

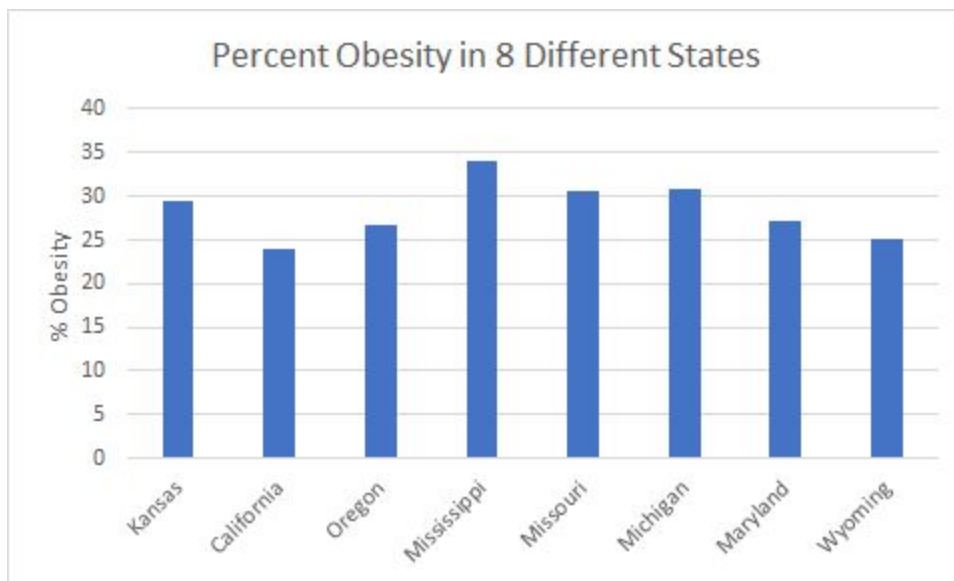
74

a)

Stem	Leaf
5	1
6	
7	6, 7, 8, 9
8	1, 2, 4, 6
9	9

- b) The potential outliers are 51 and 99. They are considered outliers because they are less than  $(1.5)(IQR)$  below the first quartile and more than  $(1.5)(IQR)$  above the third quartile respectively.

75



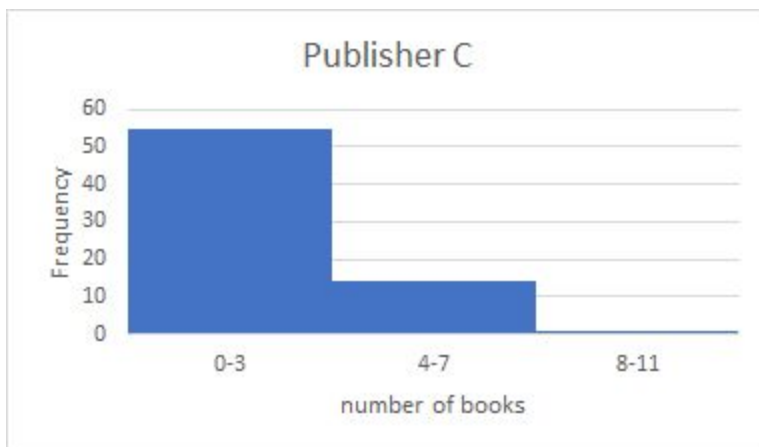
a)

1	# of books	Freq.	Rel. Freq.
2	0	10	0.14706
3	1	12	0.17647
4	2	16	0.23529
5	3	12	0.17647
6	4	8	0.11765
7	5	6	0.08824
8	6	2	0.02941
9	8	2	0.02941
10		68	
11			
12	# of books	Freq.	Rel. Freq.
13	0	18	0.15126
14	1	24	0.20168
15	2	24	0.20168
16	3	22	0.18487
17	4	15	0.12605
18	5	10	0.08403
19	7	5	0.04202
20	9	1	0.0084
21		119	
22			
23	# of books	Freq.	Rel. Freq.
24	0-1	20	0.28571
25	2-3	35	0.5
26	4-5	12	0.17143
27	6-7	2	0.02857
28	8-9	1	0.01429
29		70	

a)



