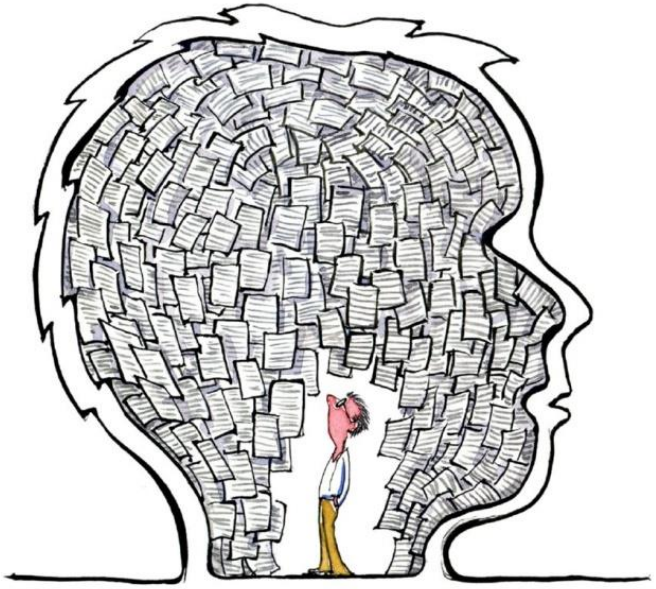
Introduction to Visualization

COMP8503

Advanced Topics in Visual Analytics

What is Visualization?

- The communication of information using graphical representation.
- “A picture is worth a thousand words.”



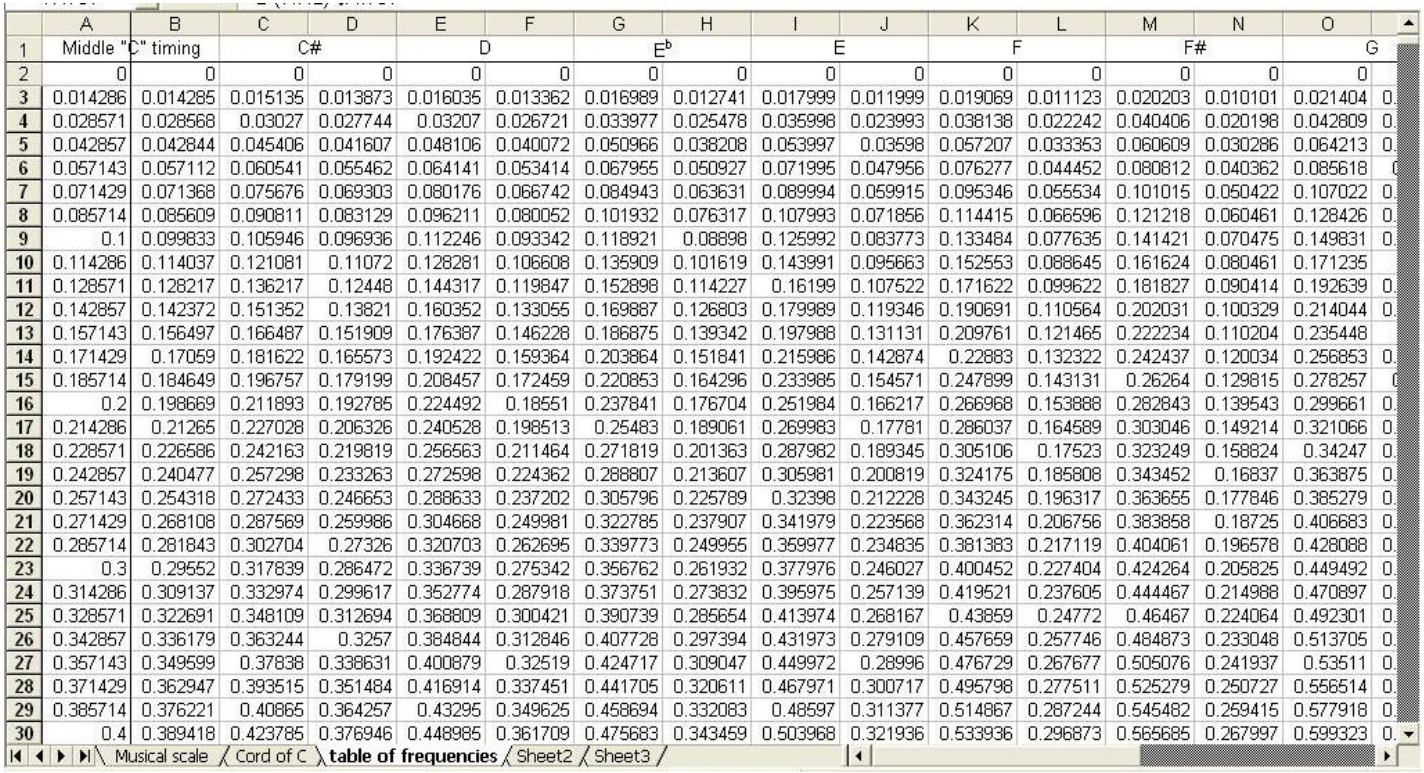
Images vs. Text

- Images/pictures
  - Interpreted in parallel by the human perceptual system
  - Independent of local language
- Text/words
  - Sequential process of reading
  - Limited by the use of different languages

The Information Age

- We are now at the Information Age (or Digital Era), there is rapid emergence of new data capture and generation technologies such as sensor networks.
- IDC (International Data Corporation) estimates that “The digital universe will reach 4ZB, almost 50% more than 2012 volumes and nearly a quadrupling of 2010 volumes.”
- 4ZB = 240 million times the content of the Library of Congress in the US
- Data needs interpretation (explain process and concepts)
- Visualization as an effective tool to assist in analysis and communication

Data in Tabular Form?



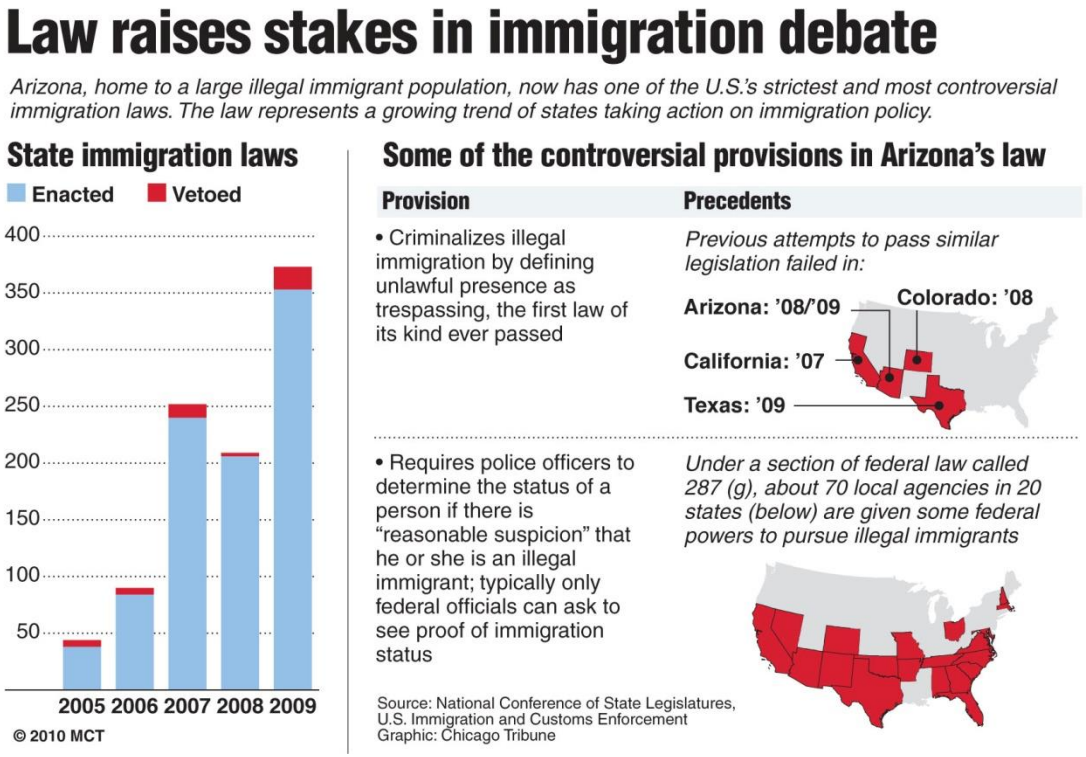
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Middle "C" timing		C#		D		E^b		E		F		F#		G
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0.014286	0.014285	0.015135	0.013873	0.016035	0.013362	0.016989	0.012741	0.017999	0.011999	0.019069	0.011123	0.020203	0.010101	0.021404
4	0.028571	0.028568	0.03027	0.027744	0.03207	0.026721	0.033977	0.025478	0.035998	0.023993	0.038138	0.022242	0.040406	0.020198	0.042809
5	0.042857	0.042844	0.045406	0.041607	0.048106	0.040072	0.050966	0.038208	0.053997	0.03598	0.057207	0.033353	0.060609	0.030286	0.064213
6	0.057143	0.057112	0.060541	0.055462	0.064141	0.053414	0.067955	0.050927	0.071995	0.047956	0.076277	0.044452	0.080812	0.040362	0.085618
7	0.071429	0.071368	0.075676	0.069303	0.080176	0.066742	0.084943	0.063631	0.089994	0.059915	0.095346	0.055534	0.101015	0.050422	0.107022
8	0.085714	0.085609	0.090811	0.083129	0.096211	0.080052	0.101932	0.076317	0.107993	0.071856	0.114415	0.066596	0.121218	0.060461	0.128426
9	0.1	0.099833	0.105946	0.096936	0.112246	0.093342	0.118921	0.08898	0.125992	0.083773	0.133484	0.077635	0.141421	0.070475	0.149831
10	0.114286	0.114037	0.121081	0.11072	0.128281	0.106608	0.135909	0.101619	0.143991	0.095663	0.152553	0.088645	0.161624	0.080461	0.171235
11	0.128571	0.128217	0.136217	0.12448	0.144317	0.119847	0.152898	0.114227	0.16199	0.107522	0.171622	0.099622	0.181827	0.090414	0.192639
12	0.142857	0.142372	0.151352	0.13821	0.160352	0.133055	0.169887	0.126803	0.179989	0.119346	0.190691	0.110564	0.202031	0.100329	0.214044
13	0.157143	0.156497	0.166487	0.151909	0.176387	0.146228	0.186875	0.139342	0.197988	0.131131	0.209761	0.121465	0.222234	0.110204	0.235448
14	0.171429	0.17059	0.181622	0.165573	0.192422	0.159364	0.203864	0.151841	0.215986	0.142874	0.22883	0.132322	0.242437	0.120034	0.256853
15	0.185714	0.184649	0.196757	0.179199	0.208457	0.172459	0.220853	0.164296	0.233985	0.154571	0.247899	0.143131	0.26264	0.129815	0.278257
16	0.2	0.198669	0.211893	0.192785	0.224492	0.18551	0.237841	0.176704	0.251984	0.166217	0.266968	0.153888	0.282843	0.139543	0.299661
17	0.214286	0.21265	0.227028	0.206326	0.240528	0.198513	0.25483	0.189061	0.269983	0.17781	0.286037	0.164589	0.303046	0.149214	0.321066
18	0.228571	0.226586	0.242163	0.219819	0.256563	0.211464	0.271819	0.201363	0.287982	0.189345	0.305106	0.17523	0.323249	0.158824	0.34247
19	0.242857	0.240477	0.257298	0.233263	0.272598	0.224362	0.288807	0.213607	0.305981	0.200819	0.324175	0.185808	0.343452	0.16837	0.363875
20	0.257143	0.254318	0.272433	0.246653	0.288633	0.237202	0.305796	0.225789	0.32398	0.212228	0.343245	0.196317	0.363655	0.177846	0.385279
21	0.271429	0.268108	0.287569	0.259986	0.304668	0.249981	0.322785	0.237907	0.341979	0.223568	0.362314	0.206756	0.383858	0.18725	0.406683
22	0.285714	0.281843	0.302704	0.27326	0.320703	0.262695	0.339773	0.249955	0.359977	0.234835	0.381383	0.217119	0.404061	0.196578	0.428088
23	0.3	0.29552	0.317839	0.286472	0.336739	0.275342	0.356762	0.261932	0.377976	0.246027	0.400452	0.227404	0.424264	0.205825	0.449492
24	0.314286	0.309137	0.332974	0.299617	0.352774	0.287918	0.373751	0.273832	0.395975	0.257139	0.419521	0.237605	0.444467	0.214988	0.470897
25	0.328571	0.322691	0.348109	0.312694	0.368809	0.300421	0.390739	0.285654	0.413974	0.268167	0.43859	0.24772	0.46467	0.224064	0.492301
26	0.342857	0.336179	0.363244	0.3257	0.384844	0.312846	0.407728	0.297394	0.431973	0.279109	0.457659	0.257746	0.484873	0.233048	0.513705
27	0.357143	0.349599	0.37838	0.338631	0.400879	0.32519	0.424717	0.309047	0.449972	0.28996	0.476729	0.267677	0.505076	0.241937	0.53511
28	0.371429	0.362947	0.393515	0.351484	0.416914	0.337451	0.441705	0.320611	0.467971	0.300717	0.495798	0.277511	0.525279	0.250727	0.556514
29	0.385714	0.376221	0.40865	0.364257	0.43295	0.349625	0.458694	0.332083	0.48597	0.311377	0.514867	0.287244	0.545482	0.259415	0.577918
30	0.4	0.389418	0.423785	0.376946	0.448985	0.361709	0.475683	0.343459	0.503968	0.321936	0.533936	0.296873	0.565685	0.267997	0.599323

