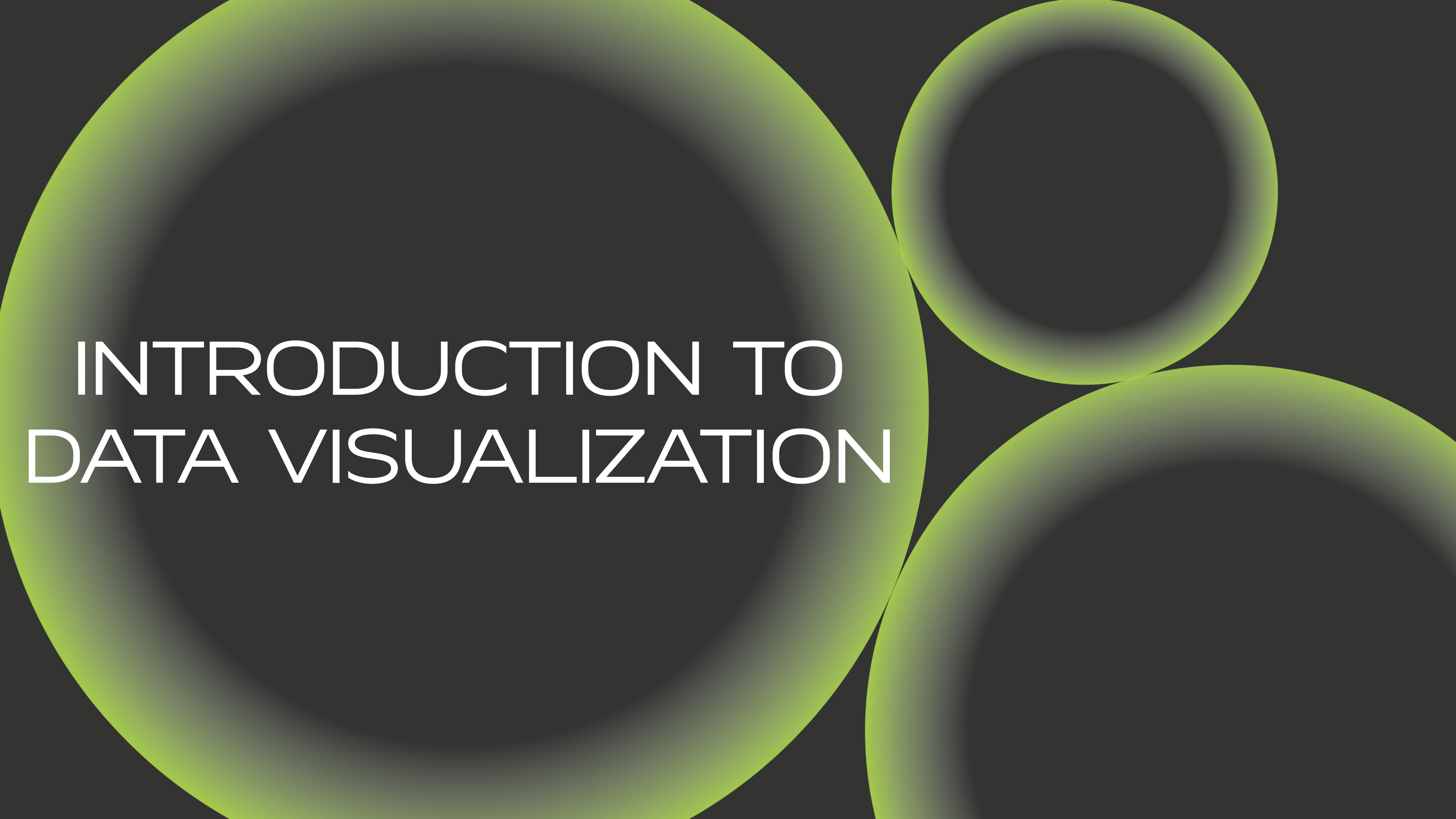




INTRODUCTION TO DATA VISUALIZATION

MODULE INFORMATION

DV.KRATA.SPACE

LEARNING OBJECTIVES

- Collect and structure data from different sources
- Prepare data to train own machine learning models
- Critically analyze datasets and visualizations
- Identify how bias enters through collection, labeling and design choices
- Explain the core principles of machine learning and generative AI
- Build interactive, web-based visualizations using APIs and AI-assisted workflows

LEARNING OBJECTIVES

- Individual Project (5.10 - 14.10) — 20%
- Group Project (17.11 - 05.01) — 40%
- Documentation (11. 01) - 40%

Basel'fied

Scene Mode

Grid Mode

add image

Generate

Unknown	41	Altstadt Grossb...	9
Gundeldingen	5	Altstadt Kleinba...	0
Vorstädte	0	Am Ring	0
Breite	0	St. Johann	0
St. Alban	0	Iselin	0
Bruderholz	0	Bachletten	0
Gotthelf	0	Clara	0
Wettstein	0	Hirzbrunnen	0
Rosental	0	Matthäus	0
Klybeck	0	Kleinhüningen	0

