What is Visualization?

Dr. Gaurav Bansal

Visualization

“Computer based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively”

Visualization Analysis & Design, Munzner (2014), p. 1

Visualization helps people transform data into actionable insights.
(adapted from Tableau.com)

A picture is worth a thousand words – especially when you are trying to find relationships and understand your data – which could include thousands or even millions of variables.
(from SAS.com)

The eye and the visual cortex of the brain form a massively parallel processor that provides the highest-bandwidth channel into human cognitive centers;**10 Mbps equivalent to high speed Internet**, and faster than any other sensory system; around 100 times faster than ears!

Information Flow in Sensory Systems and Conscious Perception

Sensory System	Total Bandwidth (Bits/second)	Conscious Bandwidth (Bits per second)
Eyes	10,000,000	40
Ears	100,000	30
Skin	1,000,000	5
Taste	1,000	1
Smell	100,000	1

Cf. *The User Illusion* by Norretranders

Why should we be interested in visualization?

At higher levels of processing, **perception and cognition** are closely interrelated, which is the reason why the words ‘understanding’ and ‘seeing’ are synonymous.

However, the visual system has its own rules. We can easily see patterns presented in certain ways, but if they are presented in other ways, they become invisible…**The more general point is that when data is presented in certain ways, the patterns can be readily perceived.**

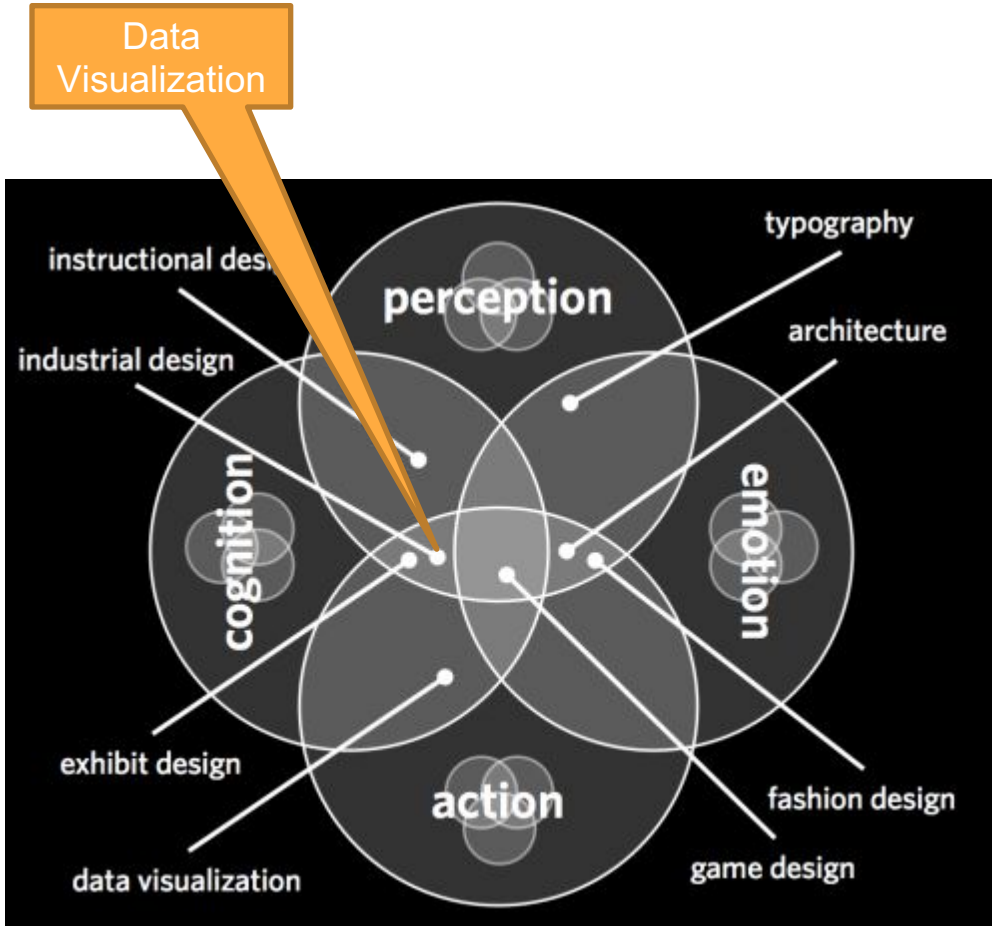
If we can understand how perception works, our knowledge can be translated into rules for displaying information.

Following perception based rules, we can present our data in such a way that the important and informative patterns stand out. If we disobey the rules, our data will be incomprehensible or misleading.

(*Information Visualization, Second Edition,* Colin Ware, Morgan Kaufmann Publishers, 2004, page xxi)

Collin Ware argues that perception and cognition are related, and Tableau.com suggests that visualization leads to actionable insights. Hence, it could be argued that: data visualization sits at the intersection of **perception, cognition and action**.

