Visualization for Data Science
CMPT 733

Steven Bergner

sbergner@cs.sfu.ca

Outline

- Visualization: What, Why, and How?
- Motivational example
- Design principles

Defining Visualization (Vis)

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

[“Visualization Analysis and Design” by T. Munzner, 2014]

Defining Visualization (Vis)

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

[“Visualization Analysis and Design” by T. Munzner, 2014]

Why have a human in the loop?

Defining Visualization (Vis)

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

["Visualization Analysis and Design" by T. Munzner, 2014]

Why have a human in the loop?

- Not needed when automatic solution is trusted

Defining Visualization (Vis)

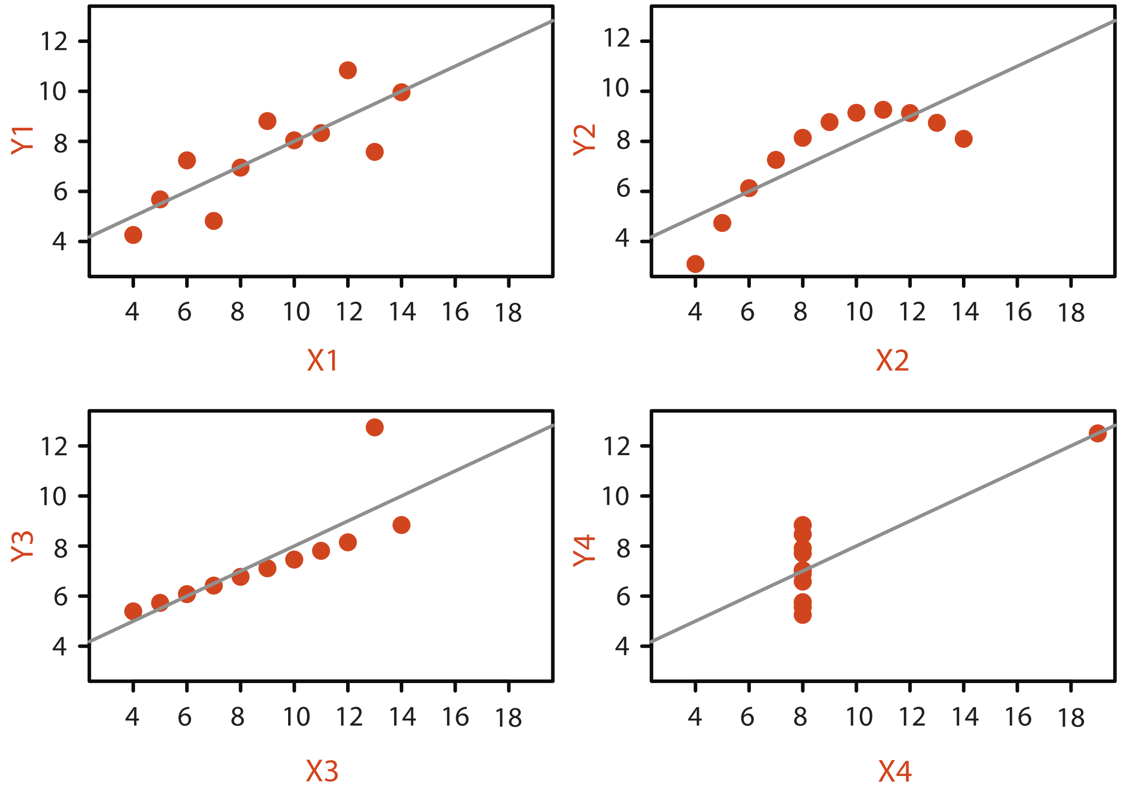
Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

["Visualization Analysis and Design" by T. Munzner, 2014]

Why have a human in the loop?

- Not needed when automatic solution is trusted
- Good for ill-specified anlaysis problems
  - Common setting: "What questions can we ask?"

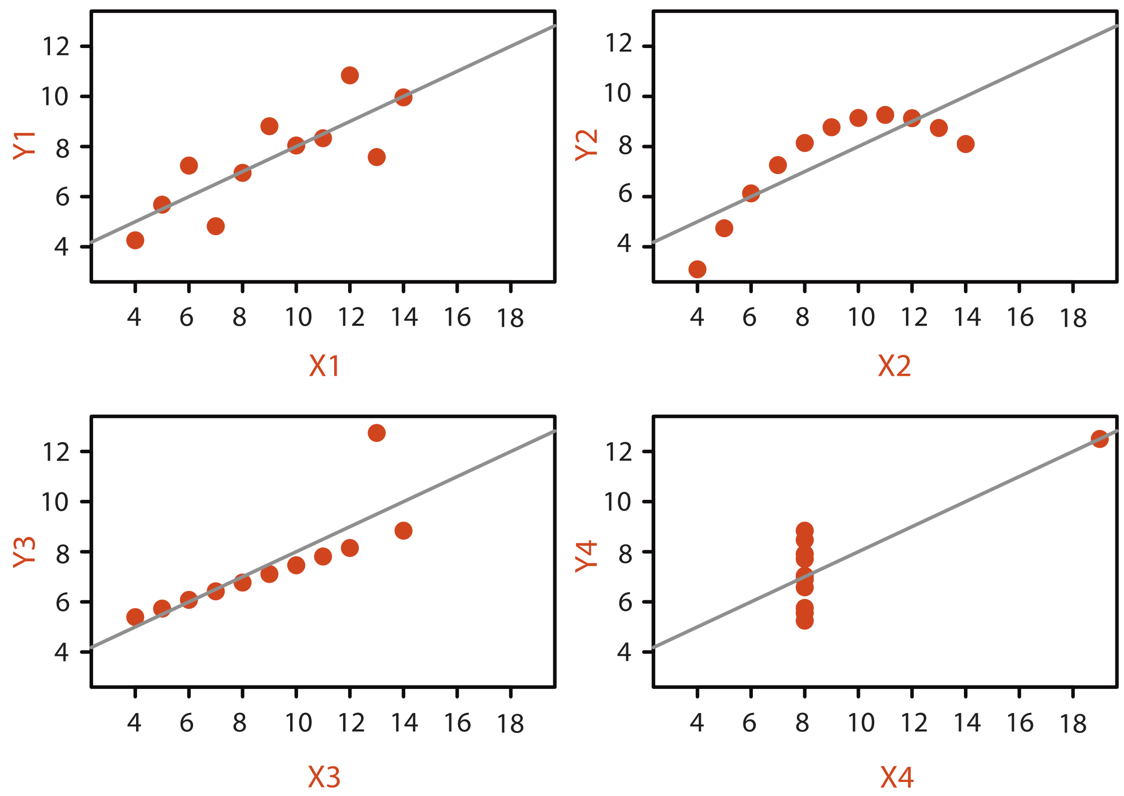
“Numerical calculations are exact, but Graphs are rough”



- Same relationship among each pair of variables?

[F. J. Anscombe, 1973]

"Numerical calculations are exact, but Graphs are rough"



- Same relationship among each pair of variables?
- Identical statistics

X mean	9
X variance	10
Y mean	7.5
Y variance	3.75
<X,Y> correlation	0.816

[F. J. Anscombe, 1973]

Why have a human in the loop?

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

Munzner, T. (2014)

Why have a human in the loop?

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

Munzner, T. (2014)

Long-term use

- Exploratory analysis of scientific data
- Presentation of known results

Why have a human in the loop?

