

Graphical information display

Summer 2014

Brent Stephens, Ph.D.

Civil, Architectural and Environmental Engineering



The Built Environment Research Group

advancing energy, environmental, and sustainability
research within the built environment
at Illinois Institute of Technology



web www.built-envi.com **email** brent@iit.edu **twitter** @built_envi

BERG summer meetings and presentations

Jun 9 - Brent - Zeineb's (former student) filter modeling

Jun 10 - Parham - Mapping outdoor particle size distributions to HVAC filtration efficiency for PM2.5 and ultrafine particles

Jun 11 - Torkan - Hot box measurement methods and results

Jun 12 - Hoaran/Jihad - Outdoor pollutant transport measurements

Jun 13 - n/a

Jun 16 - ~~Zack - OSBSS: In depth on 2-3 sensors/techniques (dual beam break; battery life extension; Builduino; etc.)~~

Jun 17 - Brent - Graphical display of information

Jun 18 - ****White Sox game 12 pm**** I have 10 tickets reserved!

Jun 19 – Akram - OSBSS: General overview of purpose, progress, example results, and path forward

Jun 23 -

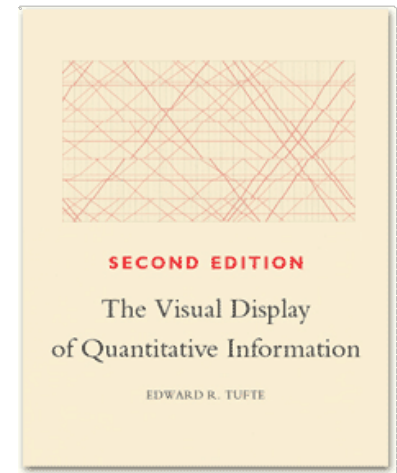
Jun 24 - Brent - Technical writing

Jun 25 - Rou Yi and Torkan - Energy and filter modeling

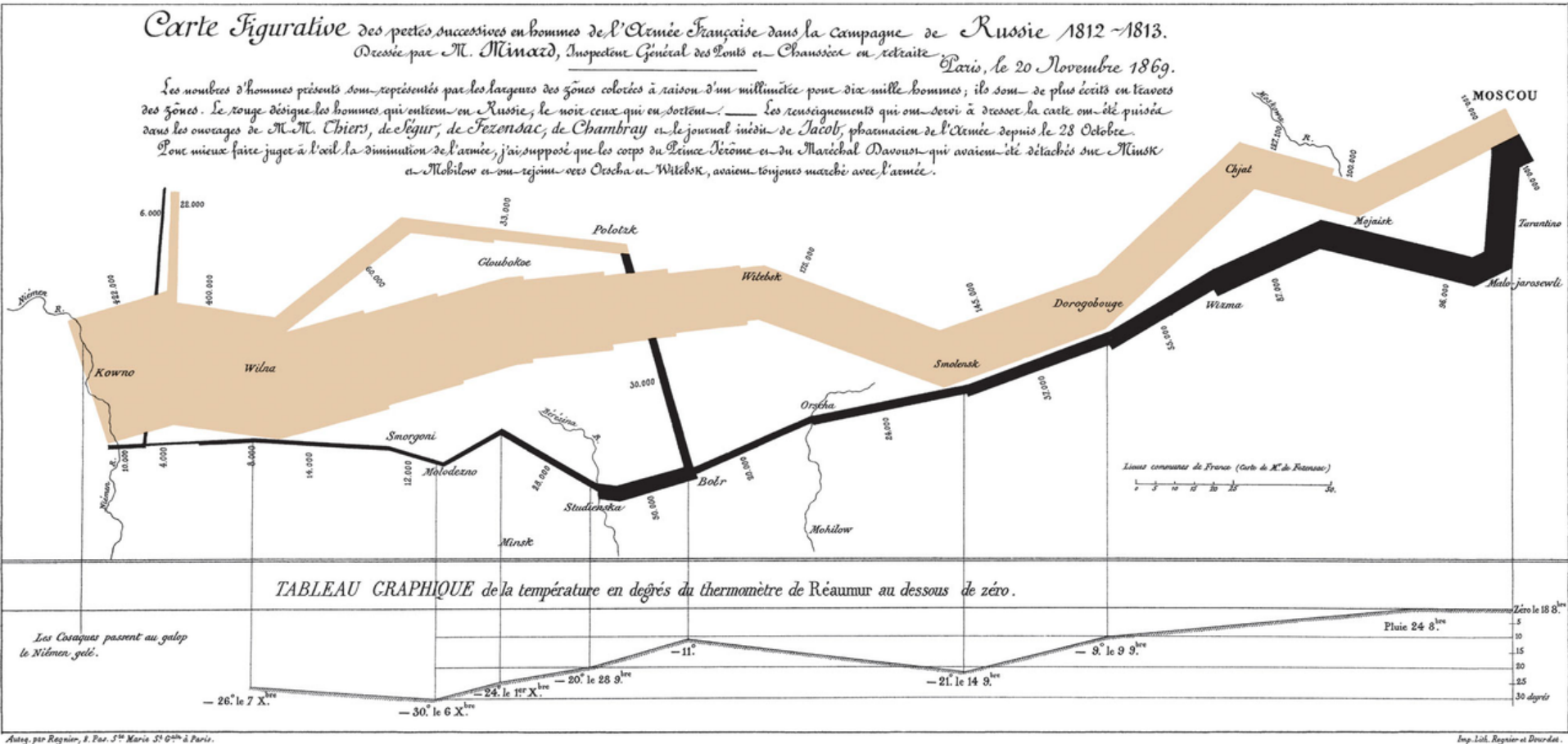
Jun 26 - Parham - Infectious risk modeling

Guiding principles for graphical excellence

- Graphical excellence is the well-designed presentation of interesting data – a matter of *substance, statistics, and design*
- Graphical excellence consists of complex ideas communicated with clarity, precision, and efficiency
- Graphical excellence is that which gives to the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space



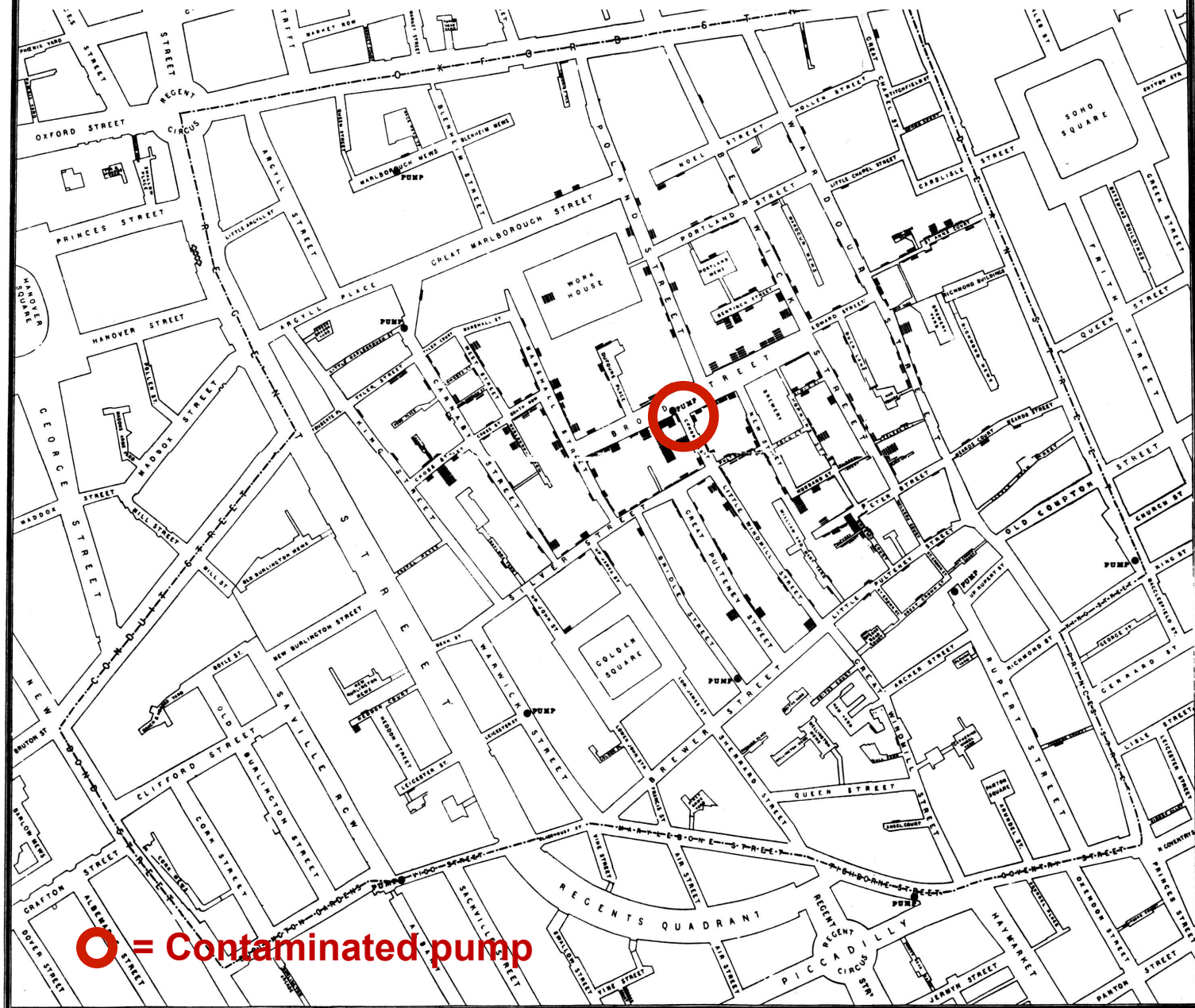
Perhaps the most famous infographic



Charles Minard's 1869 chart showing the number of men in Napoleon's failed 1812 Russian invasion, their movements, as well as the temperature they encountered on the return path

Shows space, time, and magnitude of 2 variables

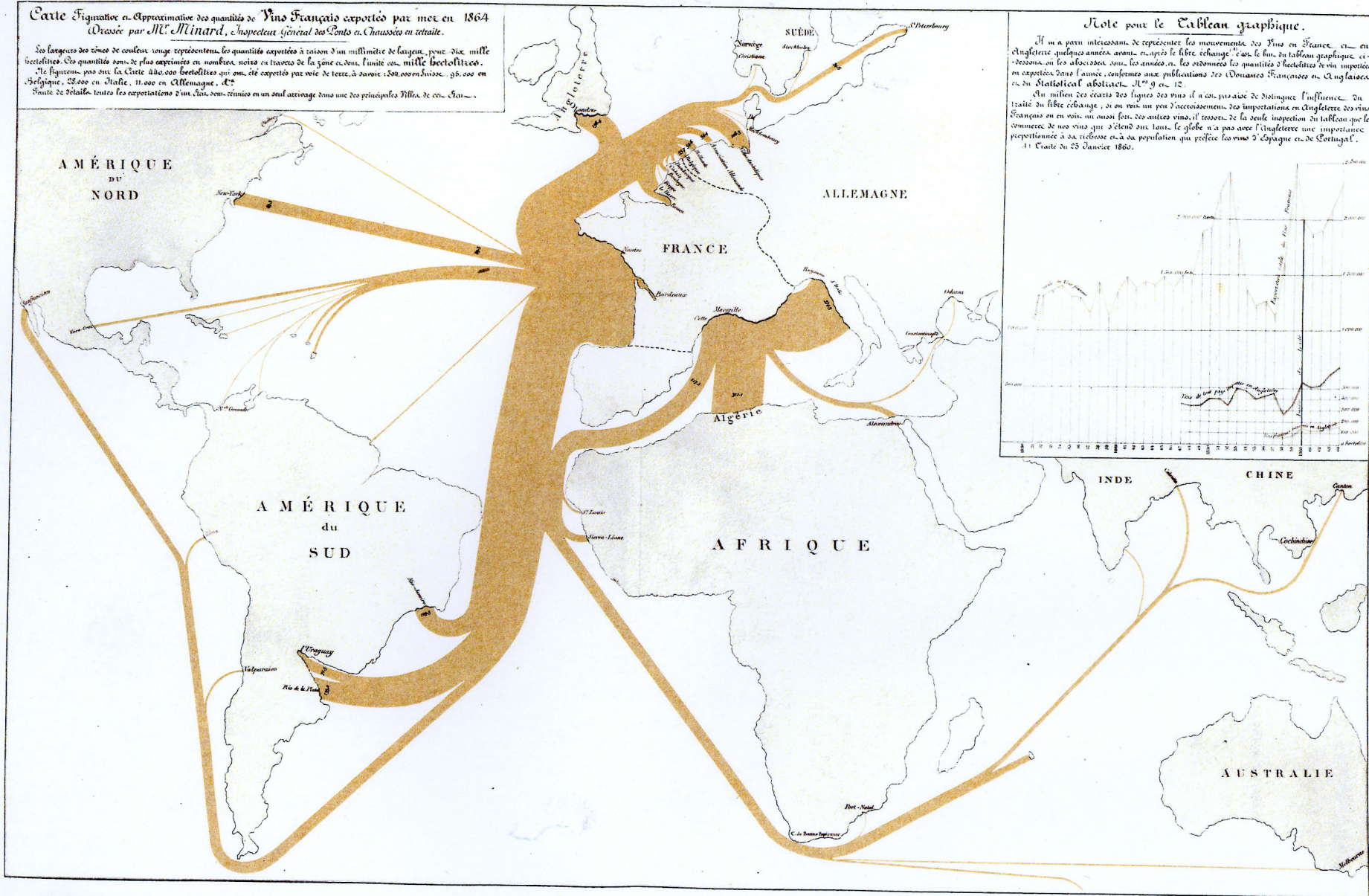
1854 London cholera outbreak, Dr. John Snow



○ = Contaminated pump

Carte Figurative et Approximative des quantités de Vins Français exportés par mer en 1864
 Dressée par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite.

