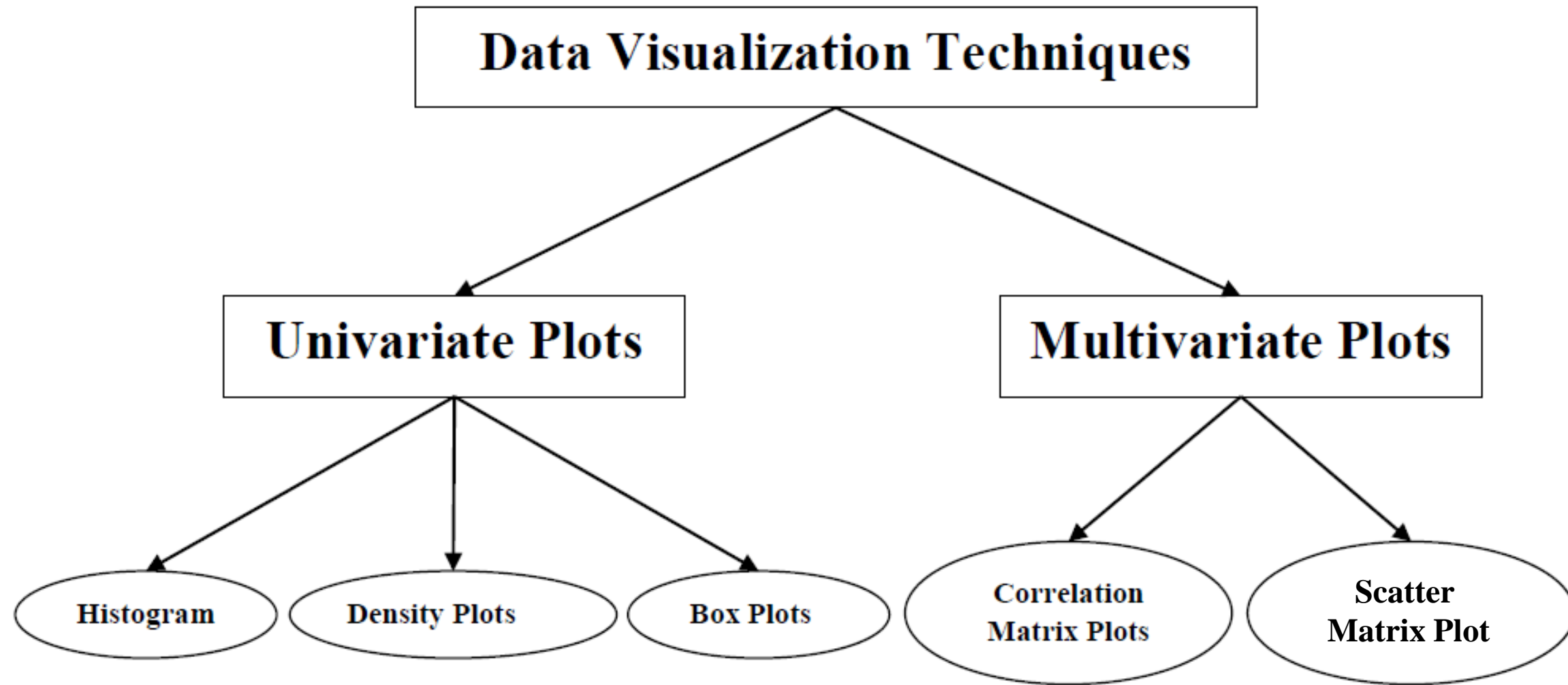




Data Visualization: Understanding Patterns and Relationships

- Data visualization គឺជា tool ដ៏មានតម្លៃ ដែលប្រើដើម្បីស្វែងយល់កាន់តែស៊ីជម្រៅពីទិន្នន័យ។
- វាបកប្រែលេខទៅជាតំណាងដែលមើលឃើញ។
- ជួយក្នុងការយល់ដឹងពី trends, patterns, និង relationships ក្នុងសំណុំទិន្នន័យ។
- ក្នុង machine learning, data visualization មានសារៈសំខាន់សម្រាប់ការរុករកវិភាគ និងបង្កើតគំរូដែលមានប្រសិទ្ធភាព។

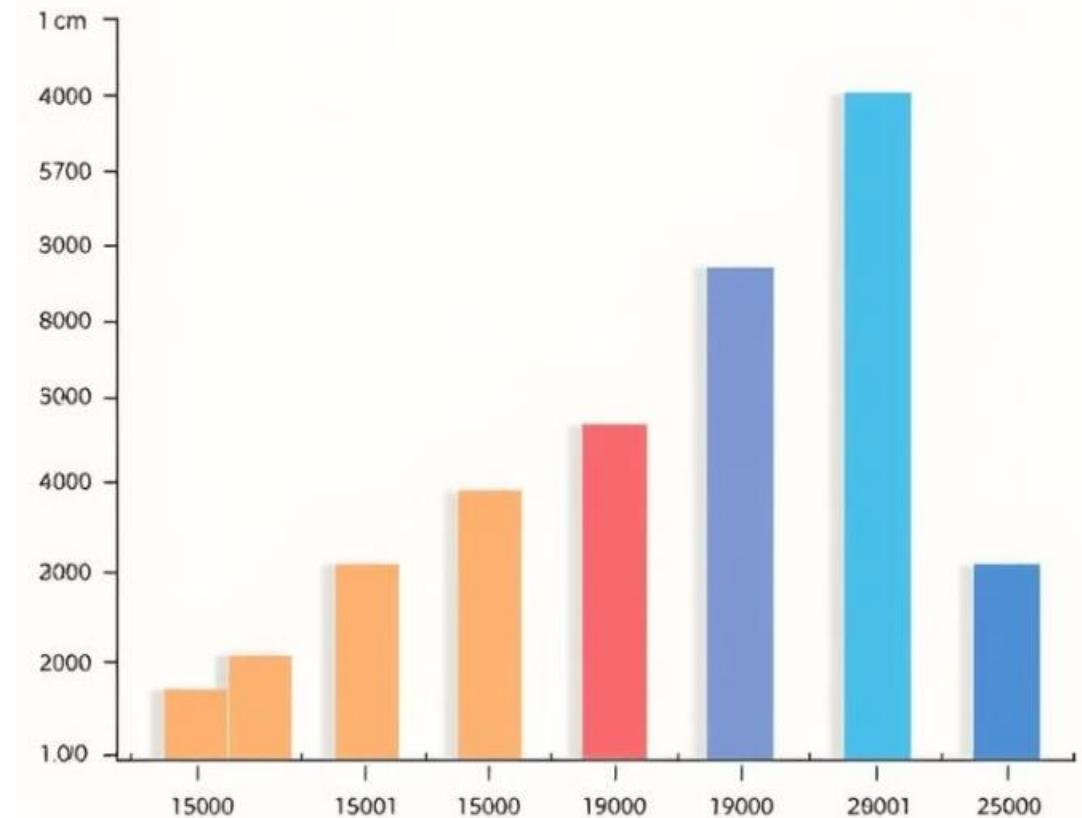




Univariate Visualization: Exploring Individual Attributes

1 Histograms

- Histograms មានសារៈសំខាន់សម្រាប់ការស្វែងយល់ពីទិន្នន័យដែលមានអថេរតែមួយ (single variable) ។
- ជួយក្នុងការអង្កេតមើលរូបរាងនៃការចែកចាយទិន្នន័យ (symmetric, skewed).
- មានប្រយោជន៍សម្រាប់កំណត់អត្តសញ្ញាណ outliers ដែលជាចំណុចទិន្នន័យដែលខុសគ្នាខ្លាំងពីទិន្នន័យភាគច្រើន។

