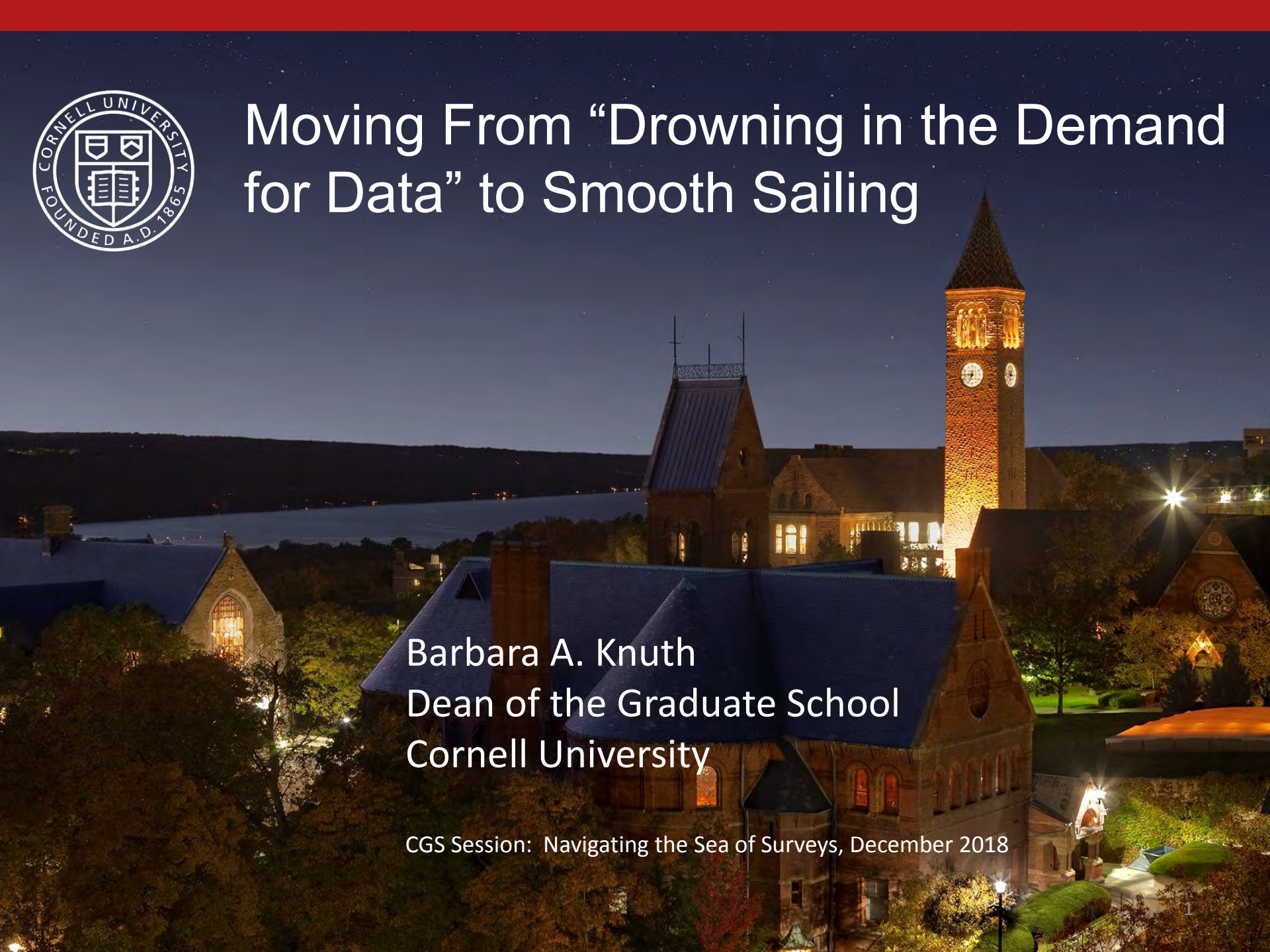


Moving From "Drowning in the Demand for Data" to Smooth Sailing

Barbara A. Knuth
Dean of the Graduate School
Cornell University

CGS Session: Navigating the Sea of Surveys, December 2018

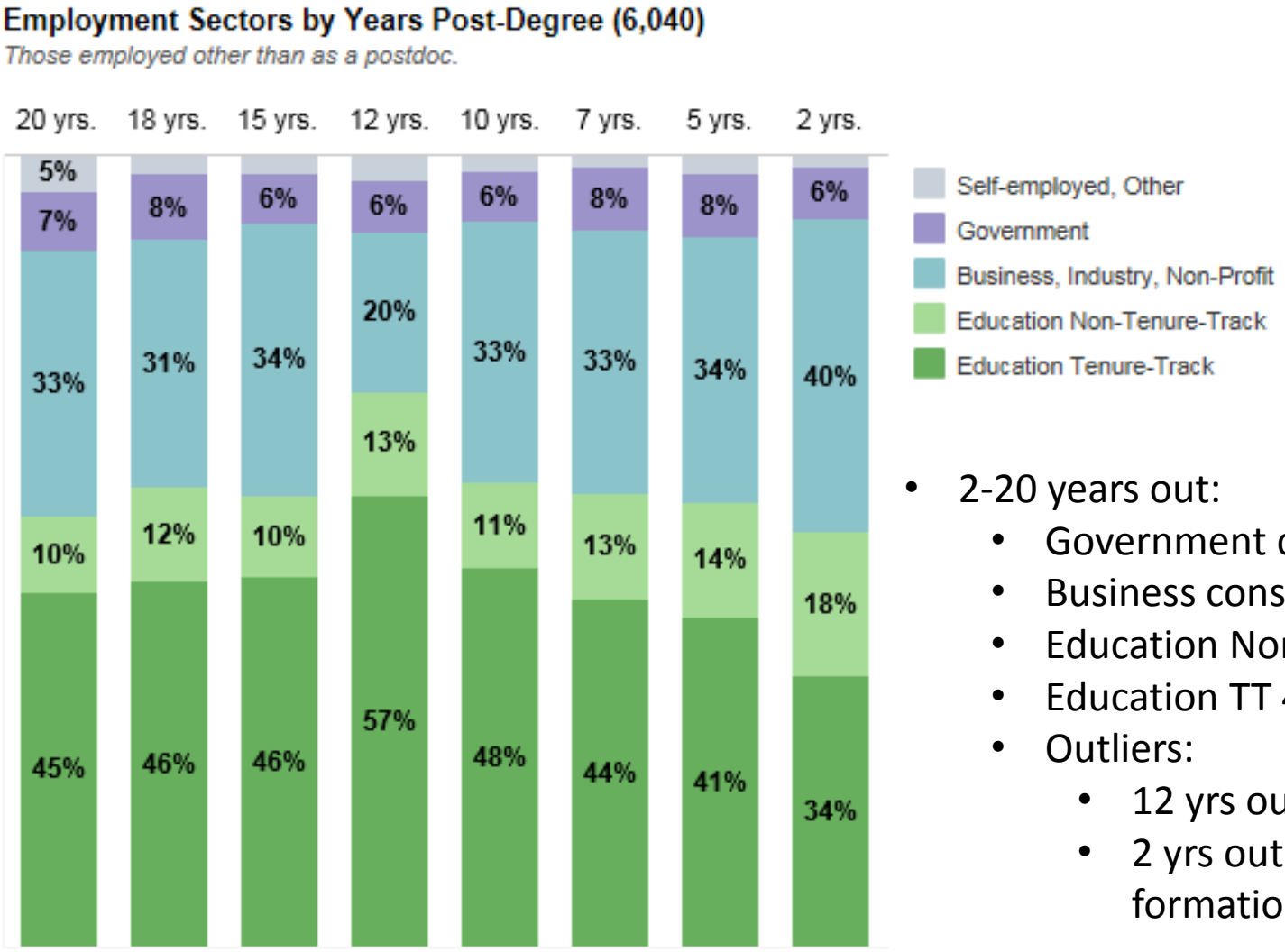
"Drowning in the Demand for Data"

- Growing expectations regarding data collection & transparency
  - External
  - Internal
- Strategizing and prioritizing with limited resources
  - Identify potential goals, uses, impacts
  - Identify potential audiences
- Achieving impact through data
  - Improve programs
  - Influence decisions (e.g., resource allocation, resource capture)
- Local & national context and implications

