Data Science

- What is data science?
  - "Study of data to extract meaningful insights for business. Combines principles and practices from fields of mathematics, statistics, artificial intelligence, and computer engineering to analyze large amounts of data." (AWS Amazon)
  - "Data science combines math and statistics, specialized programming, advanced analytics, artificial intelligence (AI), and machine learning with specific subject matter expertise to uncover actionable insights hidden in an organization's data." (IBM)
  - "Interdisciplinary field that uses scientific methods, processes, and systems to extract knowledge and insights from data across various domains." (Textbook)
- Common themes?
  - Discovering insightful information from data
  - Multi-field work (Math, Stats, CS)

Where is Data Science Used?

- Wherever data is created
- Wherever information can be extracted from data
- A few tasks:
  - Machine learning
  - Visualization
  - Predictions
  - Clustering
  - Trend analysis
- Fields that use data science:
  - Healthcare
  - Financial
  - Sports
  - Natural Language Processing (Text)
  - If there is data in a field, then data science will most likely be there too

Something to Think About

- Consider what data is considered useful / insightful
- What can we learn from that data?
- Where does the data come from?
- What story is in the data?

Why is Understanding Data so Important?

- Visualizations of data may be misleading.
- Even if the information is correct, changing the x or y scale on a graph can lead to different interpretations of the data presented.

Causation vs Correlation

- The increase in ice cream sales occurs at the same time that the temperature is higher
  - Stop eating ice cream $\rightarrow$ Climate change solved
- Snowstorms are more frequent when hockey games in Canada are more frequent
  - Hockey games cause snowstorms