

National Trends in Undergraduate Statistics and Data Science Courses from a Stratified Sample

Amira Colón (California State University - Monterey Bay)

Alana Unfried (California State University - Monterey Bay)

April Kerby-Helm (Winona State University)

Douglas Whitaker (Mount Saint Vincent University)

Leyla Batakci (Elizabethtown College)

Marjorie E. Bond (Penn State University)

Michael A. Posner (Villanova University)



DUE-2013392

A man with a beard and a black turtleneck, standing with his arms outstretched, looking slightly to the side.

How many schools have

Data Science

courses and programs?

How many schools have

Statistics

courses and programs?



Motivational Attitudes in Statistics and Data Science Education Research

- NSF grant-funded project to investigate trends and attitudes in statistics and data science education.
- Developing attitudinal surveys for students and instructors
- Understanding educational trends to inform schools and educators of areas in need of growth or allocation of resources

What percent of US colleges and universities offer Intro Stat, Intro Data Science, and/or Statistics or Data Science Programs?

Sampling Design

- **Population of Interest:** US colleges and universities
 - Excluding: Non-US Schools, service schools, graduate only schools, under 300 enrollment, some special focus schools
- **Data Source:** The Dept. of Education College Scorecard Database, a national database of institutional characteristics of US colleges combined with the Carnegie Classification Database.
- **Stratified Sample:** Eight strata based on their selectivity, highest level of degree, and minority-serving status.

Defining Strata

- **Highest Degree Offered** - **Grad, *, Bach, Assoc**
 - Using Carnegie basic definitions then predominant degree offered then highest degree offered
 - Grad and Bach combined (*) for MSI due to small stratum size
- **Selectivity** - **Selective+, Inclusive, X (Associates)**
 - Selective and highly selective into "Selective +" due to small sample
 - College scorecard primary, Carnegie secondary
- **Minority Serving Institution** - **MSI, Non-MSI**
 - HBCU, HSI, Tribal Colleges

| Strata | Definition | N_i | n_i | Ex. School |
|--------|------------------------------------|-------|-----|--|
| GSN | Grad, Selective+, not MSI | 438 | 19 | Ithaca College, East Tennessee State University |
| GIN | Grad, Inclusive, not MSI | 386 | 18 | King's College, University of New England |
| *SM | Bach/Grad, Selective+, MSI | 64 | 12 | The University of Texas at Austin, Spelman College |
| *IM | Bach/Grad, Inclusive, MSI | 207 | 12 | Fayetteville State University, Cal State Monterey Bay |
| BSN | Bach, Selective+, not MSI | 232 | 11 | Swarthmore College, Franklin and Marshall College |
| BIN | Bach, Inclusive, not MSI | 315 | 16 | Wilmington College, Miami University-Middletown |
| AxM | Associate, All (Inc/Sel+), MSI | 259 | 11 | Middlesex College, Las Positas College |
| AxN | Associate, All (Inc/Sel+), not MSI | 850 | 48 | Cape Cod Community College, Pennsylvania Institute of Technology |
| | Total: | 2,751 | 147 | |

Database

- G. Sheets database was designed to house university and course level data collected from sampled institutions.
 - Variables collected include course name, description, units, prerequisites, if they offer a major or minor in either subject, etc.
- Information was collected from publicly-available college and university websites and catalogs and stored into the database.

| Level | Variable Name | Description |
|------------|------------------------------|---|
| University | Unit ID | The university's ID |
| University | School Name | Make sure it matches |
| University | Website URL | Include the university's website URL |
| University | Stat Major | Does the university offer a statistics major |
| University | Stat Minor | Does the university offer a statistics minor |
| University | DS Major | Does the university offer a data science major |
| University | DS Minor | Does the university offer a data science minor |
| University | Catalog URL | Include the university's catalog URL |
| University | HSI | Is the university a Hispanic Serving Institution |
| Course | University | Unit ID |
| Course | University | School (official from catalog) |
| Course | Course # | The course number |
| Course | Course Name | Name of the course |
| Course | Department | Department from catalog |
| Course | Classification | Is the course an introductory, intermediate, or advanced course |
| Course | Course Description | Description of course |
| Course | Course Prerequisites | Does the course have prerequisites |
| Course | Exam/Placement Prerequisites | Does the course have exam or placement prerequisites |
| Course | Units | Number of units the course is worth |
| Course | # Sections | Number of sections offered |
| Course | FA23 Schedule Available | Is the Fall 2023 course schedule available |

