Chart Deception

Main Source: How to Lie with Charts, by Gerald E. Jones

(Click icon for audio)
Information

Information = Data + Labels

5056465238 = data
(505) 646-5238 = information
Summaries Can Hide Needed Detail
Figure 1.2  Two views of sales results: The use of segmented bars reveals a hidden story.
Pie Charts

• Pie charts denote percentages
• Label slices with percentages, not raw values
  – True of two pie figures
• Most important slice
  – Right-hand slice
  – Exploded slice
• Bottom slice of 3-D pie seems overly important
Figure 2.2 This pie chart shows values instead of percentages for the pie slices. It gives the possibly incorrect impression that most people on the staff work 7.5-hour days.

Staff Productivity Study
Average Work Hours per Task

- Travel: 0.9
- Meetings: 1.6
- Mail: 1.8
- Phone: 2.5
- Other: 0.7
Figure 2.4 Including dollar amounts in pies that show sales results is usually a bad idea because it distracts from the real story—the proportions of the slices to the whole.

Total Sales by Company
Year to Date (Millions)

- Aardvark: $162
- Aero: $83
- Acme: $124
- Able: $95
- Archer: $143
- Argo: $78
- Ace: $231
FIGURE 1. A pie chart showing the relative number of international flights from Newark Airport over a five-year period. The graphics software didn’t squawk at all when asked to make a pie chart with such inappropriate content.
Pie Charts

• Pie charts denote percentages
• Label slices with percentages, not raw values
  – True of two pie figures
• Most important slice
  – Right-hand slice
  – Exploded slice
• Bottom slice of 3-D pie seems overly important
Figure 2.6  The Two-Pie Story is told by information gluttons who can't resist the temptation to have that second pie!
Pie Charts

• Pie charts denote percentages
• Label slices with percentages, not raw values
  – True of two pie figures
• Most important slice
  – Right-hand slice
  – Exploded slice
• Bottom slice of 3-D pie seems overly important
Figure 2.9 In a pie-column chart, the column itemizes the exploded slice.

Market Share by Company
Showing Aardvark Product Lines

- Sporting Goods
- Games
- Toys

- Aardvark
- Aero
- Acme
- Able
- Archer
- Argo
- Ace
Pie Charts

- Pie charts denote percentages
- Label slices with percentages, not raw values
  - True of two pie figures
- Most important slice
  - Right-hand slice
  - Exploded slice
- **Bottom slice of 3-D pie seems overly important**
Figure 2.3  This 3D pie really distorts the importance of the bottom slice. The thick edge makes the Meetings slice look much more substantial than the Phone slice, even though its percentage is smaller.
Pie Charts

• Multi-sized pies on same figure (to denote trend) biased by dissimilar diameter-area rates of growth

• Large ‘Other’ slice can hide needed detail

• Use pie-column chart for more detail on a slice

• Donut charts confusing

• Omitting missing data increases relate size of all slices
Figure 2.7 Using different-size pies just makes a two-pie chart that much more confusing.
Pie Charts

• Multi-sized pies on same figure (to denote trend) biased by dissimilar diameter-area rates of growth
• Large ‘Other’ slice can hide needed detail
• Use pie-column chart for more detail on a slice
• Donut charts confusing
• Omitting missing data increases relate size of all slices
Figure 2.9 In a pie-column chart, the column itemizes the exploded slice.

Market Share by Company
Showing Aardvark Product Lines

- Sporting Goods
- Games
- Toys

- Aardvark  
- Aero  
- Acme  
- Able  
- Archer  
- Argo  
- Ace
Pie Charts

- Multi-sized pies on same figure (to denote trend) biased by dissimilar diameter-area rates of growth
- Large ‘Other’ slice can hide needed detail
- Use pie-column chart for more detail on a slice
- Donut charts confusing
- Omitting missing data increases relate size of all slices
Figure 2.10 A donut chart is a set of concentric, hollowed-out pies.

