One of the most important skills for you to have as a data analyst is to be able to describe your data so that others are able to understand the population your sample came from, the procedures that were used to collect the data, and the measures that you used in your statistical analyses. This background information is critical for understanding who or what you're analyzing and how you're analyzing the data. It is also important for helping you and others understand the limitations of your analysis.

Every report that you write about your data analysis should include a Methods section that describes how the research was conducted. The Methods section usually comprises discussions of your sample, measures, and procedures.

Sample. Identify who or what was studied (people, animals, etc.). Identify the level of analysis studied (individual, group, or aggregate). Describe observations vividly so your reader can distinguish them clearly. If you group observations, use meaningful names ("Low-Income Women") rather than abbreviations ("PPM100") or variable names ("INCOME_GRP"). The following is successful section of a sample description:

> The sample of 1,203 pregnant women was drawn from two public prenatal clinics in Texas and Maryland. The ethnic composition was African American (n = 414, 34.4%), Hispanic, primarily Mexican American (n = 412, 34.2%), and White (n = 377, 31.3%). Most women were between the ages of 20 and 29 years; 30% were teenagers. All were urban residents, and most (94%) had incomes below the poverty level as defined using each state's criteria for Women, Infants, and Children (WIC) eligibility.

This sample description is successful because it identifies both the observations (1,203 pregnant women) and the location (two prenatal clinics in Texas and Maryland). Furthermore, it describes the composition of the group ethnically and by income using language consistent with writing standards for the empirical research.

Procedures. Explain what participants/observations experienced. Discuss whether data were collected by surveillance, survey, experiment, or another method. Discuss where data were collected and the period over which they were collected. The following is a successful section of a procedures discussion:

> Random sampling was used to recruit participants for this study. Surveyors went to considerable lengths to secure a high completion rate, including up to four call-backs, letters, and monetary incentives. Trained researcher assistants conducted face-to-face interviews with all study participants. Sensitive questions about substance use and sexual behavior were asked using computer assisted interviewing to increase the reliability of responses.

Measures. Describe the questions or measures of your participants/observations and relate these to the type of data you collected (quantitative or categorical). Again, provide meaningful descriptions ("number of cigarettes smoked per day"), rather than variable names ("SQB8A") that will have no meaning to the reader. Discuss how you managed the measures for your analysis. The following is a successful section of a measures discussion:

> The measure of tuberculosis (TB) was drawn from country level surveillance data compiled by the World Health Organization in their Global Tuberculosis Database (www.who.int/globalatlas/dataQuery/default.asp), and made available for download through the Gapminder web site (www.gapminder.org). It measures the estimated number of new TB cases (all forms) among 100,000 residents in each country during 2008. For the current analysis, it was binned into four categories based on a quartile split.

This discussion is successful because it indicates what the variable measured (estimated number of new TB cases (all forms) among 100,000 residents), and how the data were managed (binned into four categories based on a quartile split for the study analysis).