

Université Paris-Est Marne-la-Vallée
08/10/2015

Traitement et visualisation de données ouvertes



Philippe Gambette

3° étape de la révolution numérique ?

**Révolution
informatique**

**Révolution
internet**

**Révolution
de la donnée**



La révolution des données

« data scientist » :
informatique,
mathématiques,
stratégie

**Révolution
informatique**

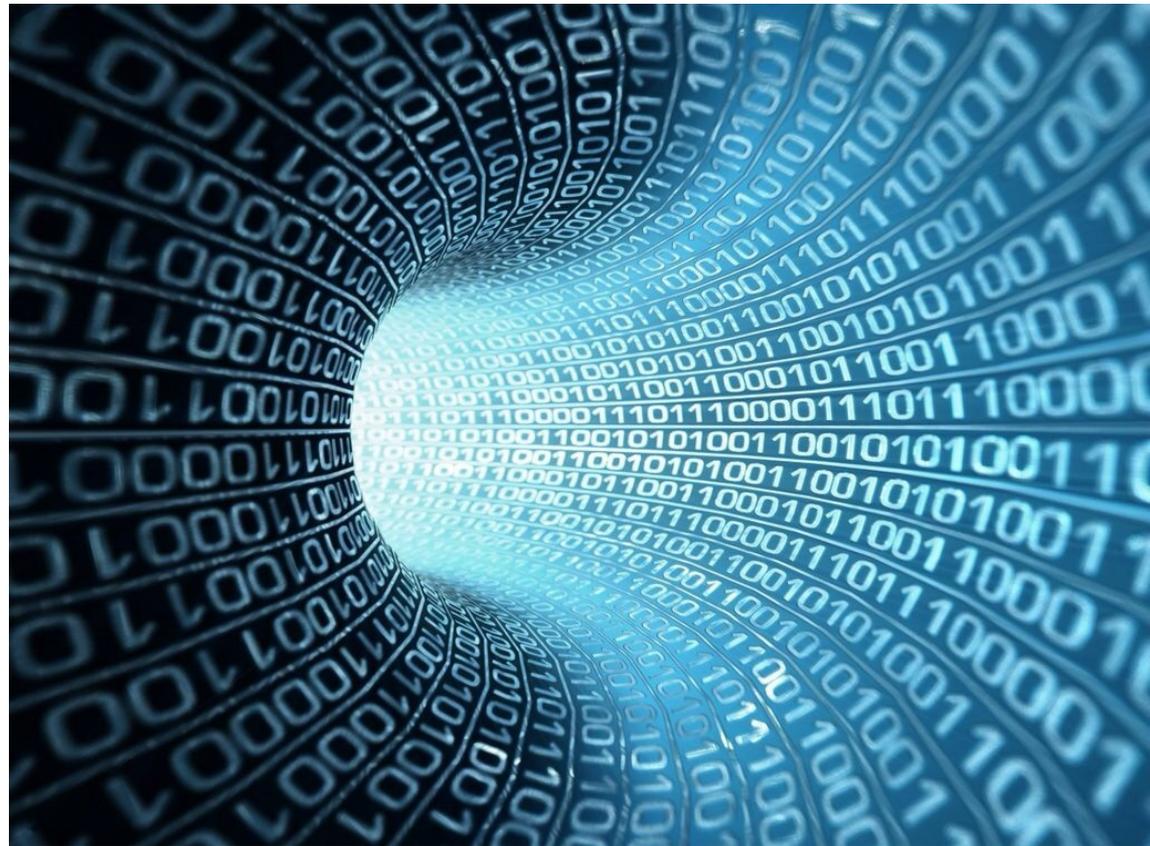
**Révolution
internet**

**Révolution
de la donnée**



Des données de plus en plus :

- accessibles
- réutilisables
- stockables
- ... traitables !



