Design World

Graphical Excellence

largely from Edward Tufte,
Tufte’s Graphical Theory

1st set of guidelines
• Show the data
• Induce viewer to think about the substance
• Avoid distorting the data
• Present many numbers in a small space
• Make large data sets coherent
• Encourage visual comparisons
• Reveal data at several levels of details
• Serve a clear purpose (description, explanation, tabulation, decoration)
• Integrate closely with text descriptions
Graphical Excellence

Fundamental graphic designs include:

- Data maps: these involve placement of additional information spatially situated on a spatially explicit diagram.
- Time series: these involve plotting some data as it changes across time.
- Space-time narrative: plotting changes across both space and time.
- Relational graphics: designed to show the relationship between two or more data aspects.
Graphical Excellence

Start with reasonable data

A. New York stock prices
B. Solar radiation inverted,
C. London stock prices
For all months 1929
Age-adjusted death rates by cancer type for USA (each some 21,000 numbers)

Can be considered at many levels from overall pattern to county by county detail

- High death rates in north east and around great lakes
- Low rates in band down middle
- Higher rates for men than women in south
- Hot spots; in Minnesota, Iowa, Nebraska, along the Missouri River
- Differences in cancer types by regions

Data map: 1864 Exports of French Wine

E. Tufte "Visual Display of Quantitative Information" p 25,
Time Series

E. J. Marey. 1885. Train schedules from Paris to Lyon
Stations spaced according to distances, time from left to right

E.J. Marey, “La Methode Graphique,” (Paris 1885), p.20. This method is attributed to the French engineer, Irby (Tufte, 1883, p.31)
Time Series

E. J. Marey. 1885. Train schedules from Paris to Lyon
Stations spaced according to distances, time from left to right

E.J. Marey, “La Methode Graphique,” (Paris 1885), p.20. This method is attributed to the french engineer, Irby (Tufte, 1883, p.31)

1981 – new express train – trip now 3 hours instead of 9
Time Series

W. Playfair. 1759-1823. 3 time series – prices, wages and reigns of Kings and Queens

Time Series

Diagrams of motion

Using white tape and black velvet, Marey created time series images.

*E. J. Marey, (1830 – 1904)*

Space-time story

Small multiples

Learn once

Invite comparisons

Los Angeles Times, July 22, 1979; based on work of G. McRae, California Institute of technology. (Tufte, 1983, p.42)
Relational Graphics

Relationship between temperature and thermal conductivity of copper

Gathers data from several laboratories

Makes a clearer and stronger point by the collection

Connected points are from one publication

Different answers result from different impurities levels
Graphical Excellence - Summary

Designed for the presentation of interesting data – matter of substance, of statistics, and of 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.

• > ideas
• < time
• < ink
• < space

Graphical excellence is nearly always multivariate.

And is all about truth and integrity.
Graphical integrity

• Graphics can be a powerful communication tool
• Lies and falsehoods are possible
• Much focus on this ‘how to lie with maps’ or ‘statistics’
Examples of misleading graphics

Where is the bottom line? What is happening in 1970?

E. Tufte: Visual Display of Quantitative Information
Misleading graphics


What is the first impression of the airlines relative success in 1978?

Order of numbers?

Magnitude of numbers? Impression?

Achieving graphical Integrity

A graphic does not distort if the visual representation is consistent with the numerical representation.

• Is the magnitude of ‘visual representations’ as physically measured on the graphic?

• Or the perceived magnitude?

Approach

Conduct a study of visual perception of the graphics.

Circles – perceived area grows more slowly than measured area

reported perceived area = (actual area)^x, where x = 0.8+-0.3

Lines -
Lie Factors

• ‘the representation of numbers, as physically measured on the surface of the graphic itself, should be directly proportional to the numerical quantities represented.’

• ‘Clear, detailed and thorough labeling should be used to defeat graphical distortion and ambiguity. Write out explanations of the data on the graphic itself. Label important events in the data.’

\[
\text{Lie Factor} = \frac{\text{size of effect shown in graphic}}{\text{size of effect in data}}
\]

– Lie factor of 1 – is desirable

– lie factors > 1.05 or < 0.95 go beyond plotting errors
Fuel economy standards for automobiles
18 miles/gallon in 1978 to 27.5 miles/gallon in 1985
Increase of 53% = \((27.5 - 18.0)/(18.0) \times 100\)
Extreme example

Graphic increase

783% = \((5.3 - 0.6)/(0.6)\) x 100

Lie Factor = 783/53 = 14.8

Additional confounding factors

- Usually the future is in front of us
- Dates remain same size and fuel factors increase
- Includes perspective distortion – how to read change in perspective
Extrapolation

a graphic generates visual expectations – deception can result from incorrect extrapolation of visual expectations

1st seven intervals are 10 years
The last interval is 4 years
Gives a false sense of decline

Nobel Prizes Awarded in Science, for Selected Countries, 1901-1974

(Number of Prizes)

United States

United Kingdom

Germany

France

U.S.S.R.