Market Share by Company and Product Line

- Aardvark
- Argo
- Acme

Legend:
- Toys
- Games
- Sporting Goods
Pie Charts

• Multi-sized pies on same figure (to denote trend) biased by dissimilar diameter-area rates of growth
• Large ‘Other’ slice can hide needed detail
• Use pie-column chart for more detail on a slice
• Donut charts confusing
• Omitting missing data increases relate size of all slices
A pie-chart liar's trick is to omit miscellaneous slices. If the miscellaneous items make up 10 percent of the actual data sample, the net effect will be to expand the remaining 90 percent to become 100 percent of the pie. In other words, the percentage of each of the slices will be exaggerated by making each of them 10 percent bigger. (See Figure 7.2.)

Figure 7.2  The pie on the left includes all the data in an actual sample. The Not Recorded slice has been omitted from the pie on the right, thereby distorting the percentages of all the other slices.
Attend to Cultural Bias in Orienting Charts
Retrograde Motion

Progress

More

Less
Argo Sales Increase

1995
1996
1997
1998
1999

Responsive customer service
Argo Sales Increase 1995-1999

- Sales, in units or dollars, listed along the y axis
- Avoid 3D effects
Figure 3.5 Using a horizontal orientation to show sales volumes fails to capitalize on the intuitive notion that up means more in a vertical chart.
Bar Chart—Avoid 3D and Sideways

How did you find your last job?

- Networking: 55.2%
- Print ad: 18.3%
- Online recruitment site: 15.4%
- Placement firm: 9.6%
- Temporary agency: 1.5%
Figure 4.17 Even if the x axis does not represent the flow of time, viewers will regard the bars on the right as the most recent or important. That’s not necessarily bad, just something you should be aware of.
Figure 3.13 To meet viewers' expectations about direction and geography, the maker of this chart should have put the West on the left, Central in the middle, and East on the right.
Figure 4.12 Turning a conventional vertical bar chart on its side confuses the intuitive meanings of up-down and left-right.

**National Sales Volume by Region**  
(Millions of Units)

- West
- North
- South
- East

**National Sales Volume by Region**  
(Millions of Units)

- West
- North
- South
- East

LIAR'S TRICK
Figure 3-16. The broken scale is contrary to the one-dimensional character of the bar and column chart. It is never an acceptable practice. Also, from a mechanical point of view, apparently little thought was given to the size of lettering in relation to the amount of reduction. [From Selma Taffel, Congenital Anomalies and Birth Injuries Among Live Births: United States, 1973-1974, (DHEW Publication; No. (PHS) 79-1909, National Center for Health Statistics, Washington, D.C.: Government Printing Office, 1978, p. 7.]
Bar and Area Charts

• Stacked vs. clustered bar charts
  – Mixing cumulative and clustered bars can confuse viewers

• When stacking, place least variable component on bottom

• Avoid paired-bar charts

• Viewers can misread 3-D bars
Stacked Bar Chart

- Group 1 — "Fun in the Sun"
- Group 2 — "Serious Tanning"
- Group 3 — "At the Lake"
Clustered Bar Chart

- Group 1 – “Fun in the Sun”
- Group 2 – “Serious Tanning”
- Group 3 – “At the Lake”
Figure 7.16  The 100-percent bar style can be used instead of multiple pies to compare ratios.
Bar and Area Charts

• Stacked vs. clustered bar charts
  – Mixing cumulative and clustered bars can confuse viewers

• When stacking, place least variable component on bottom

• Avoid paired-bar charts

• Viewers can misread 3-D bars
Figure 4.5 Area charts are usually shown with stacked data series. That is, the starting point of one area is the top of the area below it. Putting the least "bumpy" data series at the bottom minimizes distortion of the other series.
Bar and Area Charts

• Stacked vs. clustered bar charts
  – Mixing cumulative and clustered bars can confuse viewers
• When stacking, place least variable component on bottom
• Avoid paired-bar charts
• Viewers can misread 3-D bars
Figure 7.17 A paired-bar chart is doubly mistaken—by using horizontal bars to show magnitude and by using leftward-going bars to show positive values.
Bar and Area Charts

- Stacked vs. clustered bar charts
  - Mixing cumulative and clustered bars can confuse viewers
- When stacking, place least variable component on bottom
- Avoid paired-bar charts
- Viewers can misread 3-D bars
Figure 4.16 In this set of 3D bars, the vanishing point is above the bars, making it difficult for the audience to estimate the true heights.
Bar Graph—Avoid 3D
Avoid 3-D Effects

Lawyers\textsuperscript{a} (per 100,000 population)

- Chile
- India
- Malaysia
- Japan
- United Kingdom
- Germany
- United States

\textsuperscript{a} Latest years available.

