

Data Visualization



For Graphs & Charts

Peggy Parskey

September 1, 2009

KnowledgeAdvisors

Global Leader
in Learning and
Human Capital
Analytics
Solutions

We focus on **learning and human capital analytics solutions** and technologies that help organizations measure, communicate and improve the impact of their employees, customers and partners by better **managing processes through reliable metrics.**



Webinar Objectives

By the end of this webinar you will be able to:

- Describe, at a high level, the field of Data Vis
- Articulate why it is important to understand data visualization concepts
- Articulate how we perceive information
- Take theory to action
- Match your graphs to your story



*Visual representation of 718,000 book results from an Amazon search on the word "Data."
-- Amaztype*



1-Sep-09

© 2009, KnowledgeAdvisors

POLL #1

Page: 3

The Field



*"Data visualization is an evolving concept whose boundaries are continually expanding and, as such, is best defined in terms of loose generalizations."
---Wikipedia*



1-Sep-09

© 2009, KnowledgeAdvisors

Page: 4

What is data visualization?

The process of representing abstract business or scientific data as images that can aid in understanding the meaning of the data.

Important points

- It's a process
- The goals:
 - Communicate information clearly and effectively through graphical means
 - Quickly reveal the meaning in the data
- Aesthetic form and functionality must go hand-in-hand

Who are the respected voices today?

- Edward Tufte: *Visual Display of Quantitative Information*
- Stephen Few: *Show Me the Numbers; Now You See It*
- Gene Zelazny: *Say it With Charts*
- Naomi Robbins: *Creating More Effective Graphs*
- Howard Wainer: *Graphic Discovery*
- William Cleveland: *Visualizing Data*
- Robert Horn: *Visual Language*



The Challenge



"I would not give a fig for the simplicity this side of complexity, but I would give my life for the simplicity on the other side of complexity."

--Oliver Wendell Holmes, Jr (1841-1935)



1-Sep-09

© 2009, KnowledgeAdvisors

Page: 7

Information and data trends

- Technical information is **doubling every 2 years**
- In 2008, humans on earth collectively generated **4 exabytes of unique information** (4.0×10^{19})
- This is more information than humans generated in the **previous 5,000 years**

The incredible growth of the digital universe will cause individuals to face an information explosion on an unprecedented scale.



1-Sep-09

© 2009, KnowledgeAdvisors

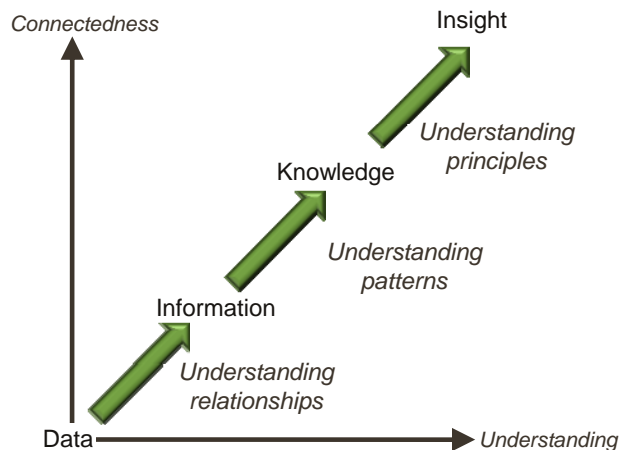
Page: 8

The challenge of the last mile...

- A telecom phrase that refers to “the final leg of delivering connectivity from a communications provider to a customer.”
- It is an expensive and complex step
- And we deal with it all the time with our data



The goal: transform data into insights



Your job:

- Create new knowledge
- Spark new connections
- Reduce the time to insight
- Improve decision-making
- Shift the focus from reflecting on the past to peering into the future

How We Perceive Information



"I do believe that there are some universal cognitive tasks that are deep and profound - indeed, so deep and profound that it is worthwhile to understand them in order to design our displays in accord with those tasks."

