IC 272: DATA SCIENCE - III
LAB ASSIGNMENT – III
Data visualization and statistics from data

Student's Name: Gaurav Kumar **Mobile No: 8529143452**

Roll Number: B20197 **Branch:**EE

1

Table 1 Mean, median, mode, minimum, maximum and standard deviation for all the attributes

S. No.	Attributes	Mean	Median	Mode	Min.	Max.	S.D.
1	pregs	3.845	3.0	1	0	17	3.370
2	plas	120.895	117.0	99	0	199	31.973
3	pres (in mm Hg)	69.105	72.0	70	0	122	19.356
4	skin (in mm)	20.536	23.0	0	0	99	15.952
5	test (in mu U/mL)	79.799	30.5	0	0	846	115.244
6	BMI (in kg/m^2)	31.993	32.0	32.0	0	67.1	7.884
7	pedi	0.472	0.373	0.254	0.078	2.42	0.331
8	Age (in years)	33.241	29.0	22	21	81	11.760

Inferences:

1. Yes, if std. deviation is close to zero, then the mean, median and mode are closer to each other generally. This can be observed from pregs ; (3.84,3,1) , BMI ; (31.99,32,32).

2 a.

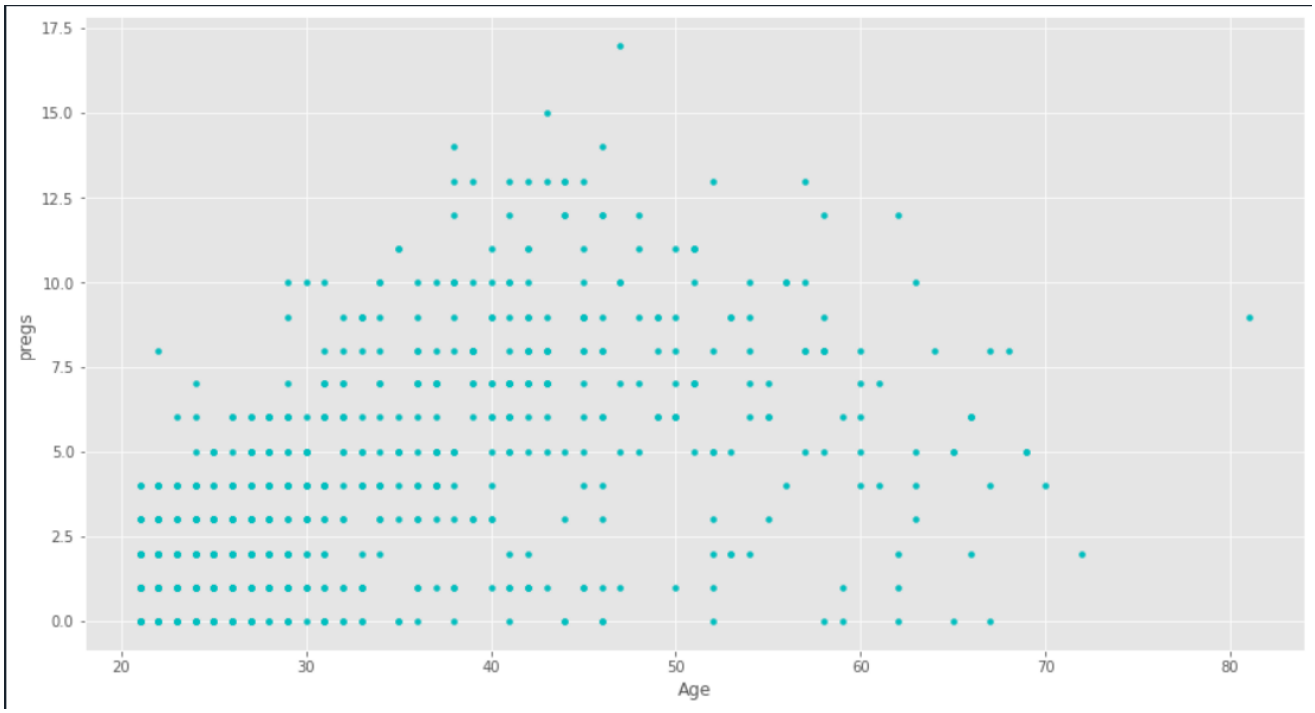

Figure 1 Scatter plot: Age (in years) vs. pregs

Inferences:

1. Age (in years) and Pregs are positively Co-Related.
2. Pregs are mostly present between the age gap of 20-40 years.
3. Pregs present in the age gap of 20-30 years are less in number but they are largely populated in small age differences. Whereas, in higher age gaps the number of pregs are huge but the age difference is less.

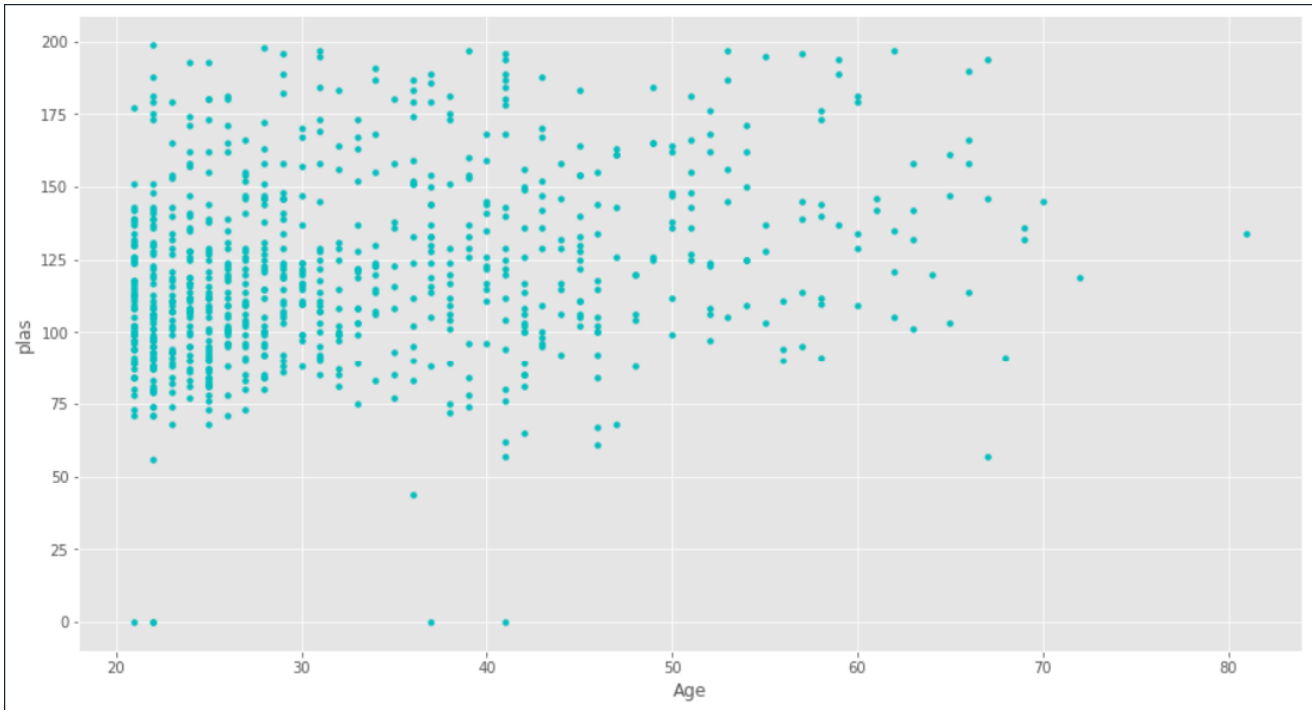

Figure 2 Scatter plot: Age (in years) vs. plas

Inferences:

1. Age (in years) and Plas are positively Co-Related.
2. The population of Plas is densely spread in the age gap of 20-30.

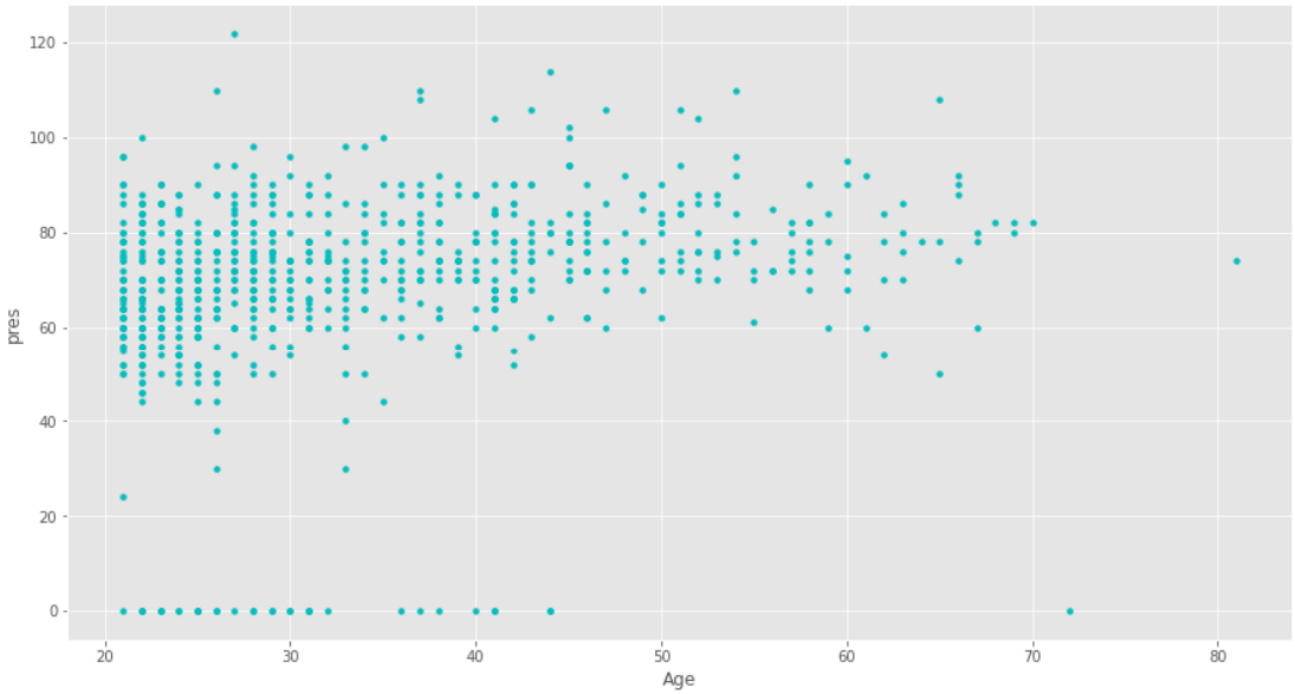

Figure 3 Scatter plot: Age (in years) vs. pres (in mm Hg)

Inferences:

1. Age (in years) and Pres are positively Co-Related.
2. Pres are mostly present in the age gap of 20-40. After this age gap the number scatters.