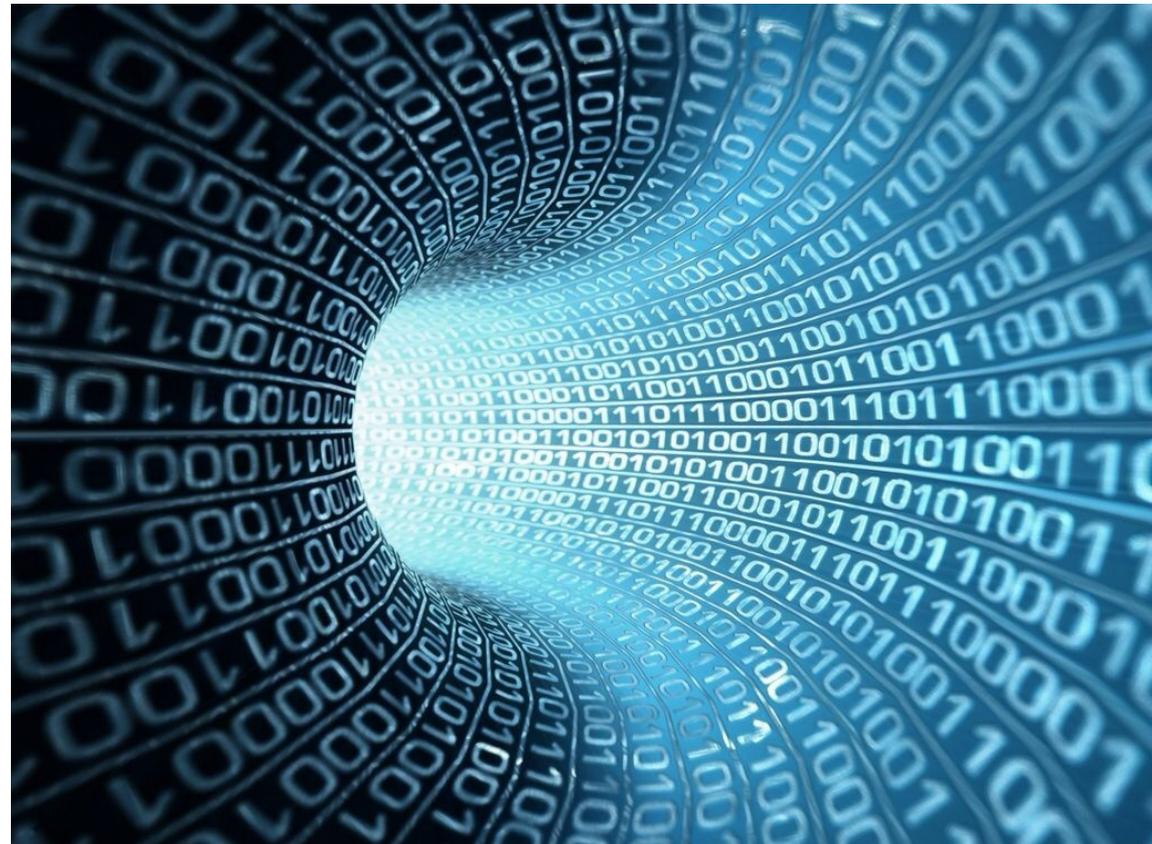
Données ouvertes, «*open data*»

Des données de plus en plus :

- accessibles
- réutilisables
- stockables
- ... traitables !



data.gouv.fr



Source : http://www.economie.gouv.fr/files/eco_numerique2.png



Henri Verdier, Chief Data Officer français,
directeur d'Etalab (<https://www.etalab.gouv.fr/>)

Des données sur :

- **Google Maps : « mashups »**
- **Open Street Map**

Des données sur :

- **Google Maps : « mashups »**
- **Open Street Map**
- **Base Adresse Nationale (en open data) :**
<http://adresse.data.gouv.fr/>



