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 2018.

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

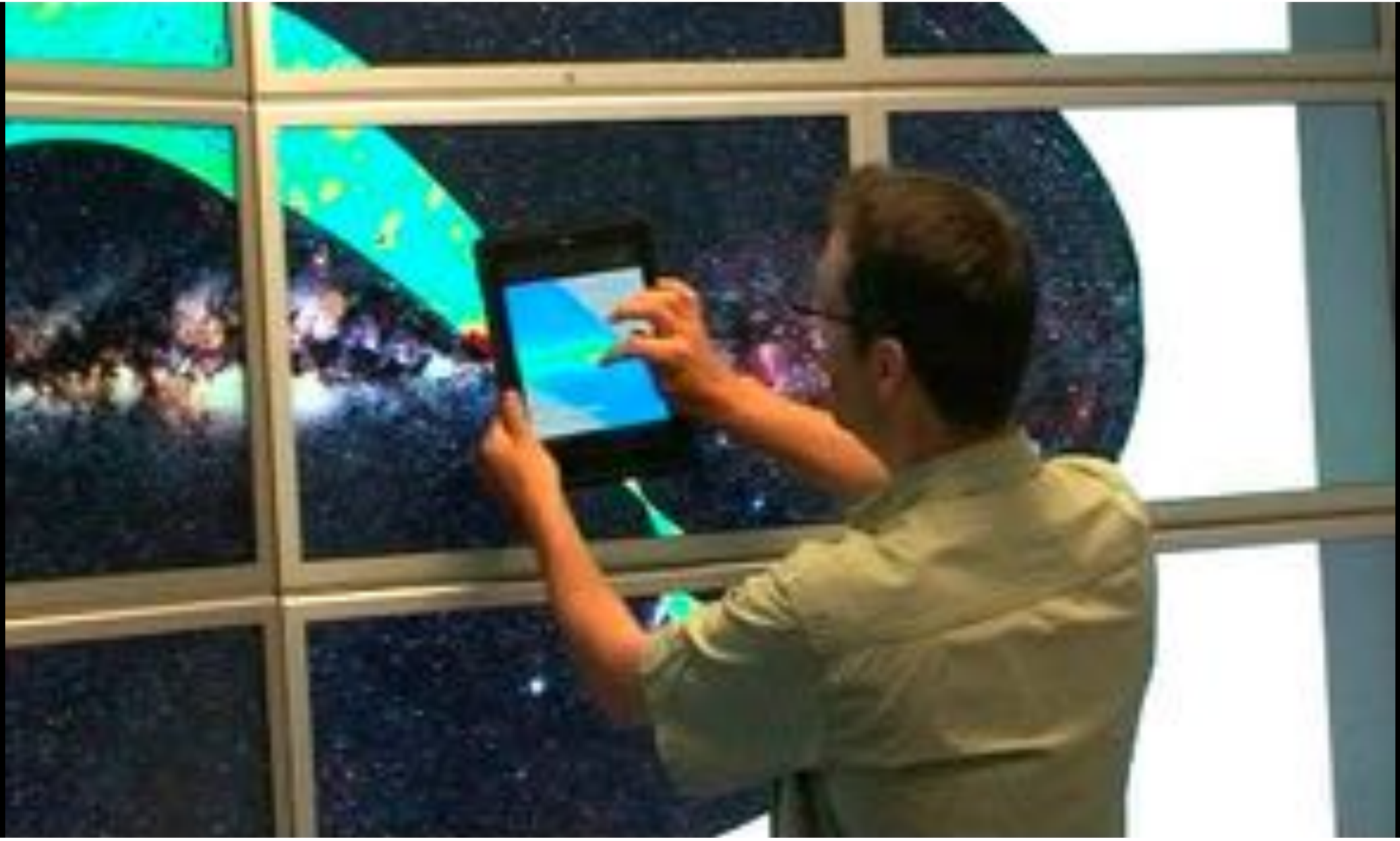
Research

Human-Computer Interaction

Information Visualization

Multi-surface Interaction

3



Class overview

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.

7

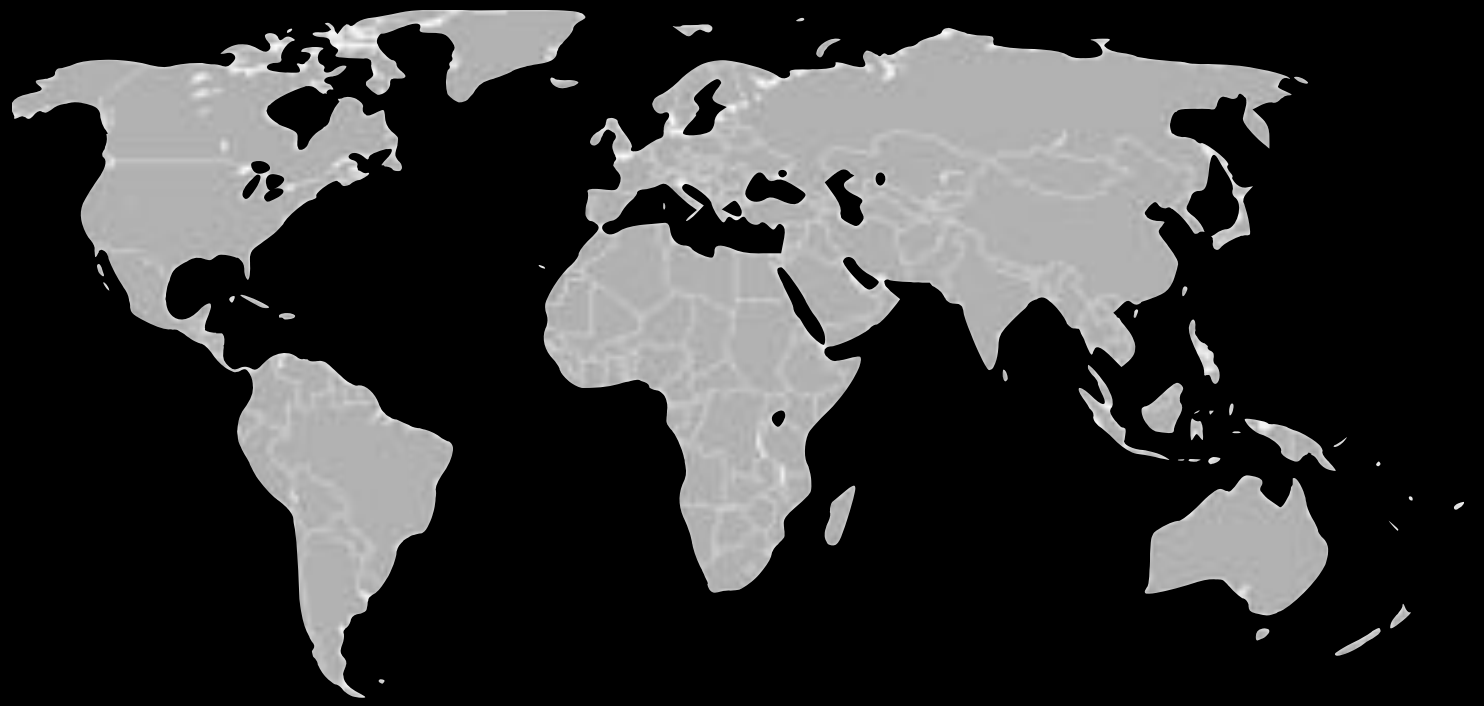
How much data?

