

---

---

# Information Visualization Basics

— People-Oriented Computing —

30.09.2019

---

---

# Announcements

- First lab is WEDNESDAY, OCTOBER 2
- Exercise is posted on OLAT

# Lab Sessions

- Not mandatory but highly recommended
- Format:
  - Time to work independently or collaboratively on the exercise
  - Explanation of exercise and sample solutions by TAs/Tutors
  - Additional independent work time and Q&A with TAs/Tutors
- You may find it beneficial to do/think about the exercise prior to the lab
  - Lab session alone may not be sufficient time for completing the exercise
  - Know ahead of time what you may need help with
- This is the best time to get help with the exercises
- Presentation slides will be made available on OLAT

# Learning Goals

After this lecture, you should:

- Have a foundational understanding of the field of information visualization
- Be familiar with key examples of effective and ineffective real world visualization examples
- Have a basic understanding of visual features, visual search, and human processing of visual information
- Be familiar with basic principles and guidelines for information visualization design
- Be able to identify strengths and weaknesses of information designs and justify arguments using concepts and principles of information visualization

# **WHAT IS INFORMATION VISUALIZATION AND WHY IS IT IMPORTANT?**

# Information Overload



Image courtesy of Petra Isenberg

# How Much Information?

Information produced in 2012: 2.8 zettabytes

Projected information produced in 2020: 40 zettabytes

1 ZB = 1 trillion gigabytes =  $10^{21}$  bytes = 1000000000000000000000 bytes

# With So Much Information...

- How do we understand and make sense of data?
- How do we extract meaning and value from data?
- How do we communicate data?
- How can we use data to gain new insight and knowledge?



# What is Information Visualization?

*The use of computer-supported,  
interactive visual representations of  
abstract data to amplify cognition*

-Stu Card et al., 1999

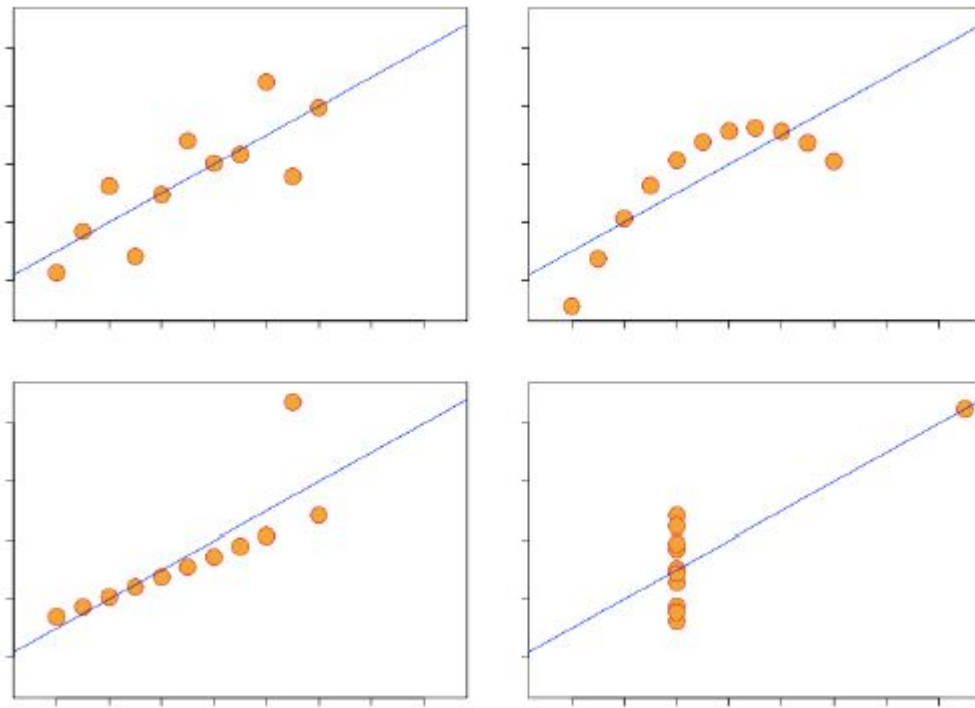
# Why Visualize Data?

- To record information
- To analyze information
- To present and communicate information
- To discover new information and gain insight

# Anscombe's Quartet

I		II		III		IV	
x	y	x	y	x	y	x	y
10,00	8,04	10,00	9,14	10,00	7,46	8,00	6,58
8,00	6,95	8,00	8,14	8,00	6,77	8,00	5,76
13,00	7,68	13,00	8,74	13,00	12,74	8,00	7,71
9,00	8,81	9,00	8,77	9,00	7,11	8,00	8,84
11,00	8,33	11,00	9,26	11,00	7,81	8,00	8,47
14,00	9,96	14,00	8,10	14,00	8,84	8,00	7,04
6,00	7,24	6,00	6,13	6,00	6,08	8,00	6,26
4,00	4,26	4,00	3,10	4,00	5,39	19,00	12,50
12,00	10,84	12,00	9,13	12,00	8,15	8,00	5,56
7,00	4,82	7,00	7,26	7,00	6,42	8,00	7,91
5,00	5,68	5,00	4,74	5,00	5,73	8,00	6,89

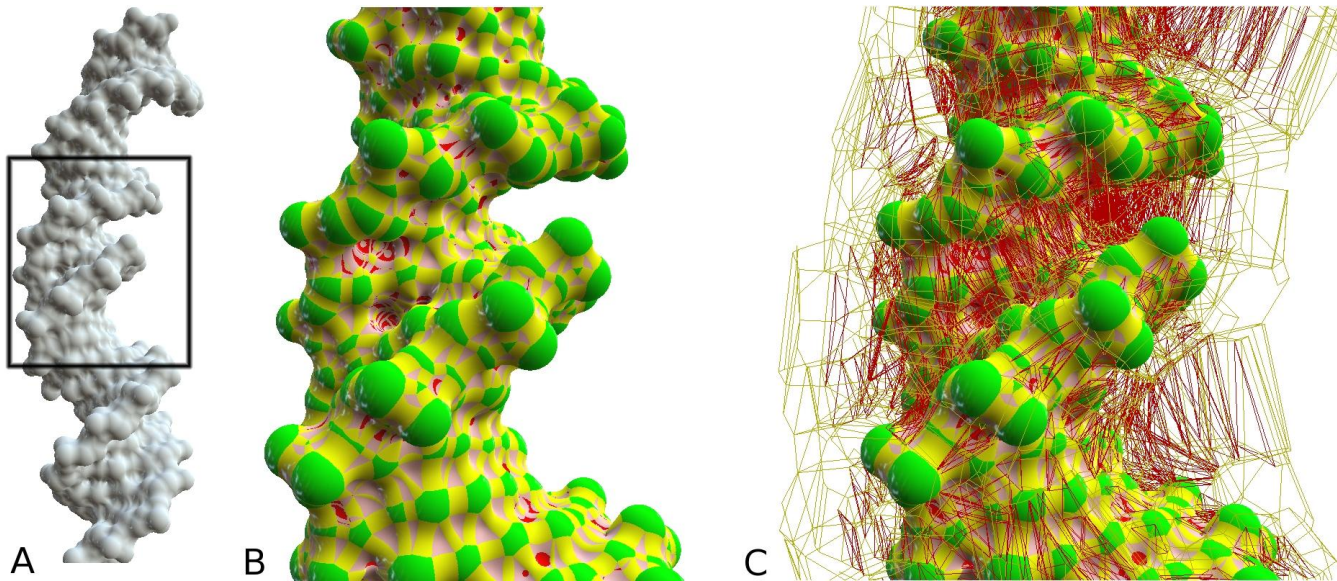
# Anscombe's Quartet



*"Contained within the data of any investigation is information that can **yield conclusions to questions not even originally asked**. That is, there can be surprises in the data... To regularly miss surprises by failing to probe thoroughly with visualization tools is terribly inefficient because **the cost of intensive data analysis is typically very small compared with the cost of data collection.**"*

– W.S. Cleveland, 1985

# Information vs. Scientific Visualization



Source: inria.fr

# Information vs. Scientific Visualization

## InfoVis

- Abstract data with no physical correspondence
- Free mapping of data to 2D or 3D space

## SciVis

- Scientific data corresponding to physical phenomena
- Fixed positions in space for visualizations

# Key Challenges in InfoVis

- Creating meaningful and useful mappings of abstract data onto 2D or 3D space
- Representing extremely large sets of data in a finite amount of space
- Representing diverse types and forms of data within a visualization