adresse.data.gouv.fr

BÊTA

INFOS DONNÉES CONTRIBUER OUTILS ACTU

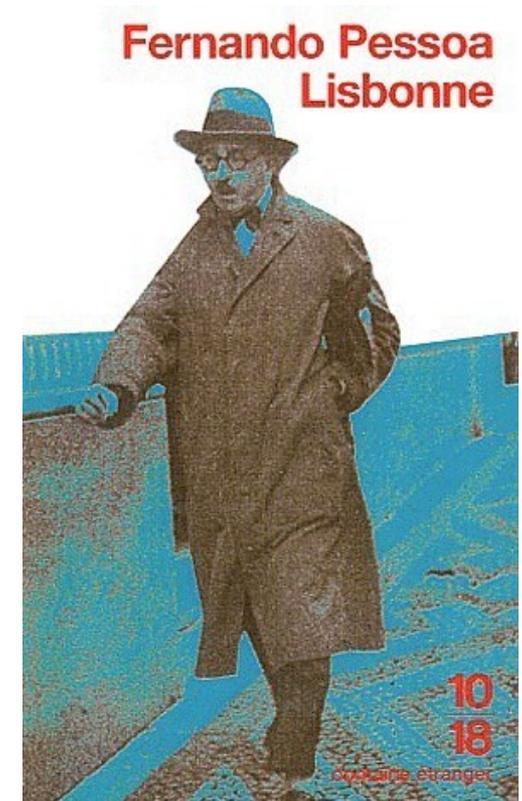
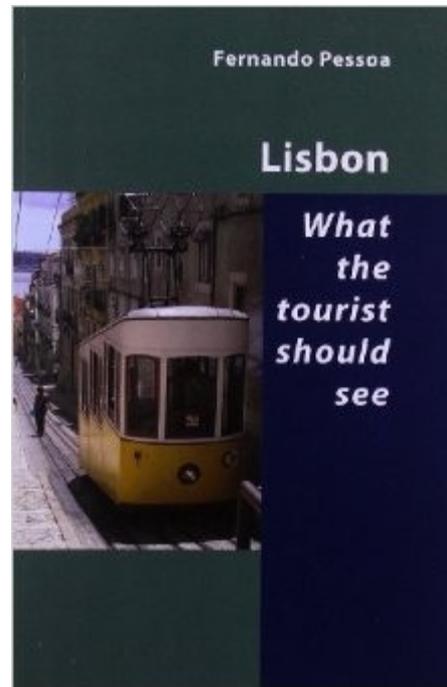
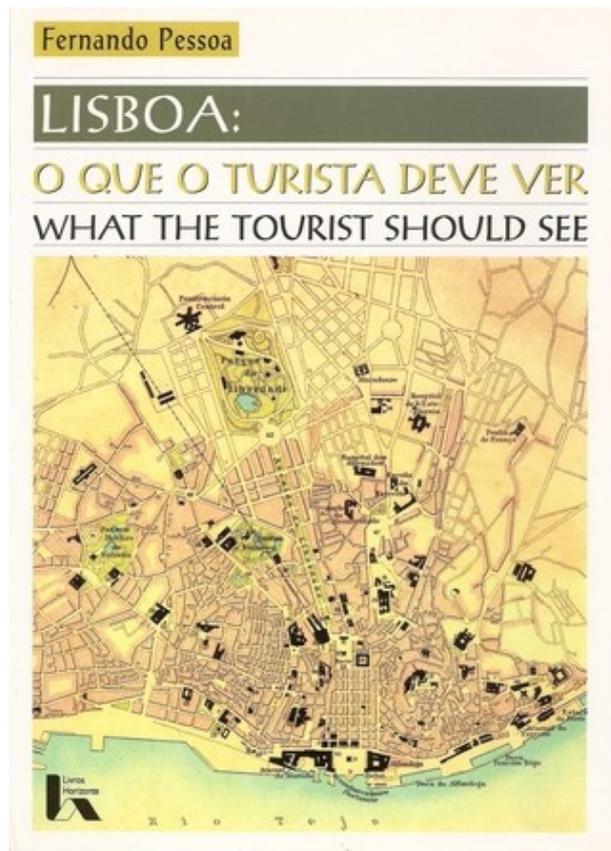
LA BASE ADRESSE NATIONALE

