What You Can Do

Nick Eubank

Modeling and Representation of Data

Develop Models to Faithfully Represent Data

Modeling and Representation of Data

Develop Models to Faithfully Represent Data

Practical Data Science

Wrestle Data into Workable Forms

Modeling and Representation of Data

Develop Models to Faithfully Represent Data

Practical Data Science

Wrestle Data into Workable Forms

Natural Language Processing

Convert Text into Machine Interpretable Data

Modeling and Representation of Data

Develop Models to Faithfully Represent Data

Practical Data Science

Wrestle Data into Workable Forms

Natural Language Processing

Convert Text into Machine Interpretable Data

Data Engineering

Learned Docker

Modeling and Representation of Data

Develop Models to Faithfully Represent Data

Practical Data Science

Wrestle Data into Workable Forms

Natural Language Processing

Convert Text into Machine Interpretable Data

Data Engineering

Learned Docker

Machine Learning

Learned the Fit and Evaluate Machine Learning Models

Modeling and Representation of Data

Develop Models to Faithfully Represent Data

Practical Data Science

Wrestle Data into Workable Forms

Natural Language Processing

Convert Text into Machine Interpretable Data

Data Engineering

Learned Docker

Machine Learning

Learned the Fit and Evaluate Machine Learning Models

$\Rightarrow$ Your toolboxes are now phenomenally powerful.

But we don't do data science to accumulate tools (as much fun as they are).

But we don't do data science to accumulate tools (as much fun as they are).

We do data science to solve problems.

This semester, we've learned to:

This semester, we've learned to:

- Identify core problems,

This semester, we've learned to:

- Identify core problems,
    - (managing and working with stakeholders)

This semester, we've learned to:

- Identify core problems,
    - (managing and working with stakeholders)
- Articulate and refine a question to answer,

This semester, we've learned to:

- Identify core problems,
    - (managing and working with stakeholders)
- Articulate and refine a question to answer,
- Recognize the purpose of different types of questions,

This semester, we've learned to:

- Identify core problems,
  - (managing and working with stakeholders)
- Articulate and refine a question to answer,
- Recognize the purpose of different types of questions,
- And then use the appropriate amazing tool to generate an answer.

I recognize that at the moment, parts of this process may feel... unnecessary or superfluous.

I recognize that at the moment, parts of this process may feel... unnecessary or superfluous.

$\Rightarrow$ As you move into your capstones, and eventually into the real world, you'll find that proficiency in these skills is what determines the success (or failure) of your projects.

I recognize that at the moment, parts of this process may feel... unnecessary or superfluous.

$\Rightarrow$ As you move into your capstones, and eventually into the real world, you'll find that proficiency in these skills is what determines the success (or failure) of your projects.

$\Rightarrow$ These are also the critical reasoning skills that won't be supplanted by chatGPT any time soon.

Causal Inference

Causal Inference

"Correlation does not imply causation"

Causal Inference

~~"Correlation does not imply causation"~~

Correlation does not necessarily imply causation, but...

Causal Inference

~~"Correlation does not imply causation"~~

Correlation does not necessarily imply causation, but...

- when certain assumptions are met, correlation does imply causation.

Causal Inference

~~"Correlation does not imply causation"~~

Correlation does not necessarily imply causation, but...

- when certain assumptions are met, correlation does imply causation.

And now you know those assumptions and how to evaluate them!

