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

Student's Name: Aryan Tiwari **Mobile No: 8982562898**

Roll Number: B20187 **Branch:** Electrical 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	3.37
2	plas	120.89	117.0	99, 100	0	199	31.97
3	pres (in mm Hg)	69.105	72	70	0	122	19.35
4	skin (in mm)	20.53	23	0	0	99	15.95
5	test (in mu U/mL)	79.79	30.5	0	0	846	115.244
6	BMI (in kg/m^2)	31.99	32	32	0	67.1	7.88
7	pedi	0.47	0.37	0.254, 0.258	0.078	2.42	0.33
8	Age (in years)	33.24	29	22	21	81	11.76

Inferences:

1. Closer the values of mean, median and mode lower is the value of standard deviation and vice versa.
2. Standard deviation is the spread of distribution wrt mean.

2 a.

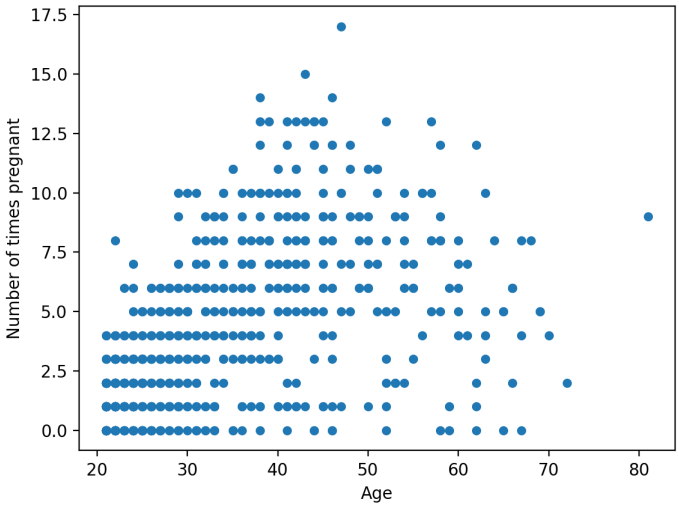

Figure 1 Scatter plot: Age (in years) vs. pregs

Inferences:

1. Number of times pregnant rises with the increase in age (positively correlated) and is highest among the mid value of 35-50 years.
2. This distribution is left skewed.

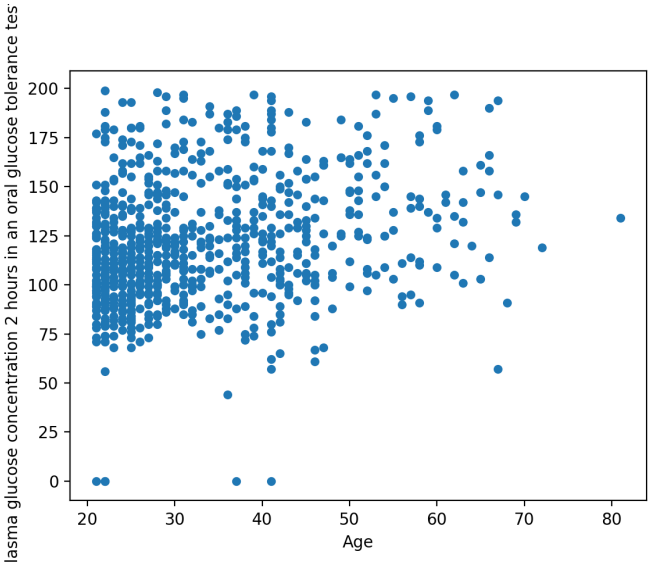

Figure 2 Scatter plot: Age (in years) vs. plas

Inferences:

1. Age is positively correlated to blood glucose level in glucose tolerance test.
2. The density of very high for the lower values of age and becomes less dense with increasing age.