Organize Images

1



Nicolas Birrer & Oliver Wick

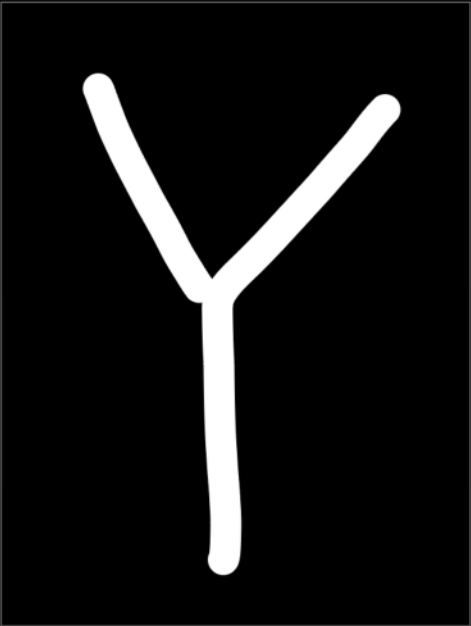
Input Mode

INPUT MODE

CONTROL MODE

CONFIRM

CLEAR

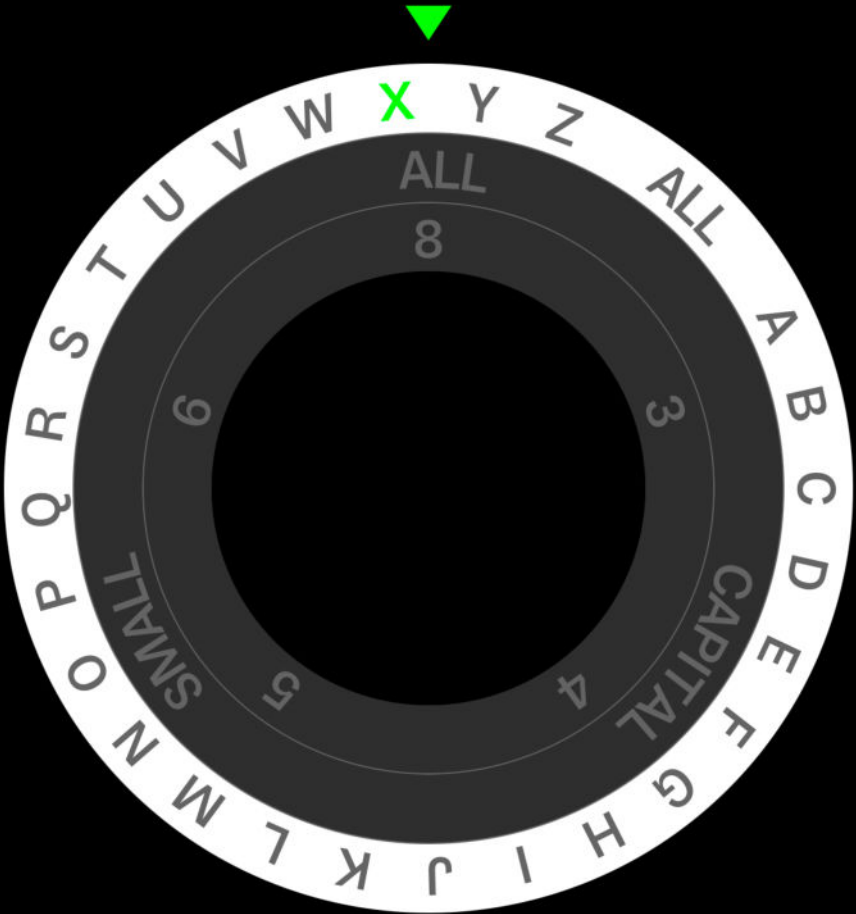


DRAW A LETTER IN THE BOX
CONFIRM YOUR DRAWING

Control Mode

INPUT MODE

CONTROL MODE



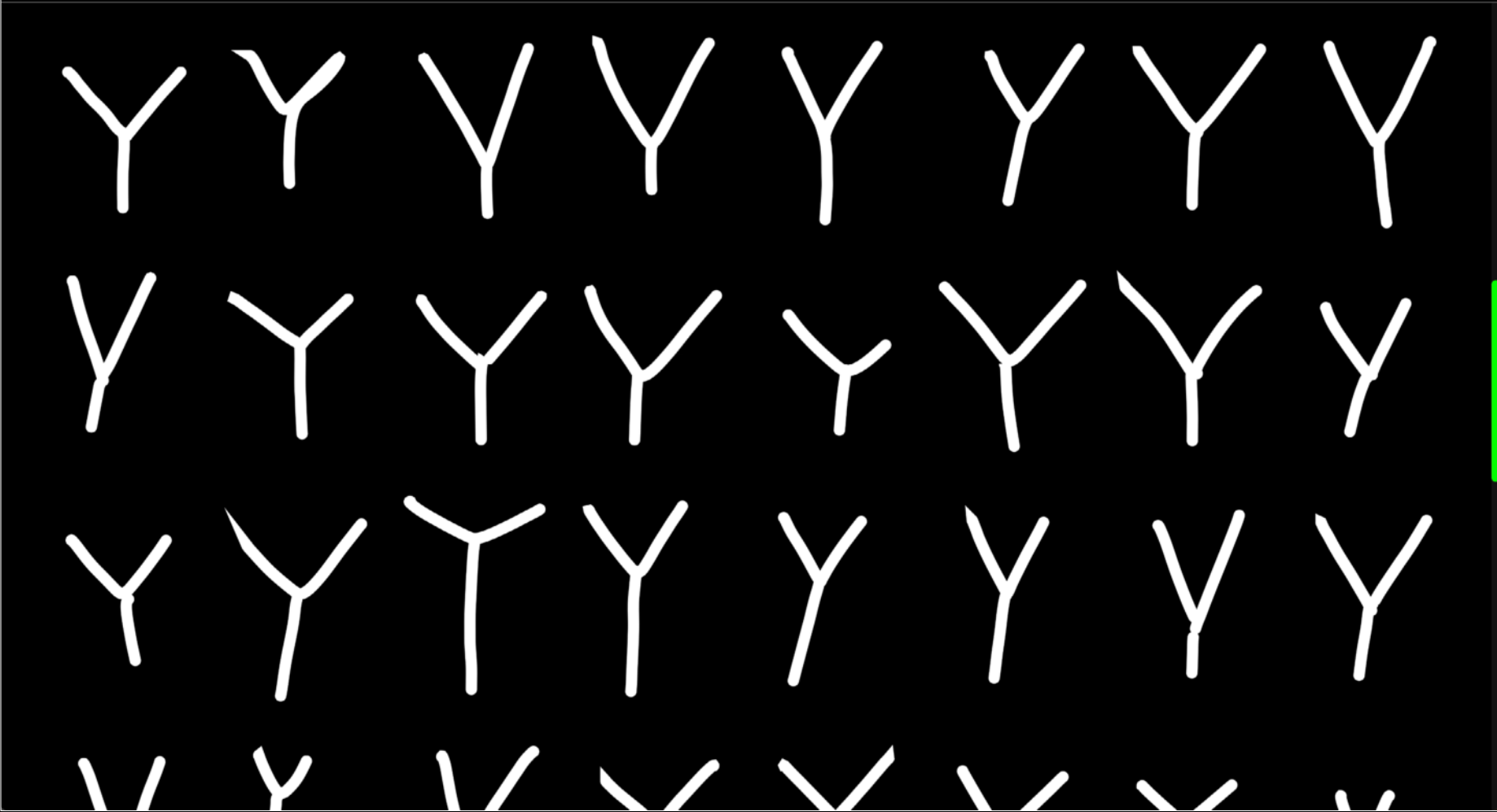
LETTER: X
CASE: ALL
COLUMNS: 8

TAP AND TURN

SCROLL

Matrix

110 OF 2043 LETTERS



Detected Pharma

Region: Albanteich Muehlegraben
Location: 47.554N, 7.604E
Date: 2014-06-17

Antibiotics: 0.025 µg/L

Sulfamethoxazol

Antiepileptics: 0.214 µg/L

Carbamazepin
Gabapentin
Lamotrigin

Analgesics: 0.000 µg/L

Psychotropic: 0.011 µg/L

Venlafaxin

Antidiabetic: 0.470 µg/L

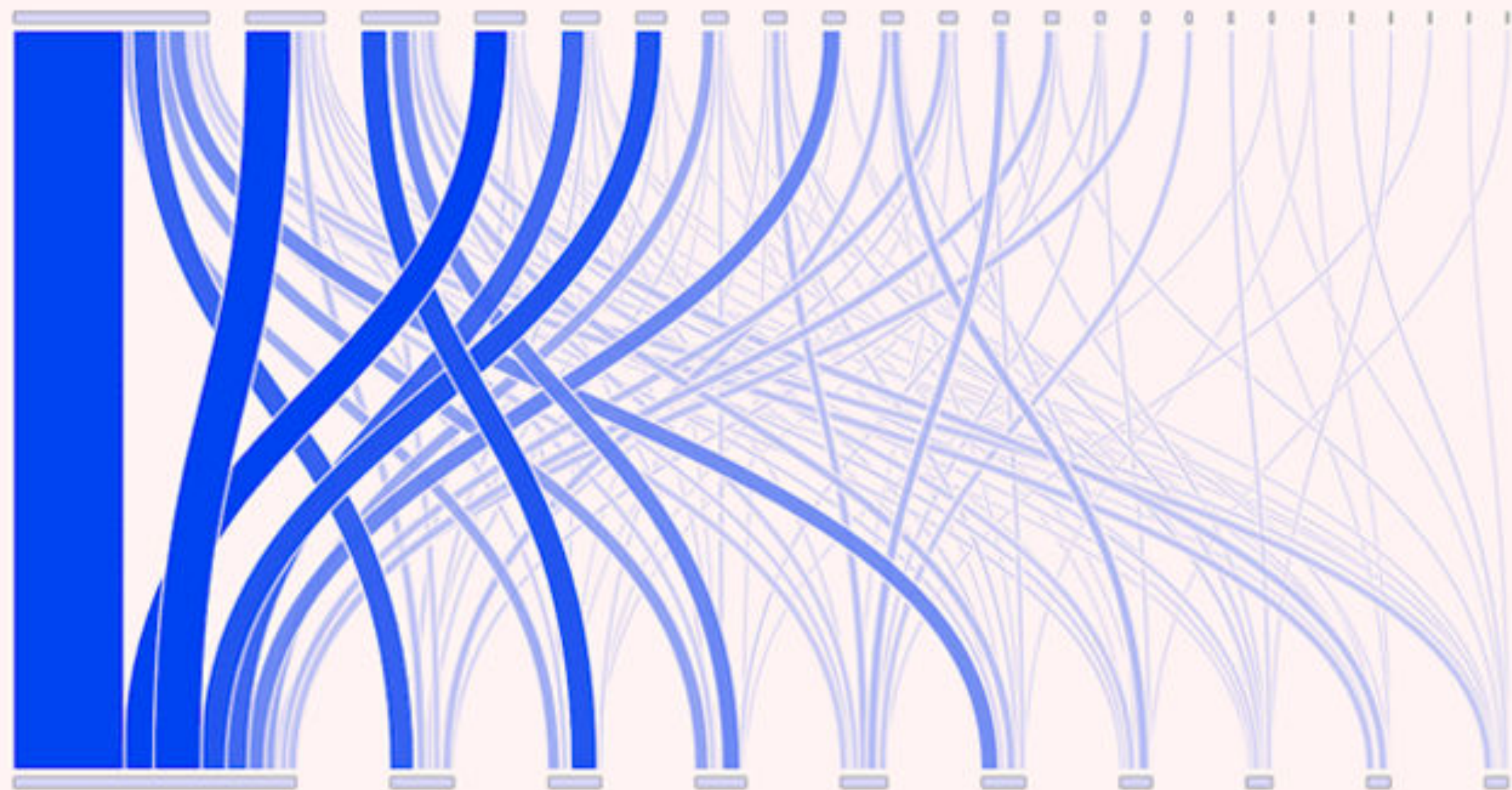
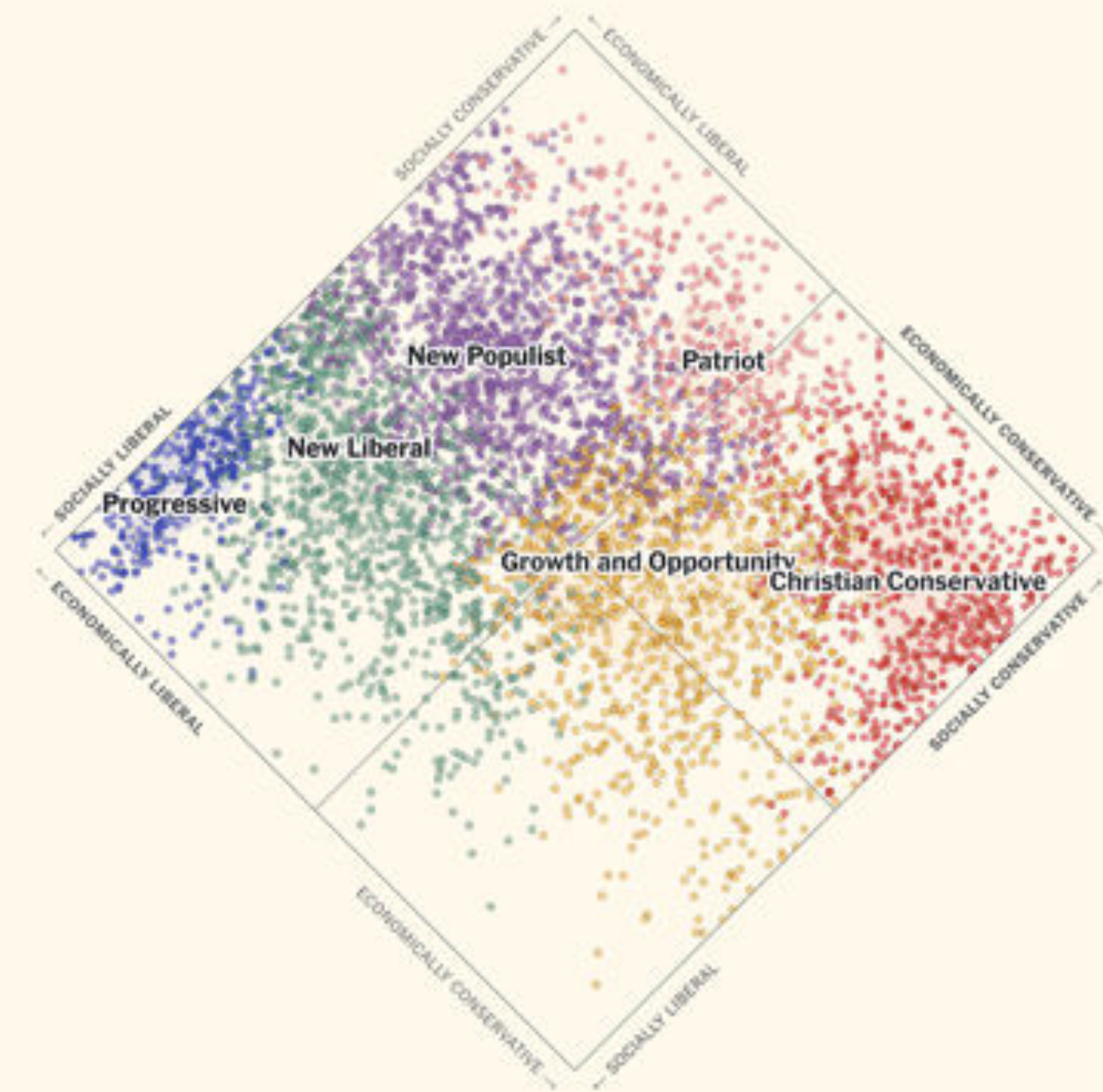
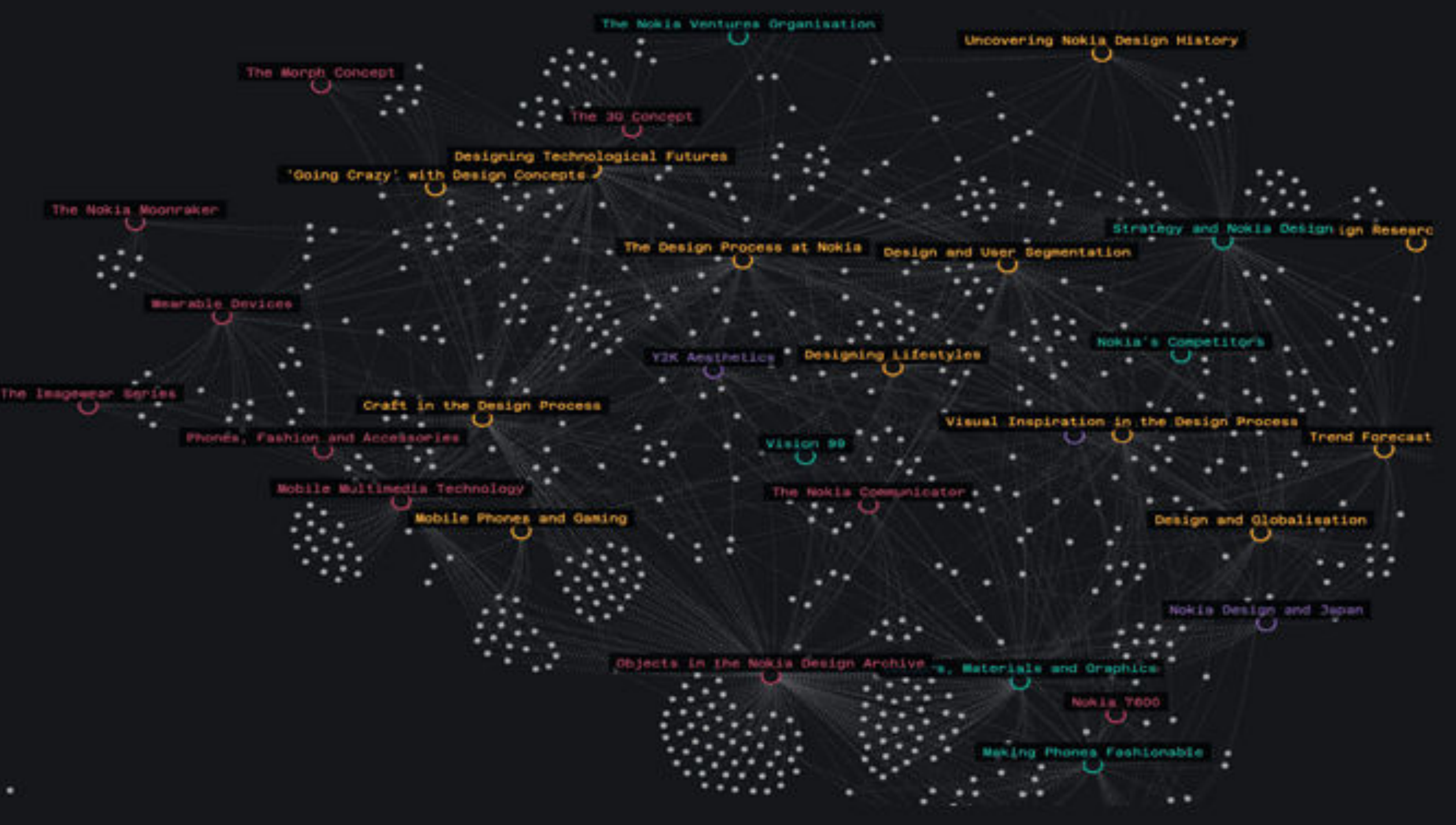
Metformin

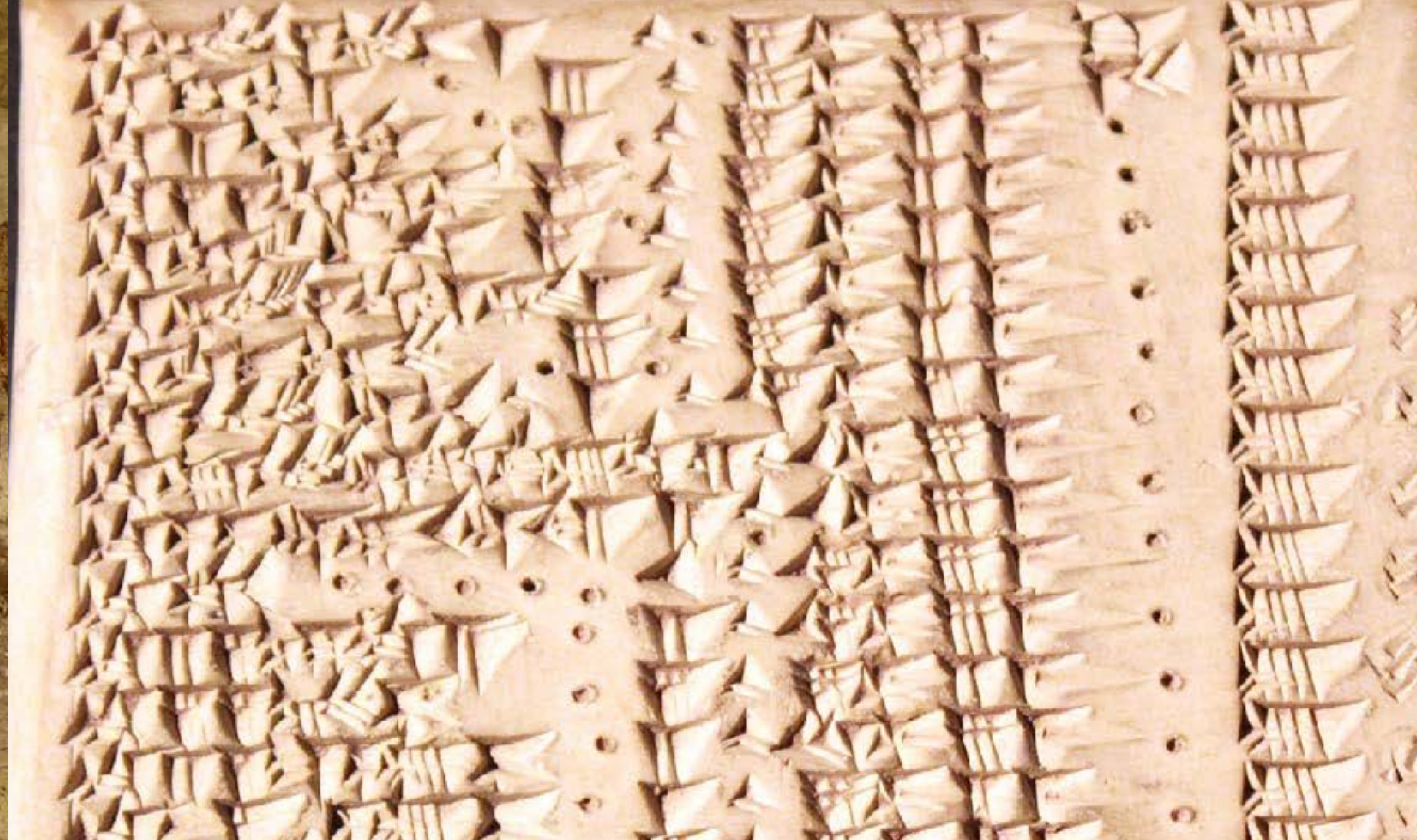
Antihypertensives: 0.107 µg/L

Aliskiren
Hydrochlorothiazid
Metoprolol
Valsartan

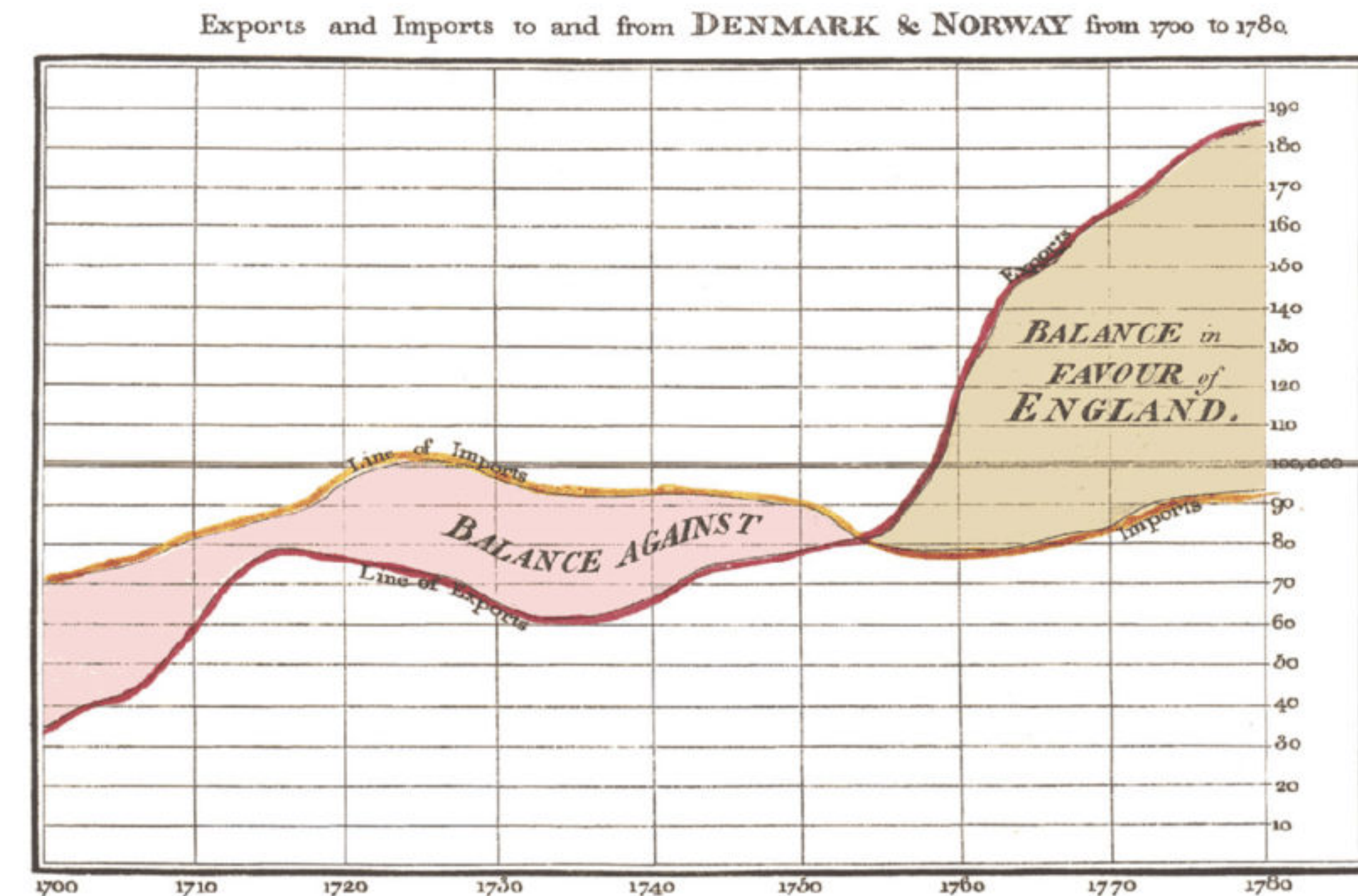
Angela Ammann & Anna Rieder

A BRIEF HISTORY OF DATA VISUALIZATION





WILLIAM PLAYFAIR

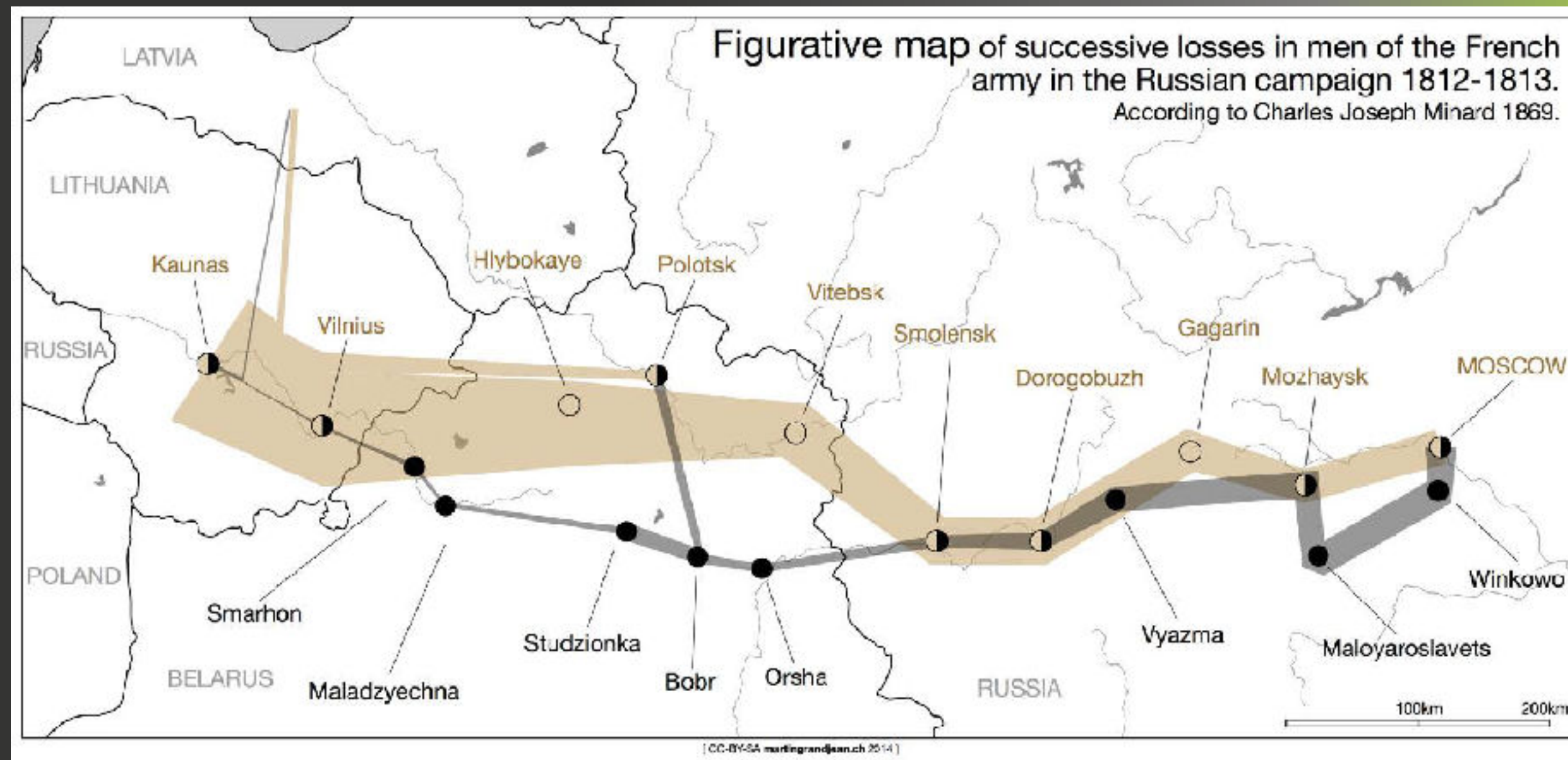


The Bottom line is divided into Years, the Right hand line into £10,000 each.
Published as the Act directs, 1st May 1786. by W. Playfair. No. 502, Strand, London.

