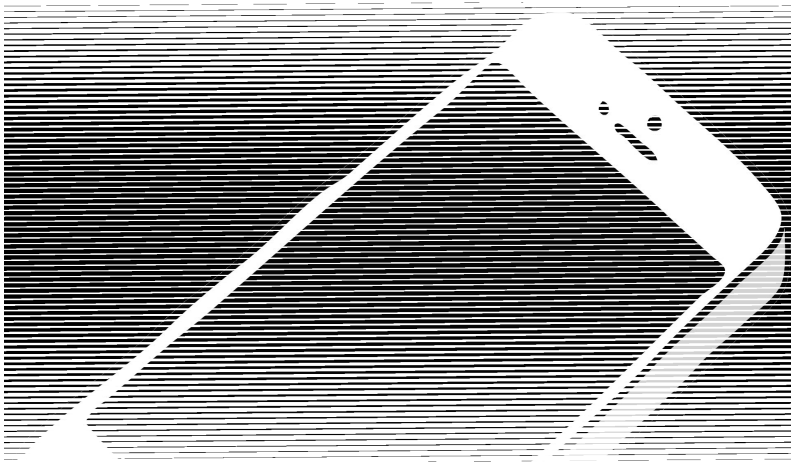
Data

Professor Ernesto Lee

Where is all this data?

- Cell Phones
- Cameras
- Log Files

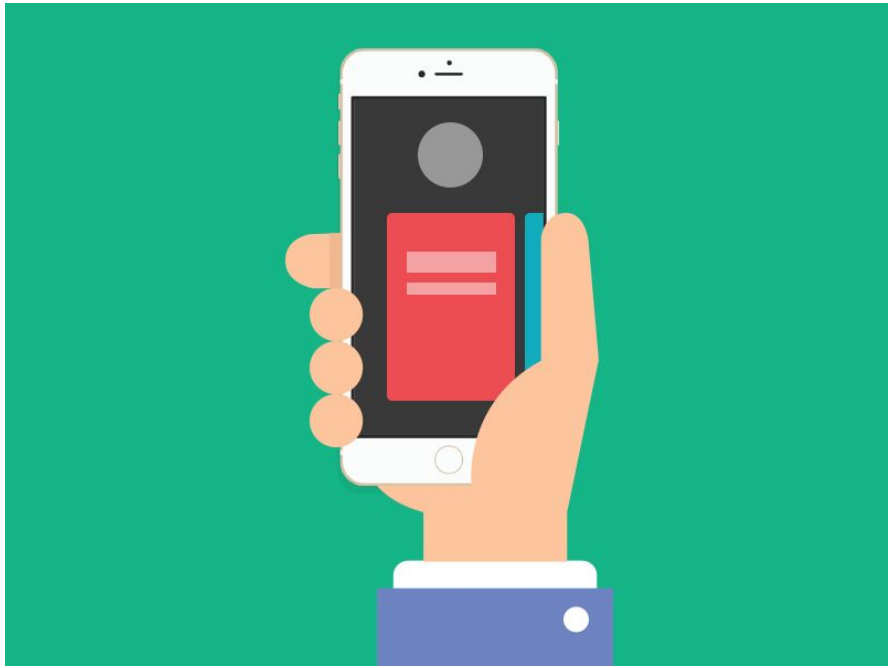
Let's Look at Cell Phone Data



Is your cel phone turned off?

IMEI

“The IMEI is the SSSN for the physical phone hardware.



IMSI

This is the SSN for your sim card. It gives you the right to use the number.