The image was taken from pinimg.com; however based on argument above, data visualization could be rightfully characterized at the intersection of perception, cognition and action as shown by the orange arrow.



https://s-media-cache-ak0.pinimg.com/564x/eb/e0/f8/ebe0f835e17e18d75571350dc3835771.jpg

Eyes Surpass Even Memory and Other Assurances!

Remember Simpson's Ride at Universal Studios where we know, and our memory keeps reminding us that we are sitting in a chair, and the visuals on the screen are just an act of fiction; however the visual effect surpasses the assurance provided by the memory and the physical chair, and many of us start believing that we are actually falling.

Visualization Key ideas

(inspired by Visualization Analysis & Design by Munzner 2014)

1. Human factor
2. Computer factor
3. External Representation
4. Depend on Vision (as opposed to Sound!)
5. Show data in detail
6. Use interactivity
7. Visualization Idiom design space
8. Focus on tasks
9. Focus on Effectiveness
10. Most designs are ineffective (only some are effective!)
11. Resource limitations
12. Analyze Visualization

Human Factor

Visualization systems are appropriate for use when your goal is to augment human capabilities, rather than completely replace the human in the loop.

Computers can proces raw tables, they do not suffer from perception issues, but they also cannot recognize **interesting** trends either!

Computer Factor

Democratization of visualization

- Excel
- R / ggplot
- Tableau
- Microsoft Power BI (to name a few)

Prevalence of computer tools have “revolutionized” the world of visualization, by giving the power in the hand of “masses”

Big Data (too big to be done manually!)

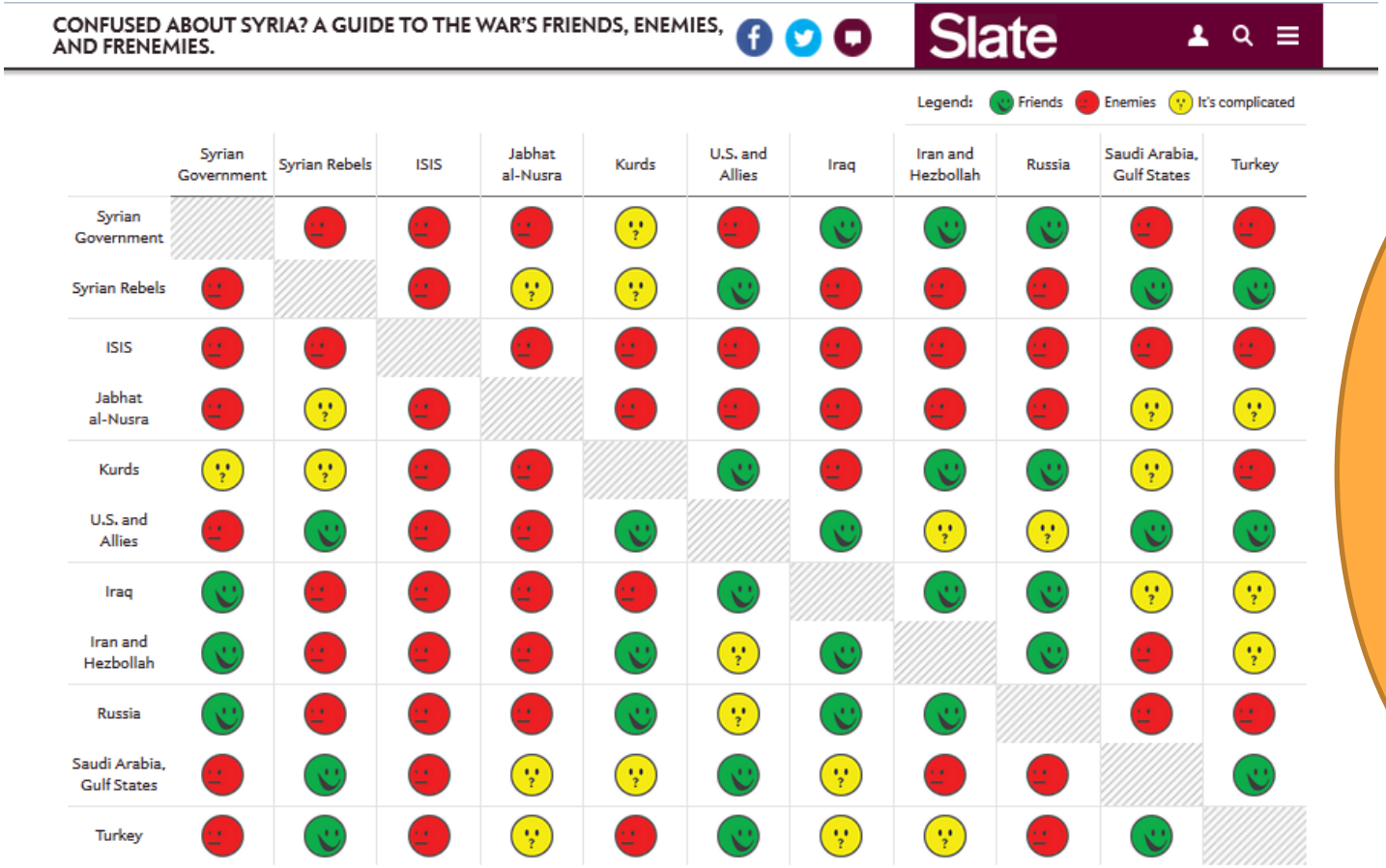
Prevalence of computers led to “big data." Big data, on the other hand, needs computers for “visualization!"

