

ARC<sup>2</sup>S Group

Applied Research on Computational Complex Systems

# Design and Evaluation

Prof. Giancarlo **Ruffo**

“Analisi e Visualizzazione di Reti Complesse” (9 credits)

Laurea Magistrale in **Informatica**

Università degli Studi di Torino

A.A. 2018/19

@giaruffo



# **GRAPHICAL REDESIGN**

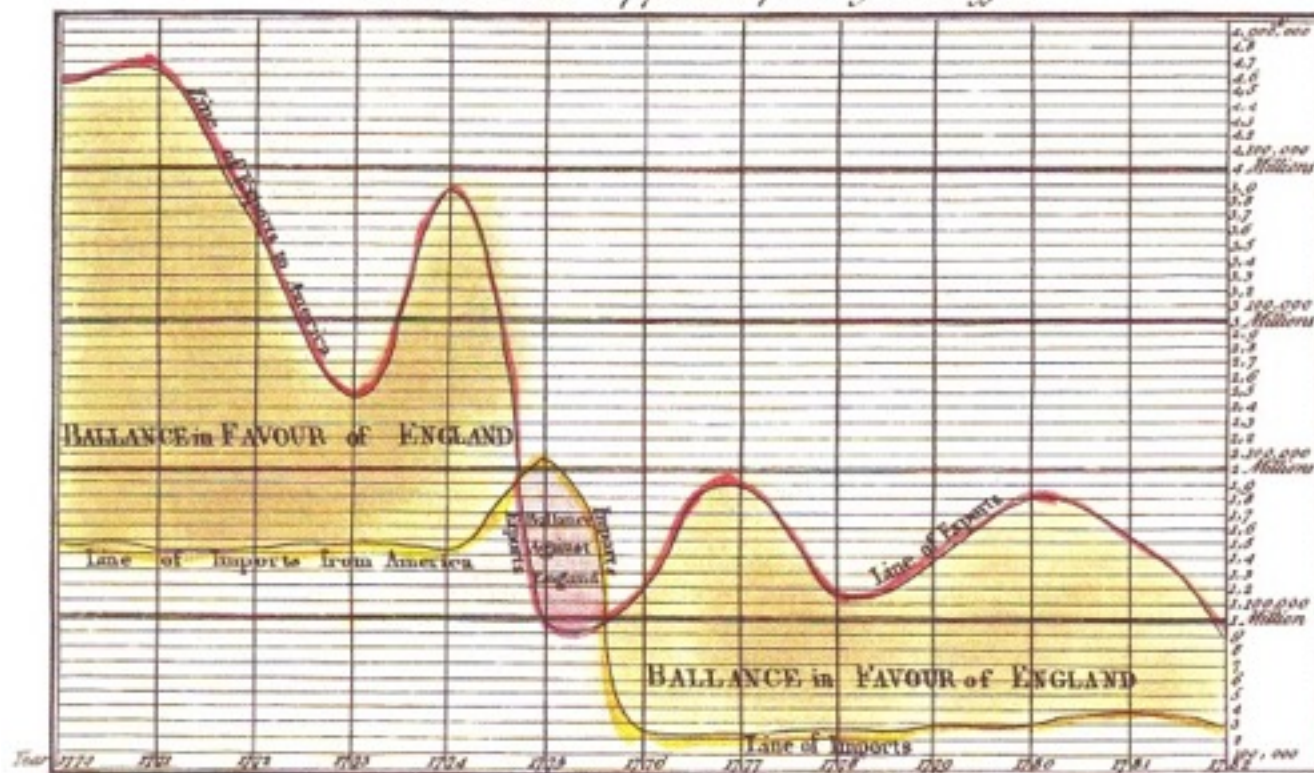
VDQI Chapter 4

# Graphical Redesign

“Above all else, show the data.”

# VDQI Example (p91)

*CHART of IMPORTS and EXPORTS of ENGLAND to and from all NORTH AMERICA  
From the Year 1770 to 1782 by W. Playfair*



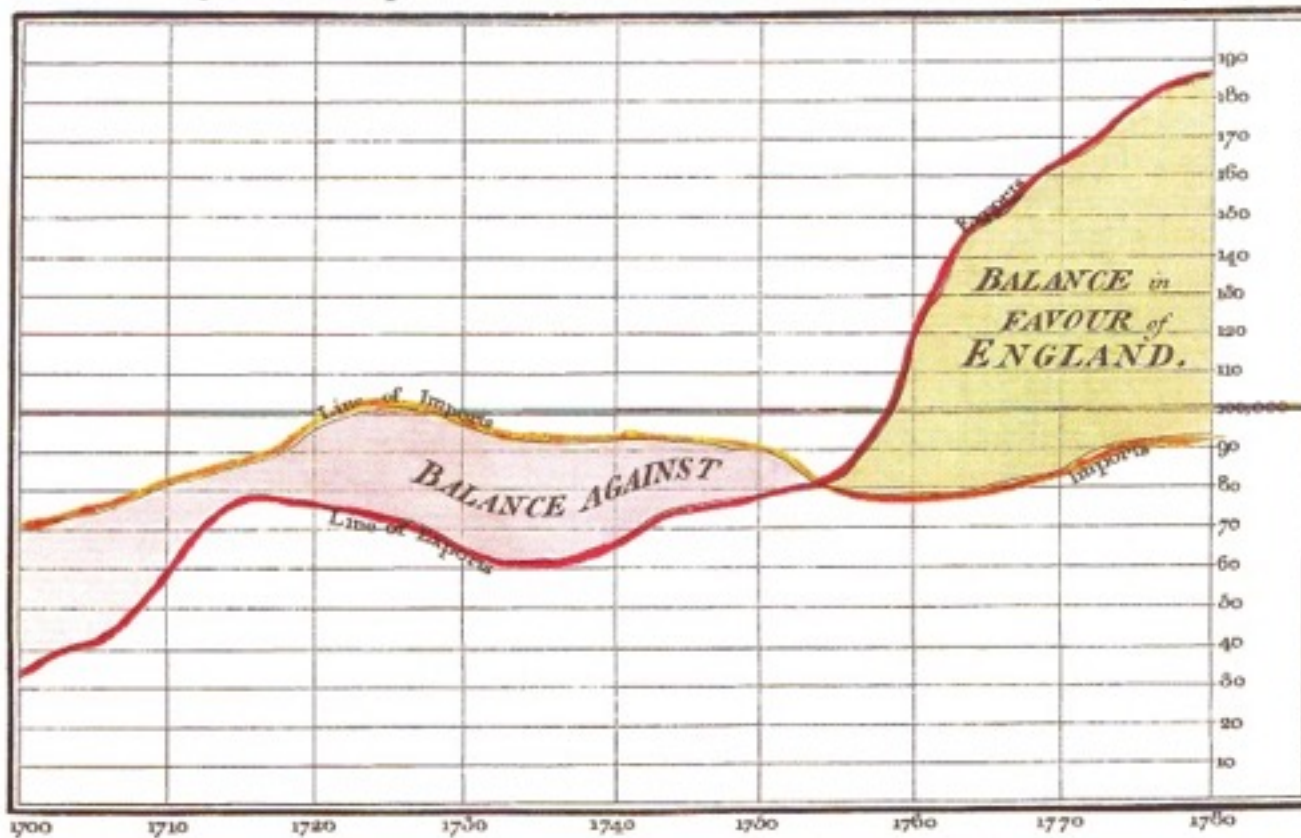
*The Bottom Line is divided into Years the right-hand Line into HUNDRED THOUSAND POUNDS*

*J. Smith Sculp.*

*Published as the Act direct 10<sup>th</sup> Aug<sup>r</sup> 1785.*

# VDQI Example (p91)

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.*

*Published as the Act stands, 17 May 1786, by W. P. Fryer.*

*Printed and Sold by S. D. Fryer, London.*

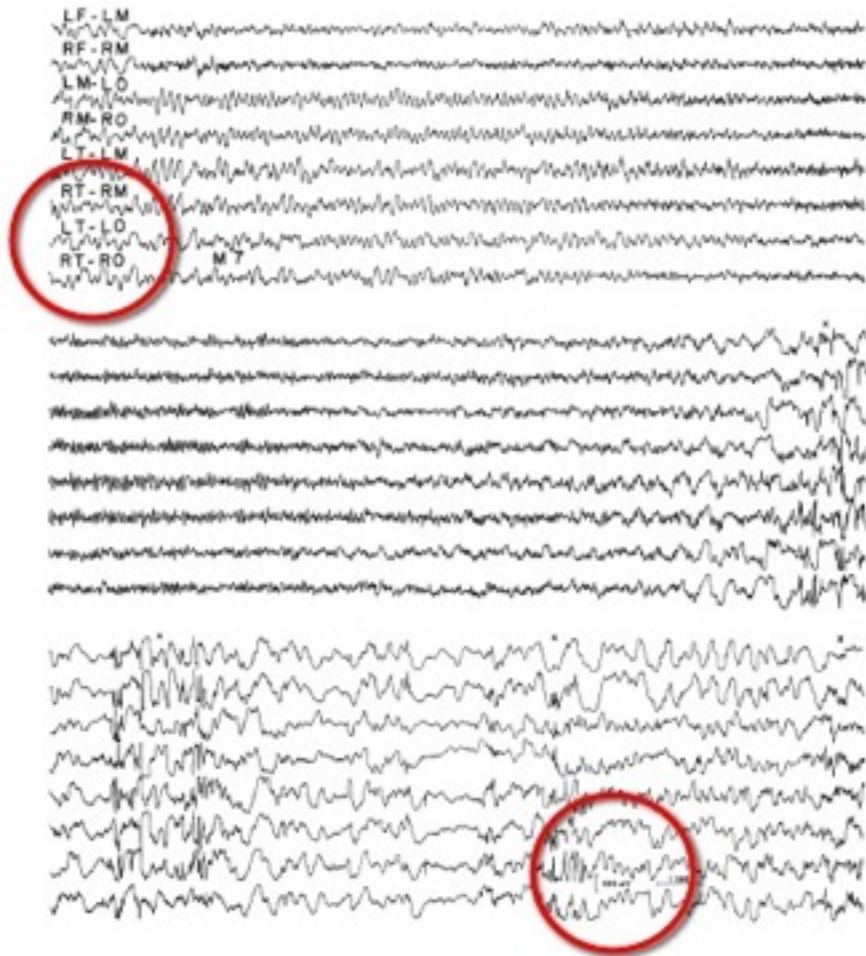
# Data-Ink Ratio

$$\text{Data-Ink Ratio} = \frac{\text{data-ink}}{\text{total ink used in graphic}}$$

= proportion of a graphic's ink devoted to the non-redundant display of data-information

= 1.0 – proportion of a graphic that can be erased without loss of data-information

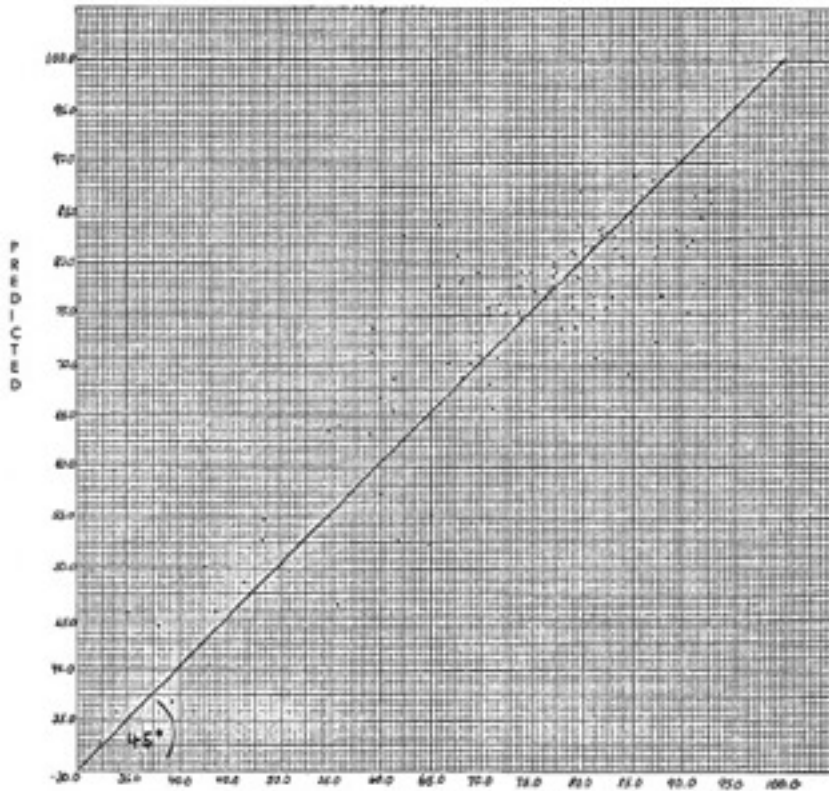
# VDQI Example (p93)



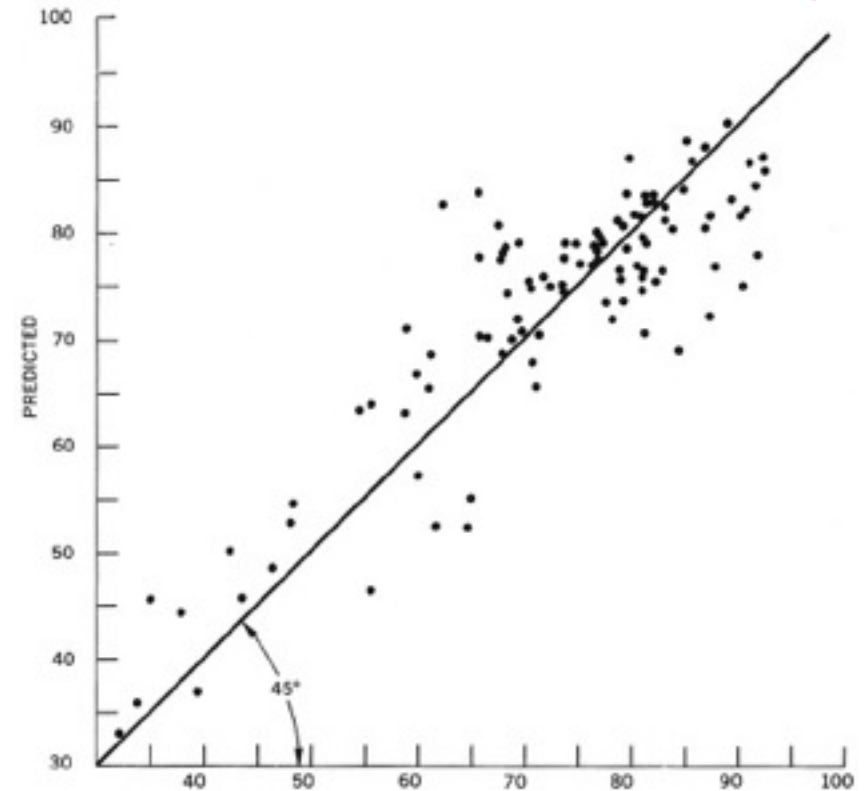
- Nothing can be erased without losing information

# VDQI Example (p94)

data-ink:low



data-ink:high





# Graphic Redesign

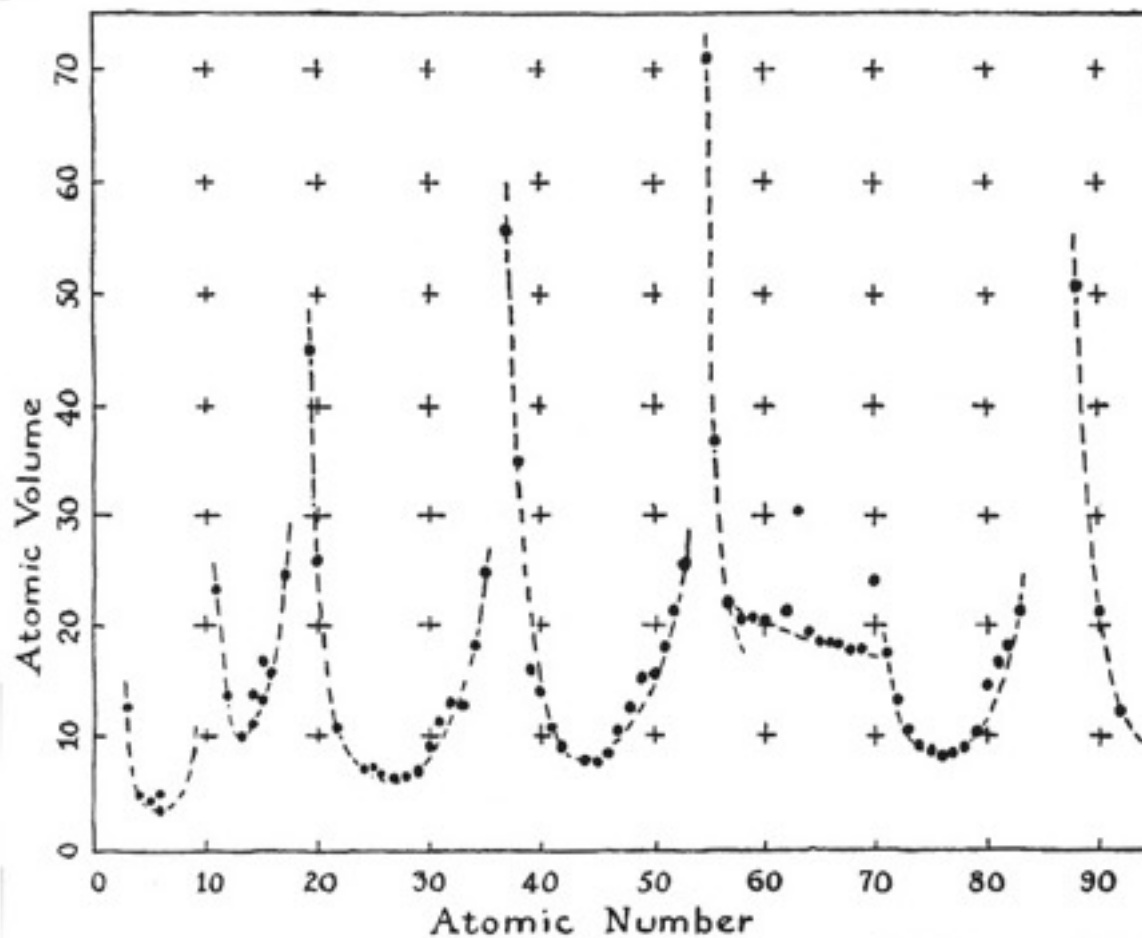
- Maximize the data-ink ratio, within reason.
- Erase non-data-ink, within reason.
- Erase redundant data-ink, within reason.
- Revise and edit.