Used abstract shapes, such as lines, bars and slices, to show numbers that had nothing to do with physical space, such as trade and debt over time.

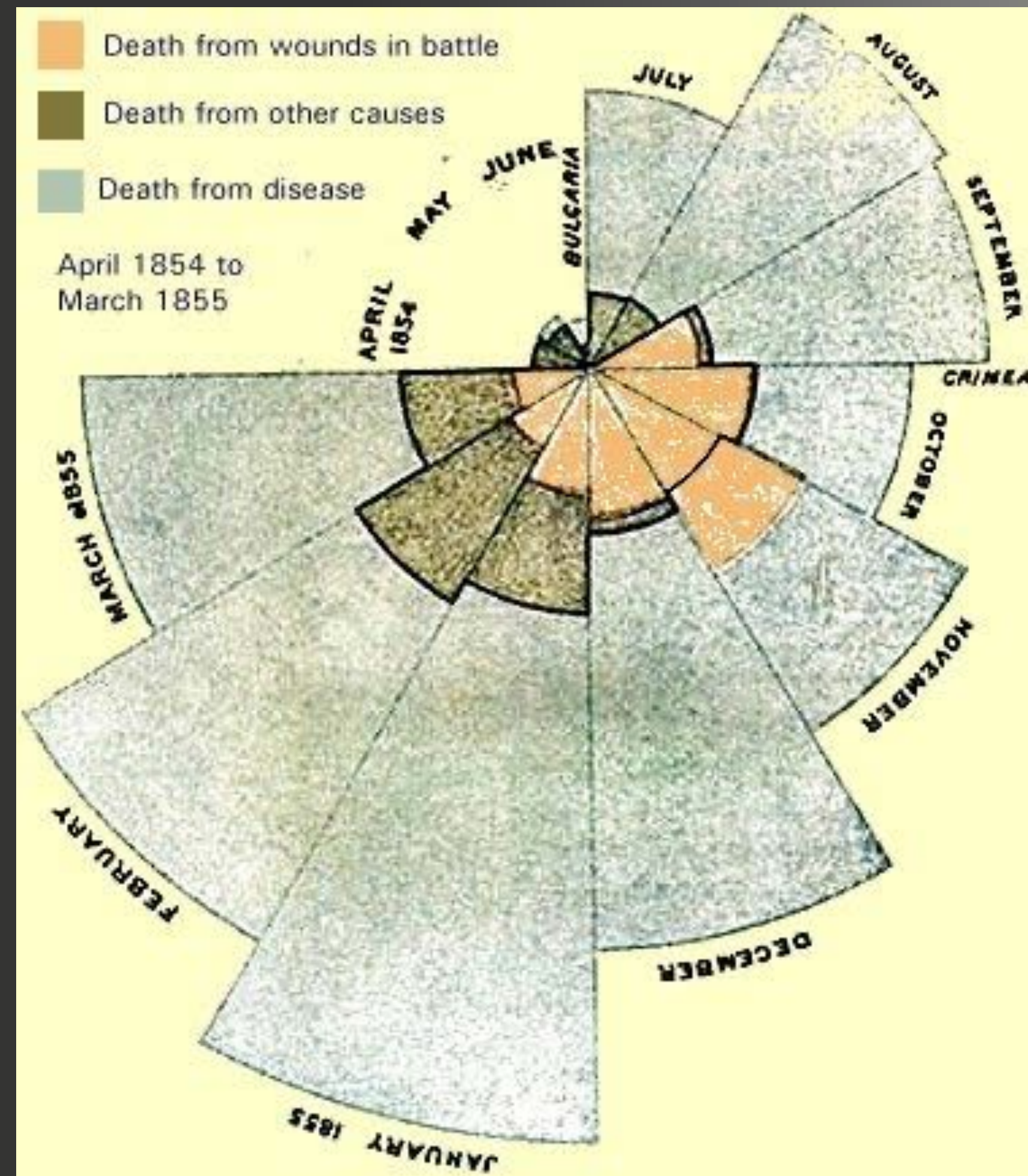
- line chart (1786)
- bar chart (1786)
- pie chart (1801)

CHARLES MINARD



In 1869, Charles Minard mapped Napoleon's 1812 Russian campaign by showing the army's size, route, direction, key locations, temperature and dates together in one image.

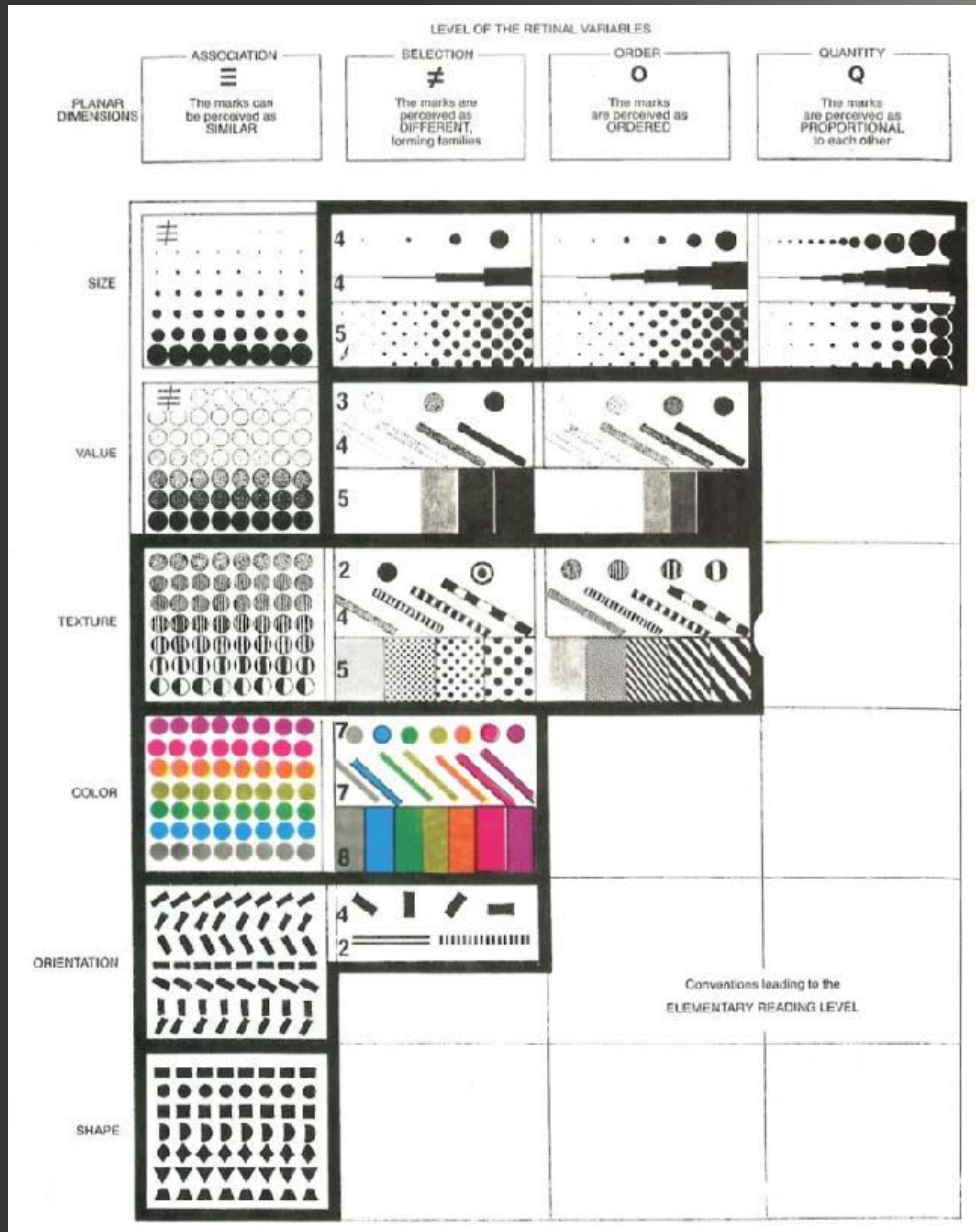
FLORENCE NIGHTINGALE



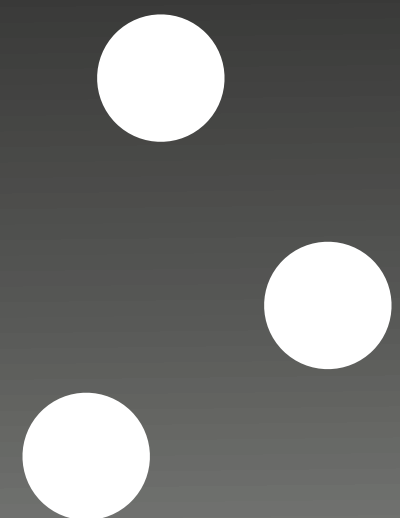
In 1854 he proved that cholera spread through water, not air, like previously believed. He created a map where he showed that deaths clustered around Broad Street water pump.

JACQUES BERTIN

In 1967 he created “Semiology of Graphic”, which became the first the first systematic theory of visual encoding.



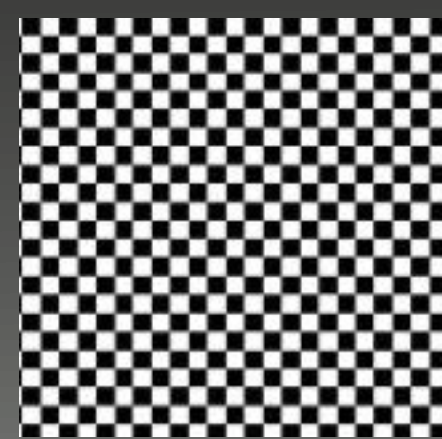
ENCODINGS



Position



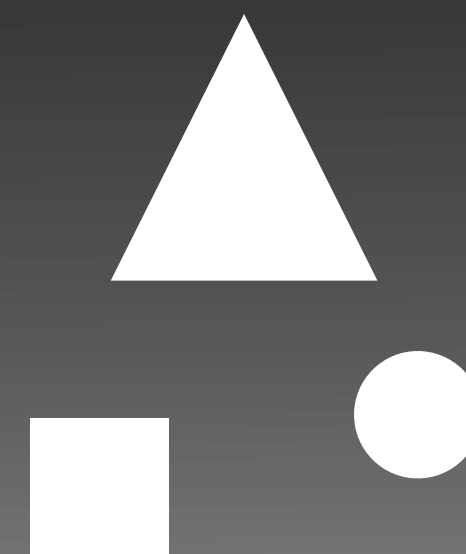
Size



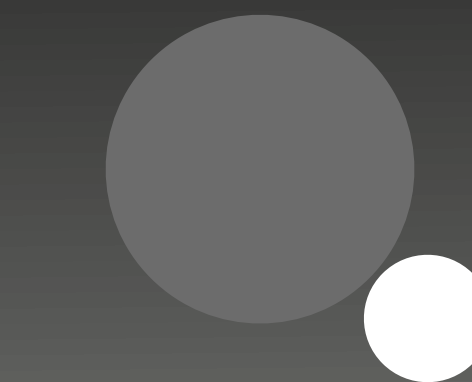
Texture



Slope



Shape



Value



Color

TIMELESS PRINCIPLES

1

Maximize ink, minimize 'chartjunk'

2

Context is everything

3

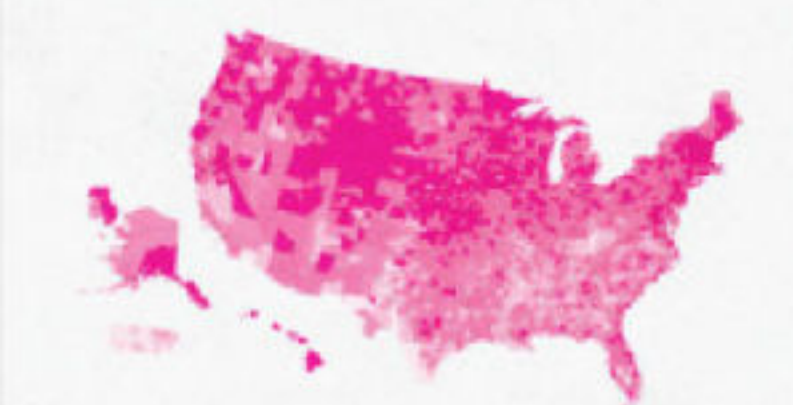
Above all else show the data

4

Use "Sparkles"

READING, WRITING, AND EARNING MONEY

The latest data from the U.S. Census's American Community Survey paints a fascinating picture of the United States at the county level. We've looked at the educational achievement and the median income of the entire nation, to see where people are going to school, where they're earning money, and if there is any correlation.



