



datascope

Iterative design for data science projects

Bo Peng • @bo_p

for QCon San Francisco • Nov 7, 2016

approach

case study: heritage health prize

**Goal: Create an algorithm
that predicts how many
days a patient will spend in
a hospital in the next year.**

approach

case study: heritage health prize

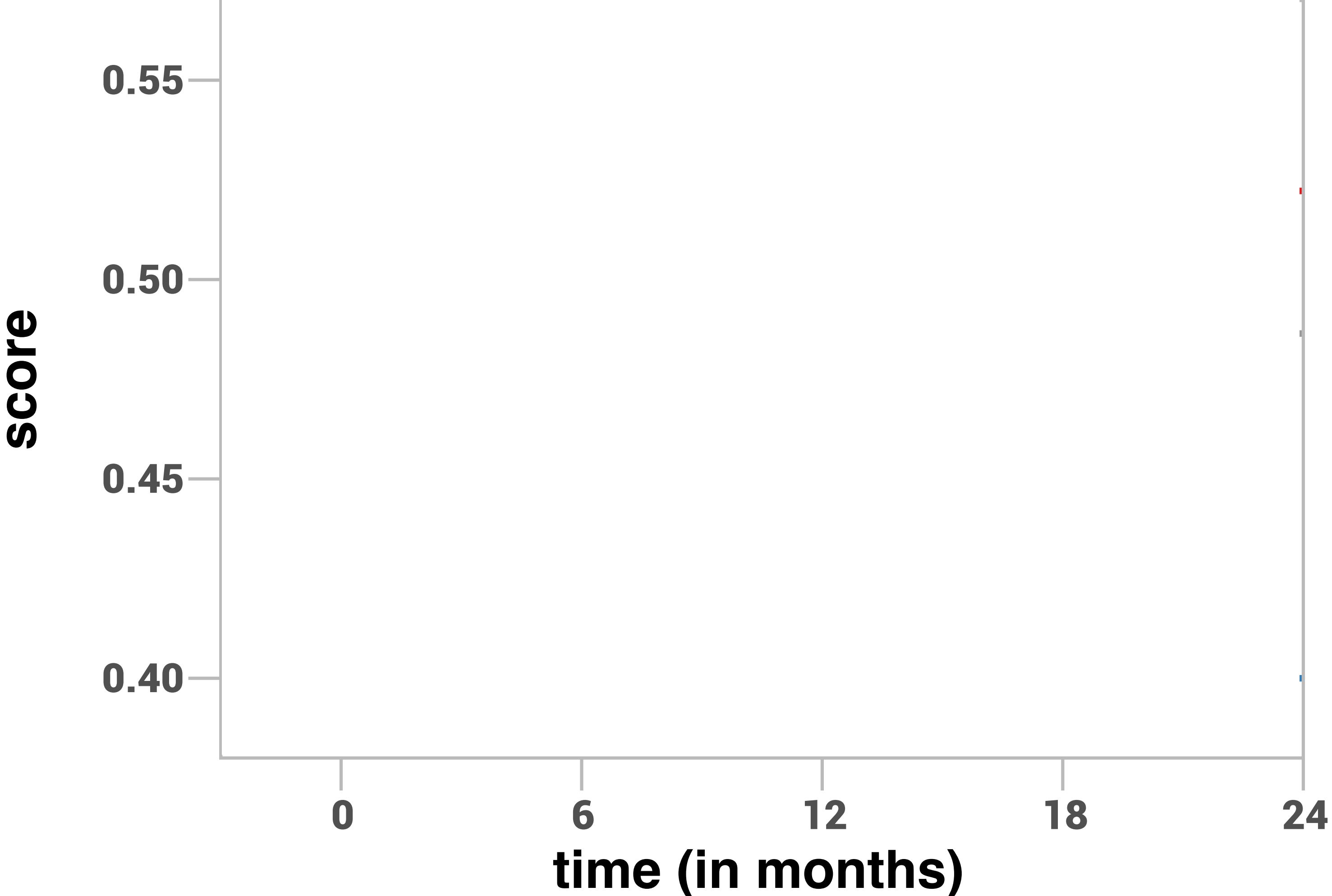
2 years

1,363 teams

25,316 entries

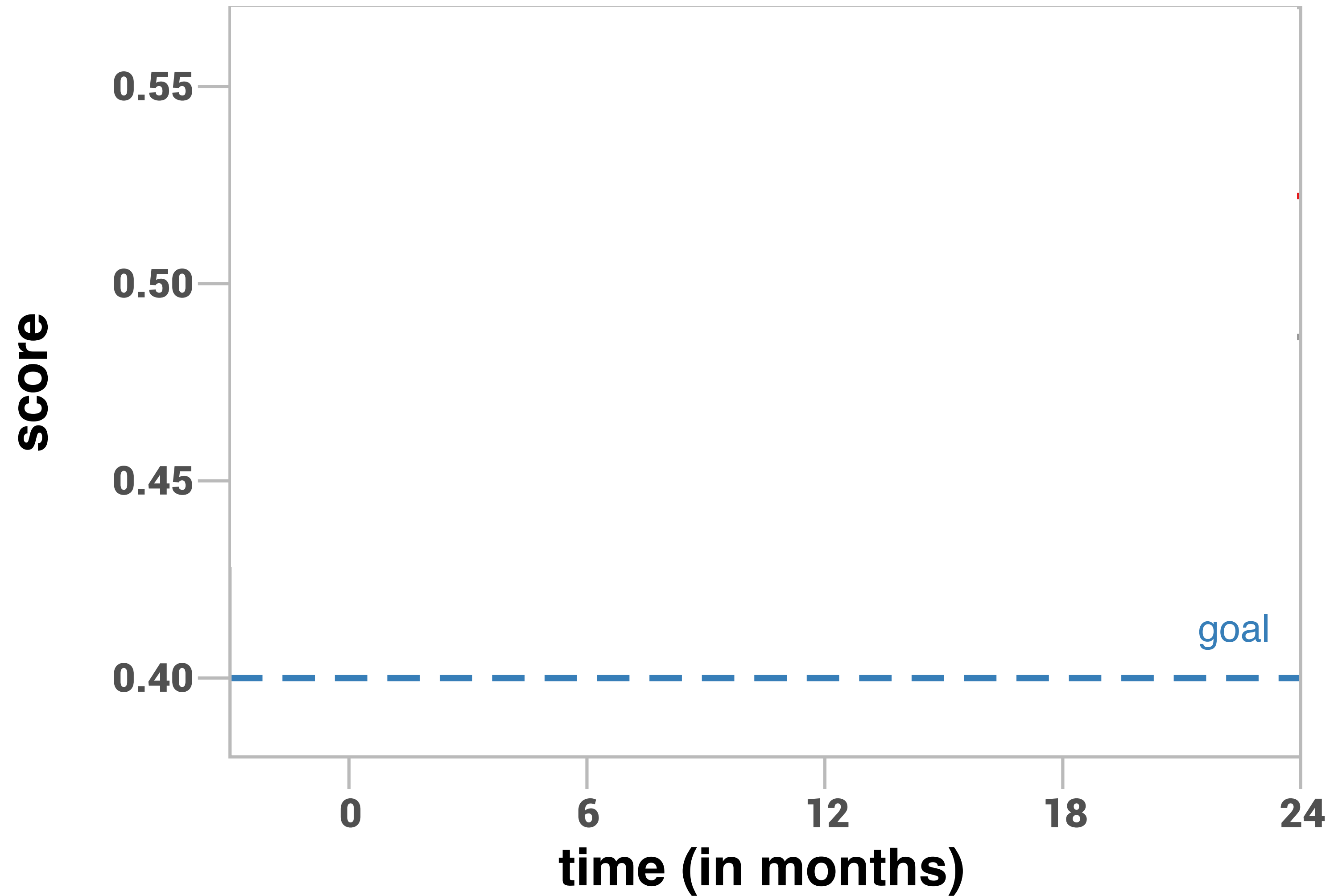
approach

case study: heritage health prize



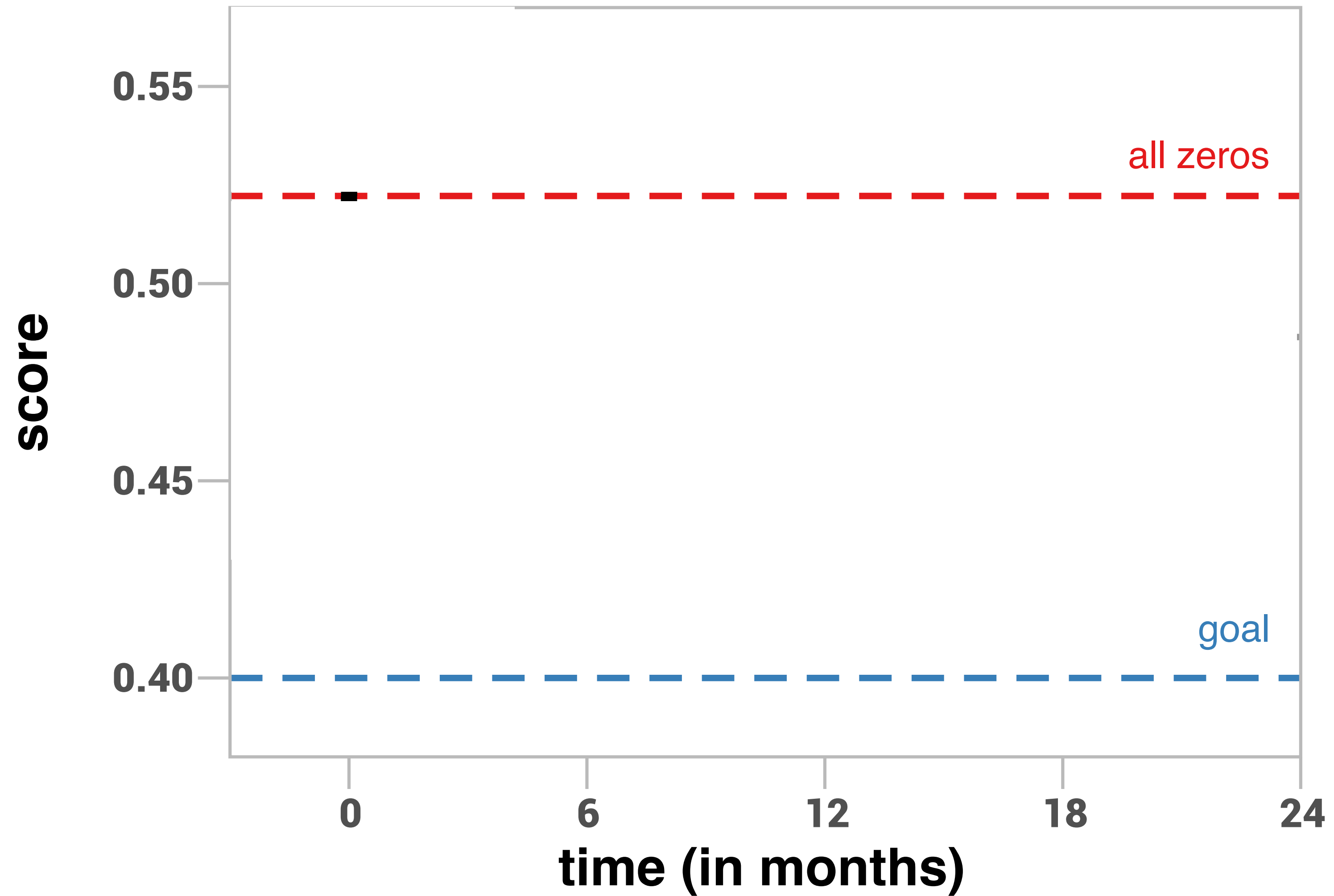
approach

case study: heritage health prize



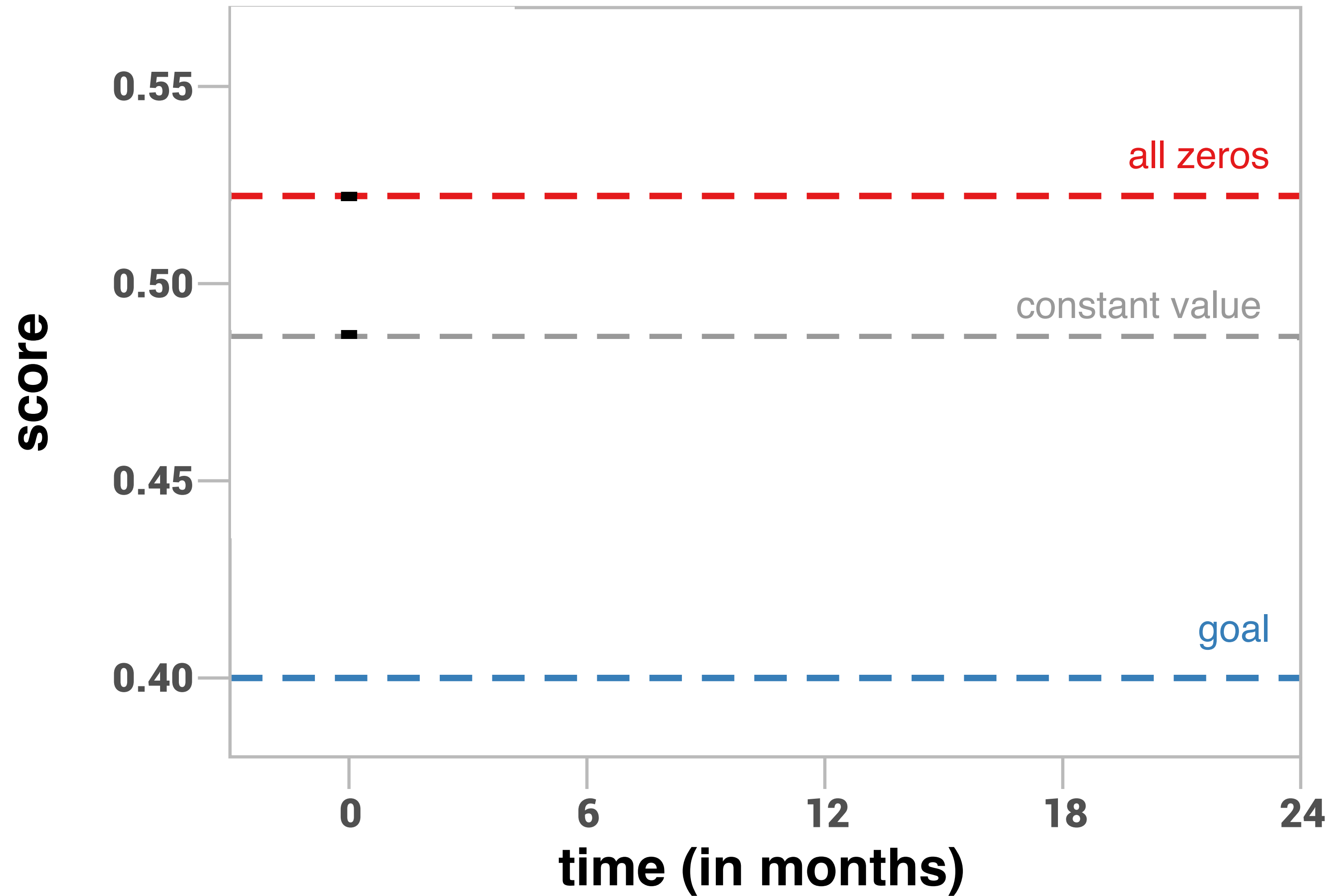
approach

case study: heritage health prize



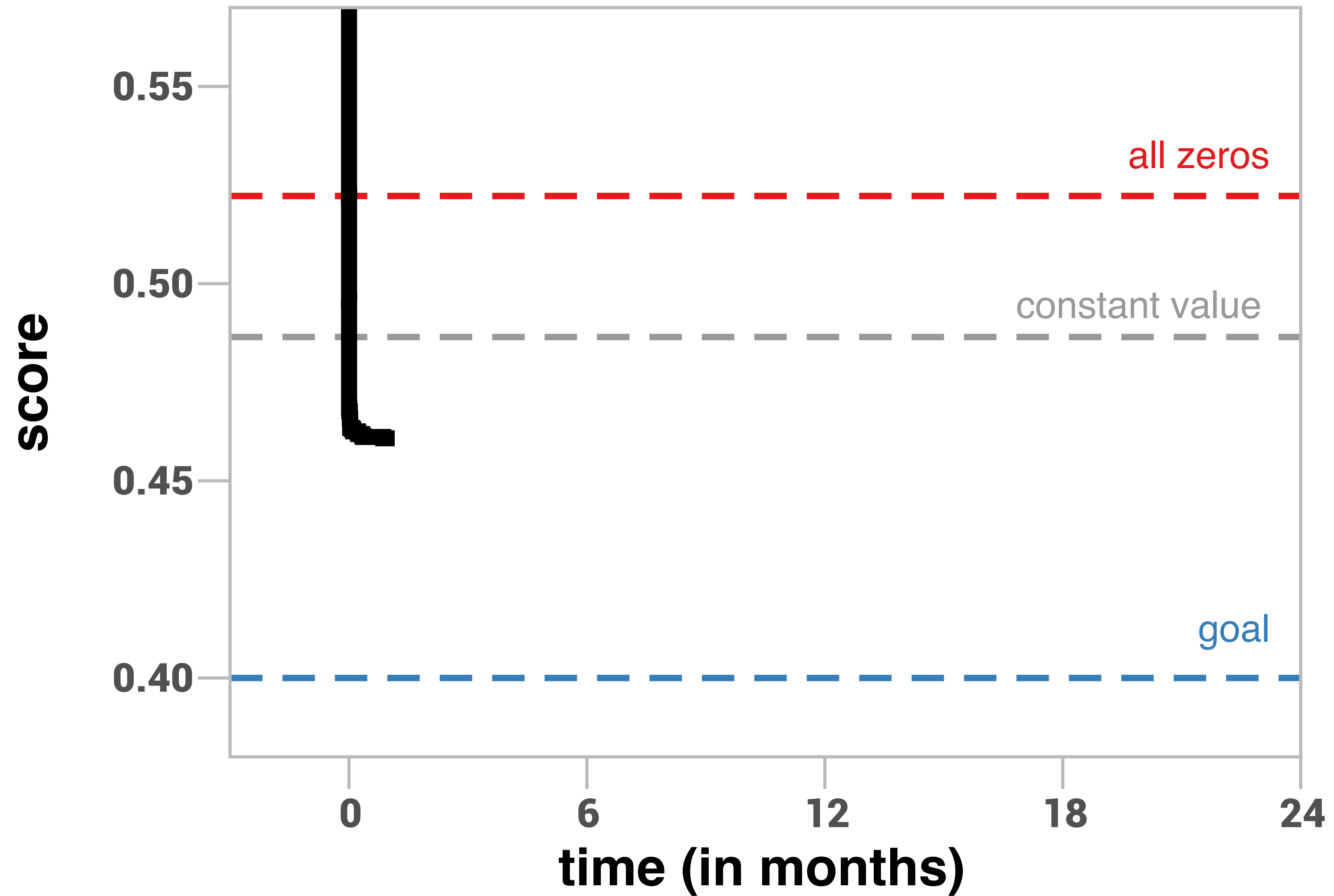
