III. Data into Information

High Information Content Graphics

There is often too much information available to be analyzed with simple x-y plots. Data acquisition requires resources (time, effort, and money). It is important to extract as much information as possible from this data and to communicate it effectively. This additional investment in time and thought is a high value activity and can lead to better insight into the analysis of the problem.

The examples below show illustrate the principles of good design and also the consequences of ineffective graphs

Principles of an excellent visual (E. Tufte Presentation of Visual Information)

Shows the data individually

Induces the reader to think about the substance, not anything else

Encourages the eye to compare different pieces of data

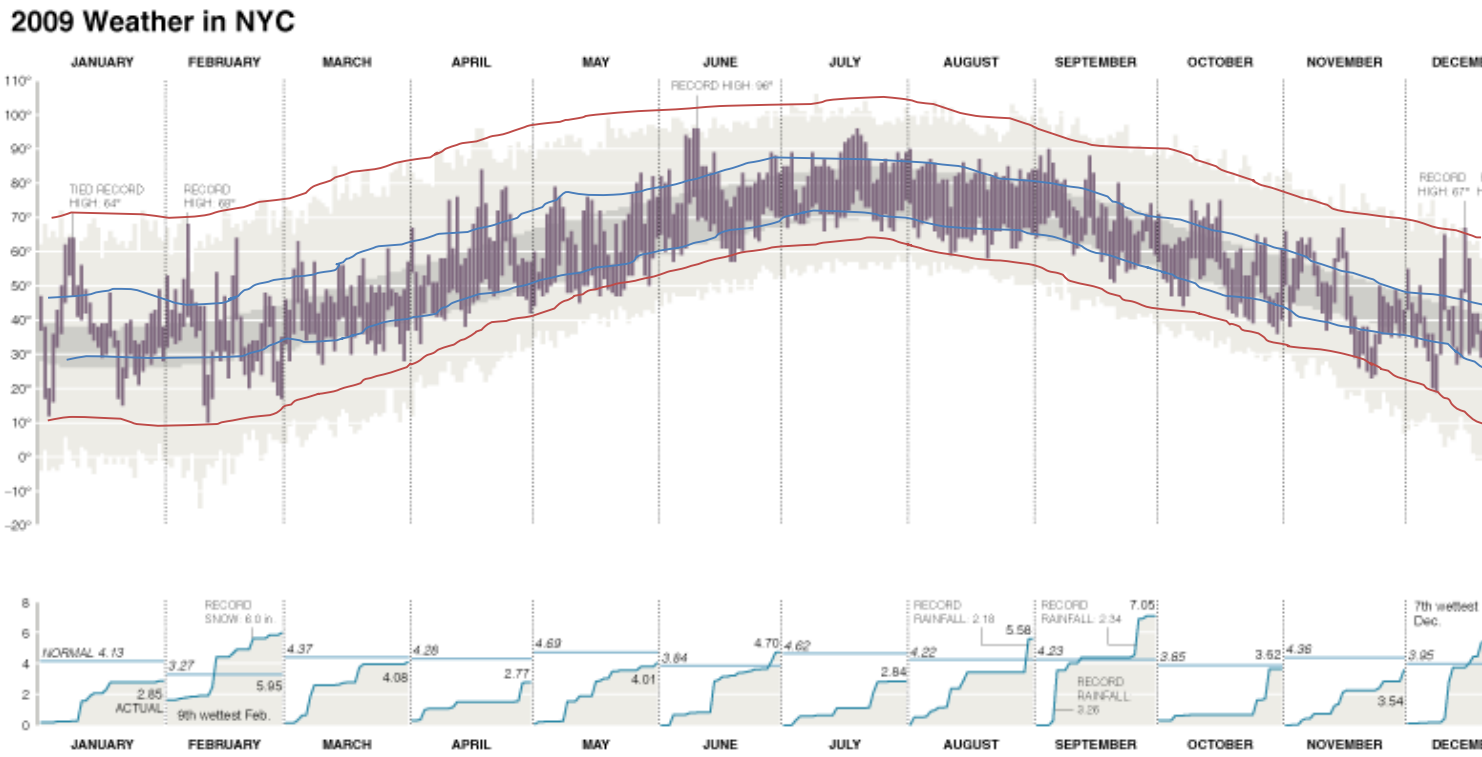
Reveals the data at several levels of detail

Makes large data sets coherent

Avoids content free decoration

Example of Excellence: New York Times Annual Weather Graphic

Each year the NYT, publishes an annual review of the weather in New York. Examine the graphic specifically in terms of the way that the principles above are followed. Make a written comment of an example for each one. Extract as much information as you can from the graphic. Then compare your list to the one of the following page. Describe your own process of observation and comparison.



