

G53FIV: Fundamentals of Information Visualization

Lecture 2: The Value of Visualization

Ke Zhou
School of Computer Science
Ke.Zhou@nottingham.ac.uk

Overview

- What are the Key Values of IV?
 - Record
 - Communicate
 - Reason
- What is the Process of IV?

What are the Key Values of IV?

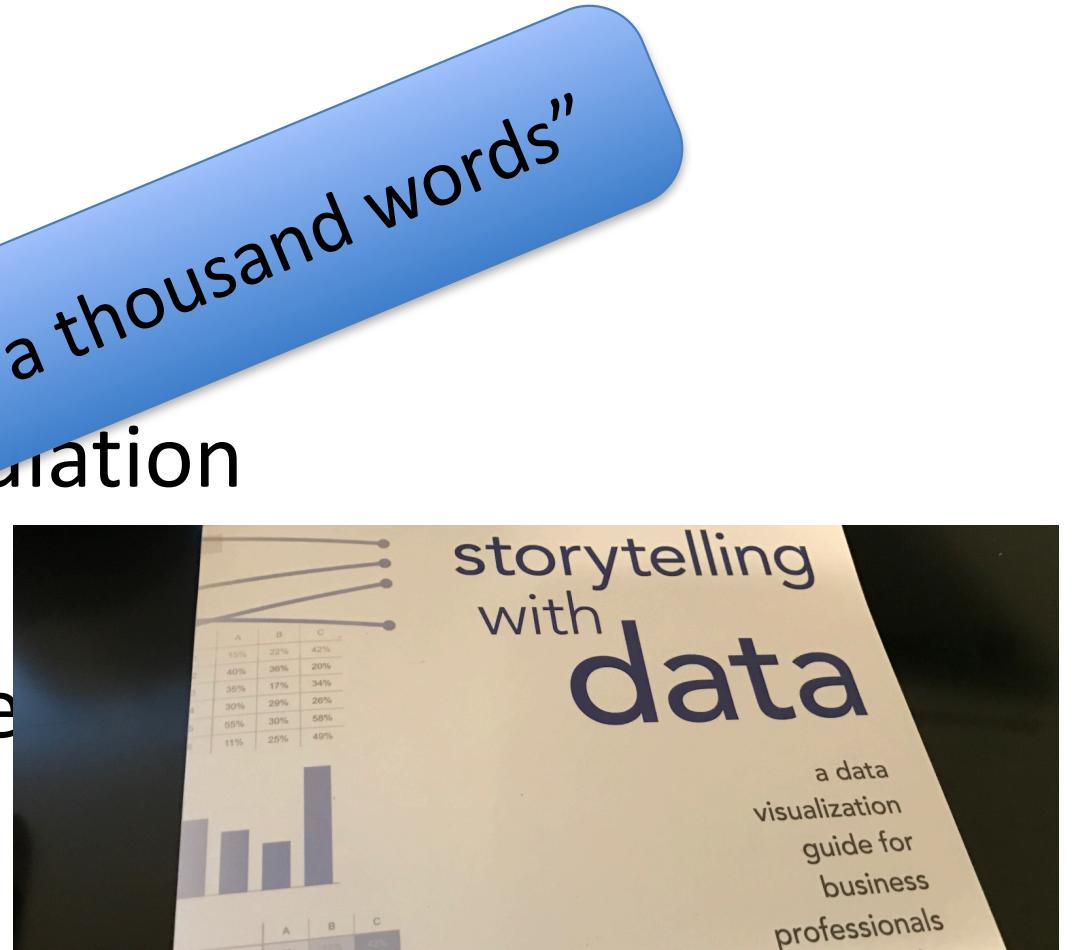
Why Create Visualization?

Why Create Visualization?

- Answer questions (or discover them)
- Make decisions
- See data in context
- Expand memory
- Support graphical calculation
- Find patterns
- Present argument or tell a story
- Inspire

Why Create Visualization?

- Answer questions (or discover them)
- Make decisions
- See data in context
- Expand memory
- Support argument or persuasion
- Find patterns “A picture is worth a thousand words”
- Present argument or tell story
- Inspire



Summary of Reasons

- **Record** information
 - Blueprints, photographs, seismographs, ...
- **Communicate** information to others
 - Share and persuade
 - Collaborate and revise
- Analyze data to **support reasoning**
 - Find patterns / Discover errors in data
 - Expand memory
 - Develop and assess hypotheses

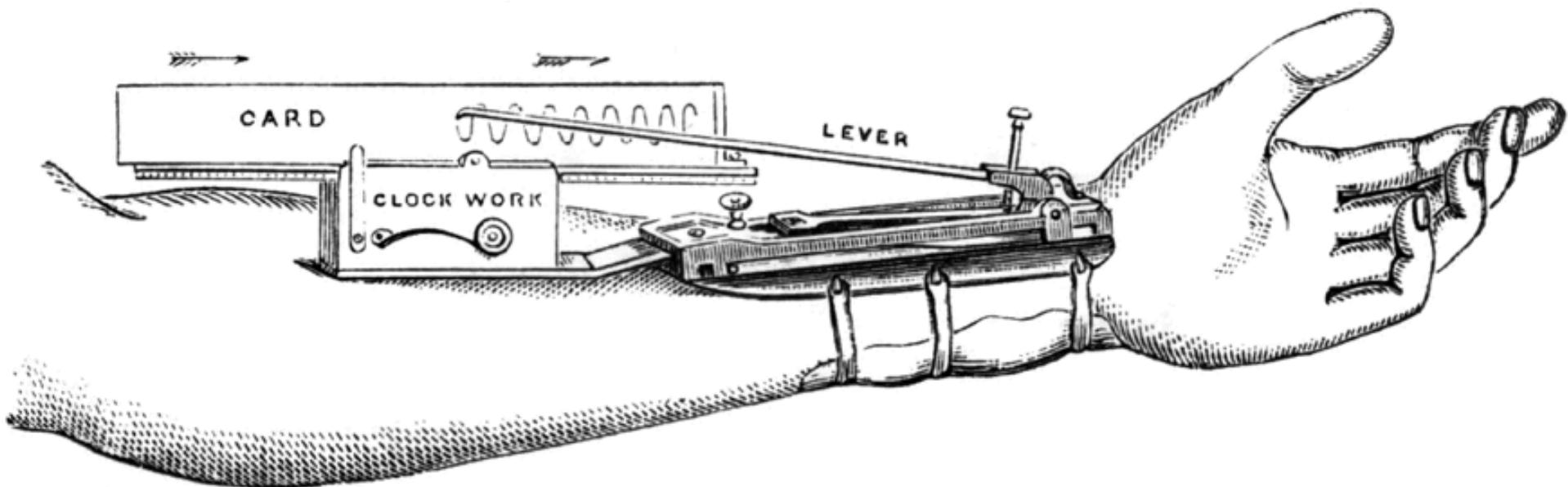
Record Information

- Egyptian hieroglyphs



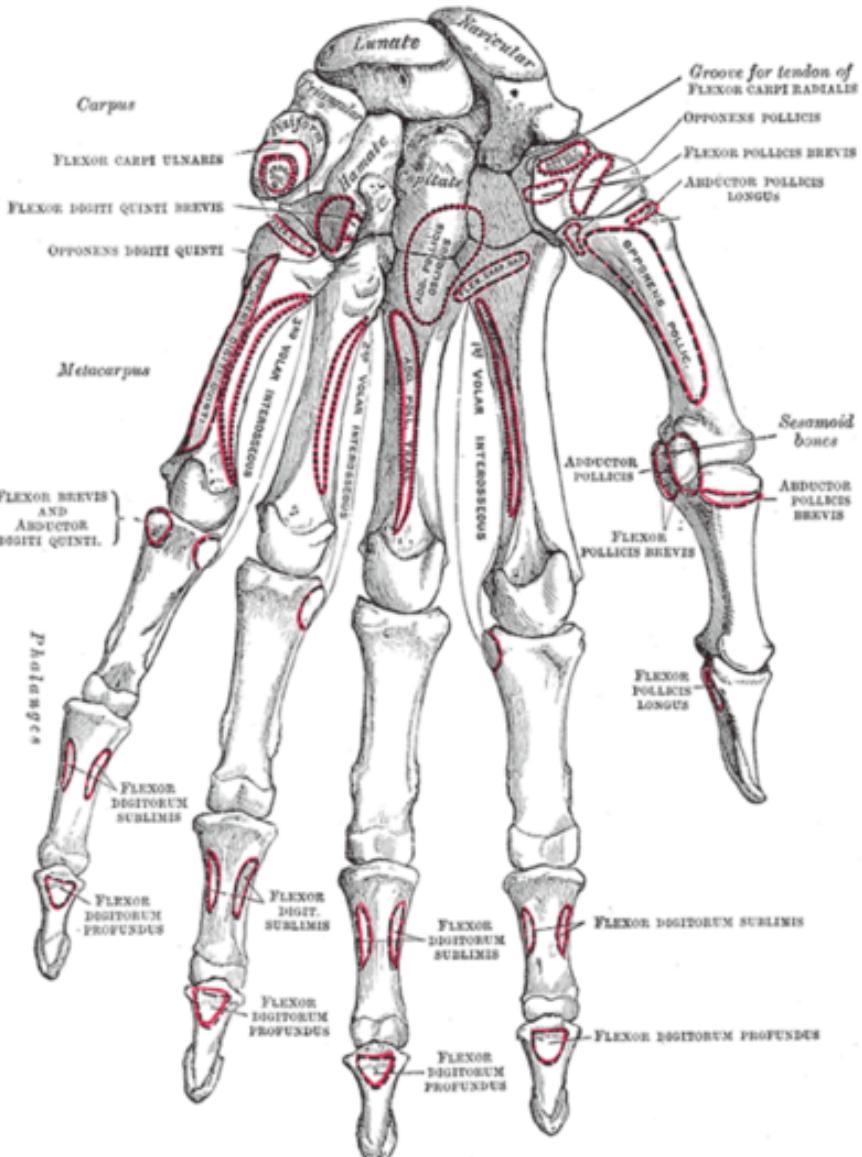
Record Information

- E.J. Marey's sphygmograph (1854)
 - an instrument which produces a line recording the strength and rate of a person's pulse.

