

Big Data Case Answers for Chapter 2

1. Cross Tabulation Tables:

- a.) Region and Control – the following crosstabulation table was constructed using the Pivot Table in Excel. In the database, both Region and Control have been converted to categorical values that were then used in the creation of this table. The analyst can see the number of hospitals in the various regions, the number of hospitals under the various types of control, and the intersection of each. Pie charts and other visuals can be used to display the various types of control, for example, in the Mid Atlantic region.

Count of Control		Column Labels				
Row Labels	Gov Fed	Gov		InvestorOwned		Grand Total
		NonFed	ForProfit	NonProfit	Nonprofit	
East North Central	7	36	46	38	135	262
East South Central	2	42	34	10	48	136
Mid Atlantic	8	20	18	11	111	168
Mountain	6	28	29	15	82	160
New England	3	6	10	3	78	100
Pacific	2	48	31	38	140	259
South Atlantic	9	44	46	10	143	252
West North Central	6	54	32	27	95	214
West South Central	3	64	233	41	108	449
Grand Total	46	342	479	193	940	2000

b.) Region and Service – This Excel-produced crosstabulation table between region and service was created using the Pivot Table. It shows which types of service dominate the hospital database (General Medical & Surgery, followed by Psychiatric, and Long-term Acute), and how the types of service breakdown by Region.

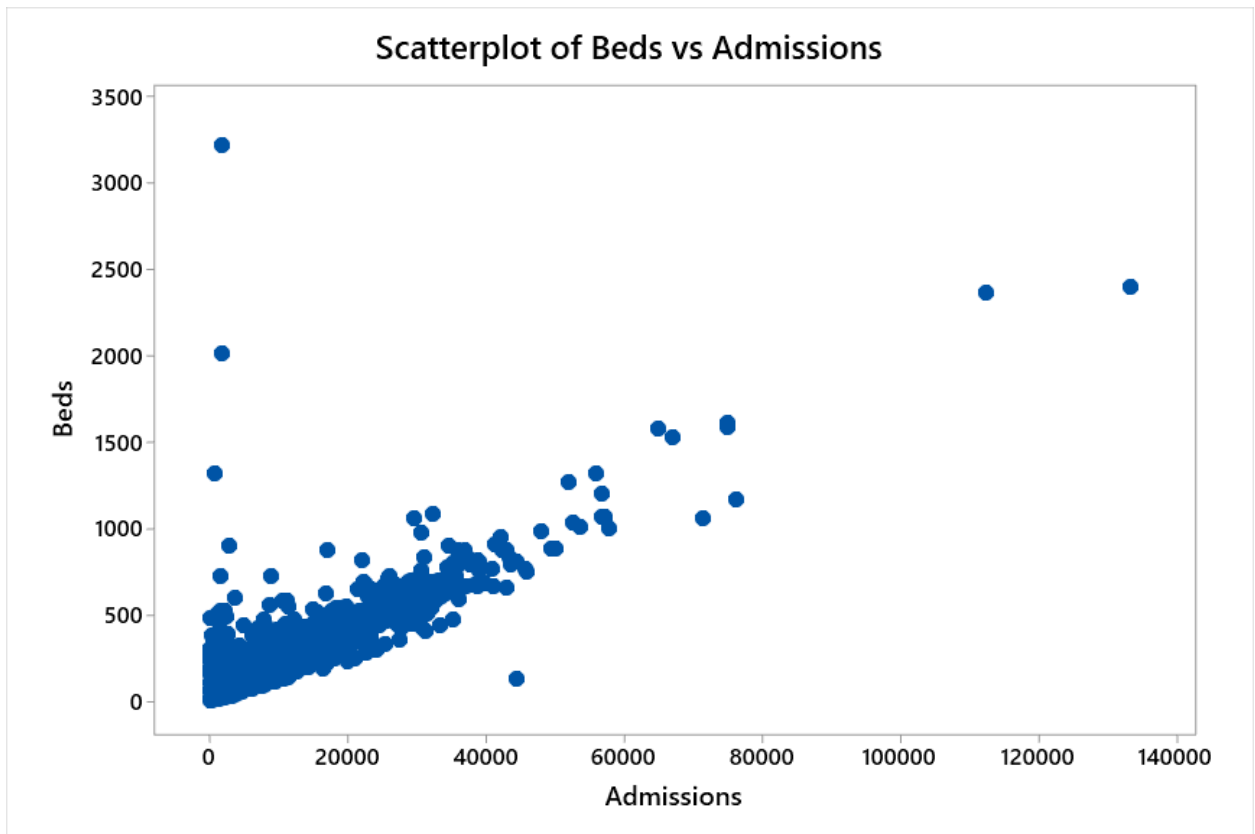
Row Labels	East North Central	East South Central	Mid Atlantic	Mountain	New England	Pacific	South Atlantic	West North Central	West South Central	Grand Total
Alcoh & Chem Dependency	3						3	1		7
Cancer	2		2			3	1		1	9
Children's	10	3	6	5	5	14	4	12	22	81
Chronic Disease					1		1			2
Gen Med & Surg	170	101	112	127	65	218	179	147	197	1316
Heart	3								4	7
Intellectual Disab								2		2
Long-term Acute	17	6	9	3	5	3	18	17	86	164
Obstetrics & Gyn	1	1	1		1	1			3	8
Orthopedic	2		2	1	1		1	1	4	12
Other Specialty			2	1	1	1	5	2	1	13
Psychiatric	38	14	19	12	13	13	20	18	58	205
Rehab	15	11	14	9	8	4	19	11	58	149
Surgical	1		1	2		2	1	3	15	25
Grand Total	262	136	168	160	100	259	252	214	449	2000

