Software Engineering Tools Lab

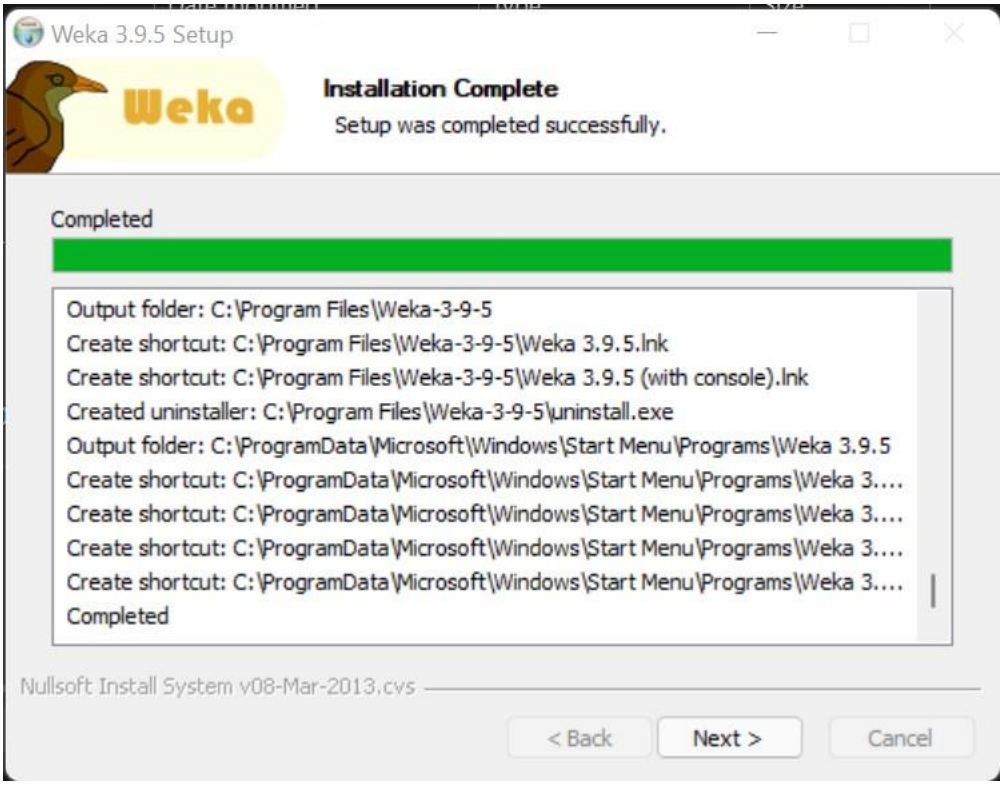
Assignment No-1

(Module 1- Introduction to OSS)

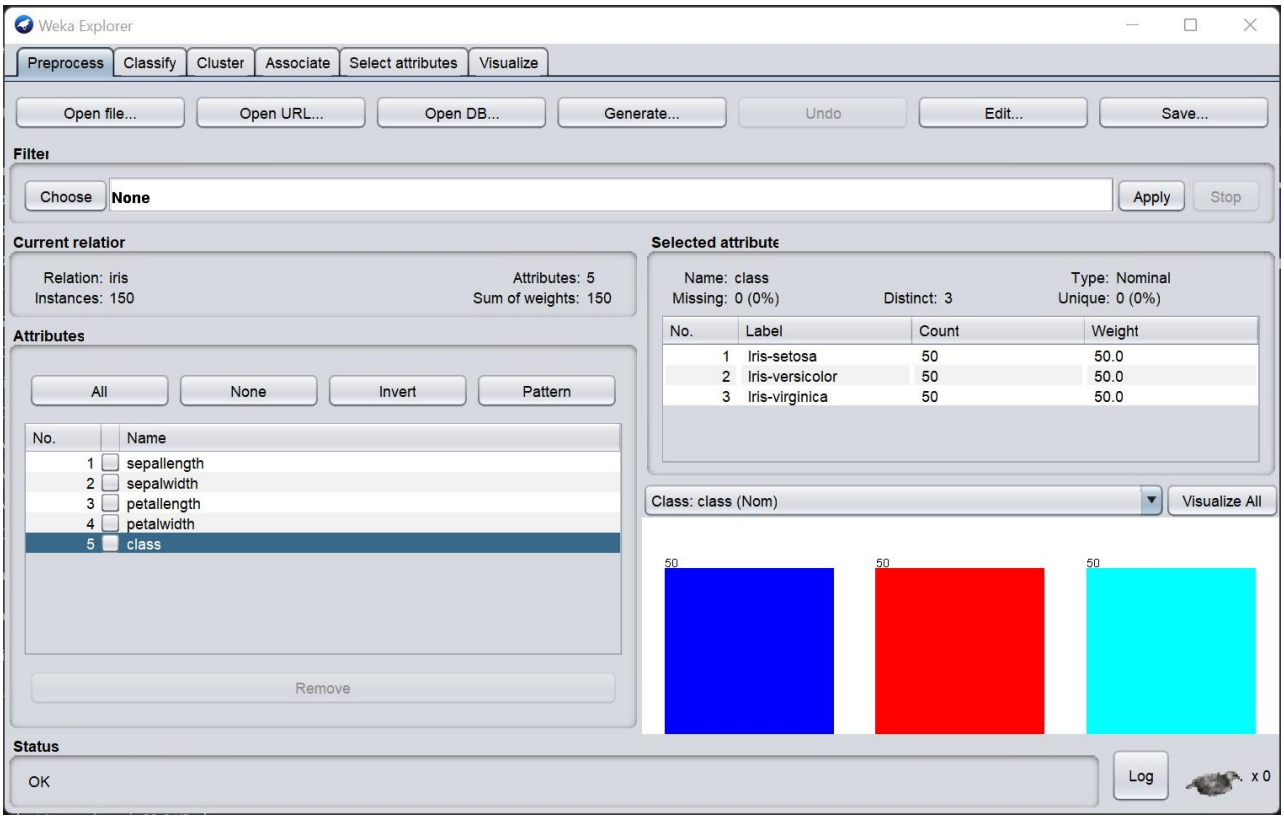
1.Weka is a GUI workbench that empowers data wranglers to assemble machine learning pipelines, train models, and run predictions without having to write code. Using Weka tool perform below tasks such as data pre-processing, data classification (use any appropriate ML algorithm) and data visualization efficiently on given dataset.
Use the Iris dataset given:
https://drive.google.com/file/d/1A3Fxsfzm6BSfhFZGDrjI47RTe45bSgYP/view

Ans:

First of all, we install the weka software on our machine.

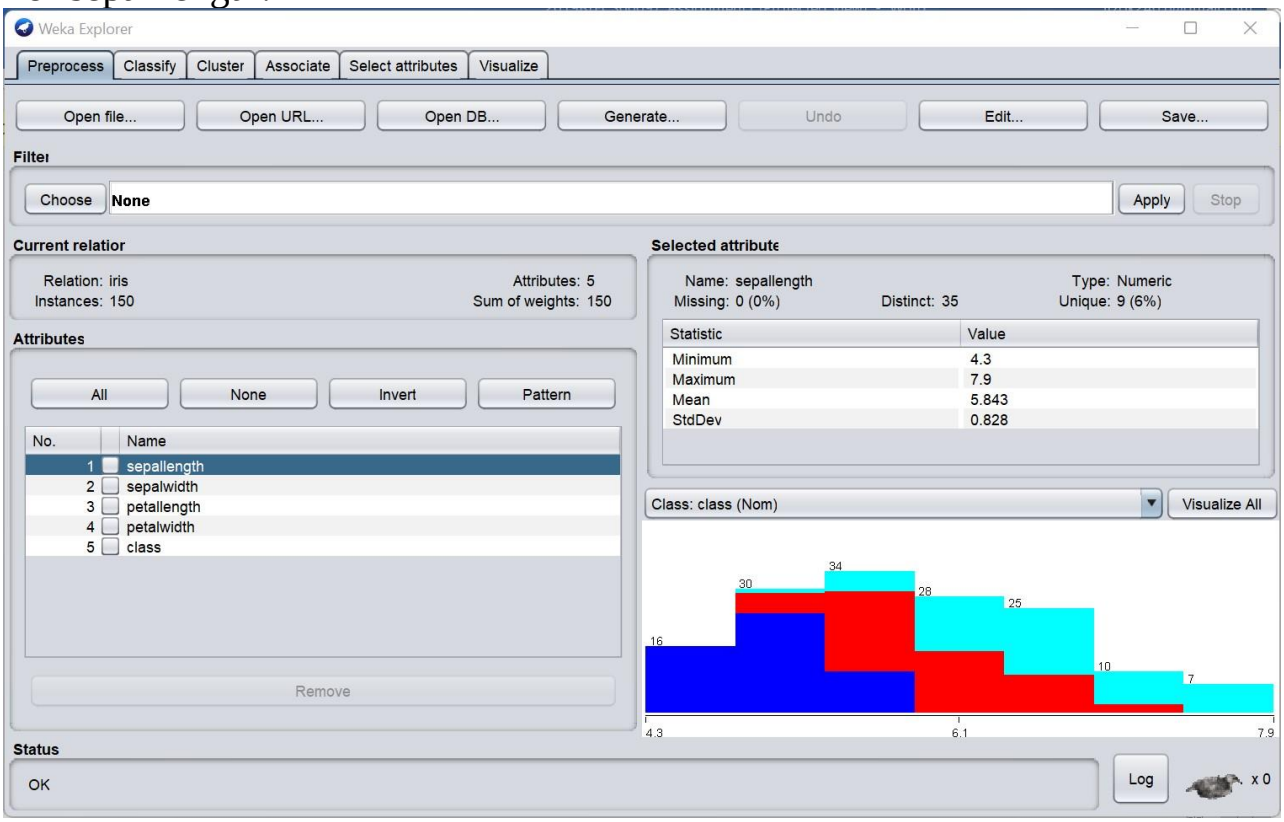


Now once this is done, we open the weka app and load our dataset. This comes under Pre-process tab. After loading, it looks like below image.

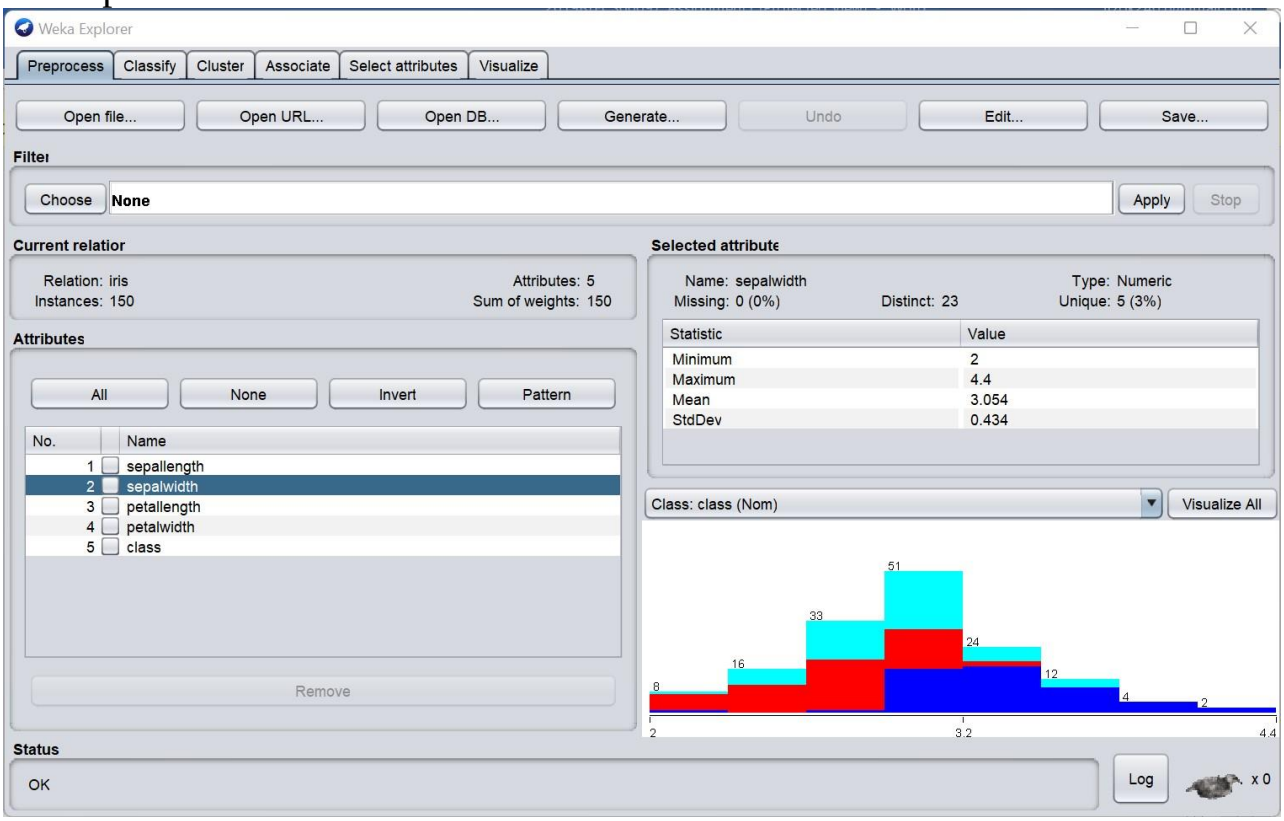


Under this tab, we can pre-process all our data like viewing and visualizing the data and the attributes including editing features. The five attributes here are sepal length, sepal width, petal length, petal width and class. All the information regarding the dataset can be viewed under this tab. Let's see the outcome for each attribute

