NEXT: Crowdsourcing, machine learning and cartoons

Scott Sievert
UW–Madison ECE
@stsievert

+

#pydata =

https://speakerdeck.com/stsievert/next-crowdsourcing-machine-learning-and-cartoons

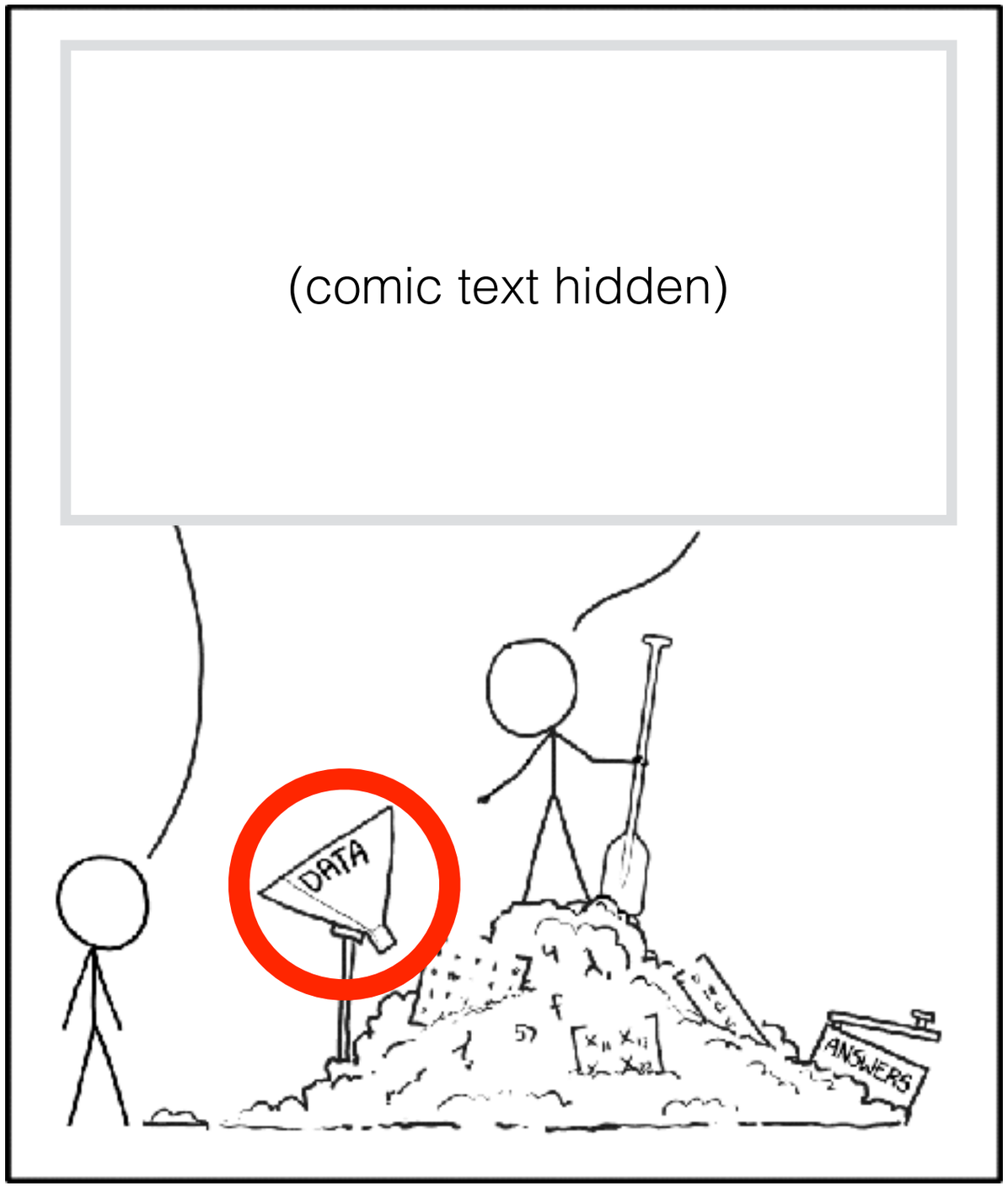
https://tinyurl.com/next-pydata

NEXT enables research
by producing
better results for the
same time or money

NEXT is implemented in pure Python

https://tinyurl.com/next-pydata

Problem



Data collection can be costly

https://xkcd.com/1838/

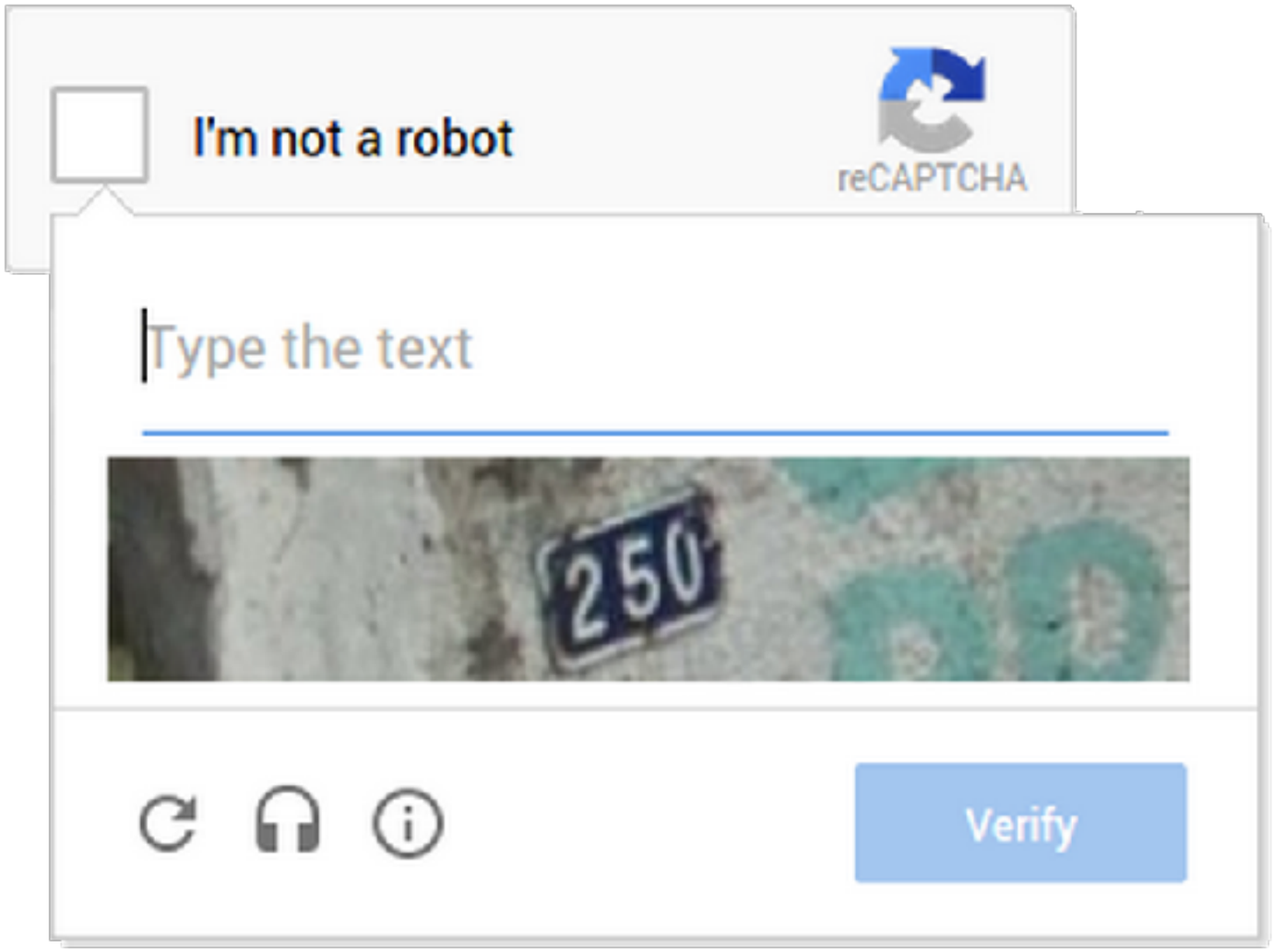
https://tinyurl.com/next-pydata

Problem

Data collection with
crowdsourcing can be expensive

https://tinyurl.com/next-pydata

Example Problem

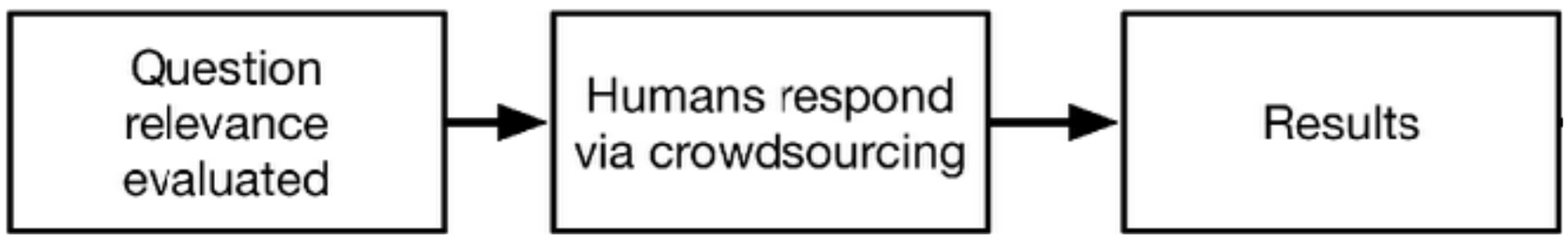


Goal: achieve goal with minimal responses

https://tinyurl.com/next-pydata

One solution

Existing crowdsourcing systems are *passive*

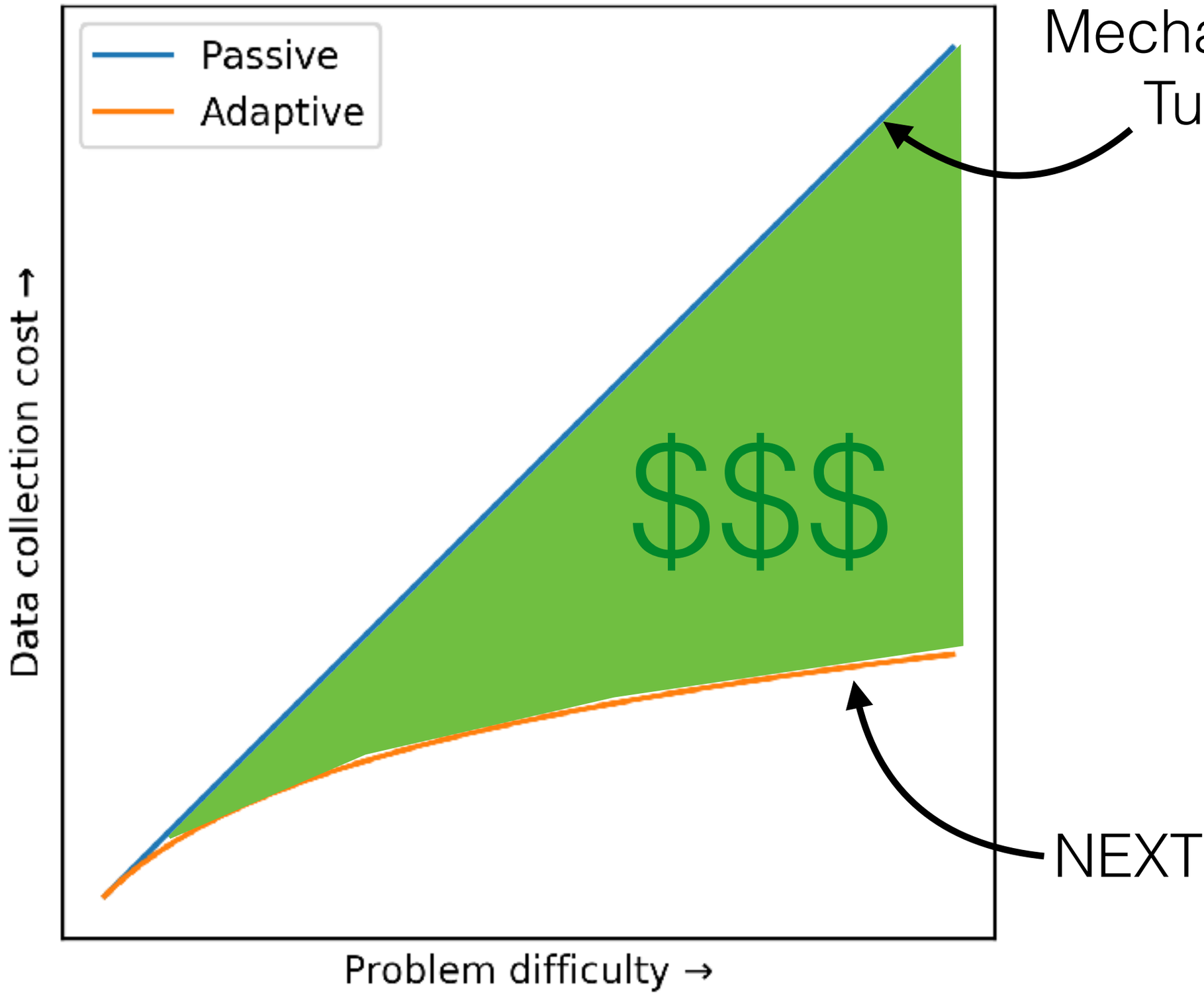


Adapting to previous responses finds the best questions

Goal: adapt to previously collected responses

https://tinyurl.com/next-pydata

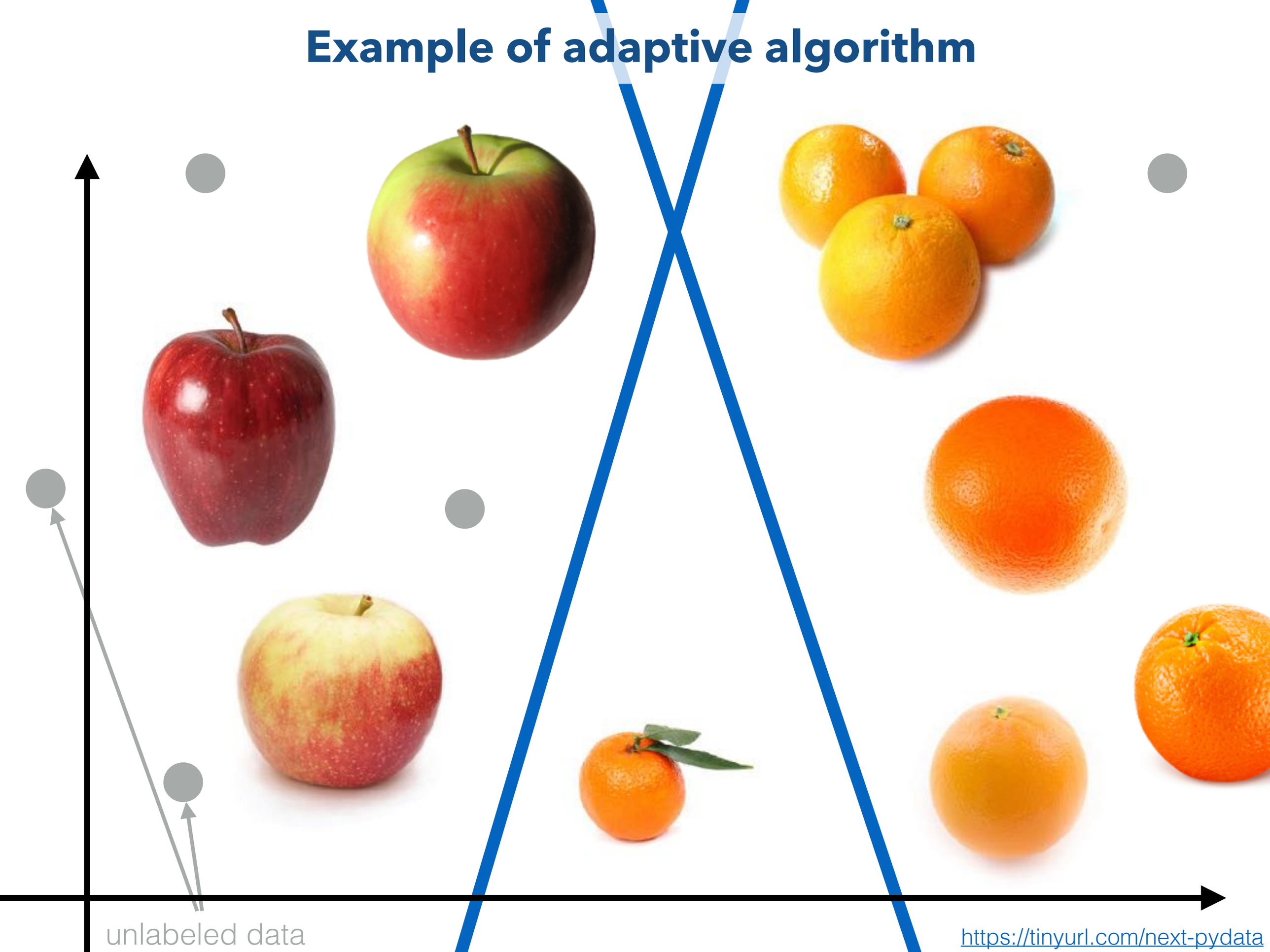
Benefits



Adaptive sampling can have large benefits

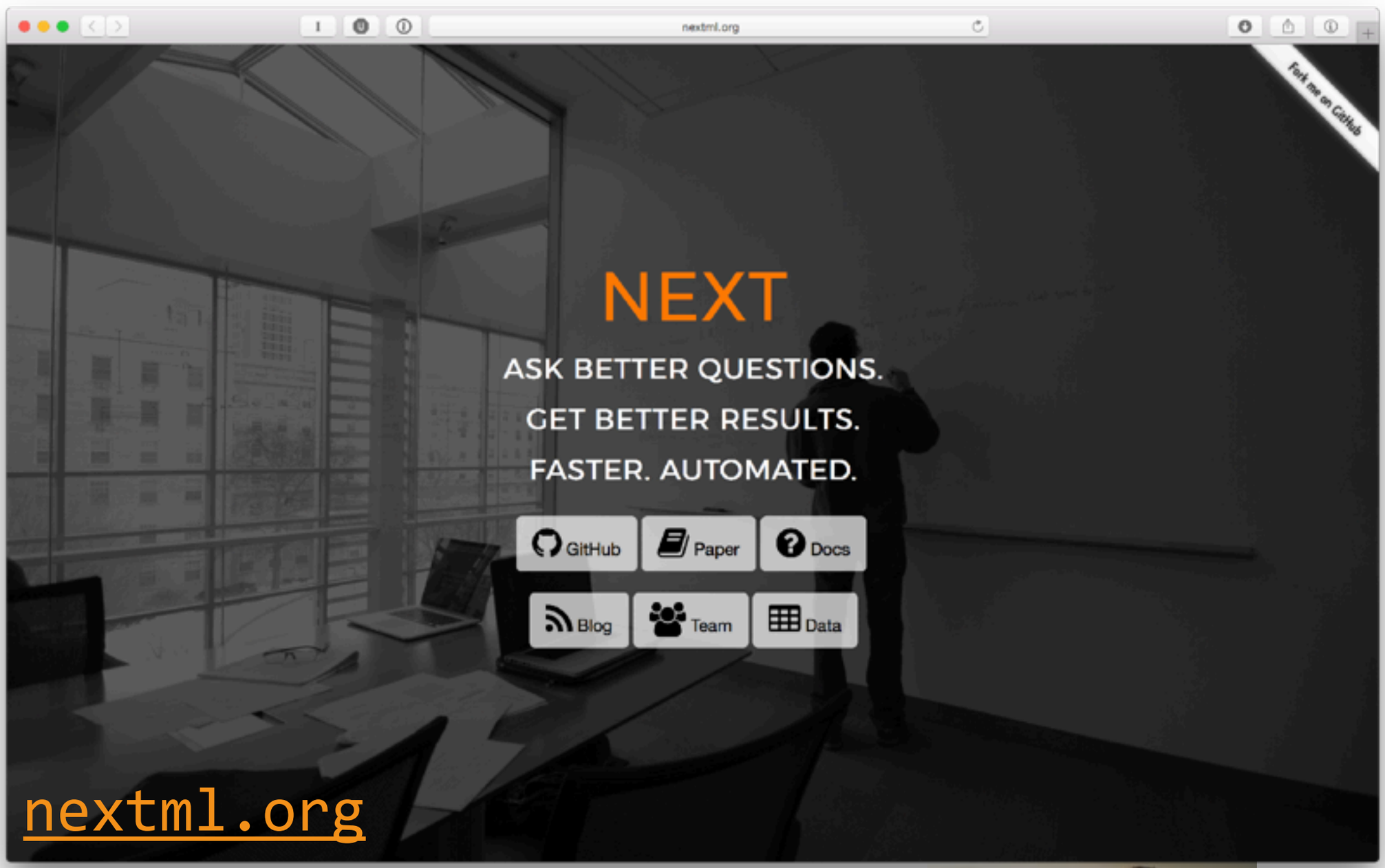
https://tinyurl.com/next-pydata

Example of adaptive algorithm



unlabeled data

https://tinyurl.com/next-pydata



Lalit Jain

Daniel Ross

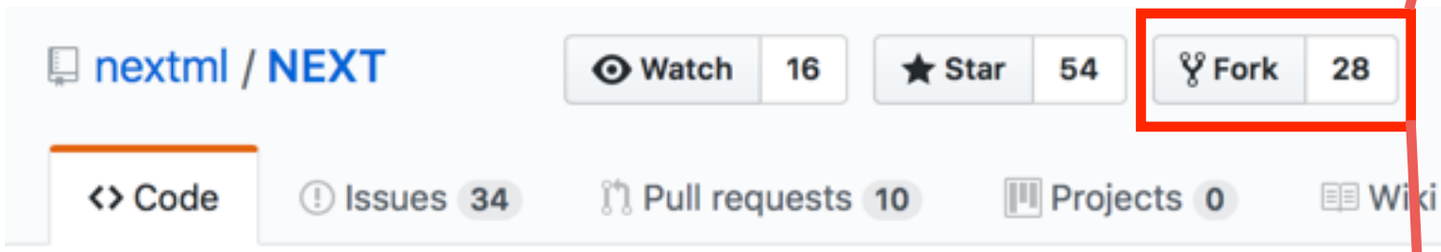
Prof. Rob Nowak

Kevin Jamieson

Homepage: http://nextml.org
Source: https://github.com/nextml/NEXT
Documentation: https://github.com/nextml/NEXT/wiki

https://tinyurl.com/next-pydata

NEXT users

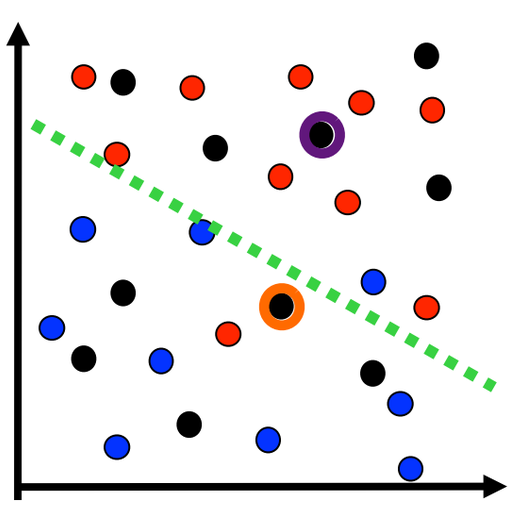


Theory

Practice

ML Researchers

Air Force Research Lab uses NEXT for active image classification.



Experimentalists

UW Psychology uses NEXT to find the best algorithms for adaptive data collection in cognitive science.

Practitioners

The New Yorker uses NEXT to crowd-source the weekly cartoon caption contest.

