

Principles of Visualisation

Vik Gopal

*Everything should be made as simple as possible,
but no simpler.*

① Introduction

② Graphical Practice

- Good Graphs
- Bad Graphs

③ A Theory of Data Graphics

- Data-Ink
- Chart Junk
- Small Multiples
- Aesthetics and Technique
- Applications of Data-Ink Principles
- Thinking About Your Graph

④ Summary

About This Chapter

- Data graphics visually display measured quantities by means of the combined use of points, lines, a coordinate system, numbers, symbols, words, shading and colour.
- What we learn about in this class will not cover all the possible types of graphs you could (or should not) make.
- What we will try to do in this lecture, is this:
 - ▶ Show some good graphics.
 - ▶ Show some bad graphics.
 - ▶ Try to list down some principles of data graphic design.

Reference Material

- Most of this material is taken from books by *Edward Tufte*:
 - ▶ *Visual Display of Quantitative Information*
 - ▶ *Envisioning Information*
 - ▶ *Visual Explanations*
 - ▶ *Beautiful Evidence*
- You might not agree with some/most/all of them, but principles in this book are closely adhered to by the ggplot2 authors.
- The principles serve as a good guide when we need to draw up our own graphic.
- Another source for the material in this chapter are the these two books by Alberto Cairo:
 - ▶ *The functional art*
 - ▶ *The truthful art*
- The final source for this chapter is the excellent blog by *Nathan Yau*:
<https://flowingdata.com/>.
 - ▶ The “Guides” section is extremely useful for practitioners.

① Introduction

② Graphical Practice

- Good Graphs
- Bad Graphs

③ A Theory of Data Graphics

- Data-Ink
- Chart Junk
- Small Multiples
- Aesthetics and Technique
- Applications of Data-Ink Principles
- Thinking About Your Graph

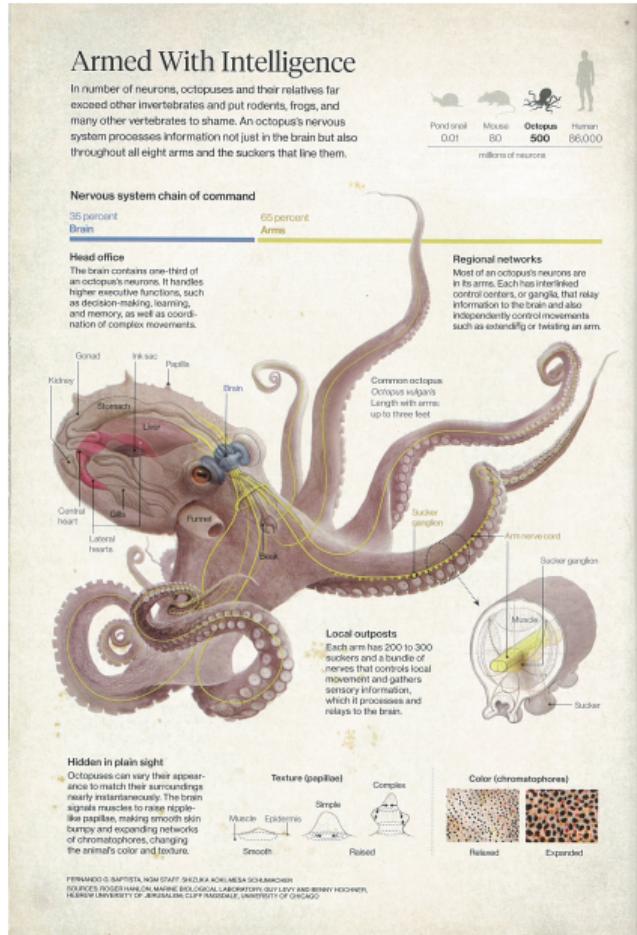
④ Summary

Some Loose Definitions

Let's work with these terms, so that we have some common understanding.

- A **visualisation** is any kind of visual representation of information designed to enable communication, analysis, discovery, exploration, etc.
- A **graph** or a **chart** is a display in which data are encoded with symbols that have different shapes, colours or proportions.
- A **map** is a depiction of a geographical area or a representation of data that pertains to that area.
- An **infographic** is a multi-section visual representation of information intended to communicate one or more specific messages.

- This is an example of an infographic.
- What do you think is good or bad about it? And why?



What Makes a Good Graph?

We should have certain expectations of a graph, that we study or make ourselves.

Graphical displays should:

- ① show the data.
- ② induce the viewer to think about the substance rather than about methodology, graphic design, the software used, etc.
- ③ avoid distorting what the data have to say.
- ④ present many numbers in a small space.
- ⑤ make large data sets coherent.
- ⑥ encourage the eye to compare different pieces of data.
- ⑦ reveal the data at several levels of detail, from a broad overview to the fine structure.
- ⑧ serve a reasonably clear purpose: description, exploration, tabulation or decoration.
- ⑨ be closely integrated with the statistical and verbal descriptions of a dataset.

Types of Graphical Displays

We can broadly categorize graphics into the following groups:

- **Data maps:** These were the first types of graphics drawn. They superimposed attributes onto geographical data, allowing us to study clusters, hotspots, etc.
- **Time series:** This is probably the most frequently used form of graphic design.
- **Narrative graphics of space and time:** These add spatial dimensions to time series displays. When done well, viewers will hardly be aware that they are looking at a world of four or even five dimensions.
- **Relational data:** These are more abstract, putting one variable (not necessarily latitude or time) on the X-axis and another (not longitude) on the Y-axis.

Let us take a look at some examples of these types of graphics.

① Introduction

② Graphical Practice

- Good Graphs
- Bad Graphs

③ A Theory of Data Graphics

- Data-Ink
- Chart Junk
- Small Multiples
- Aesthetics and Technique
- Applications of Data-Ink Principles
- Thinking About Your Graph

④ Summary

Cholera in Central London



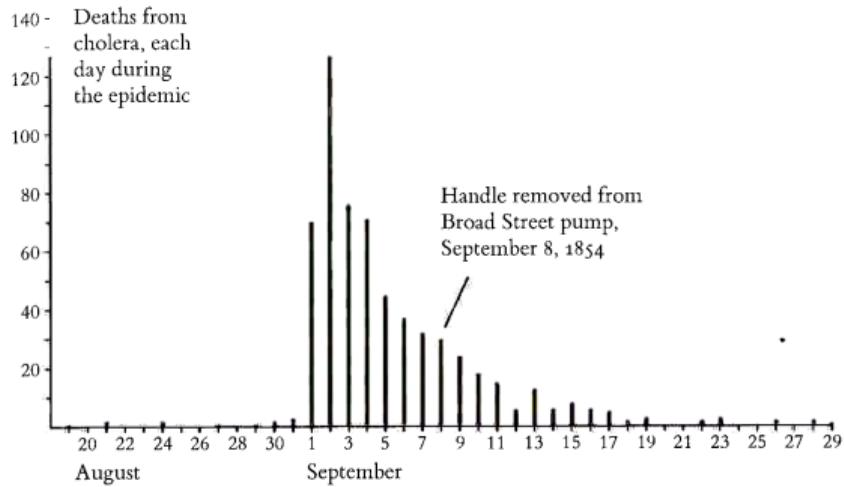
Cholera in Central London

cont'd

- Dr John Snow plotted the locations of deaths from cholera in central London in September 1854.
- He observed that cholera occurred almost entirely among those who lived near the Broad Street water pump.
- He ordered its removal, after which the epidemic ended.
- In the graph, X marks the location of pumps, and dots mark the location of cholera cases.
- Notice how we stacks up cases if they occurred in the same house/unit.
- Over the years, stories of the impact of this graphic have been rather embellished. I encourage you to read more about this case on your own.
- Nonetheless, it is accurate, and beautifully captures the situation at the time.

Cholera in Central London

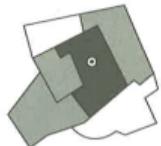
time series of deaths



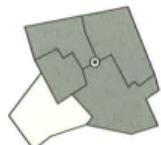
- The time series of deaths show that the number of deaths had declined greatly even before his instruction to remove the Broad Street pump.
- This was due to mass evacuation from London.

Cholera in Central London

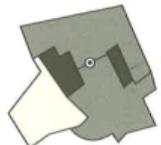
distortions



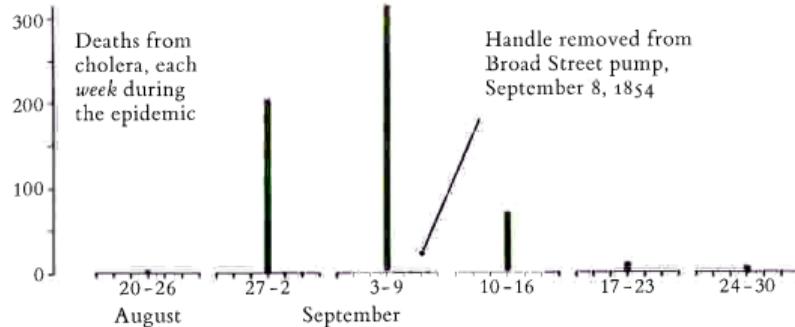
In this aggregation of individual deaths into six areas, the greatest number is concentrated at the Broad Street pump.



Using different geographic subdivisions the cholera numbers are nearly the same in four of the five areas.



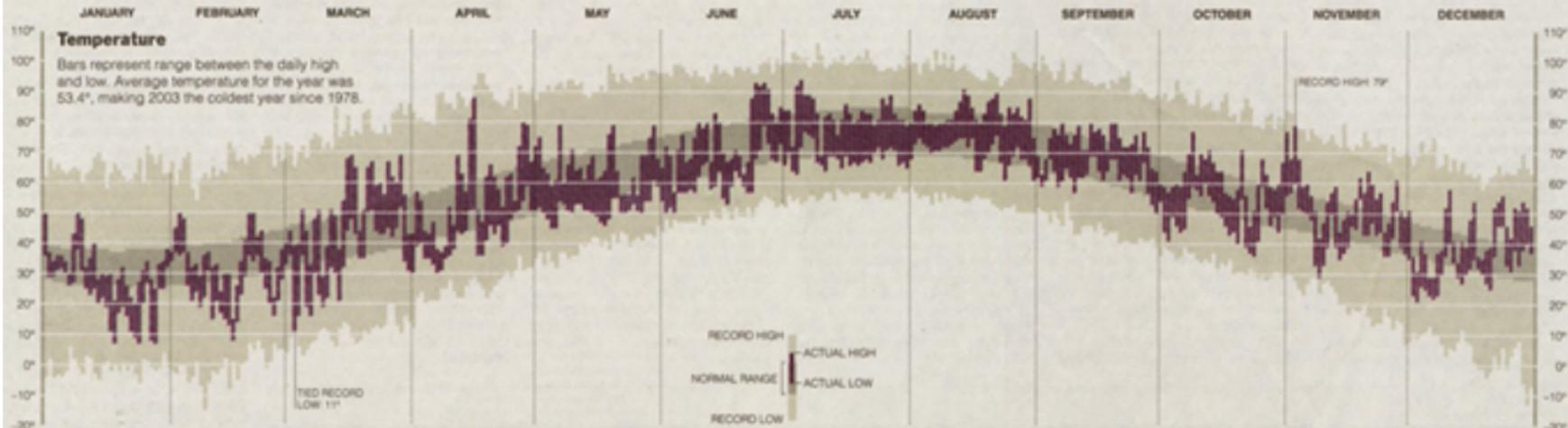
In this aggregation of the deaths, the two areas with the most deaths do not even include the infected pump!



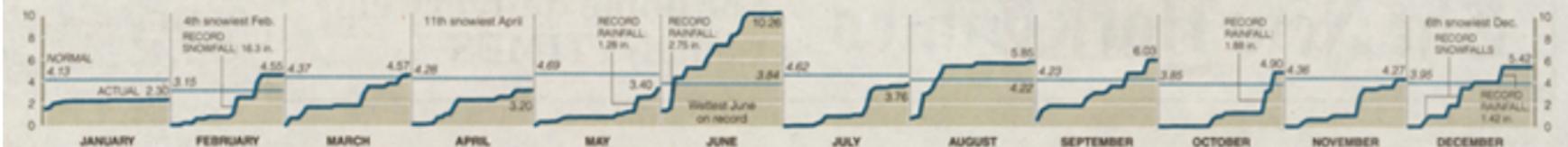
- Above, a temporal aggregation credits John Snow for **ending** the epidemic.
- On the left, a careless choice of spatial aggregation would fail to identify the guilty pump.

New York Weather

New York City's Weather in 2003



Precipitation Cumulative monthly precipitation in inches compared with normal monthly precipitation. Total precipitation in 2003 was 58.51 inches, 8.82 inches more than normal, which makes the year the sixth wettest on record.



