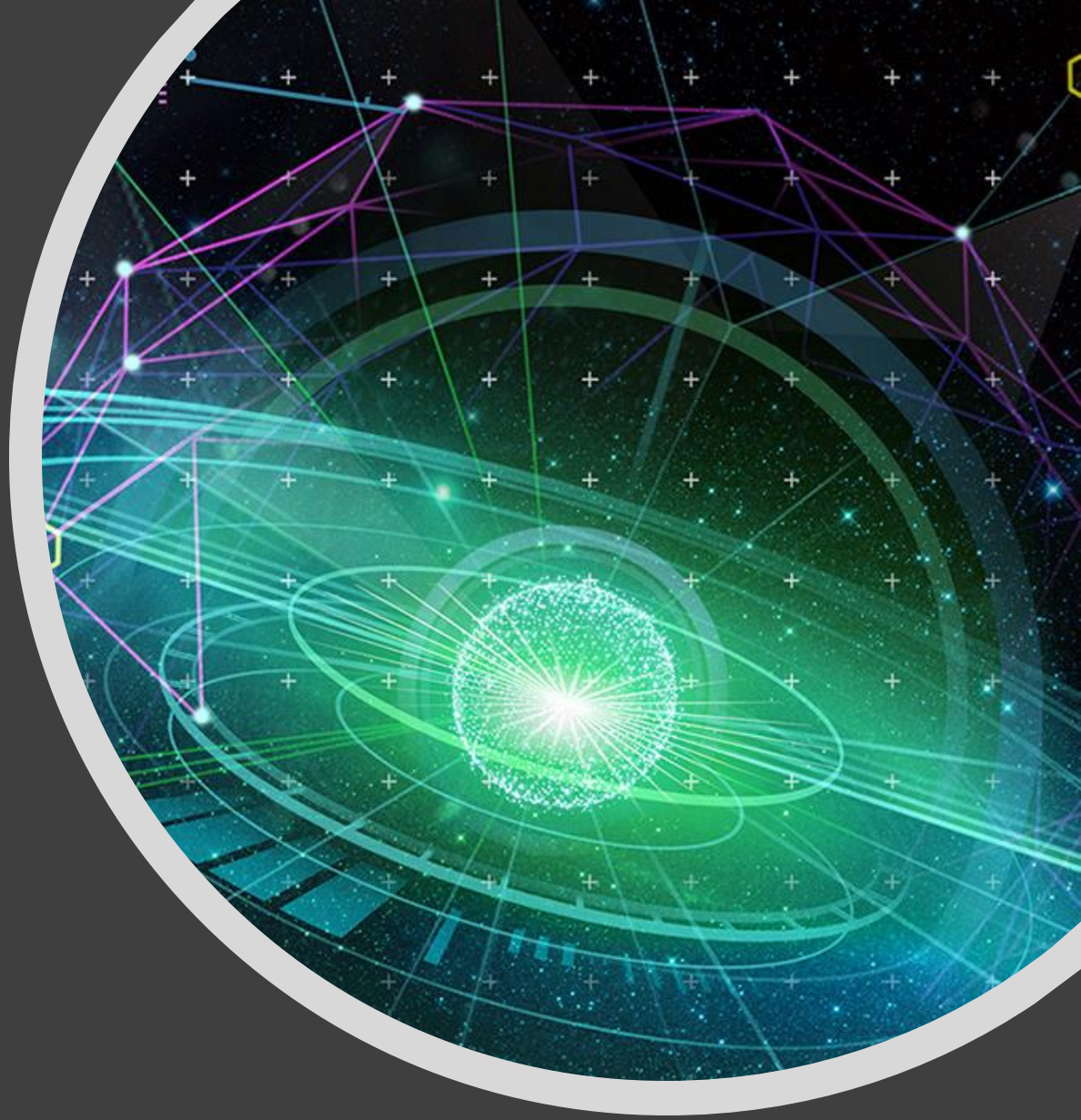
The Future of Data Science

Dr. Robert Buscaglia | High School Math Day | November 2nd, 2022

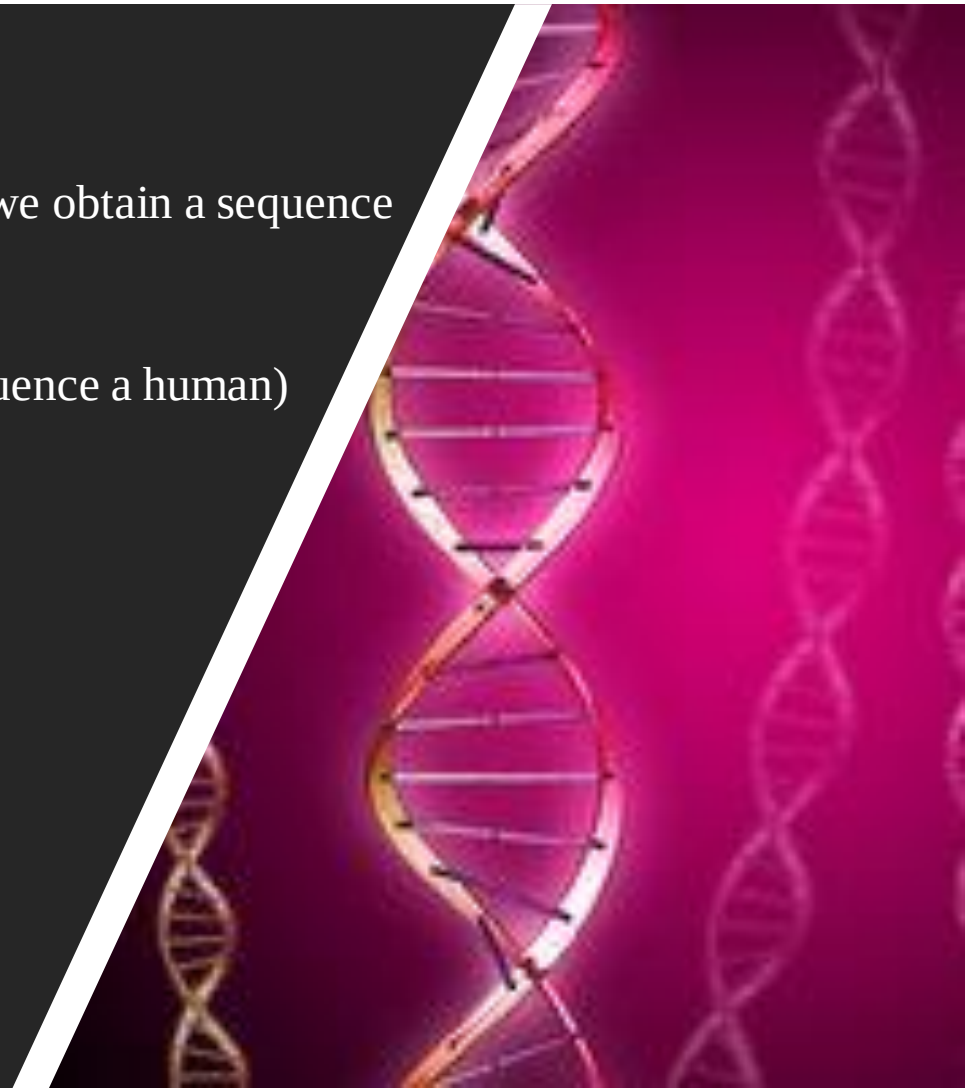
What is Data Science?

