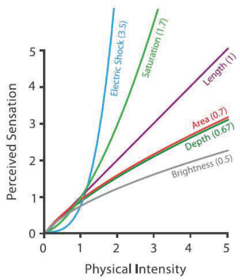
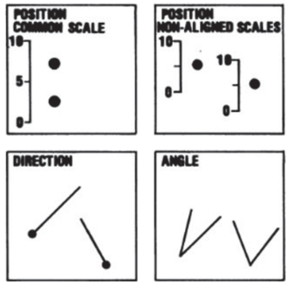
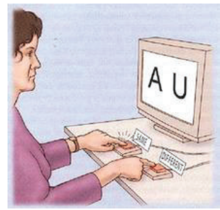






Human Factors Research: How to tell what works

Michael Friendly
Psych 6135

http://euclid.psych.yorku.ca/www/psy6135/

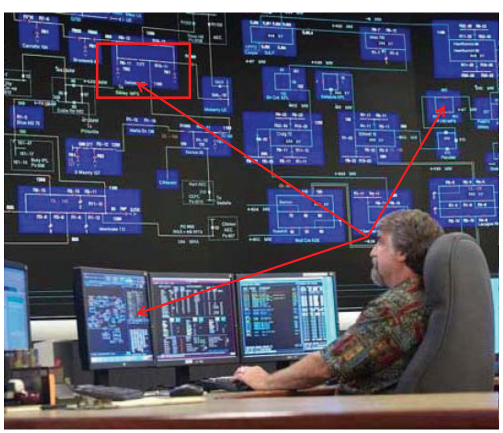
Topics

- Why consider human factors in graphic & information design?
- Empirical study of graphs
- Experimental methods
  - Psychophysical methods
  - Eye tracking
  - Computer, web-based experiments
- Visual inference

Control room display

A control room for a nuclear power plant or electrical system for a large city

- How does visual design support important decisions?
- How to warn or know when something fails?



How many things can the operator attend to at one time?

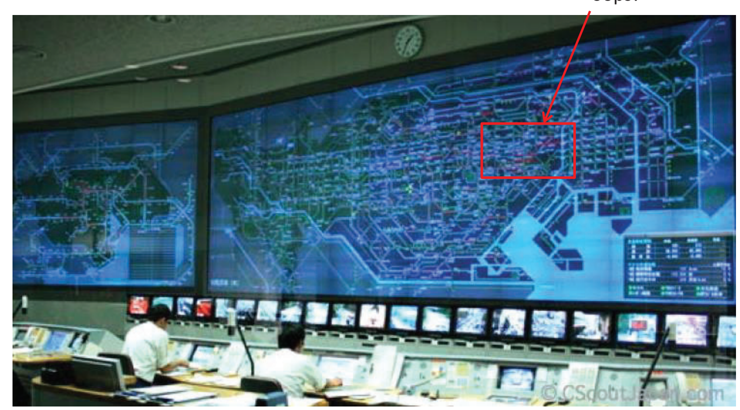
What visual design factors make important events salient?

Traffic control display

A traffic control system for a large metropolitan city (Tokyo)

- How does visual design support important decisions?
- How to warn or know when something fails?
- When is there too much information?

oops!



Navigation

An early digital navigation display panel, incorporating visual gauges and charts for a variety of functions– combines separate variables into a single display

What visual features make it easier or harder to navigate?

Garmin Txi touchscreen device for a small jet

How does the pilot combine the map view with the heads-up view and all the visual dials?

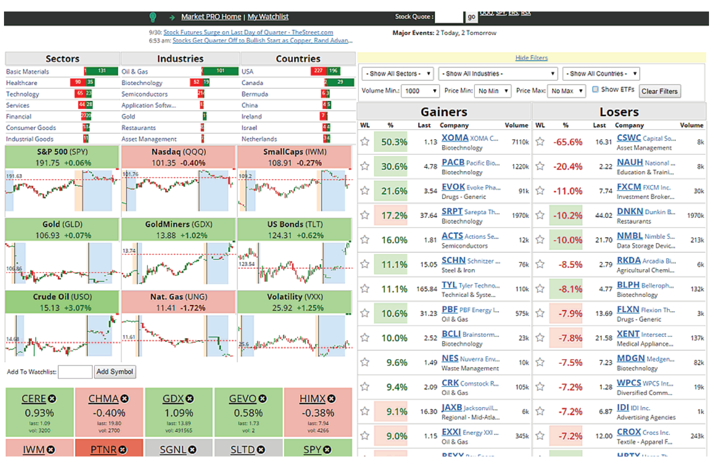
Navigation

A more modern design integrates a wider variety of displays in a touch-screen device
- Are there too many options or features?
- What features demand the most pilot attention or distract attention from flying?
- How to study the efficacy of alternative visual display designs?

Rockwell-Collins Air King control panel

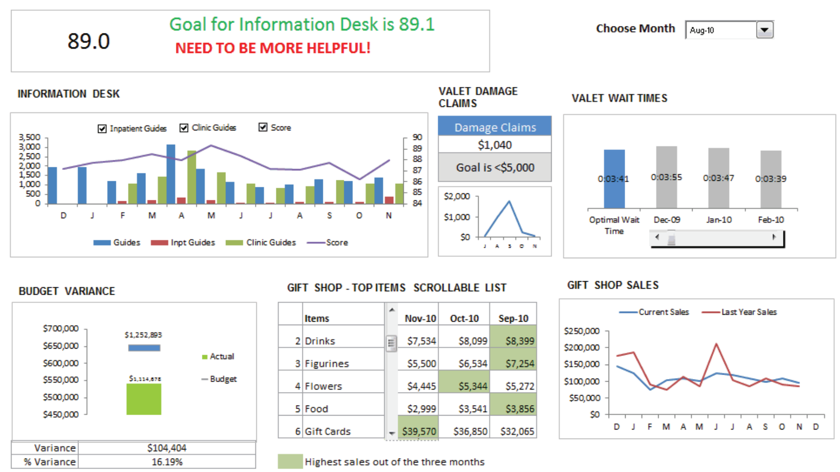
Dashboards: Financial trading

Dashboards combine visual information for decisions on a single screen
Good dashboards are: dynamic, interactive, customizable



Dashboards: Customer service

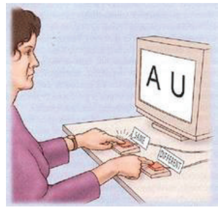
Interactive dashboards use sliders, buttons, pick lists, etc.



From: https://chandoo.org/wp/customer-service-dashboard/

Empirical studies of graphs

- Human factors vs. Psychology
- Experimental methods
  - Psychophysical methods
  - Task analysis
- Running graph perception studies
- Some results



Human factors vs. Psychology

- Human factors research often motivated by applied problems in engineering, design, computer science
  - A/B testing (booking.com, Netflix, …) for features of user interaction
  - navigation controls: pilot testing, flight simulators
- Psychological research often motivated by more basic perceptual & cognitive questions.
  - accuracy of judgments of graphs
  - judgments of trend, correlation, etc.

Psychophysical methods

- Psychophysical methods are used in studies of graphical perception to study the relationship between
  - properties of a stimulus (position, length, area, angle, …)
  - and a perceived response (how big? which is greater?)
- Magnitude estimation: rated (0-100) size or %
- Matching – adjust size of B so it is as large as A
- Discrimination
  - same/different?
  - which is larger?