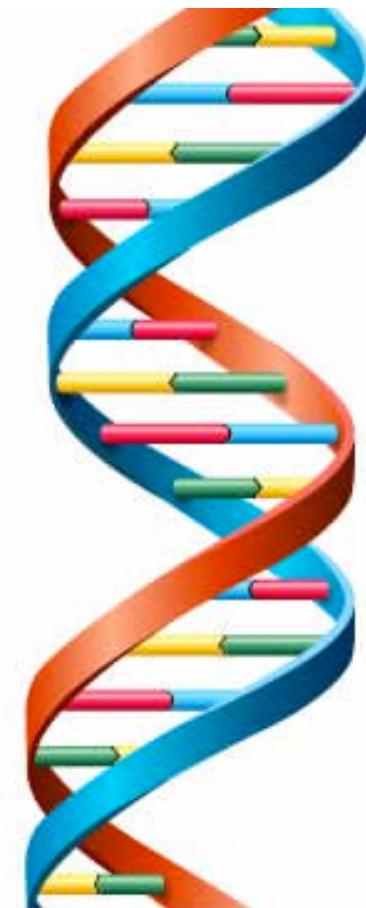


Communicate: convey information to others

Share and collaborate



Bones in hand (drawn in 1918)



DNA Helix

Persuade: Nightingale's Graph

2.
APRIL 1855 to MARCH 1856.

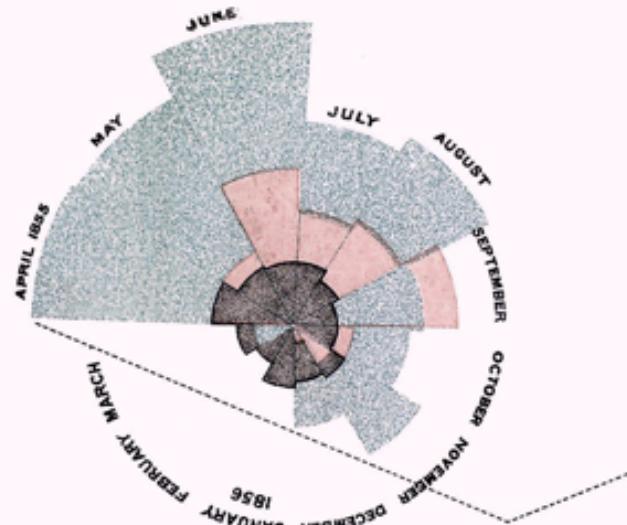
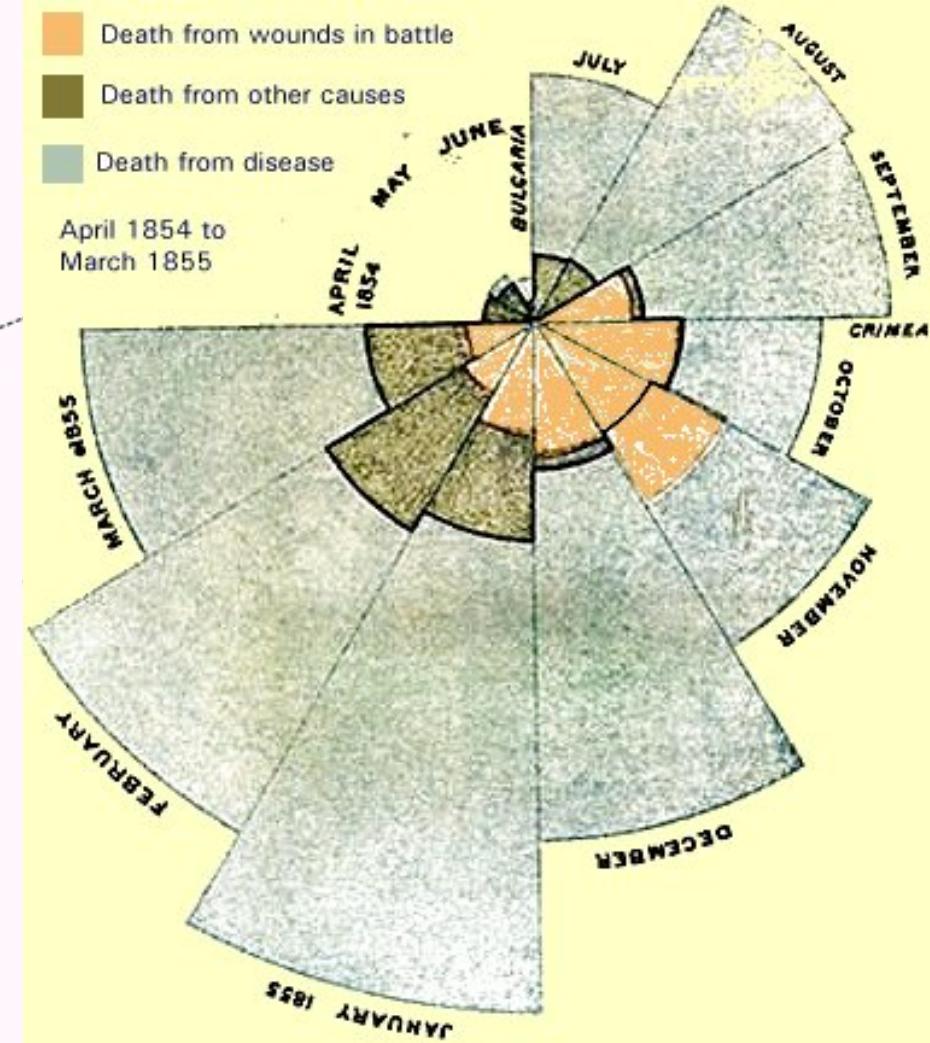


DIAGRAM OF THE CAUSES OF MORTALITY
IN THE ARMY IN THE EAST.

