Norm Matloff University of California,

Davis
MenloAtherton High

School

## Careers in Data Science (You Know, Statistics)

Norm Matloff<br>University of California, Davis

Menlo-Atherton High School

May 24, 2016
http://heather.cs.ucdavis.edu/MenloAtherton.pdf

Data Science<br>(You Know,<br>Statistics)

## Confusing Terms

## Norm Matloff

 University of California,Davis
Menlo-
Atherton High
School

# Confusing Terms 

## Norm Matloff

University of
California,
Davis
Menlo-
Atherton High
School
 Data Science (You Know, Statistics)

## Confusing Terms

Norm Matloff
University of
California,
Davis
Menlo-
Atherton High
School


- We live in the Age of Buzzwords.

Careers in Data Science (You Know, Statistics)

Norm Matloff
University of California,

Davis

MenloAtherton High School

## Confusing Terms



- We live in the Age of Buzzwords.
- Trust me:

Careers in Data Science (You Know, Statistics)

Norm Matloff University of California,

Davis
MenloAtherton High School

## Confusing Terms



- We live in the Age of Buzzwords.
- Trust me: Data Science $=$ Statistics.


## Confusing Terms

Norm Matloff
University of
California,
Davis
MenloAtherton High School


- We live in the Age of Buzzwords.
- Trust me: Data Science $=$ Statistics.
- But isn't Data Science $=$ Computer Science + Statistics?


## Confusing Terms

Norm Matloff

MenloAtherton High School


- We live in the Age of Buzzwords.
- Trust me: Data Science $=$ Statistics.
- But isn't Data Science $=$ Computer Science + Statistics?
- I'm a computer scientist and a statistician - and I say No.


## Confusing Terms

Menlo-


- We live in the Age of Buzzwords.
- Trust me: Data Science $=$ Statistics.
- But isn't Data Science $=$ Computer Science + Statistics?
- I'm a computer scientist and a statistician - and I say No.
- Statisticians have always had to be highly skilled with computers.

Careers in Data Science (You Know, Statistics)

## More Confusing Terms

## Norm Matloff

 University of California,Davis
Menlo-
Atherton High
School

Careers in Data Science (You Know, Statistics)

## More Confusing Terms

## Norm Matloff

 University of California,Davis
MenloAtherton High School

- Big Data:


## More Confusing Terms

Norm Matloff
University of
California,
Davis
MenloAtherton High School

- Big Data:
- Yes, have many huge data sets these days.


## More Confusing Terms

Norm Matloff University of
California,
Davis
MenloAtherton High School

- Big Data:
- Yes, have many huge data sets these days.
- Yes, typically requires parallel computation (one of my areas).


## More Confusing Terms

Norm Matloff University of
California,
Davis
MenloAtherton High School

- Big Data:
- Yes, have many huge data sets these days.
- Yes, typically requires parallel computation (one of my areas).
- But still, not really a new paradigm.


## More Confusing Terms

Norm Matloff University of California,

Davis
MenloAtherton High School

- Big Data:
- Yes, have many huge data sets these days.
- Yes, typically requires parallel computation (one of my areas).
- But still, not really a new paradigm.
- Machine Learning:


## More Confusing Terms

Norm Matloff University of California,

Davis
MenloAtherton High School

- Big Data:
- Yes, have many huge data sets these days.
- Yes, typically requires parallel computation (one of my areas).
- But still, not really a new paradigm.
- Machine Learning:
- Fancy new term for use of data for prediction.


## More Confusing Terms

Norm Matloff University of California,

Davis
MenloAtherton High School

- Big Data:
- Yes, have many huge data sets these days.
- Yes, typically requires parallel computation (one of my areas).
- But still, not really a new paradigm.
- Machine Learning:
- Fancy new term for use of data for prediction. Statisticians have been doing that since 1804, thank you.


## More Confusing Terms

Norm Matloff University of California,

Davis
MenloAtherton High School

- Big Data:
- Yes, have many huge data sets these days.
- Yes, typically requires parallel computation (one of my areas).
- But still, not really a new paradigm.
- Machine Learning:
- Fancy new term for use of data for prediction. Statisticians have been doing that since 1804, thank you.
- Methods either invented by statisticians (e.g. Random Forests) or statistically motivated.
Careers in
Data Science
(You Know,
Statistics)
Norm Matloff
University of
California,
Davis
Menlo-
Atherton High
School

Careers in Data Science (You Know, Statistics)

Norm Matloff University of California,

Davis
MenloAtherton High School

## How to Become a GOOD Data

 Scientist- I'll get to examples of Data Science shortly.

