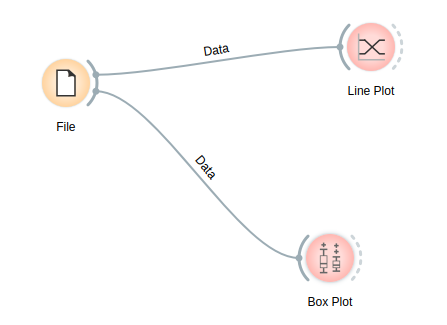
EXPERIMENT -4

**Aim:** Given a case study of predicting diabetics from a data set. You are expected to perform data visualization using Orange Tool & Python/R/Java.

Quote your observations after the visualization.

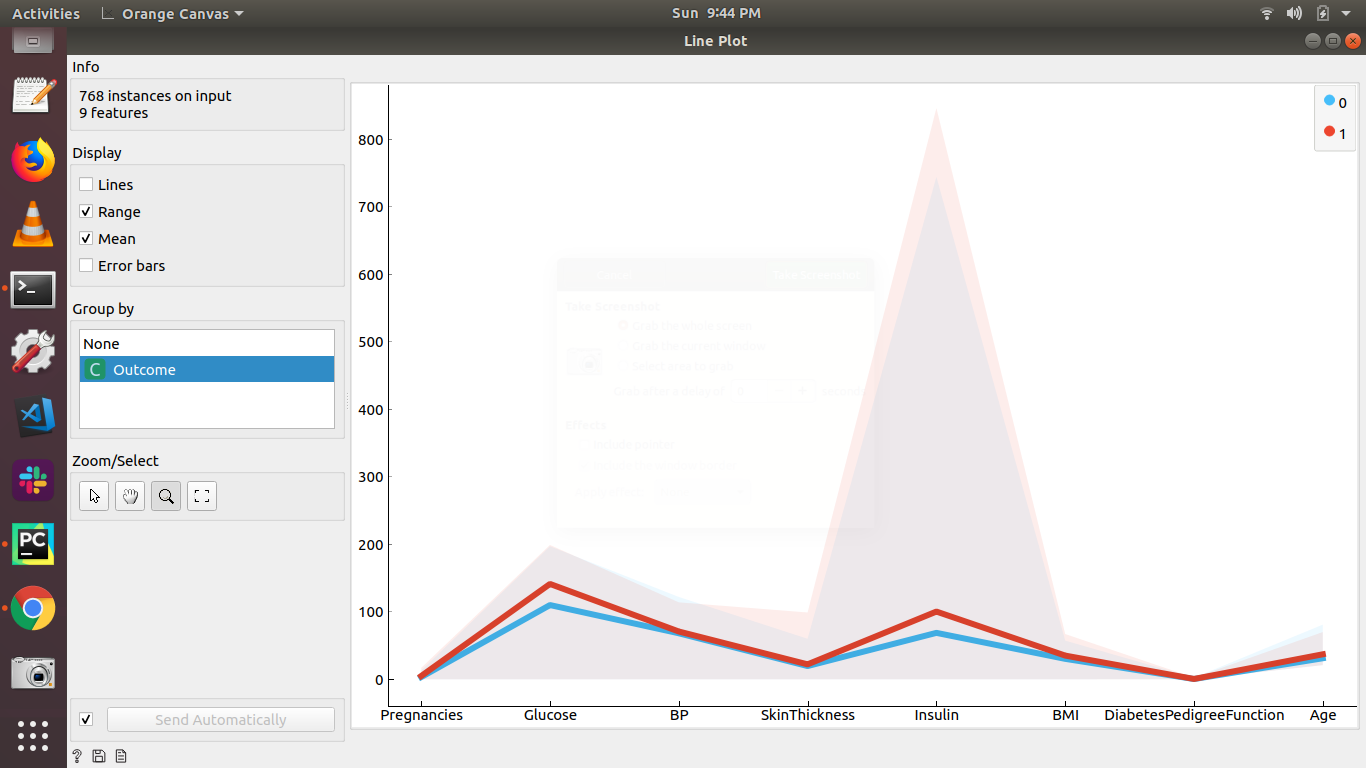
**Theory:** Data visualization is very important as it helps us to understand the data in a better format To see relationship between 2 variables we can use scatter plots. To see frequency of variables we can use histogram or bar plot. To see the changes in data we can use a line graph.

