# Mathematics in the Real World: Some Recent Successes and Open Challenges

Emmanuel Candès

*Moonshot International Symposium, Tokyo, December 18, 2019*

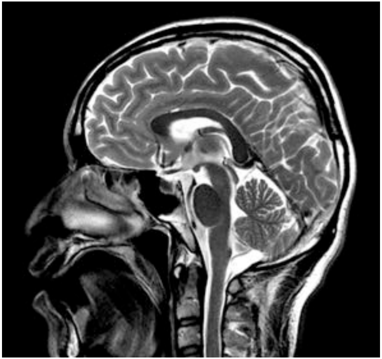
# Diagnostic medicine then...

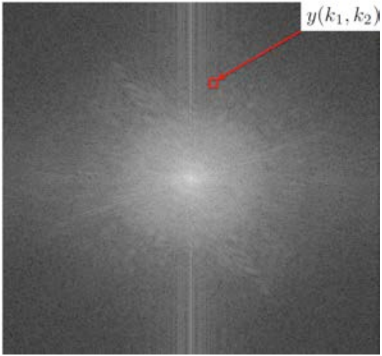
*The anatomy lesson of Dr. Nicolaes Tulp*, Rembrandt, 1632

# Diagnostic medicine now


Lauterbur & Mansfield (Nobel Prize in Medicine)

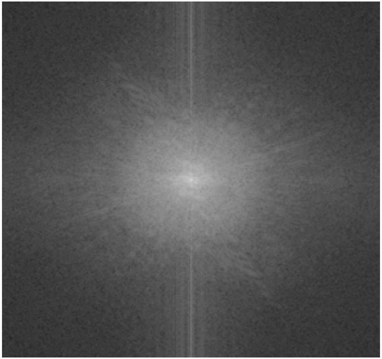
# What an MRI machine sees






Measured data $y(k_1, k_2)$ $\longleftarrow$ Fourier transform of image $f(x_1, x_2)$

# How do we form an image?

## Inverse Fourier transform

$$f(x_1, x_2) \approx \sum \sum y(k_1, k_2) \mathrm{e}^{i2\pi(k_1 x_1 + k_2 x_2)}$$

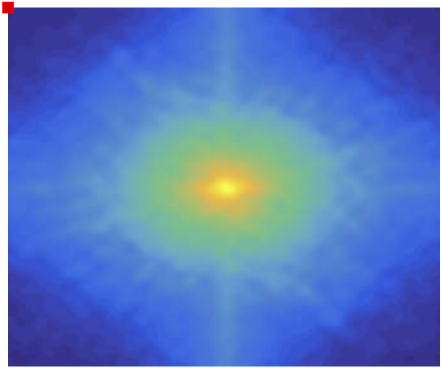
# MR Imaging: body examples

abdominal blood vessels

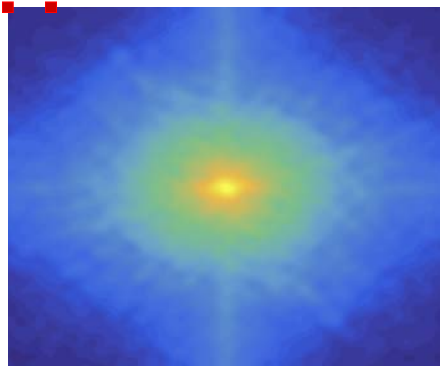

knee

*K. Pauly, G. Gold, RAD220*

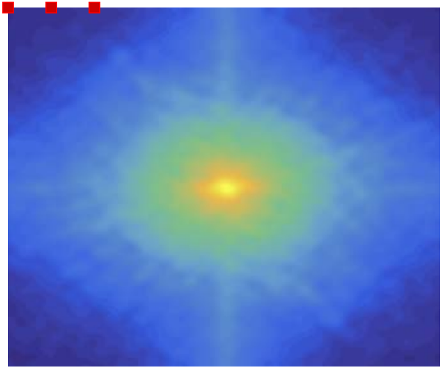
# MRI data collection is inherently slow



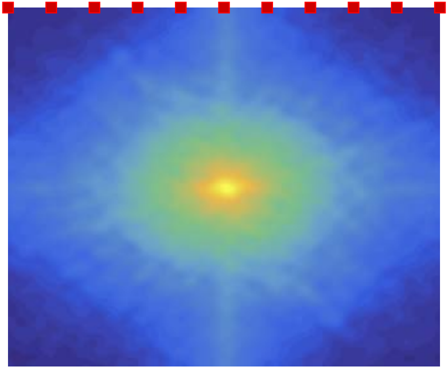
# MRI data collection is inherently slow



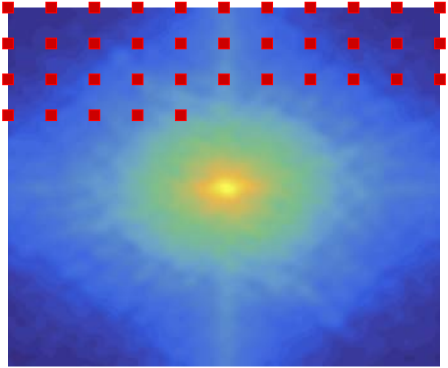
# MRI data collection is inherently slow