Using Data Effectively & With Impact @ Cornell:

- Annual report to the provost
  - Document change over time
  - Progress made
  - Areas for improvement, resource needs
- Biennial internal program review meetings
- Interactive dashboards for faculty directors of graduate students & staff assistants
- Public interactive dashboards & reports:
  - Informed prospective students
  - Transparency for enrolled students
  - Accountability by graduate programs
- Situating ourselves in national conversations

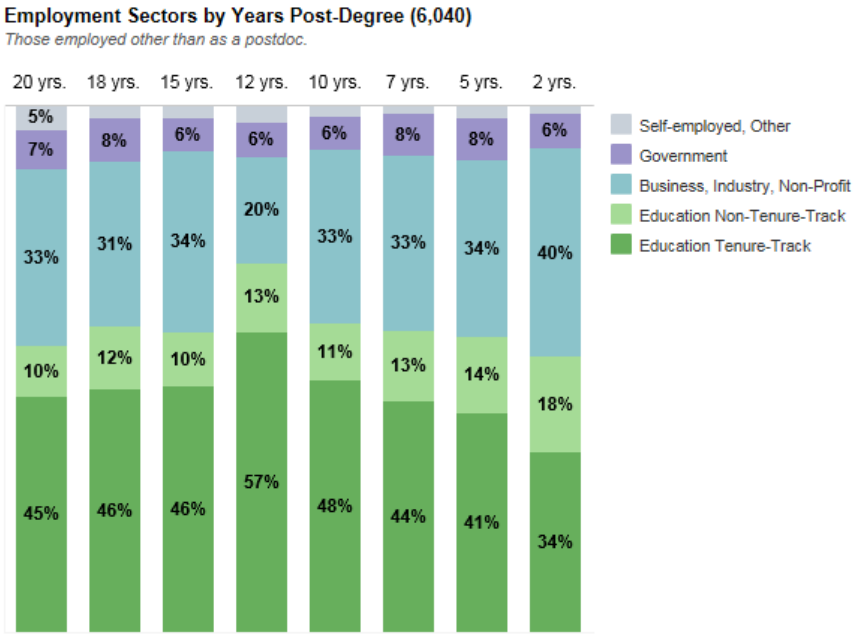
Cornell Example: Doctoral Career Outcomes vs. National Narratives (shift away from TT, into non-academic)



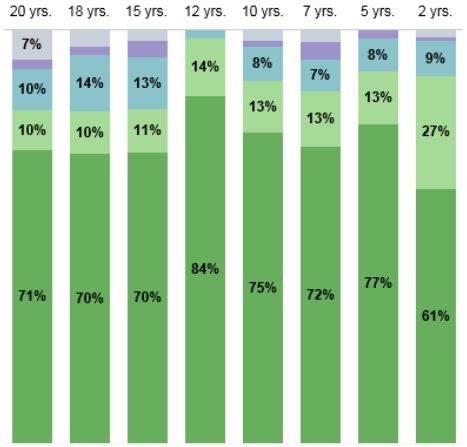
- 2-20 years out:
  - Government consistently 6-8%
  - Business consistently 31-34%
  - Education Non-TT 10-14%
  - Education TT 41-48%
  - Outliers:
    - 12 yrs out (??)
    - 2 yrs out (early career formation? Or sea change?)

Cornell Example: Doctoral Career Outcomes

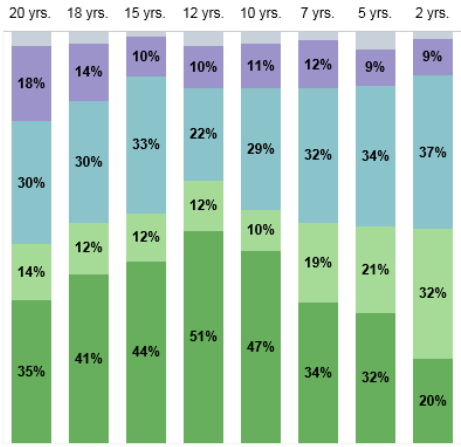
All Fields



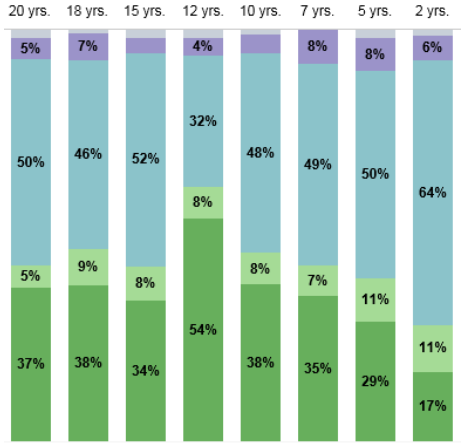
Humanities & Arts



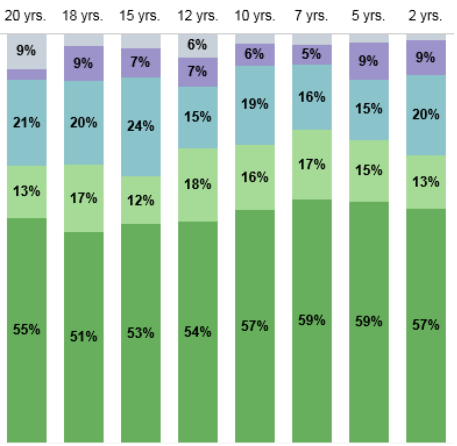
Life Sciences



Physical Sciences



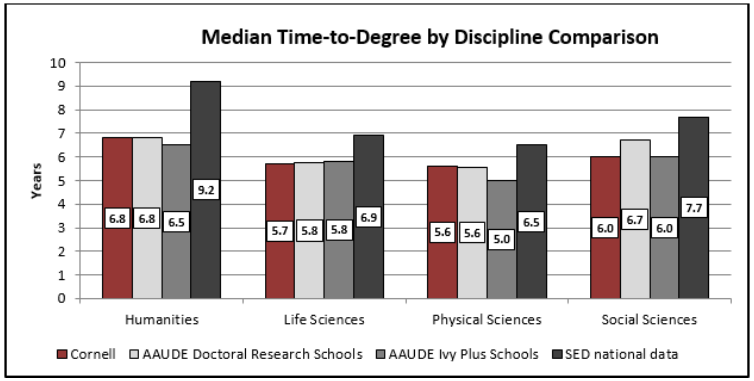
Social Sciences



Cornell's Multi-institutional Data-Related Engagements

Program Evaluation and Improvement

- AGEP (diversity for the future professoriate, campus climate)
- BEST (career exploration)
- CIRTL (preparing future faculty)



"Private" Collaborative Data Exchange

- AAUDE (data comparisons)

"Public" Data Transparency

- CNGLS (graduate students & postdocs)
- Council of Graduate Schools (enrollment, international, etc.)
- SED, etc.

Multi-institutional Benefits

+ Generate comparative data (vs. peers)
+ Contextualize graduate education (nationally)
+ Share what works for evidence-based improvement
+ Reduce risk from unilateral data transparency

Multi-institutional Benefits and Challenges

+ Generate comparative data (vs. peers)
+ Contextualize graduate education (nationally)
+ Share what works for evidence-based improvement
+ Reduce risk from unilateral data transparency

– Agree to shared or flexible data definitions
– Reconfigure data for multiple internal and external needs
– Factor in existing data collected prior to multi-institution agreements
– Reconcile lumpers vs. splitters
– Coordinate data across different institutional structures
  - internally (e.g., Graduate Schools vs. Postdoc Offices vs. HR vs. IR)
  - externally (e.g., CIP code variation, data sharing restrictions vs. sunshine laws)

Weighing Pros & Cons of Participation

- What is the potential benefit?
  - To my graduate school
  - To my institution
  - To my students/faculty/staff
  - To higher education

- How much will it cost to participate?
  - $$
  - Staff time
  - Student/faculty time

- What are the costs of not participating?
  - Institutional reputation, visibility
  - Informed decisions: individually, locally, regionally, nationally

"Private" Data Example:

AAUDE: The Association of American Universities Data Exchange

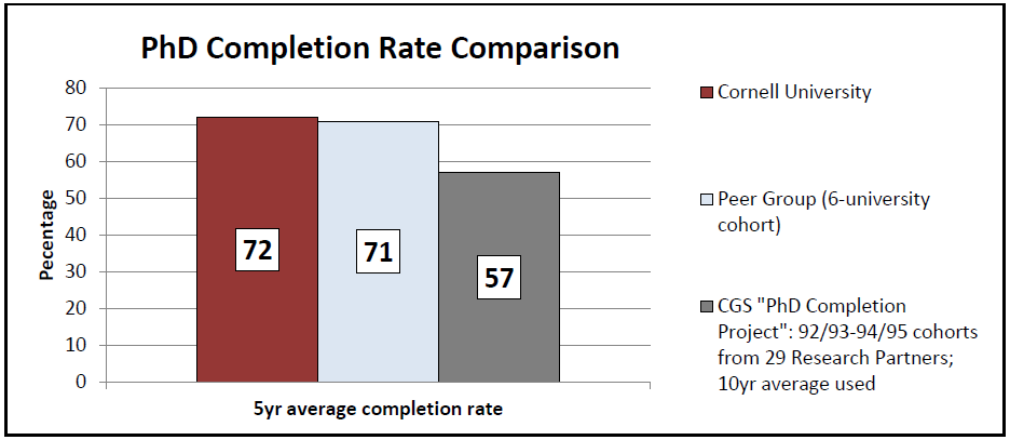
- AAU institutions
- Participate in exchanging data/information to support decision-making for graduate education at institution
- Graduate education data are not public
- If you submit data, you have access to data

✓ Time to degree

✓ Degree completion rate

✓ PhD exit survey

✓ PhD career outcomes

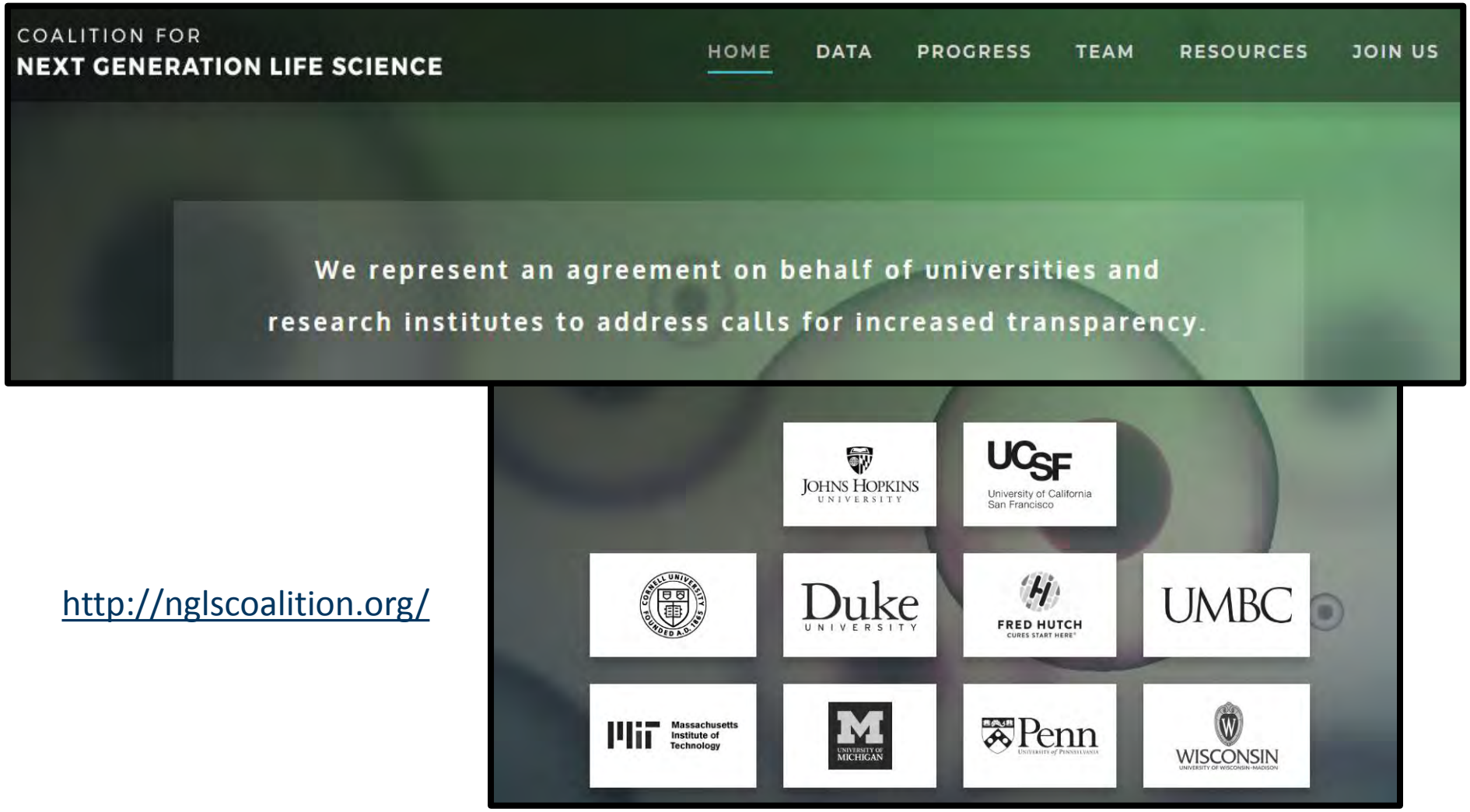


- Well-developed protocols and definitions for each

AAUDE Experience:

- ✓ Data protocols explicitly consider compatibility issues:
  - Clear definition for each data element
  - Well-documented data dictionary
  - Compare and contrast similar data elements commonly available
  - Discipline crosswalks
  - Relation to IPEDS, SED, CGS and other data collection systems
- ✓ Rules for small cell sizes; safeguard individual privacy
- ✓ Rules to enable meaningful aggregation
- ✓ Not for rankings
- ✓ Peer comparisons
  - Internally: group member names only, not individual schools
  - Externally: masked, group name only

Multi-institution Public Data Example: Coalition for Next Generation Life Science



http://nglscoalition.org/

Multi-Institution Data Transparency Example: Coalition for Next Generation Life Science (CNGLS)

- Voluntary agreement by universities and research institutes (26+) to address calls for increased transparency re: life science trainees
  - Graduate students
  - Postdoctoral scholars
- Post data using common definitions (by demographic groups)
  - Admissions
  - Matriculation
  - MTTD and MT in postdoc status
  - Completion
  - Career outcomes (taxonomy by job sector & career type)

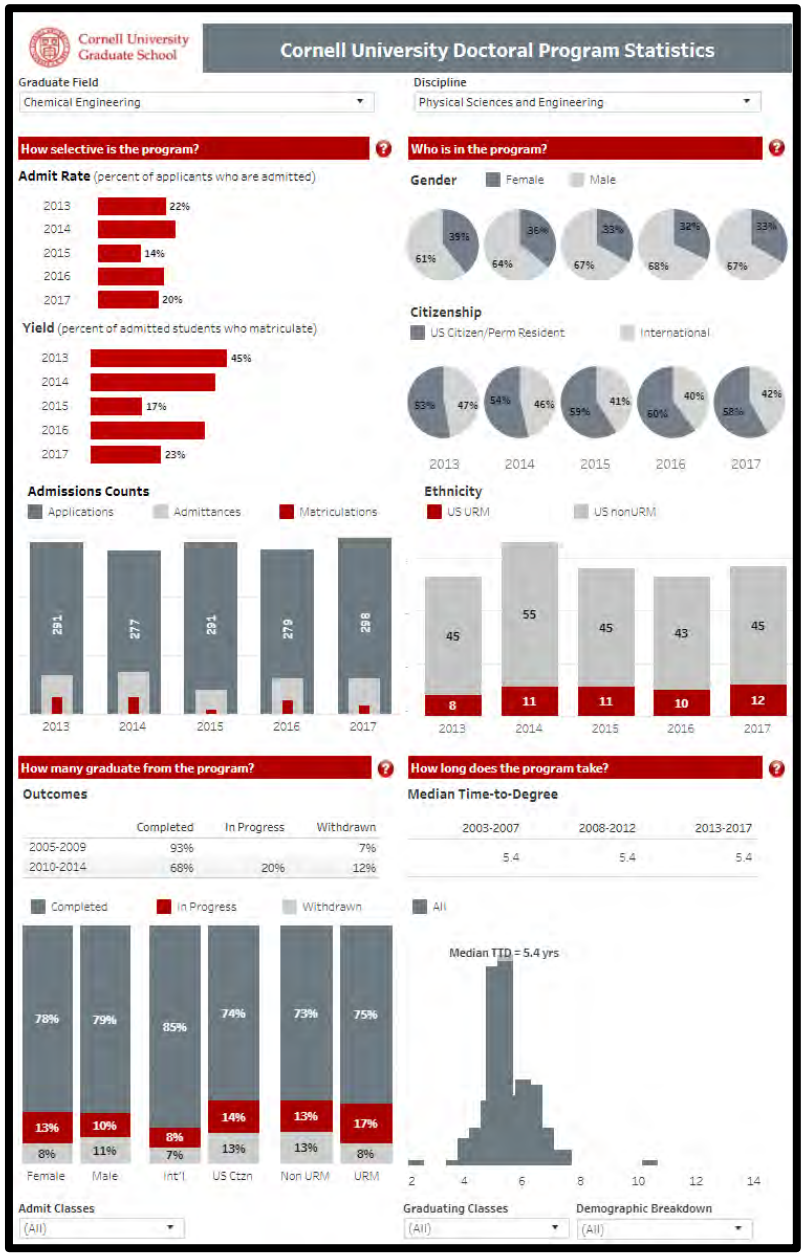
Multi-Institution Data Transparency Example: Coalition for Next Generation Life Science (CNGLS)

