

Graphical Misrepresentation

In today's high information/data consuming society use of statistics is wide spread. Especially, today's electronic and print media very frequently invoke statistics to create attention grabbing news items. But unfortunately, such frequent use of statistics, have led to misuse of statistics. Nothing accentuates the distortion of data than a visual display of data. Use of graphical representation of statistics produces powerful messages. But it is also leaves the door wide open to distort and manipulate graphs to supporting a particular point or portrays wrong interpretation of the data. In today's business community not a day passes without with a meeting with graphs and other visual aid of data. With powerful graphical tools in common use, it becomes very easy for anyone to produce visual presentation of data, again this opens the door for much abuse. Our class leadership presentation will focus what is graphical misrepresentation, how to detect it and how to create "representing graphics". Graphical misrepresentation falls into to two categories. A) Unintentional misrepresentation. B) Intentional misrepresentation.

Unintentional misrepresentation occurs because of oversight or incompetence in the part of the presenter or the creator of the graphical information. And obviously, Intentional misrepresentation occurs with the knowledge of the author. These graphs are created with an intend to mislead the viewers of the actual conclusion of the data or to over dramatize the effects of small changes (make a small increase in sales into a huge visual effect) and in some cases to over simplify the big changes (make a large loss in to small change visually).

There are many ways to manipulate a graph. Below is a list most commonly used tactic to distort and misrepresent the data.

1. Alter the scale (not start with zero, change the intervals)
2. Alter the axis (Change intervals, pick unusual variable for x axis)
3. Avoid showing t he whole context (take it out of context)
4. Add distractions and fillers (take the attention away from the data)
5. Show 1 dimensional data with 2& 3 dimensional figures, which will distort the

increase in data by higher proposition.

6. Bad choice of graphs (Unsuitable graph for the given data)
7. Bad wording and labeling (creates vagueness)
8. Bad use of color (creates false perceptions)

So far we have talked about how can data be misrepresented by graphs. It is essential to put an end to such practices, regardless if it is intentional or unintentional. In order to reduce the number of misrepresenting graphs, the consumer and the creator of the graphs should become graphical competent.

A consumer of the graph should look for the commonly used distortion tactics in graphical misrepresentation. If consumers can notice scale changes, out of context graphs and non-zero starting point, they can easily identify bad graphs. There are many formulas to calculate the effectiveness or lack of it. Statisticians and other users of graphical data have created ratios and percentages such as Lie factor, Graph discrepancy index, Data-ink ratio and data Density to quantify the distortion. These calculations aid a consumer in evaluating and understanding graphical information.

Displays of Statistical information should always reveal the data at several levels of detail, from a broad overview to the fine structure. It should serve a reasonably clear purpose: description, exploration, tabulation and should be closely integrated with statistical and verbal descriptions of a data set. Authors of a graph should be aware of following details when creating a graph.

- There should primacy given to the data
- Graphs should be rich in information.
- Graphs should reveal the data they are trying to display
- Author should always remember the audience he presenting

Authors also should make and attempt to

- Induce the viewer to think about substance rather than about methodology, graphic design the technology of graphic production, or something else
- Avoid distorting the data

- Have a properly chosen format and design
- Reflect a balance, a proportion, a sense of relevant scale
- Display an accessible complexity of detail
- Encourage the eye to compare different pieces of data
- Be drawn in a professional manner

