

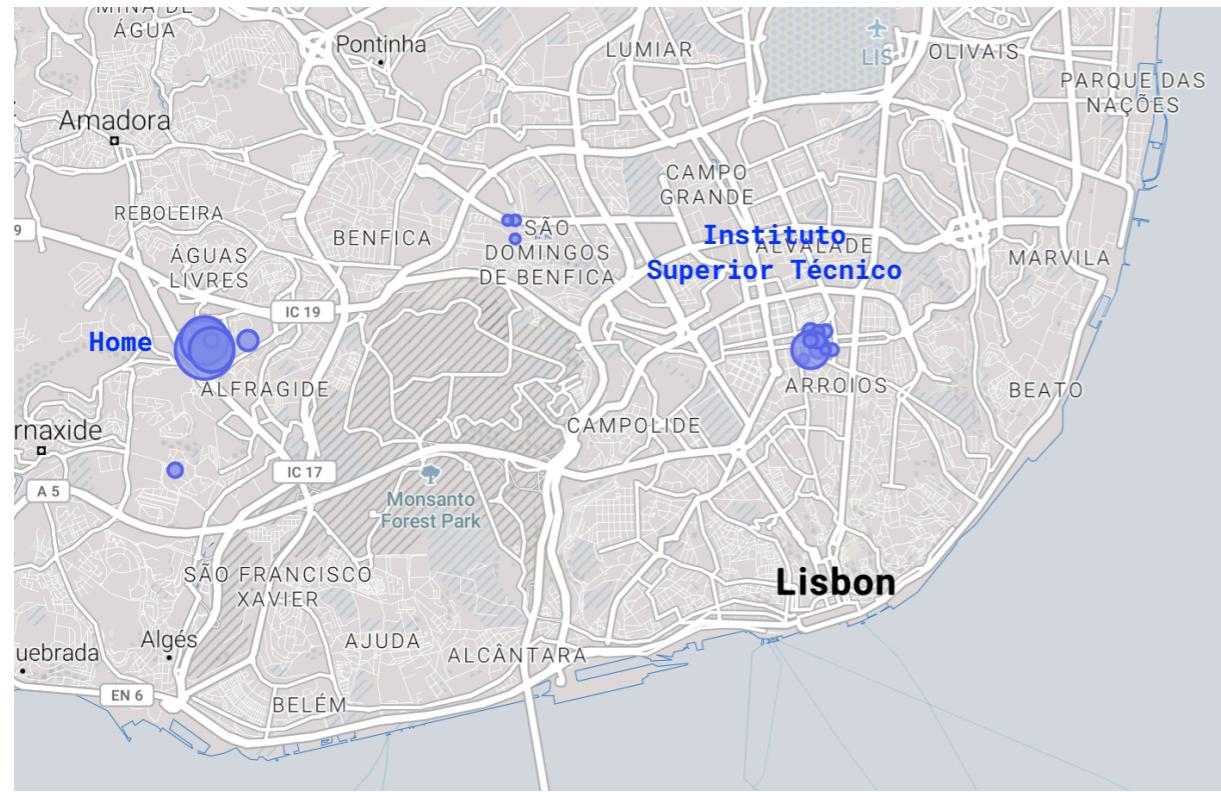
# VISUAL EXPLORATIONS

THE WORLD OF DATAVIZ & ITS CHALLENGES IN PRODUCT

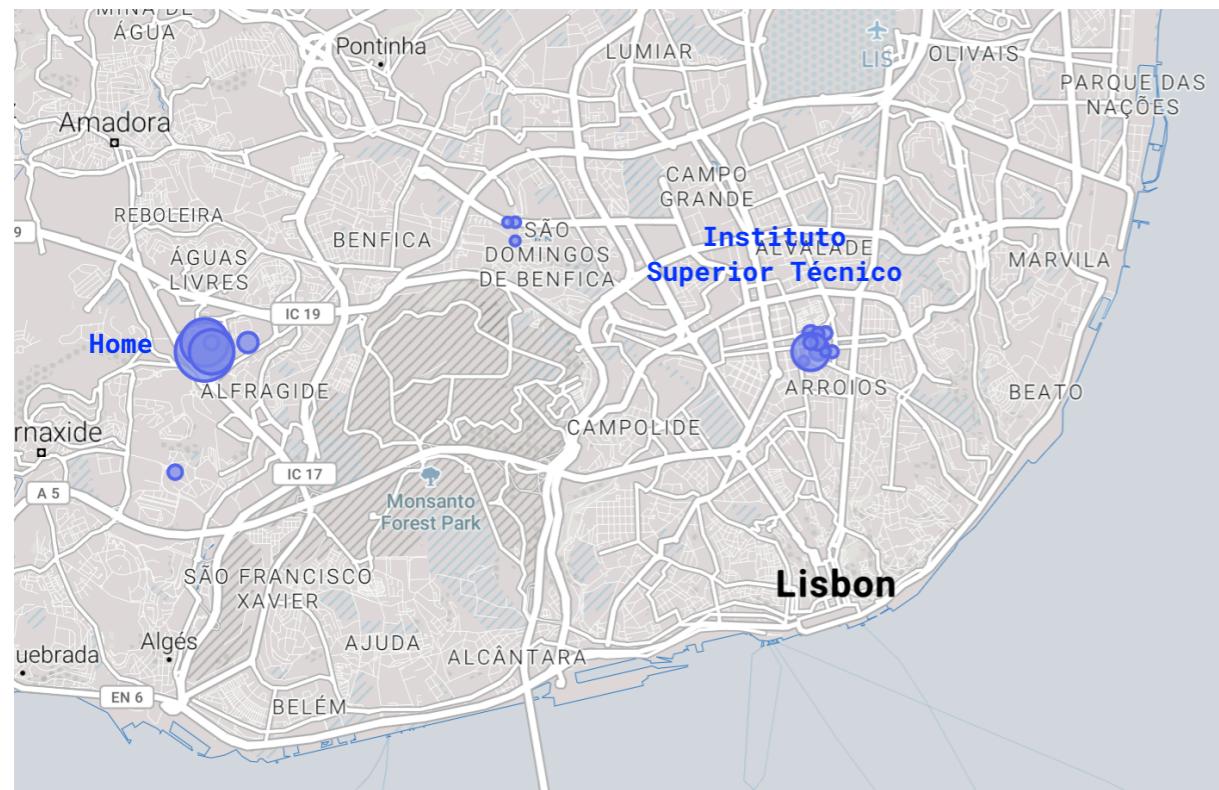
BEATRIZ MALVEIRO

MATHEMATICIAN + GRAPHIC DESIGN ENTHUSIAST = DATAVIZ ENGINEER

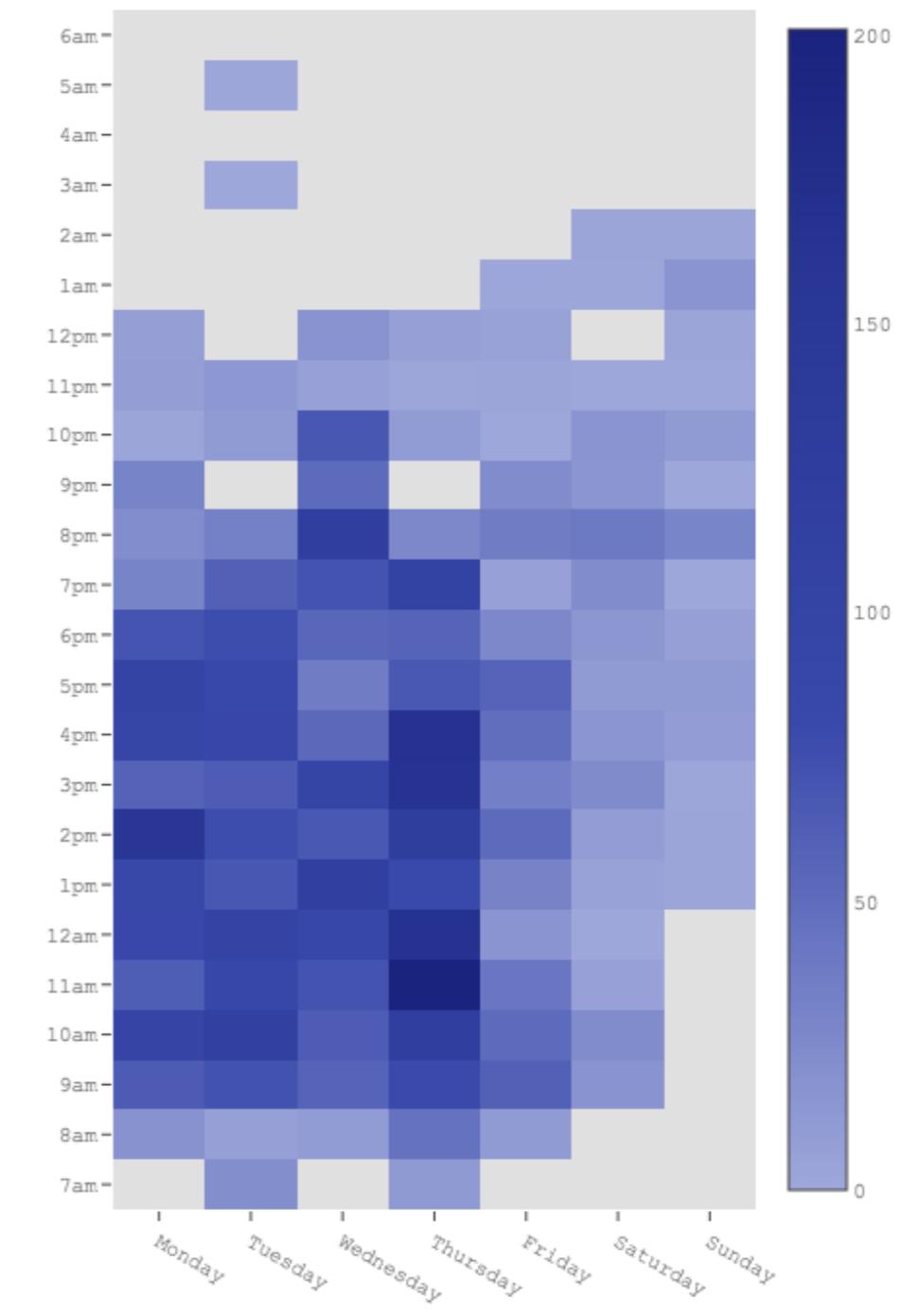
feedzai 

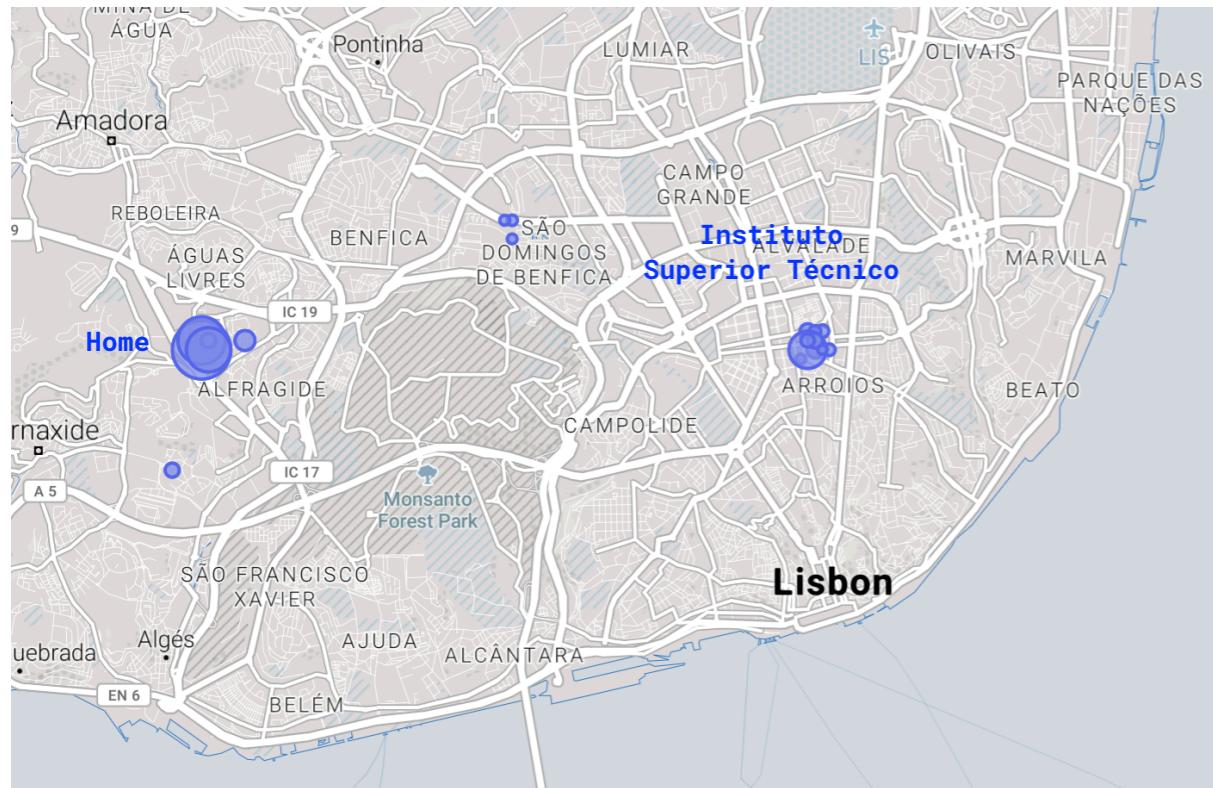


2014



- 2014

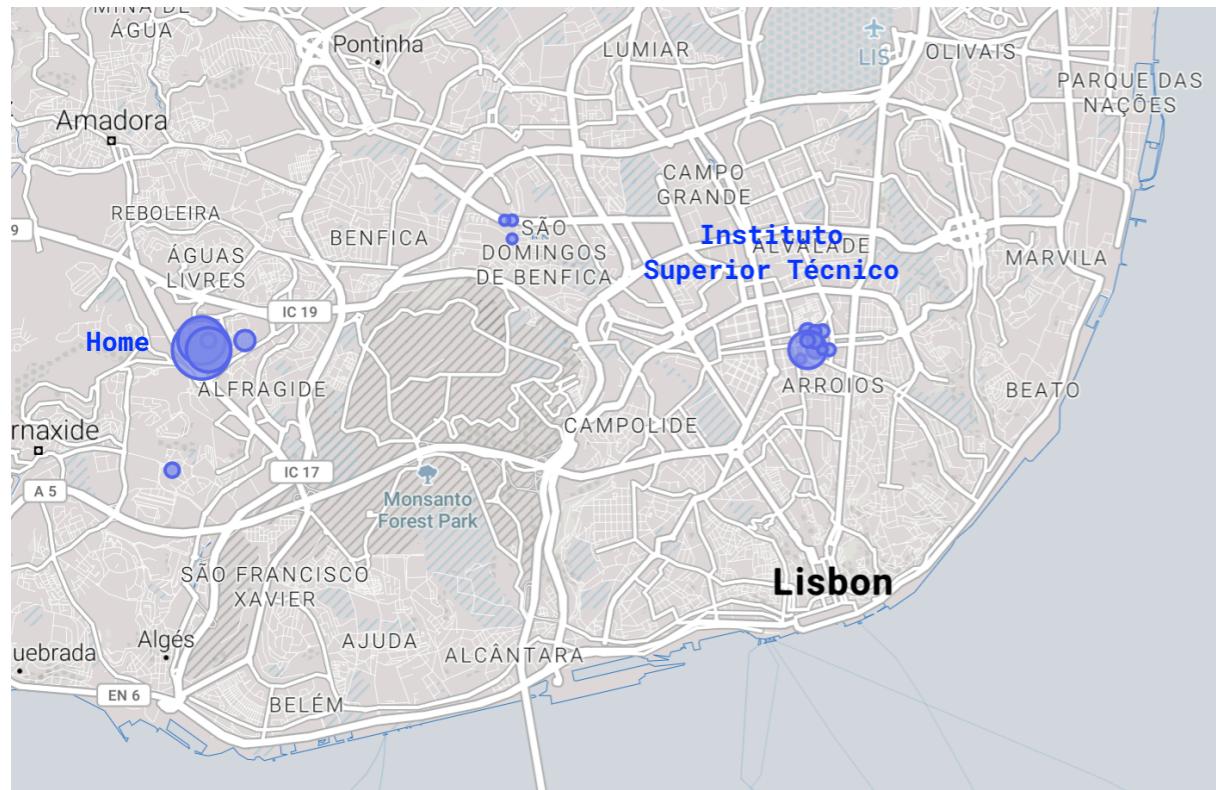




2014



2015



2014



2015



2016



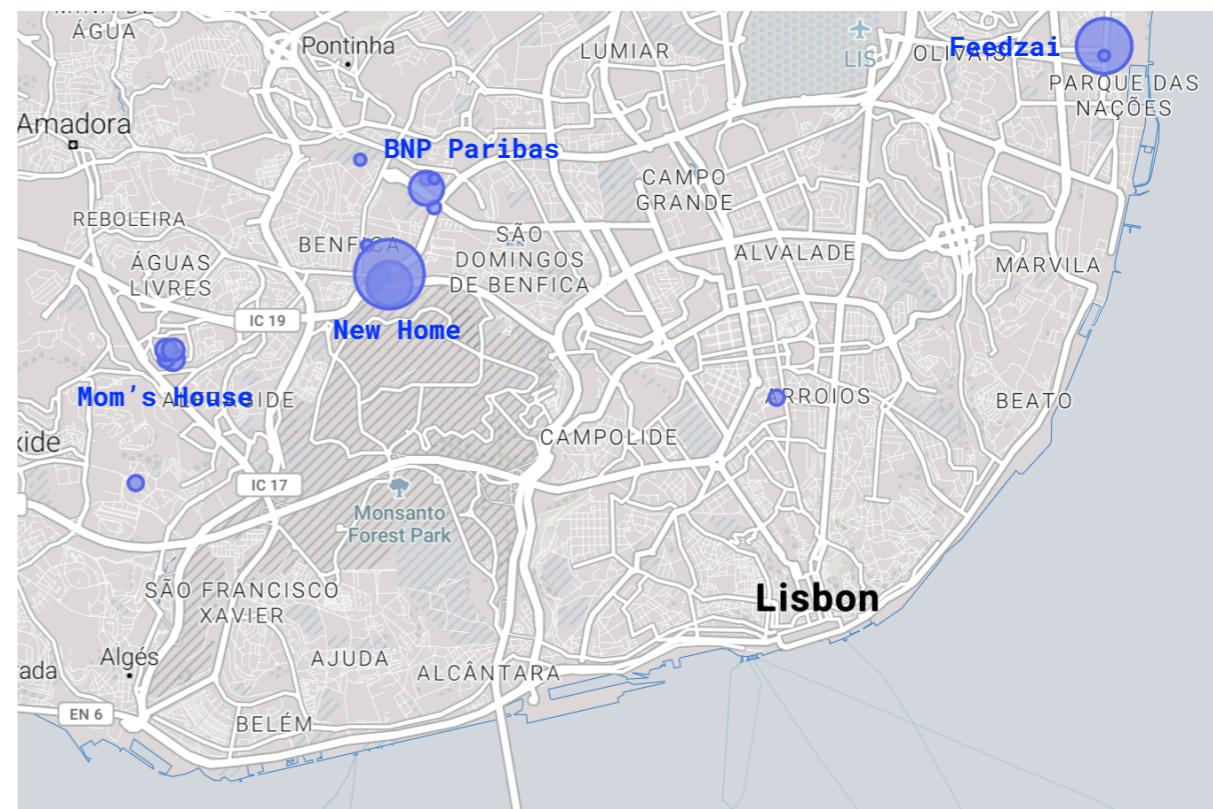
2014



2015



2016



2017



## Star Wars Characters Statistics

Character	Episode 4	Episode 5	Episode 6	Episode 7	Episode 8
Lando Clarissien	0	9	5	0	0
Yoda	0	13	4	0	2
Obi-Wan Kenobi	18	0	2	0	0
Poe Dameron	0	0	0	9	12
Kylo Ren	0	0	0	19	15
Darth Vader	9	13	14	0	0
R2-D2	18	11	7	2	1
Chewbacca	1	17	11	9	2
Finn	0	0	0	32	18
C3PO	19	15	14	1	3
Leia Organa	14	23	21	6	9
Rey	0	0	0	43	30
Han Solo	20	24	18	21	0
Luke Skywalker	37	35	34	1	21

## Star Wars Characters Statistics

Character	Episode 4	Episode 5	Episode 6	Episode 7	Episode 8
Lando Clarissien	0	9	5	0	0
Yoda	0	13	4	0	2
Obi-Wan Kenobi	18	0	2	0	0
Poe Dameron	0	0	0	9	12
Kylo Ren	0	0	0	19	15
Darth Vader	9	13	14	0	0
R2-D2	18	11	7	2	1
Chewbacca	1	17	11	9	2
Finn	0	0	0	32	18
C3PO	19	15	14	1	3
Leia Organa	14	23	21	6	9
Rey	0	0	0	43	30
Han Solo	20	24	18	21	0
Luke Skywalker	37	35	34	1	21

How much screen time did **Leia** have in **Episode VII**?

## Star Wars Characters Statistics

Character	Episode 4	Episode 5	Episode 6	Episode 7	Episode 8
Lando Clarissien	0	9	5	0	0
Yoda	0	13	4	0	2
Obi-Wan Kenobi	18	0	2	0	0
Poe Dameron	0	0	0	9	12
Kylo Ren	0	0	0	19	15
Darth Vader	9	13	14	0	0
R2-D2	18	11	7	2	1
Chewbacca	1	17	11	9	2
Finn	0	0	0	32	18
C3PO	19	15	14	1	3
Leia Organa	14	23	21	6	9
Rey	0	0	0	43	30
Han Solo	20	24	18	21	0
Luke Skywalker	37	35	34	1	21

How much screen time did **Leia** have in **Episode VII**?

⊕ Expresses values precisely

## Star Wars Characters Statistics

Character	Episode 4	Episode 5	Episode 6	Episode 7	Episode 8
Lando Clarissien	0	9	5	0	0
