



Amine BENHENNI

albenhenni@gmail.com

Institut Villebon
Georges Charpak

5.2 : Analyse de données

De la donnée partout !

**Harvard
Business
Review**

DATA

Data Scientist: The Sexiest Job of the 21st Century

by **Thomas H. Davenport** and **D.J. Patil**

FROM THE OCTOBER 2012 ISSUE

- **Systèmes de recommandation en ligne**
- **Traduction automatique**
- **Diagnostiques médicaux**
- **Pilotage d'usines**
- **Voitures autonomes**
- **Chatbots**
- **Assistants intelligents**
- **...**



A person riding a motorcycle on a dirt road.



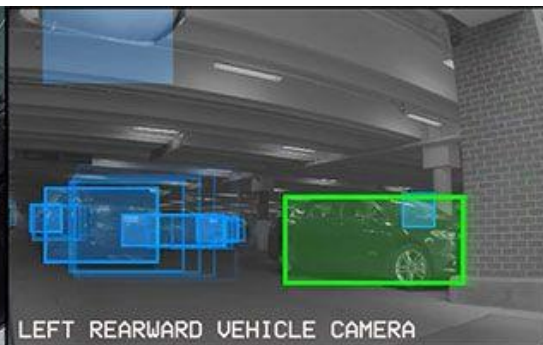
Two dogs play in the grass.



A group of young people playing a game of frisbee.



Two hockey players are fighting over the puck.



ROAD FLOW

IN-PATH OBJECTS

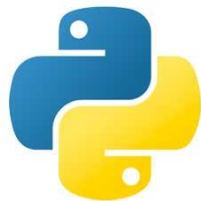
ROAD LIGHTS

OBJECTS

ROAD SIGNS

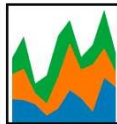
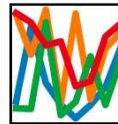


Python : langage star de la data



pandas

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



pythonTM



SymPy



Flask



django



Institut Villebon
Georges Charpak

Le cours

Roadmap

Concepts scientifiques

Statistiques, vecteurs, algèbre linéaire,
éléments de Machine Learning

UE 5.2

Gestion de projet

Git, kanban, SMART

Environnement technique

Python, librairies data
(numpy, sklearn, matplotlib, etc)

Déroulement du semestre

Pour fin septembre:

- Premières bases techniques et scientifiques
- Mise en place des projets

Pour fin octobre :

- Régression linéaire, notion de fit/apprentissage
- Roadmap des projets définie

Pour fin novembre :

- Réduction de dimension (PCA), éléments de Machine Learning
- Premiers résultats projets

Modalités de contrôle des connaissances

Examen de 1h30 à la mi-novembre (coeff 3)

Contrôles continus tout le long du module (coeff 4)

Présentation des projets (coeff 3) mi-décembre

A propos des projets

Comprendre et s'approprier un sujet

Illustrer le sujet avec un **code de démonstration disponible sur github** pour les camarades

Petit cours (30') pour expliquer les concepts et le cas d'illustration*

Gestion de projet pour le travail en équipe (**note de CC**)

Projets proposés les années précédentes

- Reconnaissance d'activités par accéléromètre
- Relation entre pollution et mortalité
- Détecteur de baleine
- Reconnaissance de caractères
- Prédiction de notes de films
- Reconnaissance faciale
- Diagnostique cancer du sein
- Détection de crises d'épilepsie
- ...



Institut Villebon
Georges Charpak

Outils

Jupyter notebooks

The image shows two overlapping Jupyter Notebook windows. The background window displays the 'Welcome to the Jupyter Notebook Server' page, which includes a warning message: 'WARNING: Don't rely on this server for production use. Your server is hosted there' and instructions on how to run Python code. The foreground window is titled 'Lorenz Differential Equations (autosaved)' and contains the following content:

Exploring the Lorenz System

In this Notebook we explore the [Lorenz system](#) of differential equations:

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

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.