Musical scale / Cord of C / **table of frequencies** / Sheet2 / Sheet3 /

- Hard to discover relationship, trends, patterns

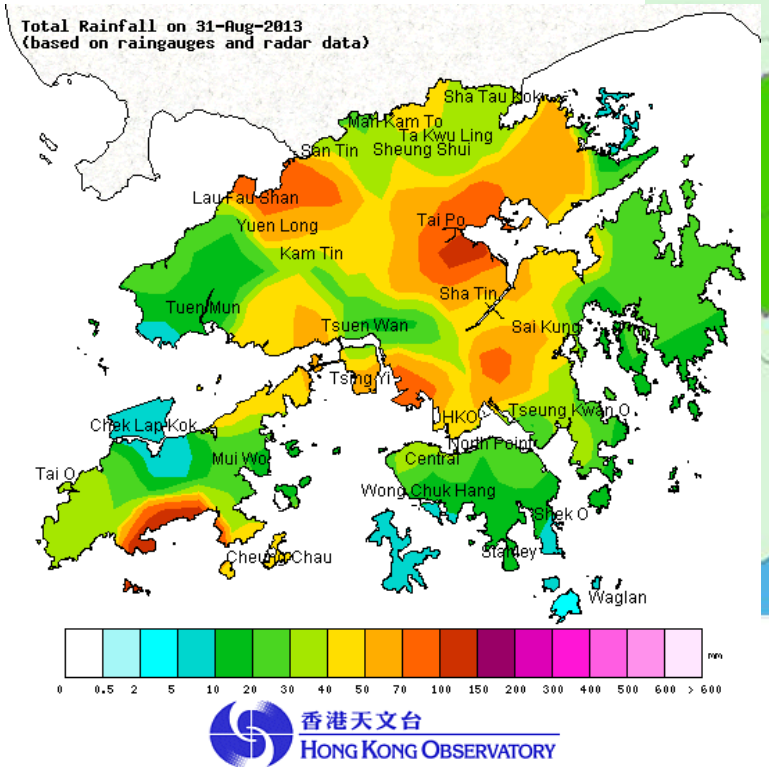
Daily Life Examples of Visualization

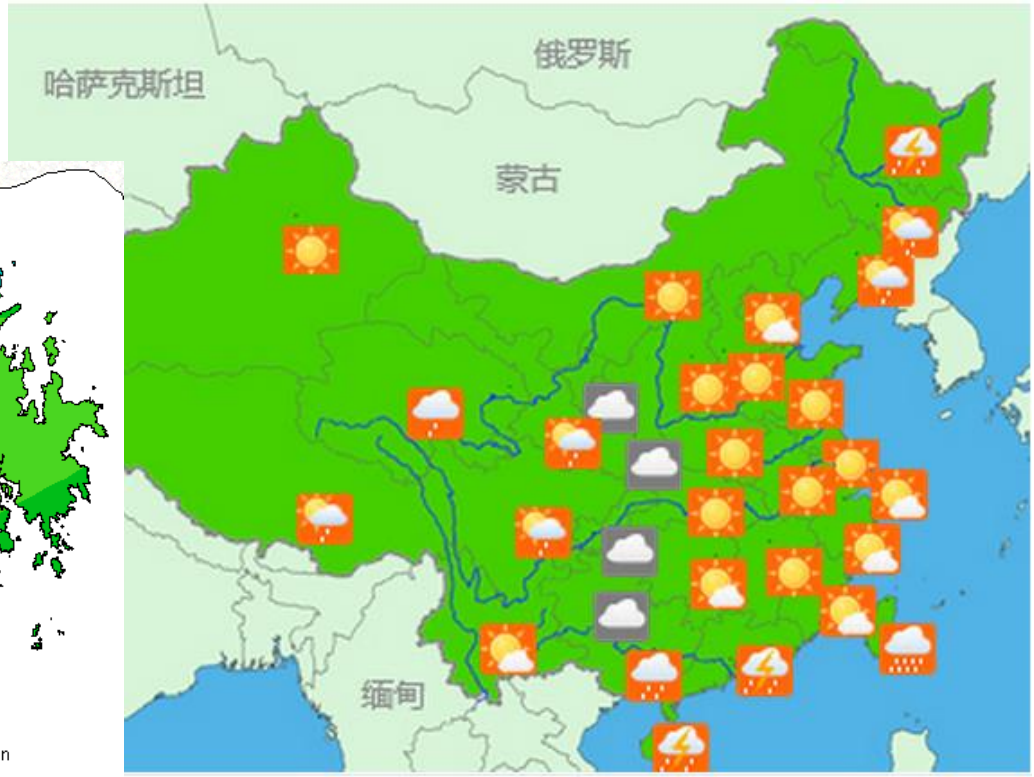
- Newspaper / Magazine – in articles accompanying text to present data