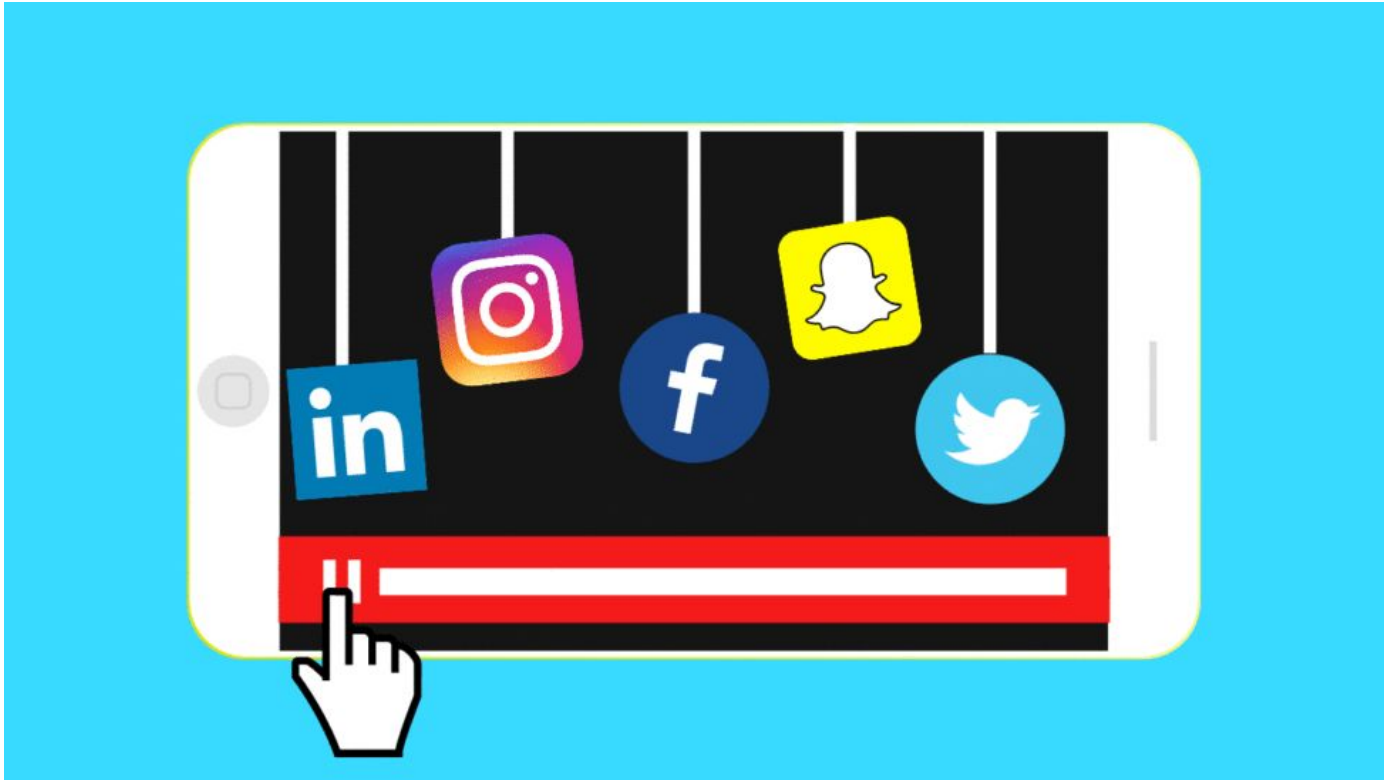
Cell Phone Tower

Your Phone / Tower / Identity

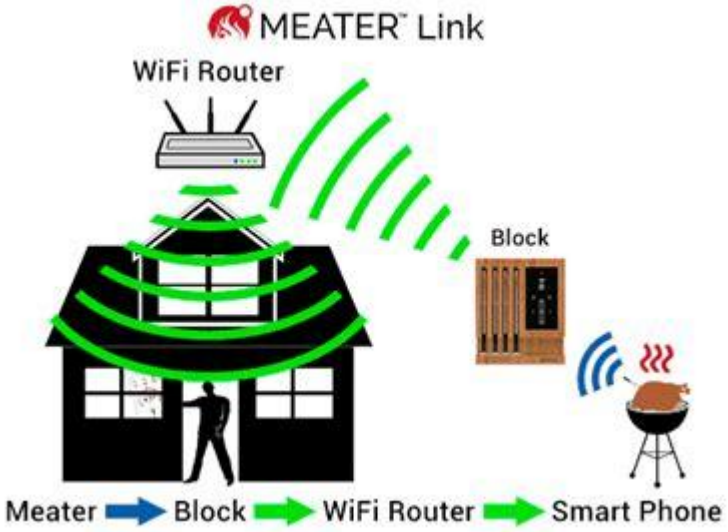
ALL OF THIS DATA IS BEING COLLECTED

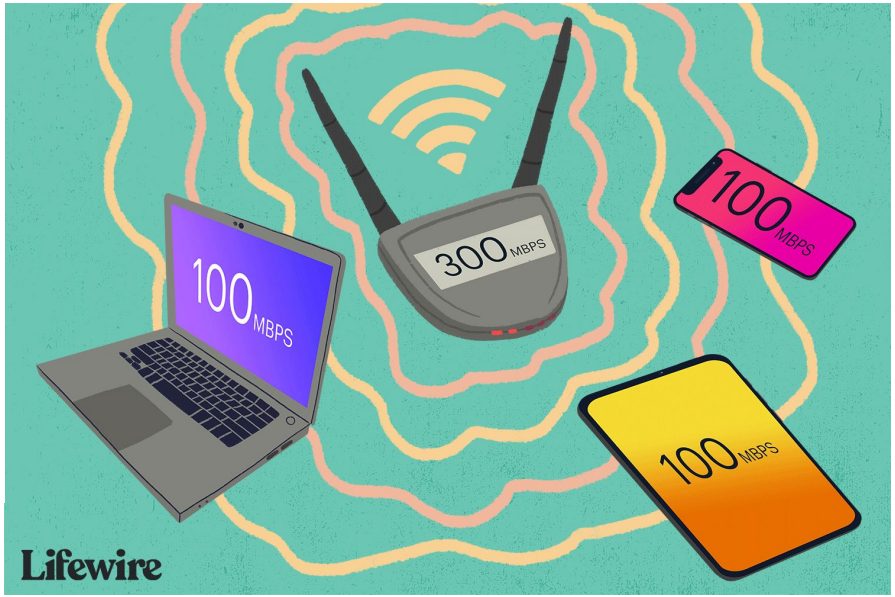
https://timeline.google.com/

What about the other Apps?