Un référentiel national ouvert : de l'adresse à la coordonnée géographique

Données géographiques

Géolocalisation de *Lisbonne* par Pessoa

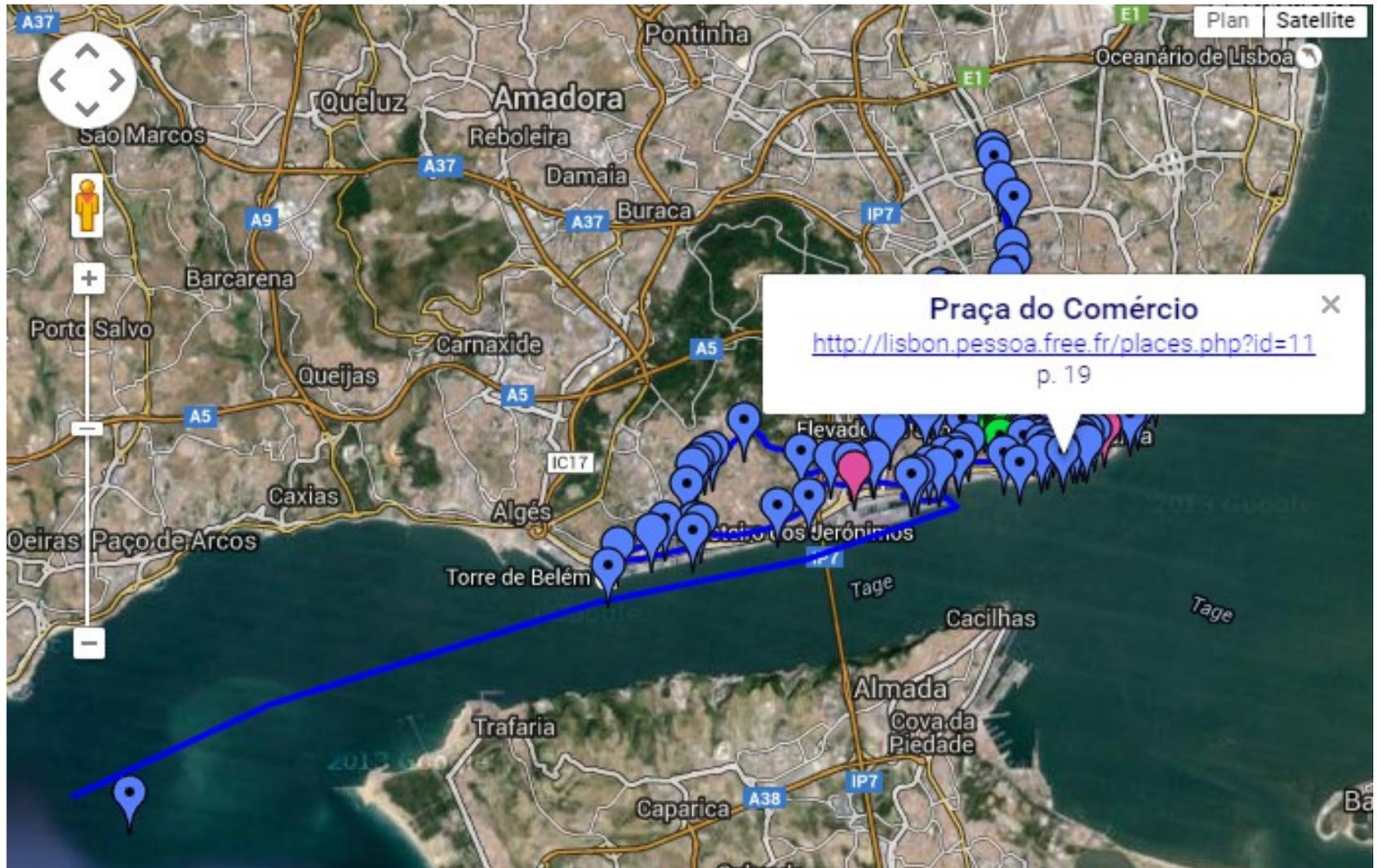
Guide touristique écrit en 1925 par Fernando Pessoa, en anglais



<http://lisbon.pessoa.free.fr>

Géolocalisation de *Lisbonne* par Pessoa

Géolocalisation manuelle Google Maps :



Géolocalisation automatique Google Maps :

On the North side of the square, facing the river, there are three parallel streets; the middle one issues from a magnificent [triumphal arch](#)¹² of great dimensions, indubitably one of the largest ones in Europe. It is dated 1873, but it was designed by Veríssimo José da Costa and began to be built in 1755.



The allegoric group which crowns the arch, sculptured by Calmels, personifies Glory crowning Genius and Valour; and the recumbent figures, which represent the rivers Tagus and Douro, as well as the statues of Nun'Alvares, Viriato, Pombal and Vasco da Gama, are due to the sculptor Victor Bastos.

The Terreiro do Paço is one of the places where boats are taken to cross the river; on the right-hand side, facing the river and on it, is the provisional station of the Southern Railways. It also often happens that tourists land here, as commonly do the crews of foreign men-of-war which visit the port. There is also a carriage and motor-car stand in this square. The general aspect of the square is of a kind to give a very agreeable impression to the most exacting of tourists.



