17-803 Empirical Methods

Bogdan Vasilescu, Institute for Software Research

Research Production

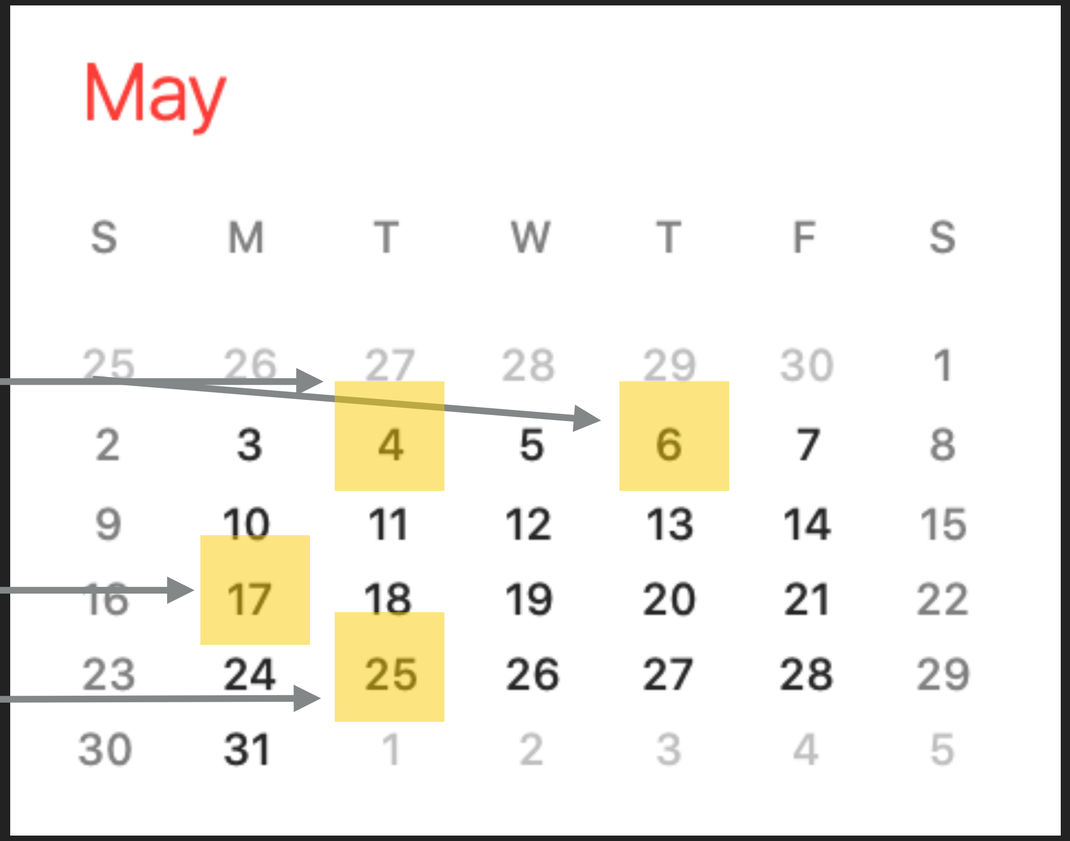
Thursday, April 22, 2021

Final Report

Final presentations

Final report due

Grading deadline



Where We Are in the Semester

Tue, Apr 20	Mixed-methods designs	slides • video
Thu, Apr 22	Stepping up your paper production	
Tue, Apr 27	Agree to disagree	
Thu, Apr 29	Agree to disagree / Retrospective	
Tue, May 4	Final presentations (part I)	
Thu, May 6	Final presentations (part II)	

Plan for Next Week

- Breznau, N., Rinke, E. M., Wuttke, A., Adem, M., Adriaans, J., Alvarez-Benjumea, A., ... & van der Linden, M. (2021). Observing Many Researchers using the Same Data and Hypothesis Reveals a Hidden Universe of Data Analysis.

- Shepperd, M., Bowes, D., & Hall, T. (2014). Researcher bias: The use of machine learning in software defect prediction. IEEE Transactions on Software Engineering, 40(6), 603-616.

- AlShebli, B., Makovi, K., & Rahwan, T. (2020). RETRACTED ARTICLE: The association between early career informal mentorship in academic collaborations and junior author performance. Nature communications, 11(1), 1-8.

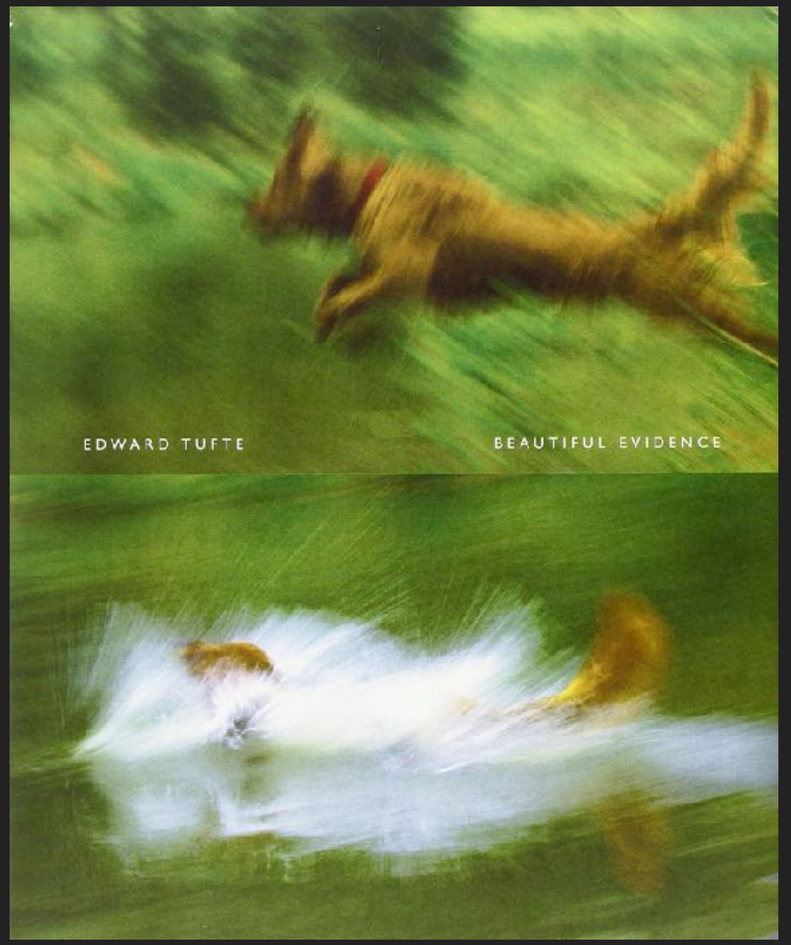
Plan for Next Week

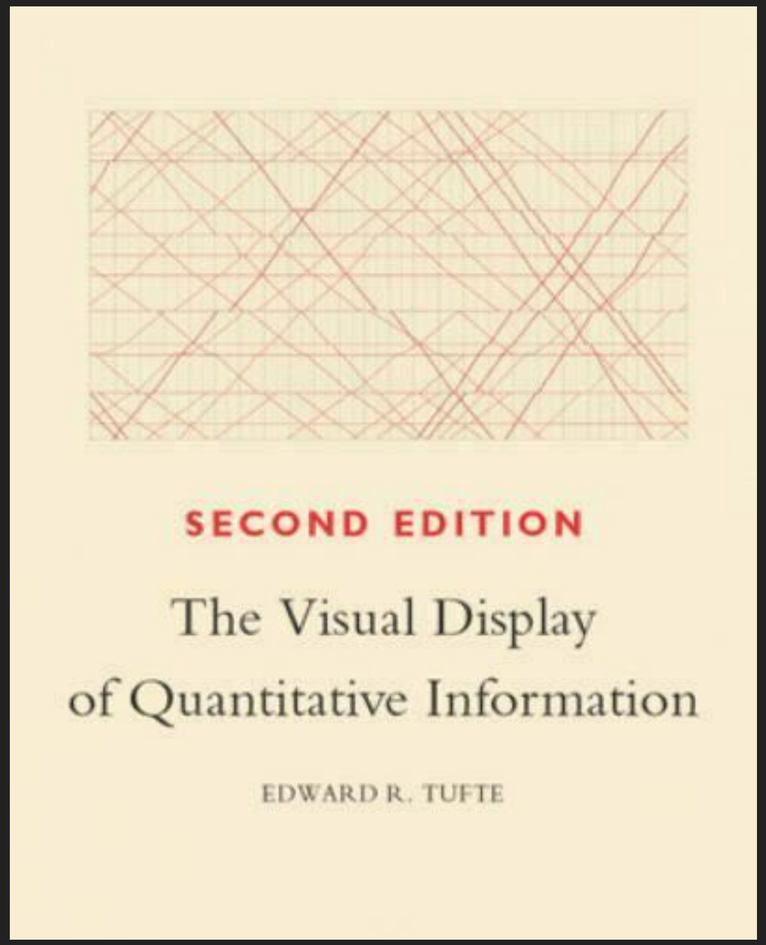
- Ray, B., Posnett, D., Filkov, V., & Devanbu, P. (2014, November). A large scale study of programming languages and code quality in github. In Proceedings of the 22nd ACM SIGSOFT International Symposium on Foundations of Software Engineering (pp. 155-165).

- Berger, E. D., Hollenbeck, C., Maj, P., Vitek, O., & Vitek, J. (2019). On the impact of programming languages on code quality: a reproduction study. ACM Transactions on Programming Languages and Systems (TOPLAS), 41(4), 1-24.

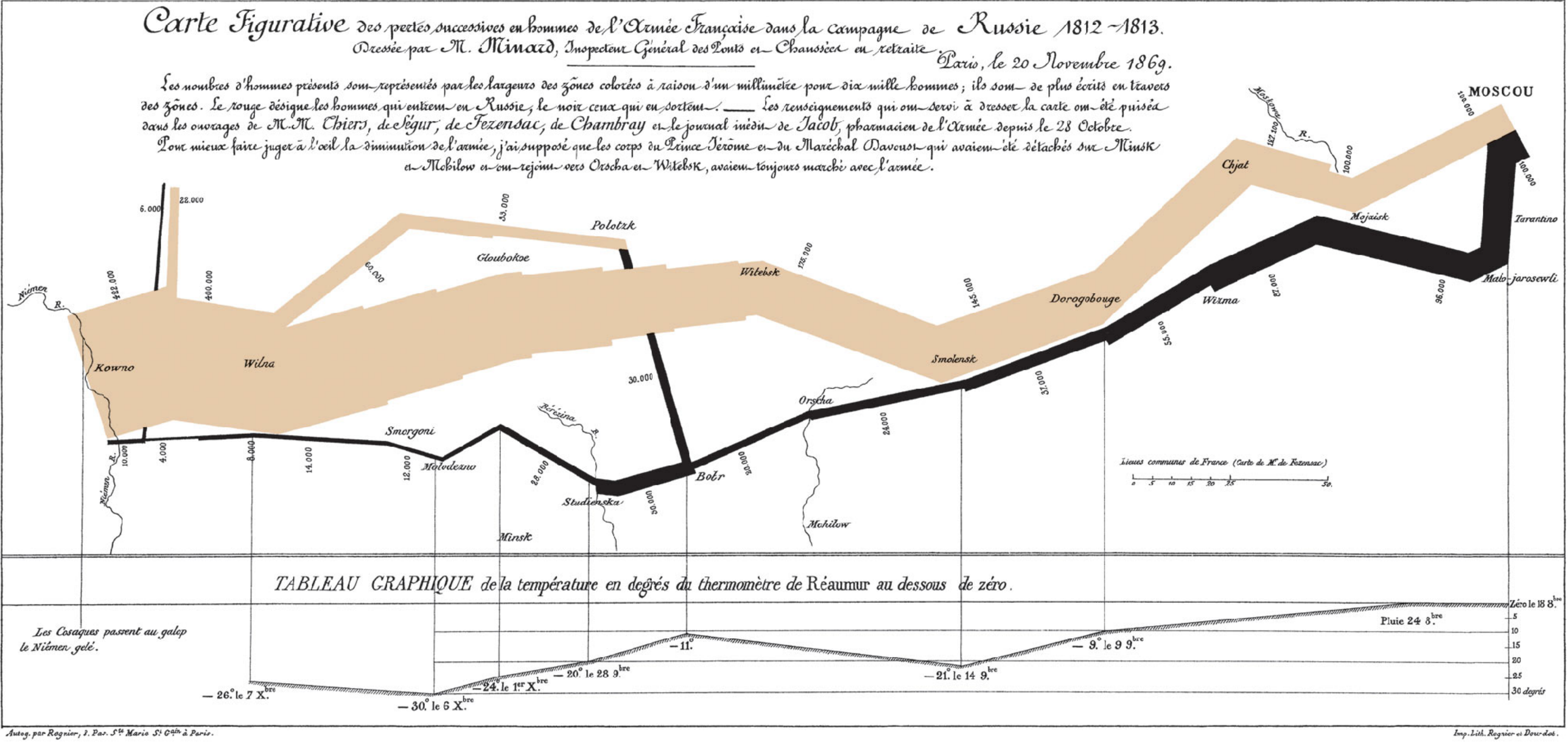
Plan for Today

- Writing
- Presenting
- Graphics





Part I: Next-level graphics



Minard's depiction of the devastating losses suffered in Napoleon's Russian campaign of 1812

Tufte: Above all Else Show the Data

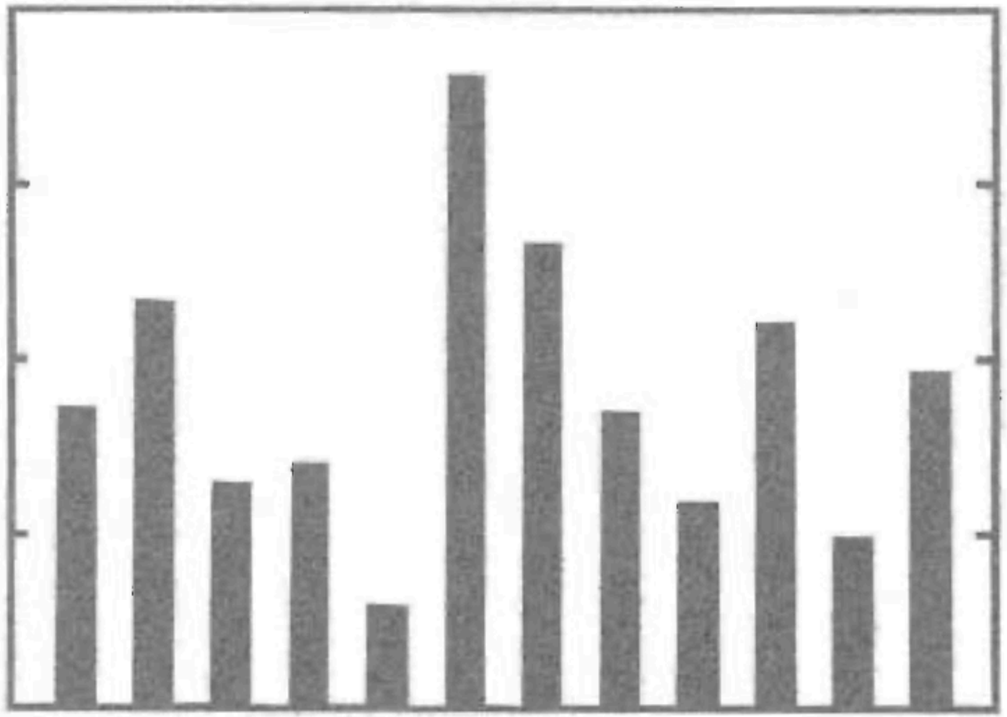
- Data graphics (& tables) should draw the viewer's attention to the sense and substance of the data, not to something else.

