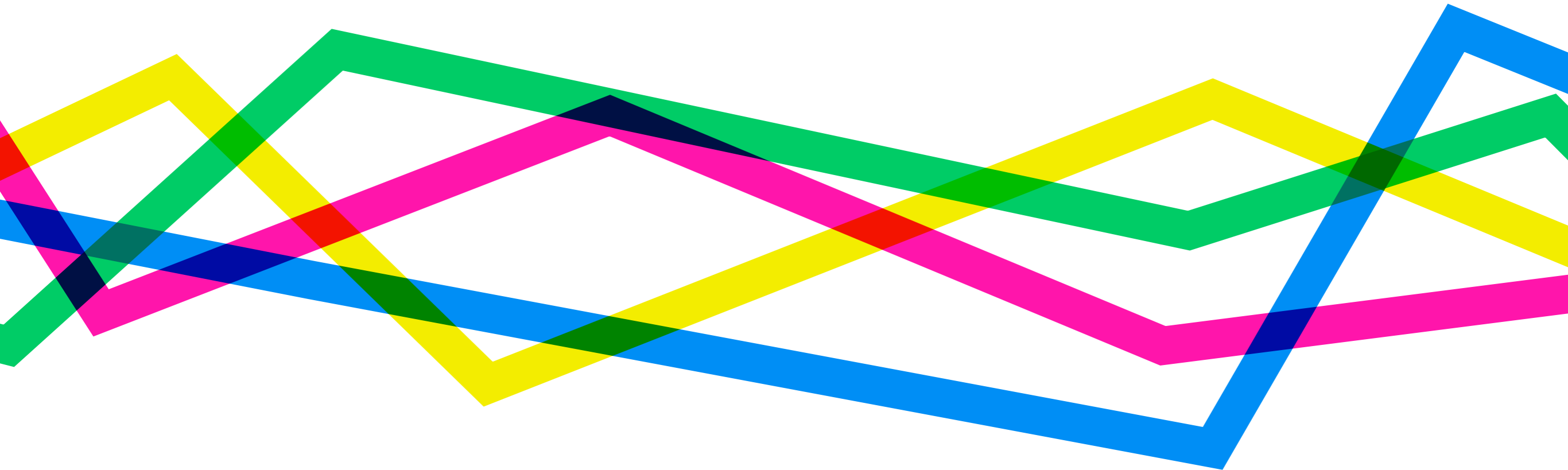


datascope



Data science in the wild wild midwest

Bo Peng • @bo_p

for Big Data Wisconsin • Aug 22, 2016

Agenda

- A common approach to data science
- How we do data science
- How we teach others to do data science