A HIGH SCHOOL GRADUATES 65% 75% 82% 88%

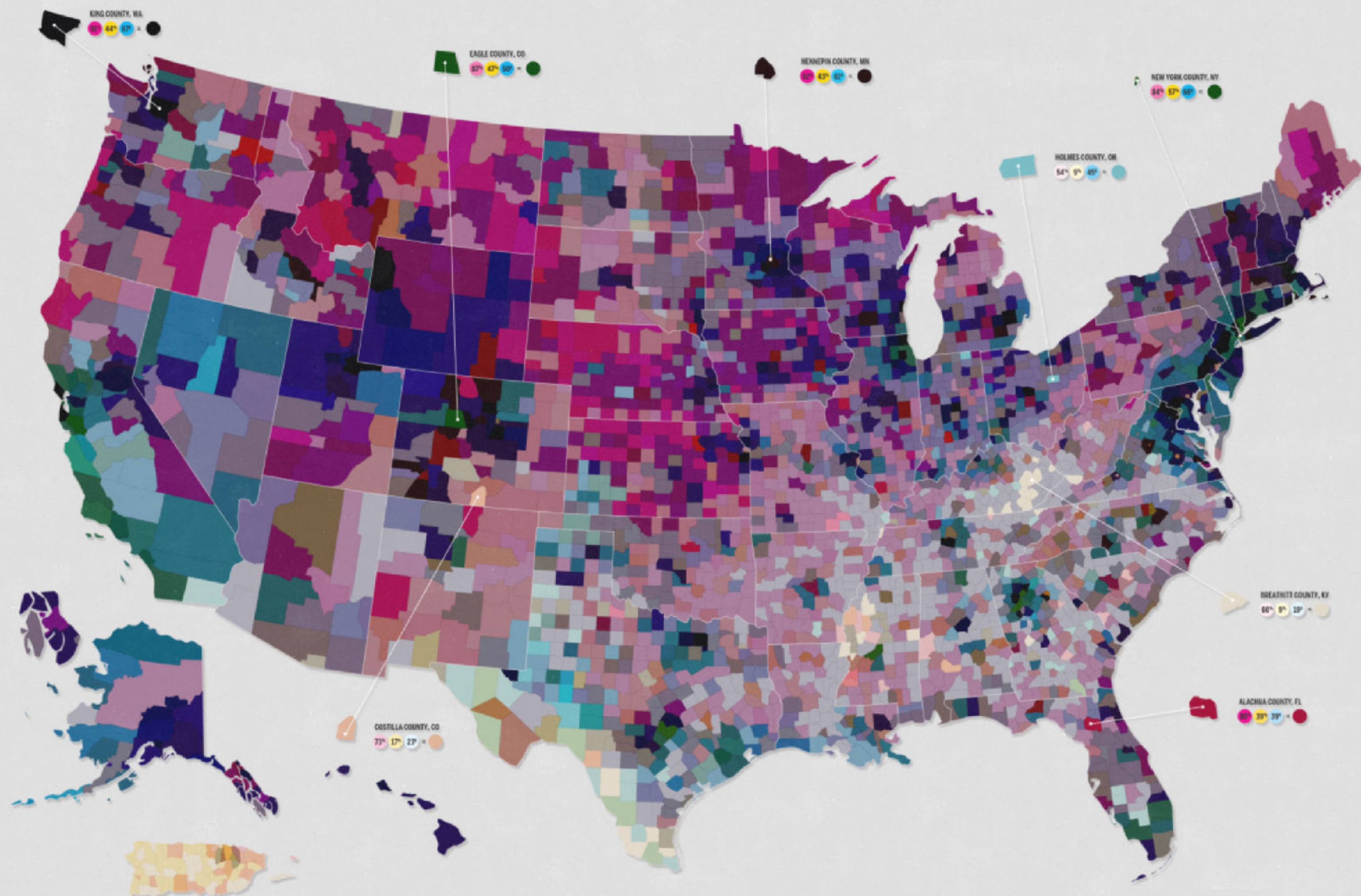


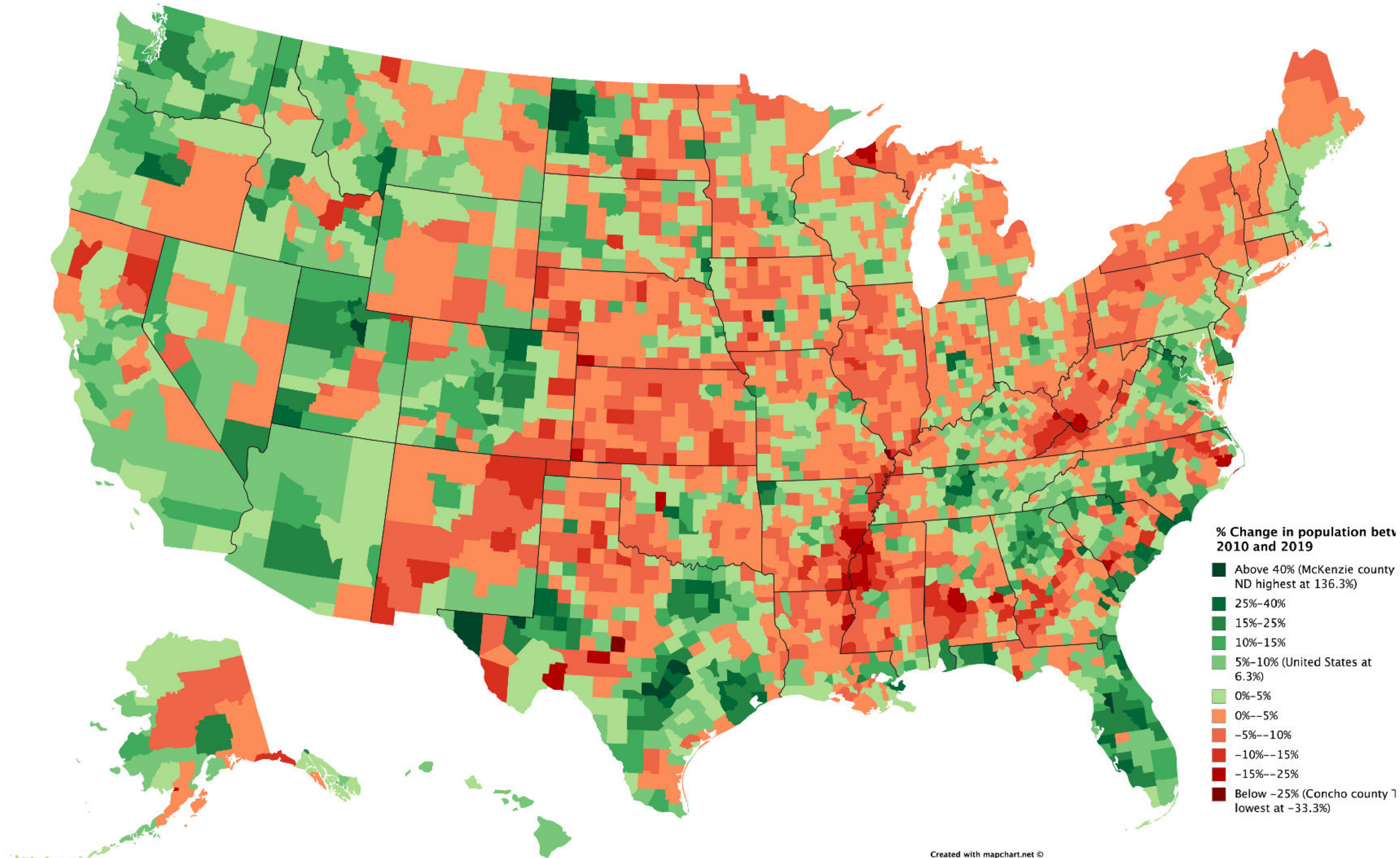
B COLLEGE GRADUATES 15% 22% 30% 40%

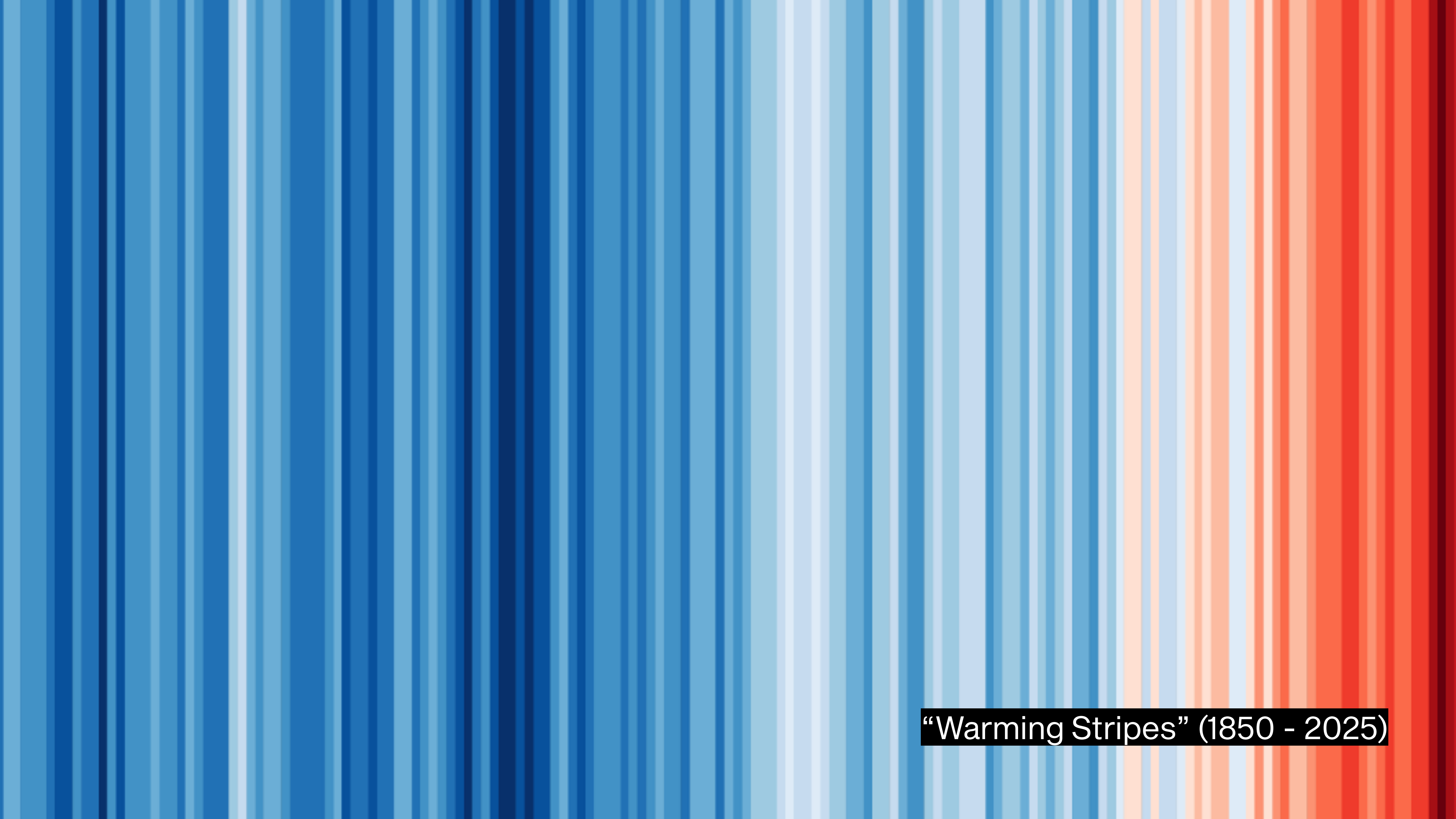


C MEDIAN HOUSEHOLD INCOME 25K 40K 50K 65K

The map at right is a product of overlaying the three sets of data. The variation in hue and value has been produced from the data shown above. In general, darker counties represent a more educated, better paid population while lighter areas represent communities with fewer graduates and lower incomes.







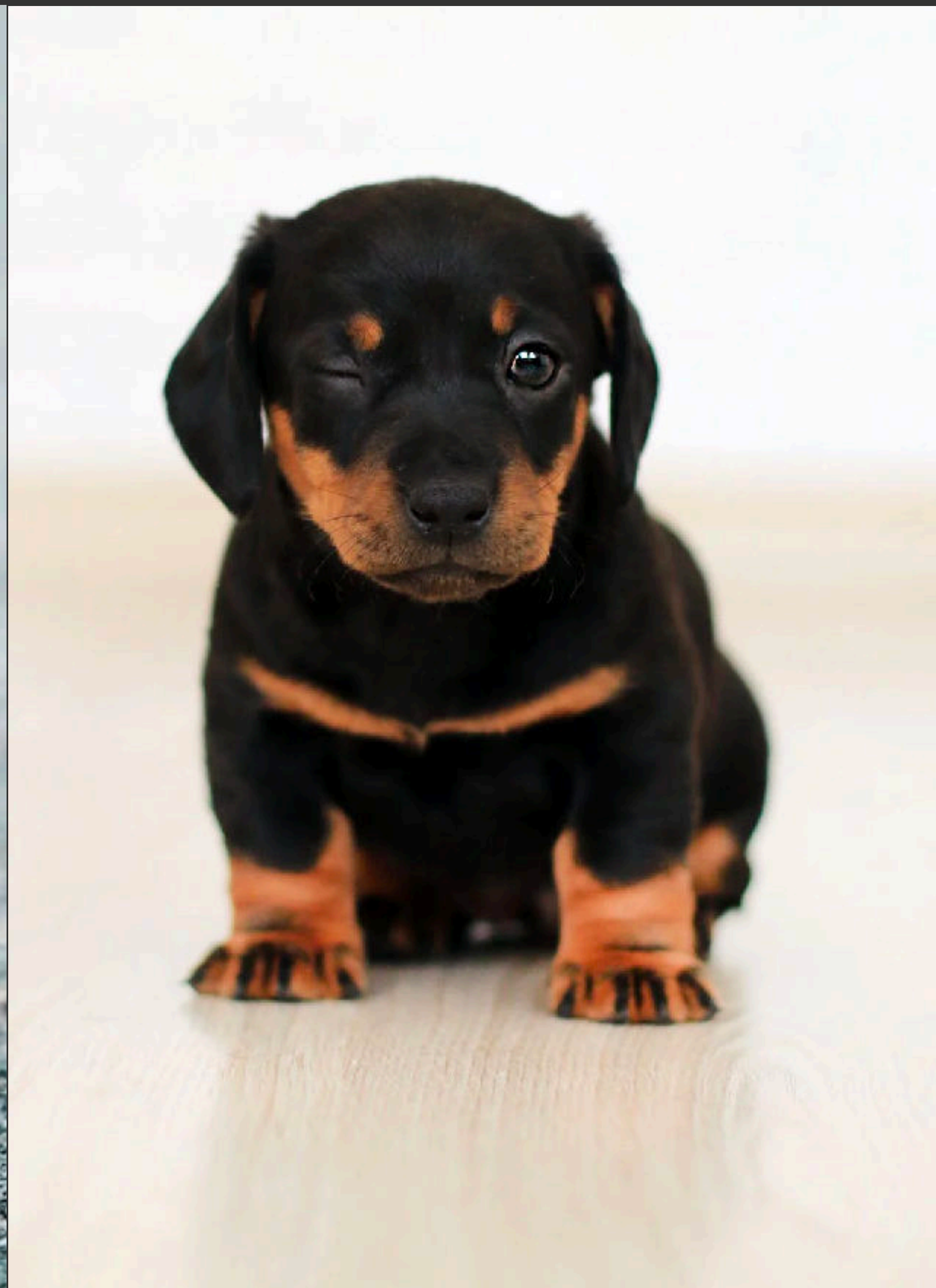
“Warming Stripes” (1850 - 2025)



DATASET AS DATA VISUALISATION



How would you explain
the difference between
a cat and a dog?





WHY DATASET?

~~"A cat has pointy ears."~~

→ So does a Pomeranian.

~~"A dog is bigger than a cat."~~

→ Not a Chihuahua.

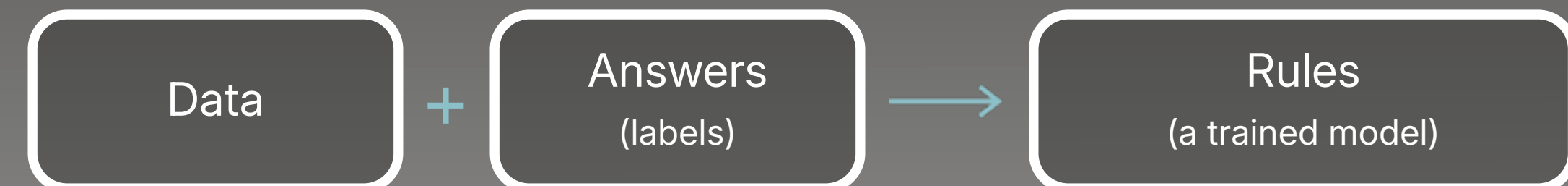
~~"A cat has whiskers."~~

→ So does a dog.

CLASSIC PROGRAMMING



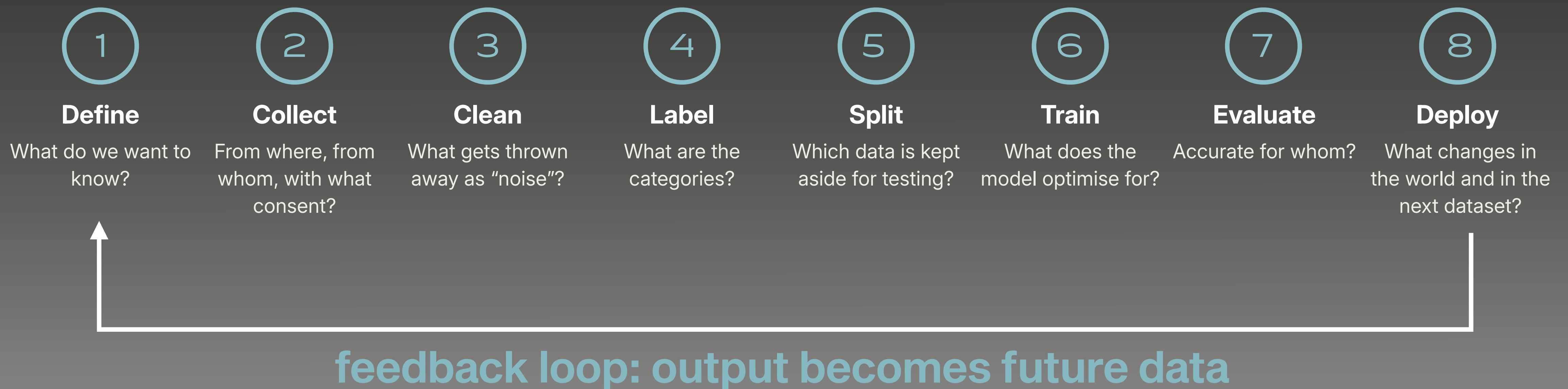
MACHINE LEARNING



WHY DATASET?

- Helps to recognize complicated patterns
- Allows models to generalize
- Explores different facets of a problem
- Captures change over time

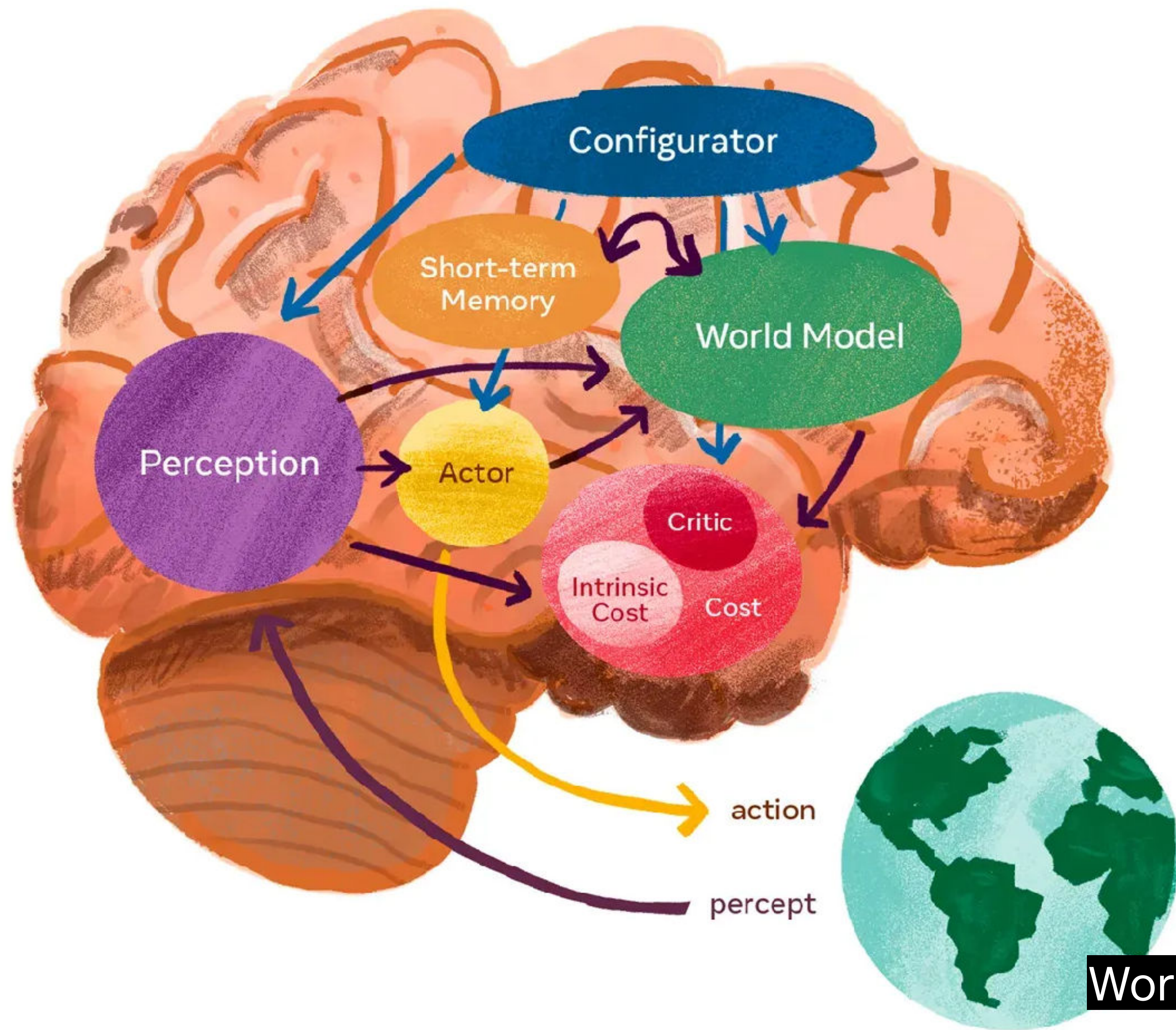
FROM WORLD TO MODEL



train 70%

val 15%

test 15%



World Model (Meta AI)