Daily Life Examples of Visualization

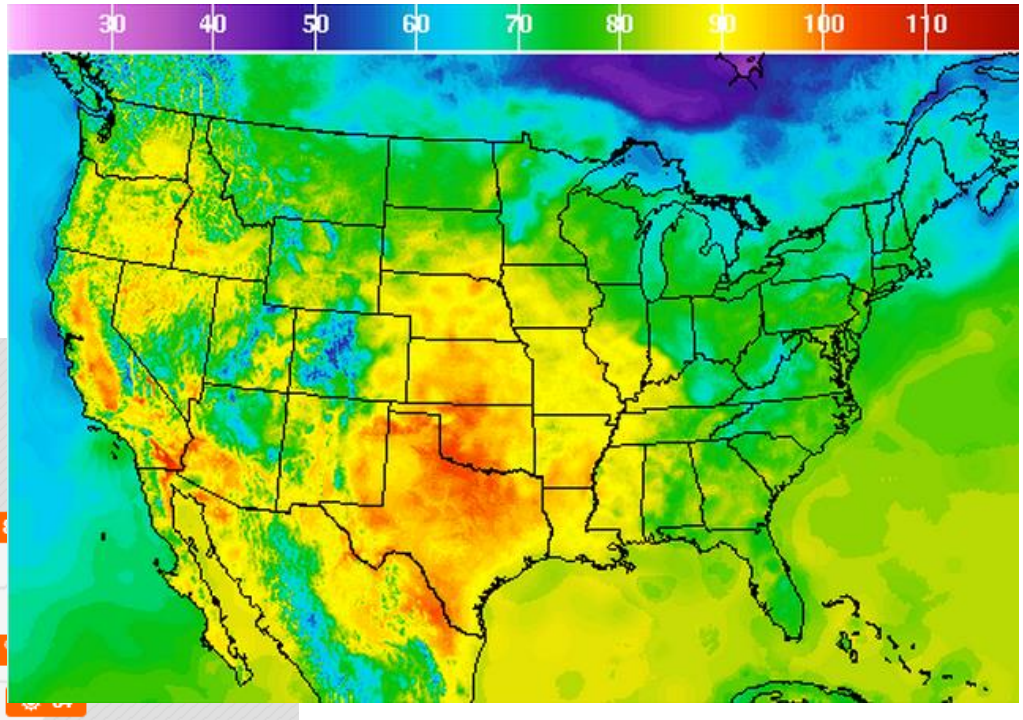
- Weather chart

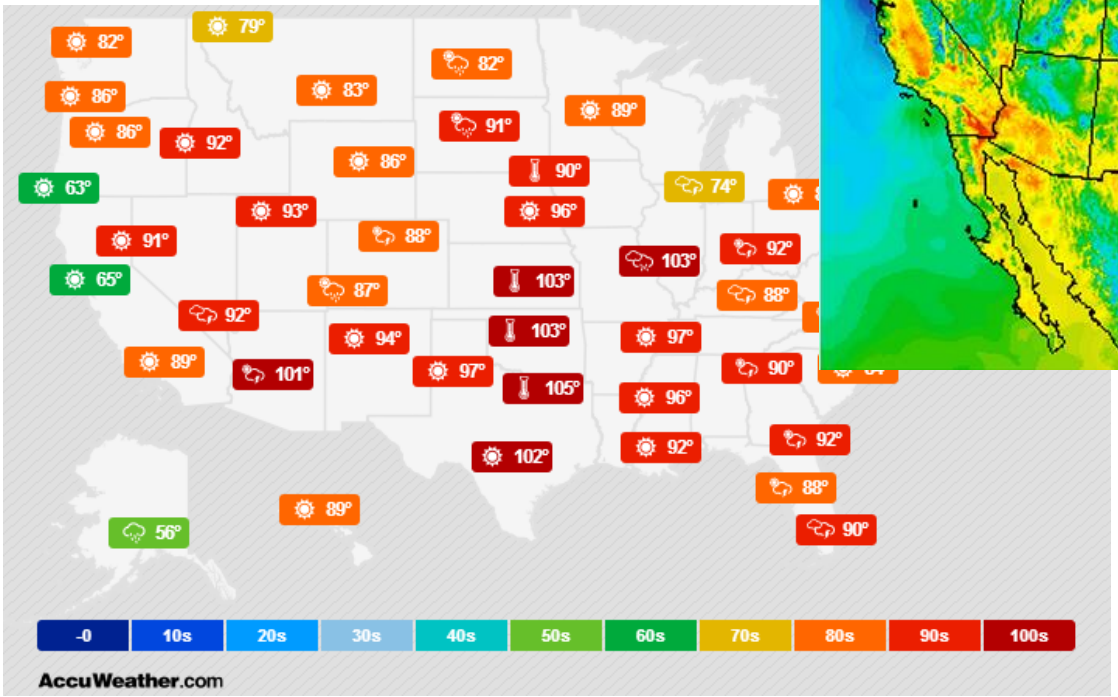




Daily Life Examples of Visualization

- Weather chart





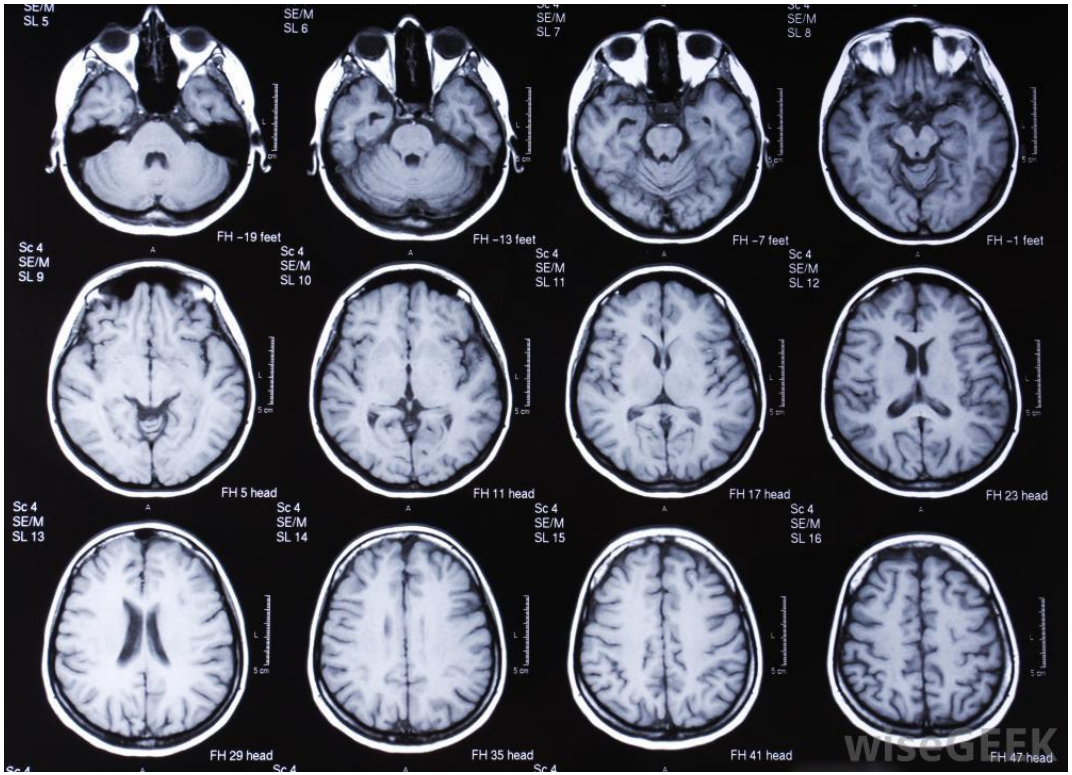
Daily Life Examples of Visualization

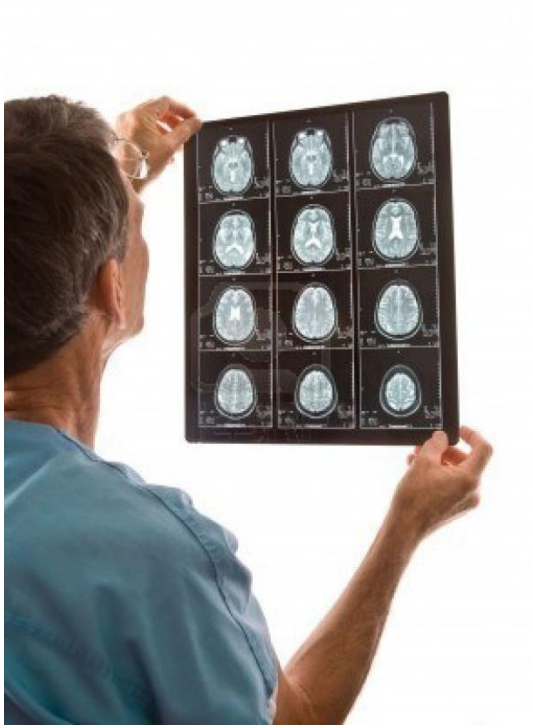
Daily Life Examples of Visualization

- Stock market

Daily Life Examples of Visualization

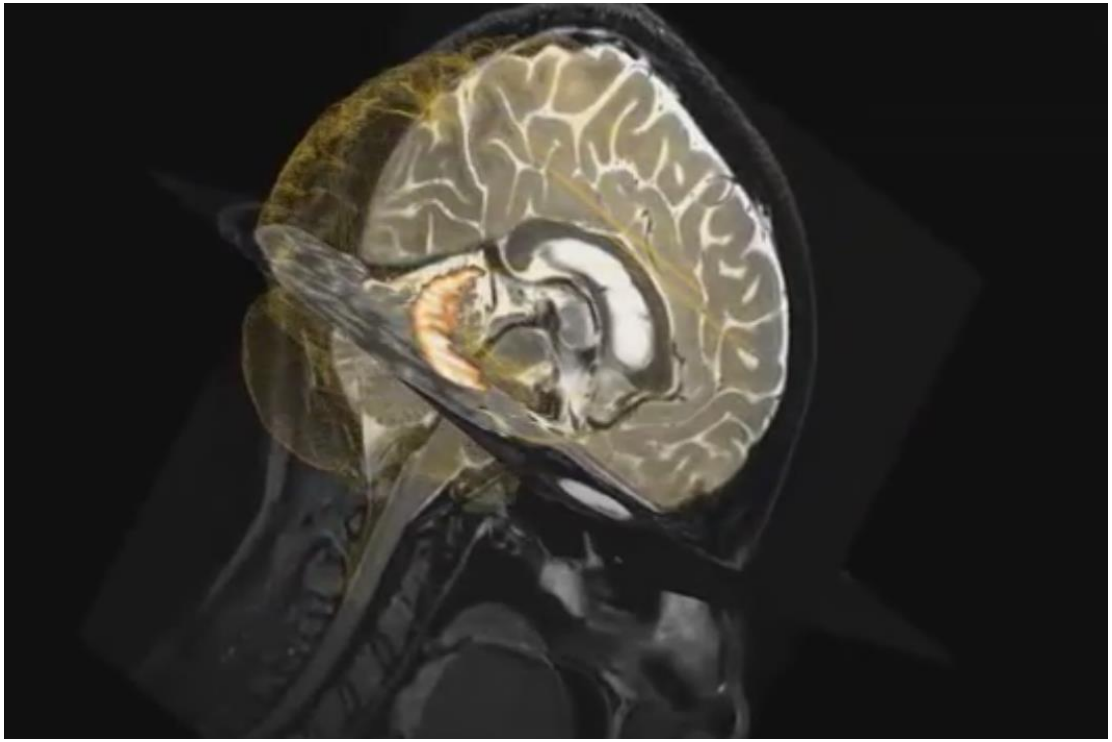
- Medical Care (Brain MRI 2D)





Daily Life Examples of Visualization

- Medical Care (Brain MRI 3D Visualization)



[VUmc Amsterdam, http://www.youtube.com/watch?v=wjvDDH-uJ0s]

Daily Life Examples of Visualization

- Medical care (Ultrasound)

2D

https://www.youtube.com/watch?v=8if_IVf0yHw

3D/4D

https://www.youtube.com/watch?v=nD_l0EtkNYU

Some Questions

- Why visualization can explain better than text?
- Is visualization always effective?
- Is there any type of data/information that cannot be visualized?
- Given the same data, are there different ways to visualize it? Which is the best?