Yoda	0	13	4	0	2
Obi-Wan Kenobi	18	0	2	0	0
Poe Dameron	0	0	0	9	12
Kylo Ren	0	0	0	19	15
Darth Vader	9	13	14	0	0
R2-D2	18	11	7	2	1
Chewbacca	1	17	11	9	2
Finn	0	0	0	32	18
C3PO	19	15	14	1	3
Leia Organa	14	23	21	6	9
Rey	0	0	0	43	30
Han Solo	20	24	18	21	0
Luke Skywalker	37	35	34	1	21

Did R2-D2 have more screen time than C3PO?

## Star Wars Characters Statistics

Character	Episode 4	Episode 5	Episode 6	Episode 7	Episode 8
Lando Clarissien	0	9	5	0	0
Yoda	0	13	4	0	2
Obi-Wan Kenobi	18	0	2	0	0
Poe Dameron	0	0	0	9	12
Kylo Ren	0	0	0	19	15
Darth Vader	9	13	14	0	0
R2-D2	18	11	7	2	1
Chewbacca	1	17	11	9	2
Finn	0	0	0	32	18
C3PO	19	15	14	1	3
Leia Organa	14	23	21	6	9
Rey	0	0	0	43	30
Han Solo	20	24	18	21	0
Luke Skywalker	37	35	34	1	21

Did R2-D2 have more screen time than C3PO?

Who has the most screen time?

## Star Wars Characters Statistics

Character	Episode 4	Episode 5	Episode 6	Episode 7	Episode 8
Lando Clarissien	0	9	5	0	0
Yoda	0	13	4	0	2
Obi-Wan Kenobi	18	0	2	0	0
Poe Dameron	0	0	0	9	12
Kylo Ren	0	0	0	19	15
Darth Vader	9	13	14	0	0
R2-D2	18	11	7	2	1
Chewbacca	1	17	11	9	2
Finn	0	0	0	32	18
C3PO	19	15	14	1	3
Leia Organa	14	23	21	6	9
Rey	0	0	0	43	30
Han Solo	20	24	18	21	0
Luke Skywalker	37	35	34	1	21

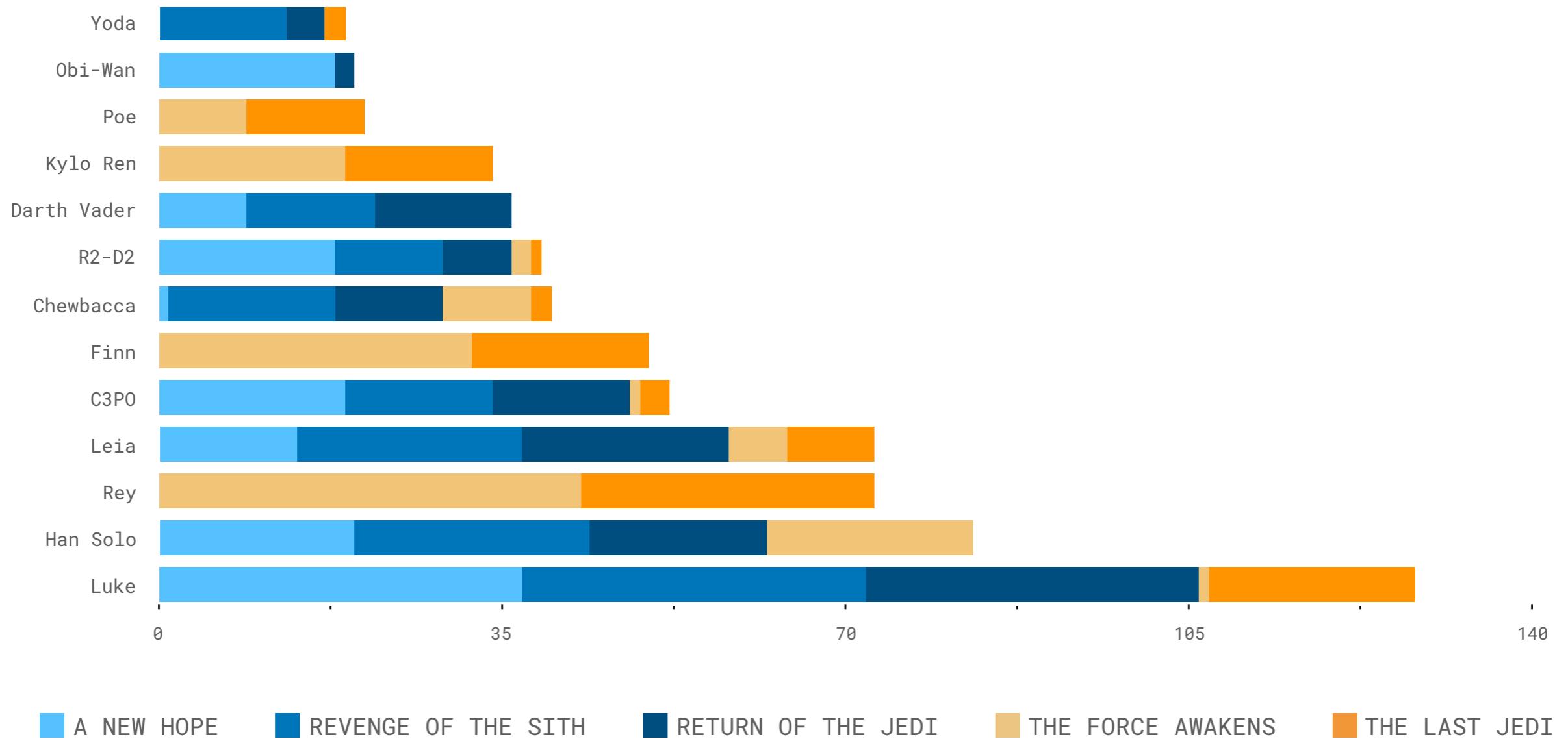
Did R2-D2 have more screen time than C3PO?

Who has the most screen time?

✖ Compare whole sets of numbers

PATTERNS | TRENDS | STORY

## Who has the most screen time?



## ABSTRACT INFORMATION

## COGNITION

SLOW



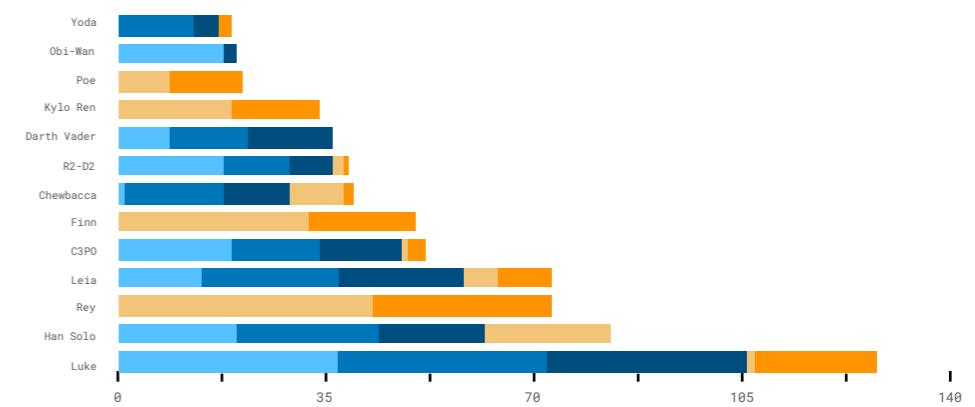
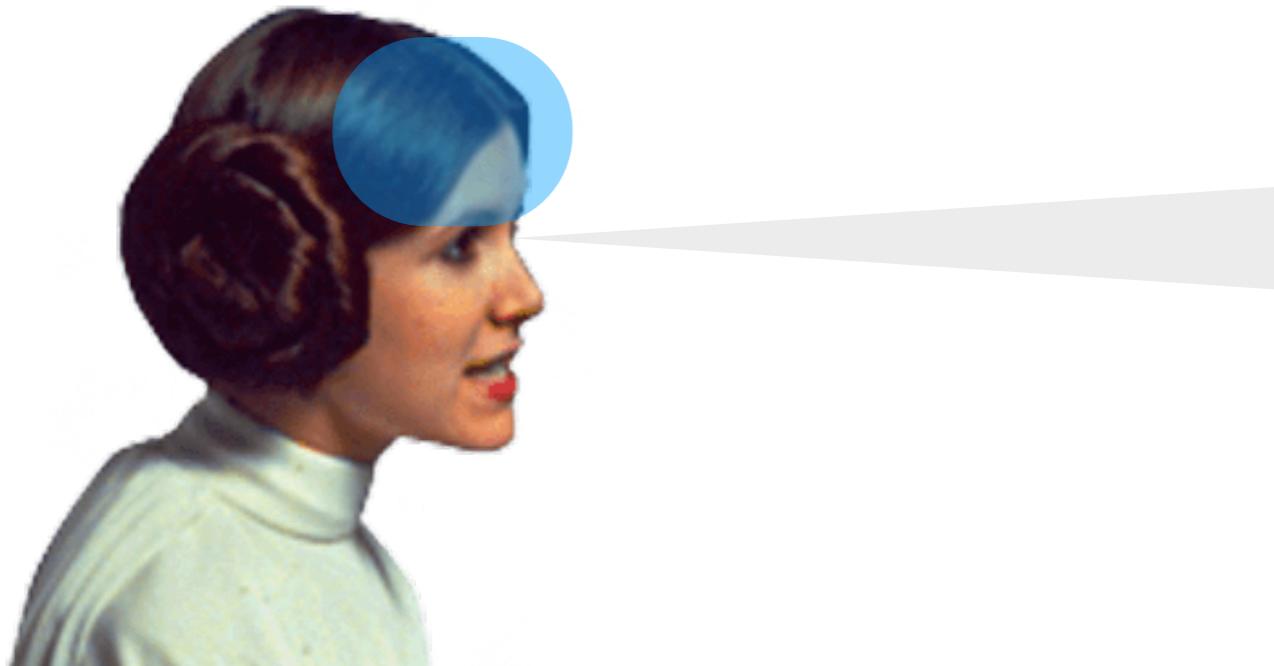
Character	Height	Episode 4	Episode 5	Episode 6	Episode 7	Episode 8
Lando Calrissian	179	0	9	5	0	0
Yoda	66	0	13	4	0	2
Obi-Wan Kenobi	182	18	0	2	0	0
Poe Dameron	175	0	0	0	9	12
Kylo Ren	182	0	0	0	19	15
Darth Vader	203	0	13	14	0	0
R2-D2	189	18	11	7	2	1
Chewbacca	229	1	17	11	9	2
Finn	178	0	0	0	32	18
C-3PO	171	19	15	14	1	2
Leia Organa	158	14	23	21	6	9
Rey	170	0	0	0	43	30
Han Solo	180	20	24	18	21	0
Luke Skywalker	172	37	35	34	1	21

