CSC2537 / STA2555

INFORMATION VISUALISATION

Fanny CHEVALIER

UNIVERSITY OF TORONTO

ABOUT ME...

Assistant Professor at University of Toronto
Formerly Research Scientist at Inria

Work in:

- HCI
- Data Visualization
- Visual Analytics

http://fannychevalier.net

COURSE OBJECTIVES

After following this course, you will be able to:

- **know** the scientific foundation of Infovis;
- **analyze** data sets using visualization techniques;
- **build** visualizations that convey information and ideas.

http://www.cs.toronto.edu/~csc2537h

ASSIGNMENTS

- **Scientific research paper presentation** — 40%
- **Project** — 60%

http://www.cs.toronto.edu/~csc2537h

GUEST SPEAKERS

Isabel MEREILLES

OCAD UNIVERSITY

Justin MATEJKA

AUTODESK RESEARCH

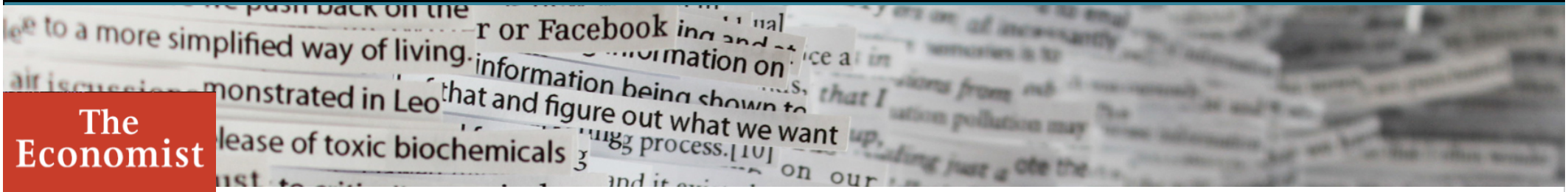
SCHEDULE

9 JAN.	WELCOME - INTRODUCTION / PROSPECTIVE PROJECTS
16 JAN.	VISUAL PERCEPTION & DATA MODELS
23 JAN.	GUEST SPEAKER : JUSTIN MATEJKA
30 JAN.	EXPLORATORY DATA ANALYSIS + ***STUDENTS PAPER PRESENTATIONS***
6 FEB.	GRAPHS & NETWORKS + ***STUDENTS PAPER PRESENTATIONS***
13 FEB.	**PROJECT: MID-TERM REVIEW** 
20 FEB.	READING WEEK
27 FEB.	GUEST SPEAKER : ISABEL MEREILLES
6 MAR.	INTERACTION & ANIMATION + ***STUDENTS PAPER PRESENTATIONS***
13 MAR.	***STUDENTS PAPER PRESENTATIONS***
20 MAR.	***STUDENTS PAPER PRESENTATIONS***
27 MAR.	***STUDENTS PAPER PRESENTATIONS***
3 APR.	**PROJECT: FINAL PRESENTATIONS** + WRAP UP 

http://www.cs.toronto.edu/~csc2537h

INTRODUCTION

WHY VISUALIZATION?



A special report on managing information | February 27th 2010

Special Report | Data, data everywhere

Information has gone from scarce to superabundant. That brings huge new benefits, says Kenneth Cukier (interviewed here)—but also big headaches

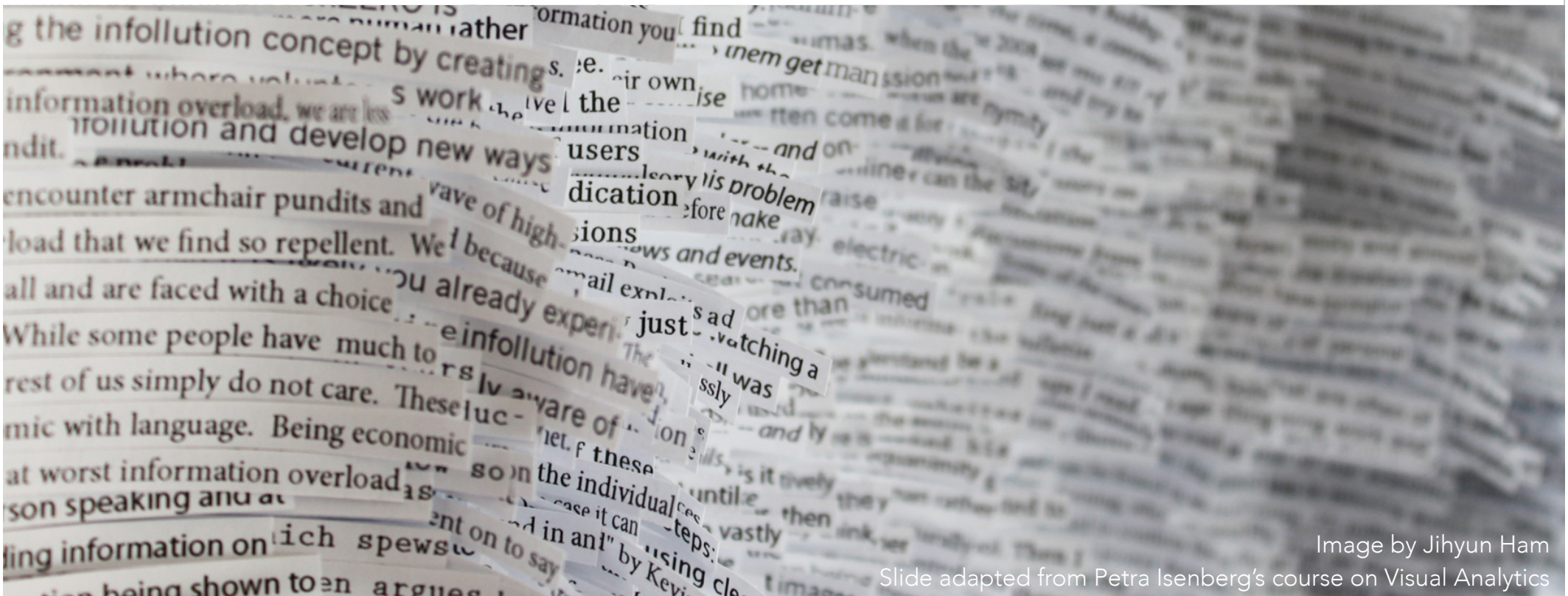
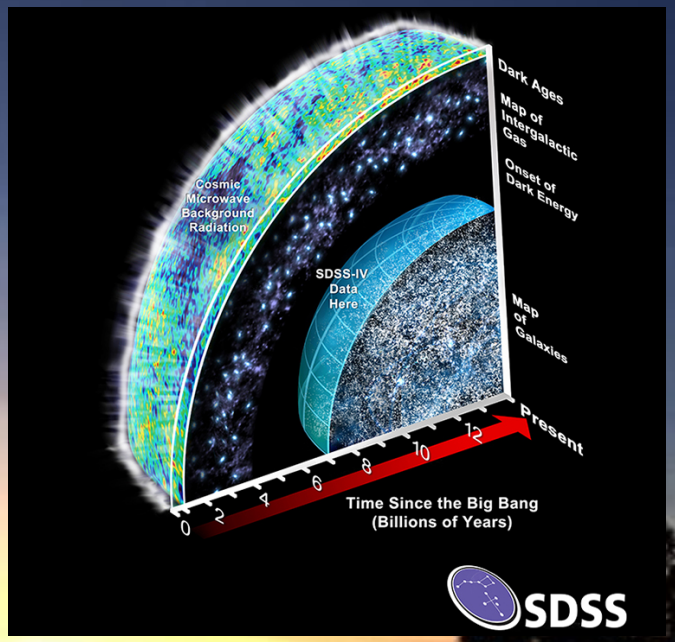


Image by Jihyun Ham
Slide adapted from Petra Isenberg's course on Visual Analytics

SLOAN DIGITAL SKY SURVEY

- started in 2000 http://www.sdss.org/
- in first weeks, collected more data than entire history of astronomy before



Images: http://www.sdss.org/press-releases/the-sdss-expands-its-reach/

Slide adapted from Petra Isenberg's course on Visual Analytics

WALMART

- 1 million customer transactions per hour
- likely has information on >145 million Americans [1]

Image: http://saultonline.com/2016/06/walmart-canada-to-stop-accepting-visa/

[1] http://centerformediajustice.org/wp-content/uploads/2014/06/WALMART_PRIVACY_.pdf

Slide adapted from Petra Isenberg's course on Visual Analytics

AND MUCH MORE...

- Youtube users upload more than 100 hours of new video every minute
  https://youtube.googleblog.com/2013/05/heres-to-eight-great-years.html
- Facebook has currently on average 1.13 billion active users daily
  http://newsroom.fb.com/company-info/
- the Library of Congress adds 12,000 items to their collection every day
  https://www.loc.gov/about/fascinating-facts/