- nextml / NEXT
  - aashish24 / NEXT
  - abiswas3 / NEXT
  - alphaprime / NEXT
  - aniruddhajb / NEXT
  - AvinWangZH / NEX
  - ayonsn017 / NEXT
  - caomw / NEXT
  - connectthefuture /
  - crcox / NEXT
  - dconathan / NEXT
    - robinsonkwame
  - jattenberg / NEXT
  - jimwmg / NEXT
  - justicelee / NEXT
  - juthawong / NEXT
  - liamim / NEXT
  - mllewis / NEXT
  - NandanaSengupta
  - naveendennis / NE
  - pedmiston / NEXT
  - samim23 / NEXT
  - stsievert / NEXT
    - BhargavaA / NE
  - suchow / NEXT
  - sumeetsk / NEXT-1
  - widoptimization-wi
  - worldbank / NEXT

https://tinyurl.com/next-pydata

Example problem

[Bob Mankoff]

THE NEW YORKER

YOUR CAPTION

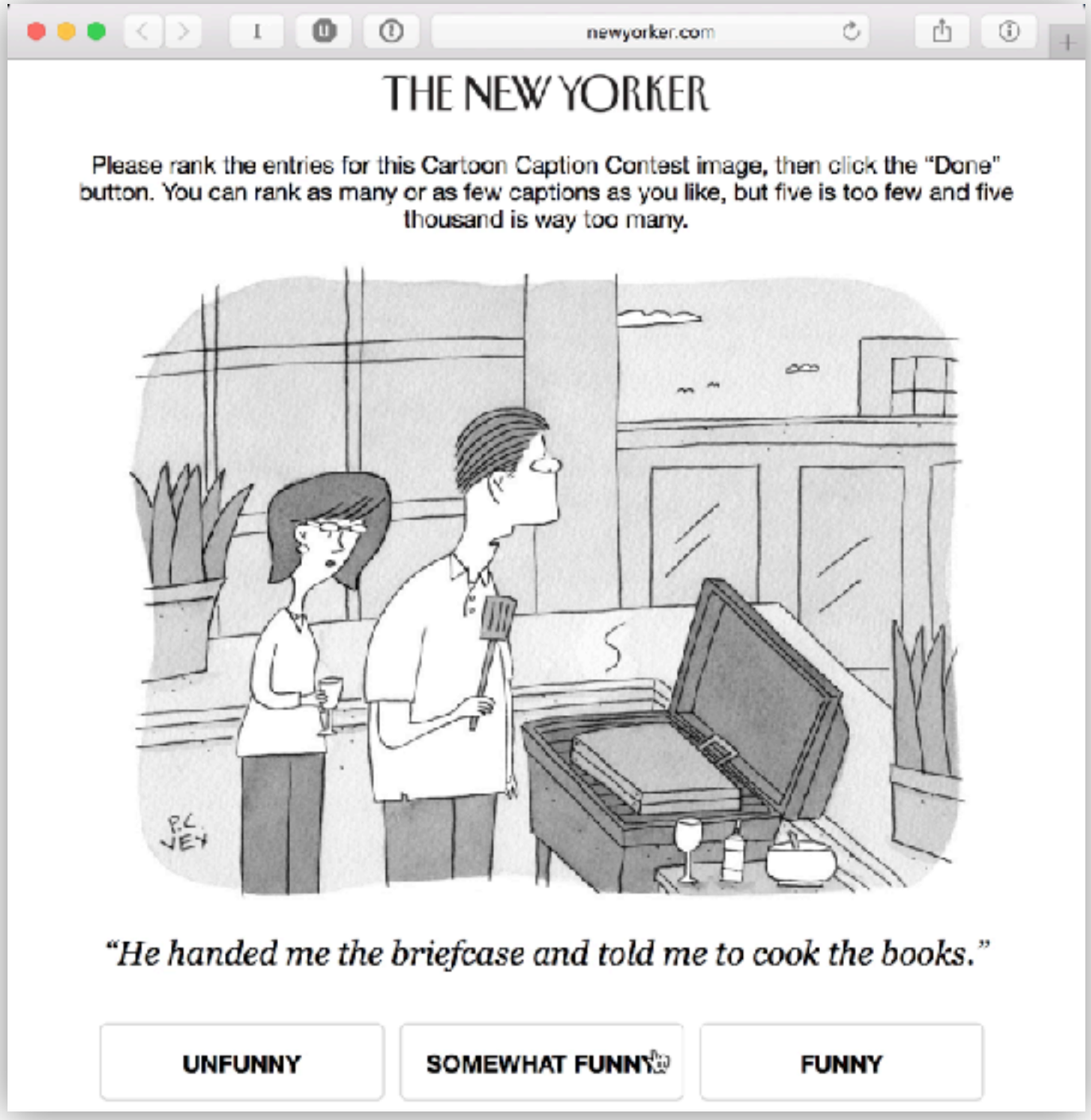
Enter your caption (250 characters or fewer):

The New Yorker has to find the funniest caption from ~5,000 captions

Comic by [P. C. Vey]

https://tinyurl.com/next-pydata

Interface



http://www.newyorker.com/cartoons/vote

http://nextml.org/captioncontest

Comic by P. C. Vey

https://tinyurl.com/next-pydata

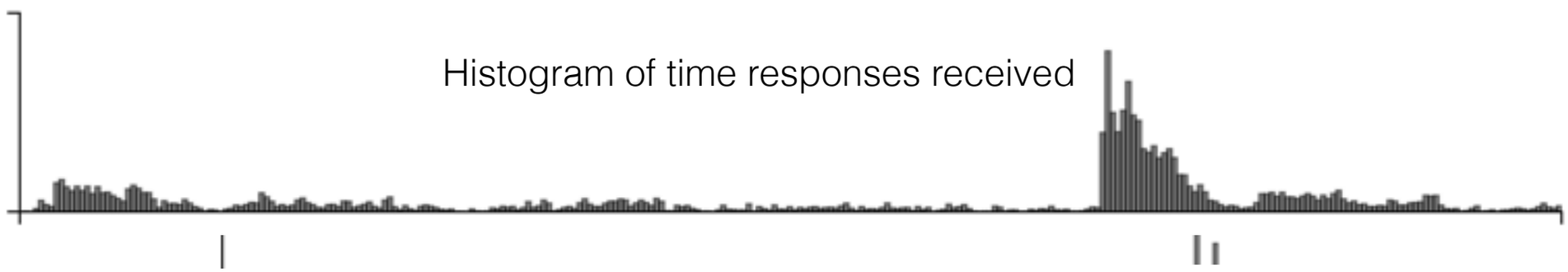
Dashboard

ec2-34-208-101-150.us-west-2.compute.amazonaws.com

NEXT Experiment Dashboard | NEXT - b85faed0ebbd9c6256bdcaa5f2951a

NEXT

EXPERIMENT SYSTEM



Rankings

KLUCB

Download CSV

Experiment Info

Start date: 2017-06-26 14:39:13.745959 UTC
Dashboard data generated: <1 minute ago
Number of participants: 14111
Number of answers: 423367

Rank	Target	Score	Precision	Count	Unfunny	Somewhat Funny	Funny
1	I'd better give it a little longer. It's a really tough case.	2.0921	0.0135	3107	751	1319	1037
2	He told me to cook					1922	1256
3	Hopefully cooki laundering the money.					768	535
4	Something tells me that there are two prime rib steaks sitting on my desk.	1.9767	0.0121	4341	1437	1568	1336
5	No, I won't be going back to work on Monday. Why do you ask?	1.9679	0.0123	3864	1193	1602	1069