- Just like any other science: *it is the study of data*.
- The field of study is broadly defined
  - Developing methods for data storage
  - Enhancing data recording
  - Analyzing and extracting useful information
  - Mining data for new insights
- Data Science is a mixture of Mathematics and Computer Science
  - Algorithm development
  - Statistical testing
  - Programming is almost always a key component
  - High-performance computing (super computers)



Why Data Science Now?

- We live at a time where data collection rates far exceed data analysis.
  - Example : Human Genetics
    - For every single human who has their genetic information scanned, we obtain a sequence of approximately 3,000,000,000 nucleic acids.
    - The technology to sequence a human genome now takes ***hours***.
      - Human Genome Project (~2003, $1 Billion and 13 years to sequence a human)
      - 10-year update (~2013, $3000-5000 and 1-2 days)
      - Recent (~2018, $100 and 1-2 hours)
  - We are only learning from a fraction of what is being collected
- Data Science helps bridge the information gap
  - Methods that improve our speed and critical level of understanding.
  - Making new connections that were not otherwise possible
  - Advancements in Statistical and Machine Learning
    - Computational Methods
    - Neural Networks & Artificial Intelligence



How did I end up in Data Science?

- Built first computer when 12 years old
- “Hobby” has always been computers (A+ cert, programming)
- BS Applied Mathematics and BS Biochemistry @ NAU
  - Introduced to Mathematical Programming – Mathematica
  - Studied unique DNA structures : Q-quadruplexes
  - Found an interest in modeling complex systems
- PhD Biochemistry and Molecular Biology @ University of Louisville
  - Experimental Biophysics
  - Modeling of complex equilibrium through algorithm development
  - Anti-cancer therapeutic design
- PhD Applied Mathematics @ Arizona State University
  - Introduced to the concepts of “Data Science”
  - Designed algorithm to diagnose Lupus

Biochemistry and Analysis

- Data science was integral to my PhD Biochemistry
  - We just didn't call it data science yet!

$$\ln\left(K_i\right) = \frac{-\Delta H_i}{R}\left(\frac{1}{T_{m,i}} - \frac{1}{T}\right)$$

1) $F \rightleftharpoons U$

2) $F_1 \rightleftharpoons U \rightleftharpoons F_2$

3) $F \rightleftharpoons I \rightleftharpoons U$

4) $F \rightleftharpoons I_1 \rightleftharpoons I_2 \rightleftharpoons U$

5) $F_1 \rightleftharpoons I \rightleftharpoons U \updownarrow F_2$

F = folded state(s)
I = intermediate state(s)
U = unfolded ensemble

Three Transitions (Mechanism 4)

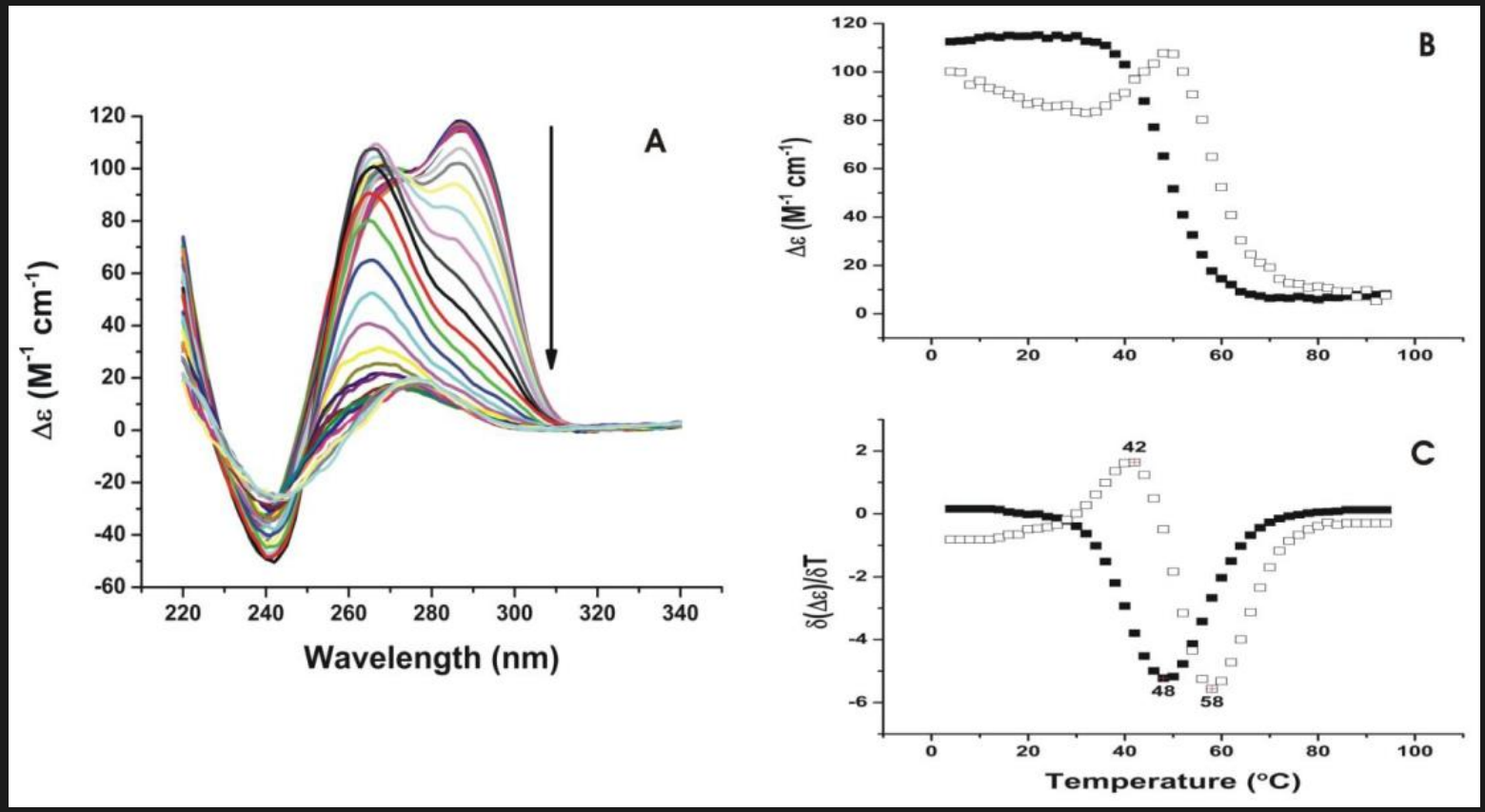
$$S_{obs_j} = \frac{S_{n_j} + S_{1_j} e^{\frac{-dH_1}{R}\left(\frac{1}{Tm1}-\frac{1}{T}\right)} + S_{2_j} e^{\frac{-dH_1}{R}\left(\frac{1}{Tm1}-\frac{1}{T}\right)} e^{\frac{-dH_2}{R}\left(\frac{1}{Tm2}-\frac{1}{T}\right)} + S_{u_j} e^{\frac{-dH_1}{R}\left(\frac{1}{Tm1}-\frac{1}{T}\right)} e^{\frac{-dH_2}{R}\left(\frac{1}{Tm2}-\frac{1}{T}\right)} e^{\frac{-dH_3}{R}\left(\frac{1}{Tm3}-\frac{1}{T}\right)}}{1 + e^{\frac{-dH_1}{R}\left(\frac{1}{Tm1}-\frac{1}{T}\right)} + e^{\frac{-dH_1}{R}\left(\frac{1}{Tm1}-\frac{1}{T}\right)} e^{\frac{-dH_2}{R}\left(\frac{1}{Tm2}-\frac{1}{T}\right)} + e^{\frac{-dH_1}{R}\left(\frac{1}{Tm1}-\frac{1}{T}\right)} e^{\frac{-dH_2}{R}\left(\frac{1}{Tm2}-\frac{1}{T}\right)} e^{\frac{-dH_3}{R}\left(\frac{1}{Tm3}-\frac{1}{T}\right)}}$$

