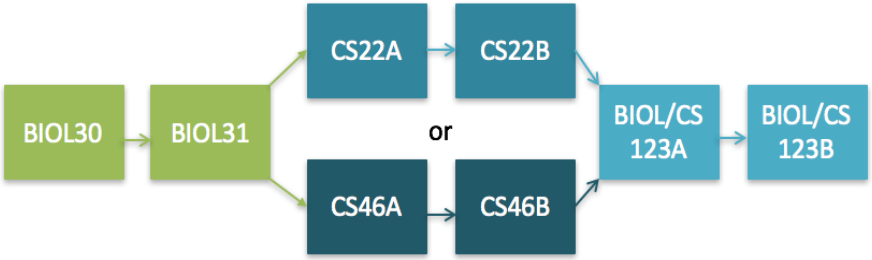
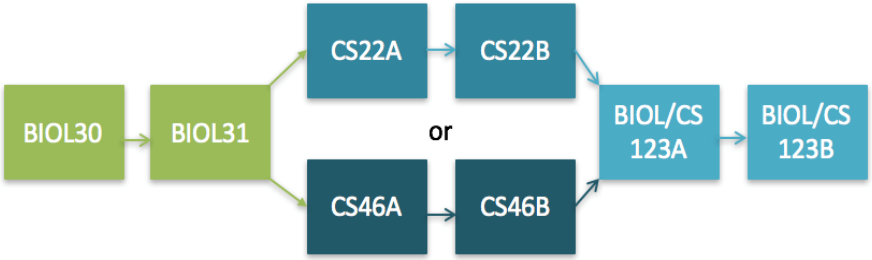




An Interdisciplinary Minor in Bioinformatics

Sami Khuri
San José State University
Department of Computer Science

©2018 Sami Khuri

Outline of Presentation

- CS Employment Projections
- Can we count on K-12?
  - CS in K-12 in USA
- Can we count on Colleges/Universities?
  - CS Degree Production
- Gender Equity in CS
- Current Trends in CS
- Solution: Minor in Bioinformatics at SJSU
- Description of the Courses
- Conclusion

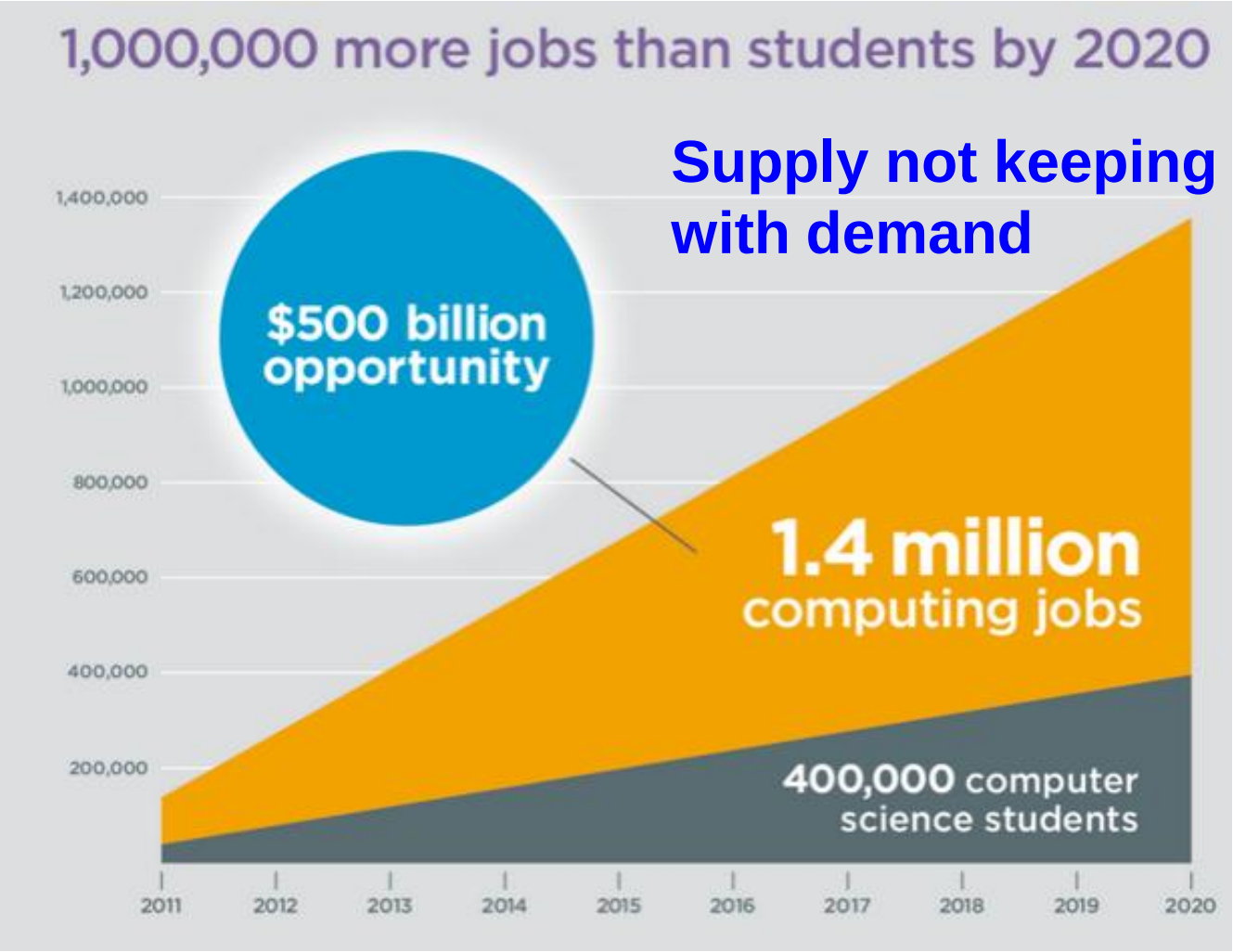
©2018 Sami Khuri

Programming Skills in High Demand

- **This deficit is pronounced in the fast growing biotechnology industry.**
  - High-throughput technologies generate data in unprecedented amounts, creating demand for training of biologists in data management, manipulation, mining and analysis.
  - Training is needed:
    - to use and gain familiarity with bioinformatics, data science tools and resources
    - to automate common tasks and understand fundamental theoretical and practical computer science concepts.
- **Other fields**
  - Cybersecurity
  - Data Science

©2018 Sami Khuri

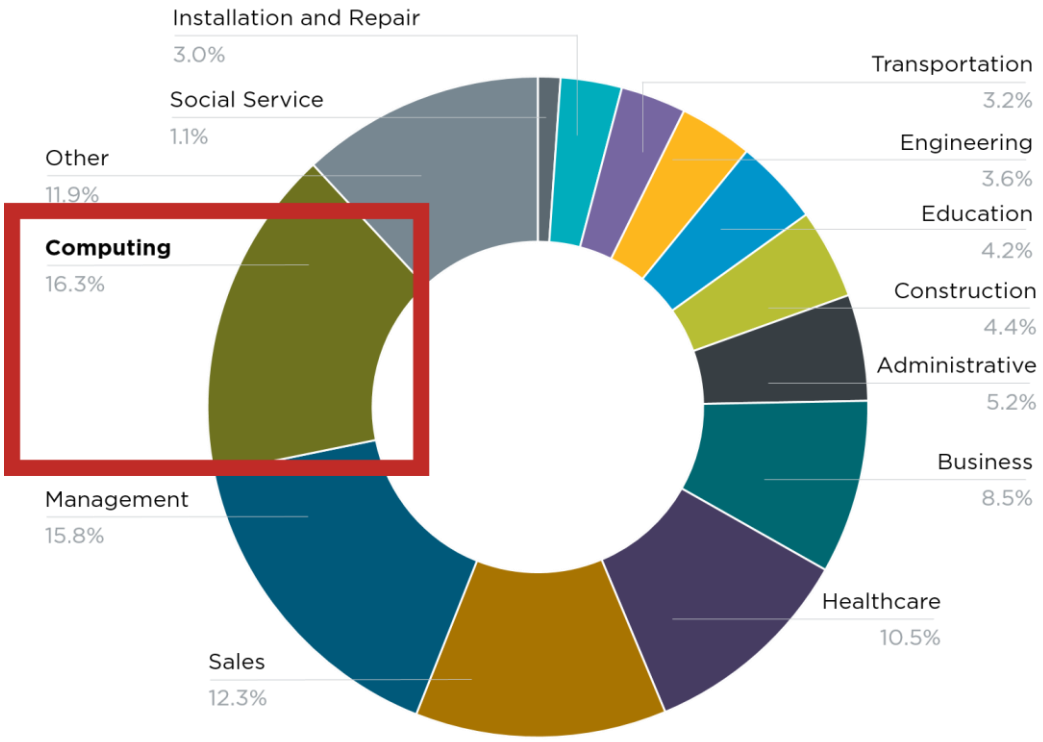
Jobs Growing Faster than National Average Production