Data-ink ratio = data-ink / total ink used to print the graphic (table)

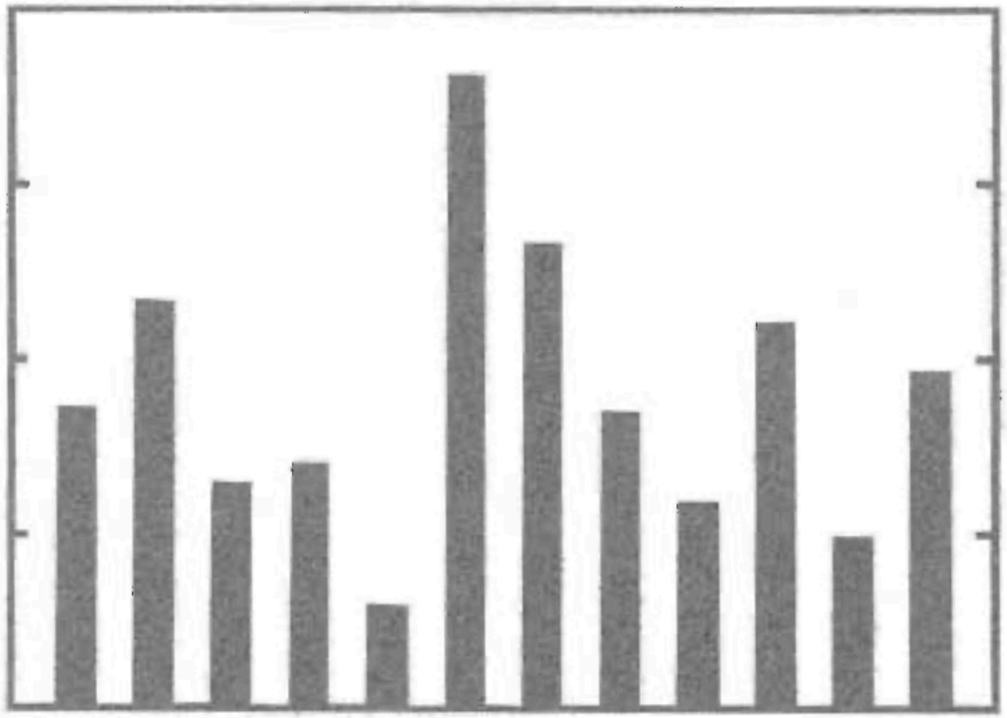
= proportion of a graphic's ink devoted to the non-redundant display of data-information

= 1.0 - proportion of a graphic that can be erased without loss of data-information

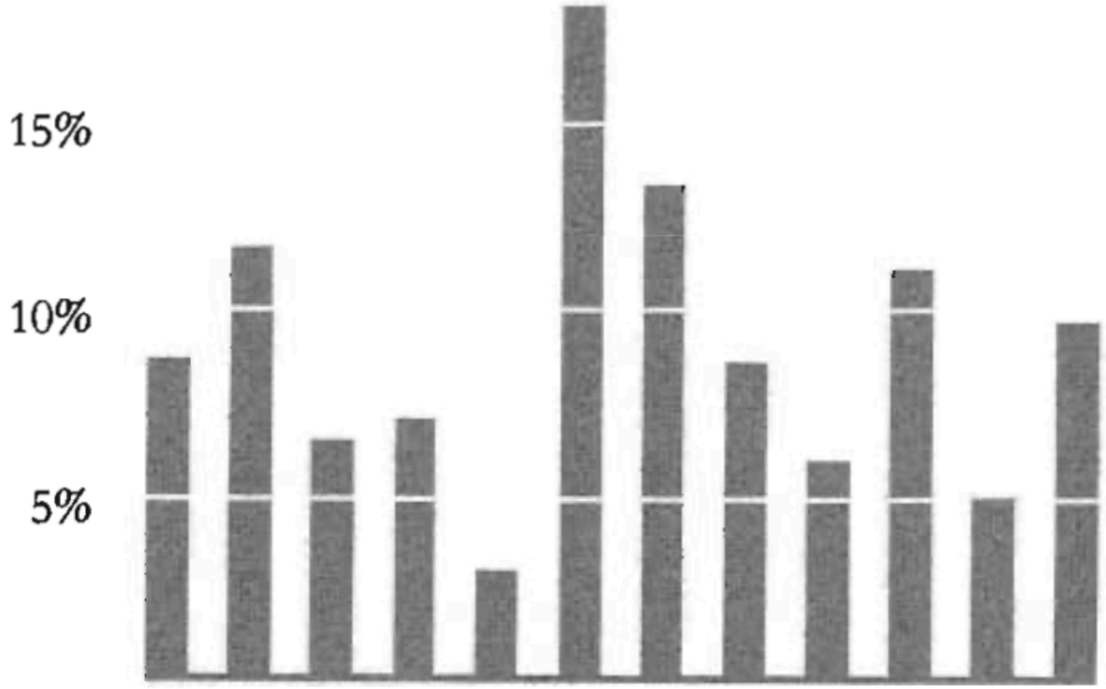
Barplots



Barplots

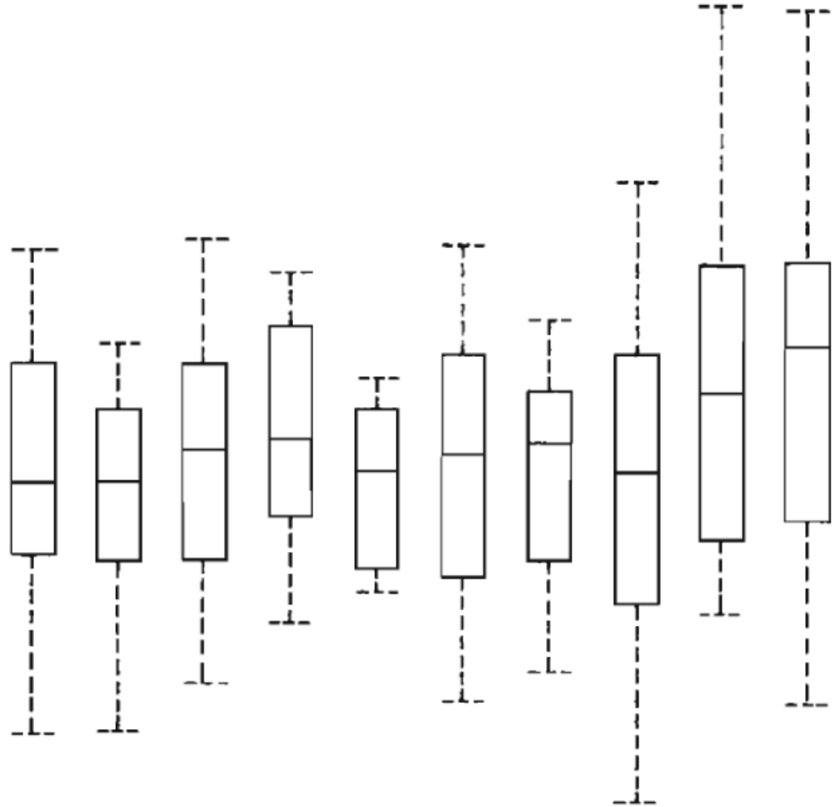


Before

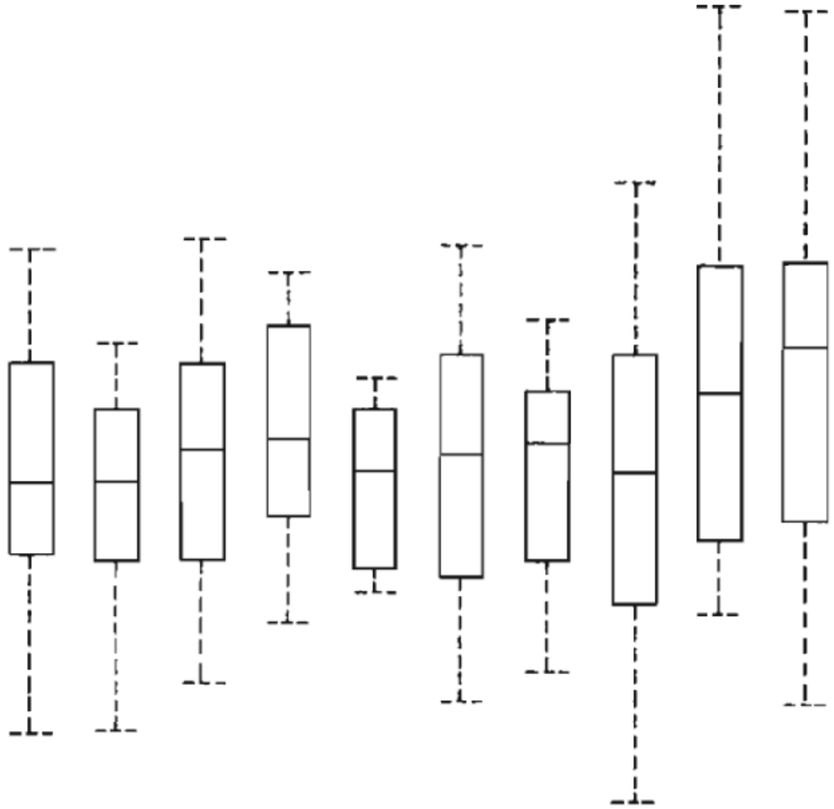


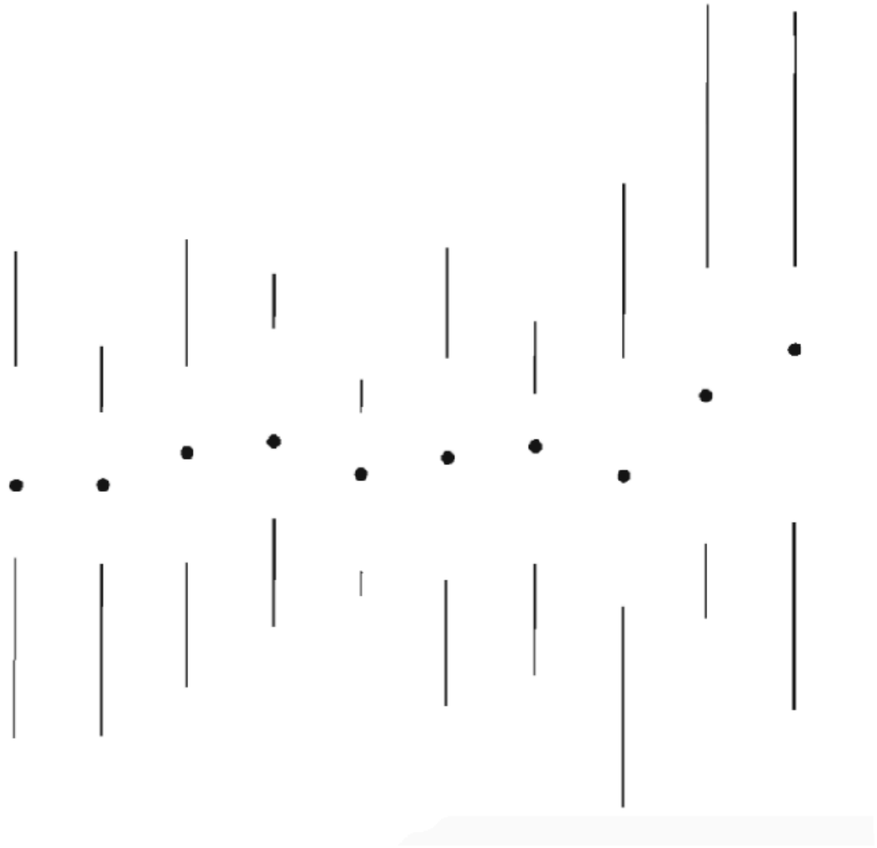
After: No box, no vertical axis, no ticks; added white grid

