DSCI 554 LECTURE 1

COURSE OVERVIEW, INTRODUCTION TO DATA VISUALIZATION

Dr. Luciano Nocera

USC Viterbi
School of Engineering
Integrated Media Systems Center

OUTLINE

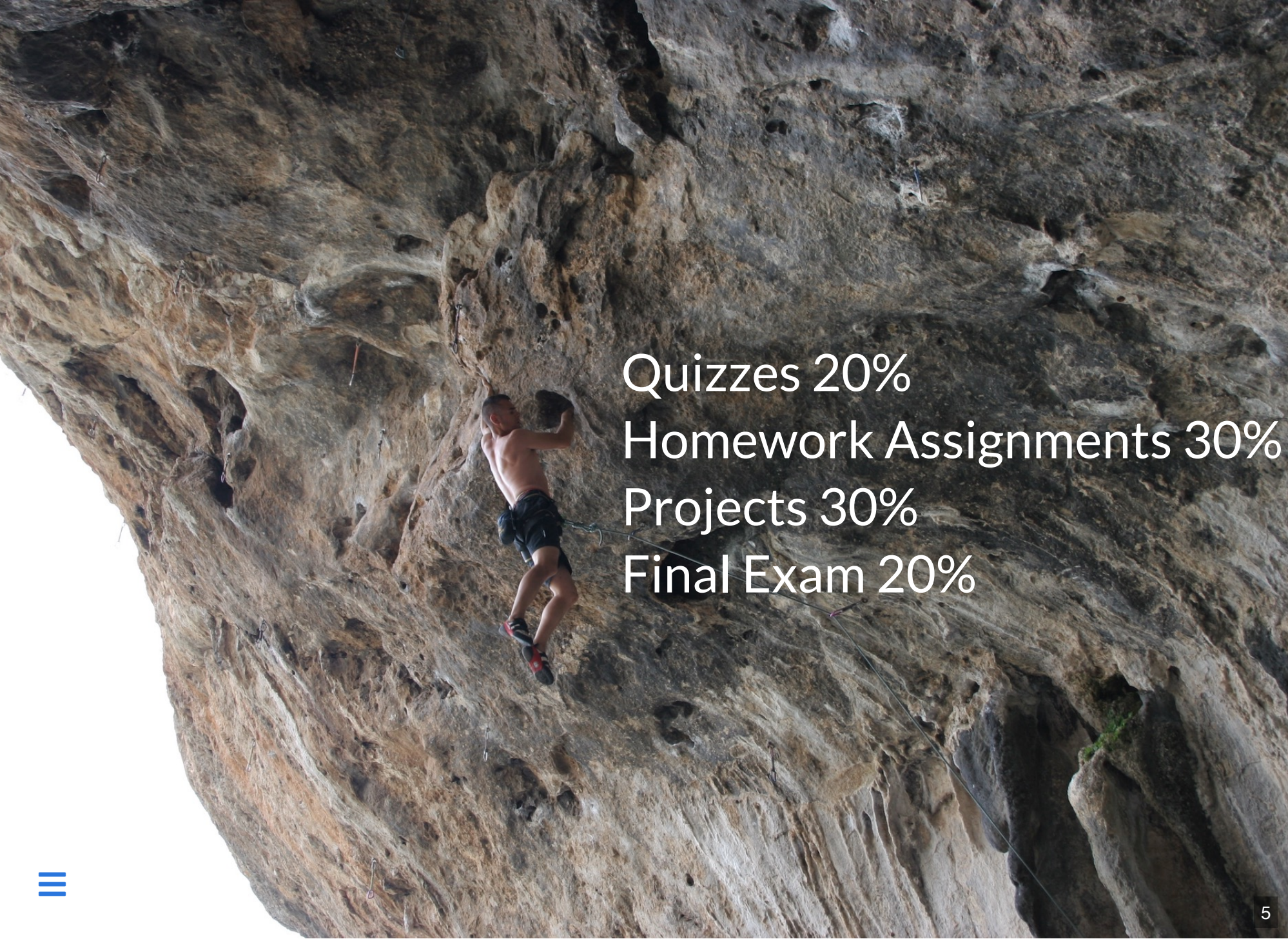
- Course information
- Data visualization
- Examples and uses
- Tools and software

COURSE OBJECTIVE

- Learn design principles and guidelines
- Learn to critique and evaluate visualizations
- Understand visualization tools and techniques
- Learn how to create interactive visualizations

COURSE MATERIALS

http://pdms.usc.edu/dsci-554


Quizzes 20%
Homework Assignments 30%
Projects 30%
Final Exam 20%

QUIZZES

- 20% of grade
- 20min
- On previous class content & readings
- The worst quiz score will not count. No retake!
- MCQ
- Some coding questions

HOMEWORK

- 30% of grade
- 1-4 hours to complete
- One week to complete
- In GitHub **using your USC email!**
- See rubric in starter repositories for more details

FINAL EXAM

- 20% of grade
- Cumulative
- Quiz questions
- According to University schedule: https://classes.usc.edu/term-20213/finals/

> *Classes meeting 2 or 2:30 MWF Friday, December 10, 2-4 p.m.*

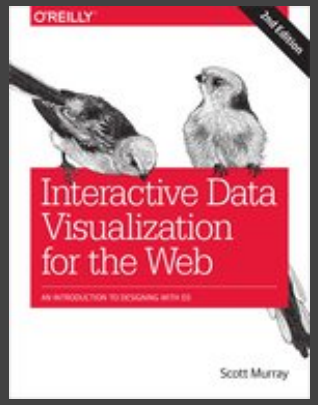
PROJECTS

- 30% of grade
- Groups of 3-4: start to form groups!
- Design and implement interactive, responsive data visualizations (e.g., dashboard)
- Website, video, paper and presentation

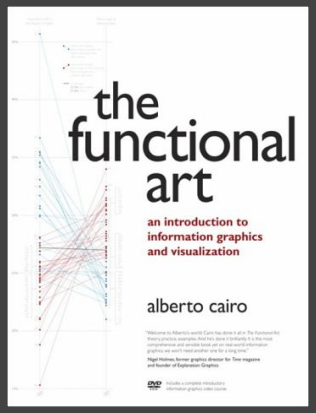
CLASS COMMUNICATION

On Slack at #fall21-dsci-554-general

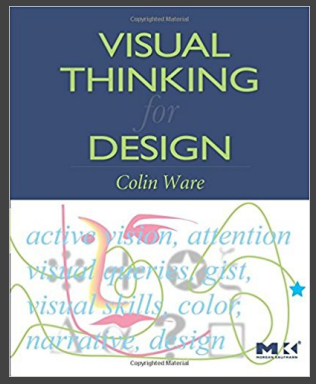
REQUIRED READINGS



Murray S. Interactive Data Visualization for the Web, 2nd Edition. 2nd ed. O'Reilly Media, Inc; 2017. †



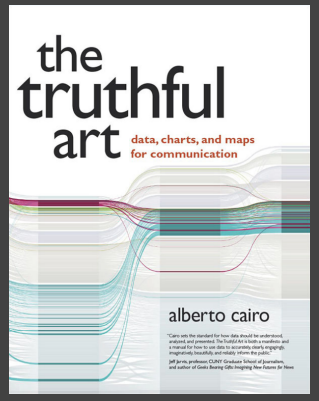
Alberto Cairo. The Functional Art: An Introduction to Information Graphics and Visualization. First. New Riders; 2012. †



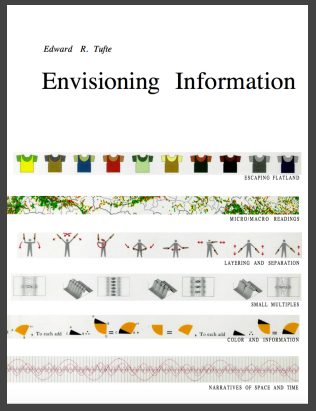
Colin Ware. Visual Thinking: For Design. 1st ed. Morgan Kaufmann Publishers Inc; 2008. †

† online through USC Libraries

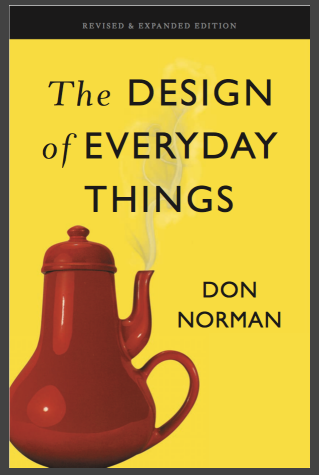
OTHER RELATED READINGS



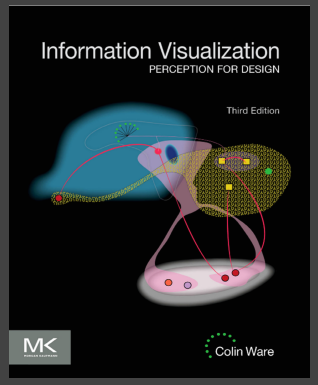
Cairo A. The Truthful Art. Pearson Education; 2016. †



Tufte ER. Envisioning Information . Graphics Press; 1990.



Norman DA. The Design of Everyday Things . 1st Basic paperback ed. Basic Books; 2002.

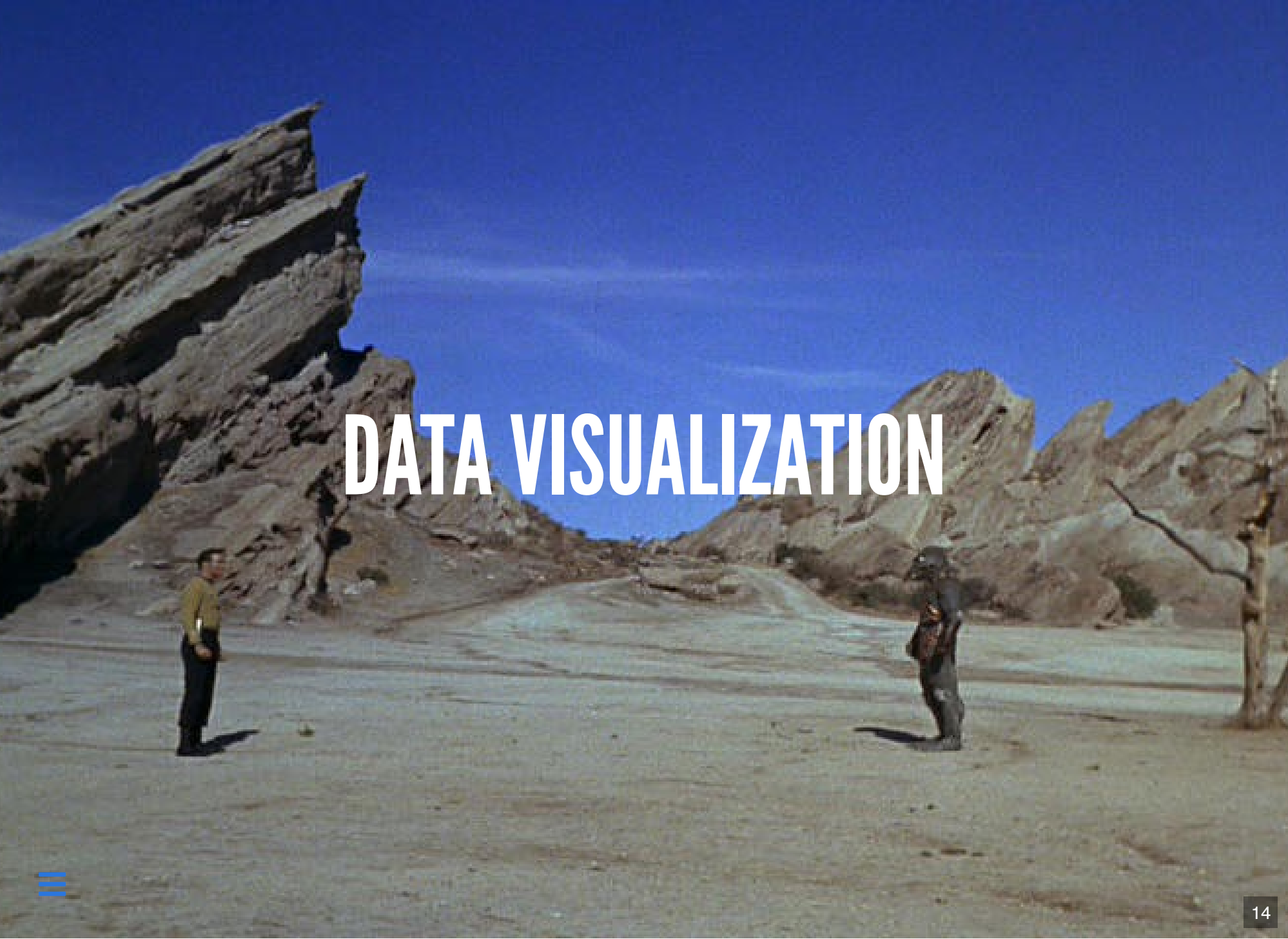


Ware C. Information Visualization Perception for Design . 3rd ed. Elsevier/MK; 2013. †