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

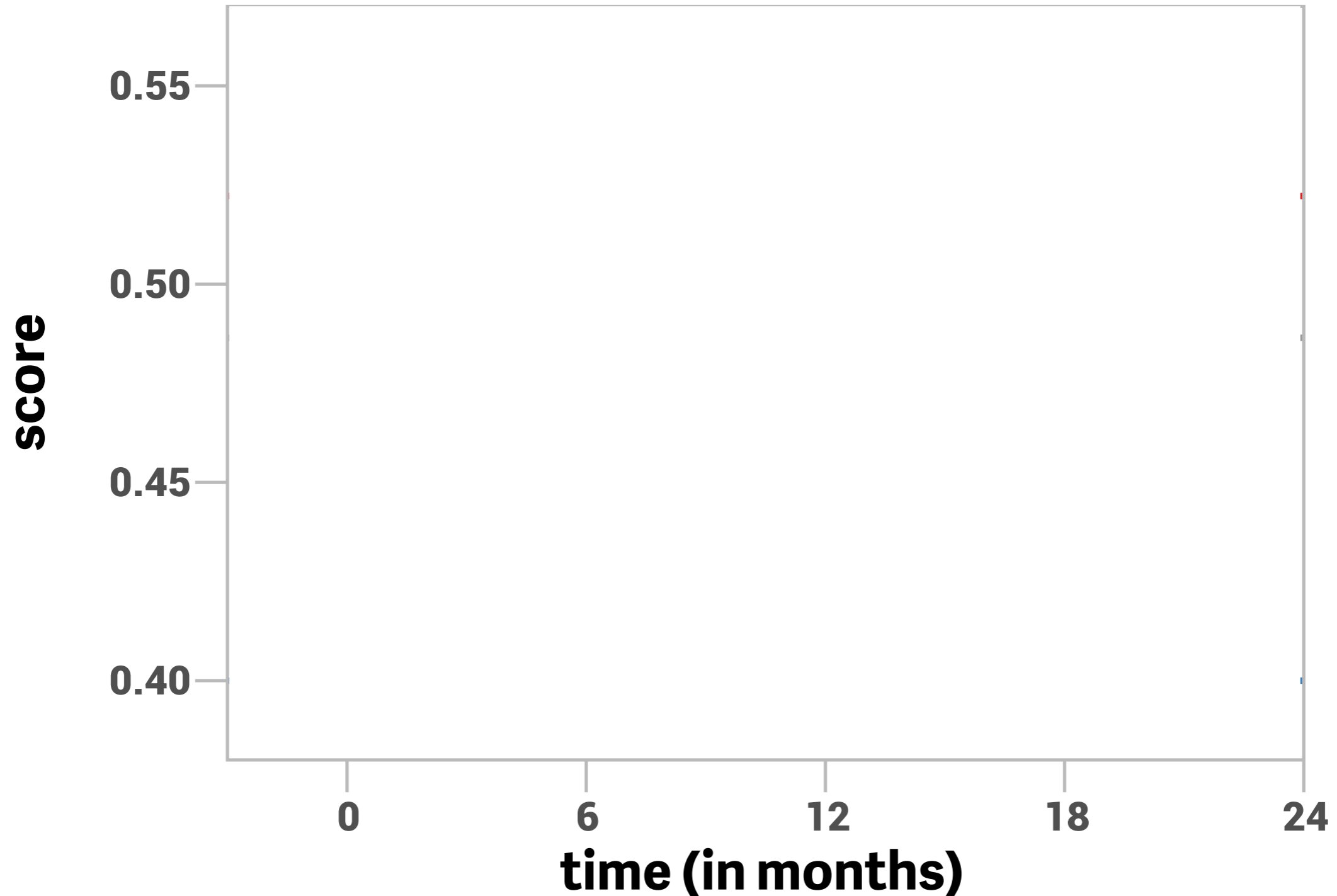
2 years

1,363 teams

25,316 entries

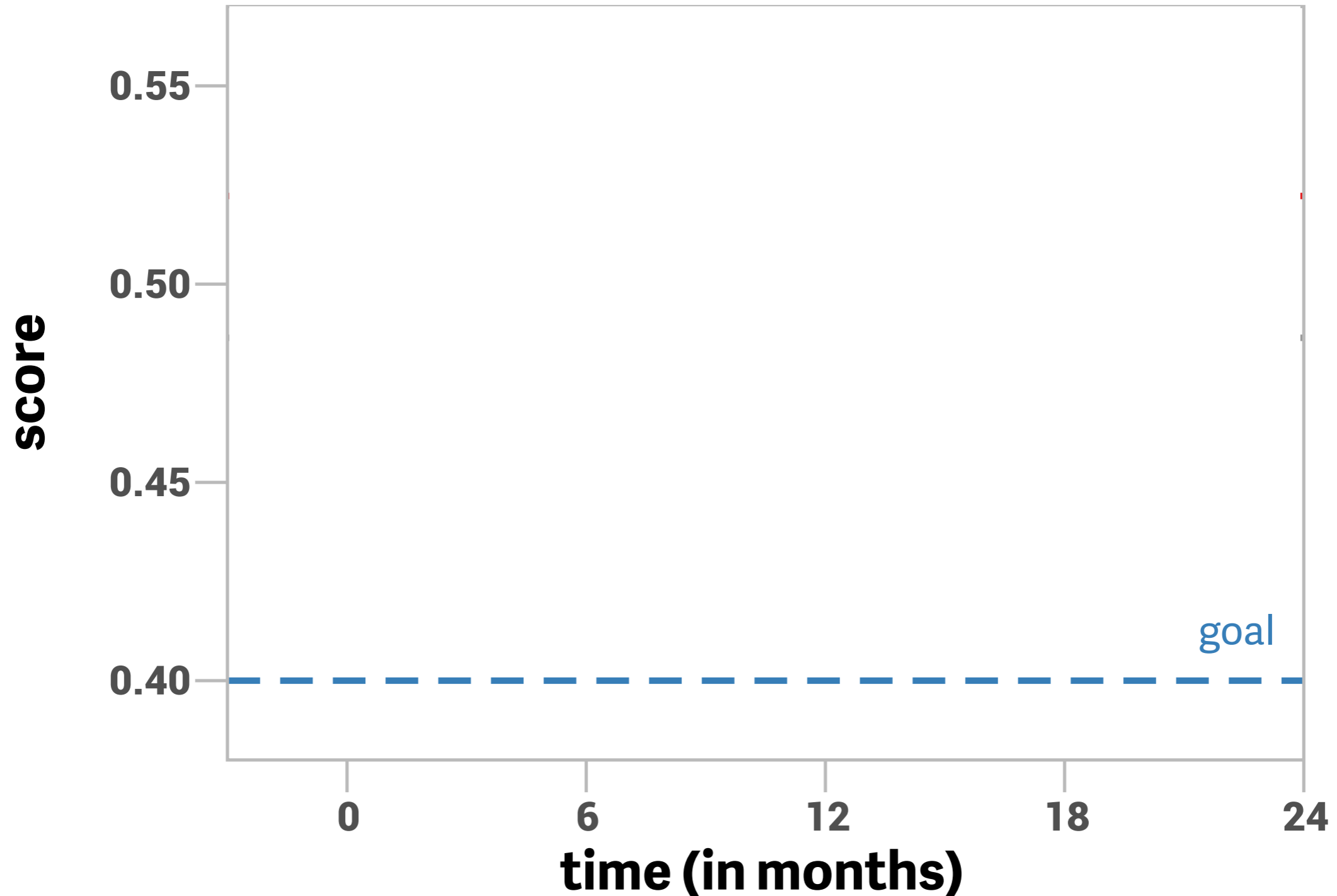
approach

case study: heritage health prize



