

Visualization for Data Science

CMPT 733

Slides by Steven Bergner

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)

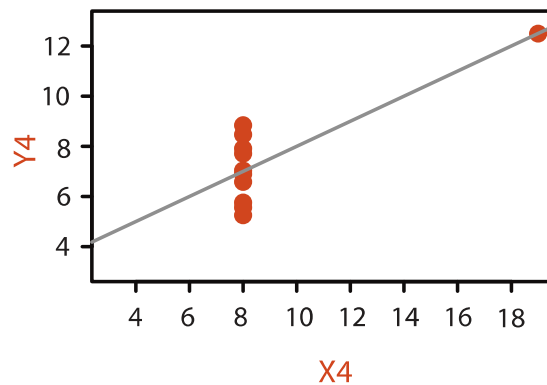
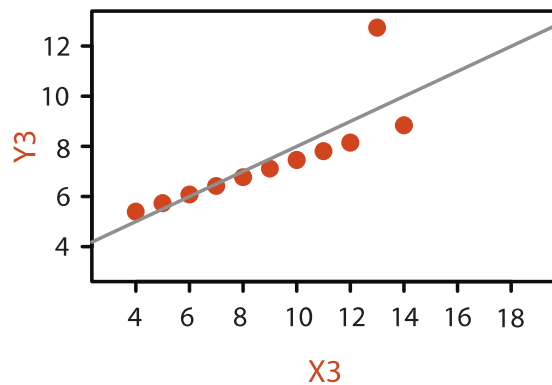
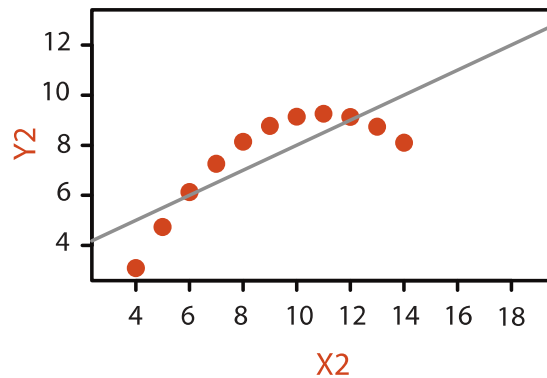
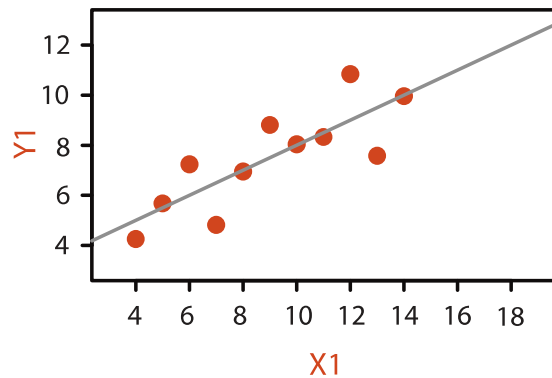
Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.

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

Why have a human in the loop?

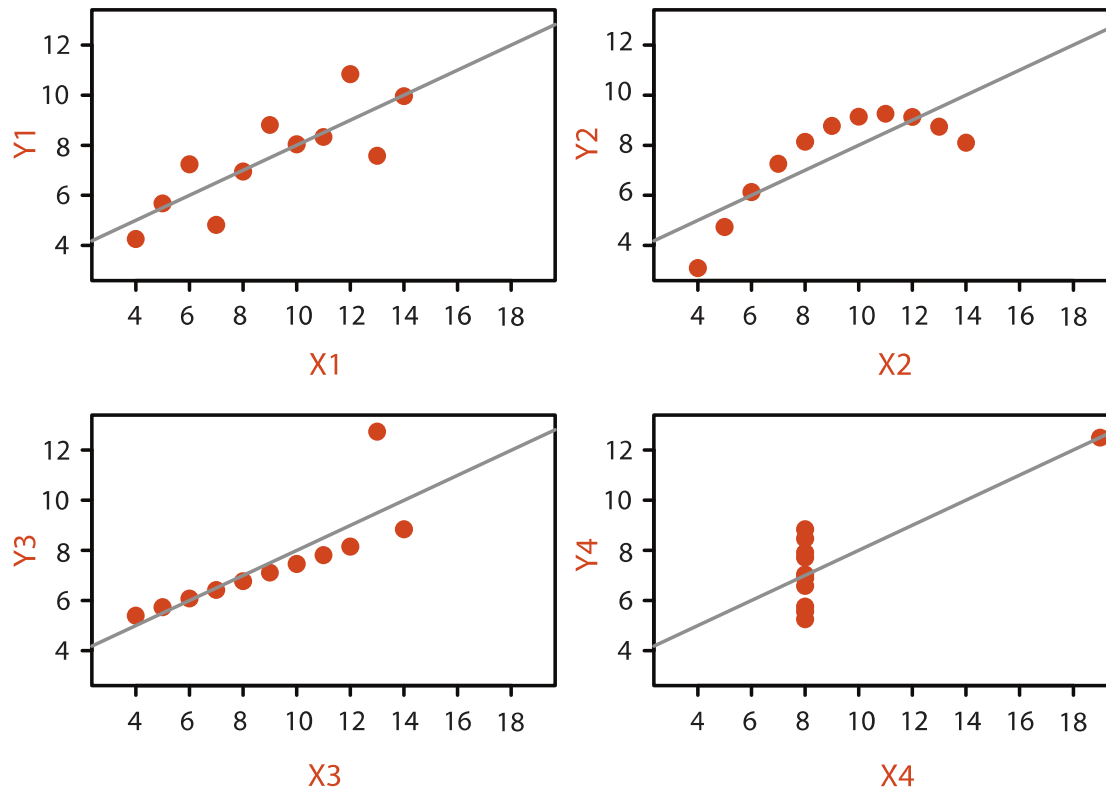
- Not needed when automatic solution is trusted
- Good for ill-specified analysis problems
 - Common setting: “What questions can we ask?”