- Between 1 and 2 exabytes of unique info produced per year
  - 10000000000000000 (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`

8

2008



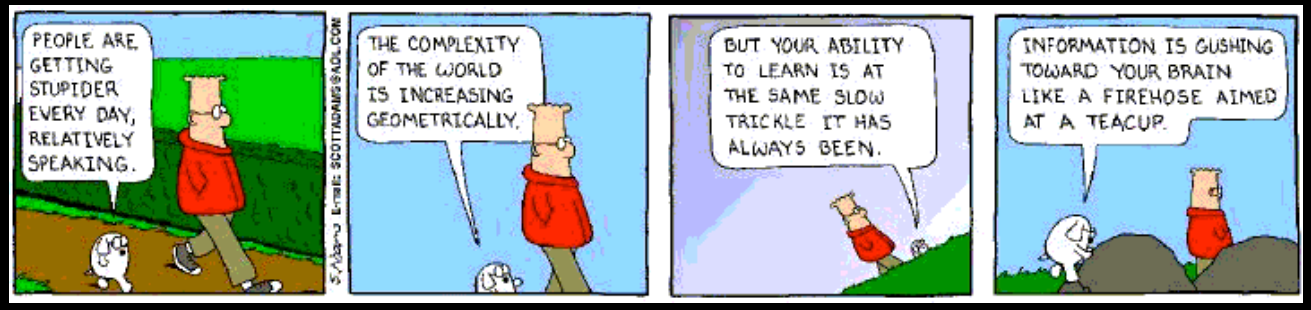
800 exabytes per year

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


2017
900 zetabytes per year
[10 Key Marketing Trends for 2017, IBM, 2017]
10

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

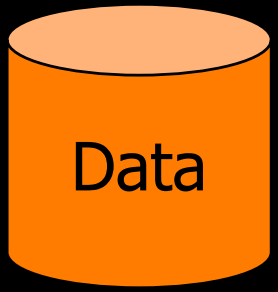
[The McKinsey Quarterly, January 2009]

The Challenge

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

13

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]

15

Example

Questions: Which state has the highest income?
Is there a relationship between income and education?
Are there any outliers?

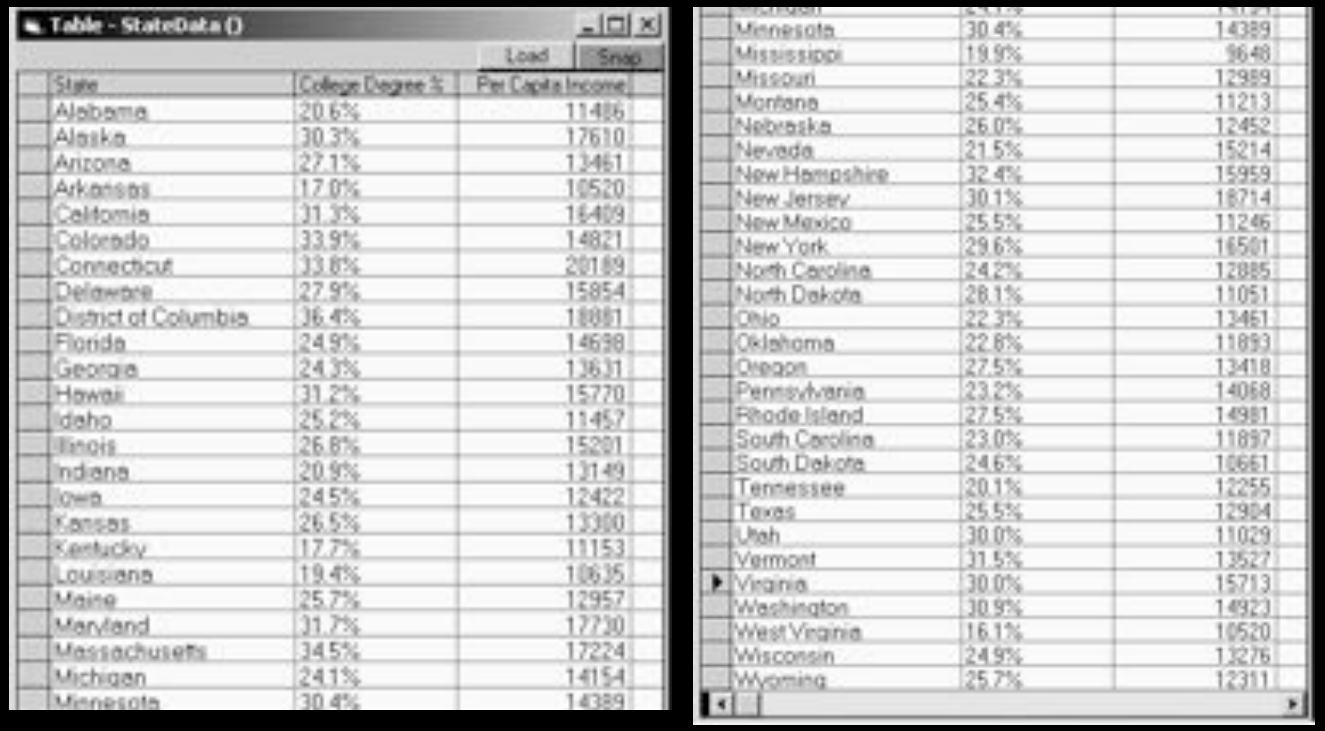


Table - StateData ()

Load | Snap

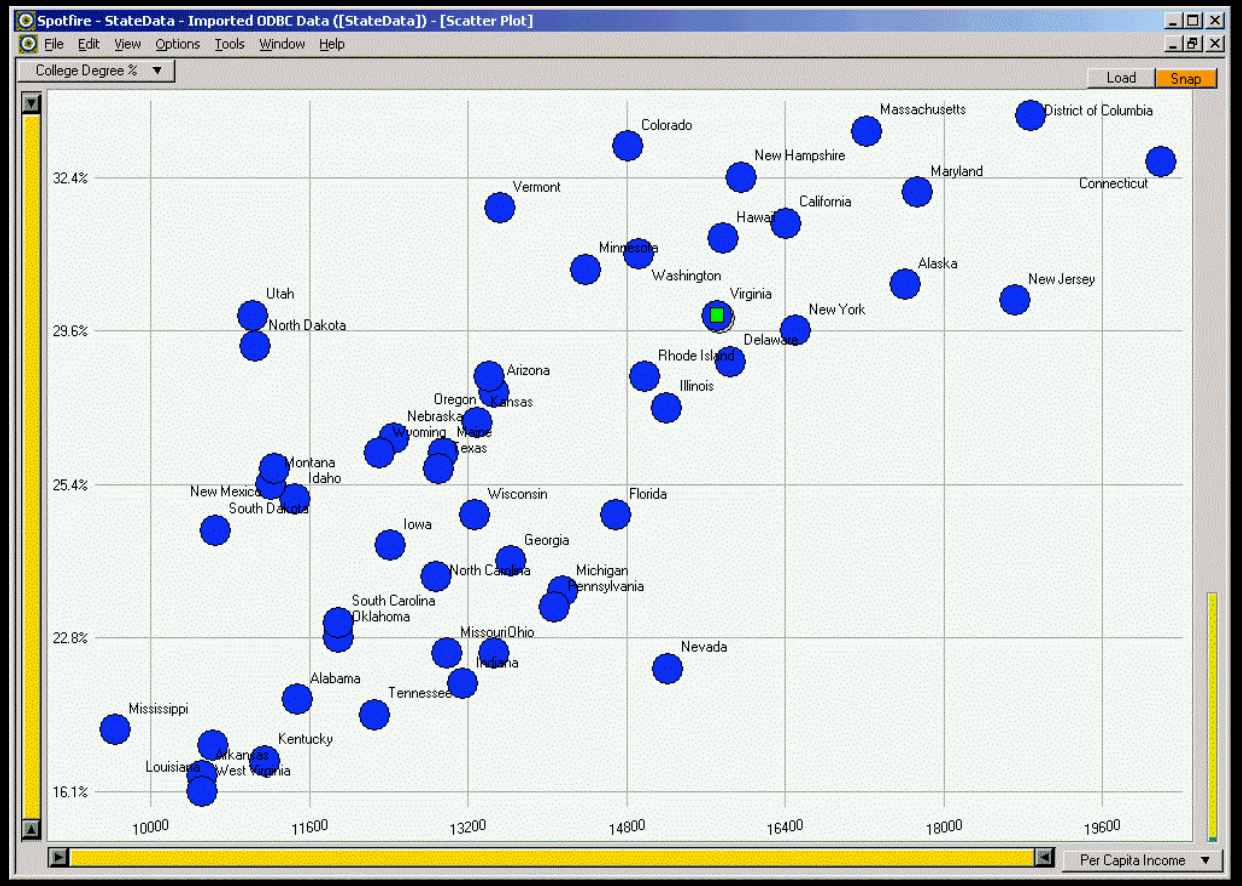
State	College Degree %	Per Capita Income
Alabama	20.6%	11486
Alaska	30.3%	17610
Arizona	27.1%	13461
Arkansas	17.0%	10520
California	31.3%	16409
Colorado	33.9%	14821
Connecticut	33.8%	20189
Delaware	27.9%	15854
District of Columbia	36.4%	18881
Florida	24.9%	14698
Georgia	24.3%	13631
Hawaii	31.2%	15770
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]

16

Visualize the Data

College Degree %



Per Capita Income

[Courtesy of Chris North, Virginia Tech]

17

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?

