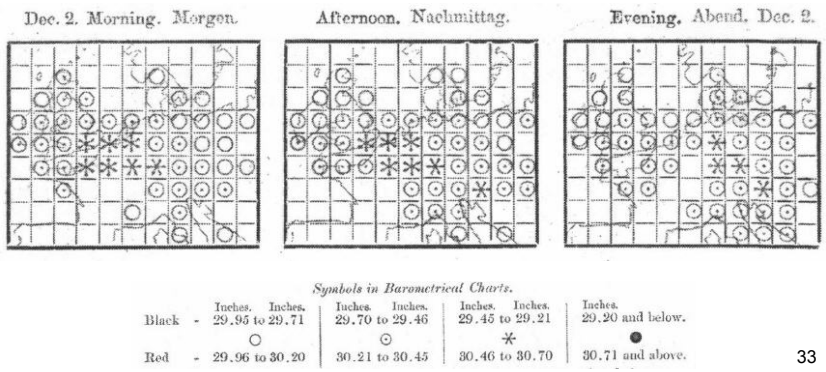
Graphic analysis: 3x31=93 maps, each with multivariate glyphs showing all variables

Visual ideas:
- Iconic symbols
- Multivariate glyphs (stamps!)



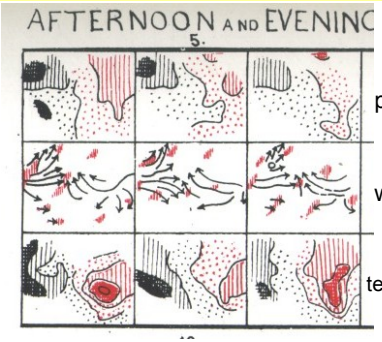
Visual abstraction → Patterns

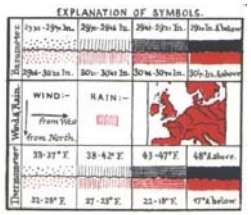
How to see patterns of geographical variation over time?
- Iconic symbols on a geographical grid
- "Small multiples:" separate graphs laid out for direct comparison



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





AM noon PM

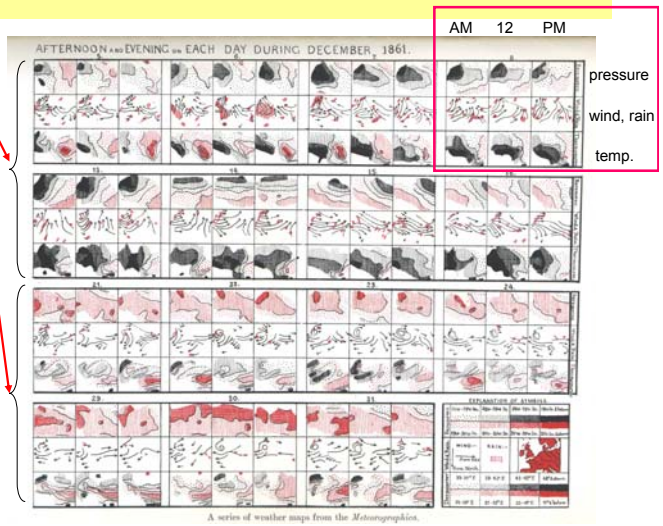
Data for Dec 5, 1861

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}

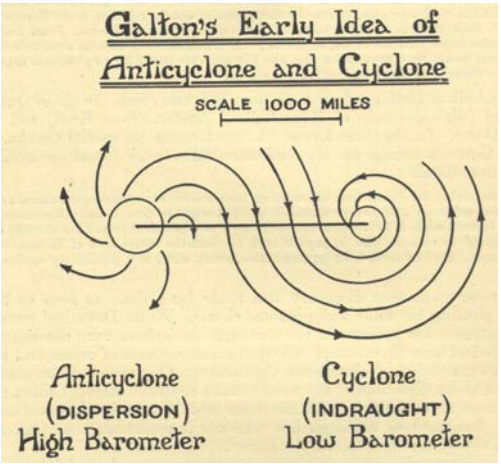
(try this with your software!)



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"

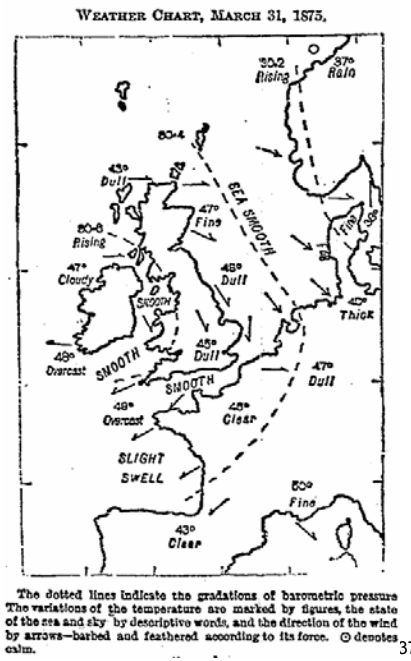
Theory:
- Explained by Dove's 'Law of Gyration'
- Prediction: reversed pattern (CW/CCW) in southern hemisphere – confirmed!