From the Praça do Comercio we can go on to the centre of city by any of the three streets which go North from there - [Rua do Ouro](#)¹³ on the left, Rua Augusta (the one with the arch) in the middle, and Rua da Prata on the right. Let us choose Rua do Ouro, which, owing to its commercial importance is the main street of the city. There are several banks, restaurants, and shops of all kinds in this street; many of the shops, especially towards the upper end of the artery will be found to be as luxurious as their Parisian equivalents.

Almost at the upper end of the street, on the left-hand side as we go up, there is the [Santa Justa Elevator](#)¹⁴, so called because the transversal street in which it is built is called Rua de Santa Justa. This is one of the "sights" of Lisbon and always compels great admiration from tourists from everywhere.



It is due to a French engineer, Raoul Mesnier, to whom other interesting projects are also due. The elevator is all built in iron, but it is extremely distinctive, light and safe. There are two lifts, worked by electricity. It goes up to Largo do Carmo, where there are the ruins of [Carmo Church](#)¹⁵, now the Archeological Museum. Authority is needed to go right up to the top, above where the lifts themselves stop; from there a magnificent panorama is got of the whole city and the river. The elevator belongs to the Electric Tramway Company.



Géolocalisation de *Lisbonne* par Pessoa

Géolocalisation automatique Google Maps :



river, there are three parallel streets; the middle one has the greatest dimensions, indubitably one of the largest squares designed by Verissimo José da Costa and began to



which, sculptured by Calmels, personifies Glory. The figures, which represent the rivers Tagus and Douro, Pombal and Vasco da Gama, are due to the

where boats are taken to cross the river; on the right-hand side, facing the river are the main Railway. It also often happens that tourists land here, as commonly do the boats. There is also a carriage and motor-car stand in this square. The general impression is one of a most exacting of tourists.

So we can go on to the centre of city by any of the three streets which go North from there - [Rua do Ouro](#)¹³ on the left, Rua Augusta (the one with the arch) in the middle, and Rua da Prata on the right. Let us choose Rua do Ouro, which, owing to its commercial importance is the main street of the city. There are several banks, restaurants, and shops of all kinds in this street; many of the shops, especially towards the upper end of the artery will be found to be as luxurious as their Parisian equivalents.



Almost at the upper end of the street, on the left-hand side as we go up, there is the [Santa Justa Elevator](#)¹⁴, so called because the transversal street in which it is built is called Rua de Santa Justa. This is one of the "sights" of Lisbon and always compels great admiration from tourists from everywhere.



It is due to a French engineer, Raoul Mesnier, to whom other interesting projects are also due. The elevator is all built in iron, but it is extremely distinctive, light and safe. There are two lifts, worked by electricity. It goes up to Largo do Carmo, where there are the ruins of [Carmo Church](#)¹⁵, now the Archeological Museum. Authority is needed to go right up to the top, above where the lifts themselves stop; from there a magnificent panorama is got of the whole city and the river. The elevator belongs to the Electric Tramway Company.



Géolocalisation de *Lisbonne* par Pessoa

Géolocalisation automatique Google Maps :



river, there are three parallel streets; the middle one is of great dimensions, indubitably one of the largest in the city. It was designed by Verissimo José da Costa and began to

which, sculptured by Calmels, personifies Glory and the three figures, which represent the rivers Tagus and Douro, Pombal and Vasco da Gama, are due to the

where boats are taken to cross the river; on the right-hand side, there are the main Railways. It also often happens that tourists land here, as a point of departure. There is also a carriage and motor-car stand in this square. It has a very agreeable impression to the most exacting of tourists.

to go we can go on to the centre of city by any of the three streets which

North from there - [Rua do Ouro](#)¹³ on the left, Rua Augusta (the one with the arch) in the middle, and Rua da Prata on the right. Let us choose Rua do Ouro, which, owing to its commercial importance is the main street of the city. There are several banks, restaurants, and shops of all kinds in this street; many of the shops, especially towards the upper end of the artery will be found to be as luxurious as their Parisian equivalents.



Almost at the upper end of the street, on the left-hand side as we go up, there is the [Santa Justa Elevator](#)¹⁴, so called because the transversal street in which it is built is called Rua de Santa Justa. This is one of the "sights" of Lisbon and always compels great admiration from tourists from everywhere.



It is due to a French engineer, Raoul Mesnier, to whom other interesting projects are also due. The elevator is all built in iron, but it is extremely distinctive, light and safe. There are two lifts, worked by electricity. It goes up to Largo do Carmo, where there are the ruins of [Carmo Church](#)¹⁵, now the Archeological Museum. Authority is needed to go right up to the top, above where the lifts themselves stop; from there a magnificent panorama is got of the whole city and the river. The elevator belongs to the Electric Tramway Company.