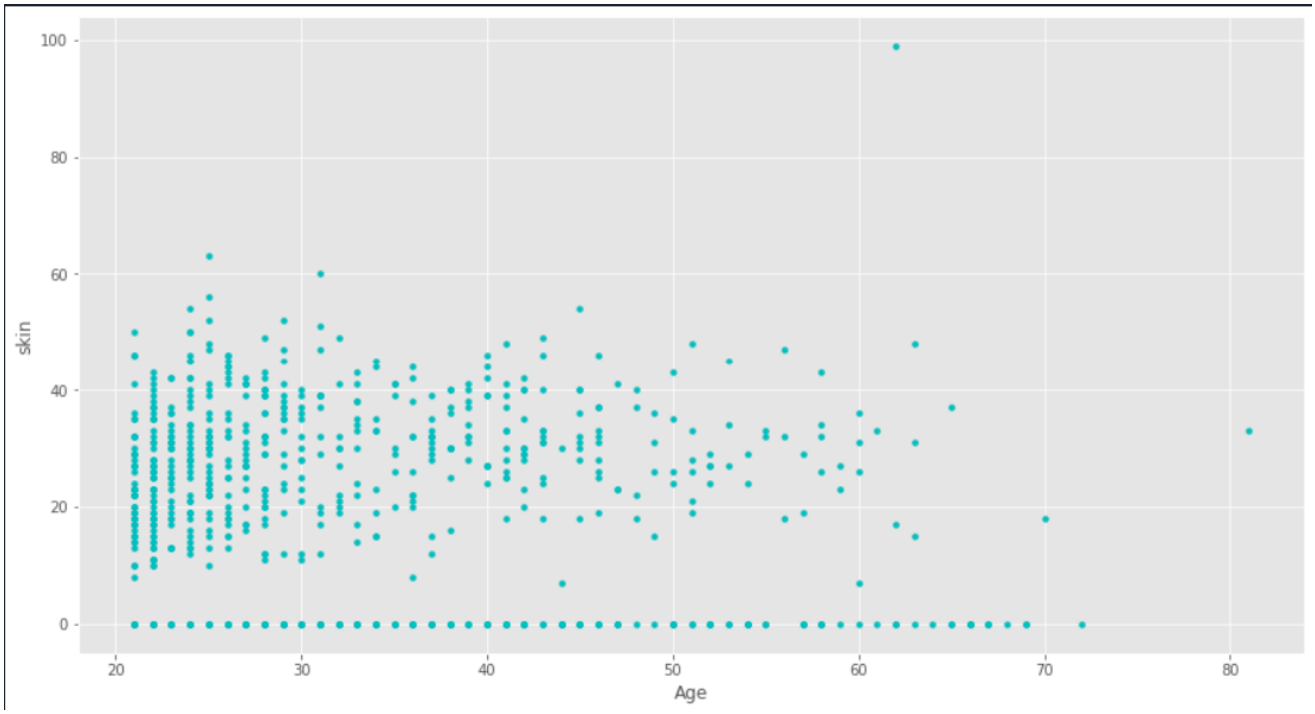

Figure 4 Scatter plot: Age (in years) vs. skin (in mm)

Inferences:

1. Age (in years) and Skin (in mm) are negatively Co-Related.
2. In the age gap of 20-30 years, the skins are hugely and densely present.
3. In further age gaps, this number gradually decreases.

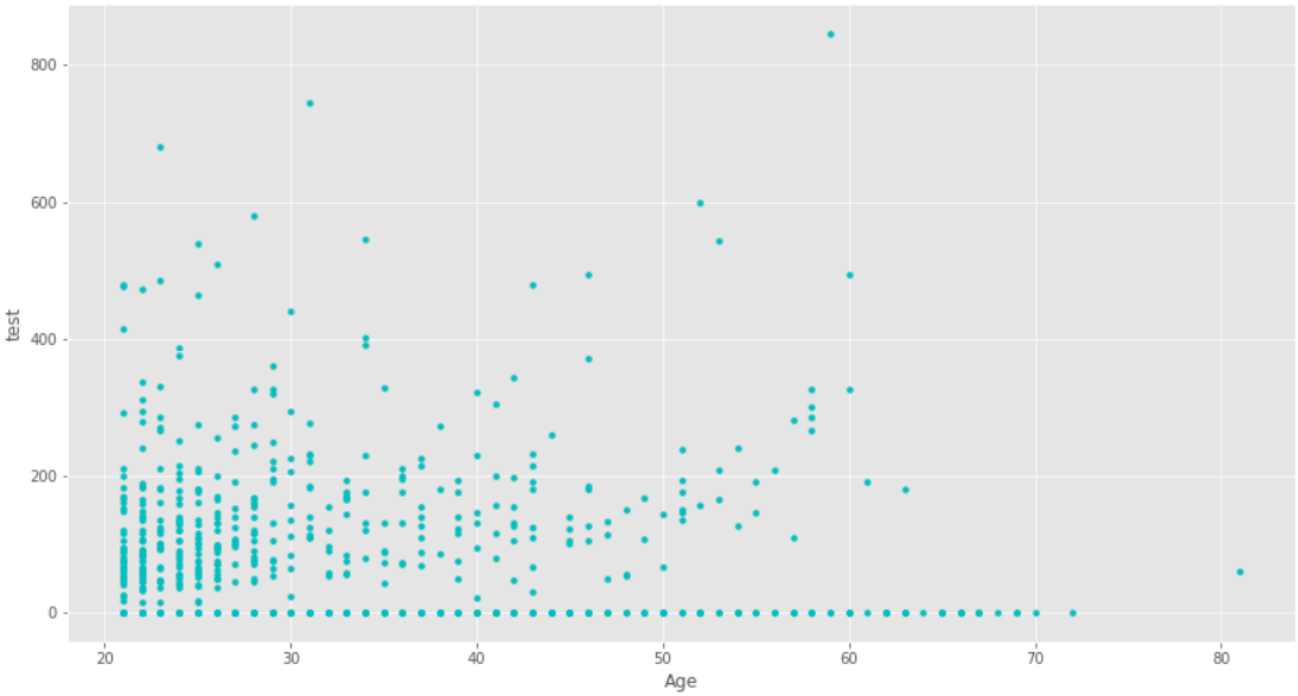

Figure 5 Scatter plot: Age (in years) vs. test (in mm U/mL)

Inferences:

1. Age (in years) and Test are negatively Co-Related.
2. Most tests are present between the age gap of 20-30 years.
3. The density then lessens in the higher age groups.

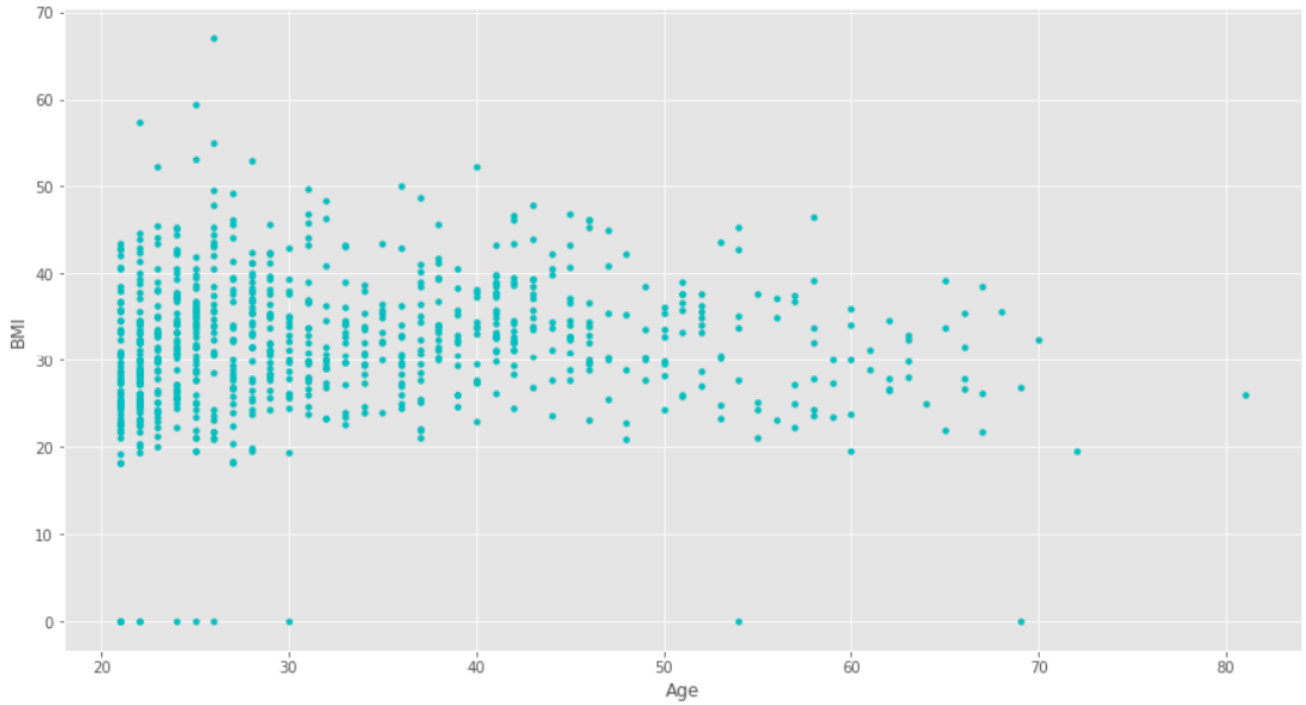

Figure 6 Scatter plot: Age (in years) vs. BMI (in kg/m^2)

Inferences:

1. Age (in years) and BMI are positively Co-Related.
2. **BMI = (Weight in Kg) / (Height in meters)^2**
3. Between the age gap of 20-30 years the BMI ranges from 20-30.
4. The BMI higher than the age of 30 years ranges from 20-35.

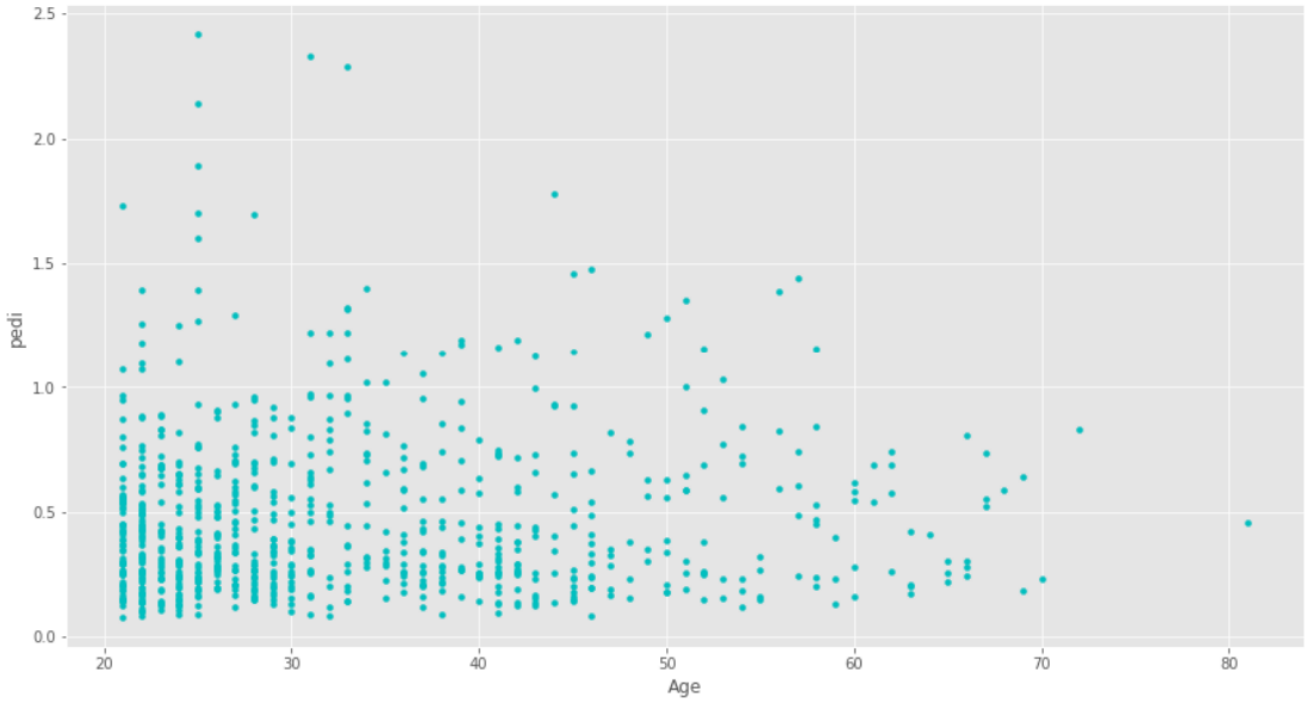

Figure 7 Scatter plot: Age (in years) vs. pedi

Inferences:

1. Age (in years) and Pedi are positively Co-Related.
2. Number of Pedi are densely present between the age gap of 20-30 but the age difference is very less.

b.