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)



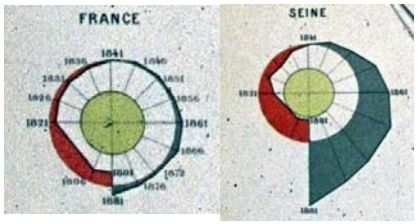
Statistical albums: Data → practice & Graphical excellence

- 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*
  - Gemany: local albums (Berlin, Frankfurt, etc.)
  - Switzerland: *Atlas graphique de la Suisse*
  - USA: Census atlases: 1870/80/90

Album de statistique graphique

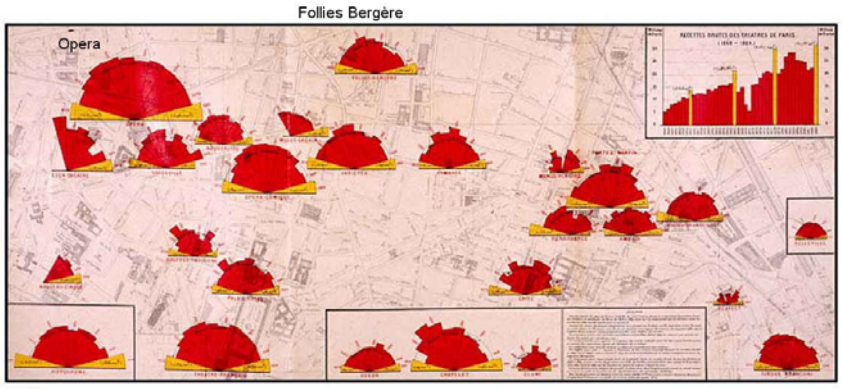
- The pinnacle of the Golden Age
- 18 volumes published 1879-1899

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



Album de statistique graphique

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



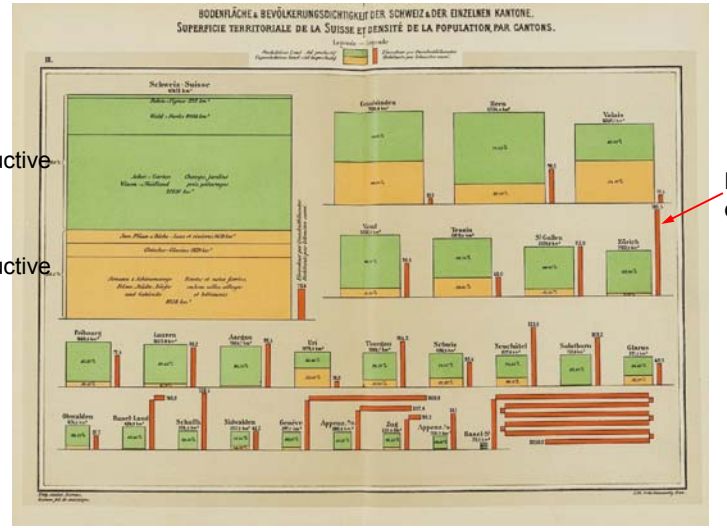
Atlas graphique de la Suisse (1897, 1914)

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

productive

unproductive

Pop. density



Atlas graphique de la Suisse (1897, 1914)

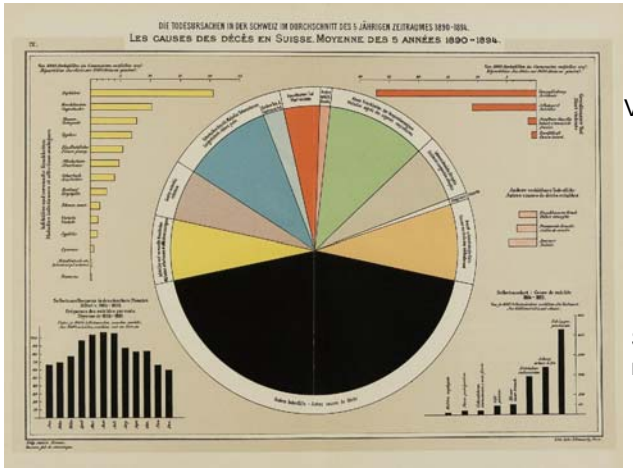
Causes of death, 1890—1894, color linking (Atlas, 1897)

Diseases

Violent death

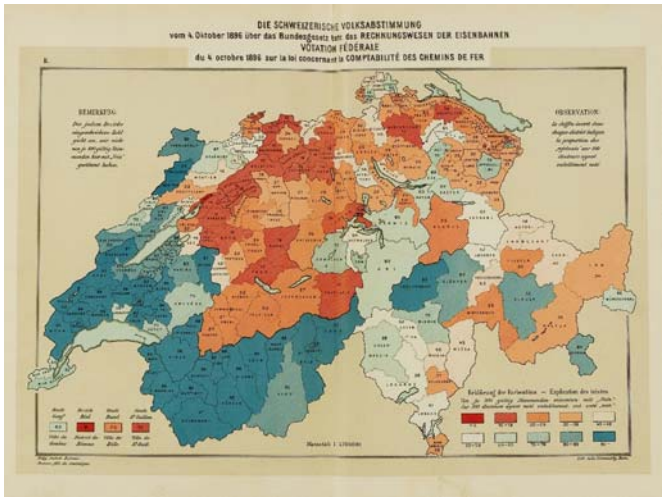
Suicide, by month

Suicide, by method



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

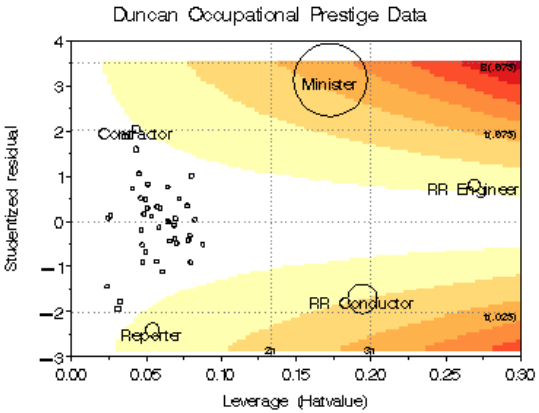
# of response variables	Classical linear models	Generalized linear models
1	LM family: $E(\mathbf{y})=\mathbf{X}\beta$, $V(y\|\mathbf{X})=\sigma^2\mathbf{I}$ ANOVA, regression, … Many graphical methods: effect plots, spread-leverage, influence, …	GLM: $E(\mathbf{y})=g^{-1}(\mathbf{X}\beta)$, $V=V[g^{-1}(\mathbf{X}\beta)]$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

