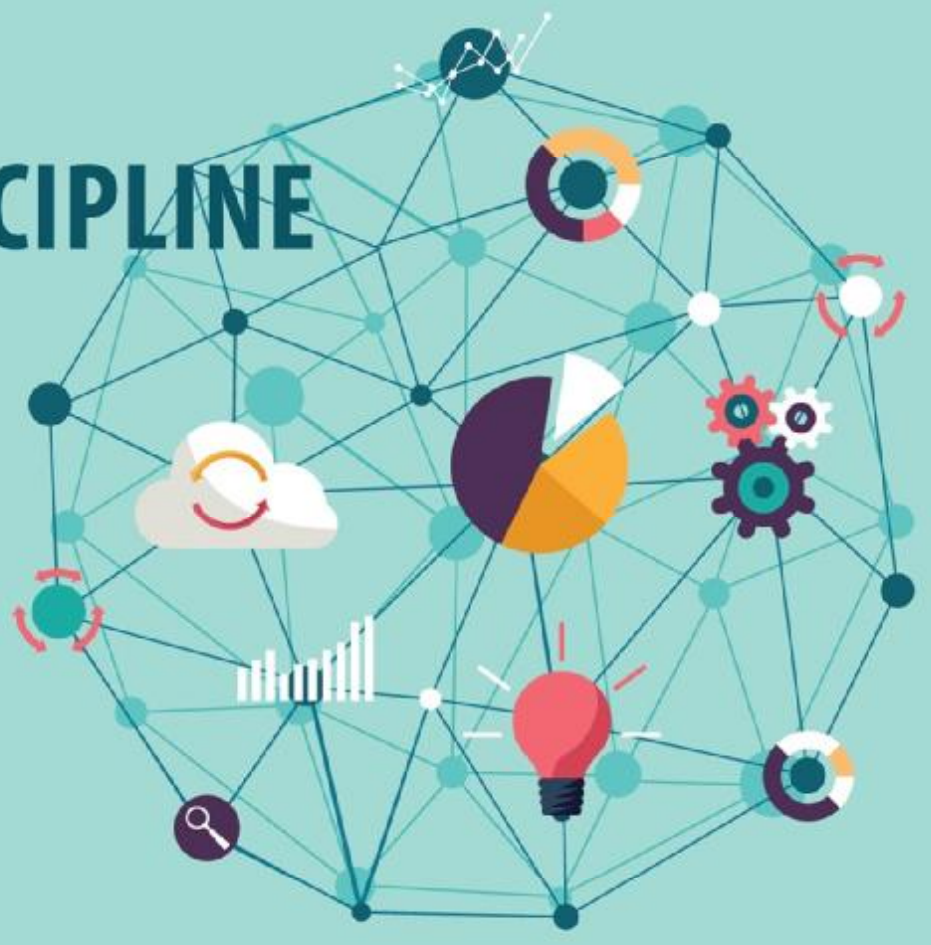
Envisioning the

DATA SCIENCE DISCIPLINE

The Undergraduate Perspective



Webinar Series
Fall 2017

The National Academies of | SCIENCES ENGINEERING MEDICINE

nas.edu/EnvisioningDS

Envisioning the
DATA SCIENCE DISCIPLINE
The Undergraduate Perspective

9/12/17 – Building Data Acumen *(recording posted)*

9/19/17 – Incorporating Real-World Applications *(recording posted)*

9/26/17 – Faculty Training and Curriculum Development *(recording posted)*

10/3/17 – Communication Skills and Teamwork *(recording posted)*

10/10/17 – Inter-Departmental Collaboration and Institutional Organization *(recording posted)*

10/17/17 – Ethics *(recording posted)*

10/24/17 – Assessment and Evaluation for Data Science Programs *(recording posted)*

11/7/17 – Diversity, Inclusion, and Increasing Participation *(recording posted)*

11/14/17 – Two-Year Colleges and Institutional Partnerships

Provide input, download the interim report, and learn more about the study at www.nas.edu/EnvisioningDS

Envisioning the DATA SCIENCE DISCIPLINE

The Undergraduate Perspective

Two-Year Colleges and Institutional Partnerships

Brian Kotz,
Montgomery College
Professor, Mathematics and Statistics

Suzanne Smith,
Johnson County Community College
Assistant Professor, Computer Information Systems

Provide input and learn more about the study at www.nas.edu/EnvisioningDS

Envisioning the DATA SCIENCE DISCIPLINE

The Undergraduate Perspective

Two-Year Colleges and Institutional Partnerships

Brian Kotz,
Montgomery College
Professor, Mathematics and Statistics

Developing a Two-Year College Certificate Program in Data Science

Frequently Asked Questions

- Why create a program?
- How was the program created?
- Who assisted?
- What's next?

Frequently Asked Questions

- Why create a program?
- How was the program created?
- Who assisted?
- What's next?

Presenter's Background:

- Former programmer/consultant
- 23 years teaching college level statistics and mathematics
- American Statistical Association (ASA) and American Mathematical Association of Two-Year Colleges (AMATYC) appointed committee member.

Why Pursue This?

Plenty going on in TYCs: completion, redesign, remediation,...

Teaching!

Why Pursue This?

Plenty going on in TYCs: completion, redesign, remediation,...

Teaching!

Dr. DJ Patil

"I'm the U.S. Chief Data Scientist — and I got my start in community college."

https://obamawhitehouse.archives.gov/blog/2015/05/06/email-dj-patil-how-i-became-chief-data-scientist

Why Pursue This?

Plenty going on in TYCs: completion, redesign, remediation,…

Teaching!

Dr. DJ Patil

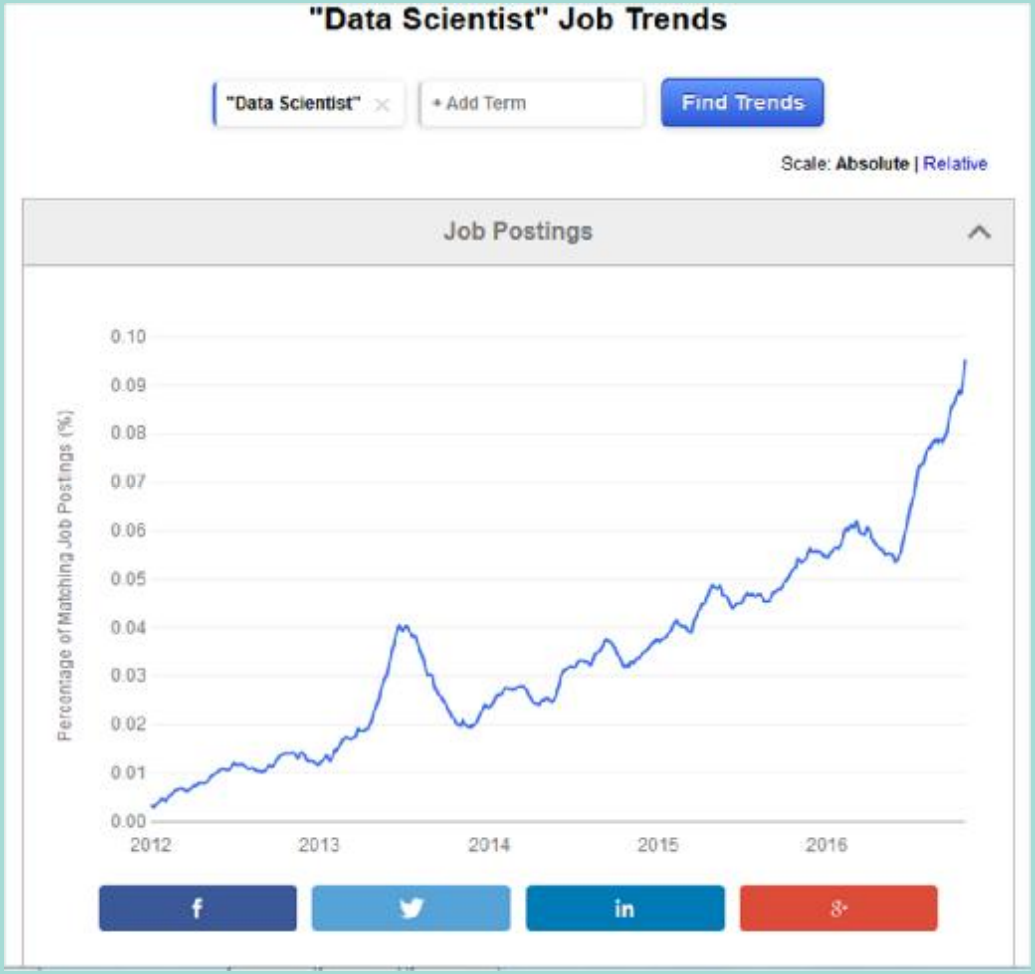
"I'm the U.S. Chief Data Scientist — and I got my start in community college."