Parallel Intermediate Model (Mechanism 5)

$$S_{obs_j} = \frac{e^{\frac{-dH_3}{R}\left(\frac{1}{Tm1}-\frac{1}{T}\right)}\left(S_{n1_j} e^{\frac{dH_1 Tm_2(T-Tm_1)+dH_3 Tm_1(T-Tm_2))}{R \cdot T \cdot Tm_1 \cdot Tm_2}} + S_{n2_j} e^{\frac{(dH_1+dH_3)(T-Tm_2)}{R \cdot T \cdot Tm_2}} + S_{1_j} e^{\frac{dH_3\left(\frac{1}{Tm2}-\frac{1}{T}\right)}{R}} + S_u\right)}{1 + e^{\frac{dH_1}{R}\left(\frac{1}{Tm1}-\frac{1}{T}\right)} + e^{\frac{dH_2}{R}\left(\frac{1}{Tm2}-\frac{1}{T}\right)} + e^{\frac{-dH_3}{R}\left(\frac{1}{Tm3}-\frac{1}{T}\right)}}$$

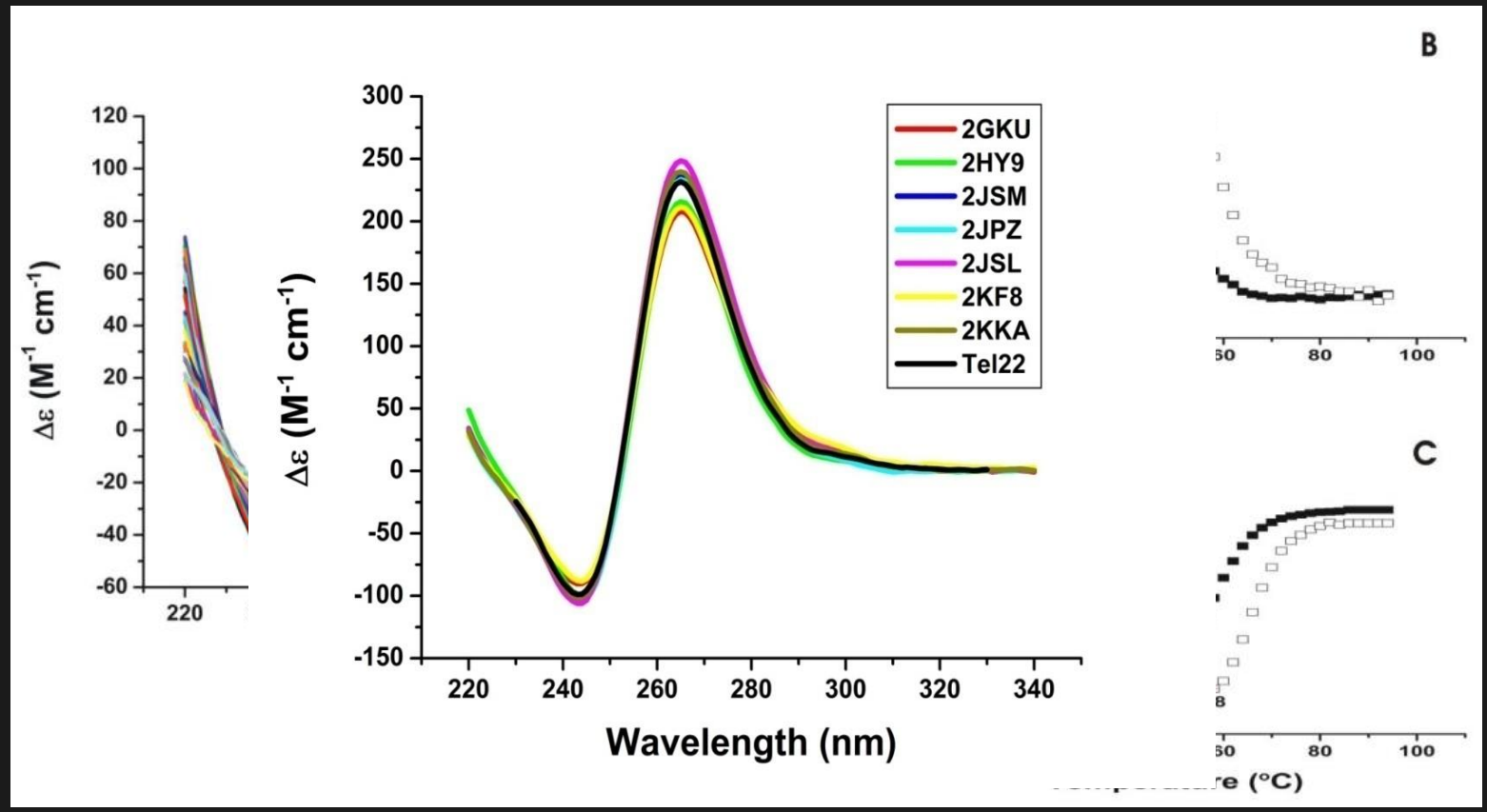
Thermodynamic Deconvolution

- Modern spectroscopy allows for simultaneous collection of full-spectra across the controlled variable (here: temperature).
- Thermodynamic characterization is enriched by incorporation of the full-spectra in the final analysis.
- Provided insight into folding mechanisms.
  - DNA acting like proteins – complex 3-dimensional structure.
  - Elucidated new targets for anti-cancer therapies that are still under development!