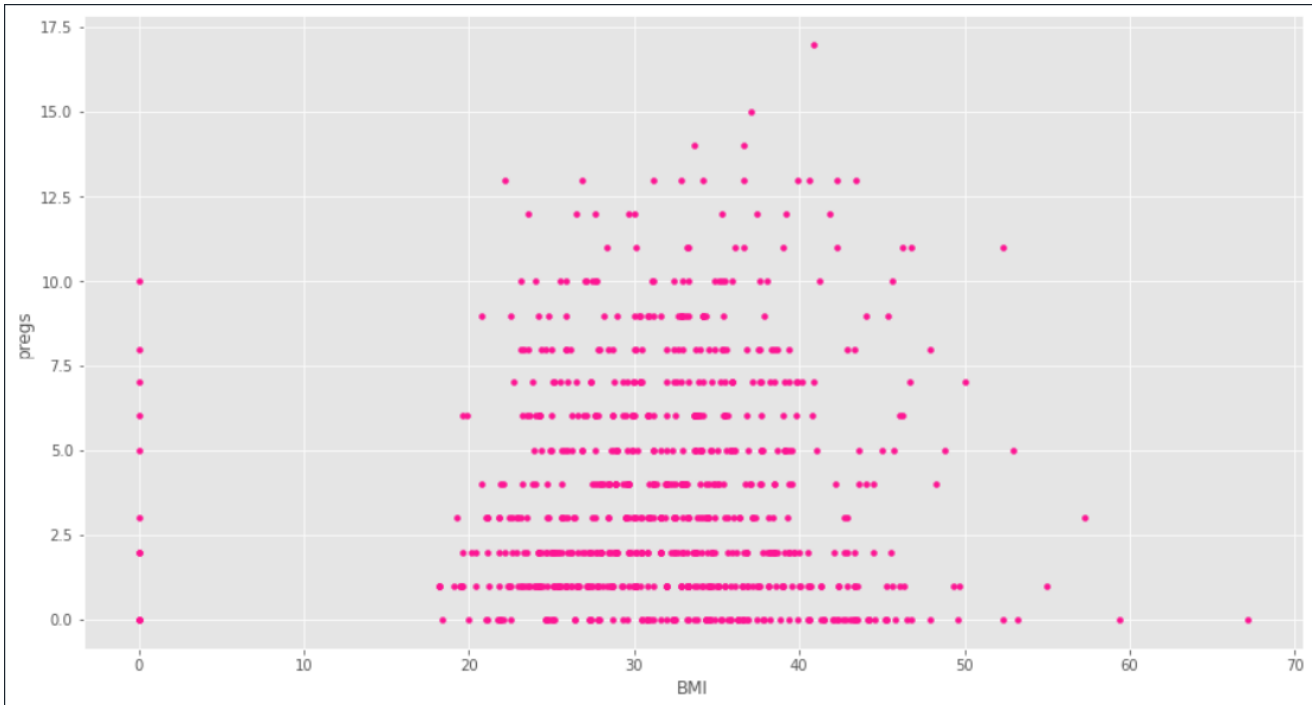

Figure 8 Scatter plot: BMI (in kg/m²) vs. pregs

Inferences:

1. BMI and Pregs are positively Co-Related.
2. In the interval of 20-50 of BMI, the greatest number of Pregs are present.
3. For BMI= 60 and 70 there are no regs present because these observations are due to the noises present in the data.

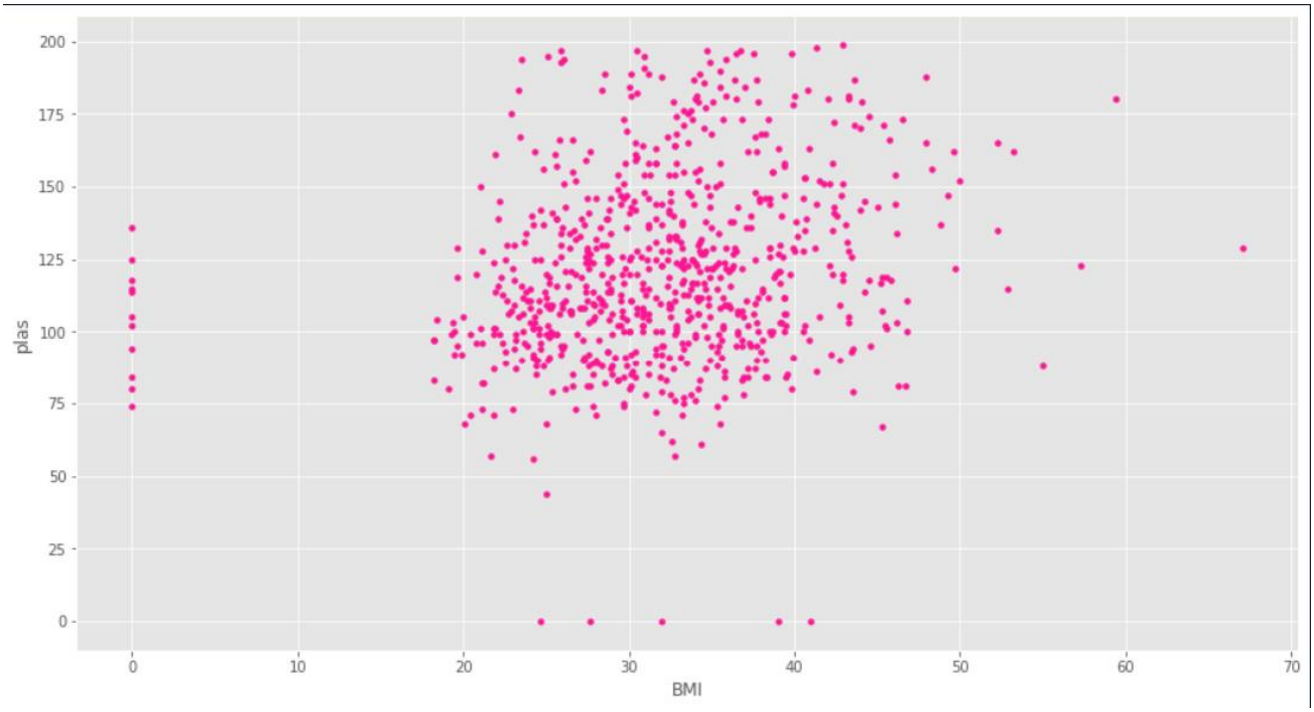

Figure 9 Scatter plot: BMI (in kg/m²) vs. plas

Inferences:

1. BMI (Kg/m_2) and Plas are positively Co-Related.
2. Large number of Plas are present in the BMI range of 20-40.

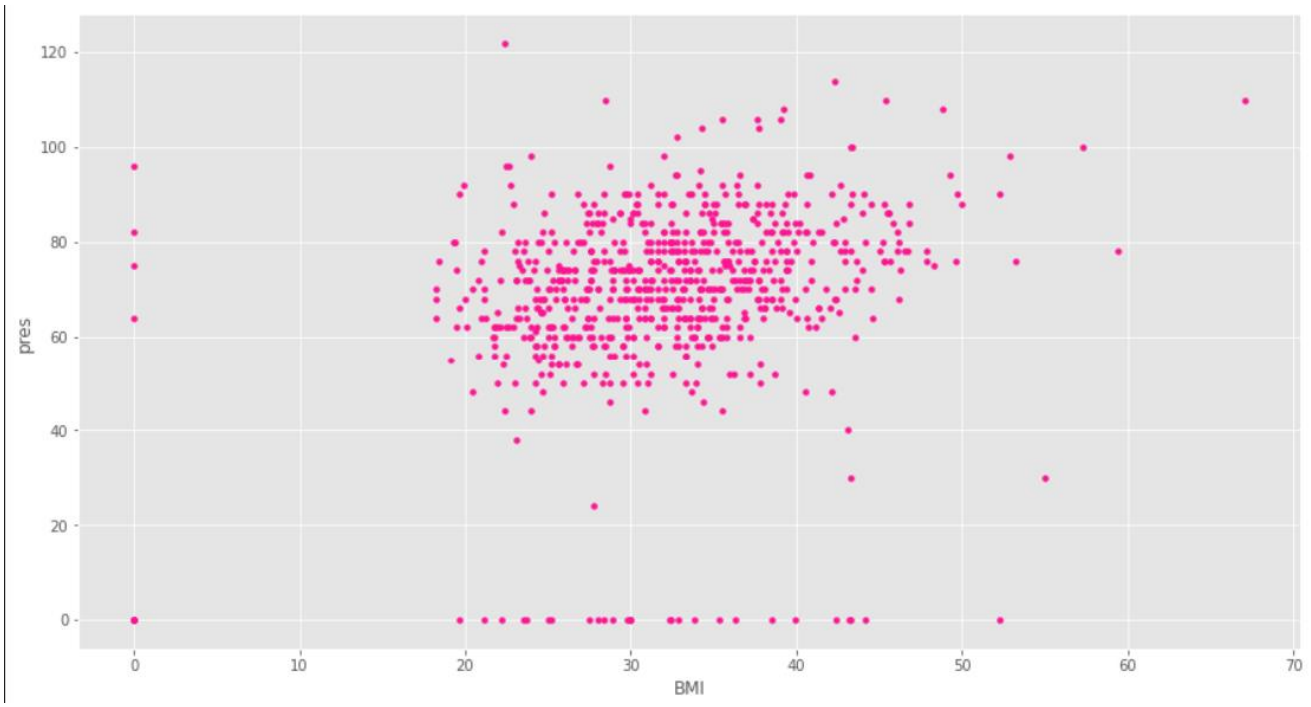

Figure 10 Scatter plot: BMI (in kg/m^2) vs. pres (in mm Hg)

Inferences:

1. BMI (Kg/m$_2$) and Pres are positively Co-Related.
2. Pres is mostly present in the BMI range of 20-40.

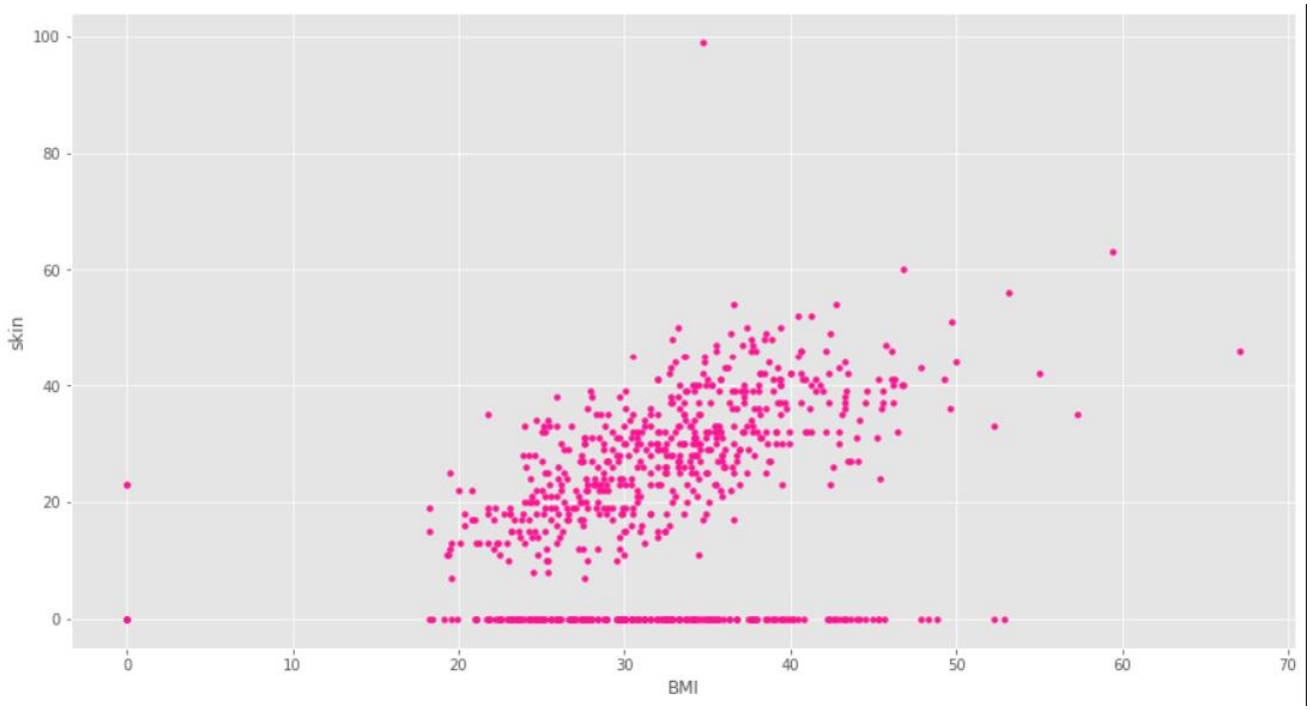

Figure 11 Scatter plot: BMI (in kg/m^2) vs. skin (in mm)

Inferences:

1. BMI and Skin are positively Co-Related.
2. Skins are hugely present between the BMI range of 30-40.

Figure 12 Scatter plot: BMI (in kg/m^2) vs. test (in mm U/mL)

Inferences:

1. BMI and Test are positively Co-Related.
2. The number of tests is present even in the small BMI differences but the number of tests revolve around 200 only in the BMI gap of 20-40.

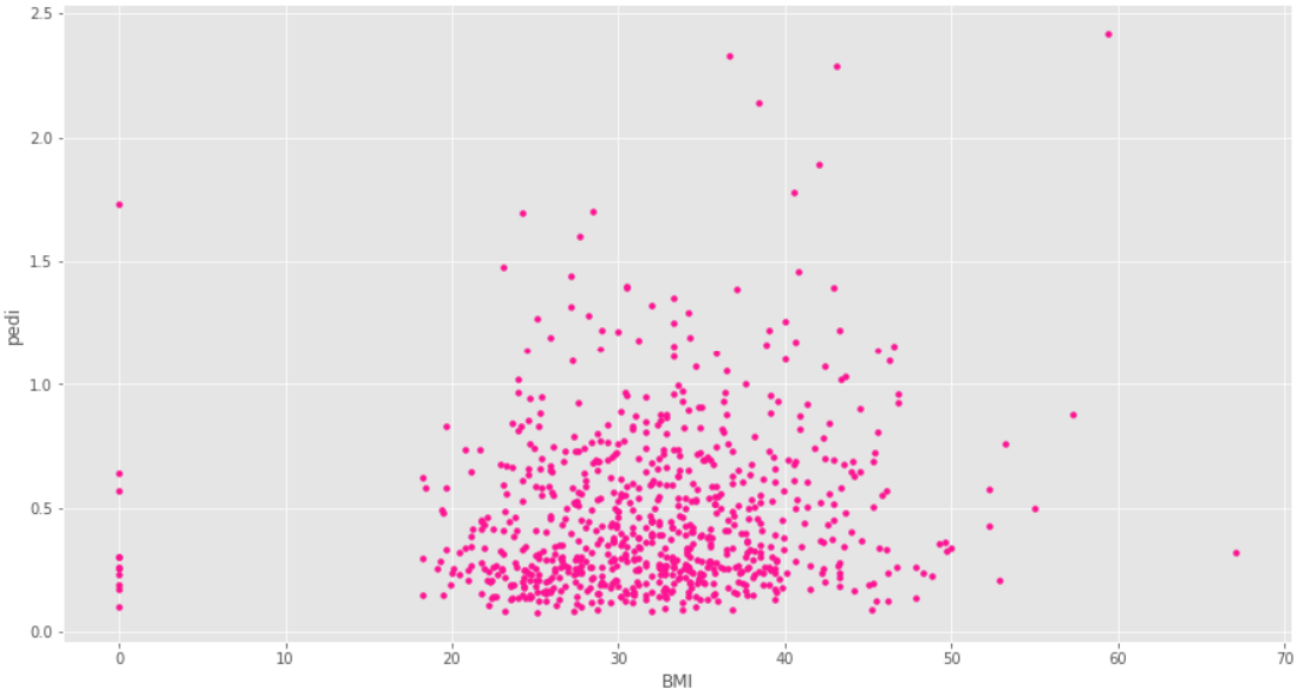

Figure 13 Scatter plot: BMI (in kg/m^2) vs. pedi

Inferences:

1. BMI and Pedi are Positively Co-Related.
2. The number of pedi are densely present in the BMI range of 20-40, after this BMI range the number scatters and becomes sparsely populated.

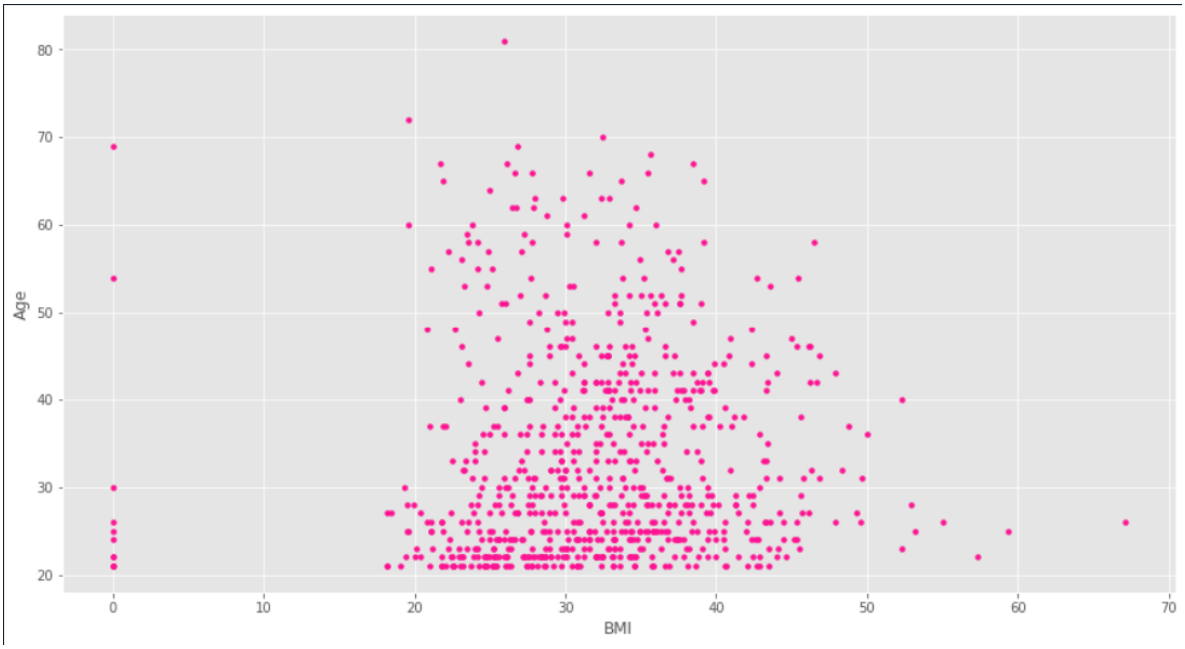

Figure 14 Scatter plot: BMI (in kg/m²) vs. Age (in years)

Inferences:

1. BMI and Age are positively Co-Related.
2. The BMI range of 20-30 is mainly maintained by the people of age-gap of 20-30 years.

3 a.

Table 3 Correlation coefficient value computed between age and all other attributes

S. No.	Attributes	Correlation Coefficient Value
1	pregs	0.544
2	plas	0.264
3	pres (in mm Hg)	0.240
4	skin (in mm)	-0.114
5	test (in mu U/mL)	-0.042
6	BMI (in kg/m^2)	0.036
7	pedi	0.034
8	Age (in years)	1.0

Inferences:

1. From the magnitude of correlation coefficient value, comment on the degree of correlation between age and each of the attribute.
2. From the sign of correlation coefficient value, comment whether with increase or decrease in age each of the attributes will increase or decrease.
3. Relate and comment on the value of correlation coefficient with corresponding scatter plot.
4. Inference 4(You may add or delete the number of inferences)

b.

Table 4 Correlation coefficient value computed between BMI and all other attributes

S. No.	Attributes	Correlation Coefficient Value
1	pregs	0.018
2	plas	0.221
3	pres (in mm Hg)	0.282
4	skin (in mm)	0.393
5	test (in mu U/mL)	0.198
6	BMI (in kg/m^2)	1.0
7	pedi	0.141
8	Age (in years)	0.036

Inferences:

1. From the magnitude of correlation coefficient value, comment on the degree of correlation between age and each of the attribute.
2. From the sign of correlation coefficient value, comment whether with increase or decrease in age each of the attributes will increase or decrease.
3. Relate and comment on the value of correlation coefficient with corresponding scatter plot.
4. Inference 4(You may add or delete the number of inferences)

4 a.

