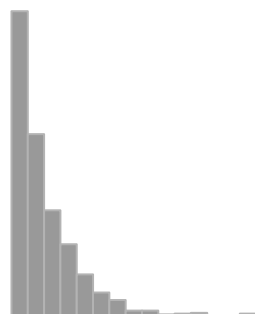
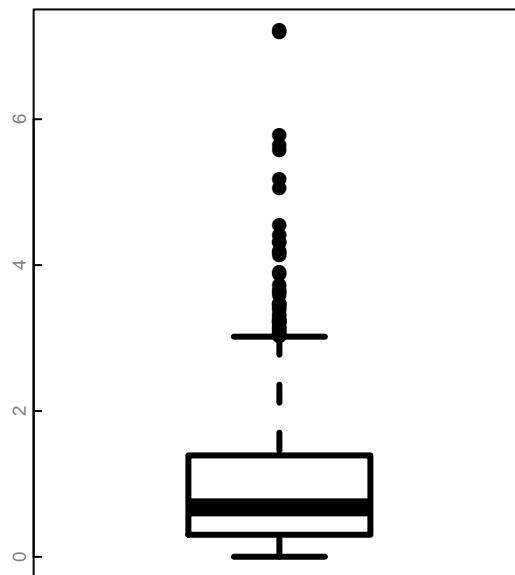


BIOS 600 · Practice for Quiz 2: Summary Methods

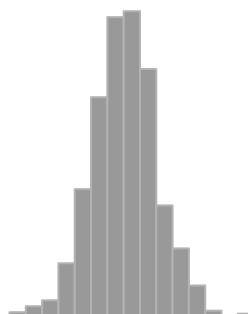
1. Suppose we measure the X (a generic characteristic) from 1000 individuals. Create a boxplot of X (without the outliers) with the following summary information:

Min	Q1	Median	Mean	Q3	Max
0.002	0.304	0.679	0.968	1.391	7.217

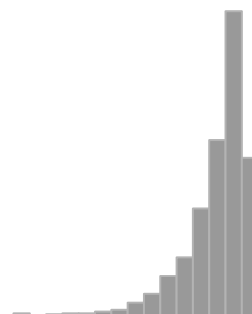
2. Which histogram represents the same data summarized in the boxplot?



(A)



(B)

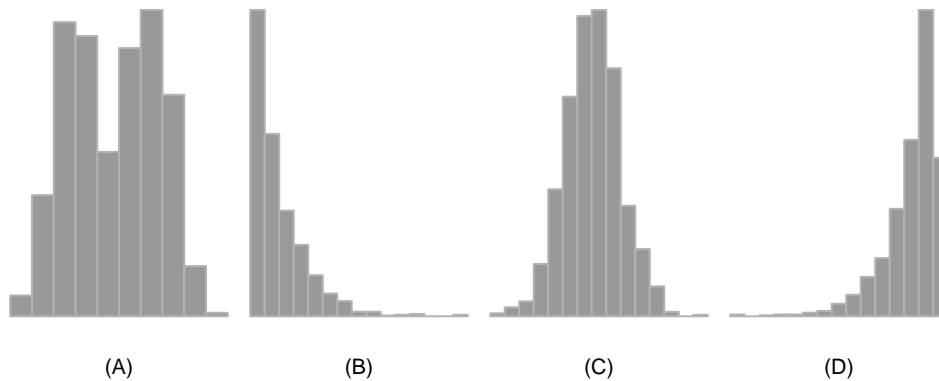


(C)

3. Find the min, median, mean, max, 25th percentile, and sample variance of the following numbers.

-1, 2, 3, 0, 3, 4

4. Which data is skewed? symmetric? bimodal?



5. A die is a six sided cube. The sides are labeled with dots: \square , \square , \square , \square , \square , \square . (The plural of die is dice.) Suppose we roll two dice and sum the outcomes. (Example: $\square + \square = 7$.) Call the sum S . What is the variable range of S ?

6. (Continued from 5.) The following frequency table comes from 1000 rolls. What must the proportion of 2s be? What proportion of the outcomes are odd? What proportion of the outcomes are greater than 10?

Sum	%
2	???
3	0.056
4	0.083
5	0.111
6	0.139
7	0.167
8	0.139
9	0.111
10	0.083
11	0.055
12	0.027