University

| School (official from college scorecard) | Strata | Website URL | Intro Stat Course | Intro DS Course | Include in Samp | Stat major | Stat Mino |
|--|--------|---|-------------------|-----------------|-----------------|------------|-----------|
| St Philip's College | AxM | https://www.alan | Yes | No | Yes | No | No |
| Middlesex College | AxM | https://www.midc | Yes | Yes | Yes | No | No |
| Dodge City Community College | AxM | https://dc3.edu | Yes | No | Yes | No | No |
| Remington College - Nashville Campus | AxN | https://www.remi | No | No | No | No | No |
| City Colleges of Chicago-Malcolm X College | AxM | https://www.ccc | Yes | No | Yes | No | No |
| Las Positas College | AxM | https://www.lasp | Yes | No | Yes | No | No |
| Norwalk Community College | AxM | https://norwalk.e | Yes | No | Yes | No | No |
| Skyline College | AxM | https://www.skyli | Yes | No | Yes | No | No |
| CUNY Queensborough Community College | AxM | https://www.qcc | Yes | Yes | Yes | No | No |
| University of New Mexico-Taos Campus | AxM | https://taos.unm | Yes | No | Yes | No | No |
| Arkansas State University Mid-South | AxN | https://www.asur | Yes | No | Yes | No | No |
| Cape Cod Community College | AxN | https://live-capec | Yes | No | Yes | No | No |
| Grand Rapids Community College | AxN | https://www.grcc | Yes | Yes | Yes | No | No |

Course

| School (official from c | Course # | Course Name | Department | Classification | Course Descrip | Prerequisites | Calc Pre |
|-------------------------|------------|-------------------------------------|----------------|----------------|--------------------|-------------------|----------|
| Alice Lloyd College | BUS 295 | Business Statistics and Research | Business | Intro to Stat | This course exa | Math 112, 113 or | No |
| American University | STAT 202 | Basic Statistics | Statistics | Intro to Stat | Basic Statistics (| No | No |
| American University | STAT 204 | Introduction to Business Statistics | Business | Intro to Stat | Introduction to B | No | No |
| American University | DATA 413 | Data Science | Data Science | Intro to DS | Data Science (3 | DATA-412. | No |
| Arkansas State Unive | MATH 2133 | Introduction to Statistics | Mathematics | Intro to Stat | An algebra-base | MATH 1213 Rea | No |
| Bennington College | MAT2246.01 | Presentation of Statistics | Mathematics | Intro to Stat | This course will : | No | No |
| Bennington College | CS2132.01 | The Data Science Foundations | Computer Scien | Intro to DS | The Data Scienc | No | No |
| Birmingham-Souther | CAC 350 | Data Science | Computer Scien | Intro to DS | Applies concept | MA 209 or instru | No |
| Birmingham-Souther | MA 207 | QA General Statistics | Mathematics | Intro to Stat | An investigation | No | No |
| Birmingham-Souther | MA 209 | Data Science Statistics | Mathematics | Intro to DS | Introduces stude | CAC 181 or instru | No |
| Bridgewater College | MATH 140 | Introduction to Statistics | Mathematics | Intro to Stat | Basic descriptive | MATH-110 or MA | No |
| Bryant & Stratton Coll | MATH 309 | Statistics | Mathematics | Intro to Stat | This course is de | MATH 201 | No |

What is an Introductory Course?

Statistics

- Statistics, Elementary Statistics, Probability and Statistics, Statistical Methods, etc.
- Excludes courses with other statistics prerequisites
 - Prerequisites allowed and collected include Calc I and any Algebra courses.

Data Science

- Data Analytics, Data Science, Machine Learning, Introduction to Data, Data Visualization, etc.
- Excludes courses such as Intro Comp. Science and Data Structures.
- Keywords
 - Data modeling, big data, algorithmic problem solving, data wrangling, data driven decision making.