Boxplots



Boxplots

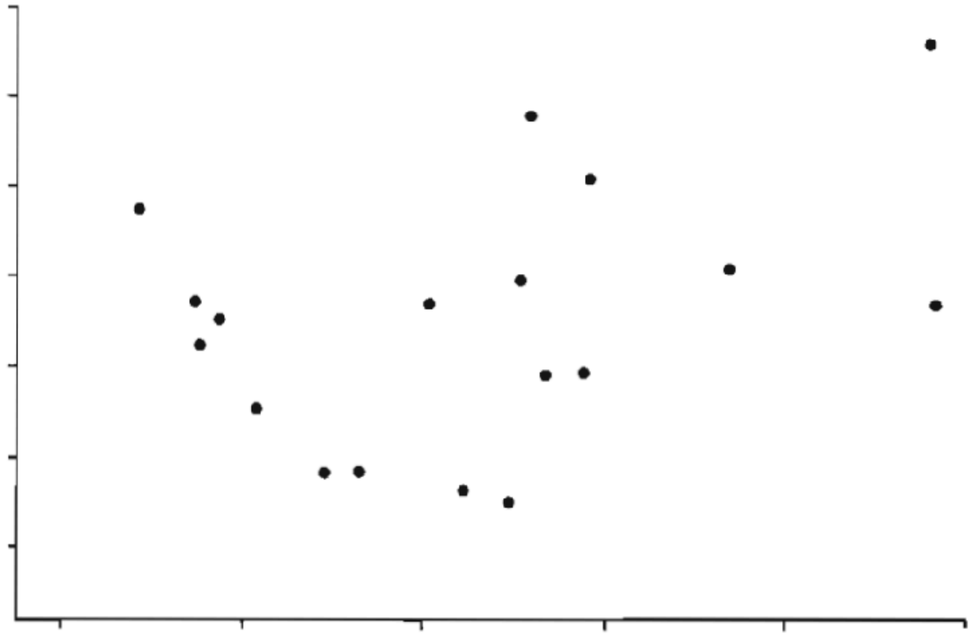




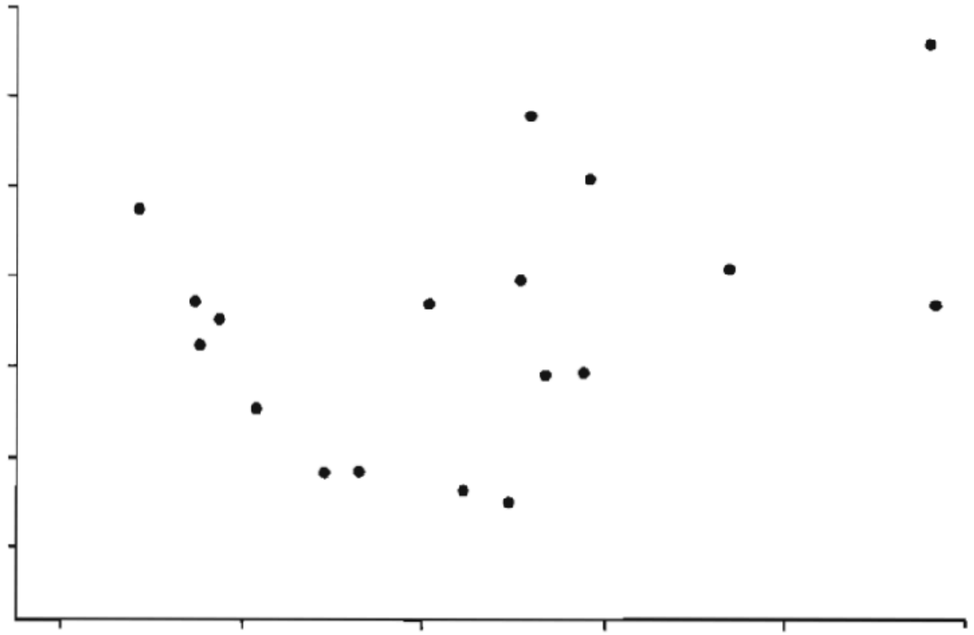
Before: 50 horizontals and 30 verticals

After: 10 verticals

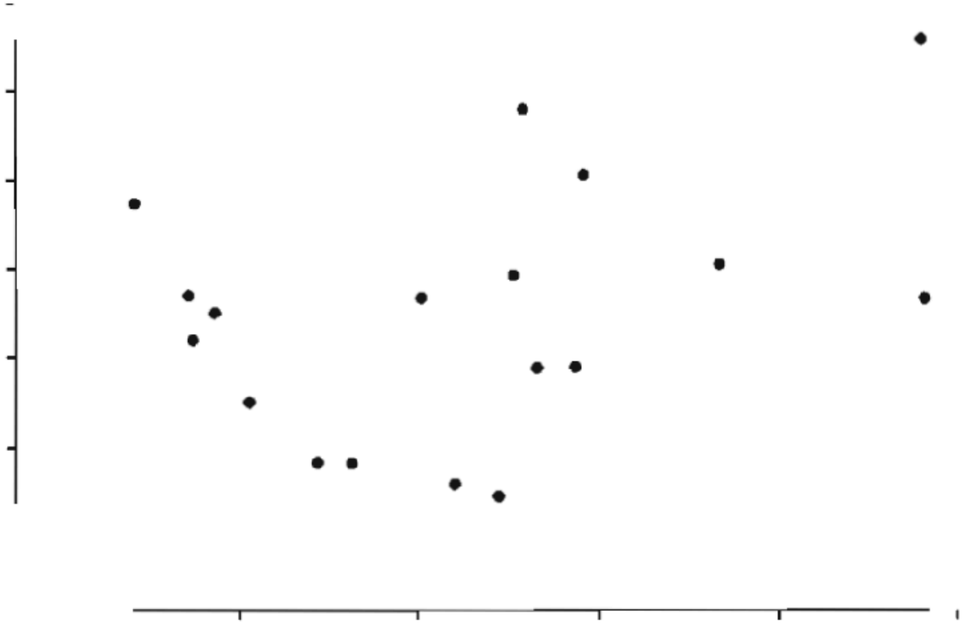
Scatterplots



Scatterplots

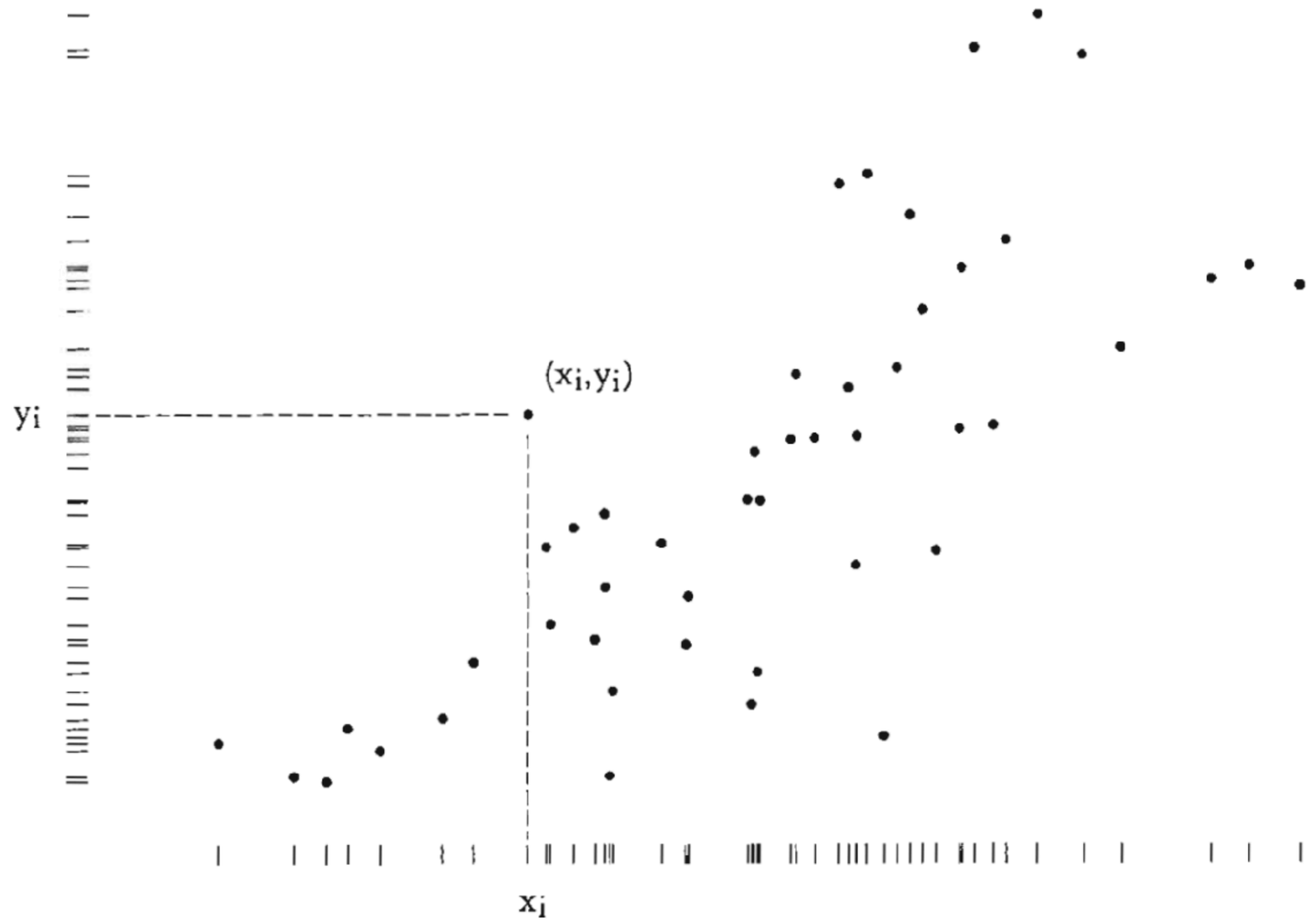


Conventional



Range-frame:
Show max and min of both variables

Scatterplots



Dot-dash-plot: Add marginal frequency distributions

Tables

Data Density and Size of Data Matrix, Statistical Graphics in Selected Publications, Circa 1979–1980

	Data Density (Numbers per square inch) median	minimum	maximum	Size of Data Matrix median	minimum	maximum
Nature	48	3	362	177	15	3780
Journal of the Royal Statistical Society, B	27	4	115	200	10	1460
Science	21	5	44	109	26	316
Wall Street Journal	19	3	154	135	28	788
Fortune	18	5	31	96	42	156
The Times (London)	18	2	122	50	14	440
Journal of the American Statistical Association	17	4	167	150	46	1600
Annuaire Statistique de la France	6	1	25	96	12	540
Scientific American	5	1	69	46	14	652
Statistical Abstract of the United States	5	2	23	38	8	164
American Political Science Review	2	1	10	16	9	40
Pravda	0.2	0.1	1	5	4	20

No vertical lines

Pro Tip: Use Booktabs for Tables

https://www.ctan.org/pkg/booktabs

gnats	gram	$13.65
	each	.01
gnu	stuffed	92.50
emu		33.33
armadillo	frozen	8.99

No

Item		
Animal	Description	Price ($)
Gnat	per gram	13.65
	each	0.01
Gnu	stuffed	92.50
Emu	stuffed	33.33
Armadillo	frozen	8.99

Yes

Pro Tip: Use Booktabs for Tables

https://www.ctan.org/pkg/booktabs

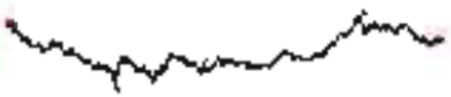
	Item	
Animal	Description	Price ($)
Gnat	per gram	13.65
	each	0.01
Gnu	stuffed	92.50
Emu	stuffed	33.33
Armadillo	frozen	8.99

Also no: \hline

	Item	
Animal	Description	Price ($)
Gnat	per gram	13.65
	each	0.01
Gnu	stuffed	92.50
Emu	stuffed	33.33
Armadillo	frozen	8.99

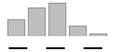
Yes:
\toprule \midrule \bottomrule

Sparklines: Intense, Simple, Word-Sized Graphics

		1999.1.1	65 months	2004.4.28	low	high
Euro foreign exchange	$	1.1608		1.1907	.8252	1.2858
Euro foreign exchange	¥	121.32		130.17	89.30	140.31
Euro foreign exchange	£	0.7111		0.6665	.5711	0.7235

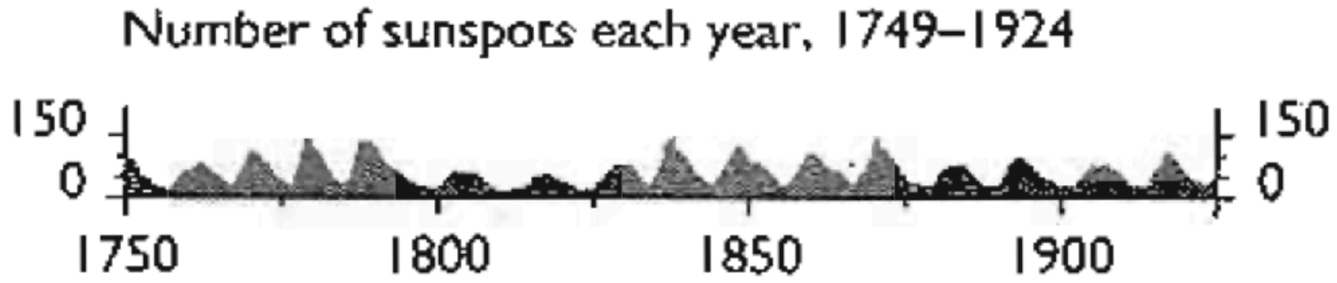
Embed overall trend in data table

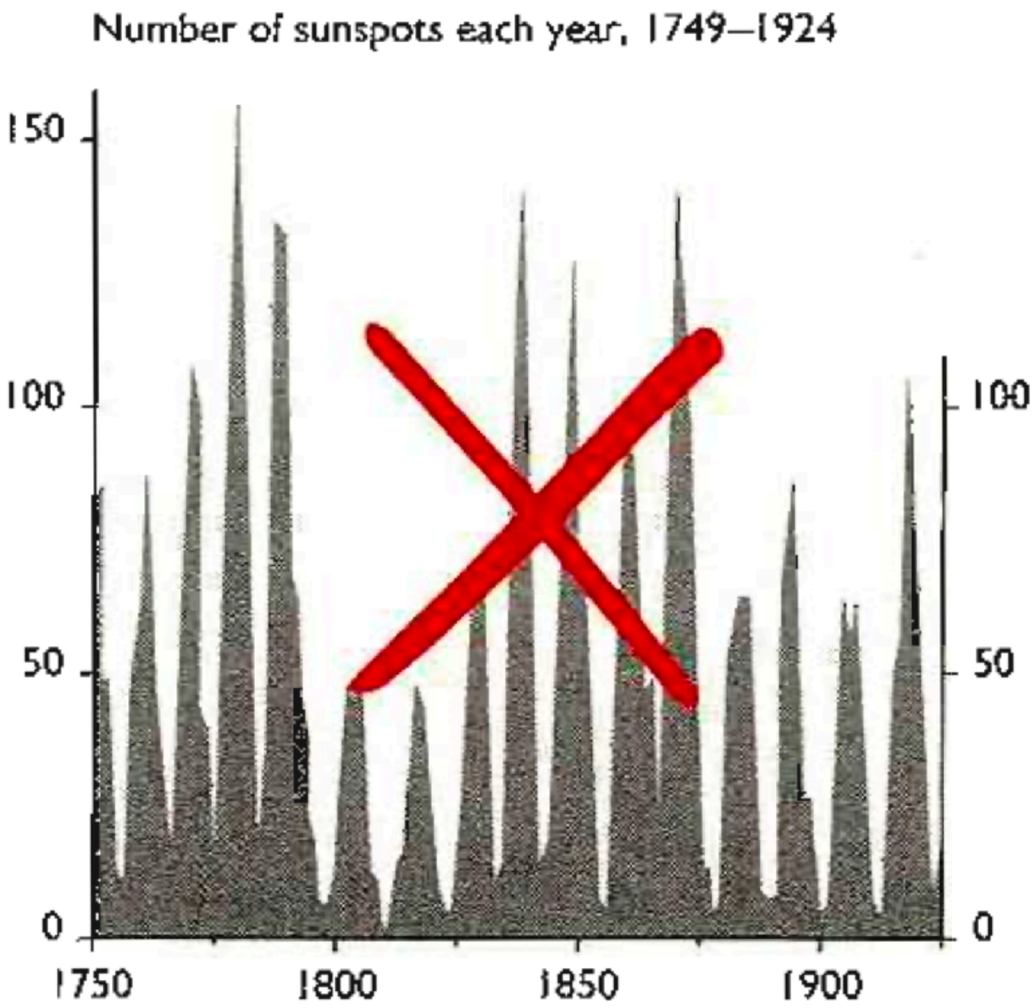
Sparklines: Intense, Simple, Word-Sized Graphics

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16		H	L+M	N	R	U/P
Technical factors																						
Software system																						
Brown- or green-field	•	•	•	•	•	•					•		•			•		10%	55%	18%	✓	*
Functionality																						
Size and complexity	•	•		•	•		•				•							16%	71%	11%		P
Fit for purpose	•	•		•	•		•	•		•			•		•	•		36%	52%	7%		U
Quality																						
Alignment w/ architecture					•		•		•	•			•					21%	69%	5%		U
Usability		•	•				•	•	•		•					•		55%	42%	3%		P
Documentation	•	•	•		•	•	•	•	•	•					•	•		51%	47%	1%		U/P
Security	•	•			•		•				•		•		•	•		39%	53%	4%		U/P
Performance		•						•	•			•			•			32%	63%	5%		P
Well tested	•									•	•				•			11%	68%	18%		U
Release																						
Active maintenance	•	•	•		•		•		•	•	•	•	•			•		44%	47%	7%		U
Maturity and stability	•	•				•			•	•					•			62%	37%	0%		U
Release cycle frequency	•			•			•				•		•					3%	74%	22%		U

Larios Vargas, E., Aniche, M., Treude, C., Bruntink, M., & Gousios, G. (2020). Selecting third-party libraries: The practitioners' perspective. In ESEC/FSE (pp. 245-256).

