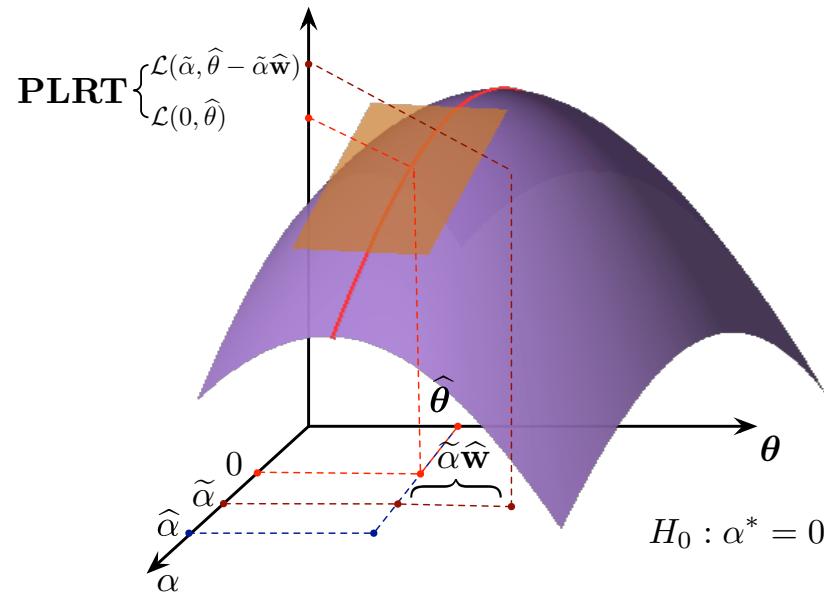


Statistics at Penn State



- ▶ Joiner, Brian J. (1975), Living Histograms, *International Statistical Review* 43(3): 339–340.

High-Dimensional Inference

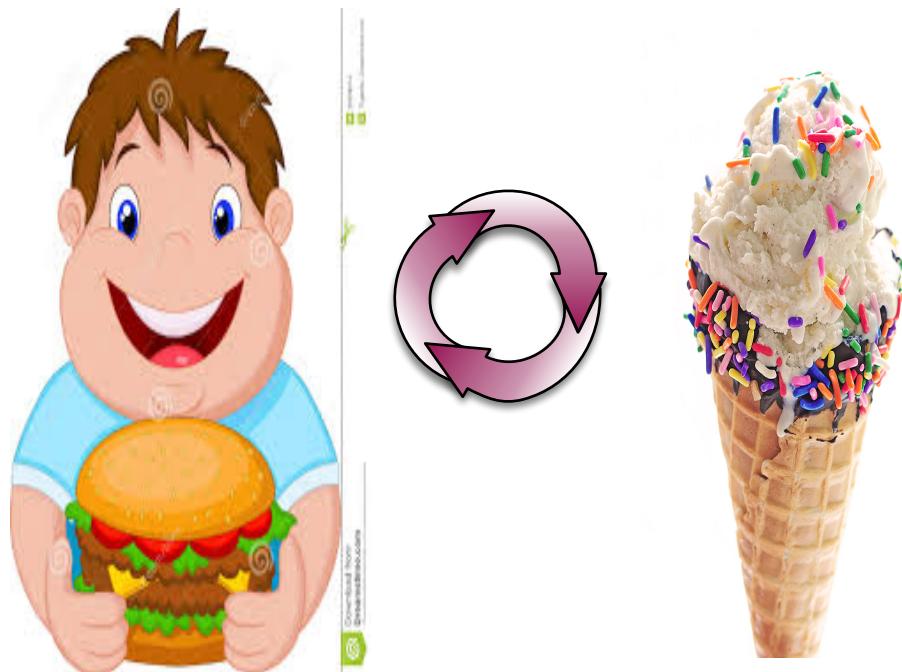


Optimization in Machine Learning

Prison Break	House of Cards	Game of Thrones	True Detective	Big Bang Theory

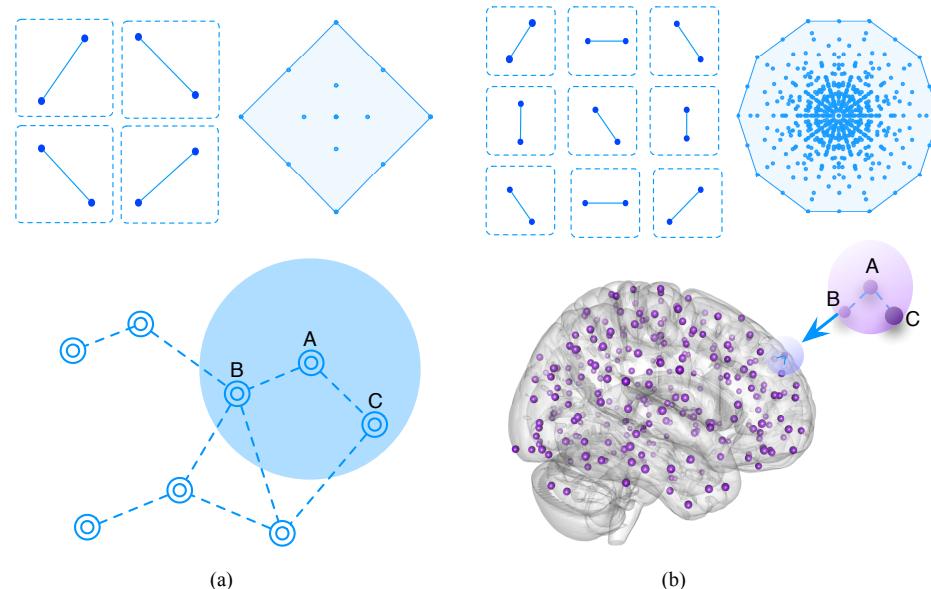
Ethan				5
Edward	2			4
Iain	4	3	3	5
Adam	2			5
Cagin	5	4	1	3