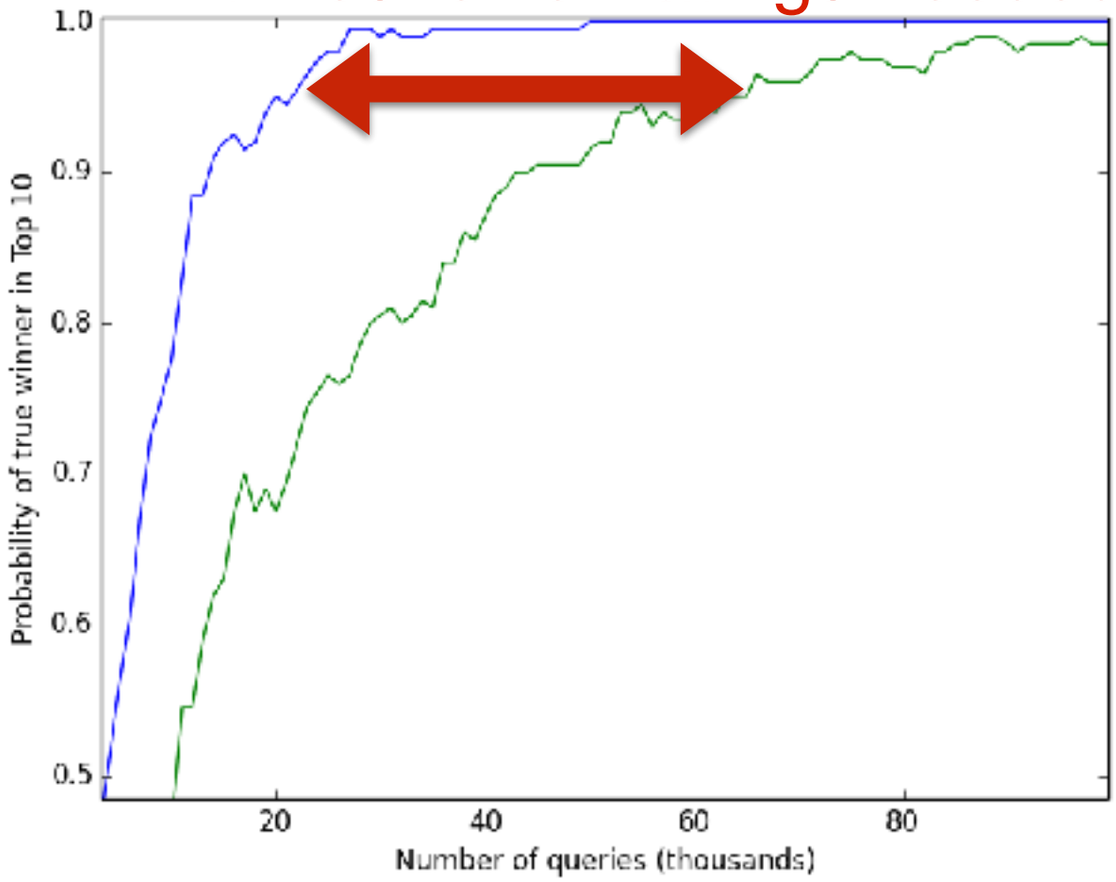
Data from contests:
https://github.com/nextml/caption-contest-data

https://tinyurl.com/next-pydata

I'd better give it a little longer. It's a really tough case.

Experimentalist Benefits

4x times fewer ratings needed!



Papers

A KL-LUCB algorithm for Large-Scale Crowdsourcing

E Tanczos, R Nowak, B Mankoff - Advances in Neural Information ..., 2017 - papers.nips.cc

Abstract This paper focuses on best-arm identification in multi-armed bandits with bounded rewards. We develop an algorithm that is a fusion of lil-UCB and KL-LUCB, offering the best qualities of the two algorithms in one method. This is achieved by proving a novel anytime confidence bound for the mean of bounded distributions, which is the analogue of the LIL-type bounds recently developed for sub-Gaussian distributions. We corroborate our ...

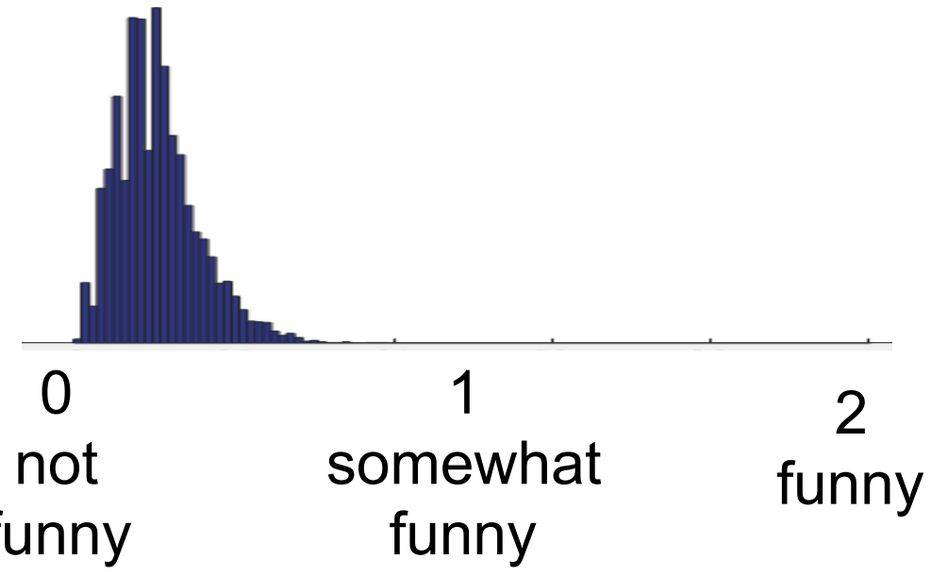
☆ 99 Cited by 1 All 2 versions

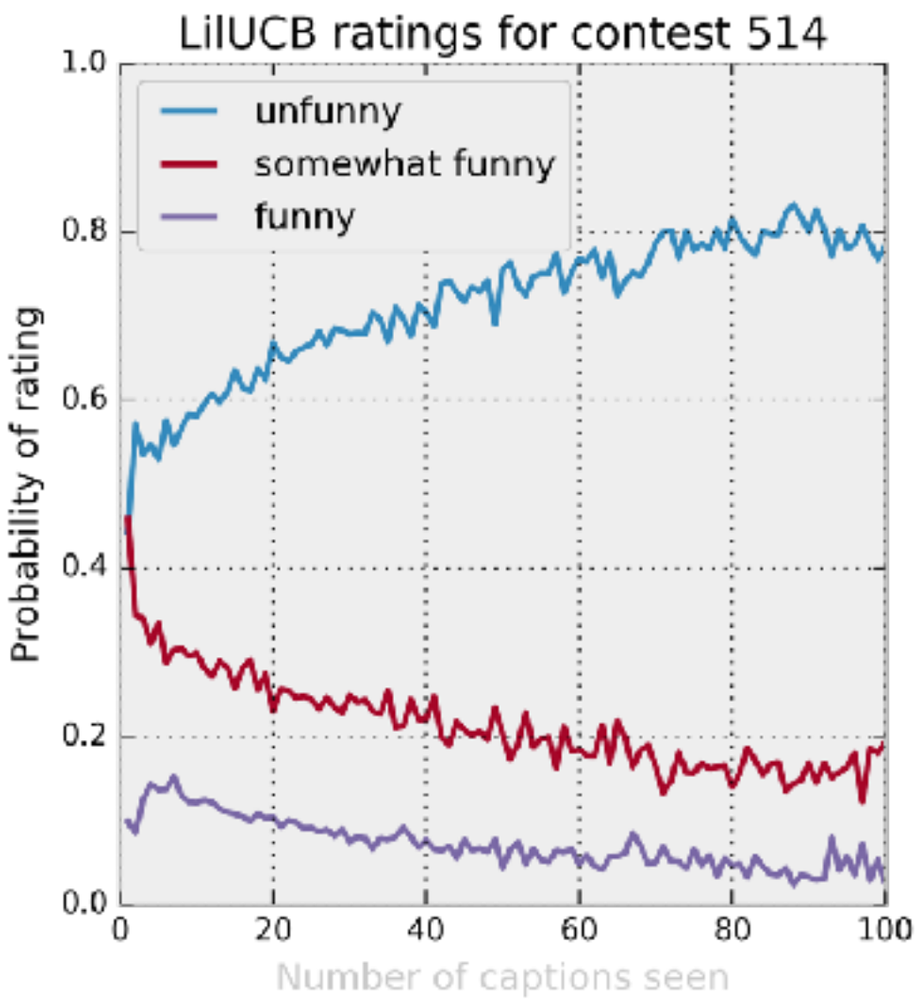
[HTML] Top arm identification in multi-armed bandits with batch arm pulls

KS Jun, K Jamieson, R Nowak, X Zhu - Artificial Intelligence and Statistics, 2016 - jmlr.org

Abstract We introduce a new multi-armed bandit (MAB) problem in which arms must be sampled in batches, rather than one at a time. This is motivated by applications in social media monitoring and biological experimentation where such batch constraints naturally arise. This paper develops and analyzes algorithms for batch MABs and top arm

Data

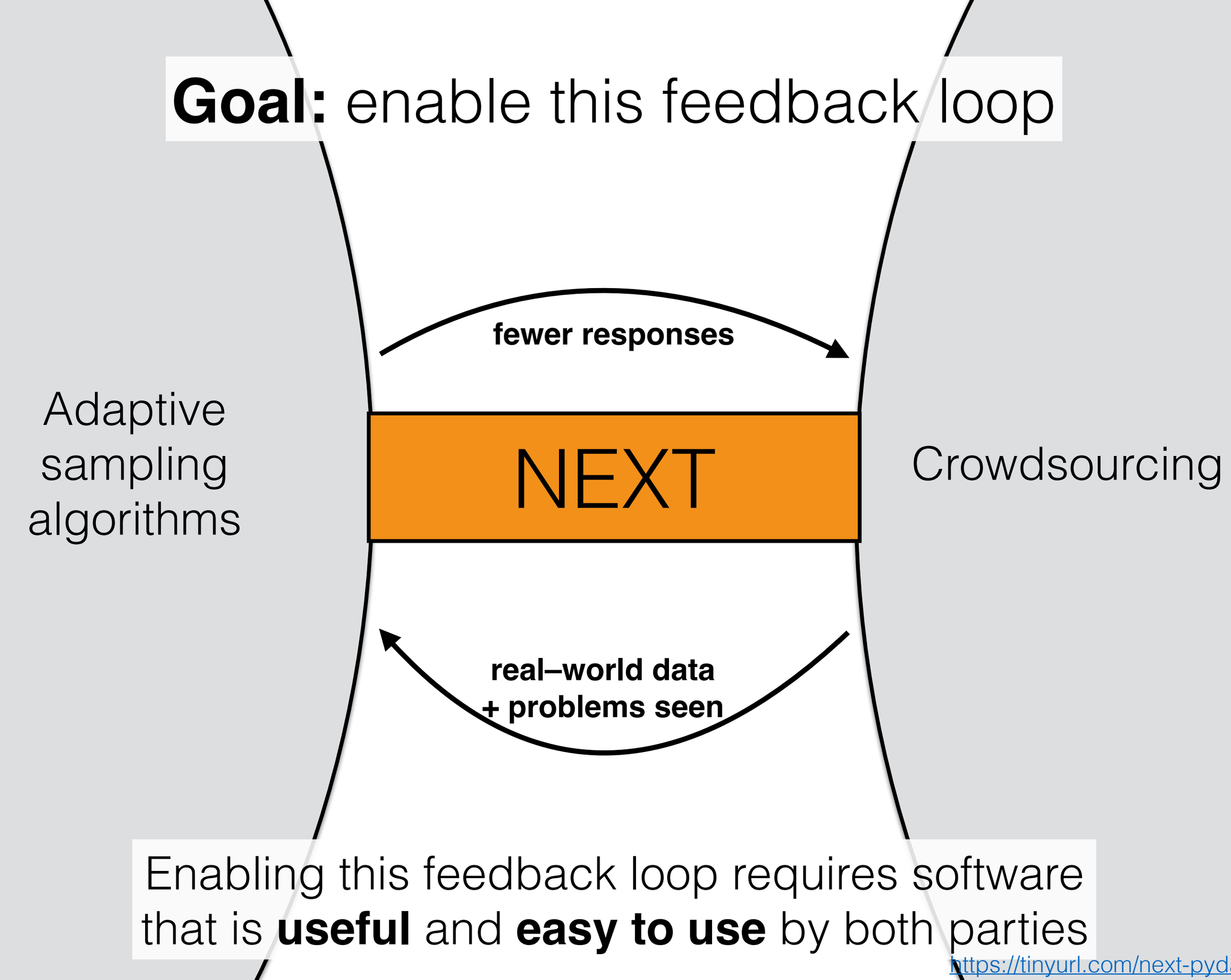




+ software enhancements

https://tinyurl.com/r...t-pydata

Goal: enable this feedback loop



Adaptive sampling algorithms

Crowdsourcing

Enabling this feedback loop requires software that is **useful** and **easy to use** by both parties