External Representation

External representations augment human capacity by allowing us to surpass the limitations of our own internal cognition and memory.

Visualization as External Memory (offloads internal cognition / aids perceptual inferences): Our memory is of two types, long term and short term. Short term memory easily gets clogged up, hence optimal diagrams can ease up the cognitive load by freeing up the short term memory!

Optimal diagrams can group relevant information together, or place them at same location, thus enabling easy visual contrast and comparison (surpassing the need to remember information, and easing the cognitive load).



An example of examining multiple data points at one place.

Information overload is seriously curtailed by this "external representation"

http://www.slate.com/blogs/the_slatest/2015/10/06/syrian_conflict_relationships_explained.html Dec 23, 2016

Vision

- Very high bandwidth channel to our brain (about 10 Mbps, the same as high-speed Internet)
- Preattentive (Parallel processing)
- Attentive Processing (Sequential processing)
- Visuals can be parallel / sound is sequential

There are two cognition systems: System 1 and System 2 (you can also term them as preattentive and attentive; or parallel and sequential)

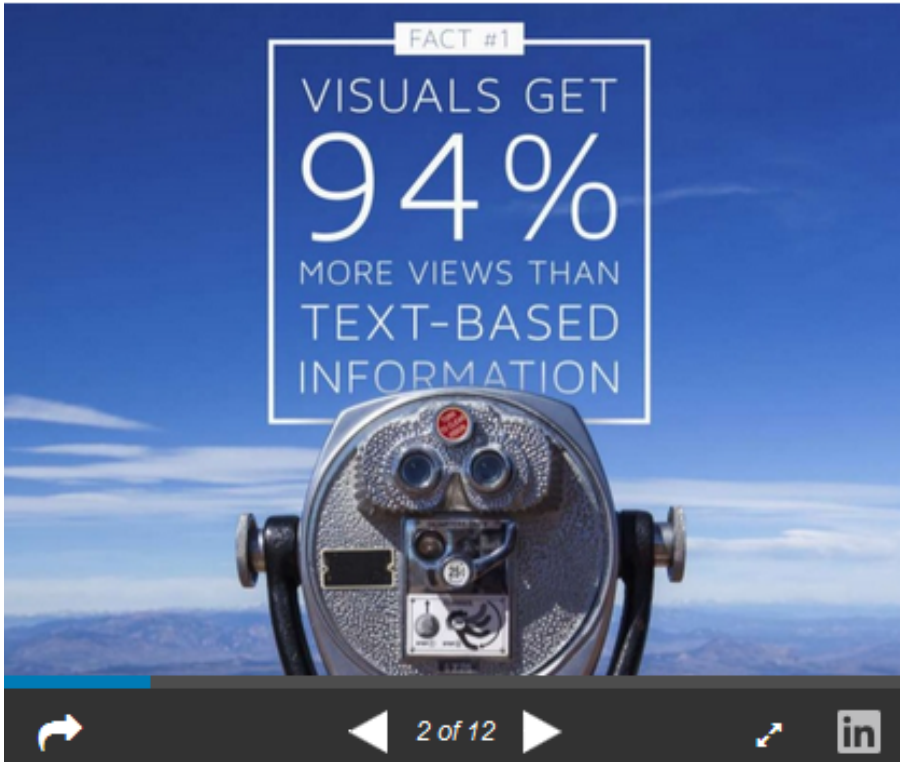
To understand that the person shown in figure below is angry we do not need to process any information. We know it instantaneously. This subconscious system is called System 1 (or preattentive, or parallel processing).

For multiplying 17x24 we need to apply conscious effort. It is not instantaneous. This system is called System 2 (or attentive or sequential processing).

17 x 24 = ?