ABSTRACT INFORMATION

VISUAL REPRESENTATIONS

PERCEPTION

FAST



ABSTRACT INFORMATION

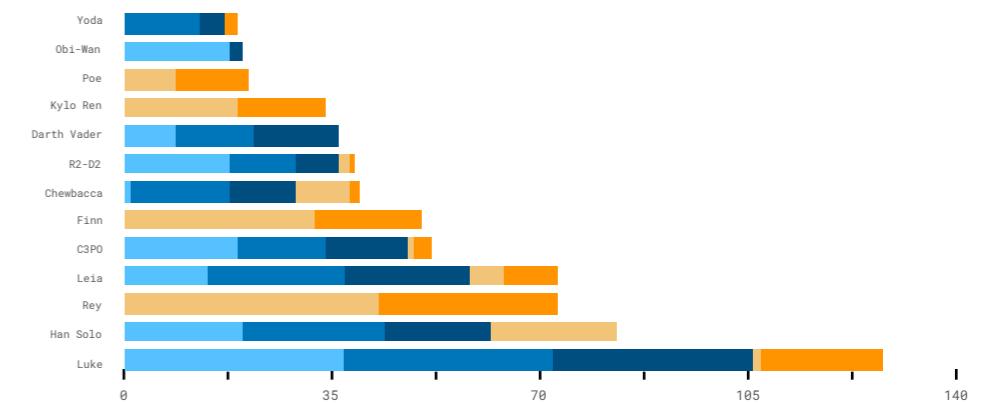
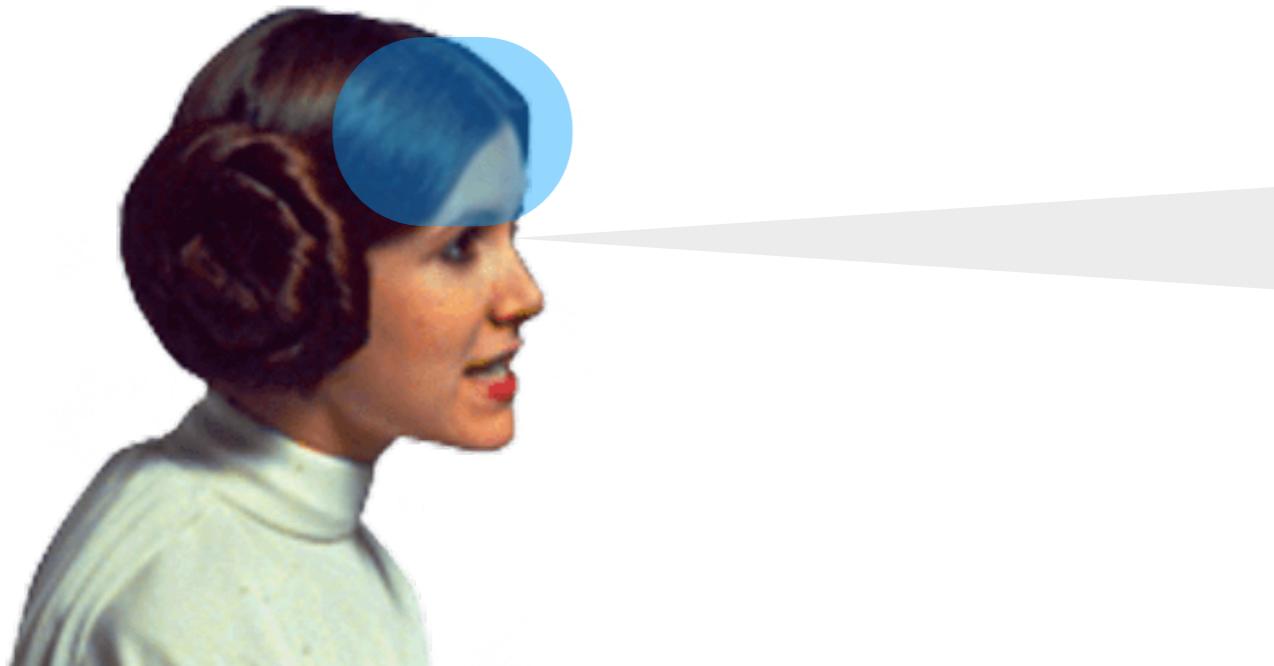
PHYSICAL ATTRIBUTES OF VISION

VISUAL REPRESENTATIONS

easily, efficiently, accurately, and meaningfully decoded

PERCEPTION

FAST



ABSTRACT INFORMATION

PHYSICAL ATTRIBUTES OF VISION

VISUAL REPRESENTATIONS

easily, efficiently, accurately, and meaningfully decoded

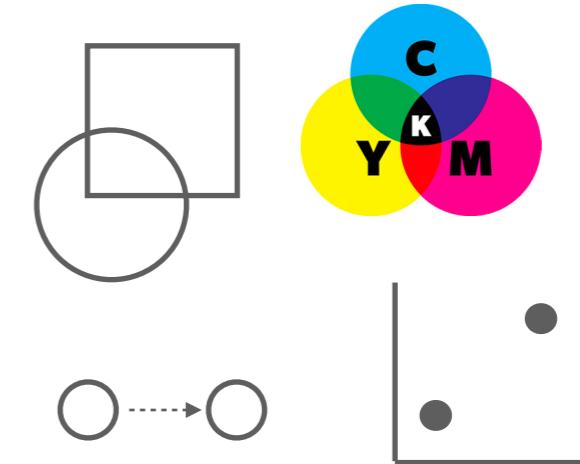
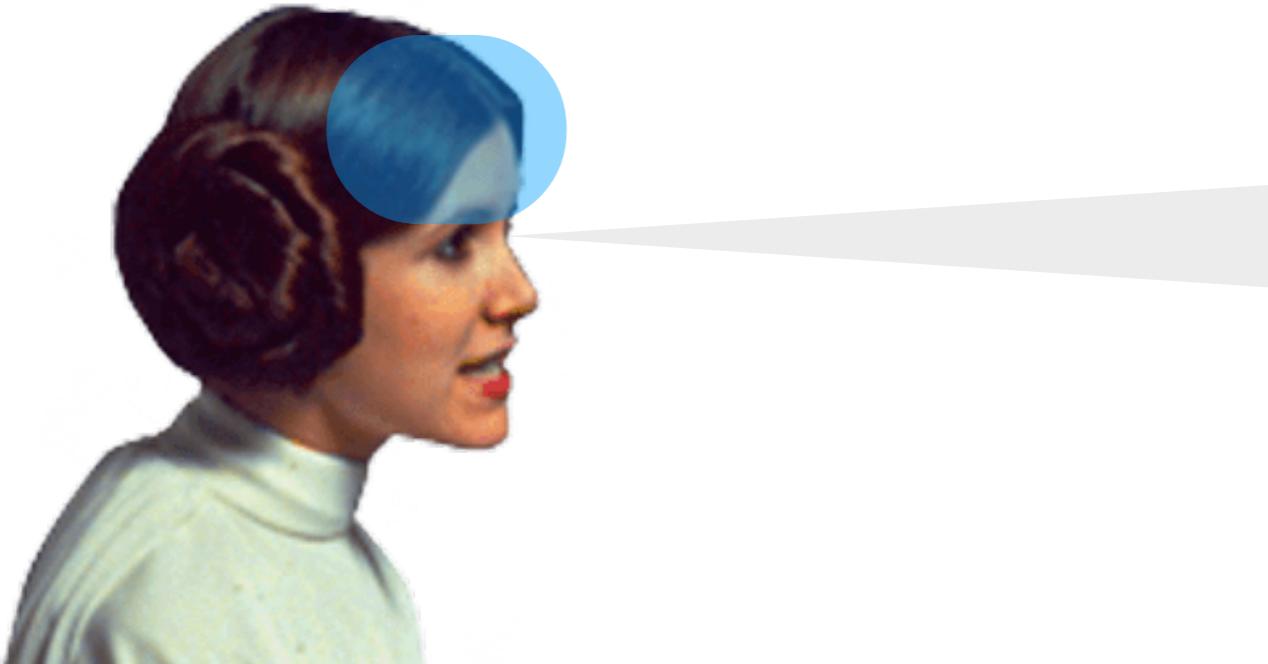
### PREATTENTIVE VISUAL PROCESSING

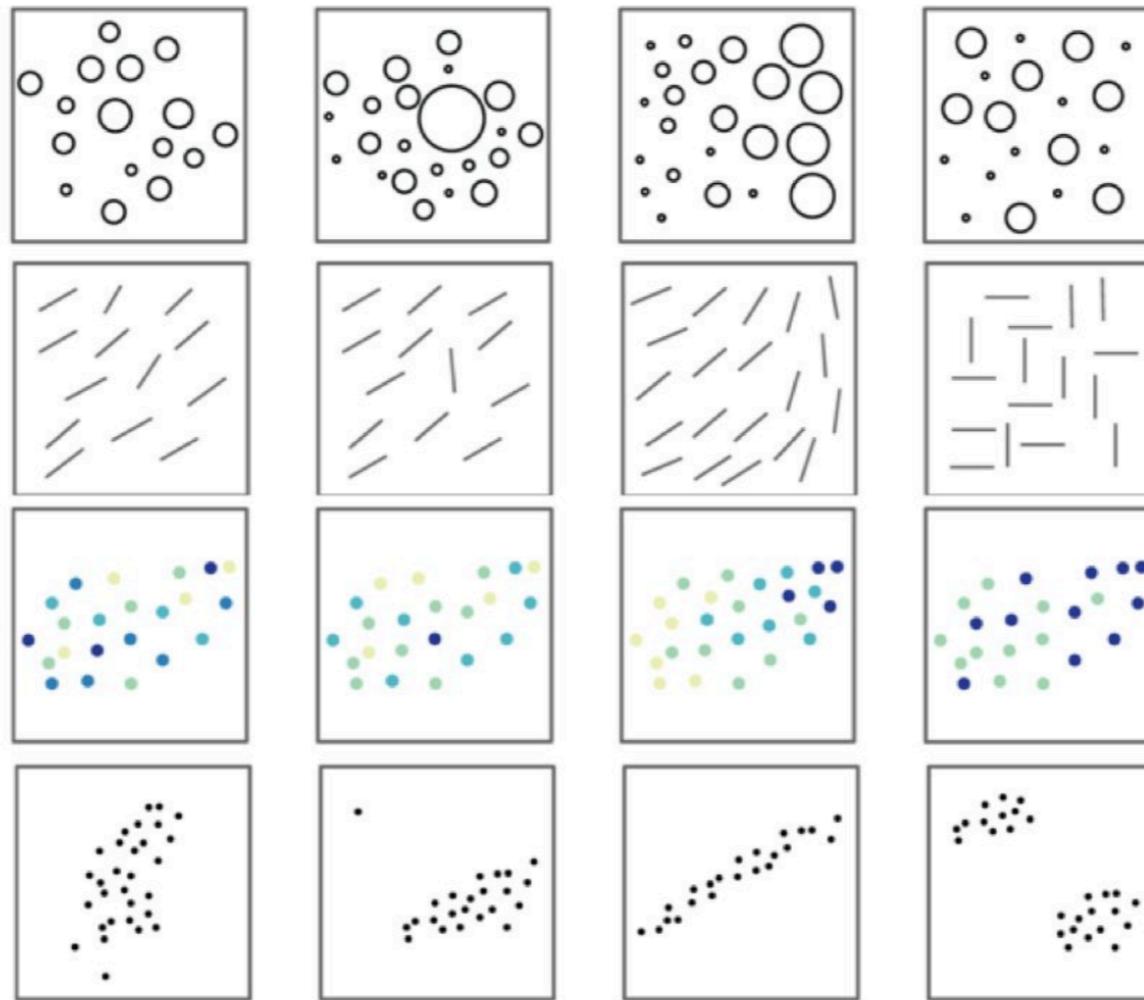
*"The features that pop out are hardwired in the brain, not learned."*

- Colin Ware, Visual Thinking: for Design

### PERCEPTION

FAST





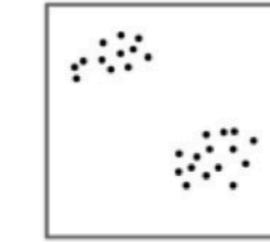
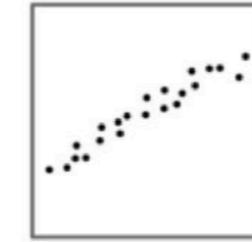
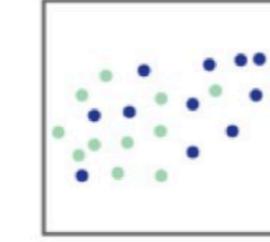
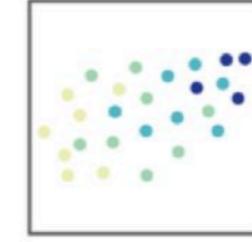
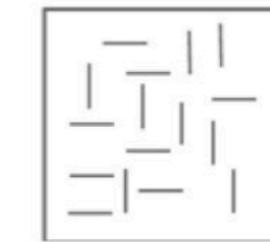
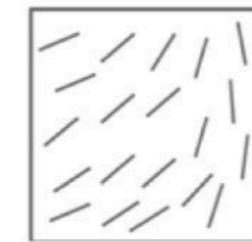
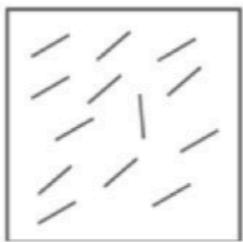
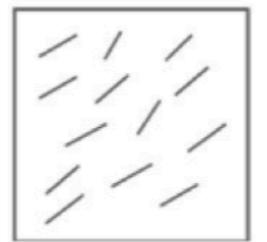
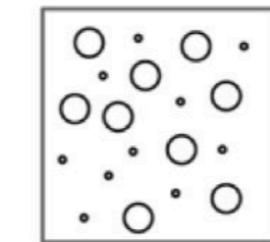
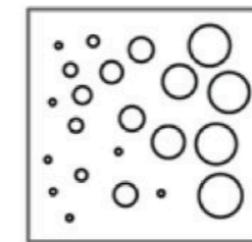
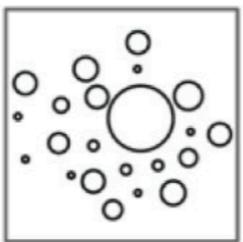
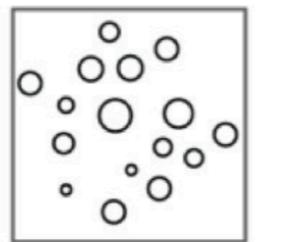
SUMMARY  
(MEAN)

IDENTIFICATION  
(OUTLIER)

RECOGNITION  
(TRENDS)

SEGMENTATION  
(CLUSTERING)

<100ms



SUMMARY  
(MEAN)

IDENTIFICATION  
(OUTLIER)

RECOGNITION  
(TRENDS)

SEGMENTATION  
(CLUSTERING)

<100ms

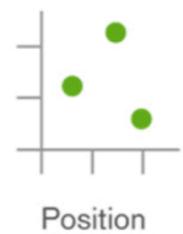
Let's jump to lightspeed!