https://obamawhitehouse.archives.gov/blog/2015/05/06/email-dj-patil-how-i-became-chief-data-scientist

"The Certificate in Practical Data Science is designed for undergraduate students…"

Great Bay Community College

Why Pursue This?

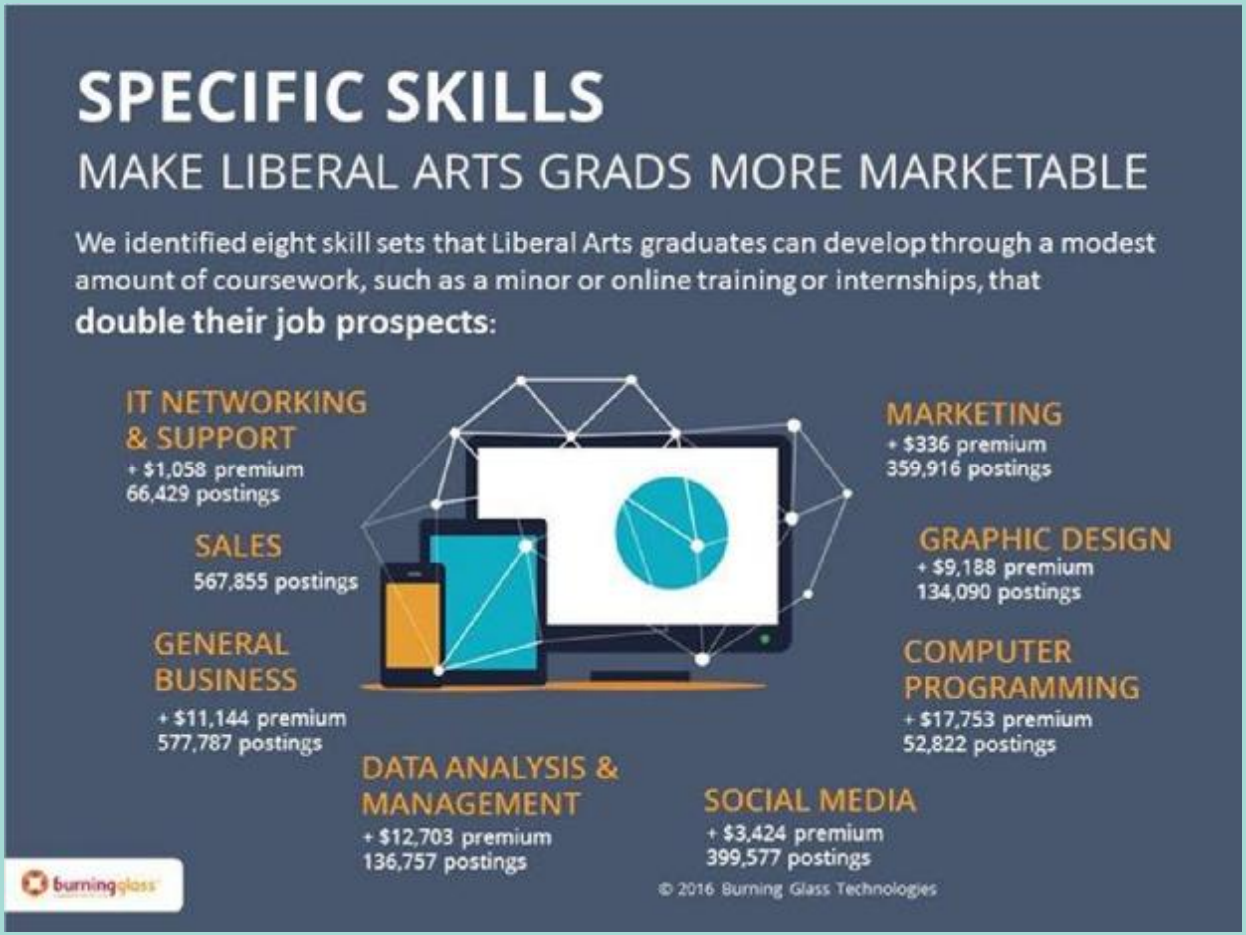


Better get involved!

https://www.indeed.com/jobtrends/q-%22Data-Scientist%22.html (accessed January 3, 2017)

Provide input and learn more about the study at www.nas.edu/EnvisioningDS

Why Pursue This?



http://www.chronicle.com/img/photos/biz/liberal-arts-skills_683x512.jpeg

Why Pursue This?

Provide input and learn more about the study at www.nas.edu/EnvisioningDS

Why Pursue This?

"By constantly using information in completely new ways, we're cracking the cancer code."

Dana-Farber Cancer Institute, discovercarebelieve.org

Why Mathematics and Statistics?

Part of data science and of other disciplines teaching data science (e.g., business, computer science)

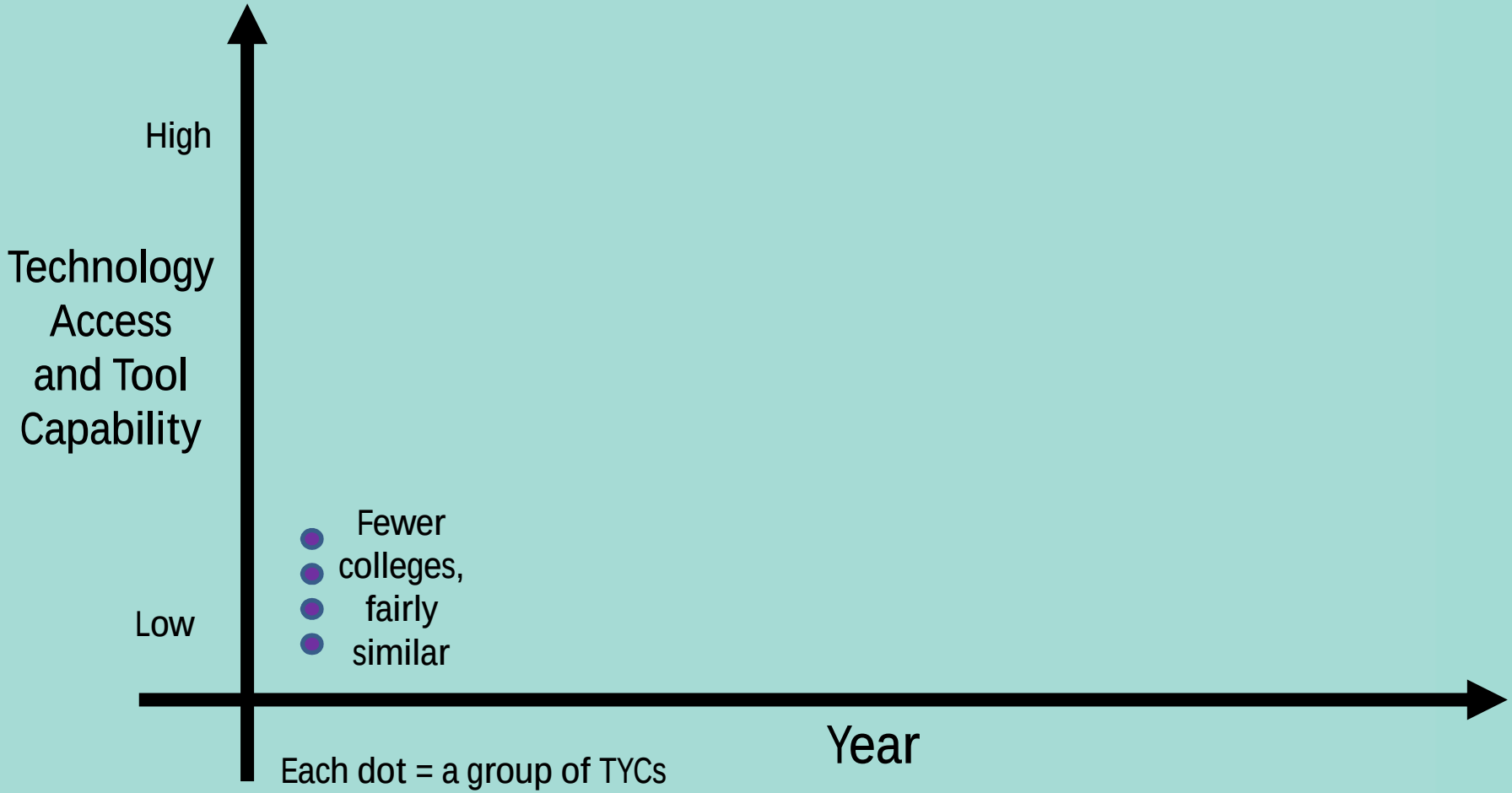
Why Mathematics and Statistics?