- Voluntary agreement by universities and research institutes (26+) to address calls for increased transparency re: life science trainees
  - Graduate students
  - Postdoctoral scholars

- Post data using common definitions (by demographic groups)
  - Admissions
  - Matriculation
  - MTTD and MT in postdoc status
  - Completion
  - Career outcomes (taxonomy by job sector & career type)

- BUT:
  - Most graduate schools have responsibilities for multiple disciplines
  - Many graduate schools don't curate postdoctoral scholar data
  - Some graduate schools already have internal & other partnership protocols for these data

Cornell Example: CNGLS Data

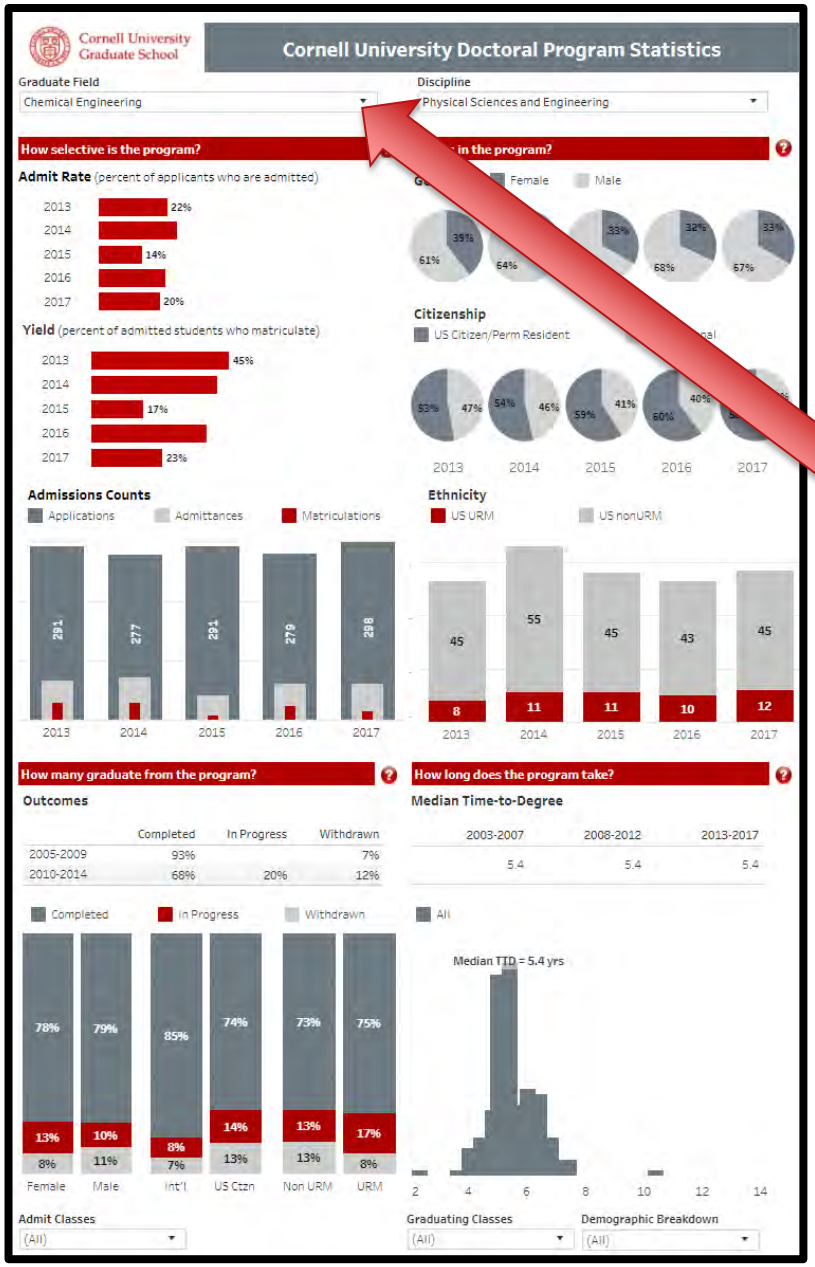


- Posted on websites:
  - Graduate School Academics Info
  - Graduate School Catalog of Degree Programs
  - Individual Degree Programs
  - Office of Postdoctoral Studies (postdoc data)

- No required format
- No single posting location to compare schools

https://tableau.cornell.edu/views/CornellUniversityGraduateSchoolDoctoralProgramStatistics/TheOneDashboard?iframeSizedToWindow=true&:embed=y&:showAppBanner=false&:display_count=no&:showVizHome=no

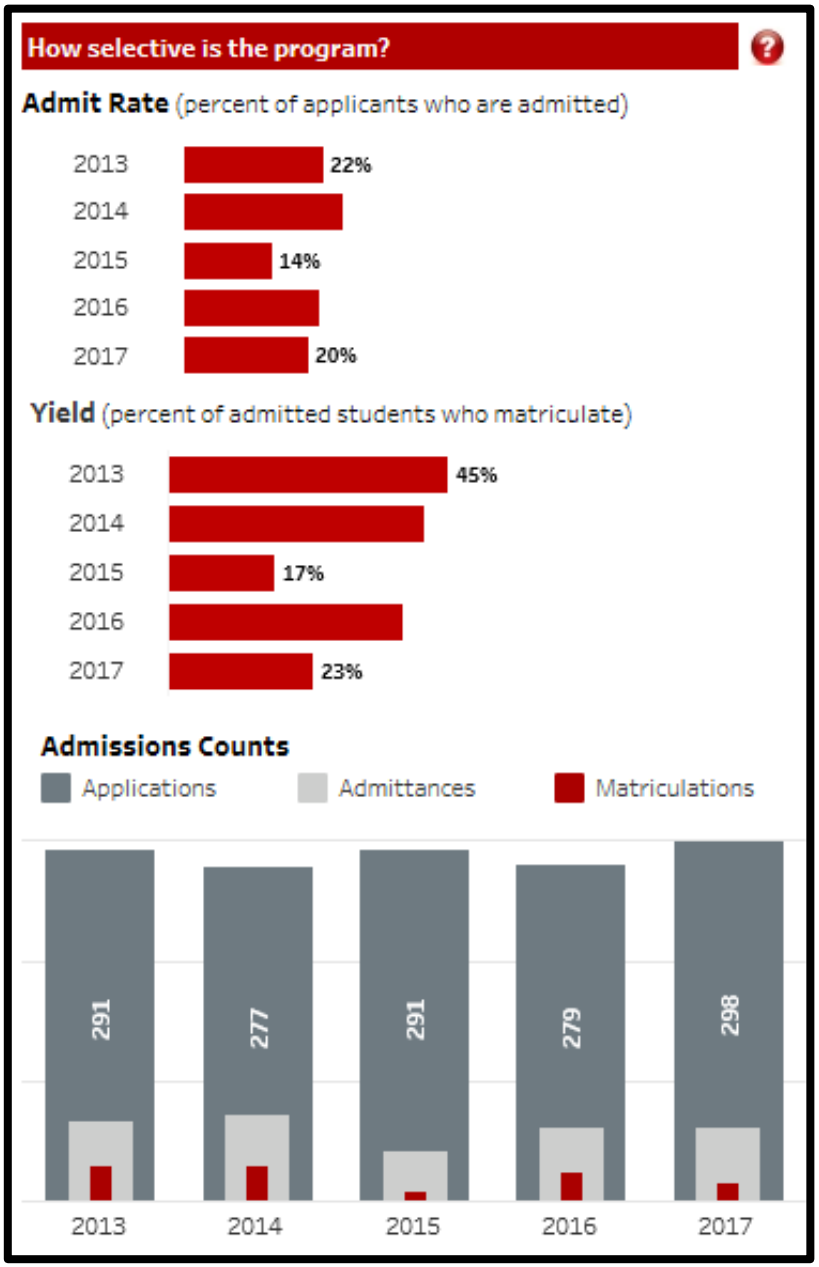
Cornell Example: CNGLS – graduate student demographics