New York Weather

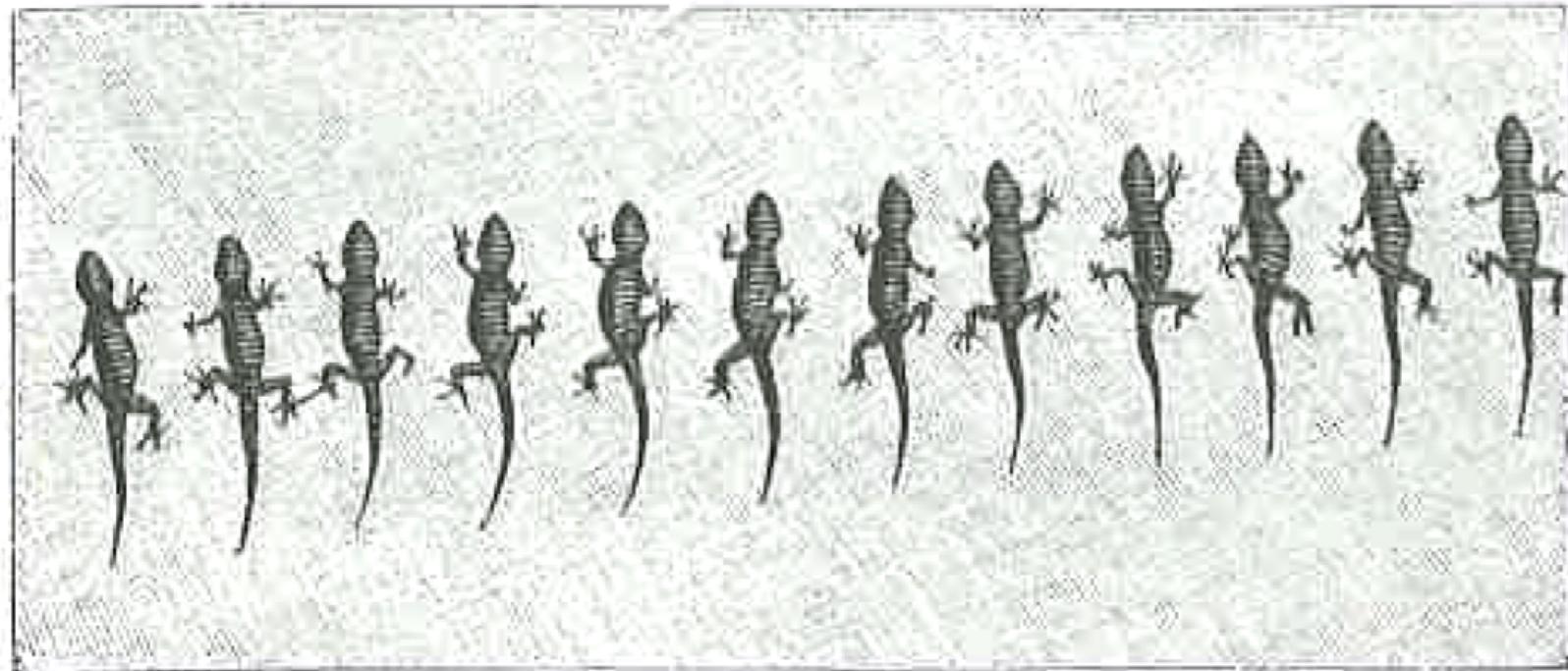
cont'd

- The time series graphic conveys different scales of information:
 - ▶ How the temperature varies over the year.
 - ▶ How rainfall varies from month to month.
- On *any day*, we are given the opportunity to compare the weather in 2003 with typical weather in New York, as well as the extreme weather on that day.
- The displays encourages comparisons within and between the time series presented.

Gecko Advancement

time series?

A different sort of progression through time is depicted in this superb graphic:



Gecko Advancement

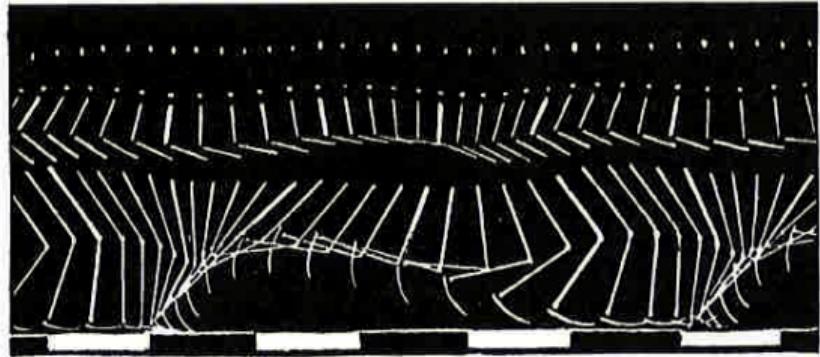
cont'd

- The use of small multiples of the same gecko emphasize the changes in foot position.
- It is even possible to tell how far it travels with each cycle.
- The repeated initial sequences convey the continuity of the motion.

Man In Black

time series

A similar graphic portrays the movements involved in running.



Man In Black

cont'd

This is a 3D version of the running man. It was at the iLight Festival 2022 (near MBS).

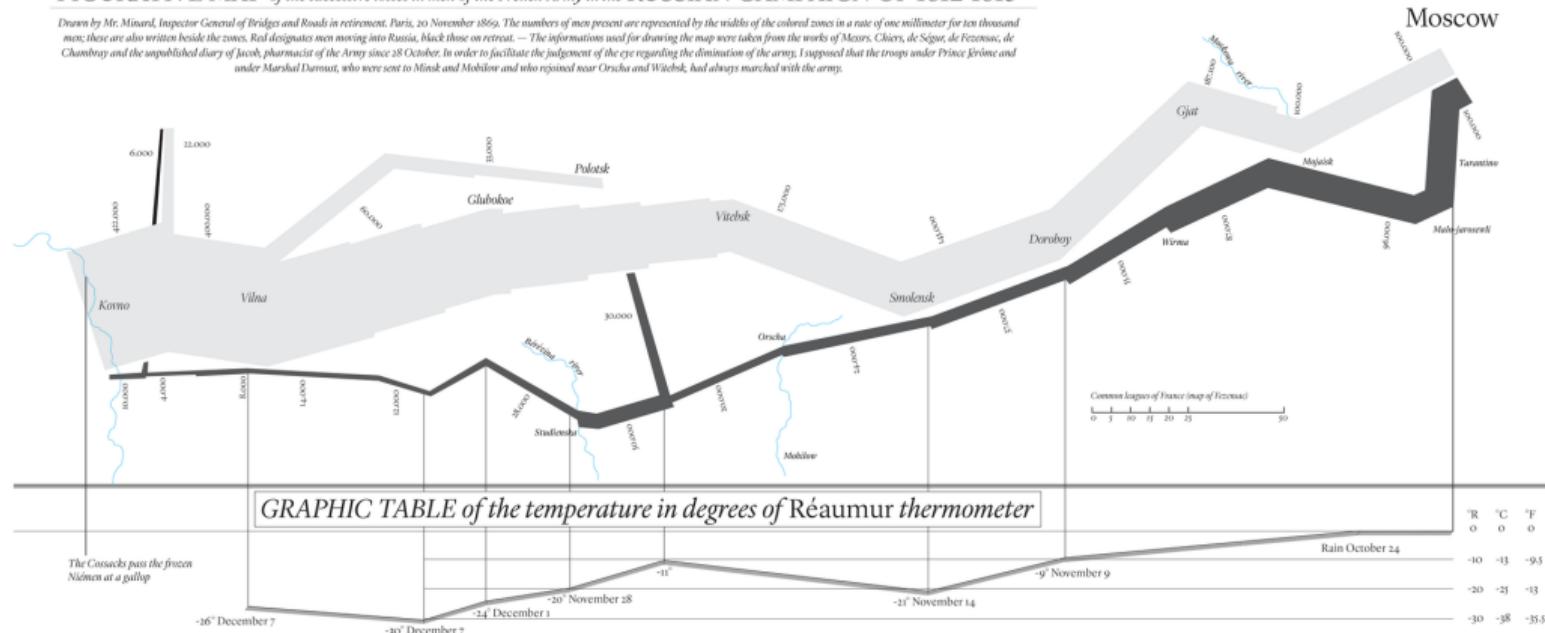


Napoleon's Army in Russia

space-time narrative

FIGURATIVE MAP of the successive losses in men of the French Army in the RUSSIAN CAMPAIGN OF 1812-1813

Drawn by Mr. Minard, Inspector General of Bridges and Roads in retirement, Paris, 20 November 1869. The numbers of men present are represented by the widths of the colored zones in a rate of one millimeter for ten thousand men; these are also written beside the zones. Red designates men moving into Russia, black those on retreat. — The informations used for drawing the map were taken from the works of Messrs. Clairs, de Ségur, de Férouzeac, de Chambry and the unpublished diary of Jacob, pharmacist of the Army since 28 October. In order to facilitate the judgment of the eye regarding the diminution of the army, I supposed that the troops under Prince Jérôme and under Marshal Davout, who were sent to Minsk and Mohilev and who rejoined near Orscha and Witebsk, had always marched with the army.



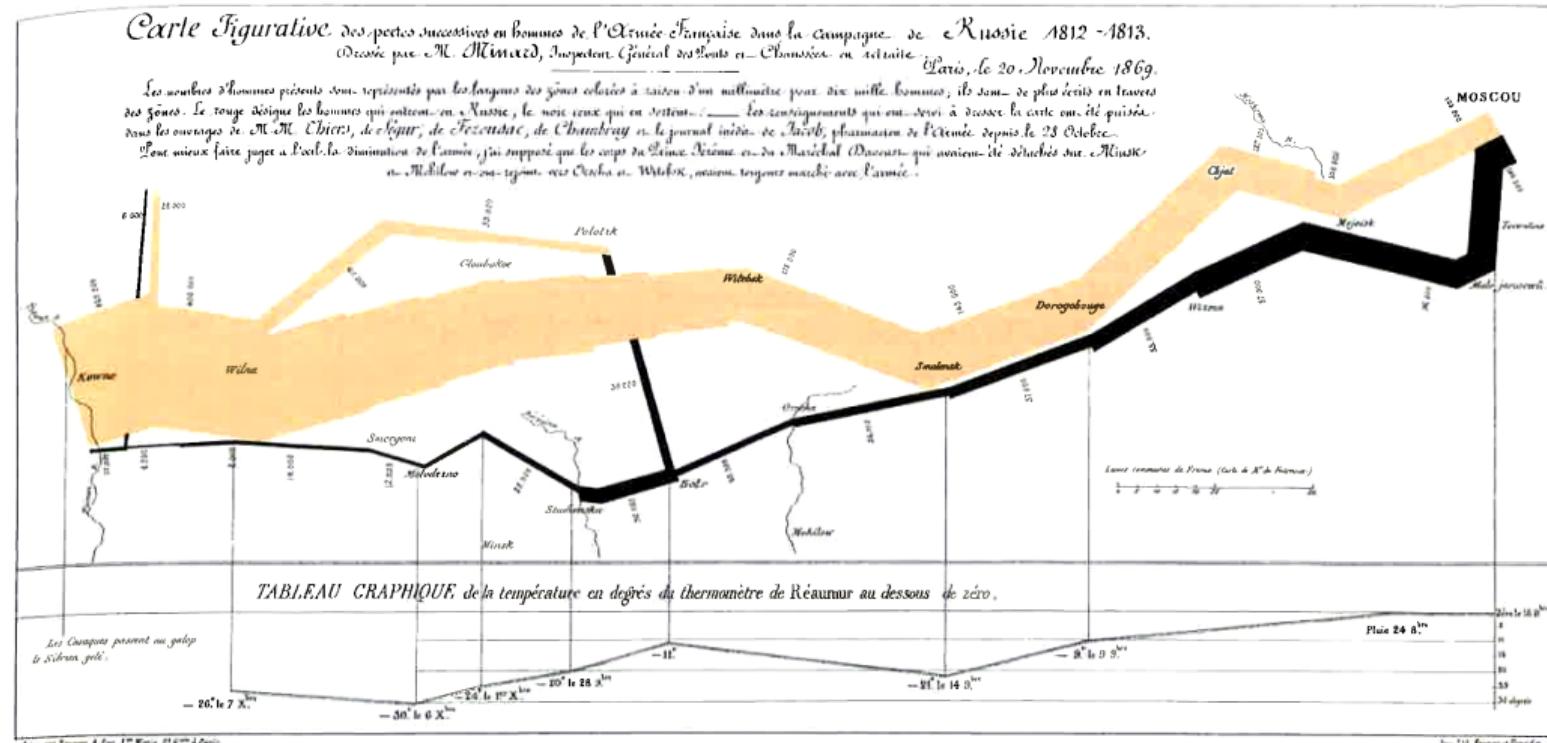
Napoleon's Army in Russia

cont'd

- The graphic by Charles Minard shows the losses suffered by Napoleon's Russian campaign in 1812.
- At the beginning, he has 422,000 soldiers, indicated by the width of the band.
- When he reaches Moscow, there are only 100,000.
- On the way back, the numbers dwindle even more. Close to home, crossing one particular river wipes out almost half of what remains.
- Six variables are plotted: size of army, location on a map, direction of movement, temperature on various dates.

