

Topics in Performance Evaluation

Dror Feitelson
Hebrew University

Lecture 2 – Graphs

“Few of us escape being indoctrinated with these notions:

- (1) Numerical calculations are exact, but graphs are rough;
- (2) For any particular kind of statistical data there is just one set of calculations constituting a correct statistical analysis;
- (3) Performing intricate calculations is virtuous, whereas actually looking at the data is cheating.”

F. J. Anscombe

The American Statistician **27(1)** Feb 1973

Anscombe's example of 4 datasets:

10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.1	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.1	4	5.39	19	12.5
12	10.84	12	9.13	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89

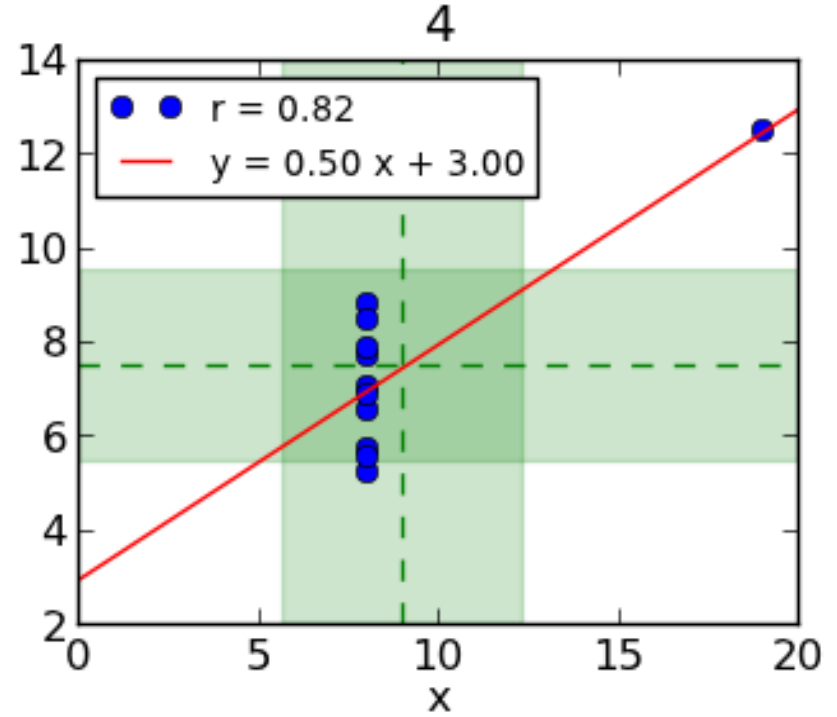
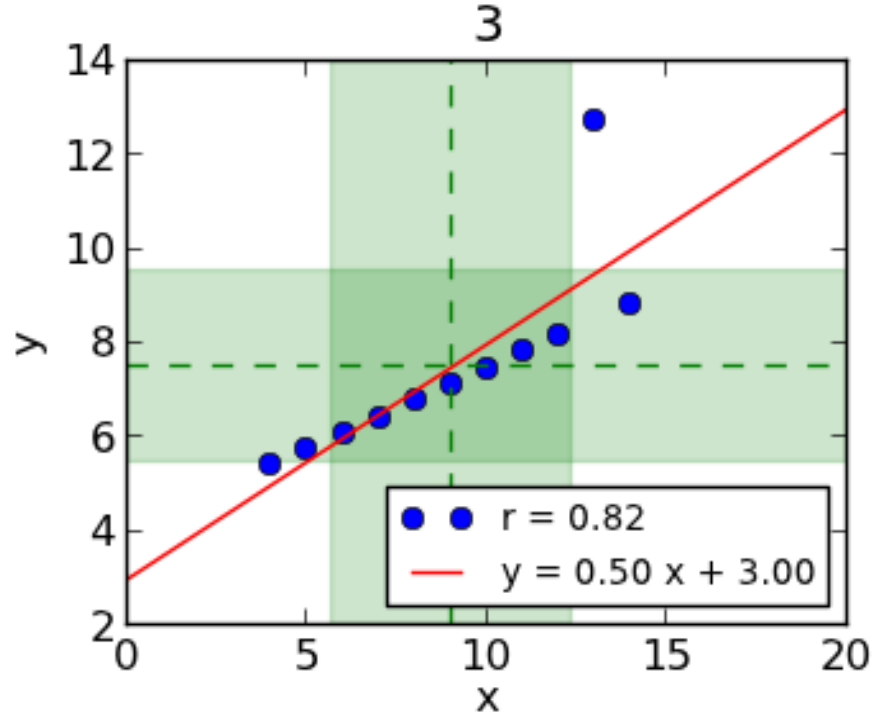
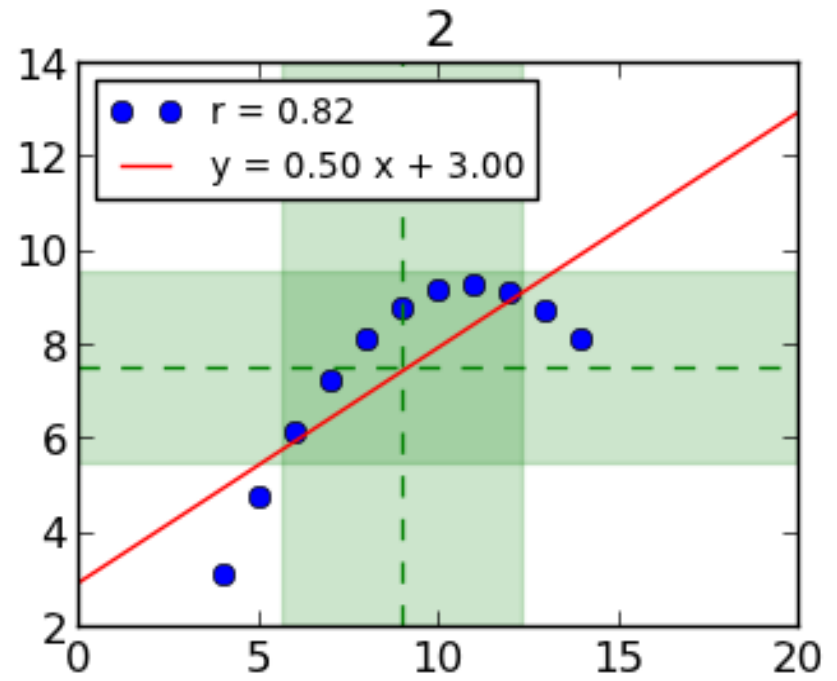
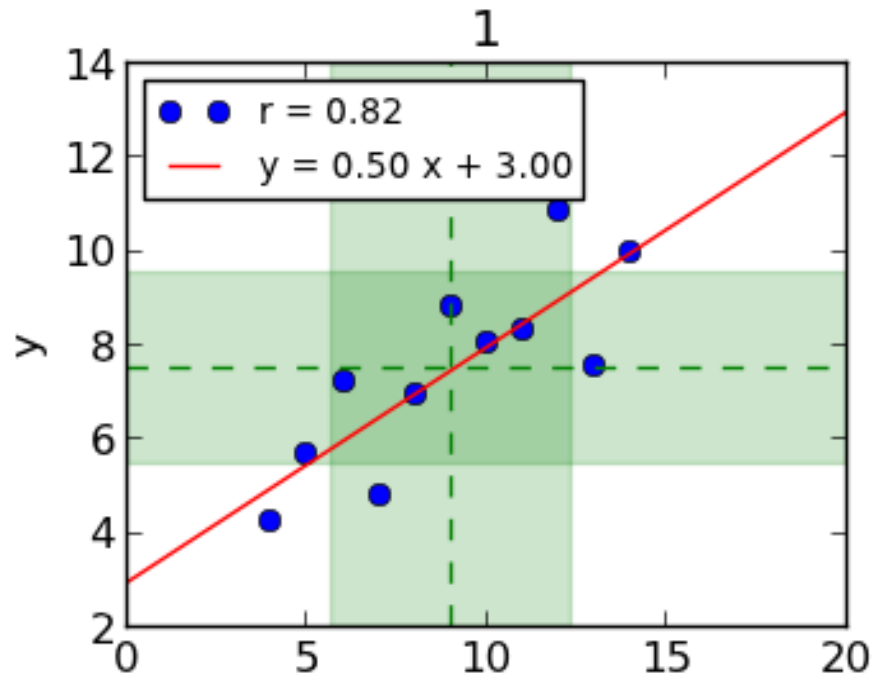
What can you say about them?

Let's calculate some descriptive statistics for dataset #1:

number of observations:	11
mean \pm std dev of x :	9.0 \pm 3.32
mean \pm std dev of y :	7.5 \pm 2.03
linear regression:	$y = 3 + 0.5x$
R^2 :	0.667
correlation coefficient:	0.82
sum of squares of $x - \text{avg}(x)$:	110.0
regression sum of squares:	27.5
residual sum of squares of y :	13.75
estimated std. error of slope:	0.118

For the other data sets we get the same results!!!
so they are all similar, right?

Anscombe's Quartet



Conclusion:

look at the data!

- Discover what the data has to say

John W. Tukey, *Exploratory Data Analysis*,
Addison-Wesley, 1977

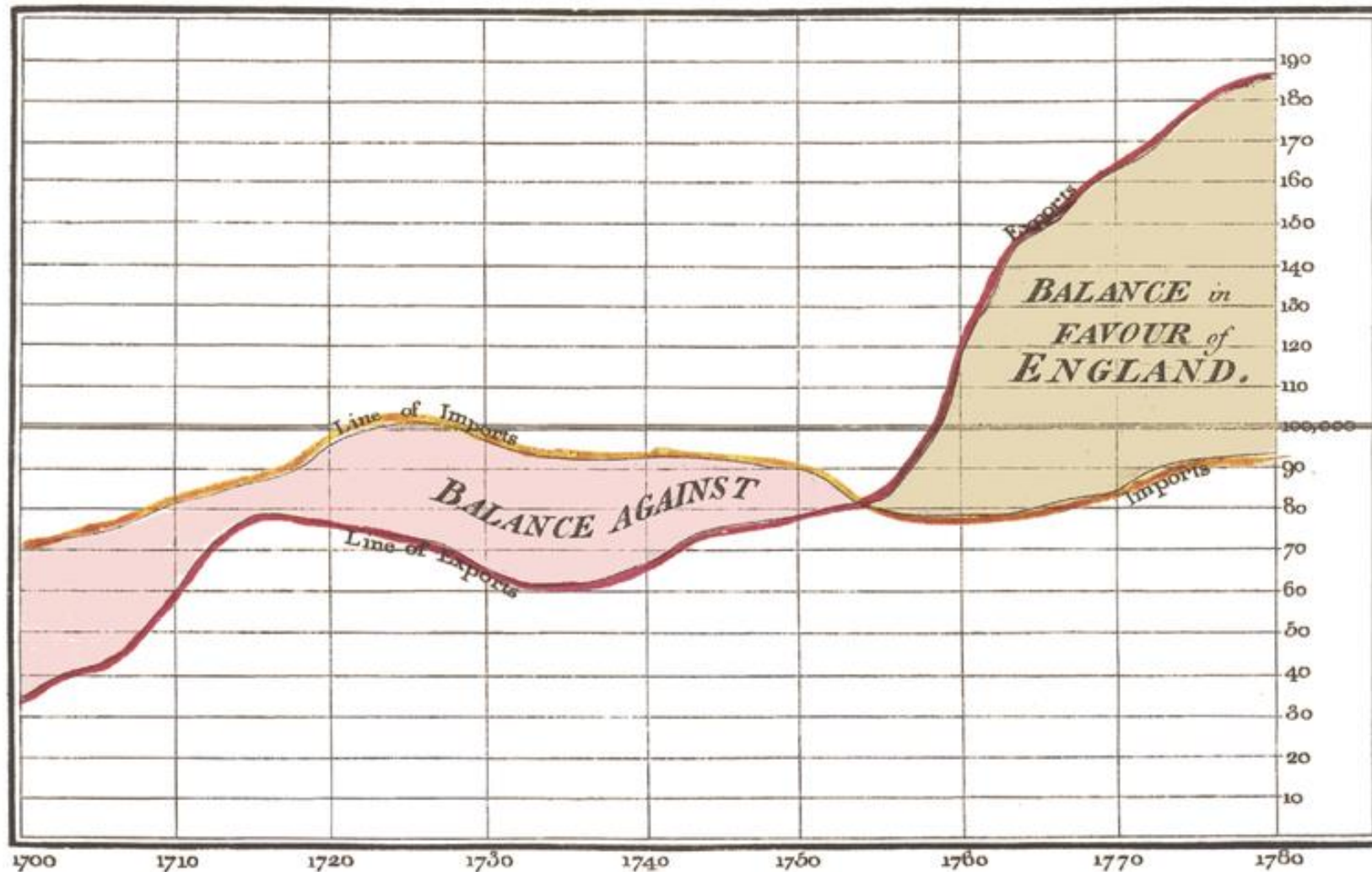
- Display your conclusions in the most convincing manner

Graphs that Made History or Illuminate Data

Michael Friendly's Gallery of data visualization

William Playfair, *The Commercial and Political Atlas*, 1786: invented most graphs used today

Exports and Imports to and from DENMARK & NORWAY from 1700 to 1780.



The Bottom line is divided into Years, the Right hand line into £10,000 each.

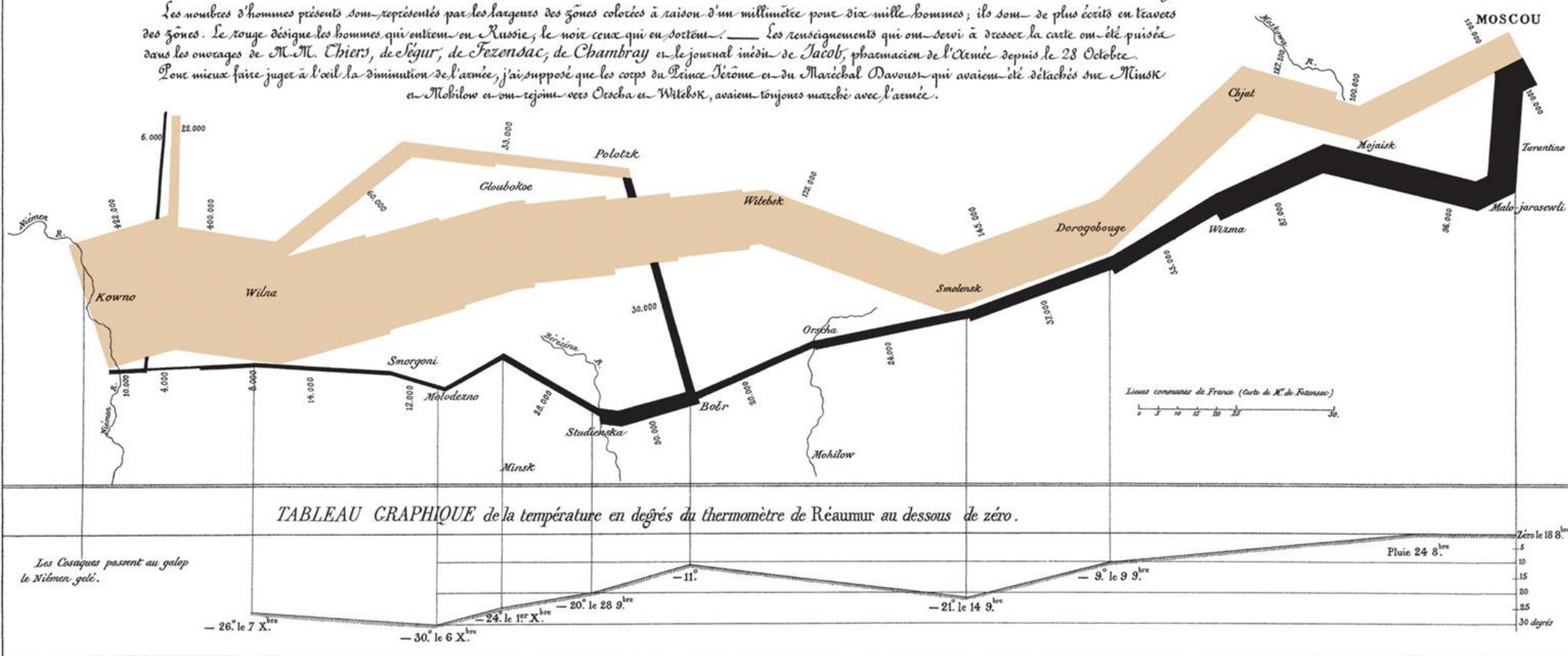
Charles Minard, plot of Napoleon's failed campaign in Russia, 1812

Carte Figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813.

Dressée par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite. Paris, le 20 Novembre 1869.

Les nombres d'hommes présents sont représentés par les largeurs des zones colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en travers des zones. Le rouge désigne les hommes qui ont été en Russie, le noir ceux qui en sont sortis. Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de M. M. Chiers, de Légar, de Fezensac, de Chambrey et le journal inédit de Jacob, pharmacien de l'Armée depuis le 28 Octobre.

Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Prince Jérôme et du Maréchal Davout qui avaient été détachés sur Minsk et Mohilow et ont rejoint vers Orscha et Witebsk, avaient toujours marché avec l'armée.