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

Munzner, T. (2014)

Long-term use

- Exploratory analysis of scientific data
- Presentation of known results

Short-term use

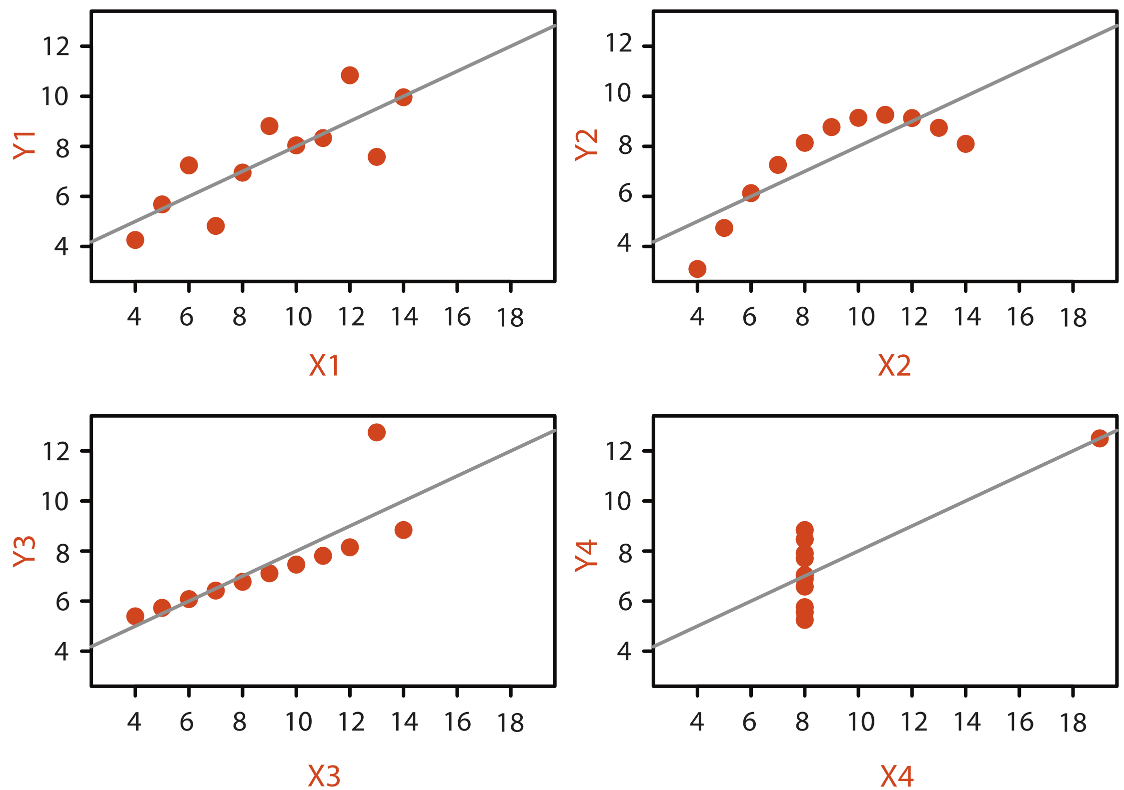
- For **developers** of automatic solutions:
  - Understand requirements for model development
  - Refine/debug and determine parameters
- For **end users** of automatic solutions: verify, build trust

Why use an external representation?

	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
SUM	99,00	82,51	99,00	82,51	99,00	82,50	99,00	82,51
AVG	9,00	7,50	9,00	7,50	9,00	7,50	9,00	7,50
STDEV	3,32	2,03	3,32	2,03	3,32	2,03	3,32	2,03

[F. J. Anscombe, 1973]

Why use an external representation?

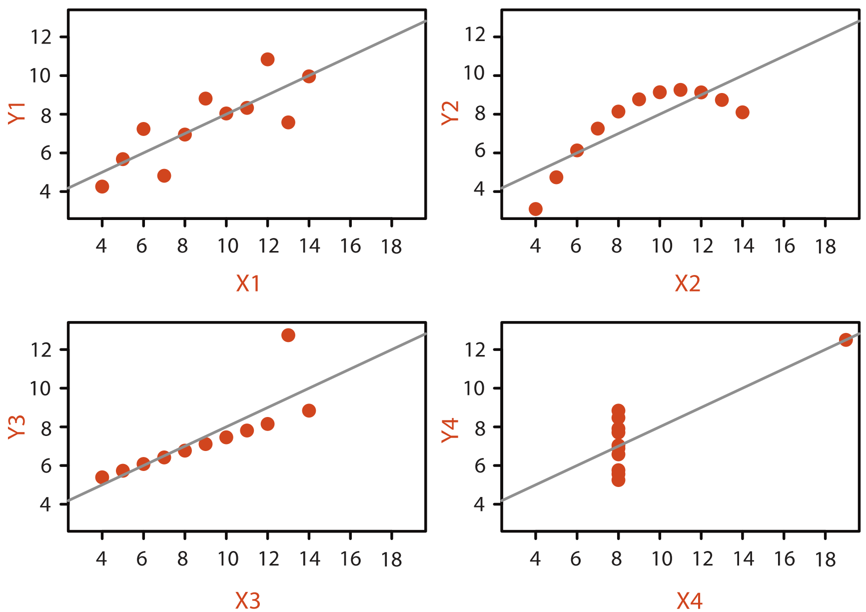


- Replace cognition with perception

	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
SUM	99,00	82,51	99,00	82,51	99,00	82,50	99,00	82,51
AVG	9,00	7,50	9,00	7,50	9,00	7,50	9,00	7,50
STDEV	3,32	2,03	3,32	2,03	3,32	2,03	3,32	2,03

[F. J. Anscombe, 1973]

Why represent all the data?



	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
SUM	99,00	82,51	99,00	82,51	99,00	82,50	99,00	82,51
AVG	9,00	7,50	9,00	7,50	9,00	7,50	9,00	7,50
STDEV	3,32	2,03	3,32	2,03	3,32	2,03	3,32	2,03

- Summaries lose information, details matter
  - Confirm expected and find unexpected patterns
  - Assess validity of statistical model

Analysis framework: four *levels*

[A Nested Model of Visualization Design and Validation.
Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

Analysis framework: four *levels*

- ***Domain*** situation: Who are the target users?

Analysis framework: four *levels*

- ***Domain*** situation: Who are the target users?
- ***Abstraction***: Translate from specifics of domain to vocabulary of vis

Analysis framework: four *levels*

- ***Domain*** situation: Who are the target users?
- ***Abstraction***: Translate from specifics of domain to vocabulary of vis
- **What** is shown? *Data abstraction*
  - Don't just draw what you're given: transform to new form

[A Nested Model of Visualization Design and Validation.
Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

Analysis framework: four *levels*

- ***Domain*** situation: Who are the target users?
- ***Abstraction***: Translate from specifics of domain to vocabulary of vis
- **What** is shown? *Data abstraction*
  - Don't just draw what you're given: transform to new form
- **Why** is the user looking at it? *Task abstraction*

[A Nested Model of Visualization Design and Validation.
Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

Analysis framework: four *levels*

- ***Domain*** situation: Who are the target users?
- ***Abstraction***: Translate from specifics of domain to vocabulary of vis
- **What** is shown? *Data abstraction*
  - Don’t just draw what you’re given: transform to new form
- **Why** is the user looking at it? *Task abstraction*
- **How** is it shown? ***Idiom***
  - Visual encoding idiom: How to draw
  - Interaction idiom: How to manipulate

[A Nested Model of Visualization Design and Validation.
Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

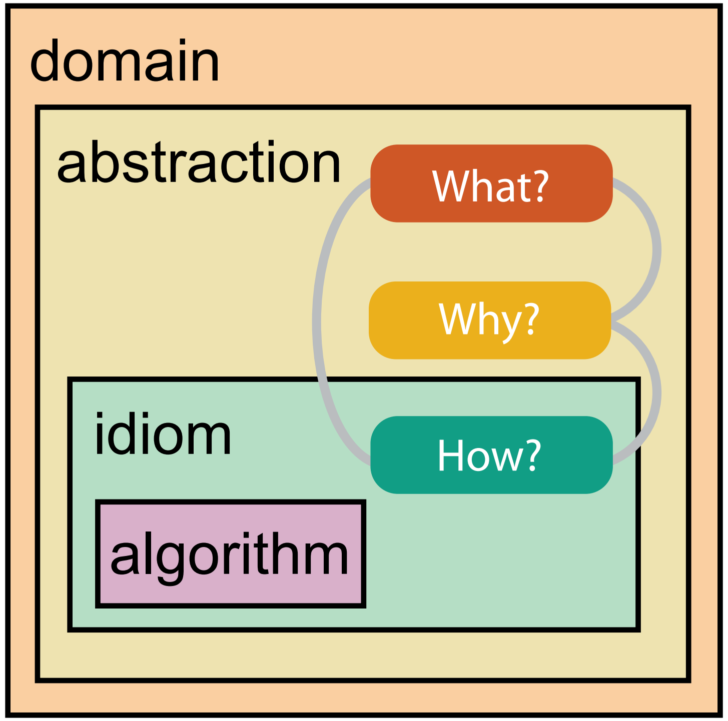
Analysis framework: four *levels*

- ***Domain*** situation: Who are the target users?
- ***Abstraction***: Translate from specifics of domain to vocabulary of vis
- **What** is shown? *Data abstraction*
  - Don't just draw what you're given: transform to new form
- **Why** is the user looking at it? *Task abstraction*
- **How** is it shown? ***Idiom***
  - Visual encoding idiom: How to draw
  - Interaction idiom: How to manipulate
- ***Algorithm***: efficient computation