“Numerical calculations are exact, but Graphs are rough”



- Same relationship among each pair of variables?

“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

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)

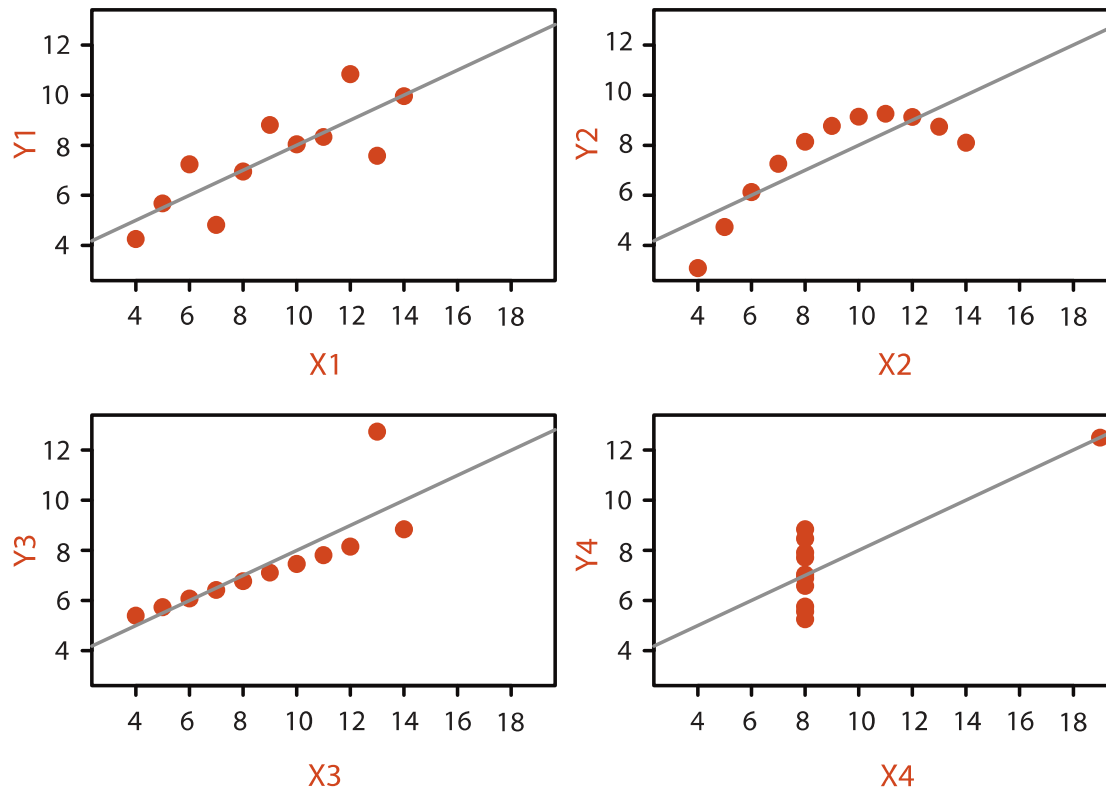
- Long-term use**
 - Exploratory analysis of scientific data
 - Presentation of known results
- Short-term use**
 - For **developers** of automatic solutions:
 - Understand requirements for model development
 - Refine/debug and determine parameters
 - For **end users** of automatic solutions: verify, 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

Why use an external representation?

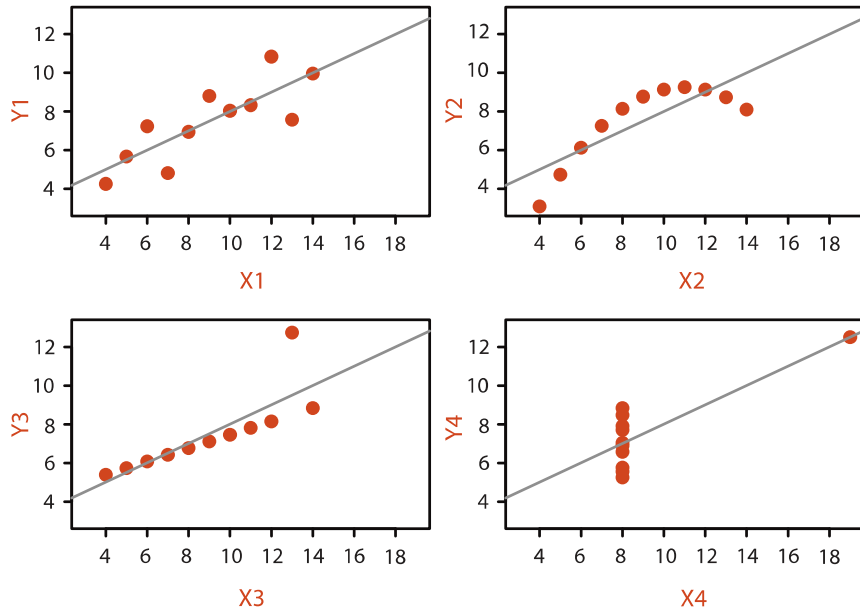
- Replace cognition with perception



	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

Why represent all the data?