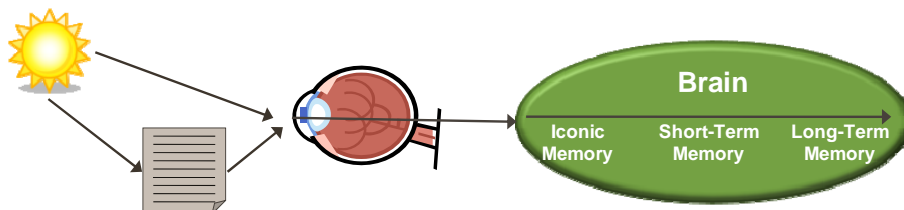
--Edward Tufte



Understanding how we perceive

"The way in which problems are represented has much to do with the quality of the solutions that are found."

– Herbert Simon, 1978 Nobel Prize winner in Economics



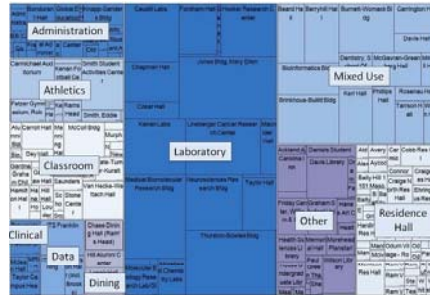
★ The goal: Move your information from the eyeball into long term memory



How to achieve the goal? Think design

- Design is crafting communications to answer audience needs in the most effective way
- The more you focus on design, the more you'll speak to your audience - so you will be more effective with your data presentation.

It's about the audience, not you.



Greenhouse gas emissions by buildings on the campus of the University of North Carolina at Chapel Hill



Graphical interpretation capabilities

In 1984, William Cleveland conducted experiments to measure the accuracy of viewers to perform elementary perceptual tasks. He found accuracy skills rank as follows:

- Position along a common scale
- Position along identical, non-aligned scales
- Length
- Angle-slope
- Area
- Volume
- Color: hue, saturation, density

Conclusion:
Design graphics according to the abilities of viewers to *decode* the graphic elements.



From Theory To Action



“Computers are incredibly fast, accurate and stupid; humans are incredibly slow, inaccurate and brilliant; together they are powerful beyond imagination.”

– Albert Einstein

1-Sep-09

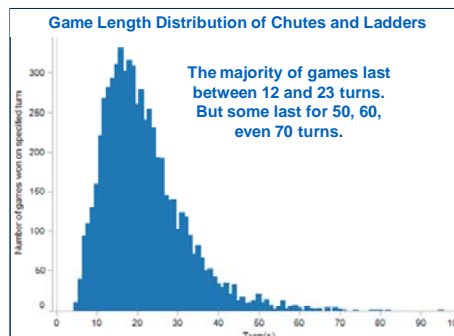
© 2009, KnowledgeAdvisors

Page: 15

Tufte's display principles

Well designed displays of statistical information:

- Have a properly chosen format and design
- Use words, numbers and drawing together
- Reflect a balance, a proportion, a sense of relevant scale
- Display an accessible complexity of detail
- Often have a narrative quality, a story to tell
- Avoid chart junk

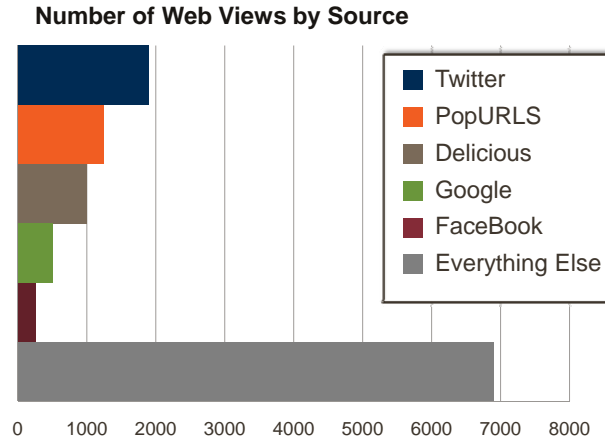


1-Sep-09

© 2009, KnowledgeAdvisors

Page: 16

What's wrong with this chart?



