



TEXAS ADVANCED COMPUTING CENTER

WWW.TACC.UTEXAS.EDU



TEXAS

The University of Texas at Austin

# What is Data Science

**PRESENTED BY:**

Charlie Dey

Director, Training and Professional Development

Texas Advanced Computing Center

# Data Science 101

What is Data?

# Data Science 101



What is Data |



- what is data **in computer**
- what is data **in statistics**
- what is data **mining**
- what is data **on a phone**
- what is data **analysis**
- what is data **in dbms**
- what is data **collection**
- what is data **roaming**
- what is data **processing**
- what is data **science**

*Report inappropriate predictions*

# Data Science 101

Data is a set of values of subjects with respect to qualitative or quantitative variables. Data and information or knowledge are often used interchangeably; however **data becomes information when it is viewed in context or in post-analysis.** [Wikipedia](#)

# Data Science 101

Data is everywhere!

Where is your data?

# Data Science 101

What makes data important?

# Data Science 101



Data Science is |



data science is **different now**  
data science is **a branch of**  
data science is **the future**  
data science is **overrated**  
data science is **hard**  
data science is **a fad**  
data science is **just statistics**  
data science is **not science**  
data science is **dead**  
data science is **a team sport**

*Report inappropriate predictions*

# Data Science 101

Data science is a multi-disciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from structured and unstructured data. [Wikipedia](#)



# Data Science 101

Data Science is the ability to understand that there is a story hidden in the data.

# Data Science 101

100 Million Dollars - Southwest Airlines saved by reducing the time their airplanes sat idle on the tarmac

39 Million Gallons - the amount of fuel UPS saved by optimizing its fleet

32,000 Dollars the amount of money it costs TACC to have our machines sitting idle

# Data Science 101

Data is worth money.

# Data Science 101

## BIG DATA

There isn't a readily available definition of Big Data because you can't "see it"

Examples of Big Data?

# Data Science 101

We are in the era of Big Data

There was a road to get to this moment with a few important stops along the way, and it's a road on which we're probably still nowhere near the end. To get to the data driven world we have today, we needed **scale**, **speed**, and **ubiquity**.