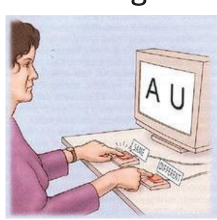
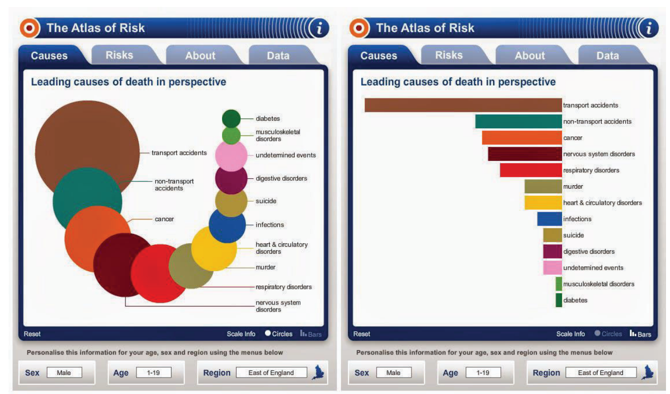


Image:
http://www.cns.nyu.edu/~david/courses/perception/lecturenotes/psychophysics/psychophysics.html

Magnitude estimation

How large are transport accidents?
How much bigger than non-transport accidents?



Estimation of length or ratios of length are more accurate than the same judgments of area.

Pies vs. dot charts

Which sector is the largest?
Which is the smallest?

Judgement of position along a scale is more accurate than that of area or angle

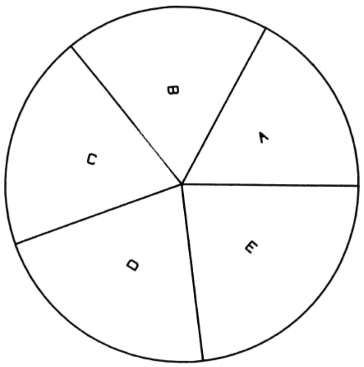


Figure 22. Pie chart.

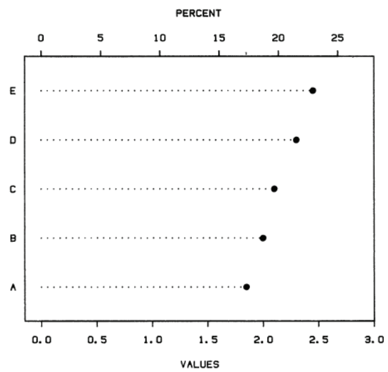


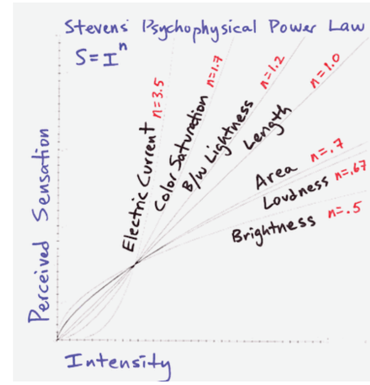
Figure 23. Dot chart.

Stevens' Power Law

- How does perceived magnitude of a sensation relate to stimulus intensity?
- S. S. Stevens (1957) showed that, for many domains

  Sensation $\propto$ Intensity^p

- These provide ways to assess the **accuracy** of magnitude estimation for visual encodings
  - length judgments most accurate
  - area: less so
- But: graph perception is not always a matter of estimating magnitudes.



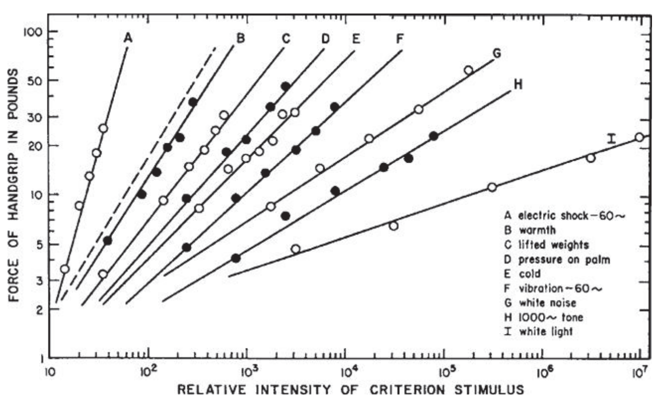
From: https://santhoshsoundar.blog/power-law/

Stevens' Power Law: log-log form

The exponential form, S ~ Intensity^p is more easily understood when both are plotted on log scales, where it is linear

$\log(S) = p \log(I)$

p is the multiplier effect of a multiple of Intensity

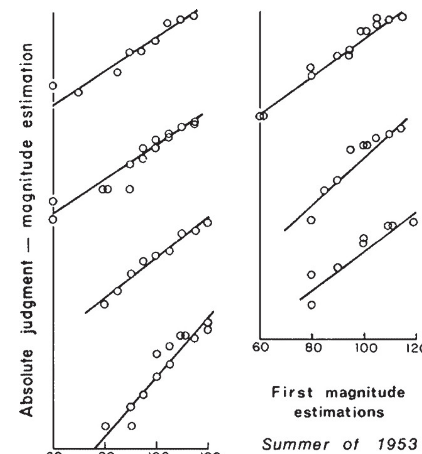


Power Law: Origin story

In 1953, S. S. Stevens carried out experiments on magnitude estimation of loudness of sounds – measured in db (a log scale)

This graph shows results for 7 individual subjects, offset to show the separate data

The idea of an average slope, *p*, arose later, as the effect for an average observer.



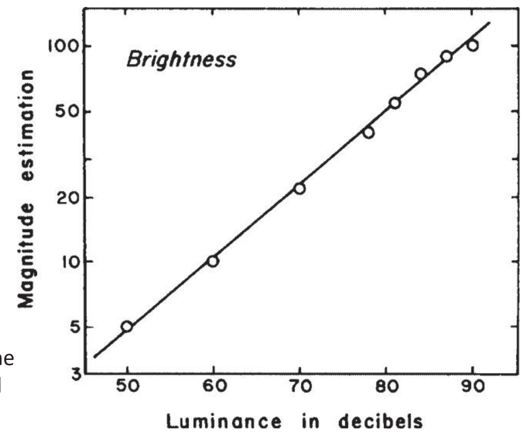
See: (2009) Stevens' Power Law. In: *Sensory Neuroscience: Four Laws of Psychophysics*. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-84849-5_1

Power Law: Brightness

What about other stimulus domains?

Luminance of a light source could be measured (on a log scale)

Averaging over observers, log of the magnitude estimates was again linear with log intensity.



Stevens (1975) asserted that he had discovered a fundamental law of the relation between physical properties and perception.

