



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

Student's Name: Ishaan Gupta

Mobile No: 9179242114

Roll Number: B20292

Branch: Mechanical Engineering

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	1	0	17	3.370
2	plas	120.895	117	100	0	199	31.973
3	pres (in mm Hg)	69.105	72	70	0	122	19.356
4	skin (in mm)	20.536	23	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	32	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	22	21	81	11.760

Inferences:

1. If standard deviation is close to zero; then mean, median and mode are close to each other.
2. If standard deviation is lesser than mean and median then the values of mean and median are close to each other.

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

2 a.

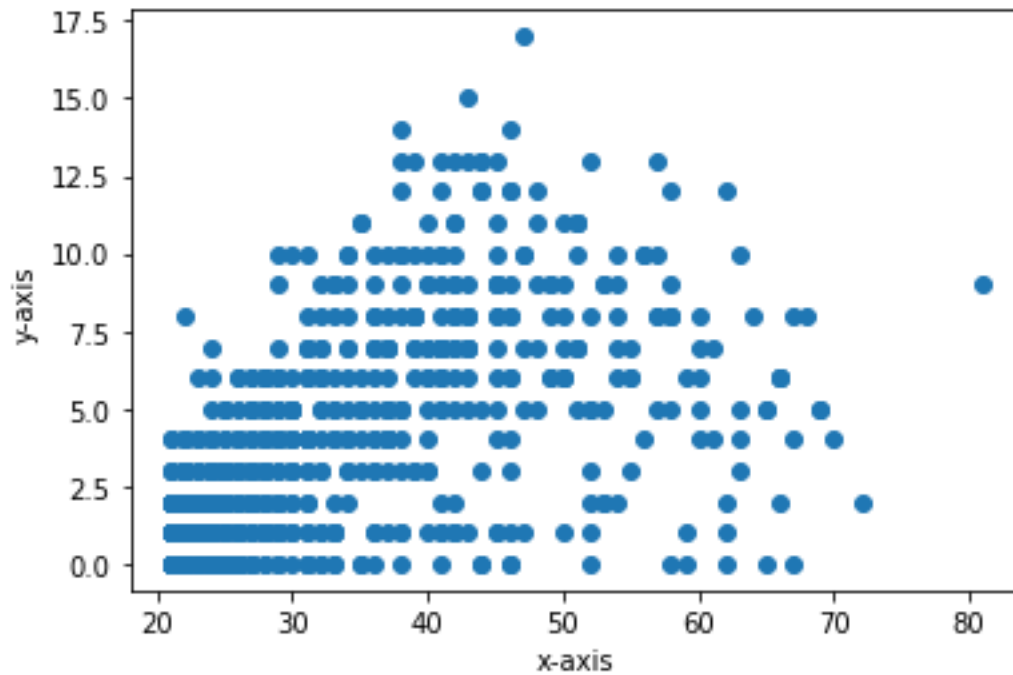


Figure 1 Scatter plot: Age (in years) vs. pregs

Inferences:

1. There is no correlation between both the attributes as when age increases from 20 to 50 years the number of times pregnant does not go up or down.
2. High density is seen when age is 20-30 years and number of times pregnant is 0-4 times.

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

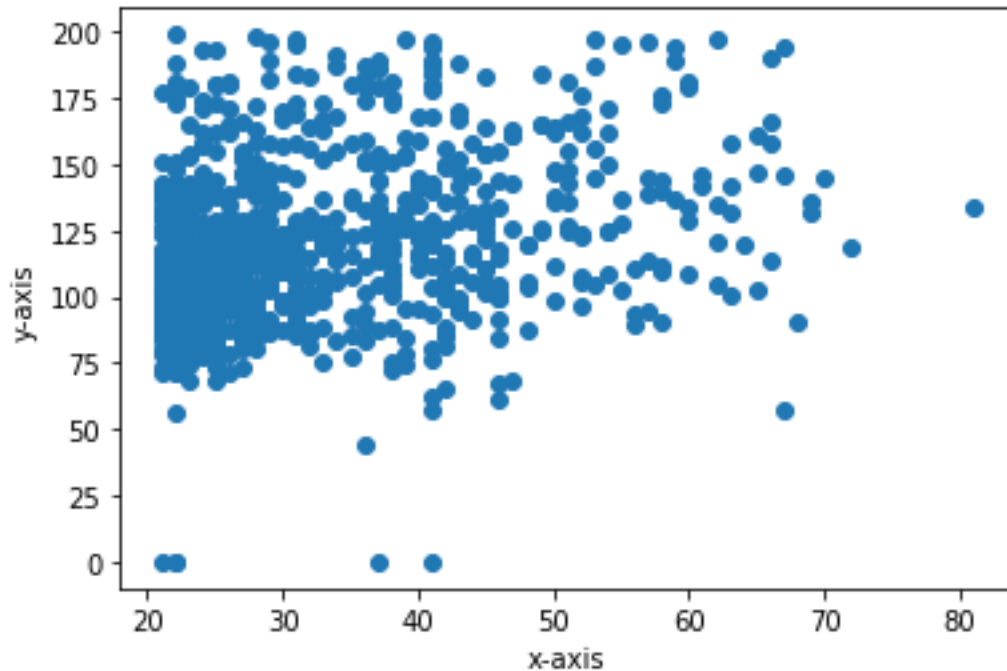


Figure 2 Scatter plot: Age (in years) vs. plas

Inferences:

1. There is no correlation between both the attributes as when age increases Plasma glucose concentration 2 hours in an oral glucose tolerance test is not increasing or decreasing.
2. High density is seen when age is 20-35 years and Plasma glucose concentration 2 hours in an oral glucose tolerance test is between 70-140.

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

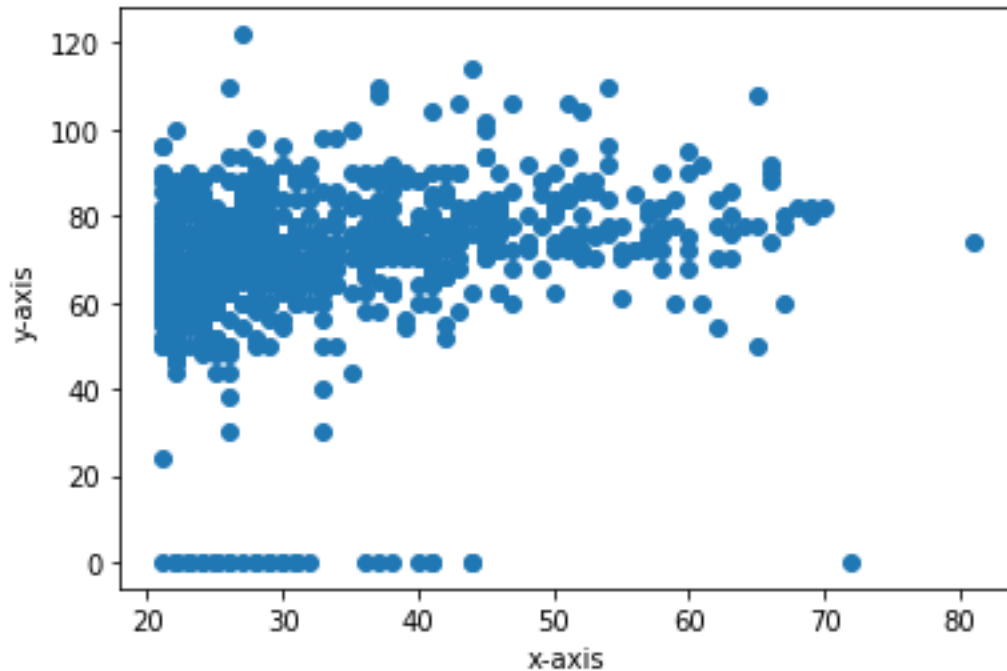


Figure 3 Scatter plot: Age (in years) vs. pres (in mm Hg)

Inferences:

1. There is no correlation between both the attributes as when age increases Diastolic blood pressure is not increasing or decreasing.
2. High density is seen when age is 20-35 years and Diastolic blood pressure is between 50-90 mm Hg.