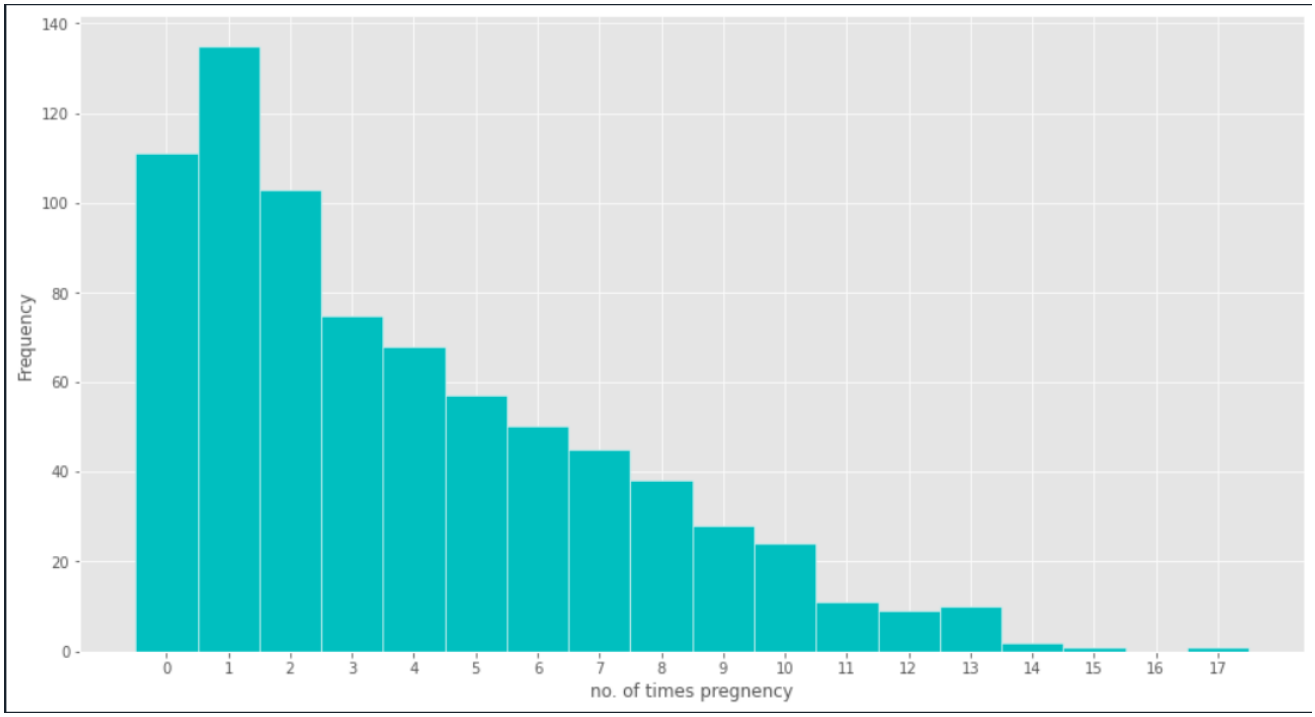

Figure 15 Histogram depiction of attribute pregs

Inferences:

1. Infer the frequency of each bin referring to its height.
2. From the histogram, infer in which of the bins mode of the attribute pregs lies.
3. Inference 3(You may add or delete the number of inferences)
   Note: The histogram plot above is for illustration purpose. Replace it with the histogram plot obtained by you. Rename x-axis legend and y-axis legends with appropriate attribute names with units.

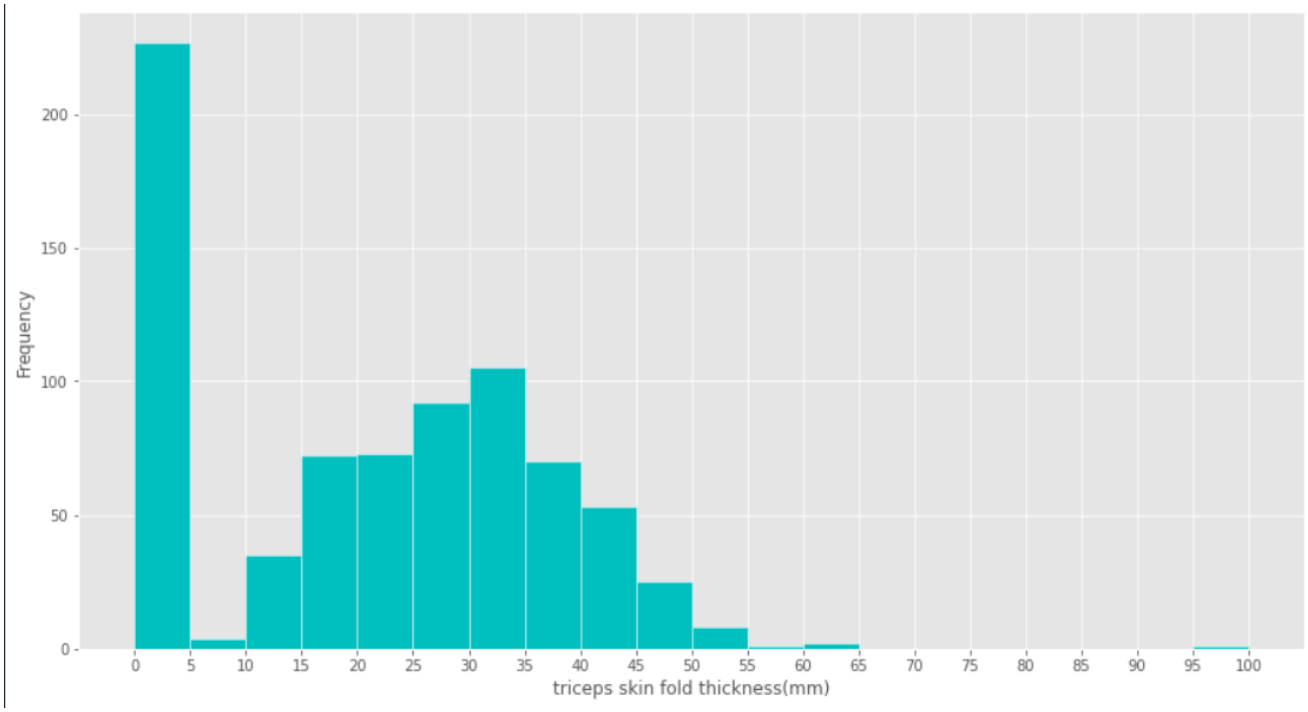

Figure 16 Histogram depiction of attribute skin

Inferences:

1. Infer the frequency of each bin referring to its height.
2. From the histogram, infer in which of the bins mode of the attribute skin lies.
3. Inference 3(You may add or delete the number of inferences)

Note: The histogram plot above is for illustration purpose. Replace it with the histogram plot obtained by you. Rename x-axis legend and y-axis legends with appropriate attribute names with units.

5

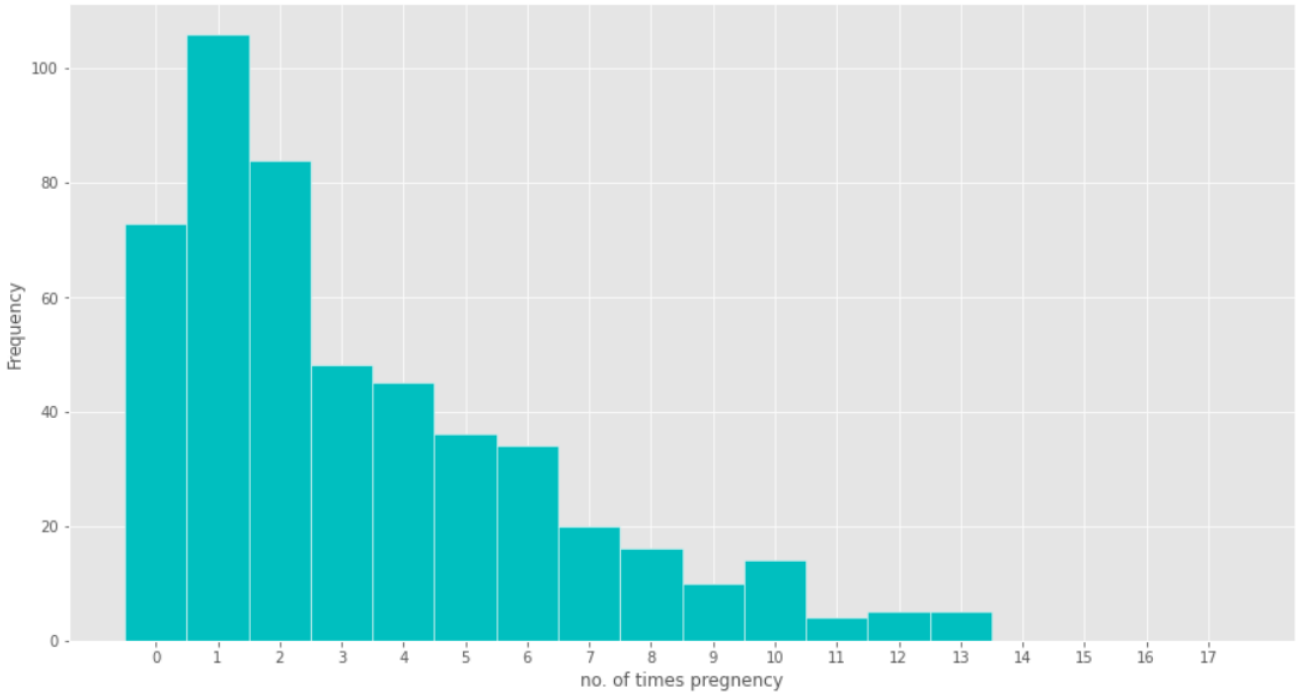

Figure 17 Histogram depiction of attribute pregs for class 0

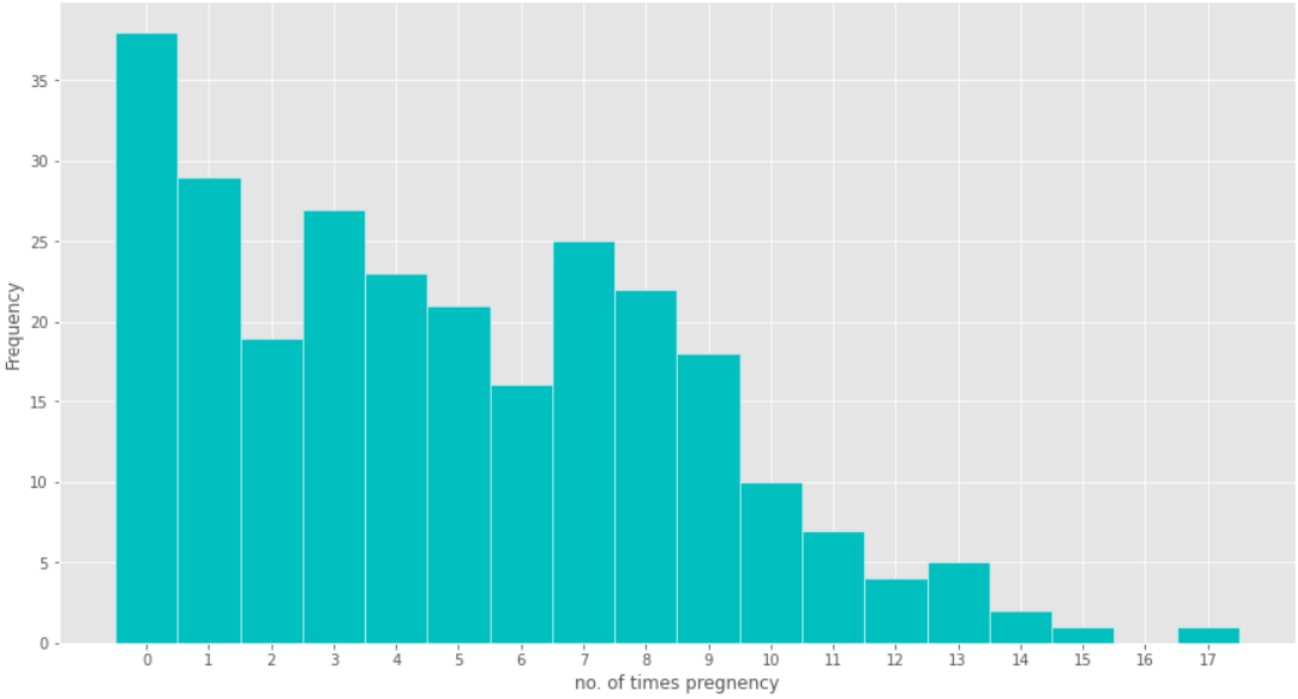

Figure 18 Histogram depiction of attribute pregs for class 1

Inferences:

1. From the histogram, infer in which of the bins mode of the attribute pregs lies for class 0 and 1.
2. Compare and contrast the frequency referring to the height of each bin for class 0 and 1
3. Inference 3(You may add or delete the number of inferences)
   Note: The histogram plot above is for illustration purpose. Replace it with the histogram plot obtained by you. Rename x-axis legend and y-axis legends with appropriate attribute names with units.