\textsuperscript{b} Includes England and Wales only.
Figure 8-12. Another example of a three-dimensional chart which does not exhibit any special qualities that make it superior to the conventional composite bar chart as a visual medium of communication. [From U.S. Department of Labor, Bureau of Labor Statistics, Chartbook on Occupational Injuries and Illnesses, 1974, Report 460, 1976, p. 6.]
Line Graph—Avoid 3D
Computer Graphic Illustrating Wiretap Surveillance Interception by State
XY (or Scatter) Charts

- **Label data points**
- Start axes at coordinates (0,0)
- Avoid distorting axes
- Avoid broken axes
- Cumulative charts seem optimistic because upward sloping
- Avoid dissimilar data on same chart
- Semi-log charts can minimize trends
- If x axis isn’t time axis, connecting dots can mistakenly suggest a trend
- Interpolation and extrapolation are subjective
Figure 3.2 This crude chart, which has no numeric scales or labels to help you interpret it, has a powerful implied message.
Labeling Data Points

A basic tool of chart designers is to label the values of data points on a line plot or set of bars:
XY (or Scatter) Charts

- Label data points
- **Start axes at coordinates (0,0)**
- Avoid distorting axes
- Avoid broken axes
- Cumulative charts seem optimistic because upward sloping
- Avoid dissimilar data on same chart
- Semi-log charts can minimize trends
- If x axis isn’t time axis, connecting dots can mistakenly suggest a trend
- Interpolation and extrapolation are subjective
Figure 6.1  The range of $y$-axis values is 3.5–8, and the bounds of this $y$-axis scale are 3.0 and 9.0.
Figure 6.2  Expanding the y-axis range here minimizes fluctuations in the data plot.
Figure 6.4  Making the axis range narrower exaggerates fluctuations in the plot, which suits my deceitful purposes. I'm also pleased that the high point of the plot extends beyond the maximum value on the scale, which unnecessarily exaggerates the magnitude of that data point.
Figure 6.3  Here's another example where expanding the axis range flattened the plot. Unfortunately, that wasn't the effect I wanted.
XY (or Scatter) Charts

- Label data points
- Start axes at coordinates (0,0)
- **Avoid distorting axes**
- Avoid broken axes
- Cumulative charts seem optimistic because upward sloping
- Avoid dissimilar data on same chart
- Semi-log charts can minimize trends
- If x axis isn’t time axis, connecting dots can mistakenly suggest a trend
- Interpolation and extrapolation are subjective
Distortion from Treating Unequal Time Intervals as Equal

(a) $ millions

(b) $ millions
Distortion by Alternating Scales

Changing the Visual Image
Contracting or expanding vertical (amount) scale or horizontal (time) scale tends to change the visual picture.

- Original Scale Arrangement
- Expanding Vertical
- Contracting Horizontal
- Expanding Horizontal
- Expanding Vertical and Contracting Horizontal
- Contracting Vertical
- Contracting Vertical and Expanding Horizontal
XY (or Scatter) Charts

- Label data points
- Start axes at coordinates (0,0)
- Avoid distorting axes
- **Avoid broken axes**
- Cumulative charts seem optimistic because upward sloping
- Avoid dissimilar data on same chart
- Semi-log charts can minimize trends
- If x axis isn’t time axis, connecting dots can mistakenly suggest a trend
- Interpolation and extrapolation are subjective
Distortion of Chart from Broken Vertical Scales

(a) Distortion: Broken Vertical Scales

Sales ($ millions)

(b) Correct: Full Vertical Scales

Sales ($ millions)