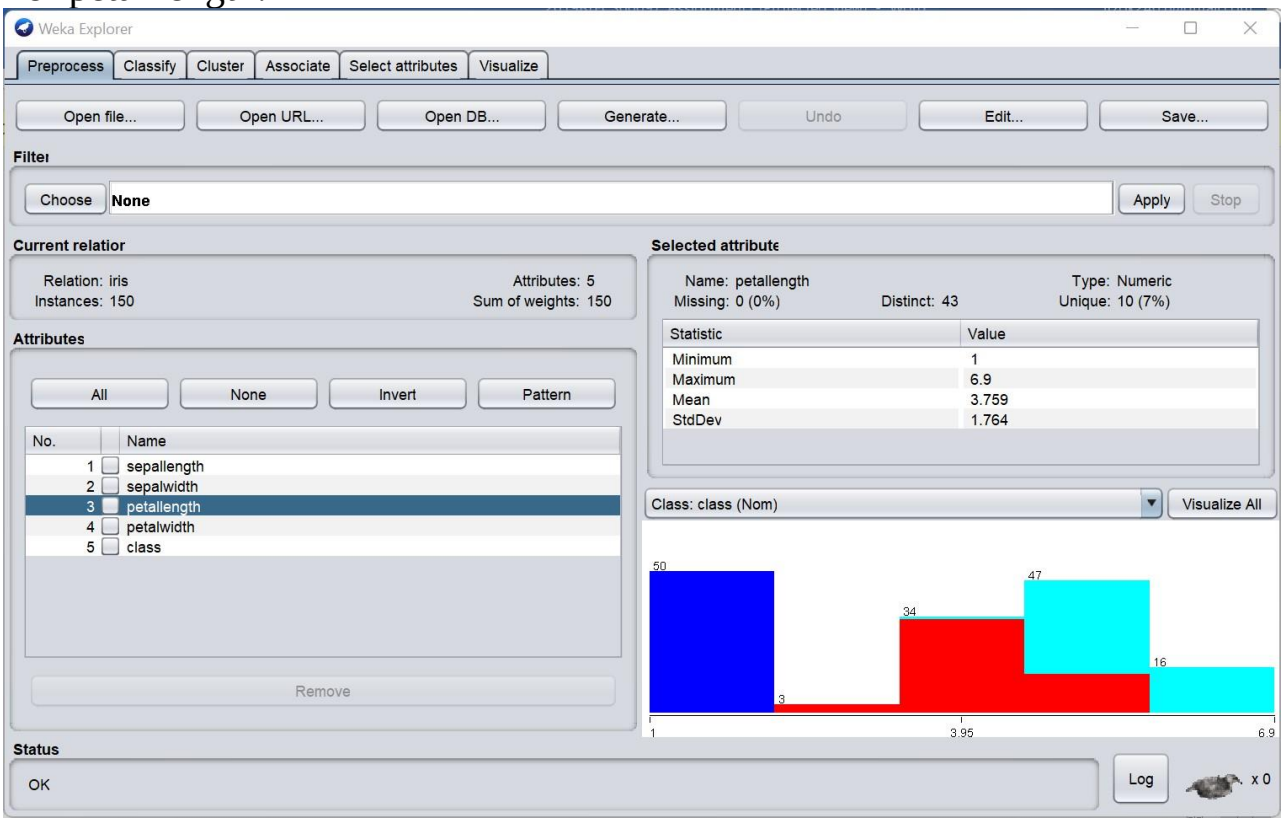
For sepal length:



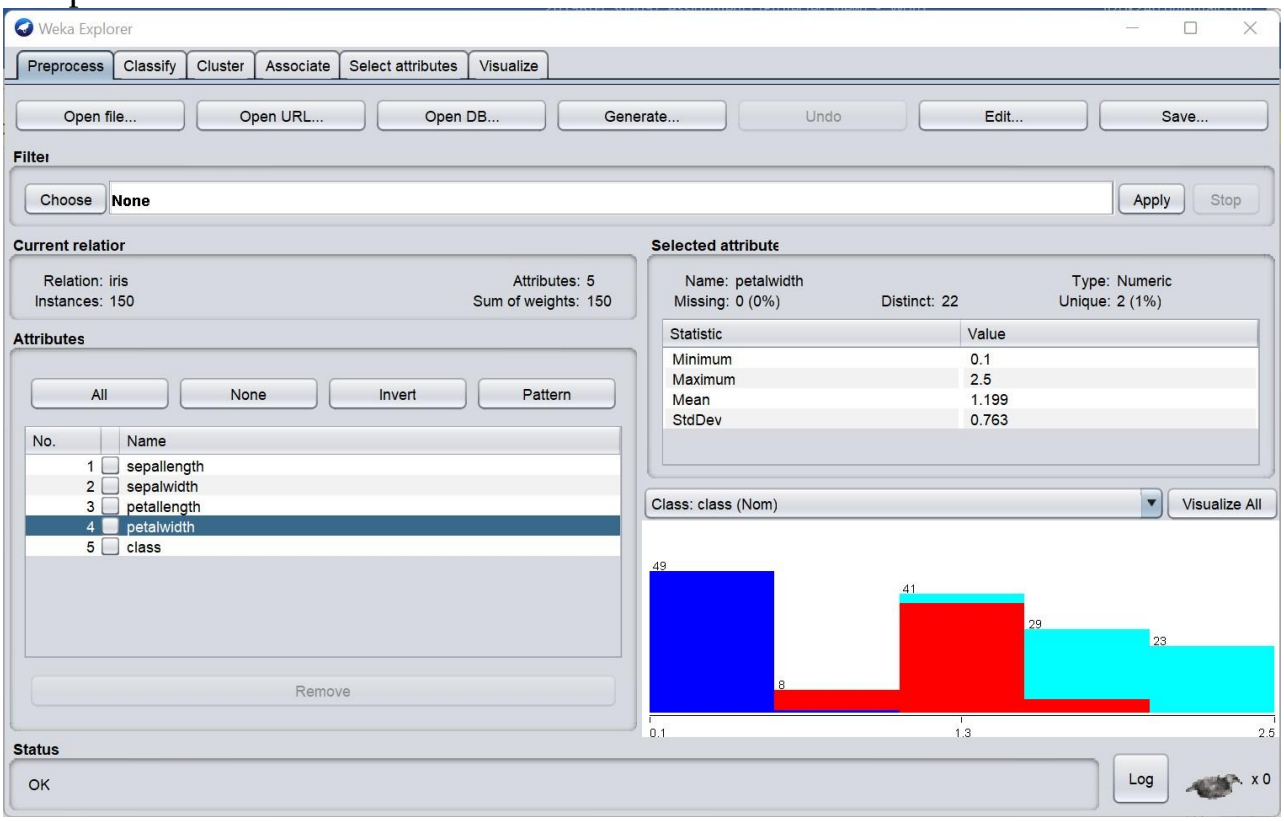
For sepal width:



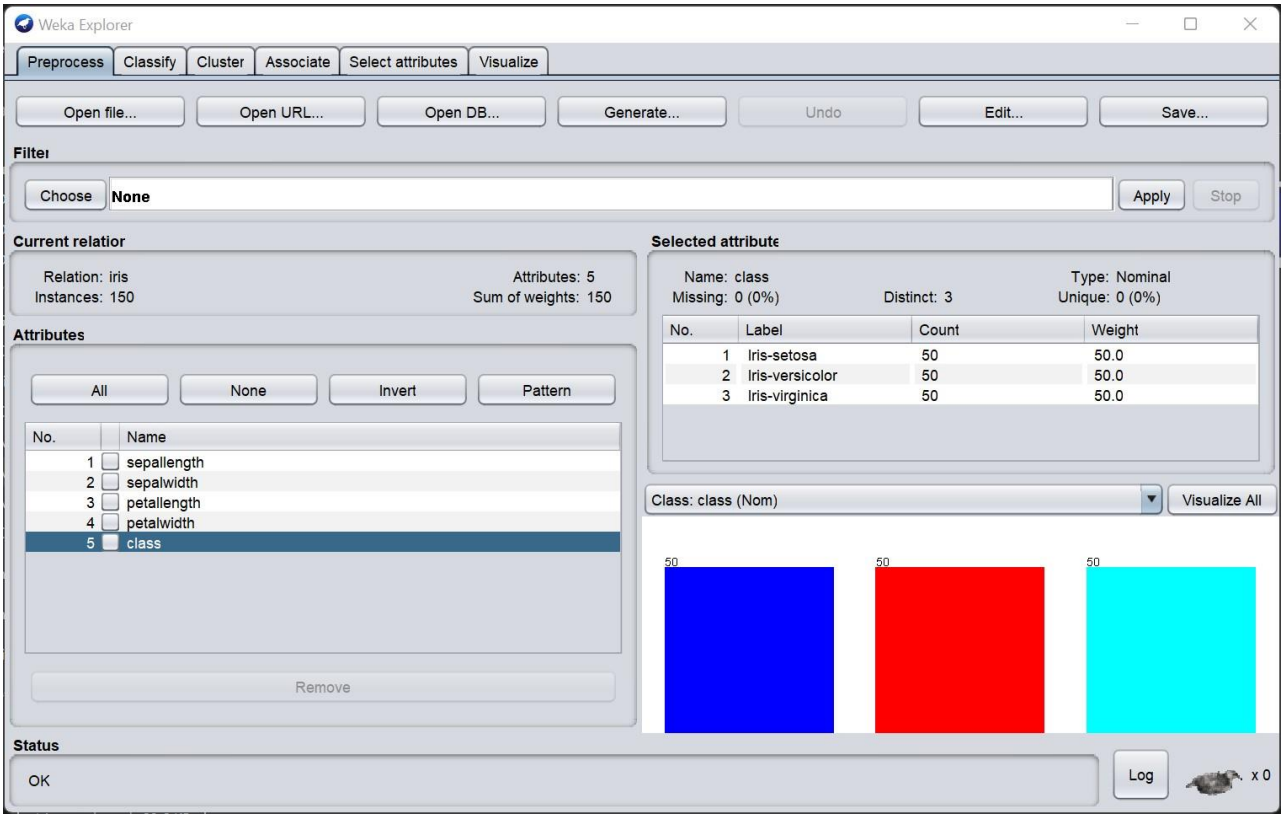
For petal length:



For petal width:

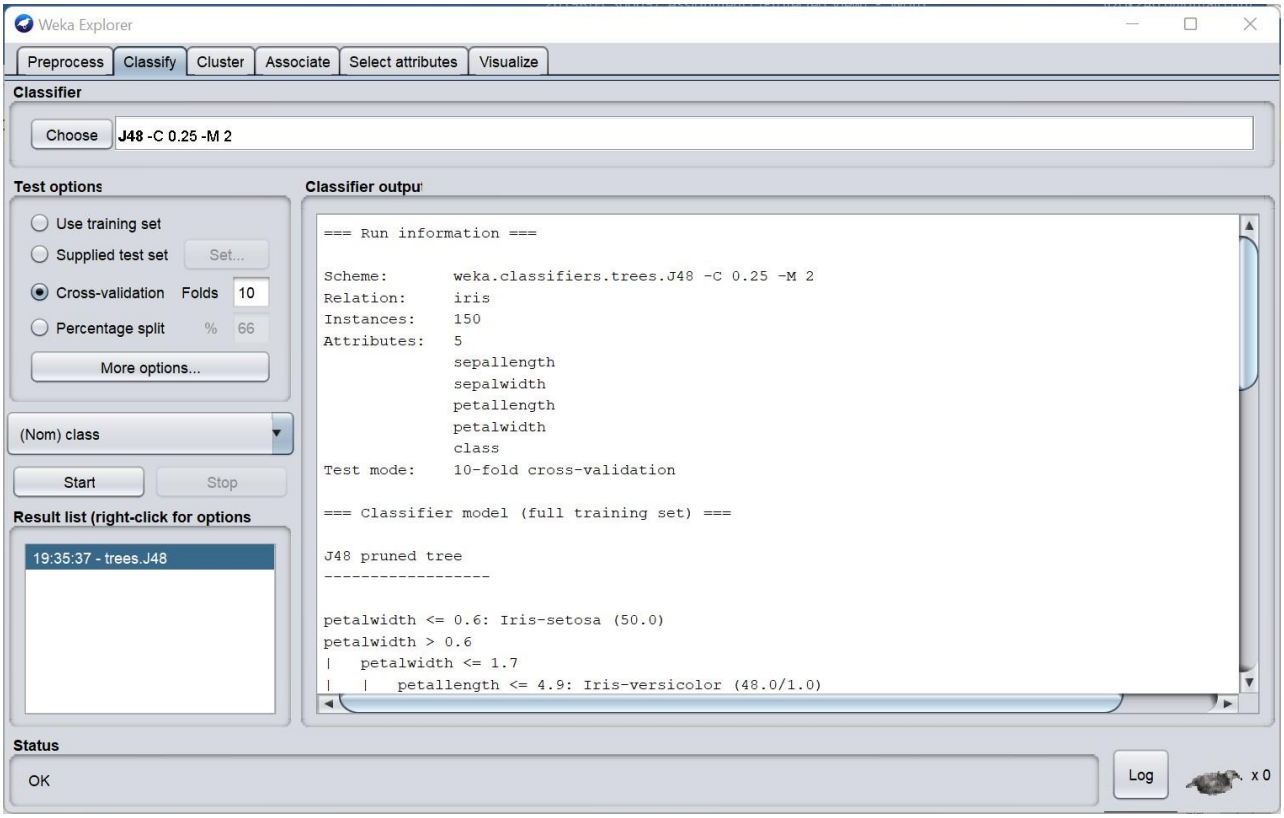


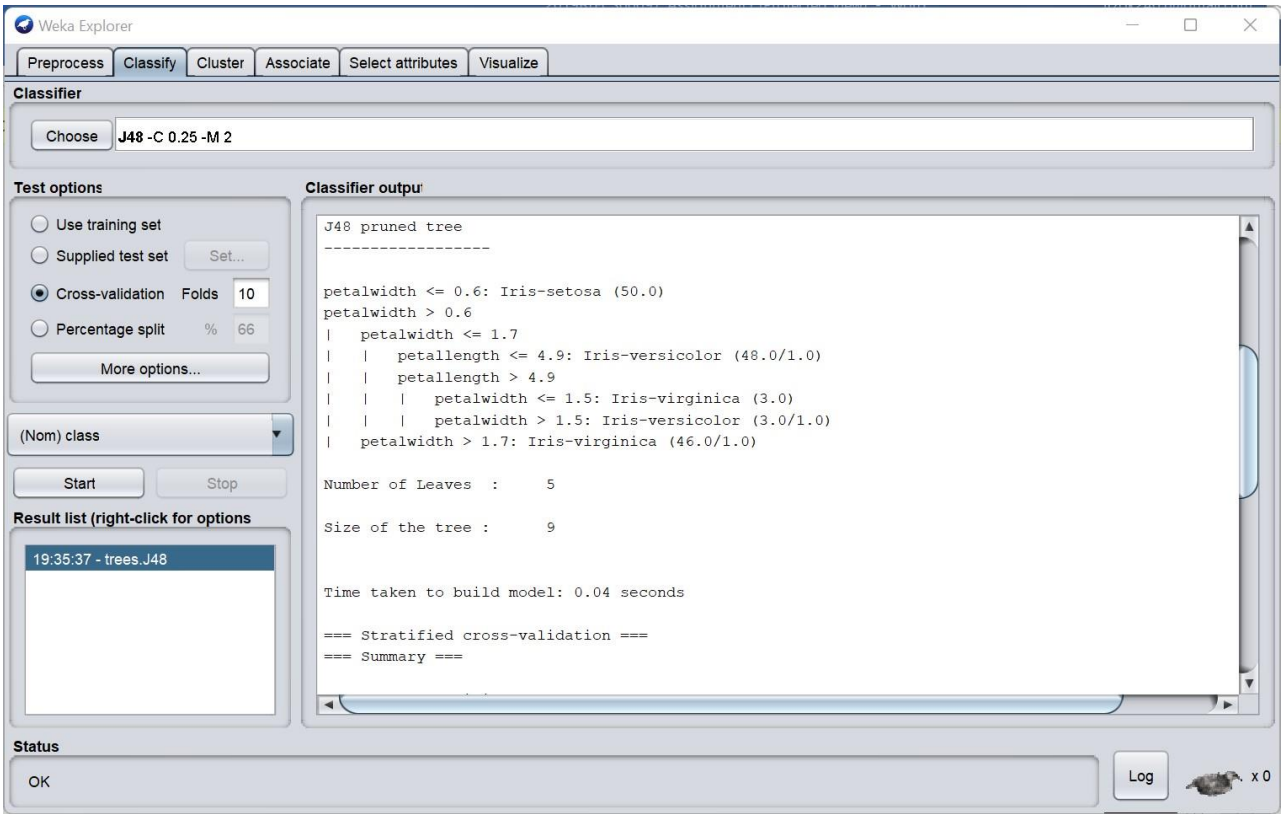
For class:

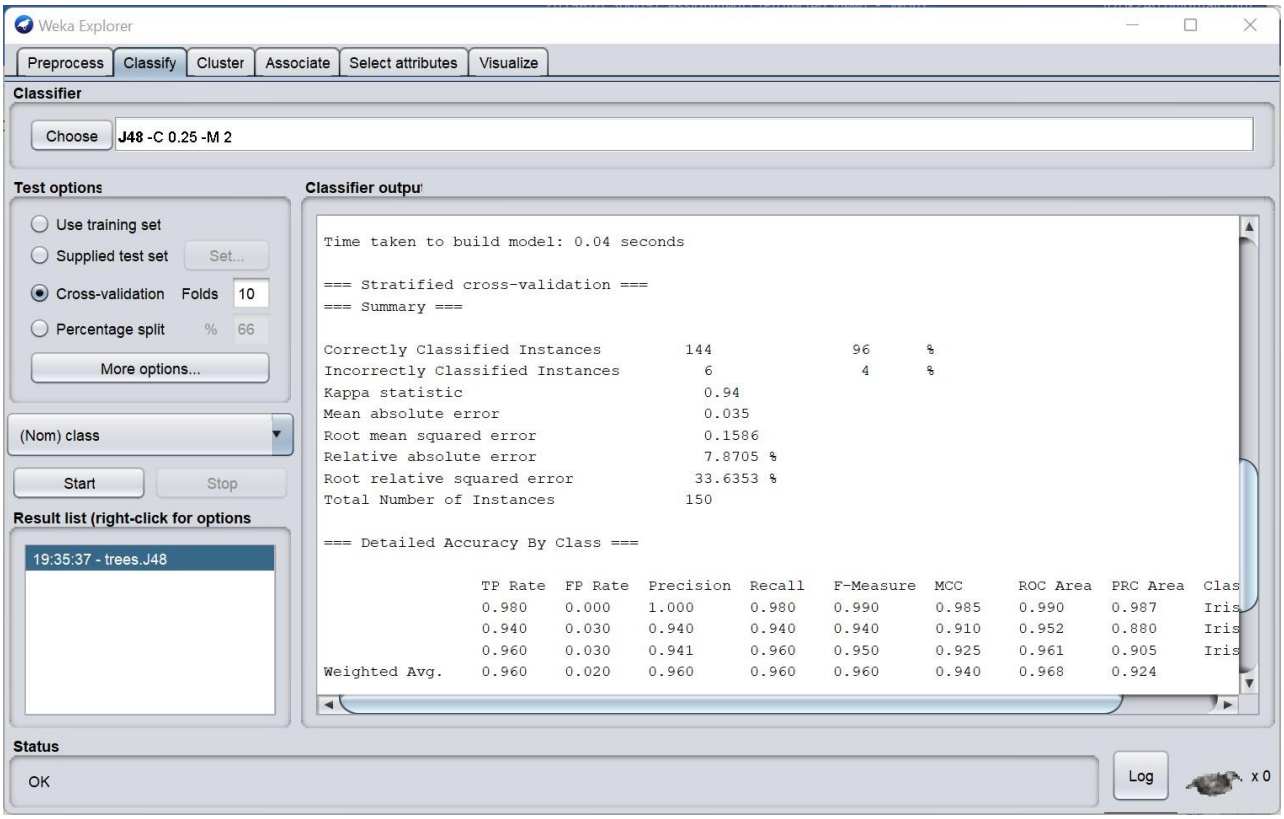


You can also view specifics of each attributes in this tab.

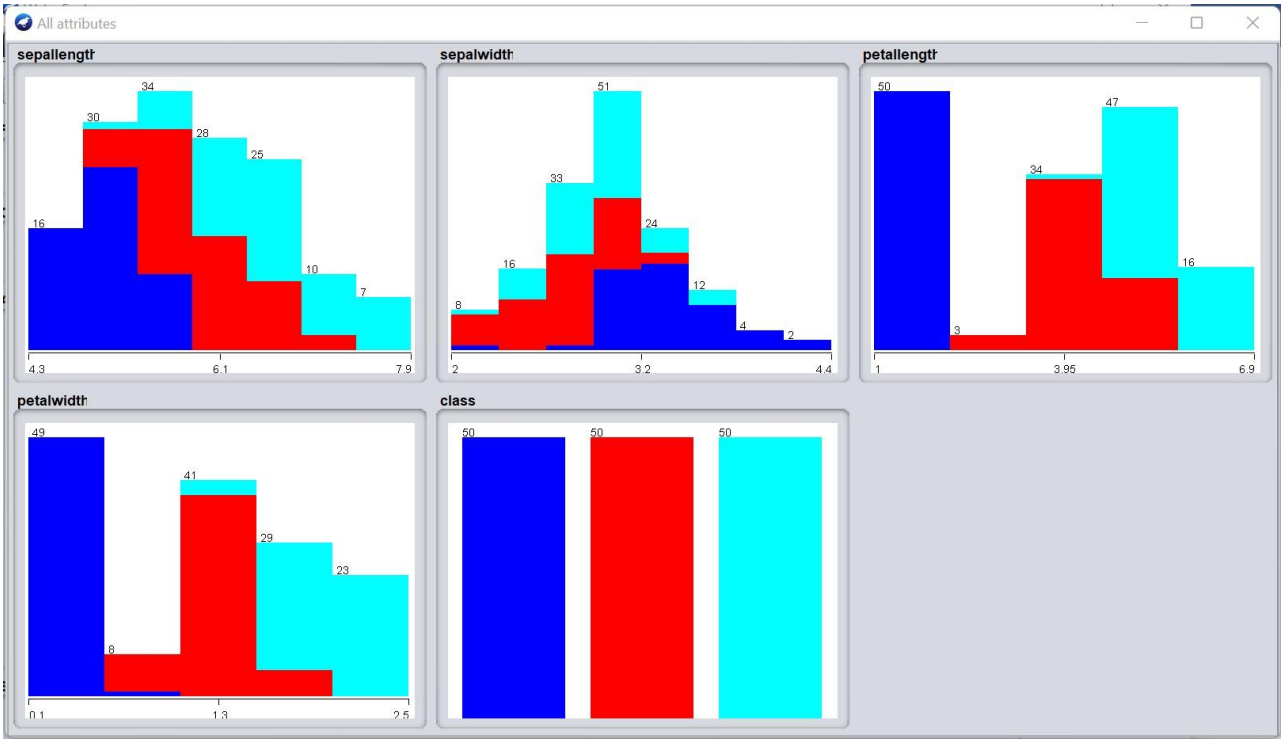
Now we move to our next tab Classify where the applying algorithm part is done. We are going to apply decision tree algorithm to our dataset. We will be using J48 classifier. This is the outcome.