Les largeurs des cônes de couleur orange représentent les quantités exportées à raison d'un millimètre de largeur pour dix mille hectolitres. Les quantités sont de plus exprimées en nombres, noirs et tracées de la zone côtière, l'unité est mille hectolitres.
 Le figurement par cette carte 480,000 hectolitres qui ont été exportés par voie de terre, à savoir : 300,000 en Suisse, 90,000 en Belgique, 20,000 en Italie, 10,000 en Allemagne, etc.
 Faute de détails, toutes les exportations d'un vin sont réunies en un seul atterrissage dans une des principales villes de ces pays.



Note pour le Tableau graphique.

Il m'a paru intéressant de représenter les mouvements des vins en France, et en Angleterre, quelques années avant, et après le libre échange, c'est le but du tableau graphique ci-dessous, on les a comparés, sous les années, et les ordonnées les quantités d'hectolitres de vin importés ou exportés dans l'année, conformément aux publications des Douanes Françaises en Angleterre, et du Statistical abstract, 1869 et 1870.

Au milieu des écarts des lignes des vins il a été possible de distinguer l'influence du traité du libre échange, et on voit un peu d'accroissement des importations en Angleterre des vins Français en en voit un aussi fort des autres vins, il ressort de la seule inspection du tableau que le commerce de nos vins qui s'étend sur tout le globe n'a pas avec l'Angleterre une importance proportionnée à sa richesse en sa population qui préfère les vins d'Espagne et de Portugal.

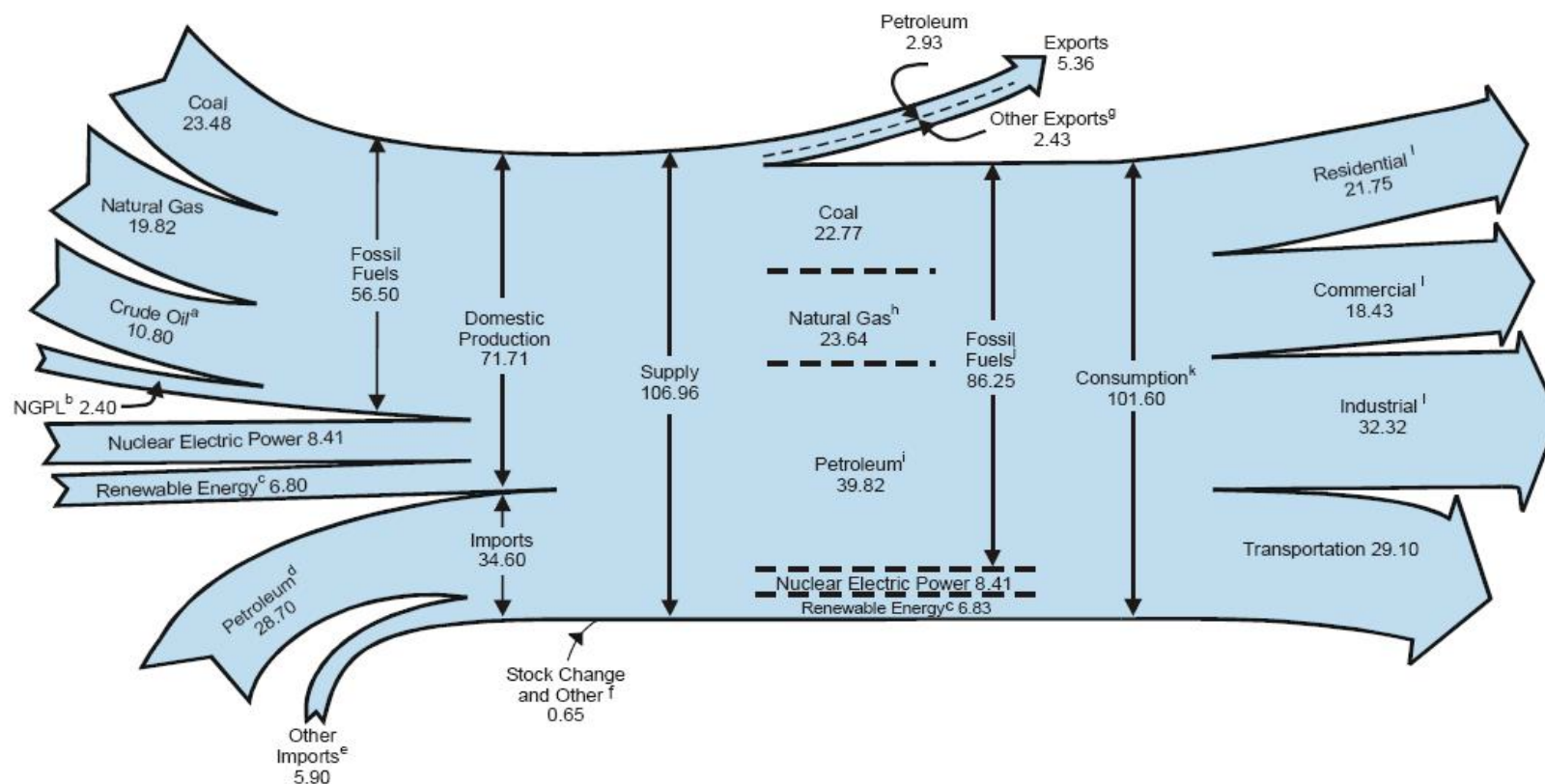
1) Traité du 25 Janvier 1860.

Charles Joseph Minard, *Tableaux Graphiques et Cartes Figuratives de M. Minard, 1845-1869*, a portfolio of his work held by the Bibliothèque de l'École Nationale des Ponts et Chaussées, Paris.

More C. Minard: Map of French exports (space and magnitude)

Diagram 1. Energy Flow, 2007
(Quadrillion Btu)

Flow diagrams are still used today



^a Includes lease condensate.

^b Natural gas plant liquids.

^c Conventional hydroelectric power, biomass, geothermal, solar/photovoltaic, and wind.

^d Crude oil and petroleum products. Includes imports into the Strategic Petroleum Reserve.

^e Natural gas, coal, coal coke, fuel ethanol, and electricity.

^f Adjustments, losses, and unaccounted for.

^g Coal, natural gas, coal coke, and electricity.

^h Natural gas only; excludes supplemental gaseous fuels.

ⁱ Petroleum products, including natural gas plant liquids, and crude oil burned as fuel.

^j Includes 0.03 quadrillion Btu of coal coke net imports.

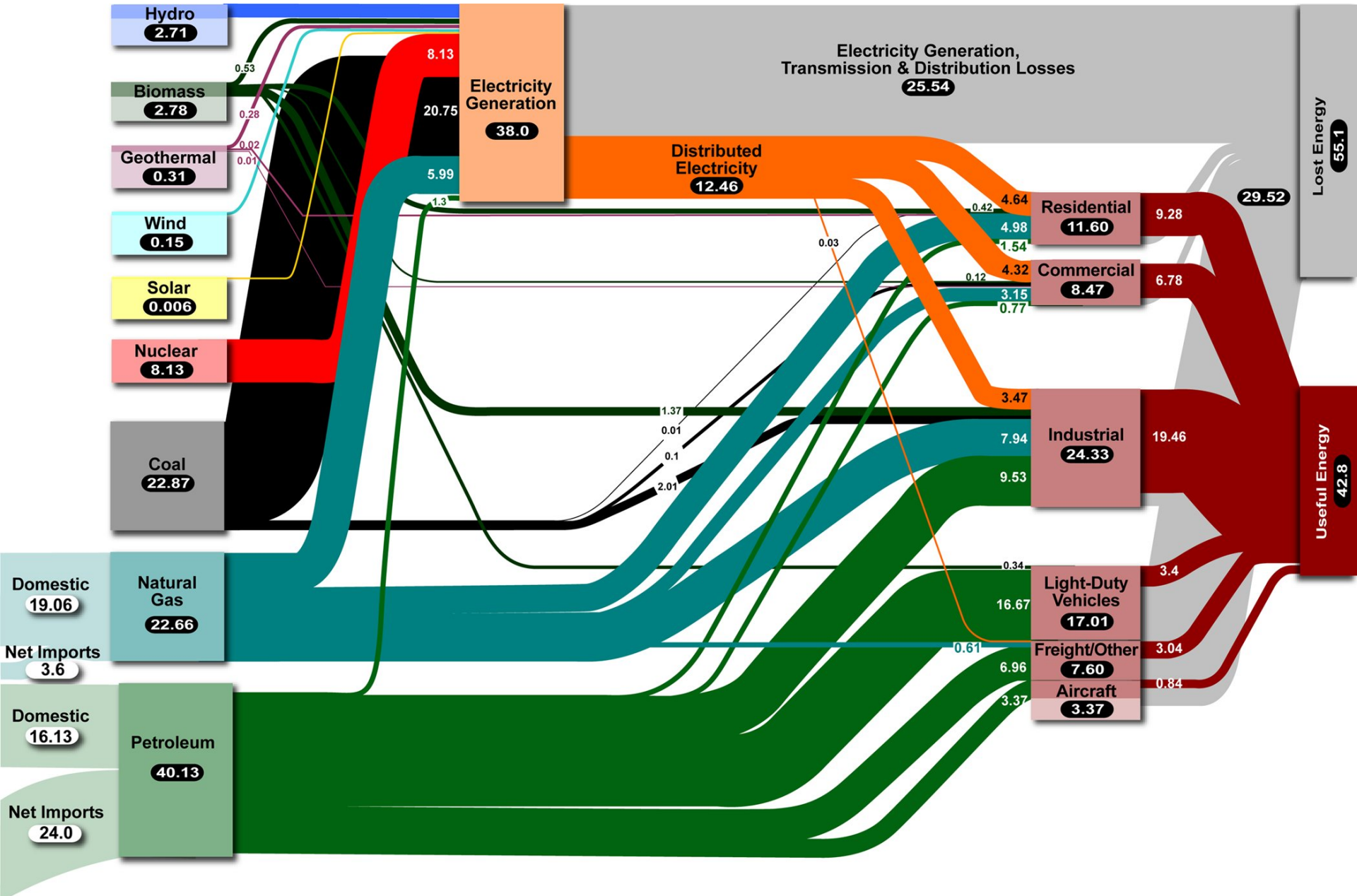
^k Includes 0.11 quadrillion Btu of electricity net imports.

^l Primary consumption, electricity retail sales, and electrical system energy losses, which are allocated to the end-use sectors in proportion to each sector's share of total electricity retail sales. See Note, "Electrical Systems Energy Losses," at end of Section 2.

Notes: • Data are preliminary. • Values are derived from source data prior to rounding for publication. • Totals may not equal sum of components due to independent rounding.

Sources: Tables 1.1, 1.2, 1.3, 1.4, and 2.1a.