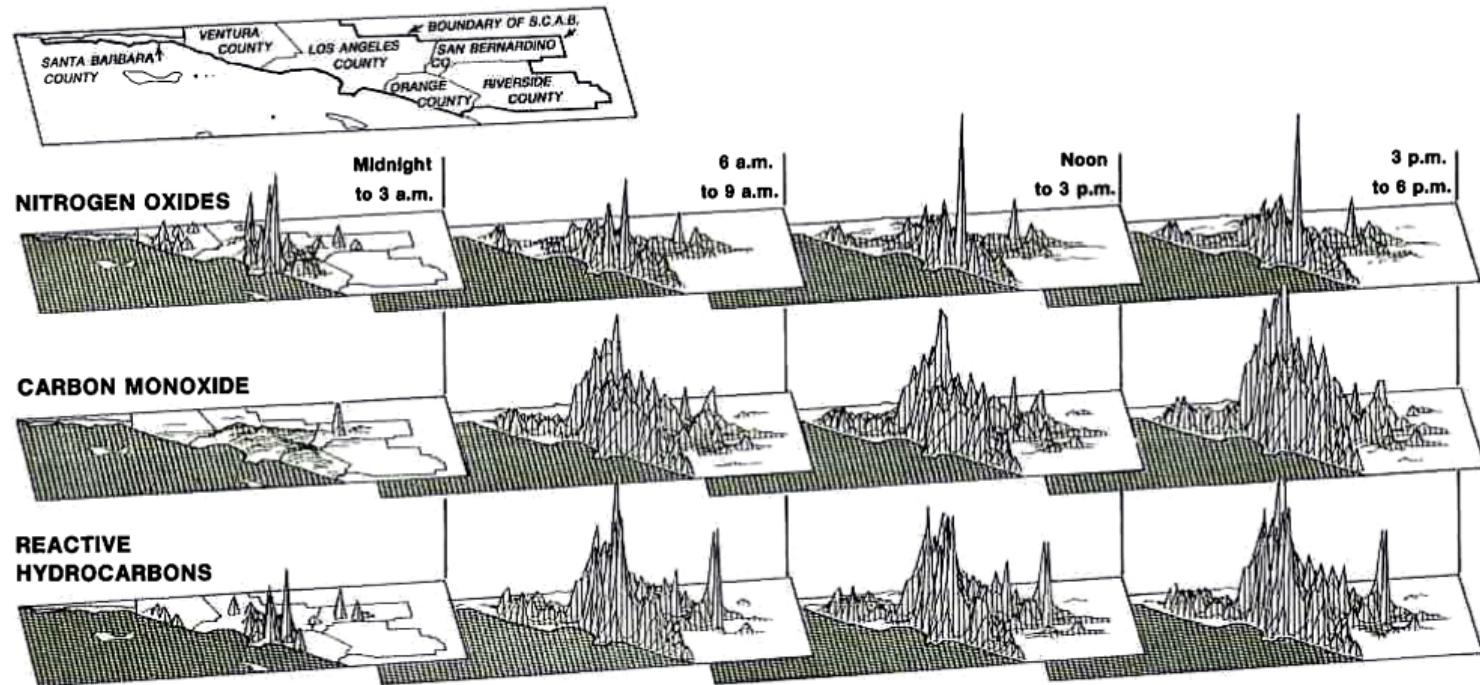
Napoleon's Army in Russia

original

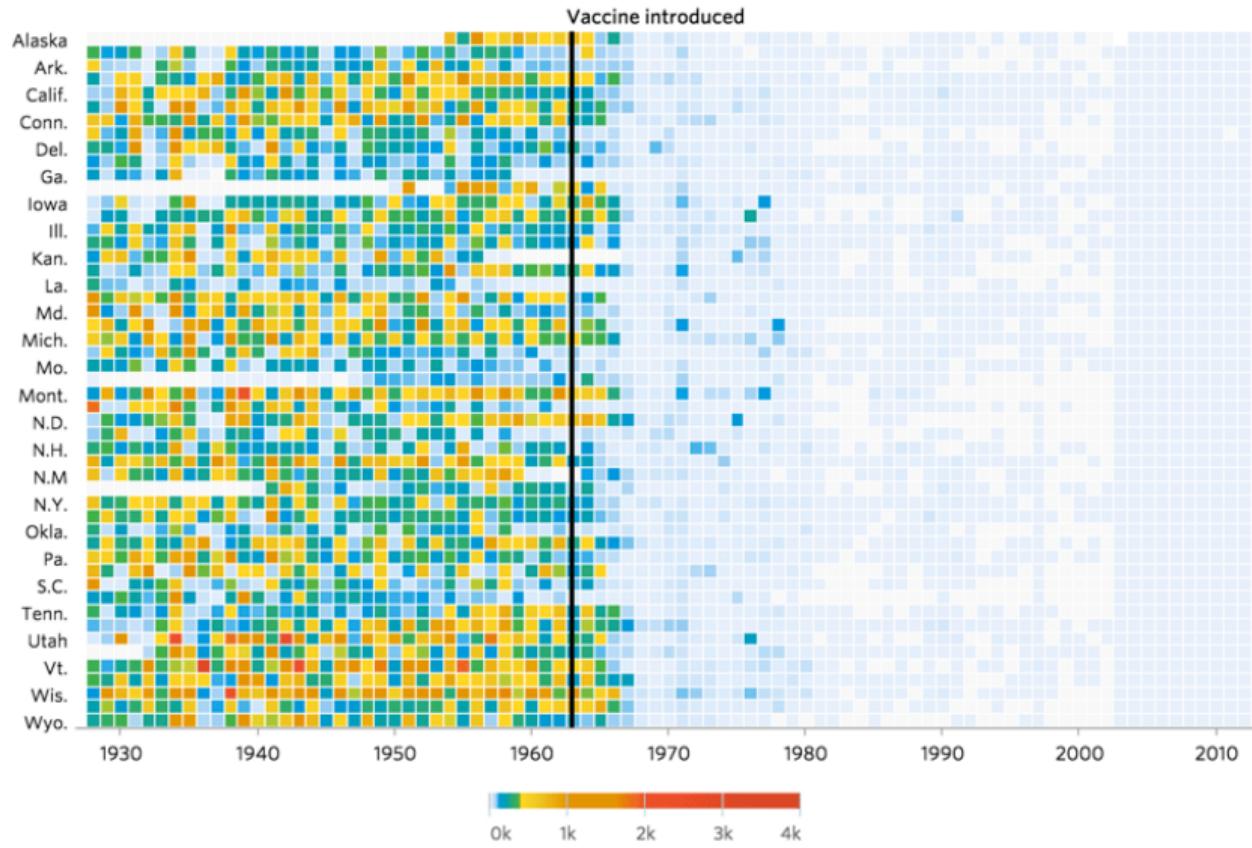


Pollution in LA

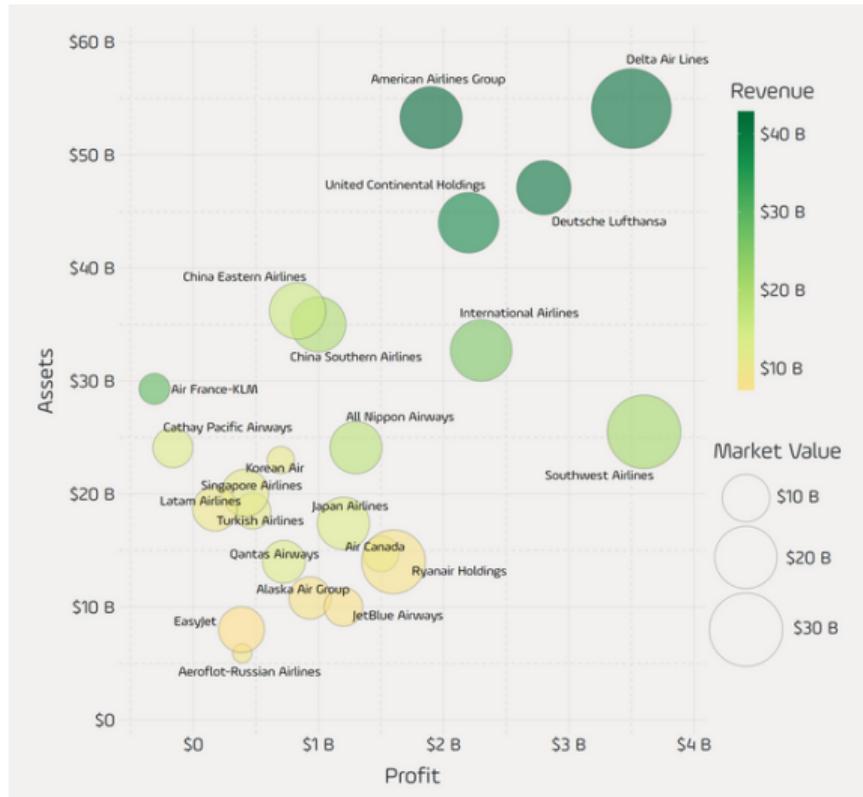
space-time narrative



Measles

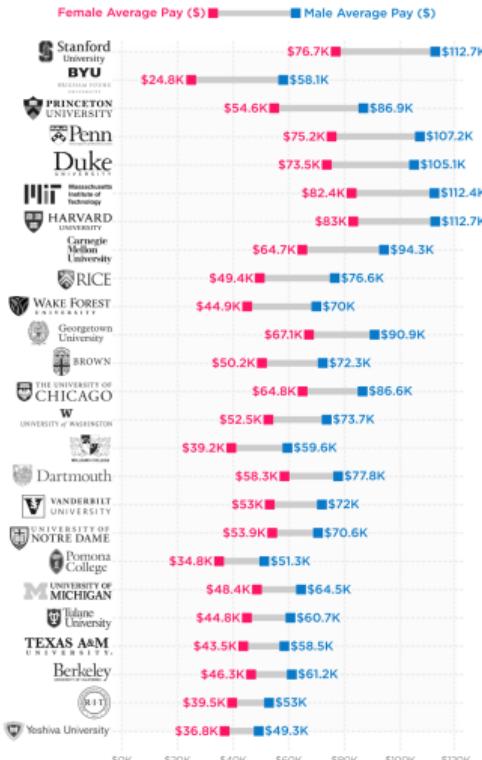


Airline Revenue



Gender Gap

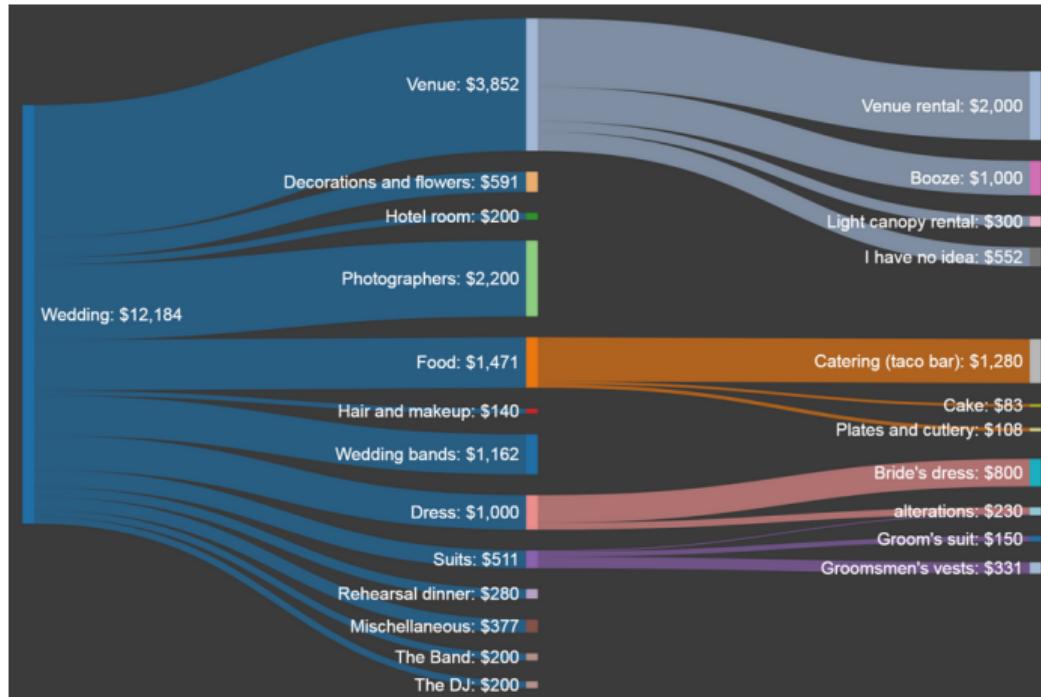
Colleges with the Biggest Gender Gap in Earnings from Graduates
Median Pay by Gender for Alumni 6 Years from Entry to School



Article & Sources:
<https://howmuch.net/articles/americas-colleges-gender-wage-gap>
<https://www.businessstudent.com>

howmuch.net

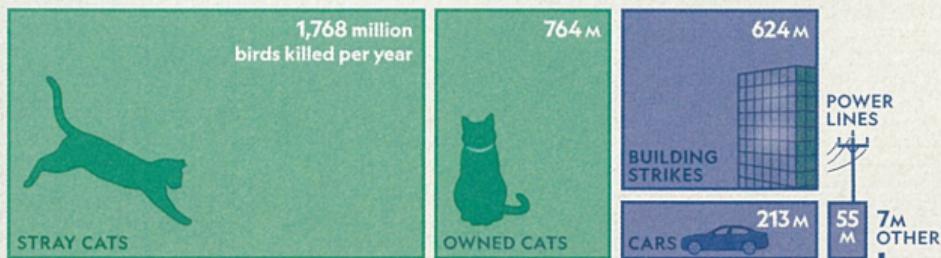
Sankey Chart



Tree Map

Death Toll

Researchers compiled data from multiple studies to estimate the number of birds in the U.S. and Canada killed in a year as a result of human activity.



*Communication towers and wind turbines

DIANA MARQUES, NGM STAFF; RYAN WILLIAMS. SOURCE: SCOTT LOSS, TOM WILL, AND PETER MARRA, ANNUAL REVIEW OF ECOLOGY, EVOLUTION, AND SYSTEMATICS, 2015

Principles of Graphical Excellence

Graphical excellence

- is the well-designed presentation of interesting data.
- consists of complex ideas communicated with clarity, precision and efficiency.
- is that which gives the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space.
- is nearly always multivariate.
- requires telling the truth about the data.

Let's take a look at some bad graphs.

① Introduction

② Graphical Practice

- Good Graphs
- Bad Graphs

③ A Theory of Data Graphics

- Data-Ink
- Chart Junk
- Small Multiples
- Aesthetics and Technique
- Applications of Data-Ink Principles
- Thinking About Your Graph

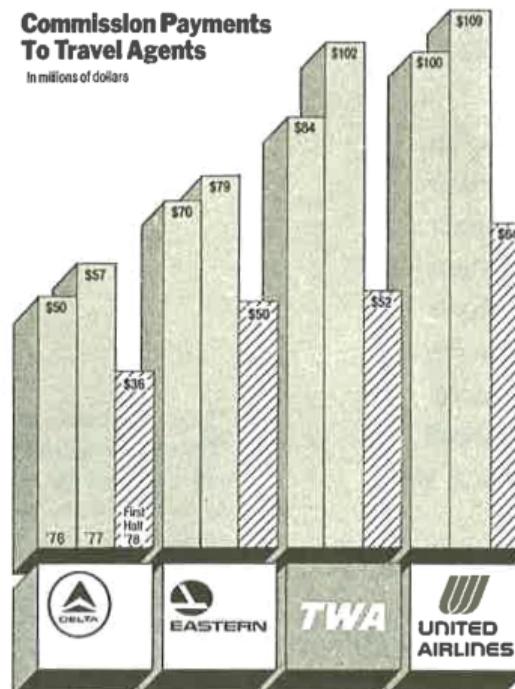
④ Summary

The Dominant Myth Around Statistical Graphics

- Until recently, statistical graphics were regarded with great suspicion.
- This was because many graphics makers worked with the assumption that graphics had to be “alive”, to be “communicatively dynamic”.
- If not, the belief was that the audience would fall asleep.
- This belief led to deceptive graphics, that in turn led to more suspicion of graphics.