Analysis Methods

Nationwide estimates

- Estimating proportions and confidence intervals from a stratified random sample.

$$\hat{p}_{st} = \frac{1}{N} \sum_{i=1}^L N_i \hat{p}_i \quad \hat{V}(\hat{p}_{st}) = \frac{1}{N^2} \sum_{i=1}^L N_i^2 \cdot \frac{\hat{p}_i(1 - \hat{p}_i)}{n_i - 1} \left(1 - \frac{n_i}{N_i}\right)$$

- *Survey* package in R

Stratum Estimates

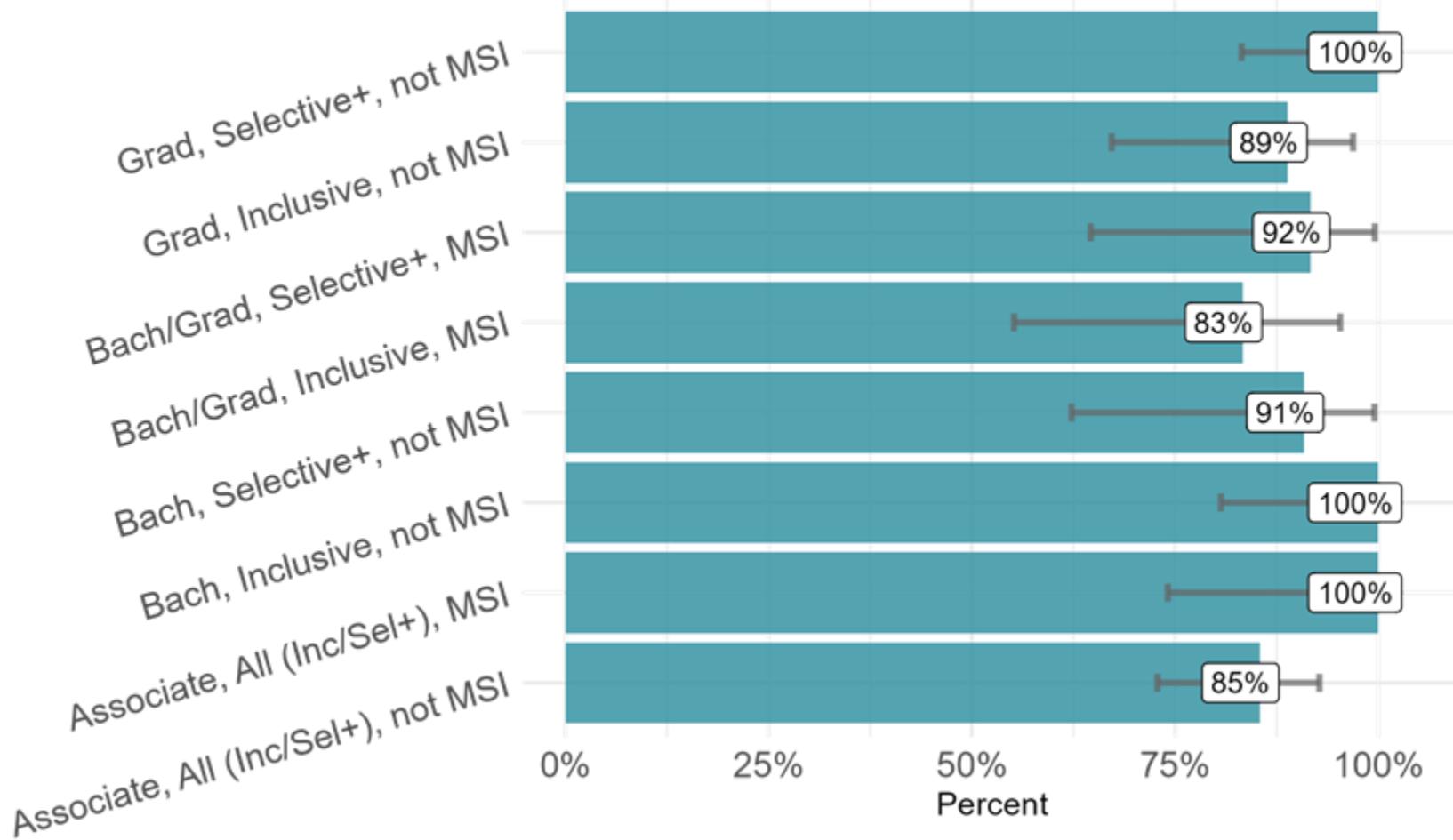
- Proportions and binomial confidence intervals were calculated for each stratum.
 - *Hmisc* package in R

Prevalence of Introductory Courses

92% (86%, 95%)

The estimated percent of colleges and universities in the United States offering an **introductory statistics course**.