# Graphic Redesign

- Maximize the data-ink ratio, within reason.
- Erase non-data-ink, within reason.
- Erase redundant data-ink, within reason.
- Revise and edit.
- Some ornamentation, axis labels, etc. is okay.
- Some redundancy is useful.

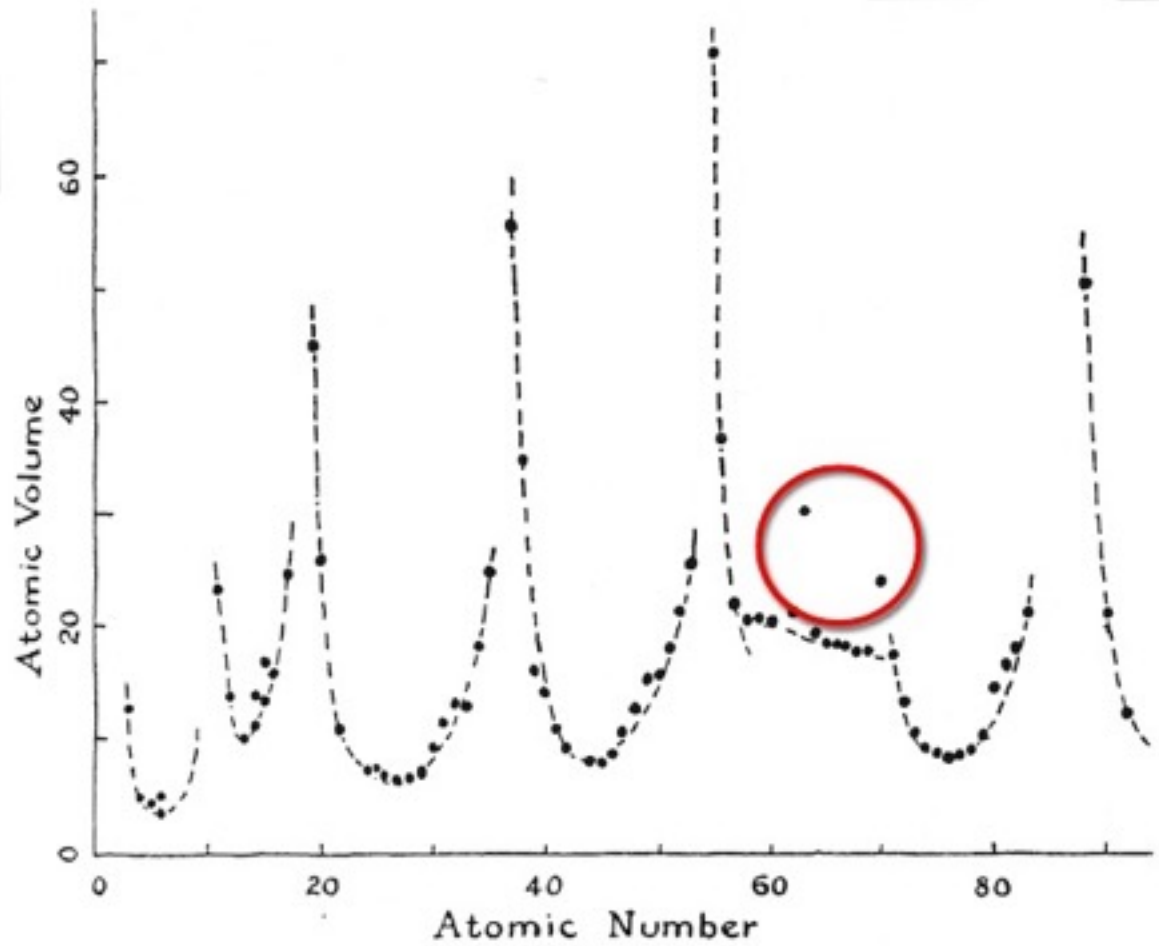
# VDQI Example (p102)

data-ink:  
low



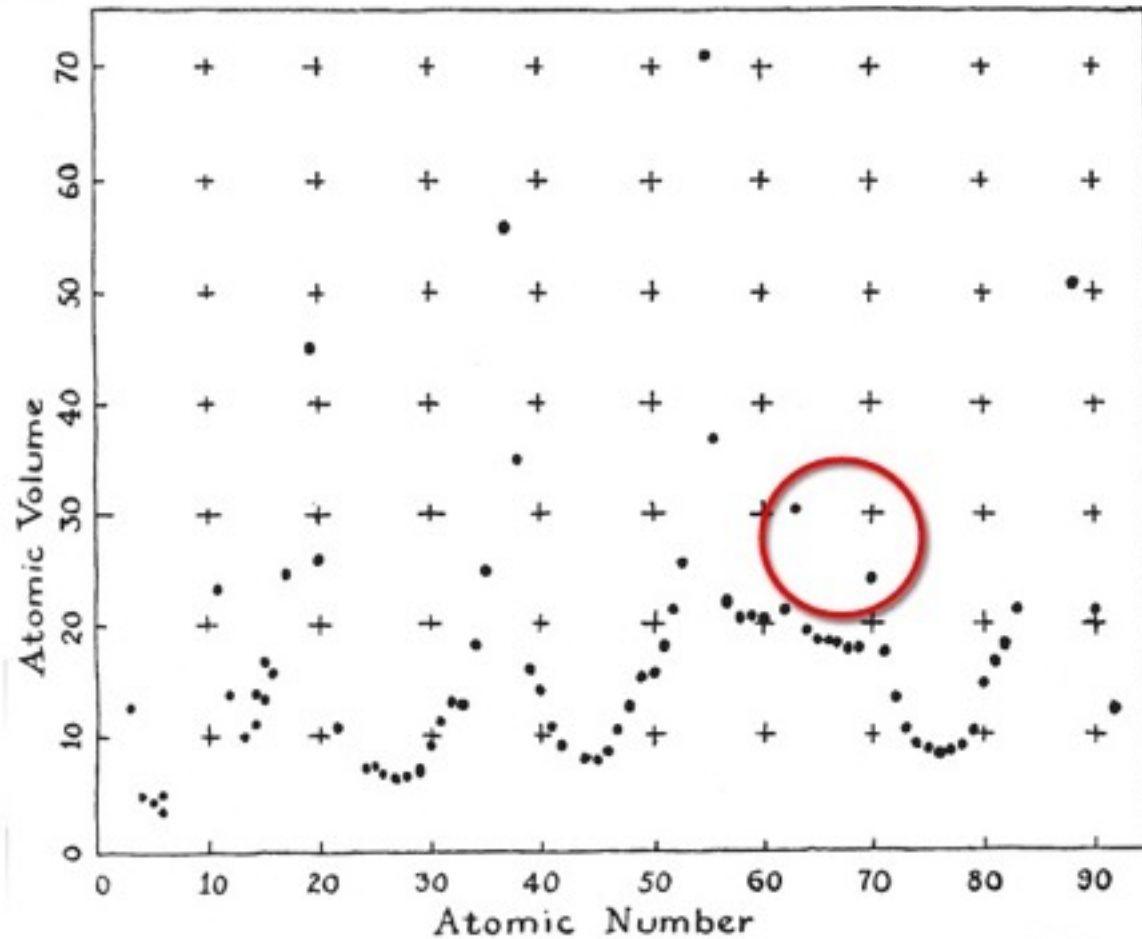
# VDQI Example (p102)

remove  
grid?



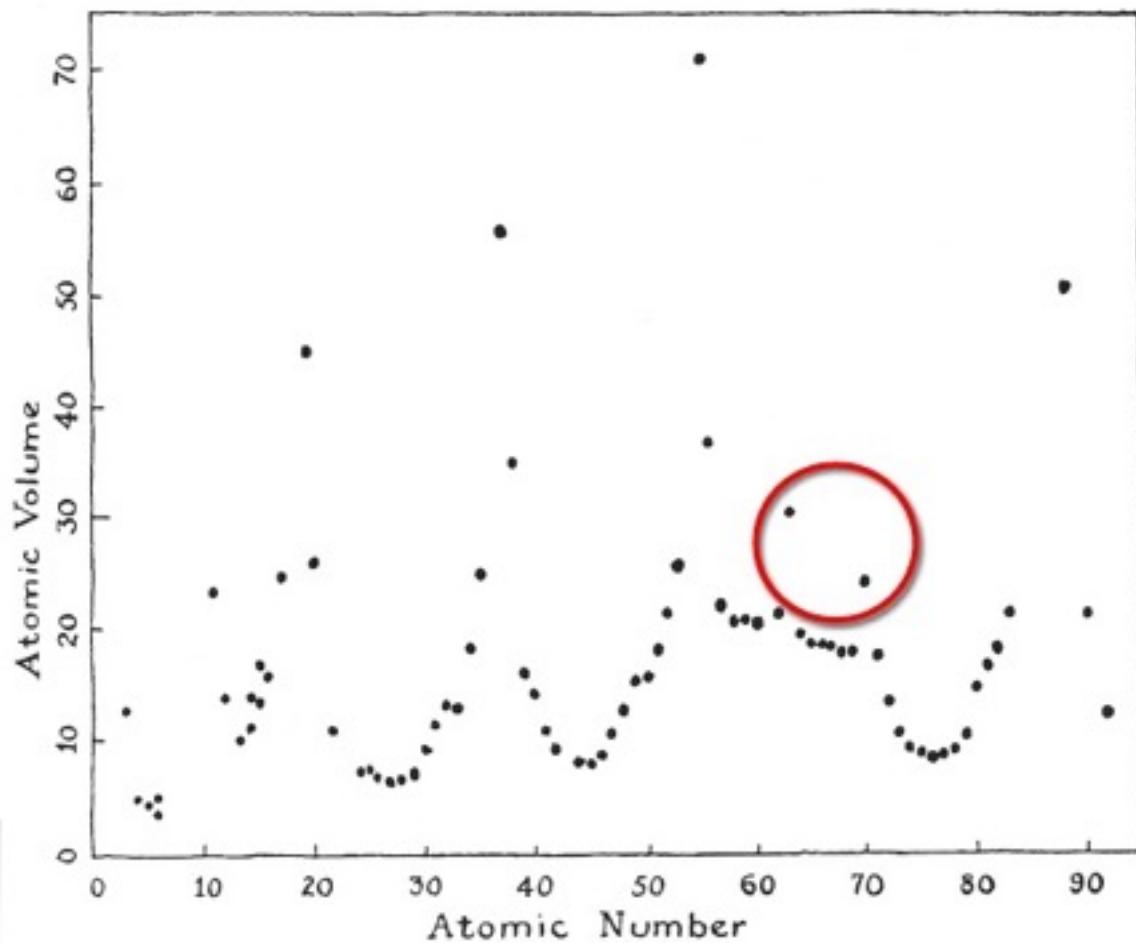
# VDQI Example (p102)

remove  
lines?



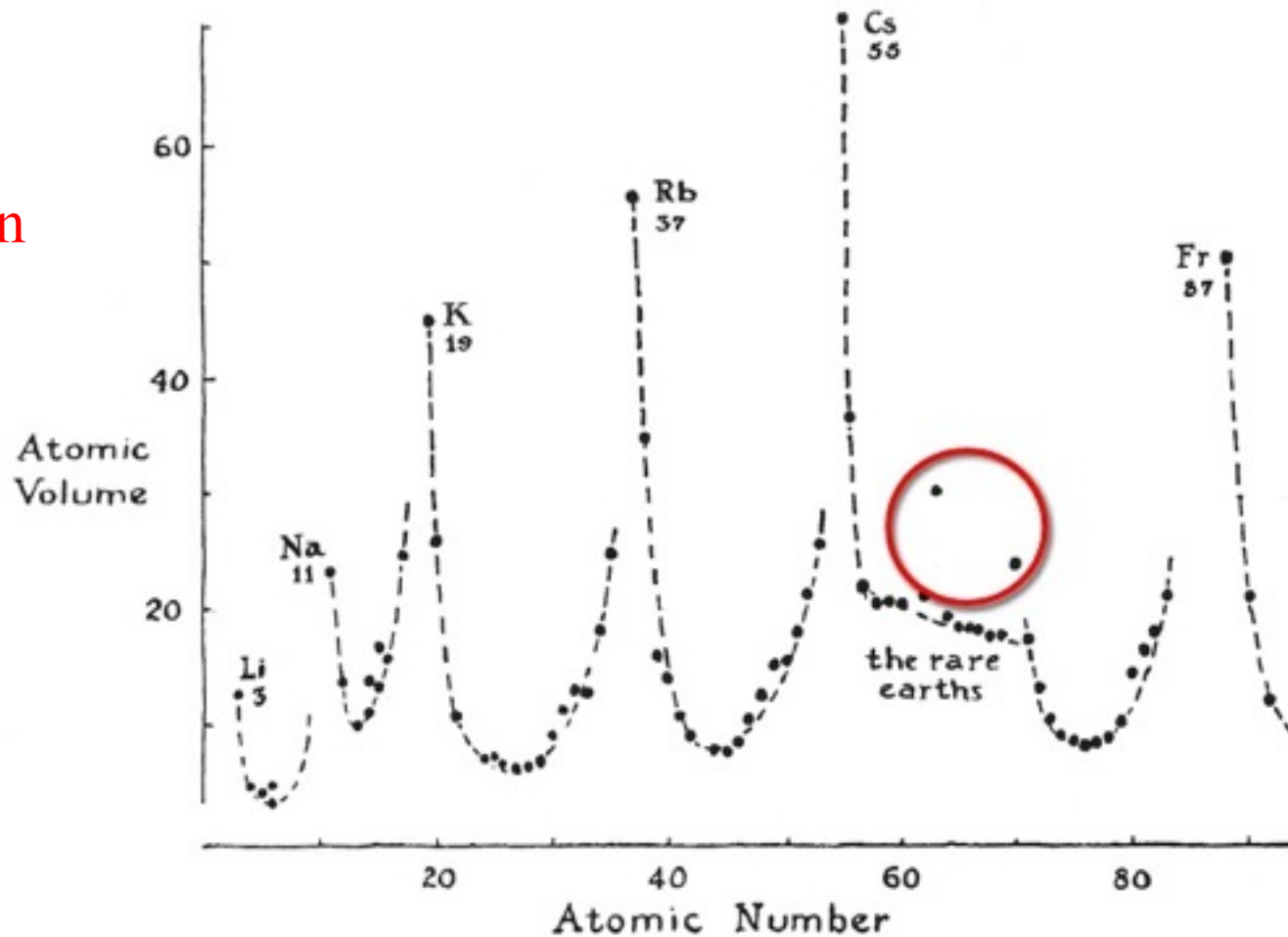
# VDQI Example (p102)

remove  
both?