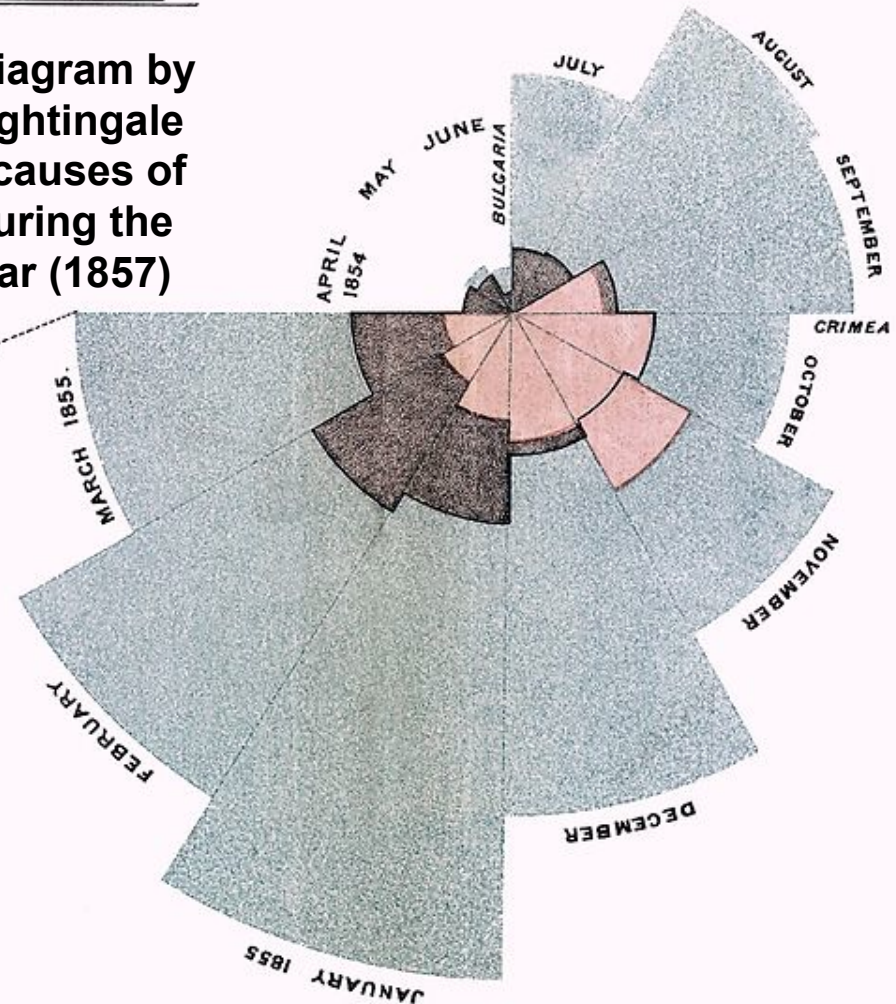
Flow diagrams are still used today



Florence Nightingale (1858)

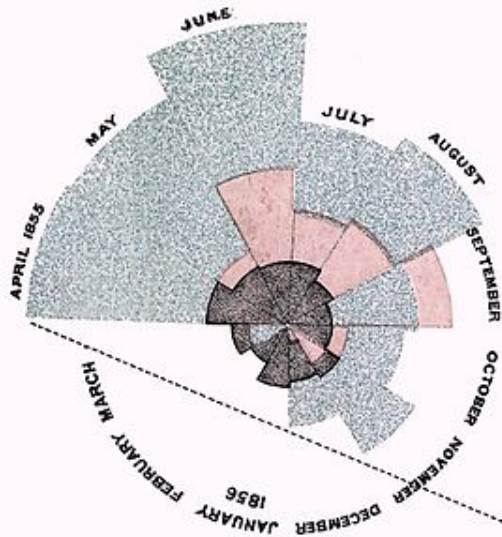
DIAGRAM OF THE CAUSES OF MORTALITY IN THE ARMY IN THE EAST.

1.
APRIL 1854 TO MARCH 1855.



2.
APRIL 1855 TO MARCH 1856.

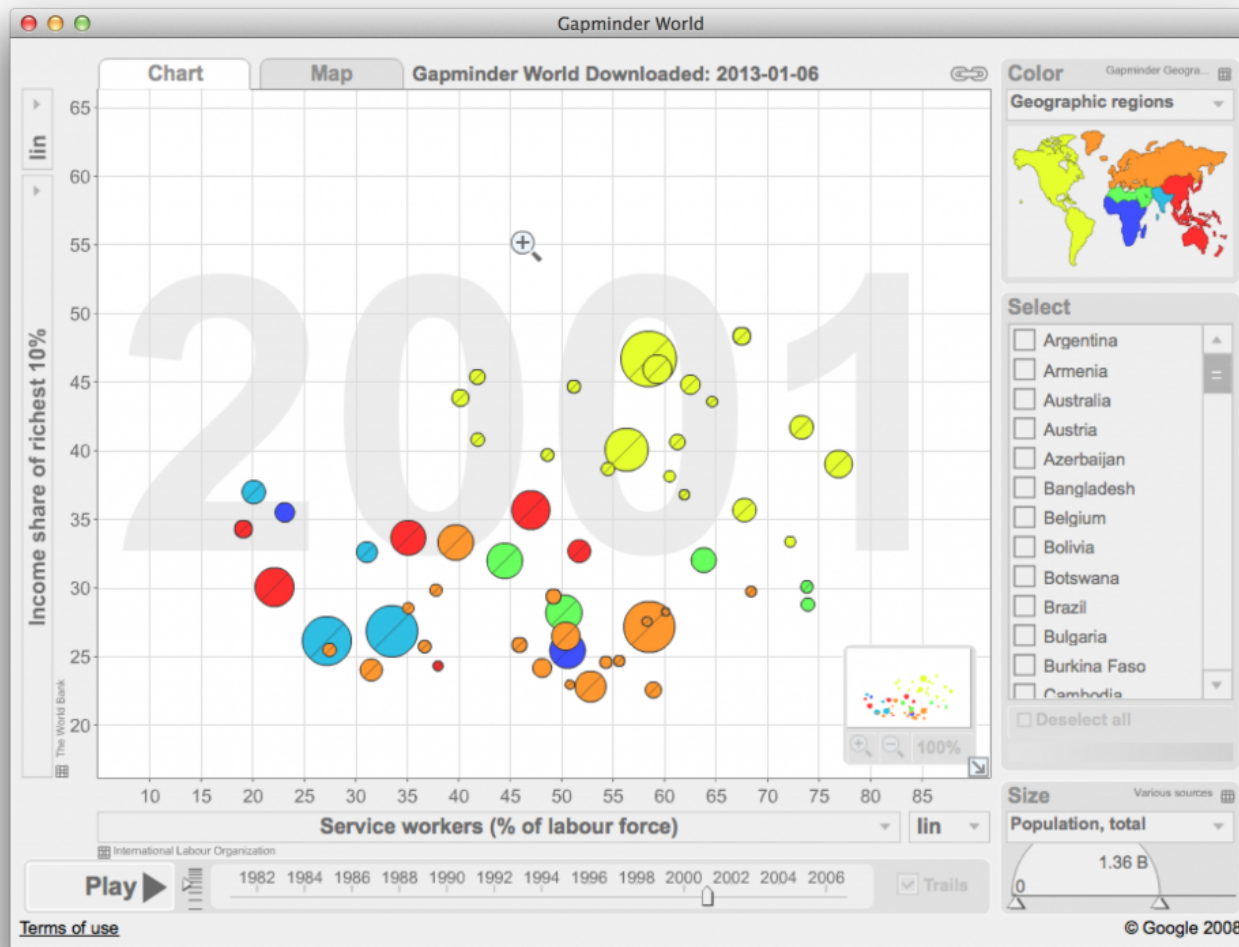
Polar area diagram by
Florence Nightingale
illustrating causes of
mortality during the
Crimean War (1857)

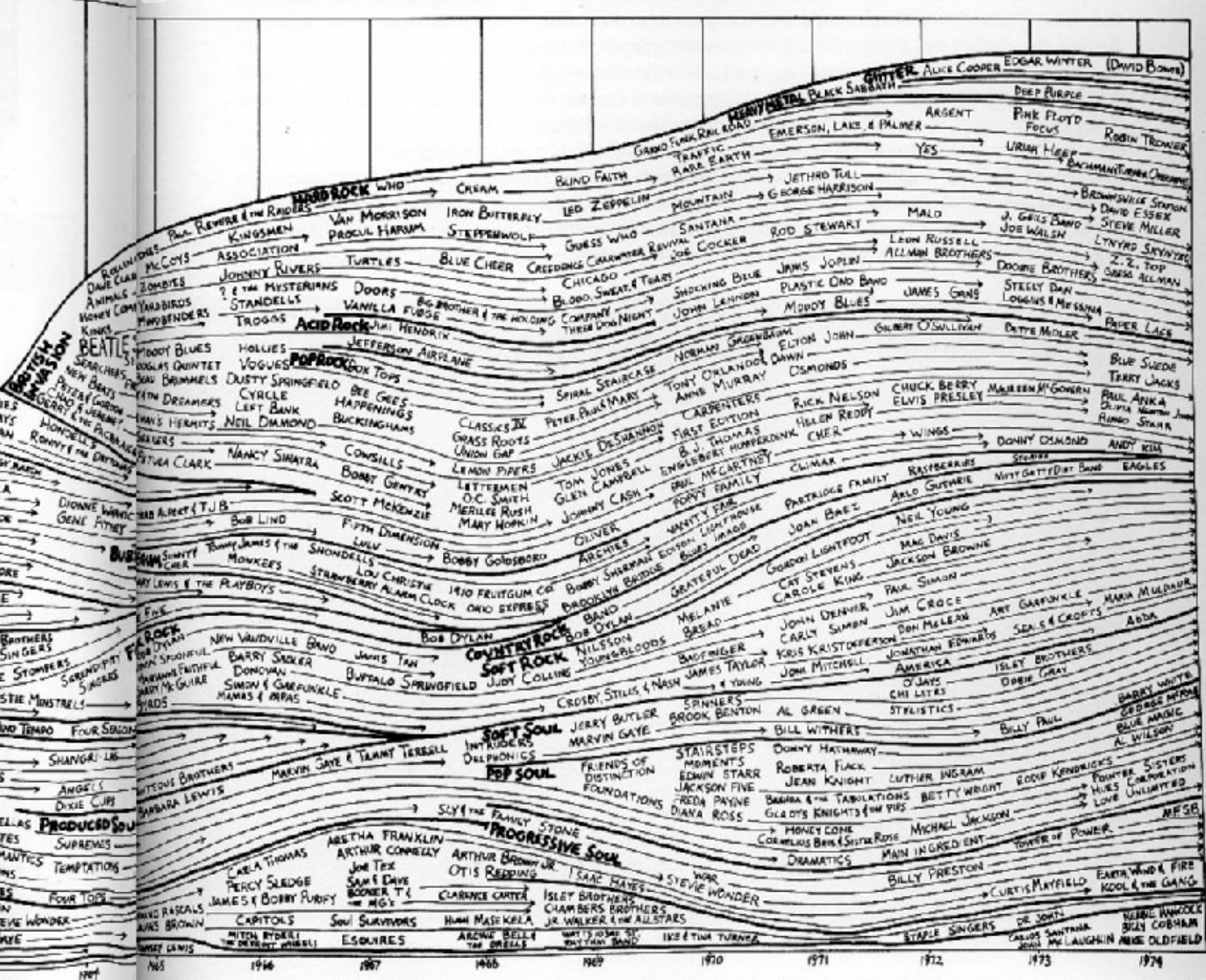


The Areas of the blue, red, & black wedges are each measured from the centre as the common vertex.
The blue wedges measured from the centre of the circle represent area for area the deaths from Preventable or Mitigable Zymotic diseases; the red wedges measured from the centre the deaths from wounds; & the black wedges measured from the centre the deaths from all other causes.
The black line across the red triangle in Nov. 1854 marks the boundary of the deaths from all other causes during the month.
In October 1854, & April 1855, the black area coincides with the red; in January & February 1856, the blue coincides with the black.
The entire areas may be compared by following the blue, the red & the black lines enclosing them.

Modern day: Gapminder and *The Joy of Stats*

GAPMINDER for a fact-based world view



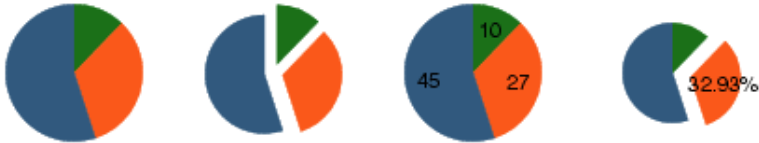


Deciding how to show information for engineering

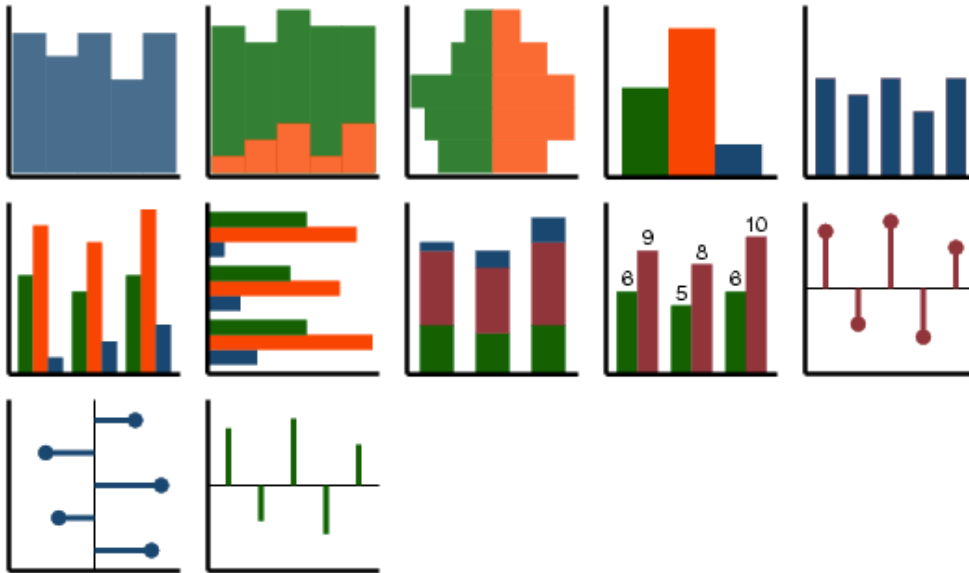
- In what context would you like to show your data?
 - Temporal (time) variations?
 - Spatial (space) variations?
 - Statistical distributions?
 - Summary statistics?
 - Means, medians, max, min, percentiles?
- Are tables an appropriate option?
- What plot types are available?

