

# A Glimpse into Data Science in Industry

NCSSM Data Science Summer Institute  
Summer Research Programs July 14, 2023

Dr. Rachel Levy (Ray)  
Executive Director, NC State Data Science Academy

# Data science: turning information into insight

## Let's start with a game.

With your neighbors, come up with three places your students might work in the future. (you have about 2 minutes)

Next, I will call on some groups to suggest a place.

Then I'll share some ways that data science plays a role in that workplace.

**Take home message:  
Every workplace has data.**

## More and more, workplaces...

- Want to collect the data they need to gain insight.
- Hire people who use data to make predictions or decisions.
- Need to manage their data (organization, security, privacy).
- Consider sharing their data so it can be more useful.
- Incorporate the use of data into their budget and strategy.

**Connecting with Industry has been a recurring  
theme in my career...**

# Carolina Friends School – 4 yrs MS 4 yrs HS + Dean



1990's – how can new tools like graphing calculators allow us to teach math by asking questions, exploring concepts and assessing learning in new ways?

**Fast forward: same  
questions for data science.**

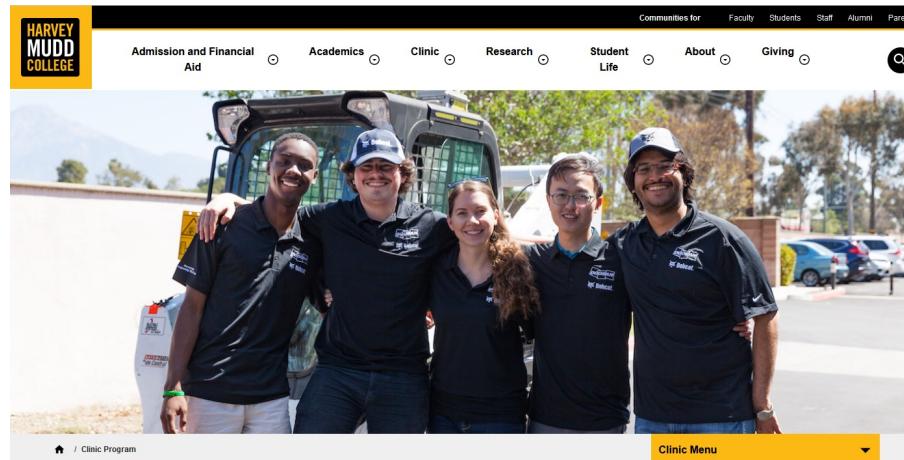


**I had a Master's Degree in Educational Media  
and Instructional Design and 2 kids.  
Why get a MS & PhD in Applied Math?  
The credentials opened many doors.**



Student research experience at  
NASA put the idea in my head  
to learn more...

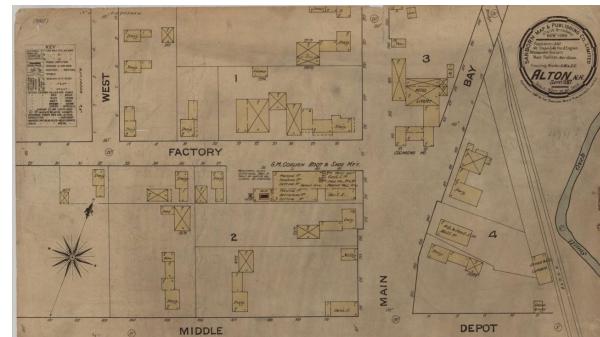
# Harvey Mudd College – prof for 12 years and Associate Dean for Faculty Development



## Clinic Program



# Student Clinic Projects at Harvey Mudd College



# Non-profits: Professional Societies

**siam.**  
Society for Industrial and Applied Mathematics

Input your search...

Become a Member  Login  Get Involved

Home Publications Research Areas Conferences Careers Students & Education Membership Prizes & Recognition

Students & Education Programs & Initiatives Thinking of a Career in the Mathematical Sciences? Profiles Rachel Levy

Rachel Levy | Professor of Mathematics and Associate Dean for Faculty Development

Print

---

  
Harvey Mudd College  
Claremont, California

**Department:** Mathematics  
**Education:** B.A. English and Mathematics, 1989, Oberlin College; M.A. Education Media and Instructional Design, 1996, University of North Carolina at Chapel Hill; M.S. Applied Mathematics, 2003, and Ph.D. Applied Mathematics, 2005, North Carolina State University

**Career stage:** Late—28 years post Bachelor's



Vice President for Education

Deputy Executive Director

## Non-profit data challenges

- Who can we recruit as members/participants?
- How can we prioritize work/expenditures?
- How can we demonstrate impact?
- How can we identify potential donors / funding?

# Government: AAAS Policy Fellow for US Senator Maggie Hassan – New Hampshire



The logo for ORMS Today, which is the membership magazine of INFORMS. The word "ORMS" is in large white letters, with "TODAY" in smaller letters to its right. To the right of "ORMS" is the INFORMS logo, which consists of the word "informs" in a lowercase sans-serif font with a registered trademark symbol, and a stylized arrow pointing upwards and to the right. Below the main title are navigation links: NEWS, FEATURES, PODCASTS, and DEPARTMENTS.