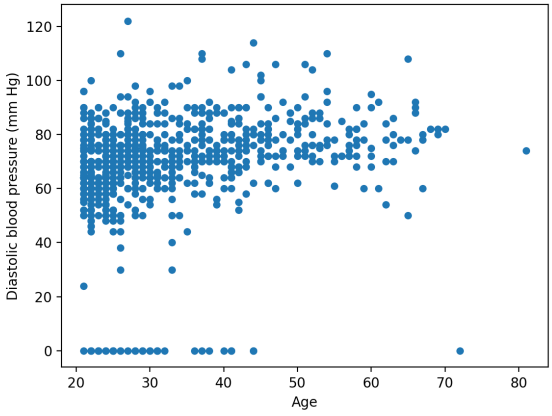

Figure 3 Scatter plot: Age (in years) vs. pres (in mm Hg)

Inferences:

1. There is a positive correlation between age and blood pressure
2. The density of points is very high for the lower values of age and becomes less dense with increasing age.

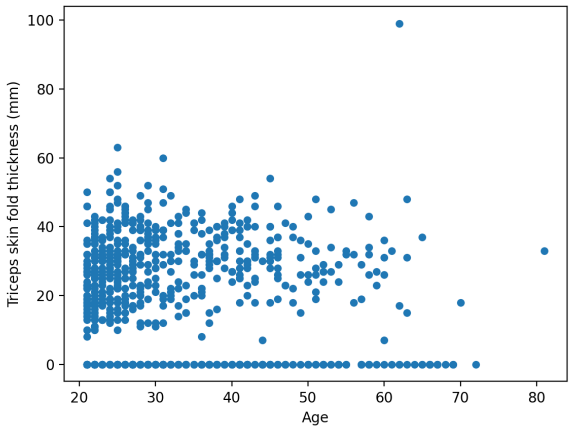

Figure 4 Scatter plot: Age (in years) vs. skin (in mm)

Inferences:

1. Thickness of skin fold decreases with increasing age.
2. The plot is dense towards the origin(lower age and lower thickness), which shows that thickness increases slightly an initial ages and then decreases.

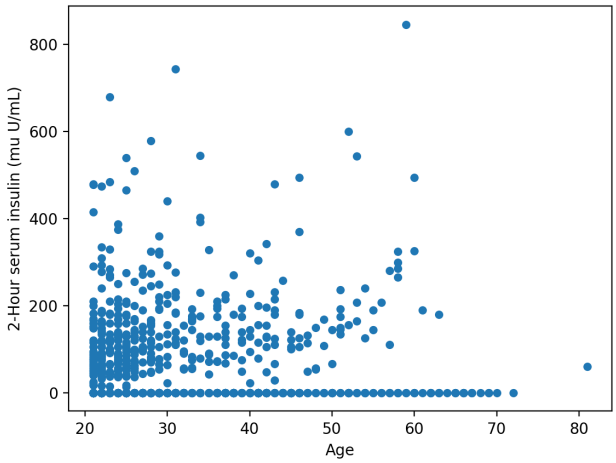

Figure 5 Scatter plot: Age (in years) vs. test (in mm U/mL)

Inferences:

1. The plot of insulin levels is almost un-corelated to age.
2. The density is very high at origin showing lower levels of insulin at younger ages.

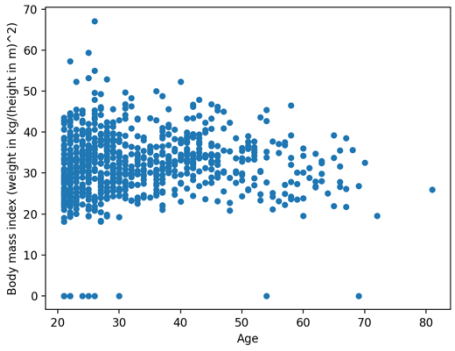

Figure 6 Scatter plot: Age (in years) vs. BMI (in kg/m²)

Inferences:

1. The BMI is almost un-corelated to age.
2. The plot is dense at mid values of bmi for younger ages.

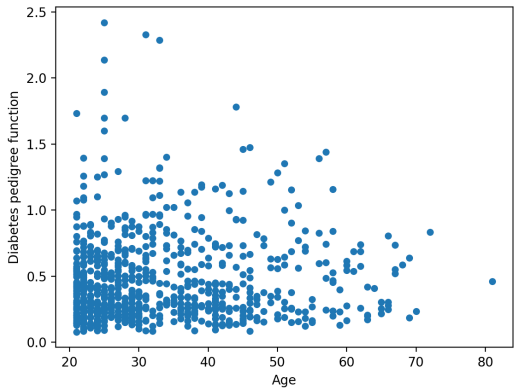

Figure 7 Scatter plot: Age (in years) vs. pedi

Inferences:

1. The pedigree function of diabetes increases slightly with increasing age
2. Dense at origin, showing very low pedigree function at younger age.

b.