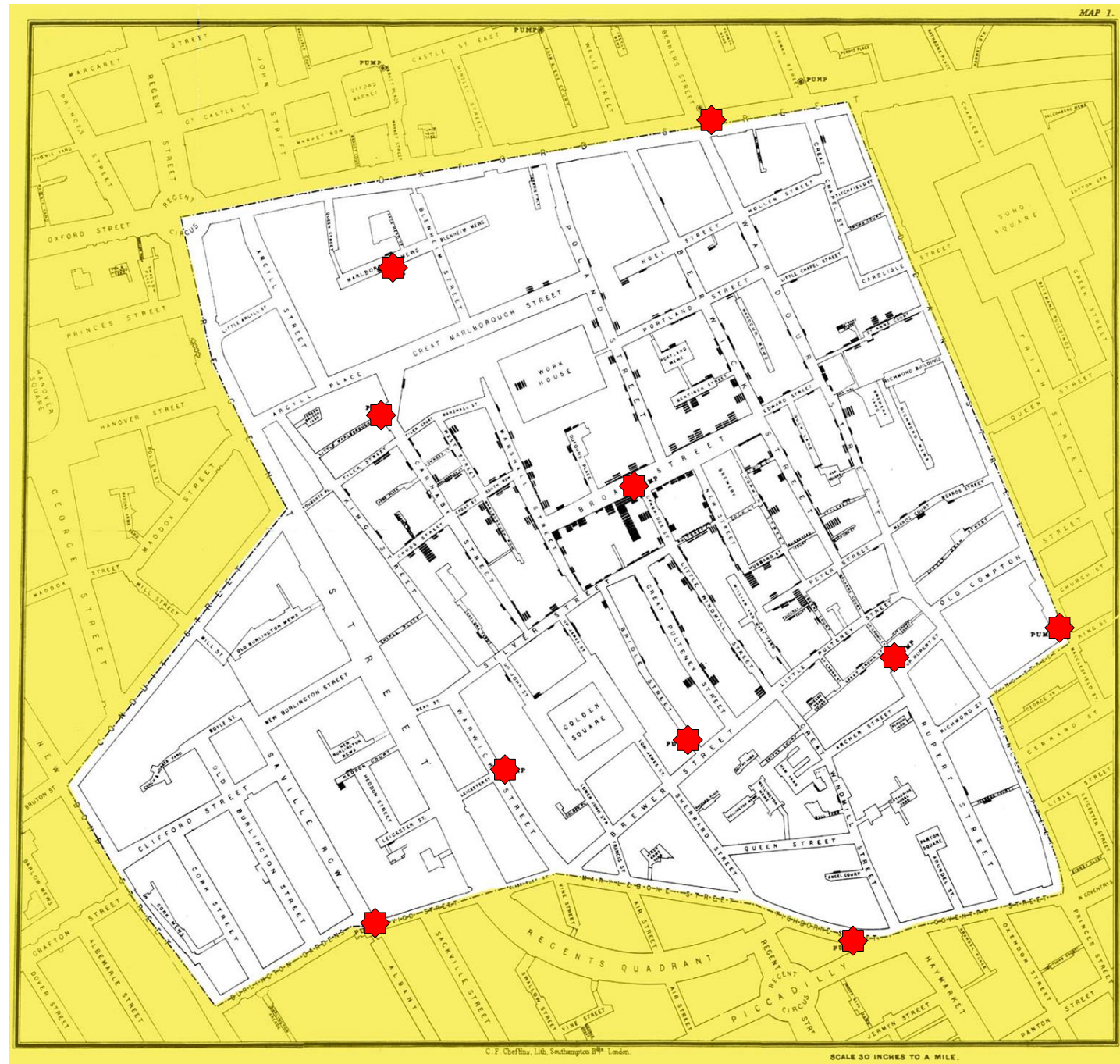


Popularized by Tufte as the best graphic ever

John Snow, deaths in London Cholera epidemic, 1854

Established link between water quality and health

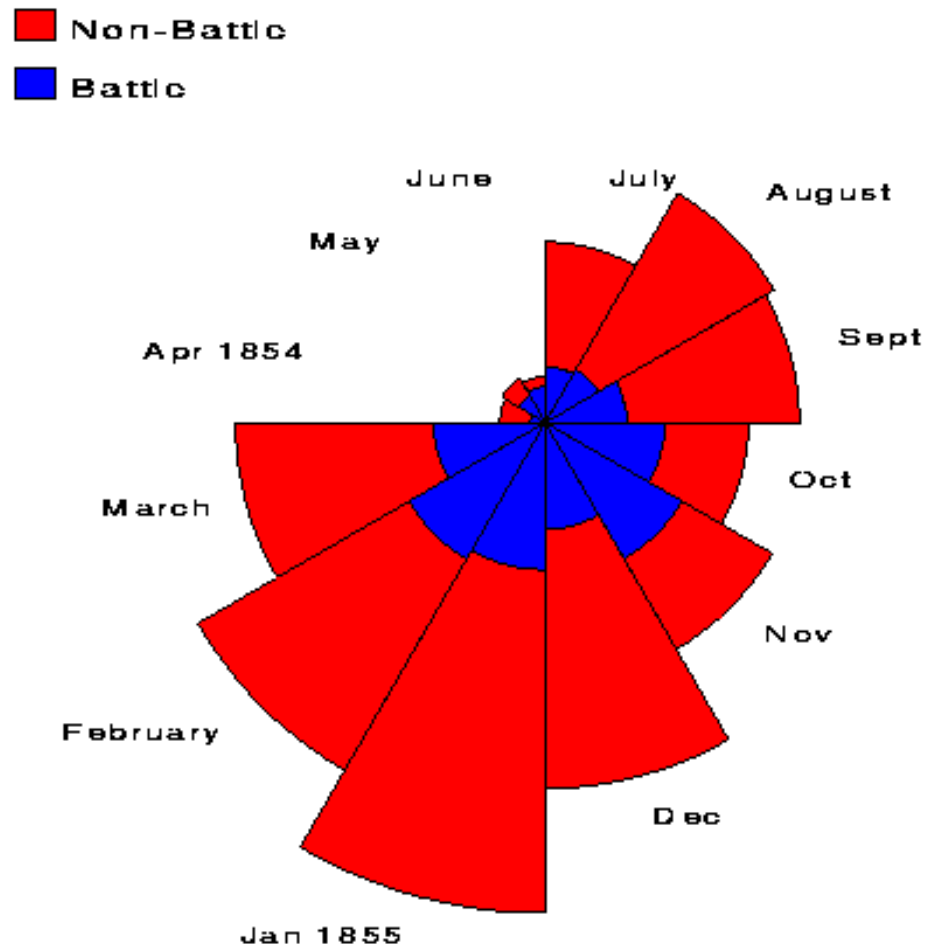
Precursor of modern GIS systems



Florence Nightingale, British casualties in Crimean war, 1858

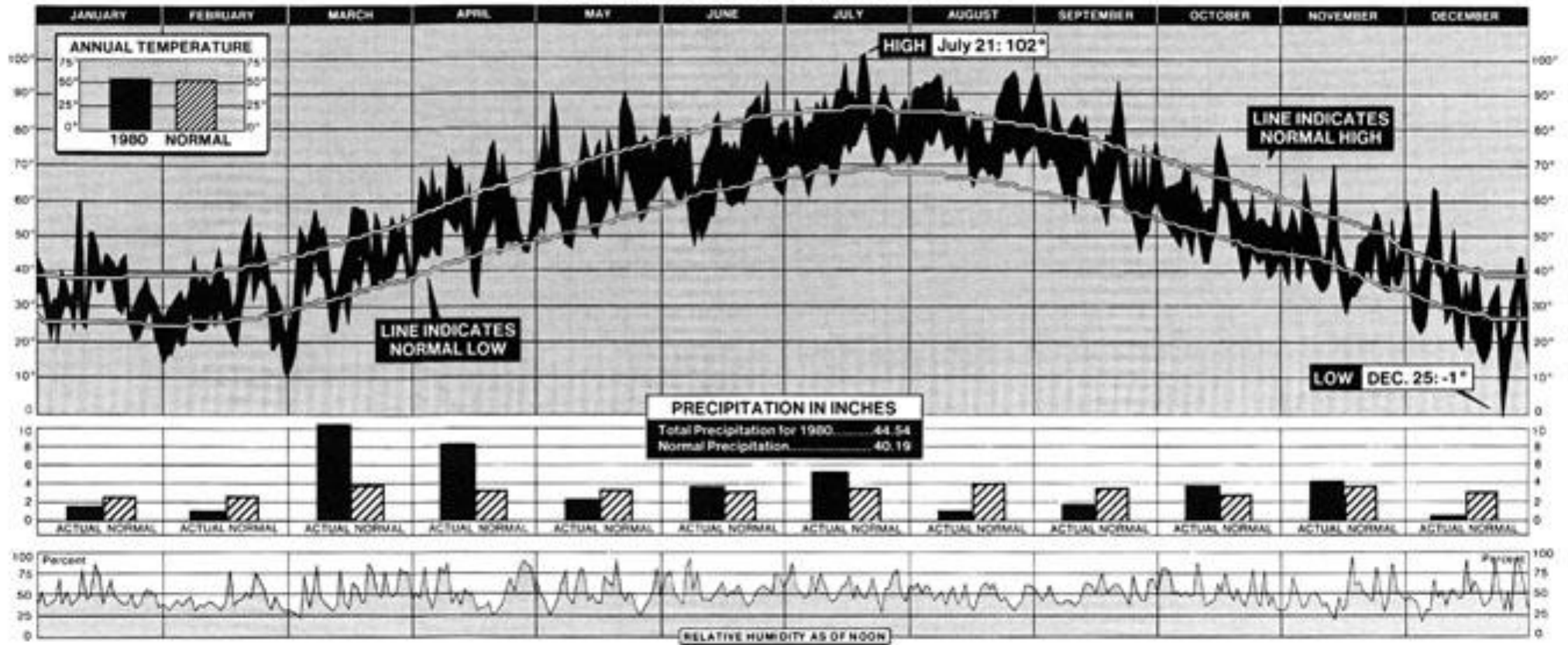
Established sanitation as a decisive factor in hospital operation

Causes of Mortality in the Army in the East
April, 1854 to March 1855



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

NEW YORK CITY'S WEATHER FOR 1980



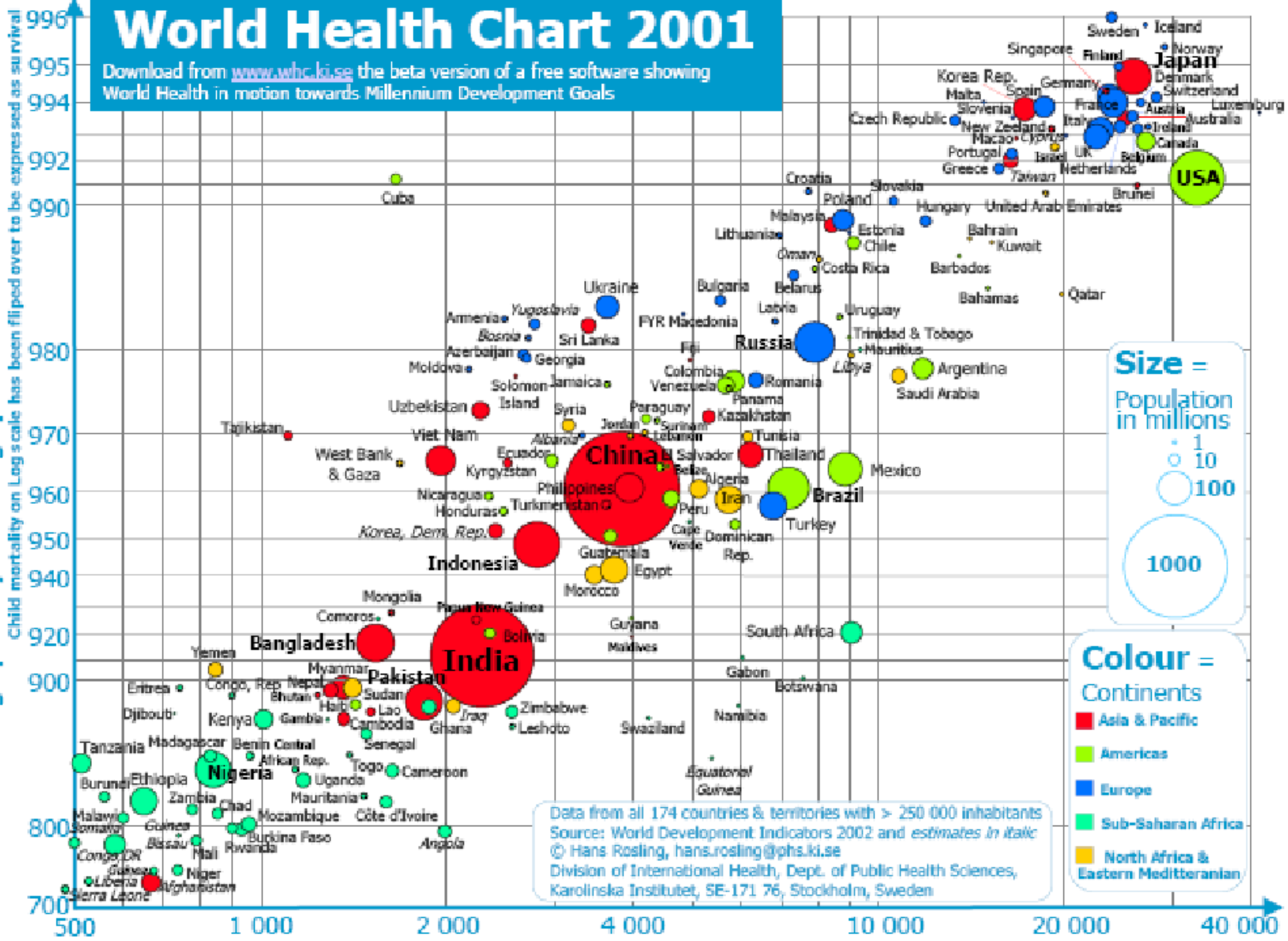
Summary of a whole year's weather

- Lots of numbers (daily max/min + average max/min + humidity)
- Use of parallel graphs for correlation
- Callouts to emphasize special points

Children surviving up to 5 years of age per 1000 live births = Health

World Health Chart 2001

Download from www.whc.ki.se the beta version of a free software showing World Health in motion towards Millennium Development Goals



Size =
Population in millions

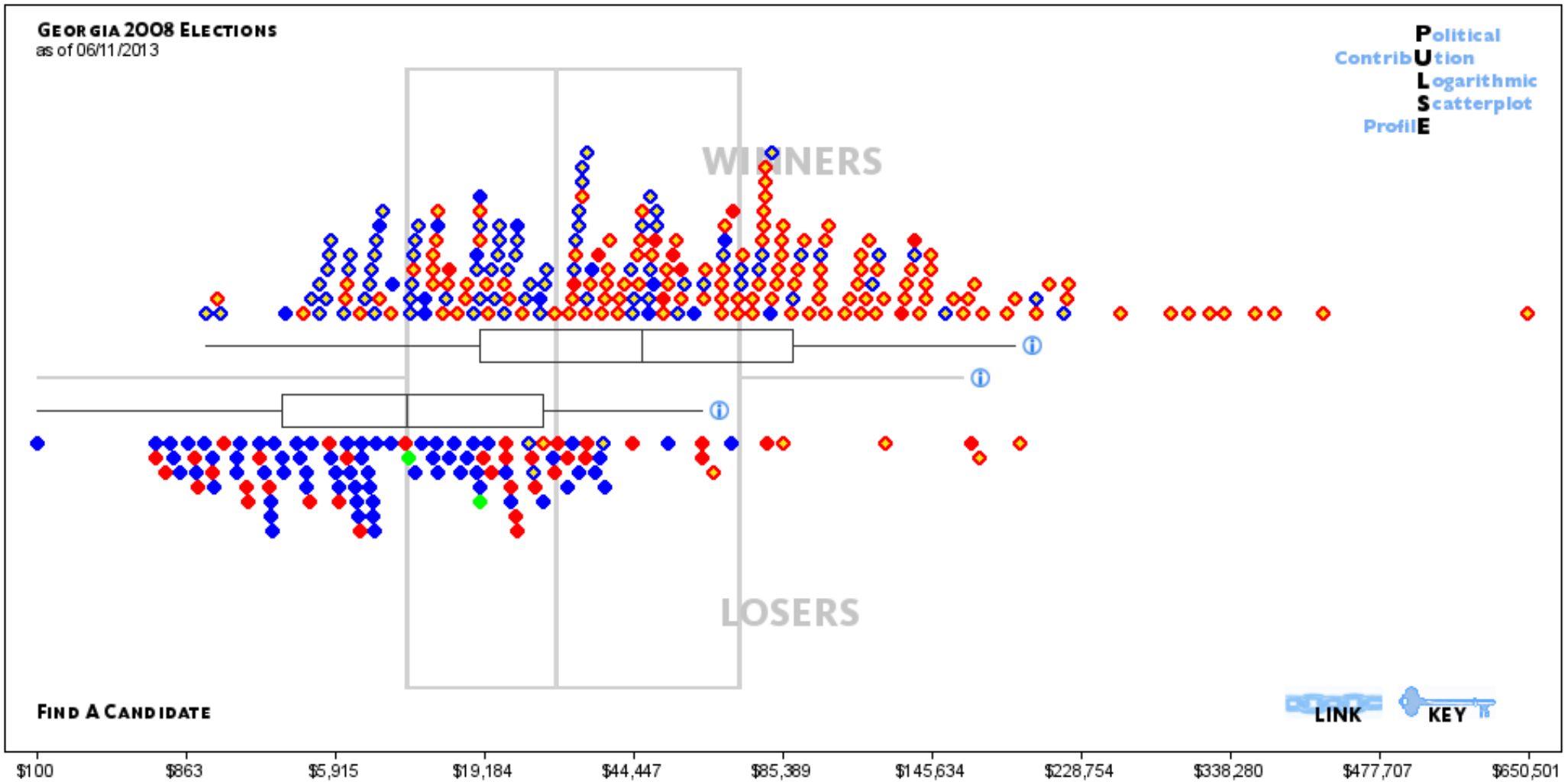
- 1
- 10
- 100
- 1000

Colour =
Continents

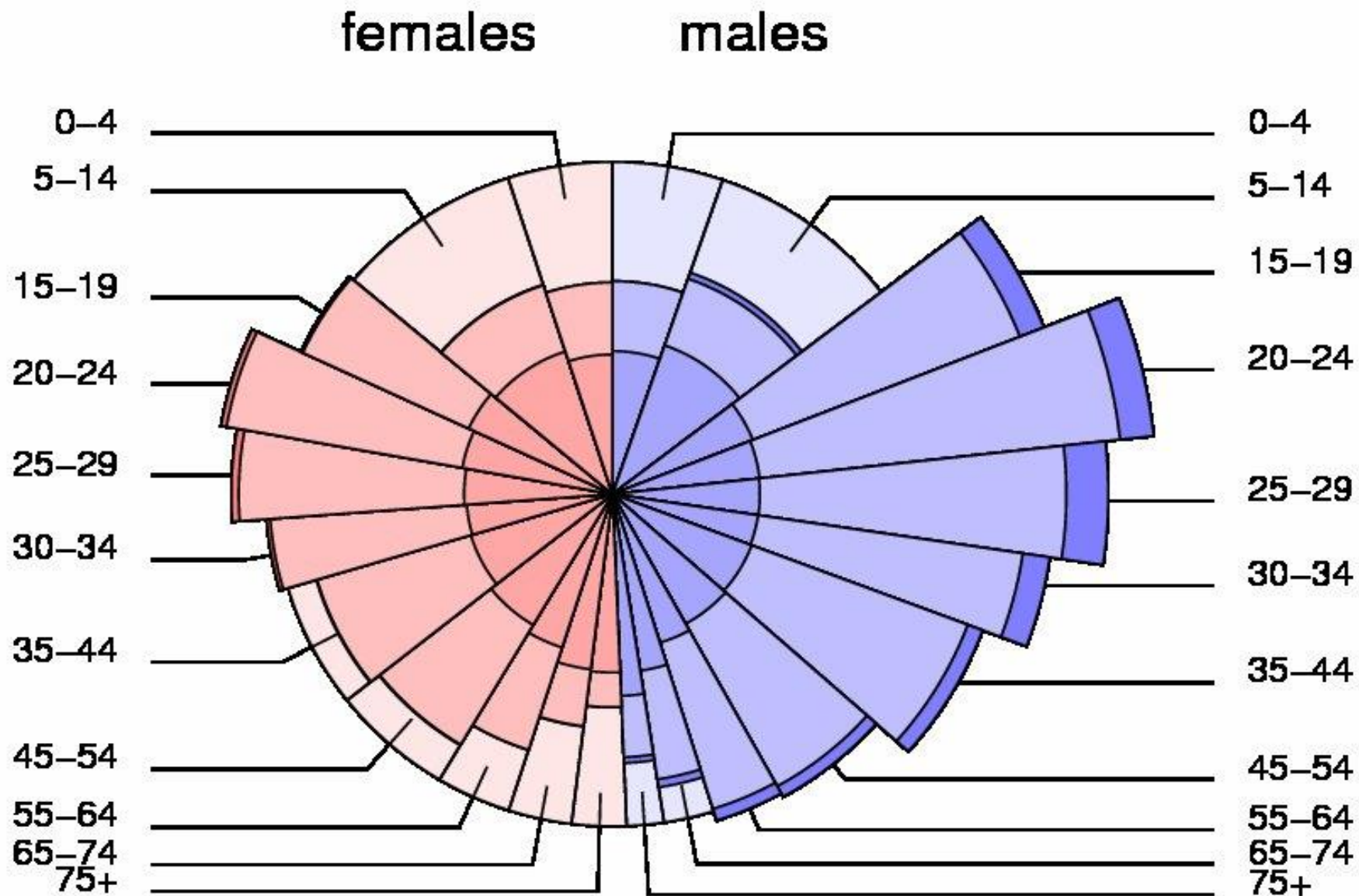
- Asia & Pacific
- Americas
- Europe
- Sub-Saharan Africa
- North Africa & Eastern Mediterranean

Data from all 174 countries & territories with > 250 000 inhabitants
Source: World Development Indicators 2002 and estimates in italic
© Hans Rosling, hans.rosling@ghs.ki.se
Division of International Health, Dept. of Public Health Sciences, Karolinska Institutet, SE-171 76, Stockholm, Sweden

Gross Domestic Product per capita in US dollar purchasing power parity (log scale) = Money