Length Versus Height





Prefer lumpy over spiky

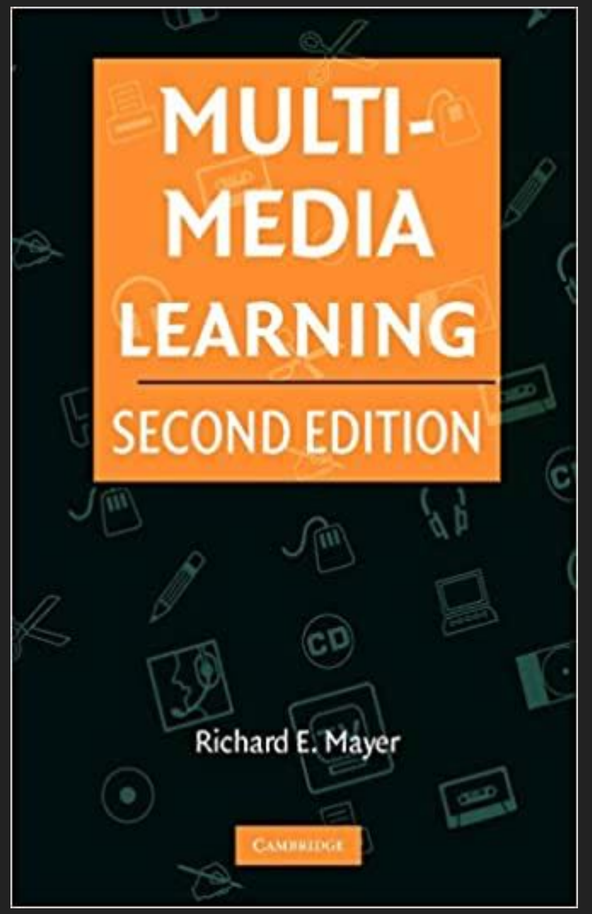
Part II: Some of the science of presenting and slide design

(How to Avoid “Death by PowerPoint”)

The Design of PowerPoint Presentations Should Be Compatible With How People Learn.

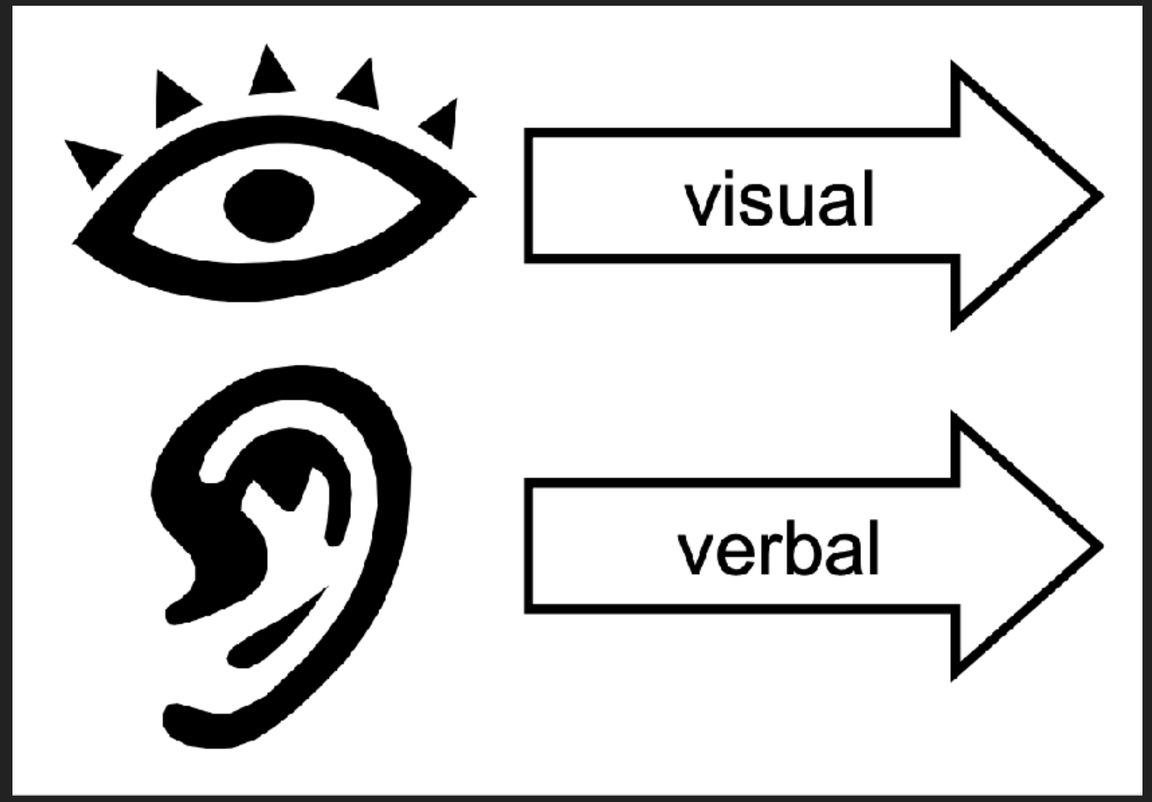
"Most PowerPoint presentations look a particular way because the PowerPoint tool has features that make particular tasks easy. PowerPoint makes it easy to use templates, so we use templates. PowerPoint makes it easy to use bulleted lists, so we use bulleted lists. PowerPoint makes it easy to paste many items on a screen, so we paste them onto the screen. By using these features, we are making specific assumptions about the way people learn.

Unfortunately, many PowerPoint features and techniques contradict current research in cognitive science. We can no longer expect our audiences to adapt to our PowerPoint features; instead we have to change our own thinking to conform to [how people learn]."



Atkinson, C., & Mayer, R. E. (2004). Five ways to reduce PowerPoint overload. Creative Commons, 1.

Cognitive Scientists Say the Mind Processes Information in 2 Channels

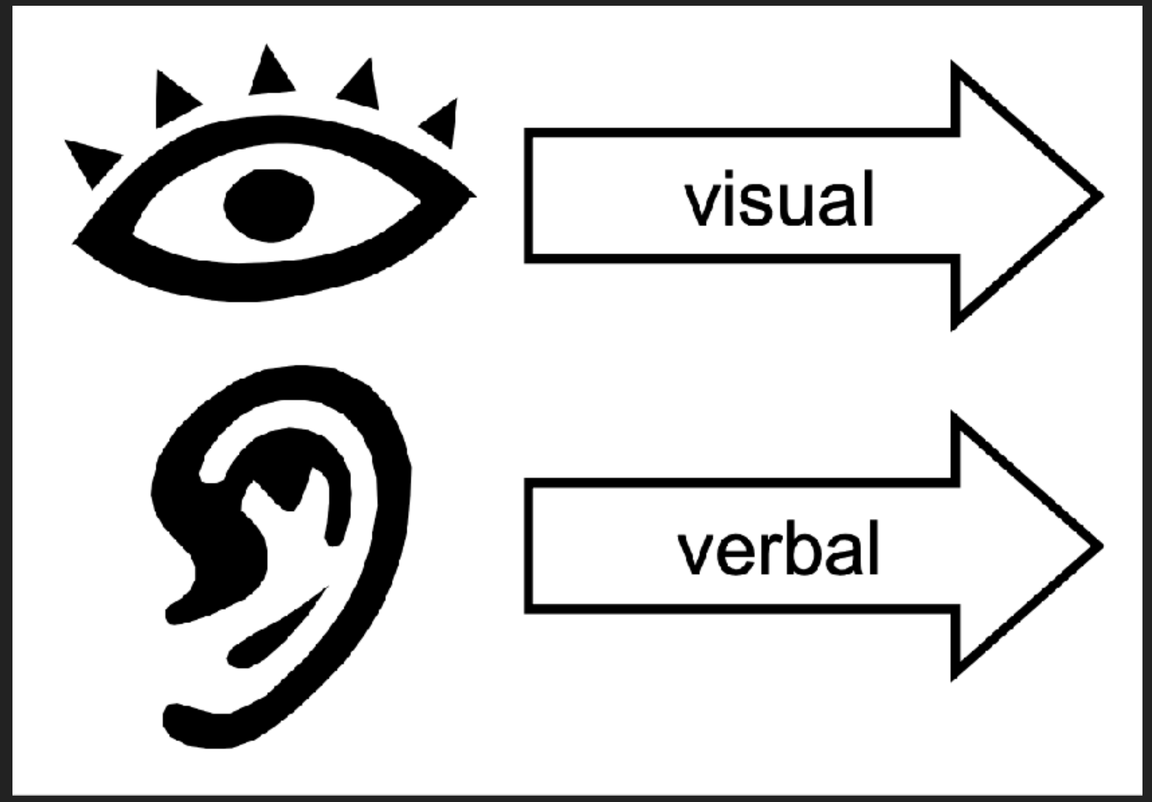


Dual channel

The **visual channel** handles information presented to the eyes (such as illustrations, animation, video, or on-screen text).

The **verbal channel** handles information presented to the ears (such as narration or nonverbal sounds).

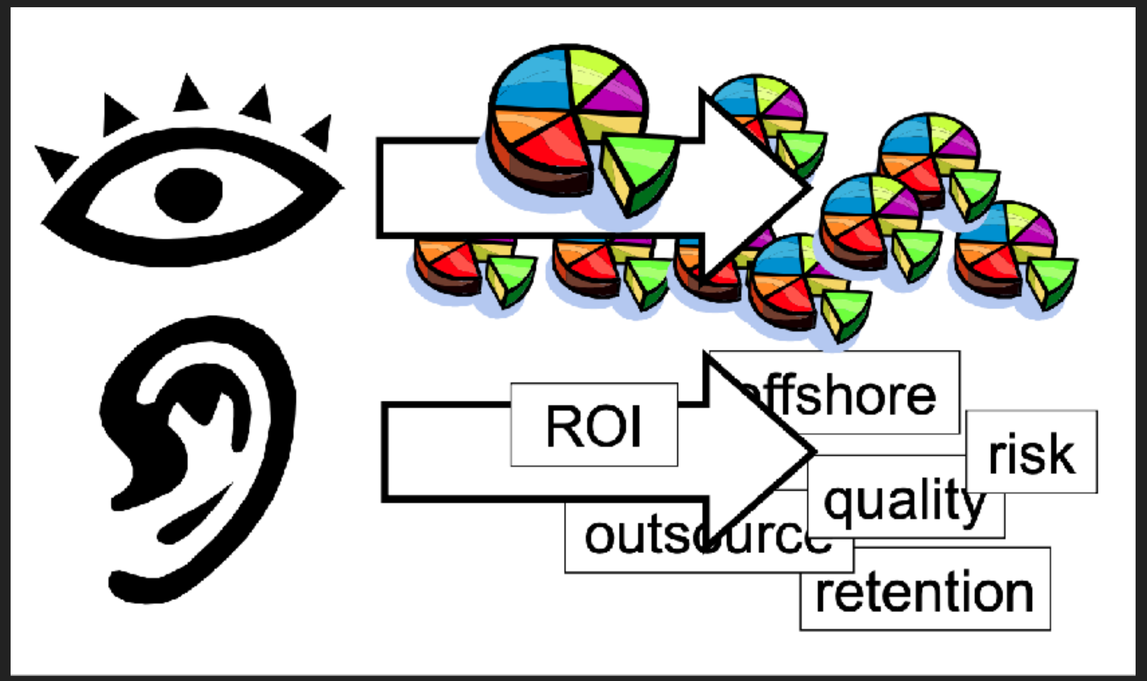
Cognitive Scientists Say the Mind Processes Information in 2 Channels



Dual channel

The constraints on our processing capacity force us to make decisions about which pieces of incoming information to pay attention to, and the degree to which we should build connections between selected pieces of information and our existing knowledge.

The Mind Pays Attention to Only a Few Pieces of Information in Each Channel

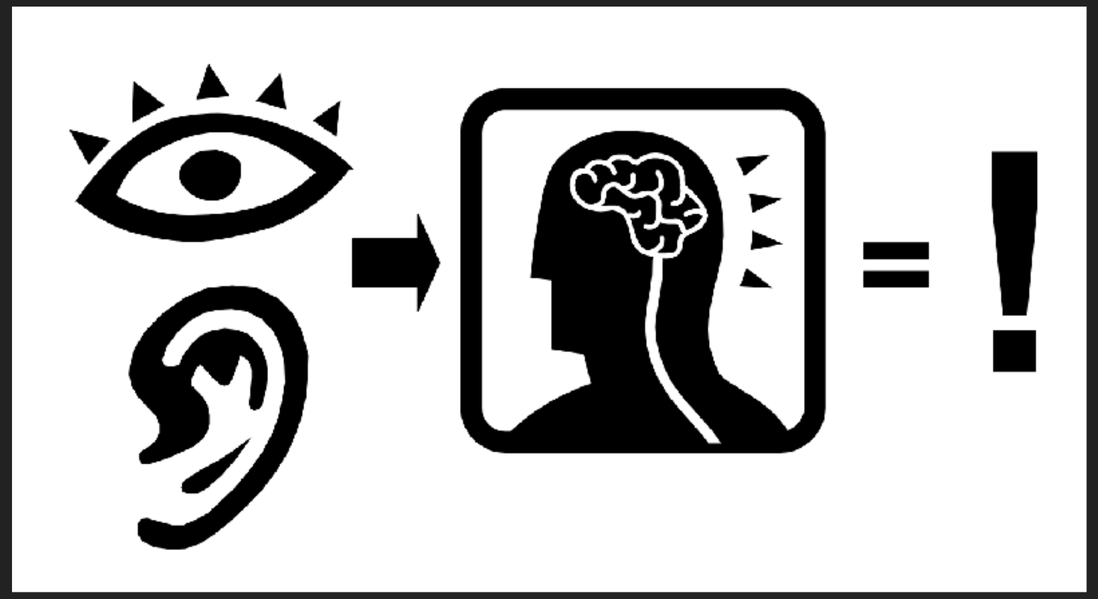


When an illustration or animation is presented, the learner is able to hold only a few images in working memory at any one time.

When a narration is presented, the learner is able to hold only a few words in working memory at any one time.

Limited capacity

The Mind Needs Space To Select, Organize & Integrate What's Important



Active processing

People understand the presented material when they pay attention to the relevant material, organize it into a coherent mental structure, and integrate it with their prior knowledge.

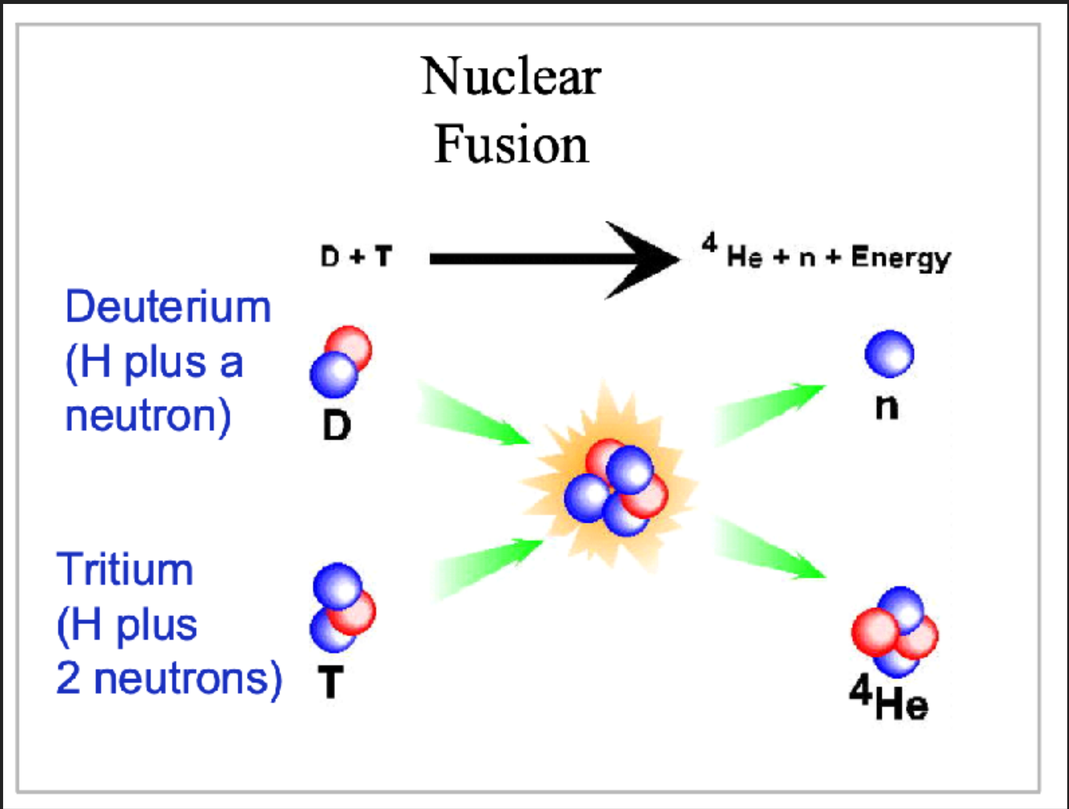
This view of humans as active processors conflicts with a common view of humans as passive processors who seek to add as much information as possible to memory.