# VDQI Example (p102)

final  
version



# **DATA DENSITY**

VDQI Chapter 8

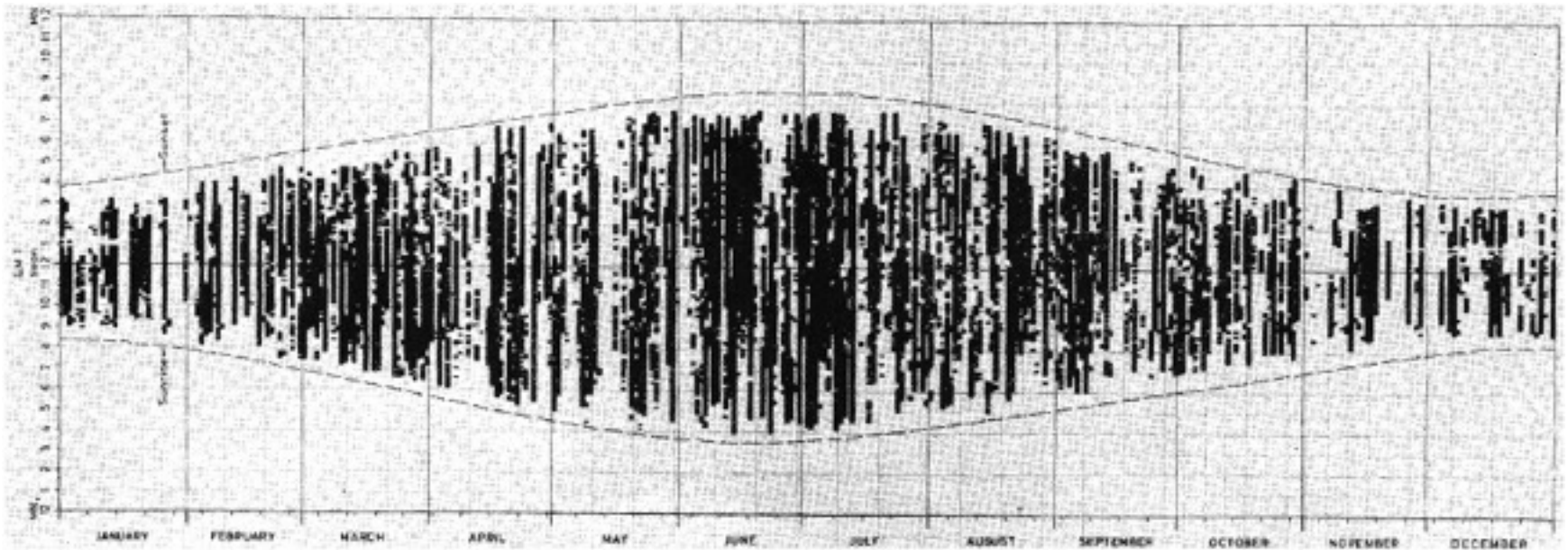


# Data Density

$$\text{Data Density} = \frac{\text{number of entries in data matrix}}{\text{area of data graphic}}$$

# VDQI Example (p163)

annual  
sunshine record



1,000 numbers per  
square inch

# VDQI Example (p160)

**NO. 1450. STEEL PRODUCTS—NET SHIPMENTS, BY MARKET CLASSES: 1960 TO 1978**  
 [In thousands of short tons. Comprises carbon, alloy, and stainless steel. "N.e.c." means not elsewhere classified]

MARKET CLASS	1960	1965	1970	1973	1974	1975	1976	1977	1978
<b>Total</b> <sup>1</sup> .....	<b>71,149</b>	<b>92,666</b>	<b>90,798</b>	<b>111,430</b>	<b>109,472</b>	<b>79,957</b>	<b>89,447</b>	<b>91,147</b>	<b>97,935</b>
Steel for converting and processing.....	2,928	3,932	3,443	4,714	4,486	3,255	4,036	3,679	4,612
Independent forgers, n.e.c.....	841	1,250	1,048	1,213	1,339	1,098	952	998	1,192
Industrial fasteners <sup>2</sup> .....	1,071	1,234	1,005	1,278	1,331	675	912	848	870
Steel service centers, distributors.....	11,125	14,813	16,025	20,383	20,400	12,700	14,615	15,346	17,333
Construction, incl. maintenance.....	9,664	11,836	8,913	10,731	11,360	8,119	7,508	7,553	9,612
Contractors' products.....	3,602	5,018	4,440	6,459	6,249	3,927	4,502	4,500	3,480
Automotive.....	14,610	20,123	14,475	23,217	18,928	15,214	21,351	21,490	21,253
Rail transportation.....	2,525	3,805	3,098	3,228	3,417	3,152	3,056	3,238	3,549
Freight cars, passenger cars, locomotives.....	1,763	2,875	2,005	1,997	2,097	1,794	1,428	1,709	2,188
Rails and all other <sup>3</sup> .....	762	930	1,093	1,231	1,320	1,358	1,628	1,529	1,361
Shipbuilding and marine equip.....	622	1,051	859	1,019	1,339	1,413	969	869	845
Aircraft and aerospace.....	78	94	56	69	79	69	59	63	60
Oil and gas industries.....	1,759	1,936	3,550	3,405	4,210	4,171	2,653	3,650	4,140
Mining, quarrying, and lumbering.....	288	392	497	534	644	596	536	486	508
Agricultural, incl. machinery.....	1,003	1,483	1,126	1,772	1,859	1,429	1,784	1,743	1,805
Machinery, industrial equip., tools	3,958	5,873	5,169	6,351	6,440	5,173	5,180	5,566	5,992
Electrical equipment.....	2,078	2,985	2,694	3,348	3,242	2,173	2,671	2,639	2,811
Appliances, utensils, and cutlery...	1,760	2,179	2,160	2,747	2,412	1,653	1,950	2,129	2,094
Other domestic commercial equip.....	1,959	2,179	1,778	1,990	1,941	1,390	1,813	1,846	1,889
Containers, packaging, shipping....	6,429	7,331	7,775	7,911	8,218	6,053	6,914	6,714	6,595
Cans and closures.....	4,976	5,867	6,239	6,070	6,349	4,859	5,290	5,173	4,950
Ordnance and other military.....	165	289	1,222	918	654	405	219	193	207
Exports (reporting companies only)	2,563	2,078	5,985	3,138	3,961	1,755	1,839	1,076	1,224

<sup>1</sup> Total includes nonclassified shipments, and, beginning 1970, data include estimates for a relatively small number of companies which report raw steel production but not shipments. <sup>2</sup> Bolts, nuts, rivets, and screws.

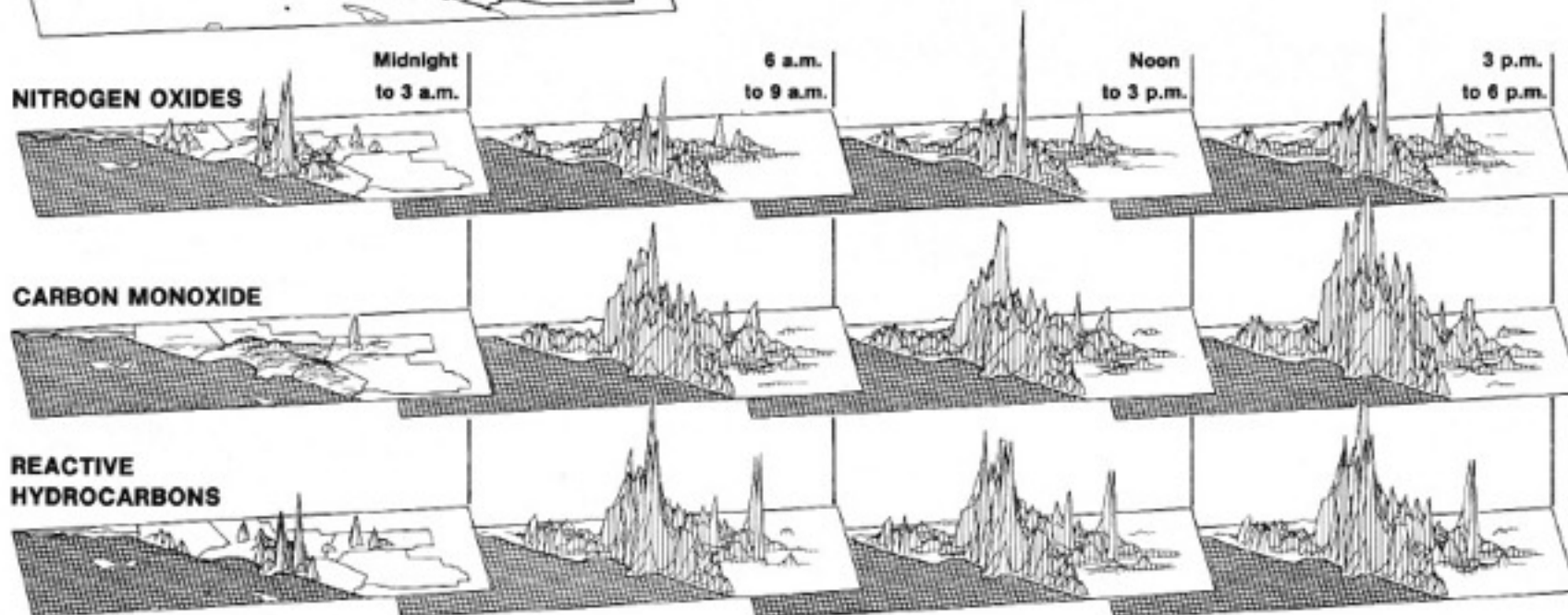
<sup>3</sup> Includes railways, rapid transit systems, railroad rails, trackwork, and equipment.

tables/  
matrices  
are  
dense

# VDQI Example (p42)

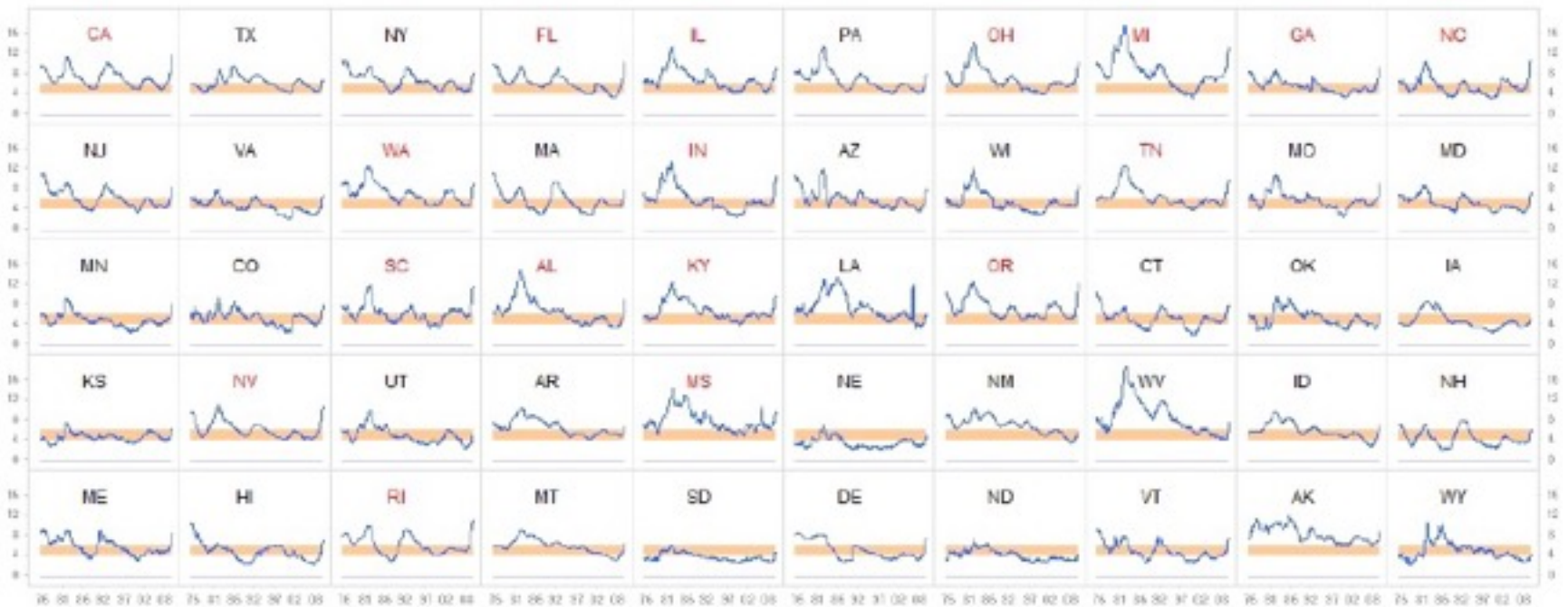


Small multiples  
are dense



# Small Multiples Example

Monthly Unemployment Rates by State, Jan 1976 - Apr 2009




Source: Bureau of Labor Statistics

Notes: The orange band denotes a "normal" unemployment rate (4%-6%);  
State code in red: unemployment rate in April 2009 is higher than the US average

<http://www.excelcharts.com/blog/charts-monthly-unemployment-rates-by-state-1976-2009/>

# Resources

- Envisioning Information  
*by Edward R. Tufte, Graphics Press, 1990*
  - Visual Explanations  
*by Edward R. Tufte, Graphics Press, 1997*
  - The Visual Display of Quantitative Information  
*by Edward R. Tufte, Graphics Press, ,2001*
- 

# Evaluation Recap

- Lie Factor
  - Size of effect shown in graphic versus size of effect shown in data
- Data-Ink Ratio
  - Proportion of ink devoted to non-redundant display of data information
- Data Density
  - Amount of data entries versus graphic area

# **LOSSES IN FRENCH ARMY**

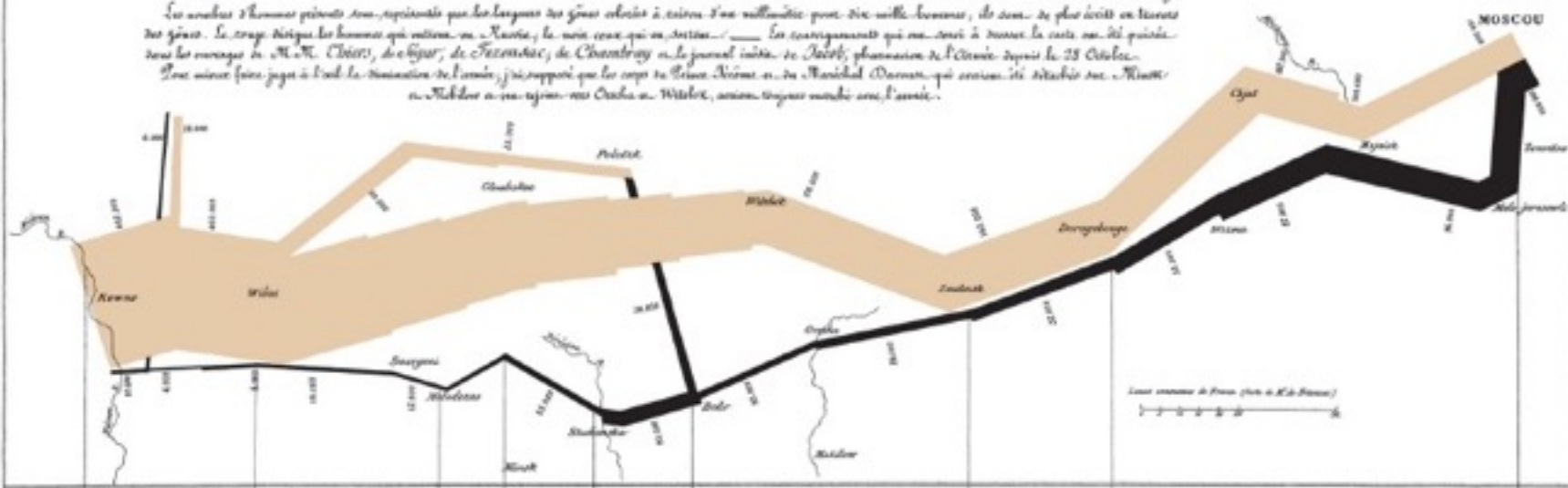
By Charles Joseph Minard, 1869



# Carte Figurative des postes successives en hommes de l'Armée Française dans la Campagne de Russie 1812-1813.

Ordonné par M. MINARD, Supérieur Général des Forts et Chantiers de la Marine. Paris, le 20 Novembre 1869.

Les analyses d'hommes privés sont supposées que les longues des gens choisis à extraire l'air vicié pour six mille hommes, de deux de plus écrit en lettres des gens. Le crayon indique les hommes qui entrent en Russie, le noir ceux qui en sortent. Les arrangements qui ont servi à travers la carte ont été prisés sous les ordres de M. M. Chéris, de Béjar, de Fessendic, de Chevrolat et le journal inédit de Jolly, pharmacien de l'Armée depuis le 25 Octobre. Les mines font jusqu'à l'oubli la destination de l'armée, j'ai supposé que les corps de Blaise Népom et du Maréchal Davout qui avaient été établis sur Moskva et Nihilus à un rayon des Orcha et Witebsk, avaient toujours marché avec l'armée.