18

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

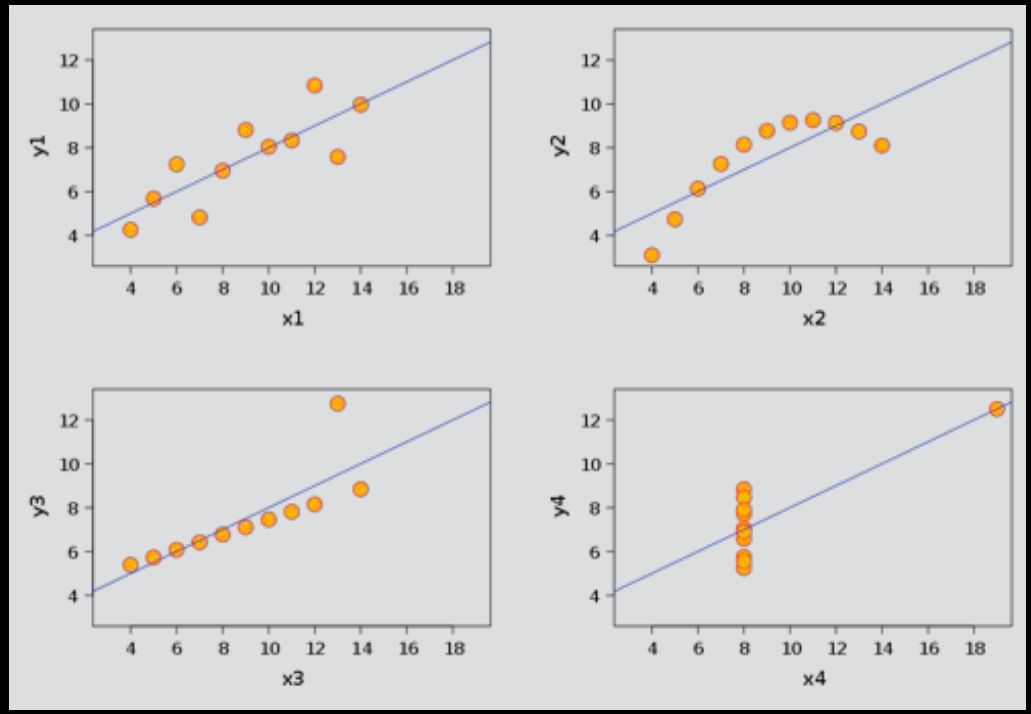
Anscombe's Quartet

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

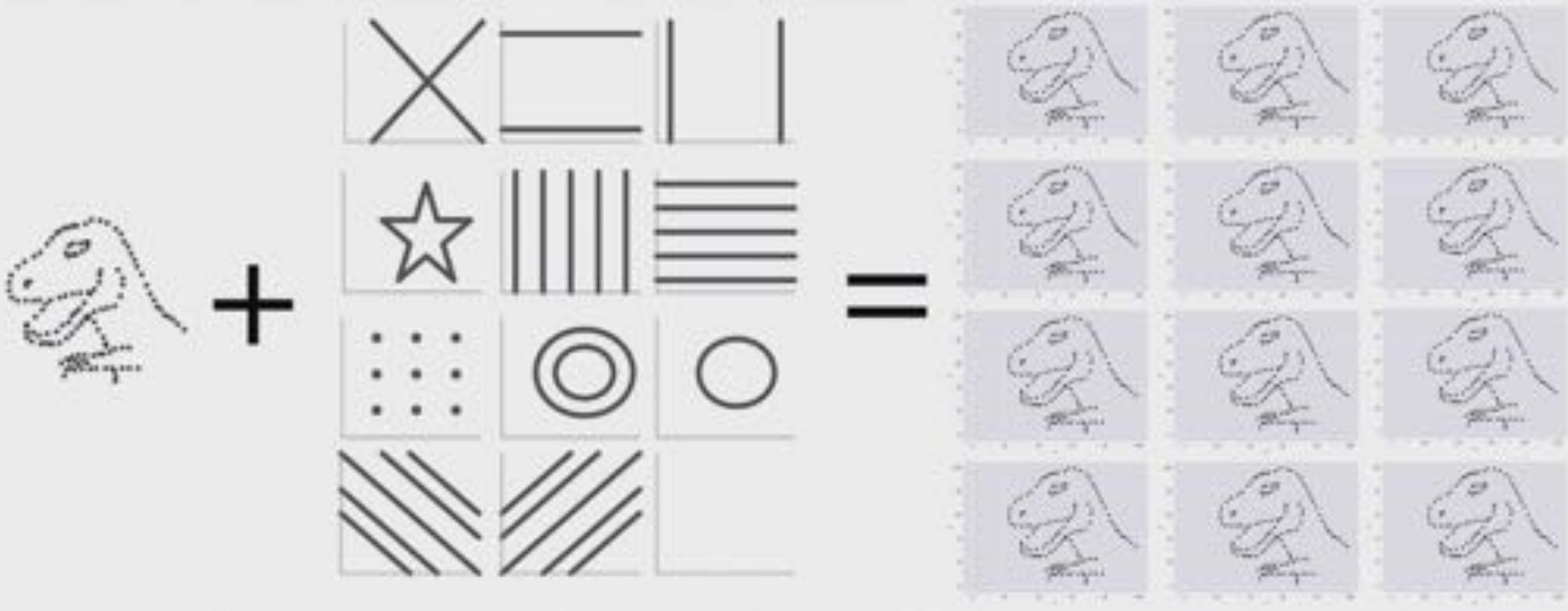
Mean of x	9.0
Variance of x	11.0
Mean of y	7.5
Variance of y	4.12
Correlation between x and y	0.816
Linear regression line	$y = 3 + 0.5x$

Anscombe's Quartet

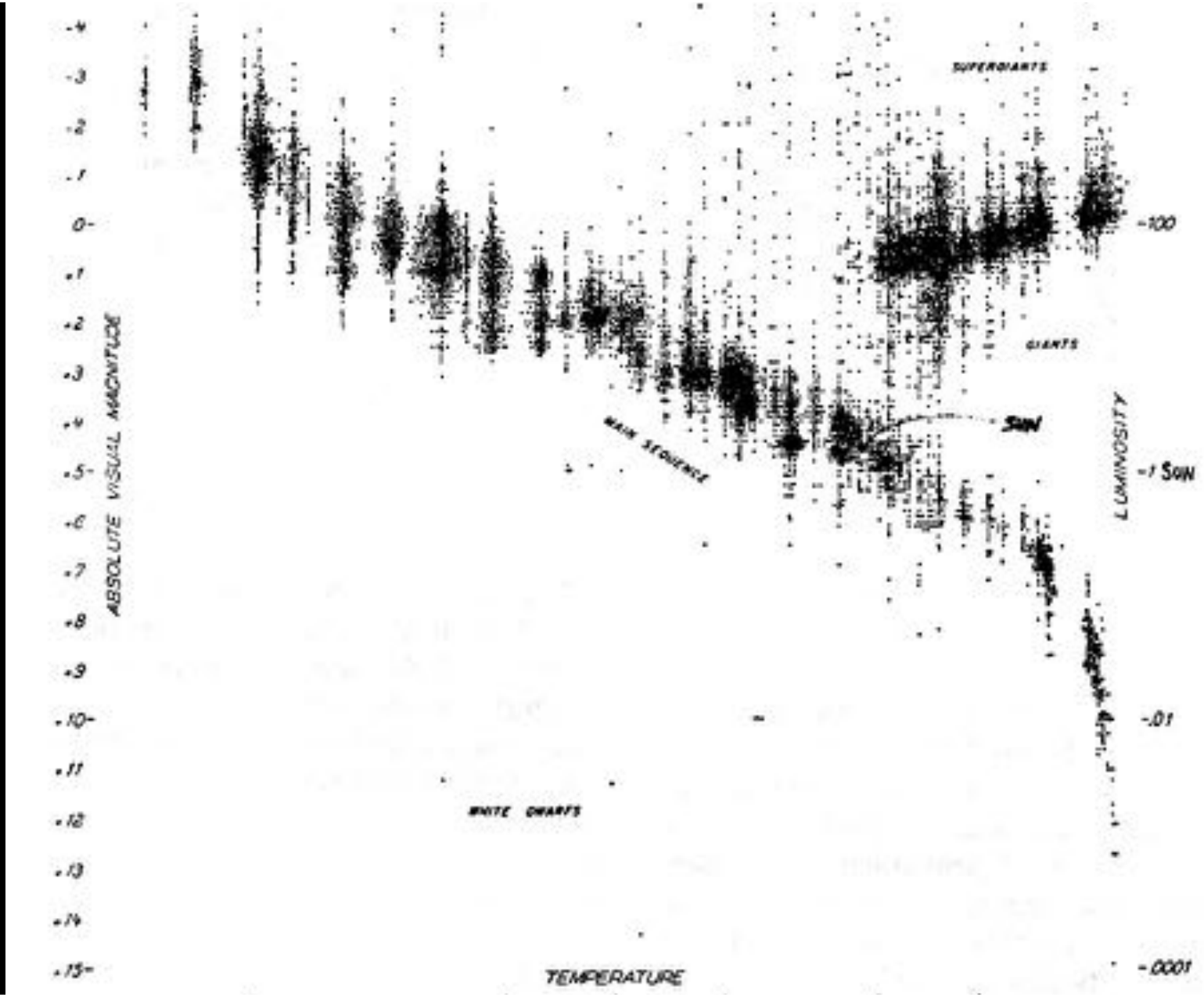
I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	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