Part of data science and of other disciplines teaching data science (e.g., business, computer science)

Winners asked: “Will you have data science classes soon?”

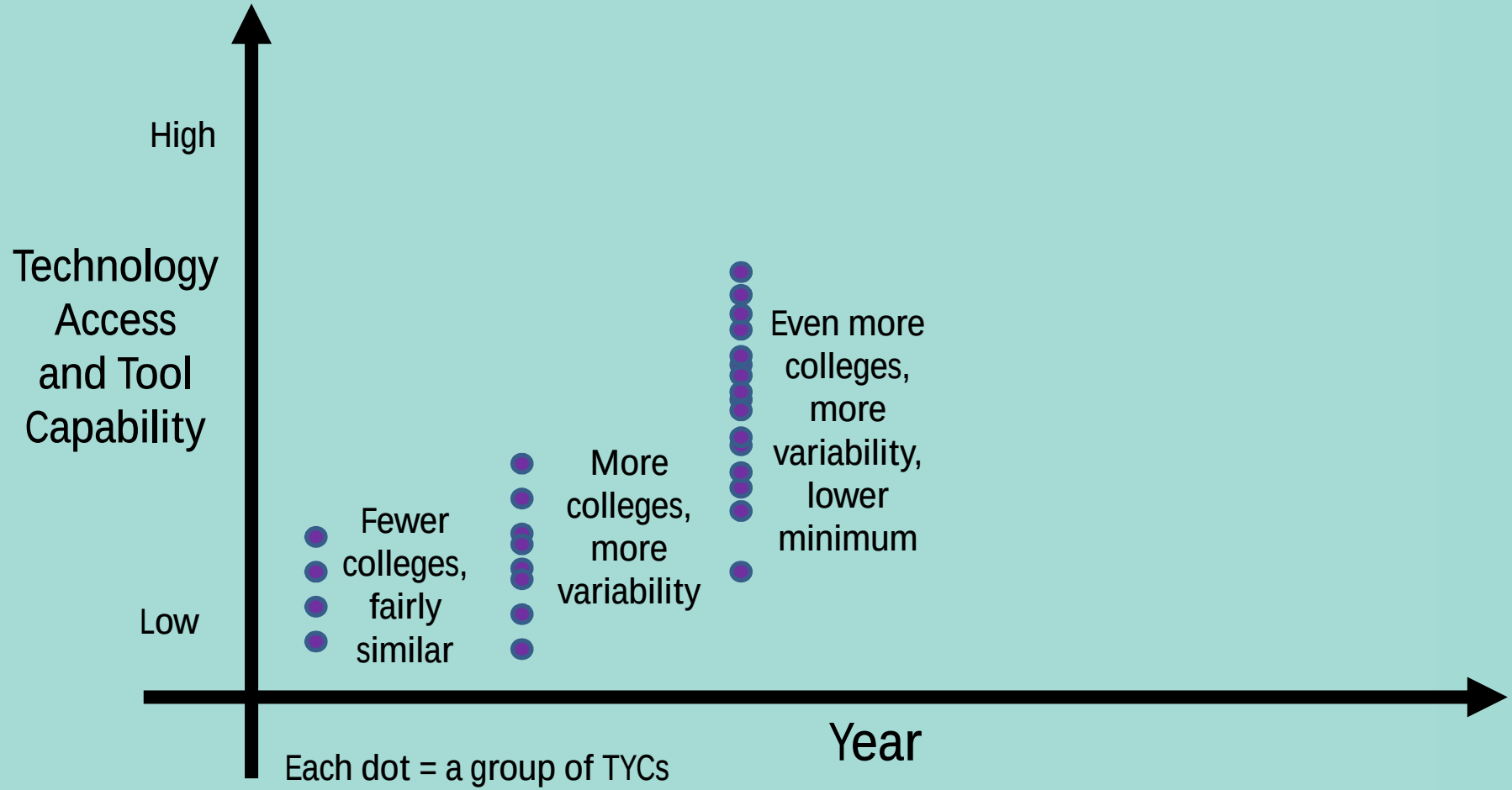
Why Mathematics and Statistics?

History of Two-Year College Statistics Instruction (in my opinion)



Why Mathematics and Statistics?

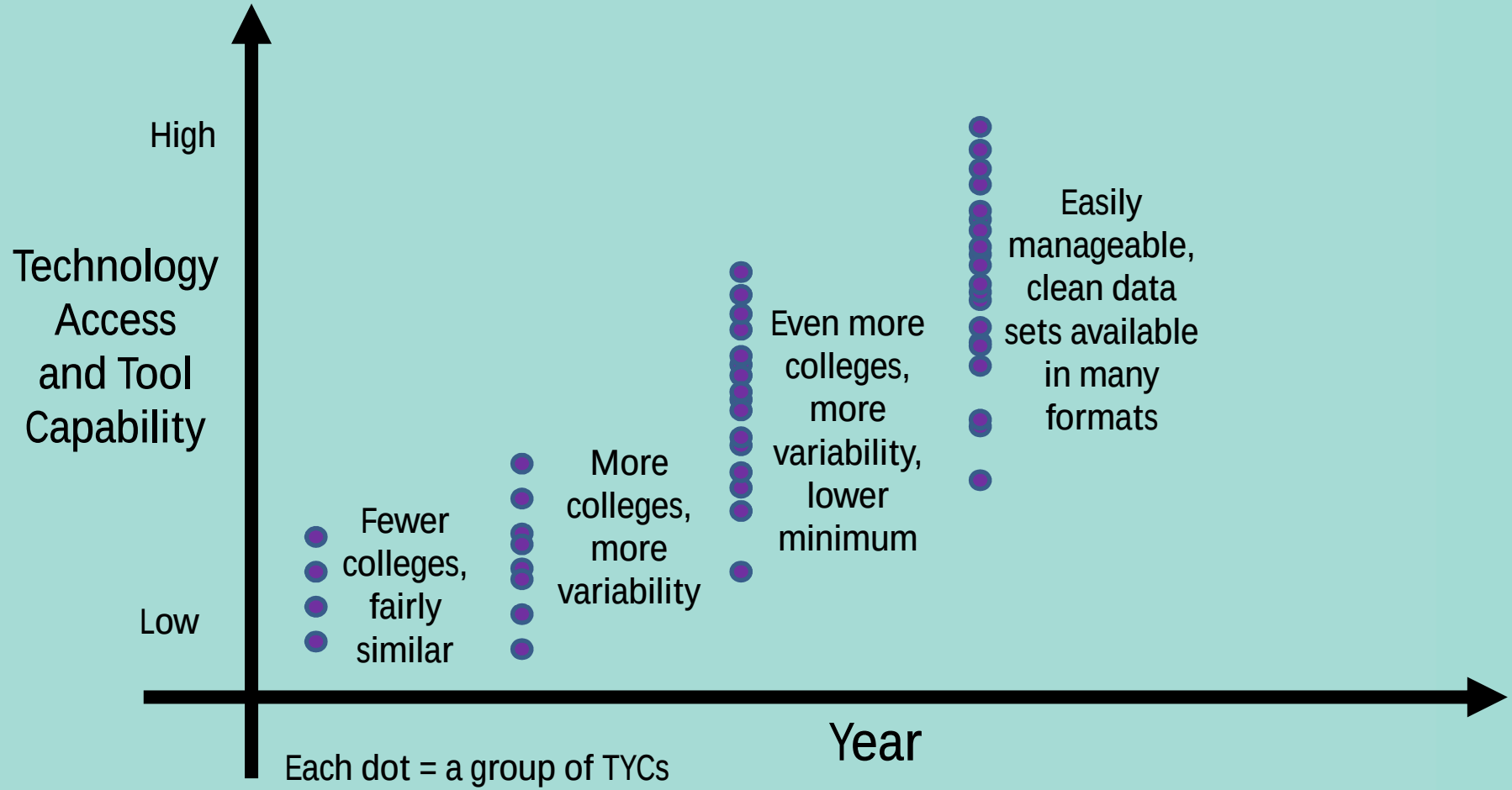
History of Two-Year College Statistics Instruction (in my opinion)



