CSC2524 L0101—TOPICS IN INTERACTIVE COMPUTING

INFORMATION VISUALISATION

Fanny CHEVALIER

UNIVERSITY OF TORONTO

Inria informatics mathematics

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.

CREDIT

- **Scientific research paper presentation** — 40%
- **Project** — 60%
  - implementation and demo : 40% (mid-term evaluation is worth 10 of the 40%)
  - project report (1-4 pages) and presentation: 20%

PLANNING

WEEK 1	WELCOME - INTRODUCTION
WEEK 2	VISUAL PERCEPTION, DATA MODELS / PROSPECTIVE PROJECTS
WEEK 3	NETWORKS / MULTIDIMENSIONAL DATA
WEEK 4	INTERACTION / ANIMATION
WEEK 5	~HOLIDAY ~
WEEK 6	GUEST SPEAKER : CHRIS COLLINS — TEXT VISUALISATION
WEEK 7	MID-TERM REVIEW
WEEK 8	GUEST SPEAKER : RICHARD BRATH — VISUALIZATION IN INDUSTRY
WEEK 9	STUDENTS PAPER PRESENTATION
WEEK 10	GUEST SPEAKER : JUSTIN MATEJKA — VISUALIZATION AESTHETICS
WEEK 11	GUEST SPEAKER : ISABEL MEREILLES — DESIGN
WEEK 12	STUDENTS PAPER PRESENTATION
WEEK 13	FINAL PRESENTATIONS + WRAP UP

GUEST SPEAKERS

Christopher
COLLINS

UNIVERSITY OF ONTARIO INSTITUTE OF TECHNOLOGY

Isabel
MEREILLES

OCAD UNIVERSITY

Richard
BRATH

uncharted

Justin
MATEJKA

AUTODESK. RESEARCH

INTRODUCTION

WHY VISUALIZATION?

The Economist

A special report on managing information | February 27th 2010

Special Report | Data, data everywhere

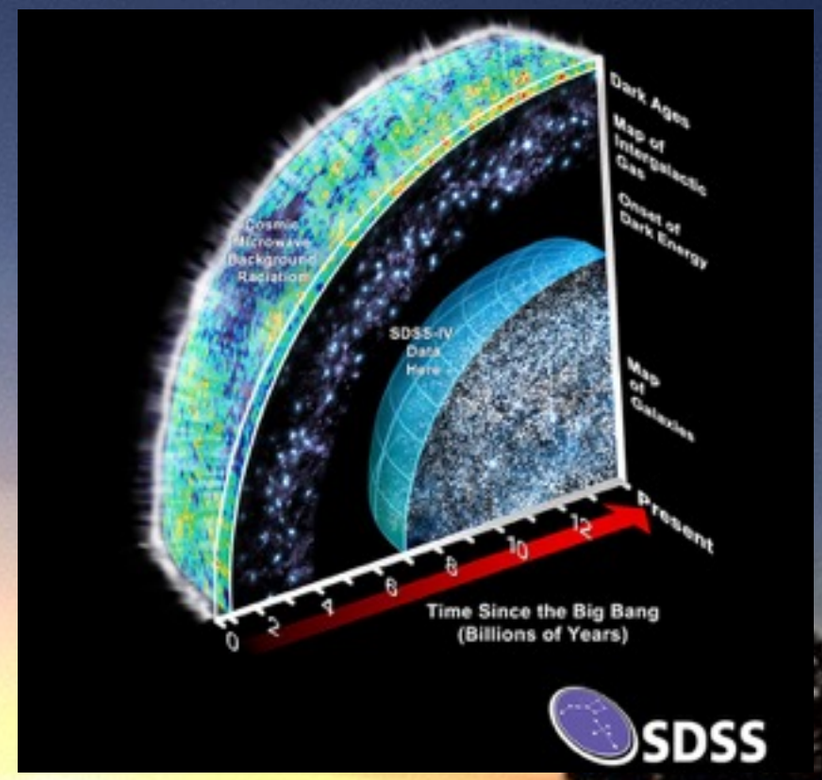
Information has gone from scarce to superabundant. That brings huge new benefits, says Kenneth Cukier (interviewed here)—but also big headaches

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

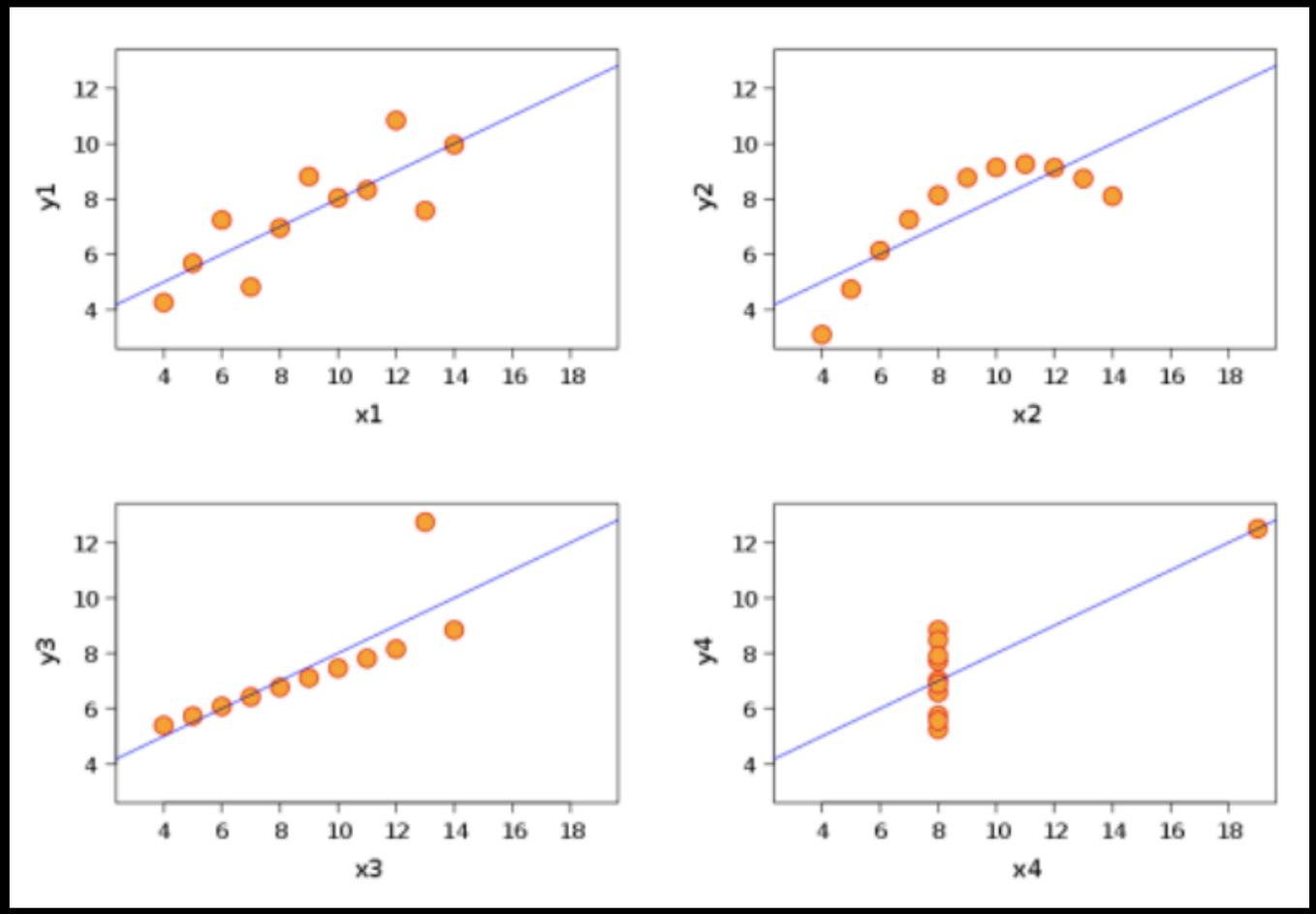
Mean of x	9
Sample variance of x	11
Mean of y	7.50
Sample variance of y	4.12
Correlation between x and y	0.816
Linear regression line	y = 3.00 + 0.500x

Source: Anascombe's Quartet, Wikipedia

VISUALIZATION

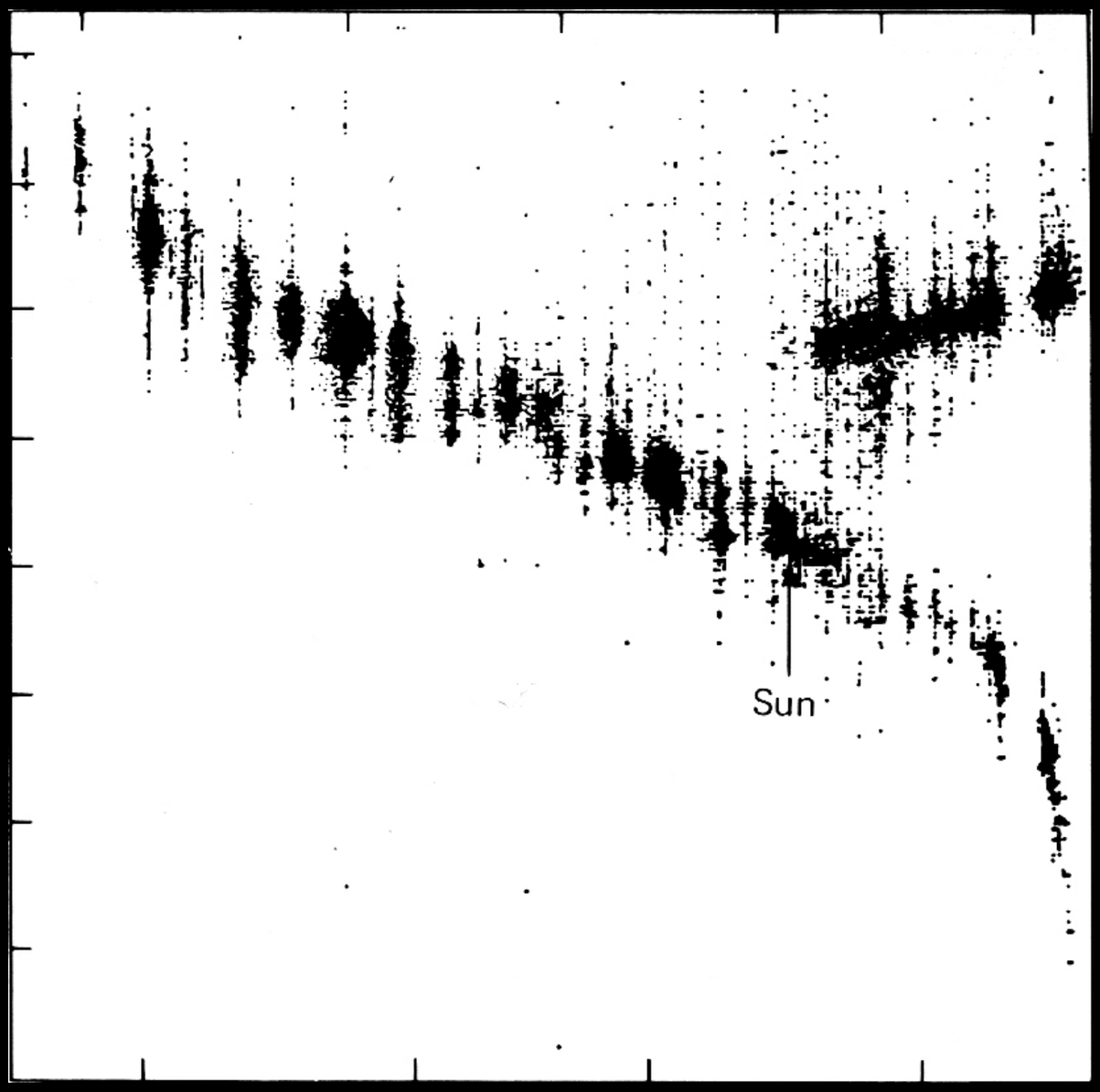
the visual representations tell a complete different story…

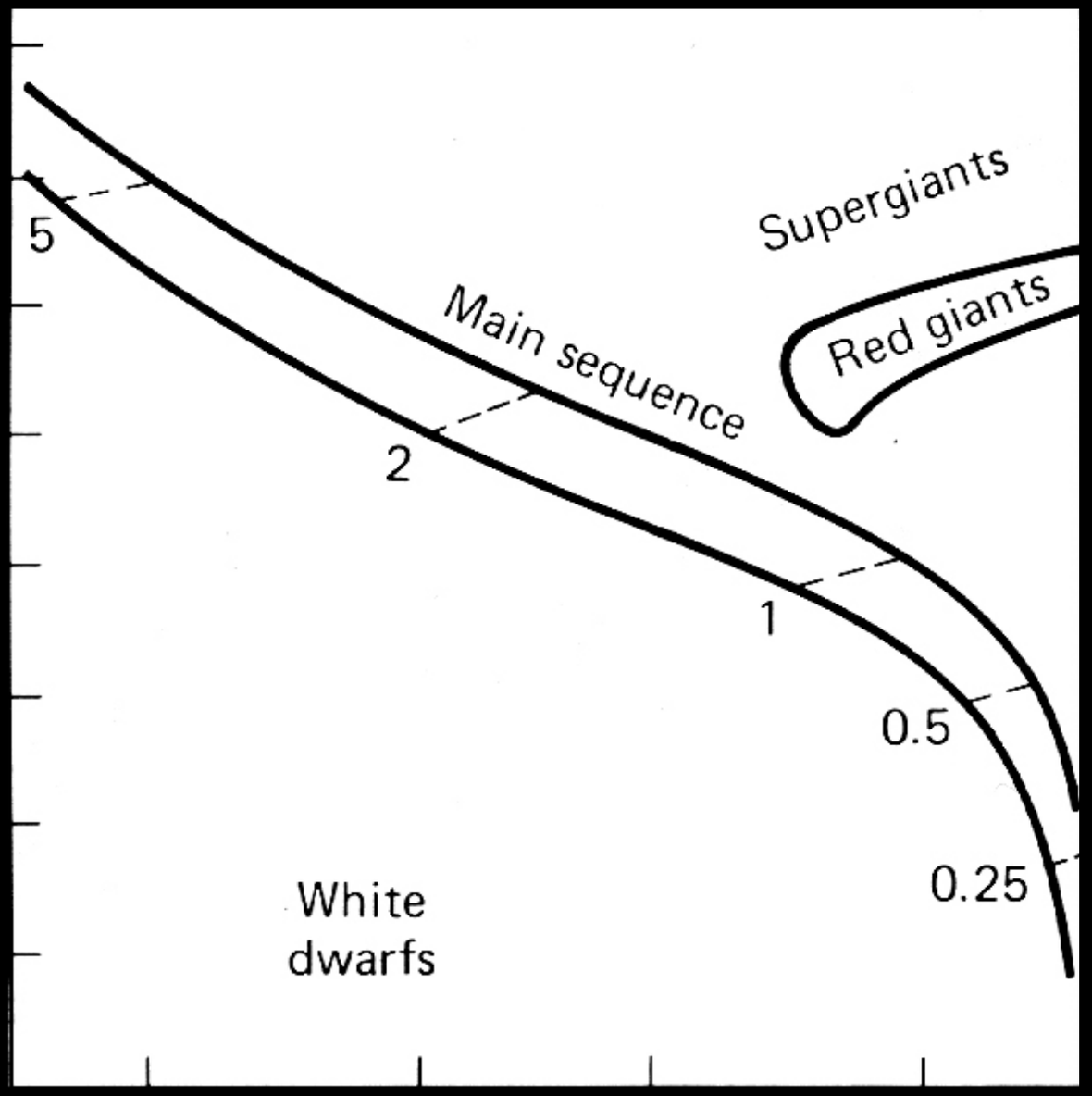
I		II		III		IV	
x	y	x	y	x	y	x	y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0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

AUTOMATIC ABSTRACTION CAPABILITY





Hertzsprung Russell Diagram and its interpretation

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

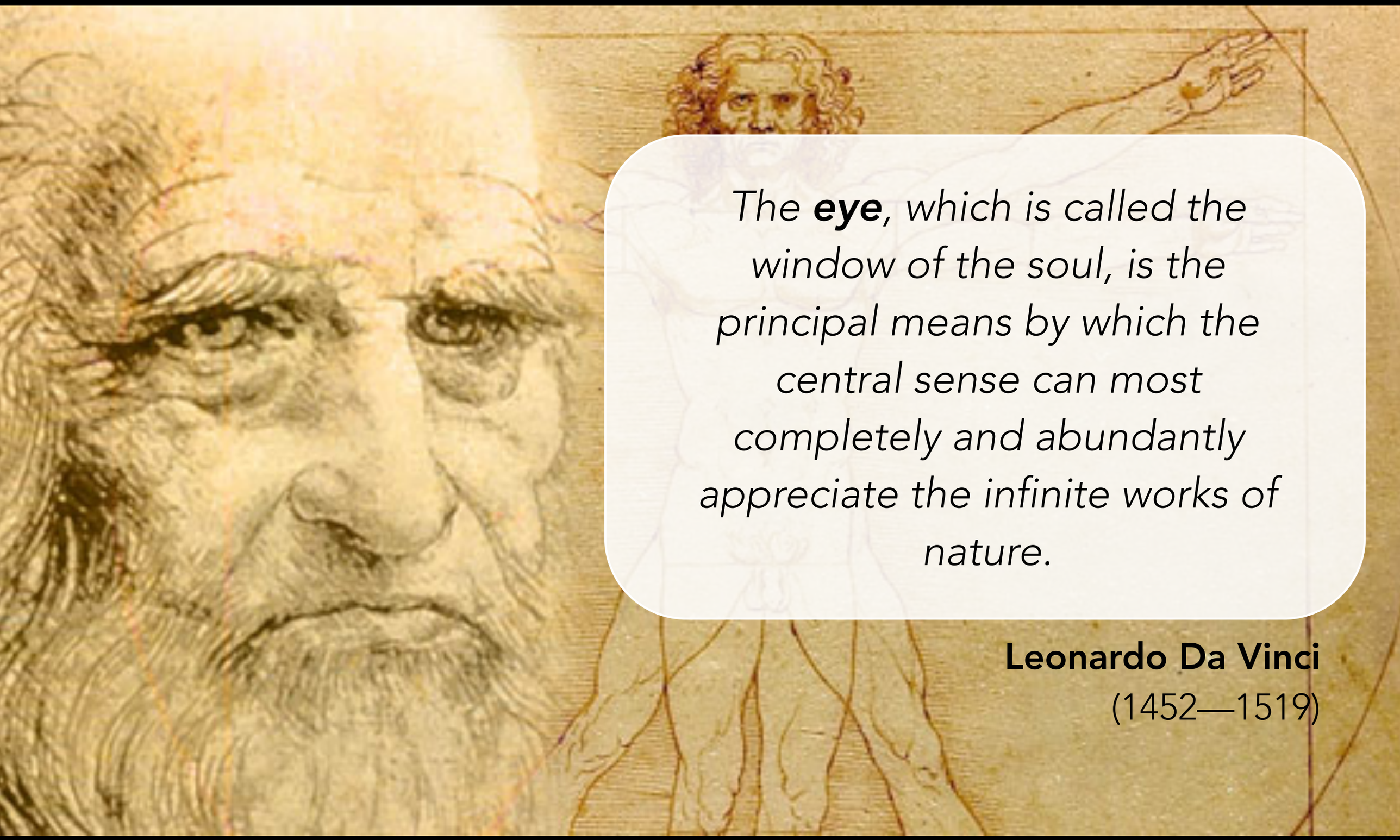
WHY VISUAL REPRESENTATIONS?

