Developing a Two-Year College Data Science Certificate Program

Brian Kotz
Montgomery College
Germantown, Maryland
AMATYC – San Diego
November 9, 2017
brian.kotz@montgomerycollege.edu
Frequently Asked Questions:

- Why?
- How?
- Who?
- What was I thinking?
Why pursue this?

(Enough going on: completion, redesign, remediation, etc.)
Why pursue this?

(Enough going on: completion, redesign, remediation, etc.)

Teaching!
Why pursue this?

(Enough going on: completion, redesign, remediation, etc.)

“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

Dr. DJ Patil
Why pursue this?
(Enough going on: completion, redesign, remediation, etc.)

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

Dr. DJ Patil

“...The Certificate in Practical Data Science is designed for undergraduate students...”

Great Bay Community College
"Data Scientist" Job Trends

Percentage of Matching Job Postings (%)


Better get involved!

Lock Ad - Final- 30
“By constantly using information in completely new ways, we’re cracking the cancer code.”

Dana-Farber Cancer Institute, discovercarebelieve.org
SPECIFIC SKILLS
MAKE LIBERAL ARTS GRADS MORE MARKETABLE

We identified eight skill sets that Liberal Arts graduates can develop through a modest amount of coursework, such as a minor or online training or internships, that double their job prospects:

- **IT Networking & Support**: + $1,058 premium 66,429 postings
- **Marketing**: + $336 premium 359,916 postings
- **Sales**: 567,855 postings
- **Graphics Design**: + $9,188 premium 134,050 postings
- **General Business**: + $11,144 premium 577,787 postings
- **Computer Programming**: + $17,753 premium 52,822 postings
- **Data Analysis & Management**: + $12,703 premium 136,757 postings
- **Social Media**: + $3,424 premium 398,377 postings

Why have Statistics lead?

(Enough going on: GAISE, simulation based inference, etc.)
(Other disciplines: Business, Computer Science)
Why have Statistics lead?

(Enough going on: GAISE, simulation based inference, etc.)
(Other disciplines: Business, Computer Science)

May 2016
Why have Statistics lead?

(Enough going on: GAISE, simulation based inference, etc.)
(Other disciplines: Business, Computer Science)

May 2016

Students asked: “Will you have data science classes soon?”
History of Two-Year College Statistics Instruction (in my opinion)

- Each dot = a group of TYCs
History of Two-Year College Statistics Instruction (in my opinion)

Technology Access and "Sophistication"

Year

Each dot = a group of TYCs

- Fewer colleges, fairly similar
- n is a perfect square

Low

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

Each dot = a group of TYCs

- Fewer colleges, fairly similar
- More colleges, more variability
History of Two-Year College Statistics Instruction (in my opinion)

- Fewer colleges, fairly similar
- More colleges, more variability
- Even more colleges, more variability, lower minimum
- Textbooks with data sets (floppy)

Each dot = a group of TYCs
History of Two-Year College Statistics Instruction (in my opinion)

Each dot = a group of TYCs

- Fewer colleges, fairly similar
- More colleges, more variability
- Even more colleges, more variability, lower minimum
- Easily manageable, clean data sets available in many formats

Technology Access and “Sophistication”

Year

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

Technology Access and "Sophistication"

- Fewer colleges, fairly similar
- More colleges, more variability
- Even more colleges, more variability, lower minimum
- Easily manageable, clean data sets available in many formats
- LARGE, uncleaned, raw data sets (publicly) available!

Each dot = a group of TYCs

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

The realm of Data Science

- Fewer colleges, fairly similar
- More colleges, more variability
- Easily manageable, clean data sets available in many formats
- Even more colleges, more variability, lower minimum
- LARGE, uncleaned, raw data sets (publicly) available!

Each dot = a group of TYCs
History of Two-Year College Statistics Instruction (in my opinion)

The realm of Data Science

Each dot = a group of TYCs

Working with genuine DATA

Technology Access and “Sophistication”

Year

Low

High

Fewer colleges, fairly similar

More colleges, more variability

Even more colleges, more variability, lower minimum

Easily manageable, clean data sets available in many formats

LARGE, uncleaned, raw data sets (publicly) available!
MC
MONTGOMERY COLLEGE

... a two-year college in Montgomery County, Maryland (borders Washington, DC)

Fall 2016 Credit Enrollment: 23,916 students; 217,117 credit hours

MC Fall Profile, 2016

Black, Hispanic, Asian, Multi-Race, and Native American/Pacific Islander students make up roughly two-thirds of the students.