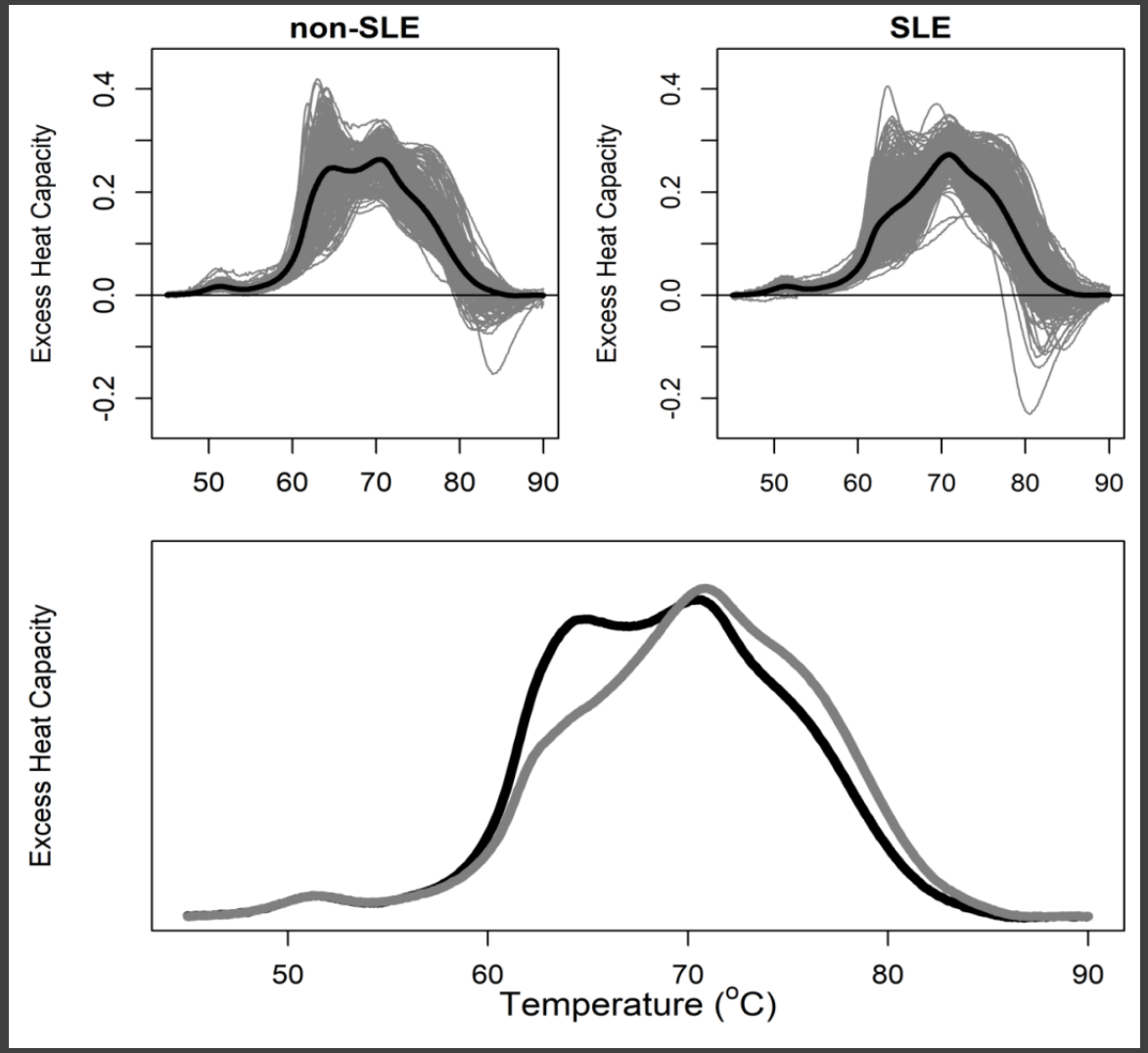
Thermodynamic Deconvolution

- Spectral reconstruction resulted in identification of common intermediate species.
  - This intermediate species was identified only by use of mathematical decomposition (singular value decomposition) and reconstruction using linear algebra approaches.
  - The intermediate is known as triplex-DNA and can be specifically targeted by anti-cancer compounds.
  - The intermediate is present regardless of starting state making it a robust target for treatment.

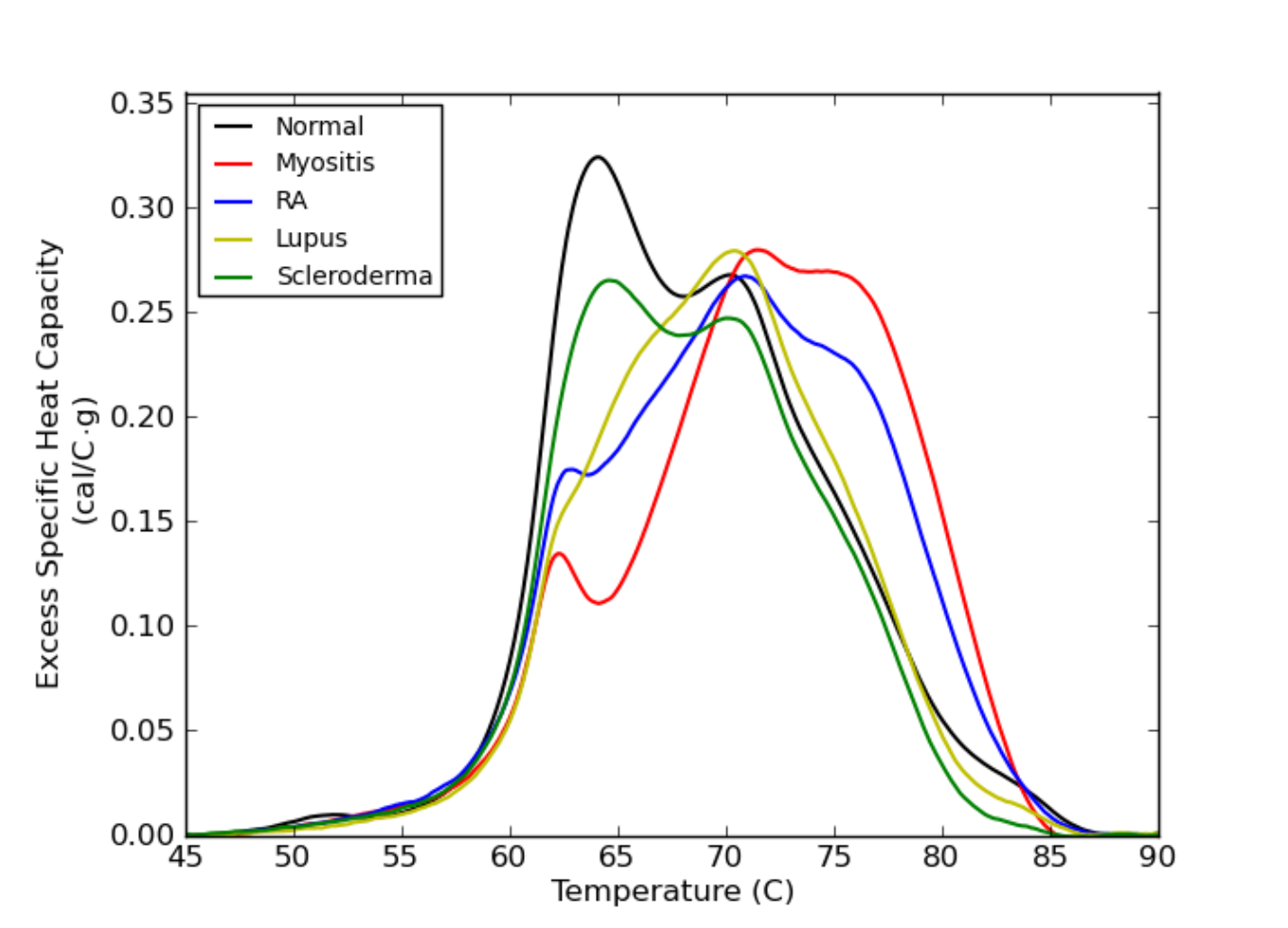


The Power of Data

- A Challenging Analysis Problem and the Role of Data Science
- Can we identify the presence of lupus against non-lupus alternatives?