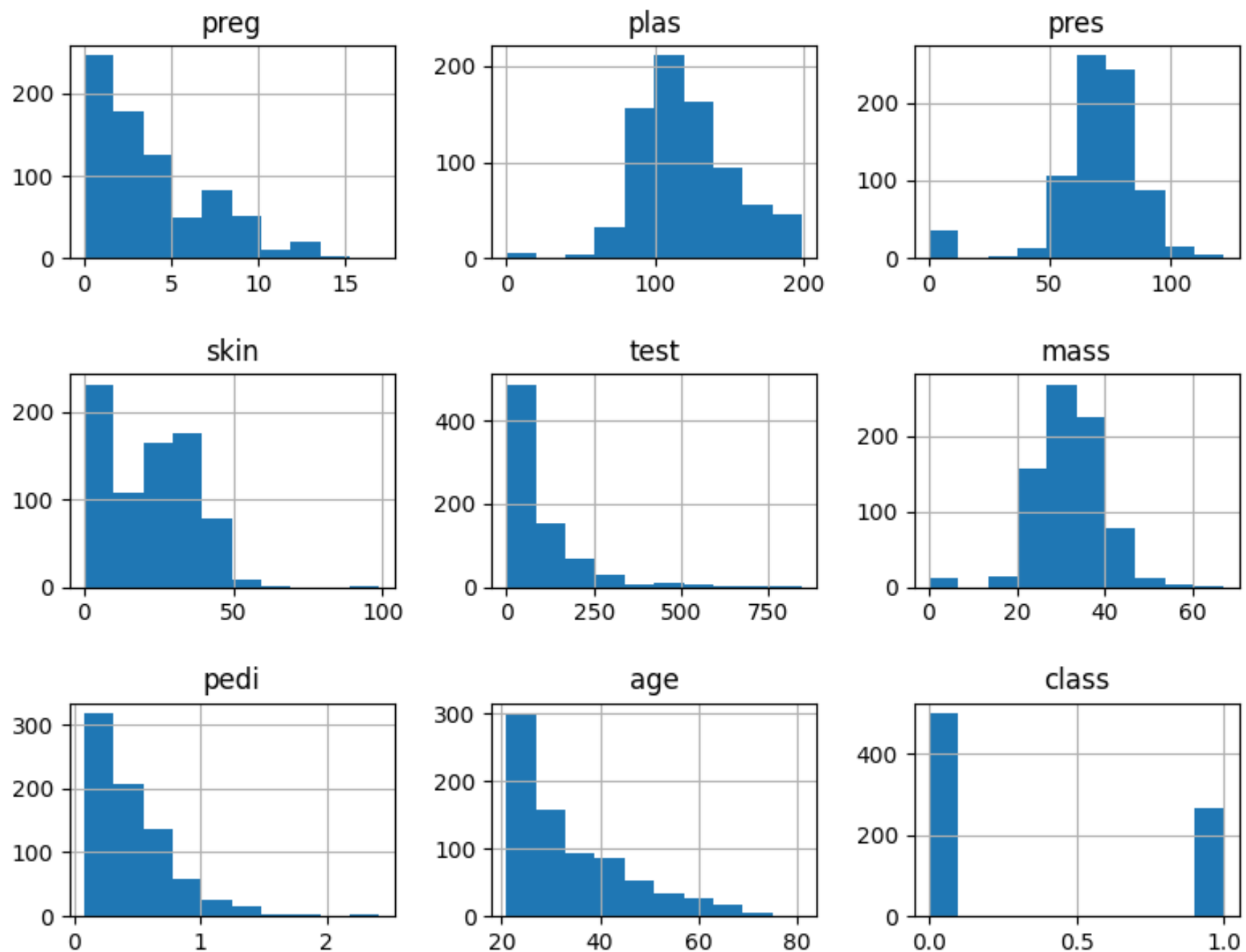


Univariate Visualization: Exploring Individual Attributes

1 Histograms

```
from matplotlib import pyplot
from pandas import read_csv
path = "pima-indians-diabetes.csv"
names = ['preg', 'plas', 'pres', 'skin', 'test', 'mass',
         'pedi', 'age', 'class']
data = read_csv(path, names=names)
data.hist()
pyplot.show()
```

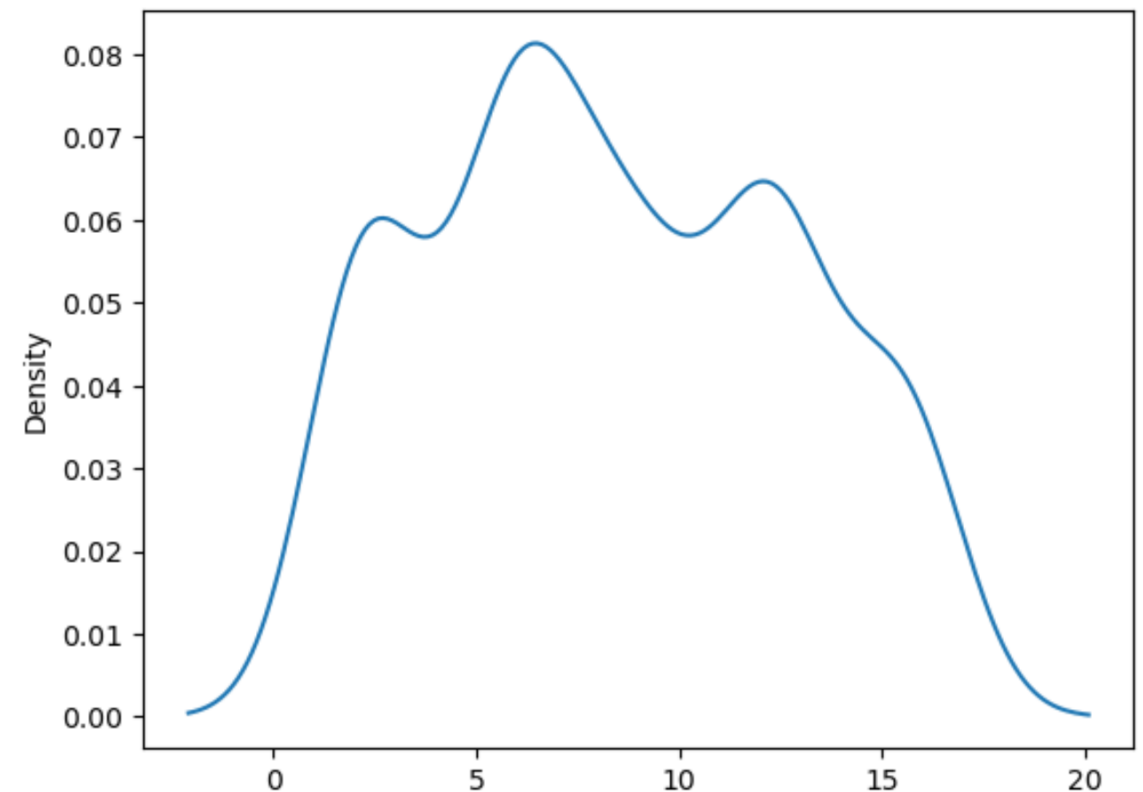
Univariate Visualization: Exploring Individual Attributes



Univariate Visualization: Exploring Individual Attributes

2 Density Plots

- Density plots ផ្តល់នូវតំណាងយ៉ាងរលូននៃការចែកចាយទិន្នន័យបើប្រៀបធៀបទៅនឹង Histograms។
- ពួកគេបង្កើតខ្សែកោងបន្តដែលប៉ាន់ស្មានដង់ស៊ីតេប្រូបាប៊ីលីតេនៃអថេរនៅចំណុចនីមួយៗ។
- Density plots មានប្រយោជន៍សម្រាប់ការមើលឃើញការចែកចាយស្មុគស្មាញនៃទិន្នន័យ។