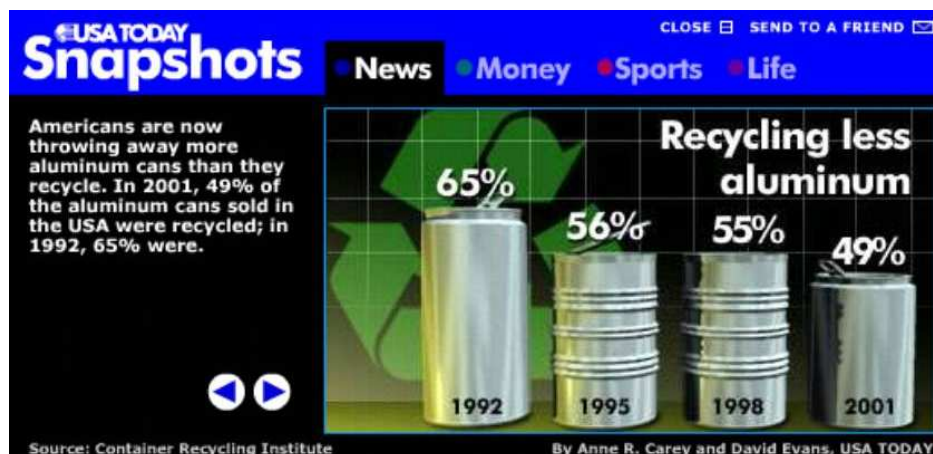
Analysis of Bad Graphs

In the previous section, a number of recommendations were provided for how to rate the effectiveness of charts. In this section, that knowledge will be applied to real world examples. Charts found in newspapers, financial reports, stock reviews and on government web sites will be analyzed using the methods already described. As you'll see, it's surprising which institutions display good graph competence and which don't.

Lie Factor Analysis – USA Today

The USA Today is an excellent source of “bad” graphs. The front page of every issue contains a colorful chart: both in meaning and in visual artistry. These charts seem to exemplify the questionable advice given by Gene Zelazny in his book Say it with Charts, "Choosing the correct chart form depends completely on ... your message. It is not the data that determines the chart. It is not the measure...that determines the chart. Rather, it is your message, what you want to show, the specific point you want to make."

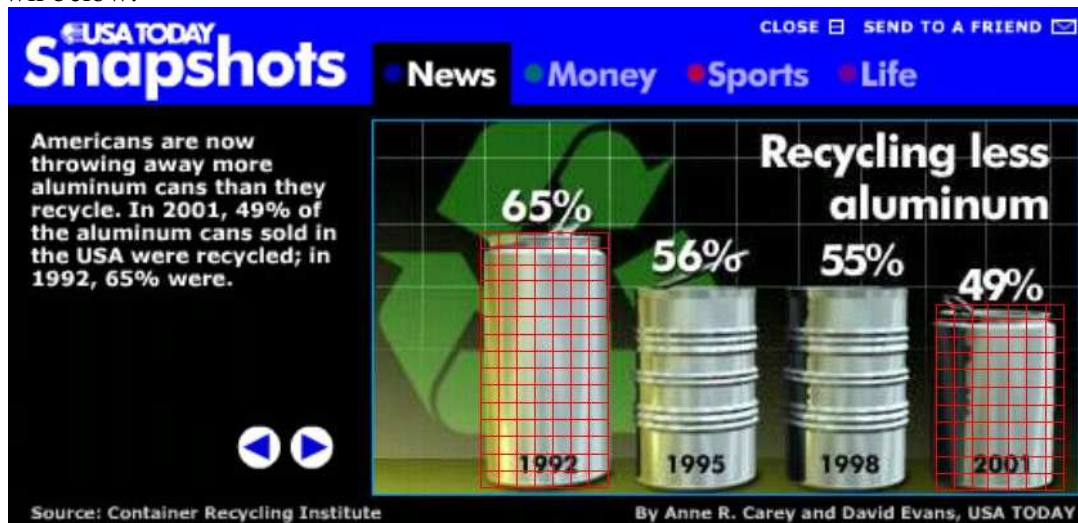
All charts provided in USA Today's Snapshot gallery follow this advice. The charts have a message, and it seems the charts are altered in any way to get that point across. Below is a typical example:



This chart depicts the decrease in percentage of aluminum cans over the years. When reviewing this chart, it breaks many of the recommendations provided in the previous section. First, the baseline is questionable. The starting point of the graph is not labeled; therefore, the reader could assume the baseline is zero. Second, what is the context of the

graph? What was the trend before 1992? Only data points are shown. Continuing with them of data points, only four data points are shown on this entire graph. All the artwork and text can't hide this point. Hence, the data-to-ink ratio is poor and the data density could be improved. Other questions are why are some years represented by vegetable cans (1995 & 1998) and some years represented by soda-pop cans (1992 & 2001). Should this indicate something to the reader? Another artwork question is why are the cans more shaded as the percentages decrease? The 1992 can is well lit, while the 2001 can seems to be in the dark. Is this important?