Clinical DSC



Mean curves shown for N≥25 thermograms

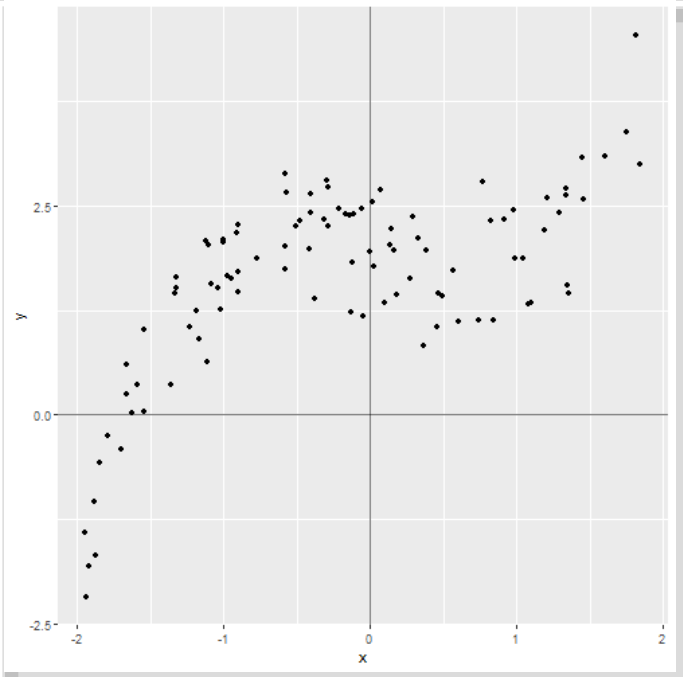
- ESFuNC algorithm
  - Tuned test set accuracy of 94.2% for lupus identification.
- Medical directions
  - Clinically driven
  - Improvement / Early-stage diagnosis
  - Treatment progression
- Data Driven Biochemistry!
  - BART
  - Deep Neural Networks
  - Random Forests

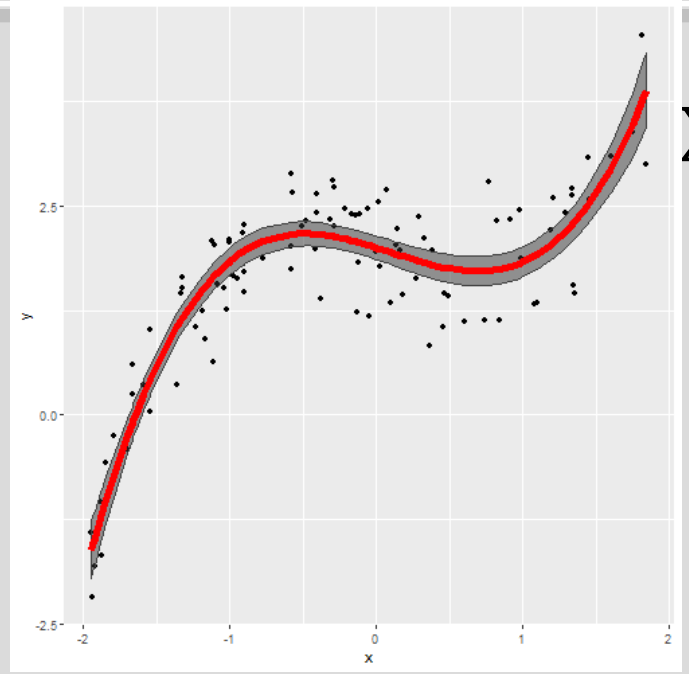
Getting Starting in Data Science

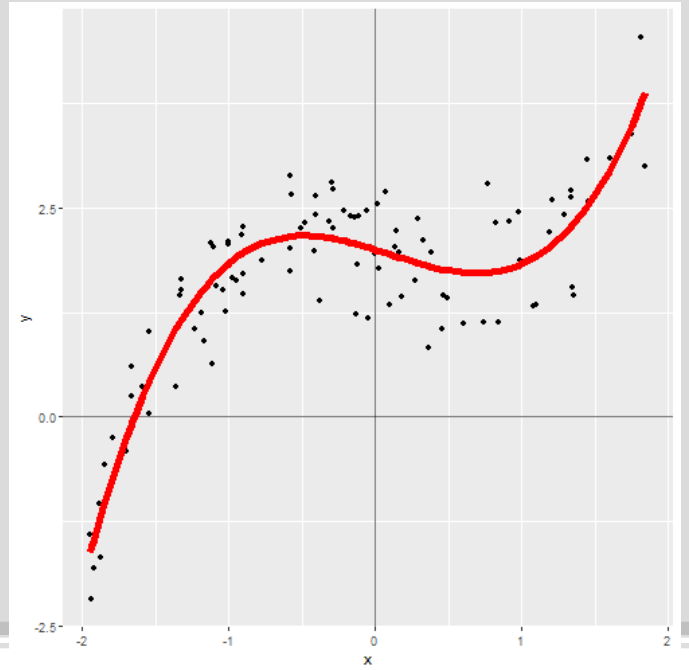
- The core knowledge base are skills learned during High School
  - **Algebra** sets the foundation for most data science techniques
  - **Calculus** is a requirement only for deeper investigation / research / development
    - Calculus is the foundation of theoretical statistics and the study of probability.
  - **Statistics** is an essential component of learning algorithms

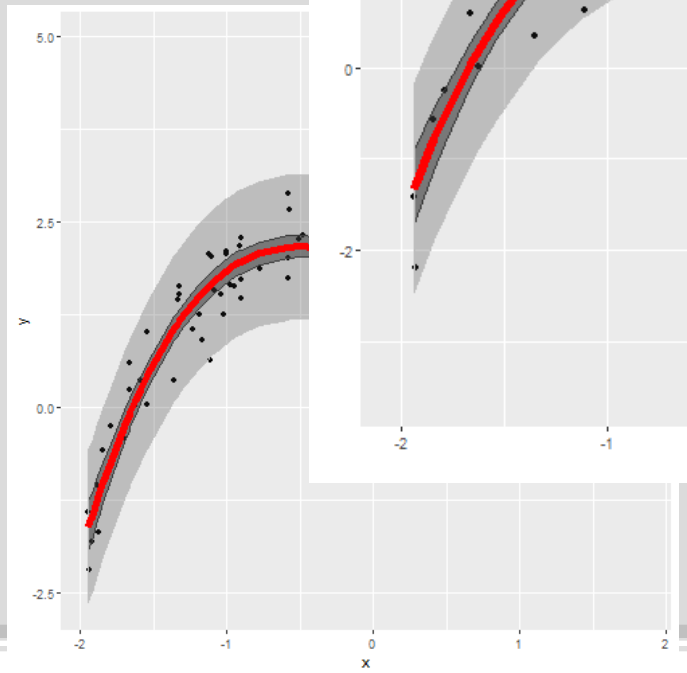
- **Computer science** and programming are critical skills to start practicing
  - High school computer courses lay important groundwork that will be used during a data science career.
  - It is never too early to start learning programming skills (C, R, Python are all free!!).
  - Being comfortable with computing is always a good thing but requires practice!