August 8, 2022 in STEM Education

## Historic U.S. legislation to modernize STEM education

*INFORMS supports the passage of the Mathematical and Statistical Modeling Education Act in the U.S. House of Representatives*

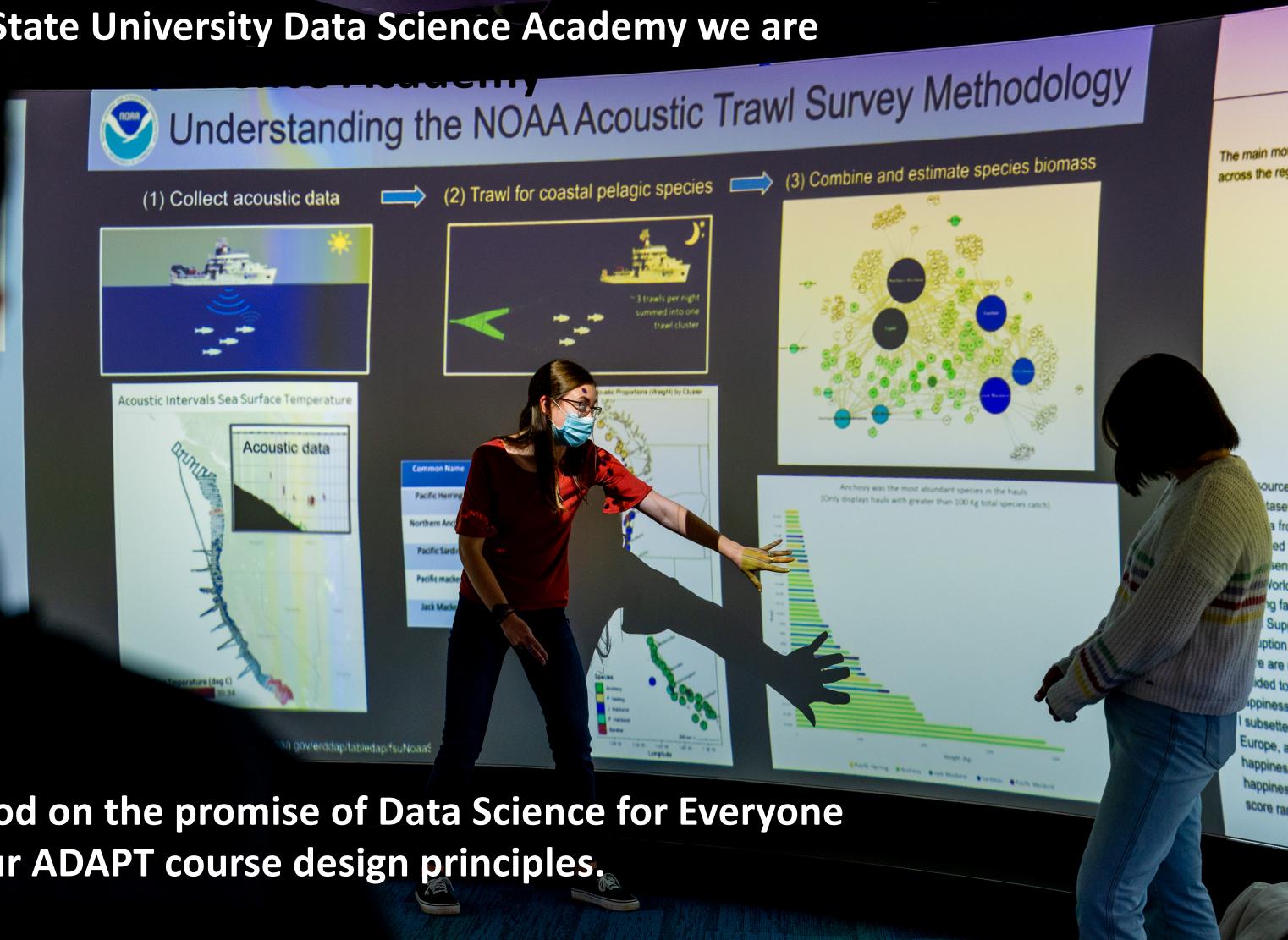
SHARE: [f](#) [in](#) [t](#) [e](#) [m](#) PRINT ARTICLE: [p](#) <https://doi.org/10.1287/orms.2022.04.24n>



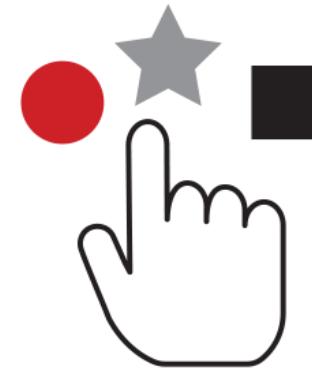
# Government data challenges

- Data from many sources – how to gather and combine
- Large systems that need updating but could break if you only change one part
- Estimating the cost of proposed legislation (and arguing why it should or should not pass)
- Data security, privacy, sharing as well as modeling

At the NC State University Data Science Academy we are



making good on the promise of Data Science for Everyone  
through our ADAPT course design principles.



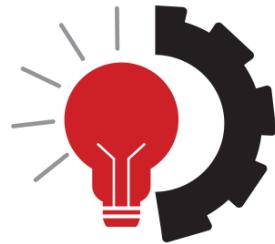
Project-based Learning | 10 Commons Learning Elements | Identity-conscious Choices

# ADAPT

All-campus Data science through Accessible Project-based Teaching and learning model

## We invite you to test the ADAPT model and contribute to its development!

- You can pick just one idea
- Let us know you are trying it
- Report how it goes
- We are happy to provide support and encouragement



## Project-based teaching and learning

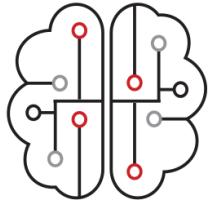
- No quizzes, tests or exams
- Homework and projects
- 1-credit courses
- Workshop model (like writing workshop)



# 10 Common Learning Elements

## Data Perspectives

1. Recognizing **data as information - not truth** - with error, variability, and degrees of inclusion/exclusion,
2. Explaining **what it means to be a data scientist** and data-enabled,
3. Observing a **diverse collection of data scientist role models and careers**,



# 10 Common Learning Elements

## Data Practices

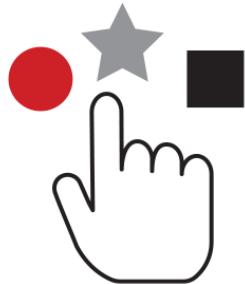
4. Examining **how data are created**, and the related assumptions and collection practices,
5. Practicing **data curation, wrangling, and cleaning**,
6. **Assessing validity** of data, methods, results, and communication,
7. Employing **accessibility practices** (e.g., alt text, color choices, GUIs, sonification, visualization, commenting, captioning),
8. Investigating **ethical issues** and ways to approach them,



## 10 Common Learning Elements

### Data Discoveries

9. Articulating **current issues or open questions** in data science,  
and
10. Specifying **exciting discoveries or impacts** of data science.