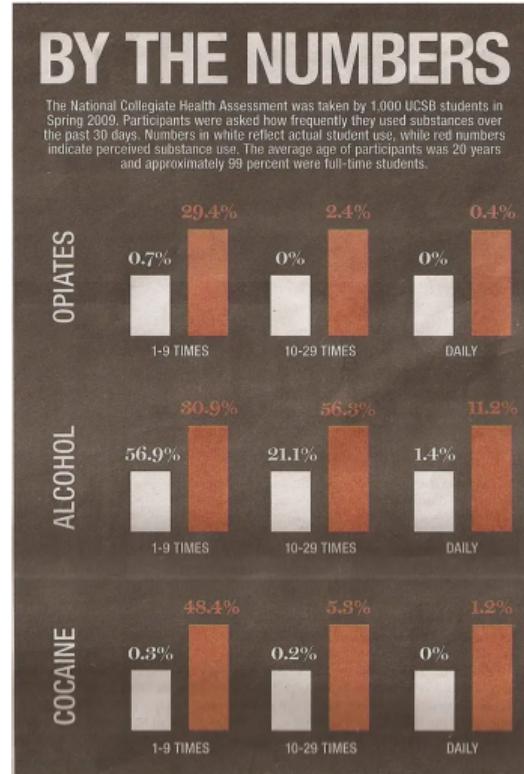
Inconsistent Basis of Comparison

half a year compared to full years



Design Variation

design variation not consistent with data variation



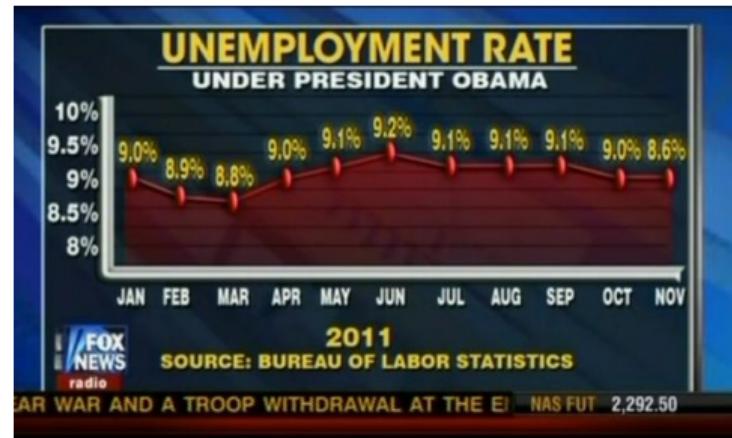
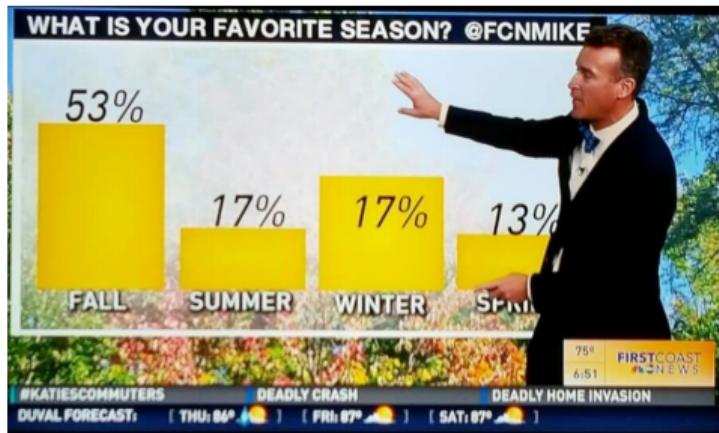
Distortion in Data Graphics

- A graphic does not distort if the visual representation of the data is consistent with the numerical representation.
- When we present a graphic, we cannot alter the visual perception of the data:
 - ▶ It has been established that the reported perception of something as clear and simple as line length depends on the context and what other people have already said about the lines.
- However, it is our duty to ensure that the visual image represents the underlying numbers.

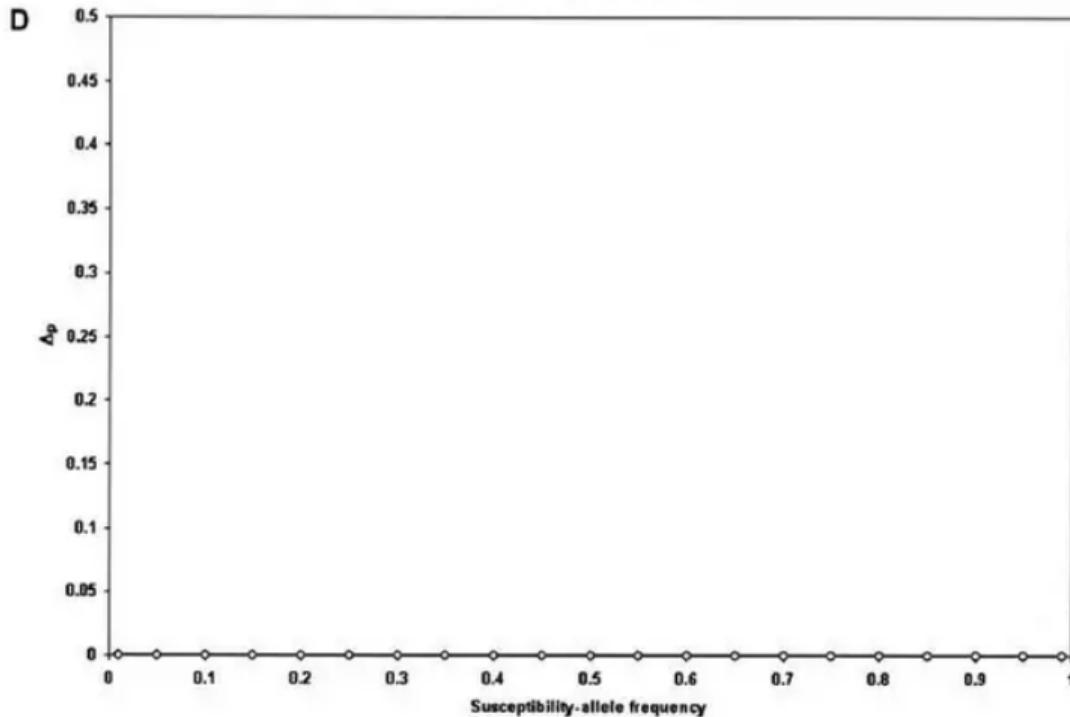
Design and Data Variation

- Each part of a graphic generates visual expectations about its other parts.
- For instance, a length of 1 cm on the axis on the left side of the graph should represent the same distance on the right side.
- If it does not, then we have design variation. This is bad, since the viewer can be easily deceived if he/she does not catch the variation.

Graphs with Dubious Integrity



An Unnecessary Chart



What Can You Understand From This Infographic?

WHERE YOUR FOOD COMES FROM

Imports of foods have doubled in a decade and now account for a fifth of what Americans eat.

A large percentage of these foods that Americans eat are imported.

FRUIT AND NUTS

51%



HONEY
61%



FRESH VEGETABLES/
MELONS
20%



LAMB
52%



SEAFOOD
88%



These countries are the largest exporters of each food.

FRUIT

MEXICO



CHILE



26% SUPPLIED

VEGETABLES

MEXICO



CANADA



50%

22%

SEAFOOD

CHINA



Vietnam

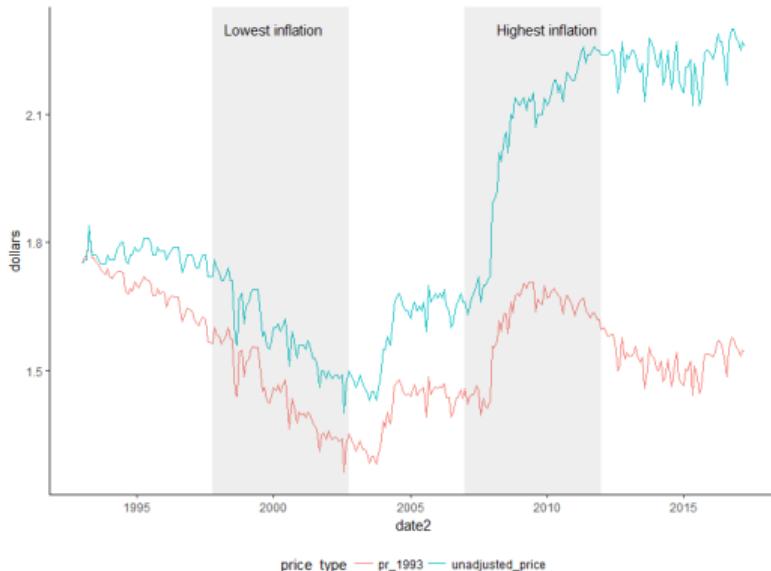


16%

6.5%

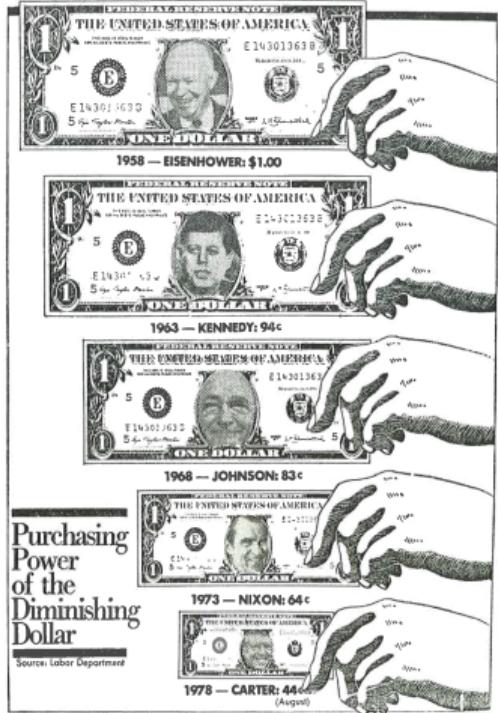
Source:
USDA Economic Research Service

Instant Noodles in Singapore



- On the left is a chart from a previous year's class.
- We used data from MTI to adjust the price of instant noodles in Singapore.
- Failing to account for the high inflation in those 10 years would have depicted a worrying trend in the increase in price of cup noodles.

Visual Area and Numerical Measure



- Should we be comparing their areas, lengths or breadths?

Spoken Languages



Reasons for Incompetent Graphics

- A possible lack of quantitative skills among the graphic designers.
- The doctrine that statistical data are boring. This leads to the exaggeration of evidence that we saw earlier. Here are some choice quotes:
 - ① *The challenge is to present statistics as a visual idea rather than a tedious parade of numbers.*
 - ② *Why are statistics so boring?*
- The doctrine that graphics are only for the unsophisticated reader:
 - ① *If you have to explain it, don't use it.*
 - ② *(We placed more emphasis) on graphics than on information. We had feared children might be overwhelmed by too many facts.*
- This contempt for graphics and their audience leads to unsatisfactory displays.

Graphical Integrity

Here are some guides to ensure the integrity of the graphic you create:

- ① The representation of numbers, as physically measured on the surface of the graphic itself, should be directly proportional to the numerical quantities represented.
- ② Clear, detailed and thorough labeling should be used to defeat graphical distortion and ambiguity.
If necessary,
 - ▶ write out explanations of the data on the graphic itself.
 - ▶ label important events in the data.
- ③ Show data variation, not design variation.
- ④ In time-series displays of money, deflated and standardised units of monetary measurements are nearly always better than nominal units.
- ⑤ The number of information-carrying dimensions depicted should not exceed the number of dimensions in the data.
- ⑥ Graphics must not quote data out of context.

Common Visualisation Cheats

- The Flowing Data blog has a good section on what could be misleading in visualisations.
- <https://flowingdata.com/2017/02/09/how-to-spot-visualization-lies/>

Common Visualisation Cheats

cont'd

- Stay away from dual axes.

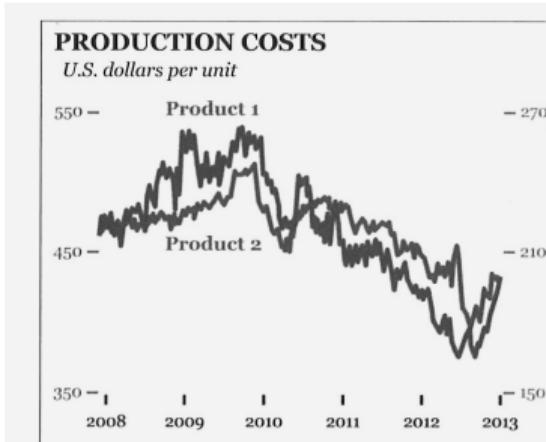


Figure 2.4 Dual-axis charts can be easily misinterpreted.

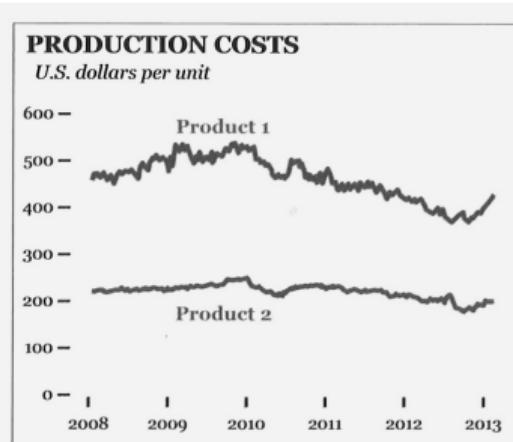


Figure 2.5 The same data, plotted on the same scale. As a general rule, avoid dual-axis charts.

