Approximate order of difficulty:

**Descriptive, Exploratory, Inferential, Predictive, Causal and Mechanistic**

**Descriptive** – The first kind of data analysis performed. The commonly applied to census data. The description and interpretation are different steps. Descriptions can usually not be generalized without additional statistical modeling.

**Exploratory –** Models are good for discovering new connections. They’re also useful for defining future studies. Usually not the final say and should not be used solely for generalizing or predicting.

**Inferential -** Commonly the goal of statistical models. It involves estimating both the quantity you care about and your uncertainty about your estimate.

**Predictive –** If x predicts Y it does not mean that X causes Y. Accurate prediction depends heavily on measuring the right variables. Although there are better and worse prediction models, more data and a simple model works really well. Prediction is very hard, especially about the future references.

**Causal –** Usually randomized studies are required to identify causation. There are approaches to inferring causation in non-randomized studies, but they’re complicated and sensitive to assumptions. Causal relationships are usually identified as average effects, but may not apply to every individual. Causal models are usually the “gold standard” for data analysis.

**Mechanistic –** Incredibly hard to infer, except in simple situations. Usually modeled by a deterministic set of equations. Generally the random component of the data is measurement error. If the equations are known but the parameters are not, they may be inferred with data analysis.

**What is Data** – Data are values of qualitative or quantitative variables, belonging to a set of items.

Data is actually the second most important thing; the question is the most important. Often the data will limit what question you can answer. But having data can’t save you if you don’t have a question.

**Formulate your question in advance.**