† online through USC Libraries

OUTLINE

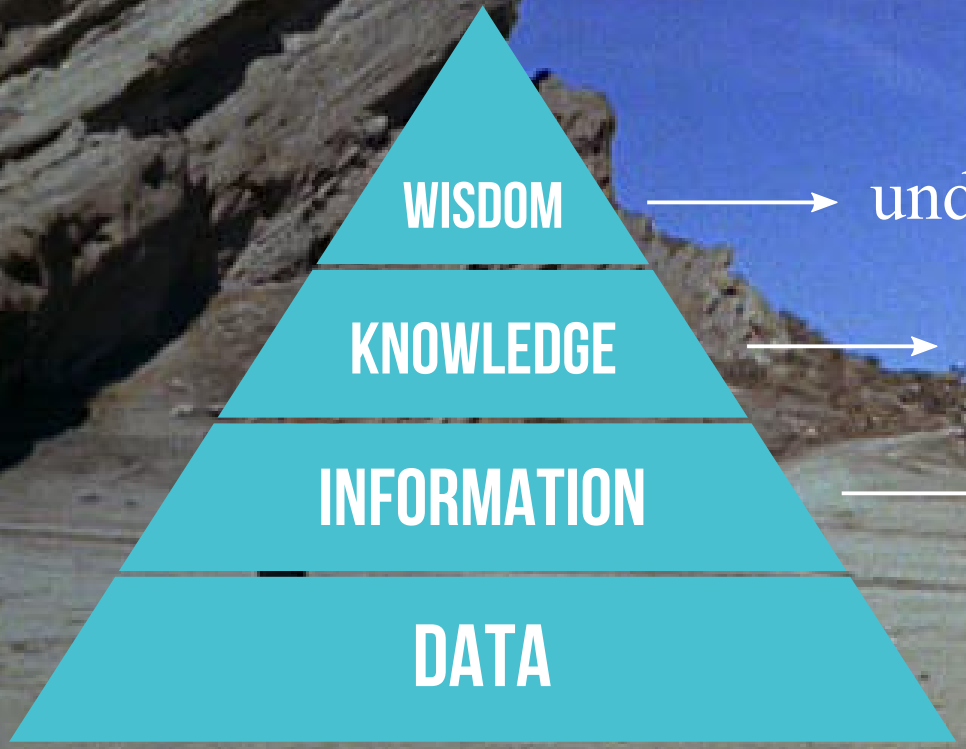
- Course information
- Data visualization
- Examples and uses
- Tools and software


DATA VISUALIZATION

DATA VISUALIZATION

DIKW MODEL [ACKOFF 1989]

DATA VISUALIZATION



understanding of principles

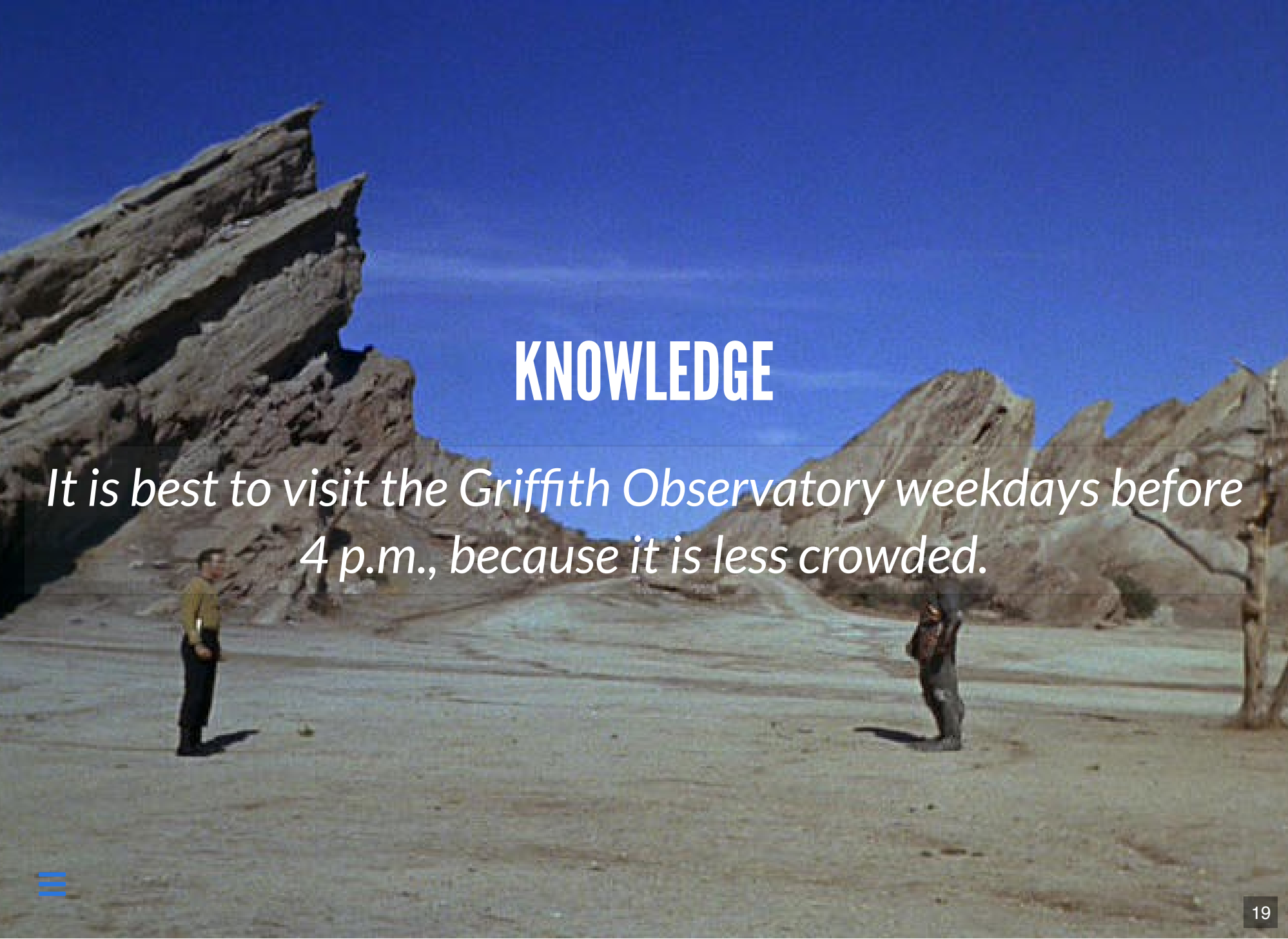
understanding of patterns

understanding of relationships


DATA
The Griffith Observatory is open Tuesday to Friday during 12:00 noon - 10:00 p.m., admission to building and grounds is always FREE.

INFORMATION

The Griffith Observatory is open Tuesday to Friday during 12:00 noon - 10:00 p.m., admission to building and grounds is always FREE.

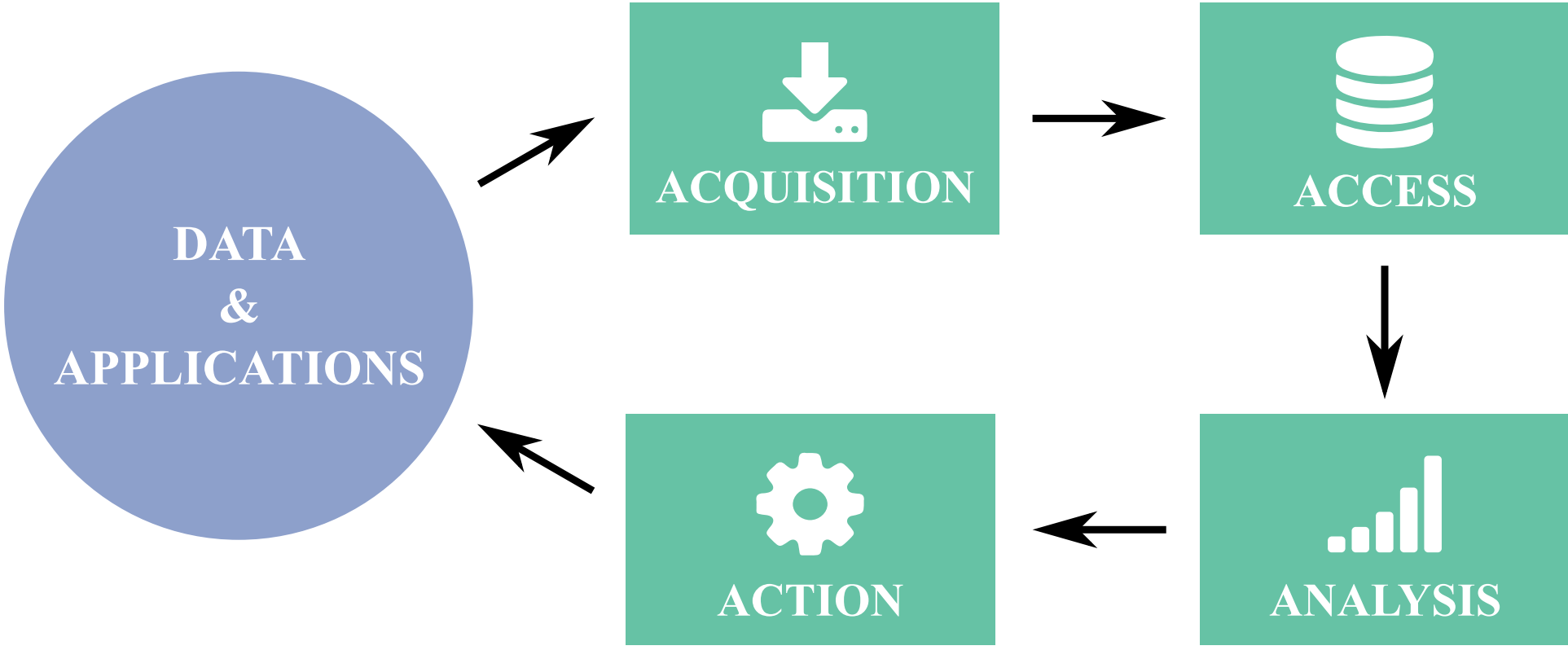


KNOWLEDGE

It is best to visit the Griffith Observatory weekdays before 4 p.m., because it is less crowded.


WISDOM
To go from USC Park Campus to the Griffith Observatory at this time takes 45 min with traffic.