- c.) Service and Control – This Excel-produced crosstabulation table between service and control was created using the Pivot Table. It shows which types of service dominate the hospital database (General Medical & Surgery, followed by Psychiatric, Long-term Acute, and Rehab), and how the types of service breakdown by type of control. For example, Government Federal Control (mostly veteran’s hospitals) are mainly General Medical & Surgery hospitals. Whereas, the modal control of rehab hospitals is Owned by Investors and are For Profit.

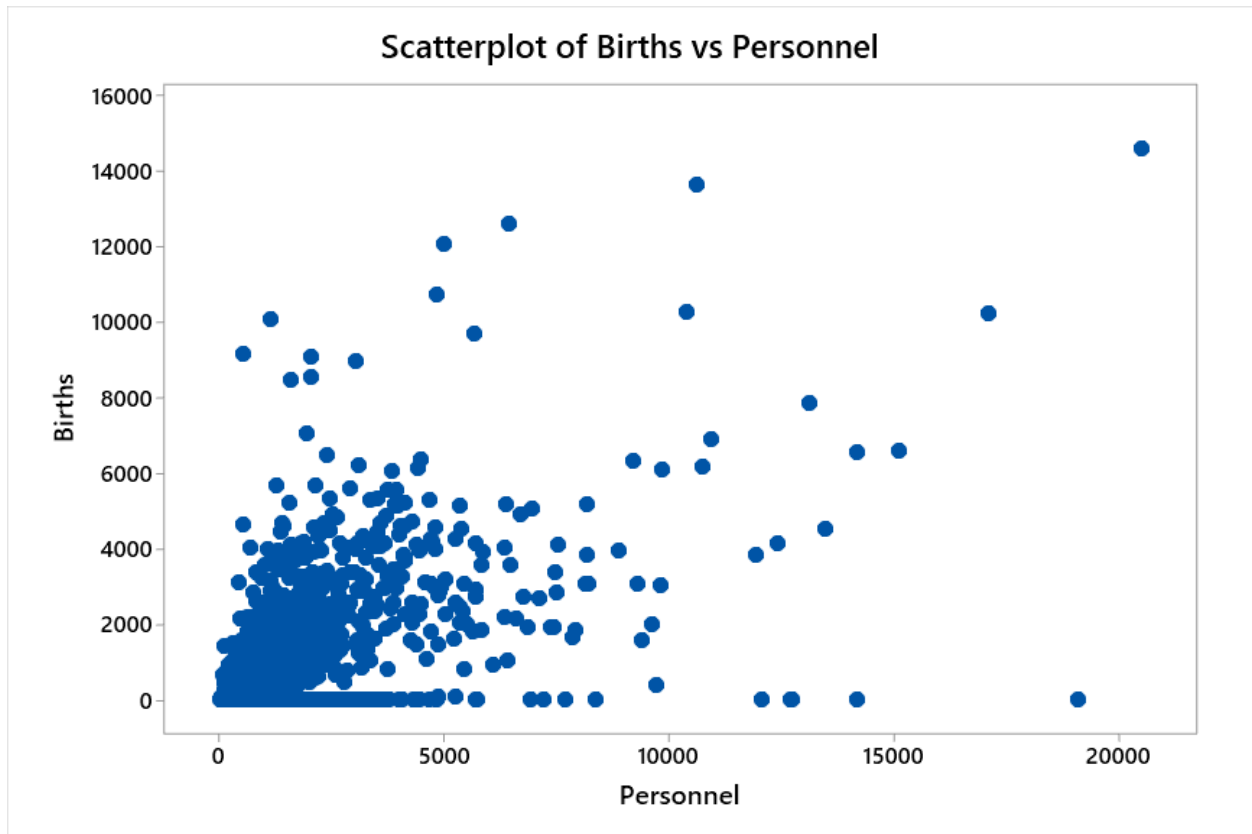
Count of Control		Column Labels				
Row Labels	Gov Fed	Gov NonFed	InvestorOwned ForProfit	NonGov NonProfit	Other Nonprofit	Grand Total
Alcoh & Chem Dependency		1	3		3	7
Cancer		3			6	9
Children's		2	11	2	66	81
Chronic Disease		1			1	2
Gen Med & Surg	43	249	127	167	730	1316
Heart			6		1	7
Intellectual Disab		2				2
Long-term Acute		6	113	13	32	164
Obstetrics & Gynecology			3		5	8
Orthopedic			8		4	12
Other Specialty	1	2			10	13
Psychiatric	2	70	81	4	48	205
Rehab		4	107	6	32	149
Surgical		2	20	1	2	25
Grand Total	46	342	479	193	940	2000