Some available plot types

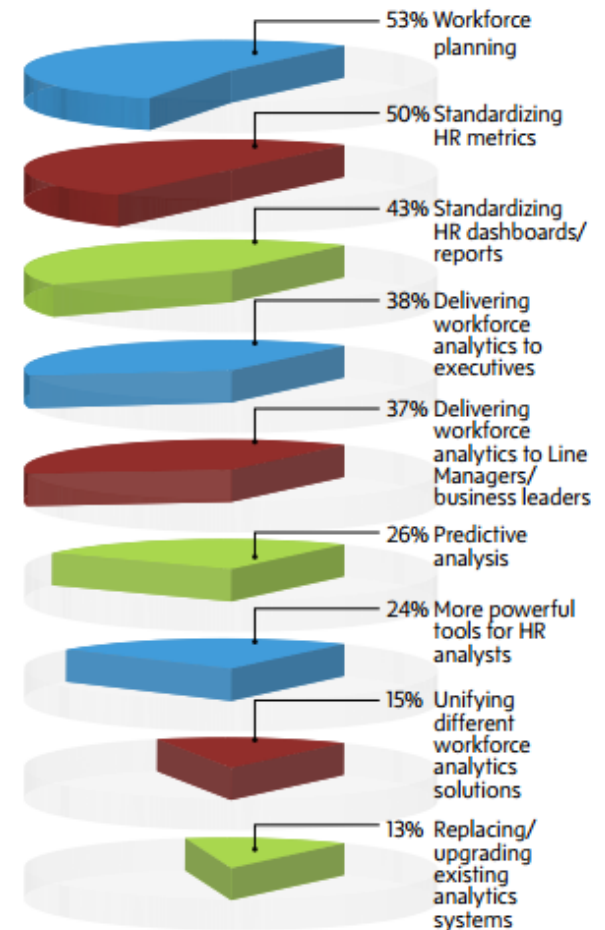
Pie charts



Bar and dropped-line charts

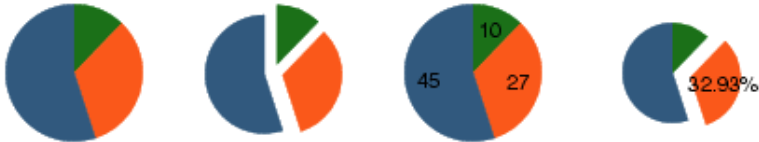


All plot types can be used ...and abused!

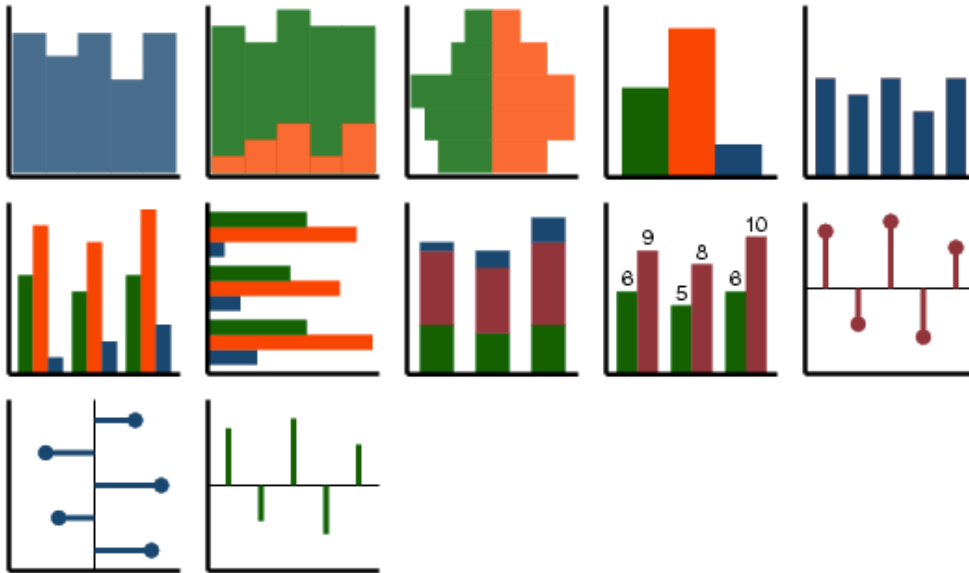


Some available plot types

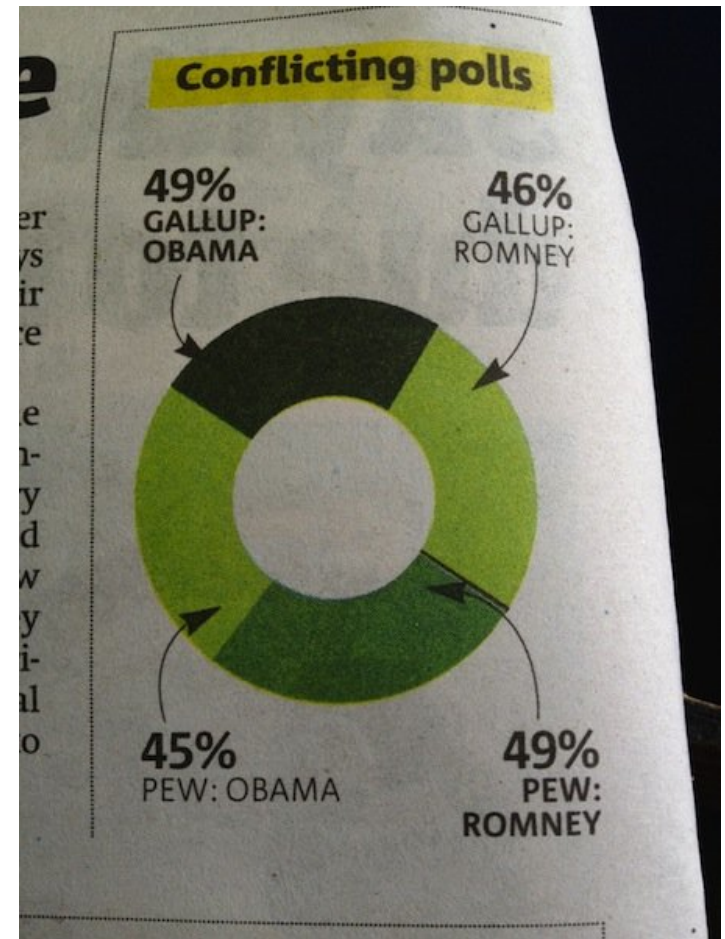
Pie charts



Bar and dropped-line charts

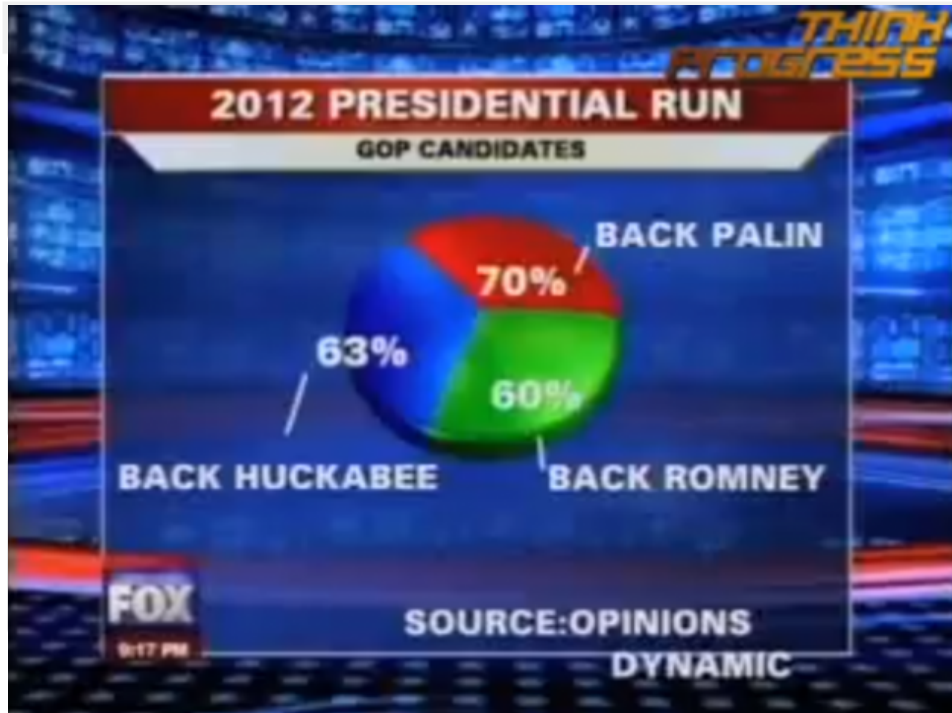
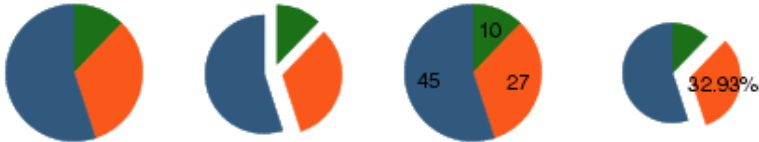


All plot types can be used
...and abused!

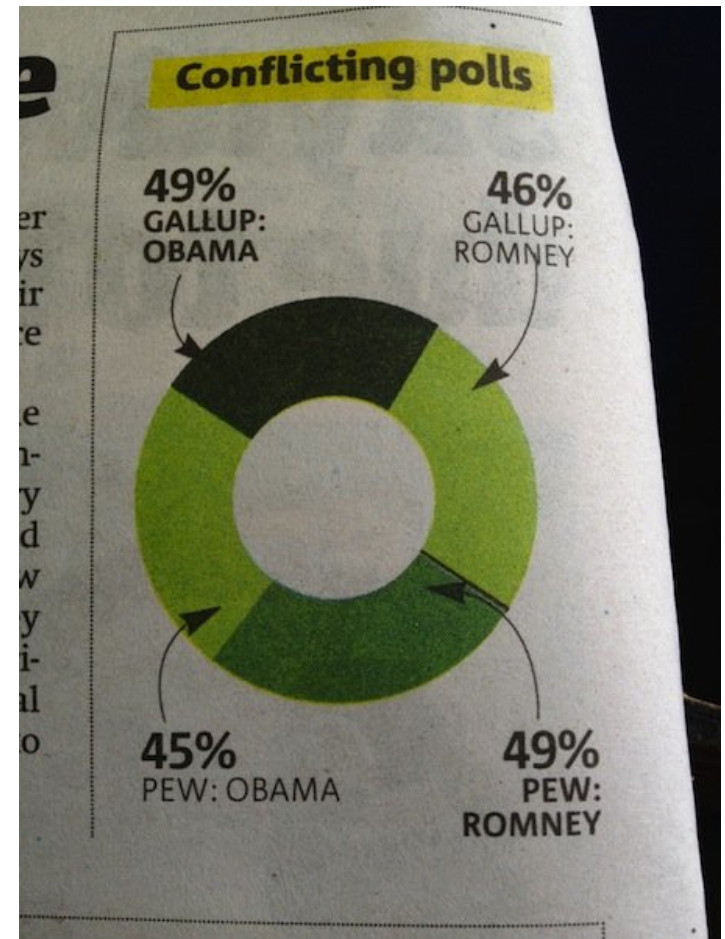


Some available plot types

Pie charts



All plot types can be used
...and abused!