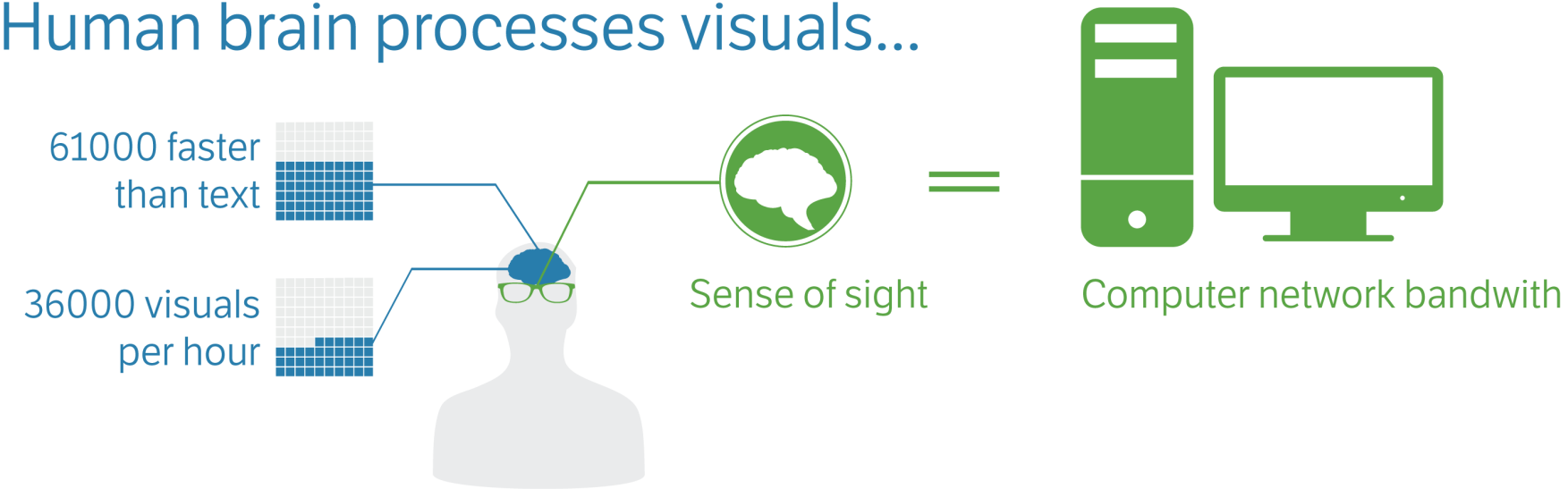
http://www.creativebloq.com/how-design-better-data-visualisations-8134175

Effective visualization can help understand the data by invoking our **System 1** Memory (or **preattentive** **or** **parallel** processing)



https://blog.slideshare.net/2014/11/10/the-power-of-visuals-10-facts-you-need-to-know

Human brain processes visuals...



https://d1wgc1sljrdxgx.cloudfront.net/blog/visual.png

Show Data in Detail

- Anscombe's Quartet
- A simple summary is often an oversimplification that hides the true structure of the dataset

Anscombe's Quartet

4 different sets of data with identical correlation coefficients (.81).
What does this say about each data set? Not much.

data set 1		data set 2		data set 3		data set 4	
X1	Y1	X2	Y2	X3	Y3	X4	Y4
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.1	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.1	4	5.39	19	12.5
12	10.84	12	9.13	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89
	0.82		0.82		0.82		0.82

Correlation Coefficient

Wikipedia : dataMind.co

All three datasets have same correlation values i.e. .81.

However, that is only part of the story. Turn to the next slide to see the actual differences.

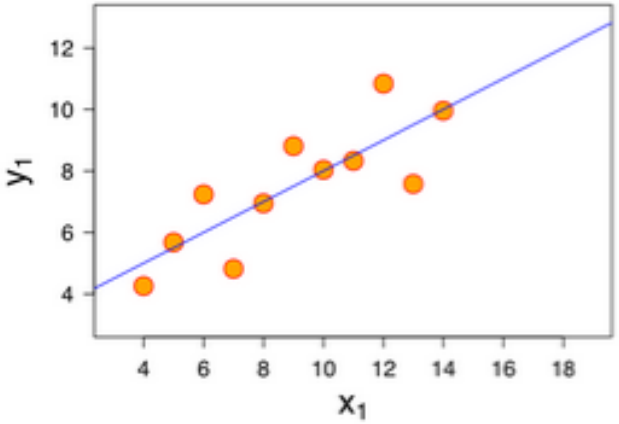
I		II		III		IV		
x	y	x	y	x	y	x	y	
10	8.04	10	9.14	10	7.46	8	6.58	
8	6.95	8	8.14	8	6.77	8	5.76	
13	7.58	13	8.74	13	12.74	8	7.71	
9	8.81	9	8.77	9	7.11	8	8.84	
11	8.33	11	9.26	11	7.81	8	8.47	
14	9.96	14	8.1	14	8.84	8	7.04	
6	7.24	6	6.13	6	6.08	8	5.25	
4	4.26	4	3.1	4	5.39	19	12.5	
12	10.84	12	9.13	12	8.15	8	5.56	
7	4.82	7	7.26	7	6.42	8	7.91	
5	5.68	5	4.74	5	5.73	8	6.89	
9.00	**7.50**	**9.00**	**7.50**	**9.00**	**7.50**	**9.00**	**7.50**	**<< Mean**
3.32	**2.03**	**3.32**	**2.03**	**3.32**	**2.03**	**3.32**	**2.03**	**<< Standard Deviation**
0.82		**0.82**		**0.82**		**0.82**		**<< Correlation between x and y**

