**CHAPTER 2**

**Answers to End of Chapter Problems**

2.1 Cross Section – Answers vary by student – for example a sample of 200 firm’s profits in 2013, a sample of 50 state’s population, etc.

Time Series – Answers vary by student – for example labor force participation from 1980 to 2010, productivity monthly from 1994 to 2013.

Panel Data – Answers vary by student – for example high school and beyond that surveys students which samples the same individuals from when they are in high school and then every two years for 20 years.

2.2 a. Answers vary by student – for example, to find out how many hours of exercise students at your campus get.

b. Stand on campus and ask every third individual, use survey monkey or a facebook survey

c. The email survey to friends will suffer from selection bias as their responses will likely be similar to yours while surveying people at a local grocery store will still suffer from selection bias but not by the same degree

d. If the survey is biased in some manner then the analysis that is based on the survey will also be biased (not truly representative of the population).

2.3 Data management techniques are important because simple mistakes with data can lead to hours upon hours of work to reconstruct a data set. The goal of data management is to” (1) reduce the chances of overwriting existing data, which would require us to spend unwanted hours re-creating our data from the beginning and (2) provide clear intuition as to our process, thereby enabling us to quickly get back up to speed if we return to our project after a long hiatus.

We recommend saving one master file so that if we make a mistake and overwrite or otherwise change our data, we can easily go back and reconstruct our correct data without having to start at square one with our internet search, data downloading, and so on.

We recommend including a worksheet identifying all data sources so that if we make a mistake and overwrite or otherwise change our data, we can more easily go back and reconstruct our exact data set without having to re-perform internet searches, which can at later dates provide different results and make our reconstructing the data nearly impossible.

We recommend making file and variable names as intuitive as possible because we are often forced to put our project aside for longer periods of time and we need to be able to get back up-to-speed as quickly as possible when returning to it. If our file and variable names are not intuitive, then doing so is much more difficult and time-consuming.

Calculations often require changing the values of existing data. Unfortunately, when making such changes, it is possible to change the values of the initial data and when doing so, it is possible to make mistakes that cannot be easily undone. If we make such mistakes in a new worksheet, we can easily go back to the initial data and start over again, making sure not to make the same calculation mistake again.

**Answers to End of Chapter Exercises**

E2.1



GDP per capita is smallest in 1992 and largest in 2012. It is out of order in 2007-2010 (due to the recession).

E2.2 a-d.



e.



E2.3 Please see uploaded spreadsheet for data details.



E2.4 a. highest Equatorial Guinea

lowest Zimbabwe

b. highest Mauritius

lowest it was a tie for Botswana, Mauritania, and Nambia

c. The sum function was used to obtain both of these values.