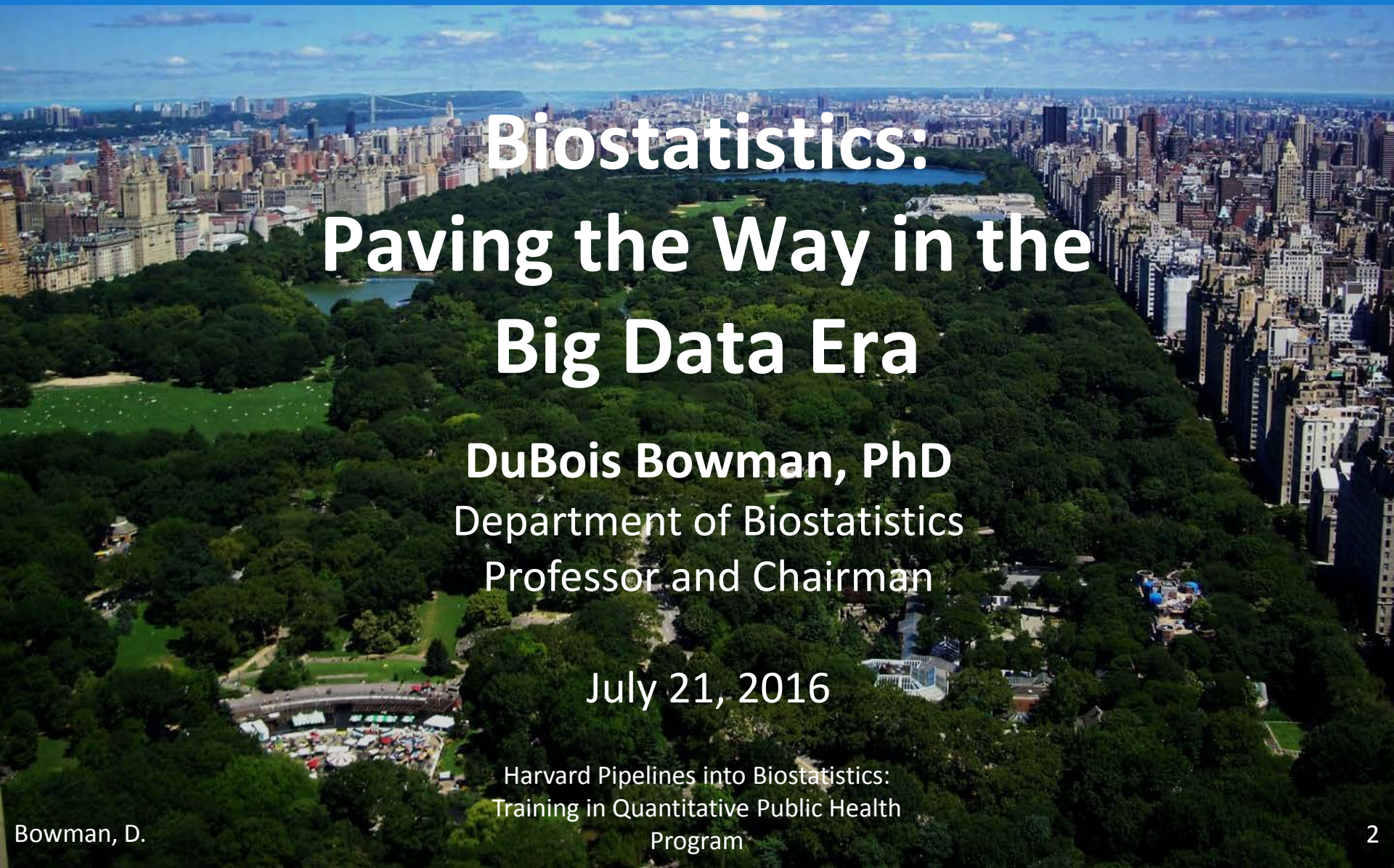
Harvard Pipelines into Biostatistics: Training in Quantitative Public Health Program

Bowman, D.



Biostatistics: Paving the Way in the Big Data Era

DuBois Bowman, PhD
Department of Biostatistics
Professor and Chairman

July 21, 2016

Harvard Pipelines into Biostatistics:
Training in Quantitative Public Health
Program

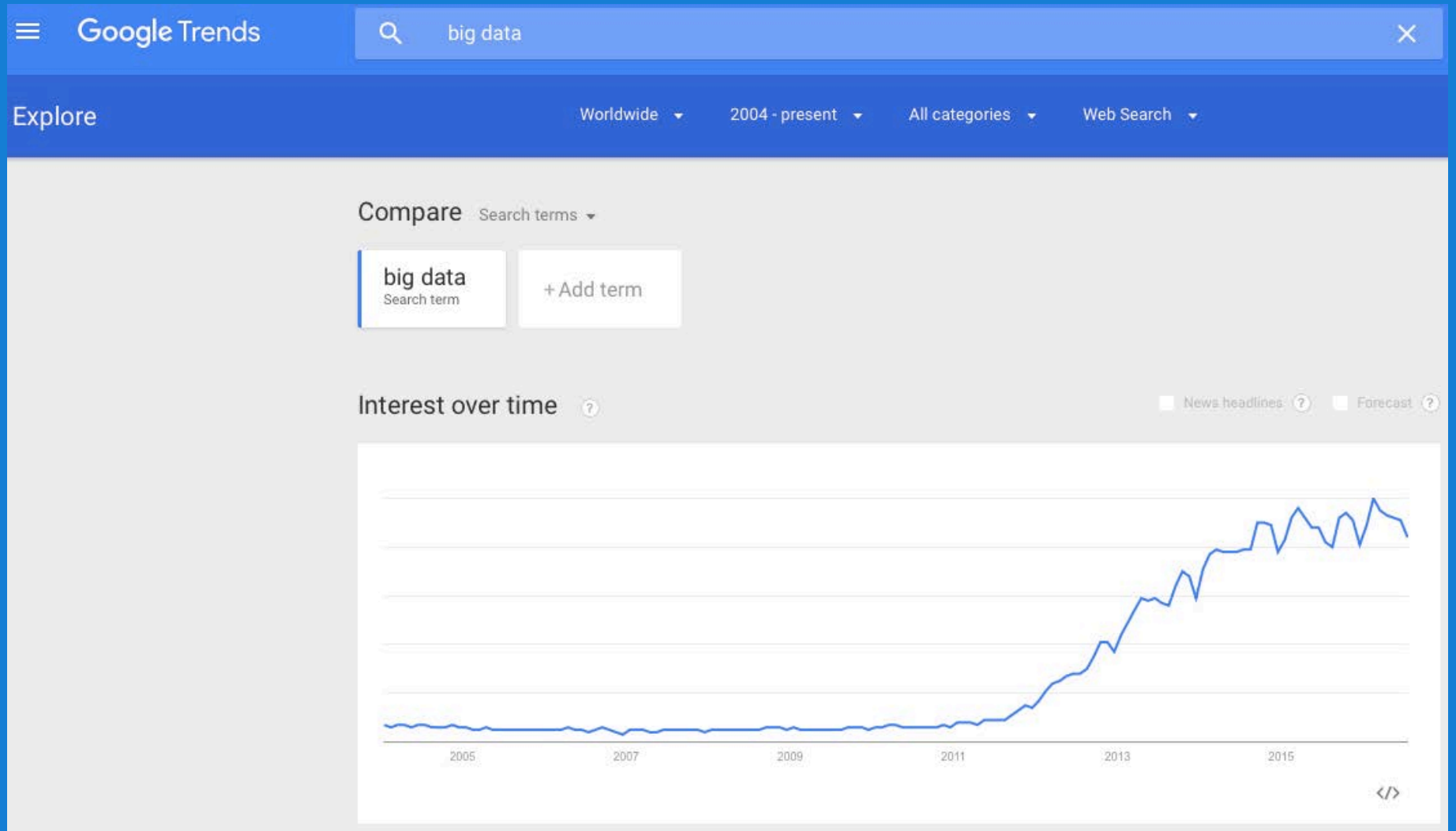
Data

Data Are …

- Ubiquitous
- Increasing in complexity
- Rapidly expanding in scale
  - Quickly and inexpensively generated by emerging technologies
  - Increased storage capacity
  - Increased number of experimental units
  - Increased temporal scales
  - Increased spatial scales

Big Data

Big Data


Google Trends
big data
Explore
Worldwide
2004 - present
All categories
Web Search
Compare
Search terms
big data
Search term
+ Add term
Interest over time
News headlines
Forecast
2005
2007
2009
2011
2013
2015

Big Data

A Revolution

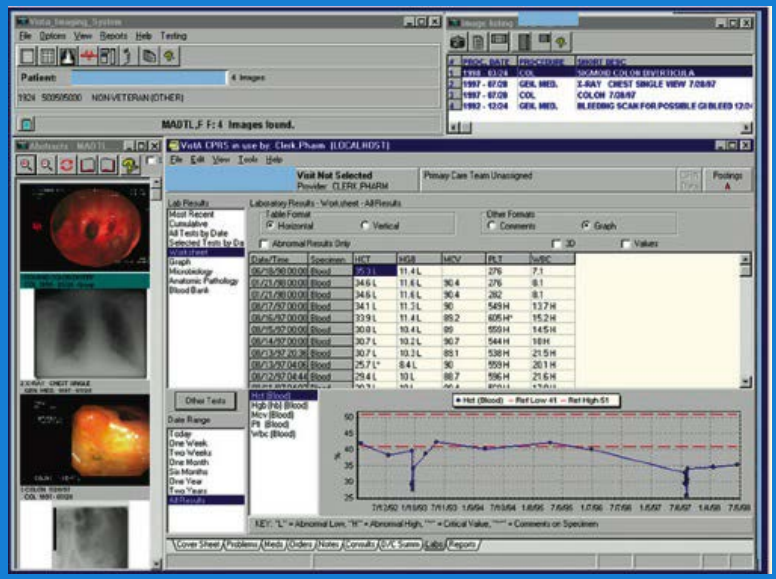
- A total of 2.5 *quintillion* bytes of data were generated every day **in 2012 alone**
- Decoding...
  - 2.5 trillion megabytes
  - If these units (megabytes) were time...say seconds…

    It would take 4125 years to elapse

Big Data

Electronic medical/health records

- What type of care would improve health outcomes?
- How to minimize risk for readmission?
- Identify high risk patients
- from data.