All these flaws are minor, compared to the lie factor. The calculations for lie factor are shown below.



$$\text{Lie factor} = \frac{\text{Size of effect shown in graphic}}{\text{Size of effect in data}}$$

$$65\% = 7 \times 15 = 105$$

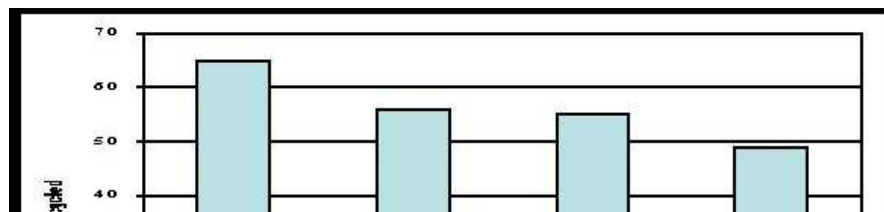
$$49\% = 6.5 \times 10.5 = 68.25$$

$$\text{Lie factor} = \frac{(105 - 68.25)/68.25}{(65 - 49)/49}$$

$$\text{Lie factor} = 1.64$$

Using the red grids laid over years 1992 & 2001, you can see that both the height and width are different. As a result, the area, not just the height needs to be used to calculate the lie factor. The resulting lie factor for the above graph is 1.64. Given the rule provided in the previous analysis that any lie factor above 1.05 is tainted, this is a tainted chart.

Correcting this chart is relatively simple. Removing the errors previously described, results in a simpler graph. Also, the lie factor for this new chart is .77.



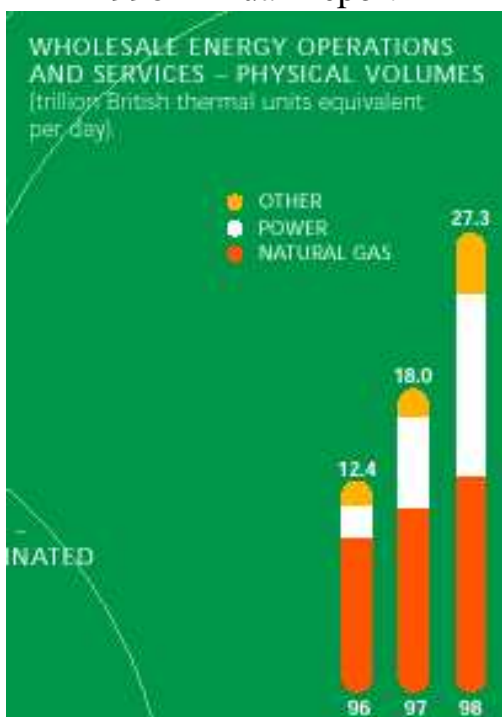
Lie Factor Analysis – Enron Corporation

Finding flaws with USA Today's charts was easy. It also might not be of any ethical concern. If USA distorts the public's opinion about aluminum recycling, is anyone going to be seriously harmed?

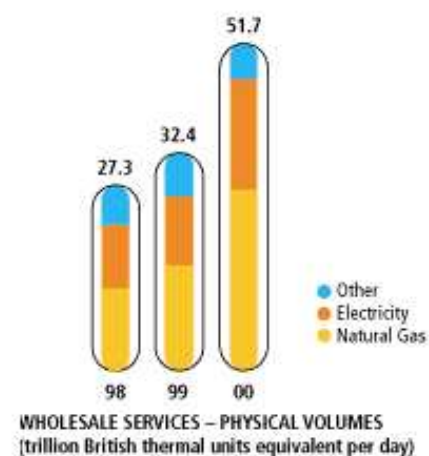
Per the previous section of this paper, much is being written about the ethical use of charts and graphics in financial reviews. Also, the word *ethics* can hardly be mentioned in today's culture without also mentioning Enron. It seemed obvious to combine these two trains of thought and analyze Enron's annual financial reviews for poor graphical use.