Slide adapted from Petra Isenberg's course on Visual Analytics

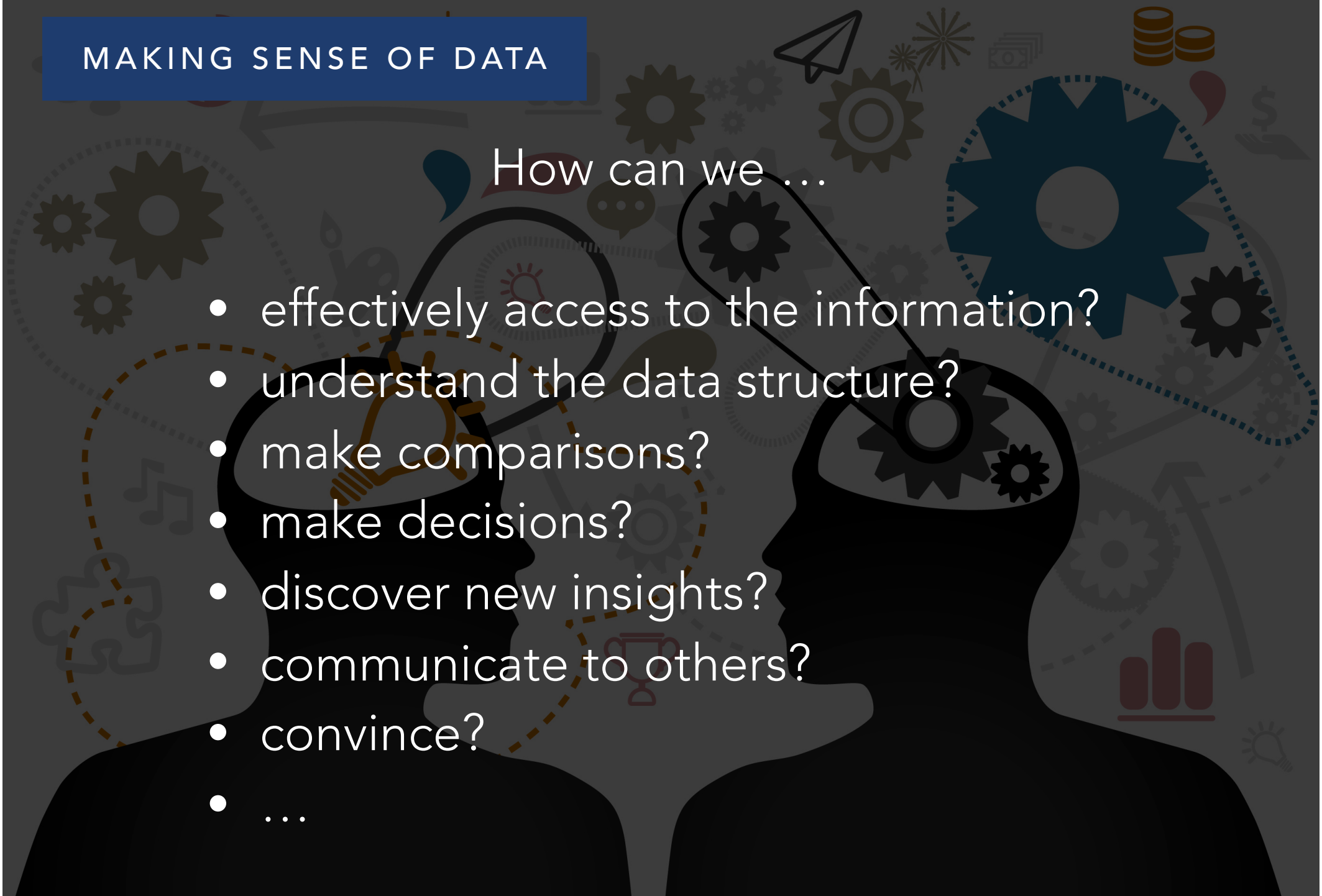
CHALLENGES

- data != useful information
- you want insights

Analysis is needed

Slide adapted from Petra Isenberg's course on Visual Analytics

MAKING SENSE OF DATA



How can we …

- effectively access to the information?
- understand the data structure?
- make comparisons?
- make decisions?
- discover new insights?
- communicate to others?
- convince?
- …

Anascombe'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

Source: Anascombe's Quartet, Wikipedia

STATISTICAL ANALYSIS

suggests that all datasets are equivalent w.r.t. some metrics

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

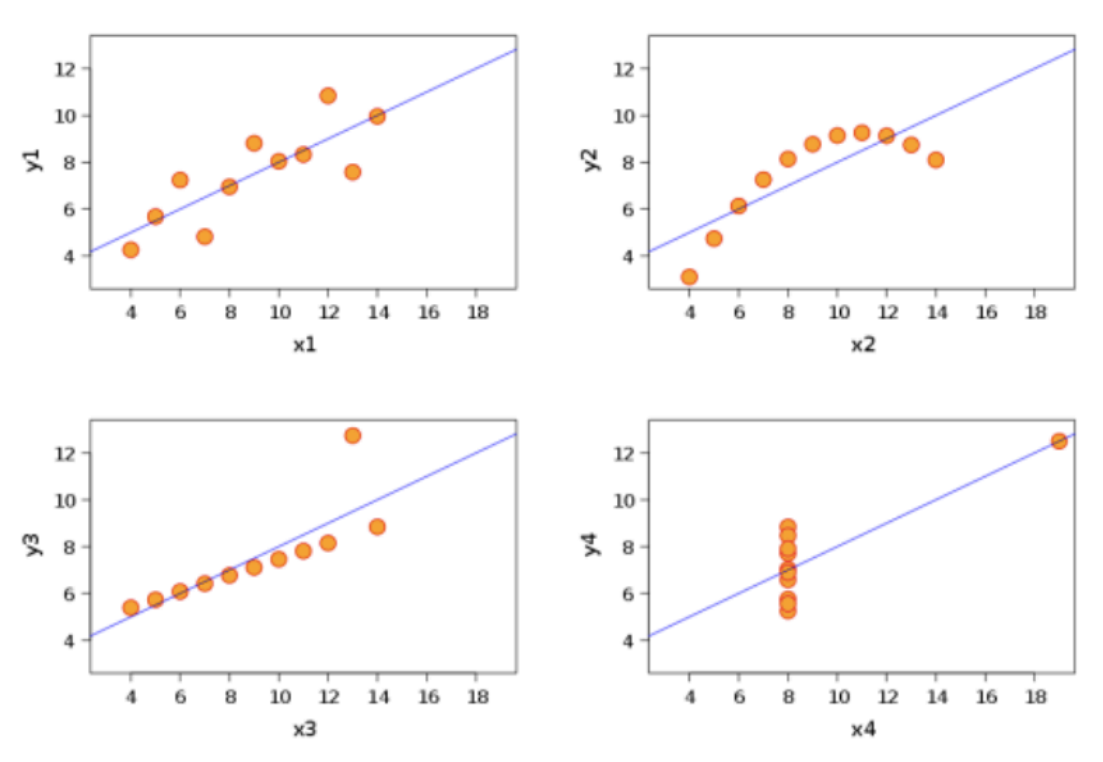
Mean of x	9
Sample variance of x	11
Mean of y	7.50
Sample variance of y	4.12
Correlation between x and y	0.816
Linear regression line	y = 3.00 + 0.500x

Source: Anascombe's Quartet, Wikipedia

VISUALIZATION

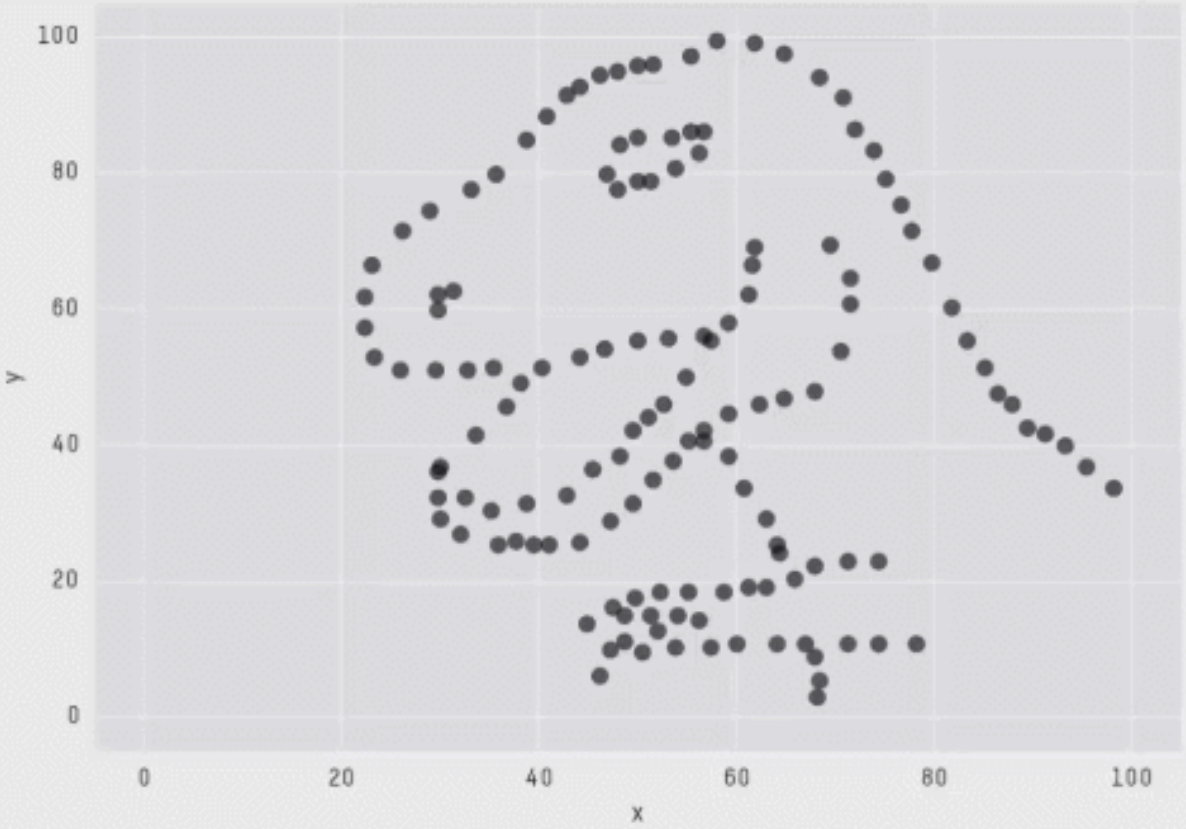
the visual representations tell a complete different story…