- **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)

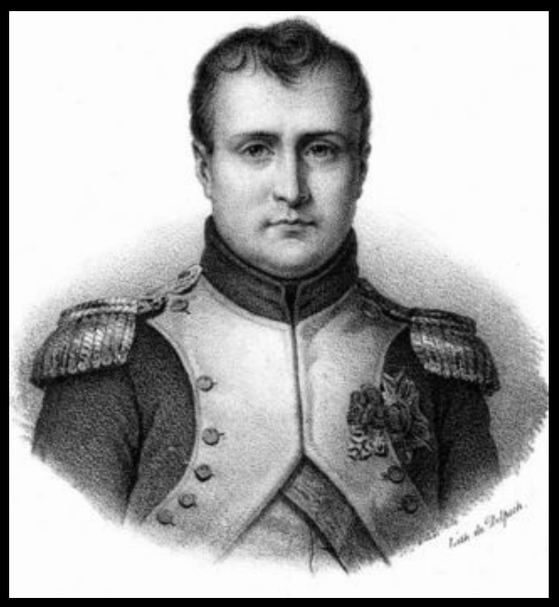
“A PICTURE IS WORTH A THOUSAND WORDS"

(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.

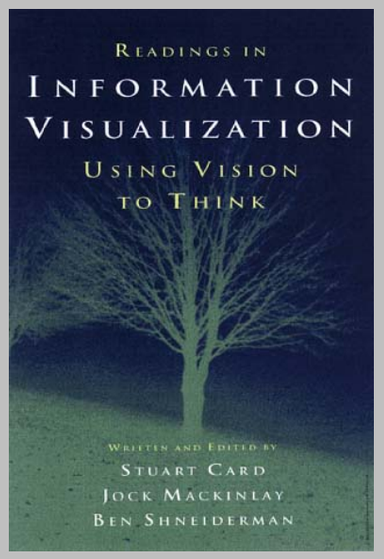
INFORMATION VISUALIZATION : Infovis

- 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

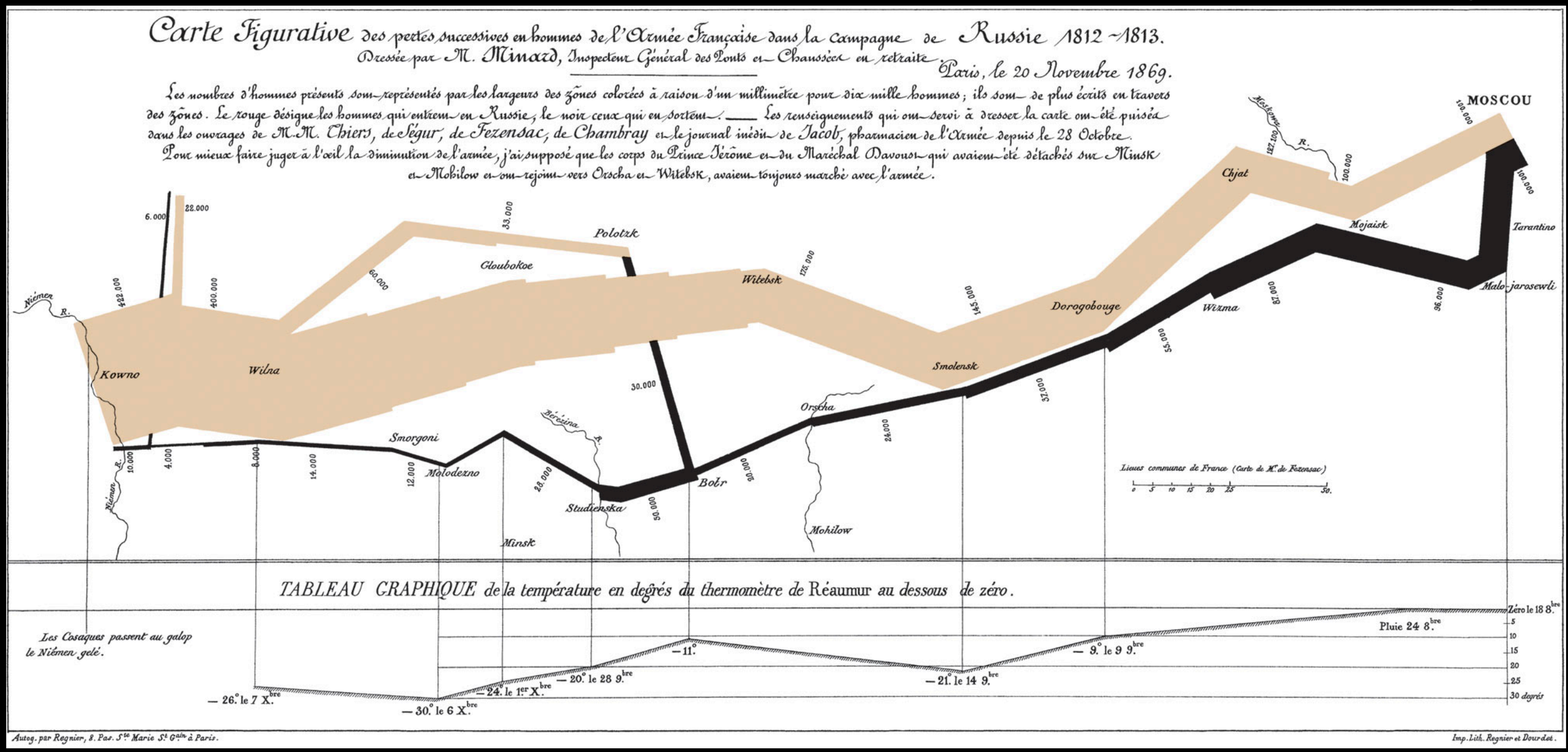
WHY VISUAL REPRESENTATIONS?

- **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)

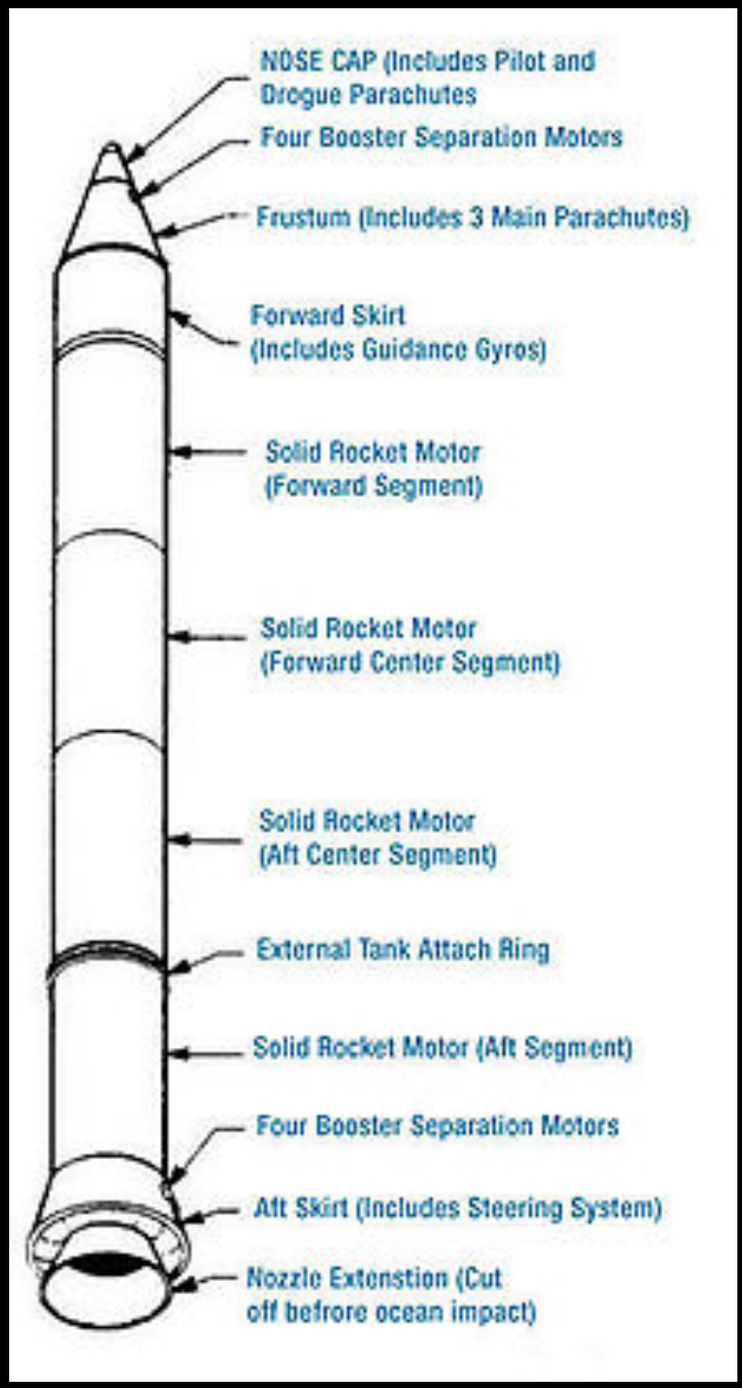
ice on the launch tower, hours before *Challenger* launch

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 Heit 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 part - 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°)

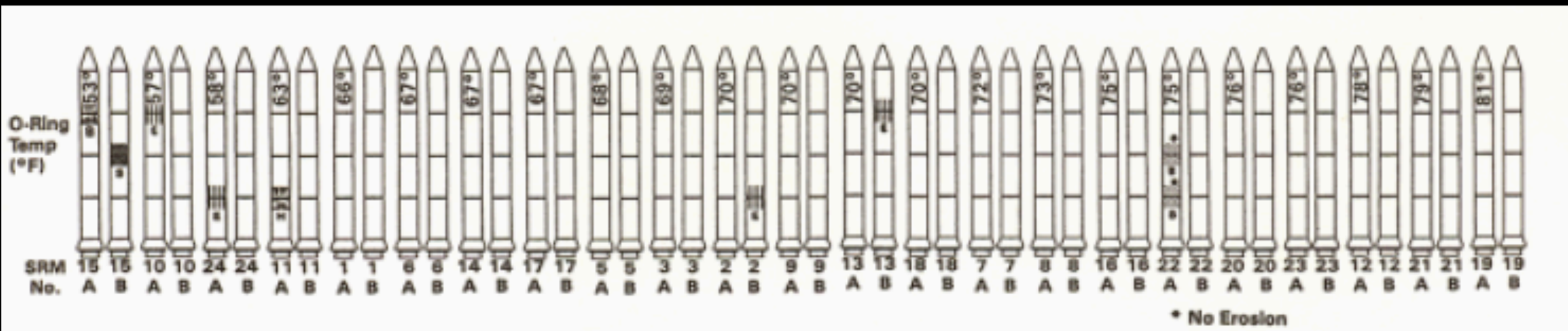
SRM-13A, 15, 16A, 18, 23A 24A
- NOZZLE BLOW-BY

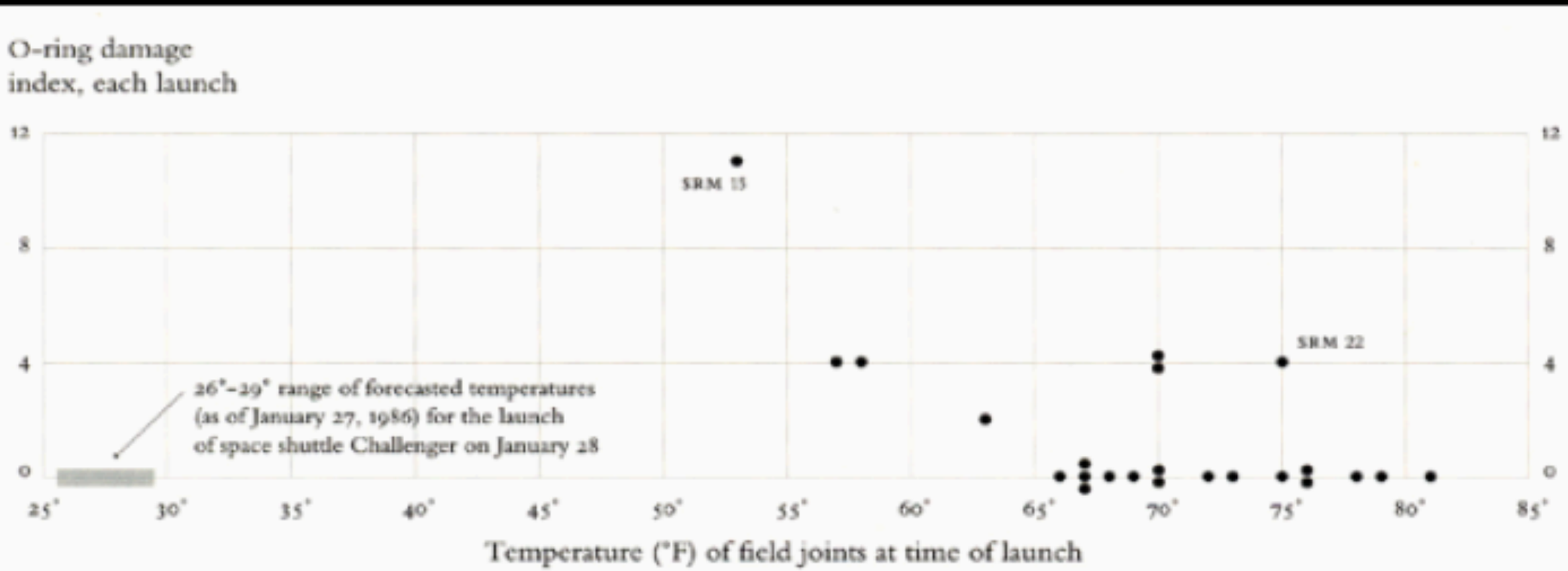
HISTORY OF O-RING TEMPERATURES (DEGREES - F)

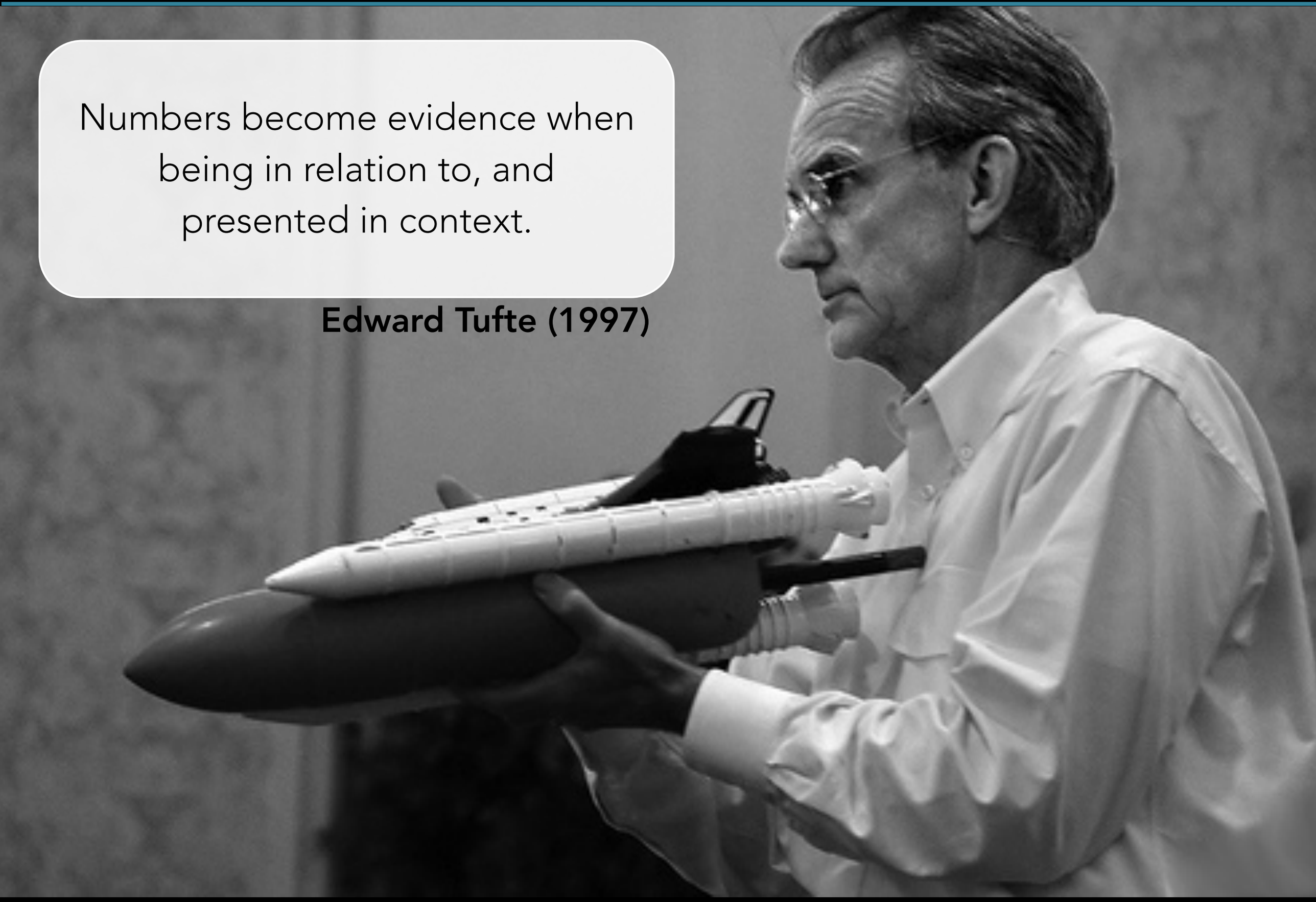
MOTOR	MBT	AMB	O-RING	WIND
QM-1	68	36	47	10 MPH
QM-2	76	45	52	10 MPH
QM-3	72.5	40	48	10 MPH
QM-4	76	48	51	10 MPH
SRM-15	52	64	53	10 MPH
SRM-22	77	78	75	10 MPH
SRM-25	55	26	29 27	10 MPH 25 MPH

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).