Provide input and learn more about the study at www.nas.edu/EnvisioningDS

Why Mathematics and Statistics?

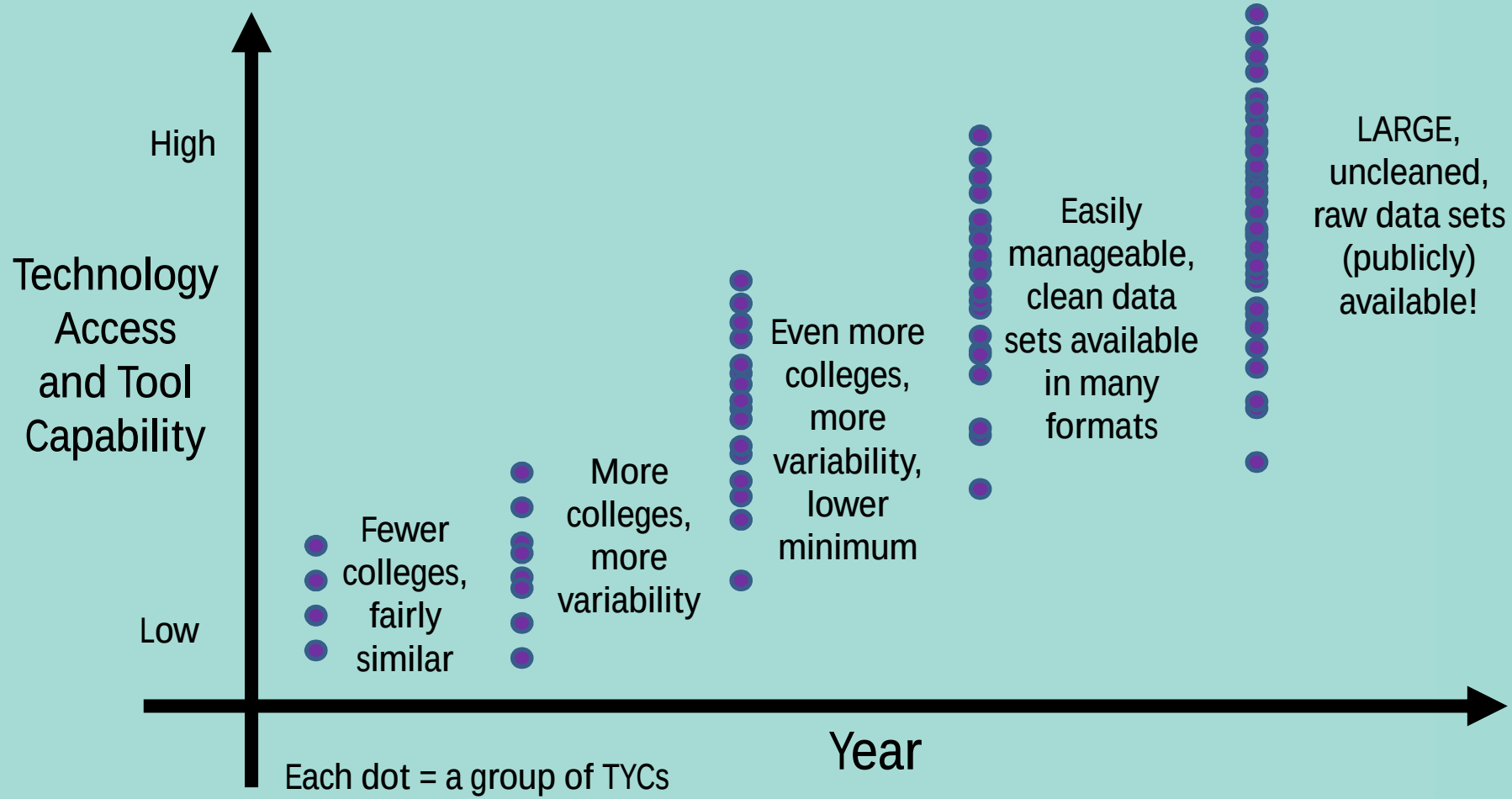
History of Two-Year College Statistics Instruction (in my opinion)



Provide input and learn more about the study at www.nas.edu/EnvisioningDS

Why Mathematics and Statistics?