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



Source: Anascombe's Quartet, Wikipedia

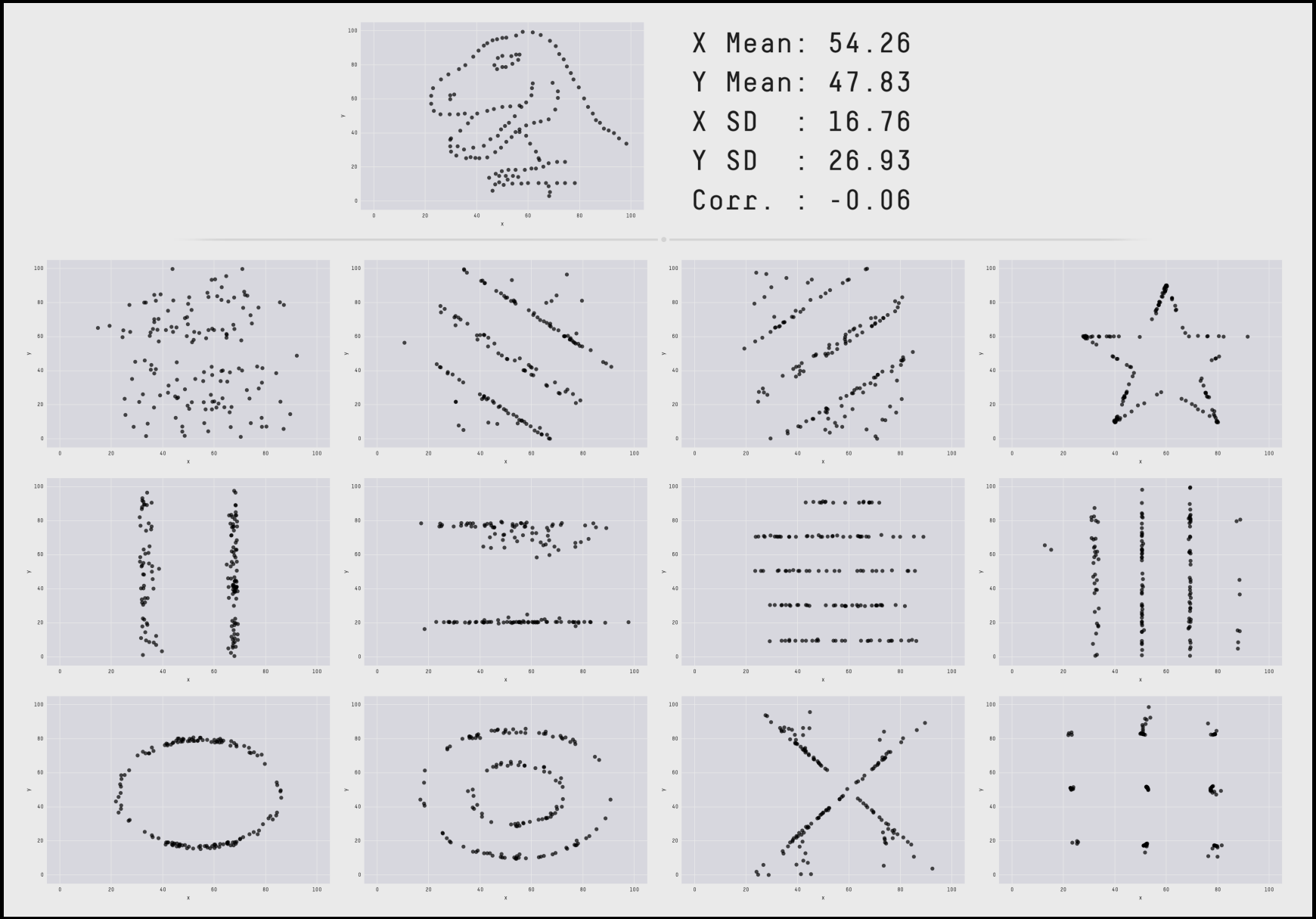
DATASAURUS DOZEN



X Mean: 54.2659224
Y Mean: 47.8313999
X SD : 16.7649829
Y SD : 26.9342120
Corr. : -0.0642526

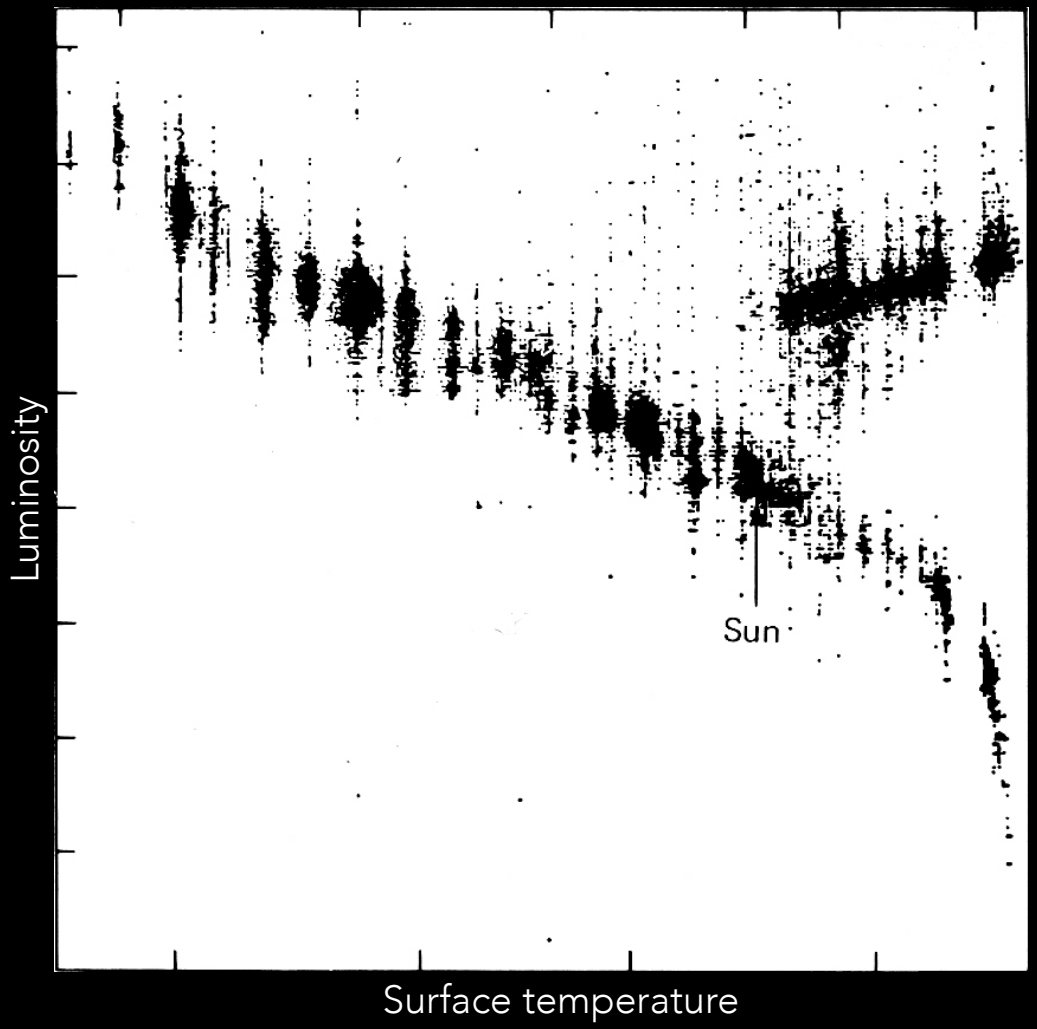
Source: J. Matejka and G. Fitzmaurice. Same Stats, Different Graphs, CHI 2017

DATASAURUS DOZEN



Source: J. Matejka and G. Fitzmaurice. Same Stats, Different Graphs, CHI 2017

AUTOMATIC ABSTRACTION CAPABILITY



Hertzsprung Russell Diagram and its interpretation

[Source: J.D. Fekete et al. "The Value of Information Visualization", 2008]

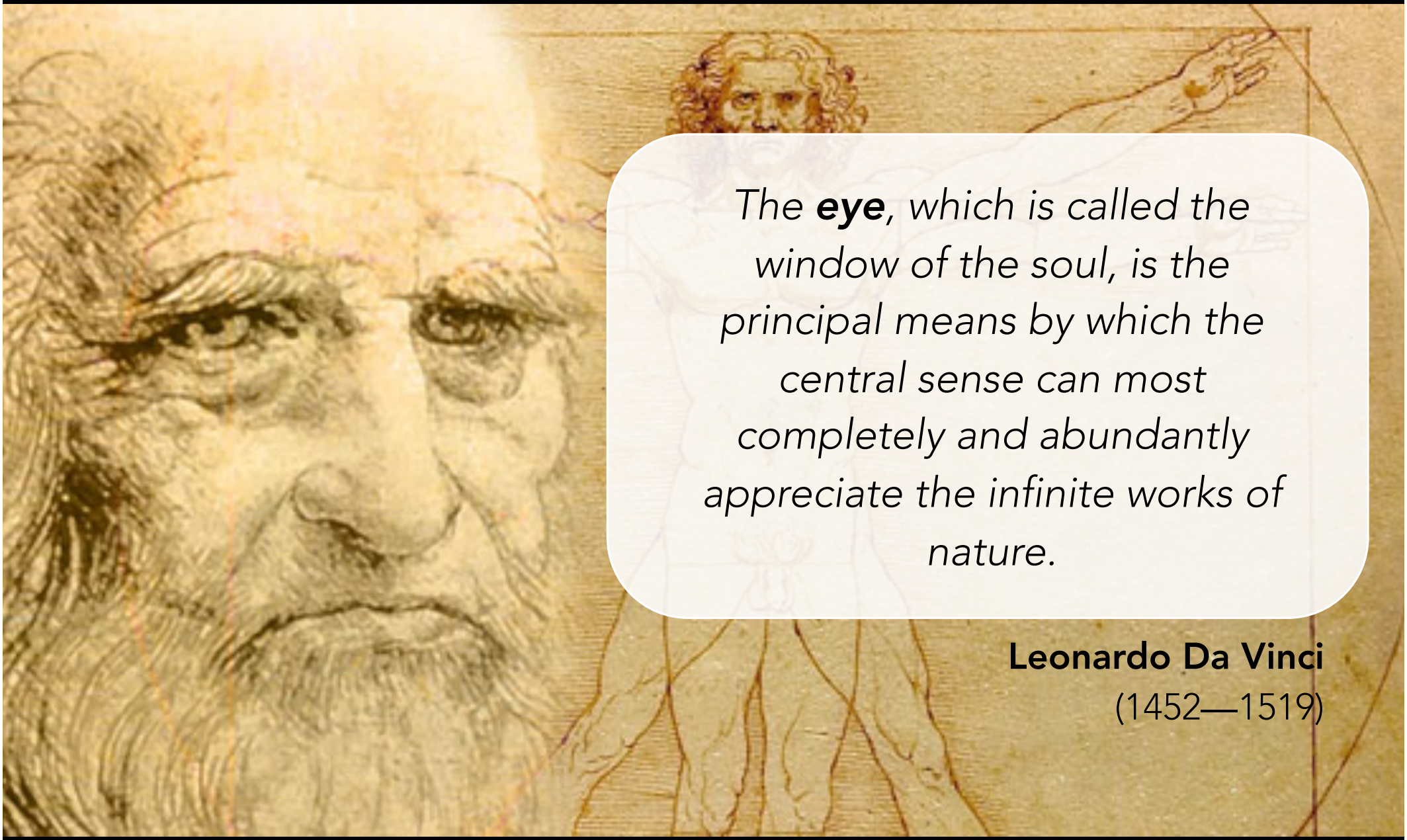
WHY VISUAL REPRESENTATIONS?

