Information Visualization



Tobias Isenberg tobias.isenberg@inria.fr

After today you will...

- have gained an overview of the research area
- learned basic principles of data representation and interaction

INFORMATION VISUALIZATION

Why



It is estimated that 10 zettabytes (10 x 10²²) of digital information will be generated this year

The Big Data Revolution

Sensors and loggers generate more and more data





Source: IDC's Digital Universe Study, sponsored by EMC, December 2012





It's not easy to get a handle on jobs in data science. However, data from O'Reilly Research shows a steady year-over-year increase in Hadoop and Cassandra job listings, which are good proxies for the "data science" market as a whole. This graph shows the increase in Cassandra jobs, and the companies listing Cassandra positions, over time.

"The ability to take data -- to be able to understand it, to process it, to extract value from it, to visualize it, to communicate it that's going to be a hugely important skill in the next decades."

Hal Varian, chief economist at Google

Question

how can we effectively access data?

- understand its structure?
- make comparisons?
- make decisions?
- gain new knowledge?
- convince others?

-...

Many possible ways to address...



Example

I		II		Ш		IV		
x	У	x	У	х	У	х	У	
10.0	8.04	10.0 9.14		10.0	7.46	8.0	6.58	
8.0	6.95	8.0 8.14		8.0	6.77	8.0	5.76	
13.0	7.58	13.0	13.0 8.74		13.0 12.74		7.71	
9.0	8.81	9.0 8.77		9.0 7.11		8.0	8.84	
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47	
14.0	9.96	14.0 8.10		14.0 8.84		8.0	7.04	
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25	
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50	
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56	
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91	
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89	

Raw Data from Anscombe's Quartet

Statistical Analysis

For all four columns, the statistics are identical

I		I	I	I	II	IV		
x	У	x	У	x	У	x	У	
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58	
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76	
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71	
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84	
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47	
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04	
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25	
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50	
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56	
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91	
5.0	5.68	5.0 4.74		5.0	5.73	8.0	6.89	

Mean of x	9.0
Variance of <i>x</i>	11.0
Mean of <i>y</i>	7.5
Variance of y	4.12
Correlation between x and y	0.816
Linear regression line	<i>y</i> = 3 + 0.5 <i>x</i>

Visual Representation of the Data

Visual representation reveals a different story

4

	12-								
• 1	10-	IV		Ш		I II			
•	년 8-	У	x	У	x	У	x	У	x
	6-	6.58	8.0	7.46	10.0	9.14	10.0	8.04	10.0
•	4-	5.76	8.0	6.77	8.0	8.14	8.0	6.95	8.0
	- 47	7.71	8.0	12.74	13.0	8.74	13.0	7.58	13.0
4 6 8 10 12 14 x1		8.84	8.0	7.11	9.0	8.77	9.0	8.81	9.0
XI		8.47	8.0	7.81	11.0	9.26	11.0	8.33	11.0
•		7.04	8.0	8.84	14.0	8.10	14.0	9.96	14.0
	12 -	5.25	8.0	6.08	6.0	6.13	6.0	7.24	6.0
	10-	12.50	19.0	5.39	4.0	3.10	4.0	4.26	4.0
	<u>қ</u> 8-	5.56	8.0	8.15	12.0	9.13	12.0	10.84	12.0
	6-	7.91	8.0	6.42	7.0	7.26	7.0	4.82	7.0
	4-	6.89	8.0	5.73	5.0	4.74	5.0	5.68	5.0
· · · · · ·	I I								



12 [Source: Anscombe's quartet, Wikipedia]

Why visual data representations?

- Vision is our most dominant sense
- We are very good at recognizing visual patterns
- We need to see and understand in order to explain, reason, and make decisions



Other benefits of visualization

- expand human working memory
 - offload cognitive resources to the visual system,
- reduce search
 - by representing a large amount of data in a small space,
- enhance the recognition of patterns
 - by making them visually explicit
- aid monitoring of a large number of potential events
- provides a manipulable medium & allows exploration of a space of parameter values.