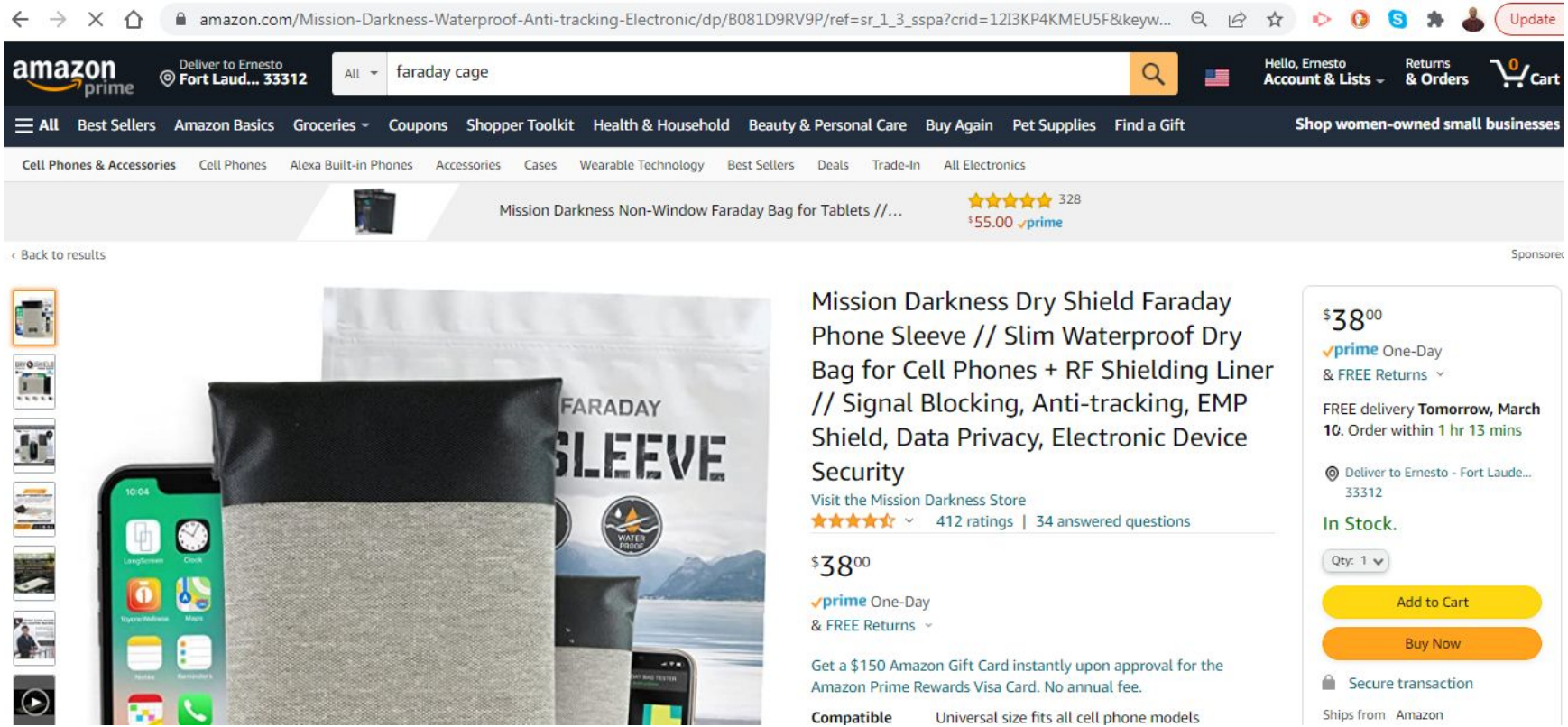


What about Wireless Access Points?





How to Mitigate Data Collection?