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



	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

Analysis framework: four *levels*

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

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*

Analysis framework: four *levels*

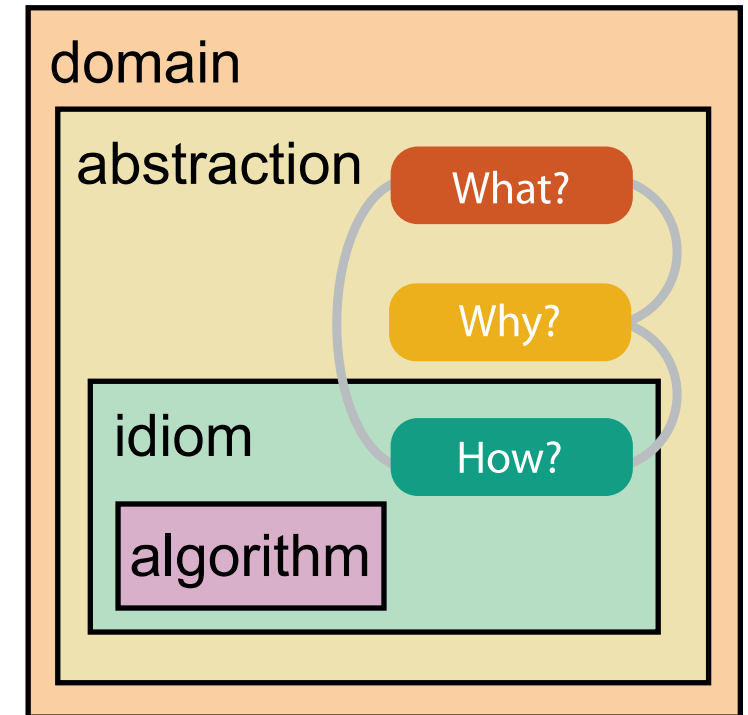
- **Domain** situation: Who are the target users?
- **Abstraction**: Translate from specifics of domain to vocabulary of vis
- **What** is shown? *Data abstraction*
 - Don't just draw what you're given: transform to new form
- **Why** is the user looking at it? *Task abstraction*
- **How** is it shown?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

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

Analysis framework: four *levels*

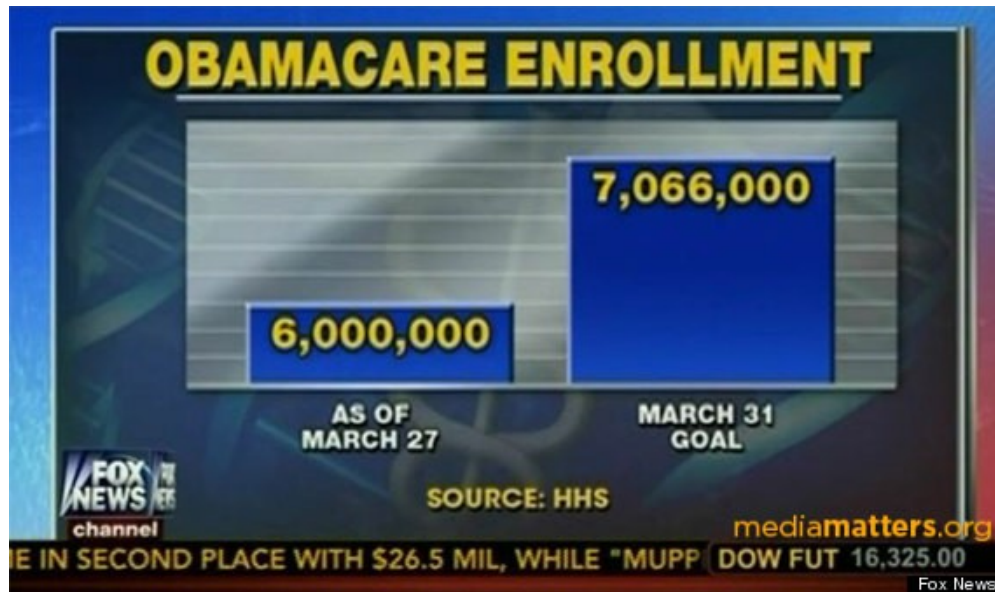
- **Domain** situation: Who are the target users?
- **Abstraction**: Translate from specifics of domain to vocabulary of vis
- **What** is shown? *Data abstraction*
 - Don't just draw what you're given: transform to new form
- **Why** is the user looking at it? *Task abstraction*
- **How** is it shown?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