[A Nested Model of Visualization Design and Validation.
Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

Analysis framework: four *levels*

- ***Domain*** situation: Who are the target users?
- ***Abstraction***: Translate from specifics of domain to vocabulary of vis
- **What** is shown? *Data abstraction*
  - Don't just draw what you're given: transform to new form
- **Why** is the user looking at it? *Task abstraction*
- **How** is it shown? ***Idiom***
  - Visual encoding idiom: How to draw
  - Interaction idiom: How to manipulate
- ***Algorithm***: efficient computation

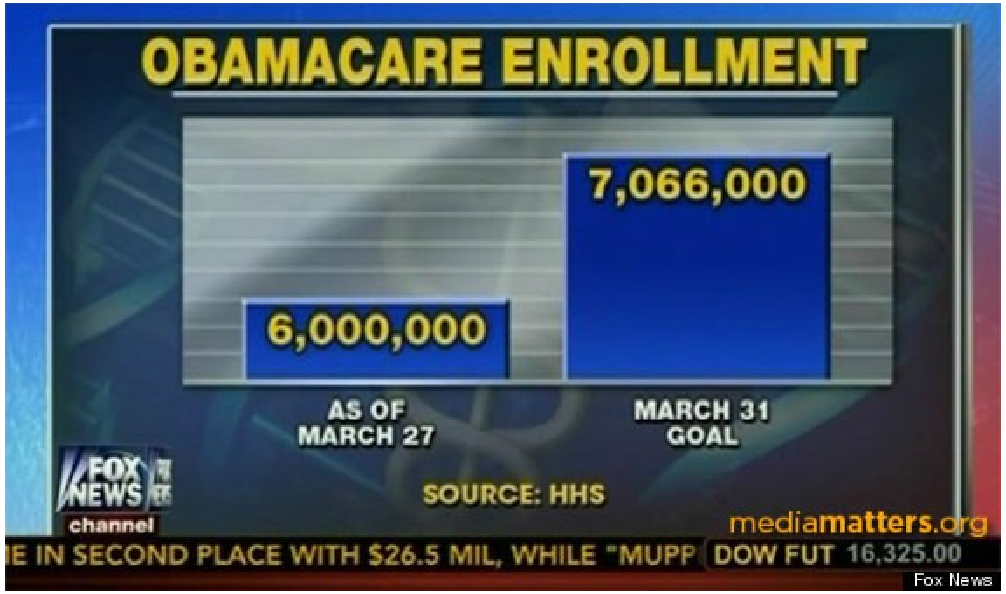


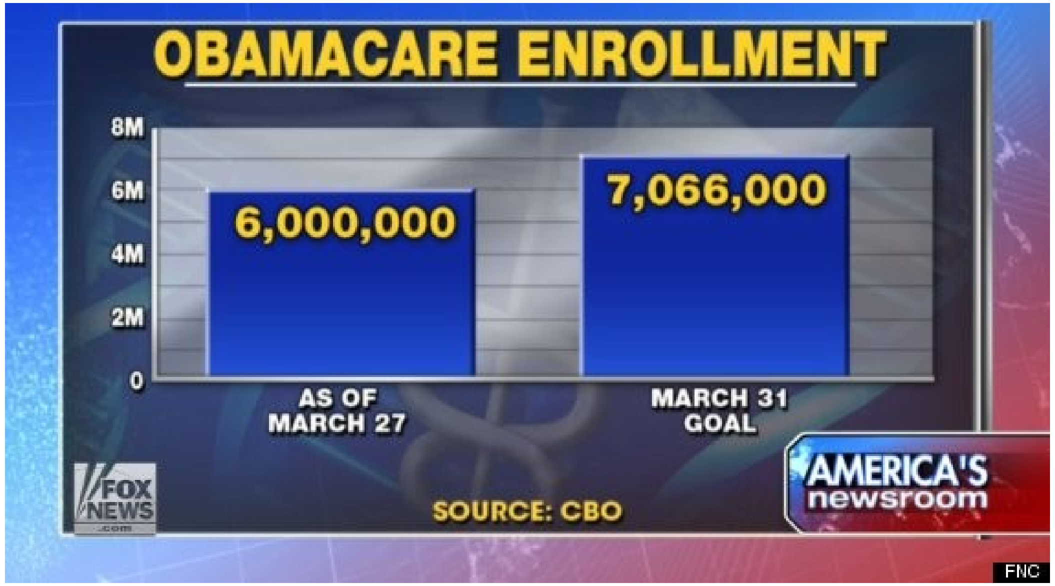
[A Nested Model of Visualization Design and Validation.
Munzner. IEEE TVCG 15(6):921-928, 2009 (Proc. InfoVis 2009).]

Examples

Motivation

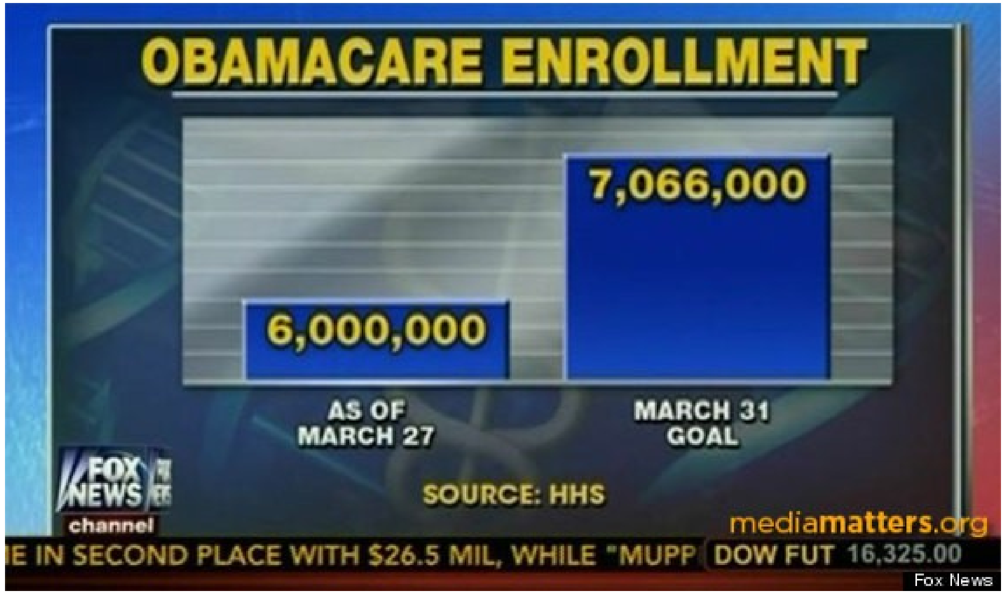
- WTF Visualizations (http://viz.wtf)