2. Four Different Scatter Plots – As an example, presented here are four different scatter plots from this AHA large database.

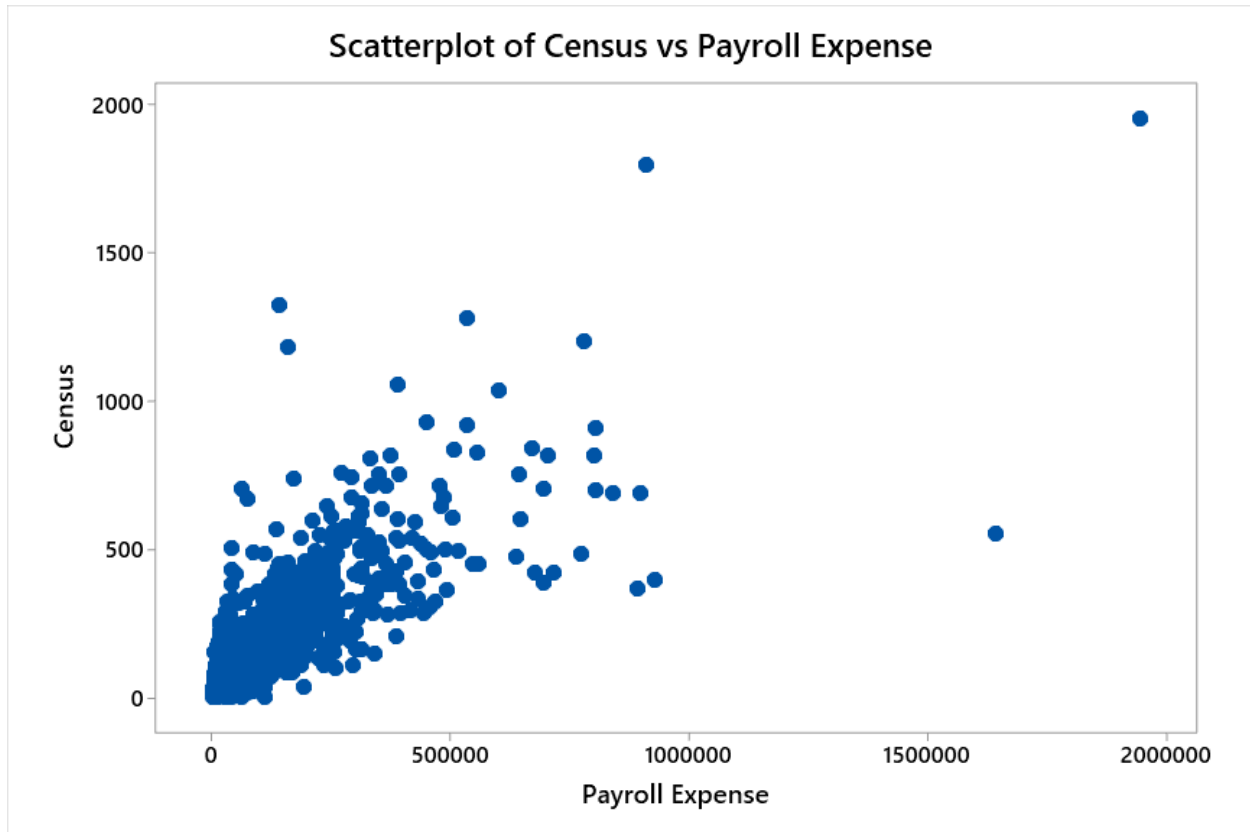
a.) Shown here is a Minitab-produced scatterplot of Number of Beds vs. Number of Annual Admissions. Of the four scatter plots, this one between beds and admissions, most indicates a potential strong correlation.