- **Vision** is the sense with the **highest bandwidth** (≈ 100MB/s, then ears <100b/s);
- **Vision extends** memory and cognition
- people **think visually**

HUMAN IN THE LOOP

- it is sometimes dangerous to rely on purely automated analyses
- **human judgment** and **intervention** often needed
  - for: background information, flexible analysis (unintended directions), creativity
  - because: data can be incomplete, inconsistent, or deceptive

Slide adapted from Petra Isenberg's course on Visual Analytics



*The **eye**, which is called 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.*

Leonardo Da Vinci
(1452—1519)

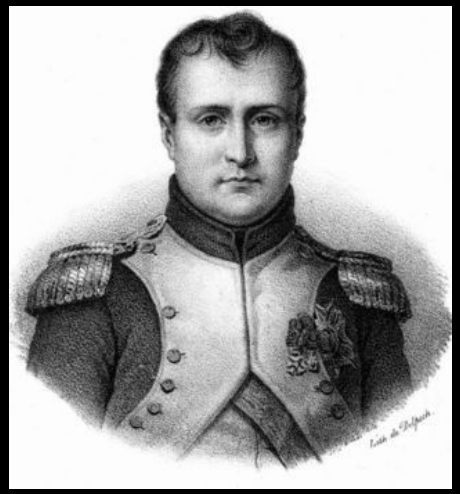
"A PICTURE IS WORTH A THOUSAND WORDS"

(Anonymous, 1911)

百聞不如一見

"One hundred rumors are not comparable to one look."

An Old Chinese Inscription



Napoleon Bonaparte (18xx)

"Un petit dessin vaut mieux qu'un long discours"

FOUNDATION OF INFORMATION VISUALISATION

DEFINITION & HISTORICAL EXAMPLES

WHAT IS VISUALIZATION?

1. The representation of an object, situation, or set of information as a chart or other image.

2. The formation of a mental image of something.

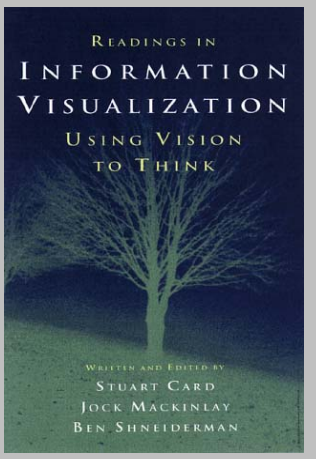
INFORMATION VISUALIZATION : Infovis

- Design **visual representations**
- Concerns **abstract data**
- Includes **interaction**

Official definition:

The use of computer-supported, interactive, visual representations of abstract data to amplify cognition.

[Card Mackinlay & Shneiderman, 1998]



Involves many fields:

- **graphics** (millenuiums of history)
- **cognitive psychology** (centuries of history)
- **Human-computer interaction** (decades of history)

SCIENTIFIC VISUALIZATION : SciViz

Visualization of data sets captured from real world, having a **given spatialization**.

Key differences with Information Visualization:

- concern data with a physical existence in the world
- limited set of application domains
- smaller design space

VISUAL ANALYTICS

Visual Analytics combines **automated analysis** techniques with **interactive visualizations** for an effective understanding, reasoning and decision making on the basis of **very large and complex data** sets.

Key differences with Information Visualization:

- involves automated data mining, information retrieval, data retrieval

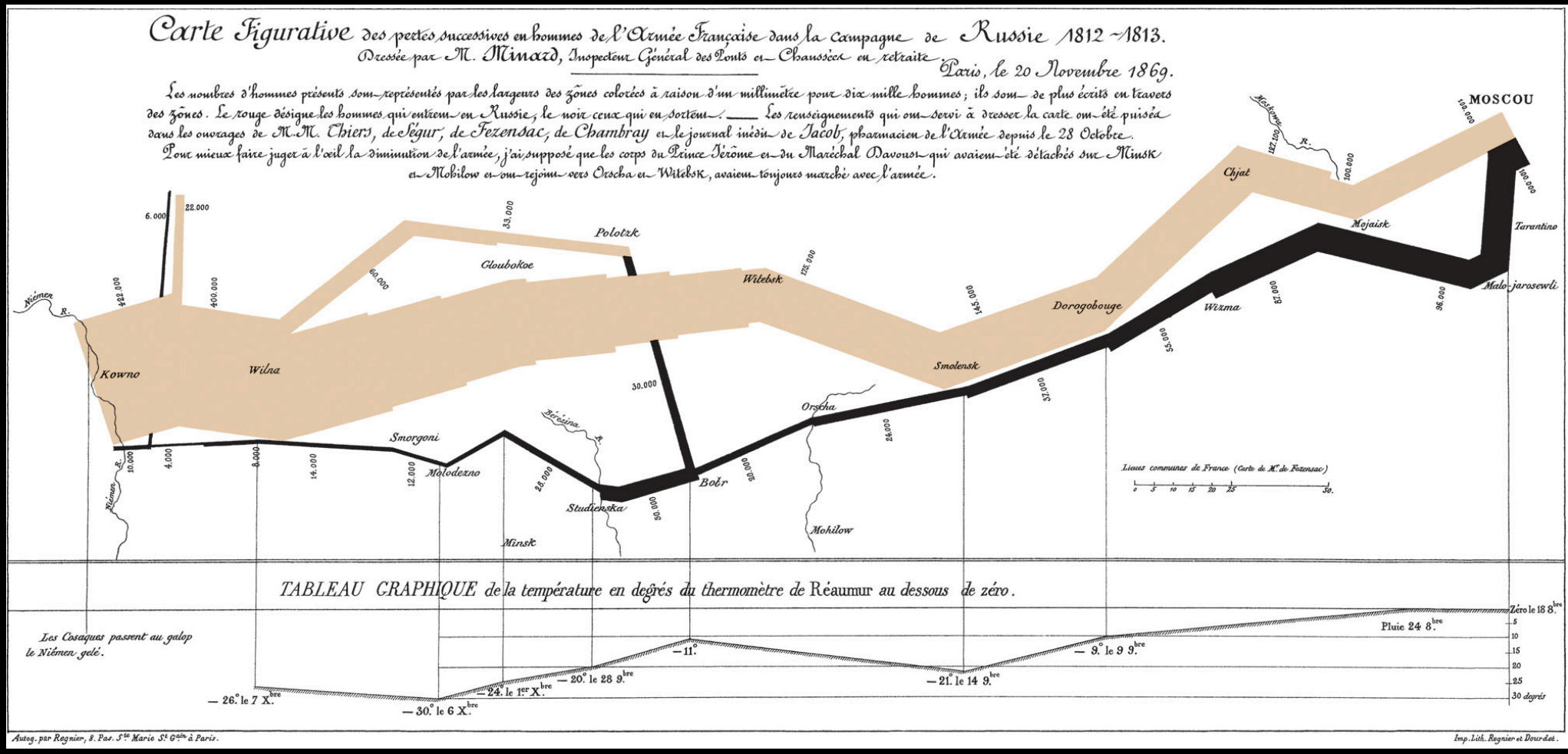
WHY VISUAL REPRESENTATIONS?

- **Vision** is the sense with the **highest bandwidth** (≈ 100MB/s, then ears <100b/s);
- **Vision extends** memory and cognition
- people **think visually**

VISUAL THINKING: NAPOLEON'S MOSCOW MARCH

Qualified by Edward Tufte as the best statistical representation ever.

Charles Minard, 1869



More about this: The Visual Display of Quantitative Information (Tufte)

VISUAL THINKING: BROAD STREET CHOLERA OUTBREAK (1854)

"The most terrible outbreak of cholera which ever occurred in the kingdom"

– John Snow

Major cholera outbreak in London in 1854

- 127 deaths within 3 days, close to Broad Street
- 616 deaths within 30 days

Dr. John Snow was the first to make the link between contaminated water pumps and the disease propagation

How did he do?

- Talked to local residents
- Hypothesized water pumps as potential source
- Used annotated maps to illustrate his theory
- Convinced authorities to condemn pumps

More about this: The Visual Display of Quantitative Information (Tufte)

VISUAL THINKING: BROAD STREET CHOLERA OUTBREAK (1854)

More about this: The Visual Display of Quantitative Information (Tufte)

VISUAL THINKING: CHALLENGER SPACE SHUTTLE (1986)

Source: Space Shuttle Challenger Disaster, Wikipedia

VISUAL THINKING: CHALLENGER SPACE SHUTTLE (1986)

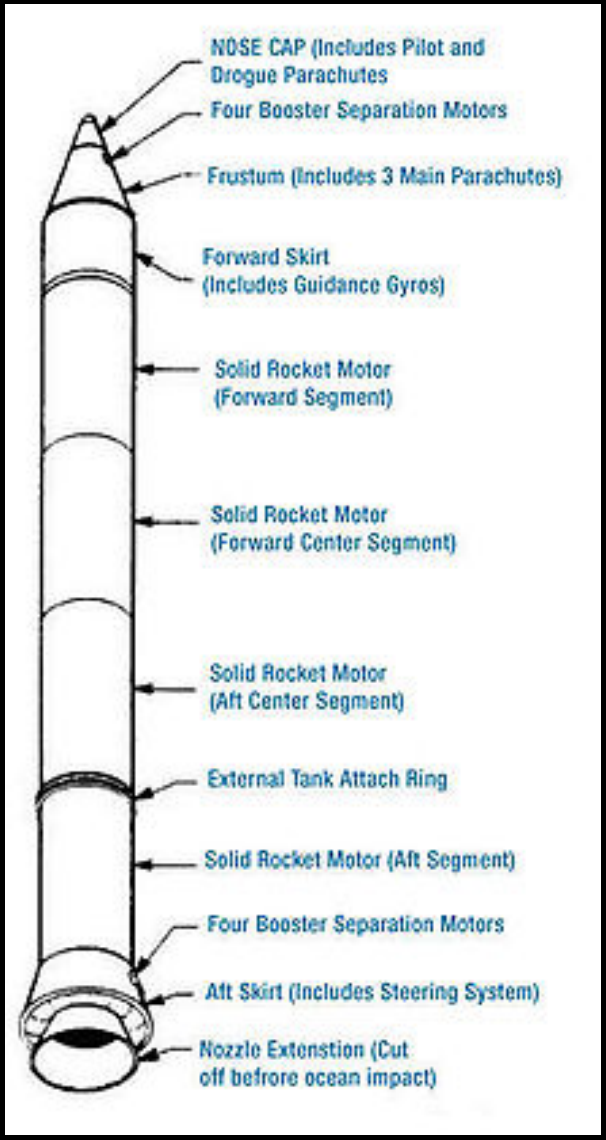
ice on the launch tower, hours before *Challenger* launch

