Information Visualization

James EAGAN

james.eagan@telecom-paristech.fr

Includes slides adapted from John Stasko (Georgia Tech), Petra Isenberg & Jean-Daniel Fekete (INRIA), Chris North (Virginia Tech), Tamara Munzner (UBC)

Dernière mise à jour : avril 2019.

Who am I?

James EAGAN

MAÎTRE DE CONFÉRENCES EN INTERACTION HOMME-MACHINE

Associate Prof. at Télécom ParisTech
Adjunct Researcher at LTCI

Ph.D. 2008 — Georgia Tech
Computer Science, Human-Computer Interaction

B.A. 2000 — Lawrence University
Mathematics/Computer Science

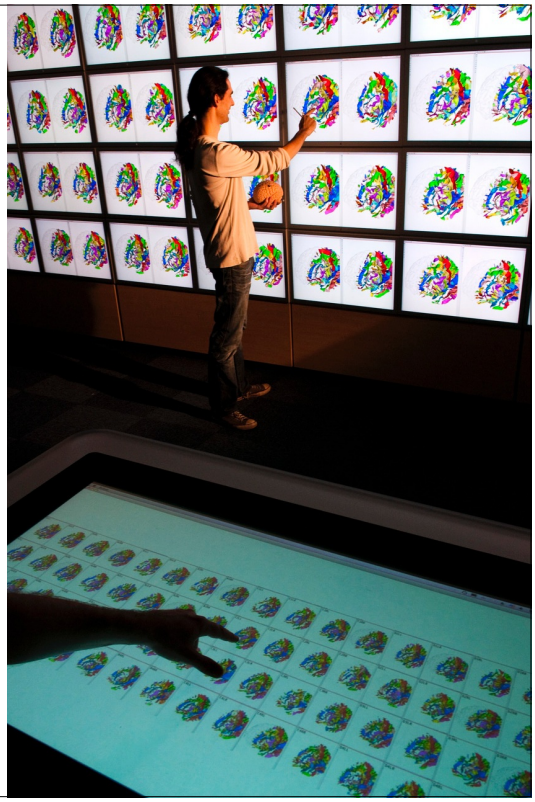
james.eagan@telecom-paristech.fr

Research

Human-Computer Interaction

Information Visualization

Multi-surface Interaction



5

Class overview

[www.telecom-paristech.fr/~eagan/class/igr204] 6

Data Exploration

Society is more complex

There is simply more “stuff”

Computers, internet, and web give people access to an incredible amount of data

news, sports, financial, sales, demographics, etc.

pollution, computer logs, weather, photos, videos, etc.

How much data?

Between 1 and 2 exabytes of unique info produced per year

1000000000000000000 (10^{18}) bytes

250 meg for every man, woman and child

Printed documents only .003% of total

Peter Lyman and Hal Varian, 2000
Cal-Berkeley, Info Mgmt & Systems
`www.sims.berkeley.edu/how-much-info`

2008

800 exabytes per year

[The Diverse and Exploding Digital Universe, IDC, 2008]

9

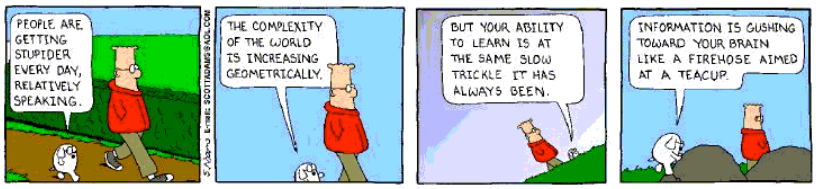
Data Overload

How can we make use of the data?

How do we make sense of the data?

How do we harness this data in decision-making processes?

How do we avoid being overwhelmed?



11

The need is there

"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's going to be a hugely important skill in the next decades."

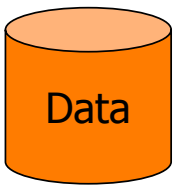
— Hal Varian, chief economist, Google

[The McKinsey Quarterly, January 2009]

The Challenge

Transform the data into information (understanding, insight) thus making it useful to people

Web,
Books,
Papers,
Game scores,
Scientific data,
Biotech,
Shopping
People
Stock/finance
News



Data Transfer

How?

Vision: 100 MB/s
Ears: <100 b/s
Telepathy
Haptic/tactile
Smell
Taste

[Courtesy of Chris North, Virginia Tech]

Human Vision

Highest bandwidth sense

Fast, parallel

Pattern recognition

Pre-attentive

Extends memory and cognitive capacity

(Multiplication test)

People think visually

Impressive. Lets use it!

[Courtesy of Chris North, Virginia Tech]

Example

Questions:

Which state has the highest income?

Is there a relationship between income and education?

Are there any outliers?

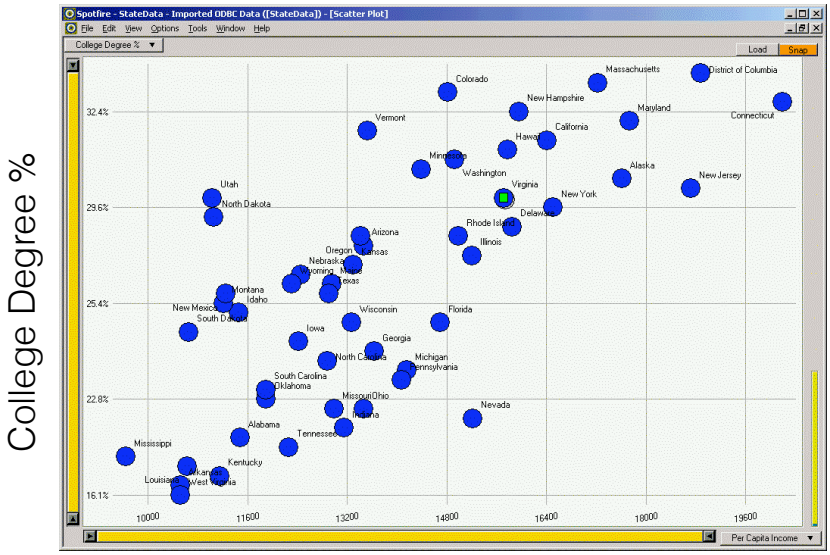
Table - StateData ()