## DATA VISUALISATION

*“Any kind of graphical representation to enable exploration, discovery, analysis and communication”*

- Alberto Cairo



Position



Length



Angle/Slope



Area



Volume



Difference



Color hue



Color Saturation



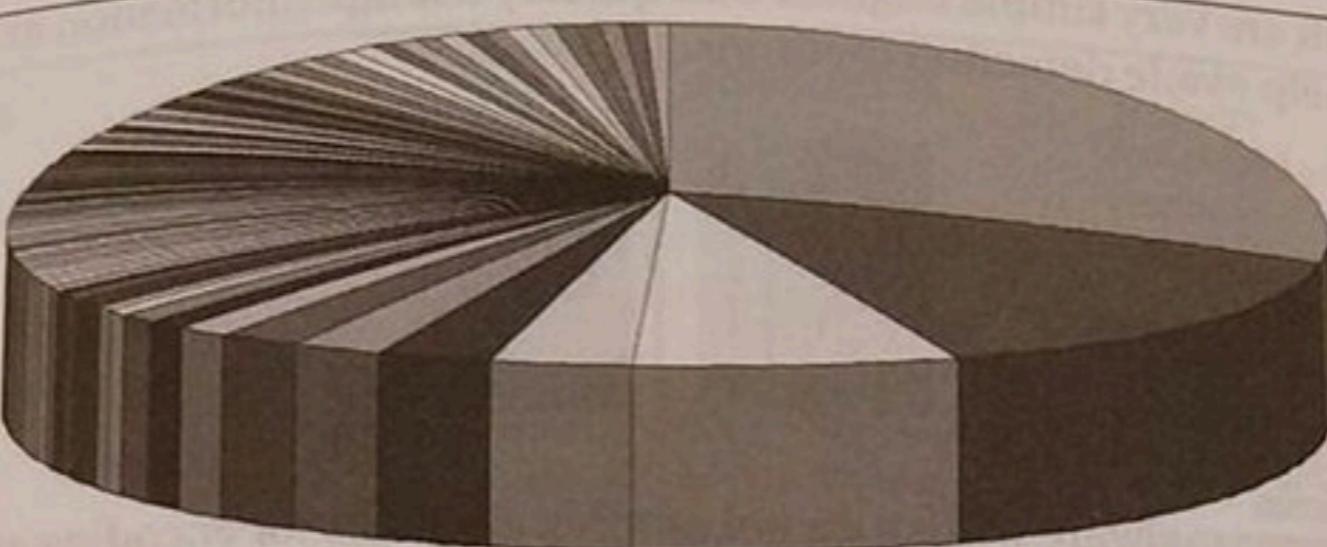
Contrast



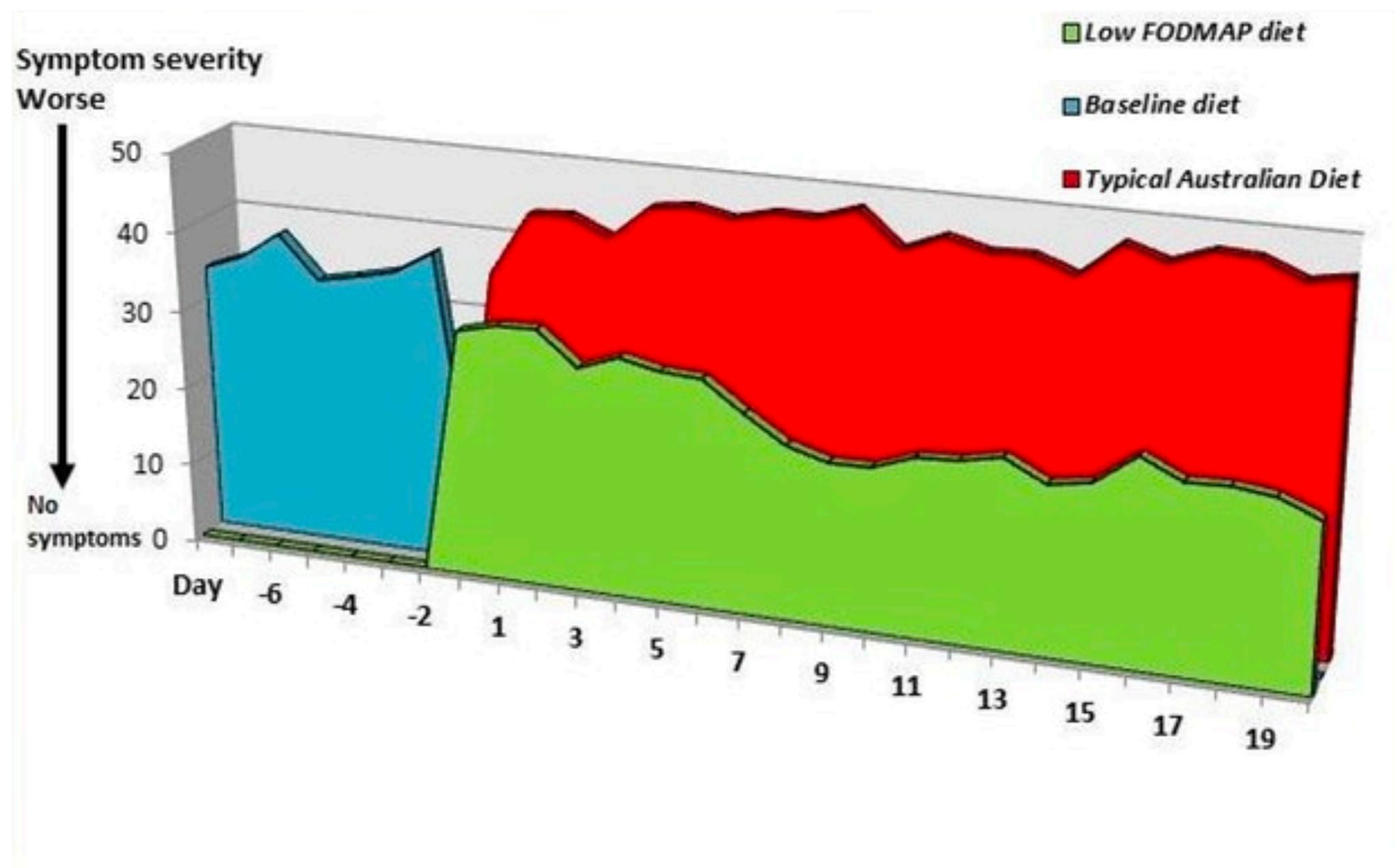
Texture







- Other Packers
- Other Installers
- ASProtect.b
- Aspack
- FSG
- NsPack
- MEW
- RLPack
- RP Crypt
- Armadillo
- PKUTE32
- PE-Pack
- NTKml
- Petite
- NSPack
- PE-Armour
- Yoda
- Hmimys
- BeRoEXEProtect
- nPack
- PEI
- Other
- UPX
- PE-Compact
- Unpack2
- Themida
- Morphine
- TeLock
- NSPack
- YodaProtect
- PEBundle
- SVK Protect
- EXE Crypter
- VMProtect
- PEspin
- PE-Compact
- Molebox
- Unpack1
- NakedPack
- Expressor
- ASProtect
- Kkrunchy
- PG.C



## DATA VISUALISATION

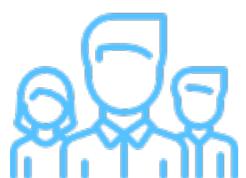
COMMUNICATION

DATA ANALYSIS

DATAVIZ FOR

COMMUNICATION

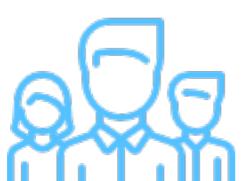
DATA  
TRENDS & PATTERNS



AUDIENCE

DATAVIZ FOR **COMMUNICATION**

**DATA**  
TRENDS & PATTERNS

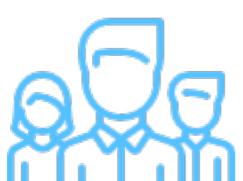


AUDIENCE

DATAVIZ FOR **COMMUNICATION**

**DATA**  
TRENDS & PATTERNS

Storytelling By **VISUALIZATION**



AUDIENCE

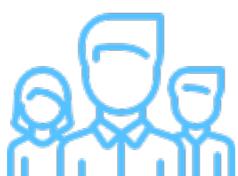
## DATAVIZ FOR

## COMMUNICATION

DATA  
TRENDS & PATTERNS

Storytelling By VISUALIZATION , example:

Annotated  
Crafted  
Custom  
Memorable



AUDIENCE

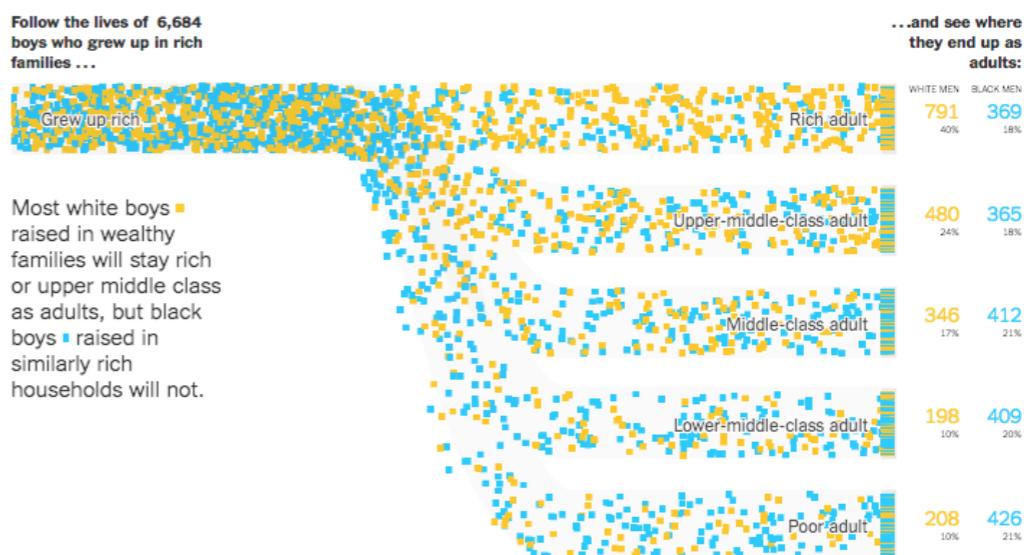


## Extensive Data Shows Punishing Reach of Racism for Black Boys

By EMILY BADGER, CLAIRE CAIN MILLER, ADAM PEARCE and KEVIN QUEALY MARCH 19, 2018

Black boys raised in America, even in the wealthiest families and living in some of the most well-to-do neighborhoods, still earn less in adulthood than white boys with similar backgrounds, according to a sweeping new study that traced the lives of millions of children.

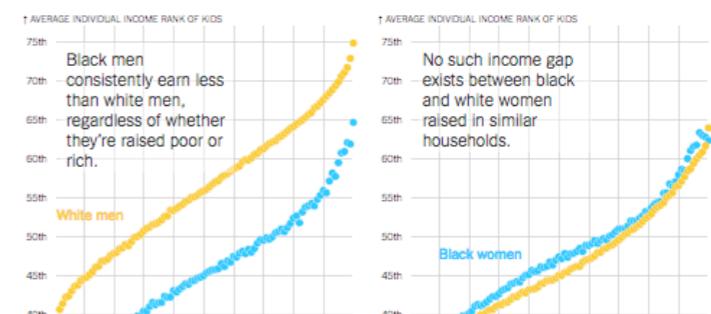
White boys who grow up rich are likely to remain that way. Black boys raised at the top, however, are more likely to become poor than to stay wealthy in their own adult households.



Even when children grow up next to each other with parents who earn similar incomes, black boys fare worse than white boys in 99 percent of America. And the gaps only worsen in the kind of neighborhoods that promise low poverty and good schools.

According to the study, led by researchers at Stanford, Harvard and the Census Bureau, income inequality between blacks and whites is driven entirely by what is happening among these boys and the men they become. Though black girls and women face deep inequality on many measures, black and white girls from families with comparable earnings attain similar individual incomes as adults.

Large income gaps persist between men — but not women.



