Introduction to Machine Learning Program Assignment #1

Goals
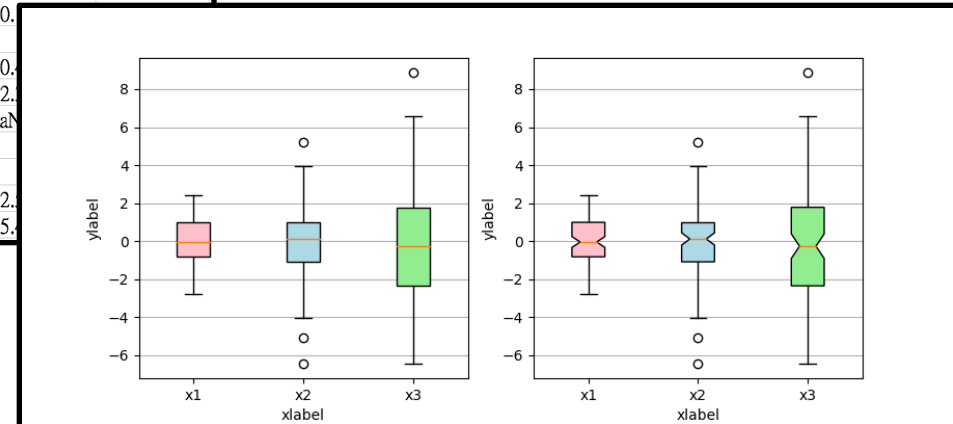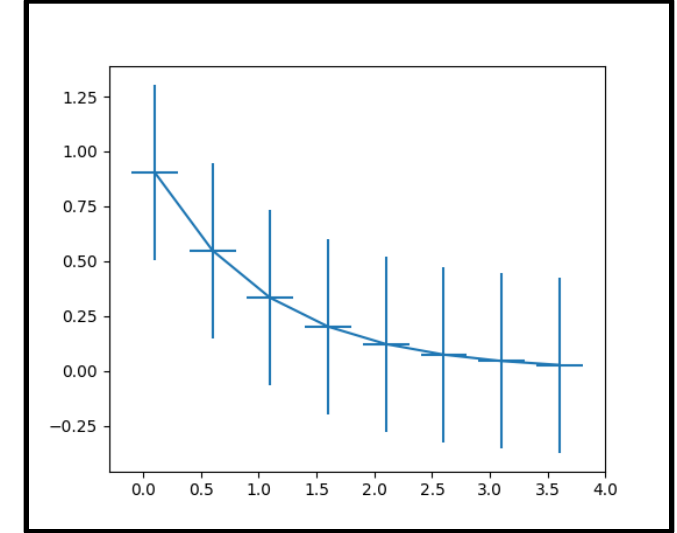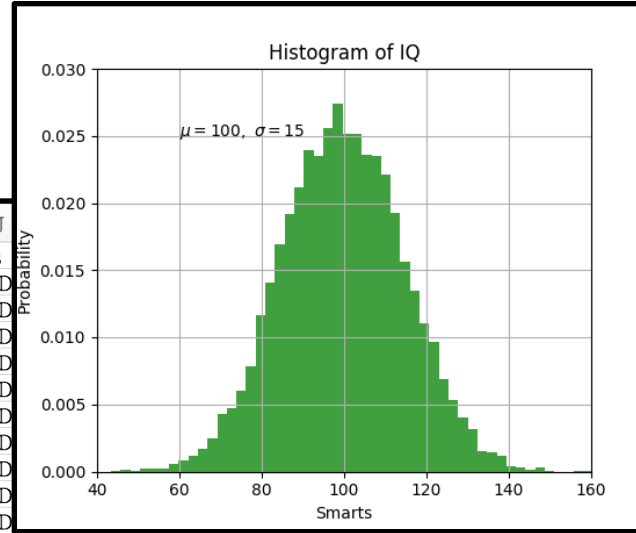
1. Environment Setup

2. Basic Workflow of Machine Learning / Data Mining Understanding

Dataset (front window):