L'occhio, che si dice finestra dell'anima, è la principale via donde il comune senso può piú copiosamente e magnificamente considerare le infinite opere di natura.

> Leonardo da Vinci (1452 - 1519)

The eye... the window of the soul, is the principal means by which the central sense can most completely and abundantly appreciate the infinite works of nature.

写開不如一見 "One hundred rumors are not comparable to one look." An Old Chinese Inscription

Via Brinton, Graphic Presentation, 1939

Information visualization

- Create visual representation
- Concentrates on abstract data
- Includes interaction

Official Definition:

The use of computer-supported, interactive, visual representations of abstract data to amplify cognition. [Card et al., 1999]



Functions of Visualizations

- Recording information
 - Tables, blueprints, satellite images
- Processing information
 - needs feedback and interaction
- Presenting information
 - share, collaborate, revise
 - for oneself, for one's peers and to teach
- Seeing the unseen

Visualization of abstract data has been practiced for hundreds of years...

HISTORICAL EXAMPLES

Napoleon's March on Moscow Charles Minard, 1869

Named the best statistical graphic ever drawn (by Edward Tufte)

- Includes: spatial layout linked with stats on: army size, temperature, time
- Tells a story in one overview



Autog. par Regnier, 8. Pas. Ste Marie St Gain à Paris.

More info: The Visual Display of Quantitative Information (Tufte)

The Broadway Street Pump

- In 1854 cholera broke out in London
 - 127 people near Broad Street died within 3 days
 - 616 people died within 30 days
- "Miasma in the atmosphere"
- Dr. John Snow was the first to link contaminated water to the outbreak of cholera
- How did he do it?
 - he talked to local residents
 - identified a water pump as a likely source
 - used maps to illustrate his theory
 - convinced authorities to disable the pump





John Snow, 1854

... AND MORE RECENTLY

TrashTrack



Winner of the NSF International Science & Engineering Visualization Challenge! http://senseable.mit.edu/trashtrack/

Artificial Intelligence



http://www.turbulence.org/spotlight/thinking/chess.html

Open Data

- Movement making government data freely available
- Encourage participation by everyone



Many Eyes

Visualizations : US government expenses 1962-2004

- Upload data, create visualizations, discuss •
- Distributed asynchronous collaboration •



http://www-958.ibm.com/software/data/cognos/manyeyes/

Specific Visualization Environments



Molecular visualisation in the Reality Cube University of Groningen, NL



Tabletops for Visualization University of Calgary



WILD Wall, INRIA

Software Visualization

EZEL: a Visual Tool for Performance Assessment of Peer-to-Peer File-Sharing Networks (Voinea et al., InfoVis, 2004)



Text Visualization

Parallel Tag Clouds to Explore Faceted Text Corpora (Collins et al., VAST 2009)