## Identity-conscious choices

- Context of the project
- Data set
- Which question to ask about a large data set
- Which tool to use
- Who/what the solution is optimized to help
- How to communicate solutions

# ADAPT

All-campus Data science through Accessible Project-based Teaching and learning model

- In 2 years 800 units delivered
- 24 sections scheduled for Fall 2023
- Over 100 majors participated
- All 10 colleges of the university
- Undergraduates, graduate students, faculty, staff as learners
- Demographics similar to that of university

**The Data Science Academy courses include**

**mathematical and statistical modeling  
data science  
artificial intelligence...**

**How are these topics related and different?**

**Let's start with**  
**mathematical and statistical modeling**  
**data science**  
**artificial intelligence**

**Many modelers have favorite models of modeling.**

**(Do you?) Here's a simple one.**

- Objective (hypothesis, question)
- Constraints
- Data
- Model
- Solution
- Check: is it a useful solution?  
Iterate as necessary.

**Let's discuss this modeling question:  
How much toilet paper should I buy?**

## How much toilet paper should I buy?

**Answer: It depends on the CONTEXT.**

- Why am I buying toilet paper? For whom?
- Does it matter what kind of toilet paper I buy?
- How much money do I have?
- How much storage room do I have?
- How much do I need?
- How does the price change if I buy more?
- Which is worse, to buy too little or too much?
- Do I need a flexible answer that works for different situations?



<https://www.gettyimages.com/photos/funny-toilet-paper>

# Simple model for toilet paper purchasing

- **Objective** – not run out.
- **Constraints** – budget, space, time for shopping.
- **Data** – seems like we use about 1 roll every 5 days (unless grandkids are visiting in which case it is one roll per waking hour) and we go shopping once a week, pack I buy has 12 rolls.
- **Model** –  $5 \text{ days/roll} * 12 \text{ rolls/pack} = 60 \text{ days/pack}$   
 $60 \text{ days/pack} * 1 \text{ month} / 30 \text{ days} = 2 \text{ months/pack}$
- **Solution** - buy 1 pack every other month. Keep an extra pack as backup at all times. **Test/Check:** is it a useful solution?



Modeling question.

## How much toilet paper should I buy?

**Key point: it depends on who the answer serves**

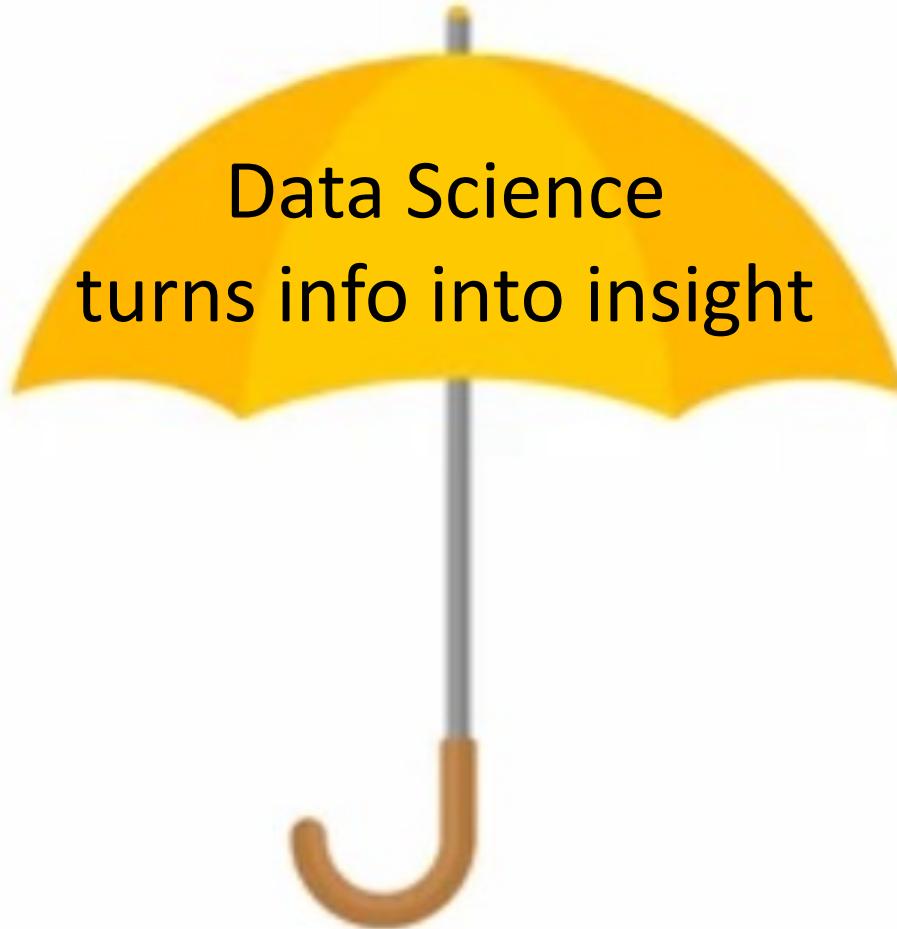
- Who will use the answer? For what purpose?
- How general does the solution have to be?  
Are you answering for 1 client/person or creating a recommendation system that can work for anyone?
- How certain is this answer? We expect 80 guests +/- 5 guests.
- Is the answer a range? I sleep 7-10 hours a night.
- **Is this kind of modeling part of data science? Short answer:**  
...