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



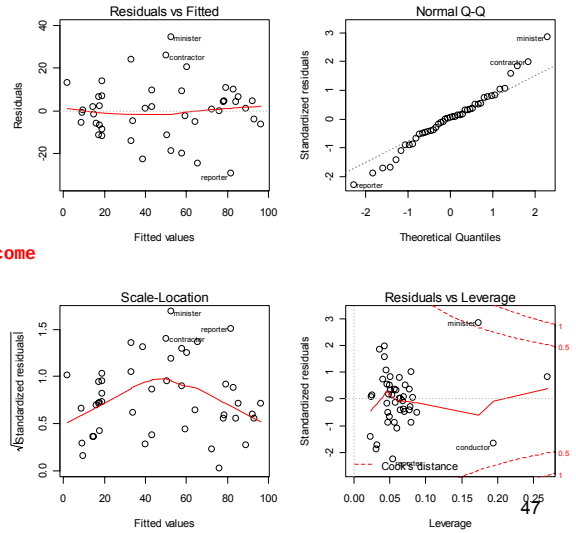
Model diagnosis: regression quartet

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)

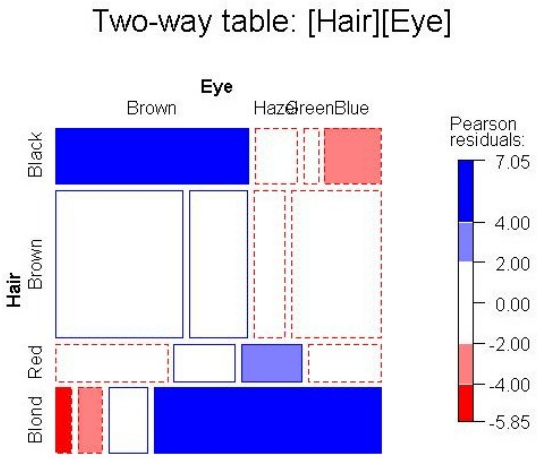


Graphical methods for GLMs

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

Visualizing categorical data: mosaic plots

Model: loglin(~ Hair + Eye)



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

Shading: ~ Pearson residual,

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

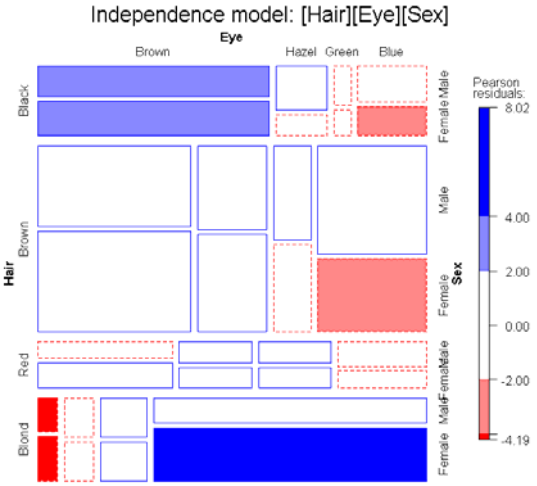
Color:

- 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)

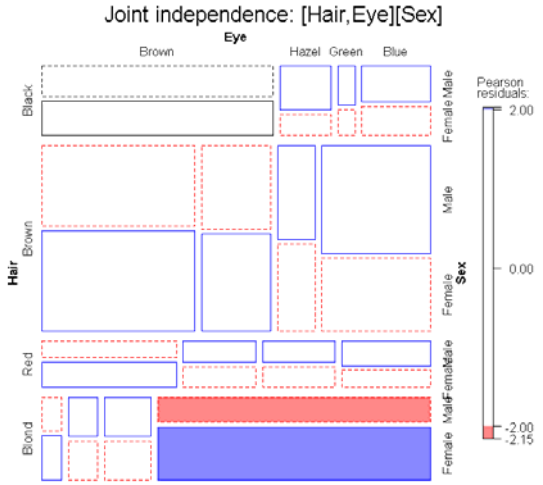


3+ way tables: split each tile ~ conditional proportions of the next variable.

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



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.