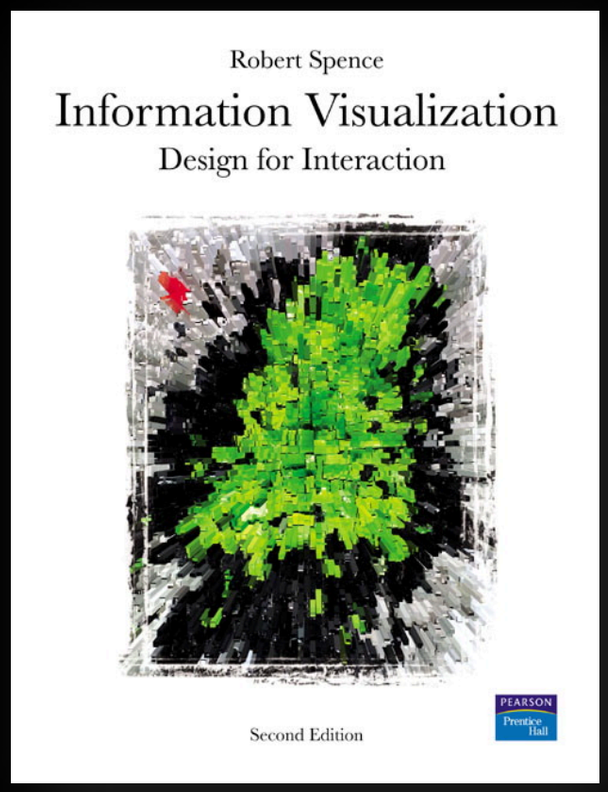
RESOURCES

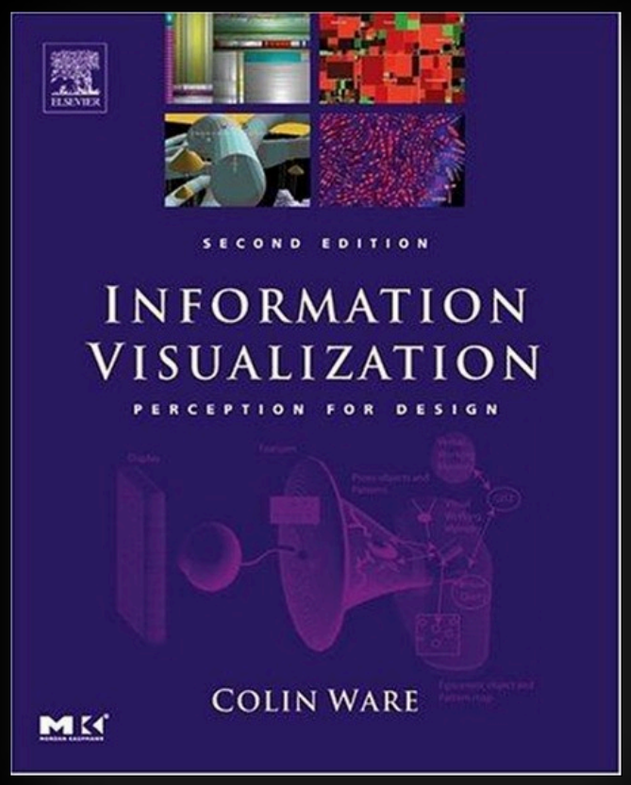
Blogs

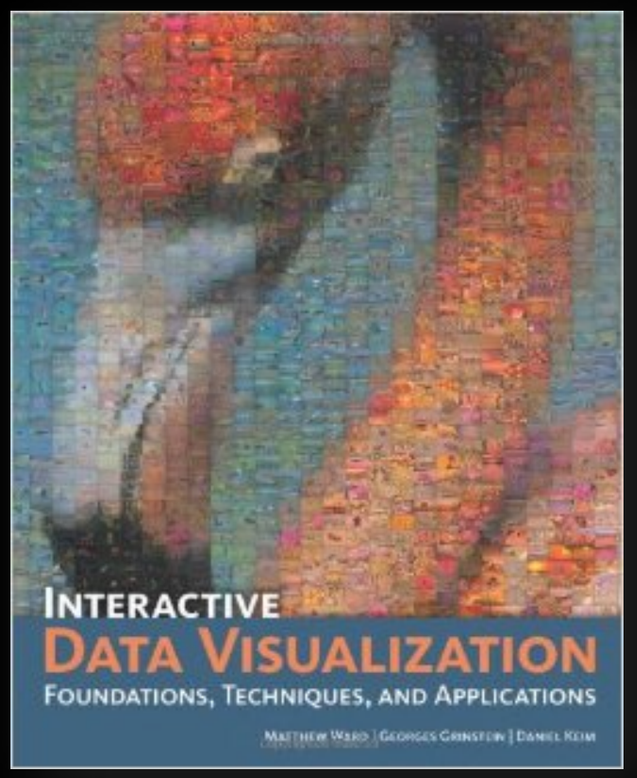
http://flowingdata.com/
http://fellinlovewithdata.com/
http://eagereyes.org/
http://infosthetics.com/

Tufte's collection

Books

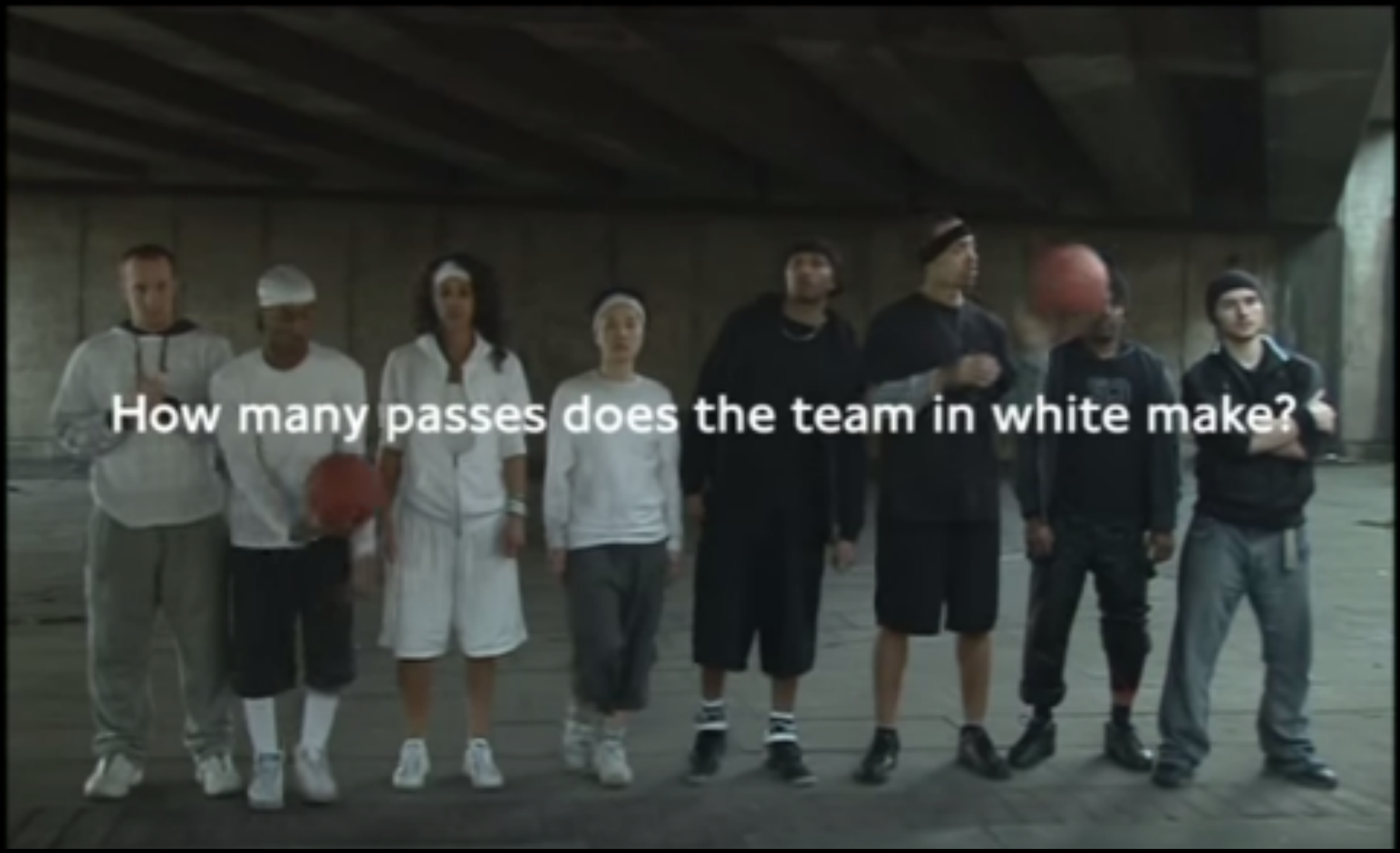


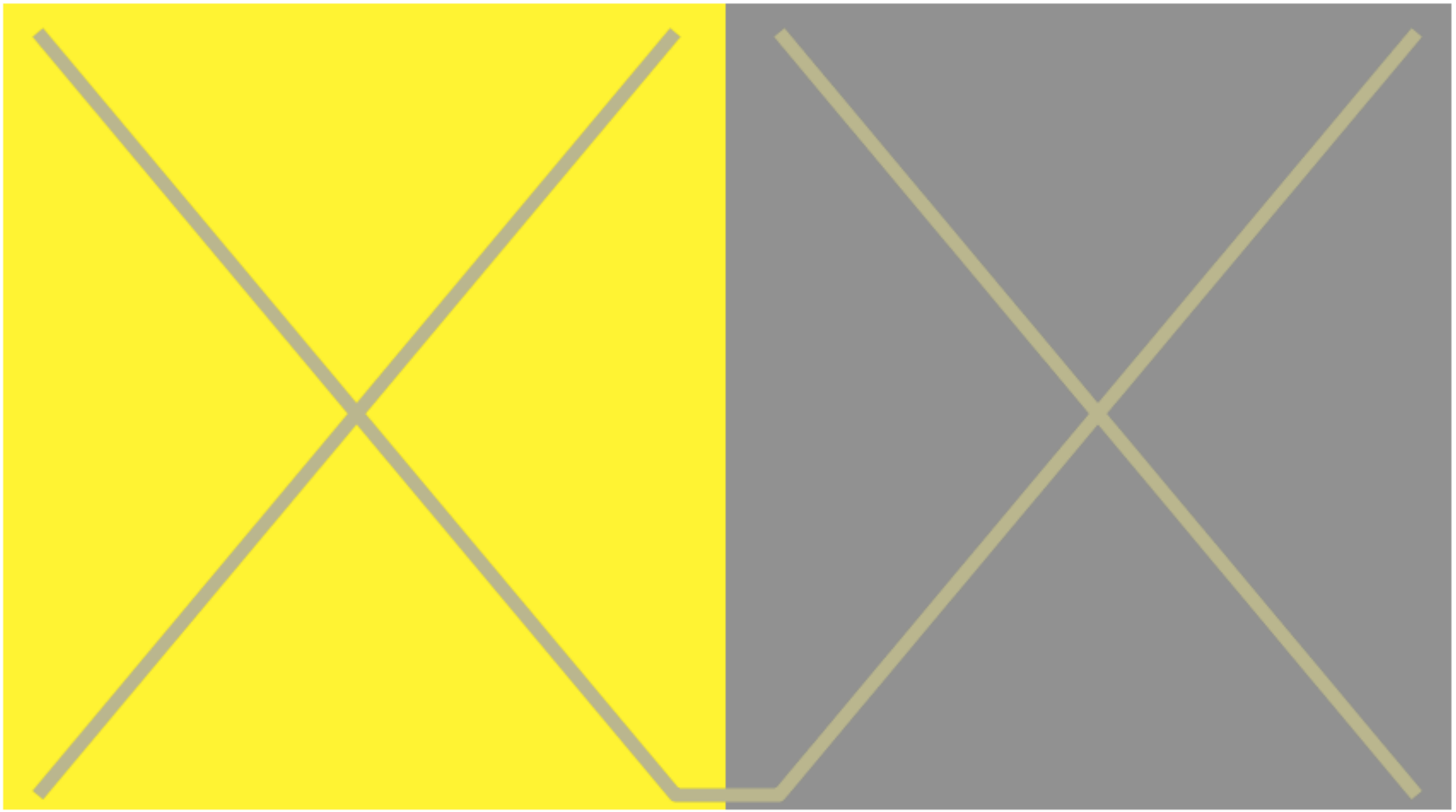




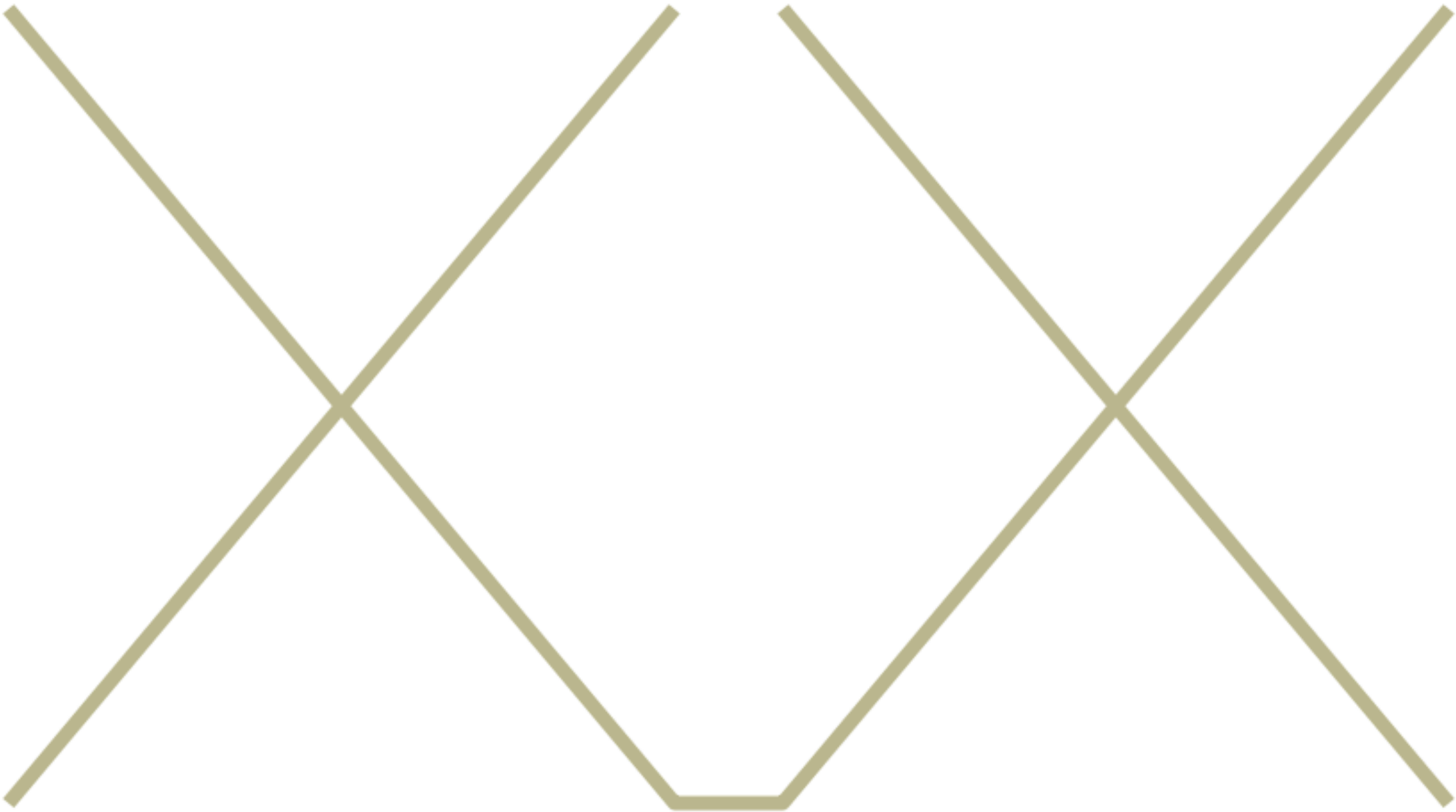
"I ONLY BELIEVE IN WHAT I SEE WITH MY OWN EYES"