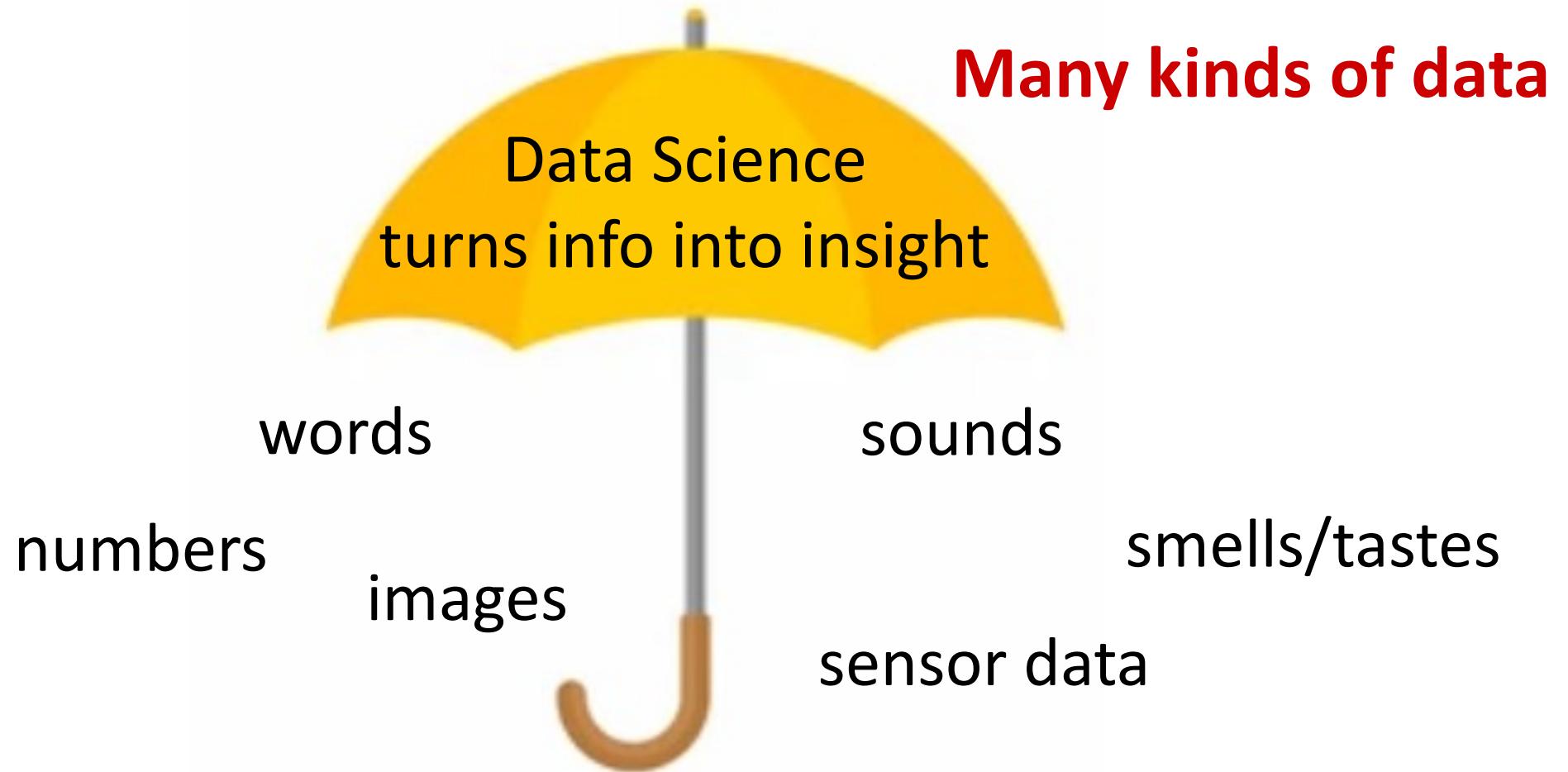
## In the workplace, there's always a client (or multiple stakeholders) and objective(s)

- The answer might be a range of numbers
- The answer might have uncertainty (uncertainty quantification is a huge field)
- The answer might need to be flexible – circumstances change
- Modular solutions might have re-useable parts