Big Data

Personal smart devices

- Physical activity
- Medication use / compliance
- Monitor symptoms of medical conditions

Big Data

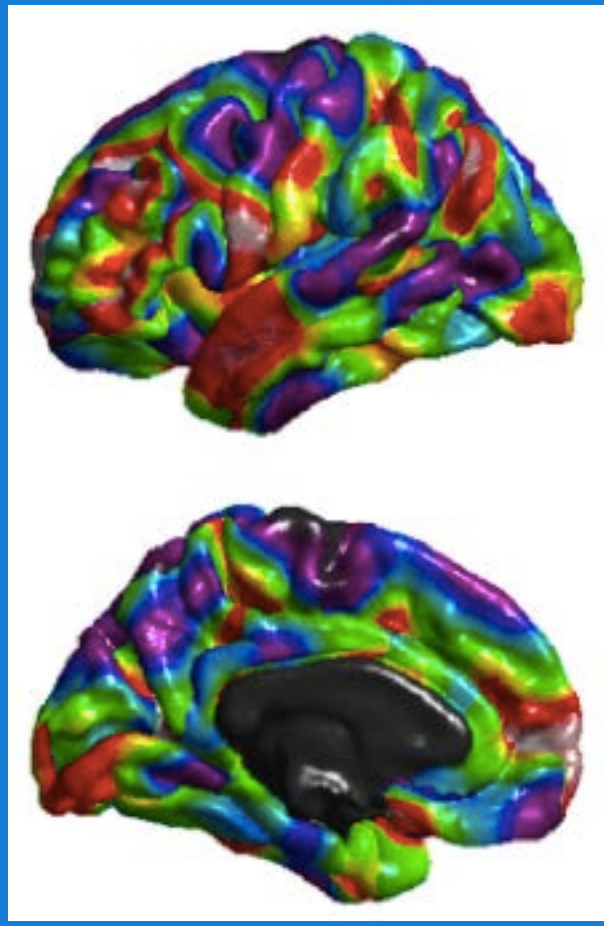
Genetic/Genomic Analyses

- Genetic basis of human diseases and traits

Big Data

Neuroimaging

- Localize brain regions altered in neurological or psychiatric diseases.
- Identify precursors of disease prior to clinical symptoms.
- Determine which patients are likely to respond to particular treatments.



Need More than Just Data

"Smart" devices are creating a flood of data. Now hospitals need to figure out how to manage it all.

Hospitals: Big data use is a 'significant challenge'

August 13, 2013 | By Dan Bowman

7:11 am ET
Mar 26, 2014 LEADERSHIP

Six Challenges of Big Data

Second, companies are struggling to find the right talent capable of both working with new technologies and of interpreting the data to find meaningful business insights.

Need More than Just Data

The 3 big problems in big data (hint: They all involve people)

2. Data Scientists

A typical enterprise generally has 10x more IT employees than analysts or data scientists. The process of analysis starts with a line of business request. IT collects data from various databases and transfers it to data scientists. Large teams of data scientists are deployed who spend months (or sometimes years) querying the data. Hiring data scientists (with advanced background in statistics, computer science, and some functional expertise) to accelerate the process is difficult because people with these skills are extremely scarce. The demand and interest in data scientists is skyrocketing, as Google Trends can attest, while we are producing fewer of them. What we need is a new class of technologies that amplify the impact data scientists and allow more people to become data scientists.

Need: Biostatistics / Statistics

Biostatistics

- Recent focus on data science, analytics, etc.
- Statistical thinking is fundamental to all of these areas
  - Rests upon well-established principles of estimation, inference, and prediction
  - Head start:
    - Long history…dating back to 1749
    - American Statistical Association was founded in 1839 (Over 175 years!)
    - Many seminal results established in early 1900's

What is Biostatistics?

What is Biostatistics?

statistics

noun plural but singular or plural in construction | sta·tis·tics | \stə-'tis-tiks\

Definition of STATISTICS

Popularity: Top 10% of words

1 : a branch of mathematics dealing with the collection, analysis, interpretation, and presentation of masses of numerical data

What is Biostatistics?

statistics *(Concise Encyclopedia)*

Branch of mathematics dealing with gathering, analyzing, and making inferences from data. Originally associated with government data (e.g., census data), the subject now has applications in all the sciences. Statistical tools not only summarize past data through such indicators as the mean (*see* MEAN, MEDIAN, AND MODE) and the standard deviation but can predict future events using FREQUENCY DISTRIBUTION functions. Statistics provides ways to design efficient experiments that eliminate time-consuming trial and error.

What is Biostatistics?

biostatistics

noun plural but singular in construction | bio·sta·tis·tics | \ˌbī-ō-stə-ˈtis-tiks\

Definition of BIOSTATISTICS

Popularity: Bottom 30% of words

: statistics applied to the analysis of biological data

Biostatistics

Helps to gain insights into health and medicine.

- **Quantify uncertainty**: how likely or unlikely something is to occur (or via variability).

- **Detect differences**, e.g. between patients treated with a new cholesterol lowering medication and those receiving standard therapy.

Biostatistics

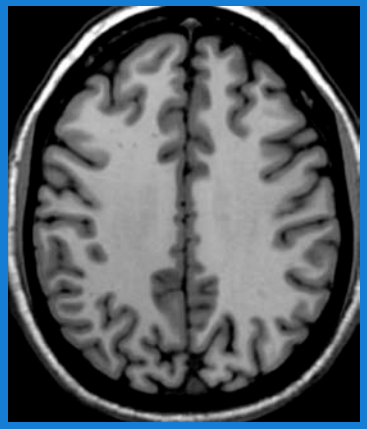
- **Estimate unknown quantities:**
  - **WHO** is at high risk for developing Alzheimer's disease?
  - **WHAT** is the safest and most effective dose for a targeted cancer therapy?
  - **WHERE** (in which neighborhoods) are there disproportionate levels of environmental pollutants causing respiratory diseases and other illnesses?
  - **WHEN** is the optimal time to intervene with a vaccination to protect against widespread outbreak of a potentially fatal infectious disease?
  - **WHY** are the rates for autism increasing?
- **Predict**, e.g. health outcomes (progression of diseases, relapse, cure).

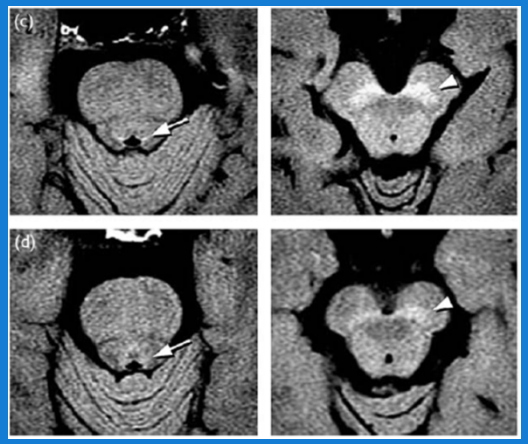
Big Data

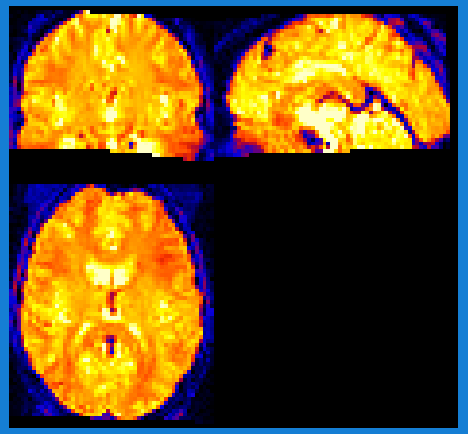
Magnetic Resonance Imaging (MRI)

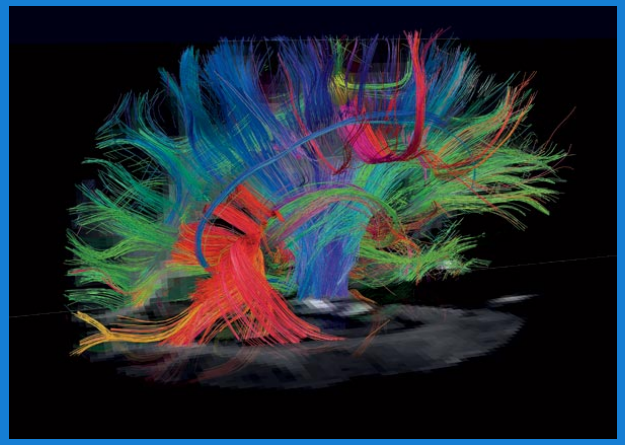
Big Data

Magnetic Resonance Imaging (MRI)



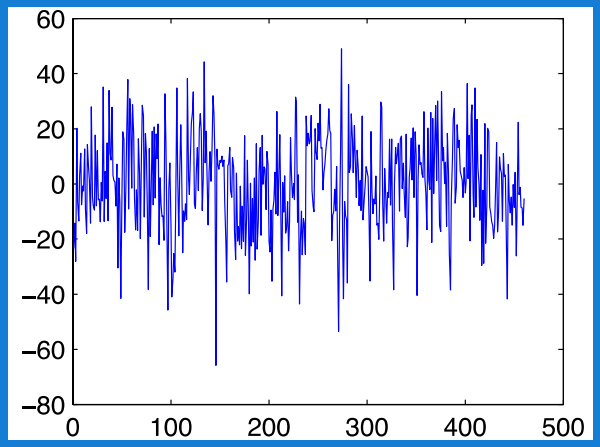






$$\oint E \cdot ds = -\frac{d}{dt}\int B \cdot \cdot dA$$

$$\oint B \cdot dA = 0$$



Big Data

Brain Imaging

- Depression
- Schizophrenia
- Bipolar Disorder
- Social Anxiety
- Cocaine dependence
- Parkinson's Disease
- Alzheimer's Disease
- Stroke Recovery
- Autism
- Moral Reasoning