Load | Snap

State	College Degree %	Per Capita Income
Alabama	20.6%	11486
Alaska	30.3%	17610
Arizona	27.1%	13461
Arkansas	17.0%	10520
California	31.3%	16409
Colorado	33.9%	14821
Connecticut	33.8%	20189
Delaware	27.9%	15854
District of Columbia	36.4%	18881
Florida	24.9%	14698
Georgia	24.3%	13631
Hawaii	31.2%	15770
Idaho	25.2%	11457
Illinois	26.8%	15201
Indiana	20.9%	13149
Iowa	24.5%	12422
Kansas	26.5%	13300
Kentucky	17.7%	11153
Louisiana	19.4%	10635
Maine	25.7%	12957
Maryland	31.7%	17730
Massachusetts	34.5%	17224
Michigan	24.1%	14154
Minnesota	30.4%	14389
Mississippi	19.9%	9648
Missouri	22.3%	12989
Montana	25.4%	11213
Nebraska	26.0%	12452
Nevada	21.5%	15214
New Hampshire	32.4%	15959
New Jersey	30.1%	18714
New Mexico	25.5%	11246
New York	29.6%	16501
North Carolina	24.2%	12885
North Dakota	28.1%	11051
Ohio	22.3%	13461
Oklahoma	22.8%	11893
Oregon	27.5%	13418
Pennsylvania	23.2%	14068
Rhode Island	27.5%	14981
South Carolina	23.0%	11897
South Dakota	24.6%	10661
Tennessee	20.1%	12255
Texas	25.5%	12904
Utah	30.0%	11029
Vermont	31.5%	13527
Virginia	30.0%	15713
Washington	30.9%	14923
West Virginia	16.1%	10520
Wisconsin	24.9%	13276
Wyoming	25.7%	12311

[Courtesy of Chris North, Virginia Tech]

Visualize the Data



[Courtesy of Chris North, Virginia Tech]

Even Tougher?

What if you could only see one state's data at a time? (e.g. U.S. Census Bureau's website)

What if I read the data to you?

I		II		III		IV	
x	y	x	y	x	y	x	y
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

19

Anscombe's Quartet

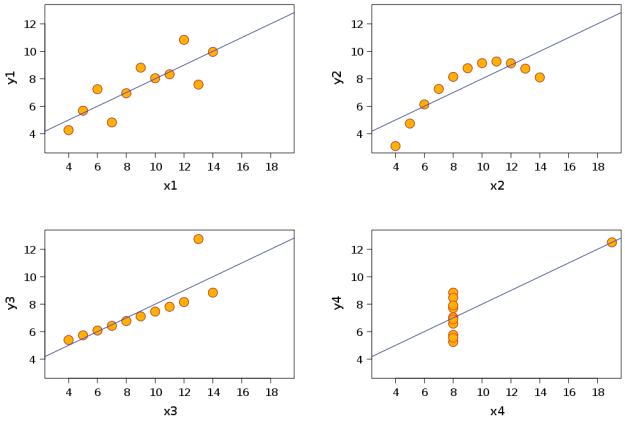
I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	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 x	11.0
Mean of y	7.5
Variance of y	4.12
Correlation between x and y	0.816
Linear regression line	$y = 3 + 0.5x$

20

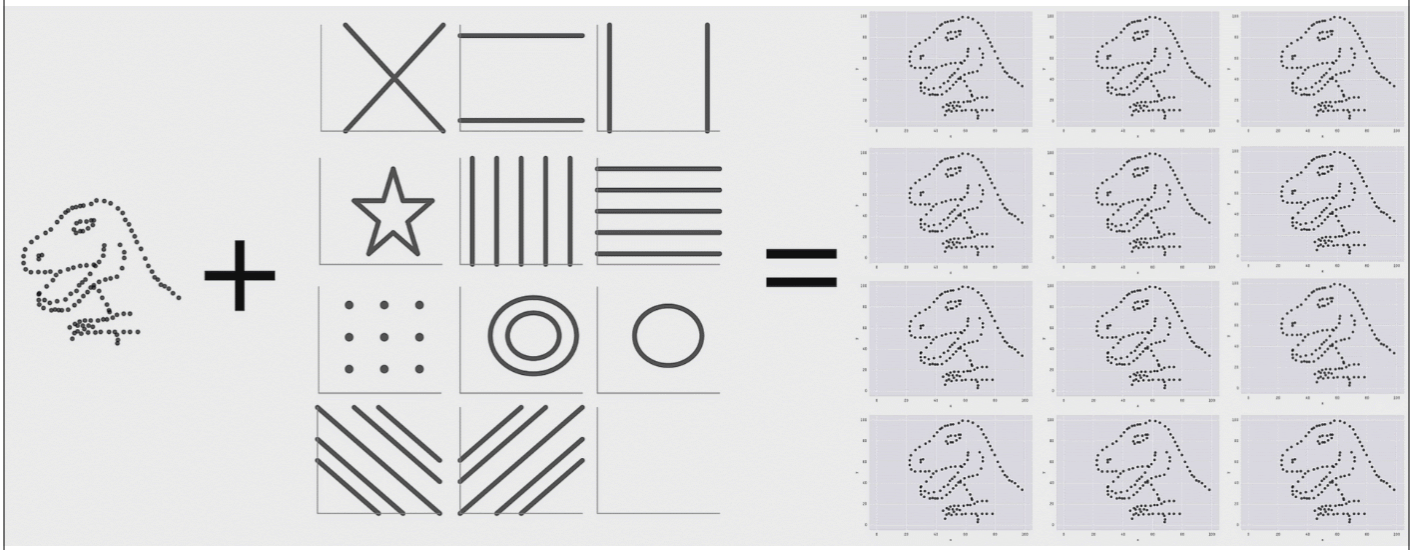
Anscombe's Quartet

I		II		III		IV	
x	y	x	y	x	y	x	y
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