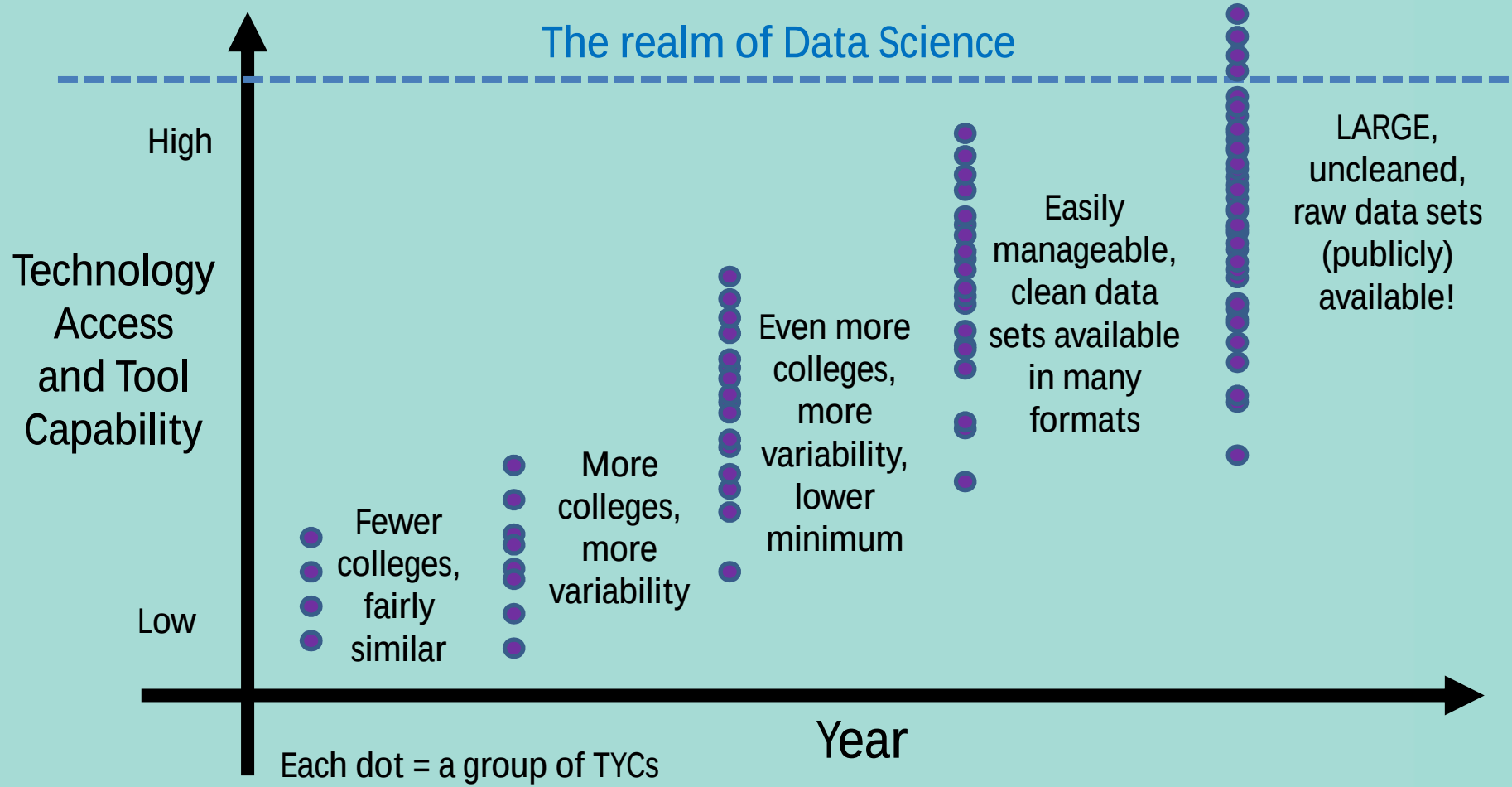
History of Two-Year College Statistics Instruction (in my opinion)



Provide input and learn more about the study at www.nas.edu/EnvisioningDS

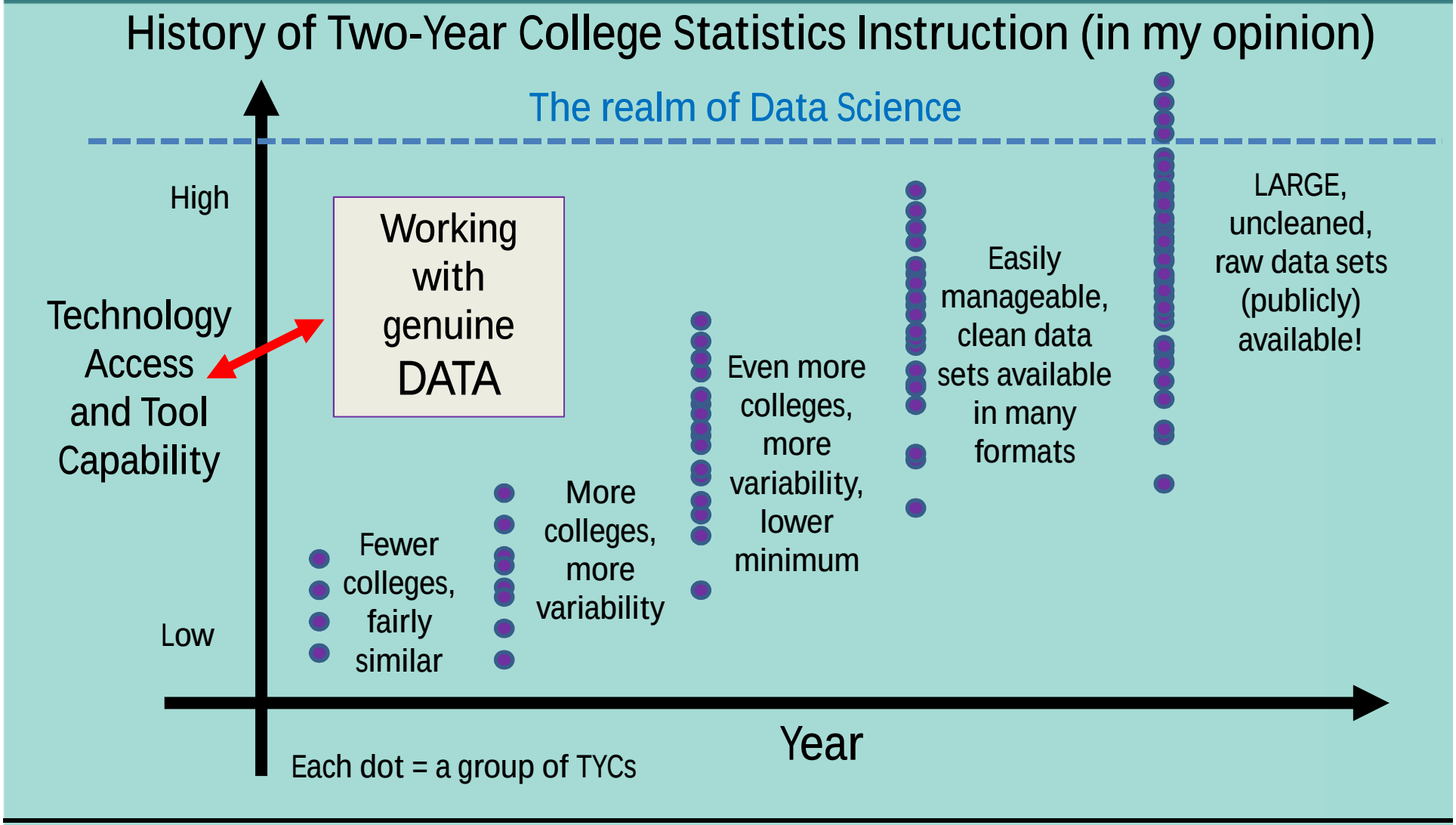
Why Mathematics and Statistics?

History of Two-Year College Statistics Instruction (in my opinion)



Provide input and learn more about the study at www.nas.edu/EnvisioningDS

Why Mathematics and Statistics?



Provide input and learn more about the study at www.nas.edu/EnvisioningDS

How Was It Built?

PROPOSALS….

- Forms, Forms, and More Forms
- Support from ASA


Sincerely,

Jessica Utts

Jessica Utts
President, American Statistical Association

- Support from local businesses *(such as DataFest sponsors)*


summit

Letter of Support for a Data Science Curriculum at Montgomery College

- Show data for local market demand, opportunities, and many 4-year programs
- Mention DataFest students, Chief Data Scientist, and other information.

How Was It Built?

PROPOSALS….

- Forms, Forms, and More Forms
- Support from ASA

Sincerely,

Jessica Utts

Jessica Utts
President, American Statistical Association

Approval of Collegewide Curriculum Committee
(multi-discipline, multi-academic unit)
and Maryland Higher Education Commission

- Support from local businesses
  (such as DataFest sponsors)

summit

Letter of Support for a Data Science Curriculum at Montgomery College

- Show data for local market demand, opportunities, and many 4-year programs
- Mention DataFest students, Chief Data Scientist, and other information.

Who Are the TYC Students?

Three distinct groups:

Would like a credit certificate (or coursework) from an accredited state college to supplement career, research, CV, etc. (might already have a degree)

Current student or "Visiting" 2-year or 4-year college student with room in schedule and interest

Transfer student ("I would like to major in data science at a 4-year school and enter that school, preferably as a junior.")

Who Are the TYC Students?

