idies

The Institute for Data Intensive Engineering and Science

JOHNS HOPKINS UNIVERSITY

The Science of Big Data

The Institute for Data Intensive Engineering and Science will foster education and research in applying data-intensive technologies to problems of national interest in physical and biological sciences and engineering.

IDIES will provide faculty, researchers and students with the structure and resources needed to accomplish these goals.

FROM BALTIMORE TO THE STARS WITH DATA

Tamas Budavari / Applied Math & Stats, JHU

Seed Funding Awardees

Fall 2014

Urban Planning in Baltimore City

Tamas Budavari (Dept. of Applied Mathematics & Statistics), Kathryn Edin (Dept. of Sociology), and Michael Braverman (Dept. of Housing & Community Development, Housing Authority of Baltimore City)

Our Vacant Housing Dynamics in Baltimore City Project aims to improve the quality of city life by integrating data-driven science with redevelopment-policy and administration. Working with City officials, our goal is to better understand the dynamics of vacant housing in Baltimore City, measure the impact of current interventions, and hone decision- and policy-making with statistical analyses of available data. Addressing the vacancy crisis is essential to attracting and retaining people in Baltimore, a key goal formalized in the **Grow Baltimore** program.

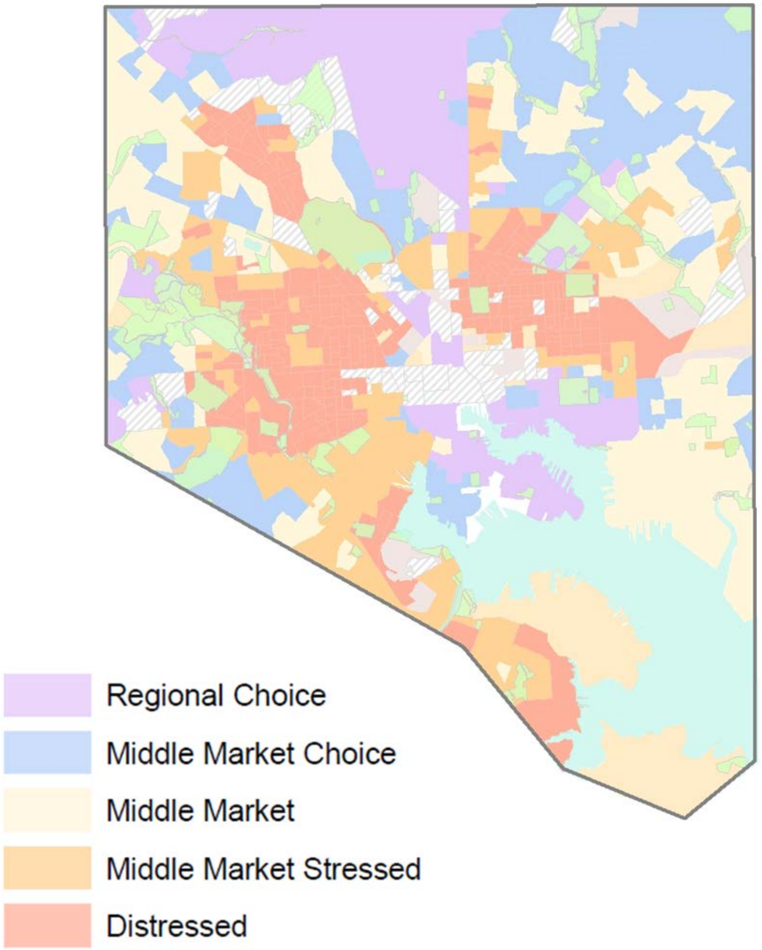
Breaking the Divestment Cycle: Predicting Abandonment & Fostering Neighborhood Revitalization in Baltimore

Tamás Budavári

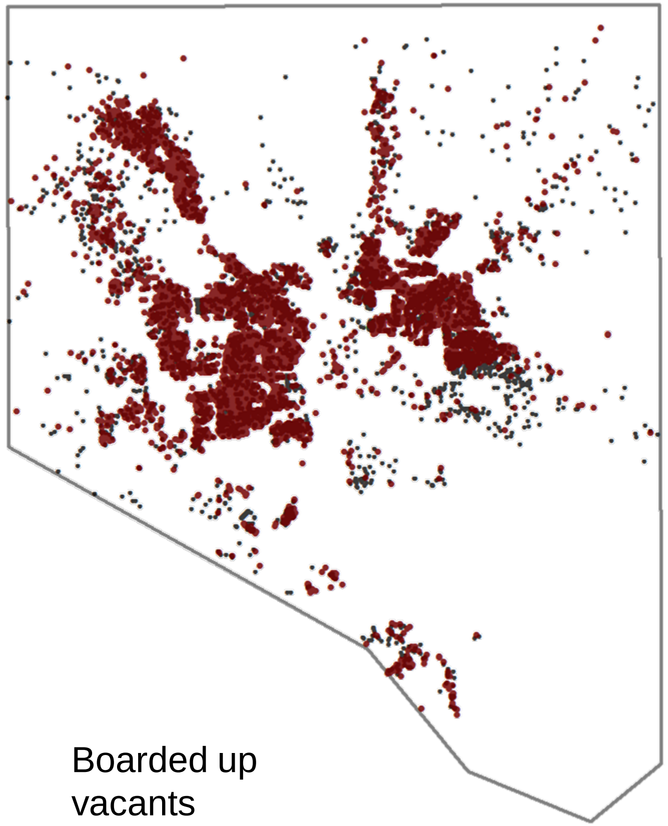
Applied Mathematics & Statistics – The Johns Hopkins University

Baltimore overview

- Baltimore has lost 1/3 of its population since 1950
- Today, we have 16,500 boarded up vacant buildings
- Of these, 13,000 are in distressed markets



Regional Choice
Middle Market Choice
Middle Market
Middle Market Stressed
Distressed