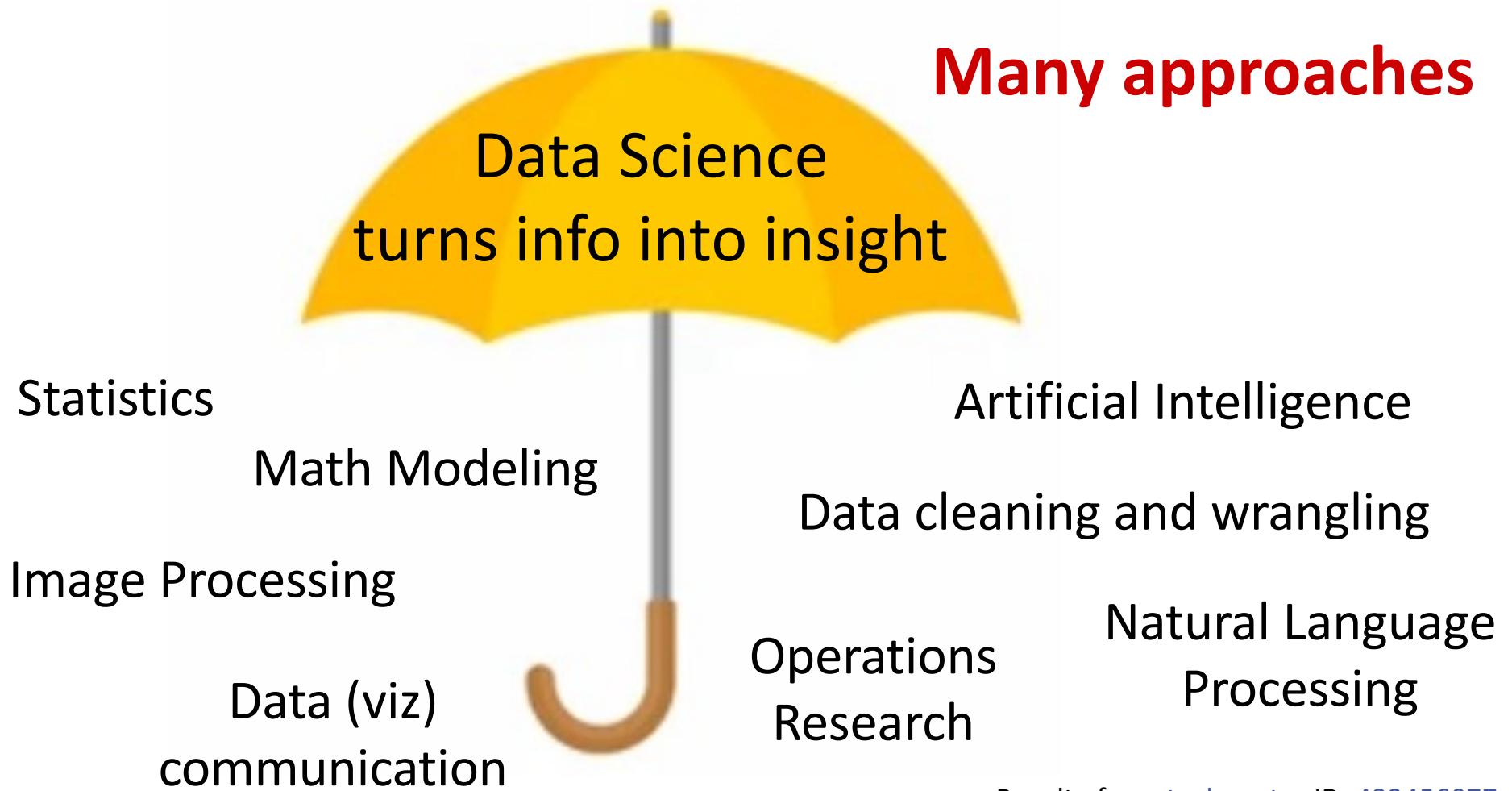
**Now let's consider**  
**mathematical and statistical modeling**  
**data science**  
**artificial intelligence**



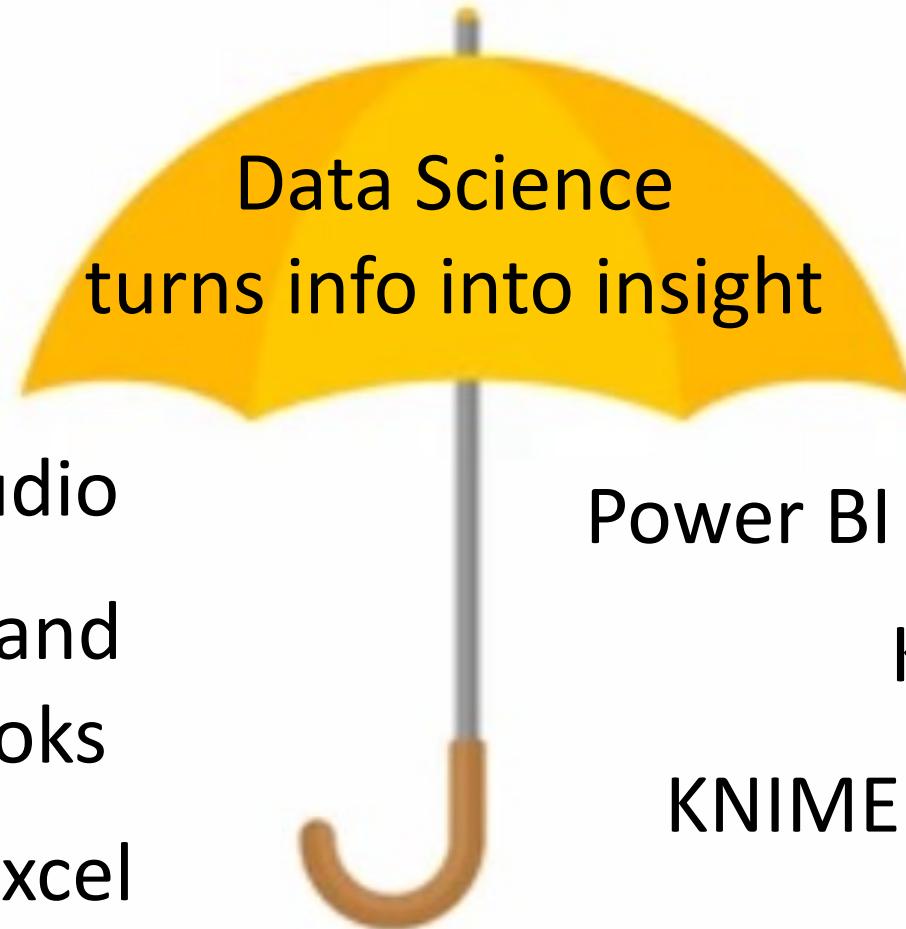
Royalty-free [stock vector](#) ID: [422456077](#)