Why Visualization is Important?

- provides an ability to comprehend huge amounts of data
- allows the perception of emergent properties that were not anticipated (new insight)
- often reveals problems with the data itself quickly (anomalies)
- facilitates understanding of both large-scale and small-scale features of the data
- facilitates hypothesis formation

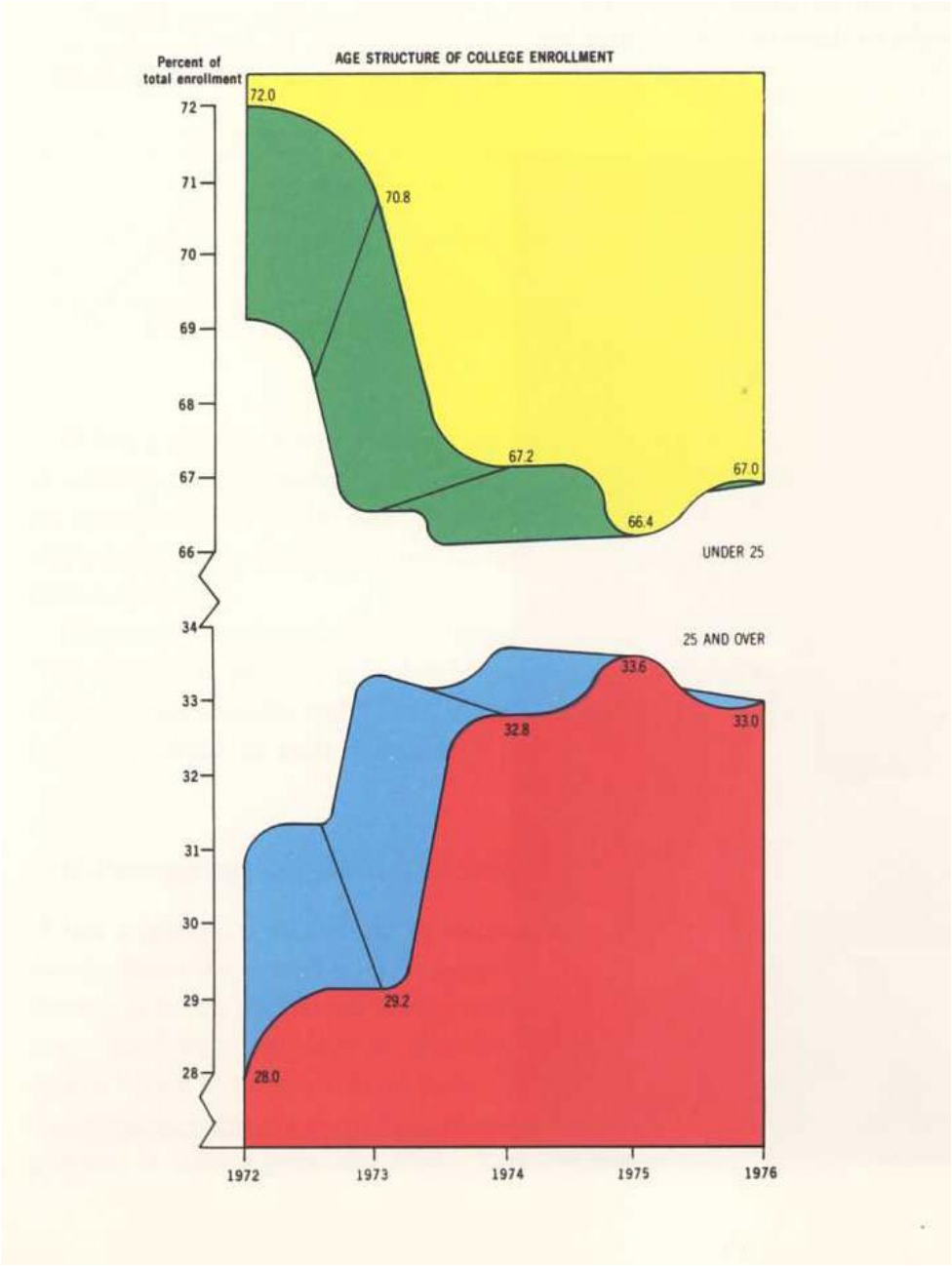
What Should a Good Visualization Achieve?

- Show the data
- Induce the viewer to focus on the substance rather than the methodology, graphic design, the technology of graphic production, etc.
- Avoid distorting the data
- Present many numbers in a small space
- Encourage the eye to compare different pieces of data
- Reveal the data at several levels of detail, from a broad overview to the fine structure
- Serve a reasonably clear purpose: description, exploration, tabulation, or decoration

Some Bad Examples

- Data:
  - the percentage of college students in US that were under 25 from 1972 through 1976

Year	Percentage
1972	72.0
1973	70.8
1974	67.2
1975	66.4
1976	67.0



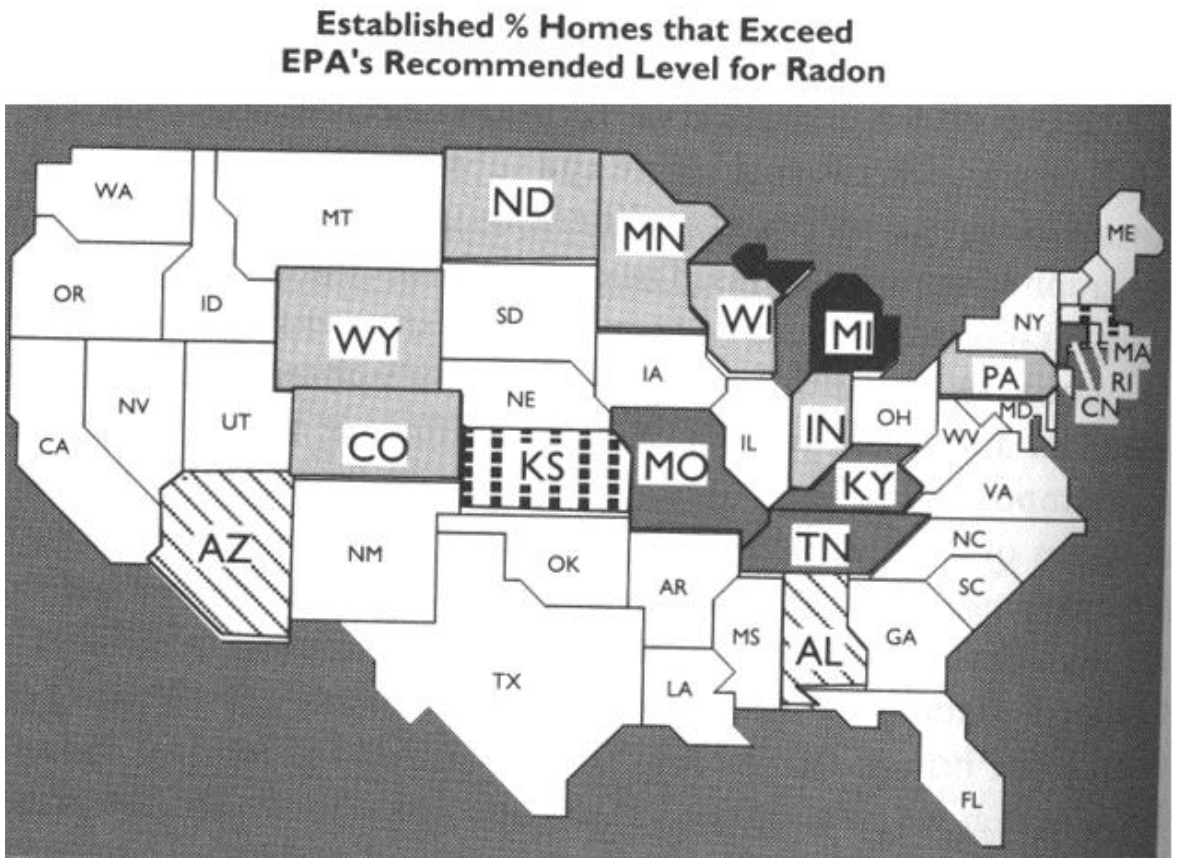
Some Bad Examples

Tufte (1983, p.118) says, "This may be the worst graphic ever to find its way into print."