Boarded up vacants

M. Braverman

1

data science

flexible data platform

predictive modeling & optimization

data fusion

geometry + history
highly extensible

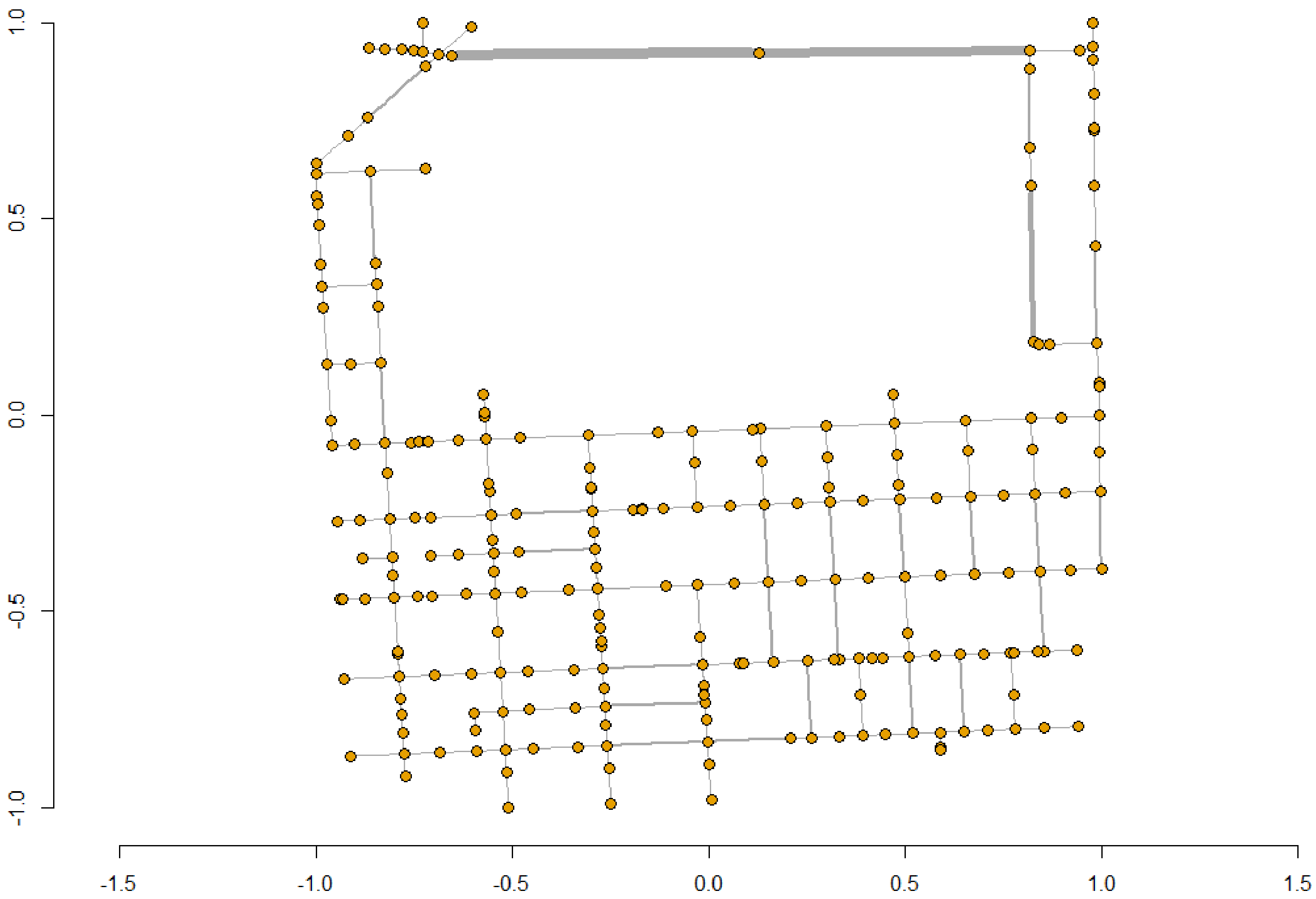
2

social science

modeling transition

estimating
externalities

evaluating policy



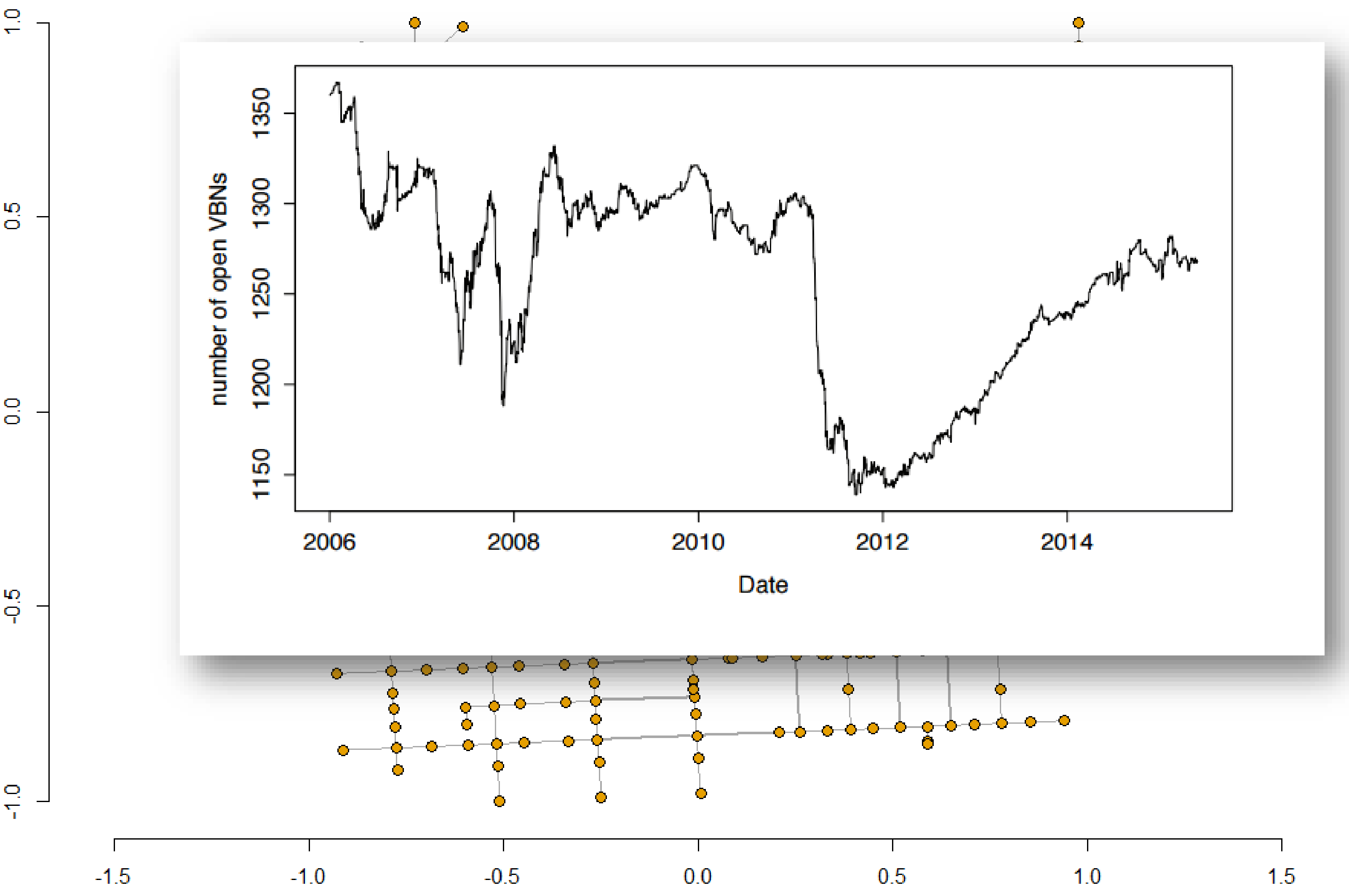
2

social science

modeling transition

estimating
externalities

evaluating policy



3

government

rapid response queries

assisting with strategic investments

mapping "unoccupancy"

Data in Baltimore

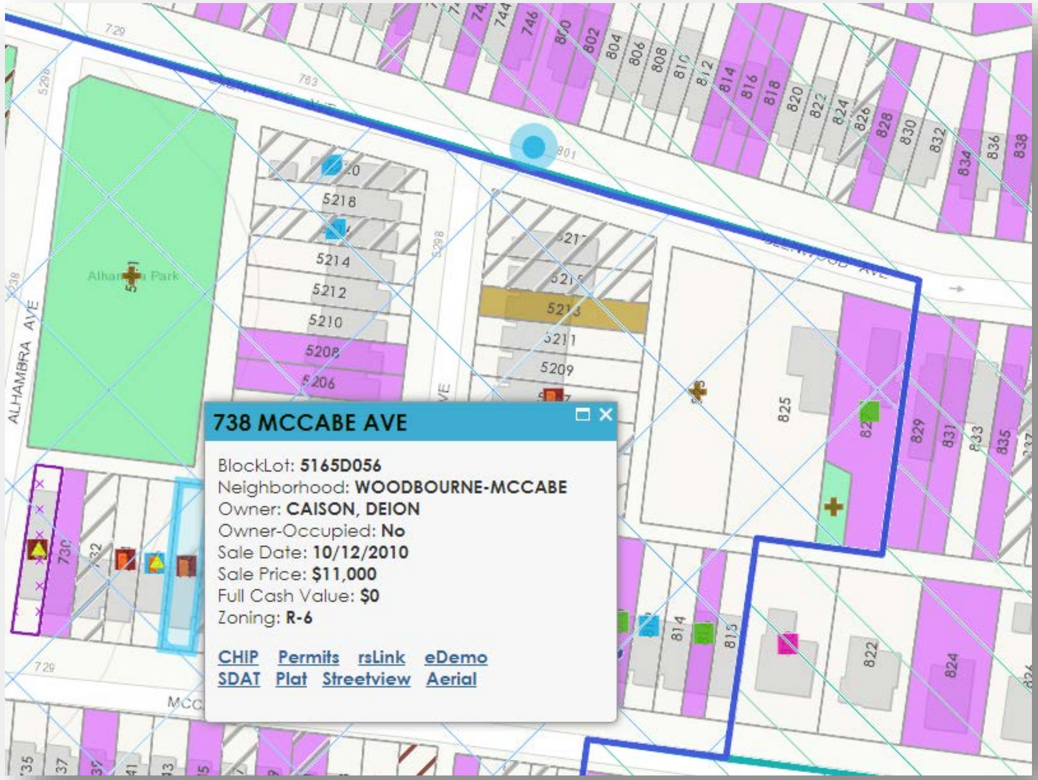
- OpenBaltimore
  - Hundreds of public datasets online
    http://data.baltimorecity.gov

- Plus more administrative data

DHCD's Data Infrastructure

- Dept. of Housing & Community Dev
  - Study changes over time
  - Support decision making

M. Braverman
J. D. Evans



- Statistics to help?
  - Inference & prediction

Jim Gray's 20 Questions

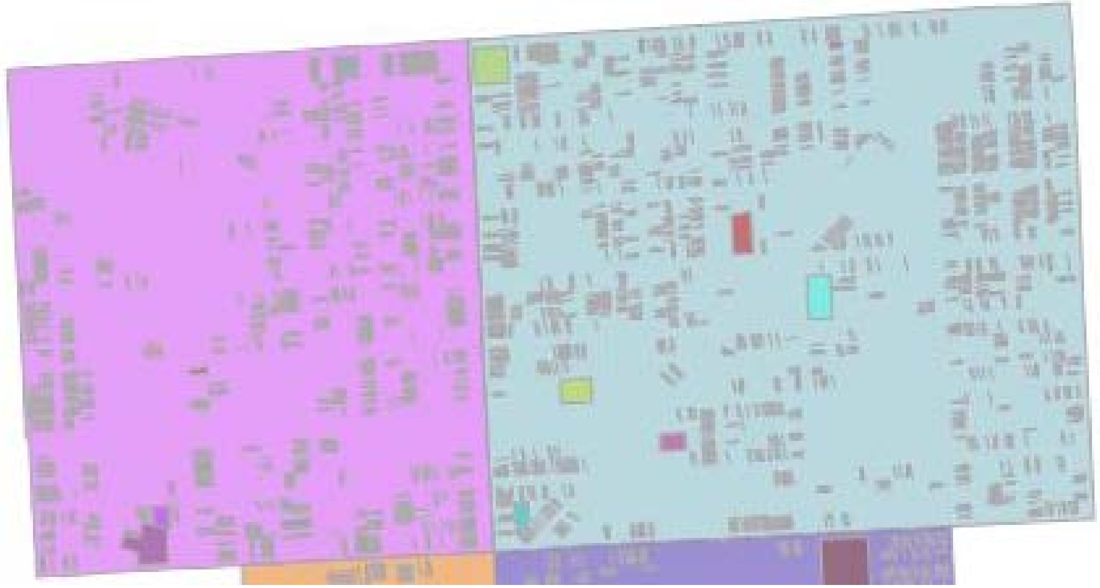
- Data-driven studies
  - Low-level questions
    - What we see
- High-level questions
  - Help hone policy making
    - Interventions

Built a Unique Solution

- Database of Baltimore City
  - Geospatial info for all parcels
  - Time history of real properties
- Easily extendable
  - On the IDIES's Data-Scope
  - Novel indexing for fast links