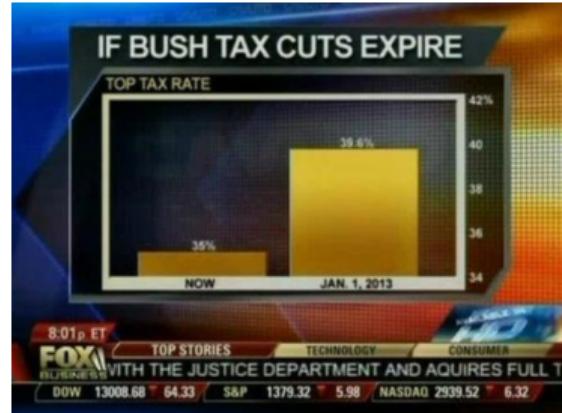
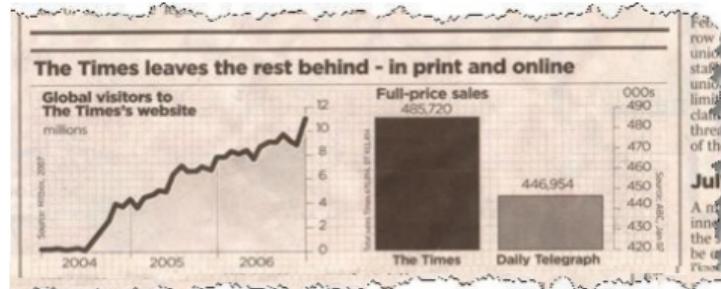
- Avoid pie charts.

"The only worse design than a pie chart is several of them."
—Edward Tufte.

Common Visualisation Cheats

cont'd

- Bar charts that do not begin with 0 can mislead.



① Introduction

② Graphical Practice

- Good Graphs
- Bad Graphs

③ A Theory of Data Graphics

- Data-Ink
- Chart Junk
- Small Multiples
- Aesthetics and Technique
- Applications of Data-Ink Principles
- Thinking About Your Graph

④ Summary

① Introduction

② Graphical Practice

- Good Graphs
- Bad Graphs

③ A Theory of Data Graphics

- Data-Ink
- Chart Junk
- Small Multiples
- Aesthetics and Technique
- Applications of Data-Ink Principles
- Thinking About Your Graph

④ Summary

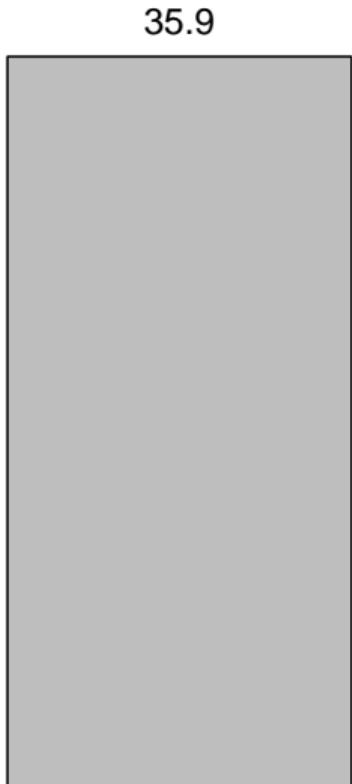
Data-Ink and Data-Ink Ratio

- *Data-ink* is the non-erasable core of a graphic, the non-redundant ink arranged in response to variation in the numbers represented.
- A large share of ink on a graphic should be used as data-ink.
- Tufte defines

$$\text{Data-ink ratio} = \frac{\text{data-ink}}{\text{total ink used to print graphic}}$$

- We should try to maximise the data-ink ratio, within reason.
- To do this, we could erase non-data-ink, also within reason.
 - ▶ In particular, redundant representations of the data should be erased.

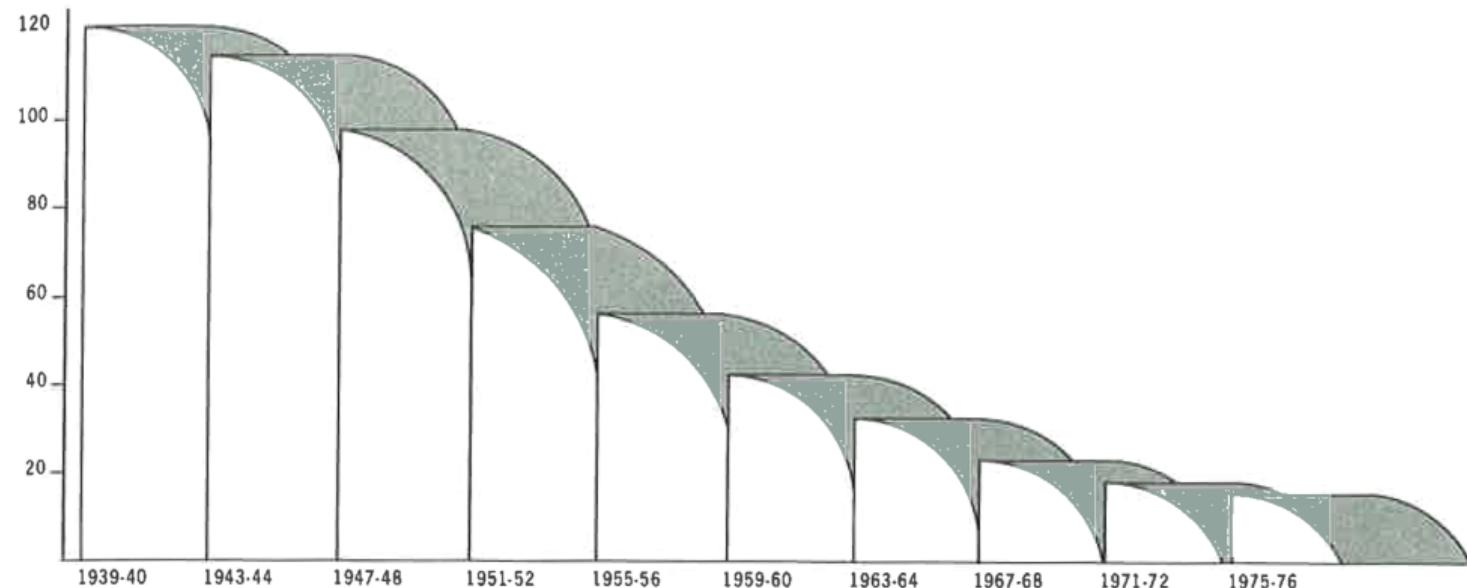
Redundant Data-Ink on A Bar



The bar represents the altitude in 6 different ways:

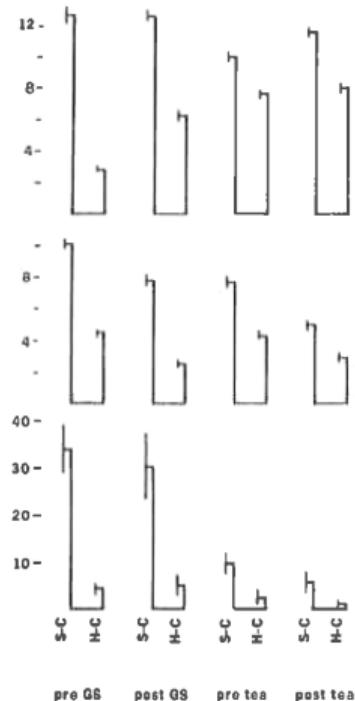
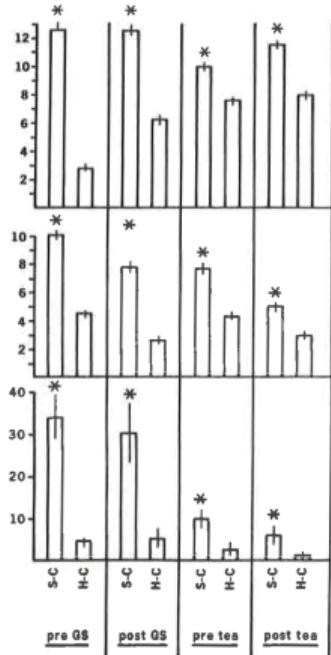
- ① Height of left lines.
- ② Height of right lines.
- ③ Height of shading.
- ④ Position of top horizontal lines.
- ⑤ Position of “35.9” text.
- ⑥ The number 35.9 text.

Redundant Ink for an Unnecessary Dimension



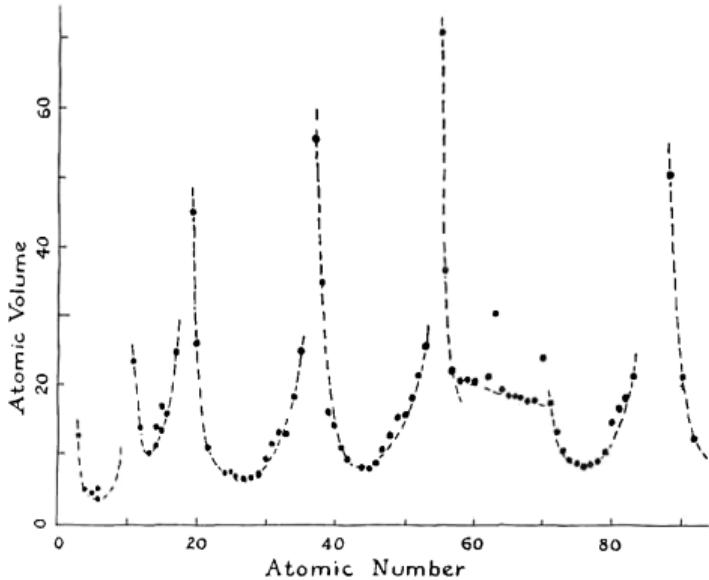
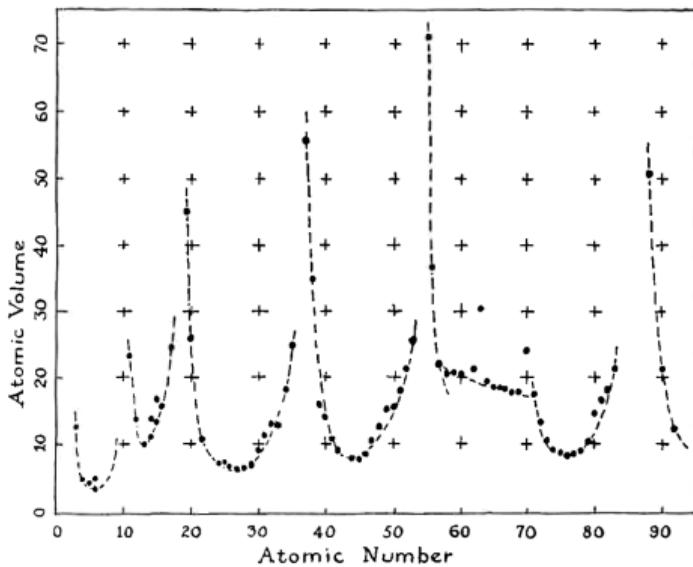
Application of Erase Principle

redesign of barcharts



Application of Erase Principle

removal of heavy grid



Data-Ink Principles

- ① Above all else show the data.
- ② Maximize the data-ink ratio.
- ③ Erase non-data-ink.
- ④ Erase redundant data-ink.
- ⑤ Revise and edit.

① Introduction

② Graphical Practice

- Good Graphs
- Bad Graphs

③ A Theory of Data Graphics

- Data-Ink
- **Chart Junk**
- Small Multiples
- Aesthetics and Technique
- Applications of Data-Ink Principles
- Thinking About Your Graph

④ Summary

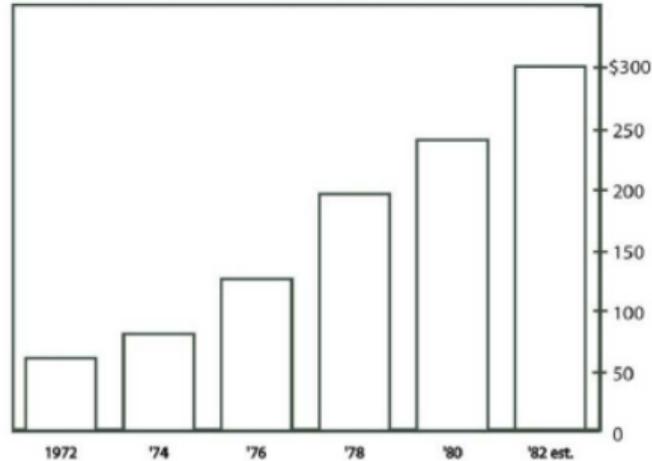
Introduction

- Most of the time, interior decoration of graphics is all non-data-ink or redundant data-ink, and it is referred to by Tufte as *chartjunk*.
- Graphical decoration, ... “comes cheaper than the hard work required to produce intriguing numbers and secure evidence.”
- The most common types of chart junk are
 - ▶ over-busy grid lines.
 - ▶ excess ticks
 - ▶ redundant representations of the simplest data
 - ▶ distracting patterns from computer software. (moire patterns)

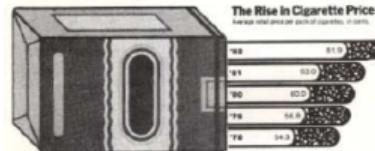
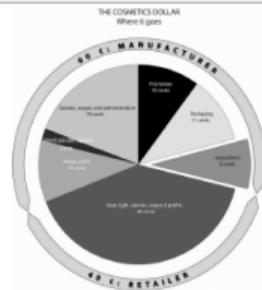
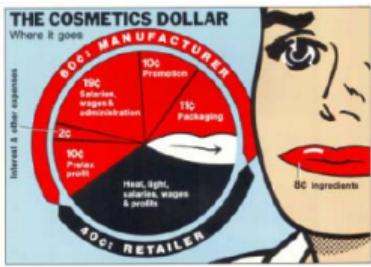
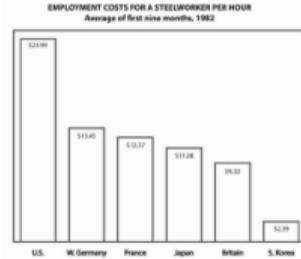
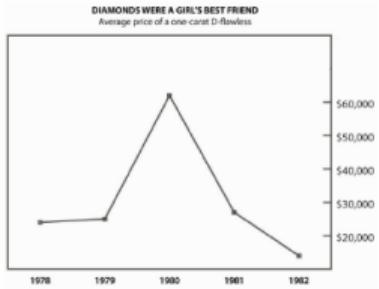
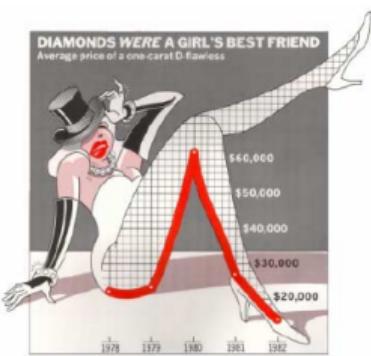
Which Is More Memorable to You?