Collaborative Research: Obesity and Sweets



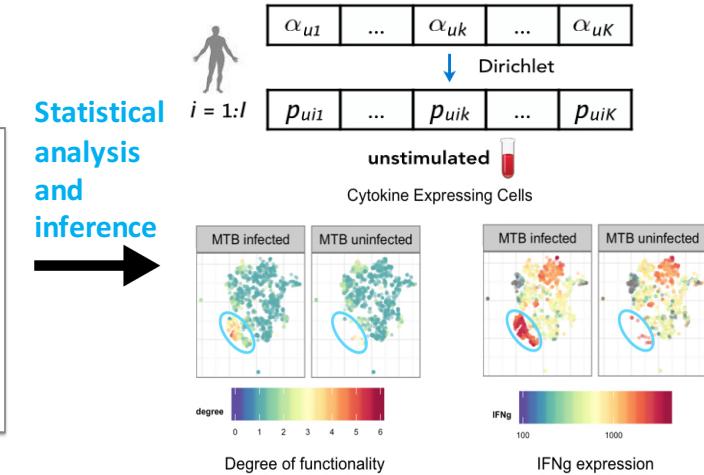
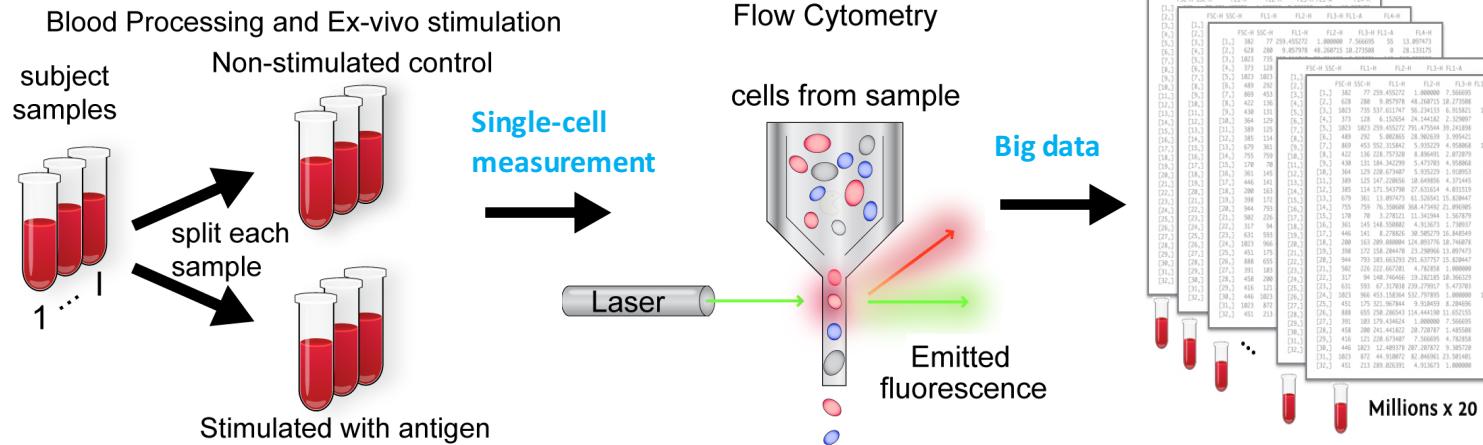
Nonconvex Optimization in Statistics

$$\min_{\beta_j} \sum_j \mathcal{L}_j(\beta_j) \text{ s.t. } \sum_j \|\beta_j\|_0 \leq k$$



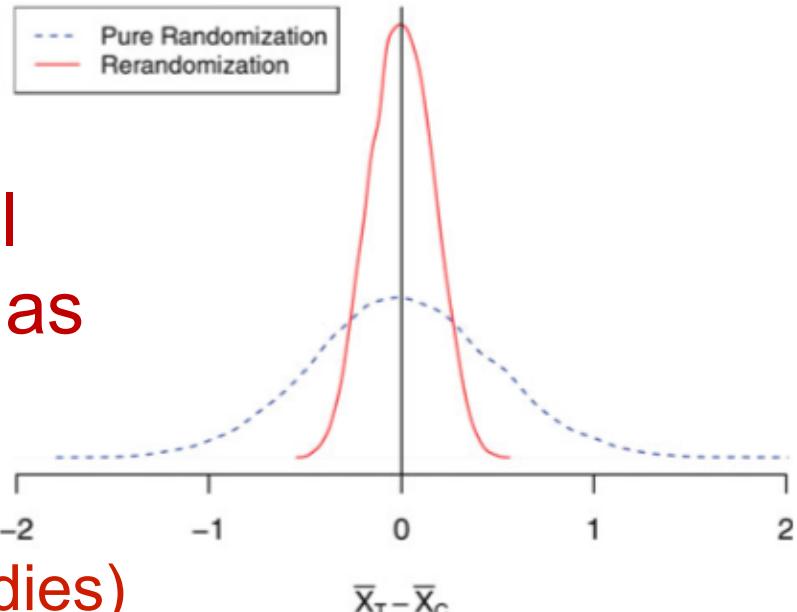
Lynn Lin's Research Interests

- Bayesian statistics and computation, integrative analysis, statistical methods for high throughput single-cell data
- Applications in immunology and vaccine research:



- How many cell types?
- How to visualize cell-to-cell heterogeneity?
- Biomarker discovery, e.g., what cell types are in response to certain therapy?
- All these above lead to more effective disease therapy and better vaccine development.

Kari Lock Morgan

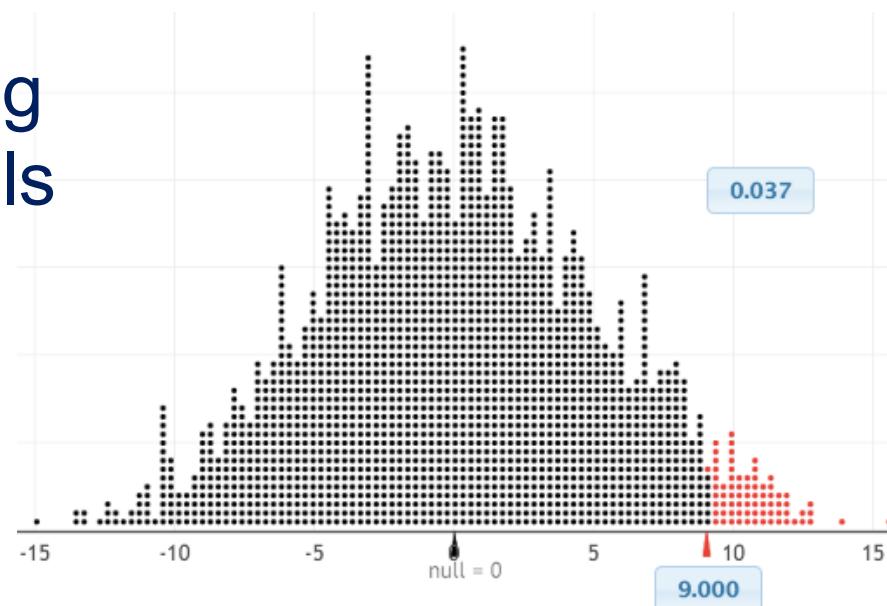


Causal inference: better causal conclusions by making groups as similar as possible at baseline