Simple guideline:

- Use table to show small data set with < 20 numbers
- Use graphic for large data set

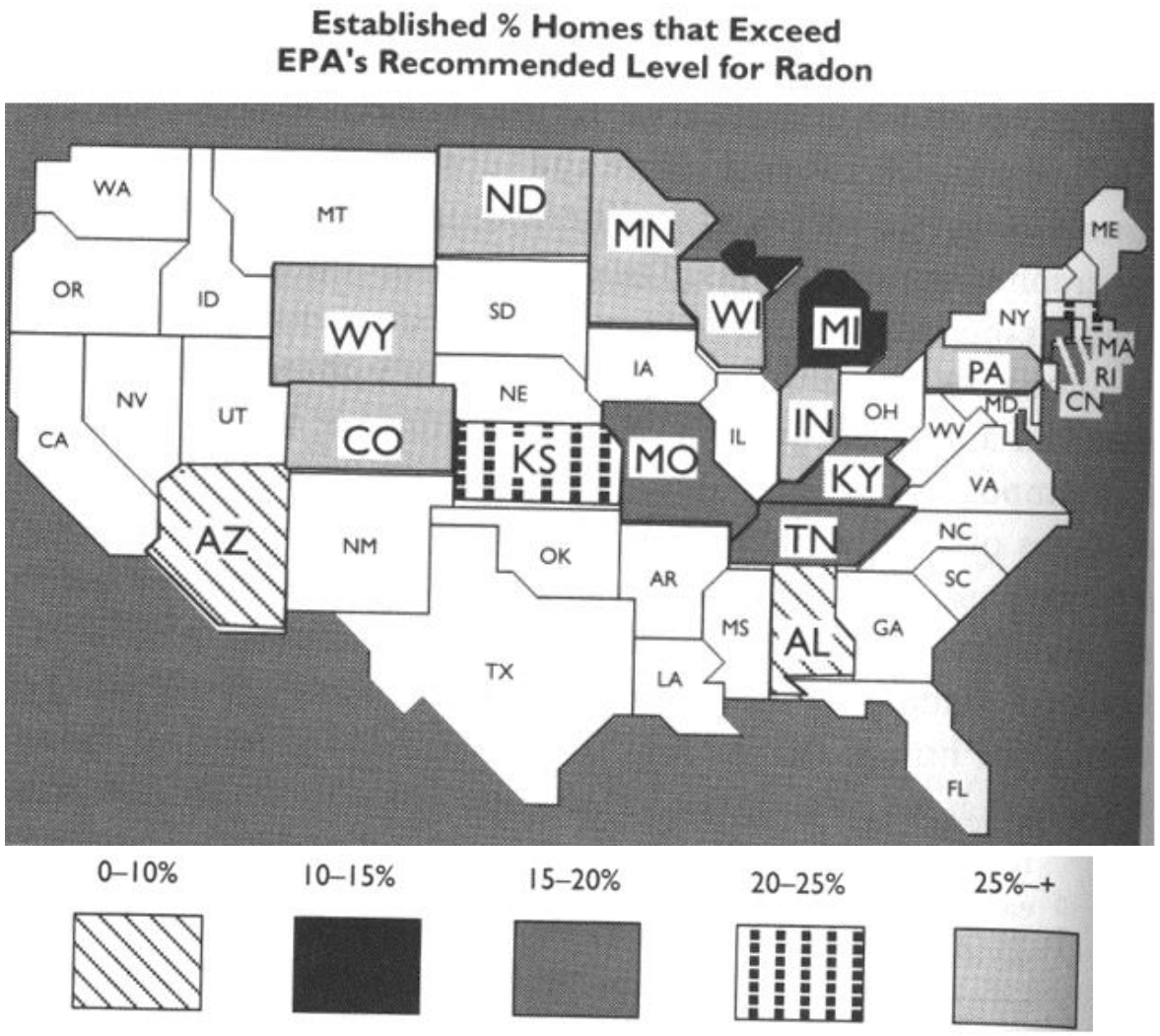
Some Bad Examples



Which state in US has the highest % of homes exceeding the recommended Radon level? Which one has the lowest?

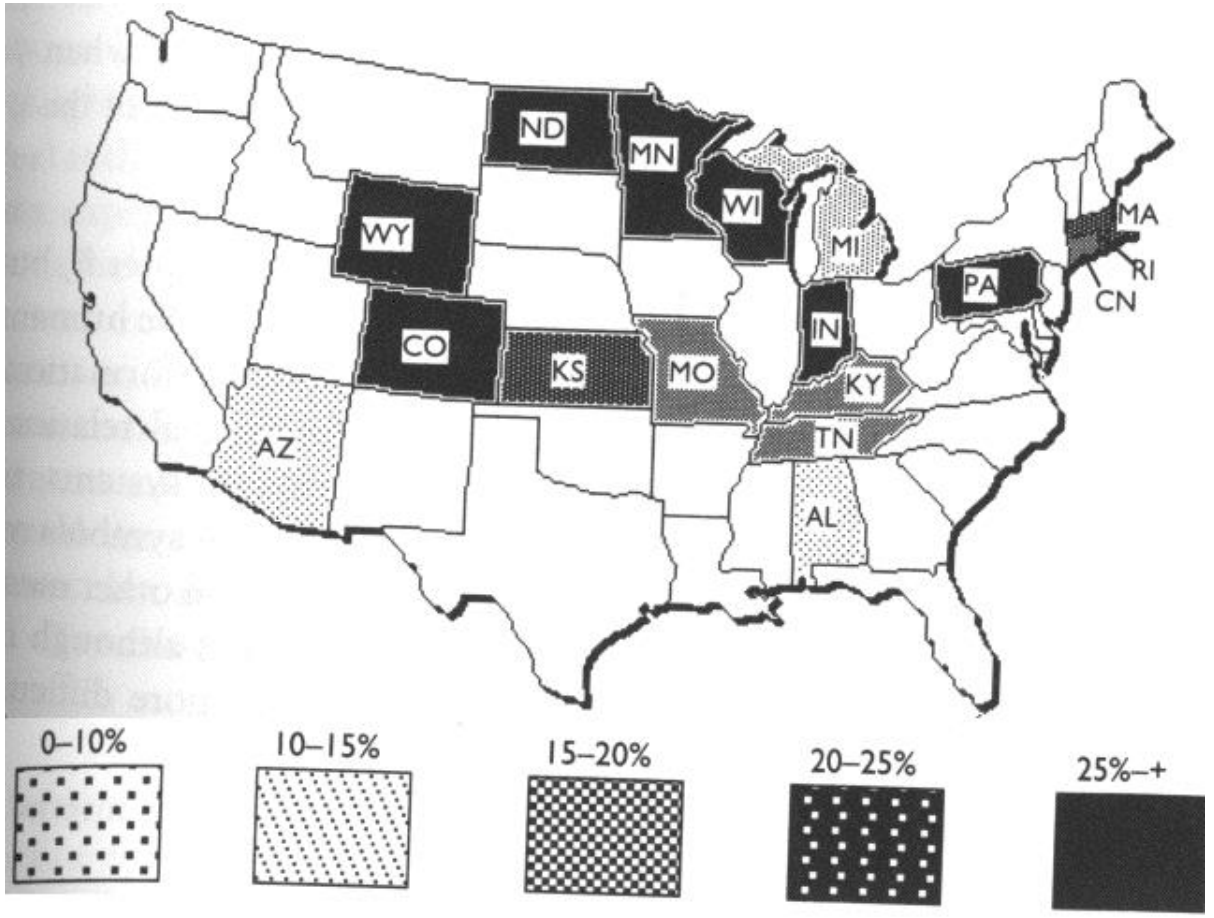
[Things that Make Us Smart, p70-71]

Some Bad Examples



Some Bad Examples

- Better presented this way



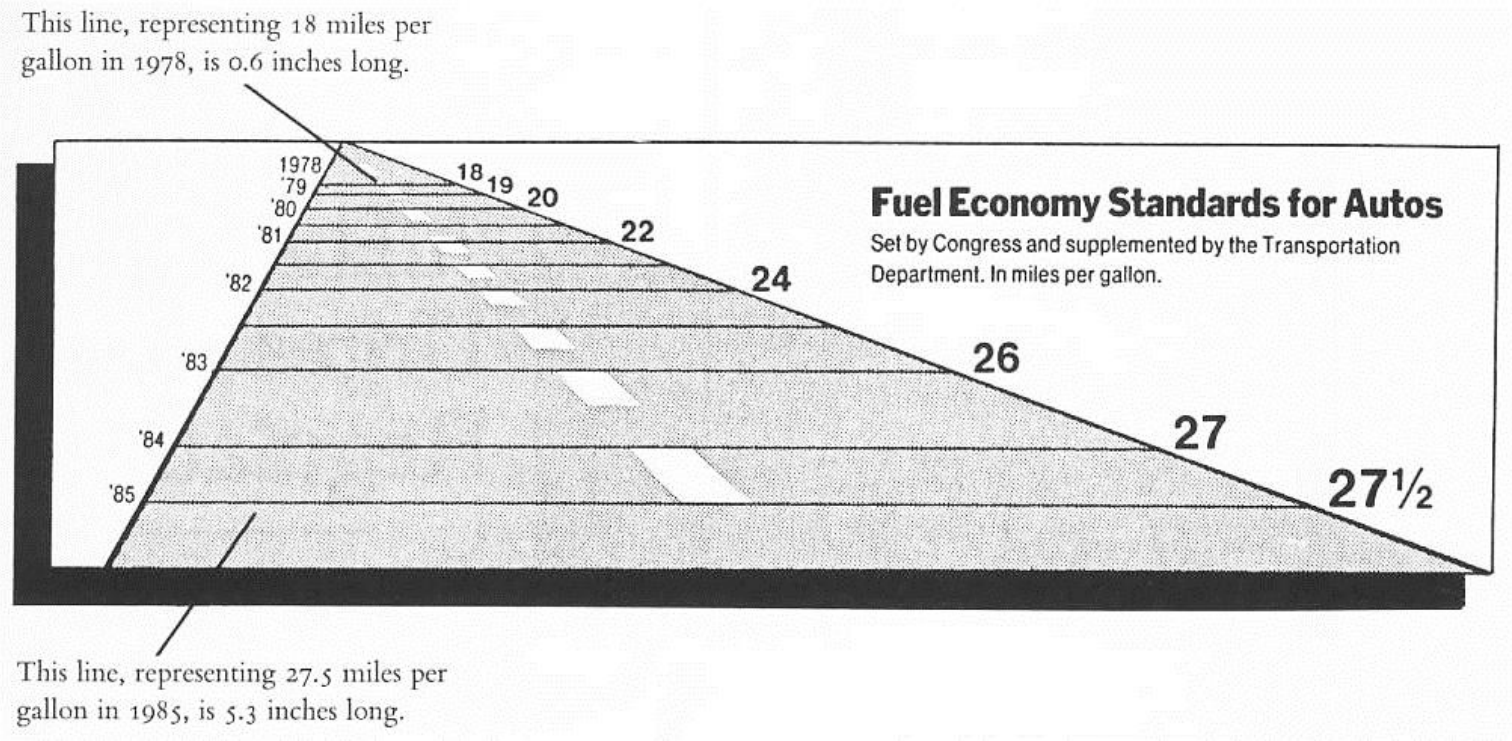
Principles for Graphical Excellence

- From Tufte's:
- the well-designed presentation of interesting data – a matter of substance, of statistics, and of 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
- requires telling the truth about the data

