The Cycle of Data

Data Dogs meets every first Monday of the month at Hawthorne Lucky Lab from 6-8pm
Run by: Ian Frantz(@IanFrantz) and John David Smith(@SmithJD)

Let's talk about everything that happens before you get data into your hands for clean-up, analysis, reporting and decision making.

- **Sources** – Where did the data come from?
    - Tools, people, devices, departments or systems.
    - Is there a record of a measurement plan, knowns, unknowns, assumptions?
    - What are the boundaries on your sources? (10 batches of 20 people testing a new feature could produce 1 positive batch just by random chance.)
    - Clarify similar terms, are "views" on Facebook equal to "views" on YouTube, Twitter, LinkedIn?
    - Are you combining things into your source? Will your results be combined with some other source later? (Combined data can reveal more than intended.)

- **Generation** - Data sets are the results of their means of collection.
    - If you collect data in two ways, you get two different results.
    - Data sets are not objective they are the subjective result of human decisions.
    - The same survey sent Monday at 8am, Wed at 12pm, Friday at 5pm will yield different results.
    - Be aware of attenuation, data collected over long distances can slowly lose signal.

- **Ownership** – Does the person, company, state or service own the data?
    - Who gains, benefits or is impacted by the data being collected? Challenge the implications and consequences.
    - What are the responsibilities of the beneficiary?
    - Can the data be destroyed? (Destruction test)
    - Are there incentives to shape the phenomena as an input for the measurement or model?

Group norms:
- We look forward to your participation in the discussion and as a member of this community we expect you to be **Open, Considerate,** and **Respectful**.
- This is a conversational environment -- not presenter-oriented. Everyone is invited to contribute.
- Mutual respect gets us to the nitty gritty inquiry much faster, so listen carefully!

Ian's Philosophical Tidbits:
- Goodhart's Law: "When a measure becomes a target, it ceases to be a good measure."
- Hyman's Maxim for data: Don't try to explain something with data, until you are sure there really is something there to be explained.
- Questioning Causality and David Hume:
    - If indeed causation is the basis for all empirical inference, then all empirical claims will follow from measuring causality. When A causes B, they are constantly conjoined in whenever we find A, we also find B. We have feelings of this certainty based on our past experiences. Be aware of your own limitations in inferring outcomes you have not experienced. Just because one event does follow another, it will not always mean there is a direct cause between the two.