1.
APRIL 1854 to MARCH 1855.



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

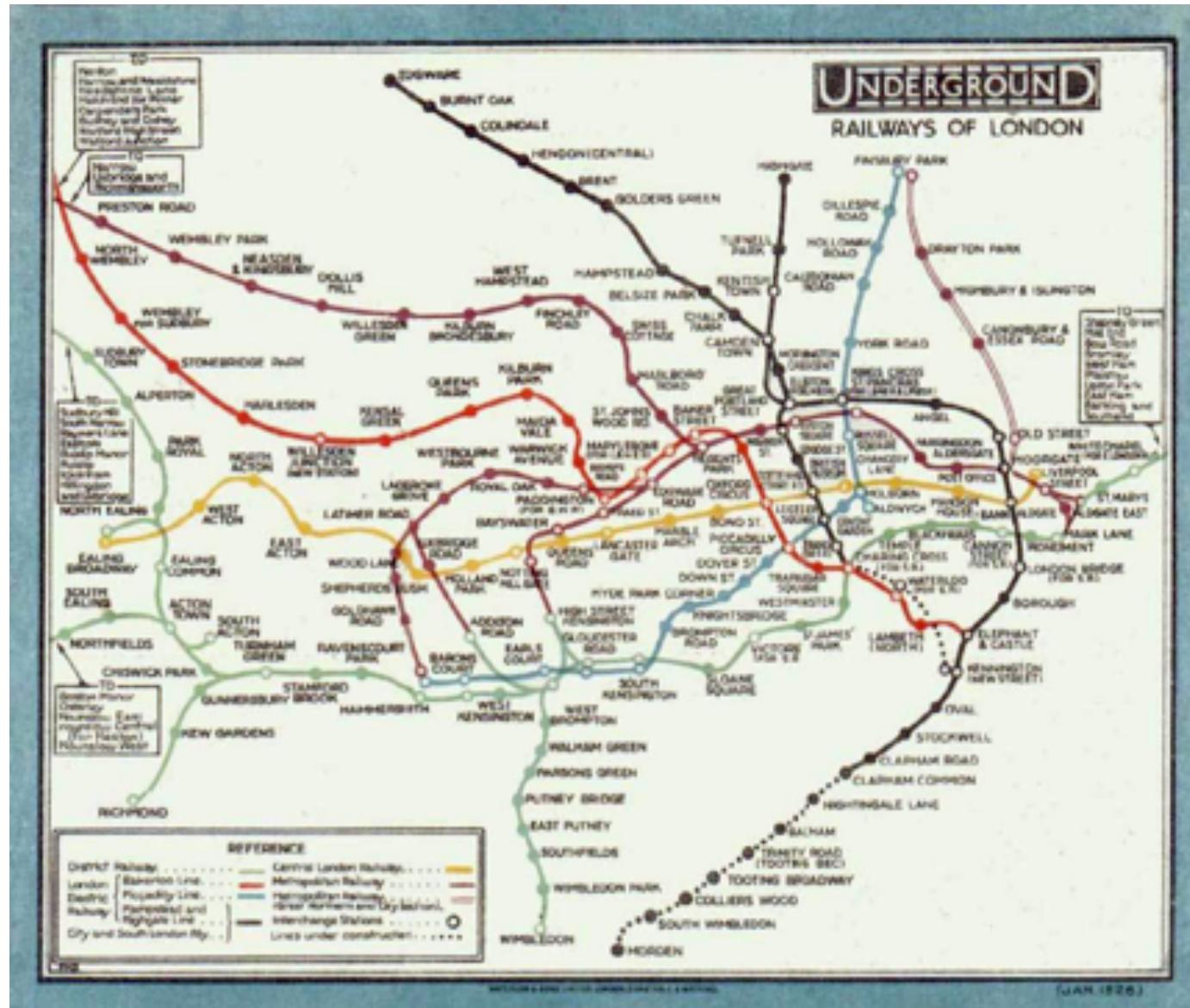
The blue wedges measured from the centre of the circle represent area for area the deaths from Preventible or Mitigable Zymotic diseases; the red wedges measured from the centre the deaths from wounds; & the black wedges measured from the centre the deaths from all other causes.

The black line across the red triangle in Nov. 1854 marks the boundary of the deaths from all other causes during the month.

In October 1854, & 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.

Clarify/Revise: London's underground map

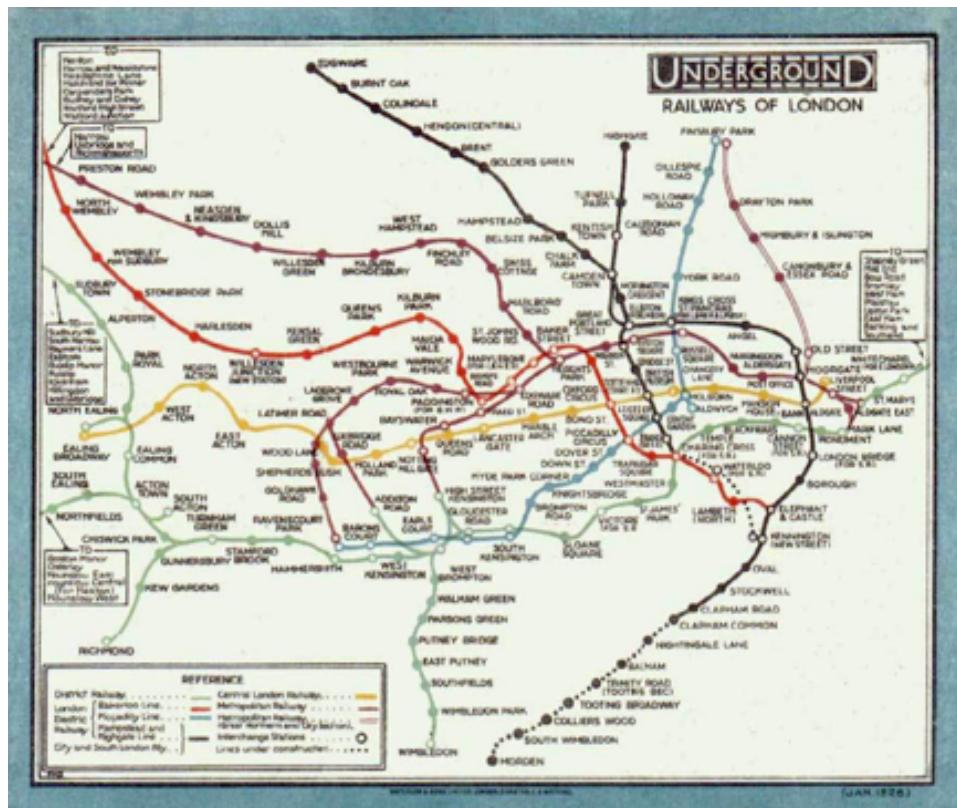


1926

Dr. Ke Zhou (<http://www.cs.nott.ac.uk/~pszkz/>)

Clarify/Revise: London's underground map

- Horizontal, vertical and 45° segments
 - Key insight: topology and relative location of stations



1926

vs.