approach

case study: heritage health prize



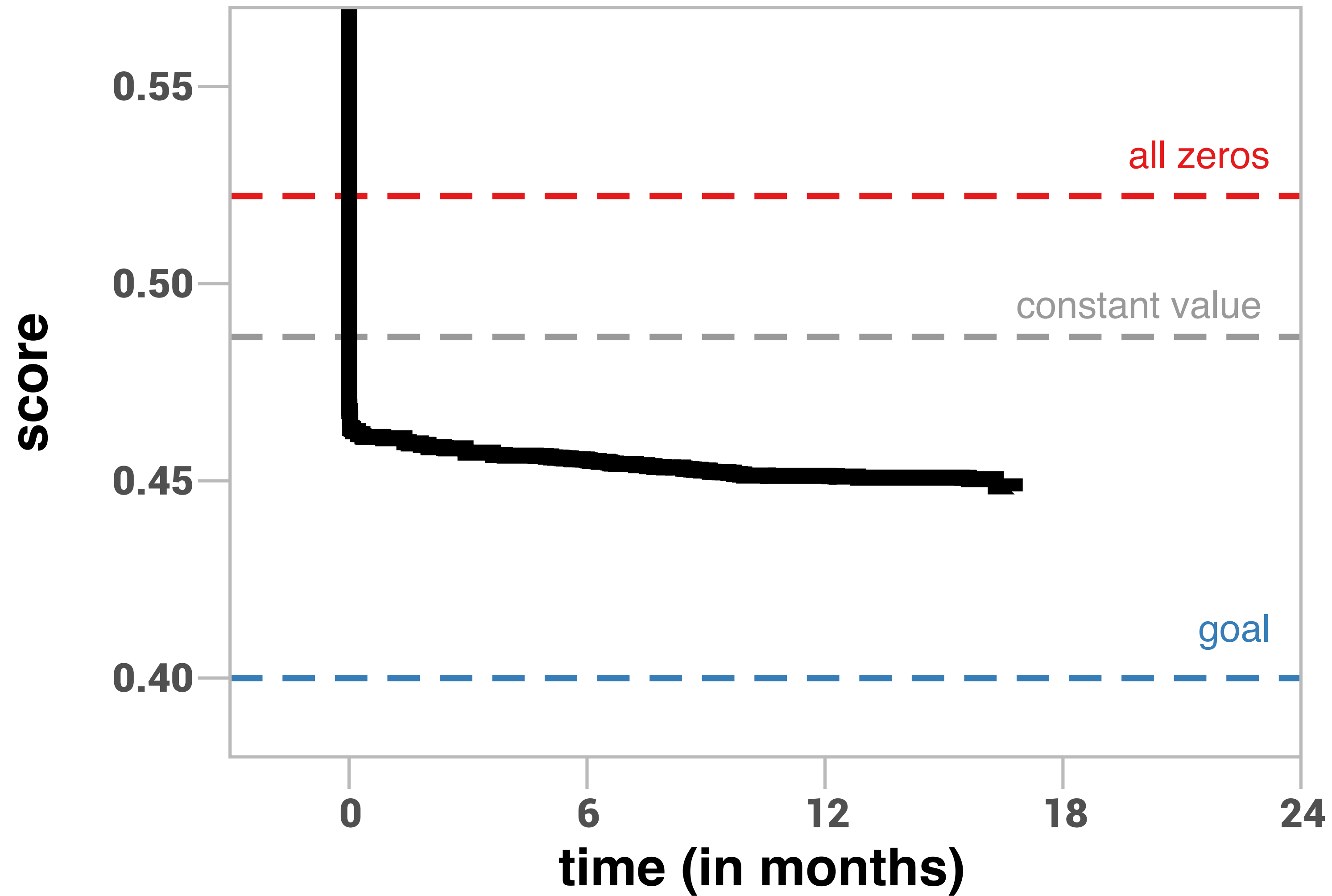
approach

case study: heritage health prize

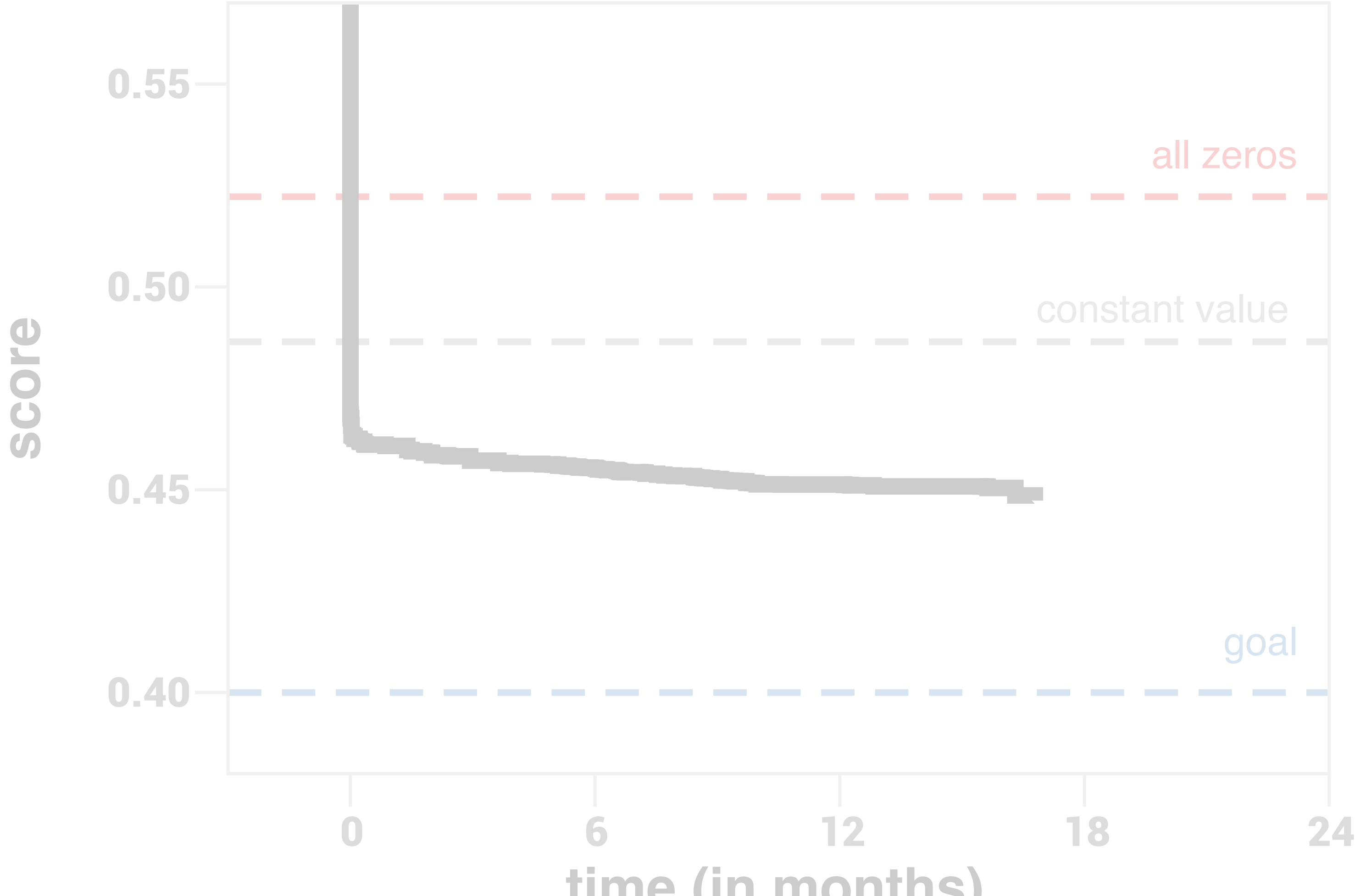


approach

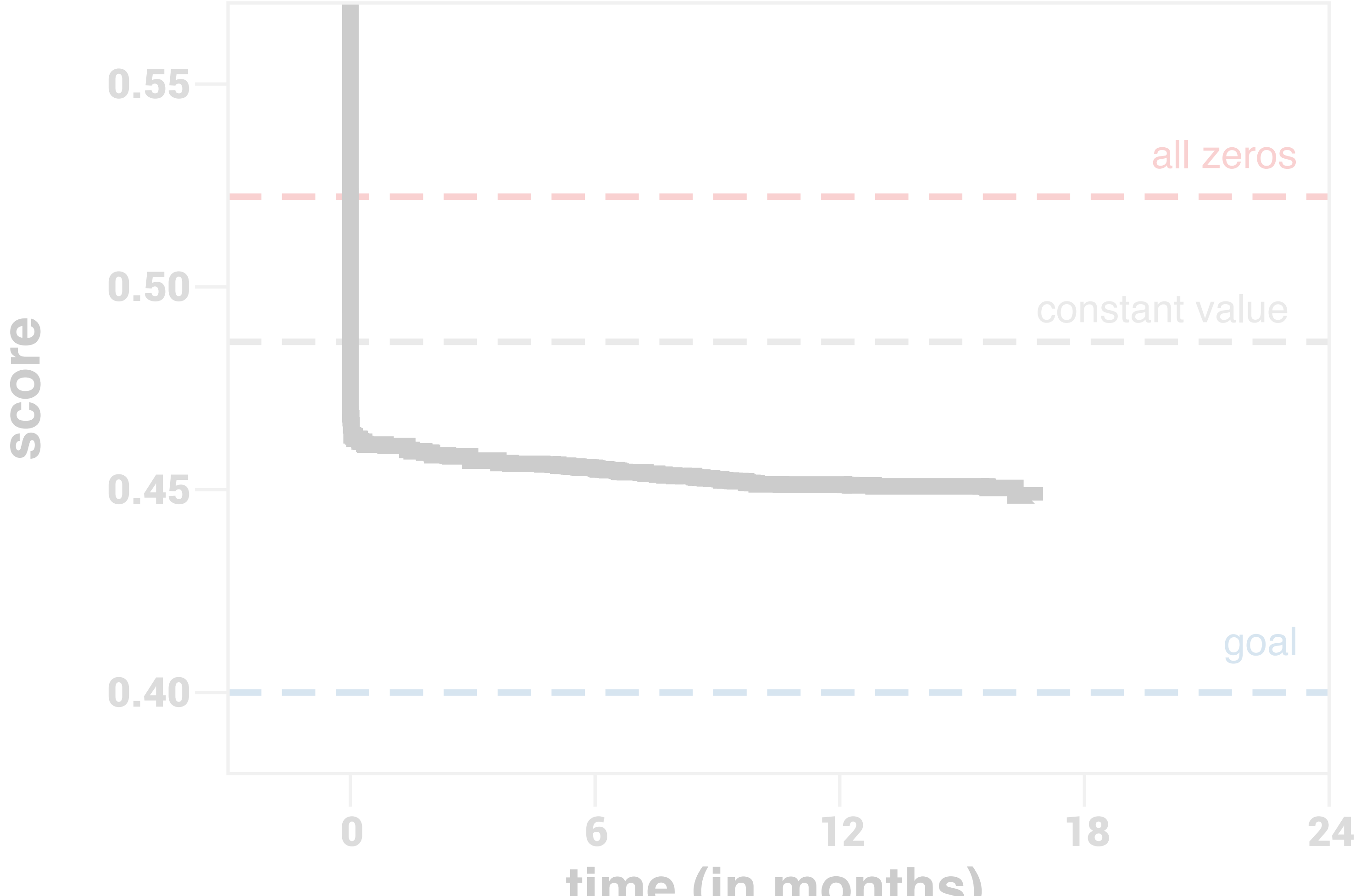
case study: heritage health prize



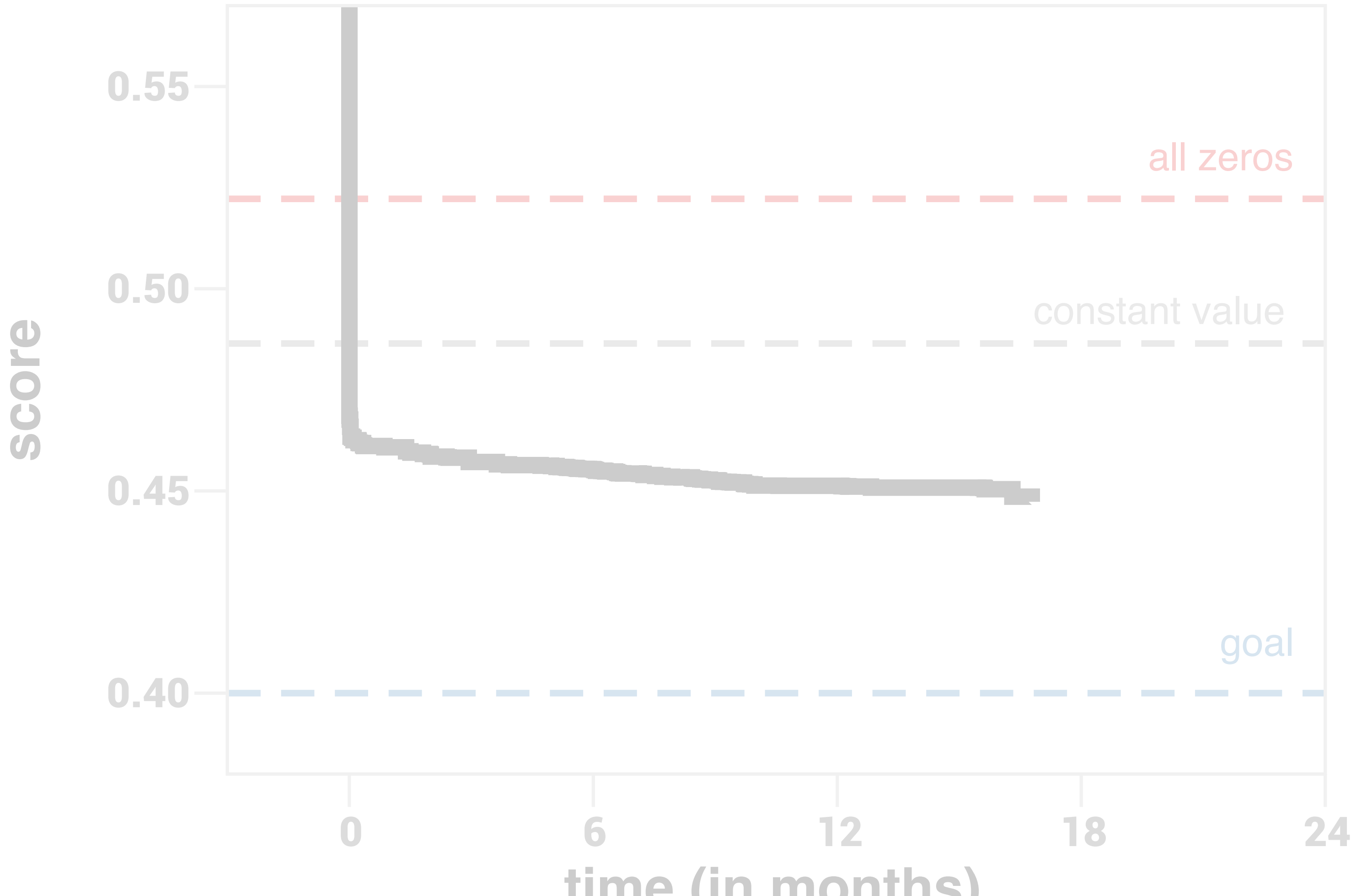
What can we learn from this?



Contests are fun.

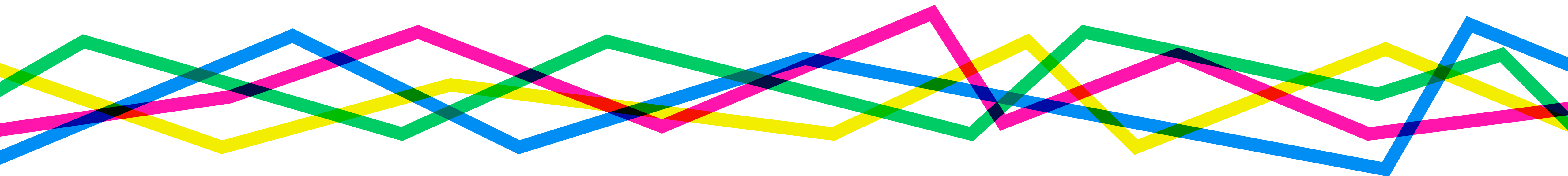


Contests are fun.
Solving business problems can rarely be
reduced to minimizing a model's RMSE.



agenda

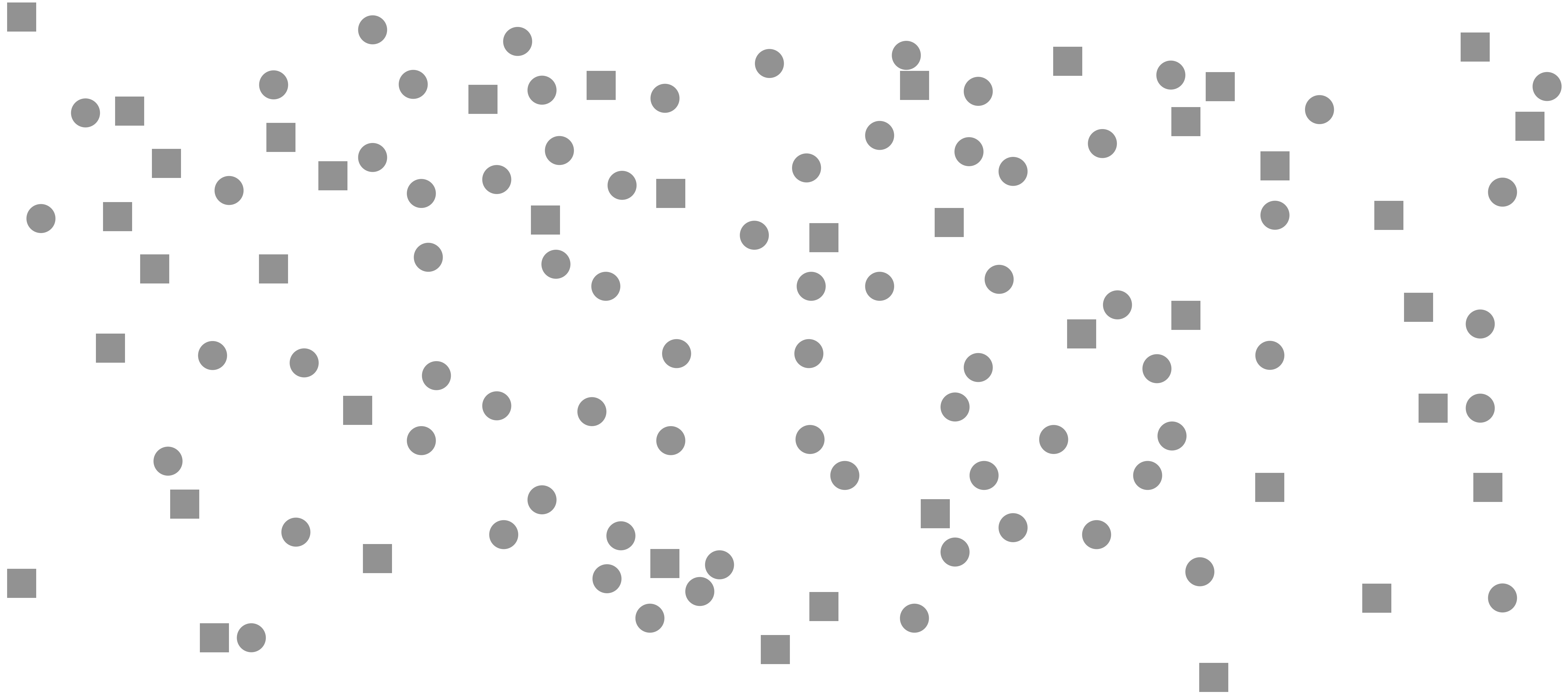
- A common approach to data science
- The design approach:
 - a simple model goes along way (eDiscovery)
 - finding & recommending experts within P&G