Examples

Motivation

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

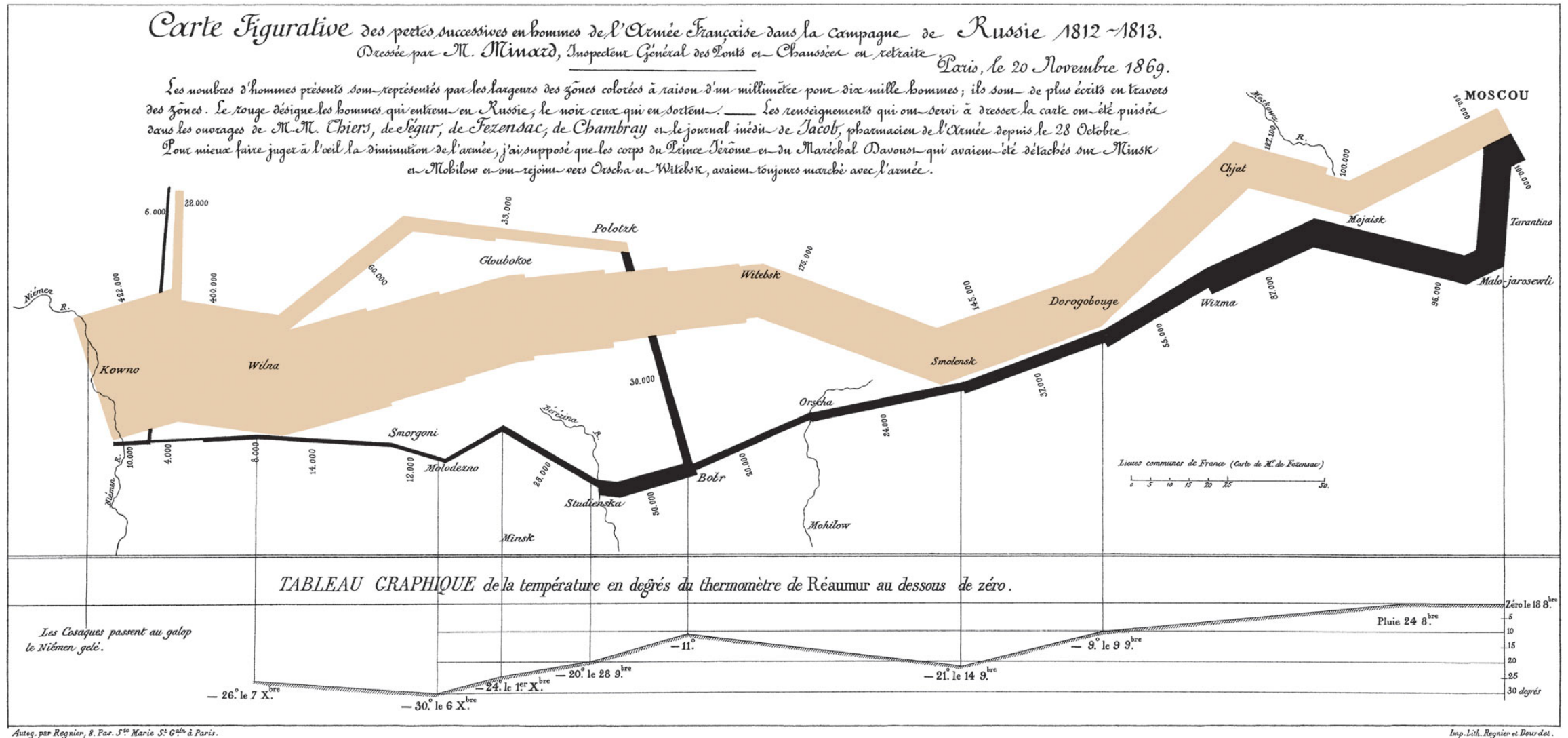


Motivation

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



Cartographic depiction of numeric data



Successive losses in men of the French Army in the Russian campaign 1812-1813 by Charles Minard

Understand Data, Task, and Encoding

What?

Datasets

➔ Data Types

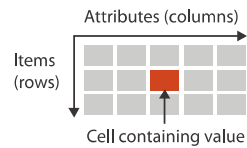
→ Items → Attributes → Links → Positions → Grids

➔ Data and Dataset Types

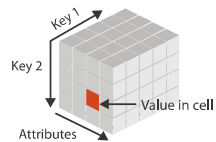
Tables	Networks & Trees	Fields	Geometry	Clusters, Sets, Lists
Items	Items (nodes)	Grids	Items	Items
Attributes	Links	Positions	Positions	
	Attributes	Attributes		

➔ Dataset Types

→ Tables



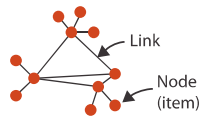
→ Multidimensional Table



→ Geometry (Spatial)



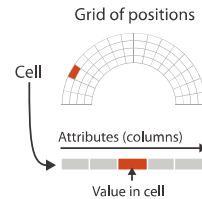
→ Networks



→ Trees



→ Fields (Continuous)



➔ Dataset Availability

→ Static



→ Dynamic



Attributes

➔ Attribute Types

→ Categorical



→ Ordered

→ Ordinal



→ Quantitative



➔ Ordering Direction

→ Sequential



→ Diverging



→ Cyclic



Data Types

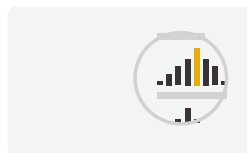
- Items and attributes as rows and columns of tables
- Position and time are special attributes
- Spatial data on grids makes computation easier

Why?

Actions

➔ Analyze

➔ Consume



➔ Present



➔ Enjoy



➔ Produce

➔ Annotate



➔ Record



➔ Derive



➔ Search

	Target known	Target unknown
Location known	<i>Lookup</i>	<i>Browse</i>
Location unknown	<i>Locate</i>	<i>Explore</i>