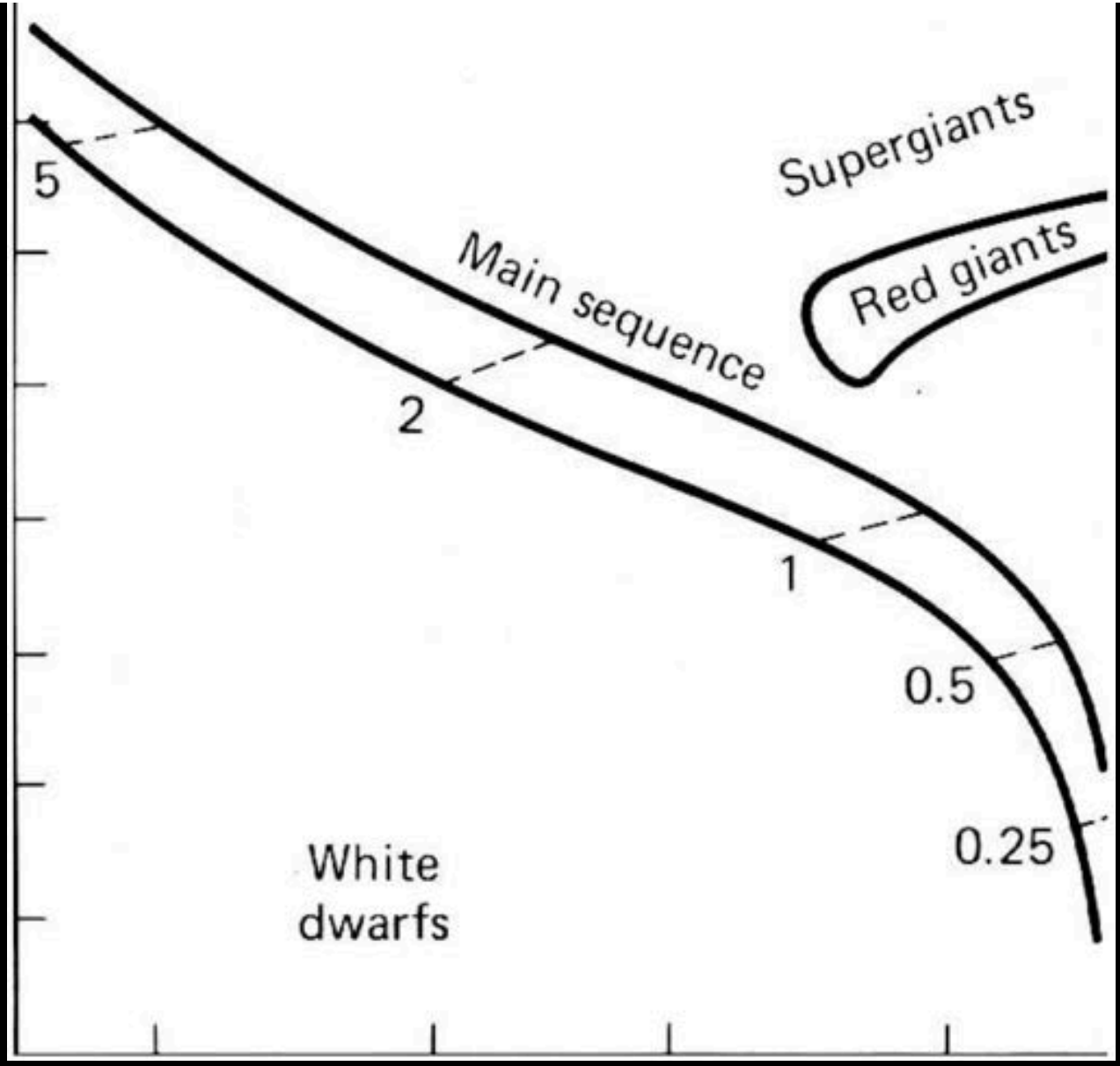
21



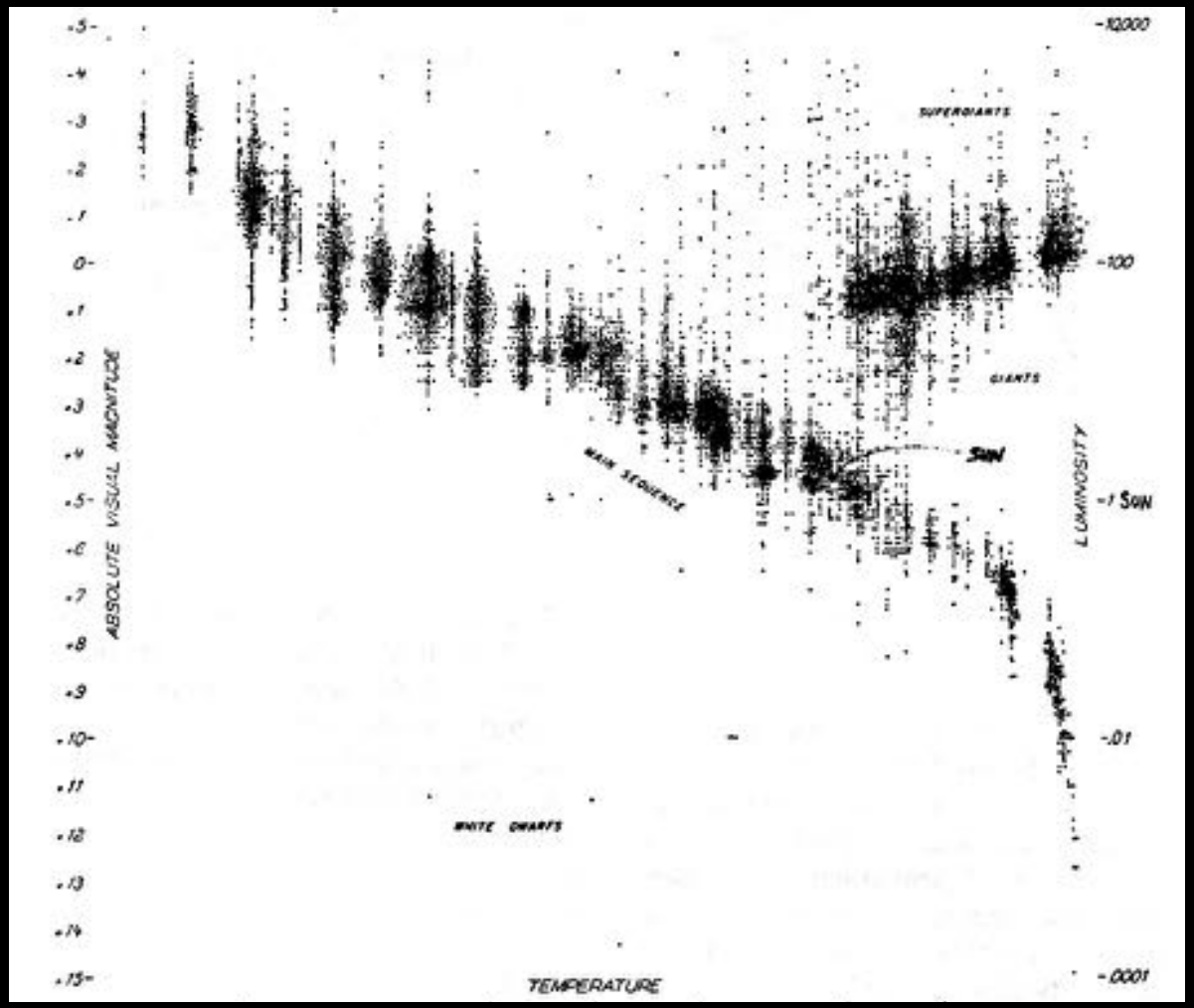
22

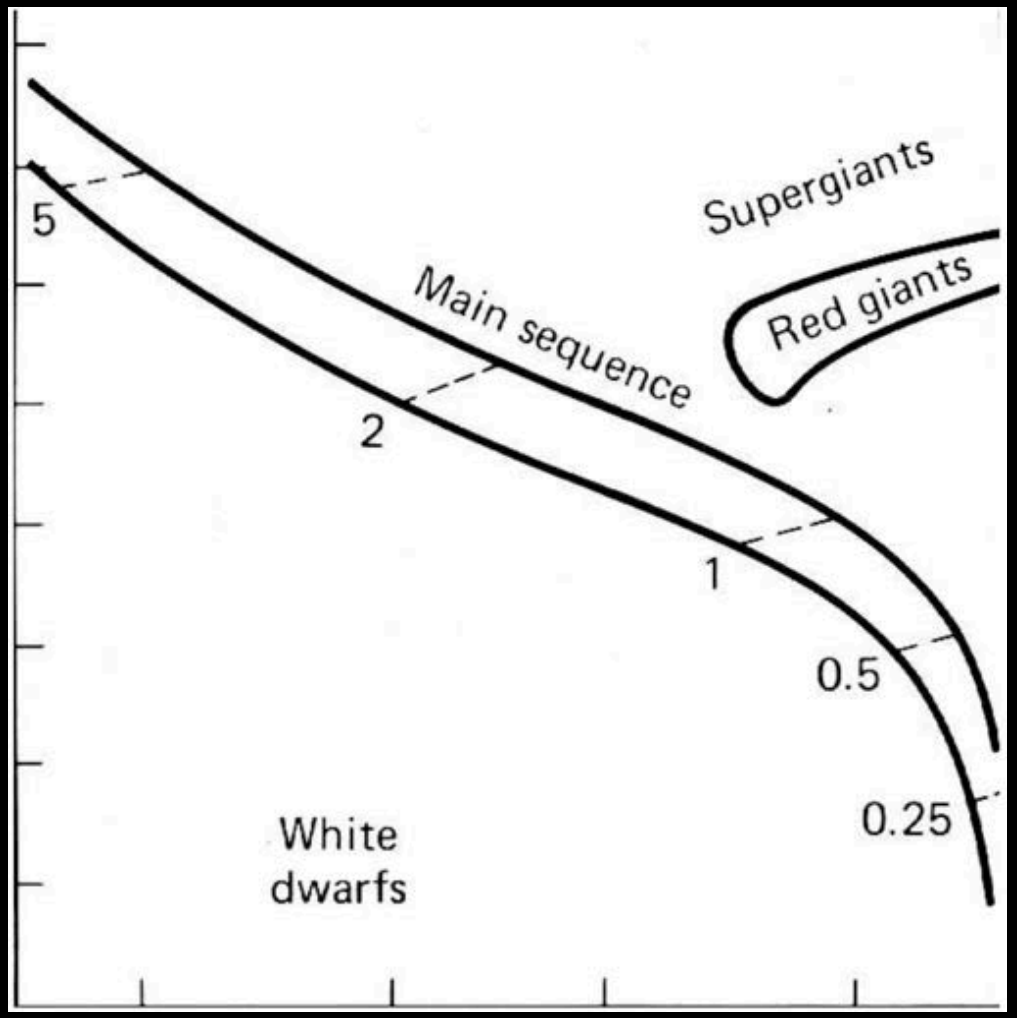

ABSOLUTE VISUAL MAGNITUDE
-4
-3
-2
-1
0
+1
+2
+3
+4
+5
+6
+7
+8
+9
+10
+11
+12
+13
+14
+15
SUPERGIANTS
GIANTS
MAIN SEQUENCE
SUN
WHITE DWARFS
TEMPERATURE
LUMINOSITY
100
1 SUN
.01
.0001

23


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

24


ABSOLUTE VISUAL MAGNITUDE
LUMINOSITY
TEMPERATURE
SUPERGIANTS
GIANTS
MAIN SEQUENCE
SUN
WHITE DWARFS
-10000
-100
-1 SUN
-.01
-.0001


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

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]

29

Information Visualization

- What is “visualization”?
  - The use of computer-supported, interactive visual representations of data to amplify cognition.
  - From [Card, Mackinlay, Shneiderman ’98]