http:/code.org/stats

©2018 Sami Khuri

Computing Jobs: #1 Source of New Wages in the US

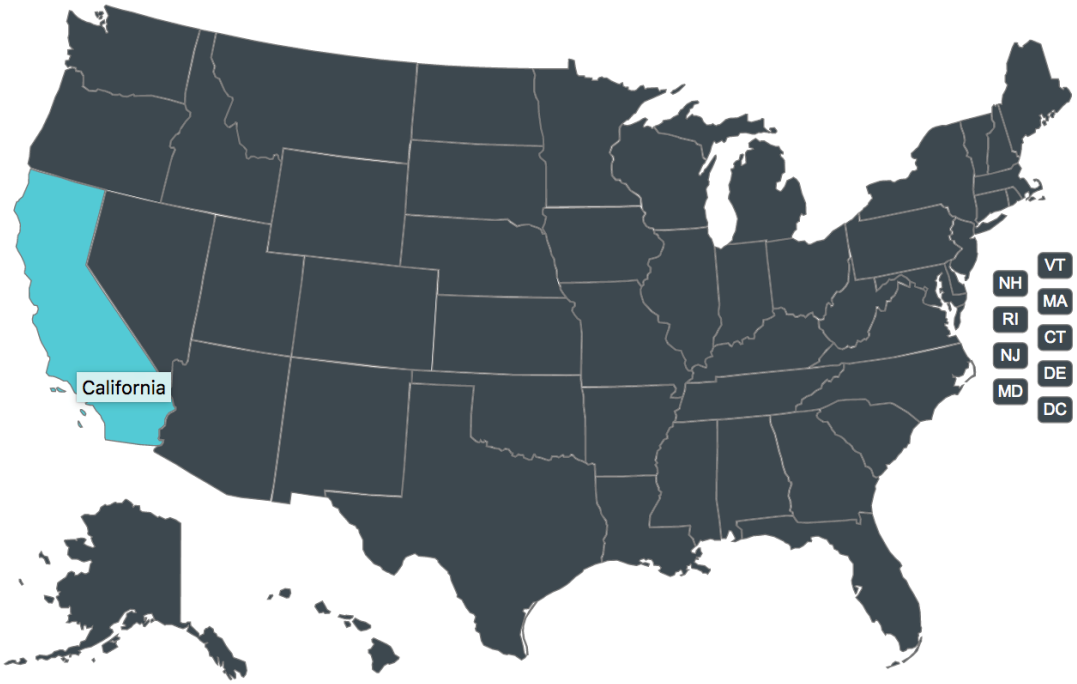


500,000 current openings:
These jobs are in **every** industry and **every** state, and they're projected to grow at twice the rate of all other jobs.

©2018 Sami Khuri

Can we Count on K-12?

Choose a State



Computer science drives innovation throughout the US economy, but it remains marginalized throughout K-12 education.

13

states have adopted a policy to give all high school students access to CS courses (and of those, only 5 states give all K-12 students access)

546,480

Open computing jobs nationwide

49,291

Computer science students graduated into the workforce last year

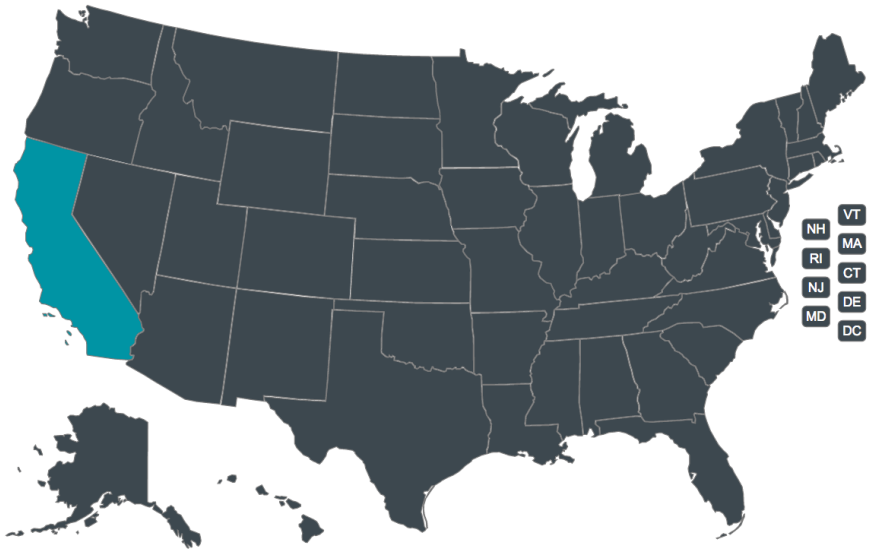
See a summary of state efforts

Sources: **The Conference Board** and **National Center for Education Statistics**. **More info**.

https://code.org/promote

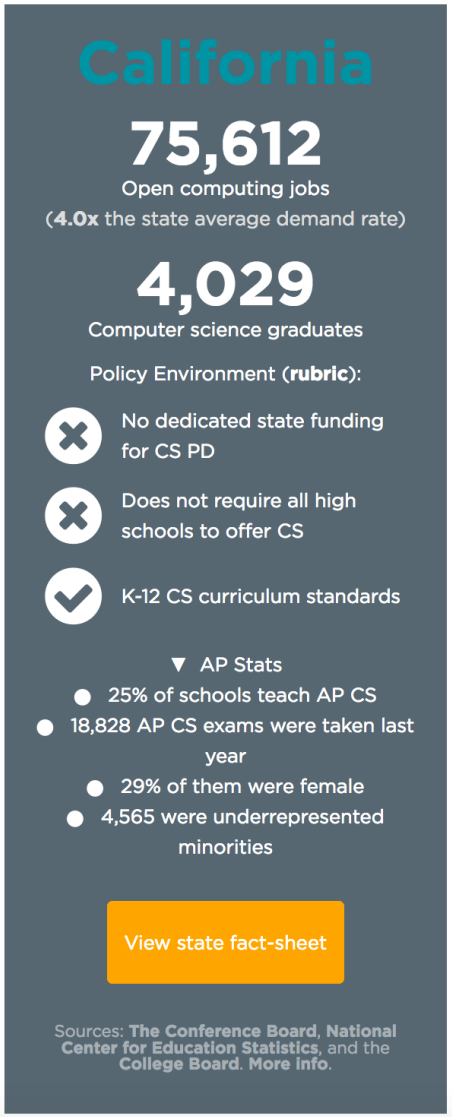
©2018 Sami Khuri

What about California?



AP Statistics:

25% of schools teach AP CS
18,828 AP CS exams were taken last year
29% of them were females
4,565 were underrepresented minorities