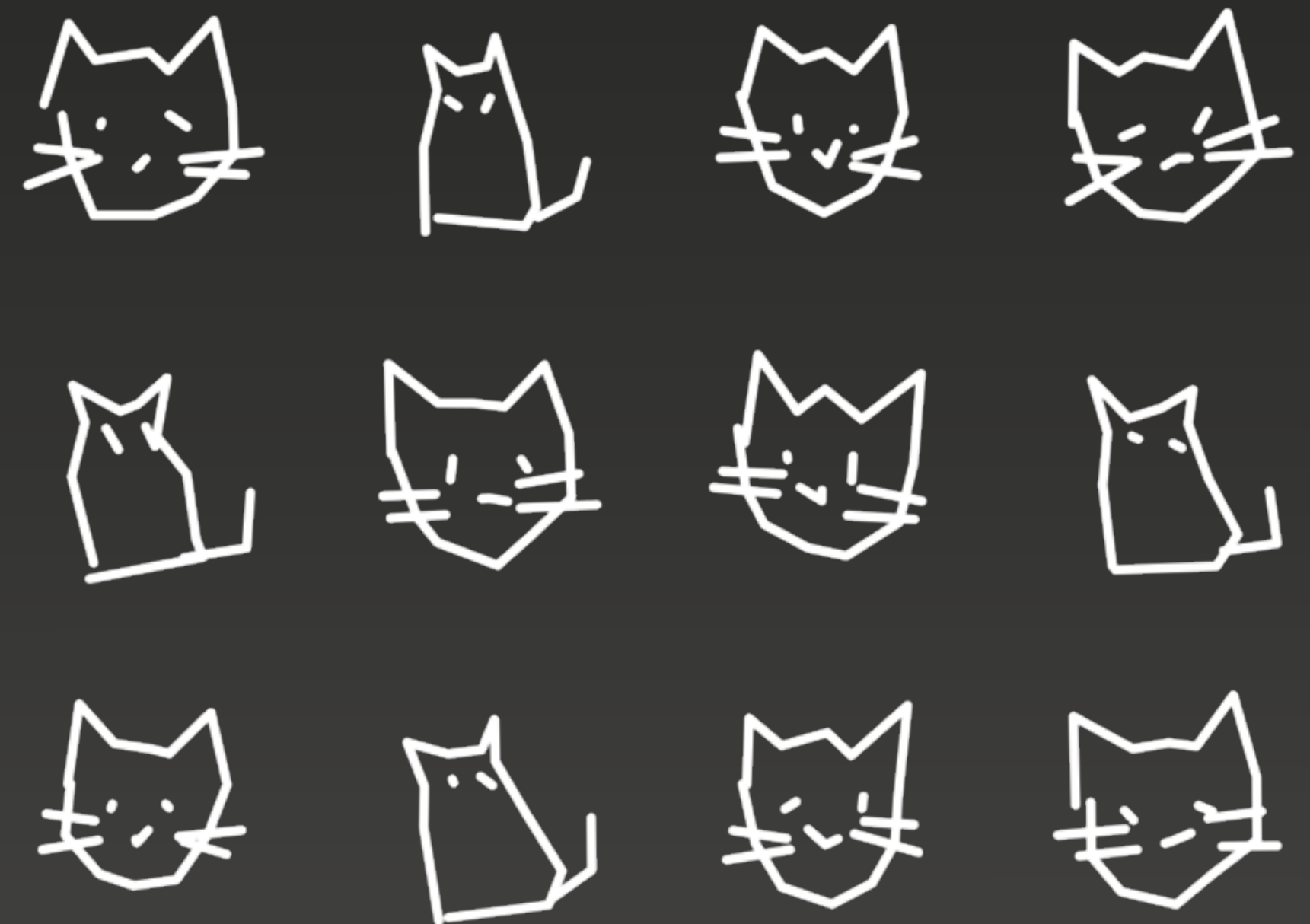
QUICK, DRAW!

Go to:

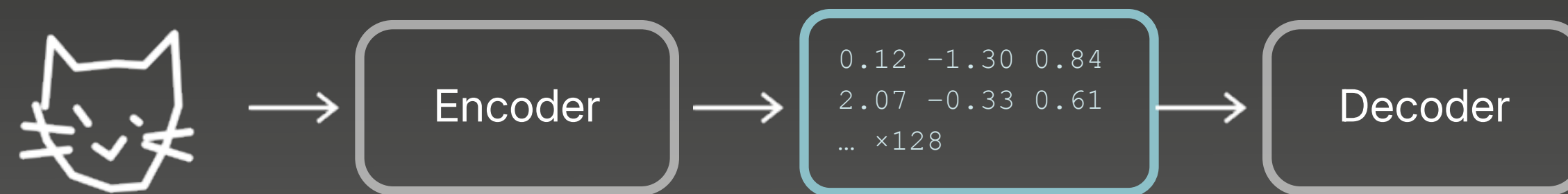
quickdraw.withgoogle.com

Analyse the data:

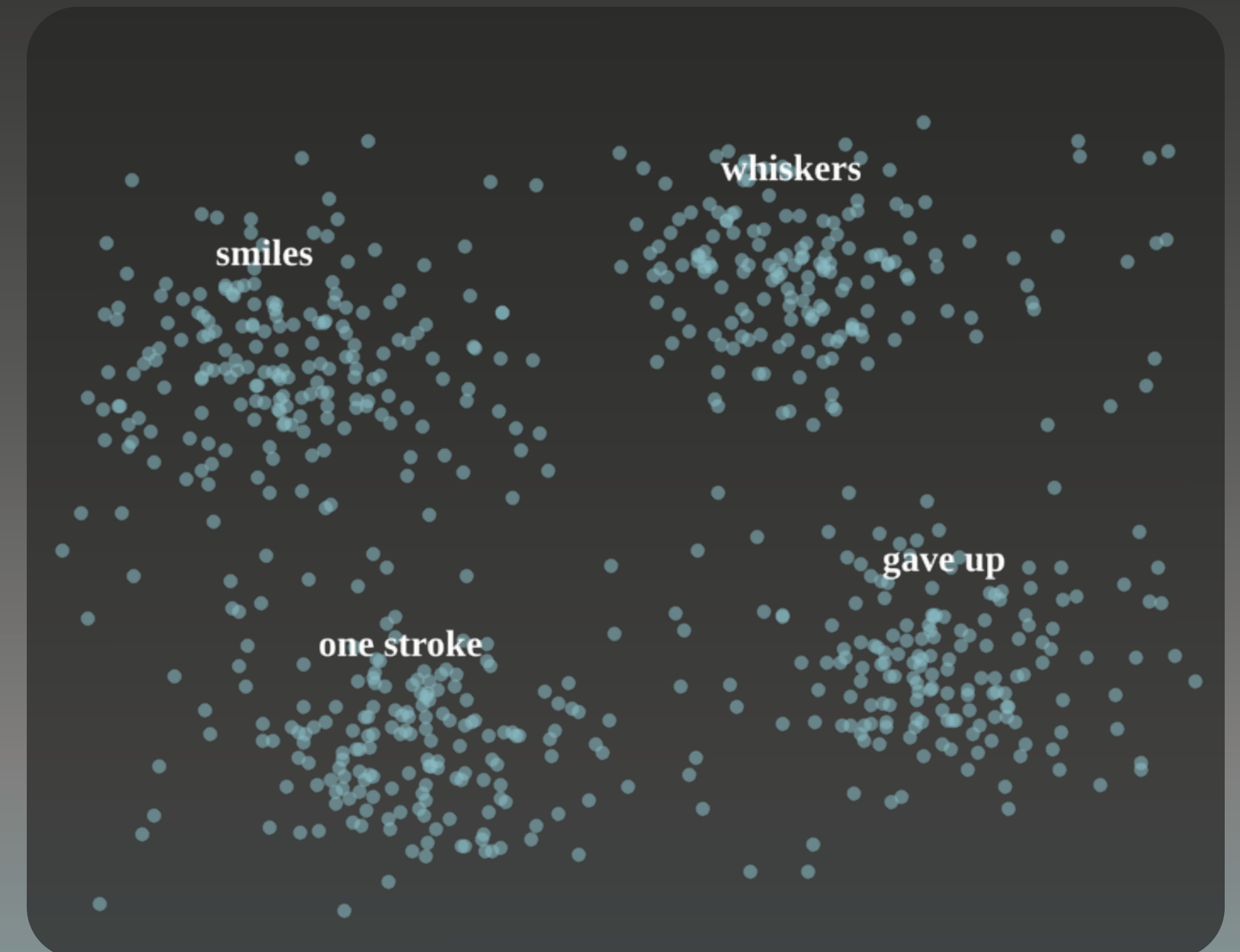
[https://github.com/googlecreativelab/
quickdraw-dataset](https://github.com/googlecreativelab/quickdraw-dataset)



LATENT SPACE



Sketch-RNN compresses each drawing into a **latent vector** of 128 numbers and can redraw it from them. **t-SNE** squeezes those 128 dimensions onto a 2D map where similar drawings sit together.

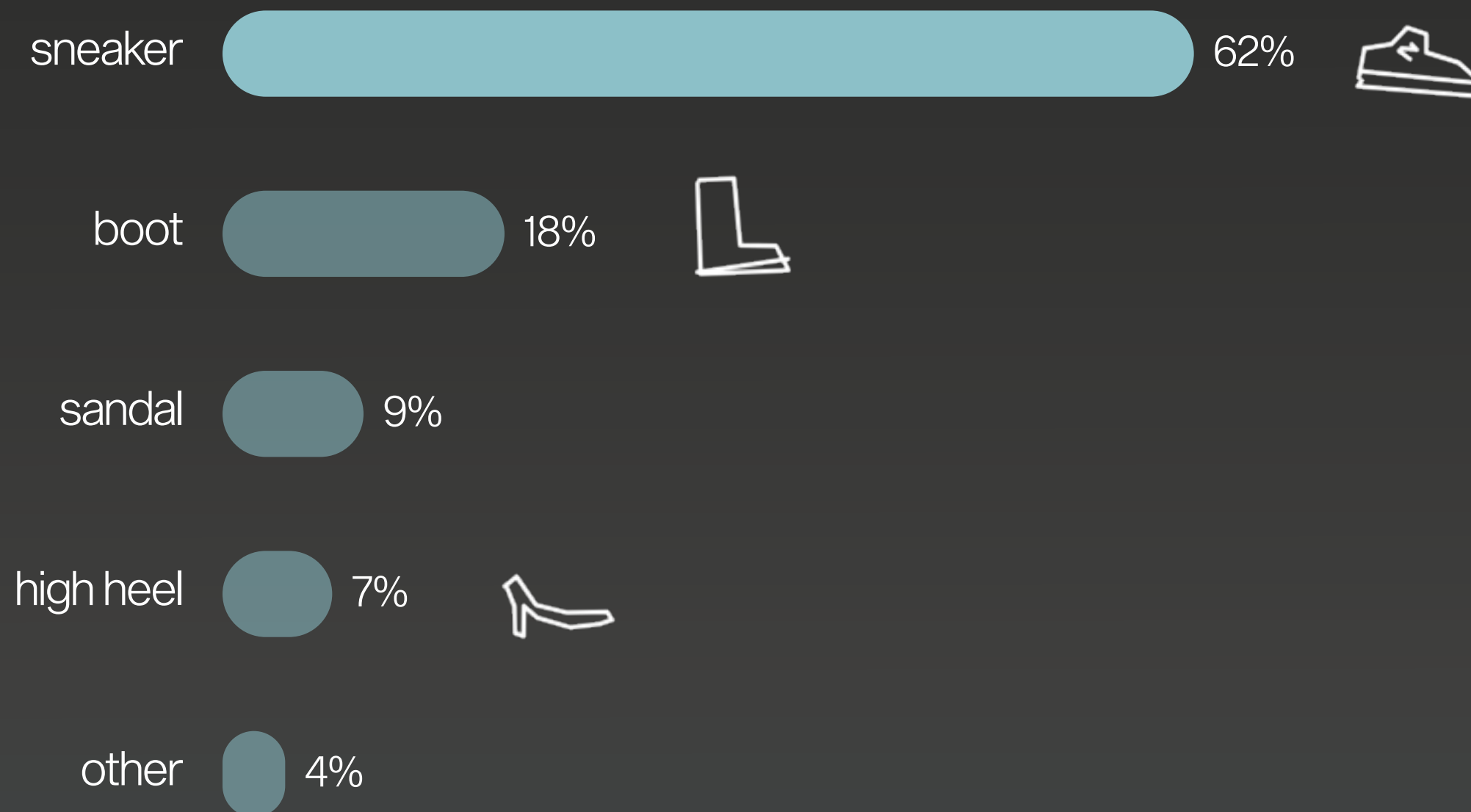


ANATOMY OF A DATASET

key_id	word	recognized	countrycode	timestamp	drawing
5629180...	cat	true	CH	2017-03-01 14:22	
4814377...	cat	true	JP	2017-03-02 09:05	
6021554...	shoe	true	US	2017-03-02 21:47	
5107236...	shoe	false	FR	2017-03-03 11:12	
4930871...	power outlet	true	DE	2017-03-04 18:30	

WHAT IS TYPICAL?

“DRAW A SHOE” — SHARE OF DRAWINGS



A model learns what appears most often leading to a data bias. Also, think about the people who participated: the ones with a device, internet access and free time.

ANATOMY OF A DATASET

Direction matters



counter-clockwise
(US)

clockwise (Japan)

The direction varies by country, shaped by how people first learned to write.

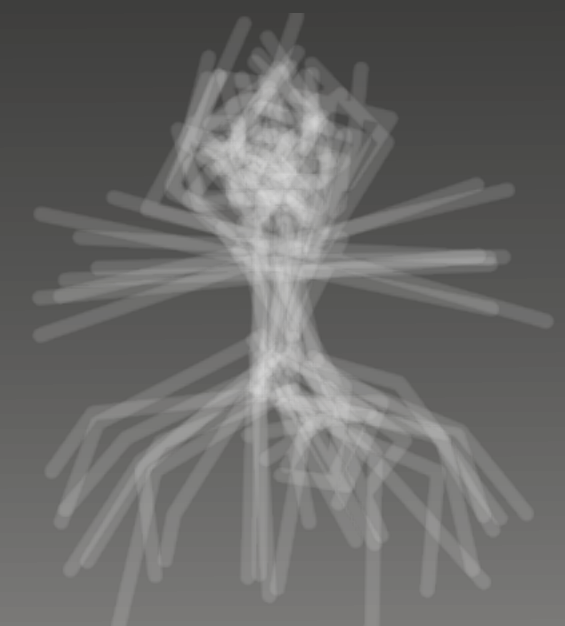
Visual averages

Germany

Japan

Thousands of drawings per country reveal local defaults.

Edge cases



Average results blur: some categories cannot be “averaged”

OPEN DATASETS

IMAGE

ImageNet

14M photos divided in 21k categories.

IMAGE

COCO

330k of everyday scenes, and objects

IMAGE

CIFAR

80M tiny Image dataset

HANDWRITING

MNIST

70,000 handwritten digits by US Census employees and students.

SPEECH

Common Voice

Volunteers reading sentences aloud in 100+ languages.

SOUND

Freesound Labs

Community library of Creative Commons sound recordings.

TEXT

Common Crawl

Regular snapshots of the public web since 2008.

IMAGE + TEXT

LAION-5B

5.85 billion image—caption pairs scraped from the web.

SWITZERLAND

opendata.swiss

Open government data from federal, cantonal and city offices.

BASEL

data.bs.ch

Open Data Basel-Stadt: mobility, environment, population, water...

COMMERCIAL WEBSITES

Hugging Face · Kaggle

Hubs with hundreds of thousands of datasets, many with dataset cards.

IMAGE + ANNOTATION

Open Images

9M images annotated with image-level labels, object bounding boxes, etc.

DATASET AUTOPSY

1. Pick one dataset from the list below.
2. Spend 15 minutes on the website.
3. Answer the three questions.
 1. Who made the dataset and why?
 2. What is in it and what is missing?
 3. Who labelled it, and how?