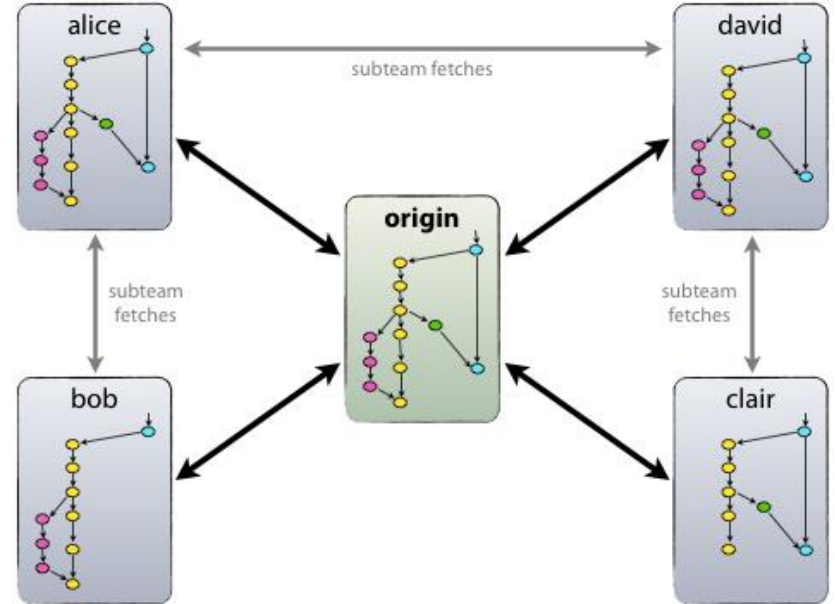
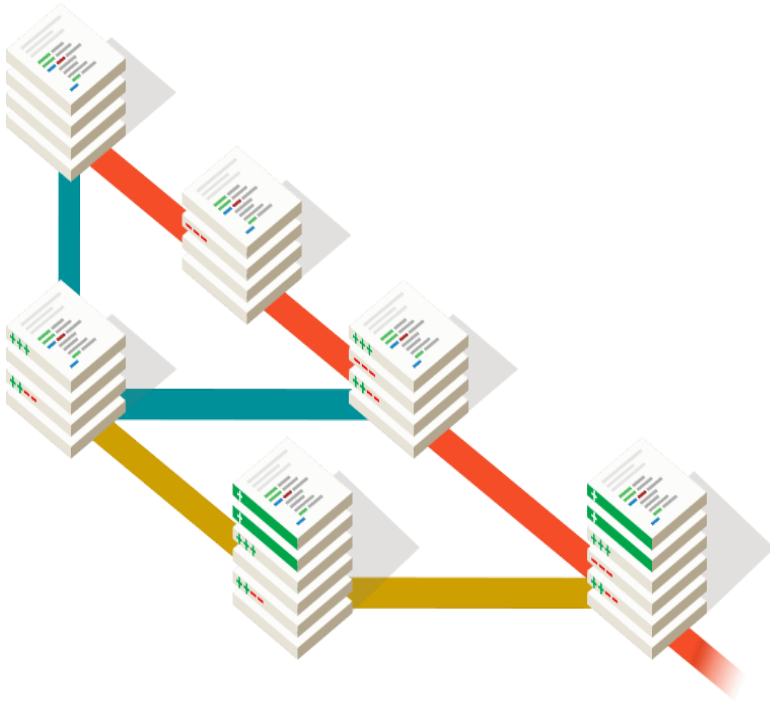
In [7]: `interact(Lorenz, N=fixed(10), angle=(0.,360.),
sigma=(0.0,50.0),beta=(0.,5), rho=(0.0,50.0))`

The interaction interface shows sliders for the following parameters:

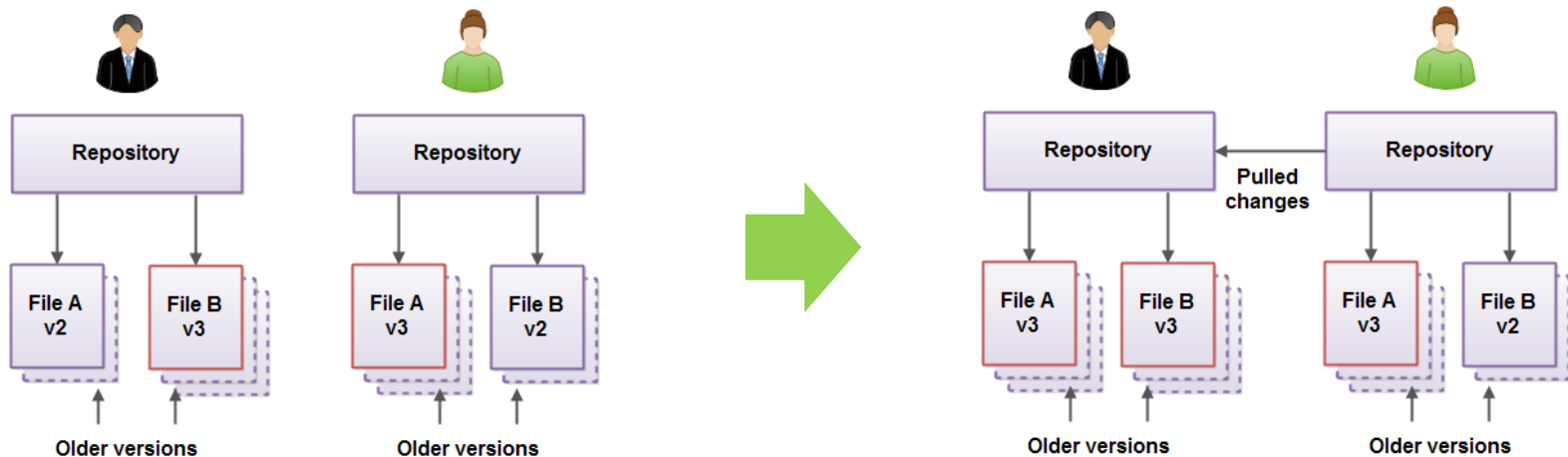
- angle: 308.2
- max_time: 12
- σ : 10
- β : 2.6
- ρ : 28

Below the sliders is a 3D plot of the Lorenz attractor, showing the characteristic butterfly shape with multiple colored trajectories.