https://code.org/promote/ca

©2018 Sami Khuri

Computer Science in High Schools: Lack of Access

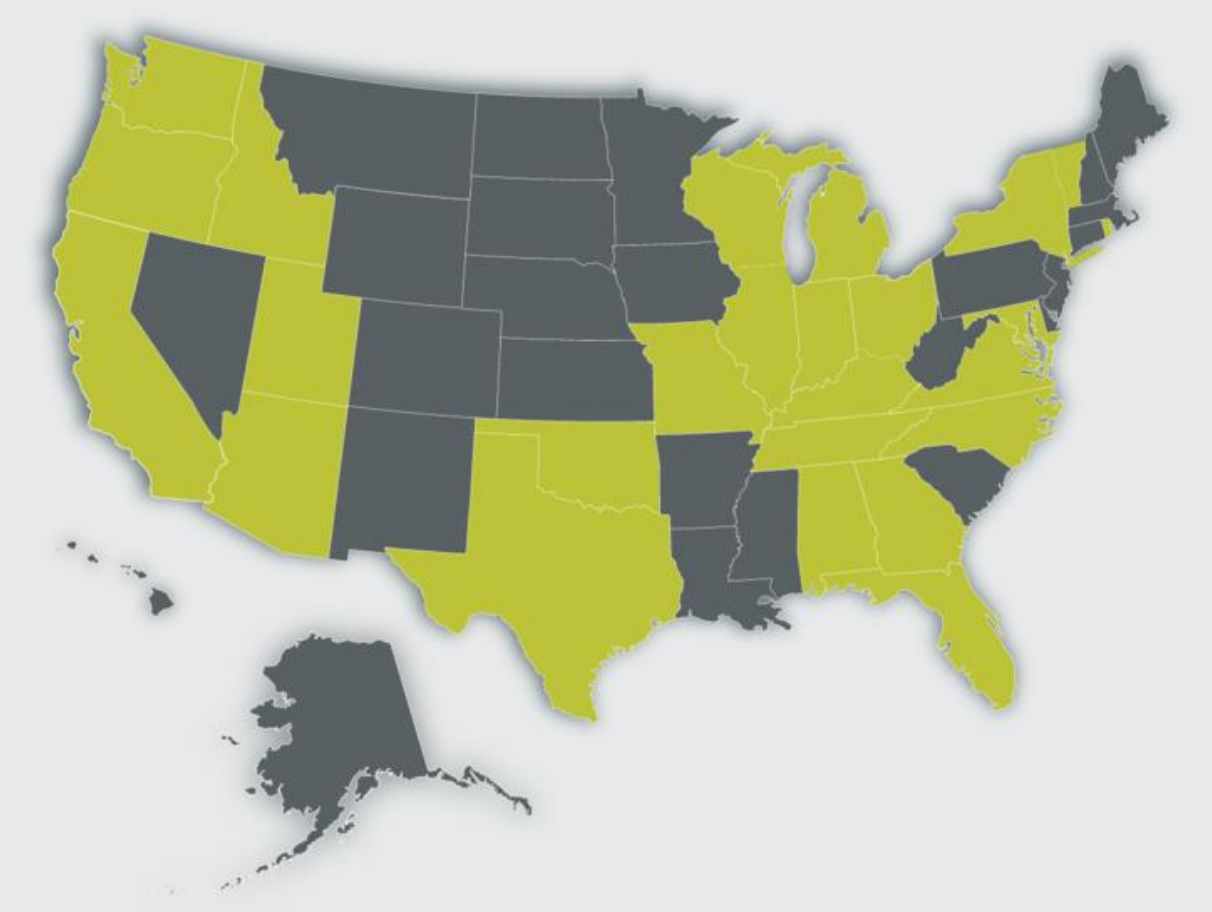
15 out of 18 American high schools don't offer computer science or programming classes.

The majority of schools **don't even offer computer programming classes.**

http:/code.org/stats

©2018 Sami Khuri

HS Graduation Requirements



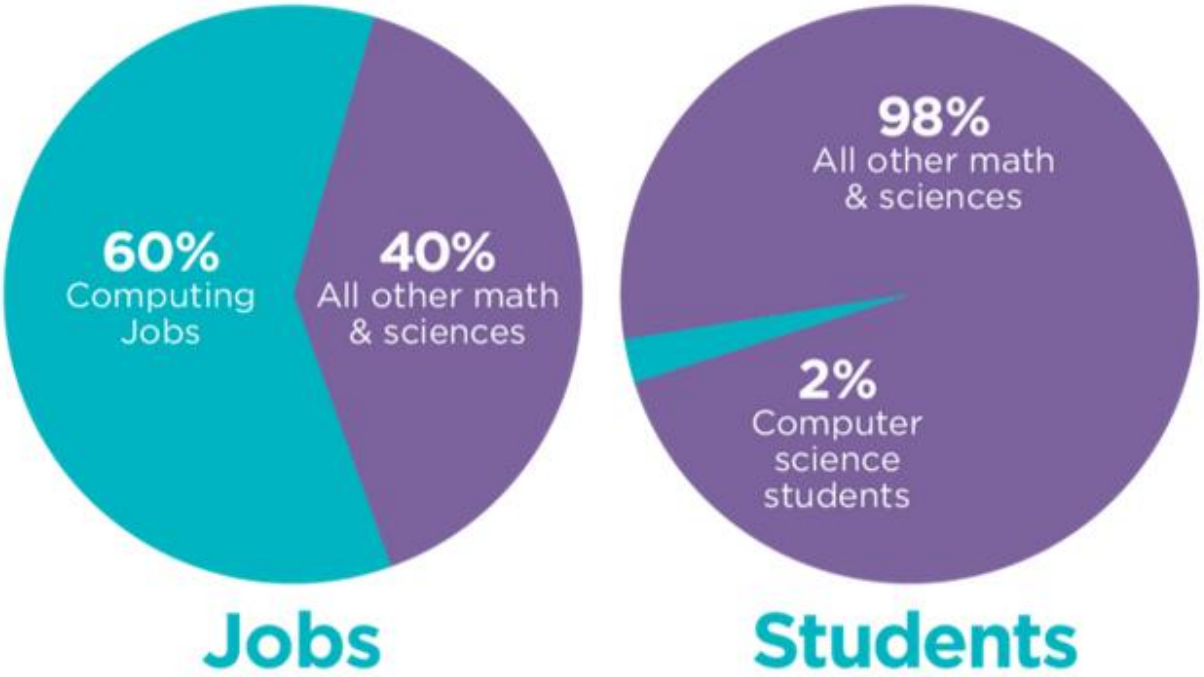
25 states don't count CS toward high school math or science graduation requirements.

Recent change: CA now allows districts to count CS as satisfying a 3rd math requirement.

http:/code.org/stats

©2018 Sami Khuri

Can we Count on Universities?
STEM Jobs: Student Gap in CS



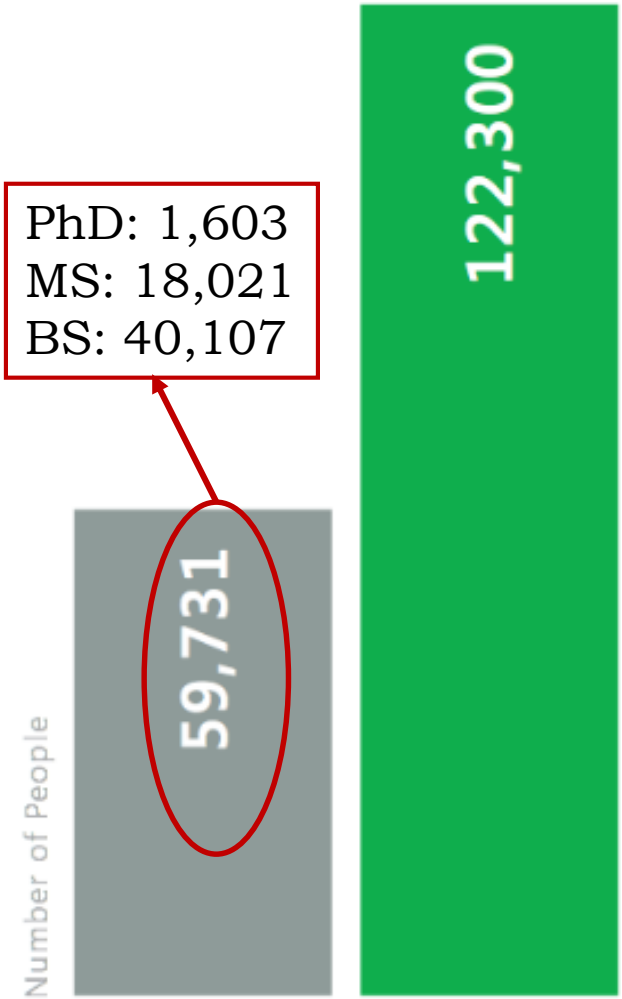
Less than 2.4% of college students graduate with a degree in computer science.