Data driven e-discovery for Daegis
How simple models + design go a long way

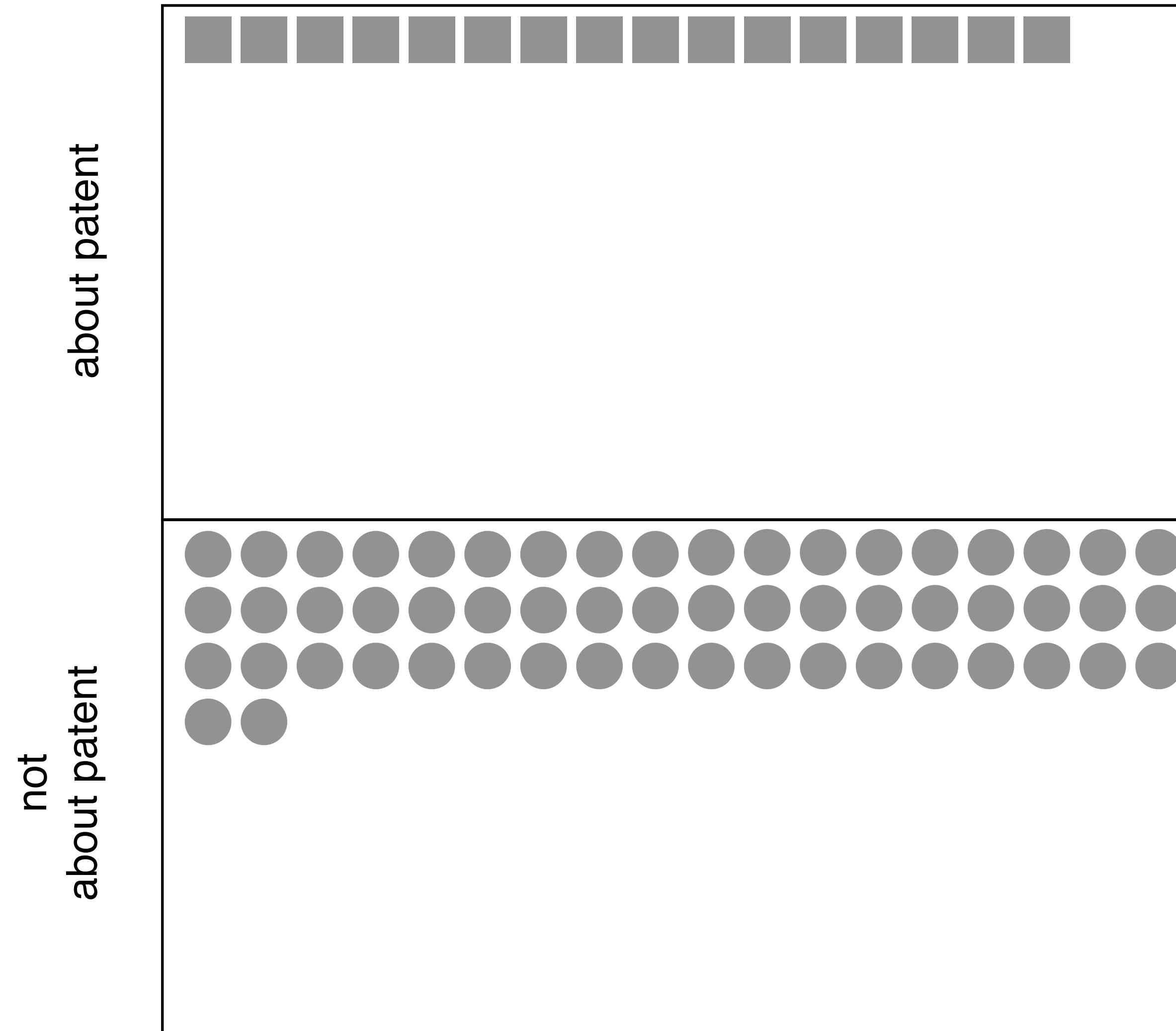
data-driven e-discovery

daegis



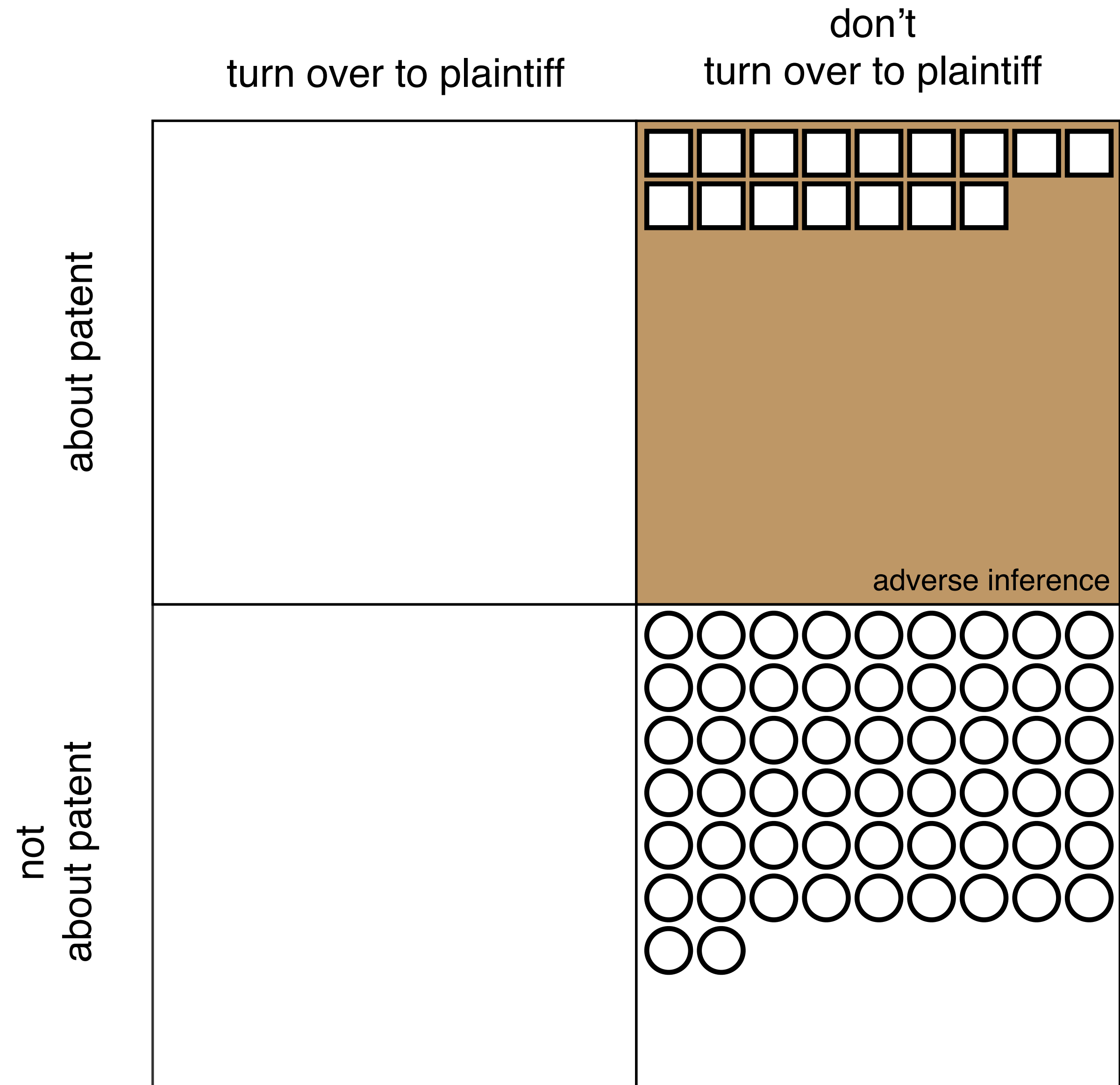
data-driven e-discovery

daegis



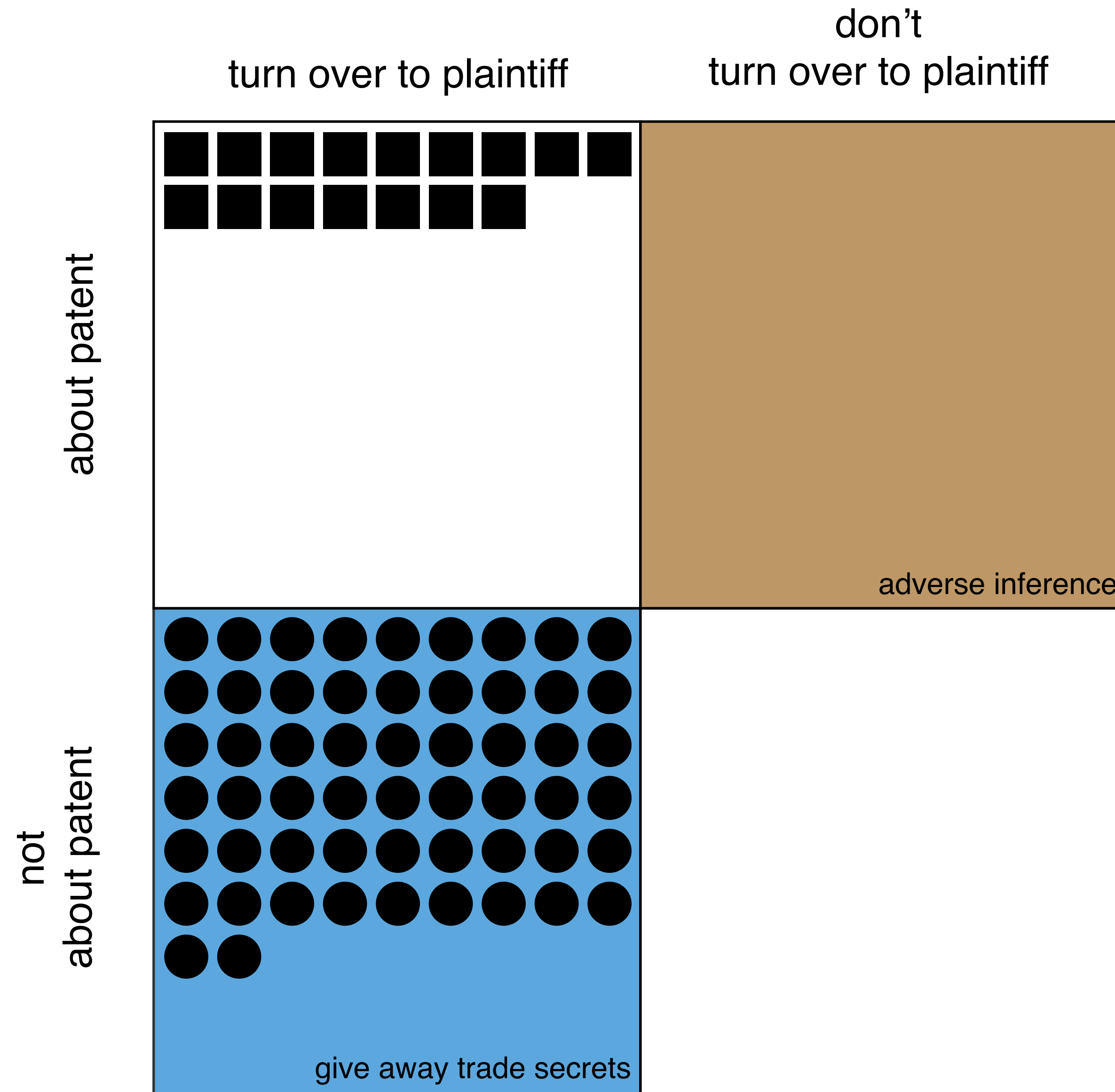
data-driven e-discovery

daegis



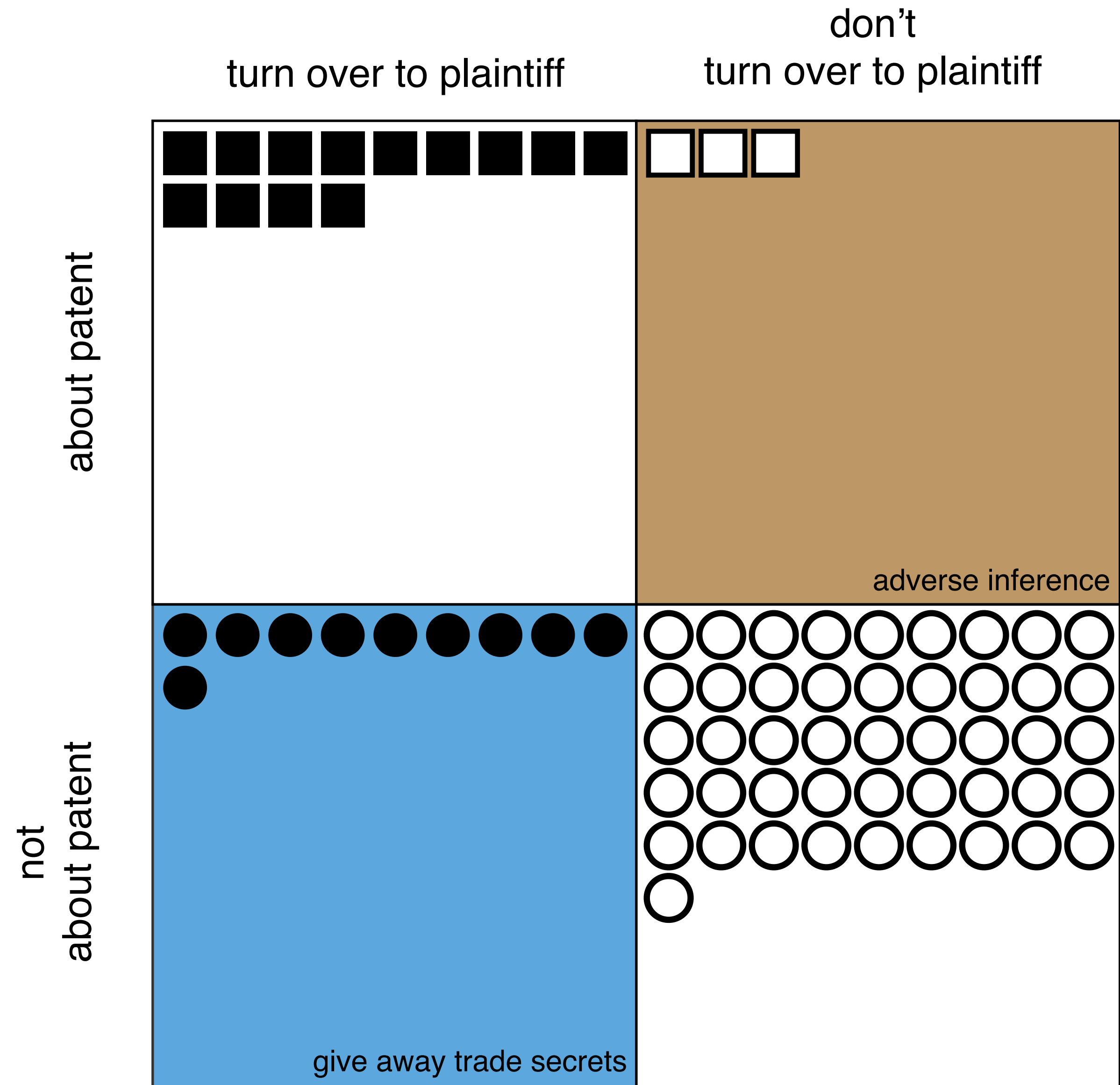
data-driven e-discovery

daegis



data-driven e-discovery

daegis

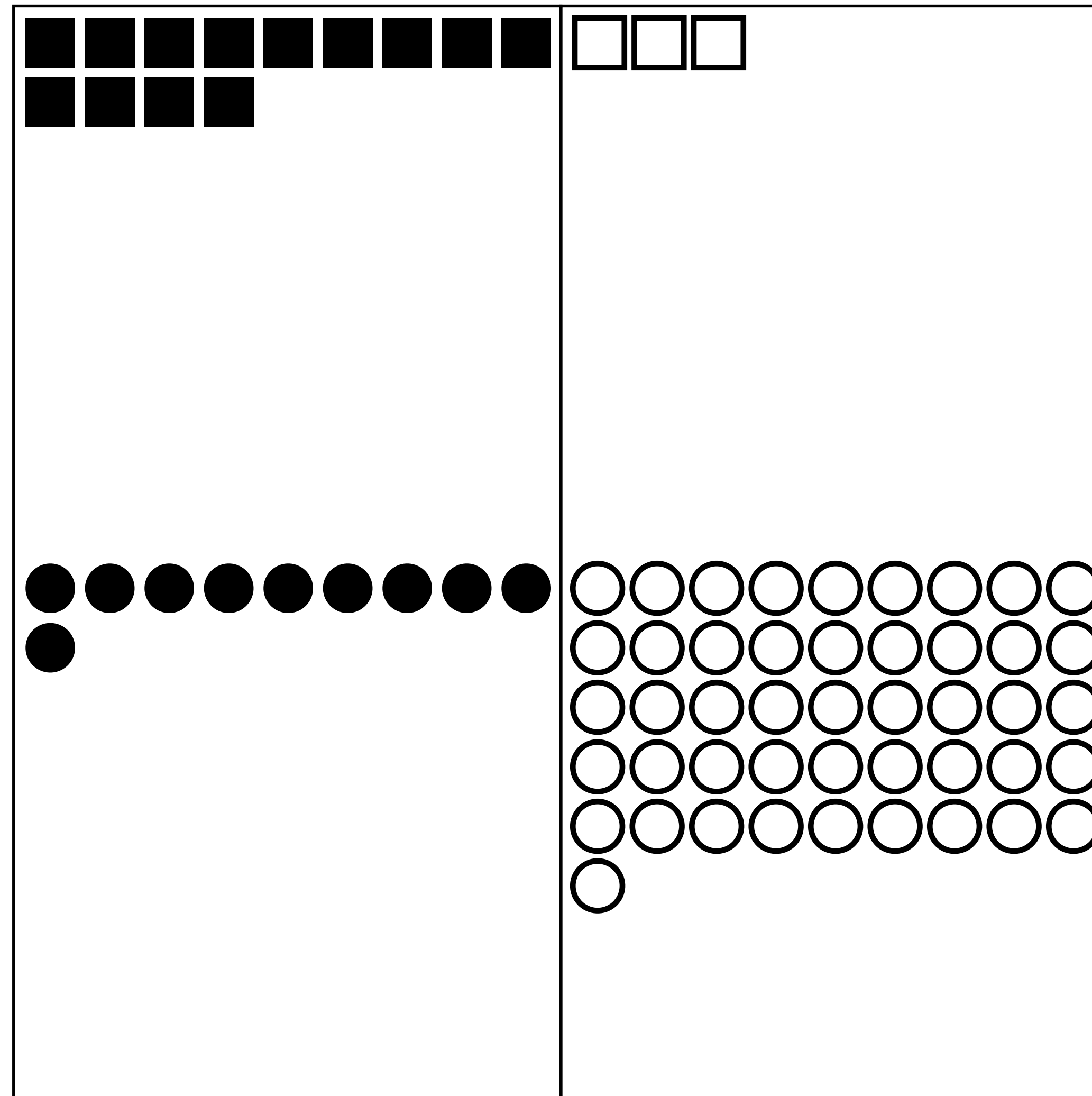


data-driven e-discovery

daegis

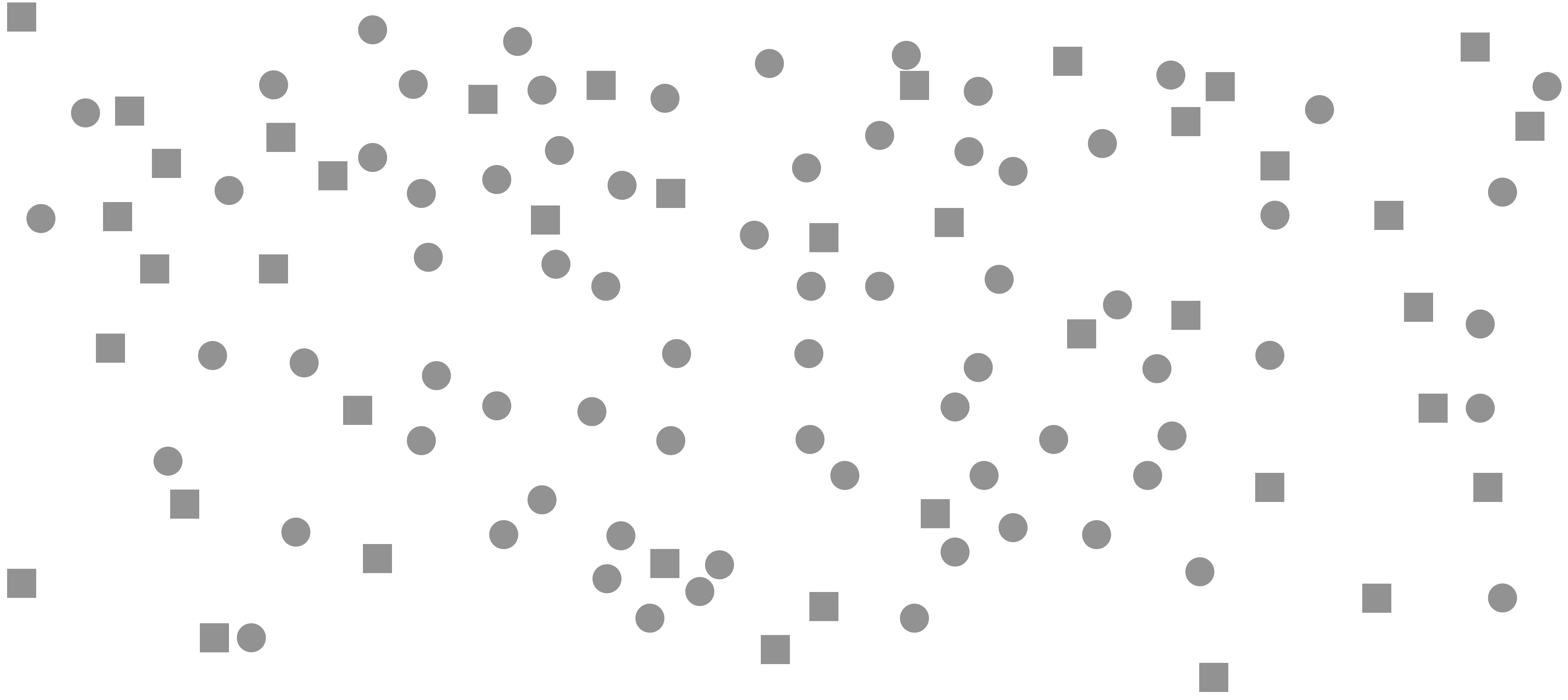
turn over to plaintiff

don't
turn over to plaintiff



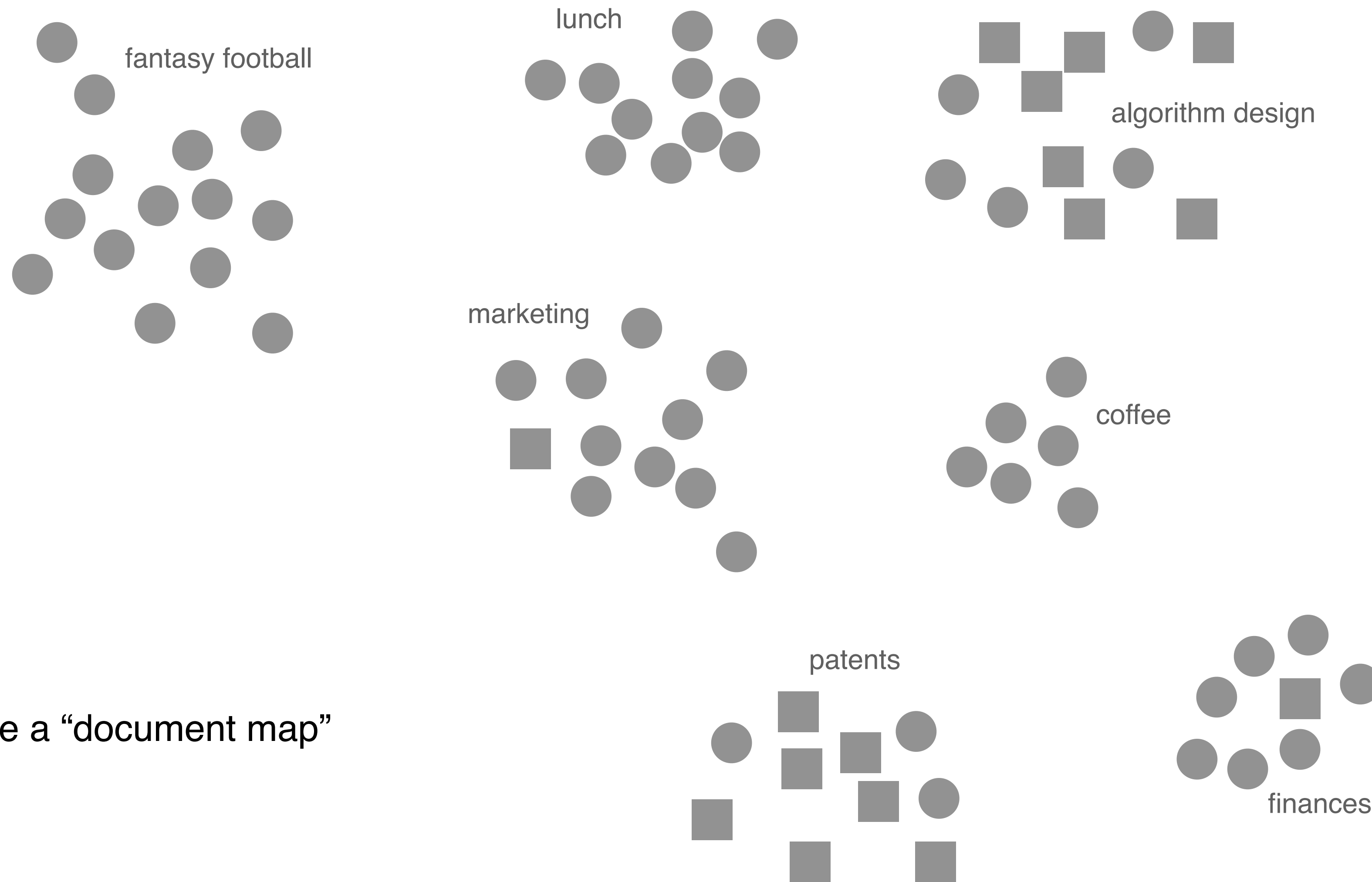
data-driven e-discovery

daegis



data-driven e-discovery

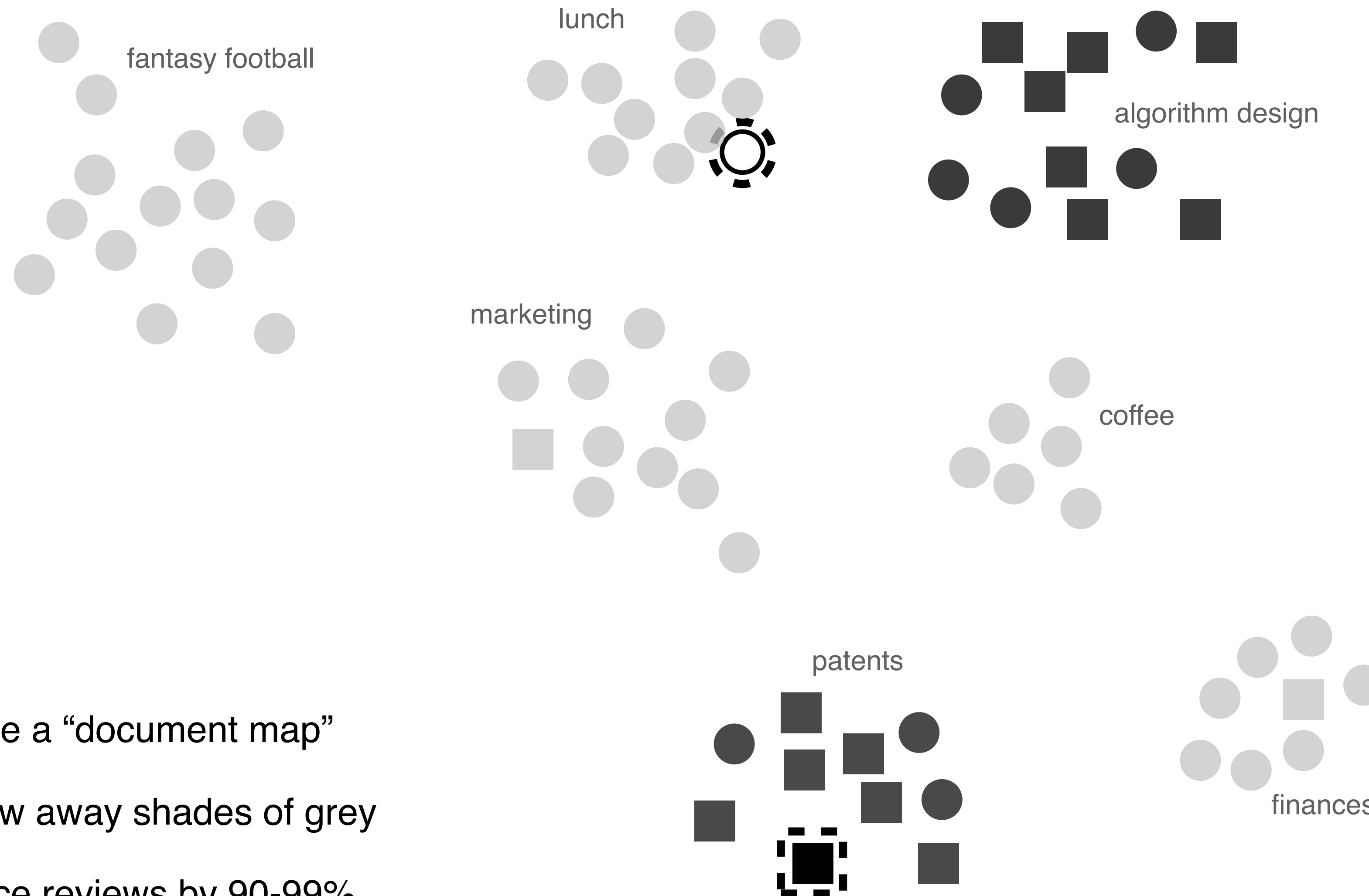