Careers in Data Science (You Know, Statistics)

Norm Matloff University of California,

Davis
MenloAtherton High School

## How to Become a GOOD Data

Scientist

- I'll get to examples of Data Science shortly. But first, what does one need to do Data Science well?


## How to Become a GOOD Data

 Scientist- I'll get to examples of Data Science shortly. But first, what does one need to do Data Science well?
- Study Stat vs. CS?


## How to Become a GOOD Data

 Scientist- I'll get to examples of Data Science shortly. But first, what does one need to do Data Science well?
- Study Stat vs. CS?


MenloAtherton High School

## How to Become a GOOD Data

 Scientist- I'll get to examples of Data Science shortly. But first, what does one need to do Data Science well?
- Study Stat vs. CS?

- A bit of tug-of-war between these two fields these days.


## How to Become a GOOD Data

 Scientist- I'll get to examples of Data Science shortly. But first, what does one need to do Data Science well?
- Study Stat vs. CS?

- A bit of tug-of-war between these two fields these days. See Statistics Losing Ground to Computer Science, N. Matloff, AMSTAT News, Nov. 2014,


## How to Become a GOOD Data

 Scientist- I'll get to examples of Data Science shortly. But first, what does one need to do Data Science well?
- Study Stat vs. CS?

- A bit of tug-of-war between these two fields these days. See Statistics Losing Ground to Computer Science, N. Matloff, AMSTAT News, Nov. 2014, and heated arguments on R vs. Python on Quora.


## How to Become a GOOD Data

 Scientist- I'll get to examples of Data Science shortly. But first, what does one need to do Data Science well?
- Study Stat vs. CS?

- A bit of tug-of-war between these two fields these days. See Statistics Losing Ground to Computer Science, N. Matloff, AMSTAT News, Nov. 2014, and heated arguments on R vs. Python on Quora.
- My bias is Stat, with lots of CS, at least a CS minor.


## How to Become a GOOD Data

 Scientist- I'll get to examples of Data Science shortly. But first, what does one need to do Data Science well?
- Study Stat vs. CS?

- A bit of tug-of-war between these two fields these days. See Statistics Losing Ground to Computer Science, N. Matloff, AMSTAT News, Nov. 2014, and heated arguments on R vs. Python on Quora.
- My bias is Stat, with lots of CS, at least a CS minor. Also, advanced linear algebra (matrix theory).

```
Careers in
Data Science
(You Know,
Statistics)
```


## Norm Matloff

``` University of California,
Davis
Menlo-
Atherton High
School
```


## What Really Counts

Careers in Data Science (You Know, Statistics)

Norm Matloff University of California,

Davis
MenloAtherton High School

## What Really Counts

- FAR more important than Stat vs. CS: Depth of insight, not rote memorization.

Careers in Data Science (You Know, Statistics)

Norm Matloff University of California,

Davis
MenloAtherton High

School

## What Really Counts

- FAR more important than Stat vs. CS: Depth of insight, not rote memorization. QUESTION THINGS!


## What Really Counts

Norm Matloff University of California,

Davis
MenloAtherton High School

- FAR more important than Stat vs. CS: Depth of insight, not rote memorization. QUESTION THINGS!
- E.g., if you've studied calculus: To minimize f, we set $\mathrm{f}^{\prime}=$ 0.


## What Really Counts

Norm Matloff University of California,

Davis
MenloAtherton High School

- FAR more important than Stat vs. CS: Depth of insight, not rote memorization. QUESTION THINGS!
- E.g., if you've studied calculus: To minimize $f$, we set $f^{\prime}=$ 0 . Then check $\mathrm{f}^{\prime \prime}>0$.


## What Really Counts

Norm Matloff University of California,

Davis
MenloAtherton High School

- FAR more important than Stat vs. CS: Depth of insight, not rote memorization. QUESTION THINGS!
- E.g., if you've studied calculus: To minimize $f$, we set $\mathrm{f}^{\prime}=$ 0 . Then check $\mathrm{f}^{\prime \prime}>0$. BUT DO YOU KNOW WHY?