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

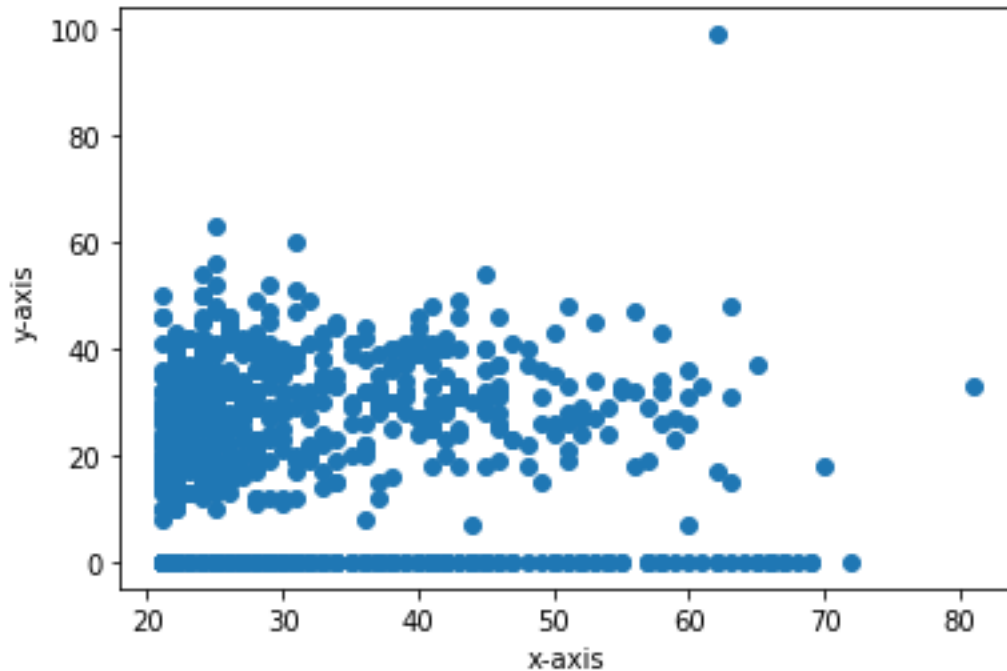


Figure 4 Scatter plot: Age (in years) vs. skin (in mm)

Inferences:

1. There is no correlation between both the attributes as when age increases Triceps skin fold thickness is not increasing or decreasing.
2. High density is seen when age is 20-35 years and Triceps skin fold thickness is between 10-40 mm.

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

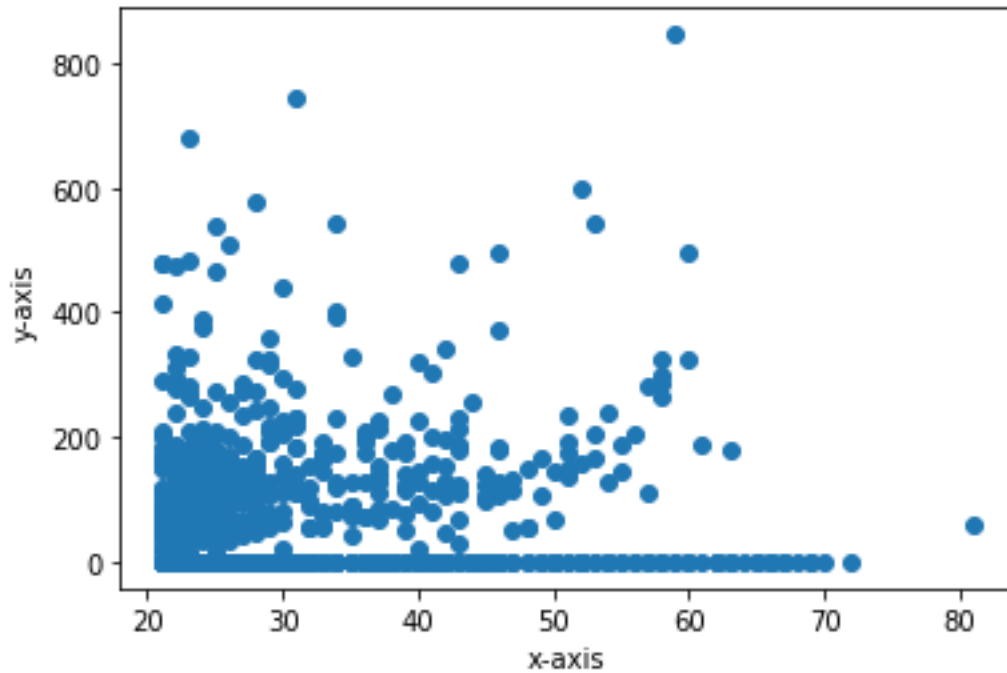


Figure 5 Scatter plot: Age (in years) vs. test (in mm U/mL)

Inferences:

1. There is a positive correlation between both the attributes as when age increases 2-Hour serum insulin is also increasing.
2. High density is seen when age is 20-30 years and 2-Hour serum insulin is between 0-200 mm U/ml.

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

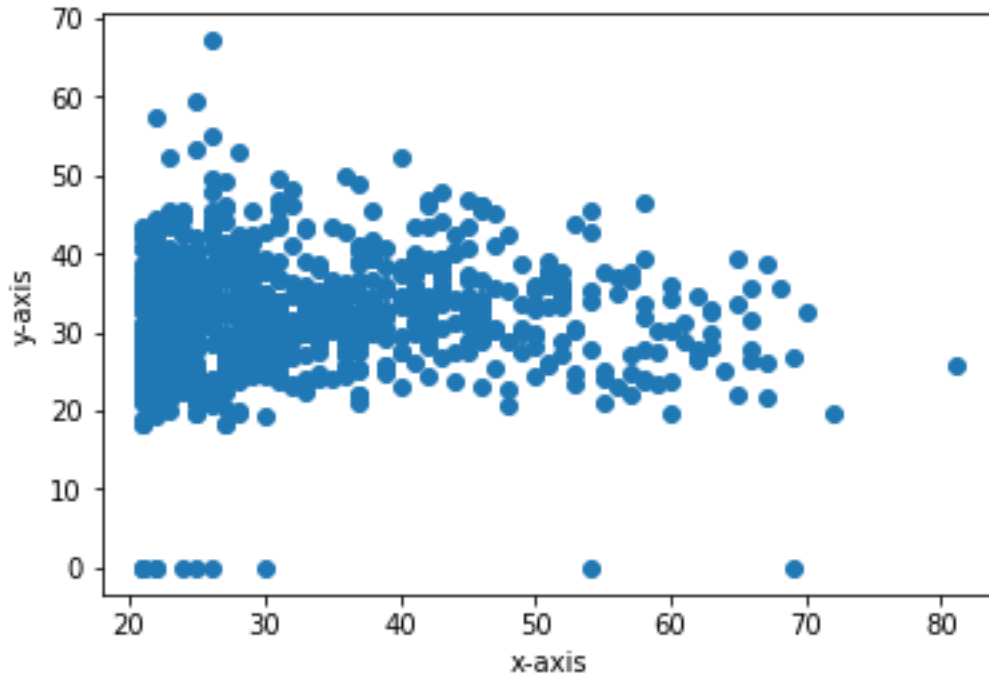


Figure 6 Scatter plot: Age (in years) vs. BMI (in kg/m²)

Inferences:

1. There is no correlation between both the attributes as when age increases BMI is not increasing or decreasing.
2. High density is seen when age is 20-40 years and BMI is between 20-40 kg/m².