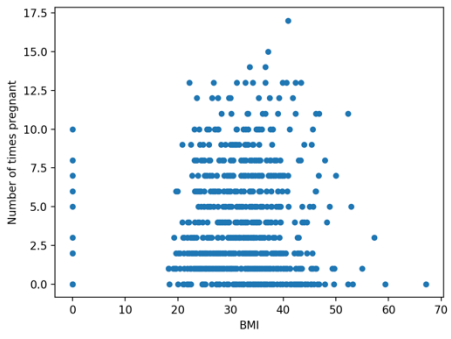

Figure 8 Scatter plot: BMI (in kg/m^2) vs. pregs

Inferences:

1. BMI is uncorrelated to pregs.
2. The plot is evenly scattered.

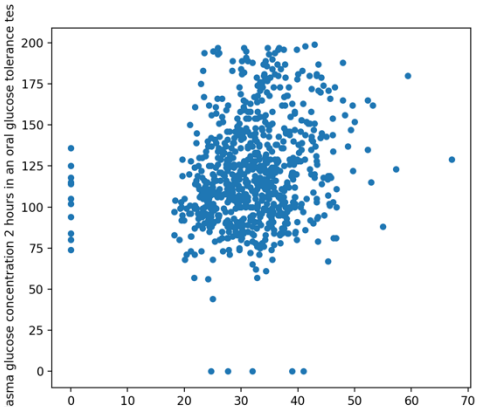

Figure 9 Scatter plot: BMI (in kg/m²) vs. plas

Inferences:

1. Plasma glucose concentration increases with increasing BMI.
2. The plot is very dense at intermediate BMI and medium glucose levels.

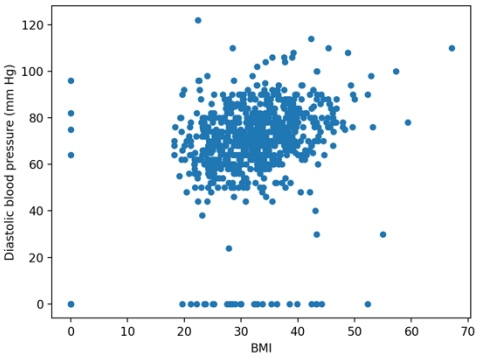

Figure 10 Scatter plot: BMI (in kg/m²) vs. pres (in mm Hg)

Inferences:

1. BP increases with increasing BMI.
2. Highly dense in mid BMI values and mid BP, showing mid BMI has optimum BP.

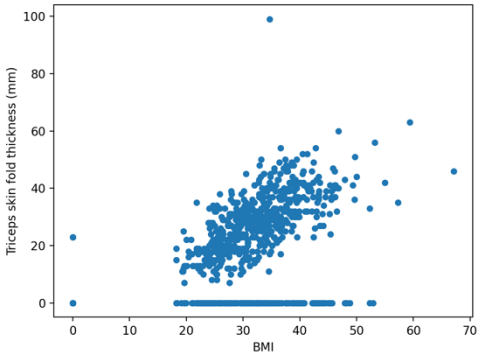

Figure 11 Scatter plot: BMI (in kg/m^2) vs. skin (in mm)

Inferences:

1. Thickness in skin fold increases with higher bmi. As body fat increases.
2. Gradual density.

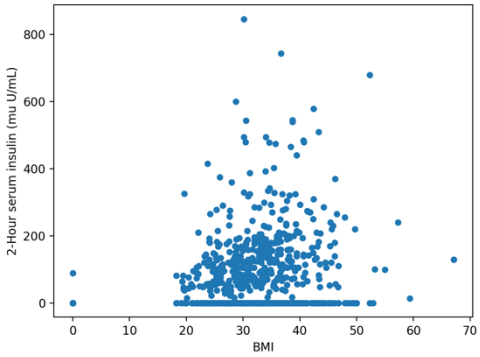

Figure 12 Scatter plot: BMI (in kg/m^2) vs. test (in mm U/mL)

Inferences:

1. Insulin levels are higher for higher bmi.
2. Dense at low values of insulin showing that insulin in low in most individuals.