- By design (experiments)
- By analysis (observational studies)

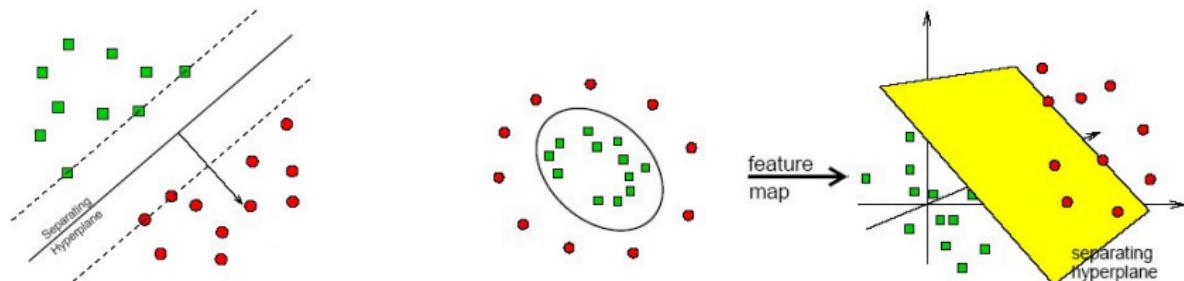
Statistics education: improving statistics education at all levels

- Simulation-based inference
- Teacher preparation
- Pedagogy



Bharath Sriperumbudur's research overview

- ▶ **Research interests:** Non-parametric statistics, statistical learning, optimization, machine learning and signal processing.
- ▶ Primarily, I study the mathematics of learning from data.

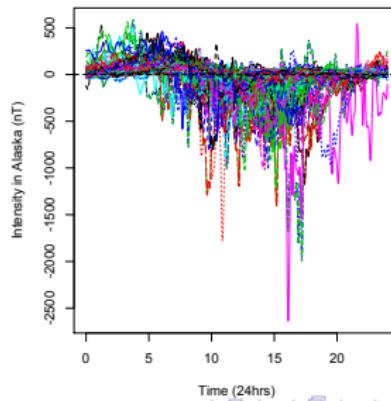
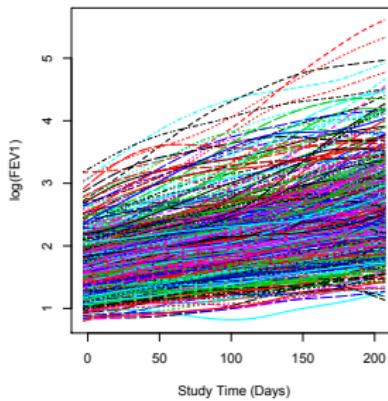


Linear and non-linear classification

- ▶ Contact: 314 Thomas (Office)
bks18@psu.edu
<http://www.personal.psu.edu/bks18>

Matthew Reimherr - Functional Data Analysis

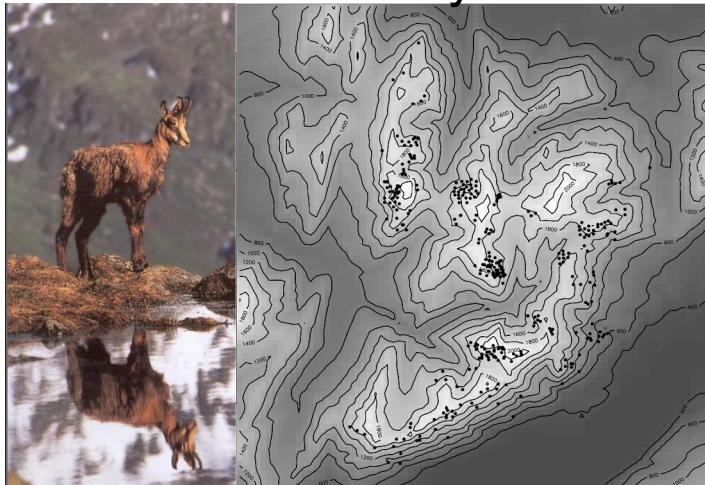
- Data: $\{X_n(t) : 1 \leq n \leq N, t \in [0, 1]\}$
- Varied Applications:
 - Childhood asthma management project (left slide)
 - Framingham heart study
 - Geomagnetic storms in magnetosphere (right slide)
 - Precipitation and temperature curves
 - Stocks and other financial instruments
 - Kinesiology study involving soccer kicks
- Fun theory! (probability and statistics in function spaces).



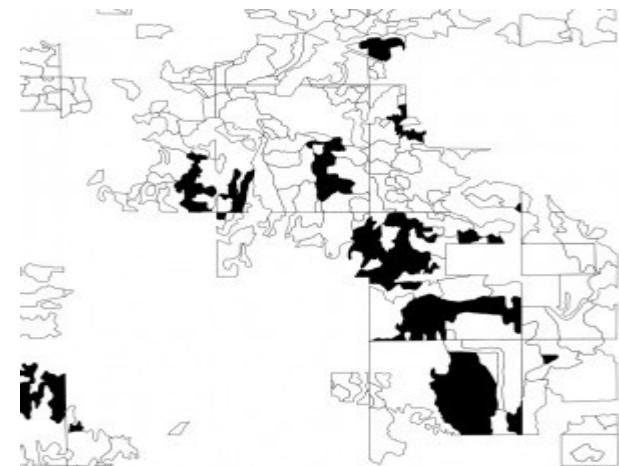
Ephraim Hanks - Thomas Rm 310

Research: Spatial and spatio-temporal statistics. Bayesian statistics.