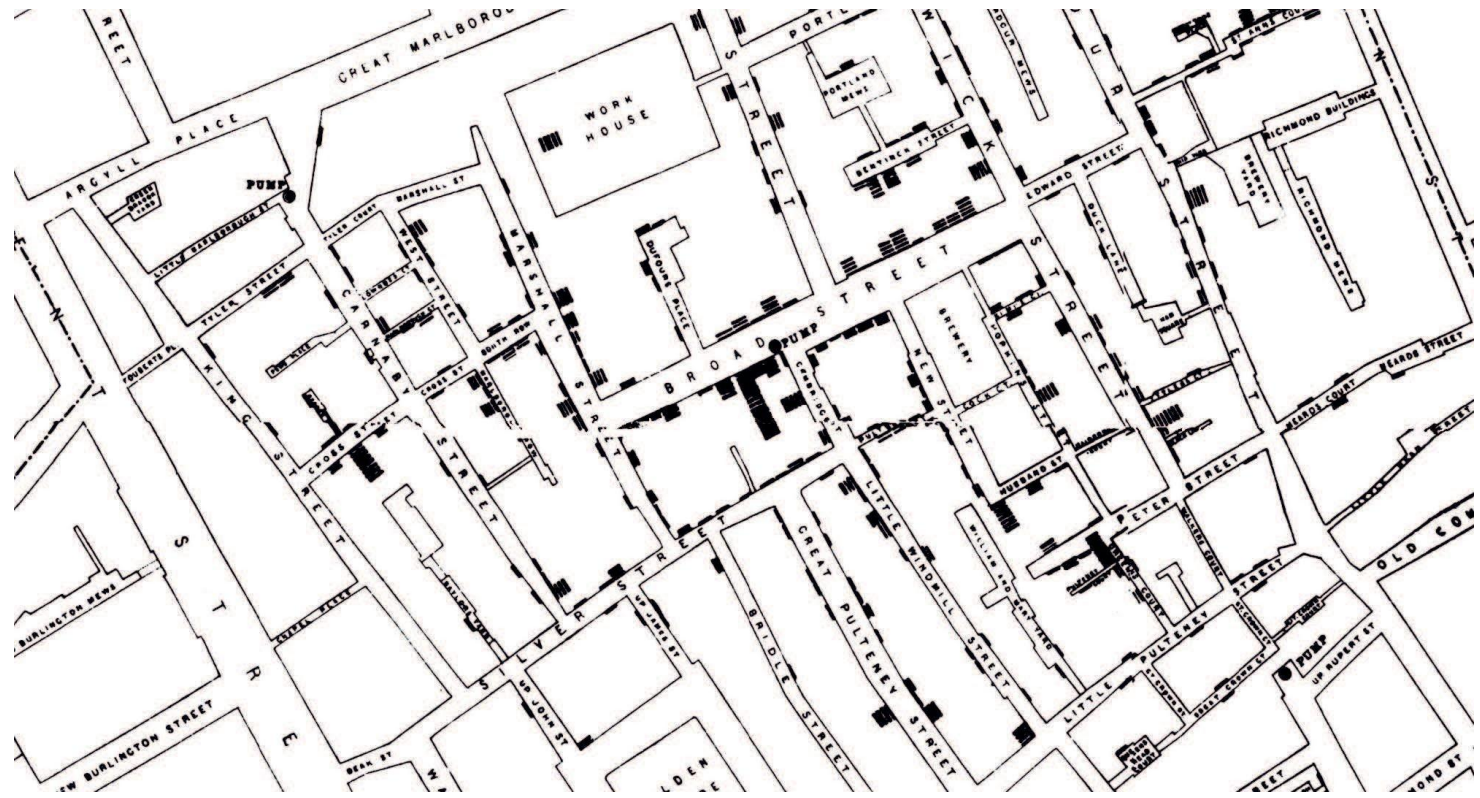
# **CLASSIC VISUALIZATION EXAMPLES: SUCCESSES AND FAILURES**

# 1845 Broad Street Cholera Outbreak

- Severe outbreak of cholera in London
  - 127 died within 3 days, 616 by the end of the outbreak
- Dominant belief at the time was “miasma theory” – deaths caused by bad air

# 1845 Broad Street Cholera Outbreak

- Dr. John Snow was skeptical of miasma theory and investigated the cause by:
  - Talking to local residents
  - Identifying a water pump as a potential source
  - Using maps to visualize cases of cholera and locations of water pumps
  - Communicating findings to local council who disabled the pump



# 1845 Broad Street Cholera Outbreak

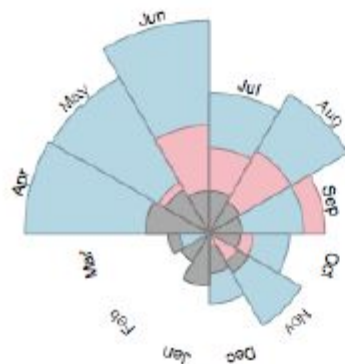
- Combination of data and visualizations explained cases and exceptions
- Visualizations and studies were convincing even though local inspection of the pump was inconclusive

# Nightingale's Rose

Florence Nightingale created a visualization to convince the British military that sanitation was a greater hazard to soldiers than battlefield combat



# DIAGRAM OF THE CAUSES OF MORTALITY IN THE ARMY OF THE EAST

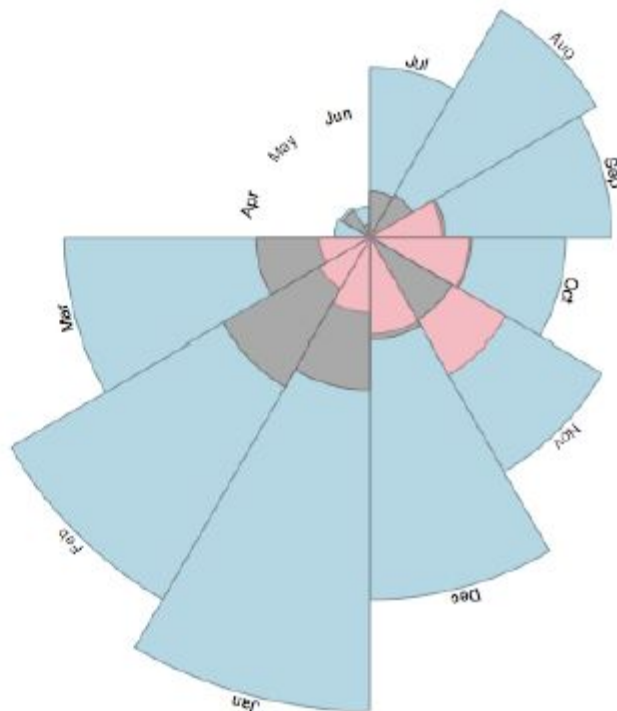


*The Areas of the blue, red, & black wedges are each measured from the centre as the common vertex*

*The blue wedges measured from the centre of the circle represent area for area the deaths from Preventible or Mitigable Zymotic Diseases, the red wedges measured from the center the deaths from wounds, & the black wedges measured from the center the deaths from all other causes*

*In October 1844, & April 1855, the black area coincides with the red, in January & February 1856, the blue coincides with the black*

*The entire areas may be compared by following the blue, the red & the black lines enclosing them.*



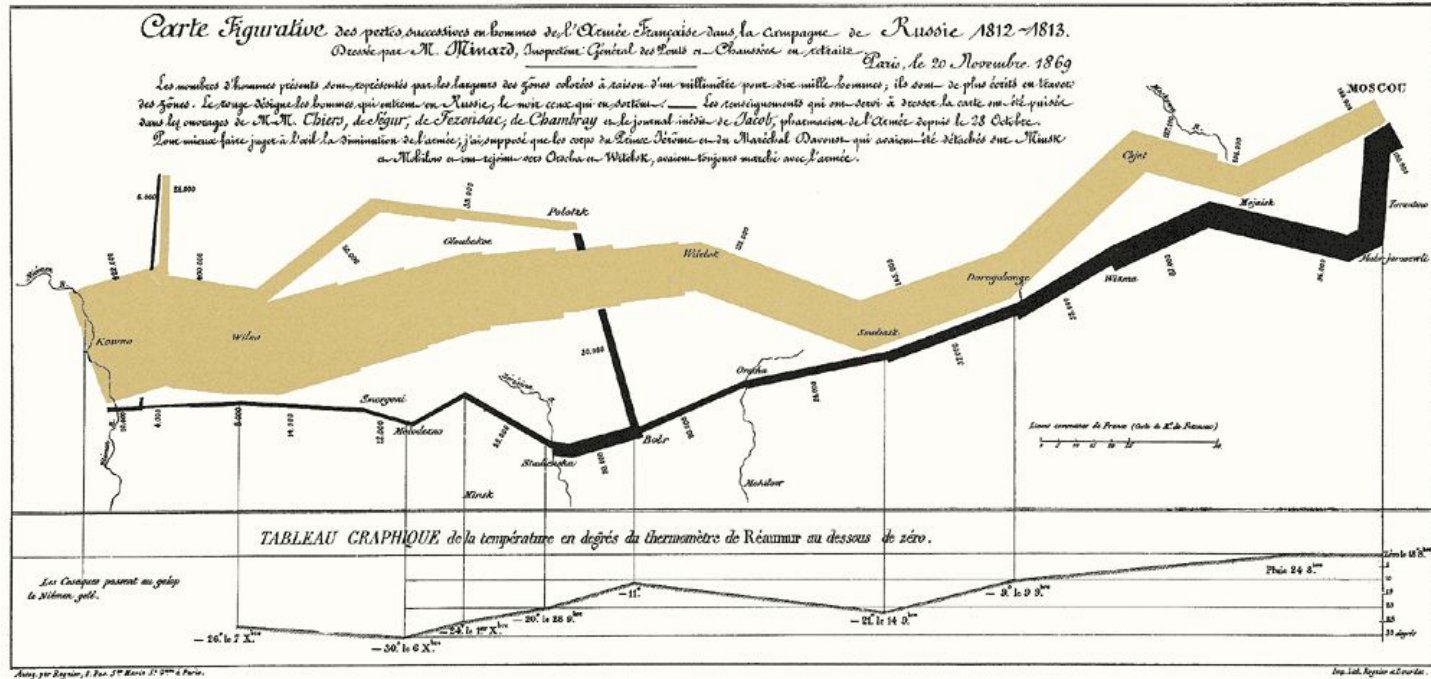
Source: <http://mbostock.github.com/protovis/ex/crimea-rose.html>

# Napoleon's March

- In 1869, Charles Minard created a visualization of Napoleon's march into Russia
- Widely admired for the richness of information conveyed and succinctness of visualization



# Napoleon's March



# 1986 Challenger Explosion

- Rocket engineers recommended cancellation because of predicted low temperature's potential effect on O-rings
- NASA officials were unconvinced of the connection
- O-rings became brittle and leaked
- Spacecraft exploded killing all 7 astronauts



# 1986 Challenger Explosion

## BLOW BY HISTORY

SRM-15 WORST BLOW-BY

○ 2 CASE JOINTS (80°), (110°) ARE

○ MUCH WORSE VISUALLY THAN SRM-22

SRM 22 BLOW-BY

○ 2 CASE JOINTS (30-40°)