# Data Science 101

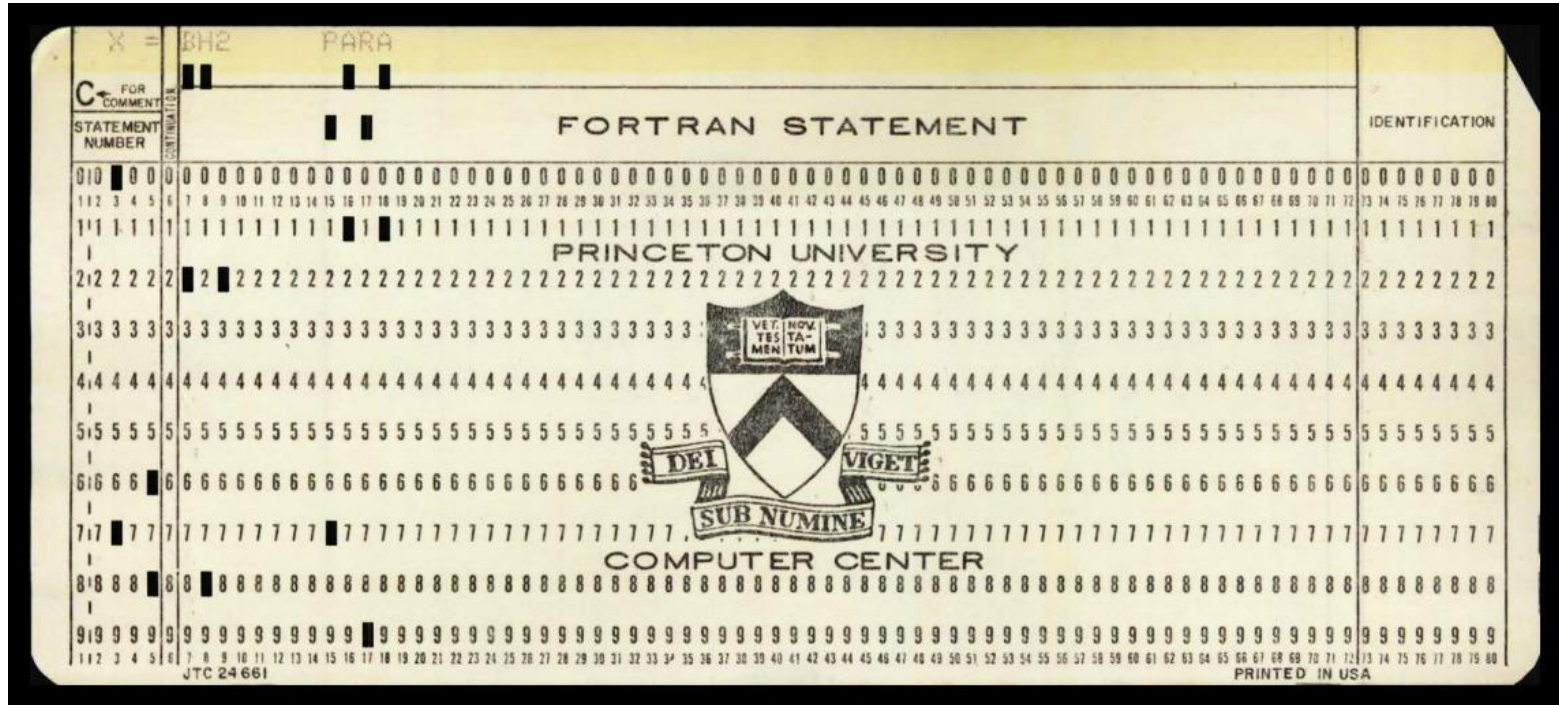
## Scale

Data started with the punch card

introduced by Herman Hollereith in 1890

7.34 inches wide by 3.25 inches high and approximately .07 inches thick, a punch card was a piece of paper or cardstock containing holes in specific locations that corresponded to specific meanings.

# Data Science 101



# Data Science 101

## Scale

Coding up data and programs through a series of holes in a piece of paper can only *scale* so far

it was revolutionary for its day because the existence of semi autotic data tallying allowed for faster and more accurate computation.



# Data Science 101

## Speed

the second prong of the big data revolution involves how fast we can move around and compute with data.

# Data Science 101

## Ubiquity

Definition: the fact of appearing everywhere or of being very common.

# Data Science 101

Putting the science in data science

it's short answer to what you can do with the billions upon billions upon of data points being collected

# Data Science 101

The data is there.

It exists

There's something valuable in it.

But what does it mean? What's going on? What can you learn? How can you use it to make better science?

**Data analysis** is all about asking these types of questions.

# Data Science 101

Here's the catch

You have to understand how the data came to be and what the goals of the process are in order to do good analytic work.

# Data Science 101

**Experimentation** has been around for a long time.

People have been testing out new ideas for far longer than data science has been a thing.

Experimentation is at the heart of a lot of modern data work.

# Data Science 101

## Machine Learning

Data scientists define machine learning as the process of using machines to better understand a process or system, and recreate, replicate or augment that system.

# Data Science 101

## Machine Learning, Supervised Learning

Supervised learning is probably the most well known of the branches of data science.

All about predicting something you've seen before.

You try to analyze what the outcome of the process was in the past and build a system that tries to draw out what matters and build predictions for the next time it happens.



# Data Science 101

## Machine Learning, Unsupervised Learning

You can do a lot of machine learning work without an observed outcome or target.

Unsupervised learning is less concerned about making predictions than understanding and identifying relationships or associations that might exist within the data.

# Data Science 101

## Machine Learning, Unsupervised Learning

The K Means algorithm.

This technique, calculates the distance between different points of data and groups similar data together.

This The “suggested new friends” feature on Facebook

# Data Science 101

## Machine Learning, Reinforcement Learning

Reinforcement learning requires an active feedback loop.

Reinforcement learning requires a dynamic dataset that interacts with the real world.

# Data Science 101

## Artificial Intelligence

Artificial Intelligence wants some kind of human interaction and is intended to be somewhat human or “intelligent” in the way it carries out those interactions. Therefore, that interaction becomes a fundamental part of the product a person seeks to build. Data science is more about insight and building systems. It places less emphasis on human interaction and more on providing intelligence, recommendations, or insights.

# Data Science 101

In a nutshell

Data is important.

We need to understand what the data is

What the data means

To find the underlying story

# Data Science 101

Demo.