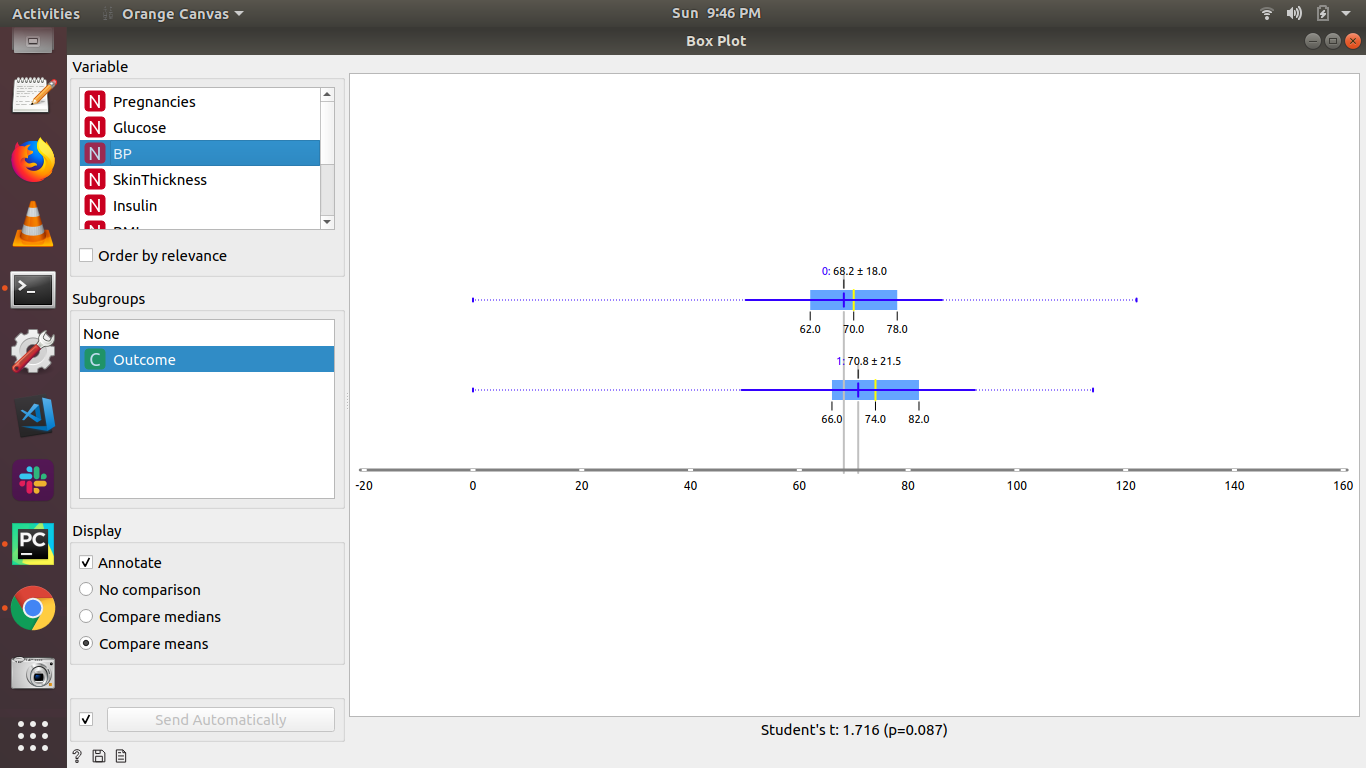
**Using Orange Tool O/p:**

****

We draw line plot to see the average value of all the numeric data . The line plots are drawn groupwise . We have a line plot for group1 and one for group 0.