BRAIN BUGS

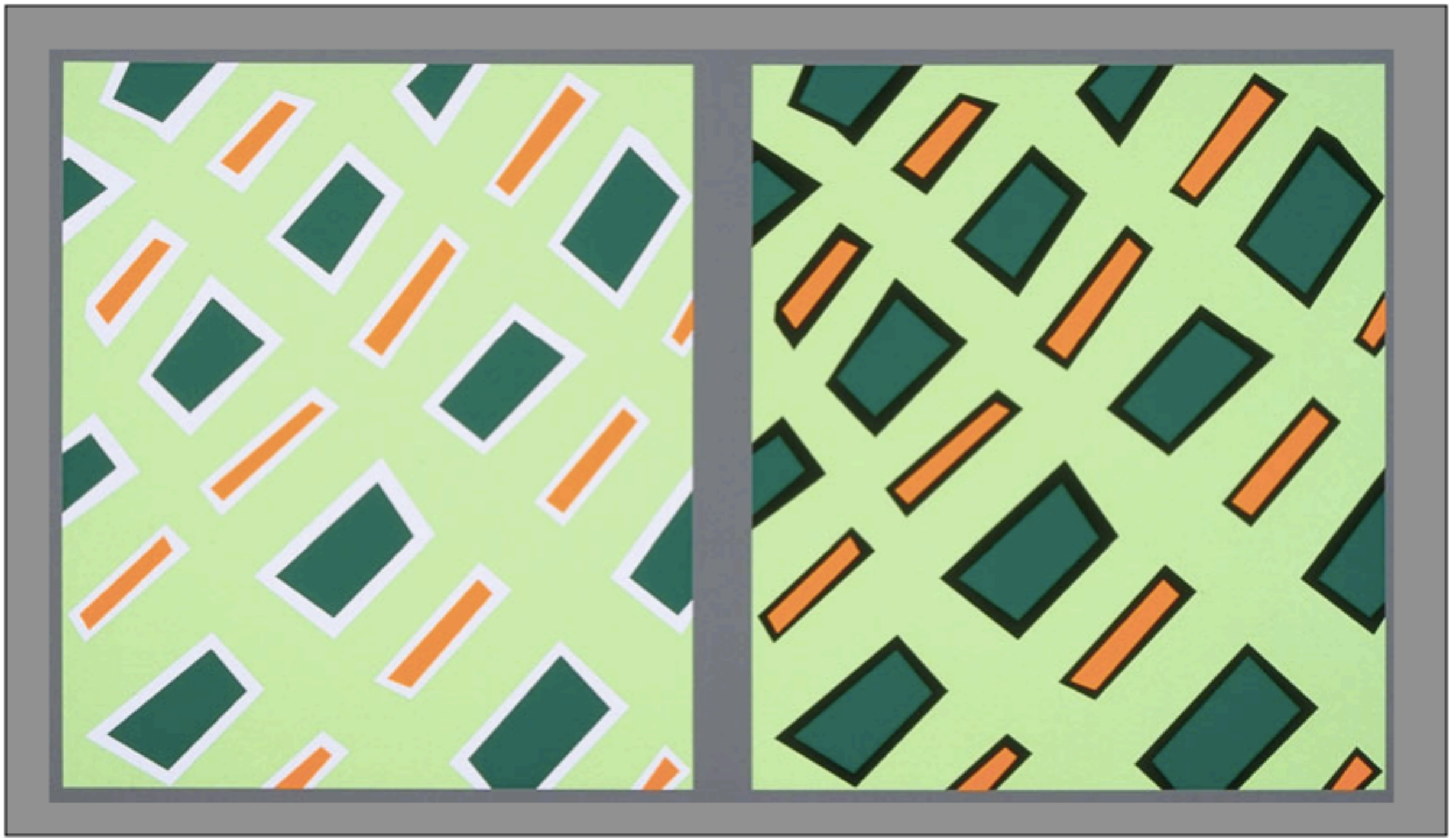

How many passes does the team in white make?