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

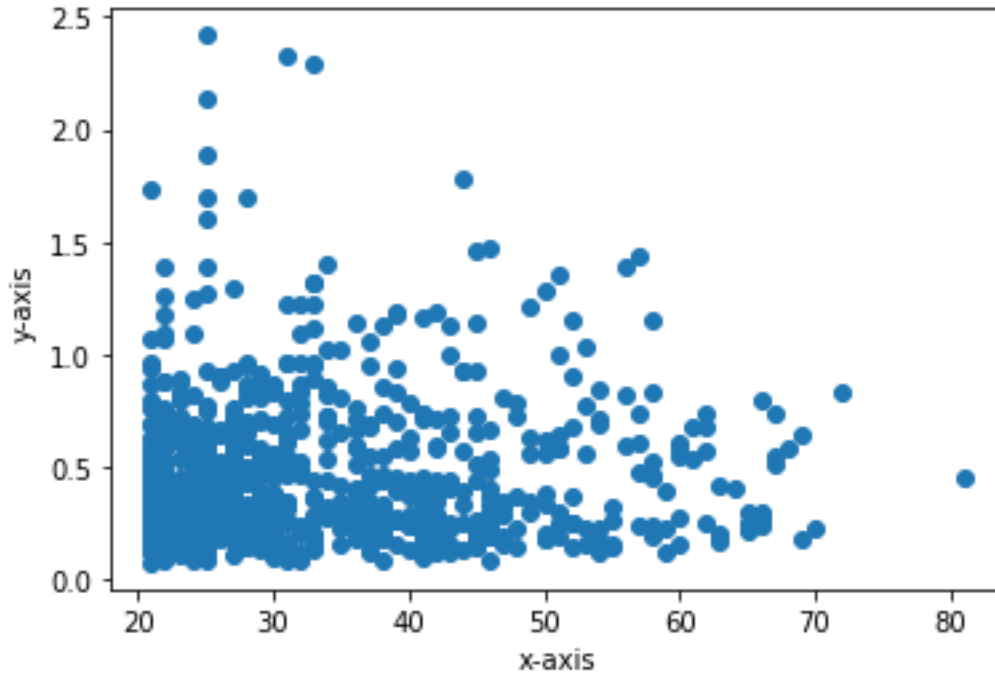


Figure 7 Scatter plot: Age (in years) vs. pedi

Inferences:

1. There is weak negative correlation between both the attributes as when age increases Diabetes pedigree function is decreasing with a few points only.
2. High density is seen when age is 20-40 years and Diabetes pedigree function is 0-1 .

b.

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

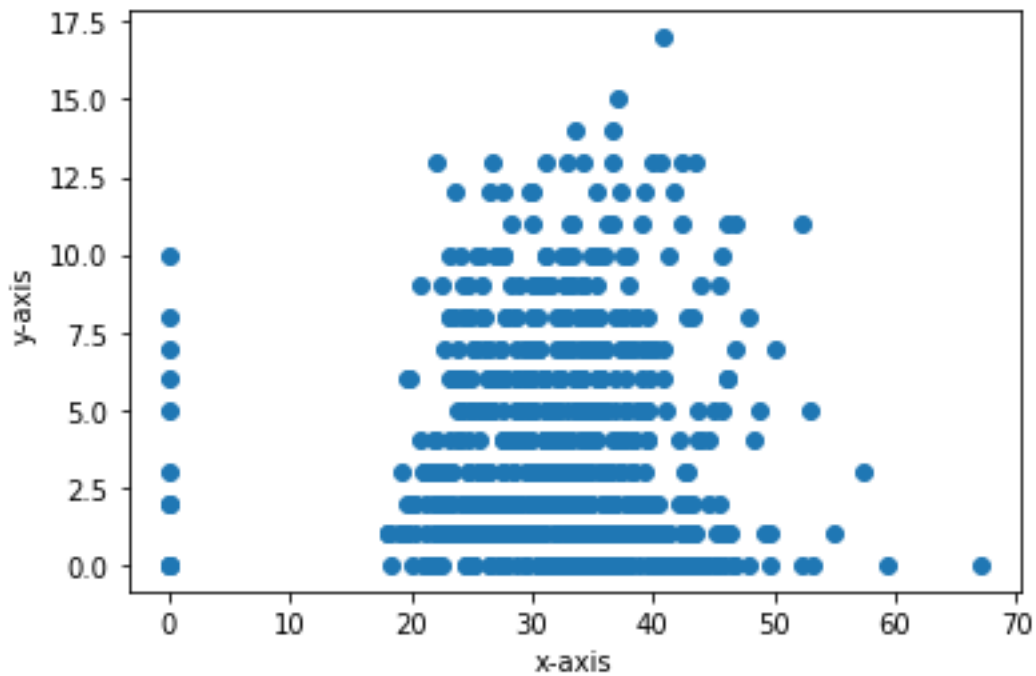


Figure 8 Scatter plot: BMI (in kg/m²) vs. pregs

Inferences:

1. There is no correlation between both the attributes as when BMI increases number of times pregnant is not increasing or decreasing.
2. High density is seen when BMI is 20-40 kg/m² and number of times pregnant is between 0-8.

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

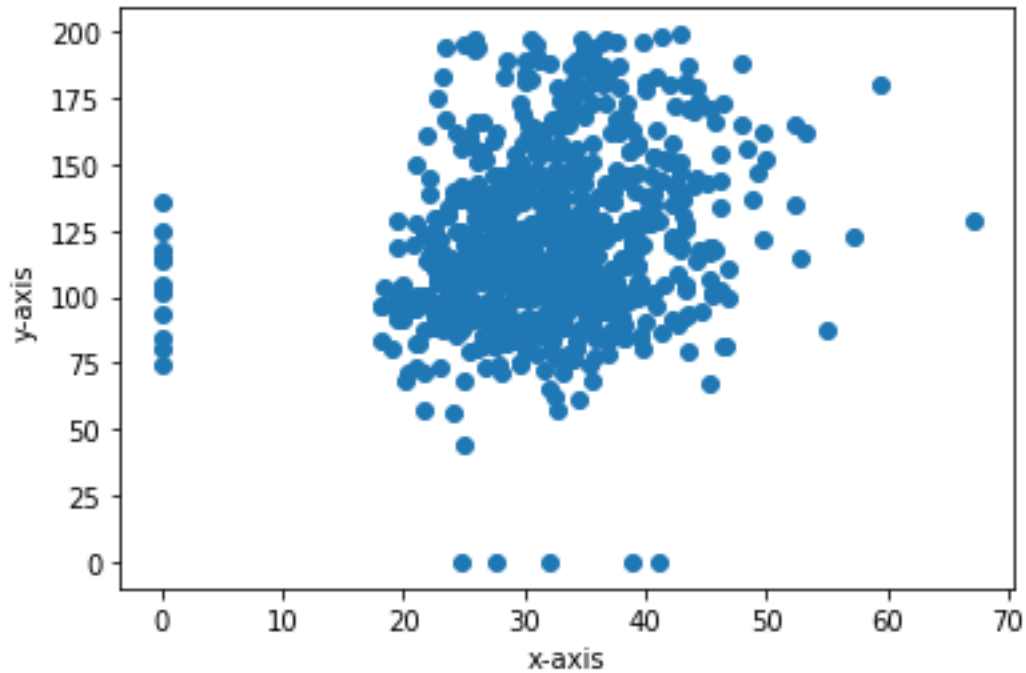


Figure 9 Scatter plot: BMI (in kg/m²) vs. plas

Inferences:

1. There is no correlation between both the attributes as when BMI increases Plasma glucose concentration 2 hours in an oral glucose tolerance test is not increasing or decreasing.
2. High density is seen when BMI is 20-45 kg/m² and Plasma glucose concentration 2 hours in an oral glucose tolerance test is between 75-165 .