https://tinyurl.com/next-pydata

Default uses

By default, NEXT has adaptive algorithms for the 3 default question types

Select expression on the bottom most similar to the face on top

Pool based triplets

Cardinal Bandits

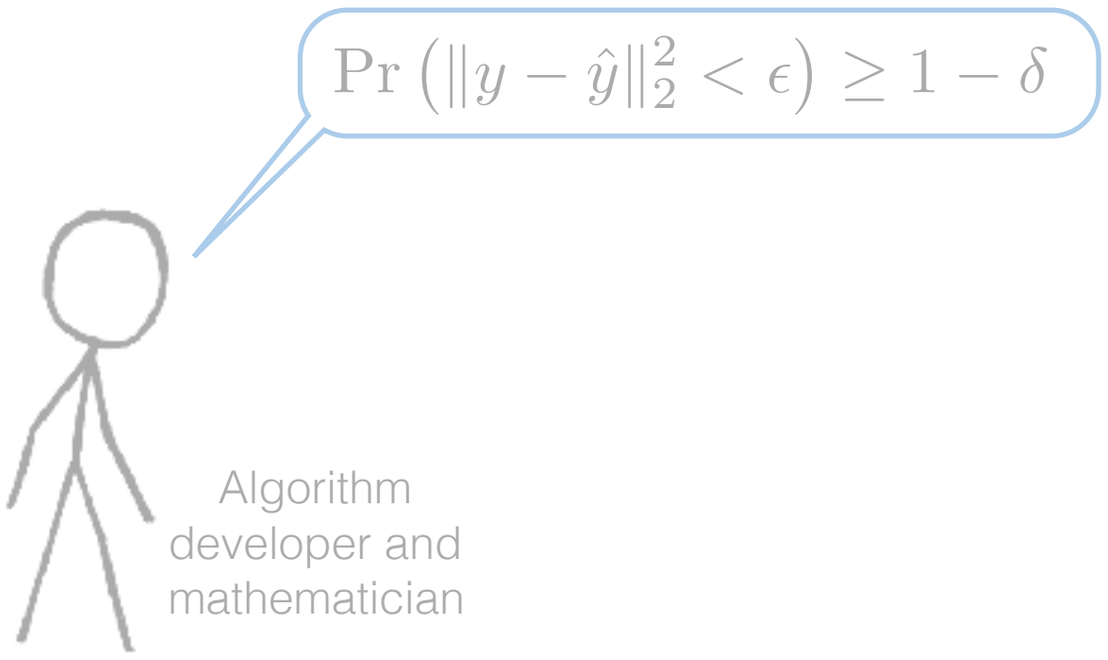
comic by P. C. Vey

Dueling Bandits

Select the street that looks safer

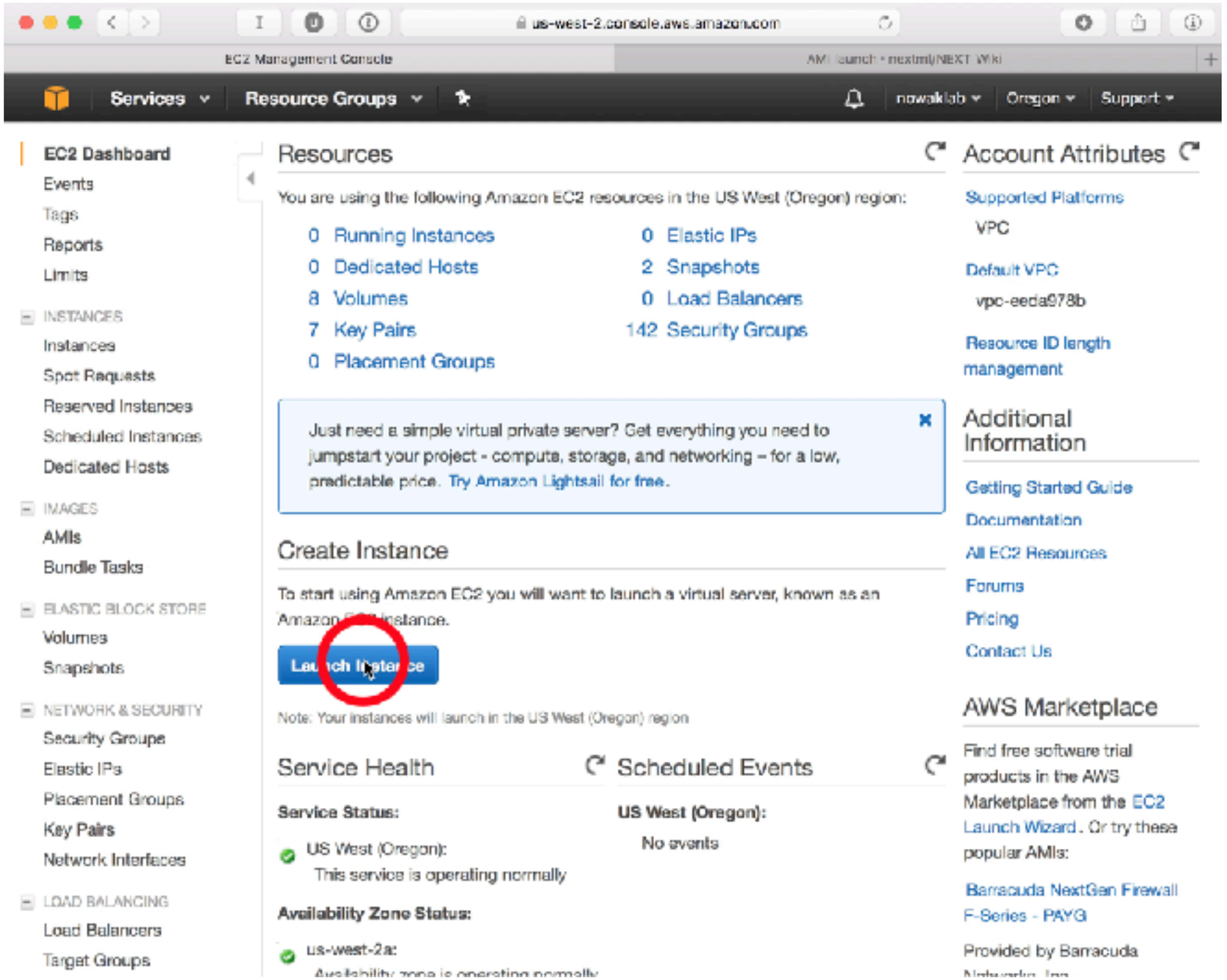
NEXT also has a REST API

https://tinyurl.com/next-pydata



https://tinyurl.com/next-pydata

Launching NEXT via Amazon EC2 AMI

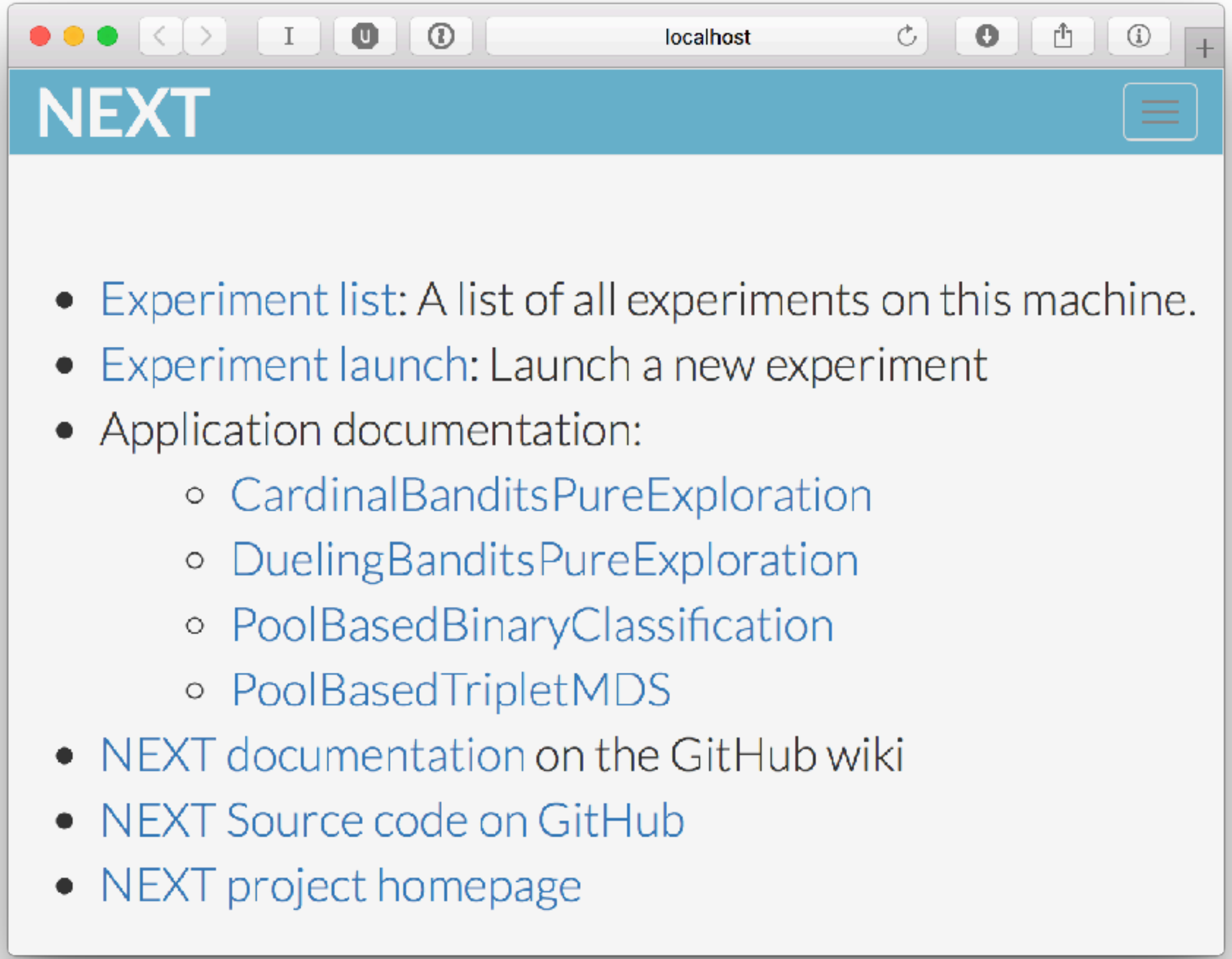


(more detail in [SciPy 2017 proceedings] and on docs)

See https://github.com/nextml/NEXT/wiki for details and more launching options

https://tinyurl.com/next-pydata

NEXT startup page at `http://[ec2-dns]:8000/home`



https://tinyurl.com/next-pydata

Start an experiment

localhost:8000/ass

Launching experiment:

Argument	Input
Experiment arguments	Browse... No file selected.
Experiment targets (optional)	Browse... No file selected.
AWS Bucket ID*	
AWS Key ID*	
AWS Secret Key*	

* = Required for target upload to Amazon S3.

Launch

Upload image (or videos/etc) targets by including many different files in the ZIP file. These are uploaded to S3 using the credentials.

Upload text or JSON targets by

- Leaving the AWS fiel
- Uploading a ZIP file t
  - CSV or TXT fil
  - JSON files are arguments

Status

Ready for file upload

- Application documentation:
  - PoolBasedTripletMDS
  - PoolBasedBinaryClassifi
  - DuelingBanditsPureExpl
  - CardinalBanditsPureExpl
- Examples on GitHub

Docs

localhost:8000/assistant/doc/Cardin

initExp(args, app_id) : None

JSON fo
exp_uid

Argume

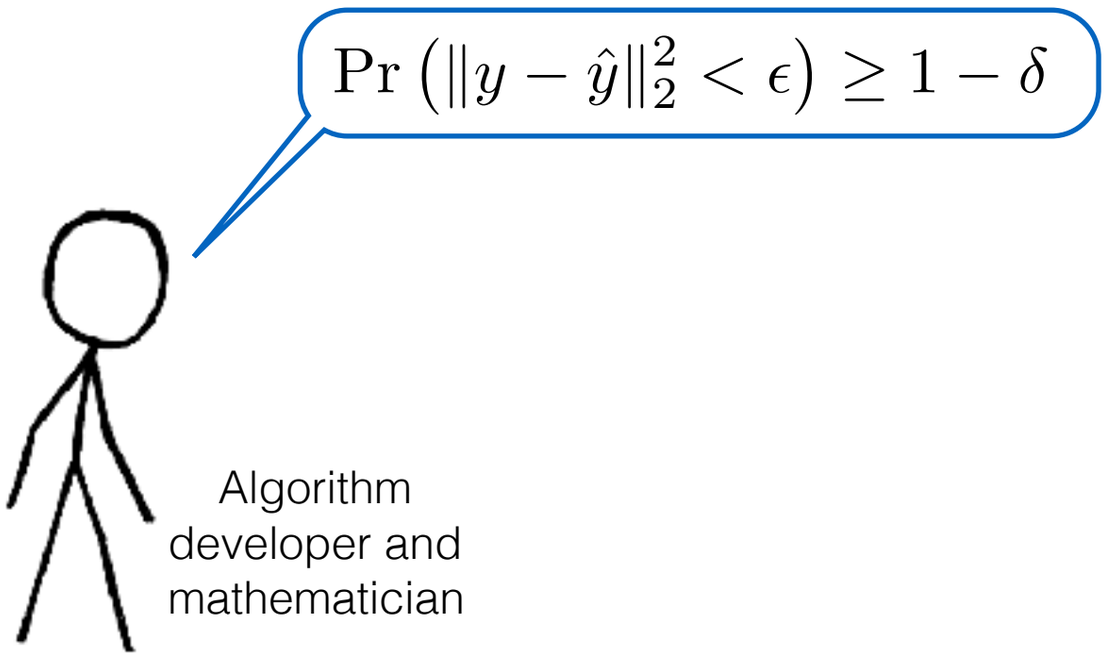
- arg

NEXT/init.yaml at master · nextr

GitHub, Inc. (US) | https://g

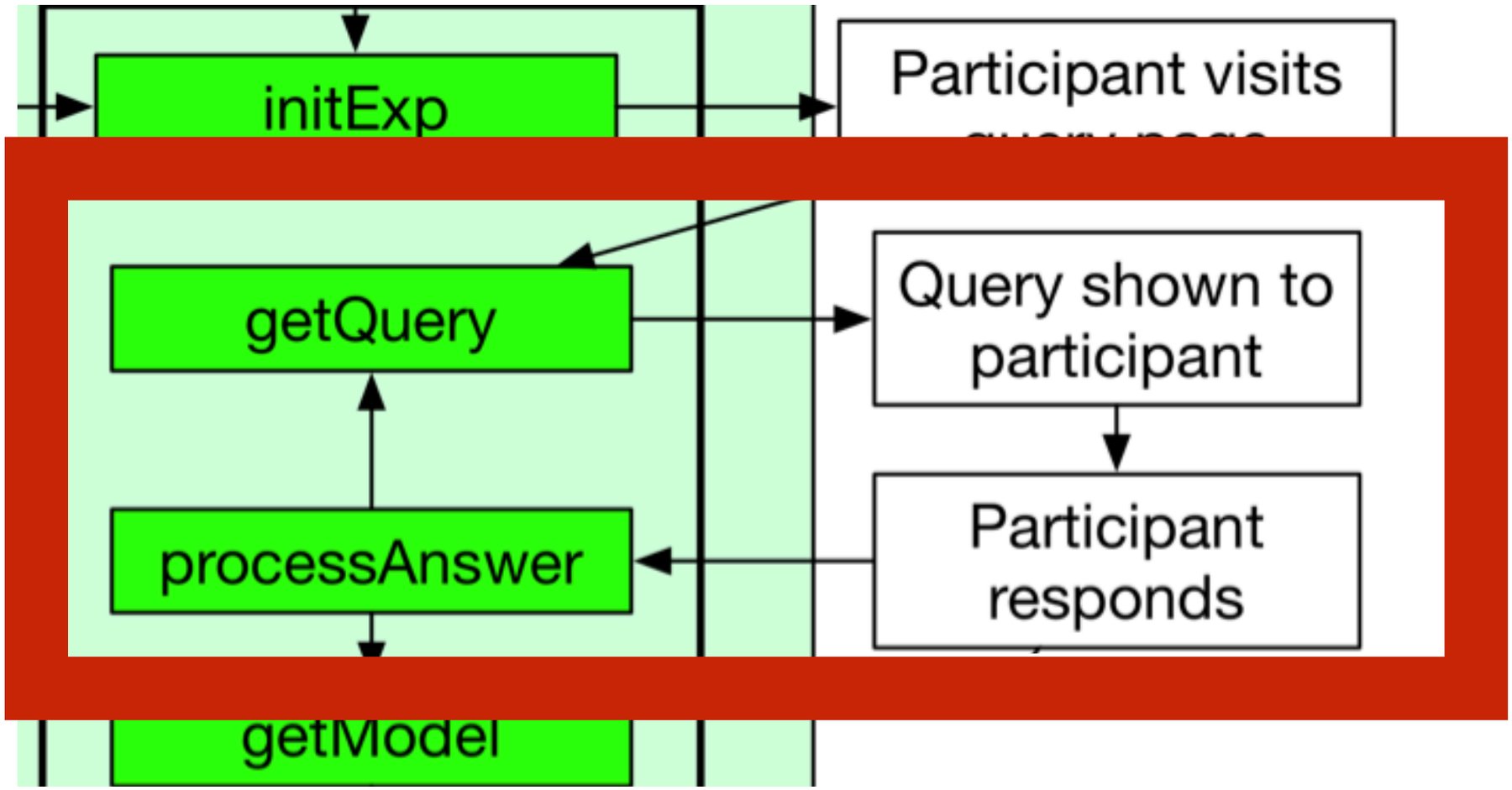
22 lines (21 sloc) | 644 Bytes

```yaml
app_id: CardinalBanditsPureExploration
args:
  R: 1.0
  alg_list:
  - {alg_id: LilUCB, alg_label: LilUCB}
  - {alg_id: RoundRobin, alg_label: RoundRobin}
  algorithm_management_settings:
    mode: fixed_proportions
    params:
    - {alg_label: LilUCB, proportion: 0.5}
    - {alg_label: RoundRobin, proportion: 0.5}
context: http://778252f0106cfd6d26f14c76a2a399.
context_type: image
failure_probability: 0.05
num_tries: 500
participant_to_algorithm_management: one_to_mar
rating_scale:
  labels:
  - {label: unfunny, reward: 1}
  - {label: somewhat funny, reward: 2}
  - {label: funny, reward: 3}
```



https://tinyurl.com/next-pydata

Adaptive data flow



https://tinyurl.com/next-pydata

What choices were made in implementation?