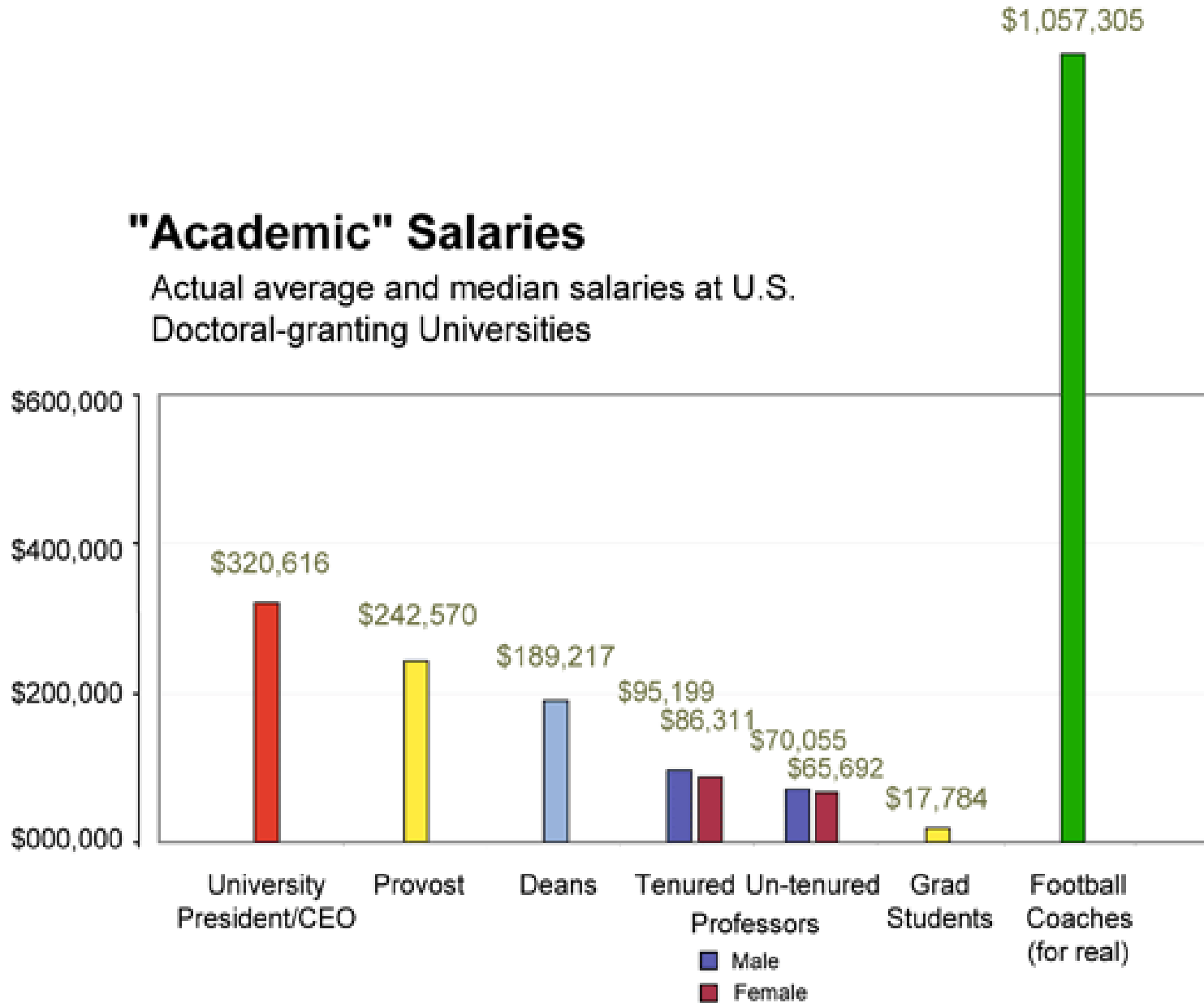
Winners in house election are mainly republican incumbents who raised more money



Road casualty data relative to population.
 Radial sections are pedestrians, cars, and motorcycle

"Academic" Salaries

Actual average and median salaries at U.S. Doctoral-granting Universities



JORGE CHAM © 2008

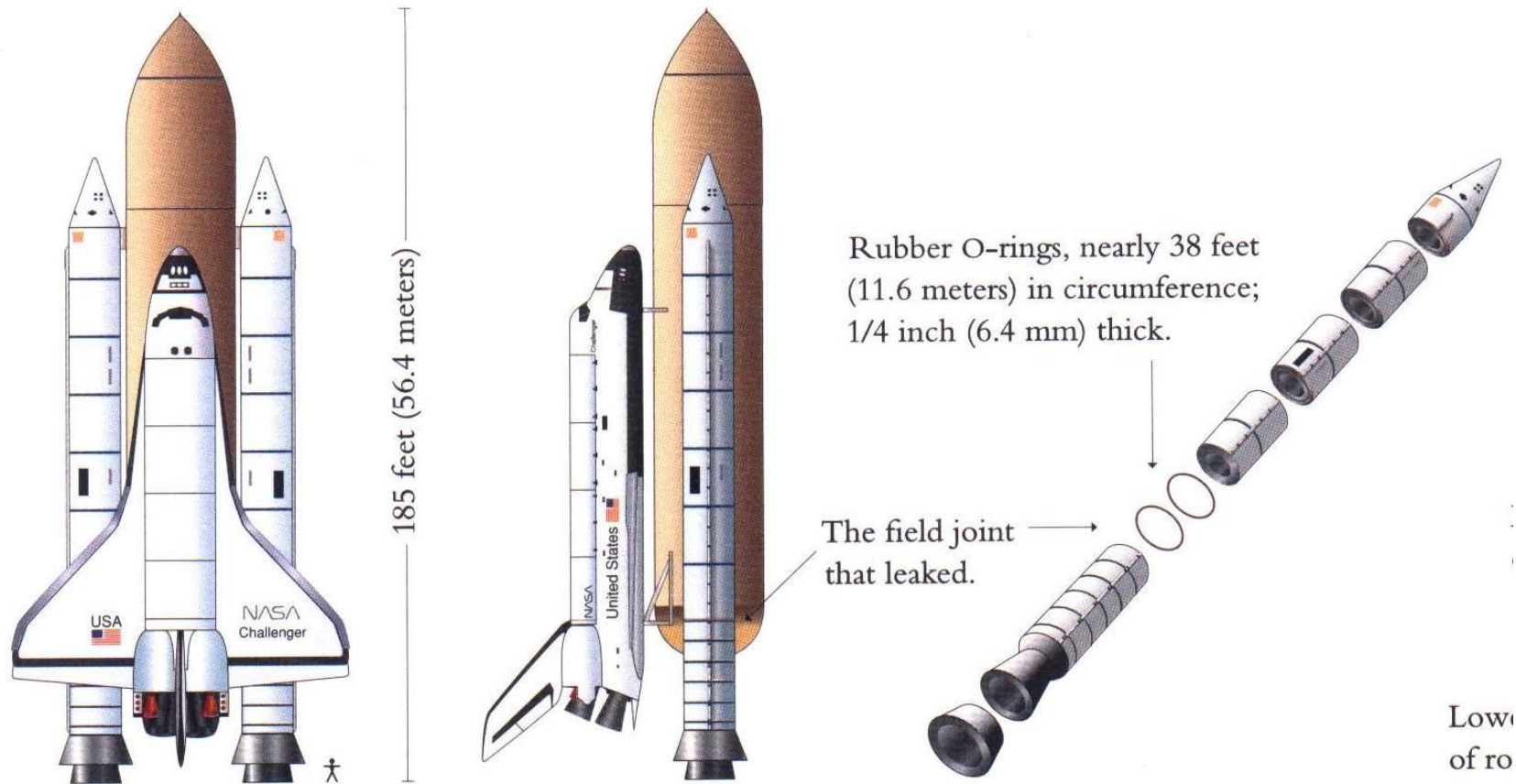
Notes: Administrator figures are medians salaries, the rest are averages. All figures in 2008 dollars. Sources: College and University Professional Association for Human Resources 2005 Survey; American Association of University Professors 2007 Survey; The Chronicle of Higher Education 2001 Survey of Graduate Assistants; USA Today Survey of Div. I-A College Football Coaches Compensation 2007.

WWW.PHDCOMICS.COM

The Harm of Bad Graphics

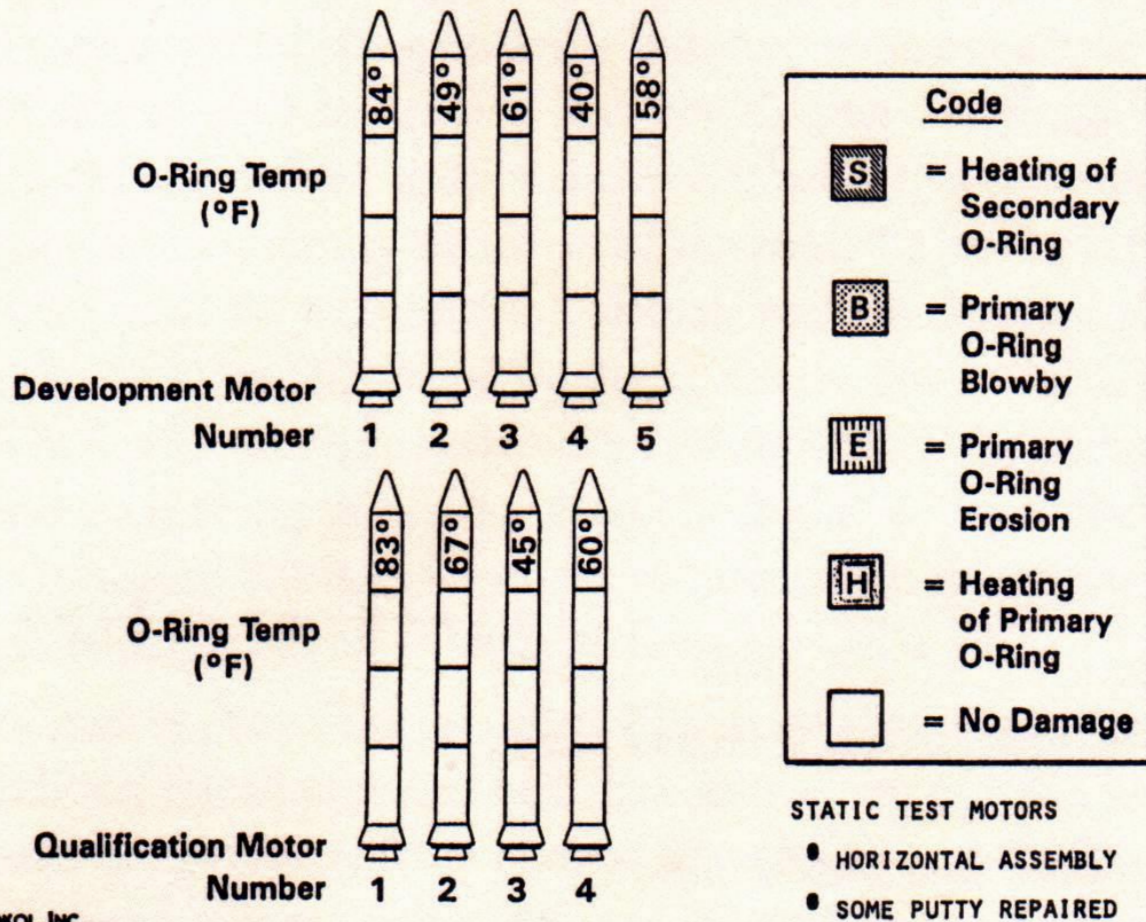
Tufte, *Visual Explanations*
using graphs by Morton Thiokol, Inc.

Background: launch of the Challenger space shuttle on 27 January 1986, amid concerns regarding O-ring function in cold weather



Data regarding test rockets from the manufacturer (chart prepared later; charts used in discussions prior to the launch contained less data)

History of O-Ring Damage in Field Joints



MORTON THIOKOL, INC.
Wasatch Operations

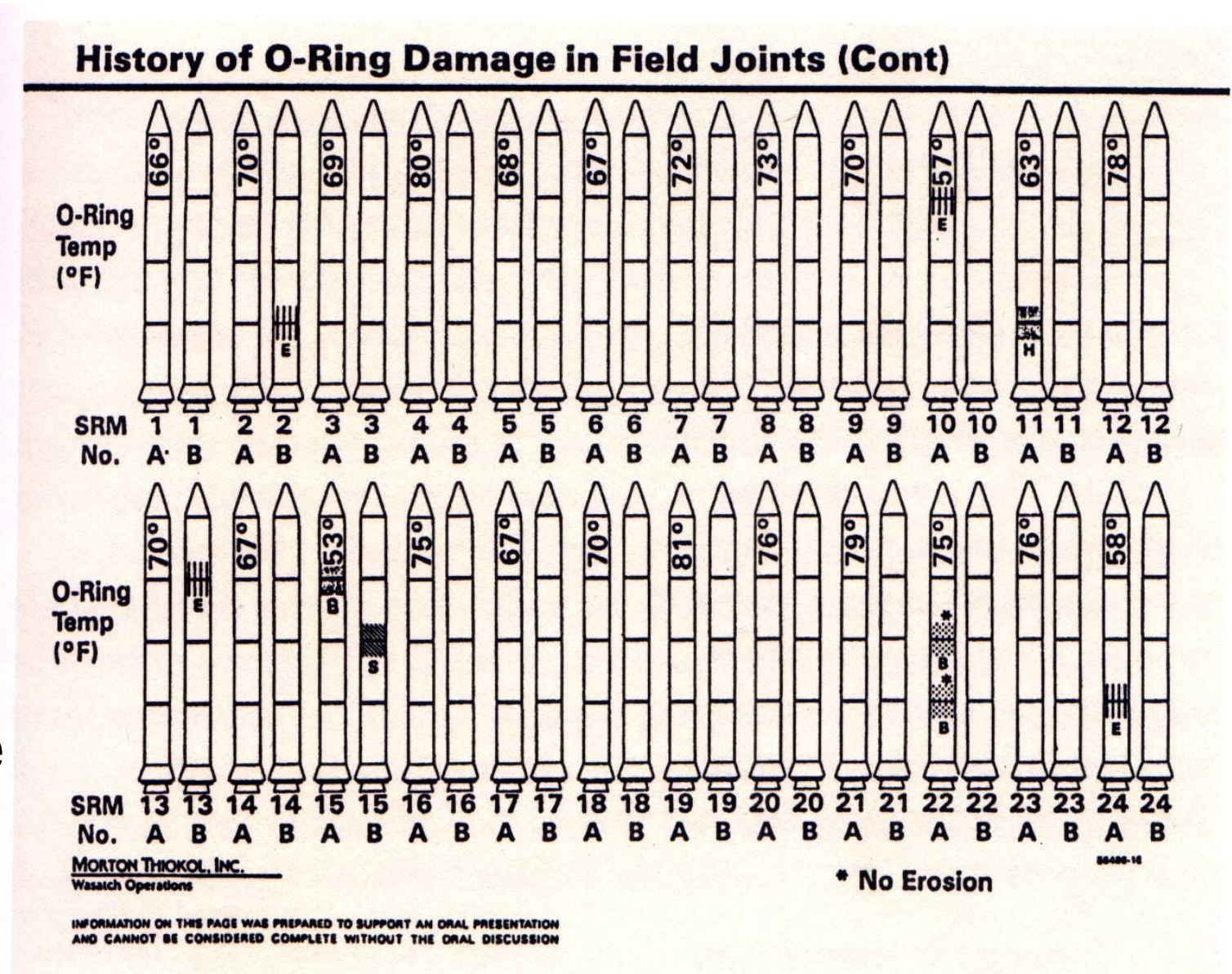
INFORMATION ON THIS PAGE WAS PREPARED TO SUPPORT AN ORAL PRESENTATION AND CANNOT BE CONSIDERED COMPLETE WITHOUT THE ORAL DISCUSSION

Data regarding prior launches

Note that legend is missing (appeared previously)

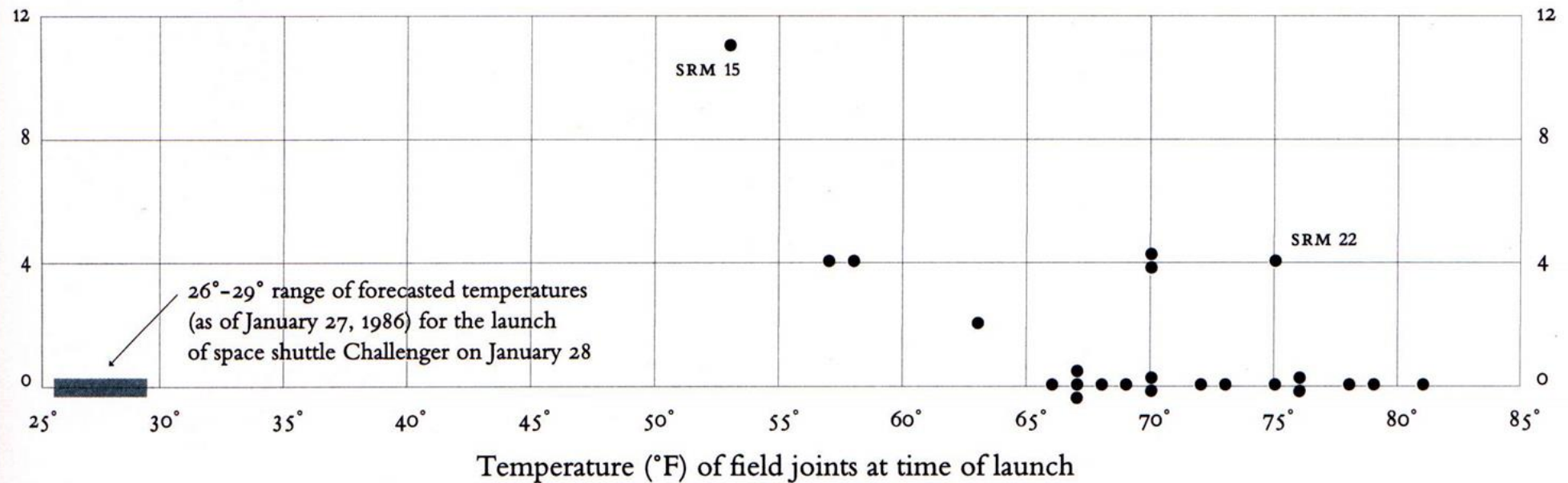
Contains too much irrelevant data

Does not clarify effect of temperature



Tufte's alternative rendering of the data

O-ring damage
index, each launch

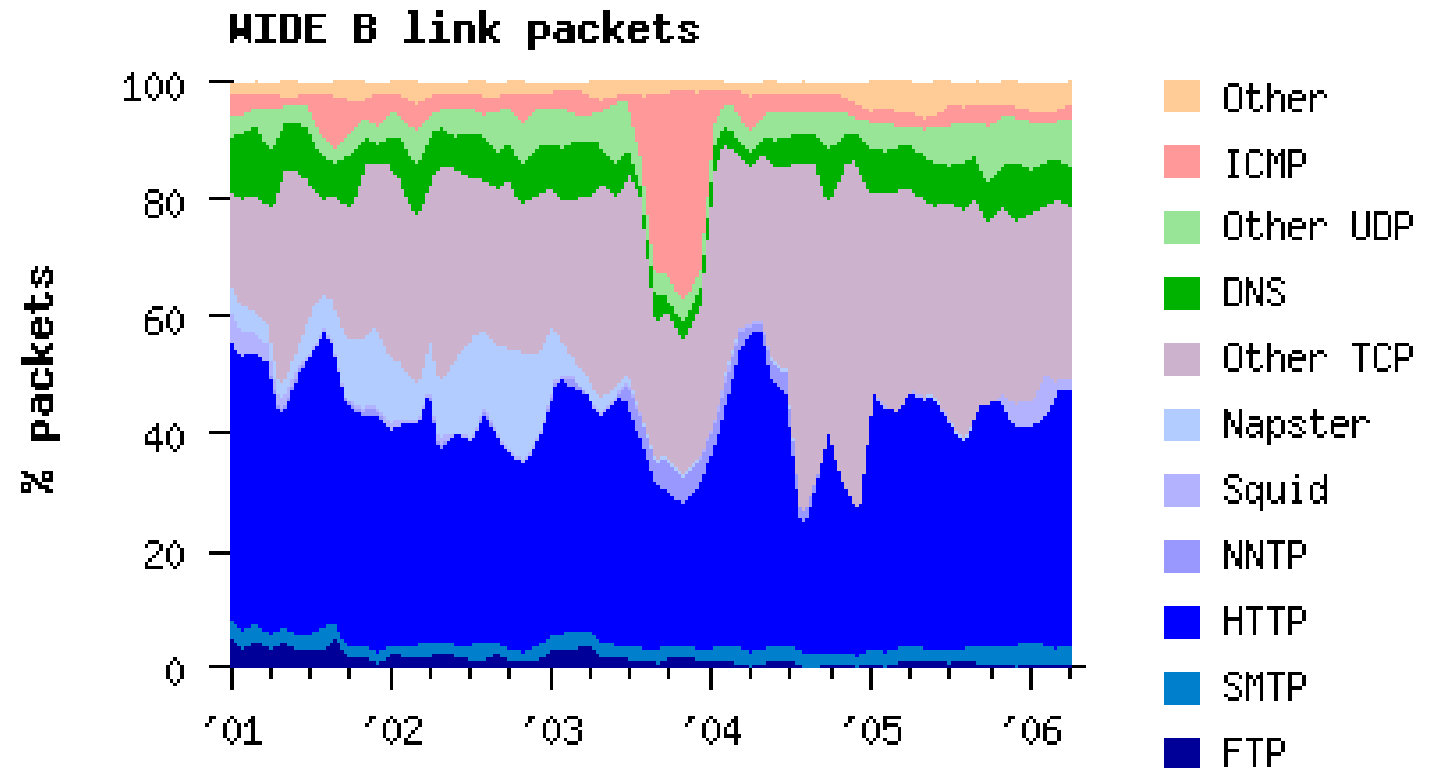


Note exaggerated X scale for emphasis

Principles and Examples

A graph should be independent and provide full information

- Title (if relevant)
- Axis labels (including units)
- Tics indicating values
- Legend

