Whose Analysis, Whose Expertise: Partnering for Data Analytics for Small Cities

Kelly Joyce, Drexel University
Susan Sterett, University of Maryland, Baltimore County

Big Data

- The 3 Vs: Velocity, Variety and Volume

- Data veracity added later

- Data is   er activity

The Growth of Data Science Programs

MASTER'S IN DATA SCIENCE

Arizona State's MS in

online GMAT waivers available!

The Harvard Faculty of Arts and Sciences is pleased to announce the launch of a new Master of Science (SM) degree in Data Science. The new degree, under the joint academic leadership of the Computer Science and Statistics faculties and administered through the Institute for Applied Computational Science (IACS) at the John A. Paulson

The **Data Science Initiative at Brown offers a new master's program (ScM)** that will prepare students from a wide range of disciplinary backgrounds for distinctive careers in Data Science. Rooted in a research collaboration among four very strong academic departments (**Applied Mathematics**, **Biostatistics**, **Computer Science**, and **Mathematics**), the master's

Syracuse University

Online Master's in Data Science

And Data Science Boot Camps

Data Science Initiatives

Big Data Regional Innovation Hubs (NSF)

White House National Big Data Research and Development Initiative

$5 million awarded in 2015, builds on previous solicitation

Big Data Spokes

Big Data Regional Innovation Hubs: Establishing Spokes to Advance Big Data Applications (BD Spokes)
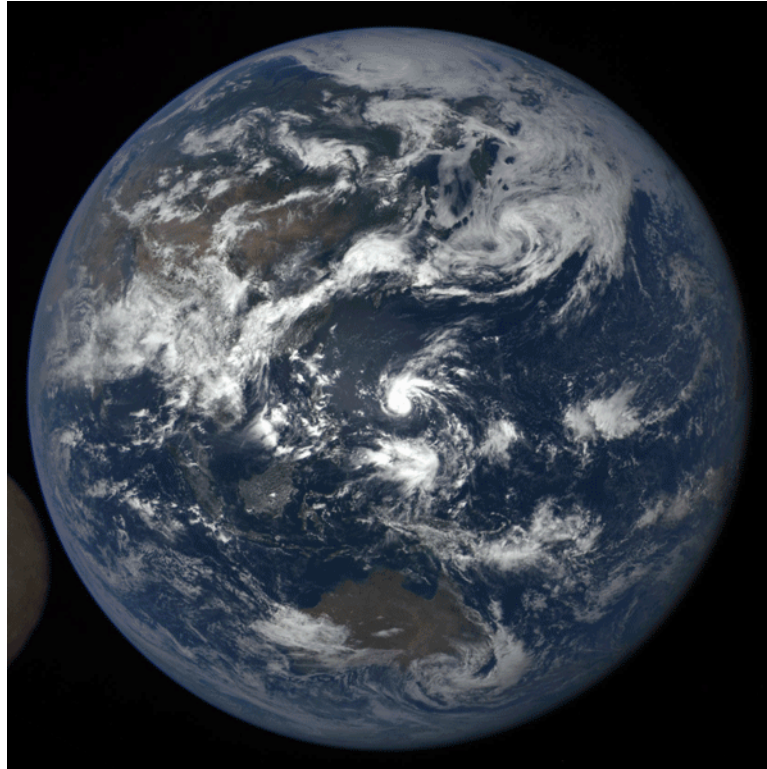
PROGRAM SOLICITATION
NSF 16-510

Upcoming deadlines:
Jan 12, 2016; February 25, 2016

Separate Worlds? Or Partnerships? If partnerships, what kind of partnerships exist?

ANALYTICS TEAM

The Citywide Analytics Team is the City of Boston's central data organization. We use data and technology to improve our City.

Translational Data Science
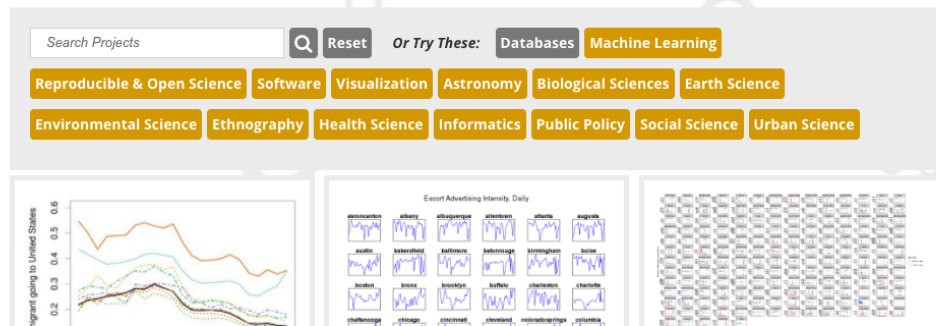
Data Science for Social Good

Partners

MetroLab Network includes 41 cities, 4 counties, and 55 universities, organized in more than 35 regional city-university partnerships. Partners focus on research, development, and deployment (RD&D) projects that offer technological and analytically-based solutions to challenges facing urban areas including: inequality in income, health, mobility, security and opportunity; aging infrastructure; and environmental sustainability and resiliency.