6,435 non-U.S. Citizens (27% of our students) enrolled in Fall 2016 representing 159 individual countries ... 696 on student (F-1) visas.

Average age: 25.02 years
PROPOSALS....

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

As president of the American Statistical Association, I write in support of your effort to initiate a data science curriculum with an “Introduction to Data Science” course at Montgomery College.

Sincerely,
Jessica Utts
President, American Statistical Association

- Support from local businesses (such as DataFest sponsors)

- Show data for local market demand, opportunities, and many 4-year programs
- Mention DataFest students and Chief Data Scientist
Our Certificate (16 Credits)

MATH 117 – Elements of Statistics*
DATA 101 – Introduction to Data Science
DATA 110 – Writing and Communication in Data Science
DATA 201 – Statistical Methods in Data Science
DATA 205 – Capstone Experience in Data Science

(all materials at no cost)

www.montgomerycollege.edu/datascience
(draft now)

* - Students may also take MATH 217 – Biostatistics or BSAD 210 – Statistics for Business and Economics
We are all uniquely responsive to our communities. Your results may vary—and they should.

Our Certificate (16 Credits)

MATH 117 – Elements of Statistics*
DATA 101 – Introduction to Data Science
DATA 110 – Writing and Communication in Data Science
DATA 201 – Statistical Methods in Data Science
DATA 205 – Capstone Experience in Data Science

(all materials at no cost)

www.montgomerycollege.edu/datascience
(draft now)
What am I doing?

• Development Team Lead; Program Coordinator (*de facto*)

• “CEO of a Flight School” Does the CEO ever get to fly?
  Does the CEO even know how to fly?
  

  “1. Very few schools make their owners rich.
  2. You’re not in the flying business.
  3. …
  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.”

A Bright Future?

At least 10 TYC programs

At least 5 NSF funded projects

- National Academies of Sciences
- American Statistical Association
- South Big Data Hub

AMATYC special meeting *(Saturday 2:15 pm, Marina 5)*

People want to work with us! All of us!
A Bright Future?

At least 10 TYC programs

At least 5 NSF funded projects
- National Academies of Sciences
- American Statistical Association
- South Big Data Hub

AMATYC special meeting *(Saturday 2:15 pm, Marina 5)*

People want to work with us! All of us!

And not just **PROGRAMS:**

- Infusing data science in STAT 101
- Offering DATA 101 (single course)
- Offering a Certificate
- Offering an Associate's Degree
A Bright Future?

At least 10 TYC programs

At least 5 NSF funded projects
- National Academies of Sciences
- American Statistical Association
- South Big Data Hub

AMATYC special meeting *(Saturday 2:15 pm, Marina 5)*

People want to work with us! All of us!

https://apcentral.collegeboard.org/
The Analytics Train Has Already Left the Station!

With all the data around these days, and abundant software tools, data analysis is open to people who may not have all the necessary statistical background. There is a sense among some that the statistics profession has been left behind. But in his address at JSM, ASA President Barry Nussbaum said this is the best of times for statistics. He encouraged us to be active, collaborate, get to the table early, and learn new skills to ensure a bright future for the statistics profession.

Below is a clip from Barry’s address in which he talks about what we—as statisticians—should and shouldn't do. Watch the full address online.
The Analytics Train Has Already Left the Station!

With all the data around these days, and abundant software tools, data analysis is open to people who may not have all the necessary statistical background. There is a sense among some that the statistics profession has been left behind. But in his address at JSM, ASA President Barry Nussbaum said this is the best of times for statistics. He encouraged us to be active, collaborate, get to the table early, and learn new skills to ensure a bright future for the statistics profession.

Below is a clip from Barry’s address in which he talks about what we—as statisticians—should and shouldn’t do. Watch the full address online.
THANK YOU!

Brian Kotz
Montgomery College
Germantown, Maryland
AMATYC – San Diego
November 9, 2017
brian.kotz@montgomerycollege.edu