6

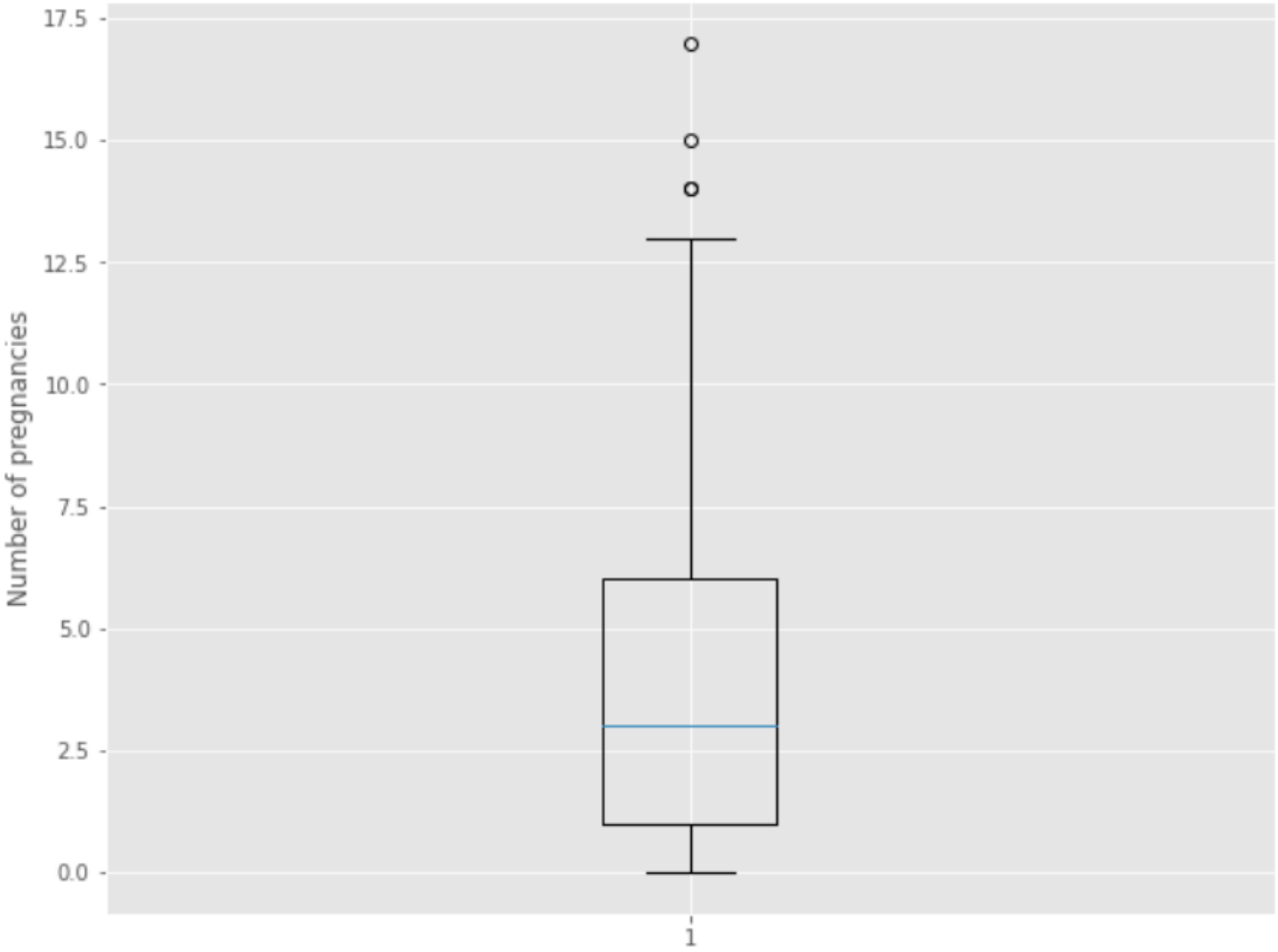

Figure 19 Boxplot for attribute pregs

Inferences:

1. Inference on outliers and their values.
2. Infer the Inter quartile range.
3. Infer the variability of attribute.
4. Infer the skewness of the data.
5. Relate with the values from Q1. for this attribute.
6. Inference 6(You may add or delete the number of inferences)

   Note: The boxplot above is for illustration purpose. Replace it with the boxplot obtained by you. Rename x-axis legend and y-axis legends with appropriate attribute names with units.

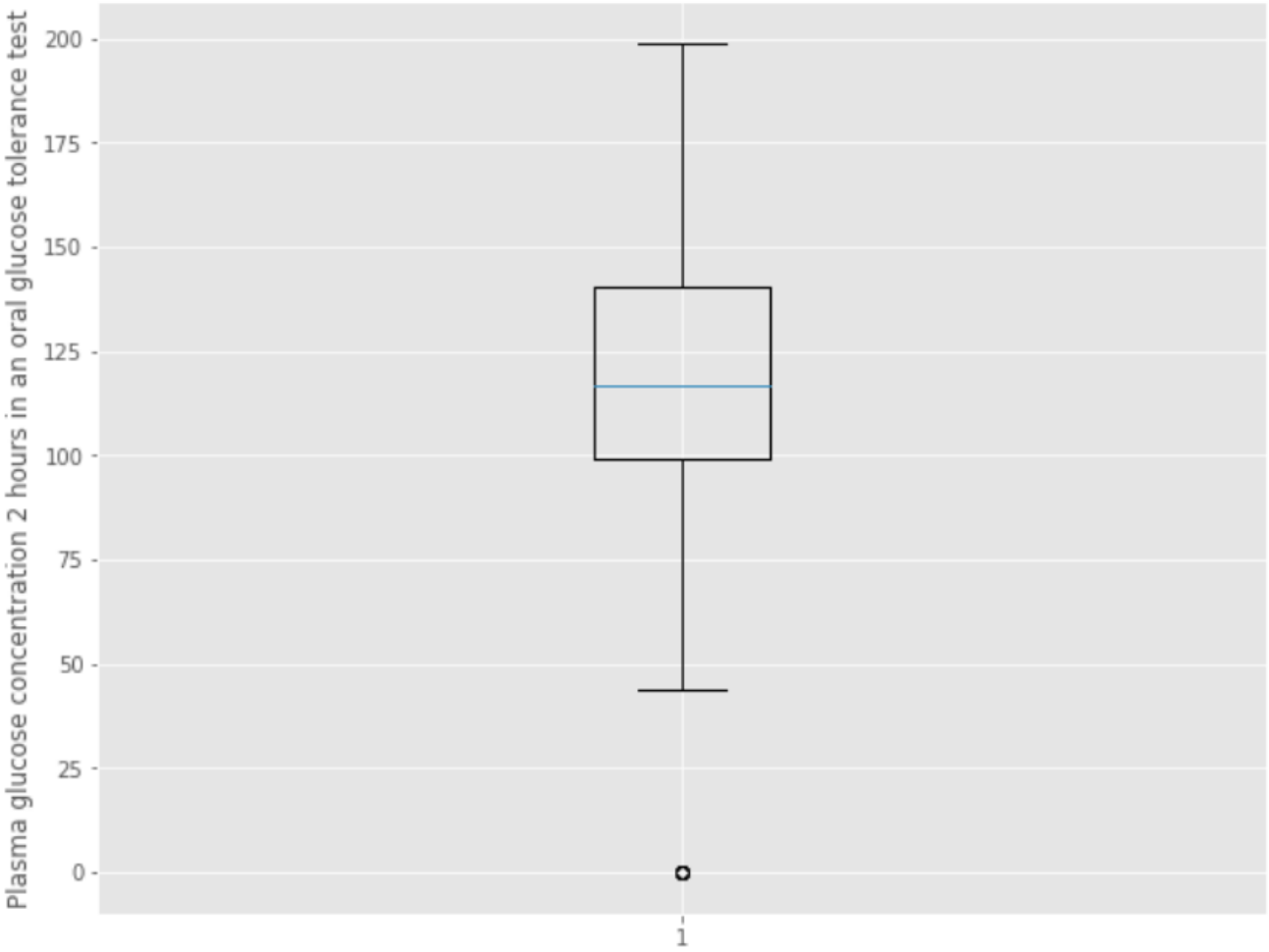

Figure 20 Boxplot for attribute plas

Inferences:

1. Inference on outliers and their values.
2. Infer the Inter quartile range.
3. Infer the variability of attribute.
4. Infer the skewness of the data.
5. Relate with the values from Q1. for this attribute.
6. Inference 6(You may add or delete the number of inferences)

Note: The boxplot above is for illustration purpose. Replace it with the boxplot obtained by you. Rename x-axis legend and y-axis legends with appropriate attribute names with units.

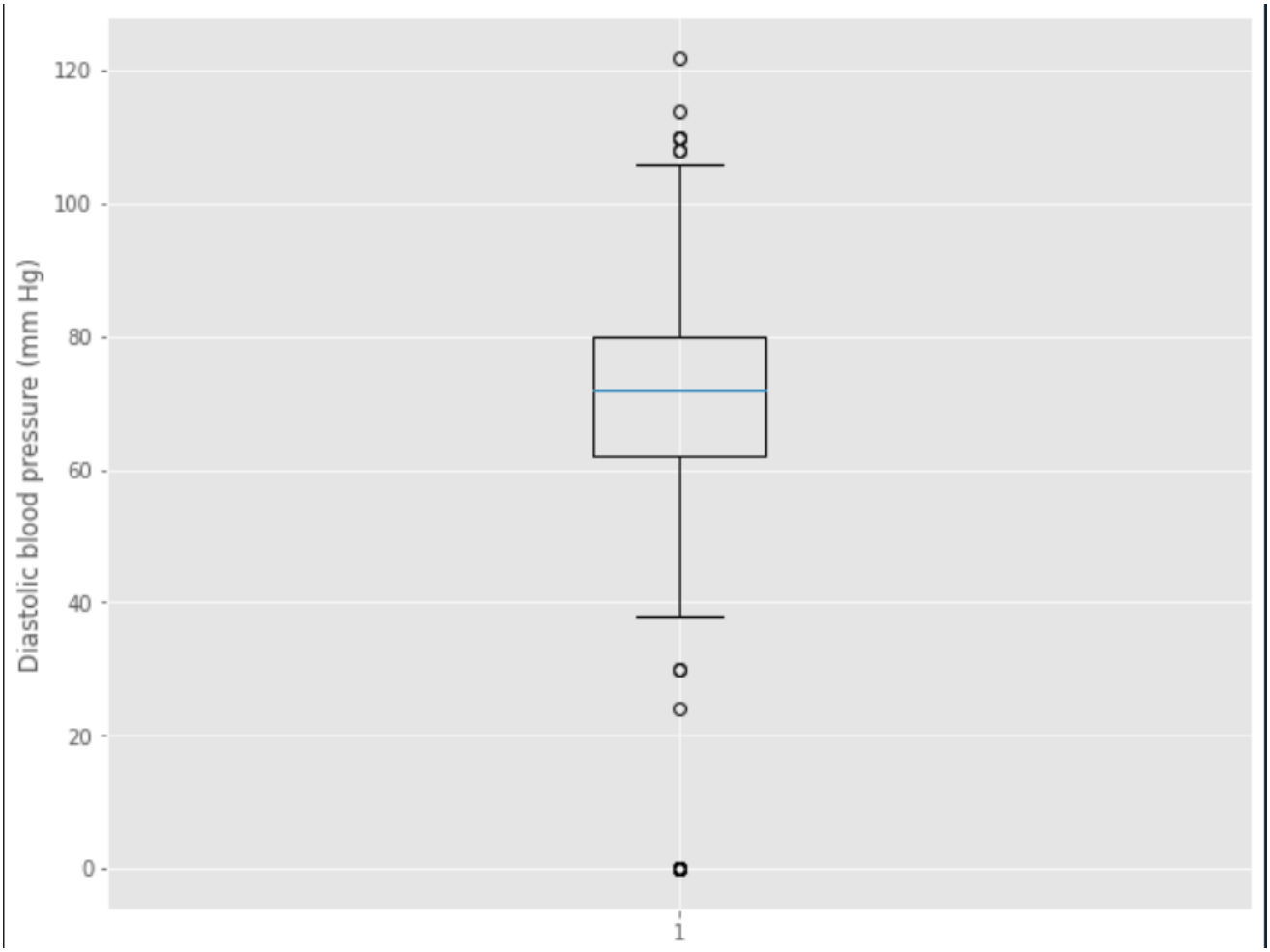

Figure 21 Boxplot for attribute pres(in mm Hg)

Inferences:

1. Inference on outliers and their values.
2. Infer the Inter quartile range.
3. Infer the variability of attribute.
4. Infer the skewness of the data.
5. Relate with the values from Q1. for this attribute.
6. Inference 6(You may add or delete the number of inferences)

Note: The boxplot above is for illustration purpose. Replace it with the boxplot obtained by you. Rename x-axis legend and y-axis legends with appropriate attribute names with units.

