

# Practices of Business Intelligence

## 描述性分析 I :

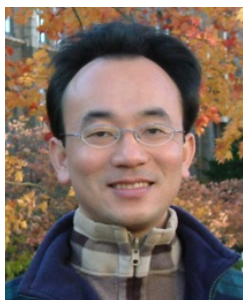
### 數據的性質、統計模型與可視化

#### (Descriptive Analytics I: Nature of Data, Statistical Modeling, and Visualization)

1071BI04

MI4 (M2084) (2888)

Wed, 7, 8 (14:10-16:00) (B217)



Min-Yuh Day

戴敏育

Assistant Professor

專任助理教授

Dept. of Information Management, Tamkang University

淡江大學 資訊管理學系

<http://mail.tku.edu.tw/myday/>

2018-10-03



# 課程大綱 (Syllabus)

- | 週次 (Week) | 日期 (Date)  | 內容 (Subject/Topics)   |
|-----------|------------|---|
| 1         | 2018/09/12 | 商業智慧實務課程介紹<br>(Course Orientation for Practices of Business Intelligence)                                   |
| 2         | 2018/09/19 | 商業智慧、分析與資料科學<br>(Business Intelligence, Analytics, and Data Science)  |
| 3         | 2018/09/26 | 人工智慧、大數據與雲端運算<br>(ABC: AI, Big Data, and Cloud Computing)   |
| 4         | 2018/10/03 | 描述性分析I：數據的性質、統計模型與可視化<br>(Descriptive Analytics I: Nature of Data, Statistical Modeling, and Visualization) |
| 5         | 2018/10/10 | 國慶紀念日 (放假一天) (National Day) (Day off)   |
| 6         | 2018/10/17 | 描述性分析II：商業智慧與資料倉儲<br>(Descriptive Analytics II: Business Intelligence and Data Warehousing)                 |



# 課程大綱 (Syllabus)

- | 週次 (Week) | 日期 (Date)  | 內容 (Subject/Topics)  |
|-----------|------------|--|
| 7         | 2018/10/24 | 預測性分析I：資料探勘流程、方法與演算法<br>(Predictive Analytics I: Data Mining Process, Methods, and Algorithms) |
| 8         | 2018/10/31 | 預測性分析II：文本、網路與社群媒體分析<br>(Predictive Analytics II: Text, Web, and Social Media Analytics)       |
| 9         | 2018/11/07 | 期中報告 (Midterm Project Report)  |
| 10        | 2018/11/14 | 期中考試 (Midterm Exam)  |
| 11        | 2018/11/21 | 處方性分析：最佳化與模擬<br>(Prescriptive Analytics: Optimization and Simulation)                          |
| 12        | 2018/11/28 | 社會網絡分析<br>(Social Network Analysis)  |

# 課程大綱 (Syllabus)

- | 週次 (Week) | 日期 (Date)  | 內容 (Subject/Topics)  |
|-----------|------------|--|
| 13        | 2018/12/05 | 機器學習與深度學習<br>(Machine Learning and Deep Learning)  |
| 14        | 2018/12/12 | 自然語言處理<br>(Natural Language Processing)  |
| 15        | 2018/12/19 | AI交談機器人與對話式商務<br>(AI Chatbots and Conversational Commerce)                               |
| 16        | 2018/12/26 | 商業分析的未來趨勢、隱私與管理考量<br>(Future Trends, Privacy and Managerial Considerations in Analytics) |
| 17        | 2019/01/02 | 期末報告 (Final Project Presentation)  |
| 18        | 2019/01/09 | 期末考試 (Final Exam)  |

# Business Intelligence (BI)

**1** Introduction to BI and Data Science

**②** Descriptive Analytics

**3** Predictive Analytics

**4** Prescriptive Analytics

**5** Big Data Analytics

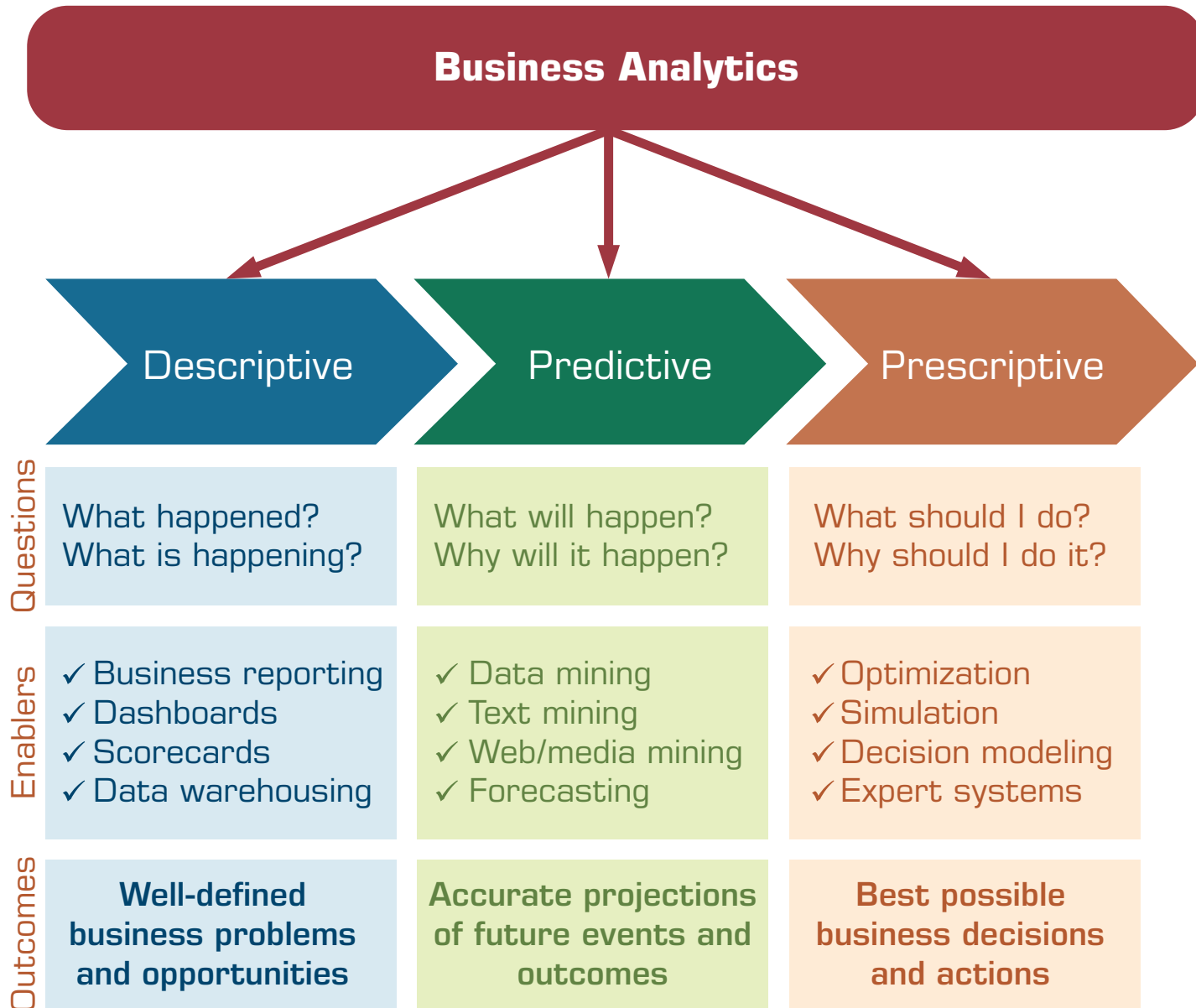
**6** Future Trends

# **Descriptive Analytics I: Nature of Data, Statistical Modeling, and Visualization**

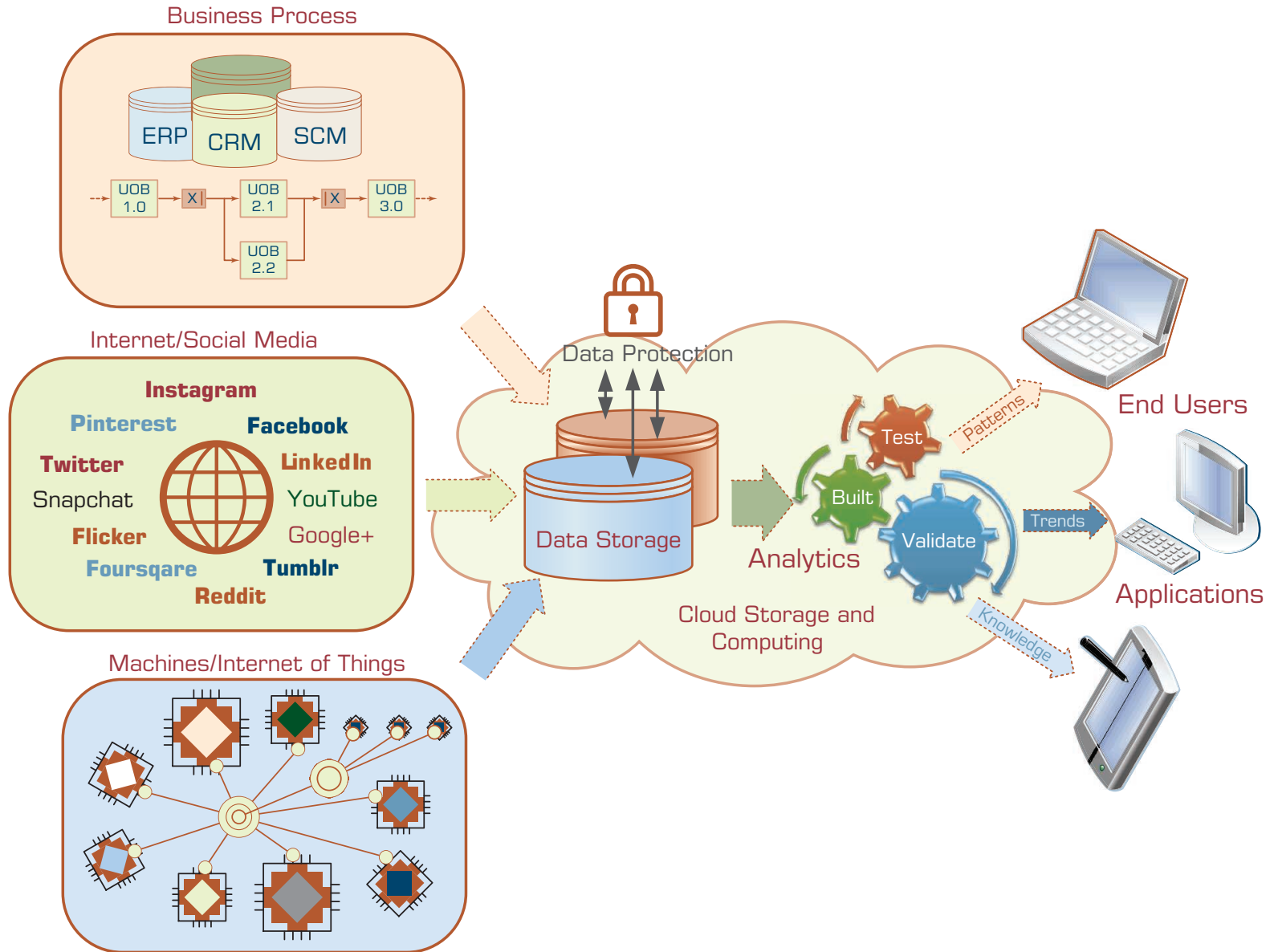
# Outline

- Descriptive Analytics I
- Nature of Data
- Statistical Modeling
- Visualization