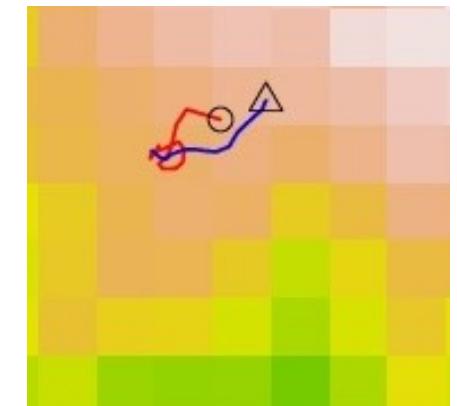
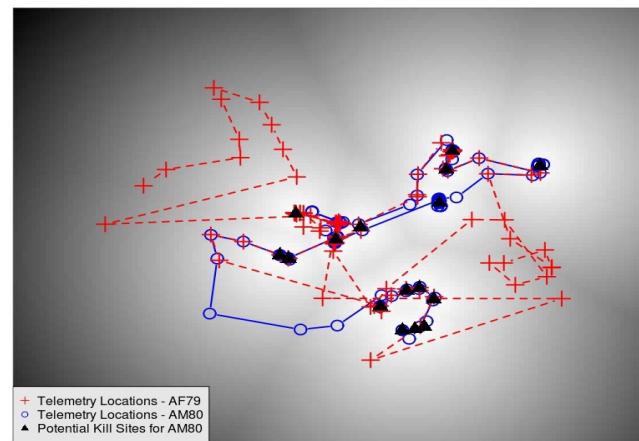
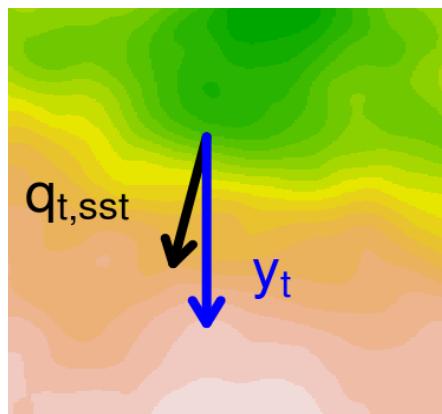
(1) Gene flow for landscape connectivity



(2) Spatial epidemiology

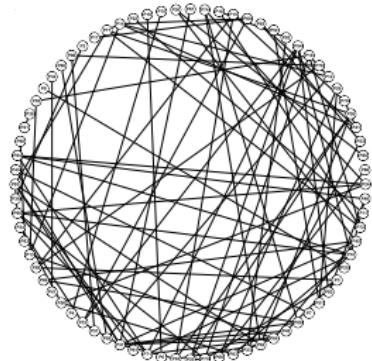


(3) Animal movement in response to environmental stimuli

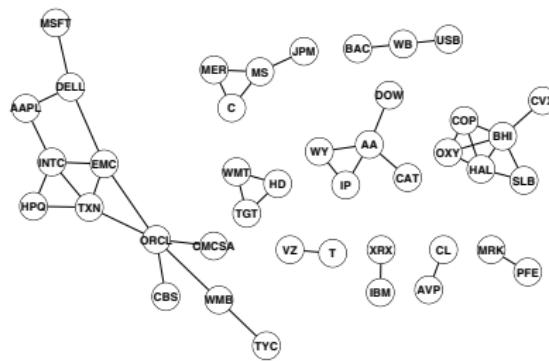


Lingzhou Xue's research overview

- ▶ **Research interests:** statistical learning, high dimensional inference, large-scale optimization algorithms, statistical methods in finance, health science, network analysis, ...



example 1: genetic network

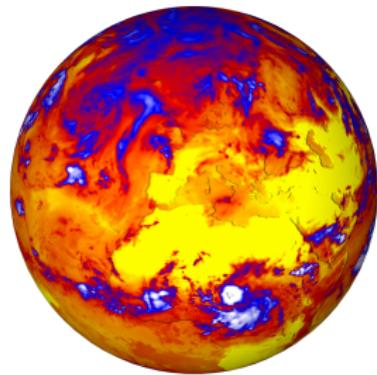


example 2: stock network

- ▶ **Contact:** (office) 318 Thomas; (email) lingzhou@psu.edu; (group webpage) <http://sites.psu.edu/sldm/>

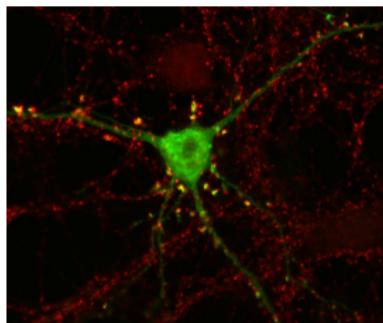
Ben Shaby's research overview

- ▶ I study spatial statistics, especially rare events that live in space (like heat waves or hurricanes). I work a lot with climate scientist types.



2003 European heat wave

- ▶ I also develop methods for large scale simultaneous inference for automated high throughput screens. This work is also acquiring a genetics component.