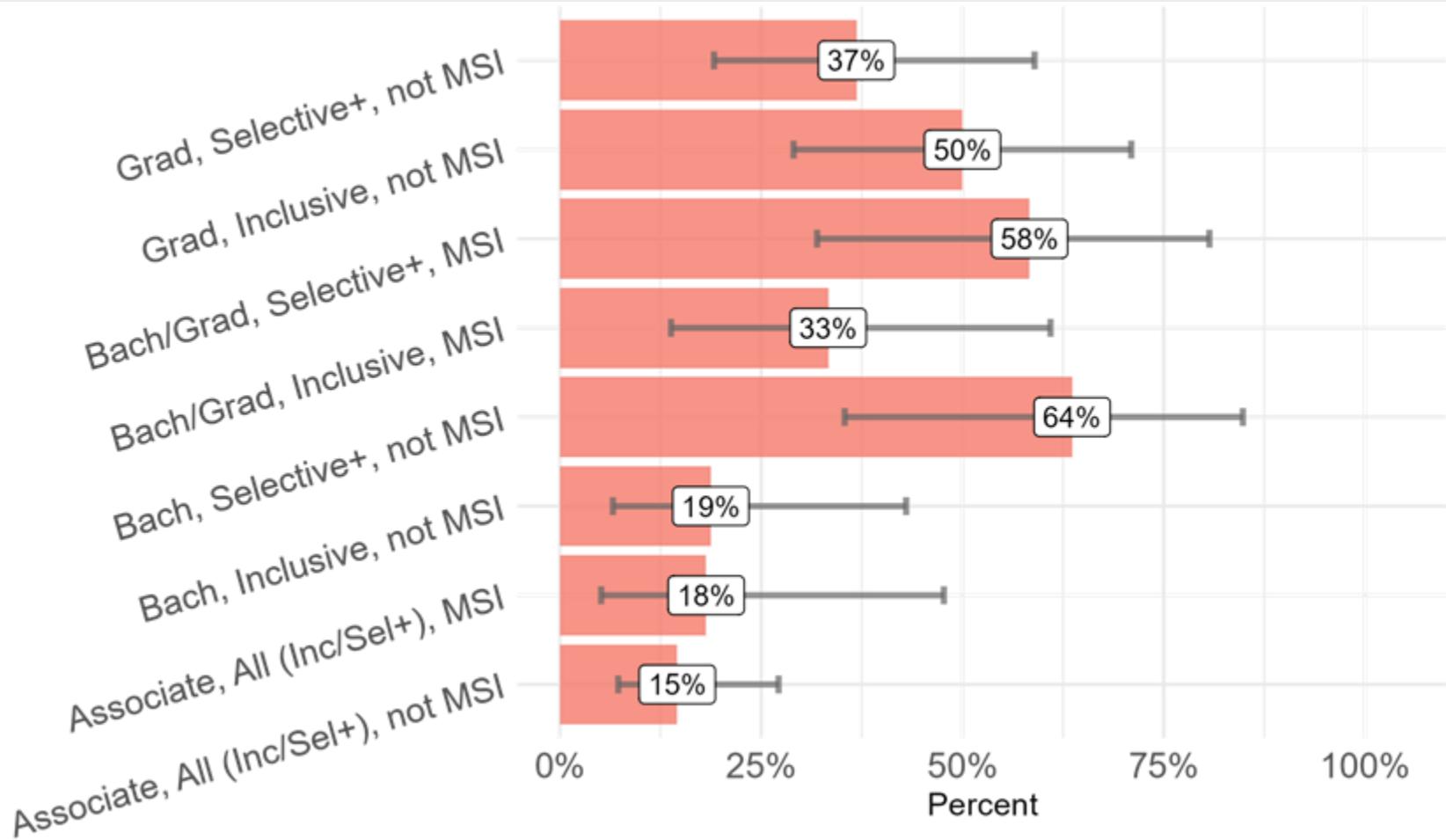
Percent of Schools Offering Introductory Statistics by Stratum



31% (24%, 38%)

The estimated percent of colleges and universities in the United States offering an **introductory data science course**.