Royalty-free [stock vector](#) ID: [422456077](#)

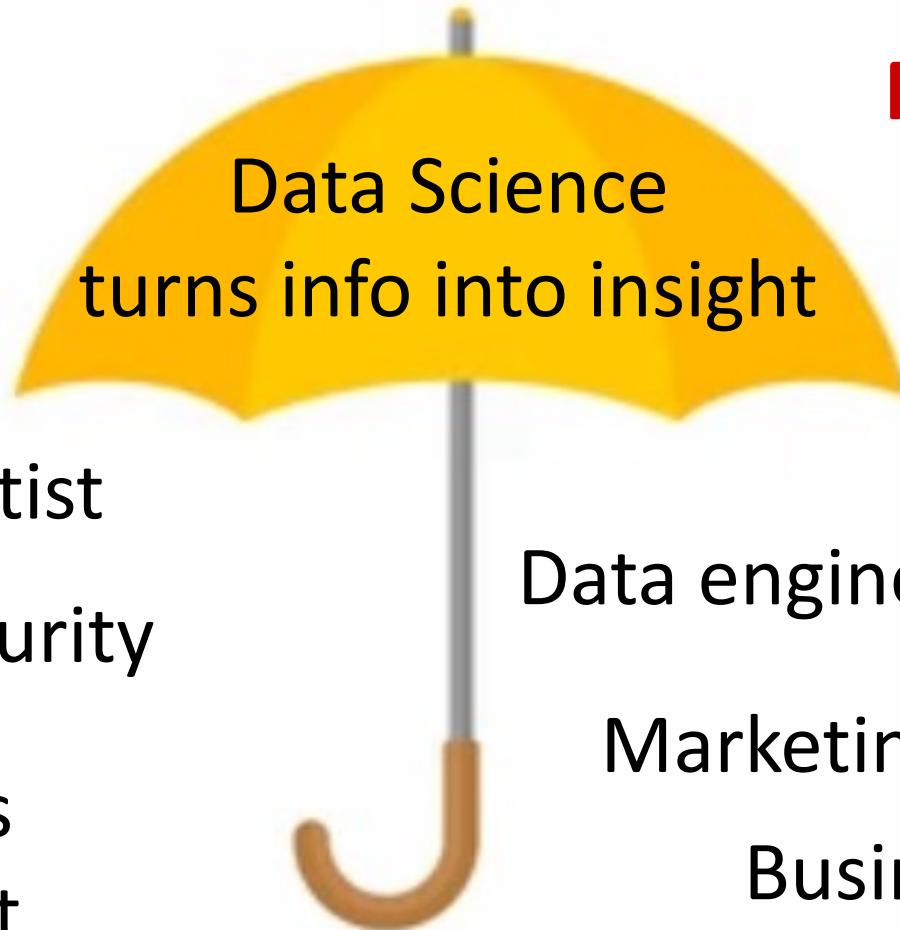


R and RStudio  
Python and  
notebooks  
SQL      Excel



**Many tools**  
**(these are examples)**

Power BI      SAS/JMP  
KNIME      Tableau  
Kubernetes



Data scientist  
Cybersecurity  
Clinical trials  
Data support

**Many job titles**

Data analyst  
Data engineer  
Marketing/Communication  
Business intelligence



### [Senior Director, Data Science \(R-14726\)](#)

Dun & Bradstreet  
Jacksonville, FL (Hybrid)  
Vision, 401(k)

1 company alum works here

Promoted



### [Adjunct Instructor for Data Science Courses](#)

San Francisco Bay University  
Fremont, CA (On-site)

Message the job poster directly

Promoted · Easy Apply



### [Data Science Architect](#)

SAIC  
Chantilly, VA (On-site)  
 5 company alumni work here

Promoted · 2 applicants



### [Data Technical Analyst](#)

Resolvit  
Cincinnati, OH (Hybrid)  
\$110/yr - \$120/yr · 401(k) benefit  
 Message the job poster directly

Promoted · Easy Apply



### [Data Science Intern](#)

Brown-Forman  
Louisville, KY (Hybrid)  
Medical, 401(k)

Actively recruiting



### [Data Science Platform Engineering Manager](#)

Dice  
Rockville, MD

Promoted · 5 applicants



### [Principal Data Science \(multiple openings\) - IHM](#)

Discover Financial Services  
Riverwoods, IL (On-site)  
\$146K/yr - \$204.8K/yr (from job description)

Actively recruiting

Promoted · 6 applicants



### [Principal Machine Learning Engineer](#)

Fidelity Investments  
Boston, MA (Hybrid)

1 connection works here

Promoted · 7 applicants



### [A&AMPLIFY Data Science Manager](#)

Alvarez & Marsal  
Chicago, IL  
\$75K/yr - \$135K/yr (from job description)

Actively recruiting

Promoted



### [Market Strategist, Light Ends](#)

bp  
Chicago, IL  
 3 company alumni work here

Promoted · 13 applicants



### [Data Analyst](#)

Meta  
United States (Remote)  
\$101K/yr - \$143K/yr

187 company alumni work here

12 minutes ago



### [Lead Data Scientist - TS/SCI](#)

Harnham  
Washington DC-Baltimore Area (On-site)  
\$170K/yr - \$190K/yr

Message the job poster directly

Promoted · Easy Apply



### [Principal Data Scientist](#)

NextGen Federal Systems  
Aberdeen Proving Ground, MD (Hybrid)  
 Your profile matches this job

Promoted · Easy Apply



### [Decision Science Manager](#)

tms  
Greater Chicago Area (Hybrid)  
 Message the job poster directly

Promoted · Easy Apply



### [Business Systems Analyst](#)

TD SYNNEX  
Phoenix, AZ  
 2 company alumni work here

Promoted · 22 applicants

## What do employers want?

- Technical and people skills
- T-shaped applicants (breadth and depth)
- Experiences in addition to coursework
- Being able to explain how you as an individual have brought value to a team and organization
- NOT everything on the wish list in the job ad (but some things with evidence to back it up)

**Let's dig a little more into  
mathematical and statistical modeling  
data science  
artificial intelligence**

**What do you think of when you hear the words  
“Artificial Intelligence (AI)”?**

# What do you think of when you hear the words “Artificial Intelligence (AI)”?



IGN  
Alienware Aurora R15 RTX 4090 Gamin...



computer  
robot?  
tech?  
machine?