Our Understanding of the Way the Mind Works Has Three Implications for PowerPoint:

1. PowerPoint presentations should use both visual and verbal forms of presentation;
2. Filling the slides with information will easily overload people's cognitive systems;
3. The presentations should help learners to select, organize, and integrate presented information.

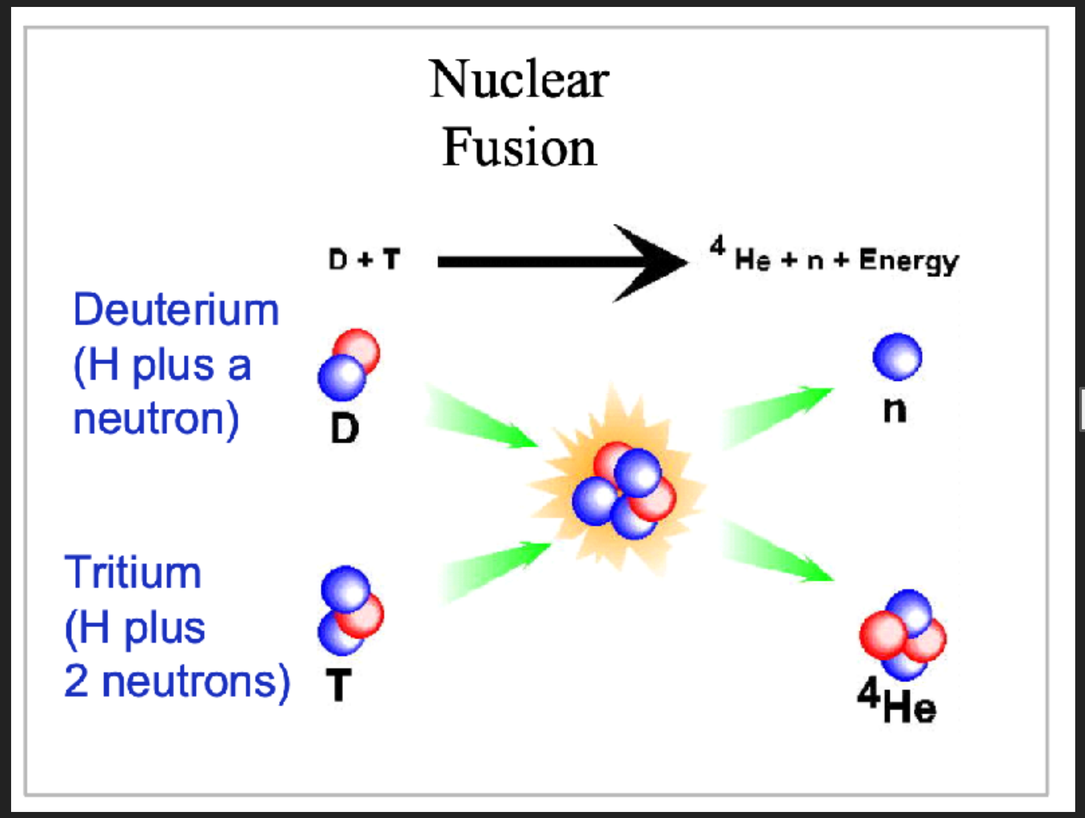
The following set of research-based techniques take these implications into account, and can help reduce cognitive load in PowerPoint.

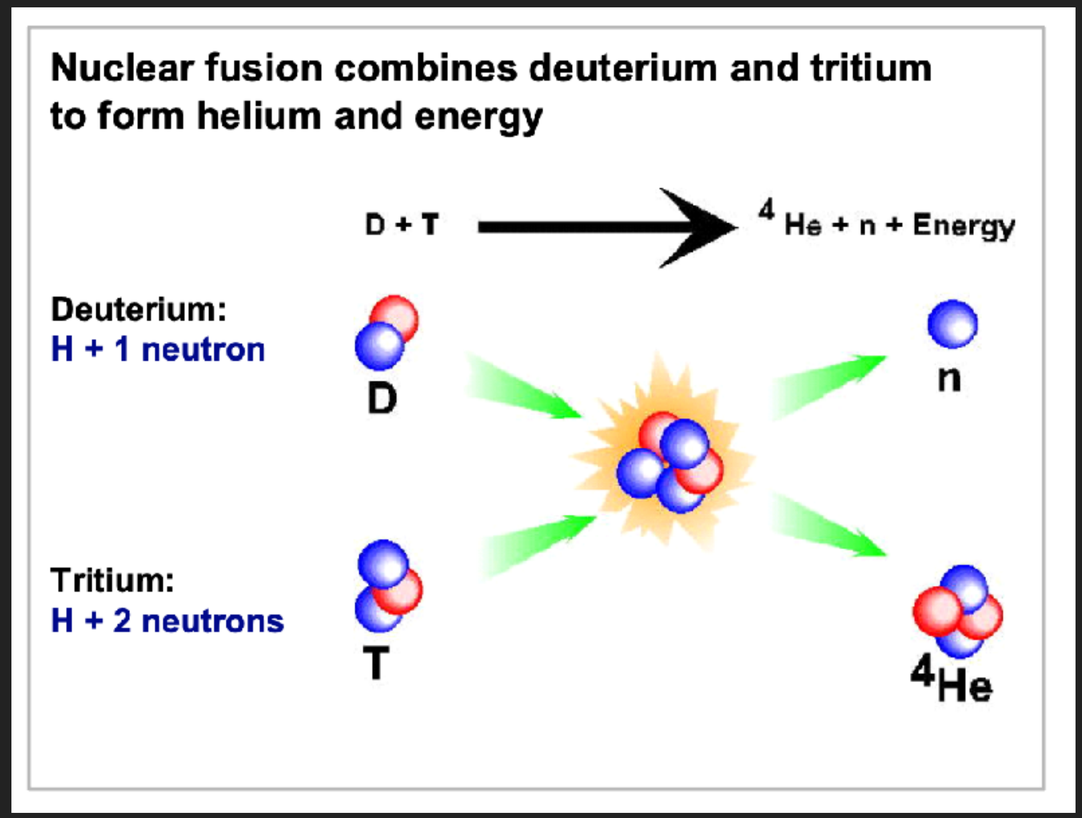
1. Write a Clear Headline That Explains the Main Idea of Every Slide (the "Assertion–Evidence" Model)



1. Write a Clear Headline That Explains the Main Idea of Every Slide (the "Assertion–Evidence" Model)

Improvement of test scores 55% → 78% to the Q:
What is the chemical representation for nuclear fusion?





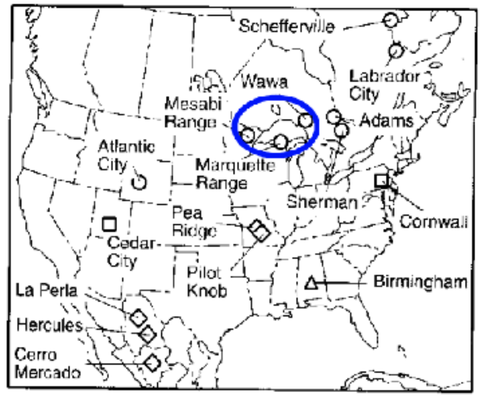
Alley, M., Schreiber, M., & Muffo, J. (2005, October). Pilot testing of a new design for presentation slides to teach science and engineering. In *Proceedings Frontiers in Education 35th Annual Conference* (pp. S3G-7). IEEE.

1. Write a Clear Headline That Explains the Main Idea of Every Slide (the "Assertion–Evidence" Model)

Iron

- An abundant metal, makes up 5.6% of earth's crust
- Properties:
  - shaped, sharpened, welded
  - strong, durable
- Accounts for >95% of metals used
- Iron ores discovered in 1844 in Michigan's Upper Peninsula
- Soon found other ores in upper Wisconsin and Minnesota

Iron Ore Distribution



Kesler 1994

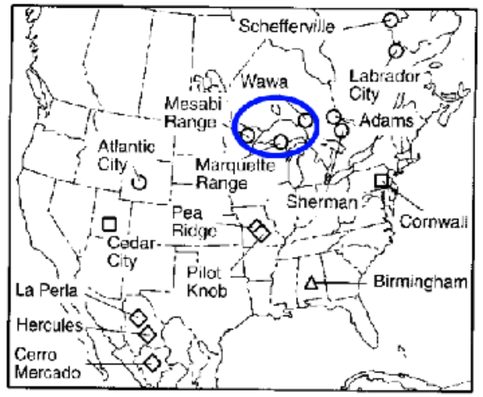
1. Write a Clear Headline That Explains the Main Idea of Every Slide (the "Assertion–Evidence" Model)

Improvement of test scores 59% → 77% to the Q: *How abundant is iron in the earth's crust?*

Iron

- An abundant metal, makes up 5.6% of earth's crust
- Properties:
  - shaped, sharpened, welded
  - strong, durable
- Accounts for >95% of metals used
- Iron ores discovered in 1844 in Michigan's Upper Peninsula
- Soon found other ores in upper Wisconsin and Minnesota

Iron Ore Distribution



Kesler 1994

Iron ores make up 5.6% of the earth's crust and account for 95% of the metals used

Iron Ore Distribution

Is strong and durable

Can be shaped, sharpened, and welded

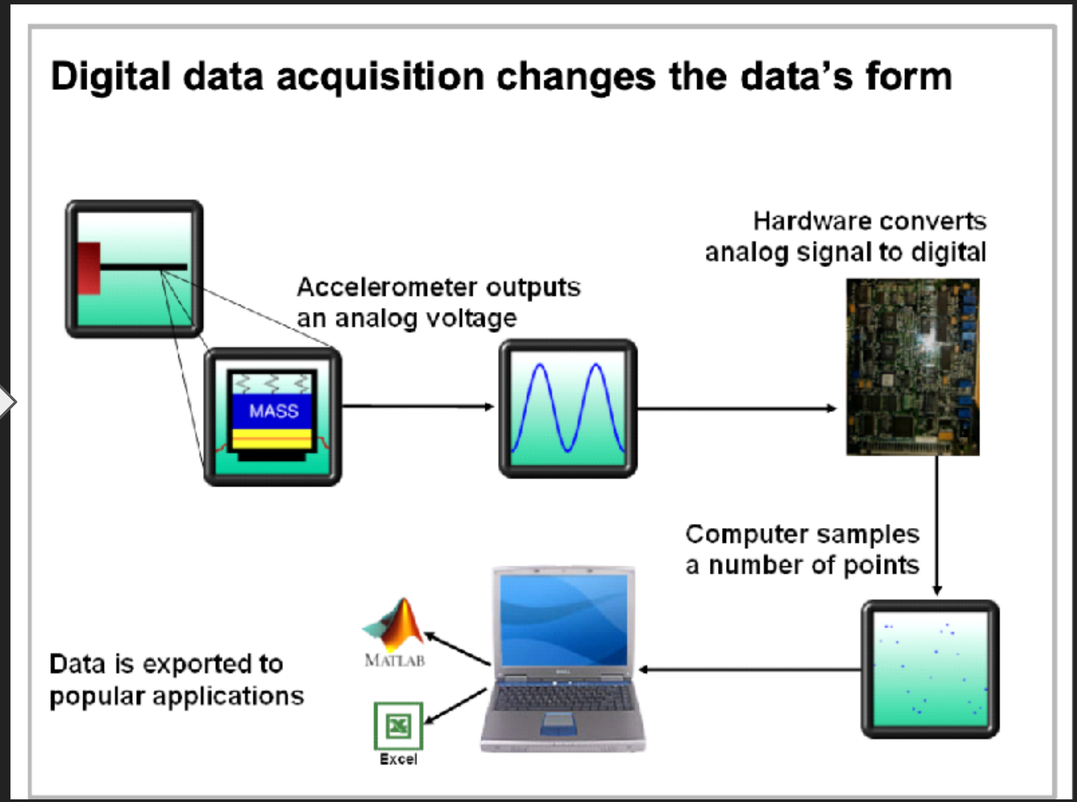
[Kesler 1994]

Alley, M., Schreiber, M., & Muffo, J. (2005, October). Pilot testing of a new design for presentation slides to teach science and engineering. In Proceedings Frontiers in Education 35th Annual Conference (pp. S3G-7). IEEE.

1. Write a Clear Headline That Explains the Main Idea of Every Slide (the "Assertion–Evidence" Model)

Digital Acquisition System

- Accelerometer outputs an analog voltage
- Hardware converts analog signal to digital
- Computer samples a number of points
- Data is exported to popular applications
  - o Microsoft Excel
  - o Matlab



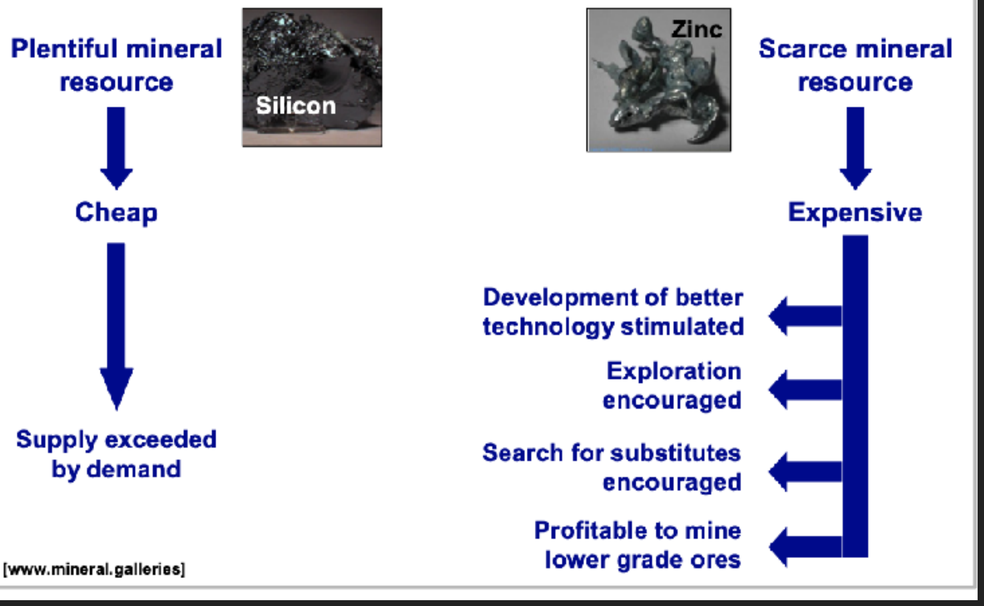
Alley, M., Schreiber, M., & Muffo, J. (2005, October). Pilot testing of a new design for presentation slides to teach science and engineering. In Proceedings Frontiers in Education 35th Annual Conference (pp. S3G-7). IEEE.