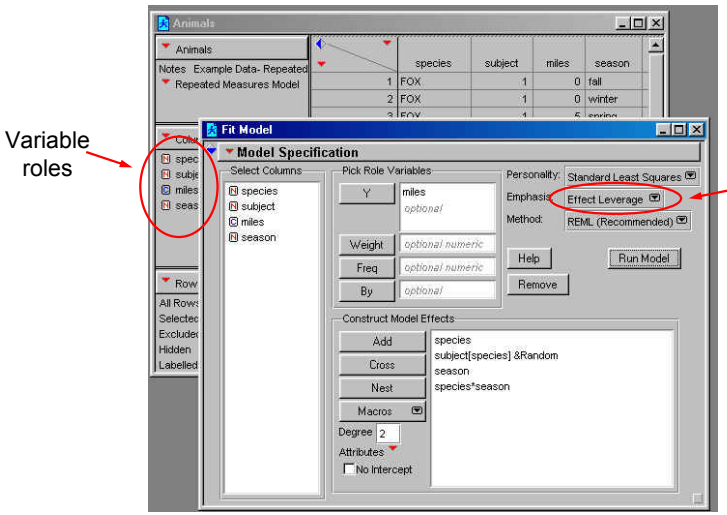
Mosaic displays: visual assessment of model fit:

Better model → "cleans the mosaic"

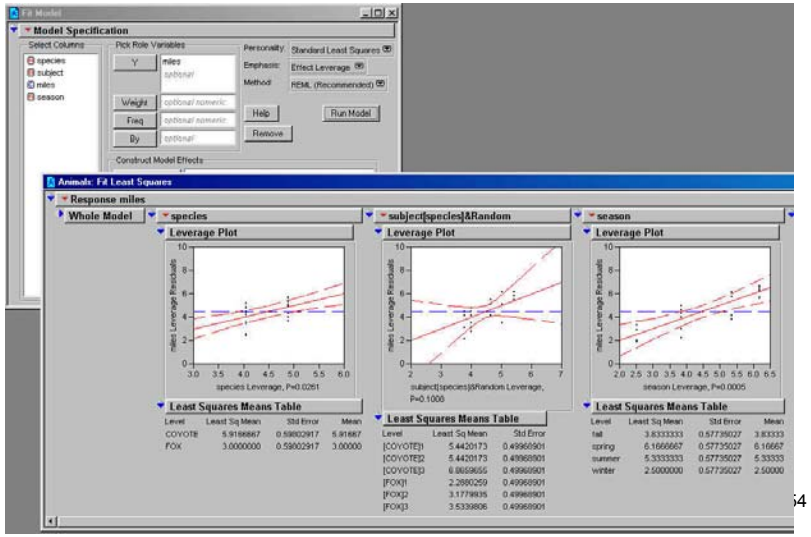
Cognitive interfaces for data exploration & model-fitting