We also draw a boxplot to see the variance in the bp data . That plot is also drawn group wise. We can see the outliers in this boxplot too.



****

**Using Python/R/Java:**

**import pandas as pd**

from sklearn.preprocessing import StandardScaler

from sklearn.decomposition import PCA

import numpy as np

# import matplotlib

# matplotlib.use('agg')

import seaborn as sns

import matplotlib.pyplot as plt

file = '/home/lavina/Desktop/pima-indians-diabetes-database/diabetes.csv'

# load data set into frame

dataFrame = pd.read\_csv(file, names=['Pregnancies', 'Glucose', 'BloodPressure', 'SkinThickness', 'Insulin', 'BMI', 'DiabetesPedigreeFunction', 'Age', 'Outcome'])

plt.scatter(x = dataFrame['Insulin'], y =dataFrame['BloodPressure'])

plt.xlabel('Insulin')

plt.ylabel('BloodPressure')

plt.show()

plt.hist(dataFrame['Pregnancies'])

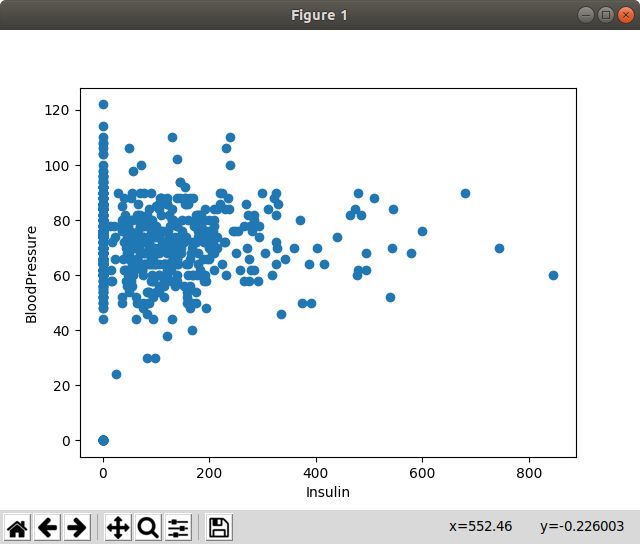
plt.xlabel('Pregnancies')

plt.ylabel('Frequency')

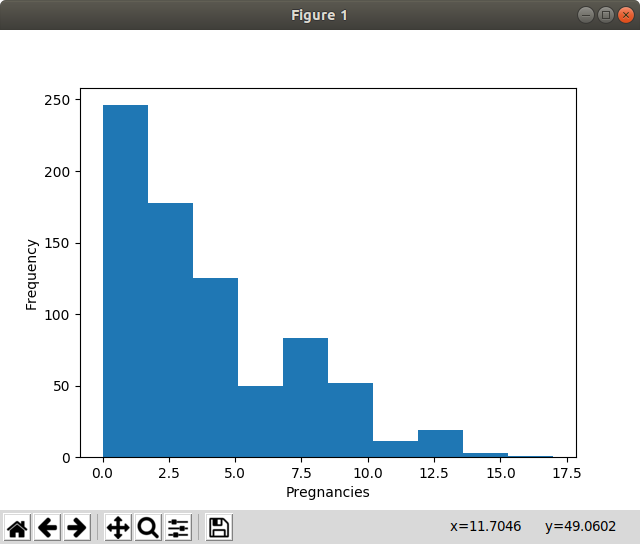
plt.show()