## TABLEAU GRAPHIQUE de la température en degrés de thermomètre de Réaumur au dessous de zéro.



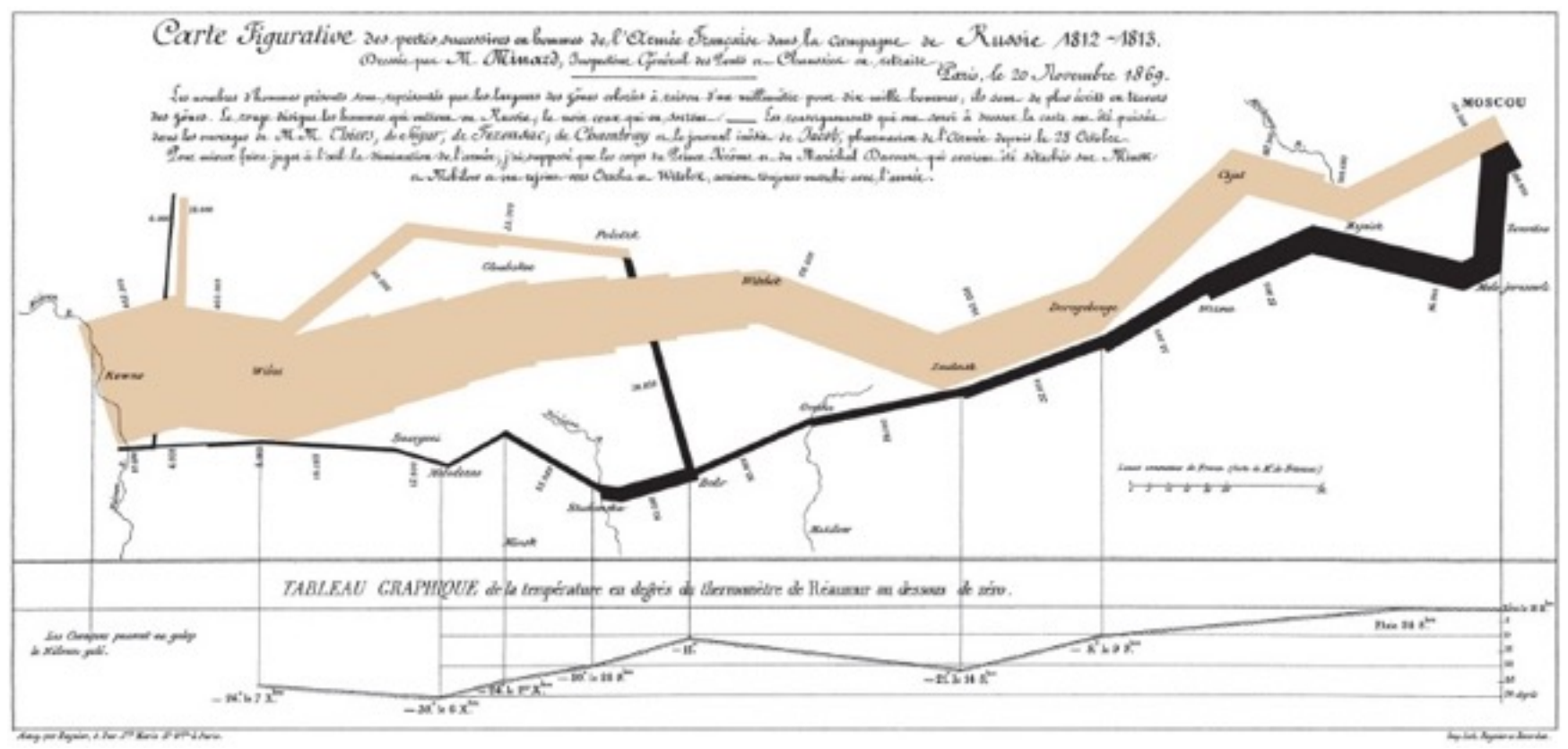
Along par Agnès, à Paris 17 Rue de la Harpe.

Après les Agnès et les autres.

# Interpretation

- Illustrates losses suffered by Napoleon's army during the Russia campaign of 1812
- Important geographical markers, like rivers, are provided
- Width of tan band is size of army at location
- Black band shows retreat of army
- Temperature during retreat shown at bottom

# Lie Factor



# Lie Factor

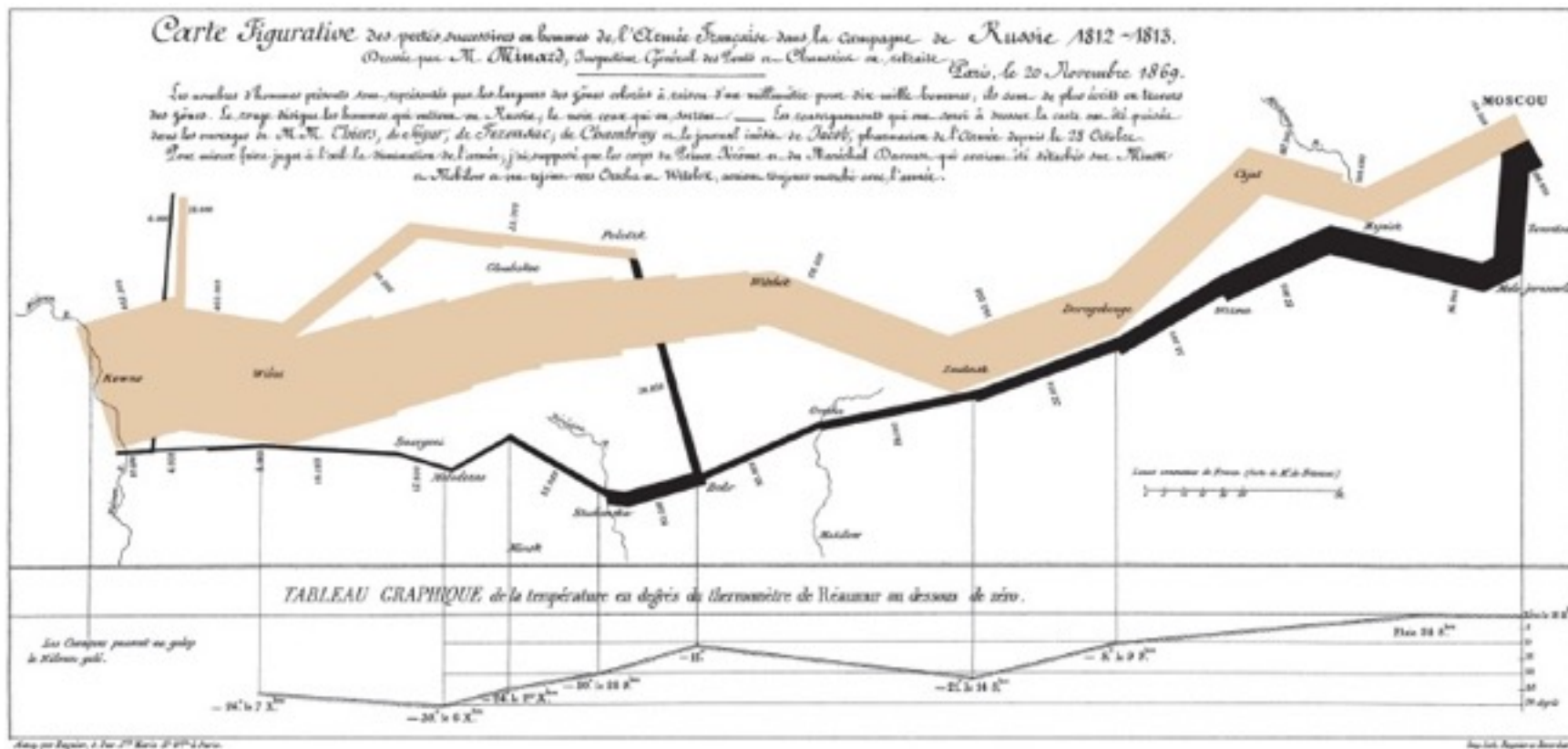
- Size is given by width which is perceived fairly accurately
- Some issues likely due to hand-drawn nature of the figure
- Time is related to position
  - Horizontal movement does not take equal amounts of time



# Data-Ink Ratio

- Can maybe remove some lines
  - But often provide context and easy reference
- Can maybe remove some text
  - But often provides context and easy reference
- Can remove shading of temperature line
  - But provides distinction between foreground and background

# Data Density



# Data Density

- Size of army (width of bands)
- Location of army (underlying map)
  - Latitude and longitude values
- Direction of army (tan versus black bands)
- Temperature change over time
  - Temperature value
  - Time (given by annotation and position)



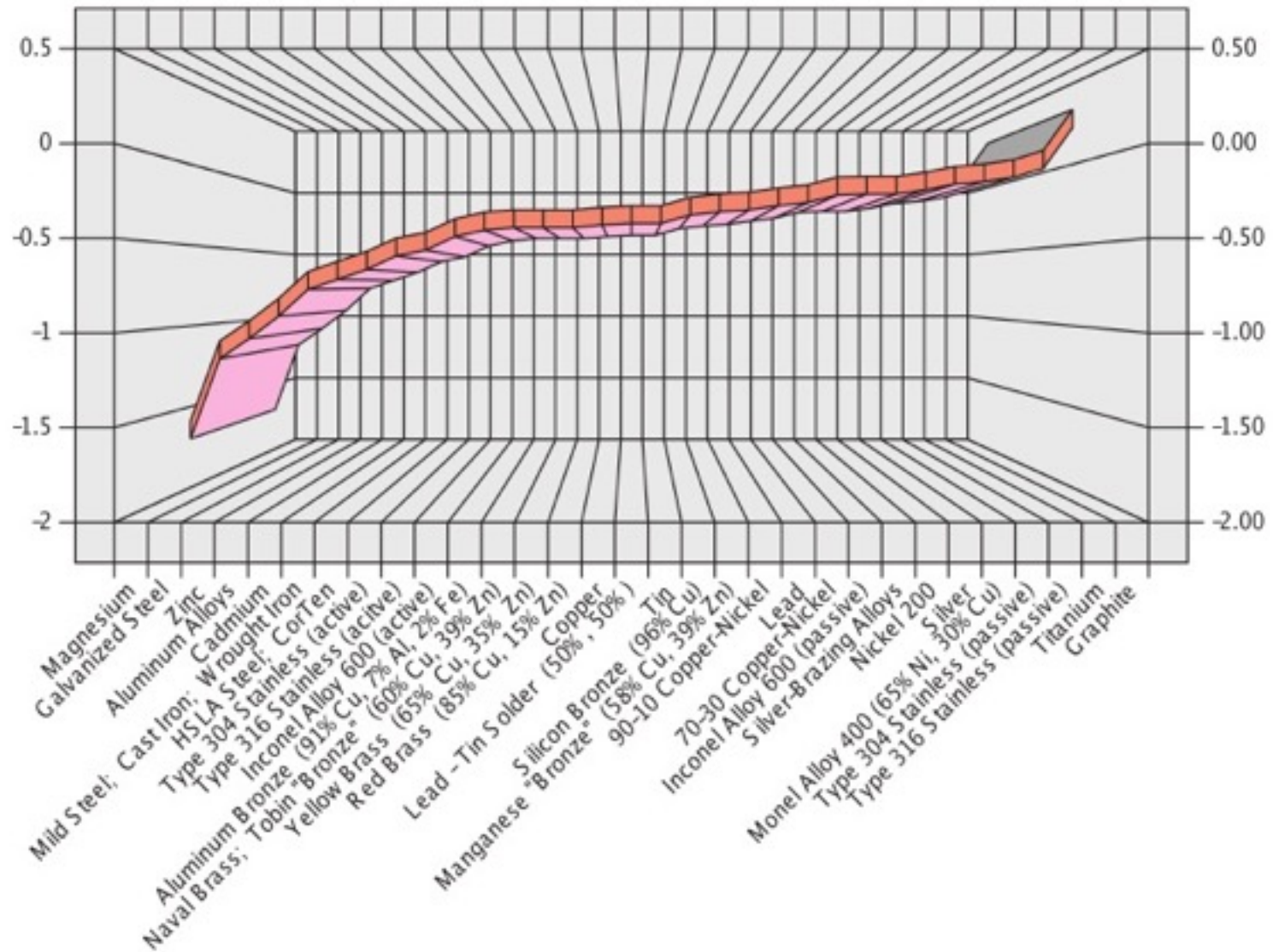
# Summary

- Excellent visualization
  - Lie factor reasonable
  - Data-ink ratio very high
  - Data density very high
- Tufte: “It may well be the best statistical graphic ever drawn.”

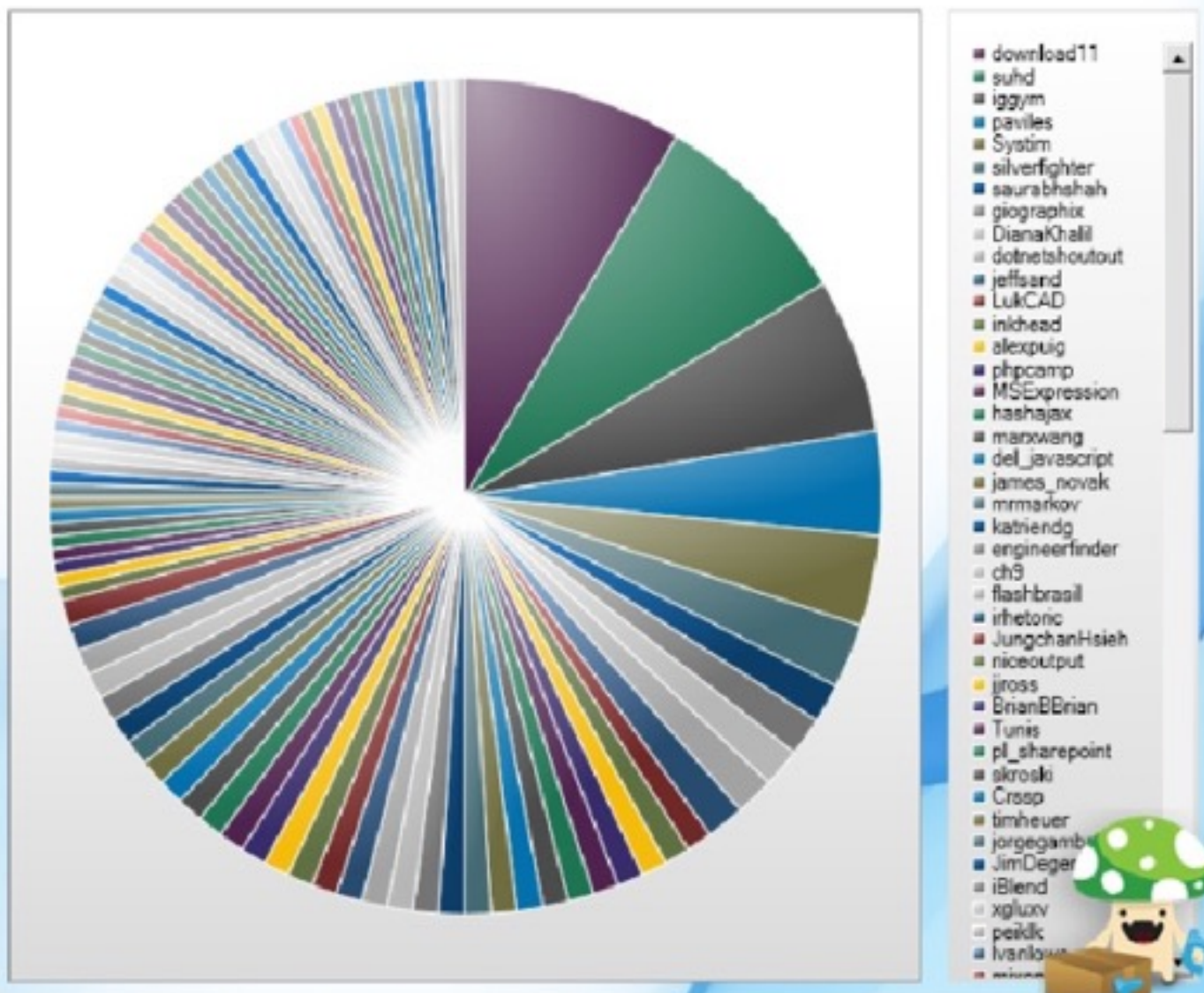
# OTHER EXAMPLES

From <http://www.datavis.ca/gallery/>

## Average Voltage in Seawater



## 100 Most Active Tweeters



[http://readwrite.com/2009/08/10/10\\_ways\\_to\\_archive\\_your\\_tweets](http://readwrite.com/2009/08/10/10_ways_to_archive_your_tweets)

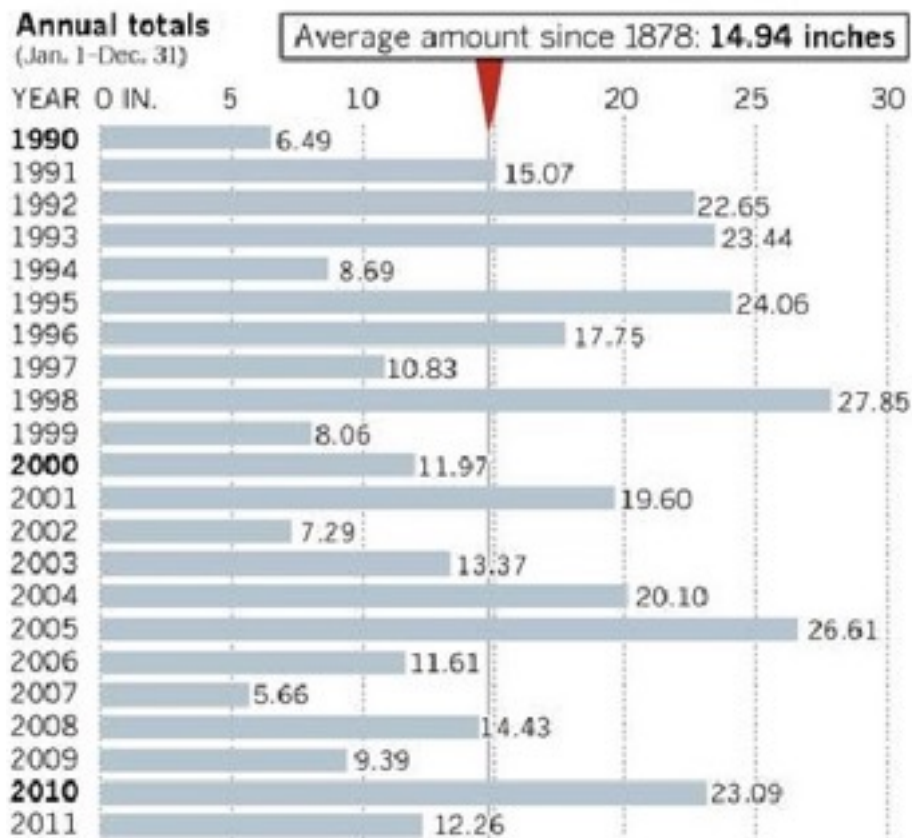
# OTHER EXAMPLES

<http://junkcharts.typepad.com/>



## L.A. annual rainfall, 1990-2011

A total of 12.26 inches of rain was recorded in 2011 at the National Weather Service's downtown Los Angeles weather station, located at USC.