Below are two charts taken from the 1998 & 2000 Annual Reports provided on Enron's homepage. Both analyze the same data. Both use similar graphics. The difference, other than color choice, is that the 2000 data is graphed with ovals surrounding each bar.

1998 Annual Report



2000 Annual Report



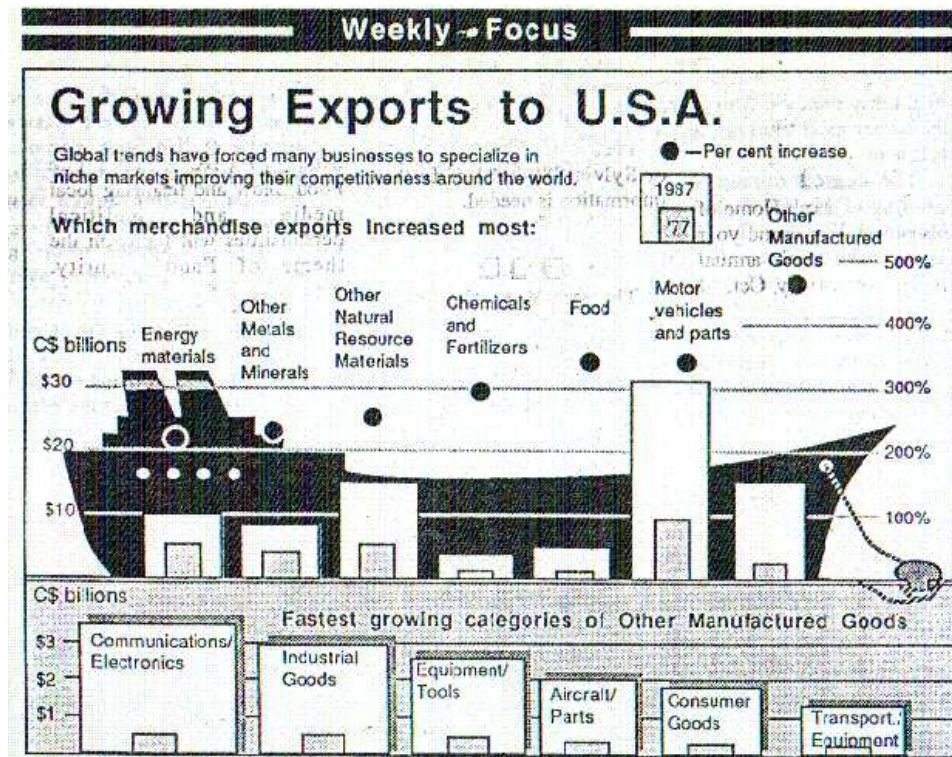
To analyze if this could have any affect on how readers interpret the data, the lie factor is calculated for both years. The calculations follow:

1998 Annual Report	2000 Annual Report
$\text{Lie factor} = \frac{(77.5 - 35)/35}{(27.3 - 12.4)/12.4} = 1.01$	$\text{Lie factor} = \frac{(88 - 52)/52}{(31.7 - 27.3)/27.3} = 0.77$

There is a difference year-to-year. The 1998 data has the most accurate lie factor I had calculated (even graphs I produced using excel did not approach 1.00). What is surprising is that the lie factor for 2000 was 0.77. This, in essence, left readers who compared both reports with the impression that the 2000 growth was not as substantial as the 1998 growth. If I were consulting Enron, I would have urged them to improve their lie factor towards 1.00 to take credit for their growth. But, since hindsight is 20/20, one could also assume that Enron was subconsciously balancing out the over-estimates of the 1998 growth. The most truthful version of any chart will be the one that has the least amount of misrepresentation and distortion. Graphical competence will determine whether the information presented will actually inform or confuse the audience. The most truthful version of any chart will be the one that has the least amount of misrepresentation and distortion. Graphical competence will determine whether the information presented will actually inform or confuse the audience. actual numbers with an under-estimated chart.