Three distinct groups:

Would like a credit certificate (or coursework) from an accredited state college to supplement career, research, CV, etc. (might already have a degree)

Yes ü

Current student or "Visiting" 2-year or 4-year college student with room in schedule and interest

Yes ü

Transfer student ("I would like to major in data science at a 4-year school and enter that school, preferably as a junior.")

Programs elsewhere!

Who Are the TYC Students?

Three distinct groups:

Would like a credit certificate (or coursework) from an accredited state college to supplement career, research, CV, etc. (might already have a degree)

Current student or "Visiting" 2-year or 4-year college student with room in schedule and interest

Transfer student ("I would like to major in data science at a 4-year school and enter that school, preferably as a junior.")

Average age of all students at our college: 25

Just graduated high school

Masters in psychology

Senior discount

What Now?

- Development Team Lead
- "CEO of a Flight School"

Does the CEO ever get to fly?
Does the CEO have to know how to fly?

https://www.aopa.org/training-and-safety/flight-schools/flight-school-business/newsletter/2016/june/13/ten-things-you-must-know; clker.com

"1. Very few schools make their owners rich.
2. You're not in the flying business.
…
8. Get used to wearing many hats.
9. Your most important role is CEO.
10. The rewards you see will be many."

- "Yobidashi"

- Builds the stage
- Beats the drums announcing matches
- Calls people in

"…it would be specious to attempt to draw up a complete list of the yobidashi's responsibilities."

http://www.lemondedusumo.com/english/

What's Next?

A Bright Future…

- At least 10 TYC programs (and in many departments!)
- At least 5 NSF funded projects involving TYC data science
  - National Academies of Sciences
  - American Statistical Association
  - South Big Data Hub
- New AMATYC Data Science Subcommittee

People want to work with Two-Year Colleges on this!

Thank You

brian.kotz@montgomerycollege.edu

www.montgomerycollege.edu/datascience
(in progress)

Envisioning the

DATA SCIENCE DISCIPLINE

The Undergraduate Perspective

Two-Year Colleges and Institutional Partnerships

Data Analytics Certificate Program at JCCC

Suzanne Smith,
Johnson County Community College
Assistant Professor, Computer Information Systems

Data Analytics Certificate Program

Launched this semester!

Suzanne Smith
Assistant Professor
Computer Science Information Systems
Johnson County Community College

Previously: Software Developer (10 yrs), Mathematics Professor (10 yrs)

The NSF Grant:

Creating Pathways for Big Data Careers

The NSF Grant:

Creating Pathways for Big Data Careers

To address the urgent need for workers who have "big data" skills, the Oceans of Data Institute (ODI) of the Education Development Center will work with four community colleges that are leaders in the area of big data, data science, and data analytics to create a career pathway model for big data careers, with special attention to the "middle skill" jobs that can be filled by graduates of community college programs.

Certificate Details

The Data Science Certificate at JCCC will allow students to learn each stage of the data science pipeline.

Students will learn how to **access data** from a variety of sources including relational databases, NoSQL data stores, and web-based APIs. Next, students will learn how to tackle the important job of **cleaning and formatting data**, also known as data wrangling. Students will then be exposed to the rapidly changing fields of **statistics and machine learning** to achieve further insight into data. As the most complex and revealing data analysis is meaningless without effective communication, this certificate program will go past the challenge of simply visually communicating quantitative information and **help students tell stories with data to impact the organizational status quo.** These skills will be brought together and applied in each course through **real-world projects** coordinated with local industries.

Certificate Details

The Data Science Certificate at JCCC will allow students to learn each stage of the data science pipeline.

Students will learn how to access data from a variety of sources including relational databases, NoSQL data stores, and web-based APIs. Next, students will learn how to tackle the important job of cleaning and formatting data, also known as data wrangling. Students will then be exposed to the rapidly changing fields of statistics and machine learning to achieve further insight into data. As the most complex and revealing data analysis is meaningless without effective communication, this certificate program will go past the challenge of simply visually communicating quantitative information and help students tell stories with data to impact the organizational status quo. These skills will be brought together and applied in each course through real-world projects coordinated with local industries.

Certificate Details

The Data Science Certificate at JCCC will allow students to learn each stage of the data science pipeline.

Students will learn how to access data from a variety of sources including relational databases, NoSQL data stores, and web-based APIs. Next, students will learn how to tackle the important job of cleaning and formatting data, also known as data wrangling. Students will then be exposed to the rapidly changing fields of statistics and machine learning to achieve further insight into data. As the most complex and revealing data analysis is meaningless without effective communication, this certificate program will go past the challenge of simply visually communicating quantitative information and help students tell stories with data to impact the organizational status quo. These skills will be brought together and applied in each course through real-world projects coordinated with local industries.

Certificate Details

The Data Science Certificate at JCCC will allow students to learn each stage of the data science pipeline.

Students will learn how to access data from a variety of sources including relational databases, NoSQL data stores, and web-based APIs. Next, students will learn how to tackle the important job of cleaning and formatting data, also known as data wrangling. Students will then be exposed to the rapidly changing fields of statistics and machine learning to achieve further insight into data. As the most complex and revealing data analysis is meaningless without effective communication, this certificate program will go past the challenge of simply visually communicating quantitative information and help students tell stories with data to impact the organizational status quo. These skills will be brought together and applied in each course through real-world projects coordinated with local industries.

Certificate Details

The Data Science Certificate at JCCC will allow students to learn each stage of the data science pipeline.

Students will learn how to access data from a variety of sources including relational databases, NoSQL data stores, and web-based APIs. Next, students will learn how to tackle the important job of cleaning and formatting data, also known as data wrangling. Students will then be exposed to the rapidly changing fields of statistics and machine learning to achieve further insight into data. As the most complex and revealing data analysis is meaningless without effective communication, this certificate program will go past the challenge of simply visually communicating quantitative information and help students tell stories with data to impact the organizational status quo. These skills will be brought together and applied in each course through real-world projects coordinated with local industries.

Certificate Details

The Data Science Certificate at JCCC will allow students to learn each stage of the data science pipeline.

Students will learn how to access data from a variety of sources including relational databases, NoSQL data stores, and web-based APIs. Next, students will learn how to tackle the important job of cleaning and formatting data, also known as data wrangling. Students will then be exposed to the rapidly changing fields of statistics and machine learning to achieve further insight into data. As the most complex and revealing data analysis is meaningless without effective communication, this certificate program will go past the challenge of simply visually communicating quantitative information and help students tell stories with data to impact the organizational status quo. These skills will be brought together and applied in each course through real-world projects coordinated with local industries.

Certificate Details

The Data Science Certificate at JCCC will allow students to learn each stage of the data science pipeline.

Students will learn how to access data from a variety of sources including relational databases, NoSQL data stores, and web-based APIs. Next, students will learn how to tackle the important job of cleaning and formatting data, also known as data wrangling. Students will then be exposed to the rapidly changing fields of statistics and machine learning to achieve further insight into data. As the most complex and revealing data analysis is meaningless without effective communication, this certificate program will go past the challenge of simply visually communicating quantitative information and help students tell stories with data to impact the organizational status quo. These skills will be brought together and applied in each course through real-world projects coordinated with local industries.

Certificate Details

Many companies cannot afford to hire Masters and PhD level Data Scientists.

Our students will be more affordable to hire and can fill that gap.

Certificate Details

A student successfully completing this certificate will have:

- An understanding of the components of data science.
- The ability to analyze various data sets using statistics and computer programming.
- The ability to create visualizations of the data.
- The ability to communicate the findings with a variety of audiences.
- An ePortfolio detailing their solutions and accomplishments during the completion of the certificate.

Certificate Details

A student successfully completing this certificate will have:

- An understanding of the components of data science.
- The ability to analyze various data sets using statistics and computer programming.
- The ability to create visualizations of the data.
- The ability to communicate the findings with a variety of audiences.
- An ePortfolio detailing their solutions and accomplishments during the completion of the certificate.