Cardiac Imaging

- Recovery following a heart attack

Cancer imaging

- Breast cancer
  - Early diagnosis
  - Monitoring treatment
- Prostate cancer

Biostatistics: Neuroimaging

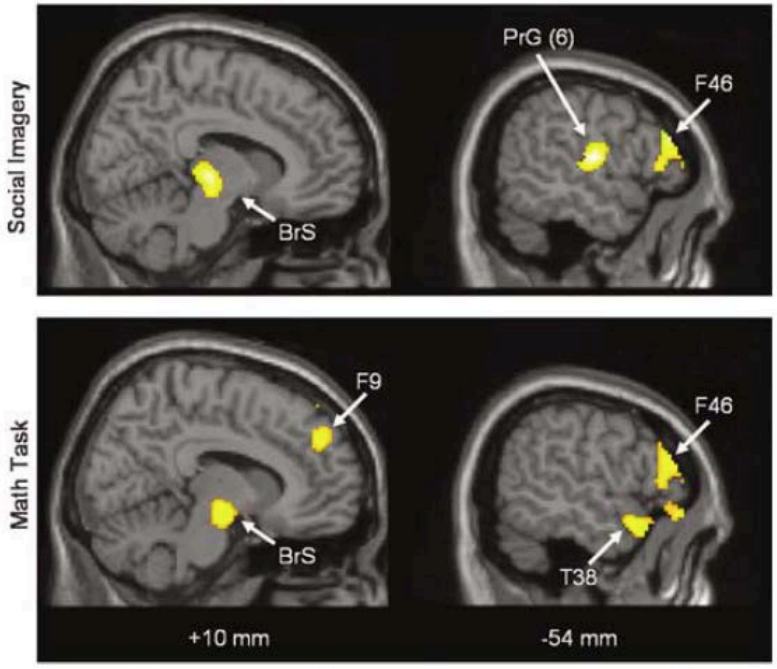


Figure 3 Greater ($p<0.005$, uncorrected) provoked social anxiety-related rCBF before compared to after treatment of individuals with gSAD with nefazodone. Common effects are illustrated in parasagittal views for the social anxiety imagery (top) and confrontational mental arithmetic task provocation (bottom) conditions. Abbreviations refer to the right brainstem/midbrain (BrS), left middle (F46), and right medial (F9) frontal gyri, left precentral gyrus (PrG6), and left temporal pole (T38).

Neuropsychopharmacology (2006) 31, 2243–2253

www.neuropsychopharmacology.org

The Neural Correlates of Social Anxiety Disorder and Response to Pharmacotherapy

Clinton D Kilts*[,1,2], Jeffrey E Kelsey[1], Bettina Knight[1], Timothy D Ely[1], F DuBois Bowman[3], Robin E Gross[1], Amy Selvig[1], Angelita Gordon[3], D Jeffrey Newport[1] and Charles B Nemeroff[1]

[1]Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta, GA, USA; [2]Emory Center for Positron Emission Tomography, Emory University School of Medicine, Atlanta, GA, USA and [3]Department of Biostatistics, Emory University School of Public Health, Atlanta, GA, USA

Biostatistics: Neuroimaging

Predicting brain activity using a Bayesian spatial model
Gordana Derado, F DuBois Bowman and Lijun Zhang
Stat Methods Med Res published online 28 June 2012

Predicting future brain activity in patients with Alzheimer's disease

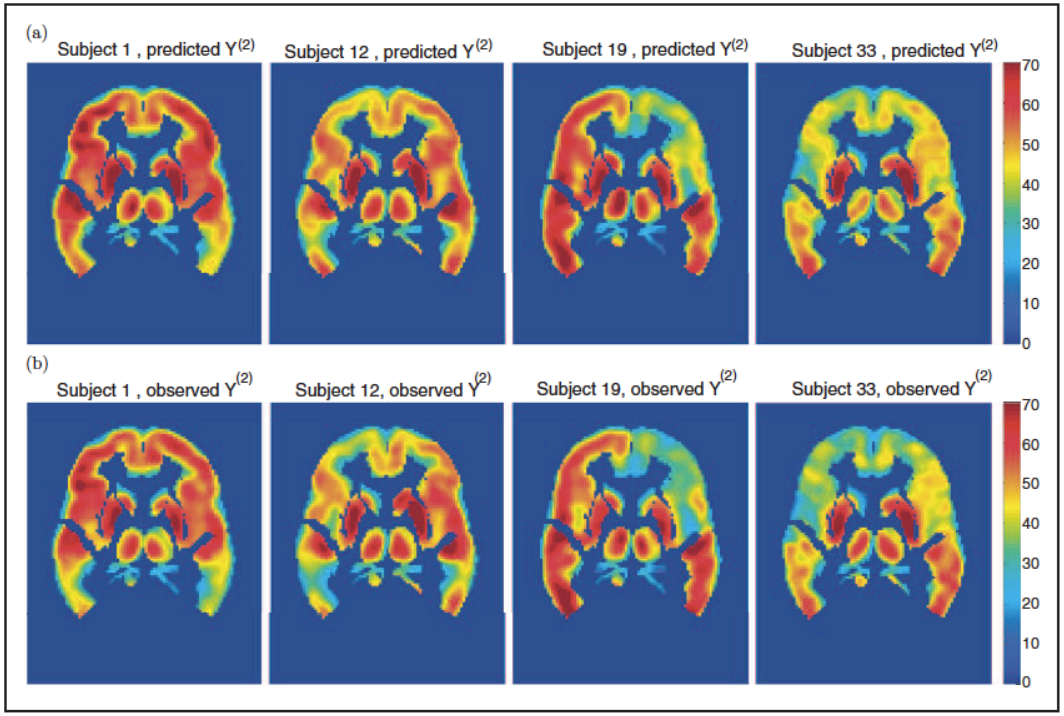
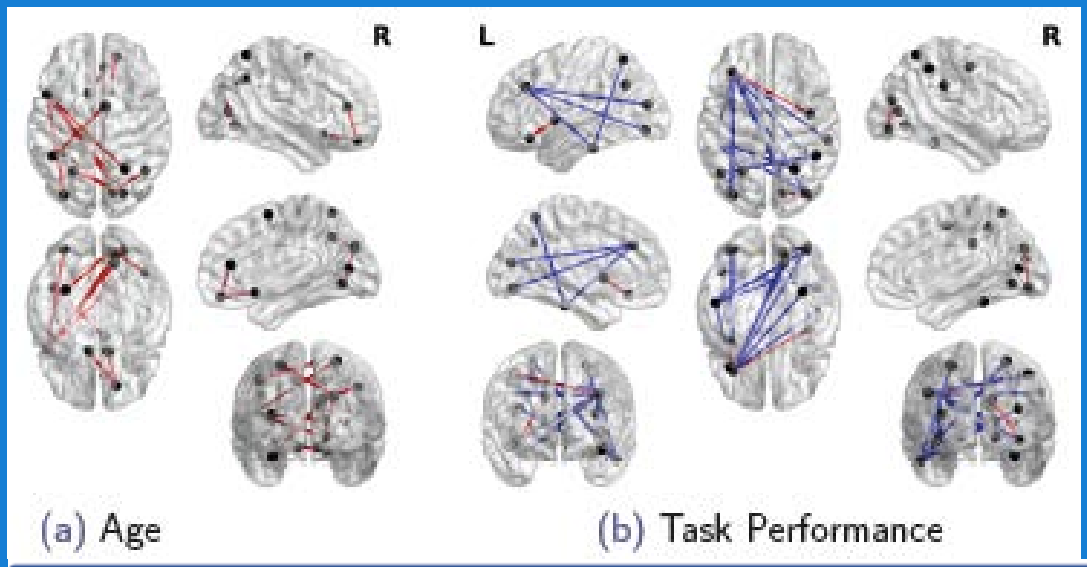


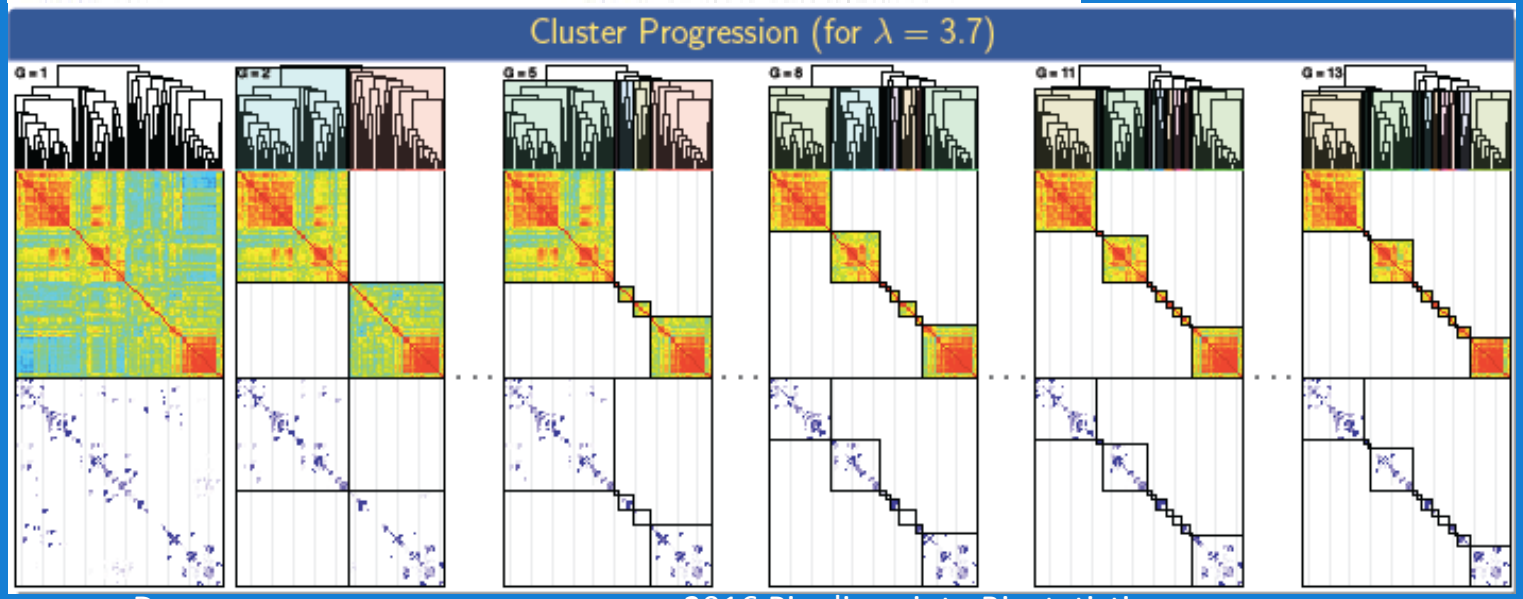
Figure 1. Individualized predicted (a) and observed (b) 6-month follow-up regional glucose uptake measurements for four AD patients from the test dataset.
Axial slice 40 is shown in radiological view. There is a satisfactory agreement between the observed and predicted 6-month regional glucose uptake.