7 crew members died during the explosion

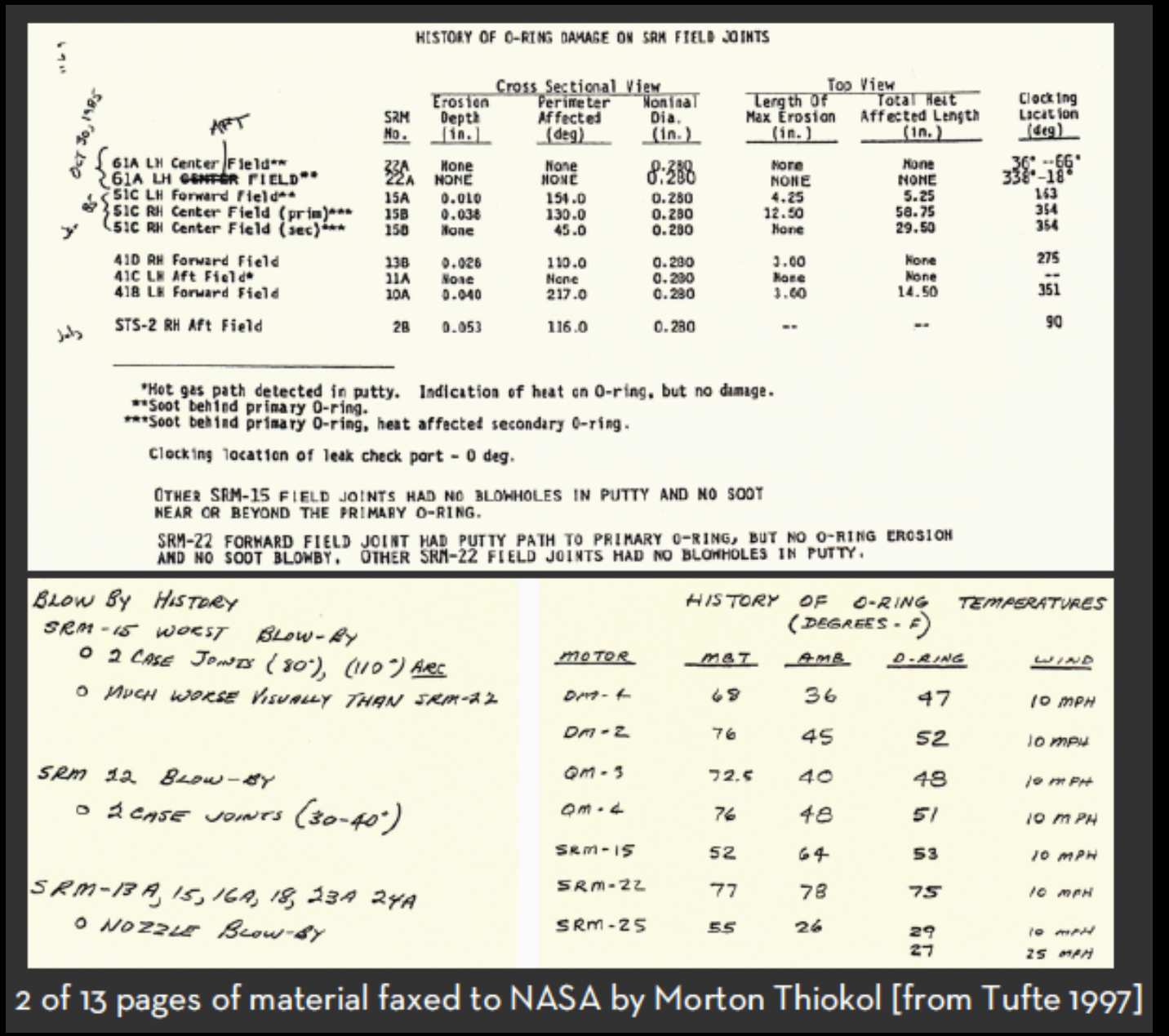
The disaster could have been avoided

- Weather forecast for Jan. 28th announced exceptionally cold morning, with temperatures close to -0.5°C
- Morton Thiokol engineers, in charge of the solid rocket booster (SRB), were concerned about low temperatures
- Engineers feared the effect of low temperature on the joint resistance

Solid rocket booster provides thrust during the first two minutes of flight space shuttle



VISUAL THINKING: CHALLENGER SPACE SHUTTLE (1986)



HISTORY OF O-RING DAMAGE ON SRM FIELD JOINTS

	SRM No.	Cross Sectional View: Erosion Depth (in.)	Cross Sectional View: Perimeter Affected (deg)	Cross Sectional View: Nominal Dia. (in.)	Top View: Length Of Max Erosion (in.)	Top View: Total Heat Affected Length (in.)	Clocking Location (deg)
61A LH Center Field**	22A	None	None	0.280	None	None	36° --66°
61A LH ~~CENTER~~ FIELD**	22A	NONE	NONE	0.280	NONE	NONE	338°-18°
51C LH Forward Field**	15A	0.010	154.0	0.280	4.25	5.25	163
51C RH Center Field (prim)***	15B	0.038	130.0	0.280	12.50	58.75	354
51C RH Center Field (sec)***	15B	None	45.0	0.280	None	29.50	354
41D RH Forward Field	13B	0.028	110.0	0.280	3.00	None	275
41C LH Aft Field*	11A	None	None	0.280	None	None	--
41B LH Forward Field	10A	0.040	217.0	0.280	3.00	14.50	351
STS-2 RH Aft Field	2B	0.053	116.0	0.280	--	--	90

*Hot gas path detected in putty. Indication of heat on O-ring, but no damage.
**Soot behind primary O-ring.
***Soot behind primary O-ring, heat affected secondary O-ring.

Clocking location of leak check port = 0 deg.

OTHER SRM-15 FIELD JOINTS HAD NO BLOWHOLES IN PUTTY AND NO SOOT NEAR OR BEYOND THE PRIMARY O-RING.

SRM-22 FORWARD FIELD JOINT HAD PUTTY PATH TO PRIMARY O-RING, BUT NO O-RING EROSION AND NO SOOT BLOWBY. OTHER SRM-22 FIELD JOINTS HAD NO BLOWHOLES IN PUTTY.

BLOW BY HISTORY

SRM-15 WORST BLOW-BY
- 2 CASE JOINTS (80°), (110°) ARC
- MUCH WORSE VISUALLY THAN SRM-22

SRM 12 BLOW-BY
- 2 CASE JOINTS (30-40°)

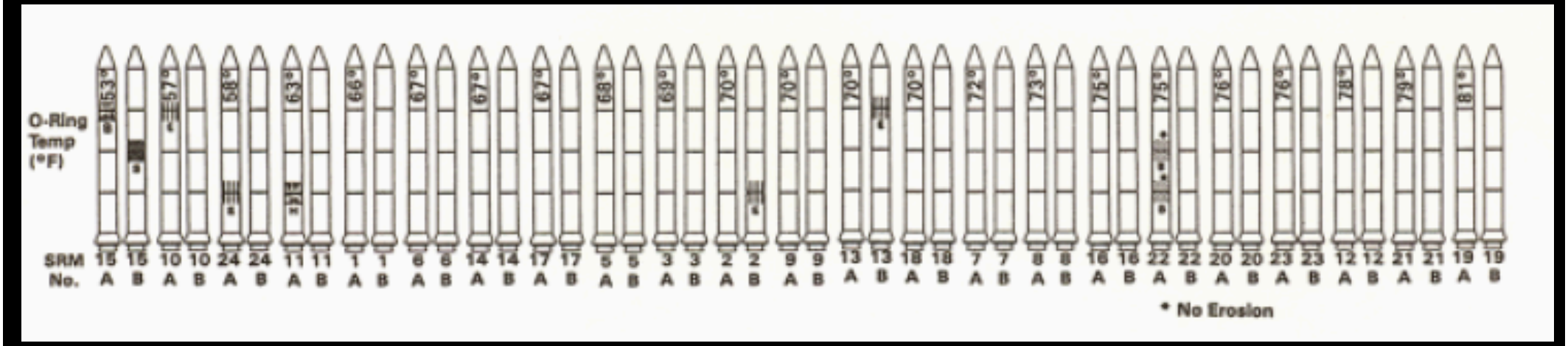
SRM-13A, 15, 16A, 18, 23A 24A
- NOZZLE BLOW-BY

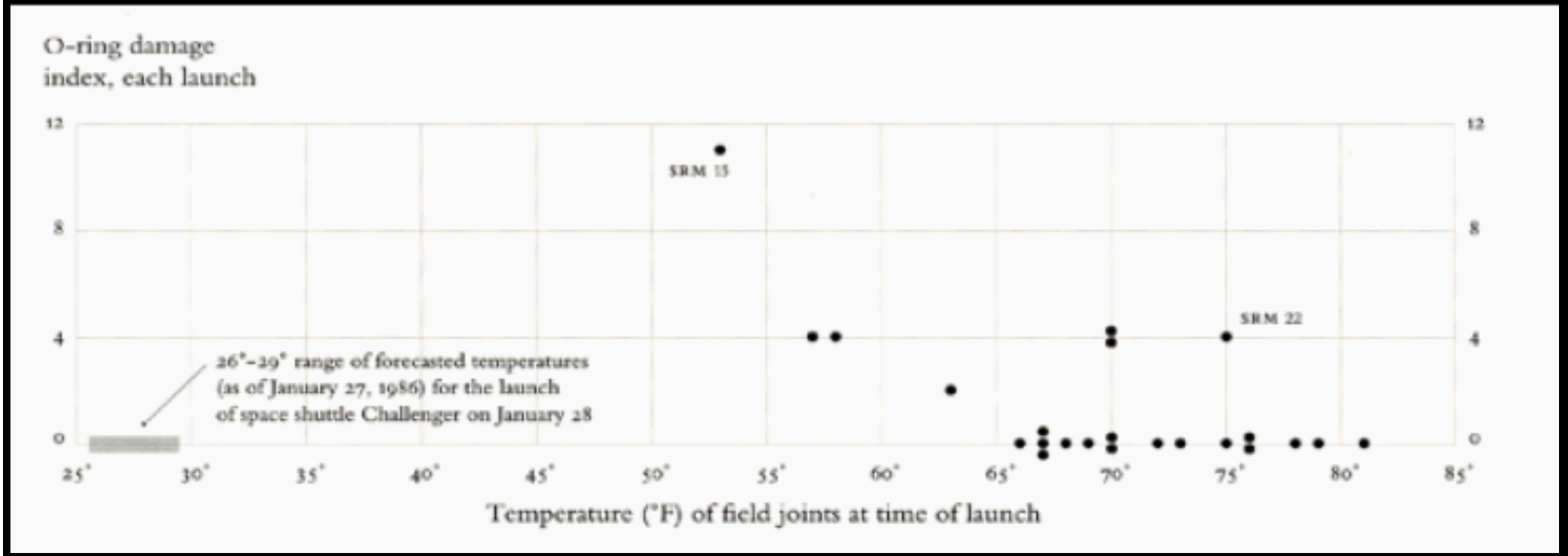
HISTORY OF O-RING TEMPERATURES (DEGREES - F)

