

Can big data tell us what clinical trials don't?

Learning objectives

1. **Behold!** Big data is already impacting clinical care & research.
2. **Understand** concepts behind two machine learning techniques.
3. **Accept** our responsibility to make sense of big data.



LIFE



STANFORD
HOSPITAL & CLINICS

Topics

Subscribe

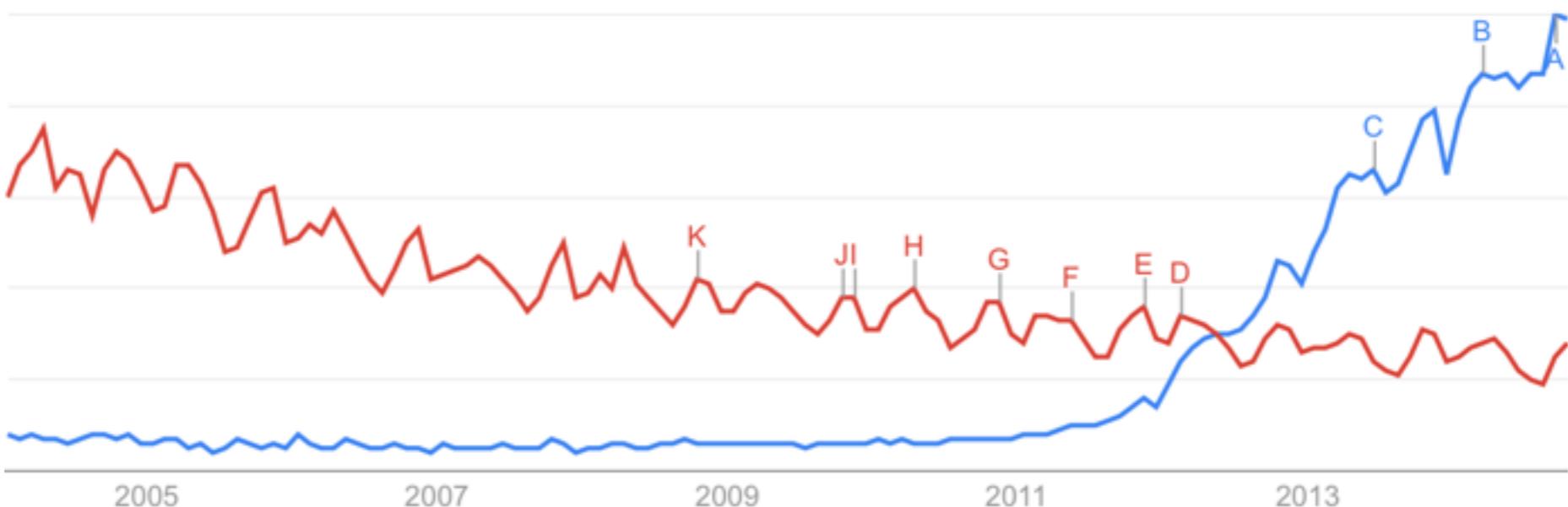


big data
Search term

kinase
Search term

+ Add term

Interest over time



Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC II)