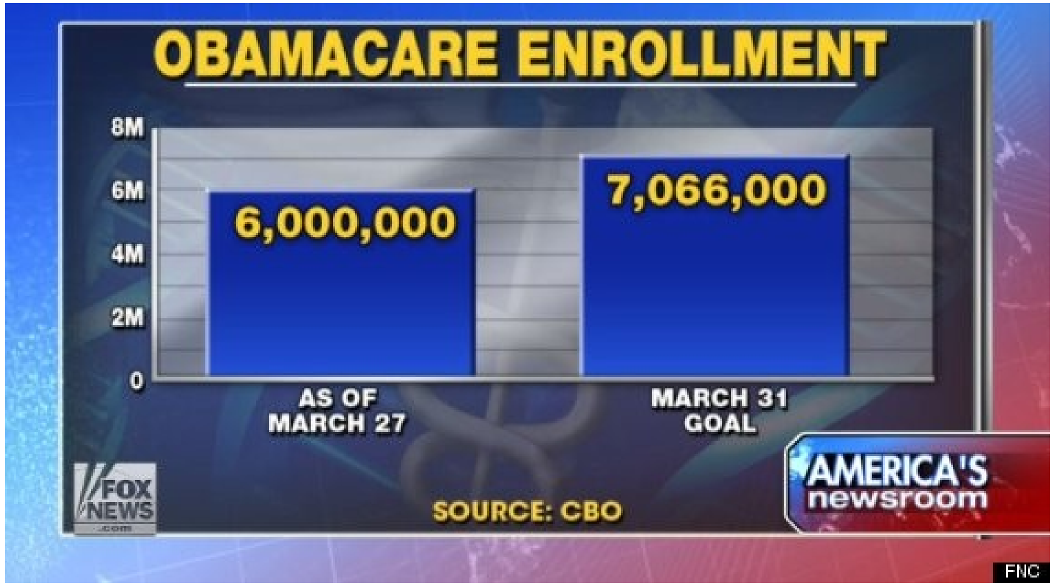




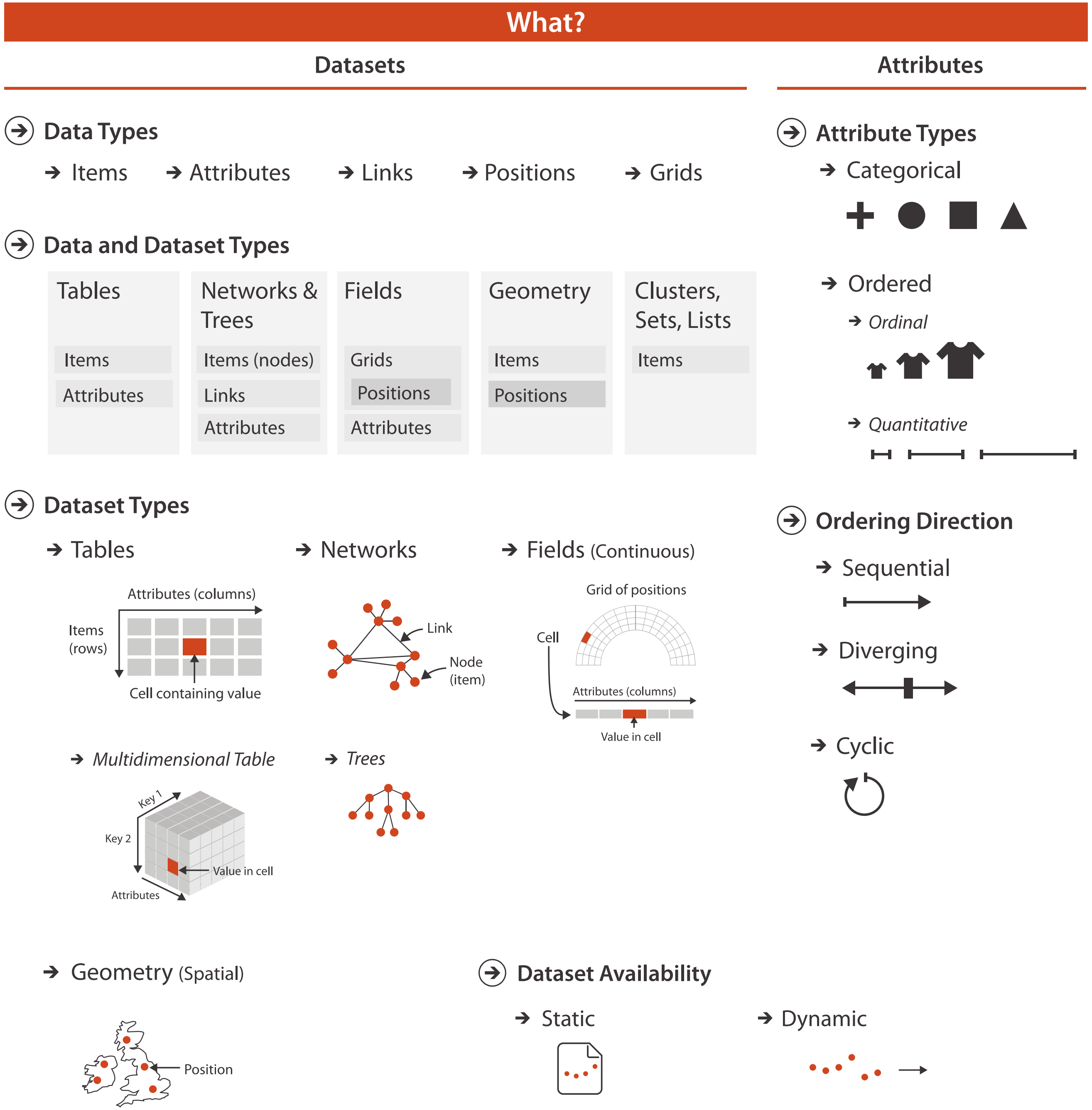
Motivation

- WTF Visualizations (http://viz.wtf)
- Without knowing the principles, you might make a lot of mistakes like this!