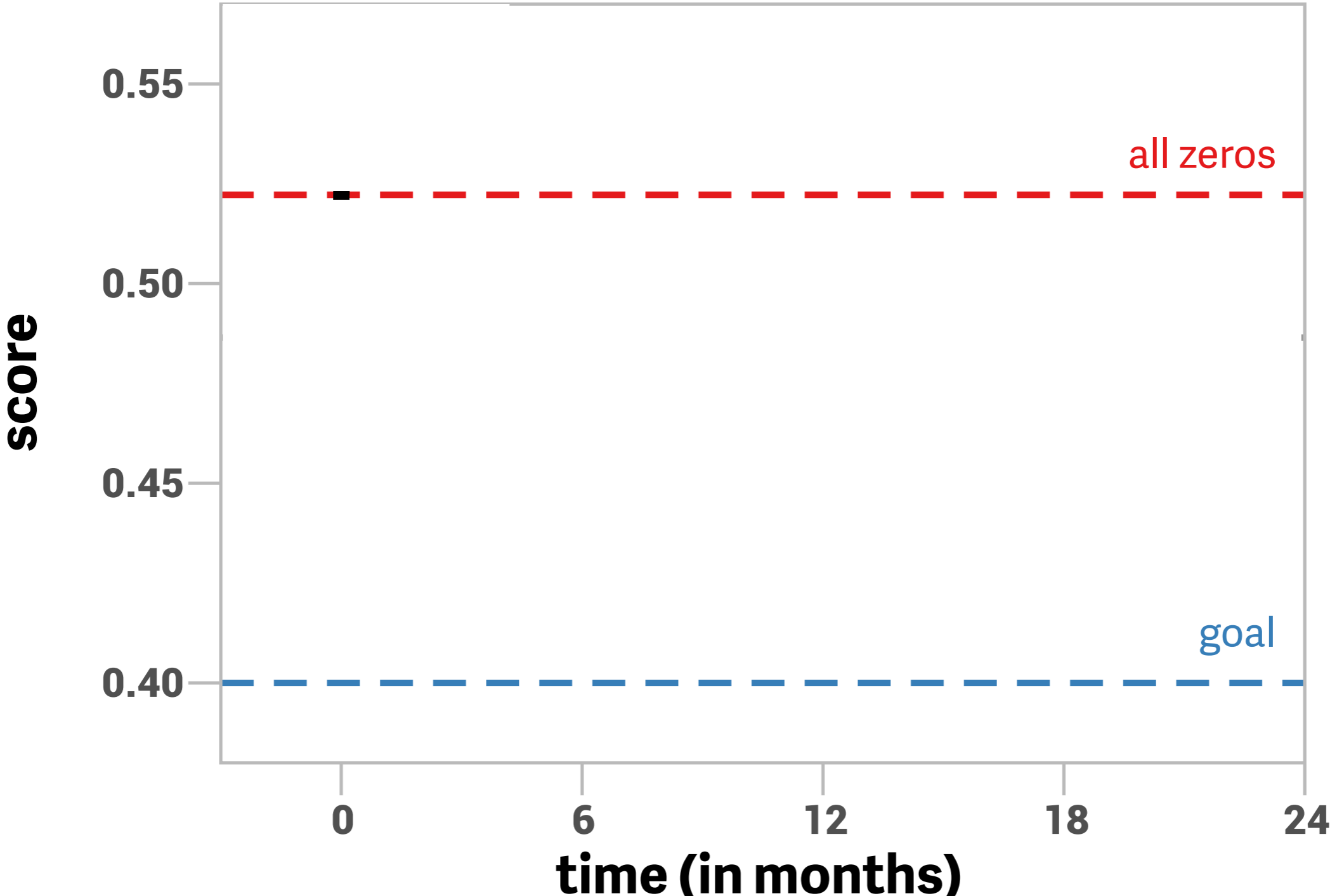
approach

case study: heritage health prize



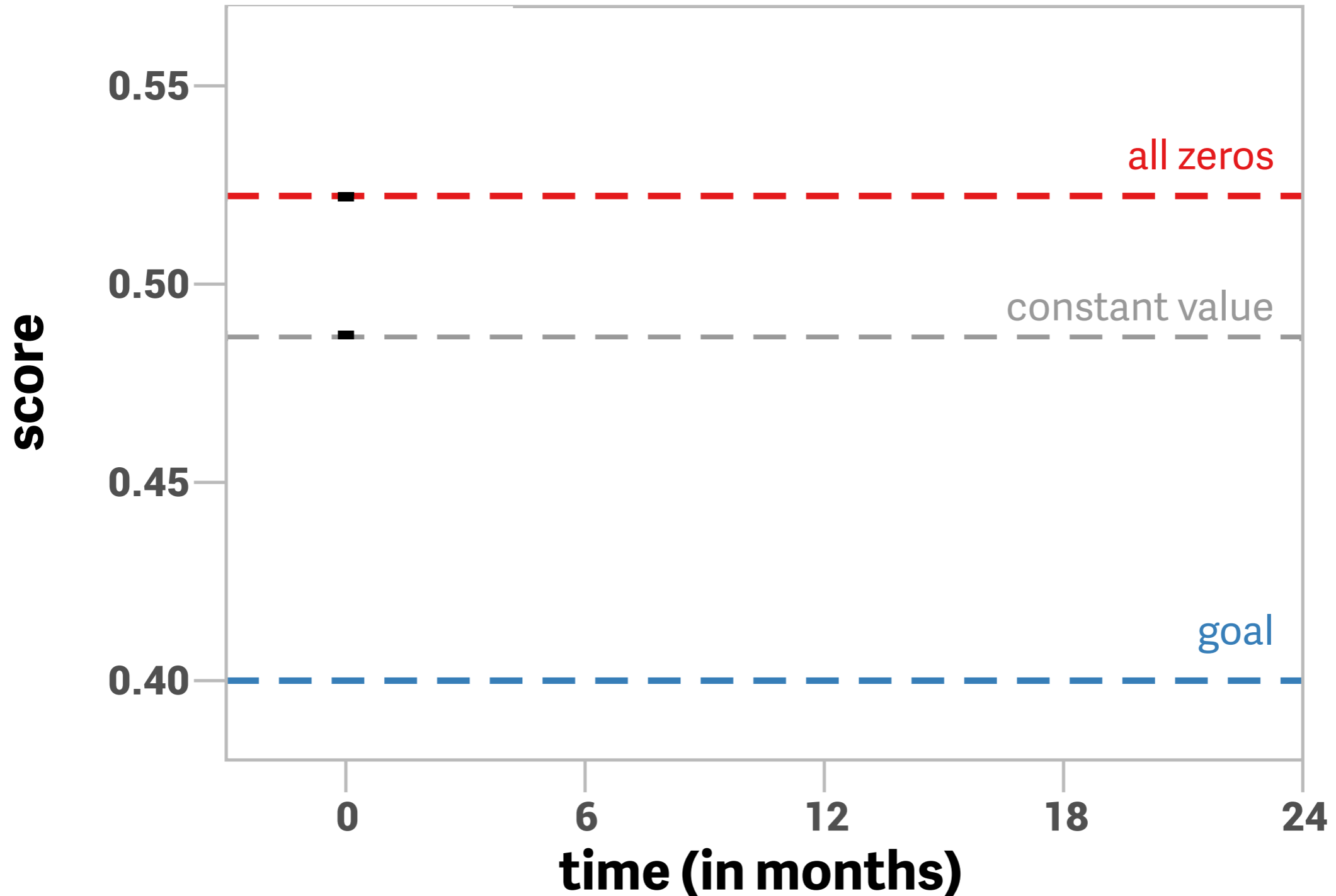
approach

case study: heritage health prize



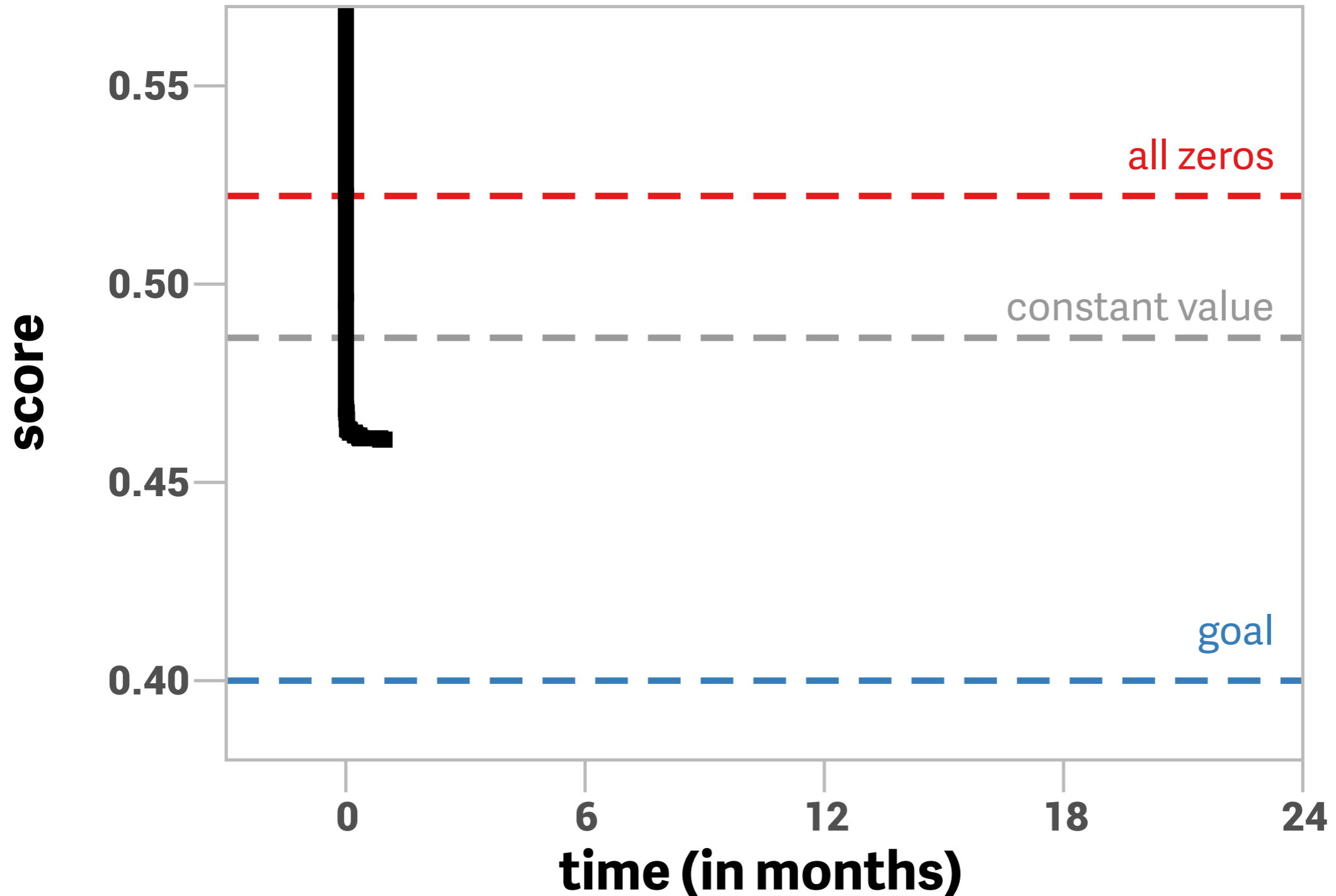
approach