What have these choices enabled?

https://tinyurl.com/next-pydata

Algorithm design decisions

0. Use a high–level language (Python)
1. Treat algorithms as black boxes
    - For each function, inputs and outputs are documented and type-checked
2. Use wrapper to allow *easy* access to experiment information and background jobs
3. Objects are abstracted to integers
    - There is a mapping from integers to complete object details

(more detail in SciPy 2017 proceedings and on docs)

https://tinyurl.com/next-pydata

ec2-34-208-101-150.us-west-2.compute.amazonaws.com

NEXT Experiment Dashboard | NEXT - b85faed0ebbd9c6256bdcaa5f2951a

NEXT

EXPERIMENT SYSTEM

Histogram of time respon

Client-side timing

KLUCB

KLUCB - response time

KLUCB - network delay

Algorithm timing comparison

getQuery

processAnswer

predict

Individual algorithm timing breakdown

BR_LilUCB_b2 BR_Thompson_b2

BR_LilUCB_b2 - getQuery

BR_LilUCB_b2 - processAnswer

BR_LilUCB_b2 - predict

compute dbGet dbSet admin enqueued

Test Error Plot

Triplet Test Error

CrowdKernel
RandomSampling
UncertaintySampling
Test

Error on hold out set

0.70
0.65
0.60
0.55
0.50

5
10
15

Number of answered triplets

.0123 3864 1193 1602 1069

do you ask?

https://tinyurl.com/next-pydata

Use cases enabled by this implementation

Psychology triplets

Select expression on the bottom most similar to the face on top

Image search

Starting image:

Is this the kind of image you are looking for?

No Yes

Images rated: 1

Street Score

Select the street that looks safer

https://tinyurl.com/next-pydata

Psychology triplets

Problem: use humans to generate “similarity” map of facial emotions

https://tinyurl.com/next-pydata

People

Prof. Robert Nowak
ECE

Prof. Tim Rogers
Psychology

Kevin Jamieson
ECE

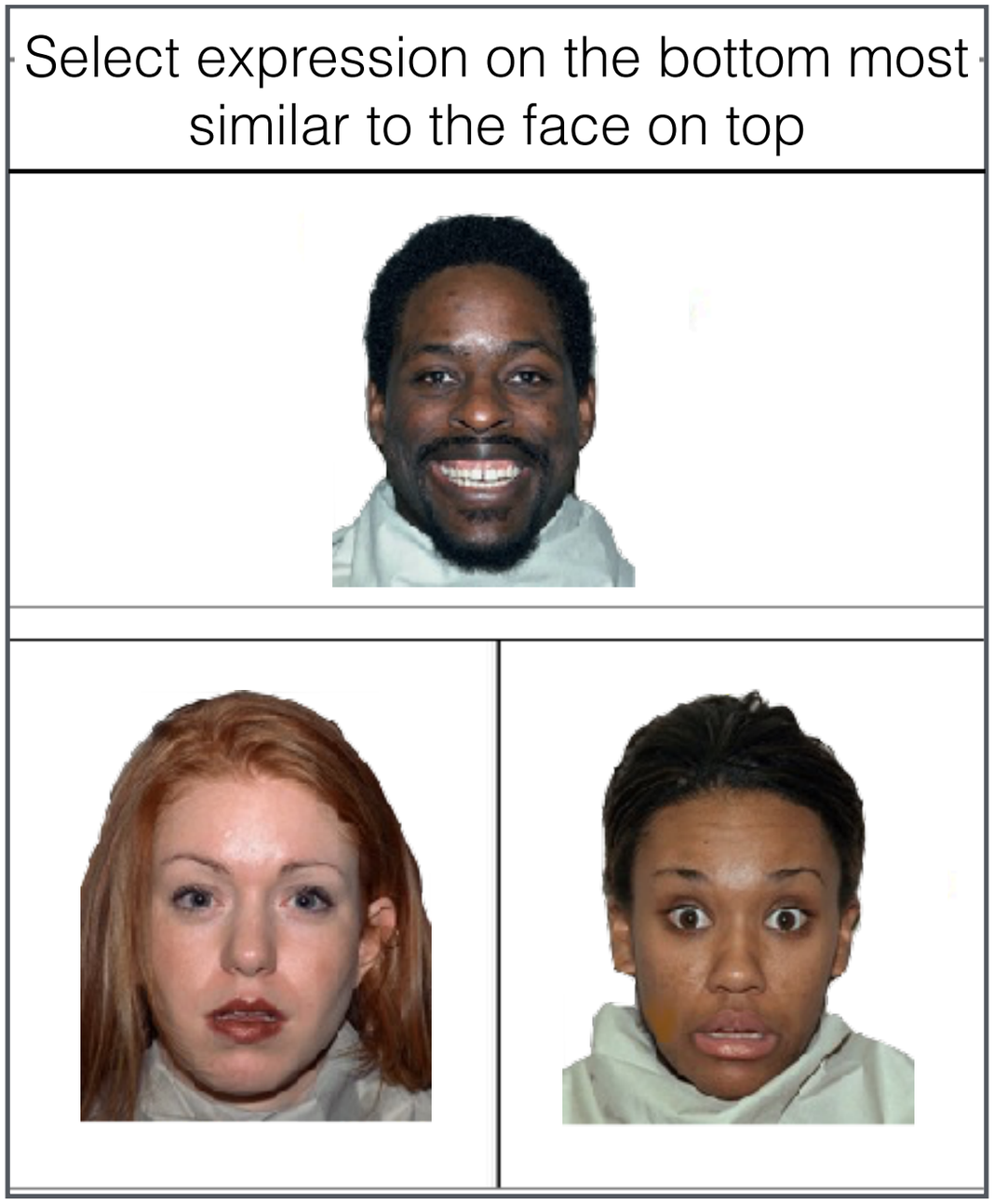
Lalit Jain
Math

April Murphy
Psychology

https://tinyurl.com/next-pydata

Psychology triplets

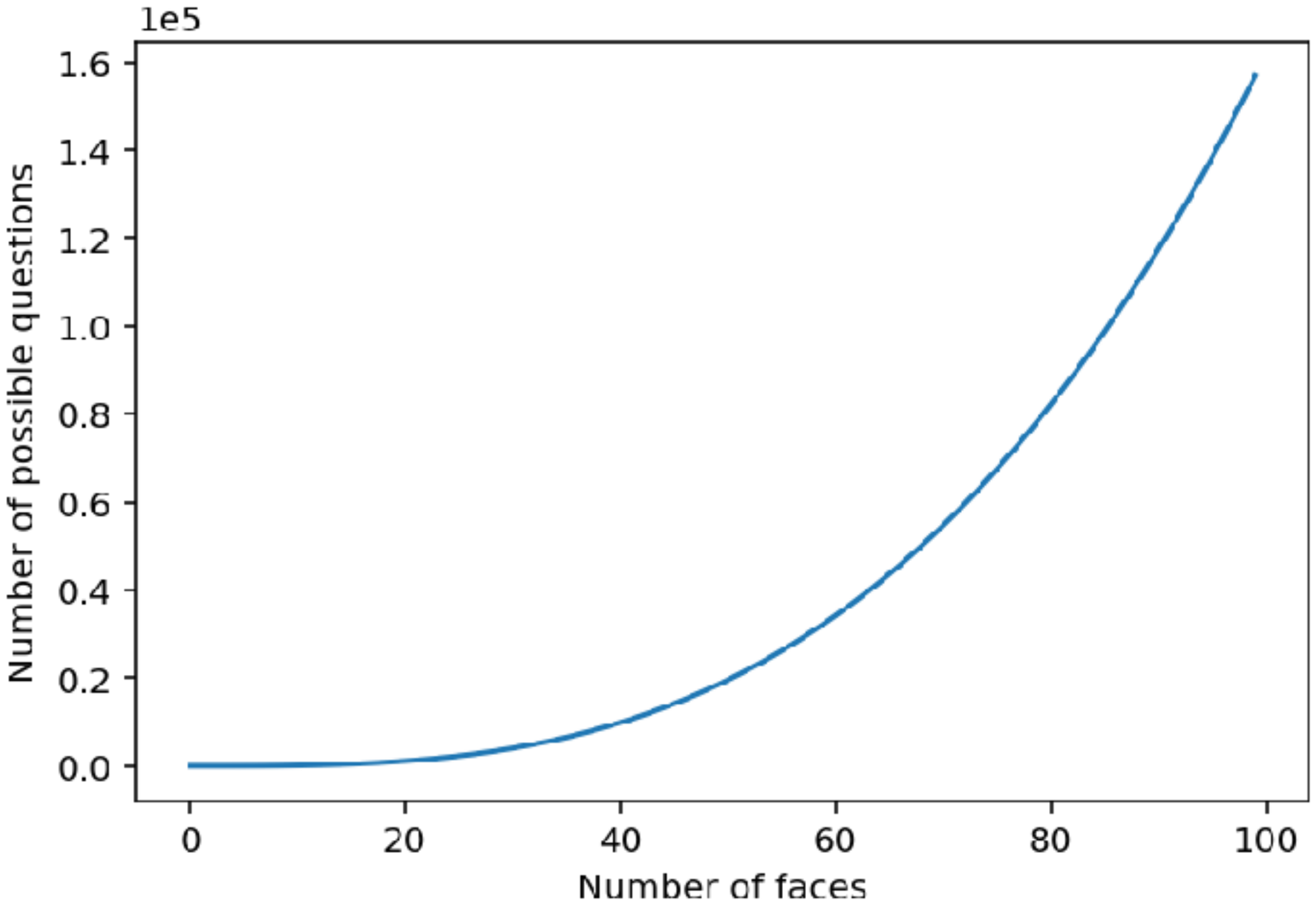
Best question for similarity map:



https://tinyurl.com/next-pydata

Psychology triplets

Number of questions scale poorly $(\approx n^3)$



Finding the *best* questions gets more difficult

https://tinyurl.com/next-pydata

Low-dimensional embedding using adaptively selected ordinal data

KG Jamieson, RD Nowak - Communication, Control, and …, 2011 - ieeexplore.ieee.org

Abstract: Low-dimensional embedding based on non-metric data (eg, non-metric multidimensional scaling) is a problem that arises in many applications, especially those involving human subjects. This paper investigates the problem of learning an embedding of n objects into d-dimensional Euclidean space that is consistent with pairwise comparisons of the type" object a is closer to object b than c." While there are O (n 3) such comparisons …

☆ 99 Cited by 29 Related articles All 9 versions

Finite sample prediction and recovery bounds for **ordinal embedding**

L Jain, KG Jamieson, R **Nowak** - Advances in Neural Information …, 2016 - papers.nips.cc

Abstract The goal of ordinal embedding is to represent items as points in a low-dimensional Euclidean space given a set of constraints like``item $ i $ is closer to item $ j $ than item $ k $". Ordinal constraints like this often come from human judgments. The classic approach to

☆ 99 Cited by 9 Related articles All 7 versions

[PDF] Learning Low-Dimensional Metrics

L **Jain**, B Mason, R **Nowak** - Ann Arbor, 2017 - papers.nips.cc

Abstract This paper investigates the theoretical foundations of metric learning, focused on three key questions that are not fully addressed in prior work: 1) we consider learning general low-dimensional (low-rank) metrics as well as sparse metrics; 2) we develop upper

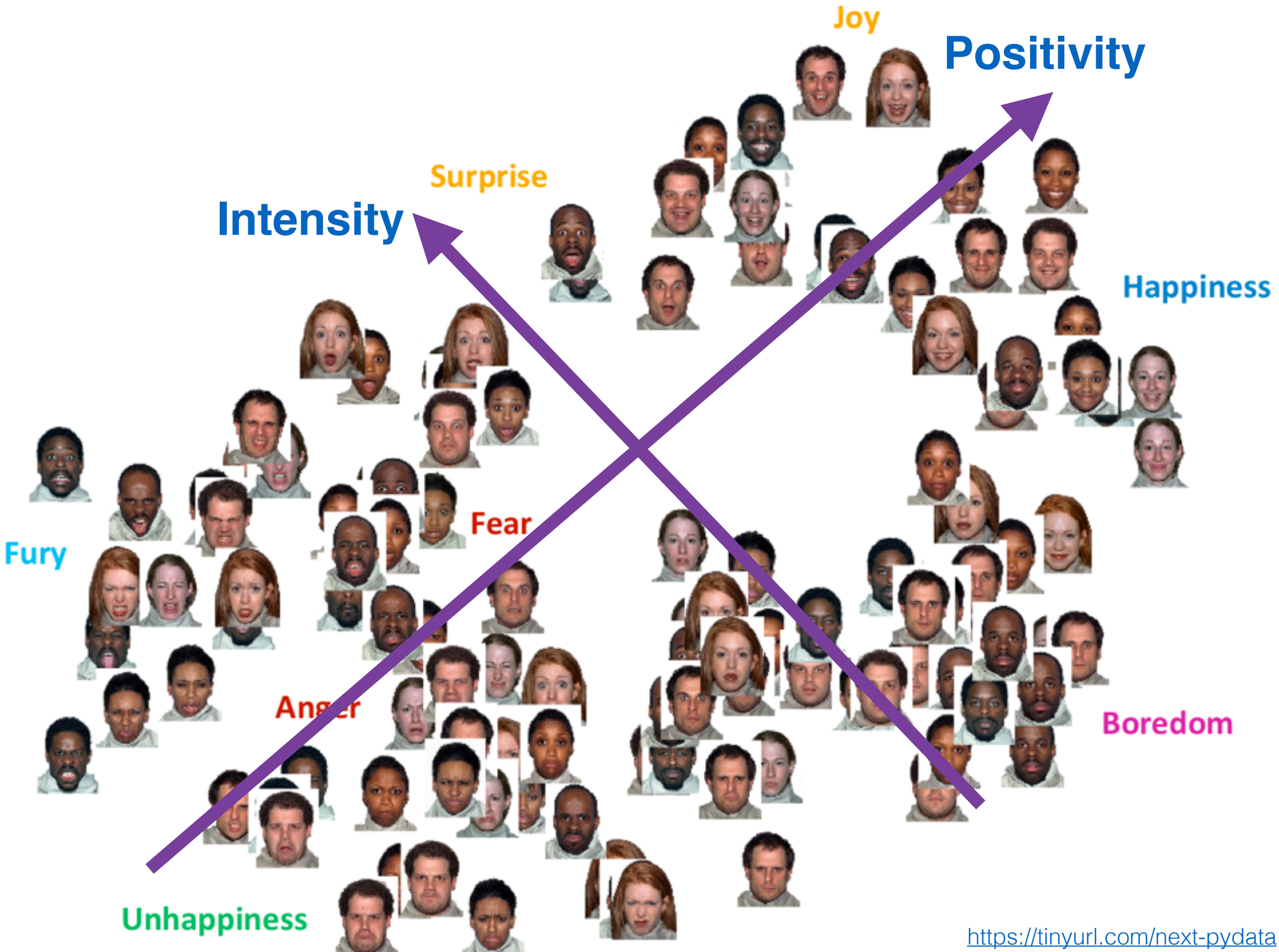
☆ 99 All 4 versions »

[PDF] **How** to **Model** Implicit **Knowledge**? **Similarity Learning** Methods to Assess Perceptions of Visual Representations.