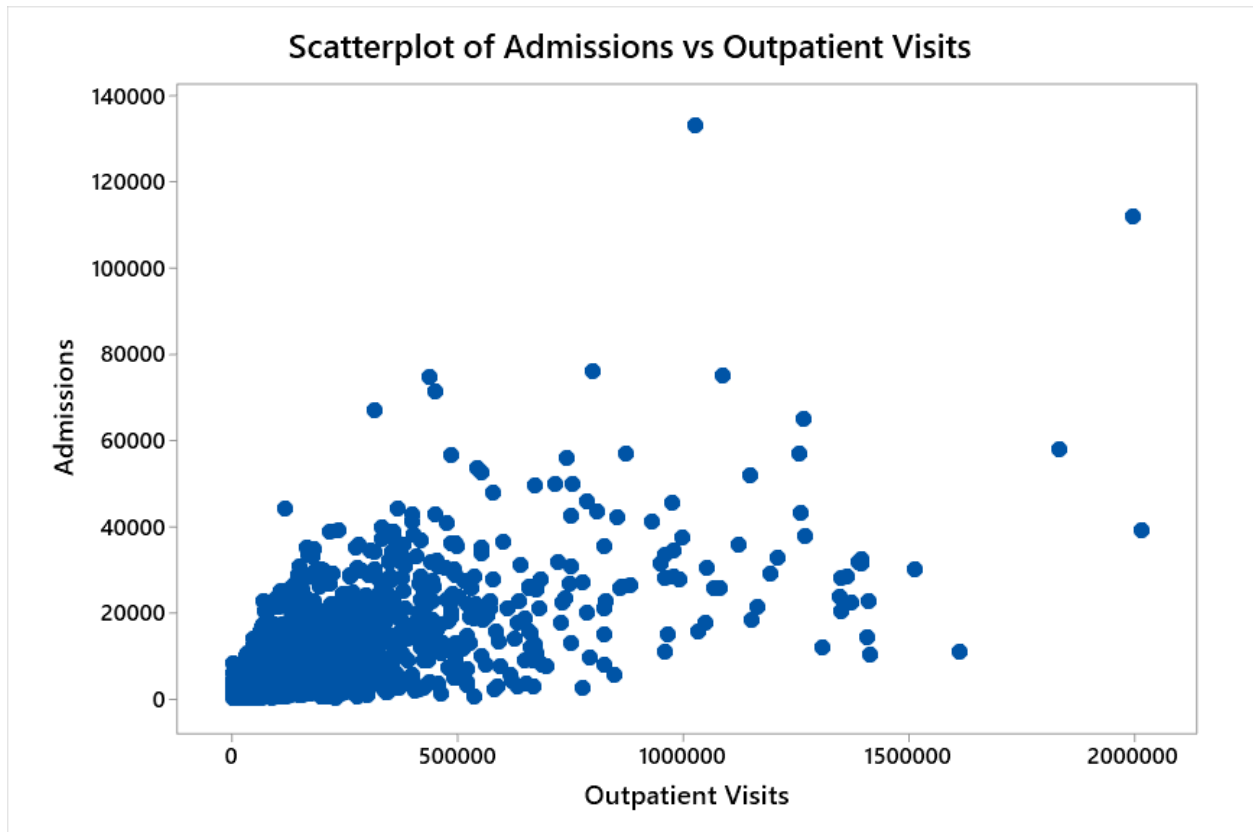
- b.) Shown here is a Minitab-produced scatterplot of Number of Births vs. Number of Personnel. This scatter plot indicates that the relationship between births and personnel is probably the weakest of the four scatter plots.