Number of customers

Number of customers

100

50

0
XY (or Scatter) Charts

- Label data points
- Start axes at coordinates (0,0)
- Avoid distorting axes
- Avoid broken axes
- **Cumulative charts seem optimistic because upward sloping**
- Avoid dissimilar data on same chart
- Semi-log charts can minimize trends
- If x axis isn’t time axis, connecting dots can mistakenly suggest a trend
- Interpolation and extrapolation are subjective
Figure 4.8 In a cumulative chart, each point within a data series is a cumulative total of all preceding points in the same series. Cumulative charts look optimistic by nature.
Figure 7.14 Cumulative series, clustered bars. (Note the greatly expanded y-axis, which now spans 0–100.)
XY (or Scatter) Charts

- Label data points
- Start axes at coordinates (0,0)
- Avoid distorting axes
- Avoid broken axes
- Cumulative charts seem optimistic because upward sloping
- Avoid dissimilar data on same chart
- Semi-log charts can minimize trends
- If x axis isn’t time axis, connecting dots can mistakenly suggest a trend
- Interpolation and extrapolation are subjective
Figure 6.7 It’s not necessarily incorrect to show dissimilar data sets on the same dual-\(y\) chart. But is the comparison valid?
Figure 7.4 Here’s a really slick lie. I plotted something that really is a trend against a secondary axis, implying that one set of data somehow influences the other.
XY (or Scatter) Charts

- Label data points
- Start axes at coordinates (0,0)
- Avoid distorting axes
- Avoid broken axes
- Cumulative charts seem optimistic because upward sloping
- Avoid dissimilar data on same chart
- **Semi-log charts can minimize trends**
- If x axis isn’t time axis, connecting dots can mistakenly suggest a trend
- Interpolation and extrapolation are subjective
Figure 6.9 An exponential trend can be made to look less dramatic by plotting it on a semi-log chart, where it resembles a linear trend.
XY (or Scatter) Charts

• Label data points
• Start axes at coordinates (0,0)
• Avoid distorting axes
• Avoid broken axes
• Cumulative charts seem optimistic because upward sloping
• Avoid dissimilar data on same chart
• Semi-log charts can minimize trends
• If x axis isn’t time axis, connecting dots can mistakenly suggest a trend
• Interpolation and extrapolation are subjective
Figure 7.3  Connecting the data "dots" creates a line graph, which might wrongly imply some kind of trend.
XY (or Scatter) Charts

- Label data points
- Start axes at coordinates (0,0)
- Avoid distorting axes
- Avoid broken axes
- Cumulative charts seem optimistic because upward sloping
- Avoid dissimilar data on same chart
- Semi-log charts can minimize trends
- If x axis isn’t time axis, connecting dots can mistakenly suggest a trend
- **Interpolation and extrapolation are subjective**
Radar Charts

Beware non-uniform scales on axes
Figure 5.5  Here’s the same data I showed you in Figure 5.1, but with the scales adjusted to improve the shape of the plot.
Depicting Confidence Intervals
Figure 10-3. Additional designs for portraying confidence limits for simple column and bar charts. Certain features based on suggestions of Albert Biderman and staff, Bureau of Social Science Research, Washington, D.C.

SUGGESTED DESIGNS FOR PORTRAYING CONFIDENCE LIMITS
Figure 10-4. The black bars on this chart have been designed to indicate 99-percent confidence intervals. This range is almost certain to include the estimate averaged over all possible repetitions of the sample. [Redrawn from Maria E. Gonzalez et al., "Standards for Discussion and Presentation of Errors in Data," Journal of the American Statistical Association, 70 (351), Part II (1975), 1-23.]

A. PERCENT UNEMPLOYED BY RACE AND SEX

- WHITE
  - TOTAL
  - MALES 20 AND OVER
  - FEMALES 20 AND OVER

- NEGRO AND OTHER RACES
  - TOTAL
  - MALES 20 AND OVER
  - FEMALES 20 AND OVER

B. DIFFERENCES IN PERCENT UNEMPLOYED BY RACE AND SEX

- NEGRO AND OTHER MINUS WHITE
  - TOTAL
  - MALES 20 AND OVER
  - FEMALES 20 AND OVER
  - FEMALES 20 AND OVER MINUS MALES 20 AND OVER