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Ratin	Genres	Last Updated	Current Ver	Android Ver
2	Photo Editor	ART_AND_D	4.1	159	19M	10,000+	Free	0	Everyone	Art & Design	7-Jan-18	1.0.0	4.0.3 and up
3	Coloring book	ART_AND_D	3.9	967	14M	500,000+	Free	0	Everyone	Art & Design;	15-Jan-18	2.0.0	4.0.3 and up
4	U Launcher L	ART_AND_D	4.7	87510	8.7M	5,000,000+	Free	0	Everyone	Art & Design	1-Aug-18	1.2.4	4.0.3 and up
5	Sketch - Draw	ART_AND_D	4.5	215644	25M	50,000,000+	Free	0	Teen	Art & Design	8-Jun-18	Varies with d	4.2 and up
6	Pixel Draw -	ART_AND_D	4.3	967	2.8M	100,000+	Free	0	Everyone	Art & Design;	20-Jun-18	1.1	4.4 and up
7	Paper flowers	ART_AND_D	4.4	167	5.6M	50,000+	Free	0	Everyone	Art & Design	26-Mar-17	1	2.3 and up
8	Smoke Effect	ART_AND_D	3.8	178	19M	50,000+	Free	0	Everyone	Art & Design	26-Apr-18	1.1	4.0.3 and up
9	Infinite Painte	ART_AND_D	4.1	36815	29M	1,000,000+	Free	0	Everyone	Art & Design	14-Jun-18	6.1.61.1	4.2 and up
10	Garden Color	ART_AND_D	4.4	13791	33M	1,000,000+	Free	0	Everyone	Art & Design	20-Sep-17	2.9.2	3.0 and up
11	Kids Paint Fre	ART_AND_D	4.7	121	3.1M	10,000+	Free	0	Everyone	Art & Design;	3-Jul-18	2.8	4.0.3 and up
12	Text on Photo	ART_AND_D	4.4	13880	28M	1,000,000+	Free	0	Everyone	Art & Design	27-Oct-17	1.0.4	4.1 and up
13	Name Art Pho	ART_AND_D	4.4	8788	12M	1,000,000+	Free	0	Everyone	Art & Design	31-Jul-18	1.0.15	4.0 and up
14	Tattoo Name	ART_AND_D	4.2	44829	20M	10,000,000+	Free	0	Teen	Art & Design	2-Apr-18	3.8	4.1 and up
15	Mandala Colo	ART_AND_D	4.6	4326	21M	100,000+	Free	0	Everyone	Art & Design	26-Jun-18	1.0.4	4.4 and up
16	3D Color Pixe	ART_AND_D	4.4	1518	37M	100,000+	Free	0	Everyone	Art & Design	3-Aug-18	1.2.3	2.3 and up
17	Learn To Dra	ART_AND_D	3.2	55	2.7M	5,000+	Free	0	Everyone	Art & Design	6-Jun-18	NaN	4.2 and up
18	Photo Designe	ART_AND_D	4.7	3632	5.5M	500,000+	Free	0	Everyone	Art & Design	31-Jul-18	3.1	4.1 and up
19	350 Diy Roon	ART_AND_D	4.5	27	17M	10,000+	Free	0	Everyone	Art & Design	7-Nov-17	1	2.3 and up
20	FlipaClip - Ca	ART_AND_D	4.3	194216	39M	5,000,000+	Free	0	Everyone	Art & Design	3-Aug-18	2.2.5	4.0.3 and up
21	ibis Paint X	ART_AND_D	4.6	224399	31M	10,000,000+	Free	0	Everyone	Art & Design	30-Jul-18	5.5.4	4.1 and up

2 Datasets

1. Visualization

2. Preprocessing

Type	Last Updated	Android Ver
Free	7-Jan-18	4.0.3 and up
Free	15-Jan-18	4.0.3 and up
Free	1-Aug-18	4.0.3 and up
Free	8-Jun-18	4.2 and up
Free	20-Jun-18	4.4 and up

3. Decision Tree

With any package you want

4. Random Forest

(1) Randomly select data to train a decision tree

(2) Repeat (1) to generate multiple trees

(3) Take the prediction of every tree as a vote.
The prediction with the most votes wins.