Temperatures:

For context: The normal high and low for each day is shown

The record high and lows are indicated. (In this reproduction, these are indicated by the blue lines.)

Annotations are made for record highs and lows (5 record highs, 0 record lows)

Seasonal temperature trends are shown

Differences in daily temperature variations in summer and winter are evident

Rainfall

For context: The magnitude of each storm is shown

Annotations are made for record accumulations

The monthly accumulation is shown for comparison

Annotations are made for months that are near records

Not shown: Annotation for annual rainfall compared to average.

Summary: There is a high density of both data and information. The information is placed into a larger context. The format invites the reader to really engage with the graphic and extract more information than is available from a casual look. There is no excess ink

Data Represented in Non-Traditional Formats to Encourage Analysis

Example: Mapping the 1854 London Cholera Outbreak

Dr. John Snow is regarded as one of the founding fathers of modern epidemiology. As London suffered a series of cholera outbreaks during the mid-19th century, Snow theorized that cholera reproduced in the human body and was spread through contaminated water. This contradicted the prevailing theory that diseases were spread by "miasma" in the air.

London's water supply system consisted of shallow public wells where people could pump their own water to carry home, and about a dozen water utilities that drew water from the Thames to supply a jumble of water lines to more upscale houses. London's sewage system was even more ad hoc: privies emptied into cesspools or cellars more often than directly into sewer pipes. So the pervasive stench of animal and human feces combined with rotting garbage made the miasma theory of disease seem very plausible. Disease was more prevalent in lower-class neighborhoods because they stank more, and because the supposed moral depravity of poor people weakened their constitutions and made them more vulnerable to disease.

The September 1854 cholera outbreak was centered in the Soho district, close to Snow's house. Snow mapped the 13 public wells and all the known cholera deaths around Soho, and noted the spatial clustering of cases around one particular water pump on the southwest corner of the intersection of Broad (now Broadwick) Street and Cambridge (now Lexington) Street. He examined water samples from various wells under a microscope, and confirmed the presence of an unknown bacterium in the Broad Street samples. Despite strong skepticism from the local authorities, he had the pump handle removed from the Broad Street pump and the outbreak quickly subsided.

Snow subsequently published a map of the epidemic to support his theory. A detail from this map is shown below. The complete map shows the locations of the 13 public wells in the area, and the 578 cholera deaths mapped by home address, marked as black bars stacked perpendicular to the streets.

Some anomalies are worth noting. Although the large workhouse just north of Broad Street housed over 500 paupers, it suffered very few cholera deaths because it had its own well (not shown on the map). Likewise, the workers at the brewery one block east of the Broad Street pump could drink all the beer they wanted; the fermentation killed the cholera bacteria, and none of the brewery workers contracted cholera. Many of the deaths further away from the Broad Street pump were people who walked to work or market on the Broad Street and drank from that well. The water from the Broad Street well reportedly tasted better than water from most of the neighboring wells, particularly the smelly water from the Carnaby Street/Little Marlborough Street well a few blocks to the northeast.

Steven Johnson's 2006 book *The Ghost Map: the Story of London's Most Terrifying Epidemic, and How it Changed Science, Cities and the Modern World* (available in paperback) is a highly entertaining account of the epidemic and Snow's analysis of it.