http:/code.org/stats

©2018 Sami Khuri

Graduation Rates not keeping up with Job Creation

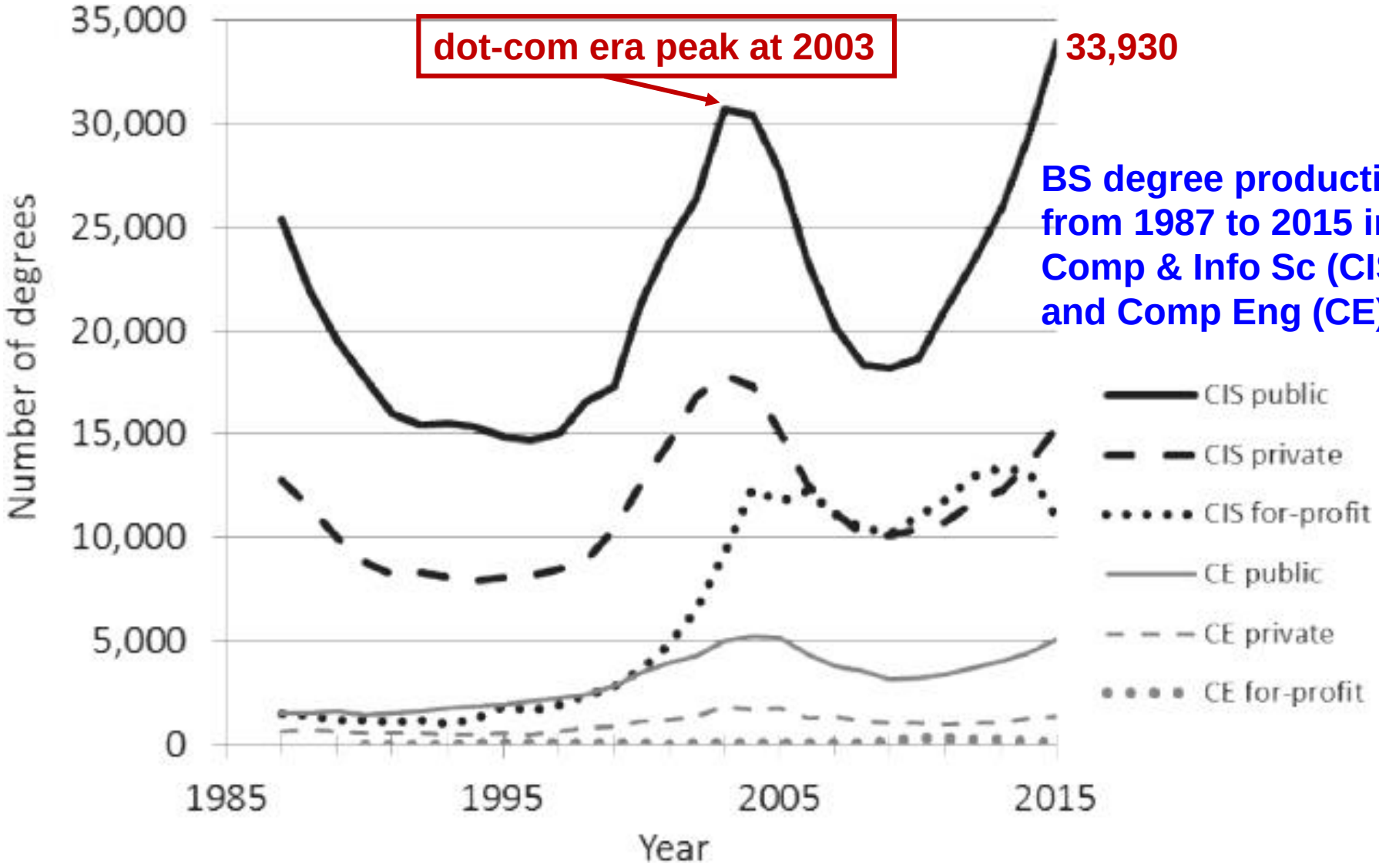
- Number of graduates in **Computer Science** versus projected job openings in computing requiring at least a bachelor's degree.



Based on U.S. Bureau of Labor Statistics forecast of 1.22 million for 2010 -2020

©2018 Sami Khuri

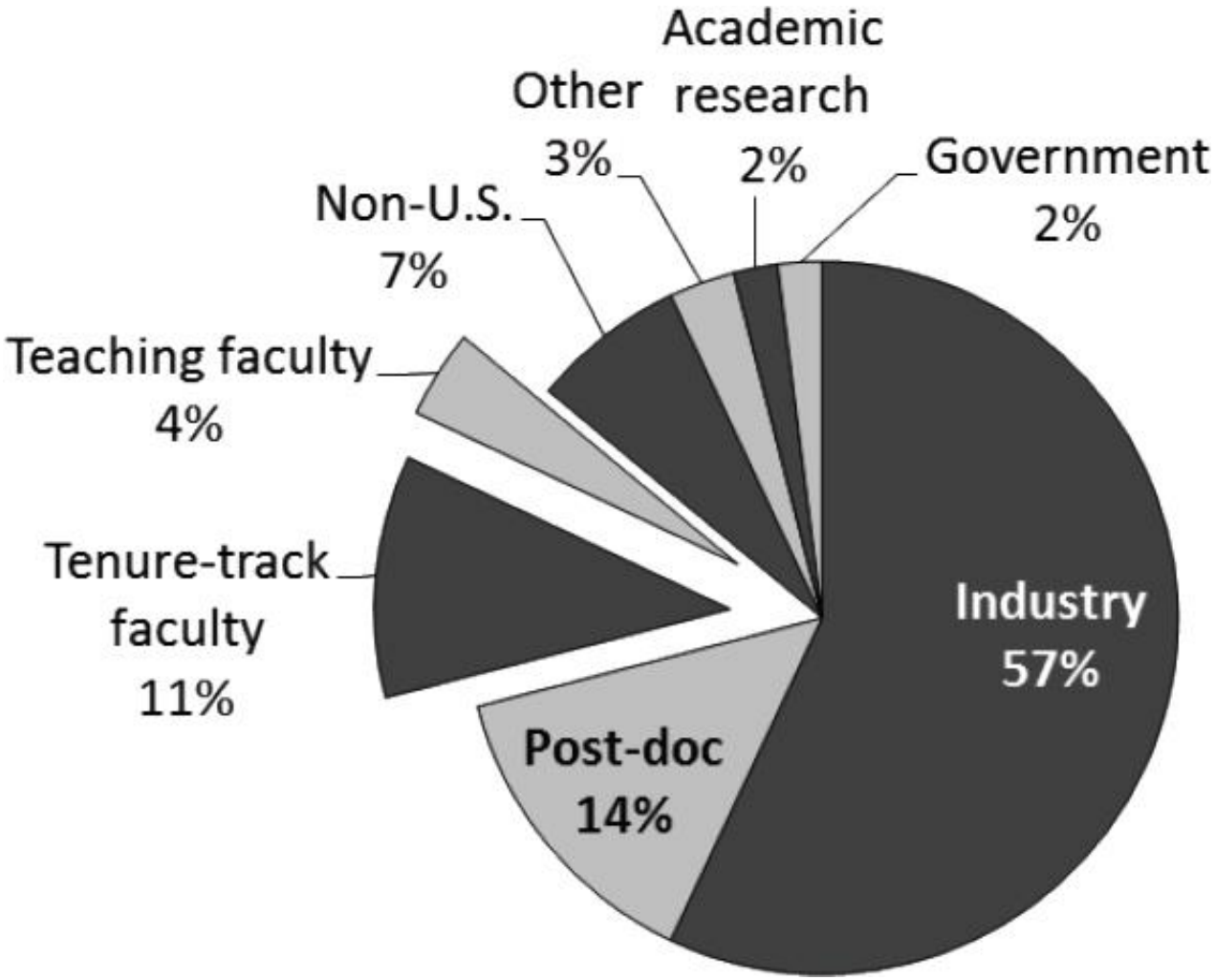
Role of Public Institutions



https://www.nap.edu/read/24926/chapter/4#29

©2018 Sami Khuri

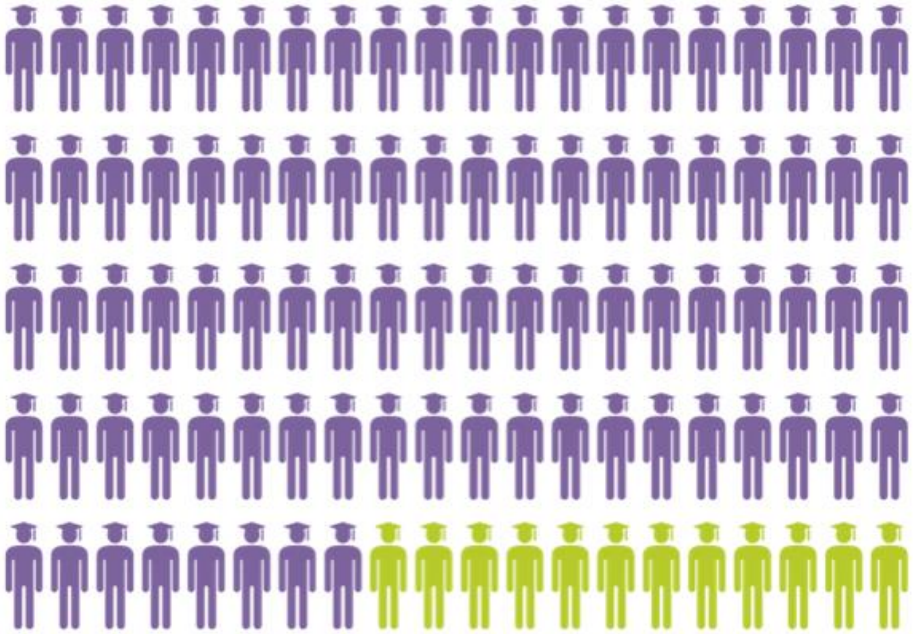
Where have the 2016 CS PhD Graduates gone?