daegis



create a “document map”

data-driven e-discovery

daegis



create a “document map”

review away shades of grey

reduce reviews by 90-99%

CURRENT ACCURACY

75%
Recall [?] 83% of goal achieved (90%)

84%
Precision [?] 120% of goal achieved (70%)

79%
F1 [?] 101% of goal achieved (79%)

[edit goal](#)

[confusion matrix](#)

COST CALCULATOR

1,002,312
total documents

51,580
eyes-on reviews so far

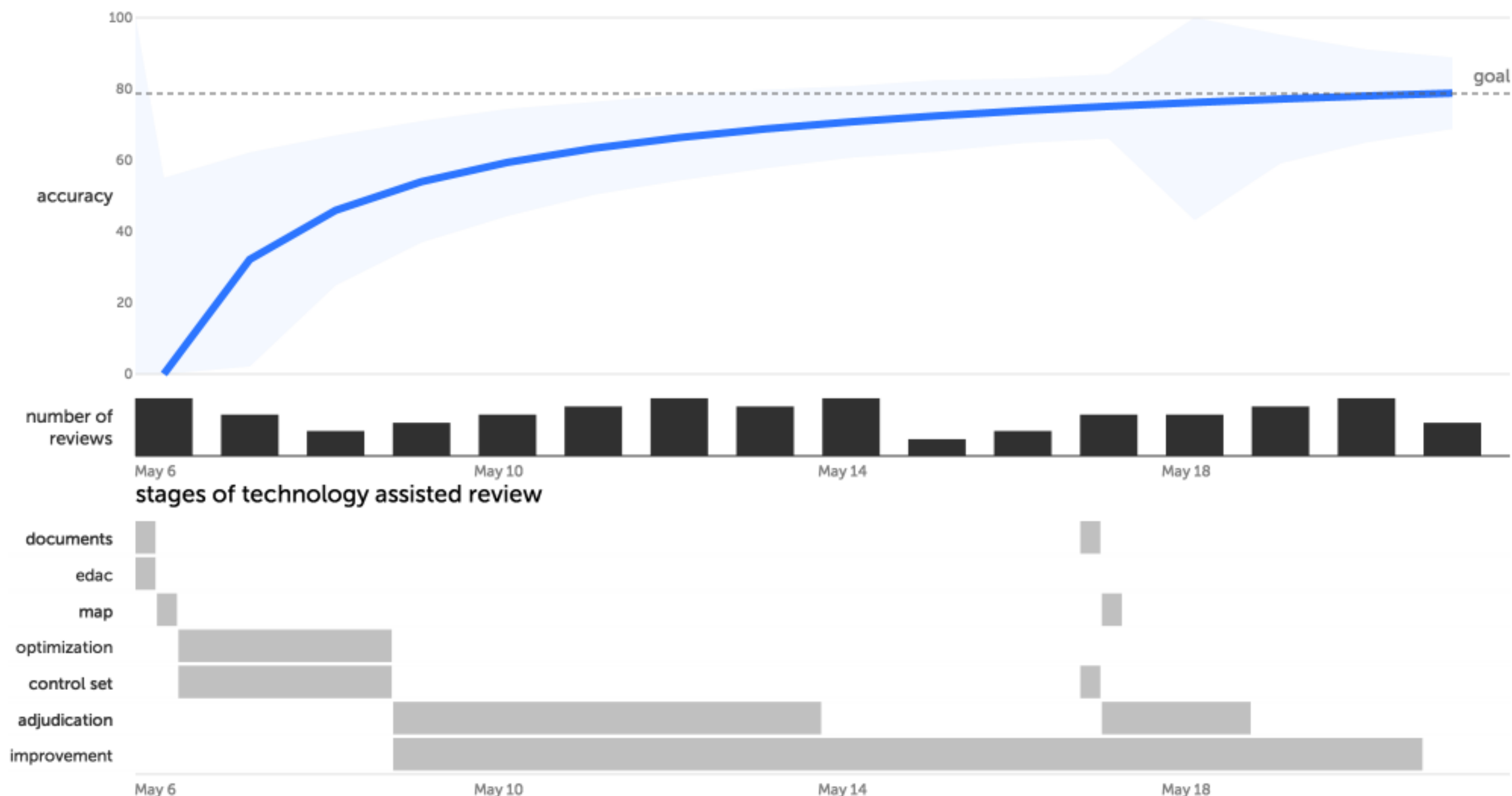
\$237,633
savings over linear review

\$23,345 cost of last 1% gain in F1

200 additional eyes-on reviews [?]

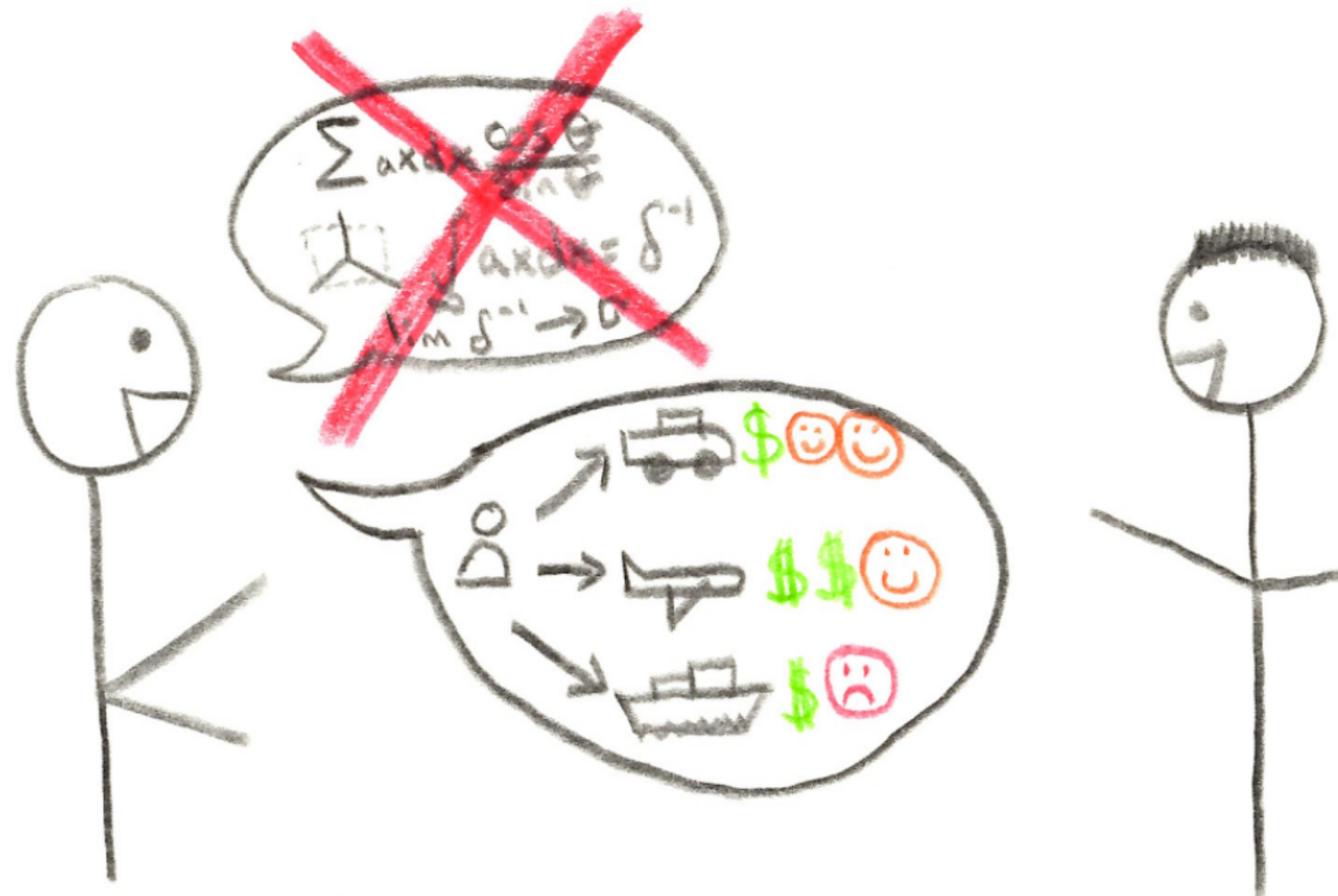
\$1.00 per eyes-on review [?]

\$0.75 per unreviewed document [?]



care about design.