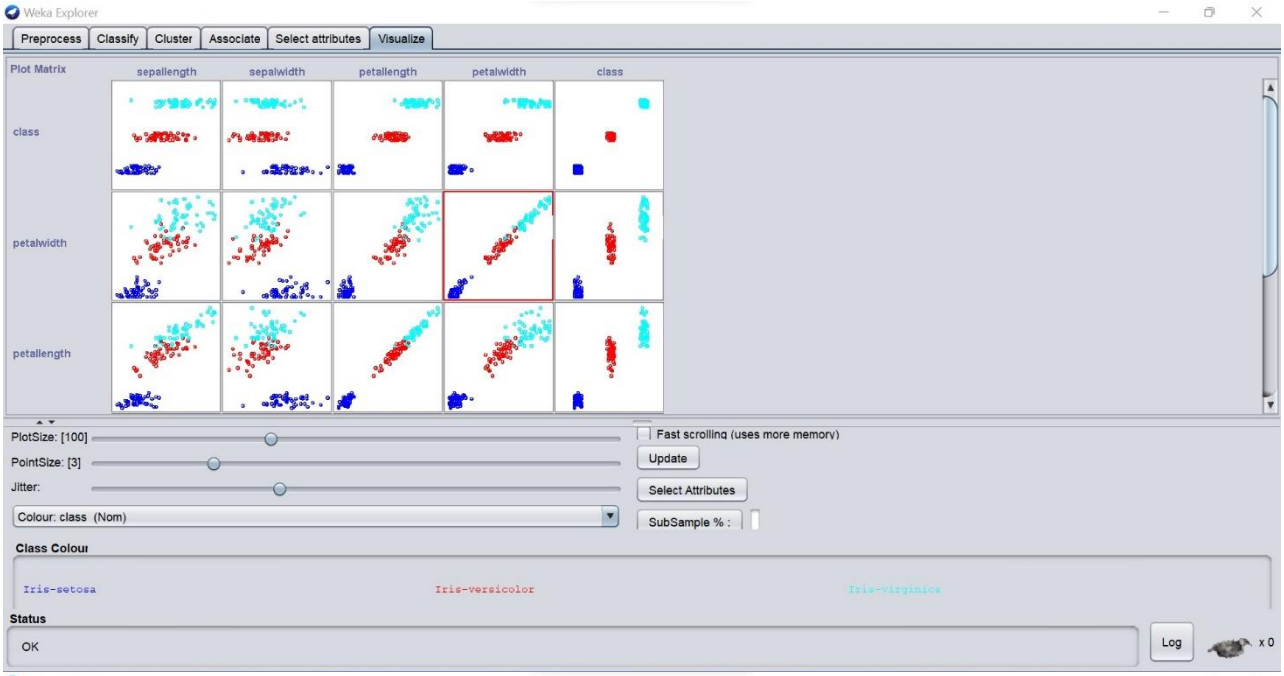


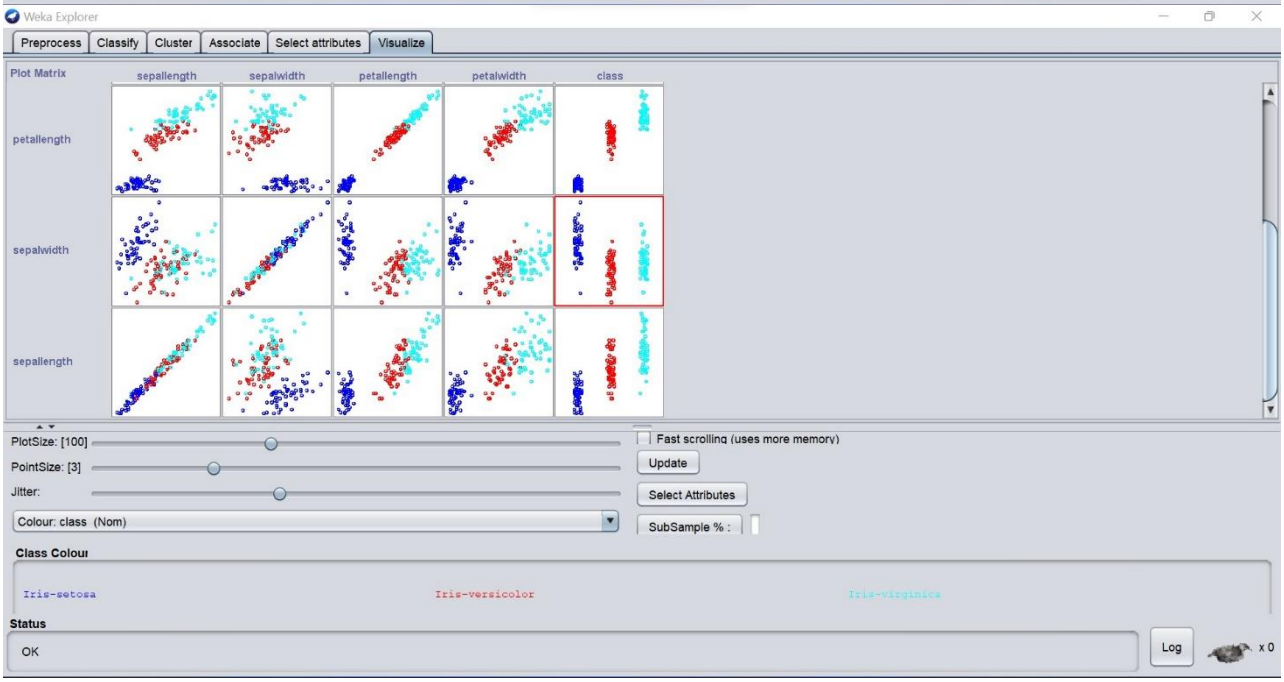


Now let’s visualize our data.



More options can be explored in the last visualize tab.





2.Orange is an easy to use data visualization tool with a large toolkit. In spite of being a GUI-based beginner-friendly tool, you mustn't mistake it for a light-weight one. It can do statistical distributions and box plots as well as decision trees, hierarchical clustering and linear projections.