## DATA VISUALISATION

COMMUNICATION

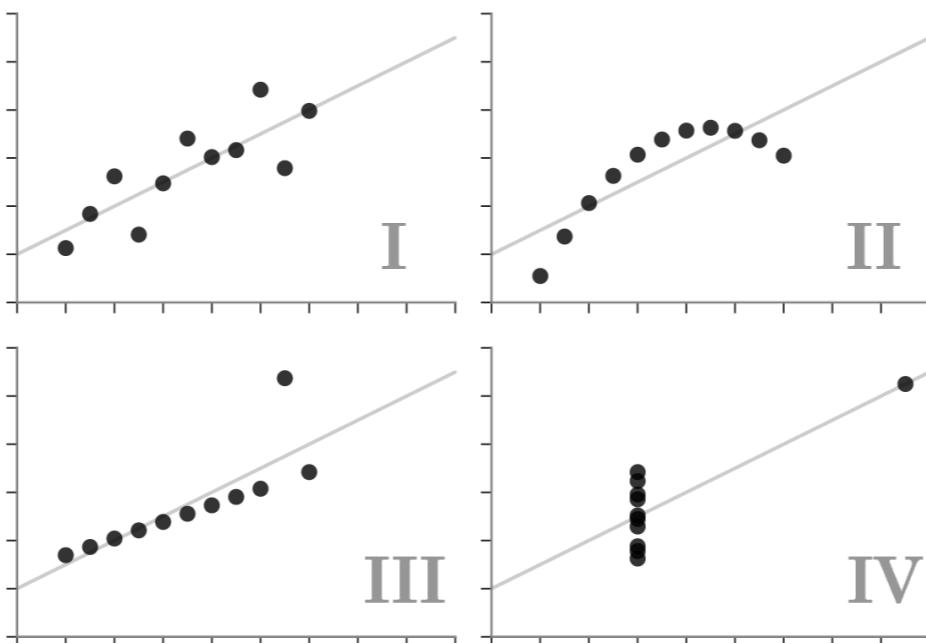
DATA ANALYSIS

DATAVIZ FOR

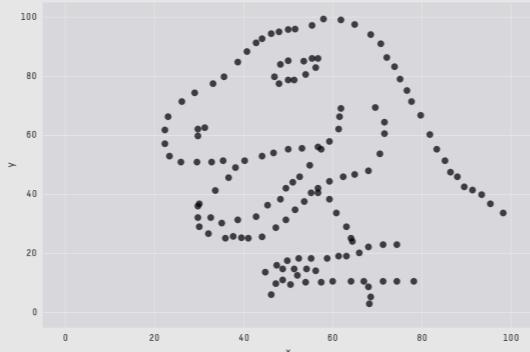
**DATA ANALYSIS**

## ANSCOMBE'S QUARTET

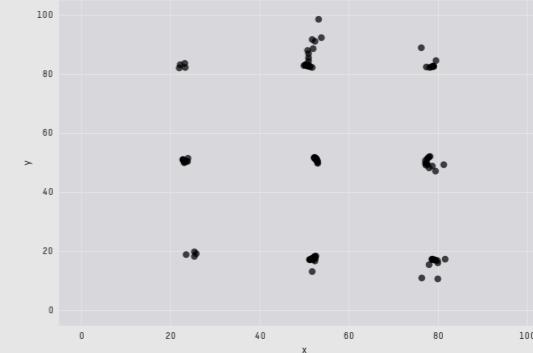
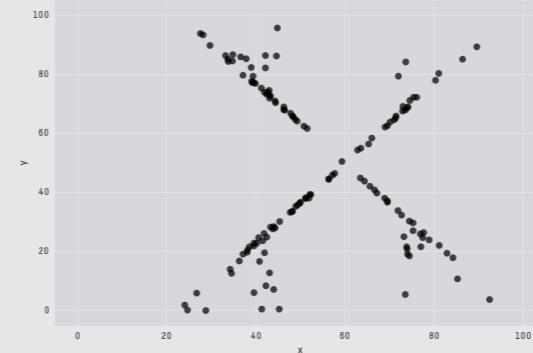
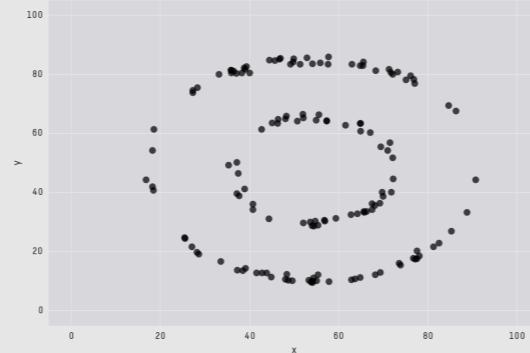
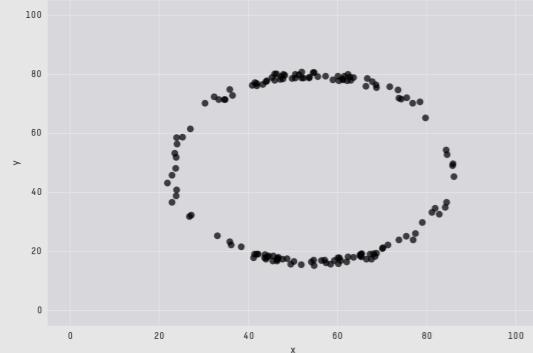
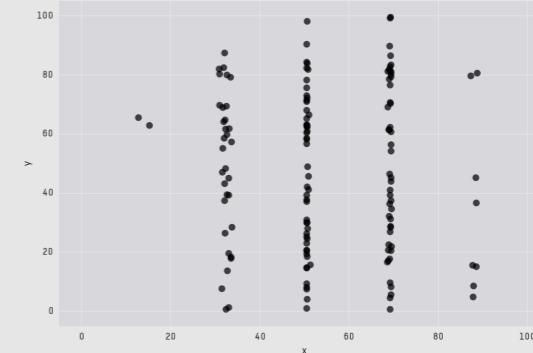
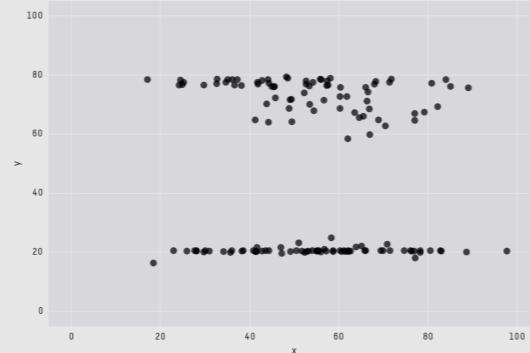
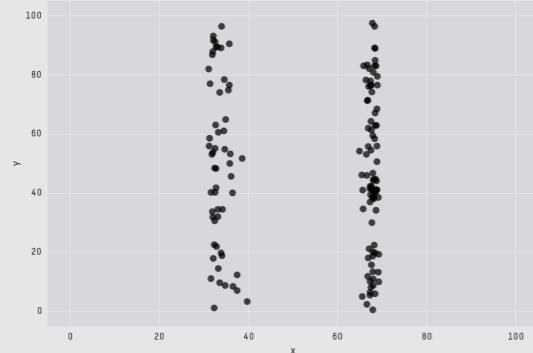
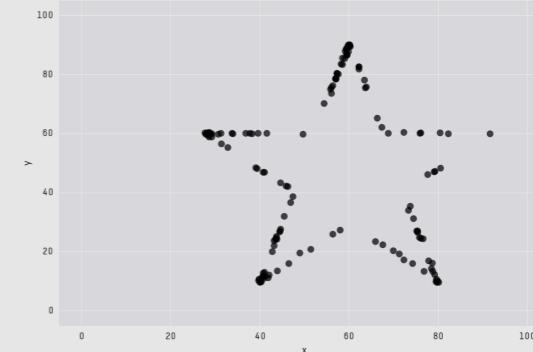
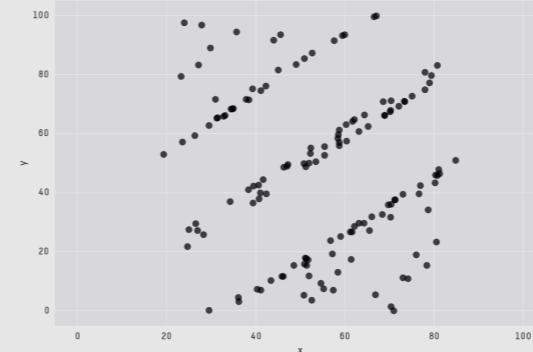
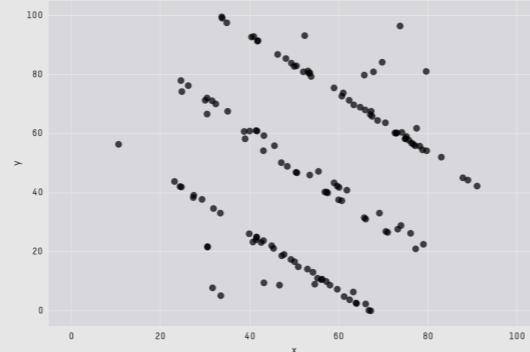
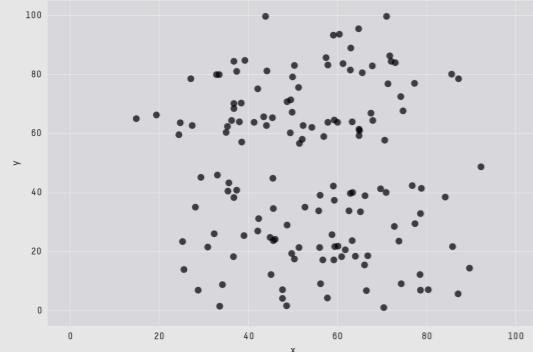
SAME MEAN, STD DEVIATION, CORRELATION



# DATASSAURUS DOZEN



X Mean: 54.26  
Y Mean: 47.83  
X SD : 16.76  
Y SD : 26.93  
Corr. : -0.06



DATAVIZ FOR

DATA ANALYSIS

**NOT BECAUSE IT LOOKS GOOD**

DATAVIZ FOR

DATA ANALYSIS

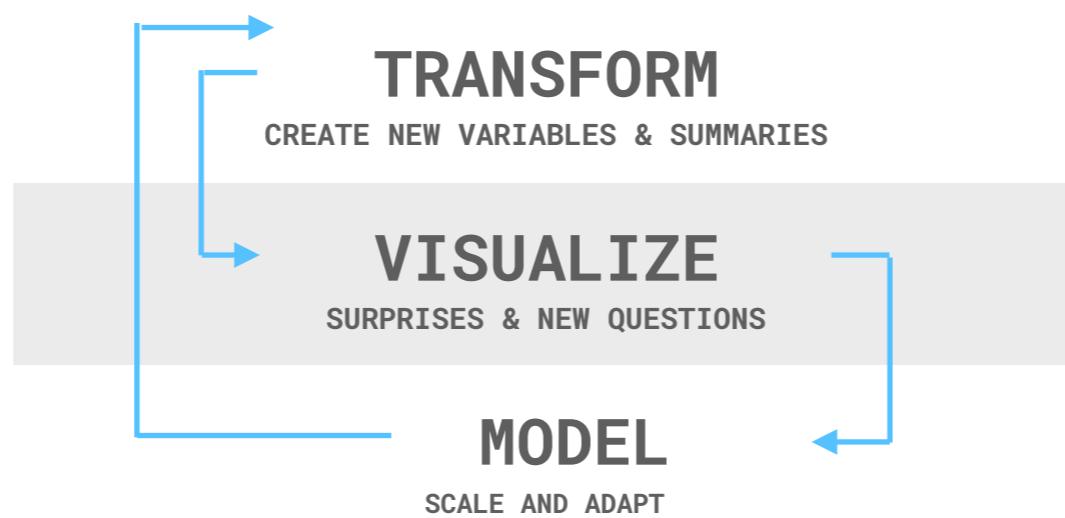
NOT NICE TO HAVE

DATAVIZ FOR

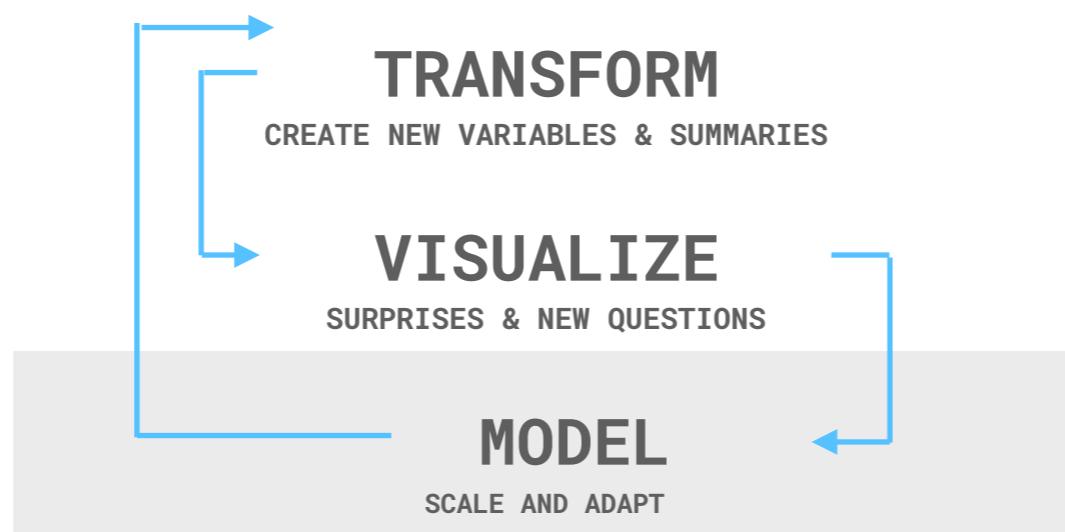
DATA ANALYSIS

IT'S FUNDAMENTAL

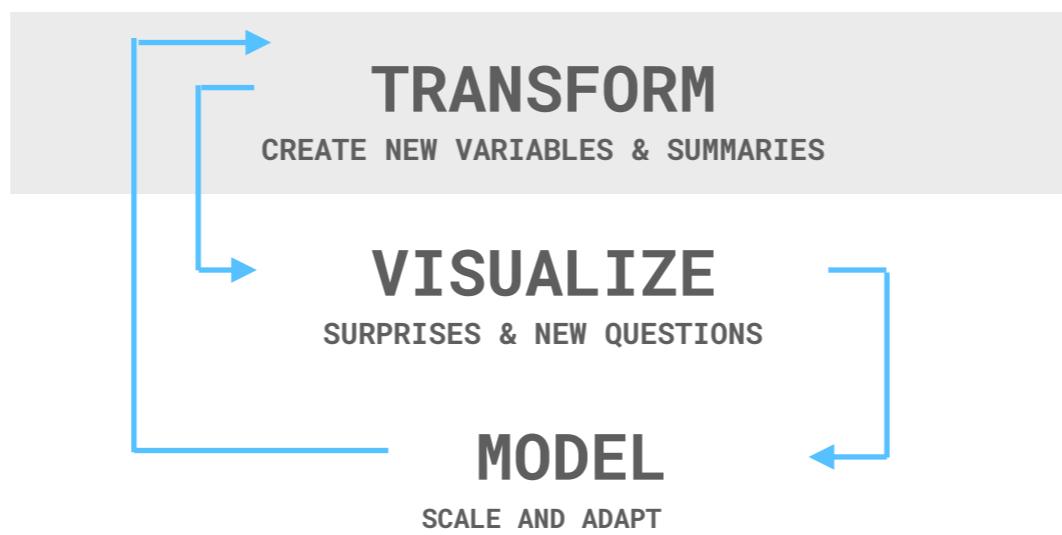
DATAVIZ FOR

**DATA ANALYSIS**

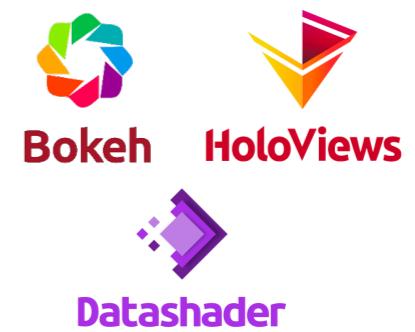
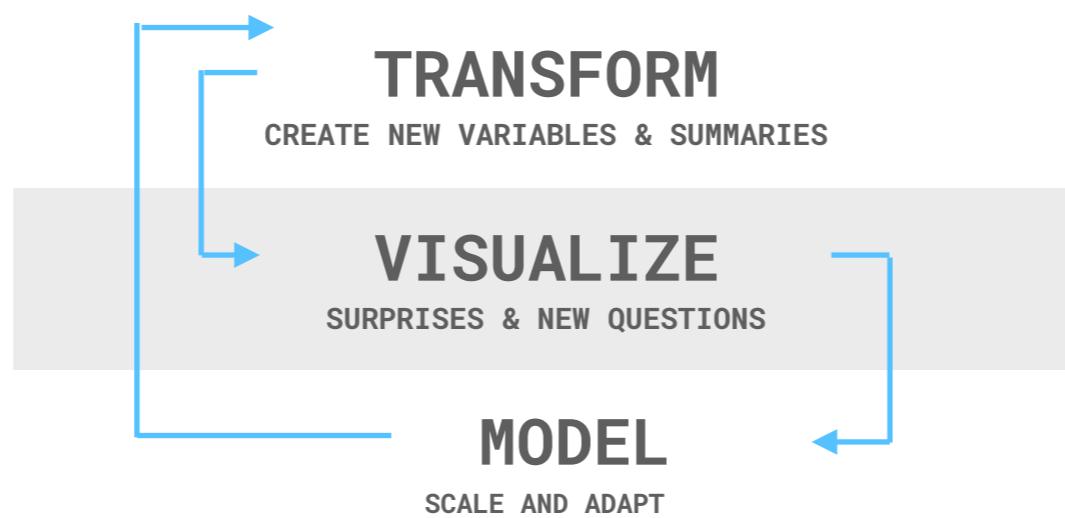
DATAVIZ FOR

