Assignment 2

1. You decide to test the impact of caffeine intake on dream vividness. You measure the average daily amount caffeine per day (in ounces) of a group. You also obtain (average) dream vividness ratings. The data looks like –

Participant	Caffeine Intake	Dream Vividness Rating
1	24	3
2	16	7
3	20	7
4	28	7
5	14	4
6	10	4
7	6	5
8	20	7
9	23	4
10	28	6

1. What plot would you use to best express the data? **Scatterplot (3pts)**
2. Create a plot of caffeine intake against dream vividness rating **(3pts)**
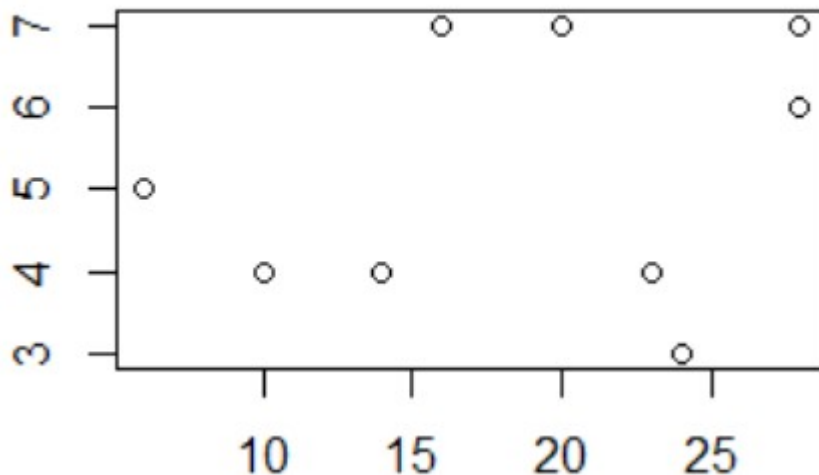
3. Compute a correlation coefficient
   r = 0.24 (1 pt)
4. Interpret the data
   There's a positive association between the two (2pts)

2. You run a personality inventory on 10 people and assess each person's level of "open-mindedness". The ratings range from 1 (not at all open-minded) to 100 (completely open minded). You get the following data –

Participant	Rating
1	94
2	65
3	46
4	79
5	69
6	12
7	67
8	22
9	47
10	79

1. Compute the mean, median and mode for the data **(1pt each)**
   Mean = 58, median = 66, mode = 79
2. Compute the standard deviation (population and sample) **(2pts each)**
   Sample sd = 26.05, the pop. sd is more complicated, give 2 pts.
3. Which standard deviation is the more appropriate one to use? **(1pt)** Why? **(3pts)**
   Sample sd is more appropriate
   The data is sample data. The statistics are based on a sample
4. What r command would you use to compute the variance of the data? **(1pt)**
   var()

3. You are asked to analyze data from a study that looked at the association between an organizational self-rating ("how organized would you rate yourself?") and social network (how many "friends" would you say you currently have at this moment?") for 15 people. The data is –

Participant	Organizational Self Rating	Number of Friends
1	9	14
2	81	12
3	57	6
4	59	9
5	51	5
6	48	19
7	32	16
8	82	7
9	44	17
10	19	20
11	10	15
12	54	13
13	20	11
14	18	3

15	15	4

1. Run a correlation on the data. What's Pearson's r? **- 0.15 (1pt)**
2. Interpret the data **(3pts) There is a negative association between the two vari-ables. As organization goes up, the size of the social network goes down**

3. Relist the correlation by rank order below **(3pts)**

Participant	Organizational Self Rating	Organizational Self Rating, Rank	Number of Friends	Number of Friends, Rank
1	9	1	14	10
2	81	14	12	8
3	57	12	6	4
4	59	13	9	6
5	51	10	5	3
6	48	9	19	14
7	32	7	16	12
8	82	15	7	5
9	44	8	17	13
10	19	5	20	15
11	10	2	15	11
12	54	11	13	9
13	20	6	11	7
14	18	4	3	1
15	15	3	4	2

4. Looking at the data, what do you anticipate about a Spearman's correlation analysis? **(3pts) This is based on "eyeballing" – anything reasonable can be accepted**
5. Run a Spearman's correlation. What is the output? **-0.16 (1pt)**
6. How does it compare to Pearson's r**? It's negative(-0.15, and almost the same (2pts)**
7. Create a scatter plot of the original data **(3pts)**
8. Using lm, find the linear model for the best fit line. **(1pt)** What is the value of the slope? **(-.03, 1pt)** What is the value of the intercept? **(12.79, 1pt)**
9. Using ABline, place trendline on your scatterplot **(2pts)**

Advanced work (...if you're interested)
Create a normal distribution

Create a data set in any range using the "seq" command. For a range of numbers, this creates a sequence from the first, to the second number, by step value you select. Try -
> x = seq(-5, 5, by = .1)

Next find the probability values, based on the normal distribution, that is associated with each value in x. (The values in 'x' will appear on the x-axis). To find what probability value, based on the normal distribution, is associated with each value in x use the "dnorm" command. The "dnorm" command returns a probability (again from the normal distribution) associated with the values of the x-axis, and based on where the mean is. (Since our range is -5 to 5, an intuitive mean could be 0.)

> y = dnorm(x, mean = 0, sd = 1.5)

- Now, plot x and y. The plot shows the probability (along the y-axis. The higher the height, the greater the likelihood) of each value (on the x-axis) of the original data set.
> plot(x,y)

- Using ABline, place a line at mean (x = 0) of the figure. The line should be in any color you choose.

- Using AB line, place a line that runs parallel to the x-axis and crosses the y-axis at 0.2 This line should be in a different color than the previous line and have a different thickness

- Replot the figure, but change the symbol used to an "X", and change the color

- Replot the figure one more time and change the symbols to a single line

Extra, extra R stuff
    Add a title to the figure
    Change the x-axis title to "Observations"
    Change the y-axis to "Probability"
    Change the range of the y-axis from 0 to 1 (rather than 0.0 to 0.4)