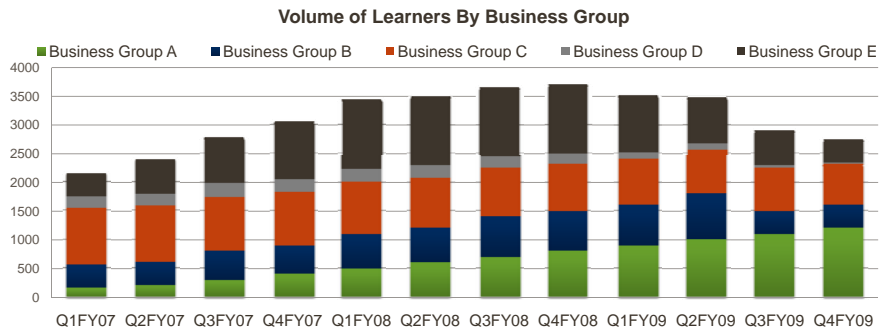
POLL #4

1-Sep-09

© 2009, KnowledgeAdvisors

Page: 17

What's wrong with this chart?



1. How much did volume for Business Group C (orange) change from quarter to quarter?
2. How many times do you have to reference the legend to figure out which color is what label?
3. Is there even a value of Business Group D (grey) after Q3FY09?



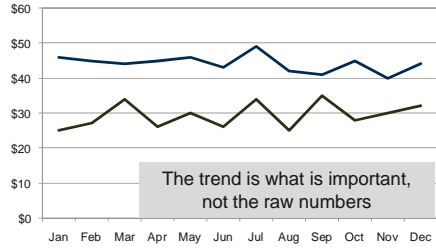
1-Sep-09

© 2009, KnowledgeAdvisors

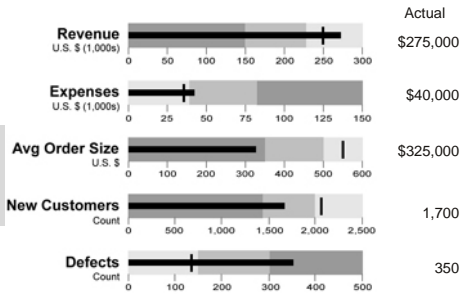
Page: 18

Enhance the data “ink”

- Eliminate unnecessary data ink
- Highlight the most important that remains
 - Information that is ‘always important’
 - Information that is important now (e.g. exceptions, below target)



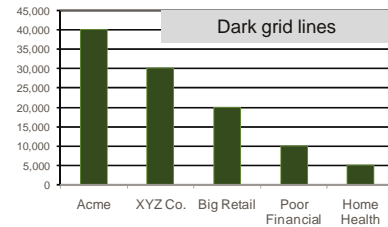
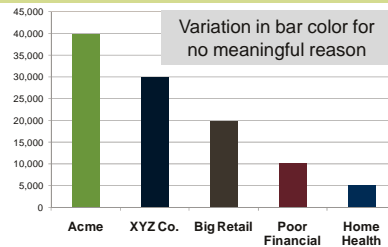
Only “Actuals” are shown numerically. Performance is compared to target and depicts it against mins and acceptable performance via shading.



Reduce non-data “ink”

- Eliminate unnecessary chart junk
- De-emphasize and regularize the non data ink that remains

Non-data “ink”: anything on your charts and graphs that is not actual data, including scales and legends

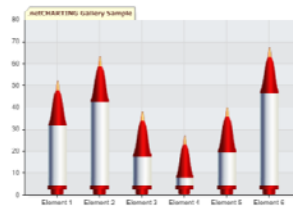


Graphics solely for decorative purposes



Above all else show the data*

Do this	Not this
Focus on the data and the message	<ul style="list-style-type: none"> Clutter the output with useless information Misuse or overuse color
<ul style="list-style-type: none"> Provide context Just enough detail Clean presentation 	<ul style="list-style-type: none"> Inadequate context for the data Excessive detail / precision Poor data arrangement Meaningless variety
<ul style="list-style-type: none"> Encode your data appropriately Point to what's important 	<ul style="list-style-type: none"> Encoding quantitative data inaccurately No or ineffective highlighting of important data

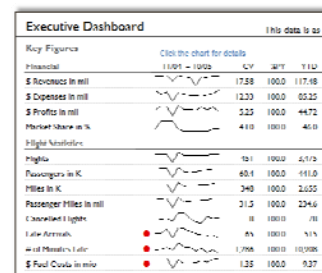
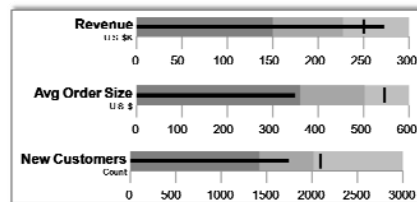