Mapping Vacancy

- 2010



- 2015

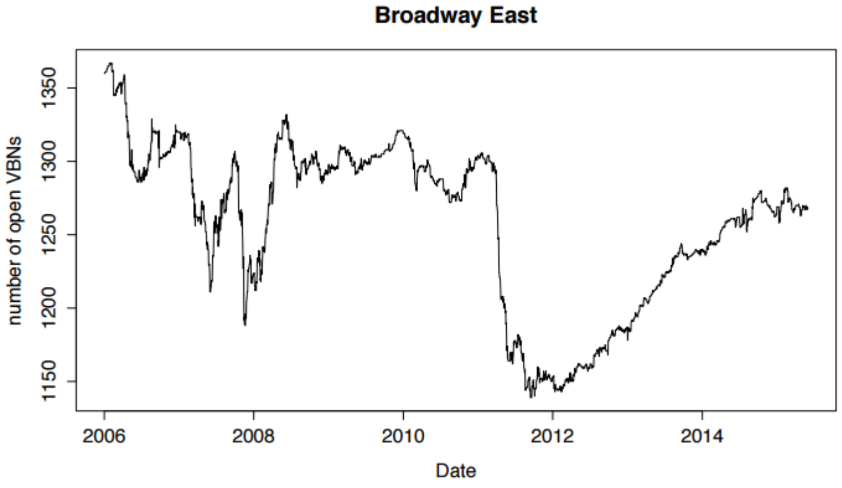
Phil Garboden

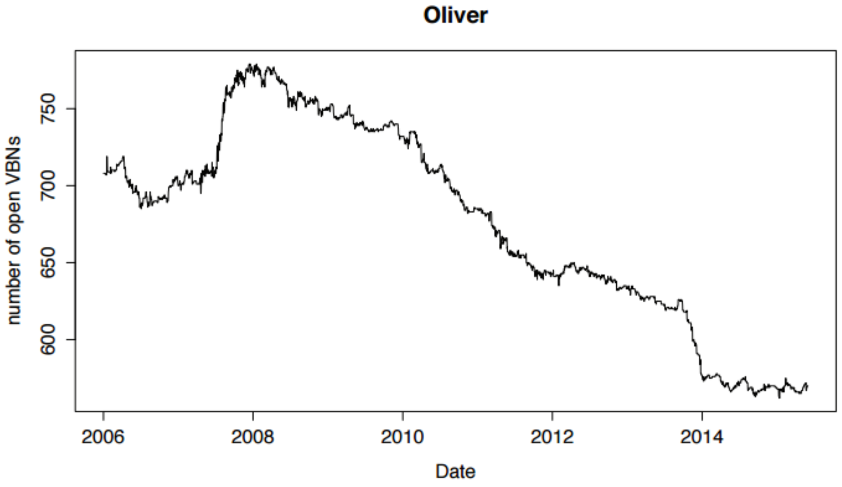
Map

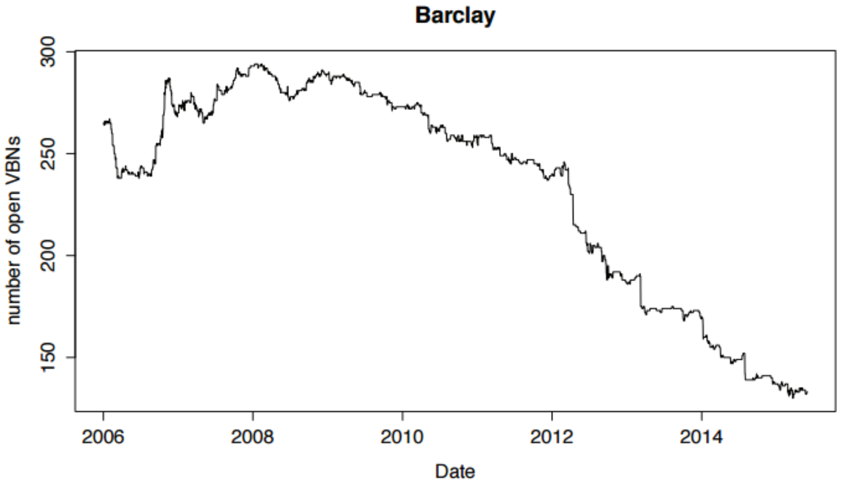
- 201

OPEN VACANT BUILDING NOTICES *in* BALTIMORE NEIGHBORHOODS

January 2006 - June 2015







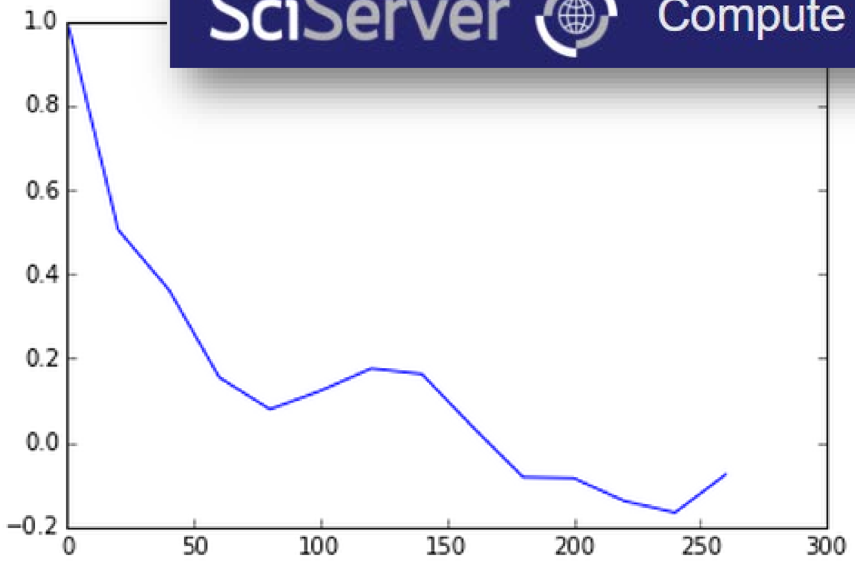
Phil Garboden

Clustering of Vacancy

- Probability of finding a vacant next to another
- Quantitative comparison
  - Over time
  - Across town

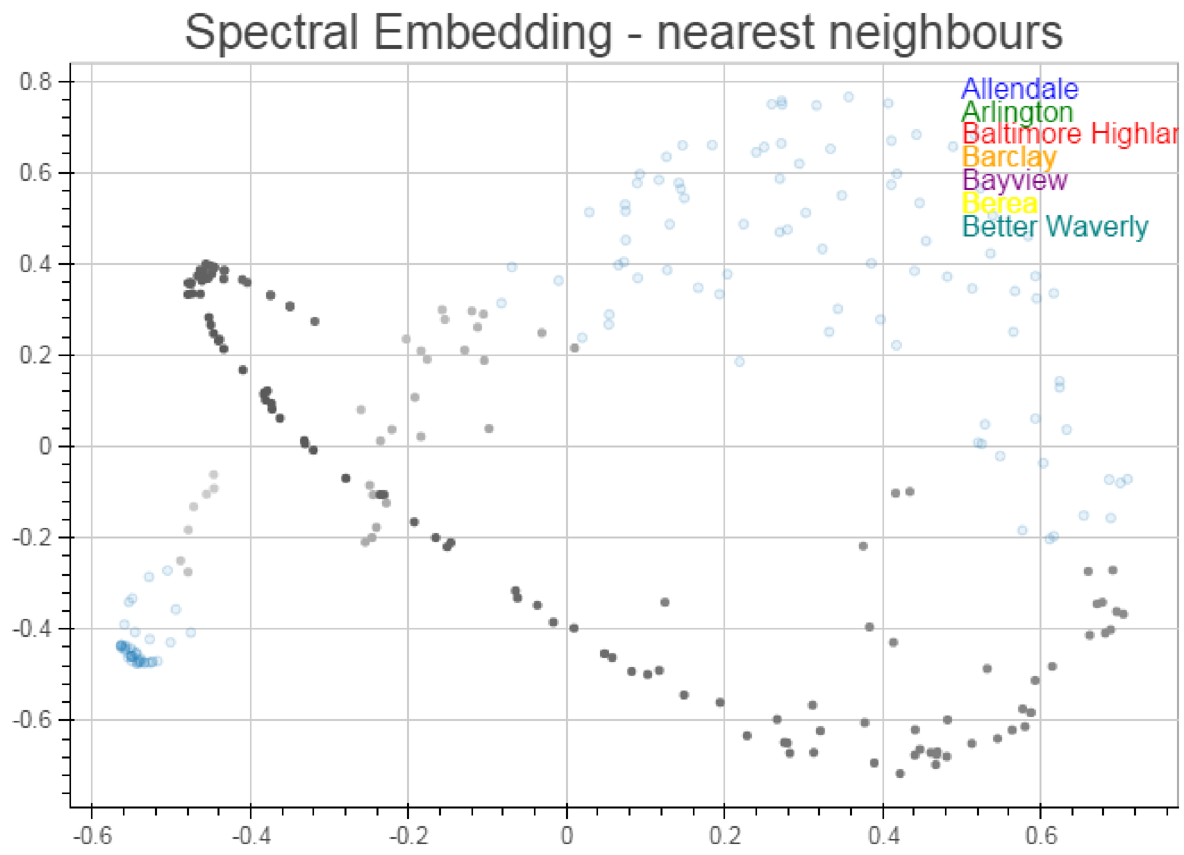
```
In [12]: a,b = pairCorrelation(M,3,300,15,50)

In [13]: plot(b,a)
         #hlines(0,0,250)

Out[13]: [<matplotlib.lines.Line2D at 0x7a51c88>]
```

SciServer Compute



Similar Neighborhoods

- Similarity graphs & eigenmaps

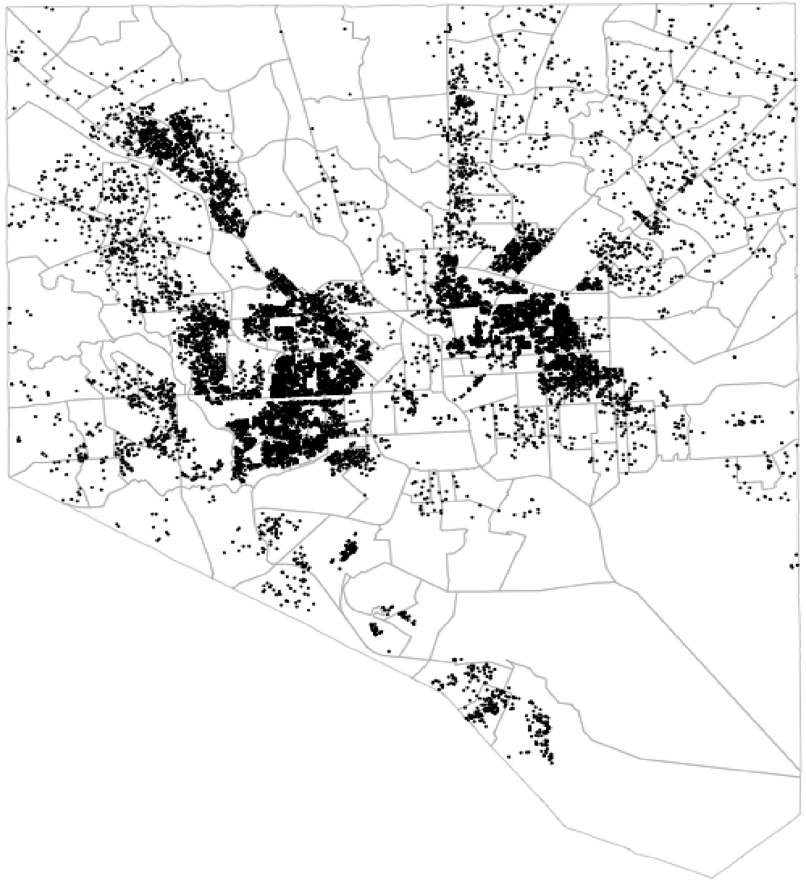


What is a Neighborhood?

- Are neighborhood boundaries meaningful?
- Better grouping of houses?
  - Trends on a finer scale

Collapsed Vacants

Map 2: Abandoned Properties In Baltimore City



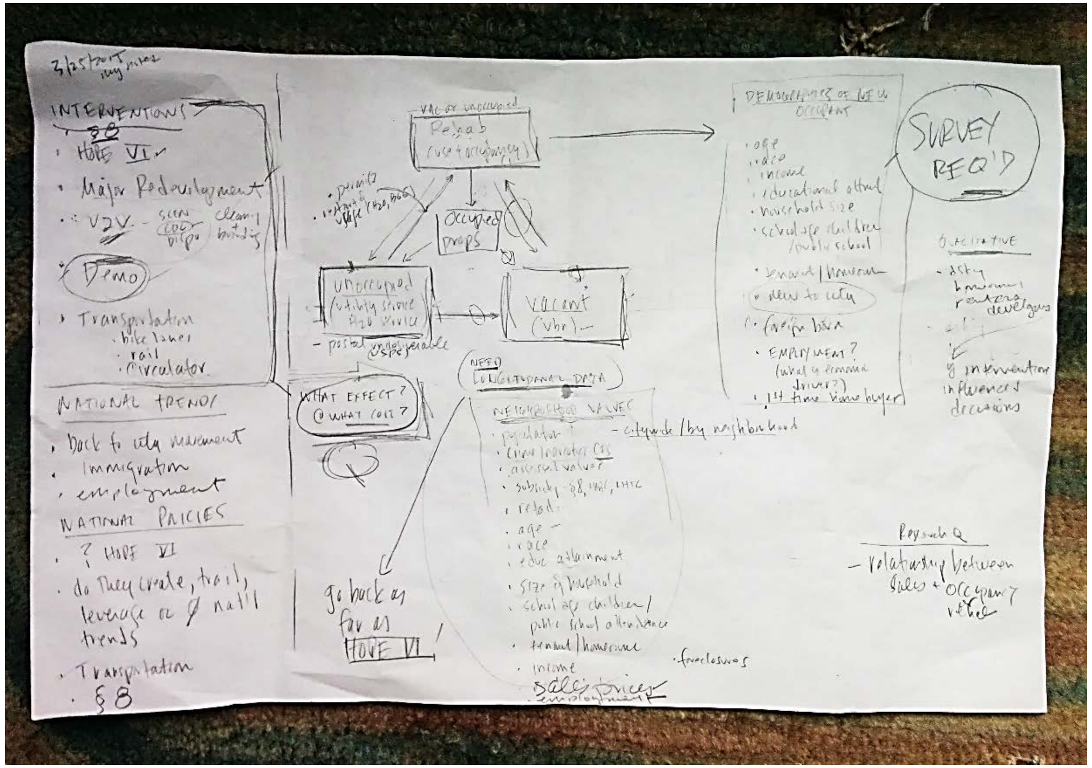
Map 3: Collapsed Properties In Baltimore City