Certificate Details

A student successfully completing this certificate will have:

- An understanding of the components of data science.
- The ability to analyze various data sets using statistics and computer programming.
- The ability to create visualizations of the data.
- The ability to communicate the findings with a variety of audiences.
- An ePortfolio detailing their solutions and accomplishments during the completion of the certificate.

Certificate Details

A student successfully completing this certificate will have:

- An understanding of the components of data science.
- The ability to analyze various data sets using statistics and computer programming.
- The ability to create visualizations of the data.
- The ability to communicate the findings with a variety of audiences.
- An ePortfolio detailing their solutions and accomplishments during the completion of the certificate.

Certificate Details

A student successfully completing this certificate will have:

- An understanding of the components of data science.
- The ability to analyze various data sets using statistics and computer programming.
- The ability to create visualizations of the data.
- The ability to communicate the findings with a variety of audiences.
- An ePortfolio detailing their solutions and accomplishments during the completion of the certificate.

Certificate Details

A student successfully completing this certificate will have:

- An understanding of the components of data science.
- The ability to analyze various data sets using statistics and computer programming.
- The ability to create visualizations of the data.
- The ability to communicate the findings with a variety of audiences.
- An ePortfolio detailing their solutions and accomplishments during the completion of the certificate.

Certificate Details

All courses must be completed with a grade of C or above.

Semester One

- DS 210 Introduction fo Data Science (3 hr)
- MATH 181 Statistics (3 hr)
- DS 220 Data Visualization (3 hr)
- DS 230 SQL for Data Analysis (3 hr)
- DS 240 Statistical Programming (3 hr)

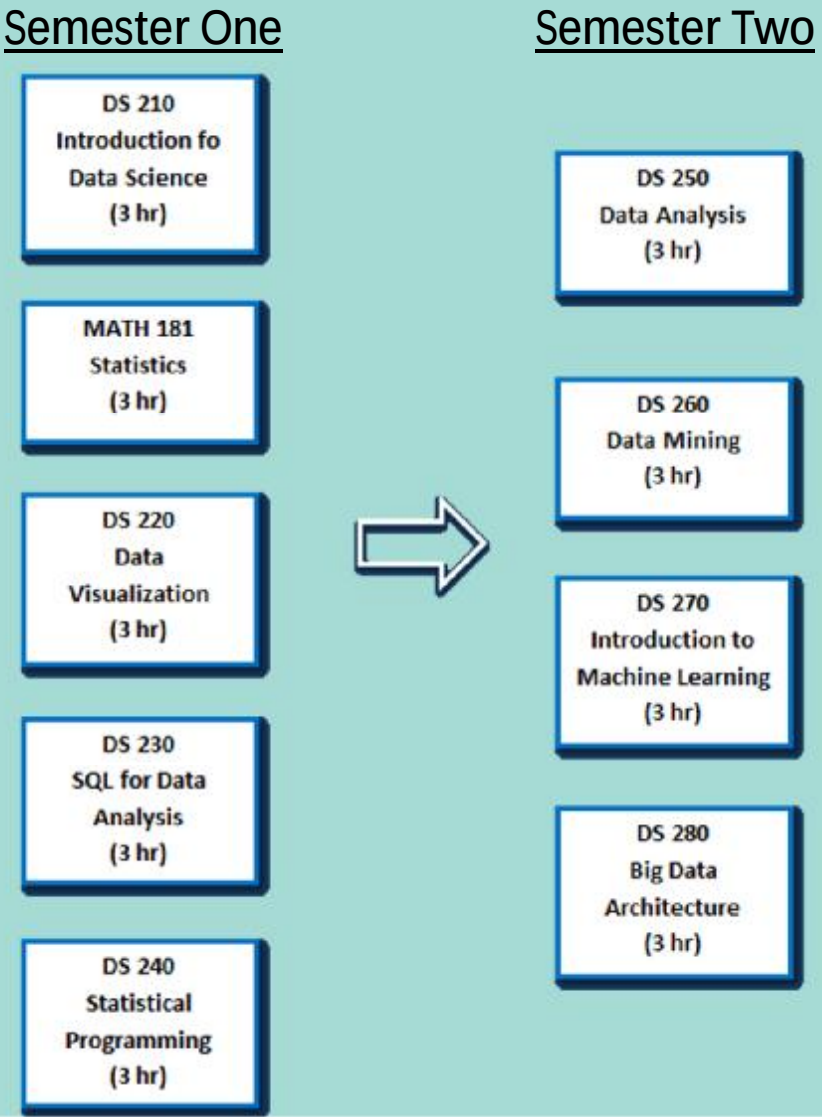
Semester Two

- DS 250 Data Analysis (3 hr)
- DS 260 Data Mining (3 hr)
- DS 270 Introduction to Machine Learning (3 hr)
- DS 280 Big Data Architecture (3 hr)

Certificate Details

All courses must be completed with a grade of C or above.

There are no prerequisites except for Statistics which requires the completion of College Algebra.

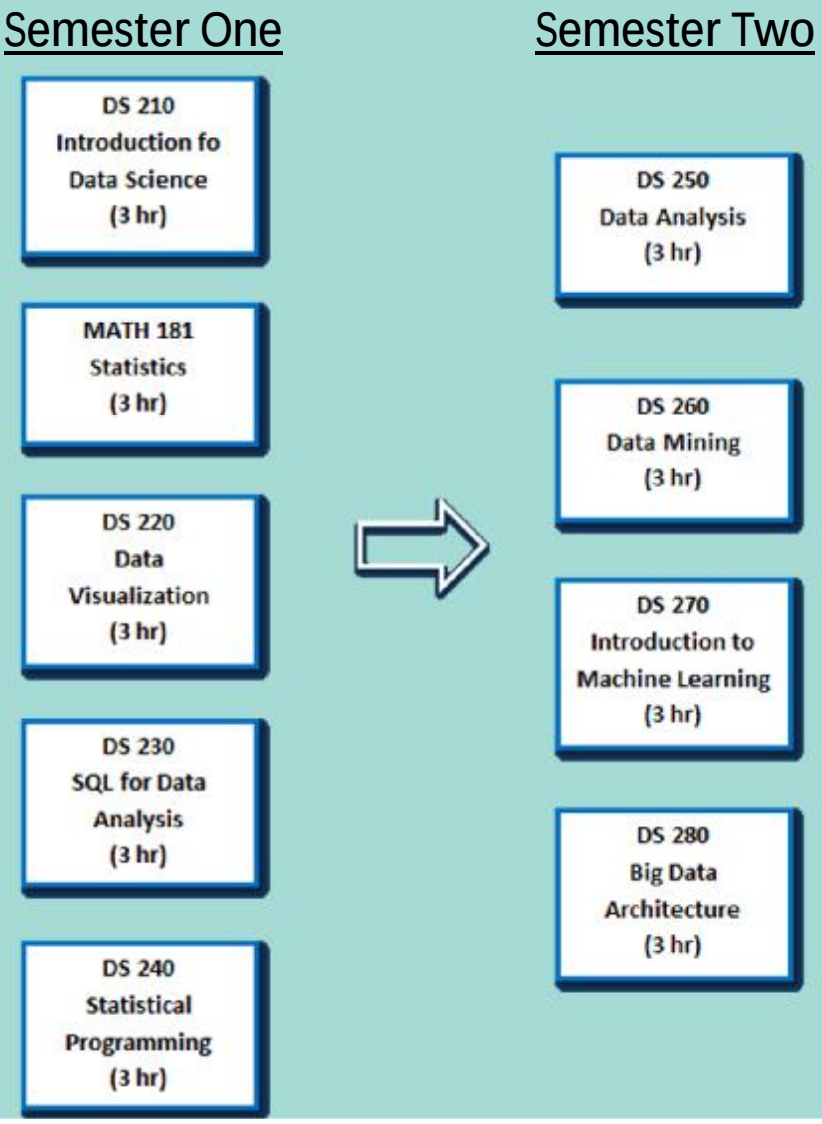


Certificate Details

All courses must be completed with a grade of C or above.