adverted	adjourned	allocatur	adequate	bankruptcy	bargaining	about	abuse	abuse	appeal	ballot	accused	agency		
anent	alia	analysis antitrust	affirmed	barge	benefit brief	asked	affirmed	aliens	assistant	black	annuity	agency's		
appellant	arbitration	app	aid	capital cargo	Cost	called	appellee	appropriate	attached ^{binding}	boat	antidumping	air seitear	1996	33/25269
	asbestos	arbitration	appeal	charter	cocaine	CMPR	argued	asylum	brief	candidate	application	authority bargaining		
appellant's	closure	asbestos	appalies	coverage	court	cocaine	baleve	circuit	cited	case	art	baiganing		
appellee	commenced	bankruptcy	argument	creature.	defendant	conspiracy could deal	cocaine	cited	collateral	class	board	broadcast		
asseveration	complaint	believe	before 🥒	damages death	defendant's	defendant	crack denied	contended	сору	commerce	claim	capricious		and the second sec
b e l ow	copyright	benefit	coal	debtor	delivered	disciplinary	disability	court courts crime	deficient	compinecy	composition compounds comprising	carrier		
brief	date	bottlers	cocaine	drilling	denied	enough	distribution	dba declared	determine	county	construction	competition		
carlacking	defendant	can	conspira coal	*						death	contract	COSTS		
commonwealth	disenfranchised	class		re (cases)		606.22	Relative Frequen	cv (cases)	4.02 per 1000	desegregation	costs clubs	disposition		
defendant del	foreign	context	court	· 1						isenfranchised	decision description	emissions		
ensued	fraud	debtor	crack							dozer	device	employees exemption		
error	heroin	cliet	decisio 🐜	Second Third Fourth Fifth	Soth Seventh Eighth Neth Tenth I	Beventh Federal DC	First Second Third Posth	Fith Sidh Severth Eighth N	rth Terth Eleverth Paderal DC	election	disclosed	explanatory		
event	injunction	exercise	denied Sco	re (occurrences)		4505.33	Relative Frequen	ry foccurrences]	0.25 per 1000	electors	embodied	gas		
ferritin	inter	fiduciary	disclos	ine (occurrences)		4000.00	Relative Frequen	ey [occurrences]	0.20 per 1000	mmunities	equivalent	hazard	1997	17/19340
guidelines	internal	have	dispen 💳						_	insurance	inequitable	interpretation		
here incarcerative	10.0°	here	distributio	Second Third Forth Fifth	Sidh Seventh Eighth Neth Tenth I	Deventh Federal DC	First Second Third Pourts	Fith Such Serverth Eagth N	rth Tarth Deverth Palarat DC	ivory	infringement	intervenor		
inference	keeplock	inasmuch insurance	distri	партну	judgment	job			judicata	jail	invalid	labor license		
jury	marks	interest	fact	lien Ioan	magistrate	judge	jury	his	material	law	invention	memoranda		
limned	millions	jurisdiction	from	marihuana	magistrate's	Judge	medical	iurisdiction	nevertheless	migrant		operation		
Ist	narcotics	legislation	his	maritime	marijuana	kilograms m	ethamphetan		now	mitigation	inventor	petitioner		
might	omitted payment	liability	interlocutory joined	mitigation	motion	lawyer	months	may	opinion	nonstatutory ordinance	layer Insted	pipelines		
more mortgage	plaintiff	ligation	legal	negligence	office	might	office	native	oral	payday	means	preceding promulgated		
pedicant	plaintiff's	majority market	magistrate	nre	panel	more	opinion	novo	order	phase	merchandise	proposed	1998	4/13693
plausible	principal	_	ranjor by	offshare	paupers	one	pain	panel	persuasive	preceding	m eth o d	quality		_
point receipt	quotation	notes	material	parish	plaintiff	out	pounds	persecution	paintiff's	qualified	noninfringement	rate		
sold	racketeering	our	merits miner's	pet	plaintiff's	police	quantity	petition	precedential	racial	patent	regulations		
say	reinsurance	pension	mining	platform	pneumoconiosis	prisoner	reversed search	political	record	section	•	rehearing		
see	respect	plaintiff	opinion	Policy y	price	say	sentence	prisoner	remained	sentence	patentee	reprinted		
some	security	plan	oral	refd	pulmonary	she	sexual	public	submitted	sheriff	product	rulemaking		
supra	See	plenary	order	removed	pursuant	suit	she	pursuant	suspended	students	rected	section		
think	shares	policy	pneumoconiosi		recommendation	supra	subd	section	taib Rocolom	trial	retirement	see		
tit	sterile stock	product	present	seaman	search	tentative than	suppression	suitable	tit	turtle	said	service		
town	subway	recognized recognized	process	servitude	sentence	thought	testified	inter .	unanimous	tusks vessel	signal skill	shipper		
trialworthy	summation	section	published	suit	sitting	told	told	tribal	unfavorable	vote	specific	tariff		
vessel	trade	settlement	pulmonary	usury	unanimous	Bo	trial	tribe	unpublished	voters	structure	transmission	1999	8/7760
viz	vacated	syrup	sentence	vessel	union	want	unpublished	unanimous	value	white	surface	union		
whom	waybill	which	wrote	writ	warrant	what sto	verdict	water	vehicle VOI	-	vaccination	united		
WIIOIII	where	would	wrote	VVITL	warrant	would	work	without	VUI	zone	veterans	waste		
			Fourth		Sixth									
Cases Score	e 🔘 Cases Rar	nk 🔘 Occurren	ice Score 💿 O	ccurrence Rank					Search:					
Options	Supreme (337) First (8311) Second (14623	Fourth (234	32) Seventh	(17557) Tenth	(15635) DC (6	ral (10610) 6454) own (518)	Start year:	1990 💭 End	year: 2007 🔹	Visualize				.