a. Install orange

b. Show data distribution

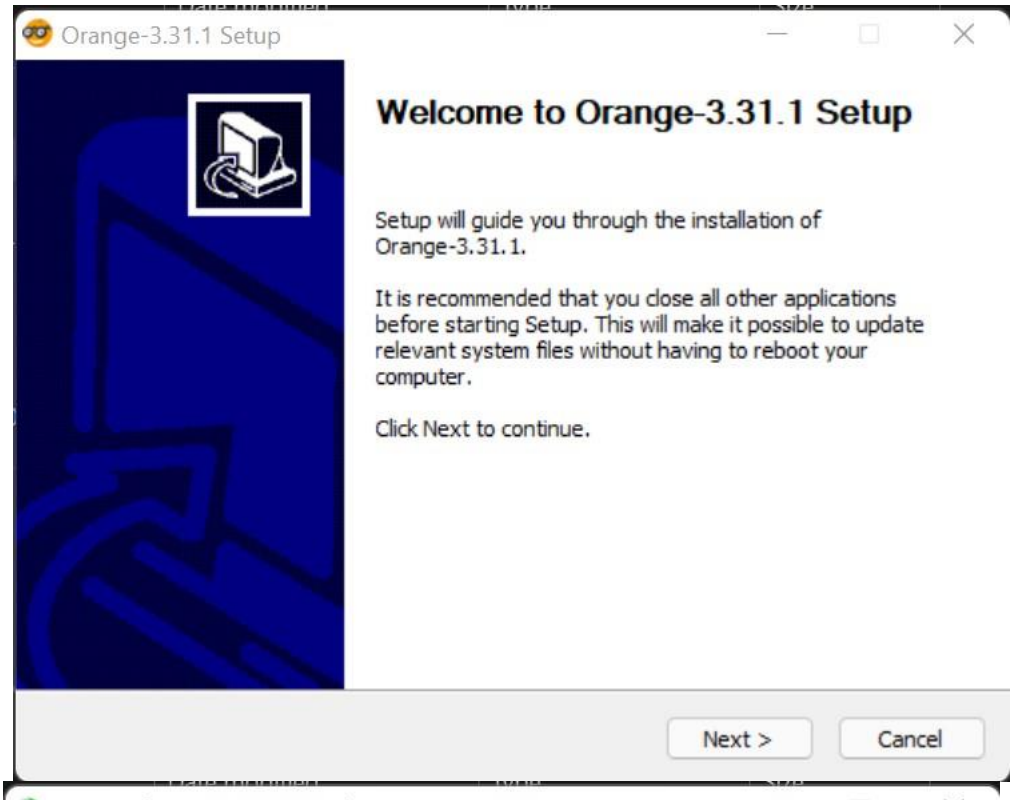
c. Show linear projection

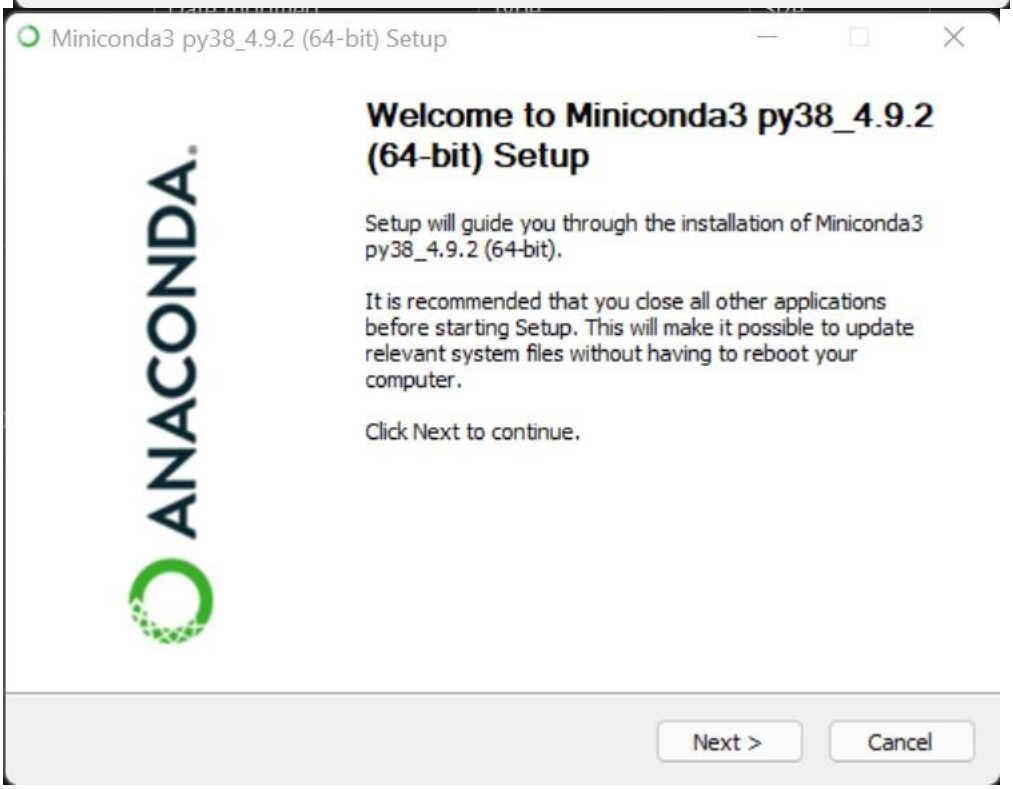
d. Show FreeViz

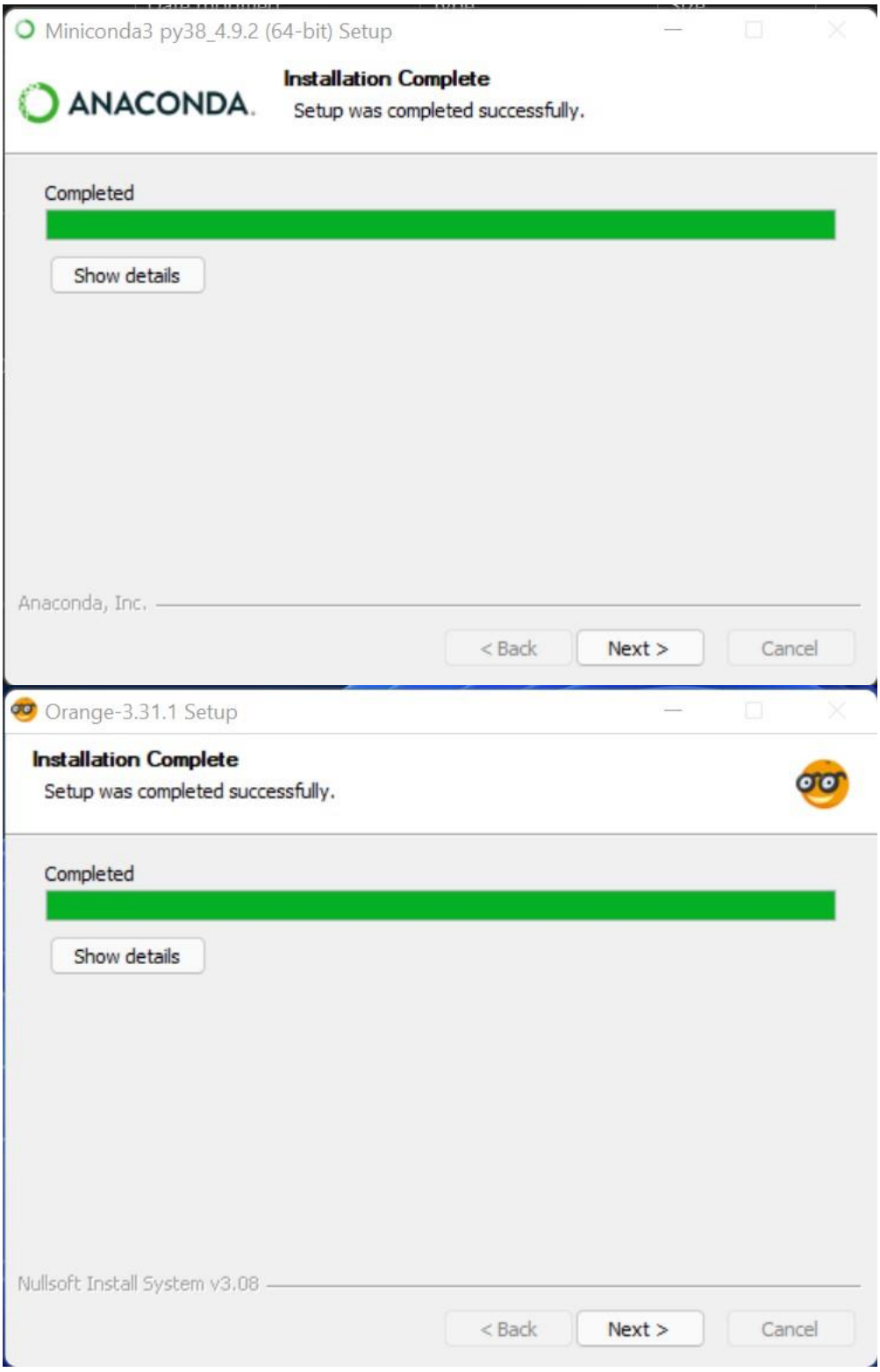
Use dataset:

https://drive.google.com/file/d/1m6sKI1Dap0XK6Bw1edUd5PohwpPwXnd9/view

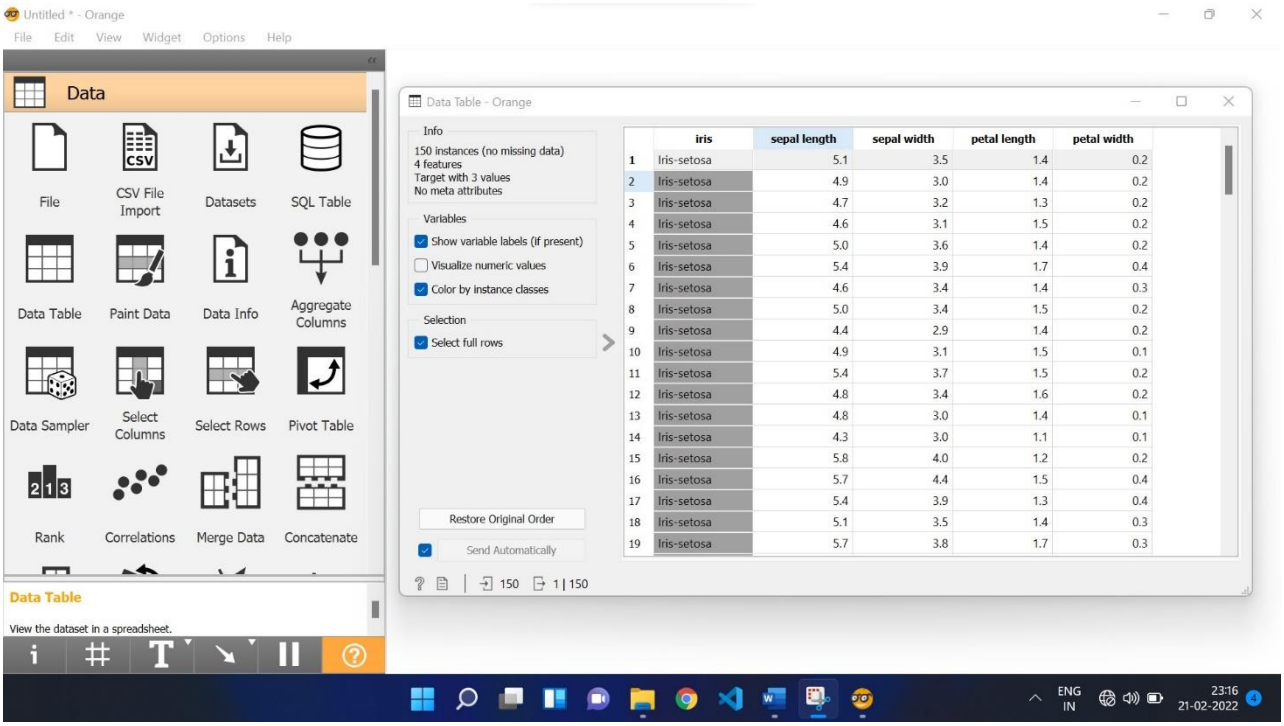
Installation of orange:



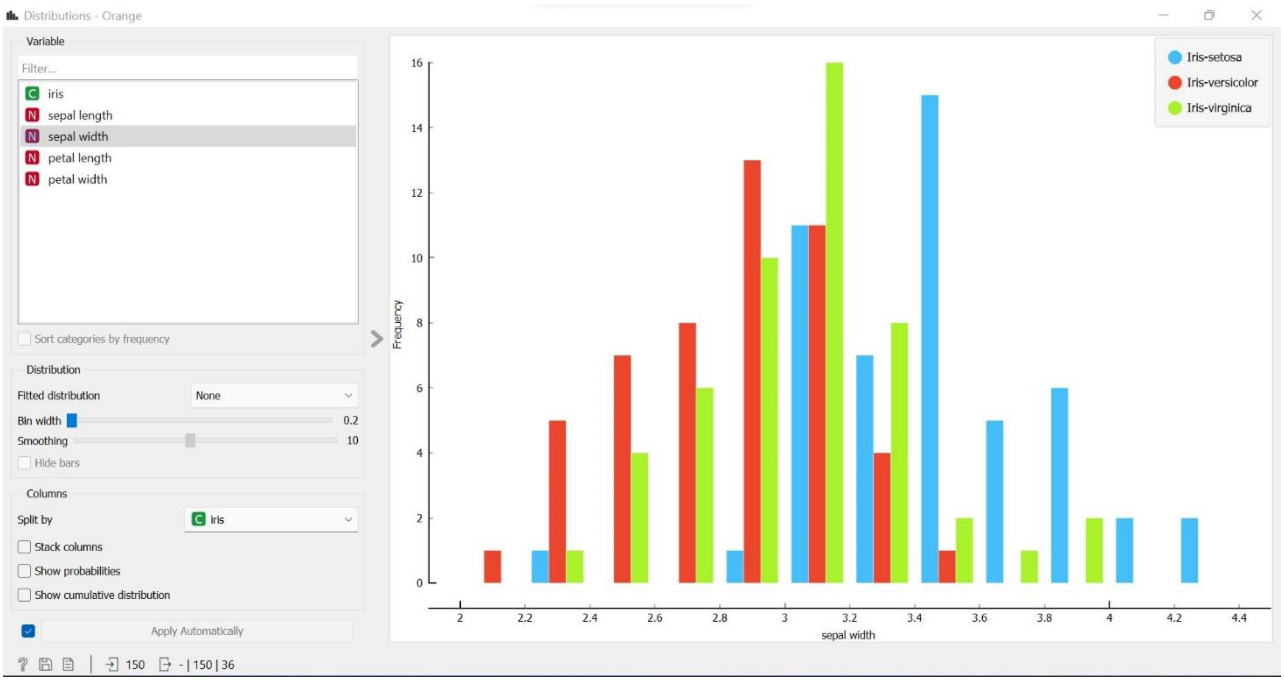


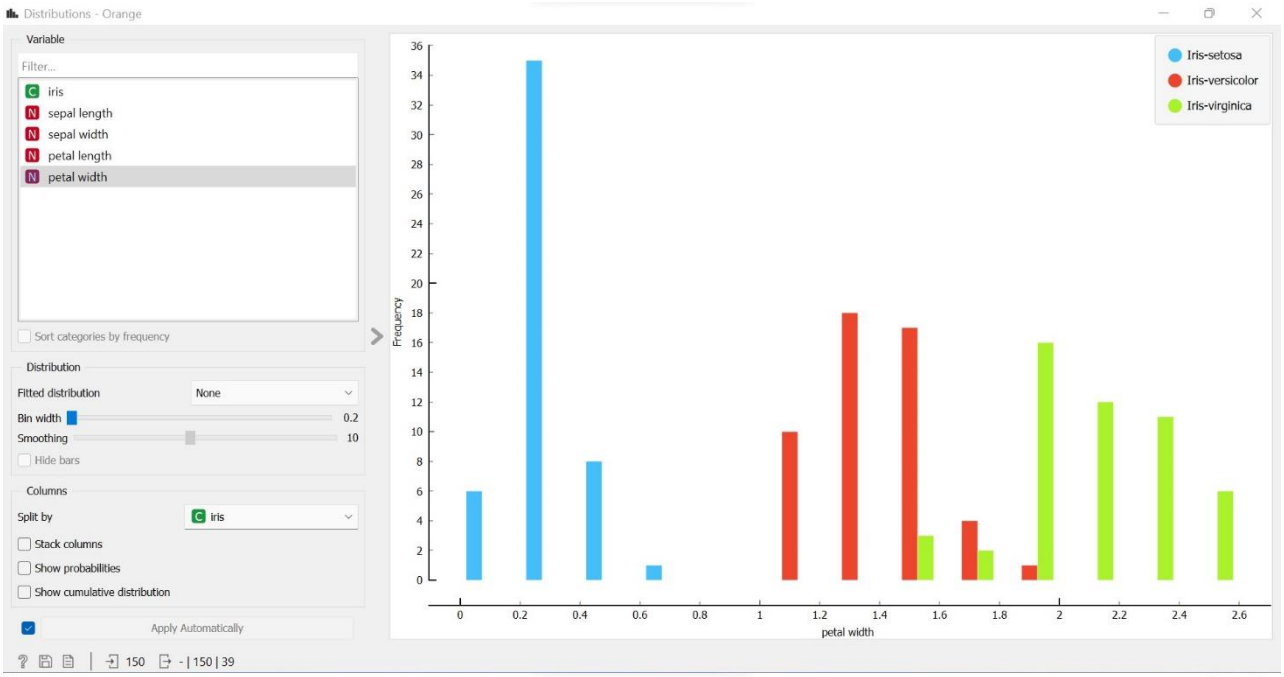

Miniconda3 py38_4.9.2 (64-bit) Setup
ANACONDA.
Installation Complete
Setup was completed successfully.
Completed
Show details
Anaconda, Inc.
< Back
Next >
Cancel
Orange-3.31.1 Setup
Installation Complete
Setup was completed successfully.
Completed
Show details
Nullsoft Install System v3.08
< Back
Next >
Cancel