Magnitude estimation & memory

How does remembered size relate to perceived size?
How does it relate to stimulus intensity?
Kerst & Howard (1984) propose another power law:

Table 1
Power Function Exponents (n) and Correlations (r) for Perceptual and Memorial Estimates of Line Length and Shape Area (Experiment 1)

		Group Data		Median Individual Data	
		Perception	Memory	Perception	Memory
Line Length	n	.93	.84 (.86)	.90	.81 (.81)
	r	.99	.99	.99	.99
Shape Area	n	.77	.61 (.59)	.74	.65 (.55)
	r	.99	.99	.99	.98

Note—Values in parentheses are those predicted by the relation that the memory exponent equals the square of the perceptual exponent.

$$\text{Sensation} \propto \text{Intensity}^{p}$$

$$\text{Memory} \propto \text{Sensation}^{p'}$$

therefore

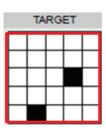
$$\text{Memory} \propto \text{Intensity}^{p \times p'}$$

If p ≈ p′, the exponent for memory ≈ p^2
→ remembered area less accurate than area itself

Kerst & Howard, *Bulletin of the Psychonomic Society* 1984, 22 (6), 517-520

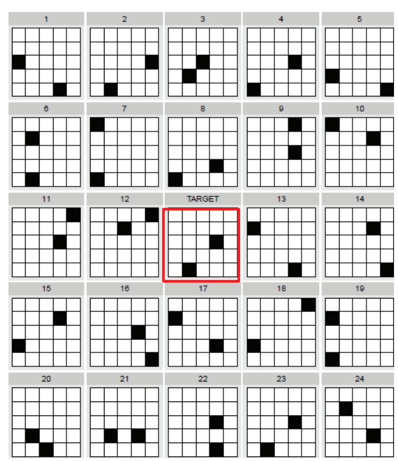
Visual search tasks

Find a target item among the many distractors in this display



Measure:
- reaction time (RT)
- accuracy (% correct)

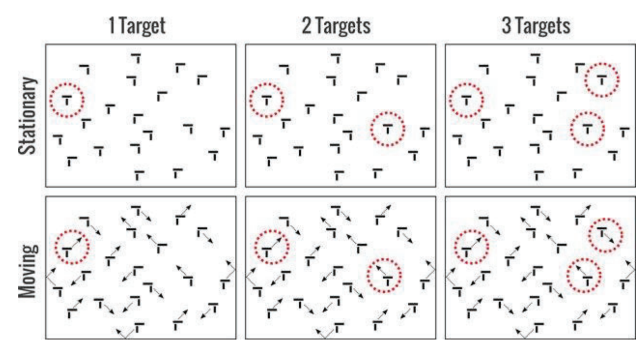
Vary:
- # distractors
- # of targets
- complexity (# black squares)
- display format
- …



This paradigm is often used in evaluating complex visual displays

Visual search experiment

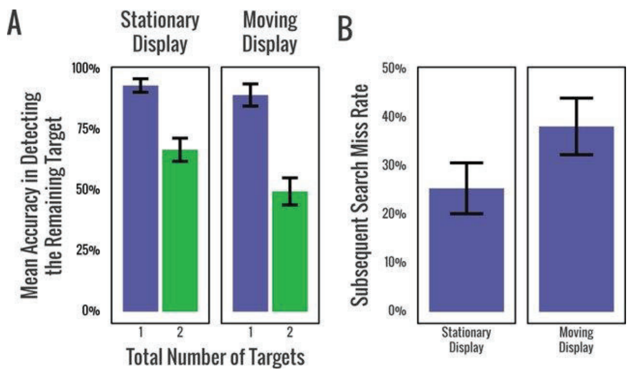
Visual search for a diagnostic signal (radiology) can vary with the # of targets and whether these are shown stationary or moving.



The response measure here is % accuracy in detecting a given target.

Visual search experiment

Results: Better for 1 target vs. 2, whether moving or not
A 2nd search is less accurate when the target is moving



From: Stothart et al (2018). Satisfaction in motion: Subsequent search misses are more likely in moving search displays. *Psychonomic Bulletin & Review* 25(1):409-415

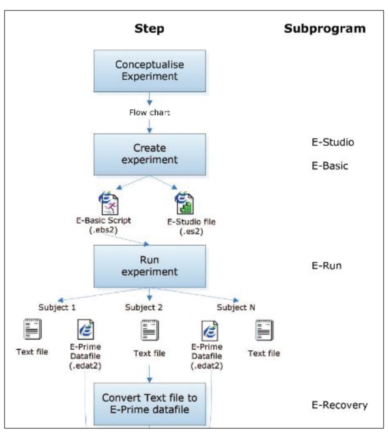
Running graph perception experiments

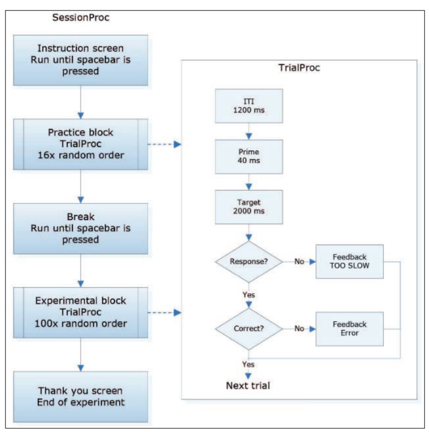
- Paper & pencil tasks
  - little control of experimental presentation features
  - can't measure RT
- Lab software: run on lab computers
  - e-prime ($$$), https://pstnet.com/
  - matlab ($) & Psychtoolbox: http://psychtoolbox.org/
  - PsychoPy – free, open source, see: http://www.psychopy.org/
- Web-based
  - Survey Monkey / Qualtrix
  - Amazon Mechanical Turk

e-prime

e prime is a polished software system for designing psychology experiments

- E-Studio GUI → E-Basic script; E-Run: runs experiments

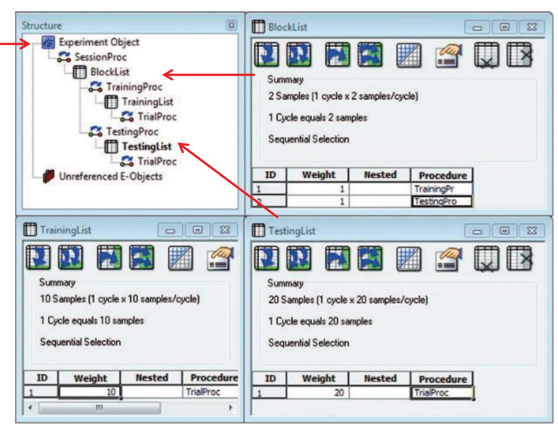




e-prime

Create experiment structure by drag-and-drop in E-Studio
Stimuli: text, audio, images, video, …
Response devices: keyboard, mouse, external devices, …

Experiment structure

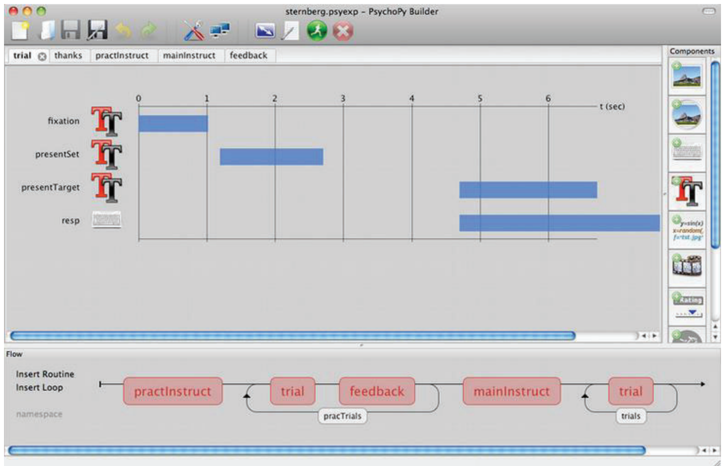


Block of trials

PsychoPy

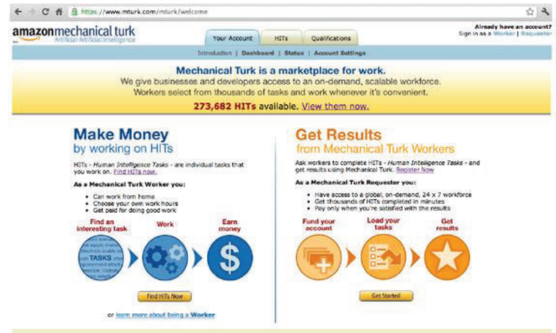
PsychoPy provides a GUI for constructing online experiments https://www.psychopy.org/