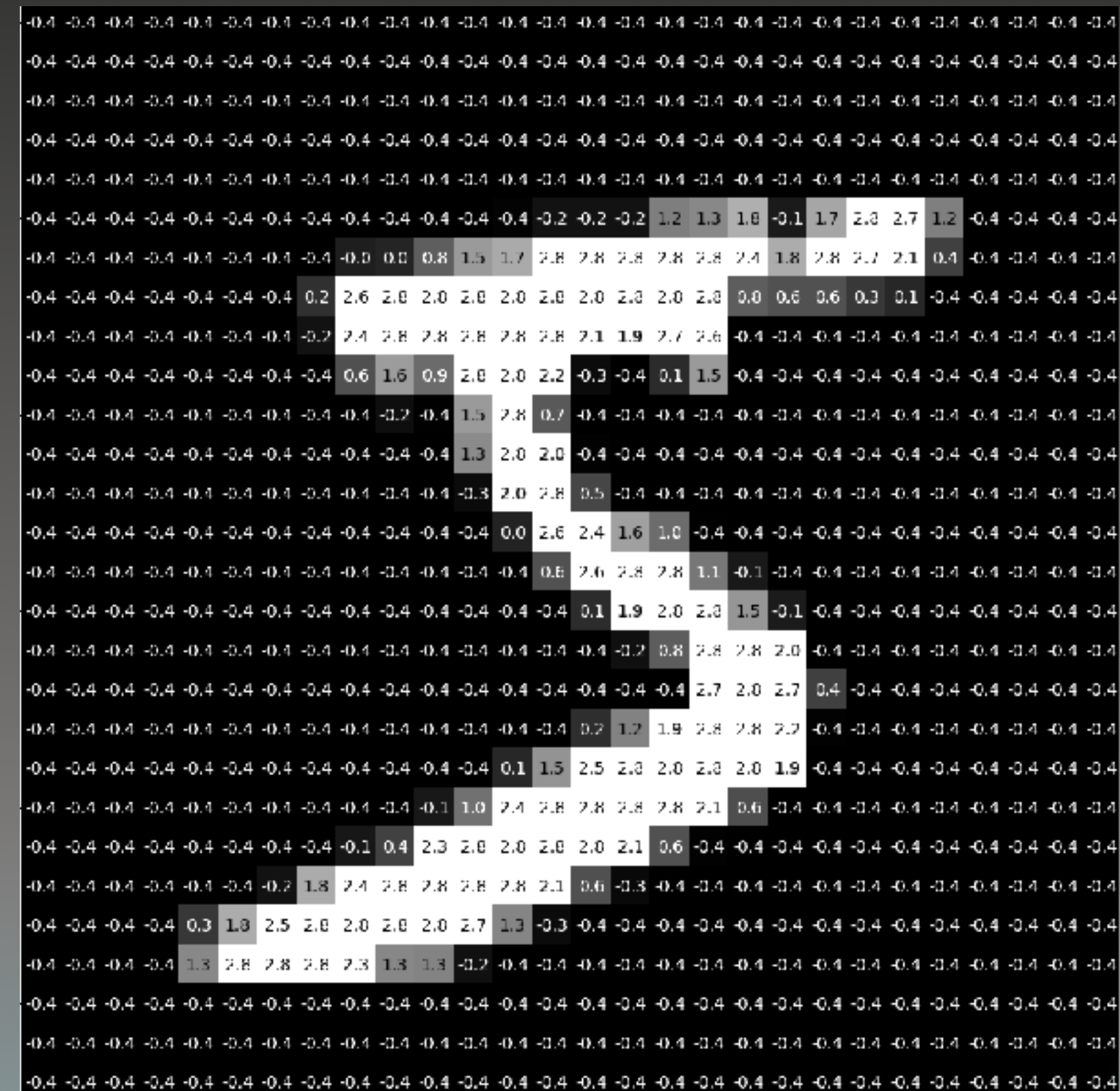
Datasets

CIFAR

MNIST

Common Voice

COCO





IM GENET

ImageNet (2012 - 2021)

Bisexual, bisexual person

A person who is sexually attracted to both sexes

- supernumerary (0)
- inhabitant, habitant, dweller, denizen, indweller (485)
- debaser, degrader (1)
- achiever, winner, success, succeder (5)
- contemplative (0)
- Cancer, Crab (0)
- national, subject (18)
- interpreter (0)
- namer (0)
- hoper (0)
- gainer (0)
- buster (0)
- biter (1)
- sensualist (12)
 - cocksucker (0)
 - erotic (0)
 - epicure, gourmet, gastronome, bon vivant, epicurean, foodie (0)
 - voluptuary, sybarite (0)
 - hedonist, pagan, pleasure seeker (1)
 - playboy, man-about-town, Corinthian (0)
 - bisexual, bisexual person (3)
 - hermaphrodite, intersex, gynandromorph, androgyne, epicene, epicene person (0)
 - pseudohermaphrodite (0)
 - switch-hitter (0)
 - wanton (1)
 - light-o'-love, light-of-love (0)
- acquirer (42)
- admirer (2)
- bad guy (0)
- ensor (0)
- deliverer (1)
- rich person, wealthy person, have (11)
- case (14)

Bondwoman, bondswoman, bondmaid

A female slave

268
pictures

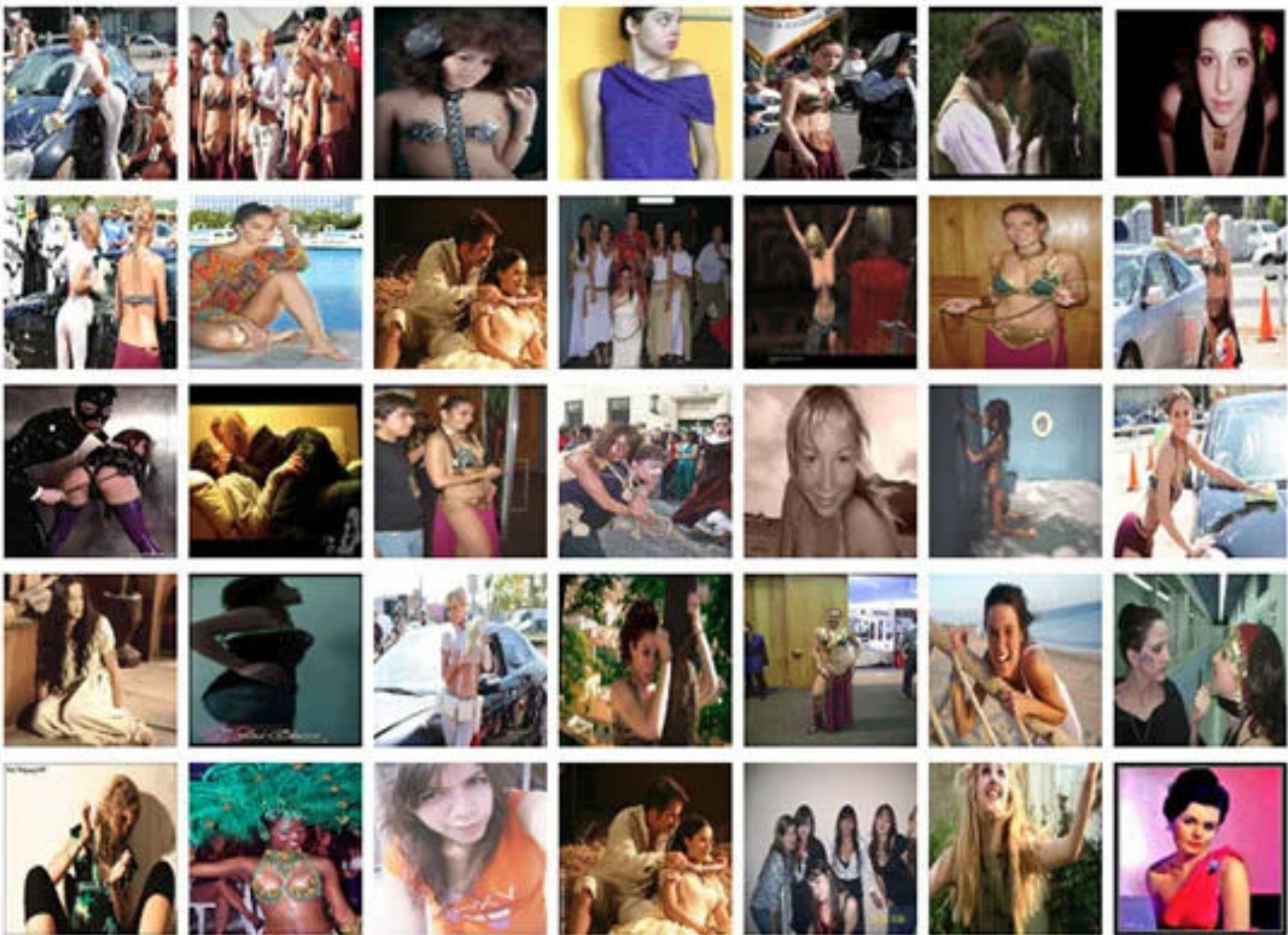
7.01%
Popularity
Percentile

Wordnet
IDs

Treemap Visualization

Images of the Synset

Downloads



*Images of children synsets are not included. All images shown are thumbnails. Images may be subject to copyright.

Prev 1 2 3 4 5 6 7 8 Next

IMAGENET IN NUMBERS

**14M
images**

collected from the web

**21,841
categories**

taken from the WordNet hierarchy

**~50,000
crowd workers**

on Mechanical Turk, from 167
countries

**45%
from the US**

of geolocated images

2009

Published by Fei-Fei Li's team

2010–17

Annual ImageNet challenge

2012

AlexNet wins — deep
learning takes off

2019

“Excavating AI” & ImageNet
Roulette expose offensive
person labels; ~600,000
images removed

2021

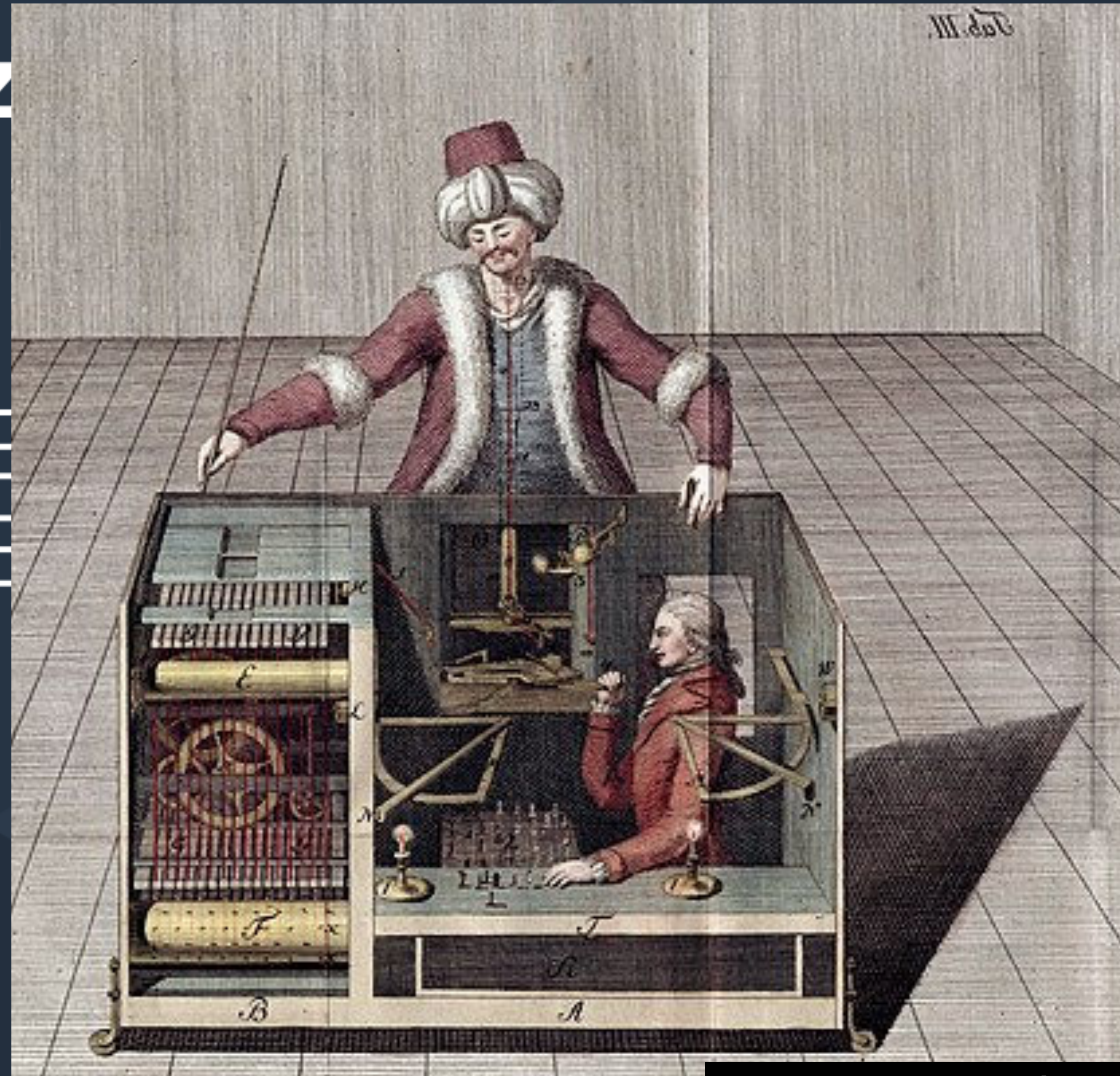
Faces in the dataset blurred

WHO FEDS THE MODELS?

- Scraping
- User Data
- Undisclosed sources
- Pirated material
- Generated & licensed material
- Bought & scanned material

Amazon

Access a



ce



Amazon Mechanical Turk (2005 - 2026)



Anthropic Project Panama



The AI factory: the rewiring of India's tech industry



People Sans People - Synthetic Data

English (United States) ▾

Updated: March 13, 2026

How your data is used to improve model performance

Learn more about how OpenAI uses content from
our services to improve and train our models.

DATA BIAS

Data bias in machine learning is a type of error in which certain elements of a dataset are more heavily weighted and/or represented than others.

DATA BIAS

DATA GENERATION

Historical

Some parts of the world do not have sufficient data that survived

Representation

Some groups are under-sampled or missing entirely.

Measurement

What is measured is a proxy?

MODEL BUILDING

Aggregation

One model for groups that behave differently.

Learning

The training goal rewards accuracy on the majority.

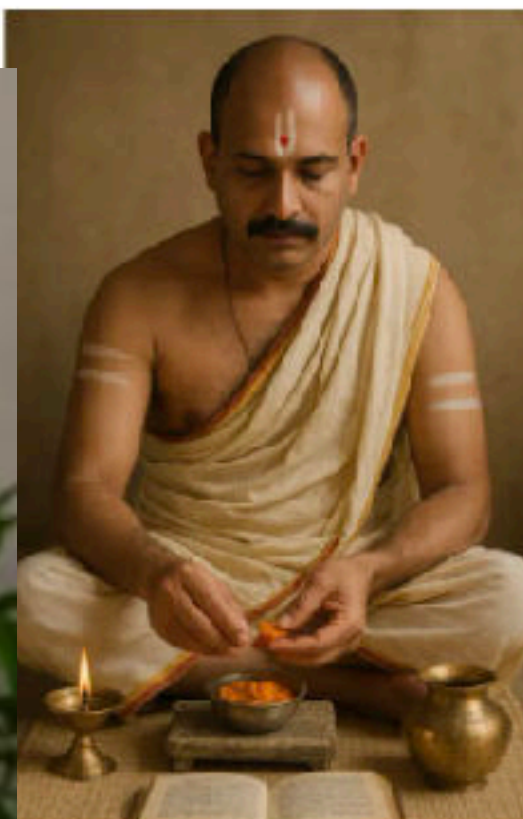
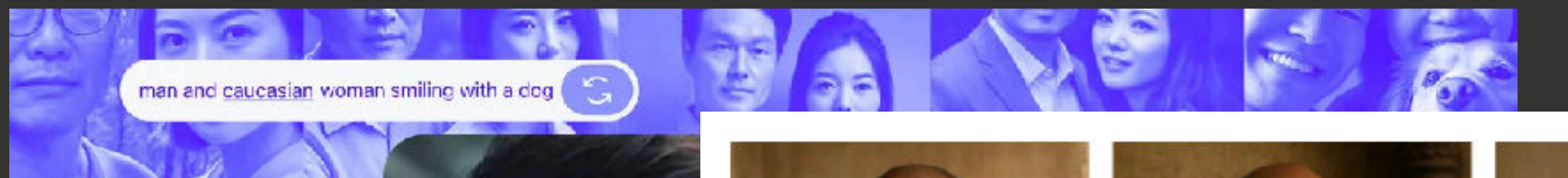
Evaluation

The test set doesn't look like the people who will use it.

DEPLOYMENT

Deployment

The system is used for something it wasn't built for.



ntly produced

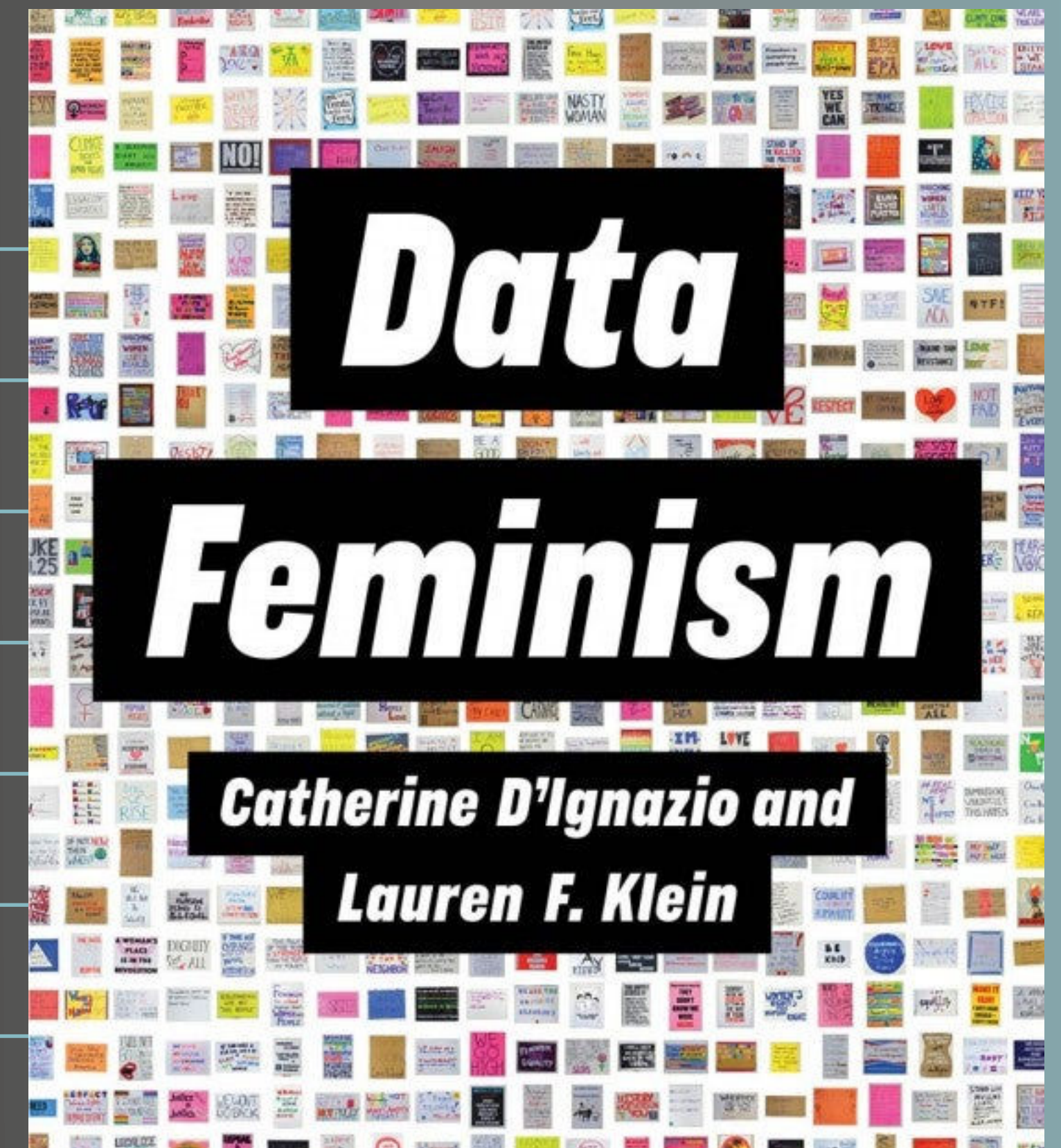


Buolamwini & Gebru, "Gender Shades", 2018

DATA FEMINISM

Catherine D'Ignazio & Lauren Klein (2020): seven principles for doing data science with power in mind.