Biostatistics: Neuroimaging

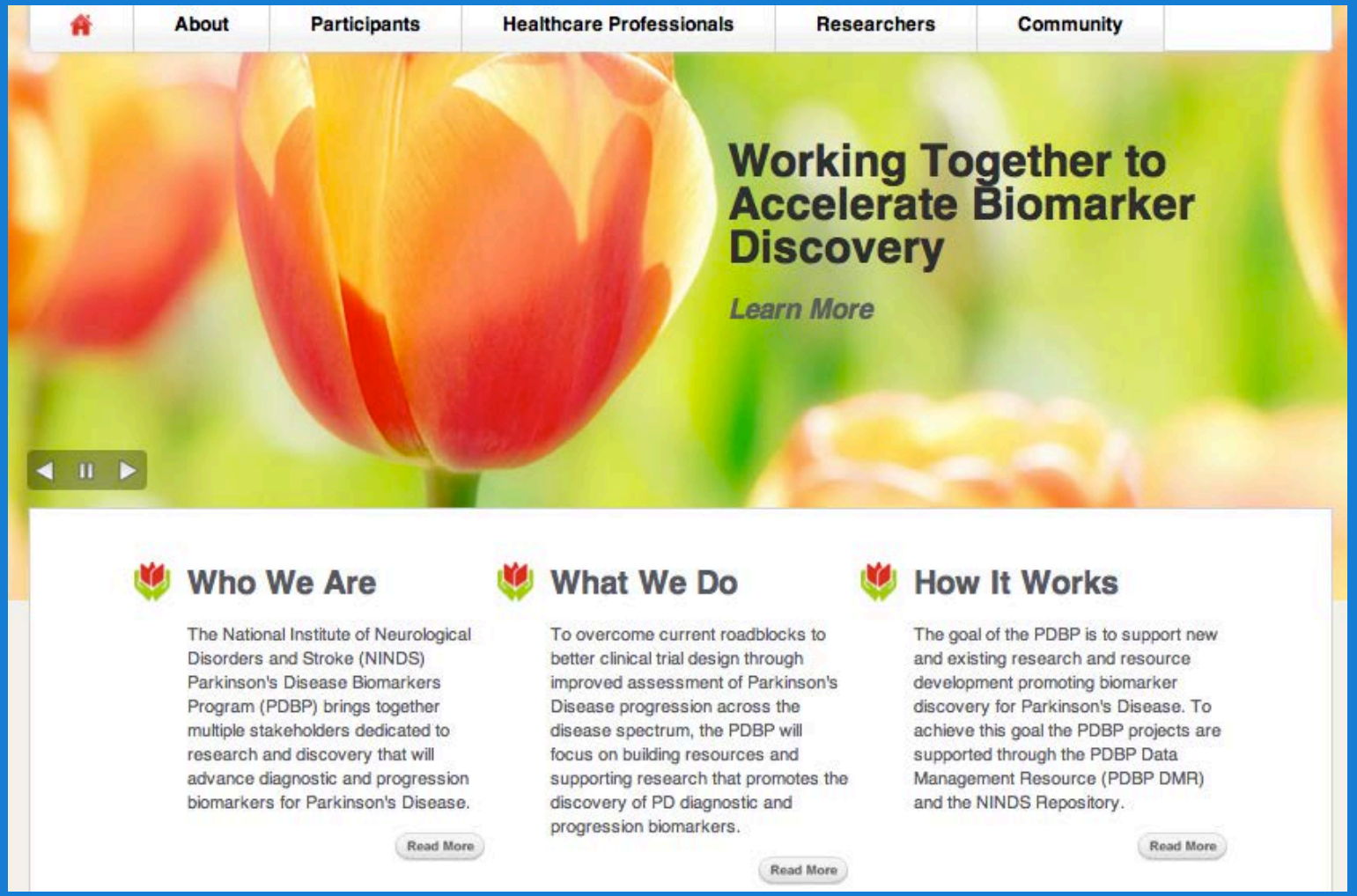


Age and task-related differences in functional brain networks using awFC

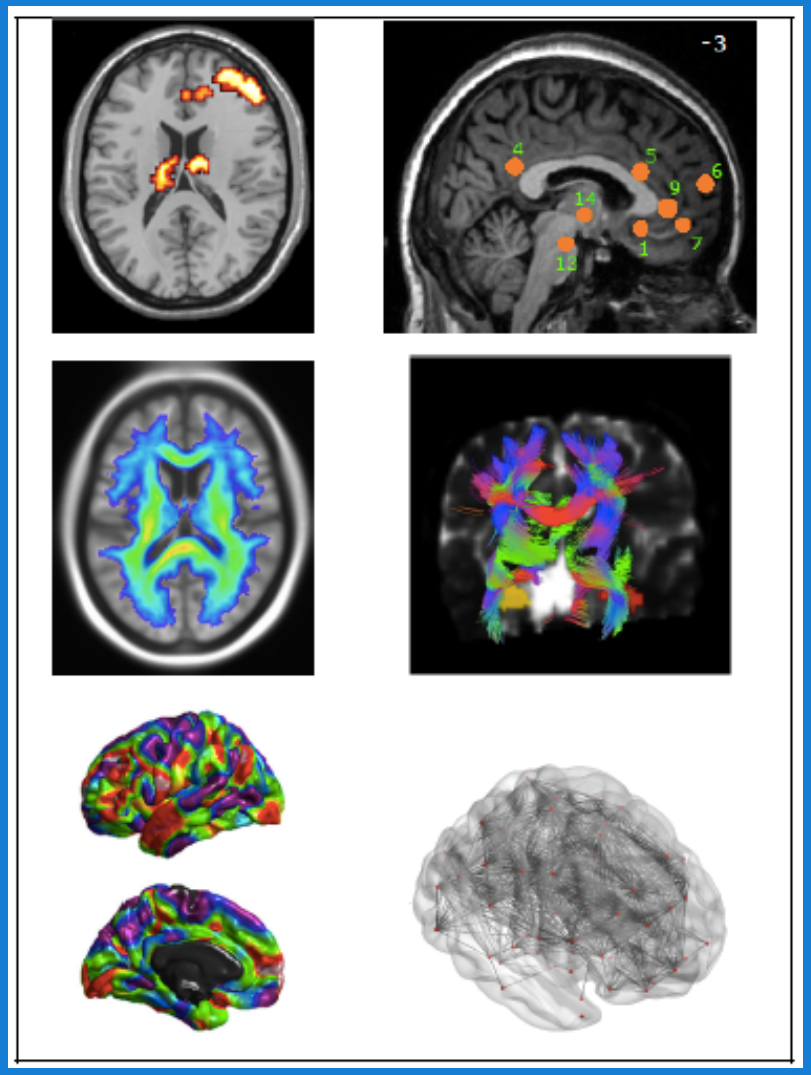
Bowman et al., 2012, NeuroImage



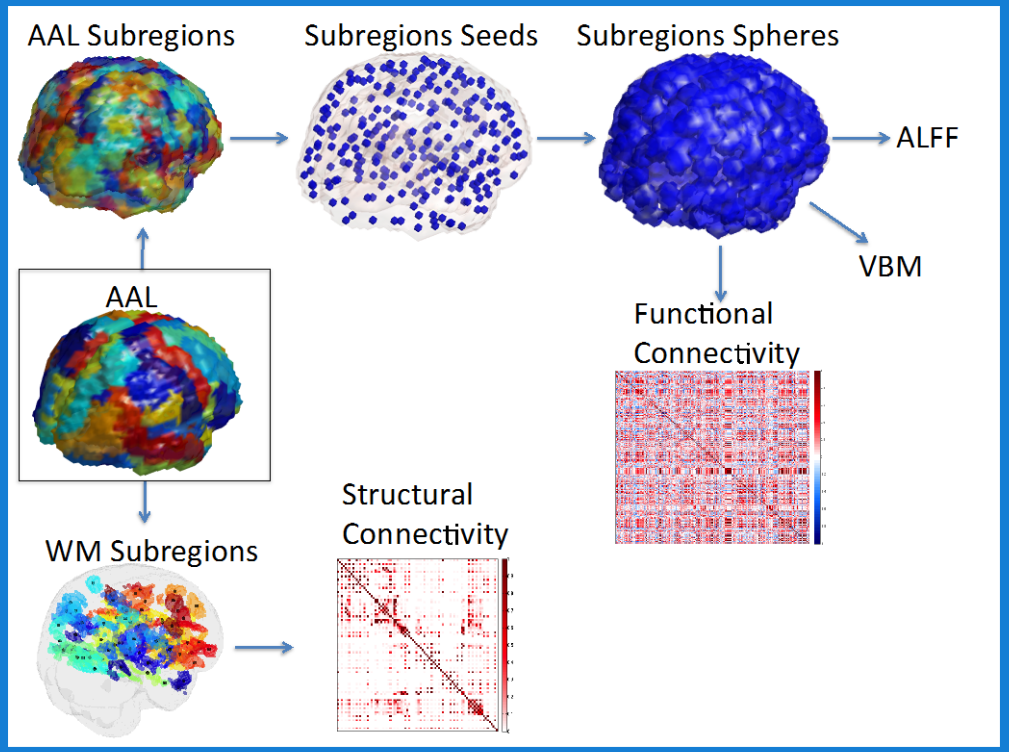
Biostatistics: Neuroimaging



Biostatistics: Neuroimaging

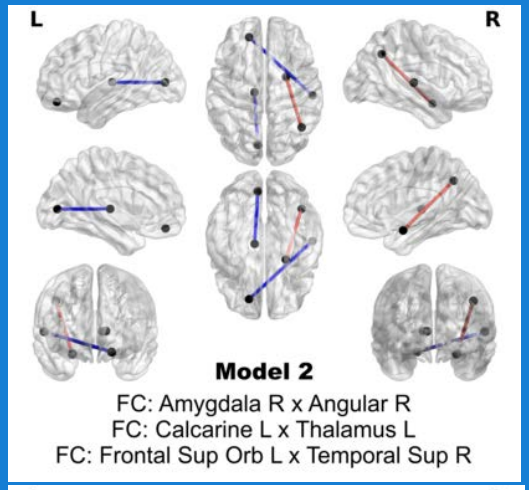


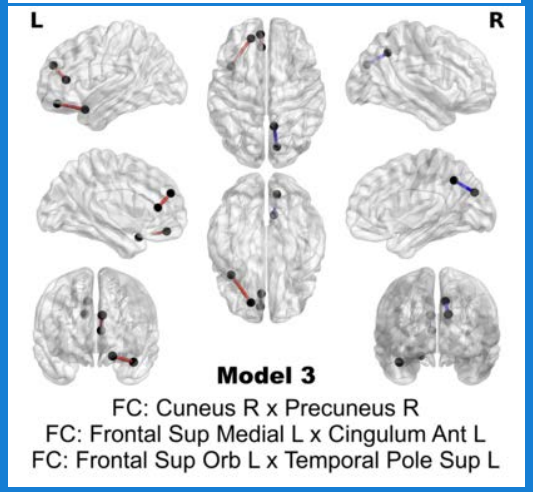
Multimodal neuroimaging data

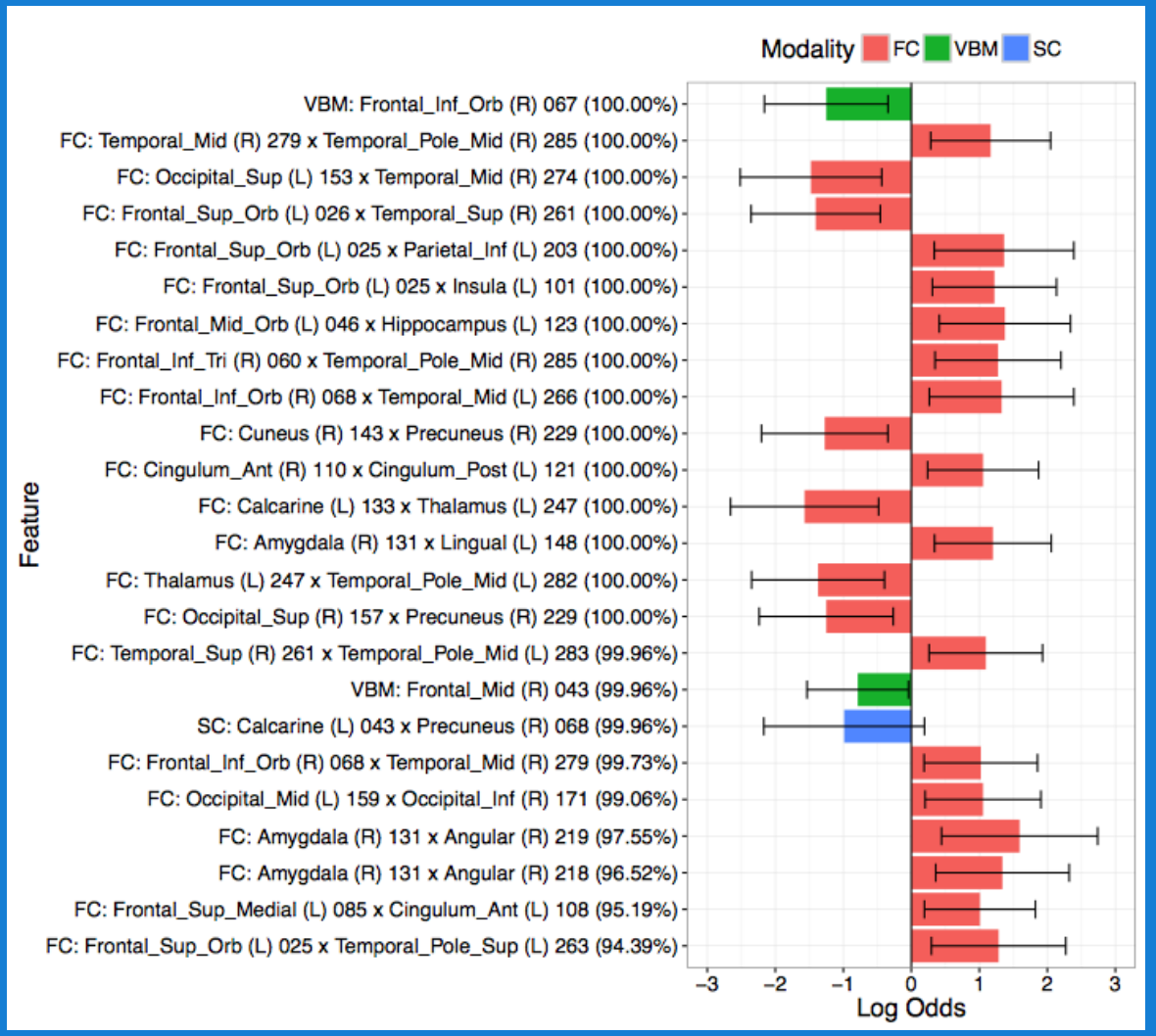


Biostatistics: Neuroimaging

Markers of PD:







Bowman et al., 2016, Front in Neurosc

The Future is Bright

Biostatistics

Best Graduate Degrees for Jobs in 2016:

- #1: Master's in Biostatistics
  - Median Salary: $106K
  - Projected Job Growth by 2024: 23%
  - Percent Highly Satisfied: 85%
  - Low Stress: 57%
- 2. Master's, Statistics
- 3. Ph.D., Computer Science
- 4. Ph.D., Economics
- 5. Master's, Applied Mathematics

Fortune magazine, 2016

Statisticians

Summary | What They Do | Work Environment | How to Become One | Pay | Job Outlook | State & Area Data | Similar Occupations | More Info

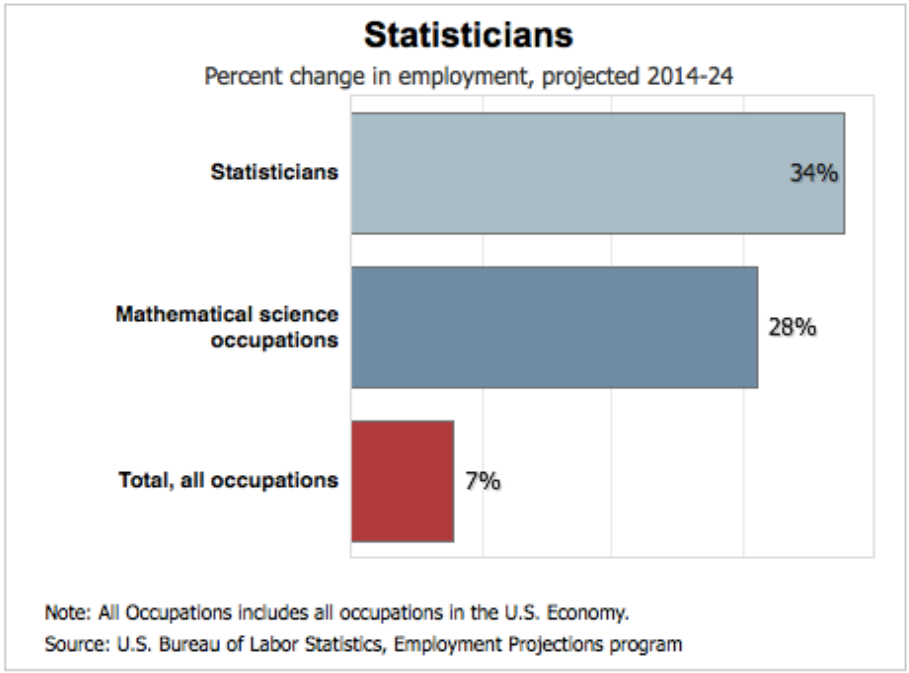
Job Outlook

About this section

Employment of statisticians is projected to grow 34 percent from 2014 to 2024, much faster than the average for all occupations. Growth is expected to result from more widespread use of statistical analysis to make informed business, healthcare, and policy decisions. In addition, the large increase in available data from the Internet will open up new areas for analysis.

A substantial amount of data is generated from Internet searching and the use of social media, smartphones, and other mobile devices. Businesses, particularly those in the retail, finance, and insurance industries, will increasingly need statisticians to organize, analyze, and sort through the data for commercial reasons. Analyses will help companies improve their business processes, design and develop new products, and advertise products to potential customers.

In addition, statisticians will be needed in the pharmaceutical industry. The aging of the U.S. population will encourage pharmaceutical companies to develop new treatments and medical technologies. Biostatisticians will be needed to conduct the research and clinical trials necessary for companies to obtain approval for their products from the Food and Drug Administration.



The occupation will also see growth in research and development in the physical, engineering, and life sciences, fields in which statisticians' skills in designing tests and assessing results are highly useful.