case study: heritage health prize



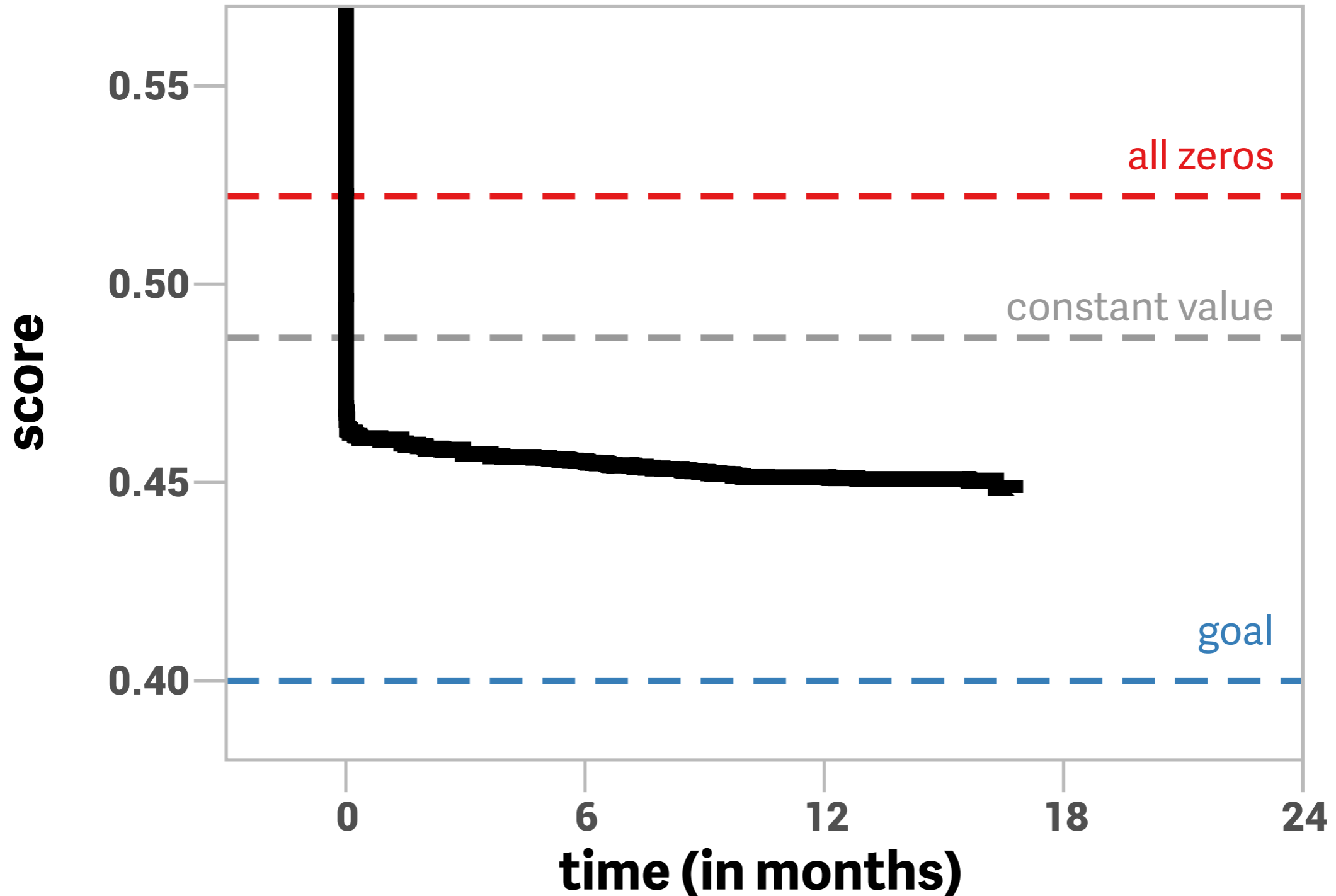
approach

case study: heritage health prize



approach

case study: heritage health prize



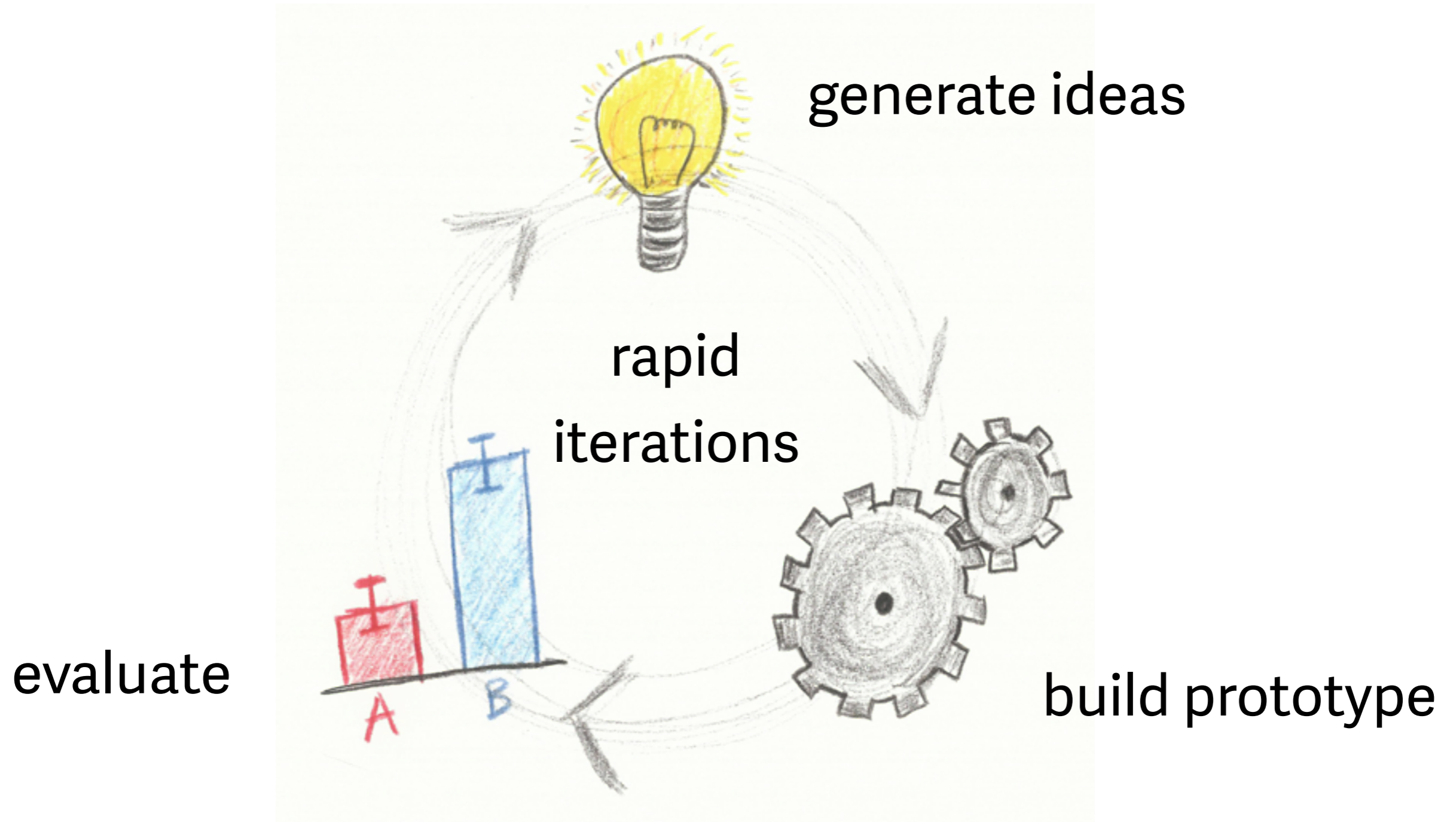
care about design.