PHILIPS



Beth Israel Deaconess
Medical Center



A teaching hospital of
Harvard Medical School

MIMIC II

2001



2008

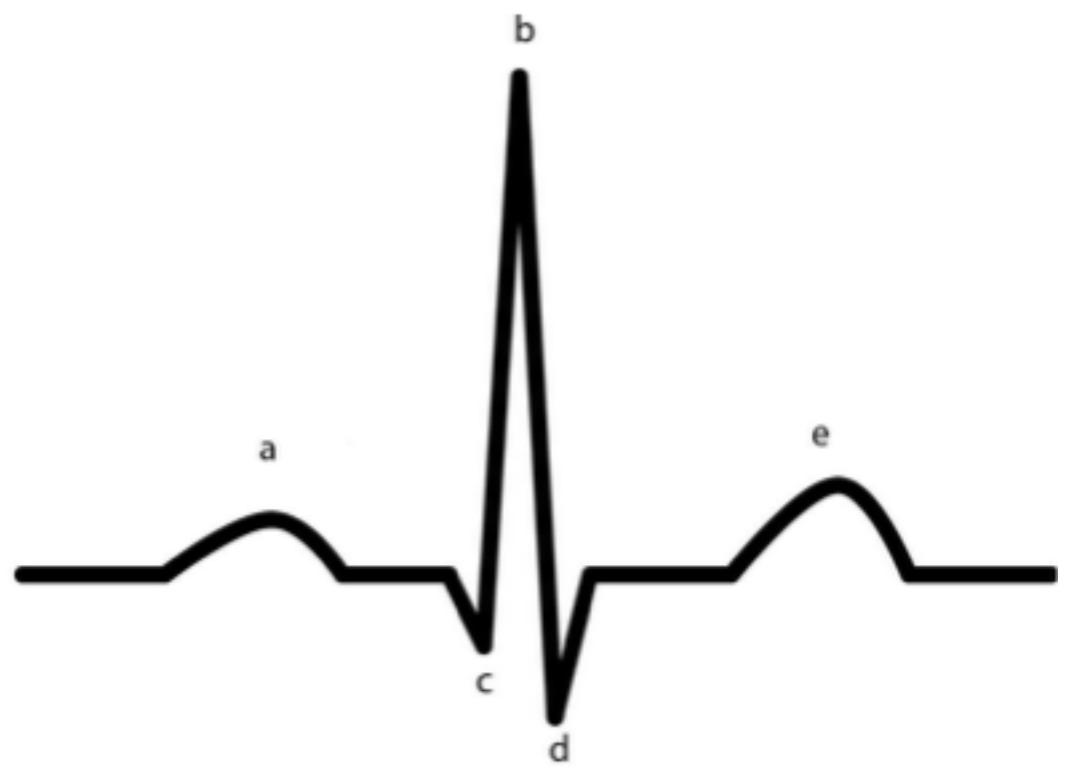
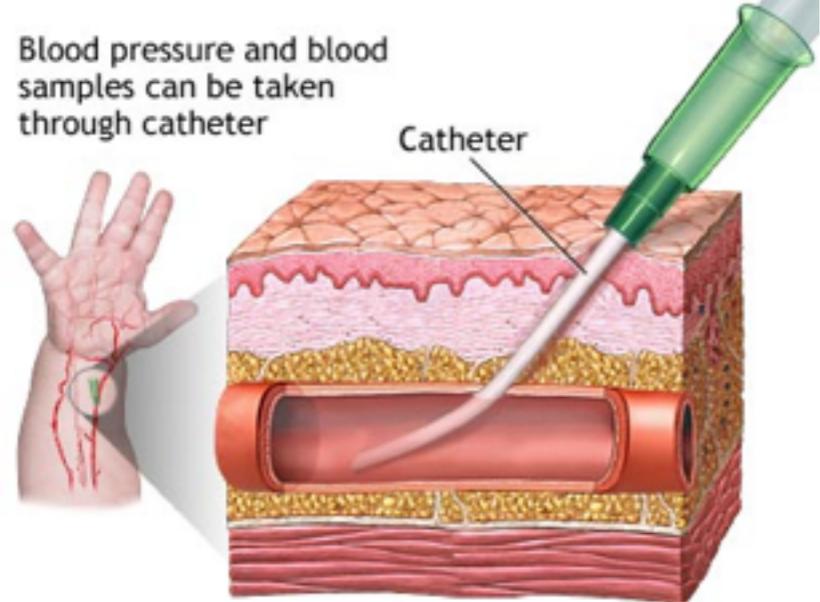


26,870 adult hospital admissions

MIMIC II data helps investigators develop clinical algorithms

Patient demographics
IV Rx drip rates
Lab values
Vital signs

De-identified & open source



RESEARCH

Open Access

An investigation of patterns in hemodynamic data indicative of impending hypotension in intensive care

Joon Lee^{1,2*}, Roger G Mark^{1,2}

Seely et al. Critical Care 2014, **18**:R65
<http://ccforum.com/content/18/2/R65>