Percentage of Schools Offering Intro Data Science by Stratum

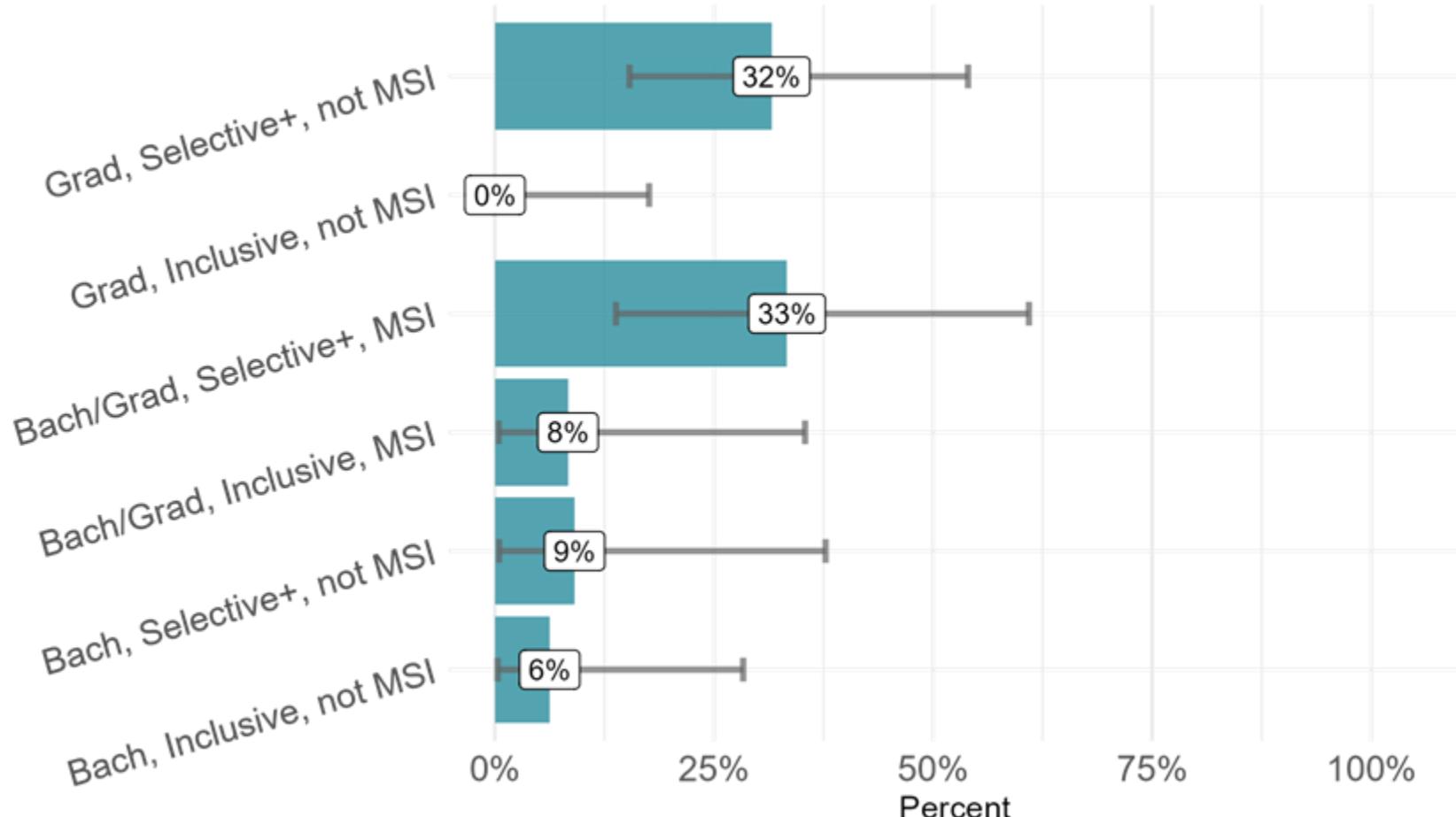


Prevalence of Majors

13% (8%, 22%)

The estimated proportion of primarily bachelors and graduate degree granting colleges and universities in the United States offering a **statistics major**.