4. Random Forest

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Ratin	Genres	Last Updated	Current Ver	Android Ver
1	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Ratin	Genres	Last Updated	Current Ver	Android Ver
2	Photo Editor	ART_AND_D	4.1	159	19M	10,000+	Free	0	Everyone	Art & Design	7-Jan-18	1.0.0	4.0.3 and up
3	Coloring book	ART_AND_D	3.9	967	14M	500,000+	Free	0	Everyone	Art & Design	15-Jan-18	2.0.0	4.0.3 and up
4	U Launcher L	ART_AND_D	4.7	87510	8.7M	5,000,000+	Free	0	Everyone	Art & Design	1-Aug-18	1.2.4	4.0.3 and up
5	Sketch - Draw	ART_AND_D	4.5	215644	25M	50,000,000+	Free	0	Teen	Art & Design	8-Jun-18	Varies with de	4.2 and up
6	Pixel Draw - I	ART_AND_D	4.3	967	2.8M	100,000+	Free	0	Everyone	Art & Design	20-Jun-18	1.1	4.4 and up
7	Paper flowers	ART_AND_D	4.4	167	5.6M	50,000+	Free	0	Everyone	Art & Design	26-Mar-17	1	2.3 and up
8	Smoke Effect	ART_AND_D	3.8	178	19M	50,000+	Free	0	Everyone	Art & Design	26-Apr-18	1.1	4.0.3 and up
9	Infinite Painte	ART_AND_D	4.1	36815	29M	1,000,000+	Free	0	Everyone	Art & Design	14-Jun-18	6.1.61.1	4.2 and up
10	Garden Color	ART_AND_D	4.4	13791	33M	1,000,000+	Free	0	Everyone	Art & Design	20-Sep-17	2.9.2	3.0 and up
11	Kids Paint Fre	ART_AND_D	4.7	121	3.1M	10,000+	Free	0	Everyone	Art & Design	3-Jul-18	2.8	4.0.3 and up
12	Text on Photo	ART_AND_D	4.4	13880	28M	1,000,000+	Free	0	Everyone	Art & Design	27-Oct-17	1.0.4	4.1 and up
13	Name Art Pho	ART_AND_D	4.4	8788	12M	1,000,000+	Free	0	Everyone	Art & Design	31-Jul-18	1.0.15	4.0 and up
14	Tattoo Name	ART_AND_D	4.2	44829	20M	10,000,000+	Free	0	Teen	Art & Design	2-Apr-18	3.8	4.1 and up
15	Mandala Colo	ART_AND_D	4.6	4326	21M	100,000+	Free	0	Everyone	Art & Design	26-Jun-18	1.0.4	4.4 and up
16	3D Color Pixe	ART_AND_D	4.4	1518	37M	100,000+	Free	0	Everyone	Art & Design	3-Aug-18	1.2.3	2.3 and up
17	Learn To Dra	ART_AND_D	3.2	55	2.7M	5,000+	Free	0	Everyone	Art & Design	6-Jun-18	NaN	4.2 and up
18	Photo Designe	ART_AND_D	4.7	3632	5.5M	500,000+	Free	0	Everyone	Art & Design	31-Jul-18	3.1	4.1 and up
19	350 Diy Room	ART_AND_D	4.5	27	17M	10,000+	Free	0	Everyone	Art & Design	7-Nov-17	1	2.3 and up
20	FlipaClip - Ca	ART_AND_D	4.3	194216	39M	5,000,000+	Free	0	Everyone	Art & Design	3-Aug-18	2.2.5	4.0.3 and up
21	ibis Paint X	ART_AND_D	4.6	224399	31M	10,000,000+	Free	0	Everyone	Art & Design	30-Jul-18	5.5.4	4.1 and up