RESEARCH

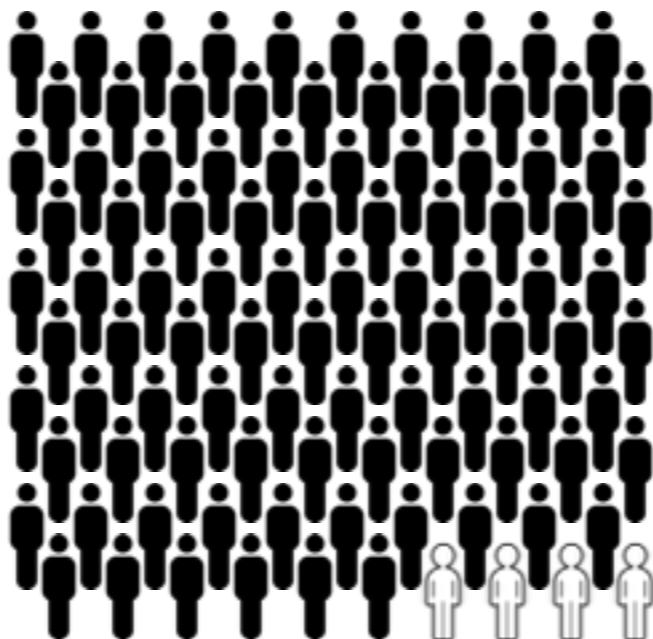
Open Access

Do heart and respiratory rate variability improve prediction of extubation outcomes in critically ill patients?

Andrew JE Seely^{1,2,11*}, Andrea Bravi², Christophe Herry¹, Geoffrey Green¹, André Longtin², Tim Ramsay¹, Dean Fergusson¹, Lauralyn McIntyre¹, Dalibor Kubelik¹, Donna E Maziak¹, Niall Ferguson³, Samuel M Brown⁴, Sangeeta Mehta⁵, Claudio Martin⁶, Gordon Rubenfeld⁷, Frank J Jacono⁸, Gari Clifford⁹, Anna Fazekas¹ and John Marshall¹⁰

FLATIRON

Flatiron's mission is to
organize the world's oncology data
and **make it useful** for patients,
physicians, life sciences & researchers.



4%

Current clinical data infrastructure **sucks**

The screenshot shows the Epic Clinicals software running in Internet Explorer. The top navigation bar includes links for File, Edit, View, Favorites, Tools, Help, Convert, Select, Atrius Health, and Patient Log Out. The main header displays the patient's name, Xbialidocious, Fuzzy*, age 5 yrs, sex M, DOB 1/1/06, MRN 70450825, Allergies (Penicillins, Nuts, Cats, N*), Pt Type TEST PAT, LOC MILLER, JAMES, CAMBRIDGE, MyChart On, and Close patient record.

The interface is divided into several sections:

- Patient Snapshot:** Contains the Problem List (DM, Down's syndrome, etc.), Allergies/Contraindications (Penicillins, Nuts, Cats, NSAIDS, AMOXICIL-CLARITHROMY, SULFADIAZINE), and Medications (Insulin, Fluocinolone, Clonazepam, Cetirizine, Acetaminophen, Bupropion, Lisinopril, Lorazepam, Fluoxetine, Lisinopril, Simvastatin, Albuterol, Epinephrine).
- Health Maintenance:** Lists completed immunizations: HEARING SCREENING (4 YEARS), (HEDIS) HEPATITIS B (0-18 YEARS), (HEDIS) DIPHTHERIA-TETANUS-PERTUSSIS, (HEDIS) POLIOMYELITIS, (HEDIS) MEASLES-MUMPS-RUBELLA (1-5 YEARS), and (HEDIS) VARICELLA (1-18 YRS).
- Patient Lists:** Shows lists for ASTHMA BTRPEDS [552], TEST [1616], TEST [1183], TEST [3841], and TEST [92].

Big data is already familiar!



PadMapper

Big data = big **money**



*NIH Big Data to
Knowledge (BD2K)*



\$32m

\$196m

How do we look at big data?

Machine learning!