- WHITE

- NEGRO AND OTHER

PERCENT UNEMPLOYED

PERCENTAGE DIFFERENCES
Figure 10-6. A rectilinear coordinate line chart with four curves based in part on sample data. The 95-percent confidence levels for the respective series of annual rates from 1954 to 1967 are shown by vertical lines, with the upper and lower limits marked with short horizontal lines. (From Abbott L. Ferriss, Indicators of Change in the American Family, New York: Russell Sage Foundation, 1970, p. 15. Copyright 1970 by Russell Sage Foundation. By permission of Russell Sage Foundation.)
Headers and Footers

- Use footnotes to source table (avoid plagiarism charge)
- Font rules
  - Use proportional font
  - Don’t mix fonts in a chart
  - Use upper and lower case letters
- Spell out numbers 1 through 9 unless a measure
Figure 10.2  Use a footnote to disclose the source of information for a chart or the grantor of permission for reproduction.

Worker Productivity
Average Piece Rate per Day

Source: Bureau of Labor Statistics
Examples of Bad Graphics
Figure 3-20. A poorly designed chart in which the bars are arranged alphabetically. In a chart of this kind, visual comparisons are extremely difficult to make. In addition to the deficiencies in basic design, the draftsmanship and other mechanical features of this chart are of mediocre quality. [From Population Crisis Committee, “Natural Family Planning: Periodic Abstinence as a Method of Fertility Control,” Population (June 1981), 3.]
Figure 4-6. Examples of frequency polygons. One set of polygons is plotted in absolute numbers; the other set, in percentages. Note the location of the plotting points. (Chart based on data furnished by Carol B. Kalish and Margaret Heisler of the U.S. Bureau of Justice Statistics.)
COLLEGE-AGE POPULATION AND ENROLLMENT*
STATE OF WASHINGTON: 1910 TO 1980

EXPLANATORY NOTES:
ENROLLMENT INCLUDES ALL DEGREE-CREDIT STUDENTS AS DEFINED BY THE UNITED STATES OFFICE OF EDUCATION.
THE CURVES ON THIS CHART, PLOTTED ON DECENNIAL YEARS, EXCEPT 1964 AND 1975, ARE DESIGNED AS A SIMPLIFIED PORTRAYAL OF GENERAL TRENDS, WITHOUT CONSIDERATION TO ANNUAL FLUCTUATIONS.
PERCENTAGES IN PARENTHESES INDICATE PROPORTIONS OF ENROLLMENT IN COLLEGES AND UNIVERSITIES IN RELATION TO POPULATION 18 TO 24 YEARS OF AGE.

*ALL DEGREE-CREDIT STUDENTS
Median Income of Year-Round Full Time Workers
25-34 Years Old by Sex and Educational Attainment:
1968-1977

Male

Female

Constant 1977 dollars

Year
Examples of Bad and Better Graphics
Example #1

In the year 2000, which energy source is predicted to supply less power than coal?

A Petroleum  D Hydropower
B Natural Gas  E I don’t know
C Nuclear Power

FIGURE 1. A graphical test item from the National Assessment of Educational Progress. Its original source was the U.S. Department of the Interior, where it was titled "United States Energy Through the Year 2000." (© 1973 Congressional Quarterly, Inc.)
Profound increases are predicted in the use of Petroleum and Nuclear energy
Only modest increases in the use of other energy sources

**Figure 2.** One possible revision of the graph in figure 1 that clarifies many things.
SAT scores and funds for education rise together

Example #2

Figure 4. The data from figure 3 redisplayed when both vertical axes are scaled to expose the changes maximally.
SAT scores soar despite sluggish funding of education

**Figure 5.** The data from figure 3 redisplayed with both vertical axes scaled to suggest that SAT scores have increased despite sluggish spending on education.
FIGURE 38. Mixing a changed metaphor with a tiny label reverses the meaning of the data.
Travel Agents' Commissions Have Grown Since Fares Have Been Cut

*FIGURE 39.* Figure 38 redrawn with 1978 data placed on a comparable basis shows that the fare cuts have been a boon to travel agents.
OECD PETROLEUM STOCKS HAVE STABILIZED