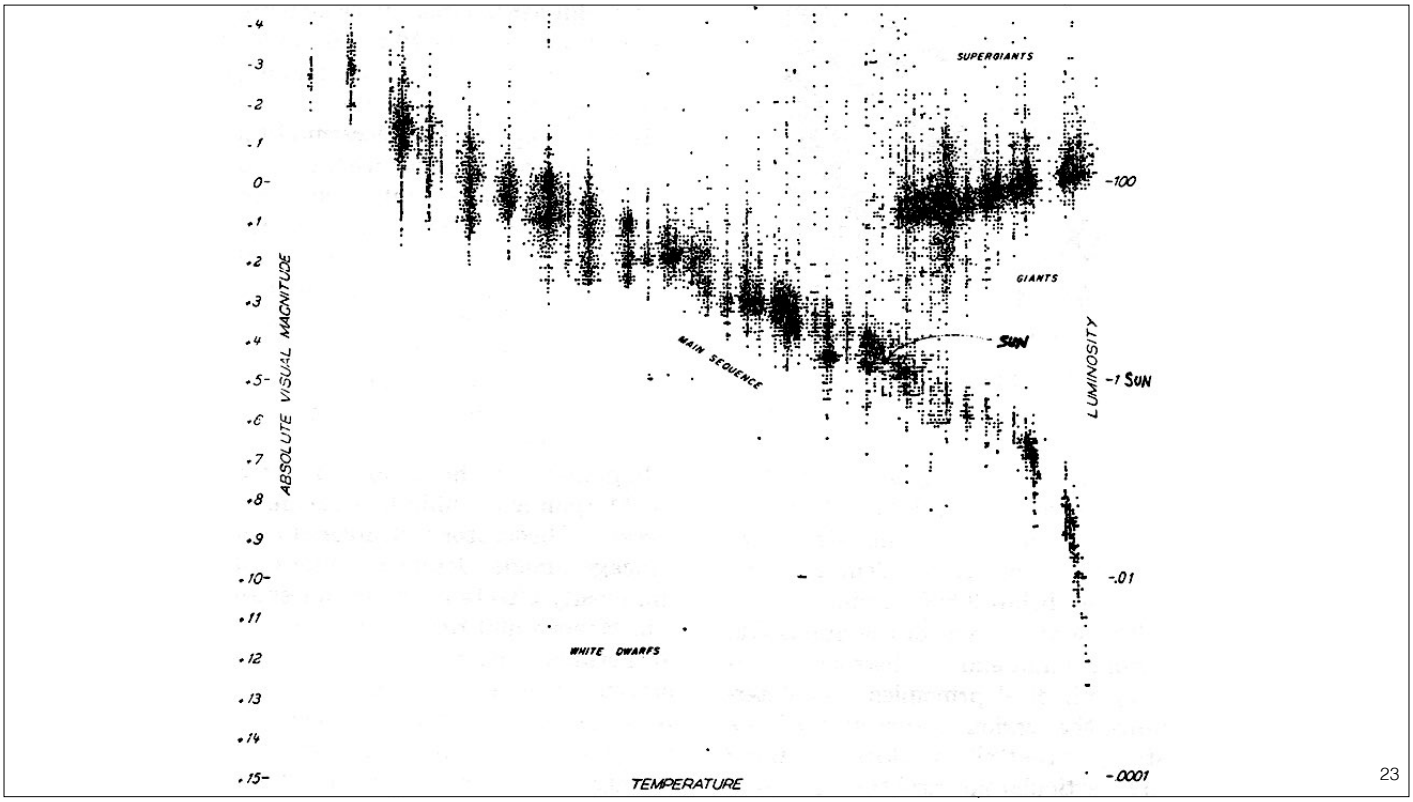
21

21



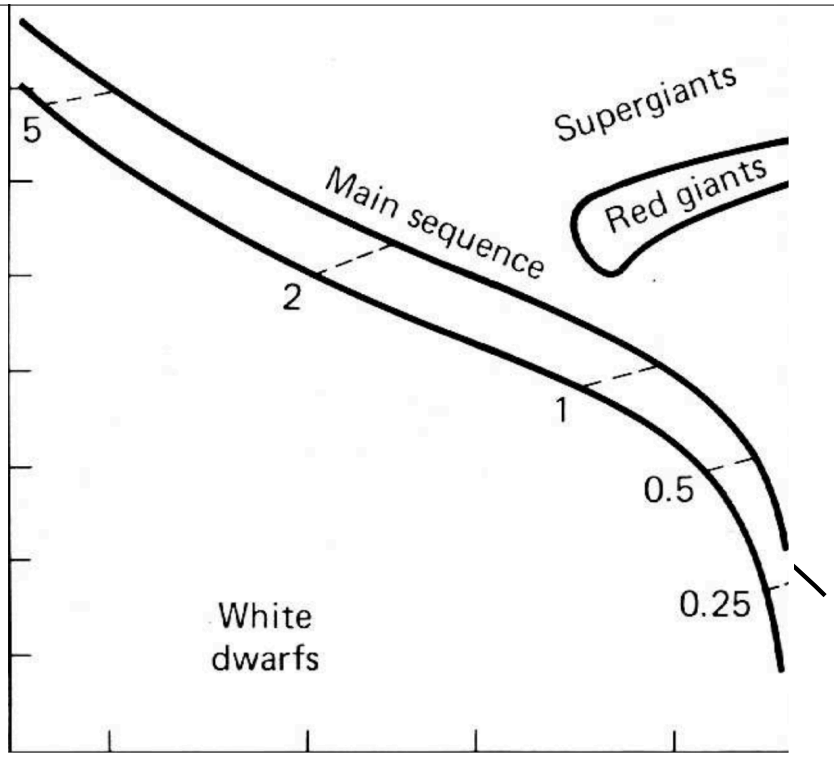
22

22


ABSOLUTE VISUAL MAGNITUDE
-4
-3
-2
-1
0-
+1
+2
+3
+4
+5-
+6
+7
+8
+9
+10-
+11
+12
+13
+14
+15-
SUPERGIANTS
GIANTS
MAIN SEQUENCE
SUN
WHITE DWARFS
TEMPERATURE
LUMINOSITY
-100
-1 SUN
-.01
-.0001