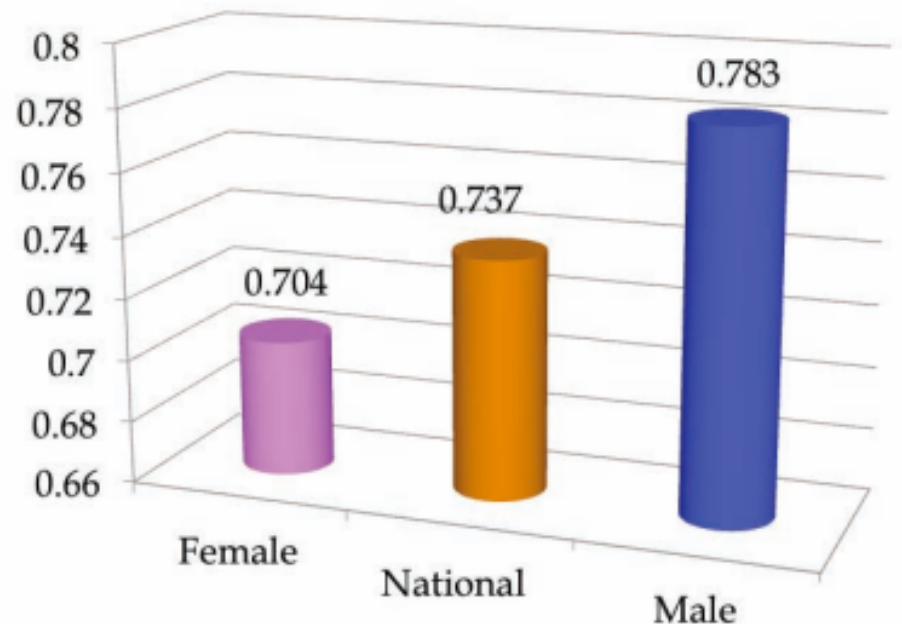
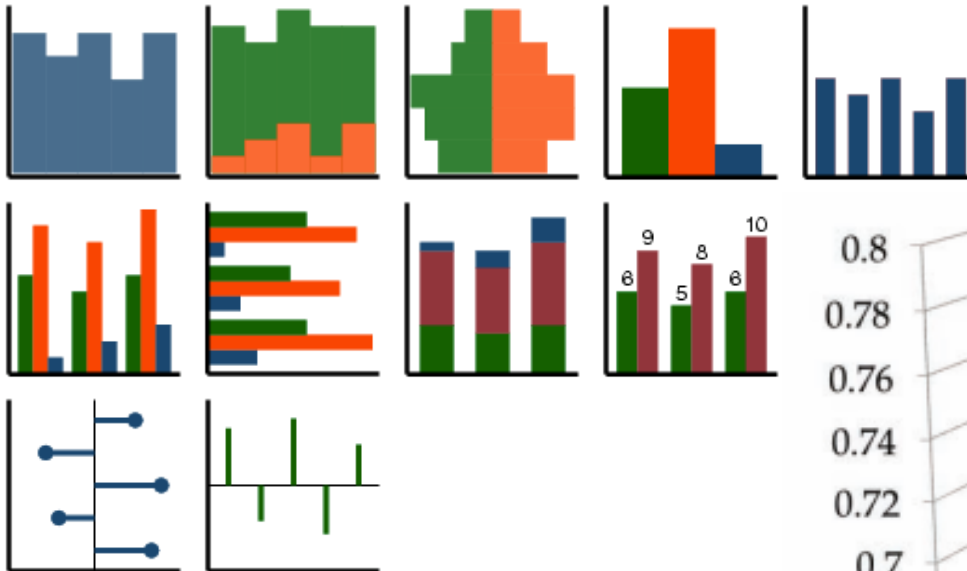
Some available plot types

Pie charts



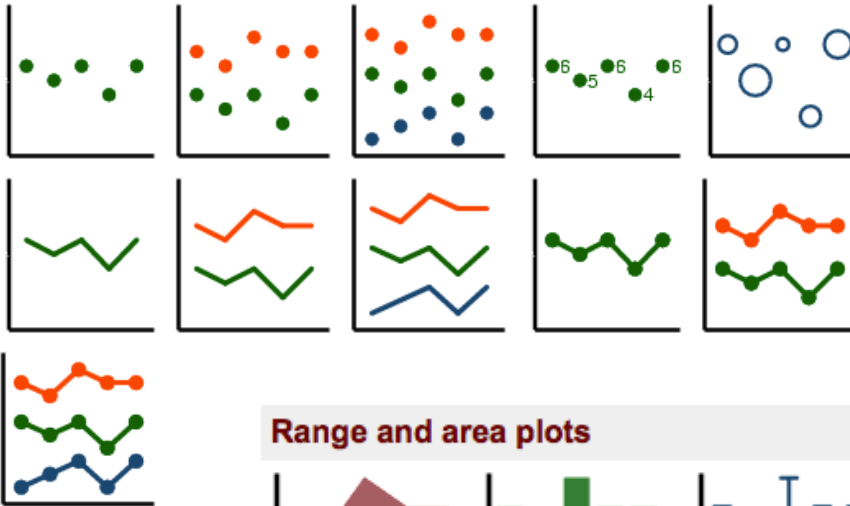
All plot types can be used
...and abused!

Bar and dropped-line charts

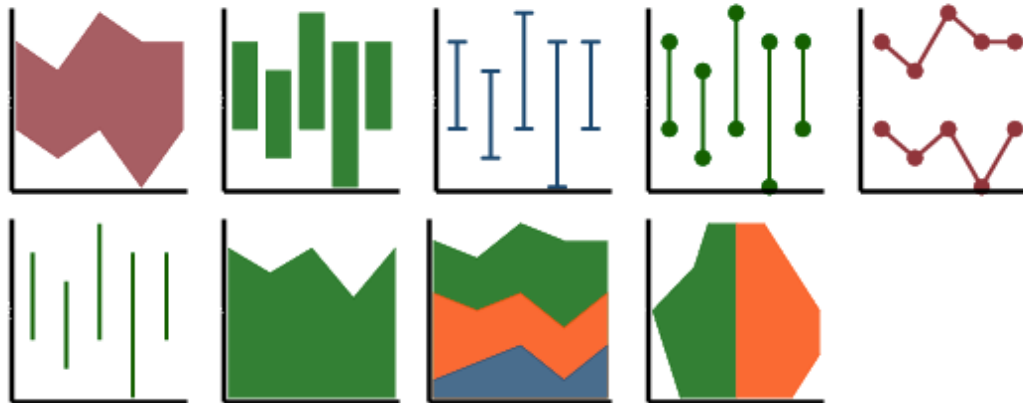


Some available plot types

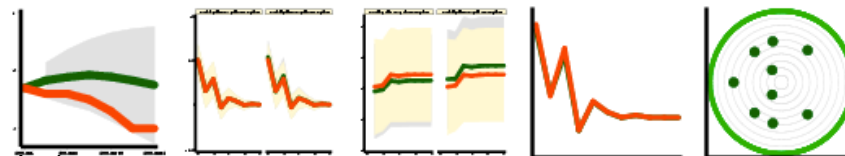
Scatter and line plots



Range and area plots

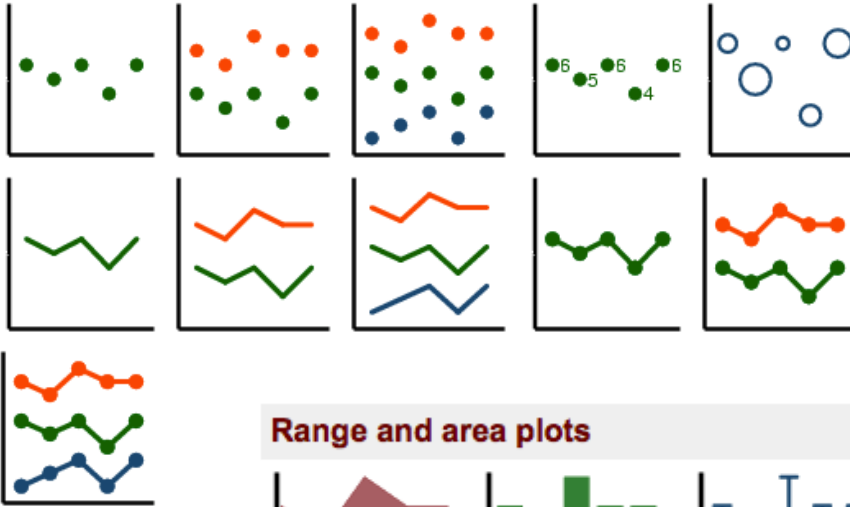


VAR and VEC

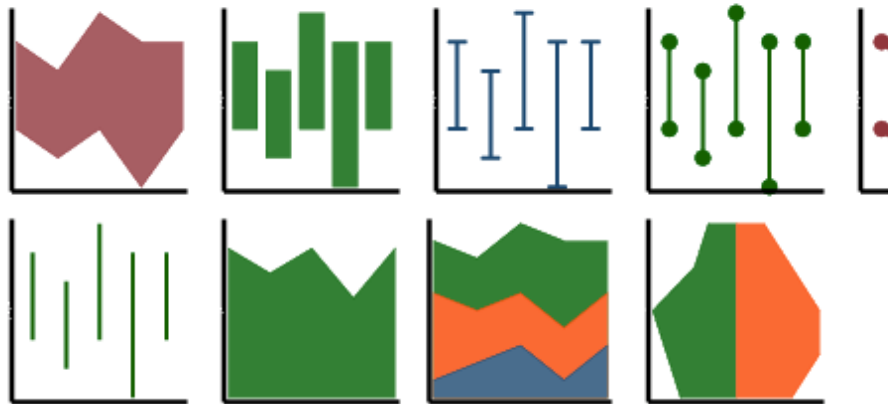


Used and abused

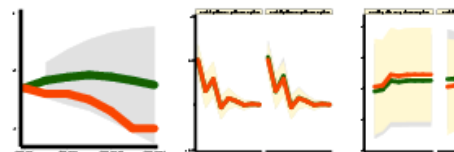
Scatter and line plots



Range and area plots



VAR and VEC



A

BINNED FREQUENCY DATA - D4S139
CHINESE, JAPANESE, KOREAN & VIETNAMESE

0.25
0.2
0.15
0.1
0.05
0

2 4 6 8 10 12 14 16 18 20 22 24 26 28 30

CHI JAP KOR VIE

B

BINNED FREQUENCY DATA - D10S28
CHINESE, JAPANESE, KOREAN, VIETNAMESE

0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0

2 4 6 8 10 12 14 16 18 20 22 24 26 28 30

CHI JAP KOR VIE

FIG. 4. Fixed bin distribution (histogram) for two loci and four Asian subpopulations (used with permission from John Hartmann): the boundaries of the 30 bins (vertical axis) are determined by the FBI; these bins are not of equal length. Sample sizes (numbers of individuals) for Chinese, Japanese, Korean and Vietnamese are 103, 125, 93 and 215 for D4S139 and 120, 137, 100 and 193 for D10S28. The horizontal axis is the bin number; bins are not of equal length.

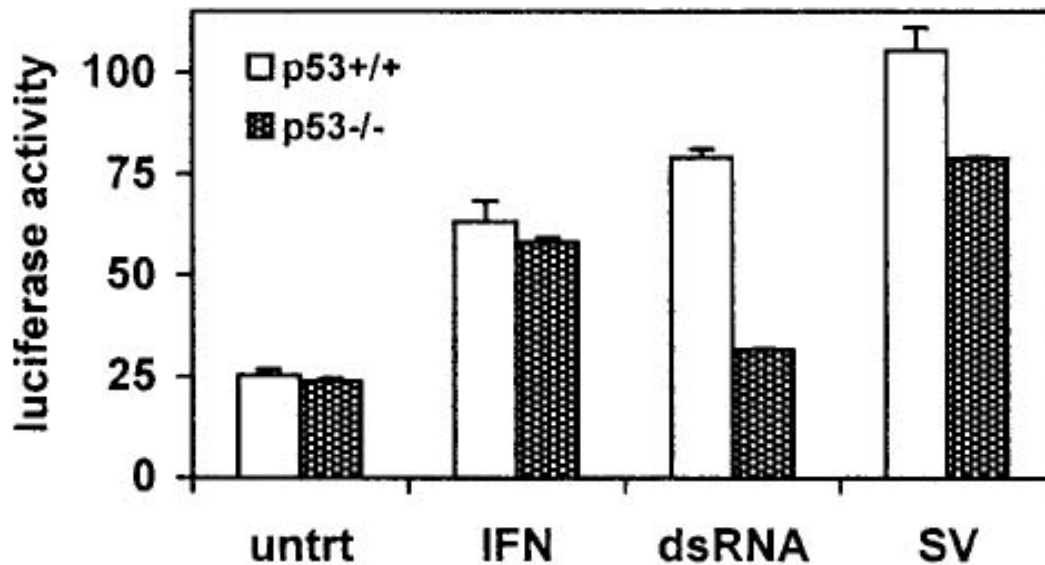


FIG. 4. ISG15 promoter activity mimics endogenous ISG15 mRNA regulation by p53, dsRNA, and virus. Cells (6×10^5 HCT 116) were seeded in 32-mm plates and allowed to attach overnight. Cells were transfected with 500 ng of pGL3/ISG15-Luc, 50 ng of pRL null (Promega), and 450 ng of pcDNA3 for carrier DNA by using Lipofectamine Plus (Life Technologies) following the manufacturer's instructions. Twenty-four hours posttransfection, the medium was aspirated and replaced with medium containing either 1,000 U of IFN- α /ml, 50 μ g of dsRNA/ml, or Sendai virus (multiplicity of infection, 10). Cells were incubated for 12 h and then lysed, and luciferase assays were performed. Luciferase activity was assessed on 20 μ l of each lysate as directed by the supplier (Dual Luciferase Kit, Promega) using a TD 20/20 luminometer (Turner Designs). Luciferase activity is presented as the ratio of firefly activity to renilla activity to control for differences in transfection efficiency. Each data point is the mean of triplicate samples \pm the standard error; the data presented are representative of four independent experiments.