Filter by Graduate Field or Discipline

(CIP codes may differ among institutions)

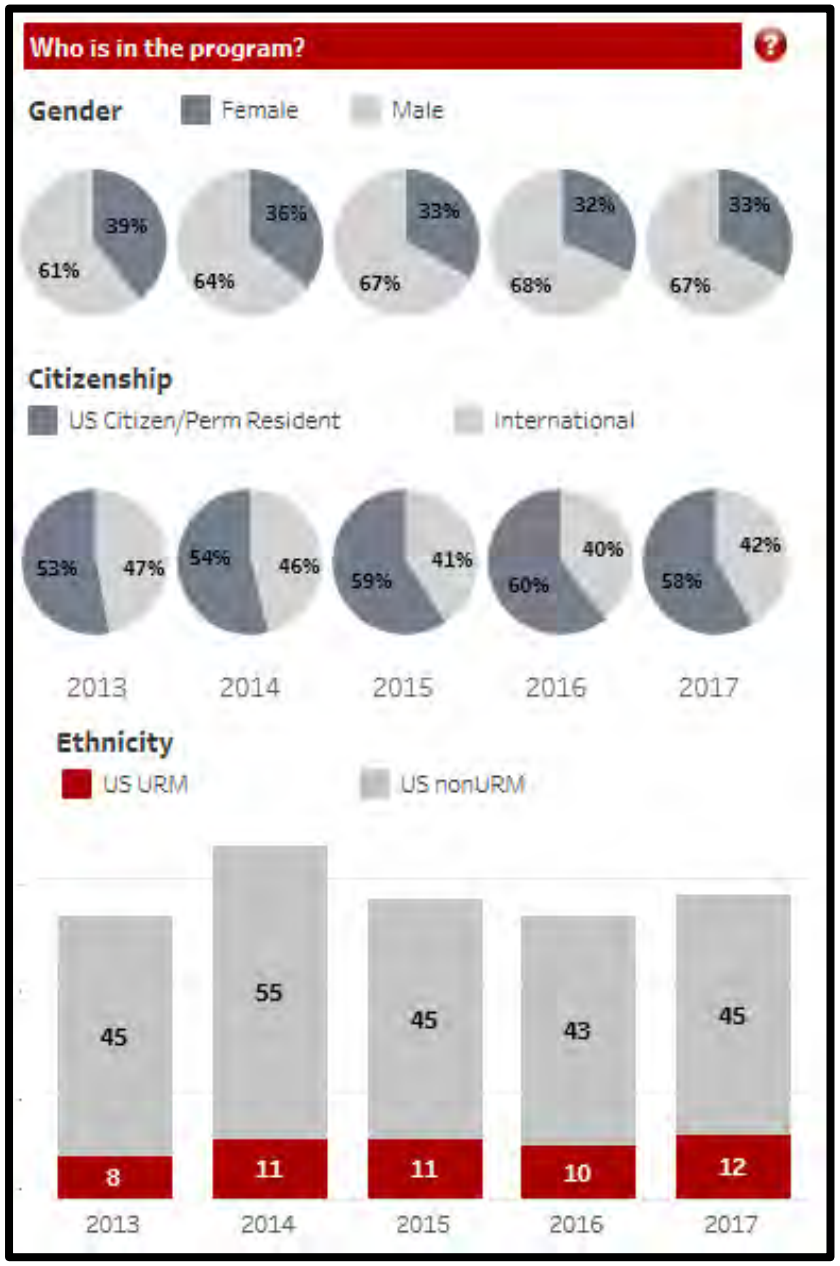
Cornell Example: CNGLS graduate students



How selective is the program?

- Admit Rate
- Yield
- Applications
- Admittances
- Matriculations

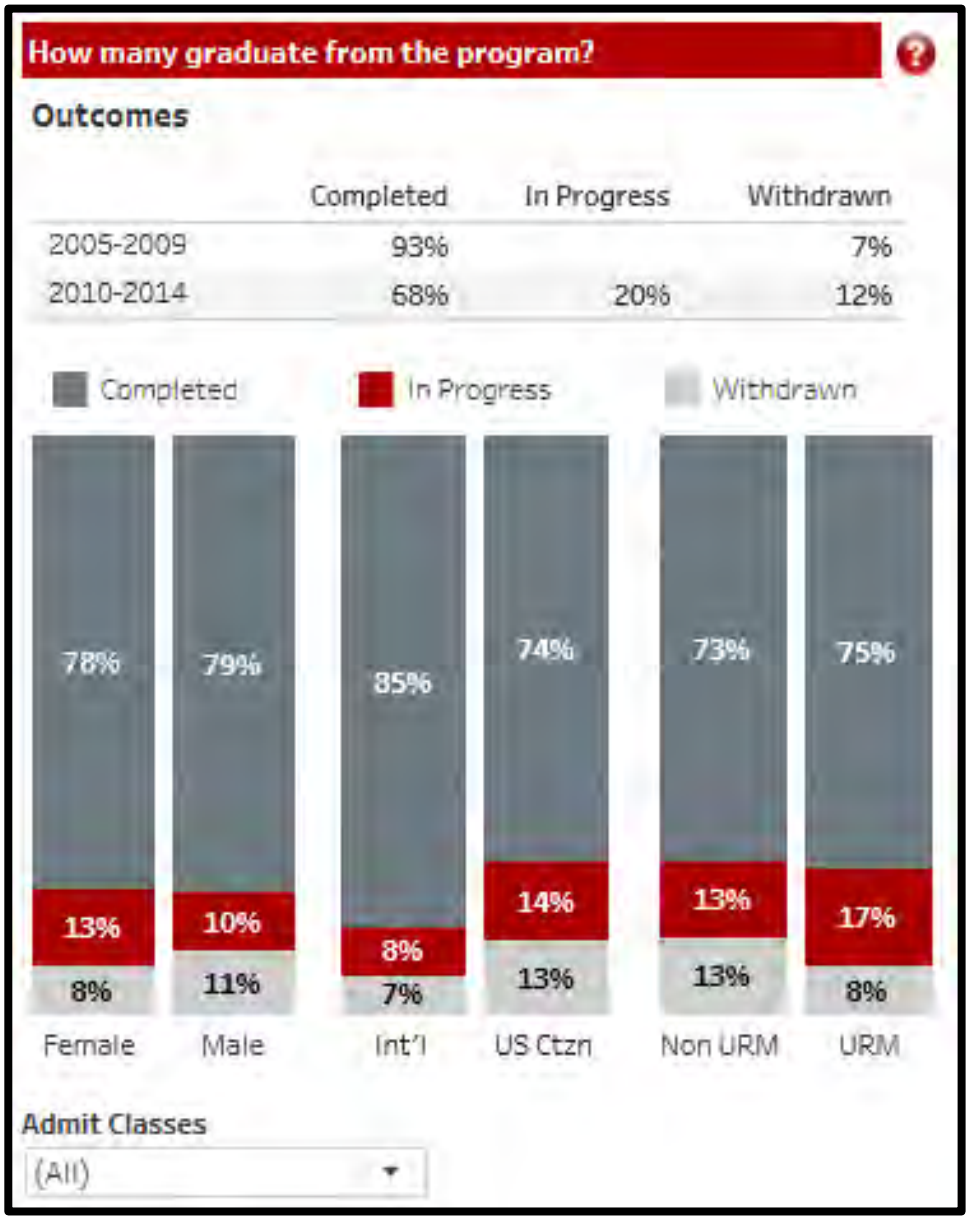
Cornell Example: CNGLS graduate students



Who is in the program?