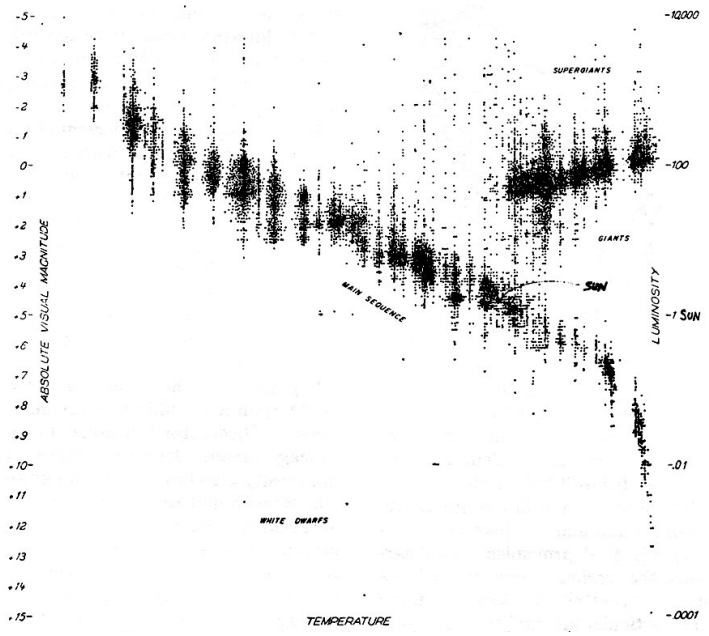
23

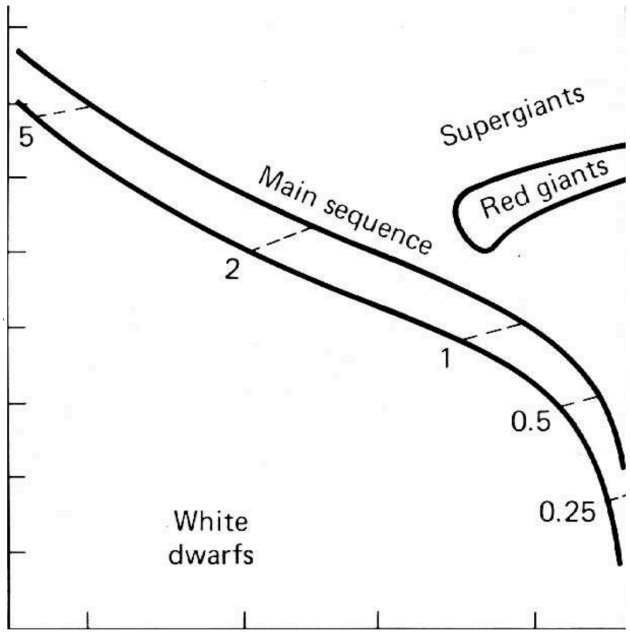
23


5
2
1
0.5
0.25
Supergiants
Red giants
Main sequence
White
dwarfs

24

24





25

25

Illustrates Our Approach

Provide tools that present data in a way to help people understand and gain insight from it

Clichés

"Seeing is believing"

"A picture is worth a thousand words"

26

Visualization

Often thought of as process of making a graphic or an image

Really is a cognitive process

- Form a mental image of something
- Internalize an understanding

“The purpose of visualization is insight, not pictures”

Insight: discovery, decision making, explanation

27

Main Idea

Visuals help us think

- Provide a frame of reference, a temporary storage area

External cognition

- Role of external world in thinking and reason

28

Information Visualization

What is “visualization”?

> The use of computer-supported, interactive visual representations of data to amplify cognition.
>
> From [Card, Mackinlay, Shneiderman ’98]

Information Visualization

What is “visualization”?

> The use of computer-supported, interactive visual representations of data to amplify cognition.
>
> From [Card, Mackinlay, Shneiderman ’98]

Three Subfields

Scientific Visualization

Information Visualization

Visual Analytics

31

Scientific Visualization

Primarily relates to and represents something physical or geometric

Examples

- Air flow over a wing
- Stresses on a girder
- Movement of clouds

32

Information Visualization

Components:

Taking items without a direct physical correspondence and mapping them to a 2-D or 3-D physical space.

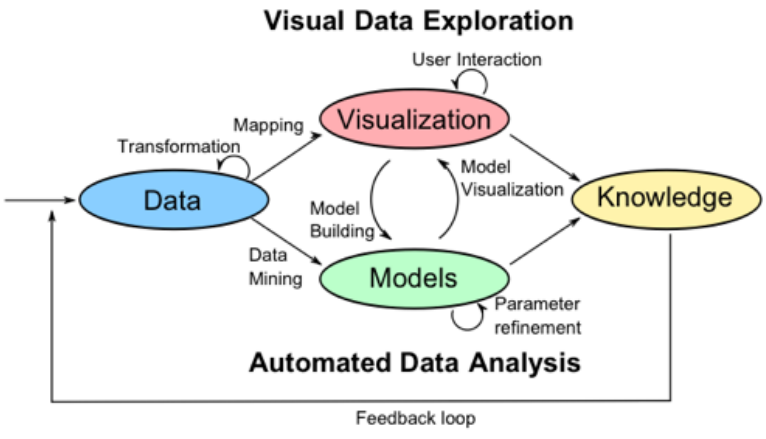
Giving information a visual representation that is useful for analysis and decision-making

33

Visual Analytics

Marry InfoVis with Data Mining

Human in control



34

Two Key Attributes

Scale

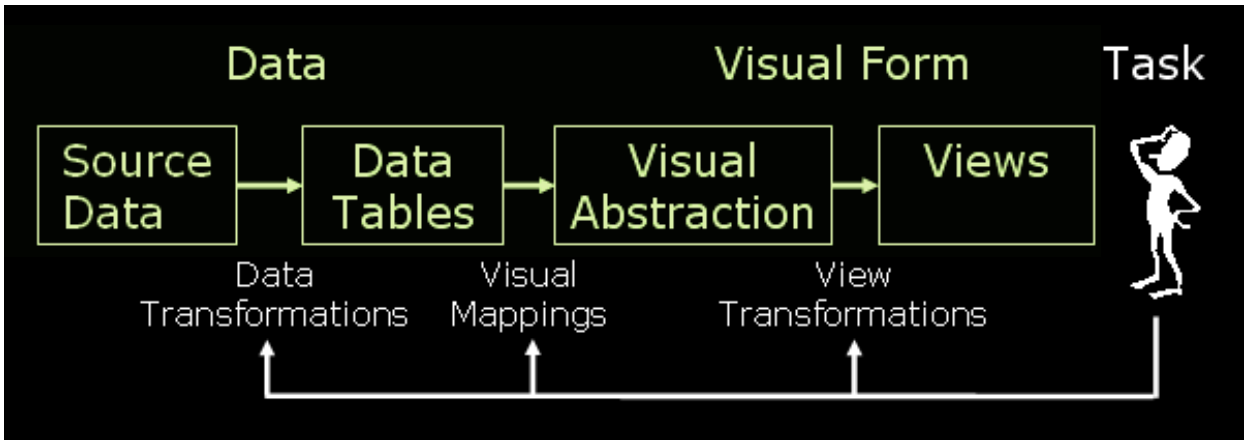
Challenge often arises when data sets become very large