Reference and additional maps: http://www.udel.edu/johnmack/frec682/cholera/

Ineffective Graphics

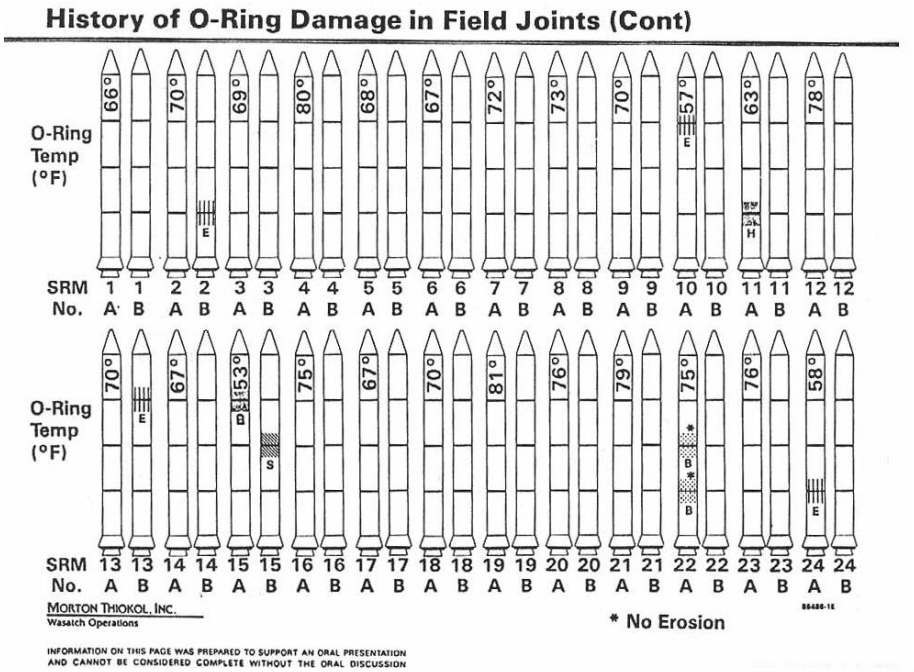
Space Shuttle Challenger O-ring Damage:

There are many articles and analyses about this disaster. The objective here is not to debate the motivations or culpability, but to examine two different graphics that represent the relevant data

According to one version: On January 28, 1986, NASA managers had to decide if the circular rubber O-rings used to join sections of the orbital spaceship could perform as designed. Unusually cold for the Florida launch pad, temperatures had dipped to about to -2 ^{o}C (28^{o}F). Since the O-rings lose flexibility as the temperature decreases, there was real concern, that there would flex enough to work. There was no performance data at temperatures below 12 ^{o}C.

There were many meetings with technical and political pressures brought to bear on both engineers and managers. The engineers at Morton Thiokol initially recommended against launch at these conditions. However, under pressure, Morton Thiokol management concluded that their engineers did not have proof the flight would fail and that the design would handle the temperature. The Shuttle was launched and exploded within 73 seconds.

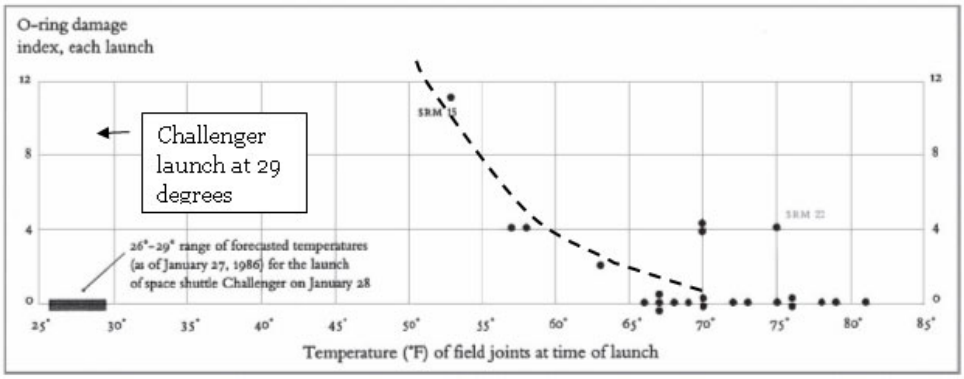
This is the diagram presented at the meeting. All of the data is contained in the diagram. The rockets are listed in chronological order. The temperature of the o-ring at launch is clearly indicated and relative damage to a seal is indicated at the appropriate location.