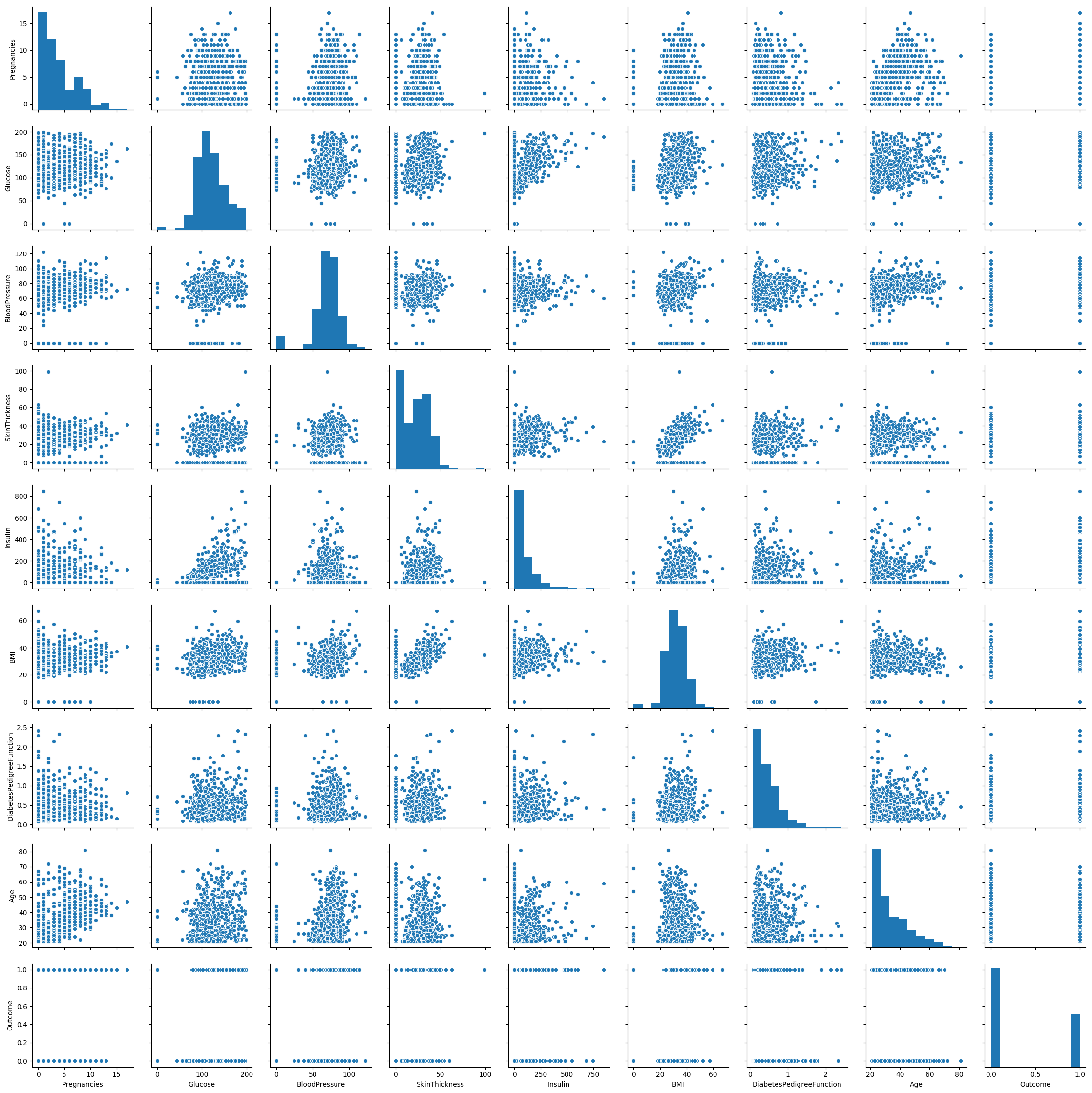
snsplot= sns.pairplot(dataFrame)

snsplot.savefig("output.png")

****

This plot shows us the covariance between insulin and blood pressure which show us they aren’t really related.

This shows us the frequency of the pregnancies . This tells us maximum people have been pregnant 0 to 2 times. And it keeps decreasing.

****

This shows us the covariance between all the features.

**Conclusion:** Hence we have successfully visualized data in orange and python.