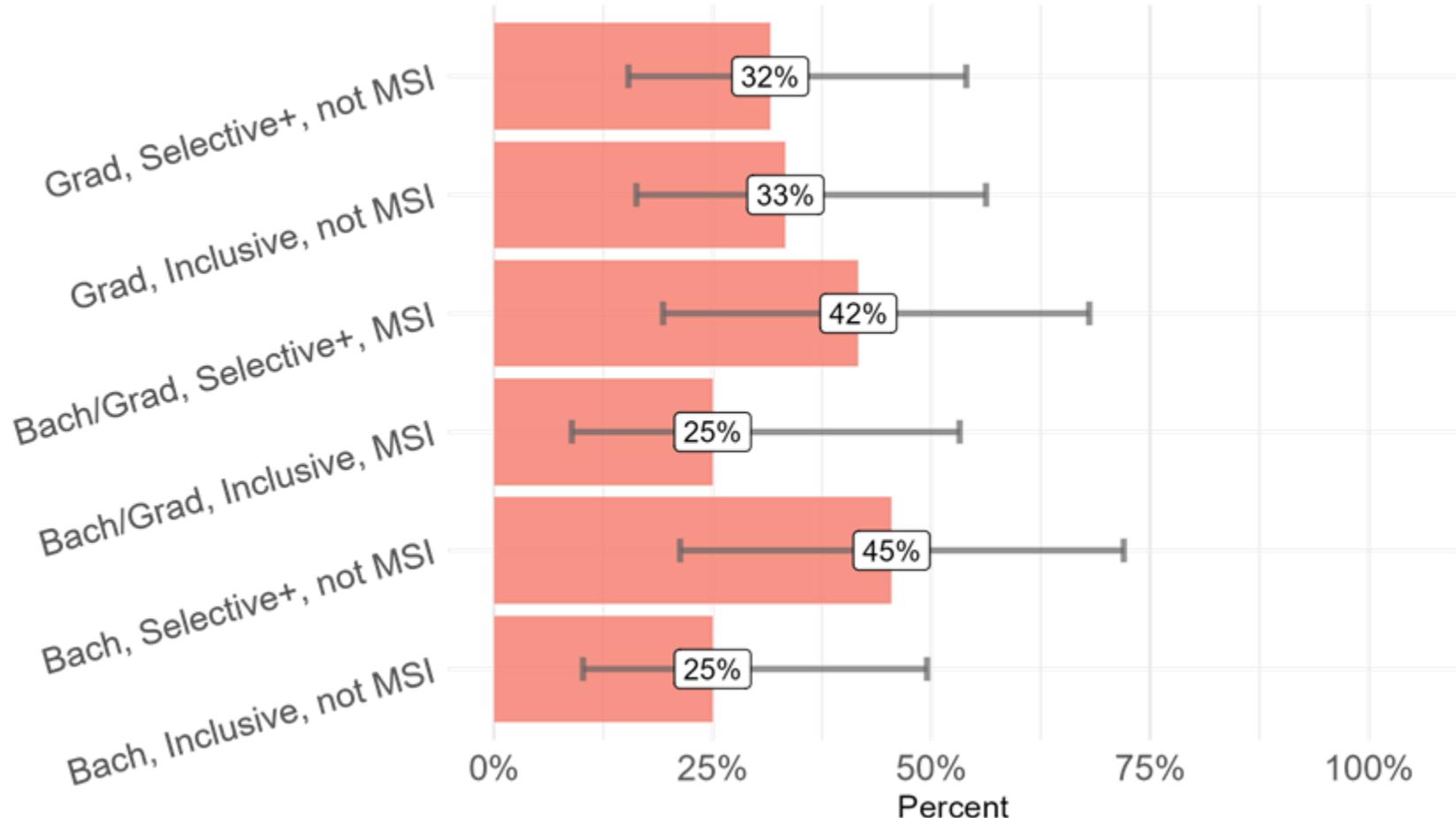
Percentage of Schools Offering a Statistics Major by Stratum



32% (23%, 43%)

The estimated proportion of primarily bachelors and graduate degree granting colleges and universities in the United States offering a **data science major**.