Data-to-Ink Ratio – U.S. Exports (Imports)

The next chart that will be analyzed has many problems. The first is the name of the graph itself, “Growing Exports to U.S.A”. Since, this graph describes the amount of goods the US is importing from other countries, a better title might be, “United States imports have increased since 1977”. Even if this suggested title is not completely accurate, it is suggested to reword the title.



The next major issue is the use of two planes of reference to display data. The data in the foreground represent smaller values than those in the background (\$30 Billion compared to \$3 Billion); however, the bar charts in the front are larger than those in the back. This is a lie-factor issue. Another lie-factor issue is that the bars for 1977 are surrounded by the bars for 1987. This creates the misperception that 1987 data largely surpasses 1977 data.

Other issues involve the artwork included in the chart. The ship has a majority of the issues. First, the bow of the ship slopes upwards. A reader might mistake this for a trend line. Similarly, the ship's anchor, the smoke stacks, and the portholes, all could be mistaken for data elements. The second issue is the ink the ship requires. To analyze this, a data-to-ink ratio is performed. In the below chart, the green squares denote any area that contains any data element. Per the previous section, any labels, bars, or numbers that could aid the viewers understanding of the chart are included as data elements. Anything else, is considered wasted ink. Included in this are grid lines and artwork. Given these definitions, the data-to-ink ratio is calculated below.



$$\text{Data-to-Ink Ratio} = \frac{\text{Data Ink}}{\text{Total Ink Used in Graphic}}$$

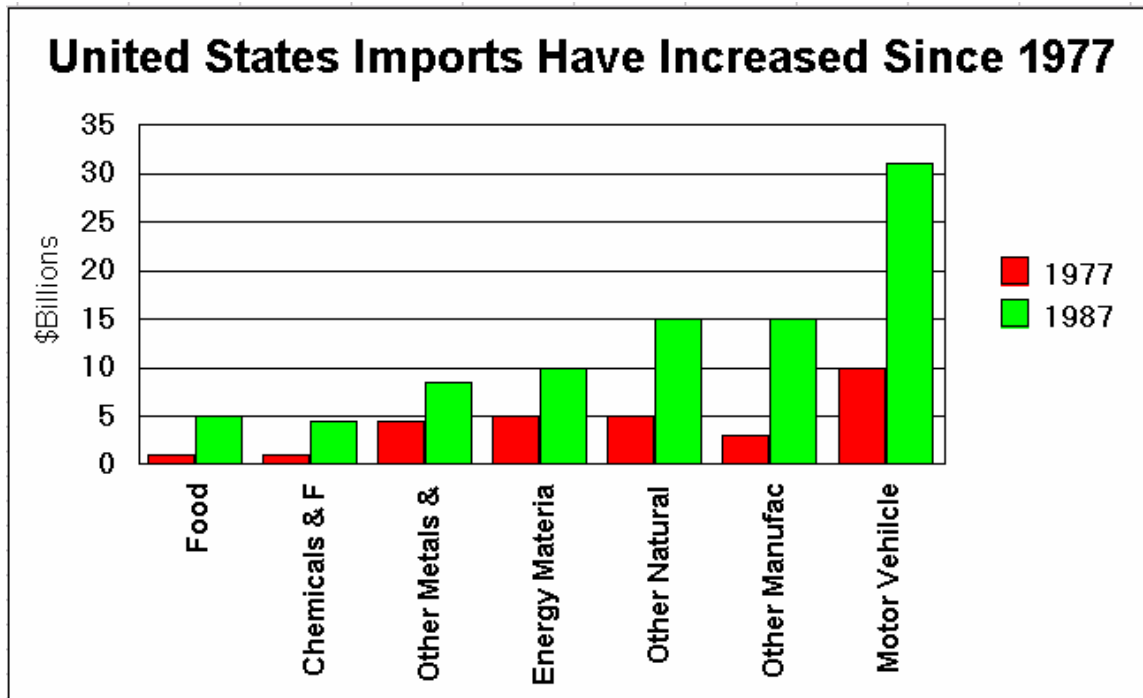
$$\text{Data Ink} = \text{Green} = 382$$