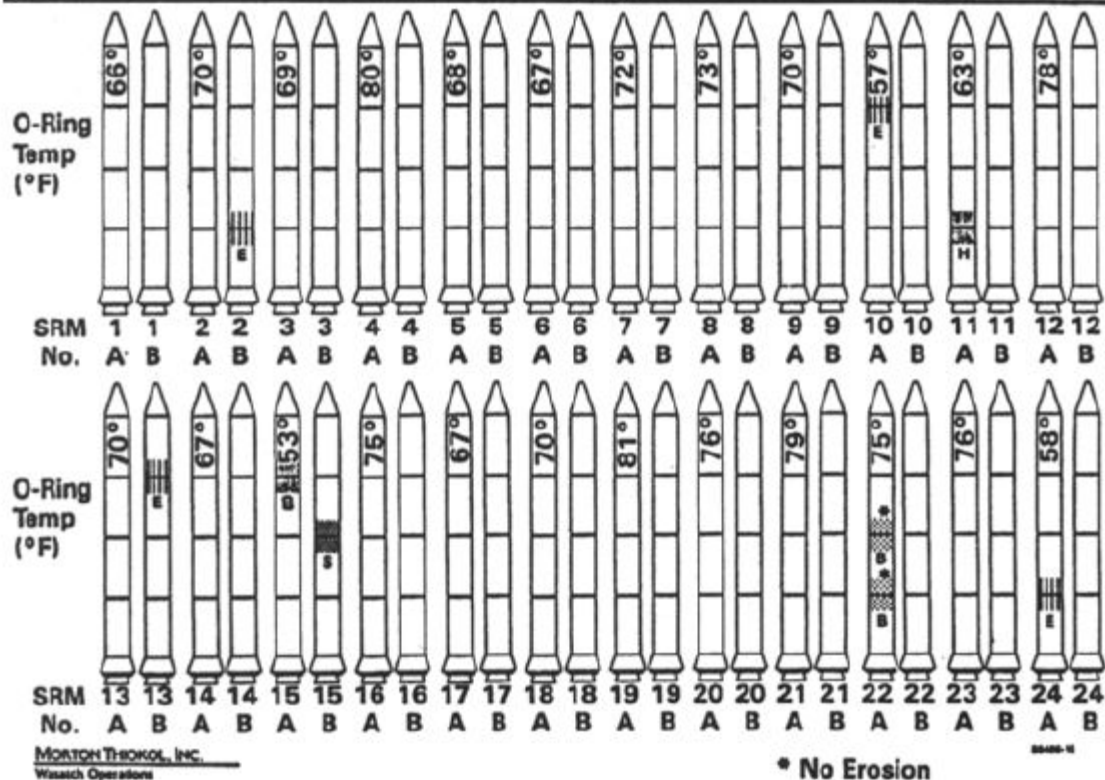
SRM-13A, 15, 16A, 18, 23A 24A

○ NOZZLE BLOW-BY

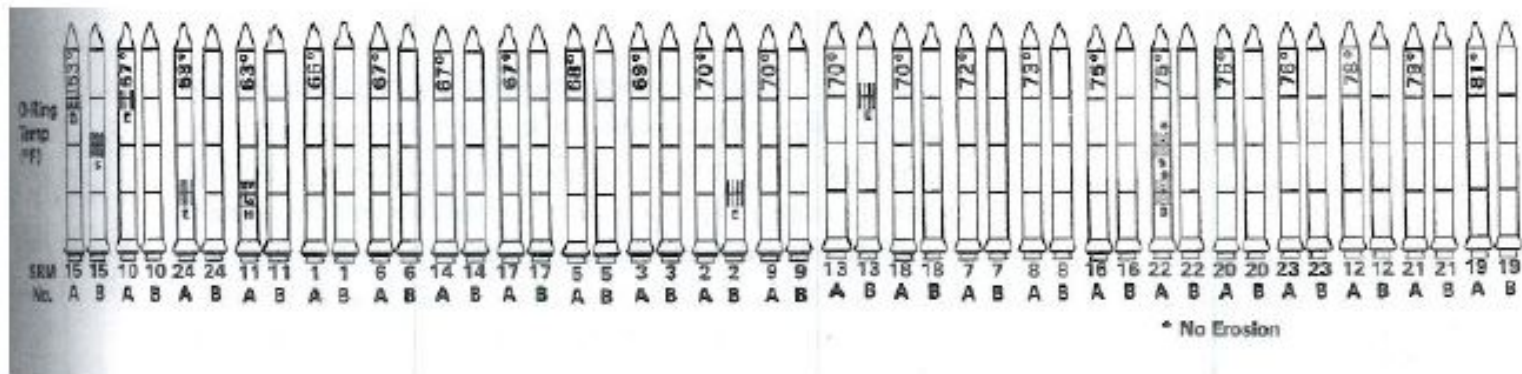
## HISTORY OF O-RING TEMPERATURES (DEGREES - F)

<u>MOTOR</u>	<u>MOT</u>	<u>AMB</u>	<u>O-RING</u>	<u>WIND</u>
DM-1	68	36	47	10 MPH
DM-2	76	45	52	10 MPH
QM-3	72.5	40	48	10 MPH
QM-4	76	48	51	10 MPH
SRM-15	52	64	53	10 MPH
SRM-22	77	78	75	10 MPH
SRM-25	55	26	29 27	10 MPH 25 MPH

## History of O-Ring Damage in Field Joints (Cont)



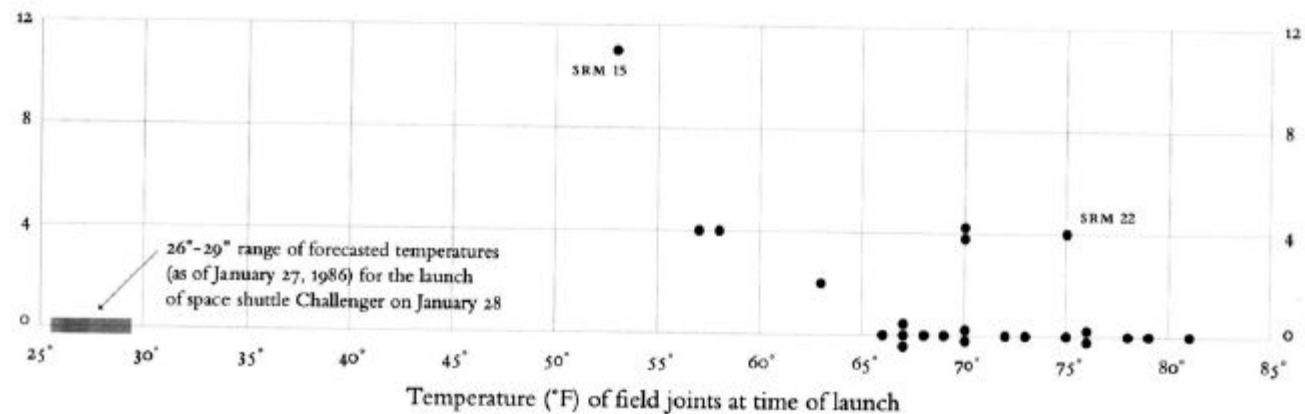
INFORMATION ON THIS PAGE WAS PREPARED TO SUPPORT AN ORAL PRESENTATION  
AND CANNOT BE CONSIDERED COMPLETE WITHOUT THE ORAL DISCUSSION



# Major Issues with O-Ring Damage Visualization

- Lack of clarity in depicting cause and effect
- Wrong ordering of data – sequential order conceals possible link between temperature and O-ring damage
- Chartjunk – strong visual presence of rocket graphics draws attention away from data

O-ring damage  
index, each launch



*“There are right ways and wrong ways to show data; there are **displays that reveal the truth and displays that do not**. And, if the matter is an important one, then getting the displays of evidence right or wrong can possibly have **momentous consequences**.”*

– E. Tufte, 1997



# VISUAL THINKING AND PERCEPTION

# What Can We See Easily?

- How does the designer ensure that all visual queries can be effectively and rapidly served?
- How do we use semantically meaningful graphic objects to create designs with the right amount of salience?
- Take perceptual laws of visual distinctness and pattern processing into account

# How Many 3s?

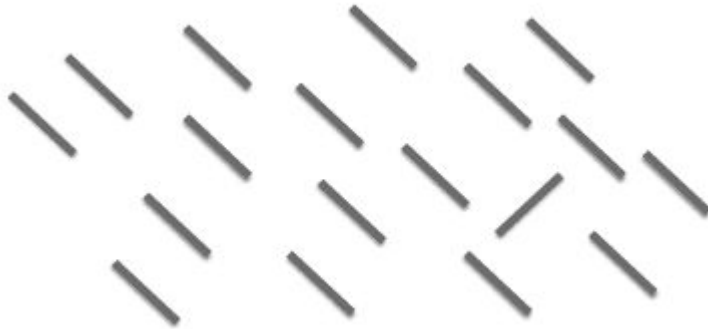
45929078059772098775972655665110049836645  
27107462144654207079014738109743897010971  
43907097349266847858715819048630901889074  
25747072354745666142018774072849875310665

# How Many 3s?

45929078059772098775972655665110049836645  
27107462144654207079014738109743897010971  
43907097349266847858715819048630901889074  
25747072354745666142018774072849875310665

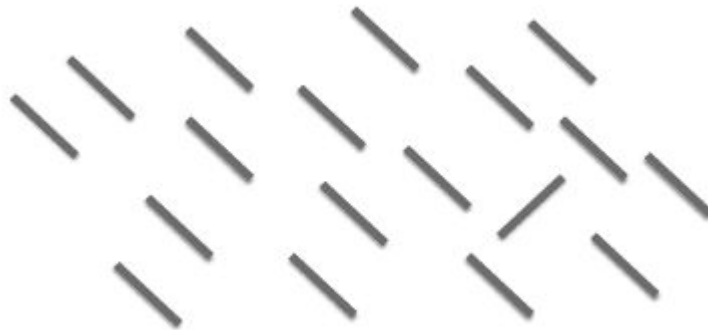
# Pre-attentive Processing

Psychologist Anne Treisman studied visual search to understand what makes some patterns easy to find

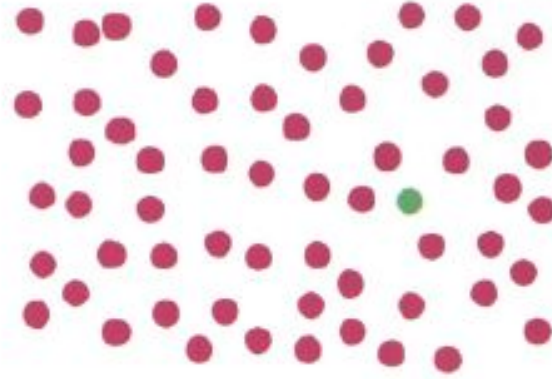


# Pre-attentive Processing

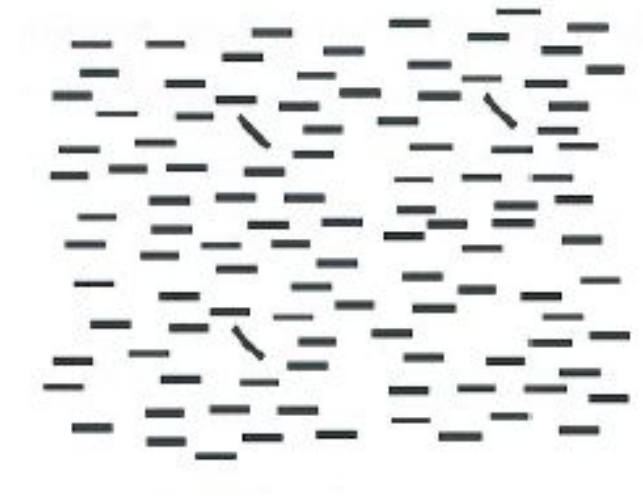
For some combinations of targets and distractors, the time to respond was independent of the number of distractors – indicative of pre-attentive processing



# Pre-attentive?



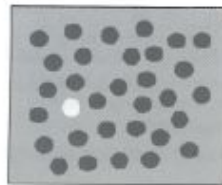
# Pre-attentive?