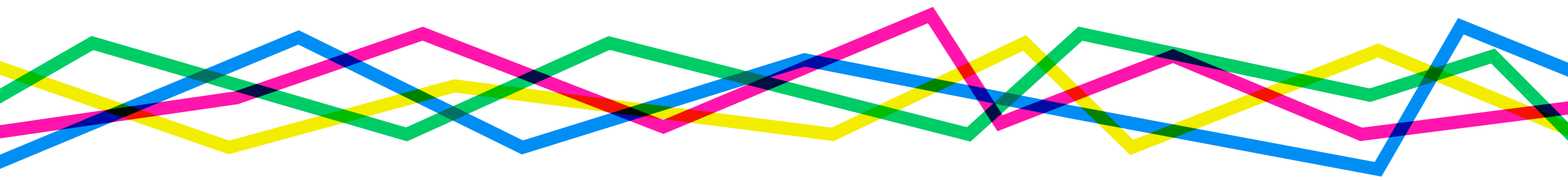
simple, powerful interfaces relay analytics better.



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



data-driven expertise exploration

procter & gamble



data-driven expertise exploration

procter & gamble



Lorem Ipsum: 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.

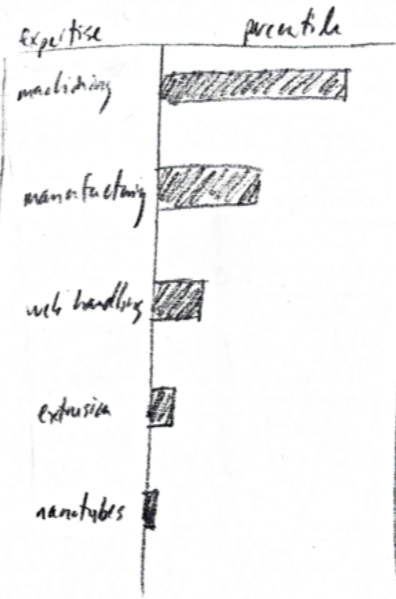
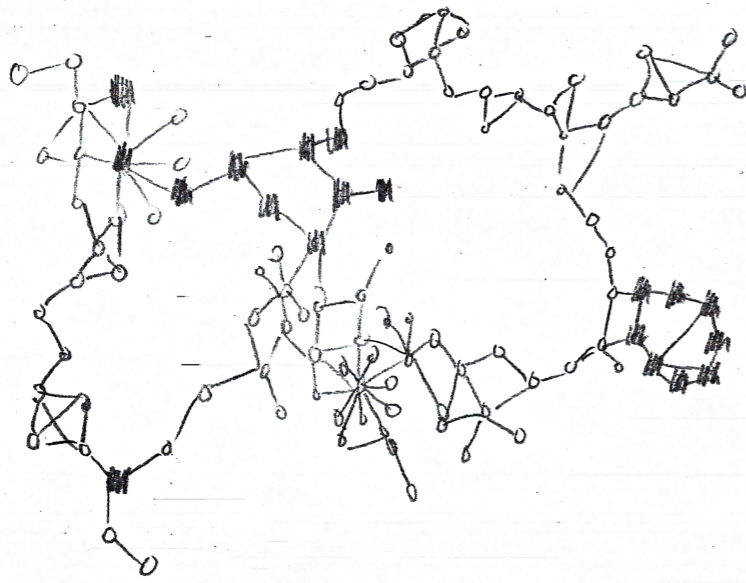
However, a placeholder text must have a natural distribution of letters and punctuation or otherwise the markup will look strange and unnatural. That's what Lorem Ipsum helps to achieve.

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

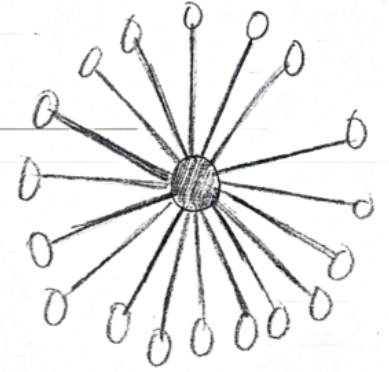
Lorem Ipsum as well as the infinite wisdom of Linus van Pelt and his willingness to use his blanket in my experiments.

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, Suttia
 Grant, Herace
 Cartwright, Bill

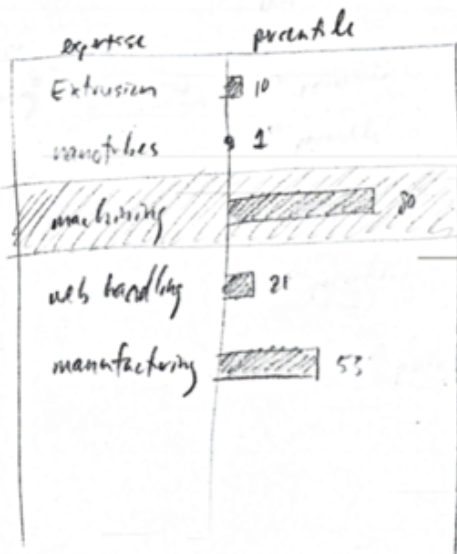


Brown, Chuck
 TU7263

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



GBO: 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:

I am robot: a review

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"

email

you've emailed Hob 4 times in the last year

data-driven expertise exploration

procter & gamble

Asker:

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: #4a90e2; border: 1px solid black;"></div>	<div style="width: 60%; background-color: #f44336; border: 1px solid black;"></div>
<input type="checkbox"/>	 sue Bruston	<div style="width: 70%; background-color: #4a90e2; border: 1px solid black;"></div>	<div style="width: 75%; background-color: #f44336; border: 1px solid black;"></div>
<input type="checkbox"/>	 Jed Bedson	<div style="width: 70%; background-color: #4a90e2; border: 1px solid black;"></div>	<div style="width: 65%; background-color: #f44336; border: 1px solid black;"></div>
<input type="checkbox"/>	 Tim Thomson	<div style="width: 70%; background-color: #4a90e2; border: 1px solid black;"></div>	<div style="width: 85%; background-color: #f44336; border: 1px solid black;"></div>

load more...