1. Predict
2. Classify
- 3....

Predict extubation outcomes using logistic regression

Seely et al. *Critical Care* 2014, **18**:R65
<http://ccforum.com/content/18/2/R65>



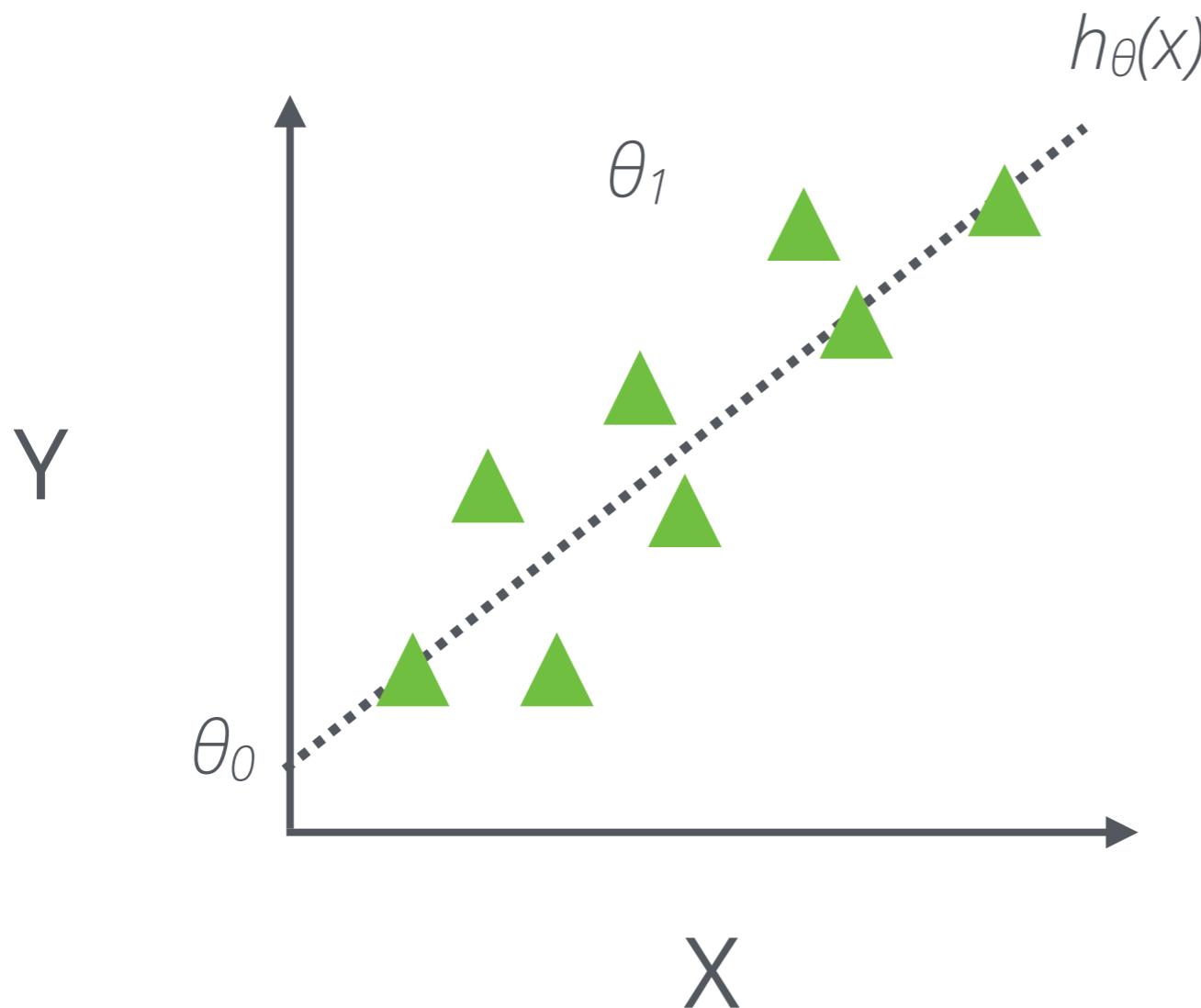
RESEARCH

Open Access

Do heart and respiratory rate variability improve prediction of extubation outcomes in critically ill patients?

