Dealing with missing or incomplete data

How we treat missing and incomplete data is important because it can adversely affect the results of our analysis. A profit increase of 10% means little if we are missing 20% of our sales data. Luckily we have a number of methods in which we can deal with missing or incomplete data.

Methods
There are four primary methods of treating missing or incomplete values.
1. **Lookup.** Where possible, we can input missing values by looking up the correct values from the same table or from other sources.
    ○ Example - the missing **Region** values in rows 2&8 of the example table below are properly recorded elsewhere in the table, enabling us to lookup the correct missing values.
2. **Remove.** Another option is to simply remove the rows with missing data. This is particularly appropriate if many variables are missing from a row. You should only consider this approach if the missing values are found to be random because if missing data only occurs in just one subset of your data, removing these rows would skew your results.
3. **Inpute.** With this method, you calculate the missing values by using the other values in the table. There are a number of ways to achieve this such as by replacing the missing values with the *median*, *average* or *mode* of related values.
    ○ Example - In row 5 of the example table, we could replace the missing **Sales** value with the *average* price of storage items.
4. **Annotate.** Rather than calculating the missing value, we could instead annotate it as missing. We should annotate any missing numerical value with a dash rather than a 0 because a 0 would still affect any summary of the data.

Row ID	State	Region	Sub-Category	Product Name	Sales
1	Kentucky	South	Bookcases	Somerset Bookcase	$ 261.96
2	Kentucky		Chairs	Deluxe Fabric Upholstered Stacking Chairs, Rounded Back	$ 731.94
3	California	West	Labels	Address Labels for Typewriters by Universal	$ 14.62
4	Florida	South	Tables	Bretford CR4500 Series Slim Rectangular Table	$ 957.58
5	Florida	South	Storage		
6	California	West	Furnishings	Eldon Expressions Wood and Plastic Desk Accessories, Cherry Wood	$ 48.86
7	California	West	Art		$ 7.28
8	California		Phones	Mitel 5320 IP Phone VoIP phone	`

Figure 1 - example table

Summary

While some of these methods have more risks than others, there is no right method for all situations. Good record-keeping (such as the inclusion of a data dictionary) is essential because it ensures that the methods underlying your findings are understood and accepted.