Graphic that fails to tell the truth

- A graphic does not distort if the visual representation of the data is consistent with the numerical representation.
- What is visual representation?
  - Physically measured?
  - Perceived visual effect?
- Perception changes with experience and is context-dependent

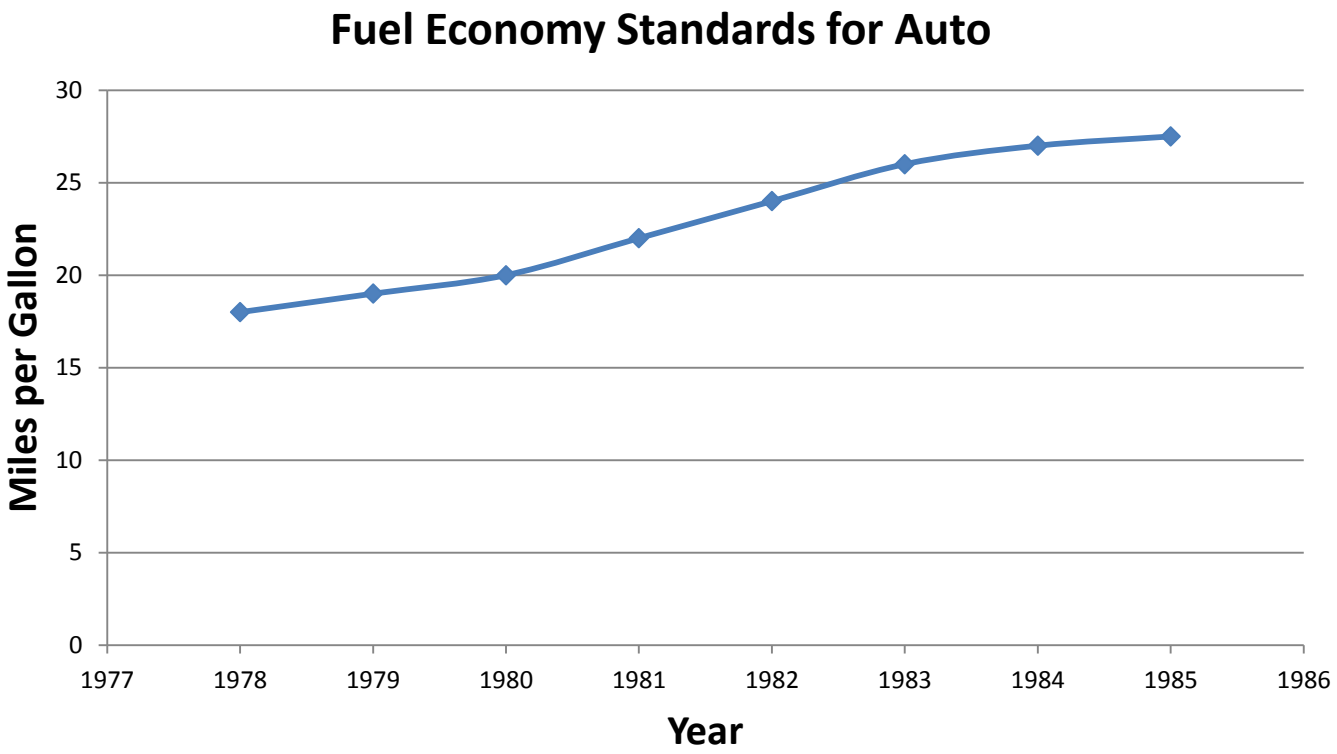
Graphic Distortion



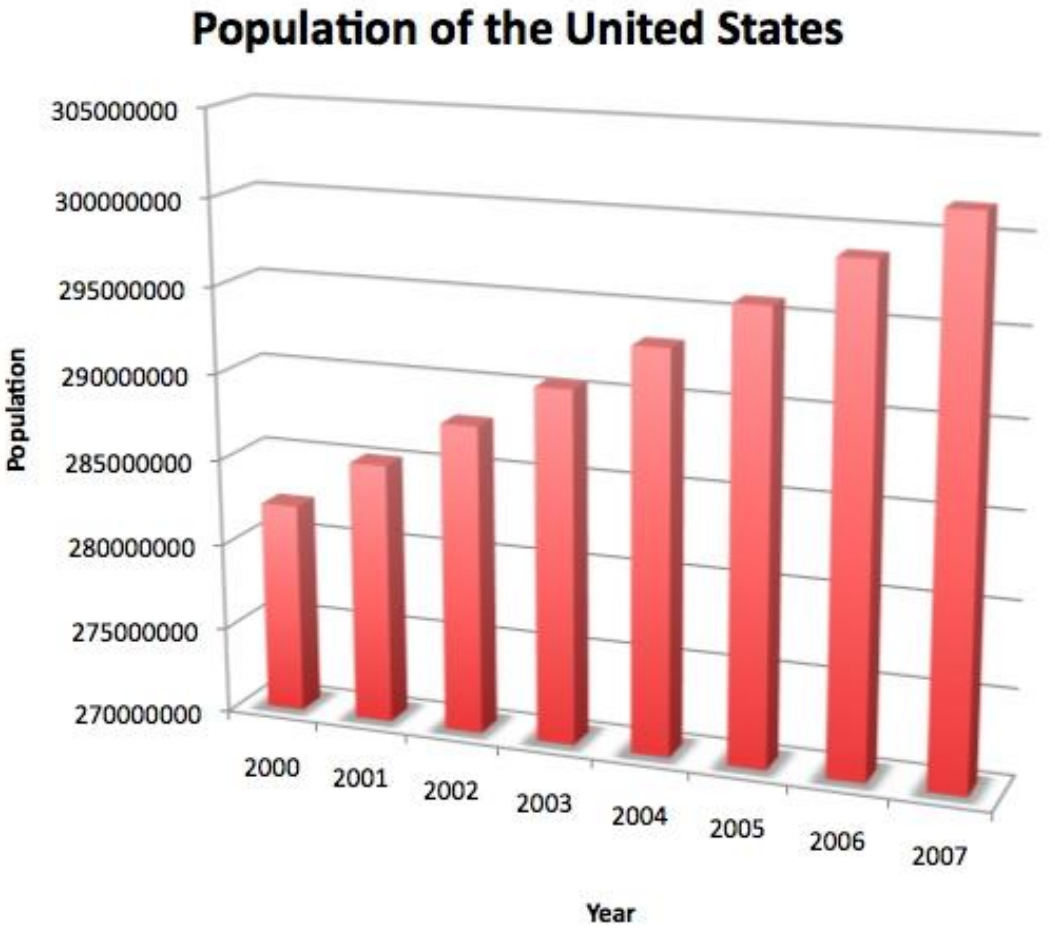
- % increase from 1978 to 1985 = (27.5-18)/18*100% = 53%
- % increase of change in graph = (5.3-0.6)/0.6 = 783%
- **Lie factor** = size of effect in graphic / size of effect in data = 783 / 53 = 14.8

Graphic Distortion

- Can be done as simple as this:



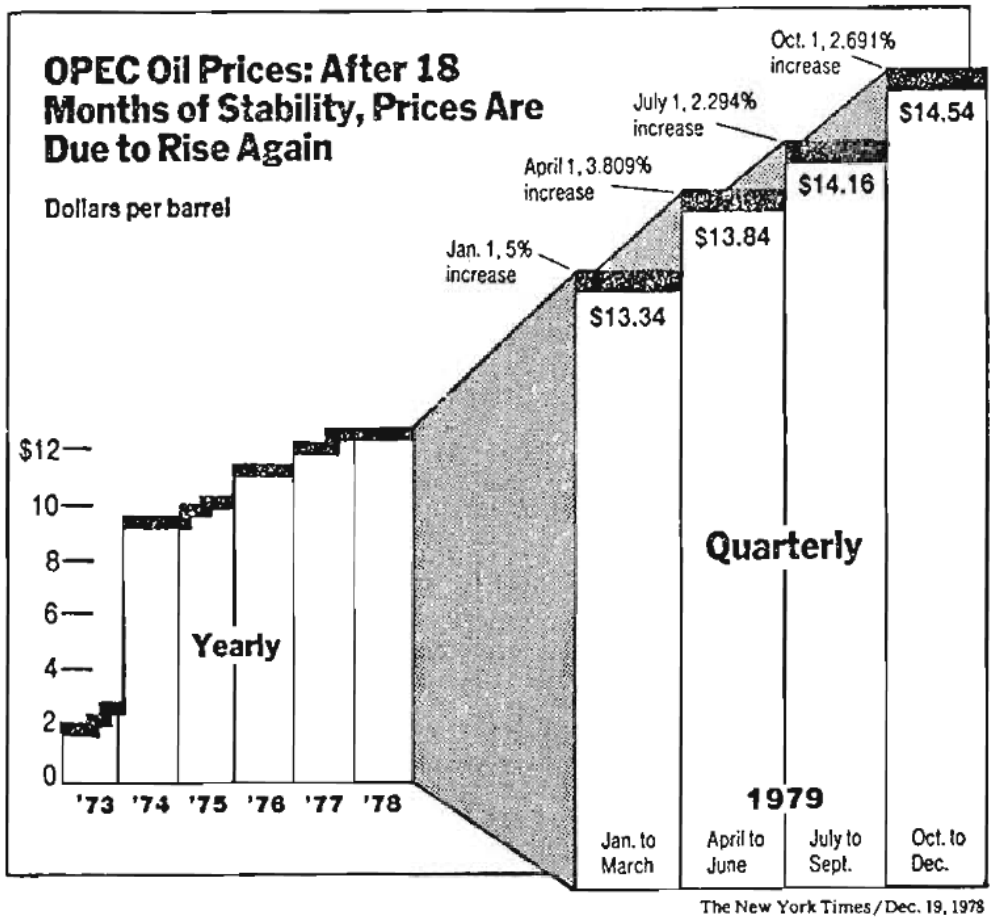
Graphic Distortion



Design vs. Data Variations

- A design is expected to be consistent over the entire graphic. E.g., the intervals depicted by an axis in a graph should be of uniform scale
- Show data variation, not design variation.