**DATA ANALYSIS**

DATAVIZ FOR

**DATA ANALYSIS**

DATAVIZ FOR

**DATA ANALYSIS**

## DATA VISUALISATION

COMMUNICATION

CRAFTED  
ANNOTATED  
TELLS A STORY

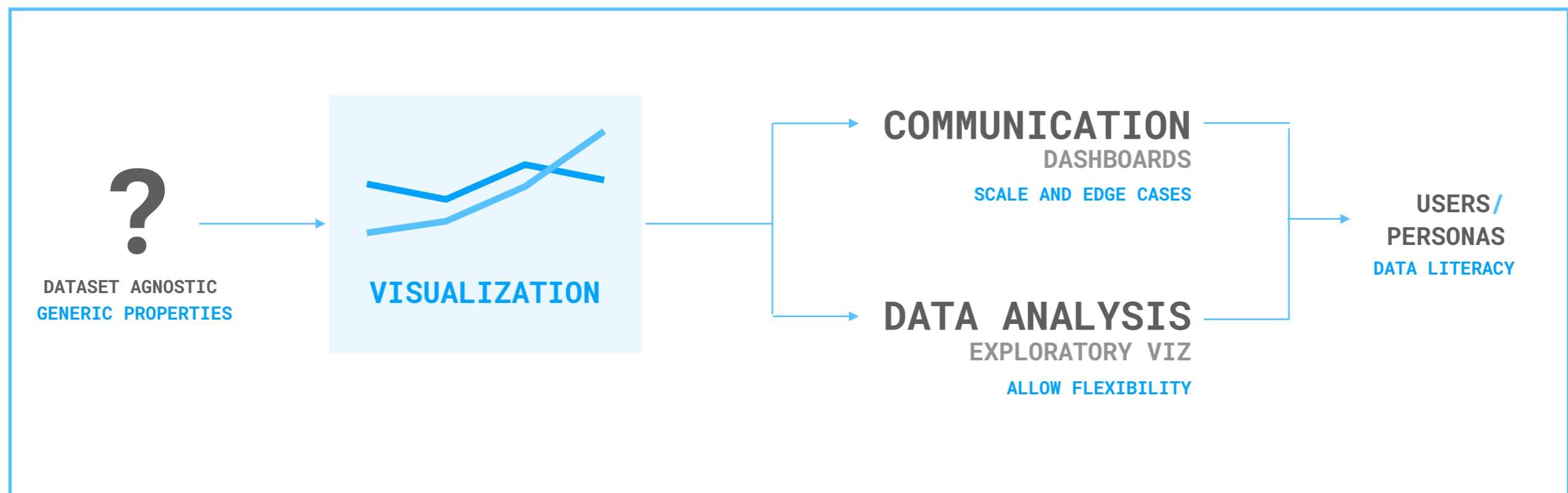
TIME TO PRODUCE PLOT

DATA ANALYSIS

EXPLORATORY  
ITERATIVE  
FAST GENERIC PLOT

TIME TO TRANSFORM PLOT

DATAVIZ FOR

**PRODUCT****CONTEXT**

DATAVIZ FOR

PRODUCT

USERS

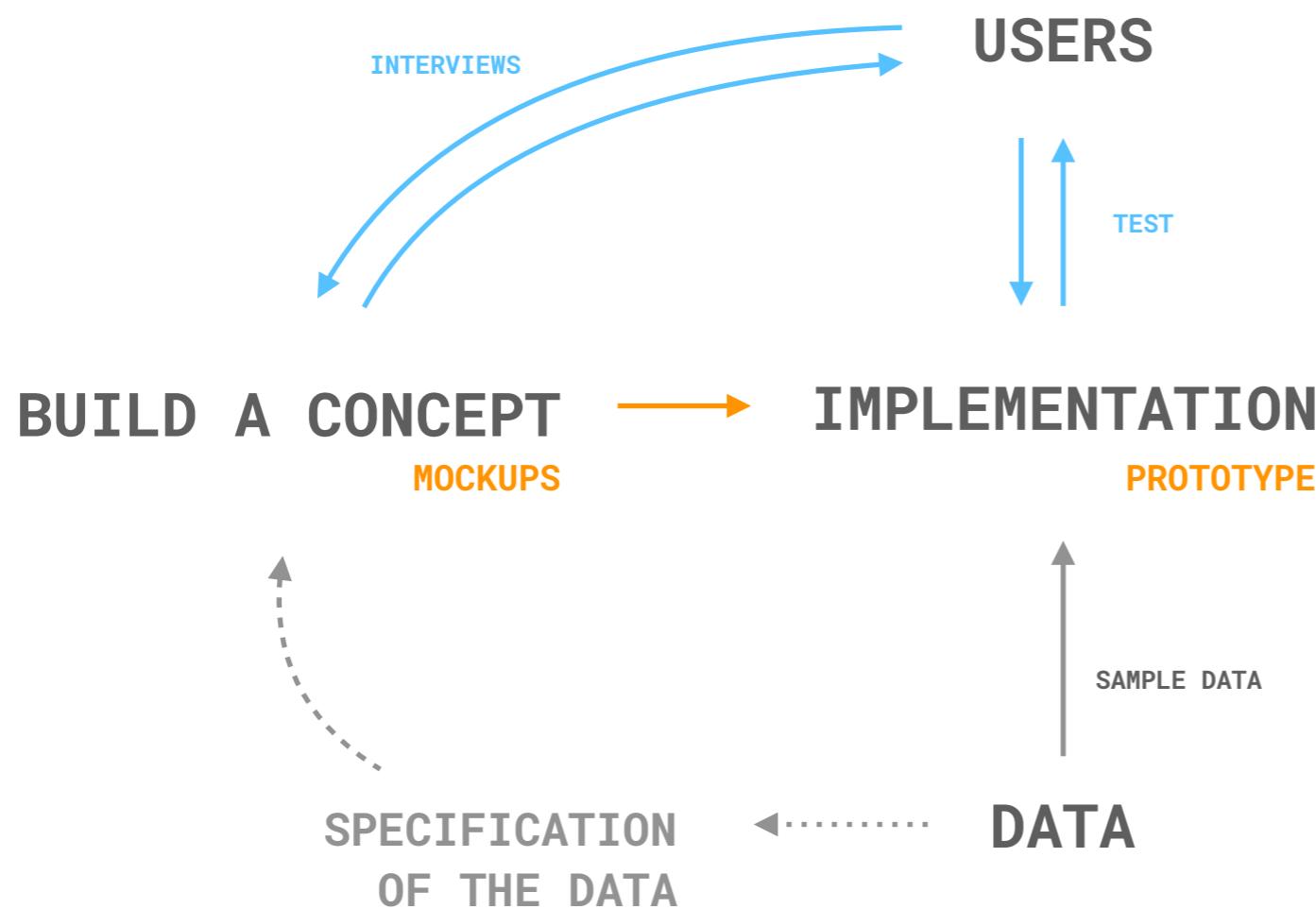
INTERVIEWS

BUILD A CONCEPT  
MOCKUPS

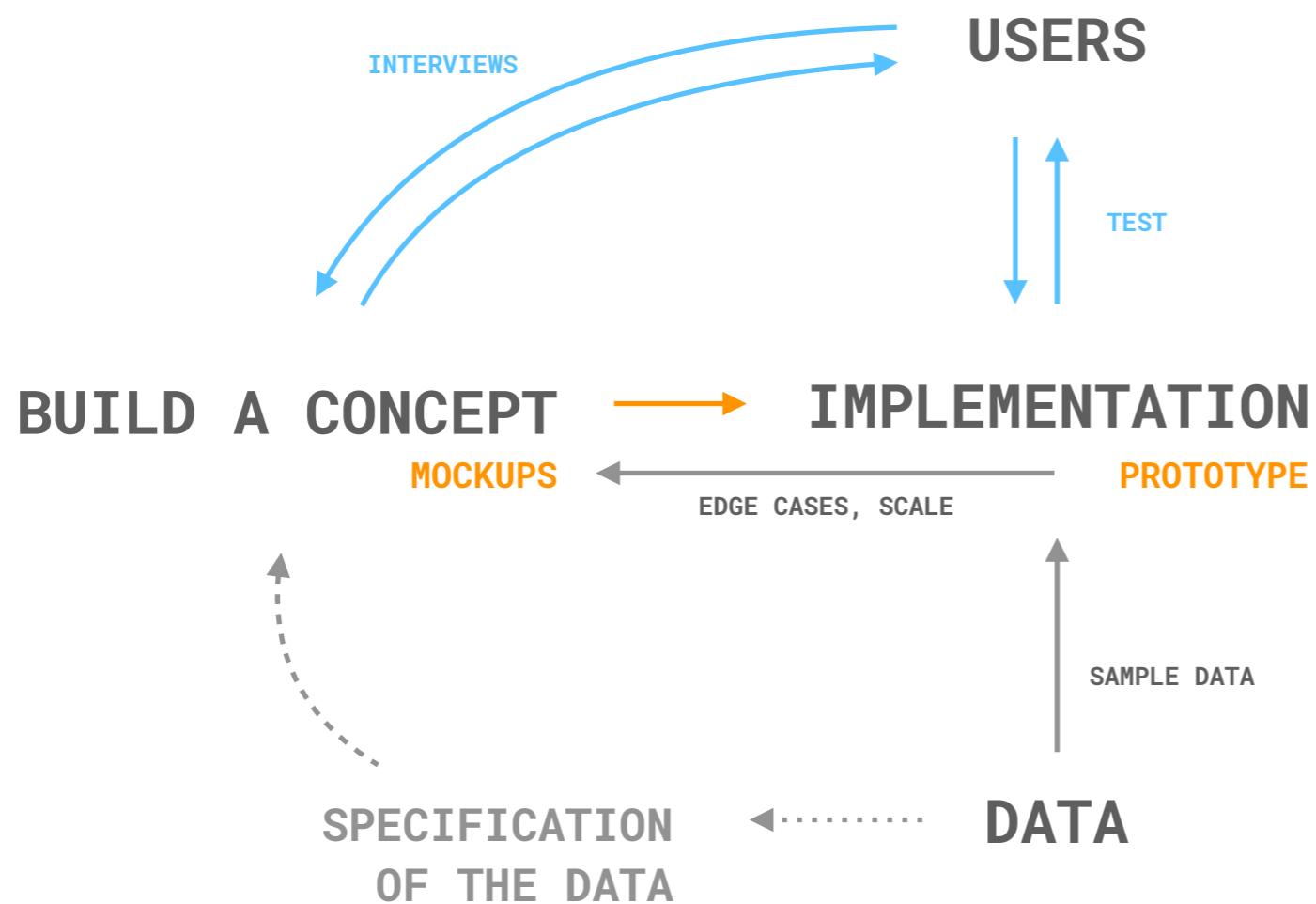
SPECIFICATION  
OF THE DATA



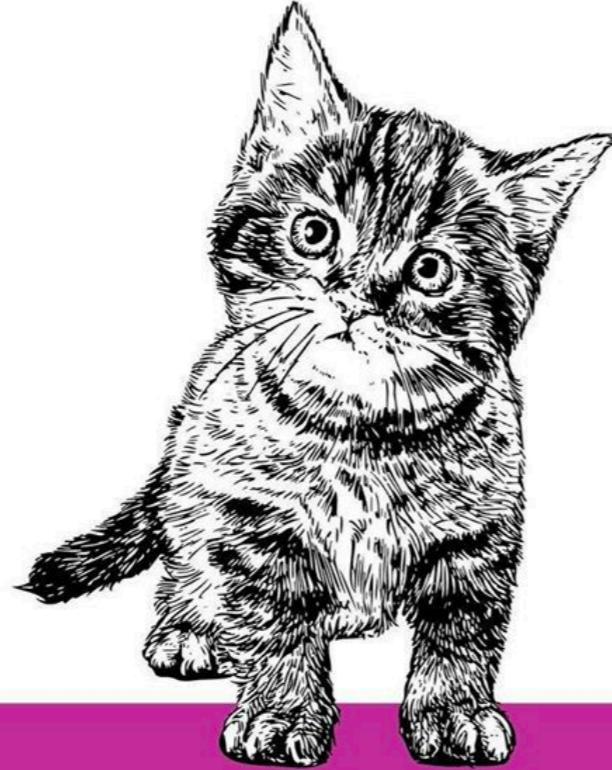
DATAVIZ FOR

**PRODUCT**

DATAVIZ FOR

**PRODUCT**

*How to actually learn any new programming concept*



*Essential*

Changing Stuff and  
Seeing What Happens

O RLY?

@ThePracticalDev

*Software can be chaotic, but we make it work*



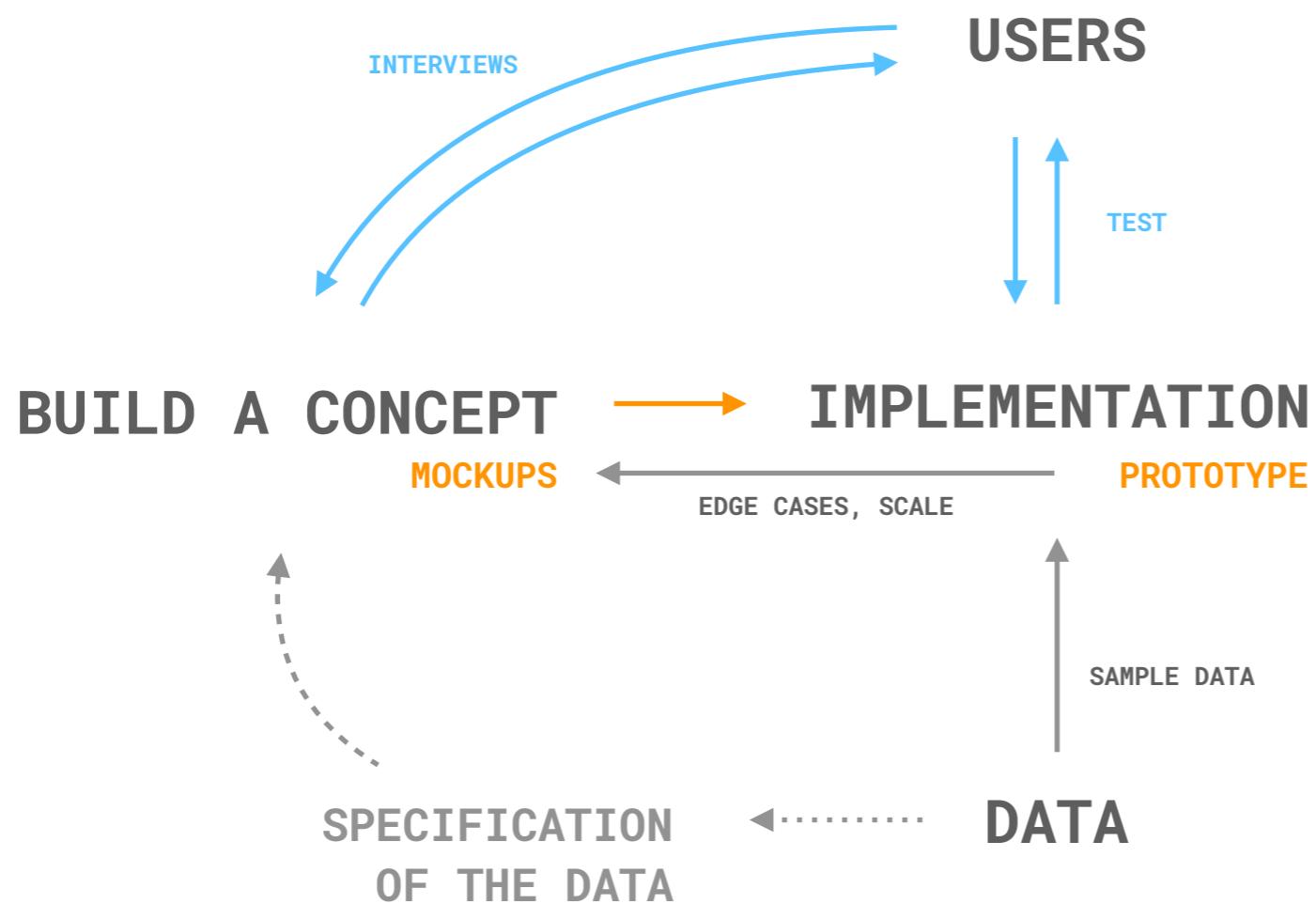
*Expert*

Trying Stuff  
Until it Works

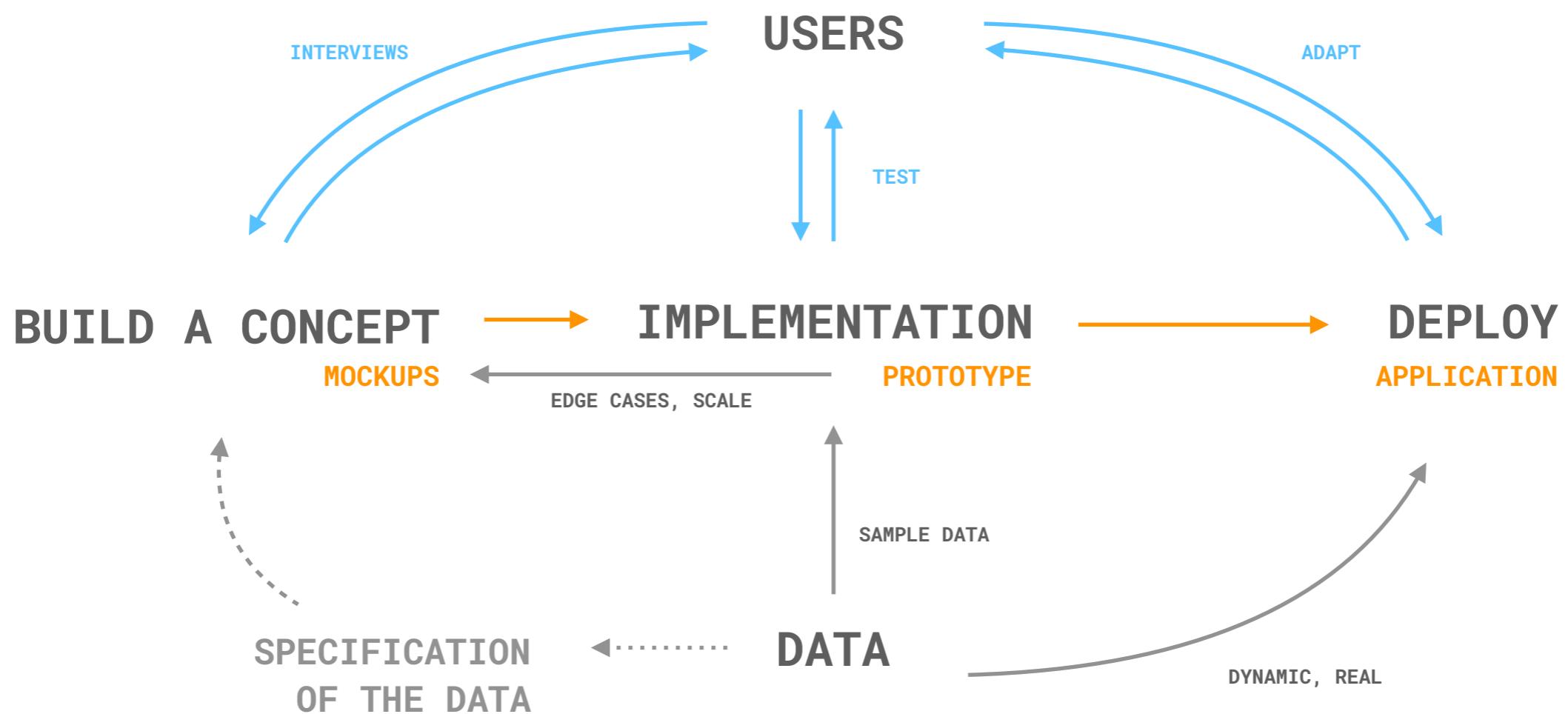
O RLY?