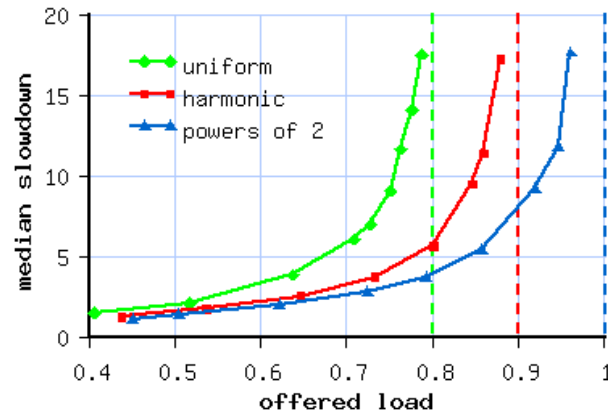


Need to also consider aesthetics

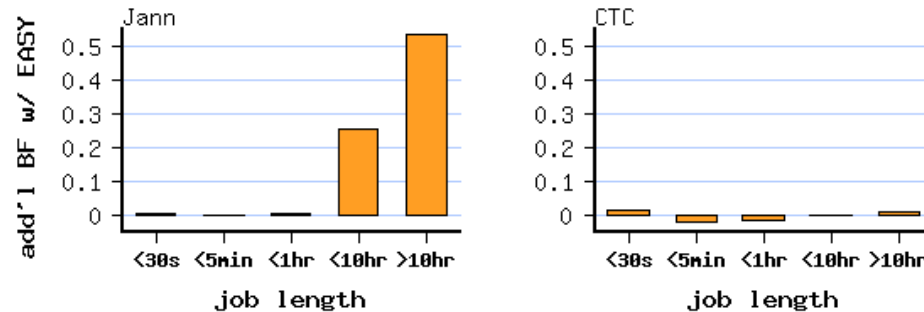
- Proportions
 - Size and placement of labels and legend
 - Size of fonts relative to graphical elements
- Use of color
 - Express gradient with deeper shades
 - Create focus for discussion
 - Should also work in black and white / colorblind
- Combination of graphical elements
 - Give full picture
 - Connections through consistent use of colors
- Order in legend matches order of graphs

Causal/functional relationship:

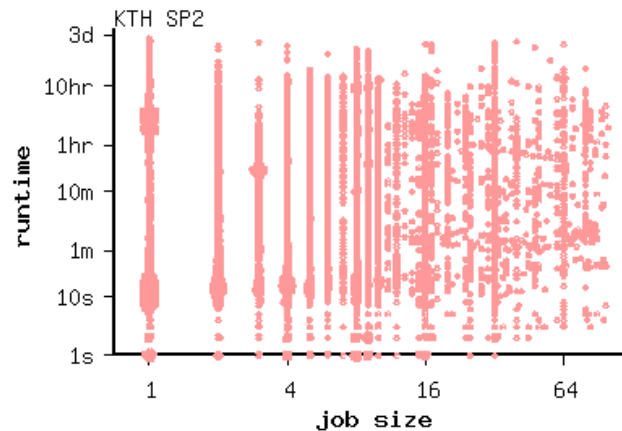
- XY plot
(continuous)



- bar chart
(categories)

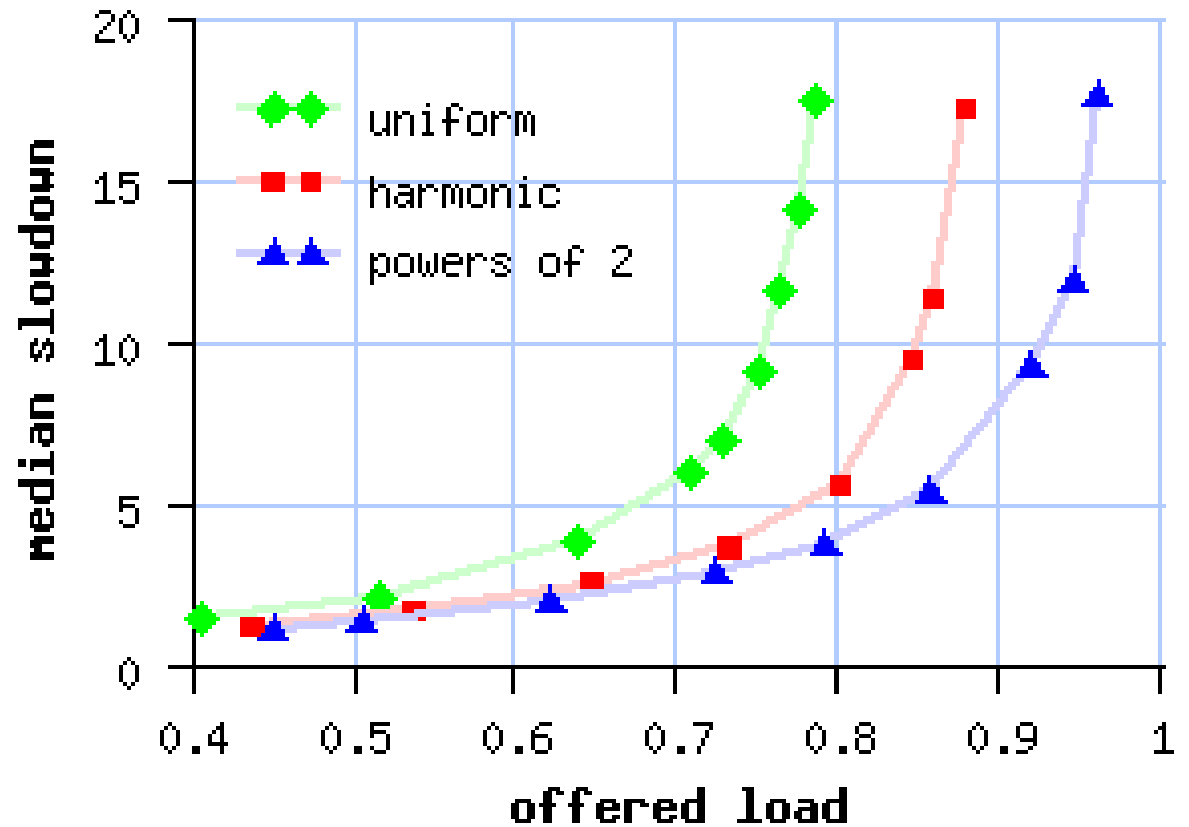


- scatter plot
(complicated)



Showing measurements

- Emphasize points
- Connect with weaker lines

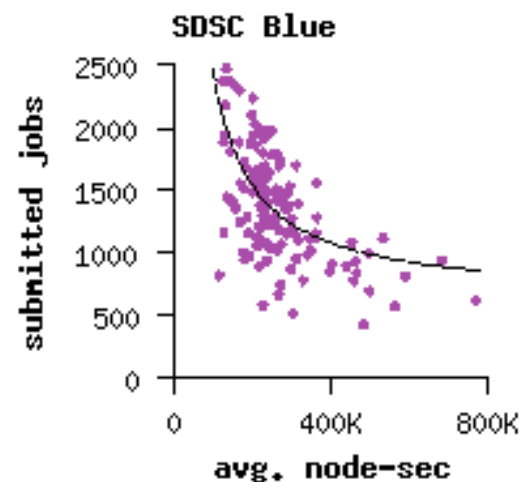
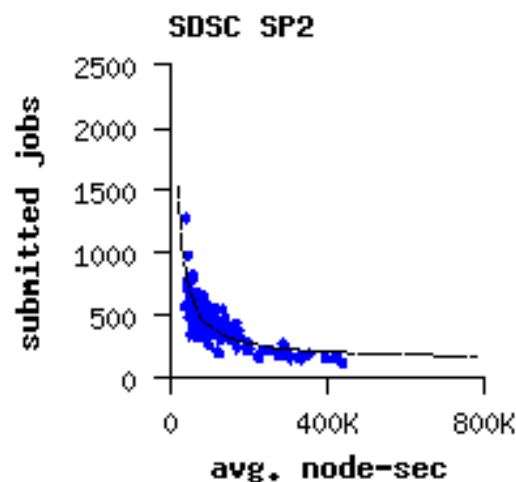
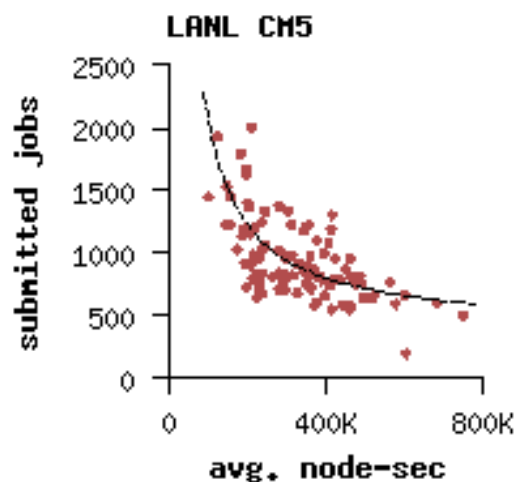
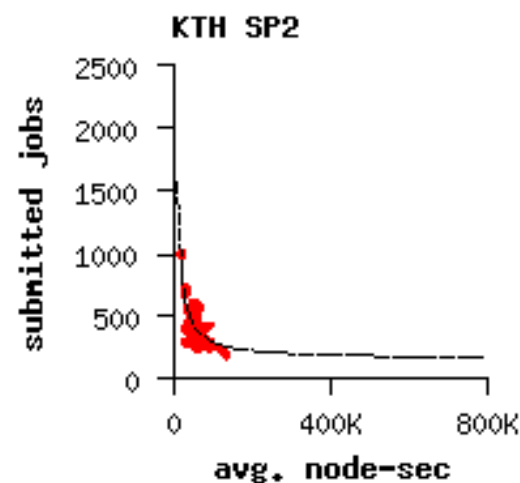
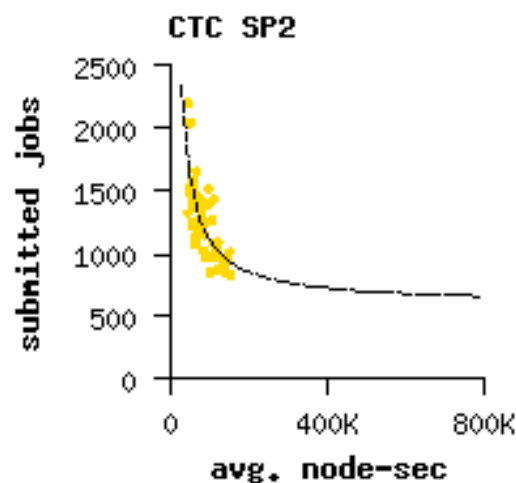
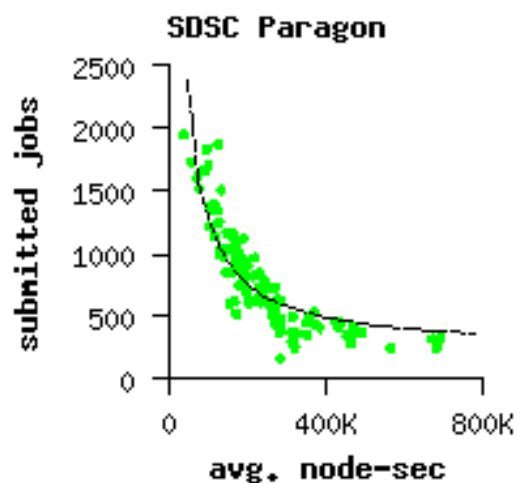


Or show fitted
model line

here model is

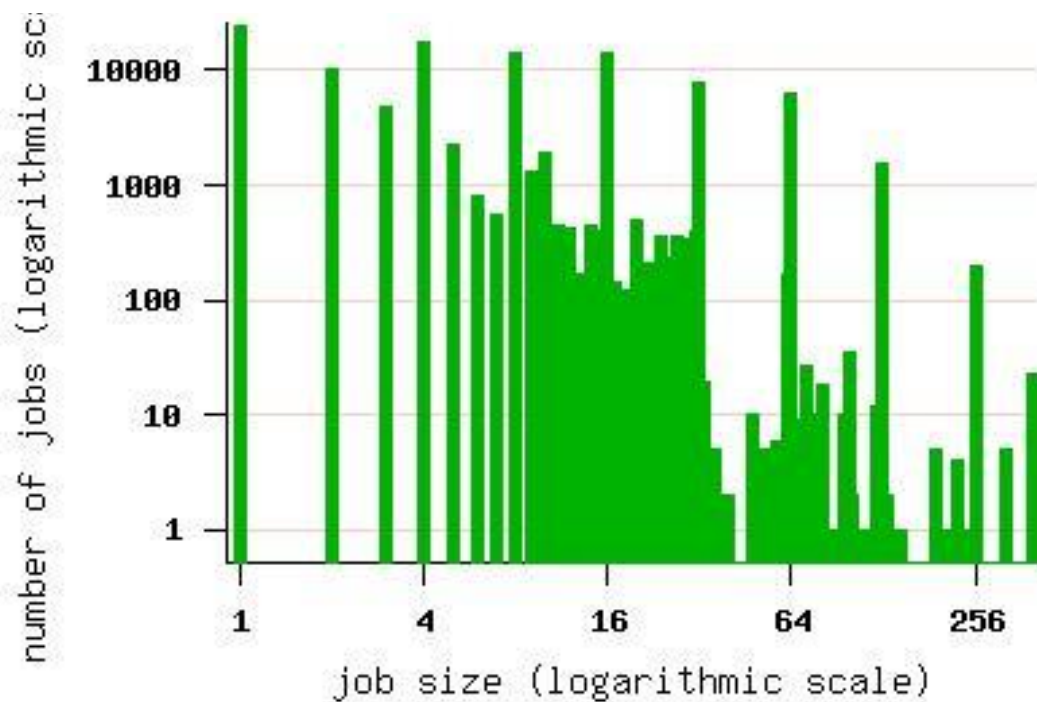
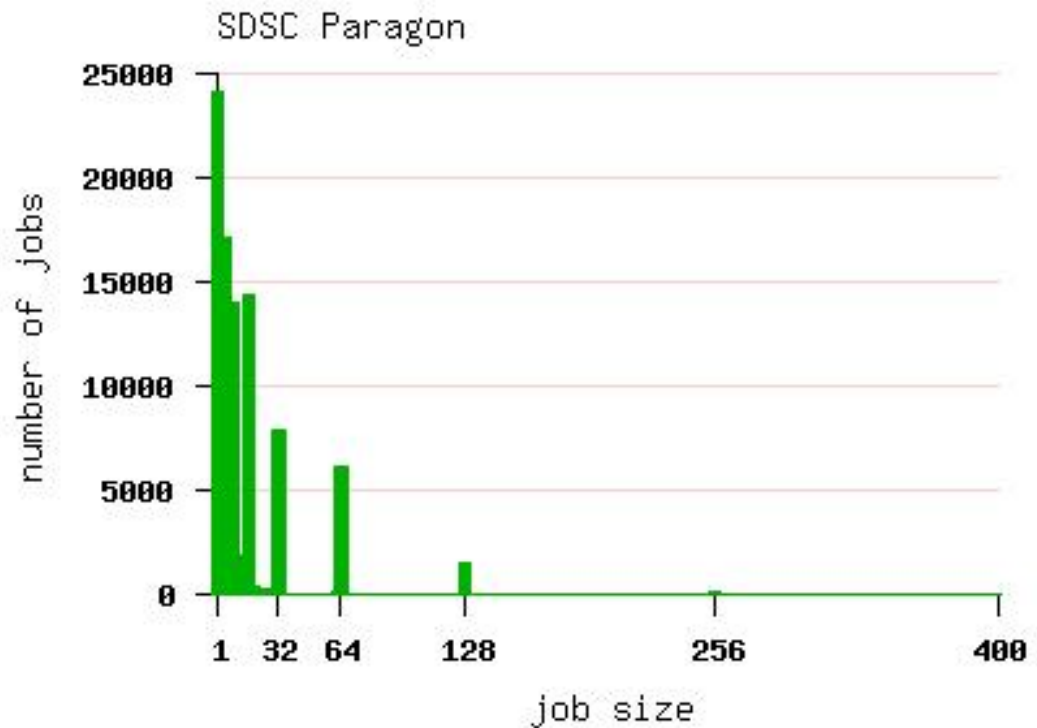
$$y \sim 1/x$$