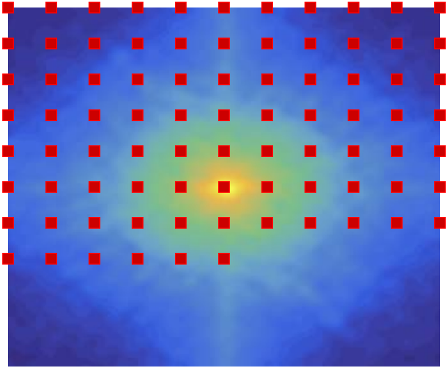
# MRI data collection is inherently slow



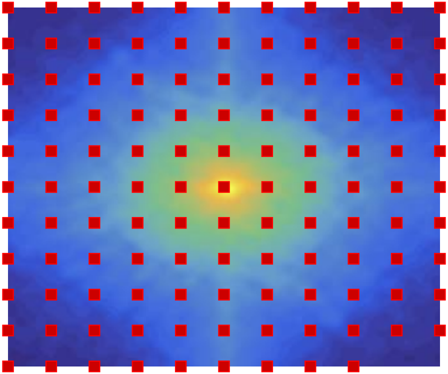
# MRI data collection is inherently slow



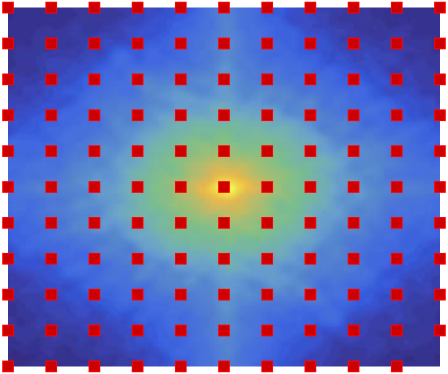
# MRI data collection is inherently slow



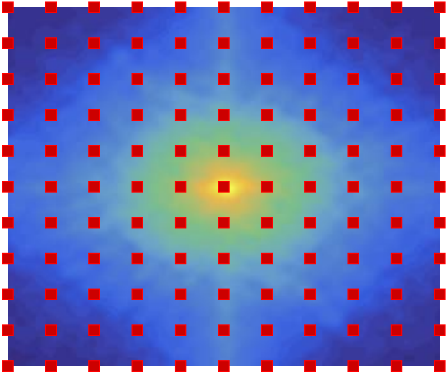
# MRI data collection is inherently slow



# MRI data collection is inherently slow



# MRI data collection is inherently slow



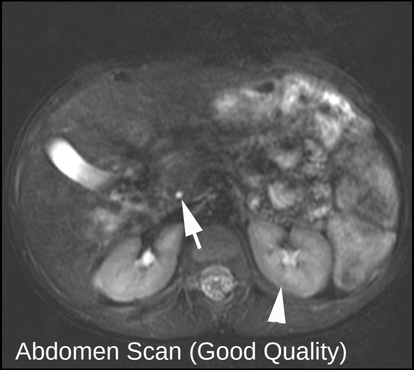
Done!

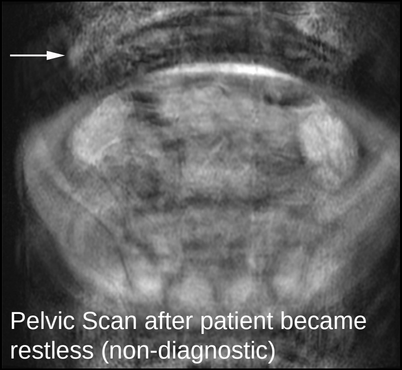
# Fact: impact of MRI on children health is limited


Children cannot stay still or breathhold!

- (deep) anesthesia required
- respiration suspension

# Need for speed in pediatric MRI




*Vasanawala and Lustig*

# Fewer equations than unknowns!

$$\mathbf{y} = \mathbf{A}\,\mathbf{x}$$

How can we possibly solve?

Thou shalt need at least as many equations as unknowns!