- | | | |
|---|---|--|
| 1 | Examine power | Who collects, who is counted, who benefits? |
| 2 | Challenge power | Use data to push back, not only to describe. |
| 3 | Elevate emotion & embodiment | Feelings and bodies are ways of knowing, too. |
| 4 | Rethink binaries & hierarchies | Question the categories you inherit. |
| 5 | Embrace pluralism | Many perspectives make better knowledge. |
| 6 | Consider context | Data is never raw. |
| 7 | Make labour visible | Credit the people who collect, clean and label it. |



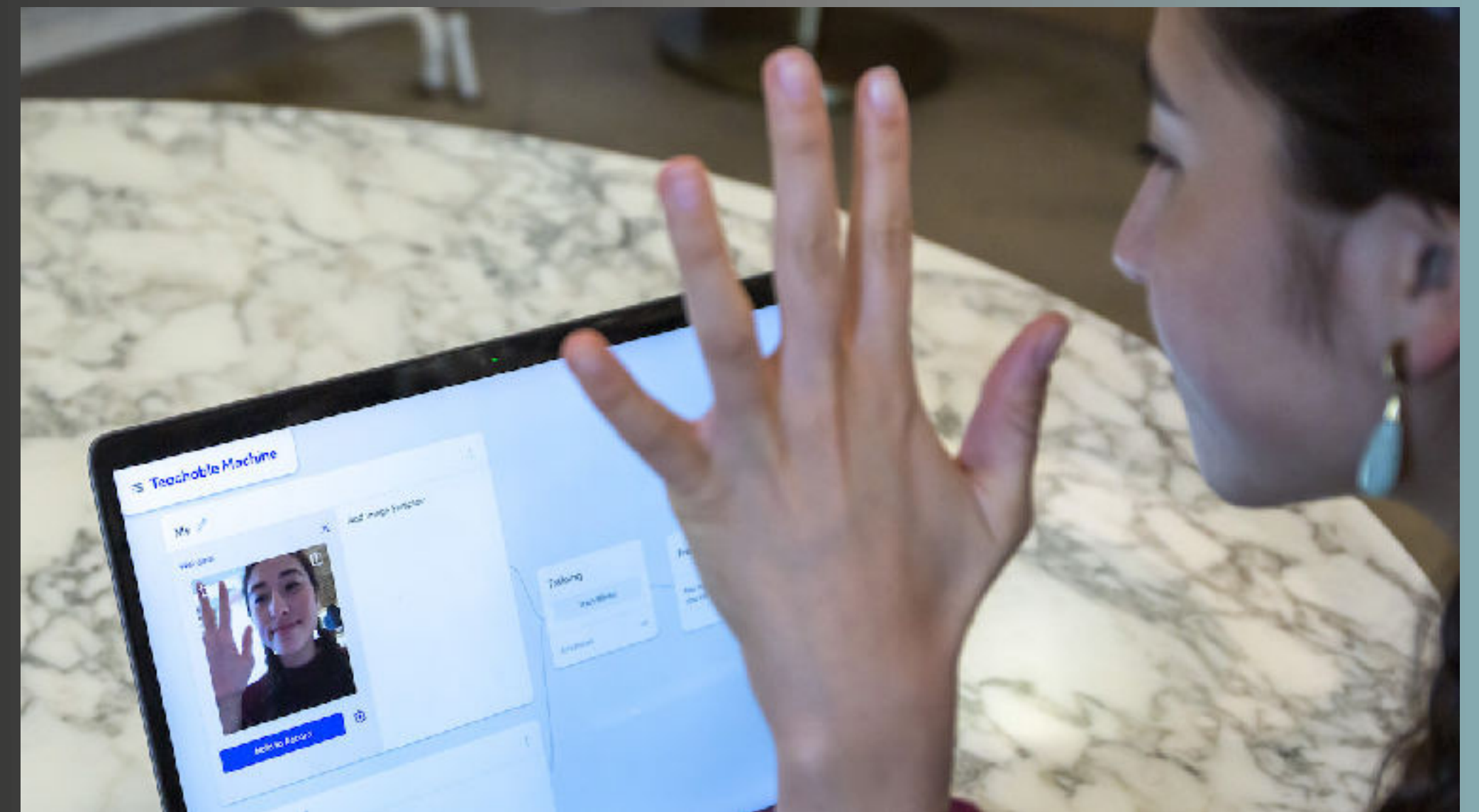
TEACHABLE MACHINE

teachablemachine.withgoogle.com/train

Choose Image —> Standard Image Model

In groups of two train your first image model using a webcam and then:

- Rotate the laptop so that the background changes
- Put jacket/scarf etc. on, take different objects

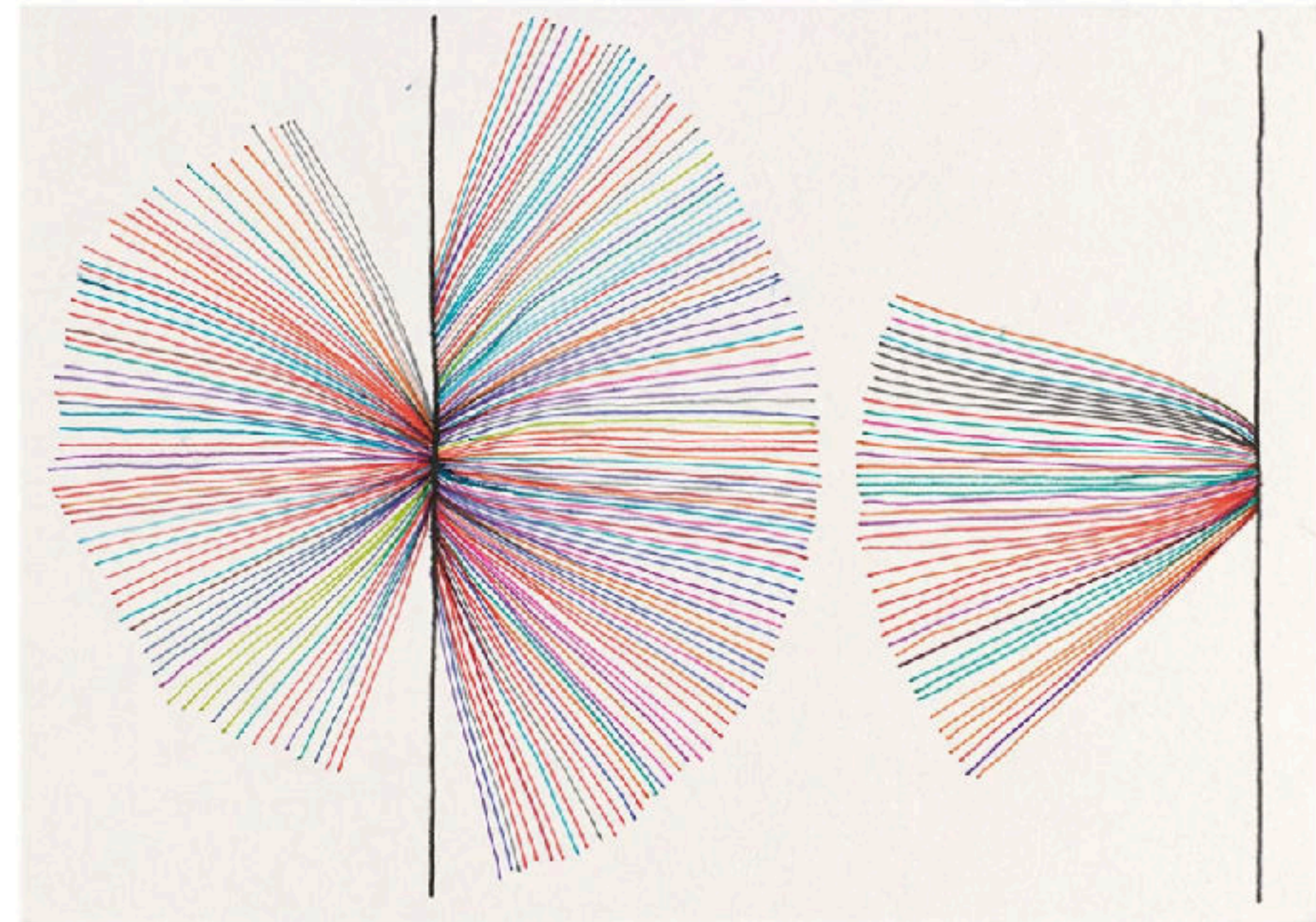
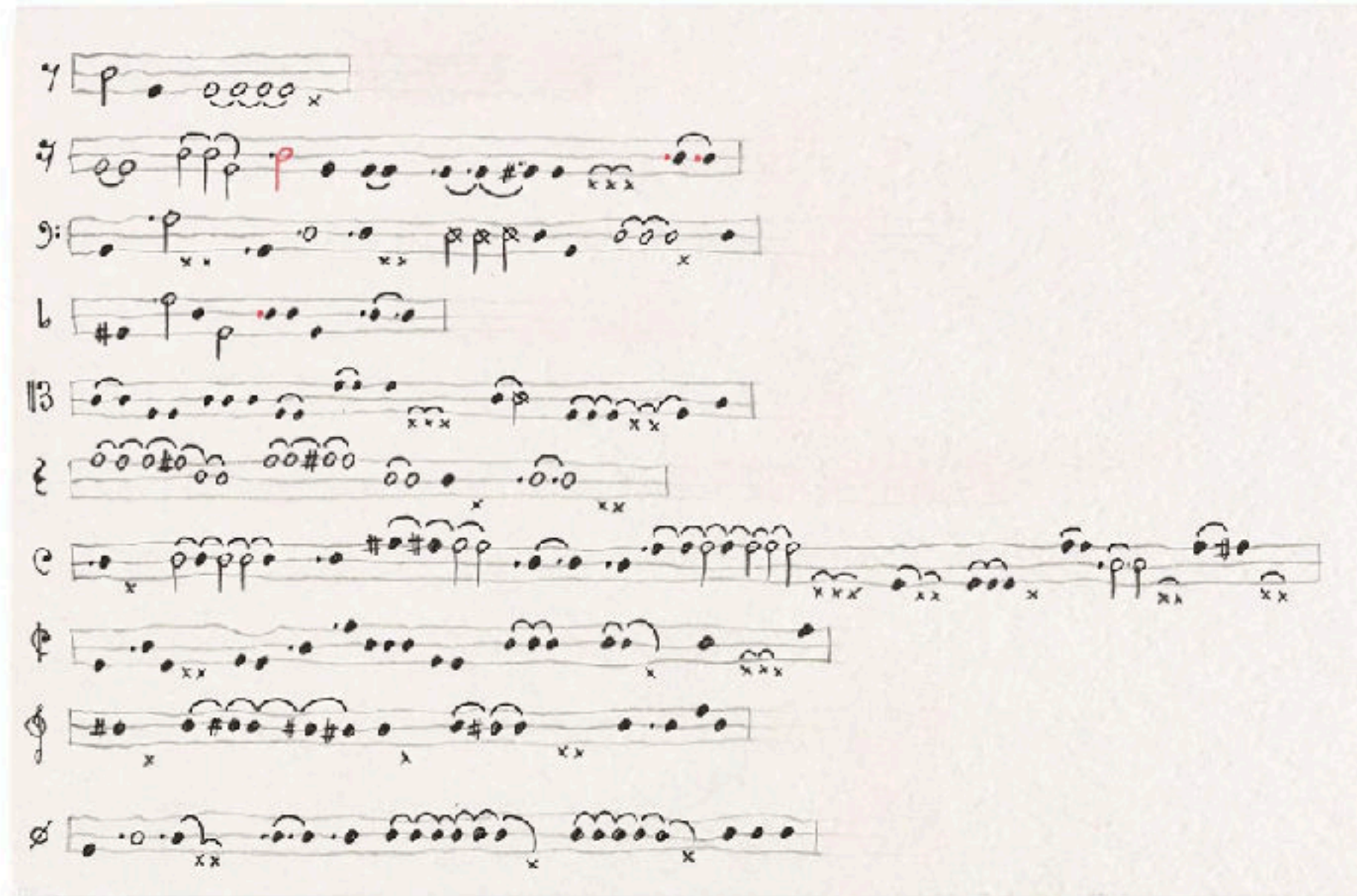




INDIVIDUAL PROJECT



Giorgia Lupi and Stefanie Posavec "Dear Data", 2016



66 DEAR DATA WEEK 07: MUSICAL COMPLAINTS

HOW TO READ IT: Each "note" is a single complaint I said.
(i.e. every pingee of time I expressed dissatisfaction or annoyance about a situation or particular thing)
Each "score" represents a typology of things I complained about, featuring complaints in chronological order.

- SCORES:
- 4 - ME AS A PERSON (e.g. "I am so... ngly / obsessive...")
 - 4 - ME AT WORK (e.g. "I should've done...")
 - 2 - WORK (e.g. "this project isn't going well!")
 - 6 - TECHNOLOGY (e.g. "the scanner is not working!")
 - 13 - SERVICE / FOOD (e.g. "OMG the waiter is so slow!")
 - 5 - SOMEBODY (e.g. "He's really a jerk...")
 - 5 - COLD (e.g. "I am freezing! The A.C. is crazy!")
 - 6 - HOW I FEEL (e.g. "so tired!!", "so bored!!")
 - 6 - BOYFRIEND (e.g. "you're snoring! you haven't...")
 - 15 - OTHER (e.g. "I spent 1 hour waiting for...")

POSITIONS OF NOTES:

- 1 - ACTUAL need to complain
- 2 - average " " "
- 3 - MOREAL " " "
- 4 - MISSED COMPLAINTS: thought of complaining but didn't do!

ATTRIBUTES

- to boyfriend
- to friend / family
- to stranger
- in english (all the others were in ITA)
- via txt / email (digital life)
- #0 adding emphasis
- oo o close on time (same situation)
- p to stefanie
- about s.thing related to DEAR DATA

FROM:
GEORGIA LUPI
BROOKLYN
NY - USA

SEND TO:
STEFANIE POSAVEC
LONDON
- UK -
ENGLAND

DELIVERED BY
HAND (SPECIAL NYC
DELIVERY!)

DEAR DATA - WEEK 07

A WEEK OF COMPLAINTS*
AND GENERAL GRUMPINESS

HOW TO READ IT: (1) THREW DOWN MY PENS WHEN I FINISHED
(COMPLAINT #7) WHAT IS WRONG W/ ME??
ALL COMPLAINTS IN CHRONOLOGICAL ORDER EXCEPT 'COMPLAINTS TO ME' BECAUSE (COMPLAINT #1) I MESSED UP THE ORDER BY THAT TIME I HAD ALREADY SPENT 1.5 HOURS ON THIS... AARGH! AND (COMPLAINT #2) THIS PART OF THE DRAWING IS NOT SYMMETRICAL WHICH BOTHERS ME

TYPE OF COMPLAINT:

- WEATHER
- HUSBAND
- ANIMALS
- FAMILY
- SOCIETY / THE WORLD TODAY
- ACQUAINTANCES / STRANGERS
- MY APPEARANCE
- FRIENDS
- WORK
- HEALTH
- HUNGER
- MYSELF
- TECHNOLOGY / MEDIA
- MONEY
- INANIMATE OBJECTS
- TRANSPORT

MAIN STATS

- PRIVATE COMPLAINTS: 67
- OUTWARD COMPLAINTS: 100
- COMPLAINTS TO ME: 43
- # OF PEOPLE WHO COMPLAINED TO ME: 5
- # OF PEOPLE I COMPLAINED TO: 6
- COMPLAINED TO: NOT WORKING
- I COMPLAINED MOST ABOUT MY FAMILY (SPOKE MOM + DAD, I QU VERB!) + SECOND MOST COMPLAINED ABOUT MY HUSBAND (WHICH STEVE K. DELIVERED)