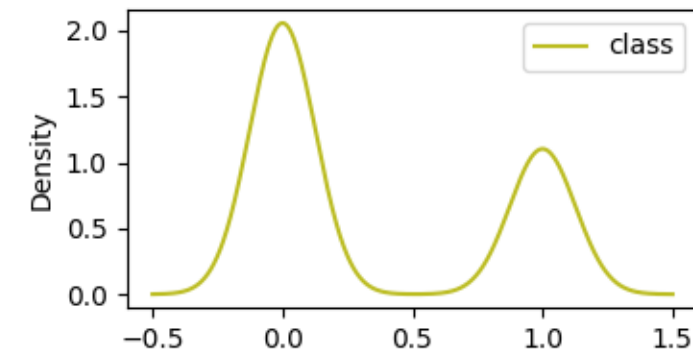
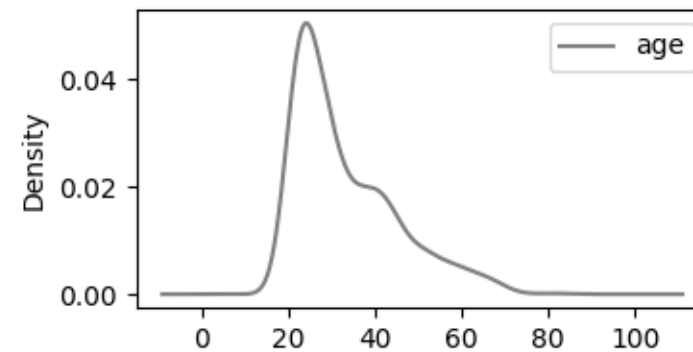
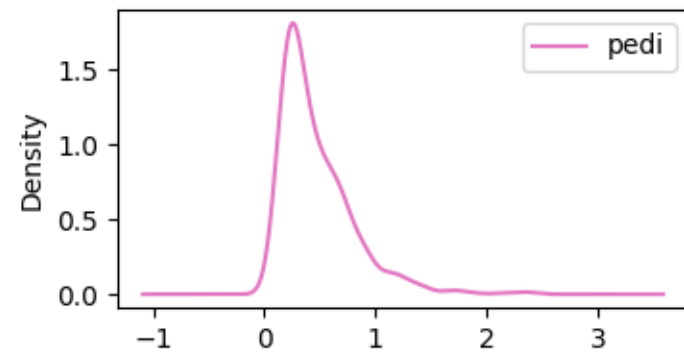
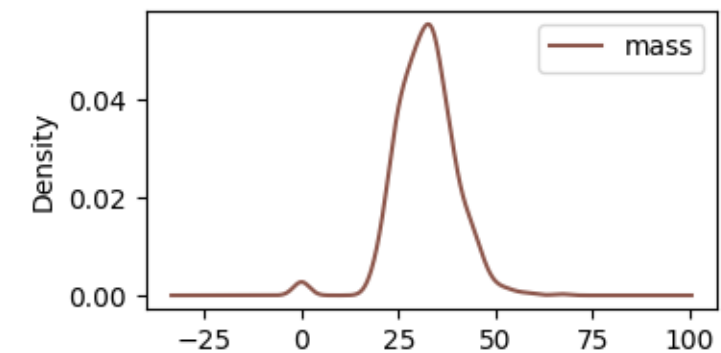
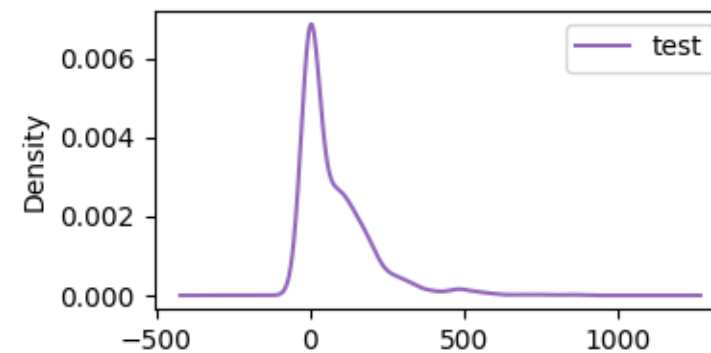
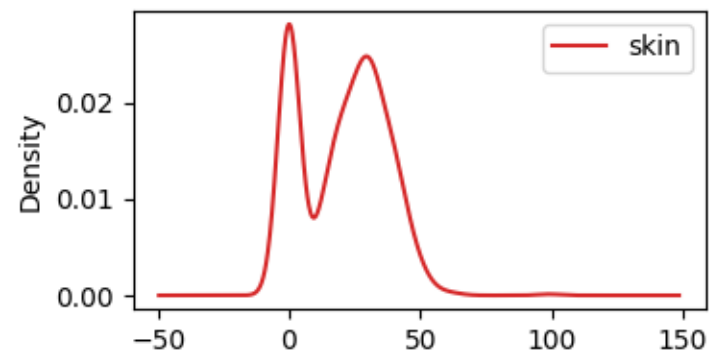
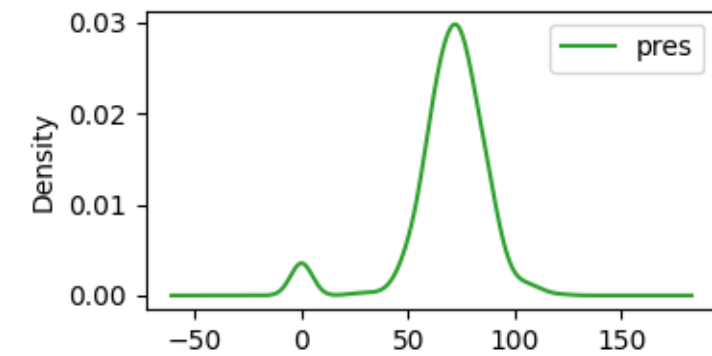
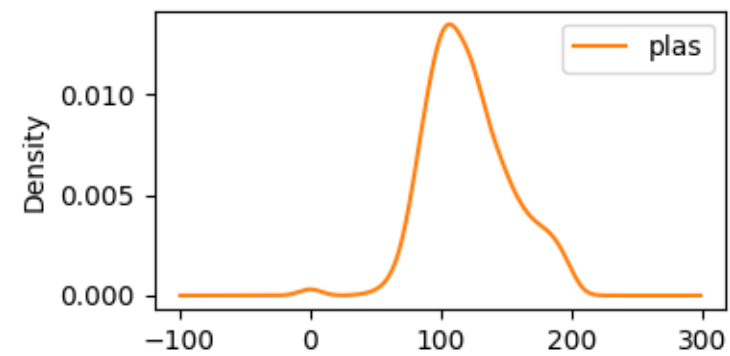
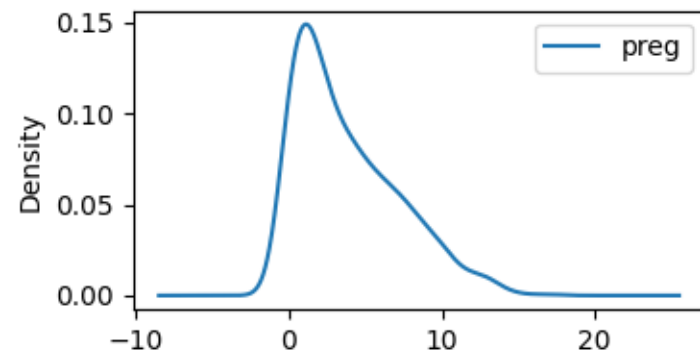
Univariate Visualization: Exploring Individual Attributes

2 Density Plots

```
from matplotlib import pyplot
from pandas import read_csv
path = "pima-indians-diabetes.csv"
names = ['preg', 'plas', 'pres', 'skin', 'test', 'mass',
         'pedi', 'age', 'class']
data = read_csv(path, names=names)
data.plot(kind='density', subplots=True, layout=(3,3),
         sharex=False)

pyplot.show()
```