1. Write a Clear Headline That Explains the Main Idea of Every Slide (the "Assertion–Evidence" Model)

Mineral Economics

- Free market:
  - plentiful mineral resource is cheap when supply exceeds demand.
  - When resource becomes scarce, price increases =>
    - encourages exploration
    - stimulates development of better technology
    - makes it profitable to mine lower grade ores
    - encourages search for substitutes
    - promotes conservation

In a free market, the supply of the resource determines the price



Alley, M., Schreiber, M., & Muffo, J. (2005, October). Pilot testing of a new design for presentation slides to teach science and engineering. In Proceedings Frontiers in Education 35th Annual Conference (pp. S3G-7). IEEE.

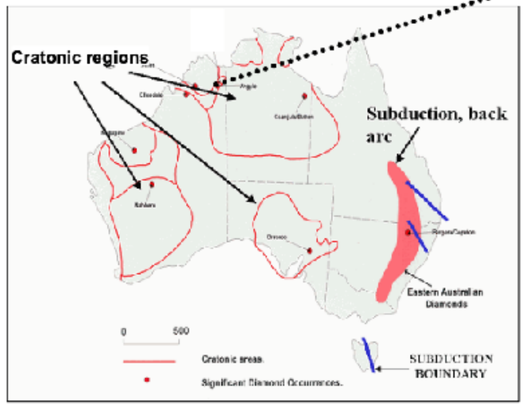
1. Write a Clear Headline That Explains the Main Idea of Every Slide (the "Assertion–Evidence" Model)

Diamonds in Australia

- During 1980s, became world's largest producer of diamonds
- First discovery in 1851, but major kimberlites not discovered until 1976 in western Australia.
- Largest pipe: Argyle mine, 60% control by Rio Tinto (British Co.).
- Single largest mine in world, produces 34 million carats a year. Most are small (average 0.08 carat), only 5% of gem quality. Unique feature: has small but consistent number of pink, red, and purple diamonds, very rare.
- Opted out of CSO, marketing in competition with DeBeers

During 1980s, Australia became the world's largest producer of diamonds

In 1976, major kimberlites were found in western Australia



Map of Australian diamond occurrences

Largest pipe is Argyle mine:

single largest mine in world, producing 34 million carats a year

in competition with DeBeers

small but consistent number of pink, red, and purple diamonds

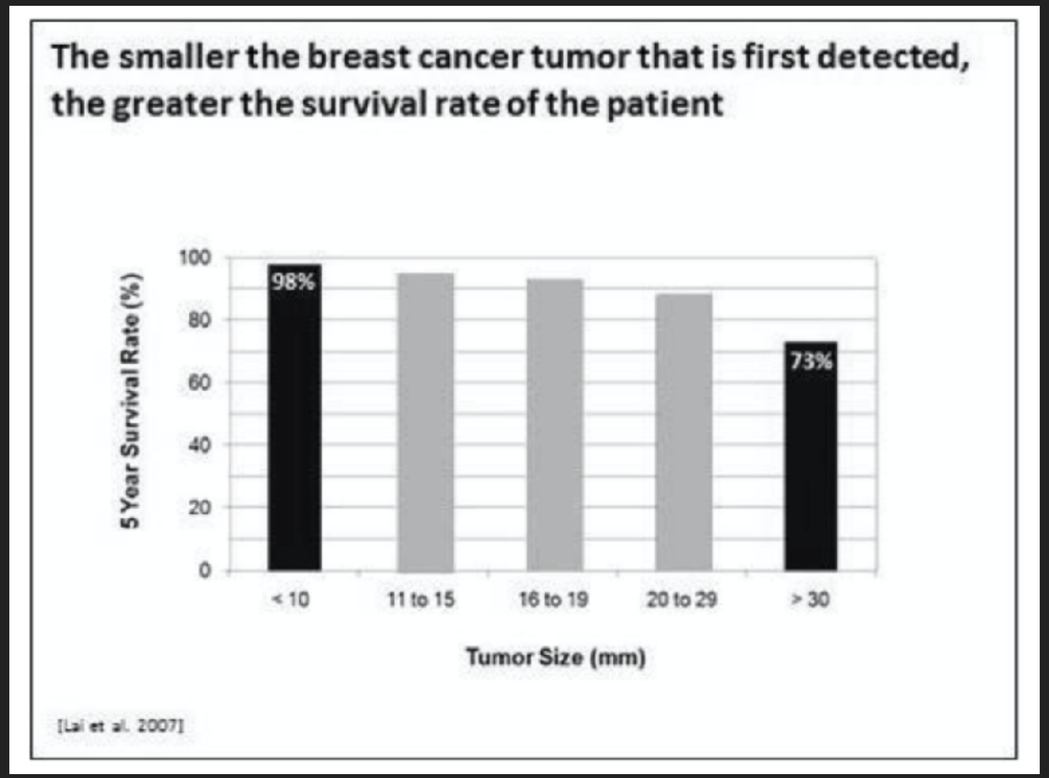
Alley, M., Schreiber, M., & Muffo, J. (2005, October). Pilot testing of a new design for presentation slides to teach science and engineering. In Proceedings Frontiers in Education 35th Annual Conference (pp. S3G-7). IEEE.

1. Write a Clear Headline That Explains the Main Idea of Every Slide (the "Assertion–Evidence" Model)

Tumor Size vs. Survival Rate

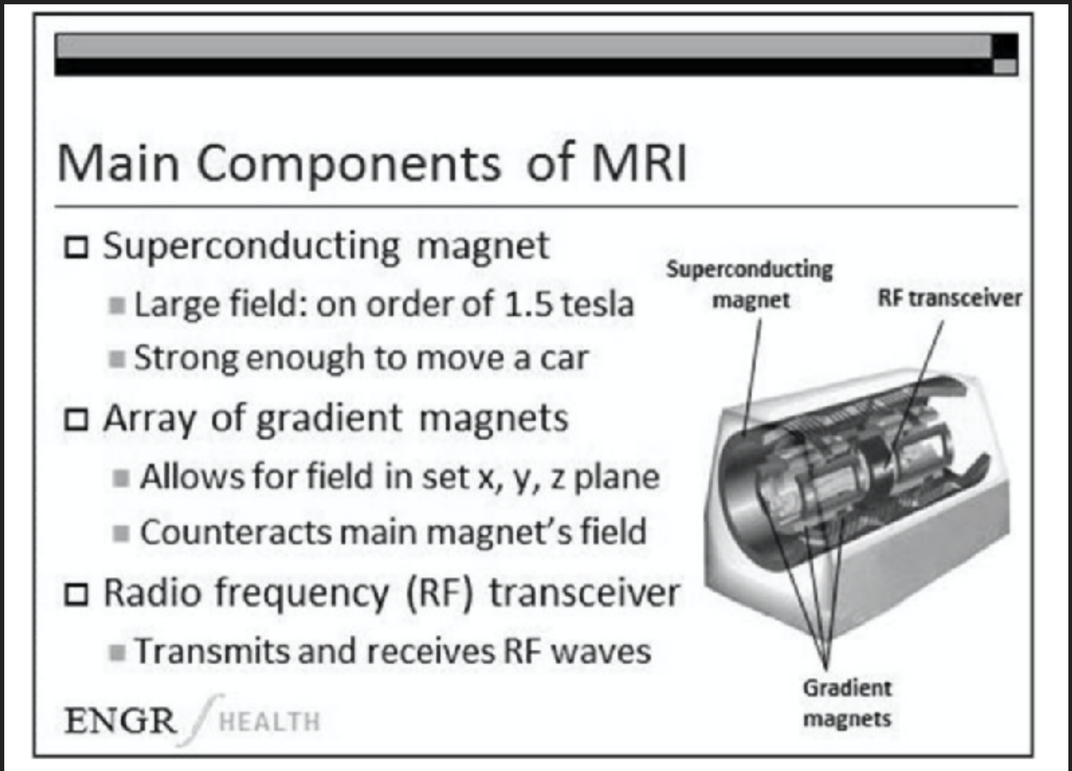
- Tumor size a key in predicting survival rates
- Survival *increases* as discovered tumor size *decreases*
- 2007 study of 10,000 Australian women:
  - > 30 mm: 73% chance for 5-year survival
  - < 10 mm: 98% chance for 5-year survival

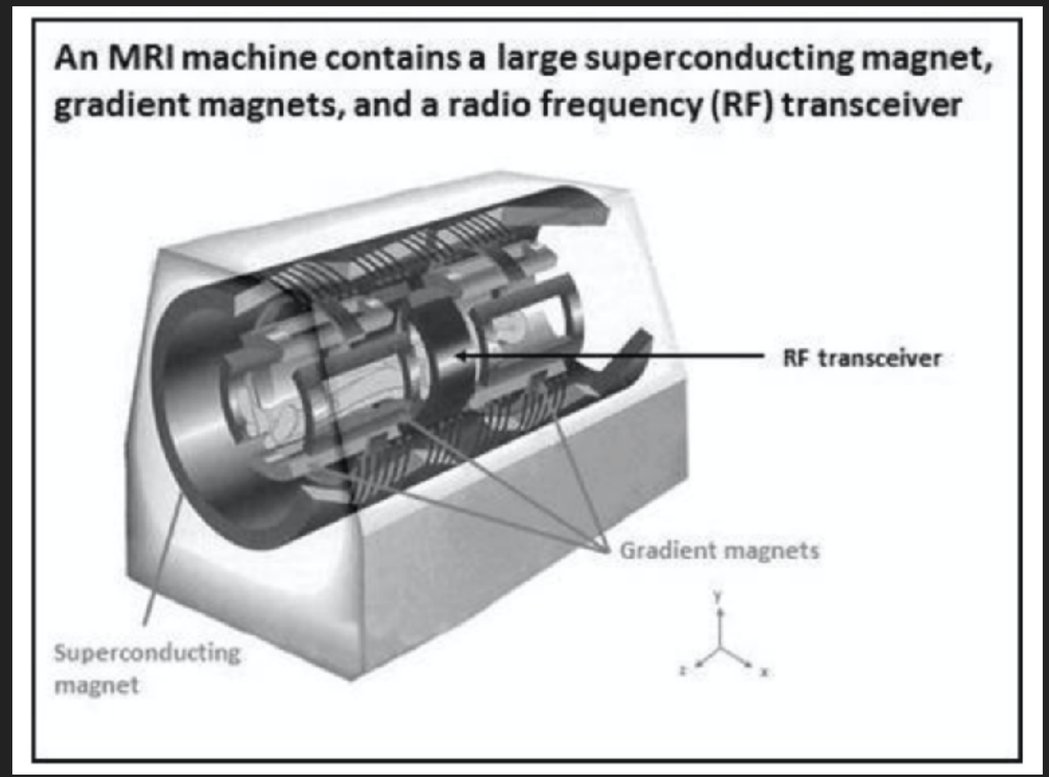
ENGR / HEALTH



Garner, J., & Alley, M. (2013). How the design of presentation slides affects audience comprehension: A case for the assertion-evidence approach. International Journal of Engineering Education, 29(6), 1564-1579.

1. Write a Clear Headline That Explains the Main Idea of Every Slide (the "Assertion–Evidence" Model)

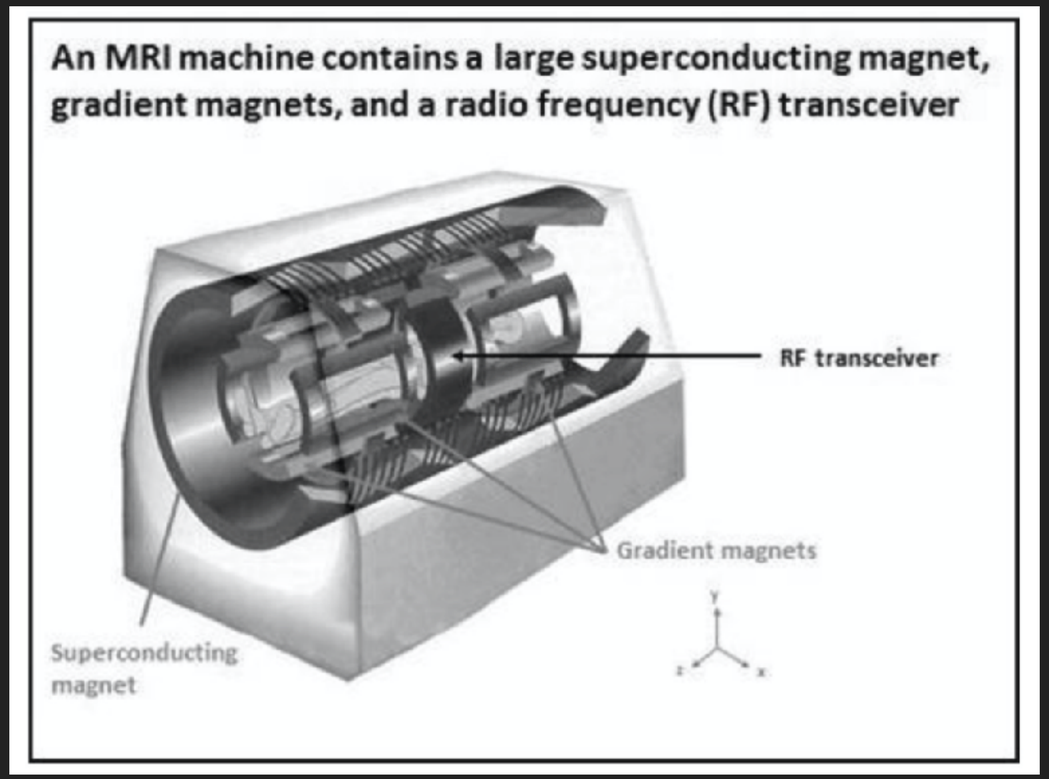




Garner, J., & Alley, M. (2013). How the design of presentation slides affects audience comprehension: A case for the assertion-evidence approach. International Journal of Engineering Education, 29(6), 1564-1579.

1. Write a Clear Headline That Explains the Main Idea of Every Slide (the "Assertion–Evidence" Model)

"Essay responses from the 110 engineering students revealed **superior comprehension** and **fewer misconceptions** for the assertion-evidence group as well as **lower perceived cognitive load**. In addition, **stronger recall** occurred in this assertion-evidence group at delayed post-test."