MA Rau, B **Mason**, RD Nowak - EDM, 2016 - educationaldatamining.org

ABSTRACT To succeed in STEM, students need to learn to use visual representations. Most prior research has focused on conceptual knowledge about visual representations that is acquired via verbally mediated forms of learning. However, students also need perceptual fluency: the ability to rapidly and effortlessly translate among representations. Perceptual fluency is acquired via nonverbal, implicit learning processes. A challenge for …

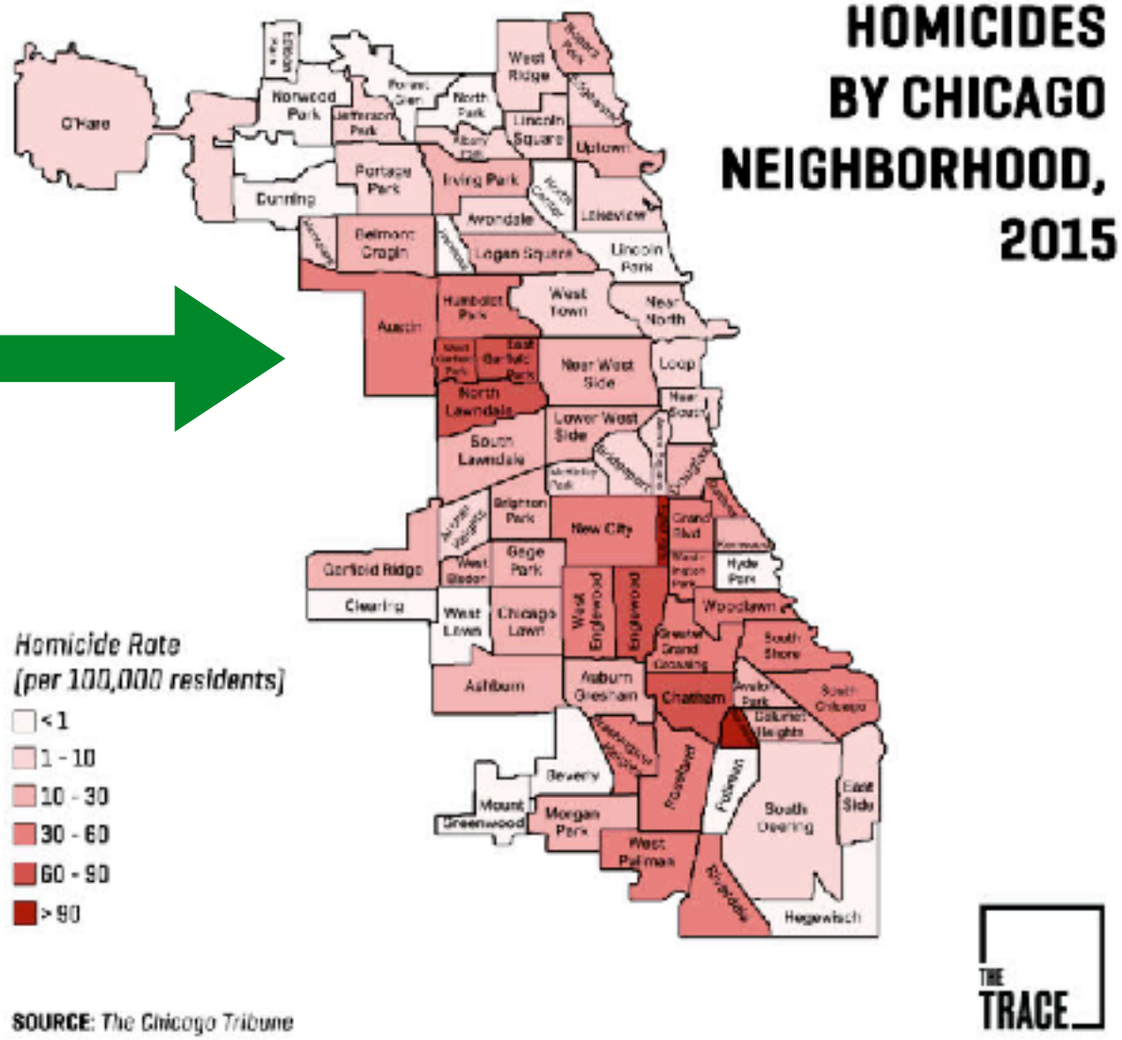
☆ 99 Cited by 4 Related articles All 6 versions »



https://tinyurl.com/next-pydata

Street score

Goal: find humans perception of
the safest streets in Chicago



https://tinyurl.com/next-pydata

Street score

Question: pairwise comparisons

These use Google StreetView images

https://tinyurl.com/next-pydata

Street score

Issues

- Clustered ranking

Advantages

Similar to sorting

safe



somewhat safe



unsafe


https://tinyurl.com/next-pydata

Paper (@AISTATS 2018)

Adaptive Sampling for Clustered Ranking

Sumeet Katariya
katariya@wisc.edu

Lalit Jain
lalitj@umich.edu

Nandana Sengupta
nandana@uchicago.edu

James Evans
jevans@uchicago.edu

Robert Nowak
rdnowak@wisc.edu

[Sumeet Katariya](#)
ECE

[Lalit Jain](#)
Math

[Nandana Sengupta](#)
Sociology

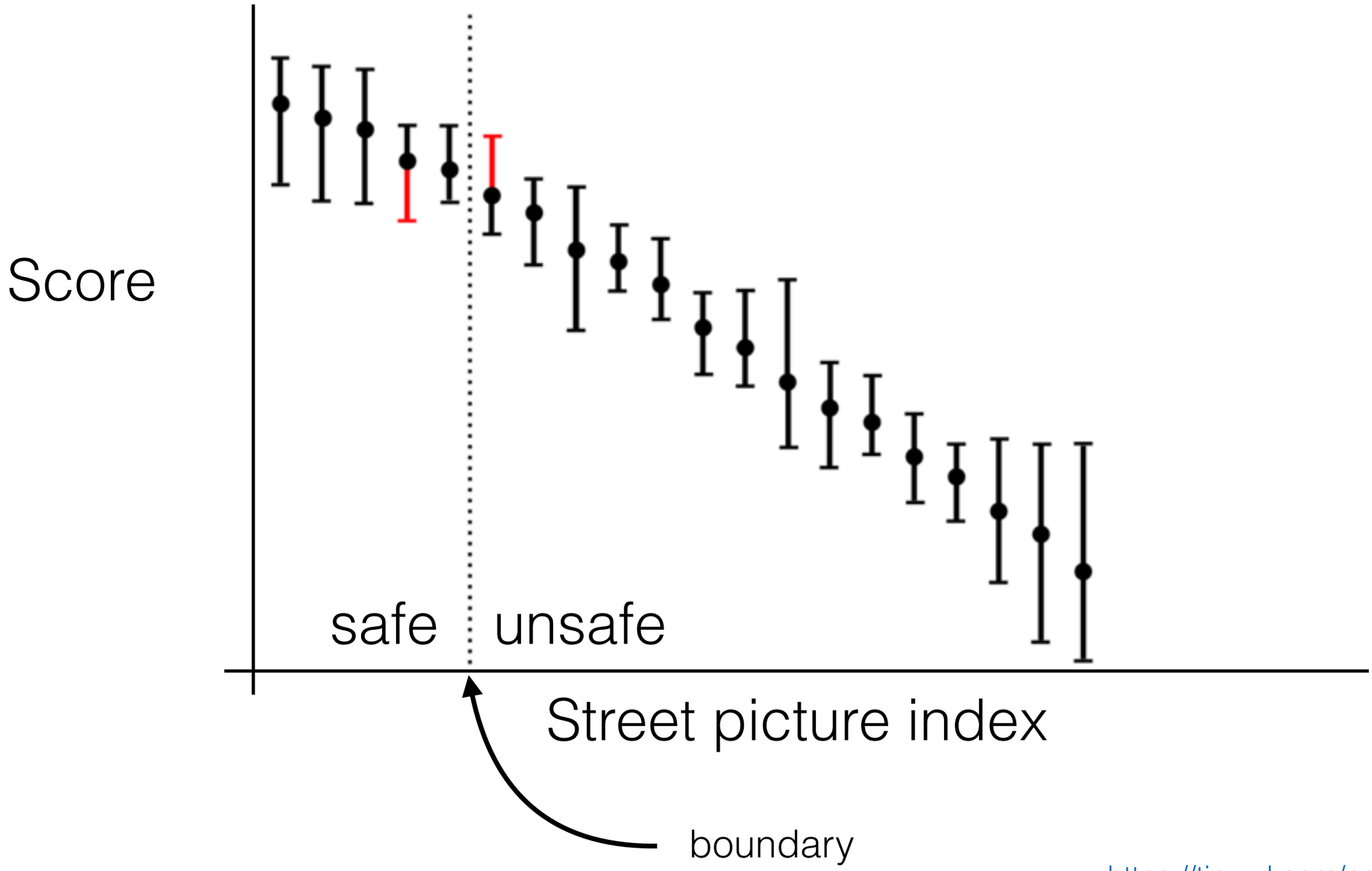
[Prof. James Evans](#)
Sociology

[Prof. Robert Nowak](#)
ECE

https://tinyurl.com/next-pydata

Layman's adaptive algorithm

`getQuery`: choose streets where *most uncertain* about which cluster the street belongs too

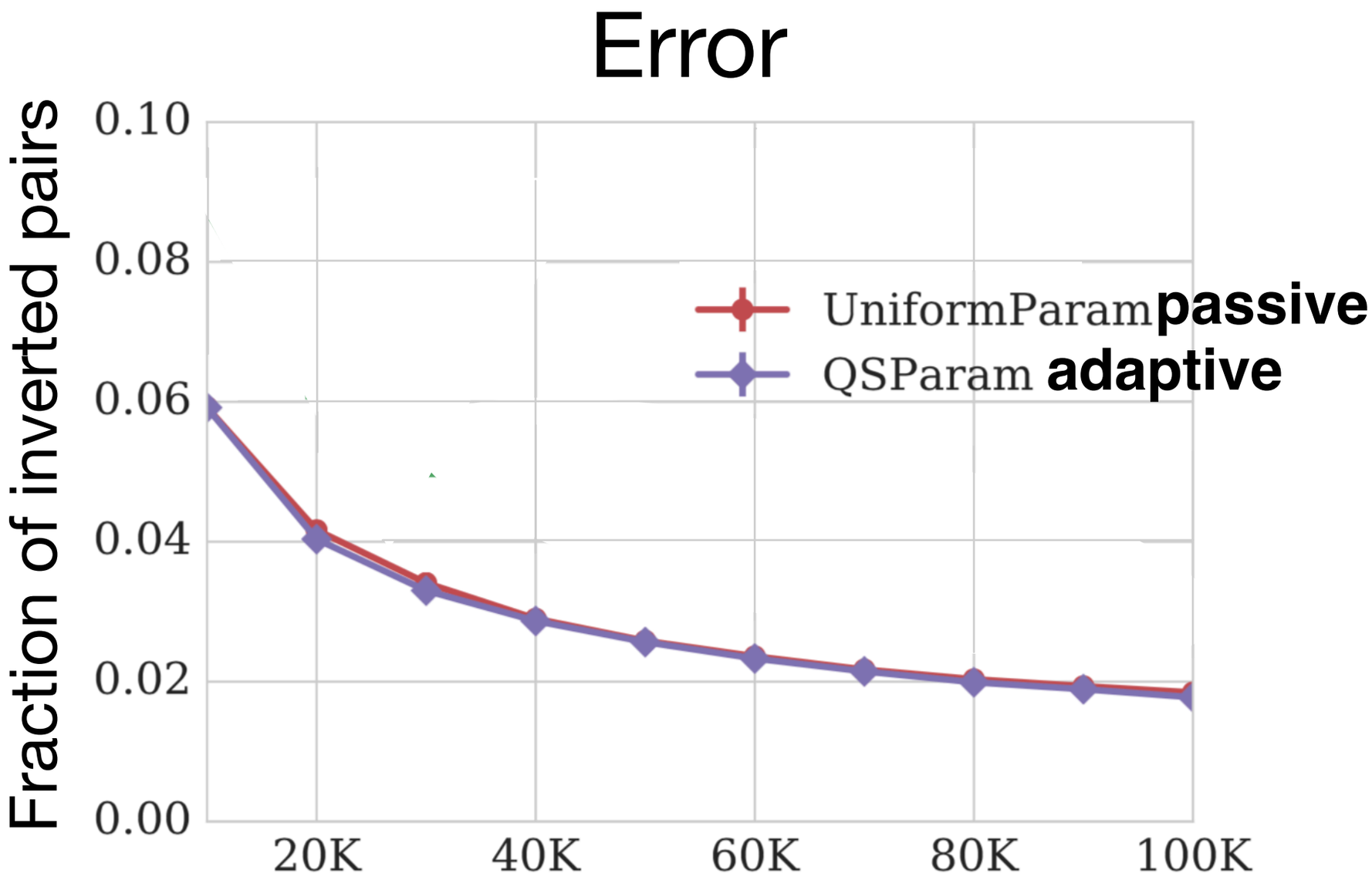


https://tinyurl.com/next-pydata

Results

Actual: 

Predicted: 



No gains from active learning!