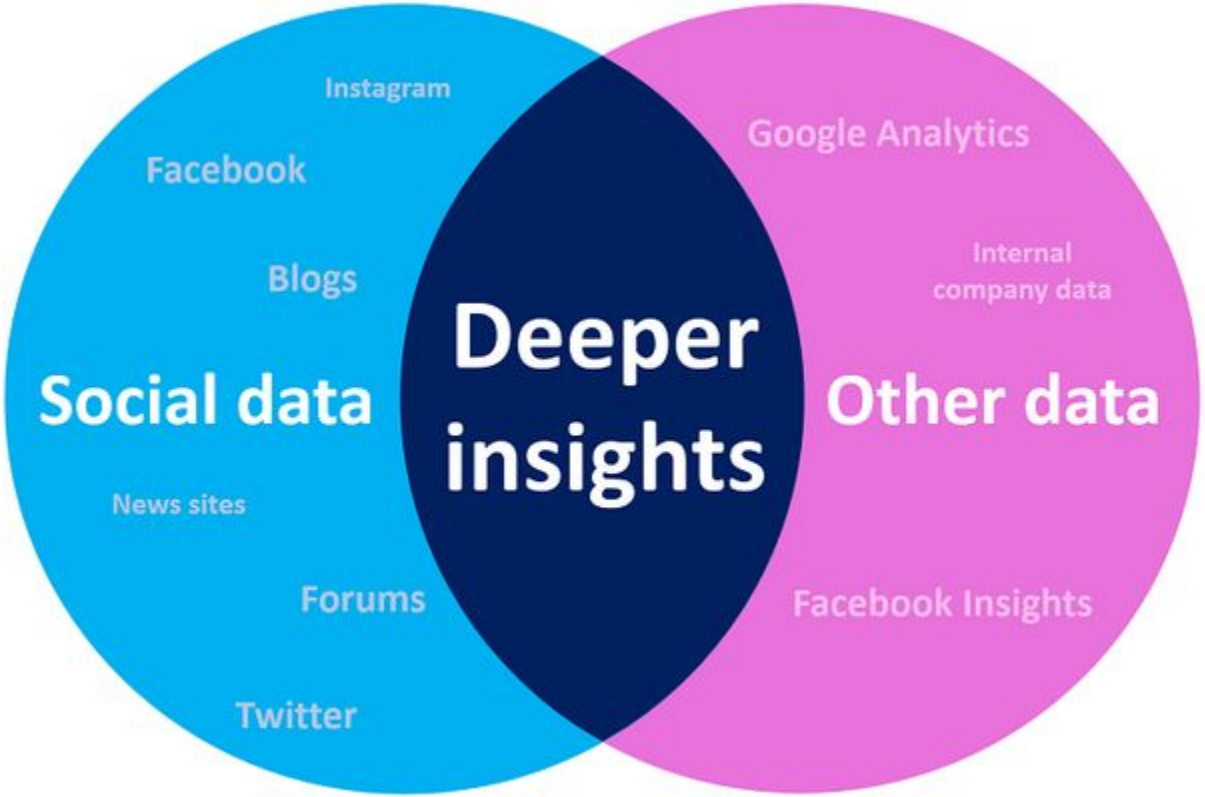


The Central Problem with Data

You don’t know what your phone is connected to…

Why Can't We Turn off FaceBook?