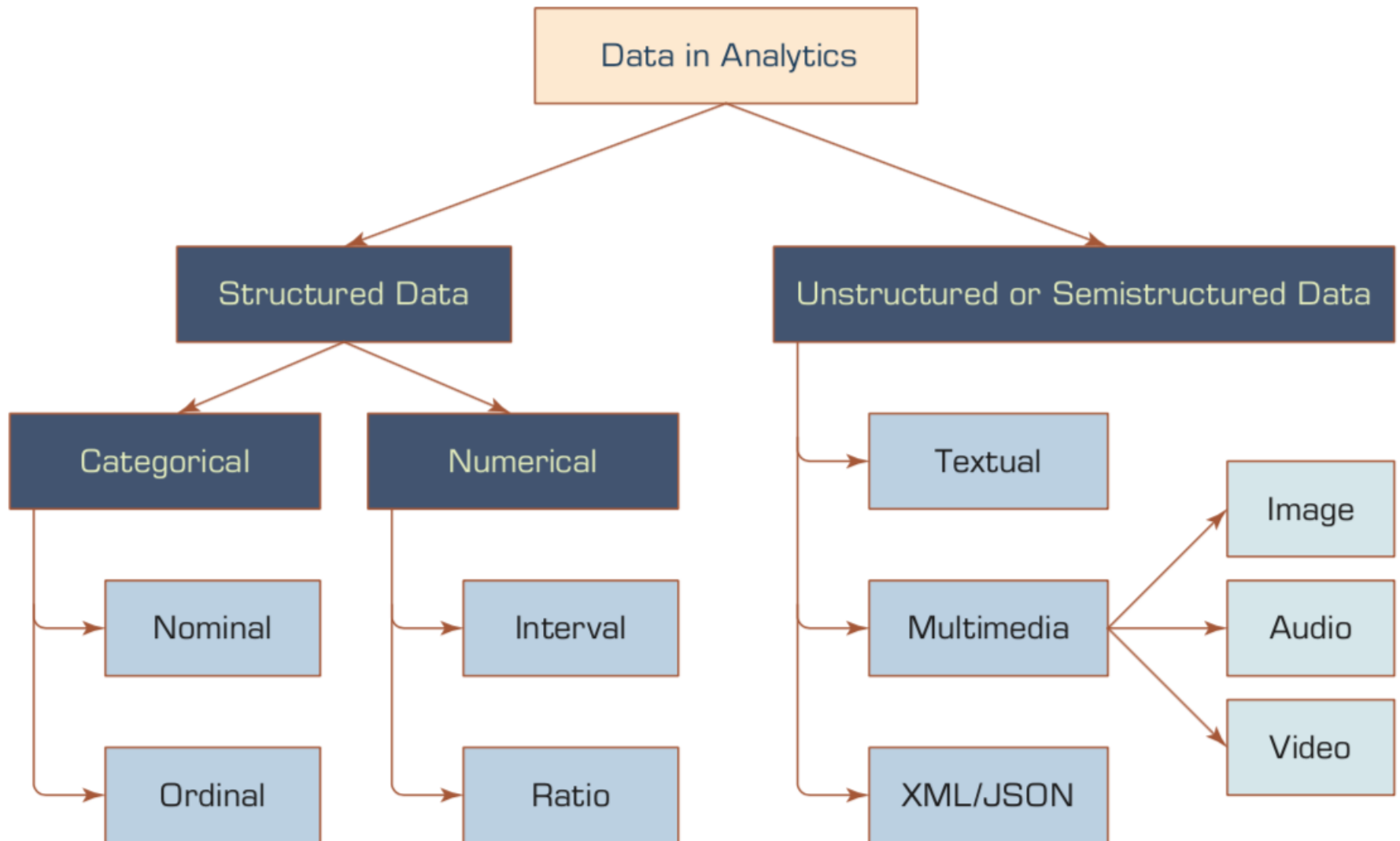
# Three Types of Analytics



# A Data to Knowledge Continuum

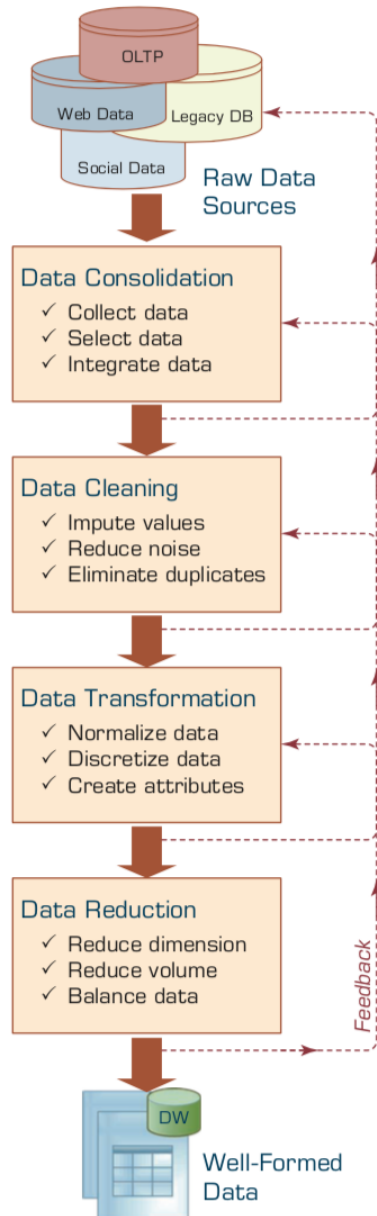


# A Simple Taxonomy of Data

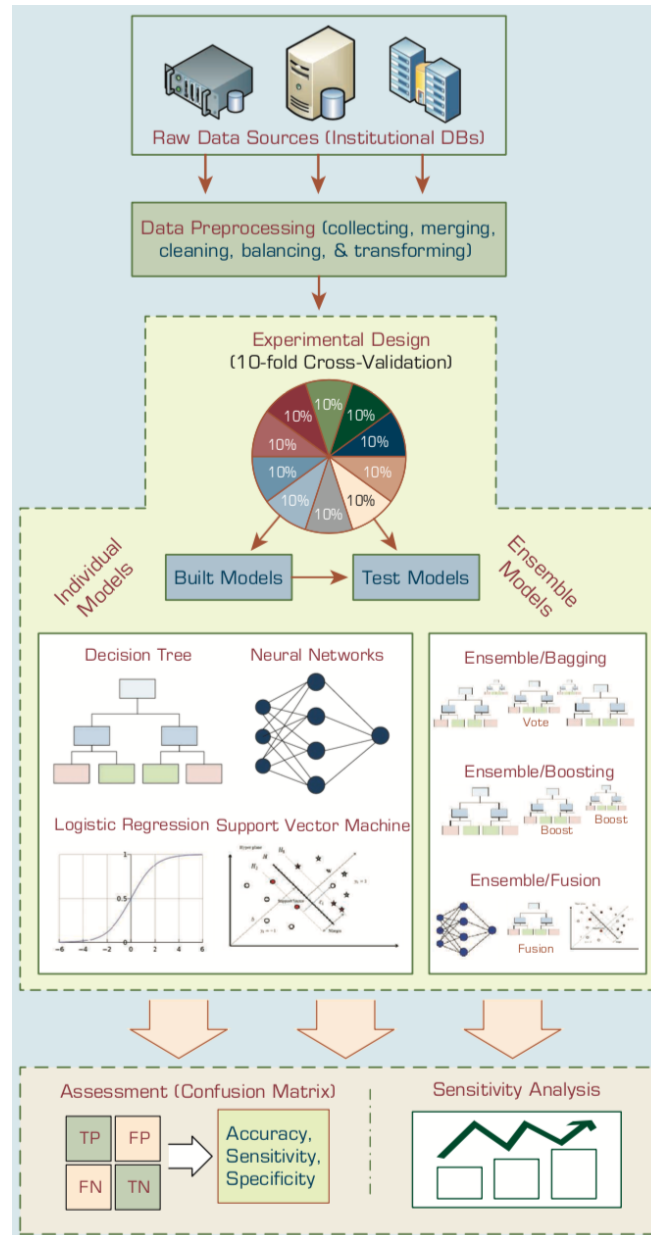




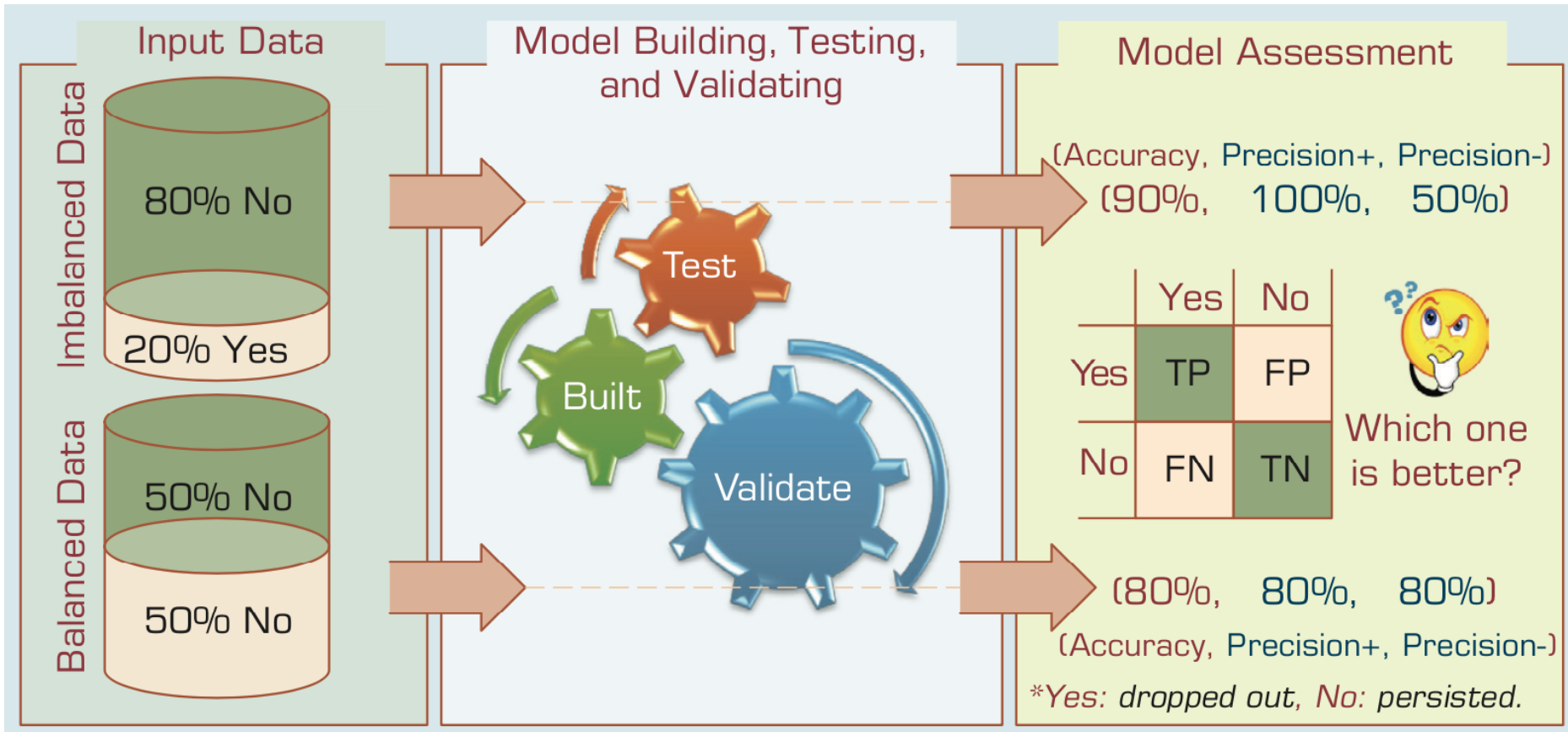
# Data Preprocessing Steps



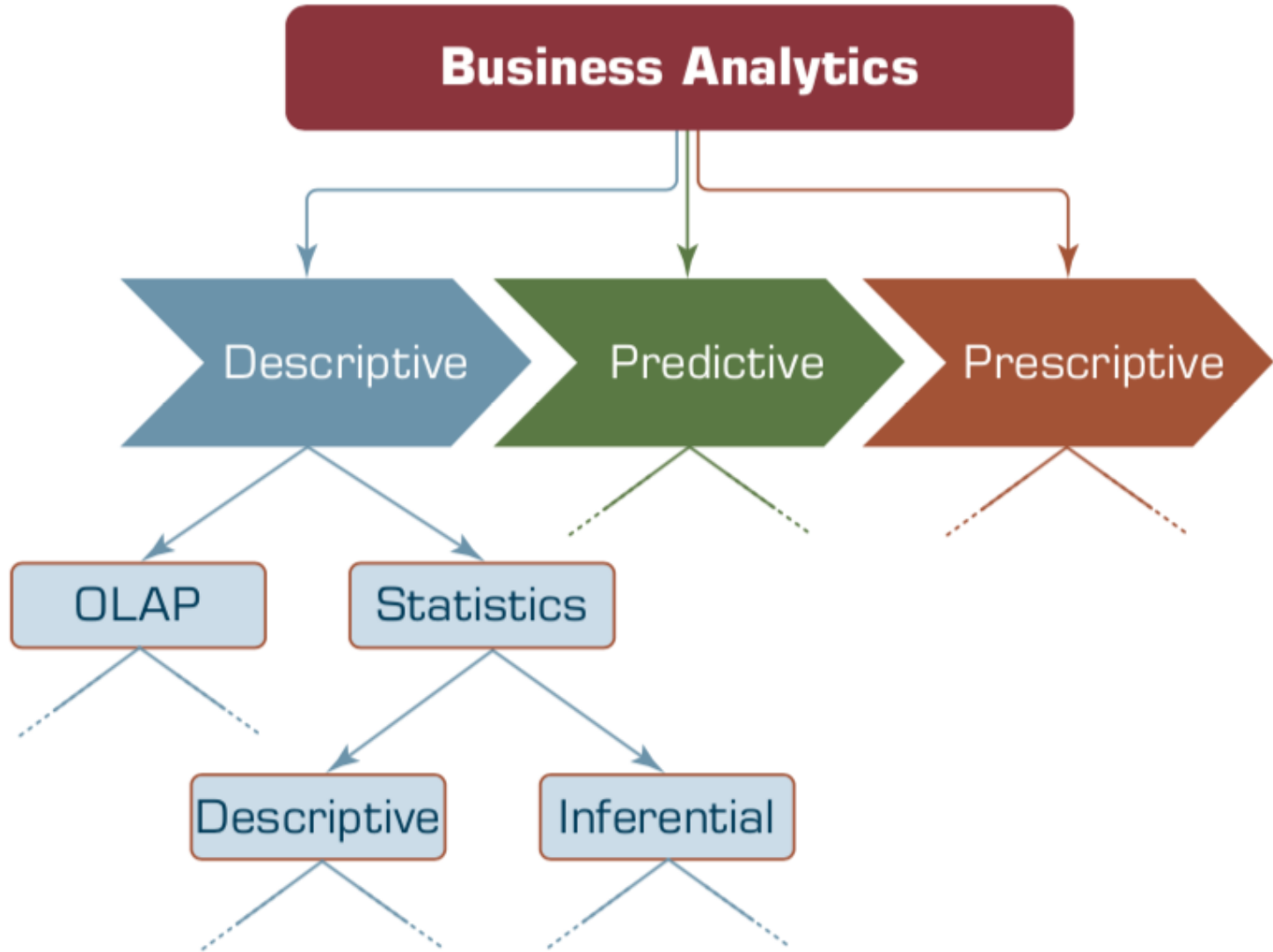
# An Analytics Approach to Predicting Student Attrition



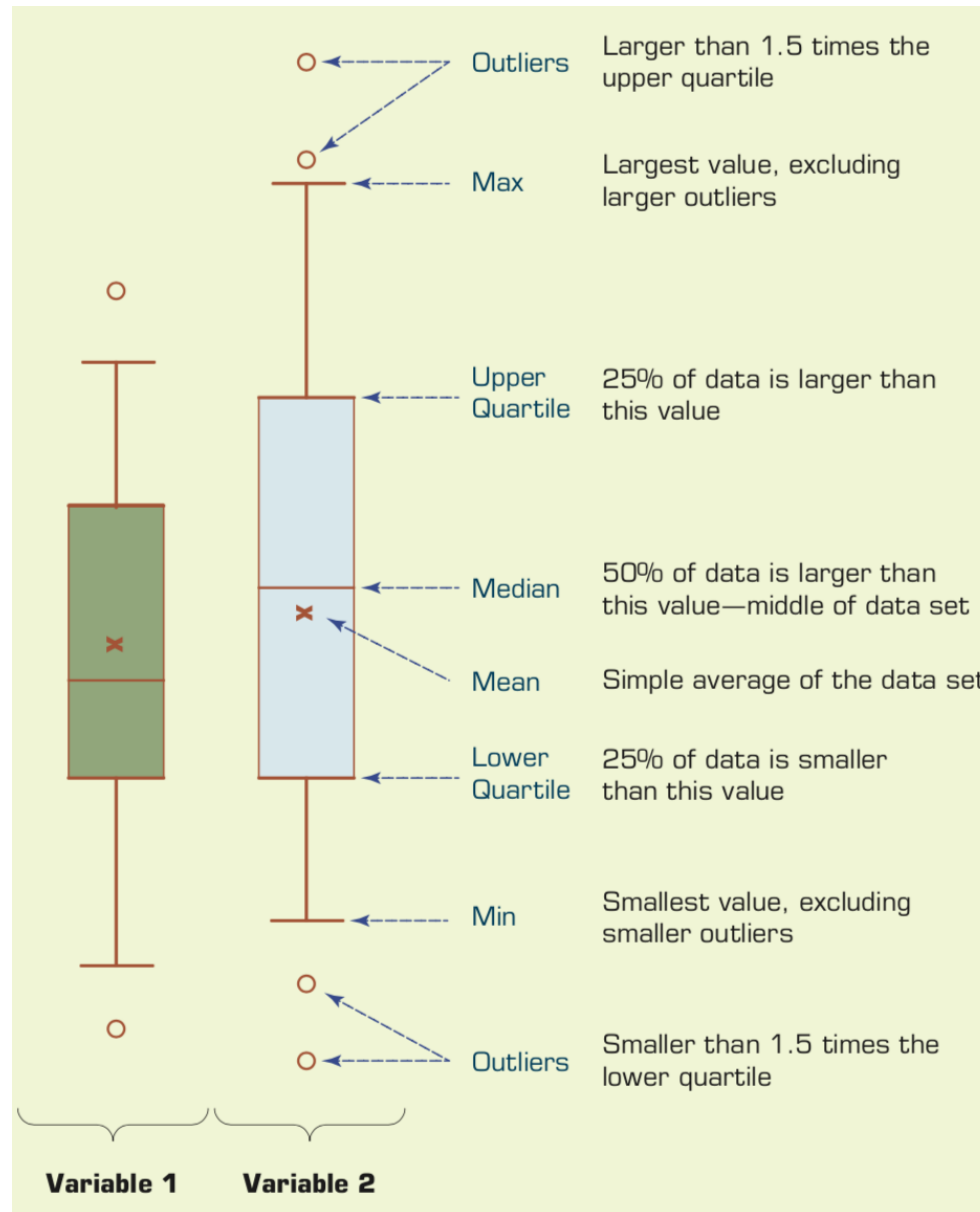
# A Graphical Depiction of the Class Imbalance Problem



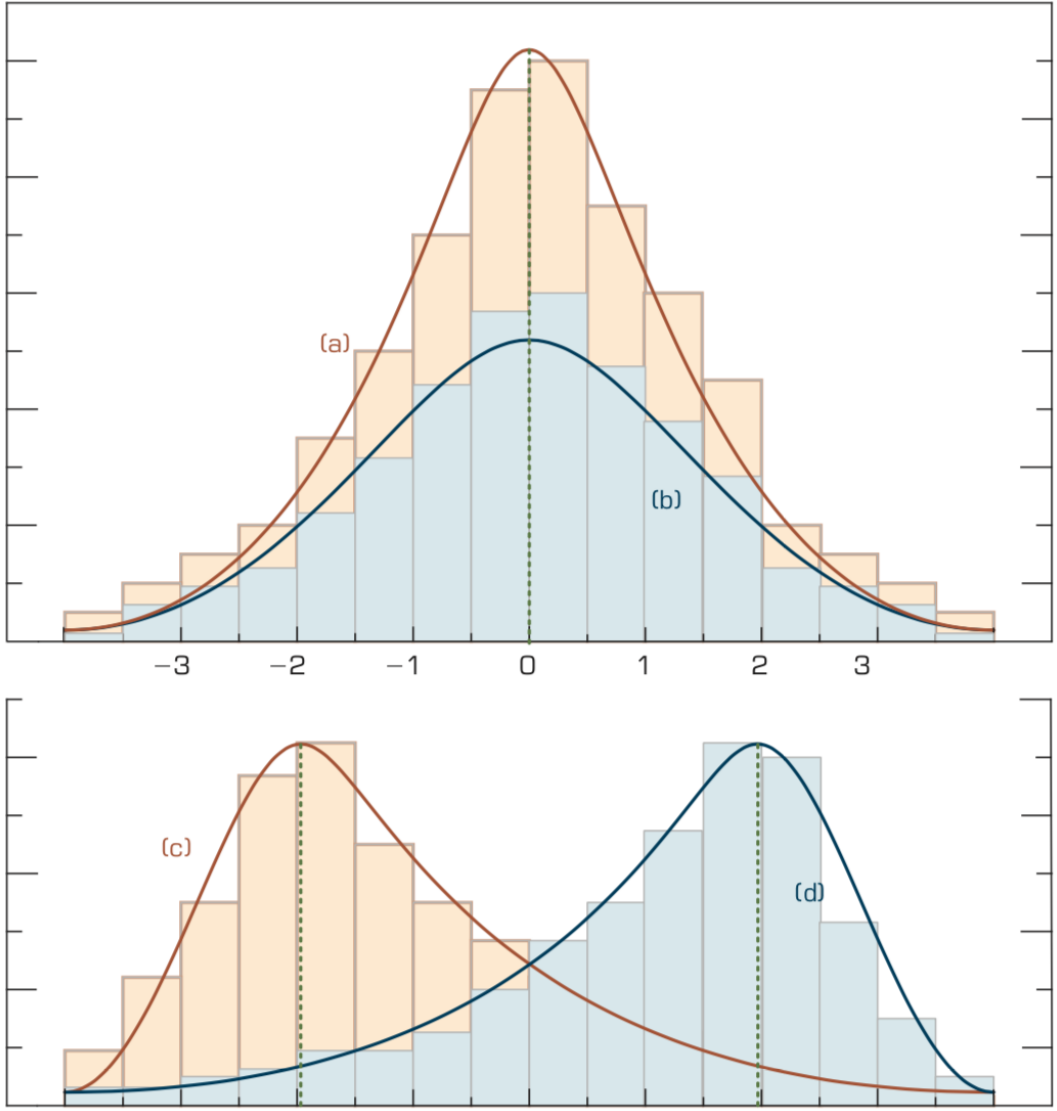
# Relationship between Statistics and Descriptive Analytics