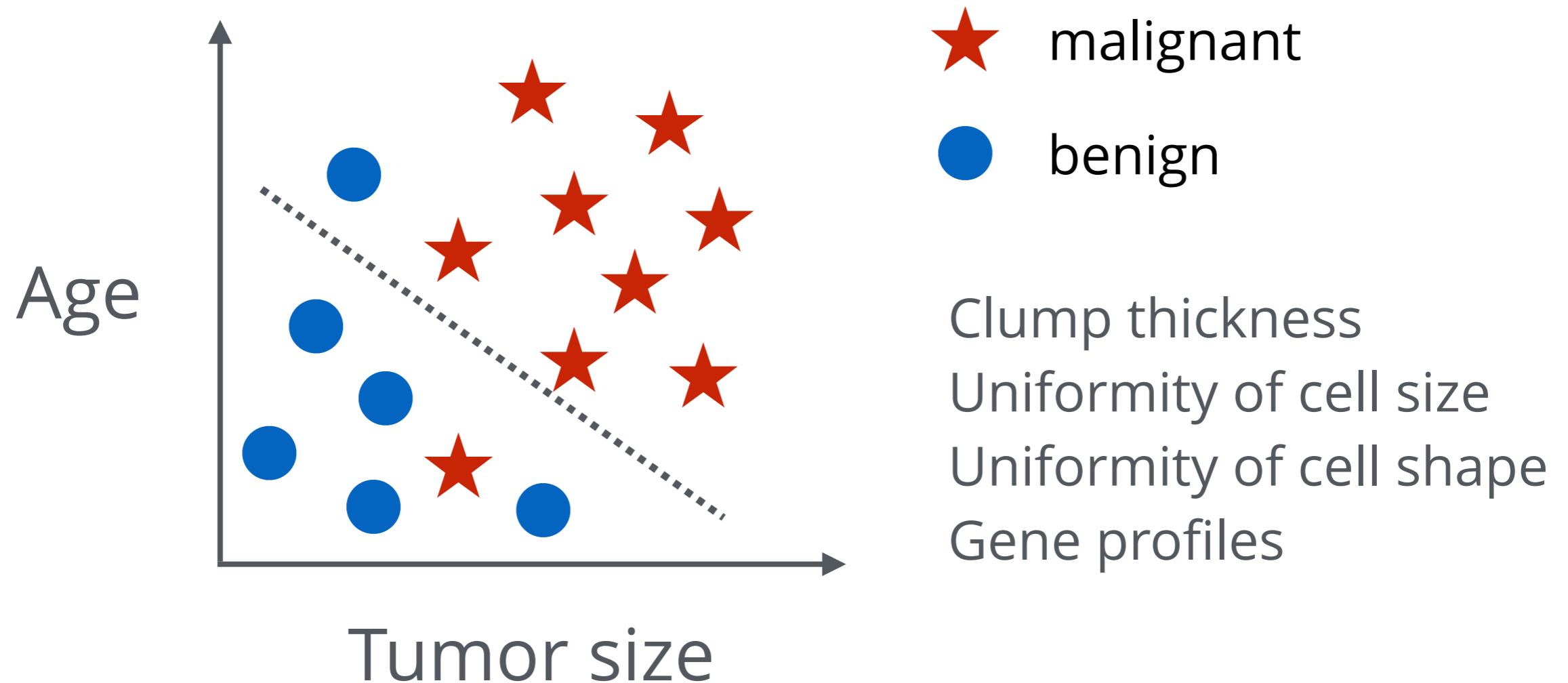
Andrew JE Seely^{1,2,11*}, Andrea Bravi², Christophe Herry¹, Geoffrey Green¹, André Longtin², Tim Ramsay¹, Dean Fergusson¹, Lauralyn McIntyre¹, Dalibor Kubelik¹, Donna E Maziak¹, Niall Ferguson³, Samuel M Brown⁴, Sangeeta Mehta⁵, Claudio Martin⁶, Gordon Rubenfeld⁷, Frank J Jacono⁸, Gari Clifford⁹, Anna Fazekas¹ and John Marshall¹⁰