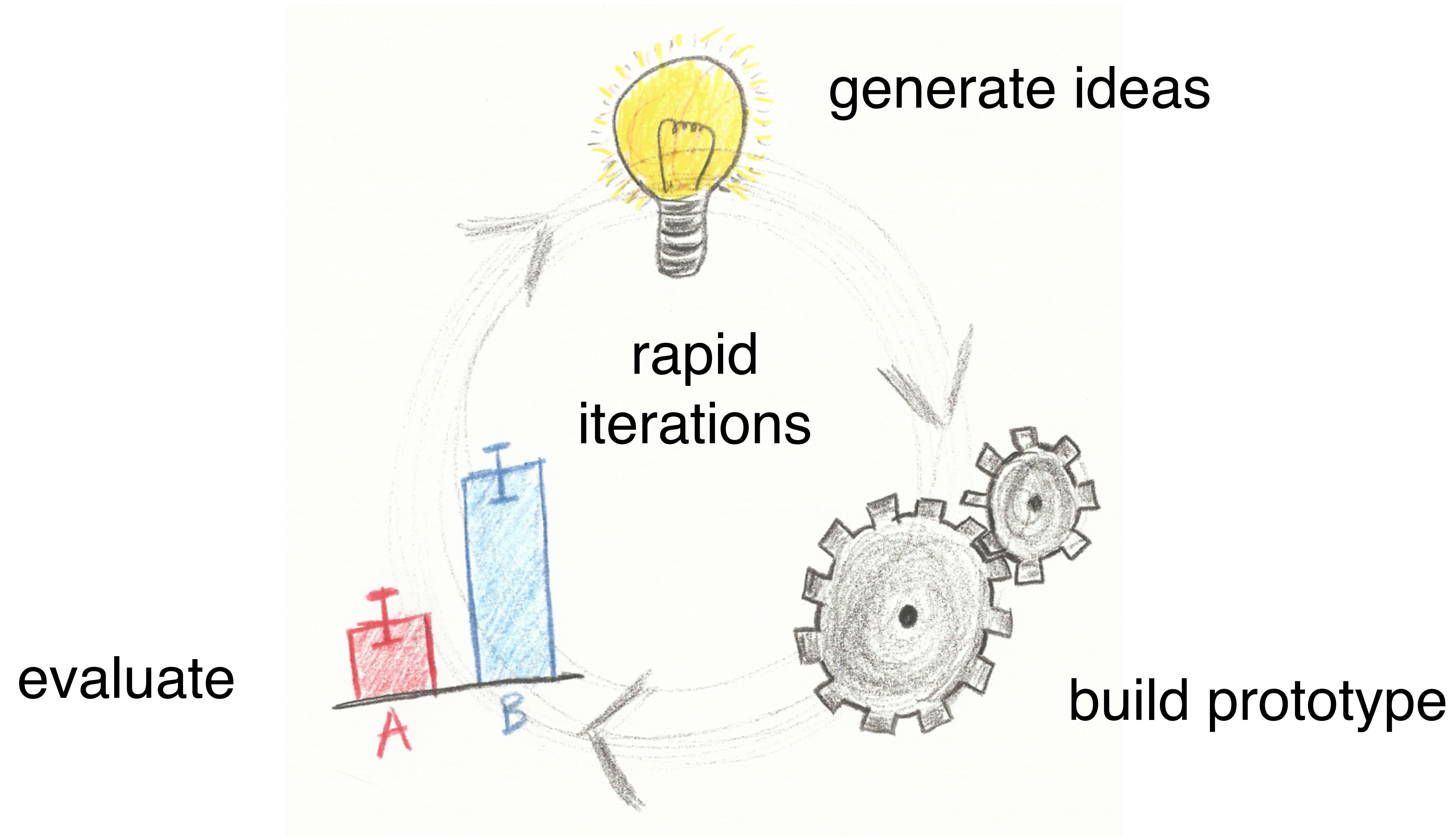
simple, powerful interfaces relay analytics better.

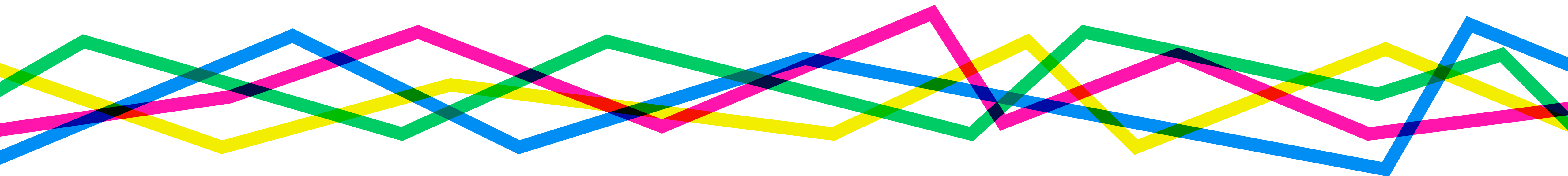




iterative problem solving

plan, build, test, and iterate as quickly as possible





Data driven expertise exploration

Procter & Gamble

data-driven expertise exploration

procter & gamble



data-driven expertise exploration

procter & gamble



High level goals:

- reveal areas of expertise
- evaluate connectivity within experts

data-driven expertise exploration

procter & gamble



data-driven expertise exploration

procter & gamble



Lorem Ipsum is a narrative about blankets.

Author: **Charlie Brown**

Date: **31 Jan 2012**

Lorem Ipsum is a dummy text used when typesetting or marking up documents. It has a long history starting from the 1500s and is still used in digital millennium for typesetting electronic documents, page designs, etc.

In itself, the original text of Lorem Ipsum might have been taken from an ancient Latin book that was written about 50 BC. Nevertheless, Lorem Ipsum's words have been changed so they don't read as a proper text.

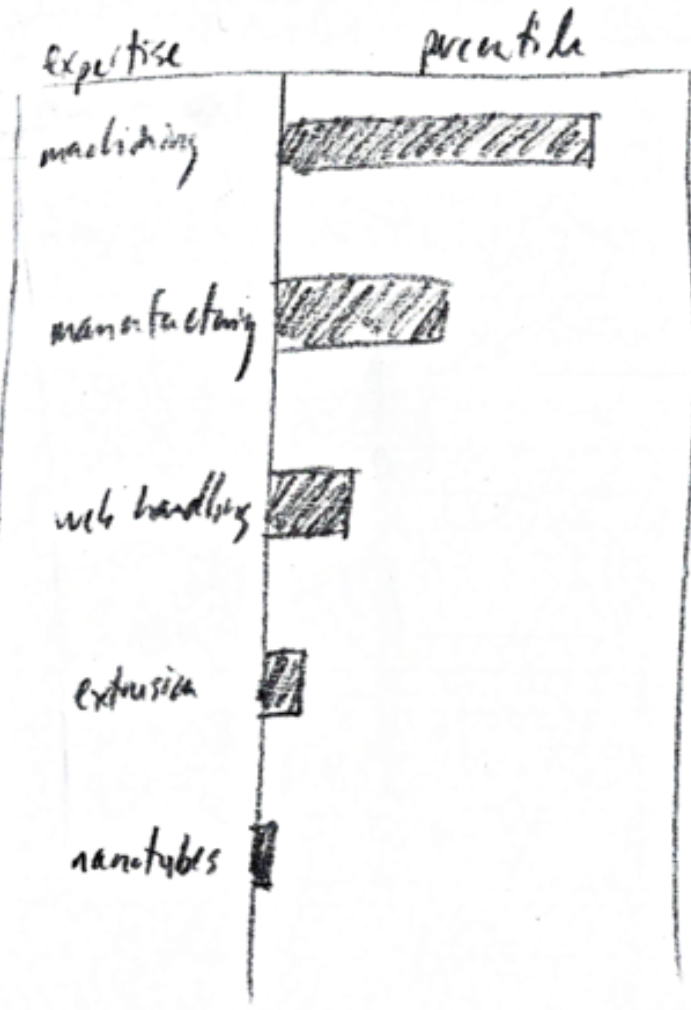
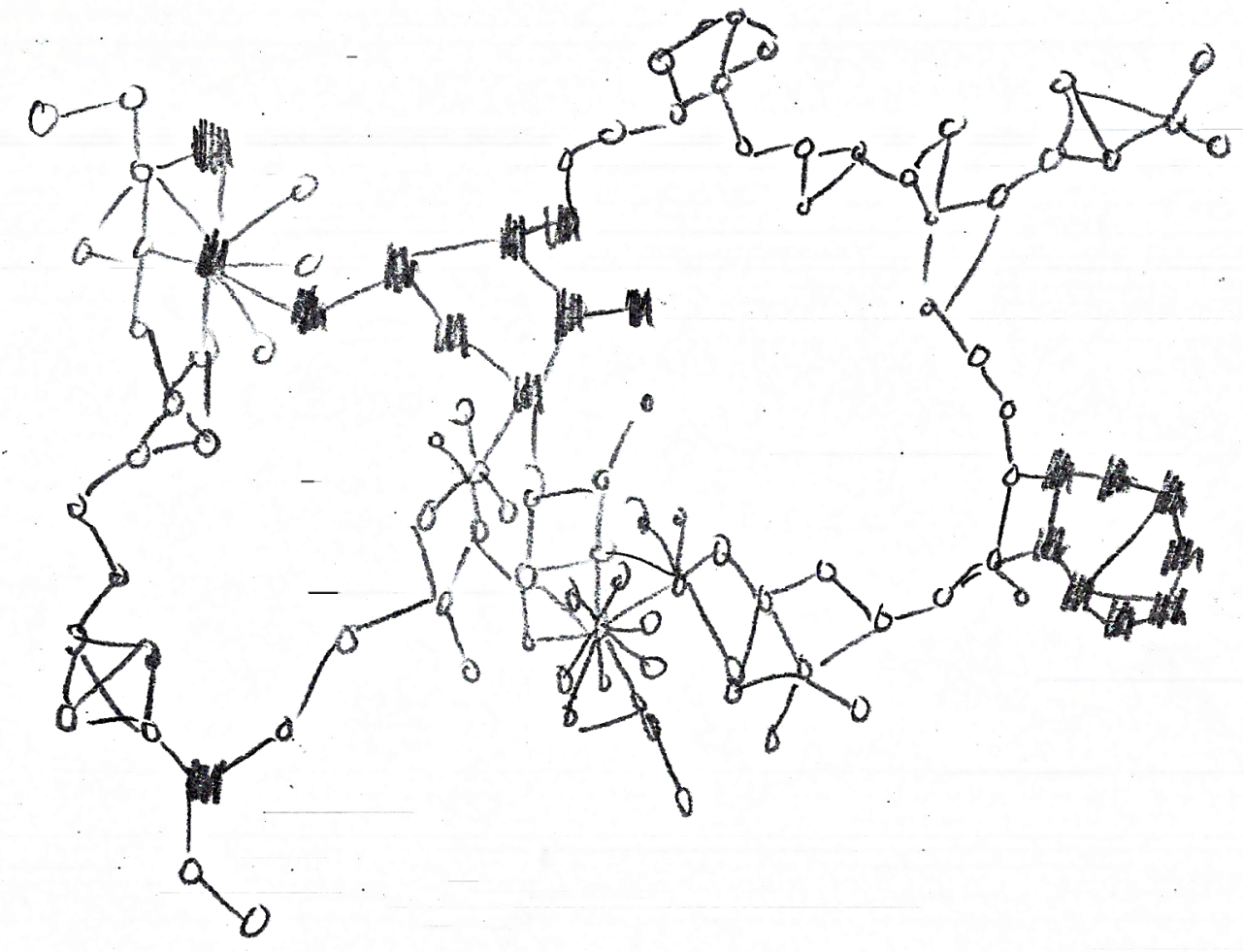
Naturally, page designs that are made for text documents must contain some text rather than placeholder dots or something else. However, should they contain proper English words and sentences almost every reader will deliberately try to interpret it eventually, missing the design itself.

However, a placeholder text must have a natural distribution of letters and punctuation or otherwise the markup will look strange and unnatural. That's what Lorem Ipsum helps to achieve.

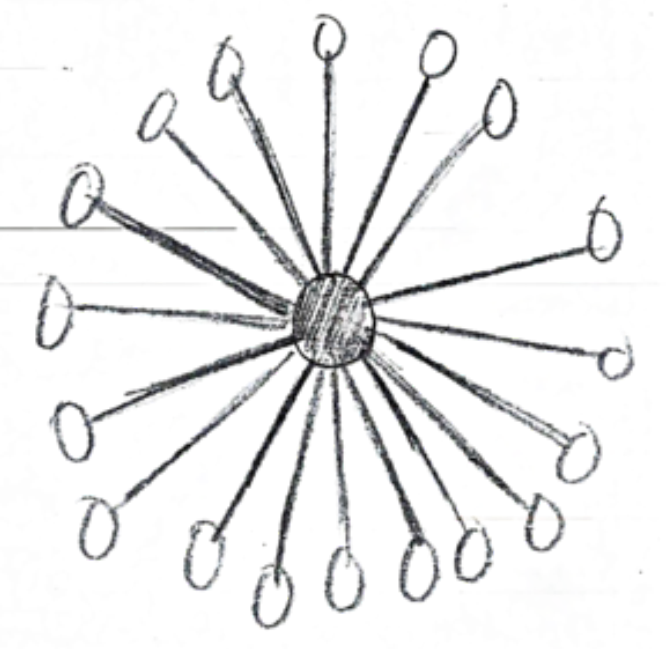
I would like to thank **Peppermint Patty** for her support on studying Lorem

Ipsum as well as the infinite wisdom of **Linus van Pelt** and his willingness to use his blanket in my experiments.

- Subject matter experts
- Extrusion
 - Nanotubes
 - Machining
 - Web handling
 - Manufacturing

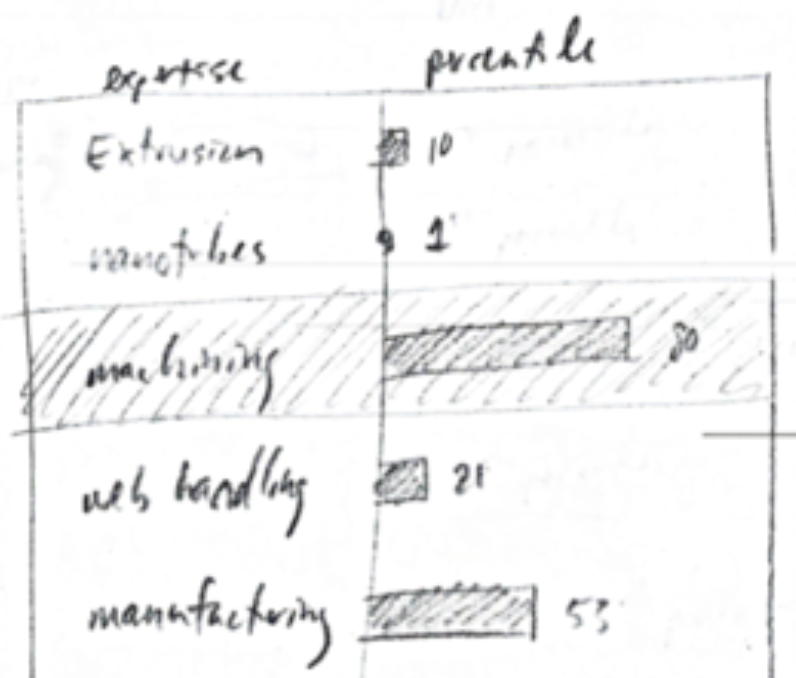


- connected by
- survey
 - coauthors
 - coexpertise
- Russell, Bill
 Wootny, James
 Lebron, James
 Jordan, Michael
 Lange, Brian
 Pippin, Suthia
 Grant, Horace
 Cartwright, Bill



Brown, Chuck
 TU7263
 Location: Cincinnati, OH USA
 Education: University of Maryland
 Band: 4


 GBU: Design
 Products: Tide



content search

Machining is as easy as 1-2-3!

Coffee: manufacturing of the future

on the extrusion machine


21st century machines:

Asker: **Jim Jamson**

Anyone done work on using nonwoven fabrics in packaging?
 #Transformations


Find

Top recommended:

 **Hob Jobson**

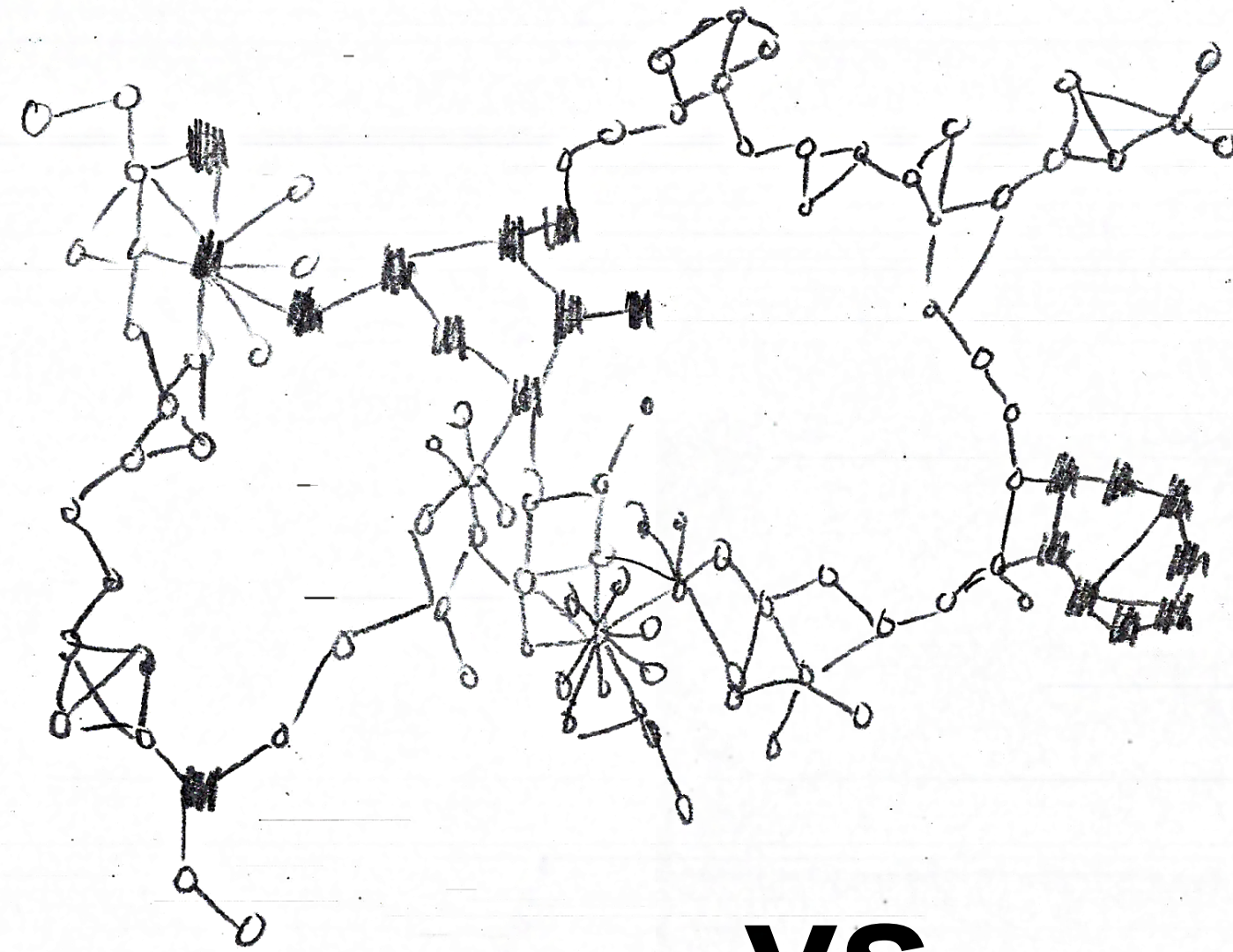
3 SLS reports on nonwoven fabrics

2 wiki answers on "non wovens"