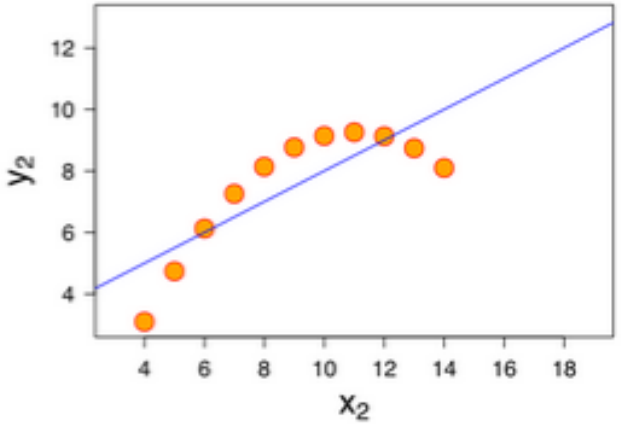
The four datasets have two variables each: X and Y..

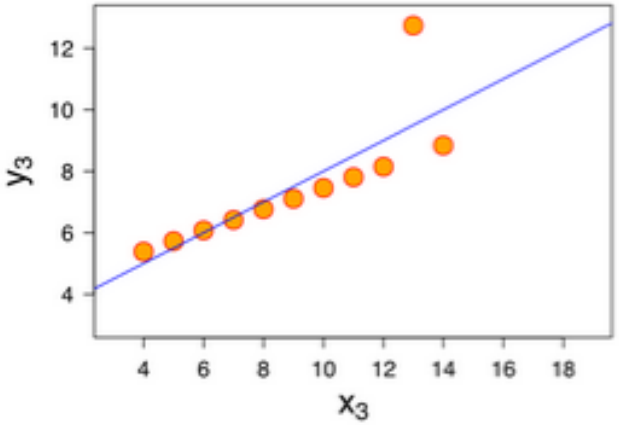
X and Y have same mean and median values across the four datasets.

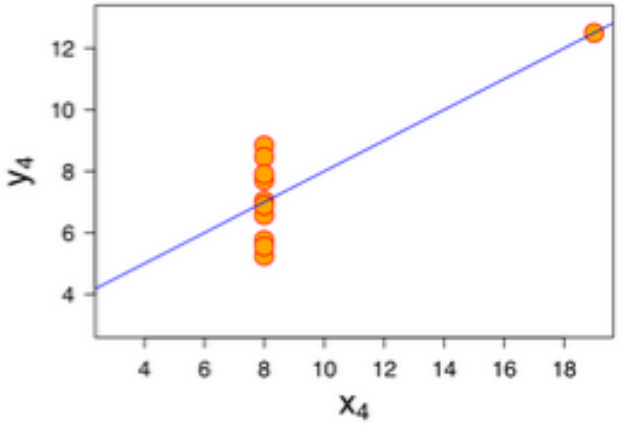
All four datasets have the same correlation values as well.

However, still they are not the same!!









Visualization highlights the differences in the four datasets.

Dataset1 shows a linear relationship.

Dataset 2 shows a curvilinear relationship.

Dataset 3 shows a linear relationship with outlier.

Dataset 4 also shows no relationship with an outlier.

Anscombe's Quartet (set of four datasets, developed in 1973) shows that **visualization** unearths the hidden characteristics which are not discernible by descriptive statistics such as mean, median and correlation alone!

Interactivity

Crucial for visualization tools that handle complexity.

Interactivity supports many possible queries.

- Different ways of representing.
- Different ways of summarizing.
- Multiple level of investigation.
- Ability to comb through big data.

Visualization Idioms

Things to consider:

According to Munzner, visualization idioms are many chart types such as scatter plots, bar charts, line plots etc. These idioms can be static or interactive.

Choice of idiom depends upon following critical questions:

- How to represent information with spatial position.
- How to facet data between multiple views.
- How to reduce the amount of data shown by filtering and aggregation (again useful technique when dealing with big data!).