Collapsed Vacant

- Ends of contiguous blocks of rowhomes
  - Alleys, gaps and demos break rows
- Need “sub-blockface” analysis
  - Time-dependent

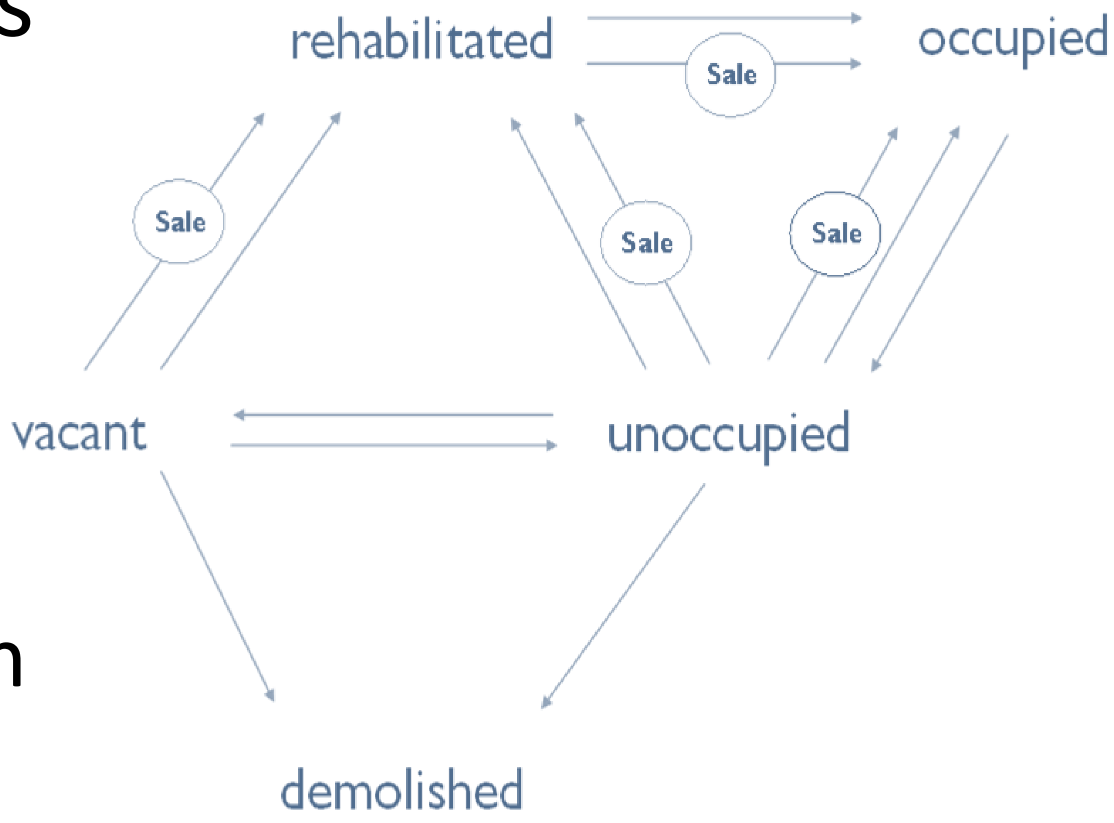
Neighborhood Revitalization

- Modeling urban transitions
  - What factors catalyze reinvestment?
  - Disinvestment?
- Innovative use of data
  - New sources of information
    - Zillow? Cell phone usage?



Neighborhood Revitalization

- Modeling urban transitions
  - What factors catalyze reinvestment?
  - Disinvestment?
- Innovative use of data
  - New sources of information
    - Zillow? Cell phone usage?



Strategic Investments

- Governor's budget
  - Unprecedented $75M
- City scheduling
  - Spring 2016
- JHU map of targets!

Strategic Investments

- Combinatorial Optimization
  - Improve some objective, e.g.,

$$C_0(\boldsymbol{x}) = \sum_{i \in H} x_i \quad \text{or} \quad C_1(\boldsymbol{x}) = \sum_{i \in H} v_i \, x_i$$

  - Within a limited budget

$$B(\boldsymbol{x}) = \sum_{i \in H} c_i \, x_i + \sum_{(i,j) \in N} w_{ij} \left(x_i \oplus x_j \right)$$

- Best objective? How to solve?

Environment	\$1,500
2-story demo	\$13,000
3-story demo	\$22,000
2-story wall	\$14,000
3-story wall	\$25,000
Renter relocation	\$85,000
Owner relocation	\$170,000

Table 1: *Approximate cost of demolition*

Optimize the Impact

- Different objectives
  - Same budget

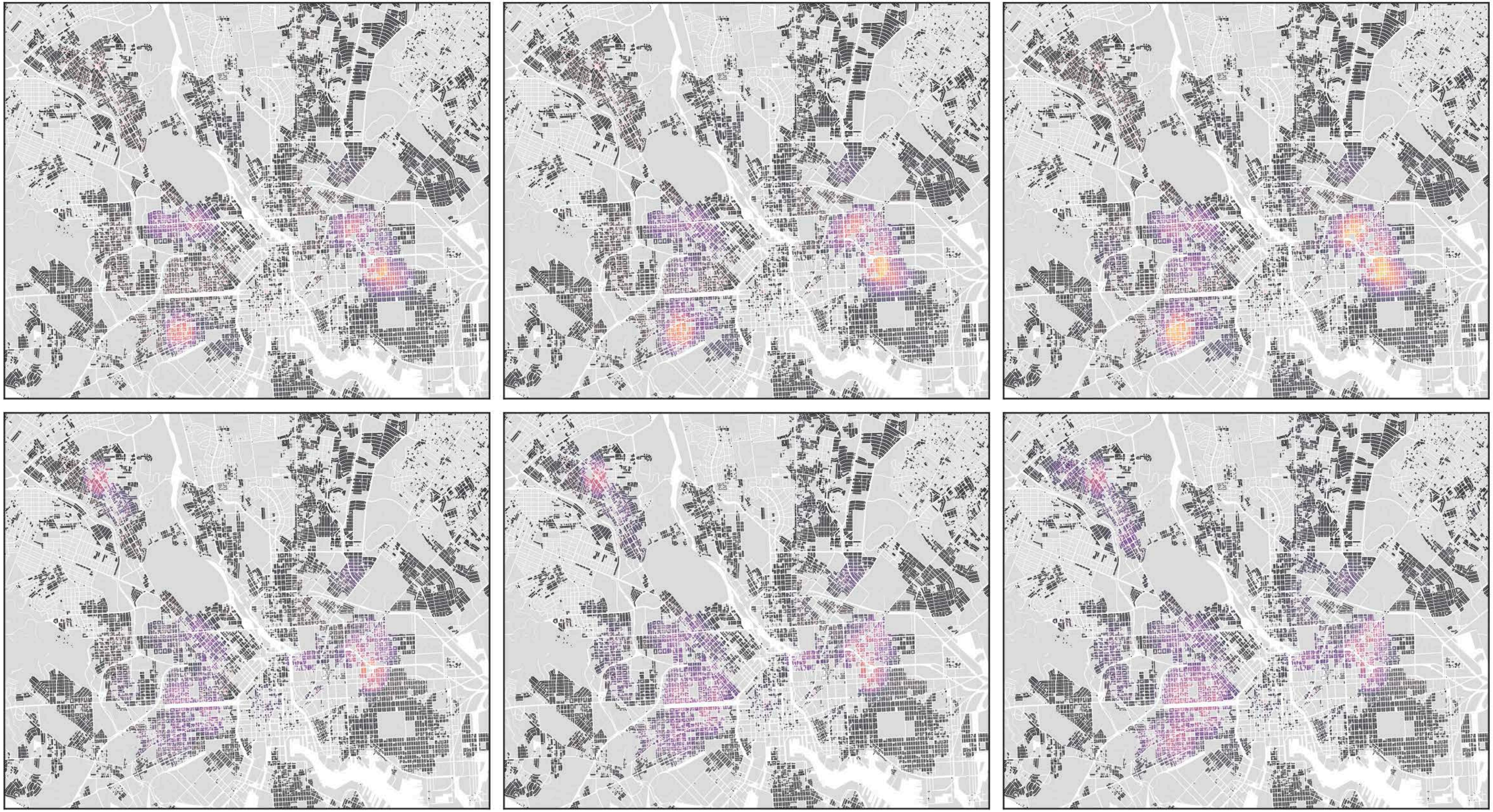
- Advanced tools
  - For decision makers