often linear
regression



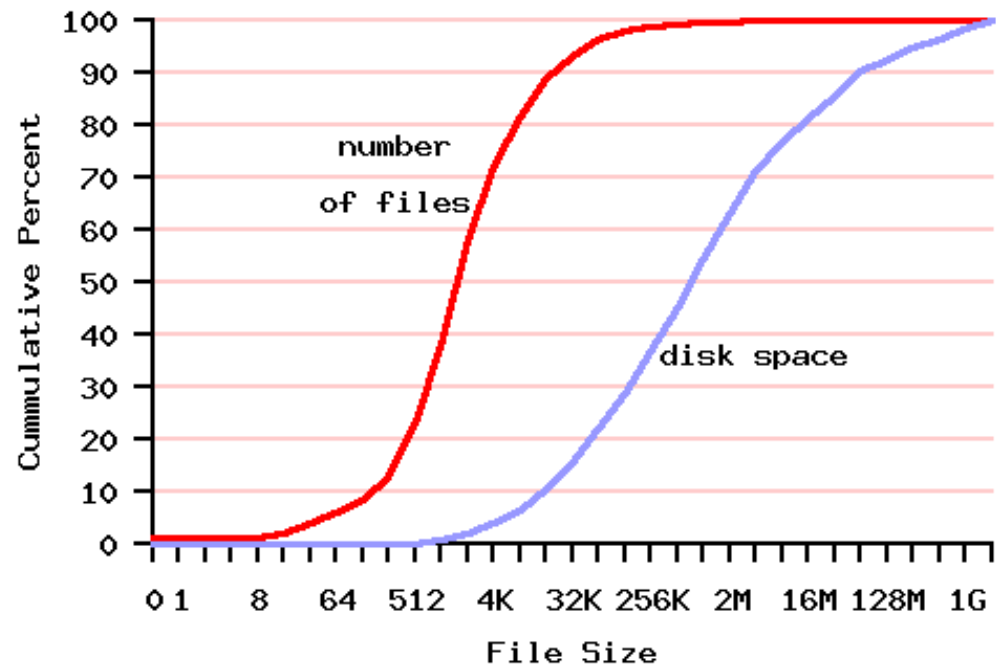
Scales

- Linear is best
- Logarithmic if needed



Scales

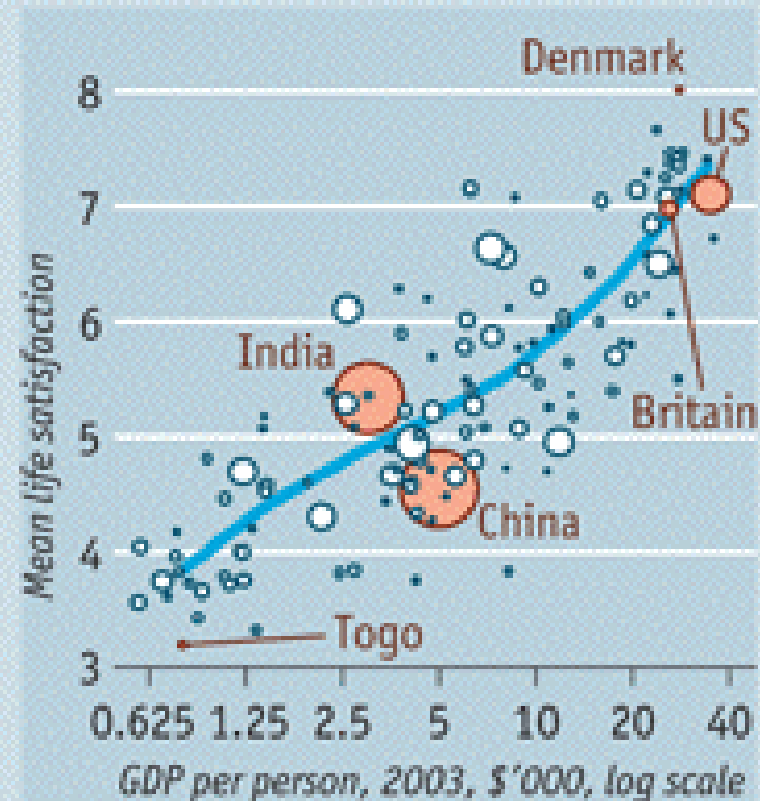
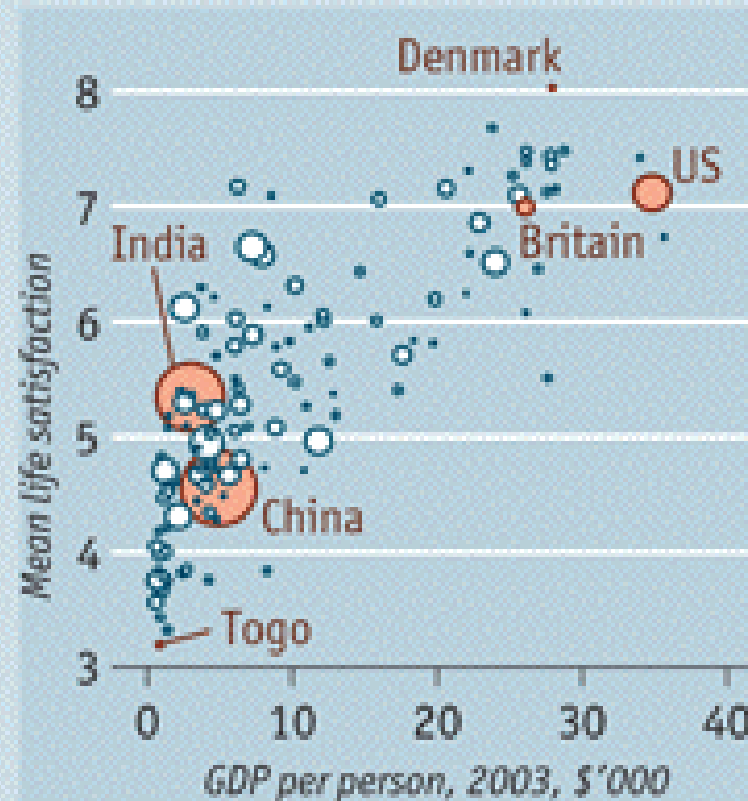
- Logarithmic if needed
- Show values, not their log, in stubs



Happiness is a logarithmic scale

Life satisfaction and GDP per person at PPP*

Circle size is proportional to population size



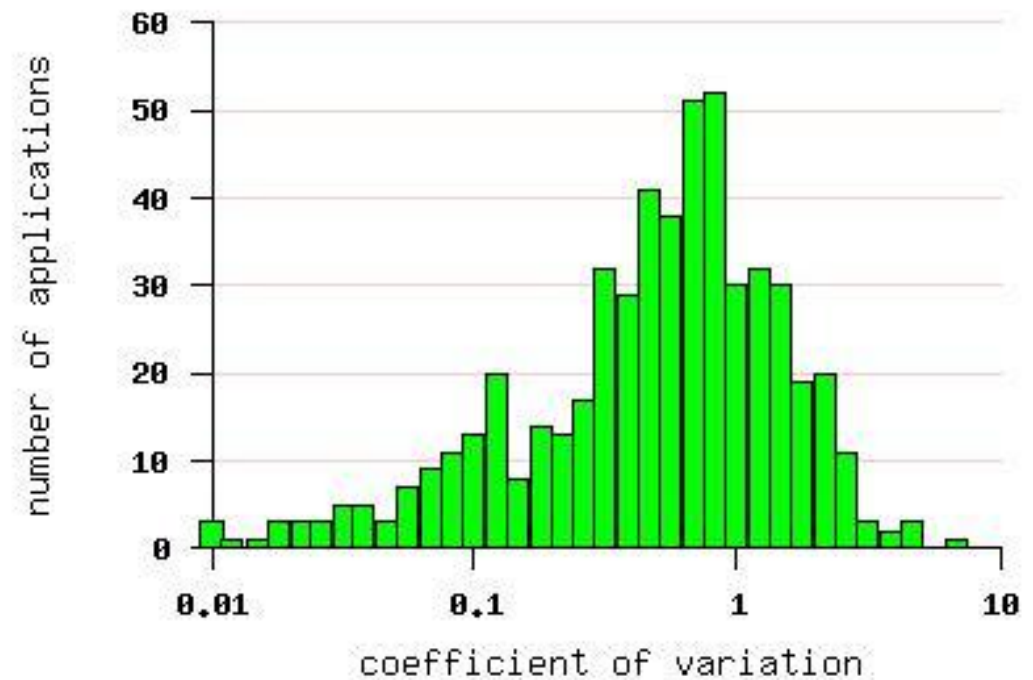
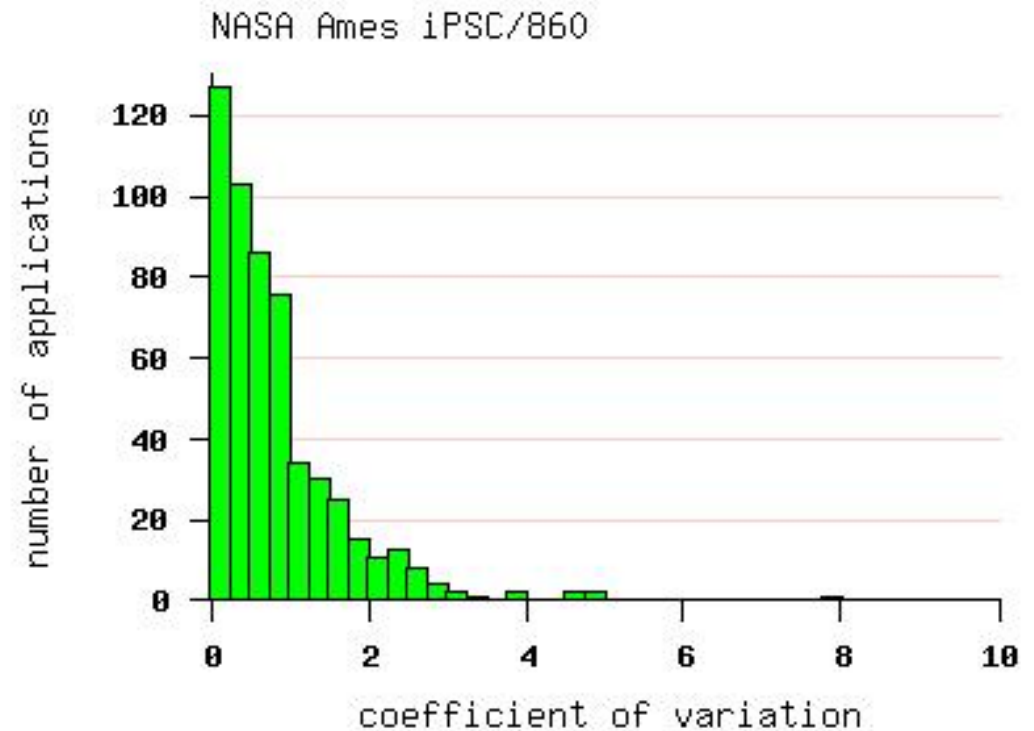
Sources: Penn World Table 6.2; Gallup World Poll, Angus Deaton

*Purchasing-power parity

Log scale can expose functional relationship

Log scale

- Beware of expansion in small values

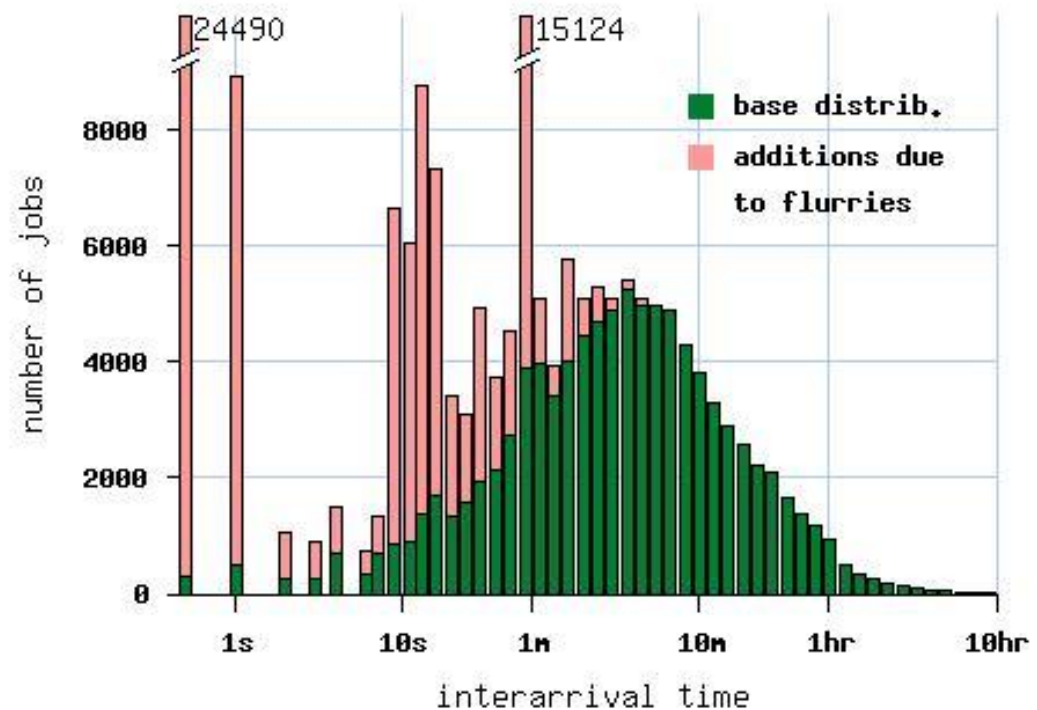
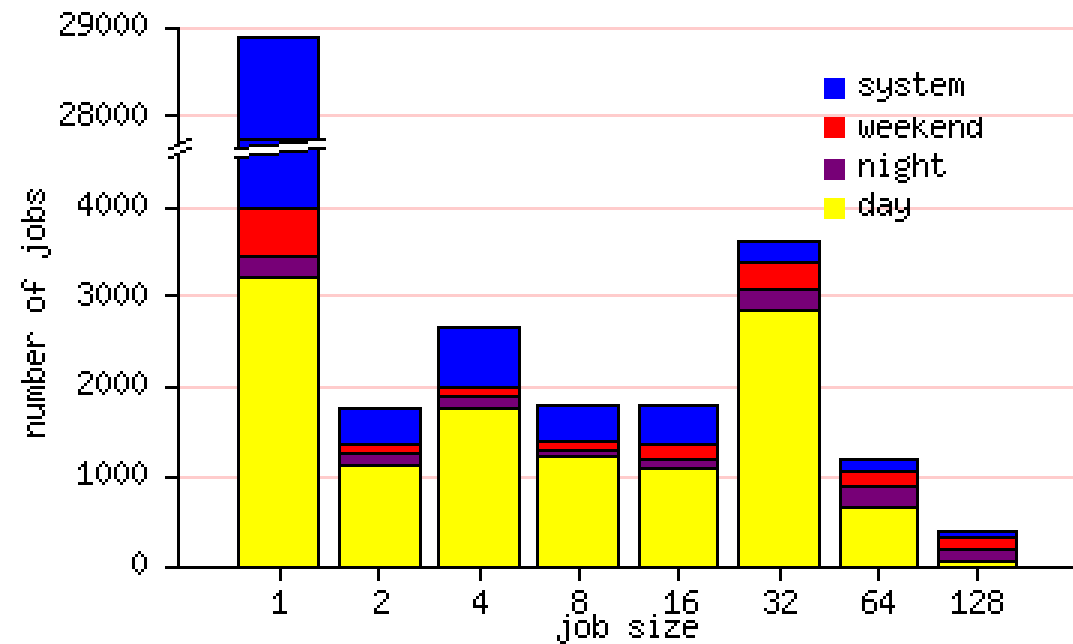


Stubs

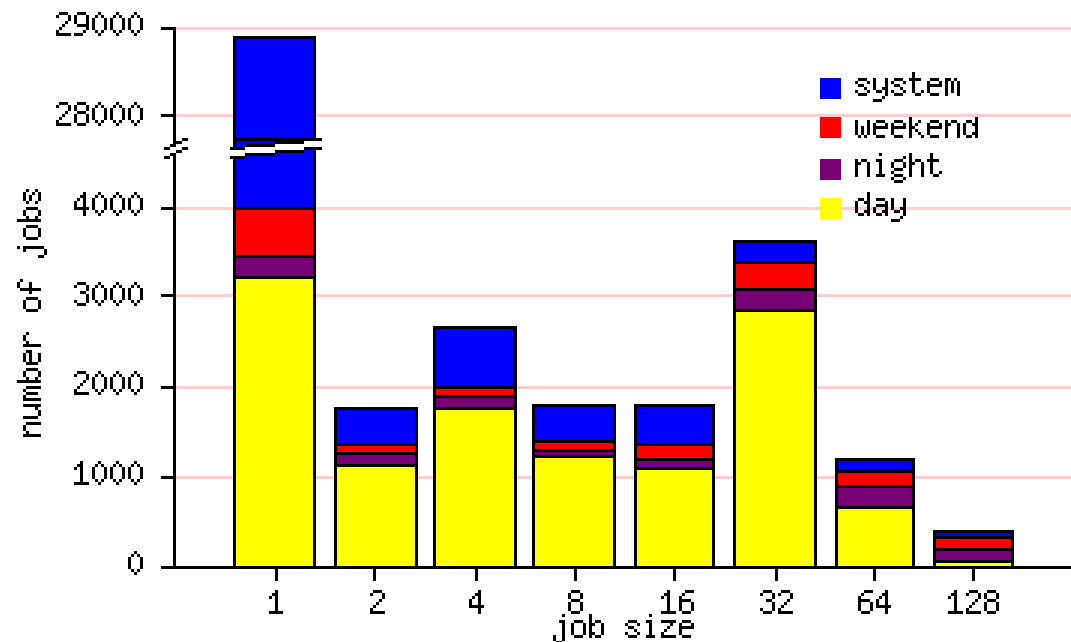
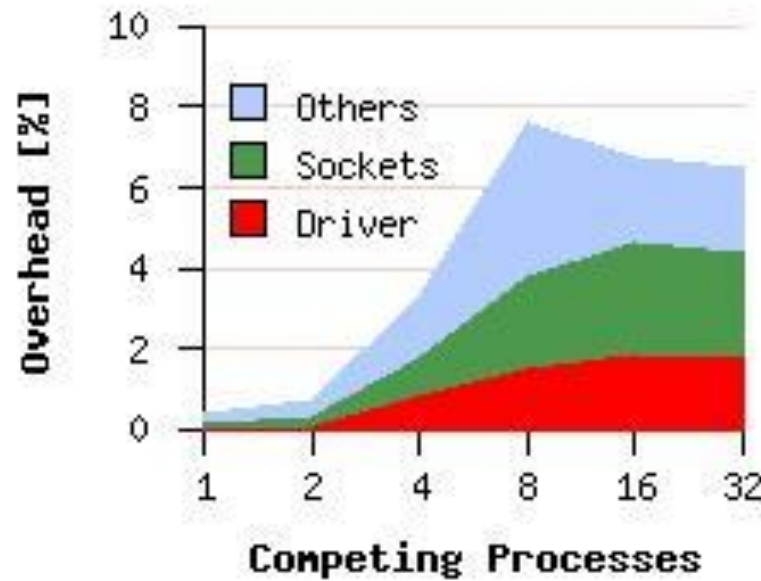
- Uniform scale (Y axis)
- Match measured values (powers of 2)
- Show important values (maximal size)