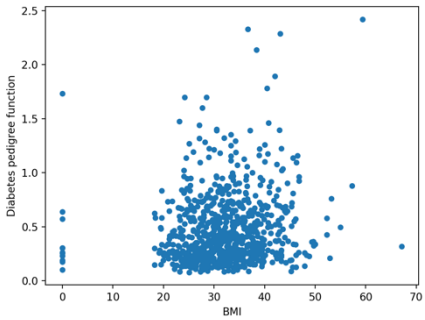

Figure 13 Scatter plot: BMI (in kg/m²) vs. pedi

Inferences:

1. Pedigree function increases with BMI above 35.
2. Lower pedigree function of diabetes for mid BMI.

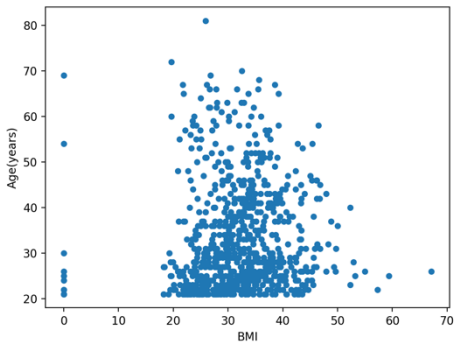

Figure 14 Scatter plot: BMI (in kg/m²) vs. Age (in years)

Inferences:

1. BMI and age are uncorrelated.
2. Mid BMI for younger individuals is dense. (Younger age, optimum BMI)

3 a.

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

S. No.	Attributes	Correlation Coefficient Value
1	pregs	0.554
2	plas	0.263
3	pres (in mm Hg)	0.239
4	skin (in mm)	-0.114
5	test (in mu U/mL)	-0.042
6	BMI (in kg/m^2)	0.036
7	pedi	0.033
8	Age (in years)	1

Inferences:

1. Pregs increase with higher rate with increasing age while plas and pres increase at moderate rate due to magnitude in correlation coefficient .
2. Preg, plas and pres increase with age and skin thickness and insulin conc. decrease with age. BMI, pedi are almost uncorrelated.

b.

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

S. No.	Attributes	Correlation Coefficient Value
1	pregs	0.017
2	plas	0.221
3	pres (in mm Hg)	0.281
4	skin (in mm)	0.392
5	test (in mu U/mL)	0.198
6	BMI (in kg/m^2)	1
7	pedi	0.146
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. Pregs and age are uncorrelated to BMI while all other attributes show slightly positive correlation to BMI.

4 a.

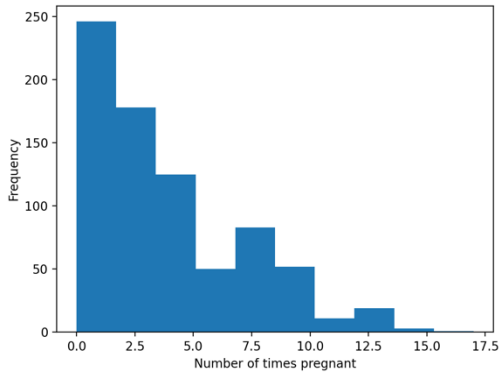

Figure 15 Histogram depiction of attribute pregs

Inferences:

1. Infer the frequency of each bin referring to its height.
2. Mode of pregs lies in 0 to 3.

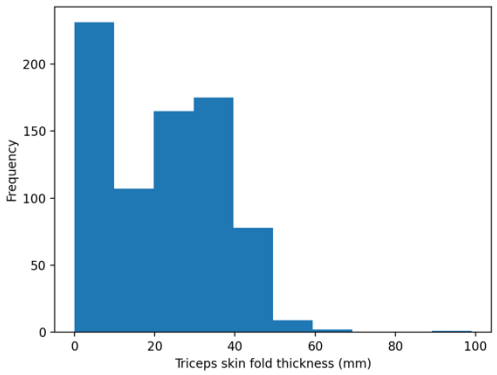

Figure 16 Histogram depiction of attribute skin

Inferences:

1. Highest frequency at start and high in middle with lowest in last.
2. Mode of skin thickness lies at 0 to 10.

5

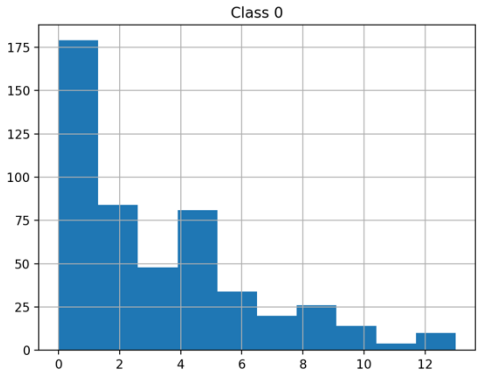

Figure 17 Histogram depiction of attribute pregs for class 0

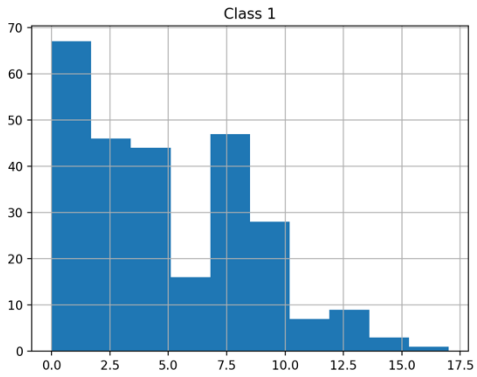

Figure 18 Histogram depiction of attribute pregs for class 1

Inferences:

1. In both classes mode lies in 0 to 2 pregs.

6

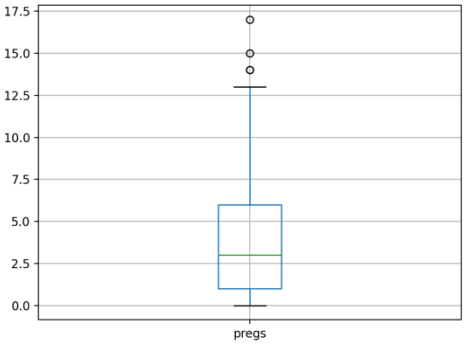

Figure 19 Boxplot for attribute pregs

Inferences:

1. Outliers lie above 12.5
2. Inter quartile range 1 to 6.
3. Left skewed.

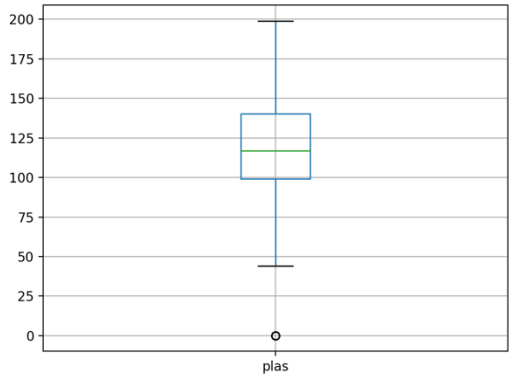

Figure 20 Boxplot for attribute plas

Inferences:

1. Outliers below 40.
2. Inter quartile range 100 to 140.
3. Mid skewed.

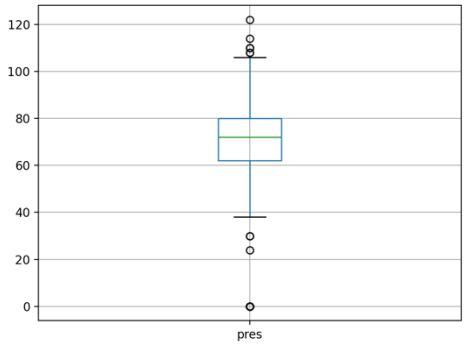

Figure 21 Boxplot for attribute pres(in mm Hg)

Inferences:

1. Outliers below 40 and above 100.
2. Inter quartile range 60 to 80.
3. Relate with the values from Q1. for this attribute.