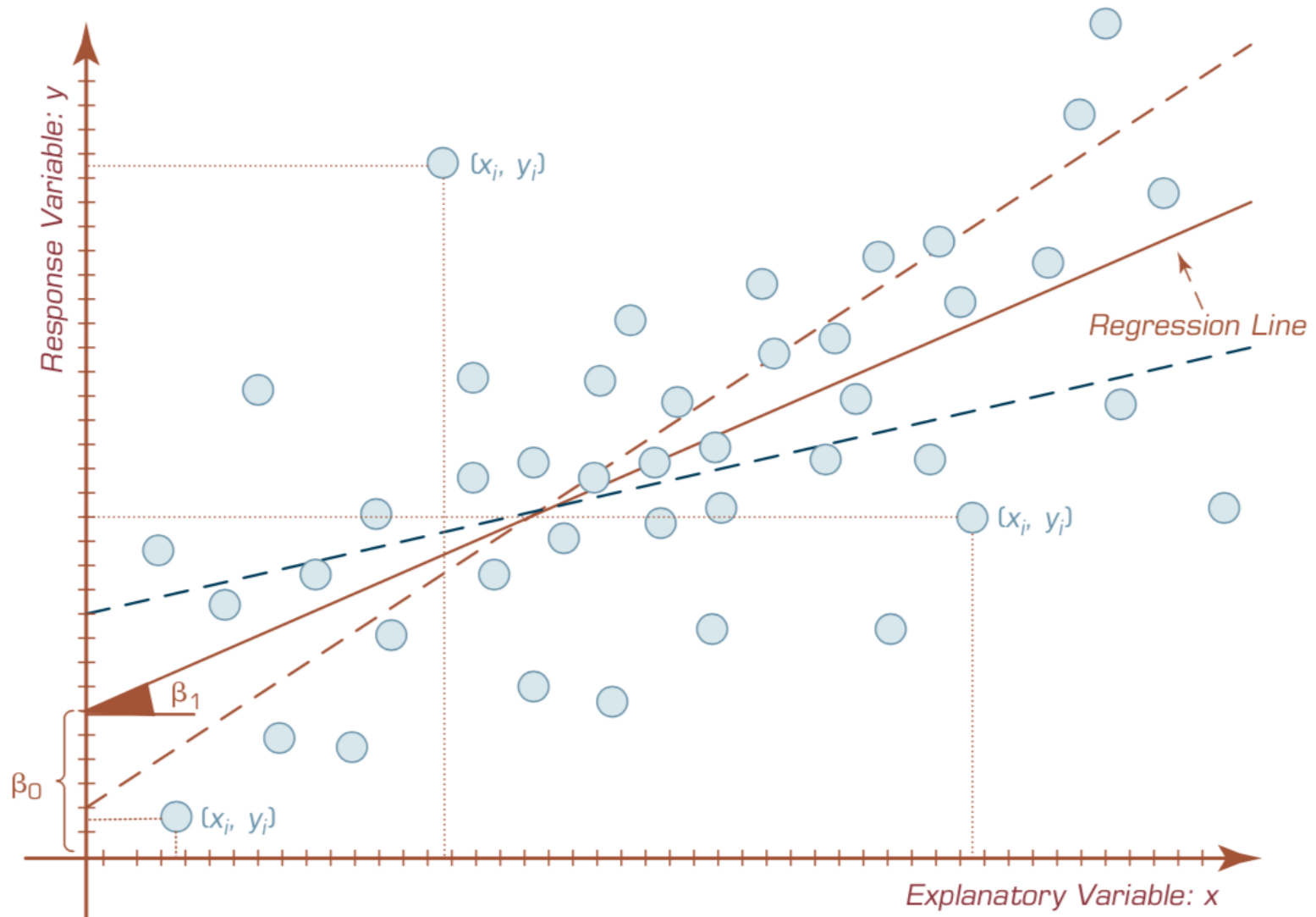
# Understanding the Specifics about Box-and-Whiskers Plots



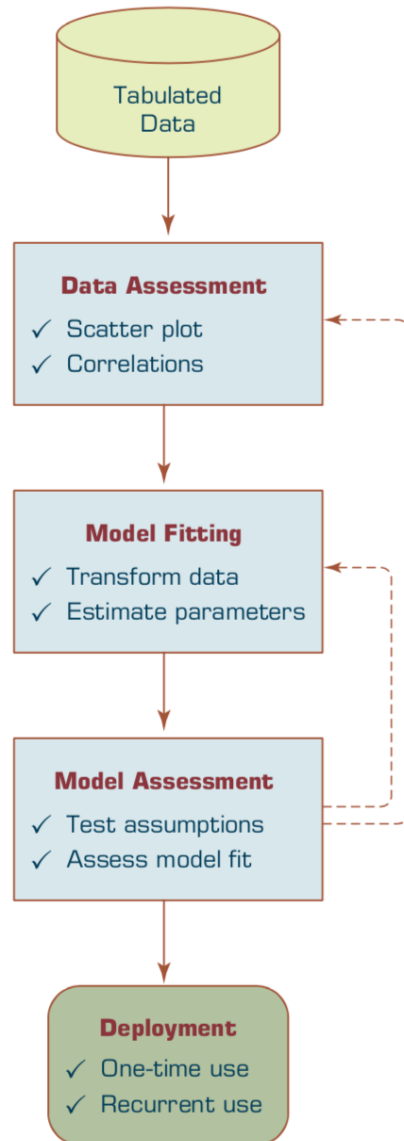
# Relationship between Dispersion and Shape Properties.



# A Scatter Plot and a Linear Regression Line



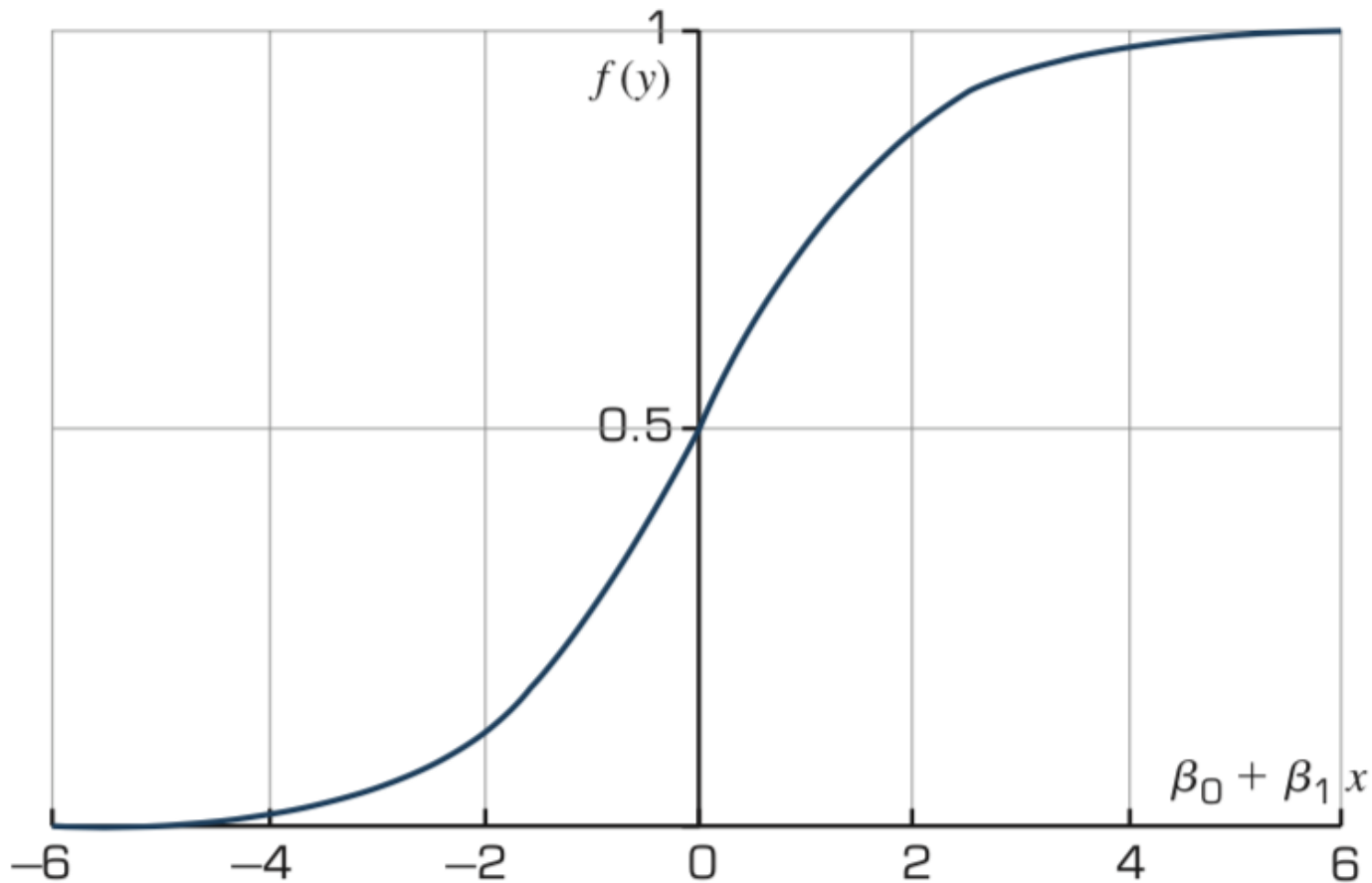
# A Process Flow for Developing Regression Models.



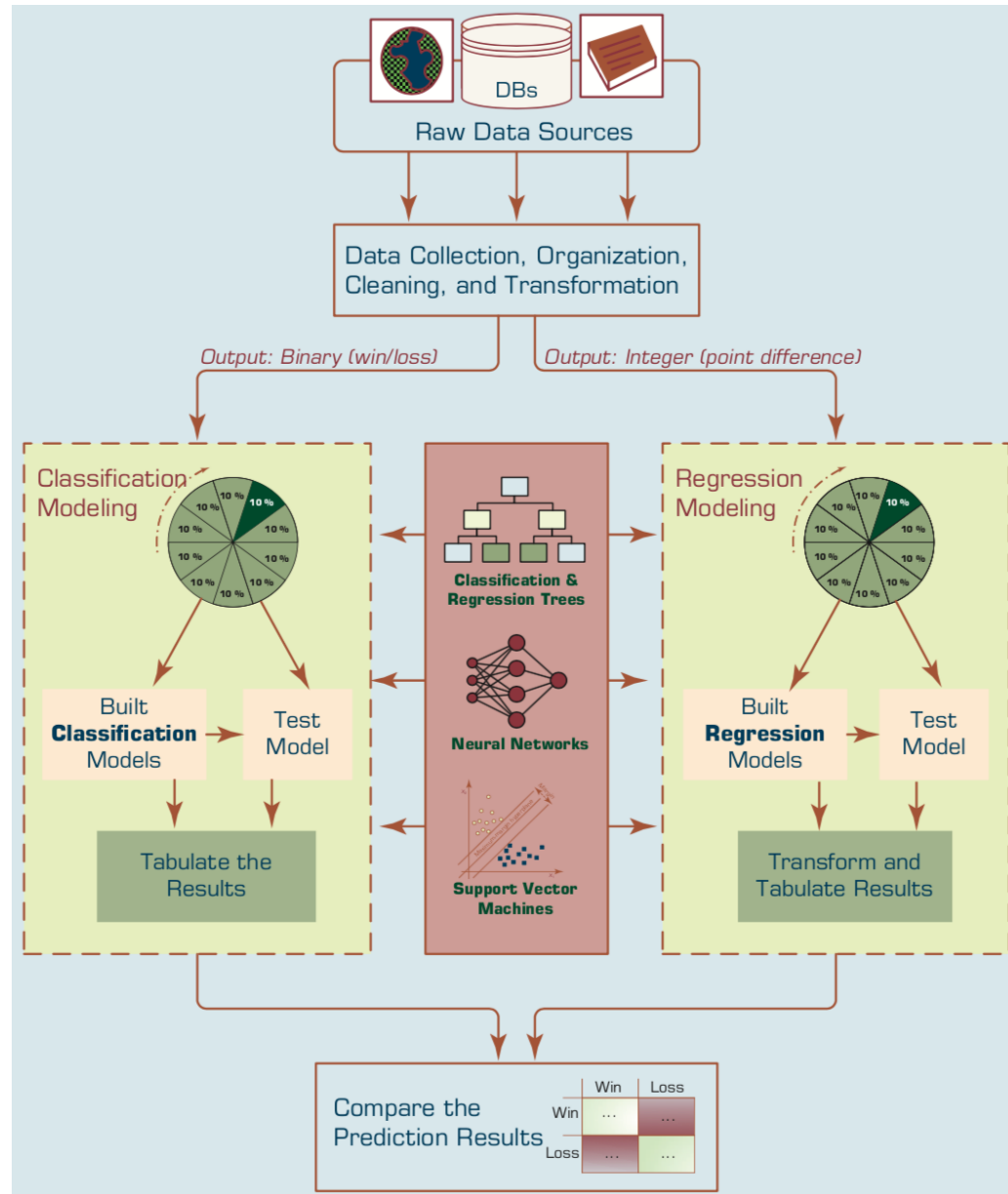


# The Logistic Function

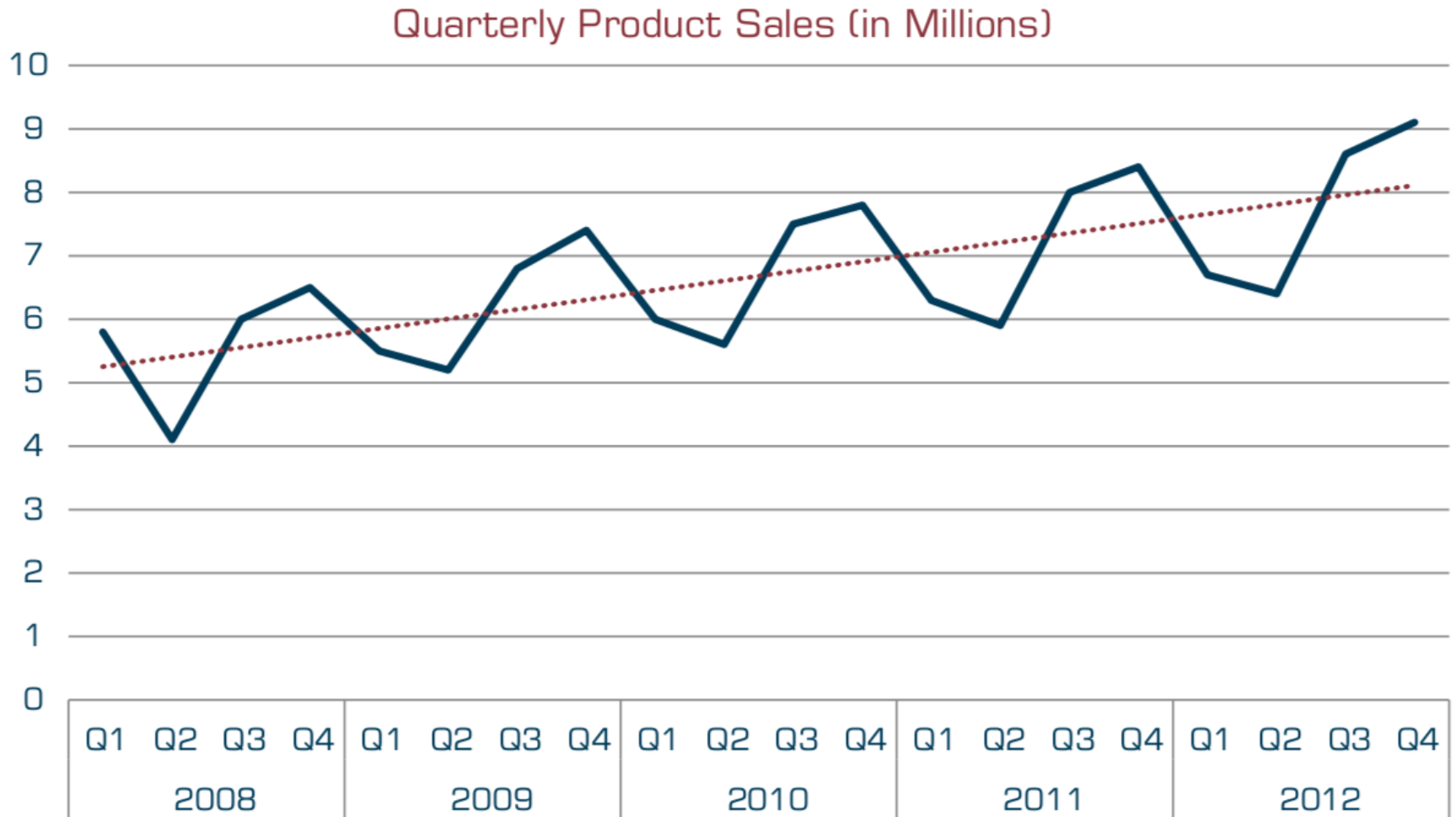
$$f(y) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$