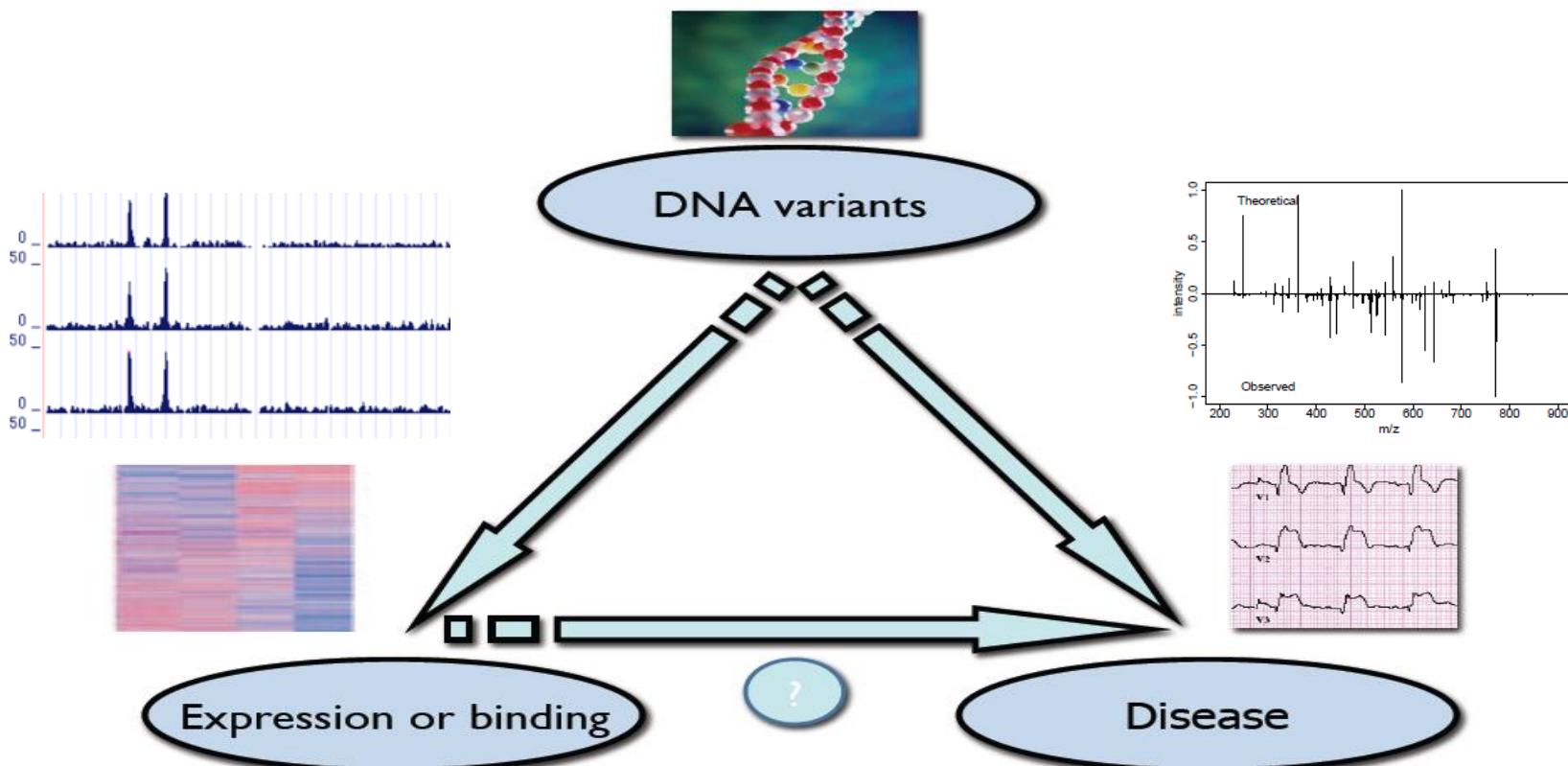


Neuron with Huntington's disease

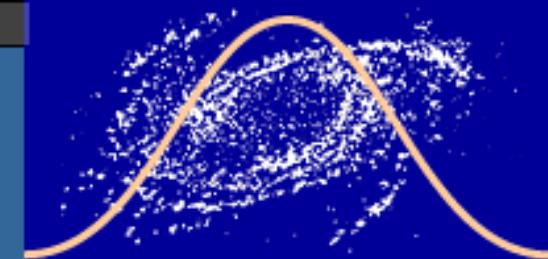
- ▶ **Contact:** (Office) 313 Thomas; (email) bshaby@psu.edu; (webpage) <http://www.personal.psu.edu/bas59/index.html>.

Qunhua Li's research area

Statistical modeling for high-throughput biological data,
mixture models, latent variable models, machine
learning, genomics, proteomics

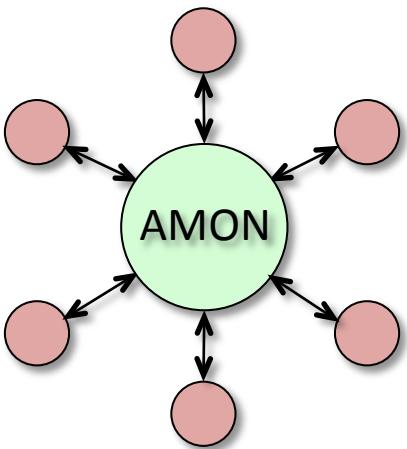
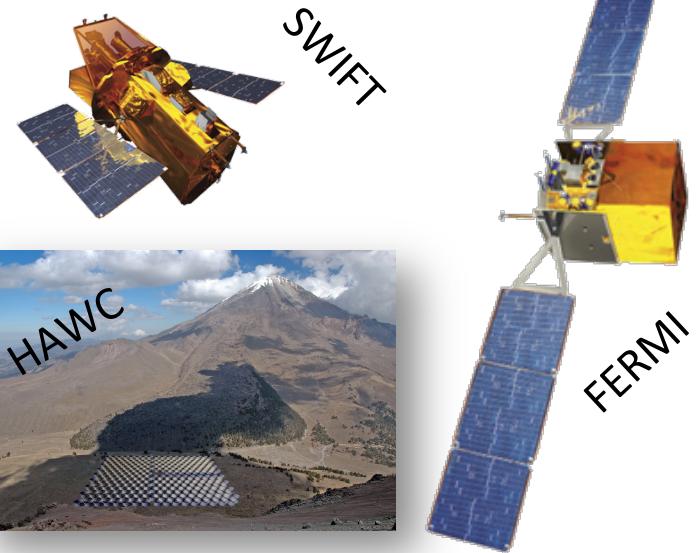