DATA VISUALIZATION IN DATA SCIENCE



Data science process flowchart

UNITS OF DATA & INFORMATION

1 byte = 8 bits

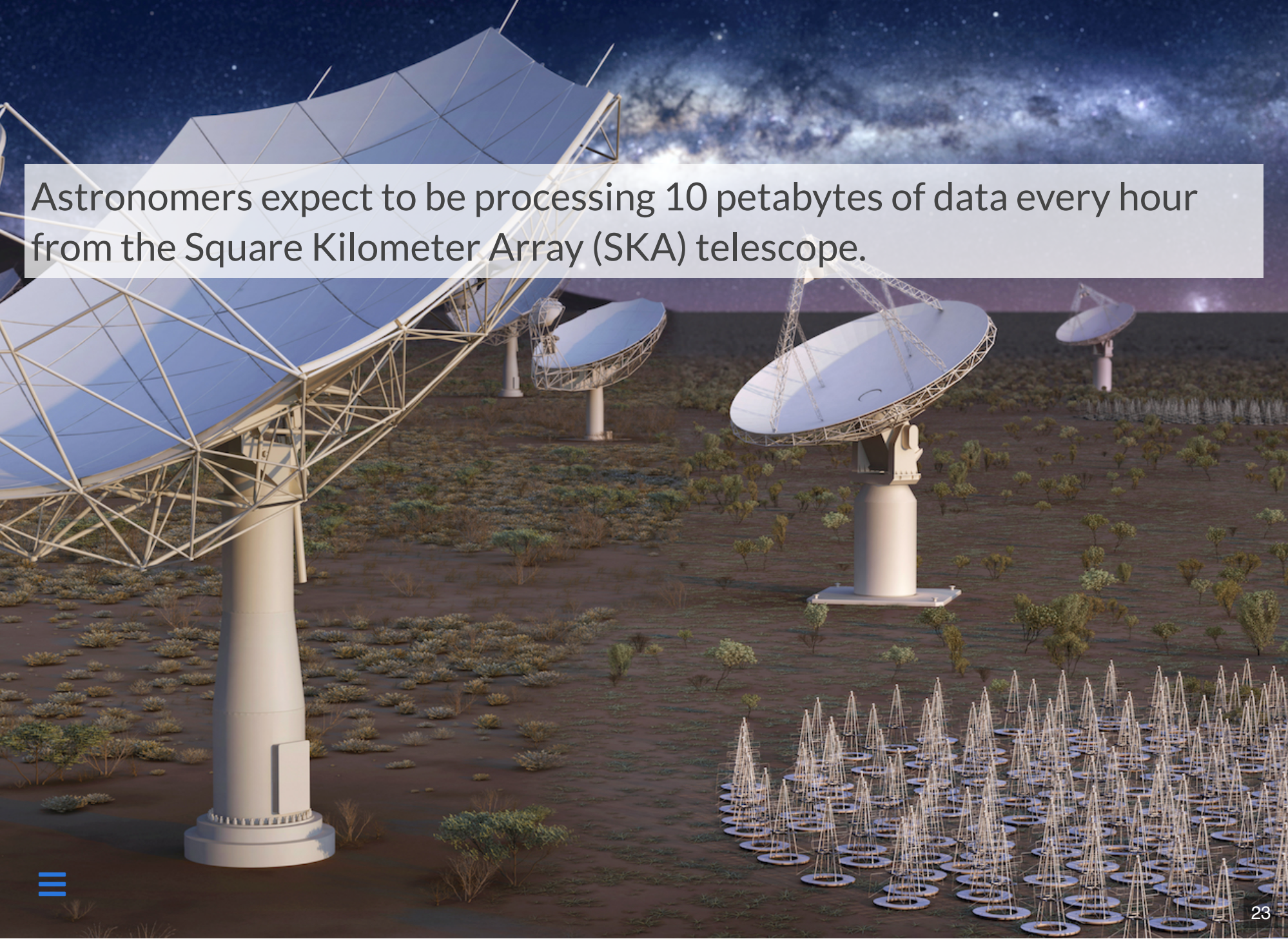
Ex: ASCII characters encoded using 1 byte: $2^8 = 256$ possible values

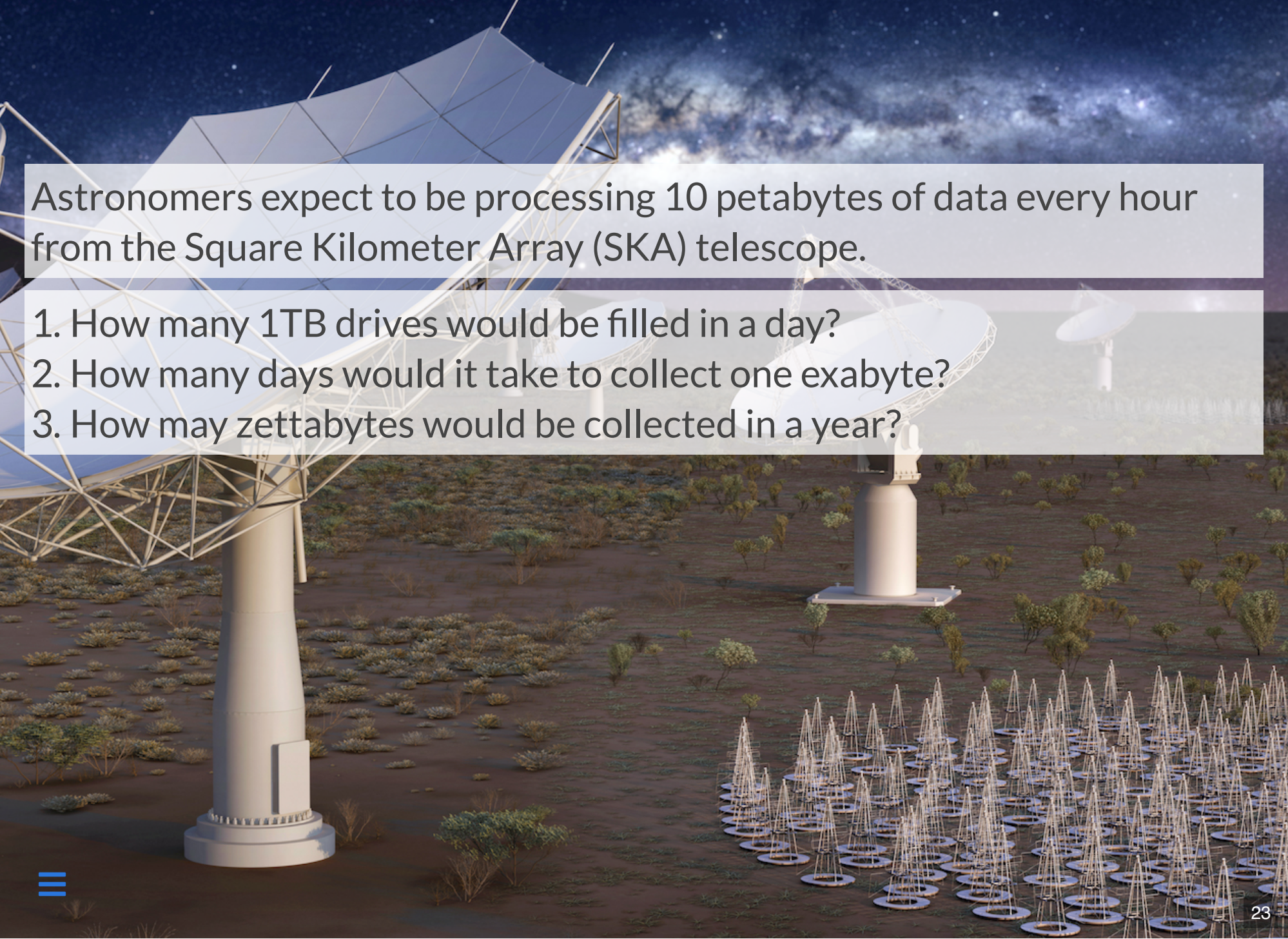
Multiples of bytes					
Decimal			**Binary**		
Value	**Metric**		**Value**	**IEC**	
1000	kB	kilobyte	1024	KiB	kibibyte
1000^2	MB	megabyte	1024^2	MiB	mebibyte
1000^3	GB	gigabyte	1024^3	GiB	gibibyte
1000^4	TB	terabyte	1024^4	TiB	tebibyte
1000^5	PB	petabyte	1024^5	PiB	pebibyte
1000^6	EB	exabyte	1024^6	EiB	exbibyte
1000^7	ZB	zettabyte	1024^7	ZiB	zebibyte
1000^8	YB	yottabyte	1024^8	YiB	yobibyte

Prefix	Symbol	Associated Value
Tera	T	4
Peta	P	5
Exa	E	6
Zetta	Z	7
Yotta	Y	8

💡 Convert between metric values using powers of kilobytes (kB)

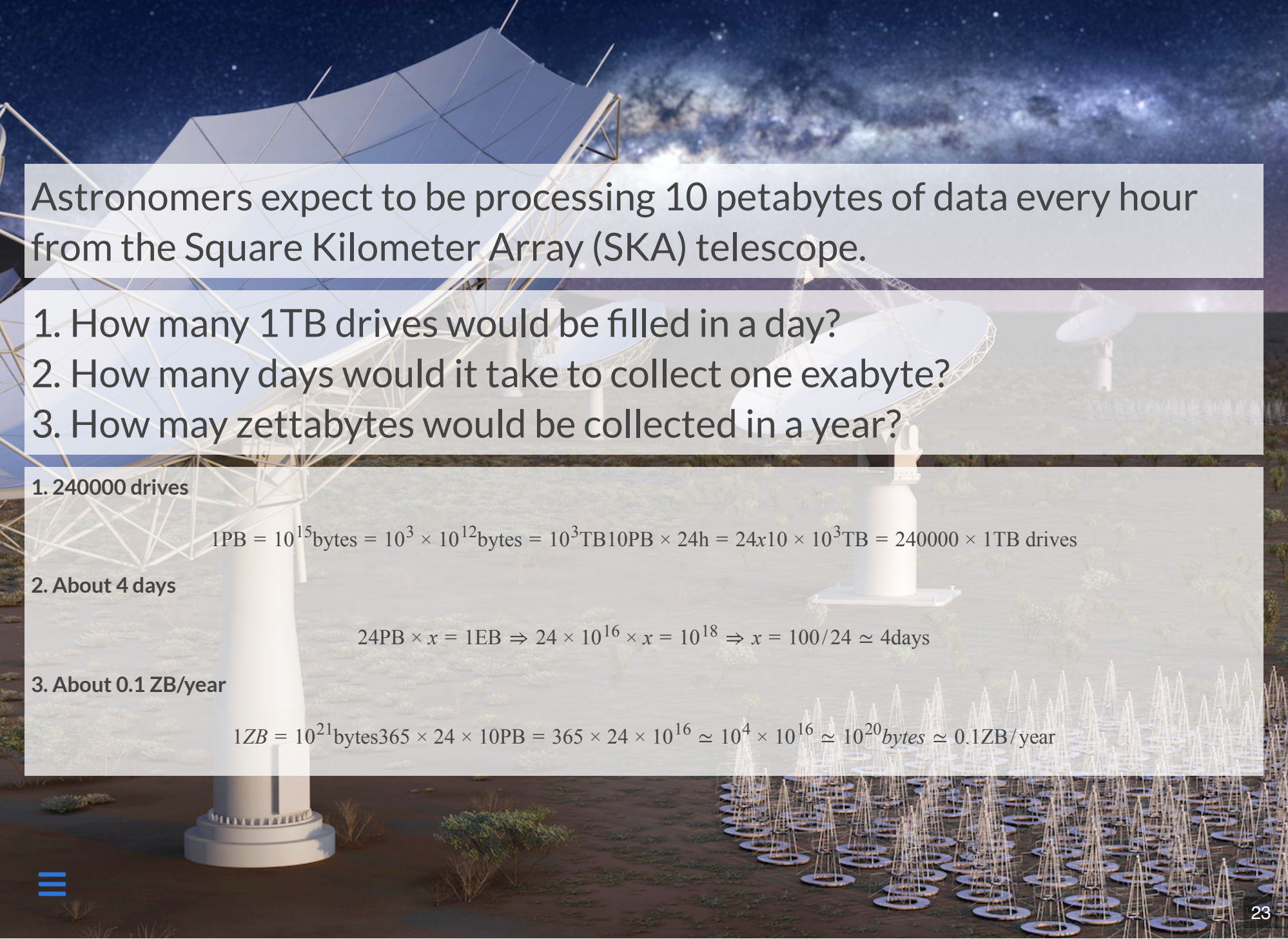
Ex: $1\text{PB} = 1000^5\text{bytes} = 10^{15}\text{bytes}$


Astronomers expect to be processing 10 petabytes of data every hour from the Square Kilometer Array (SKA) telescope.



Astronomers expect to be processing 10 petabytes of data every hour from the Square Kilometer Array (SKA) telescope.

1. How many 1TB drives would be filled in a day?
2. How many days would it take to collect one exabyte?
3. How may zettabytes would be collected in a year?



Astronomers expect to be processing 10 petabytes of data every hour from the Square Kilometer Array (SKA) telescope.

1. How many 1TB drives would be filled in a day?
2. How many days would it take to collect one exabyte?
3. How may zettabytes would be collected in a year?