Graphs



http://www.facebook.com/note.php?note_id=469716398919 Visualizing Friendships by Paul Butler on Tuesday, December 14, 2010

Family Trees



http://www.aviz.fr/geneaquilts/



Weather



http://weatherspark.com/

Data Dashboards



GLOBALL SPIROMETRY

http://globalspirometry.com

Resources for more examples

- Visualization conferences
- Blogs
 - <u>http://infosthetics.com/</u>
 - <u>http://fellinlovewithdata.com/</u>
 - <u>http://eagereyes.org/</u>
 - <u>http://flowingdata.com/</u>
 - <u>http://www.informationisbeautiful.net/</u>
- Books
 - Textbooks
 - Readings in Information Visualization: Using Vision to Think (a bit old now but good intro)
 - Information Visualization (Robert Spence a light intro, I recommend as a start)
 - Information Visualization Perception for Design (Colin Ware, focused on perception and cognition)
 - Interactive Data Visualization: Foundations, Techniques, and Applications (Ward et al. most recent)
 - Examples
 - Beautiful Data (McCandless)
 - Now You See it (Few)
 - Tufte Books: Visual Display of Quantitative Information (and others)
 - ... (many more, ask me for details)
It is difficult to create

CREATE VISUALIZATIONS



What is a representation?

- A representation is
 - a formal system or mapping by which the information can be specified (D. Marr)
 - a sign system in that it stands for something other than its self.
- for example: the number thirty-four



Presentation

• different representations reveal different aspects of the information

decimal: counting & information about powers of 10,

binary: counting & information about powers of 2,

roman: impress your friends (outperformed by positional system)

• presentation

how the representation is placed or organized on the screen

Principles of Graphical Excellence

- Well-designed presentation of interesting data a matter of *substance*, *statistics*, *design*
- Complex ideas communicated with clarity, precision, efficiency
- Gives the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space
- Involves almost always multiple variables
- Tell the truth about the data

Or a bit more simply...

- Solving a problem simply means representing it so as to make the solution transparent ... (Simon, 1981)
- Good representations:
 - allow people to find relevant information
 - information may be present but hard to find
 - allow people to compute desired conclusions
 - computations may be difficult or "for free" depending on representations

Good representation?