Contact: qunhua.li@psu.edu, 514D Wartik Lab



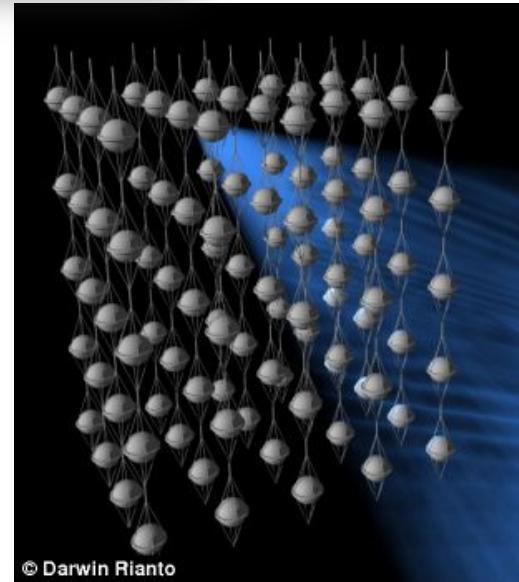
Statistics for Multi-epoch surveys

- **Electromagnetic Force:** Flashes of high energy γ -rays Swift and Fermi missions
- **Weak Nuclear Force:** Neutrinos made by interactions within an exploding fireball
- **Strong Nuclear Force:** High energy protons and neutrons ejected by explosions
- **Gravity:** Shakes the fabric of space-time. Detectable as gravitational waves –LIGO



Requires only N
connections for
completeness

G. Jogesh Babu
babu@psu.edu
 417C Thomas Building
 Group's Website
<http://astrostatistics.psu.edu>



© Darwin Rianto

Program Overview

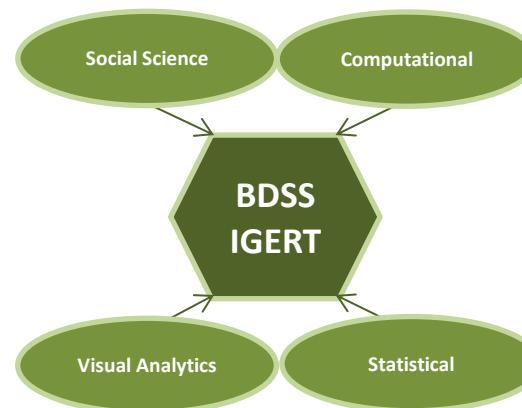
The Big Data Social Science IGERT program (BDSS-IGERT) draws together a diverse interdisciplinary team of researchers to create a new training PhD program in Social Data Analytics, aimed at enabling a new type of scientist capable of meeting emerging big data challenges. Merging statistics, computer science, visual analytics, and social science, the program is intended to establish new models for graduate education and training in a fertile environment for collaborative research that transcends traditional disciplinary boundaries, and to engage students in understanding the processes by which research is translated to innovations for societal benefit.

Sample Student Projects

- Synthetic Data and Privacy in Social Networks
- Detecting Salient Events in Video and Audio Recordings of Dyadic Interactions Within Couples
- Text Analytics of Arabic Parliamentary Speech
- Improving Disaster Response by Geo-locating Tweets
- The Geography of Immigrant Status, Minority Status, and Health Disparities Among Children
- Visualizing Uncertainty in Complex Data
- Predicting Successful Matches in Online Dating Sites
- Forecasting Violent Political Events
- Parallelizing Complex Statistical Models of Social Data



First BDSS-IGERT Trainee Cohort 2012-2014



BDSS-IGERT is a collaboration of the social, computational, statistical, and visual sciences.

Traineeships

Financial Support:

- Two year traineeship
- \$30,000 annual stipend
- Full Tuition
- Health Insurance Benefits
- Allowance for research expenditures, publication, conferences attendance

Responsibilities:

- Additional coursework
- Four semesters of research rotations at Penn State
- Two summer externships
- Competitive challenges
- Collaborative research projects

Program Curriculum

The IGERT curriculum is aligned with the student's goals and provides training, using both coursework and research involvement, to extend and supplement the core PhD discipline.

Core seminars

- Approaches & Issues in Big Social Data
- Approaches & Issues in Social Data Analytics

Analytics, e.g.:

- Data Mining
- Machine Learning
- Visual Analytics

Ethics & Scientific Responsibility. e.g.:

- Privacy in Statistical Databases
- Data Privacy, Learning, and Games
- Big Social Data and the Law
- The Information Environment

Social Data Analytics for social science students (taught in non-social science departments), e.g.:

- Network Science
- Vision-Based Tracking
- Pattern Recognition
- Information Retrieval and Organization
- Web Analytics
- Spatial Analysis

Social Data Analytics for students outside the social sciences (taught social science departments), e.g.:

- Modeling Interdependent Data
- Political Event Data and Forecasting
- Social Network Analysis
- Causal Inference
- Multilevel Modeling
- Spatial Demography
- Intensive Longitudinal Data
- Geospatial Science in Anthropology

Computation, Bioinformatics and Statistics (CBIOS) Predoctoral Training Program



The Pennsylvania State University, University Park, PA 16802, USA

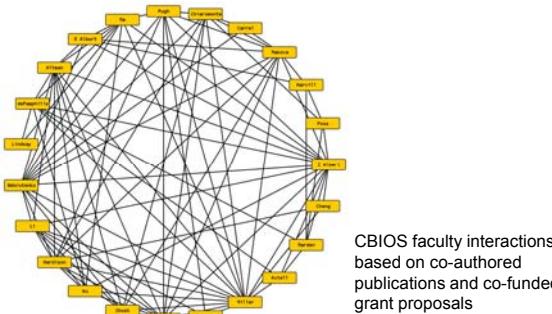


- The CBIOS training program focuses on preparing a cadre of investigators to think innovatively and keep pace with the quickly evolving landscape of high throughput genomic technologies
- The program faculty are interdisciplinary and highly collaborative, with expertise in computation, bioinformatics, statistics, functional, medical, and evolutionary genomics
- Pre-doctoral trainees selected early in their graduate program are provided with intensive training for two years with individualized curriculum, interdisciplinary and collaborative research, and expanded professional development opportunities
- Learning these discipline-crossing skills makes trainees competitive for future careers in emerging and rapidly advancing fields of comparative, systems, statistical and medical genomics