 you've emailed

Subject matter experts

- Extrusion
- Nanotubes
- Machining
- Web handling
- Manufacturing

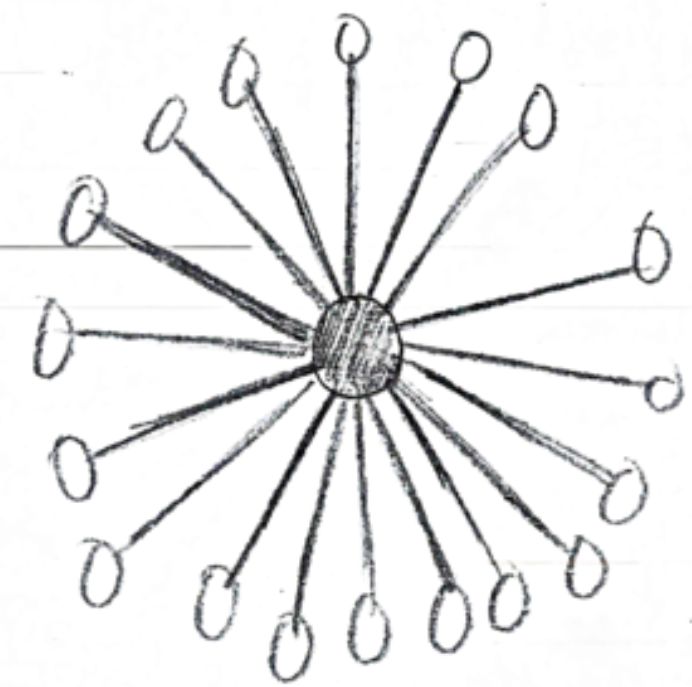


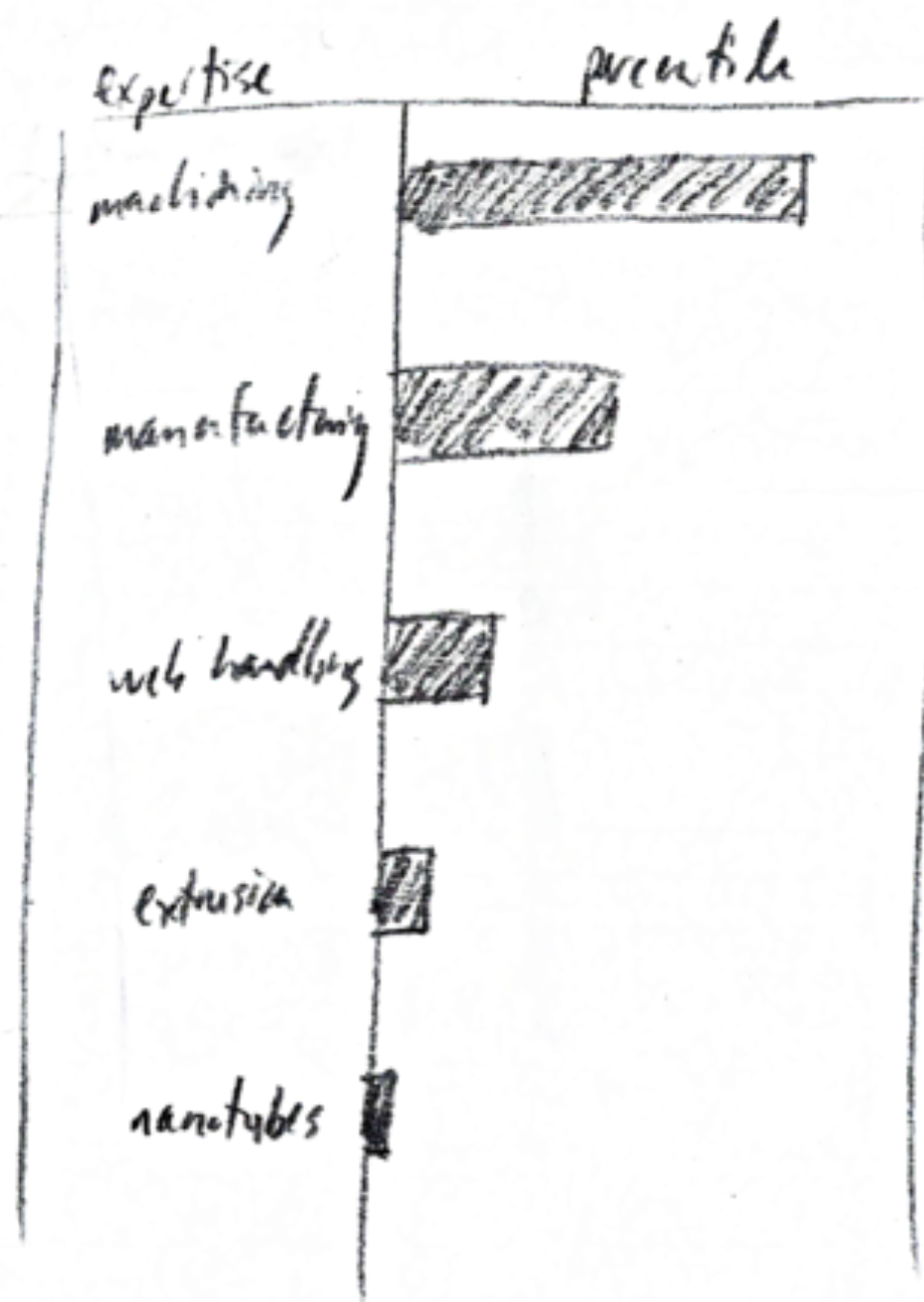
VS.

expertise	percentile
machining	<div style="width: 80%; background-color: #cccccc; border: 1px solid black;"></div>
manufacturing	<div style="width: 40%; background-color: #cccccc; border: 1px solid black;"></div>
web handling	<div style="width: 20%; background-color: #cccccc; border: 1px solid black;"></div>
extrusion	<div style="width: 10%; background-color: #cccccc; border: 1px solid black;"></div>
nanotubes	<div style="width: 5%; background-color: #cccccc; border: 1px solid black;"></div>

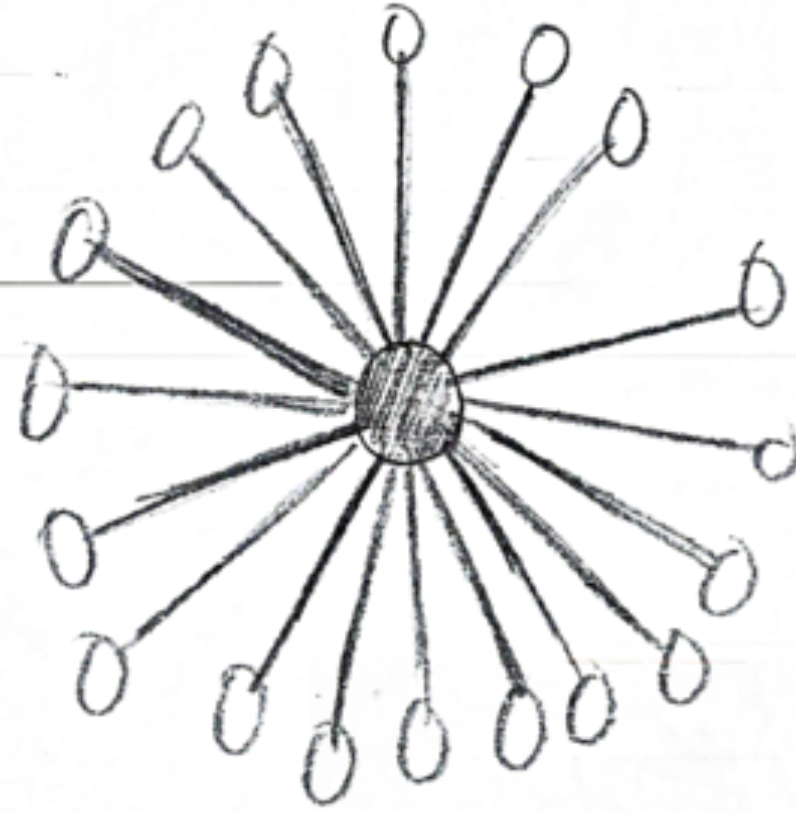
connected by ▼

- survey
- coauthors
- coexpertise
- Russell, Bill
- Wootny, James
- Lebron, James
- Jordan, Michael
- Lange, Brian
- Pippen, Scottie
- Grant, Horace
- Cartwright, Bill

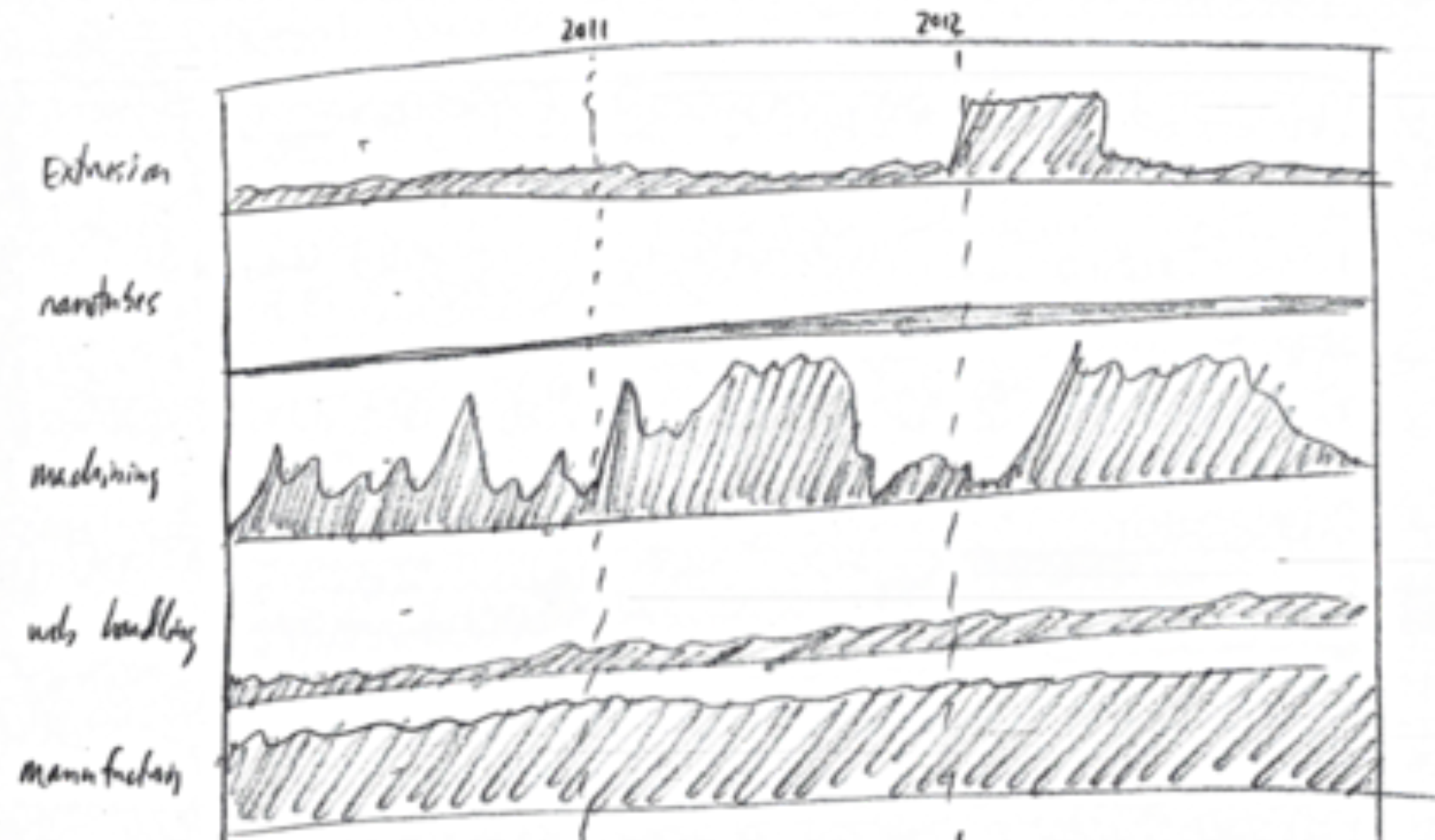




- connected by
- survey
 - coauthors
 - coexpertise
- Russell, Bill
 Worthy, James
 LeBron, James
 Jordan, Michael
 Lange, Brian
 Pippen, Scottie
 Grant, Horace
 Cartwright, Bill

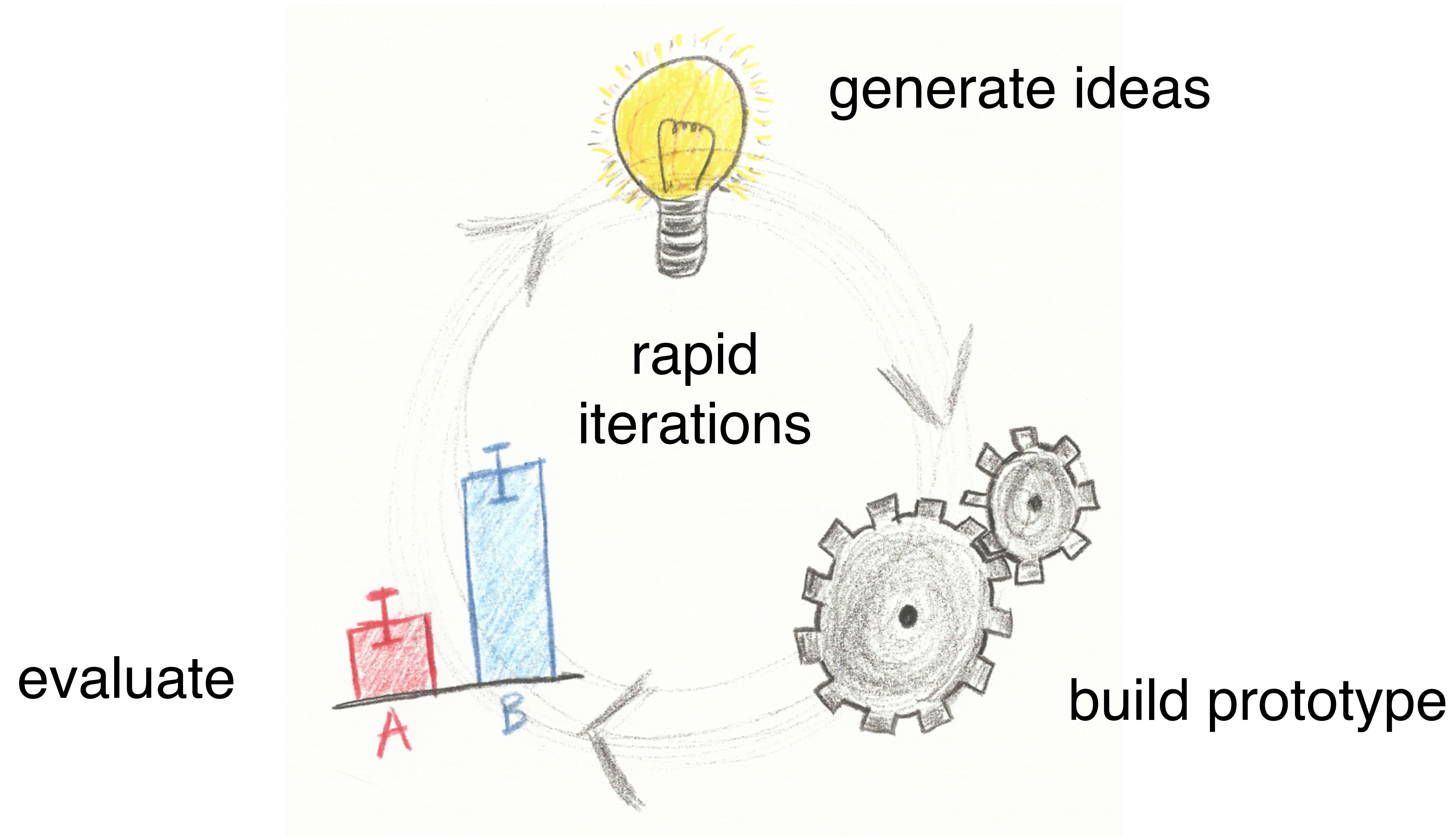


VS.



iterative problem solving

plan, build, test, and iterate as quickly as possible







Asker: Chip Dipson

Anyone have experience machining aluminum at micron-level tolerances?

Search

e-mail selected

		expertise index▼	activity index
<input type="checkbox"/>	 Chuck Brown	<div style="width: 80%; background-color: #ADD8E6; border: 1px solid black;"></div>	<div style="width: 40%; background-color: #FF8C00; border: 1px solid black;"></div>
<input type="checkbox"/>	 Sue Bruston	<div style="width: 70%; background-color: #ADD8E6; border: 1px solid black;"></div>	<div style="width: 60%; background-color: #FF8C00; border: 1px solid black;"></div>
<input type="checkbox"/>	 Jed Bedson	<div style="width: 75%; background-color: #ADD8E6; border: 1px solid black;"></div>	<div style="width: 50%; background-color: #FF8C00; border: 1px solid black;"></div>
<input type="checkbox"/>	 Tim Thomson	<div style="width: 70%; background-color: #ADD8E6; border: 1px solid black;"></div>	<div style="width: 80%; background-color: #FF8C00; border: 1px solid black;"></div>






load more...



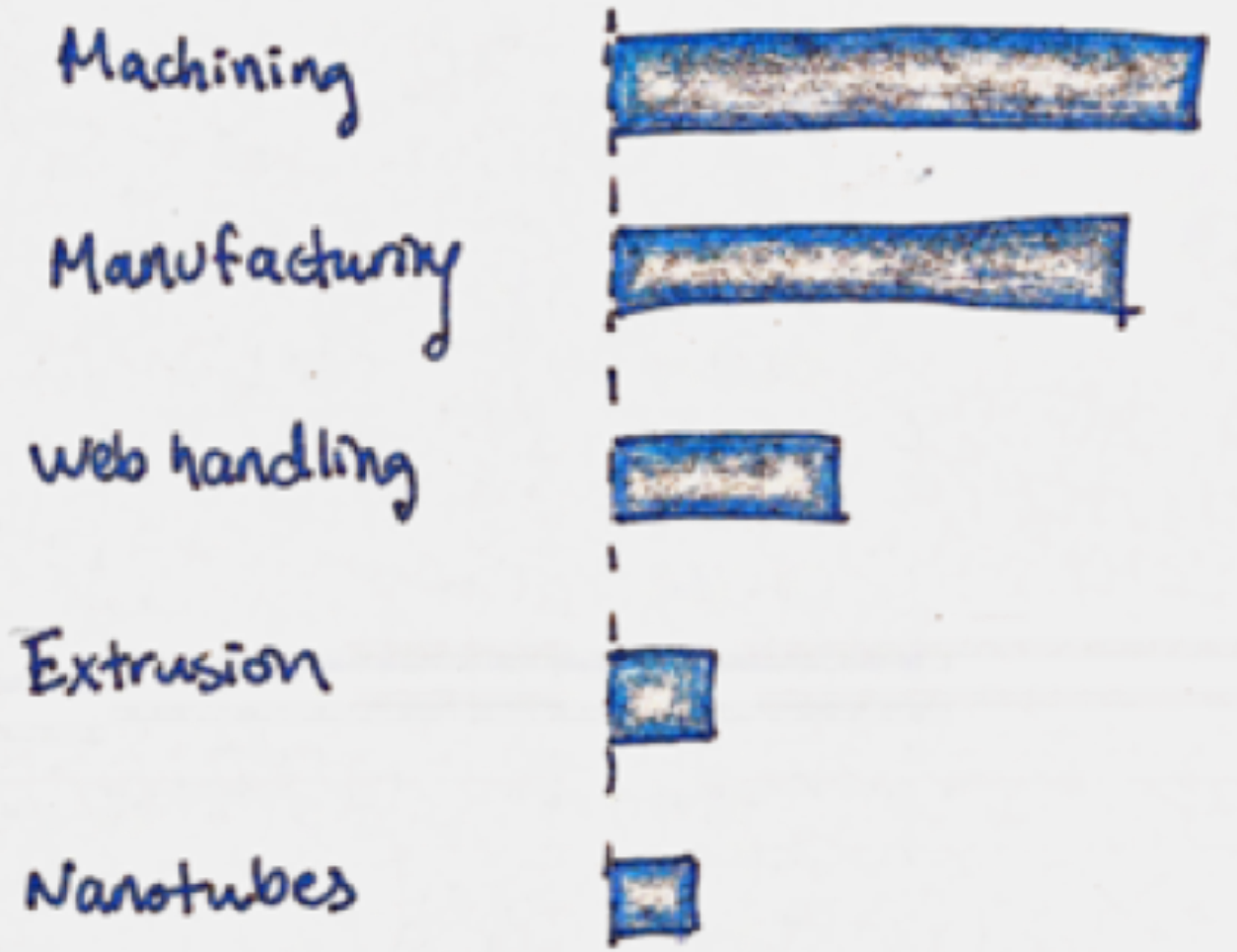
Brown, Chuck
Cambridge, MA, USA
Chemical Engineer

 email
4 emails sent recently

3 mutual connections with: question asked

 Snoopy	 Ling	 Lucy	 Patty	 Pig Pen
--	--	--	---	---

Area of Expertise



Activity







- Machining ▼
- Manufacturing ▶ 3 reports
- 5 Wiki answers
- 1 conference/presentation

asker:  **Chip Dipson**


Anyone have experience machining aluminum at micron-level tolerances?