Changes in Expertise and Accountability

How are disparate records changed into integrated data (Brayne, 2017)?

What does it mean to govern through open data? (Ruppert, 2015)
"Inevitably black boxes its own making and closes off alternatives"

How do we understand the work processes that change records into data?

What questions get asked?
What gets bracketed off?

Partnering with Universities: Arlington, VA

(with thanks to Andrew d'Huyvetter, Planner, County of Arlington)
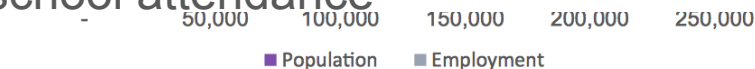
Demand for services: predictive analytics for school attendance
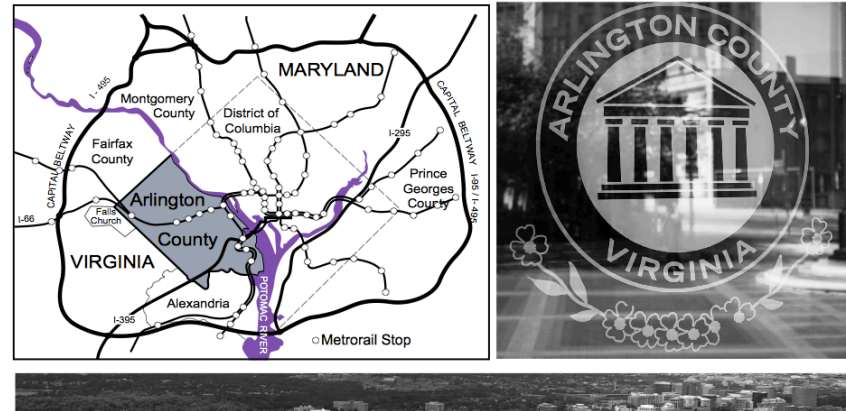
Affluent county: in the top 10 of US Counties

222,000 residents

Collaboration with Social Data Analytics Lab: computer science expertise

Databases established for one purpose: takes transform into useful data

50,000   100,000   150,000   200,000   250,000

■ Population   ■ Employment

ARLINGTON IN THE DC METRO AREA

Counting Housing, Counting Jobs

Why? Demand for schools increasing

Neither Census nor ACS accurate enough for school enrollments

Apartments,
single family homes,
multifamily condos

Data cleaning:
addresses not formatted
the same way, even for single
family homes

Permits, tax records, billing, real

HOUSING UNITS

January 1, 2017 Housing Units (Planning Division Estimate)		114,000
2010 Housing Units (U.S. Census Bureau)		105,404
2000 Housing Units (U.S. Census Bureau)		90,426

TYPE OF HOUSING UNIT (2000 and 2017)

	2000		January 1, 2017	
Single-family, detached	27,668	30.6%	28,500	25.0%
Single-family, attached	10,284	11.4%	11,100	9.7%
Multi-family (3+ units)	52,373	57.9%	74,300	65.2%
Other	101	0.1%	100	0.1%
Total Housing Units	**90,426**	**100.0%**	**114,000**	**100.0%**

Source: U.S. Census Bureau, 2000 Census of Population and Housing and Arlington Planning Division Janury 2017 estimates.

INCOME

ARLINGTON		WASHINGTON METROPOLITAN AREA	
2017 Median Household Income*	$110,700	2016 Median Family Income for Family of Four	$108,600
2017 Per Capita Income*	$89,300		

Source: Arlington County Planning Division 2017 estimates.

Source: U.S. Department of Housing and Urban Development, Estimated Median Family Incomes for FY 2016, March 2016.

2016 Effective Buying Income	$10.24 Billion

Source: ESRI

*Arlington County Planning Division 2017 estimated income is extrapolated from the 2015 ACS household income of $106,768 and the 2015 BEA per

What Works?

Time investment

Money: collaboration with SDAL

Collaboration between IT, Mapping, Housing departments in the County

Expenditure of political capital: sharing data

What's Difficult about Counting Jobs?
What's Creative?

Federal presence

Misreports of employment

Misreports of occupancy

Merging information about water use with models of how people use water: build in assumptions about how water gets used

Big Data (Unintentionally) Goes Wrong

I started to look at how apps might be reproducing this form of bias. So there's an app that's actually quite fascinating called Street Bump which is used in the Cambridge area. Now, this was designed for really good reasons. It's actually very smart. What it does is it basically tracks your accelerometer and your GPS data as you're traveling down the road so when you hit a bump it'll say, oh, there's a bump, and if a few other people hit the same bump, then the city knows to send out a road crew and–…But what I started to do is to look at who owned smartphones in the Boston area, and what, unsurprisingly enough, it tends to map with people with more disposable income and also younger audiences. Particularly if you look at the over 65s, you'll find that in some cases smartphone penetration is as low as 16%.