1987

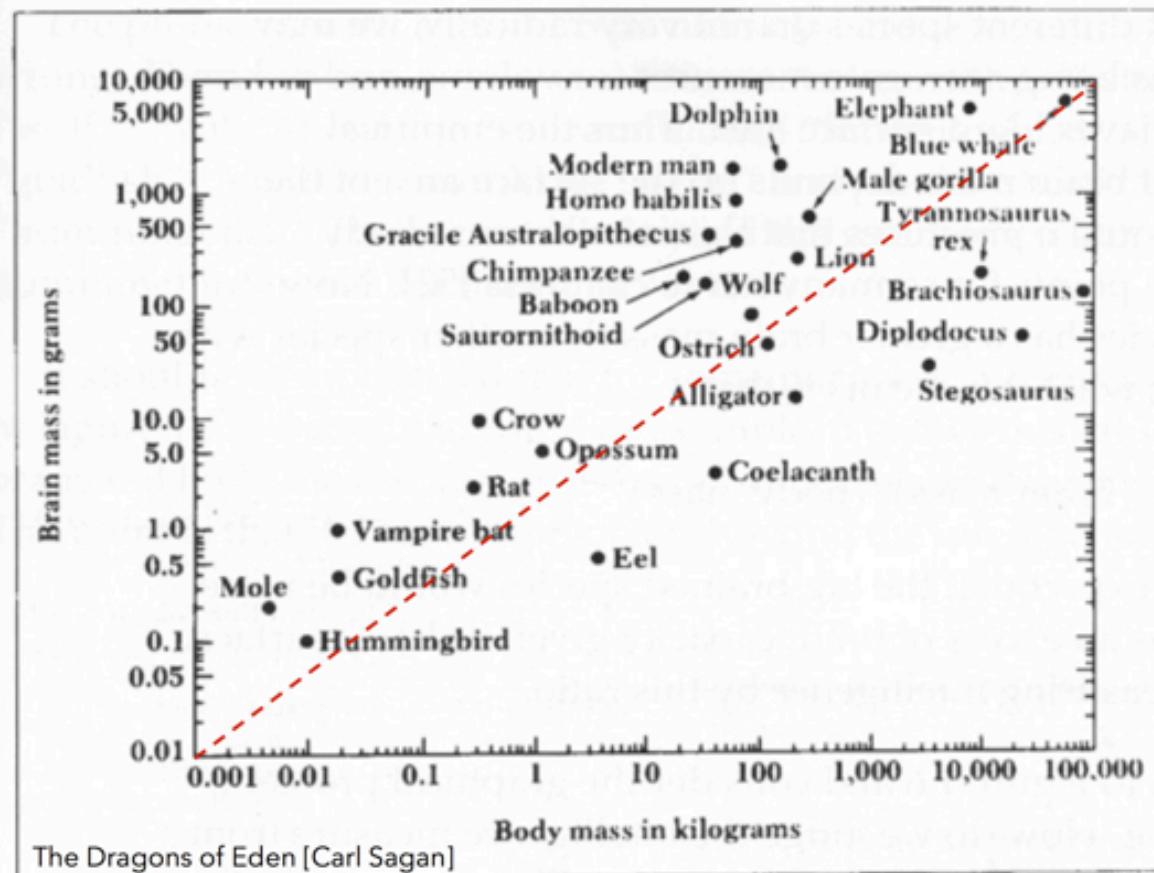
Beer Infographics



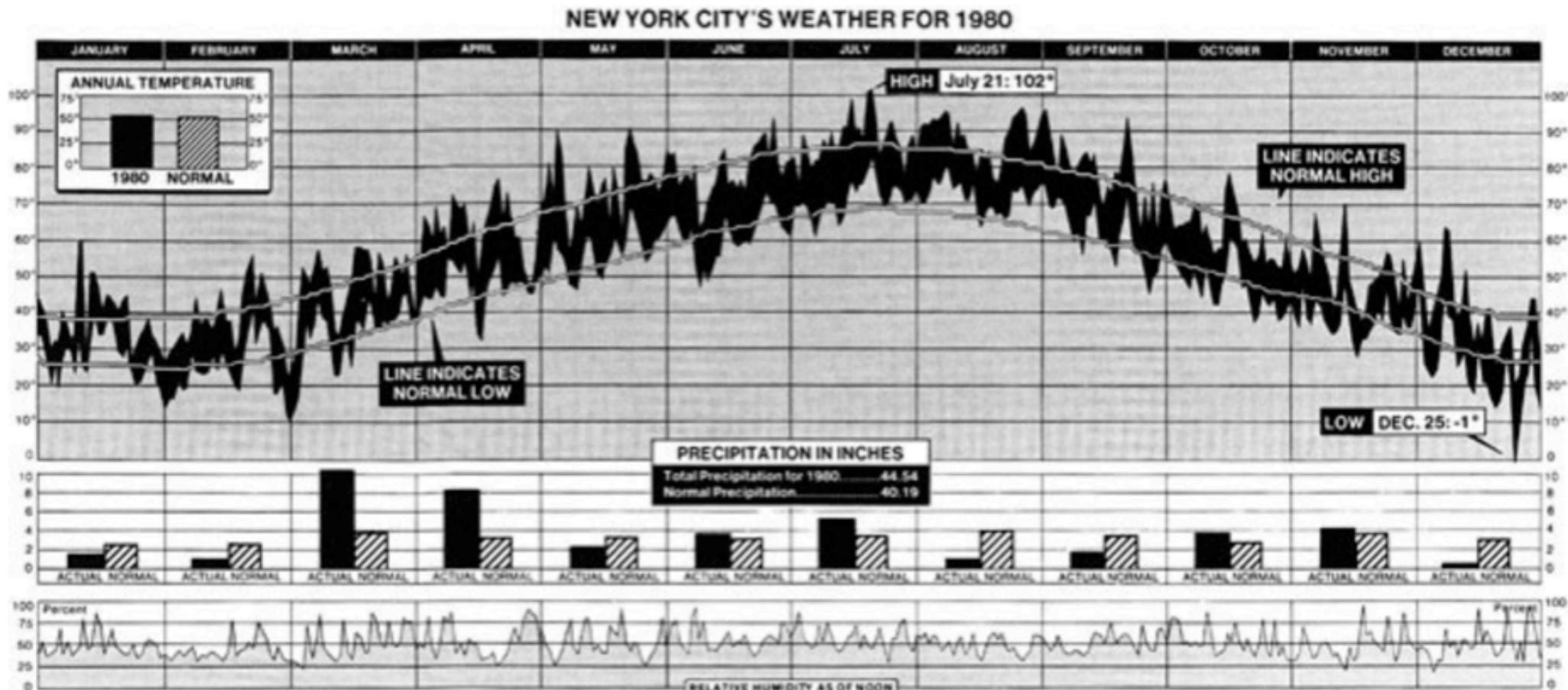
Support reasoning

Find Patterns: the most powerful brain

ID	Name	Body Weight	Brain Weight
1	Lesser Short-tailed Shrew	5	0.14
2	Little Brown Bat	10	0.25
3	Mouse	23	0.3
4	Big Brown Bat	23	0.4
5	Musk Shrew	48	0.33
6	Star Nosed Mole	60	1
7	Eastern American Mole	75	1.2
8	Ground Squirrel	101	4
9	Tree Shrew	104	2.5
10	Golden Hamster	120	1
11	Mole Rat	122	3
12	Galago	200	5
13	Rat	280	1.9
14	Chinchilla	425	6.4
15	Desert Hedgehog	550	2.4
16	Rock Hyrax (a)	750	12.3
17	European Hedgehog	785	3.5
18	Tenrec	900	2.6
19	Arctic Ground Squirrel	920	5.7
20	African Giant Pouched Rat	1000	6.6
21	Guinea Pig	1040	5.5
22	Mountain Beaver	1350	8.1
23	Slow Loris	1400	12.5
24	Genet	1410	17.5
25	Phalanger	1620	11.4



Find Patterns: NYC weather



2200 data points

Expand Memory

- Class Exercise

34

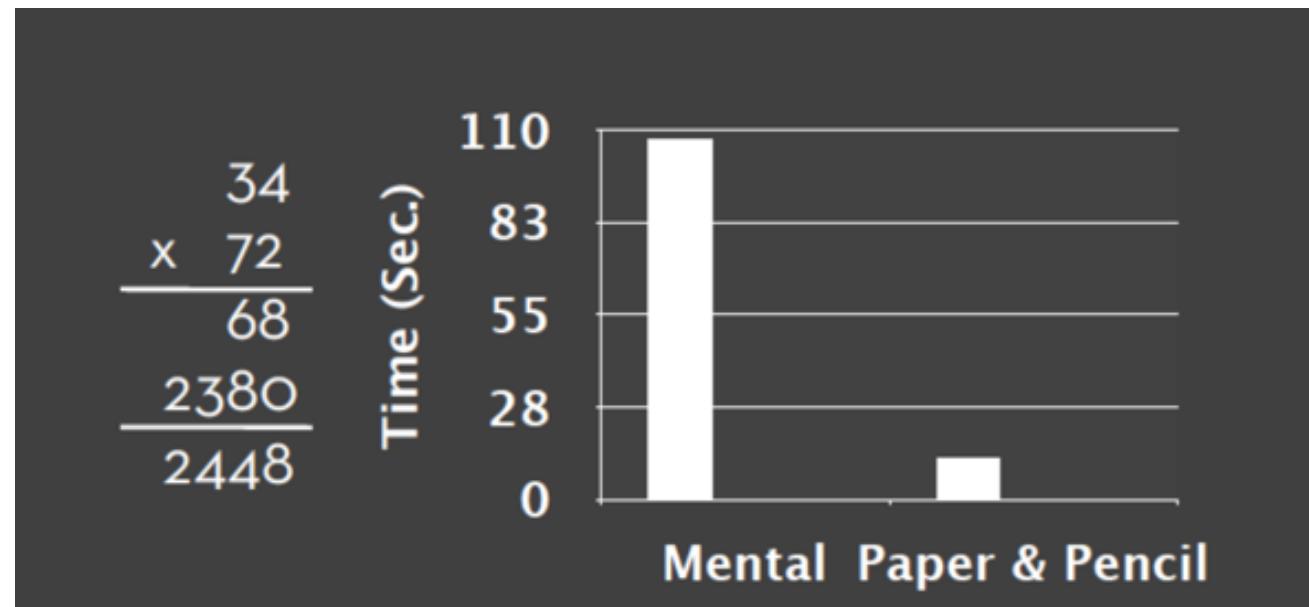
x 72

Expand Memory

- Class Exercise

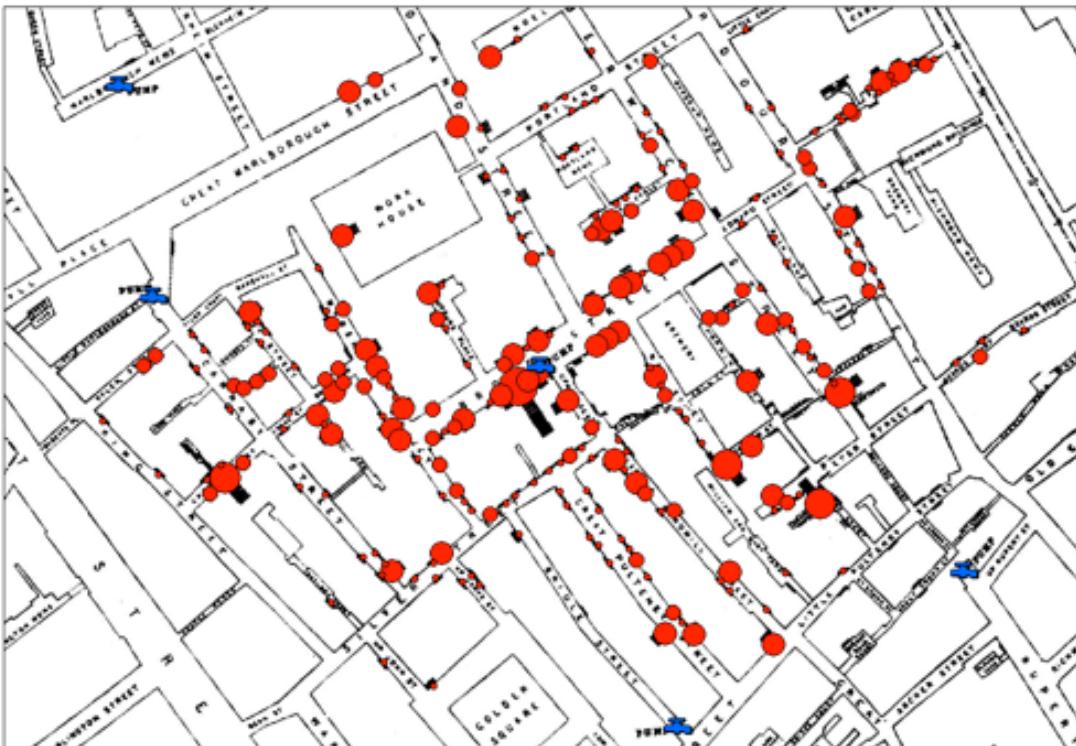
$$\begin{array}{r} 34 \\ \times 72 \\ \hline \end{array}$$

$$\begin{array}{r} 34 \\ \times 72 \\ \hline 68 \\ 2380 \\ \hline 2448 \end{array}$$



Develop and Assess Hypothesis

London Cholera Map



London Cholera Map
Visualization by John Snow, 1854.