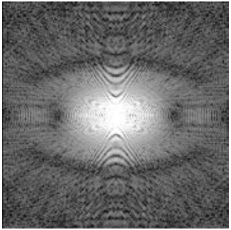
Carl Friedrich Gauss (1777 - 1855)

# A surprising experiment

C., Romberg and Tao ('04)

# A surprising experiment

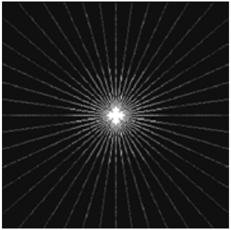
Fourier transform



C., Romberg and Tao ('04)

# A surprising experiment

Fourier transform

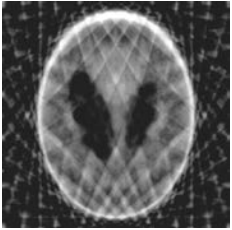


highly subsampled

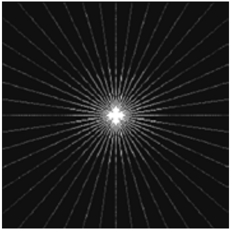
C., Romberg and Tao ('04)

# A surprising experiment

classical reconstruction



Fourier transform

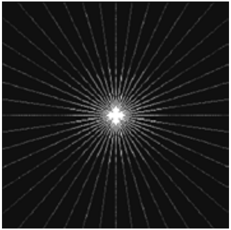


highly subsampled

C., Romberg and Tao ('04)

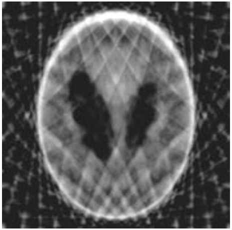
# A surprising experiment

Fourier transform



highly subsampled

classical reconstruction

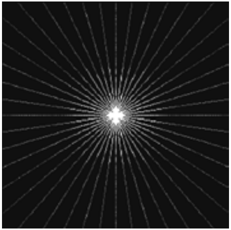


compressive sensing reconstruction

C., Romberg and Tao ('04)

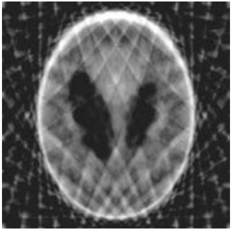
# A surprising experiment

Fourier transform



highly subsampled

classical reconstruction



compressive sensing reconstruction

Algorithm: minimize sum of gradient magnitudes subj. to data constraints

C., Romberg and Tao ('04)

# Sparse solutions

$$\underbrace{\begin{bmatrix} * \\ * \\ * \\ * \end{bmatrix}}_{y} = \underbrace{\begin{bmatrix} * & * & * & * & * & * & * & * & * \\ * & * & * & * & * & * & * & * & * \\ * & * & * & * & * & * & * & * & * \\ * & * & * & * & * & * & * & * & * \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} 0 \\ * \\ 0 \\ 0 \\ * \\ 0 \\ 0 \\ * \\ 0 \end{bmatrix}}_{x}$$

**Sparsity**

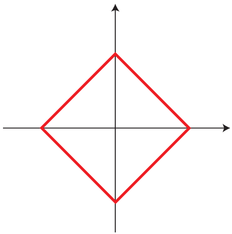
$x$ has at most $k < n$ nonzero entries

# Finding structured solutions

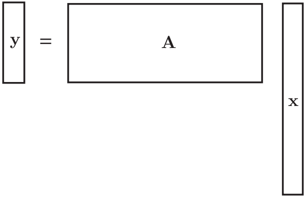
$$\underbrace{\begin{bmatrix} * \\ * \\ * \\ * \end{bmatrix}}_{y} = \underbrace{\begin{bmatrix} * & * & * & * & * & * & * & * & * \\ * & * & * & * & * & * & * & * & * \\ * & * & * & * & * & * & * & * & * \\ * & * & * & * & * & * & * & * & * \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} 0 \\ * \\ 0 \\ 0 \\ * \\ 0 \\ 0 \\ * \\ 0 \end{bmatrix}}_{x}$$

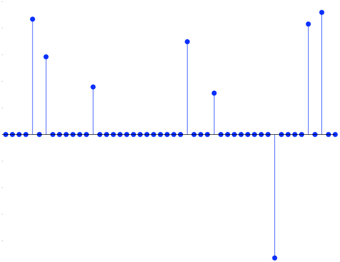
Recovery by linear programming

minimize $\quad \|x\|_{\ell_1} \quad$ subject to $\quad Ax = y$

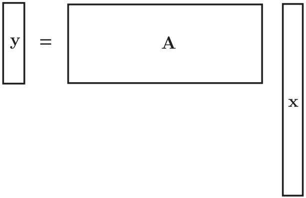


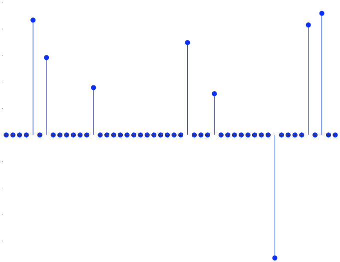
# Exact solution under sparsity constraints



# Exact solution under sparsity constraints



## Surprise [mathematical theorem by C., Romberg and Tao]

- There is a sparse solution
- Rows of $A$ are not sparse and diverse

Then mi $\ell_1$ solution is exact if

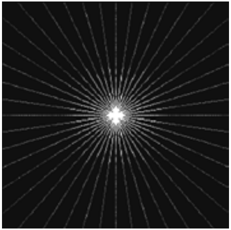
$$\#\text{ equations small multiple of } \underbrace{\#\text{ nonzero terms}}_{\text{information content}}$$

Undersampling $\rightarrow$ no information loss!