National Geographic Society  
What is a Robot?

# What do you think of when you hear the words “Artificial Intelligence (AI)”?



IGN  
Alienware Aurora R15 RTX 4090 Gamin...

computer  
robot?  
tech?  
machine?



National Geographic Society  
What is a Robot?



algorithm/data  
training/testing  
“learning” / adapting  
interacting

# What are the human roles in “Artificial Intelligence (AI)”?



IGN  
Alienware Aurora R15 RTX 4090 Gamin...

computer  
robot?  
tech?  
machine?



National Geographic Society  
What is a Robot?



Literary Hub  
Ruha Benjamin on the Tech Workers ...

algorithm/data  
training/testing  
“learning” / adapting  
interacting

# Humans are more involved in AI than many realize.

The screenshot shows a news article from The Verge. The header features the website's logo 'The Verge' in a large, stylized font. Below it, a navigation bar includes 'Tech', 'Reviews', 'Science', 'Entertainment', and 'More'. The main title 'AI Is a Lot of Work' is prominently displayed in large white text against a dark blue background. Below the title, a subtitle reads 'ARTIFICIAL INTELLIGENCE'. A date and time 'Jun 20, 2023, 8:05 AM EDT' and a comment count '24 Comments / 24 New' are visible. Social sharing icons for Twitter, Facebook, and LinkedIn are present. The main text discusses the emergence of a vast tasker underclass due to AI technology. At the bottom, bylines mention 'By Josh Dzieza' and 'Illustrations by Richard Parry for The Verge'.

<https://www.theverge.com/features/23764584/ai-artificial-intelligence-data-notation-labor-scale-surge-remotasks-openai-chatbots>

Large language models like Chat-GPT, image detection, pattern recognition, translation and other algorithms require large correctly labeled datasets.

algorithm/data  
training/testing  
“learning” / adapting  
interacting

# Future of work involves human-machine teamwork



L Literary Hub  
Ruha Benjamin on the Tech Workers ...

## QUESTIONS YOU MIGHT CONSIDER WITH YOUR STUDENTS:

- To what extent and when should AI mimic ways humans solve problems and interact?
- Will AI create ways of doing things more cheaply in ways that humans will accept/use?
- What policies or laws could help prevent harm in the creation and use of AI?
- What is the problem that the creators of AI are trying to solve?
- Who will be well-served by the answers that AI delivers?
- What will be the human “AI durable” jobs of the future?

**There has been a lot of buzz about AI  
with the recent Chat-GPT release.**

## “Powered by AI” – means many things.

**duolingo**

AI helps us analyze user responses in order to identify common grammar mistakes and make corrections and suggestions, and analyze the sounds and patterns of your speech, so we can provide targeted feedback on your pronunciation. Mar 30, 2023

<https://blog.duolingo.com/ai-improves-education/>

## “Powered by AI” – means many things.

**duolingo**

AI helps us analyze user responses in order to identify common grammar mistakes and make corrections and suggestions, and analyze the sounds and patterns of your speech, so we can provide targeted feedback on your pronunciation. Mar 30, 2023

Our responsible AI standards are best practices for ensuring accountability, transparency, fairness, privacy, and security in testing. By sharing them with the public, we hope to continue to lead the way in digital-first assessment.

<https://blog.duolingo.com/ai-improves-education/>

# These standards point to concerns about AI



Our responsible AI standards are best practices for ensuring accountability, transparency, fairness, privacy, and security in testing. By sharing them with the public, we hope to continue to lead the way in digital-first assessment.

- **Accountability – can you challenge results? Who is responsible for failures?**
- **Transparency – do you know what data were used and how?**
- **Fairness – who does the solution serve? who generated the data?**
- **Privacy – who owns and has access to the data and results?**
- **Security – are the data and results safe from leaks and hacks?**