MONSTROUS COSTS
Total House and Senate campaign expenditures, in millions

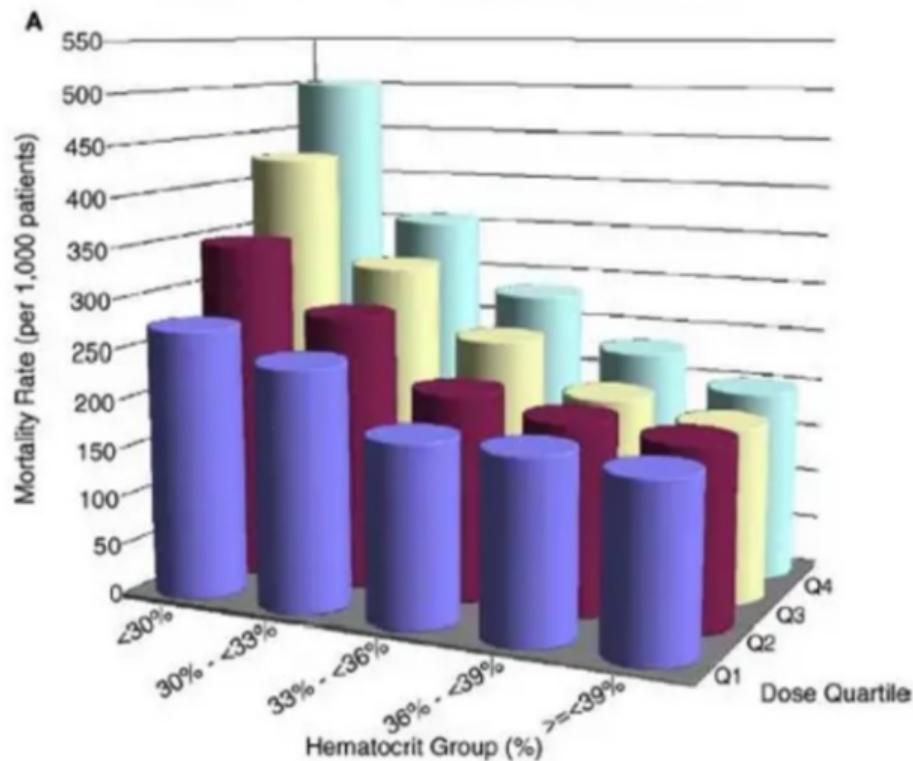


Edward Tufte vs. Nigel Holmes



3D is Almost Always Not Necessary

D.J. Coker et al. / Journal of Clinical Epidemiology 57 (2004) 1086–1095



Presenting Tables

- When presenting tables, we could reduce over-bearing grid lines by removing them, allowing the formatting to present the structure of the table.
- In most situations, we can do away with the vertical lines, unless the cells are too narrow to distinguish themselves from one another.
- On the next slide, the grid-lines seem to imprison the values in jail cells before the re-formatting.

Presenting Tables

cont'd

Before:

Train No.	3701	XM	3301	3801	A 67	3 3803	3201	3 51	3703	3807	3 3203	A 61	3809	A 47	3901	3811	3 3903	3813	3205	3815	3817	3819	3207	3821	3823	3825	3209	3827	3829	3831	
New York, N.Y.	12.10	12.40	1.30	3.52	4.50	6.10	6.25	6.35	6.50	7.10	7.30	7.33	7.45	7.47	7.59	8.04	8.19	8.39	8.54	9.04	9.24	9.54	10.24	10.39	10.54	11.24	11.54	12.04	12.24	12.54	1.24
Newark, N.J. P North Elizabeth Elizabeth	12.24	12.55	1.44	4.07	5.04	6.24	6.38	6.49	7.04	7.24	7.45	7.47	7.59	8.04	8.19	8.39	8.54	9.04	9.24	9.54	10.24	10.39	10.54	11.24	11.54	12.04	12.24	12.54	1.24		
Linden North Rahway Rahway	12.31	1.03	1.51	...	5.11	6.31	...	6.56	7.11	7.32	...	7.54	...	8.10	...	8.13	8.26	8.46	9.01	9.11	9.31	10.01	10.31	10.46	11.01	11.31	12.01	12.11	12.31	1.01	1.31
Metro Park (Iselin) Metuchen	12.44	...	2.04	4.26	5.24	...	6.56	7.10	7.25	...	8.04	8.07	8.15	...	8.40	...	9.14	...	9.44	10.14	10.44	...	11.14	11.44	12.14	...	12.44	1.14	1.44		
Edison New Brunswick Jersey Avenue	12.51	...	2.11	...	5.35	...	7.05	7.21	7.35	...	8.14	...	8.47	...	9.21	...	10.21	...	11.21	...	12.21	...	1.21	...	1.21	...	1.25	1.25	1.54		
Princeton Jct. S Trenton, N.J.	2.31	...	5.50	...	7.19	...	7.50	...	8.34	8.41	...	9.05	...	9.41	...	10.09	10.41	11.09	...	11.41	12.09	12.41	...	1.09	1.41	2.09			
	12.48	...	2.08	...	5.28	...	7.14	7.29	...	8.11	...	8.44	...	9.18	...	9.48	10.18	10.48	...	11.18	11.48	12.18	...	12.48	1.18	1.48	...	1.22	1.52	2.20	

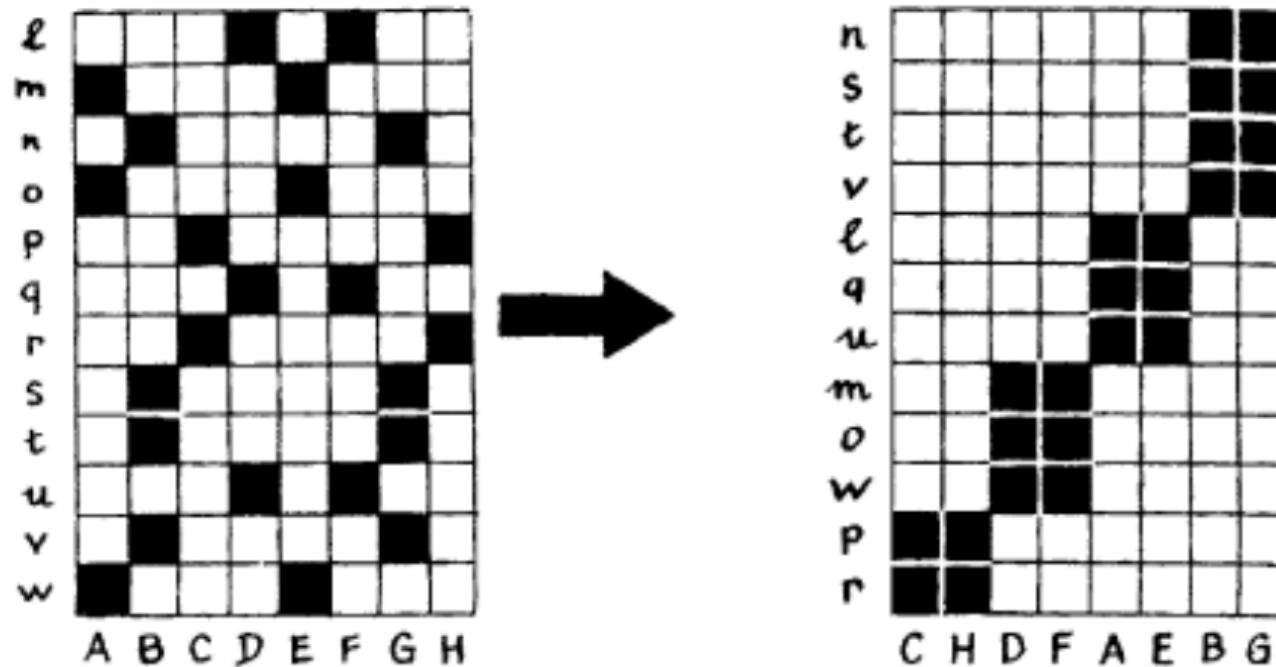
After:

	am	pm	12.10	12.40	1.30	3.52	4.50	6.10	6.25	6.35	6.50	7.10	7.30	7.33	7.45	7.50	* 8.05	8.25	8.40	8.50	9.10	9.40	10.10	10.25	10.40	11.10	11.40			
New York, NY	12.24	12.55	1.03	1.44	4.07	5.04	6.24	6.38	6.49	7.04	7.24	7.45	7.47	7.59	8.04	8.19	8.39	8.54	9.04	9.24	9.54	10.24	10.39	10.54	11.24	11.54	12.04	12.24	12.54	1.24
Newark, NJ P North Elizabeth Elizabeth	12.31	1.03	1.51	...	5.11	6.31	...	6.56	7.11	7.32	...	7.54	...	8.13	8.26	8.46	9.01	9.11	9.31	10.01	10.31	10.46	11.01	11.31	12.01	12.11	12.31	1.01	1.31	
Linden North Rahway Rahway	12.36	1.56	...	5.16	6.36	...	7.01	7.15	7.37	...	7.59	...	8.20	8.33	8.54	...	8.57	9.10	9.18	9.40	10.10	10.40	10.53	11.10	11.40	12.10	12.18	12.40	1.10	1.40
Metro Park (Iselin) Metuchen	12.44	2.04	4.26	5.24	...	6.56	7.10	7.25	...	8.04	8.07	8.15	...	8.40	...	9.14	...	9.44	10.14	10.44	...	11.14	11.44	12.14	...	12.44	1.14	1.44		
Edison New Brunswick Jersey Avenue	12.51	2.11	...	5.35	...	7.05	7.21	7.35	...	8.18	...	8.25	...	8.50	...	9.25	...	9.54	10.25	10.54	...	11.25	11.54	12.25	...	12.54	1.25	1.54		
Princeton Junction S Trenton, N.J.	...	2.31	...	5.50	...	7.19	...	7.50	...	8.34	8.41	...	9.05	...	9.41	...	10.09	10.41	11.09	...	11.41	12.09	12.41	...	1.09	1.41	2.09			
	12.48	2.42	4.58	6.03	...	7.28	...	8.01	...	8.31	8.44	8.52	...	9.16	...	9.52	...	10.19	10.52	11.19	...	11.52	12.19	12.52	...	1.22	1.52	2.20		

Presenting Tables

an aside

Ordering rows and columns can bring out the insights in the data.



Presenting Tables

an aside, cont'd

TEN TESTS OF EFFICIENCY

	CHILDREN IN SCHOOL	\$1000. IN PLANT	EXPENSE PER CHILD	30000 DAYS PER CHILD	SCHOOL YEAR	ATTEND- ANCE	EXPEND- ITURE AND WEALTH	DAILY COST	HIGH SCHOOLS	SAL- ARIES
1 WASHINGTON										
2 MASSACHUSETTS										
3 NEW YORK										
4 CALIFORNIA										
5 CONNECTICUT										
6 OHIO										
7 NEW JERSEY										
8 ILLINOIS										
9 COLORADO										
10 INDIANA										
11 RHODE ISLAND										
12 VERMONT										
13 NEW HAMPSHIRE										
14 UTAH										
15 OREGON										
16 MONTANA										
17 MICHIGAN										
18 N DAKOTA										
19 IDAHO										
20 MINNESOTA										
21 IOWA										
22 MAINE										
23 PENNSYLVANIA										
24 KANSAS										
25 NEBRASKA										
26 S. DAKOTA										
27 NEVADA										
28 WISCONSIN										
29 WYOMING										
30 ARIZONA										
31 OKLAHOMA										
32 MISSOURI										
33 W. VIRGINIA										
34 FLORIDA										
35 DELAWARE										
36 MARYLAND										
37 TENNESSEE										
38 TEXAS										
39 LOUISIANA										
40 NEW MEXICO										
41 VIRGINIA										
42 KENTUCKY										

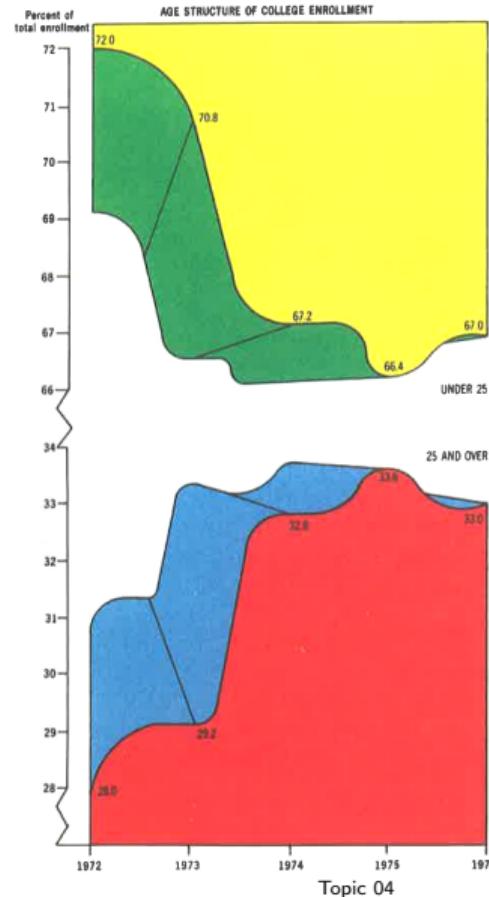
The Duck

- When a graphic is taken over by decorative forms, Tufte calls it a *duck*, in honour of this structure:



- The whole building is itself decoration.
- Similarly, the duck data graphic is one that is taken over by decoration.