### Annual records since 1878

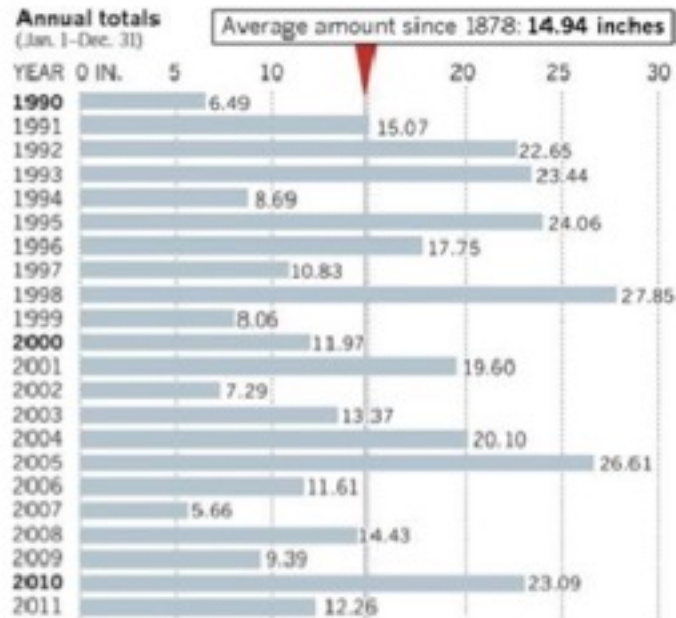
**Wettest** 40.29 (1884) **Driest** 4.08 (1953)

Source: National Weather Service

Los Angeles Times

## L.A. annual rainfall, 1990-2011

A total of 12.26 inches of rain was recorded in 2011 at the National Weather Service's downtown Los Angeles weather station, located at USC.



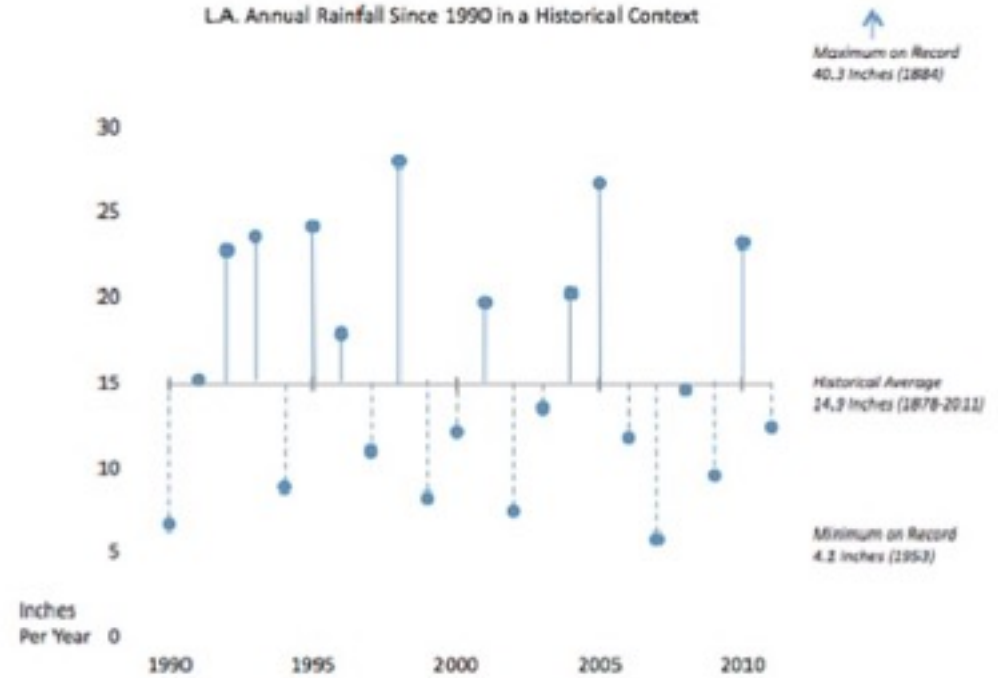
### Annual records since 1878

**Wettest** 40.29 (1884) **Driest** 4.08 (1953)

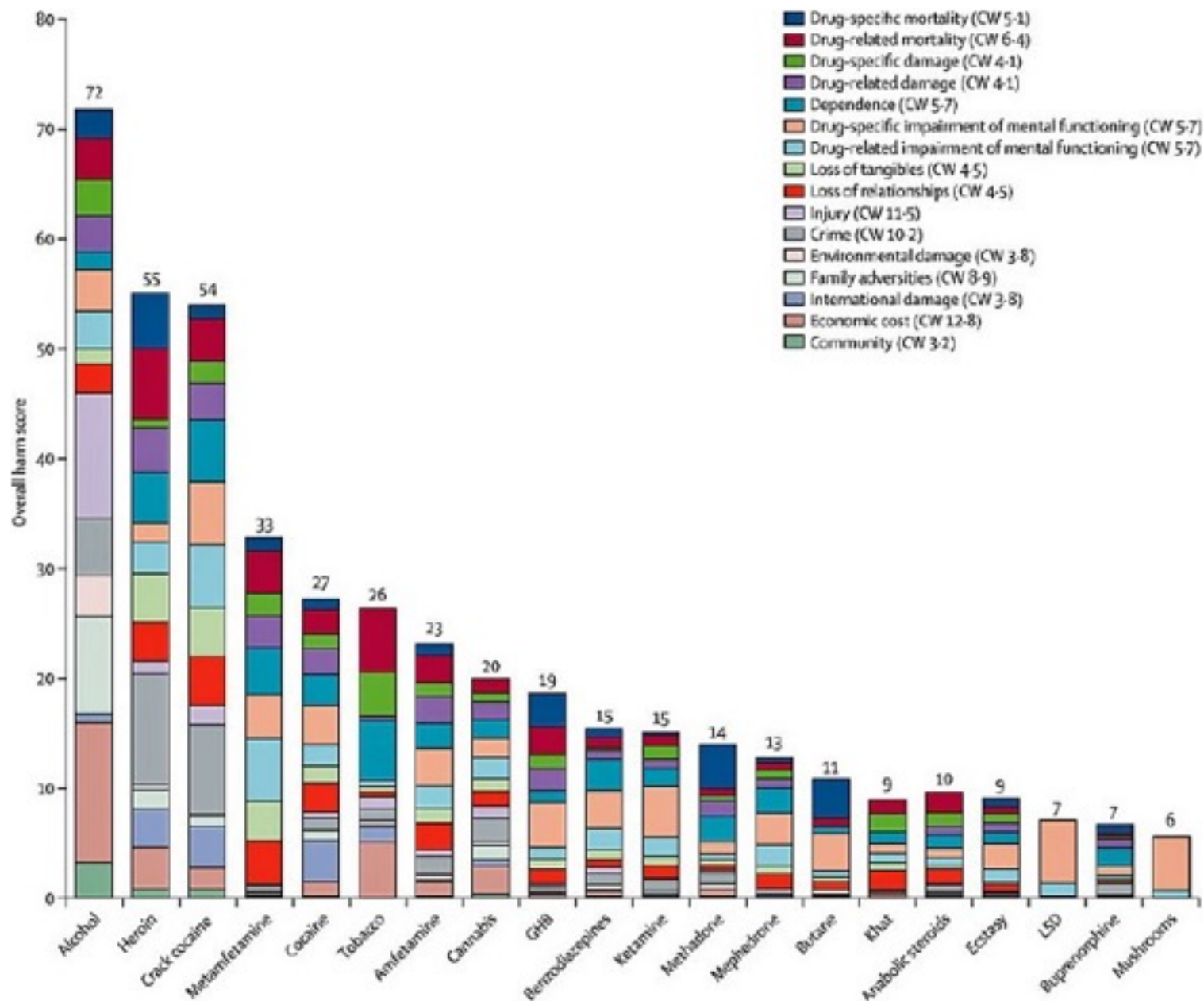
Source: National Weather Service

Los Angeles Times

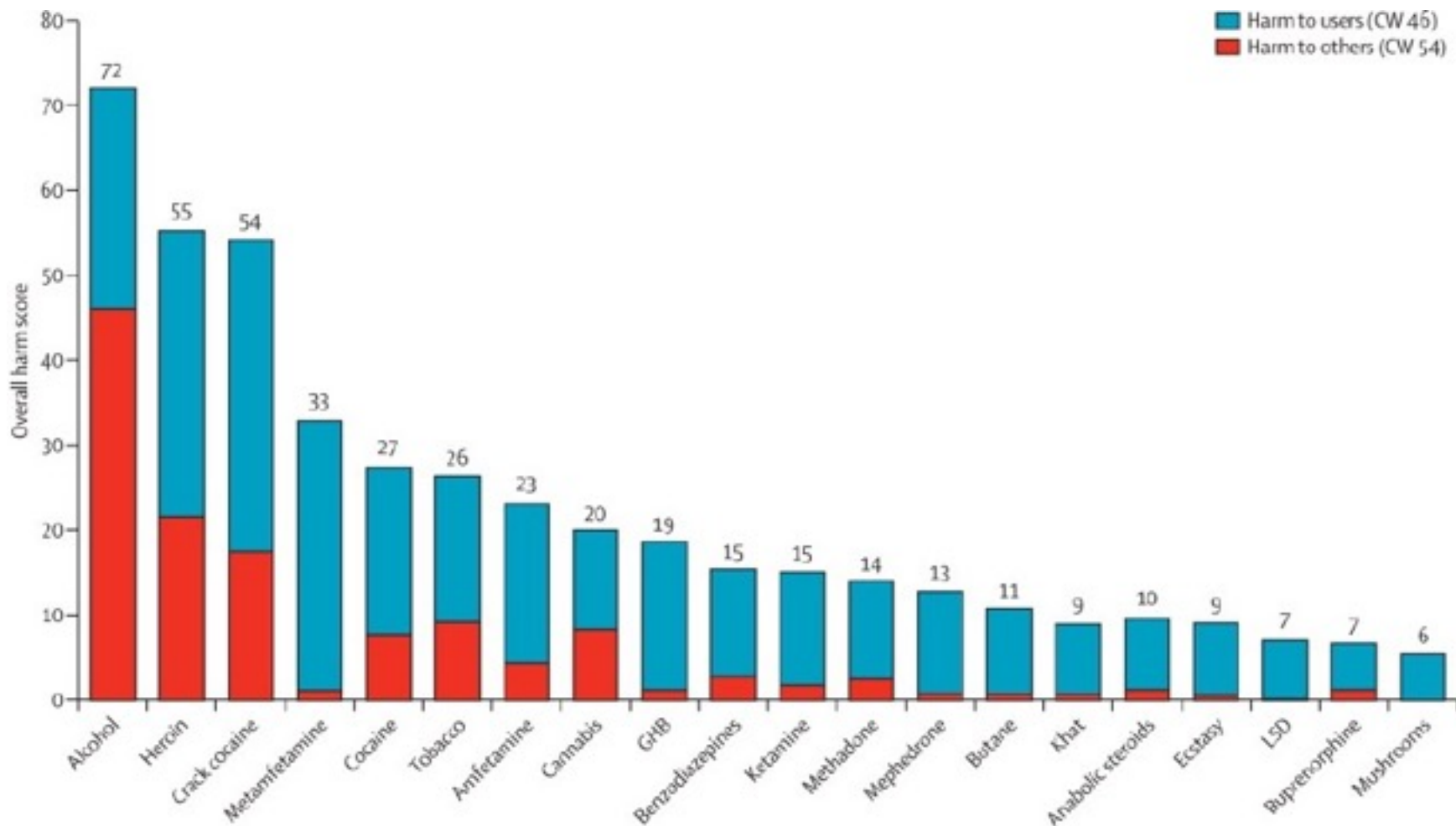
LA. Annual Rainfall Since 1990 in a Historical Context







[http://junkcharts.typepad.com/junk\\_charts/2012/07/staggering-excess.html](http://junkcharts.typepad.com/junk_charts/2012/07/staggering-excess.html)



[http://junkcharts.typepad.com/junk\\_charts/2012/07/staggering-excess.html](http://junkcharts.typepad.com/junk_charts/2012/07/staggering-excess.html)

# **OTHER EXAMPLES**

<http://andrewgelman.com/>

# HOW 2012 STACKS UP

THE WARMEST YEARS ON RECORD

CONTIGUOUS U.S.



Source: NOAA's National Climatic Data Center - State of the Climate National Overview

CLIMATE  CENTRAL

Sort states by:

POLITICS

REGION

ALPHABETICAL

**Laws on file**

If no colour appears, there is no such law on file

- 2012 election results
- Background check law
- Permit required to purchase
- Licence required to sell
- Records kept on file
- Firearms banned from workplace

**New York**

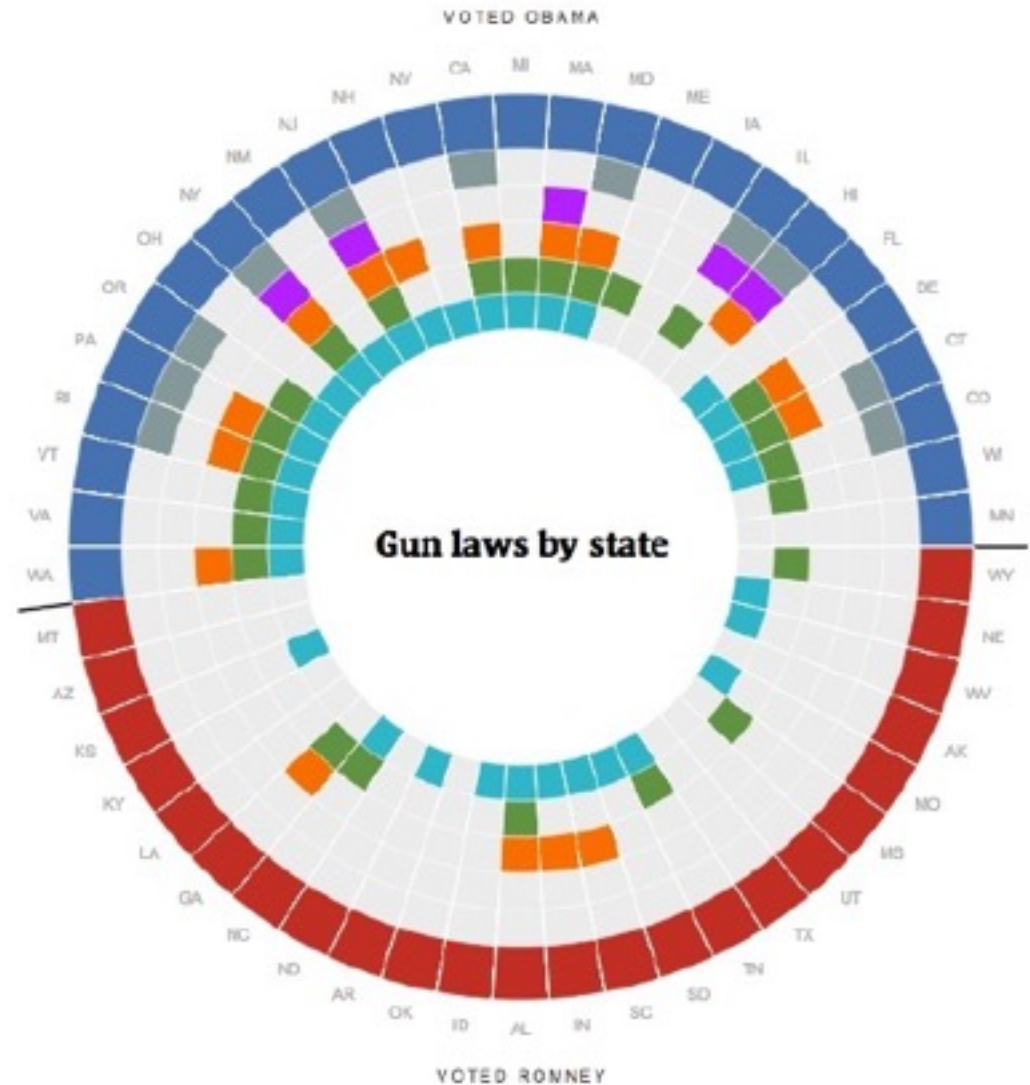
- Voted for Obama in the 2012 election
- Background check:** required for handguns
- Permit:** required to buy firearms
- Licence:** required for dealers
- Records:** kept on file for handgun owners
- Workplace:** firearms not allowed in parking lots

**Overall gun control score: 62**

New York has a **Brady Campaign score** of 62, which is higher than the national average of 16. The score comes from measuring these and other gun laws according to a weighted points system.