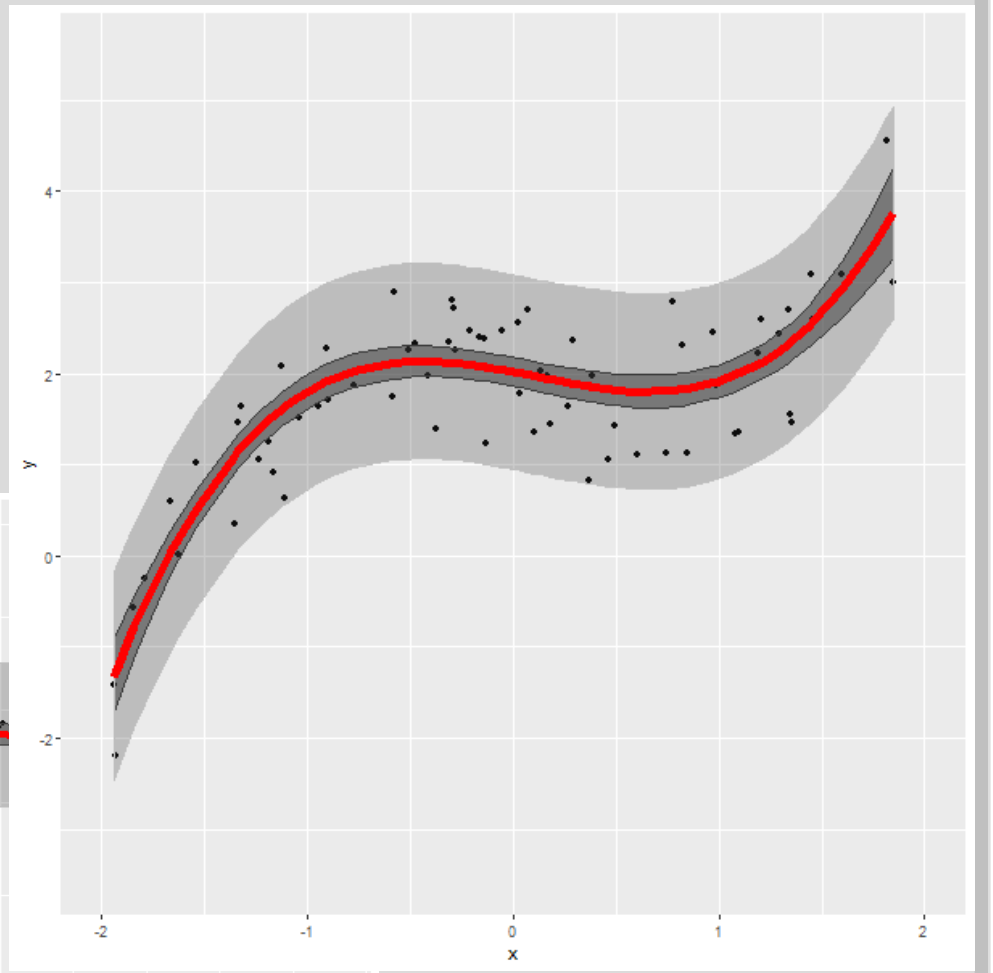
xample - Bootstrap





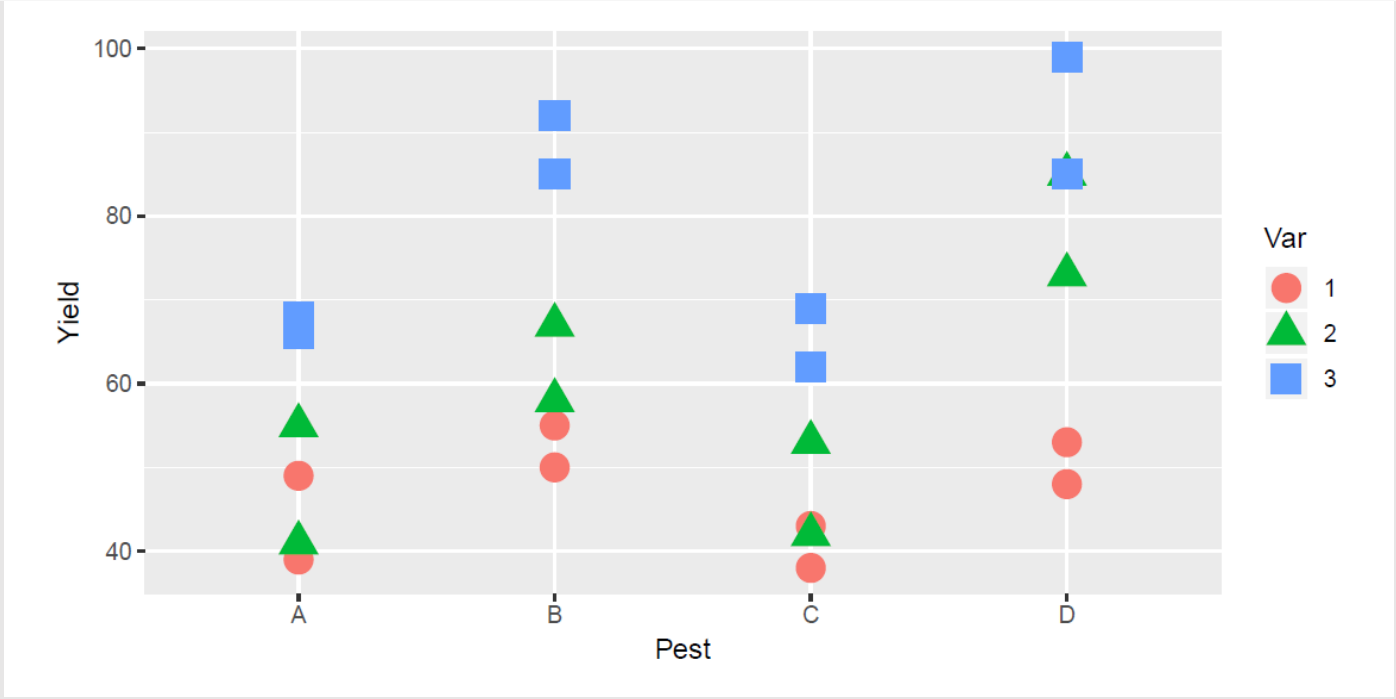






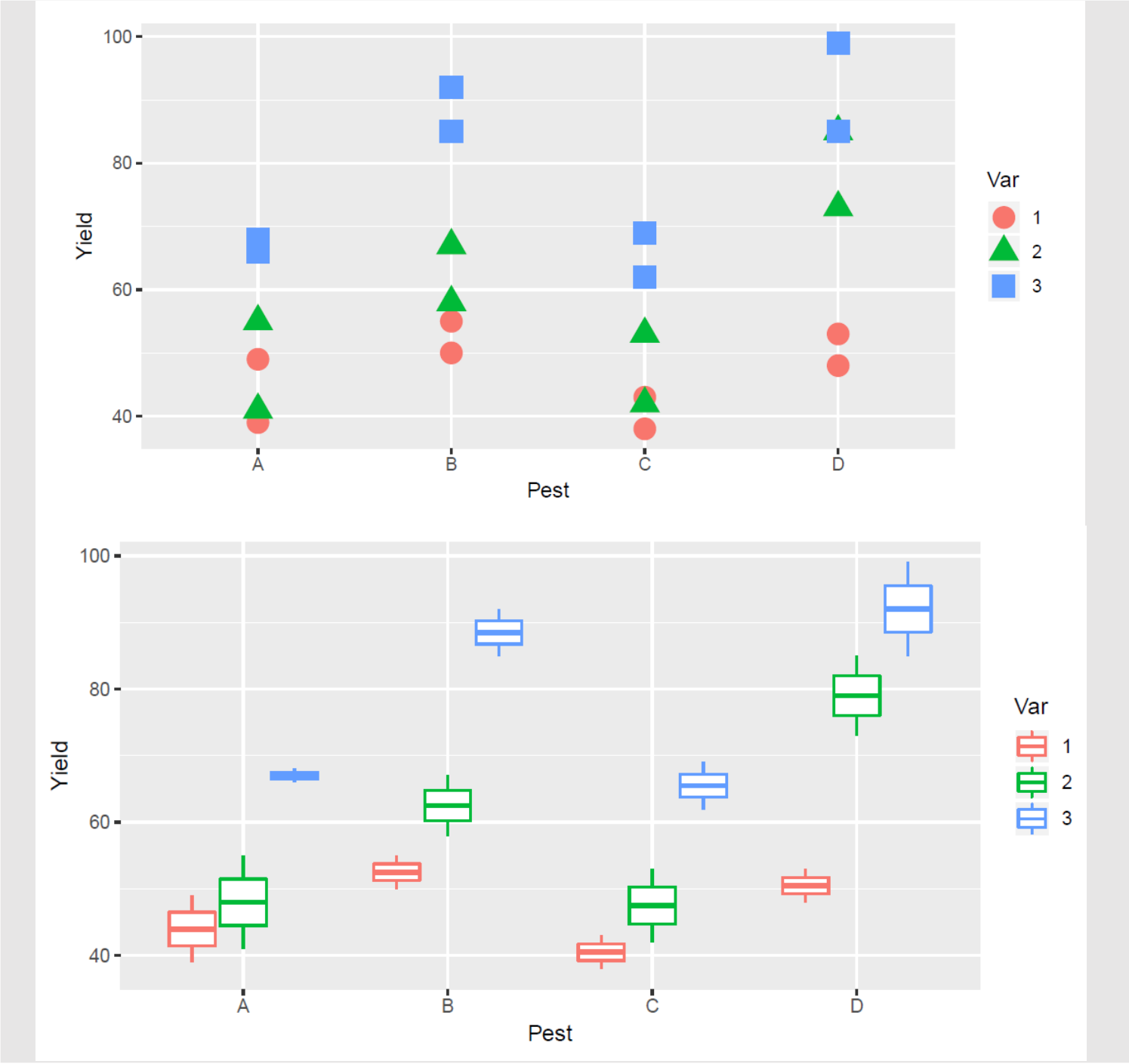
What skills can you practice today?

- Learning the basics of programming
  - Writing basic code
    - for-loops, functions, control
  - Data Handling
    - Commands for ‘cleaning’ data
  - **Visualization**
    - Honing your intuition.
    - What can we learn from correct visualization of results?
  - Data availability
    - www.kaggle.com
    - World of Data
- Recommended Programming Languages
  - **R, Python**
  - Free to learn on-line with many freely available tutorials.



What skills can you practice today?

- Learning the basics of programming
  - Writing basic code
    - for-loops, functions, control
  - Data Handling
    - Commands for ‘cleaning’ data
  - **Visualization**
    - Honing your intuition.
    - What can we learn from correct visualization of results?
  - Data availability
    - www.kaggle.com
    - World of Data
- Recommended Programming Languages
  - **R, Python**
  - Free to learn on-line with many freely available tutorials.



The Future of Data Science

Impact of Data and Benefits of Analysis

How Data Science is Growing
(It is Everywhere)

- Corporations & Industry
  - Amazon, Google, Facebook, Uber
  - Making use of data to improve customer interactions.
  - Tesla – make cars drive themselves!
- Interdisciplinary Science
  - Research combined with improved data analysis.
  - National Science Foundation : Data-Driven Chemistry
- Local improvements and production.
  - A small restaurant could use data to inform them how to organize their staff.
  - Transportation data to improve roadwork and highway congestion.
- Data Science & Arts!

Data Science is Growing

- Specialization
  - A likely change in the future of data science is more importance to ‘specialized’ talents.
    - Data Cleaning
    - Data Analytics
    - Algorithm Design
  - This can mean combining information from multiple disciplines.
  - **For college** : minors are a good thing! Apply Data Science to an area you are interested in!
    - Physical Science: Chemistry, Biochemistry
    - Forestry, Engineering, Physics
    - Art & Music