t types Used and abused

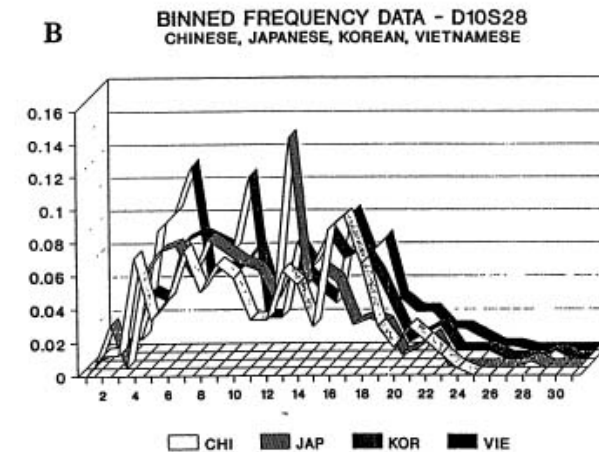
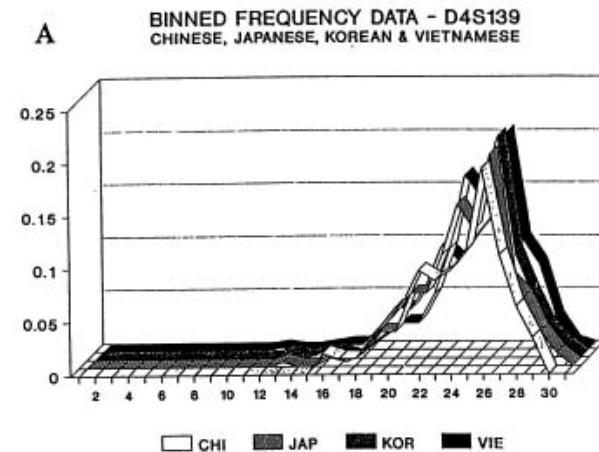
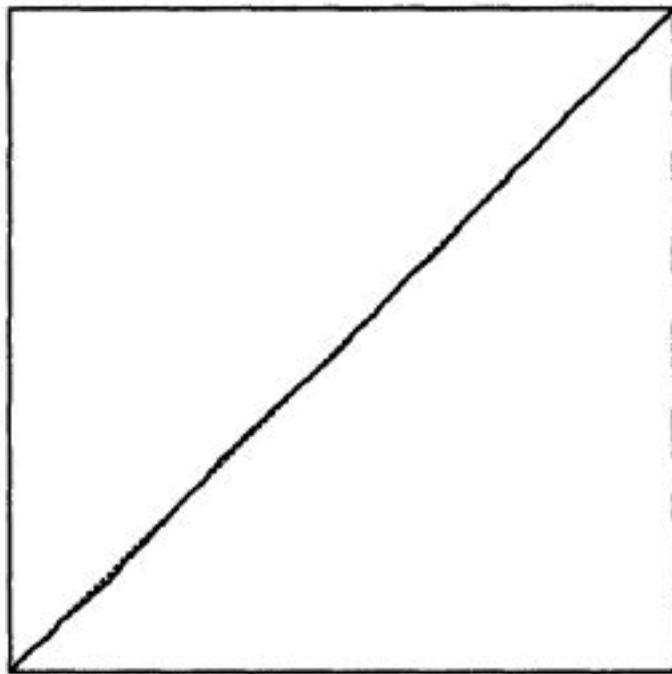
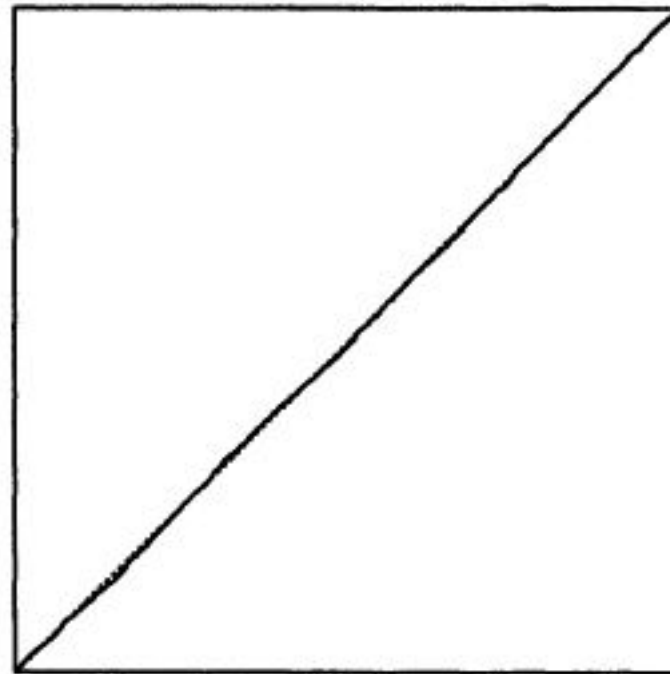


FIG. 4. Fixed bin distribution (histogram) for two loci and four Asian subpopulations (used with permission from John Hartmann): the boundaries of the 30 bins (vertical axis) are determined by the FBI; these bins are not of equal length. Sample sizes (numbers of individuals) for Chinese, Japanese, Korean and Vietnamese are 103, 125, 93 and 215 for D4S139 and 120, 137, 100 and 193 for D10S28. The horizontal axis is the bin number; bins are not of equal length.

Some available plot types Used and abused



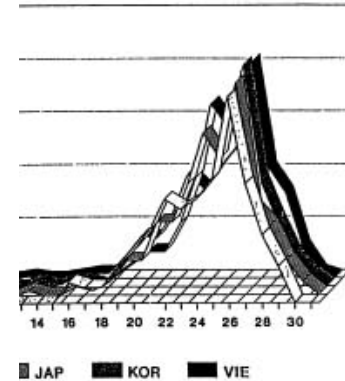
(a)



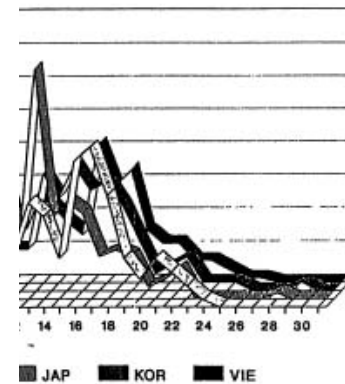
(b)

Figure 1. SRQ Plots of T_i/T_n (Vertical Axes) Against i/n (Horizontal Axes) for the Gibbs Sampler (a) and an Alternating Gibbs/Independence Sampler (b) for the Pump Failure Data Based on Runs of Length 5,000. Lines through the origin with unit slope are shown dashed; axis ranges are from 0 to 1 for all axes.

FREQUENCY DATA - D4S139
JAPANESE, KOREAN & VIETNAMESE



FREQUENCY DATA - D10S28
JAPANESE, KOREAN, VIETNAMESE

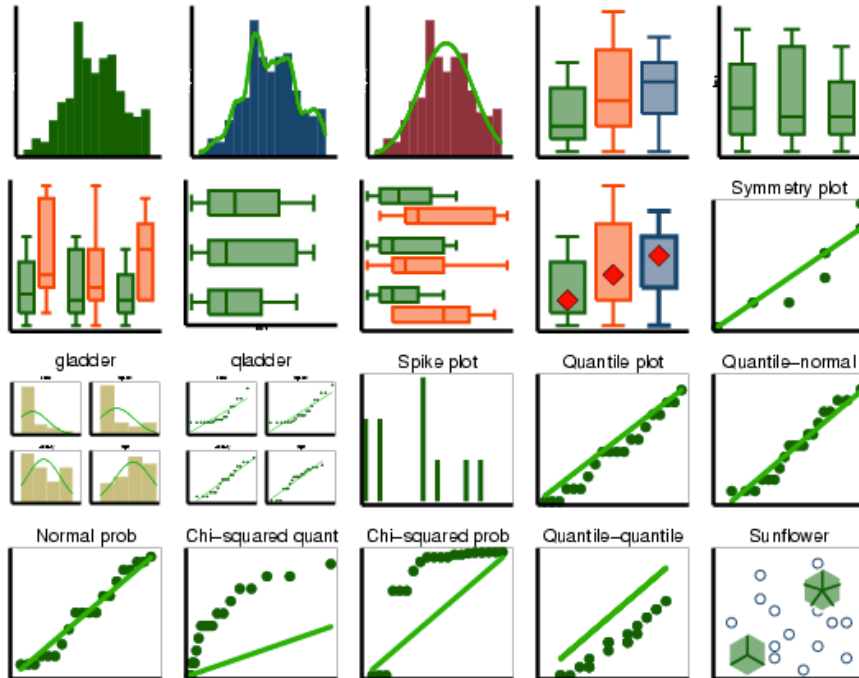


tion (histogram) for two loci and four d with permission from John Hart- e 30 bins (vertical axis) are determined t of equal length. Sample sizes (num- ese, Japanese, Korean and Vietnamese r D4S139 and 120, 137, 100 and 193

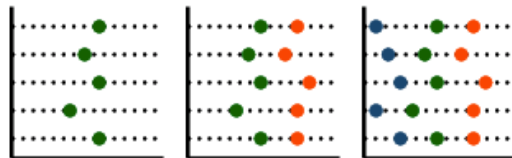
for D10S28. The horizontal axis is the bin number; bins are not of equal length.

Some available plot types

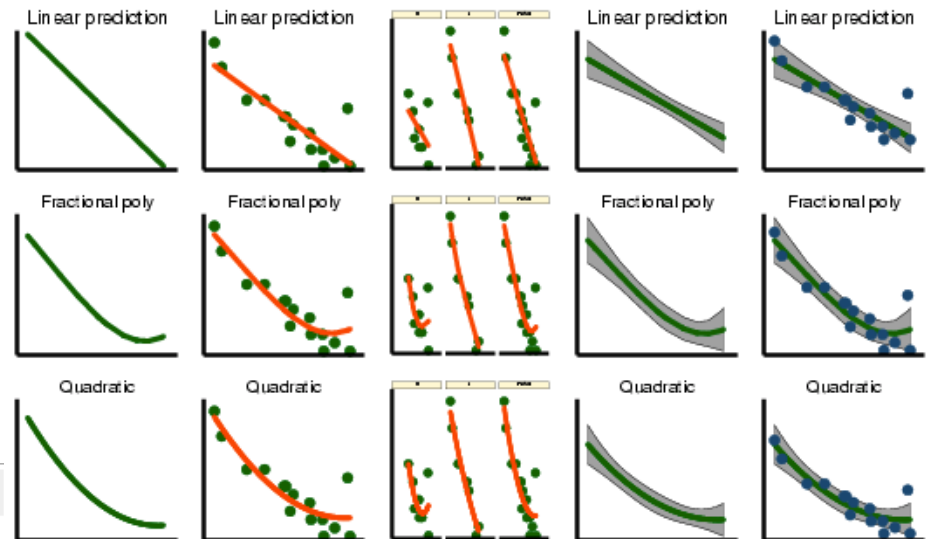
Distribution plots



Dot charts

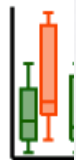
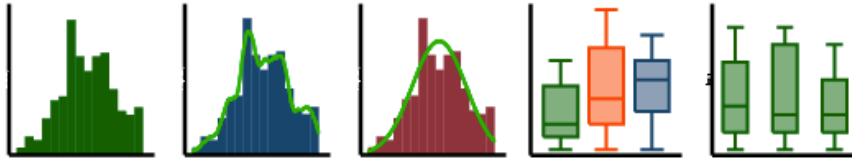