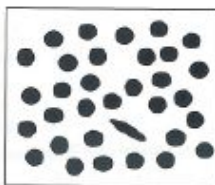


# Pre-attentive?





Grey value



Elongation



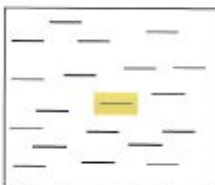
Curvature



Added surround box



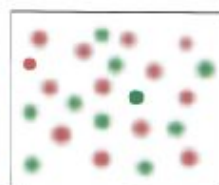
Shape



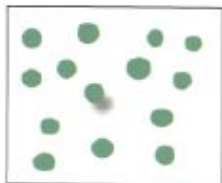
Added surround color



Filled



Sharpness



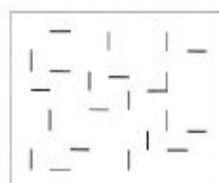
Cast shadow



Convex and concave



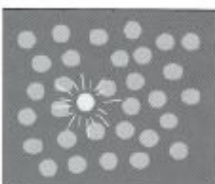
Sharp vertex



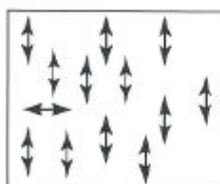
Joined lines



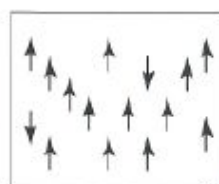
Misalignment



Blinking



Direction of motion



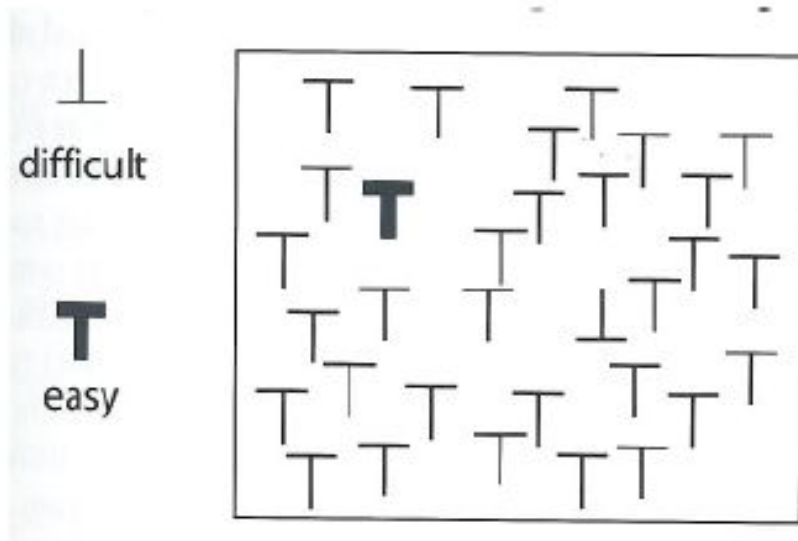
Phase of motion

C.

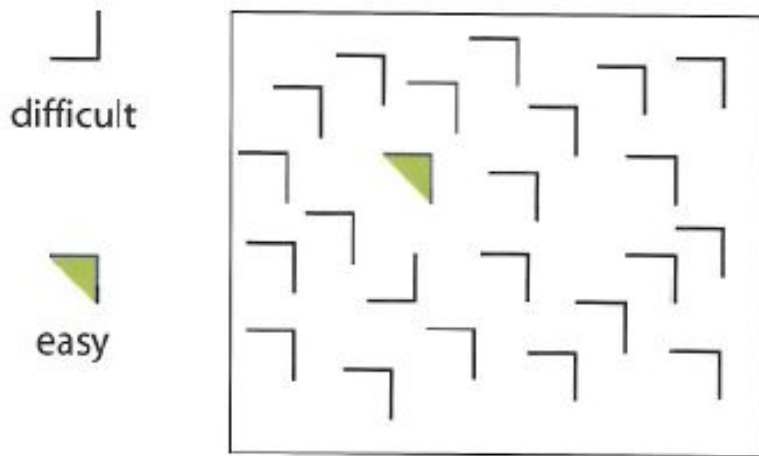
# Pre-attentive?



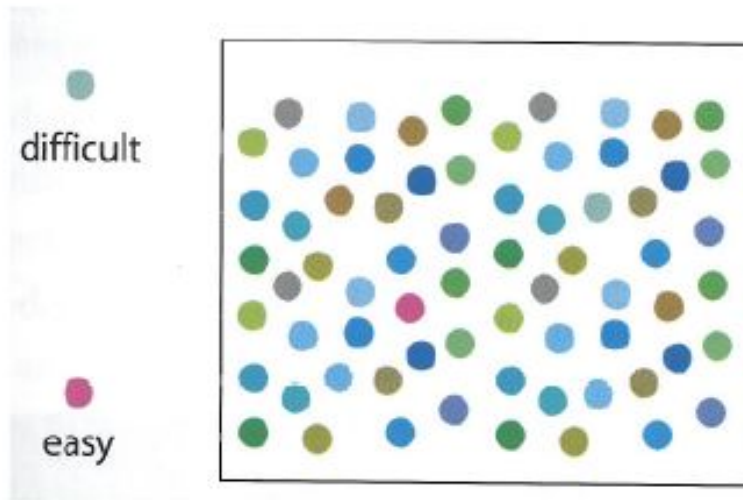
# Similar Feature Sets and Visual Search



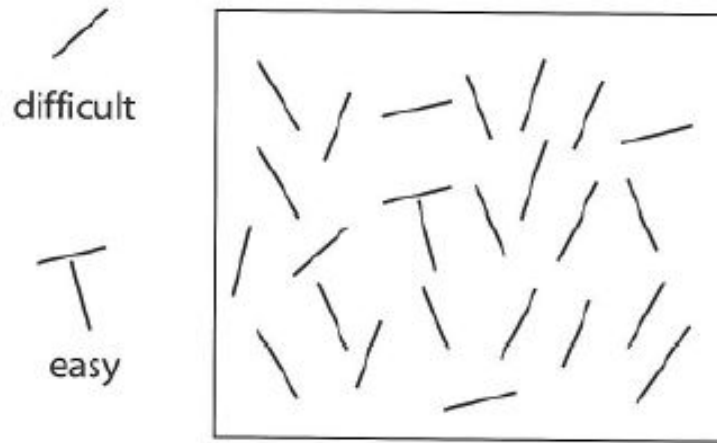
# Similar Feature Sets and Visual Search



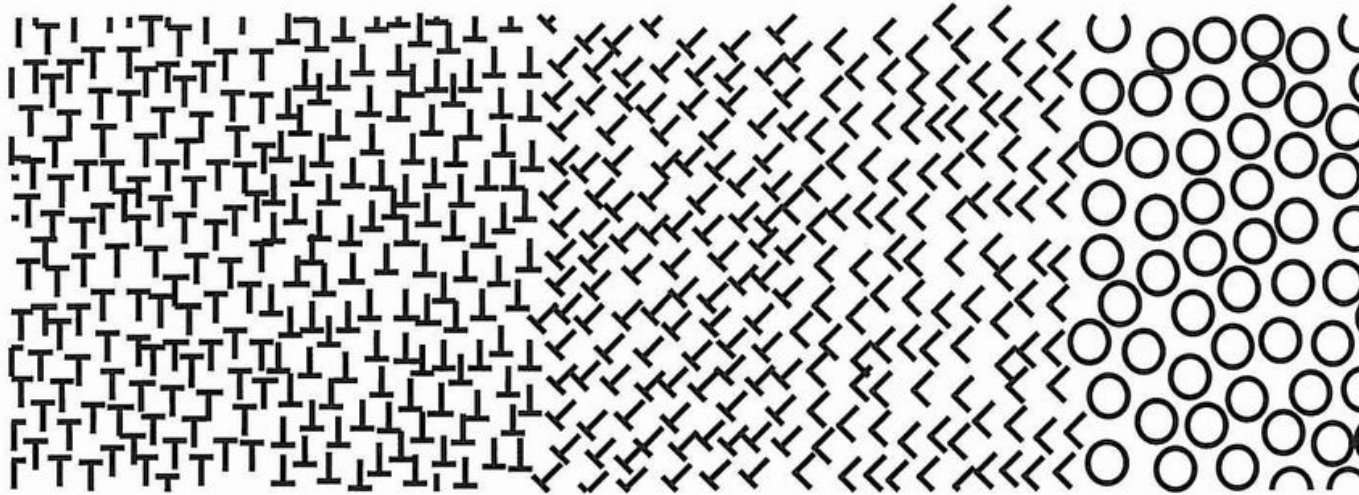
# Similar Feature Sets and Visual Search