. Table - StateData ()					
		Load Snap		30.4%	14389
State	College Degree %	Per Capita Income	Mississippi	19.9%	9648
Alabama	20.6%	11486	Missouri	22.3%	12989
Alaska	30.3%	17610	Montana	25.4%	11213
			Nebraska	26.0%	12452
Arizona	27.1%	13461	Nevada	21.5%	15214
Arkansas	17.0%	10520	New Hampshire	32.4%	15959
California	31.3%	16409	New Jersey	30.1%	18714
Colorado	33.9%	14821	New Mexico	25.5%	11246
Connecticut	33.8%	20189	New York	29.6%	16501
Delaware	27.9%	15854	North Carolina	24.2%	12885
District of Columbia (36.4%	18881	North Dakota	28.1%	11051
Florida	24.9%	14698	Ohio	22.3%	13461
Georgia	24.3%	13631	Oklahoma	22.8%	11893
			Oregon	27.5%	13418
Hawaii	31.2%	15770	Pennsylvania	23.2%	14068
Idaho	25.2%	11457	Rhode Island	27.5%	14981
Illinois	26.8%	15201	South Carolina	23.0%	11897
Indiana	20.9%	13149	South Dakota	24.6%	10661
lowa	24.5%	12422	Tennessee	20.1%	12255
Kansas	26.5%	13300	Texas	25.5%	12904
Kentucky	17.7%	11153	Utah	30.0%	11029
Louisiana	19.4%	10635	Vermont	31.5%	13527
Maine	25.7%	12957	Virginia	30.0%	<u>15713</u> 14923
			Washington West Virginia	30.9%	14923
Maryland	31.7%	17730	West Virginia Wisconsin	24.9%	13276
Massachusetts	34.5%	17224	Wisconsin	25.7%	42311
Michigan	24.1%	14154	Wyoming	20.776	42311
Minnesota	30.4%	14389			

Good representation!



How do we arrive at a visualization?



Interaction

The Visualization Pipeline

From [Spence, 2000]

Visualization Reference Model

Also a visualization pipeline a bit expanded



From [Card et al., Readings in Information Visualization]

Visualization pipeline in an image



[Tobiasz et al., 2009]

Knowledge Crystallization Cycle



Working with visualizations in NOT a linear process

[Card et al., 1999]

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
- Choosing the wrong representation
- Choosing the wrong presentation format
- Inappropriate interactions provided to explore the data

Recap

- So far you
 - learned what information visualization is
 - learned about the advantages of visualization
 - saw a number of examples (historical and new)
- Next
 - you will get to know your data
 - you will learn about the basic components of visualization

Data

- Data is the foundation of any visualization
- The visualization designer needs to understand
 - the data properties
 - know what meta-data is available
 - know what people want from the data

Nominal, Ordinal and Quantitative

- Nominal (labels)
 - Fruits: apples, oranges
- Ordered
 - Quality of meat: grade A, AA, AAA
 - Can be counted and ordered, but not measured
- Quantitative: Interval
 - no clear zero (or arbitrary)
 - e.g. dates, longitude, latitude
 - usually compare differences (intervals)
- Quantitative: Ratio
 - meaningful origin (zero)
 - physical measurements (temperature, mass, length)
 - counts and amounts

Nominal, Ordinal and Quantitative

- Nominal (labels)
 - Operations: =, ≠
- Ordered
 - Operations: =, ≠, <, >
- Quantitative: Interval
 - Operations: =, ≠, <, >, -, +
 - Can measure distances or spans
- Quantitative: Ratio
 - Operationrs: =, ≠, <, >, -, +, ×, ÷
 - Can measure ratios or proportions



[1989 - 1999] + [2002 - 2012]

10kg / 5kg

Data-Type Taxonomy

- 1D (linear)
- Temporal
- 2D (maps)
- 3D
- nD (relational) vis examples later
- Trees (hierarchies)
- Networks (graphs)



Why is this important?

- Nominal, ordinal, and quantitative data are best expressed in different ways visually
- Data types often have inherent tasks
 - temporal data (comparison of events)
 - trees (understand parent-child relationships)
 - ...
- But:
 - any data type (1D, 2D,...) can be expressed in a multitude of ways!

Visualization's Main Building Blocks

Marks which represent:



From Semiology of Graphics (Bertin)

Points

- "A point represents a location on the plane that has no theoretical length or area.
 This signification is independent of the size and character of the mark which renders it visible."
- a location
- marks that indicate points can vary in all visual variables



Lines

- "A line signifies a phenomenon on the plane which has measurable length but no area. This signification is independent of the width and characteristics of the mark which renders it visible."
- a boundary, a route, a connection



Areas

 "An area signifies something on the plane that has measurable size.