Why? Crowdsourcing responses are too noisy

https://tinyurl.com/next-pydata

But if we dig a little deeper with simulations…

Actual: Safe Unsafe

Predicted: Safe Unsafe

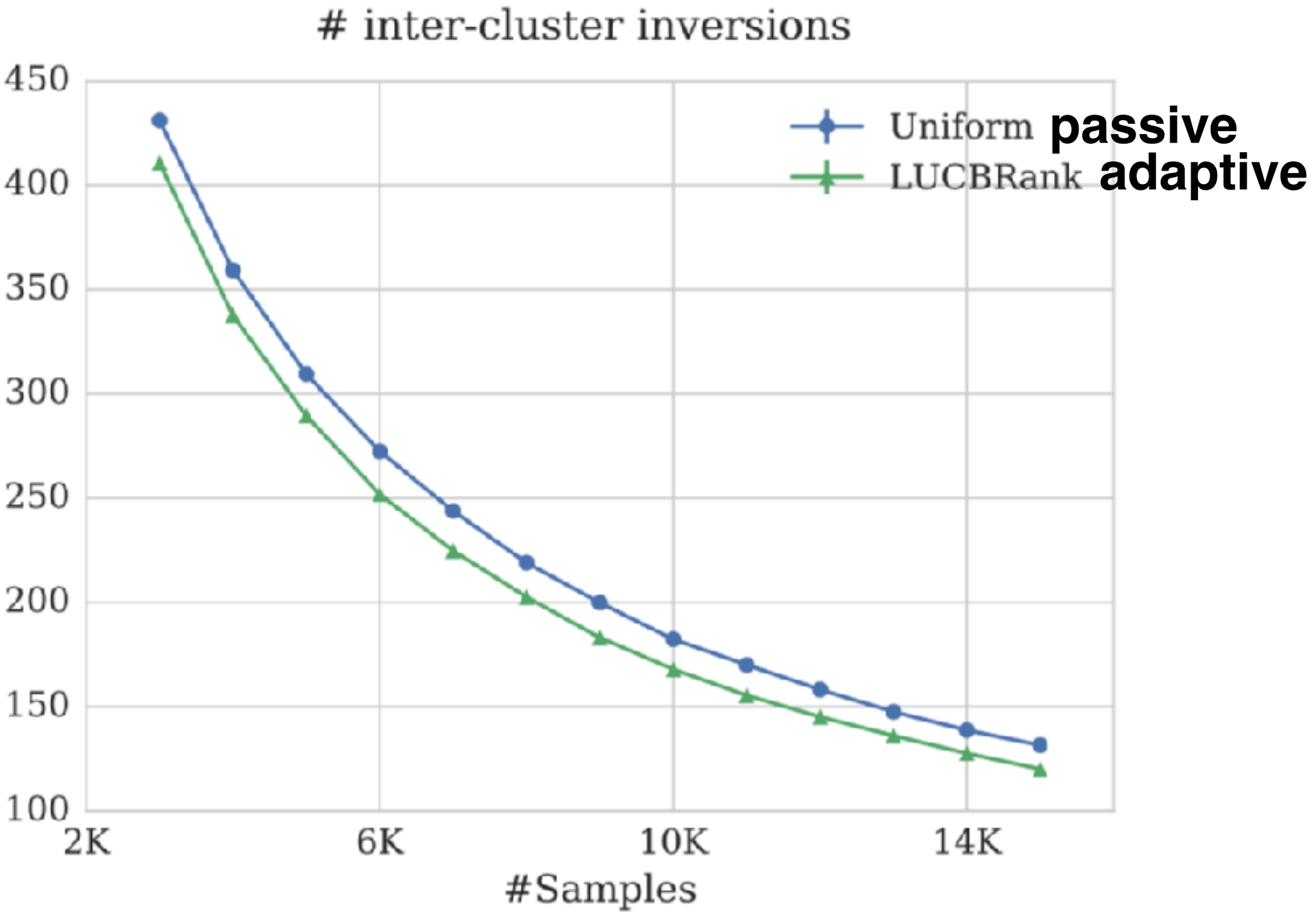
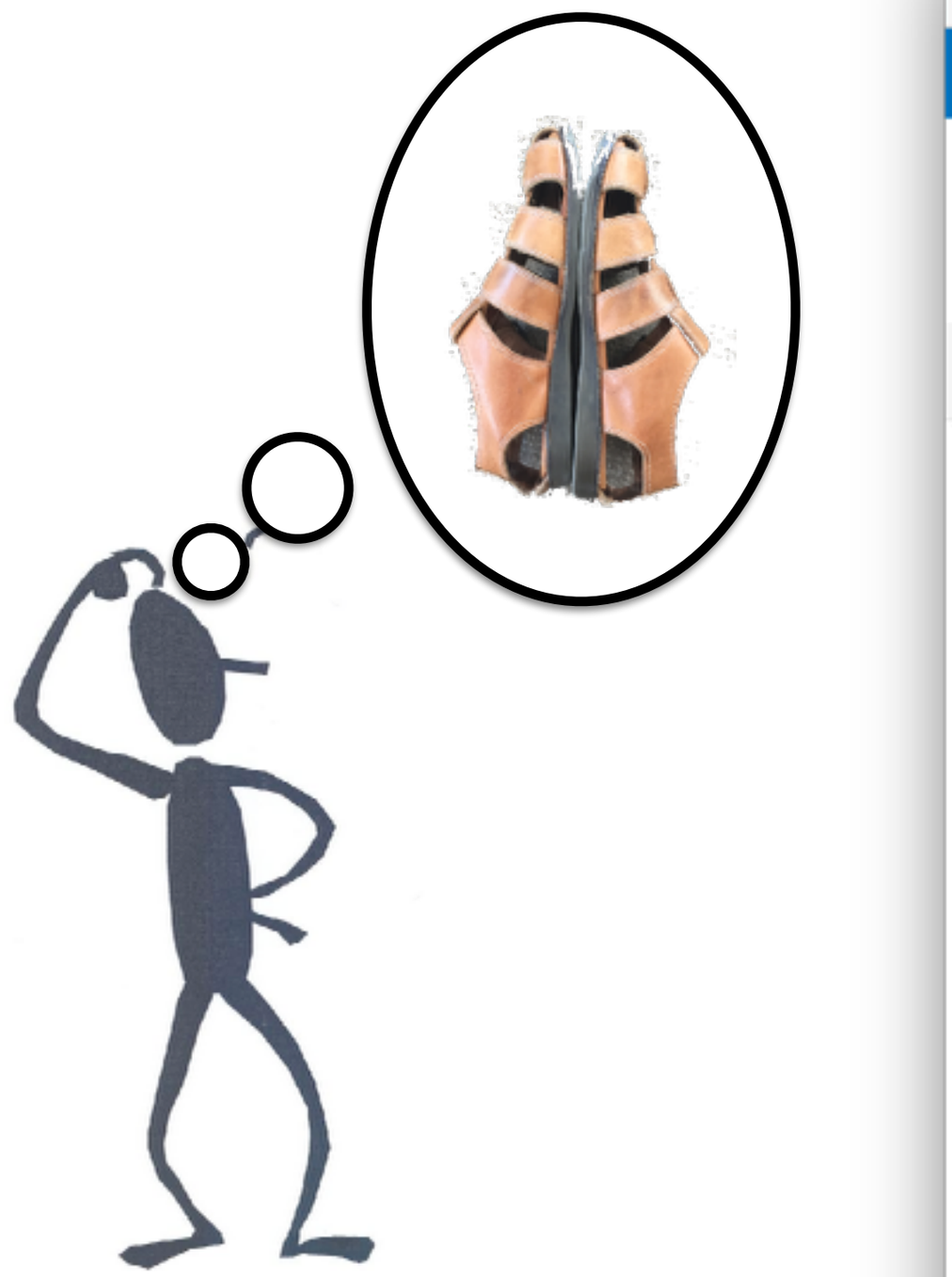
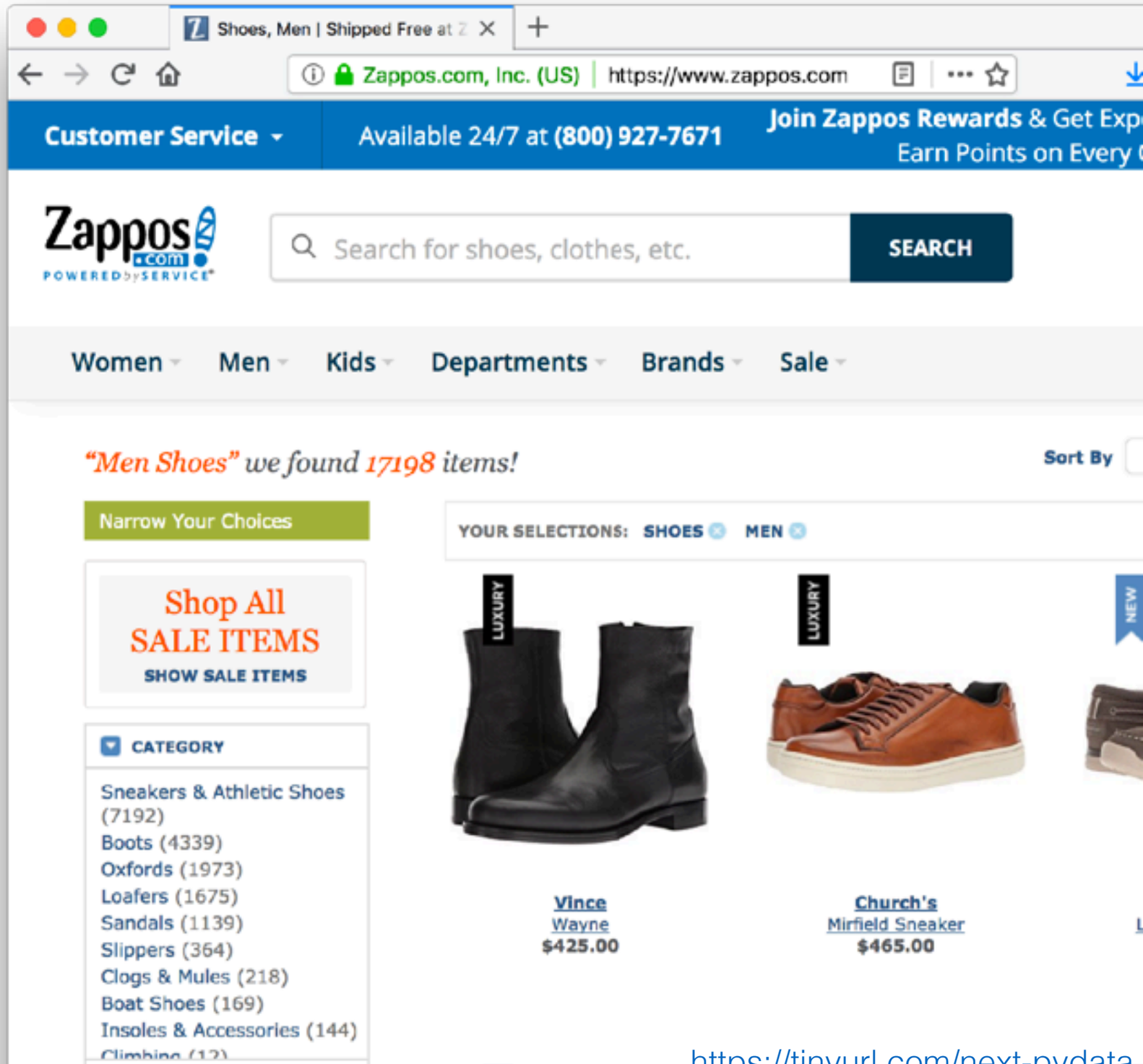


Image search

Problem: find the shoes you're interested in





https://tinyurl.com/next-pydata

Image search

Question: Pick initial shoe than rate similar shoes

In this experiment, we will show you a total of 50 images. For each image, you will be asked if it is similar to the image currently shown. To make your judgement, please look at the image and read the description below. Click on the image when you are ready to proceed. Allow for 15-20 seconds after clicking the initial image.

Pick red boots

Starting image:

Is this the kind of image you are looking for?

No | Yes

Images rated: 1

https://tinyurl.com/next-pydata

Image search problems

- Large dataset from Zappos
  - 50,000 shoes with 1,000 features each)

```
[f(shoe,👢,👍) for shoe in [
```

]

↑
entire dataset

https://tinyurl.com/next-pydata

Paper

Scalable Generalized Linear Bandits: Online Computation and Hashing

KS Jun, A Bhargava, R Nowak, R Willett - arXiv preprint arXiv:1706.00136, 2017 - arxiv.org

Abstract: Generalized Linear Bandits (GLBs), a natural extension of the stochastic linear bandits, has been popular and successful in recent years. However, existing GLBs scale poorly with the number of rounds and the number of arms, limiting their utility in practice.

☆ 99 Related articles

[Kwang-Sung Jun](#)
ECE

[Aniruddha Bhargava](#)
ECE

[Prof. Robert Nowak](#)
ECE

[Prof. Becca Willett](#)
ECE

https://tinyurl.com/next-pydata

Image search solution

exhaustively evaluate shoes

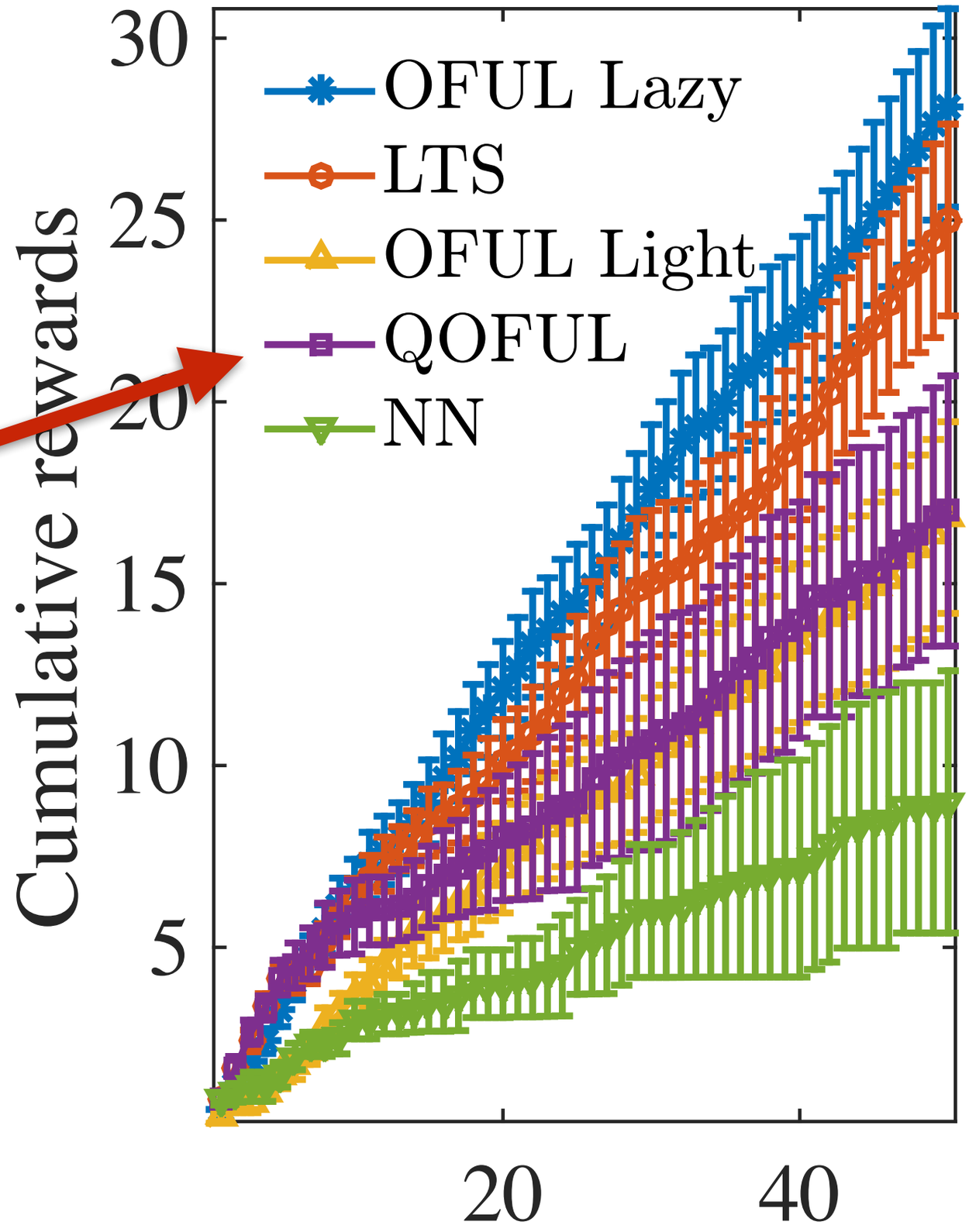
↓

selectively evaluate shoes

```
[f(shoe,👢,👍) for shoe in [👠,👡,👢]]
```

https://tinyurl.com/next-pydata

Their method



https://tinyurl.com/next-pydata

Software improvements

- Internal redesign
- {Experimentalist, developer} ease of use
- Fast in-memory database support with Redis
- Documentation improvements
- Many bugs fixes

Key messages

1. Adaptive sampling reduces data collection cost.

2. NEXT is a crowdsourcing data collection tool that can use adaptive sampling techniques

3. NEXT is easy* to use by experimentalists, algorithm developers and practitioners, and a mathematical background is not required.

4. NEXT developers use experimentalist engagement to aid research and to gain feedback to improve the software

* NEXT has been created by an academic research group in collaboration with psychologists https://tinyurl.com/next-pydata

Questions?