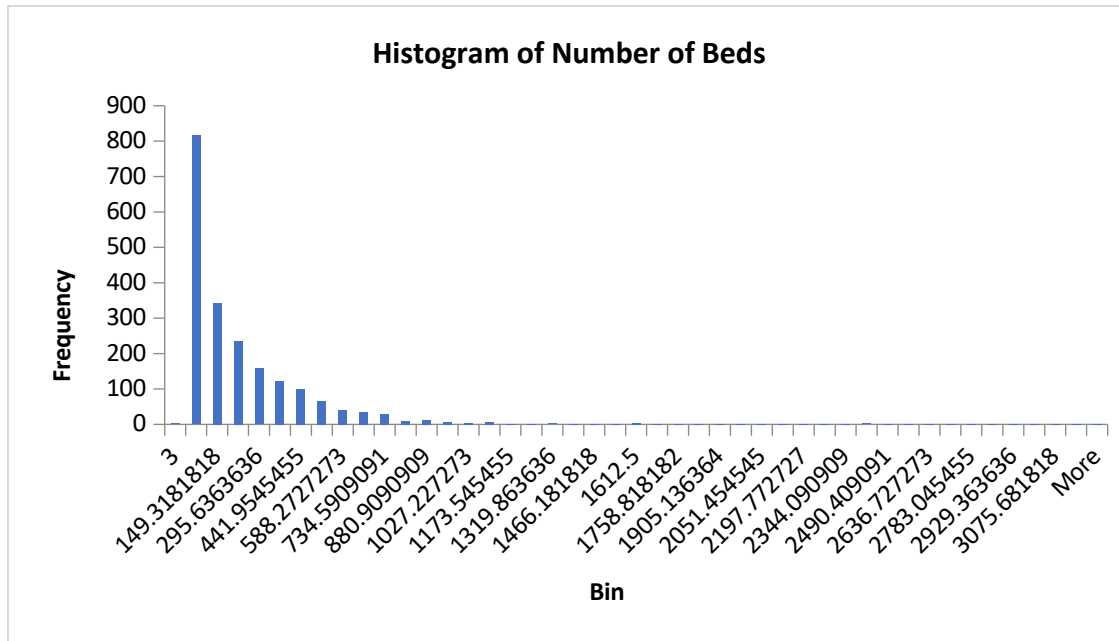
- c.) Shown here is a Minitab-produced scatterplot of Census vs. Payroll Expense. This scatter plot indicates that the relationship between census and payroll expense is stronger than births vs. personnel and weaker than beds vs. admissions.



- d.) Shown here is a Minitab-produced scatterplot of Admissions vs. Outpatient Visits. This scatter plot indicates that the relationship between admissions and outpatient visits is about as weak as for births vs. personnel.