Accurate data for the next 10 years

Design Variation vs Data Variation

Design Variation vs Data Variation

5 different vertical scales show price

<table>
<thead>
<tr>
<th>Year</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>1973-1978</td>
<td>$8.00</td>
</tr>
<tr>
<td>Jan. – Mar. 1979</td>
<td>$4.73</td>
</tr>
<tr>
<td>Apr. – June 1979</td>
<td>$4.37</td>
</tr>
<tr>
<td>Jul. – Sept. 1979</td>
<td>$4.16</td>
</tr>
<tr>
<td>Oct. – Dec. 1979</td>
<td>$3.92</td>
</tr>
</tbody>
</table>

2 different horizontal scales show passage of time

<table>
<thead>
<tr>
<th>Year</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1973-1978</td>
<td>3.8 years</td>
</tr>
<tr>
<td>1979</td>
<td>0.57 years</td>
</tr>
</tbody>
</table>

With both scales shifting the distortion is multiplicative
Context is Essential

Graphics must not quote data out of context

Data sparse graphics should provoke suspicion

Graphics often lie by omission

Nearly all important questions are left unanswered by this graphic

Connecticut Traffic Deaths, Before (1955) and After (1956) Stricter Enforcement by the Police Against Cars Exceeding Speed Limit
Context is Essential

Graphics must not quote data out of context

A few more data points tell a more complete story
Context is Essential

Graphics must not quote data out of context

Different data points would tell a different stories
Context is Essential

Graphics must not quote data out of context

Comparisons with adjacent states give more context
Graphical Integrity - Summary

- ‘The representation of numbers, as physically measured on the surface of the graphic itself, should be directly proportional to the numerical quantities represented.

- Clear, detailed, and thorough labeling should be used to defeat graphical distortion and ambiguity. Write out explanations of the graphic itself. Label important events.

- Show data variation, not design variation.

- In time-series displays of money, deflated and standardized units of monetary measurement are nearly always better than nominal units.

- The number of information-carrying (variable) dimensions depicted should not exceed the number of dimensions in the data.

- Graphics must not quote data out of context.’

(Tufte, 1983, p77)
Tufte’s Graphical Theory

Tufte presents
  • 1\textsuperscript{st} graphical excellence
  • 2\textsuperscript{nd} graphical disasters

Then discusses the causes and provides guidelines
1. Causes of poor graphics
2. Guidelines
  1. Minimize ink while maximizing data ink
  2. Data density
  3. Avoid chartjunk
  4. Create multifunctioning graphics
  5. Make use of parallelism (small multiples)
  6. Consider 3D carefully
  7. Consider colour carefully
Causes of Poor Graphics

Question: why are there so few examples of good graphics?

- One possible answer is lack of training
- Another is societal attitudes

Common Attitudes

- Data is boring – in spite of the fact that more and more time, money, and people hours are spent gathering data (fire hoses of data)
- Graphics are for people who can not understand the data

Consequences

- Simplification
- Decoration
Causes of Poor Graphics

Using occurrence of relational graphics as evidence

In newspapers and magazines

Some in Japan

Little in Europe

Almost none in USA

(Tufte, 1983, p. 83)
Guidelines: Minimize ink while maximizing data ink

Data-ink ratio = data-ink / total ink used to print the graphic

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

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

Above all else show the data

Maximize the data-ink ratio

Erase non-data ink

Revise and edit
Guidelines: Minimize ink while maximizing data ink

William Playfair’s first graphics used considerably more ink in general – August 1785
Guidelines: Minimize ink while maximizing data ink

William Playfair’s first graphics used considerably more ink in general.

Less grid lines
Space data scale
Cleaner entry lines
Succinct comments
Guidelines: Minimize ink while maximizing data ink

Most ink is being used for the data

Guidelines: Minimize ink while maximizing data ink

Actual registration compared to predicted registration

Over simplified – no data present

Data easy to see


Slides by: Sheelagh Carpendale
Guidelines: Minimize ink while maximizing data ink

heavy grid lines – non data ink – actually cause irritation – chartjunk
Guidelines: data density

Data density = number of entries in the data matrix / area of the data graphic

Data can be designed to have several viewing depths

The following map has at least 3

1. What is seen from a distance, an overall structure usually aggregated from an underlying microstructure
   • The overall pattern the concentrations

2. What is seen close up and in detail; the fine structure of the data
   • The size etc. of the cities
   • The spread in the sparse areas

3. What is seen implicitly – the interrelations between the data
   • The effect of traffic and landscape corridors

Each dot represents 500 people in rural areas – 400,000 data points
Guidelines: data density

Space data scale
Cleaner entry lines
Succinct comments
Guidelines: data density

Data points represent sun shining

Annual sunshine record reports about 1,000 numbers per square inch
Guidelines: data density

New York Weather History - 1980
- Data density - 181 numbers/sq inch

Guidelines: Avoid chartjunk

Adding frills does not help
Data graphics get sold on their content
Avoid decoration
Moiré vibration
Over powering grids
Extraneous additions
Guidelines: Avoid chartjunk

Cotton production in Brazil, 1927 – vibrating textures
Chart Junk: A common error

Information visualization is not just pretty graphics
• graphical re-design by amateurs on computers gives us
  - “fontitis,” “chart-junk,” etc.