# Similar Feature Sets and Visual Search

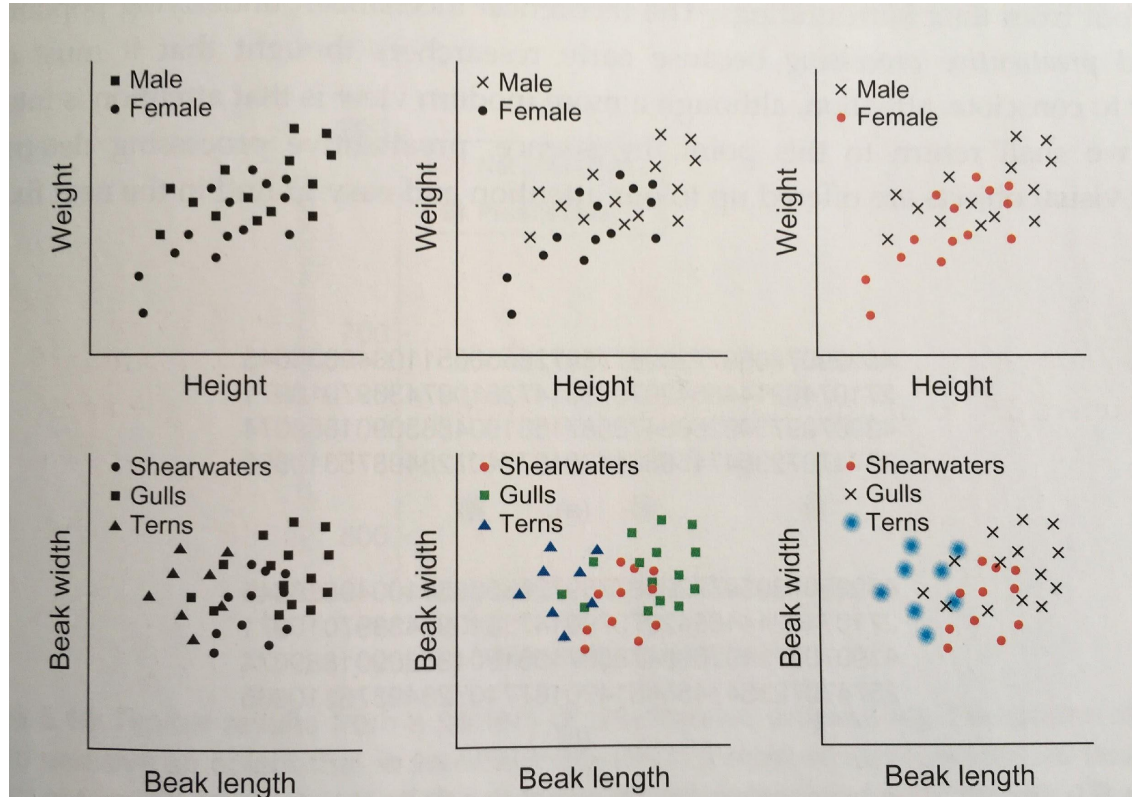


# Do You See 5 Boundaries?





# Distinct Features



# Gestalt Laws

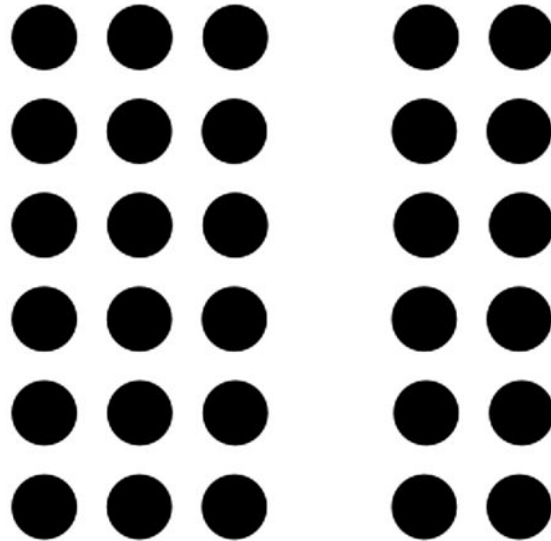
Germany psychologists Max Westheimer, Kurt Koffka, and Wolfgang Kohler from the Gestalt school of psychology undertook to understand how people perceive patterns (1935)

The resulting laws they created serve as valuable principles for displaying information

# Gestalt Laws

- Proximity
- Similarity
- Connectedness
- Continuity
- Symmetry
- Closure

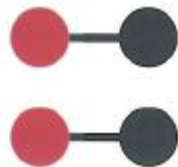
# Gestalt Laws: Proximity



# Gestalt Laws: Connectedness



(a)



(b)

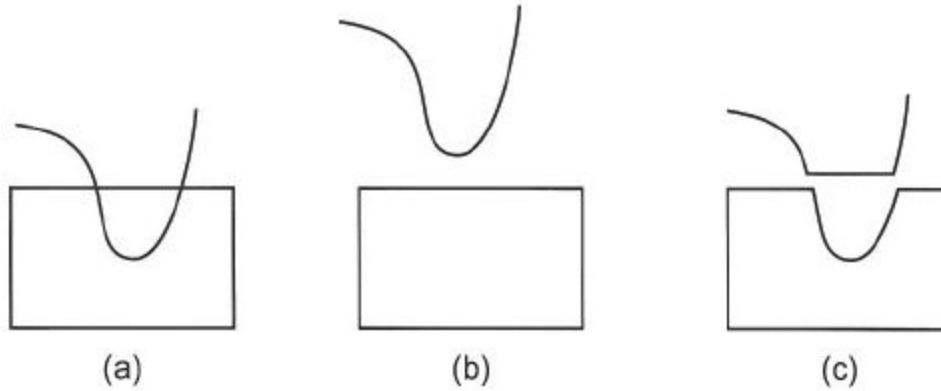


(c)

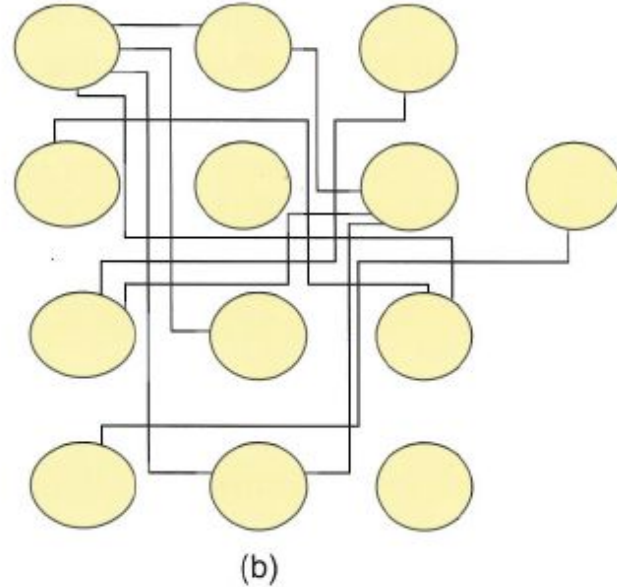
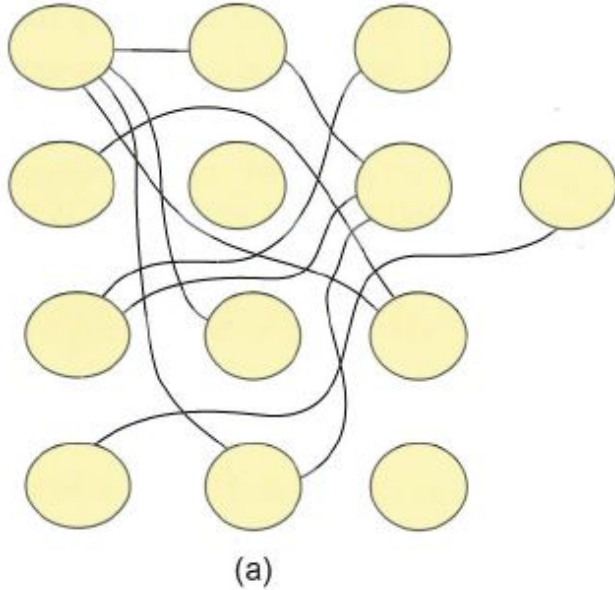


(d)

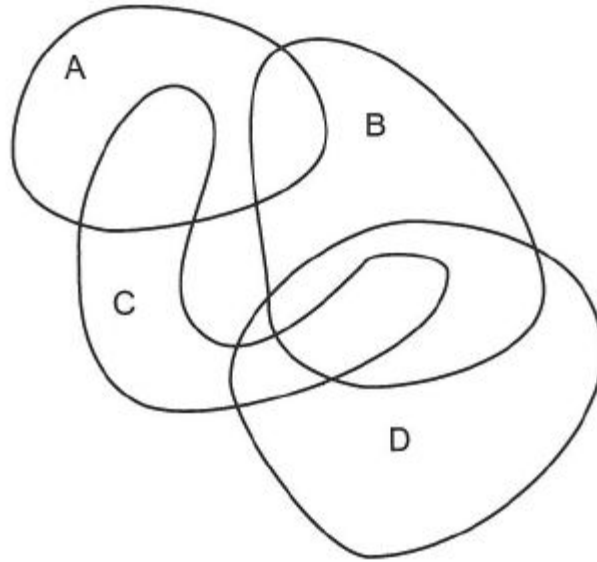
# Gestalt Laws: Continuity



# Gestalt Laws: Continuity

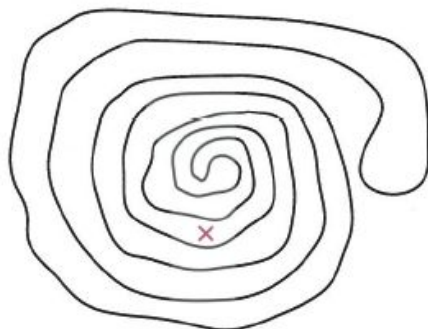


# Gestalt Laws: Closure

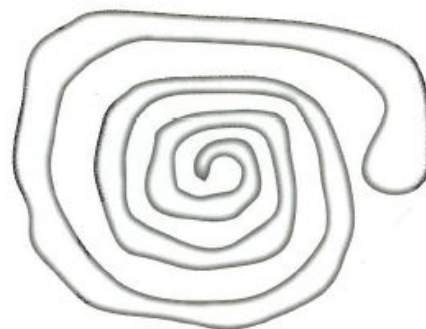




# Gestalt Laws: Closure



(a)



(b)

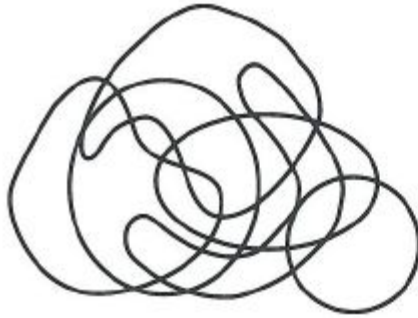


(c)

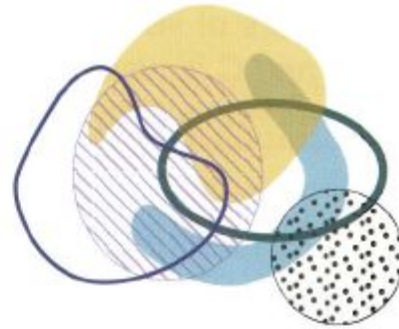


(d)