Percentage of Schools Offering a Data Science Major by Stratum

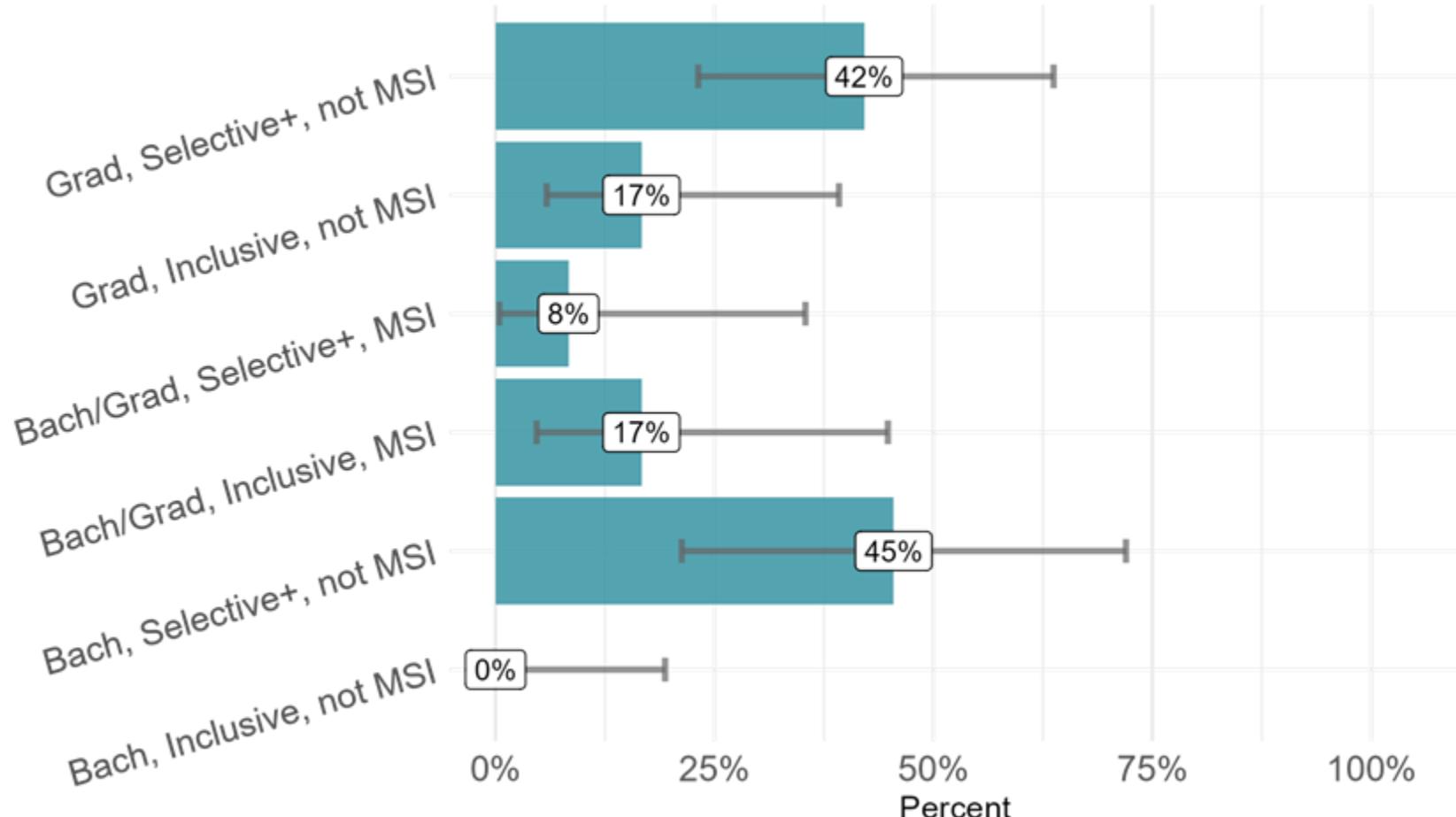


Prevalence of Minors

24% (16%, 34%)

The estimated percent of primarily bachelors and graduate degree granting colleges and universities in the United States offering a **statistics minor**.