So what you end up doing is having an app which in many cases is accentuating the signals from areas that have younger, wealthier residents, and we're not getting the signals from areas where we have people who have less money or if they're generally older populations. So what that could mean is cities could start to be resourced and repaired based on signals that are really skewing the population towards those who are already are pretty privileged base." Kate Crawford, Microsoft
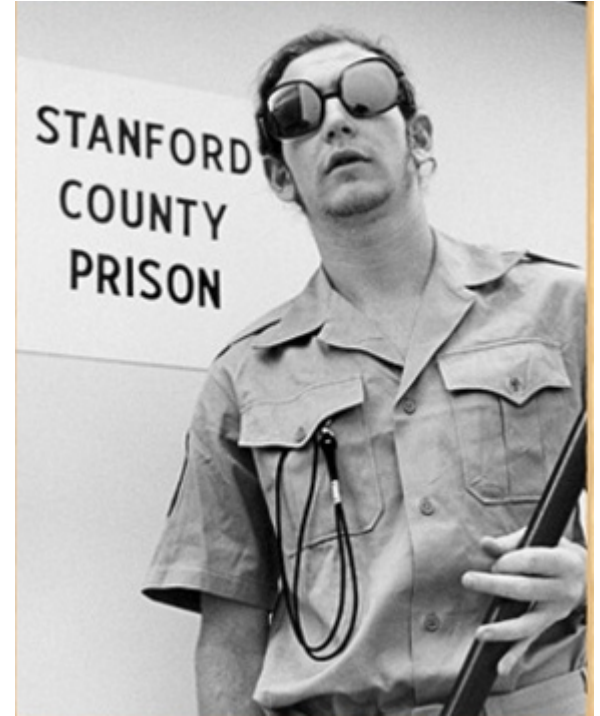
Where are the Social Sciences?

Philadelphia Behavioral Science Initiative

100 Plus Years of Learning and Mistakes

Stanford Prison Experiment (1971)

Bureau of Justice Statistics (BJS)

*Obligated to only collect data that they will use (Fair Information Practices Principles)

*Long standing practices to protect privacy

Interviewed people in prison about assaults

Include filler questions so people who were assaulted and people who were

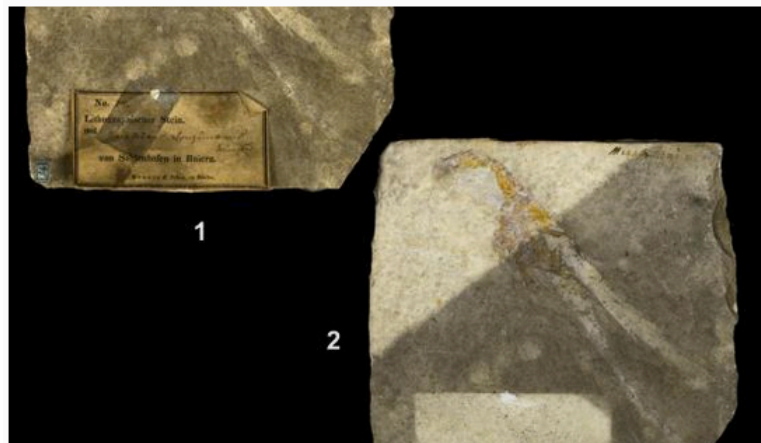not assaulted were in the interview room for the same length of time

Community-Based Participatory Research

Action-based Research

Applied-learning Research

Philadelphia Health & Environment Ethnography Lab

Home   About   People   Projects   Teaching   Contact

Asthmatic Spaces

NSF Funding for The Asthma Files

1

2

Queries

What questions get dropped out? Schooling and inequality and demand for schools

When might data analytics discriminate (Barocas et al)?

Where are the people in data analytics?

Merged datasets can reveal people or individual institutions

Principles of fitness for use (Jones, 2014)
Uncertainty and fragility of conclusions
Theorizing how people and organizations use information

Is Collaboration an Ethical Issue?

*Data Validity

*Meaningful Data

*Data that addresses inequalities, not exacerbates them

*Human subjects

Multidisciplinary Research

Challenging

    Words fall apart

    Disciplinary rewards, status and resources vary

Rewarding

    Different kinds of expertise

    Meaningful outputs

Questions to Consider

*What types of relationships are happening between small cities and data analytics?

*Examples of productive relationships between universities and small cities?

*Examples of pitfalls to avoid?

*Is there a role for social science?

Thanks to