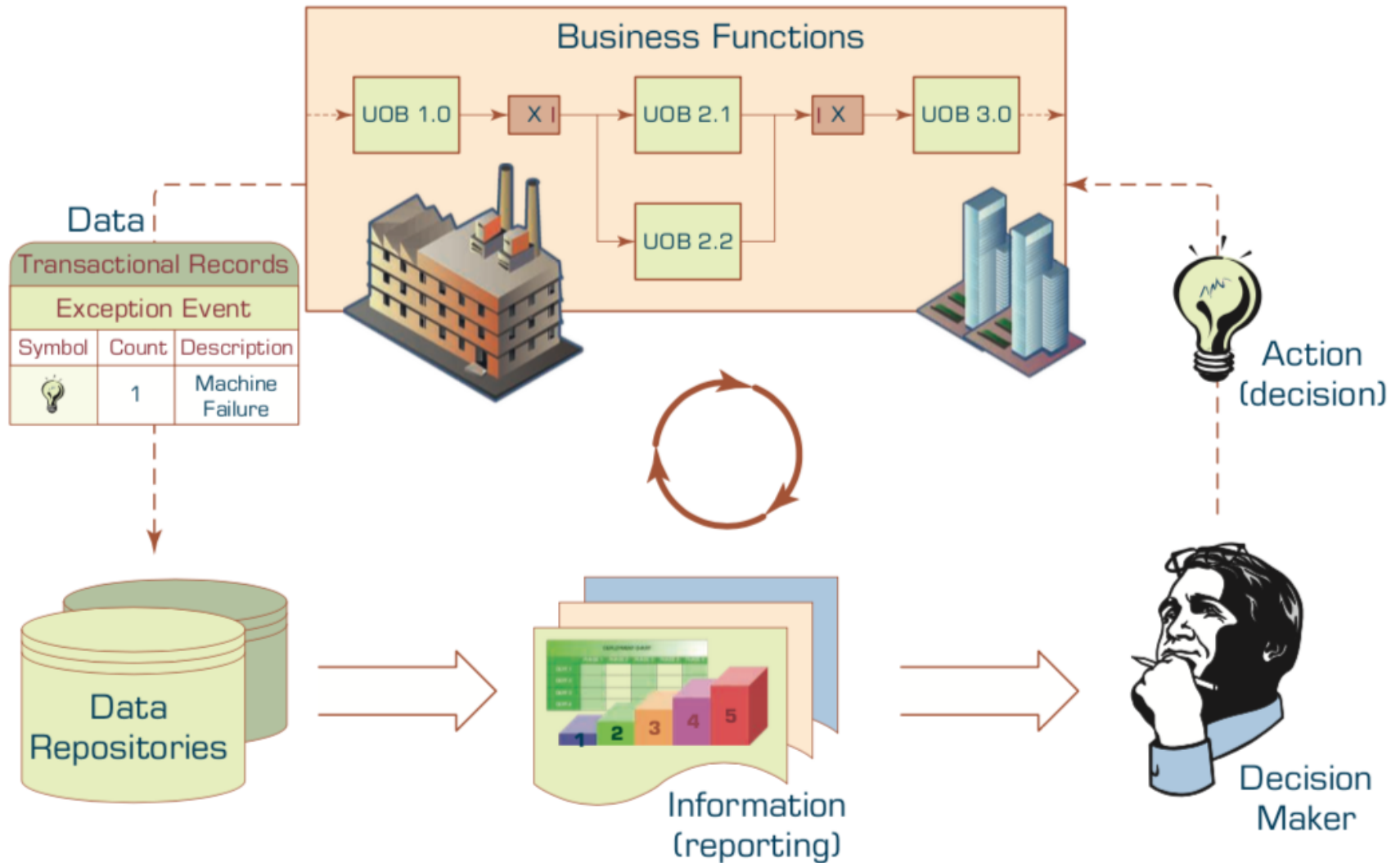
# Predicting NCAA Bowl Game Outcomes



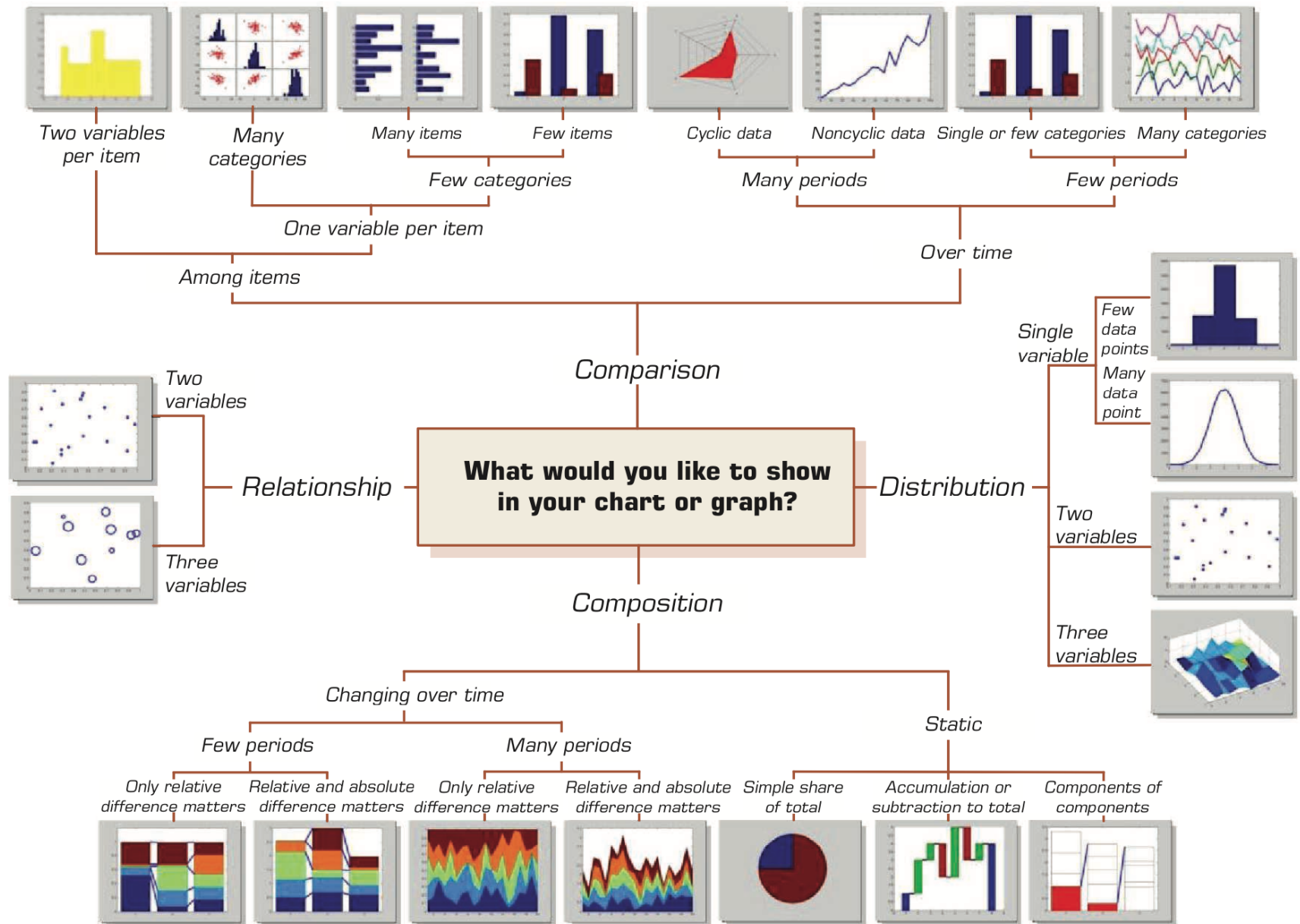
# A Sample Time Series of Data on Quarterly Sales Volumes



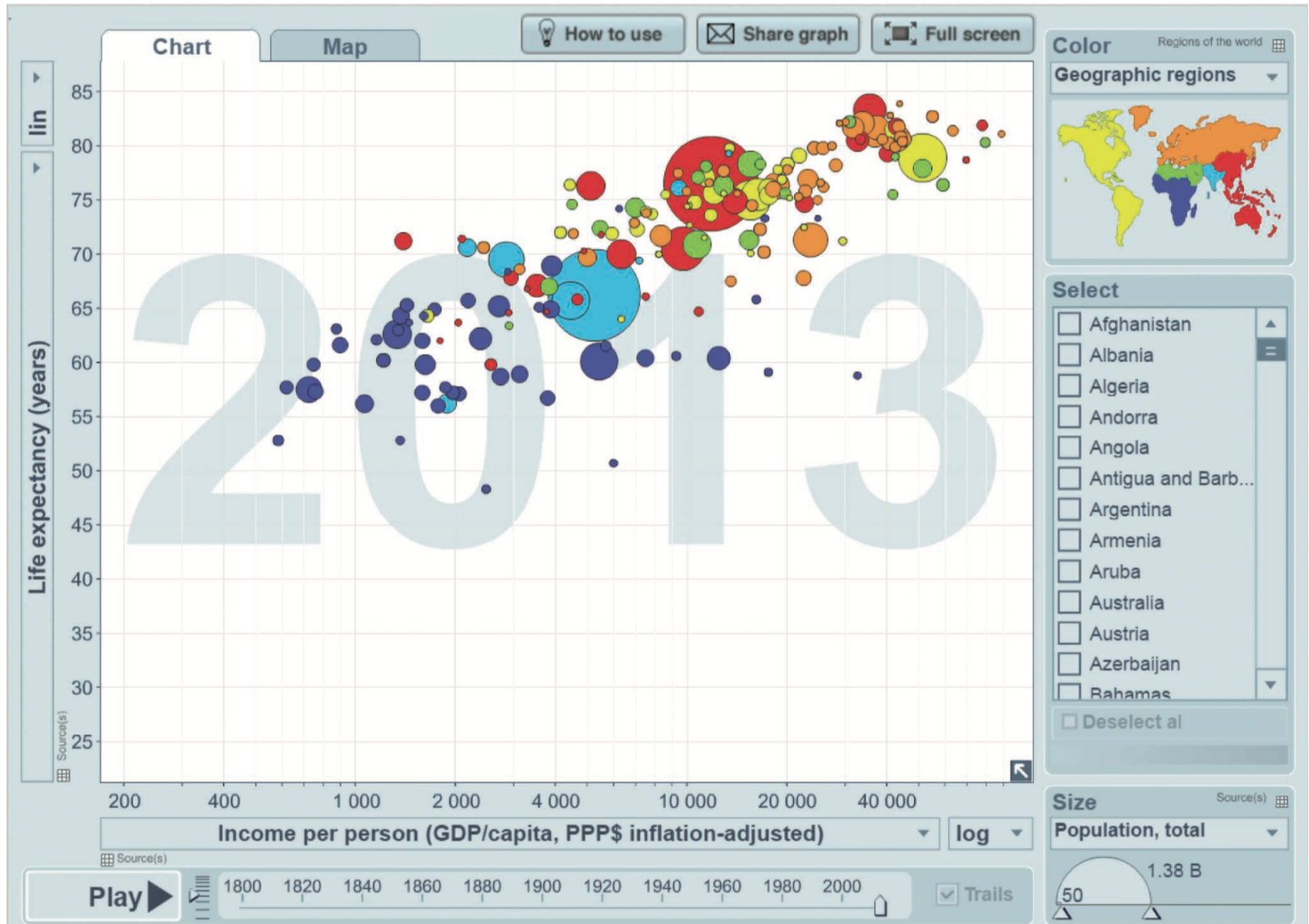
# The Role of Information Reporting in Managerial Decision Making



# A Taxonomy of Charts and Graphs



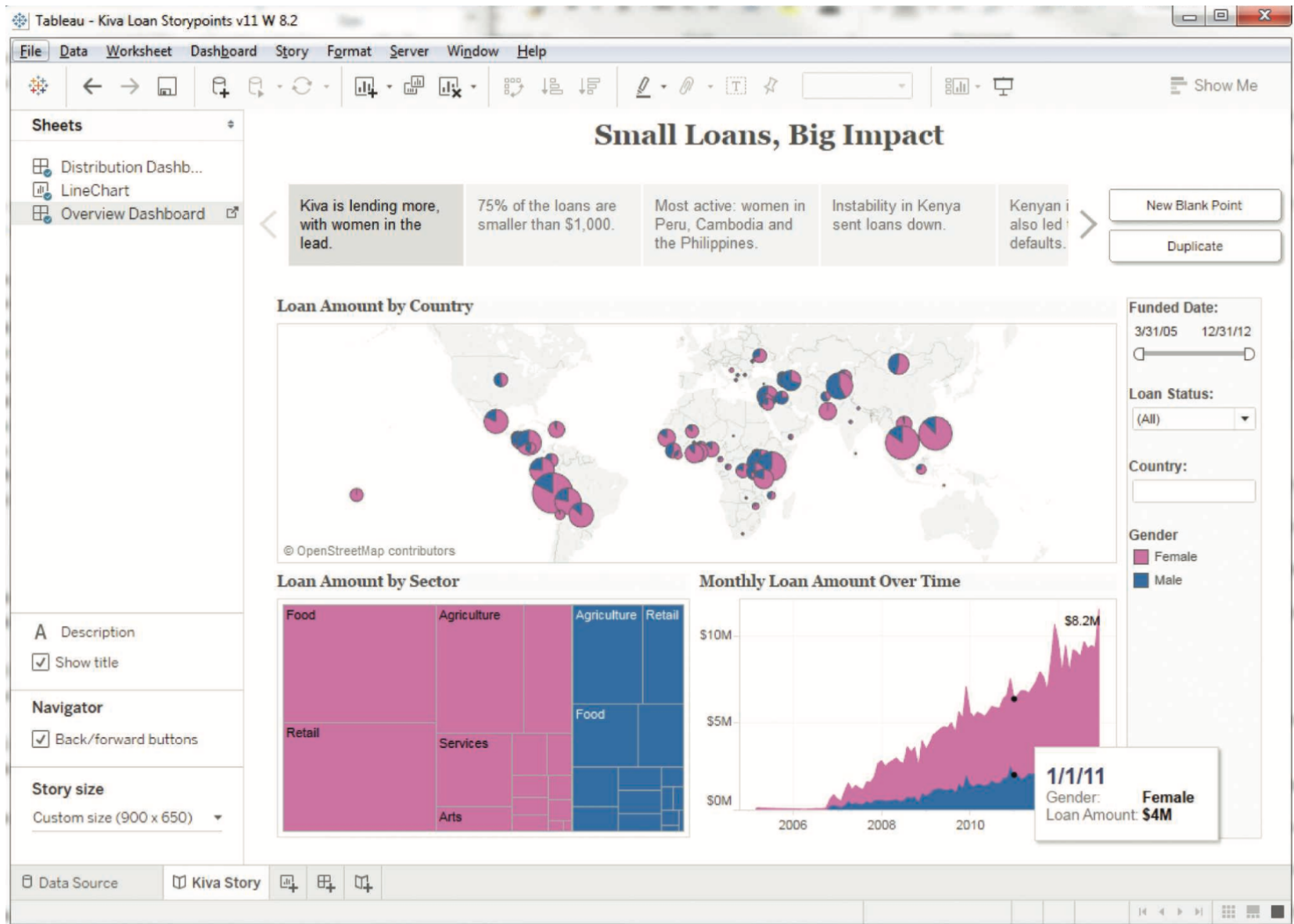
# A Gapminder Chart That Shows the Wealth and Health of Nations