My Journey

My Educational Background

Undergraduate

- Morehouse College, B.S. Mathematics
  - COR (MARC) Program

Graduate

- University of Michigan, M.S. Biostatistics
- UNC at Chapel Hill, Ph.D. Biostatistics

My Faculty Career

Department of Biostatistics

Emory University

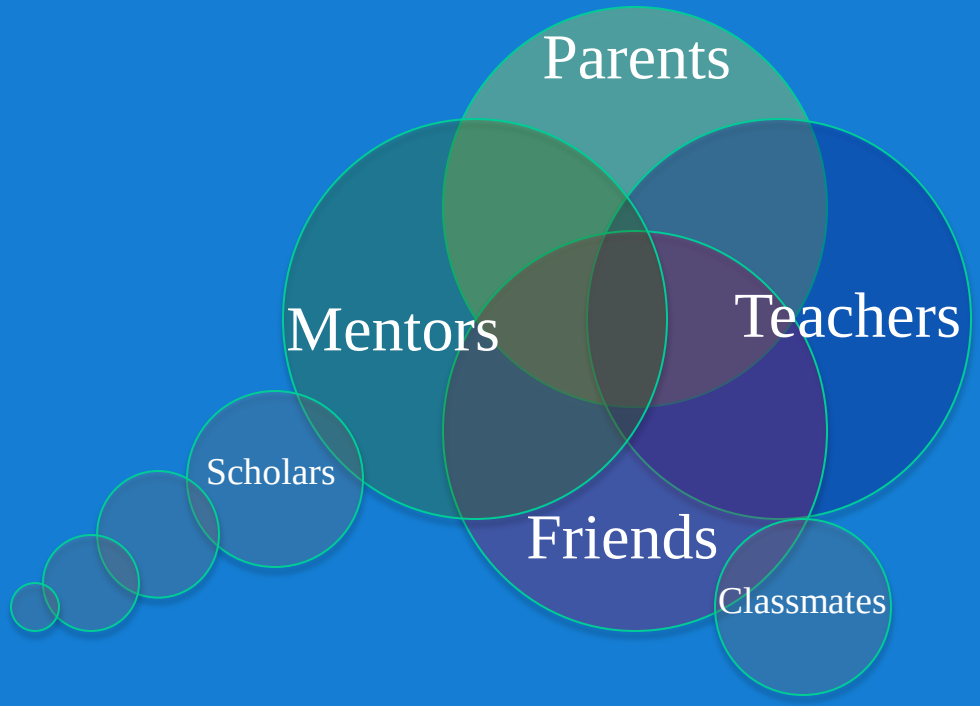
- Assistant Professor
- Associate Professor
- Full Professor

Columbia University

- Full Professor and Chairman

Networking

Professional and Personal Networks



Diversity

Underrepresentation in Academia:

Faculty

Year, sex, and academic rank	Total	White	Black	Hispanic	Asian	Pacific Islander	American Indian/ Alaska Native	Two or more races	Race/ ethnicity unknown	Non-resident alien
2013										
Total	791,391	575,491	43,188	33,217	71,038	1,208	3,538	5,291	20,013	38,407
[Pct of Faculty]		72.7%	5.5%	4.2%	9.0%	0.2%	0.4%	0.7%	2.5%	4.9%
Professors	181,530	148,577	6,665	5,604	15,247	170	573	852	2,323	1,519
[Pct of Profs]		81.8%	3.7%	3.1%	8.4%	0.1%	0.3%	0.5%	1.3%	0.8%
[Pct of Faculty]		18.8%	0.8%	0.7%	1.9%	0.0%	0.1%	0.1%	0.3%	0.2%
Associate Professors	155,095	116,817	8,812	6,381	15,626	183	591	987	2,859	2,839
[Pct of Assoc. Profs]		75.3%	5.7%	4.1%	10.1%	0.1%	0.4%	0.6%	1.8%	1.8%
[Pct of Faculty]		14.8%	1.1%	0.8%	2.0%	0.0%	0.1%	0.1%	0.4%	0.4%
Assistant Professors	166,045	112,262	10,542	7,130	18,070	332	683	1,254	5,695	10,077
[Pct of Assist Profs]		67.6%	6.3%	4.3%	10.9%	0.2%	0.4%	0.8%	3.4%	6.1%
[Pct of Faculty]		14.2%	1.3%	0.9%	2.3%	0.0%	0.1%	0.2%	0.7%	1.3%
Instructors	99,304	73,859	7,448	6,340	4,950	286	879	781	3,180	1,581
Lecturers	36,728	27,453	1,728	2,015	2,403	33	117	295	1,151	1,533
Other faculty	152,689	96,523	7,993	5,747	14,742	204	695	1,122	4,805	20,858

Underrepresentation in Academia:

Faculty (2013)

- 791,391 full-time faculty at degree-granting institutions
  - 43,188 (5.5%) were Black [Includes faculty at HBCUs]
  - 33,217 (4.2%) were Hispanic
- Full professor: 6,665 (3.7%) Black / 5,604 (3.1%) Hispanic (of 181,530)
- Associate professor: 8,812 (5.7%) Black / 6,381 (4.1%) Hispanic (of 155,095)
- Assistant professor: 10,542 (6.3%) Black / 7,130 (4.3%) Hispanic (of 166,045)

The Pool (2012)

- 224,000 African Americans held doctorates.
- 217,000 AAs held a professional degree
- 1,364,000 AAs with a master's degree (as the highest degree).
- 3.5 million AAs with a bachelor's degree (but no graduate degree)
- 10.1% of all degrees granted at 4 year institutions

[Faculty, Racial Gap, Research & Studies on July 12, 2012]

Underrepresentation in Academia:

First-Year Enrollments of African Americans at High-Ranked Universities, Fall 2012

Institution	All Applicants	Total Students Accepted	Overall Student Acceptance Rate	Black Applicants	Blacks Accepted	Black Acceptance Rate	Black Enrollees	Black Student Yield+	Black % of First-Year Class
Columbia University	31,851	2,362	7.4	3,881	**	**	201	**	14.2
Duke University	31,566	4,087	12.9	3,215	**	**	198	**	11.5
Univ. of Pennsylvania	31,218	3,935	12.6	2,766	**	**	272	**	11.1
Univ. of N.C Chapel Hill	29,947	7,847	26.2	3,254	799	24.6	382	47.8	9.8
Vanderbilt University	28,348	4,034	14.2	2,507	411	16.4	155	37.7	9.6
Brown University	28,742	2,759	9.6	2,348	331	14.1	146	44.1	9.5
Harvard University	34,302	2,032	5.9	**	**	**	150	**	9.4
Yale University	28,977	2,043	7.1	2,035	**	**	122	**	9.0
Emory University	17,500	4,599	26.3	2,337	460	19.7	120	26.1	8.6
Mass. Inst. of Technology	18,109	1,620	8.9	**	**	**	91	**	8.0
Stanford University	36,632	2,423	6.6	**	233	**	142	60.9	8.0
Cornell University	37,808	6,259	16.6	3,201	597	18.7	257	43.0	7.9
Dartmouth College	23,110	2,262	9.8	**	**	**	87	**	7.9
Northwestern University	32,069	4,924	15.4	2,043	465	22.8	161	34.6	7.9
Johns Hopkins University	20,502	3,626	17.7	1,711	381	22.3	102	26.8	7.7
Princeton University	26,664	2,094	7.9	1,546	**	**	102	**	7.5
University of Chicago	25,307	3,363	13.3	2,309	**	**	113	**	7.4
Wake Forest University	11,408	3,875	34.0	1,612	299	18.5	87	29.1	7.0
Rice University	15,133	2,528	16.7	948	200	21.1	64	32.0	6.8
Carnegie Mellon Univ.	17,313	4,748	27.4	812	326	40.1	96	29.4	6.7
Washington University	27,265	4,876	17.9	2,149	364	16.9	104	28.6	6.3
University of Virginia	28,251	8,031	28.4	1,708	508	29.7	188	37.0	5.5
University of Notre Dame	16,957	3,947	23.3	830	267	32.2	110	41.2	5.4
Univ. of Southern Calif.	54,387	11,957	22.0	4,470	696	15.6	248	35.6	5.3
Tufts University	16,378	3,504	21.4	948	233	24.6	67	28.8	5.1
University of Michigan	42,544	15,551	36.6	**	**	**	323	**	5.2
Georgetown University	20,116	3,413	17.0	1,931	335	17.3	126	37.6	4.9
Univ. of Calif. Los Angeles	52,231	9,263	17.7	3,036	348	11.5	181	52.0	4.4
Univ. of Calif. Berkeley	41,159	9,348	22.7	2,394	324	13.5	152	46.9	3.7

Universities are listed by highest percentage of black freshmen.
**Declined to provide information to the JBHE research department.

Data for UCLA and Berkeley are for California residents only.
+Yield = percentage of accepted students who enroll.

Source: Survey conducted by JBHE RESEARCH DEPARTMENT.

Opportunities

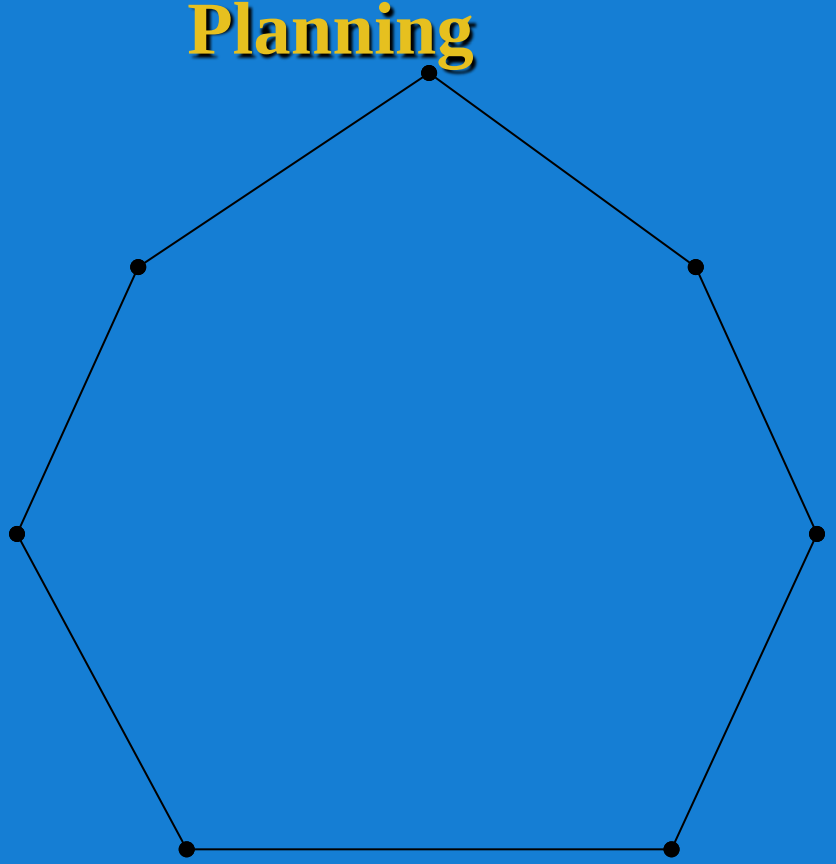
- Clear demand for biostatisticians
- Challenging road, but very rewarding
- Opportunity to make a difference!!