Thank you!

https://xkcd.com/1543/

https://tinyurl.com/next-pydata

Extras…

Algorithm inputs and outputs

- Documented exactly in `apps/[app-id]/algs/Algs.yaml`

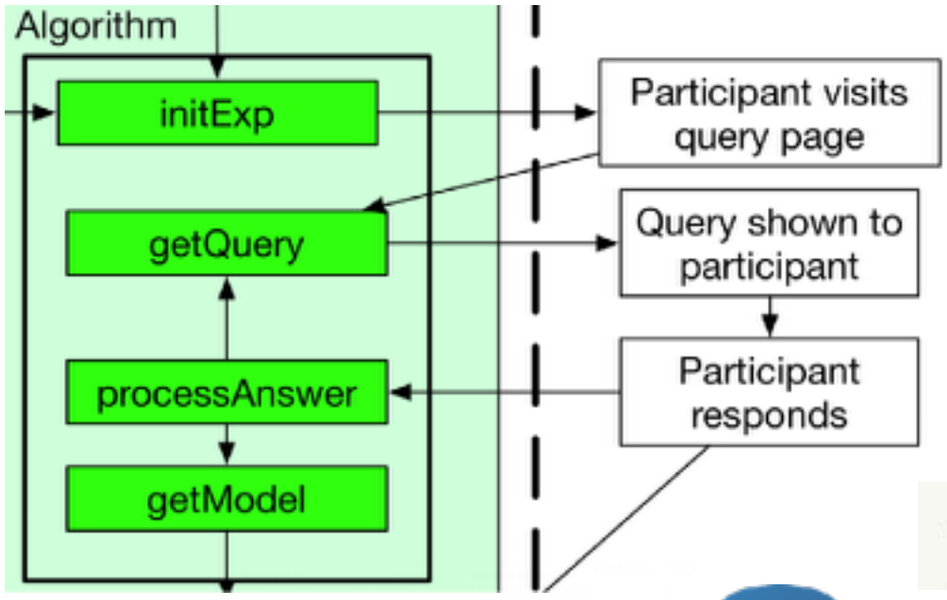
```
getQuery:
  args:
    participant_uid:
      type: string
      description: ID of the participant answering the query
  rets:
    description: The index of the target to ask about
    type: num
```

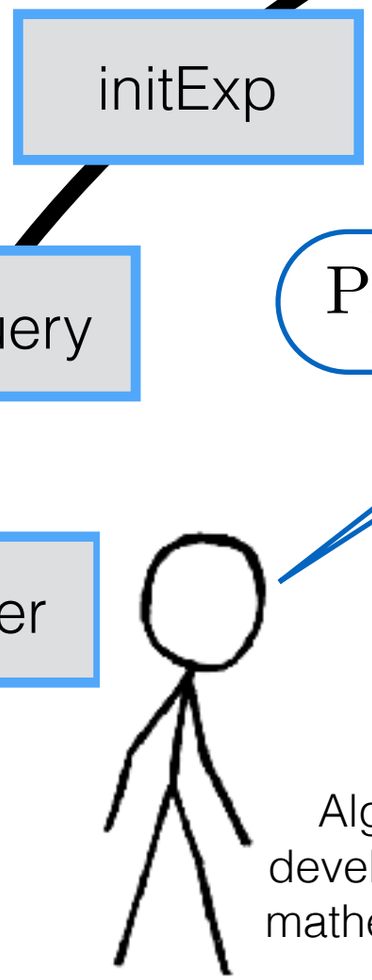
- Function implementation

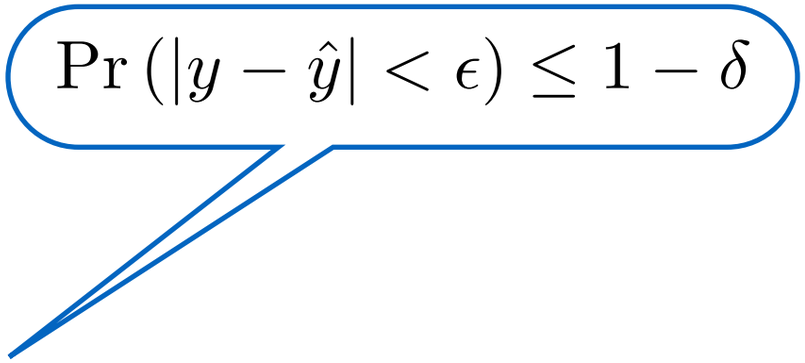
```
import random

def getQuery(self, butler, participant_uid):
    n = butler.algortithms.get(key='n')
    return random.choice(n)
```

Depends on a library we developed:
https://github.com/daniel3735928559/pijemont

NEXT stack



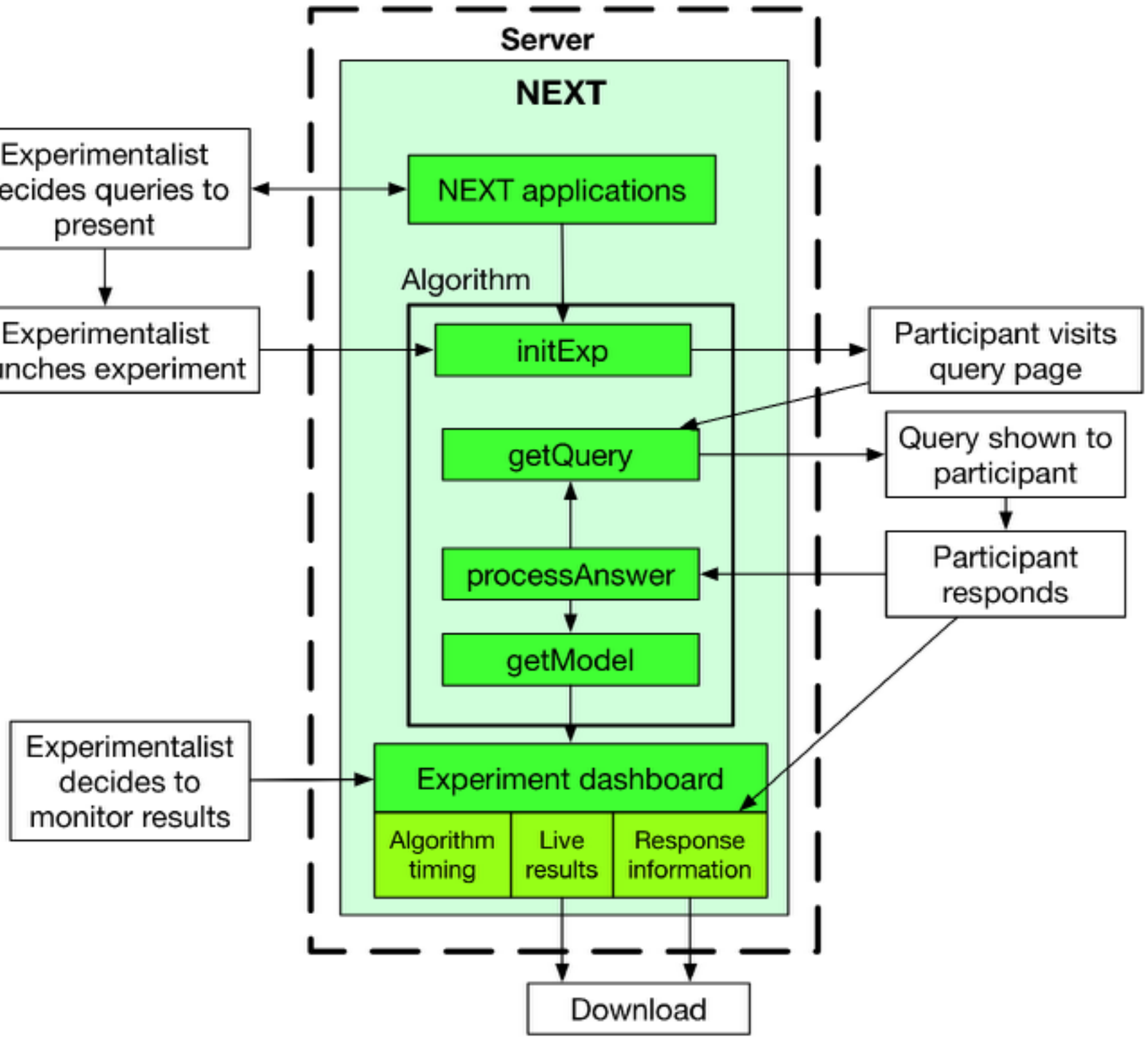




Algorithm developer and mathematician

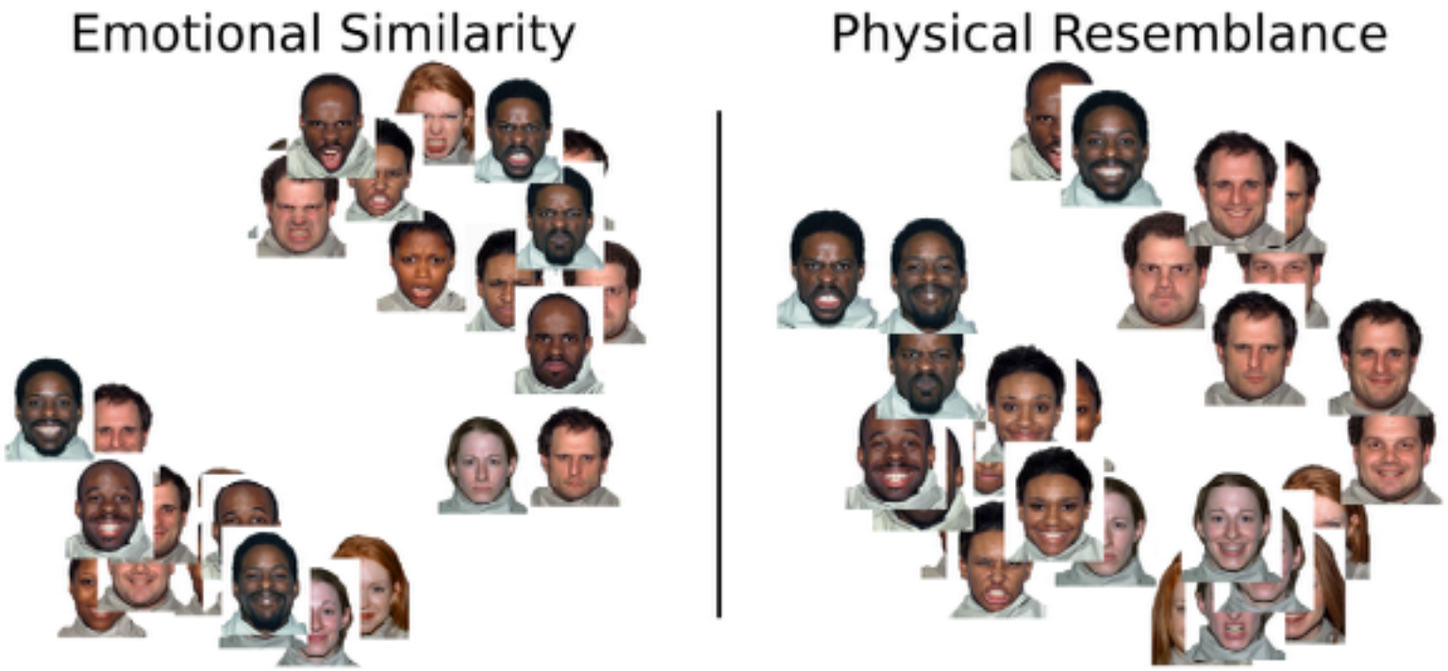
tinyurl.com/scipy-next

NEXT diagram



tinyurl.com/scipy-next

After NEXT link sent to crowdsourcing service, results can be generated!

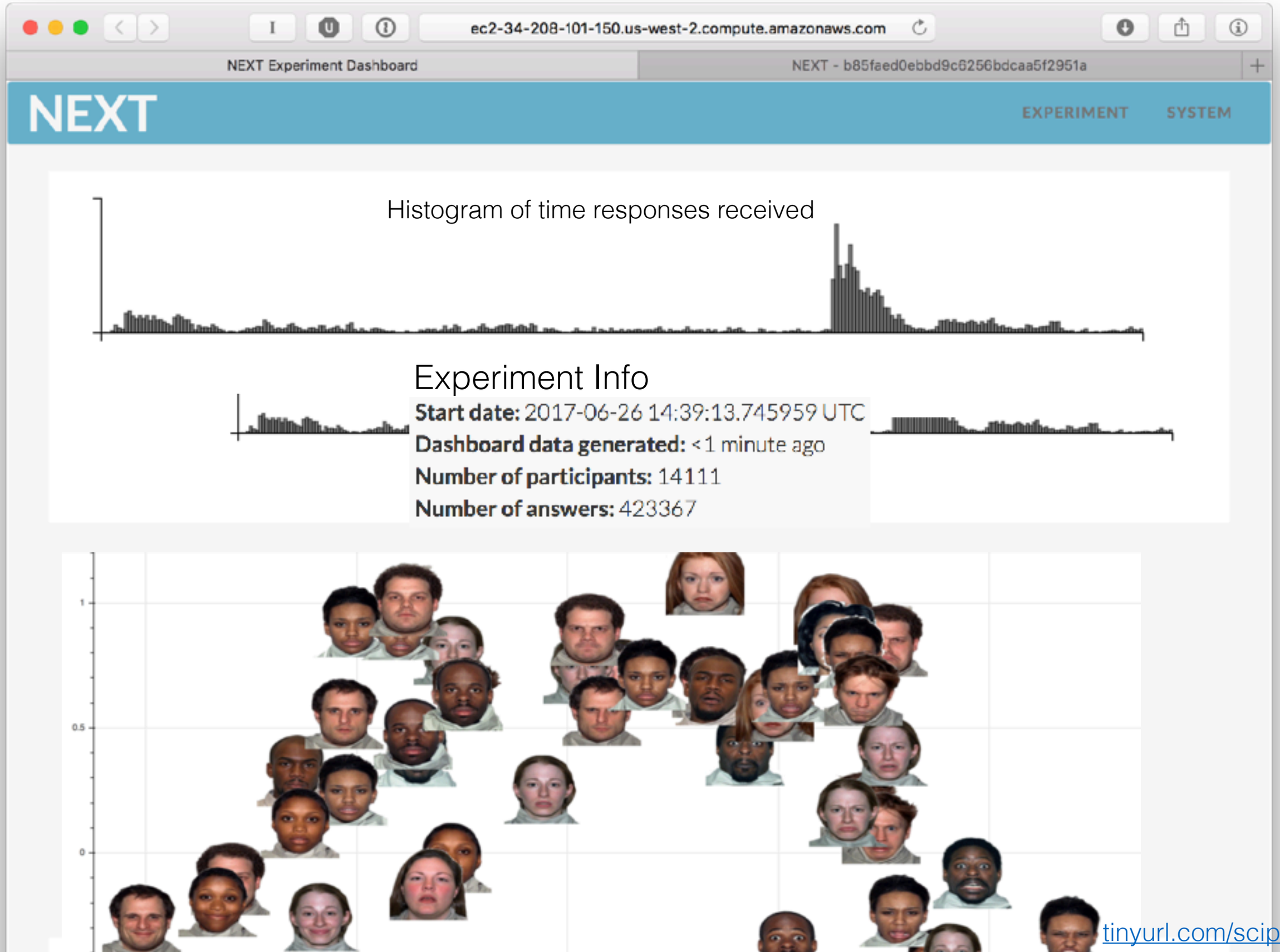


Result requirements

0. Web browser
1. Amazon AWS account
2. ZIP of targets (e.g., images)
3. Experiment description (which has good documentation!)

More detail on documentation:
https://github.com/nextml/NEXT/wiki

Dashboard



tinyurl.com/scipy-next