- Builder interface → python code that runs the experiment



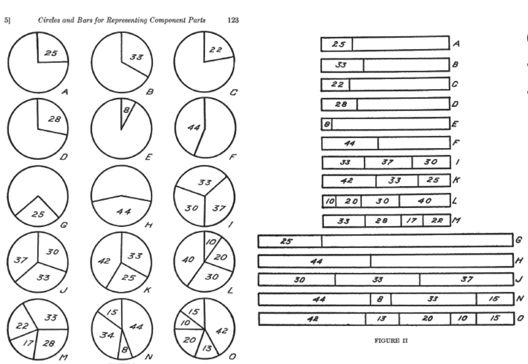
Amazon Mechanical Turk

- Web-based experiments, hosted on Amazon servers
  - requester jobs: Human Intelligence Tasks (HITs)
  - worker pool: Turkers, get paid for doing tasks ($0.01 – 0.10 per item)
  - each cell of a design can be a separate HIT
  - Amazon provides a markup language for presenting text, movies, images, … (HTML, javascript)



Early studies: Circles vs. bars

Eells (1929) studied magnitude estimation of proportions of a whole, presented as circles vs. bars
"What number represents the proportion for each marked segment?"

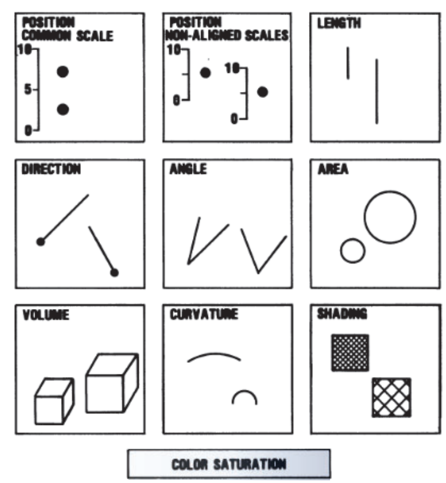


Conclusions:
- 1 segment: circles ≈ bars
- >1 segment: circles better than bars

Graph perception: Elementary perceptual tasks

Cleveland & McGill (1984) proposed that graphical perception could be studied in terms of 10 elementary perceptual tasks involved in most common graphs.

Their study was one of the first modern ones, and set a standard for magnitude estimation tasks.

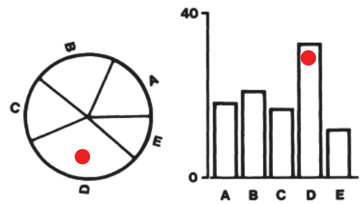


Cleveland & McGill experiments

Pie charts vs. bar charts – position-angle experiment

Elementary perceptual tasks:
- bar chart: position along a scale
- pie chart: angle? area? arc length?

Experiment:
50 graphs (½ pie, ½ bar), random order
largest marked ●
"What percent is each of the others?"
"Make a quick visual judgment"
Response on an answer sheet

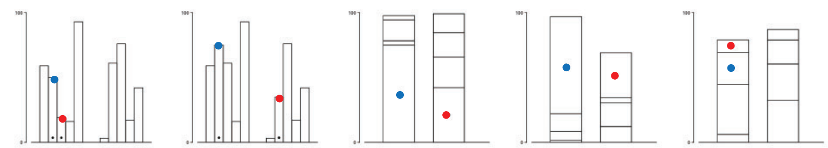


Measures:
accuracy: $\log_2(|\text{judged \%} - \text{true \%}| + 1/8|)$
bias: judged % - true %

NB: log scale estimates relative error; +1/8 handles zero values

Cleveland & McGill experiments

Bar charts tasks – position-length experiment



(a) Adjacent Bars (Type 1) (b) Separated Bars (Type 3) (c) Aligned Stacked Bars (Type 2) (d) Unaligned Stacked Bars (Type 4) (e) Divided Bar (Type 5)

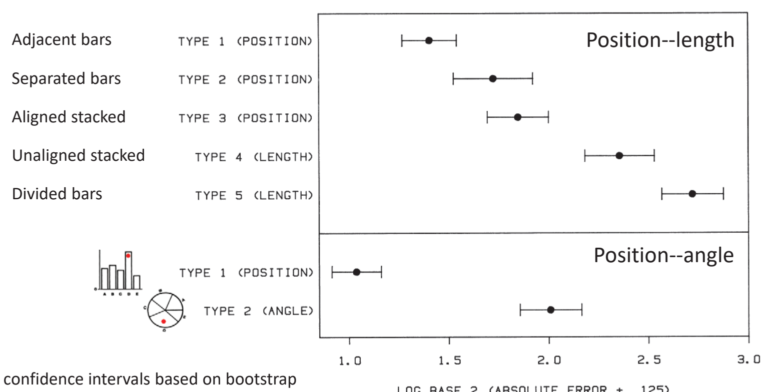
Experiment:
- 50 graphs, 10 × 5 types
- "What percent is smaller ● of the larger ● ?"

Measures:
accuracy: $\log_2(|\text{judged \%} - \text{true \%}| + 1/8|)$
bias: judged % - true %

Discussion Q:
- What are the elementary perceptual tasks involved in each of these?
- What is the statistical analysis?

Cleveland & McGill experiments

Cleveland & McGill summarized these experiments in this figure, comparing absolute error in the tasks in these two experiments.



NB: confidence intervals based on bootstrap of log |error| because of (a) non-normality; (b) repeated measures

Heer & Bostock: MTurk experiments

- Replicated Cleveland & McGill T1—T5, T6 (angle)
- Added area judgment tasks:
  - T7: Bubble chart
  - T8: Center-aligned rectangles
  - T9: Treemap

Task:
"What % is area A of area B?"

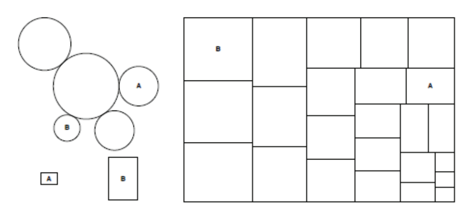


Figure 2: Area judgment stimuli. Top left: Bubble chart (T7), Bottom left: Center-aligned rectangles (T8), Right: Treemap (T9).

Heer & Bostock: results

Results largely confirmed Cleveland & McGill (1984) with respect to relative order
The area judgment tasks are shown to give even larger errors

Details:
H & B use a between-S design, n=50 per chart type, 10 charts of each type

10 × 7 = 70 separate HITs (each S responds to 1 chart)

Response: type in a # (% of smaller)

This graph of these results is a great model for data display

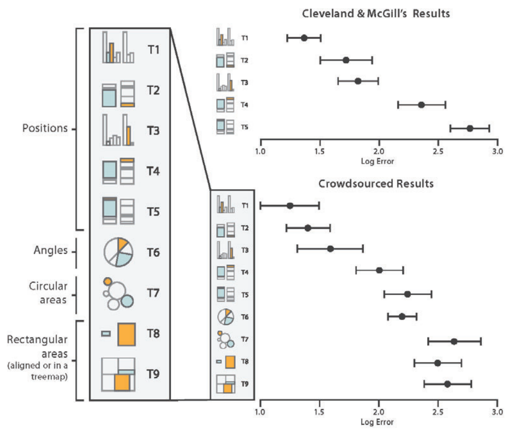


Fig. 5.8 from: Munzner, *Visualization Analysis & Design*

Heer & Bostock: results

Other findings:
For a given graph type, judgments are most accurate when the true difference is extreme

This within-graph effect is larger than the differences between graphs

Discussion Q:
What are some problems with this graph?
How could it be improved?