Logistic regression = predict

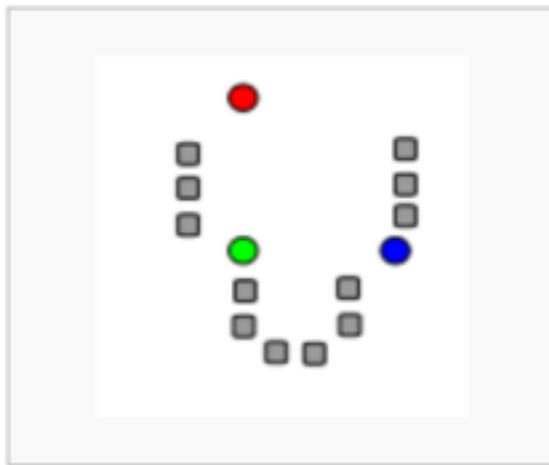


Choose θ_0, θ_1 so $h_{\theta}(x)$ is close to y for our **training examples** (x, y)

Classify tumors using clustering



K-means clustering is used to **classify**



1) k initial "means" (in this case $k=3$) are randomly generated within the data domain (shown in color).



**American Board
of Medical Specialties®**

Higher standards. Better care.™

In September 2011, the ABMS approved Clinical Informatics as a board-eligible subspecialty.

Big data is revolutionizing clinical care & biomedical research



Learn more:



Stanford
University

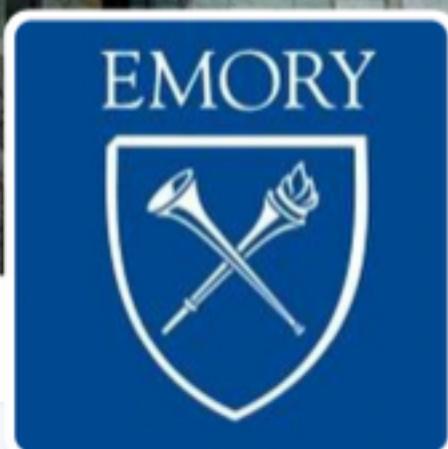




Search Twitter



Have an account? Sign in ▾



Emory MSTP

@EmoryMSTP

The Medical Scientist Training Program at [@EmoryMedicine](#) trains outstanding MD-PhDs who will become leaders in academic medicine and research.

Atlanta, GA

[EmoryMSTP.com](#)

[4 Photos and videos](#)



TWEETS
66

FOLLOWING
180

FOLLOWERS
42

FAVORITES
24

[Follow](#)

Tweets

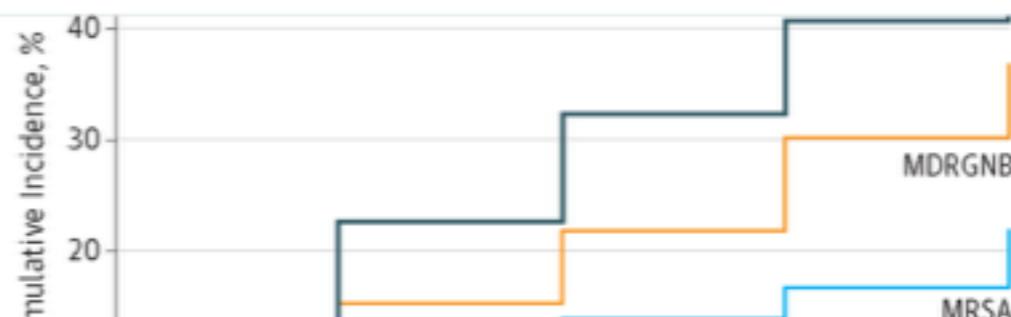
Tweets & replies

Photos & videos

Emory MSTP retweeted

JAMAInternalMed @JAMAInternalMed · Oct 14

#Antimicrobials heavily prescribed for advanced #dementia, despite no criteria to support use. [jama.md/1px0NGe](#)