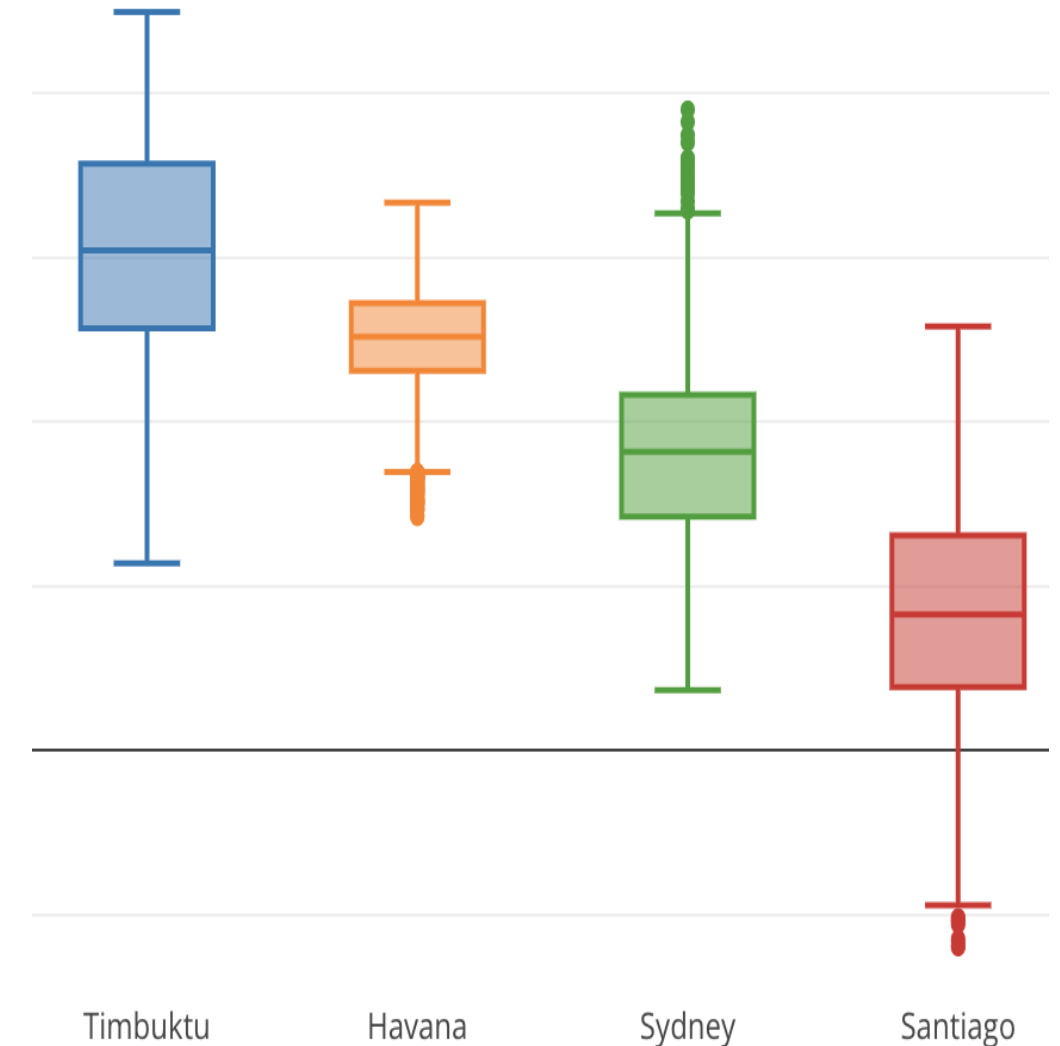
Univariate Visualization: Exploring Individual Attributes



Univariate Visualization: Exploring Individual Attributes

3 Box and whiskers Plots

- វាគូរបន្ទាត់សម្រាប់តម្លៃកណ្តាល ពេលគឺ មេដ្យាន។
- វាគូប្រអប់ជុំវិញ 25% និង 75%។
- វាក៏គូរវីស្ត័រ (whiskers) ដែលនឹងផ្តល់ឱ្យយើងនូវគំនិតអំពីការរីករាលដាលនៃទិន្នន័យ។
- ចំណុចនៅខាងក្រៅវីស្ត័រតំណាងឱ្យតម្លៃលើស។ តម្លៃ Outlier នឹងធំជាងទំហំនៃការរីករាលដាលនៃទិន្នន័យកណ្តាល 1.5 ដង។