Difficulties:

The diagram of the rockets occupy the most ink and do not add to the analysis

The chronological listing obscures analysis of the temperature, which is the variable of most concern.



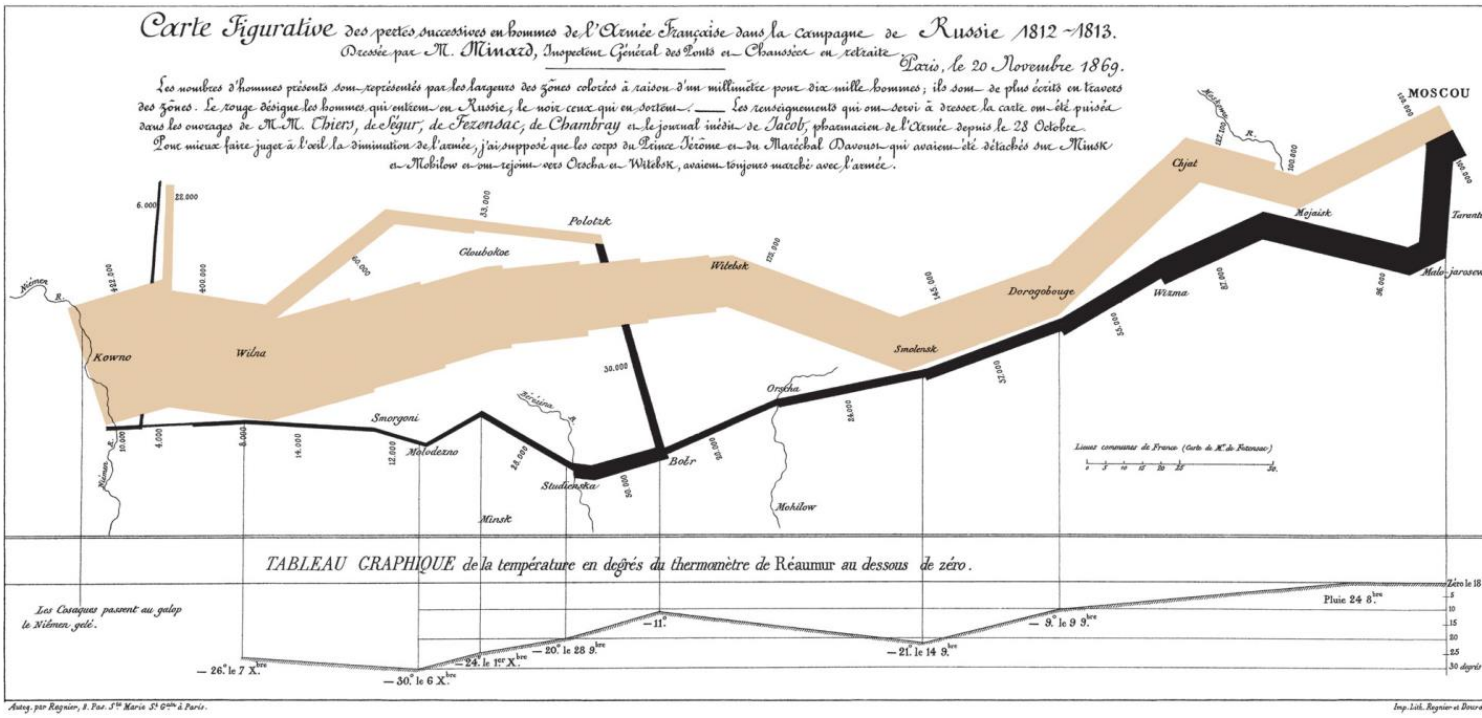
Plot of the Same Data (in Visual Display of Quantitative Information)

This representation shows that the projected launch condition is well outside of experience and that the O-ring performance deteriorates as the temperature at launch decreases.

An Effective Graphic Presents the Data to Clarify the Problem

Classic Graphic: Napolean's Russian Campaign of 1812:
Multi-Variables in a single graphic

There are 6 variables plotted here: Size of the army, location in two dimensions, direction of the army's movement, temperature on various dates.