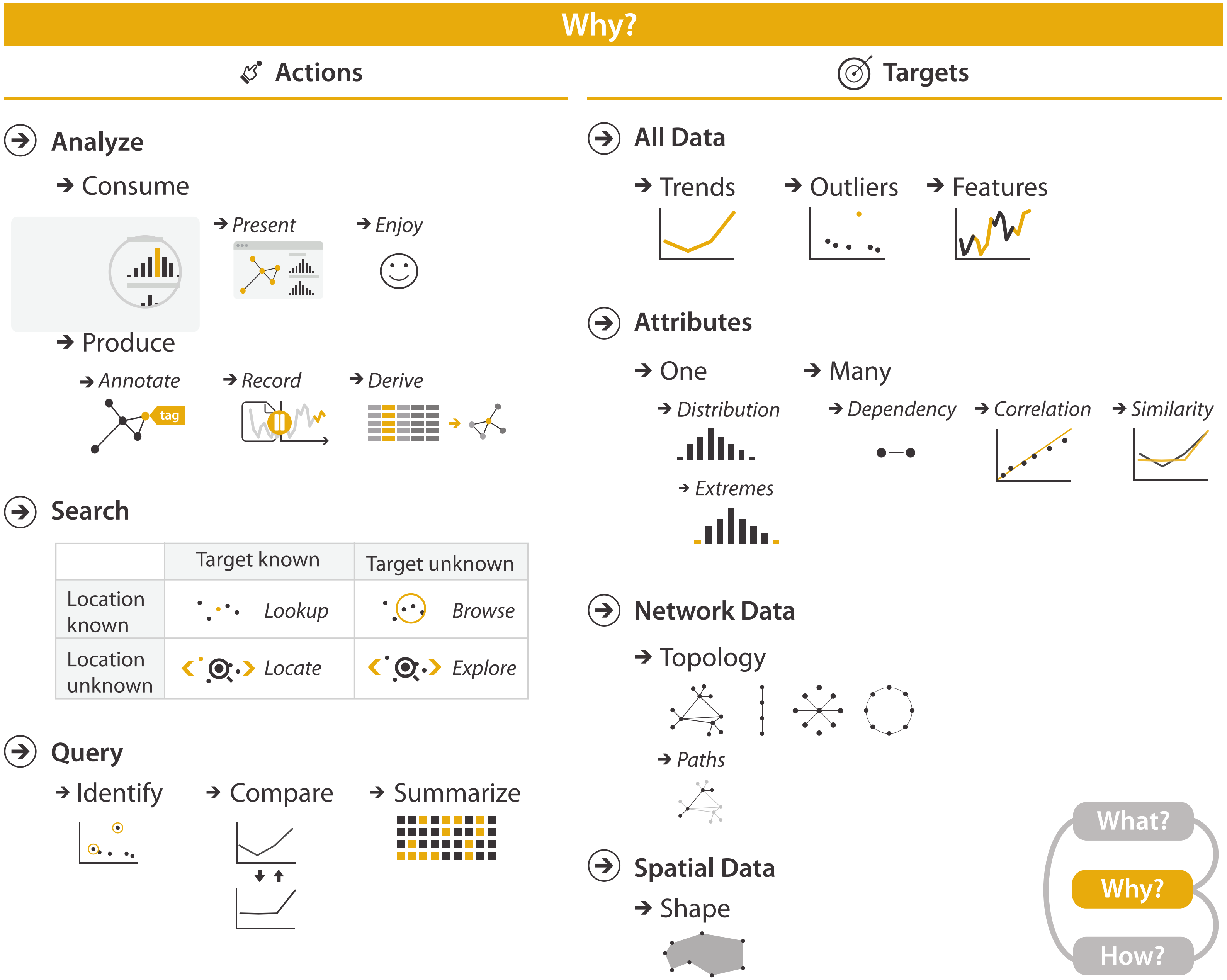
Understand
Data, Task, and Encoding



Data Types

- Items and attributes as rows and columns of tables
- Position and time are special attributes
- Spatial data on grids makes computation easier

[T. Munzner, 2014]



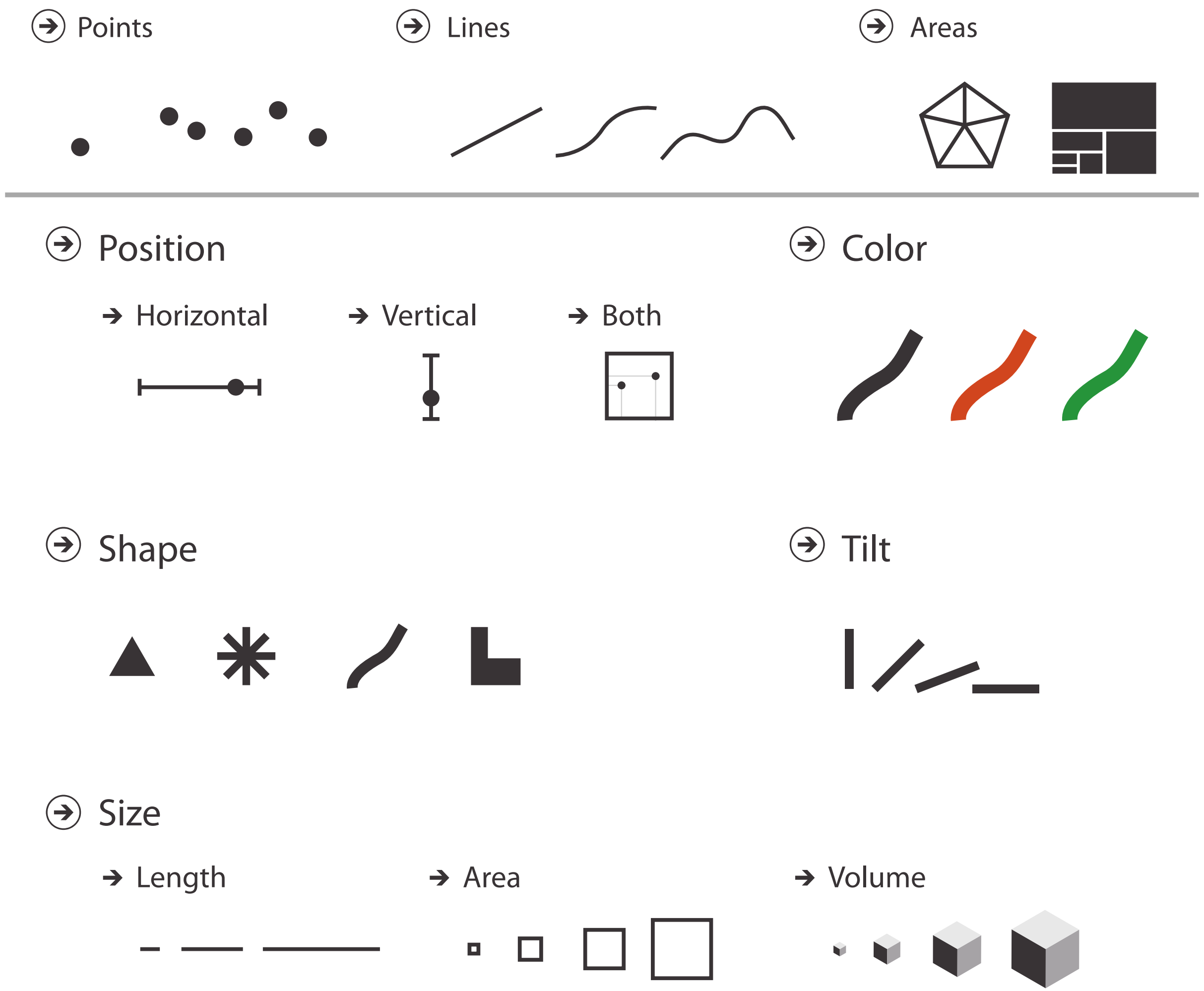
Tasks

- Actions
  - Analyze
  - Search
  - Query
- Targets
  - Item & Attributes
  - Topology & Shape

[T. Munzner, 2014]

Visual Encoding – How?

- Marks
  - Geometric primitives
- Channels
  - Appearance of marks
  - Redundant coding with multiple channels possible



[T. Munzner, 2014]

Design Principles
for Task Effective Visualization

Task and effectiveness

- Most idioms ineffective for particular task/data
  - Recast tasks from domain-specific vocabulary to abstract form
  - Systematic thinking about choices imposes structure on design space
  - Analyze existing as step to design new – iterate and compare
- What counts as effective?
  - Novel: enable entirely new kinds of analysis
  - Faster: speed up existing workflows

Resource limitations

Resource limitations

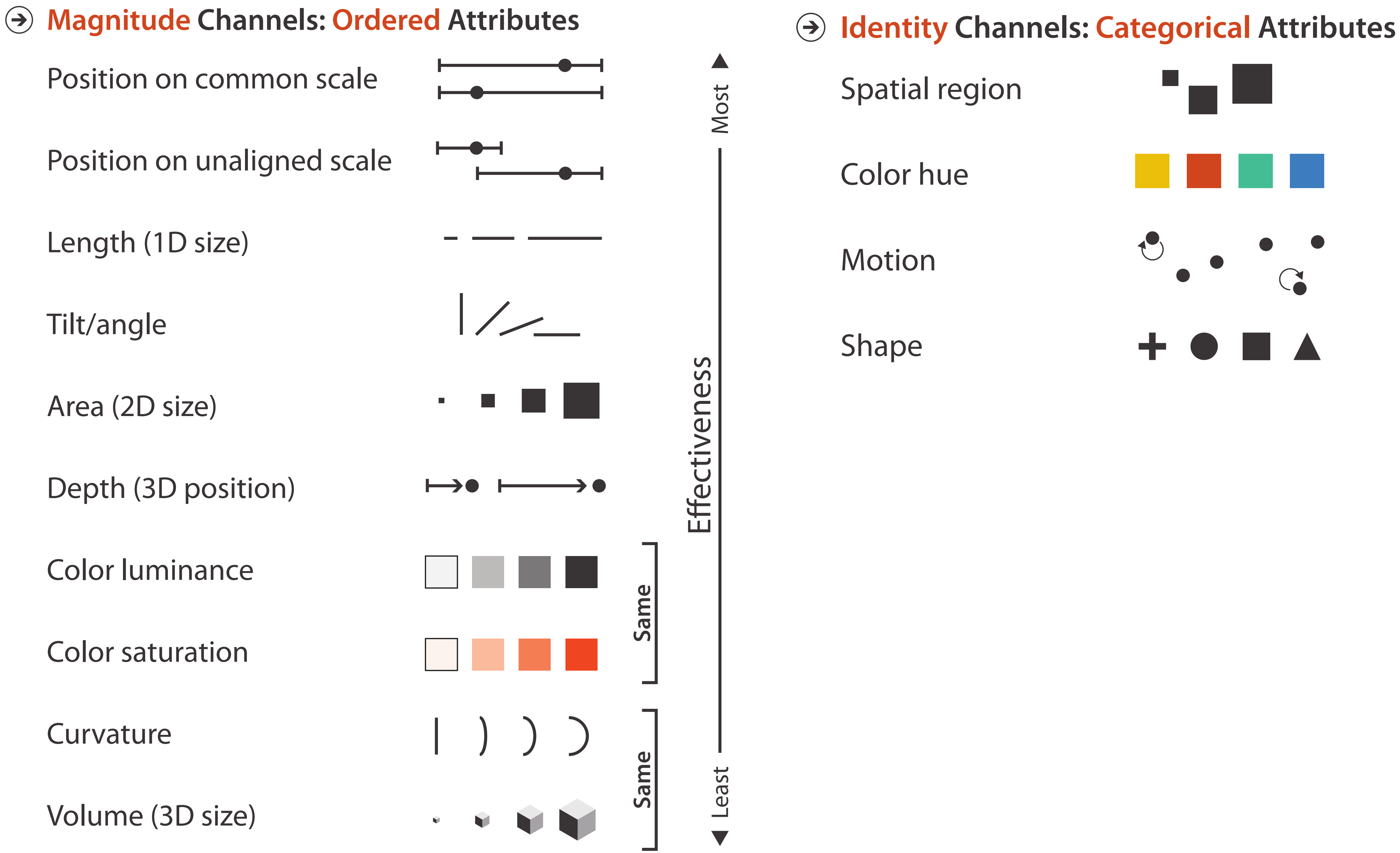
- **Computational** limits
  - Processing time and system memory