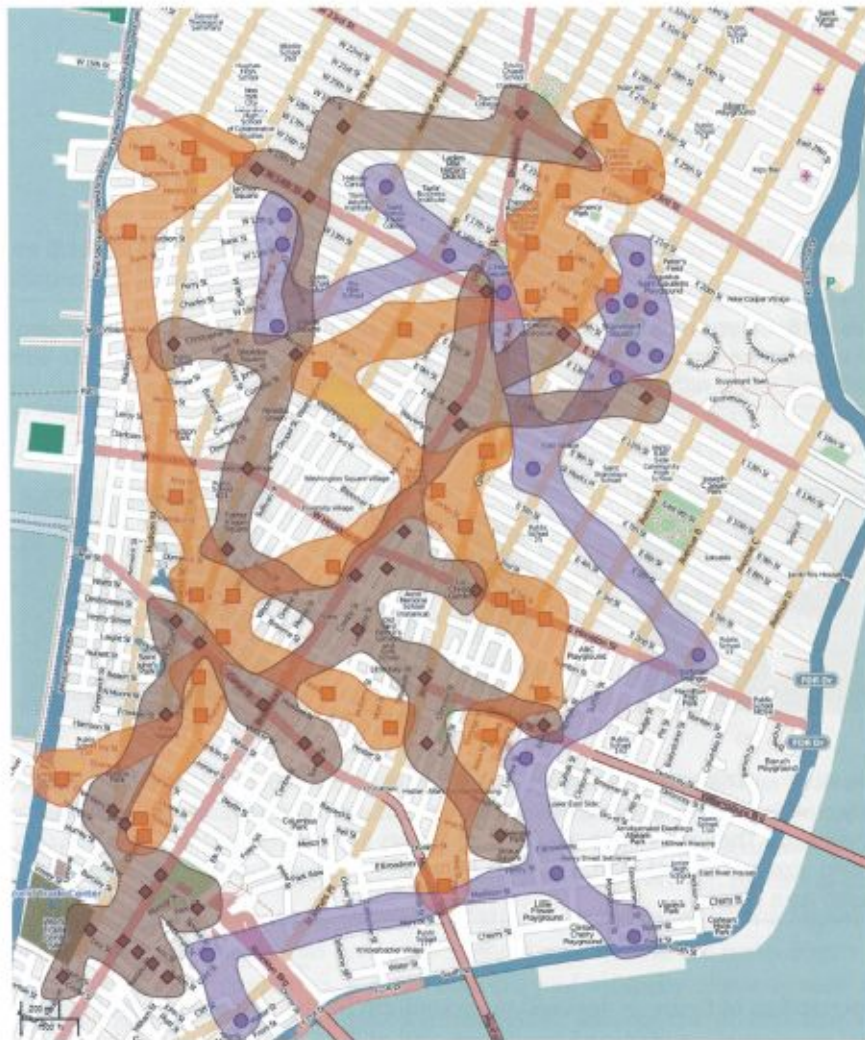
# Gestalt Laws: Closure



(a)

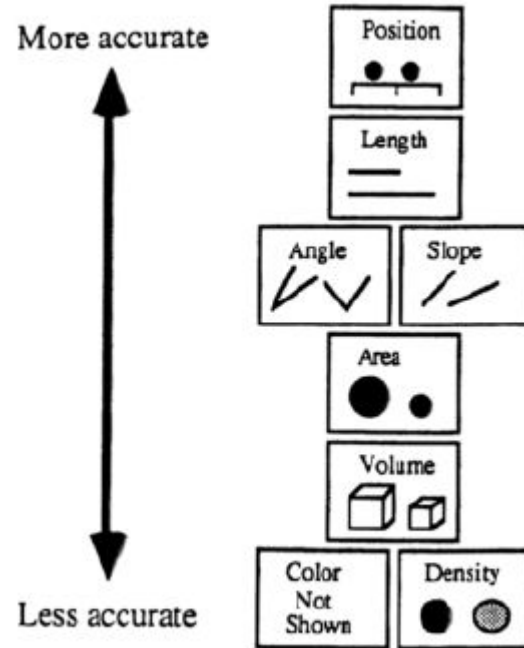


(b)

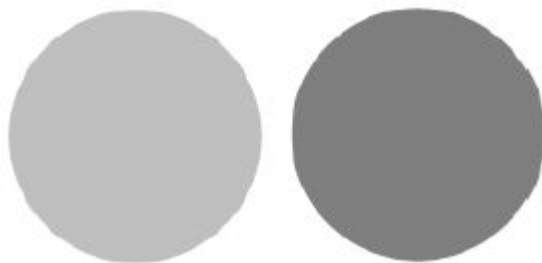


# Elementary Graphical Perception

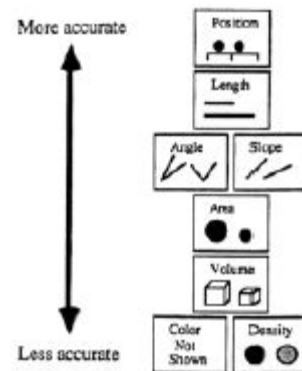
Cleveland and McGill tested ten graphical elements for accuracy of perception



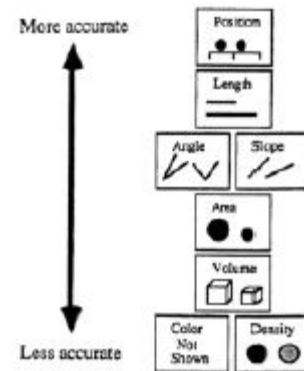
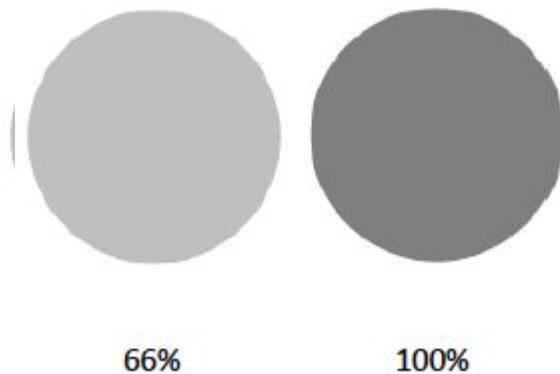
# Density



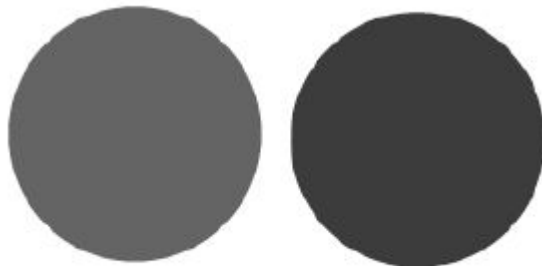
100%



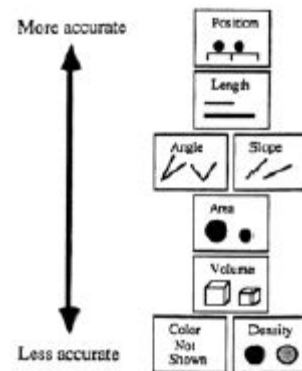
# Density



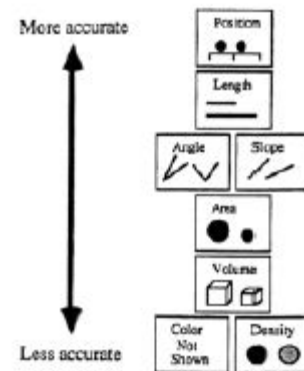
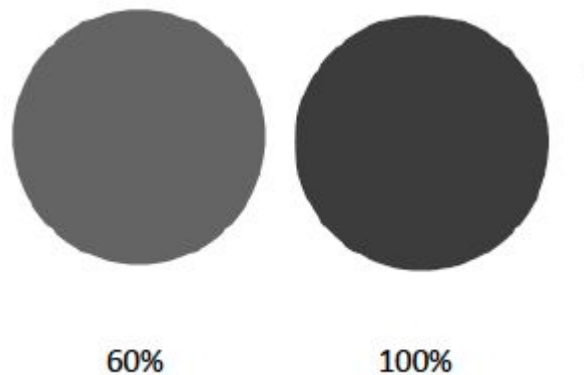
# Density



100%

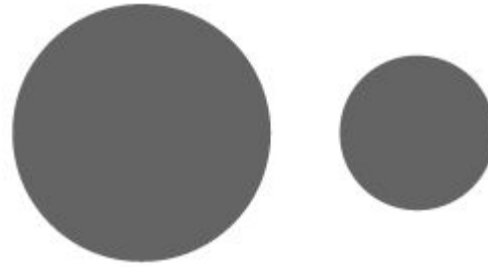


# Density

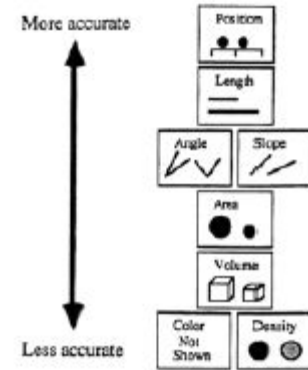




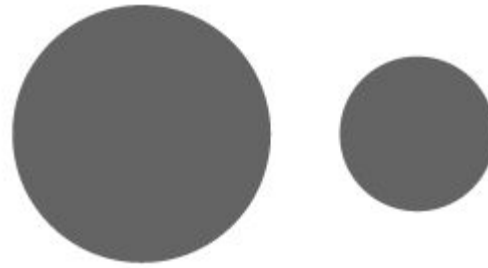
# Area



100%

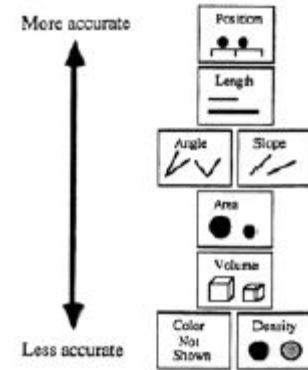


# Area

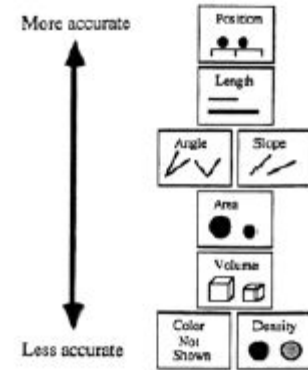
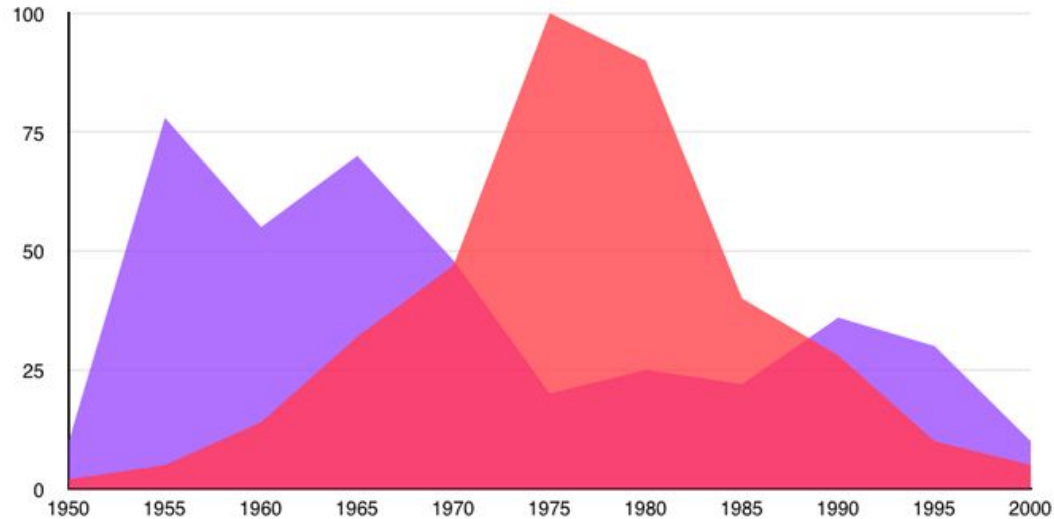