- "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



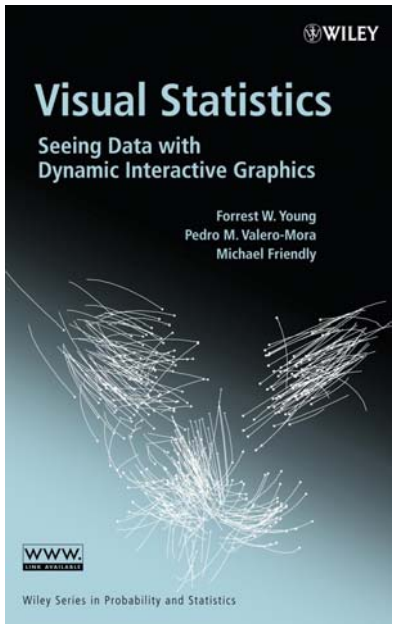
JMP: Model summary = graphs + numbers



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

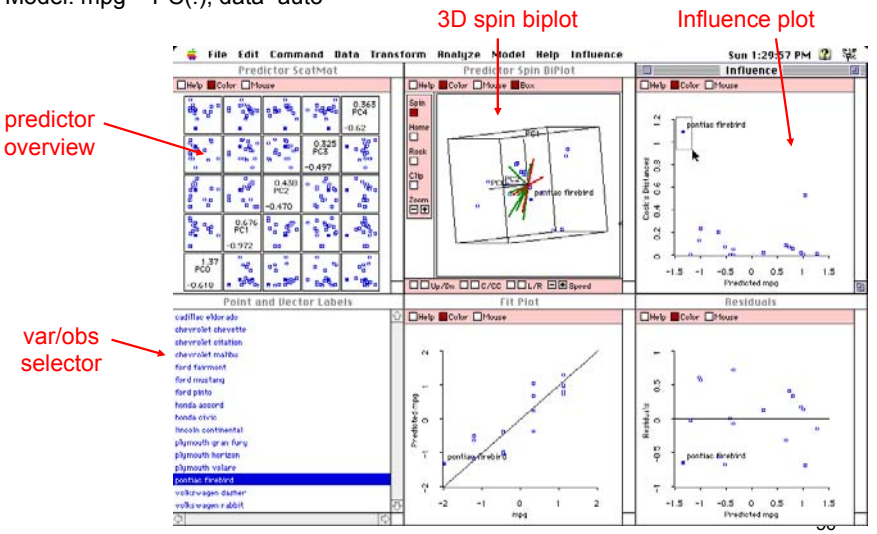
ViSta THE VISUAL STATISTICS SYSTEM

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

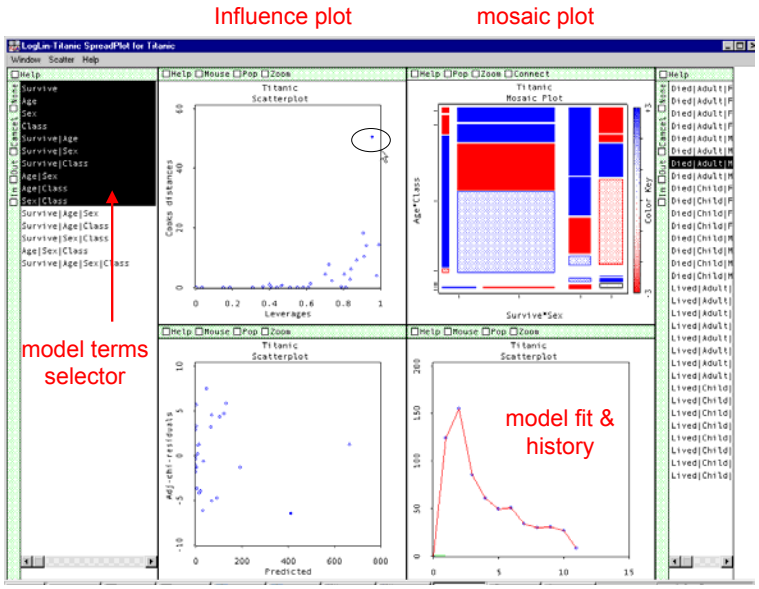


ViSta: regression spreadplots

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



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

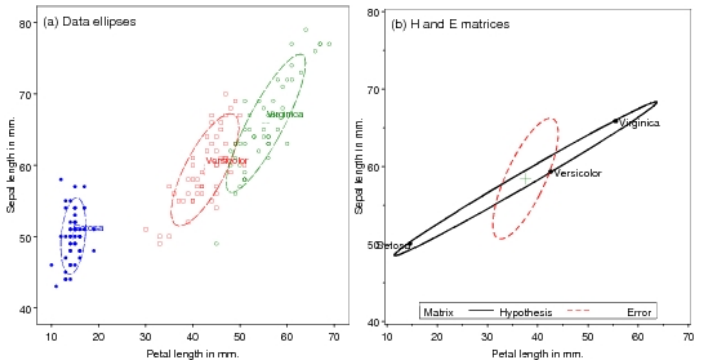
Graphical methods for MLMs

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

HE plots for MANOVA, MMReg

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.

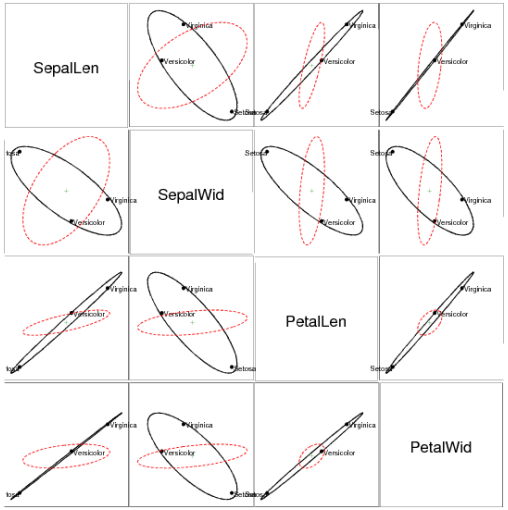
Graphic ideas: (a) Data ellipses summarize H & E (co)variation; (b) Scale H ellipse so it projects outside E ellipse *iff* effect is significant (Roy test)



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.



HE plots: 2-way MANOVA

Plastic film data: 2x2 MANOVA

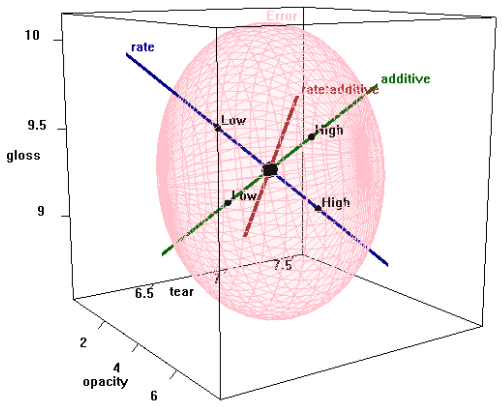
(gloss, opacity, tear) ~ rate*additive

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

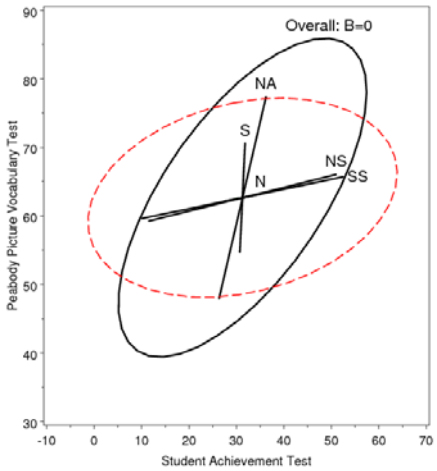
HE plots: Multivariate regression

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: **B = 0** is highly so!



