PAF 586 – Data for the Public Good
Lab 03 Solutions

PART 1 – Thin Slicing

1. Gottman's lab records 15-minute videos of each couple. How many data points are generated from those 15 minutes of footage? Stated differently, how many observations do the lab scientists record in each 15-minute interview?

   ***Gottman has taught his staff how to read every emotional nuance in people's facial expressions and how to interpret seemingly ambiguous bits of dialogue. When they watch a marriage videotape, they assign a SPAFF code to every second of the couple's interaction, so that a fifteen-minute conflict discussion ends up being translated into a row of eighteen hundred numbers—nine hundred for the husband and nine hundred for the wife.(*Note that they also collected heart rate and blood pressure data during the sessions.)**

   There are two lessons here. First, we might think we need to gather a lot of data to generate insight. Instead of gathering more data, how intensely can you use existing data?

   Second, often to predict human behavior we need to slow down time. There is no way a researcher could observe a conversation and code expressions in real-time at a rate of 1 per second. In many instances it might be hard to even observe a micro-expression without a pause or zoom.

   The insight comes from simplifying a complex world by distilling it into a few code categories, then looking for patterns in the codes. Is there any existing data within your organization that you could use to slow time and look for patterns by coding it in a systematic way?

2. What is the measured outcome in the study described by Gladwell? How would that data be collected? And consequently how long did these studies take?

   The study examined whether the couples were still married after 15 years, which would stretch the full study out for a decade and a half (or more likely the analysis was updated after 15 years to refine results after having better measures of outcomes). This attenuated timeline is really rare for a study that focuses on predictive accuracy.

3. Do you think that a marriage counselor working with couples for 20 years would be able to accurately predict those that will get divorced with 95 percent accuracy, relying on intuition from practice alone? What was unique about Gottman's approach that allowed him to achieve that kind of accuracy?

   Most marriage counselors would not be able to accurately predict the outcomes of couples they work with or those in the videos from Gottman's interviews. They would likely be better judges than the general public, but would not get anywhere near the accuracy of Gottman. There is a

broad and robust literature that demonstrates experts are rarely good at accurately predicting outcomes in their area of expertise.

The exception are experts that consistently gather empirical data, build models, and test their own assumptions. Gottman is a bit of an extreme here, coming to psychology from mathematics and physics. But he offers a good example of what is needed to improve at prediction. You can learn by thin-slicing data to slow down time. Or you can take the Charlotte Fludd approach and test new hypotheses each day. The famed Harvard business professor, Peter Drucker, suggests that leaders should make a practice of writing predictions down in a journal, and trying to find ways to improve accuracy over time ("Managing Oneself", 1999).

A good example of this is the recruiters in Moneyball. They were supposed to scout talent for the team, most were baseball players earlier in their lives, and all had many years of experience. They all had predictable limits to their ability to assess talent objectively because of their own biases and heuristics (they often paid more attention to height and size than performance statistics). The algorithm used by the A's was a way to correct for their inherent bias and better predict which players would help them win more games. This is ubiquitous across many professions – those with the most experience are rarely those that are best at accurately predicting outcomes!

Machine learning is distinctive from other quantitative social sciences because it focuses on prediction instead of causal explanations or theory. Data scientists are valuable to organizations because they are trained in building predictive models. Gottman and Charlotte Fludd are examples of managers or social scientists that achieve similar results without complex computational models.

Part II: Predicting Home Values

Identify one variable that predicts home value that is not included in the Zillow dataset. The obvious things are already present - square footage, number of bedrooms, whether there is a garage and a pool, etc. You need to be a little creative to come up with another factor.

Note that features in this case might be characteristics of houses themselves, but they also might be characteristics of neighborhoods or cities.

Variables suggested by classmates will be used in Lab 04 to demonstrate the process of feature engineering by trying to operationalize some of these measures.