# It's a mathematical truth!

Fourier transform

compressive sensing
reconstruction



highly subsampled

min $\ell_1$ norm subj. to data constraints

C., Romberg and Tao ('04)

# Impact on MR pediatrics

*Lustig (UCB), Pauly, Vasanawala (Stanford)*

6 year old
8 fold acceleration
16 second scan
0.875 mm in-plane
1.6 slice thickness

# 1 year old female with liver lesions: 8X acceleration

*Pauly, Vasanawala (Stanford), Lustig (UCB)*


Parallel imaging (PI)

Compressed sensing + PI

Lesions are barely seen with linear reconstruction

# 6 year old male abdomen: 8X acceleration

*Zhang et al., JMRI 2014*

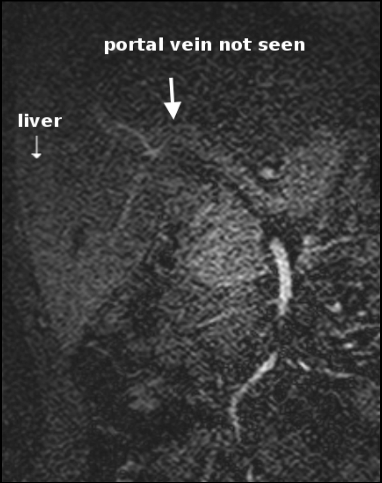


Parallel imaging (PI) | Compressed sensing + PI

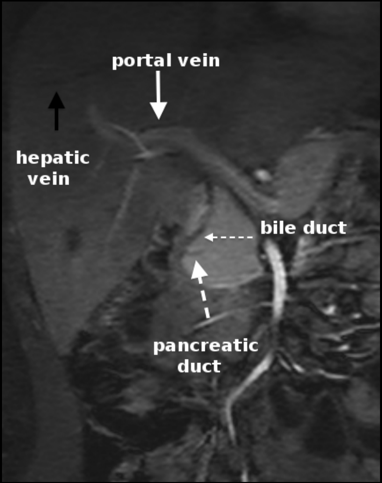
Fine structures (arrows) are buried in noise and recovered by CS

# 6 year old male abdomen: 8X acceleration

*Zhang et al., JMRI 2014*




Parallel imaging (PI)




Compressed sensing + PI

Fine structures (arrows) are buried in noise and recovered by CS

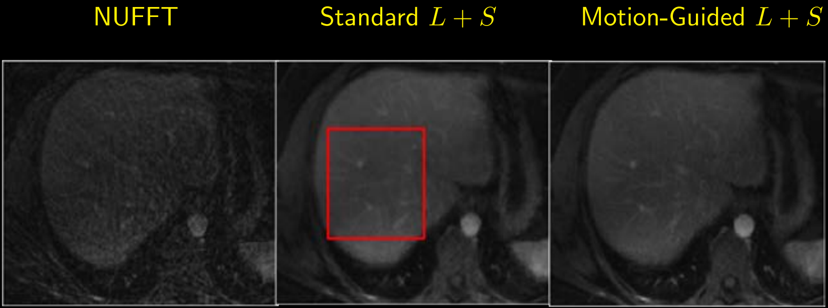
# Free-breathing MRI of the liver

NUFFT | Standard $L + S$ | Motion-Guided $L + S$

*12.8 fold acceleration*

*Joint with R. Otazo and D. Sodickson*

# Free-breathing MRI of the liver




Temporal blurring

*Joint with R. Otazo and D. Sodickson*

# Free-breathing MRI of the kidneys

NUFFT | Standard $L + S$ | Motion-Guided $L + S$

*12.8 fold acceleration*

*Joint with R. Otazo and D. Sodickson*

# Free-breathing MRI of the kidneys


*Joint with R. Otazo and D. Sodickson*

# One step further

Cheng et al., Stanford & UCB

3D phase-contrast cardiac MRI (4Dflow)

TECHNOLOGY | MAGNETIC RESONANCE IMAGING (MRI) | FEBRUARY 21, 2017

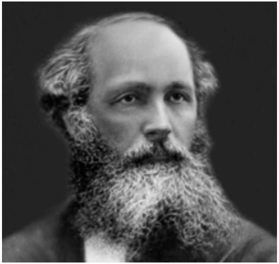
# FDA Clears Compressed Sensing MRI Acceleration Technology From Siemens Healthineers

*New technology employs iterative reconstruction to produce high-quality MR images at a rapid rate with zero diagnostic information loss*