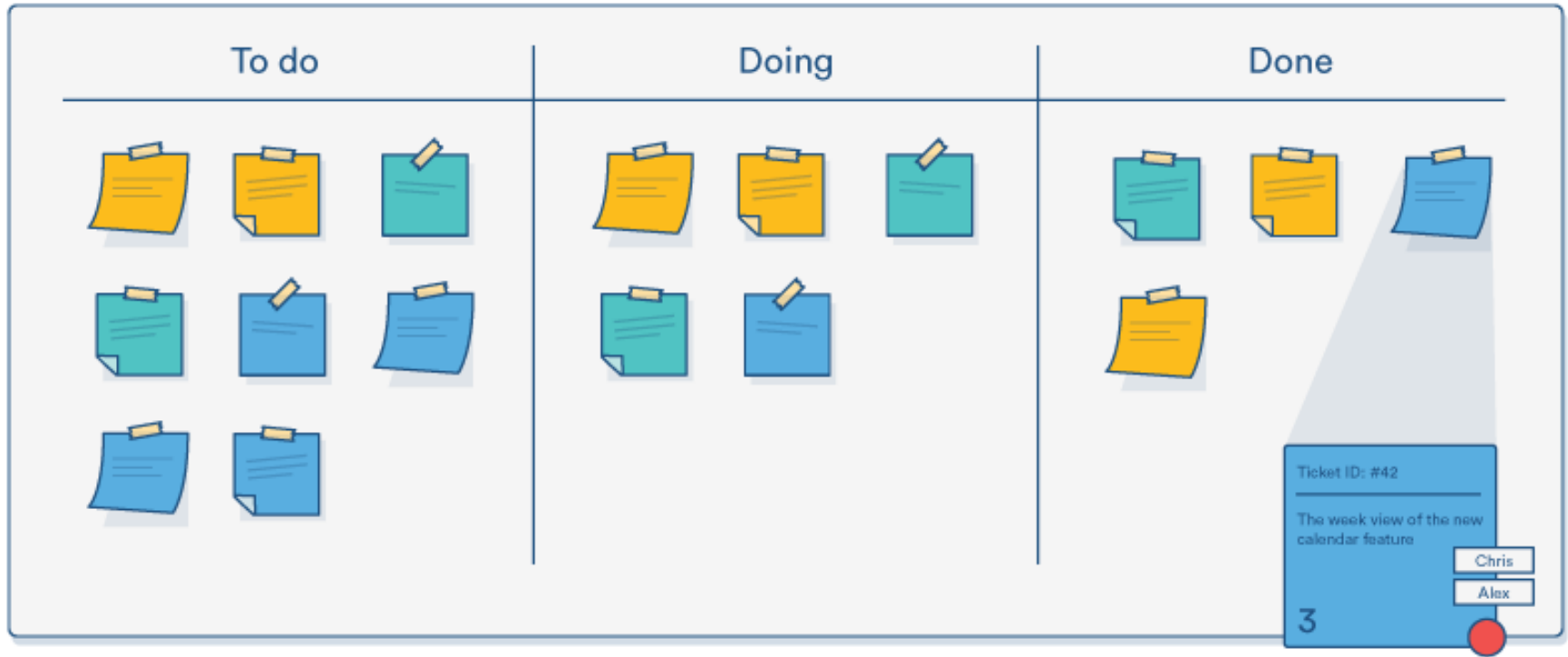
Gestion du code : versioning avec GIT



Gestion du code : versioning avec GIT



Gestion de projet : méthodes agile / Kanban



Workflow continue avec une charge max un instant t

Gestion de projet : SMART

S

Specific

M

Measurable

A

Attainable

R

Relevant

T

Timed



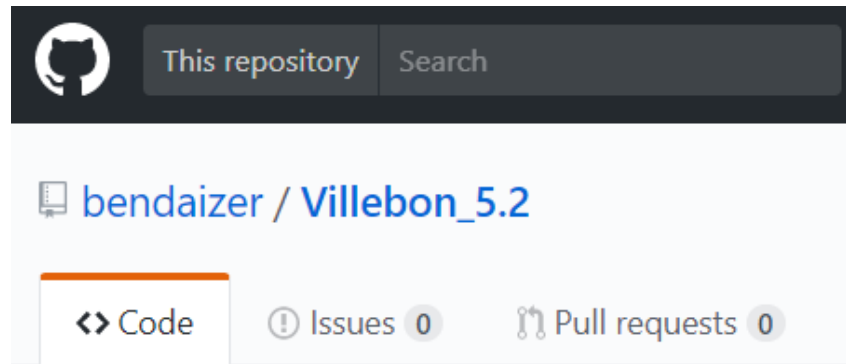
Institut Villebon
Georges Charpak

Informations pratiques

Cours disponible sur github.com

Le cours est disponible ici :

https://github.com/bendaizer/Villebon_5.2



Il sera mis à jour à chaque séance.

Vous utiliserez git pour récupérer les derniers pushes.

Création de comptes et installation



github

<http://github.com>



<http://trello.com>



<https://www.anaconda.com/download/>



<https://git-scm.com/download>



Institut Villebon
Georges Charpak

Next Steps

Envoi email > albenhenni@dataswati.com

- Nom / Prénom
- Synthèse de que vous avez compris de cette séance
- Ce que vous attendez de ce cours
- Vos aspirations, poursuites d'études, métiers, etc.

Pour la prochaine séance



<https://projecteuler.net/>

Au moins exercices 1 & 2

Python : connaître les réponses aux questions suivantes

Lecture du chapitre 1

- Comment définit-on une fonction ?
- Quelle est la différence entre une liste, un tuple, un set, un dictionnaire ?
- Qu'est-ce que numpy et un numpy array ?
- Comment est défini un vecteur dans numpy ? une matrice ?