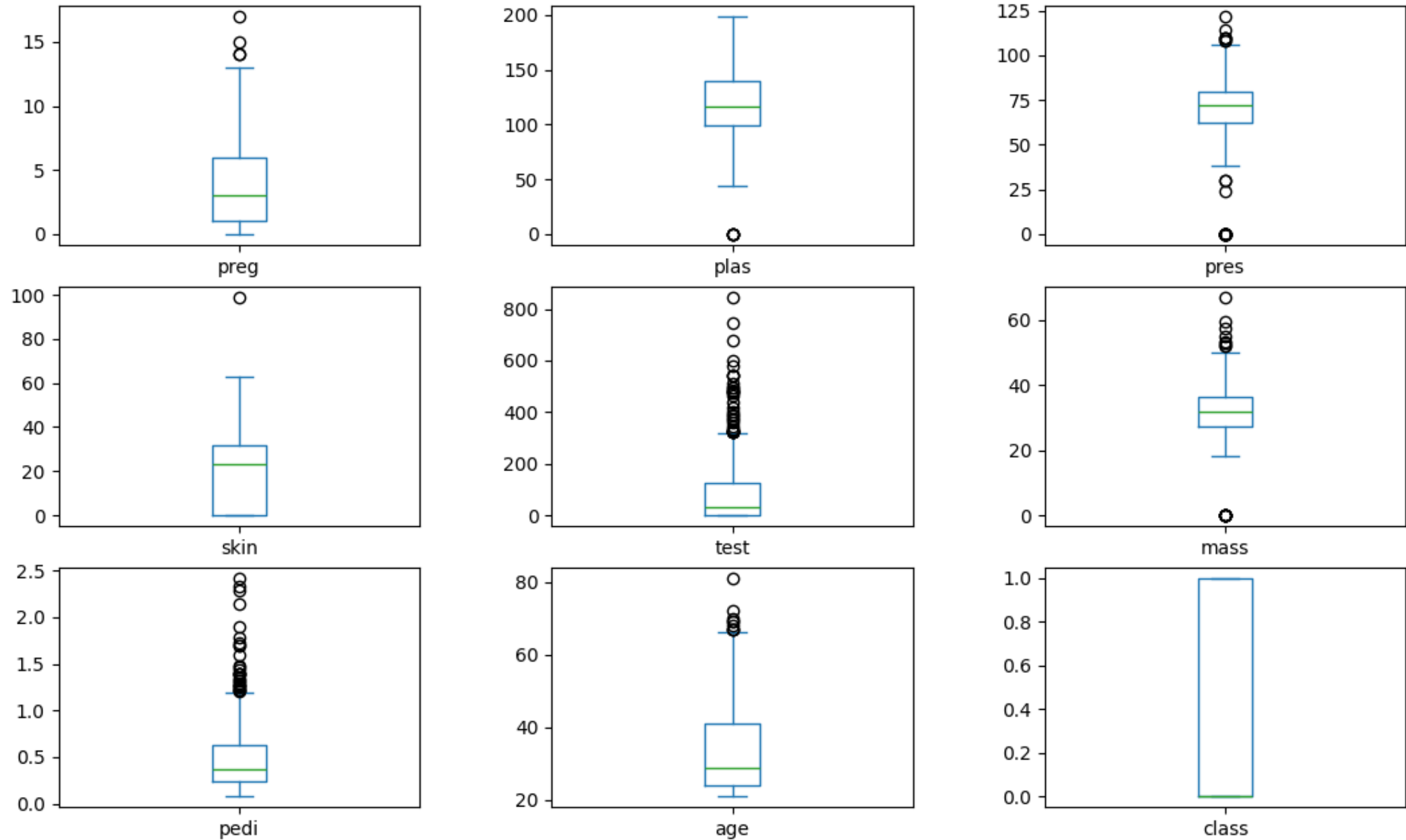


Univariate Visualization: Exploring Individual Attributes

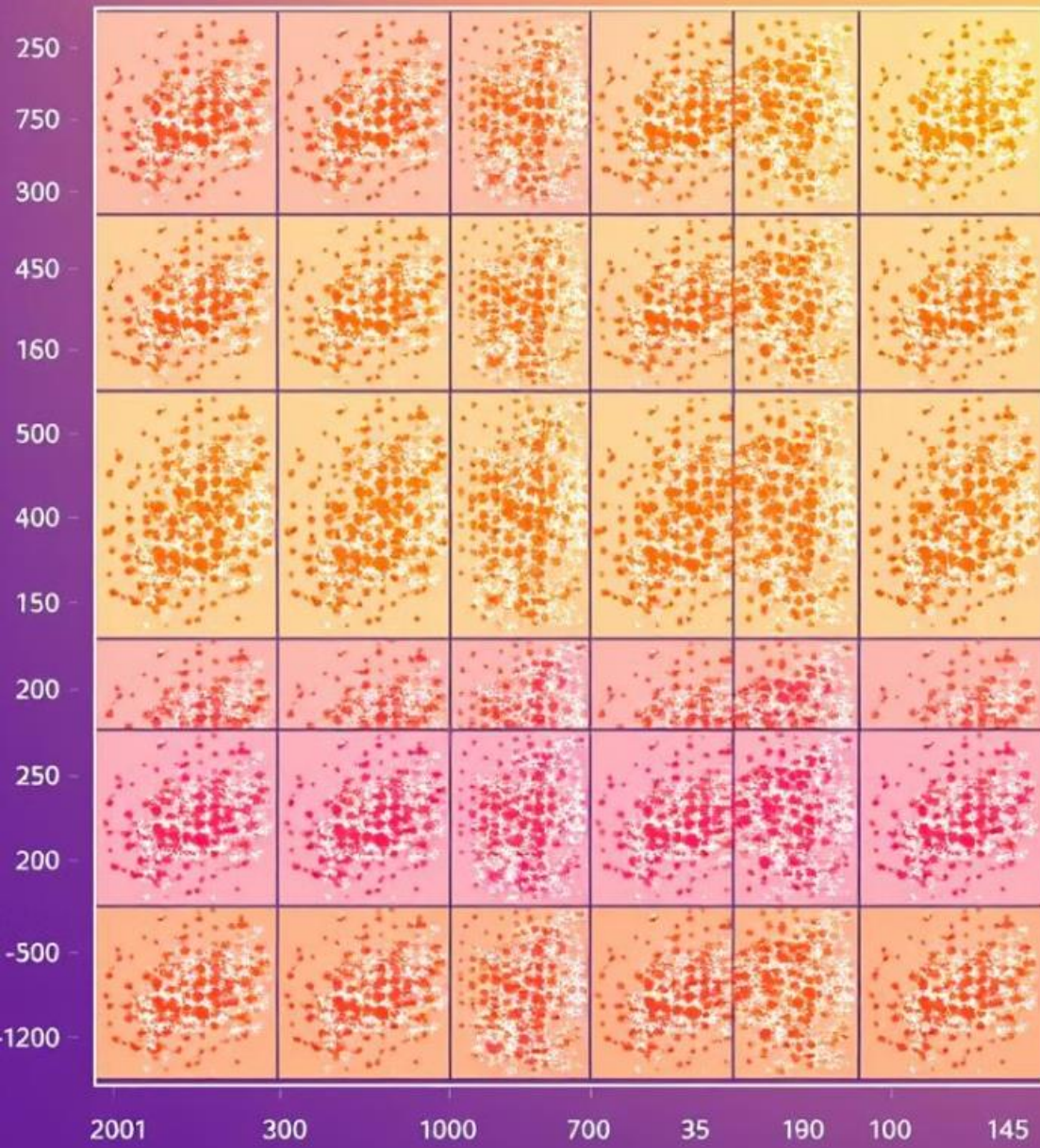
3 Box and whiskers Plots

```
from matplotlib import pyplot
from pandas import read_csv
path = "pima-indians-diabetes.csv"
names = ['preg', 'plas', 'pres', 'skin', 'test', 'mass',
         'pedi', 'age', 'class']
data = read_csv(path, names=names)
data.plot(kind='box', subplots=True, layout=(3, 3),
          sharex=False, sharey=False)
pyplot.show()
```

Univariate Visualization: Exploring Individual Attributes



scatter Matix



75 ok



Multivariate Visualization: Exploring Relationships Between Attributes

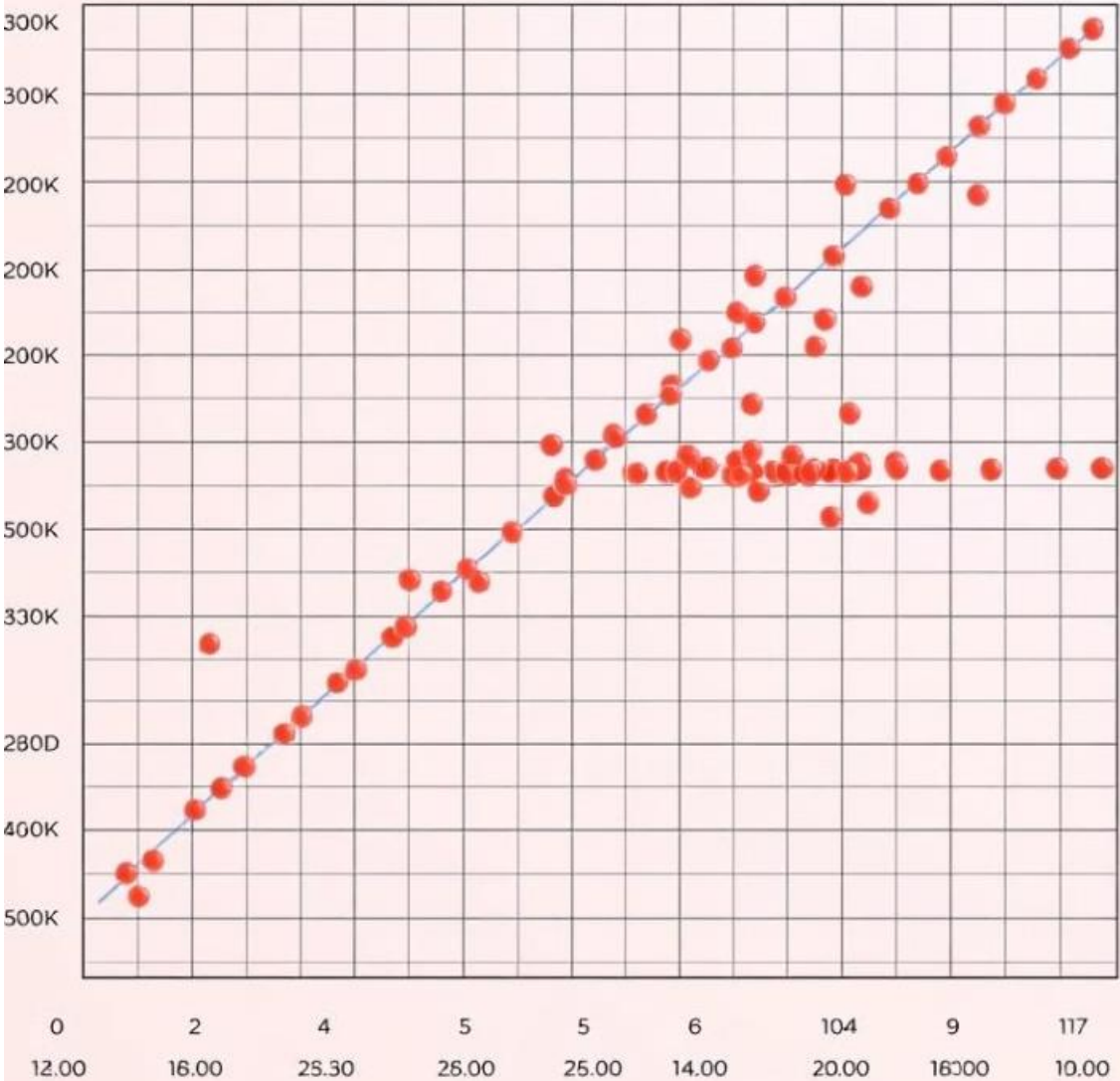
1

Correlation Matrix Plot

- A correlation matrix plot shows correlation coefficients between pairs of variables.
- Correlation coefficients range from -1 to 1. Values near 1 indicate a strong positive relationship.
- Values near -1 indicate a strong negative relationship.
- Values close to 0 indicate a weak or no relationship.