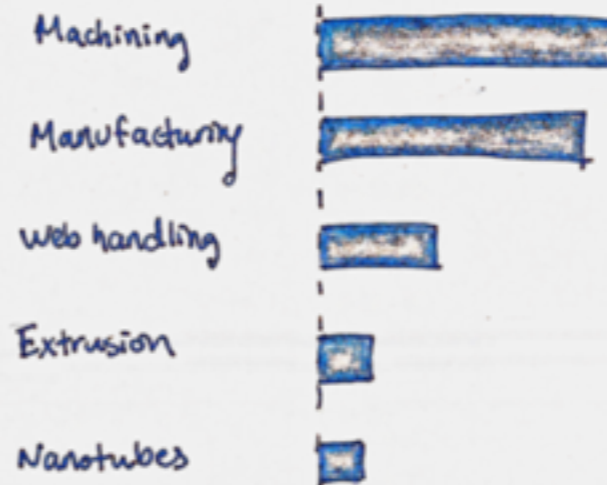
Brown, Chuck
Cambridge, MA, USA
Chemical Engineer

 email
4 emails sent recently

8 mutual connections with:

				
--	--	--	--	--

Area of Expertise




Activity

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

data-driven expertise exploration

procter & gamble







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

<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](#)

data-driven expertise exploration

wikitopic

WikipediaTools + DsA API Explorer

datascope ▾

Api Root > Document List

Document List

OPTIONS

GET ▾

API endpoint that allows documents to be created.

GET /api/v0-5/documents/

HTTP 200 OK

Content-Type: application/json

Vary: Accept

Allow: GET, POST, HEAD, OPTIONS

```
{
  "count": 4,
  "next": null,
  "previous": null,
  "results": [
    {
      "id": 36521,
      "stopwords": "",
      "created_datetime": "2016-02-05T22:22:03.860Z",
      "corpus_version": "enwiki-20140614-pages-articles.xml.bz2",
      "articles": [],
      "categories": [
        {
          "category": {
            "id": 46612,
            "title": "Masami Okui albums"
          }
        }
      ]
    }
  ]
}
```



Creating more of us

Metis Data Science Bootcamp Curriculum

bootcamp curriculum

metis

Who can be a data scientist?

Personality traits:

Curiosity

Creativity

Grit

bootcamp curriculum

metis

Who can be a data scientist?

Personality traits:

Curiosity ... Continuous drive to learn

Creativity ... Tons of different angles & ideas

Grit ... Ability to push through walls

+ Integrity

+ Humility

bootcamp curriculum

metis

Bootcamp model



Training

Repetition

Culture & community

bootcamp curriculum

metis

Bootcamp model: training

Skills

O

S

E

M

N

bootcamp curriculum

metis

Bootcamp model: training

Skills

Obtain

Scrub

Explore

Model

iNterpret

Communicate

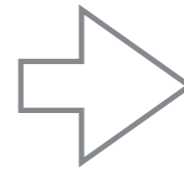
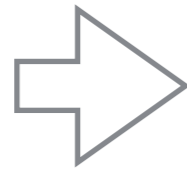
bootcamp curriculum

metis

Bootcamp model: training

Skills

Obtain Data acquisition
Scrub Data exploration
Explore Machine learning
Model & statistics
iNterpret Computer science
Communicate Web
Visualization
Domain awareness
Presentation ability
& professionalism



Tools

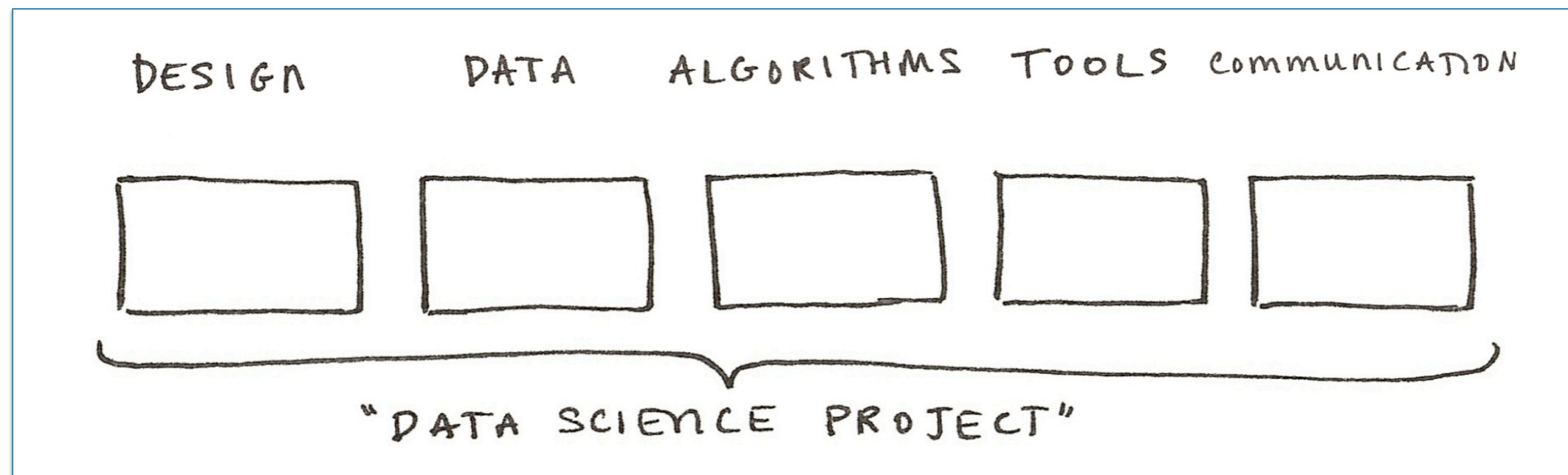
Python when possible
ipython notebook, git/github
BeautifulSoup, selenium, cron
(flatfiles), MySQL, mongoDB
numpy/scipy/matplotlib, pandas
statsmodels, scikit-learn
HTML, CSS, javascript, d3.js
DigitalOcean/AWS, Flask
Hadoop/Hive, Spark
google