4. Random Forest

	A	B	C	D	E	F	G	H	I	L	M
1	App	Category	Rating	Reviews	Size	Installs	ype	rice	Content Ratin	Genres	Last Updated	Current Ver	Android Ver
2	Photo Editor	ART_AND_[	4.1	159	19M	10,000+	ree	0	Everyone	Art & Design	7-Jan-18	1.0.0	4.0.3 and up
3	Coloring book	ART_AND_[	3.9	967	14M	500,000+	ree	0	Everyone	Art & Design	15-Jan-18	2.0.0	4.0.3 and up
4	U Launcher L	ART_AND_[	4.7	87510	8.7M	5,000,000+	ree	0	Everyone	Art & Design	1-Aug-18	1.2.4	4.0.3 and up
5	Sketch - Draw	ART_AND_[	4.5	215644	25M	50,000,000+	ree	0	Teen	Art & Design	8-Jun-18	Varies with de	4.2 and up
6	Pixel Draw - I	ART_AND_[	4.3	967	2.8M	100,000+	ree	0	Everyone	Art & Design	20-Jun-18	1.1	4.4 and up
7	Paper flowers	ART_AND_[	4.4	167	5.6M	50,000+	ree	0	Everyone	Art & Design	26-Mar-17	1	2.3 and up
8	Smoke Effect	ART_AND_[	3.8	178	19M	50,000+	ree	0	Everyone	Art & Design	26-Apr-18	1.1	4.0.3 and up
9	Infinite Painte	ART_AND_[	4.1	36815	29M	1,000,000+	ree	0	Everyone	Art & Design	14-Jun-18	6.1.61.1	4.2 and up
10	Garden Color	ART_AND_[	4.4	13791	33M	1,000,000+	ree	0	Everyone	Art & Design	20-Sep-17	2.9.2	3.0 and up
11	Kids Paint Fre	ART_AND_[	4.7	121	3.1M	10,000+	ree	0	Everyone	Art & Design	3-Jul-18	2.8	4.0.3 and up
12	Text on Photo	ART_AND_[	4.4	13880	28M	1,000,000+	ree	0	Everyone	Art & Design	27-Oct-17	1.0.4	4.1 and up
13	Name Art Pho	ART_AND_[	4.4	8788	12M	1,000,000+	ree	0	Everyone	Art & Design	31-Jul-18	1.0.15	4.0 and up
14	Tattoo Name	ART_AND_[	4.2	44829	20M	10,000,000+	ree	0	Teen	Art & Design	2-Apr-18	3.8	4.1 and up
15	Mandala Colo	ART_AND_[	4.6	4326	21M	100,000+	ree	0	Everyone	Art & Design	26-Jun-18	1.0.4	4.4 and up
16	3D Color Pixe	ART_AND_[	4.4	1518	37M	100,000+	ree	0	Everyone	Art & Design	3-Aug-18	1.2.3	2.3 and up
17	Learn To Dra	ART_AND_[	3.2	55	2.7M	5,000+	ree	0	Everyone	Art & Design	6-Jun-18	NaN	4.2 and up
18	Photo Designe	ART_AND_[	4.7	3632	5.5M	500,000+	ree	0	Everyone	Art & Design	31-Jul-18	3.1	4.1 and up
19	350 Diy Roon	ART_AND_[	4.5	27	17M	10,000+	ree	0	Everyone	Art & Design	7-Nov-17	1	2.3 and up
20	FlipaClip - Ca	ART_AND_[	4.3	194216	39M	5,000,000+	ree	0	Everyone	Art & Design	3-Aug-18	2.2.5	4.0.3 and up
21	ibis Paint X	ART_AND_[	4.6	224399	31M	10,000,000+	ree	0	Everyone	Art & Design	30-Jul-18	5.5.4	4.1 and up

