None
HW2 Solutions

Note:

1. Only one possible right answer is shown. All possible right answers will be given full credit.
2. Only the final solution is shown, and the details of actual code is not shown.
3. You may come to the office hours or the help sessions to discuss the HW solutions.
4. If you find any typos or issues, kindly contact your section instructor, or send a text @ **smujahid** on MS teams.

Table of Contents

Problem #A

```
---------------------------------------------Problem #A---------------------------------------------
```

`A-1:`

The variables in Table-1:

Variables
S. No
Game
Platform
Price
Rating
Multiplayer
Sold

Table-2 provides metadata.

`A-2:`

Variable	Type
S. No	Nominal
Game	Categorical
Platform	Categorical
Price	Numerical
Rating	Ordinal
Multiplayer	Nominal
Sold	Numerical

`A-3:`

Variable	Type
S. No	Auxilliary
Game	Independent
Platform	Independent
Price	Independent
Rating	Independent
Multiplayer	Independent
Sold	Dependent

`A-4:`

RATINGS

E 5

T 2

M 3

A-5:

PLATFORM

PC 20%

PS4 50%

Xbox 30%
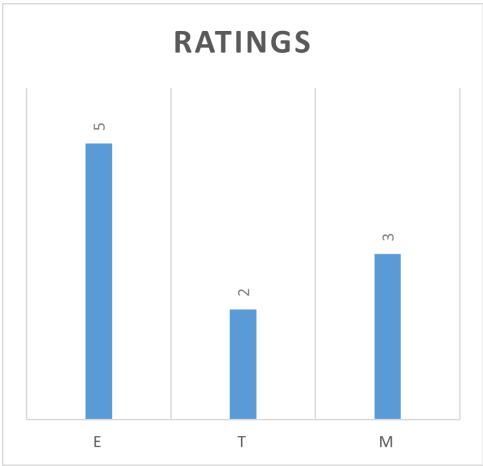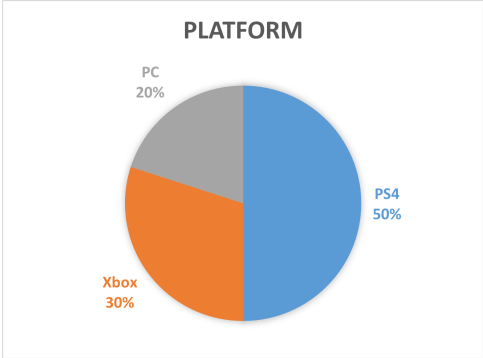
A-6:
Not symmetrical

A-7, A-8, A-9:

Here we are presenting sample variance and sample standard deviation.
The data has multiple modes, 120, 150 & 189.
Note: *The lower and upper quartile may differ based on interpolation style
 (interpolation method to use when the desired value lies between two data points).
 In the following table midpoint interpolation method is used.

statistic	value
mean	255.1
median	258
mode	289
variance	3394.1
standard deviation	58.2589
upper quartile	289*
lower quartile	210*

Here are some possible solutions using different interpolation styles.
All valid approaches will get full credit.

	midpoint	nearest	higher	lower	linear
Upper_quartile	289.000000	289	289	289	289.000000
Median	258.000000	250	266	250	258.000000
Lower_quartile	210.000000	200	220	200	205.000000

A-10:

Price
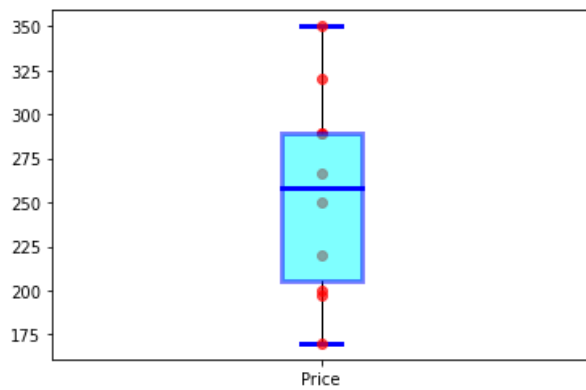
~~~~~~

Problem #B

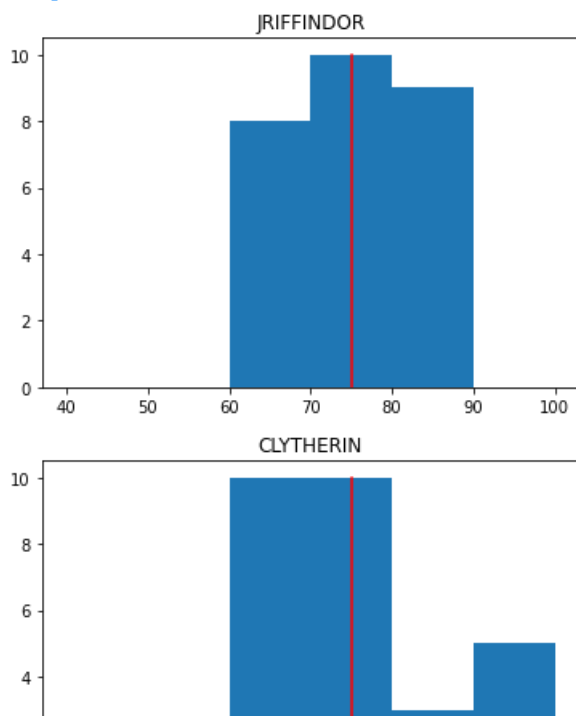---Problem #B---
B-1:

The following plots shows the histograms for each section on individual figure.
The red line depicts the mean of the data.
The bins selected for the histograms were as follows:[40,50,60,70,80,90,100].
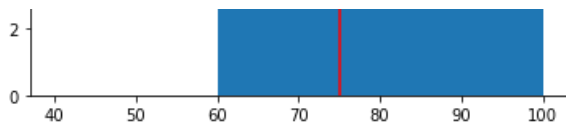Difference in bin size or number of bins will result in different looking histograms.

For each histogram (or any other matplotlib plot) on its own figure,
you need to use plt.figure() and plt.show() methods for each plot.
For example, you can use:

```
plt.figure()
plt.hist(Series1)
plt.title('Figure1')
plt.show()
plt.figure()
plt.hist(Series2)
plt.title('Figure2')
plt.show()
```
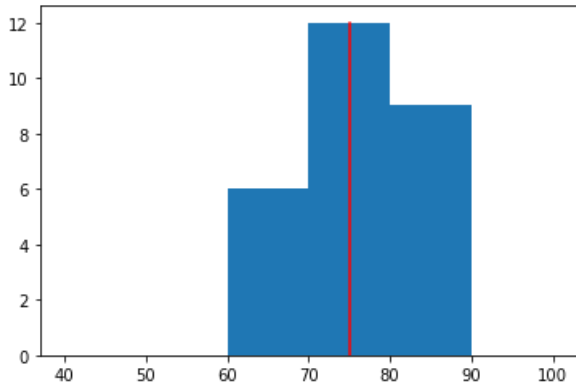
Or loop for each series as:
```
for series in listOfSeries:
    plt.figure()
    plt.hist(series)
    plt.show()
```
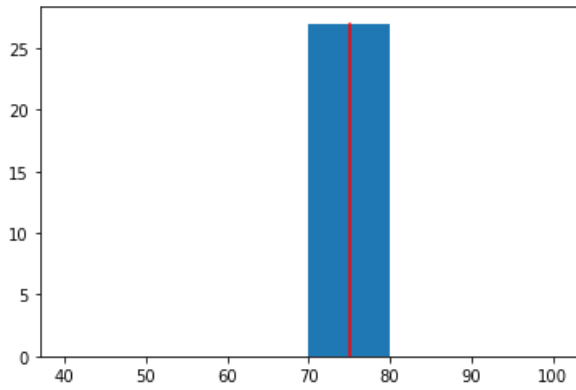

JRIFFINDOR

CLYTHERIN

HUFFLEBUFF

RAVENKLAW

B-2:
For Jriffindor the mean is 75.0, median is 75.0 and mode is 62.
For Clytherin the mean is 75.0, median is 75.0 and mode is 65.
For Hufflebuff the mean is 75.0, median is 75.0 and mode is 61.
For Ravenklaw the mean is 75.0, median is 75.0 and mode is 75.

B-3:

The population mean and variance is presented in the following output.

For Jriffindor the variance is 60.67 and standard deviation is 7.79.
For Clytherin the variance is 112.5 and standard deviation is 10.61.
For Hufflebuff the variance is 68.0 and standard deviation is 8.25.
For Ravenklaw the variance is 0.0 and standard deviation is 0.0.

B-4:
For Jriffindor the upper quartile is 81.5 and lower quartile is 68.5.
For Clytherin the upper quartile is 80.0 and lower quartile is 65.0.
For Hufflebuff the upper quartile is 80.0 and lower quartile is 70.0.
For Ravenklaw the upper quartile is 75.0 and lower quartile is 75.0.

B-5:

To draw boxplot (or other matplotlib plots) in one figure (overlapping or side by side), you can place the code for the plots between one plt.figure() and plt.show() methods. For side by side plots in one figure, you can insert sequence of arrays/lists. For example, here we used:
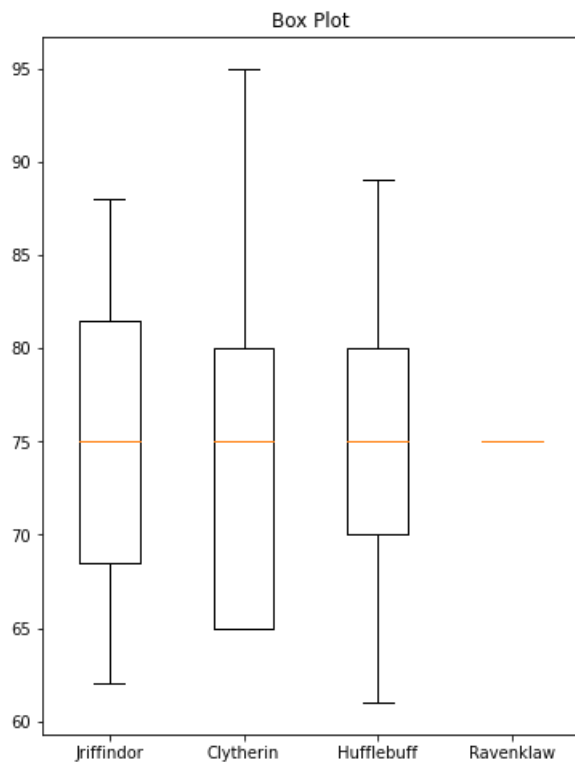
```
plt.figure()
plt.boxplot([Series1, Series2, Series3],labels=['Name1', 'Name2','Name3'])
plt.title('Side by Side')
plt.show()
```

For overlapping plots in one figure, you can insert sequence of plots. For example, you can use:

```
plt.figure()
plt.boxplot(Series1)
plt.boxplot(Series2)
plt.boxplot(Series3)
plt.title('Overlapping')
plt.show()
```

For each plot on its own figure, you need to use plt.figure() and plt.show() methods for each plot. For example, you can use:

```
plt.figure()
plt.boxplot(Series1)
plt.title('Figure1')
plt.show()
plt.figure()
plt.boxplot(Series2)
plt.title('Figure2')
plt.show()
```

B-6:

To statistically check if sample data follows normal distributon, one of the test is: scipy.stats.shapiro() method.
The null hypothesis is: "The data was drawn from a normal distribution."
For Jriffindor the hypothesis testing result is: We fail to reject null hypothesis.
For Clytherin the hypothesis testing result is: We reject null hypothesis.
For Hufflebuff the hypothesis testing result is: We fail to reject null hypothesis.
For Ravenklaw the hypothesis testing result is: We fail to reject null hypothesis.

To hypothesis test results for Ravenklaw may not be correct, since it has no range.

B-7:

To statistically compare (ttest) two samples means we use: scipy.stats.ttest_ind() method.
The null hypothesis is: "The two samples have identical means."
For Jriffindor and Clytherin the hypothesis testing result is: We fail to reject null hypothesis.
For Jriffindor and Hufflebuff the hypothesis testing result is: We fail to reject null hypothesis.
For Jriffindor and Ravenklaw the hypothesis testing result is: We fail to reject null hypothesis.
For Clytherin and Hufflebuff the hypothesis testing result is: We fail to reject null hypothesis.
For Clytherin and Ravenklaw the hypothesis testing result is: We fail to reject null hypothesis.
For Hufflebuff and Ravenklaw the hypothesis testing result is: We fail to reject null hypothesis.

B-8:

To statistically compare two samples distributions without the assumption of normality, we use: scipy.stats.ma
nnwhitneyu() method.
This method is similar to ttest when the sample is drawn from normal distribution.
The null hypothesis is: "The distribution underlying the two samples is same."
For Jriffindor and Clytherin the hypothesis testing result is: We fail to reject null hypothesis.
For Jriffindor and Hufflebuff the hypothesis testing result is: We fail to reject null hypothesis.
For Jriffindor and Ravenklaw the hypothesis testing result is: We fail to reject null hypothesis.
For Clytherin and Hufflebuff the hypothesis testing result is: We fail to reject null hypothesis.
For Clytherin and Ravenklaw the hypothesis testing result is: We fail to reject null hypothesis.
For Hufflebuff and Ravenklaw the hypothesis testing result is: We fail to reject null hypothesis.

~-~