*Opportunities and Challenges in the Data-driven Era*

# Mechanistic models

Newton



Maxwell

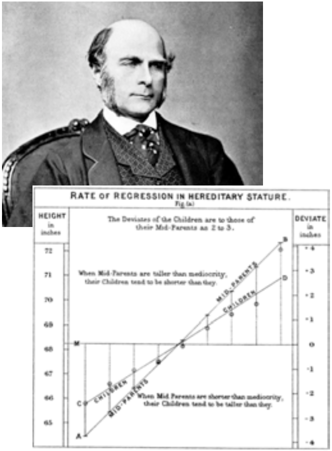
Schrödinger

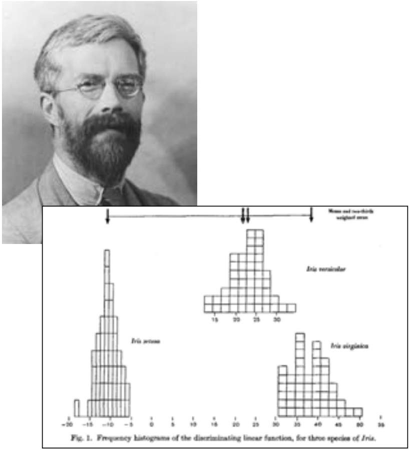
*From a long view of the history of mankind—seen from, say, ten thousand years from now—there can be little doubt that the most significant event of the 19th century will be judged as Maxwell's discovery of the laws of electrodynamics*

R. Feynman

s

# Prediction without a mechanistic model

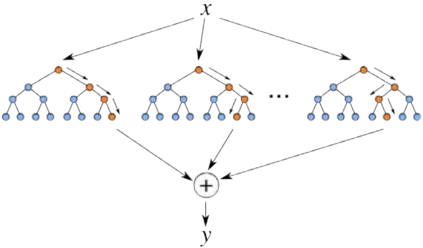


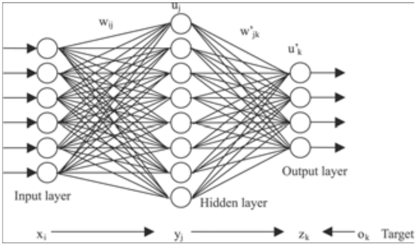

- Galton (1886): regression of sons' heights onto parents' heights
- Fisher (1936): classification of Iris species

# Today's predictive non-mechanistic algorithms

random forests, gradient boosting




neural networks




Breiman and Friedman

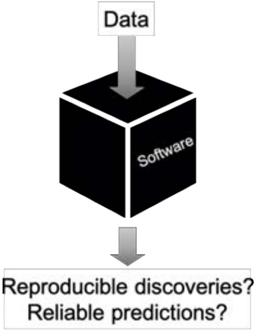
LeCun, Hinton and Bengio

# Things we cannot surrender

- Tension between interpretability and model complexity
- People become comfortable with model complexity because of tantalizing potential

# Things we cannot surrender

- Tension between interpretability and model complexity
- People become comfortable with model complexity because of tantalizing potential




- **Replicability**: Conclusions I draw today need to hold up tomorrow
- **Reliability**: Are my predictions valid? How do I communicate uncertainty?

# Houston, we have a problem...

Begley and Ellis, *Nature* (2012)

- Amgen could only replicate 6 of 53 studies they considered landmarks in basic cancer science
- HealthCare could only replicate about 25% of 67 seminal studies
- Systematic attempts to replicate widely cited priming experiments have failed