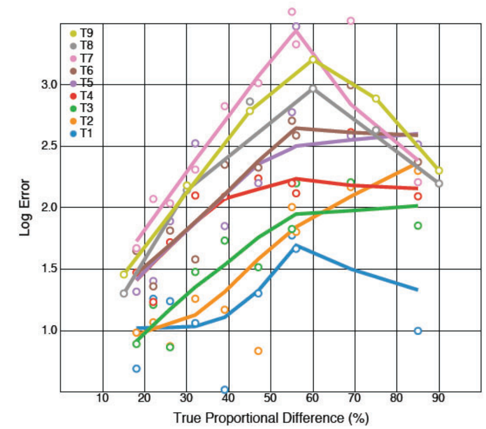
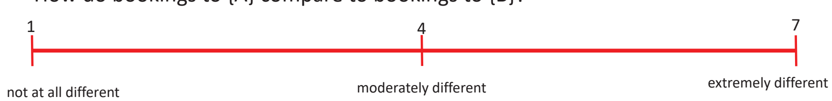


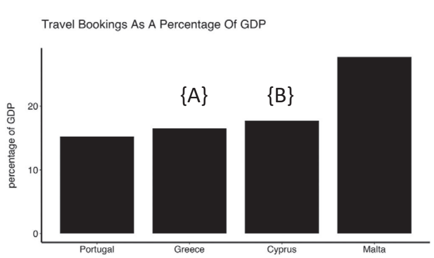
Figure 3: Midmeans of log absolute errors against true percentages for each proportional judgment type; superimposed are curves computed with *lowess*.

Effect of truncation in bar charts

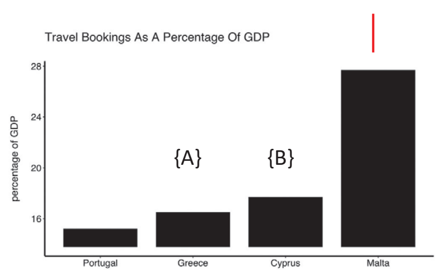
An Mturk experiment to assess the effect of Y-axis truncation on relative judgements

How do bookings to {A} compare to bookings to {B}?





Not truncated



Truncated

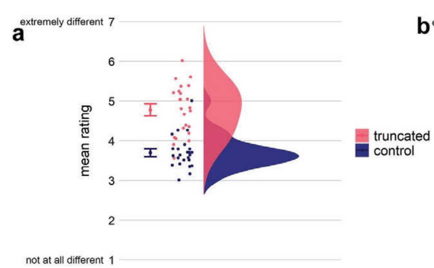
Brenda Yang et al., Truncating Bar Graphs Persistently Misleads Viewers, *Journal of Applied Research in Memory and Cognition*, 2021, https://doi.org/10.1016/j.jarmac.2020.10.002

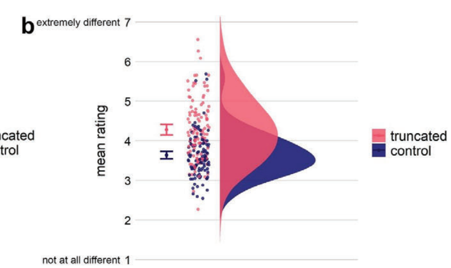
Effect of truncation in bar charts

Design: Within-S, each subject saw 20 control bar graphs and 20 truncated bar graphs

Study 1: No warning (n=24)

Study 2: Subjects warned that some graphs might be misleading (n=109)





This graph form ("raincloud plot") combines density estimates, data dots & Cis
Study 2 tests an interpretation based on task mental set / instructions

Simkin & Hastie: accuracy *and* RT

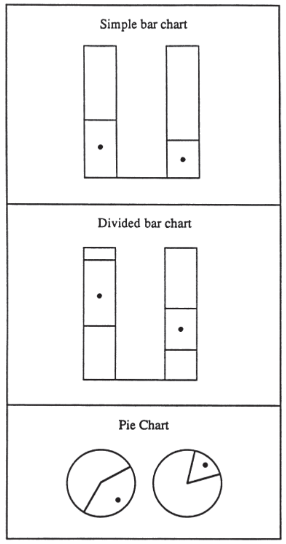
Problems with Cleveland & McGill study:

- Assessing accuracy only omits consideration of speed of judgment – should also measure reaction time
- "Elementary perceptual tasks" give no insight into the cognitive processes used by observers to perform these judgments.

Simkin & Hastie used computer-controlled experiments to measure both accuracy & RT

- Three types of stimuli x 30 of each = 90 trials
- Discrimination task: Which is larger?
- Judgment task: What % is smaller of the larger?

Analyses: Separate ANOVAs of discrimination RT, judgment RT and errors in each task



Simkin & Hastie (1987), An Information-Processing Analysis of Graph Perception, *JASA*, **82**, 454-465

Simkin & Hastie: processing stages

They propose that tasks using various graph types can be understood in terms of elementary mental processes:

anchoring: segment a component to serve as a standard for comparison
scanning: visual sweep across a distance in a graph
projection: send a ray from one point to another
superimposition: mentally move elements to a new, overlapping location
detection: detect difference in size of two components

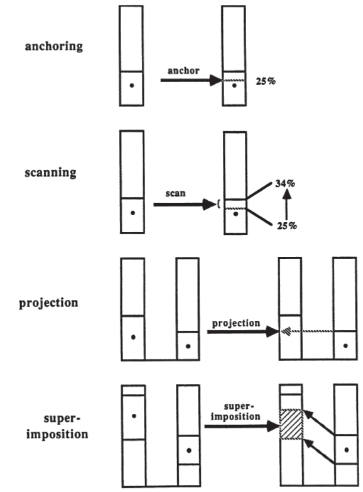


Figure 2. Schematic Summaries of Proposed Elementary Mental Processes That Can Be Combined to Explain Performance in the Experimental Tasks.

Simkin & Hastie: processing stages

Analysis of the three graph types in terms of proposed elementary mental processes

bar chart (position)

divided bar (length)

pie chart (angle)

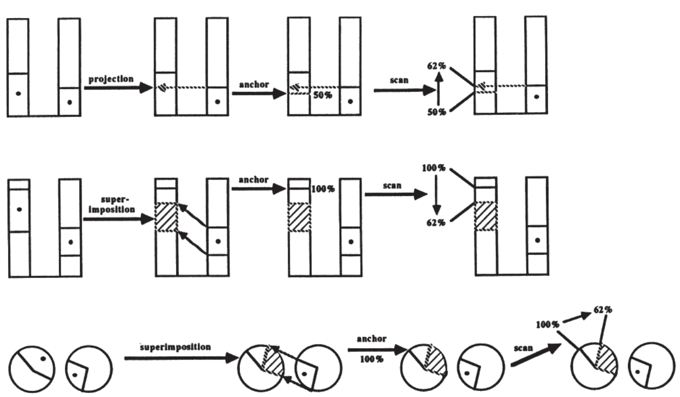


Figure 7. Proposed Sequence of Elementary Mental Processes to Explain Performance in the Comparison Judgment Task for Position (top panel), Length (middle panel), and Angle (bottom panel).

NB: If these processes are sequential, RTs should reflect additive components
AFAIK, this idea has not been tested or explored.