Correlation Matrix Plots: Unveiling Relationships Between Variables

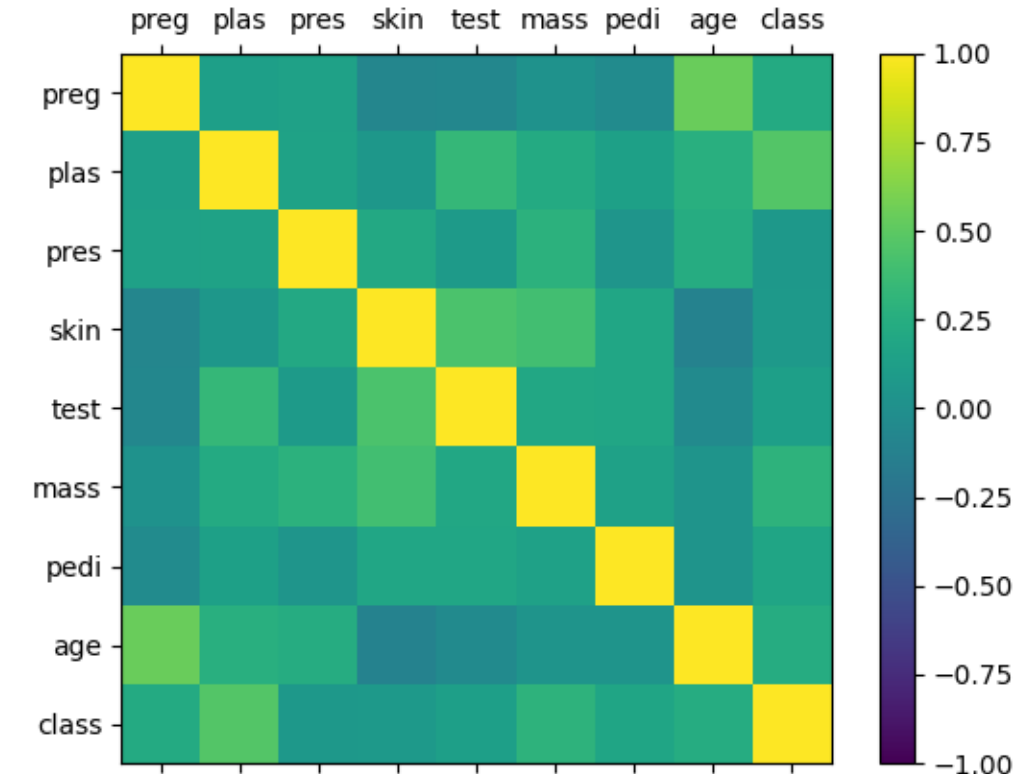
Variable 1	Variable 2	Correlation
Age	Income	0.75
Height	Weight	0.82
Temperature	Humidity	-0.53



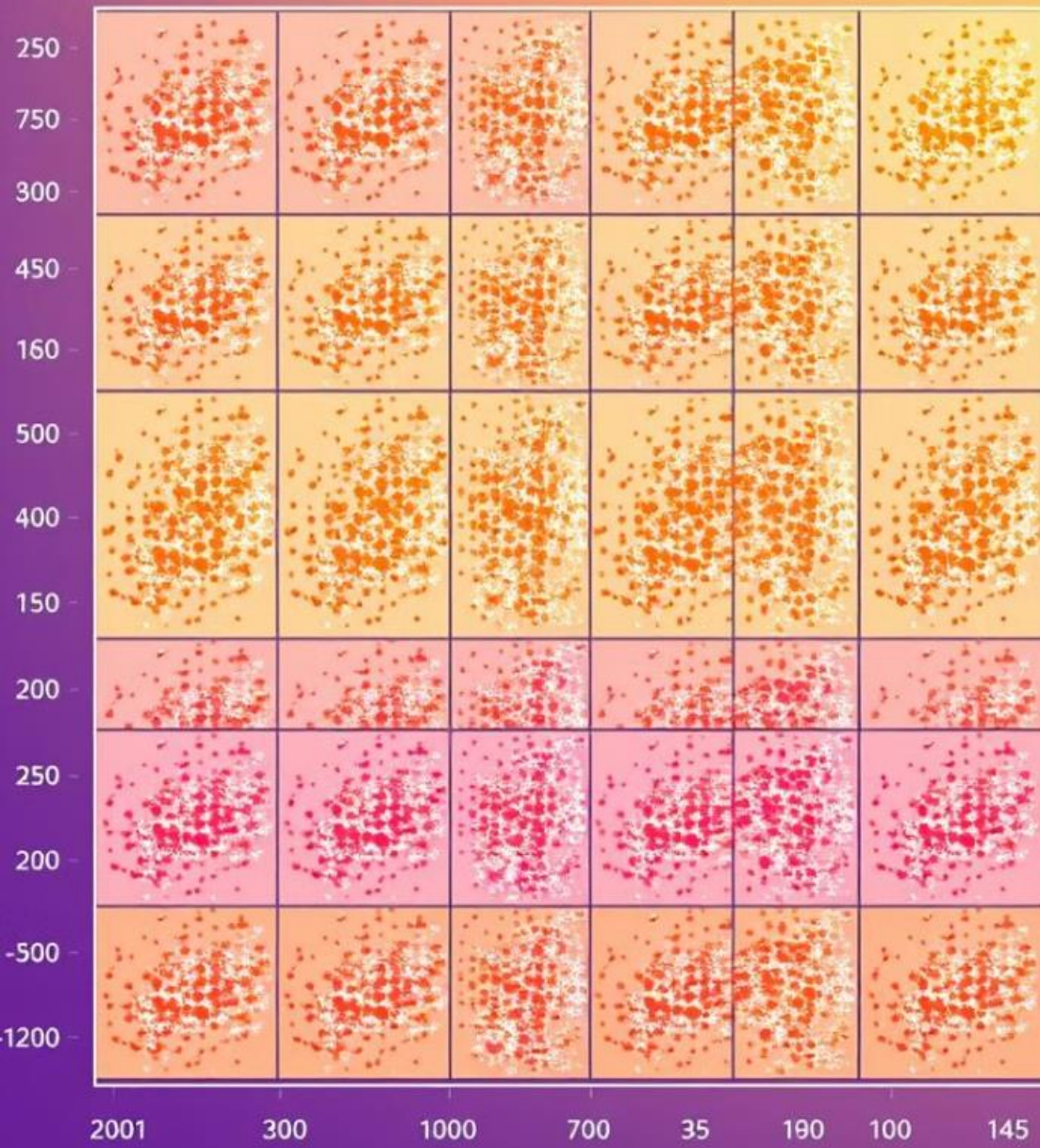
Correlation Matrix Plots:

```
from matplotlib import pyplot
from pandas import read_csv
import numpy
path = "pima-indians-diabetes.csv"
names = ['preg', 'plas', 'pres', 'skin', 'test', 'mass',
         'pedi', 'age', 'class']
```

```
data = read_csv(path, names=names, header=0)
correlations = data.corr()
fig = pyplot.figure()
ax = fig.add_subplot(111)
cax = ax.matshow(correlations, vmin=-1, vmax=1)
fig.colorbar(cax)
ticks = numpy.arange(0, 9, 1)
ax.set_xticks(ticks)
ax.set_yticks(ticks)
ax.set_xticklabels(names)
ax.set_yticklabels(names)
pyplot.show()
```



scatter Matix



Multivariate Visualization: Exploring Relationships Between Attributes

2

Scatter Matrix Plot

- A scatter matrix plot extends the correlation matrix by showing scatter plots for each pair of variables.
- Each plot in the grid displays the relationship between two variables.
- This visualization offers a visual representation of relationships between all variable pairs.
- It allows for quick identification of patterns and trends.