1. 240000 drives

$$1\text{PB} = 10^{15}\text{bytes} = 10^3 \times 10^{12}\text{bytes} = 10^3\text{TB}10\text{PB} \times 24\text{h} = 24x10 \times 10^3\text{TB} = 240000 \times 1\text{TB drives}$$

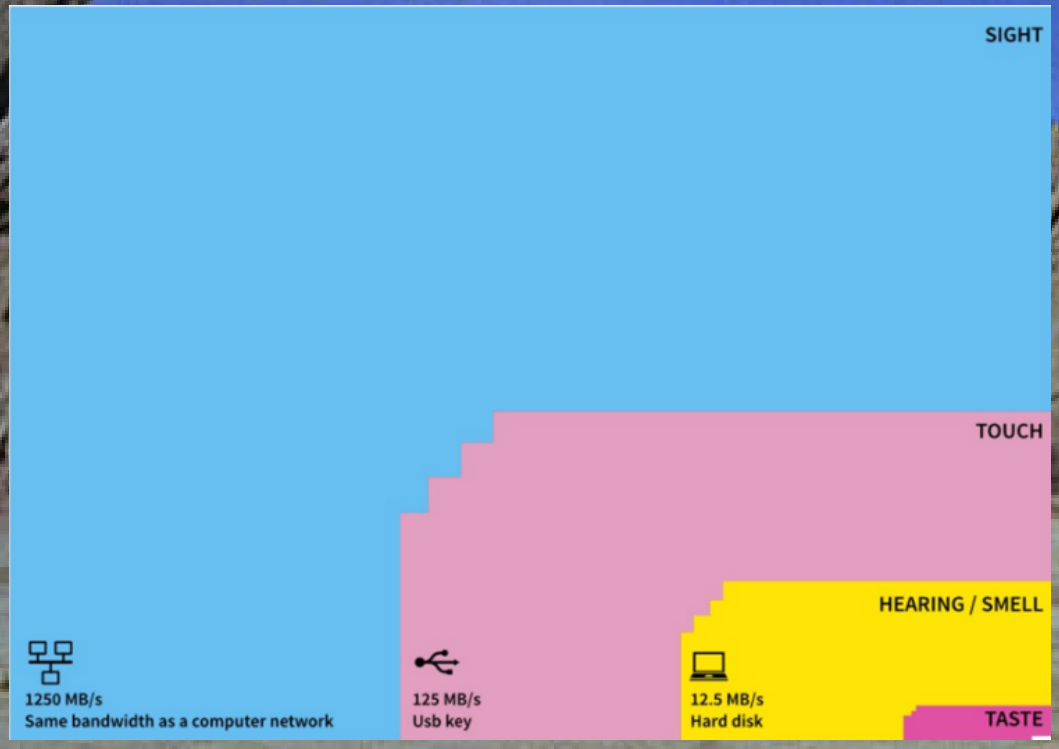
2. About 4 days

$$24\text{PB} \times x = 1\text{EB} \Rightarrow 24 \times 10^{16} \times x = 10^{18} \Rightarrow x = 100/24 \simeq 4\text{days}$$

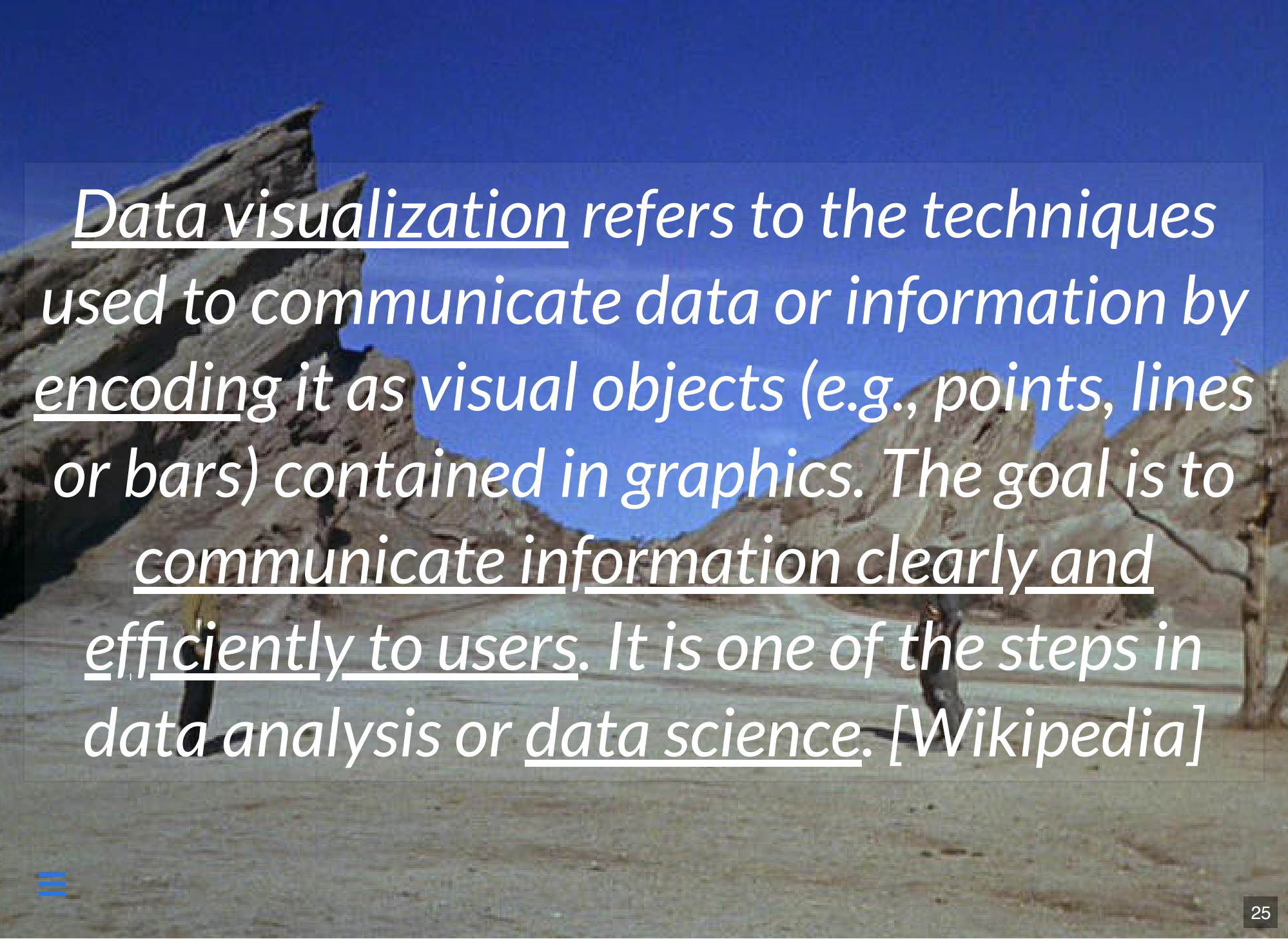
3. About 0.1 ZB/year

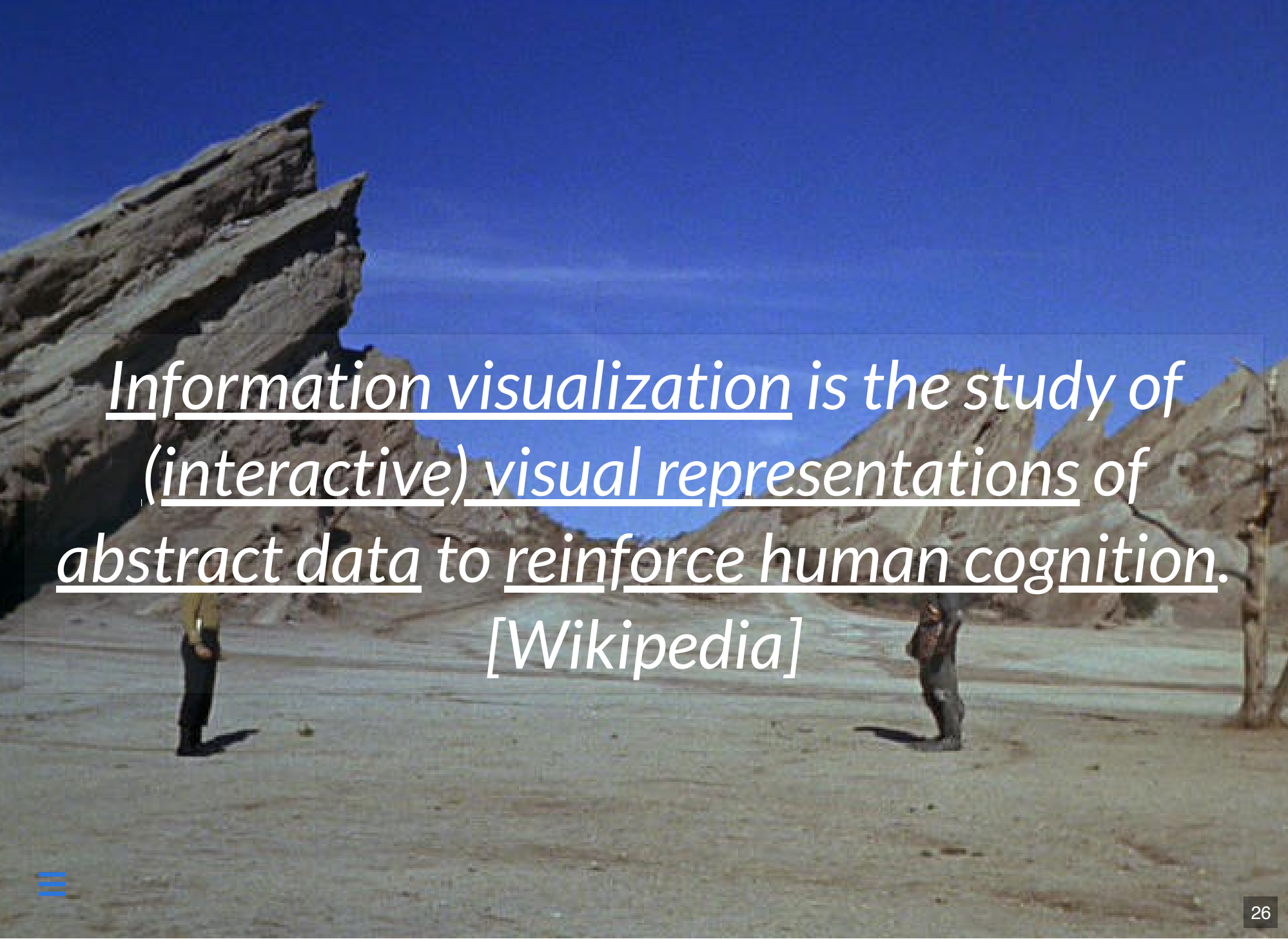
$$1ZB = 10^{21}\text{bytes}365 \times 24 \times 10\text{PB} = 365 \times 24 \times 10^{16} \simeq 10^4 \times 10^{16} \simeq 10^{20}bytes \simeq 0.1\text{ZB/year}$$

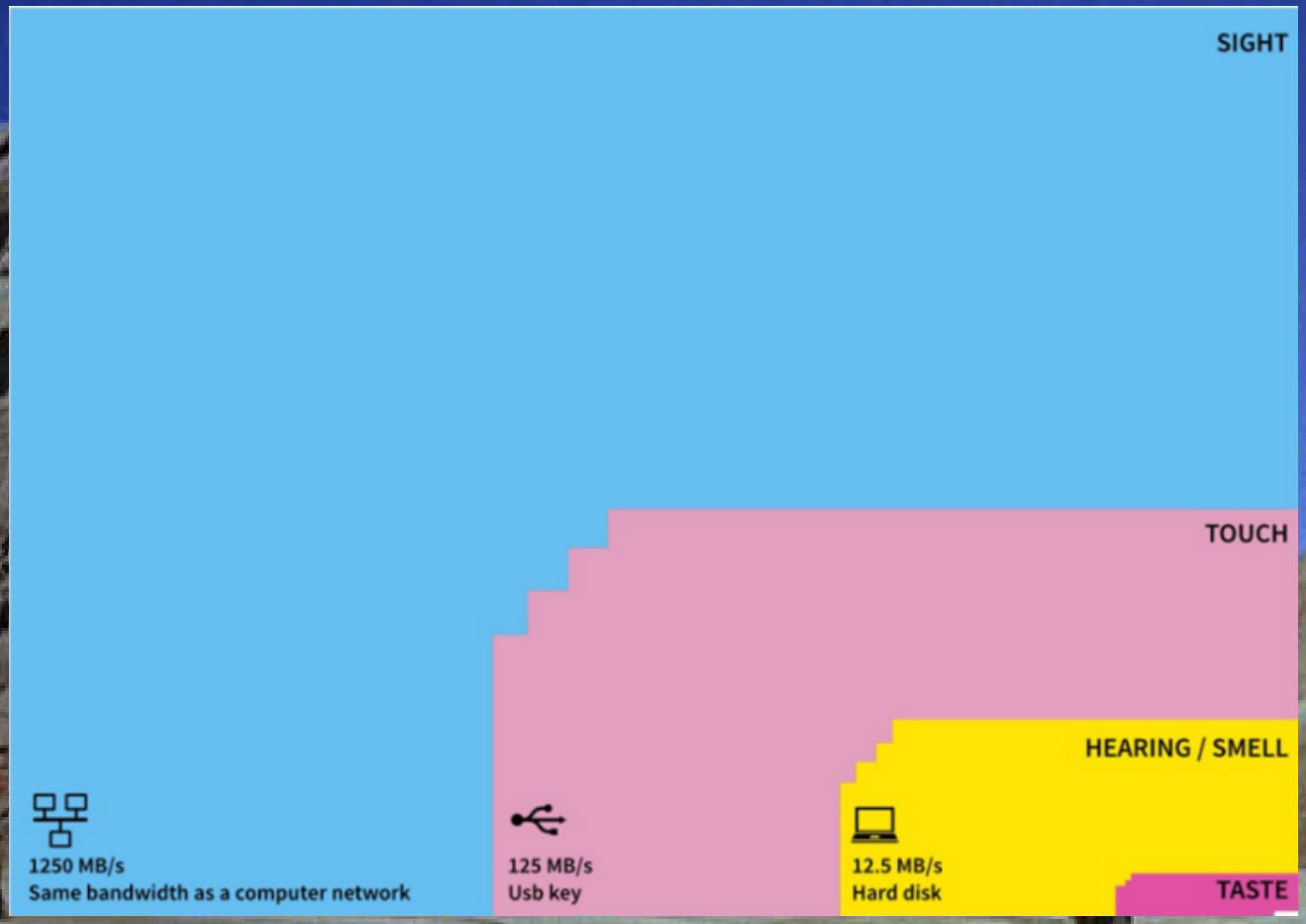
DATA VISUALIZATION



Nørretranders bandwidth of senses Graphic by David McCandless


Data visualization refers to the techniques used to communicate data or information by encoding it as visual objects (e.g., points, lines or bars) contained in graphics. The goal is to communicate information clearly and efficiently to users. It is one of the steps in data analysis or data science. [Wikipedia]