Base de données
MySQL ; PHP ;
Javascript



Diagramme de Voronoi des McDos parisiens

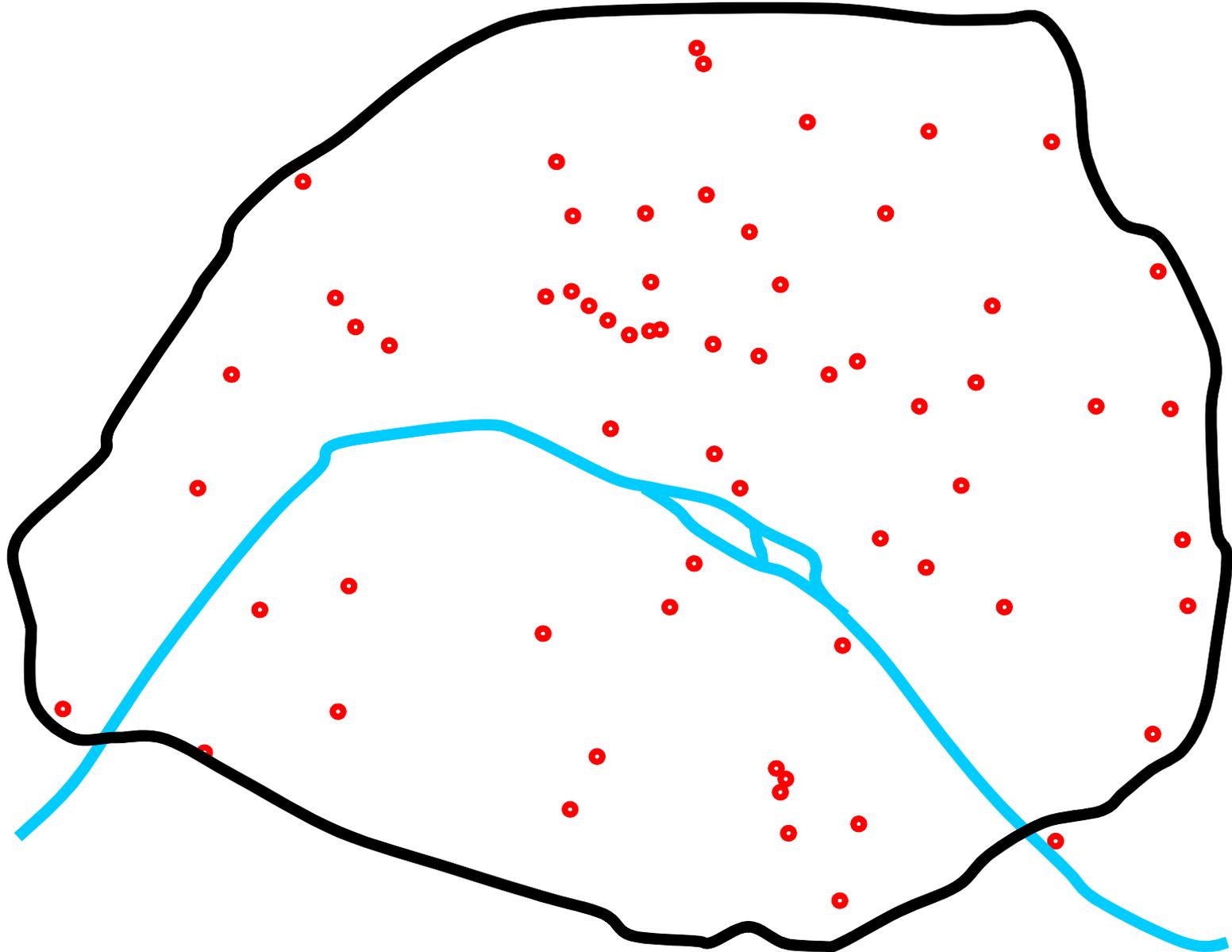