Axis break
useful for few
extreme values

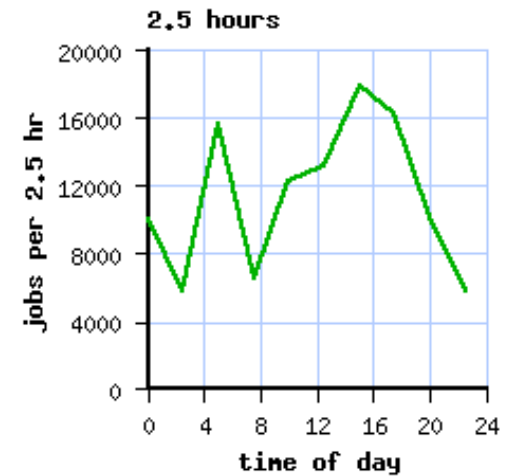
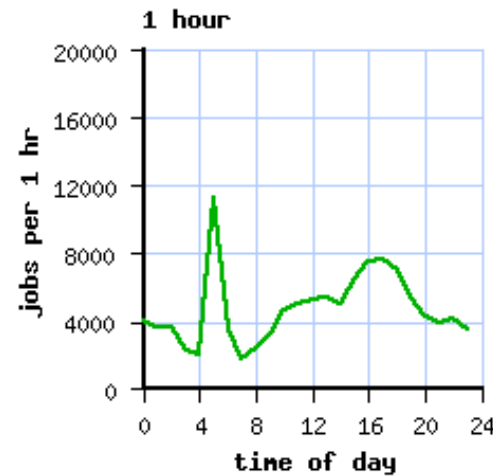
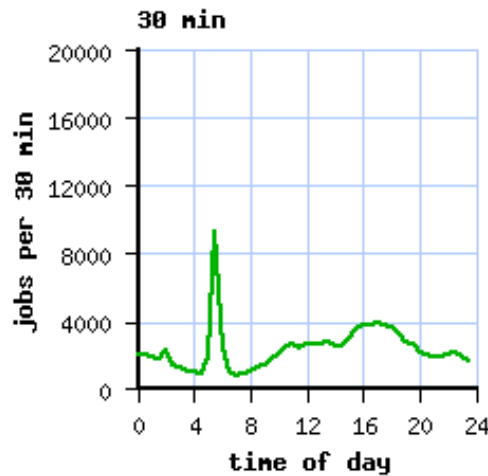
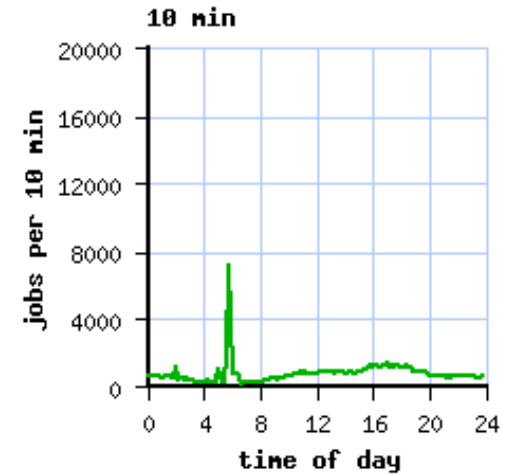
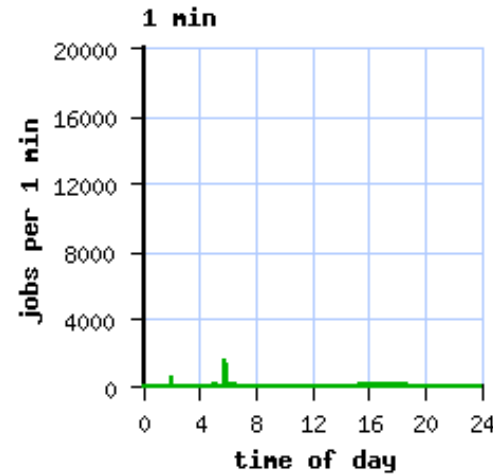
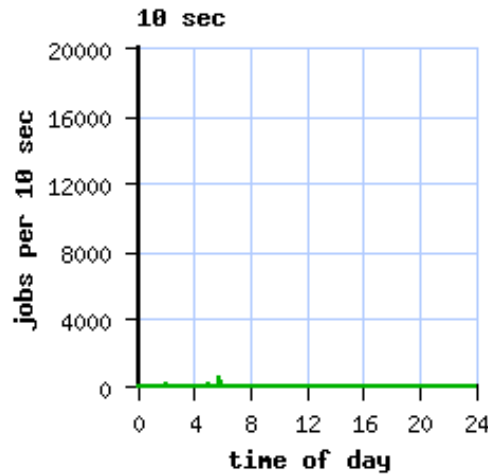


Stacking
Show
individual
components
and also
their sum



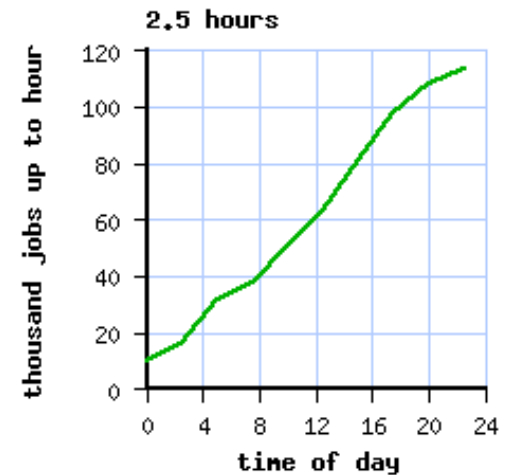
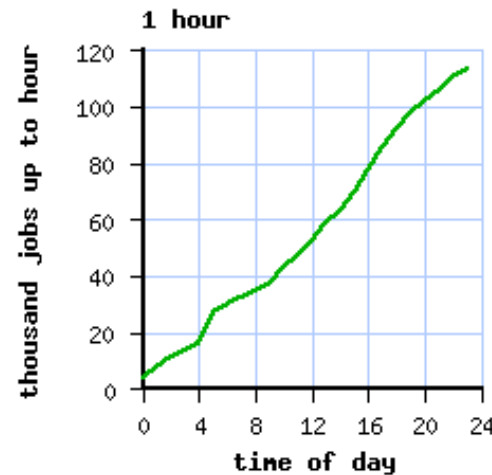
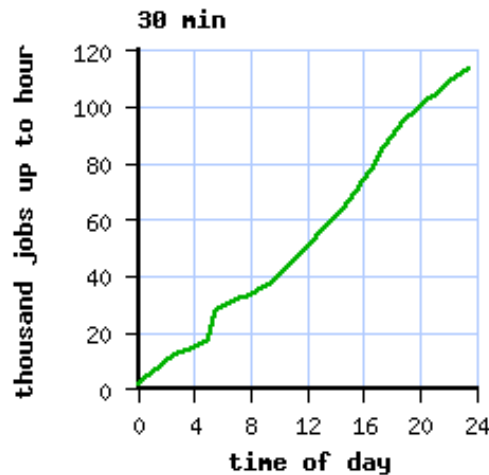
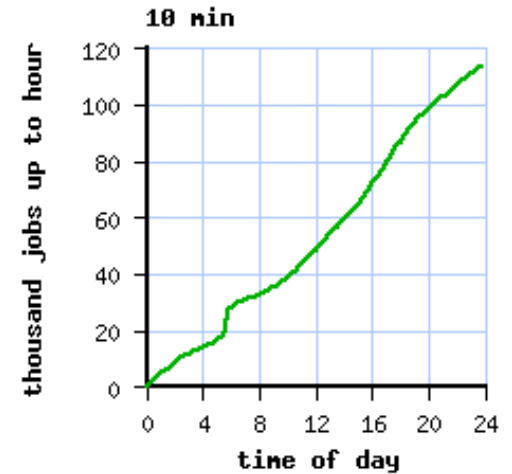
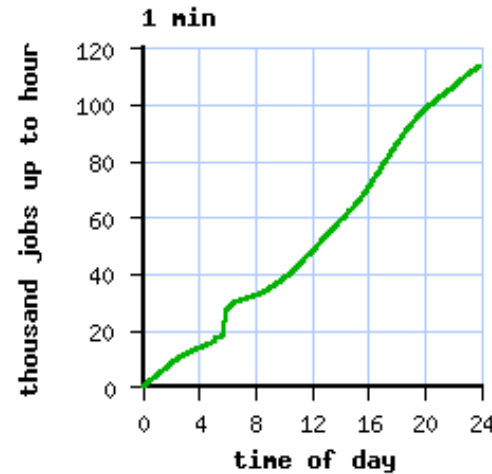
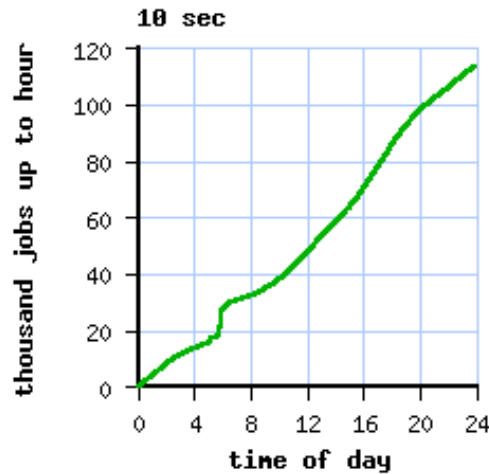
Histogram: number of samples between x and $x+\delta$

- Simplest display of a distribution
- Sensitive to bin size δ



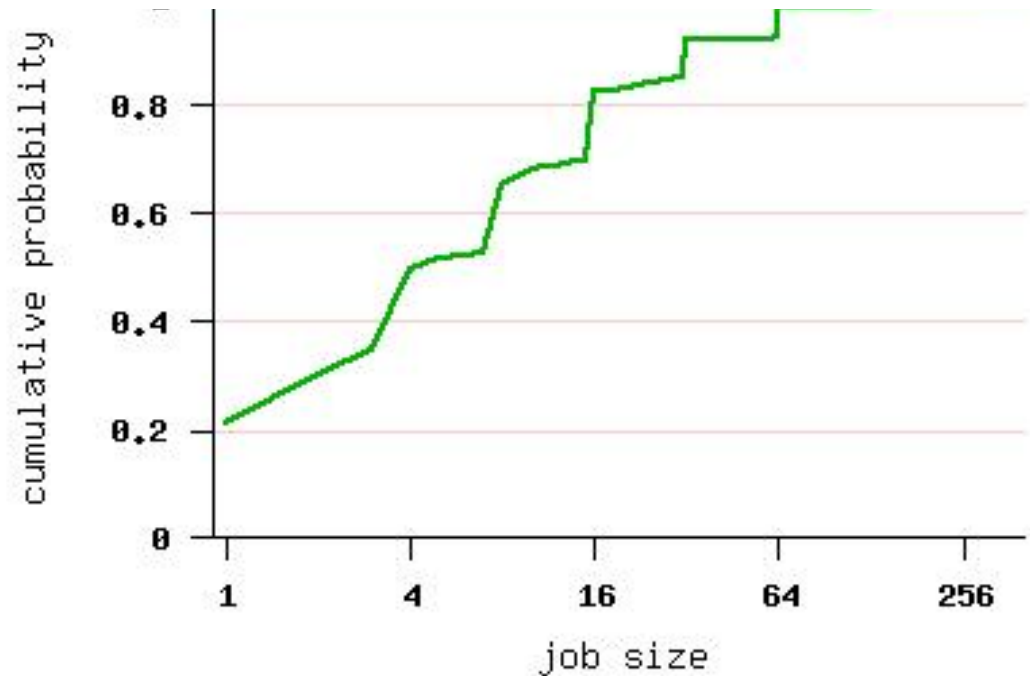
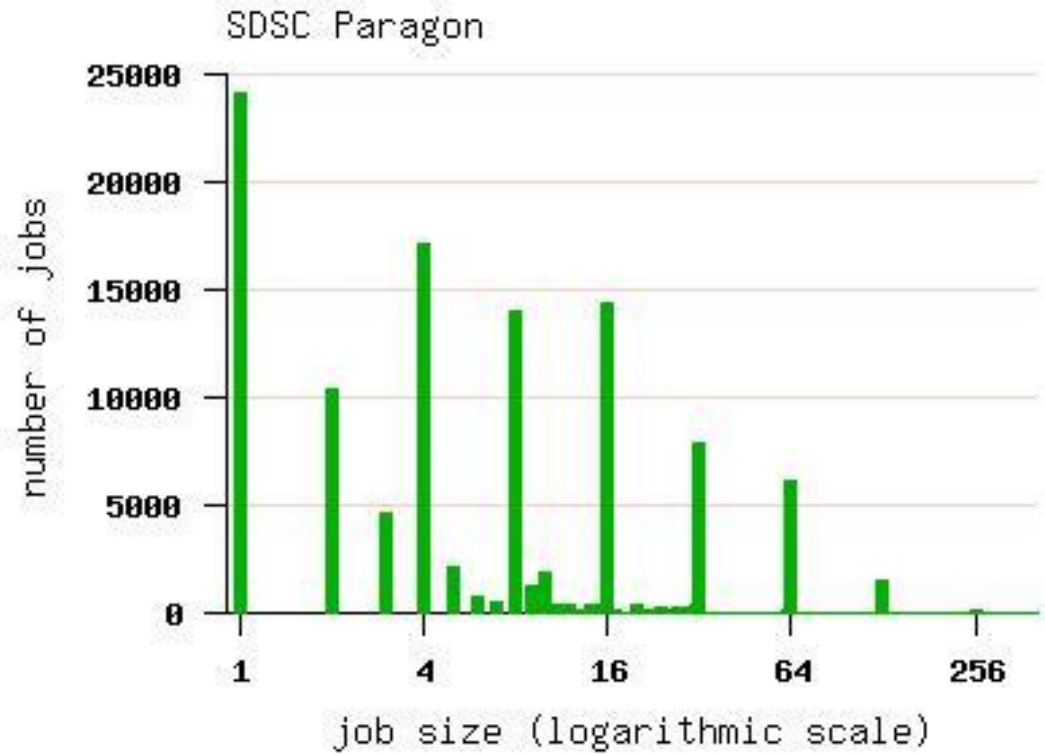
CDF: Probability of sample smaller than x

- Robust Alternative to histogram



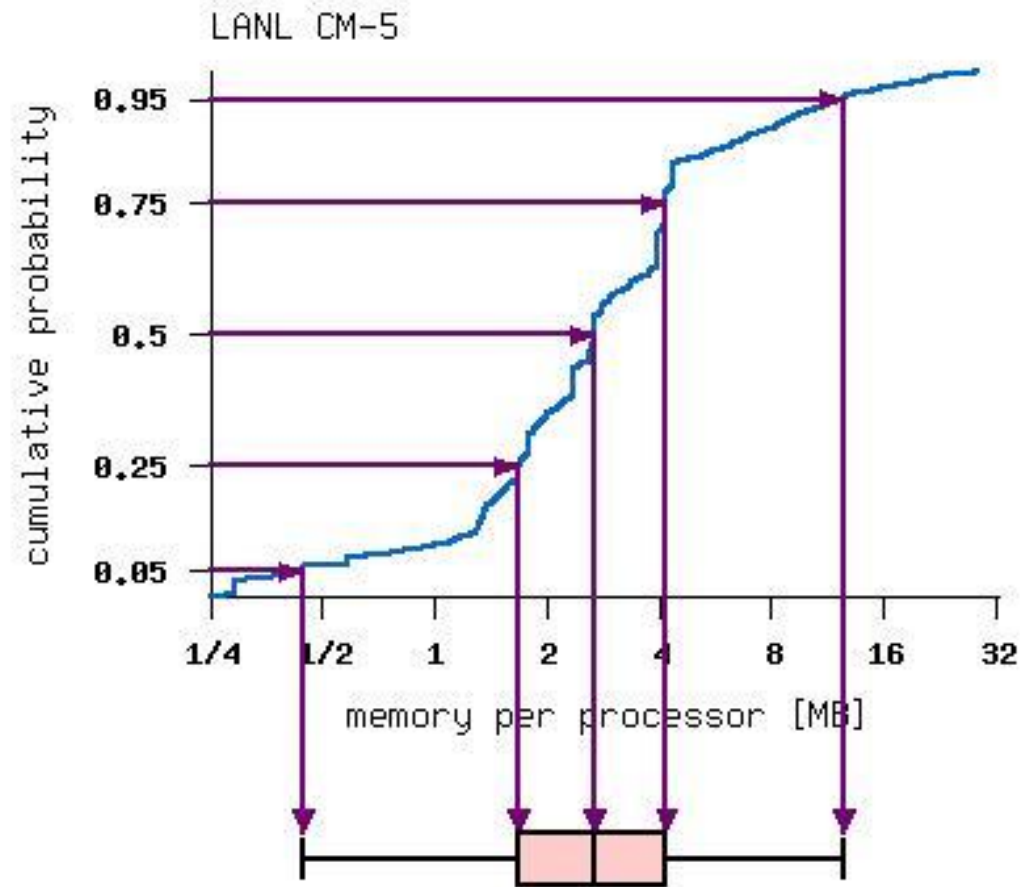
CDF

- Modes less prominent
- Easier to see where weight is concentrated

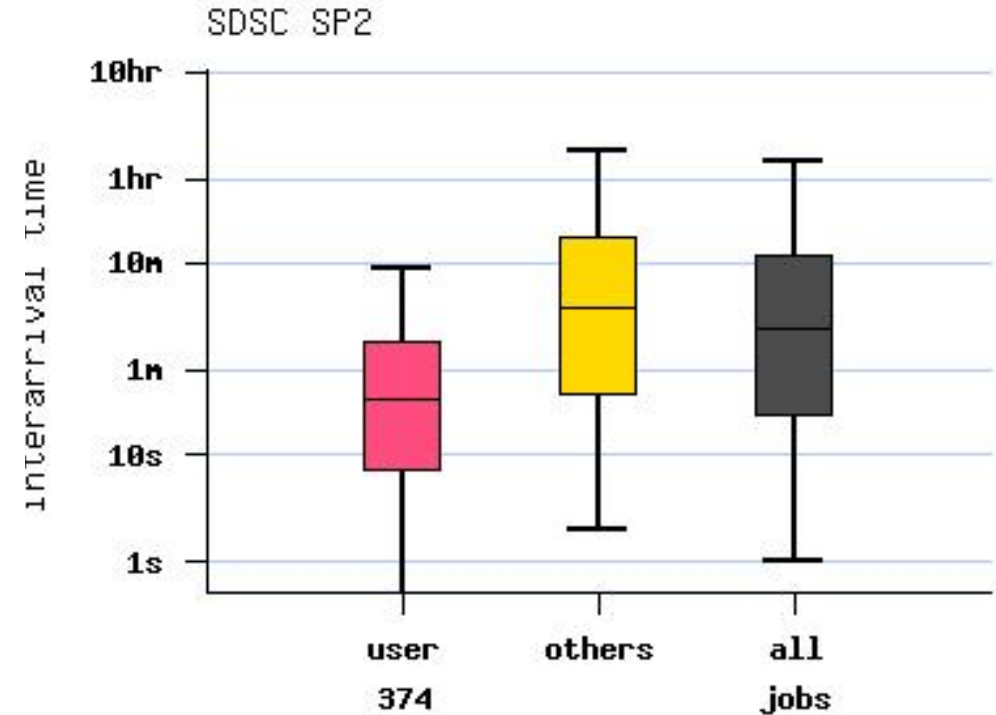
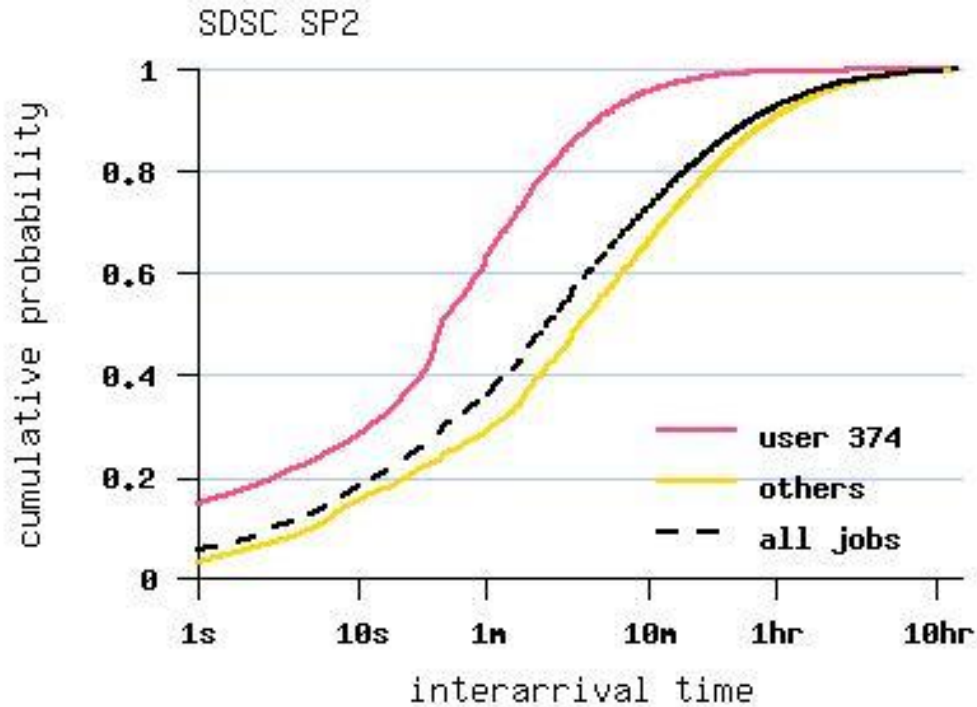
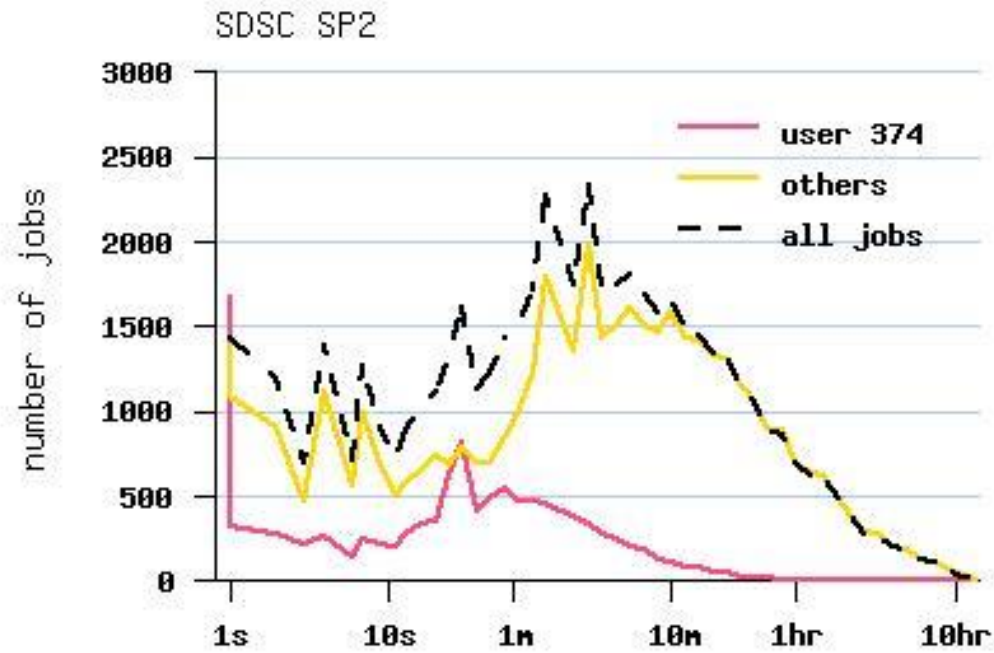