Lenny Fan
Amitabh Basu
Phil Garboden

5 million

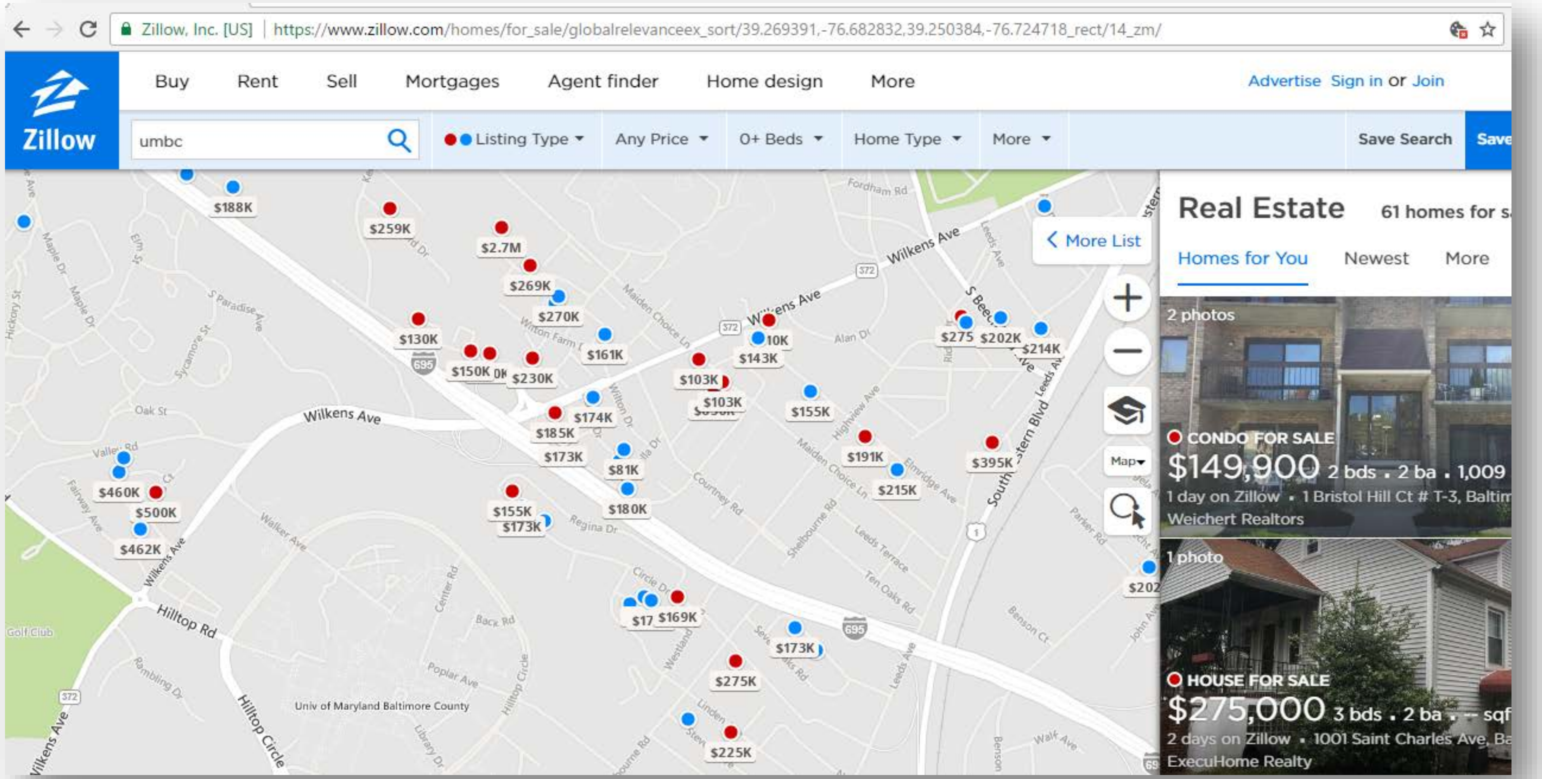
10 million

20 million



Price

- Longitudinal data
- Environment
- Prediction
- Machine Learning



Ambitious Next Steps

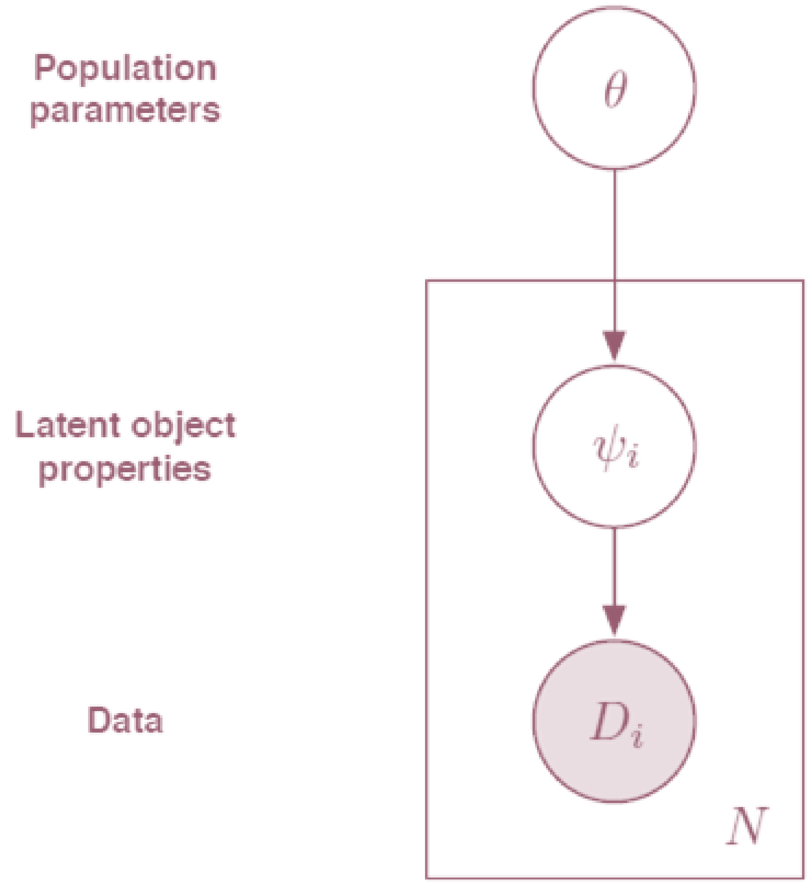
- occupied house › vacant house
- vacant house › rehabilitated house
- vacant house › vacant lot
- vacant lots › community garden
- vacant lots › urban farm
- vacant lots › park

Ben Seigel (21CC)
Katalin Szlavecz
Ben Zaitchik
Keeve Nachman
Katie O'Meara (MICA)

Spatiotemporal Multi-Level Modeling

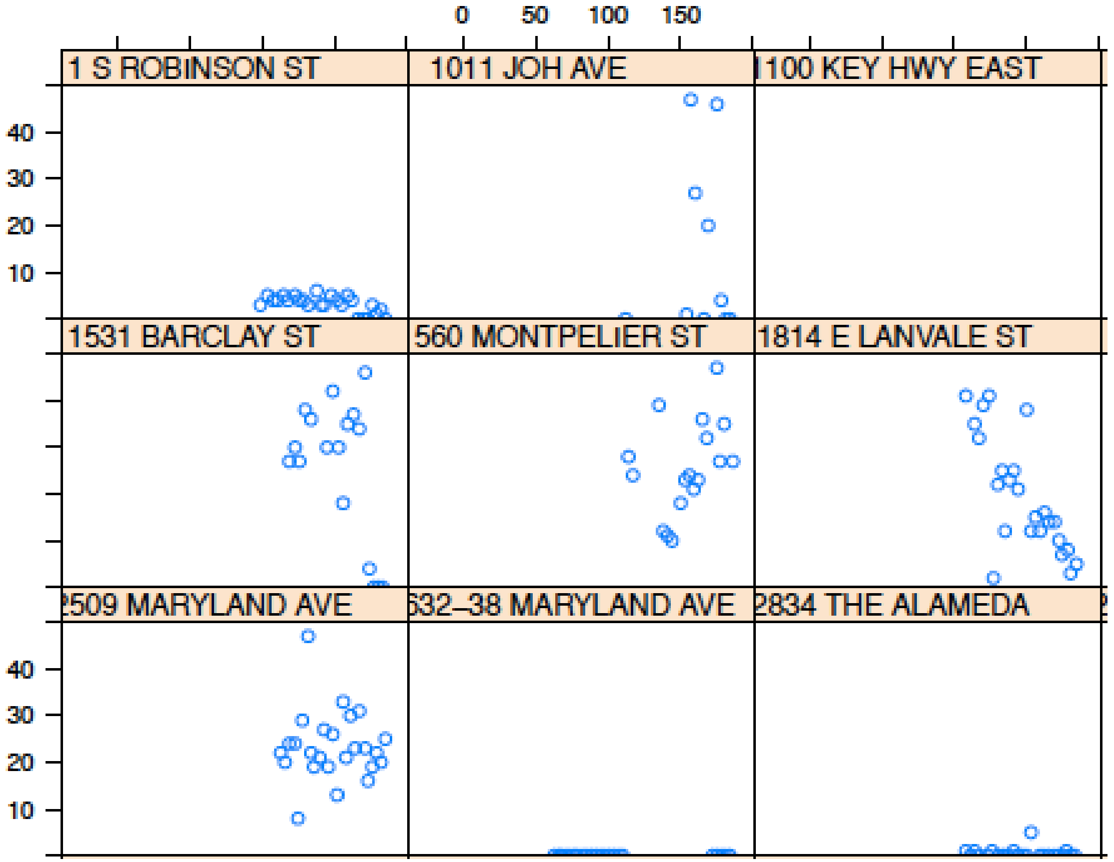
- Hierarchical Bayesian statistics
- Include all aggregated data
- Joint inference for the
  - Individual houses and
  - Ensemble distributions

Mengyang Gu



Predicting Unoccupancy

- Time-series data
  - Water usage
  - BG&E usage
  - USPS
- Proxy for occupancy



Phil Garboden
Hana Clemens

Satellite View

- Missing roof?
- Blue tarp = holes?

Image behind the Atmosphere

- Looking up!
  - Astronomy images
  - Blurred exposures
- We solve for it
  - For high-res details

Matthias Lee
Charlie Gulian
Rick White

Coadded Image

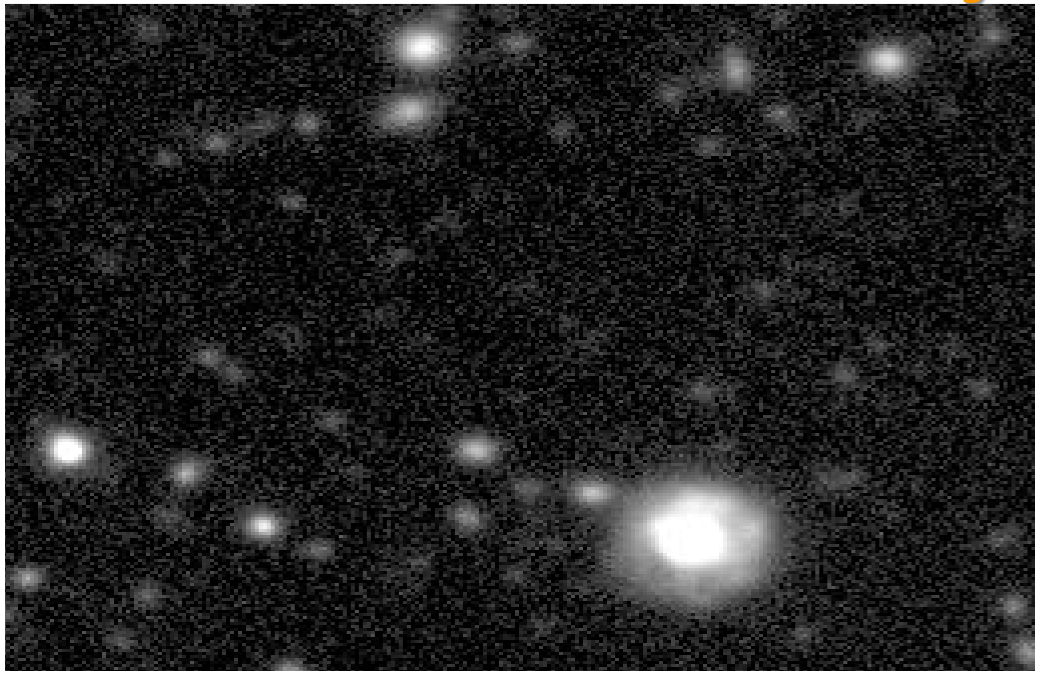
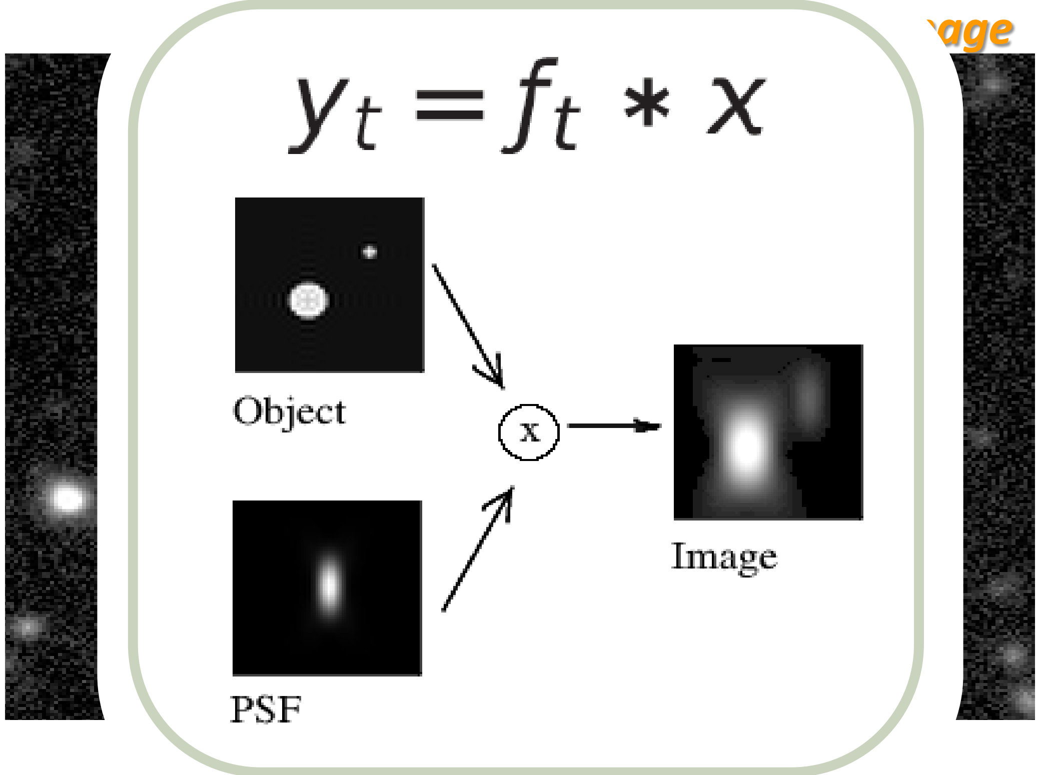