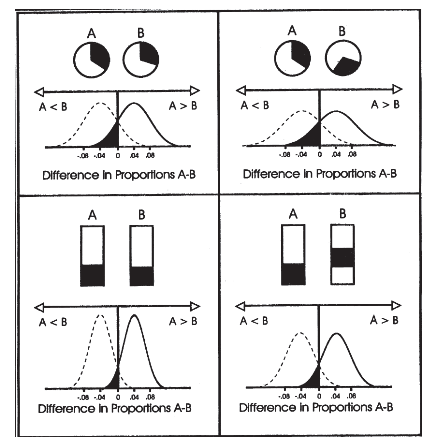
Hollands & Spence: Discrimination analysis

How do people make decisions about which is larger from different visualizations?

- aligned vs. not aligned
- pie vs. bar chart

Hollands & Spence propose an incremental estimation model to account for speed of processing:

- each stimulus evokes a distribution of a psychological response of "size"
- the response (A>B?) is determined by the separation and overlap between the two distributions
- less overlap → faster response

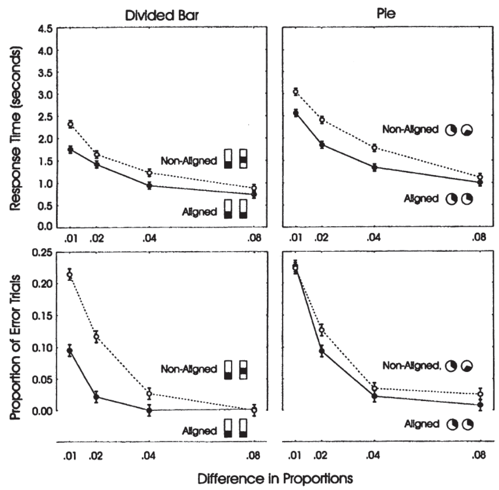


Hollands, J. & Spence, I. (2001). The discrimination of graphical elements. *Applied Cognitive Psychology*, **15**, 413-431.

Hollands & Spence: results

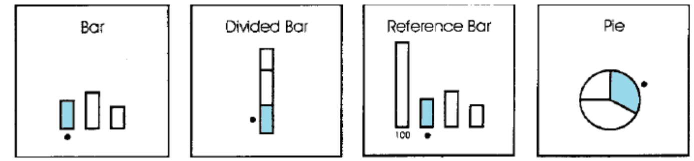
Experiment 1
- RT ↓ as Δp ↑
- RT < for aligned vs. non-aligned
- RT < for divided bars vs. pies

Similar pattern for errors

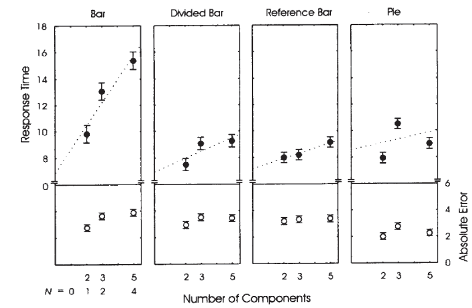


Are pies ever better?

What percent is the indicated region of the total?



Other experiments by Holland & Spence (1992, 1998) show that this judgment task is hardest for separated bars and easiest for pie charts



Hollands, J. & Spence, I. (1998). Judging Proportion with Graphs: The Summation Model. *Applied Cognitive Psychology*, **12**, 173-190

More pie studies: Skau & Kosera

Infographics use many variations of the basic pie chart to show part-whole relations
What properties do people use in making judgments: Angle, area, arc length?

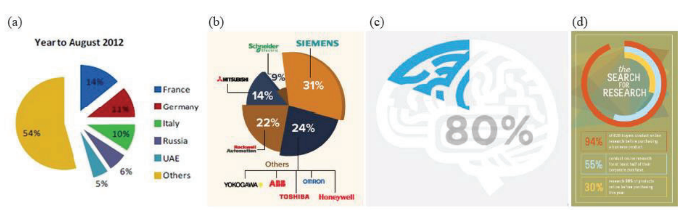
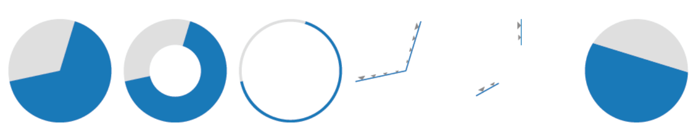


Figure 3: A sampling of pie and donut charts used in infographics, taken from examples found on Visual.ly [Vis15]. (a) exploded pie chart, (b) chart with varying segment radii, (c) pie chart constructed with an icon, and (d) nested donut chart.

Variations of these designed to test accuracy of part-whole judgments

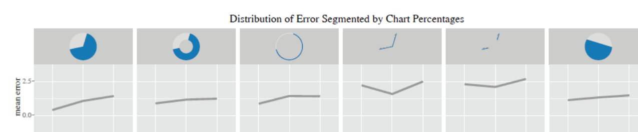


(a) Pie chart. (b) Donut chart. (c) Arc length chart. (d) Angle pie chart. (e) Angle donut chart. (f) Area chart.

Skau & Kosara (2016), Arcs, Angles, or Areas: Individual Data Encodings in Pie and Donut Charts, EuroVis.

More pie studies: Skau & Kosera

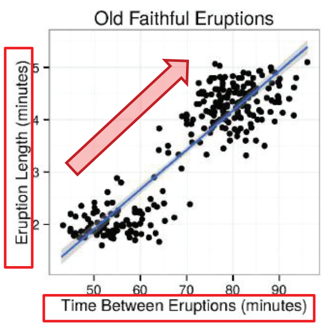
Little difference among these, except for plots showing only angle



Task analysis

Another cognitive approach is to study the visual/mental steps a viewer takes in trying to answer a question based on a graph. Sometimes uses "protocol analysis"

Q: What is the relationship between the $\boxed{\text{length of the eruption}}$ and the $\boxed{\text{time between}}$ eruptions for Old Faithful?



Mental steps:

1. Understand the Q: identify "length of eruption" & "time between eruptions" as things to search for in the graph.
2. Look at axis labels: See Y: "Eruption length"; X: "Time between Eruptions
3. Scan data: See "Y increases as X increases"
4. Answer the Q: As the time between eruptions increases, the length of the eruption seems to increase.
5. Notice: Hmm, something weird here!

Illustration from VanderPlas (2015), *Perception in statistical graphics*, PhD thesis, Iowa State U

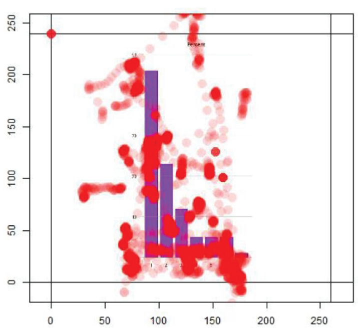
Eye-tracking studies

Where do people look when viewing graphic displays?