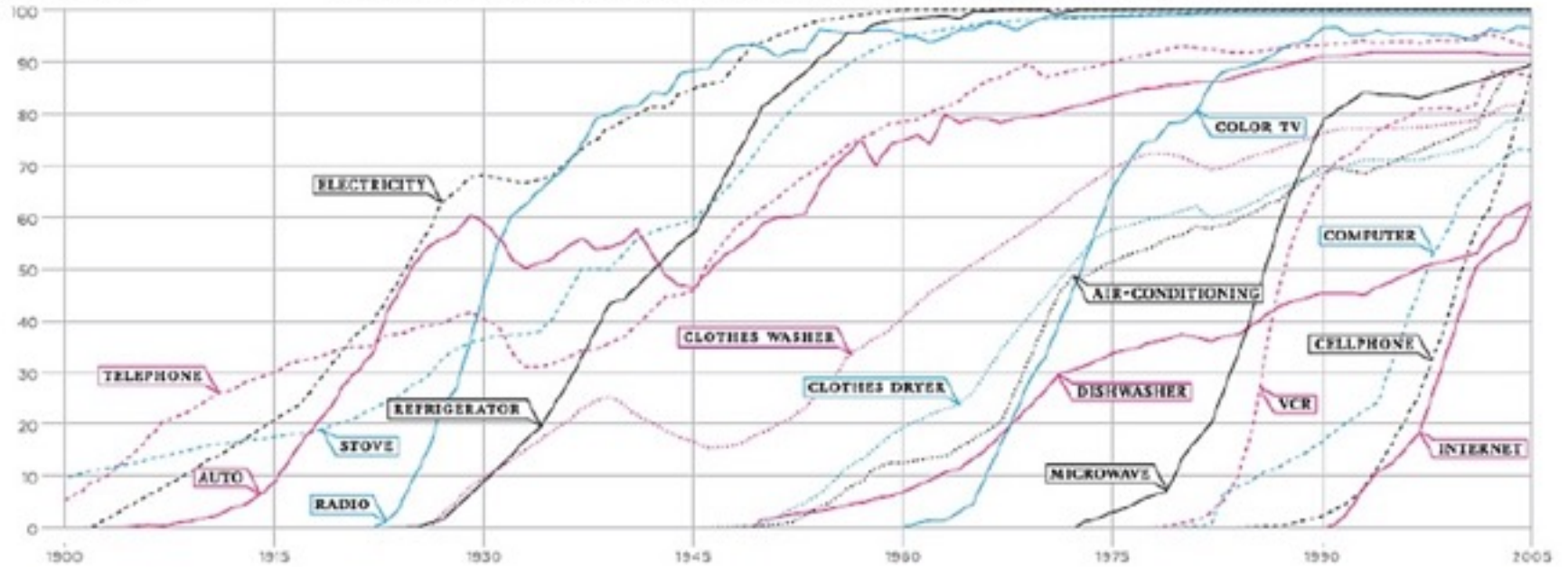
**Murder rate: 4.12**

There were 4.12 firearm murders per 100,000 people in New York during 2011, which is higher than the national average of 2.77. Overall, it is ranked #6 in murder rates out of 48 states with this data.

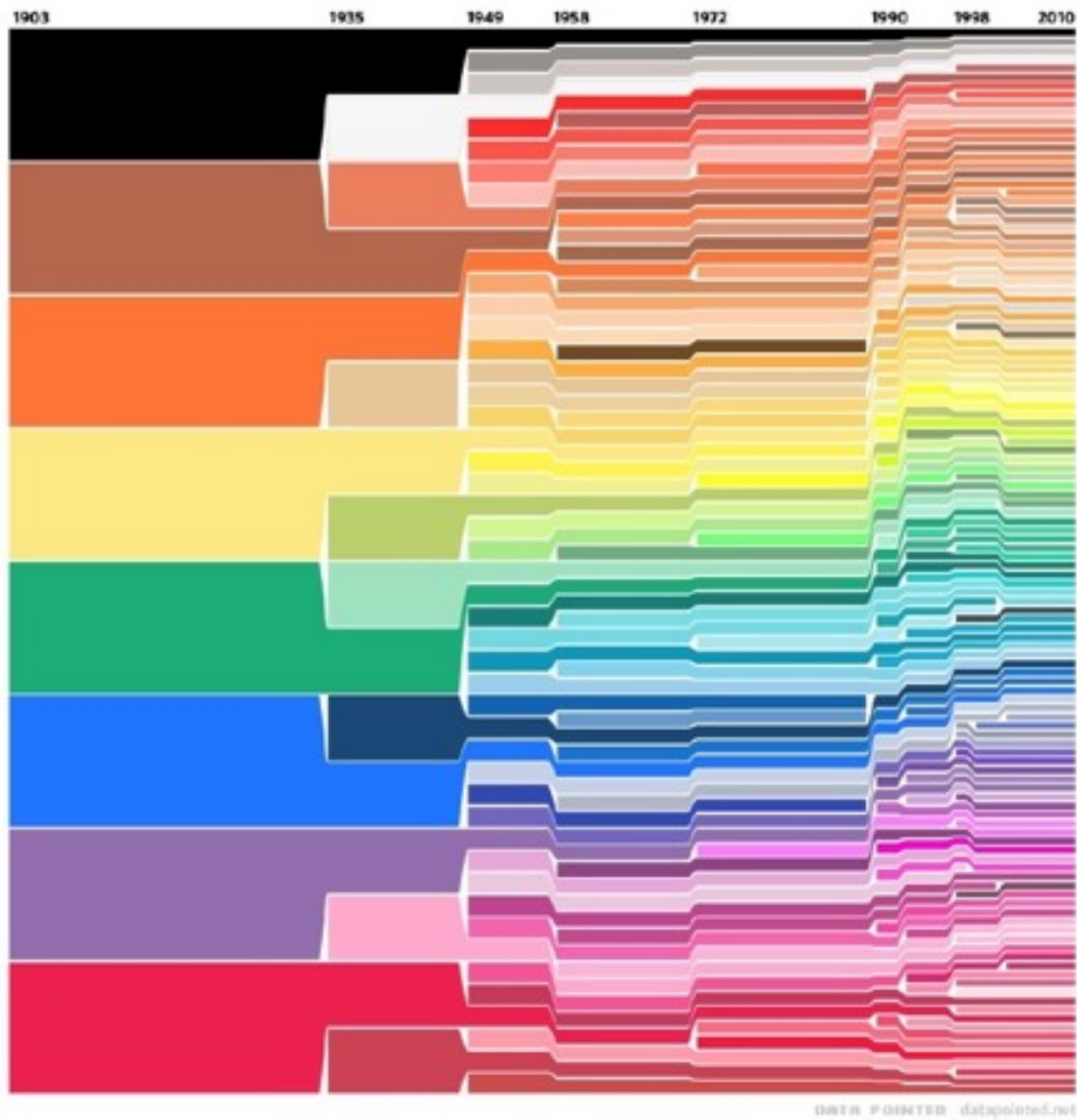


PERCENT OF  
J.S. HOUSEHOLDS

## CONSUMPTION SPREADS FASTER TODAY



<http://www.theatlantic.com/technology/archive/2012/04/the-100-year-march-of-technology-in-1-graph/255573/>



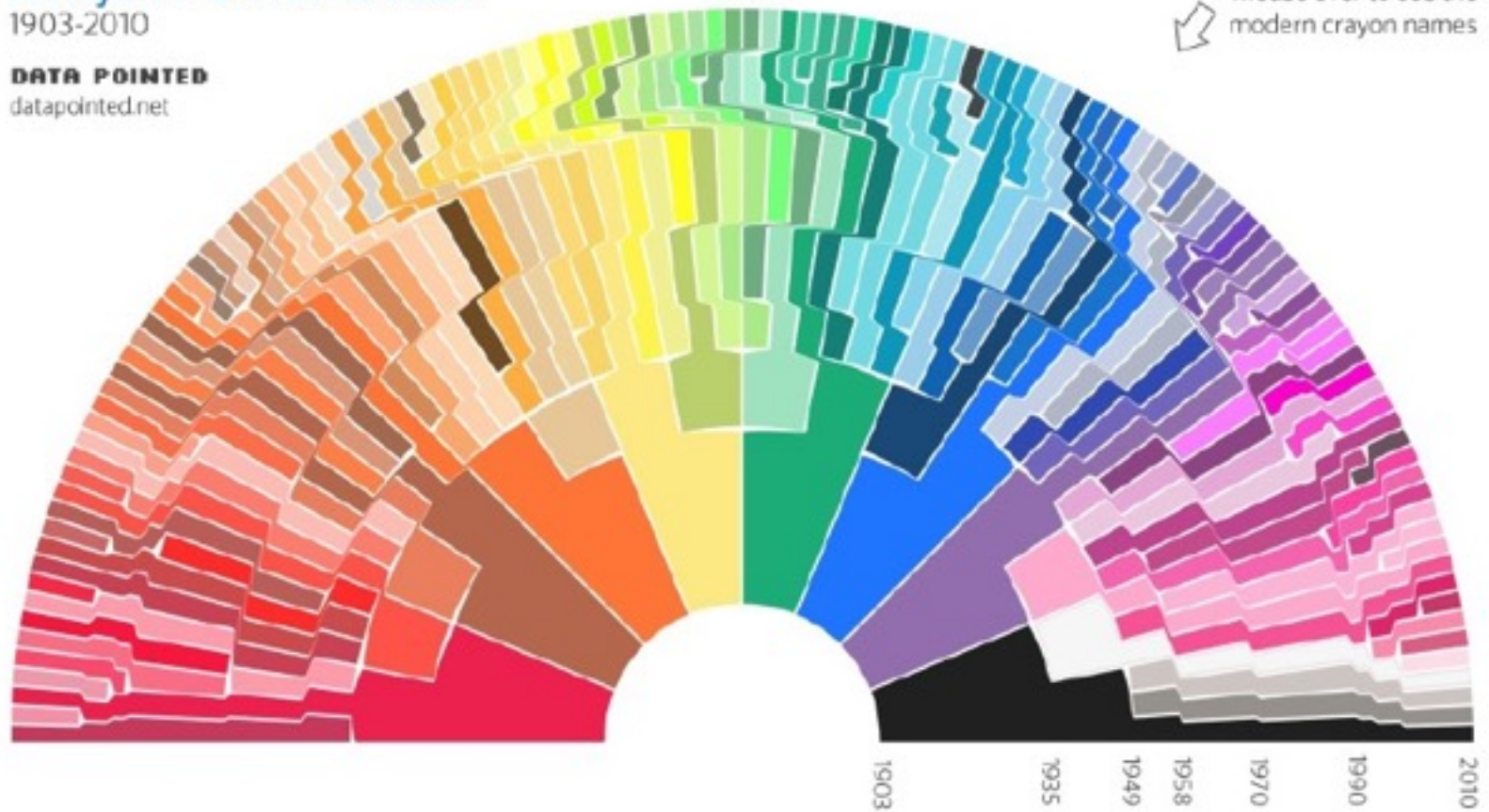
<http://www.datapointed.net/visualizations/color/crayola-crayon-chart/>

# Crayola Color Chart

1903-2010

**DATA POINTED**  
datapointed.net

mouse over to see the  
modern crayon names



<http://www.datapointed.net/visualizations/color/crayola-crayon-chart-bow/>



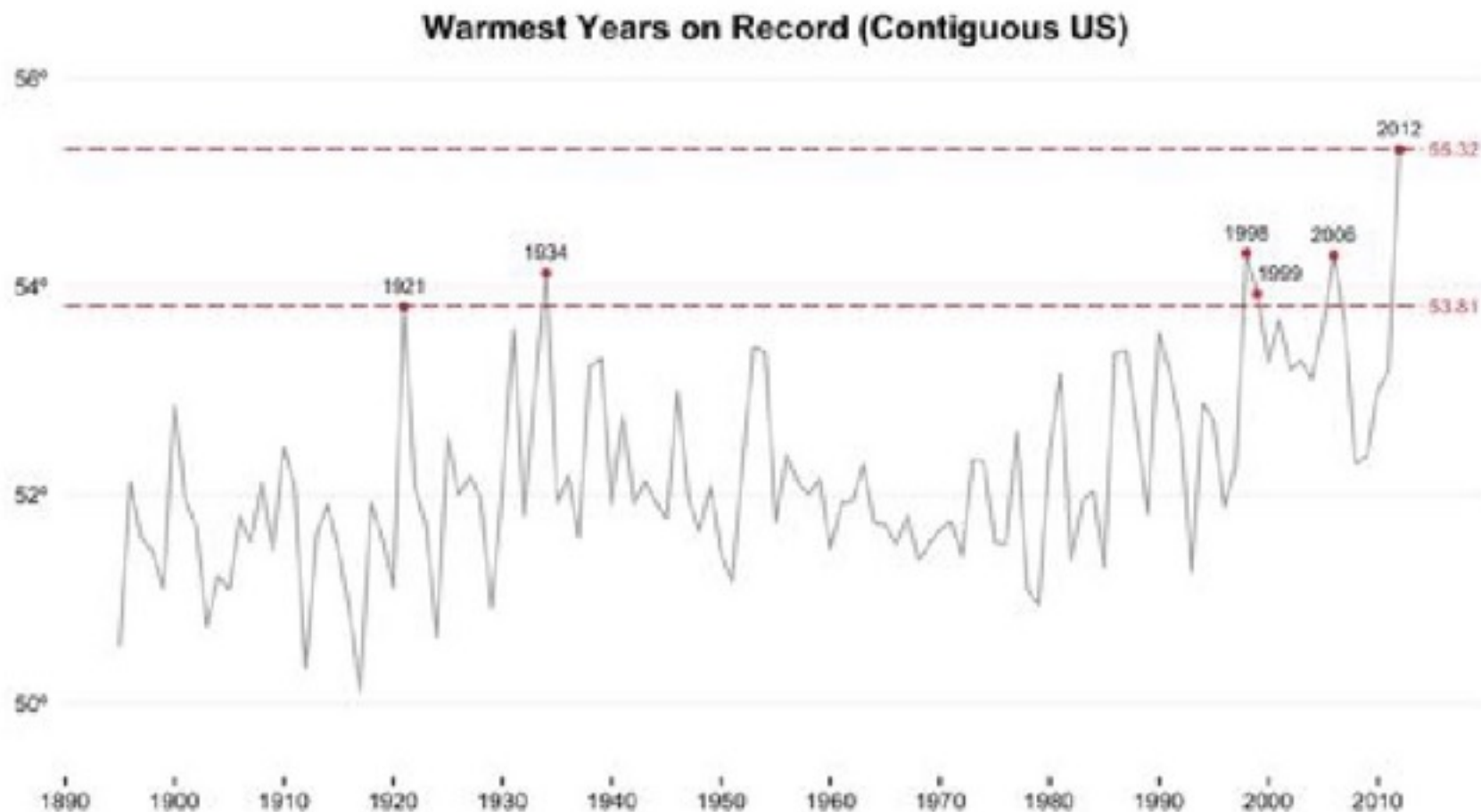
# 2D VISUALIZATION

## Basics

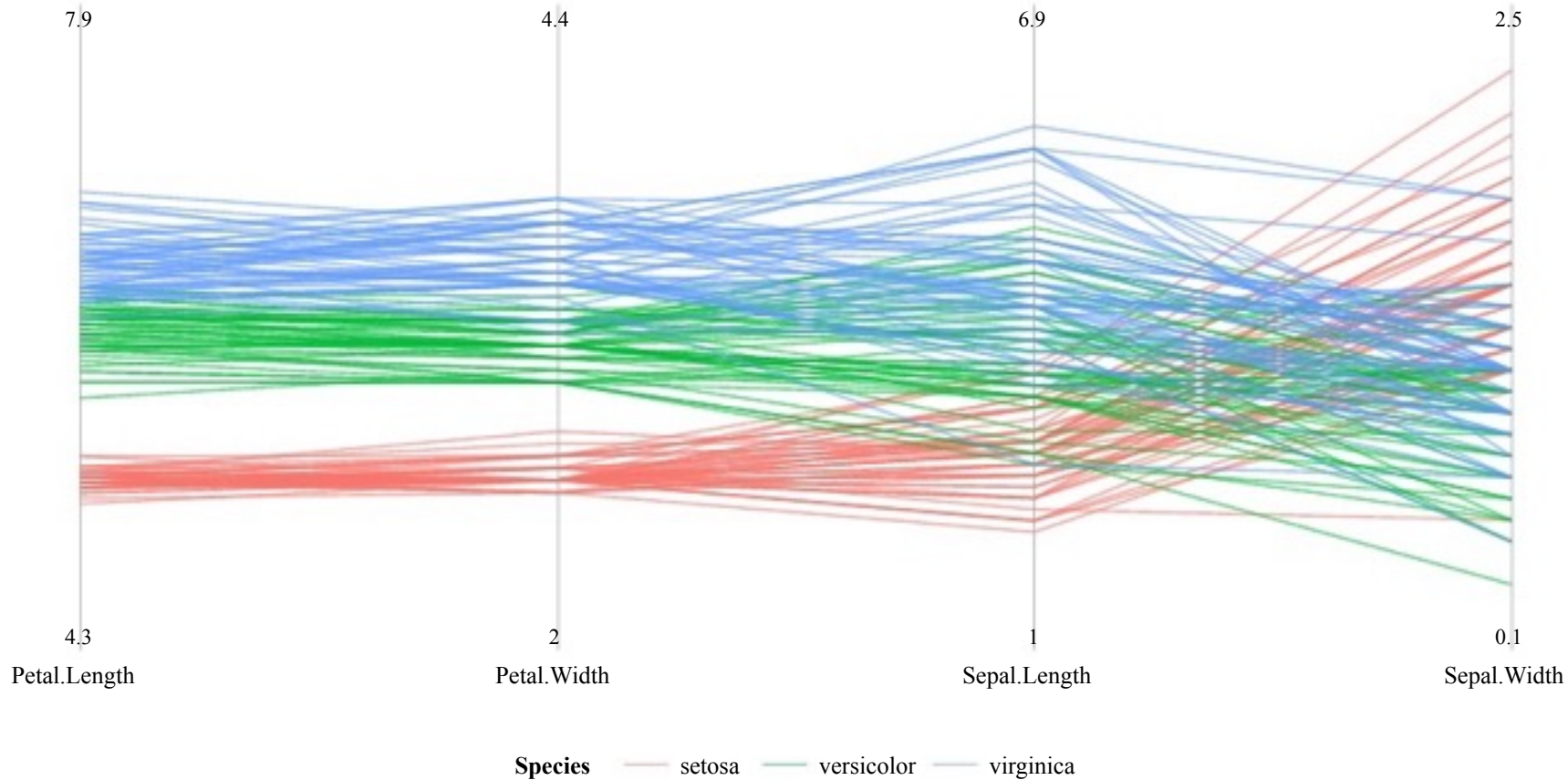
# 2D Visualization

- Only two dimensions for position
  - *Can gain additional dimensions by encoding data using other pre-attentive attributes*
- Includes many charts and graphs
- Can have issues with occlusion and overplotting
  - *Especially an issue with high-dimensional data*
  - *Especially an issue with dense data*