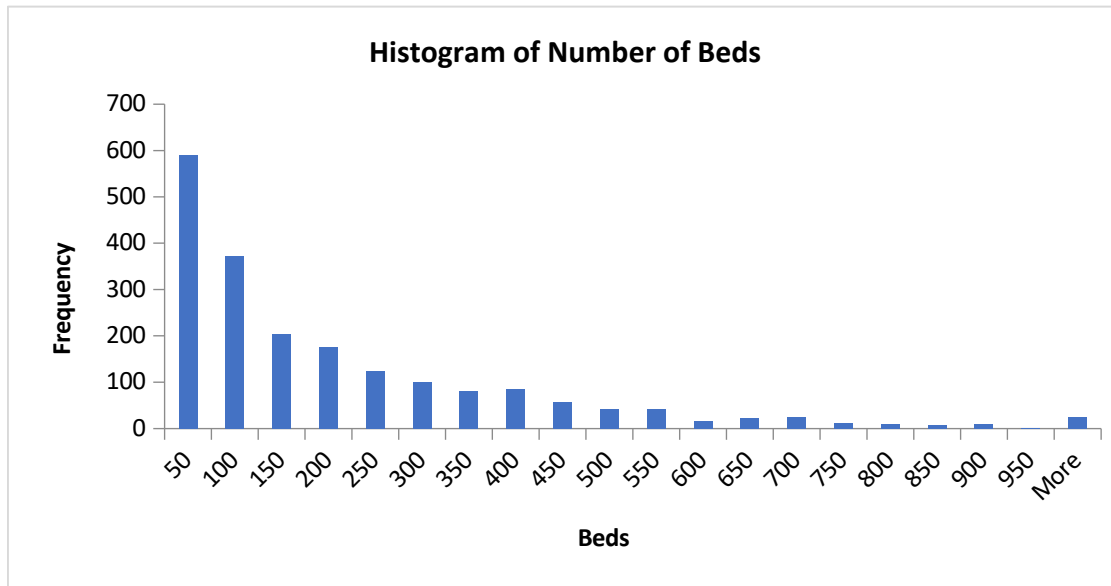


3. Histograms of Number of Beds. Shown first is a histogram of Number of Beds constructed from Excel. For this histogram, Excel created the bins.