Information visualization is the study of (interactive) visual representations of abstract data to reinforce human cognition.
[Wikipedia]



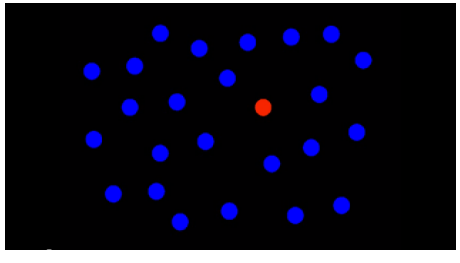
Nørretranders bandwidth of senses - Graphic by David McCandless

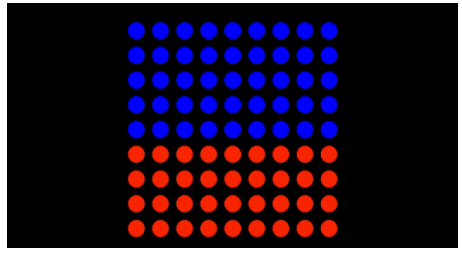
You are aware of 0.7% of what you experience:

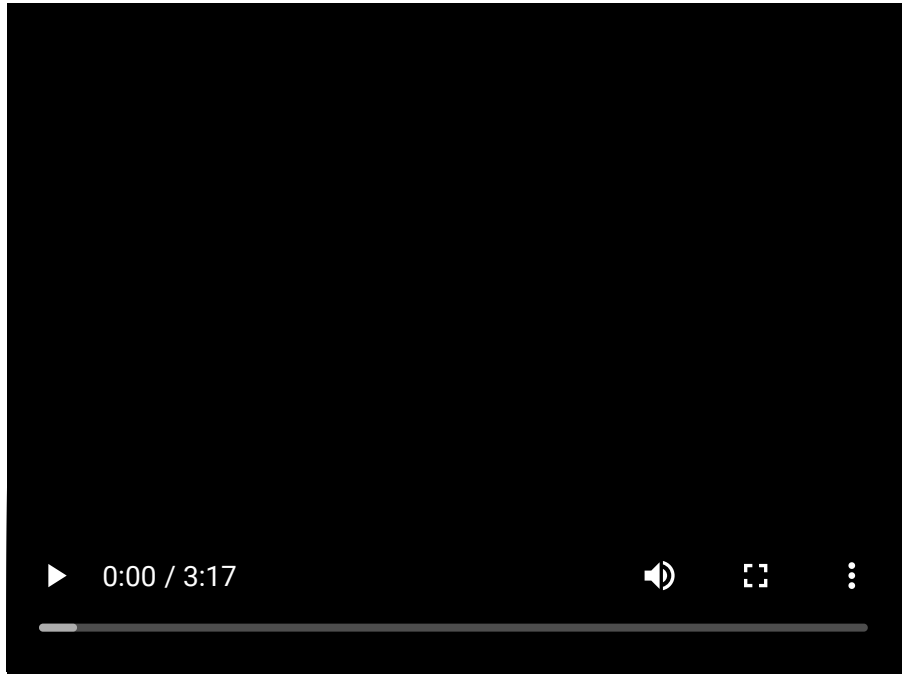
- High-res limited to central 3° of visual field
- Finite cognitive capabilities

PREATTENTIVE FEATURES (UNIVERSAL CAPABILITY)

- Typically in less than 1/10s
- Does not require eye movements
- Does not require focused attention
- Color and boundary can be detected preattentively







Christopher G. Healey - Preattentive features and tasks

Other examples of universal capabilities:

- Some color combinations are differentiated by everyone
- Some symbols are understood across cultures 🙂

COLOR INTERPRETATION (INDIVIDUAL CAPABILITY)

https://en.wikipedia.org/wiki/The_dress

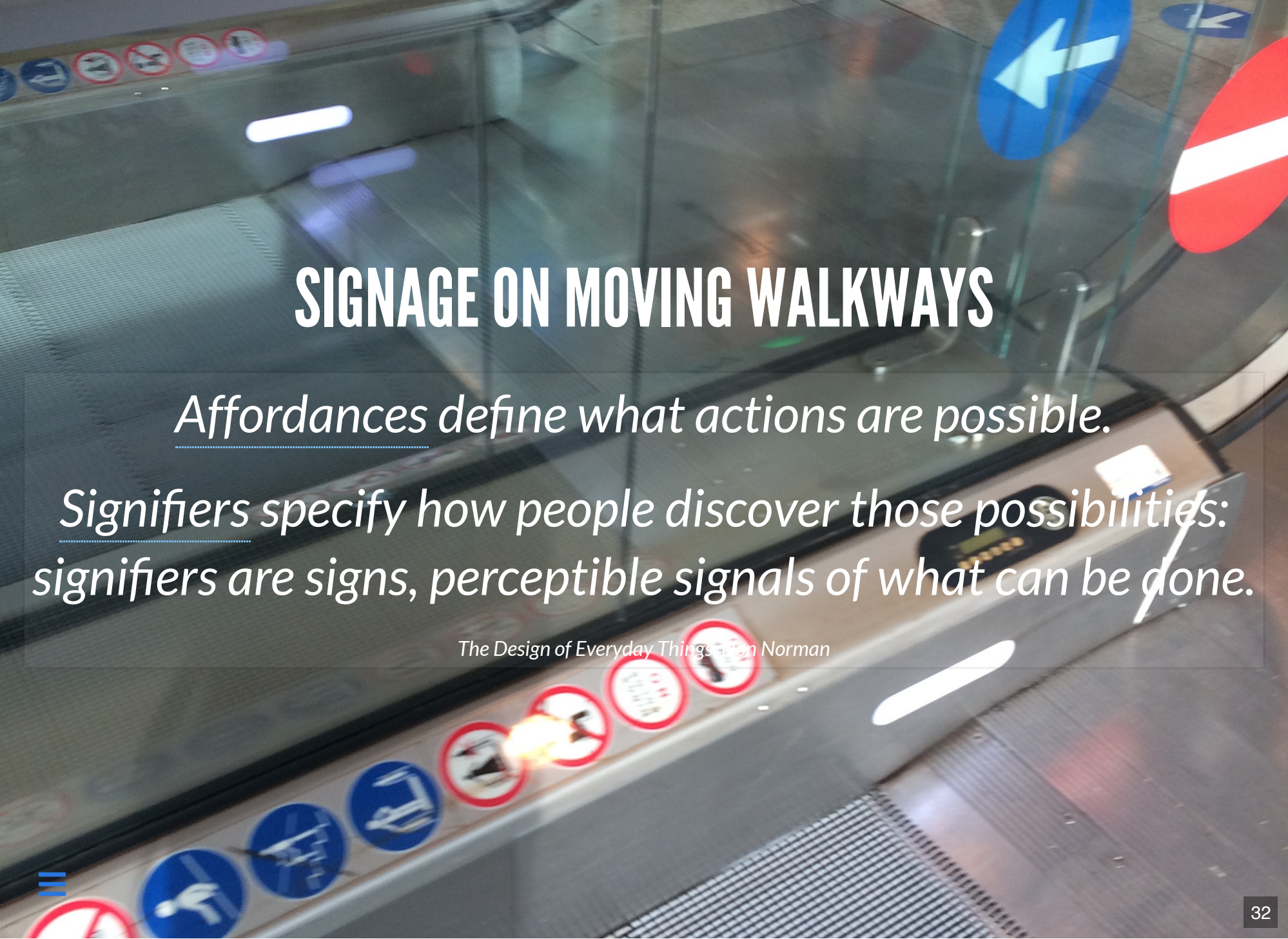
- We interpret lighting differently
- Not everyone can differentiate certain colors 
- Not everyone understands certain symbols ♳
- Not everyone can read or read small text!

INFORMATION VS. SCIENTIFIC VISUALIZATION

	Infovis	Scivis
Representation	chosen	given
Examples		

OUTLINE

- Course information
- Data visualization
- Examples and uses
- Tools and software



SIGNAGE ON MOVING WALKWAYS

Affordances define what actions are possible.

Signifiers specify how people discover those possibilities: signifiers are signs, perceptible signals of what can be done.

The Design of Everyday Things Don Norman

VISUALIZATION USES

Scope	Actions	Examples

VISUALIZATION USES

Scope	Actions	Examples
Communicate Information	◦ Inform ◦ Communicate ◦ Explain	◦ Presentations ◦ Hand-outs ◦ Instructions ◦ Infographics ◦ Signage

VISUALIZATION USES

Scope	Actions	Examples
Communicate Information	○ Inform ○ Communicate ○ Explain	○ Presentations ○ Hand-outs ○ Instructions ○ Infographics ○ Signage
Analyze & Model Data	○ Explore ○ Analyze ○ Discover ○ Decide	○ Spreadsheets ○ Dashboards ○ Notebooks ○ Interactive graphics

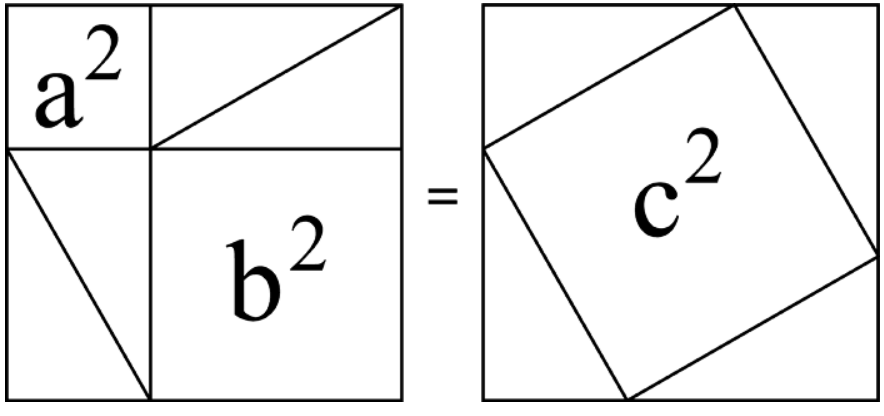
Which information visualization use most relates to communicating information?

A. Explore
B. Analyze
C. Explain
D. Decide

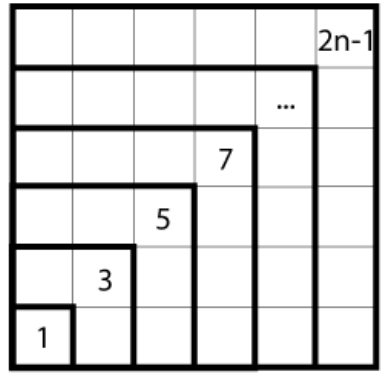
Which information visualization use most relates to communicating information?