*Edward Tufte



Condense information

- Summaries and exceptions
 - Use sums and averages
 - Present at level of precision appropriate to what's required
- Display media
 - Must be the best way to display a particular type of information
 - Must serve its purpose even when sized to fit into a small space

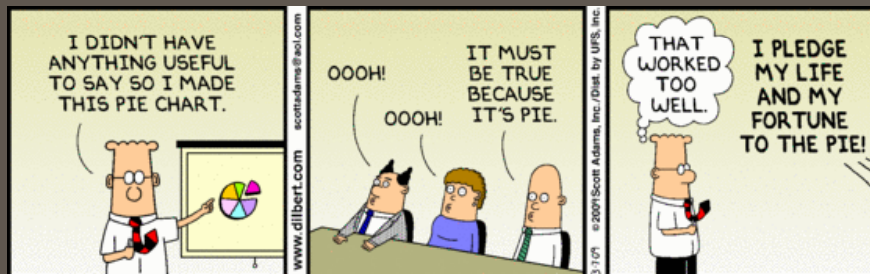


Using color: A “Few” rules

- **When to use color:** When it serves a purpose or conveys meaning
- **How to use color**
 - Use soft, natural colors to display most information
 - Use bright / dark colors to call attention to key points
 - Consider color blind people
- **Color for backgrounds:** Consistency and contrast
- **Color for non data elements:** just visible enough to perform their role

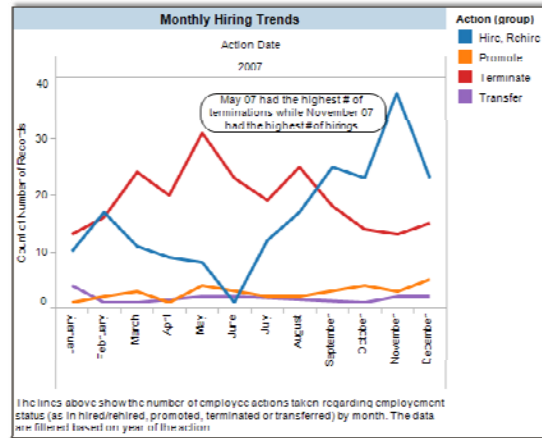


A Brief Guide to Graphs



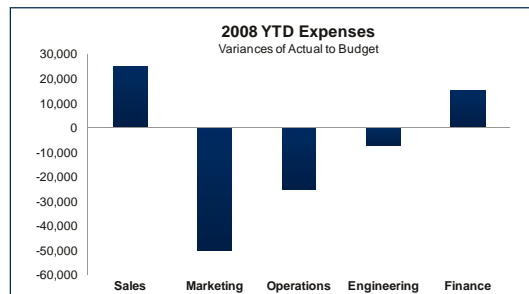
Line graphs

- Good at showing trends over time
- Show variability in the data
- Best used when you want to show the shape of the change vs individual values
- Enables predictions about the results of data not shown



Bar graphs

- Good at comparing differences among variables in small data sets
- Excellent for showing ranking and part to whole data
- Focuses on the individual values
- Can encode data along nominal, ordinal and interval scales

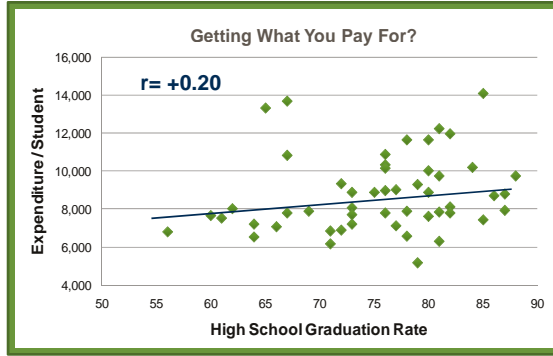


When using a bar graph, begin the scale at zero (or the lowest possible value)