Don't miss any updates from **Emory MSTP**

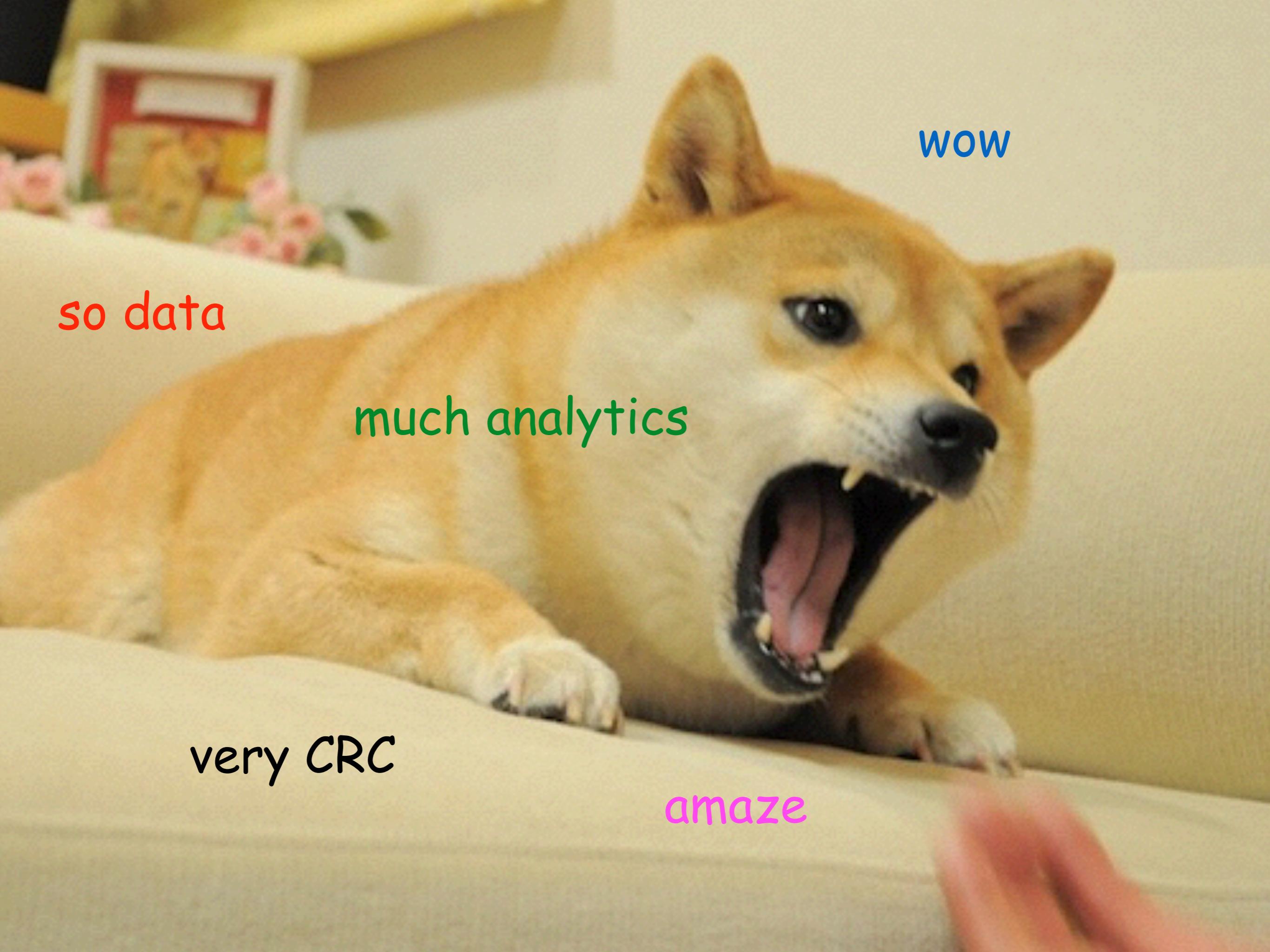
Full name

Email

Password

[Sign up for Twitter](#)

Worldwide Trends · [Change](#)



wow

so data

much analytics

very CRC

amaze