Scatter Matrix Plots: Visualizing Multivariate Relationships

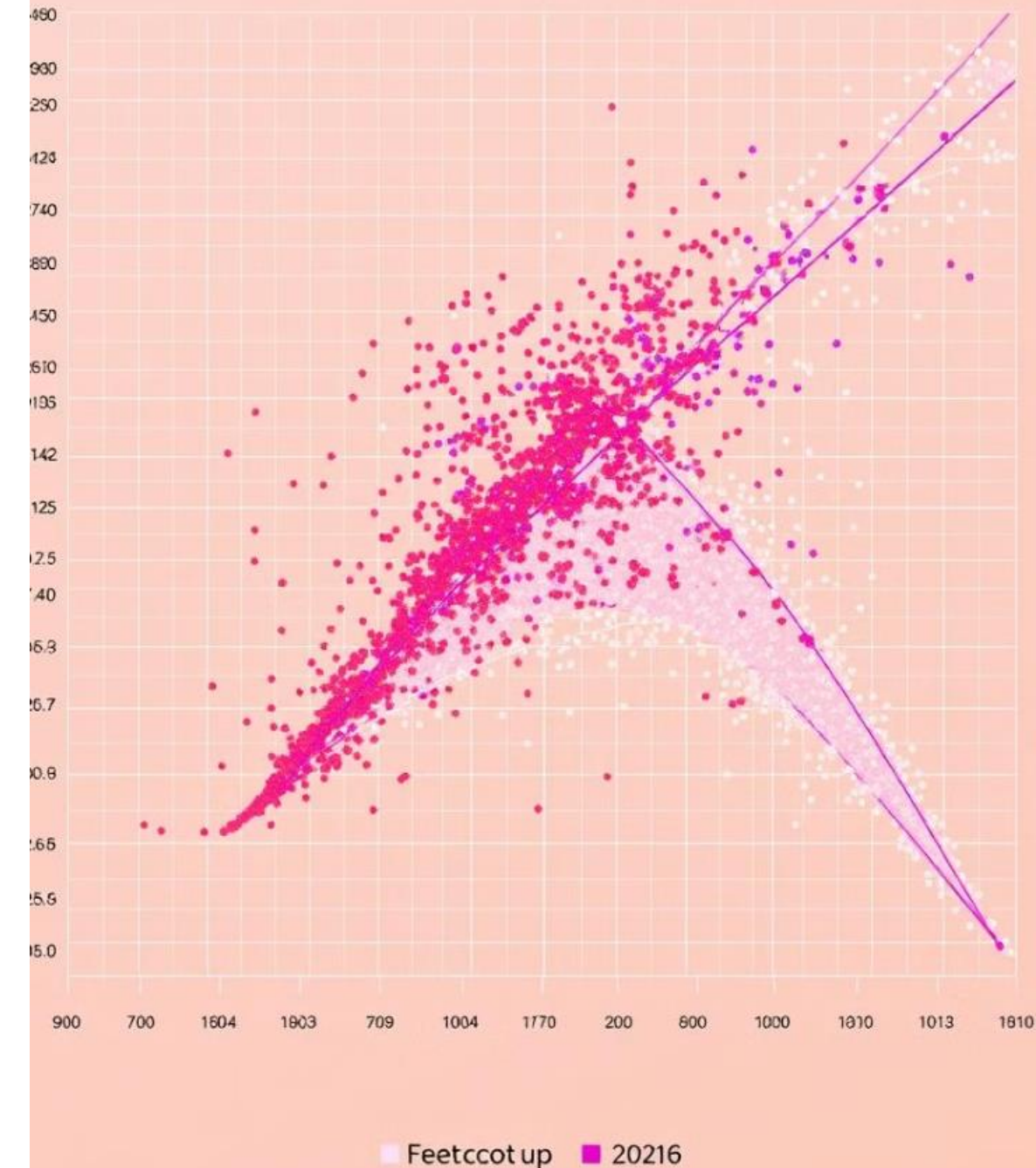
🔍 Identifying Trends

- Scatter plots បង្ហាញគំរូដែលមើលឃើញ ដែលបង្ហាញពីទំនាក់ទំនងរវាងអថេរ។
- ជាឧទាហរណ៍ ទំនាក់ទំនងលីនេអ៊ែរវិជ្ជមាននឹងអាចមើលឃើញជាបន្ទាត់ដែលចំណោតឡើងលើ ខណៈពេលដែលទំនាក់ទំនងលីនេអ៊ែរអវិជ្ជមាននឹងបង្ហាញជាបន្ទាត់ដែលចំណោតចុះក្រោម។

📊 Detecting Outliers

- Outliers ជាចំណុចទិន្នន័យដែលធ្លាក់ចុះយ៉ាងខ្លាំងពីទិន្នន័យភាគច្រើន ឬអាចត្រូវបានសម្គាល់ថាជាចំណុចដែលនៅដាច់ពីគេ ឬស្ថិតនៅខាងក្រៅនិន្នាការ (Trends) ទូទៅនៃ scatter plots។
- ចំនុច outliers នេះអាចត្រូវបានស៊ើបអង្កេតបន្ថែមទៀតដើម្បីយល់ពីមូលហេតុ និងផលប៉ះពាល់សក្តានុពលរបស់ពួកគេទៅលើការវិភាគ។

Scatter Matrix



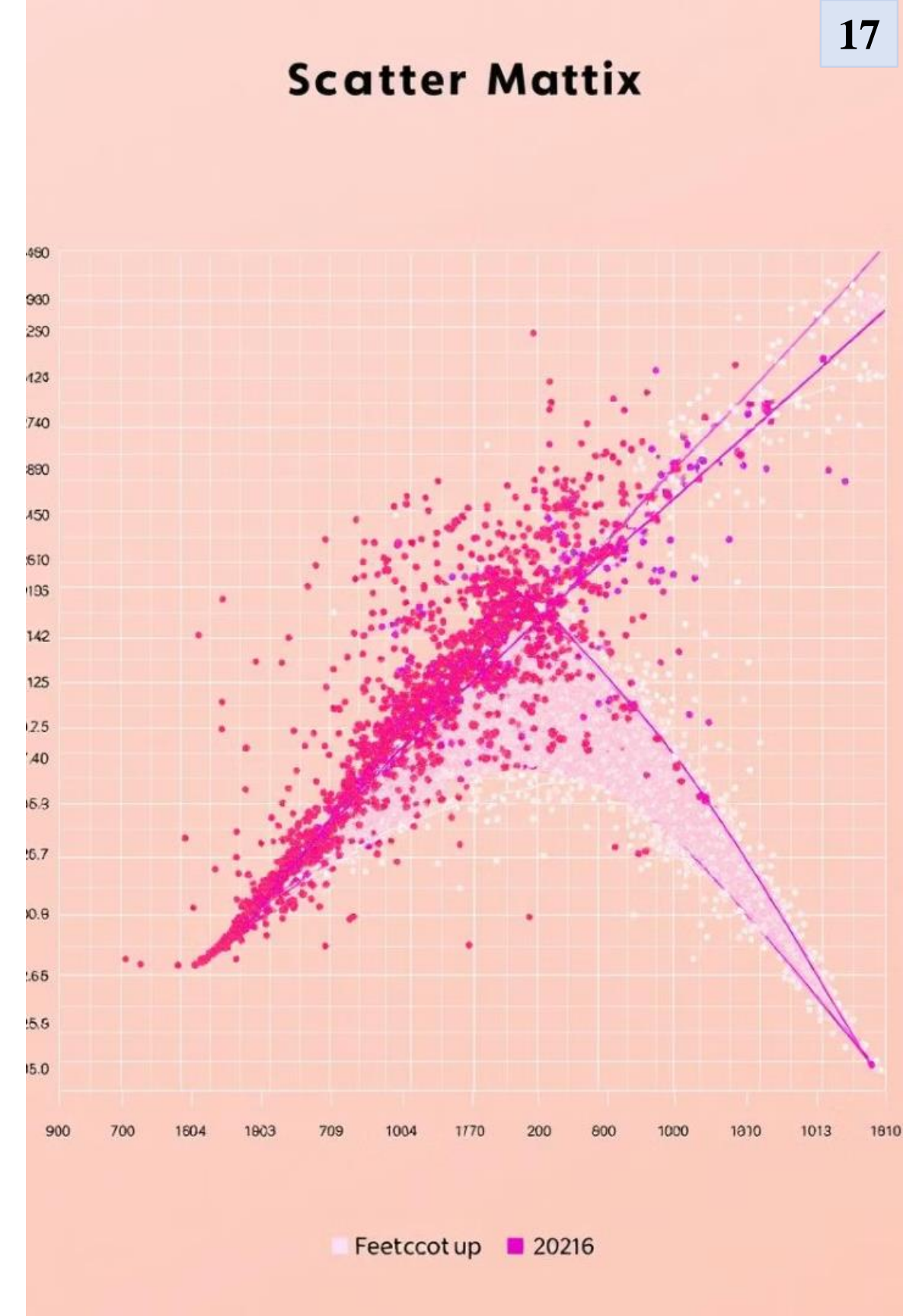
Scatter Matrix Plots: Visualizing Multivariate Relationships