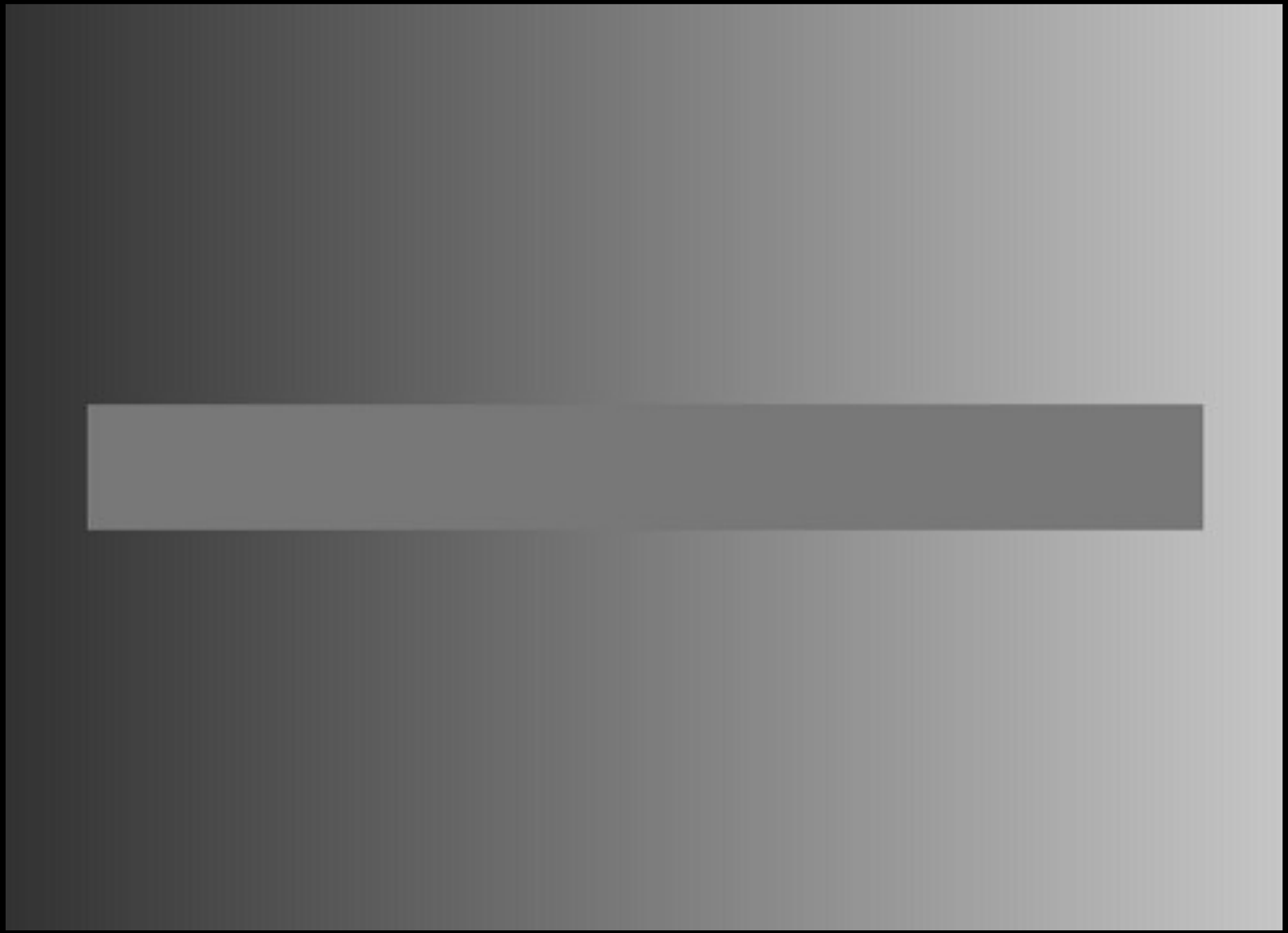


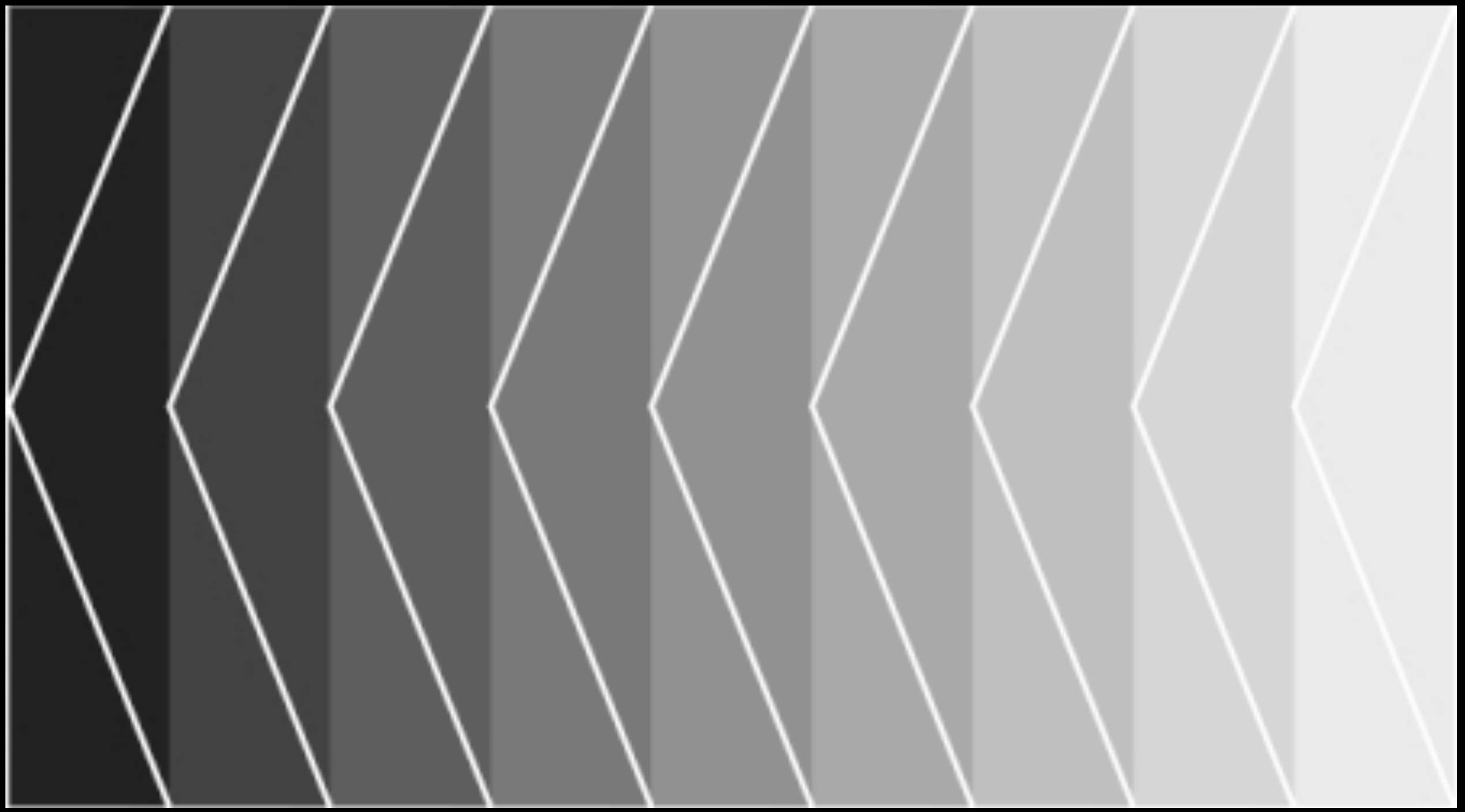
SIMULTANEOUS CONTRAST

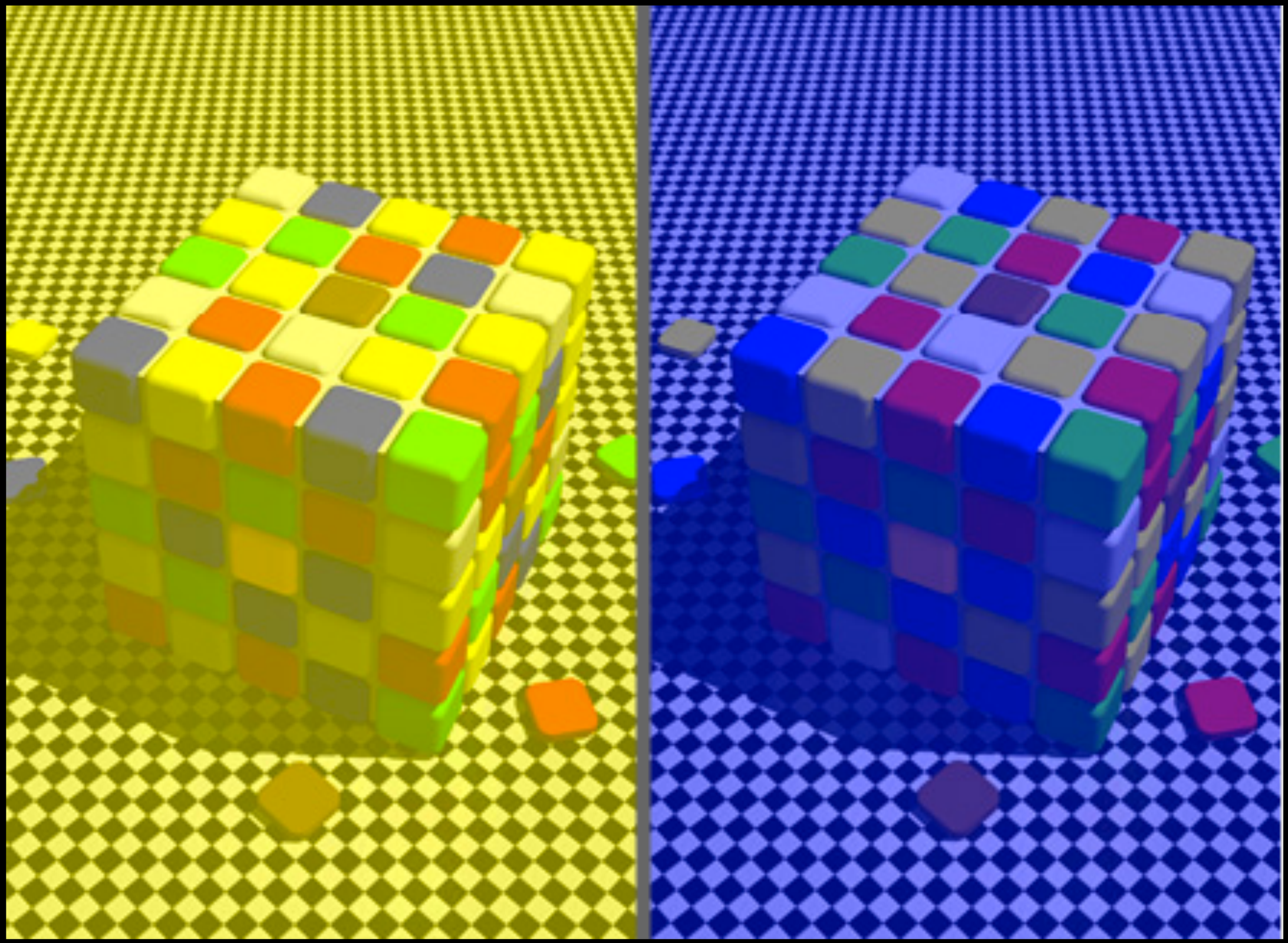


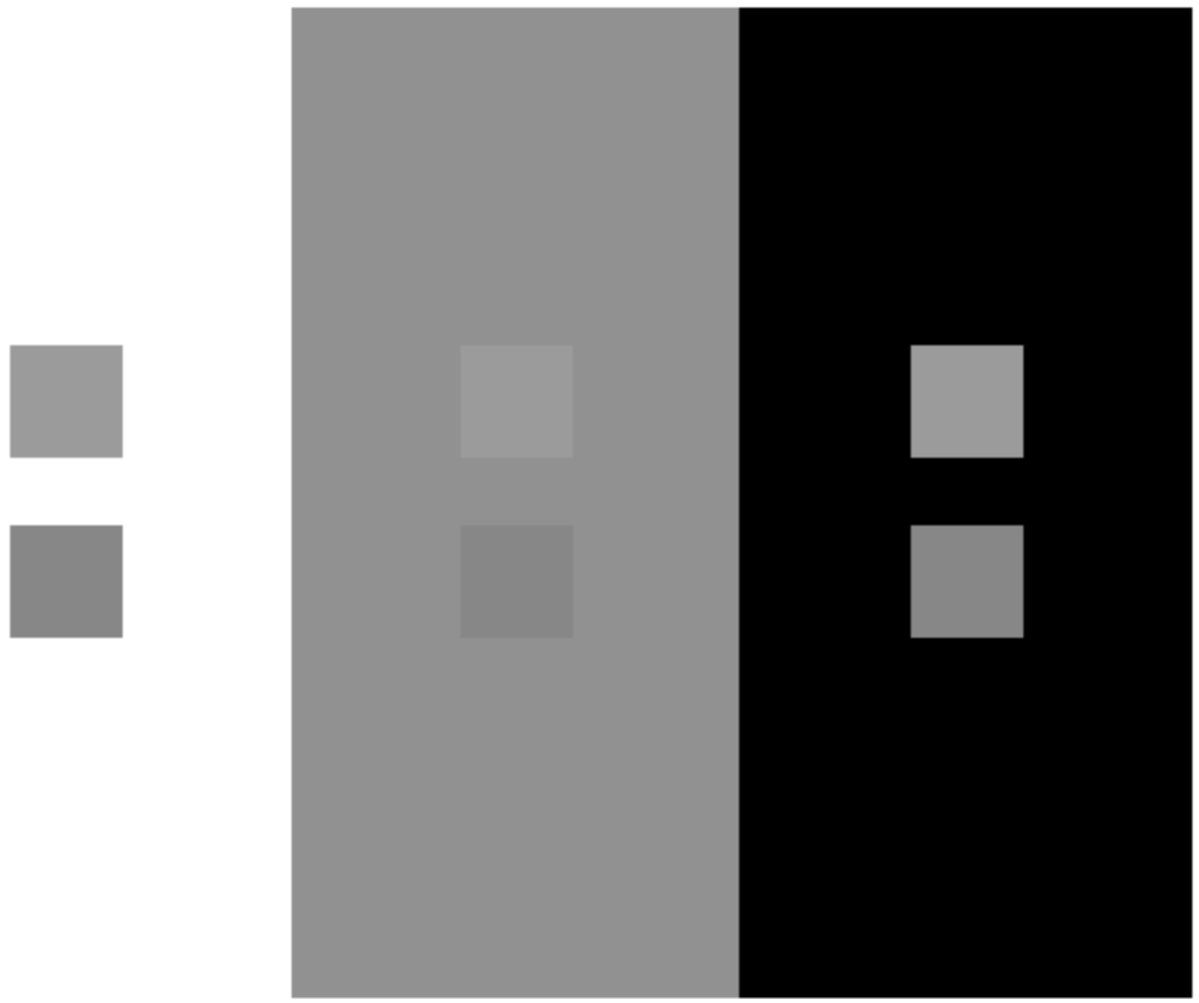
BEZOLD EFFECT



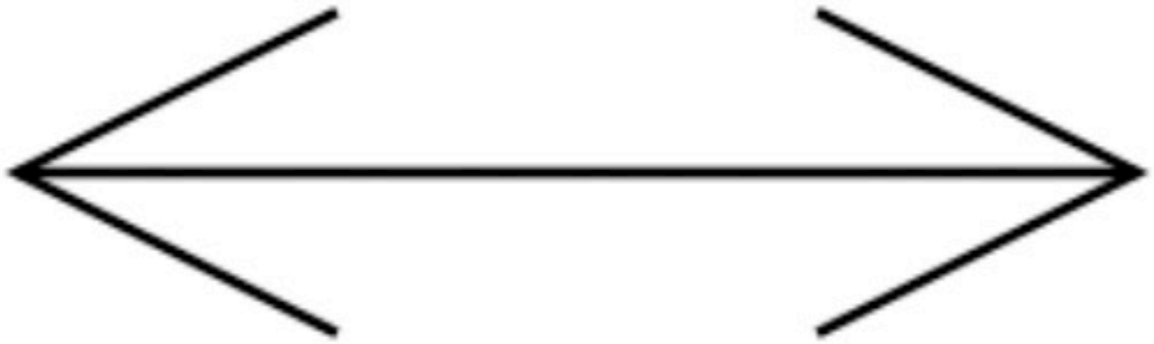


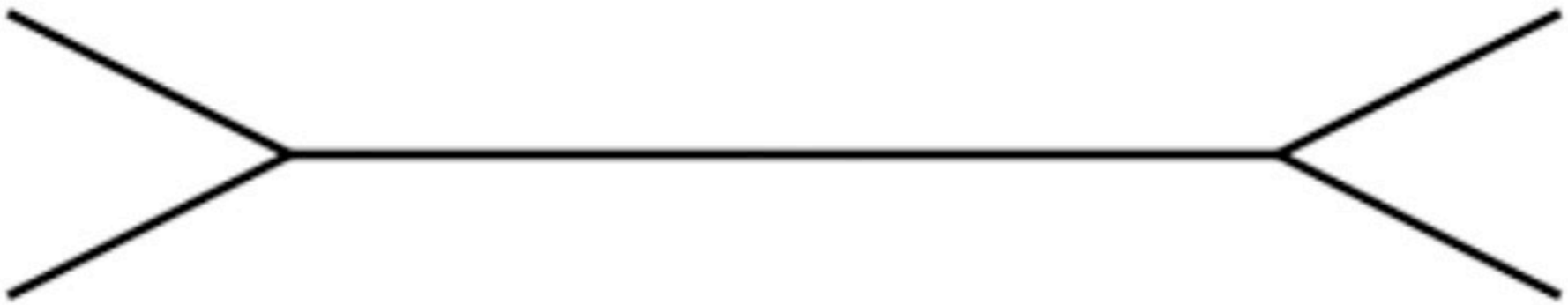


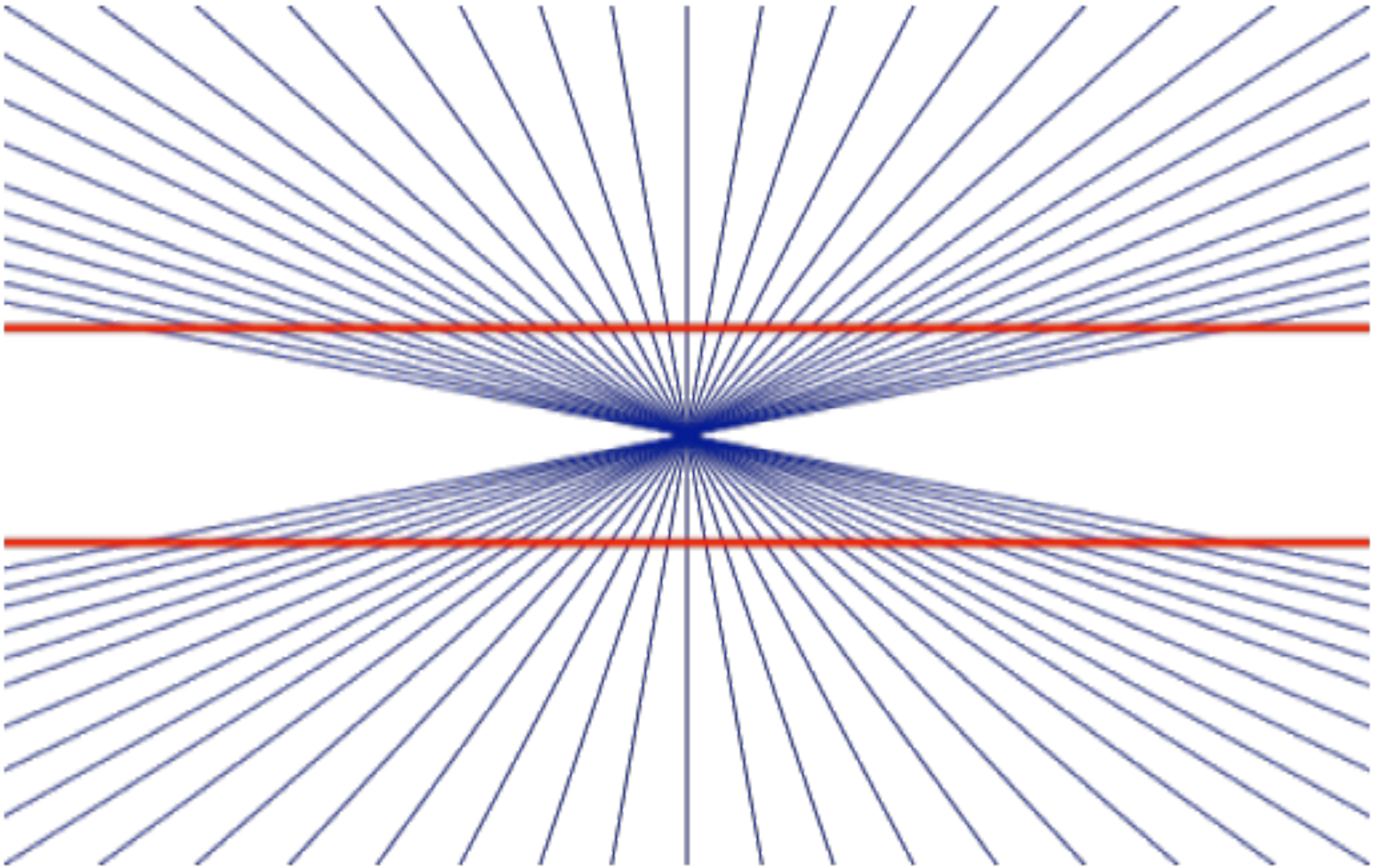


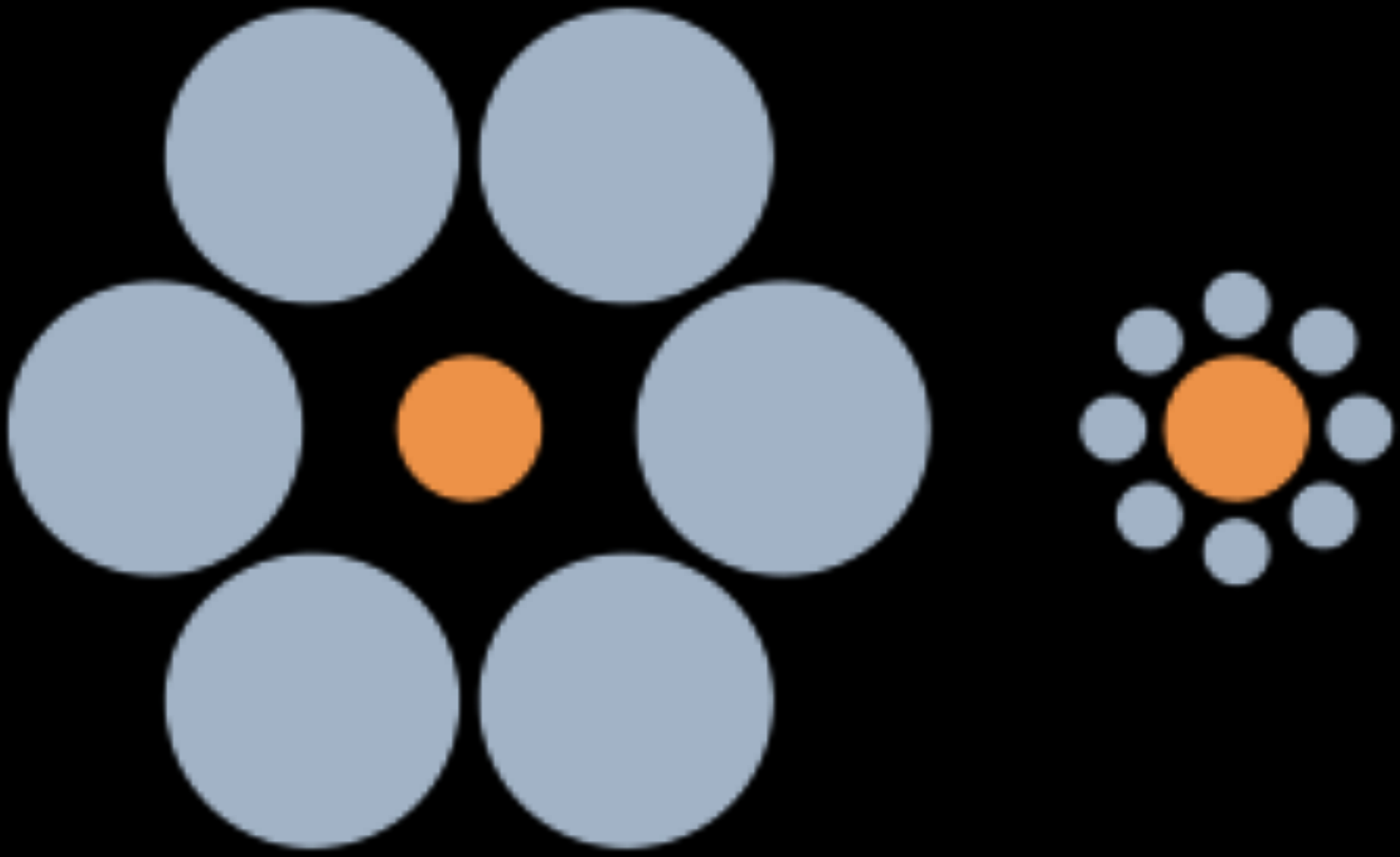


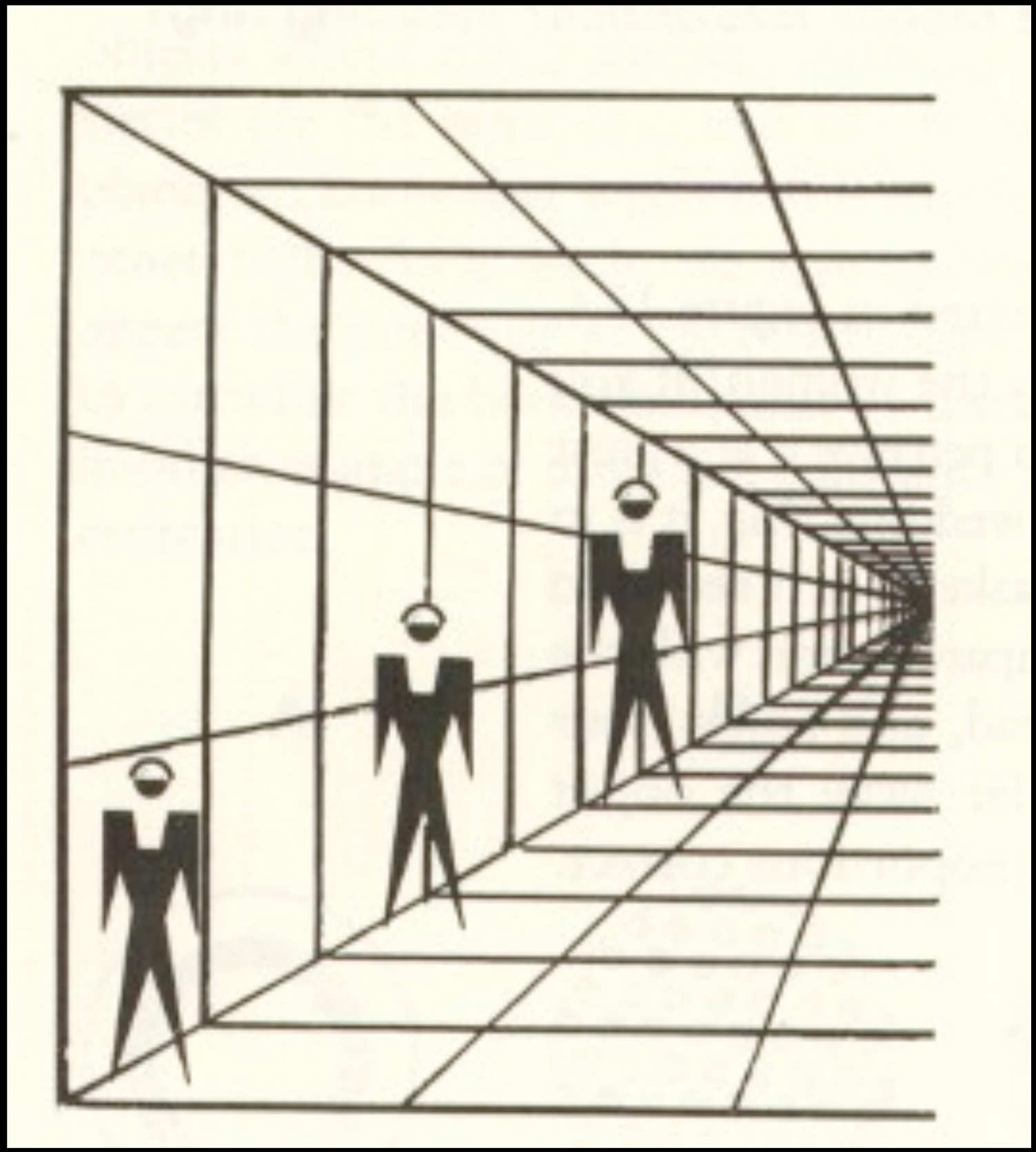
From Fairchild, *Color Appearance Models*

