Solution: Homework 2

1.

* Calclate the average volume used per book for each backpack

Average Volume = Volume / Number of books

| Backpack	Volume	Num_books	Avg_books
PM	1586	47	33.74468
MM	1810	44	41.13636
BL	2200	48	45.83333
DS	1874	50	37.48
CM	1810	49	36.93877
MR	1950	49	39.79592
MP	1670	49	34.08163
PR	1500	40	37.5
JC	1635	73	22.39726
CB	1102	38	29
EW	1700	49	34.69388
EA	1500	50	30

* Calculate the average and standard deviation of the backpack price, volume, and number of books.

$$= \frac{\Sigma(x_i)}{n} \qquad\qquad \sigma = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n}}$$

Variable	Mean	Std. Dev.
Price	45.91667	16.43974
Volume	1694.75	273.4245
Num_books	48.83333	8.579397

* What does the price distribution across manufactors look like? Draw a price histogram with bin = 5
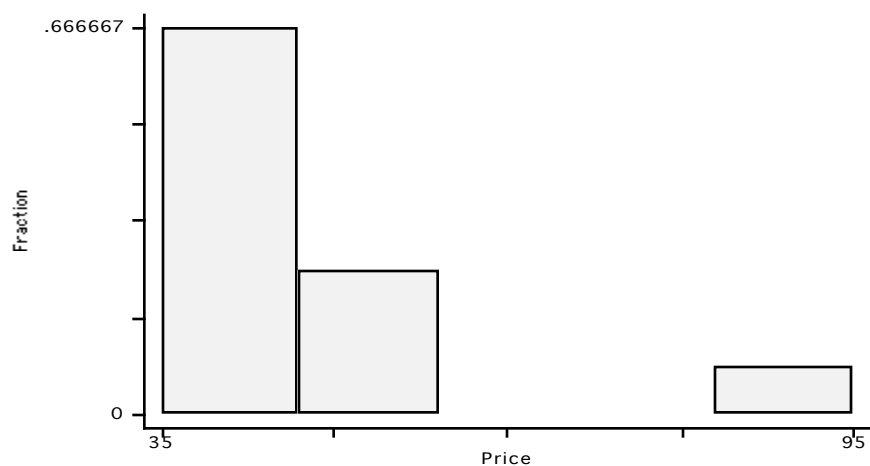
The shape of the price histogram is Unimodal and is right skewed.

* Short Report

After examining the data, most of the backpacks cost about $35 - $40; with its volume above 1500 cubic inches. Brands (or code) of PM, M, BL, DS, CM MR, MP, PR, JC, EW, EW would be good recommendations for customer, since everyone has a decent price and hold at least 44 books. JC would be the most recommended as it is the cheapest in price yet hold the most books. Other factors that we might have missed when making purchasing decision could be the style of the backpack; if the brand name is famous; comfortability when carrying it; the purchase ability of the customer; or the "endurance" of the backpack. These are possible factors that can affect the purchasing decision.

2.

*Outlier - Observation(s) in the data that are much smaller or much larger than the bulk of the data.

*Skewed Distribution - When data are not evenly spread out on either side of the center. (Data tend to concentrate on one side than the other)

*Data Collection
  Survey: Mail-in questionnaire, internet questionnaire
  Designed experiment: simple random sampling, stratified sampling
  Observational studies: lung cancer & smoking; drinking & birth defect