Resource limitations

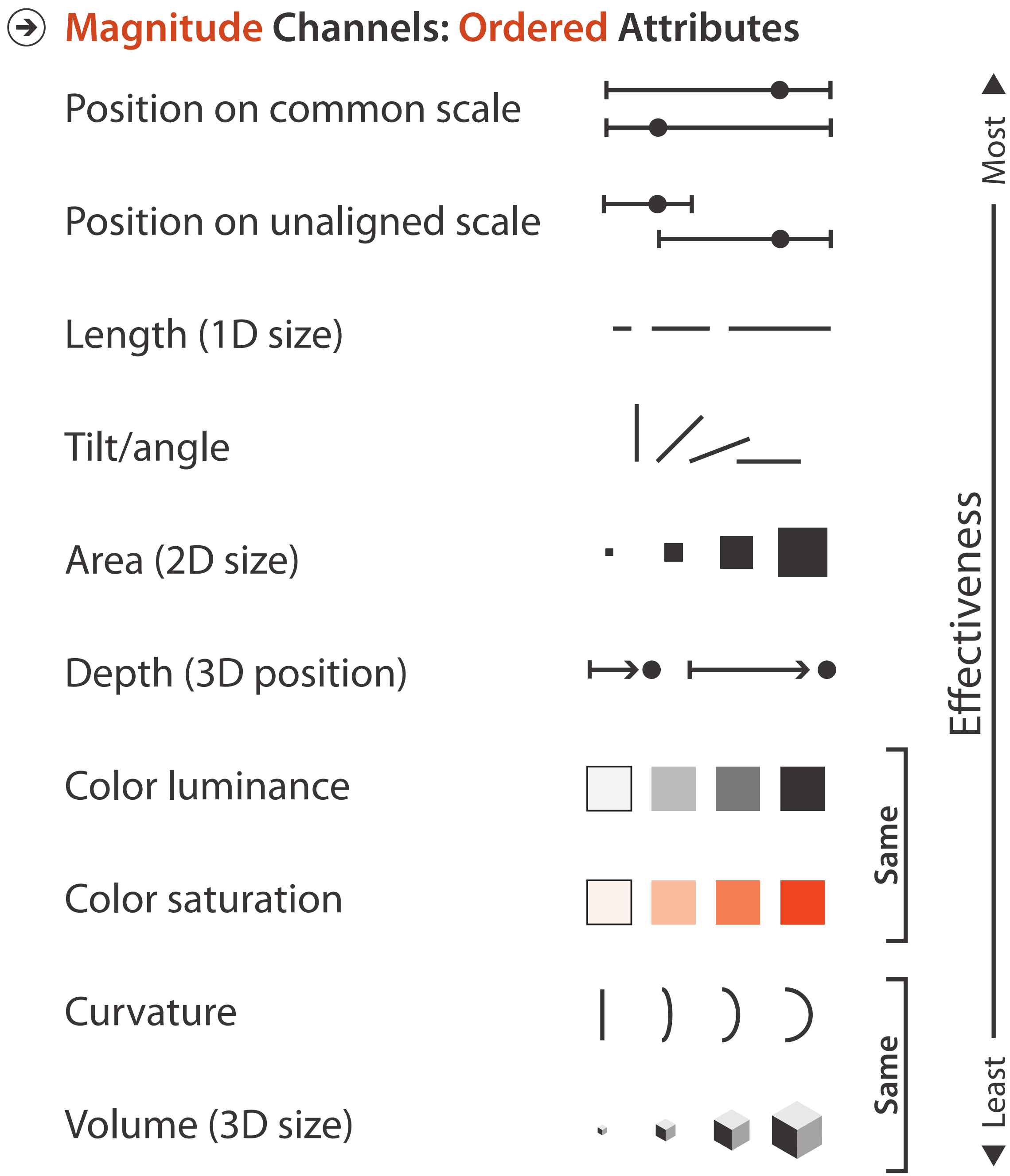
- **Computational** limits
  - Processing time and system memory
- **Human** limits
  - Human attention and memory
  - Understanding abstractions

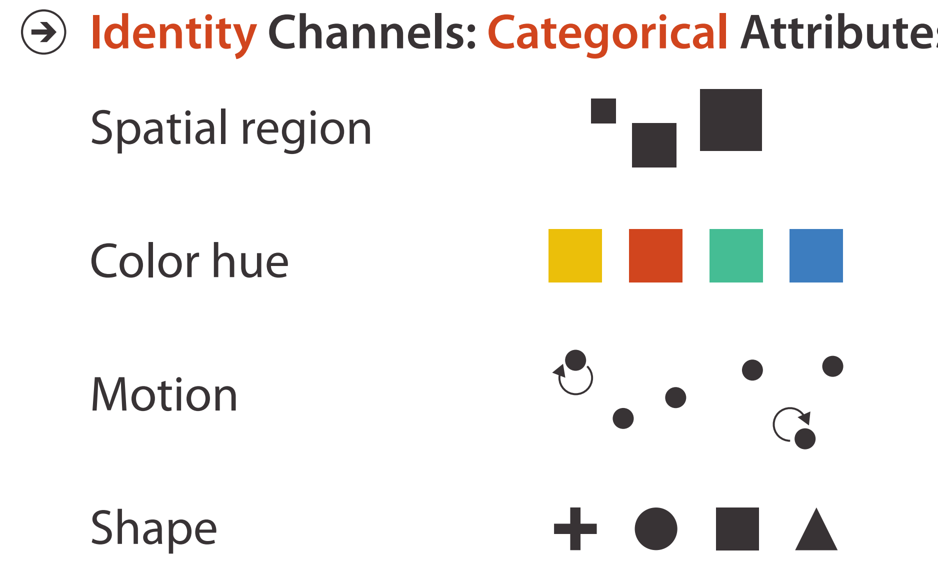
Resource limitations

- **Computational** limits
  - Processing time and system memory
- **Human** limits
  - Human attention and memory
  - Understanding abstractions
- **Display** limits
  - Pixels are precious
  - Information density tradeoff: Info encoding vs unused whitespace



[T. Munzner, 2014]





Expressiveness principle

- **Match channel and data characteristics**

Effectiveness principle

- **Encode important attributes with higher ranked channels**

[T. Munzner, 2014]