## What Really Counts

MenloAtherton High School

- FAR more important than Stat vs. CS: Depth of insight, not rote memorization. QUESTION THINGS!
- E.g., if you've studied calculus: To minimize f, we set $\mathrm{f}^{\prime}=$ 0 . Then check $\mathrm{f}^{\prime \prime}>0$. BUT DO YOU KNOW WHY?
- In statistics, in computing sample variance, we divide by $n-1$, not $n$ :

$$
s^{2}=\frac{1}{n-1} \sum_{i=1}^{n}\left(X_{i}-\bar{X}\right)^{2}
$$

## What Really Counts

- FAR more important than Stat vs. CS: Depth of insight, not rote memorization. QUESTION THINGS!
- E.g., if you've studied calculus: To minimize f, we set $\mathrm{f}^{\prime}=$ 0 . Then check $\mathrm{f}^{\prime \prime}>0$. BUT DO YOU KNOW WHY?
- In statistics, in computing sample variance, we divide by $n-1$, not $n$ :

$$
s^{2}=\frac{1}{n-1} \sum_{i=1}^{n}\left(X_{i}-\bar{X}\right)^{2}
$$

Why not divide by $n$ ? IS THERE A GOOD REASON FOR THIS?

## What Really Counts

MenloAtherton High School

- FAR more important than Stat vs. CS: Depth of insight, not rote memorization. QUESTION THINGS!
- E.g., if you've studied calculus: To minimize f, we set $\mathrm{f}^{\prime}=$ 0 . Then check $\mathrm{f}^{\prime \prime}>0$. BUT DO YOU KNOW WHY?
- In statistics, in computing sample variance, we divide by $n-1$, not $n$ :

$$
s^{2}=\frac{1}{n-1} \sum_{i=1}^{n}\left(X_{i}-\bar{X}\right)^{2}
$$

Why not divide by $n$ ? IS THERE A GOOD REASON FOR THIS?

- Can you recognize Simpson's Paradox when you see it?

```
    Careers in
    Data Science
    (You Know,
    Statistics)
Norm Matloff
One More Slide on Prep for DS
Career
University of
    California,
        Davis
    Menlo-
Atherton High
    School
```

Careers in Data Science (You Know, Statistics)

Norm Matloff University of California,

Davis
MenloAtherton High School

## One More Slide on Prep for DS

Career

- GENERAL knowledge, awareness and insight are key!

Menlo-
School

## One More Slide on Prep for DS

Career

- GENERAL knowledge, awareness and insight are key!
- Do you know terms like ameliorate, morbidity and elasticity of demand?

MenloAtherton High

School

## One More Slide on Prep for DS

Career

- GENERAL knowledge, awareness and insight are key!
- Do you know terms like ameliorate, morbidity and elasticity of demand?
- How about FDA, HOV and CPI?

MenloAtherton High

School

## One More Slide on Prep for DS

Career

- GENERAL knowledge, awareness and insight are key!
- Do you know terms like ameliorate, morbidity and elasticity of demand?
- How about FDA, HOV and CPI?
- Can't be a good data miner without understanding the data!

MenloAtherton High

School

## One More Slide on Prep for DS

Career

- GENERAL knowledge, awareness and insight are key!
- Do you know terms like ameliorate, morbidity and elasticity of demand?
- How about FDA, HOV and CPI?
- Can't be a good data miner without understanding the data! Ptolemy's epicycles fiasco.

```
Careers in
Davis
Menlo-

\section*{All Right, Then, What Do DS People Do?}
- Example: Software running in a satellite notices a bright light in a forest.
- Example: Software running in a satellite notices a bright light in a forest. Is it a fire? Maybe just a reflection?

\section*{All Right, Then, What Do DS People Do?}

\section*{All Right, Then, What Do DS People Do?}
- Example: Software running in a satellite notices a bright light in a forest. Is it a fire? Maybe just a reflection? How can previous data be used here?

MenloAtherton High School

\section*{All Right, Then, What Do DS People Do?}
- Example: Software running in a satellite notices a bright light in a forest. Is it a fire? Maybe just a reflection? How can previous data be used here?
- Example: You, humans, can spot the speaker:


MenloAtherton High School

\section*{All Right, Then, What Do DS People Do?}
- Example: Software running in a satellite notices a bright light in a forest. Is it a fire? Maybe just a reflection? How can previous data be used here?
- Example: You, humans, can spot the speaker:


But can software spot me?

MenloAtherton High School

\section*{All Right, Then, What Do DS People Do?}
- Example: Software running in a satellite notices a bright light in a forest. Is it a fire? Maybe just a reflection? How can previous data be used here?
- Example: You, humans, can spot the speaker:


But can software spot me?
```

