Below are the key ideas I would like you to understand from the first class.

1. Data science is a long history in various disciplines. A lot of mistakes have been made along the way, and we shouldn’t ignore them.
2. Data scientists have a toolkit that they apply in practice of data science. In this toolkit lie important concepts of *probability* and *statistics*.
   1. The important idea behind a *probability* is simply measuring the likelihood that some particular event occurs.
      1. For example, what is the *probability* that I get HEADS if I flip a US penny?
      2. Independence of events tells me that probabilities are not conditional on some event having occurred. For example, flip a US penny many times represents a series of independent events. Each new flip is not conditional on the prior flip. Therefore, the probability I get a HEADS if I flip a US penny does not depend on whether I flipped HEADS on the same coin five minutes ago.
      3. The only constraint we put on probabilities is that they lie between 0 and 1.
   2. *Random variables* are the tools data scientists use to operationalize our study of data.
      1. For example, a *random variable* is the outcome of flipping a US penny. Because I do not know the outcome of flipping a coin before it has occurred, the outcome can be view as a *random variable*.
      2. There are two types of *random variables*: discrete and continuous.
         1. For our purposes, discrete random variables are typically associated with outcomes that can take on only limited number of values. For example, the number of times you take the subway in a week.
         2. For our purposes, continuous random variables are typically associated with outcomes that can take on a large range of values. For example, we might view the weight of a large population of elephants to be a continuous random variable.
      3. There are key ways to summarize random variables.
         1. The mean (or average or expectation).
         2. The variance (or dispersion), which is always positive. The larger is the variance, the greater is the dispersion.
         3. The standard deviation (the square root of the variance), which is always positive.
         4. The covariance of two random variables, which measure the tendency for the random variables to “move together”.
         5. The correlation of two random variables, which we construct so that it lies between -1 and 1. (Recall the three graphs).
         6. Correlation is a linear concept which may not accurately reflect the world.
3. Data scientists represent the correlations or relationships that they want to analyze using a model of some sort. An example of such a model is the equation of a line. They combine the model with data and a statistical or machine learning algorithm to analyze the data they have obtained.
4. These algorithms are typically derived from a process of mathematical optimization, which we use as a tool. Nowadays, we let the computer do the optimizing, and the purpose of discussing optimization was to show you how this is done in practice.
5. To introduce us to the tools we use for applied data science, we started with simple bivariate linear model: . In this example, y is the random variable of interest, such as the wages earned by a sample of young people, and x is the attribute we want to explain the wages of young people. An example of an attribute is education.
6. While it is an abstraction from reality, there are a number of reasons we start with a linear model:
   1. The linear model helps to motivate many other models, including Support Vector Machines.
   2. We will come to a detailed appreciation for the quality of the statistical estimates we derive from the linear model.
   3. It allows us to explore key issues such as causation (rather than correlation).
   4. Technically, the linear model can be seen as a first-order Taylor series of any arbitrary model (when we ignore higher order terms).