Ways of Estimating Causal Effects

Ways of Estimating Causal Effects

1. Randomized Control Trials

   Law of large numbers $\rightarrow$ same potential outcomes for C & T

Ways of Estimating Causal Effects

1. Randomized Control Trials

   Law of large numbers $\rightarrow$ same potential outcomes for C & T

2. Regression

   Statistically adjust for baseline differences $\rightarrow$ same potential outcomes after adjustments

Ways of Estimating Causal Effects

1. Randomized Control Trials

   Law of large numbers $\rightarrow$ same potential outcomes for C & T

2. Regression

   Statistically adjust for baseline differences $\rightarrow$ same potential outcomes after adjustments

3. Matching

   Statistically adjust for baseline differences $\rightarrow$ same potential outcomes after adjustments

Ways of Estimating Causal Effects

1. Randomized Control Trials
   Law of large numbers $\rightarrow$ same potential outcomes for C & T
2. Regression
   Statistically adjust for baseline differences $\rightarrow$ same potential outcomes after adjustments
3. Matching
   Statistically adjust for baseline differences $\rightarrow$ same potential outcomes after adjustments
4. Differences-in-Differences
   Adjust for pre-existing baseline differences $\rightarrow$ same potential outcomes in trends

Cultivate a Skeptical Mindset

Cultivate a Skeptical Mindset

There are no statistical tests that can tell you if a causal estimate is valid.

Cultivate a Skeptical Mindset

There are no statistical tests that can tell you if a causal estimate is valid.

- Depending on how data was collected, the exact same dataset may or may not provide valid causal estimates.

You haven't just learned about these...

You haven't just learned about these...

You've done them!

You haven't just learned about these...

You've done them!

- Used randomization and A/B test data to measure effect of website design changes and racial discrimination in hiring.

You haven't just learned about these...

You've done them!

- Used randomization and A/B test data to measure effect of website design changes and racial discrimination in hiring.
- Used regression to estimate gender wage discrimination in the US

You haven't just learned about these...

You've done them!

- Used randomization and A/B test data to measure effect of website design changes and racial discrimination in hiring.
- Used regression to estimate gender wage discrimination in the US
- Used fixed effects to understand tax policy's impact on car accidents

You haven't just learned about these...

You've done them!

- Used randomization and A/B test data to measure effect of website design changes and racial discrimination in hiring.
- Used regression to estimate gender wage discrimination in the US
- Used fixed effects to understand tax policy's impact on car accidents
- Used matching to estimate labor market returns to education

You haven't just learned about these...

You've done them!

- Used randomization and A/B test data to measure effect of website design changes and racial discrimination in hiring.
- Used regression to estimate gender wage discrimination in the US
- Used fixed effects to understand tax policy's impact on car accidents
- Used matching to estimate labor market returns to education
- Used diff-in-diffs to understand impact of drug legalization on crime

You haven't just learned about these...

You've done them!

You haven't just learned about these...

You've done them!

In your projects you're studying the effects of:

- stay at home orders on the spread of covid,
- fiscal incentives on policing,
- open carry laws on violent crime,
- investment in housing infrastructure on homelessness,
- childhood trauma on depression,
- remote learning on student mental health,
- the death penalty on violent crime,
- the pandemic on student learning,
- patent protections on drug price growth, and
- same sex marriage laws on hate crimes.

You haven't just learned about these...

You've done them!

In your projects you're studying the effects of:

- stay at home orders on the spread of covid,
- fiscal incentives on policing,
- open carry laws on violent crime,
- investment in housing infrastructure on homelessness,
- childhood trauma on depression,
- remote learning on student mental health,
- the death penalty on violent crime,
- the pandemic on student learning,
- patent protections on drug price growth, and
- same sex marriage laws on hate crimes.

And you did it all while learning to be more effective colleagues.

You haven't just learned about these...

You've done them!

In your projects you're studying the effects of:

- stay at home orders on the spread of covid,
- fiscal incentives on policing,
- open carry laws on violent crime,
- investment in housing infrastructure on homelessness,
- childhood trauma on depression,
- remote learning on student mental health,
- the death penalty on violent crime,
- the pandemic on student learning,
- patent protections on drug price growth, and
- same sex marriage laws on hate crimes.

And you did it all while learning to be more effective colleagues.

And while doing Kyle's class.

erika
@brownbearika

Maybe...just maybe.... the real treasure...was the friends we made along the way... (end scene)

I hope you feel really proud of everything you've accomplished.