- The closer to the Broad Street water pump, the greater the number of deaths.
- The information helped convince the public a true sewage system was needed.

Surprises in Data

“The greatest value of a picture is when it forces us to notice what we never expected to see.”

John Tukey, 1977

“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...”

W. Cleveland
The Elements of Graphing Data

Reasoning / Exploration

“If you can articulate very precisely what you’re seeking, visualization likely isn’t your best approach”

J. Stasko, EuroVis’14

Exploration

- Don’t know what you’re looking for
- Don’t have a priori questions
- Want to know what questions to ask

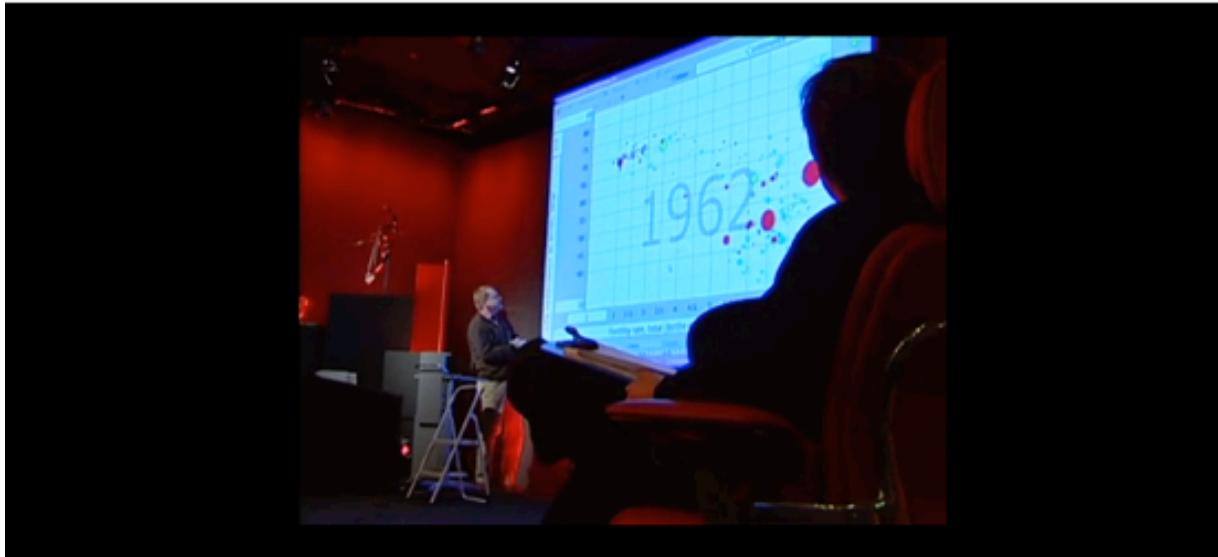


Hans Rosling's TED talk



TED

LOG IN



Hans Rosling:

The best stats you've ever seen

TED2006 · 19:50 · Filmed Feb 2006

 48 subtitle languages •

View interactive transcript



List



Download



Rate



Link

Share

11,146,687

Total views

http://www.ted.com/talks/hans_rosling_shows_the_best_stats_you_ve_ever_seen

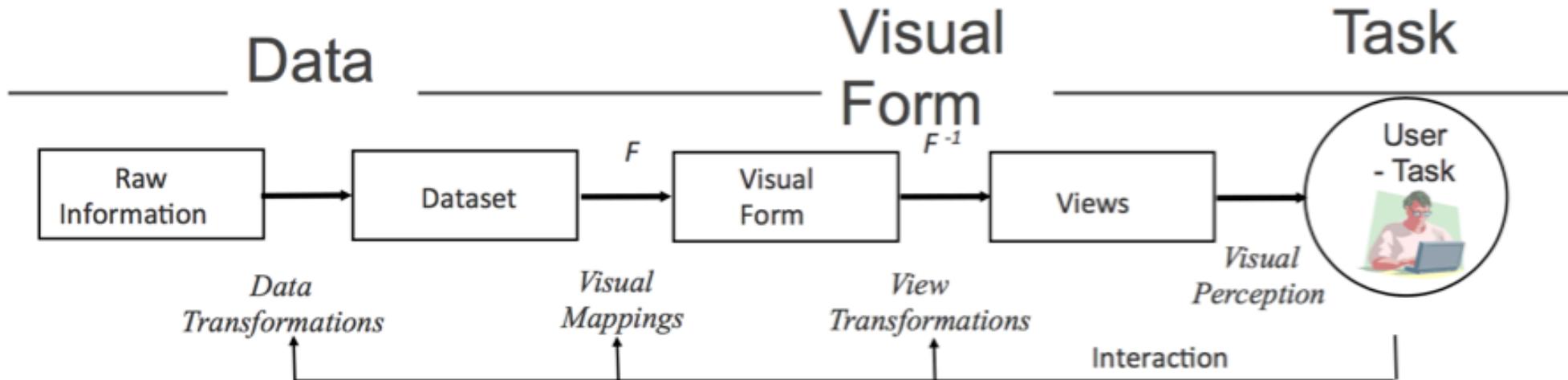
Key Applications of IV

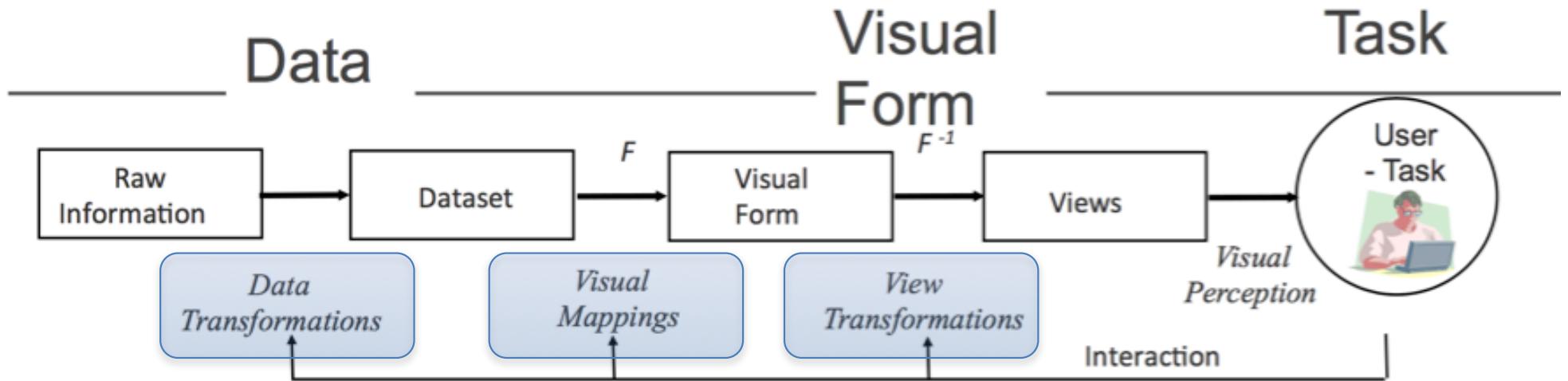
- I. Record Information
- II. Communications (Presentation)
 - Communicate data and ideas
 - Explain and inform
 - Provide evidence and support
 - Influence and persuade
- III. Reasoning (Analysis)
 - Explore the data
 - Assess a situation
 - Determine how to proceed
 - Decide what to do