*New York Times*, *Science*, November 19, 2018

# Why now? Have entered a new scientific paradigm

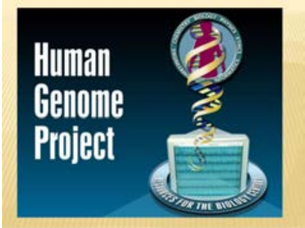

Nature, 2008,
Vol. 455, Issue 7209


Science, 2011,
Vol. 331, Issue 6018

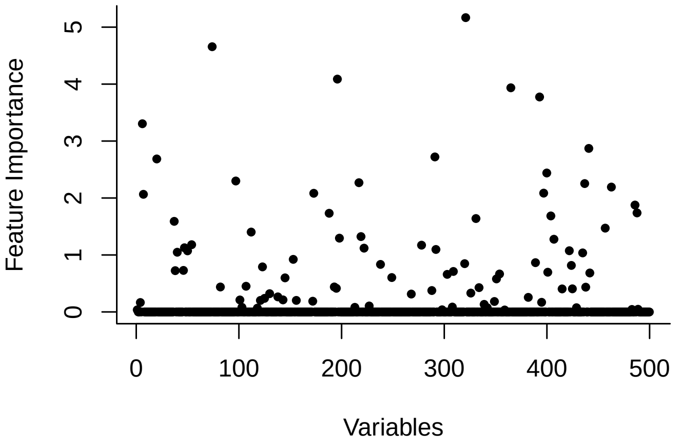

The Economist,
Feb.27th, 2010

Collect data first $\Longrightarrow$ Ask questions later

Major goal:
need new statistical tools
to address new problems

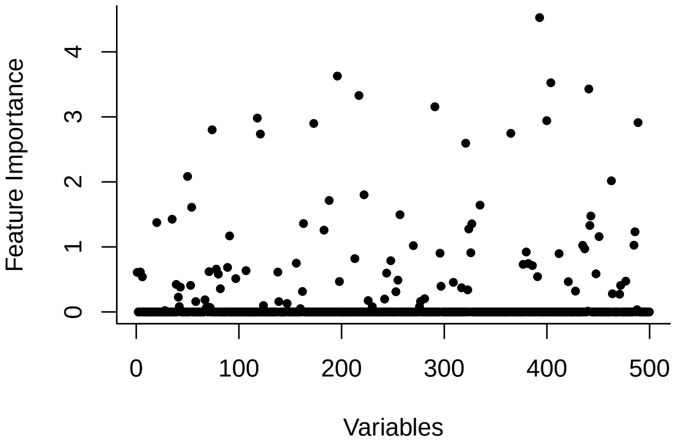


# Selecting promising leads from black-box algorithms?



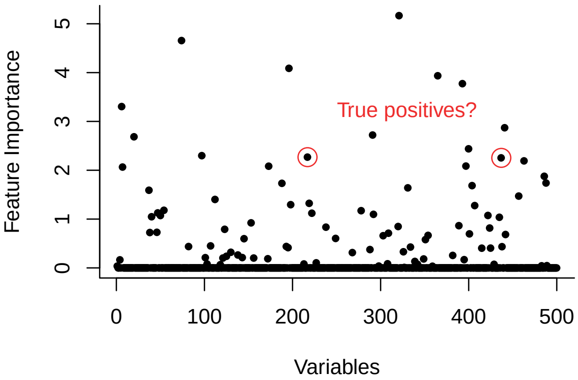

*How do we make reliable decisions in the face of unknown statistical variability?*

# Selecting promising leads from black-box algorithms?




*How do we make reliable decisions in the face of unknown statistical variability?*

# Only one dataset: what should we report?



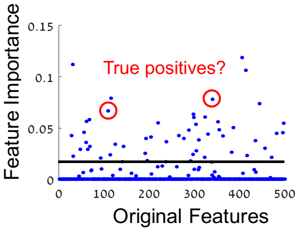

*Modern science faces the problem of selection of promising findings from the noisy estimates of many.*

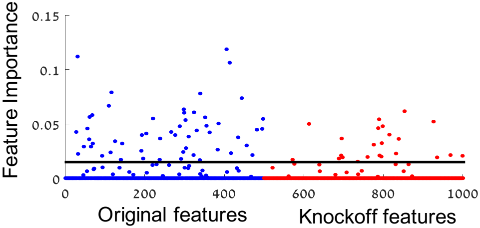
Y. Benjamini and Y. Hechtlinger

# Knockoffs (Barber and Candès, 2015)

For each variable (e.g. SNP) $X_j$, make a knockoff version (e.g. fake SNP) $\tilde{X}_j$

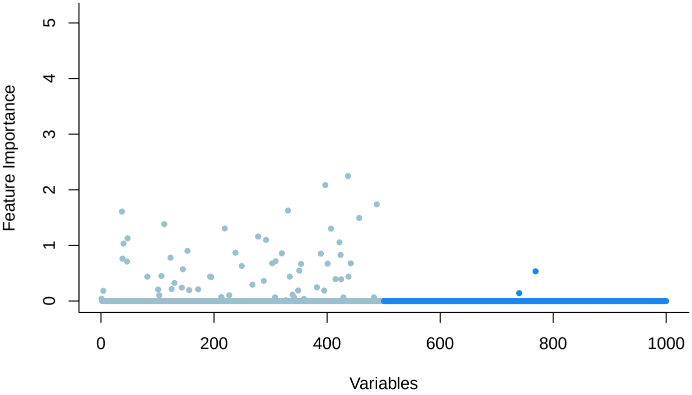
Run scoring procedure on features and knockoffs 'serving as controls'



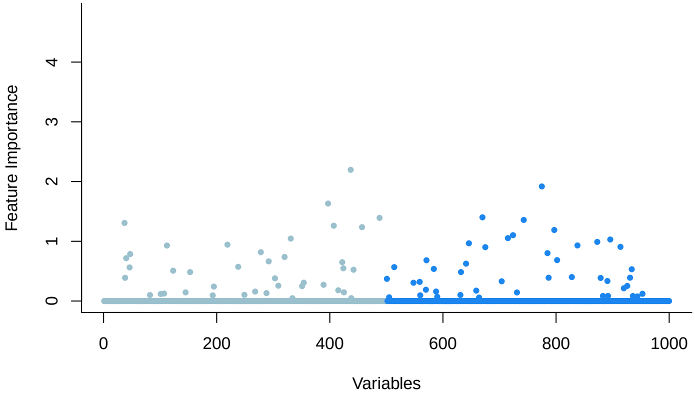

Black box "selects" 49 original features & 24 knockoff features
$\Longrightarrow$ probably $\approx$ 24 false positives among 49 original features