Image behind the Atmosphere

- Looking up!
  - Astronomy images
  - Blurred exposures
- We solve for it
  - For high-res details



Matthias Lee
Charlie Gulian
Rick White

Image behind the Atmosphere

- Looking up!
  - Astronomy images
  - Blurred exposures
- We solve for it
  - For high-res details

Matthias Lee
Charlie Gulian
Rick White

Deconvolved Image

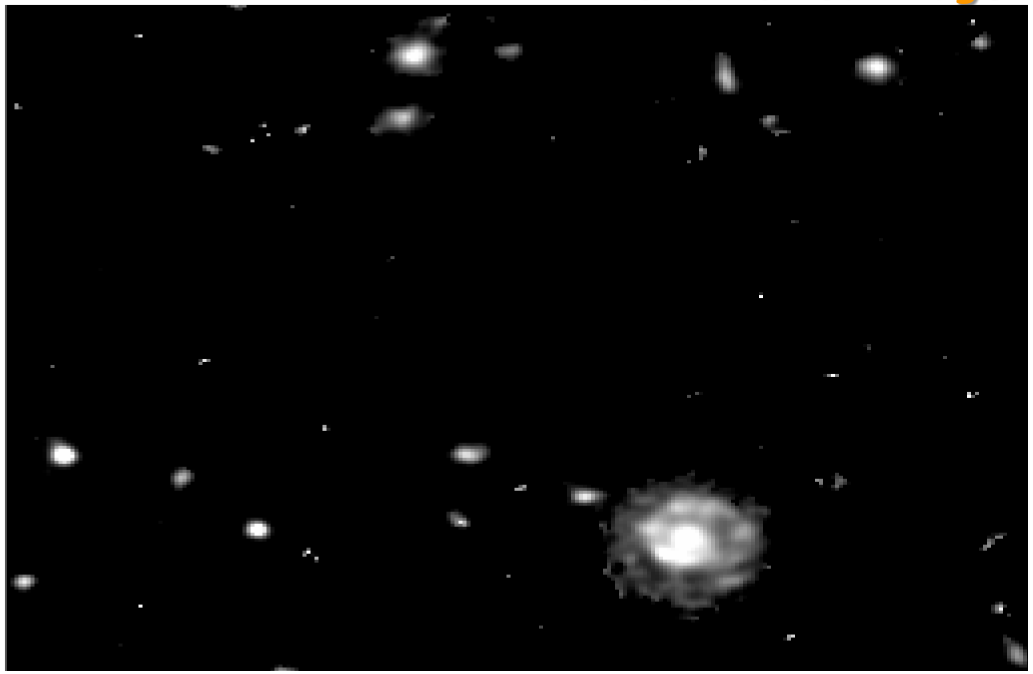
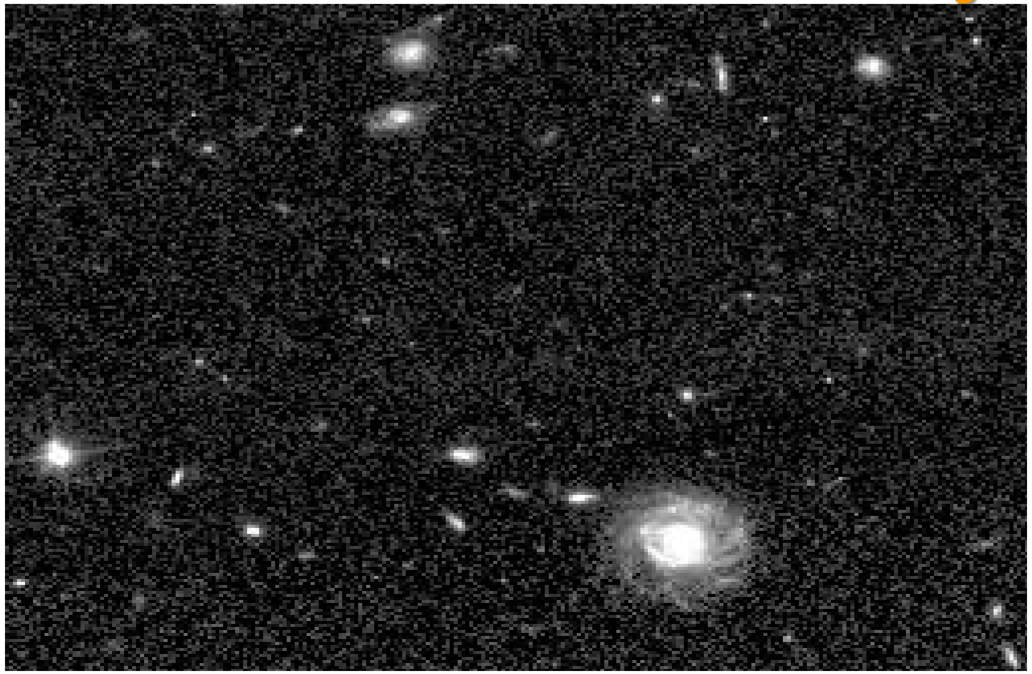


Image behind the Atmosphere

- Looking up!
  - Astronomy images
  - Blurred exposures
- We solve for it
  - For high-res details

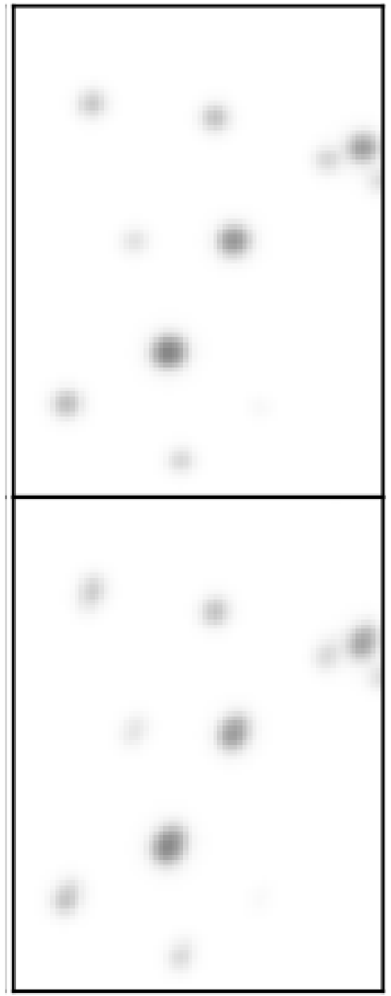
Hubble Image



Matthias Lee
Charlie Gulian
Rick White

Differential Chromatic Refraction

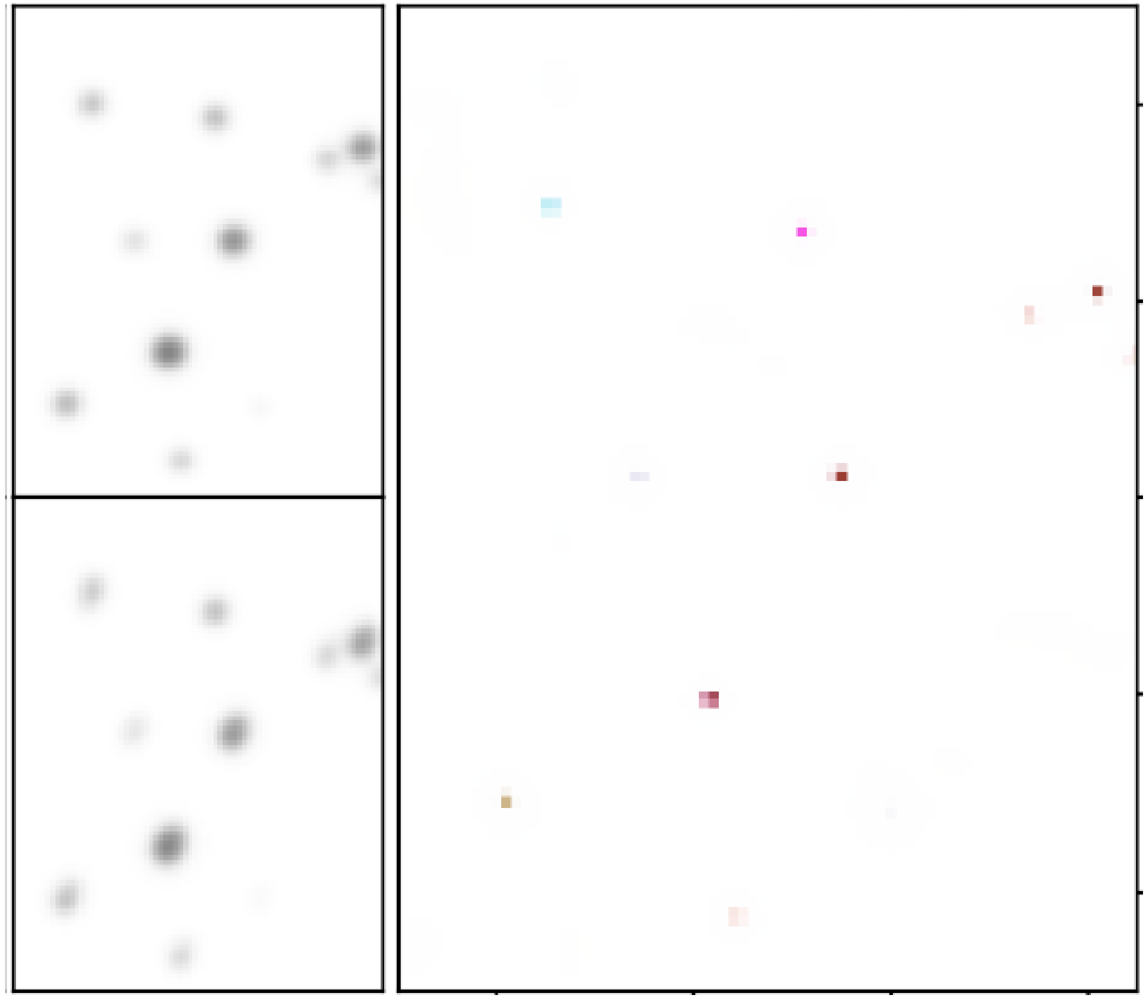
- Even colors!



