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.
Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC II)
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
Blood pressure and blood samples can be taken through catheter.
An investigation of patterns in hemodynamic data indicative of impending hypotension in intensive care

Joon Lee¹,²*, Roger G Mark¹,²

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

Andrew JE Seely¹,²,¹¹*, Andrea Bravi², Christophe Herry¹, Geoffrey Green¹, André Longtin², Tim Ramsay¹, Dean Fergusson¹, Lauralyn McIntyre¹, Dalibor Kubelik¹, Donna E Maziak³, Niall Ferguson³, Samuel M Brown⁴, Sangeeta Mehta⁷, Claudio Martin⁶, Gordon Rubenfeld⁷, Frank J Jacono⁸, Gari Clifford⁹, Anna Fazekas¹ and John Marshall¹⁰
Flatiron’s mission is to organize the world’s oncology data and make it useful for patients, physicians, life sciences & researchers.
Current clinical data infrastructure **sucks**
Big data is already familiar!
Big data = big **money**

$32m$  $196m$
How do we look at big data?
Machine learning!

1. Predict
2. Classify
3. ....
Predict extubation outcomes using logistic regression
Logistic regression = \textbf{predict}

Choose $\theta_0, \theta_1$ so $h_\theta(x)$ is close to $y$ for our training examples $(x,y)$.
Classify tumors using clustering

Age vs Tumor size scatter plot with points indicating malignancy and benign status. Points are color-coded:
- Red stars indicate malignant tumors.
- Blue circles indicate benign tumors.

Additional features:
- Clump thickness
- Uniformity of cell size
- Uniformity of cell shape
- Gene profiles
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).
In September 2011, the ABMS approved Clinical Informatics as a board-eligible subspecialty.
Big data is revolutionizing clinical care & biomedical research
Learn more:
The Medical Scientist Training Program at @EmoryMedicine trains outstanding MD-PhDs who will become leaders in academic medicine and research.

Emory Mstp
EmoryMSTP.com
so data
much analytics
very CRC
amaze
wow