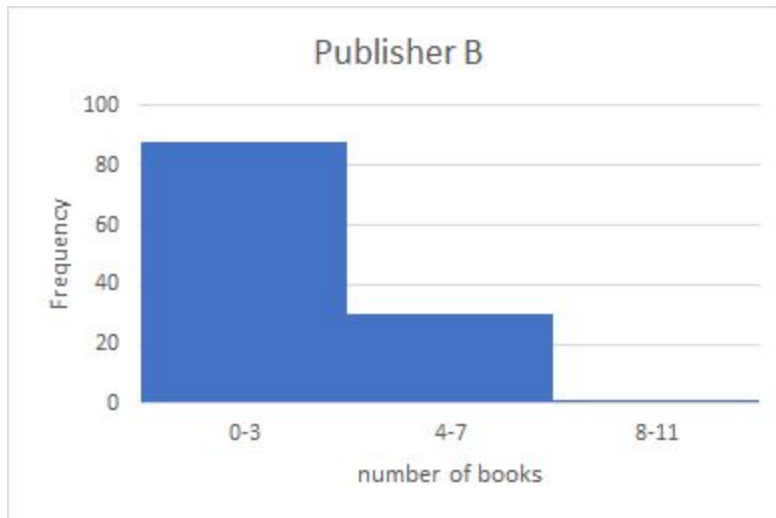
b)



- c) The graphs for A and B are not identical because they collected their data at different places. Their sample sizes are also different, so the scales of their data are also different.
- d) Yes, because there is a trend in A and B in which frequency decreases as the number of books adult purchases increases.

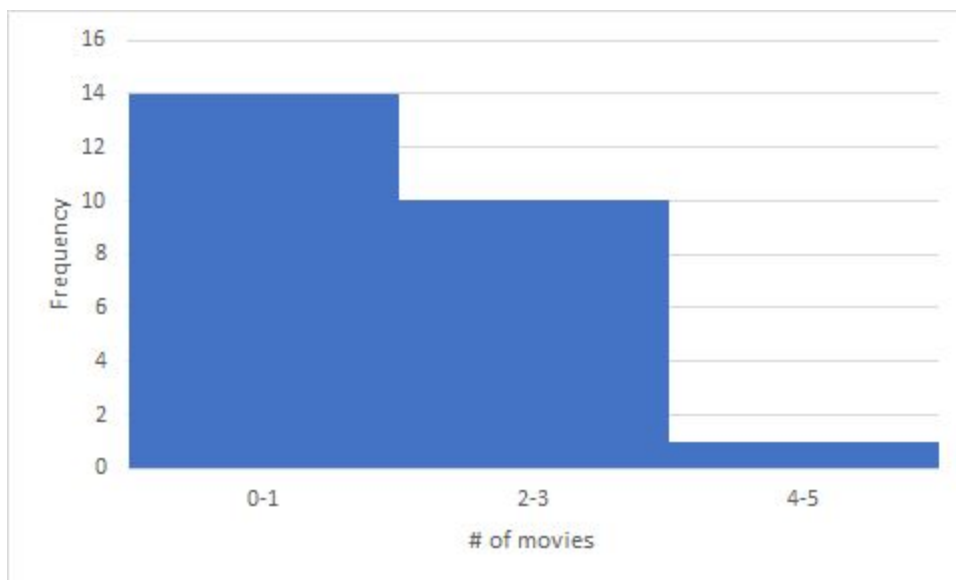


e)



- f)
- g) Publisher C's graph is more similar to Publisher A's since they both have similar frequency for each bin. In contrast, Publisher B has a much larger sample size, causing its frequency for each bin to be significantly greater than those of A and B.