    - Linguistics (speech and writing)
    - Economics (stock market)
    - Sports Science and Esports

Data Science and Art : DALL-E

- Generate images based on input text commands.
- What would character described in a book look like?

- Example 1: “Teddy bears mixing sparkling chemicals as mad scientists in a steampunk style.”

Data Science and Art : DALL-E

- Generate images based on input text commands.
- What would character described in a book look like?

- Example 1: “Teddy bears mixing sparkling chemicals as mad scientists in a steampunk style.”

Data Science and Art : DALL-E

- “An astronaut riding a horse in a photorealistic style.”

Data Science and Art : DALL-E

- “An astronaut riding a horse in a photorealistic style.”

Data Science at NAU

Our degree program

COLLEGE OF THE ENVIRONMENT, FORESTRY, AND NATURAL SCIENCES

Department of Mathematics and Statistics

Data Science, Bachelor of Science

Data science is systematic way to discover, quantify, and understand relationships hidden in data. A data scientist is an expert in acquiring, storing, accessing, and manipulating data; visualizing and modeling relationships in the data; and drawing actionable conclusions from the data. Data science is a tool applied to other areas of study such as biological sciences or business, and emphasizes the holistic process of "thinking with (real) data."

The required course work is in statistics and computer science with the upper-division statistics courses utilizing the program competency acquired. Students are encouraged to pursue a minor in another field of interest in order to gain deep understanding of the challenges and needs that can be addressed by data science.

Careers | University Requirements | Overview | Details | Campus Availability

What Can I Do with a Bachelor of Science in Data Science?

Data science is an essential tool to nearly all disciplines and businesses that collect any data. Government and industry employers need data scientists to store, manipulate, and analyze data acquired in research, engineering, finance, marketing, public health, and other fields. By combining quantitative training with a minor in another field, data scientists are employed in a wide range of fields.