$$\text{Total Ink} = \text{Green} + \text{Red} = 640$$

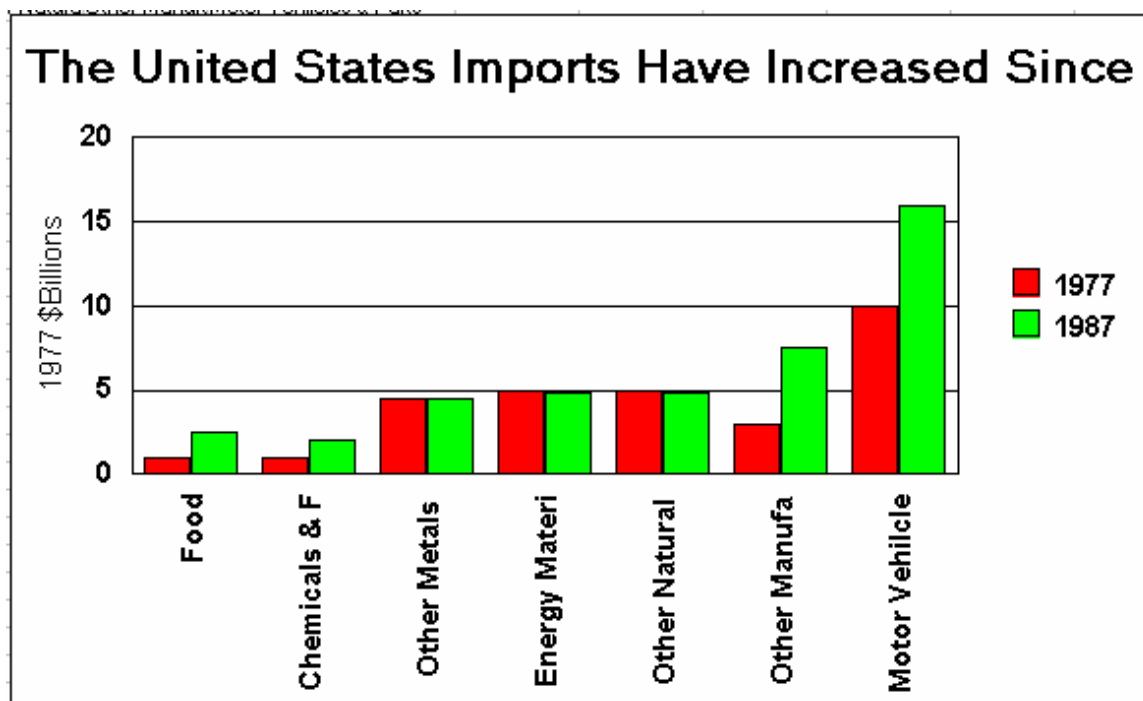
$$\text{Data-to-Ink Ratio} = 59.6\%$$

This means 40% of the information this graph displayed contained no data that was useful for the reader.

To correct this graph, another issue needs to be addressed that hasn't yet been discussed: the dollar amounts need to be corrected for inflation. Below a scrubbed chart with inflation is shown:



The differences between 1977 & 1987 still appear to be substantial. Next, a chart corrected with inflation is shown:



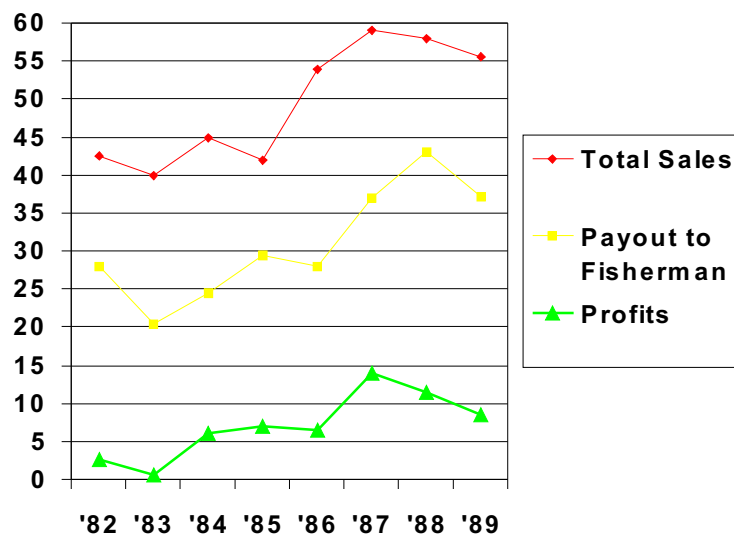
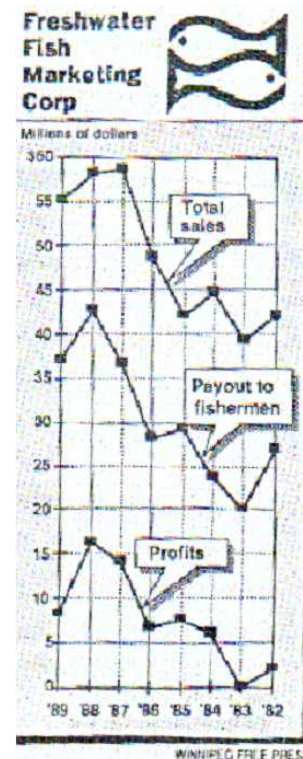
Now the data isn't as shocking. Motor Vehicle and Other Manufacturing have increased, but the other categories show less change.

Simple Mistakes with Major Impact – Freshwater Fish

This example shows how a chart that follows all the recommendations made in this paper could still be misleading. The following chart shows the changes in the fish industry between 1982 and 1989. The data has a good lie-factor. The data-to-ink ratio is not a problem. There is a zero point. There is clear context because enough years are shown. No colors are influencing the reader. The scale is constant, showing a data point for each year.

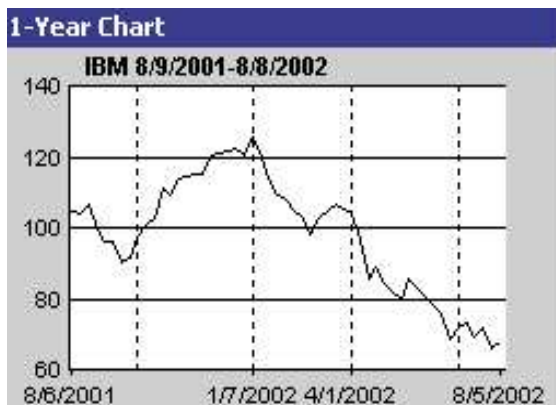
Have you found the flaw yet? If not, the date range is listed in reverse chronological order. As a result, the unobservant reader could easily mistake the trend. They might read that the past year increased after years of decline, while in reality, there was a decline after a few years of improvement.

Another minor problem with the graph is that the axes are slightly skewed making the graph taller and thinner. This gives the impression that the yearly changes are more severe than might be the case. An improved graph appears below.



Skewed Scales -- Stock Analysis

One last analysis involves the stock market. Given technological advancements, many websites help investors track stocks by offering a variety of tools. Probably the most widespread tool is a stock price tracker. This creates a chart of the closing prices of stocks given a date range specified by the user. Below are graphs of the closing price of stocks for a one-year period, between 8/9/2001 and 8/8/2002.



What may not be obvious is that all these charts refer to the same company. Looking at each again, it appears that the different graphs display slightly different trends. The top-left chart is the most square and displays a sharp drop-off in price. The top-right chart is more rectangular and the decline seems less severe. The bottom chart is more rectangular still, and the decline appears even less drastic.

This is caused by a skewed scale. Adding to this effect is that the baseline of \$0 isn't shown, but all graphs have a baseline of \$60 dollars, so are all distorted the same way. It would be interesting to see if users of these different services would react to the stock price changes in dissimilar ways, given that their data is graphed differently.