# Example 2D Visualization



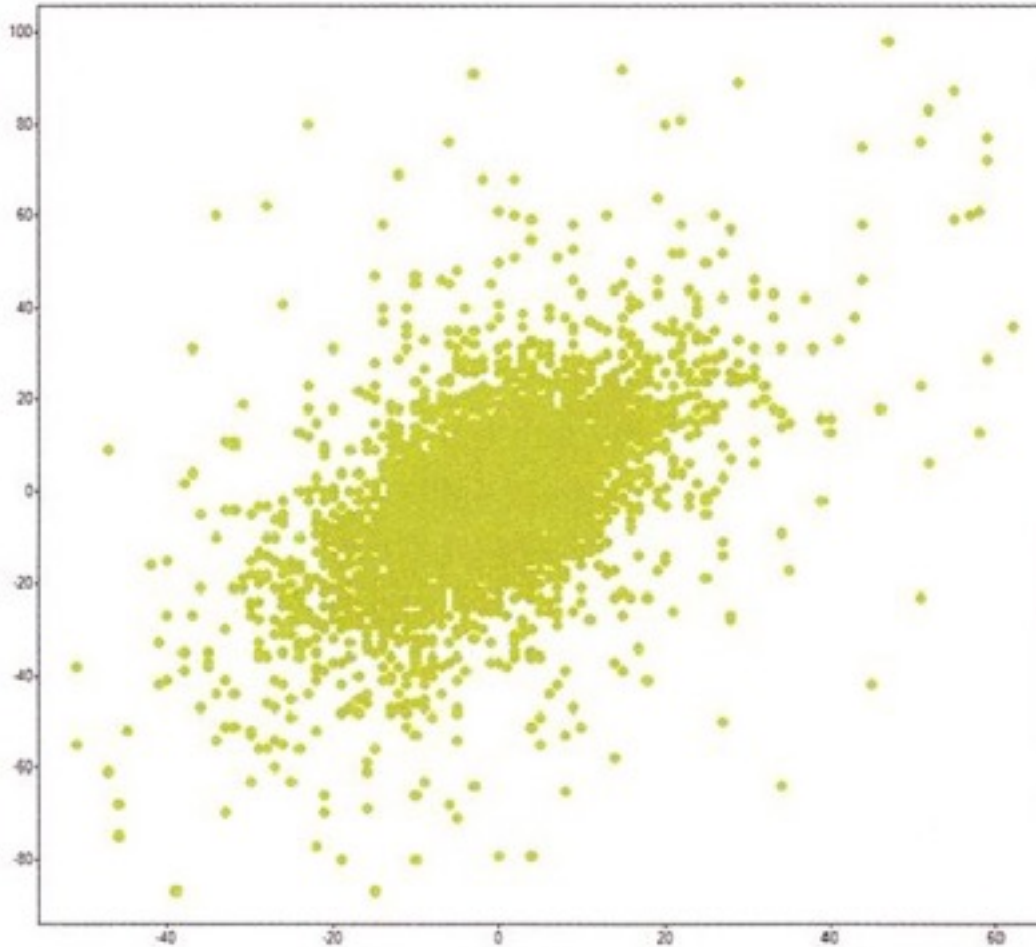
# Example 2D Visualization



# 2D VISUALIZATION

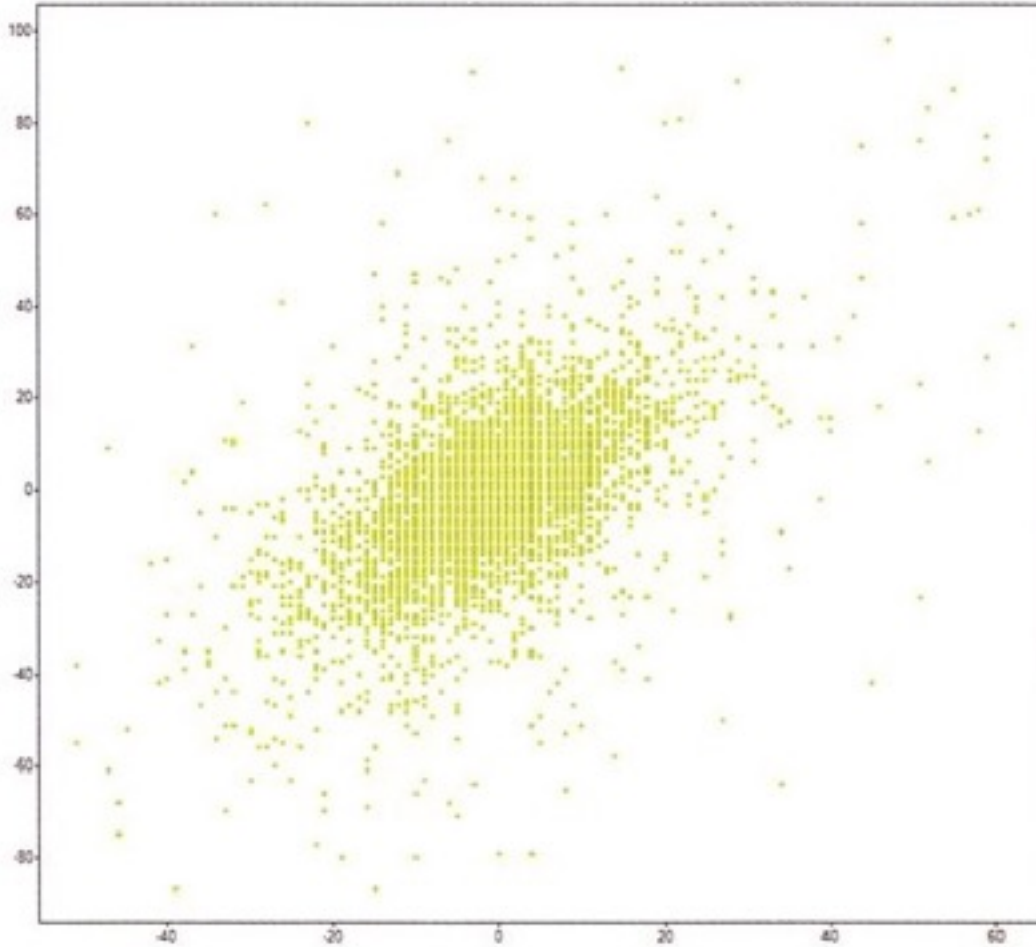
Overplotting

# Overplotting Example 1



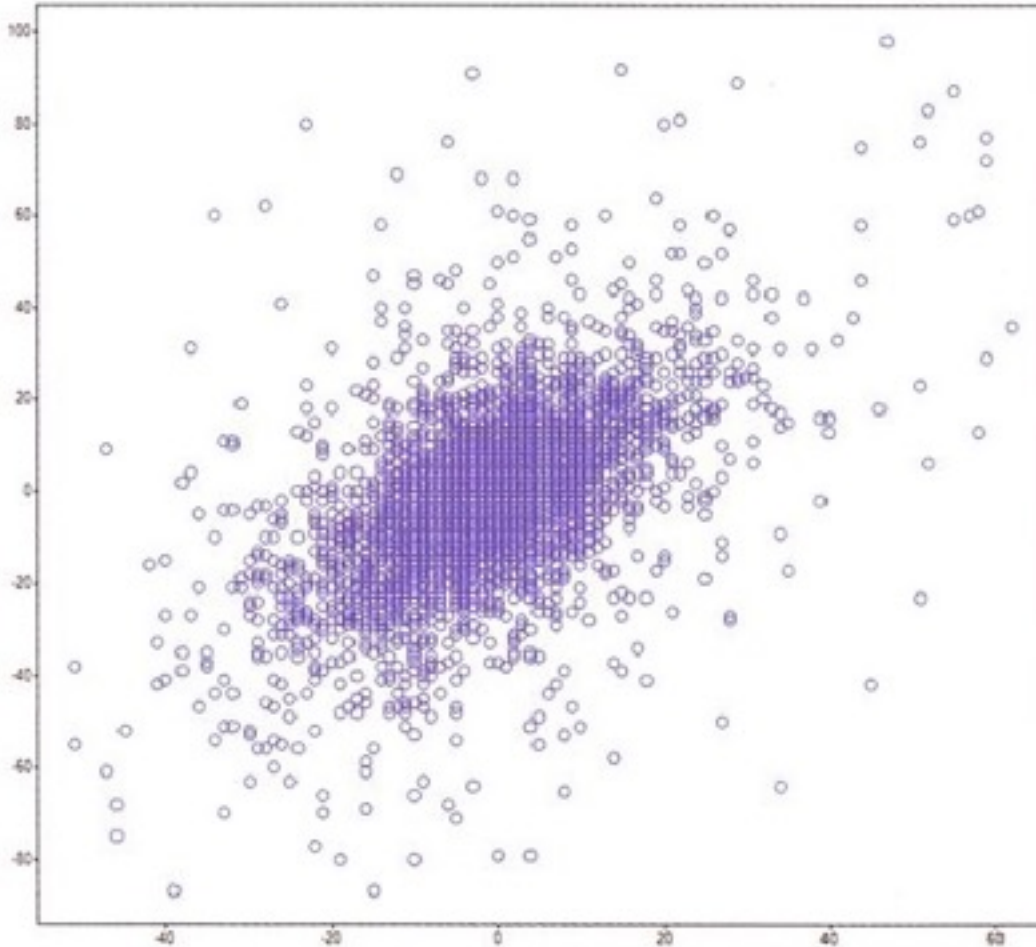
Now You See It, by Stephen Few, Analytics Press, 2009, p118.

# Overplotting – reducing size



Now You See It, by Stephen Few, Analytics Press, 2009, p118.