*The Practical Developer*  
@ThePracticalDev

DATAVIZ FOR

**PRODUCT**

DATAVIZ FOR

**PRODUCT**

DATAVIZ @feedzai

PLOTZAI

GOAL

**BIG DATA EXPLORATION**

EXPLORATORY VIZ

ITERATIVE

DATAVIZ @feedzai

PLOTZAI

GOAL

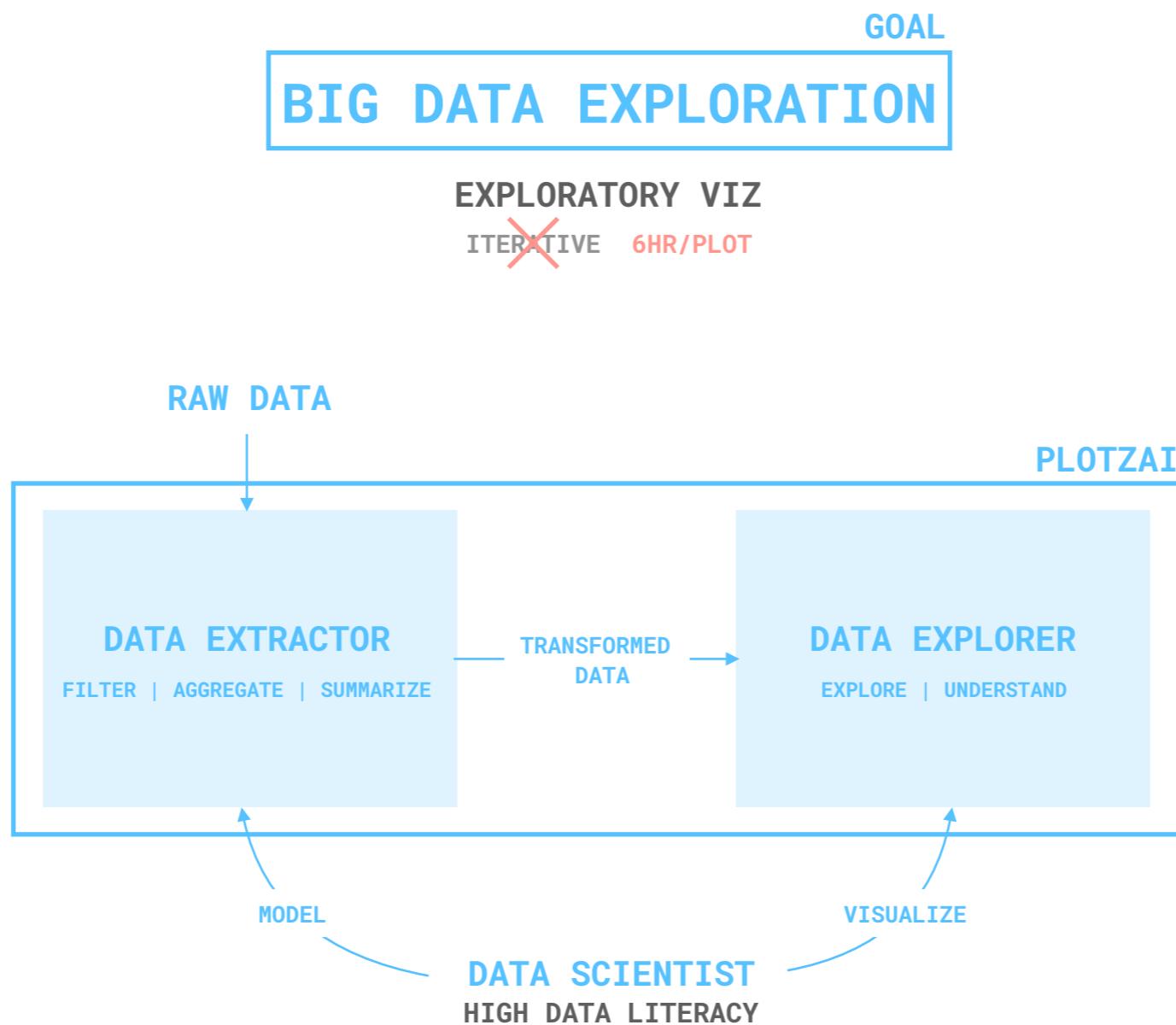
## BIG DATA EXPLORATION

EXPLORATORY VIZ

~~ITERATIVE 6HR/PLOT~~

DATAVIZ @feedzai

PLOTZAI

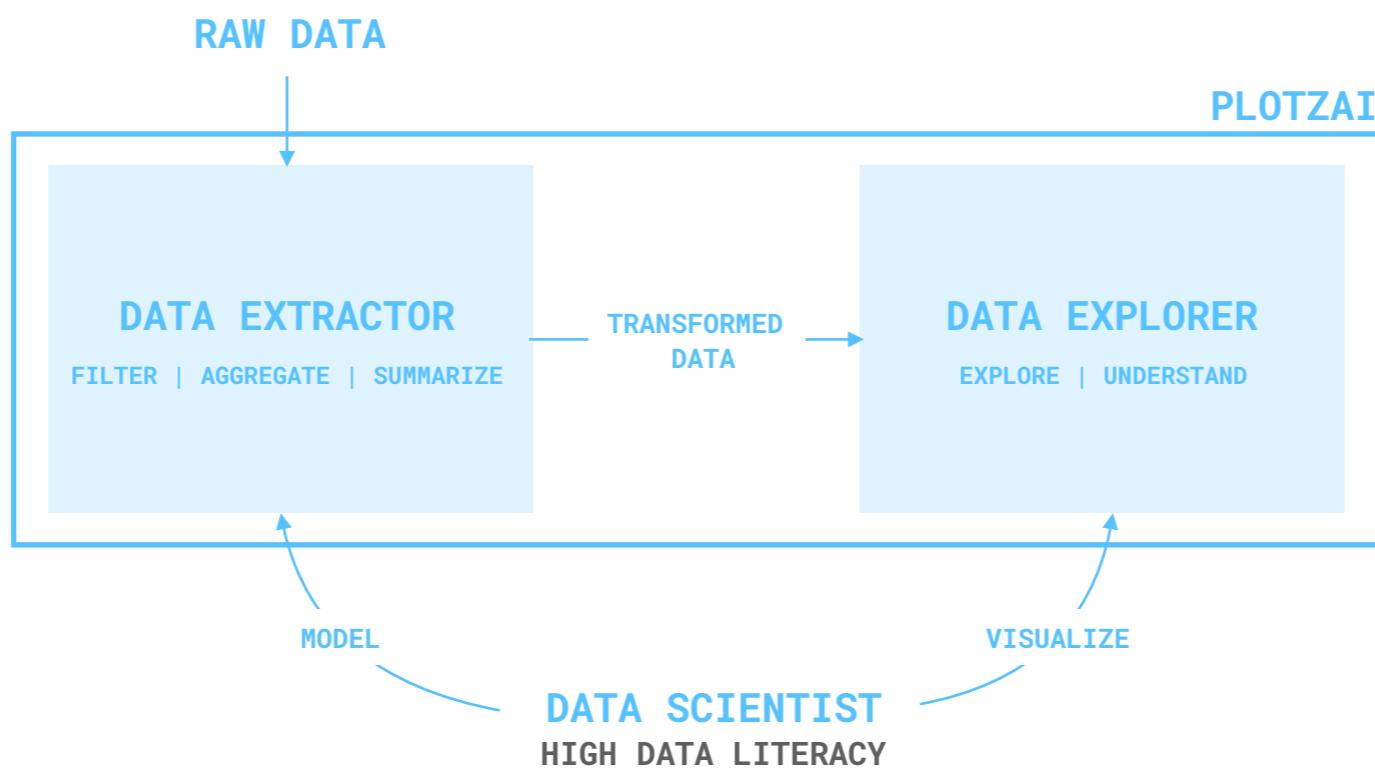


DATAVIZ @feedzai

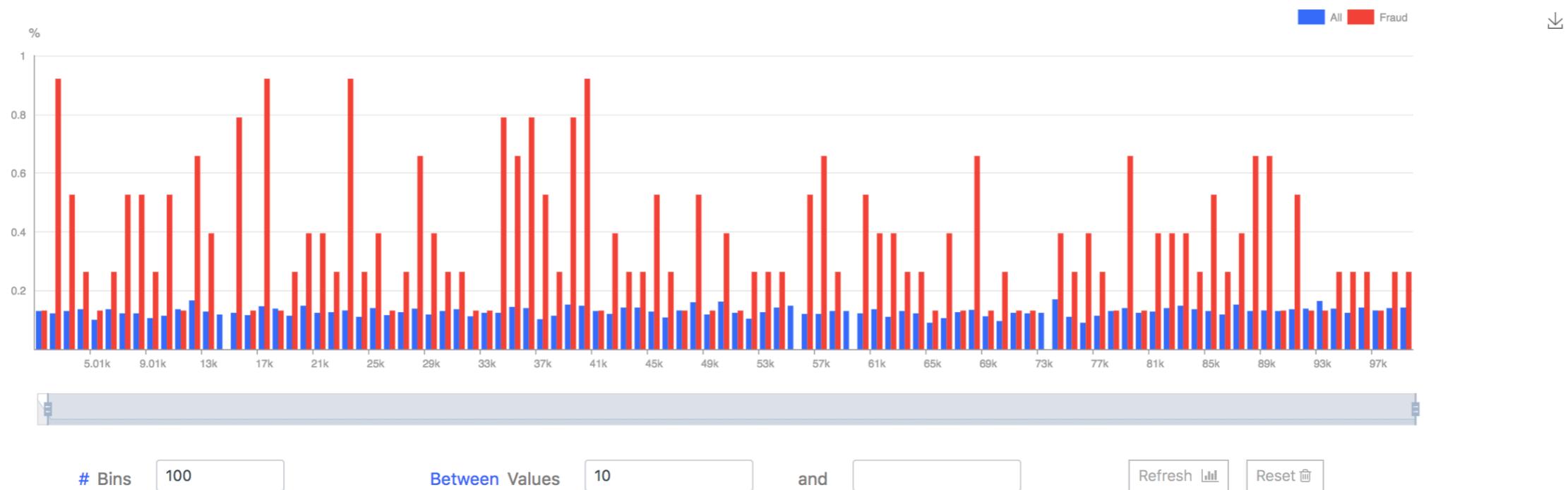
PLOTZAI

GOAL

## BIG DATA EXPLORATION + MODELS COMPARISION



## Amount



**SLAs**

Add goals and update the condition metric.

Goal: Cost-Recall = 0.75 +

Recall = 60% -

Cost-Recall = 75% -

@FPR = 10% ⚙ Settings

Plot Selected Models

**FILTER BY**

MODEL TYPE	Select... +	PARAMETERS	Select... Select... +	TIME TO RUN	Select... Select... S +
PERFORMANCE	Select... Select... +	DATASETS	TRAIN Select... +	VALIDATION	Select... + TEST Select... +

**CURRENT FILTERS:**

	Models				⚙ Customize	Performance				⚙ Customize	
	Parameters					Dataset Train	Dataset Validation	Recall Validation	Recall Over Time		
	Number Trees	Max Depth	IB Factor	Sampling							
	50	Default	0.65	0.8 undersampling		sydator_trx_train	sydator_trx_val	64.71%		78.39%	
	30	Default	0.65	0.8 undersampling		sydator_trx_train	sydator_trx_val	62.18%		71.17%	
	30	Default	0.65	0.6 undersampling		sydator_trx_train	sydator_trx_val	61.34%		78.25%	
	20	Default	0.65	0.6 undersampling		sydator_trx_train	sydator_trx_val	62.18%		70.67%	
	50	Default	0.65	0.6 undersampling		sydator_trx_train	sydator_trx_val	64.71%		82.23%	
	Number Trees										

## SLAs

Add goals and update the condition metric.

## Goals

Select... = +

@  
Select... = ↻

Recall = 60% -

Cost-Recall = 75% -

@FPR = 10% ⚙ Settings

## Models

List of selected models and their respective parameters

■ f1523276974432

Number Trees	Max Depth	IB Factor	Sampling
50	Default	0.65	0.8 undersampling

■ f1523276948911

Number Trees	Max Depth	IB Factor	Sampling
30	Default	0.65	0.8 undersampling

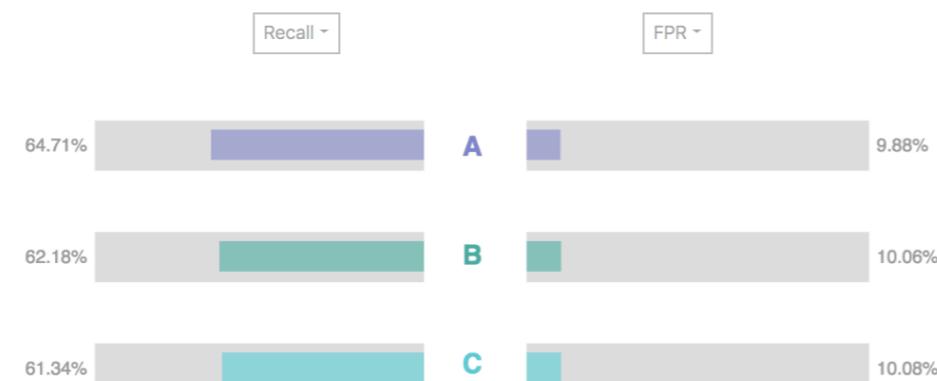
■ f1523276895714

Number Trees	Max Depth	IB Factor	Sampling
30	Default	0.65	0.6 undersampling

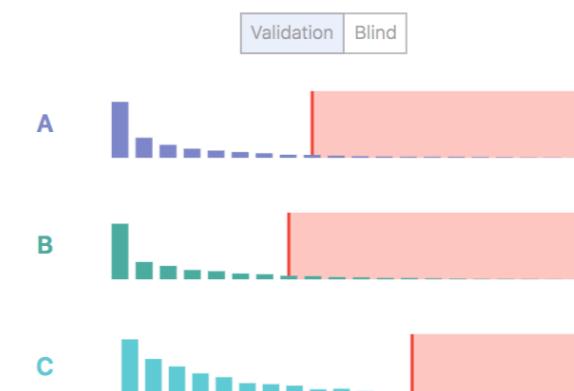
## PERFORMANCE OVER TIME



## PERFORMANCE OVERALL



## SCORE DISTRIBUTIONS

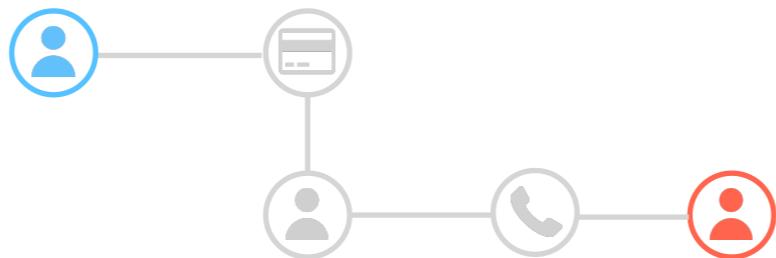


DATAVIZ @feedzai

GENOME

GOAL

## EXPLORE CONNECTED FRAUD

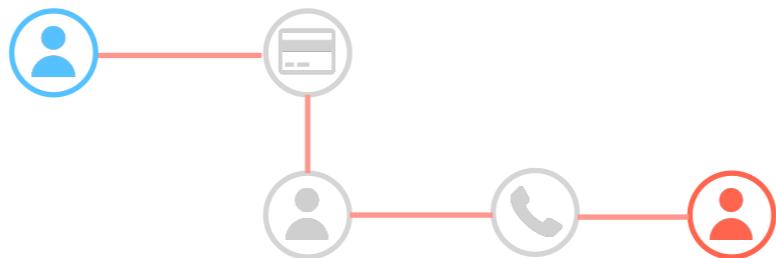


DATAVIZ @feedzai

GENOME

GOAL

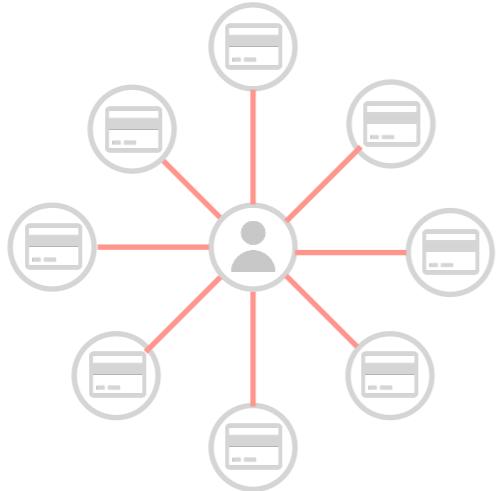
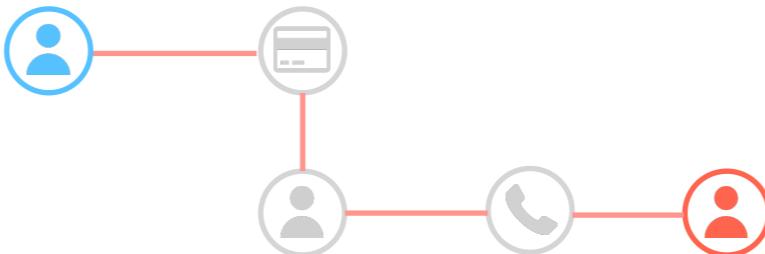
## EXPLORE CONNECTED FRAUD