Opportunities

- Programmatic efforts
  - Recruitment
  - Funding
  - Fostering environment for success
- Students: in your hands!

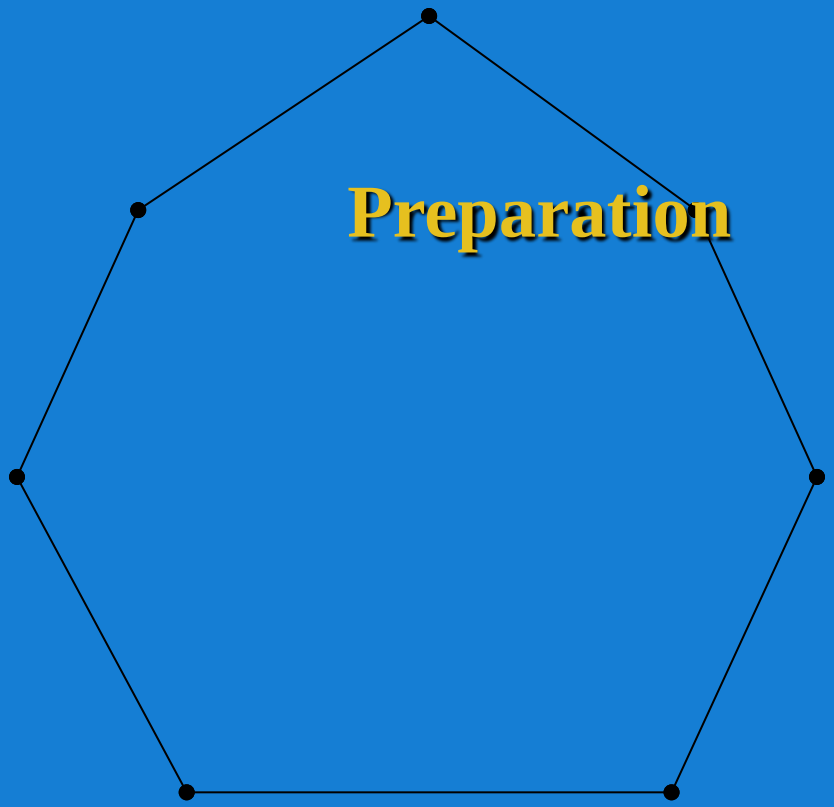
Success in graduate school: Staying the Course

- Strategy
- Navigate the maze of graduate school
  - Course selection
  - Selecting an advisor
  - Dissertation topics
- Work *smarter* AND harder
- "He who fails to plan, plans to fail."



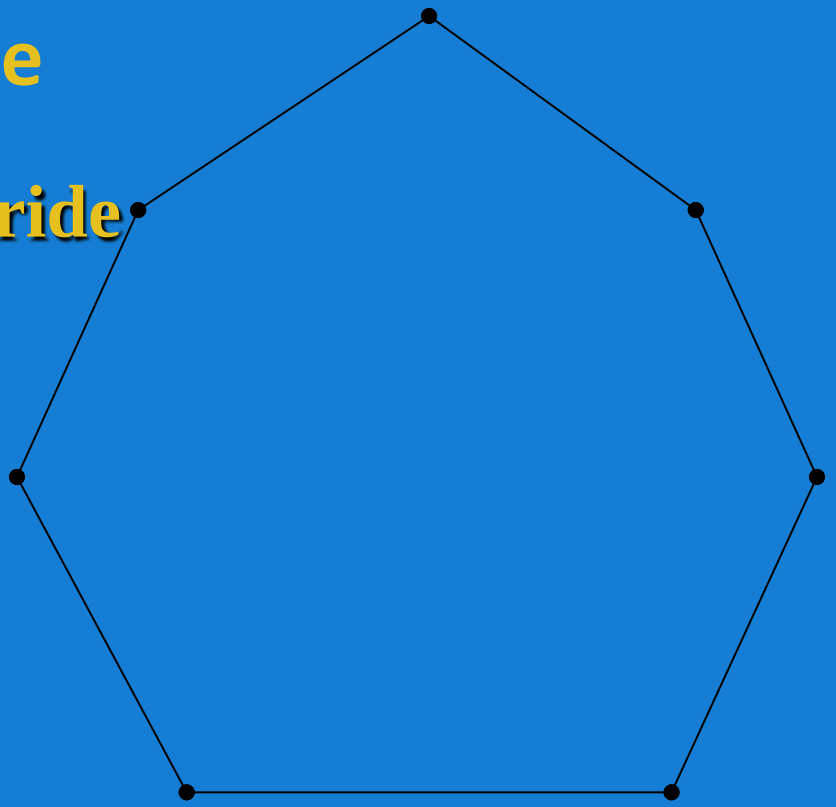
Success in graduate school: Staying the Course

- Today's preparation determines tomorrow's success.
  - Qualifying exams
  - Dissertation research
  - Coursework
  - Meetings, …
- "Luck is what happens when preparation meets opportunity."



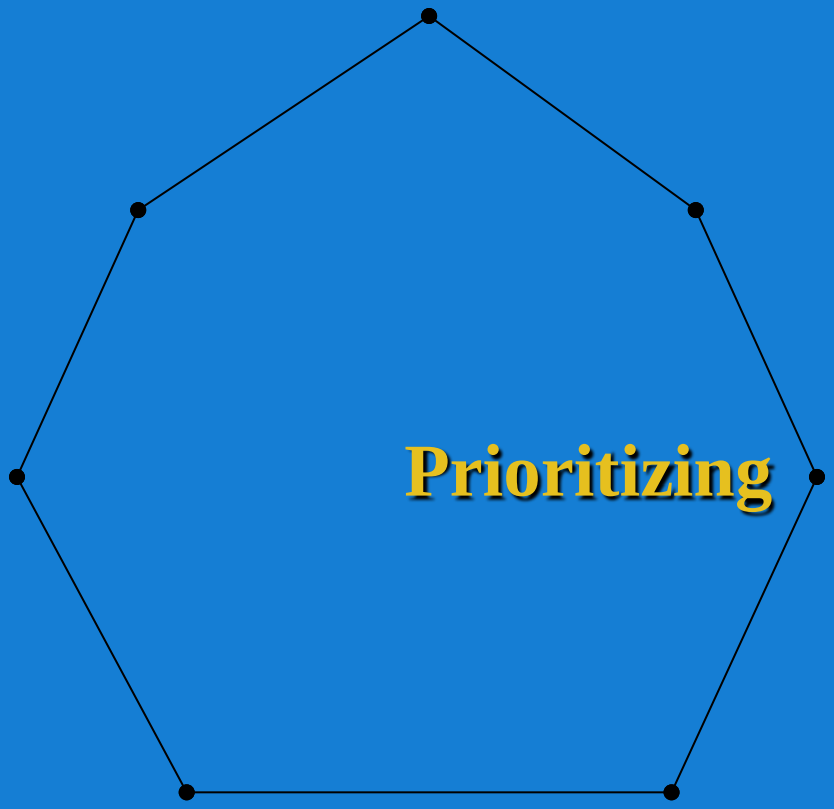
Staying the Course

- **Your ability: Confidence**
- **Your work**
  - **Personal standard of excellence**



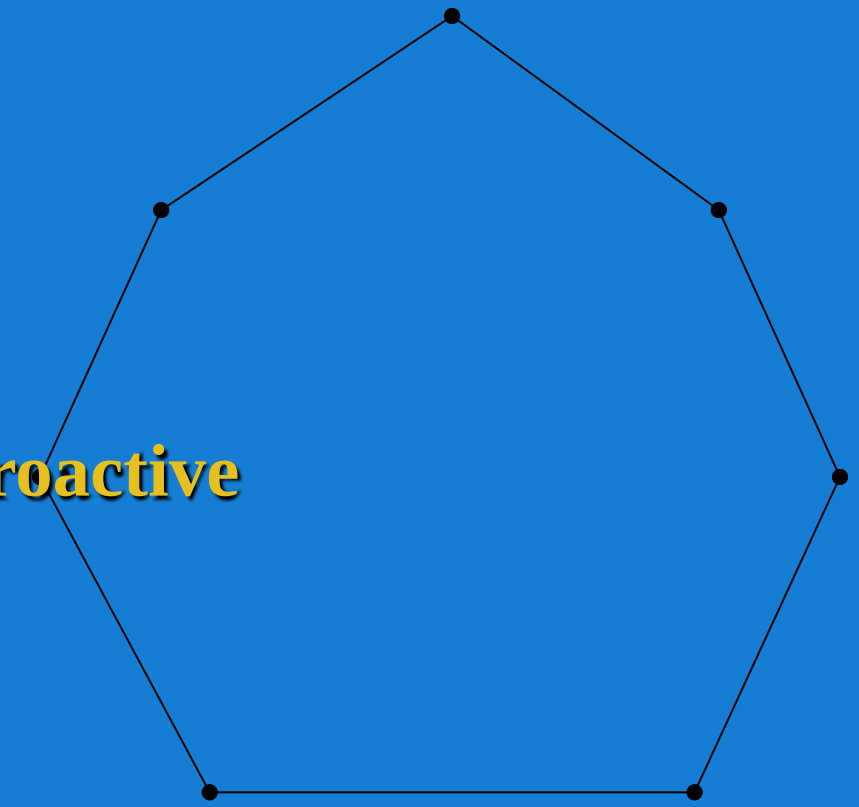
Success in graduate school: Staying the Course