<https://blog.duolingo.com/ai-improves-education/>

# Let's compare the models of modeling

## Traditional Math / Stats modeling

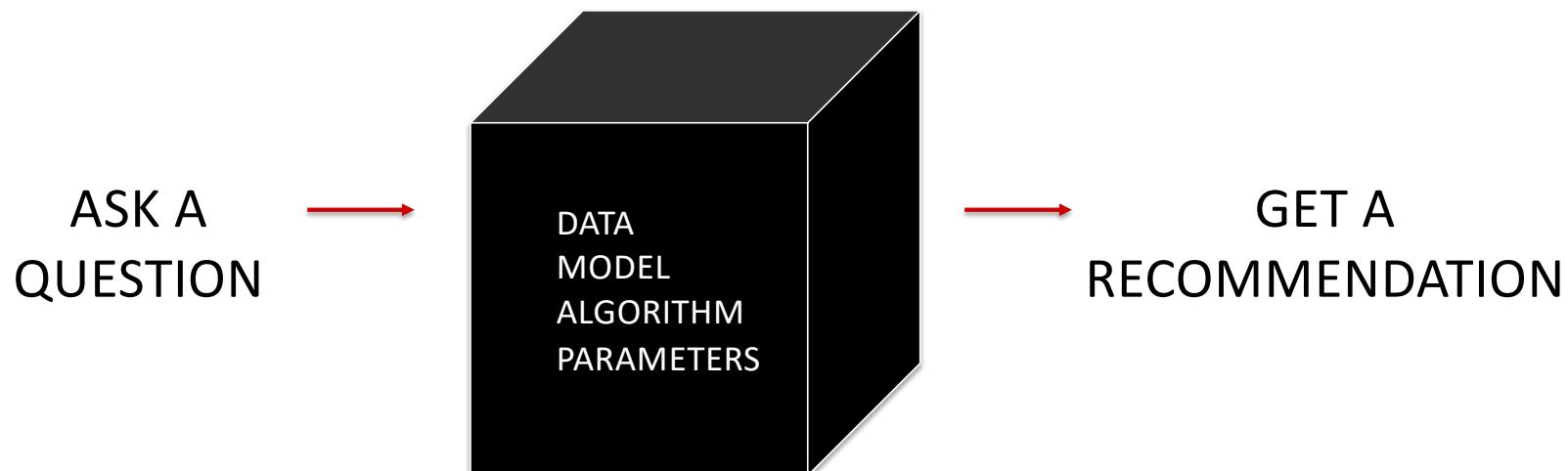
- Formulate question/hypothesis
- Gather data relevant to the problem
- Build model, generate solution
- Test model and iterate
- Communicate results

# Data science is big umbrella that can include traditional modeling, but sometimes is quite different.

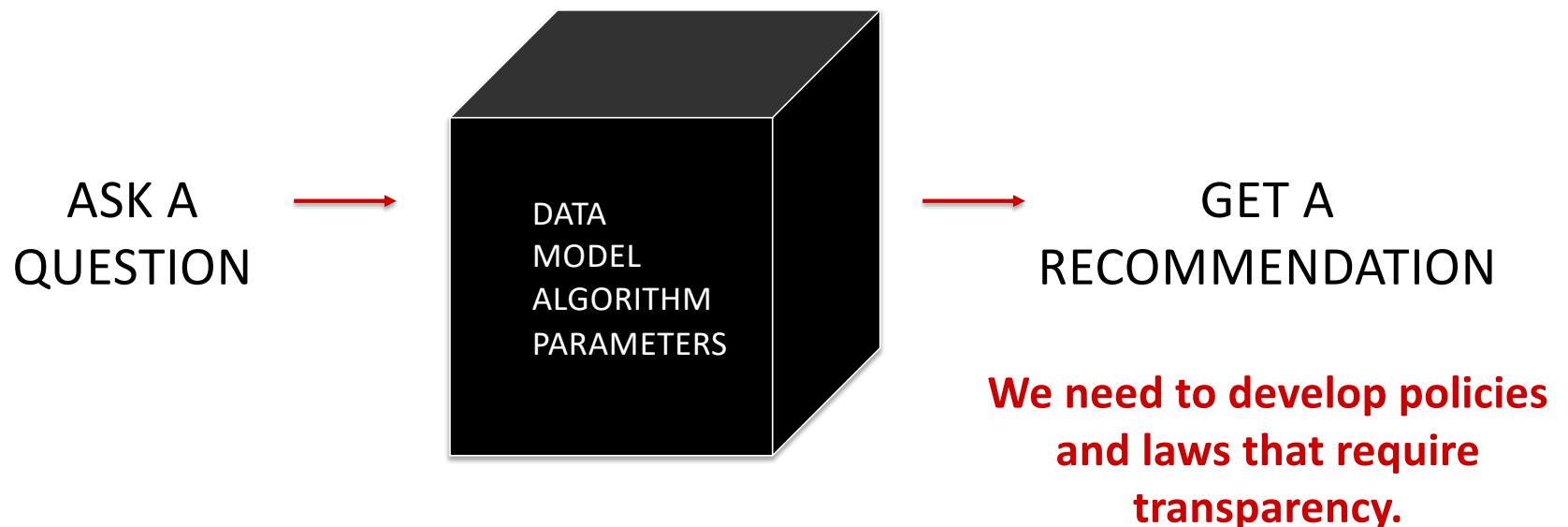
## Traditional Math / Stats modeling   Some Current Data Science / AI

- Formulate question/hypothesis
- Gather data relevant to the problem
- Build model, generate solution
- Test model and iterate
- Communicate results
- Large data sets (**can be black box**)
- Ask question
- Build/choose model(s)/algorithm(s)  
**(sometimes details are secret)**
- Generate and iterate solutions  
**(sometimes in black box)**
- Assess output **(can be hard to test)**

**SOLUTIONS ARE FAIRLY IMPOSSIBLE TO TEST OR CHALLENGE  
IF YOU DON'T KNOW WHAT IS IN THE BOX OR HOW IT WORKS.**



**SOLUTIONS are fairly impossible to test or challenge  
if you don't know what is in the box or how it works.**



# What data science future are you excited about?

- Individualized, meaningful learning experiences.
- Automating jobs that are harmful to humans.
- Personalized medicine.
- Aiding scientific exploration and discovery.
- Language learning and translation.
- Addressing environmental and health issues.
- **In general, finding solutions to important problems to build a better future.**

**Together, we are going to grow  
data science education.**

**I hope we will keep in touch. Thank you!**

**Rachel Levy (Ray)**

**web: [datascienceacademy.ncsu.edu](http://datascienceacademy.ncsu.edu)**

**email: [datascienceacademy@ncsu.edu](mailto:datascienceacademy@ncsu.edu)**

Follow Us: @NCStateDSA