Career opportunities that might be pursued:

- Data Scientist
- Data Analyst
- Research Analyst
- Big Data Analyst

With further education, one of these paths is possible:

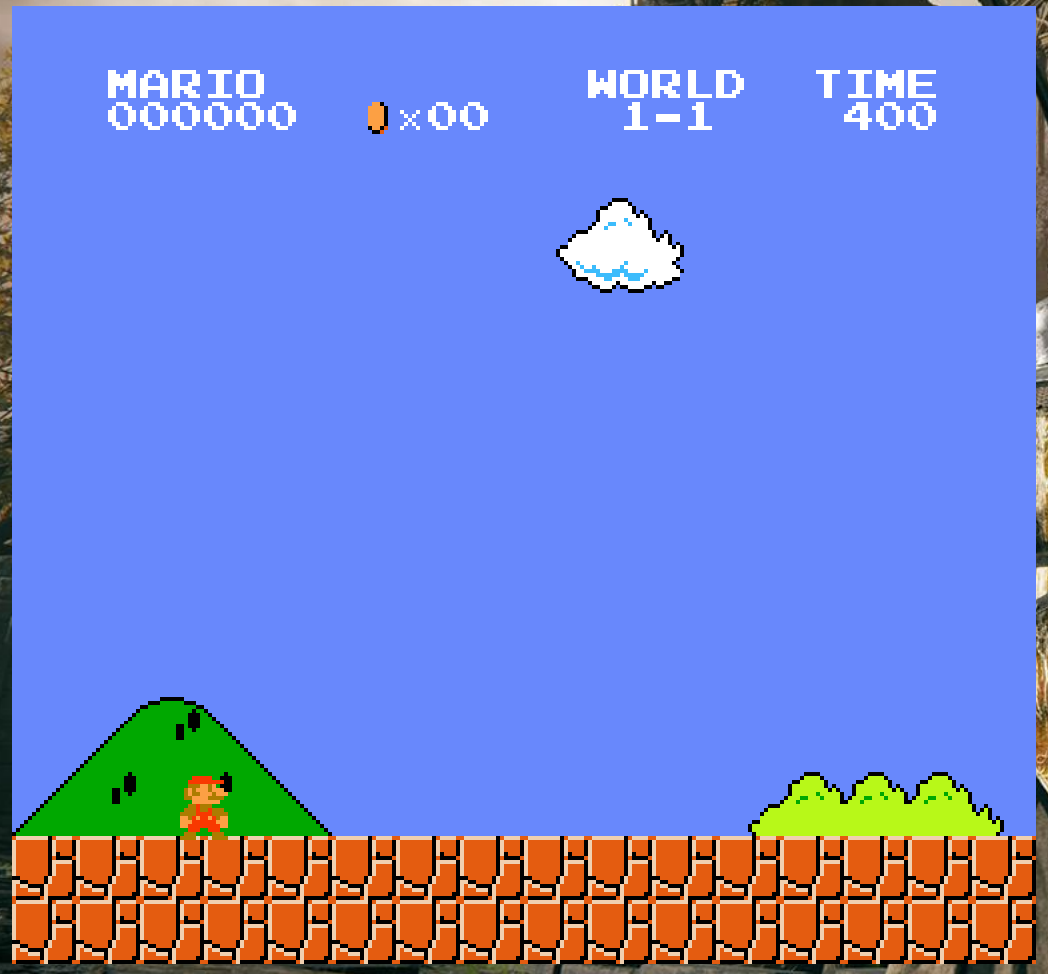
- Statistics
- Econometrics
- Actuarial Science
- Informatics

One Last Example – Reinforcement Learning

One Last Example – Reinforcement Learning

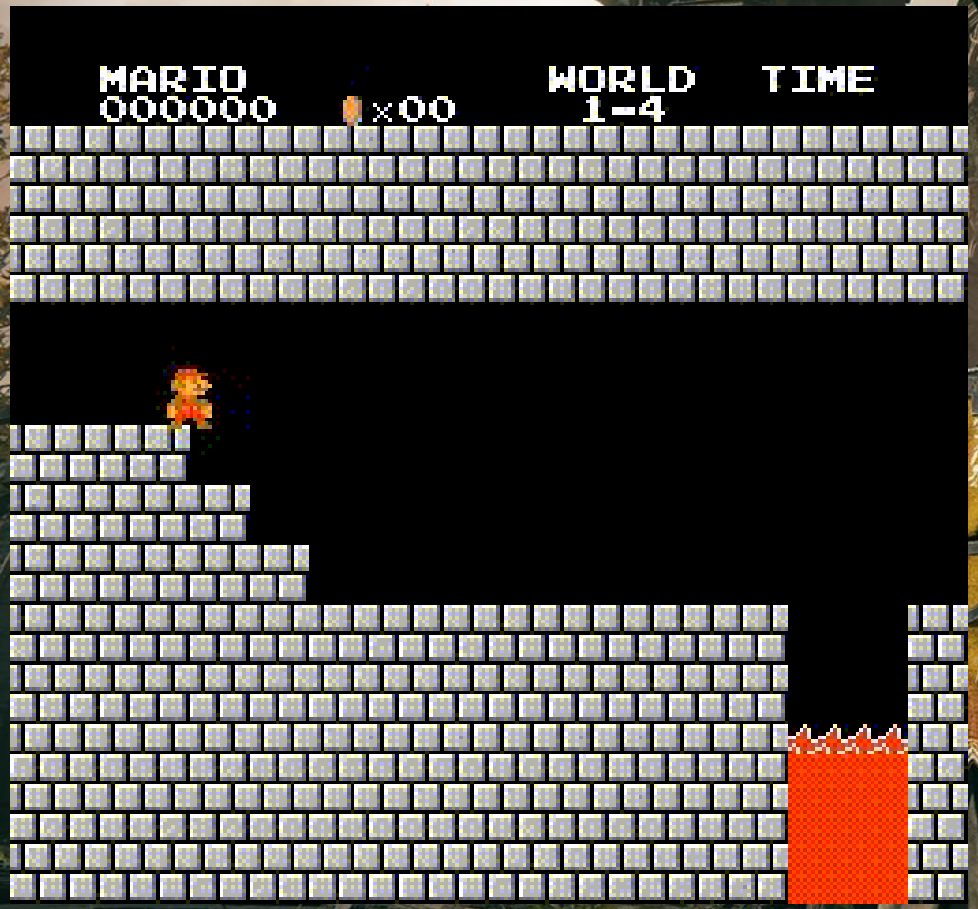
- **Training AI to a task**
  - Engineering
  - Construction
  - Self-driving Cars
- **An excellent example of reinforcement is video games!**
  - Train a network to continually improve the score based on the game architecture.
  - Is a fun example to see how SMART artificial intelligence can become.
- **This is how new game testing is being done!**
  - Bethesda, EA, Blizzard, Riot

One Last Example – Reinforcement Learning



- **Training AI to a task**
  - Engineering
  - Construction
  - Self-driving Cars
- **An excellent example of reinforcement is video games!**
  - Train a network to continually improve the score based on the game architecture.
  - Is a fun example to see how SMART artificial intelligence can become.
- **This is how new game testing is being done!**
  - Bethesda, EA, Blizzard, Riot

One Last Example – Reinforcement Learning



- **Training AI to a task**
  - Engineering
  - Construction
  - Self-driving Cars
- **An excellent example of reinforcement is video games!**
  - Train a network to continually improve the score based on the game architecture.
  - Is a fun example to see how SMART artificial intelligence can become.
- **This is how new game testing is being done!**
  - Bethesda, EA, Blizzard, Riot


THEN A MIRACLE OCCURS...
S. Harris
"I THINK YOU SHOULD BE MORE EXPLICIT HERE IN STEP TWO."

Thanks!