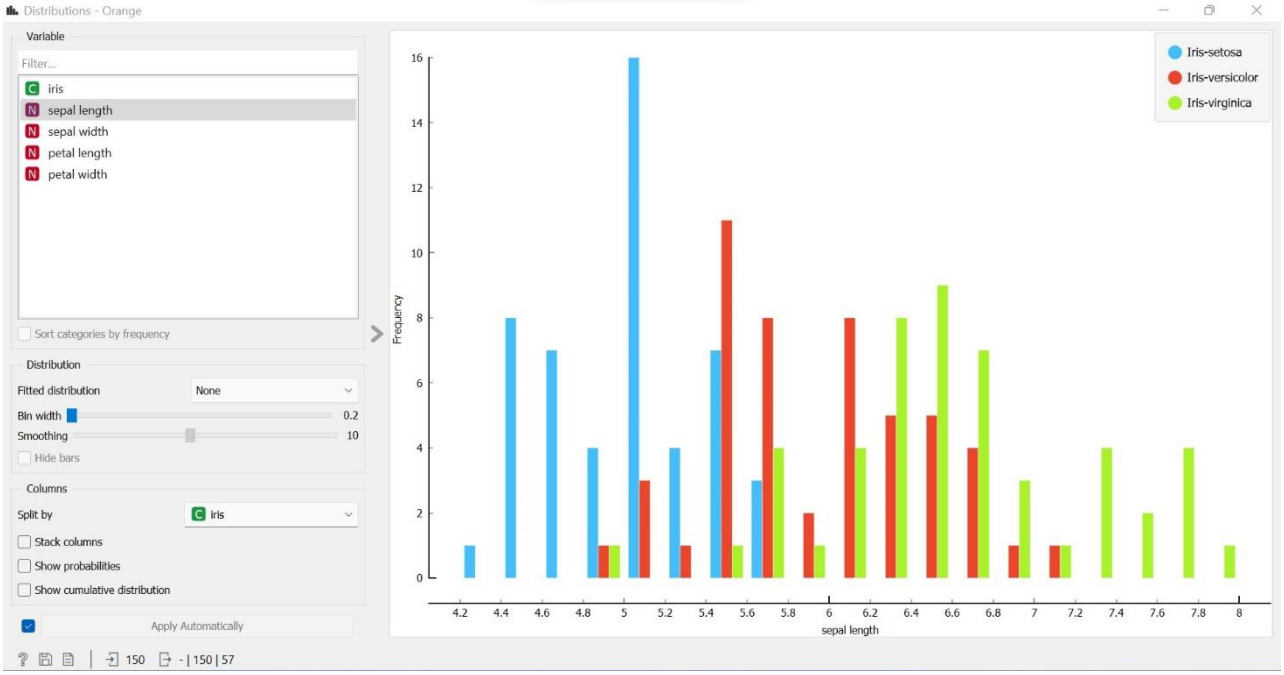
We first open orange software and then open our dataset. This is how it looks like.



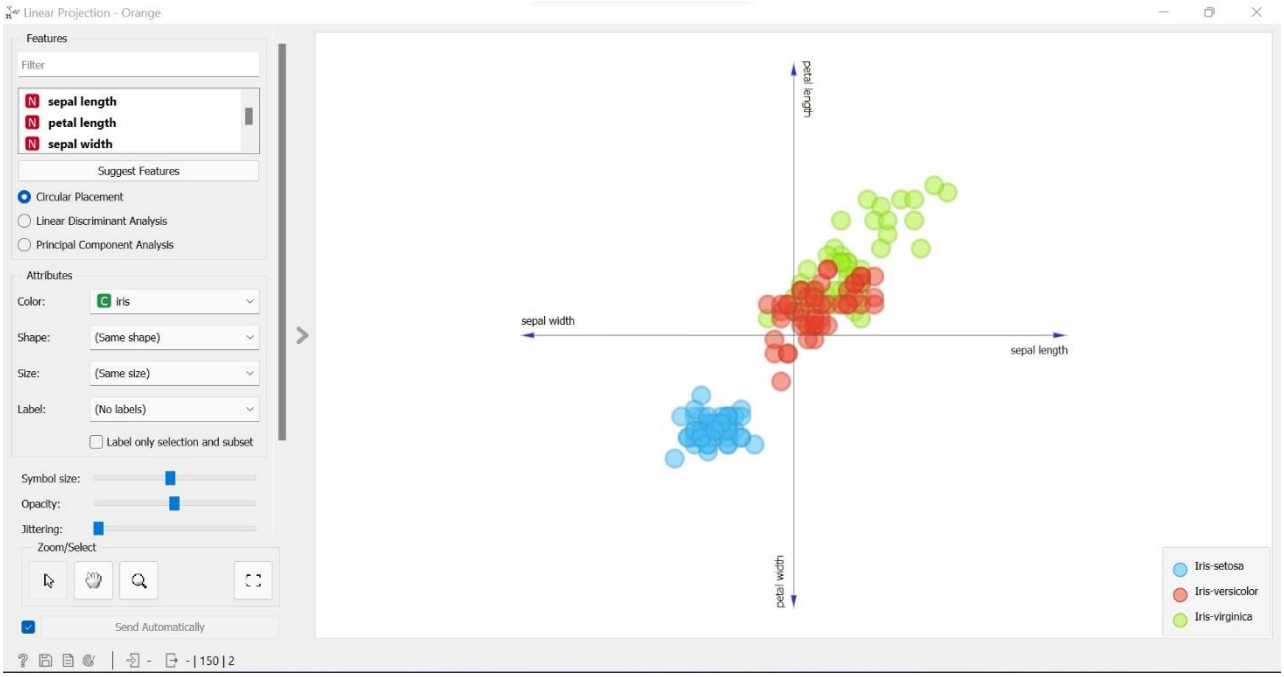
Data Distribution:

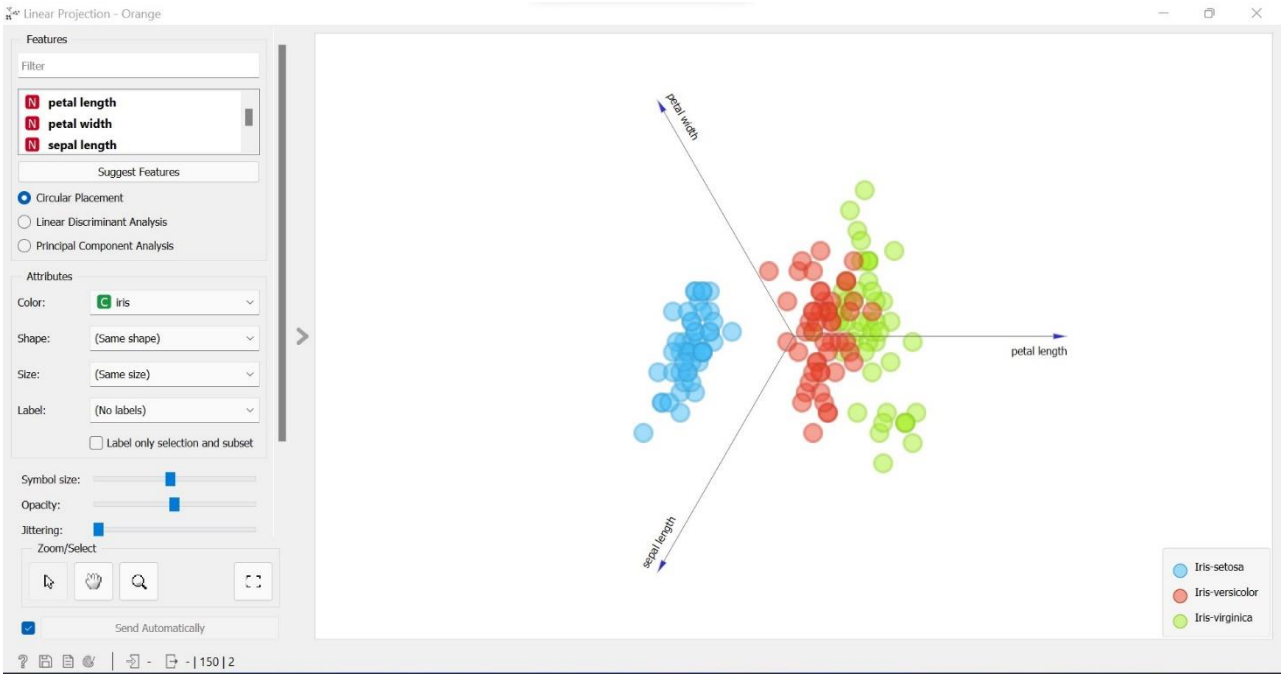



Distributions - Orange
Variable
Filter...
iris
sepal length
sepal width
petal length
petal width
Sort categories by frequency
Distribution
Fitted distribution
None
Bin width
0.2
Smoothing
10
Hide bars
Columns
Split by
iris
Stack columns
Show probabilities
Show cumulative distribution
Apply Automatically
150
- | 150 | 39
Iris-setosa
Iris-versicolor
Iris-virginica
Frequency
petal width


Distributions - Orange
Variable
Filter...
iris
sepal length
sepal width
petal length
petal width
Sort categories by frequency
Distribution
Fitted distribution
None
Bin width
0.2
Smoothing
10
Hide bars
Columns
Split by
iris
Stack columns
Show probabilities
Show cumulative distribution
Apply Automatically
150
- | 150 | 57
Iris-setosa
Iris-versicolor
Iris-virginica
Frequency
sepal length