# Magic Quadrant for Business Intelligence and Analytics Platforms



# A Storyline Visualization in Tableau Software





# An Overview of SAS Visual Analytics Architecture

Central Entry Point

Integration

Role-based Views



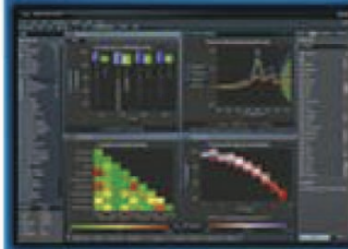
## DATA BUILDER

- Join data from multiple sources
- Create calculated and derived columns
- Load data



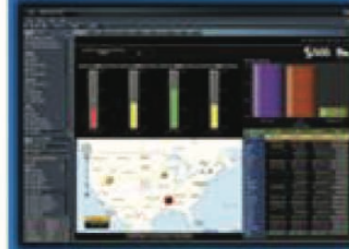
## ADMINISTRATOR

- Monitor SAS® LASR™ Analytic server
- Load/unload data
- Manage security



## EXPLORER

- Perform ad hoc analysis and data discovery
- Apply advanced analytics



## DESIGNER

- Create dashboard style reports for web or mobile

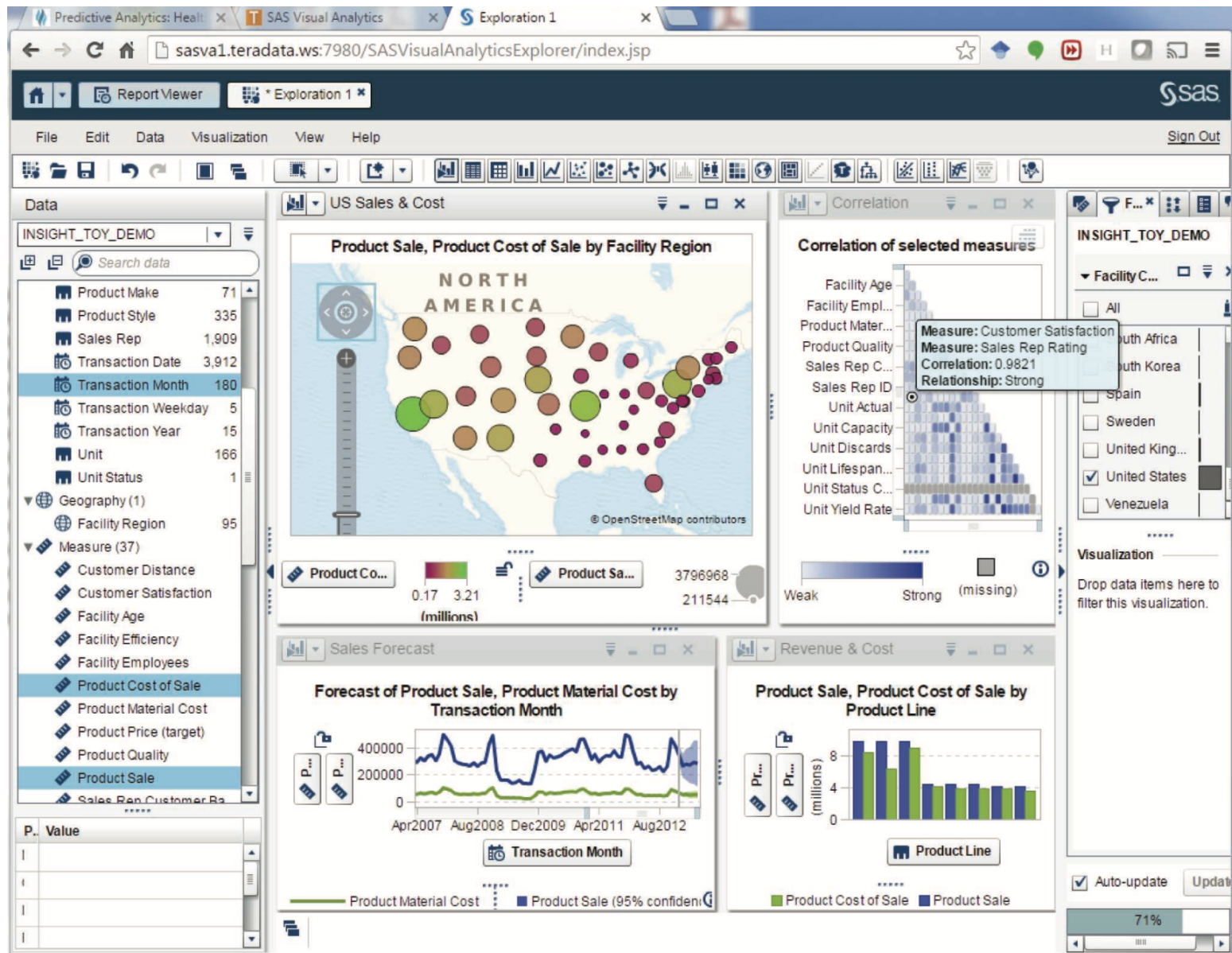


## MOBILE BI

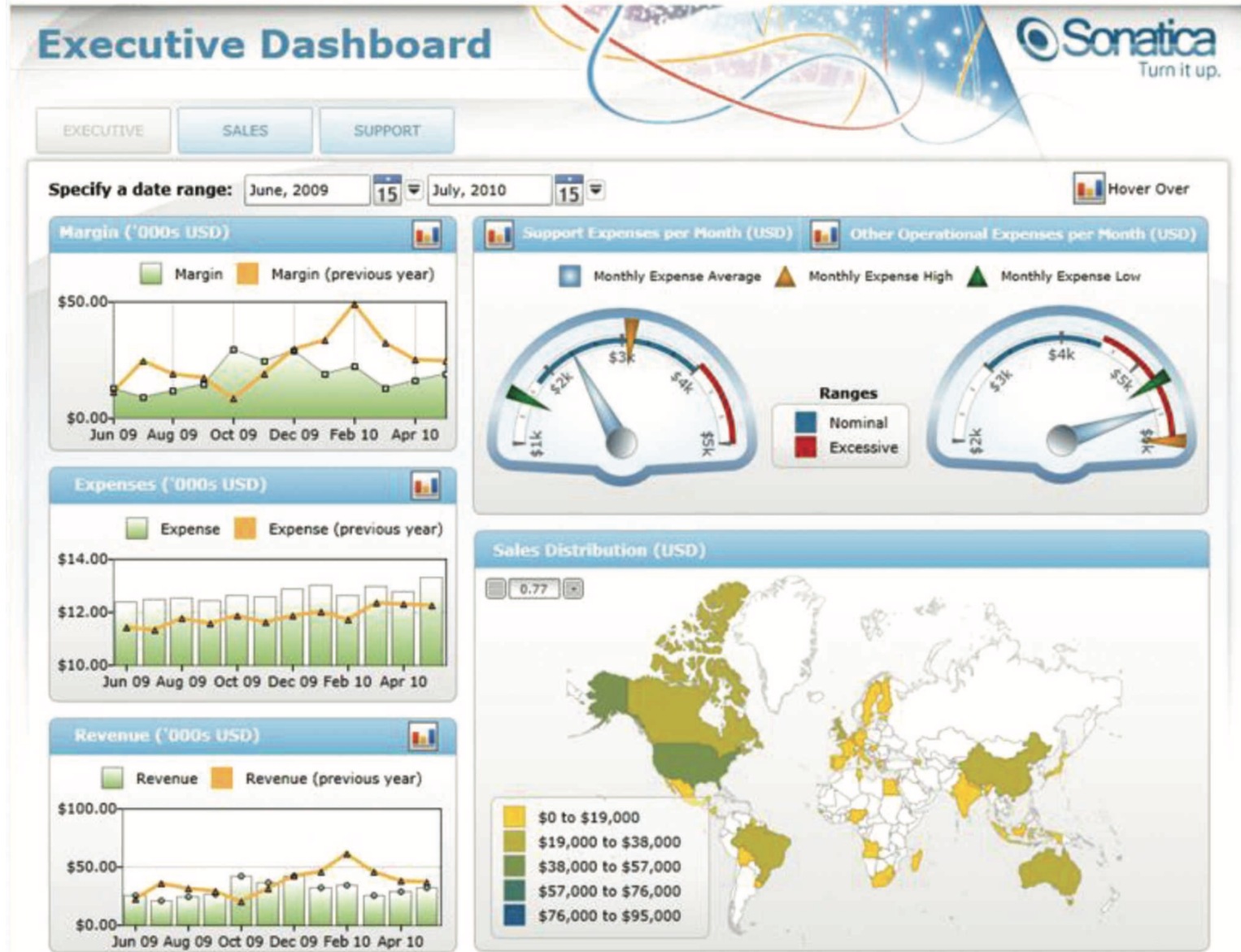
- Native iOS and Android applications that deliver interactive reports

SAS® LASR™ ANALYTIC SERVER

# A Screenshot from SAS Visual Analytics



# A Sample Executive Dashboard





# igraph



## igraph – The network analysis package

igraph is a collection of network analysis tools with the emphasis on **efficiency**, **portability** and ease of use. igraph is **open source** and free. igraph can be programmed in **R**, **Python** and **C/C++**.

igraph R package

python-igraph

igraph C library

R/igraph 1.0.0

Repositories at Github

R/igraph 0.7.1

C/igraph 0.7.1

R/igraph 0.7.0

python-igraph 0.7.0

C/igraph 0.7.0

R/igraph 0.6.5

## Recent news

### R/igraph 1.0.0

June 24, 2015

#### Release Notes

This is a new major release, with a lot of UI changes. We tried to make it easier to use, with short and easy to remember, consistent function names. Unfortunately

<http://igraph.org/redirect.html>

# Gephi



[Download](#) [Blog](#) [Wiki](#) [Forum](#) [Support](#) [Bug tracker](#)

[Home](#) [Features](#) [Learn](#) [Develop](#) [Plugins](#) [Services](#) [Consortium](#)

## The Open Graph Viz Platform

Gephi is the leading visualization and exploration software for all kinds of graphs and networks. Gephi is open-source and free.

Runs on Windows, Mac OS X and Linux.

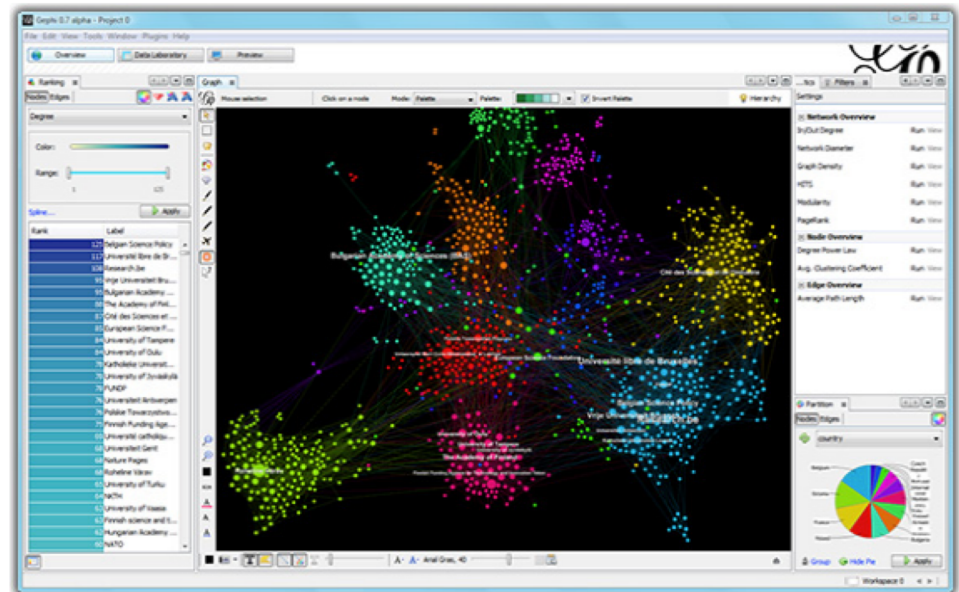
[Learn More on Gephi Platform](#)



[Release Notes](#) | [System Requirements](#)

► [Features](#)  
► [Quick start](#)

► [Screenshots](#)  
► [Videos](#)



Support us! We are **non-profit**. Help us to **innovate** and **empower** the community by donating only **8€**:

Donate



### APPLICATIONS

- ✓ **Exploratory Data Analysis:** intuition-oriented analysis by networks manipulations in real time.
- ✓ **Link Analysis:** revealing the underlying structures of associations between objects.
- ✓ **Social Network Analysis:** easy creation of social

Like Photoshop™ for graphs.

— the Community

### LATEST NEWS

► [Gephi updates with 0.9.1 version](#)

### PAPERS



<https://gephi.org/>

**Discovering,  
Analyzing,  
Visualizing and  
Presenting Data  
with Python  
in Google Colab**

# Google Colab

The screenshot shows the Google Colab web interface. At the top, the browser address bar displays the URL <https://colab.research.google.com/notebooks/welcome.ipynb>. The main header includes the Colab logo, the text "Hello, Colaboratory", and a menu with options: File, Edit, View, Insert, Runtime, Tools, and Help. On the right side of the header, there are "SHARE" and "CONNECT" buttons, along with a user profile icon. Below the header, a toolbar contains buttons for "CODE", "TEXT", "CELL", "COPY TO DRIVE", "CONNECT", and "EDITING".

The left sidebar contains a "Table of contents" with the following items: Getting Started, Highlighted Features, TensorFlow execution, GitHub, Visualization, Forms, Examples, and Local runtime support. A "+ SECTION" button is located at the bottom of the sidebar.