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

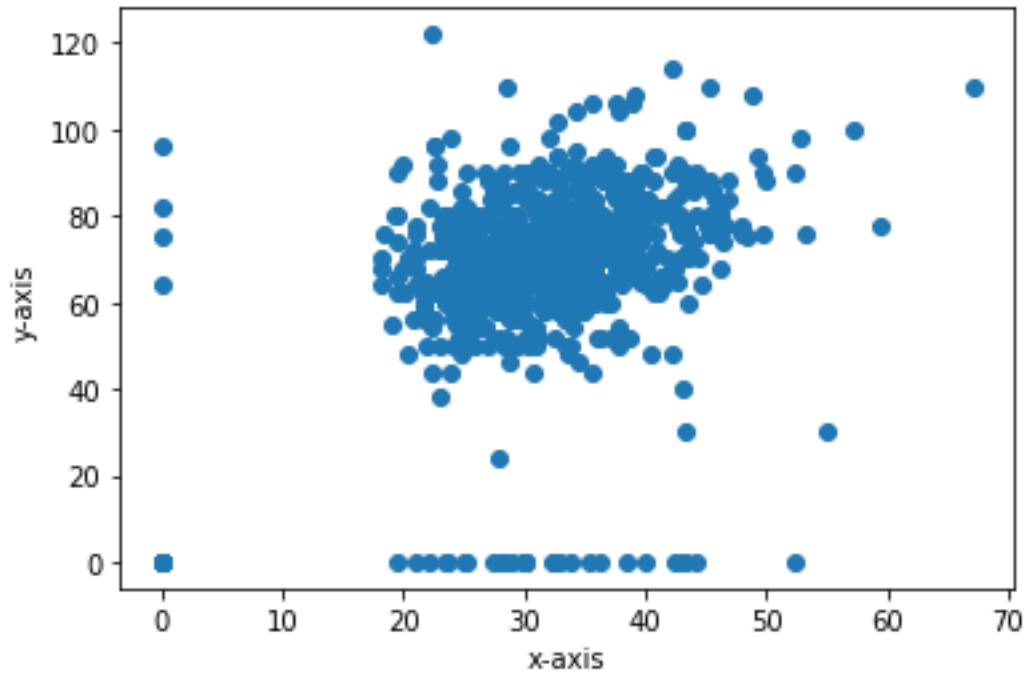


Figure 10 Scatter plot: BMI (in kg/m^2) vs. pres (in mm Hg)

Inferences:

1. There is no correlation between both the attributes as when BMI increases Diastolic blood pressure is not increasing or decreasing.
2. High density is seen when BMI is 20-40 kg/m^2 and Diastolic blood pressure is between 60-90 mm Hg .

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

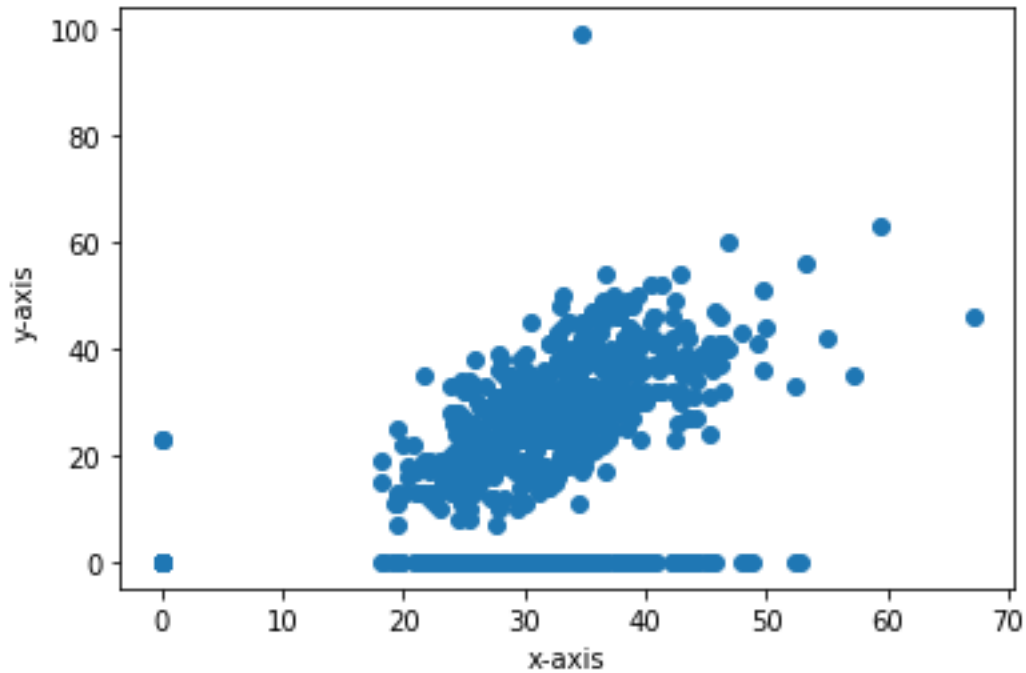


Figure 11 Scatter plot: BMI (in kg/m²) vs. skin (in mm)

Inferences:

1. There is a positive correlation between both the attributes as when BMI increases Triceps skin fold thickness is increasing .
2. High density is seen when BMI is 20-40 kg/m² and Triceps skin fold thickness is between 10-50 mm .

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

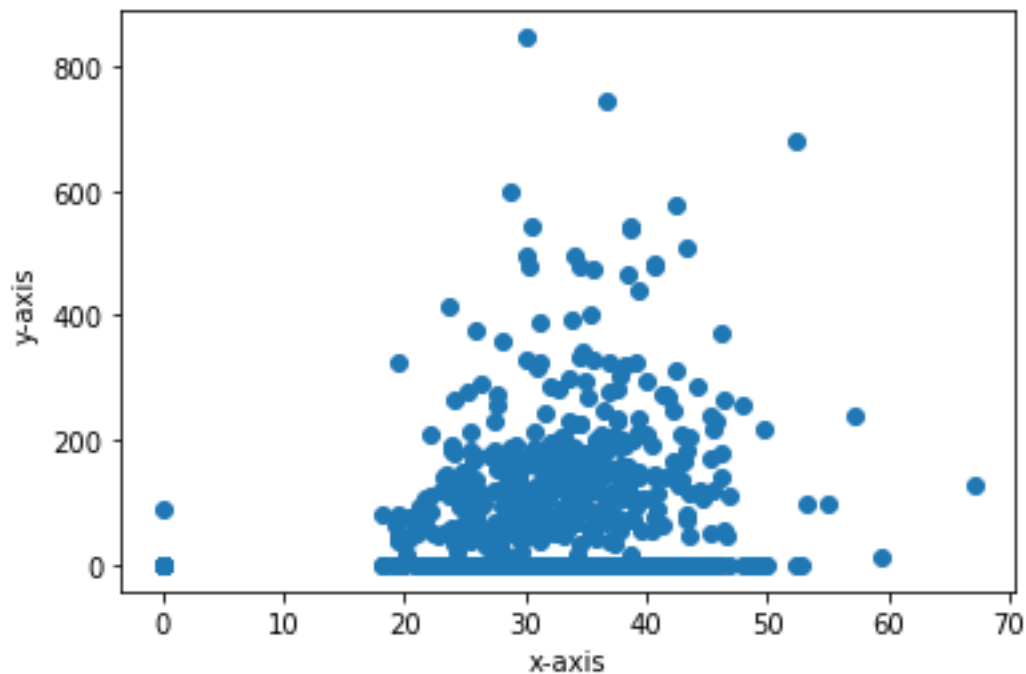


Figure 12 Scatter plot: BMI (in kg/m^2) vs. test (in mm U/mL)

Inferences:

1. There is a weak negative correlation between both the attributes as when BMI increases 2-Hour serum insulin is decreasing.
2. High density is seen when BMI is 20-40 kg/m^2 and 2-Hour serum insulin is between 0-200 mm U/mL.