But Not All Countries Are Pulling Their Own Weight

- United States: U.S. Adds 34 Million Barrels a Year
- All Other OECD: Petroleum Stock Declined 30 Million Barrels a Year
- Japan: Japan Adds 9 Million Barrels a Year
- West Germany: Reserves Fluctuated Around 250 Million Barrels
U.S. Beer Sales and Schlitz's Share

Barrels show total U.S. Sales

Example #5
U.S Beer Sales Grew Steadily Throughout the 1970s

**Figure 20.** Data from figure redone without tricks (from Wainer, 1980).
Example #6

**Purchasing Power of the Diminishing Dollar**

*Source: Labor Department*
In six Presidential administrations the dollar suffered its greatest decline during the Nixon administration.
Example #7

During the 1970s there was an increasingly Positive Balance of Trade with CHINA

Over the same time period our trade deficit with TAIWAN got much worse
<table>
<thead>
<tr>
<th>Country</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Argentina</td>
<td>56.90</td>
<td>61.40</td>
</tr>
<tr>
<td>Brazil</td>
<td>39.30</td>
<td>45.50</td>
</tr>
<tr>
<td>Canada</td>
<td>67.61</td>
<td>72.92</td>
</tr>
<tr>
<td>Iceland</td>
<td>66.10</td>
<td>70.30</td>
</tr>
<tr>
<td>Japan</td>
<td>65.37</td>
<td>70.26</td>
</tr>
<tr>
<td>Mexico</td>
<td>37.92</td>
<td>39.79</td>
</tr>
<tr>
<td>Netherlands</td>
<td>71.40</td>
<td>74.80</td>
</tr>
<tr>
<td>New Zealand</td>
<td>68.20</td>
<td>73.00</td>
</tr>
<tr>
<td>Norway</td>
<td>71.11</td>
<td>74.70</td>
</tr>
<tr>
<td>Spain</td>
<td>58.76</td>
<td>63.50</td>
</tr>
<tr>
<td>Country</td>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td>-----------------</td>
<td>------</td>
<td>--------</td>
</tr>
<tr>
<td>Netherlands</td>
<td>71</td>
<td>75</td>
</tr>
<tr>
<td>Norway</td>
<td>71</td>
<td>75</td>
</tr>
<tr>
<td>New Zealand</td>
<td>68</td>
<td>73</td>
</tr>
<tr>
<td>Canada</td>
<td>68</td>
<td>73</td>
</tr>
<tr>
<td>Iceland</td>
<td>66</td>
<td>70</td>
</tr>
<tr>
<td>Japan</td>
<td>65</td>
<td>70</td>
</tr>
<tr>
<td>Spain</td>
<td>59</td>
<td>64</td>
</tr>
<tr>
<td>Argentina</td>
<td>57</td>
<td>61</td>
</tr>
<tr>
<td>Brazil</td>
<td>39</td>
<td>46</td>
</tr>
<tr>
<td>Mexico</td>
<td>38</td>
<td>40</td>
</tr>
</tbody>
</table>
## Life Expectancy at Birth, By Sex
### Most Recent Available Year

<table>
<thead>
<tr>
<th>Women</th>
<th>Age</th>
<th>Men</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sweden</td>
<td>78</td>
<td></td>
</tr>
<tr>
<td>France, US, Japan, Canada</td>
<td>77</td>
<td>74</td>
</tr>
<tr>
<td>Finland, Austria, UK</td>
<td>76</td>
<td>73</td>
</tr>
<tr>
<td>USSR, Germany</td>
<td>75</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td>71</td>
<td>Sweden</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>Japan</td>
</tr>
<tr>
<td></td>
<td>69</td>
<td>Canada, UK, US, France</td>
</tr>
<tr>
<td></td>
<td>68</td>
<td>Germany, Austria</td>
</tr>
<tr>
<td></td>
<td>67</td>
<td>Finland</td>
</tr>
<tr>
<td></td>
<td>66</td>
<td></td>
</tr>
<tr>
<td></td>
<td>65</td>
<td></td>
</tr>
<tr>
<td></td>
<td>64</td>
<td></td>
</tr>
<tr>
<td></td>
<td>63</td>
<td></td>
</tr>
<tr>
<td></td>
<td>62</td>
<td>USSR</td>
</tr>
</tbody>
</table>