4. Random Forest

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Ratin	Genres	Last Updated	Current Ver	Android Ver
2	Photo Editor	ART_AND_D	4.1	159	19M	10,000+	Free	0	Everyone	Art & Design	7-Jan-18	1.0.0	4.0.3 and up
3	Coloring book	ART_AND_D	3.9	967	14M	500,000+	Free	0	Everyone	Art & Design	15-Jan-18	2.0.0	4.0.3 and up
4	U Launcher L	ART_AND_D	4.7	87510	8.7M	5,000,000+	Free	0	Everyone	Art & Design	1-Aug-18	1.2.4	4.0.3 and up
5	Sketch - Draw	ART_AND_D	4.5	215644	25M	50,000,000+	Free	0	Teen	Art & Design	8-Jun-18	Varies with de	4.2 and up
6	Pixel Draw - I	ART_AND_D	4.3	967	2.8M	100,000+	Free	0	Everyone	Art & Design	20-Jun-18	1.1	4.4 and up
7	Paper flowers	ART_AND_D	4.4	167	5.6M	50,000+	Free	0	Everyone	Art & Design	26-Mar-17	1	2.3 and up
8	Smoke Effect	ART_AND_D	3.8	178	19M	50,000+	Free	0	Everyone	Art & Design	26-Apr-18	1.1	4.0.3 and up
9	Infinite Painte	ART_AND_D	4.1	36815	29M	1,000,000+	Free	0	Everyone	Art & Design	14-Jun-18	6.1.61.1	4.2 and up
10	Garden Color	ART_AND_D	4.4	13791	33M	1,000,000+	Free	0	Everyone	Art & Design	20-Sep-17	2.9.2	3.0 and up
11	Kids Paint Fre	ART_AND_D	4.7	121	3.1M	10,000+	Free	0	Everyone	Art & Design	3-Jul-18	2.8	4.0.3 and up
12	Text on Photo	ART_AND_D	4.4	13880	28M	1,000,000+	Free	0	Everyone	Art & Design	27-Oct-17	1.0.4	4.1 and up
13	Name Art Pho	ART_AND_D	4.4	8788	12M	1,000,000+	Free	0	Everyone	Art & Design	31-Jul-18	1.0.15	4.0 and up
14	Tattoo Name	ART_AND_D	4.2	44829	20M	10,000,000+	Free	0	Teen	Art & Design	2-Apr-18	3.8	4.1 and up
15	Mandala Colo	ART_AND_D	4.6	4326	21M	100,000+	Free	0	Everyone	Art & Design	26-Jun-18	1.0.4	4.4 and up
16	3D Color Pixe	ART_AND_D	4.4	1518	37M	100,000+	Free	0	Everyone	Art & Design	3-Aug-18	1.2.3	2.3 and up
17	Learn To Dra	ART_AND_D	3.2	55	2.7M	5,000+	Free	0	Everyone	Art & Design	6-Jun-18	NaN	4.2 and up
18	Photo Designe	ART_AND_D	4.7	3632	5.5M	500,000+	Free	0	Everyone	Art & Design	31-Jul-18	3.1	4.1 and up
19	350 Diy Room	ART_AND_D	4.5	27	17M	10,000+	Free	0	Everyone	Art & Design	7-Nov-17	1	2.3 and up
20	FlipaClip - Ca	ART_AND_D	4.3	194216	39M	5,000,000+	Free	0	Everyone	Art & Design	3-Aug-18	2.2.5	4.0.3 and up
21	ibis Paint X	ART_AND_D	4.6	224399	31M	10,000,000+	Free	0	Everyone	Art & Design	30-Jul-18	5.5.4	4.1 and up

4. Random Forest

	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Ratin	Genres	Last Updated	Current Ver	Android Ver
1	App	Category	Rating	Reviews	Size	Installs	Type	Price	Content Ratin	Genres	Last Updated	Current Ver	Android Ver
2	Photo Editor	ART_AND_D	4.1	159	19M	10,000+	Free	0	Everyone	Art & Design	7-Jan-18	1.0.0	4.0.3 and up
3	Coloring book	ART_AND_D	3.9	967	14M	500,000+	Free	0	Everyone	Art & Design	15-Jan-18	2.0.0	4.0.3 and up
4	U Launcher L	ART_AND_D	4.7	87510	8.7M	5,000,000+	Free	0	Everyone	Art & Design	1-Aug-18	1.2.4	4.0.3 and up
5	Sketch - Draw	ART_AND_D	4.5	215644	25M	50,000,000+	Free	0	Teen	Art & Design	8-Jun-18	Varies with de	4.2 and up
6	Pixel Draw -	ART_AND_D	4.3	967	2.8M	100,000+	Free	0	Everyone	Art & Design	20-Jun-18	1.1	4.4 and up
7	Paper flowers	ART_AND_D	4.4	167	5.6M	50,000+	Free	0	Everyone	Art & Design	26-Mar-17	1	2.3 and up
8	Smoke Effect	ART_AND_D	3.8	178	19M	50,000+	Free	0	Everyone	Art & Design	26-Apr-18	1.1	4.0.3 and up
9	Infinite Painte	ART_AND_D	4.1	36815	29M	1,000,000+	Free	0	Everyone	Art & Design	14-Jun-18	6.1.61.1	4.2 and up
10	Garden Color	ART_AND_D	4.4	13791	33M	1,000,000+	Free	0	Everyone	Art & Design	20-Sep-17	2.9.2	3.0 and up
11	Kids Paint Fre	ART_AND_D	4.7	121	3.1M	10,000+	Free	0	Everyone	Art & Design	3-Jul-18	2.8	4.0.3 and up
12	Text on Photo	ART_AND_D	4.4	13880	28M	1,000,000+	Free	0	Everyone	Art & Design	27-Oct-17	1.0.4	4.1 and up
13	Name Art Pho	ART_AND_D	4.4	8788	12M	1,000,000+	Free	0	Everyone	Art & Design	31-Jul-18	1.0.15	4.0 and up
14	Tattoo Name	ART_AND_D	4.2	44829	20M	10,000,000+	Free	0	Teen	Art & Design	2-Apr-18	3.8	4.1 and up
15	Mandala Colo	ART_AND_D	4.6	4326	21M	100,000+	Free	0	Everyone	Art & Design	26-Jun-18	1.0.4	and up
16	3D Color Pixe	ART_AND_D	4.4	1518	37M	100,000+	Free	0	Everyone	Art & Design	3-Aug-18	1.2.3	2.3 and up
17	Learn To Dra	ART_AND_D	3.2	55	2.7M	5,000+	Free	0	Everyone	Art & Design	6-Jun-18	NaN	4.2 and up
18	Photo Designe	ART_AND_D	4.7	3632	5.5M	500,000+	Free	0	Everyone	Art & Design	31-Jul-18	3.1	4.1 and up
19	350 Diy Room	ART_AND_D	4.5	27	17M	10,000+	Free	0	Everyone	Art & Design	7-Nov-17	1	2.3 and up
20	FlipaClip - C	ART_AND_D	4.3	194216	39M	5,000,000+	Free	0	Everyone	Art & Design	3-Aug-18	2.2.5	4.0.3 and up
21	ibis Paint X	ART_AND_D	4.6	224399	31M	10,000,000+	Free	0	Everyone	Art & Design	30-Jul-18	5.5.4	4.1 and up