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

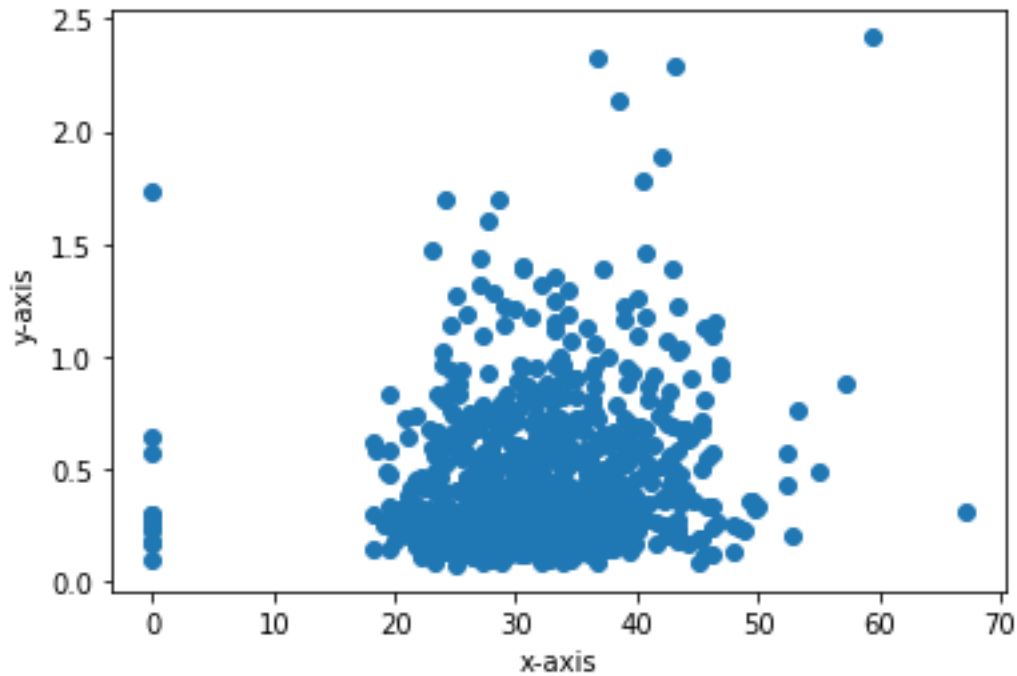


Figure 13 Scatter plot: BMI (in kg/m²) vs. pedi

Inferences:

1. There is no correlation between both the attributes as when BMI increases Diabetes pedigree function is not increasing or decreasing.
2. High density is seen when BMI is 20-45 kg/m² and Diabetes pedigree function is between 0-1 .

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

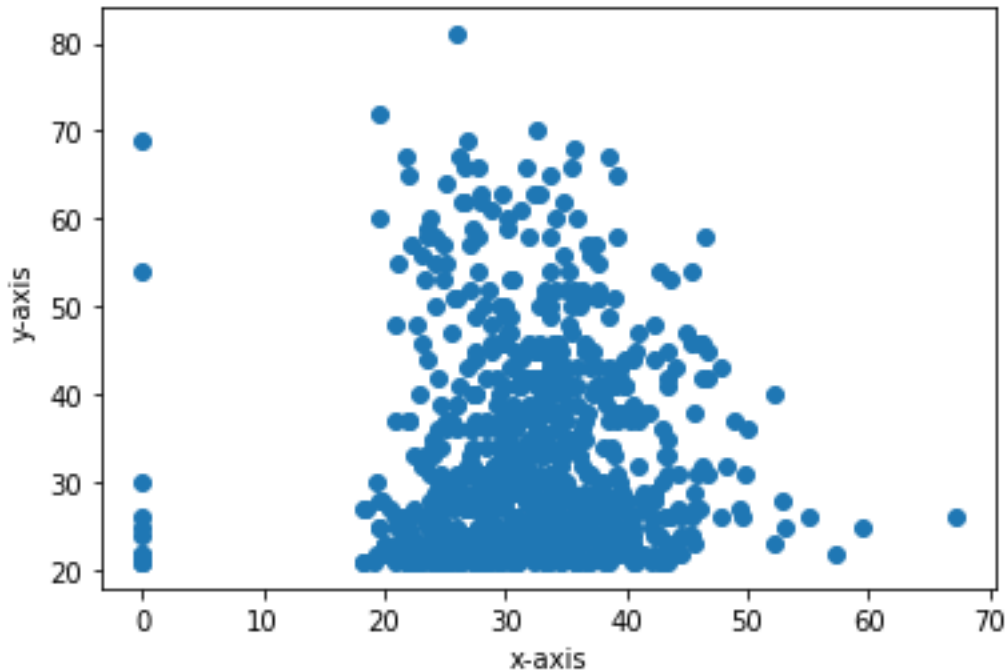


Figure 14 Scatter plot: BMI (in kg/m^2) vs. Age (in years)

Inferences:

1. There is no correlation between both the attributes as when BMI increases Age is not increasing or decreasing.
2. High density is seen when BMI is 20-40 kg/m^2 and Age is between 20-40 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

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

5	test (in μ U/mL)	-0.042
6	BMI (in kg/m^2)	0.036
7	pedi	0.034
8	Age (in years)	1.000

Inferences:

1. As age increases pregs increases.
2. As age increases plas increases.
3. As age increases pres increases.
4. As age increases skin decreases.
5. As age increases test decreases.
6. As age increases BMI increases.
7. As age increases pedi increases.
8. As age increases age increases.

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 μ U/mL)	0.198
6	BMI (in kg/m^2)	1.000
7	pedi	0.141
8	Age (in years)	0.036

Inferences:

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

1. As BMI increases pregs increases.
2. As BMI increases plas increases.
3. As BMI increases pres increases.
4. As BMI increases skin increases.
5. As BMI increases test increases.
6. As BMI increases BMI increases.
7. As BMI increases pedi increases.
8. As BMI increases age increases.

4 a.

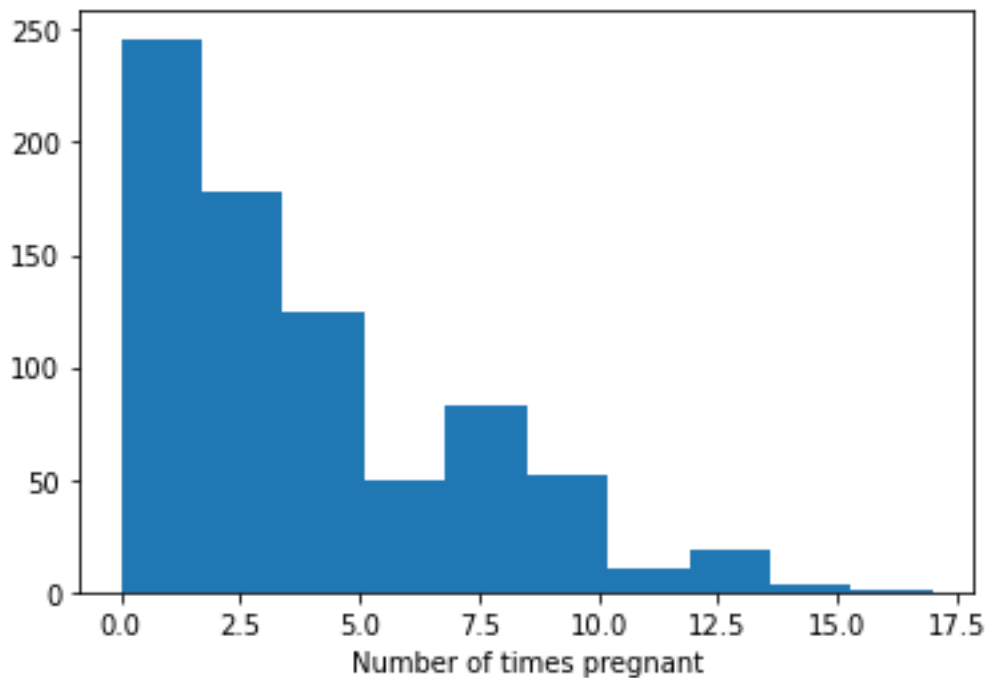


Figure 15 Histogram depiction of attribute pregs

Inferences:

1. Frequency from 0-1.67=240
2. Frequency from 1.67-3.34=175
3. Frequency from 3.34-5=125
4. Frequency from 5-6.68=50

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

5. $\text{Mode} = (1.67 - 0) / 2 = 0.835$

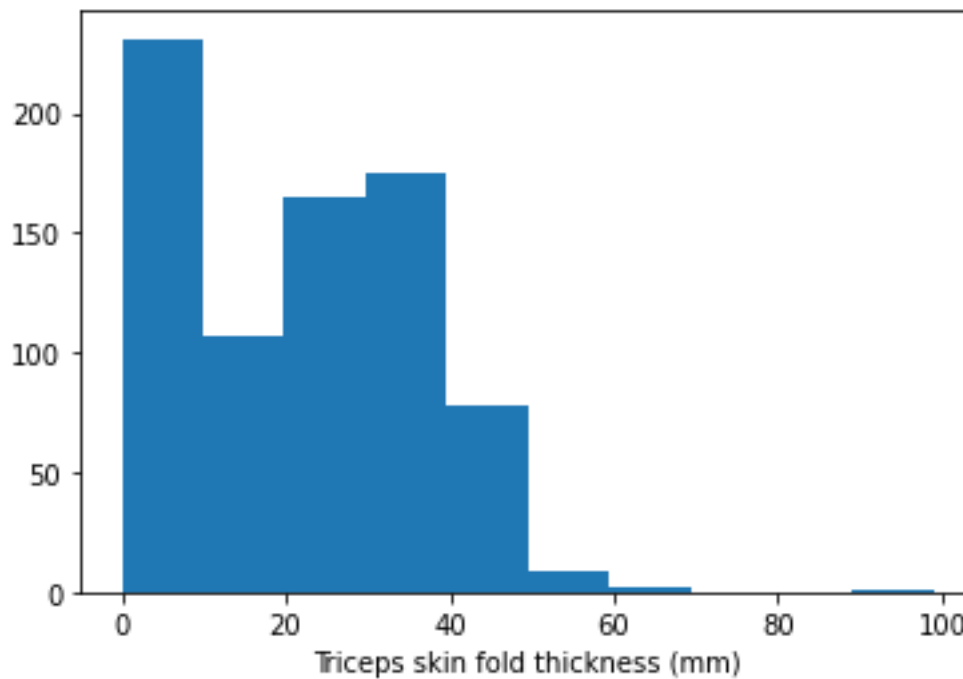


Figure 16 Histogram depiction of attribute skin

Inferences:

1. Frequency from 0-10=250
2. Frequency from 10-20=100
3. Frequency from 20-30=160
4. Frequency from 30-40=170
5. $\text{Mode} = (10 - 0) / 2 = 5$

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

5

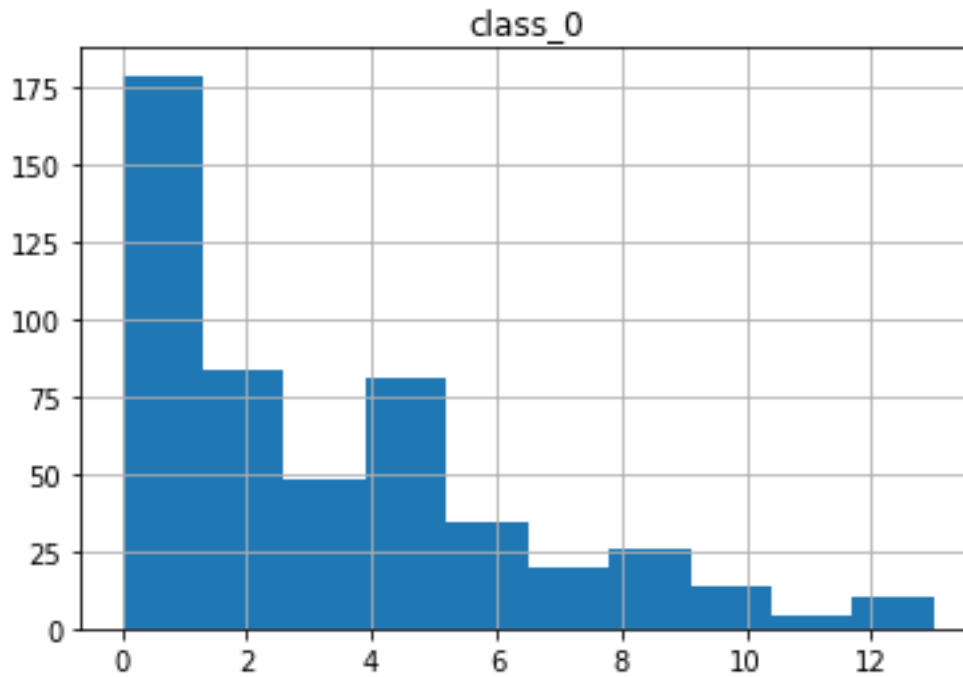


Figure 17 Histogram depiction of attribute pregs for class 0

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

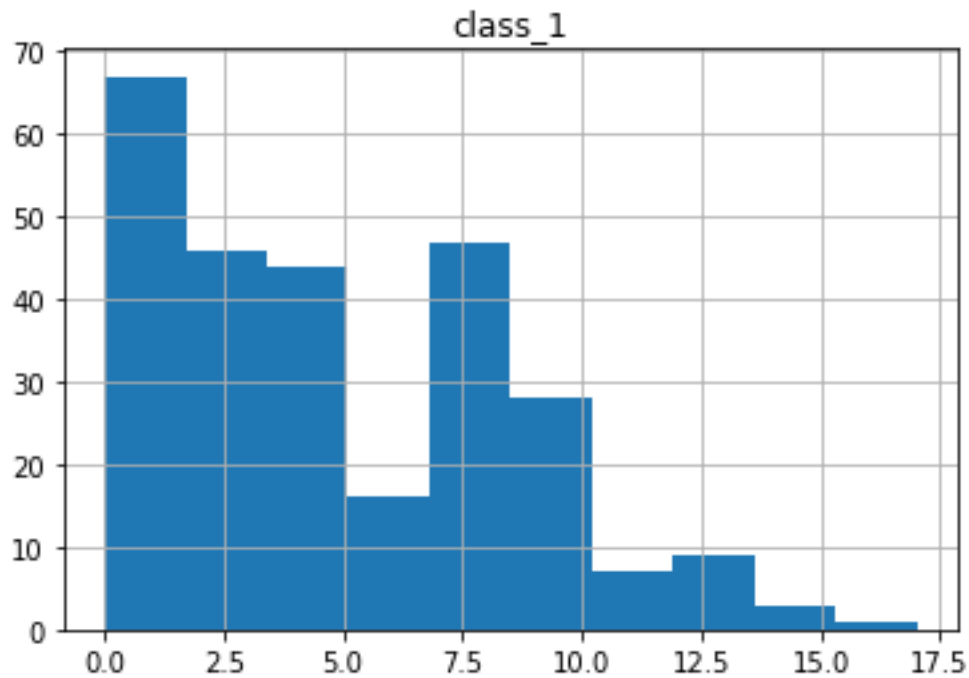


Figure 18 Histogram depiction of attribute pregs for class 1

Inferences:

1. Mode in class_0= $(1.33-0)/2=0.66$ and in class_1= $(1.67-0)/2=0.83$.
2. Frequency sharply decreases in class_0 and in case of class_1 it decreases then increases.

