Name:

Collaborators:

Instructions:

You must submit your worksheet individually by end-of-class or end-of-day. Your name must exist in your worksheet and the names of your collaborators.

Worksheets are marked mostly on completion, and partially on correctness. It will be marked either pass or fail, there will no detailed feedback on worksheets, and no opportunities for revisions and make-up.

Compute Summary Statistics and Visualize Numerical Data

1. Using R to Compute the Summary Statistics

Consider the data set given below. This data set contains one discrete numerical variable named A.

Encode the given data into R, then use the R functions to compute the mean, standard deviation, quartiles, and interquartile range (IQR). Write them –up to three decimal places– on the provided summary statistic table below. Then, answer the provided questions.

Data Table:

A
6
6
9
7
8
8
7
9
8
7
8
7
6
9
8
6
5
9
8
11

Summary Statistics:

Mean	
SD	
Q_1	
Q_2 (Median)	
Q_3	
IQR	

Questions:

a. What R function did you use to compute the mean and standard deviation?

b. What R function did you use to compute the quartiles?

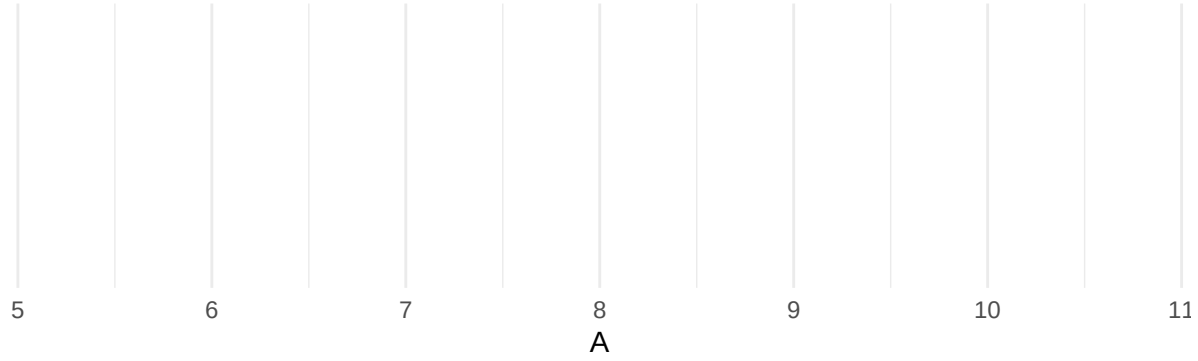
c. Which R function should you use to compute the 30th percentile, and what values should you input for its arguments?

d. What is the 30th percentile of the given data set?

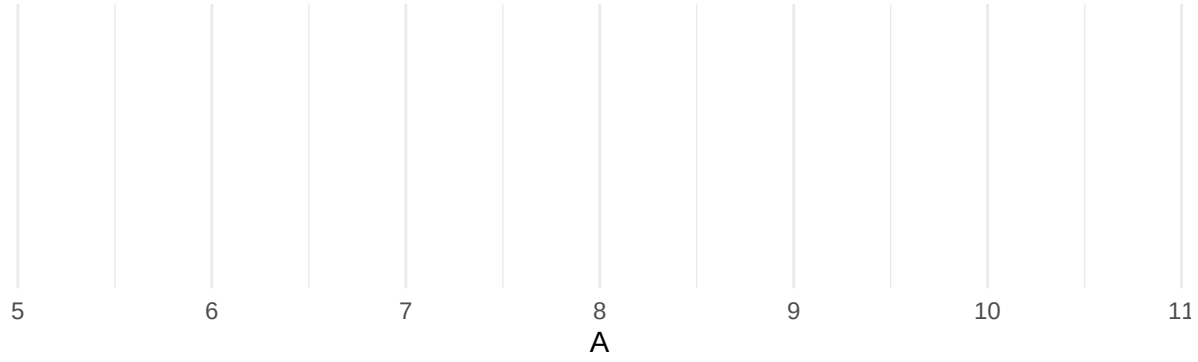
2. Visualizing Numerical Data

Consider the given data set from Problem 1 and the resulting summary statistics.

a. Draw a stacked dot plot on the provided empty axis below.



b. Draw a box plot on the provided empty axis below. Label the main parts of the box plot.



c. Describe your observations on the shape of the distribution in terms of skewness and modality. Compare the dot plot and the box plot in terms of what each reveals and what limitations they have. Explain your reasoning.

References

1. Diez DM, Barr CD, Çetinkaya-Rundel M (2012) OpenIntro statistics, OpenIntro.