Chart Design: Simplifying

Example from Tim Bray

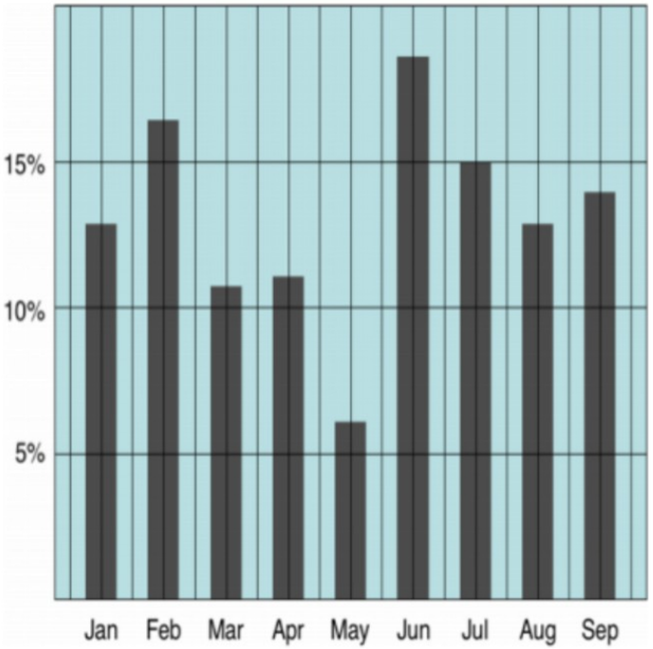
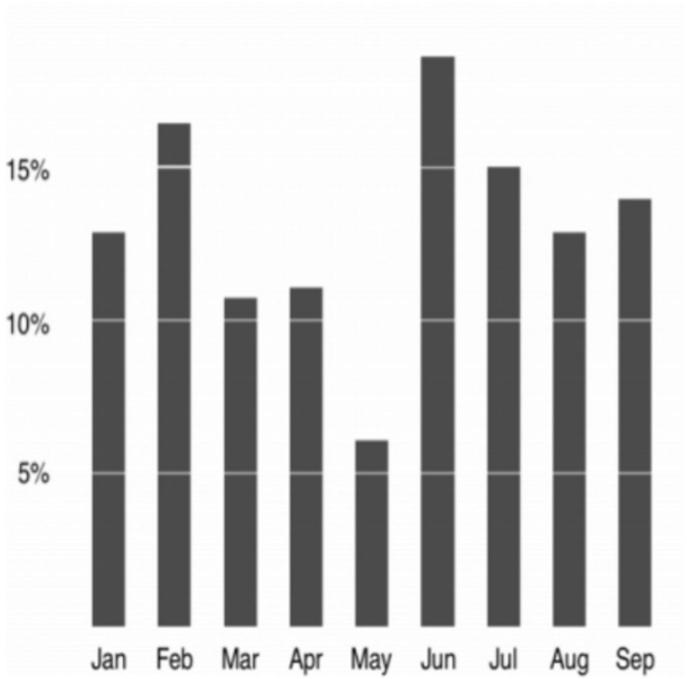


Chart Design: Simplifying

Example from Tim Bray



Principle 2: Understand Magnitudes

Which one is brighter?

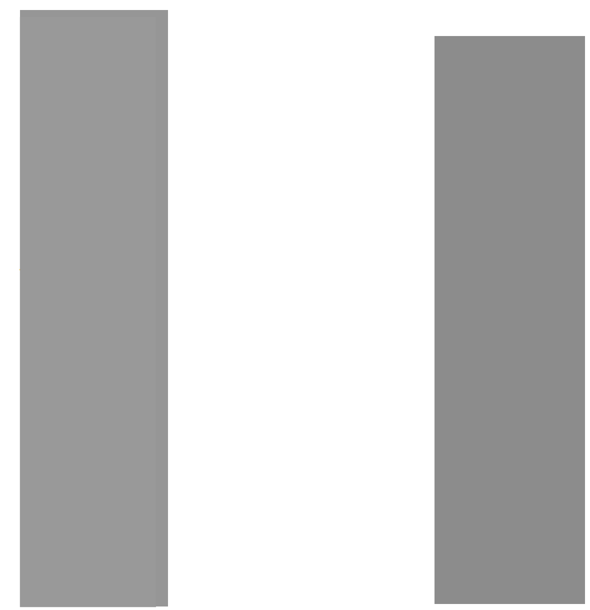
Principle 2: Understand Magnitudes

Which one is brighter?

(40, 40, 40) (38, 38, 38)

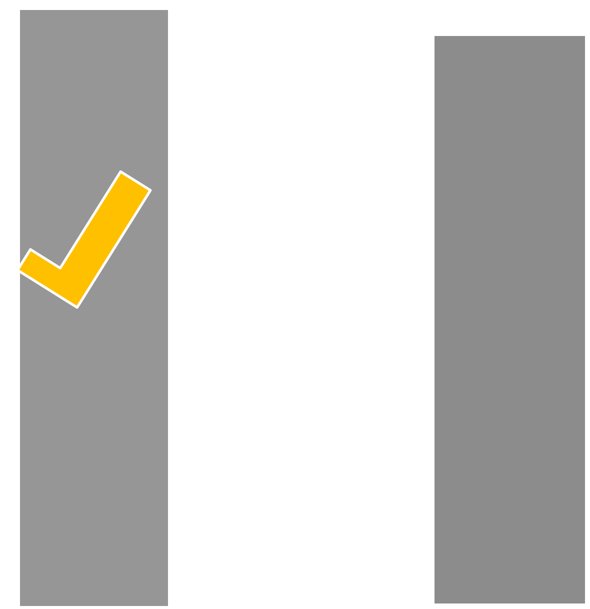
Principle 2: Understand Magnitudes

Which one is longer?



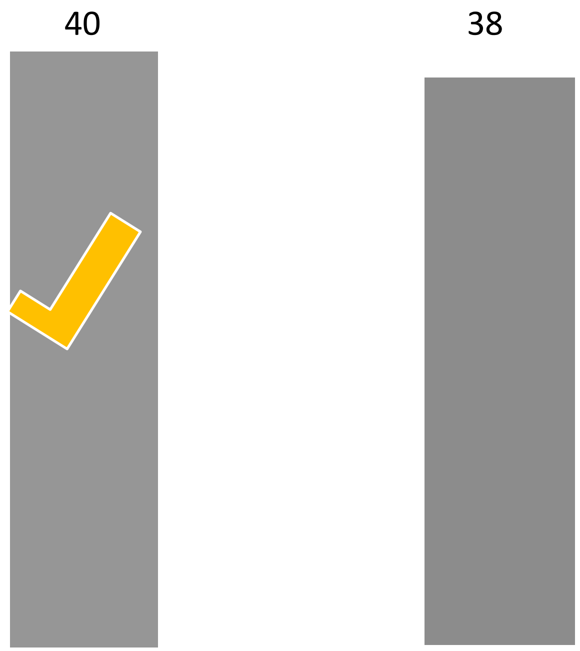
Principle 2: Understand Magnitudes

Which one is longer?