➔ Query

➔ Identify



➔ Compare



➔ Summarize



Targets

➔ All Data

➔ Trends



➔ Outliers



➔ Features



➔ Attributes

➔ One

➔ Distribution



➔ Extremes



➔ Many

➔ Dependency



➔ Correlation



➔ Similarity



➔ Network Data

➔ Topology

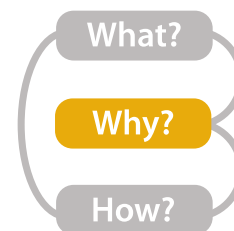


➔ Paths



➔ Spatial Data

➔ Shape

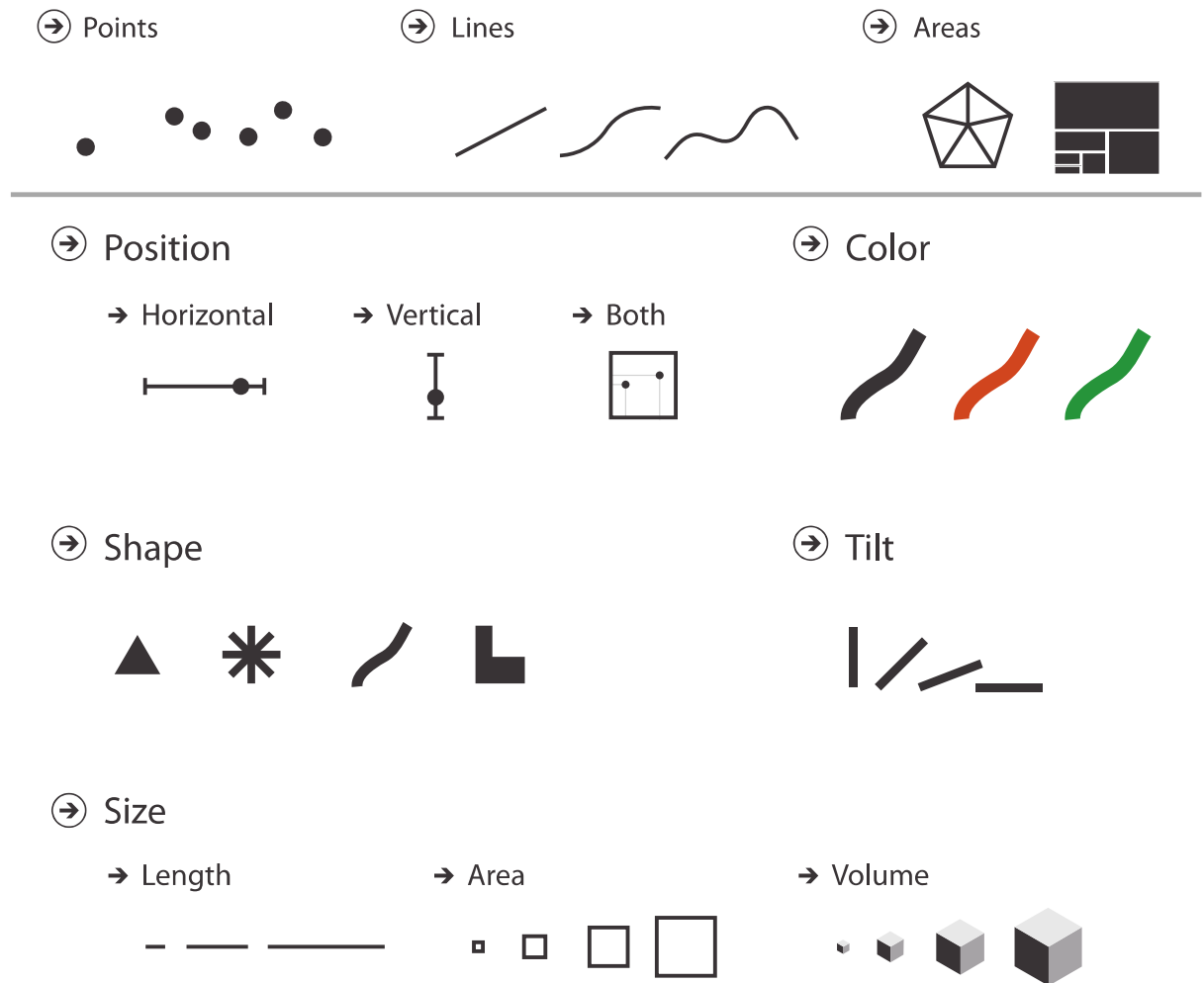


Tasks

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

Visual Encoding – How?