https://www.nap.edu/read/24926/chapter/5#62

©2018 Sami Khuri

Gender Equity
BS in CS Awarded to Women



While 57% of bachelor's degrees are earned by women, **just 12% of computer science degrees** are awarded to women.

http:/code.org/stats

©2018 Sami Khuri

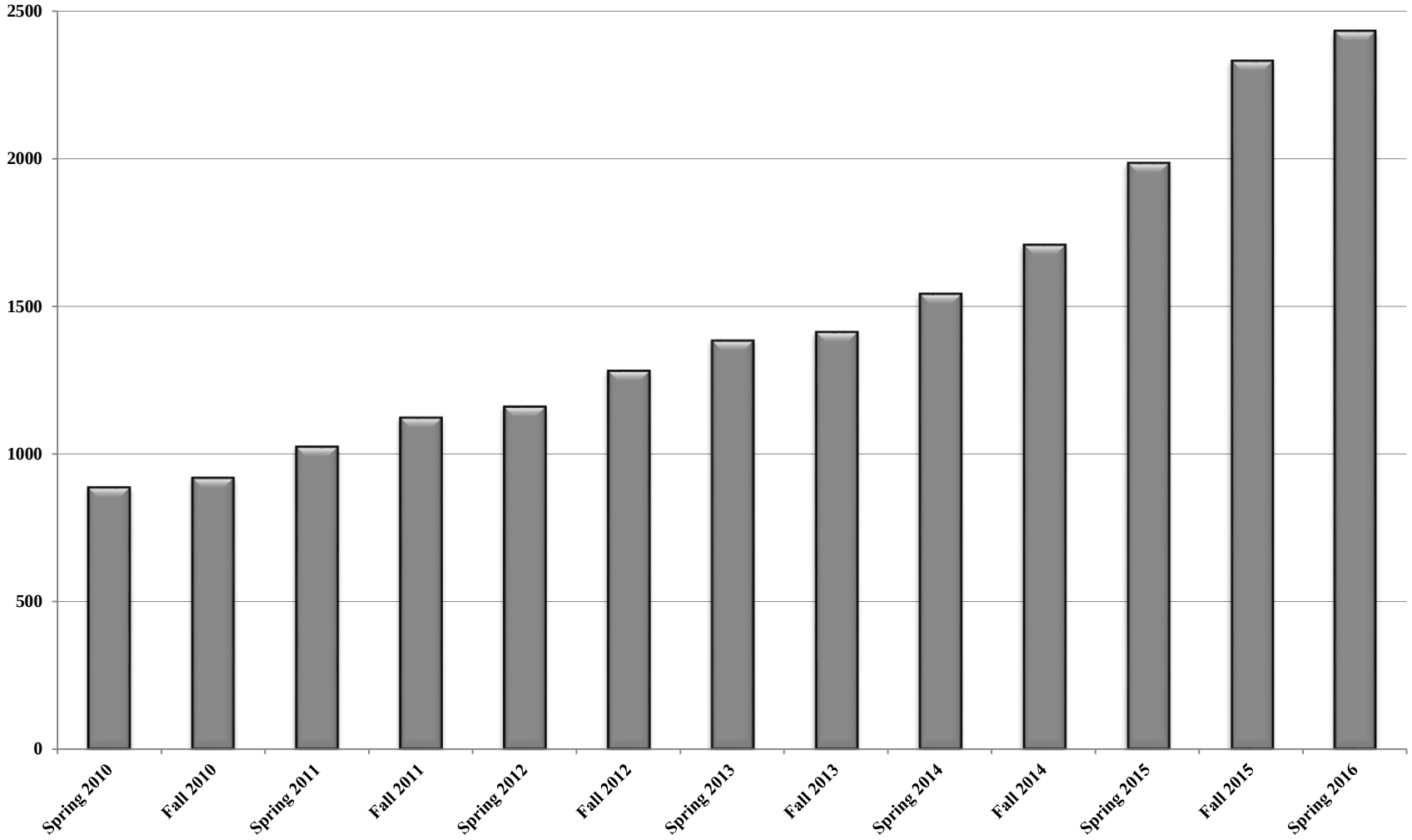
Importance of Equity in CS

- People with more diverse sources of information generate **consistently better ideas**.

  [Ron Burt, U of Chicago]
- Diversity is a **key driver of innovation** and is a critical component of being successful on a global scale. [Forbes]
- Diversity fosters, enhances, and leads to **creativity**.
- We need as **many viewpoints** as possible when it comes to solving complex problems.
- To meet the **demands of the workforce** there is a critical need to tap into and prepare underrepresented groups.

©2018 Sami Khuri

2010-2016: SJSU Student Enrollments in CS courses



©2018 Sami Khuri

CS1 and CS2 Students

- Many of the non-CS majors enroll in CS1 and CS2 seeking to transfer to the CS major or to minor in Computer Science.
- Anecdotal evidence indicates that many other non-CS majors are taking these courses to enhance their major degrees with programming and computational skills.

©2018 Sami Khuri

Enrollments in Spring 2016 at SJSU

Degree Major	CS 1 Introduction to Computer Science	CS 2 Introduction to Data Structures
Total enrollment	213	162
Computer Science	47	28
Software Engineering	47	29
Undeclared	68	56
Biology	2	1
Chemistry/Biochemistry	2	1
Other	47	47
Failure rate	26%	10%

©2018 Sami Khuri

Enrollments in Spring 2018 at SJSU

Degree Major	CS 1 Introduction to Computer Science	CS 2 Introduction to Data Structures
Total enrollment	182	179
Computer Science	10	62
Software Engineering	40	37
Undeclared	47	19
Biology	8	2
Chemistry/Biochemistry	2	0
Other	75	59
Failure rate	24%	10%

