

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

Year	GDP (billions of dollars)	GDP (millions of dollars)	Population (million of people)	GDP per Capita
1992	6539.3	6539300	256.51	25493.35309
1993	6878.7	6878700	259.92	26464.68144
1994	7308.7	7308700	263.13	27776.00426
1995	7664	7664000	266.28	28781.73351
1996	8100.2	8100200	269.39	30068.67367
1997	8608.5	8608500	272.65	31573.44581
1998	9089.1	9089100	275.85	32949.42904
1999	9665.7	9665700	279.04	34639.11984
2000	10289.7	10289700	282.16	36467.60703
2001	10625.3	10625300	284.97	37285.67919
2002	10980.2	10980200	287.63	38174.73838
2003	11512.2	11512200	290.11	39682.18951
2004	12277	12277000	292.81	41928.21283
2005	13095.4	13095400	292.81	44723.19934
2006	13857.9	13857900	295.52	46893.27287
2009	14417.9	14417900	306.77	46999.05467
2007	14480.3	14480300	301.23	48070.5773
2010	14958.3	14958300	309.35	48353.968
2008	14720.3	14720300	304.09	48407.70824
2011	15533.8	15533800	311.59	49853.33291
2012	16244.6	16244600	313.91	51749.22749

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

Chapter 02 - Collection and Management of Data

E2.2 a-d.

City	Unemployment Rate	Population	Total Unemployed
New York City	7.8	8,336,697	650262
Los Angeles	8.6	3,857,799	331771
Chicago	8.3	2,714,856	225333
Phoenix	6.8	1,488,750	101235
Houston	5.9	1,488,750	87836
Phiadelphia	7.6	1,547,607	117618
		Total	1514056

e.

City	New York City	Los Angeles	Chicago	Phoenix	Houston	Phiadelphia	
Unemployment Rate	7.8	8.6	8.3	6.8	5.9	7.6	
Population	8336697	3857799	2714856	1488750	1488750	1547607	Total
Total Unemployed	650262.366	331770.714	225333.048	101235	87836.25	117618.132	1514055.51

E2.3 Please see uploaded spreadsheet for data details.

Category	City	Amount	Year
Least miles Traveled	Brownsville TX	110	1982
Most miles Traveled	Los Angeles-Long Beach-Santa Ana CA	139,275	2004
Lowest Gas Cost	Atlanta GA	0.93	1998
Highest Gas Cost	Los Angeles-Long Beach-Santa Ana CA	3.84	2008

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.