A. Explore
B. Analyze
C. Explain ←
D. Decide

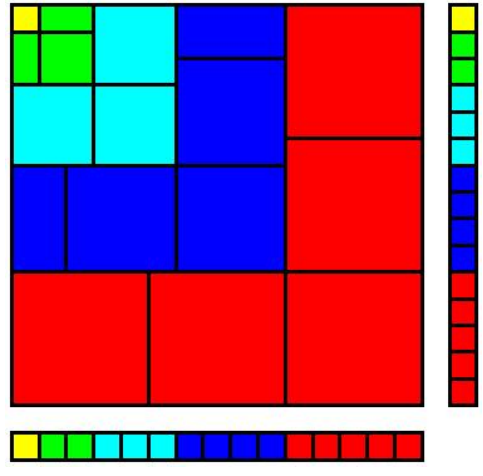
CAN REPLACE COMPLEX CALCULATIONS



$$a^2 + b^2 = c^2$$



$$1 + 3 + 5 + \cdots + (2n - 1) = n^2$$



$$\sum_{k=1}^{n} k^3 = \left(\sum_{k=1}^{n} k\right)^2$$

CAN REVEAL COMPLEX PATTERNS, TRENDS AND OUTLIERS

193

189

297

311

247

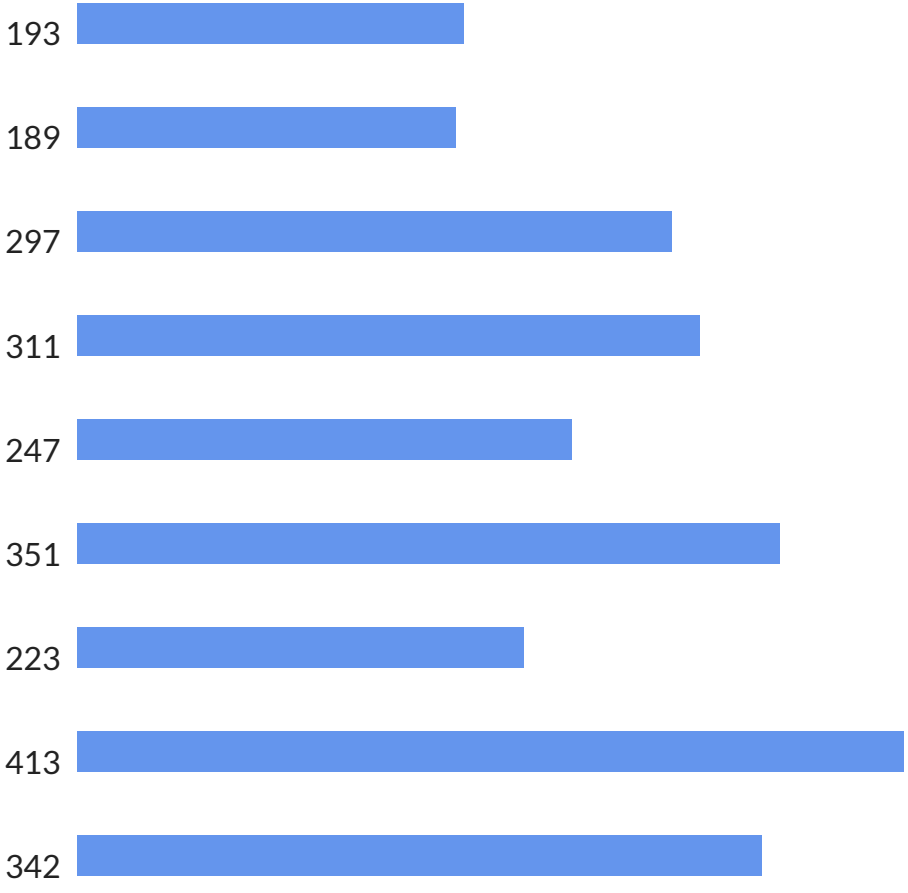
351

223

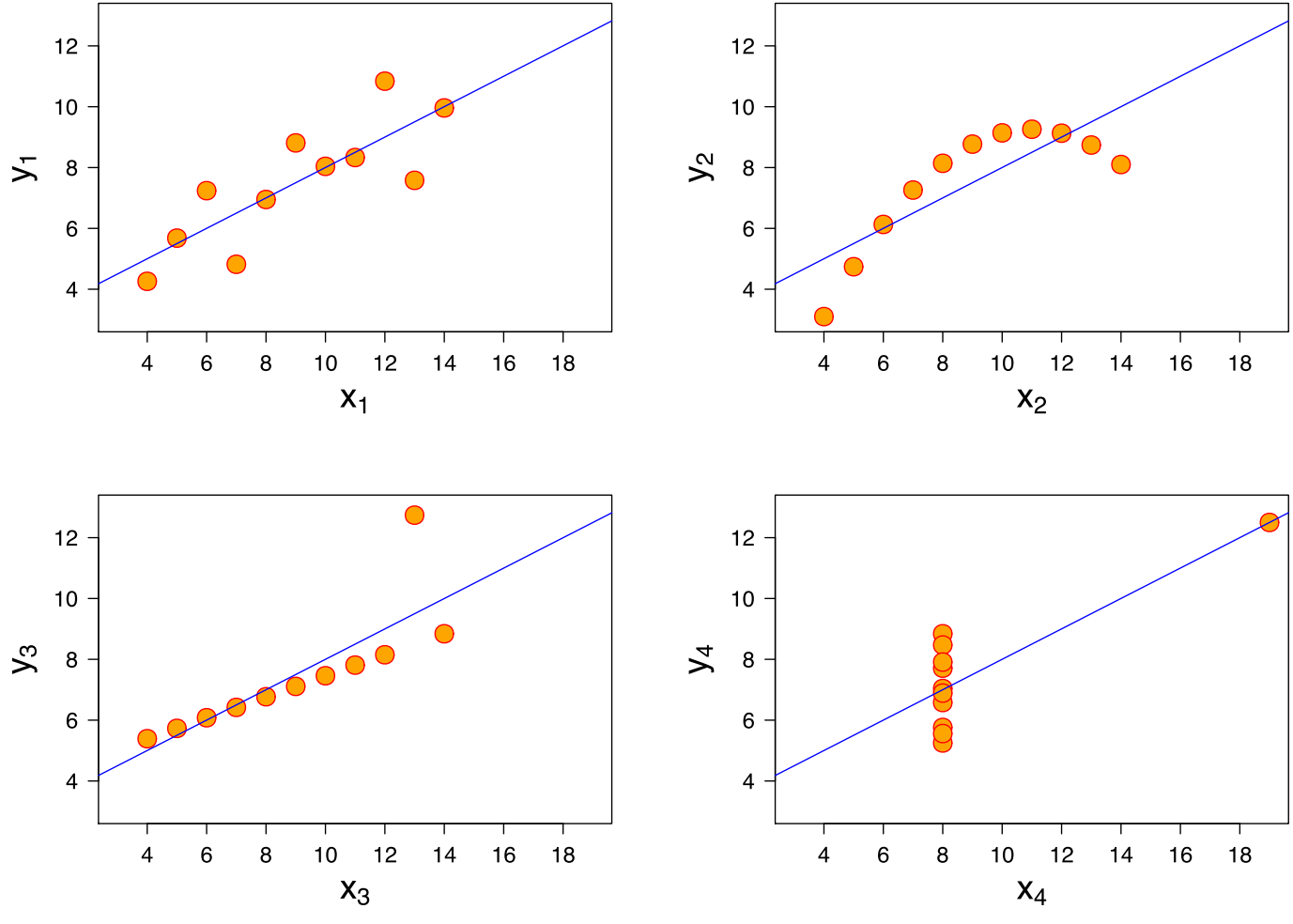
413

342

CAN REVEAL COMPLEX PATTERNS, TRENDS AND OUTLIERS

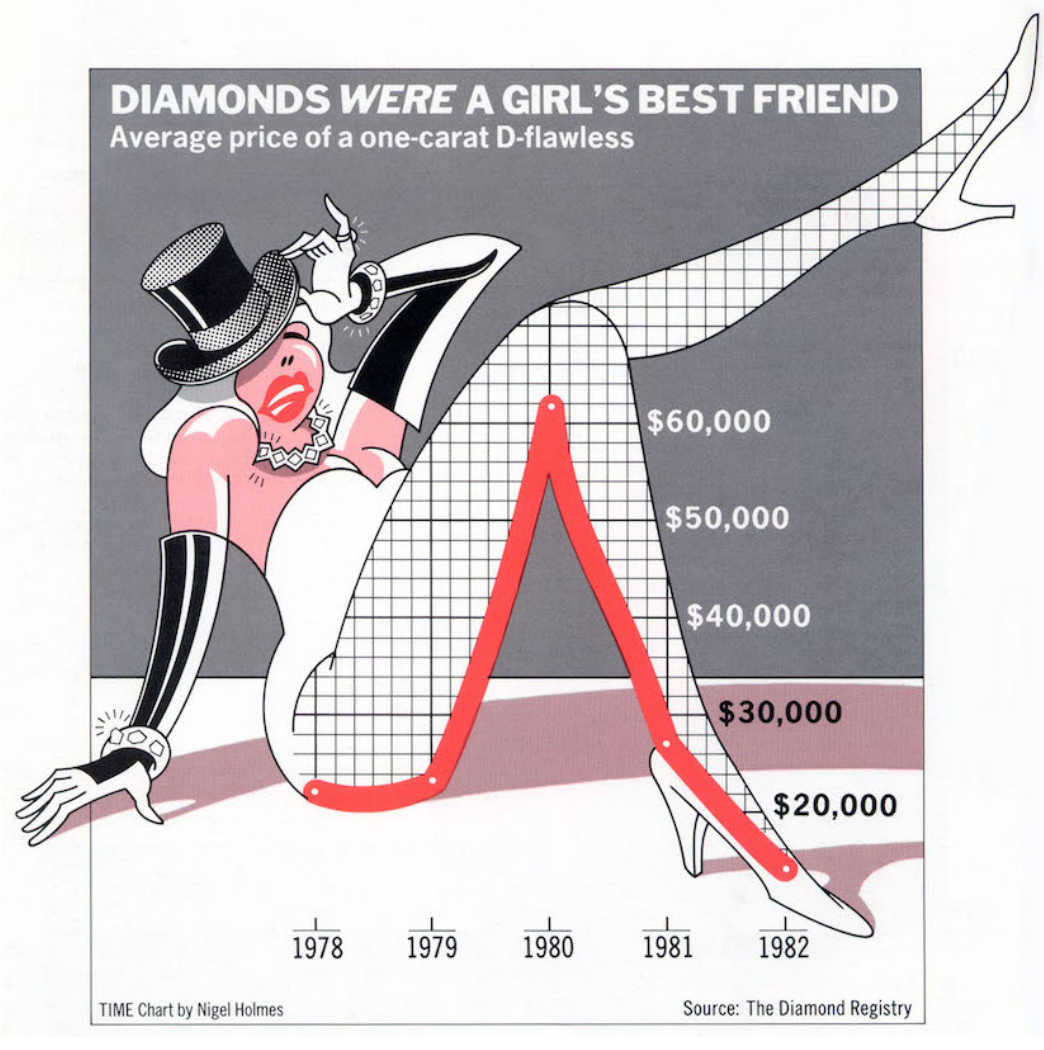


CAN REVEAL FEATURES NOT OTHERWISE APPARENT



Anscombe's quartet (1973): importance of graphing data before analysis