A Prime Example of A Chart Duck



Architecture And Construction

- A dismissive cynic said that *architecture is the decoration of construction*.
- Even if that is true, we should abide by this principle:
It is all right to decorate construction, but never construct decoration.

Conclusions

- Graphics stand or fall on their content.
- Graphics do not become attractive and/or interesting through chart junk.
- Chart junk cannot rescue a thin data set.
- The best designs are provoke thought and repeated viewing, through
 - ▶ the story they tell.
 - ▶ the immense detail they contain.
 - ▶ simple presentation of interesting data.

Forgo chartjunk, including moire vibration, the grid, and the duck.

① Introduction

② Graphical Practice

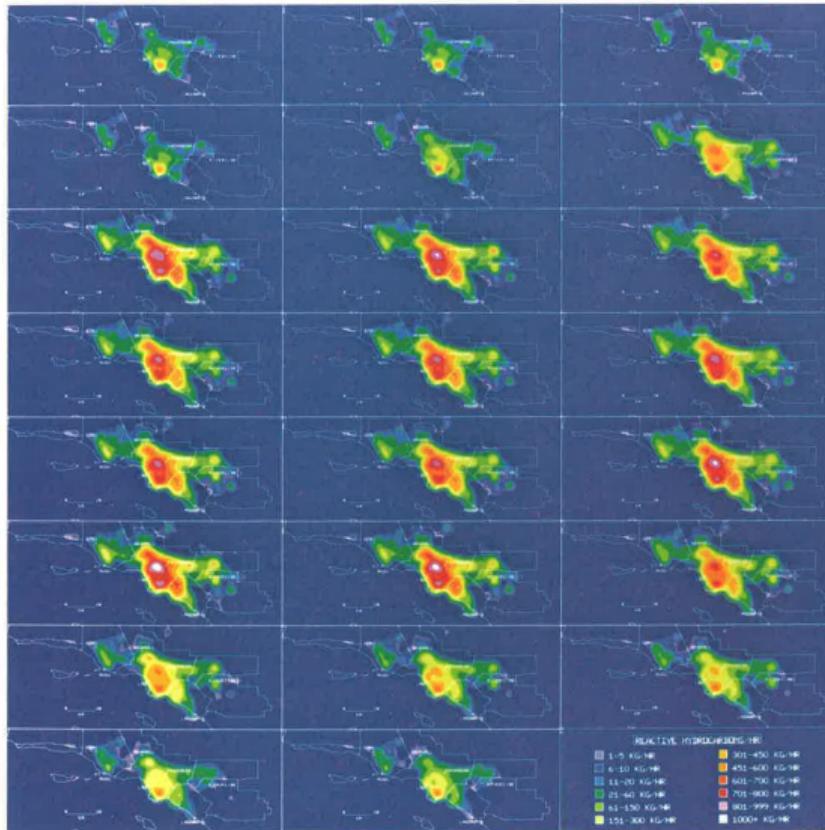
- Good Graphs
- Bad Graphs

③ A Theory of Data Graphics

- Data-Ink
- Chart Junk
- **Small Multiples**
- Aesthetics and Technique
- Applications of Data-Ink Principles
- Thinking About Your Graph

④ Summary

LA Pollution, 23 hours



LA Pollution, 23 hours

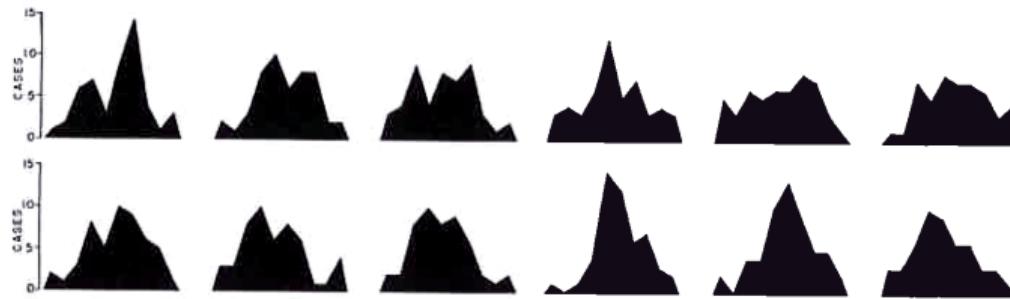
cont'd

- Small multiples are a series of graphics, showing the same combination of variables, indexed by changes in another variable.
- In the previous slide, our attention is instantly drawn to changes in data from one frame to the next.
- The design of each frame is constant.

Depicting Motion without Animation



Normal Random Variables



- The diagram shows 12 samples of 50 random normal deviates.
- The intention was to show the distributions can show such different (even multimodal) patterns despite all coming from the same distribution.
- This is conveyed, without axes, frames, colours.

① Introduction

② Graphical Practice

- Good Graphs
- Bad Graphs

③ A Theory of Data Graphics

- Data-Ink
- Chart Junk
- Small Multiples
- **Aesthetics and Technique**
- Applications of Data-Ink Principles
- Thinking About Your Graph

④ Summary

Friendly and Unfriendly Graphics

Friendly

- ① Words are spelt out whenever possible.
- ② Words run from left to right.
- ③ Little messages on the graphic explain the data.
- ④ Labels are on the graphic; no legend is required.
- ⑤ Graphic provokes curiosity of viewer.

Unfriendly

- ① Contains abbreviations.
- ② Words run vertically, or in several different directions.
- ③ Viewer needs to repeatedly refer to text to understand the graphic.
- ④ Legend is difficult to remember.
- ⑤ Filled with chart junk.

① Introduction

② Graphical Practice

- Good Graphs
- Bad Graphs

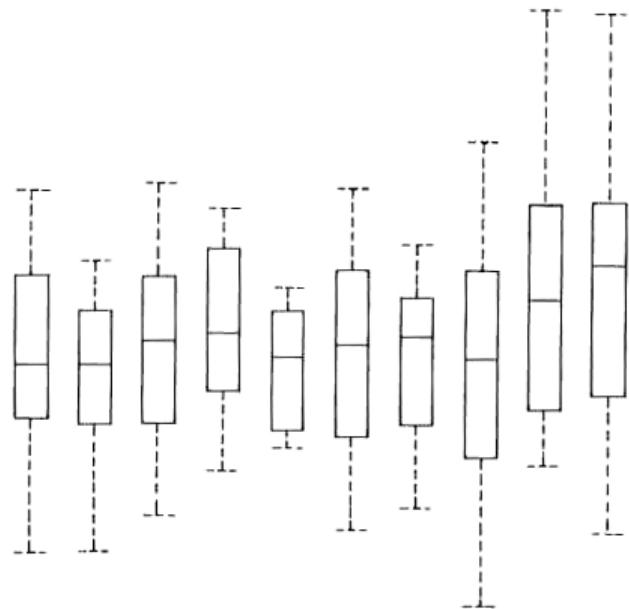
③ A Theory of Data Graphics

- Data-Ink
- Chart Junk
- Small Multiples
- Aesthetics and Technique
- Applications of Data-Ink Principles**
- Thinking About Your Graph

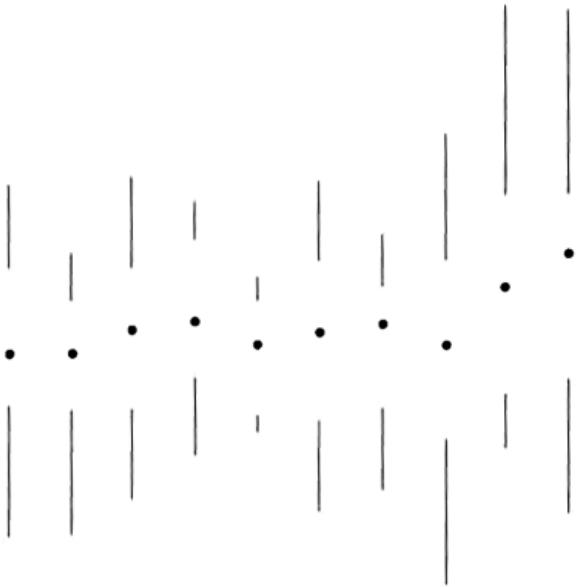
④ Summary

Redesign of Box-plots

Old

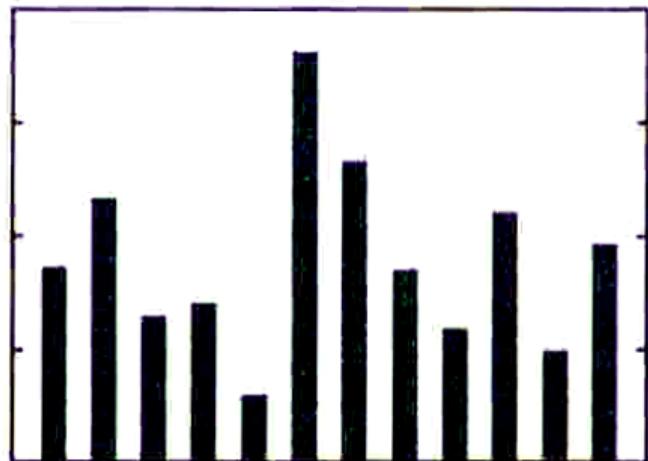


New

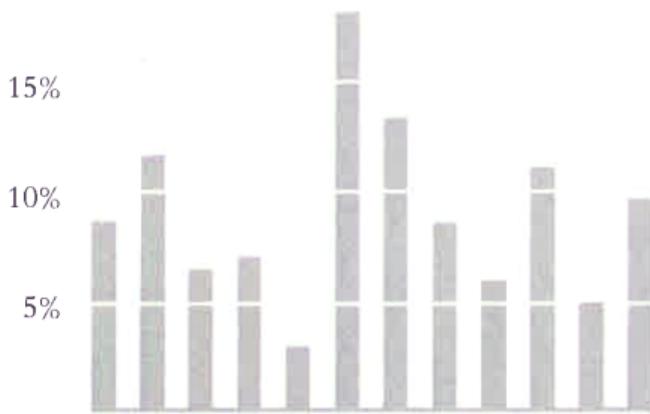


Redesign of Bar-charts

Old

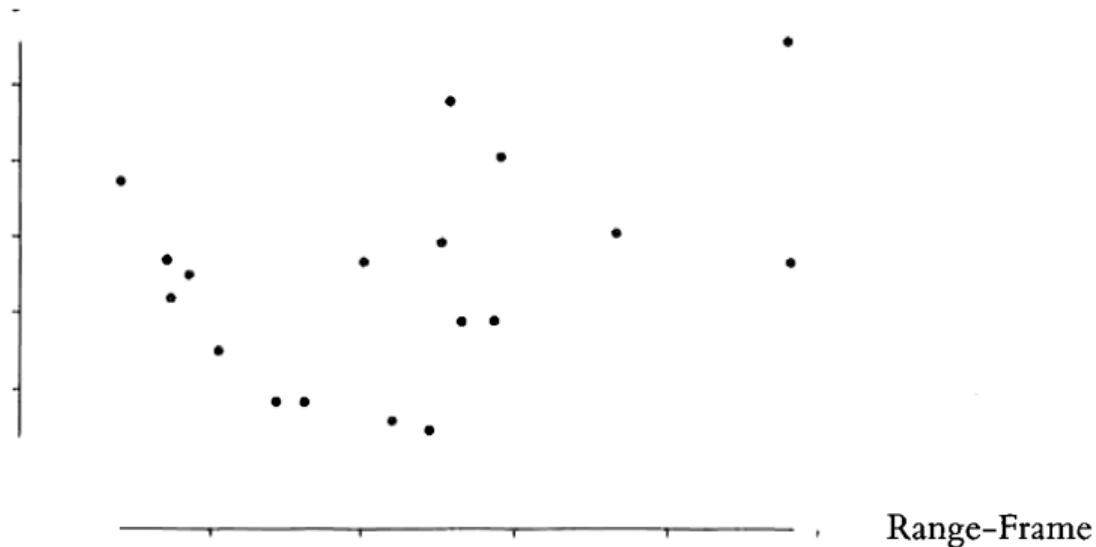


New



Redesign of Scatter Plots

range frame



The frame above depicts the marginal range of each variable.

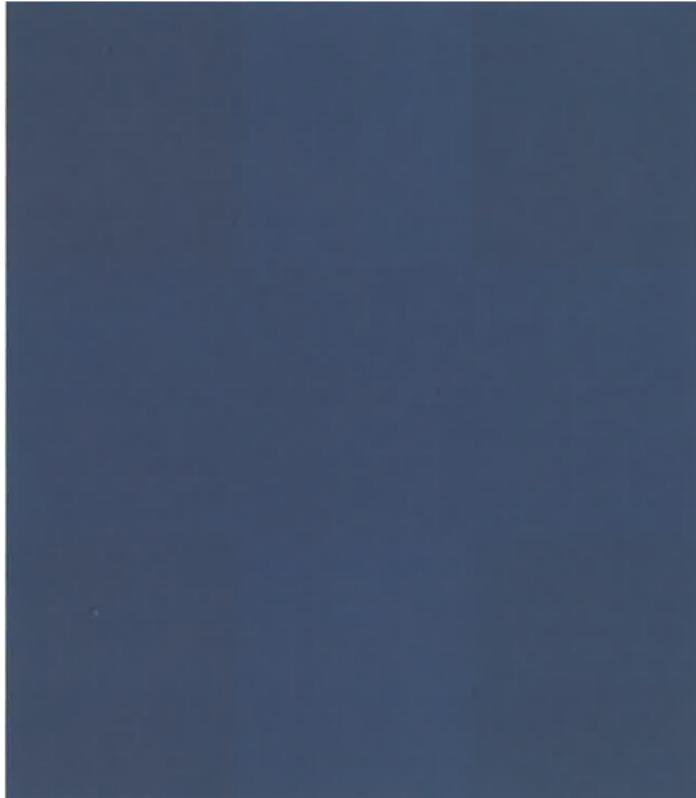
Summary

- ① The new designs maximise the data-ink ratio, so are justified in some sense.
- ② Graphics always improve as we revisit and re-edit them, so do not stop at the first plot you make.
- ③ Do not underestimate the audience. The graph should be as rich as can be. If it looks cluttered, re-think the design - don't erase data ink instead.