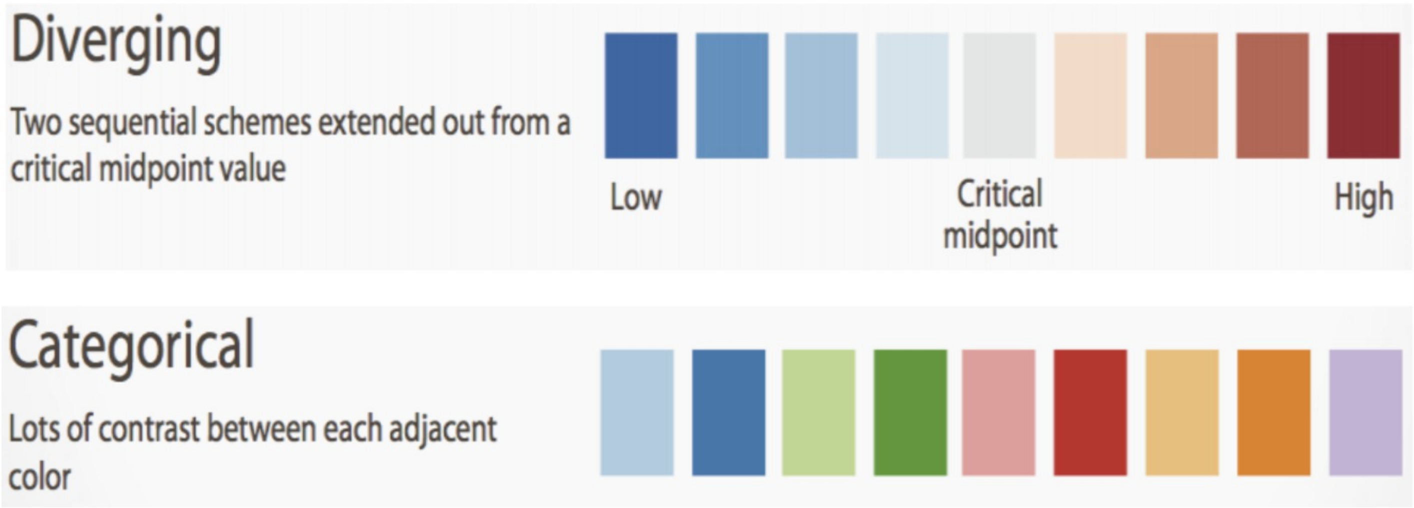
Principle 2: Understand Magnitudes

Which one is longer?



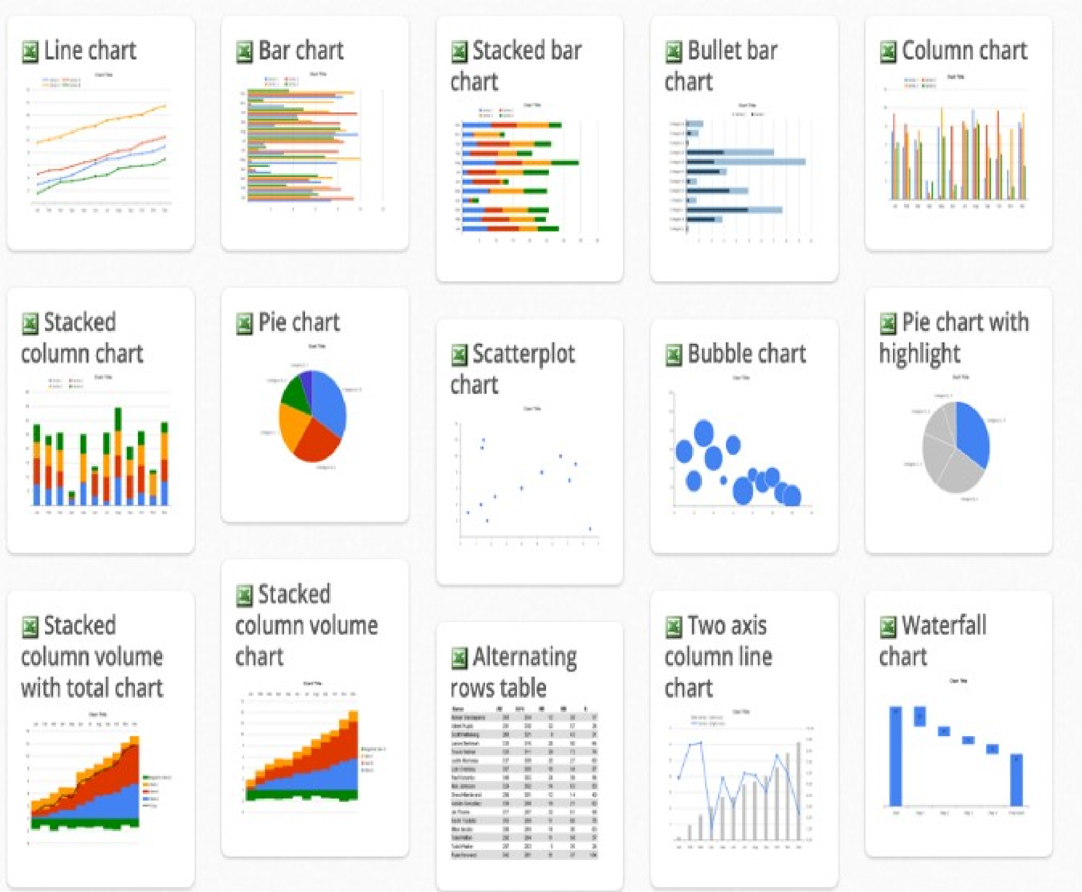
Principle 3: Use Color

- **Make your visualization look beautiful**
  - Colour Lovers: http://www.colourlovers.com
- **Work for different kinds of data**



Principle 4: Use Structure

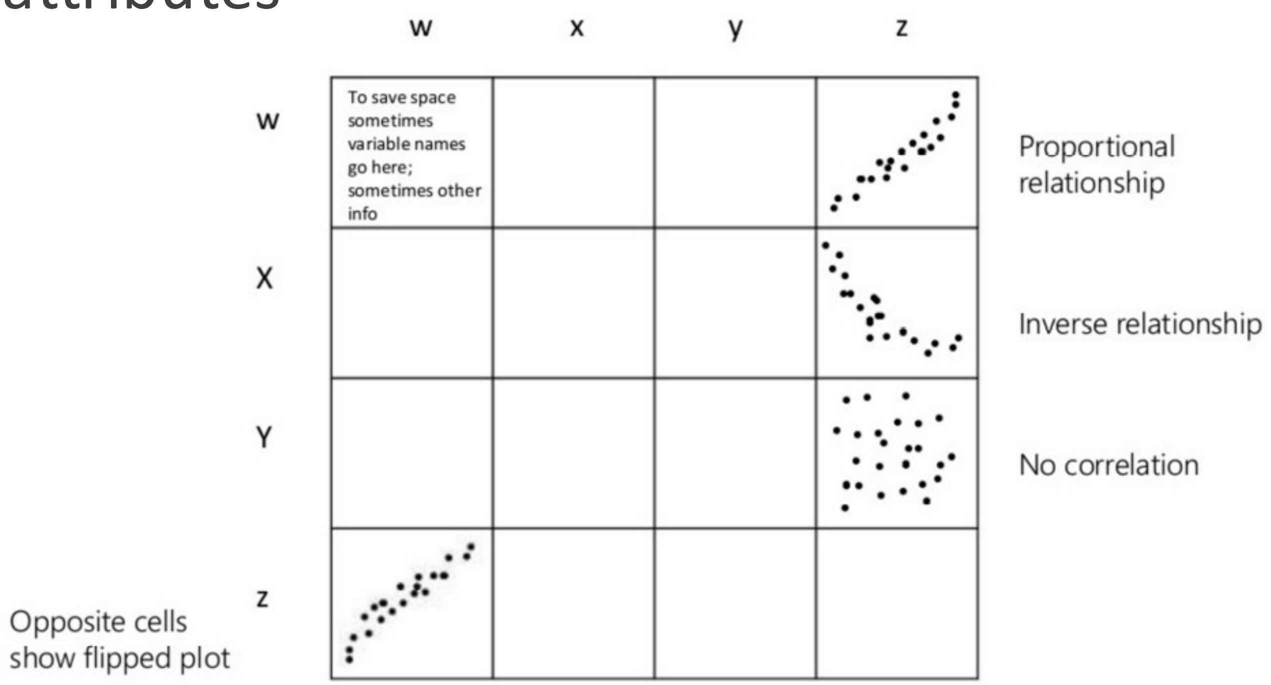
- Chart chooser: http://labs.juiceanalytics.com



Principle 4: Use Structure

Correlation Visualization

- Consider a table with n=4 attributes



Principle 4: Use Structure

Correlation Visualization

- Conduct a deeper analysis on each pair of attributes

	10 °C	20 °C	30 °C	40 °C
6 hrs of light per day				
12 hrs of light per day				
18 hrs of light per day				
24 hrs of light per day				

Sources

- Tamara Munzner's "Visualization Analysis and Design", 2014
- Jiannan Wang's CMPT 733 slides, Spring 2017