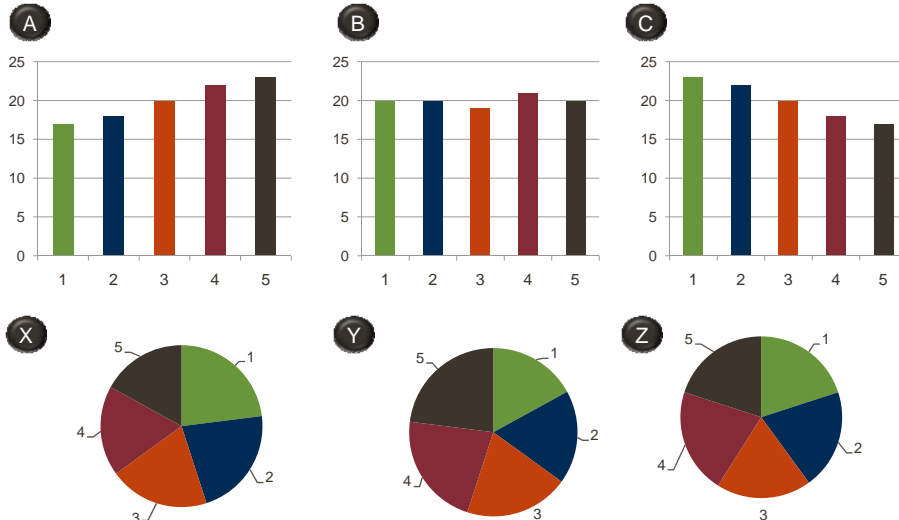


Scatter plots

- Depicts how much one variable is affected by another.
- Encourage us to notice patterns: clusters, gaps, outliers, data relationships
- Useful with a large body of data.

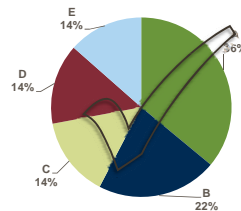
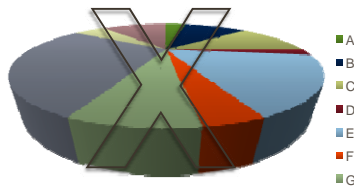


Match the pie to the equivalent bar chart



Pie charts

- Avoid pie charts
 - They take up a lot of space
 - Difficult to discern area and angles
 - Lack common baseline
- If you must use them
 - Limit to ~5 segments
 - Sort by numerical value
 - Use 2D pie charts **only**
 - Put data labels on segments vs a legend
 - Eliminate chart junk



1-Sep-09

© 2009, KnowledgeAdvisors

Page: 29

Do you even need a graph?

A table is best when:

- You need to look up specific values
- Users need precise values
- You need to compare related values (e.g. sales in Q1 vs Q2)
- You have multiple data sets with different units of measure

A graph is best when:

- The message is contained in the shape of the values
- You want to reveal relationships among multiple values (similarities and differences)
- You have large data sets

Graphs and tables serve different purposes. Choose the appropriate data display to fit your purpose. A table may be sufficient.



1-Sep-09

© 2009, KnowledgeAdvisors

Page: 30

Five key points

1. Consider your audience
2. Design data visuals to ensure your message sticks
3. Choose the right chart for the right message
4. Provide context with the data
5. Emphasize data ink, de-emphasize non-data ink; be careful when using color



Questions?



www.knowledgeadvisors.com

For further information on the information contact:

Peggy Parskey

Email: pparskey@knowledgeadvisors.com

Phone: 323-931-6589

Fax: 323-933-2961

1-Sep-09

Page 33

Poll # 1

- What involvement do you have in presenting data to your organization?
 - I create the charts, graphs and data visuals and present the findings
 - I create charts, graphs and data visuals that others present to our leaders
 - I analyze data but do not create the visualizations
 - None of the above

Poll #2

- My graphs and data visualizations are:
 - Fantastic: I'm the next Edward Tufte
 - Pretty good: People often comment on the clarity of my graphs and charts
 - Only ok: They get the point across but could be better
 - Need work: I get a lot of questions on how to interpret my graphs
 - Not good at all: My charts are in serious need of improvement
 - I have no idea about the quality of my charts

Poll #3

- When I create charts and graphs, I consider my audience and their ability to absorb and comprehend the information
- Poll answers
 - Always
 - Most of the time
 - Some of the time
 - Rarely
 - Never

Poll #4

- Color unnecessarily varied by data point.
- Use of a legend instead of axis labels.
- Forces the eyes to shift down and up to perceive the information
- Extremely large residual category
- All of the above

Match the pie to the equivalent bar chart