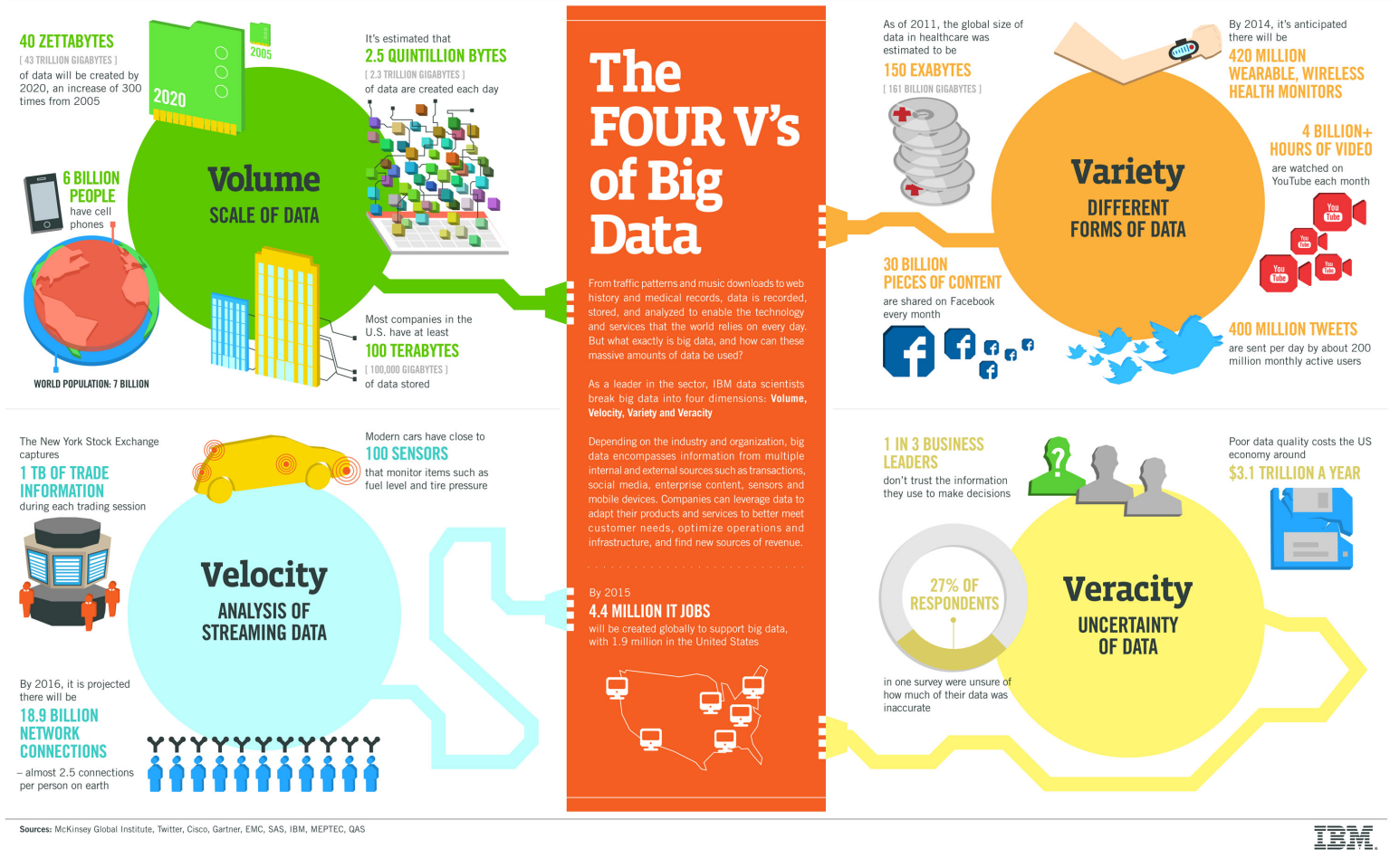
CAN SUPPORT MEMORY AND COMPREHENSION



CAN TELL A STORY

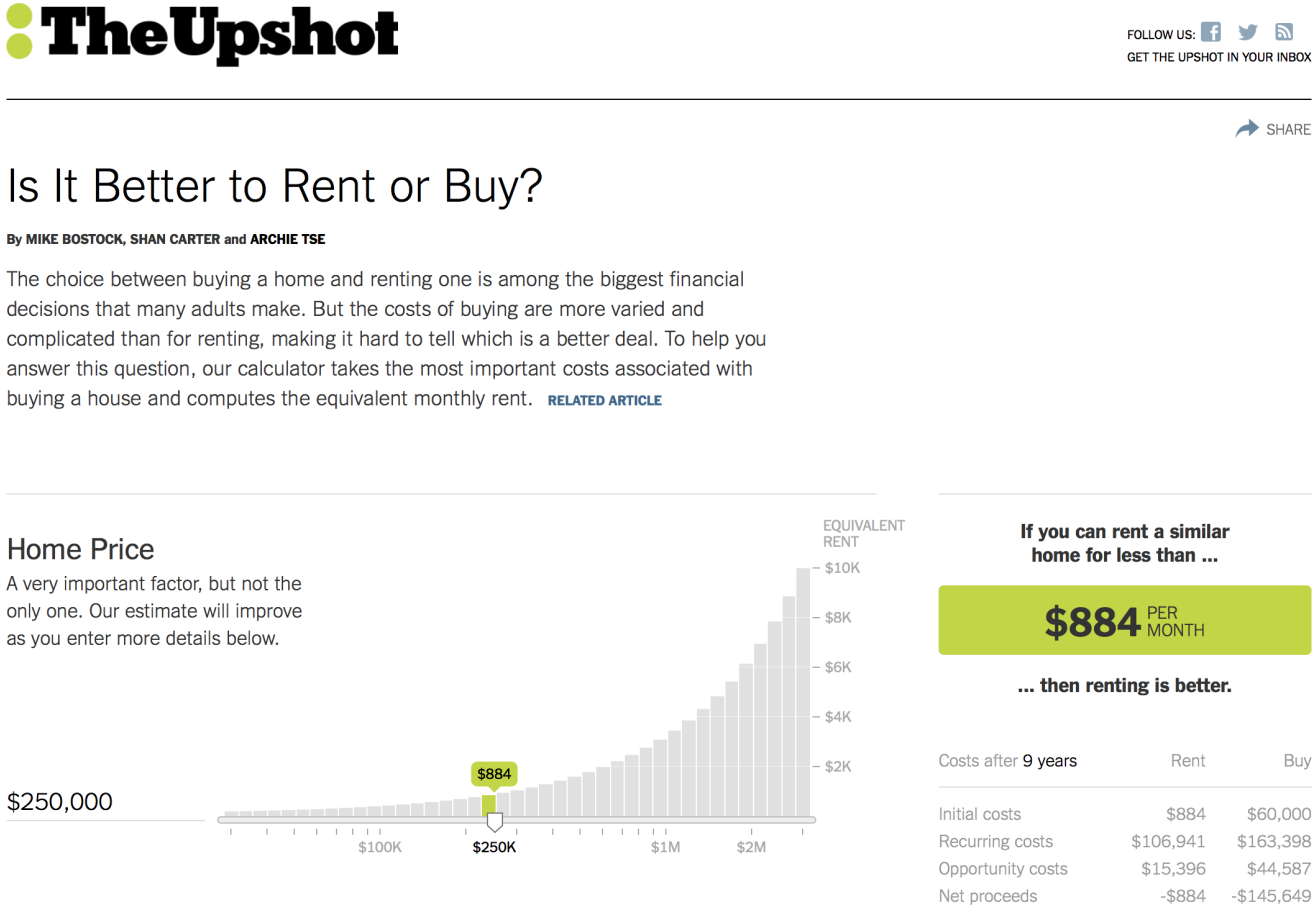
Hans Rosling's 200 Countries, 200 Years
🔗 https://youtu.be/jbkSRLYSojo

CAN INFORM AND ENGAGE MORE DIVERSE AUDIENCES



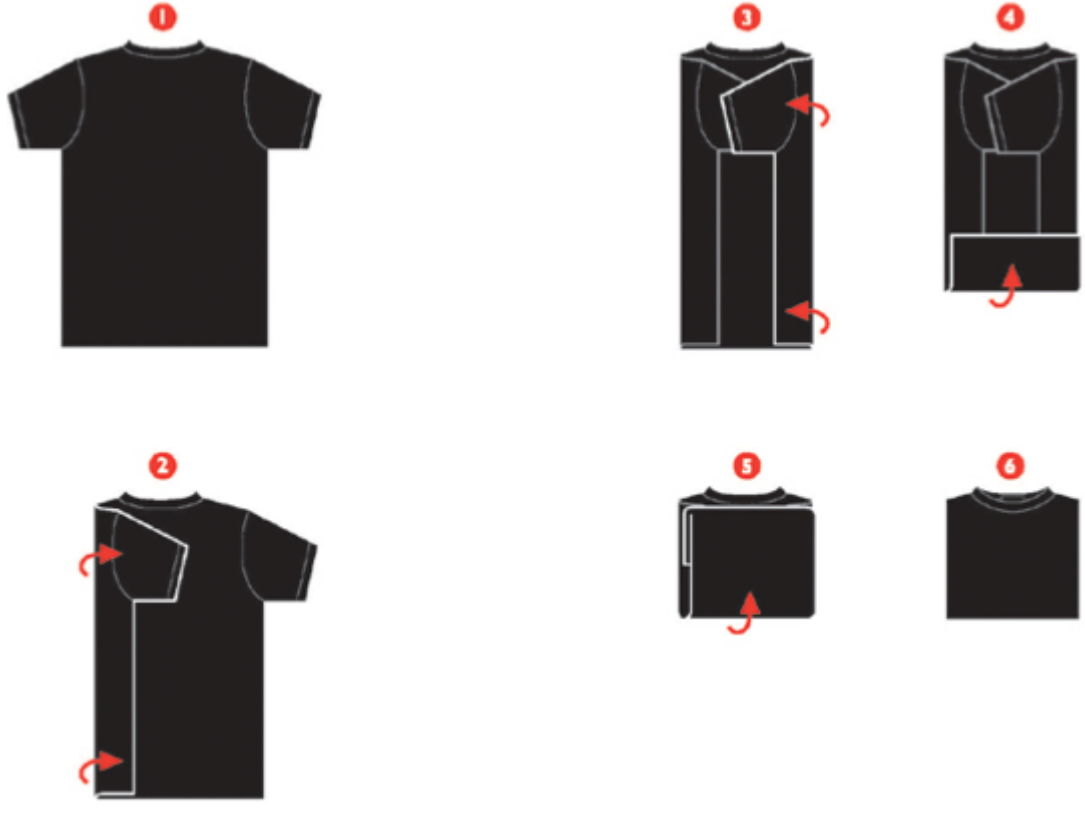
IBM Big Data & Analytics Hub - Infographics & Animations

VISUALIZATIONS ARE MEANS TO REACH GOALS



NYT Buy rent calculator

INFORMATION GRAPHICS (INFOGRAPHICS) ARE DEVICES WHOSE AIM IS TO HELP AN AUDIENCE COMPLETE CERTAIN TASKS

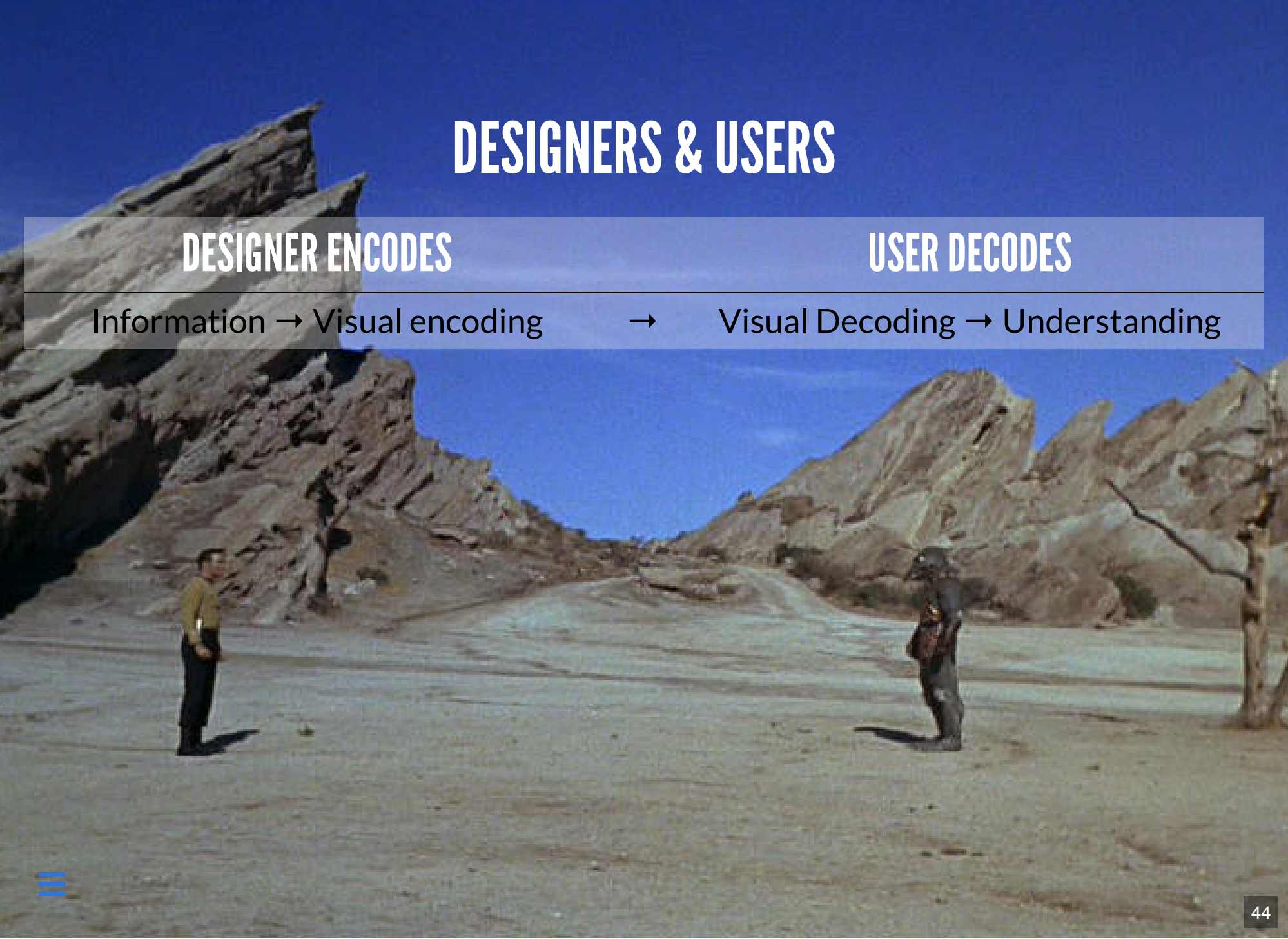


Wordless Diagrams (2005) by Nigel Holmes.

