HW 03

Stat140-04

Due date Nov 11 6PM EST on Gradescope

Notes on grading

In general, you will receive full points on the question if (1) there are no errors in your solution AND (2) the solution is written in highly articulate Statistical and English language. Point will be taken off if there are errors, your writing of the solution is incomplete, or there are issues with writing organization.

Problem 1

Estimating Summary Statistics For the dataset

$$45, 46, 48, 49, 49, 50, 50, 52, 52, 54, 57, 57, 58, 58, 60, 61$$

(a) Without doing any calculations, estimate which of the following numbers is closest to the mean:

$$60, 53, 47, 58$$

Explain why you chose your estimate.

(b) Without doing any calculations, estimate which of the following numbers is closest to the standard deviation:
$$52, 5, 1, 10, 55$$

Explain why you chose your estimate.

(c) Use R to find the mean and the standard deviation for this dataset. Write your R code here.
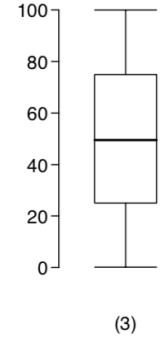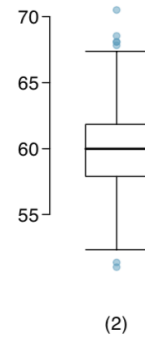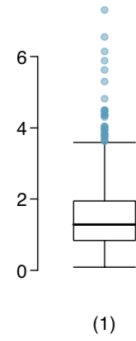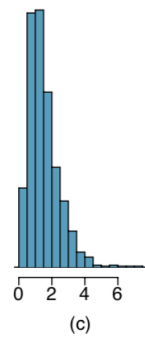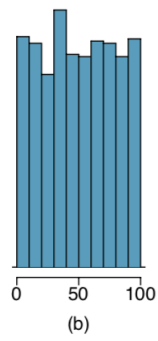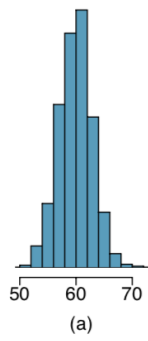
Problem 2

Will young blood help old brains? In a study, old mice were randomly assigned to receive transfusions of blood from either young mice or old mice. Researchers then measured, among other things, the number of minutes each mouse was able to run on a treadmill. The results are given in the side-by-side boxplots in the following figure.
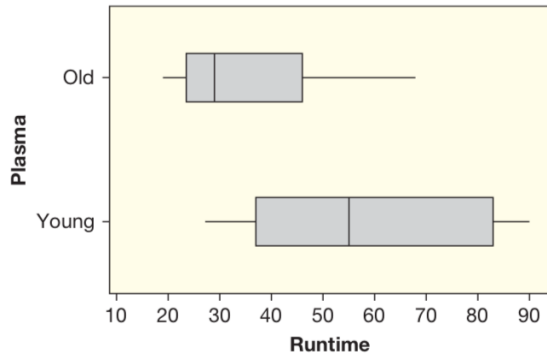
(a) Estimate the median runtime for the mice receiving old blood.

(b) Do we expect the mean runtime for the mice receiving old blood to be larger than, smaller than, or about the same as the median for these mice? Explain your answer.

(c) Which group of mice, those receiving old blood or those receiving young blood, appear to be able to run for more minutes?

(d) Does there appear to be an association between runtime and whether or not the mouse received young blood or old blood?

(e) Are there any outliers in either group? If so, which one(s)?

Problem 3

Describe the distribution in the histograms below and match them to the box plots.

Problem 4

Draw any scatterplot satisfying the following conditions

(a) $n = 10$ and $r = 1$

(b) $n = 8$ and $r = -1$

(c) $n = 5$ and $r = 0$

Problem 5

Read the first chapter in the book "Data Visualization: A Practical Introduction Book" by Kieran Healy
https://socviz.co/lookatdata.html

Find a bad statistics figure on the internet and include it in your homework. Clearly explain what makes the figure bad, and how to improve it.