©2018 Sami Khuri

Non-Majors in CS Course

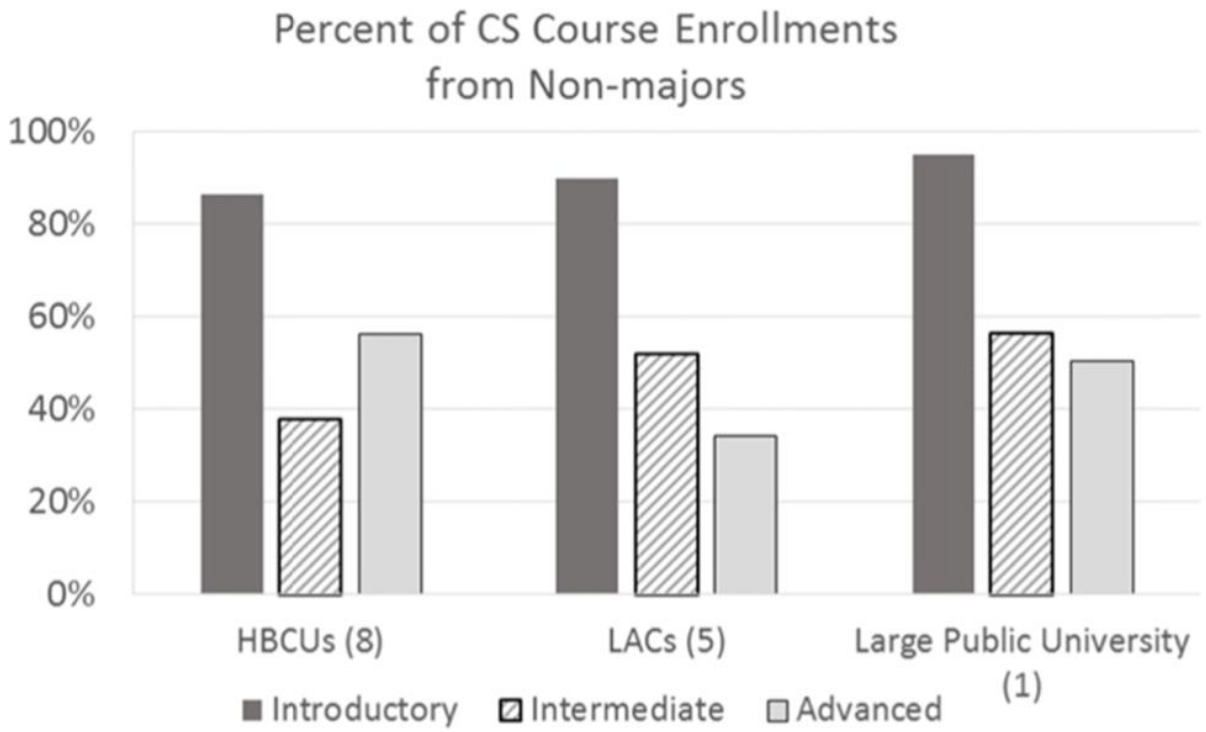
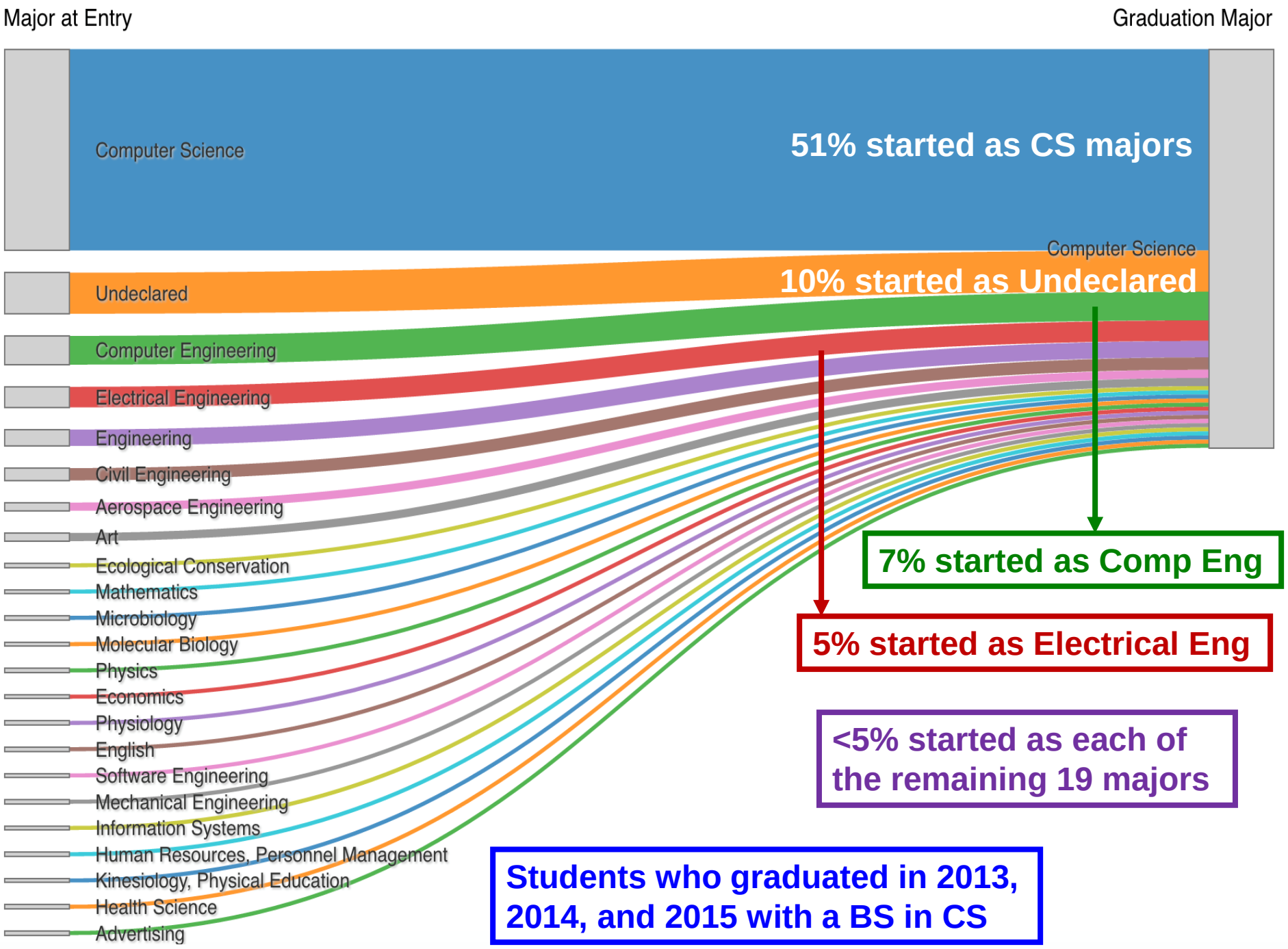


FIGURE 3.10 Percent of non-majors (STEM and non-STEM) in three levels of courses at small samples of different institution types, taken from aggregate (by institution type, averaged between 2009 and 2014) data. HBCUs are historically black colleges and universities; LACs are liberal arts colleges. SOURCE: Data from CUSTEMS.

https://www.nap.edu/read/24926/

©2018 Sami Khuri



www.csudata.calstate.edu

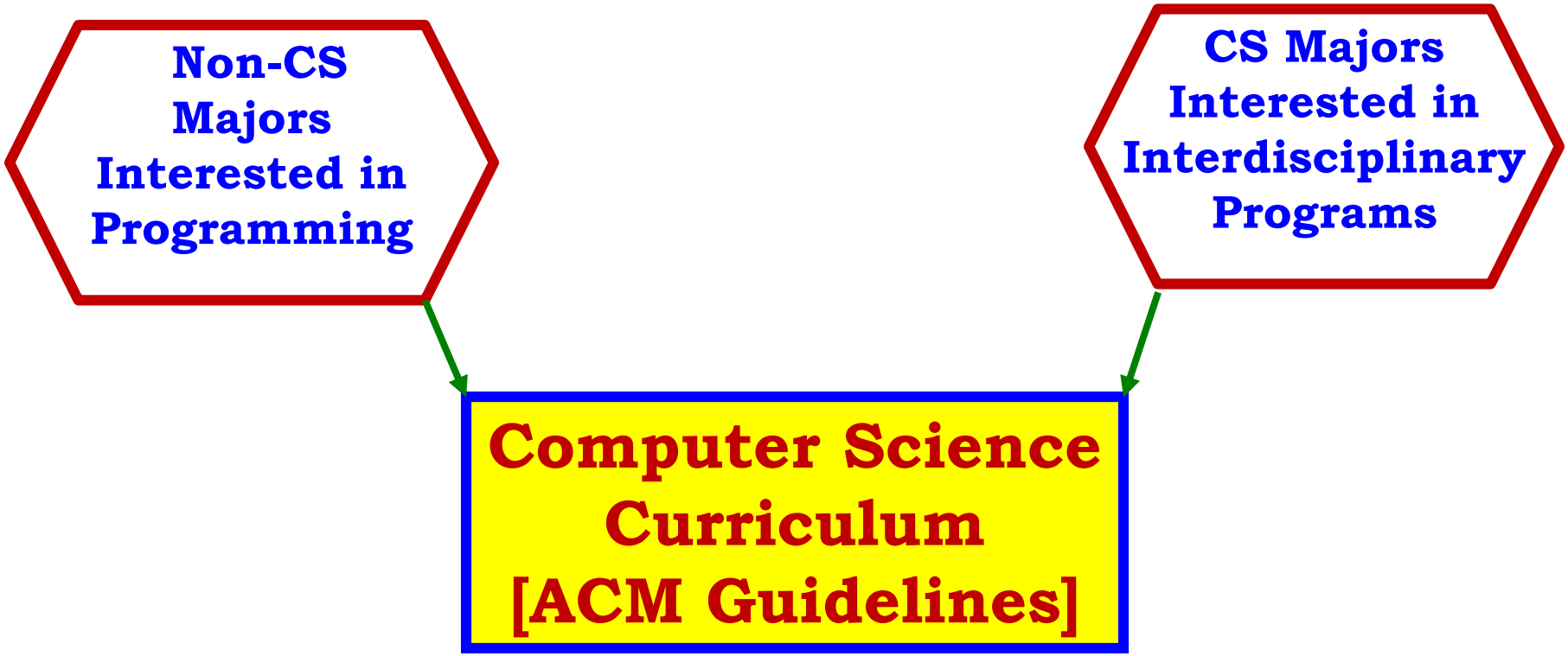
©2018 Sami Khuri

Curricular Challenges and New Pathways to STEM

- Recruitment and retention of STEM majors remains a high priority for academic institutions and requires:
  - **innovative curricular changes** and
  - **new pathways** to STEM disciplines.