The Visualization Process

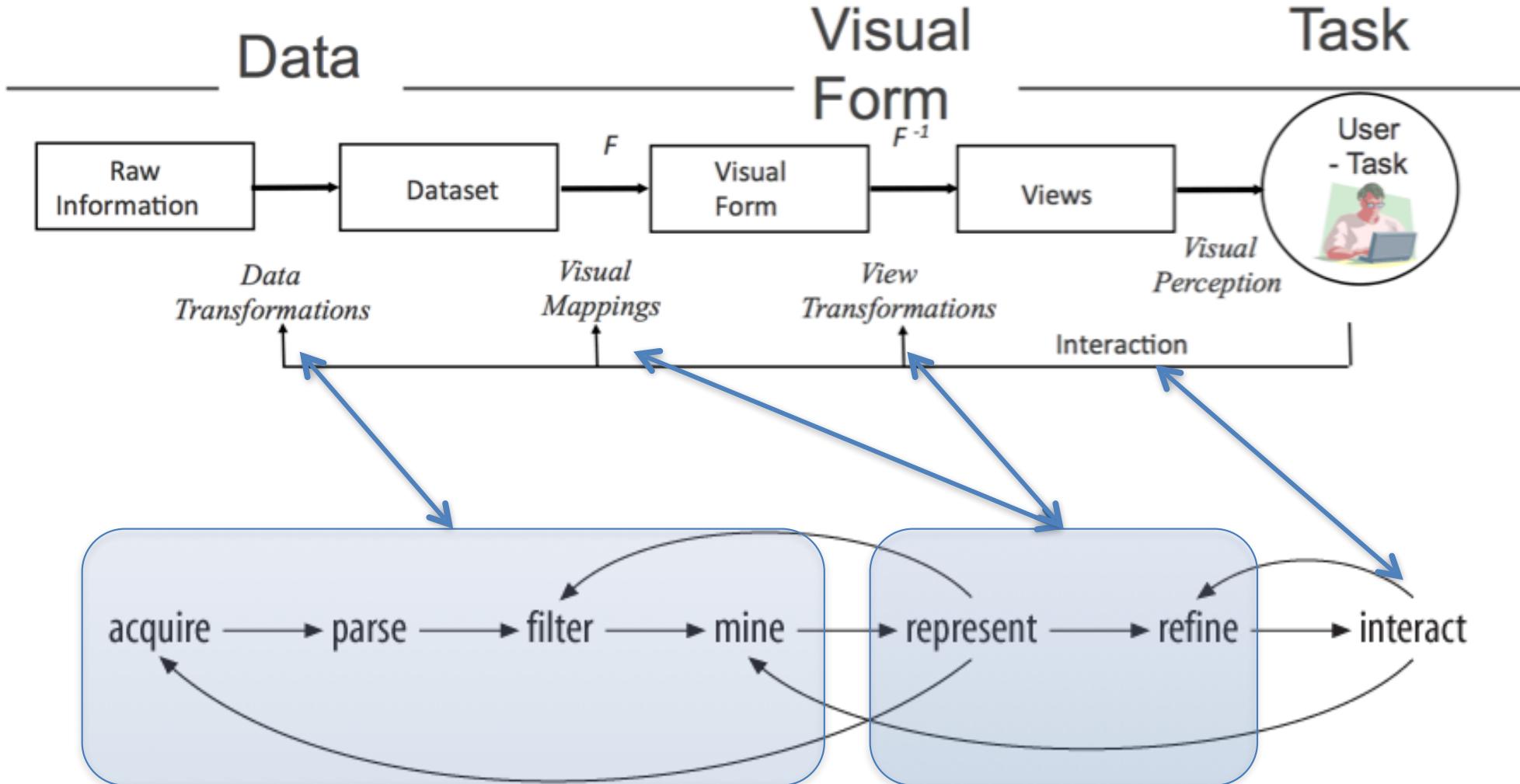
Different Stages of Visualization



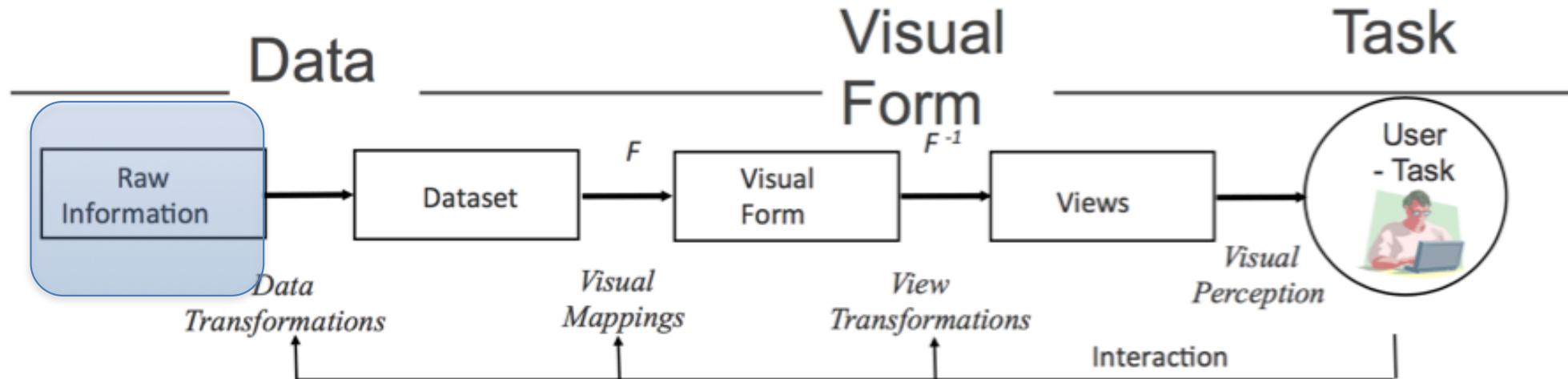


- Data transformation
 - create a structural model (schema), mapping raw data into data tables
- Visual mapping
 - create a visual spatial model, transforming data tables into visual structures
- View Transformations
 - Create views of the Visual Structures by specifying graphical parameters such as position, scaling, and clipping

Different Stages of Visualization



Seven Stages: Acquire

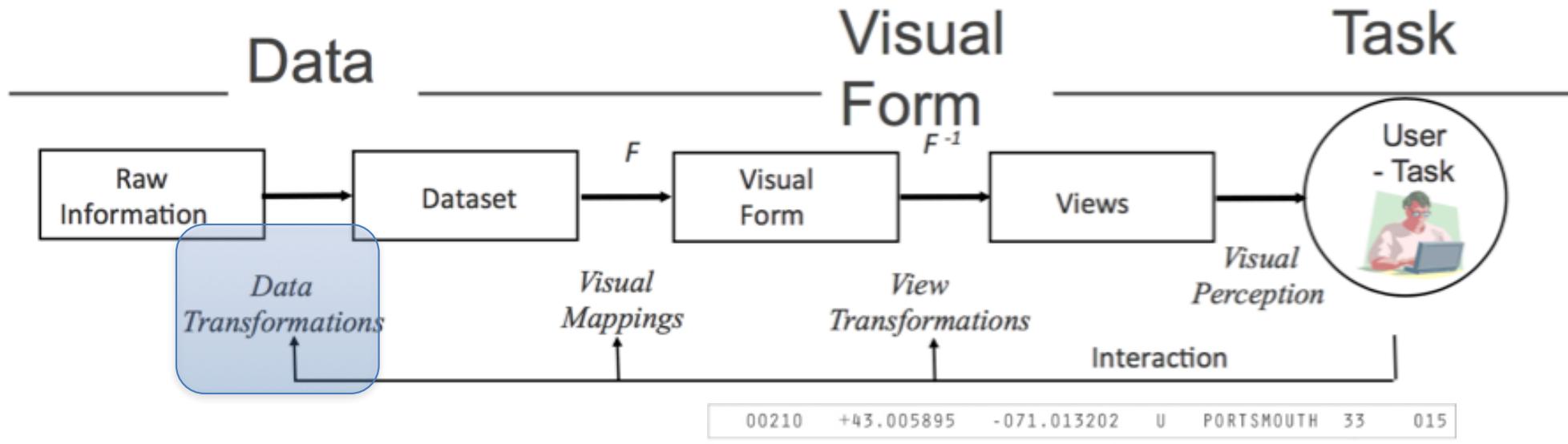


- Obtain the data, whether from a file on a disk or a source over a network

00210	+43.005895	-071.013202	U	PORTSMOUTH	33	015
00211	+43.005895	-071.013202	U	PORTSMOUTH	33	015
00212	+43.005895	-071.013202	U	PORTSMOUTH	33	015
00213	+43.005895	-071.013202	U	PORTSMOUTH	33	015
00214	+43.005895	-071.013202	U	PORTSMOUTH	33	015
00215	+43.005895	-071.013202	U	PORTSMOUTH	33	015
00501	+40.922326	-072.637078	U	HOLTSVILLE	36	103
00544	+40.922326	-072.637078	U	HOLTSVILLE	36	103
00601	+18.165273	-066.722583		ADJUNTAS	72	001
00602	+18.393103	-067.180953		AGUADA	72	003
00603	+18.455913	-067.145780		AGUADILLA	72	005
00604	+18.493520	-067.135883		AGUADILLA	72	005
00605	+18.465162	-067.141486	P	AGUADILLA	72	005
00606	+18.172947	-066.944111		MARICAO	72	093
00610	+18.288685	-067.139696		ANASCO	72	011