Your Data is Valuable



The Scandal

Third Party Doctrine

If your data goes to a third party server, that third party owns the data.

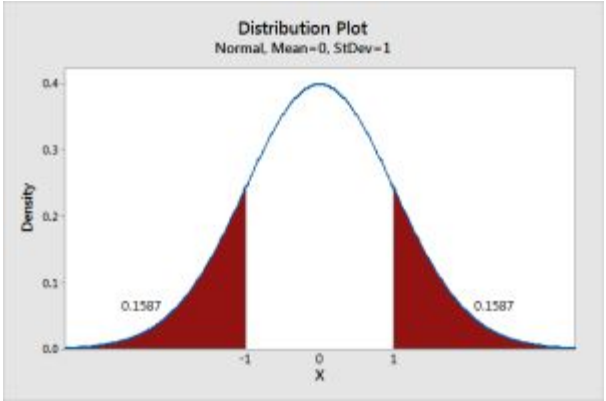
Conclusion

Week 2

The Importance of Data Analytics

- The field of analytics is the science of learning from data.
- analytical knowledge helps you use the proper methods to collect the data, employ the correct analyses, and effectively present the results.
- analytics is a crucial process behind how we make discoveries in science, make decisions based on data, and make predictions.

- Analysts are guides for learning from data and navigating common problems that can lead you to incorrect conclusions.
- Second, given the growing importance of decisions and opinions based on data, it’s crucial that you can critically assess the quality of analyses that others present to you.



Analytics Uses Numerical Evidence to Draw Valid Conclusions

- Analytics is not just numbers and facts.
  - You know, things like 4 out of 5 dentists prefer a specific toothpaste.
- Analytics is an array of knowledge and procedures that allow you to learn from data reliably.
- Analytics allow you to evaluate claims based on quantitative evidence and help you differentiate between reasonable and dubious conclusions.

How to be a good Analyst

- Producing reliable data.
- Analyzing the data appropriately.
- Drawing reasonable conclusions.

Avoid Common Pitfalls

Accuracy and Precision: Before collecting data, you must ascertain the accuracy and precision of your measurement system. After all, if you can't trust your data, you can't trust the results!

Biased samples: An incorrectly drawn sample can bias the conclusions from the start. For example, if a study uses human subjects, the subjects might be different than non-subjects in a way that affects the results. See: Populations, Parameters, and Samples in Inferential analytics.

Overgeneralization: Findings from one population might not apply to another population. Unfortunately, it's not necessarily clear what differentiates one population from another. analytical inferences are always limited, and you must understand the limitations.

Avoid Common Pitfalls

Causality: How do you determine when X causes a change in Y? analysts need tight standards to assume causality whereas others accept causal relationships more easily.

- When A precedes B, and A is correlated with B, many mistakenly believe it is a causal connection!
- However, you'll need to use an experimental design that includes random assignment to assume confidently that the results represent causality.
- Learn how to determine whether you're observing causation or correlation!

Incorrect analysis: Are you analyzing a multivariate study area with only one variable? Or, using an inadequate set of variables?

- Perhaps you're assessing the mean when the median might be a better?
- Or, did you fit a linear relationship to data that are nonlinear?
- You can use a wide range of analytical tools, but not all of them are correct for a specific situation.

Common Pitfalls

Violating the assumptions for an analysis: Most analytical analyses have assumptions. These assumptions often involve properties of the sample, variables, data, and the model.

- Adding to the complexity, you can waive some assumptions under specific conditions—sometimes thanks to the central limit theorem.
- When you violate an important assumption, you risk producing misleading results.

Data mining: Even when analysts do everything else correctly, they can produce falsely significant results by investigating a dataset for too long.

- When analysts conduct many tests, some will be analytically significant due to chance patterns in the data.
- Fastidious analysts track the number of tests performed during a study and place the results in the proper context.

Use Analytics to Make a Difference in your Field

- Data Analytics is used in almost all fields to make sense of the vast amount of data that are available.
- Even if the field of Analytics is not your primary field of study, it can help you make an impact in your chosen field.
- Chances are very high that you'll need working knowledge of analytical methodology both to produce new findings in your field and to understand the work of others.