©2018 Sami Khuri

Current Trends in Computer Science



Current curriculum is not flexible enough to accommodate both trends

Solution: Create Interdisciplinary Programs (Minors/Certificates) that require at most one additional semester to complete

©2018 Sami Khuri

Minor in Bioinformatics to the Rescue

- Interdisciplinary, cohort-based, and programming-intensive Minor in Bioinformatics (MB).
- The Minor in Bioinformatics is:
  - a step toward meeting the increased need for skilled workers in biotechnology fields
  - a means of training biologists who can program, analyze complex bioinformatics problems and identify computational solutions to these problems.

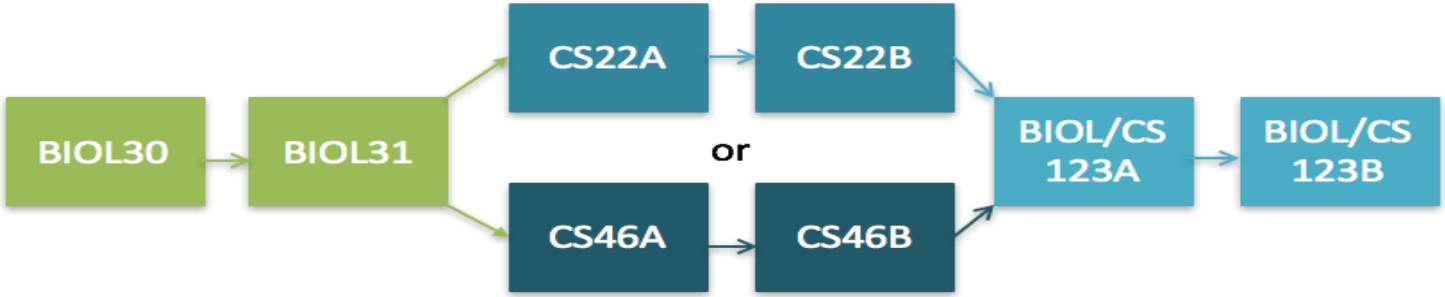
©2018 Sami Khuri

Attracting More Women in Computer Programming

The Minor in Bioinformatics at SJSU will attract more women to computer programming and bioinformatics, since the program will be most attractive to biology undergraduate students, approximately 60% of whom are women.

©2018 Sami Khuri

The Six Courses of the Bioinformatics Minor



Fall 2017: 17 Females + 14 Males = 31

Biology	Chemistry/ Biochem	Engineering	Business	Other	Undeclared
16 11 + 5	4 2 + 2	Biomed: 2 Mechanic: 1	5 2 + 3	Music: 1 Cont. Ed: 1	1

Spring 2018: 11 Females + 6 Males = 17

Biology	Chemistry/ Biochem	Engineering	Business	Other	Undeclared
11 8 + 3	1	Biomed: 1 Mechanic: 1	0	Music: 1 Cont. Ed: 2	0

©2018 Sami Khuri

Coursework in the Bioinformatics Minor

Biology Coursework	Computer Science Coursework	Bioinformatics Coursework
a) Principles of Biology I (4 units)	c) Python Programming for Non-Majors I (3 units)	e) Bioinformatics I (3 units)
b) Principles of Biology II (4 units)	d) Python Programming for Non-Majors II (3 units)	f) Bioinformatics II (3 units)
Note: All biology majors take a) and b) as required courses in their major	**Note:** CS majors will take Introduction to Computer Science and Introduction to Data Structures instead	**Note:** c) and d) will be taken before e), which is a prerequisite for f).

©2018 Sami Khuri

Developing the Minor in Bioinformatics

We developed the Minor in Bioinformatics by:

- Identifying unmet needs of program constituencies
- Defining primary target audience
- Designing the curriculum for the intended target audience
- Laying out administrative and assessment processes

©2018 Sami Khuri

The Unmet Needs of Program Constituencies

Students: program addresses need to recruit and retain life science students in STEM disciplines.

Faculty: current bioinformatics courses, *Bioinformatics I* and *II*, are not sufficient to teach biology students advanced computational skills to address biological questions.

Graduate School: graduate research in life sciences is becoming increasingly interdisciplinary, and biology students with quantitative and computational preparation are highly sought after.

Employers: Data-driven science is becoming commonplace and bioinformatics curricula should adapt to meet these new needs. Minor in Bioinformatics prepares biology graduates to program, develop workflows and pipelines, run programs in the Cloud, perform common database tasks and build simple web servers.

©2018 Sami Khuri

Student Learning Outcomes of Minor in Bioinformatics

- Be proficient in the use of Python (or Java for CS majors) to implement computational solutions to life science problems.
- Demonstrate knowledge of
  - Bioinformatics algorithms and statistical methods used in the analysis of life science problems
  - Bioinformatics database construction and interrogation
  - The role of high-throughput technologies and data in life science research.
- Integrate these skills in the analysis of complex life science problems and the design of computational solutions.

©2018 Sami Khuri

Conclusions

- US economy will generate 1.4 million CS-related jobs by 2020, but graduate only 400,000 computer science graduates.
- One way to reduce the deficit: attract women to computational field.
  - Women earn almost 60% of bachelor's degrees in the US, but <20% in CS
  - Over 60% of undergraduate biology students are females.
- Designed, developed, implemented, and will assess, a programming-intensive, cohort-based, interdisciplinary Minor in Bioinformatics.
- By rooting computational program in biological concepts & questions:
  - we plan to interest and educate biology students in computational methods that can be applied to complex questions in growing field of bioinformatics.
- MB: model for designing, developing and implementing a new minor, which could serve as an example for future minors that bring female students from other majors into Programming or even CS.

©2018 Sami Khuri

Acknowledgments

1) Some slides from presentation I gave at the 2017 SIGCSE Conference:
"Increasing the capacity of STEM workforce: Minor in Bioinformatics"
co-authors:
Miri VanHoven, Biological Sciences, SJSU
Natalia Khuri, Bioengineering, Stanford

2) Technology Pathway Initiative (TPI) grant
Center for Advancing Women in Technology

©2018 Sami Khuri

Computer Science in US Secondary Schools (I)

- Computer science courses in K-12 education are fading from the national landscape at the very moment they are needed most.
- Introductory secondary school computer science courses have decreased in numbers by 17 percent from 2005.
- The number of Advanced Placement (AP) Computer Science Courses has similarly decreased by 33 percent.

http://www.computinginthecore.org/facts-resources

©2018 Sami Khuri

Computer Science in US Secondary Schools (II)

- Despite its critical and growing importance, computer science is taught in only a small minority of U.S. schools.
- There currently are just over 42,000 high schools in the United States.
- But only 2,100 of them were certified to teach the AP Computer Science course in 2011, and in fact only 21,139 students took the AP exam.

http://news.microsoft.com/download/presskits/citizenship/msnts.pdf

©2018 Sami Khuri

2008-2015: Freshmen Biology Enrollment

Semester	Female	Male	Percent
Fall 2008	119	72	62%
Fall 2009	104	64	62%
Fall 2010	91	49	65%
Fall 2011	74	56	57%
Fall 2012	57	39	59%
Fall 2013	145	76	66%
Fall 2014	112	45	71%
Fall 2015	120	64	65%
Fall 2016	N/A	N/A	N/A

©2018 Sami Khuri

Sample Research Theme Detecting Cryptic Splice Sites

1. Describe the anatomy of a eukaryotic gene
2. Explain importance of patterns/motifs in DNA & RNA sequences
3. Name sequence patterns important to gene processing, such as 5' and 3' splice sites
4. Explain how point mutations in sites can cause genetic disease
5. Explain how available experimental data can be utilized to develop computational predictive programs
6. Write computer programs to detect 5' and 3' splice sites in DNA or RNA sequences using exact matches, regular expressions, and position weight matrices (PWMs)

©2018 Sami Khuri

Sample Research Theme Detecting Cryptic Splice Sites

7. Implement workflow/pipeline using hidden Markov Model (HMM) classifier to detect plausible splice sites in the human genome
8. Explain strength and weaknesses of PWMs and HMMs
9. Explain how to estimate the accuracy of predictions
10. Run workflows on a computer cluster.

- The research theme is recurrently addressed in four courses at increasingly more sophisticated levels of detail. (Next slide)
- Students implement increasingly more complex Python programs and work with increasingly higher volumes of data as they progress through the project.
  - Important computer science data structures and algorithms are introduced to complement their learning.