Box plot

- Summary of a distribution

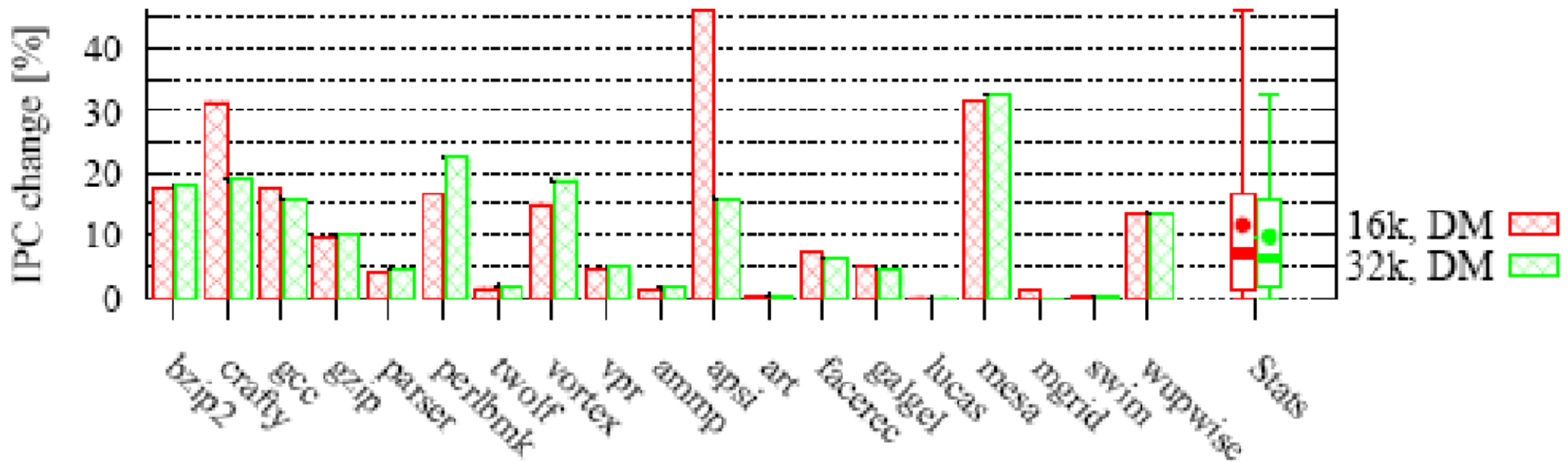


Comparison of distributions



Skewed distributions are common

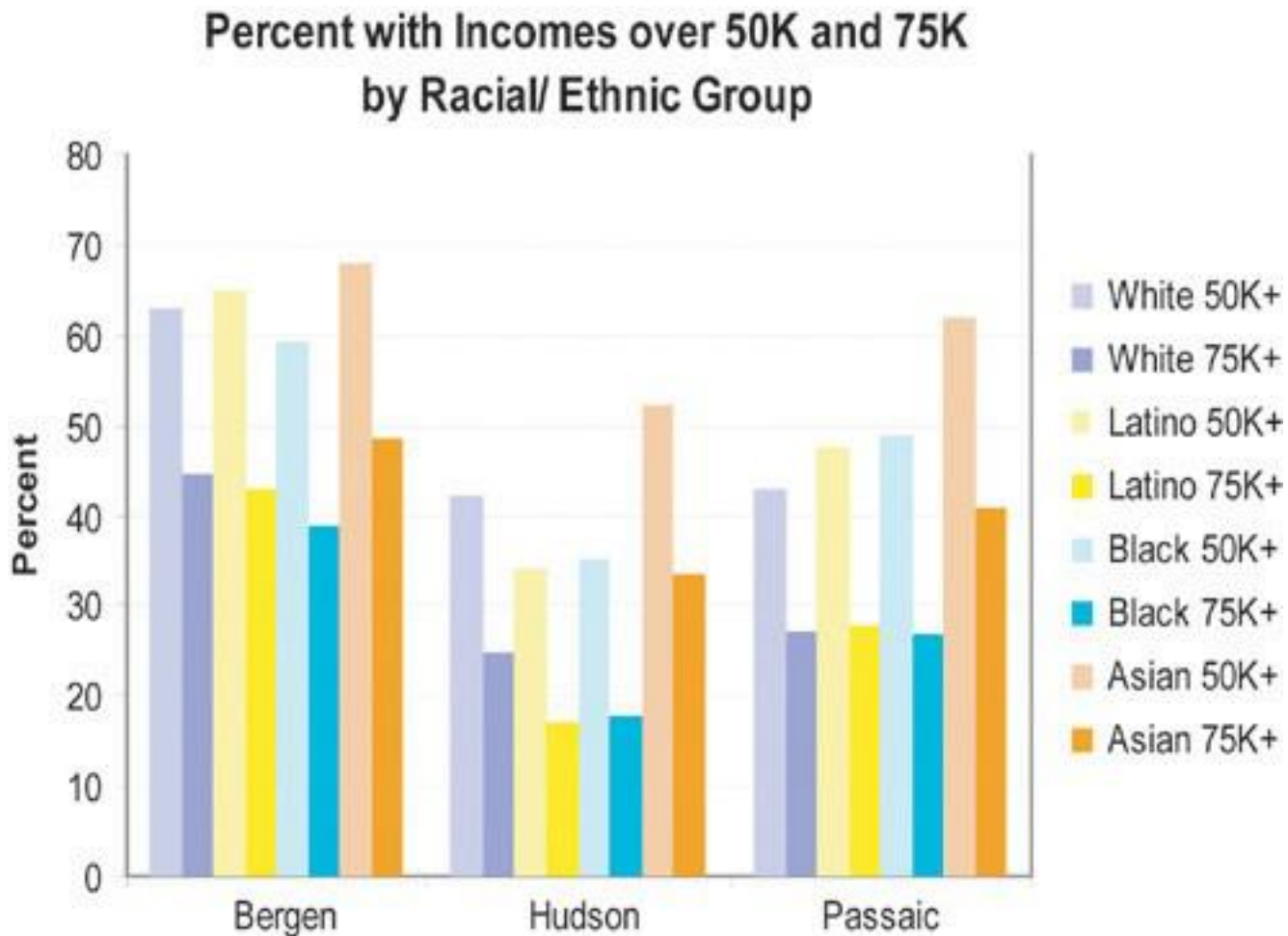
note difference between mean and median



Comparison in groups

3 locations, 4 ethnicities, 2 income levels

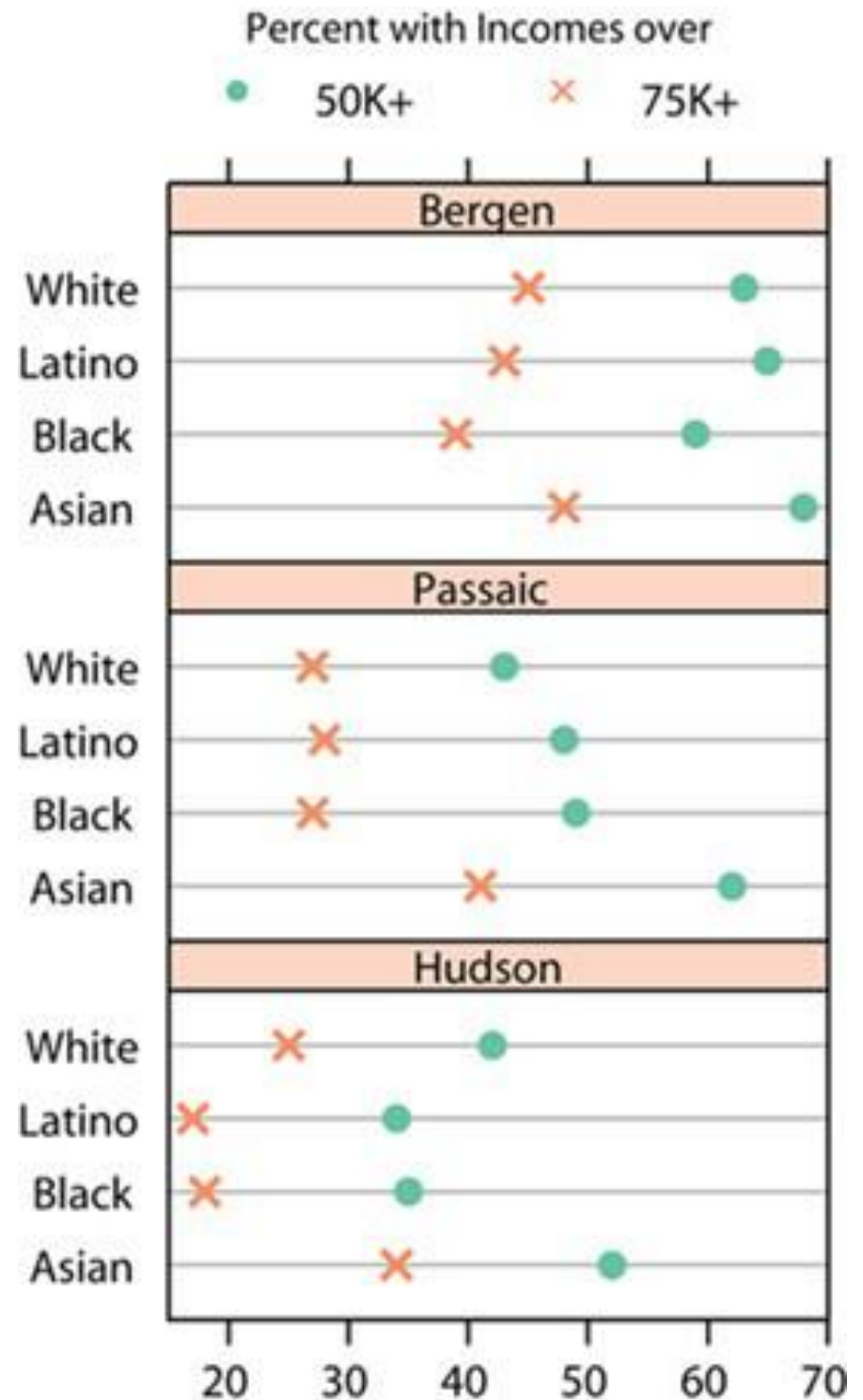
Clustered bars are generally hard to understand



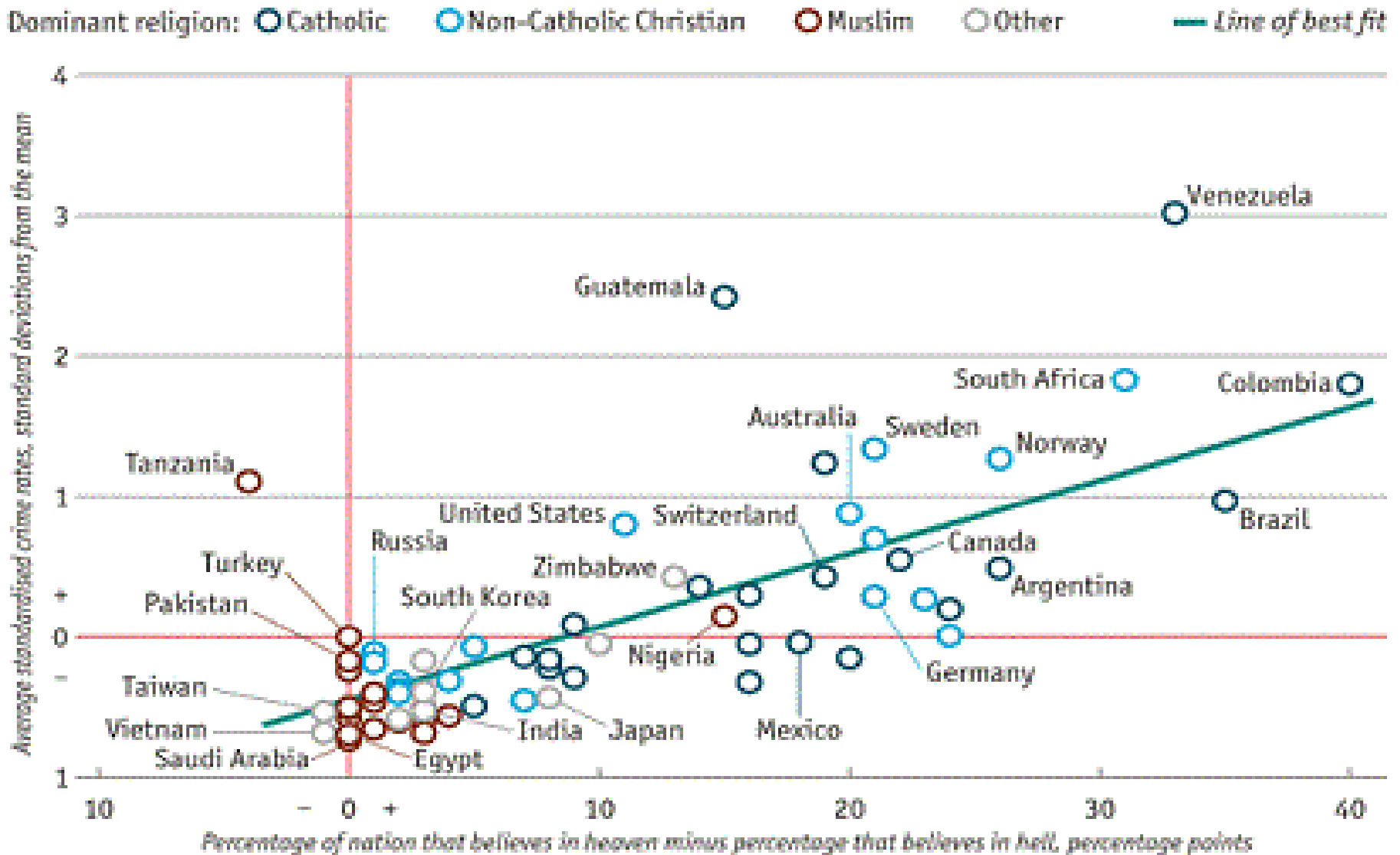
Comparison in groups

3 locations, 4 ethnicities,
2 income levels

Using dot plot makes it
clearer



Crime and religious beliefs*



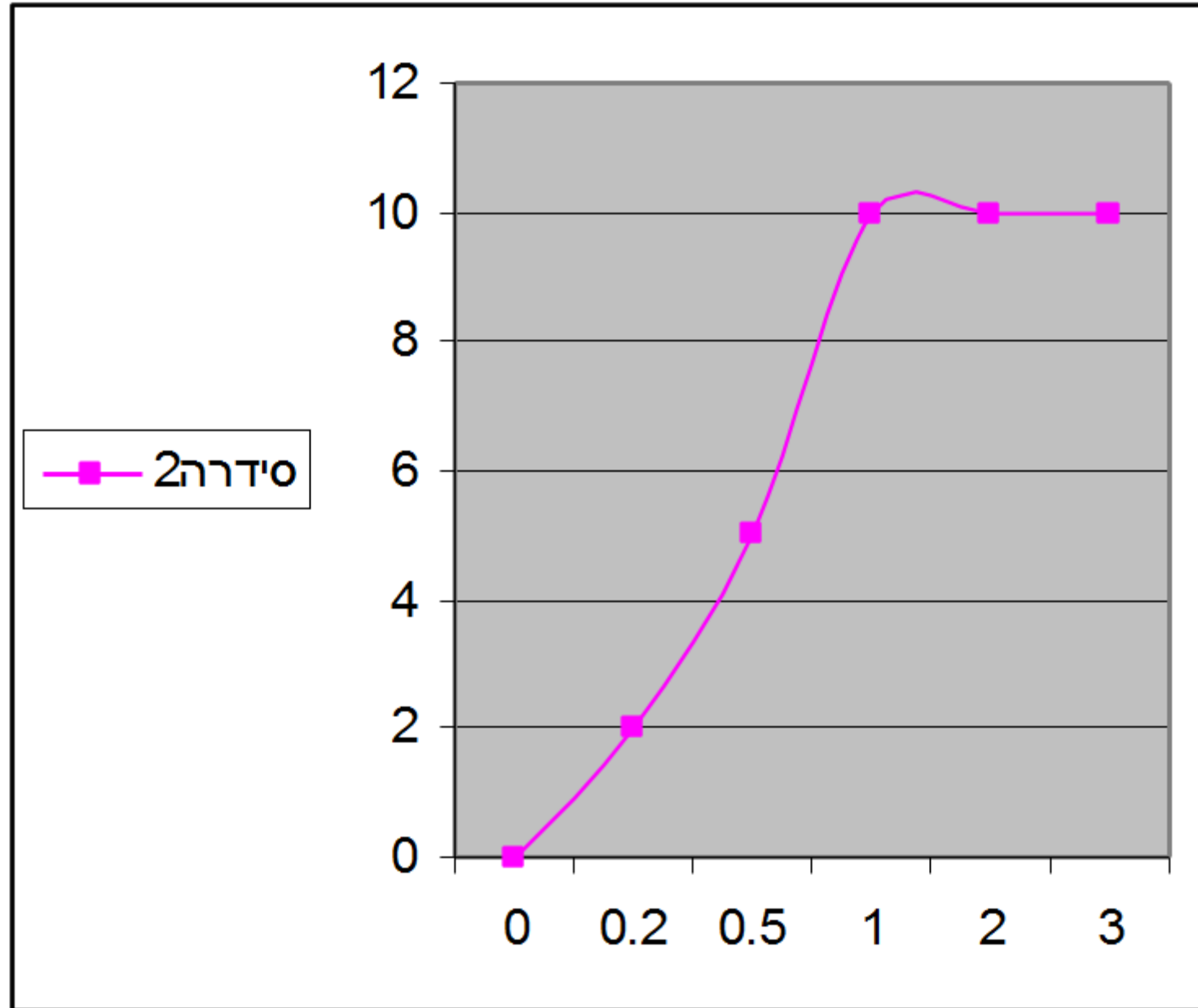
Source: "Divergent Effects of Beliefs in Heaven and Hell on National Crime Rates", by Azim Shariff and Mijke Rhemtulla, 2012

*Based on data from 67 countries, World Values and European Values surveys from 1981-2007 and UNODC 2003-10

Problems and Bad Examples

What's bad with this?

