Unit 5 Pre-Class Warm-up
Adam Yang

In our traditional approach to analyzing data, we have engrained procedures to first establish a model or hypothesis before we tackle the dataset. The idea that correlation does not equal causation drove us to figure out the underlying mechanism between variable X and Y, establish a model, which then will allow us to speak on the correlation with confidence. However, in the current era of technology, we have achieved drastically superior computational speeds, as well as the ability to store and collect petabytes of data. In this sense we can take the correlations at their face value without requiring a model to explain such correlations. The idea is that "with enough data, the numbers speak for themselves". I understand the author's argument and am inclined to agree that as the amount of data we have increases, we will have more confidence of any correlation we find, regardless of whether we know the root cause. However, the goals of the two approaches are different. Sometimes, when you analyze data, especially in the realm of science, the why is just as, if not more important than the what. It is not enough to be content with a correlation, the goal is to figure out why. In the example of Google using the computer algorithms to find correlations by sifting through petabytes of data, the goal is to just get a result. They care more about finding a correlation they can believe rather than the reason why the correlation exists. Therefore, I disagree with the author and believe that there will always be a reason to use models.