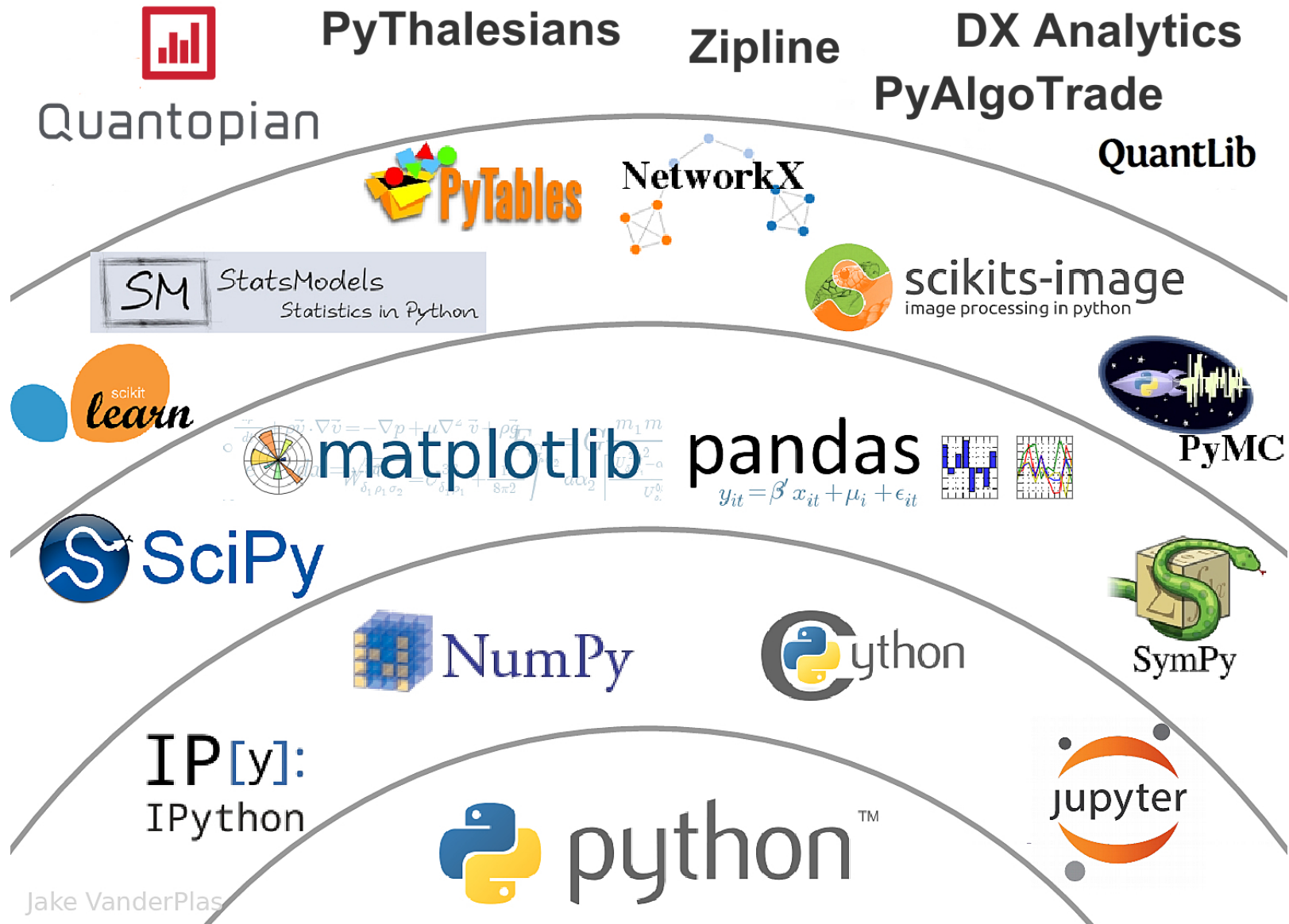
The main content area features a large "Welcome to Colaboratory!" message with the Colab logo and a link to the FAQ. Below this is a "Getting Started" section with a list of links: Overview of Colaboratory, Loading and saving data: Local files, Drive, Sheets, Google Cloud Storage, Importing libraries and installing dependencies, Using Google Cloud BigQuery, Forms, Charts, Markdown, & Widgets, TensorFlow with GPU, and Machine Learning Crash Course: Intro to Pandas & First Steps with TensorFlow.

Underneath is a "Highlighted Features" section with a "Seedbank" subsection. The Seedbank text reads: "Looking for Colab notebooks to learn from? Check out [Seedbank](#), a place to discover interactive machine learning examples."

The "TensorFlow execution" subsection contains the text: "Colaboratory allows you to execute TensorFlow code in your browser with a single click. The example below adds two matrices." followed by the matrix equation:

$$\begin{bmatrix} 1. & 1. & 1. \end{bmatrix} + \begin{bmatrix} 1. & 2. & 3. \end{bmatrix} = \begin{bmatrix} 2. & 3. & 4. \end{bmatrix}$$

# The Quant Finance PyData Stack



Jake VanderPlas

Source: [http://nbviewer.jupyter.org/format/slides/github/quantopian/pyfolio/blob/master/pyfolio/examples/overview\\_slides.ipynb#/5](http://nbviewer.jupyter.org/format/slides/github/quantopian/pyfolio/blob/master/pyfolio/examples/overview_slides.ipynb#/5)



# Python

# matplotlib

# matplotlib

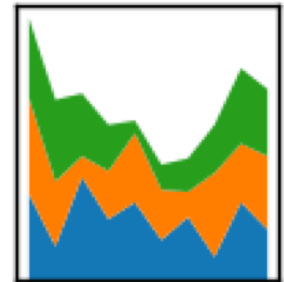
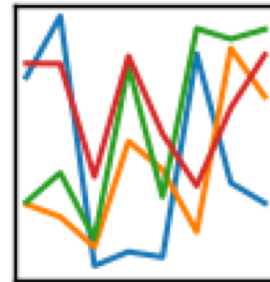
The logo for Matplotlib, which is a circular plot with a grid and several colored wedges (orange, yellow, green, blue) radiating from the center.

# Python

# Pandas

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



# Iris flower data set

**setosa**



**versicolor**



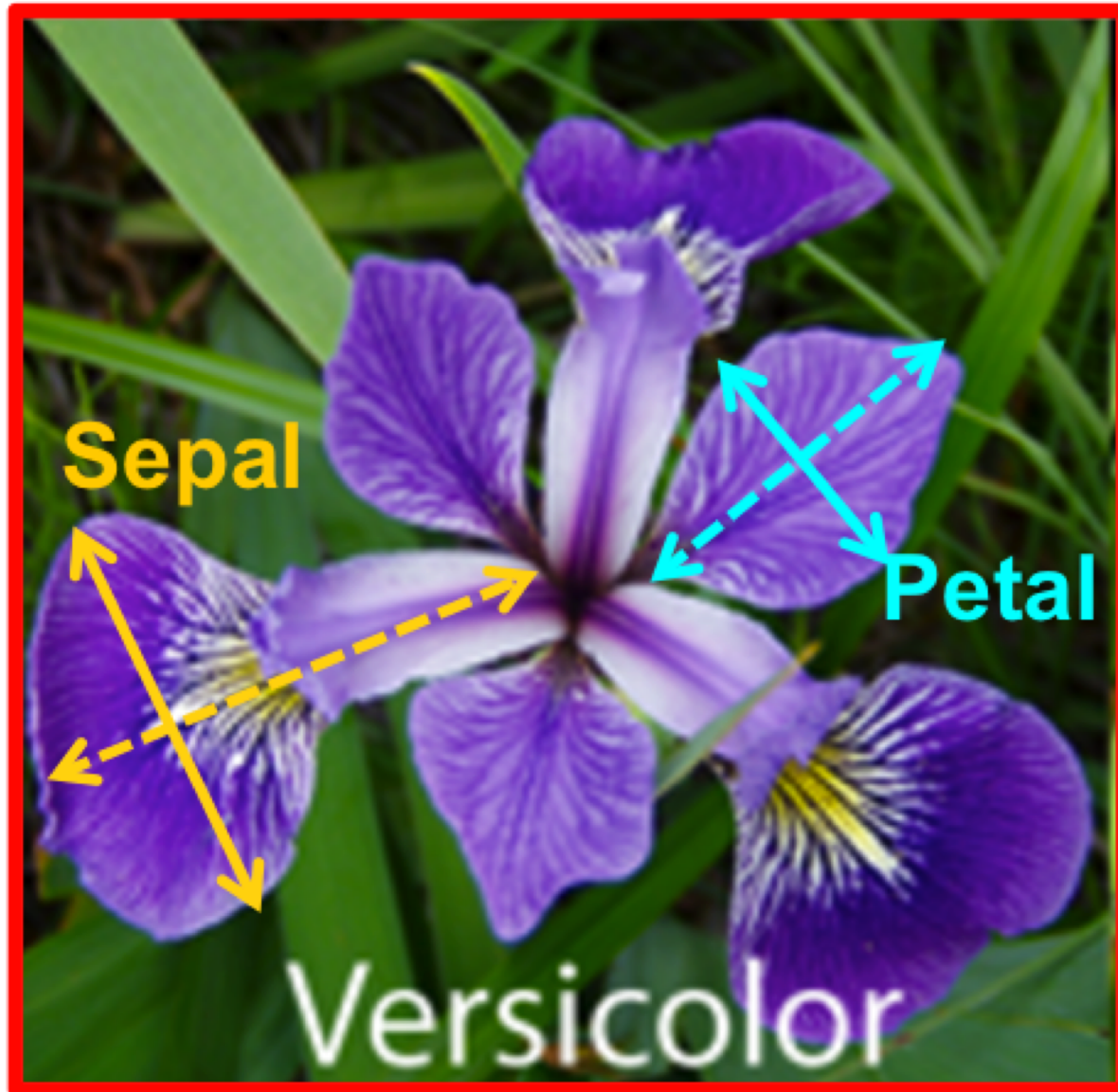
**virginica**



Source: [https://en.wikipedia.org/wiki/Iris\\_flower\\_data\\_set](https://en.wikipedia.org/wiki/Iris_flower_data_set)

Source: <http://suruchifaloke.com/2016-10-13-machine-learning-tutorial-iris-classification/>

# Iris Classification



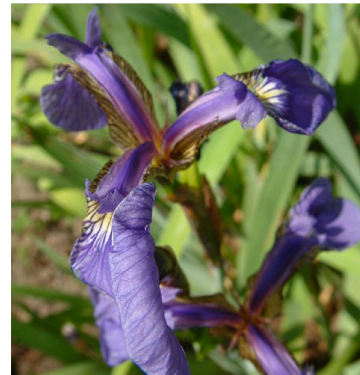


# iris.data

<https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data>

```
5.1,3.5,1.4,0.2,Iris-setosa
4.9,3.0,1.4,0.2,Iris-setosa
4.7,3.2,1.3,0.2,Iris-setosa
4.6,3.1,1.5,0.2,Iris-setosa
5.0,3.6,1.4,0.2,Iris-setosa
5.4,3.9,1.7,0.4,Iris-setosa
4.6,3.4,1.4,0.3,Iris-setosa
5.0,3.4,1.5,0.2,Iris-setosa
4.4,2.9,1.4,0.2,Iris-setosa
4.9,3.1,1.5,0.1,Iris-setosa
5.4,3.7,1.5,0.2,Iris-setosa
4.8,3.4,1.6,0.2,Iris-setosa
4.8,3.0,1.4,0.1,Iris-setosa
4.3,3.0,1.1,0.1,Iris-setosa
5.8,4.0,1.2,0.2,Iris-setosa
5.7,4.4,1.5,0.4,Iris-setosa
5.4,3.9,1.3,0.4,Iris-setosa
5.1,3.5,1.4,0.3,Iris-setosa
5.7,3.8,1.7,0.3,Iris-setosa
5.1,3.8,1.5,0.3,Iris-setosa
5.4,3.4,1.7,0.2,Iris-setosa
5.1,3.7,1.5,0.4,Iris-setosa
4.6,3.6,1.0,0.2,Iris-setosa
5.1,3.3,1.7,0.5,Iris-setosa
4.8,3.4,1.9,0.2,Iris-setosa
5.0,3.0,1.6,0.2,Iris-setosa
5.0,3.4,1.6,0.4,Iris-setosa
```

**setosa**



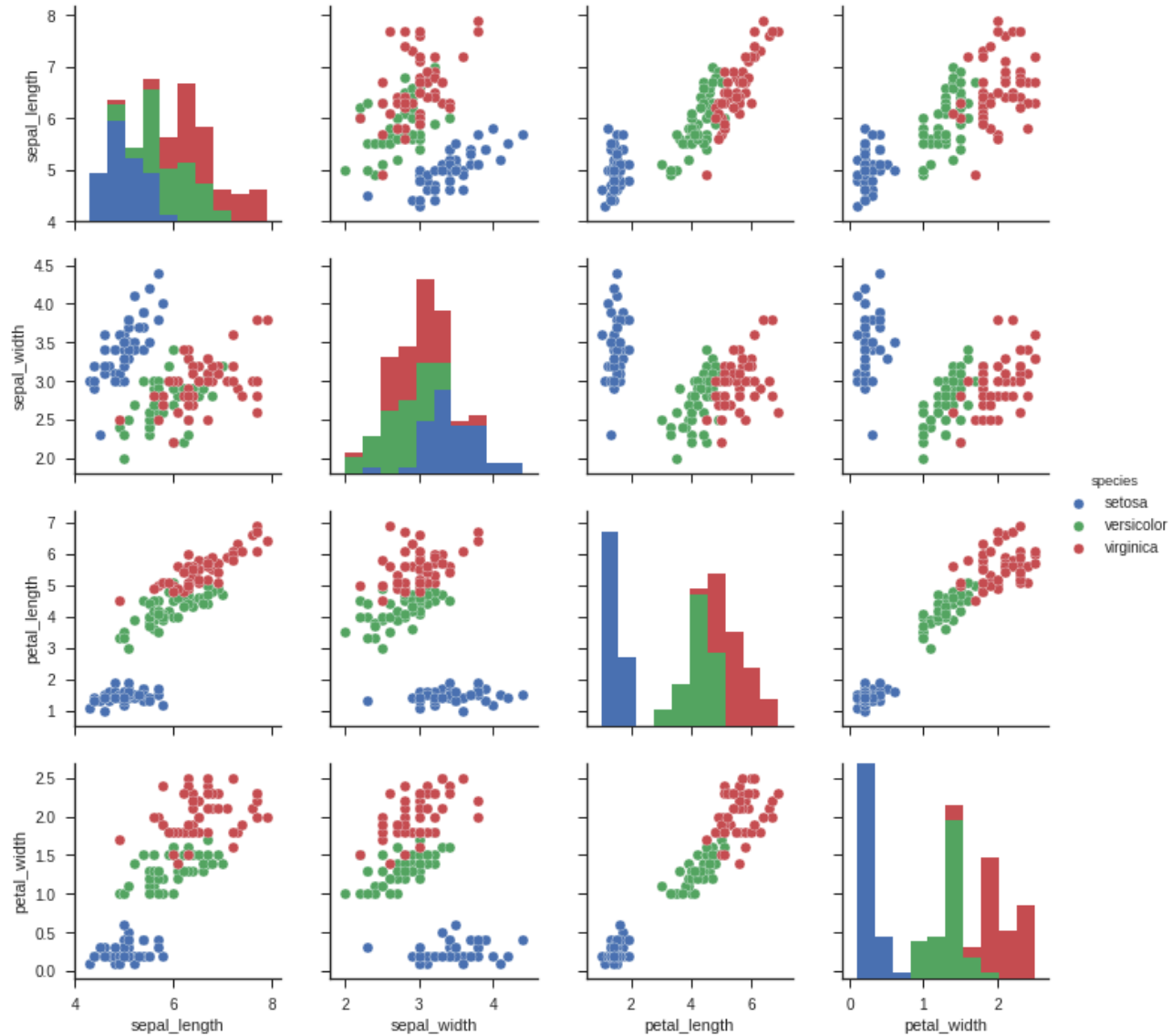
**virginica**



**versicolor**



# Iris Data Visualization



# Connect Google Colab in Google Drive

The image shows a browser window with the Google Drive interface. The address bar shows the URL `https://drive.google.com/drive/u/2/my-drive`. The main header includes the Drive logo, a search bar, and navigation icons. On the left sidebar, the 'New' button is highlighted with a red dashed box. A dropdown menu is open, listing options: 'New folder...', 'Upload files...', 'Upload folder...', 'Google Docs', 'Google Sheets', 'Google Slides', 'More', 'Google Forms', 'Google Drawings', 'Google My Maps', 'Google Sites', and 'Connect more apps'. The 'More' and 'Connect more apps' options are also highlighted with red dashed boxes. The 'Storage' section shows '0 bytes of 15 GB used' and an 'UPGRADE STORAGE' link. A 'Files' section is partially visible at the bottom.

# Google Colab








My Drive - Google Drive x +

https://drive.google.com/drive/u/2/my-drive

Drive Search Drive

Connect apps to Drive

All colab x

 <b>ZIP Extractor</b> Extract ZIP files to Google Drive Extraction complete. <a href="#">View extracted files</a> <a href="#">Share</a> <a href="#">Extract another</a>  Test.zip ZIP Extractor 307,585 users	 <b>LUMIN PDF</b> The fast and simple PDF Viewer box	 <b>cloudconvert</b> CloudConvert 373,161 users
 <b>Sejda</b> Merge PDF - Split PDF - Sejda.com ★★★★★ (1106)	 <b>DocHub</b> Edit, Send & Sign PDFs DocHub - Edit and Sign PDF Docu... 2,131,600 users	 <b>Google Forms</b> Google Forms 4,803,614 users

Get Backup and Sync for Mac

Access anywhere  
Share easily

Name ↑



# Google Colab


My Drive - Google Drive x +

https://drive.google.com/drive/u/2/my-drive

Drive Search Drive

Connect apps to Drive

All colab



**Colaboratory**  
offered by <https://colab.research.google.com>  
A data analysis tool that combines code, output, and descriptive text into one collaborative document.