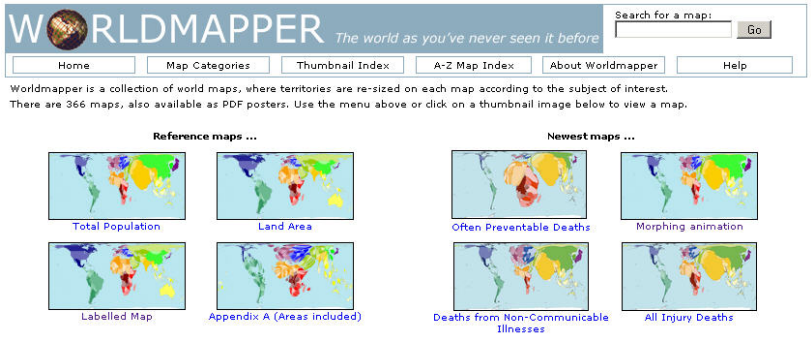
Public data, public graphics

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

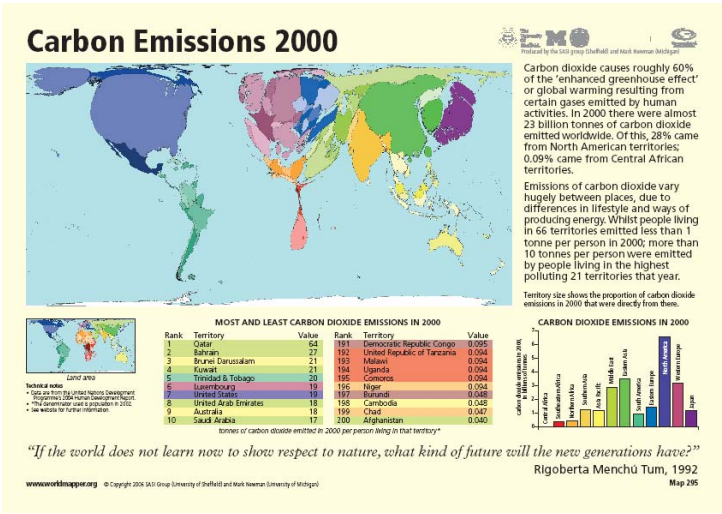
Worldmapper: The world in cartograms

How to visualize social, economic, disease, … data for geographic units?

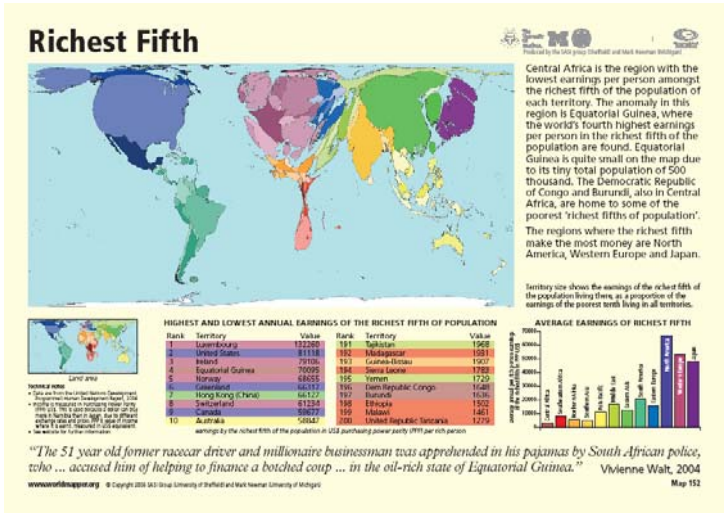
worldmapper.com : cartograms: area ~ variable of interest (350+ maps)