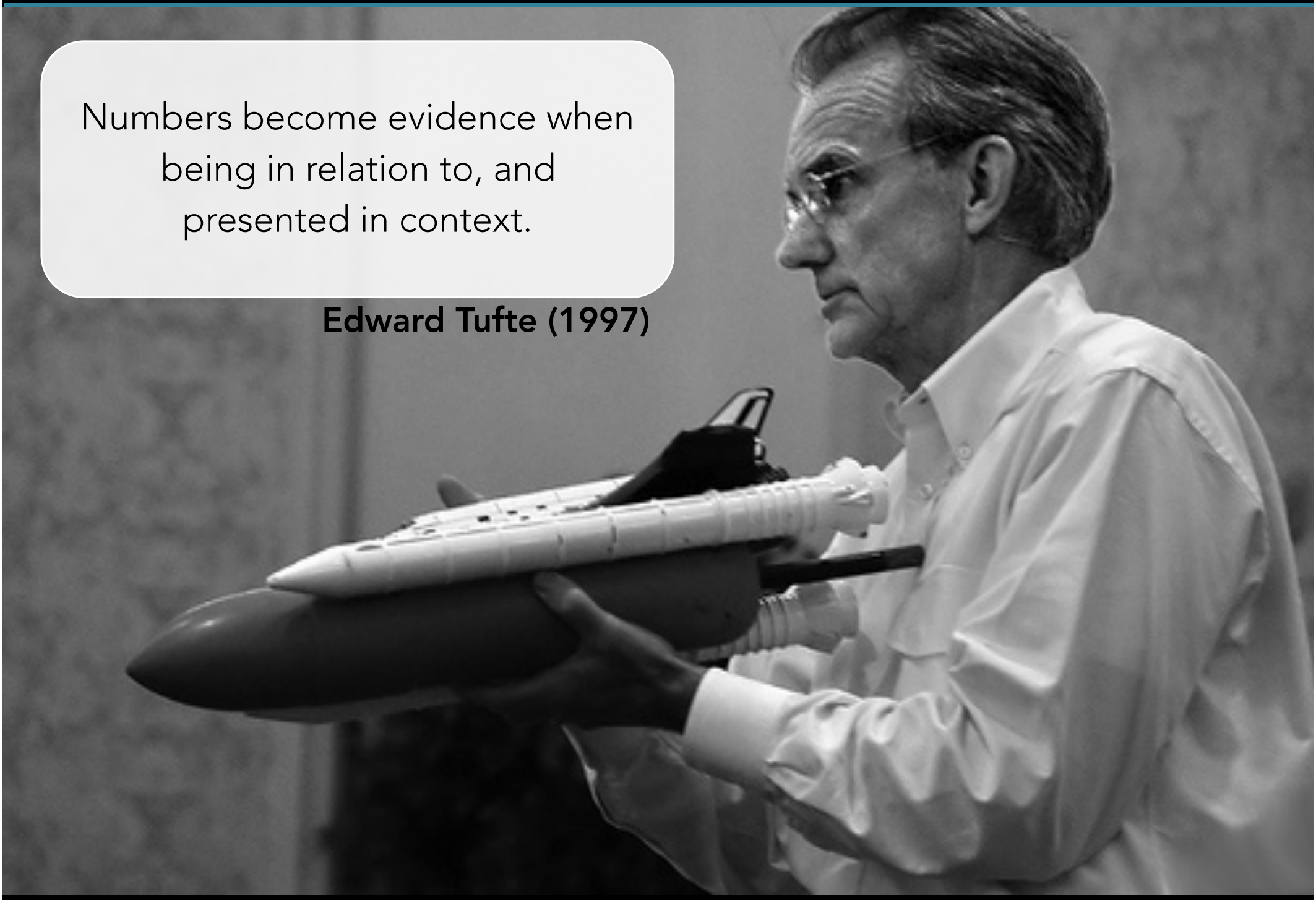
MOTOR	M&T	AMB	O-RING	WIND
DM-1	68	36	47	10 MPH
DM-2	76	45	52	10 MPH
QM-3	72.5	40	48	10 MPH
QM-4	76	48	51	10 MPH
SRM-15	52	64	53	10 MPH
SRM-22	77	78	75	10 MPH
SRM-25	55	26	29 27	10 MPH 25 MPH

2 of 13 pages of material faxed to NASA by Morton Thiokol [from Tufte 1997]

VISUAL THINKING: CHALLENGER SPACE SHUTTLE (1986)






Numbers become evidence when being in relation to, and presented in context.
Edward Tufte (1997)

CHALLENGES

data

- quantity (e.g. large and streaming data)
- quality of data is often low
- dealing with uncertainty in the data

CHALLENGES

human perception and reasoning

- understanding and supporting how humans perceive and reason about data
- create representations that are fair to the data
- create interfaces that are meaningful, clear, effective, and efficient

CHALLENGES

evaluation

- develop methods to compare novel techniques / tools to existing ones

- assess how good (effective, efficient, etc.) a technique / tool is

  very difficult for measures other than time & error, e.g. how many insights a technique / tool generates

WRAP UP

AGENDA

- We will pick a theme every week and go over representative papers in the area.
- Potential papers and themes and a schedule is on the course webpage.
- Students will present one (or two) such papers at one such session (decided by the second week).
- Students should also define groups and pick a project in consultation with the instructor. (decided by the third week).

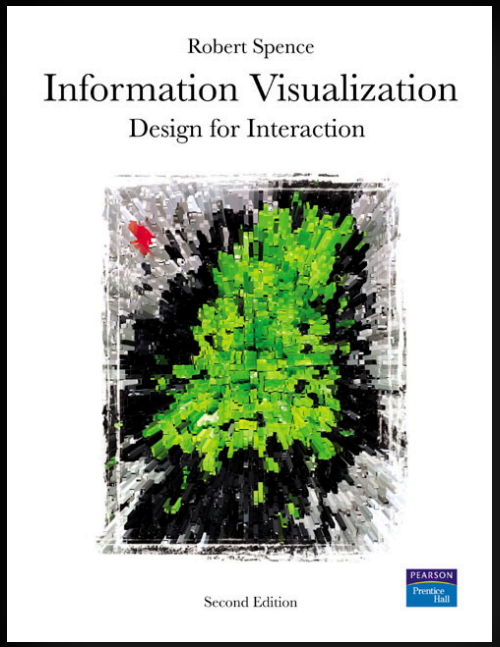
RESOURCES

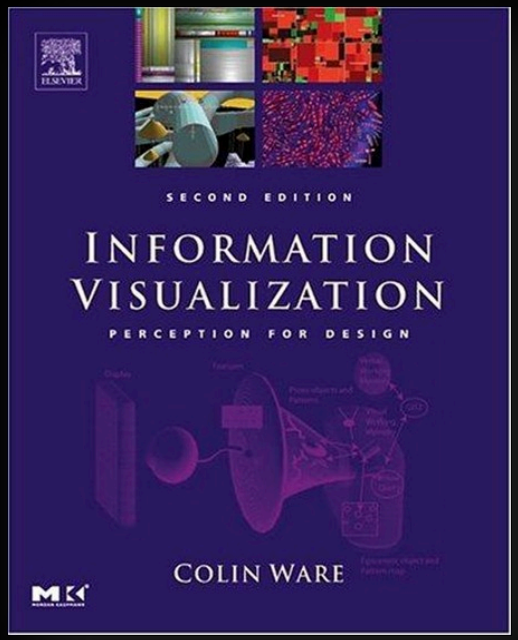
Blogs

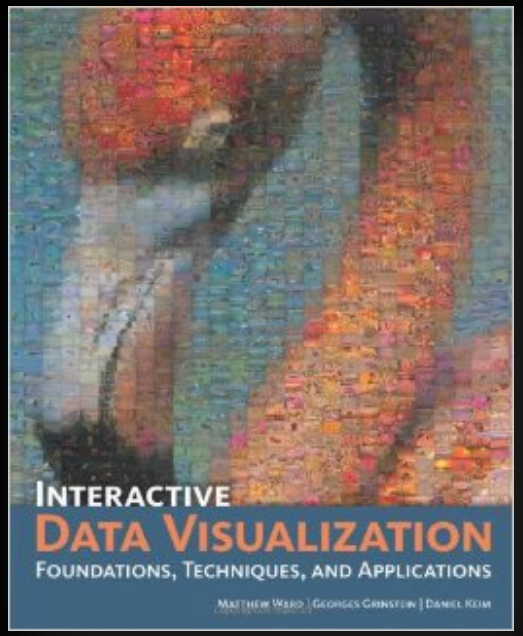
http://flowingdata.com/
http://fellinlovewithdata.com/
http://eagereyes.org/
http://infosthetics.com/

Tufte's collection

Books







ASSIGNMENTS

PAPER PRESENTATIONS & PROSPECTIVE PROJECTS

PAPERS PRESENTATION

PAPERS TO BE PICKED

Here is a list of papers to pick from, roughly organized by theme. Note that this list is not exhaustive. You can propose another academic paper for this assignment. In this case, you must first get the approval of the instructor that the paper is suitable for the assignment.

You must use this form to submit your choice (first choice, second choice). A dedicated field allows to enter your own paper proposal not included in the list below. Choice of papers submitted by another means will **not** be considered. Papers will be assigned on a first-in-first-served basis.