English: The map's French caption reads:

Figurative Map of the successive losses in men of the French Army in the Russian campaign 1812-1813. Drawn up by M. Minard, Inspector General of Bridges and Roads in retirement. Paris, November 20, 1869.

The numbers of men present are represented by the widths of the colored zones at a rate of one millimeter for every ten-thousand men; they are further written across the zones. The wider band (brown) designates number of men who enter into Russia, the thinner black line those who leave it. —— The information which has served to draw up the map has been extracted from the works of M. M. Thiers, of Segur, of Fezensac, of Chambray, and the unpublished diary of Jacob, pharmacist of the army since October 28th. In order to better judge with the eye the diminution of the army, I have assumed that the troops of prince Jerome and of Marshal Davoush who had been detached at Minsk and Moghilev and have rejoined around Orcha and Vitebsk, had always marched with the army.

The scale is shown on the center-right, in "lieues communes de France" (common French league) which is 4,444m (2.75 miles).

The lower portion of the graph is to be read from right to left. It shows the temperature on the army's return from Russia, in degrees below freezing on the Réaumur scale. (Multiply Réaumur temperatures by 1¼ to getCelsius, e.g. −30°R = −37.5 °C) At Smolensk, the temperature was −21° Réaumur on November 14th.

Problems

1. **Extracting information and conclusions from visual representations**
   This visual appeared in the New York Times earlier this year.

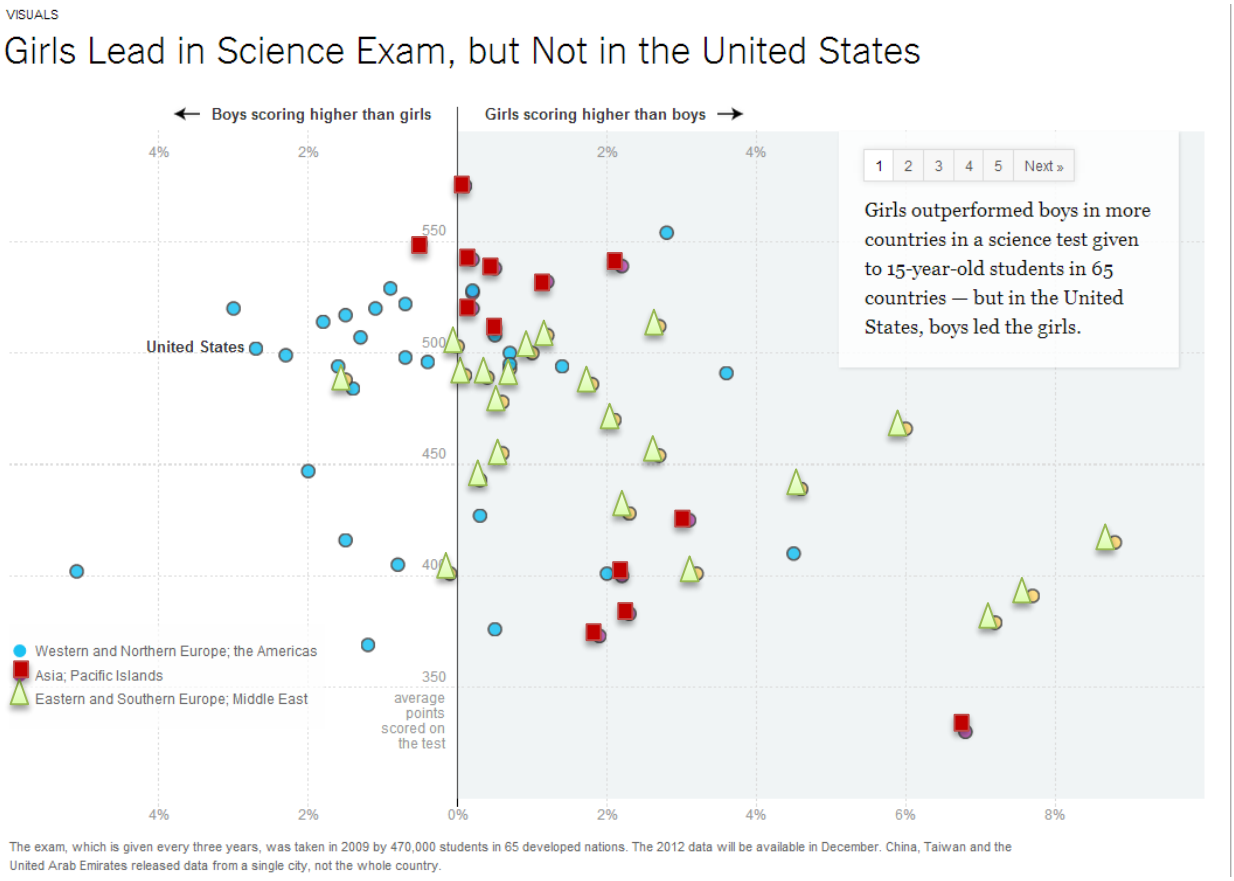
   First look closely at the graph. Observe different data sets.