**+ CONNECT**  
Productivity  
★★★★★ (195)

My Drive

Computers

Shared with me

Recent

Starred

Trash

Backups

Storage

0 bytes of 15 GB used  
[UPGRADE STORAGE](#)

Name ↑

Get Backup and Sync for Mac

Access anywhere

Share easily

# Connect Colaboratory to Google Drive

The image shows a browser window with the Google Drive interface. A dialog box titled "Connect apps to Drive" is open, displaying a search for "colab". A confirmation message from Colaboratory is shown, stating "Colaboratory was connected to Google Drive." and offering to make it the default app for files it can open. The message includes the Colaboratory logo, a checked checkbox, and an "OK" button. The background shows the Google Drive sidebar with options like "New", "My Drive", "Computers", "Shared with me", "Recent", "Starred", "Trash", "Backups", and "Storage".


My Drive - Google Drive x +

https://drive.google.com/drive/u/2/my-drive

Drive Search Drive

Connect apps to Drive

All colab

 **Colaboratory** was connected to Google Drive.

Make **Colaboratory** the default app for files it can open

OK

★ RATE IT

Productivity

★★★★★ (195)

Name ↑

Get Backup and Sync for Mac

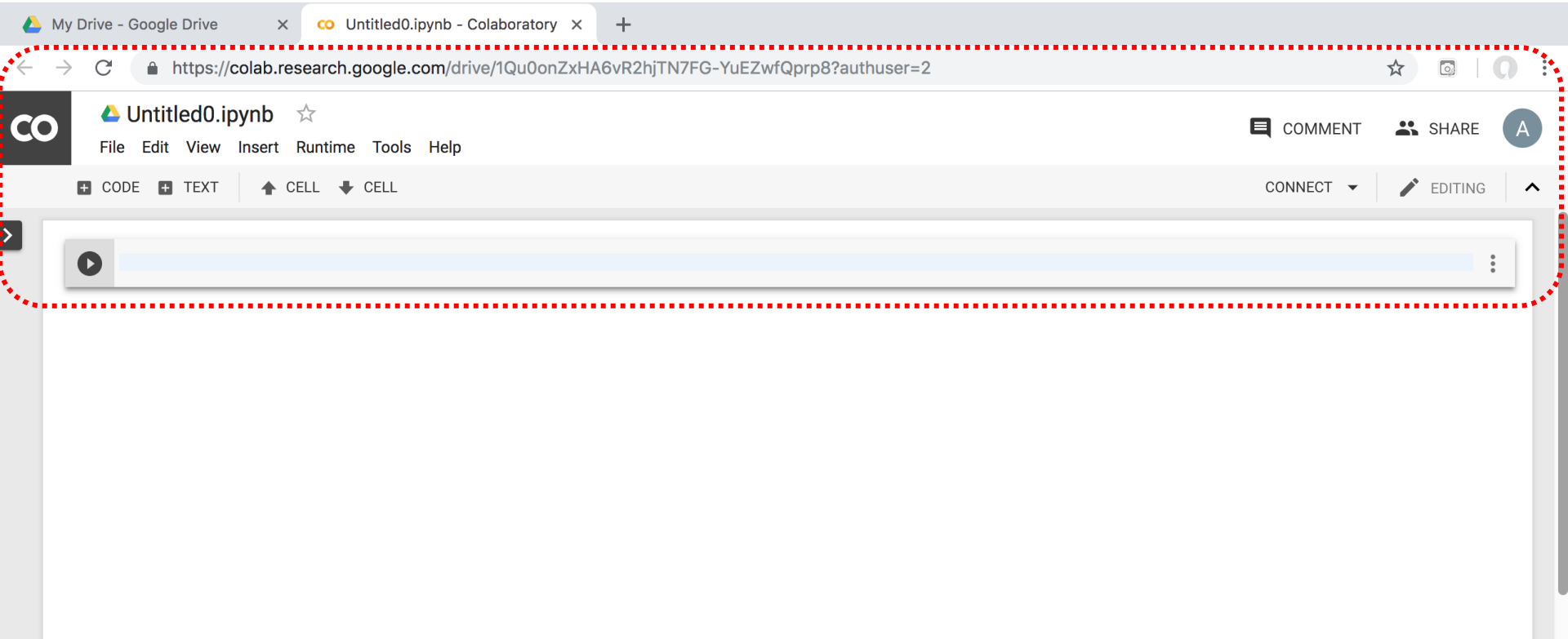
Access anywhere

Share easily

# Google Colab

The image shows a browser window with the Google Drive interface. The address bar displays 'https://drive.google.com/drive/u/2/my-drive'. The 'New' button is clicked, opening a menu with the following options: 'New folder...', 'Upload files...', 'Upload folder...', 'Google Docs', 'Google Sheets', 'Google Slides', 'More', 'Google Forms', 'Google Drawings', 'Google My Maps', 'Google Sites', 'Colaboratory', and 'Connect more apps'. The 'More' option in the first menu and the 'Colaboratory' option in the second menu are highlighted with red dashed boxes. The background shows the 'My Drive' section with a search bar and a 'Quick Access' section.

# Google Colab



# Google Colab

The image shows a browser window with two tabs: "My Drive - Google Drive" and "Untitled0.ipynb - Colaboratory". The address bar shows the URL: <https://colab.research.google.com/drive/1Qu0onZxHA6vR2hjTN7FG-YuEZwfQprp8?authuser=2>. The main content area displays the Google Colab interface for an "Untitled0.ipynb" notebook. The menu bar includes "File", "Edit", "View", "Insert", "Runtime", "Tools", and "Help". The "Runtime" menu is open, showing the following options:

- Run all (⌘/Ctrl+F9)
- Run before (⌘/Ctrl+F8)
- Run the focused cell (⌘/Ctrl+Enter)
- Run selection (⌘/Ctrl+Shift+Enter)
- Run after (⌘/Ctrl+F10)
- Interrupt execution (⌘/Ctrl+M I)
- Restart runtime... (⌘/Ctrl+M .)
- Restart and run all...
- Reset all runtimes...
- Change runtime type
- Manage sessions

The "Runtime" menu title and the "Change runtime type" option are highlighted with red dashed boxes. The interface also shows a "CONNECT" button and an "EDITING" mode indicator.

# Run Jupyter Notebook Python3 GPU Google Colab

The screenshot shows the Google Colab web interface. The browser tab is titled "Untitled0.ipynb - Colaboratory". The address bar shows the URL: <https://colab.research.google.com/drive/1Qu0onZxHA6vR2hjTN7FG-YuEZwfQprp8?authuser=2>. The interface includes a menu bar with "File", "Edit", "View", "Insert", "Runtime", "Tools", and "Help". Below the menu bar are buttons for "+ CODE", "+ TEXT", "↑ CELL", and "↓ CELL". On the right side, there are buttons for "COMMENT", "SHARE", "CONNECT", and "EDITING".

A "Notebook settings" dialog box is open in the center of the screen. It contains the following options:

- Runtime type:** A dropdown menu currently set to "Python 3".
- Hardware accelerator:** A dropdown menu currently set to "GPU".
- Omit code cell output when saving this notebook

At the bottom of the dialog box, there are two buttons: "CANCEL" and "SAVE".

# Google Colab Python Hello World

```
print('Hello World')
```

The screenshot displays the Google Colaboratory web interface. At the top, the browser tab is labeled 'Untitled0.ipynb - Colaboratory'. The address bar shows the URL: <https://colab.research.google.com/drive/1Qu0onZxHA6vR2hjTN7FG-YuEZwfQprp8?authuser=2#scrollTo=6s-m3sER8G1u>. The main header includes the 'Untitled0.ipynb' title, a star icon, and navigation links for 'File', 'Edit', 'View', 'Insert', 'Runtime', 'Tools', and 'Help'. On the right side of the header, there are buttons for 'COMMENT', 'SHARE', and a user profile icon labeled 'A'. Below the header, a toolbar contains options for '+ CODE', '+ TEXT', '↑ CELL', and '↓ CELL'. The status bar at the bottom right indicates 'CONNECTED' with a green checkmark and 'EDITING' with a pencil icon. The main workspace shows a code cell with a play button icon and the code `print('Hello World')`. Below the code cell, the output is displayed as 'Hello World' with a copy icon to its left.

# Data Visualization in Google Colab

datav.ipynb - Colaboratory

https://colab.research.google.com/drive/1KRqtEUd2Hg4dM2au9bfVQKrxWnWN3O9-?authuser=2

datav.ipynb

File Edit View Insert Runtime Tools Help

COMMENT SHARE

CONNECTED EDITING

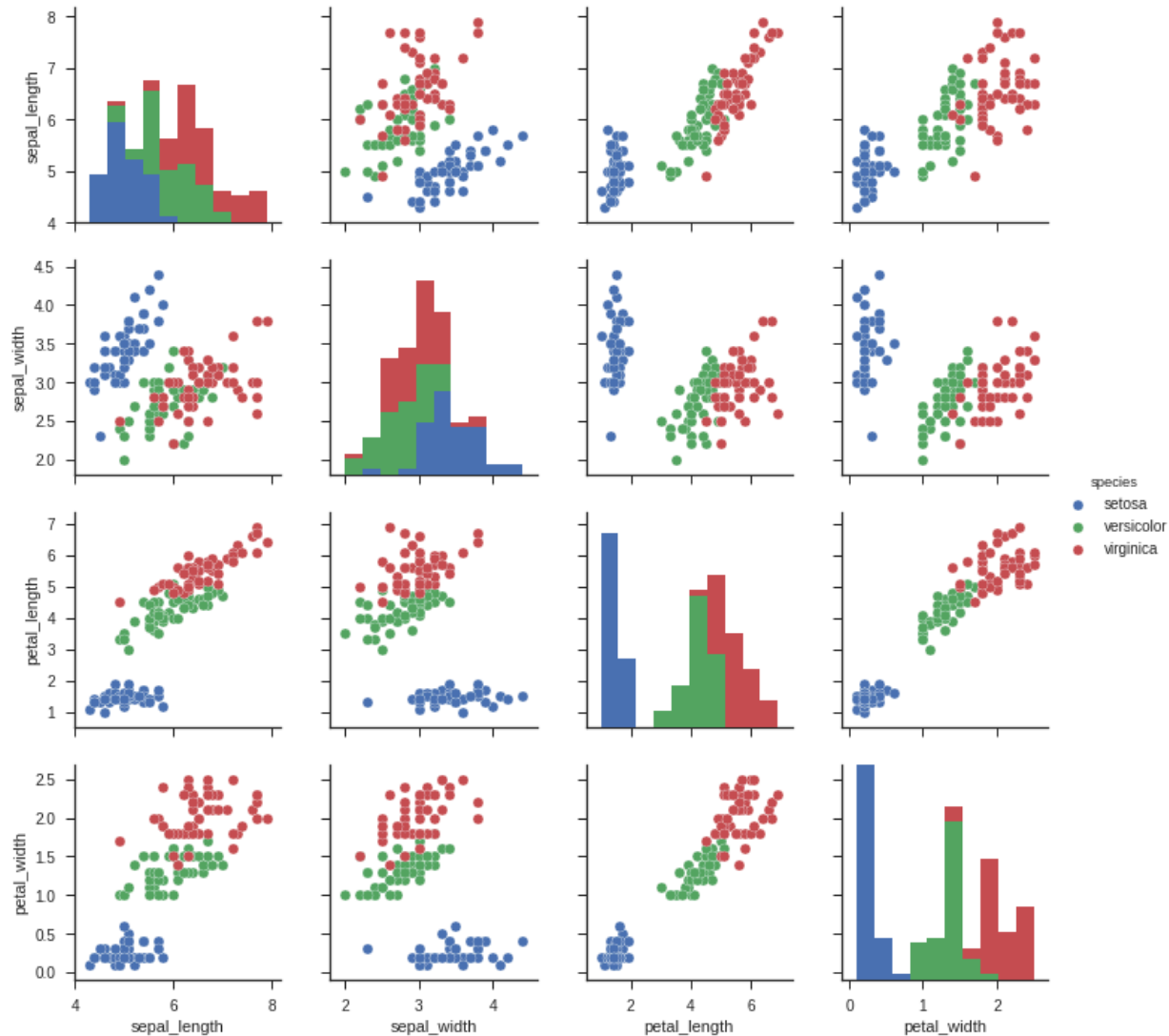
```
import seaborn as sns
sns.set(style="ticks", color_codes=True)
iris = sns.load_dataset("iris")
g = sns.pairplot(iris, hue="species").
```

species

- setosa
- versicolor
- virginica



```
import seaborn as sns
sns.set(style="ticks", color_codes=True)
iris = sns.load_dataset("iris")
g = sns.pairplot(iris, hue="species")
```



N3O9-

```
import numpy as np
import pandas as pd
%matplotlib inline
import matplotlib.pyplot as plt
import seaborn as sns
from pandas.plotting import scatter_matrix

# Load dataset
url = "https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data"
names = ['sepal-length', 'sepal-width', 'petal-length', 'petal-width', 'class']
df = pd.read_csv(url, names=names)

print(df.head(10))
print(df.tail(10))
print(df.describe())
print(df.info())
print(df.shape)
print(df.groupby('class').size())

plt.rcParams["figure.figsize"] = (10,8)
df.plot(kind='box', subplots=True, layout=(2,2), sharex=False, sharey=False)
plt.show()

df.hist()
plt.show()

scatter_matrix(df)
plt.show()

sns.pairplot(df, hue="class", size=2)
```

```
import numpy as np
import pandas as pd
%matplotlib inline
import matplotlib.pyplot as plt
import seaborn as sns
from pandas.plotting import scatter_matrix
```

```
# Import Libraries
import numpy as np
import pandas as pd
%matplotlib inline
import matplotlib.pyplot as plt
import seaborn as sns
from pandas.plotting import scatter_matrix
print('imported')
```

imported

```
url = "https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data"  
names = ['sepal-length', 'sepal-width', 'petal-length', 'petal-width', 'class']  
df = pd.read_csv(url, names=names)  
print(df.head(10))
```

```
# Load dataset  
url = "https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data"  
names = ['sepal-length', 'sepal-width', 'petal-length', 'petal-width', 'class']  
df = pd.read_csv(url, names=names)  
print(df.head(10)).
```

	sepal-length	sepal-width	petal-length	petal-width	class
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5.0	3.6	1.4	0.2	Iris-setosa
5	5.4	3.9	1.7	0.4	Iris-setosa
6	4.6	3.4	1.4	0.3	Iris-setosa
7	5.0	3.4	1.5	0.2	Iris-setosa
8	4.4	2.9	1.4	0.2	Iris-setosa
9	4.9	3.1	1.5	0.1	Iris-setosa

# df.tail(10)

```
print(df.tail(10)).
```

	sepal-length	sepal-width	petal-length	petal-width	class
140	6.7	3.1	5.6	2.4	Iris-virginica
141	6.9	3.1	5.1	2.3	Iris-virginica
142	5.8	2.7	5.1	1.9	Iris-virginica
143	6.8	3.2	5.9	2.3	Iris-virginica
144	6.7	3.3	5.7	2.5	Iris-virginica
145	6.7	3.0	5.2	2.3	Iris-virginica
146	6.3	2.5	5.0	1.9	Iris-virginica
147	6.5	3.0	5.2	2.0	Iris-virginica
148	6.2	3.4	5.4	2.3	Iris-virginica
149	5.9	3.0	5.1	1.8	Iris-virginica

# df.describe()

```
print(df.describe())
```

	sepal-length	sepal-width	petal-length	petal-width
count	150.000000	150.000000	150.000000	150.000000
mean	5.843333	3.054000	3.758667	1.198667
std	0.828066	0.433594	1.764420	0.763161
min	4.300000	2.000000	1.000000	0.100000
25%	5.100000	2.800000	1.600000	0.300000
50%	5.800000	3.000000	4.350000	1.300000
75%	6.400000	3.300000	5.100000	1.800000
max	7.900000	4.400000	6.900000	2.500000

```
print(df.info())  
print(df.shape)
```

```
print(df.info())
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 150 entries, 0 to 149  
Data columns (total 5 columns):  
sepal-length      150 non-null float64  
sepal-width       150 non-null float64  
petal-length      150 non-null float64  
petal-width       150 non-null float64  
class             150 non-null object  
dtypes: float64(4), object(1)  
memory usage: 5.9+ KB  
None
```

```
print(df.shape)
```

```
(150, 5)
```

```
df.groupby('class').size()
```