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

6

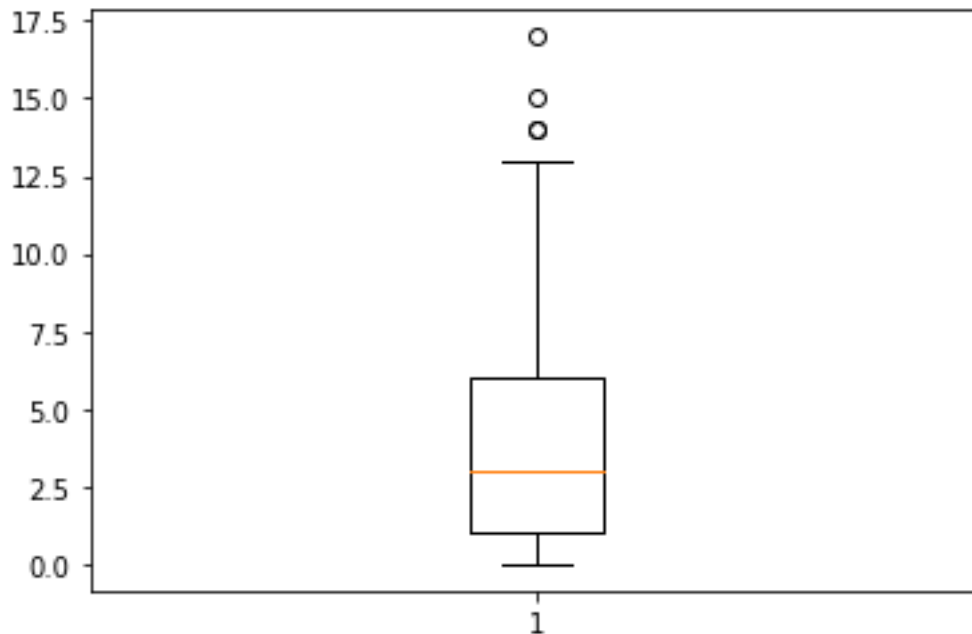


Figure 19 Boxplot for attribute pregs

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.

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

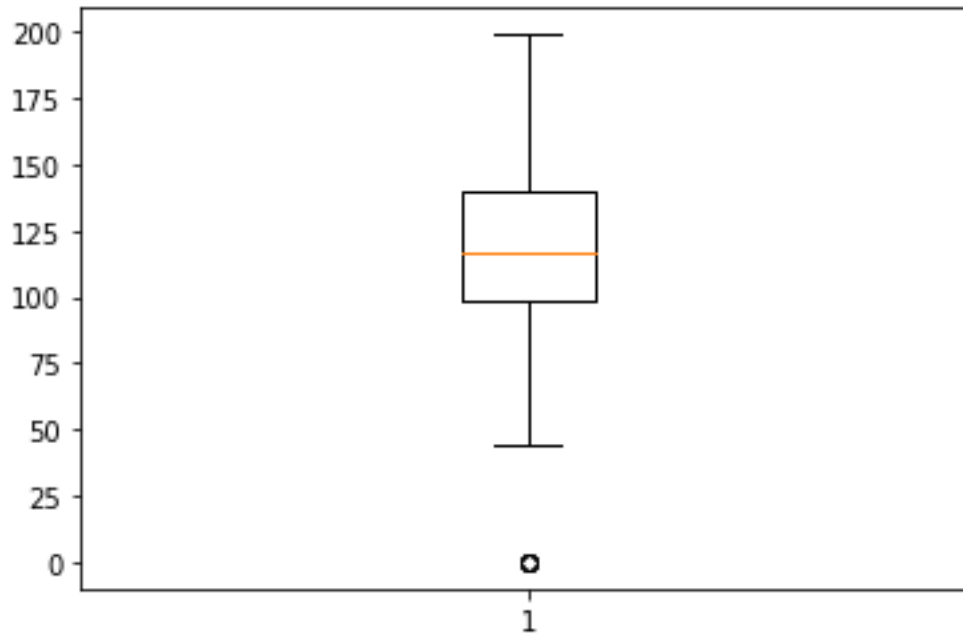


Figure 20 Boxplot for attribute plas

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.

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

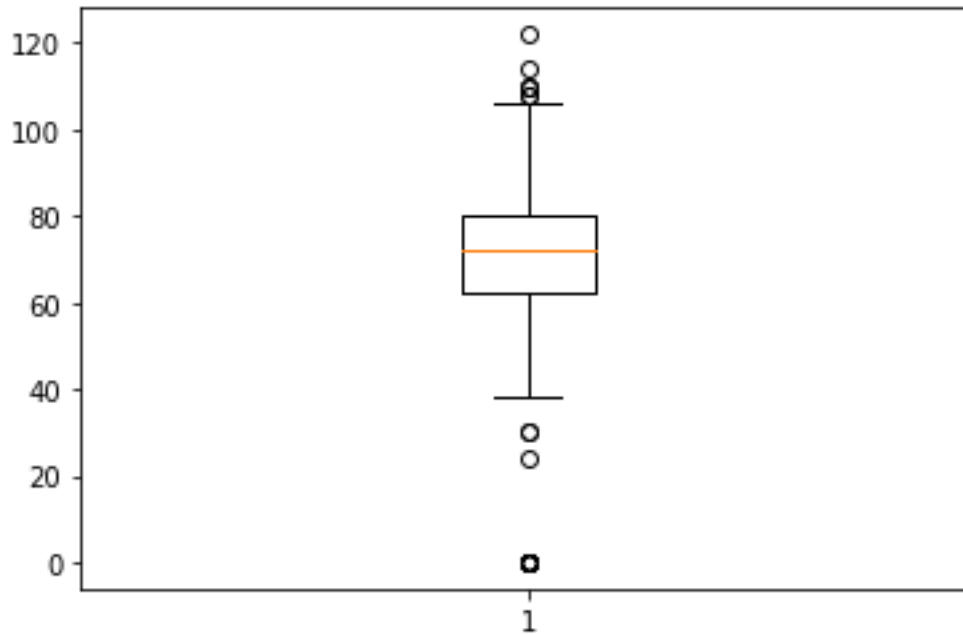


Figure 21 Boxplot for attribute pres(in mm Hg)

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.

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

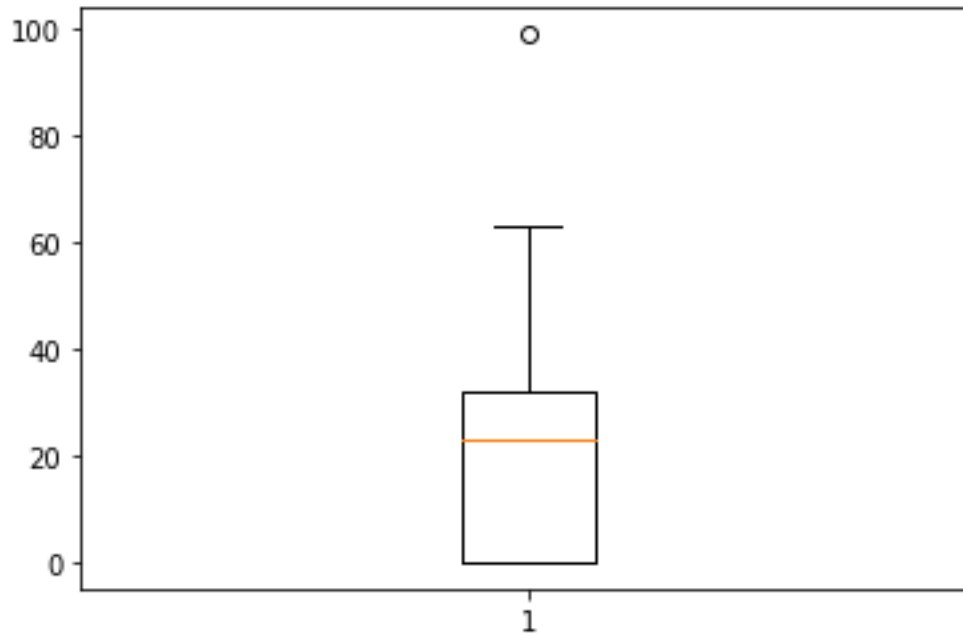


Figure 22 Boxplot for attribute skin(in mm)

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.