Interactivity

Want to show multiple different perspectives on the data

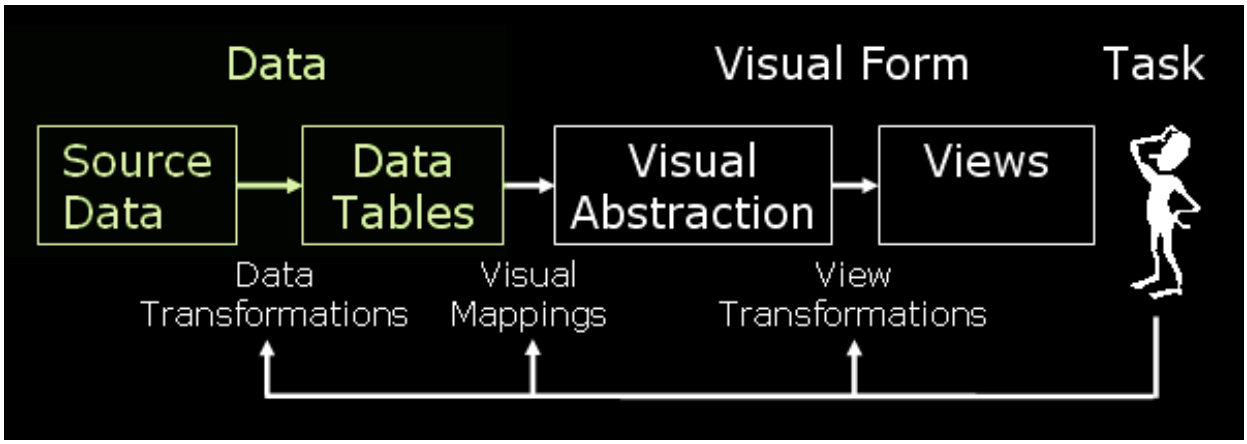
Data → Visualisation

InfoVis Pipeline



37

InfoVis Pipeline



38

Data Sets

Data comes in many different forms

Typically, not in the way you want it

How is stored (in the raw)?

Example

Cars
- make
- model
- year
- miles per gallon
- cost
- number of cylinders
- weights
- ...

Data Tables

Often, we take raw data and transform it into a form that is more workable

Main idea:

Individual items are called cases

Cases have variables (attributes)

41

Data Table Format

Dimensions

	$Variable_1$	$Variable_2$	$Variable_3$	$Variable_4$	…
$Case_1$	$Value_{1,1}$	$Value_{1,2}$	$Value_{1,3}$	$Value_{1,4}$	
$Case_2$	$Value_{2,1}$	$Value_{2,2}$	$Value_{2,3}$	$Value_{2,4}$	
$Case_3$	$Value_{3,1}$	$Value_{3,2}$	$Value_{3,3}$	$Value_{3,4}$	
$Case_4$	$Value_{4,1}$	$Value_{4,2}$	$Value_{4,3}$	$Value_{4,4}$	
⋮					

Think of as a function:

$f(case_i) = \langle value_{i,1}, value_{i,2}, \ldots, value_{i,n} \rangle$

42

Data Table Example

People in Class

	Hair	Age	GPA	ID	...
Marie	brown	23	12,3	901-12-3456	
Jean	black	17	14,6	901-12-4567	
Henri	blond	47	10,2	901-12-5678	
Bob	red	29	11,8	901-12-6789	
⋮					

43

Variable Types

Three main types of variables

Qualitative
- N-Nominal (equal or not equal to other values)
  - Example: gender
- O-Ordinal (obeys < relation, ordered set)
  - Example: fr,so,jr,sr

Q-Quantitative (can do math on them)

Can be *absolute* or *relative*
Example: age, temperature

44

Metadata

Descriptive information about the data

Might be something as simple as the type of a variable, or could be more complex

For times when the table itself just isn't enough

Example:

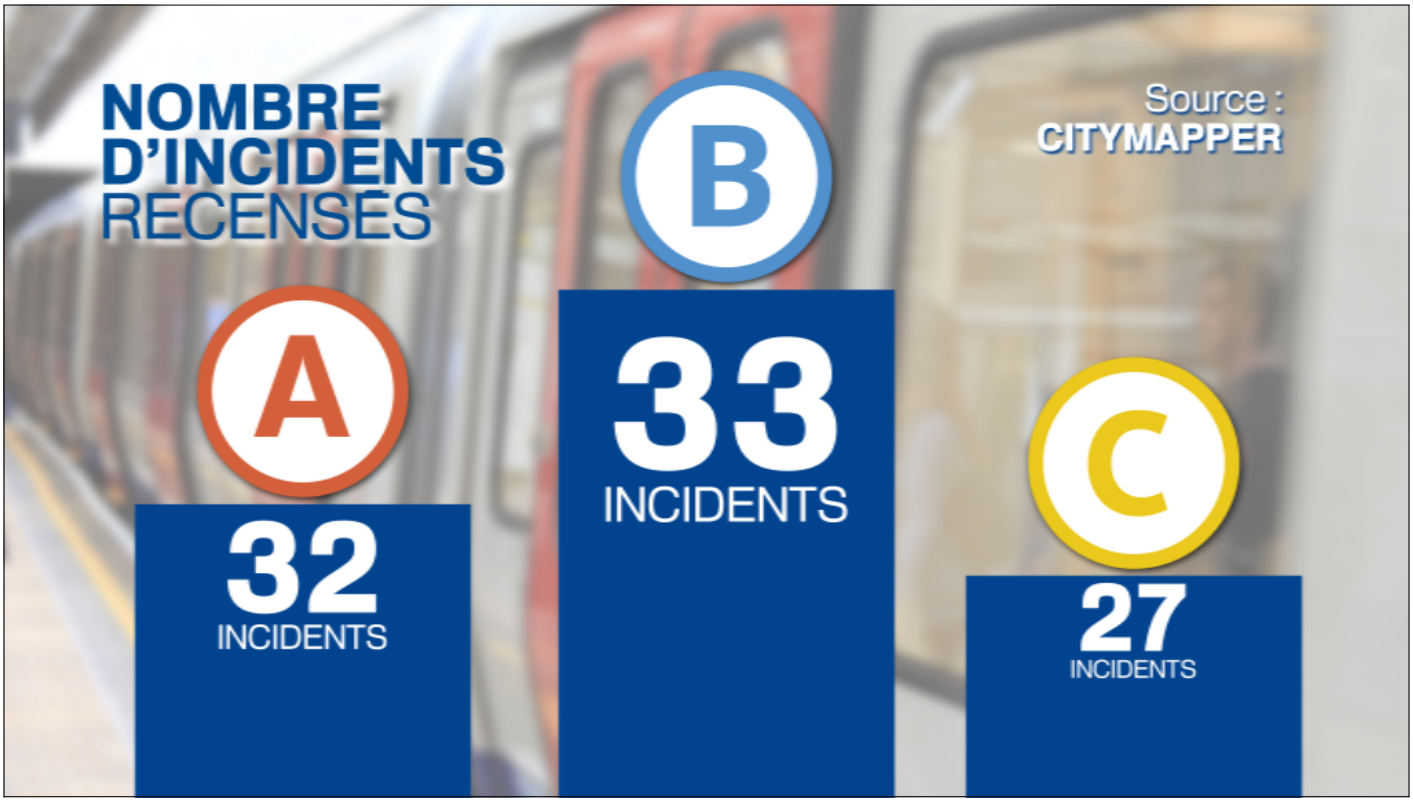
if car motor is electric, then L/100km is meaningless.

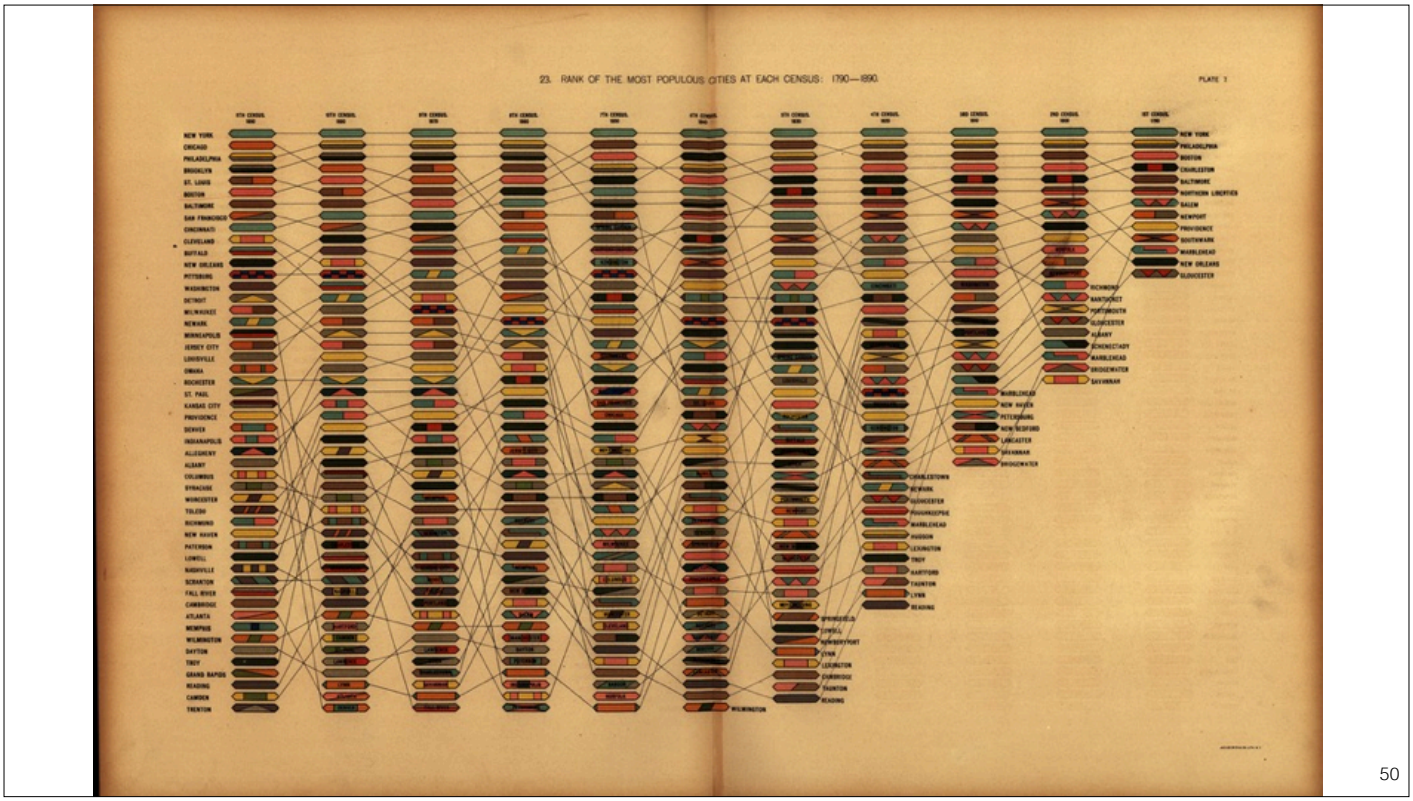
number of home runs ≤ number of at-bats

How do we show the data?