Zip codes in the format provided by the U.S. Census Bureau

Seven Stages: Parse

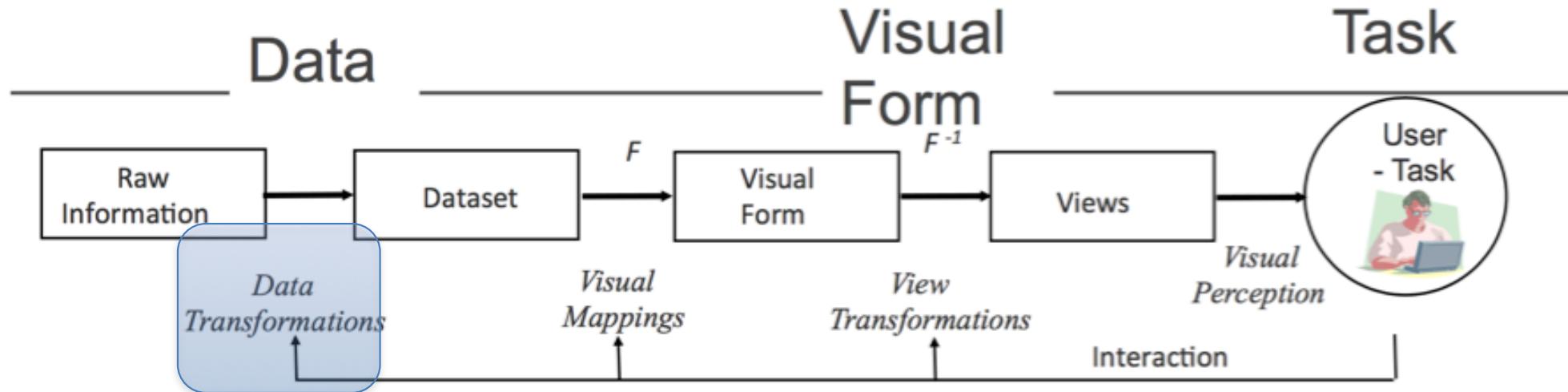


- Provide some structure for the data's meaning, and order it into categories.

01	ALABAMA	AL
02	ALASKA	AK
04	ARIZONA	AZ
05	ARKANSAS	AR
06	CALIFORNIA	CA
08	COLORADO	CO
09	CONNECTICUT	CT
10	DELAWARE	DE
12	FLORIDA	FL
13	GEORGIA	GA
15	HAWAII	HI
16	IDAHO	ID
17	ILLINOIS	IL
18	INDIANA	IN
19	IOWA	IA
20	KANSAS	KS

Structure of acquired data, formatted as a data type that we'll handle in a conversion program

Seven Stages: Filter

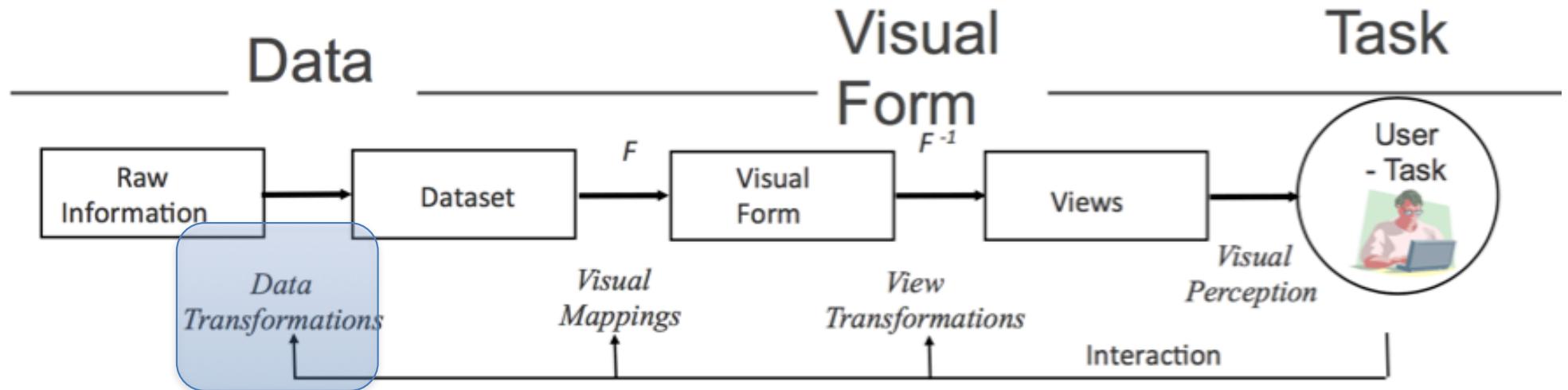


- Remove all but the data of interest.

00210	+43.005895	-071.013202	U	PORTSMOUTH	33	015
00211	+43.005895	-071.013202	U	PORTSMOUTH	33	015
00212	+43.005895	-071.013202	U	PORTSMOUTH	33	015
00213	+43.005895	-071.013202	U	PORTSMOUTH	33	015
00214	+43.005895	-071.013202	U	PORTSMOUTH	33	015
00215	+43.005895	-071.013202	U	PORTSMOUTH	33	015
00501	+40.922326	-072.637078	U	HOLTSVILLE	36	103
00544	+40.922326	-072.637078	U	HOLTSVILLE	36	103
00601	+18.165273	-066.722583		ADJUNTAS	72	001
00602	+18.393103	-067.180953		AGUADA	72	003
00603	+18.455913	-067.145780		AGUADILLA	72	005
00604	+18.493520	-067.135883		AGUADILLA	72	005
00605	+18.465162	-067.141486	P	AGUADILLA	72	005
00606	+18.172947	-066.944111		MARICAO	72	093
00610	+18.288685	-067.139696		ANASCO	72	011

Filter out some data points
remain only some data fields

Seven Stages: Mine



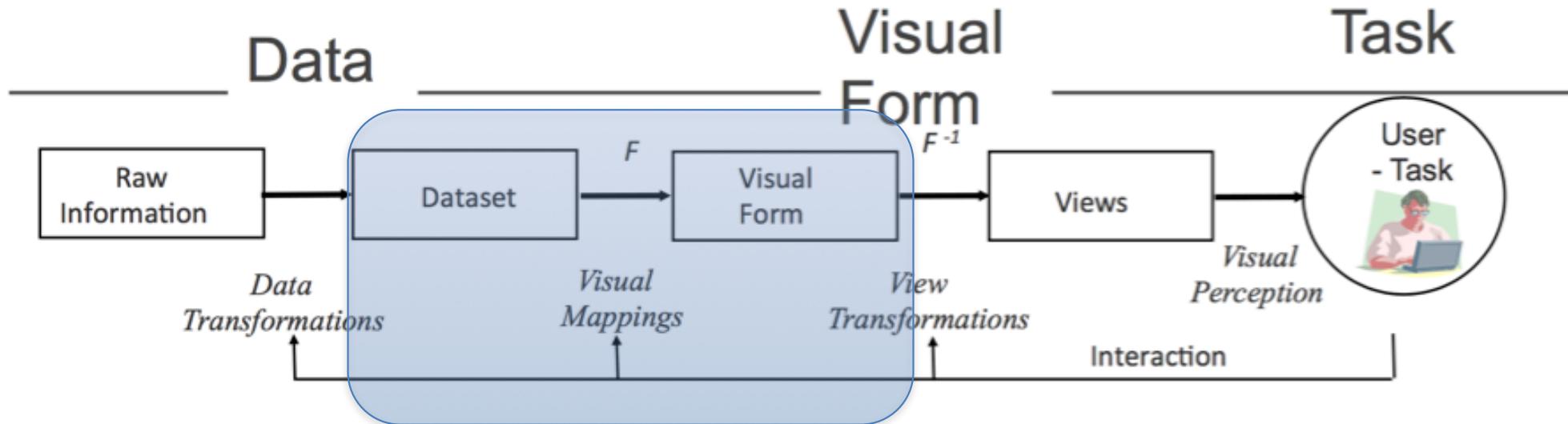
- Apply methods from statistics or data mining as a way to discern patterns or place the data in mathematical context.

