

What is Data Science?

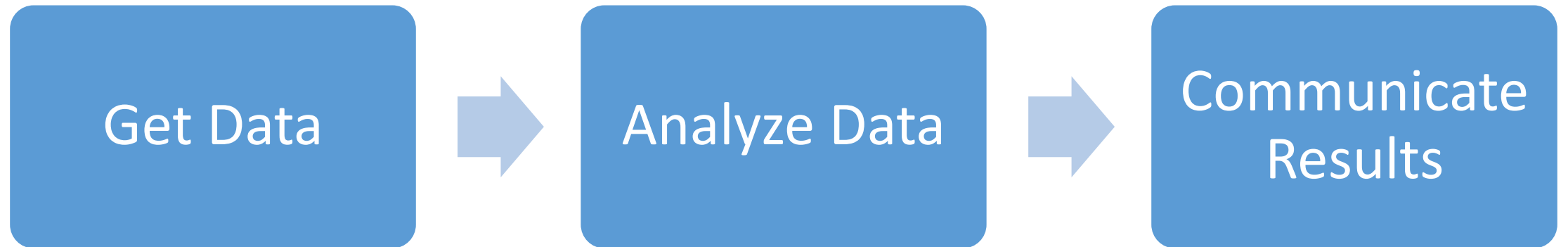
What is Data Science?

Applied mathematics

- Measurement Theory
- Data analysis
- Statistics
- Pragmatism
- Varied content domains

A cluster of principles, techniques, and skills used to get from a “real-world problem” to a “real-world solution” via quantitative modeling.

What does Data Science do?



What have we done?

Covered some common algorithmic models used in Data Science

- Conditional means
- Linear regression
- Logistic regression
- Cross-validation techniques

Added tools into your “toolbox”

- R scripts
 - Analyses
 - Visualizations

What else is there?

TONS

Some other things

Word clouds

Neural networks

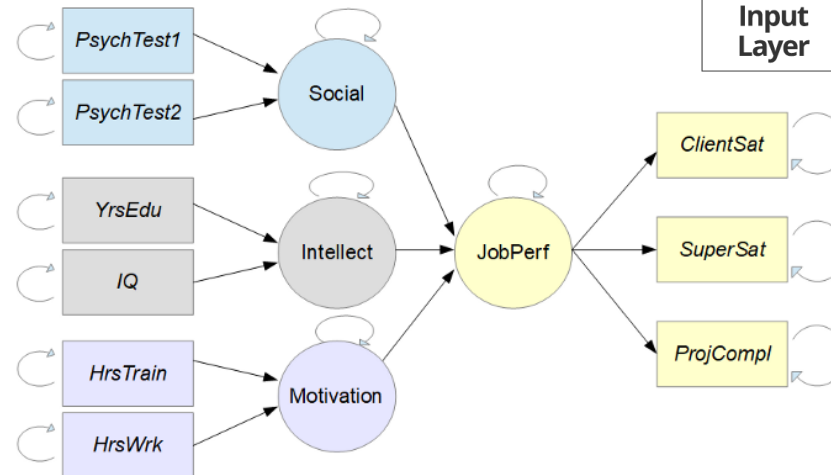
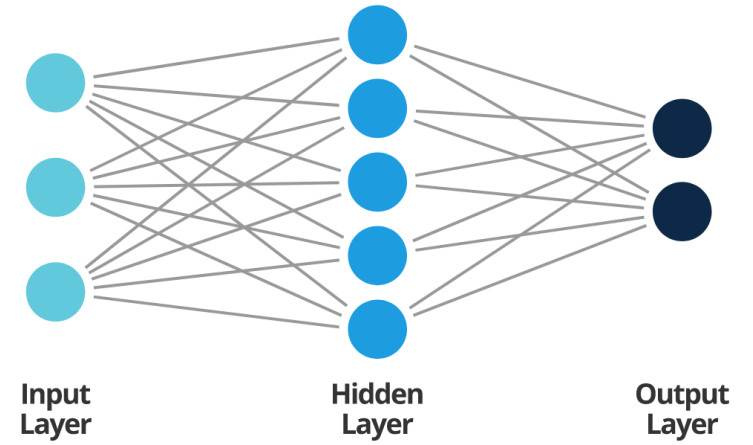
Structural Equations Modeling

Principal Component Analysis

And so many others....



Artificial Neural Network Architecture



Some other things

Word clouds

Neural networks

Structural Equations Modeling

Principal Component Analysis

And so many others....

Tableau (visualizations)

- Minimal coding skills required

Google charts (visualizations)

- Html background essential

Plotly (visualizations)

- Can use R, Python, or Matlab

Biggest takeaways

- Data doesn't **say** anything.
 - It can't lie, but it also can't tell the truth.
- Data Science is generally **prediction-driven**.
 - But theory and substantive knowledge will *constrain* its use
- Know **how** the data are acquired and **what** that may mean about conclusions you draw.
- Always go back to assumptions.

Biggest takeaways

- ***You*** are behind the wheel. Grab it.
- Your employer paid for the gas...
- *“Nothing in statistics or research methodology requires you to do something [misguided].”*
 - ~Modified from Lee’s Law #4
 - Addendum: “Keep it simple.”