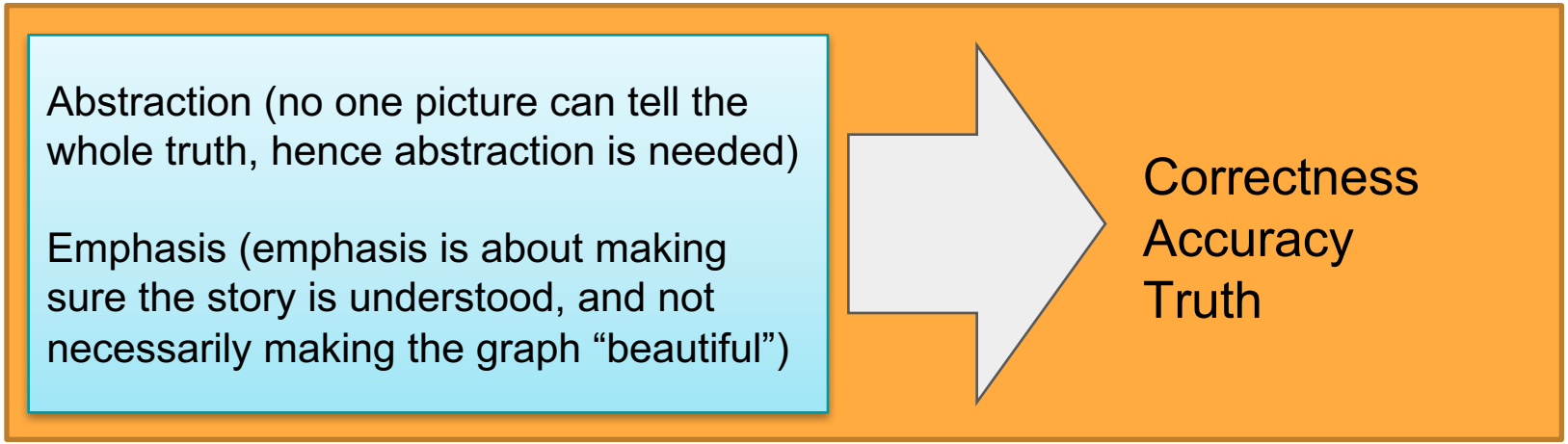
Focus on Task

A tool that serves well for one task can be poorly suited for another for exactly the same dataset.

Focus on Effectiveness

Effectiveness - “It is not just about making pretty pictures”

Choices of abstraction and emphasis must lead to revelation of correctness, accuracy, and truth. Even photographs in real-world scenes involve choices of abstraction and emphasis by including and excluding certain things.



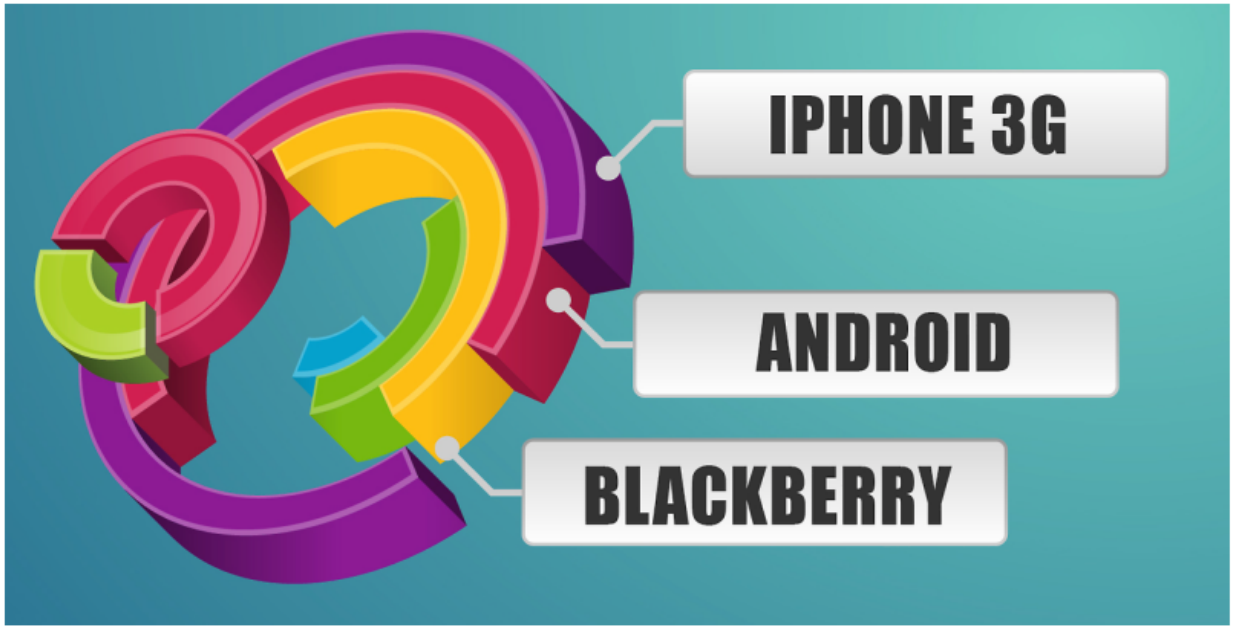
Why are most designs ineffective?

There are several factors which make a design effective in any given context.

The following need to match:

Design + Context + Human Perceptual / Cognition System + Setting + Task

The following visualization is perhaps beautiful, but not effective for sure!