EXPLORATORY DATA ANALYSIS

- Integrating Statistics and Visualization: Case Studies of Gaining Clarity during Exploratory Data Analysis (Perer and Shneiderman, CHI 2008)
- Polaris: A System for Query, Analysis and Visualization of Multi-dimensional Relational Databases (Stolte and Hanrahan, TVCG 2002)
- Voyager: Exploratory Analysis via Faceted Browsing of Visualization Recommendations (Wongsuphasawat et al., TVCG 2016)
- Orko: Facilitating Multimodal Interaction for Visual Exploration and Analysis of Networks (Srinivasan and Stasko, TVCG 2017)
- Exploratory Analysis of Time-series with ChronoLenses (Zhao et al., Infovis 2011)

GRAPHS & NETWORKS

- Supporting Handoff in Asynchronous Collaborative Sensemaking Using Knowledge-Transfer Graphs (Zhao et al., TVCG 2017)
- What Would a Graph Look Like in This Layout? A Machine Learning Approach to Large Graph Visualization (Kwon et al., TVCG 2017)
- Functional Decomposition for Bundled Simplification of Trail Sets (Hurter et al., TVCG 2017)
- Edge compression techniques for visualization of dense directed graphs (Dwyer et al., Infovis 2013)
- Visualizing Dense Dynamic Networks with Matrix Cubes (Bach et al., CHI'14)
- NodeTrix: a hybrid visualization of social networks (Henry et al., TVCG 2007)
- GeneaQuilts: A System for Exploring Large Genealogies (Bezerianos et al., Infovis 2010)
- Egocentric Analysis of Dynamic Networks with EgoLines (Zhao et al., CHI'16)
- Telling Stories about Dynamic Networks with Graph Comics (Bach et al., CHI'16)
- Annotation Graphs: A Graph-Based Visualization for Meta-Analysis of Data based on User-Authored Annotations (Zhao et al., TVCG 2016)

INTERACTION

- SketchSliders: Sketching Widgets for Visual Exploration on Wall Displays (Tsandilas et al., CHI 2015)

PAPERS PRESENTATION

12-minute presentation
3-5 minute questions

Should I stick solely to the content of the given paper?
Absolutely not! Context + related research welcome

Shall we all read the papers **before** the presentation?
Recommended

Shall we all read the papers **after** the presentation?
Absolutely!

PROJECT

Individual, or group project

It is expected that you explore and discuss

- Related work
- Analysis of domain, tasks, design goals
- Design & execution
- (Evaluation)
- Implications for visualization design

PROJECT DELIVERABLES

- Project proposal — 2%
- Mid-term presentation (5-10min presentation) — 10%
- Project report (i.e. research paper) — 20%
- Final presentation (live demos welcome!) — 10%
- Project execution (i.e. research contribution) — 18%

PROSPECTIVE PROJECT 1
VISUALIZING PIXEL HISTORY / LAYERS

PROSPECTIVE PROJECT 2
VISUALIZATION OF NUTRITIVE PROPERTIES

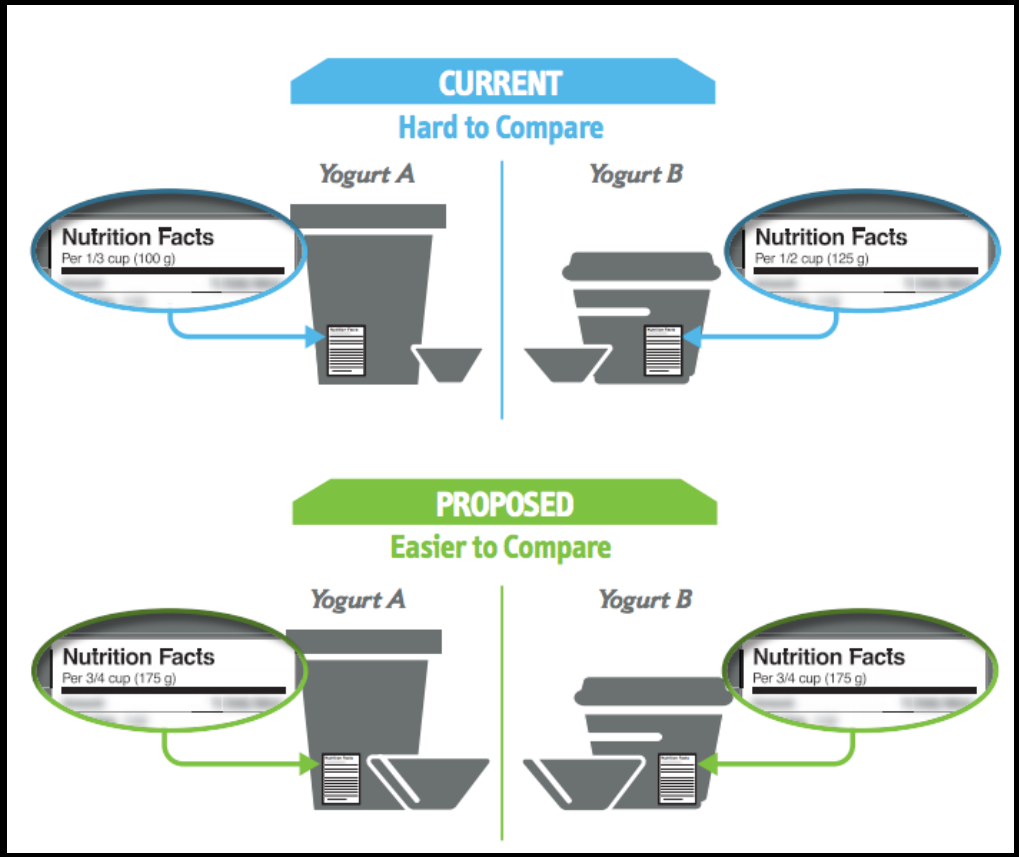
Cracker B

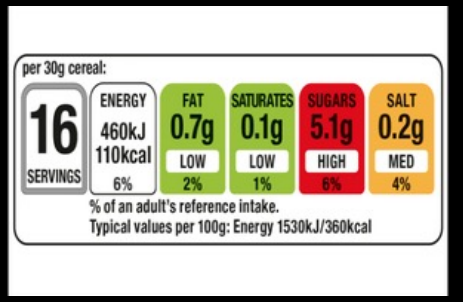
Nutrition Facts

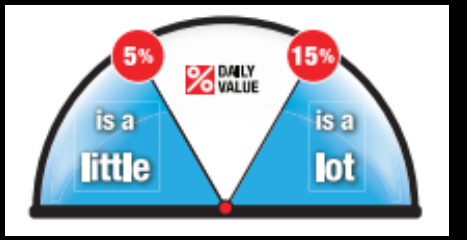
Per 4 crackers (20 g)

Amount	% Daily Value
Calories 85	
Fat 2 g	3 %
Saturated 0.3 g + Trans 0 g	2 %
Cholesterol 0 mg	
Sodium 90 mg	4 %
Carbohydrate 15 g	5 %
Fibre 3 g	12 %
Sugars 1 g	
Protein 2 g	

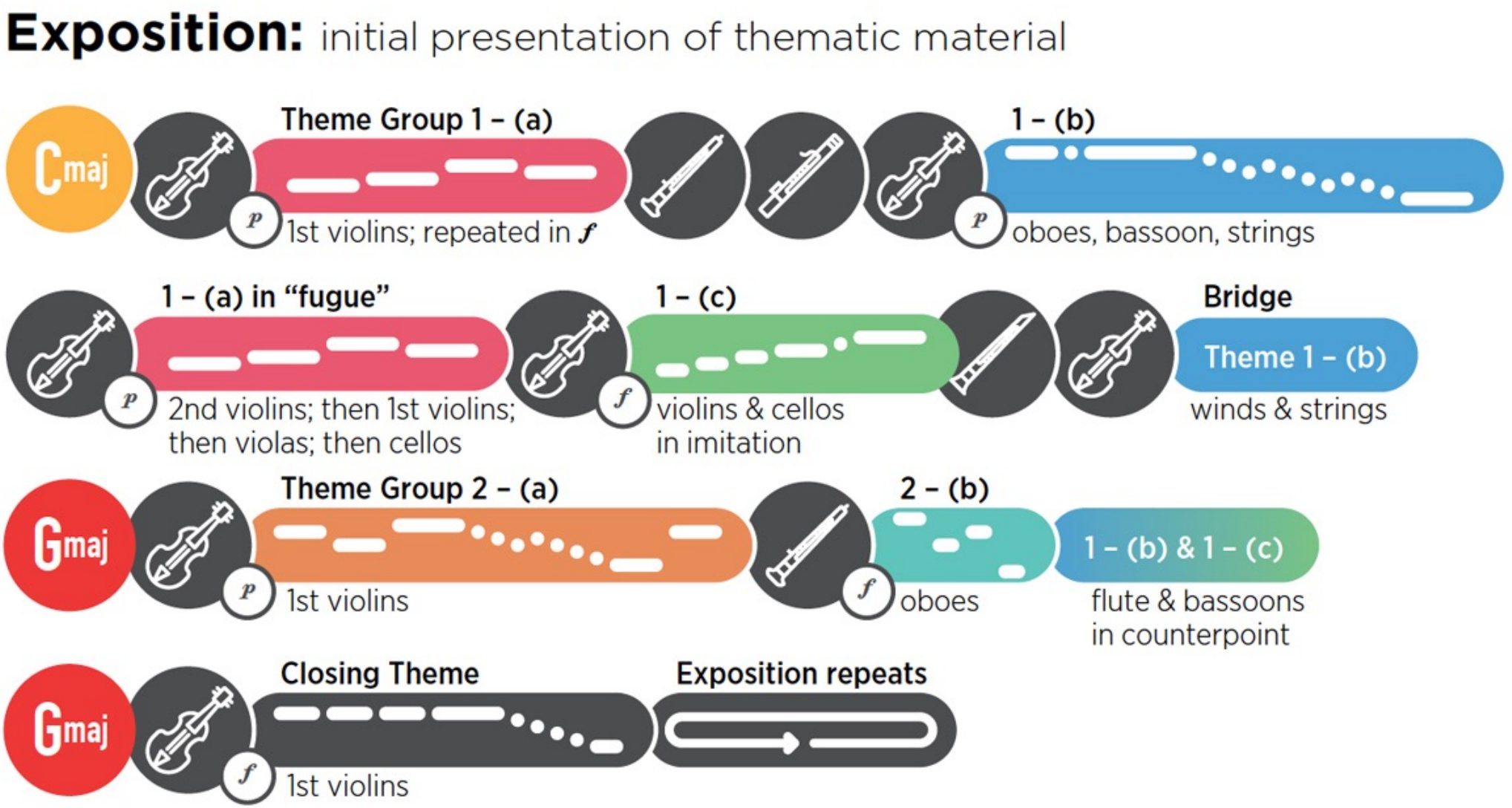
Vitamin A	0 %	Vitamin C	0 %
Calcium	2 %	Iron	7 %