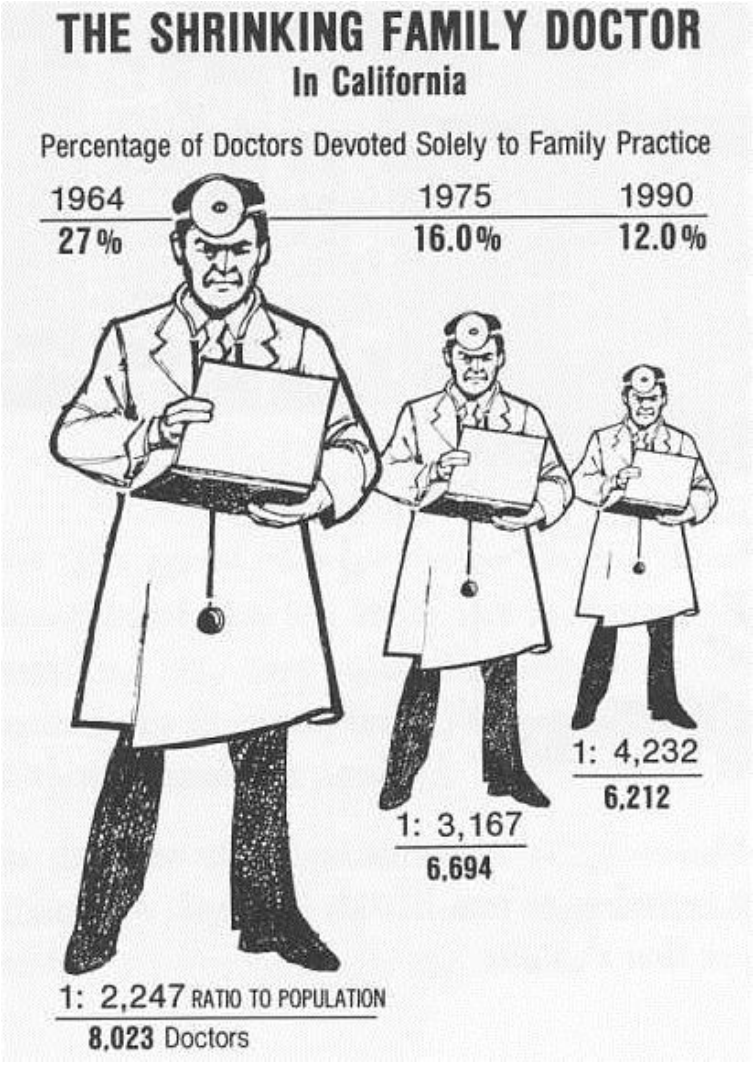
Design vs. Data Variations



- What's the problem here?
- Different horizontal scale
- Different vertical scale

Visual Area vs. Numerical Measure

- Using areas to show one-dimensional data variation is confusing.



Principles of Graphic Integrity

- The representation of numbers, as 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 defeat graphical distortion and ambiguity.
- Graphics must not quote data out of context.

Visualization in Relation to Computer Graphics

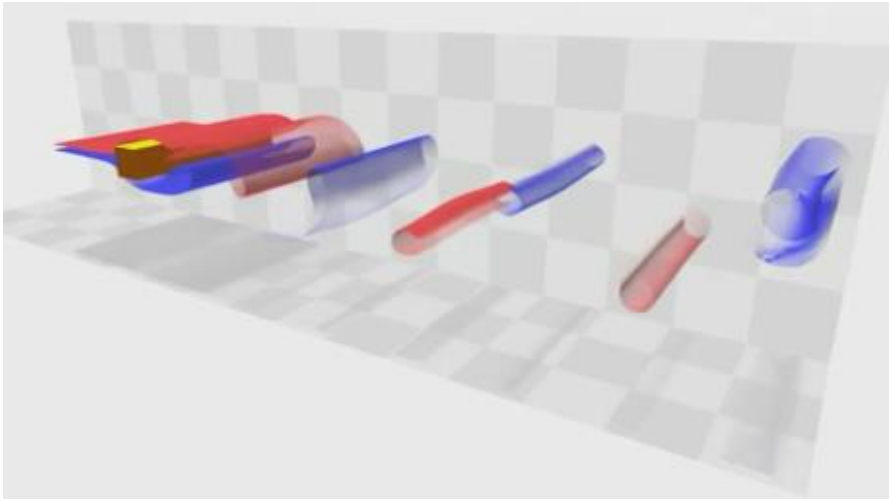
- Computer graphics
  - Digital synthesis and manipulation of visual contents (geometry/imaging/rendering/animation)
  - Visual realism is one of the primary goals
  - Big impact in animation/movies/video games

Visualization in Relation to Computer Graphics

- Visualization
  - Applies computer graphics techniques to generate visual display of data
  - Emphasizes on effective communication of information

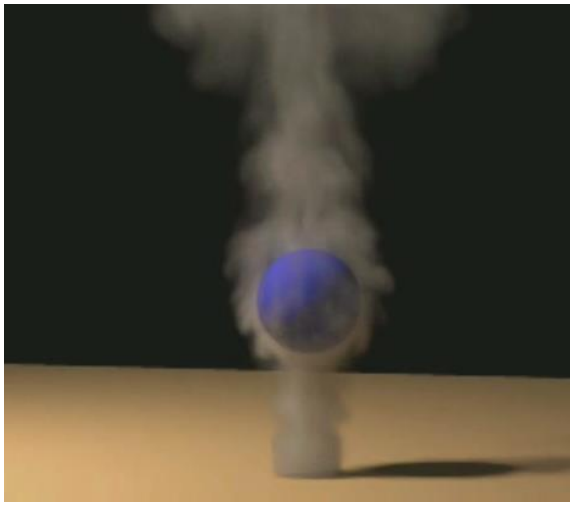
Visualization vs. Computer Graphics

Smoke visualization



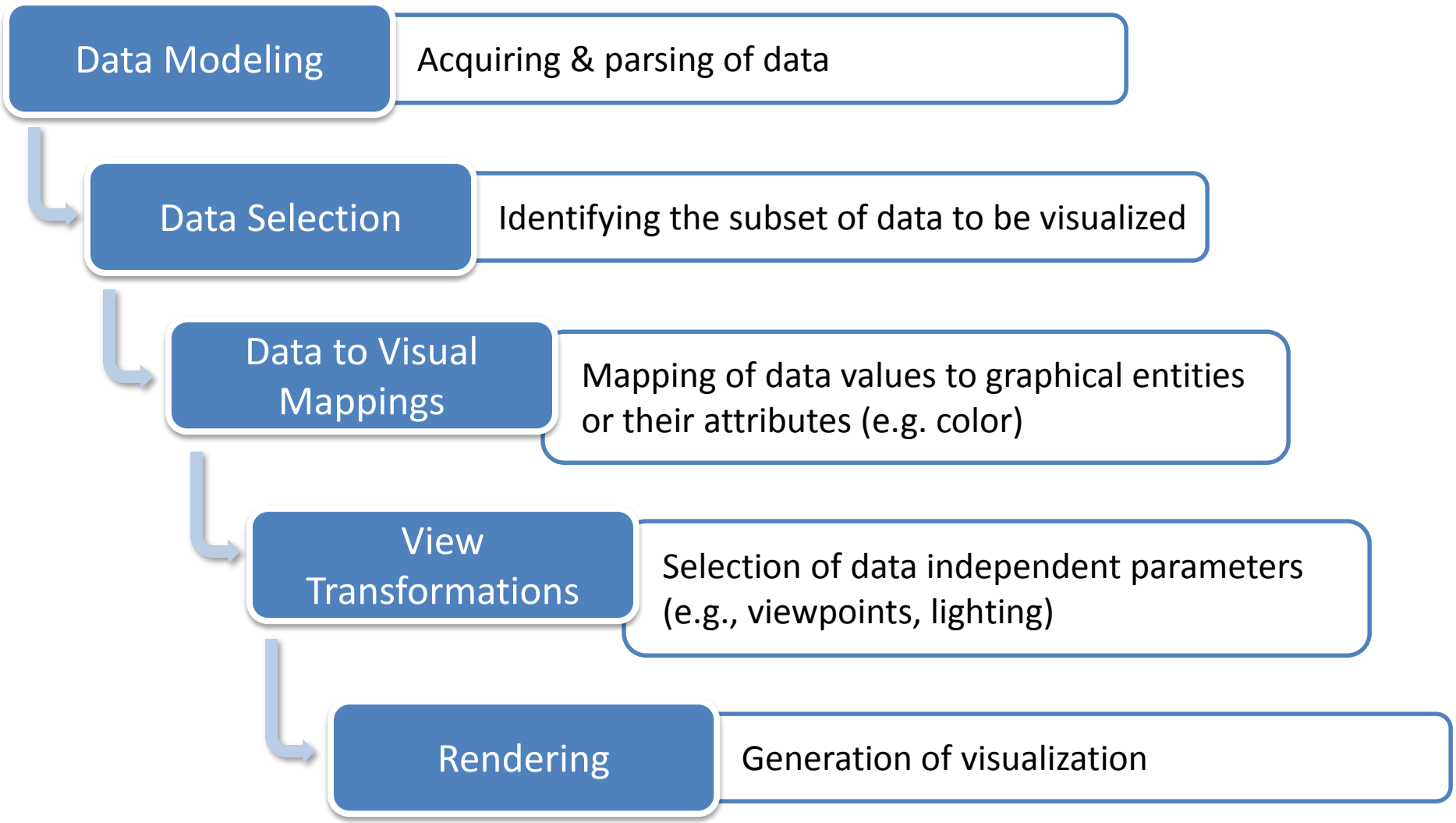
[von Funck et al., VIS 2008]

Smoke rendering



[Fedkiw et al., SIGGRAPH 2001]

The Visualization Pipeline



Role of User

- User involvement in the visualization pipeline depends on the purpose that a visualization serves.
- Visualization for **Presentation**
  - Definite idea or knowledge to convey. Creator of the visualization is fully aware of what is to be shown.