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

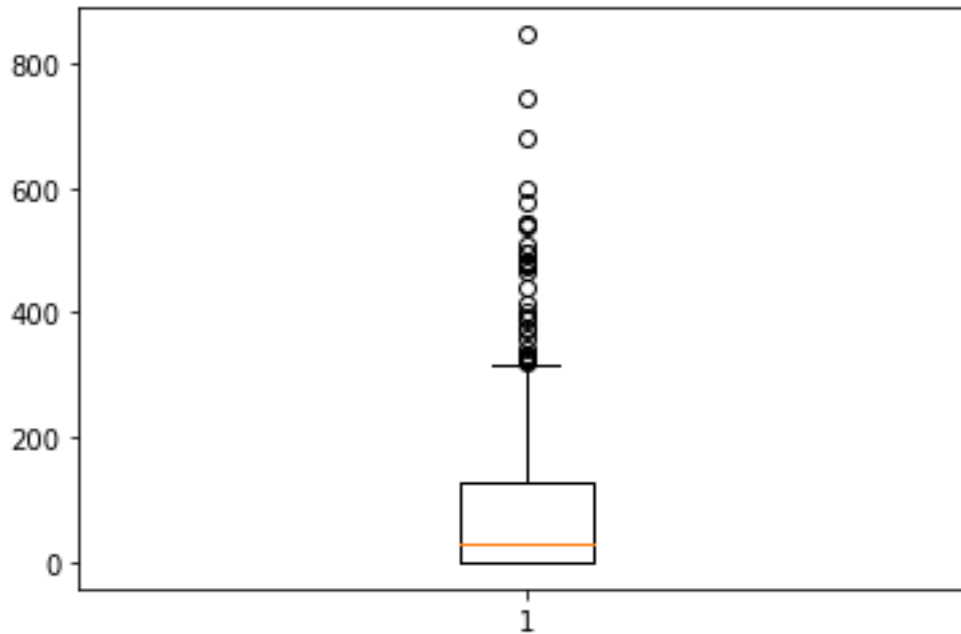


Figure 23 Boxplot for attribute test (mu U/mL)

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.

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

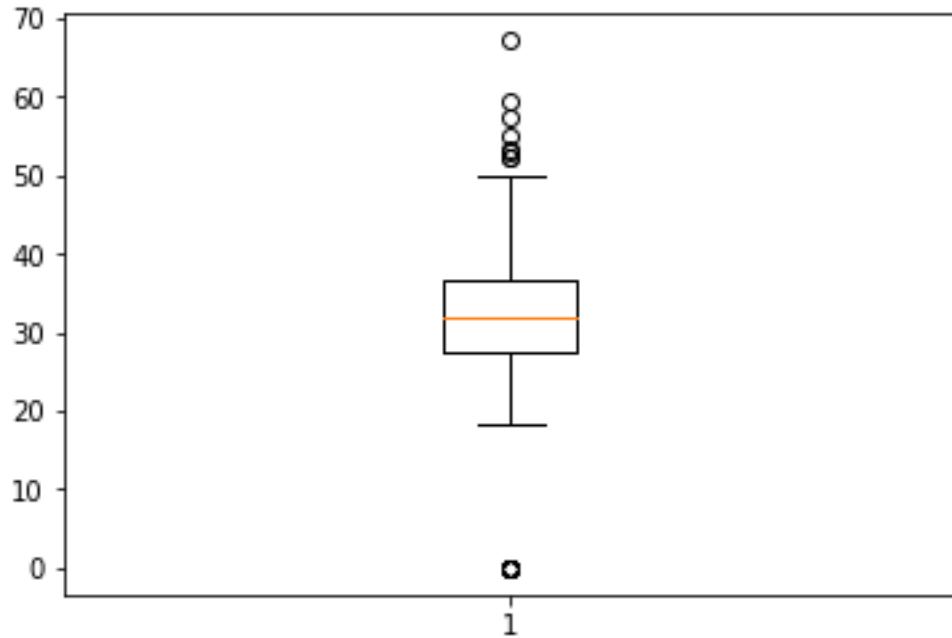


Figure 24 Boxplot for attribute BMI (in kg/m²)

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.

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

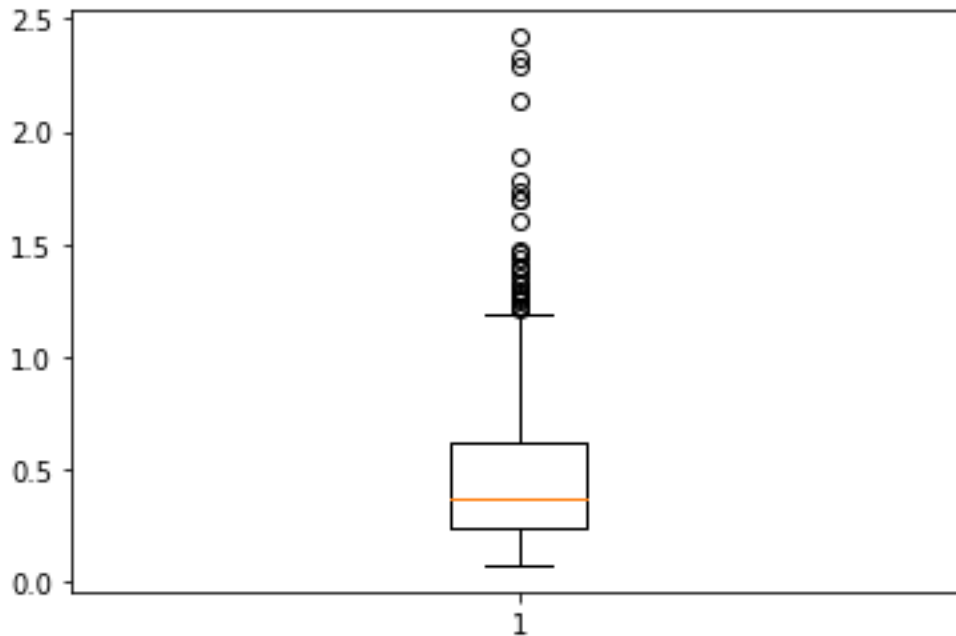


Figure 25 Boxplot for attribute pedi

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.

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

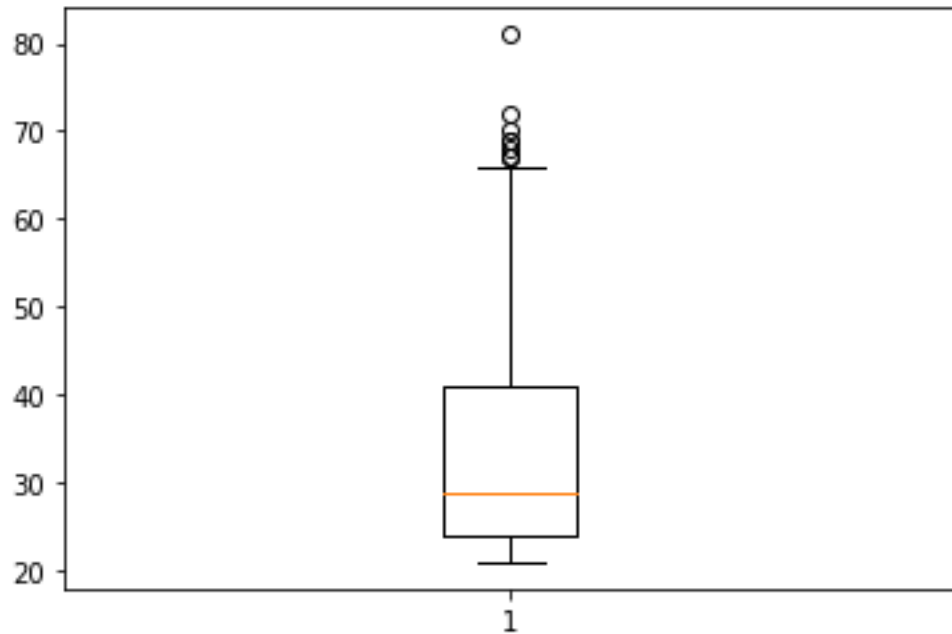


Figure 26 Boxplot for attribute Age (in years)

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