Lies, Damned Lies, and analytics: Use Analytical Knowledge to Protect Yourself

- Unscrupulous analysts can use incorrect methodology to draw unwarranted conclusions.
- That long list of accidental pitfalls can quickly become a source of techniques to produce misleading analyses intentionally.
- But, how do you know? If you’re not familiar with analytics, these manipulations can be hard to detect. analytical knowledge is the solution to this problem.
  - Use it to protect yourself from manipulation and to react to information intelligently.

Anecdotal Evidence

- Wow! I took this supplement and lost a lot of weight! This pill must work!
- I know someone who smoked for decades, and it never produced any significant illness. Those claims about smoking are exaggerated!
- This anti-aging cream took years off. It must be the best!
- The vaccine doesn’t work
- <XYZ Race> are lazy

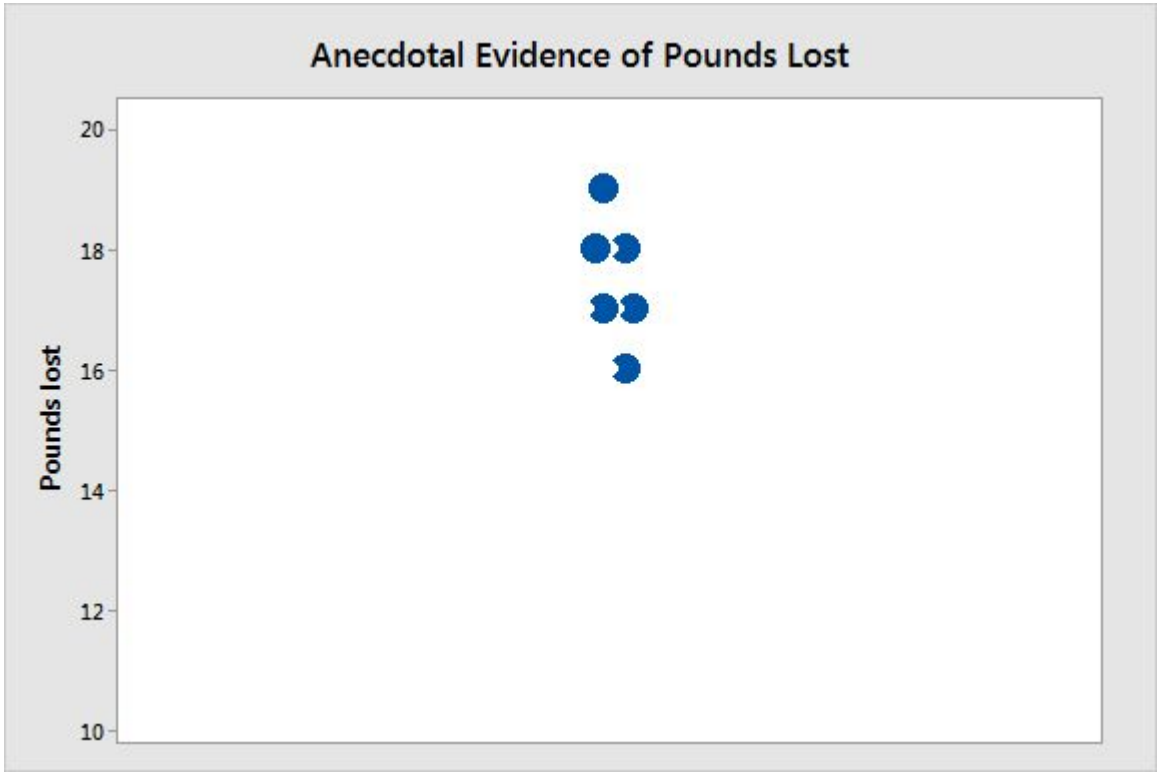
Analytical Methodology versus Anecdotal Evidence

Statistical methodology	Anecdotal evidence
Samples are large and representative. Typically, they are generalizable outside the sample.	Small, biased samples are not generalizable.
Scientists take precise measurements in controlled environments with calibrated equipment.	Unplanned observations are described orally or in writing.
Other relevant factors are measured and controlled.	Pertinent factors are ignored.
Strict requirements for identifying causal connections	Anecdotes assume causal relationships as a matter of fact.