100%

36%



# Area

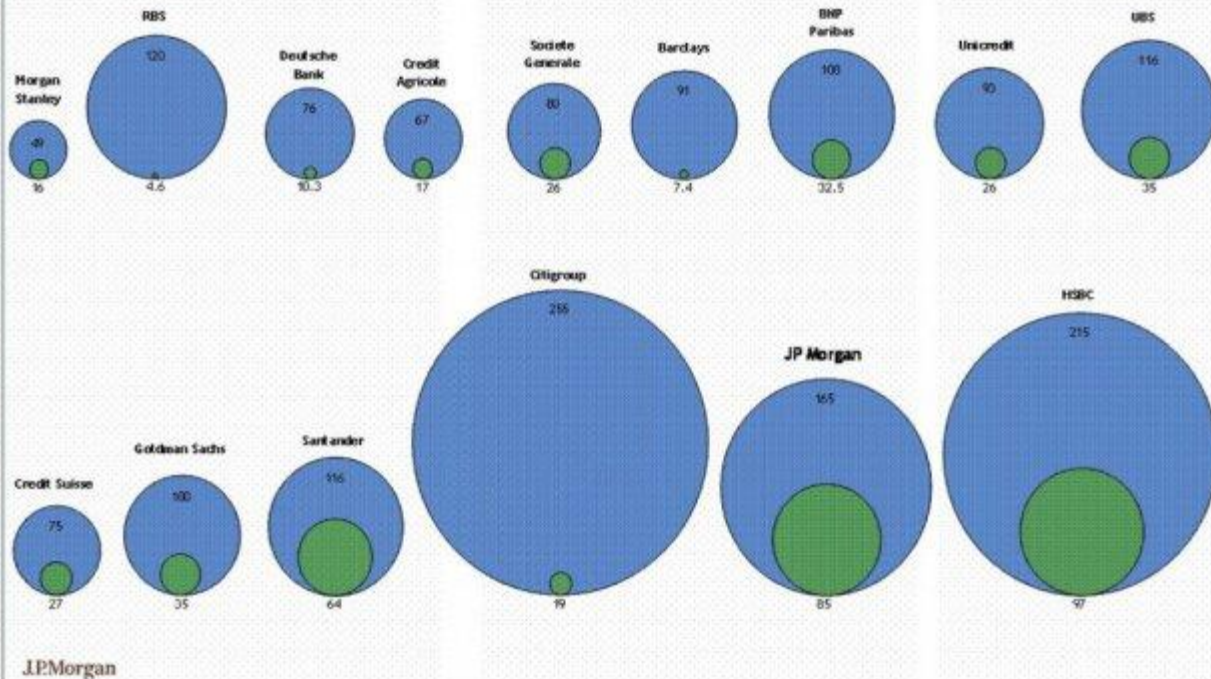


What percentage in size is the red area compared to the purple area?

## Banks: Market Cap

● Market Value as of January 20<sup>th</sup> 2009, \$Bn

● Market Value as of Q2 2007, \$Bn

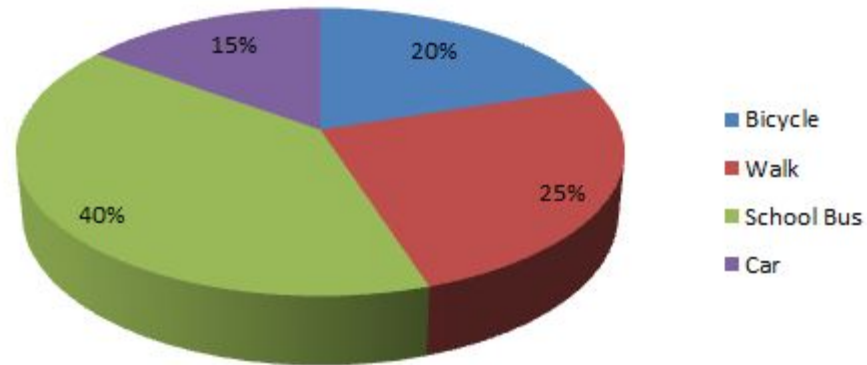


J.P.Morgan

While JPMorgan considers this information to be reliable, we cannot guarantee its accuracy or completeness.

Source: Bloomberg, Jan 20<sup>th</sup> 2009

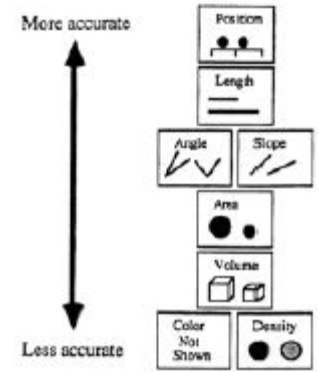
## Method of Transportation to school



Source: [tutorvista.com](http://tutorvista.com)

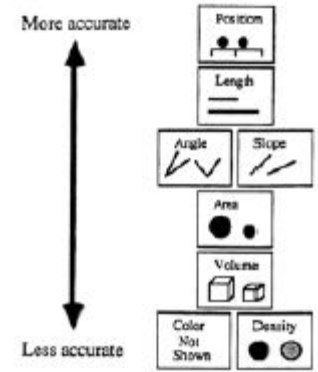
# Length

100%



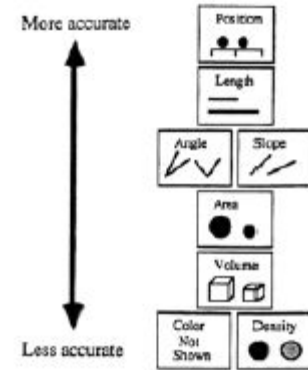
# Length

100%  
75%



# Length

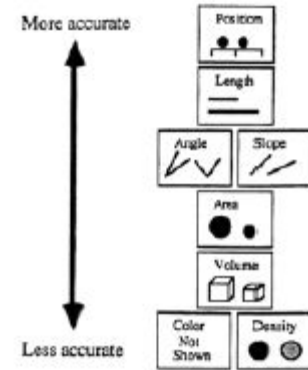
100%



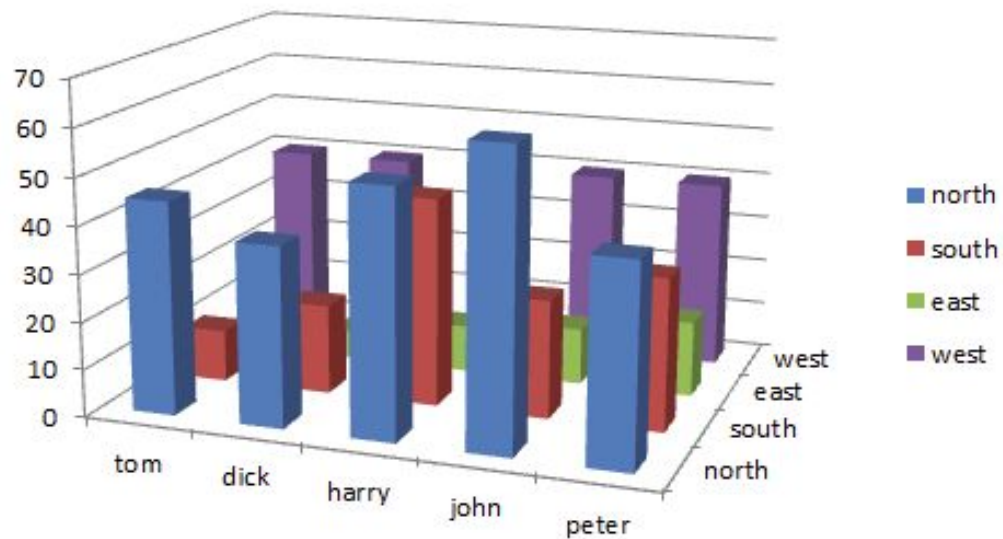


# Length

100%  
25%

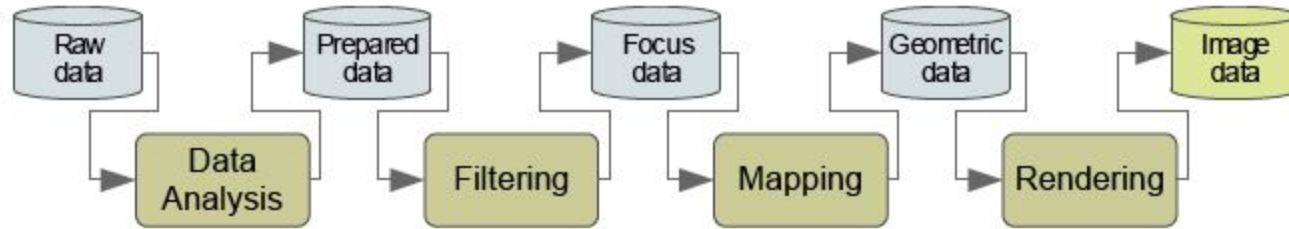


# What Do You Think?



# **GUIDELINES FOR VISUALIZATION**

# The Visualization Pipeline



# Tufte's Principles of Graphical Excellence

## *Graphical excellence*

- The well-designed presentation of interesting data – a matter of substance, statistics, and design
- Consists of complex ideas communicated with clarity, precision, and efficiency
- Gives the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space
- Is nearly always multivariate
- Tells the truth about the data