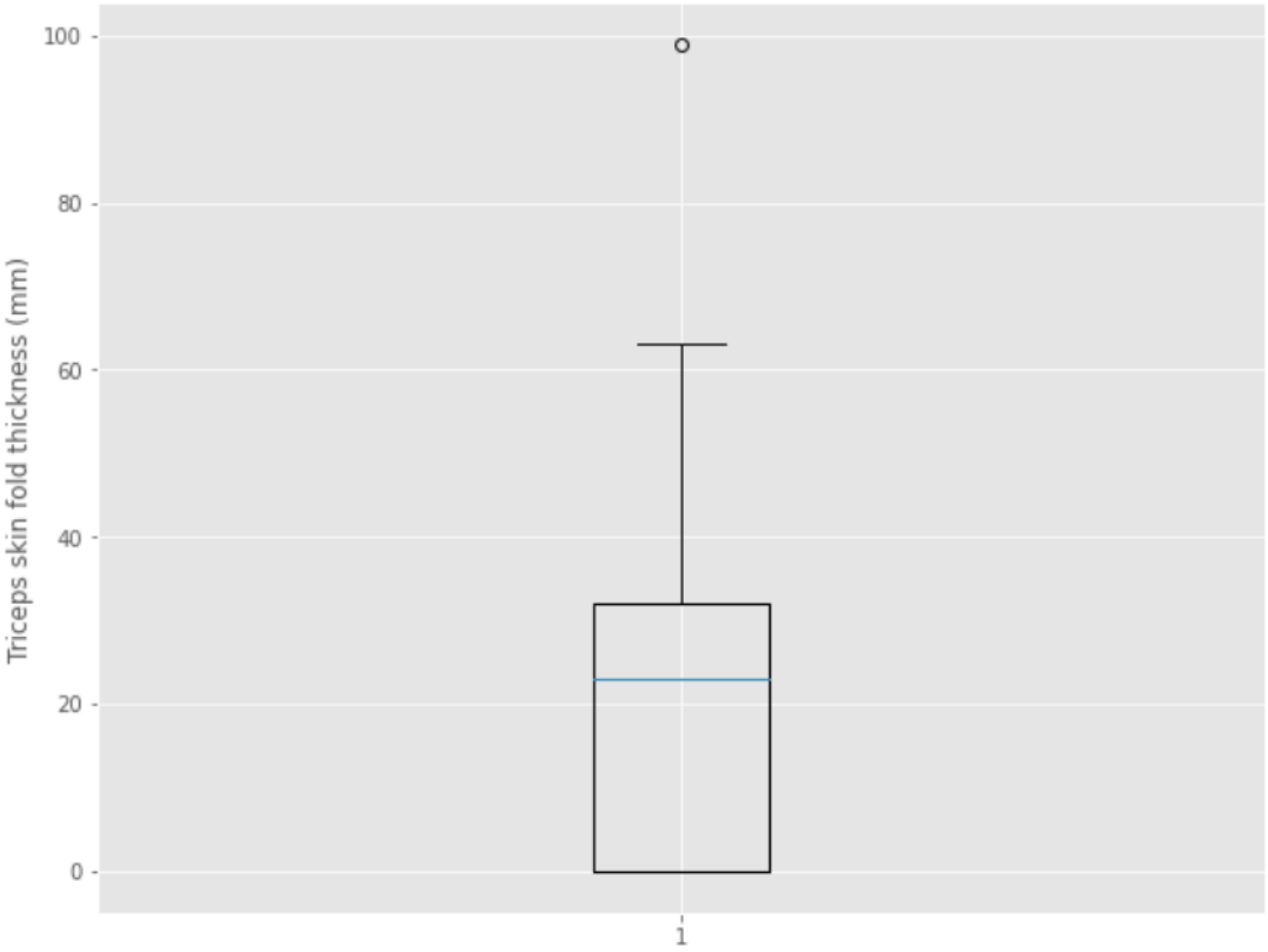

Figure 22 Boxplot for attribute skin(in mm)

Inferences:

1. Inference on outliers and their values.
2. Infer the Inter quartile range.
3. Infer the variability of attribute.
4. Infer the skewness of the data.
5. Relate with the values from Q1. for this attribute.
6. Inference 6(You may add or delete the number of inferences)

   Note: The boxplot above is for illustration purpose. Replace it with the boxplot obtained by you. Rename x-axis legend and y-axis legends with appropriate attribute names with units.

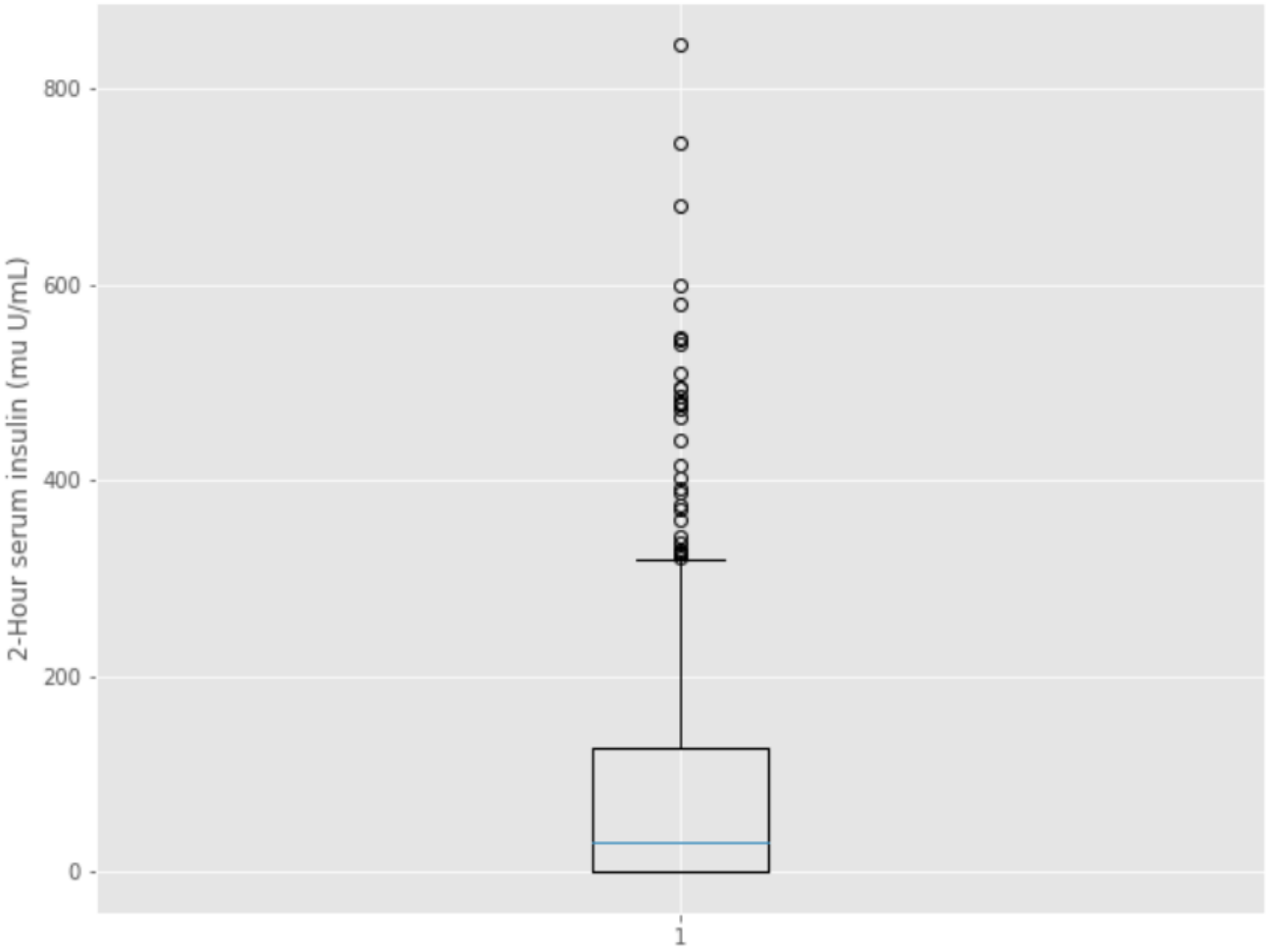

Figure 23 Boxplot for attribute test (mu U/mL)

Inferences:

1. Inference on outliers and their values.
2. Infer the Inter quartile range.
3. Infer the variability of attribute.
4. Infer the skewness of the data.
5. Relate with the values from Q1. for this attribute.
6. Inference 6(You may add or delete the number of inferences)

Note: The boxplot above is for illustration purpose. Replace it with the boxplot obtained by you. Rename x-axis legend and y-axis legends with appropriate attribute names with units.