Regression fit plots



Some available plot types

Distribution plots



Dot plots



rediction



val poly



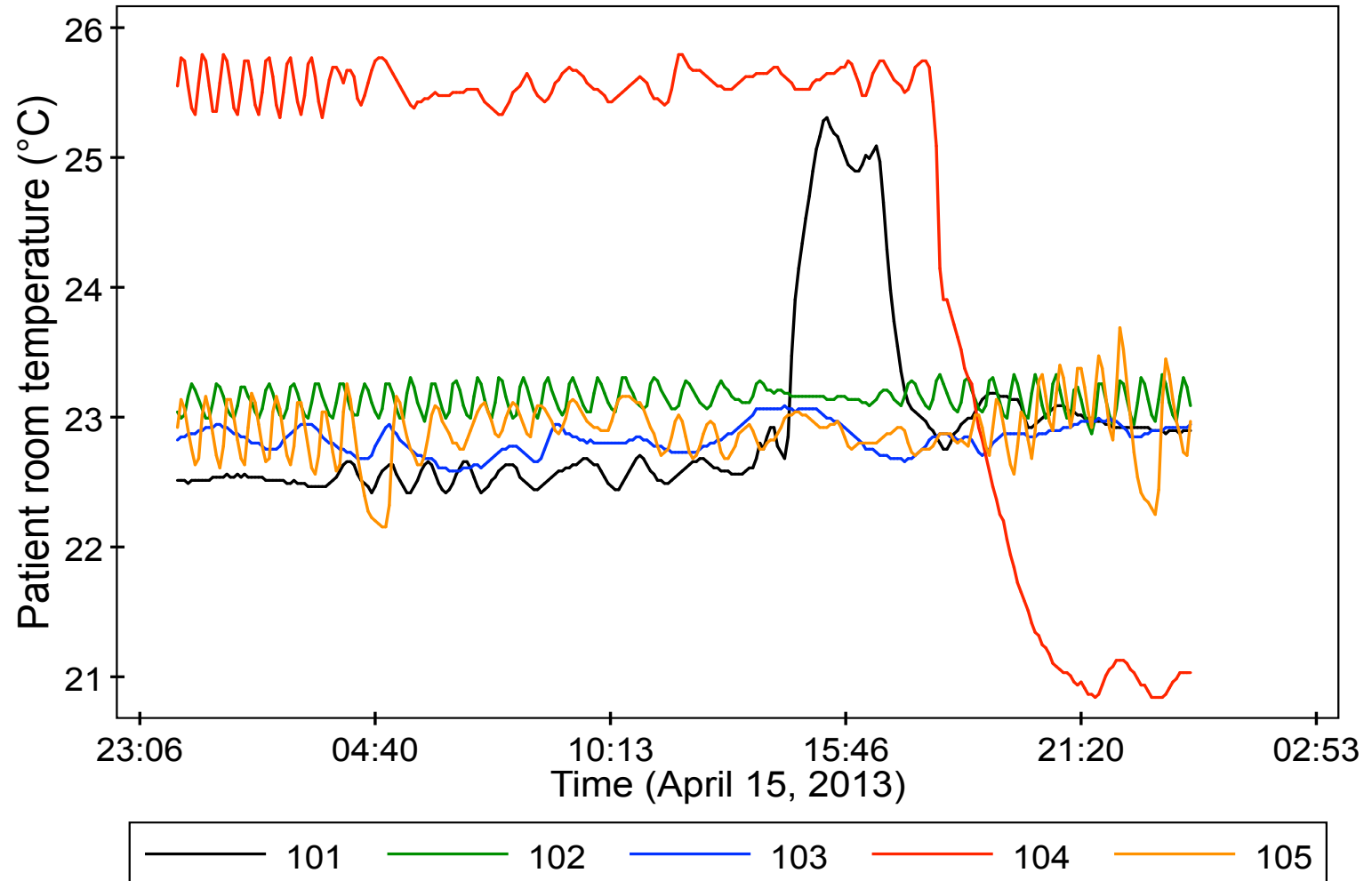
tratic



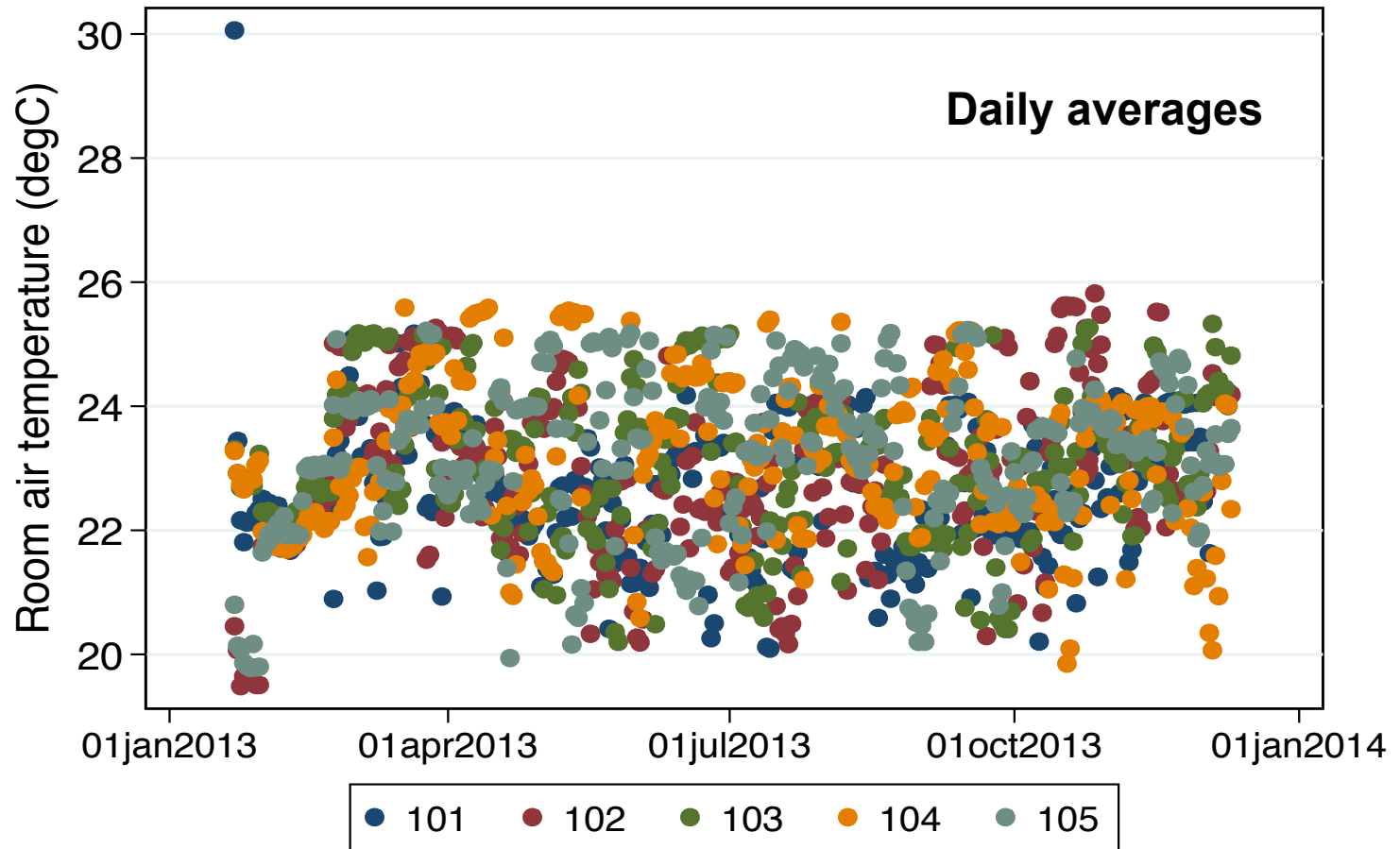
So you want to make a graph...?

- One year of temperature data in the Hospital Microbiome Project
- 5-minute intervals
- Averaged over:
 - Hourly
 - Daily
 - Weekly
- Examples in Stata:
 - Line, scatter, caps, area, dot plots, scatterplot matrices, heat map