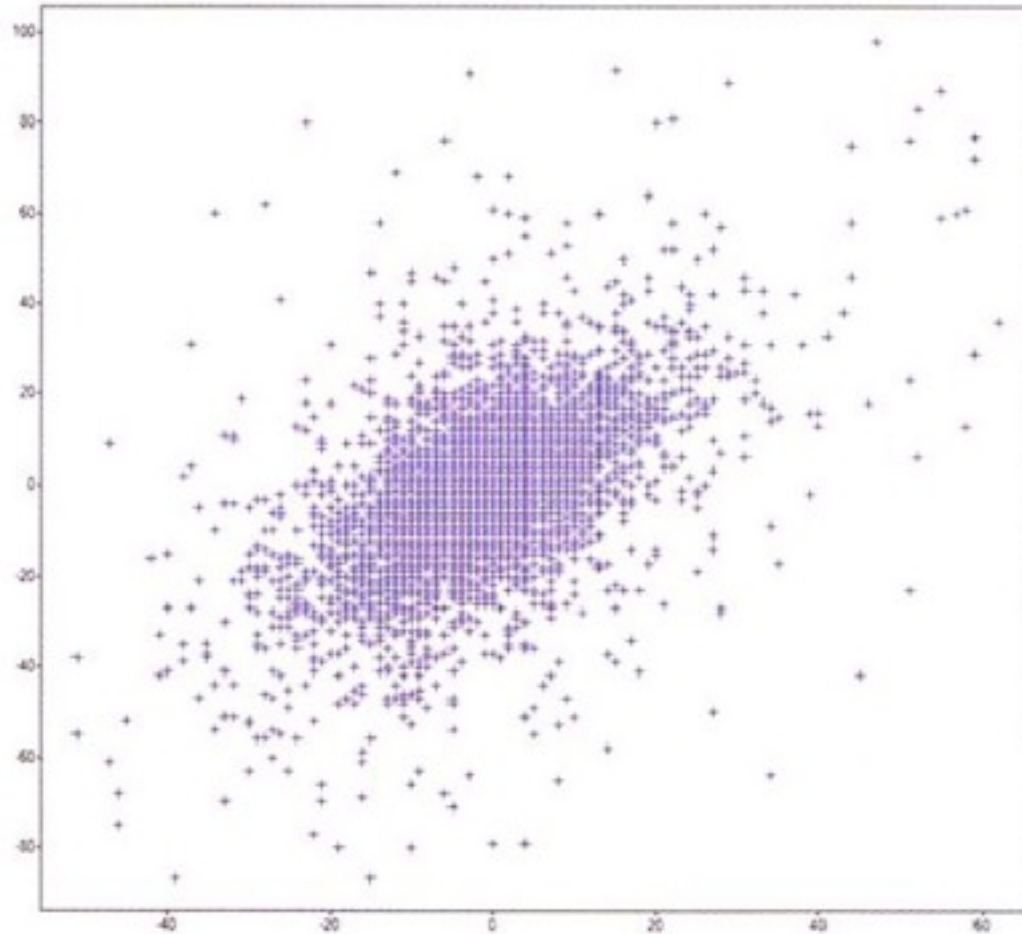
# Overplotting – removing fill color



Now You See It, by Stephen Few, Analytics Press, 2009, p118

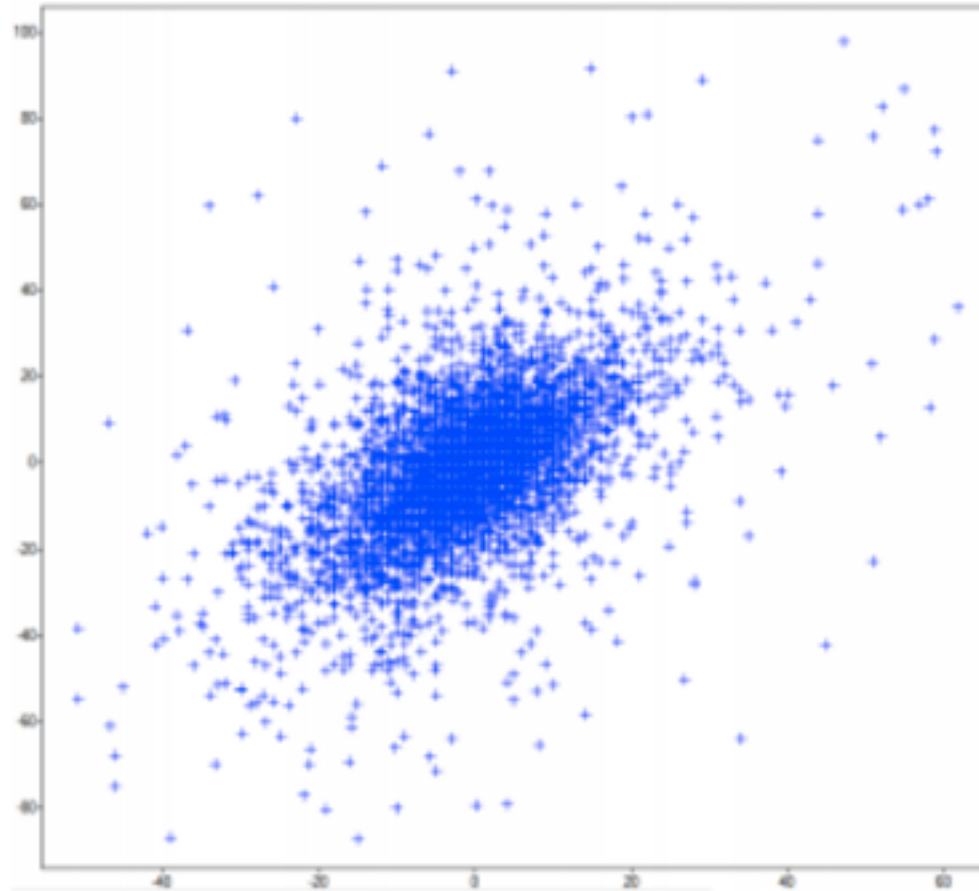


# Overplotting – changing shape



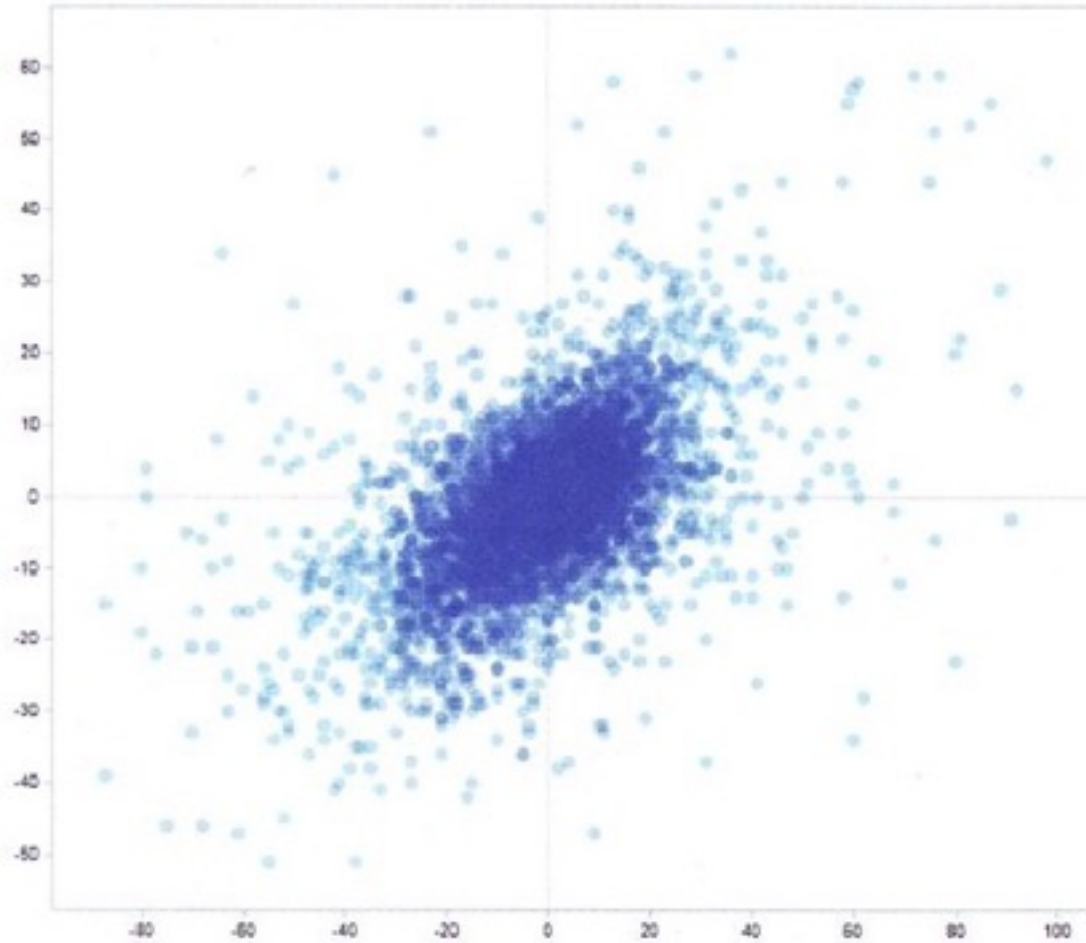
Now You See It, by Stephen Few, Analytics Press, 2009, p118.

# Overplotting – jittering



Now You See It, by Stephen Few, Analytics Press, 2009, p118.

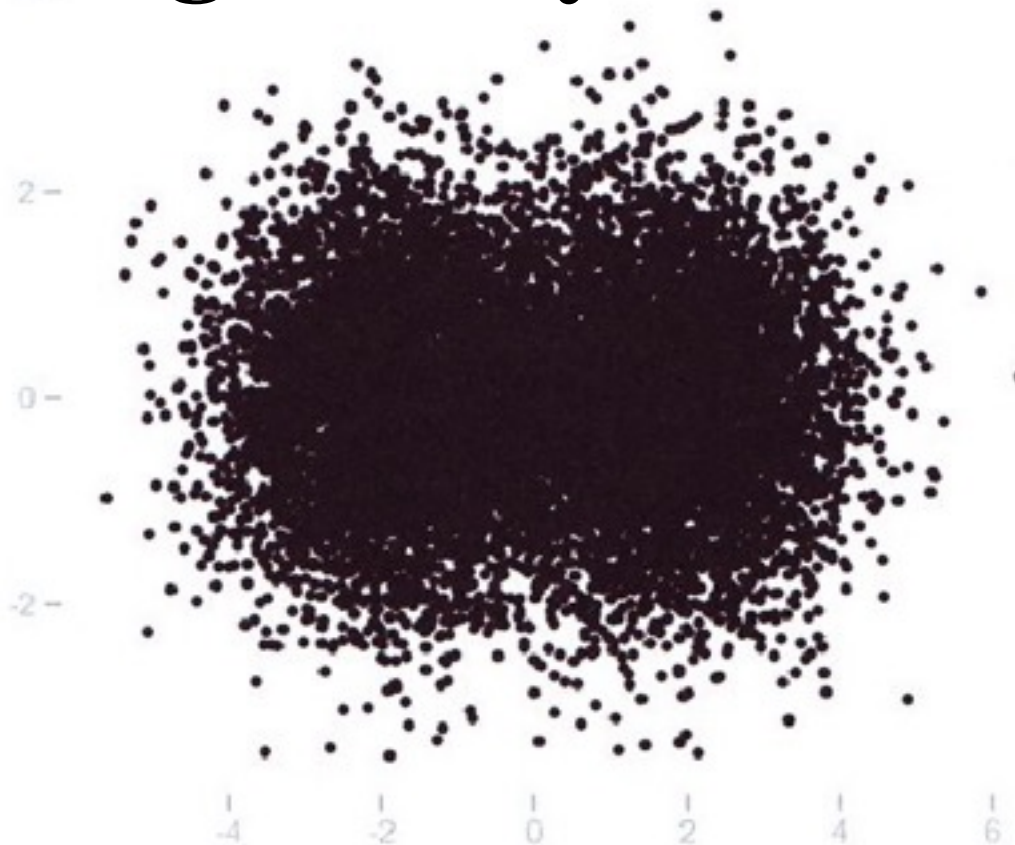
# Overplotting – transparency



Now You See It, by Stephen Few, Analytics Press, 2009, p118.

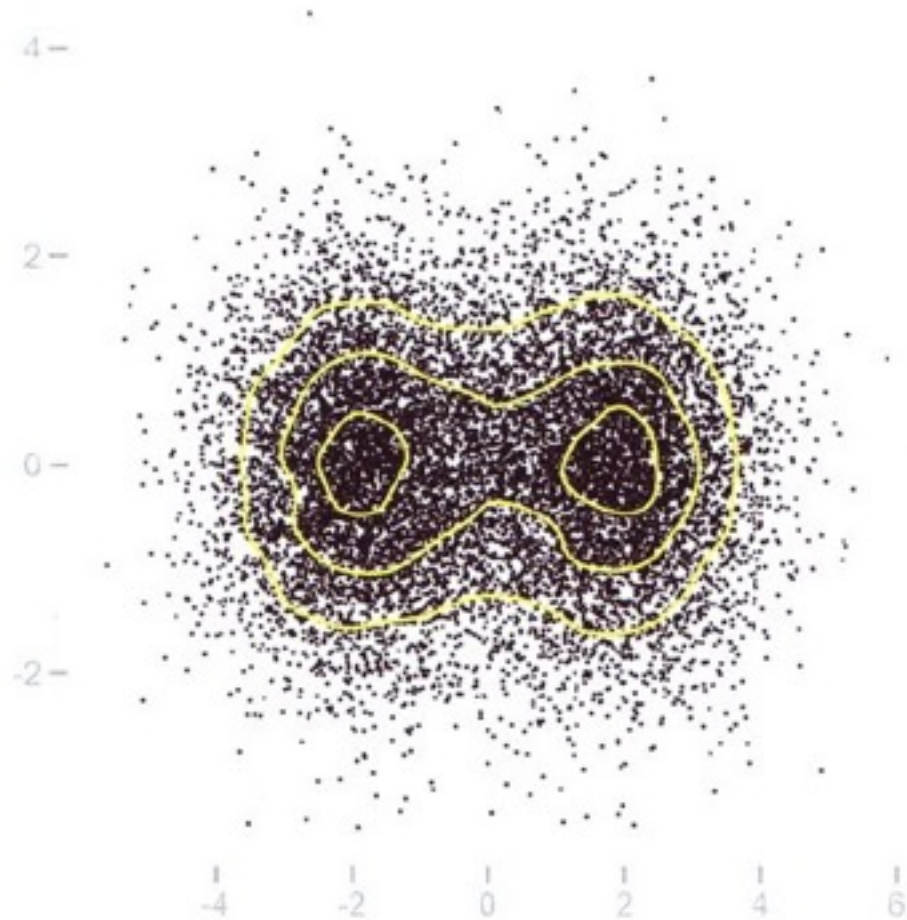
# Overplotting Example 2

## Encoding density values



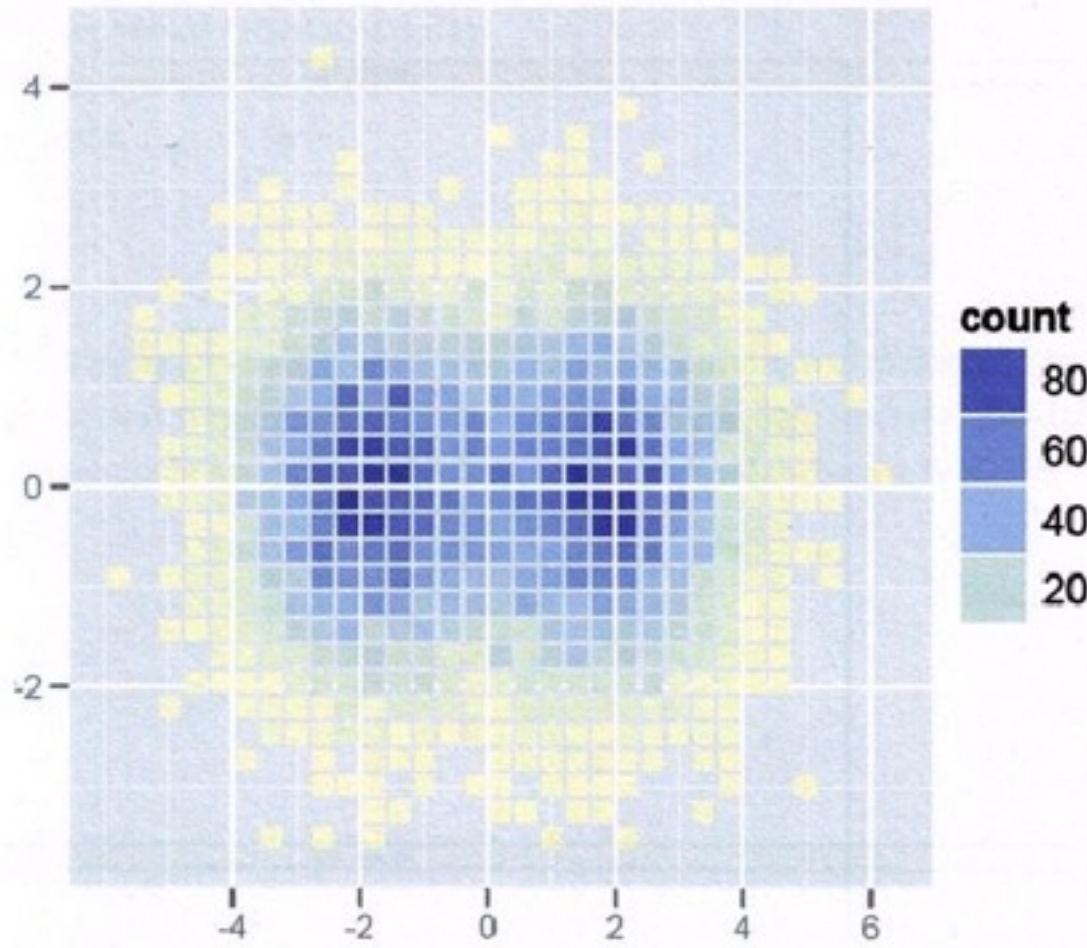
Now You See It, by Stephen Few, Analytics Press, 2009, p118.

# Overplotting – using contours



Now You See It, by Stephen Few, Analytics Press, 2009, p118.

# Overplotting – using colors



Now You See It, by Stephen Few, Analytics Press, 2009, p118.

# Overplotting Remedies

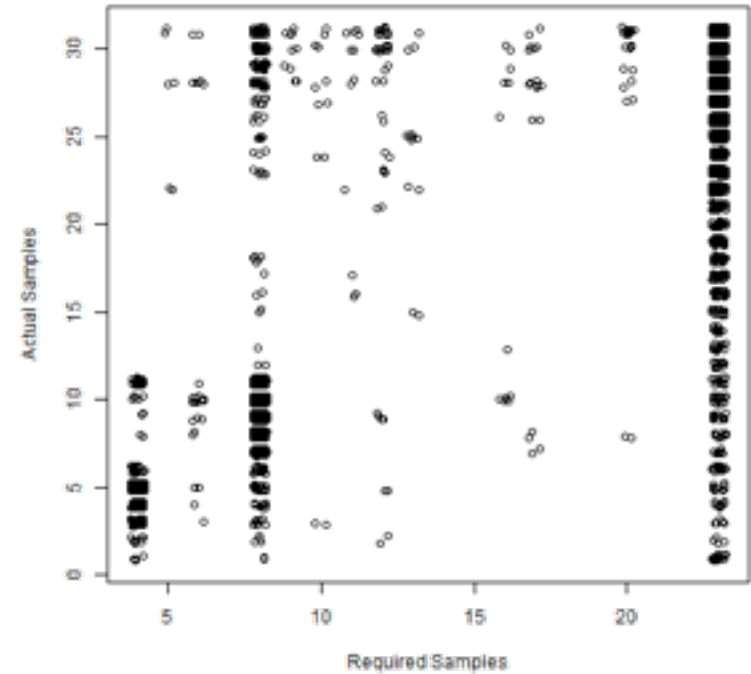
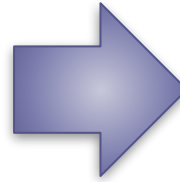
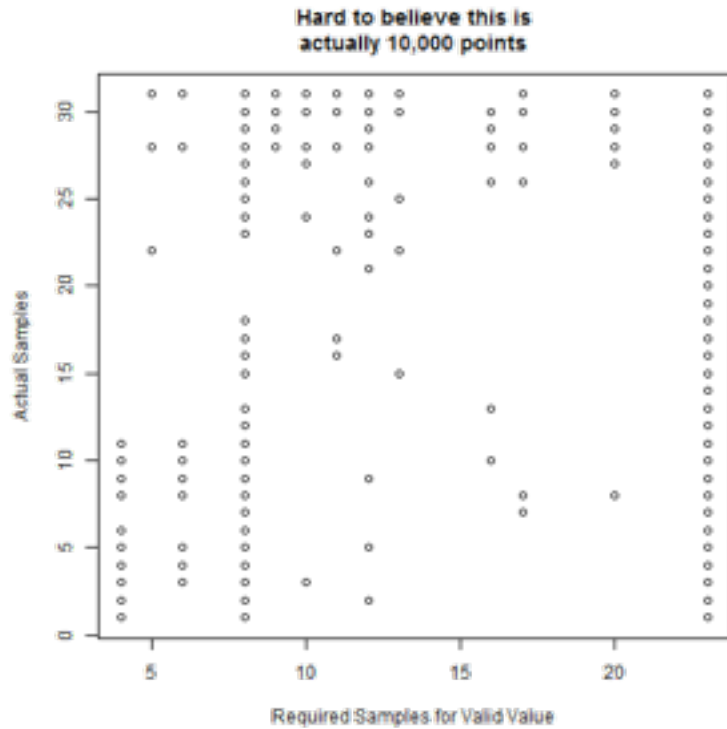
- Modify marker display
  - *Reduce marker size*
  - *Change marker shape*
  - Jittering
  - *Increase transparency*
- Indicate density of regions
  - *Use contour lines to indicate density*
  - *Use heatmap coloring to indicate density*
- Move to more dimensions?

# Jitter vs Overplotting

- You can also ‘jitter’ the data so that underlying data can be viewed making it easier to see patterns
- Keep in mind that
  - *jittering means adding additional noise to your data*
  - *can confuse your audience and lead to misinterpretation*
- jitter in-house fairly regularly but only sparingly jitter data in graphics we share with clients or the public.



# Jitter vs Overplotting



**Questions?**

*Thanks to  
Sophie J. Engle  
San Francisco University*

*for ideas, suggestions, slides, links, and much other stuff*