bootcamp curriculum

metis

Bootcamp model: repetition

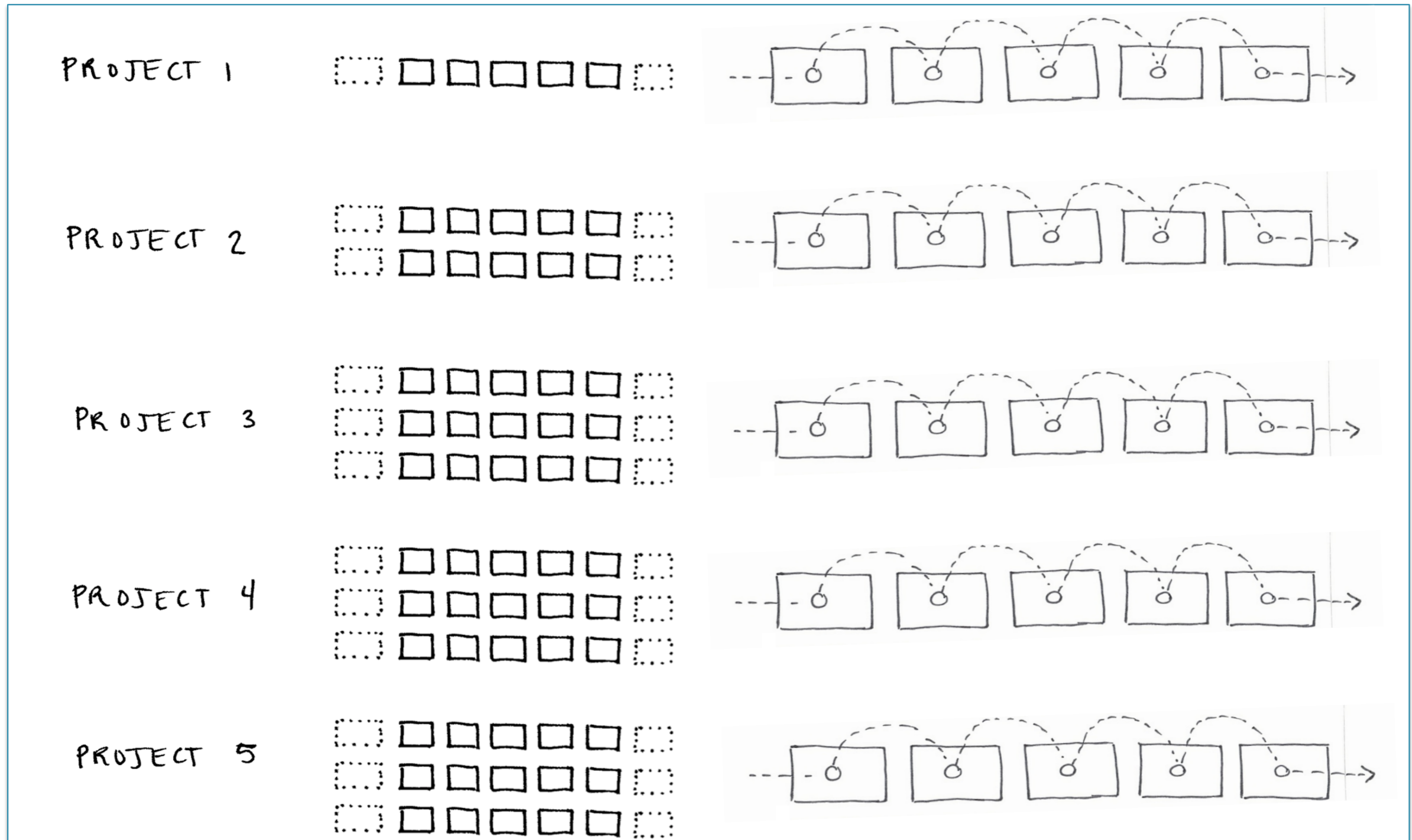
Project-based model



bootcamp curriculum

metis

Bootcamp model: repetition

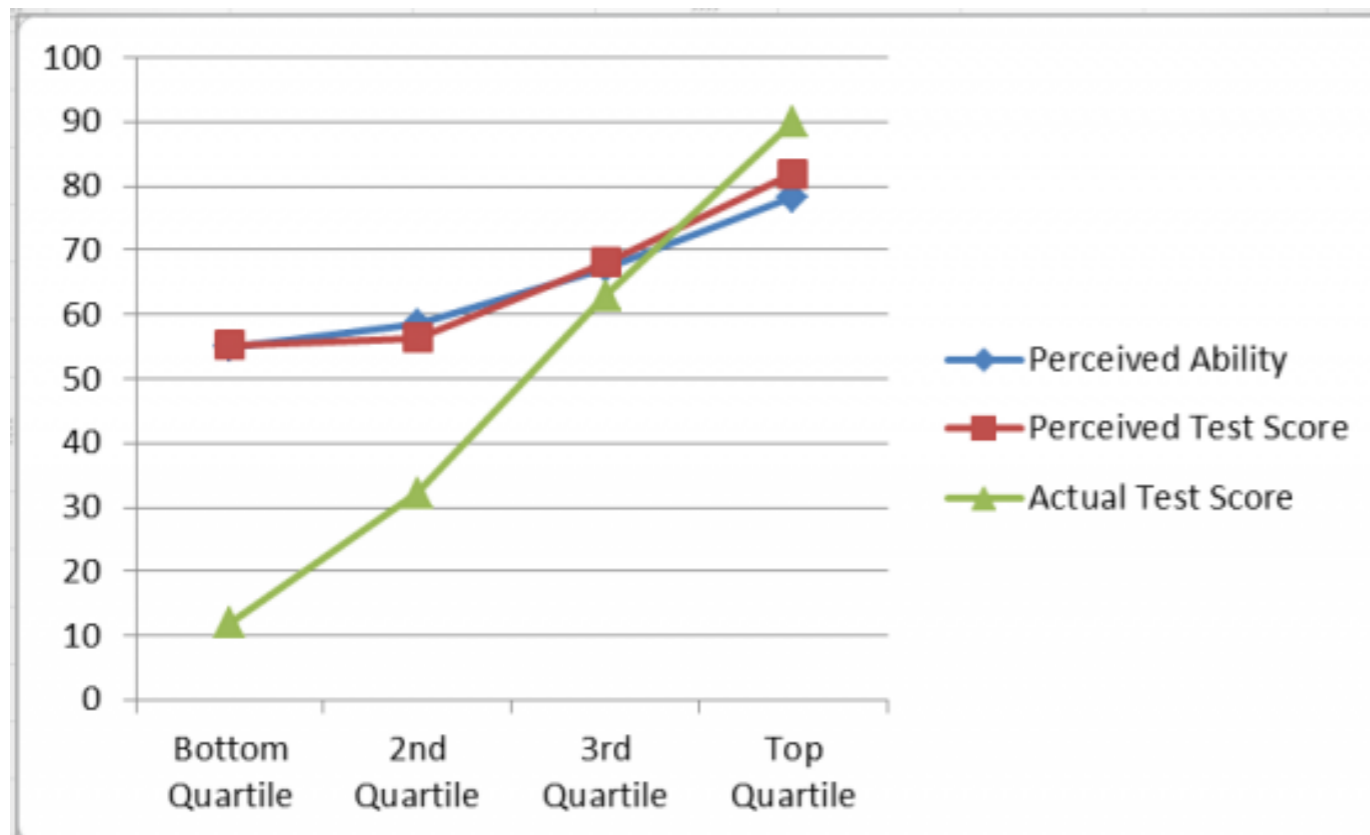


bootcamp curriculum

metis

Bootcamp model: culture & community

Fighting the Dunning-Kruger effect

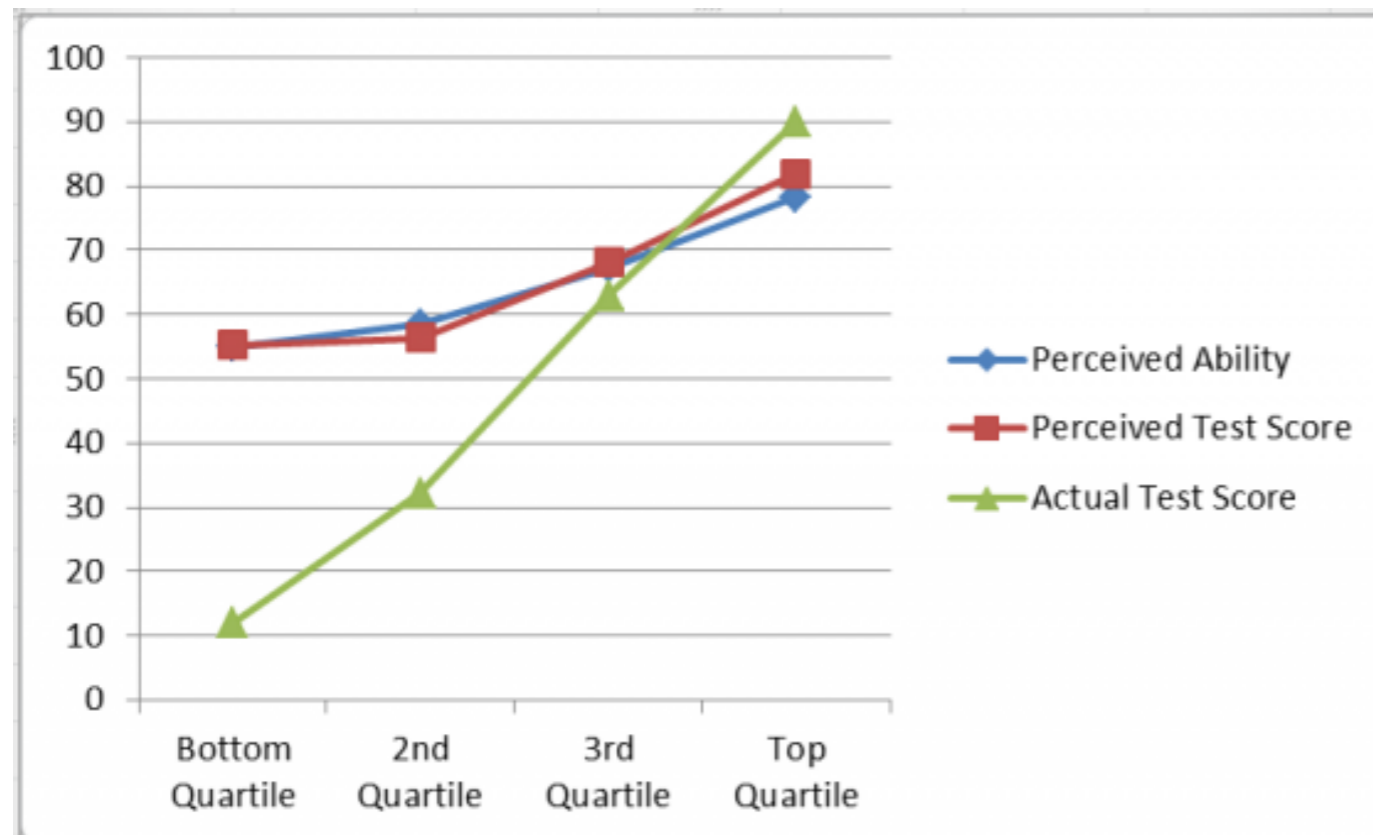


bootcamp curriculum

metis

Bootcamp model: culture & community

Fighting the Dunning-Kruger effect



Overcoming Imposter Syndrome:

- group work
- communication

Overcoming Perfectionism:

- unfairly short deadlines
- jumping into the unfamiliar
- presenting projects

bootcamp grads

metis



TESLA MOTORS



VAYNERMEDIA



PANDORA



UPTAKE



www.meetup.com/Data-Science-Madison/

Next event: Tuesday Sept. 13



Find
a Meetup Group

Start
a Meetup Group

English

Log in

Sign up

Data Science Madison

[Home](#) [Members](#) [Sponsors](#) [Photos](#) [Discussions](#) [More](#)

Join us!

data science madison



Madison, WI

Founded Nov 4, 2014

Data Scientists 437

Group reviews 1

Upcoming Meetups 3

Past Meetups 3

Our calendar