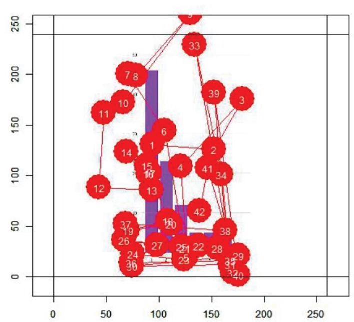
Eye-tracking hardware allows recording of gaze fixation points over time
Eye-tracking software allows some visual analysis

Viewing a bar chart

This 2016 study by Trent Fawcett was a basic test of this methodology.
Uses R and the saccades package. There is also: gazepath, emov and more…



Gaze fixation points with a scatterplot and transparency to show density



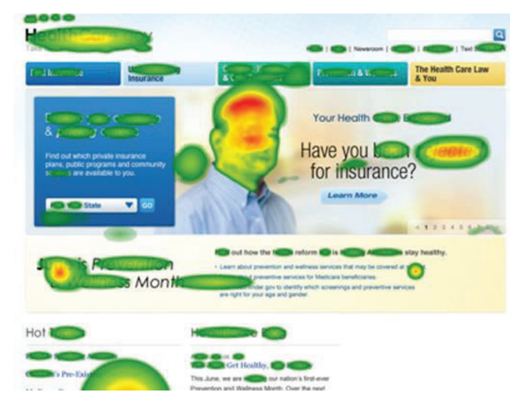
Gaze fixation points showing the temporal sequence of saccades

Fawcett, T. "The Eyes Have It: Eye Tracking Data Visualizations of Viewing Patterns of Statistical Graphics" (2016). https://digitalcommons.usu.edu/gradreports/787

Viewing web pages

This methodology is now well developed, particularly for viewing web pages

This illustration uses heatmap colors to show density of gaze locations



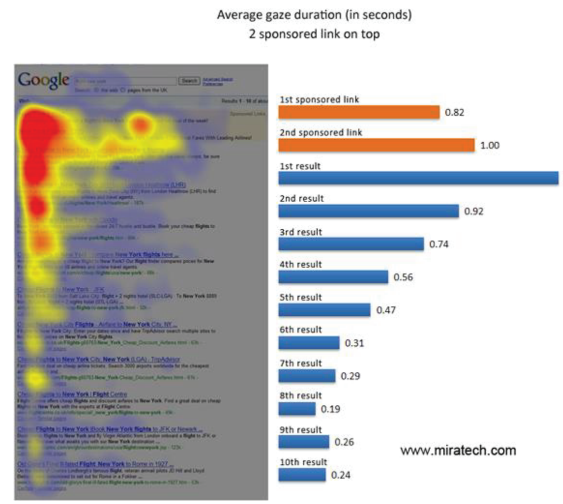
Viewing web pages

This illustration shows a path of eye-movement locations in viewing the same web page

Viewing web pages

This method is widely used to evaluate effectiveness & $ for ads on web pages

Online vendors do numerous such studies

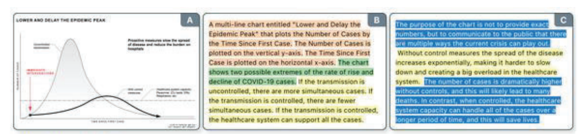


Accessibility of data visualization

- Graph design
  - Design for color deficiencies
  - Favor direct labels over legends
- Assistive technology
  - Data sonification: map data to sounds
  - Screen readers: turn text into speech
- Text, titles, captions
  - Web: use alt-text <img alt="What to understand here">
  - PPT: embed descriptive text in images
  - Papers: use better titles and captions

Accessibility of graph captions

Lundgard & Satyanarayan (2022) studied how graph captions affect understanding, particularly for disabled (blind) viewers



Long Description

Visualizations like "Flatten the Curve" (A) efficiently communicate critical public health information, while simultaneously excluding people with disabilities [11, 28]. To promote accessible visualization via natural language descriptions (B, C), we introduce a four-level model of semantic content. Our model categorizes and color codes sentences according to the semantic content they convey.

How to design figure captions / alt-text to better communicate what is to be shown?

Accessible Visualization via Natural Language Descriptions, http://vis.csail.mit.edu/pubs/vis-text-model

Accessibility of graph captions

They present a model of 4 levels of description and their semantic content

#	LEVEL KEYWORDS	SEMANTIC CONTENT	COMPUTATIONAL CONSIDERATIONS
4	contextual and domain-specific	*domain-specific insights, current events, social and political context, explanations*	contextual knowledge and domain-specific expertise (*perceiver-dependent*)
3	perceptual and cognitive	*complex trends, pattern synthesis, exceptions, commonplace concepts*	reference to the rendered visualization and "common knowledge" (*perceiver-dependent*)
2	statistical and relational	*descriptive statistics, extrema, outliers, correlations, point-wise comparisons*	access to the visualization specification or backing dataset (*perceiver-independent*)
1	elemental and encoded	*chart type, encoding channels, title, axis ranges, labels, colors*	access to the visualization specification or rasterized image (*perceiver-independent*)

Mortality rate is plotted on the vertical y-axis from 0 to 15%. Age is plotted on the horizontal x-axis in bins: 10-19, 20-29, 30-39, 40-49, 50-59, 60-69, 70-79, 80+.

For low income countries, the average life expectancy is 60 years for men and 65 years for women. For high income countries, the average life expectancy is 77 years for men and 82 years for women.

The low income countries are more scattered than the high income countries. There is a visible gap between high and low income countries, indicated by the Income-Age Divide line.

People living in low-income countries tend to have a lower life expectancy than the people living in high-income countries, likely due to many societal factors, including access to healthcare, food, other resources, and overall quality of life.

Study design

- Stimuli:
  - chart types (bar, line, scatter)
  - topics (academic, business, journalism)
  - difficulty (easy, medium, hard)
- Subjects: 90 sighted, & 30 blind (proficient with a screen reader).
- Task: rank the usefulness of 4 descriptions (Levels 1-4) for understanding

"Suppose that you are reading an academic paper about how life expectancy differs for people of different genders from countries with different levels of income. You encounter the following visualization. [Table 3.C] *Which content do you think would be most useful to include in a textual description of this visualization?"*

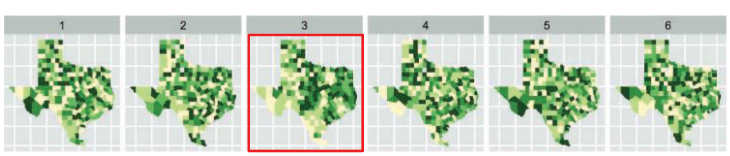
Read the paper to see the scope and content of this type of research

Visual inference

- To what extent can visual display of real data, against a background of random data, be used as a substitute for standard statistical inference?

One of these plots doesn't belong. Which one??

Six choropleth maps of cancer deaths in Texas, darker colors = more deaths.
Can you spot which of the six plots is made from a real dataset and not simulated under the null hypothesis of spatial independence?



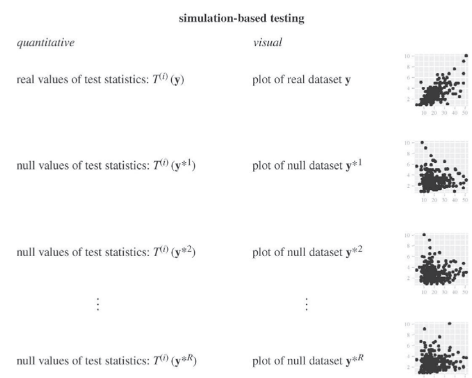
From: Wickham etal. (2010) Graphical Inference for Infovis, *IEEE Trans. Visualization & Computer Graphics*, Vol. 16, No. 6, 972-979.

Visual inference: Lineup protocol

Buja etal. (2009) propose an analogy between standard statistical inference and visual inference based on human observers

Lineup protocol:
• Generate $n-1$ decoys (null data sets).
• Make plots of the decoys, and randomly position a plot of the true data.
• Show to an impartial observer. Can they spot the real data?