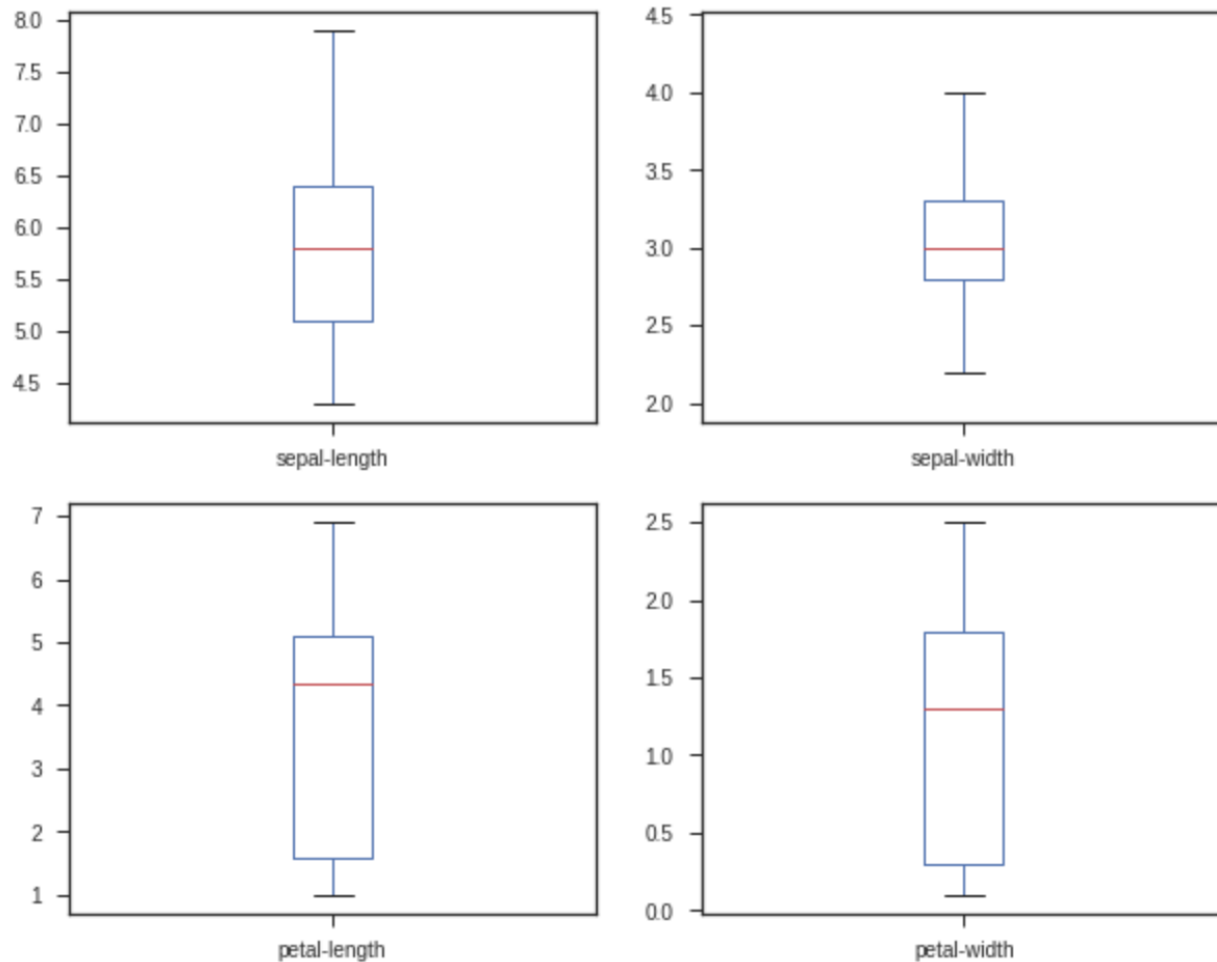
```
print(df.groupby('class').size())
```

```
class
Iris-setosa      50
Iris-versicolor 50
Iris-virginica   50
dtype: int64
```



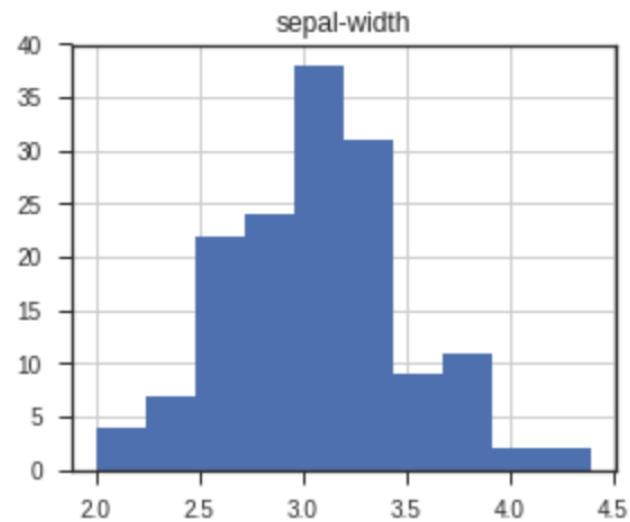
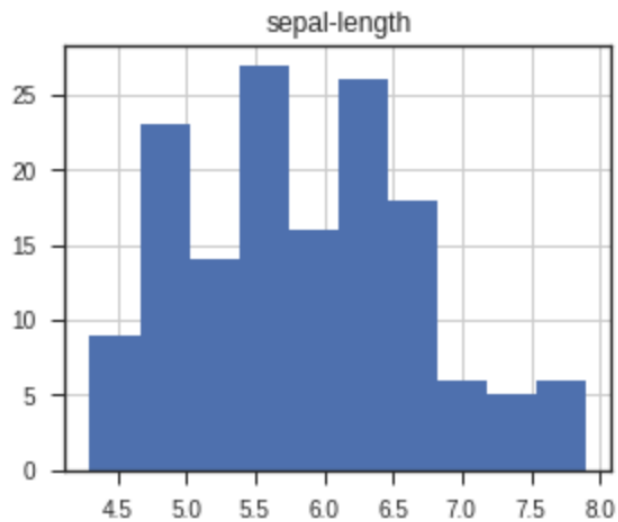
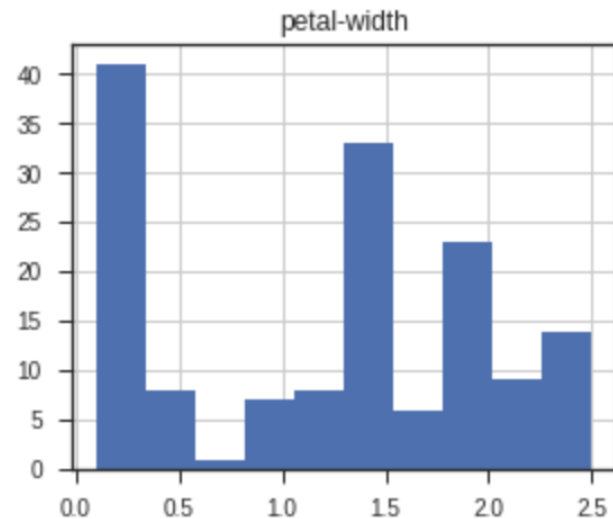
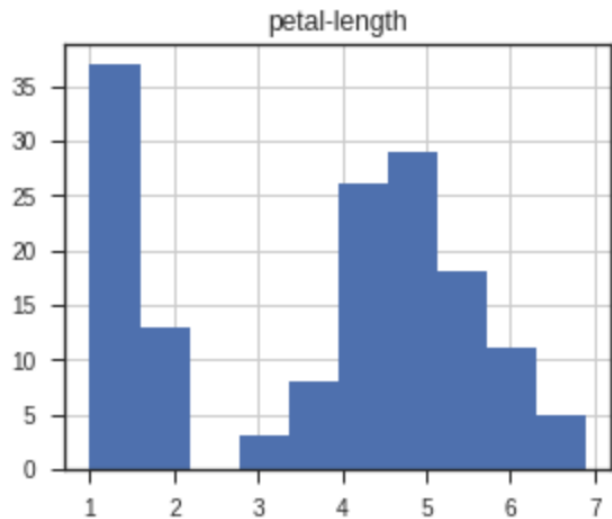
```
plt.rcParams["figure.figsize"] = (10,8)
df.plot(kind='box', subplots=True, layout=(2,2), sharex=False, sharey=False)
plt.show()
```

```
plt.rcParams["figure.figsize"] = (10,8)
df.plot(kind='box', subplots=True, layout=(2,2), sharex=False, sharey=False)
plt.show().
```



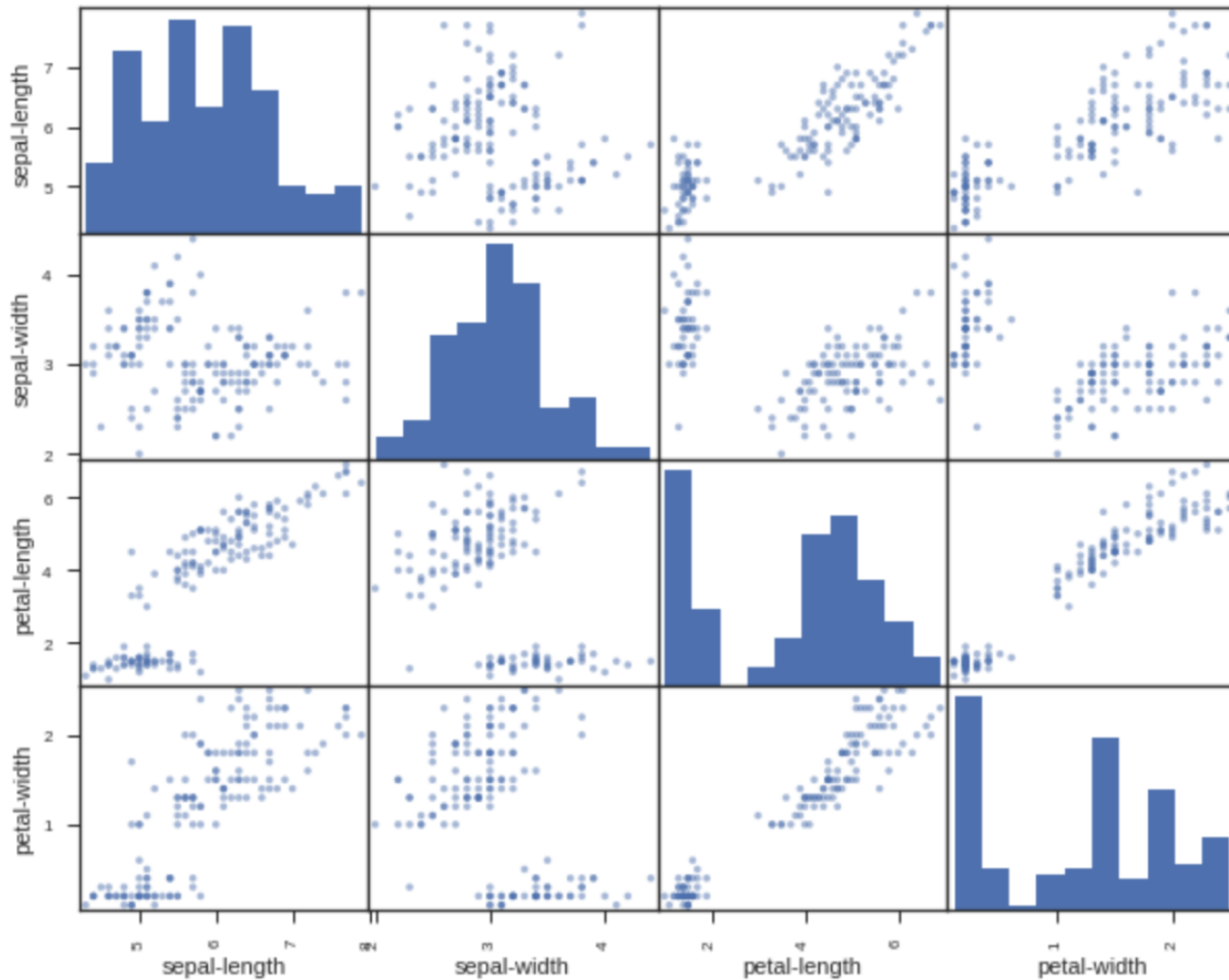
```
df.hist()  
plt.show()
```

```
df.hist()  
plt.show()
```



```
scatter_matrix(df)
plt.show()
```

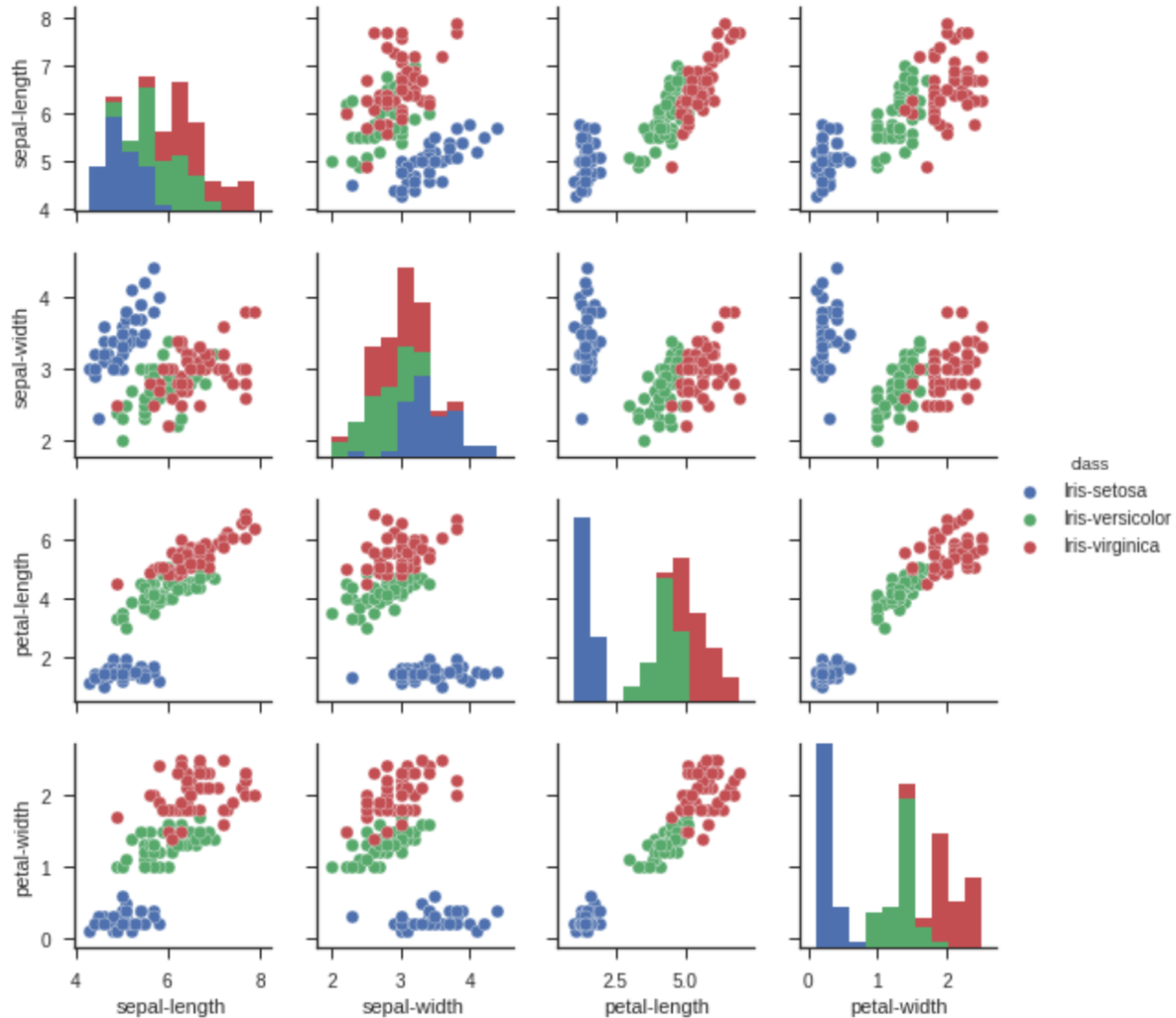
```
scatter_matrix(df)
plt.show(.
```



# sns.pairplot(df, hue="class", size=2)

```
sns.pairplot(df, hue="class", size=2)
```

```
<seaborn.axisgrid.PairGrid at 0x7f1d21267390>
```



# Summary

- Descriptive Analytics I
- Nature of Data
- Statistical Modeling
- Visualization

# References

- Ramesh Sharda, Dursun Delen, and Efraim Turban (2017), Business Intelligence, Analytics, and Data Science: A Managerial Perspective, 4th Edition, Pearson.
- EMC Education Services (2015), Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data, Wiley