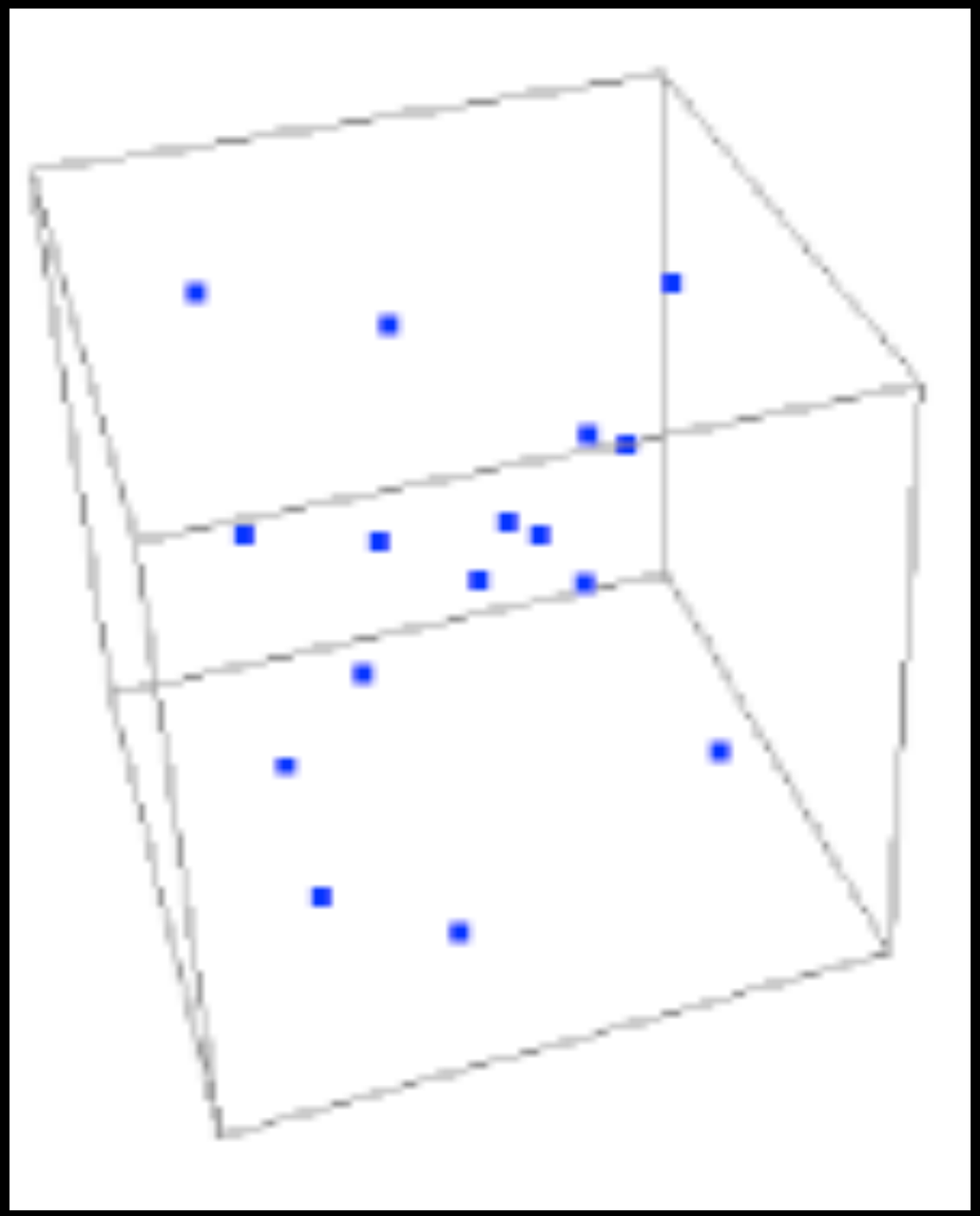


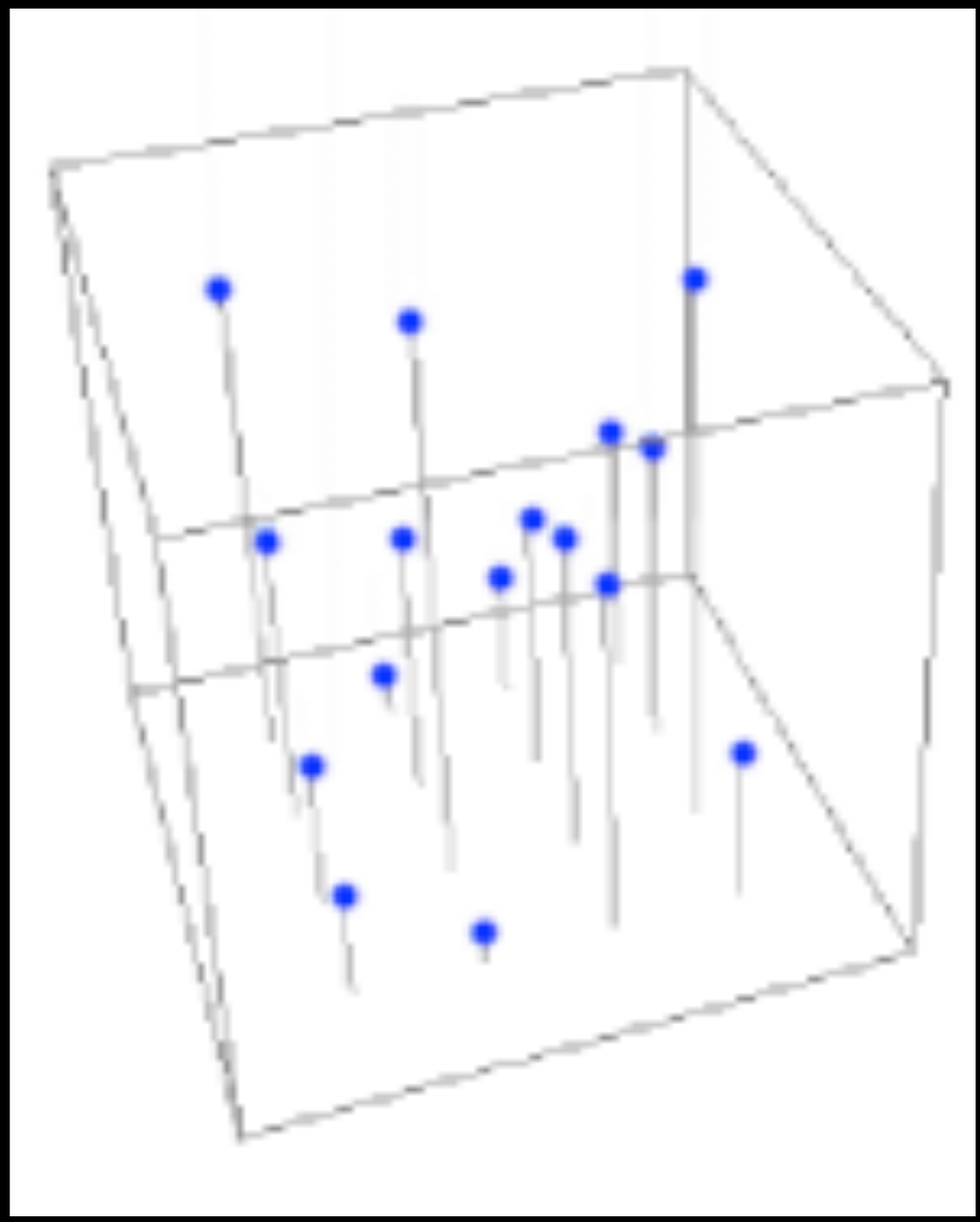


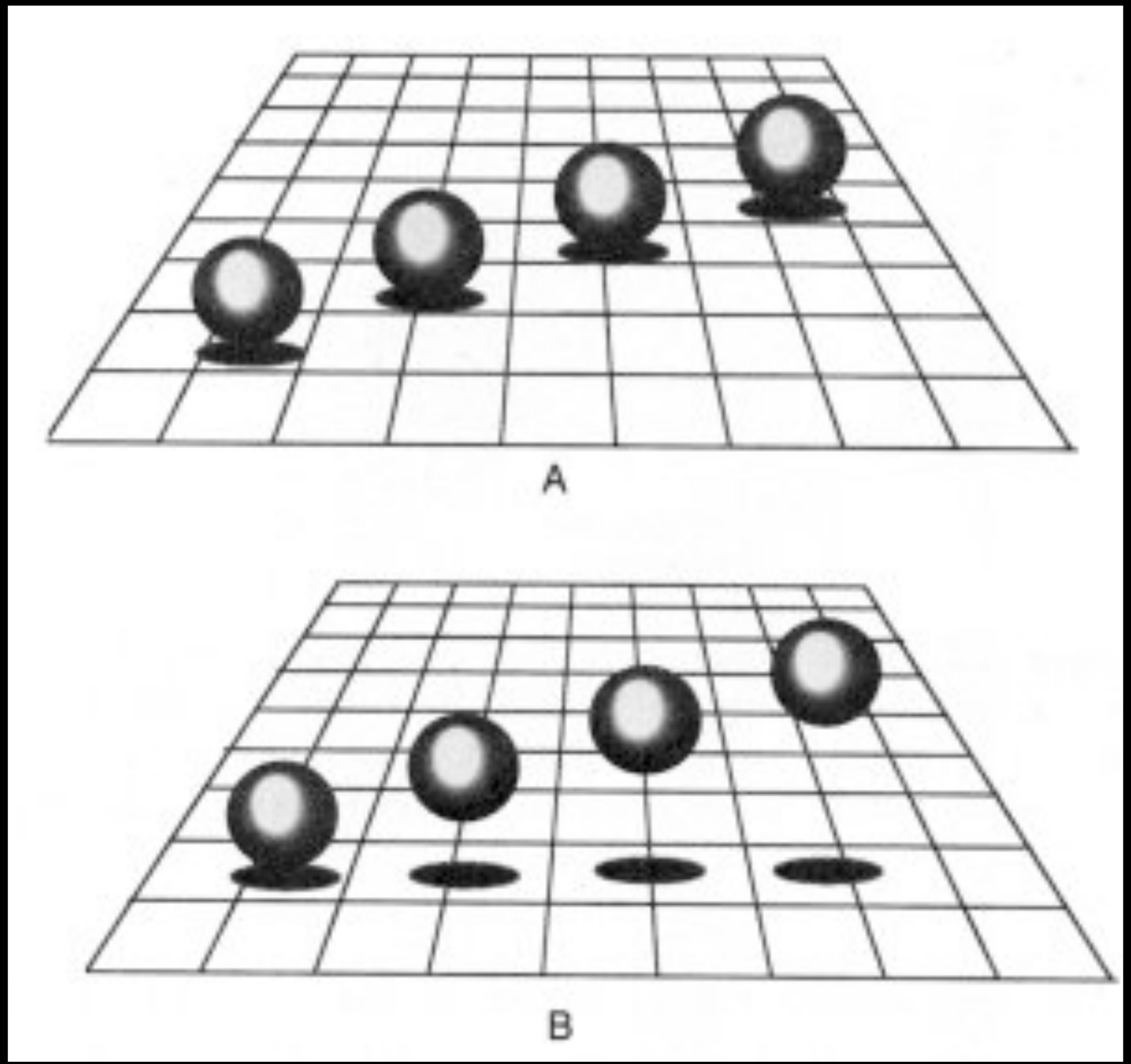




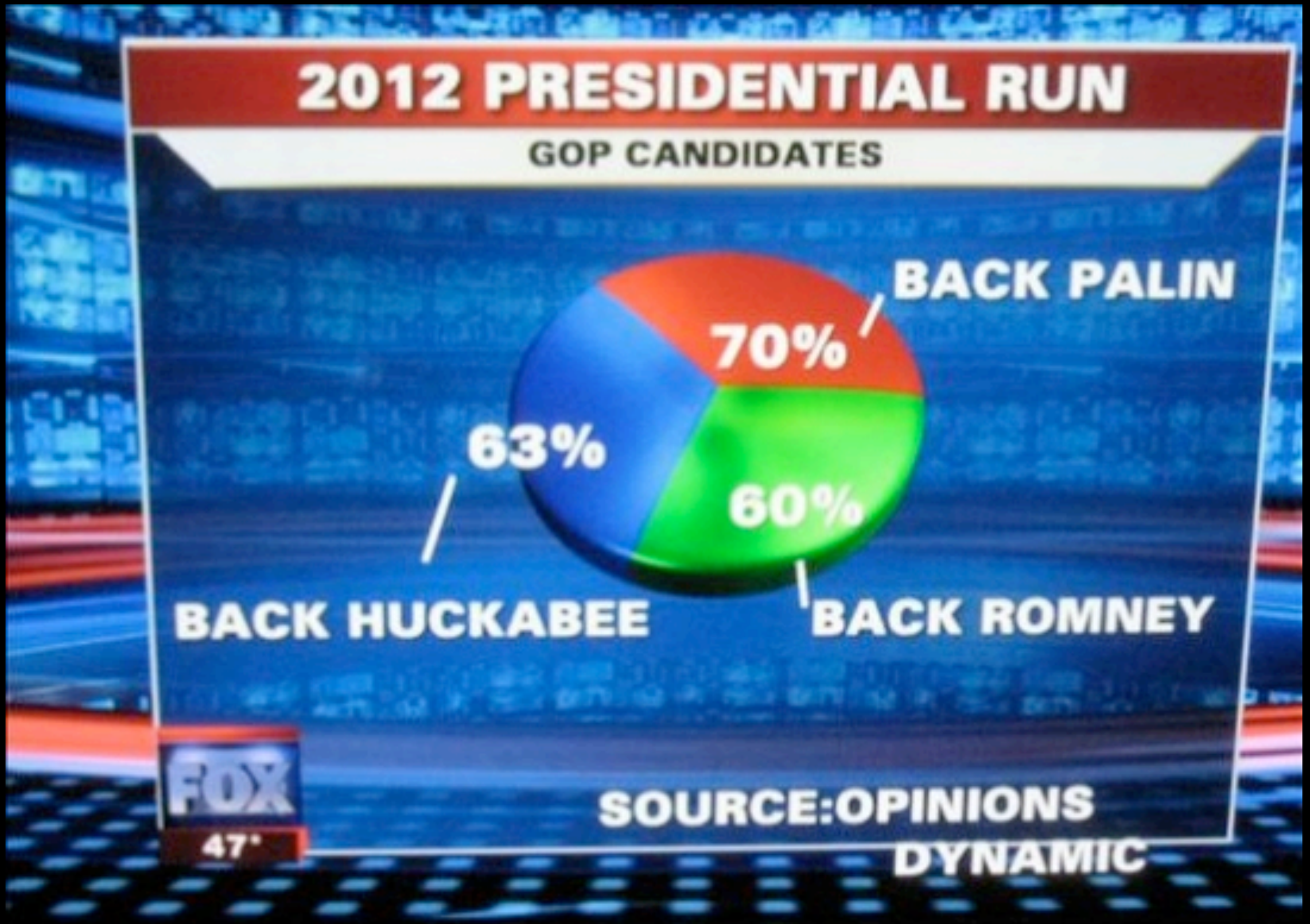


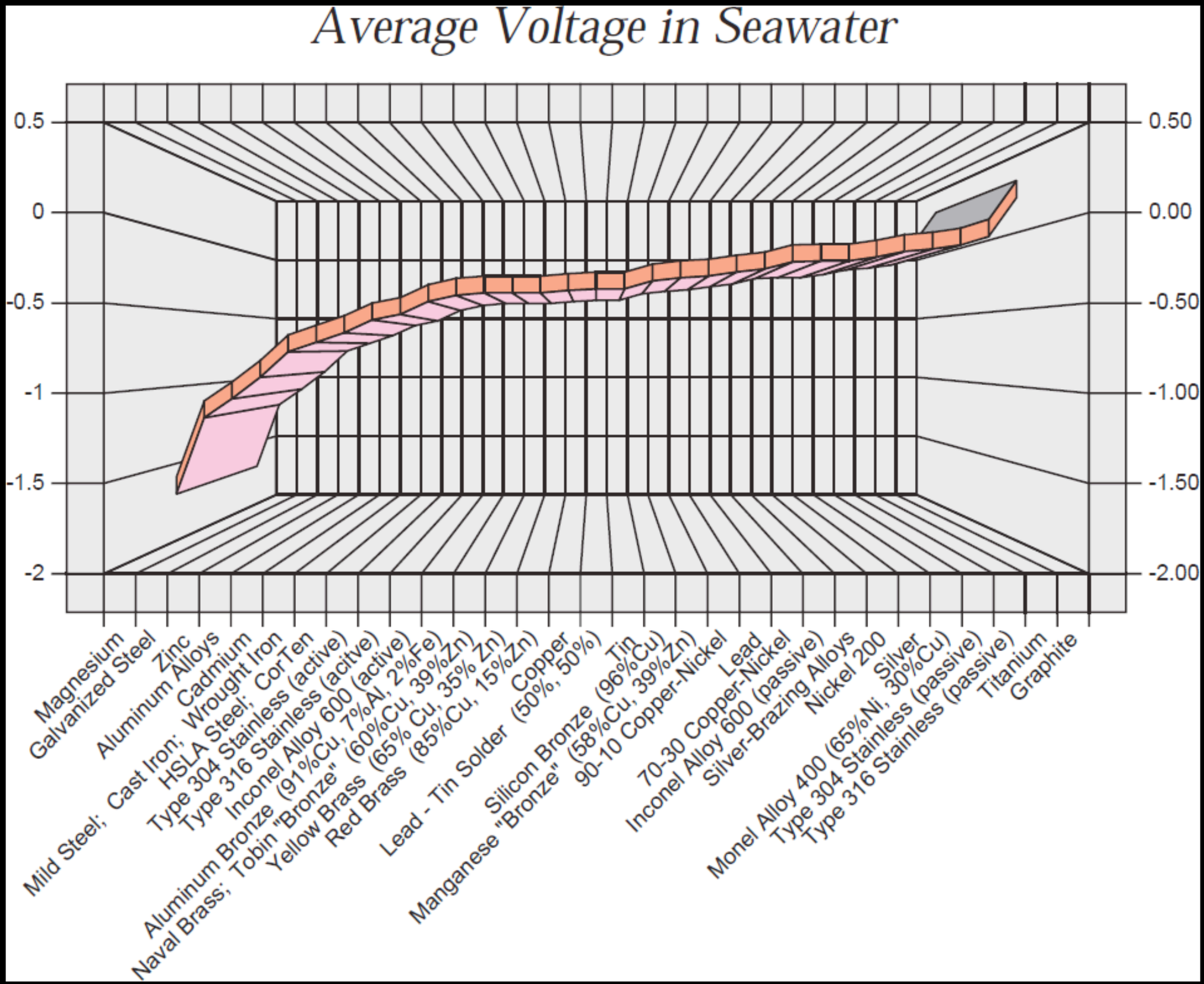


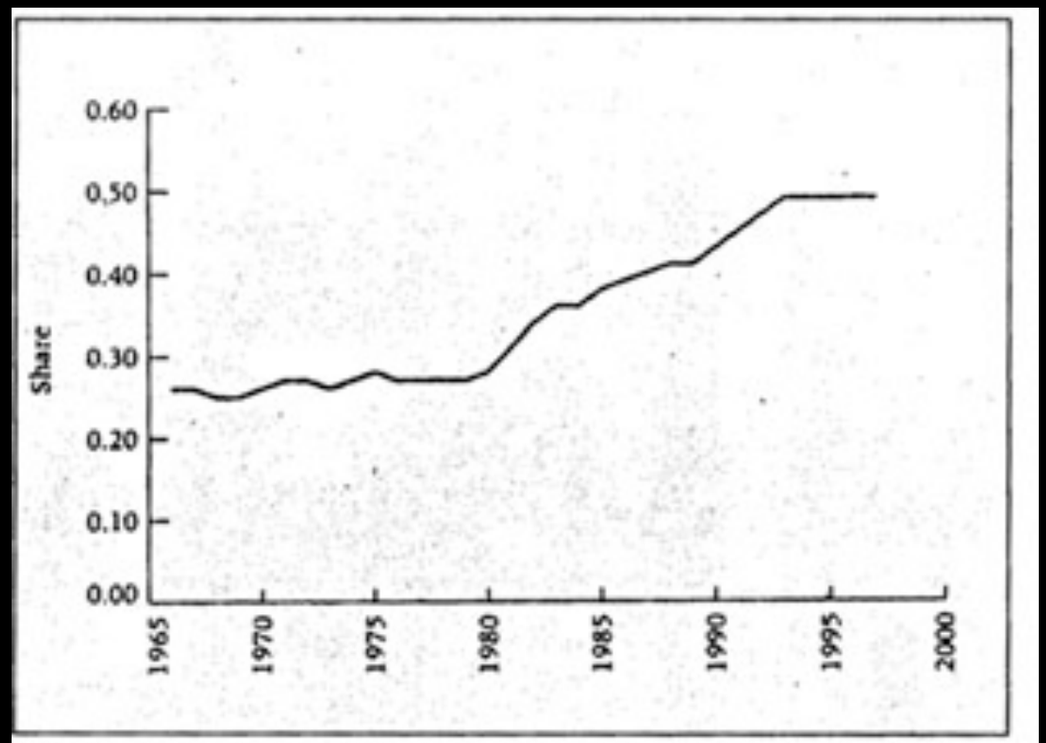



A
B

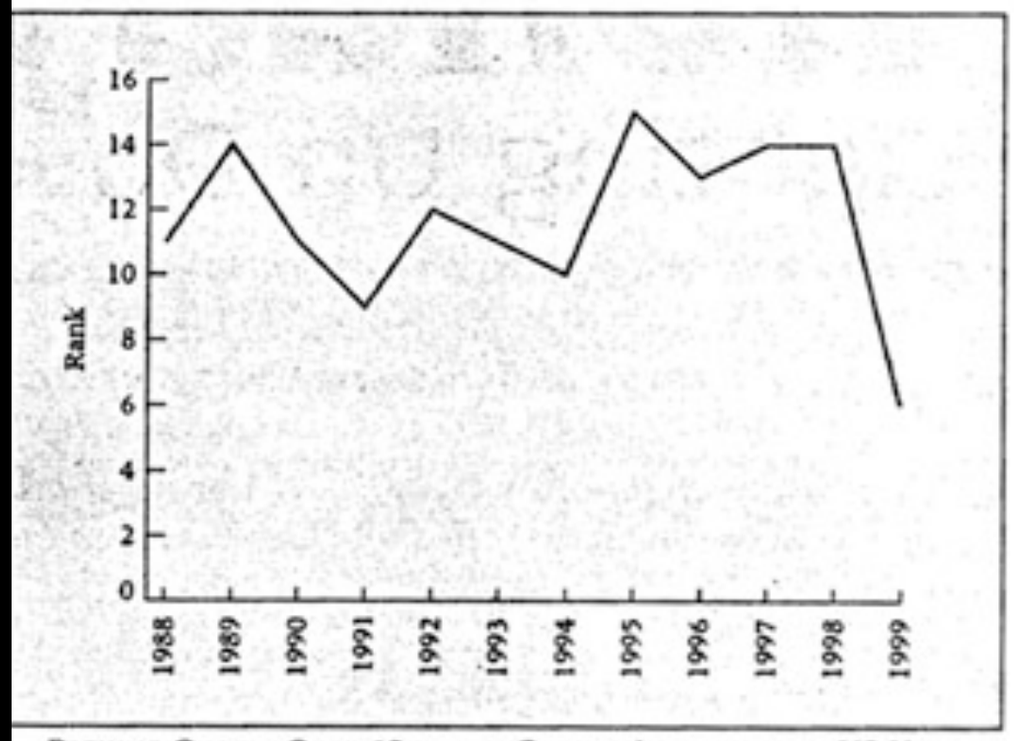
BAD VISUALISATIONS


2012 PRESIDENTIAL RUN
GOP CANDIDATES
70%
BACK PALIN
63%
BACK HUCKABEE
60%
BACK ROMNEY
FOX
47°
SOURCE:OPINIONS
DYNAMIC

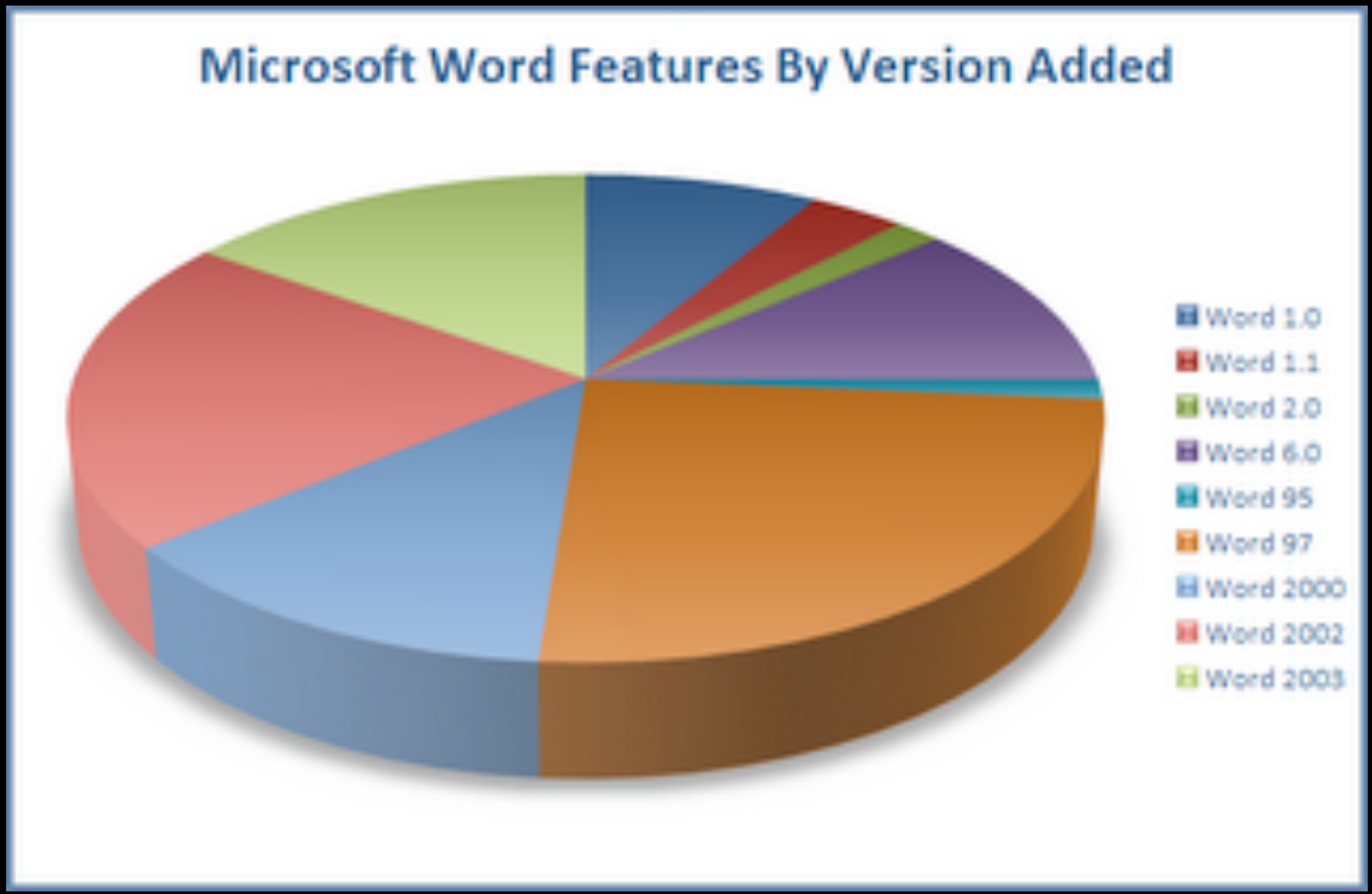

Average Voltage in Seawater
0.5
0
-0.5
-1
-1.5
-2
0.50
0.00
-0.50
-1.00
-1.50
-2.00
Magnesium
Galvanized Steel
Zinc
Aluminum Alloys
Cadmium
Wrought Iron
Mild Steel; Cast Iron; CorTen
HSLA Steel;
Type 304 Stainless (active)
Type 316 Stainless (acitve)
Inconel Alloy 600 (active)
Aluminum Bronze (91%Cu, 7%Al, 2%Fe)