There are no prerequisites except for Statistics which requires the completion of College Algebra.

This is a problem!



First Semester Courses

DS 210
Introduction to Data Science
(3 hrs)

DS 220
Data Visualization
(3 hrs)

DS 230
SQL for Data Analysis
(3 hrs)

DS 240
Statistical Programming
(3 hrs)

MATH 181
Statistics
(3 hrs)

First Semester Courses

DS 210
Introduction to Data Science
(3 hrs)

Mostly Python
Some R APIs
Web Scraping
Ethics
Life Cycle

DS 220
Data Visualization
(3 hrs)

DS 230
SQL for Data Analysis
(3 hrs)

DS 240
Statistical Programming
(3 hrs)

MATH 181
Statistics
(3 hrs)

First Semester Courses

DS 210
Introduction to Data Science
(3 hrs)

Mostly Python
Some R APIs
Web Scraping
Ethics
Life Cycle

DS 220
Data Visualization
(3 hrs)

Some R
Mostly Tableau
Taught by Data Scientist from Cerner

DS 230
SQL for Data Analysis
(3 hrs)

DS 240
Statistical Programming
(3 hrs)

MATH 181
Statistics
(3 hrs)

First Semester Courses

DS 210
Introduction to Data Science
(3 hrs)

Mostly Python
Some R APIs
Web Scraping
Ethics
Life Cycle

DS 220
Data Visualization
(3 hrs)

Some R
Mostly Tableau
Taught by Data Scientist from Cerner

DS 230
SQL for Data Analysis
(3 hrs)

SQL
(SQL Server)
Excel for Data Visualizaiton

DS 240
Statistical Programming
(3 hrs)

MATH 181
Statistics
(3 hrs)

First Semester Courses

DS 210
Introduction to Data Science
(3 hrs)

Mostly Python
Some R APIs
Web Scraping
Ethics
Life Cycle

DS 220
Data Visualization
(3 hrs)

Some R
Mostly Tableau
Taught by Data Scientist from Cerner

DS 230
SQL for Data Analysis
(3 hrs)

SQL
(SQL Server)
Excel for Data Visualizaiton

DS 240
Statistical Programming
(3 hrs)

All R
Taught by Data Scientist from Sprint

MATH 181
Statistics
(3 hrs)

First Semester Courses

DS 210
Introduction to Data Science
(3 hrs)

Mostly Python
Some R APIs
Web Scraping
Ethics
Life Cycle

DS 220
Data Visualization
(3 hrs)

Some R
Mostly Tableau
Taught by Data Scientist from Cerner

DS 230
SQL for Data Analysis
(3 hrs)

SQL
(SQL Server)
Excel for Data Visualizaiton

DS 240
Statistical Programming
(3 hrs)

All R
Taught by Data Scientist from Sprint

MATH 181
Statistics
(3 hrs)

Taught in Math Department
Various Software Tools

Second Semester Courses

DS 250
Data Analysis
(3 hrs)

DS 260
Data Mining
(3 hrs)

DS 270
Introduction to Machine Learning (3 hrs)

DS 280
Big Data Architecture
(3 hrs)

Second Semester Courses

DS 250
Data Analysis
(3 hrs)

Mostly R
Some Python
Taught by
Data Scientist
from Sprint

DS 260
Data Mining
(3 hrs)

DS 270
Introduction to
Machine
Learning (3 hrs)

DS 280
Big Data
Architecture
(3 hrs)

Second Semester Courses

DS 250
Data Analysis
(3 hrs)

Mostly R
Some Python
Taught by
Data Scientist
from Sprint

DS 260
Data Mining
(3 hrs)

To Be
Determined

DS 270
Introduction to
Machine
Learning (3 hrs)

DS 280
Big Data
Architecture
(3 hrs)

Second Semester Courses

DS 250
Data Analysis
(3 hrs)

Mostly R
Some Python
Taught by
Data Scientist
from Sprint

DS 260
Data Mining
(3 hrs)

To Be
Determined

DS 270
Introduction to
Machine
Learning (3 hrs)

Python
Data Scientist
from Hallmark
is interested

DS 280
Big Data
Architecture
(3 hrs)

Second Semester Courses

DS 250
Data Analysis
(3 hrs)

Mostly R
Some Python
Taught by
Data Scientist
from Sprint

DS 260
Data Mining
(3 hrs)

To Be
Determined

DS 270
Introduction to
Machine
Learning (3 hrs)

Python
Data Scientist
from Hallmark
is interested

DS 280
Big Data
Architecture
(3 hrs)

Hadoop, etc.
Taught by Tech
Professional
from Century
Link

Target Student

A successful student will have

- Strong analytical skills
- An interest in Computer Science and programming
- A desire to apply their knowledge to an industry such as business, medicine, marketing, etc.

Target Student

A successful student will have

- Strong analytical skills
- An interest in Computer Science and programming
- A desire to apply their knowledge to an industry such as business, medicine, marketing, etc.

Target Student

A successful student will have

- Strong analytical skills
- An interest in Computer Science and programming
- A desire to apply their knowledge to an industry such as business, medicine, marketing, etc.

Target Student

A successful student will have

- Strong analytical skills
- An interest in Computer Science and programming
- A desire to apply their knowledge to an industry such as business, medicine, marketing, etc.

Target Student

A successful student will have

- Strong analytical skills
- An interest in Computer Science and programming
- A desire to apply their knowledge to an industry such as business, medicine, marketing, etc.

Our students come from a wide variety of educational backgrounds.

Established Professionals

Our certificate could be very beneficial for someone out in industry who has specific industry knowledge but wants to develop skills in data analytics to help satisify the needs of their employer.

Our students come from a wide variety of educational backgrounds.

Established Professionals

Our certificate could be very beneficial for someone out in industry who has specific industry knowledge but wants to develop skills in data analytics to help satisify the needs of their employer.

Students with a Bachelors Degree

Some students will have a bachelors degree (in Computer Science, Business, or something else) and want to add the certificate and data analytics skills to their resume.

Our students come from a wide variety of educational backgrounds.

Established Professionals

Our certificate could be very beneficial for someone out in industry who has specific industry knowledge but wants to develop skills in data analytics to help satisify the needs of their employer.

Students with a Bachelors Degree

Some students will have a bachelors degree (in Computer Science, Business, or something else) and want to add the certificate and data analytics skills to their resume.

Students at Associates Level

Although this certificate does not lead to a degree, students starting out can use these hours as their electives while earning an associates degree.

Challenges and Successes

Questions?

Suzanne Smith, ssmit348@jccc.edu

Envisioning the DATA SCIENCE DISCIPLINE

The Undergraduate Perspective

Two-Year Colleges and Institutional Partnerships

Brian Kotz,
Montgomery College
Professor, Mathematics and Statistics

Suzanne Smith,
Johnson County Community College
Assistant Professor, Computer Information Systems

Provide input and learn more about the study at www.nas.edu/EnvisioningDS

Envisioning the
DATA SCIENCE DISCIPLINE
The Undergraduate Perspective

9/12/17 – Building Data Acumen *(recording posted)*

9/19/17 – Incorporating Real-World Applications *(recording posted)*

9/26/17 – Faculty Training and Curriculum Development *(recording posted)*

10/3/17 – Communication Skills and Teamwork *(recording posted)*

10/10/17 – Inter-Departmental Collaboration and Institutional Organization *(recording posted)*

10/17/17 – Ethics *(recording posted)*

10/24/17 – Assessment and Evaluation for Data Science Programs *(recording posted)*

11/7/17 – Diversity, Inclusion, and Increasing Participation *(recording posted)*

11/14/17 – Two-Year Colleges and Institutional Partnerships

Provide input, download the interim report, and learn more about the study at www.nas.edu/EnvisioningDS