- Gender
- Citizenship
- Ethnicity

Cornell Example: CNGLS graduate students

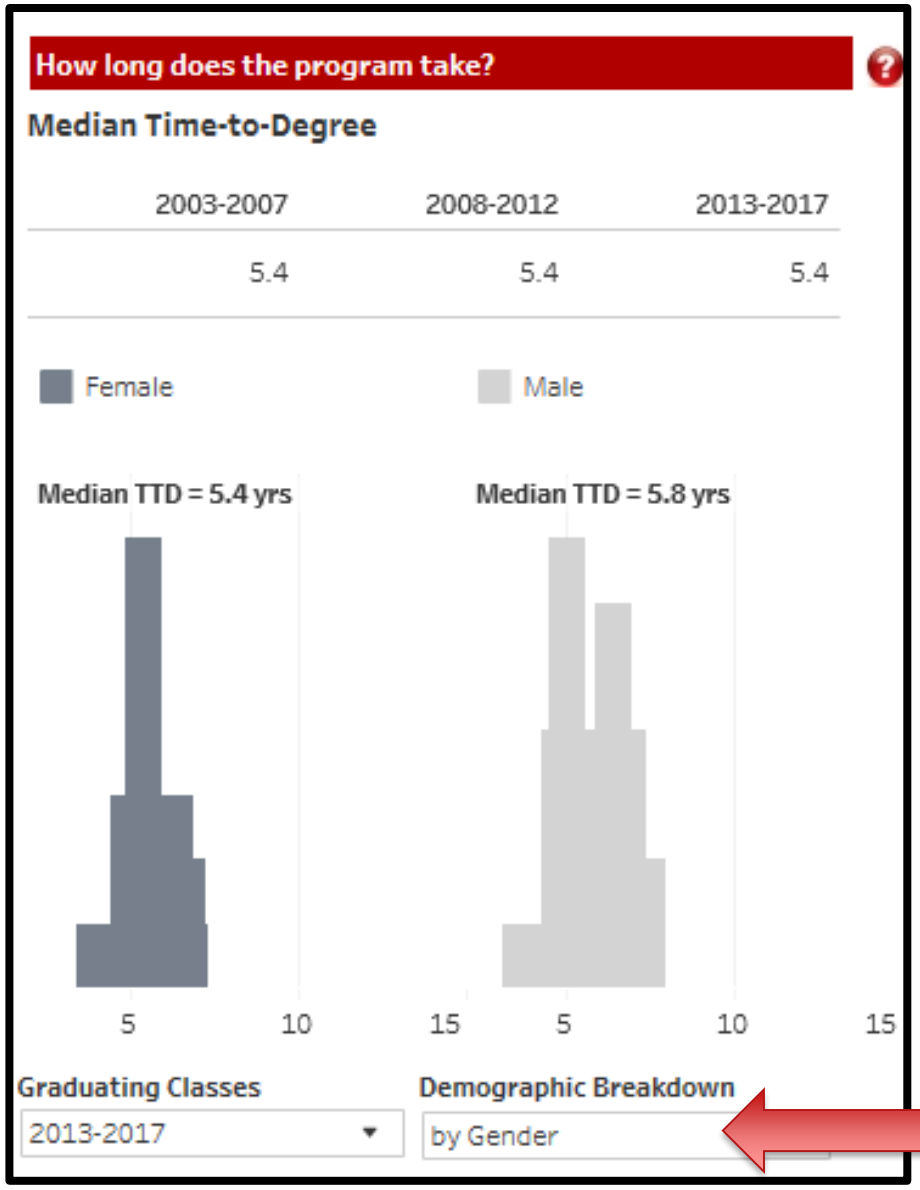


How many graduate from the program?

- Completed
- In Progress
- Withdrawn

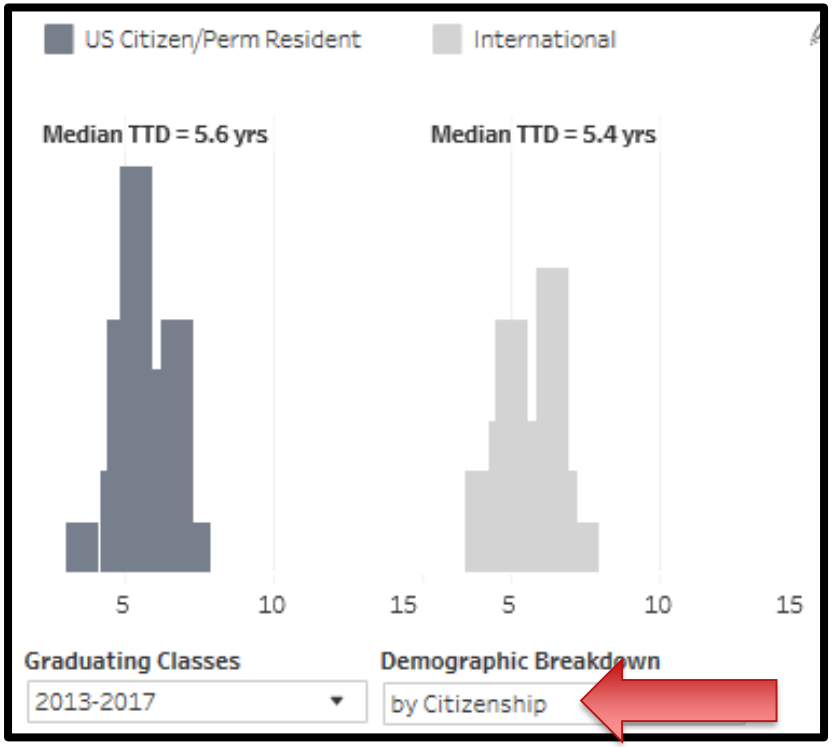
- Gender
- Citizenship
- Ethnicity

Cornell Example: CNGLS graduate students



How long does the program take?

- Median Time-to-Degree
- Filter by Graduating Classes
- Filter by Demographics



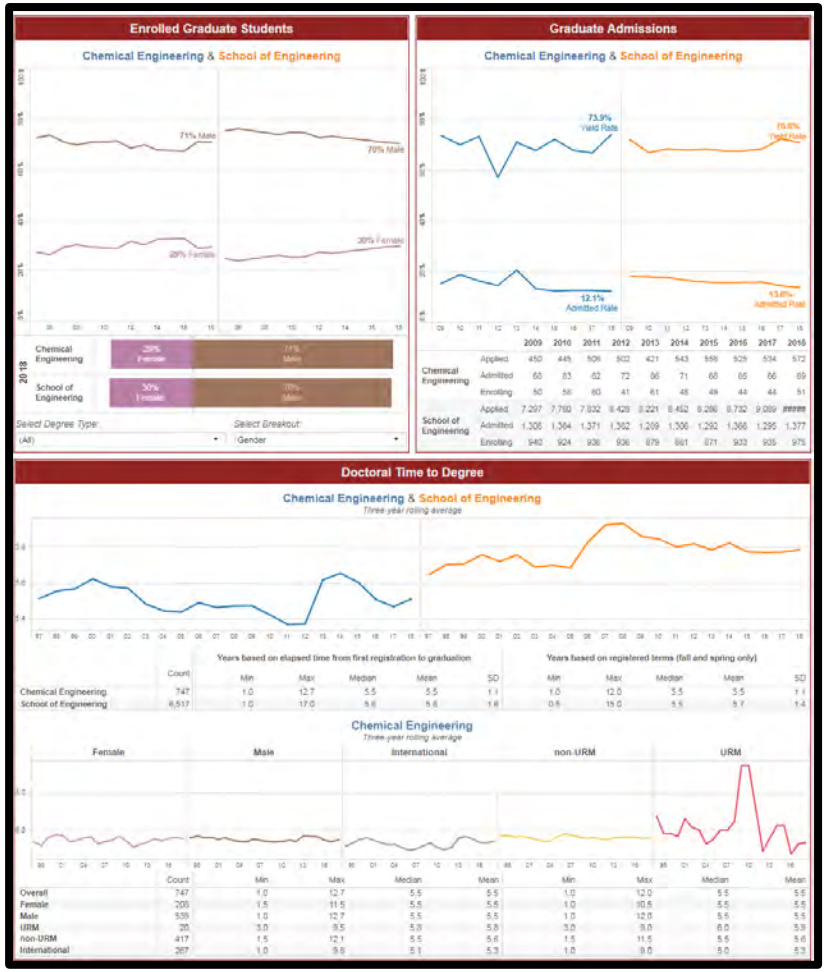
Flexible Data Visualization Approaches in Coalition:

UCSF

Flexible Data Visualization Approaches in Coalition:

UCSF

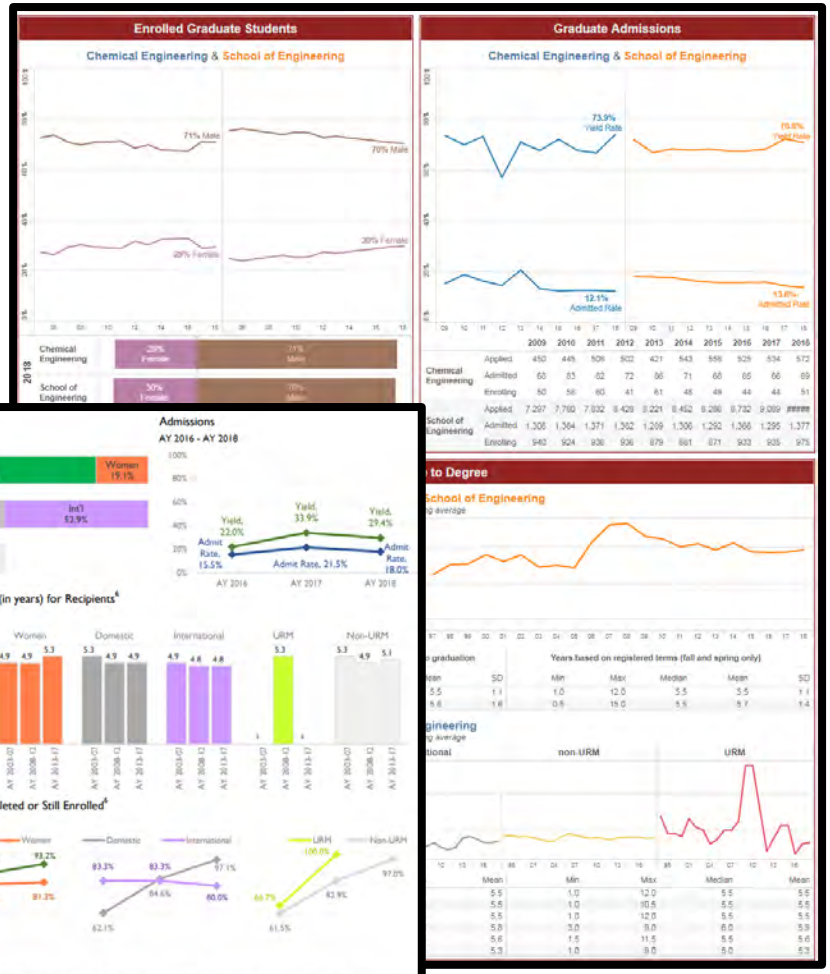
MIT



Flexible Data Visualization Approaches in Coalition:

UCSF

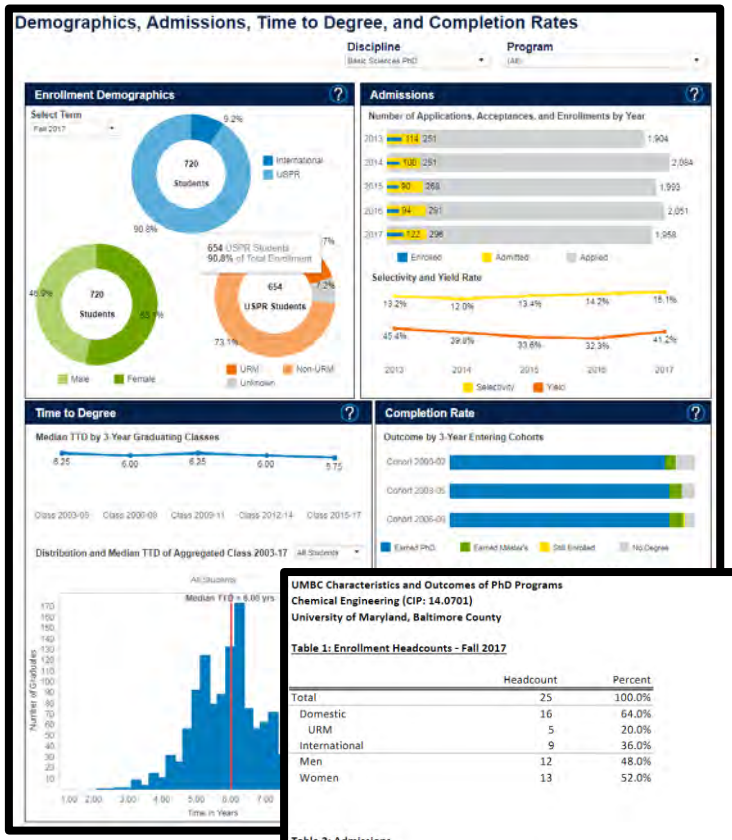
MIT



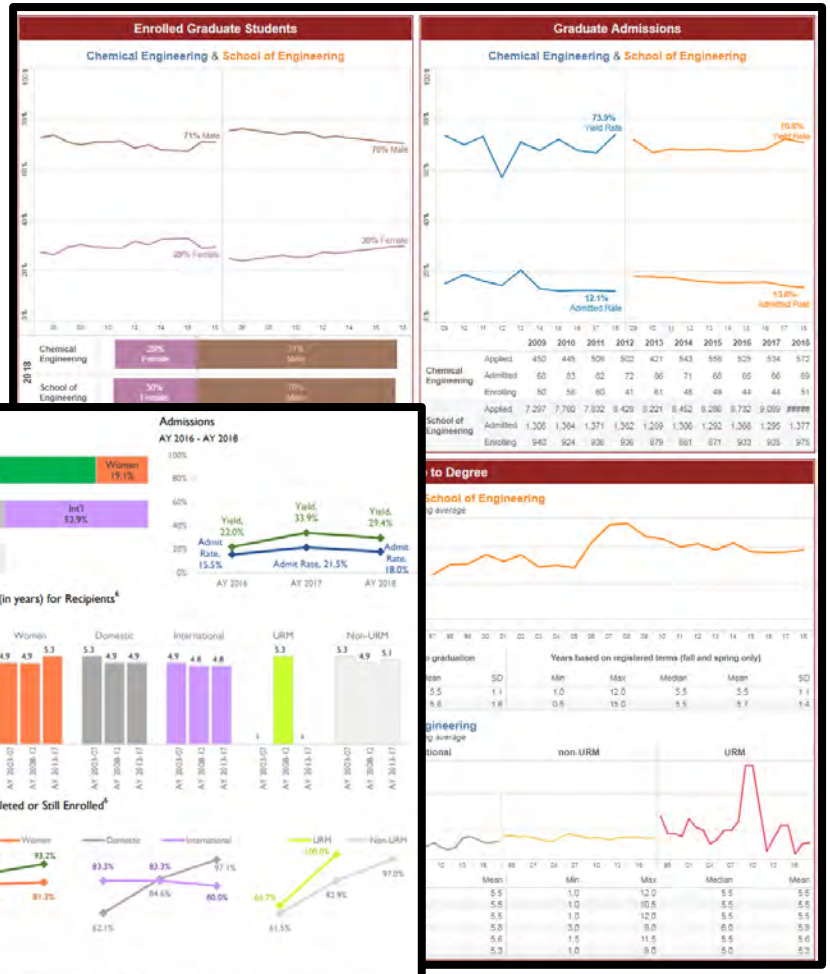
PENN

Flexible Data Visualization Approaches in Coalition:

UCSF

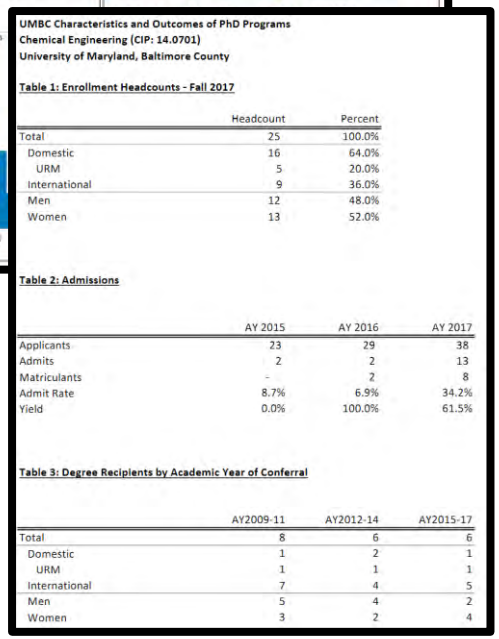


MIT



PENN

UMBC



UMBC Characteristics and Outcomes of PhD Programs
Chemical Engineering (CIP: 14.0701)
University of Maryland, Baltimore County

Table 1: Enrollment Headcounts - Fall 2017

	Headcount	Percent
Total	25	100.0%
Domestic	16	64.0%
URM	5	20.0%
International	9	36.0%
Men	12	48.0%
Women	13	52.0%

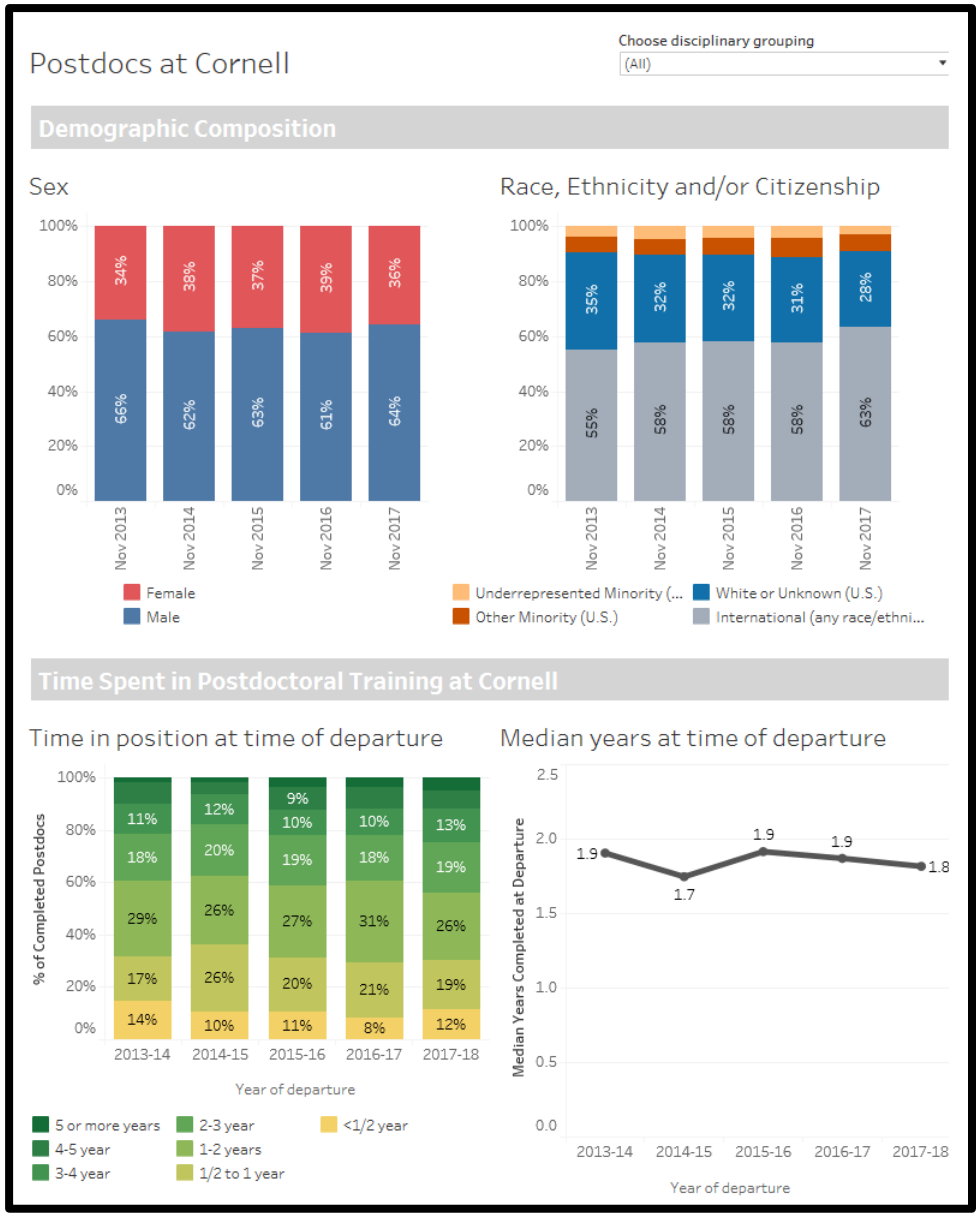
Table 2: Admissions

	AY 2015	AY 2016	AY 2017
Applicants	23	29	38
Admits	2	2	13
Matriculants	-	2	8
Admit Rate	8.7%	6.9%	34.2%
Yield	0.0%	100.0%	61.5%

Table 3: Degree Recipients by Academic Year of Conferral

	AY2009-11	AY2012-14	AY2015-17
Total	8	6	6
Domestic	1	2	1
URM	1	1	1
International	7	4	5
Men	5	4	2
Women	3	2	4

Cornell Example: CNGLS Postdoc Demographics



- CNGLS caused us to look at, and collect, data we hadn't before
- Required collaboration with other campus offices

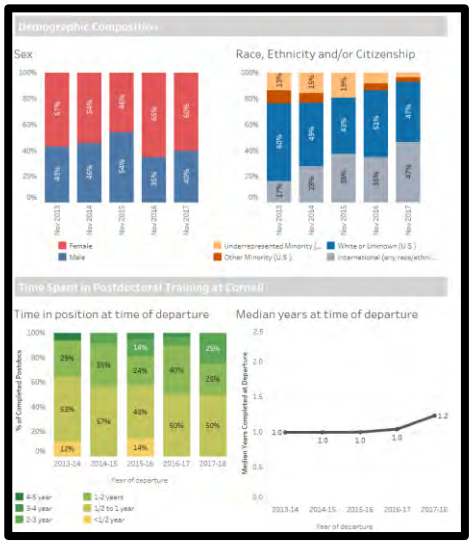
- Postdocs:
  - Sex
  - Race/ethnicity
  - Citizenship
  - Time in position
  - Median time to departure

- Coming Soon: Job after departure

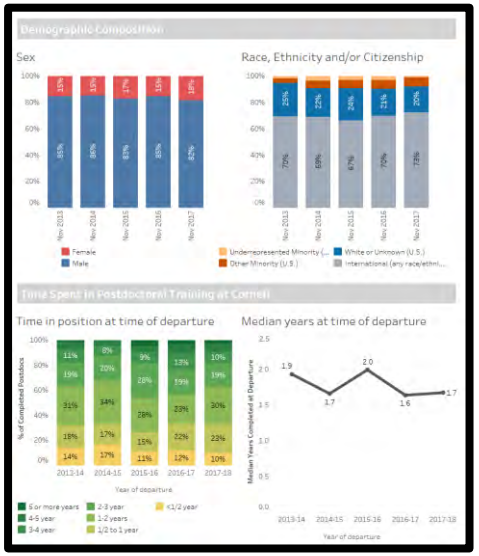
Cornell Example: CNGLS Postdoc Demographics

Plant, Animal, Life Sciences

Humanities, Arts, Design



Physical Sciences, Engineering, Math



Social & Behavioral Sciences

Lessons Learned

- Opportunity for comparative data to inform decisions & resource requests
- Promote internal communications at multiple institutional levels
- Share what works and build on progress
- Collaboration may result in better outcomes
  - Data visualizations
  - Provisions for aggregation vs. specificity
  - Data masking as appropriate
  - Consensus, within reason, on data definitions
- Contextualize graduate education in national conversations
- Flexibility and adaptability are important
  - Multiple data transparency efforts for different purposes
  - Real costs of managing multiple efforts