Naval Brass; Tobin "Bronze" (60%Cu, 39%Zn)
Yellow Brass (65% Cu, 35% Zn)
Red Brass (85%Cu, 15%Zn)
Copper
Lead - Tin Solder (50%, 50%)
Tin
Silicon Bronze (96%Cu)
Manganese "Bronze" (58%Cu, 39%Zn)
90-10 Copper-Nickel
Lead
70-30 Copper-Nickel
Inconel Alloy 600 (passive)
Silver-Brazing Alloys
Nickel 200
Silver
Monel Alloy 400 (65%Ni, 30%Cu)
Type 304 Stainless (passive)
Type 316 Stainless (passive)
Titanium
Graphite

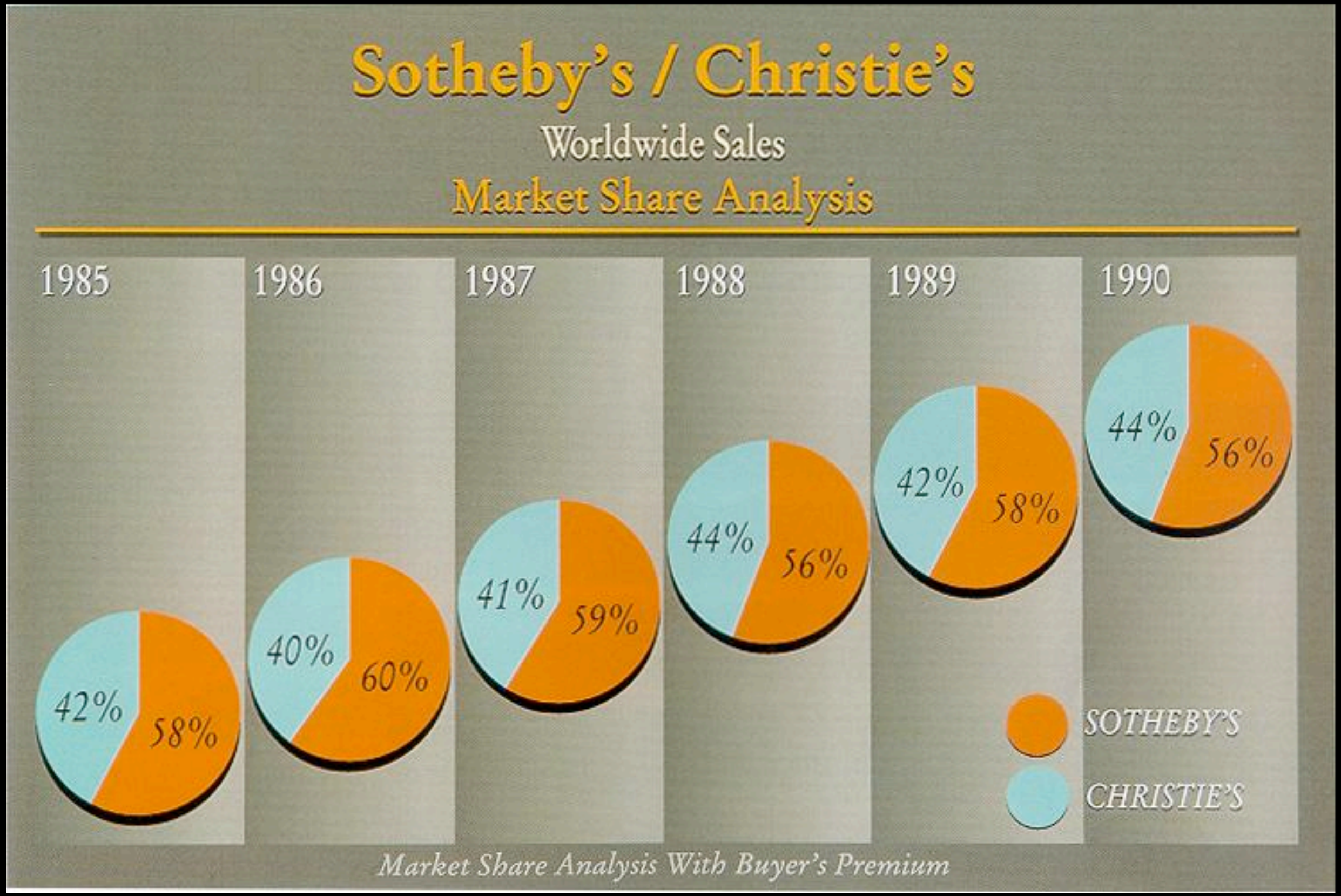


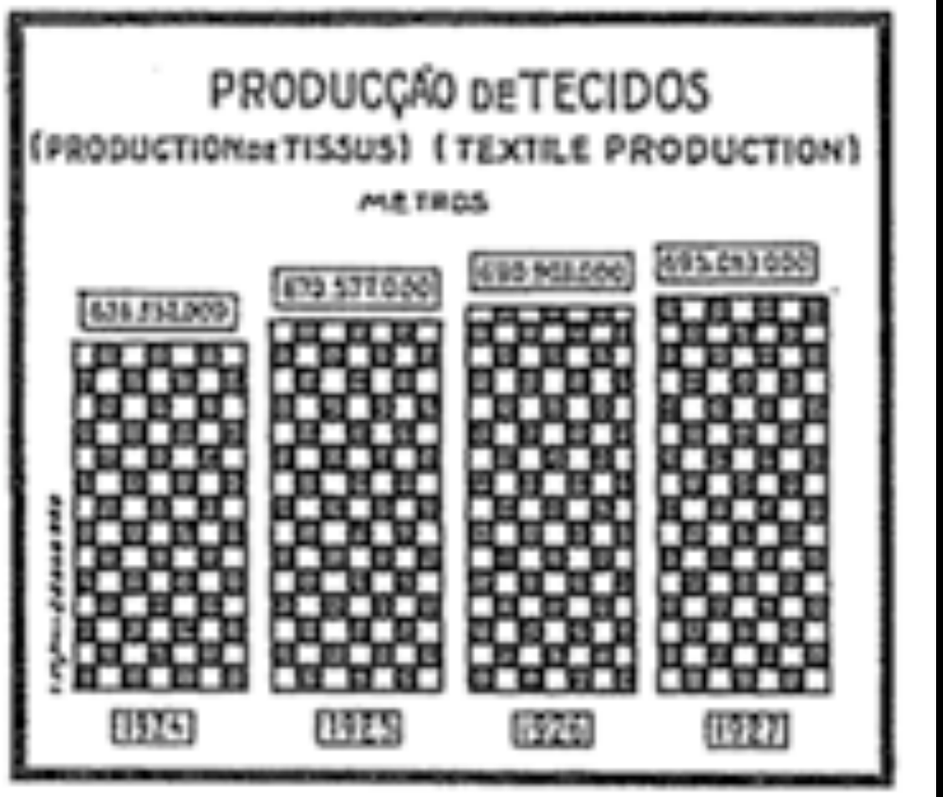
BY THE NUMBERS: OVER 35 YEARS, CORNELL'S TUITION HAS TAKEN AN INCREASINGLY LARGER SHARE OF ITS MEDIAN STUDENT FAMILY INCOME.

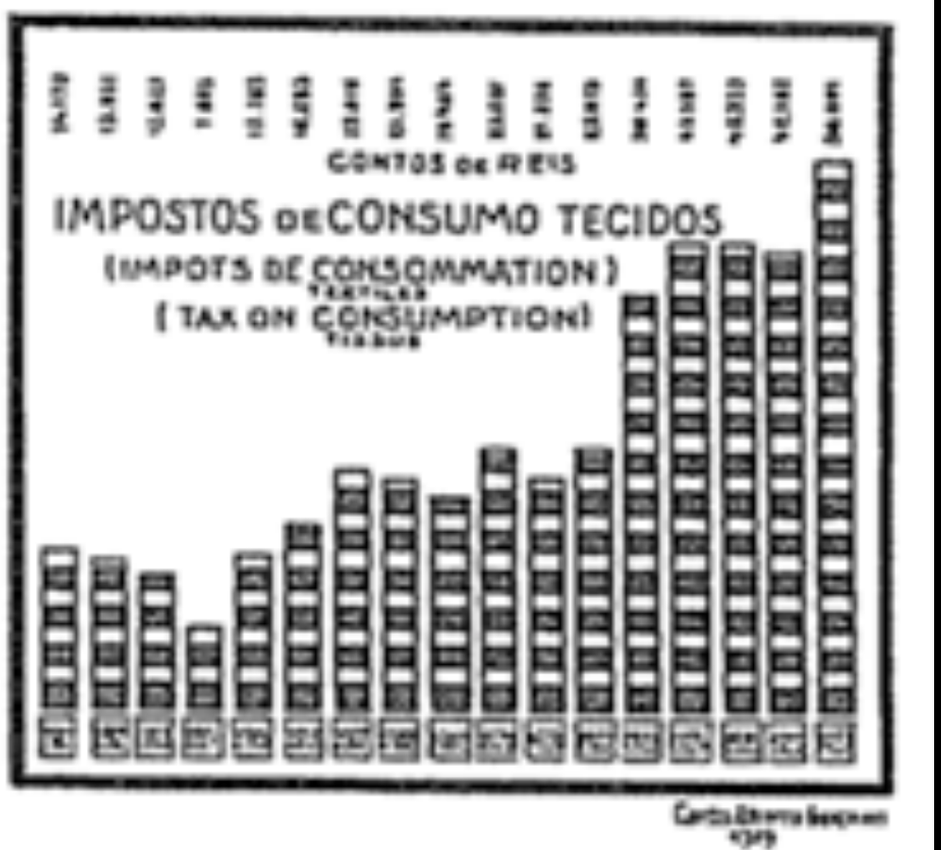


PECKING ORDER: OVER 12 YEARS, CORNELL'S RANKING IN *US NEWS & WORLD REPORT* HAS RISEN AND FALLEN ERRATICALLY.