À suivre…

Teasers


NOMBRE D'INCIDENTS RECENSÉS
Source : CITYMAPPER
A
32
INCIDENTS
B
33
INCIDENTS
C
27
INCIDENTS


23. RANK OF THE MOST POPULOUS CITIES AT EACH CENSUS: 1790—1890.
PLATE 7

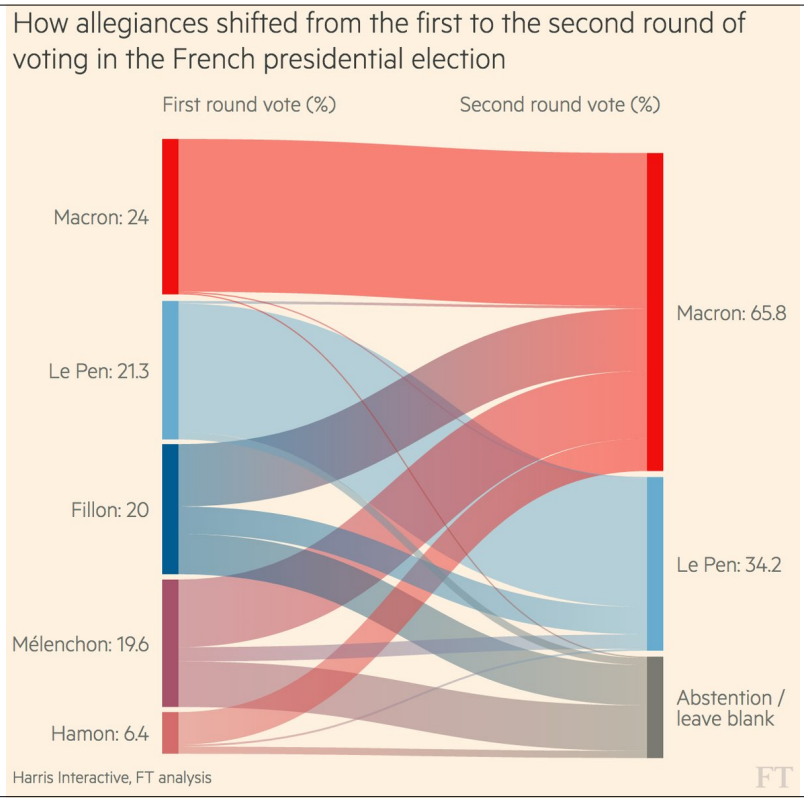
51

H 1 1.01																	He 2 4
Li 3 6.94	Be 4 9.01											B 5 10.81	C 6 12.01	N 7 14.01	O 8 16	F 9 19	Ne 10 20.18
Na 11 22.99	Mg 12 24.31											Al 13 26.98	Si 14 28.09	P 15 30.97	S 16 32.07	Cl 17 35.45	Ar 18 39.95
K 19 39.10	Ca 20 40.08	Sc 21 44.96	Ti 22 47.87	V 23 50.94	Cr 24 52	Mn 25 54.94	Fe 26 55.85	Co 27 58.93	Ni 28 58.69	Cu 29 63.55	Zn 30 65.38	Ga 31 69.72	Ge 32 72.64	As 33 74.92	Se 34 78.96	Br 35 79.90	Kr 36 83.80
Rb 37 85.47	Sr 38 87.62	Y 39 88.91	Zr 40 91.22	Nb 41 92.91	Mo 42 95.96	Tc 43 98	Ru 44 101.07	Rh 45 102.91	Pd 46 106.42	Ag 47 107.87	Cd 48 112.41	In 49 114.82	Sn 50 118.71	Sb 51 121.76	Te 52 127.60	I 53 126.90	Xe 54 131.29
Cs 55 132.91	Ba 56 137.33		Hf 72 178.49	Ta 73 180.95	W 74 183.84	Re 75 186.21	Os 76 190.23	Ir 77 192.22	Pt 78 195.08	Au 79 196.97	Hg 80 200.59	Tl 81 204.38	Pb 82 207.20	Bi 83 208.98	Po 84 209	At 85 210	Rn 86 222
Fr 87 223	Ra 88 226		Rf 104 267	Db 105 268	Sg 106 271	Bh 107 272	Hs 108 270	Mt 109 276	Ds 110 281	Rg 111 280	Cn 112 285	Nh 113 284	Fl 114 289	Mc 115 288	Lv 116 293	Ts 117 294	Og 118 294

La 57 138.91	Ce 58 140.12	Pr 59 140.91	Nd 60 144.24	Pm 61 145	Sm 62 150.36	Eu 63 151.96	Gd 64 157.25	Tb 65 158.93	Dy 66 162.50	Ho 67 164.93	Er 68 167.26	Tm 69 168.93	Yb 70 173.05	Lu 71 174.97
Ac 89 227	Th 90 232.04	Pa 91 231.04	U 92 238.03	Np 93 237	Pu 94 244	Am 95 243	Cm 96 247	Bk 97 247	Cf 98 251	Es 99 252	Fm 100 257	Md 101 258	No 102 259	Lr 103 262