Additional helpful information to keep in mind when working with charts:

I. Using the appropriate chart for your presentation

Pie charts

Pie charts should be used to show any proportional relationship between a slice and the whole pie as a percentage, a fraction, a ratio or decimal. Pie charts should never be used to show values. When actual values are given for pie slices, the audience will be tempted to add them to find how much to total is. If actual values are essential for your presentation, use another type of chart.

Stacked Bars

Stacked bars chart is one of the best ways to show relative sizes of different segments as well as their actual amounts.

Stacked Areas

Stacked areas like stacked bars can be used to show both relative sizes and actual values. If the areas are truly stacked, only the bottom data series will have a flat line. Please note, that distortions between stacked areas can be minimized by including the area with the least fluctuations on the bottom.

XY Charts

XY charts are a good tool to plot quantity versus time, amounts increase along the vertical scale, or y axis. Time progresses from left to right along the horizontal scale, or x axis.

Line Charts

Line charts can highlight trends in data better than any other chart type. Oversimplifying the data is the pitfall with this kind of chart.

II. Labeling

As a good practice, whenever a number is included in a graph, there should be a label nearby.

III. Source Identification

Use of footnotes in charts to identify sources of information increases credibility.

IV. Orientation

Viewers get very different impressions depending on whether shapes appear to go left, right, up or down. The basic assumptions behind the Cartesian coordinate system and the xy chart is that time flows from left to right, and the amount of the item being measured and charted fluctuates up or down.

V. Colors

When colors are used as codes, the color scheme must be as simple as possible, with the least number of colors necessary.

The most truthful version of any chart will be the one that has the least amount of misrepresentation and distortion. Graphical competence will determine whether the information presented will actually inform or confuse the audience. Each creator and user of graphical information should be aware of the potential misuse as well as how to offset them by being a more competent consumer.

References

- CNBC Money. (2002). Retrieved August 2, 2002, from <http://moneycentral.msn.com/investor/home.asp>
- CNN Money. (2002). Retrieved August 2, 2002, from <http://qs.money.cnn.com/>
- Enron Corporation. (2002). Retrieved August 2, 2002, from <http://www.enron.com/corp/investors/>
- Harris, Robert L. (2000). Information Graphics: A Comprehensive Illustrated Reference. Retrieved August 2, 2002, From http://www.amazon.com/exec/obidos/ASIN/0195135326/qid=1028308514/sr=1-1/ref=sr_1_1/103-2954531-0058235
- IBM. (2002). Daily stock performance snapshot (from IDD). Retrieved August 2, 2002, from <http://www.ibm.com/investor/stock/stockchart.html>
- Schwarz, Carl J. (1998). Exports to the US. Retrieved August 2, 2002, <http://www.math.sfu.ca/~cschwarz/Stat-301/Handouts/node13.html>
- Schwarz, Carl J. (1998). Sales of Seafood. Retrieved August 2, 2002, <http://www.math.sfu.ca/~cschwarz/Stat-301/Handouts/node10.html>
- U.S.A Today. (2002). Snapshots. Retrieved August 2, 2002, from <http://www.usatoday.com/>
- U.S. Census Bureau. (2002). 2001 Public Employment Data State and Local Governments UNITED STATES TOTAL. Retrieved August 2, 2002, from <http://www.census.gov/govs/apes/01stlus.txt>
- U.S. Department of Labor. (2002). A Prepared Workforce. Retrieved August 2, 2002, from http://www.dol.gov/_sec/media/reports/annual2001/goal1.pdf#page=4
- Washington Post. (2002). Retrieved August 2, 2002, from <http://www.washingtonpost.com/>
- Zelanzy, Gene. (2001). Say it with Charts: the Executive's Guide to Visual Communication. Retrieved August 2, 2002, From http://www.amazon.com/exec/obidos/ASIN/007136997X/qid=1028301647/sr=2-1/ref=sr_2_1/103-2954531-0058235
- E. Tufte, *The Visual Display of Quantitative Information*
- E. Tufte, *Envisioning Information*
- E. Tufte, *Visual Explanations*
- Gerald E Jones, (1995) *How to Lie with Charts*.