Careers in
Data Science
(You Know,
Statistics)

```

\section*{Norm Matloff}
``` University of California,
Davis
Menlo-
Atherton High
School
```


## Detailed Example

Careers in Data Science (You Know, Statistics)

## Detailed Example

## Norm Matloff

 University of California,Davis
MenloAtherton High School

## Goals:

Careers in Data Science (You Know, Statistics)

## Detailed Example

Norm Matloff University of California,

Davis
MenloAtherton High School

Goals:

- Show you something different from AP Stat.

Careers in

## Detailed Example

## Norm Matloff

 University ofCalifornia,
Davis
MenloAtherton High School

Goals:

- Show you something different from AP Stat.
- Show that serious math is involved (calculus, matrix theory).


## Detailed Example

## Norm Matloff

 University of California,Davis
MenloAtherton High School

Goals:

- Show you something different from AP Stat.
- Show that serious math is involved (calculus, matrix theory).
- This will get a little technical; don't feel that you need to follow $100 \%$.


## Detailed Example

## Norm Matloff

 University of California,Davis
MenloAtherton High School

Goals:

- Show you something different from AP Stat.
- Show that serious math is involved (calculus, matrix theory).
- This will get a little technical; don't feel that you need to follow 100\%.
- The Question: Will Mary like the movie Captain America?

Careers in Data Science (You Know, Statistics)

Norm Matloff University of California,

Davis
MenloAtherton High School

Will Mary Like the Movie Captain America?

Careers in Data Science (You Know, Statistics)

Norm Matloff University of California,

Davis
MenloAtherton High School

## Will Mary Like the Movie Captain

 America?- Mary hasn't seen the movie.

MenloAtherton High School

## Will Mary Like the Movie Captain

 America?- Mary hasn't seen the movie.
- But we have Mary's ratings on some other movies, and we have ratings of Captain America by some other people.

MenloAtherton High

School

## Will Mary Like the Movie Captain

 America?- Mary hasn't seen the movie.
- But we have Mary's ratings on some other movies, and we have ratings of Captain America by some other people.
- How do we use this data to guess Mary's rating of this movie?

Careers in Data Science (You Know, Statistics)

Norm Matloff University of California,

Davis
Menlo-
Atherton High
School

## Neighborhood-Based Approach

## Neighborhood-Based Approach

## Norm Matloff

 University of California,Davis
MenloAtherton High School

- Find people ("neighbors") in our dataset whose movie tastes are similar to Mary's.


## Neighborhood-Based Approach

## Norm Matloff

 University of California,Davis
MenloAtherton High School

- Find people ("neighbors") in our dataset whose movie tastes are similar to Mary's.
- Of those, focus on the ones that have seen Captain America.


## Neighborhood-Based Approach

Norm Matloff University of California,

Davis
MenloAtherton High School

- Find people ("neighbors") in our dataset whose movie tastes are similar to Mary's.
- Of those, focus on the ones that have seen Captain America.
- Guess Mary's rating of the movie to be the mean of the ratings in that group.