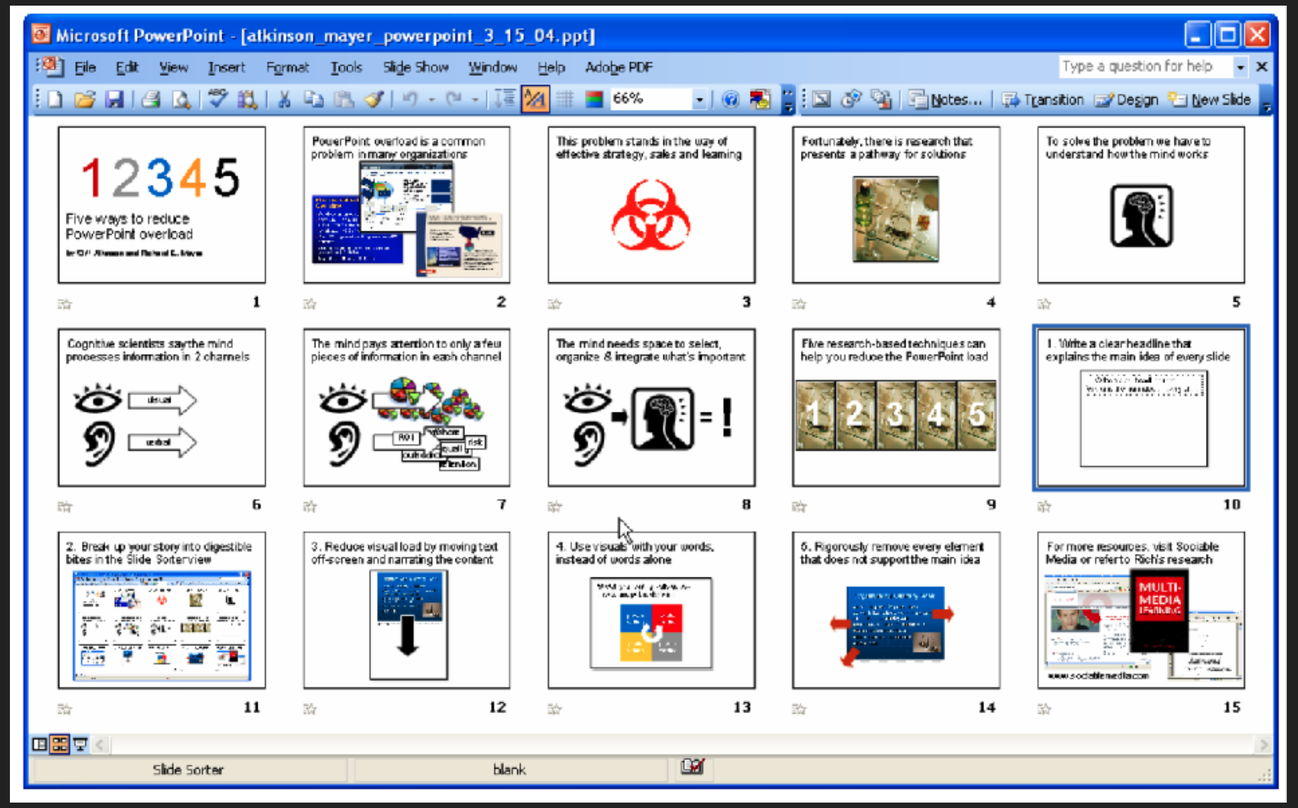
Garner, J., & Alley, M. (2013). How the design of presentation slides affects audience comprehension: A case for the assertion-evidence approach. International Journal of Engineering Education, 29(6), 1564-1579.

2. Break Up Your Story Into Digestible Bites

People learn better when information is presented in bite-size segments.

Segmentation of multimedia instruction facilitates basic (recall) and deep (application) knowledge acquisition.

The use of segmentation mediates the effects of working memory capacity (WMC) to allow learners with lower WMC to recall and apply equal to those with higher WMC.



Lusk, D. L., Evans, A. D., Jeffrey, T. R., Palmer, K. R., Wikstrom, C. S., & Doolittle, P. E. (2009). Multimedia learning and individual differences: Mediating the effects of working memory capacity with segmentation. British Journal of Educational Technology, 40(4), 636-651.

Tip: Outline First, Slides Later

- Use paper or post-its
- Write down a main idea – one sentence for each slide
- Assemble the notes into a story, and iterate as needed
- Only when happy with the story, go design slides

3. Reduce Visual Load by Moving Text Off-Screen and Narrating the Content

People understand a multimedia explanation better when the words are presented as narration rather than on-screen text (the "Modality Effect").

"Presenting instructional materials using a combination of an *auditory mode for textual information*, such as spoken text, and a *visual mode for graphical information*, such as illustrations, charts, animations, etc., is more effective than presenting all information in a visual format, such as printed text with illustrations, charts or animations."

Ginns, P. (2005). Meta-analysis of the modality effect. Learning and instruction, 15(4), 313-331.

4. Use Visuals With Your Words, Instead of Words Alone

People learn better from words and pictures than from words alone (the "Multimedia Principle")

"Students remember more if instructors speak to images on a slide, rather than images and redundant text (i.e., bullet points that reiterate what the speaker is discussing)."

Schmaltz, R. M., & Enström, R. (2014). Death to weak PowerPoint: strategies to create effective visual presentations. Frontiers in Psychology, 5, 1138.

4. Use Visuals With Your Words, Instead of Words Alone

People learn better from words and pictures than from words alone (the "Multimedia Principle")

"Students remember more if instructors speak to images on a slide, rather than images and redundant text (i.e., bullet points that reiterate what the speaker is discussing)."

BUT: "Students performed worse on recall and recognition tasks and had greater dislike for slides with pictures that were not relevant."

Schmaltz, R. M., & Enström, R. (2014). Death to weak PowerPoint: strategies to create effective visual presentations. Frontiers in Psychology, 5, 1138.

4. Use Visuals With Your Words, Instead of Words Alone

People learn better from words and pictures than from words alone (the "Multimedia Principle")

"Students remember more if instructors speak to images on a slide, rather than images and redundant text (i.e., bullet points that reiterate what the speaker is discussing)."

BUT: "Accuracy in the Image Incongruent condition was significantly worse than the Text Based condition, despite participants rating [Text Based] as much less interesting."

Tangen, J. M., Constable, M. D., Durrant, E., Teeter, C., Beston, B. R., & Kim, J. A. (2011). The role of interest and images in slideware presentations. Computers & Education, 56(3), 865-872.

Pro Tip: for a More Polished Look, Crop Images to Circles

This slide looks basic!

h/t Ellen Vitercik

Pro Tip: for a More Polished Look, Crop Images to Circles

This slide looks polished!

h/t Ellen Vitercik

5. Keep It Simple – Needless Complexity Leads to Negative Evaluations

Will deliberately increasing the complexity of one's vocabulary give the impression of intelligence?

Experiments 1–3 manipulate complexity of texts and find a **negative relationship between complexity and judged intelligence -- that is, the presenters are perceived as less intelligent**.

The negative impact of complexity is mediated by processing fluency. Experiment 4 directly manipulated fluency and found that **texts in hard to read fonts are judged to come from less intelligent authors**.

Oppenheimer, D. M. (2006). Consequences of erudite vernacular utilized irrespective of necessity: Problems with using long words needlessly. Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition, 20(2), 139-156.

Fonts Don't Matter Much, as Long as They're Not Too Exotic

Serif	Example
Garamond	The quick fox
Times New Roman	The quick fox
Souvenir Lt	The quick fox
Bookman Old Style	The quick fox
Lubalin Graph Bk	The quick fox
Sans serif	**Example**
Gill Sans	The quick fox
Futura Bk	The quick fox
Arial	The quick fox
Tahoma	The quick fox
Verdana	The quick fox

Figure 2. The 10 fonts used in this study (size = 24-point).

One study shows **no significant difference between sans serif and serif** fonts on three variables: comfortable-to-read, attractive, and interesting.

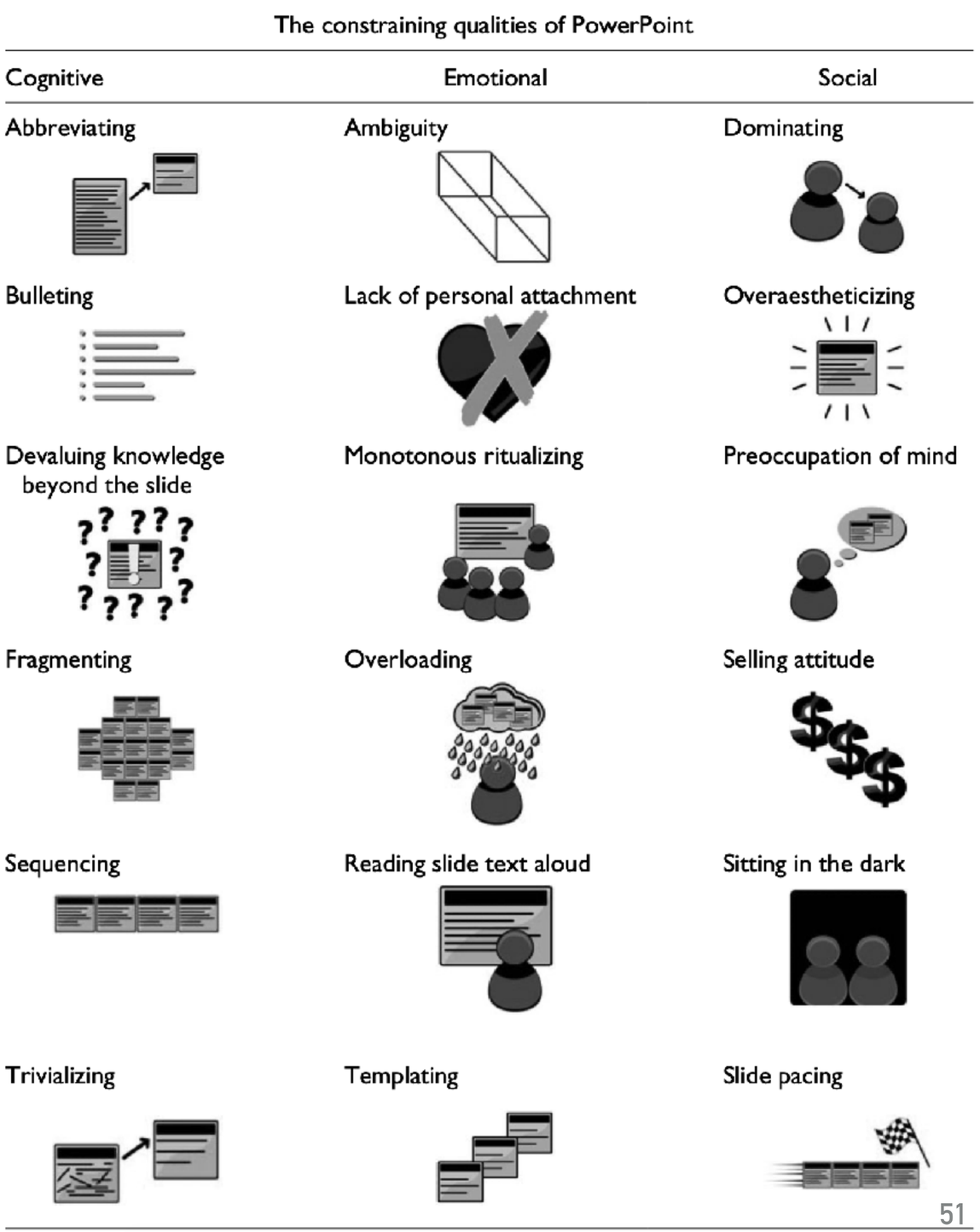
Mackiewicz, J. (2007). Audience perceptions of fonts in projected PowerPoint text slides. Technical communication, 54(3), 295-307.

More Information: 18 Major Pitfalls of PowerPoint

- **Cognitive**: abbreviating, bulleting, devaluing knowledge beyond the slide, fragmenting, sequencing, and trivializing
- **Emotional**: ambiguity, lack of personal attachment, monotonous ritualizing, overloading, reading slide text aloud, and templating
- **Social**: dominating, overaestheticizing, preoccupation of mind, selling attitude, sitting in the dark, and slide pacing

See paper for full details.

Kernbach, S., Bresciani, S., & Eppler, M. J. (2015). Slip-sliding-away: A review of the literature on the constraining qualities of PowerPoint. *Business and Professional Communication Quarterly*, *78*(3), 292-313.



More Information: the Pitfalls of Visual Representations in General (See Paper)

Table 2. List of Visual Representation Pitfalls With Brief Explanations.