Matthias Lee
Andy Connolly
Charlie Gulian

Differential Chromatic Refraction

- Even colors!



Matthias Lee
Andy Connolly
Charlie Gulian

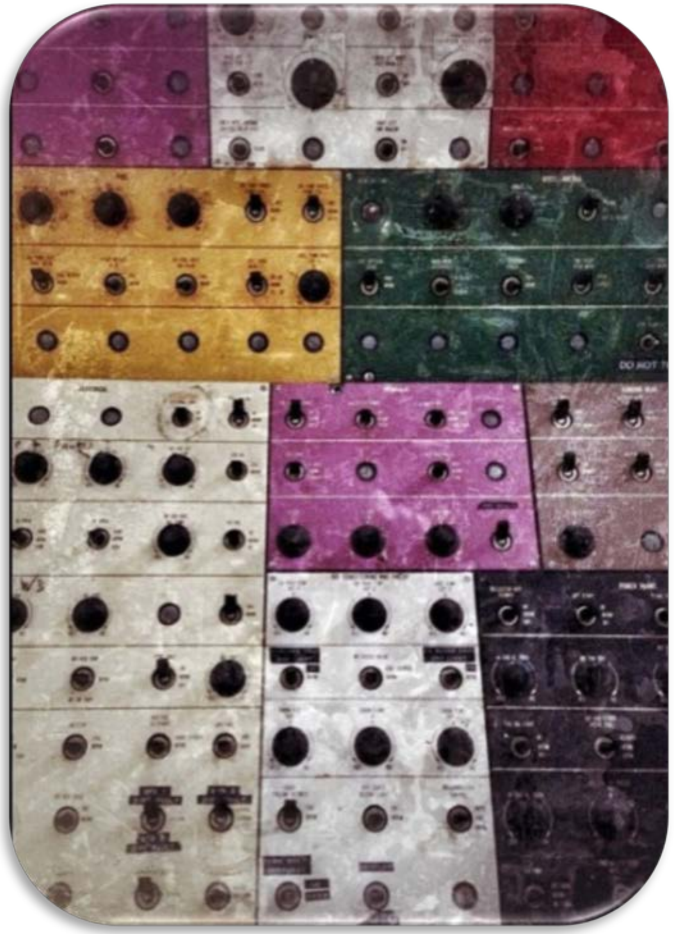
At the Heart…

- Applied Math & Stats
  - Data mining
  - Statistical modeling
  - Machine learning
  - Optimization
  - Bayesian inference
- Data-Intensive Science
  - Hardware platforms
  - Software solutions
  - Streaming algorithms
  - Database technologies
  - GIS tools & indexing

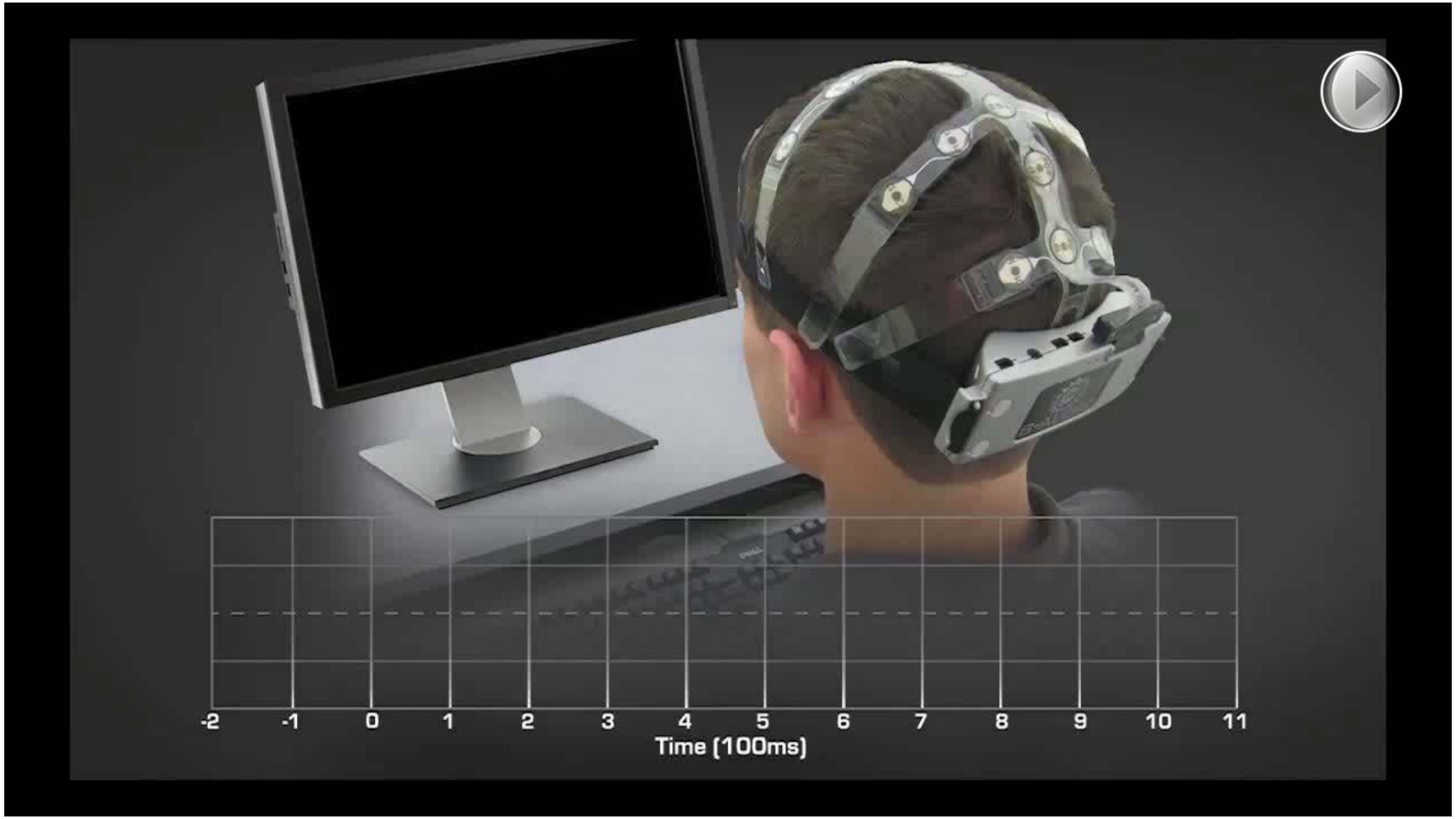
Limitations of Machine Learning

- Many methods to choose from
  - And more knobs to tweak
- Latching on known features
  - Manual intervention to refine
- What's left in the data?

Missing the Human in the Loop!

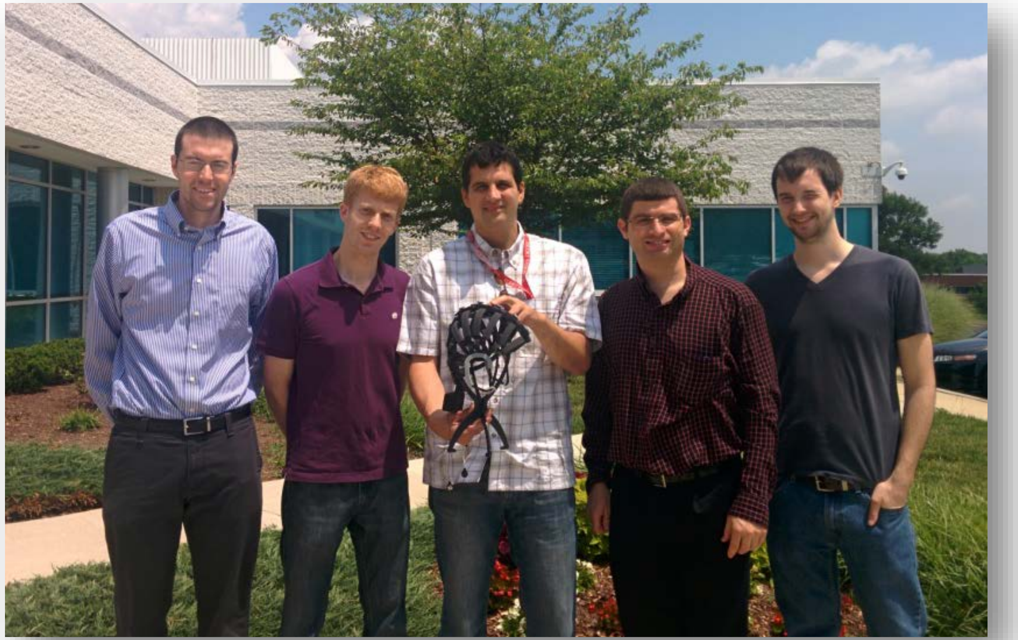


Use the Brain's Detection Power



Rapid Serial Visual Presentation

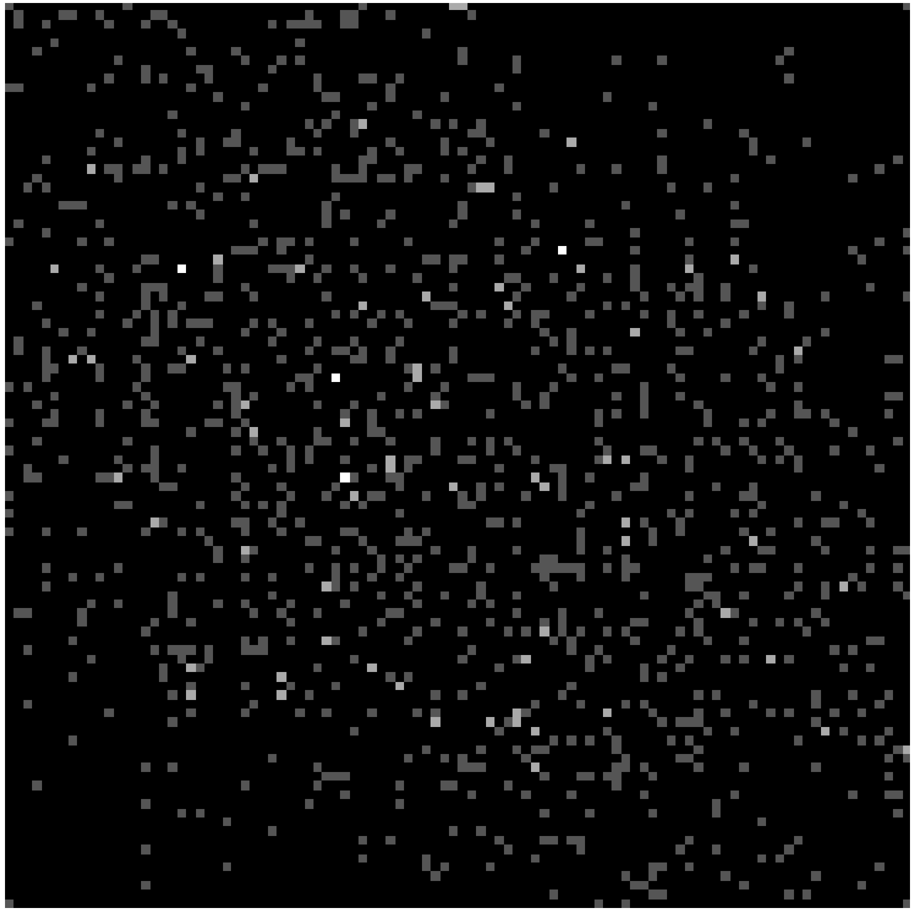
- Current state-of-the-art is binary classification
  - Target / Distractor
- We look for <u>the interesting</u>
  - Dynamic behavior of brain: looking for new