[Redo search](#)

[Email **Chip** w/ 3 selections](#)

		▼ expertise index	activity level	degrees separated
<input checked="" type="checkbox"/>	 Chuck Brown	★★★★★	● ● ●	 2
<input checked="" type="checkbox"/>	 Sue Bruston	★★★★★	●	 3
<input type="checkbox"/>	 Jed Bedson	★★★★☆	● ●	 1
<input checked="" type="checkbox"/>	 Tim Thomson	★★★☆☆	● ● ●	 2
<input type="checkbox"/>	 Bing Kingston	★★★☆☆	●	 3
<input type="checkbox"/>	 Daniel Manson	★★☆☆☆	● ●	 1


[load more...](#)



Chuck Brown
 Research Fellow
 Cambridge, MA, USA
 Hill Technical Center

17 recommendations ✓

2 mutual connections with Chip



Expertise

Machining	██
Manufacturing	████████████████████████████████████
Web Handling	██████████████████████████████
Nanotubes	██████████████████████████
Extrusion	██████████████████████

Additional Info

[Social Network Survey Page](#)

[Profile](#)
[Articles](#)

High level goals:

- reveal areas of expertise
- evaluate connectivity within experts

High level goals:

- reveal areas of expertise
- evaluate connectivity within experts

FUNC 1
FUNC 2
FUNC 3
FUNC 4

FUNCTION

FUNCTION DROP DOWN

FUNC 1

FUNC 2

FUNC 3

FUNC 4

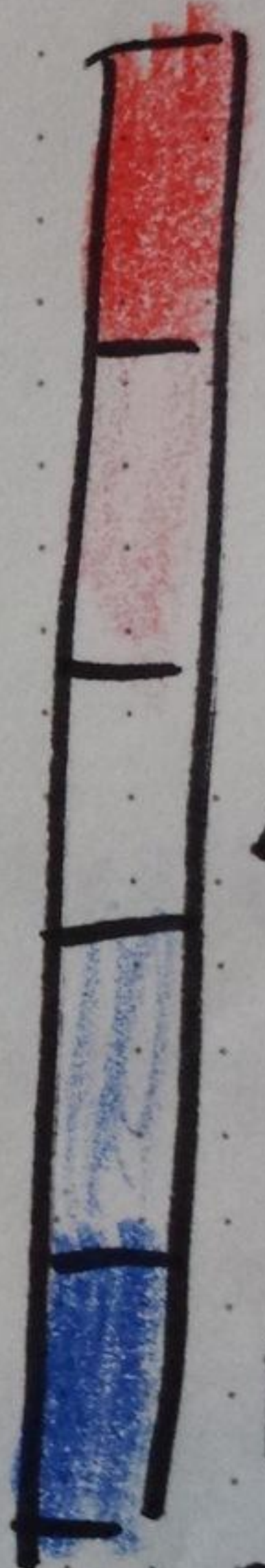
⋮

	FUNC 1			
	FUNC 2			
	FUNC 3			
	FUNC 4			
	⋮			
	⋮			

FUNCTION



↑ WIDTH BY PEOPLE



Above



Ave.

Below



let's compare countries.

~~~~~  
Author: Alice  
~~~~~  
~~~~~  
~~~~~  
Thanks Bill
~~~~~



 Alice   
~~~~~  
Country: Canada
~~~~~  
~~~~~