Revealing Clusters

- Scatter matrix plots អាចបង្ហាញពីចង្កោម (clusters) នៃចំណុចទិន្នន័យ។
- ចង្កោមណែនាំការដាក់ជាក្រុមនៅក្នុងទិន្នន័យ។
- ការដាក់ជាក្រុមទាំងនេះអាចបង្ហាញពីក្រុមរងនៅក្នុងចំនួន Population ឬទំនាក់ទំនងផ្សេងគ្នារវាងអថេរ។

? Exploring Non-Linear Relationships

- Scatter matrix plots ត្រូវបានប្រើជាចម្បងដើម្បីកំណត់ទំនាក់ទំនងលីនេអ៊ែរ។
- ពួកគេក៏អាចជួយរកឃើញទំនាក់ទំនងមិនលីនេអ៊ែររវាងអថេរ។
- ទំនាក់ទំនងដែលមិនមែនជាលីនេអ៊ែរអាចលេចឡើងជាខ្សែកោង ឬលំនាំផ្សេងទៀតនៅ scatter plots ។

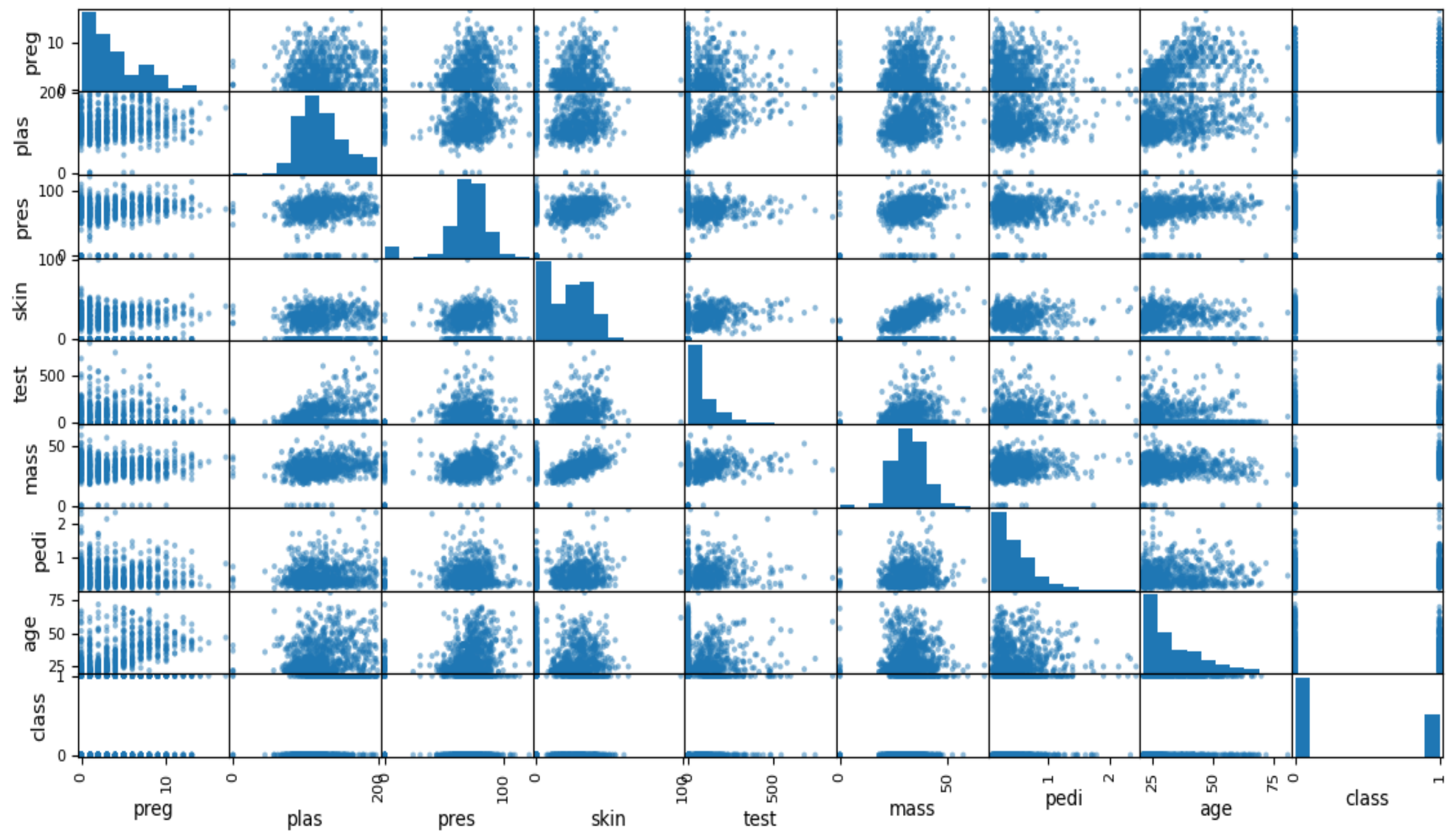


Scatter Matrix Plot

```
from matplotlib import pyplot
from pandas import read_csv
from pandas.plotting import scatter_matrix
path = "pima-indians-diabetes.csv"
names = ['preg', 'plas', 'pres', 'skin', 'test', 'mass',
         'pedi', 'age', 'class']

data = read_csv(path, names=names, header=0)
scatter_matrix(data)
pyplot.show()
```

Scatter Matrix Plot



Data Visualization: A Powerful Tool for Machine Learning

- Data visualization គឺជា tool ដ៏មានសារៈសំខាន់ក្នុងម៉ាស៊ីនស្វ័យសិក្សា។
- វាជួយឱ្យយល់ពីទិន្នន័យ កំណត់អត្តសញ្ញាណគំរូ និងវិភាគបញ្ហាគំរូ។
- Visualization ផ្តល់នូវការយល់ដឹងសម្រាប់ feature engineering, model selection, និង decision-making។
- Visualization ប្រកបដោយប្រសិទ្ធភាពគឺមានសារៈសំខាន់សម្រាប់ការកសាងគំរូម៉ាស៊ីនសិក្សាដ៏រឹងមាំ និងអាចទុកចិត្តបាន។