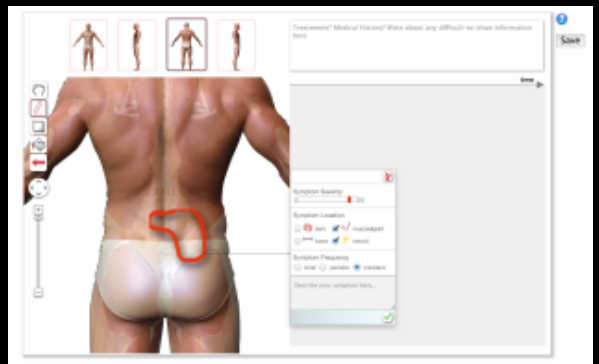
PROSPECTIVE PROJECT 3
VISUALIZATION OF MUSIC (ORCHESTRA)



PROSPECTIVE PROJECT 4
ANIMATED TRANSITIONS OF THE INTERNET

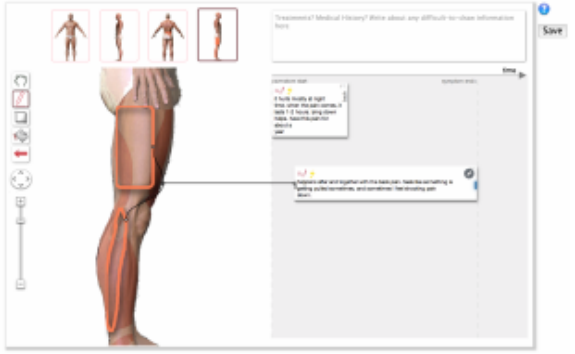
PROSPECTIVE PROJECT 5
(SKETCH-BASED) TOOL FOR AUTHORING ANIMATION

PROSPECTIVE PROJECT 6
EXPRESSING AND VISUALIZING BODY PAIN

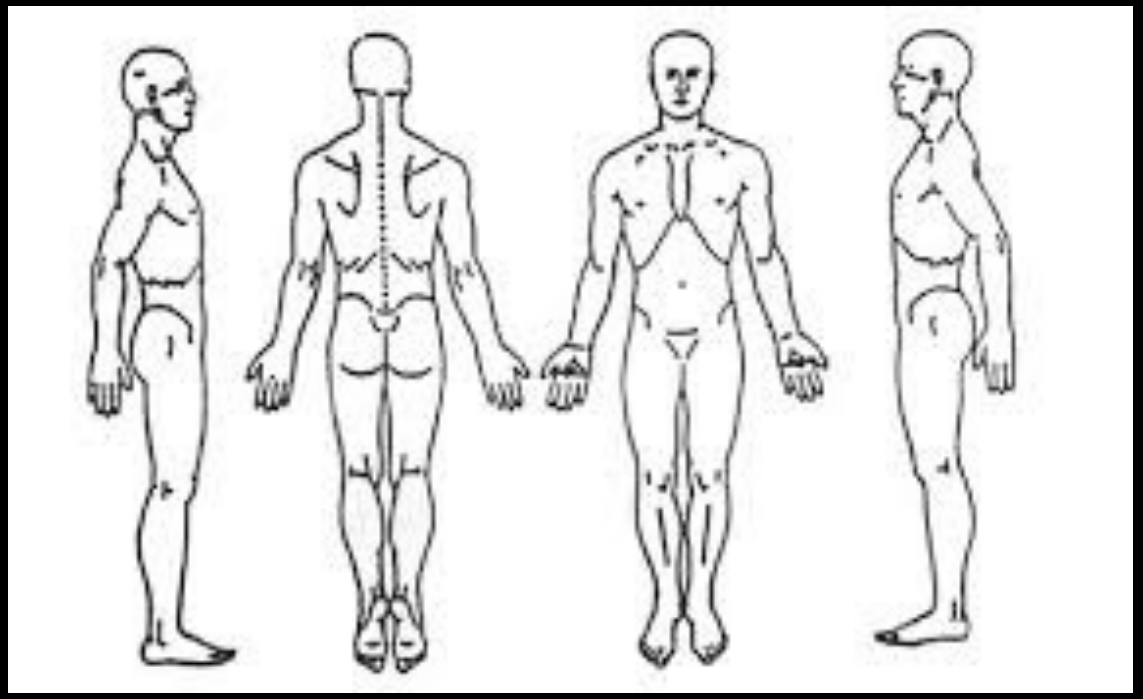


(b) Back pain drawn using pen.

(c) Grouped leg pain drawings.

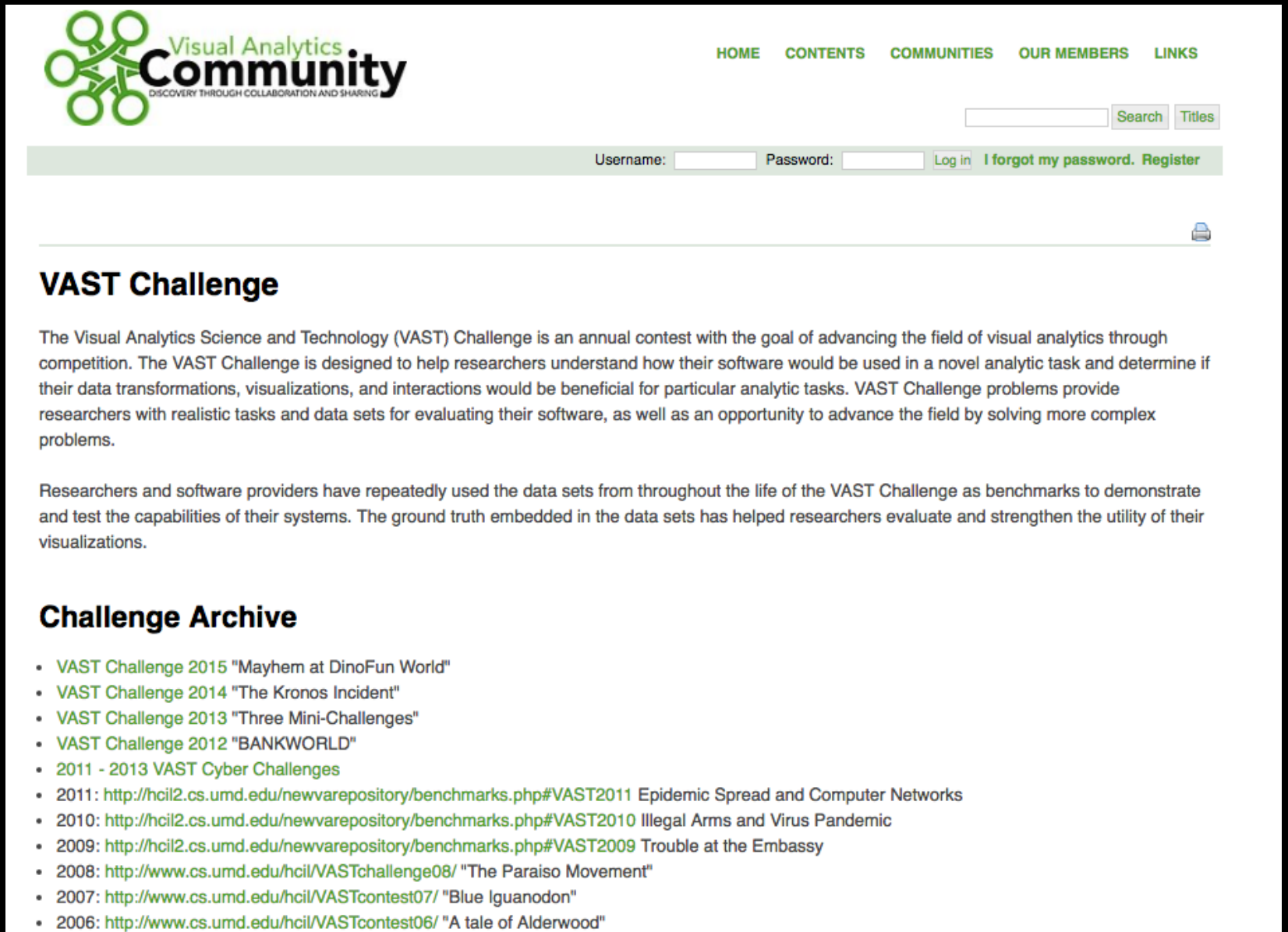


(d) Changing the duration and starting time.



PROSPECTIVE PROJECT 7
VISUALIZATION LITERACY AT SCHOOL

PROSPECTIVE PROJECT X
VAST CHALLENGES



Visual Analytics Community
DISCOVERY THROUGH COLLABORATION AND SHARING

HOME CONTENTS COMMUNITIES OUR MEMBERS LINKS

Search Titles

Username: Password: Log in I forgot my password. Register

VAST Challenge

The Visual Analytics Science and Technology (VAST) Challenge is an annual contest with the goal of advancing the field of visual analytics through competition. The VAST Challenge is designed to help researchers understand how their software would be used in a novel analytic task and determine if their data transformations, visualizations, and interactions would be beneficial for particular analytic tasks. VAST Challenge problems provide researchers with realistic tasks and data sets for evaluating their software, as well as an opportunity to advance the field by solving more complex problems.

Researchers and software providers have repeatedly used the data sets from throughout the life of the VAST Challenge as benchmarks to demonstrate and test the capabilities of their systems. The ground truth embedded in the data sets has helped researchers evaluate and strengthen the utility of their visualizations.

Challenge Archive

- VAST Challenge 2015 "Mayhem at DinoFun World"
- VAST Challenge 2014 "The Kronos Incident"
- VAST Challenge 2013 "Three Mini-Challenges"
- VAST Challenge 2012 "BANKWORLD"
- 2011 - 2013 VAST Cyber Challenges
- 2011: http://hcil2.cs.umd.edu/newvarepository/benchmarks.php#VAST2011 Epidemic Spread and Computer Networks
- 2010: http://hcil2.cs.umd.edu/newvarepository/benchmarks.php#VAST2010 Illegal Arms and Virus Pandemic
- 2009: http://hcil2.cs.umd.edu/newvarepository/benchmarks.php#VAST2009 Trouble at the Embassy
- 2008: http://www.cs.umd.edu/hcil/VASTchallenge08/ "The Paraiso Movement"
- 2007: http://www.cs.umd.edu/hcil/VASTcontest07/ "Blue Iguanodon"
- 2006: http://www.cs.umd.edu/hcil/VASTcontest06/ "A tale of Alderwood"