30

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

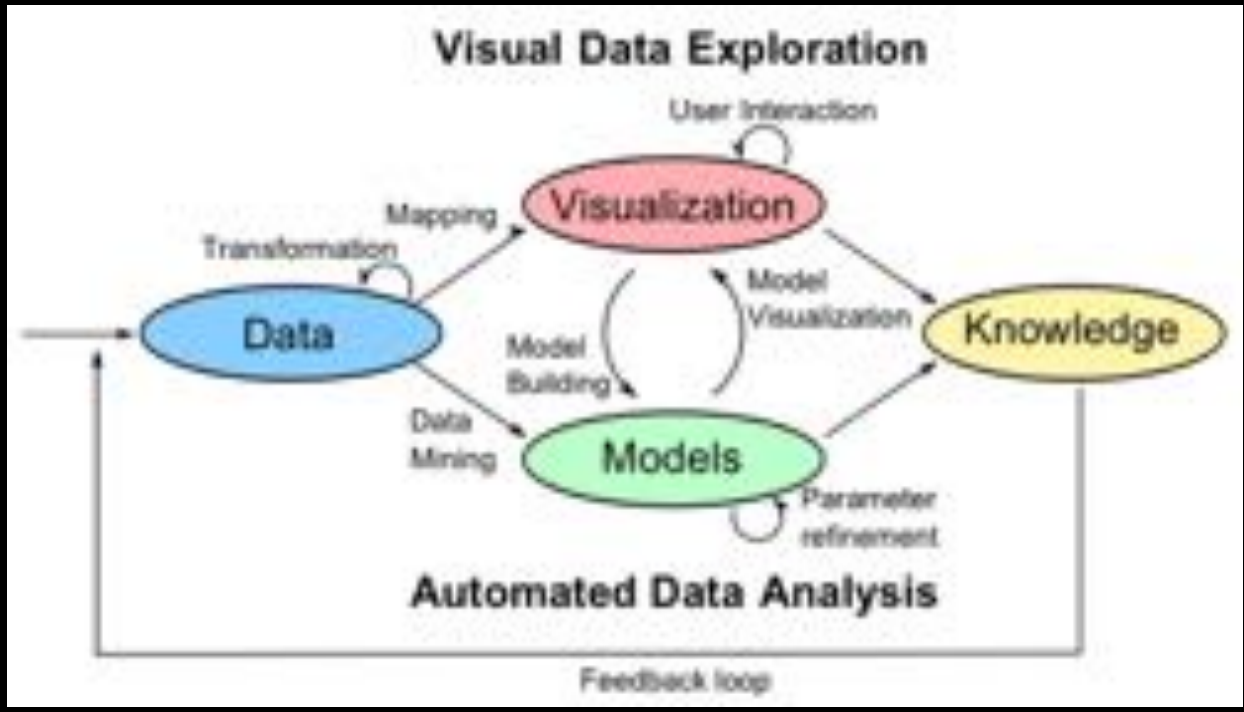
Information Visualization

- 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

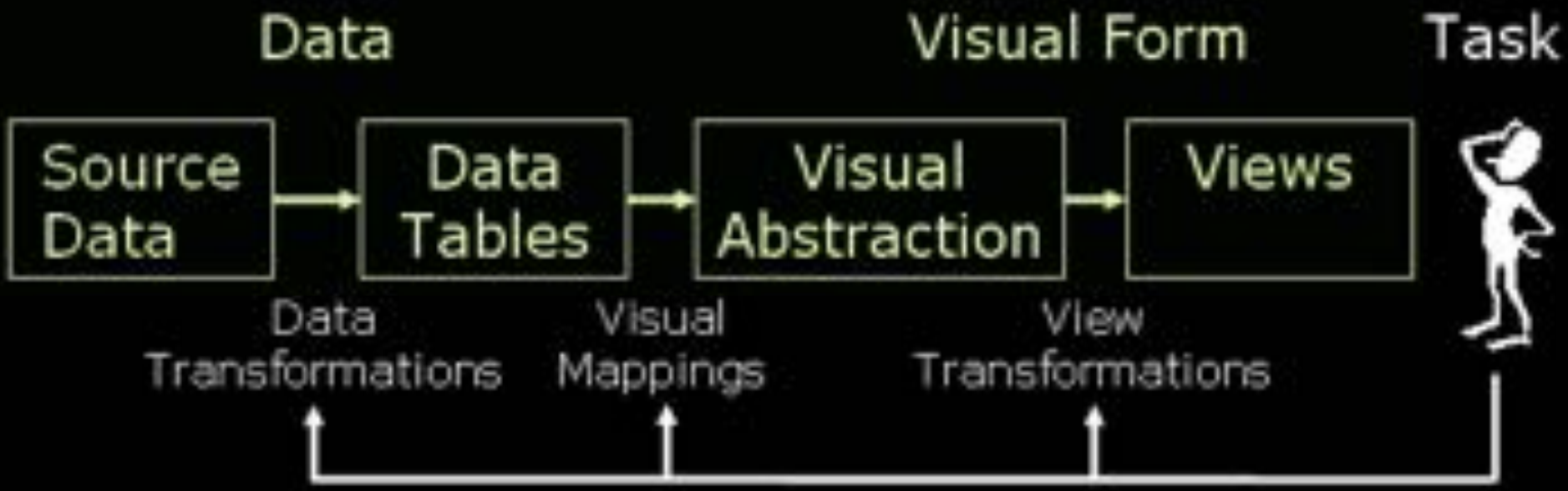
- **Interactivity**
  - Want to show multiple different perspectives on the data

35

Data → Visualisation

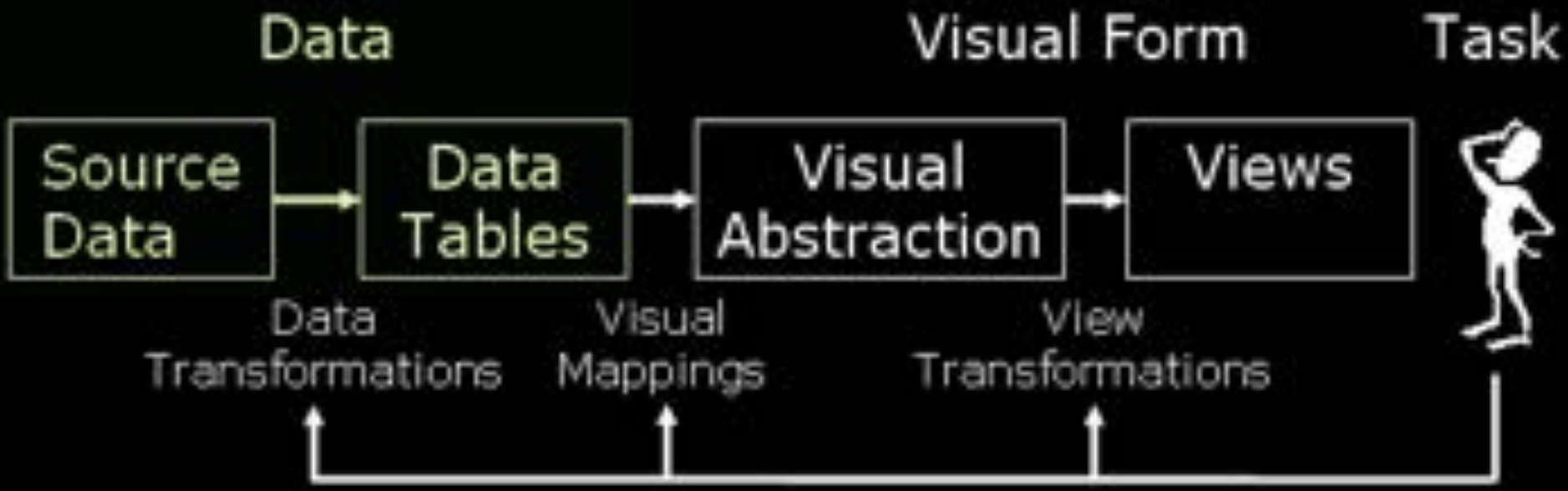
36

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)?

39

Example

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

40

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

How do we show the data?