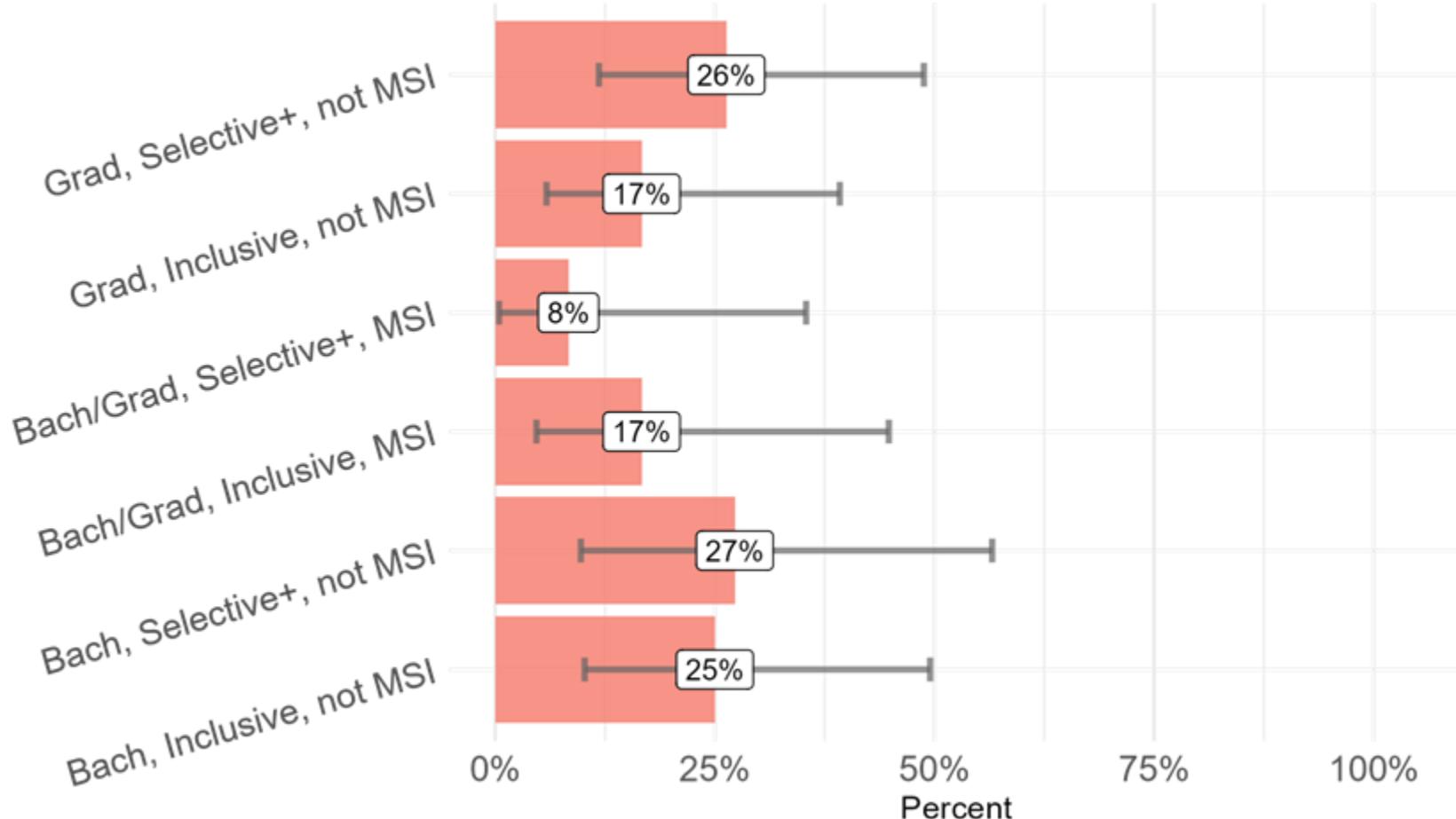
Percentage of Schools Offering a Statistics Minor by Stratum



22% (14%, 33%)

The estimated percent of primarily bachelors and graduate degree granting colleges and universities in the United States offering a **data science minor**.

Percentage of Schools Offering a Data Science Minor by Stratum



With 92% of schools offering Intro Stat
and 31% offering Intro **Data Science**,
what percentage of these courses have
prerequisites in **Math, Stat, and CS?**

Course Prerequisites

Percent of Introductory Stat and DS Courses with Prerequisites in each Subject

| Prereq | Discipline | |
|------------|-----------------------|-----------------------|
| | Statistics | Data Science |
| Math | 56% (49%, 63%) | 17% (8%, 33%) |
| Statistics | | 14% (6%, 30%) |
| Comp Sci | | 16% (7%, 32%) |
| All Three | | 9% (3%, 24%) |
| None | 44% (35%, 50%) | 44% (28%, 61%) |

Summary

| | Statistics | Data Science | Odds Ratio (Stat/DS) |
|---------------------|------------|--------------|-------------------------|
| Introductory Course | 92% | 31% | 2.97 |
| Major | 13% | 32% | 0.41 |
| Minor | 24% | 22% | 1.09 |
| Associates Degree | 0% | 5% | NA |
| Certificate | 1% | 6% | 0.16 |

What's Next?

- The database created will be made publicly accessible for other educators and researchers to utilize.
- Point estimates for the population proportion of these courses and majors set a foundation for future researchers to study changes over time.
- Other variables of interest in the future may include course unit counts, support courses offered, lab components, etc.
- The next steps of the MASDER project focus on studying student attitudes and academic performance in relation to institutional characteristics.

References

- **Dept. of Education College Scorecard Database**
 - <https://collegescorecard.ed.gov/data/>
- **Carnegie Classification of Institutions of Higher Education**
 - <https://carnegieclassifications.acenet.edu/>

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. DUE-2013392. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

- The Undergraduate Research Opportunities Center @ CSUMB
- Apple Scholars Research Program



Thank You!

Contact us:

Amira Colón

acolon@csumb.edu

SDSattitudes.com