A graph should be as sophisticated as the words and argument you are making.

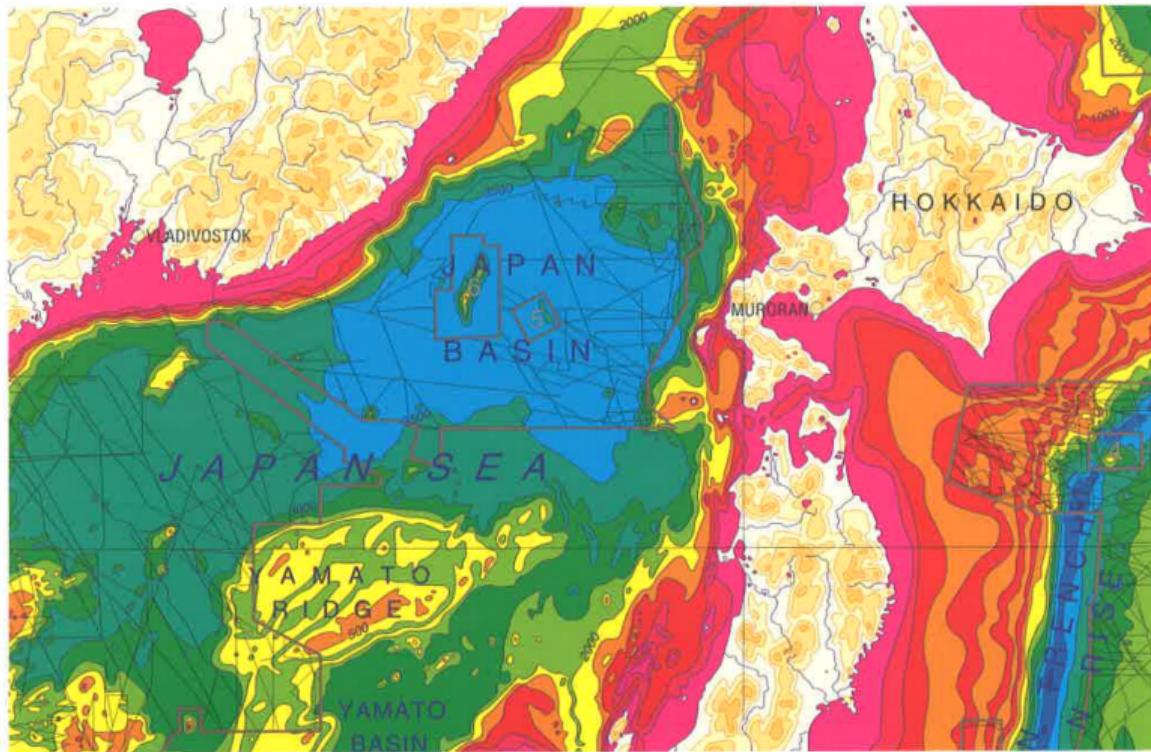
Smallest Effective Difference

Our eyes are excellent at distinguishing even small differences in colour.



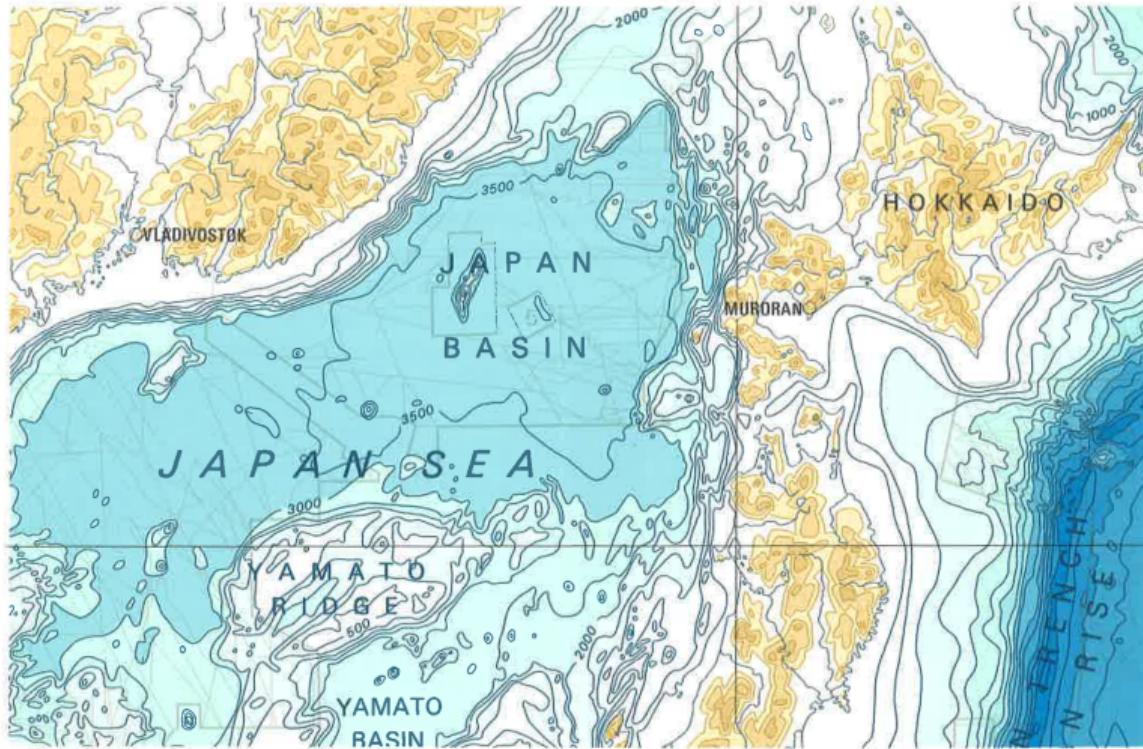
Smallest Effective Difference

no need

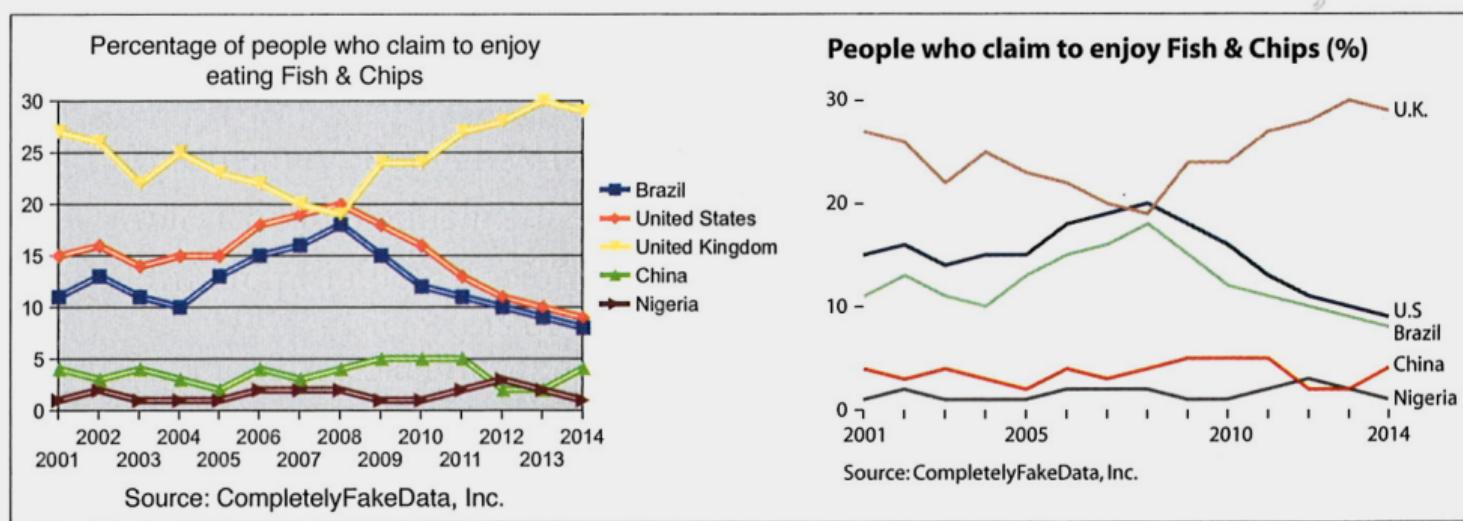


Smallest Effective Difference

this is enough



Remove Chart Junk



The chart on the right does not contain redundant elements. It is certainly the choice that is more aesthetically pleasing.

① Introduction

② Graphical Practice

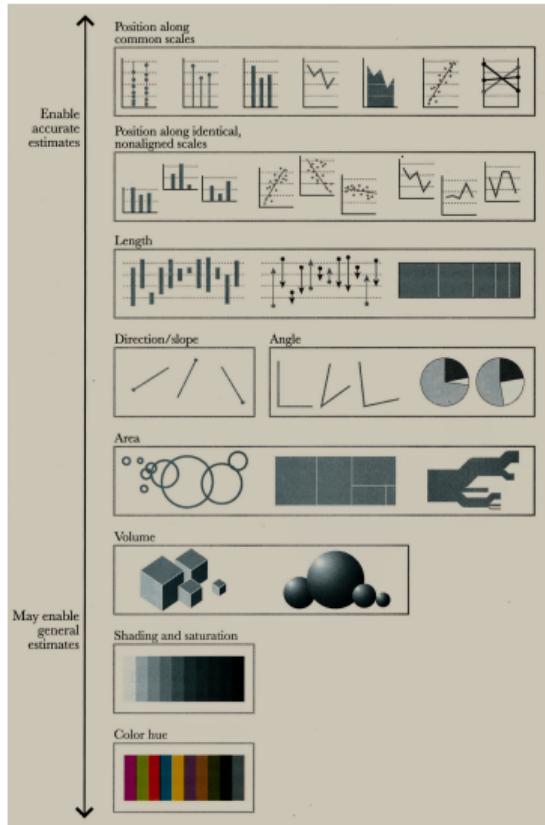
- Good Graphs
- Bad Graphs

③ A Theory of Data Graphics

- Data-Ink
- Chart Junk
- Small Multiples
- Aesthetics and Technique
- Applications of Data-Ink Principles
- Thinking About Your Graph

④ Summary

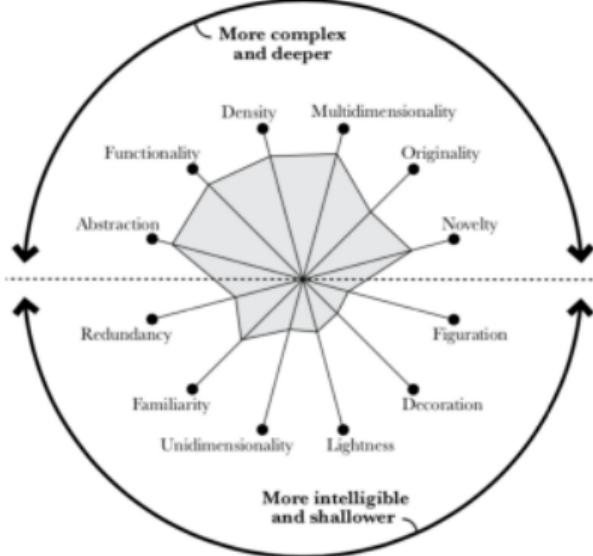
Which Type of Graph to Make?



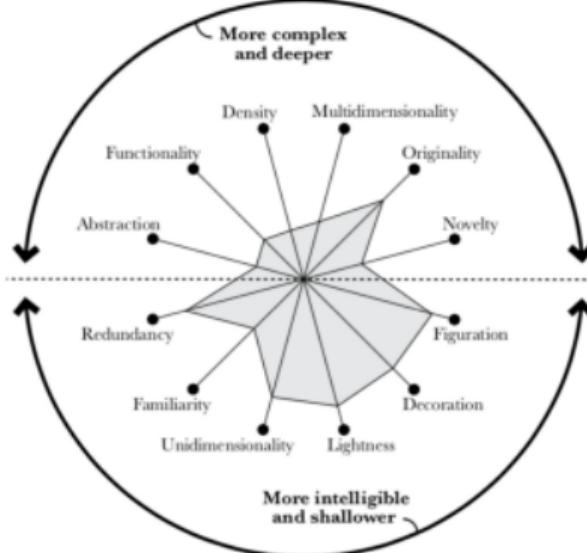
- Consider the purpose and audience for your graph.
- They may help you determine which type of graph to make, using the scale on the left.

For Whom is the Graph?

The wheel preferred by scientists and engineers



The wheel favored by artists, graphic designers, and journalists



Be Aware of Your Own Biases!

- There are three types of mistakes our minds cause us to make that we should be aware of when performing analyses:
 - ① The *patternicity* bias.
 - ② The *storytelling* desire.
 - ③ The *confirmation* bias.
- How to avoid these?

The Scientific Process

- Treat each “finding” as a conjecture.
- Do not believe it until you can find multiple ways of confirming it. Repeatedly.
- Ask your colleagues/team-mates to review your work.
- Always follow-up with more questions, investigations and plots.

The larger the island of knowledge, the longer the shoreline of wonder.

① Introduction

② Graphical Practice

- Good Graphs
- Bad Graphs

③ A Theory of Data Graphics

- Data-Ink
- Chart Junk
- Small Multiples
- Aesthetics and Technique
- Applications of Data-Ink Principles
- Thinking About Your Graph

④ Summary

Summary

- Graphical elegance is often found in simplicity of design and complexity of data.
- Data graphics are paragraphs about data and should be treated as such.