GENOME

GOAL

## EXPLORE CONNECTED FRAUD



CARDING



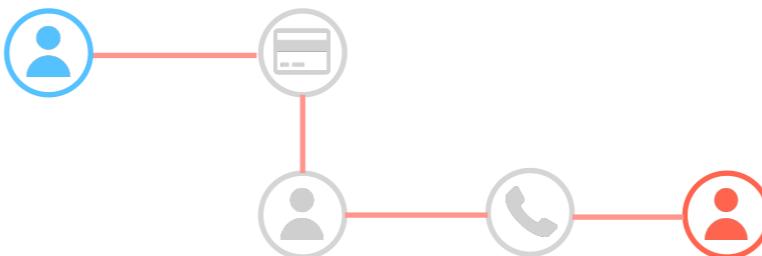
POINTS OF COMPROMISE

DATAVIZ @feedzai

GENOME

GOAL

## EXPLORE CONNECTED FRAUD



GENOME

CONTEXT  
INFORMATION  
UNDERSTAND

INTERACTION  
TOOLS

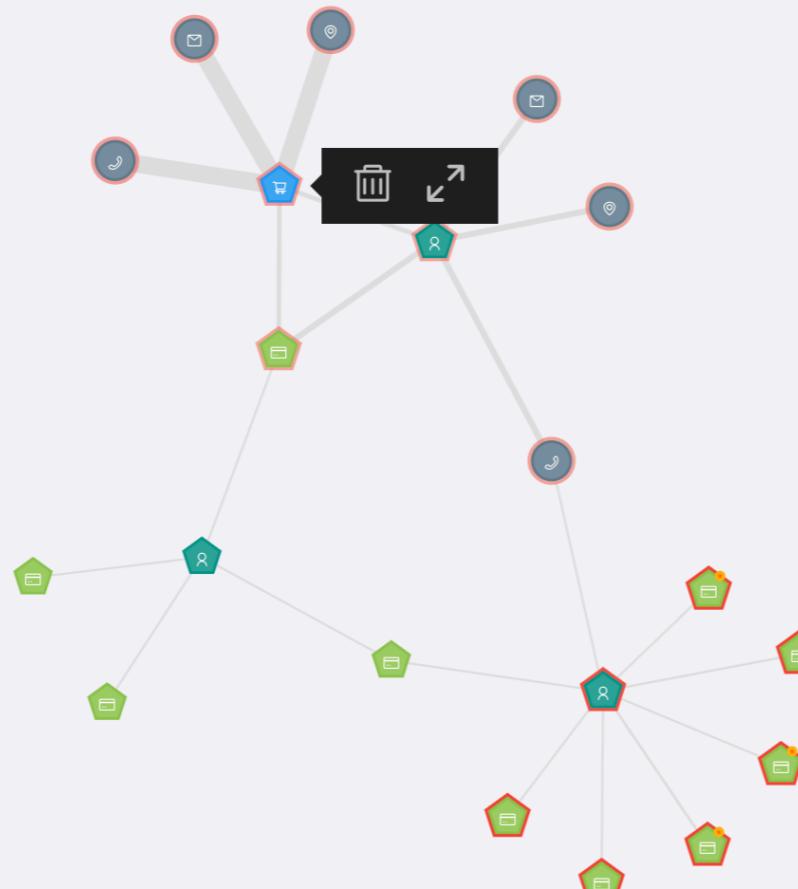
NETWORK  
NAVIGATOR  
EXPLORE

VISUALIZE

ACT

VISUALIZE

FRAUD ANALYST

*i* MAYER GROUP >

**MERCHANT**  
MAYER GROUP

72 EVENTS(#)  
21 #FRAUD  
\$5K AMOUNT(\$)  
\$3K \$FRAUD

✉ EMAIL(1) ↓↑ # \$

1. ESTELLA@EXA... 72

👤 CUSTOMER(9) ↓↑ # \$

1. JOVANI DOUG... 30  
2. MRS. KAREN ... 30  
3. ESTELLE LAN... 28

📞 PHONE(1) ↓↑ # \$

1. (252) 470-3... 72

📅 EVENT(72) ↓↑ # \$

1. D09CC234-37... 1  
2. C00B2523-79... 1  
3. 2C2ADC77-DF... 1

📍 CITY(1) ↓↑ # \$

1. GOODMAN 72

\* (20)

CARD (10)

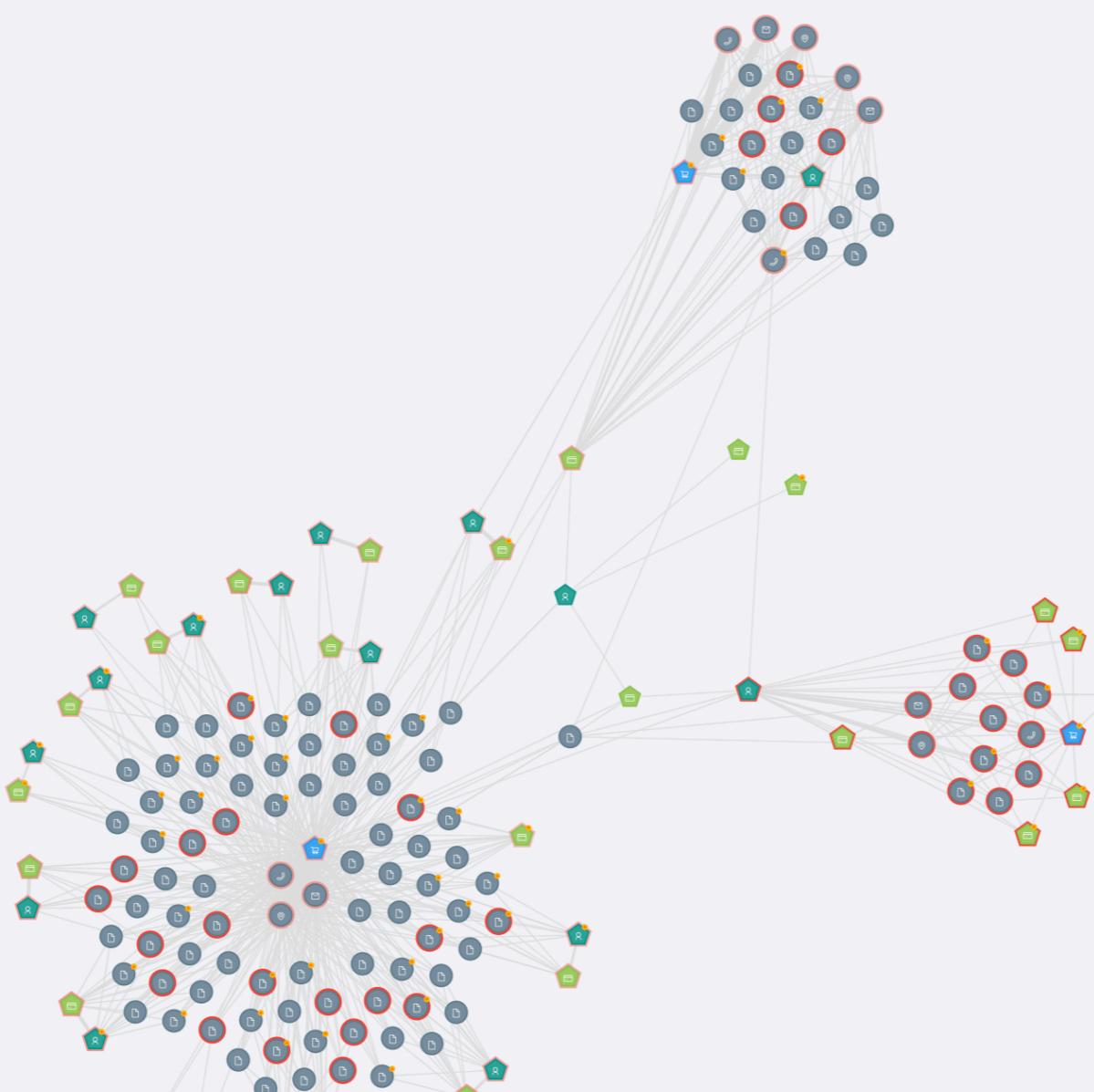
CITY (2)

CUSTOMER (3)

EMAIL (2)

MERCHANT (1)

PHONE (2)



\* (168)

CARD (25)

CITY (4)

CUSTOMER (17)

EMAIL (4)

EVENT (111)

MERCHANT (3)

PHONE (4)

i KEEBLER, MACGYVER AND ... >

#### MERCHANT

KEEBLER, MACGYVER AND LEH...

	83	20	\$6K	\$3K
EVENTS(#)	83	20	\$6K	\$3K
#FRAUD				
AMOUNT(\$)				
\$FRAUD				

✉ EMAIL(1) ↓↑ # \$

1.TITO\_WIZA@E... 83

✉ CUSTOMER(16) ↓↑ # \$

1.MS. LOWELL ... 31  
2.ESTELLE LAN... 28  
3.HUDSON WHIT... 25

☎ PHONE(1) ↓↑ # \$

1.(813) 240-4... 83

⌚ EVENT(83) ↓↑ # \$

1.5C8C5323-D2... 1  
2.5FC4C61F-D2... 1  
3.9A132D86-EC... 1

📍 CITY(1) ↓↑ # \$

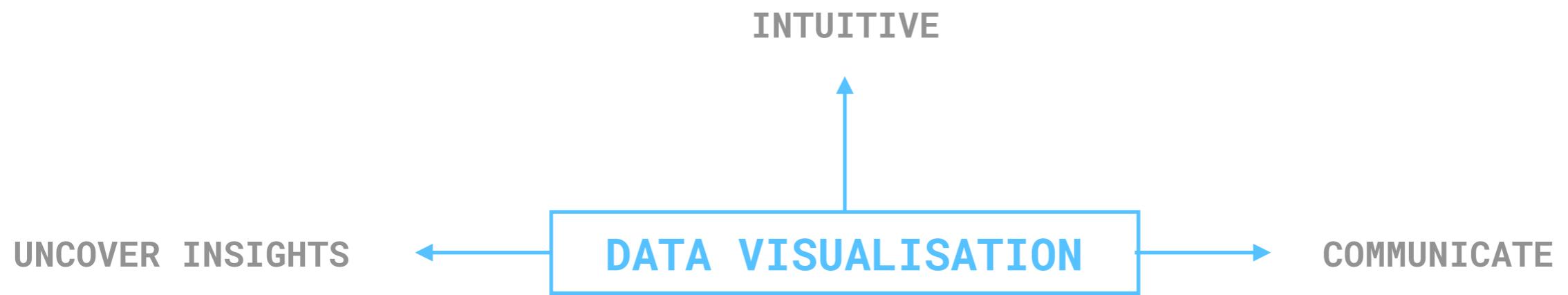
1.TORRANCE 83

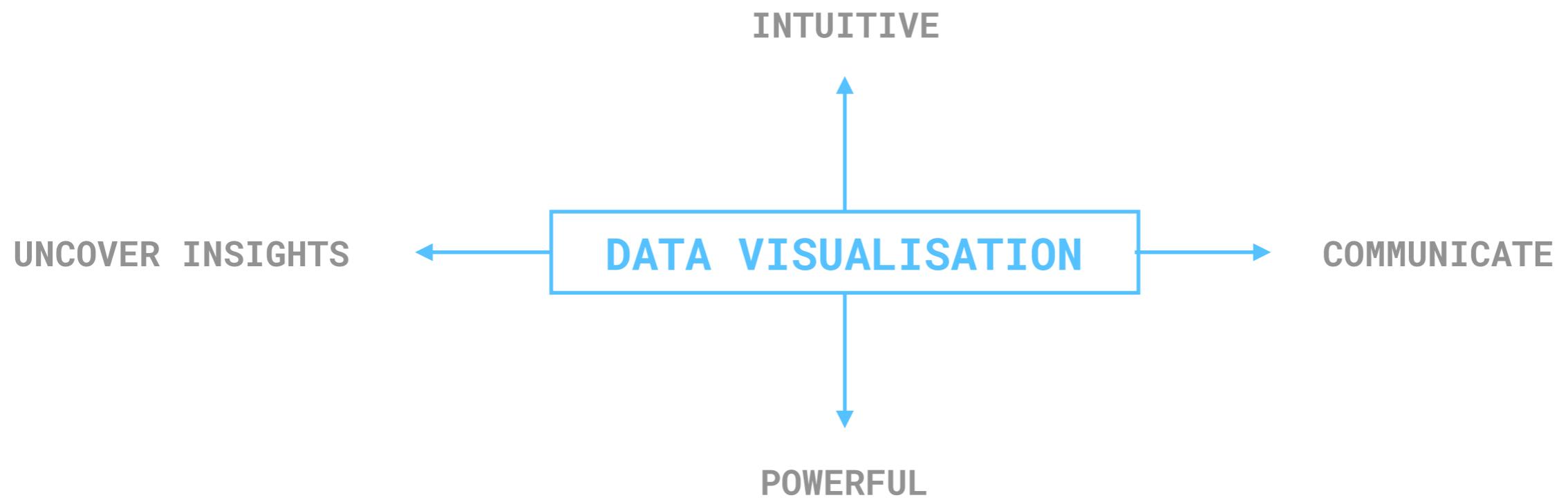
# DATA VISUALISATION

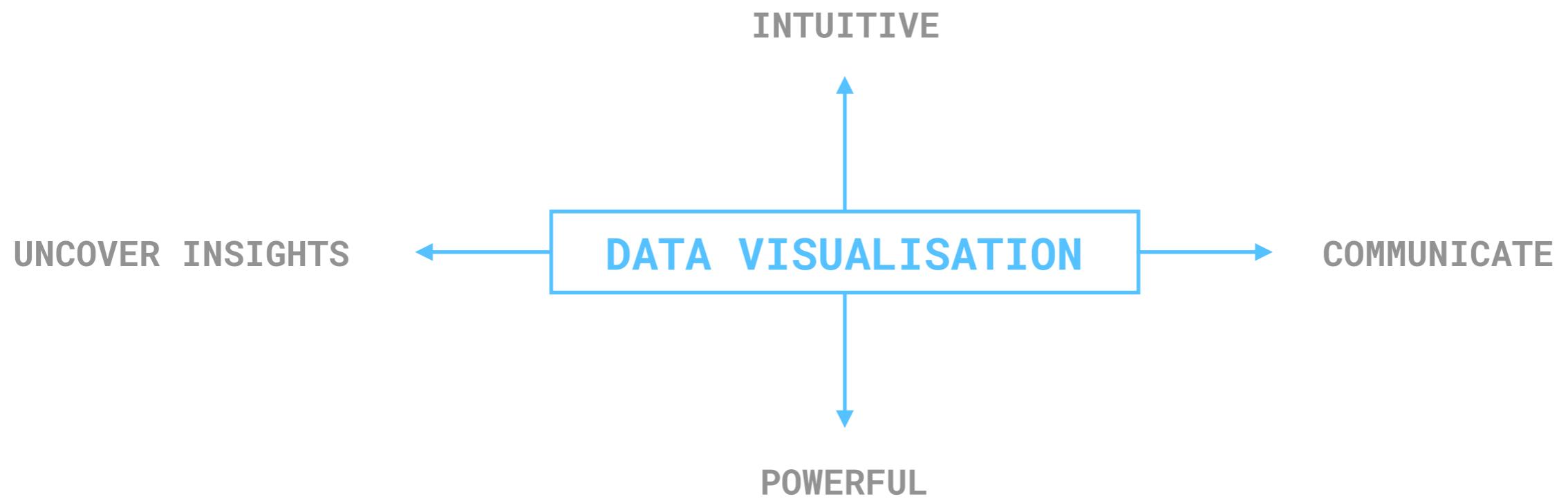
UNCOVER INSIGHTS

DATA VISUALISATION









*DATA SCIENTISTS, DON'T GO IN BLIND. SEE THE DATA*



# QUESTIONS?

THANK YOU