Training Goals

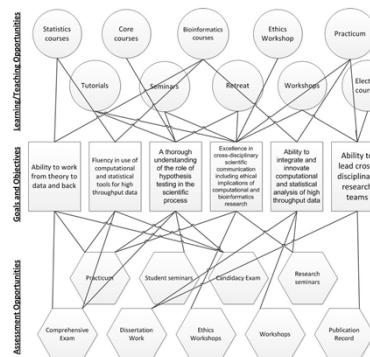
- To facilitate students passing the critical threshold from user to innovator in analysis of high throughput genomic data, and from functioning to leading in an interdisciplinary research domain
- To facilitate cross-disciplinary training of students from diverse scientific background
- To create a collaborative infrastructure of faculty who conduct collaborative research with trainees, serve on joint dissertation committees, team teach courses, and participate in journal clubs and seminar series

Collaborative Training Faculty



CBIOS faculty interactions based on co-authored publications and co-funded grant proposals

Goals, Learning and Assessments



Program Curriculum

- Required Courses
 - Critical Analysis of Bioinformatics and Genomics Research Topics
 - CBIOS Practicum
- Select one from prescribed courses in the following clusters
 - Bioinformatics Cluster
 - Statistics Cluster
 - Life Science Cluster

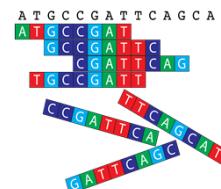
Critical Analysis Course

This course is designed to help trainees to develop oral and written critical thinking skills as they prepare for their written and oral candidacy examination. These same skills are the basis for preparation of manuscripts and grants

Course Objectives:

- Develop oral presentation skills through weekly presentations of classic and recent literature
- Develop concise writing skills through written summaries of the week's paper
- Develop the ability to analyze experimental designs used to test hypotheses
- Develop experience in asking and answering thought provoking questions through class discussions
- Dive into the fundamental literature in bioinformatics and genomics both at the computational and biological levels

CBIOS Practicum, Fall 2014



There is a major gap in scientific communication regarding genomic sequencing data and their application to the general public

The CBIOS practicum aims to develop a web based activity called *Assemblo* with three primary goals:

- Increase literacy on genome sequencing
- Demystify the data processing
- Demonstrate applications to human and animal health

Professional Development Activities

Trainees are required to participate in

- Annual Bioinformatics and Genomics Retreat
- Weekly Seminars
- Genomics Ethics Discussions

Trainees are required to perform one of the following activities:

- Organize retreat/instruct workshop
- Teaching Assistantship in a CBIOS course
- Internship
- Advanced training in grant writing and submission

Genomics Ethics



A discussion on issues related to release, withdrawal and re-release of HeLa genomic sequence data, Spring 2014

Summer Research Training

- A partnership between Howard University and CBIOS Training Faculty
- Provide a cutting-edge, research-intensive program for undergraduate students
- Continued mentorship over junior and senior years
- Prepare a diverse body of highly skilled students for graduate studies in cross-disciplinary areas



A Howard University student working in a Penn State lab, Summer 2014

To Apply:

- Be a first or second year graduate student in a life science or a quantitative science graduate program
- Have an aptitude for inter-disciplinary research involving computation, bioinformatics, statistics and genomics
- Contact a CBIOS training faculty and develop a statement of purpose
- The call for applications will be announced in early fall with a submission deadline in mid-October
- Selected trainees will be supported for two years beginning the following spring semester

For more information, visit: <http://www.huck.psu.edu/education/grants/computation-bioinformatics-statistics>

Home » Research

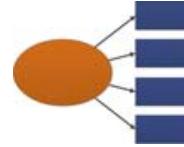
Research Areas

We conduct methodological research focusing on experimental design and data analysis for the prevention and treatment of problem behaviors, particularly drug abuse and HIV. We apply the new methods we develop in empirical research and disseminate new methods to prevention and treatment scientists.

Uncover population profiles

Latent class analysis (LCA) and latent transition analysis (LTA) allow researchers to detect subgroups within a population.

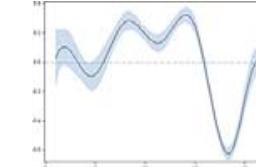
[Read more about LCA](#)



Understand complex longitudinal data

Our intensive longitudinal data (ILD) project provides tools like the time-varying effect model (TVEM) for analyzing complex data, like smartphone data.

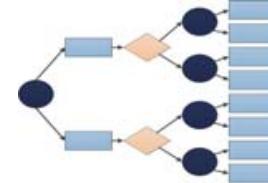
[Read more about ILD](#)



Develop adaptive interventions

The sequential, multiple assignment, randomized trial (SMART) project provides tools for building adaptive interventions so that treatments adapt to meet patients' needs.

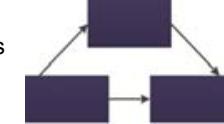
[Read more about SMART](#)



Infer causality

Our causal inference project develops methods for inferring causality in situations where randomization is not possible, practical, or ethical.

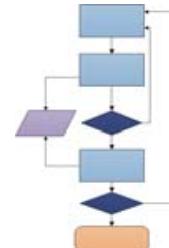
[Read more about causal inference](#)



Optimize interventions

The multiphase optimization strategy (MOST) is a framework for engineering interventions in order to get the maximum impact from an intervention.

[Read more about MOST](#)



Identify important variables in large data sets

Our high-dimensional data project provides tools for variable screening and selection in datasets with large numbers of variables, like genetics data.

[Read more about high-dimensional data](#)





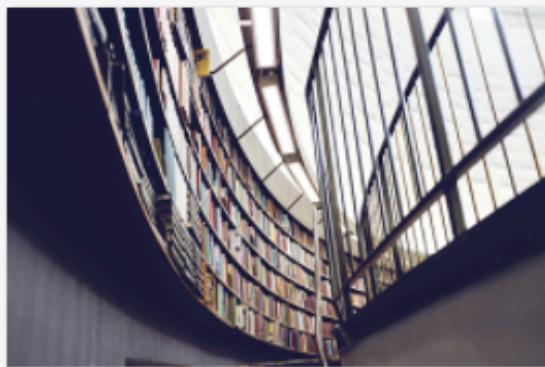
Search for...

CAUSEweb ▾

GO!

Consortium for the Advancement of Undergraduate Statistics Education

A national organization whose mission is to support the advancement of undergraduate statistics education.



Resources

Collect, review, develop, and disseminate resources for members of the undergraduate statistics education community.



Professional Development

Coordinate, develop, and disseminate opportunities, programs, and workshops for teachers.