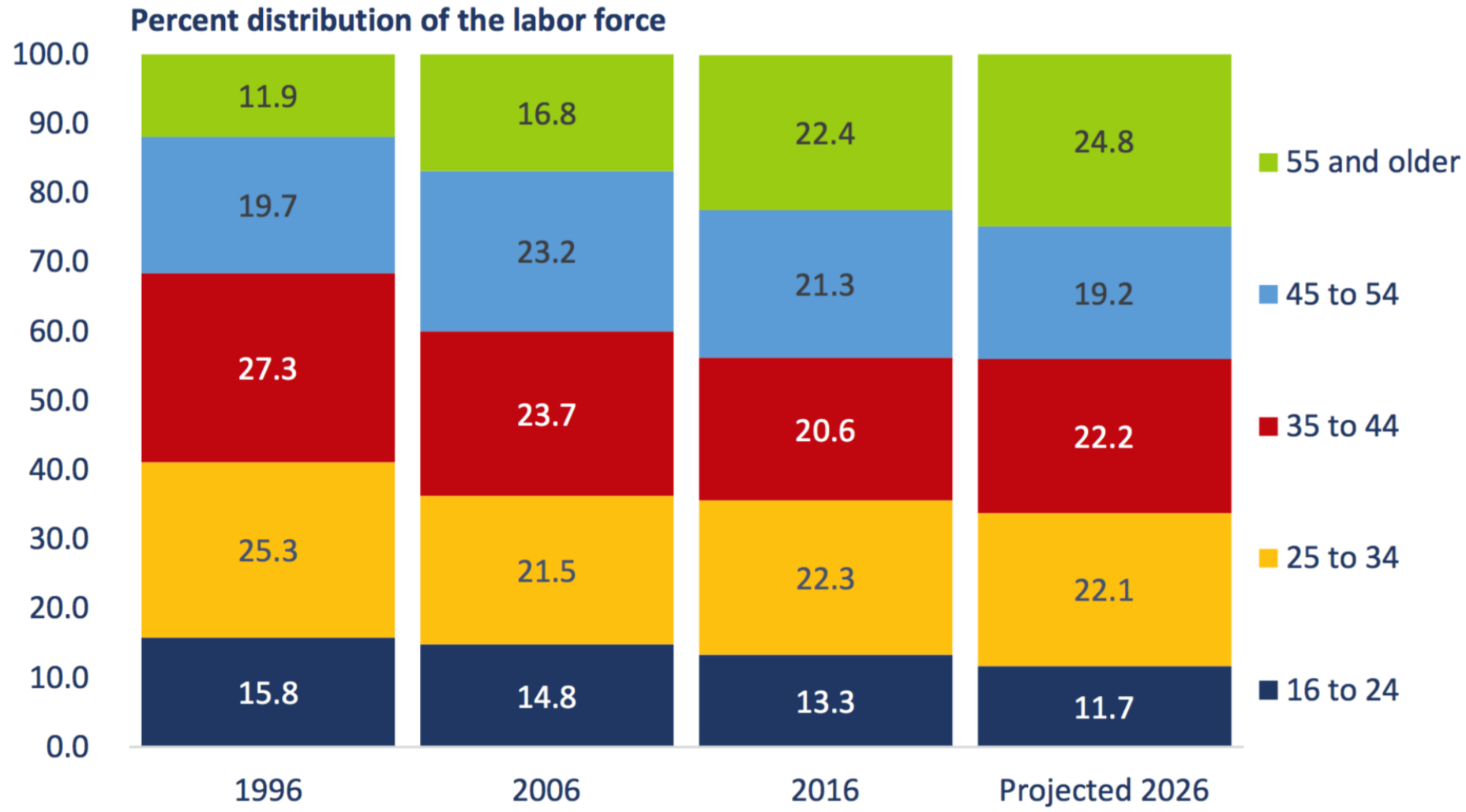
©2018 Sami Khuri

Mapping of Learning Outcomes (LO) to Courses

LO	Python I	Python II	Bioinformatics I	Bioinformatics II
(1)	✓	✓	✓	✓
(2)	✓	✓	✓	✓
(3)	✓	✓	✓	✓
(4)			✓	✓
(5)		✓	✓	✓
(6)	✓	✓	✓	
(7)				✓
(8)				✓
(9)			✓	✓
(10)				✓

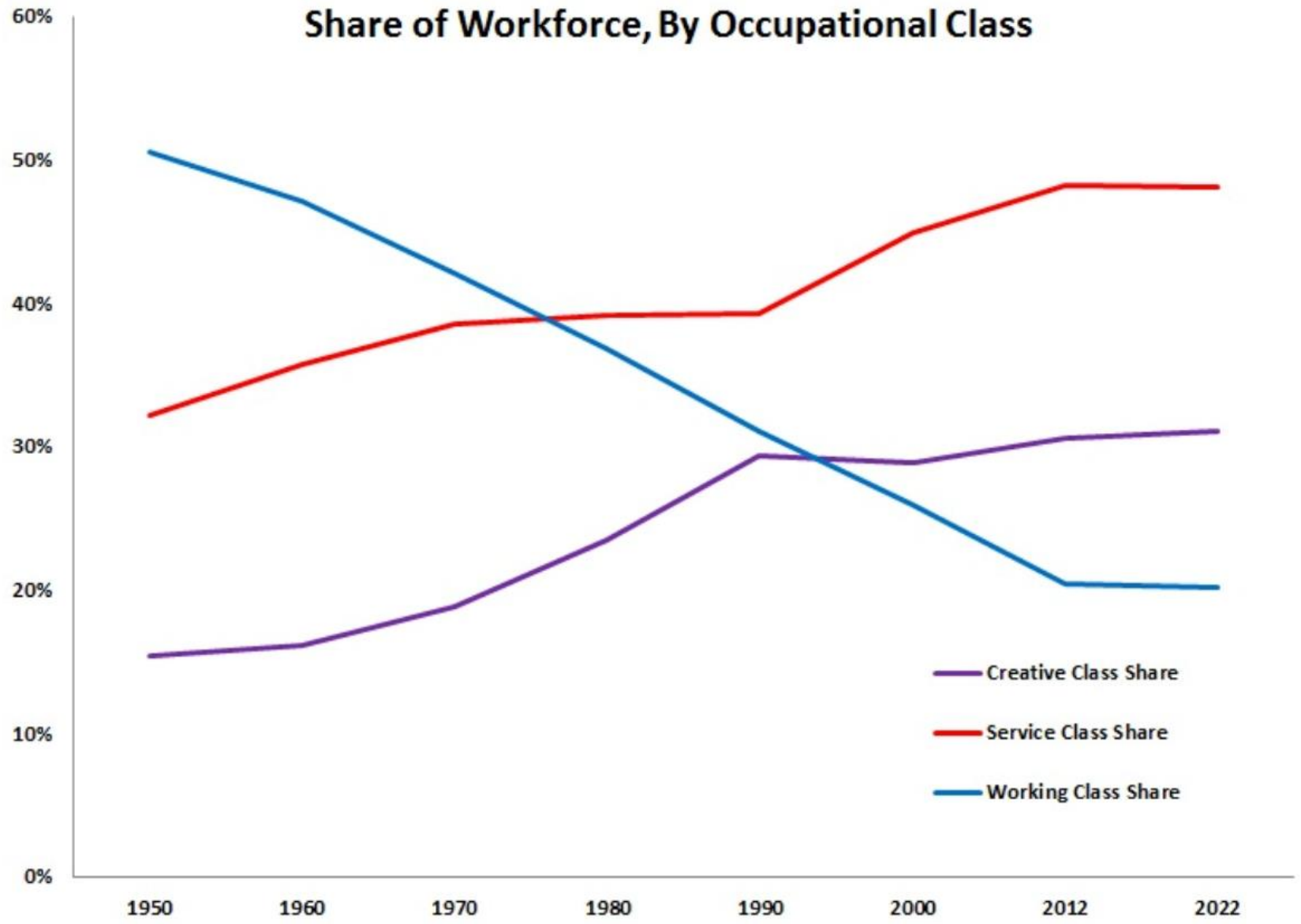
©2018 Sami Khuri

Labor Force Share by Age Group: 1996, 2006, 2016, and Projected 2026



Bureau of Labor Statistics, October 2017

©2018 Sami Khuri



https://www.citylab.com/life/2014/02/where-good-and-bad-jobs-will-be-10-years-now/8470/
and Bureau of Labor Statistics

©2018 Sami Khuri