DESIGN CONSIDERATIONS

DESIGNERS & USERS

DESIGNER ENCODES		USER DECODES
Information → Visual encoding	→	Visual Decoding → Understanding



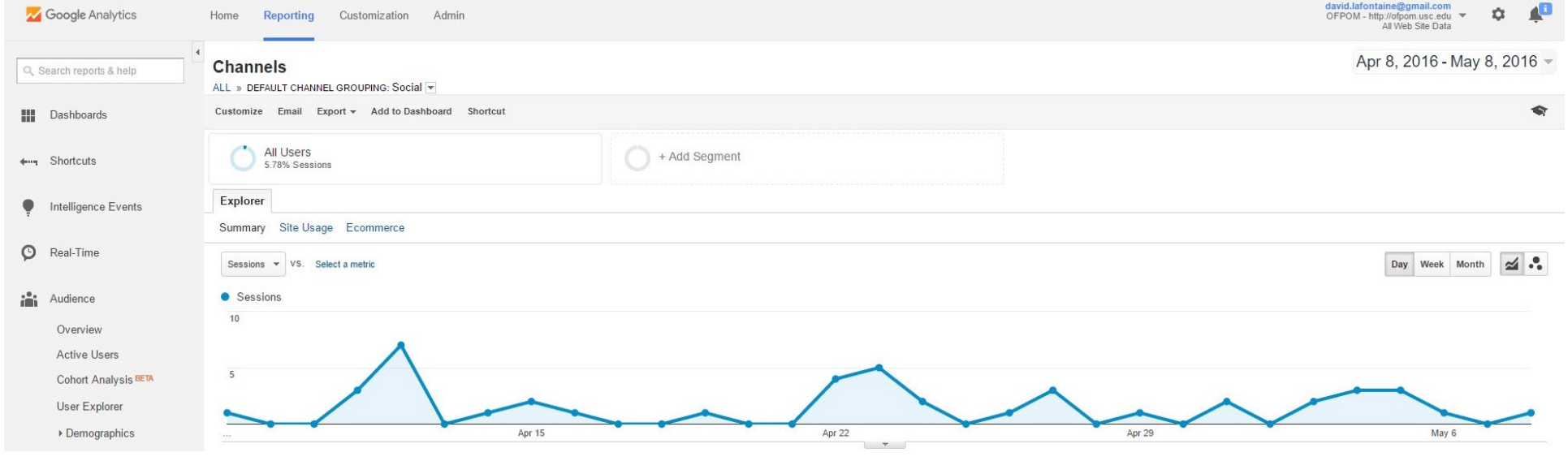
DESIGNERS & USERS

DESIGNER ENCODES		USER DECODES
Information → Visual encoding	→	Visual Decoding → Understanding

WHAT INFORMATION DESIGNERS USE

DATA RELATED	USER RELATED
Form adapted to nature of information	◦ User familiarity with form ◦ User knowledge of topic ◦ User abilities ◦ Display type and size ◦ Context where the form is used

THE FORM SHOULD BE CONSTRAINED BY THE GOALS OF THE VISUALIZATION



Google Analytics dashboard

FORM FOLLOWS FUNCTION

20th-century modernist architecture and industrial design principle

The shape of an object should primarily relate to its intended function or purpose

Sullivan, Louis H. (1896). "The Tall Office Building Artistically Considered". Lippincott's Magazine (March 1896).

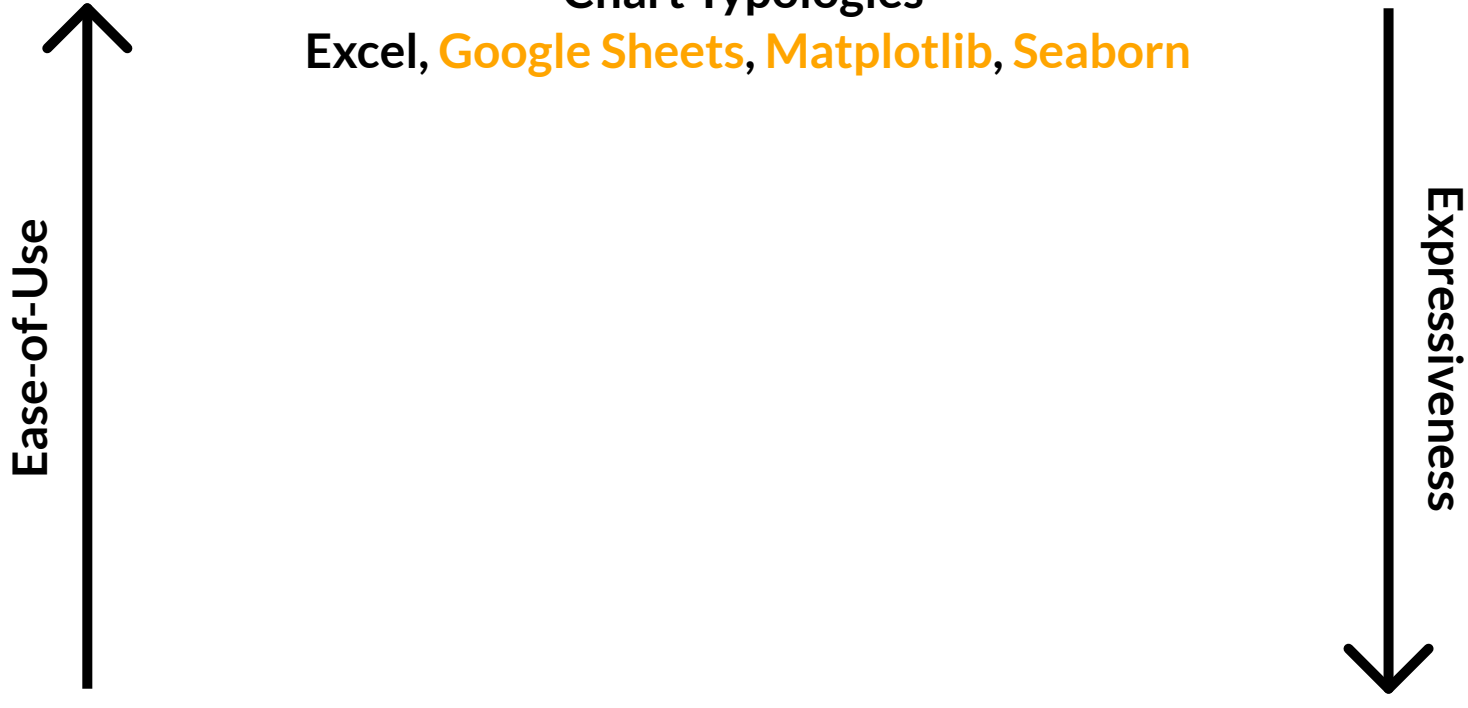
OUTLINE

- Course information
- Data visualization
- Examples and uses
- Tools and software

VISUALIZATION TOOLS

Chart Typologies

Excel, Google Sheets, Matplotlib, Seaborn



Adapted from [Heer 2014]

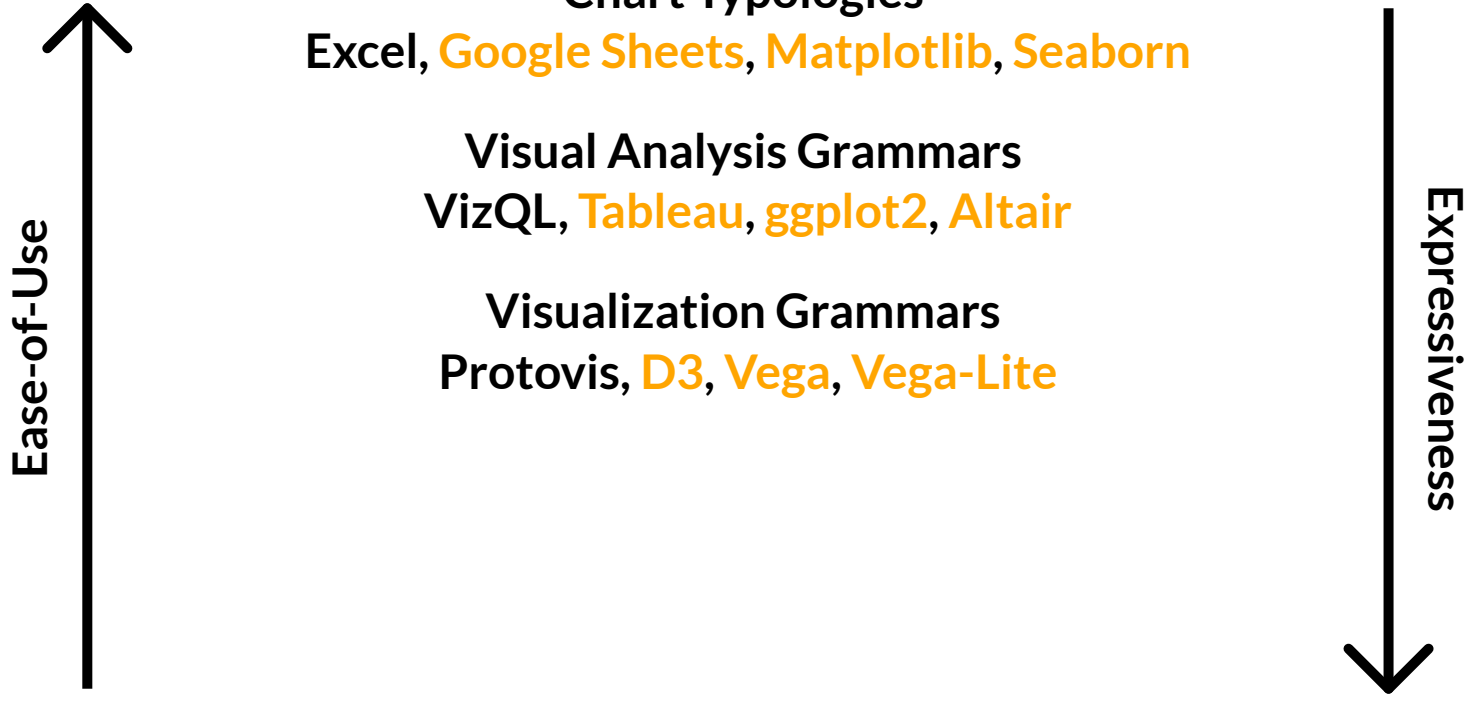
Satyanarayan, Arvind, and Jeffrey Heer. "Lyra: An interactive visualization design environment." In Computer Graphics Forum, 2014.

VISUALIZATION TOOLS

Adapted from [Heer 2014]

Satyanarayan, Arvind, and Jeffrey Heer. "Lyra: An interactive visualization design environment." In Computer Graphics Forum, 2014.

VISUALIZATION TOOLS



Adapted from [Heer 2014]

Satyanarayan, Arvind, and Jeffrey Heer. "Lyra: An interactive visualization design environment." In Computer Graphics Forum, 2014.

VISUALIZATION TOOLS

Adapted from [Heer 2014]

Satyanarayan, Arvind, and Jeffrey Heer. "Lyra: An interactive visualization design environment." In Computer Graphics Forum, 2014.

VISUALIZATION TOOLS

Adapted from [Heer 2014]

Satyanarayan, Arvind, and Jeffrey Heer. "Lyra: An interactive visualization design environment." In Computer Graphics Forum, 2014.

D3.JS

WHAT IT IS

- Javascript client-side library
- D3 stands for Data-Driven Documents
- Uses recent HTML, SVG, and CSS
- Primarily made to use SVG (not raster graphics, i.e., images)

WHAT IT DOES

- Loads data in the browser memory
- Create elements and bind data to elements within the document
- Transform and customize elements
- Transition elements in response to user input