   Formulate different questions about the relationships of the data based on your initial observation.

   (Reference: Asking good quick questions)

   Analyze the data to draw as many conclusions about the results as you can.

   Make a list of your conclusions.



2. **Constructing a Quality Graphic** (F Sp13 F12) There is a debate about which groups in an organization use the work day effectively. The data table below shows how much time taken for lunch for four different groups in the organization. Your job is too effectively communicate this information in a visual format. Show a more effective way to communicate this information in a visual format using the principles. Sketch the graphic so that your approach is clear.

Time taken for lunch break by different groups (minutes)

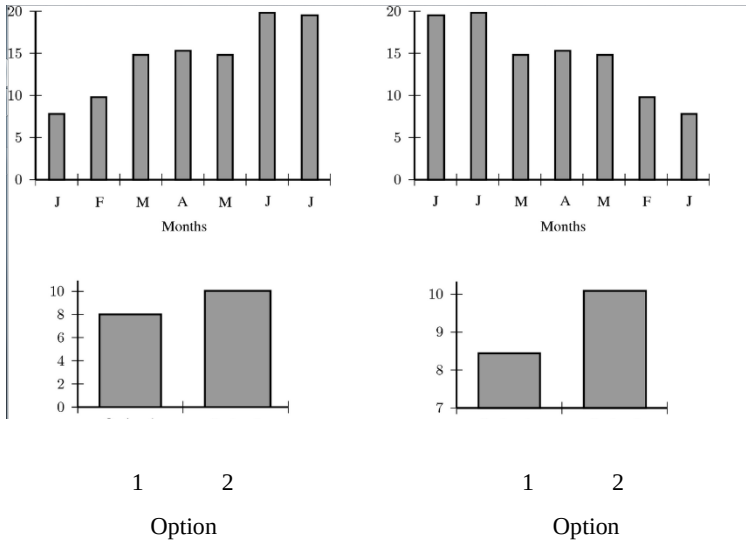
Group	0 (no lunch break)	<30minutes	30-60 minutes	>60 minutes
Marketing	5	44	38	13
Sales	10	21	41	28
Finance	6	1	21	72
Administration	0	15	81	4

3. **Critique** (Spr13 E2) Analyze each of the two sets of graphic to the right in terms of the principles. Indicate the quality of information and the strengths and weaknesses of each.

Indicate the quality of information, the conclusions, and the strengths and weakness of each graph.

a) Units Sold over a 7 month period

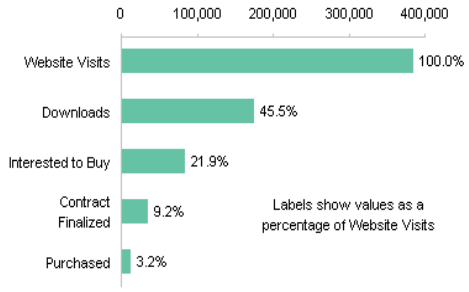
b) Comparison of Production Options



4) **Critique** (E2F12) A web site sells software apps. The product can be downloaded, but is not activated until the contract is finalized and the actual purchase completed. There are many more initial inquiries than actual sales. The two data representations below depict the same information.

Using the principles of graphical analysis, critique each graphic and give the reasons for selection of the most appropriate graphic.

Graphic A Critique



Graphic B Critique