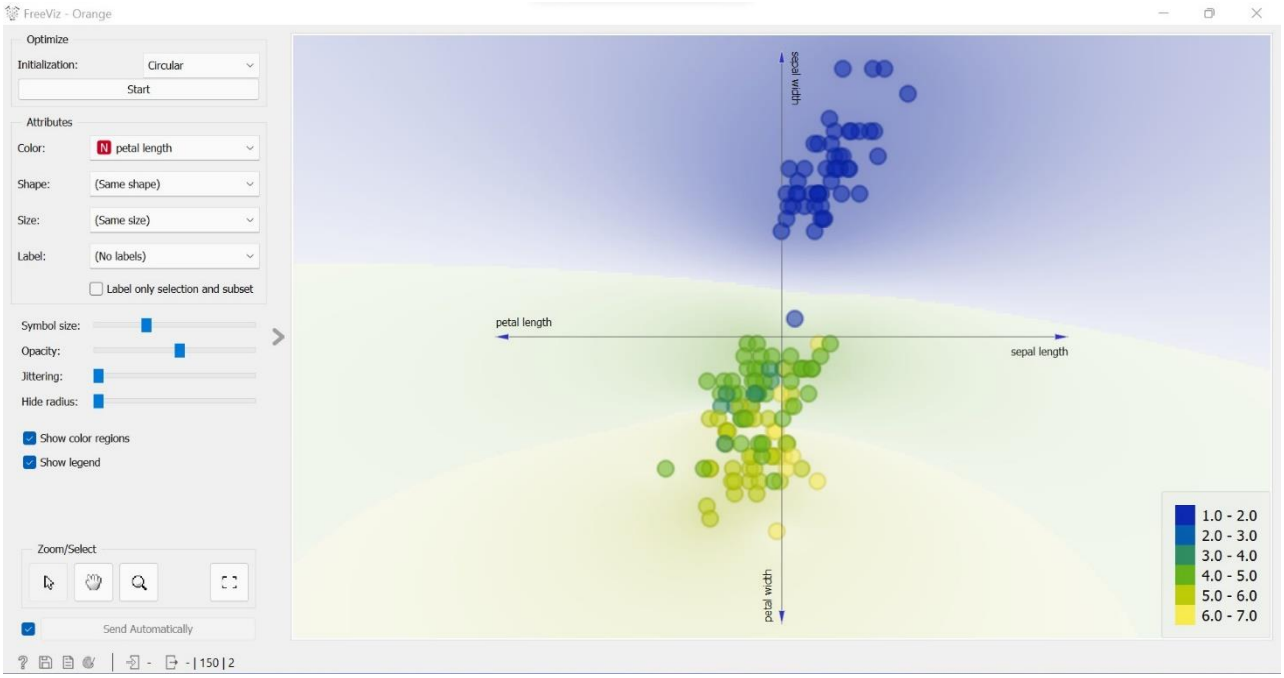
Linear Projection:

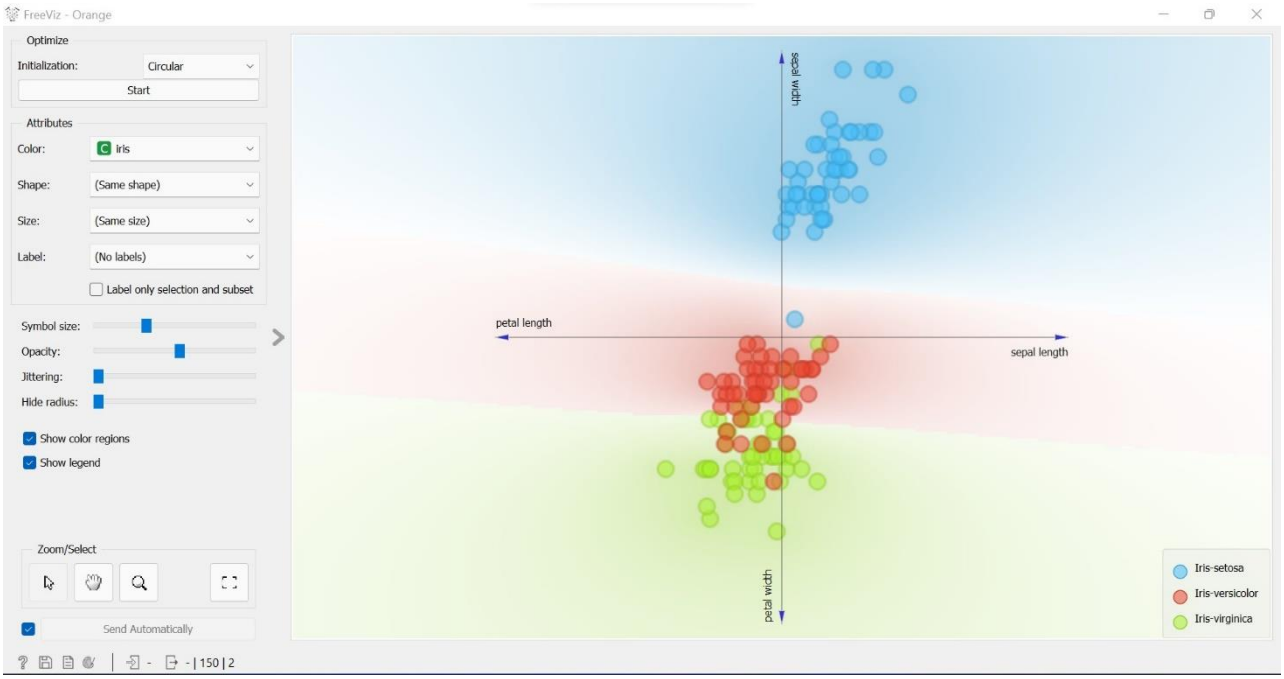





Linear Projection - Orange
Features
Filter
sepal length
sepal width
petal length
Suggest Features
Circular Placement
Linear Discriminant Analysis
Principal Component Analysis
Attributes
Color: iris
Shape: (Same shape)
Size: (Same size)
Label: (No labels)
Label only selection and subset
Symbol size:
Opacity:
Jittering:
Zoom/Select
Send Automatically
- | 150 | 2
sepal length
Iris-setosa
Iris-versicolor
Iris-virginica

FreeViz:


FreeViz - Orange
Optimize
Initialization: Circular
Start
Attributes
Color: petal length
Shape: (Same shape)
Size: (Same size)
Label: (No labels)
Label only selection and subset
Symbol size:
Opacity:
Jittering:
Hide radius:
Show color regions
Show legend
Zoom/Select
Send Automatically
- | 150 | 2
sepal width
petal length
sepal length
petal width
1.0 - 2.0
2.0 - 3.0
3.0 - 4.0
4.0 - 5.0
5.0 - 6.0
6.0 - 7.0



3. Differentiate in between free software, Open source software and proprietary software with respect to its properties.

Free software	Open software	Proprietary software
Source Code may be openly available	Code is openly available	Source Code is available with organization
Complete Freedom	Also has Freedom	No Freedom
Can be installed easily	Can be installed easily on any computer	Cannot be installed without license
Users can modify	We can modify program	Only organizations can modify
Users don’t need license	No need for license	Users need authenticat-ed license
It can be free	This is free software	Users have to pay for software

4. Using Anaconda Python create Histogram, Scatter plot and Bar plot for the dataset given below.

Dataset- https://drive.google.com/file/d/1i11BZFe8Xj9kNq7eeE9KOa_Iz1KhEdXJ/view

a. Scatter plot- Scatter plot of Price Vs Age
b. Histogram- for Kilometer and CC
c. Bar plot- Bar plot for different fuel types

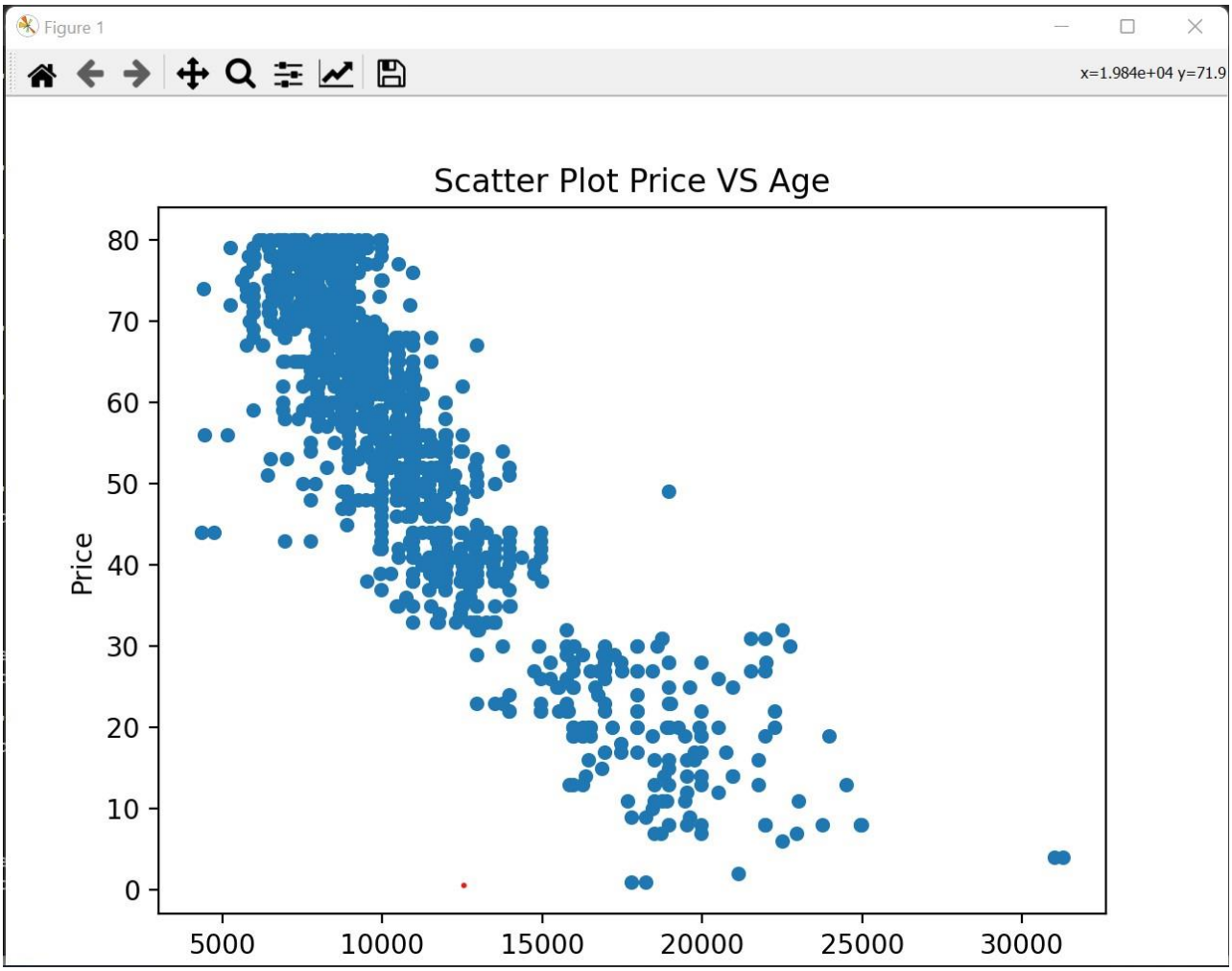
Ans:

First let's load the data and see the info.

```
(base) D:\College_Work\TY-2021-2022\SEM VI\LAB\SET\Assignment-1>python
Python 3.9.7 (default, Sep 16 2021, 16:59:28) [MSC v.1916 64 bit (AMD64)] :: Anaconda, Inc. on win32
Type "help", "copyright", "credits" or "license" for more information.
>>> import numpy as np
>>> import pandas as pd
>>> import matplotlib.pyplot as plt
>>> import seaborn as sns
>>> data=pd.read_csv('Toyota.csv')
>>> data.head()
   Unnamed: 0  Price   Age     KM FuelType  HP  MetColor  Automatic    CC  Doors  Weight
0           0  13500  23.0  46986   Diesel  90       1.0          0  2000  three    1165
1           1  13750  23.0  72937   Diesel  90       1.0          0  2000      3    1165
2           2  13950  24.0  41711   Diesel  90       NaN          0  2000      3    1165
3           3  14950  26.0  48000   Diesel  90       0.0          0  2000      3    1165
4           4  13750  30.0  38500   Diesel  90       0.0          0  2000      3    1170
>>>
```