# Tufte's Principles of Graphical Integrity

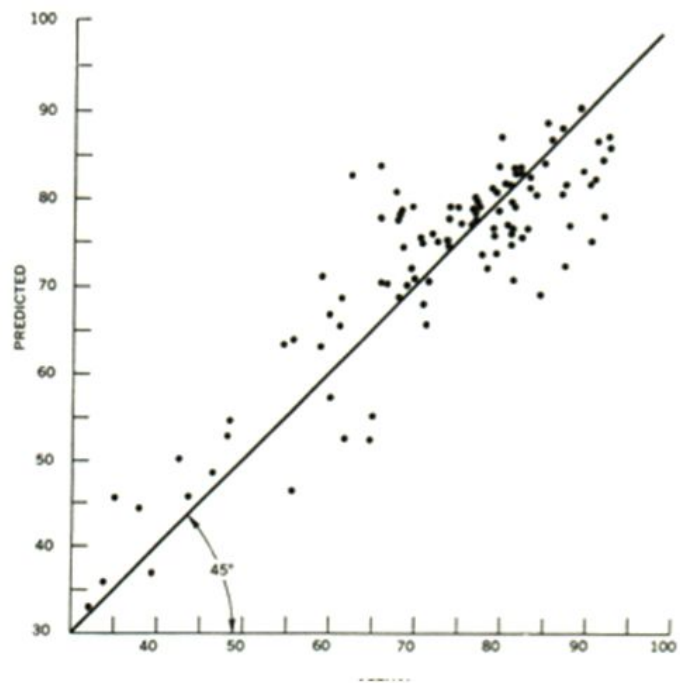
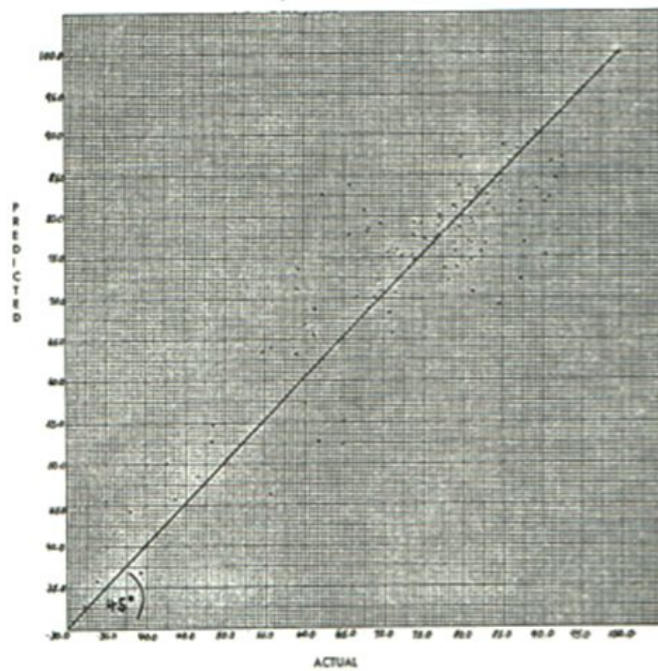
## *Graphical integrity*

- The representation of numbers, as physically measured on the surface of the graphic itself, should be directly proportional to the numerical quantities represented
- Clear, detailed, and thorough labeling should be used to prevent distortion and ambiguity
- Show data variation, not design variation
- Graphics should not quote data out of context
- The number of information-carrying (variable) dimensions depicted should not exceed the dimensions of the data

# Tufte's Principles of Data Graphics

- Above all else show the data
- Maximize the data-ink ratio
- Erase non-data-ink
- Erase redundant data-ink
- Revise and edit

Relationship of Actual Rates of Registration to Predicted Rates  
(104 cities 1960).





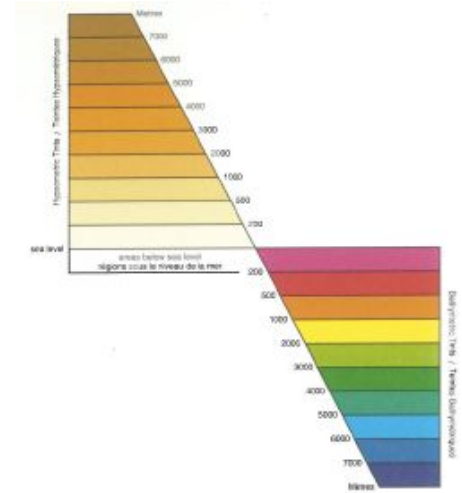
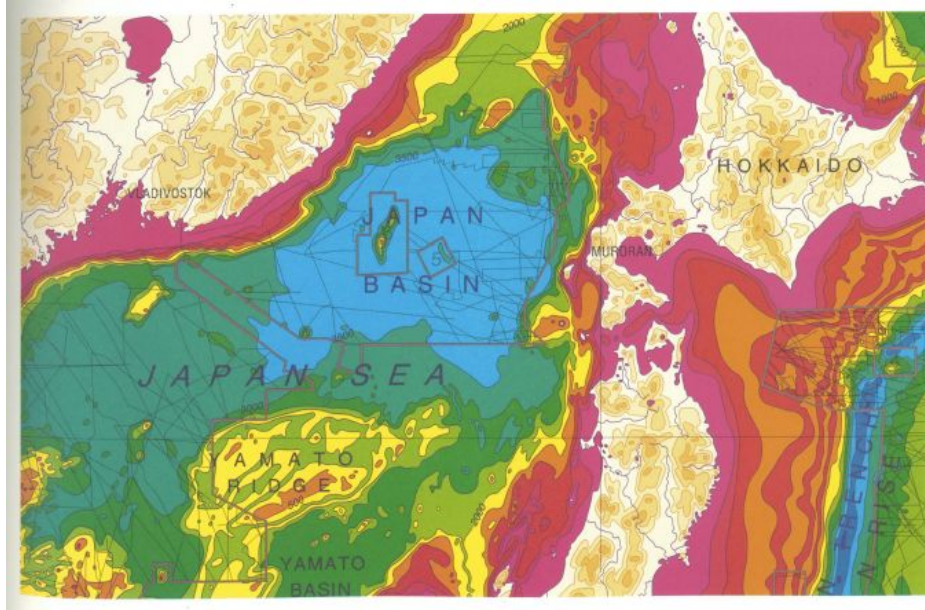
# The Friendly Data Graphic (Tufte)

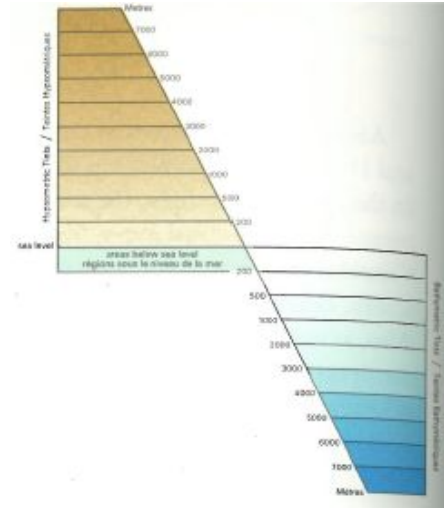
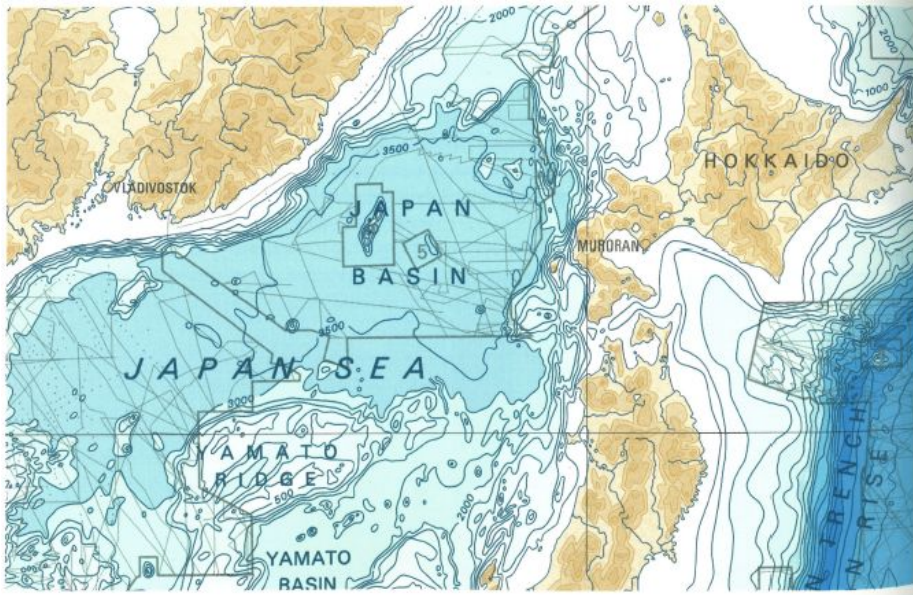
- Words are spelled out, elaborate encoding avoided
- Words run from left to right
- Little messages to help explain data
- Labels are placed on the graphics, no legend required
- Graphics attract viewer, provoke curiosity
- Colors, if used, are chosen so that the color-deficient and color-blind can make sense of the graphic
- Type is clear, precise, modest
- Type is upper-and-lower case, with serifs

# Smallest Effective Difference

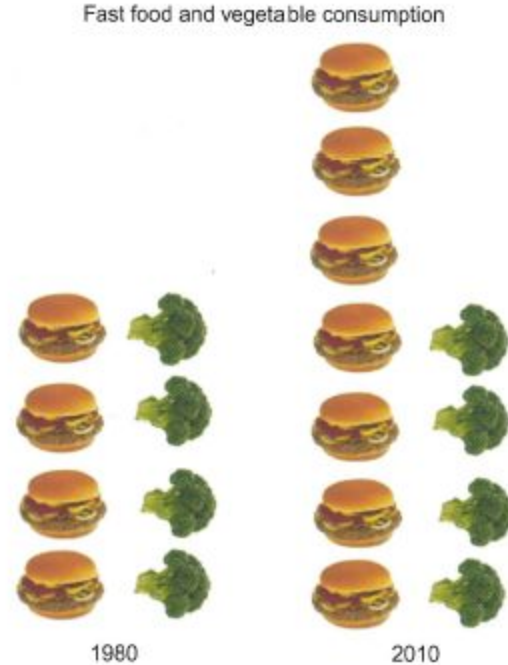
*“In designing information... the idea is to use just notable differences, visual elements that make a clear difference but no more – contrasts that are definite, effective, and minimal.”*

-Tufte 1997





# Pictures, Symbols, or Words



# Pictures, Symbols, or Words

- Pictorial icons for pedagogical purposes in infographics
  - Only when a canonical or culturally defined image is available
- Symbols when a large number of data points must be represented
- Words directly on charts where the number of symbolic objects in each category is few and space is available

# Pictures, Symbols, or Words





# What Do You Think?



Source: [cuzproduces.com](http://cuzproduces.com)



# Pitfalls

- Selecting the wrong data
- Selecting the wrong data structure
- Filtering out important data
- Failed understanding of the types of things that need to be shown
- Selecting the wrong representation
- Choosing the wrong presentation format

# Hans Rosling Visualization

TED talk: <https://www.youtube.com/watch?v=hVimVzgtD6w>

# Concluding thoughts

- Information is powerful and abundant
- Information alone is not effective for discovering and communicating ideas, trends, and insights
- Basic understanding of visualization is beneficial to anyone who deals with data or communication

# Reading Assignment

- Required reading for next week: Chapters 1 and 4 of *The Design of Everyday Things* by Donald Norman
- This week's lab and reading assignment available for download on OLAT