# Permuted dummies do not work!



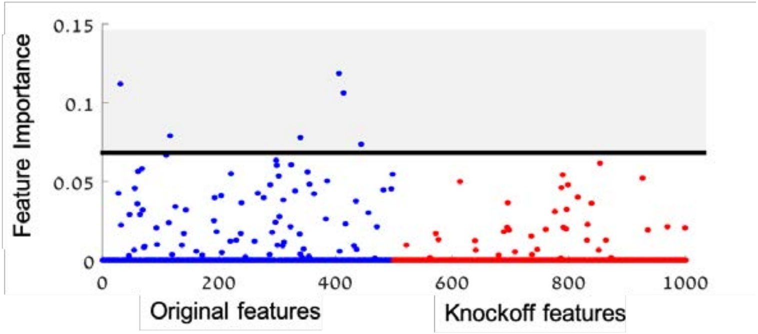

Importance of correct statistical reasoning

# Knockoff dummies work!



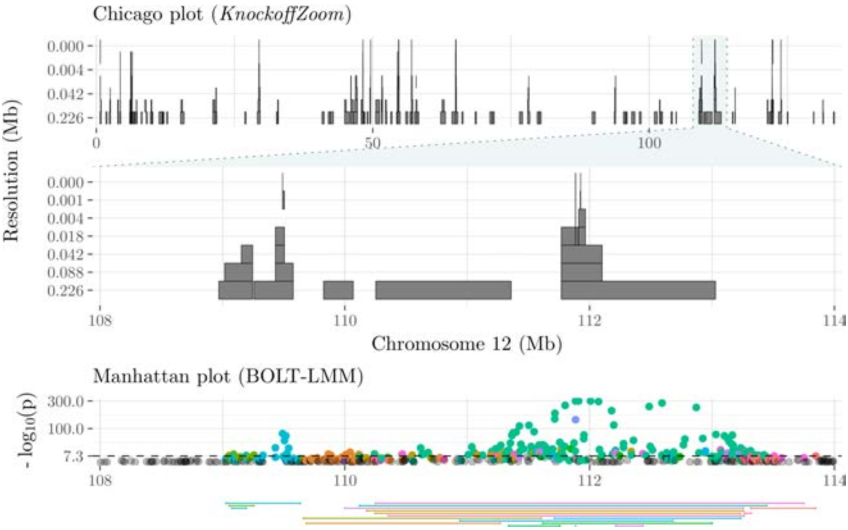

Importance of correct statistical reasoning

# Importance of mathematics

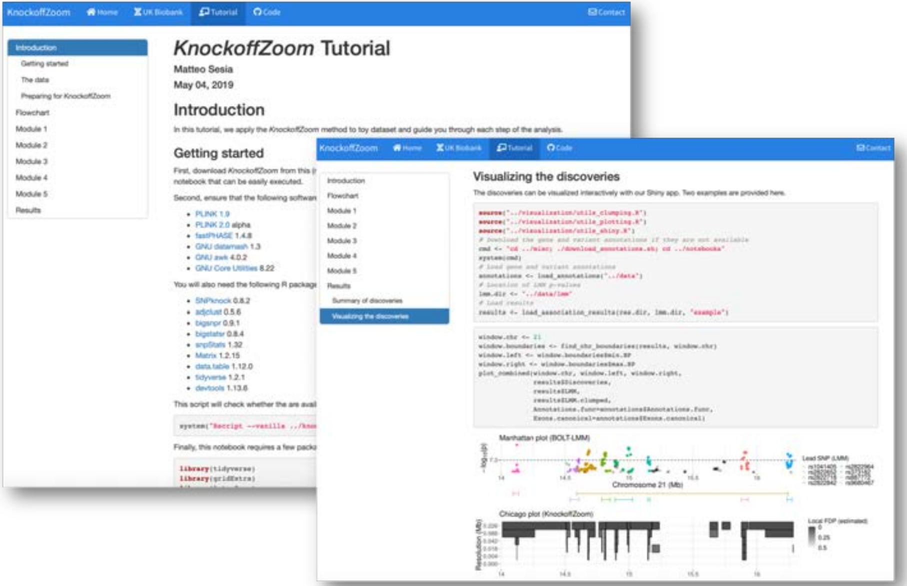



- Theory of exchangeable variables
- Theory of martingales

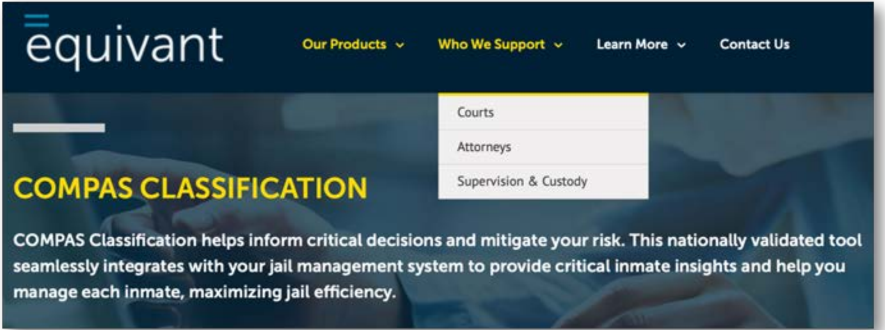
# Analysis of platelet count



# KnockoffZoom: website & code



# Machine learning in sensitive applications



# With Malice Towards None:
# Assessing Uncertainty via Equalized Coverage

Yaniv Romano* Rina Foygel Barber† Chiara Sabatti* ‡ Emmanuel J. Candès* §

August, 2019

# On the use of ML to support important decisions

- How do we communicate uncertainty to decision makers?
- How do we not overstate what can be inferred from the black box?
- How do we treat everyone equitably?