- Disciplined
  - Time management
- Help you stay afloat during overwhelming periods



Success in graduate school: Staying the Course

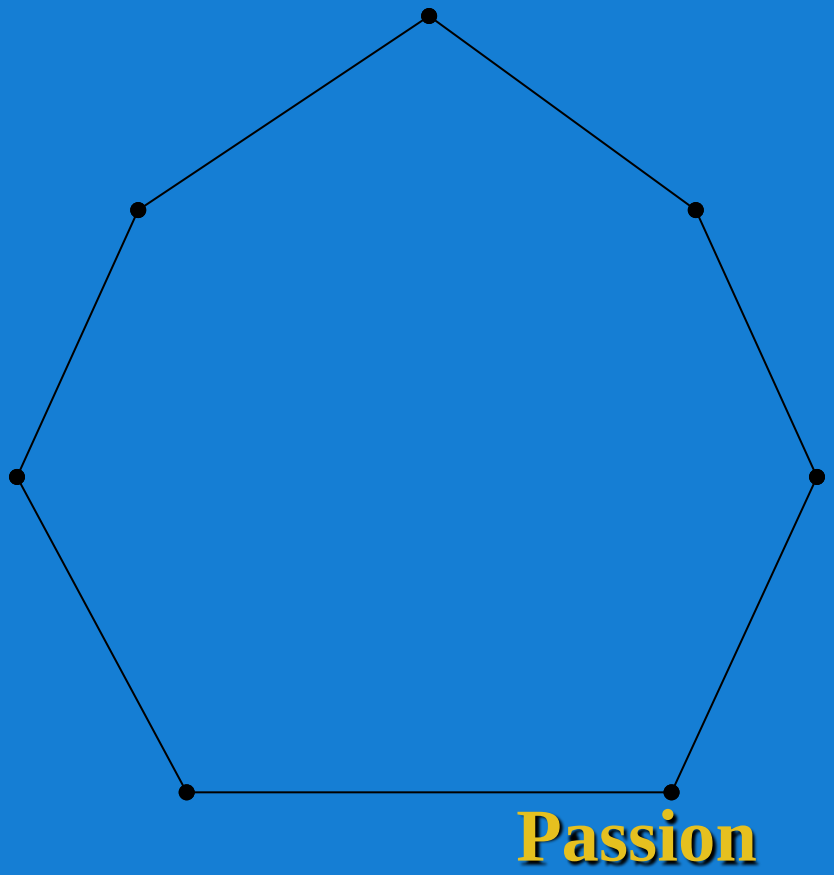
- Take ownership of your graduate education
- Take initiative
- Utilize the resources around you (students, faculty, labs, courses, jobs, potential mentors)



Success in graduate school: Staying the Course

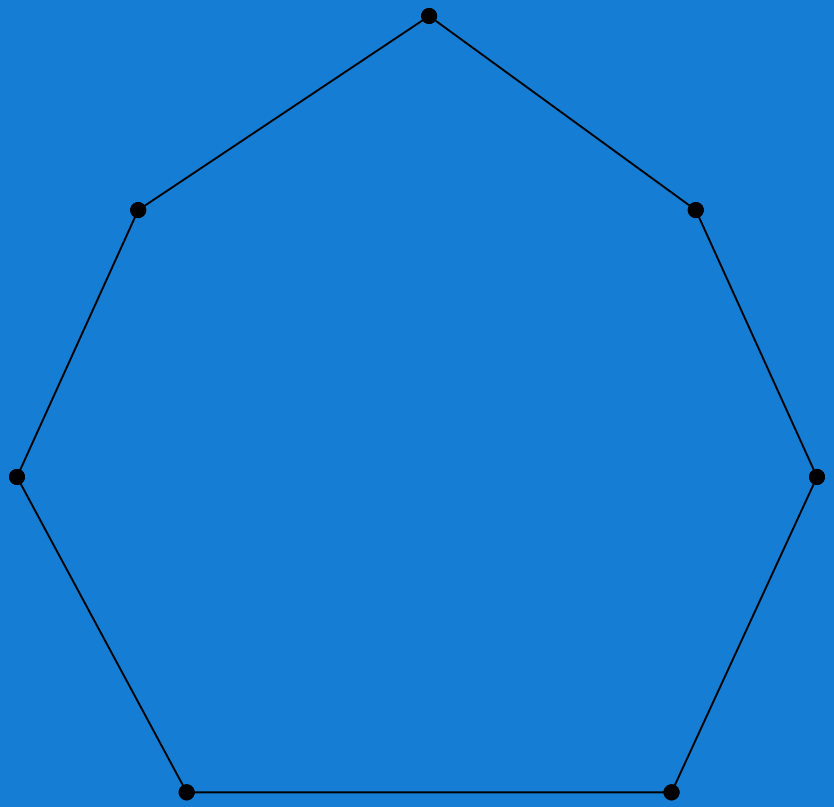
- **Determine what things motivate you. Passion will follow from these.**
  - **Learning**
  - **Discovery**
  - **Challenge**
  - **Purpose**

- **"Passion is the genesis of genius"** [Anthony Robbins]

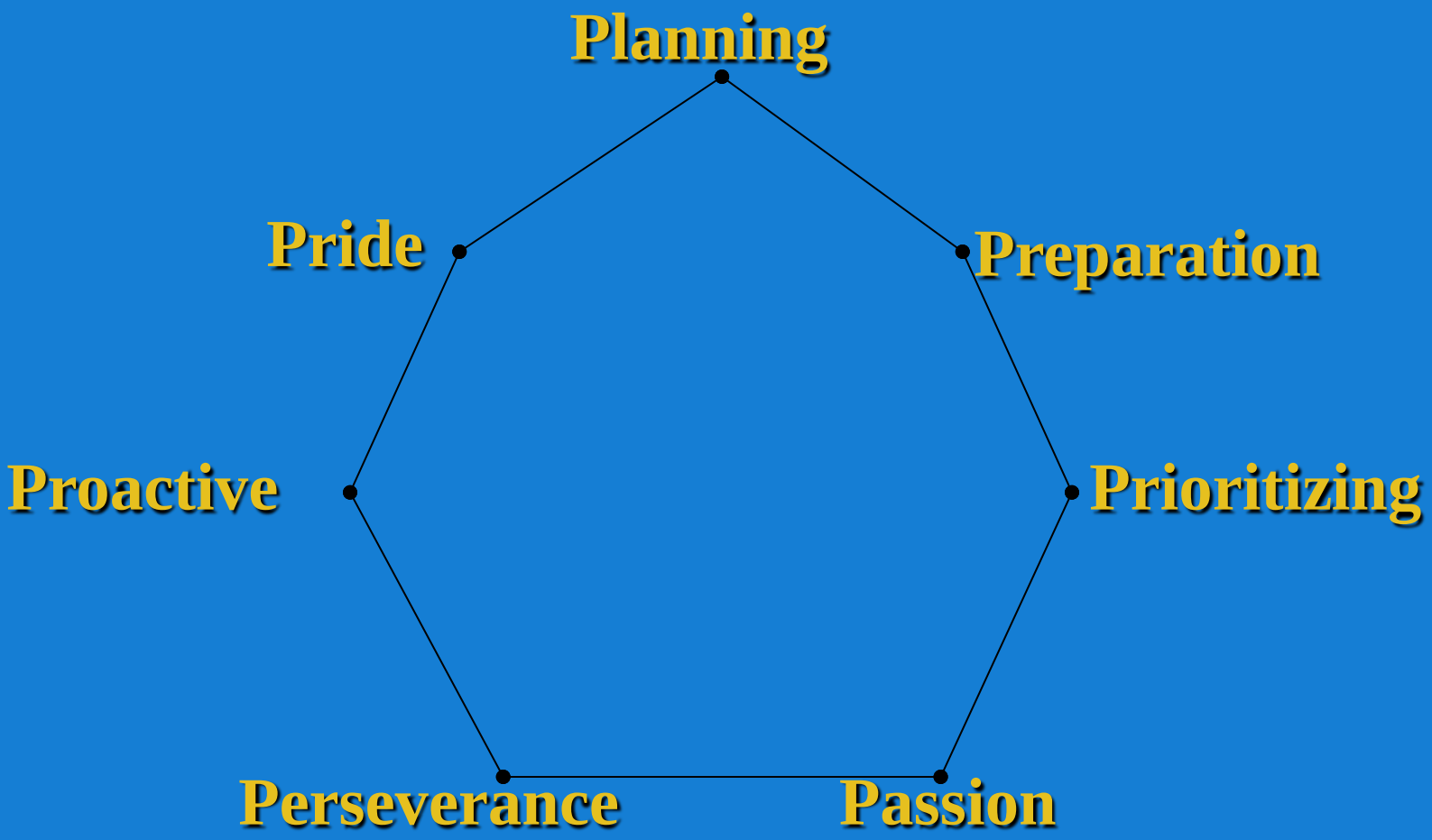


Success in graduate school: Staying the Course

- ***WHEN*** **you fall down, get up!!**
- **Your ultimate career trajectory depends heavily on how you respond to challenging times.**
- **"There are no secrets to success. It is the result of preparation, hard work, and learning from failure."** [Colin Powell]



Staying the Course



Biostatistics

- Provides a pathway to a bright future

- Increasing demand for biostatisticians, particularly in the big data era
  - Strong job market for master's and doctoral level statisticians

Biostatistics

The field needs you!

Thank you!

DuBois Bowman, Ph.D.

Columbia University

Department of Biostatistics

Professor and Chairman