Dear Sir,
This is a **really exciting** opportunity! Take advantage of it!
Guidelines: Create multifunctioning graphics

Multifunctioning graphical elements can effectively display complex data

The principle is: mobilize every graphical element, perhaps several times over, to show the data.

Avoid making puzzle graphics with encodings that can only be understood by their creator
Guidelines: Create multifunctioning graphics

Reading chart vertically – ranks 15 countries by government tax collection for 1970 and 1979

Names are spaced in proportion to percentages

Paired comparisons show how numbers changed over the years
Guidelines: Make use of parallelism

Often there exists some parallelism in the data

This can be used to clarify which aspects of the data change

Small multiples have been used for 100s of years
  • These are a series of graphics such as frames in a movie showing the same combination of variables, indexed by changes in another variable(s)

Well designed small multiples are
  • Inevitably comparative
  • Deftly multivariate
  • Shrunken, high density graphics
  • Usually based on a large data matrix
  • Drawn almost entirely with data-ink
  • Efficient in interpretation (learn once applies to all)
  • Often narrative in content – showing shifts in relationships
Guidelines: Make use of parallelism

Galileo marked and labeled sunspots

Collections of these labeled drawings make small multiples

Showing sunspot location, with time and labels and the shifting orientation of the sun in our sky
Guidelines: Make use of parallelism

Small Multiple from Huygens’ *Systema Saturnium* 1659.

Inner ellipse is earth’s orbit, outer Saturn’s orbit, outer most floating images depict Saturn as viewed from earth, explaining previous idea about Saturn’s shape.

E. Tufte: *Visual Display of Quantitative Information*
Guidelines: Make use of parallelism

J. H. Colton

Johnson’s New Illustrated Family Atlas with Physical Geography, 1864

E. Tufte: Visual Display of Quantitative Information
Small Multiples: Showing Time and Change
Small Multiples: Showing Time and Change
Guidelines: Consider 3D carefully

Why 3D?

- The world is 3D – 2D graphics flatten data
Guidelines: Consider 3D carefully

Why 3D?

- Multivariate data is difficult to draw in 2D
- Consider these early suggestions for periodic table in chemistry
Guidelines: Consider 3D carefully

Why 3D?

• Representing 3D scatterplots in 2D
Guidelines: Consider 3D carefully

3D graphic excellence
Guidelines: Consider 3D carefully

3D graphic excellence
Guidelines: Consider 3D carefully

3D graphic excellence

Actual 3D, Euclid, The Elements of Geometrie

So all their angles there joined together, make a solide angle. And for the better sight thereof, I have here a figure whereby ye shall more easily conceive it, the base of the figure is a triangle, namely, A B C, if on every side of the triangle A B C, ye raise vp a triangle, as upon the side A B, ye raise vp the triangle A F B, and upon the side A C the triangle A F C, and upon the side B C, the triangle B F C, and so bowing the triangles raised vp, that their toppes, namely, the pointes F more and joyn together in one point, ye shall easily and plainly see how these three superficial angles A F B, B F C, C F A, joyn and close together, touching the one the other in the point F, and so make a solide angle.

Direct methods for the display of three dimensions include making models, as in the construction of Euclid’s Elements, where little paper constructions tell all geometry. Models pleasingly represent the smooth surfaces of three-space, as in architectural miniatures and mathematical solids; more obstreperous statistical data however, call for computer analysis of data point clouds.

E. Tufte, Visual Display of Quantitative Information

Guidelines: Consider 3D carefully

3D graphic excellence
Guidelines:
Consider 3D carefully

3D graphic excellence

M. E. Turgot, L. Bretez, Plan de Paris (Paris 1739)
Guidelines: Consider 3D carefully

3D graphic excellence

Some 7000 pieces of space debris – points not to scale but each at least as big as 8x10 inches – 1987

Micro and macro readings
Guidelines: Consider 3D carefully

3D graphic excellence

3D explosion diagram IBM Series III copier
Guidelines: Consider colour carefully

To avoid over powering and garish colours.
Guidelines: Consider colour carefully

To avoid over powering and garish colours
Guidelines: Consider colour carefully

Use least possible emphasis colours on clam background
Guidelines:
Consider colour carefully

To avoid over powering and garish colours

Use least possible distinguishable difference
Guidelines: narratives of space and time
Guidelines: narratives of space and time
Guidelines: narratives of space and time

Use least possible distinguishable difference
Guidelines: narratives of space and time

Use least possible distinguishable difference
Graphical Integrity - Summary

• ‘The representation of numbers, as physically measured on the surface of the graphic itself, should be directly proportional to the numerical quantities represented.

• Clear, detailed, and thorough labeling should be used to defeat graphical distortion and ambiguity. Write out explanations of the graphic itself. Label important events.

• Show data variation, not design variation.

• In time-series displays of money, deflated and standardized units of monetary measurement are nearly always better than nominal units.

• The number of information-carrying (variable) dimensions depicted should not exceed the number of dimensions in the data.

• Graphics must not quote data out of context.’

(Tufte, 1983, p77)