Role of User

- Visualization for **Confirmation**
  - The user has certain hypothesis (e.g. feature in the data) to be verified by the visualization.
- Visualization for **Exploration**
  - The user wants to examine a data set and finds out if any feature is present or missing there.

Role of User

- Why do we need to know the role of the user?
  - Affects the way we design the visualization for a particular purpose
  - Each category of visualization has its own special tools

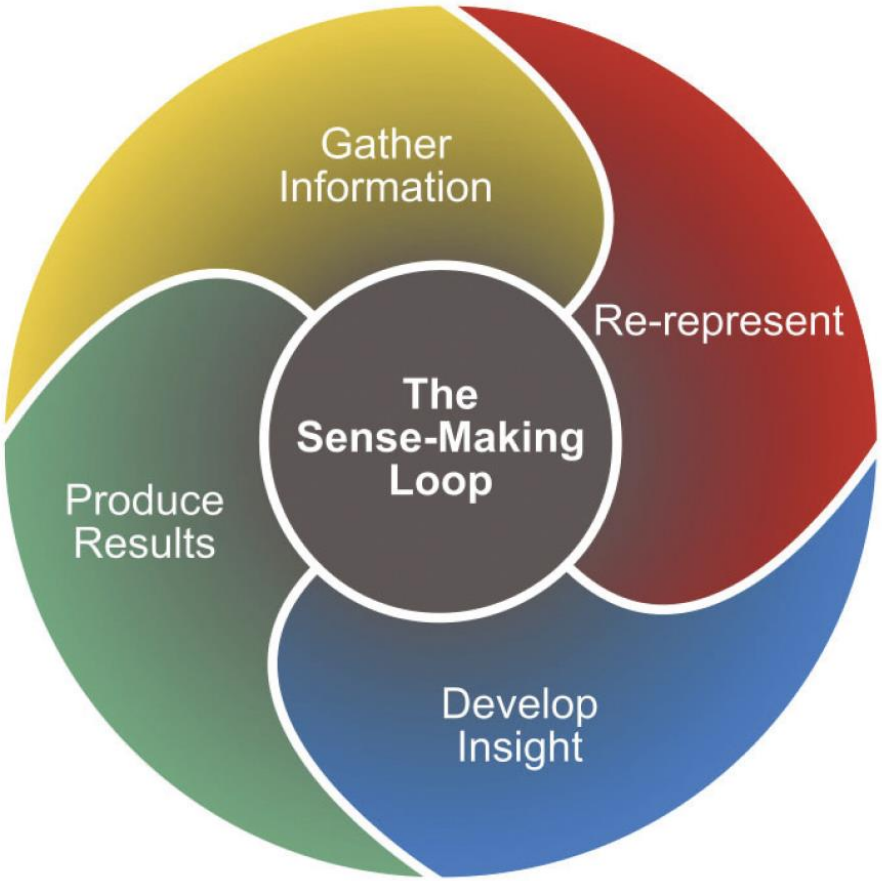
Visual Analytics

- the science of **analytical reasoning** facilitated by interactive visual interfaces
- combines automated analysis techniques with interactive visualizations for an effective understanding, reasoning and decision making on the basis of very large and complex datasets.
- Multi-disciplinary: human-computer interaction, perceptual psychology, databases, statistics, data mining, etc.

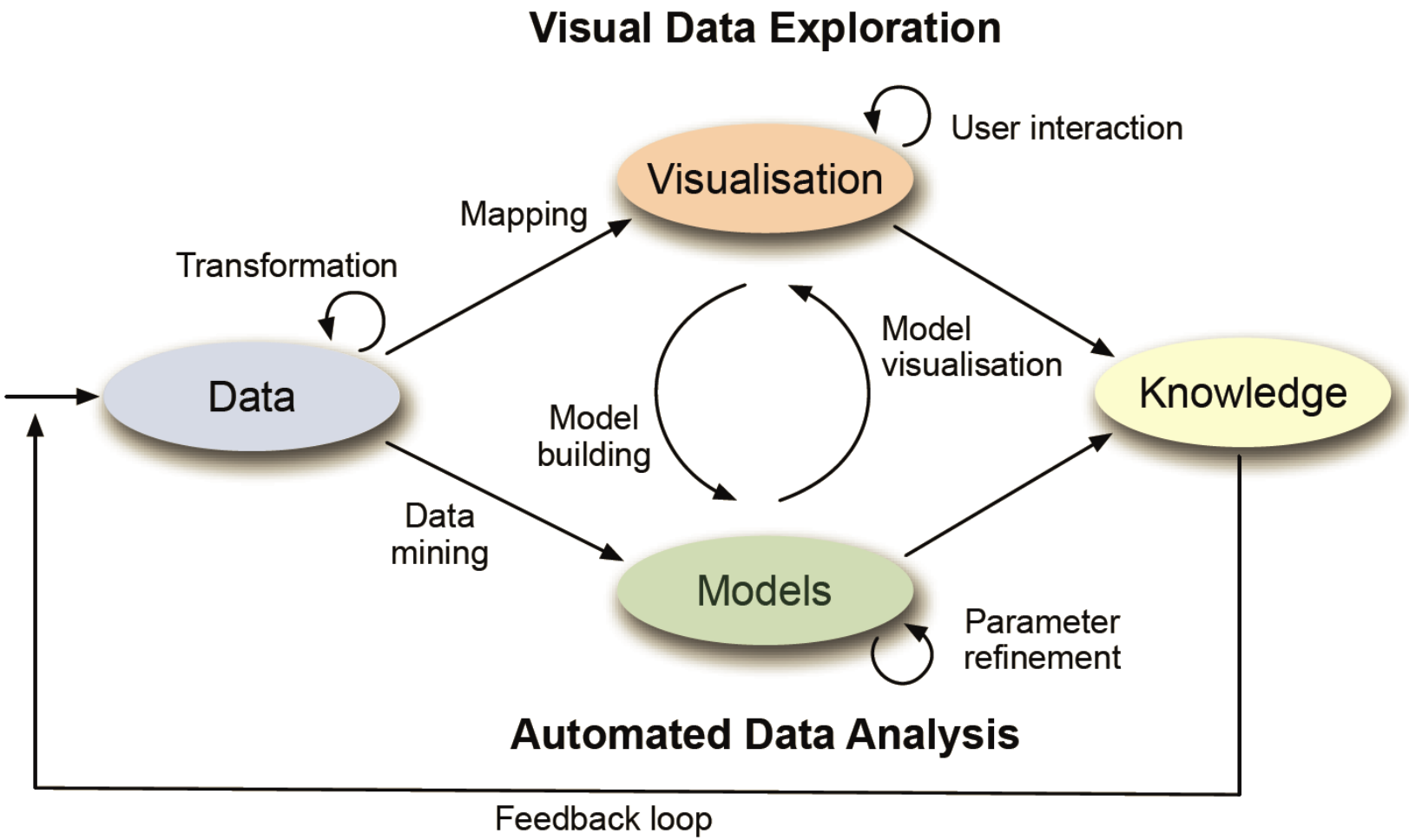
Goal of Visual Analytics

- Visual analytics is the creation of tools and techniques to enable people to
  - Use visual analytics tools and techniques to synthesize information and derive insights from massive, dynamic, ambiguous, and often conflicting data.
  - Detect the expected and discover the unexpected.
  - Provide timely, defensible and understandable assessment.
  - Communicate these assessment effectively for action.

Analytical Reasoning



The Visual Analytics Pipeline



What is Needed for a Visual Analytic Application?

- Effective information presentation
- Exploratory capability
- Intuitive user interface

Working Force in Visual Analytic

- A growing field of research in the US and Europe, with different foci (US: homeland security / Europe: visual business analytics)
- US: National Visualization and Analytics Center
  - *Thomas, J., Cook, K.: Illuminating the Path: Research and Development Agenda for Visual Analytics. IEEE-Press (2005)*
- Europe: The VisMaster Project
  - http://www.visual-analytics.eu

State-of-the-Art Research

- Visualization
  - IEEE VIS (InfoVis / SciVis) (ieeevis.org)
- Visual Analytics
  - IEEE VIS (VAST) (ieeevis.org)
  - Visual Analytics Community (http://vacommunity.org)
- Computer Graphics
  - SIGGRAPH

References

- Matthew Ward, Georges Grinstein and Daniel Keim, "*Interactive Data Visualization: Foundations, Techniques, and Applications*", 2010 [Chapter 1]
- Edware R. Tufte, "*The Visual Display of Quantitative Information*", 2001.
- D. Keim, J. Kohlhammer, G. Ellis and F. Mansmann, "*Mastering the Information Age: Solving Problems with Visual Analytics", 2010.*
- J. Thomas and K. Cook, "*Illuminating the Path*", 2006.

Multiples of bytes

Decimal		
Value	Metric	
1000	kB	kilobyte
1000^2	MB	megabyte
1000^3	GB	gigabyte
1000^4	TB	terabyte
1000^5	PB	petabyte
1000^6	EB	exabyte
1000^7	ZB	**zettabyte**
1000^8	YB	yottabyte