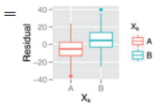
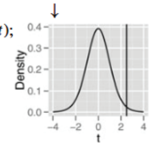
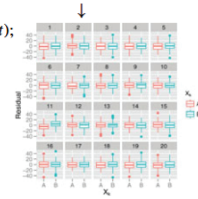
With n=19 decoys, a correct decision by chance would have p=1/20 = 0.05



From: Buja etal. (2009), Statistical inference for exploratory data analysis and model diagnostics. *Phil. Trans. R. Soc. A*, **367**, 4361–4383

Visual inference: Lineup protocol

This graphic table makes the comparison more direct:

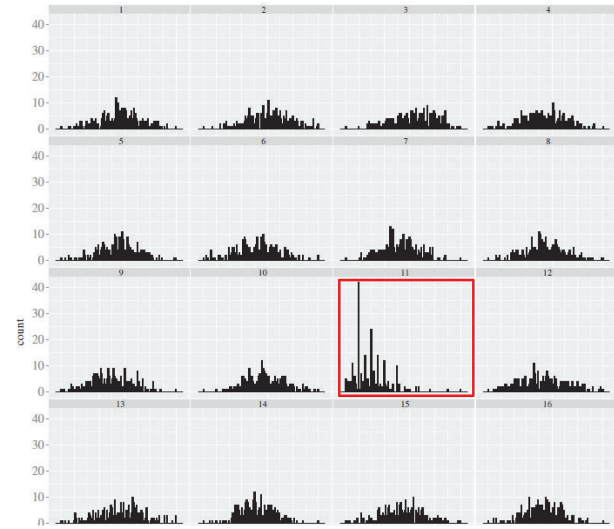
	Conventional inference	Lineup protocol
Hypothesis	$H_0 : \beta = 0$ vs $H_1 : \beta > 0$ ↓	$H_0 : \beta = 0$ vs $H_1 : \beta > 0$ ↓
Test statistic	$T(y) = \frac{\hat{\beta}}{se(\hat{\beta})}$ ↓	$T(y) =$  ↓
Sampling Distribution	$f_{T(y)}(t)$;  ↓	$f_{T(y)}(t)$;  ↓
Reject H_0 if	Actual T is extreme	Actual plot is identifiable

From: Majumder, Hofmann & Cook (2013) Validation of Visual Statistical Inference, Applied to Linear Models, *JASA*, 108:503, 942-956, DOI: 10.1080/01621459.2013.808157

Frequency distribution of tips at restaurants.

Which one is the real data?
What features lead you to this conclusion?

Panel 11:
- Skewed
- Multiple peaks
- Outliers



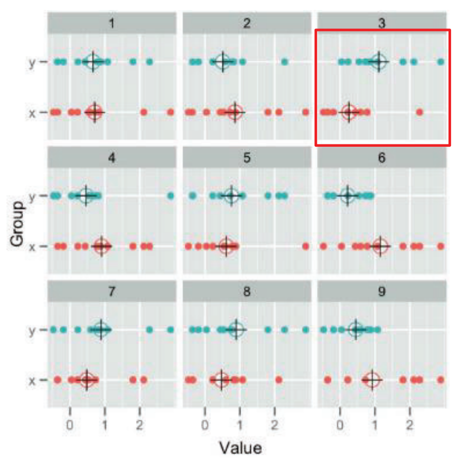
Visual t-test

For each data set, the observations are shown as points and the group means as crosses.

The accused (real data) is hidden amongst eight innocents.

Can you spot him?

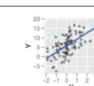
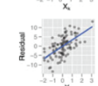
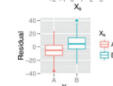
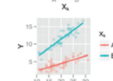
Panel 3: a larger difference among the group means



Visual tests for linear models

This idea can be extended to visual inference for a wide range of hypotheses in linear models.
Main idea: Numerical test → Visual test

Table 3. Visual test statistics for testing hypotheses related to the model $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_1 X_{i2} + \beta_3 X_{i1} X_{i2} + \cdots + \epsilon_i$

Case	null hypothesis	Statistic	Test statistic	Description
1	$H_0 : \beta_0 = 0$	Scatterplot		Scatterplot with least square line overlaid. For null plots we simulate data from fitted null model.
2	$H_0 : \beta_k = 0$	Residual plot		Residual vs X_k plots. For null plots we simulate data from normal with mean 0 variance $\hat{\sigma}^2$.
3	$H_0 : \beta_k = 0$ (for binary X_k)	Boxplot		Boxplot of residuals grouped by category of X_k. For null plots we simulate data from normal with mean 0 variance $\hat{\sigma}^2$.
4	$H_0 : \beta_k = 0$ (interaction of continuous and binary X_k)	Scatterplot		Scatterplot with least square lines of each category overlaid. For null plots we simulate data from fitted null model.

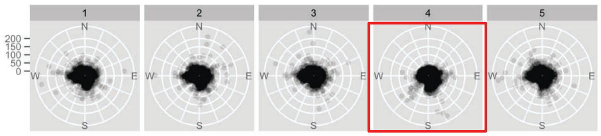
From: Majumder etal (2013) Validation of Visual Statistical Inference, Applied to Linear Models, *JASA*

Visual inference: Power

Some statistical tests have greater power to detect a non-null effect
What can be said about the power of different graphical methods for visual inference?

Graphs of wind direction and arrival delays for incoming flights to Phoenix airport.

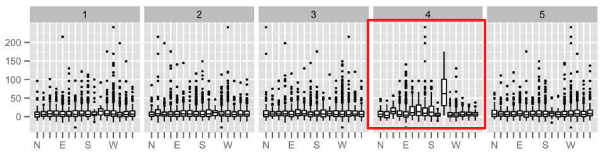
1: polar scatterplot, delay=radius, wind=angle



2: boxplots of delay grouped by angle

Which is easier?

Which is the real data?

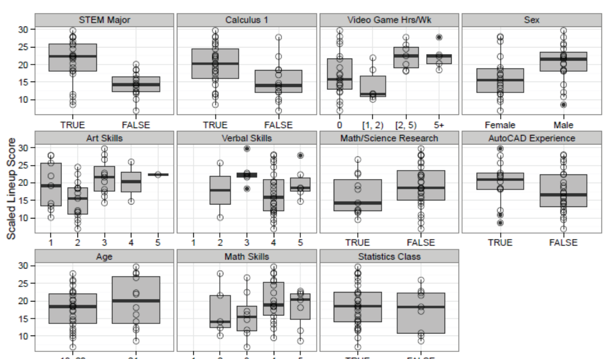


Visual inference: Discussion

- What can we learn from this approach?
  - To what extent can visual inference substitute for numerical statistical inference?
  - How to study this more?
    - How many observers to declare some effect "significant"
    - Can we use this paradigm to study observer differences?
    - Can we use this to study the effectiveness of different graph types or forms?

Online studies: Demographics

It is relatively easy to conduct online studies of graph perception.
- For validity & comprehensibility of results, it is often crucial to examine possible demographic variables that relate to the outcome



What was the effect of various subject variables on outcome in a lineup task?

Various demographic variables are ordered here by the effect size for differences among various groups.

When several visual displays are compared, important to consider interactions – some people may do better with some displays.

from VanderPlas (2015), *Perception in statistical graphics*, PhD thesis, Iowa State U, Fig. 4.5

Discussion questions

- Why is human factor research in graphics useful & important?
  - How can it make a difference?
- What methods are available to study this?
  - What is the task?
  - How to measure "performance"?
- What have we learned?
- How to go forward?