**Our take:**

Decouple the statistical problem from the policy problem

Corbett-Davis and Goel, '19

Somewhat against current thinking in "algorithmic fairness", e.g. equalized odds

Hardt, Price and Srebro, '16

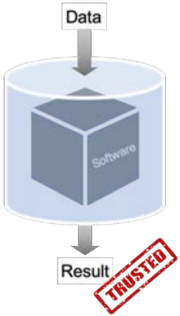
# Equalized coverage

<u>Goal</u>: construct perfectly calibrated intervals across all groups

$$\mathbb{P}\{Y_{n+1} \in C(X_{n+1}) \mid A = \male\} \geq 90\%$$
$$\mathbb{P}\{Y_{n+1} \in C(X_{n+1}) \mid A = \female\} \geq 90\%$$

Summarizes what we have learned from ML s.t.

- Rigorously quantifies uncertainty
- Treats individuals equitably



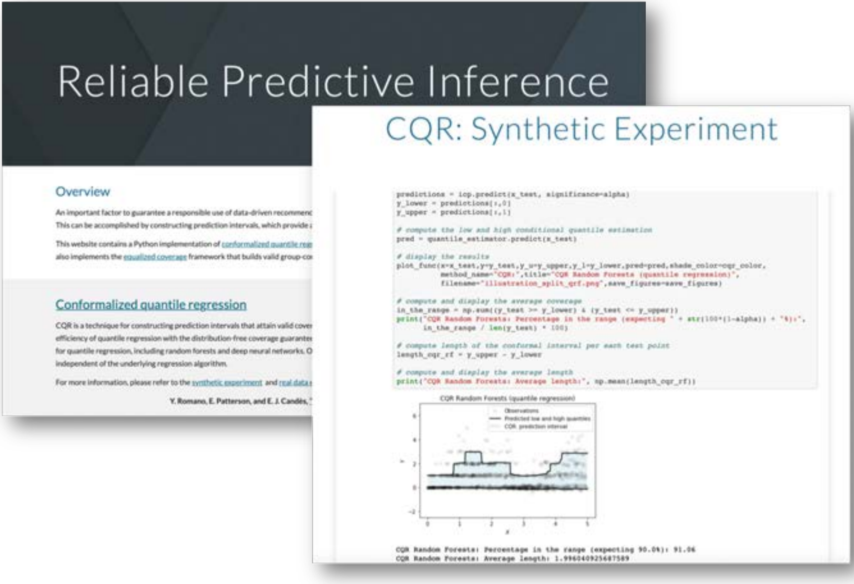
# Performance

- Average across 40 random train-test (80%/20%) splits

| Method | Group | Avg. Coverage | Avg. Length |
|---|---|---|---|
| Conformal (separate train.) | Non-white | 0.903 | 2.764 |
| | White | 0.901 | 3.182 |
| Conformal (joint train.) | Non-white | 0.904 | 2.738 |
| | White | 0.902 | 3.150 |
| CQR (separate train.) | Non-white | 0.904 | 2.567 |
| | White | 0.900 | 3.203 |
| **CQR (joint train.)** | Non-white | 0.902 | **2.527** |
| | White | 0.901 | **3.102** |

- CQR produces shorter intervals
- Join training is more powerful

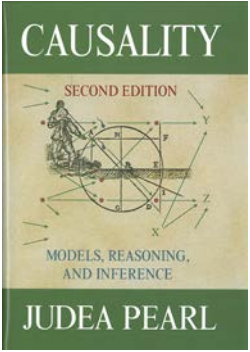
# Website & code



# The next ten years...

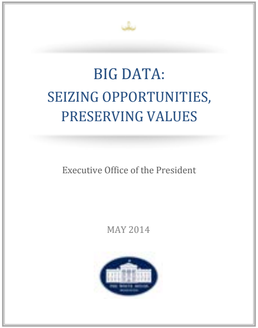

Replicability

The Economist Oct.19th, 2013




Causality

Judea Pearl(2009), Causality(2nd Edition), Cambridge University Press, ISBN-10: 052189560X




Fairness

Executive Office of the President(2014), BIG DATA: Seizing Opportunities, Preserving Values, ISBN-10: 1503016447
https://obamawhitehouse.archives.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf

## Learning from data is not trivial

- Correct statistical reasoning increasingly important
- Importance of mathematics