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



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

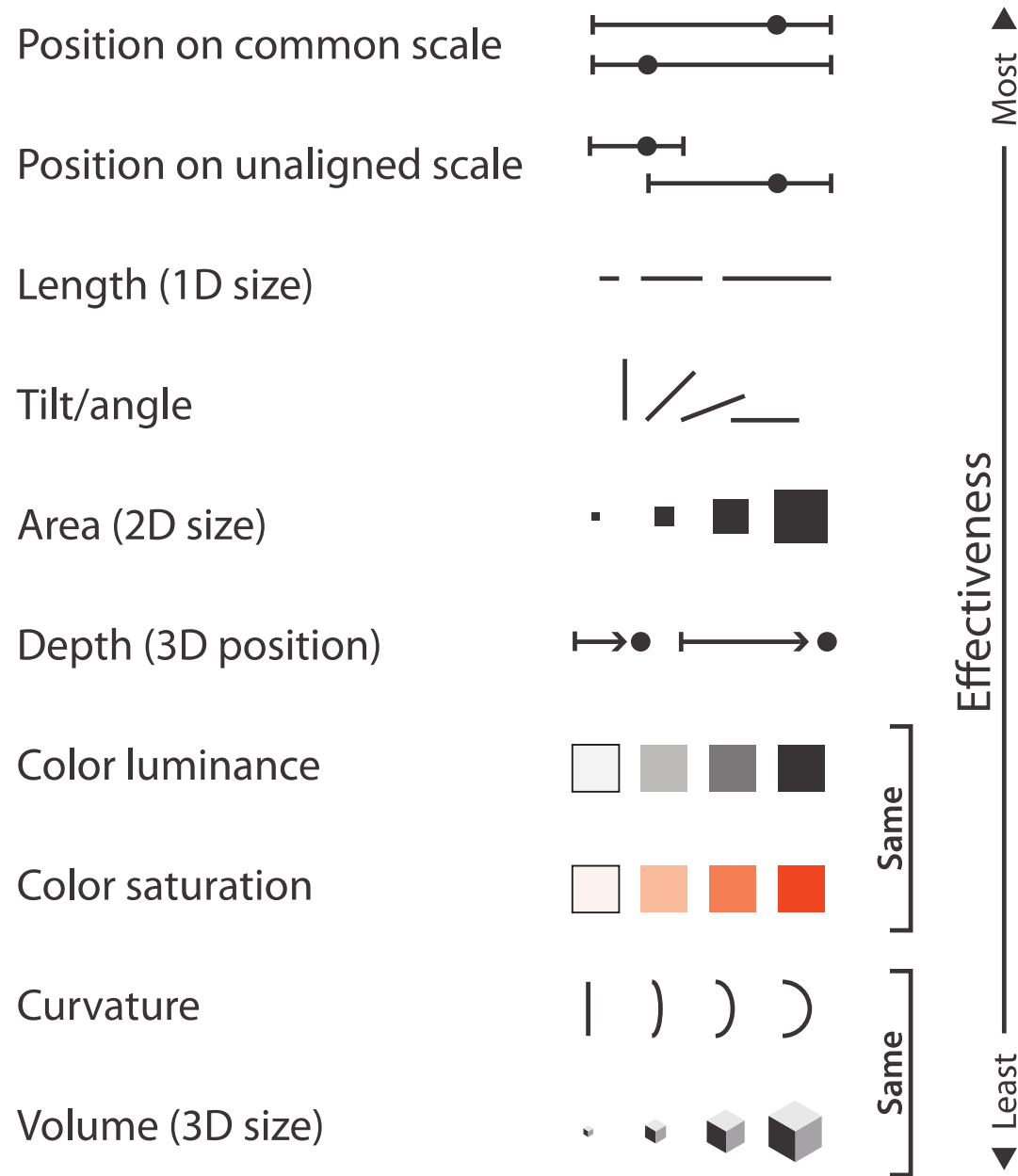
Resource limitations

- **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

➔ Magnitude Channels: **Ordered** Attributes



➔ Identity Channels: **Categorical** Attributes



➔ Magnitude Channels: **Ordered** Attributes

Position on common scale 

Position on unaligned scale 

Length (1D size) 

Tilt/angle 

Area (2D size) 

Depth (3D position) 

Color luminance 

Color saturation 

Curvature 

Volume (3D size) 

Effectiveness
Most
Least

➔ Identity Channels: **Categorical** Attributes

Spatial region 

Color hue 

Motion 

Shape 

Expressiveness principle

- **Match channel and data characteristics**

Effectiveness principle

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

Chart Design: Simplifying

Example from Tim Bray

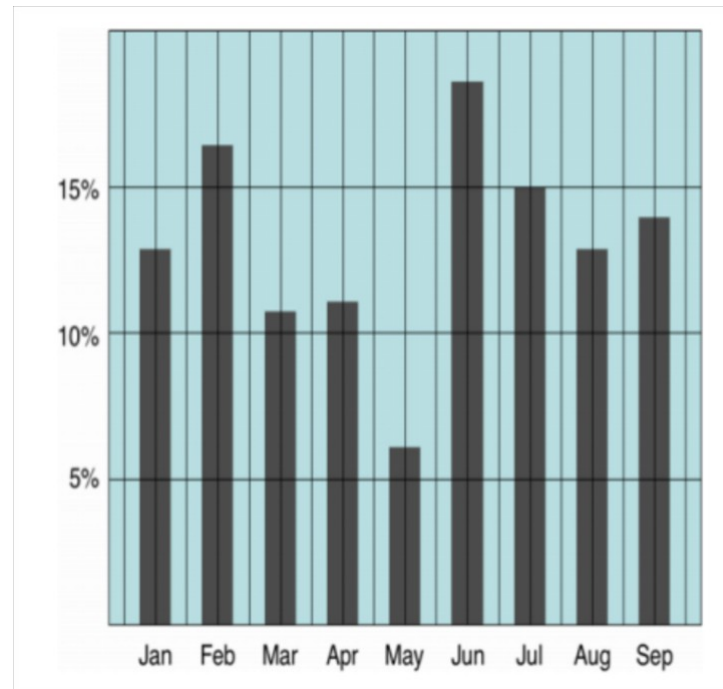
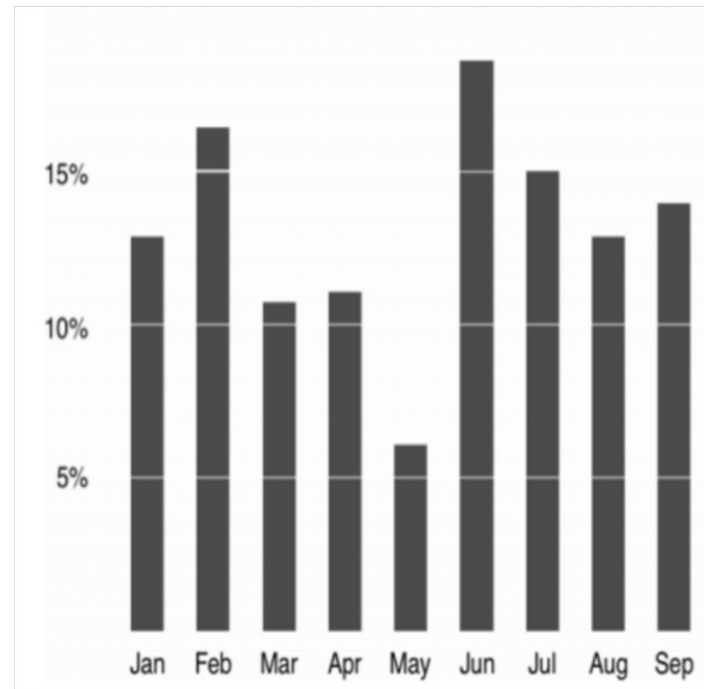


Chart Design: Simplifying

Example from Tim Bray



Principle 2: Understand Magnitudes

Which one is brighter?