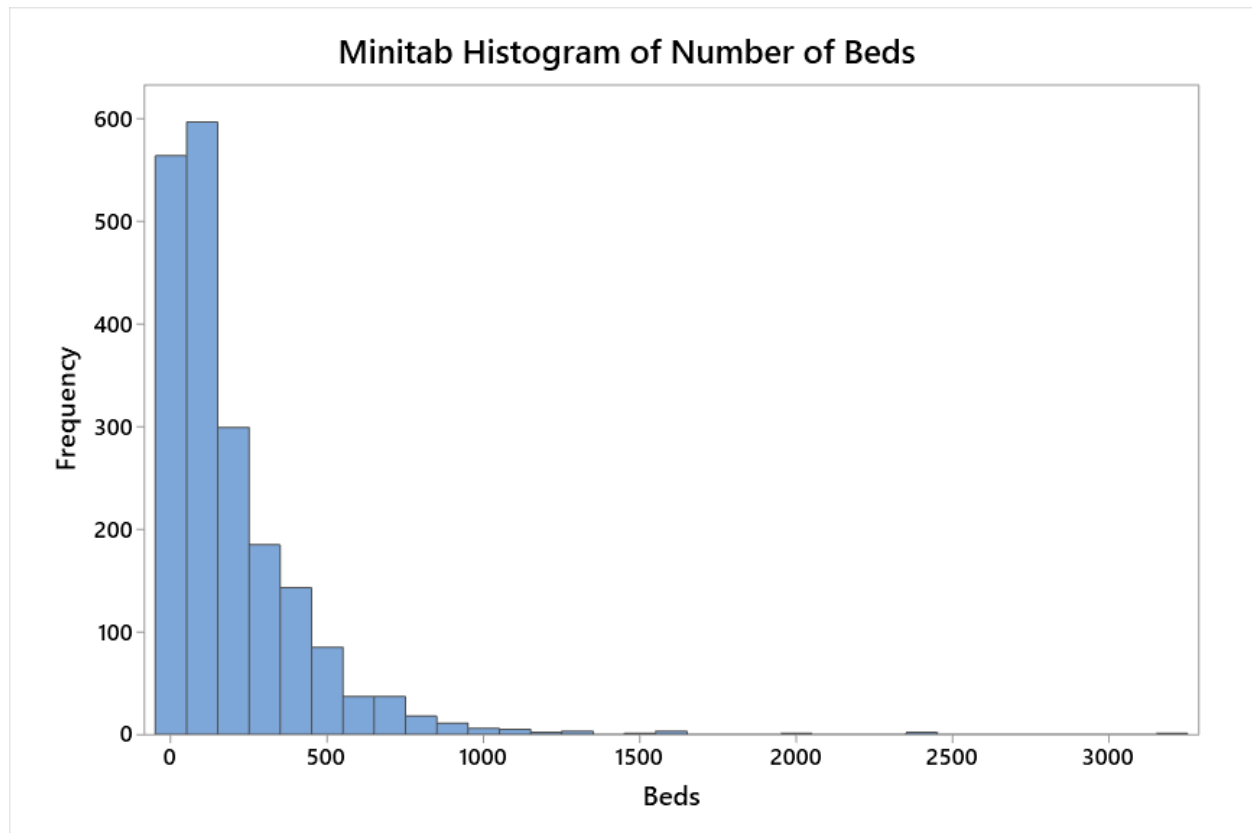


This shows that a few large hospitals are pulling the class intervals way to the right. There is a large “spike” of hospitals probably around 50 beds.

Next, we provided Excel with bins of 50, 100, 150, 200, 250, ..., 950. The resulting histogram makes the distribution a little easier to see the breaks in number of beds classes with many of the hospital having less than 150 beds.



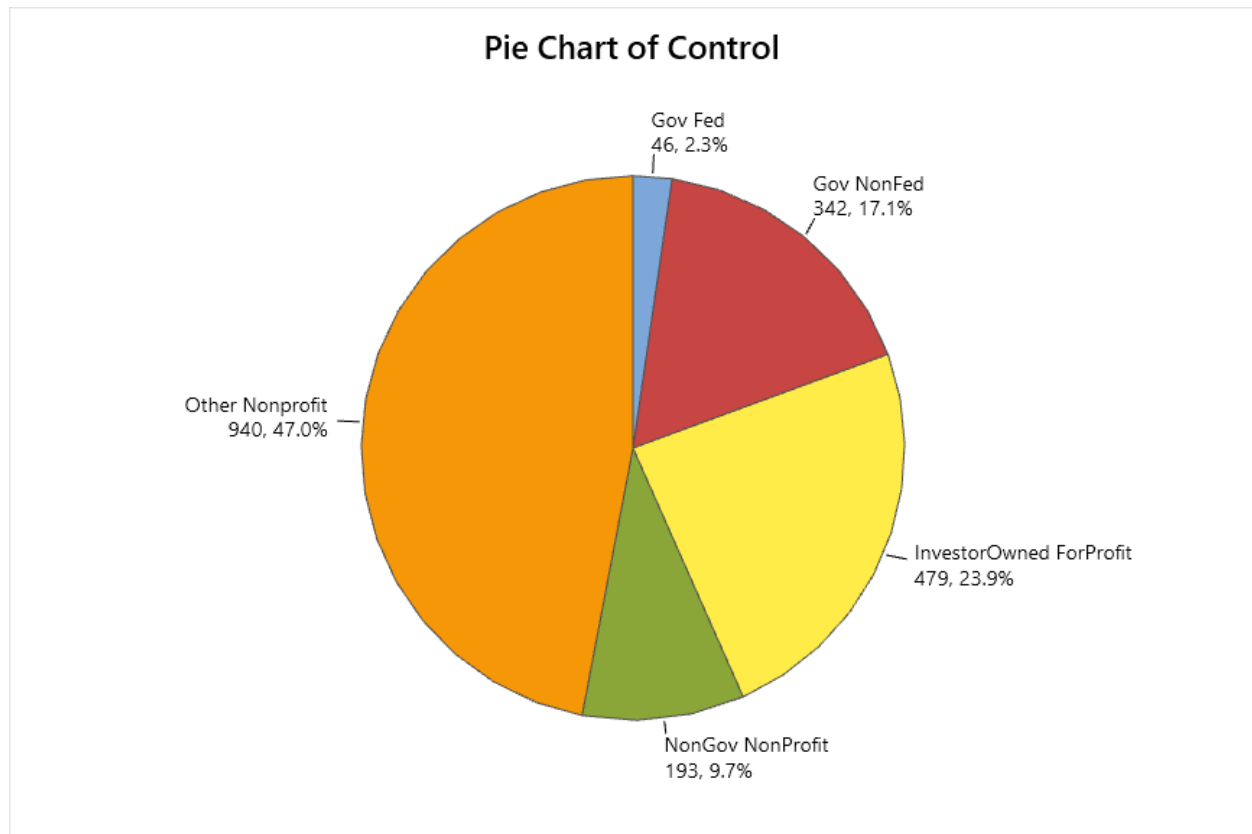
Shown next is a histogram of Number of Beds created in Minitab.