4. Random Forest

< It's definitely a **CAT**.

< It's definitely a **DOG**.

< It's definitely a **CAT**.

4. Random Forest

Note: You are NOT allowed to use package-provided random forest model for this assignment.
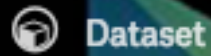
4. Validation

(1) Confusion Matrix

(2) Resubstitution validaiton

(3) K-fold cross validation

IRIS Flower Dataset

Google Play Store Apps Dataset

III.   Report & Scoring

This is a team-based program assignment, so one team should only submit one report and one source code to E3. The report should contain the following for both datasets:

1.   What environments the members are using (5%)
2.   Basic statistic visualization of the data (20%)
3.   Data preprocessing methods (20%)
4.   How you generate decision tree and random forest models (20%)
5.   The performance (20%)
6.   Conclusion (20%)

Some Rules

1. This is a **TEAM** program assignment.
   All the members must be able to execute the code.

Some Rules

2. Each member should provide **one screenshot** of the code-execution result in the report.

Some Rules

3. C / C++ / Java / Python / Matlab are allowed to use.
   For visualization, Excel or other programs are allowed.

Some Rules

4. Report format should be **PDF**.

5. One team one submission to E3 which contains one **report** and one **source code**

Some Rules

6. No cheating and plagiarizing

If you are using Python

jupyter Lorenz Differential Equations (autosaved)

Code ▾   Cell Toolbar: None ▾

Exploring the Lorenz System

In this Notebook we explore the Lorenz system of differential equations:

$$\dot{x} = \sigma(y - x)$$
$$\dot{y} = \rho x - y - xz$$
$$\dot{z} = -\beta z + xy$$

This is one of the classic systems in non-linear differential equations. It exhibits a range of complex behaviors as the parameters (σ, β, ρ) are varied, including what are known as *chaotic solutions*. The system was originally developed as a simplified mathematical model for atmospheric convection in 1963.

```
In [7]:  interact(Lorenz, N=fixed(10), angle=(0.,360.),
                  σ=(0.0,50.0),β=(0.,5), ρ=(0.0,50.0))
```

angle ————————————————————————— 308.2

max_time ———————————————————————— 12

σ ———————————————————————— 10

β ———————————————————————— 2.6

ρ ———————————————————————— 28

---

jupyter Welcome to P...

jupyter

Welcome to the

This Notebook Server wa:

WARNING
Don't rely on this sen

Your server is hosted than

Run some Python

To run the code below:

1. Click on the cell to se
2. Press SHIFT+ENTER

A full tutorial for using the

```
In [ ]:  %matplotlib inline

         import pandas as pd
         import numpy as np
         import matplotlib
```
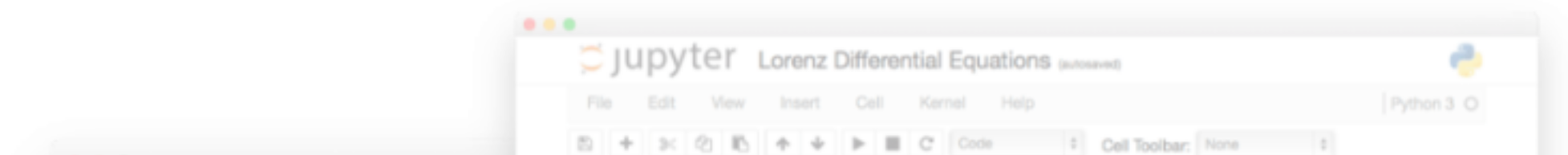
Project Jupyter exists to develop open-source software, open-standards, and services for interactive computing across dozens of programming languages.
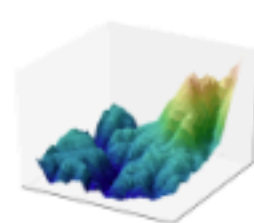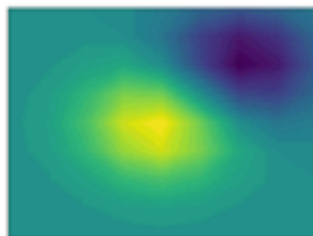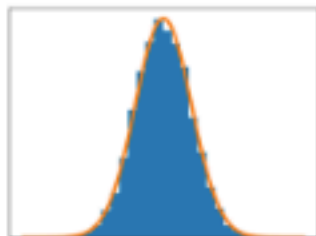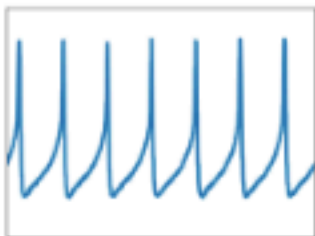
matplotlib

Version 3.0.0

Matplotlib is a Python 2D plotting library which produces publication quality figures in a variety of hardcopy formats and interactive environments across platforms. Matplotlib can be used in Python scripts, the Python and IPython shells, the Jupyter notebook, web application servers, and four graphical user interface toolkits.
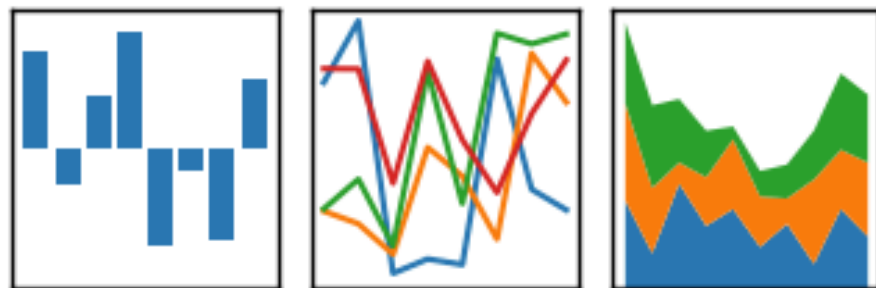
Matplotlib tries to make easy things easy and hard things possible. You can generate plots, histograms, power spectra, bar charts, errorcharts, scatterplots, etc., with just a few lines of code. For examples, see the sample plots and thumbnail gallery.

For simple plotting the `pyplot` module provides a MATLAB-like interface, particularly when combined with IPython. For the power user, you have full control of line styles, font properties, axes properties, etc, via an object oriented interface or via a set of functions familiar to MATLAB users.

pandas

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

Python Data Analysis Library

pandas is an open source, BSD-licensed library providing high-performance, easy-to-use data structures and data analysis tools for the Python programming language.

pandas is a NumFOCUS sponsored project. This will help ensure the success of development of *pandas* as a world-class open-source project, and makes it possible to donate to the project.

VERSIONS

Release

0.23.4 - August 2018

download // docs // pdf

Development

0.24.0 - 2018

github // docs