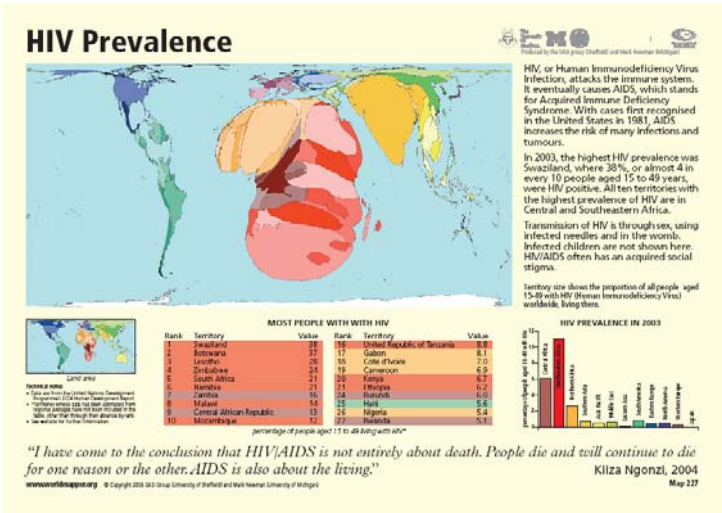
Worldmapper: The world in cartograms



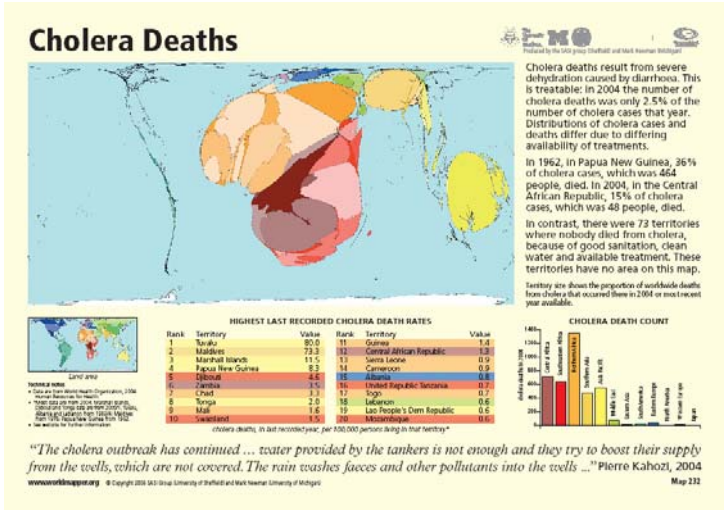
Worldmapper: The world in cartograms



Worldmapper: The world in cartograms

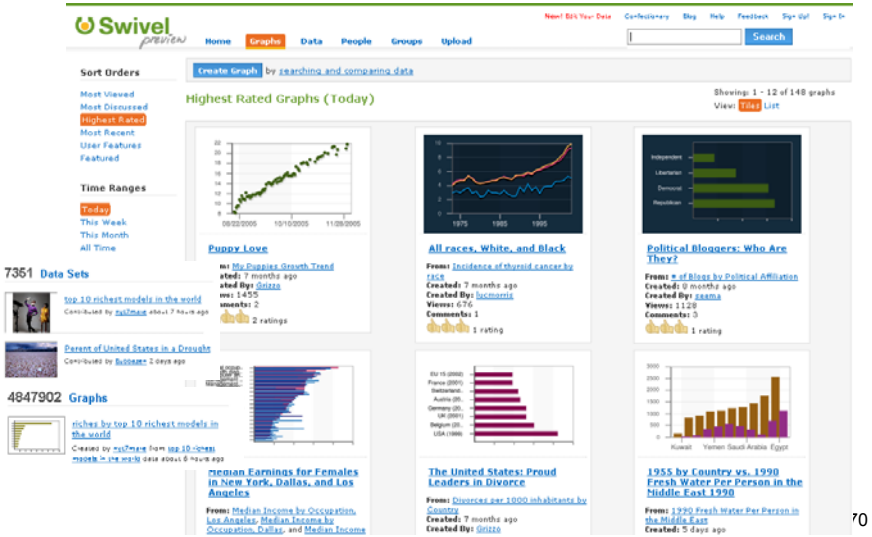


Worldmapper: The world in cartograms



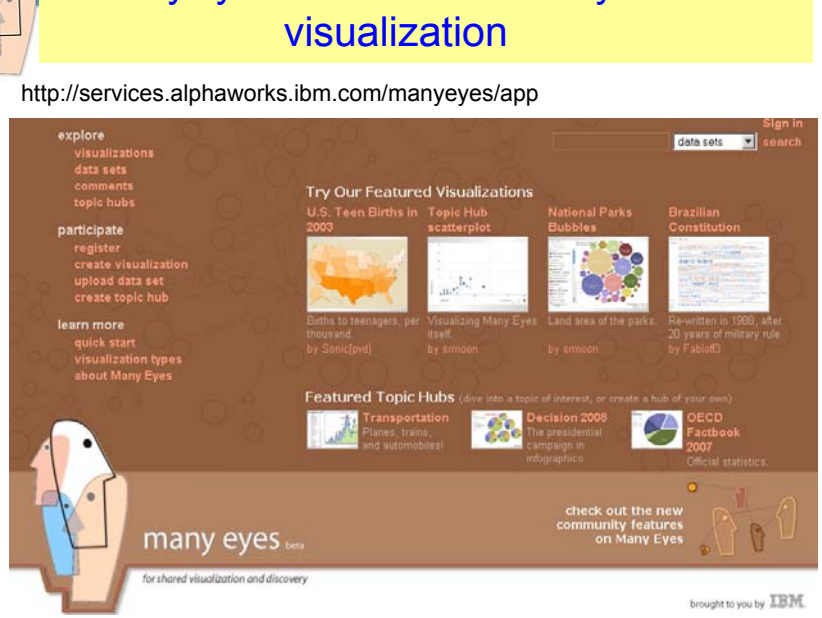
Swivel: Facebook for data & graphs

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

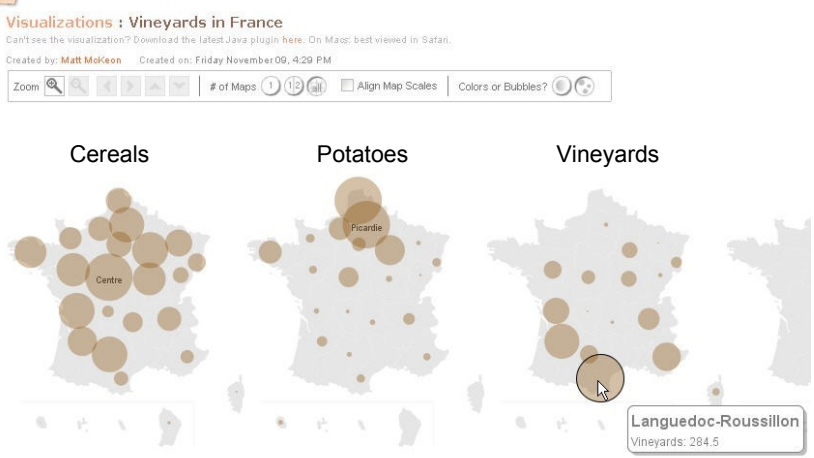


ManyEyes: Social data analysis & visualization

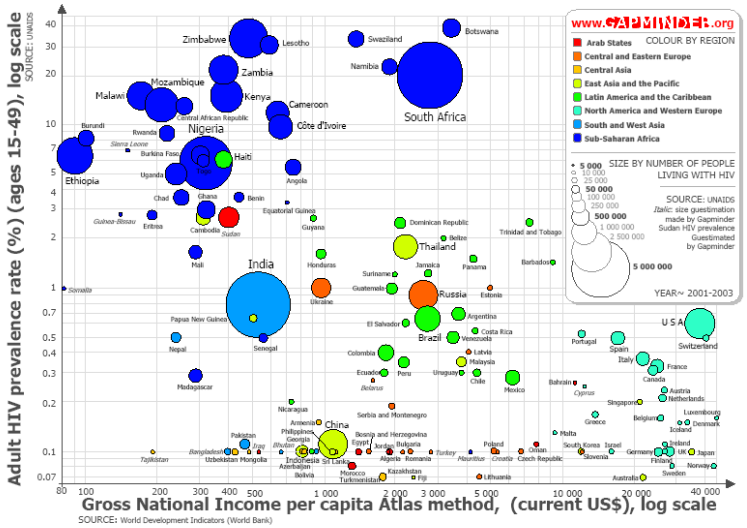
http://services.alphaworks.ibm.com/manyeyes/app