78



a)

# of movie	Frequency	Rel. Freq	Cumul. Rel. Freq.
0	5	0.2	0.2
1	9	0.36	0.56
2	6	0.24	0.8
3	4	0.16	0.96
4	1	0.04	1
	25		

b)

79

c) 41

80

d) Convenience

83

a) 6%

b) 72%

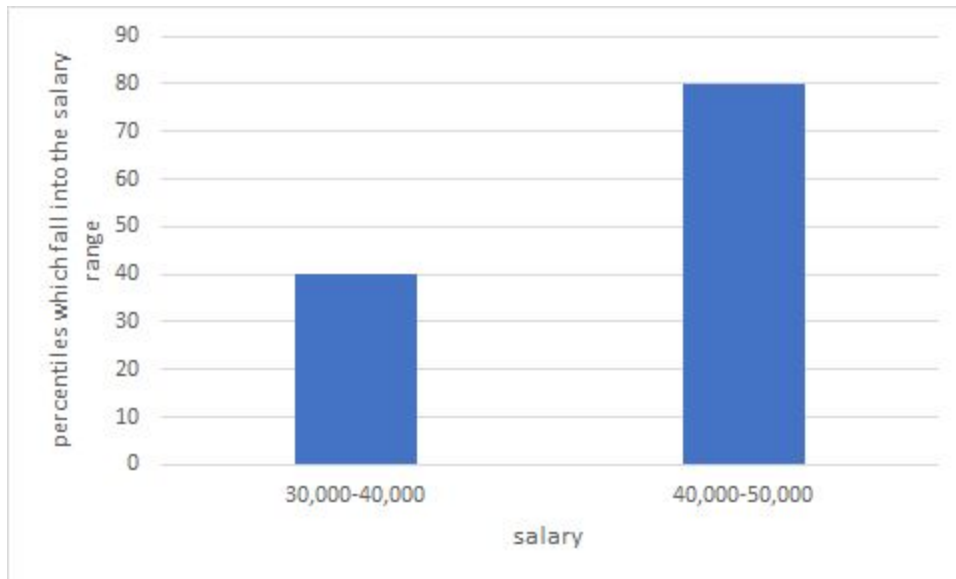


c)

i) The bin values are not required to be equal when there is a larger sample size and data range.

ii) those intervals should be handled because they are part of the data. They also show that there are people who consider those extreme salaries as a possible middle-class income range, which is important.

d) 40th lies in 30,000 to 40,000. 80th lies in 40,000 to 50,000



e)

84

- a) The fourth quarter. It has a spread of 1
- b) The second quarter. It has a spread of 8
- c)  $IQR = Q3 - Q1 = 12 - 2 = 10$
- d) No. Because 10-13 cover 50% of the data, while 5-10 covers approximately 25% of the data
- e) ii) 2-4, because we know the rest covers all the data for their respective quartiles, while 2-4 falls only part of the second quartiles, which is 2-10.

88

- a) i) They might not have the same sample size, even though they have the same minimum and maximum values
- ii) the data sets are shown shown, although they might have different median values, they can still have the same mode.
- iii)  $>4$  and  $<4$  are both the upper and lower 50% of the data, so they must have the same data values
- b) Data 2. Because the IQR of Data 1 seems larger and its  $Q3 + (1.5)IQR$  is probably greater than that of Data 2, Data 2 has a higher chance to have "7" as an outlier.

92

21.45% - 26.9%

93

- a) Median rising means there are more older ages than younger ages
- b) There is fewer smaller ages or more older ages
- c) Yes, because more people age while there were fewer newborns to meet the aging population

94) less than 50% of the total FTES. Since 1014 is the median, to have a FTES of 1014 or above means the upper 50% of the FTES. when we are taking 11 years, which is less than half of the sample size, we can assume that there is less than 50% of FTES that fall into the 11 years

95

- a) At or below the third Quartile
- b) At or above the first Quartile

96) 474

97) 50% because the 528.5 is the first quartile and 1447.5 is the third quartile. Their difference is 2 quartiles, meaning a 50%.

98) IQR is the difference between the third and first quartiles, and it represents the middle 50% of the data

99)  $(1014 - 1000)/474 = 0.03$