(40, 40, 40)



(38, 38, 38)



Principle 2: Understand Magnitudes

Which one is brighter?

(40, 40, 40)

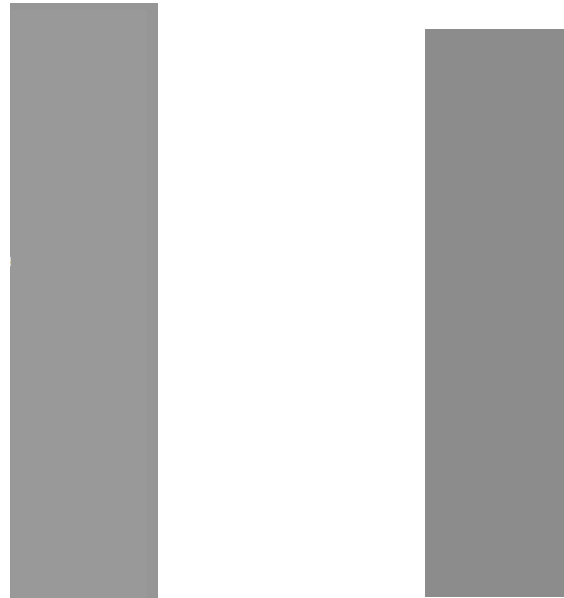


(38, 38, 38)



Principle 2: Understand Magnitudes

Which one is longer?



Principle 2: Understand Magnitudes

Which one is longer?



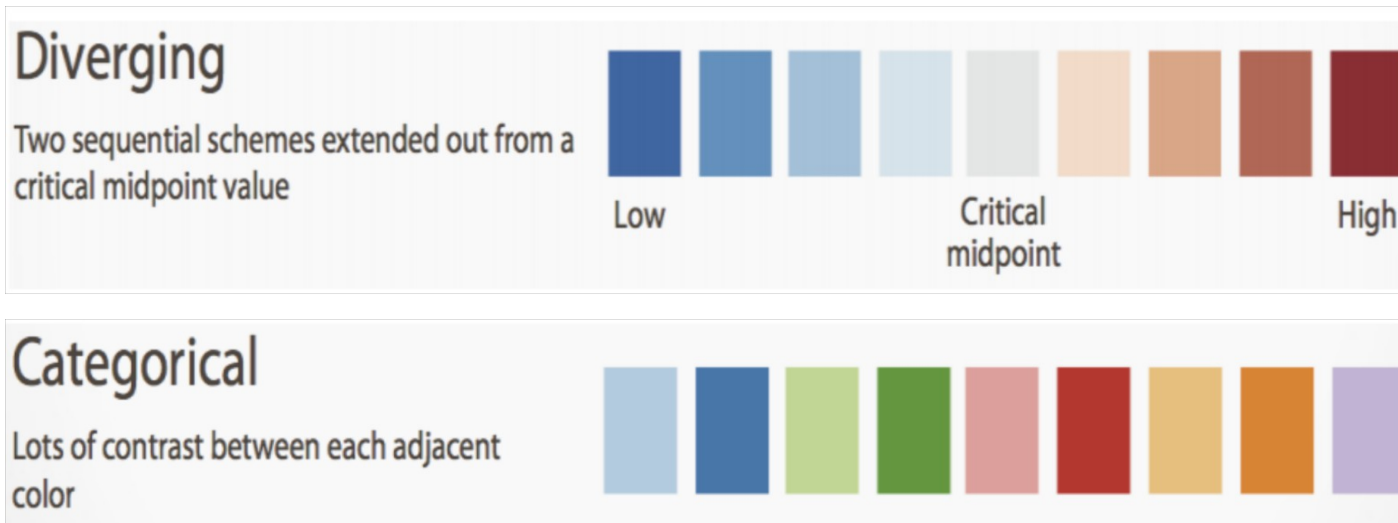
Principle 2: Understand Magnitudes

Which one is longer?