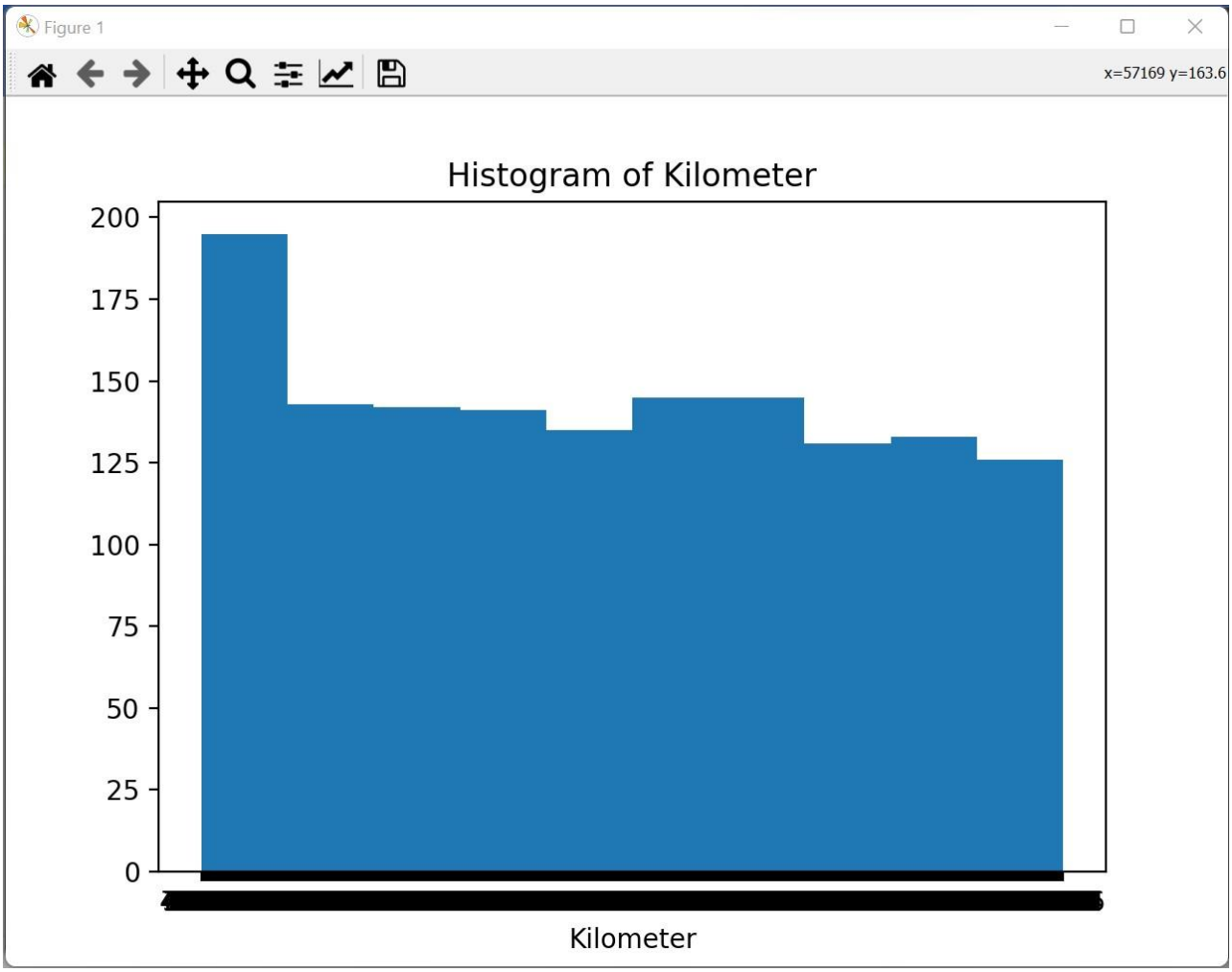
a.Scatter plot- Scatter plot of Price Vs Age

```
>>> data.plot(kind='scatter', x='Price', y='Age')
<AxesSubplot:xlabel='Price', ylabel='Age'>
>>> plt.xlabel('Age in years')
Text(0.5, 0, 'Age in years')
>>> plt.ylabel('Price')
Text(0, 0.5, 'Price')
>>> plt.title('Scatter Plot Price VS Age')
Text(0.5, 1.0, 'Scatter Plot Price VS Age')
>>> plt.show()
```

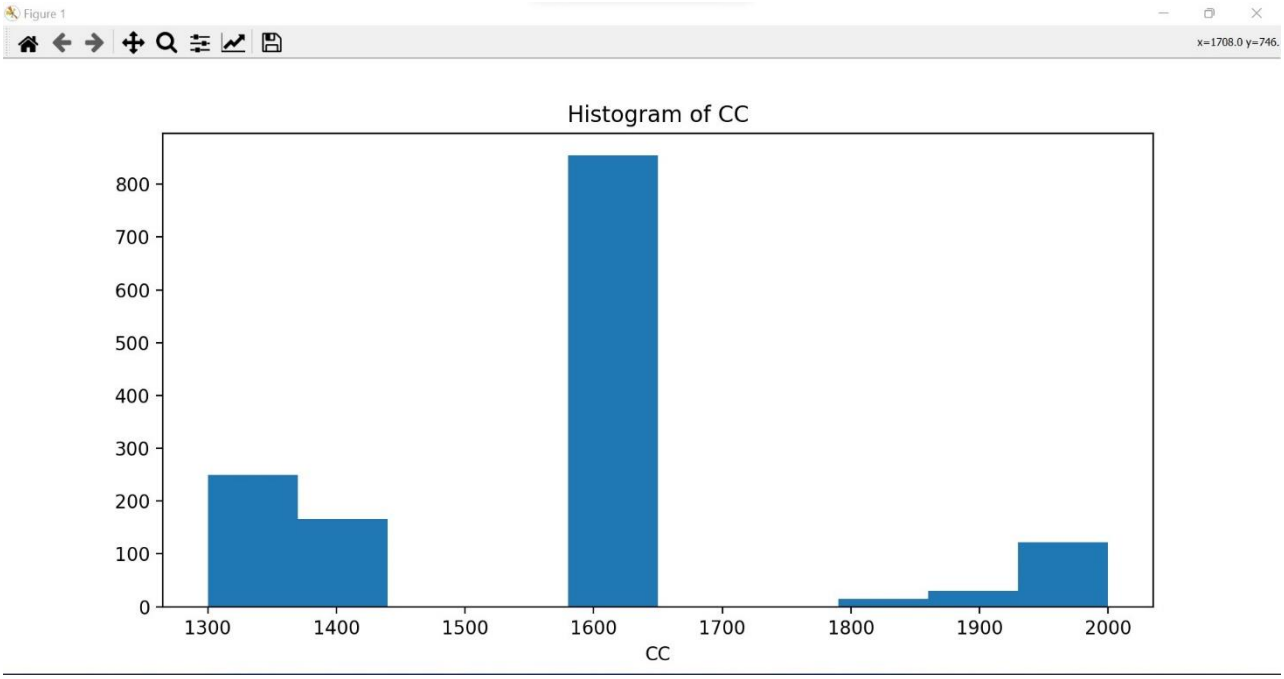


b.Histogram- for Kilometer and CC

```
>>> plt.hist(data['KM'])
(array([195., 143., 142., 141., 135., 145., 145., 131., 133., 126.]), array([   0. ,  125.5,  251. ,  376.5,  502. ,  627.5,  753. ,  878.5,
       1004. , 1129.5, 1255. ]), <BarContainer object of 10 artists>)
>>> plt.title('Histogram of Kilometer')
Text(0.5, 1.0, 'Histogram of Kilometer')
>>> plt.xlabel('Kilometer')
Text(0.5, 0, 'Kilometer')
>>> plt.show()
```

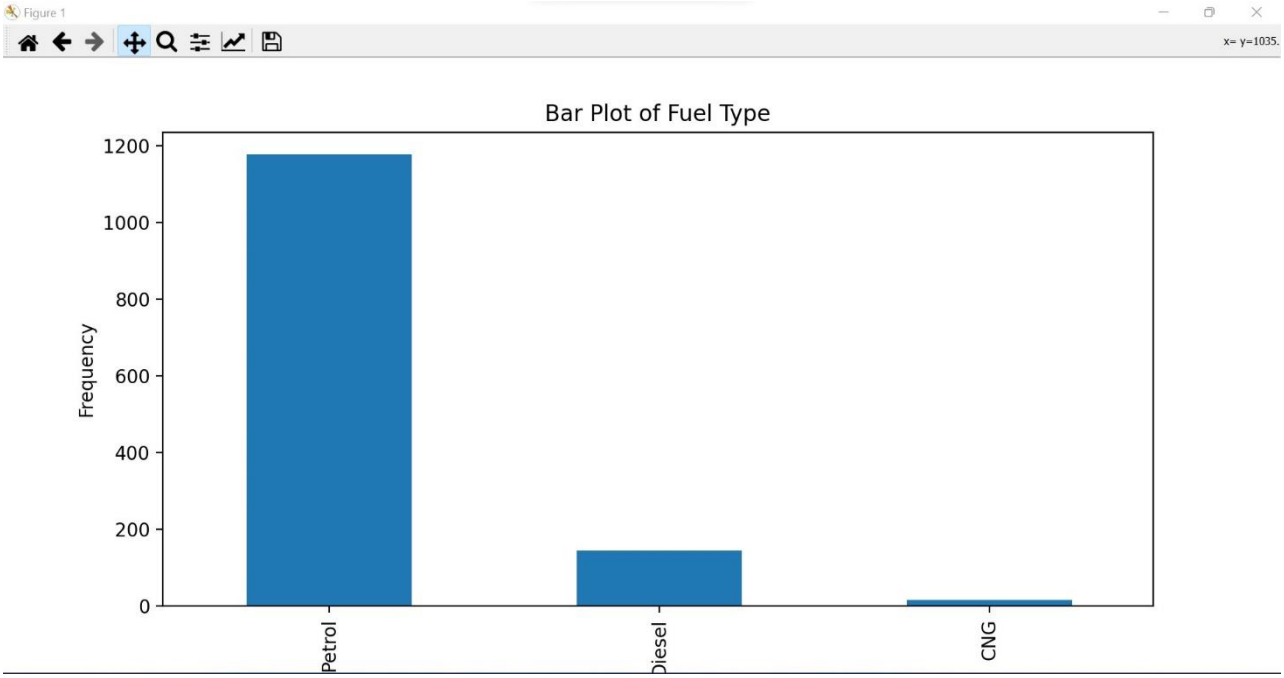



```
>>> plt.hist(data['CC'])
(array([250., 166.,   0.,   0., 854.,   0.,   0.,  14.,  30., 122.]), array([1300., 1370., 1440., 1510., 1580., 1650., 1720., 1790., 1860.,
       1930., 2000.]), <BarContainer object of 10 artists>)
>>> plt.title('Histogram of CC')
Text(0.5, 1.0, 'Histogram of CC')
>>> plt.xlabel('CC')
Text(0.5, 0, 'CC')
>>> plt.show()
```



c. Bar plot- Bar plot for different fuel types

```
>>> fuel_count=pd.value_counts(data['FuelType'].values, sort=True)
>>> plt.xlabel('Fuel Type')
Text(0.5, 0, 'Fuel Type')
>>> plt.ylabel('Frequency')
Text(0, 0.5, 'Frequency')
>>> plt.title('Bar Plot of Fuel Type')
Text(0.5, 1.0, 'Bar Plot of Fuel Type')
>>> fuel_count.plot.bar()
<AxesSubplot:title={'center':'Bar Plot of Fuel Type'}, xlabel='Fuel Type', ylabel='Frequency'>
>>> plt.show()
```



5. Enlist some examples along with its purpose and properties (at least 10) of FOSS and proprietary software with respect to database.

1.Open Source:

- Firefox—a Web browser that competes with Internet Explorer
- OpenOffice—a competitor to Microsoft Office
- Gimp—a graphic tool with features found in Photoshop
- MySQL, Ingres, and EnterpriseDB—open source database software packages that each go head-to-head with commercial products from Oracle, Microsoft, Sybase, and IBM

- SugarCRM—customer relationship management software that competes with Salesforce.com and Siebel

2. Proprietary:

- Windows and OS X operating systems
- Microsoft Office productivity suite
- Adobe Creative Suite productivity software
- Logic music creation software
- paid-for games for consoles