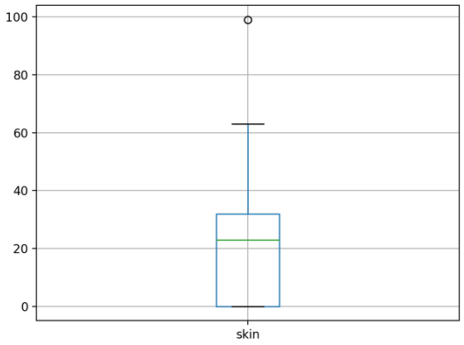

Figure 22 Boxplot for attribute skin(in mm)

Inferences:

1. Outliers lie above 60 and only one at 100.
2. Inter quartile range 0 to 35.
3. Left skewed.

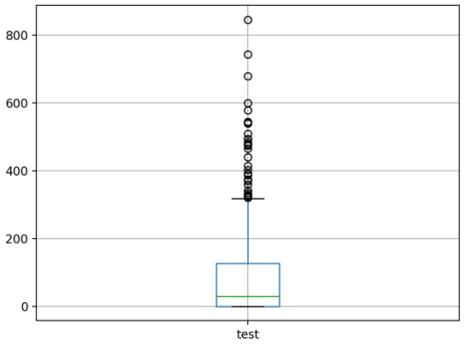

Figure 23 Boxplot for attribute test (mu U/mL)

Inferences:

1. Lots of outliers above 300 .

2. Inter quartile range 0 to 150.

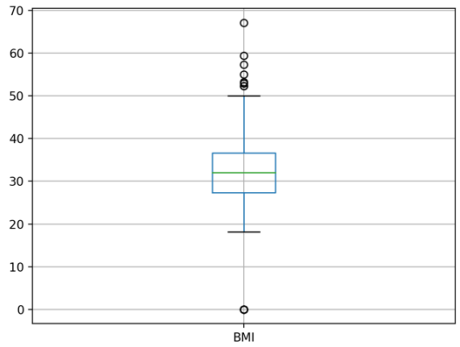

Figure 24 Boxplot for attribute BMI (in kg/m²)

Inferences:

1. Many outliers above 50 and one below 20 at 0.
2. Inter quartile range is 27 to 35.

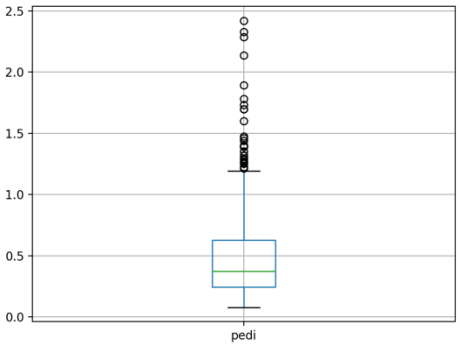

Figure 25 Boxplot for attribute pedi

Inferences:

1. Lots of outliers above 1.24.
2. Inter quartile range is 0.25 to 0.75

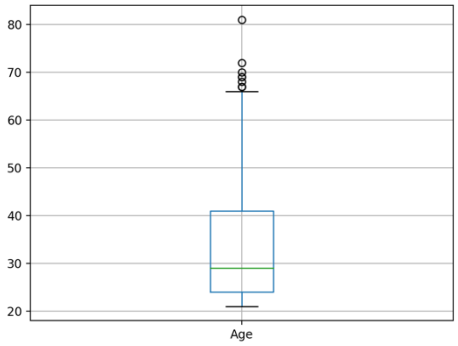

Figure 26 Boxplot for attribute Age (in years)

Inferences:

1. 6 outliers above 65 .
2. Inter quartile range is 25 to 41.