5-minute air temperatures

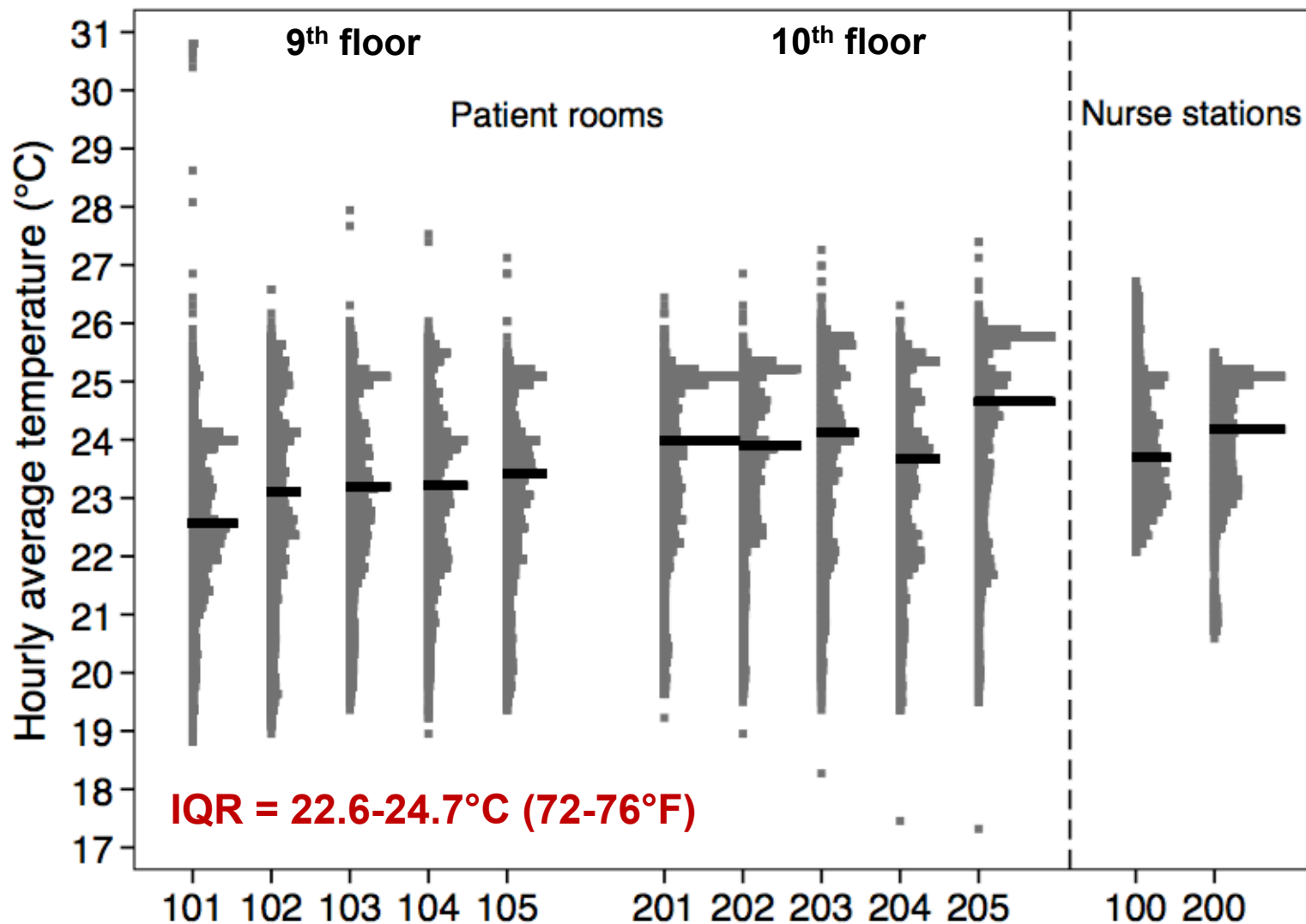


Daily average air temperatures

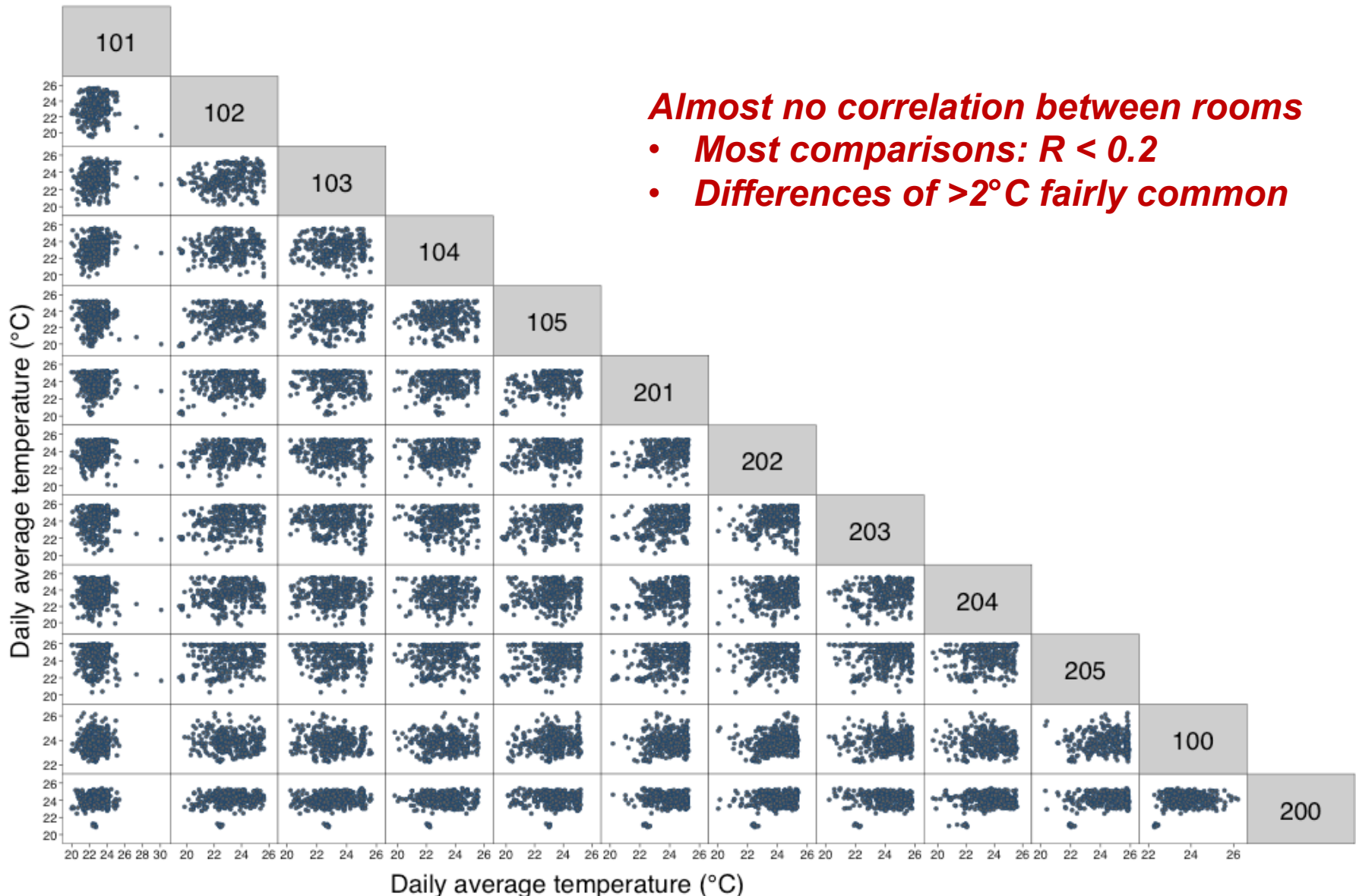


Hourly mean air temperatures: Distributions

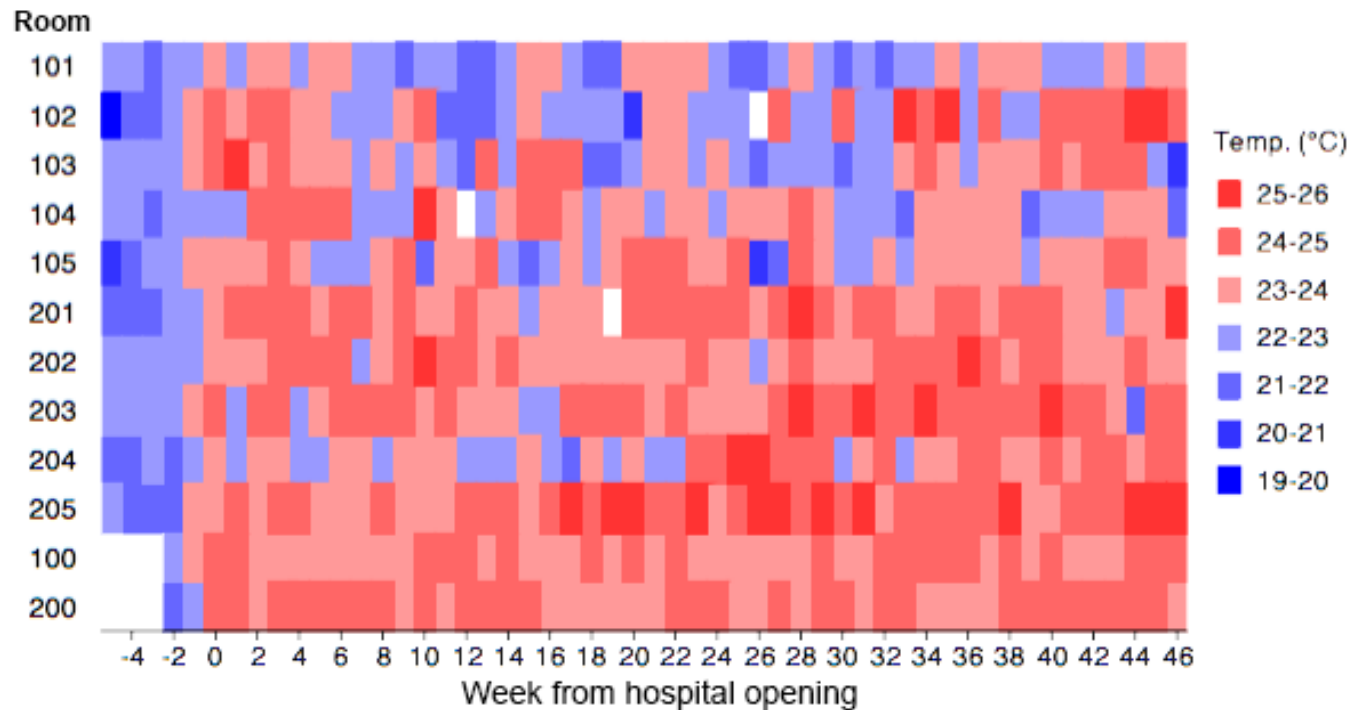
*Similar distributions across patient rooms and nurse stations | Range 19-27°C
66-80°F*



Daily mean air temperatures: Scatterplot matrices

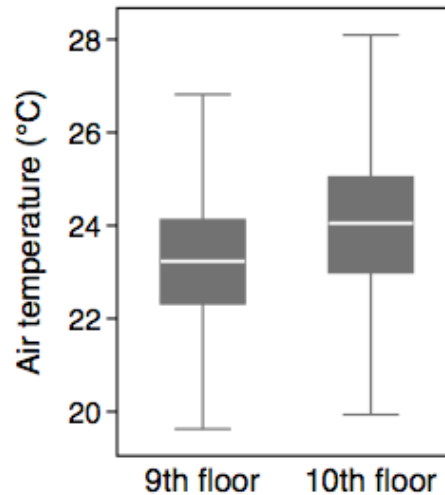


Weekly mean temperatures: Heat map

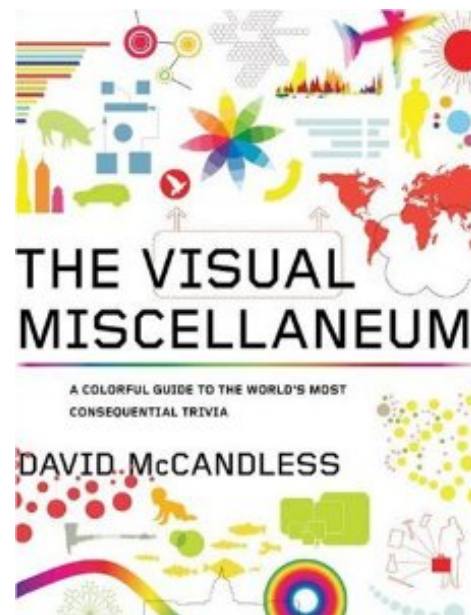
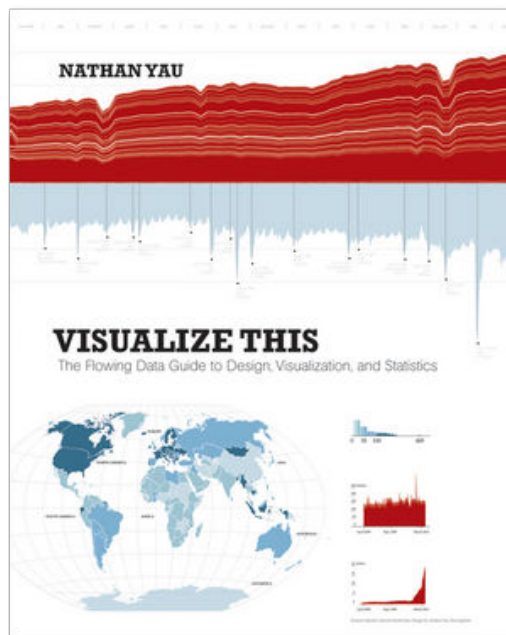
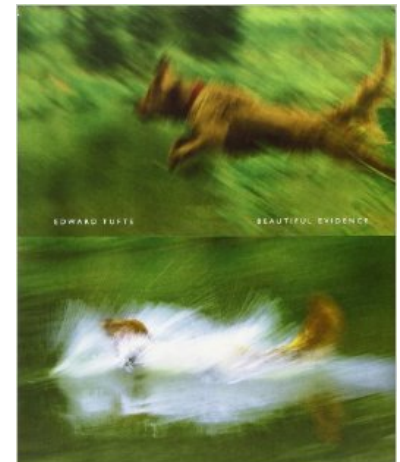
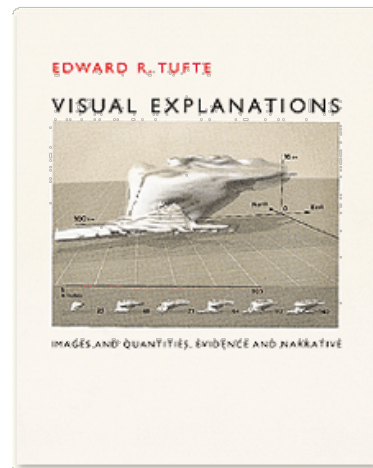
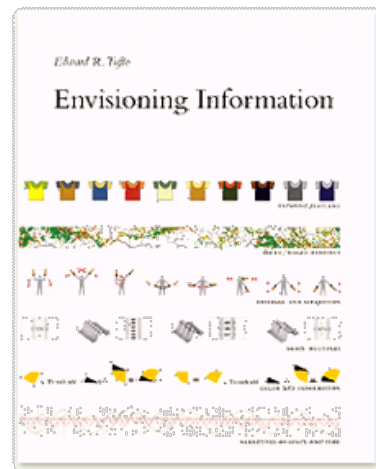
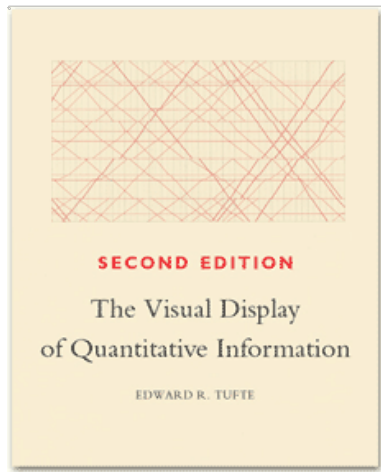


Variations between floors: Box plots

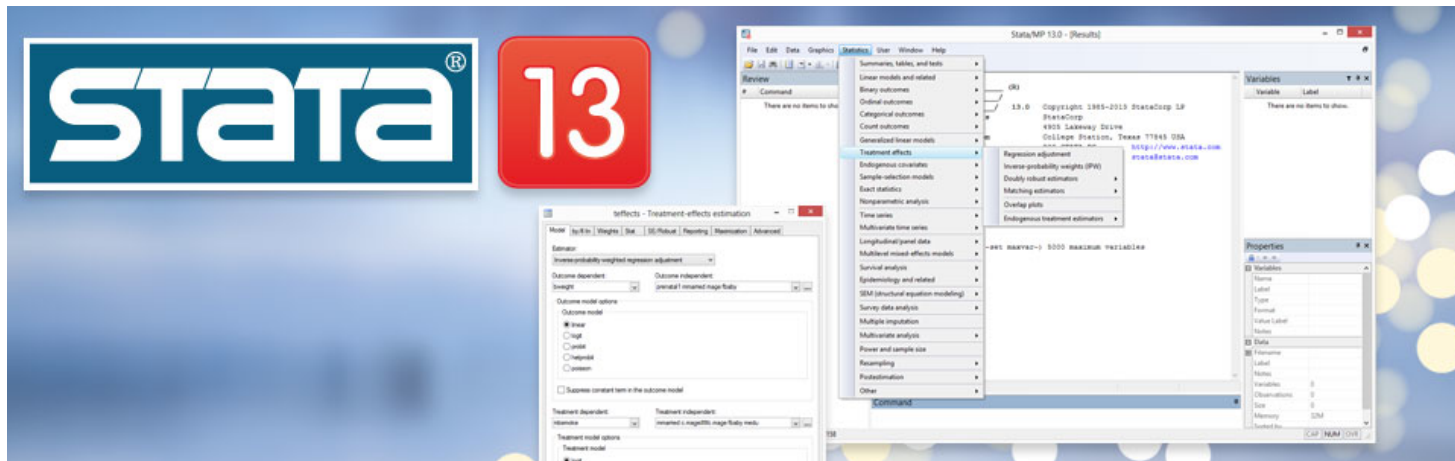
Median Air temperatures ~1°C higher on 10th floor



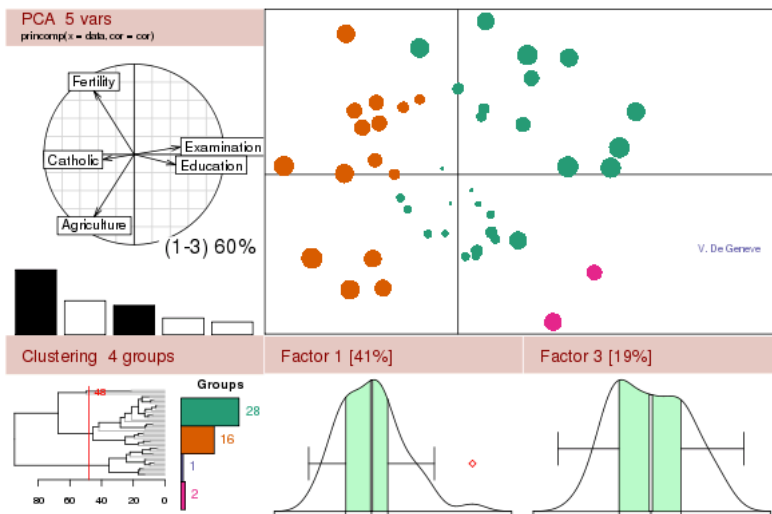
Data visualization resources



Data visualization tools



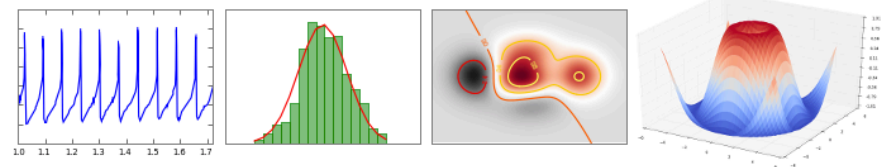
The R Project for Statistical Computing



[home](#) | [examples](#) | [gallery](#) | [pyplot](#) | [docs](#) »

Introduction

matplotlib is a python 2D plotting library which produces publication quality figures in a variety of hardcopy formats and interactive environments across platforms. matplotlib can be used in python scripts, the python and [ipython](#) shell (ala MATLAB[®] or Mathematica[®]), web application servers, and six graphical user interface toolkits.



Others: Excel, SPSS, Matlab, and many more