Assignment 3

Find a list of numbers online. Any data of your choice would do. The list should have at least 30 numbers in it.

- Calculate the arithmetic mean, median and Q1 and Q3 of the numbers
- Take these numbers into the notebook and create a box plot for these. Make sure you label the plot correctly

. For submission, include screenshots of all the code in detail. Make sure you write brief a description of each step in comments. For the calculations done by hand, include the basic formulae used and calculations done as comments at the top of the notebook. (If you know markdown, you can create a markdown cell in the notebook to show your working as well.) Create a single PDF which has the hand-written material (if any) and screenshots (if any) combined. Make sure you name your PDF file as: 1XP-XXXX-YourName.pdf (where 1XP-XXXX is your student ID in that exact format).

Calculation Arithematic Mean :

Formula = Sum of All Values / Total lenght

Arithematic mean = (1+1+1+1+1+2+2+2+2+2+2+3+3+3+3+3+3+4+4+4+4+4+4+4+5+5+5+5+5+5)/30

Arithematic mean = 3.0

Calculation Medain :

$$\text{Med}(X) = \begin{cases} \frac{X[\frac{n}{2}]+X[\frac{n+1}{2}]}{2} & \text{if n is even} \\ X[\frac{n+1}{2}] & \text{if n is odd} \end{cases}$$

Data = [3,3,3,3,4,1,4,1,5,4,3,5,2,2,4,2,1,5,4,5,5,3,2,1,5,4,2,2,4,1]

sorted Data

Data = [1,1,1,1,1,2,2,2,2,2,2,3,3,3,3,3,3,3,4,4,4,4,4,4,4,5,5,5,5,5,5]

median = 3 + 3 /2

medain = 6/2

medain = 3

Computing Quartile Q1 :

Q1 = {0.25(30+1)}th Value

Q1 = 7.75

Because of 0 index

Q1 = 7.75 -1

Q1 =6.75th value

Computing Quartile Q2 :