This signification applies to the entire area covered by the visible mark."

 an area can change in position but not in size, shape or orientation without making the area itself have a different meaning



Visual Variables Applicable to Marks



From Semiology of Graphics (Bertin)

Additional Variables for Computers

motion

direction, acceleration, speed, frequency, onset, 'personality'





saturation

 colour as Bertin uses largely refers to hue, saturation != value

Extending those from Semiology of Graphics (Bertin)

Additional Variables for Computers

flicker

- frequency, rhythm, appearance
- depth? 'quasi' 3D
 - depth, occlusion, aerial perspective, binocular disparity
- Illumination

transparency







Characteristics of Visual Variables

• Selective:

Can this variable allow us to spontaneously differentiate/isolate items from groups?

• Associative:

Can this variable allow us to spontaneously group items in a group?

• Ordered:

Can this variable allow us to spontaneously perceive an order?

• Quantitative:

Is there a numerical reading obtainable from changes in this variable?

• Length (resolution):

Across how many changes in this variable are distinctions possible?

Visual Variable: Position



From Semiology of Graphics (Bertin)

Visual Variable: Size





Visual Variable: Shape

- infinite



Shape



points

lines

areas

Visual Variable: Value



- \neq · quantitative
 - ✓ order



length (resolution)

- theoretically infinite but practically limited
- association and selection ~ < 7 and distinction ~ 10

Value



points lines areas

Value

• Ordered, cannot be reordered



Values not ordered correctly according to scale Information has to be read point by point



Values ordered correctly Image much more useful

Visual Variable: Colour



length (resolution)

- theoretically infinite but practically limited
- association and selection ~ < 7 and distinction ~ 10

Visual Variable: Orientation



- ✓ length (resolution)
 - ~5 in 2D; ? in 3D

Orientation



points

lines

areas

Visual Variable: Texture



\$

Þ

≯



- order
- length(resolution)
 - theoretically infinite



4

Texture



points

lines

areas

Visual Variable: Motion

selective

- motion is one of our most powerful attention grabbers
- associative
 - moving in unison groups objects effectively
- ✓ quantitative
 - subjective perception
- \neq order
 - **?** length (resolution)
 - distinguishable types of motion?

Motion



Visual Variables

Visual Variable	Selective	Associative	Quantitative	Order	Length	
Position	Yes	Yes	Yes	Yes	Dependant on resolution	
Size	Yes	Yes	Approximate	Yes	Association: 5; Distinction: 20	
Shape	With Effort	With Effort	No	No	Infinite	
Value	Yes	Yes	No	Yes	Association: 7; Distinction: 10	
Hue	Yes	Yes	No	No	Association: 7; Distinction: 10	
Orientation	Yes	Yes	No	No	4	
Grain	Yes	Yes	No	No	5	
Texture	Yes	Yes	No	No	Infinite	
Motion	Yes	Yes	No	Yes	Unknown	

Summary

	Quantitative		Ordinal		Nominal	
More Accurate	Position	•••	Position	•••	Position	•••
↑	Length	=	Density	•••	Hue	•••
	Angle	2	Saturation	•••	Density	• • •
	Slope	11	Hue	•••	Saturation	•••
	Area	••	Length	=	Shape	• • =
	Density	•••	Angle	2	Length	_
	Saturation		Slope	11	Angle	2
¥	Hue	•••	Area	••	Slope	1-
Less Accurate	Shape	• • =	Shape	• • •	Area	••

Jacques Bertin refined by Cleveland&McGill then by Card&Mackinlay

Summary

- Now you know the main building blocks are marks
- Marks are modified by **visual variables**
- Visual variables have **specific characteristics**
- These characteristics influence how the data will be perceived