Research

Promoting statistics education research as a recognized discipline with a visible presence.

News



Webinars

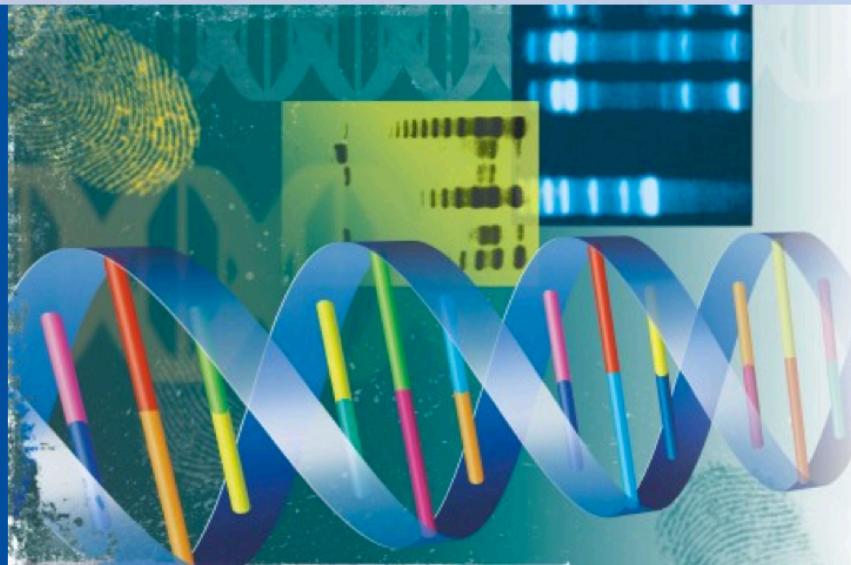
- Activity Series

Presenting individual activities with complete DIY guidance.

- Journal of Statistics Education Series

Recent Topics in Research Methods

Join the CTSI for a bi-weekly research methods seminar series throughout the Spring 2015 semester. All seminars will be videoconferenced between University Park and Hershey.

[More »](#)

Penn State CTSI

The most important challenge facing academic medical centers is translating new research findings into tangible health benefits that improve lives. Penn State has created the **Clinical and Translational Science Institute (CTSI)** to serve as an independent but highly integrated academic engine for clinical and translational science.

The CTSI builds on the strong tradition and rich experience of collaborative interdisciplinary thinking that exists at Penn State and brings this wide breadth of academic endeavor—including traditional biomedical activities, social sciences, arts, communications, economics, education, engineering, ethics, and health care delivery and policy—to bear on health.

The long-range goal of the CTSI is to develop, implement and make available to the community at large new methods to predict, prevent and effectively treat human disease.

Upcoming Events

[Back](#)[CTSI Directors' Council](#)[Next](#)

Recent CTSI News

[Latest NSF rankings: Penn State scores high in](#)

RESEARCH TOOLS:

**REDCap**

Research Electronic
Data Capture

**Penn State Profiles**

Search faculty and their
research expertise

**i2b2**

Informatics for Integrating
Biology and the Bedside

**BERD**

Biostatistics consultation
form

**eagle-i**

Database of research
resources

ACCESS SERVICES:



Penn State Hershey
Research Concierge



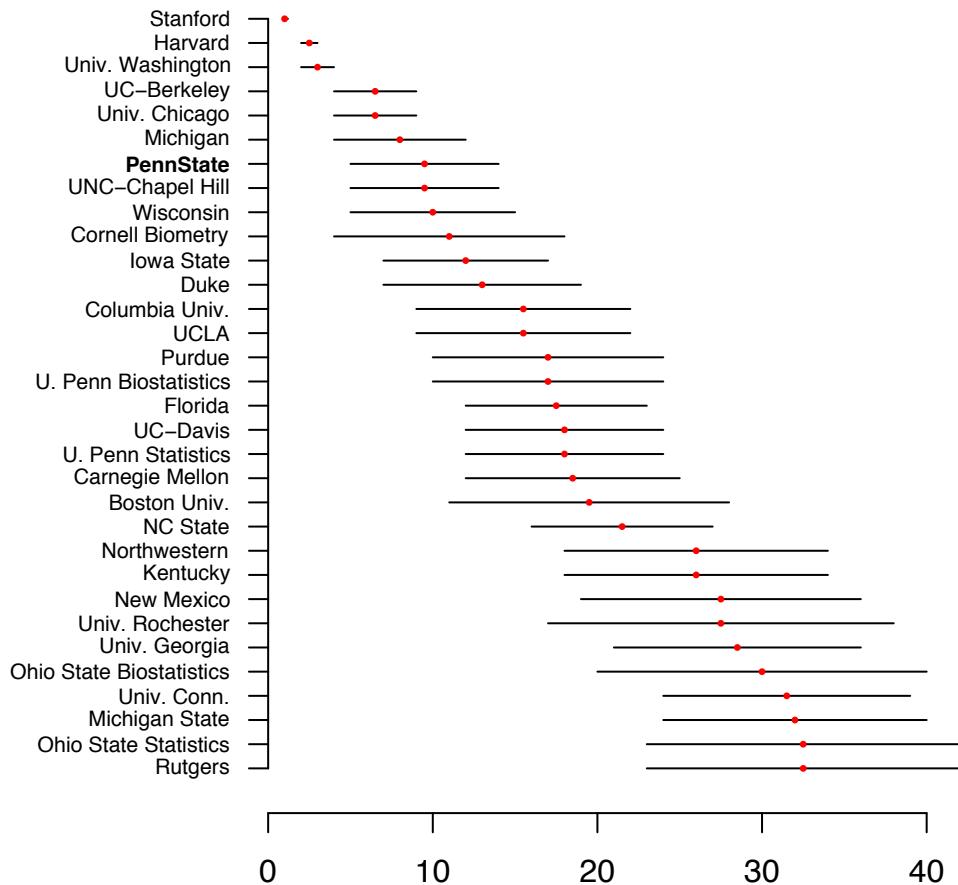
Research Penn State
University Park
Research Support

**CTSI Request Form**

FUNDING:

CTSI Pilot Project Funding
Opportunity

NRC S Rankings for Statistics



Estimated Ranks (5th and 95th Percentiles)