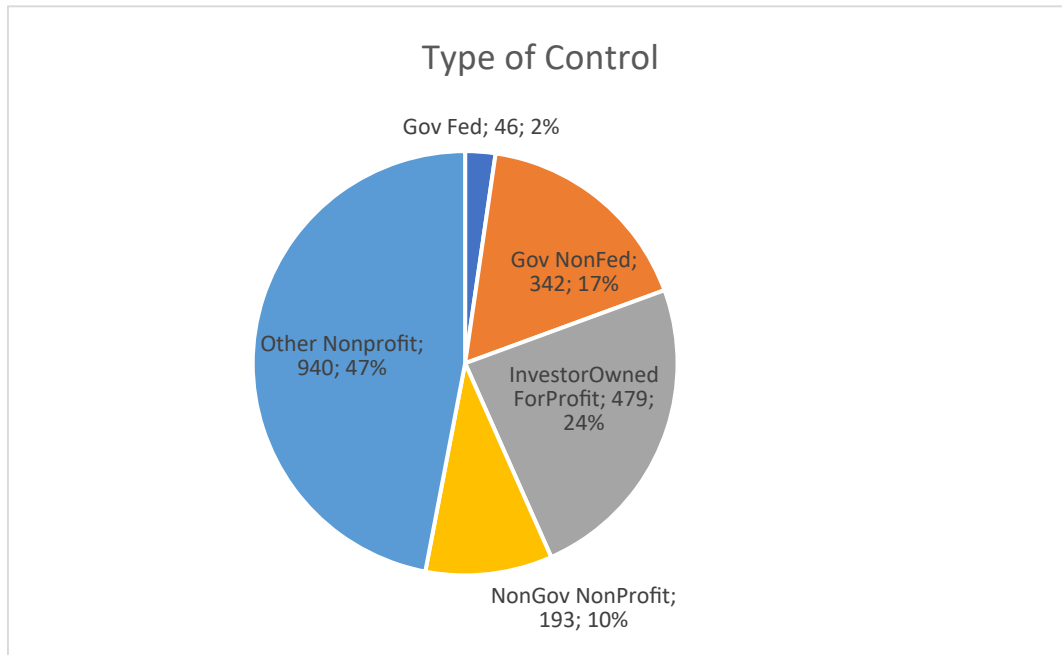
The student should be able to study their histogram and reach some conclusions about the size of the hospitals in this database. Obviously, we can see that they “pile-up” on the left of the graph and have a long thin “tail” to the right which is indicative of a graph that is skewed to the right (an opportunity to introduce a chapter 3 concept visually). Most of the hospitals in the database have less than 500 beds. In addition, one can see from these different histograms that the bin range (class endpoints) makes a difference in how the data are visualized.

4. Construct a pie chart of the variable, Control.

From Minitab,



From Excel,



The pie charts show that almost half of all the hospitals in this database are “Other Nonprofit” (47%). Next are investor owned for-profit hospitals at 24% followed by nonfederal government hospitals (state, county, city, etc.) at 17%,. Note that there were 46 federal government hospitals (2%). About 90% of these federal government hospitals were veterans’ hospitals (this information is not contained in the database but is known by the author) and there were a couple of hospitals that were dedicated to native Americans.