FROM:
S POSAVEC
LONDON
UK

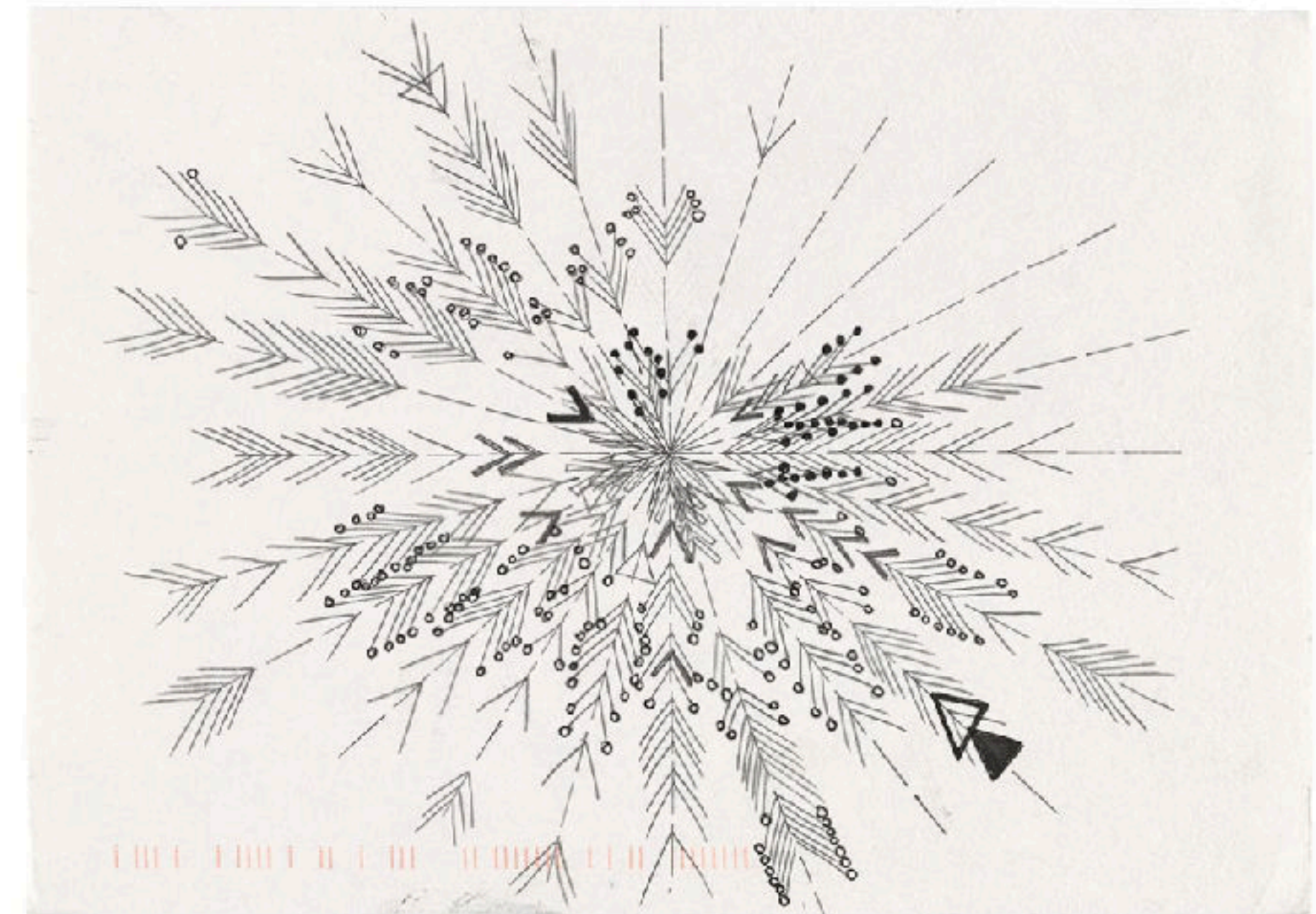
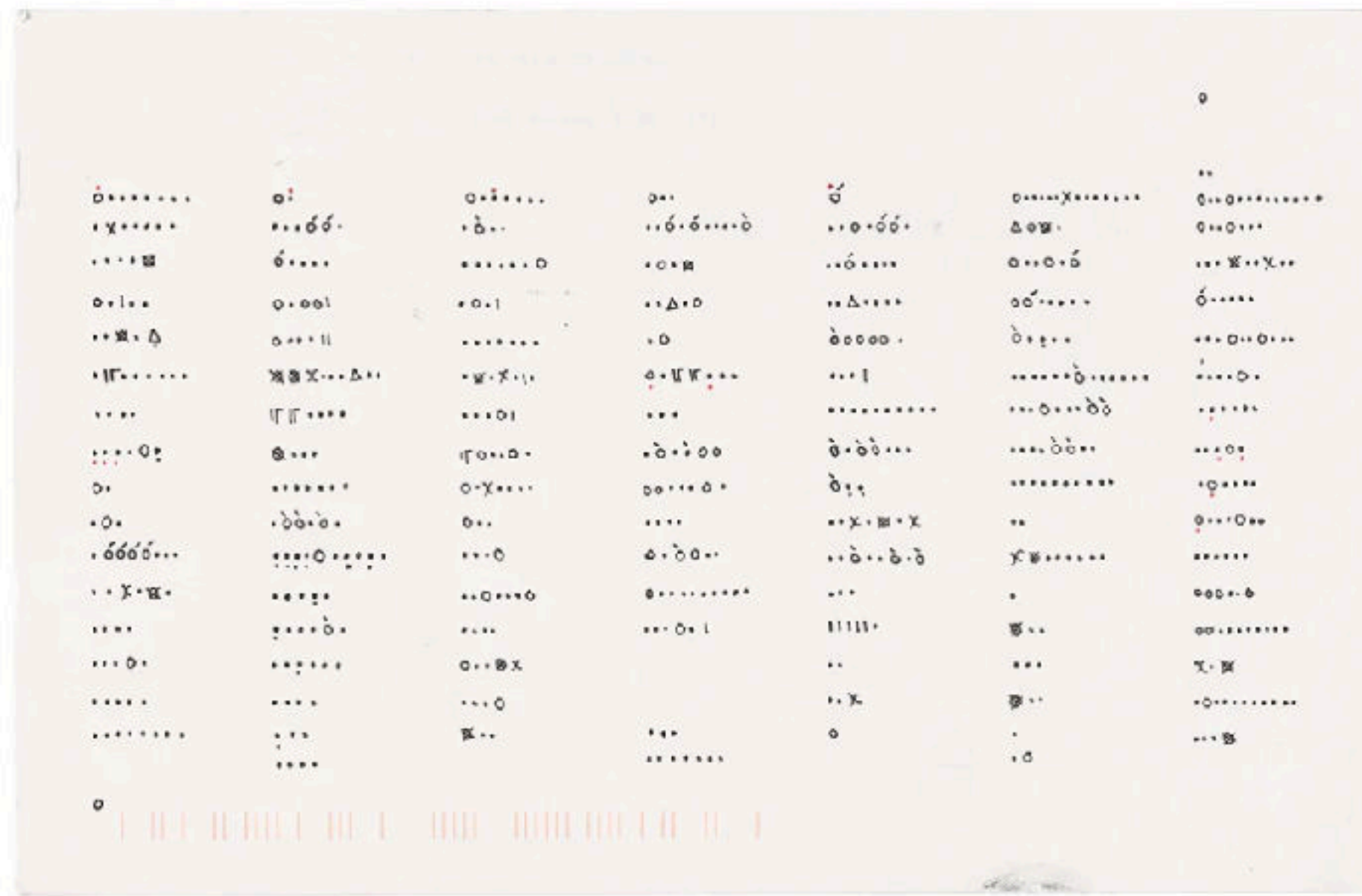


TO:
GEORGIA LUPI
BROOKLYN, NY
USA

DELIVERED BY
HAND (SPECIAL NYC
DELIVERY!)

What better visual reference than a musical score to show the repetitiveness of Georgia's protests and the "level" of complaint: whether they are justified or totally out of place.

Note the hand-drawn stamps: these postcards were delivered in person in New York!



DEAR DATA
WEEK 01: WHAT'S THE TIME?

HOW TO READ IT:



each symbol represents every moment I glanced at the clock, grouped by hours of the day.
Different symbols and attributes represent WHY and HOW I checked the time.

SYMBOLS

- o on purpose: wanted to know what the time was.
- just glanced: on a phone, mac or else
- Because I thought of this project.
- x I thought "don't look!" on but I did.
- ! Because I was Bored
- ! Because I was Hungry
- △ Heard somebody saying the time aloud.

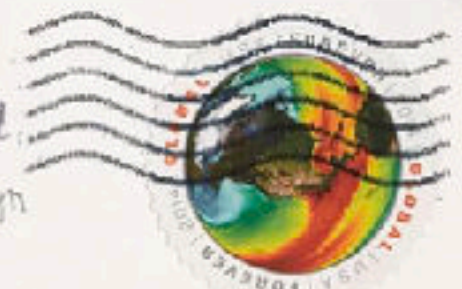
ATTRIBUTES

- f*ck! I'm late!
- oh, ok. I'm fine.
- analog support (i.e. wrist watch)
- alarm clock rang.
- glanced at the clock while texting or emailing with Stefanie. ☺

SECOND ATTEMPT

07 OCT 2014

NEW YORK NY 100
GEORGIA LUPI
BROOKLYN
-NY-



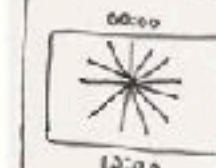
SEND TO:

STEFANIE POSAVEC
LONDON
[UK]
ENGLAND

DEAR DATA: WEEK 01:

A WEEK OF CLOCKS

LEGEND



Each line = ONE HOUR OF THE DAY, MOVING CLOCKWISE
Each line segment = ONE DAY. WEEK BEGINS IN CENTRE + MOVES OUTWARD
MONDAY SUNDAY

AN INSTANCE OF CLOCK-WATCHING IS INDICATED BY A SYMBOL

SYMBOL	SYMBOL	TOTAL INSTANCES	CAR	SYMBOL
PHONE		151	MICROWAVE	
LAPTOP		84	FRIEND'S OVEN	
TABLET		10	CHURCH CLOCK	
HUSBAND'S PHONE		3		
WATCH		11		

1124935282

FROM:

S. POSAVEC

LONDON

UK

15-09-2014

54007997

15-09-2014

54007997

GEORGIA LUPI

BROOKLYN, NY

USA

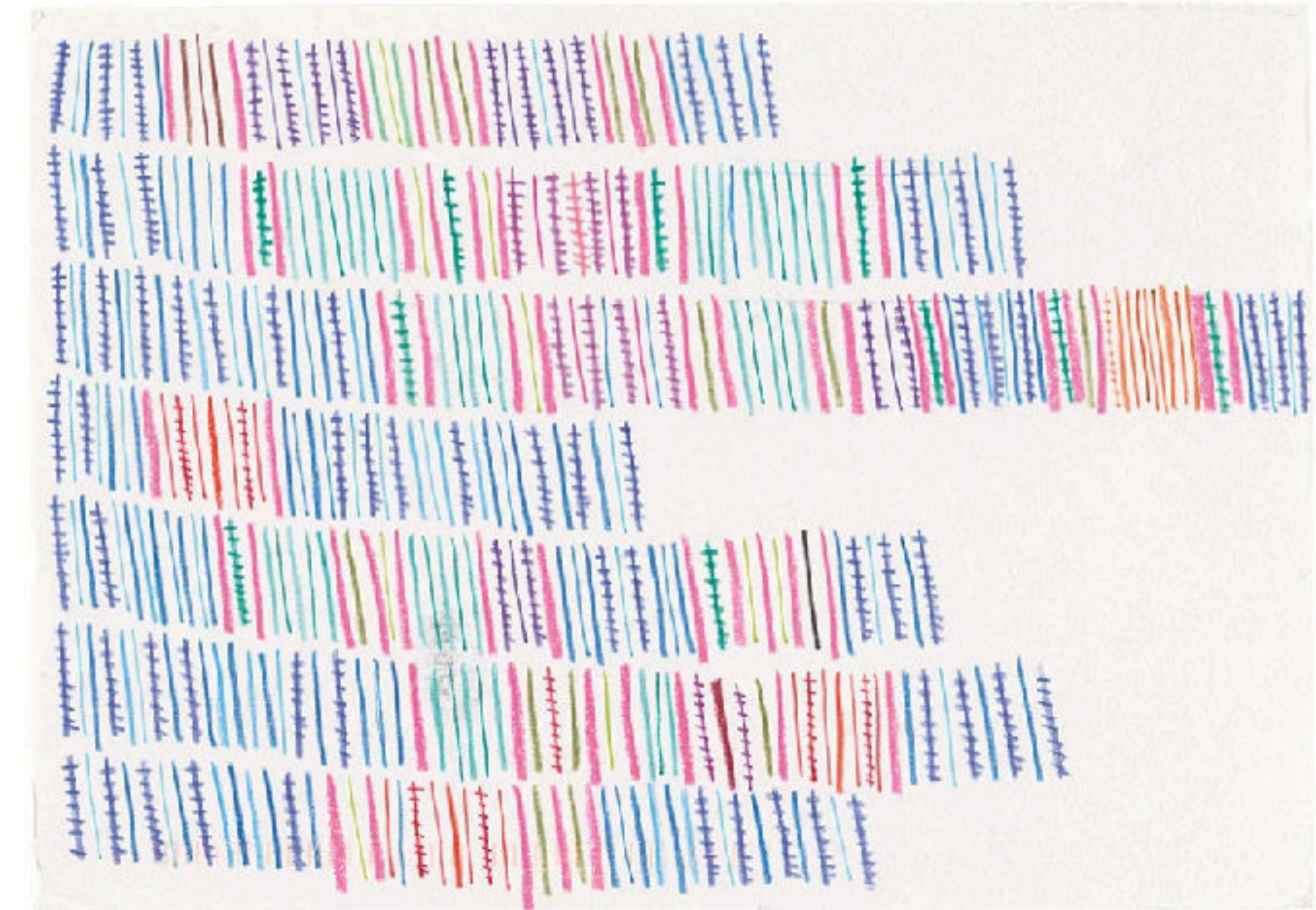
BY AIR MAIL
par avion

Royal Mail

Drawing her first postcard, Georgia had an idea for her whole collection: from now on every time she tracks something related to Stefanie, or to Dear Data, she uses a special pen to represent it!

• pink ink pen!

This week Georgia and Stefanie tried gathering data in small notebooks (tedious), but soon switched to making notes on their phones (much easier). Stefanie's favourite clock to capture: a bell tolling the time in a town in Devon.



DEAR DATA WEEK 24: DOORS' PATTERNS

HOW TO READ IT:

Every little rectangle represents a door I opened and/or passed through, in chronological order, to enter a space. (P.S. closet doors and furniture doors are not included)

TYPE OF DOOR:

MAIN LOCATION:

- my building
- work (NEW-INC)
- SHOP/STORE
- CAFE/PUB RESTAURANT
- CLIENT PLACE
- TRANSPORTATION

- extra external door (it's cold here Ha!)
- external door - entering the building
- eventual MID DOOR
- main space access (e.g. my Apt)
- when external door coincides to main space entrance
- external sliding door
- elevator automatic door

- bathroom restroom door
- bathroom entry
- the door between my bedroom and my living room
- thrash room
- glass door (e.g. meeting room)
- turnstile
- train doors

ATTRIBUTES

- so heavy!!
- boy friend opened it for me
- somebody opened it for me
- I locked it!
- I was carrying my 'Dear Data' postcard to post it to you
- cab door

FROM:
G. LUPI
NEW YORK NY 10001
24 FEB 2015
BROOKLYN NY - USA

SEND TO:

STEFANIE POSAVEC
LONDON
- UK -
ENGLAND

DEAR DATA - WEEK 24

A WEEK OF DOORS/SPACES

ABOUT THE DATA: I GATHERED DATA ON ALL OF THE SPACES I PASSED THROUGH IN THE WEEK, BOTH INTERNAL + EXTERNAL. A SPACE IS DEFINED BY WHETHER I HAD TO PASS THROUGH A DOOR OR NOT.

HOW TO READ IT:

IN CHRONOLOGICAL ORDER

EACH SPACE IS REPRESENTED BY A: OR A (CROSS-HATCHED LINES CARRY NO IMPORTANCE, MAINLY ONLY FOR DIFFERENTIATION BETWEEN SIMILAR COLOURS) THIS WAS A LAST MINUTE SPONTANEOUS DECISION ON THE 24 CARD TO MAKE THE COLOURS CLEARER... SOO AWWW!!! HAVE TO REBRAND X

THE TYPES OF SPACES I PASSED THROUGH INCLUDE:

- IN THE HOME/ PERSONAL SPACE: BEDROOM, SPACE BEDROOM (1 OF 2), GROUND FLOOR (OPEN PLAN, INCL. KITCHEN + LOUNGE), TOILET, BATHROOM, STUDIO: STUDIO BLDG., STUDIO SPACE, OUTSIDE: IN THE STREET
- DOCTOR: DOCTOR'S OFFICE, WAITING ROOM, UNIVERSITIES: UNIV. BLDG., CLASSROOM, TEAROOM, CULTURAL CENTRE: MAIN BLDG., GALLERY, TRANSPORT: BUS, TUBE/TRAIN, CAB
- SHOP: SHOP, BRYXTON MET. BLDG., PUB/RESTAURANT/CAFE/BAR, FRIEND'S HOUSE: MAIN HOUSE, BASEMENT, KITCHEN, LOUNGE, GYM: MAIN BLDG., CHANGING ROOM, BIKERIE ROOM, *CHANGING ROOMS + TOILETS ARE FOUND ACROSS DIFFERENT SPACES

FROM:
S. POSAVEC
LONDON
+840 - 2015

Royal Mail
Postbox
Mail Centre
05-05-2015
33505221

To: GIORGIA LUPI
BROOKLYN, NY
USA

BY AIR MAIL
par avion
Royal Mail®

After spending more than six hours drawing this hyper-detailed card, Georgia texted Stefanie as she posted it: "You need to know that if this one doesn't get to you I won't redraw it. You'll see what I mean."

Unfortunately, while Georgia's postcard arrived, Stefanie's postcard didn't, so she had to draw hers again (luckily it wasn't as detailed, but it was still supremely annoying).

DEAR, DATA

NO AI
at any step

Choose one everyday phenomenon. Observe it for seven days and record every occurrence as text, image or sound. Build the dataset by hand, draw it, and find the biases in your own collection.

28.09



Kick-off

Choose a topic, write your protocol



Collect

Text, image, audio



Structure

One row per observation



Draw

Hand-drawn postcards with a legend



Reflect

Find the biases in your own data



Present

Postcards, dataset, bias report

14.10

DATA PROTOCOL

“A WEEK OF DOORS”

ID	date	time	place	category	note	image
001	06.10	07:12	home	wooden · push	“slammed by the wind”	001.jpg
002	06.10	07:43	tram 8	automatic	didn't touch it	002.jpg
003	06.10	08:05	campus	heavy glass	“held it for someone”	003.jpg
004	06.10	21:30	?	fridge?	does a fridge count? → rule 2b	—

BIAS DIARY

- Missed entries (when, and why)?
- Definition changes (when did “a door” start to include gates)?
- Behaviour changes (Did you open more doors because you were counting)?
- Things left out (privacy, embarrassment, effort)

DELIVERABLES

1. Raw collection of data (images, audio recordings, drawings, texts)
2. Structured dataset (one table with one row per observation explaining every column and category).
3. Hand-drawn postcard with the data in the front and the legend in the back
4. Bias report
5. 7 min presentation on 14.10, show the postcard, walk us through the biases.