Nick Carey

Human-Machine Co-Learning

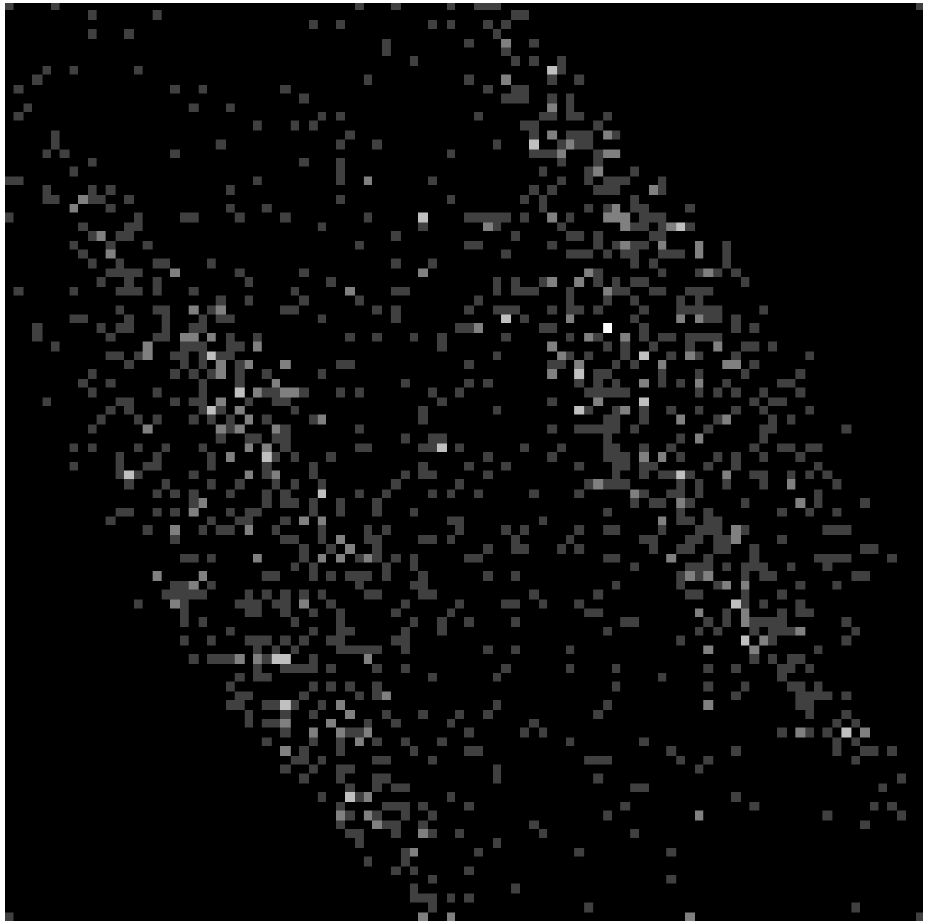
- Hide wireframe of 3D cube in high-D
  - Looks like noise
  - Random projections



Nick Carey

Human-Machine Co-Learning

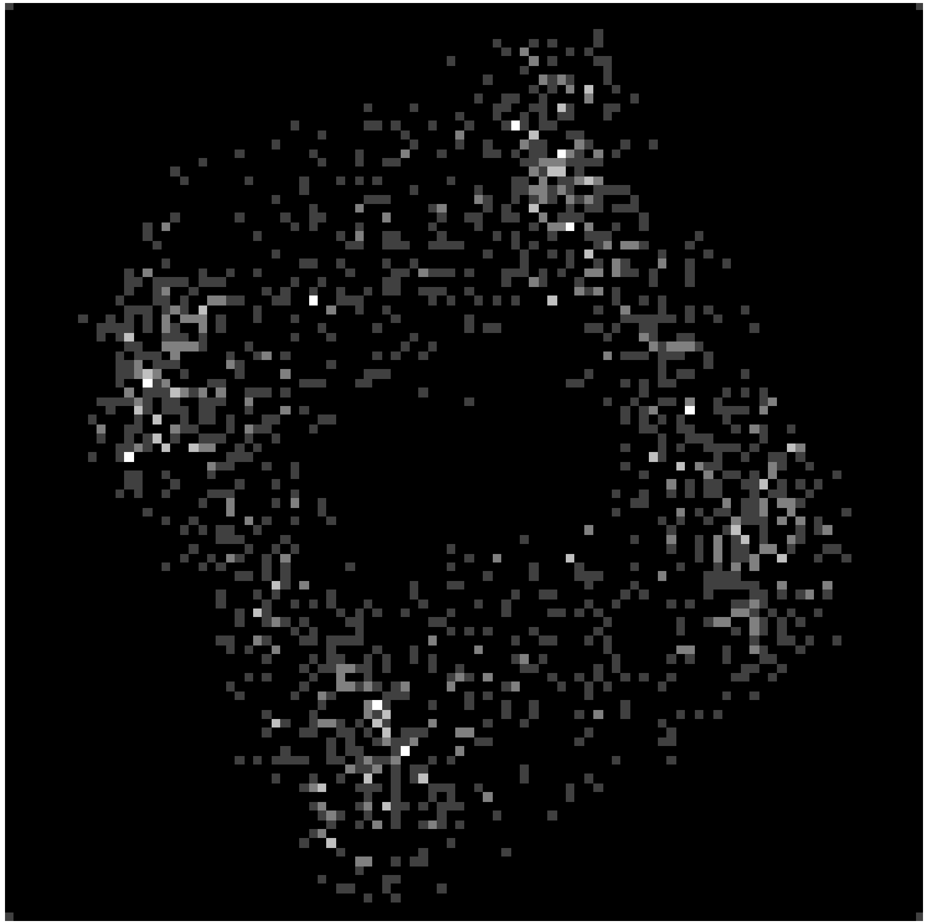
- Hide wireframe of 3D cube in high-D
  - Looks like noise
  - Random projections
  - Trigger to explore locally



Nick Carey

Human-Machine Co-Learning

- Hide wireframe of 3D cube in high-D
  - Looks like noise
  - Random projections
  - Trigger to explore locally
  - Converge on better view



Nick Carey

Human-Machine Co-Learning

- Hide wireframe of 3D cube in high-D
  - Looks like noise
  - Random projections
  - Trigger to explore locally
  - Converge on better view

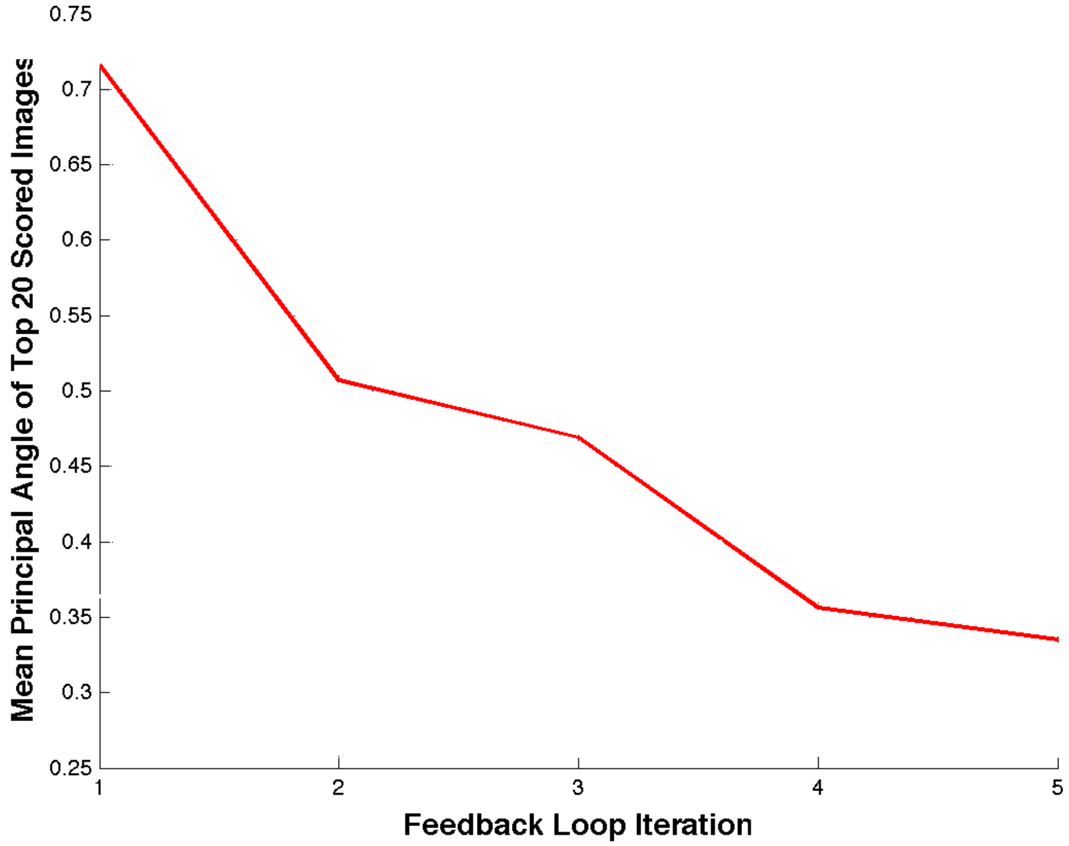
Subconscious Navigation!

Nick Carey

Human-Machine Co-Learning

- Hide wireframe of 3D cube in high-D
  - Looks like noise
  - Random projections
  - Trigger to explore locally
  - Converge on better view

Subconscious Navigation!



Nick Carey

Summary

- Promising first steps
  - With direct applications already deployed
- Common data infrastructure & approaches
  - Surprisingly similar, e.g., across astro/city
- Ambitious future plans
  - Need help! And need more data…