7. Calculate the quartiles of the following data.

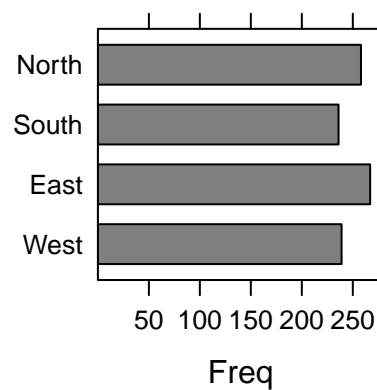
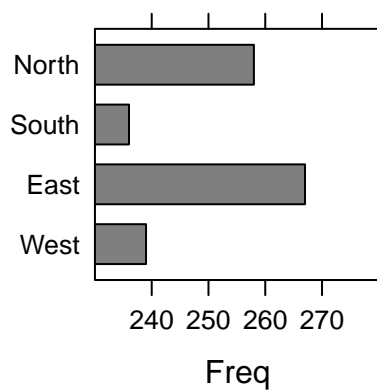
i	X_i	$X_{(i)}$
1	0.62	0.03
2	0.36	0.05
3	3.29	0.05
4	0.12	0.05
5	2.34	0.12
6	0.29	0.19
7	1.17	0.20
8	0.03	0.26
9	0.20	0.29
10	0.80	0.32
11	0.05	0.36
12	0.05	0.46
13	1.12	0.48
14	0.19	0.53
15	0.53	0.62
16	0.26	0.68
17	0.68	0.80
18	0.95	0.86
19	0.48	0.95
20	0.86	1.12
21	1.76	1.17
22	0.05	1.27
23	0.46	1.76
24	1.27	2.34
25	0.32	3.29

8. Table 1 summarizes the Digoxin clinical trial data from page 22 of our text book. This table is missing 2 important pieces of information. What is it missing? (Hint: The table is not missing another summary statistic like the median or standard deviation.)

Table 1: Variable Summary of Digoxin Clinical Trial Data

Variables	%	Min	Mean	Max
sysbp		100.00	131.40	170.00
creat		0.90	1.35	2.68
bmi		15.20	26.75	43.27
TRT				
Placebo	55			
Digoxin	45			
Race				
White	88			
Not White	12			
Sex				
Male	75			
Female	25			
Total (N=40)				

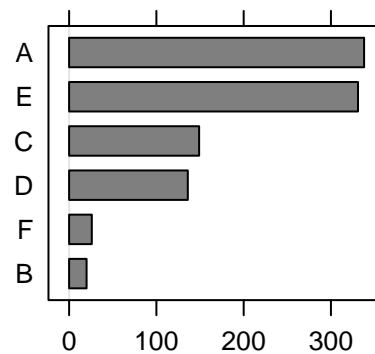
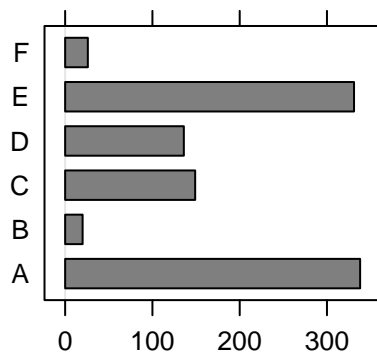
9. Consider the two barcharts. They are based on the same data, but the scales are slightly different. Which accurately represents the data? Which one is “lying with the scale”? How?



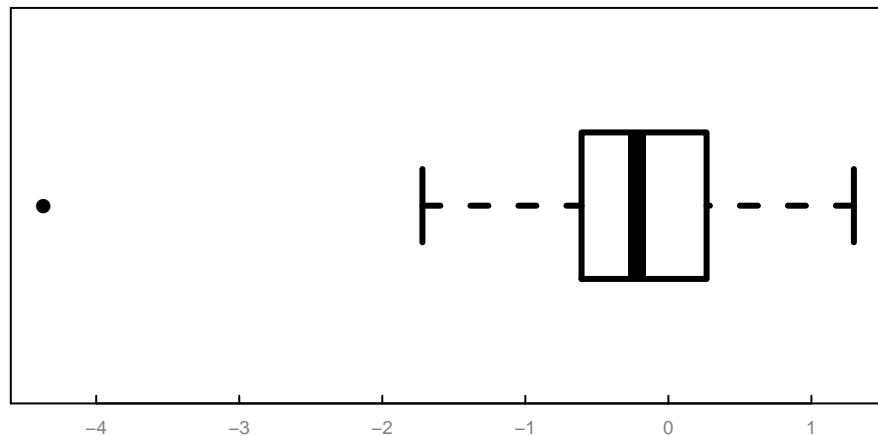
10. Suppose you collect the following data from the WHO. Which type of plot is most appropriate for this type of data?

Country	Year	X
Afghanistan	2003	3
Afghanistan	2008	6
Albania	2009	8
Albania	2008	5
Algeria	2008	1
Algeria	2006	3
Angola	2009	10
Angola	2008	3
Argentina	2008	2
Argentina	2007	2

11. When we present nominal data in a barchart, there is no natural ordering of the categories. The two plots below are barcharts of the same data. In the first, categories are ordered alphabetically. In the second, categories are ordered by their relative frequency. What are the advantages of the frequency ordering? (*I know this is not covered in the notes or text, but think about it.*)



12. Which measure of central tendency is best suited for the data summarized by this boxplot? Why?



13. It is standard convention to draw a histogram so that the total area in the histogram is 1. Use this information to answer the following question: The following figure is a histogram with the second bin missing. Calculate what the height of the second bin must be and add it to the figure. Also, write the height above the bin.