00213	43.005895	-71.013202	PORTSMOUTH	NH
00214	43.005895	-71.013202	PORTSMOUTH	NH
00215	43.005895	-71.013202	PORTSMOUTH	NH
00501	40.922326	-72.637078	HOLTSVILLE	NY
00544	40.922326	-72.637078	HOLTSVILLE	NY
+	+	+	+	+
+	+	+	+	+
+	+	+	+	+

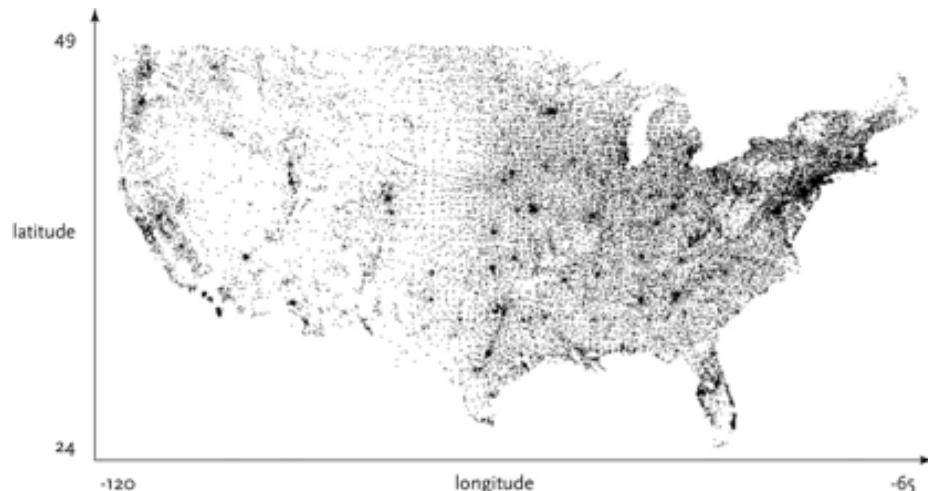
↓ ↓

min 24.655691	min -124.62608
max 48.987385	max -67.040764

Seven Stages: Represent

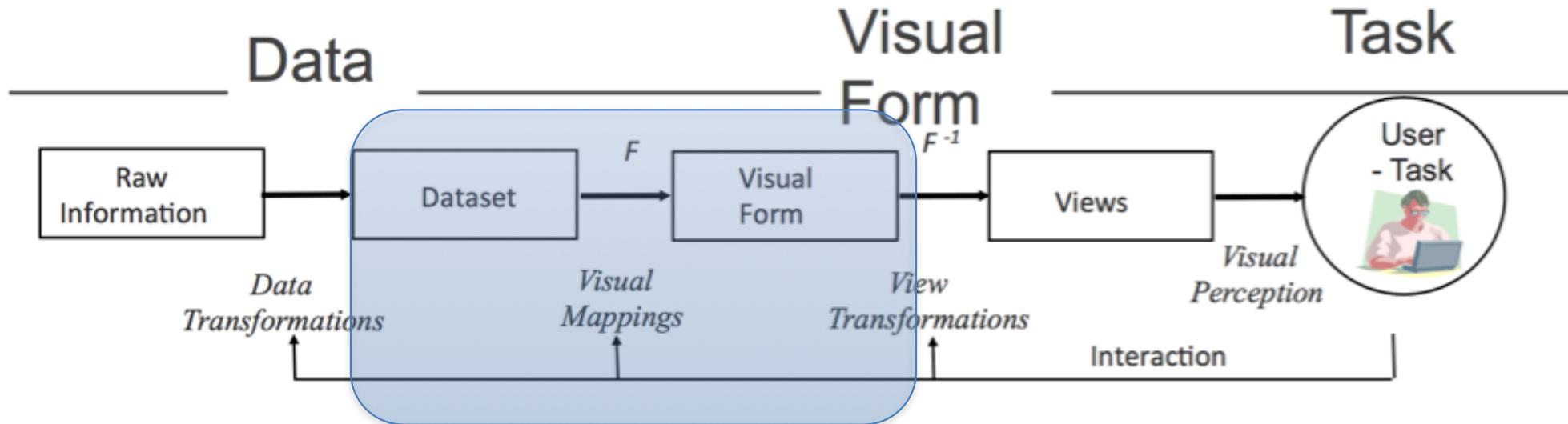


- Choose a visual model, such as a bar graph, list, or tree.

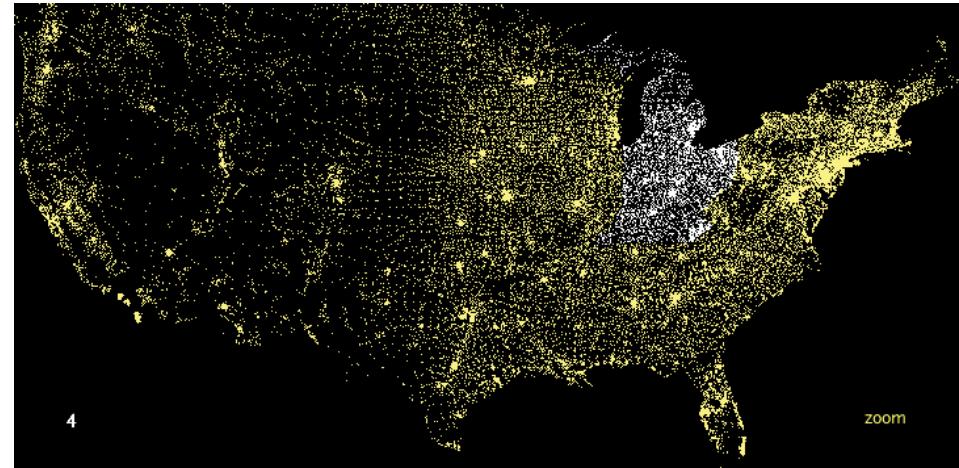


Basic visual representation of zip code data

Seven Stages: Refine

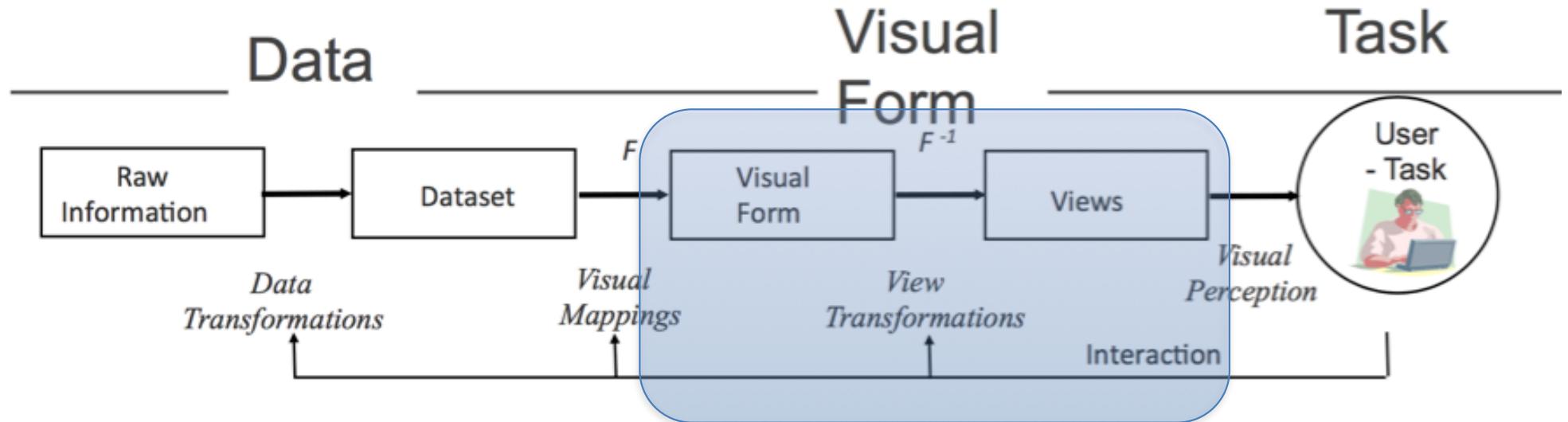


- Improve the basic representation to make it clearer and more visually engaging.

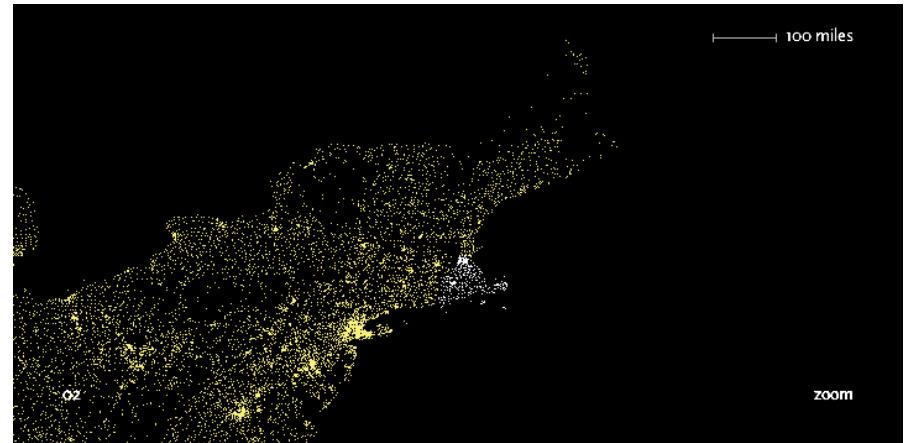


Using color to refine the representation

Seven Stages: Interact



- Add methods for manipulating the data or controlling what features are visible.



Zooming in with two digits of the post code (02)

Interaction is Vital for Exploration

- Engage in a dialog with your data
- Employ interaction in a more fundamental manner to strengthen the power of visualization
- Possible Actions
 - Select
 - Explore
 - Reconfigure
 - Encode
 - Abstract/Elaborate
 - Filter
 - Connect

Yi, et al. "Toward a deeper understanding of the role of interaction in information visualization." 2007.

Dr. Ke Zhou (<http://www.cs.nott.ac.uk/~pszkz/>)