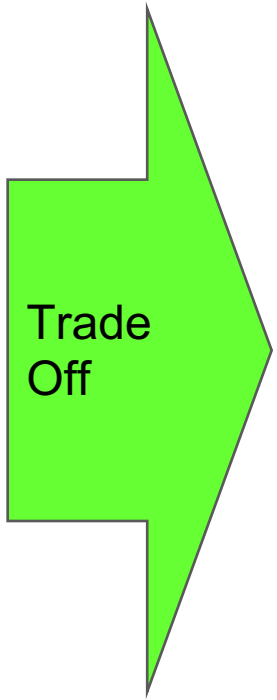
http://gizmodo.com/8-horrible-data-visualizations-that-make-no-sense-1228022038

Why are there Resource Limitations

Computational Capacity (when it comes big data!)

Human Perceptual and Cognitive Capacity (e.g., short term memory, change blindness – many of us are blind to change, see the following slide)

Display Capacity (e.g., limited computer screen resolution)

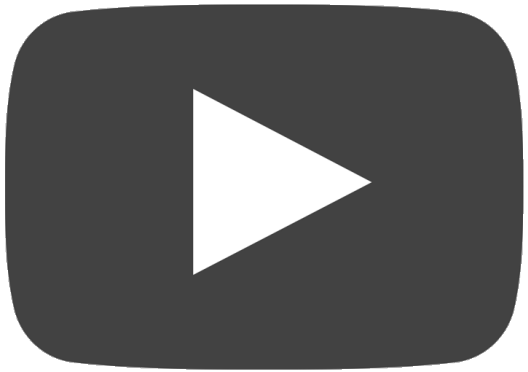


Showing as much as possible at once

vs.

Information Overload and Visual Clutter

Change Blindness Demonstration!



https://www.youtube.com/watch?v=bh_9XFzbWV8

Why Analyze?

There are endless possible combinations of data + tasks + idioms

What? - What **data** the user sees.

Why? - Why the user intends to use a visualization tool (**task**).

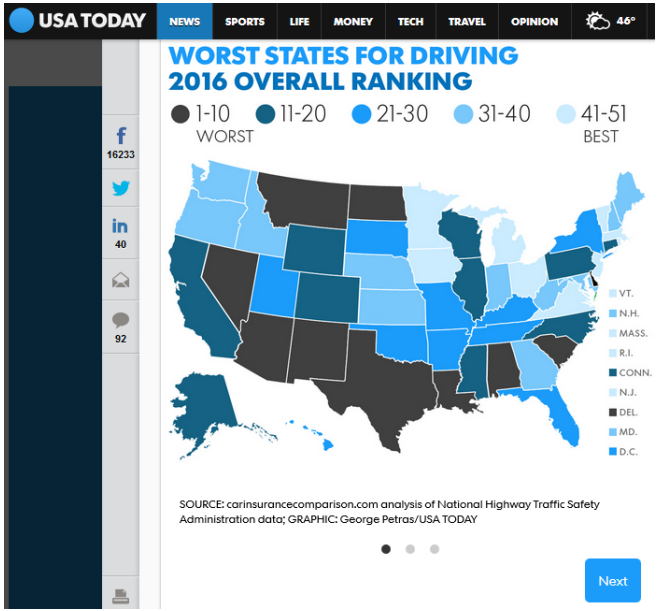
How? - How the visual encoding and interaction **idioms** are constructed in terms of design choices.

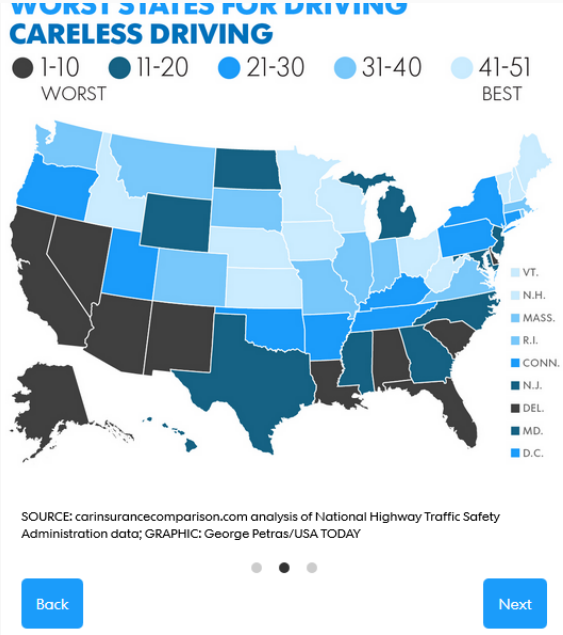
Analyze the **What**, **Why** and **How** to improve the ineffective design(s)

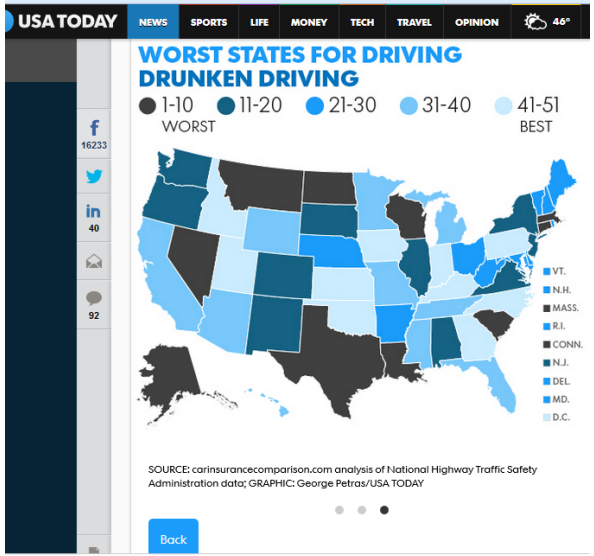
Lets Review An Example!