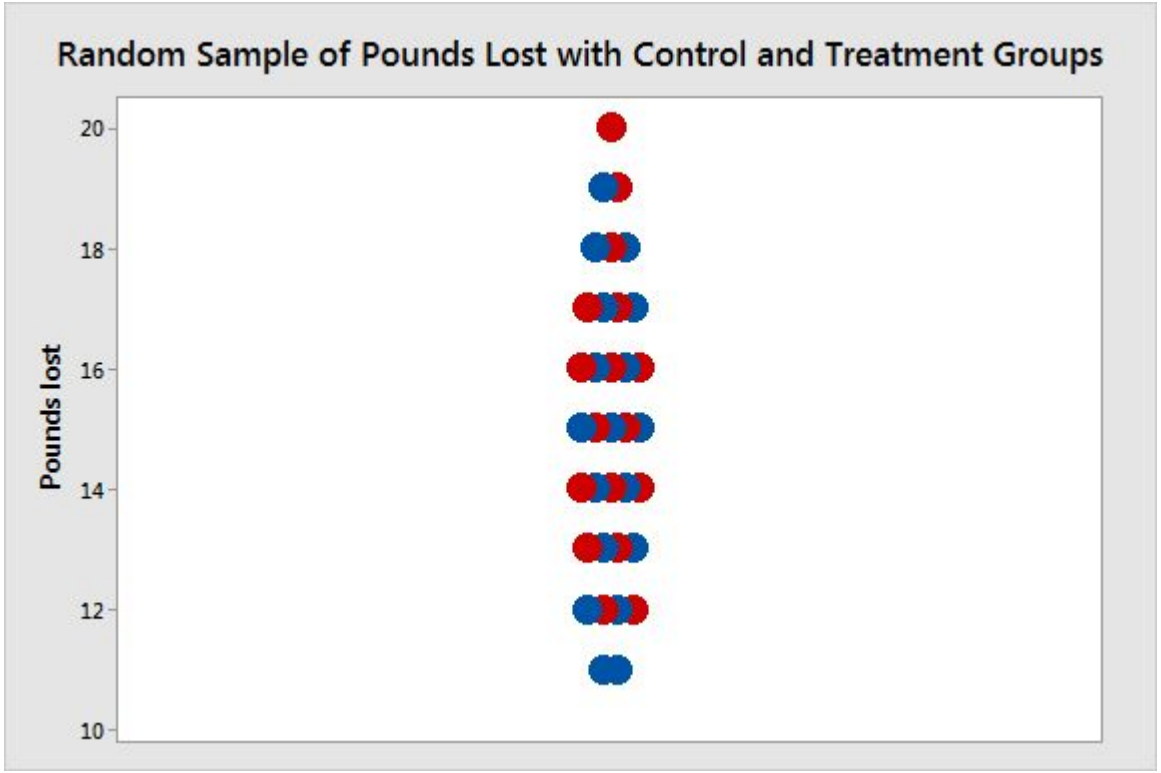
Why your anecdotal Evidence Sucks

- So, if Fred tells an emotional story about how he took a supplement and then lost a lot of weight, we'll remember Fred's story and assume that the supplement caused the weight loss.
- Unfortunately, we don't hear from the other 10 people who took the supplement and didn't lose weight.
- We also don't know what else Fred might be doing to lose weight.
- Collectively, these factors bias conclusions drawn from anecdotal evidence towards unusual outcomes and unjustified causal connections.

Graphically Illustrating the Shortcomings of Anecdotal Evidence



A SCIENTIFIC STUDY OF THE WEIGHT LOSS SUPPLEMENT



How Analytics Beats Anecdotal Evidence

In analytics, there are two basic methods for determining whether a dietary supplement causes weight loss: observational studies and randomized controlled trials (RCTs).

Has Anecdotal Evidence Ever Persuaded You?

- Making decisions based on anecdotal evidence might not always be harmful.
- For example, if you ask a friend for a restaurant recommendation, the risk is low, especially if you know his/her tastes.
- However, if you’re making important decisions about things like finances, healthcare, and fitness, don’t base them anecdotal evidence.
- Look at scientific data and expert analysis even though they’re not as flashy as emotionally charged stories presented by relatable people!
- If you find yourself being won over by anecdotal evidence, remind yourself that the results are not typical!