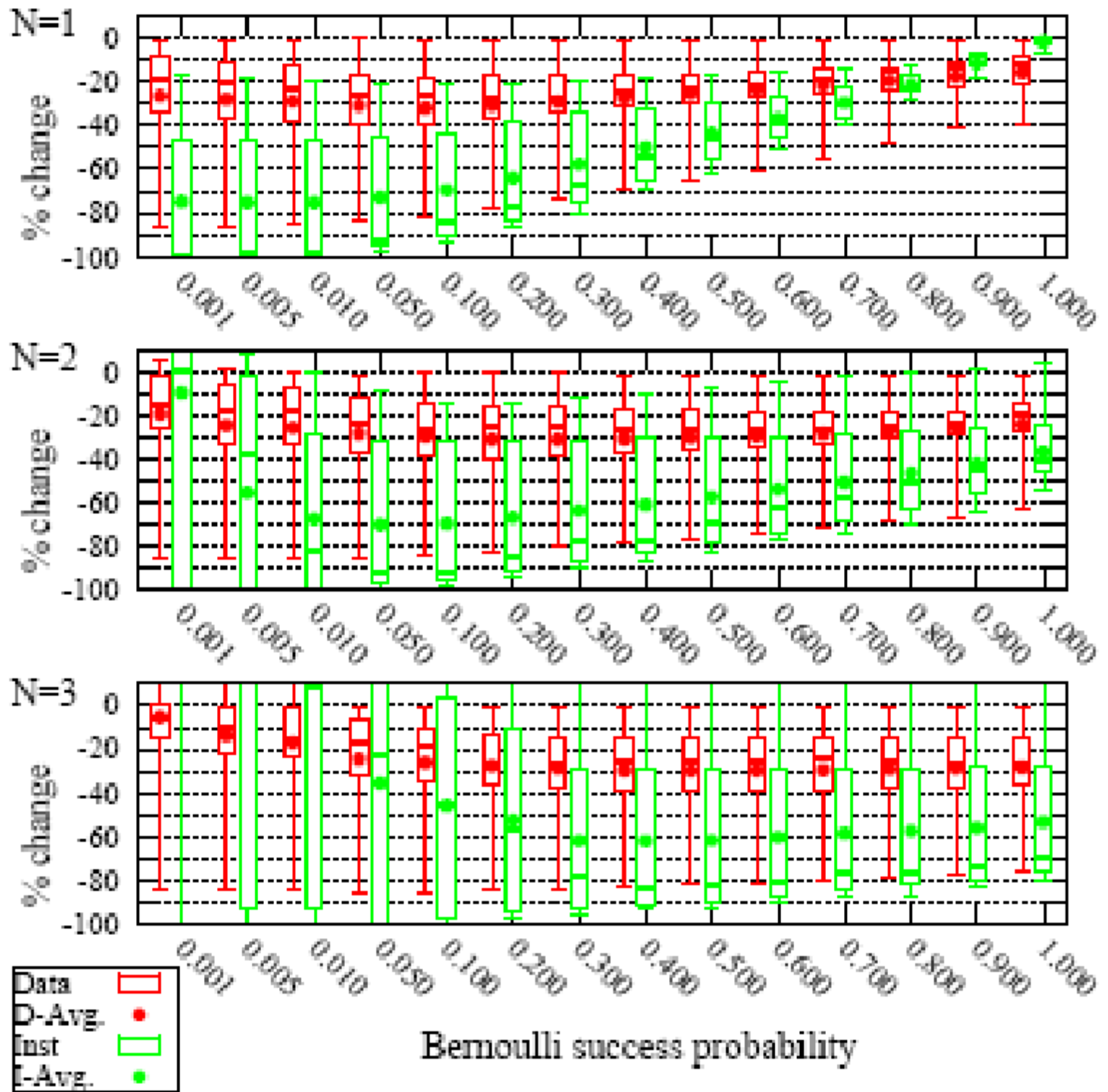
- X axis scale
- Color contrast
- Smooth line
- Legend not needed



Comparison of distribution of results for different experimental parameter values

Red/green not seen by colorblind

Grid too dark



Enlarging
glyph creates
appearance
of squared
growth

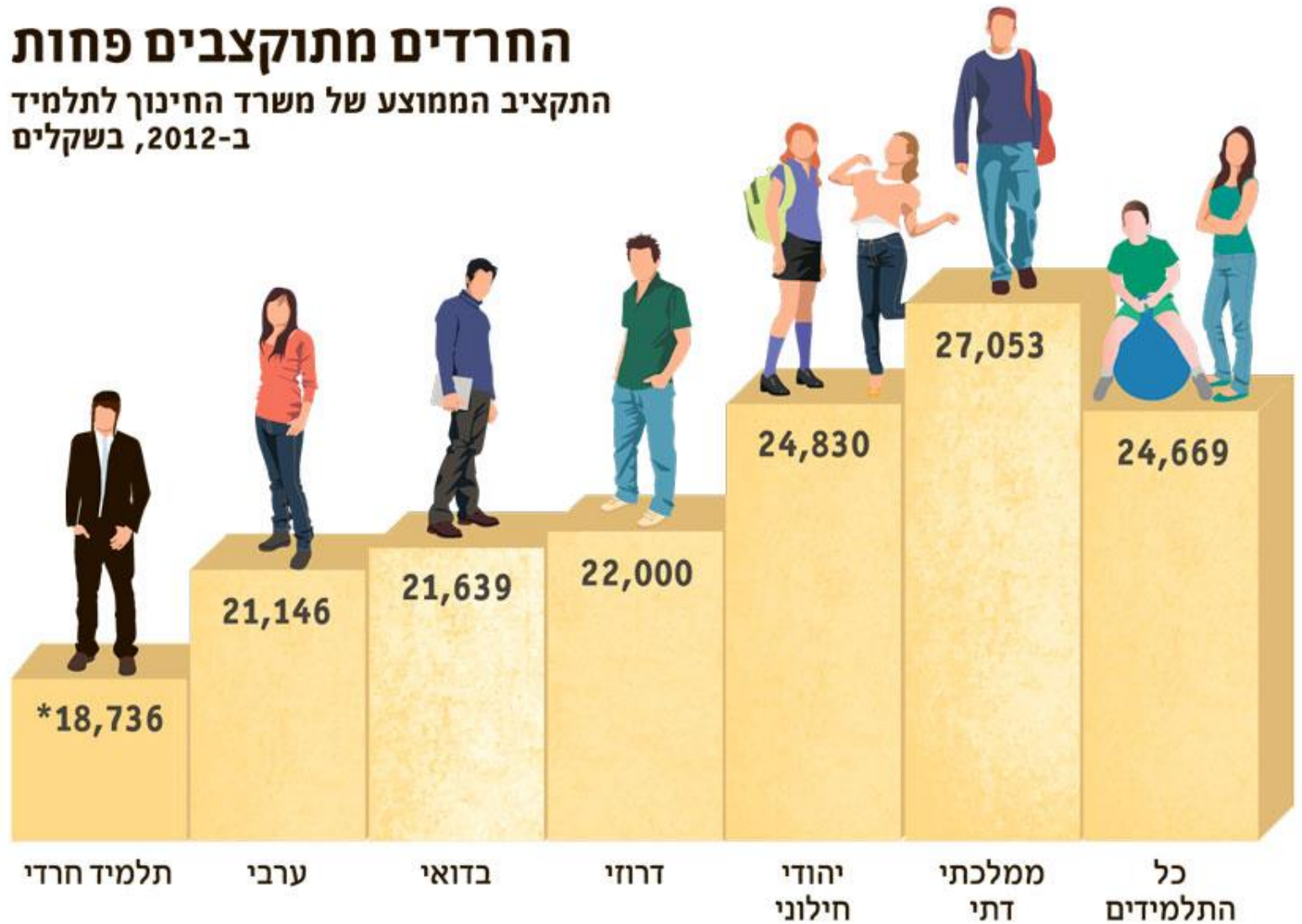


החרדים מתוקצבים פחות

התקציב הממוצע של משרד החינוך לתלמיד ב-2012, בשקלים

Start from 0 to avoid misleading

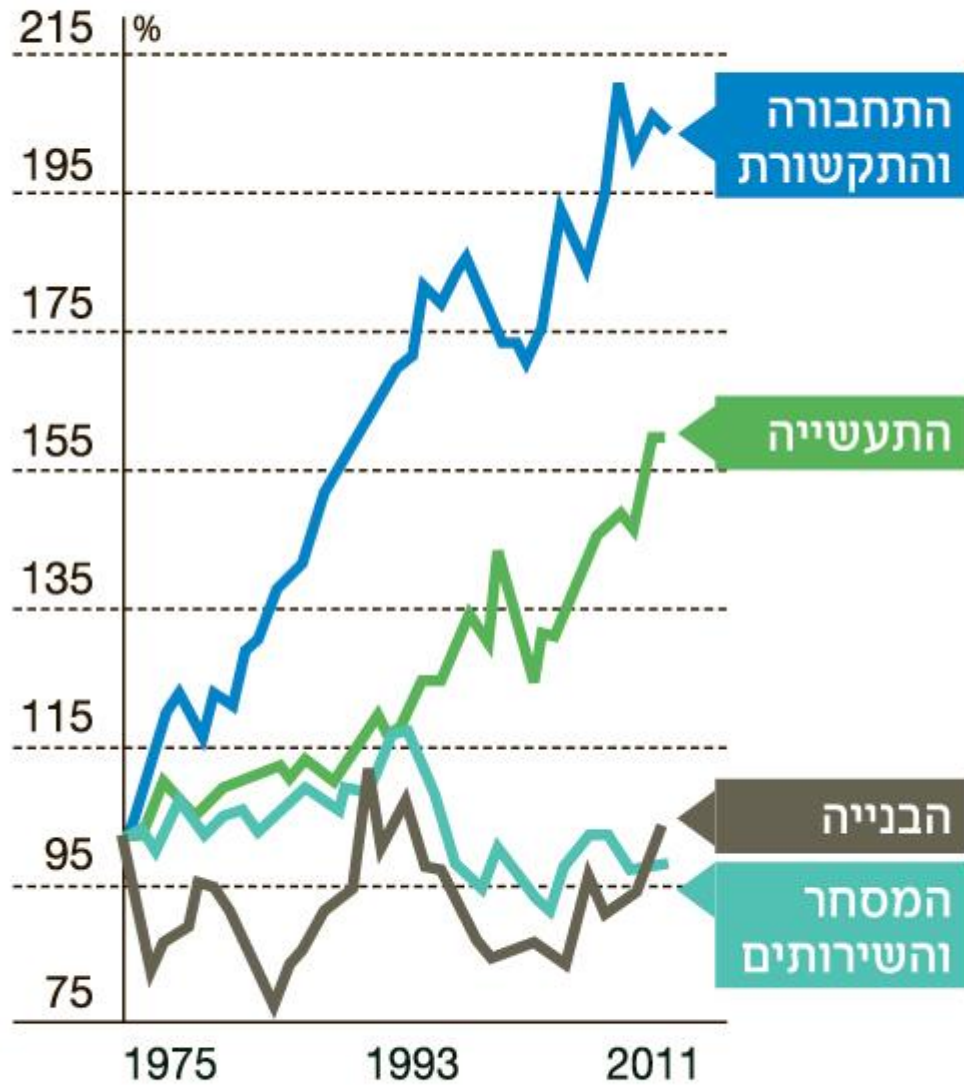
Average as line, not bar



מקור: עיבוד לנתוני משרד החינוך

*מתייחס בעיקר לסמינרים ולכמה ישיבות המגישות לבגרות

הפריון בענפי המשק העיקריים



מקור: הנספח הסטטיסטי לדו"ח בנק ישראל, 2012

TheMarker

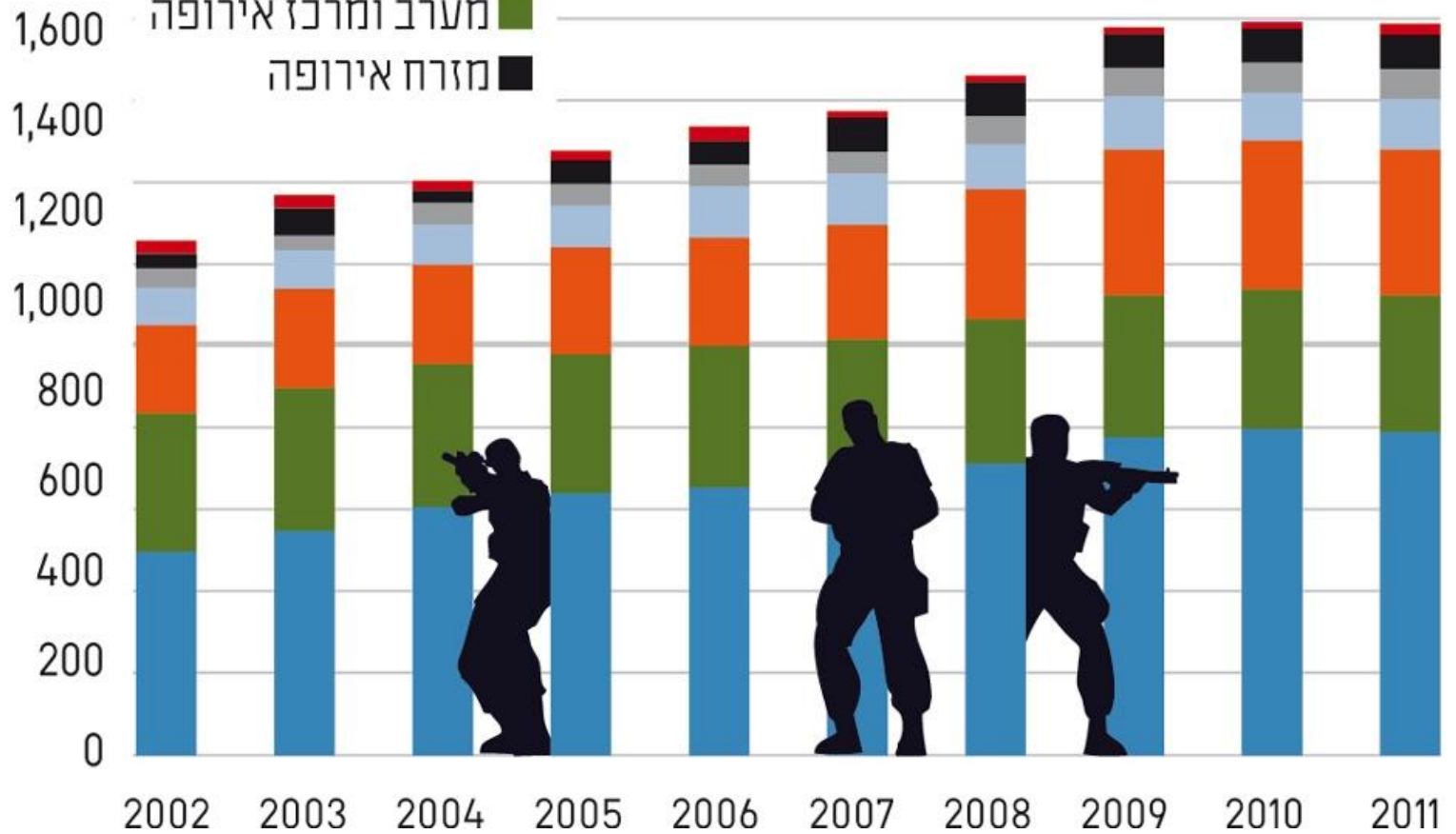
Y stubs don't match data

X stubs missing, strange choice

מלחמה זה עסק טוב

ההוצאות הביטחוניות בעולם,
במיליארדי דולרים (2010)

- צפון אמריקה
- המזרח התיכון
- אמריקה הלטינית
- אסיה ואוקיאניה
- אפריקה
- מערב ומרכז אירופה
- מזרח אירופה



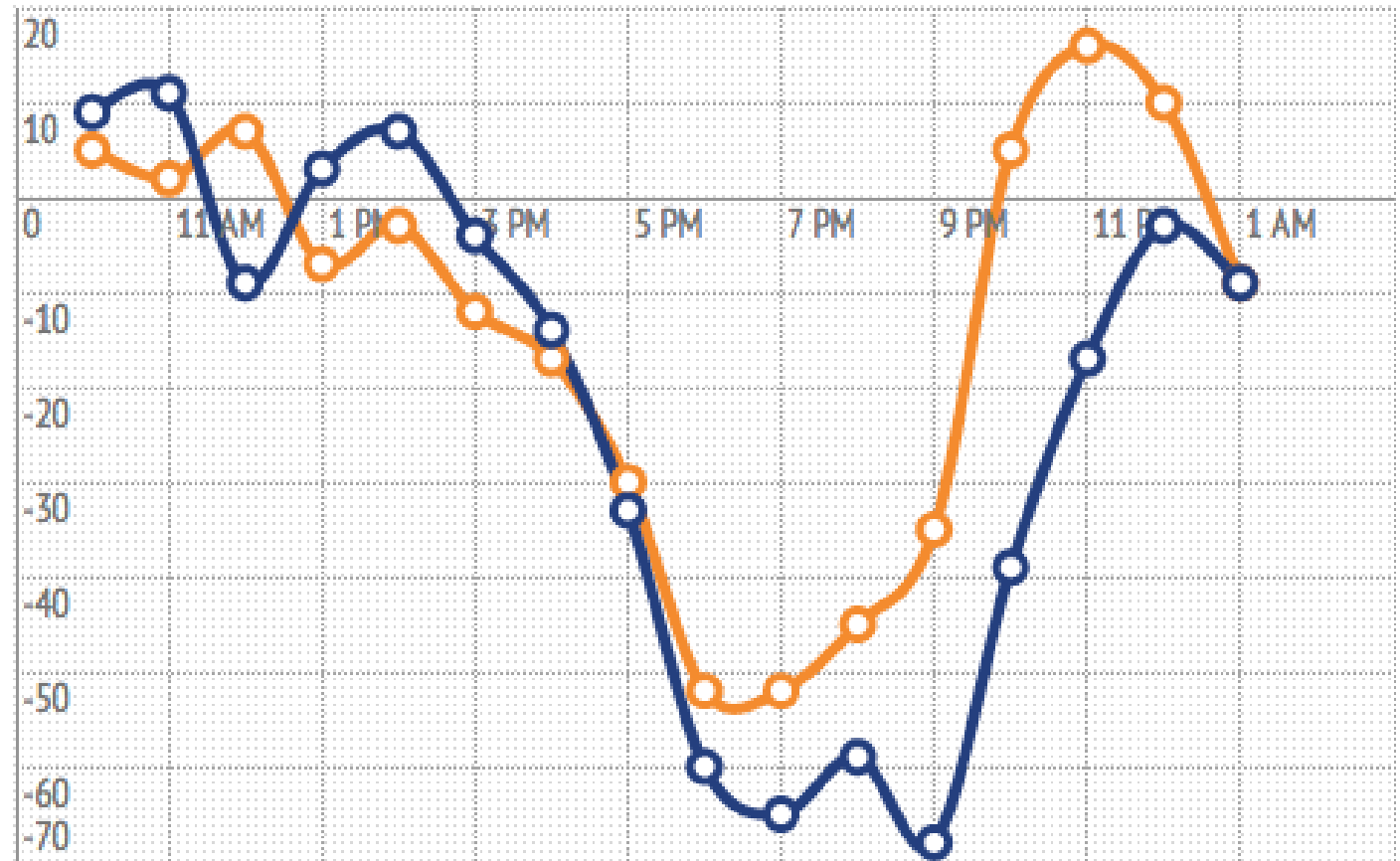
Order of legend

Stubs don't match grid lines

Pornhub - Traffic Change During Super Bowl - Seattle and Denver

X axis
placement

...Y axis is
Change in
%?



Denver, CO Seattle, WA

good for
colorblind