Microsoft Word Features By Version Added
Word 1.0
Word 1.1
Word 2.0
Word 6.0
Word 95
Word 97
Word 2000
Word 2002
Word 2003

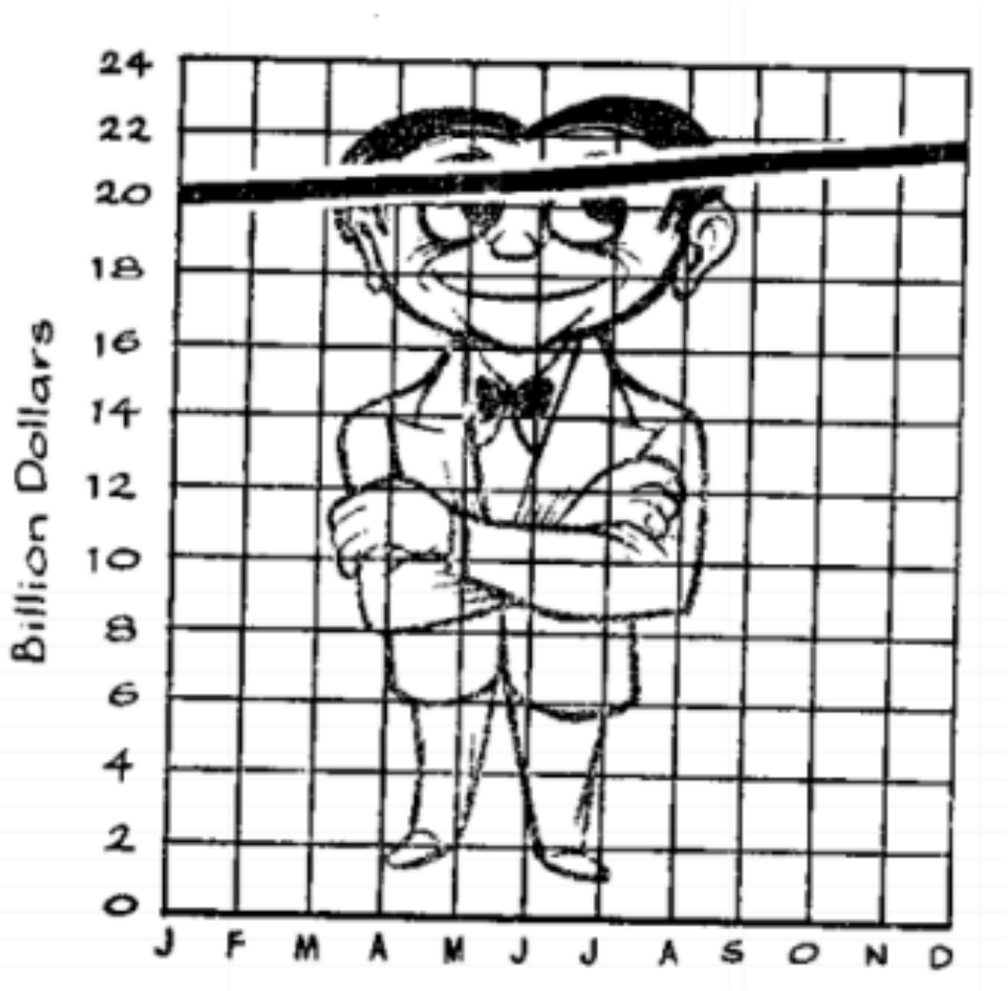

Sotheby's / Christie's
Worldwide Sales
Market Share Analysis
1985
42%
58%
1986
40%
60%
1987
41%
59%
1988
44%
56%
1989
42%
58%
1990
44%
56%
SOTHEBY'S
CHRISTIE'S
Market Share Analysis With Buyer's Premium


PRODUCÇÃO DE TECIDOS
(PRODUCTION DE TISSUS) (TEXTILE PRODUCTION)
METROS
1924
1925
1926
1927


CONTOS DE RÉIS
IMPOSTOS DE CONSUMO TECIDOS
(IMPOTS DE CONSOMMATION)
(TAX ON CONSUMPTION)

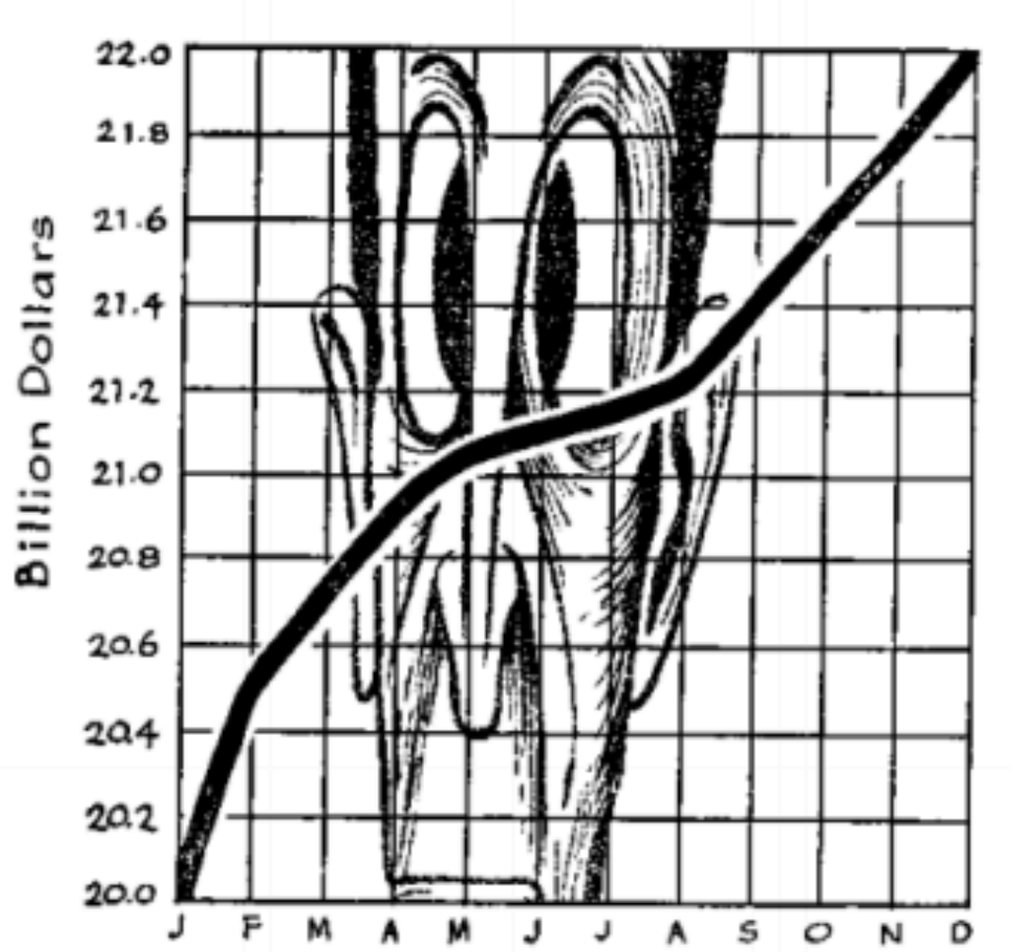
VISUALIZATION RULES

- Provide a proper baseline



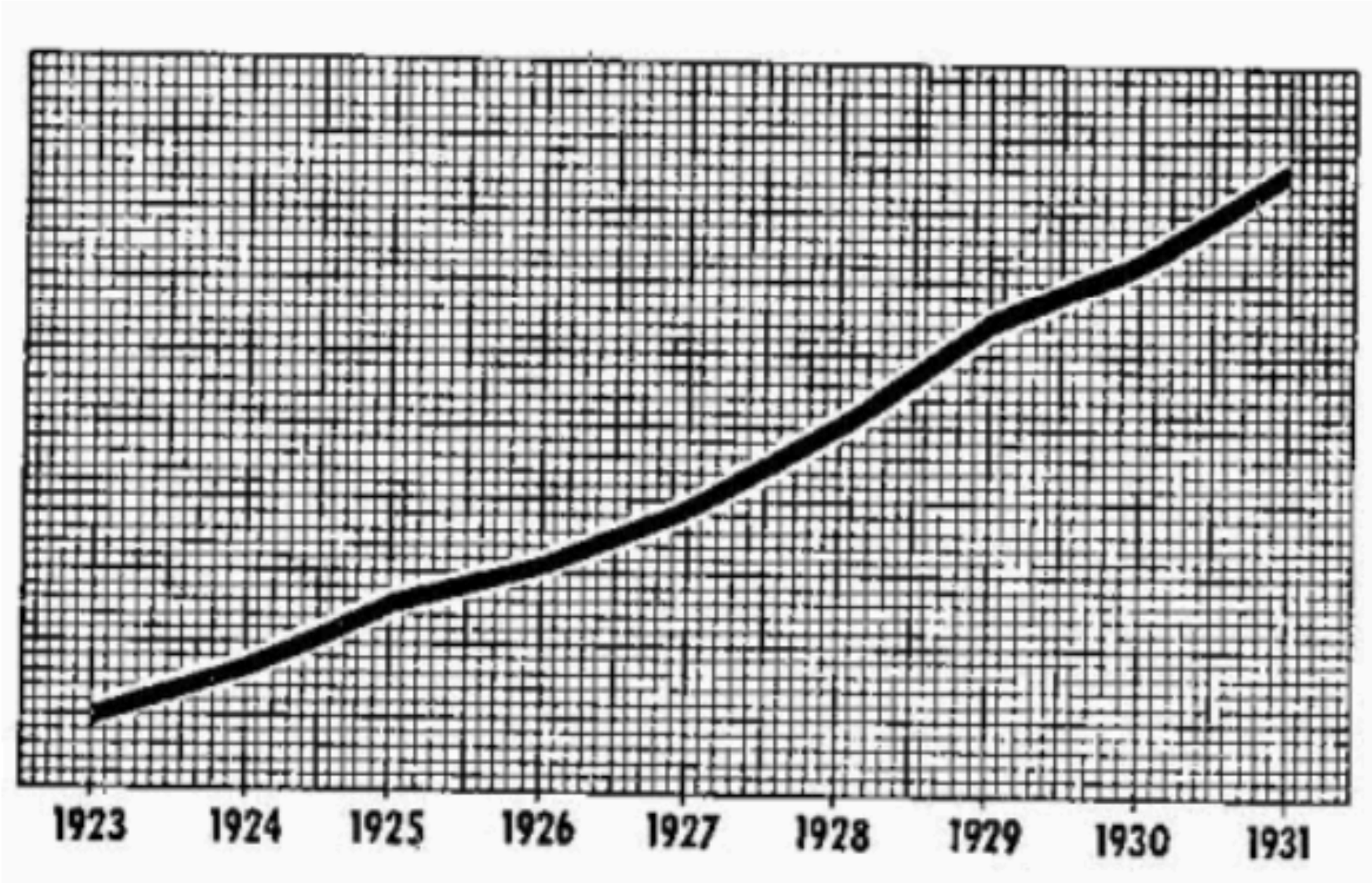
A 10% increase. Good!

Already looks more impressive

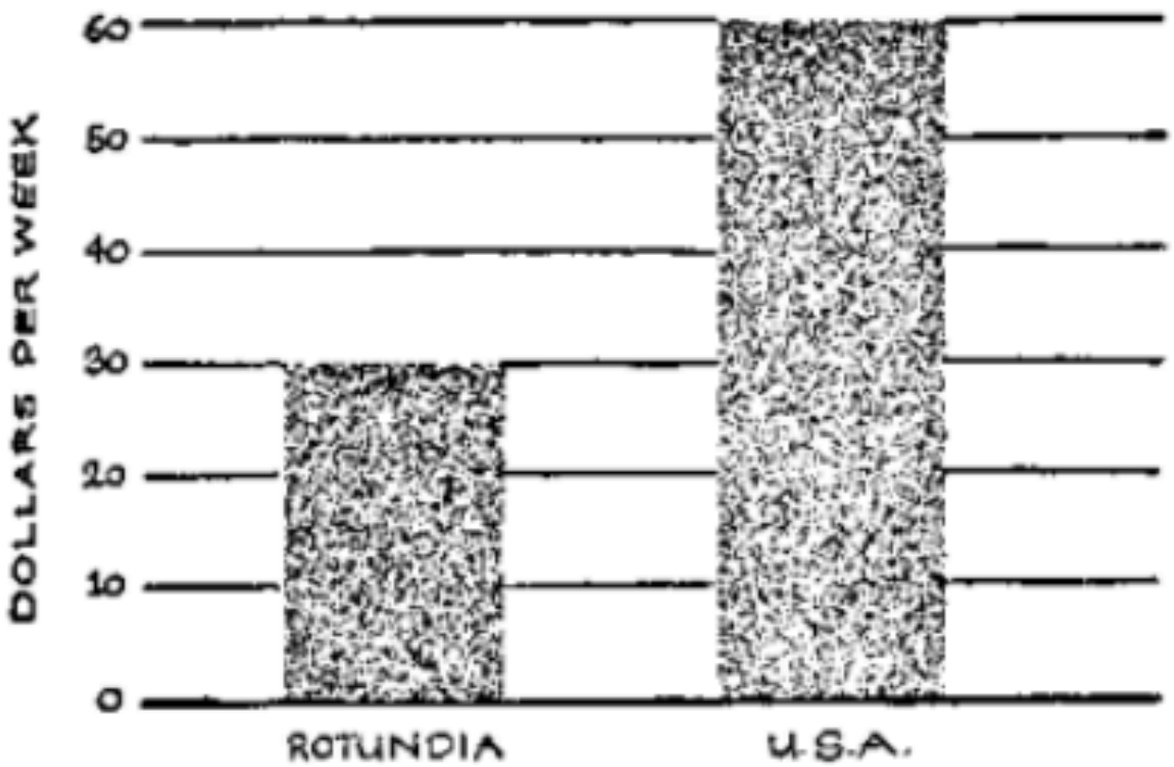


Wow!

- Provide a **proper baseline** & **label your axes**



- Provide a **proper baseline** & **label your axes**
- Avoid **eye-candy**

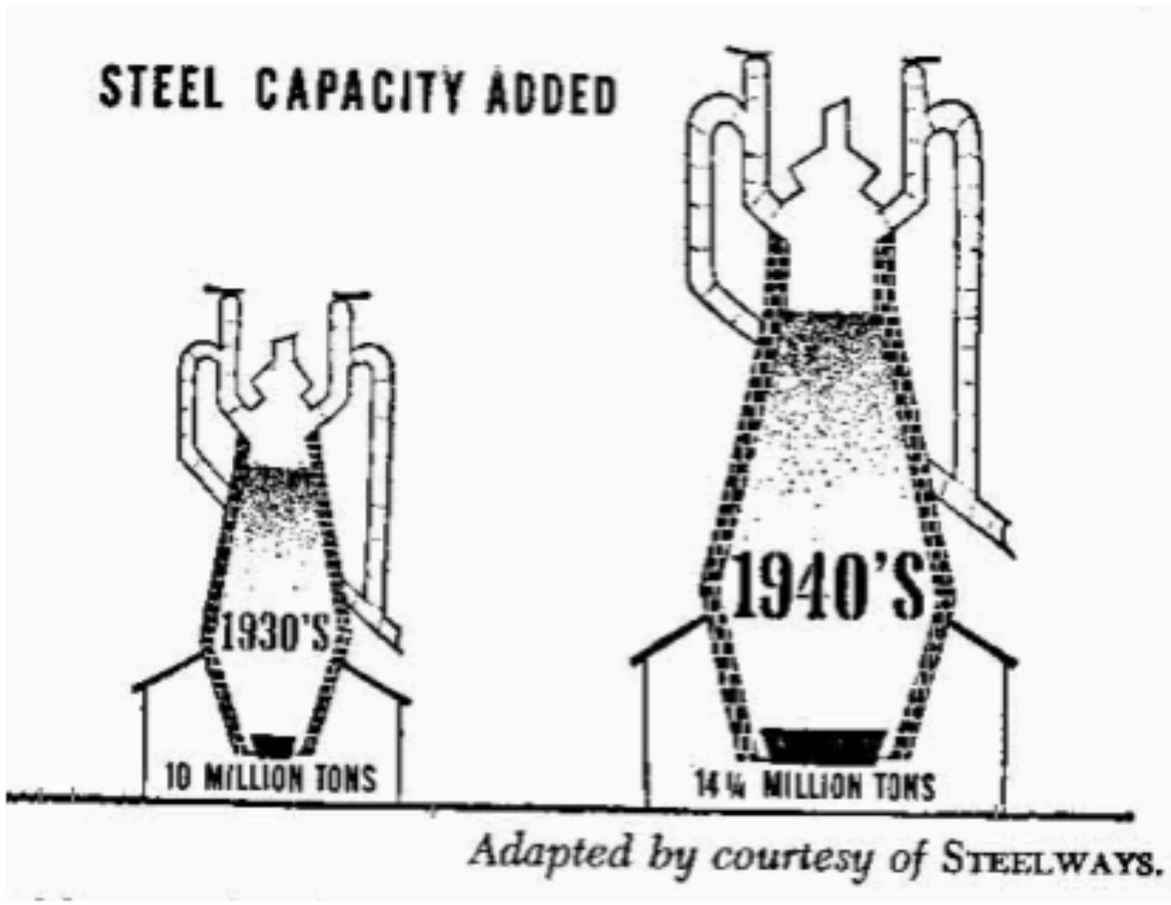


Actual data

The same data with eye-candy & no numbers … but at least it tells the same general story.

Impressive, but a lie!

- Provide a **proper baseline** & **label your axes**
- Avoid **eye-candy**
- Avoid **area comparisons** whenever possible



Adapted by courtesy of STEELWAYS.