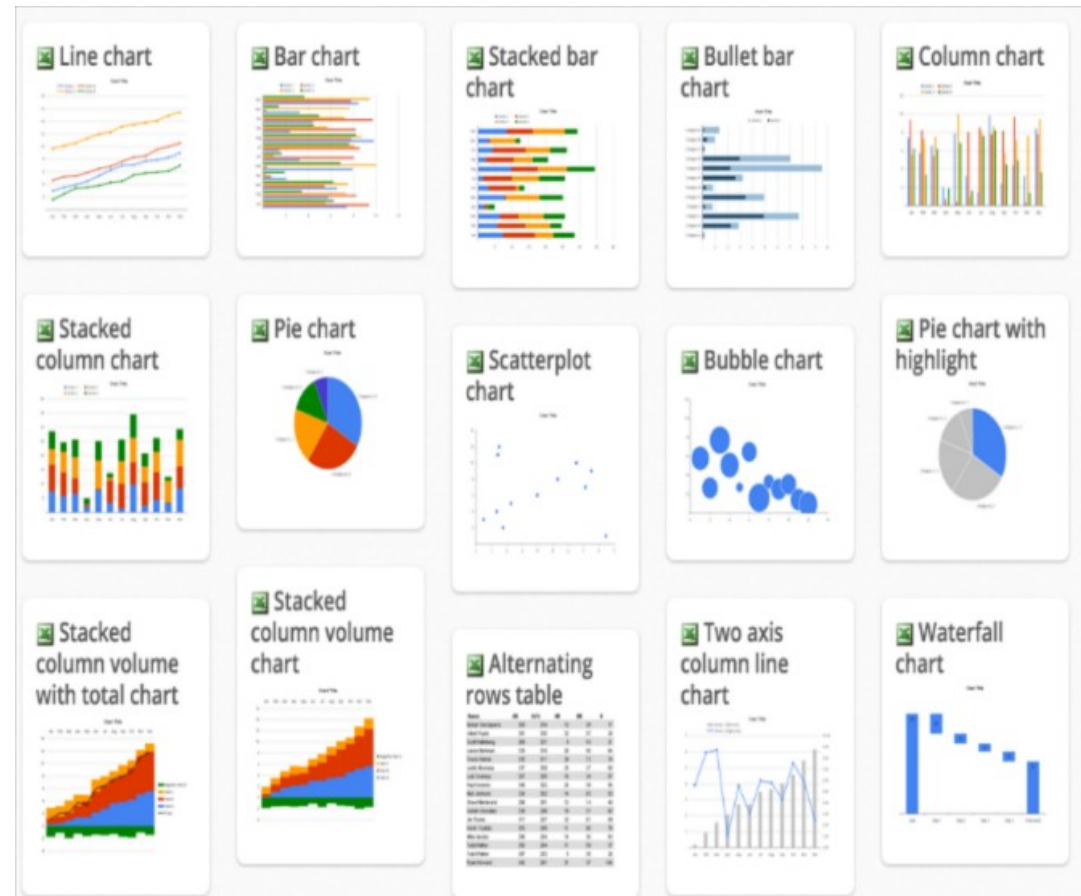
Principle 3: Use Color

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



Principle 4: Use Structure

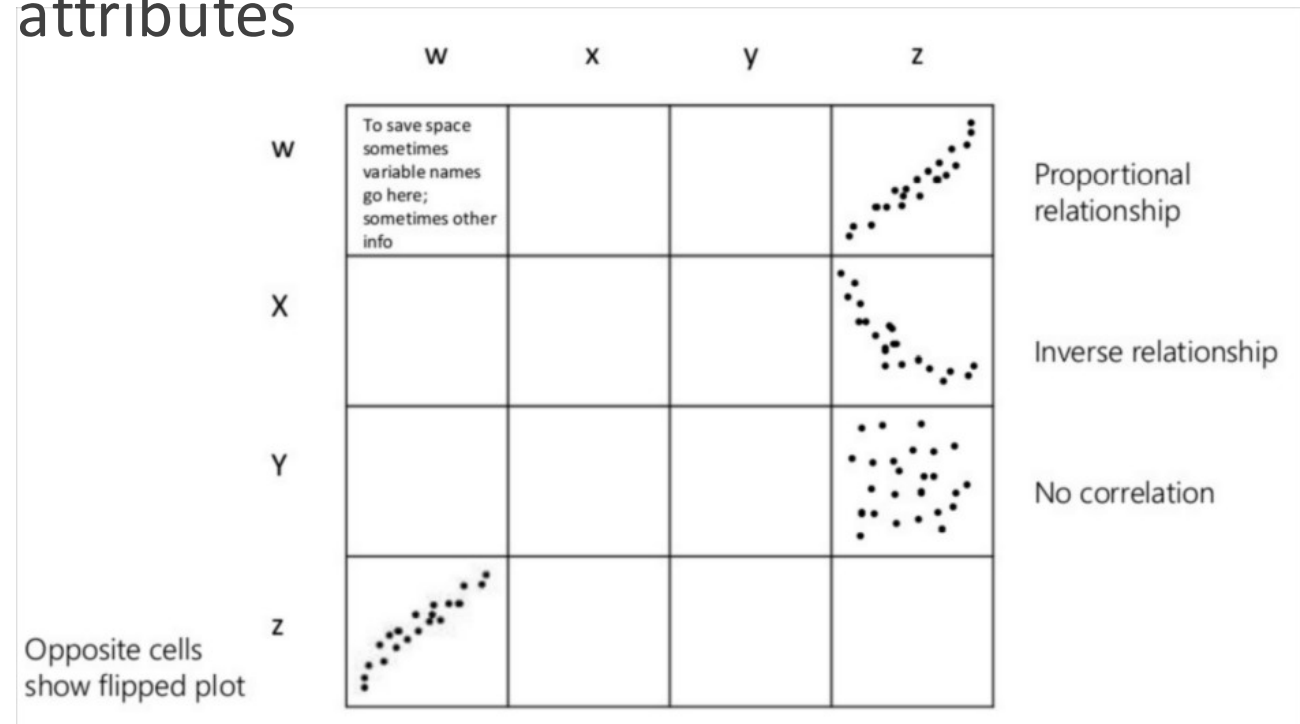
- Chart chooser: <http://labs.juiceanalytics.com>



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