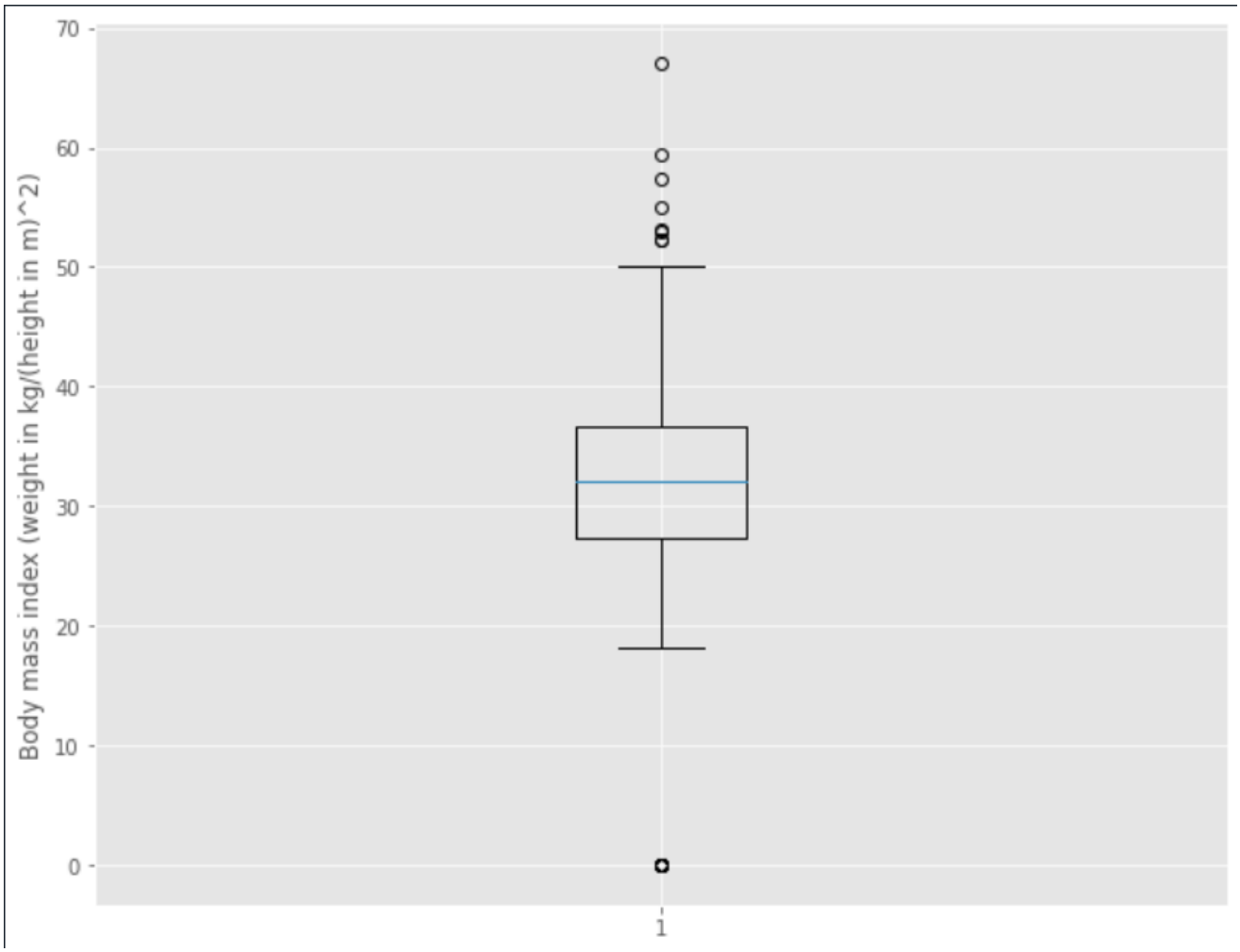

Figure 24 Boxplot for attribute BMI (in kg/m^2)

Inferences:

1. Inference on outliers and their values.
2. Infer the Inter quartile range.
3. Infer the variability of attribute.
4. Infer the skewness of the data.
5. Relate with the values from Q1. for this attribute.
6. Inference 6(You may add or delete the number of inferences)

Note: The boxplot above is for illustration purpose. Replace it with the boxplot obtained by you. Rename x-axis legend and y-axis legends with appropriate attribute names with units.

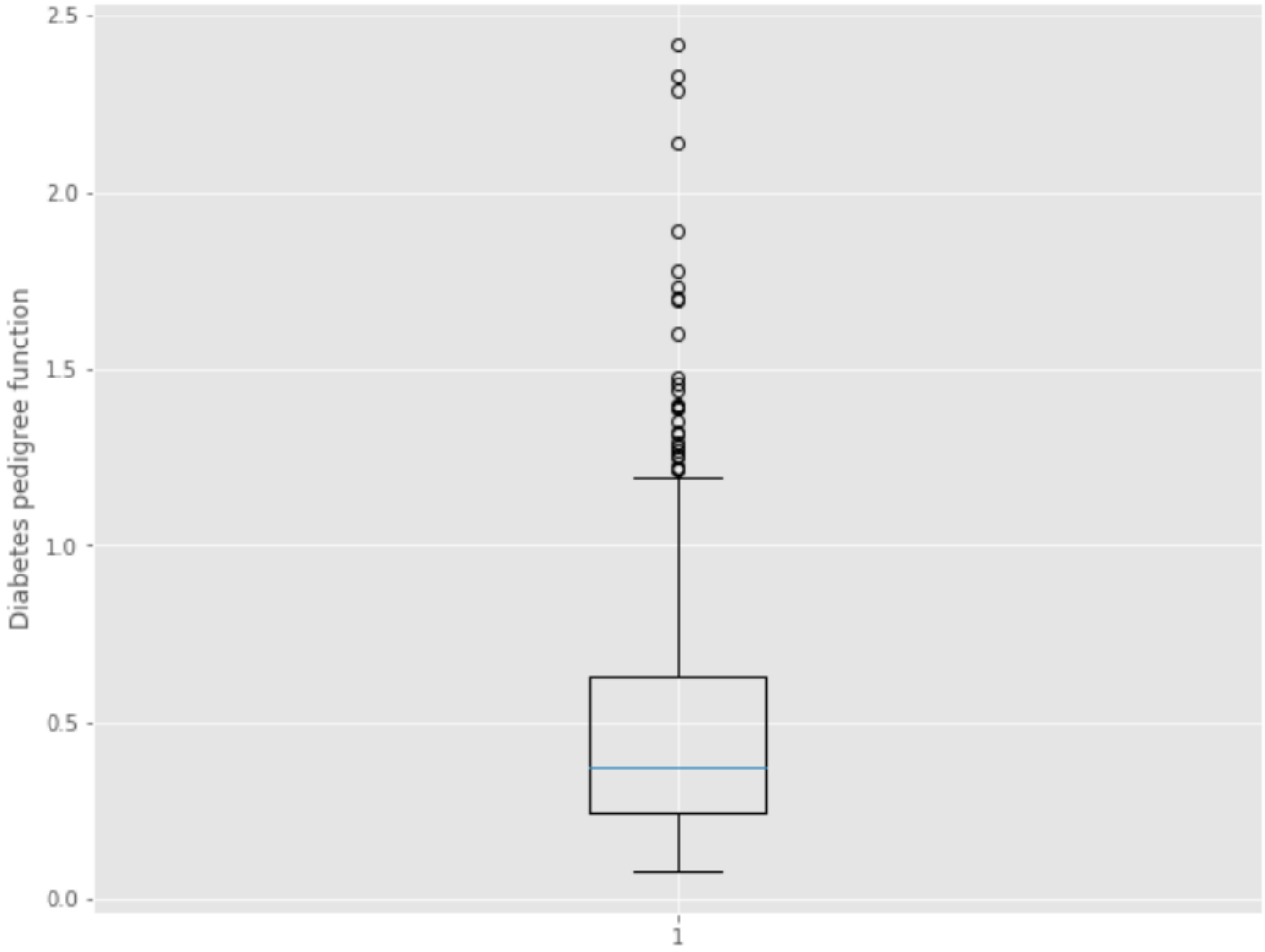

Figure 25 Boxplot for attribute pedi

Inferences:

1. Inference on outliers and their values.
2. Infer the Inter quartile range.
3. Infer the variability of attribute.
4. Infer the skewness of the data.
5. Relate with the values from Q1. for this attribute.
6. Inference 6(You may add or delete the number of inferences)

Note: The boxplot above is for illustration purpose. Replace it with the boxplot obtained by you. Rename x-axis legend and y-axis legends with appropriate attribute names with units.

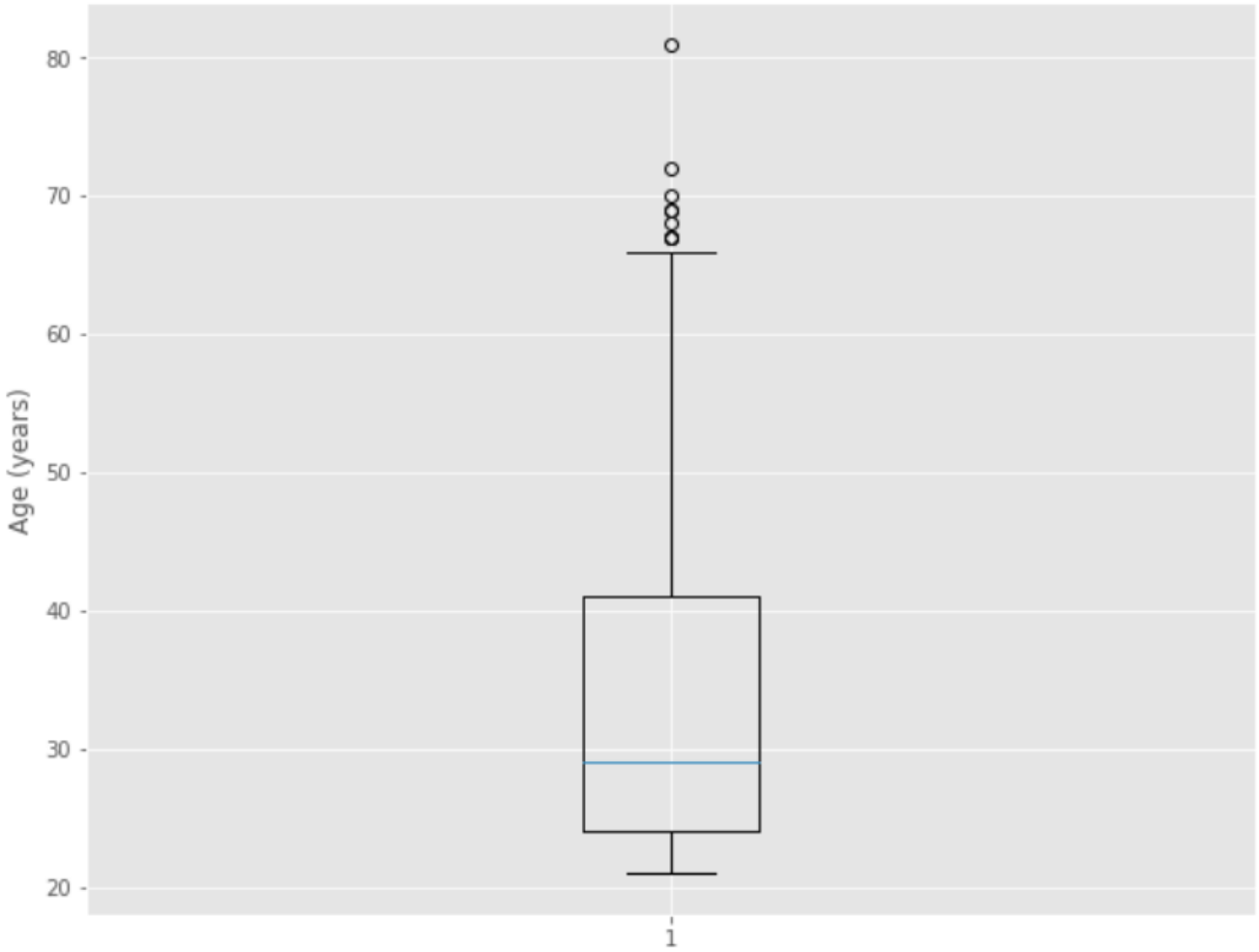

Figure 26 Boxplot for attribute Age (in years)

Inferences:

1. Inference on outliers and their values.
2. Infer the Inter quartile range.
3. Infer the variability of attribute.
4. Infer the skewness of the data.
5. Inference 5(You may add or delete the number of inferences)

   Note: The boxplot above is for illustration purpose. Replace it with the boxplot obtained by you. Rename x-axis legend and y-axis legends with appropriate attribute names with units