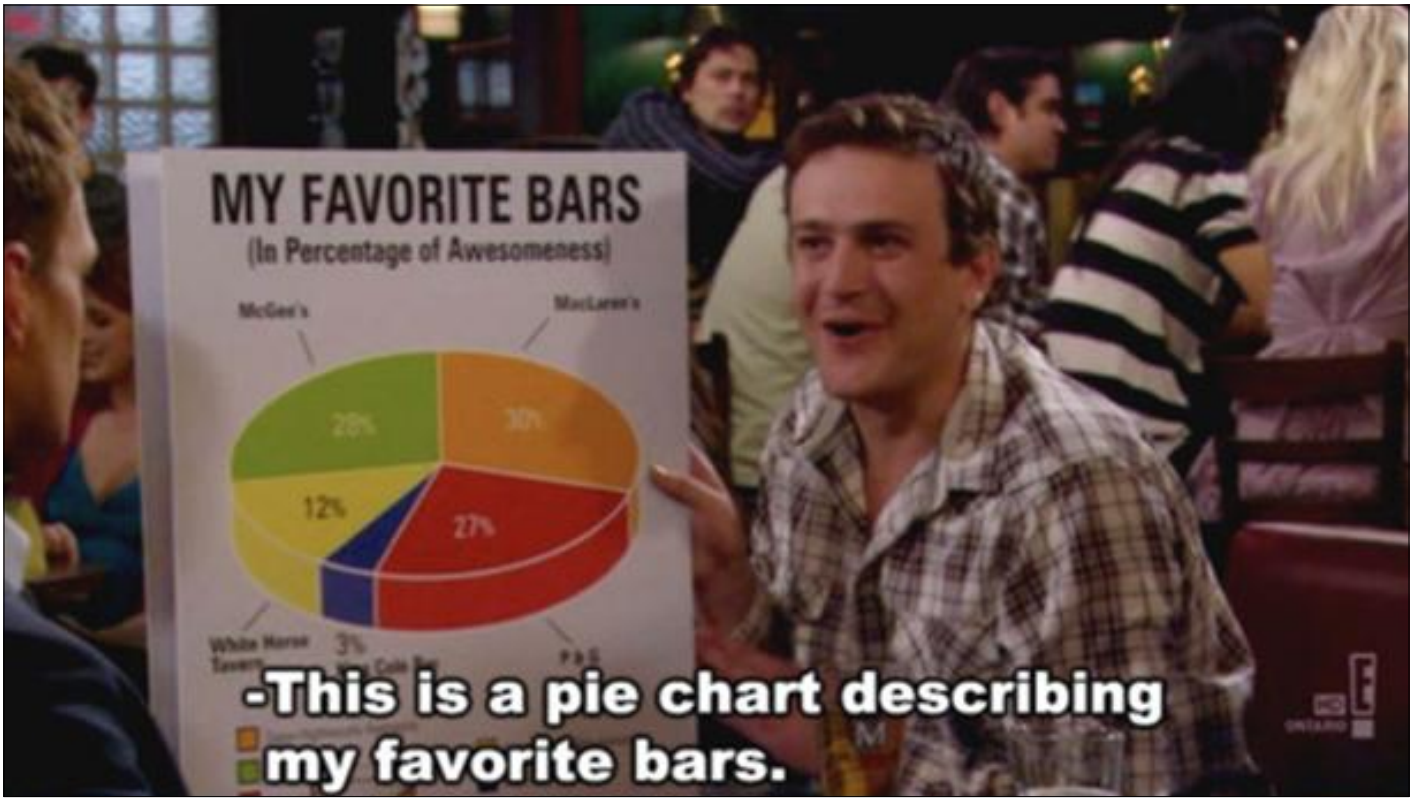
2

52

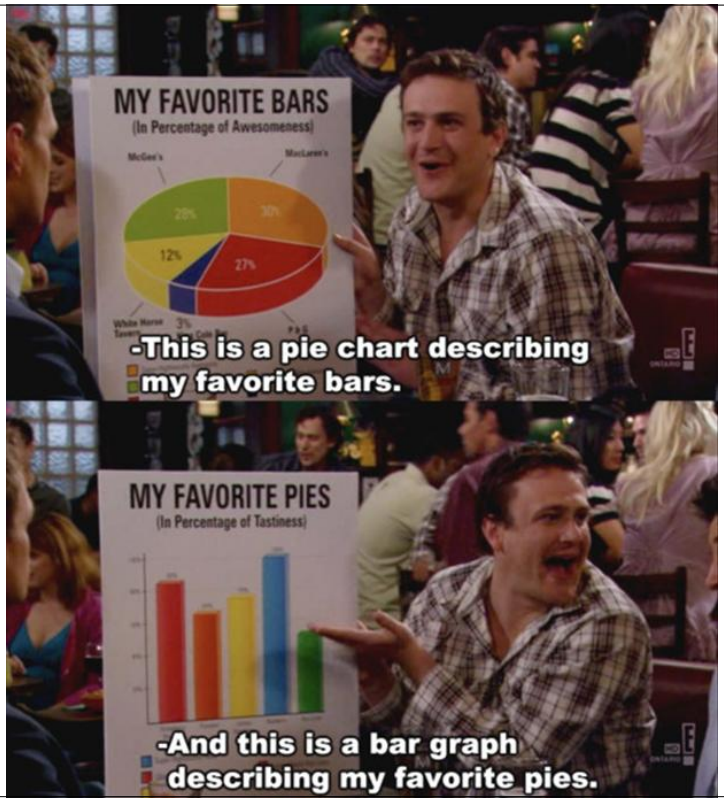

How allegiances shifted from the first to the second round of voting in the French presidential election
First round vote (%)
Second round vote (%)
Macron: 24
Le Pen: 21.3
Fillon: 20
Mélenchon: 19.6
Hamon: 6.4
Macron: 65.8
Le Pen: 34.2
Abstention / leave blank
Harris Interactive, FT analysis
FT

53

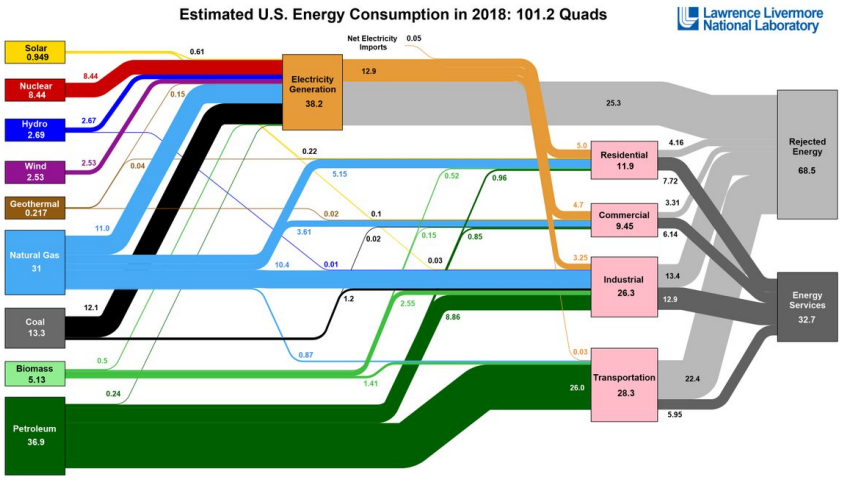
53


MY FAVORITE BARS
(In Percentage of Awesomeness)
-This is a pie chart describing my favorite bars.

54



55



[LLNL, via @carlzimmer]