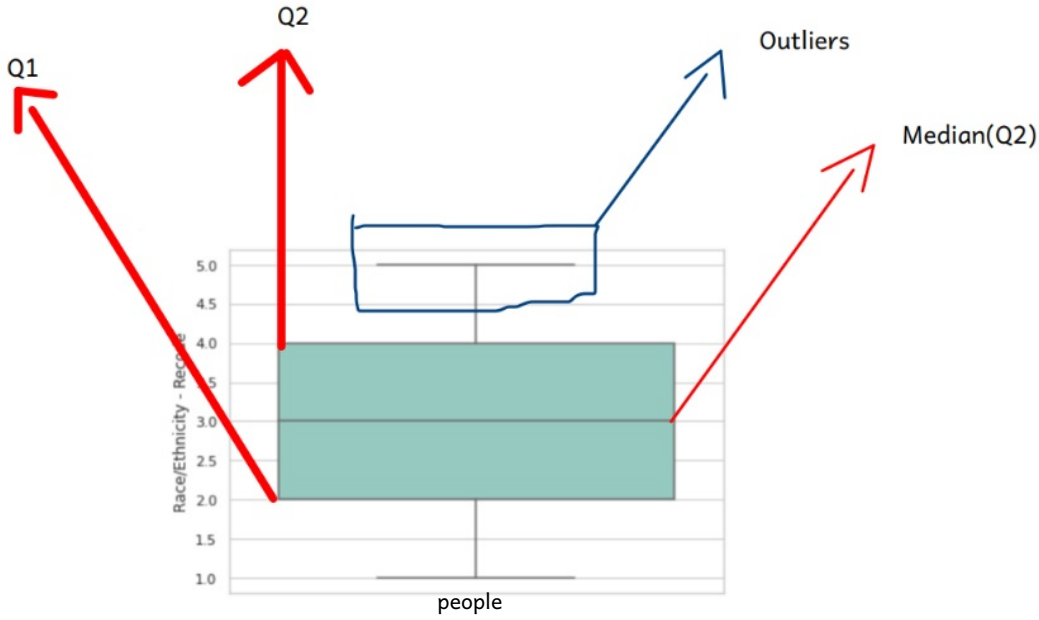
Q2 = {0.75(30+1)}th Value

Q2 = 23.25

Because of 0 index

Q2 = 23.25 -1

Q2 =22.25 th value


Q1
Q2
Outliers
Median(Q2)
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
Race/Ethnicity - Recode
people

```
#importing libararies for ploting thing Data
#pandas is usefull to draw the CSV Data and lots of more help us
import pandas as pd
#matplotlib help will help us in Calculation of Numerical
import matplotlib.pyplot as plt
#numpy help us calculation of Numpy Array
import numpy as np
#for calculating the Quartiles
import statistics as st
#import seaborn for plotting Boxplot
import seaborn as sns
sns.set(color_codes=True)
sns.set_style("white")

# i am getting data from the  nhanes_2015_2016 CSV file and i will on the RIDRETH1 which is
# Race/Ethnicity - Recode
url = "nhanes_2015_2016.csv"
da = pd.read_csv(url)
da
```

	SEQN	ALQ101	ALQ110	ALQ130	SMQ020	RIAGENDR	RIDAGEYR	RIDRETH1	DMDCITZN	DMDEDUC2	...	BPXSY2	BPXDI2	BMXWT	BMXHT	BMXBMI	BMXLEG	BMXARML	BMXARMC	BMXWAIST	HIQ210
0	83732	1.0	NaN	1.0	1	1	62	3	1.0	5.0	...	124.0	64.0	94.8	184.5	27.8	43.3	43.6	35.9	101.1	2.0
1	83733	1.0	NaN	6.0	1	1	53	3	2.0	3.0	...	140.0	88.0	90.4	171.4	30.8	38.0	40.0	33.2	107.9	NaN
2	83734	1.0	NaN	NaN	1	1	78	3	1.0	3.0	...	132.0	44.0	83.4	170.1	28.8	35.6	37.0	31.0	116.5	2.0
3	83735	2.0	1.0	1.0	2	2	56	3	1.0	5.0	...	134.0	68.0	109.8	160.9	42.4	38.5	37.7	38.3	110.1	2.0
4	83736	2.0	1.0	1.0	2	2	42	4	1.0	4.0	...	114.0	54.0	55.2	164.9	20.3	37.4	36.0	27.2	80.4	2.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
5730	93695	2.0	2.0	NaN	1	2	76	3	1.0	3.0	...	112.0	46.0	59.1	165.8	21.5	38.2	37.0	29.5	95.0	2.0
5731	93696	2.0	2.0	NaN	2	1	26	3	1.0	5.0	...	116.0	76.0	112.1	182.2	33.8	43.4	41.8	42.3	110.2	2.0
5732	93697	1.0	NaN	1.0	1	2	80	3	1.0	4.0	...	146.0	58.0	71.7	152.2	31.0	31.3	37.5	28.8	NaN	2.0
5733	93700	NaN	NaN	NaN	1	1	35	3	2.0	1.0	...	106.0	66.0	78.2	173.3	26.0	40.3	37.5	30.6	98.9	2.0
5734	93702	1.0	NaN	2.0	2	2	24	3	1.0	5.0	...	114.0	68.0	58.3	165.0	21.4	38.2	33.5	26.2	72.5	2.0

5735 rows × 28 columns

[49]:
```
Data = da["RIDRETH1"][0:30]  # working frist 30 values
mean = da['RIDRETH1'].mean() #geting mean from the frist 30 values
median = da['RIDRETH1'].median()  #getting medain from the first 30 value

print("Mean  : ",mean)
print("Median : ",median)



print(st.quantiles(Data ,n=4))  #for the Quartiles


plt.figure(figsize=(7,5))  #lenght of Diagram height into width
sns.set_style("whitegrid")


ax = sns.boxplot(y=Data , palette="Set3")
ax.set(xlabel=" People ", ylabel = "Race/Ethnicity - Recode")

#this boxplot represent us Q1 and Q2 and medain most of the people lie between the boxplot
# one 1st part of the people lie on Q1 and 1/3th part of the poeple lie in Q3 and median of the
#people represent as a line
```

```
Mean  :  3.0423714036617264
Median :  3.0
[2.0, 3.0, 4.0]
```

[49]:
```
[Text(0, 0.5, 'Race/Ethnicity - Recode'), Text(0.5, 0, ' People ')]
```