Disadvantage	Author(s)	Description
Cognitive: Encoding		
Ambiguity	(Eppler & Burkhard, 2005; Tufte, 2007)	Visual notations may contain unlabeled symbols that may be ambiguous and thus difficult to interpret.
Breaking conventions	(Ware, 2004)	A visualization may use different visual rules or symbols than normally expected.
Confusion	(Eppler & Burkhard, 2005; Few, 2006)	Visualizations that do not have a clear overall logic or accompanying text may confuse the viewers.
Cost to make explicit	(Larkin & Simon, 1987)	"Diagrammatic representations typically display information that is only implicit in sentential representations and that therefore has to be computed, sometimes at great cost, to make it explicit for use" (Larkin & Simon, 1987).
Cryptic encoding	(Tufte, 1986)	The visual format used to represent data may not be universally understandable and confuse some audiences.
Defocused	(Few, 2006; Kosslyn, 2006; Tufte, 1986; Ware, 2004)	Visualization may distract a person from the main goal he or she tries to achieve or emphasize, at the same time, using multiple items.
Hiding/obscuring	(Few, 2006; Kosslyn, 2006; Tufte, 1986; Wainer, 1984)	A visualization may hide important insights contained in data by the way that data are represented graphically (e.g., covarying height and width, changing the starting point, or varying the aspect ratio, etc.).
Inconsistency	(Cawthon & Vande Moere, 2007; Tufte, 1986)	A visualization may make inconsistent use of certain symbols, for example, changing their function or meaning without signaling this change.
Low accuracy	(Few, 2006; Kosslyn, 2006; Tufte, 1986; Wainer, 1984)	Visualization generally depicts information less precisely than number and tables.
Misleading/Distorting	(Tufte, 1986; van Wijk, 2006; Wainer, 1984)	Some visualizations are drawn in a way that may lead to incorrect conclusions.
Misuse of figure ground	(Tufte, 1986)	The figure ground and layer contrasts are not illustrated properly.
Not respected gestalt principles	(Tufte, 1986)	Some visualizations do not group related information (proximity principle) or do not represent the same kind of information with the same symbols (similarity principle).
Over-determinism	(Shimojima, 1996)	A visualization is, by its nature, inherently more specific than text in depicting concepts and relations.
Over/under-reliability appearance	(Crilly, Blackwell, & Clarkson, 2006; Green & Petre, 1996; Henderson, 1995; Whyte, Ewenstein, Hales, & Tidd, 2007)	Highly polished visualizations might prevent users from criticizing the content whereas more provisional sketches encourage suggested revisions.
Over-complexity	(Few, 2006; Kosslyn, 2006; Tversky, 2005)	The visualization depicts elements in a more complex manner than necessary.
Over-simplification	(Eppler & Burkhard, 2005; Nicolini, 2007)	Some graphic depictions leave out essential elements to simplify information, which leads to a distortion of the information.
Redundancy	(Few, 2006; Tufte, 1986)	In some graphic representations of information, the information is visualized in superfluous ways that clutter the visualization unnecessarily.
Task-visualization fit	(Al-Kassab, Ouertani, Schiuma, & Neely, 2014)	The lack of an appropriate fit between the task and the visual representation can be misleading.
Technology/template driven	(Few, 2006; Tufte, 1986)	Some visualizations are based on pre-defined forms or templates that are not adequate for the communication task at hand or the information to be represented.
Time-consuming to produce	(van Wijk, 2006)	Producing a visualization may take a disproportional amount of time for the information that is communicated.
Unclear	(Cawthon & Vande Moere, 2007)	A graphic depiction may leave too much room for interpretation regarding its purpose or main message.
Unevenness	(Blackwell, 2001)	A visualization can typically not be used in many different ways. It may privilege some activities while making others harder, thus constraining users' thoughts in one direction.
Cognitive: Decoding		
Change blindness	(Ware, 2004)	Important changes in pictures may go unnoticed by the viewers.
Channel thinking	(Mengis, 2007)	The visualization can direct thinking in an inappropriate direction (caused by a metaphor or familiarity level).
Depending on perceptual skills	(Nisbett, 2003; Tufte, 1986; van Wijk, 2006)	People see differently, depending on physical (e.g., color blindness) and cultural factors (attention to foreground or background).
Difficult to understand	(Buergi & Roos, 2003; Cawthon & Vande Moere, 2007)	Some visualizations are inherently difficult to understand because they depict many complex relationships that may not be optimally represented.
Focus on low relevance items	(Lurie & Mason, 2007)	Visual representations may accentuate biases in decision making by increasing attention to particular attributes or less diagnostic information.
High requirement on training and resources	(Chen, 2005; van Wijk, 2006)	The use of certain images or visual applications requires extensive training and support.
Knowledge of visual conventions	(Avgerinou, 2007; Knox, 2007; Machin & Van Leeuwen, 2007)	Knowing the visual conventions (e.g., reading from left to right or in a clockwise direction) is a learned skill, not a natural ability.
Misuse	(Eppler & Burkhard, 2005)	A visualization may be used for a purpose for which it was not intended or adequate.
Overload	(Eppler & Burkhard, 2005; Eppler et al., 2006; Tufte, 1997; Ware, 2004)	Some graphic depictions overload the senses of a viewer by presenting too many visual elements at the same time.
Reification	(Whyte et al., 2007)	Tendency to consider an abstract concept as concrete, for example, attributing properties of a material object to that concept.
Wrong salience	(Al-Kassab et al., 2014; Few, 2006; Green & Petre, 1996; Mengis, 2007; Ware, 2004)	The user concentrates on the wrong issue, for example, on the tool or on the visual appearance, instead of on the task.
Emotional: Encoding		
Disturbing	(Cawthon & Vande Moere, 2007; Tufte, 1990)	Some images may cause emotional harm to the viewer because of their shocking or repellent content.
Boring	(Cawthon & Vande Moere, 2007)	Some graphic representations are perceived as uninteresting and do not help to focus attention for an appropriate amount of time.
Ugly/unappealing	(Cawthon & Vande Moere, 2007)	Some graphic representations may reduce the motivation to explore them in spite of their informative content due to a sub-optimal, non-aesthetic form.
Wrong use of color	(Few, 2006; Tufte, 1986; Wainer, 1984; Ware, 2004)	The inadequate use of colors or their combinations may make an image confusing or unappealing.
Emotional: Decoding		
Visual stress	(Ware, 2004)	Some kind of patterns (striped or flickering) may cause illness in the viewer.
Personal likes and dislikes	(Tversky, 2005)	Some visualizations may get more attention than others, not because of their importance, but because they fit the cognitive preferences of a particular viewer.
Prior knowledge and experience	(Al-Kassab et al., 2014; Avgerinou & Pettersson, 2011; Chen, 2005; Dwyer, 1972)	Previous domain knowledge on how to interpret the content and positive or negative experience with a specific visualization influences the willingness of people to use it.
Social: Encoding		
Affordance conflict	(Nicolini, 2007)	A visualization may signal the wrong kind of required (inter-)activity to its viewers.
Hierarchy, exercise of power	(Ewenstein & Whyte, 2007; Henderson, 1995; Nicolini, 2007; Whyte et al., 2007)	The political use of images in collaborative settings by certain people may result in unequal possibilities to contribute (e.g., through manipulative use of visualization provisionality, facilitator choice, sequence of contributions, etc.).
Inhibit conversation	(Nicolini, 2007; Oliver, 2007)	Having one's contributions visualized (for example, in a group context) may lead to participants being less outspoken about certain issues.
Rhythm of freezing and unfreezing	(Whyte et al., 2007)	A visualization may make a certain viewpoint or idea too rigorous and too fixed, thus not leaving adequate conditions to invent alternative views or options.
Turn taking alteration	(Eppler, 2004)	Using a graphic representation to guide a team conversation can affect the natural turn taking within a group in favor of those who can directly change that visualization.
Unequal participation	(Mengis, 2007)	The use of visualizations in group contexts may lead to unequal participation on behalf of the participants.
Social: Decoding		
Altered behavior	(Eppler et al., 2006; Mengis, 2007; Nicolini, 2007)	The use of visuals in group interaction may affect the typical behavior of the user.
Cultural and cross-cultural differences	(Al-Kassab et al., 2014; Avgerinou & Pettersson, 2011; Bresciani, 2014; Ewenstein & Whyte, 2007; Forsythe, 2011; Henderson, 1995; Nisbett, 2003; Segall, Campbell, & Herskovits, 1966; Ware, 2004)	The meaning of symbols and colors is not universal, and hence, some graphic representations may be misinterpreted in other cultural contexts.
Defocused from non-verbal interaction	(DeSanctis & Gallupe, 1987)	A group's focus on a central visualization on a board or screen can take away the participants' attention from their body language and gestures, which give important information on how to interpret verbal contributions.
Different perspectives	(Heer & Agrawala, 2008)	Different people look at issues from different points of view (e.g., people from different organizational levels).
Hiding differences of opinion	(Eppler et al., 2006)	The use of one visualization in a group context may hide individual differences of opinion because of the need to find one common representation.
Recency effect	(Nisbett, 2003; Tufte, 1986)	The meaning of a visualization is not interpreted in a vacuum but as part of a broader context that depends on user's previous exposure.
Time-consuming to agree upon	(DeSanctis & Gallupe, 1987)	Group discussion based on visualization requires more time than verbal discussion.

Bresciani, S., & Eppler, M. J. (2015). The pitfalls of visual representations: A review and classification of common errors made while designing and interpreting visualizations. *Sage Open*, *5*(4), 2158244015611451.

More Resources

- Tips for Giving Clear Talks by Kayvon Fatahalian
  - http://graphics.stanford.edu/~kayvonf/misc/cleartalktips.pdf

- Presenting Research: Structure, Story, and Support by Bogdan Vasilescu
  - https://drive.google.com/file/d/1f2iscUS5NaeiMMTp68PIkE3vm_1FeMBM/view?usp=sharing

- CMU Global Communications Center's guide
  - https://www.cmu.edu/gcc/handouts-and-resources/powerpoint-design.html

Part III: Claire Le Goues' "Things I Keep Repeating About Writing"

https://clairelegoues.com/2016/08/23/things-i-keep-repeating-about-writing/

Use Clear and Precise Language

- Use short, declarative, active sentences. BANISH THE PASSIVE VOICE!
- Use adverbs and pronouns judiciously:
  - Adverbs are often imprecise: what does “incredibly” add to the phrase “incredibly important” that the word “important” lacked on its own? How much more important than important is something that is incredibly important?
- Pronouns are often unclear with respect to their antecedents, which can confuse the reader.

Use Clear and Precise Language

- Be as explicit/concrete in your statements as you can.
  - Instead of “The dataset has a few attributes.”, say “The dataset has 22 attributes.”
  - Instead of “We performed a number of experiments.” or “The cat had a number of lives.”, try “We performed four experiments.”, “The cat had nine lives.”
- (To highlight the point, consider the sentence(s) without “a number of”: “We performed experiments.”/“The cat had lives.” See how the meaning didn’t really change?)

Do Not Use More Syllables Than Necessary

- Two easy manifestations of this rule are the following transformations that can be applied universally to your draft:
  - “In order to” –> “To”
  - “Utilize” –> “Use”
- The point of writing is to communicate an idea. Using more syllables than necessary obscures the idea without adding meaning.

Present Numbers Properly

- Write out in letters all positive numbers less than or equal to 10, unless they are in a sentence with a number greater than 10.
  - No: “We analyze 2 datasets”
  - Yes: “We analyze two datasets”

  - No: “We interviewed two designers and 12 users”
  - Yes: “We interviewed 2 designers and 12 users”

Right Justify Columns of Numbers

- I will repeat this in all-caps, because I really mean it: RIGHT JUSTIFY COLUMNS OF NUMBERS.
- Ensure that the correct number of significant digits are used and that decimal points align.
- You will argue with me about this, because you really want to left-justify or center them. I don't know why. A reader should be able to quickly scan a column of numbers to get a sense of magnitude, and cannot do that if they are left-justified unless they are all (coincidentally) the same order of magnitude.
- Text in columns should be left justified. Never center anything that's not a column header.

Typesetting/Copy-Editing Minutiae

- Capitalize Table, Figure, and Section.
- Refer to sections only, never subsections, even when you're referencing an actual subsection
  - No: Subsection 4.1
  - Yes: Section 4.1
- Include a non breaking space (~) between the words Figure/Section/etc and the \ref.
- Capitalize and punctuate section/paragraph headings/captions consistently. If one ends with a period, they all should.

Typesetting/Copy-Editing Minutiae

- Do not use citations as nouns.
  - No: “In [14], Hazelwood et al. describe facts.”
  - Yes: “Hazelwood et al. [14] describe facts.”
- Citations go before punctuation, with a non-breaking space between the word and the citation. Footnotes go after the punctuation, with no space.
- Always put a comma after i.e. and e.g., and use them properly (i.e. means “put differently” or “in other words”, e.g. means “for example”).

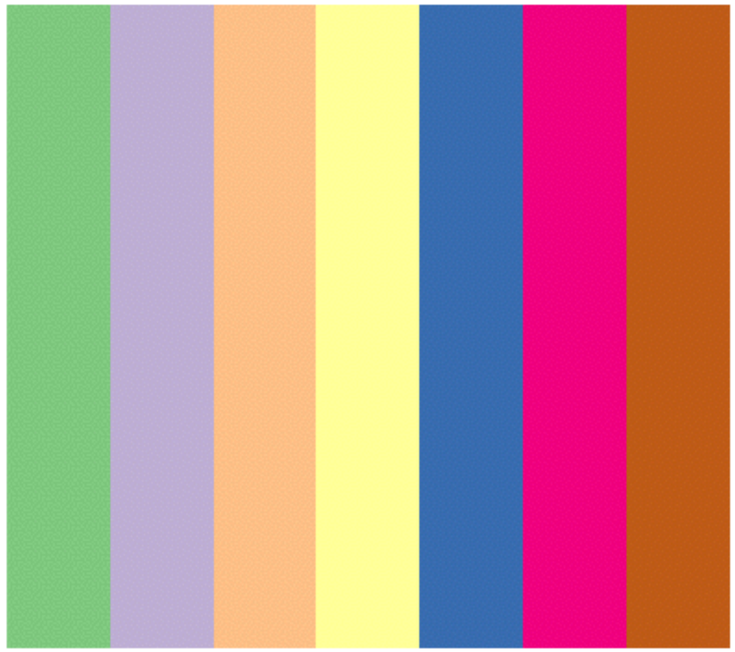
Typesetting/Copy-Editing Minutiae

- Use the Oxford/serial comma!
  - No: I love my parents, Lady Gaga and Humpty Dumpty.
  - Yes: I love my parents, Lady Gaga, and Humpty Dumpty.

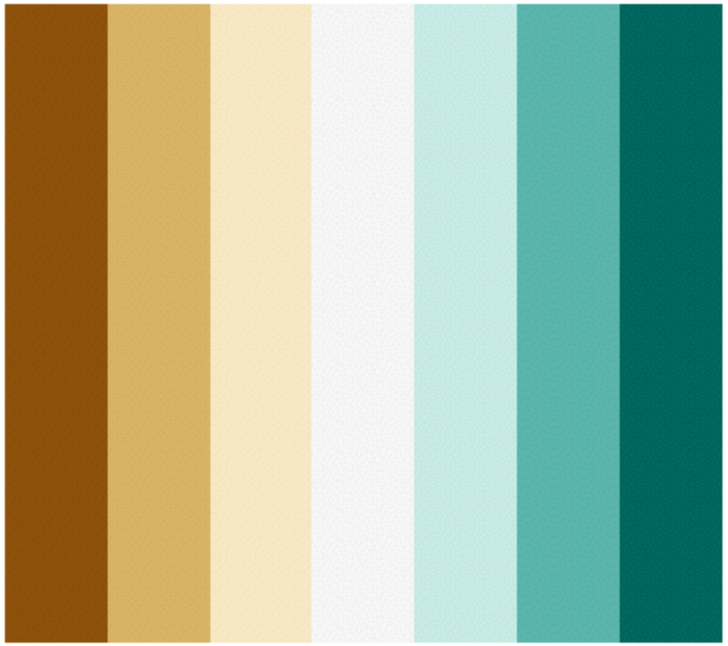
Colors

- Choose colors for graphs and figures that:
  - show up when your paper is printed in greyscale.
  - are “colorblind safe.”
- Go to http://colorbrewer2.org/ and choose “colorblind safe” and “print friendly” to find color combinations that work.

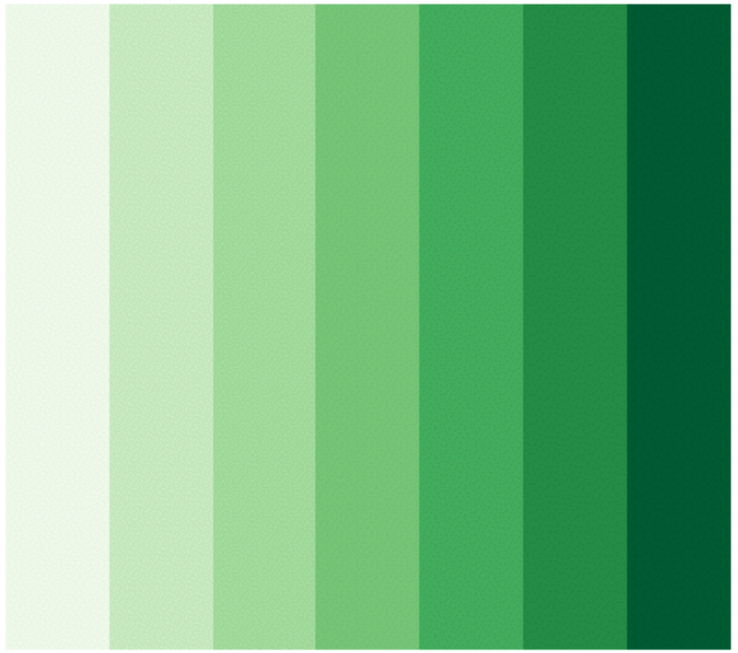
Color Palette Examples



Accent (qualitative)



BrBG (divergent)



Greens (sequential)

Bibtex

- Give your bibtex entries reasonably indicative names.
- If you cut and paste it from the web somewhere, ensure that it's done properly (some sites make everything a @misc, which is almost always wrong) and modify the bibtex so that it's reasonable.
- Definition of reasonable:
  - special characters are copied properly;
  - authors names and title are spelled/capitalized correctly;
  - includes venue, preferably spelled out along with its acronym, but you can drop the "Proceedings of the 23rd Annual ACM/IEEE blah blah" in favor of just "International Conference on Software Engineering";
  - include year and page numbers.

Credits

- Graphics: Dave DiCello photography (cover)
- Many slides from Atkinson, C., & Mayer, R. E. (2004). Five ways to reduce PowerPoint overload. Creative Commons, 1.
- Tufte, E. (2016). The visual display of quantitative information (1983). In Diagrammatik-Reader (pp. 219-230). De Gruyter (A).
- Tufte, E. R. (2006). Beautiful evidence (Vol. 1). Cheshire, CT: Graphics Press.