 Bill 
~~~~~  
Country: Germany  
~~~~~  
~~~~~



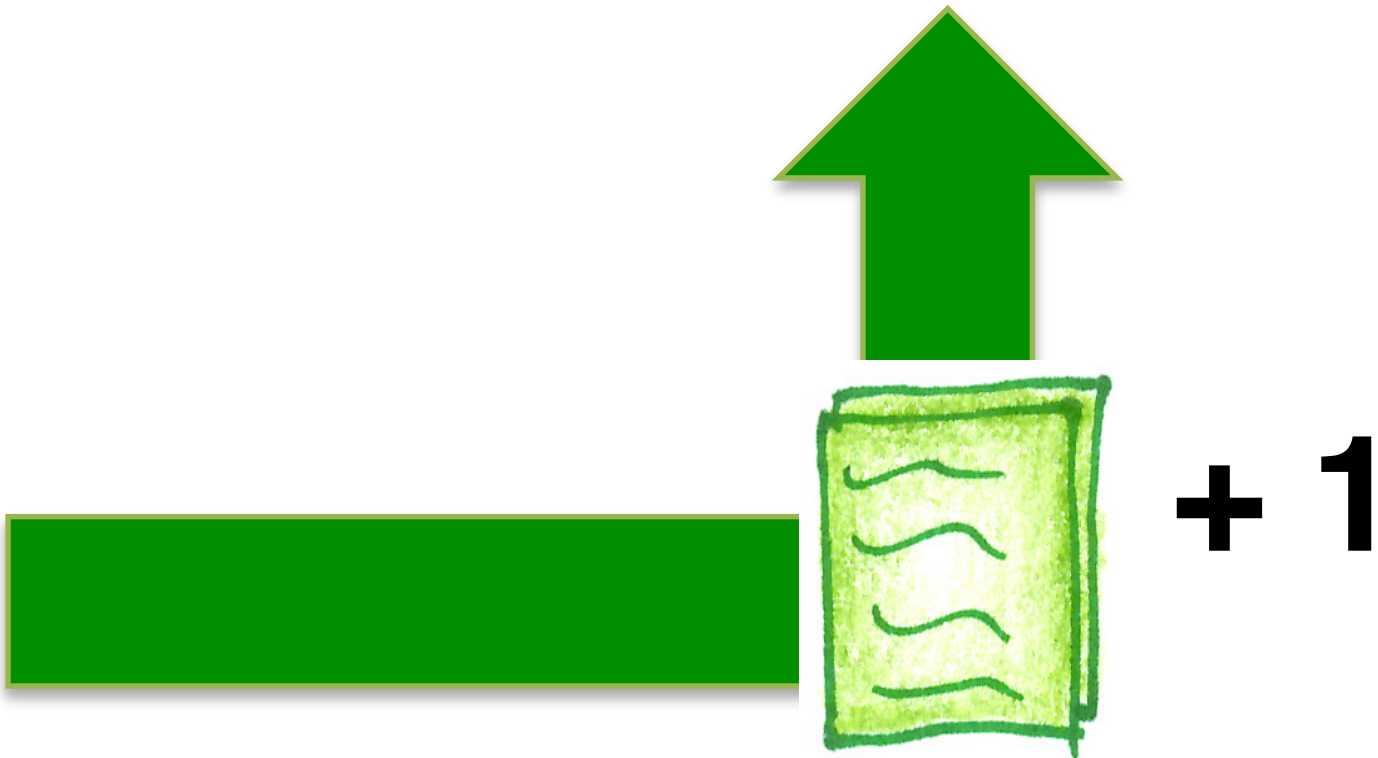
mention's country

Germany



author's country

Canada



mention's country

Brazil  
Canada  
Germany  
USA

author's country

Brazil  
Canada  
Germany  
USA

|  |  |  |  |
|--|--|--|--|
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |



mention's country

Brazil  
Canada  
Germany  
USA

author's country

Brazil  
Canada  
Germany  
USA

|    |    |    |    |
|----|----|----|----|
| 10 | 5  | 5  | 20 |
| 8  | 25 | 2  | 5  |
| 12 | 3  | 30 | 10 |
| 1  | 20 | 25 | 50 |

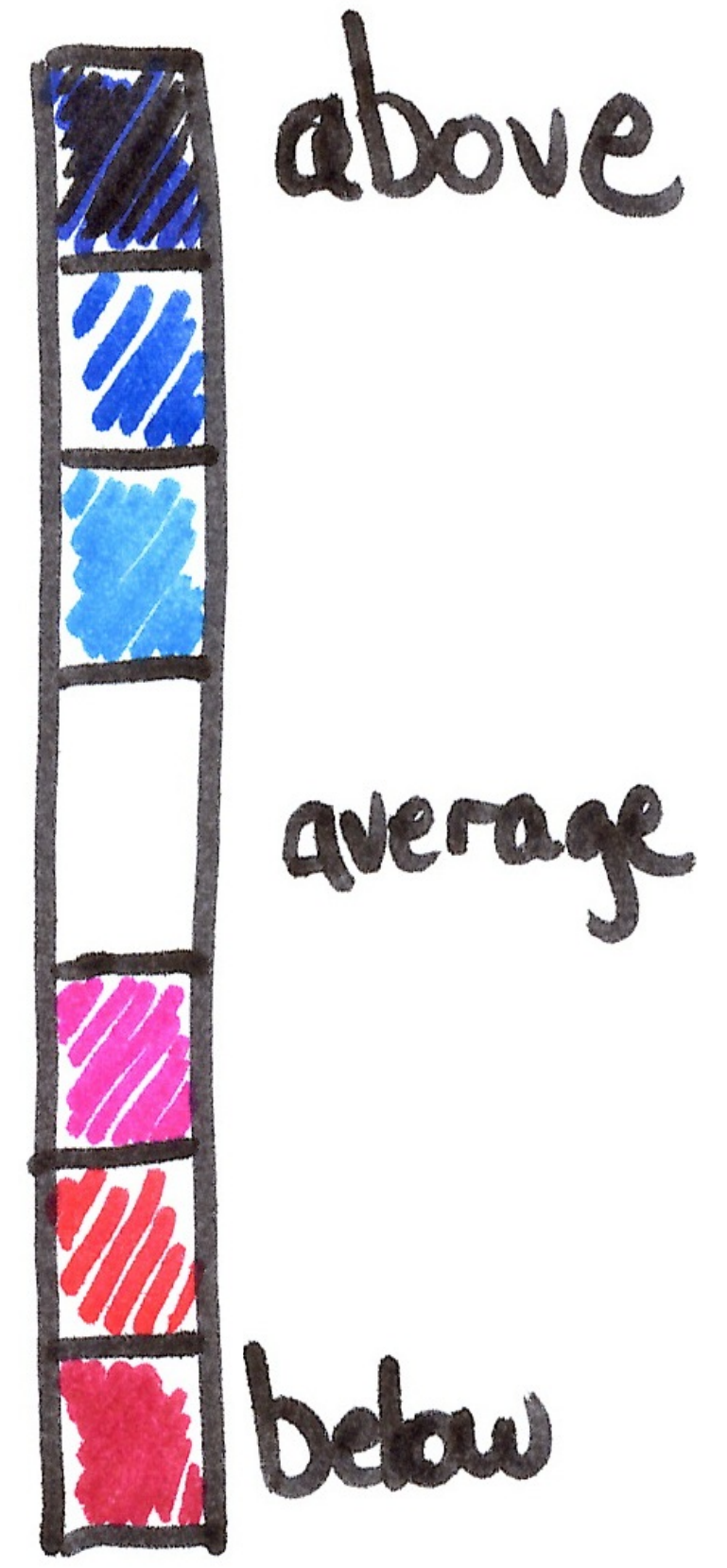
mention's country

Brazil  
Canada  
Germany  
USA

author's country

Brazil  
Canada  
Germany  
USA

|    |    |    |    |
|----|----|----|----|
| 10 | 5  | 5  | 20 |
| 8  | 25 | 2  | 5  |
| 12 | 3  | 30 | 10 |
| 1  | 20 | 25 | 50 |



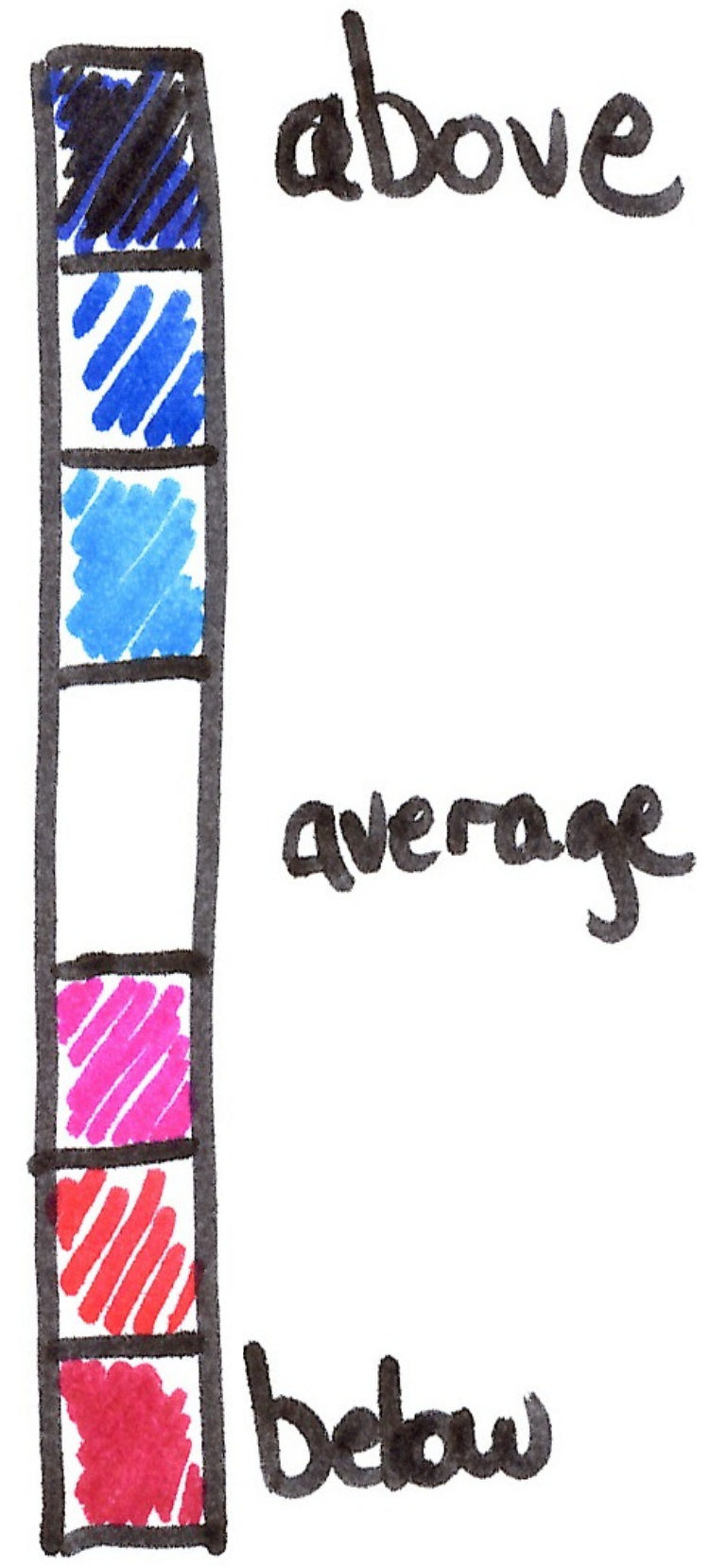
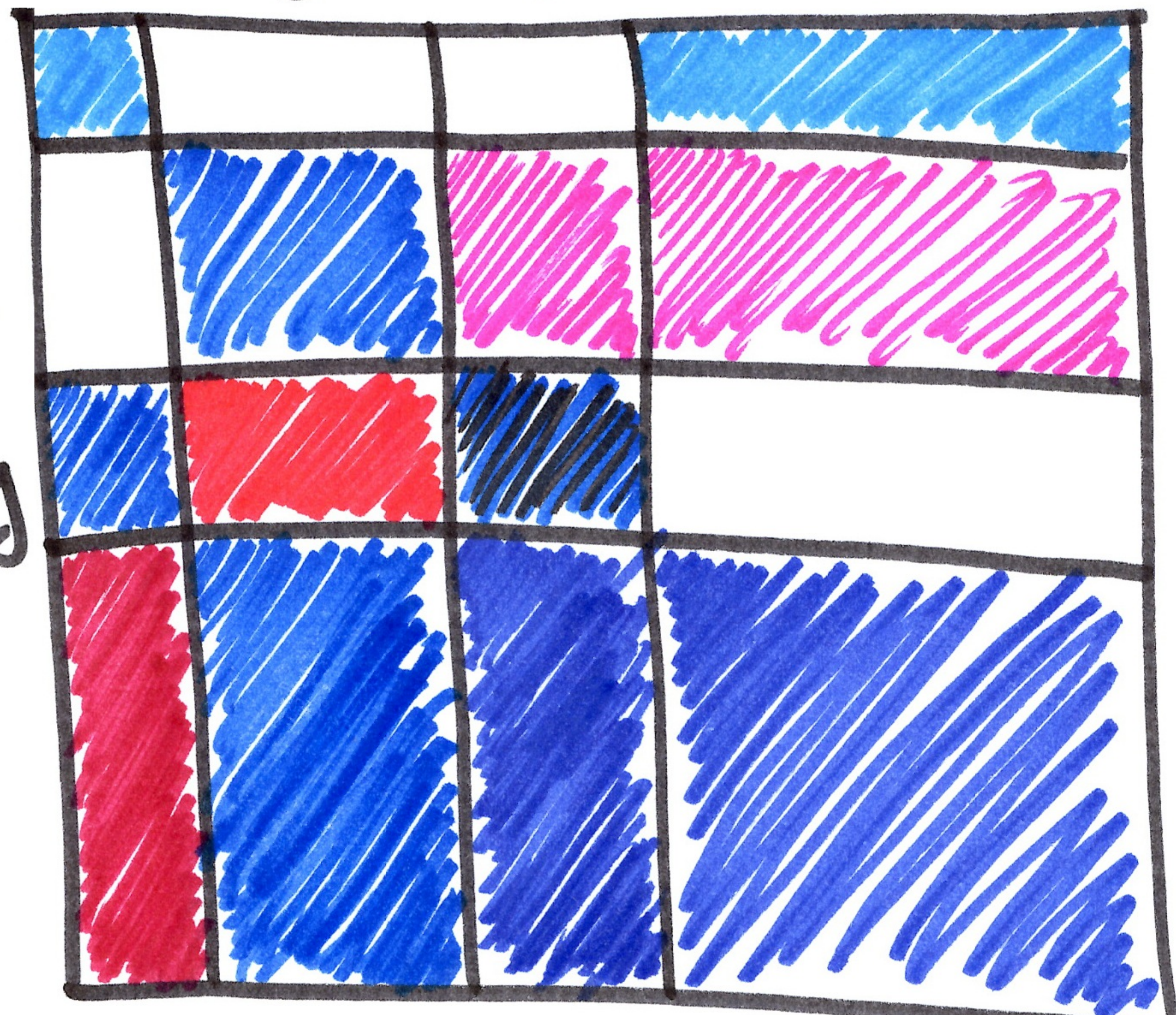


mention's country

Brazil  
Canada  
Germany  
USA

author's country

Brazil  
Canada  
Germany  
USA

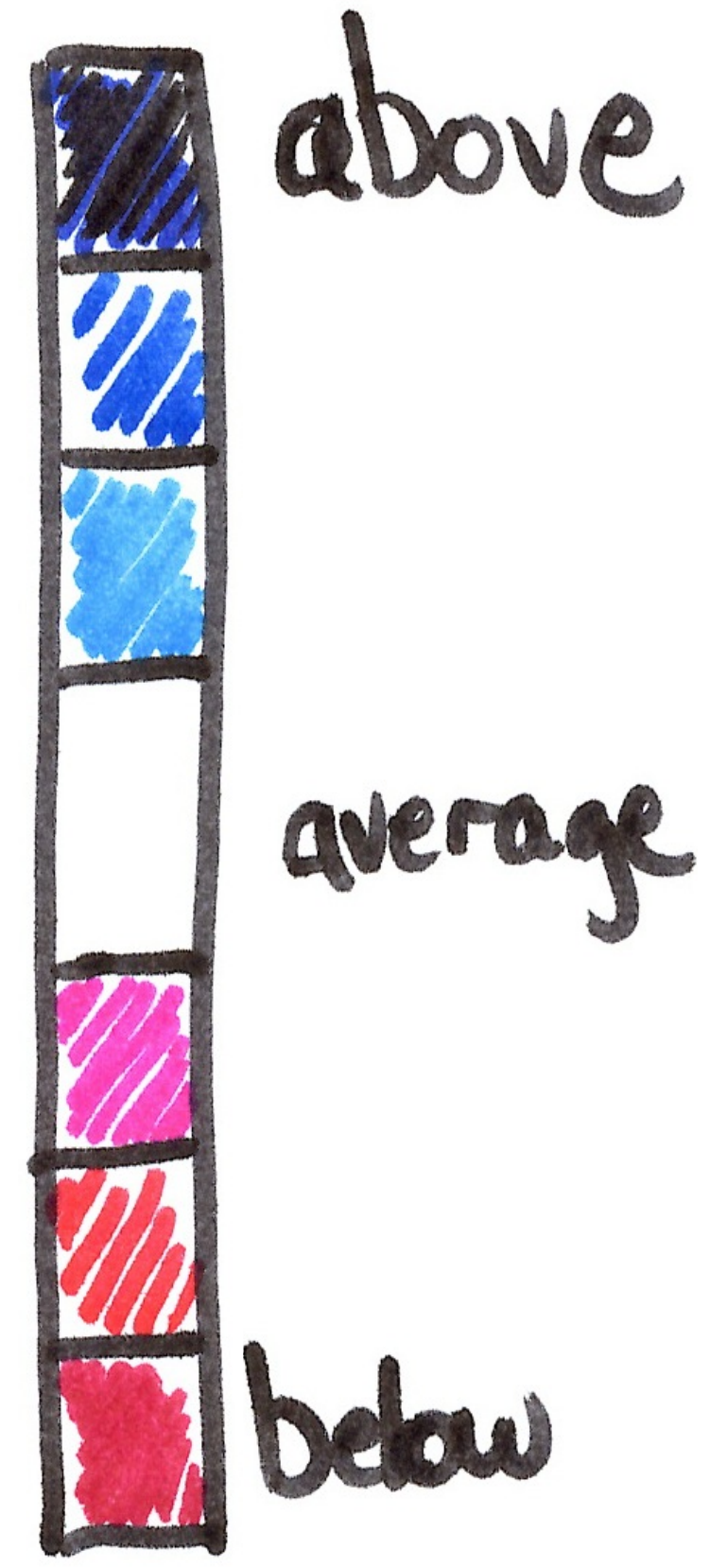
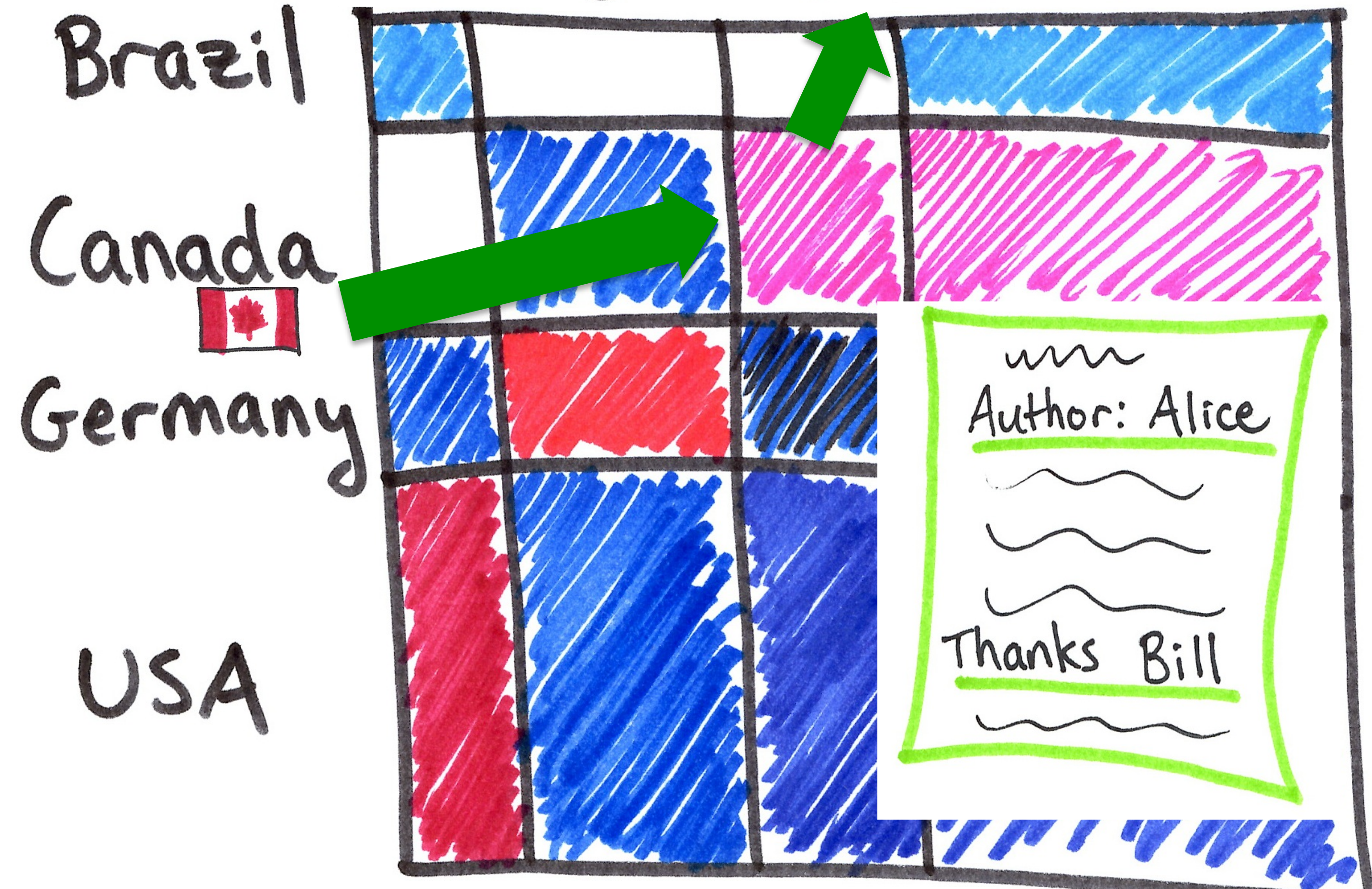




mention's country

Brazil  
Canada  
Germany   
USA

author's country

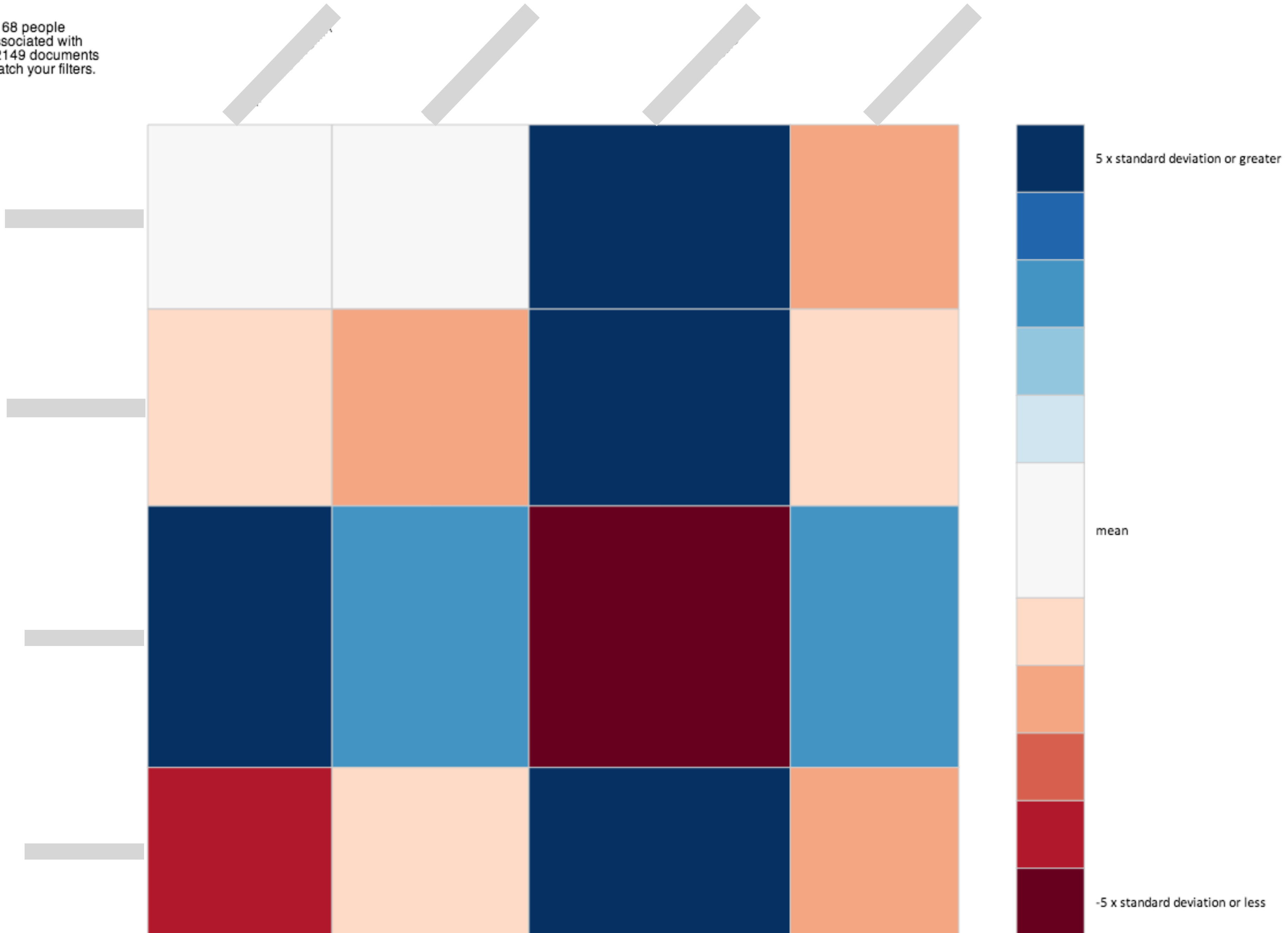


Author: Alice

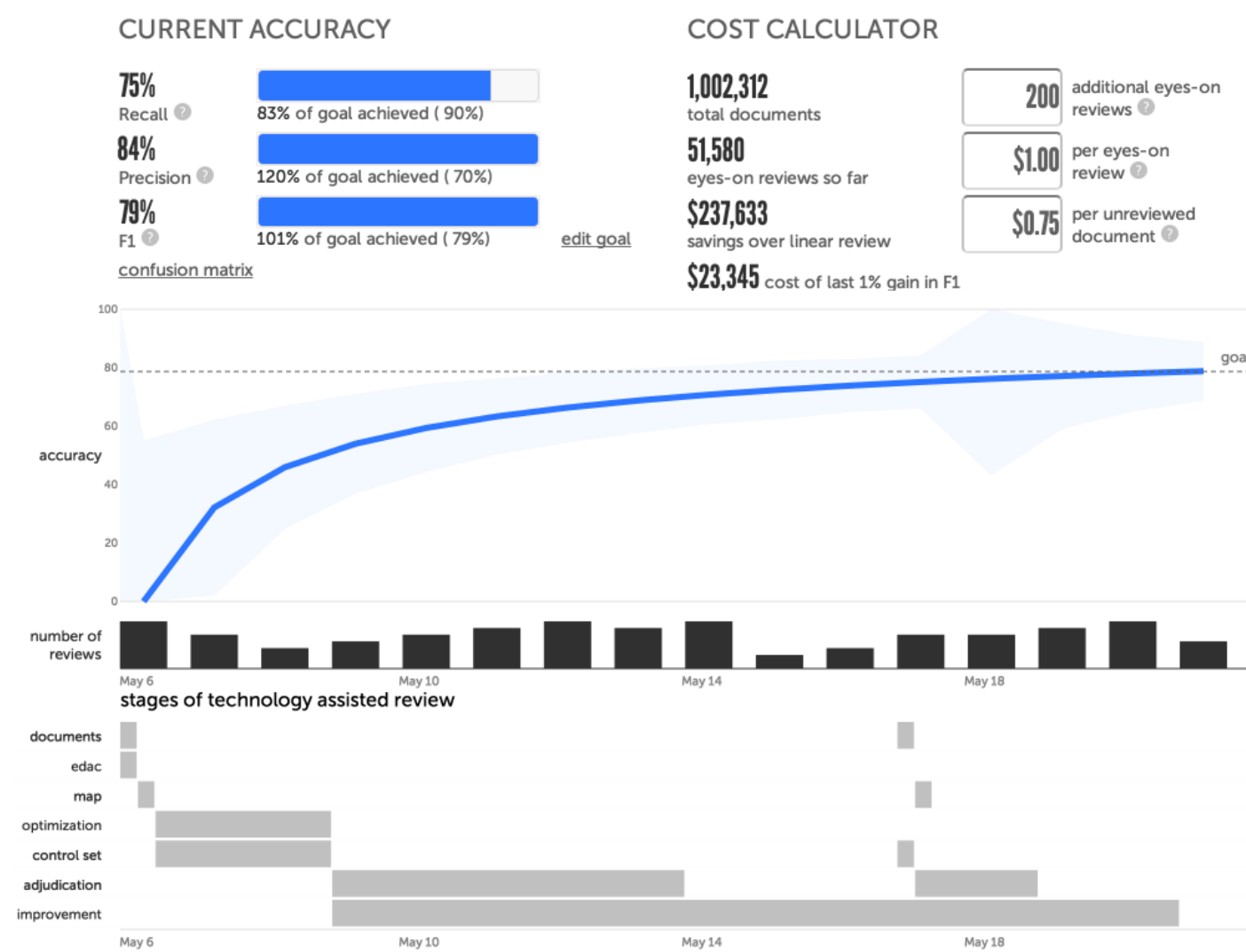
Thanks Bill



7168 people associated with 62149 documents match your filters.



# design influences data science.



*expert finder*

asker: **Chip Dipson**

Anyone have experience machining aluminum at micron-level tolerances?

[Redo search](#) [Email Chip w/ 3 selections](#)

▼ expertise index activity level degrees separated

- Chuck Brown** ★★★★★ ●●● ● **2**
- Sue Bruston** ★★★★★ ● ●●● **3**
- Jed Bedson** ★★★★★ ●●●●● ●●● **1**
- Tim Thomson** ★★☆☆● ●●●●● ●●● **2**
- Bing Kingston** ★★☆☆● ●●●●● ●●●● **3**
- Daniel Manson** ★★☆☆● ●●●●● ●●●● **1**

[load more...](#)

**Chuck Brown**  
 Research Fellow  
 Cambridge, MA, USA  
 Hill Technical Center

17 recommendations ✓

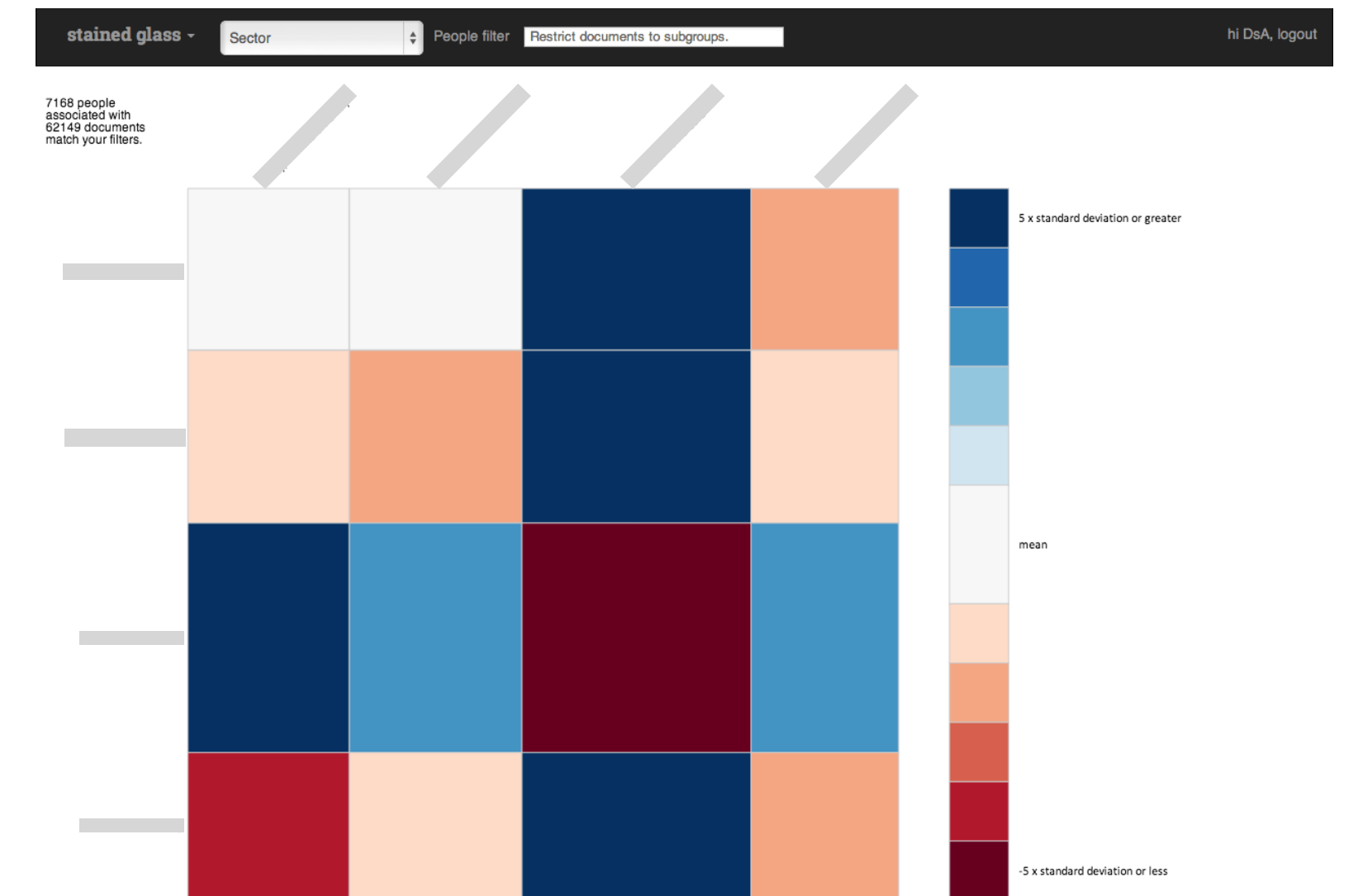
**2 mutual connections with Chip**

**Expertise**

- Machining
- Manufacturing
- Web Handling
- Nanotubes
- Extrusion

**Additional Info**

- [Social Network Survey Page](#)
- [Profile](#)
- [Articles](#)





# care about design.





datascope

Iterative design for data science projects

**Bo Peng** • @bo\_p

for QCon San Francisco • Thanks!