44

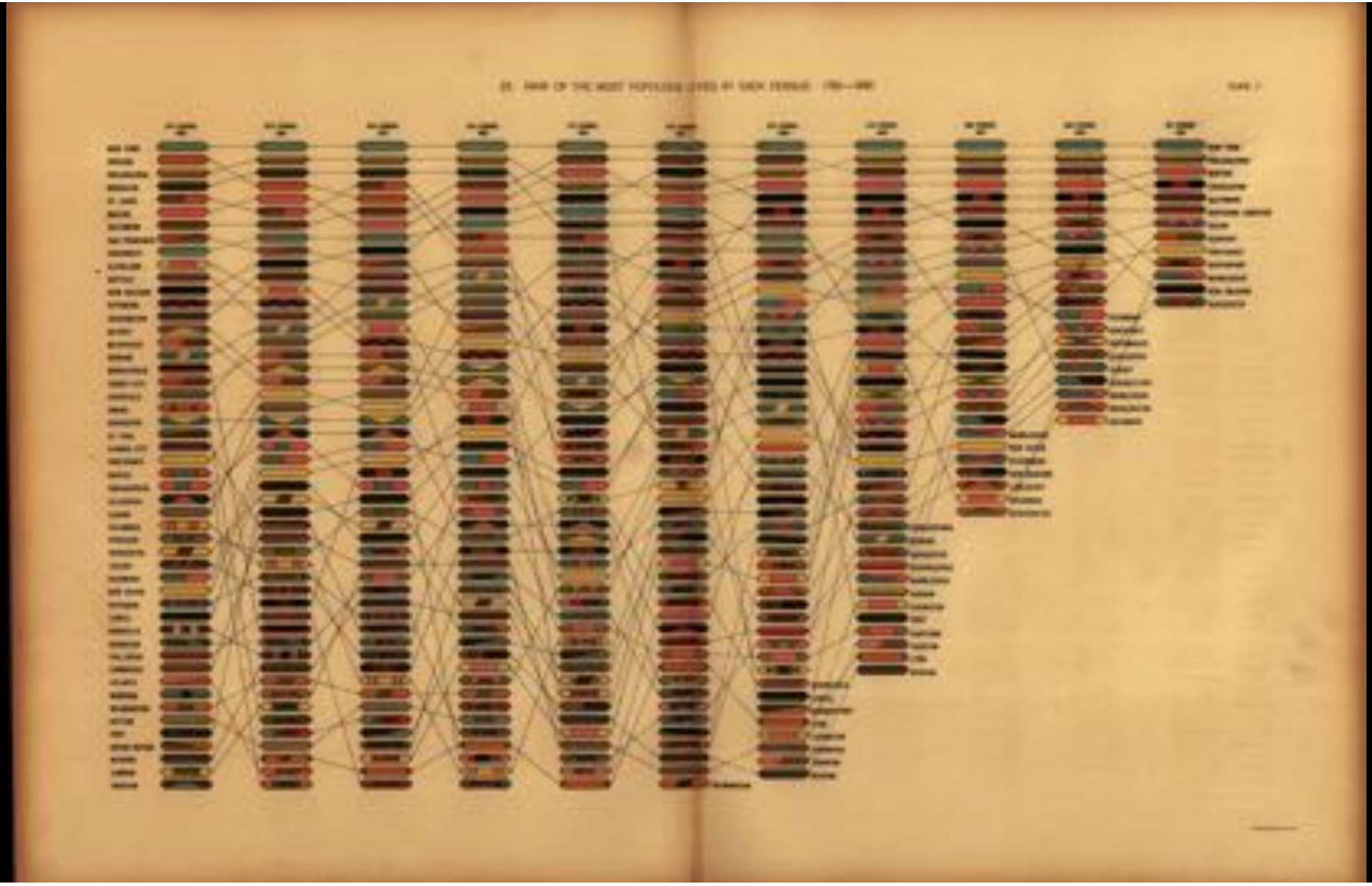
À suivre…

45

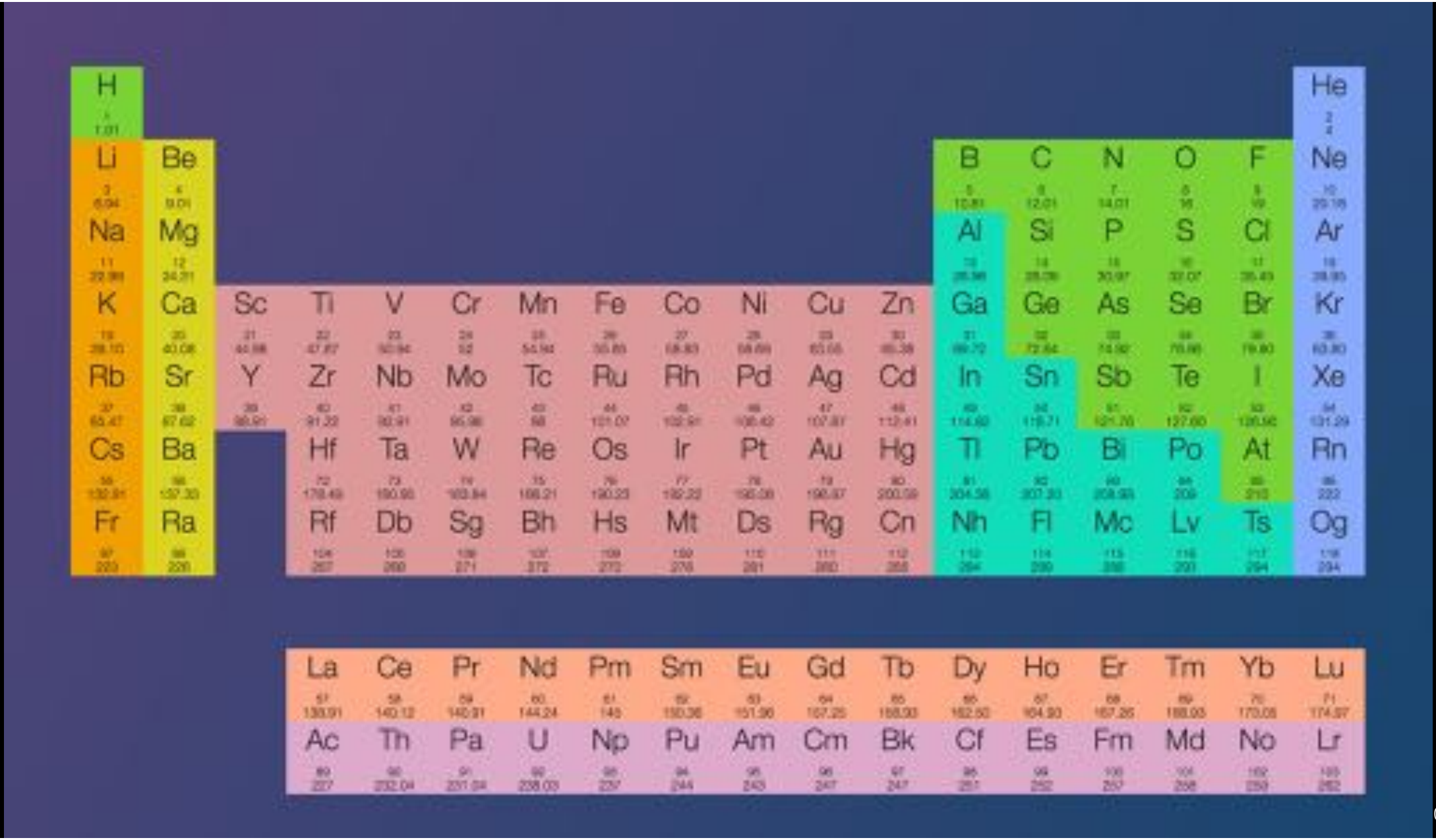
Teasers

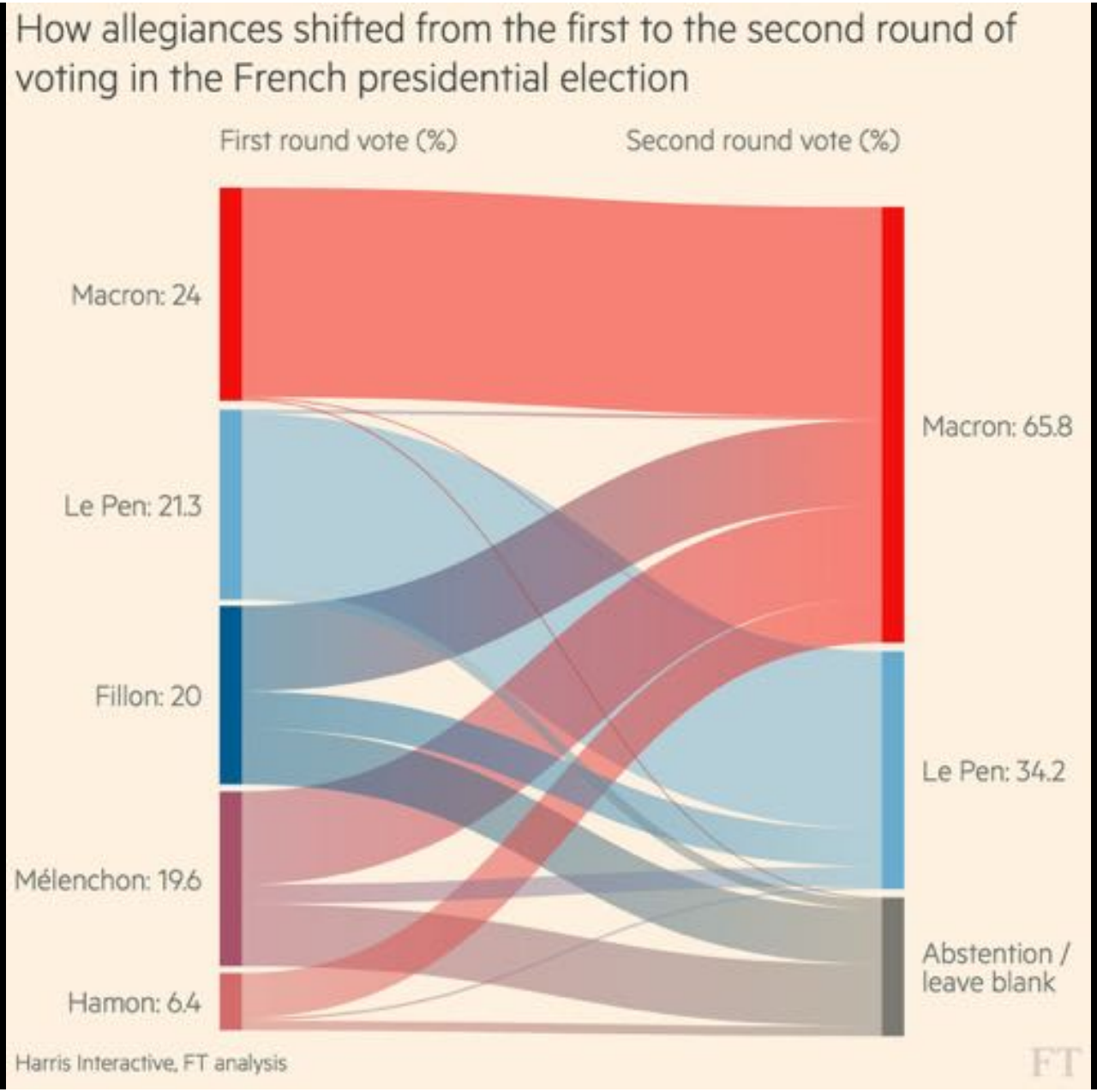
46


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

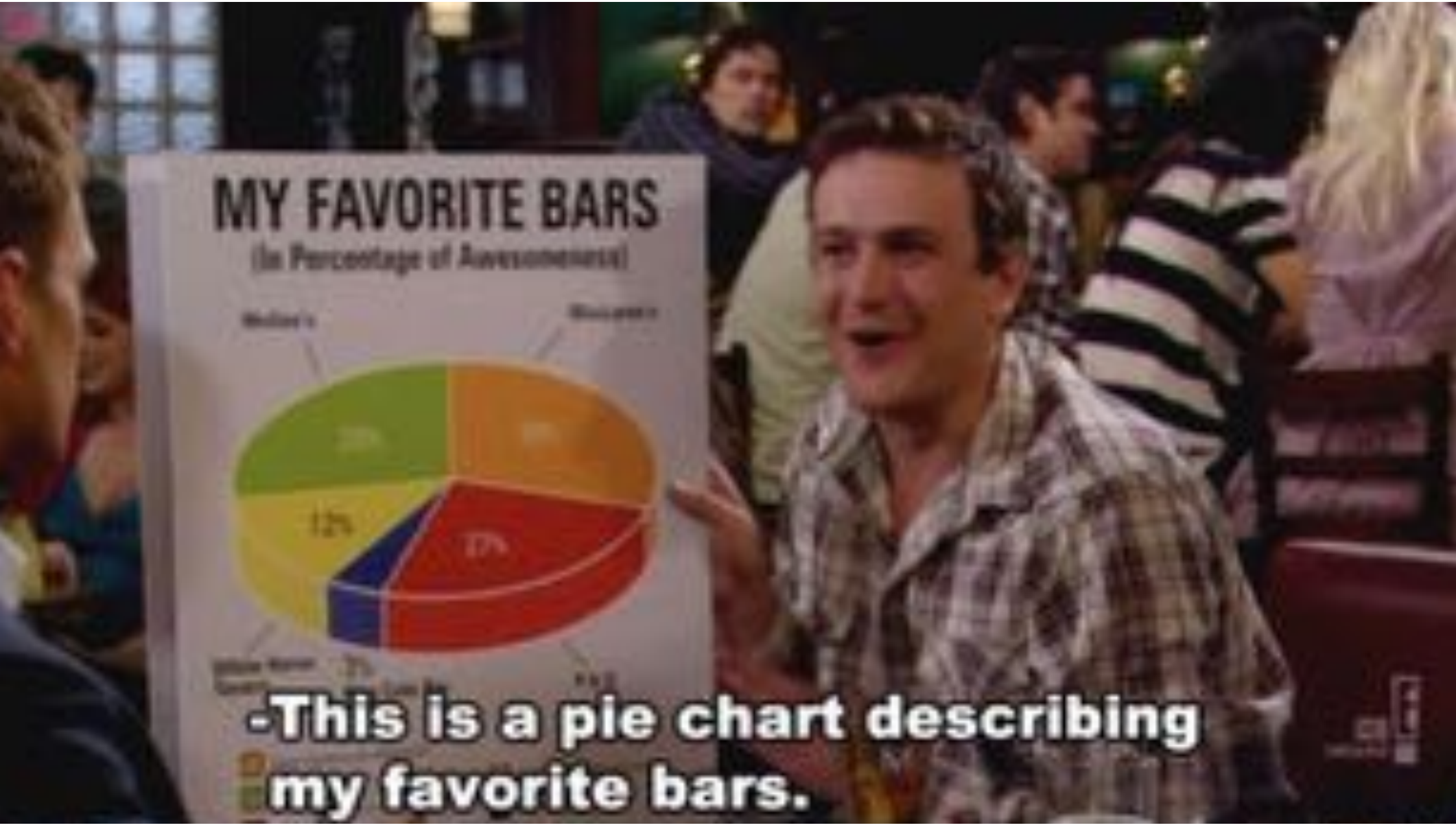


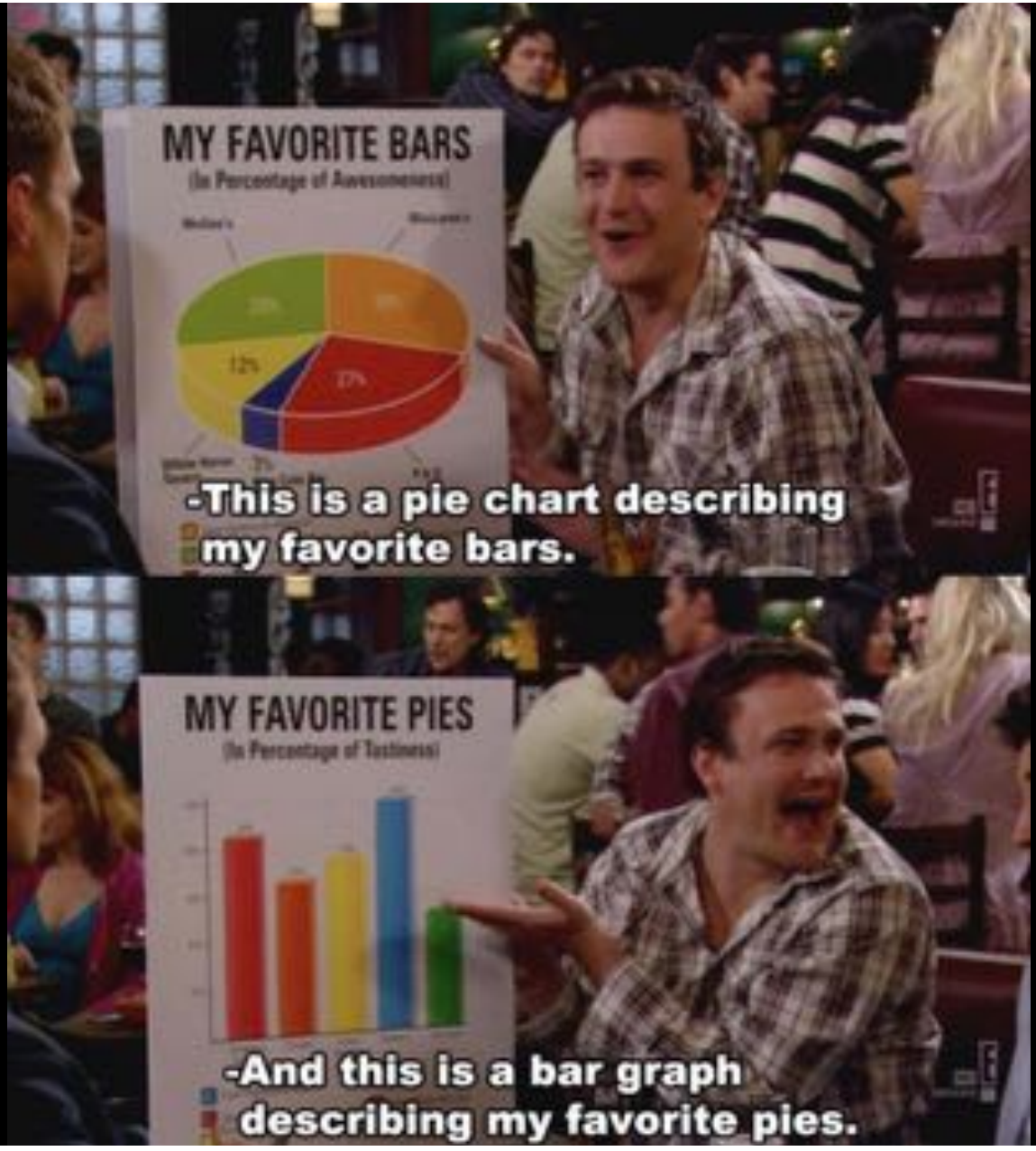
48


H
Li Be
Na Mg
K Ca Sc Ti V Cr Mn Fe Co Ni Cu Zn
Rb Sr Y Zr Nb Mo Tc Ru Rh Pd Ag Cd
Cs Ba Hf Ta W Re Os Ir Pt Au Hg
Fr Ra Rf Db Sg Bh Hs Mt Ds Rg Cn
B C N O F He
Al Si P S Cl Ne
Ga Ge As Se Br Ar
In Sn Sb Te I Kr
Tl Pb Bi Po At Xe
Nh Fl Mc Lv Ts Rn
Og
La Ce Pr Nd Pm Sm Eu Gd Tb Dy Ho Er Tm Yb Lu
Ac Th Pa U Np Pu Am Cm Bk Cf Es Fm Md No Lr


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

51


MY FAVORITE BARS
-This is a pie chart describing
my favorite bars.


MY FAVORITE BARS
-This is a pie chart describing my favorite bars.
MY FAVORITE PIES
-And this is a bar graph describing my favorite pies.

53