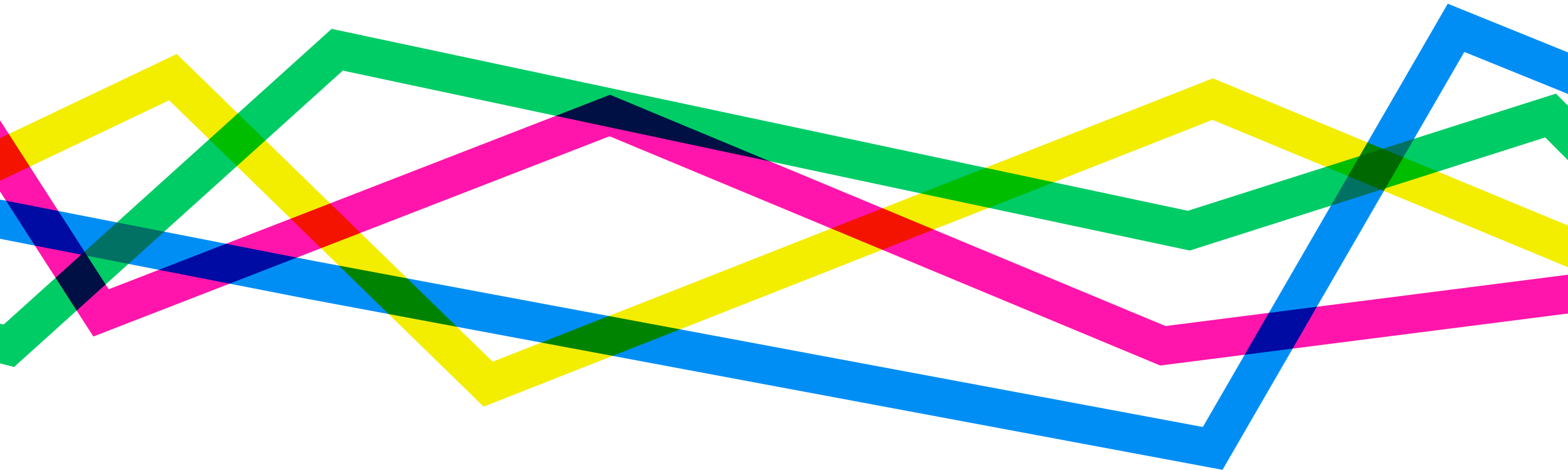
Data Science Madison brings together people interested in data science:

- the analytical "statistics and machine learning" side,
- the creative "visualization and design" side,
- the verbal "communication and story-telling" side,
- and especially the areas in between those extremes.

Members are professionals, students, consultants, enthusiasts and entrepreneurs who love data and understand the potential value it brings for society, business and government. Meeting activities are varied and range from tutorials on new technologies and their applications, success stories from practitioners, cutting-edge academic data science, and social activities. Long, boring talks should seek other venues :)

We are inclusive of all genders and races and try to foster a safe environment where everyone feels comfortable to share and learn. We welcome your ideas on how to make this group an exciting growth opportunity for everyone. Come to meet others and contribute to the community!

datascope



Data science in the wild wild midwest

Thanks!

Bo Peng • @bo_p