```
Careers in
Data Science
(You Know,
Statistics)
```


## Norm Matloff

``` University of California,
Davis
Menlo-
Atherton High
School
```


## A More Nuanced Model

## A More Nuanced Model

Norm Matloff University of California,

Davis

- $Y_{i j}$ is user i's rating of movie $j$.

MenloAtherton High School

Careers in Data Science (You Know, Statistics)

## A More Nuanced Model

Norm Matloff University of California,

Davis
MenloAtherton High School

- $Y_{i j}$ is user i's rating of movie j .
- We want to predict $Y_{\text {Mary, Cpt.Am. }}$.

Careers in Data Science (You Know, Statistics)

## A More Nuanced Model

Norm Matloff
University of California,

Davis
MenloAtherton High School

- $Y_{i j}$ is user i's rating of movie j .
- We want to predict $Y_{\text {Mary, Cpt.Am. }}$.
- Latent Factor Model:


## A More Nuanced Model

Norm Matloff
University of California,

Davis
MenloAtherton High School

- $Y_{i j}$ is user i's rating of movie $j$.
- We want to predict $Y_{\text {Mary, Cpt.Am. }}$.
- Latent Factor Model:

$$
Y_{i j}=\mu+\alpha_{i}+\beta_{j}+\epsilon_{i j}
$$

where

- $\mu=$ mean ratings over all users and all movies.


## A More Nuanced Model

Norm Matloff University of California,

Davis
MenloAtherton High School

- $Y_{i j}$ is user i's rating of movie j .
- We want to predict $Y_{\text {Mary, Cpt.Am. }}$.
- Latent Factor Model:

$$
Y_{i j}=\mu+\alpha_{i}+\beta_{j}+\epsilon_{i j}
$$

where

- $\mu=$ mean ratings over all users and all movies.
- $\alpha_{i}=$ tendency for user i to give higher/lower ratings than the typical user


## A More Nuanced Model

Norm Matloff University of California, Davis

MenloAtherton High School

- $Y_{i j}$ is user i's rating of movie j .
- We want to predict $Y_{\text {Mary, Cpt.Am. }}$.
- Latent Factor Model:

$$
Y_{i j}=\mu+\alpha_{i}+\beta_{j}+\epsilon_{i j}
$$

where

- $\mu=$ mean ratings over all users and all movies.
- $\alpha_{i}=$ tendency for user i to give higher/lower ratings than the typical user
- $\beta_{j}=$ tendency for movie j to be rated higher/lower than the typical movie


## A More Nuanced Model

MenloAtherton High School

- $Y_{i j}$ is user i's rating of movie j .
- We want to predict $Y_{\text {Mary, Cpt.Am. }}$.
- Latent Factor Model:

$$
Y_{i j}=\mu+\alpha_{i}+\beta_{j}+\epsilon_{i j}
$$

where

- $\mu=$ mean ratings over all users and all movies.
- $\alpha_{i}=$ tendency for user i to give higher/lower ratings than the typical user
- $\beta_{j}=$ tendency for movie j to be rated higher/lower than the typical movie
- $\epsilon_{i j}=$ sum of all unknown effects, e.g. user i's mood when viewing movie j

Careers in Data Science (You Know, Statistics)

## Where Does the Data Come in?

## Norm Matloff

 University of California,Davis
Menlo-
Atherton High
School

Careers in Data Science (You Know, Statistics)

Norm Matloff University of California,

Davis
MenloAtherton High School

## Where Does the Data Come in?

So, we really have a statistical estimation problem:

MenloAtherton High School

## Where Does the Data Come in?

So, we really have a statistical estimation problem:

- We'll use our data to estimate $\mu, \alpha_{\text {Mary }}$ and $\beta_{C p t . A m}$.


## Where Does the Data Come in?

## Norm Matloff

 University of California,Davis
MenloAtherton High School

So, we really have a statistical estimation problem:

- We'll use our data to estimate $\mu, \alpha_{\text {Mary }}$ and $\beta_{\text {Cpt.Am. }}$. This gives us (stat. notation) $\widehat{\mu}$ etc.


## Where Does the Data Come in?

MenloAtherton High School

So, we really have a statistical estimation problem:

- We'll use our data to estimate $\mu, \alpha_{\text {Mary }}$ and $\beta_{C p t . A m .}$. This gives us (stat. notation) $\widehat{\mu}$ etc.
- We then guess Mary's rating of Cpt. Am. to be

$$
\widehat{\mu}+\widehat{\alpha}_{M a r y}+\widehat{\beta}_{C p t . A m}
$$

## Where Does the Data Come in?

MenloAtherton High School

So, we really have a statistical estimation problem:

- We'll use our data to estimate $\mu, \alpha_{\text {Mary }}$ and $\beta_{C p t . A m .}$. This gives us (stat. notation) $\widehat{\mu}$ etc.
- We then guess Mary's rating of Cpt. Am. to be

$$
\widehat{\mu}+\widehat{\alpha}_{M a r y}+\widehat{\beta}_{C p t . A m .}
$$

- But HOW will we get those estimates?