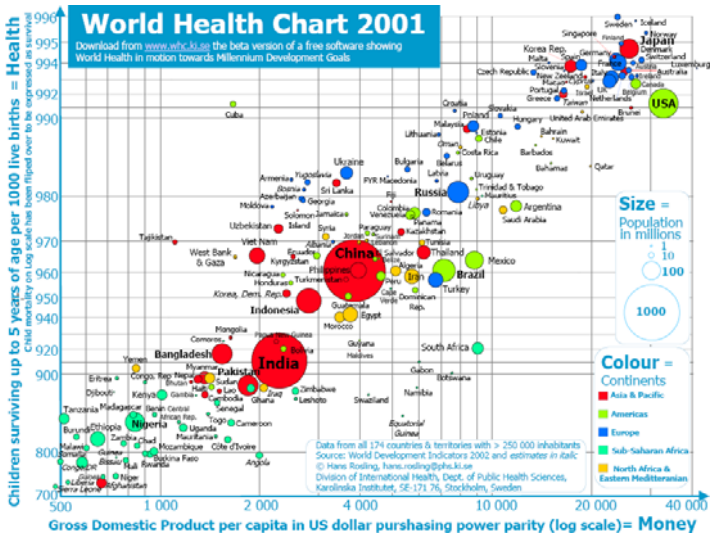
ManyEyes: Comparing maps



Gapminder: Exploring trends in world health



Gapminder: Exploring trends in world health

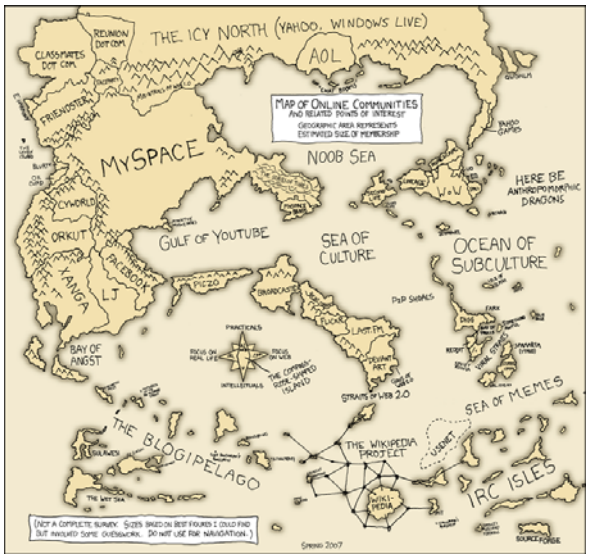


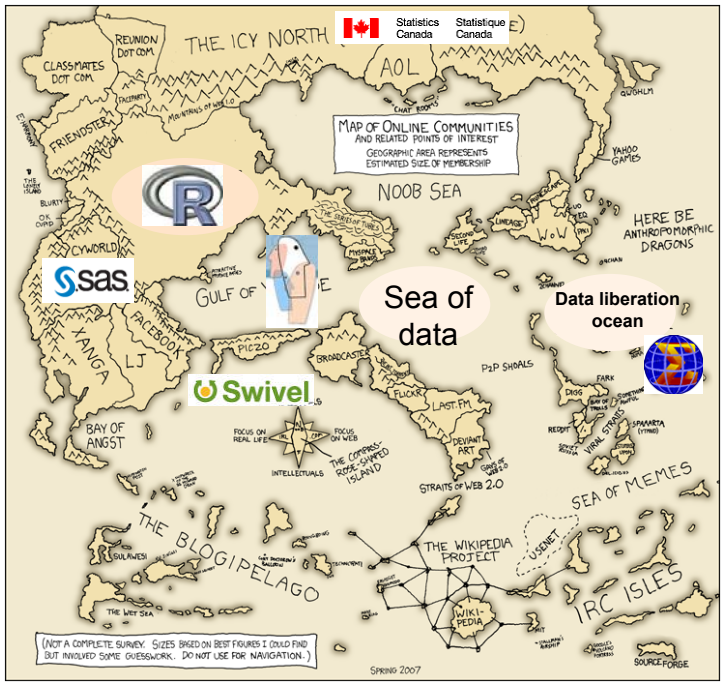
Gapminder: visualizing trends and relations

Online communities

Map by Randall Munroe of online communities. Area ~ # users

(Berners-Lee projection)

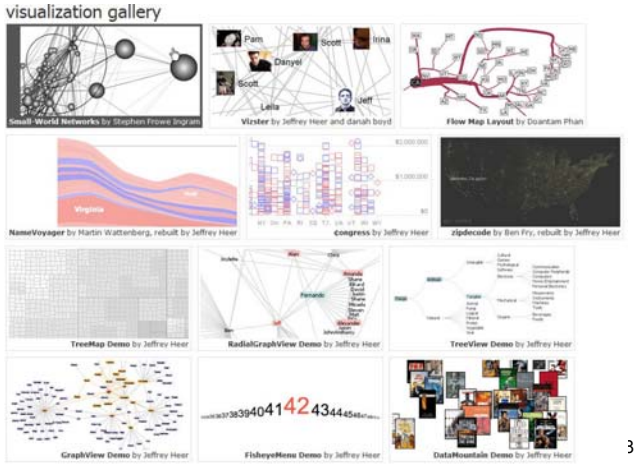




What else do we need?
Visualization Tools: Availability & Accessibility

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

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



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}