Hitting the road for the holidays? These states have the worst drivers

Bart Jansen , USA TODAY 11:22 a.m. EST November 23, 2016







Appeared in USA Today on 11/23/2016: http://www.usatoday.com/story/news/2016/11/23/worst-drivers/93470416/

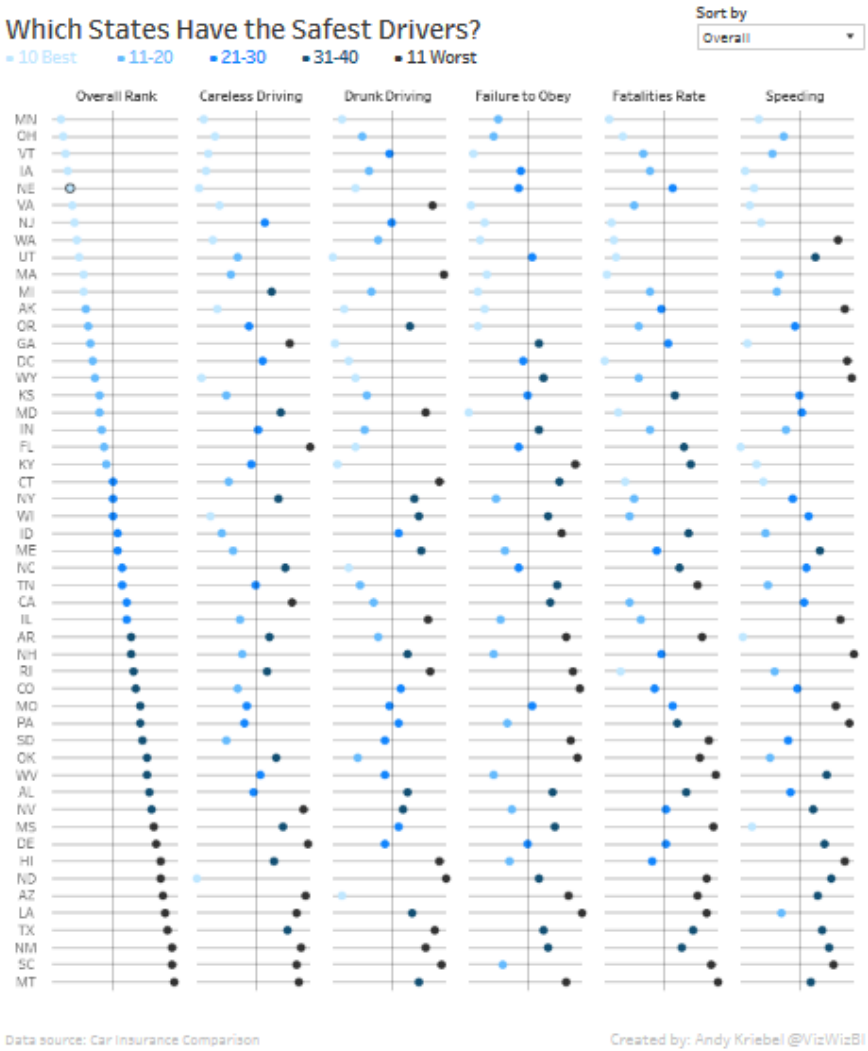
What works well?

- Grouping the States into sets of 10
- Gradual color scaling with black being the most negative
- Calling out the smaller States separately
- Labelling is clear and prominent
- Good color choices
- Simple title

What could be improved?

- A filled map doesn't add much value since there aren't any apparently regional trend. Just because you have States, that doesn't mean you MUST have a map.
- They have three separate maps, which makes it hard to compare them.
- Could have a more impactful title
- For me, it's counterintuitive to have 1 as the worst and 51 as the best.
- It's a very negative message. Wouldn't something a bit more positive be better during the holiday season?

http://www.vizwiz.com/2016/12/makeover-monday-where-are-americas-best.html



The makeover is available at Makeover Monday blog: http://www.vizwiz.com/2016/12/makeover-monday-where-are-americas-best.html

Provides easy comparison and that too at one place without the need to click “back" or "next” across states and across various factors

Interactivity lets us sort using multiple options.

Choice of Idiom: Having state data doesn’t mean that we have to use a map!

Questions? Thoughts?
Visit the course's online discussion forum.