```
Careers in
Data Science
(You Know,
Statistics)
```


## Norm Matloff

``` University of California,
Davis
Menlo-
Atherton High
School
```


## One Common Method

Careers in Data Science (You Know, Statistics)

Norm Matloff University of California,

Davis
MenloAtherton High School

## One Common Method

A popular way to obtain those estimates is matrix factorization.

Norm Matloff University of California,

Davis
MenloAtherton High

School

## One Common Method

A popular way to obtain those estimates is matrix factorization.

- Define $A$ to be the matrix of the ratings - even the unknown ones.

Norm Matloff University of California,

Davis
MenloAtherton High School

## One Common Method

A popular way to obtain those estimates is matrix factorization.

- Define $A$ to be the matrix of the ratings - even the unknown ones.
- Mary has a row in the matrix.

Norm Matloff University of California,

Davis
MenloAtherton High School

## One Common Method

A popular way to obtain those estimates is matrix factorization.

- Define $A$ to be the matrix of the ratings - even the unknown ones.
- Mary has a row in the matrix.
- Captain America has a column in the matrix.

Norm Matloff University of California,

Davis
MenloAtherton High School

## One Common Method

A popular way to obtain those estimates is matrix factorization.

- Define $A$ to be the matrix of the ratings - even the unknown ones.
- Mary has a row in the matrix.
- Captain America has a column in the matrix.
- Unfortunately, the entry in Mary's row and Captain America's column is unknown.


## One Common Method

A popular way to obtain those estimates is matrix factorization.

- Define $A$ to be the matrix of the ratings - even the unknown ones.
- Mary has a row in the matrix.
- Captain America has a column in the matrix.
- Unfortunately, the entry in Mary's row and Captain America's column is unknown.
- But some other entries in Mary's row are known. Same for Captain America's column.


## One Common Method

A popular way to obtain those estimates is matrix factorization.

- Define $A$ to be the matrix of the ratings - even the unknown ones.
- Mary has a row in the matrix.
- Captain America has a column in the matrix.
- Unfortunately, the entry in Mary's row and Captain America's column is unknown.
- But some other entries in Mary's row are known. Same for Captain America's column.
- Can use calculus, matrix theory to estimate the missing entries in the matrix.


## One Common Method

A popular way to obtain those estimates is matrix factorization.

- Define $A$ to be the matrix of the ratings - even the unknown ones.
- Mary has a row in the matrix.
- Captain America has a column in the matrix.
- Unfortunately, the entry in Mary's row and Captain America's column is unknown.
- But some other entries in Mary's row are known. Same for Captain America's column.
- Can use calculus, matrix theory to estimate the missing entries in the matrix. Gives us matrices $P$ and $Q$ such that

$$
A \approx P Q
$$

## One Common Method

A popular way to obtain those estimates is matrix factorization.

- Define $A$ to be the matrix of the ratings - even the unknown ones.
- Mary has a row in the matrix.
- Captain America has a column in the matrix.
- Unfortunately, the entry in Mary's row and Captain America's column is unknown.
- But some other entries in Mary's row are known. Same for Captain America's column.
- Can use calculus, matrix theory to estimate the missing entries in the matrix. Gives us matrices $P$ and $Q$ such that

$$
A \approx P Q
$$

Details not shown. :-)

```
Careers in
Data Science
(You Know,
Statistics)
Norm Matloff University of California,
Davis
Menlo-
Atherton High
School
```


## Summary

## Summary

Norm Matloff
University of California,

Davis
MenloAtherton High School

- Beware of the buzzwords like data science. Modern methodology is not really new.


## Summary

Norm Matloff
University of
California,
Davis
MenloAtherton High

School

- Beware of the buzzwords like data science. Modern methodology is not really new.
- But...things ARE different today.


## Summary

Norm Matloff University of California,

Davis
MenloAtherton High School

- Beware of the buzzwords like data science. Modern methodology is not really new.
- But...things ARE different today. The applications are highly engaging, and we have powerful computers to handle the large volume of data.


## Summary

- Beware of the buzzwords like data science. Modern methodology is not really new.
- But...things ARE different today. The applications are highly engaging, and we have powerful computers to handle the large volume of data.
- Don't fall into the trap of thinking that taking Course $X$ or studying Major Y is sufficient!


## Summary

- Beware of the buzzwords like data science. Modern methodology is not really new.
- But...things ARE different today. The applications are highly engaging, and we have powerful computers to handle the large volume of data.
- Don't fall into the trap of thinking that taking Course $X$ or studying Major Y is sufficient!
- You need to be strong in math - this means INSIGHT, not just facility with equations - and be very AWARE of the world around you.


## Summary

- Beware of the buzzwords like data science. Modern methodology is not really new.
- But...things ARE different today. The applications are highly engaging, and we have powerful computers to handle the large volume of data.
- Don't fall into the trap of thinking that taking Course $X$ or studying Major Y is sufficient!
- You need to be strong in math - this means INSIGHT, not just facility with equations - and be very AWARE of the world around you.
- Good luck to you!