Diagramme de Voronoi des McDos parisiens

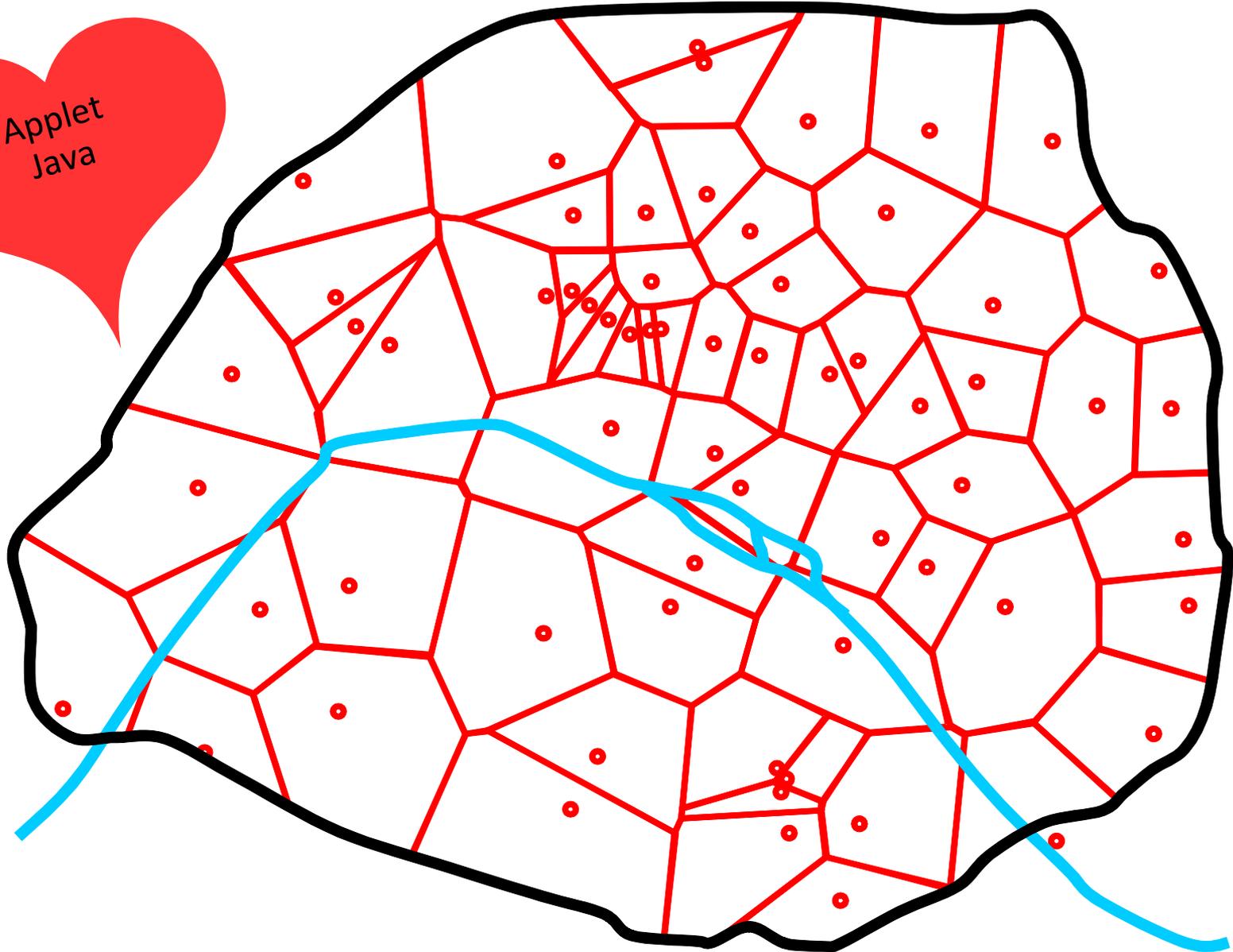
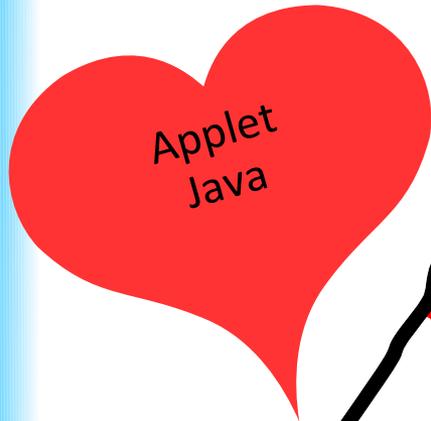


Diagramme de Voronoi des McDos parisiens

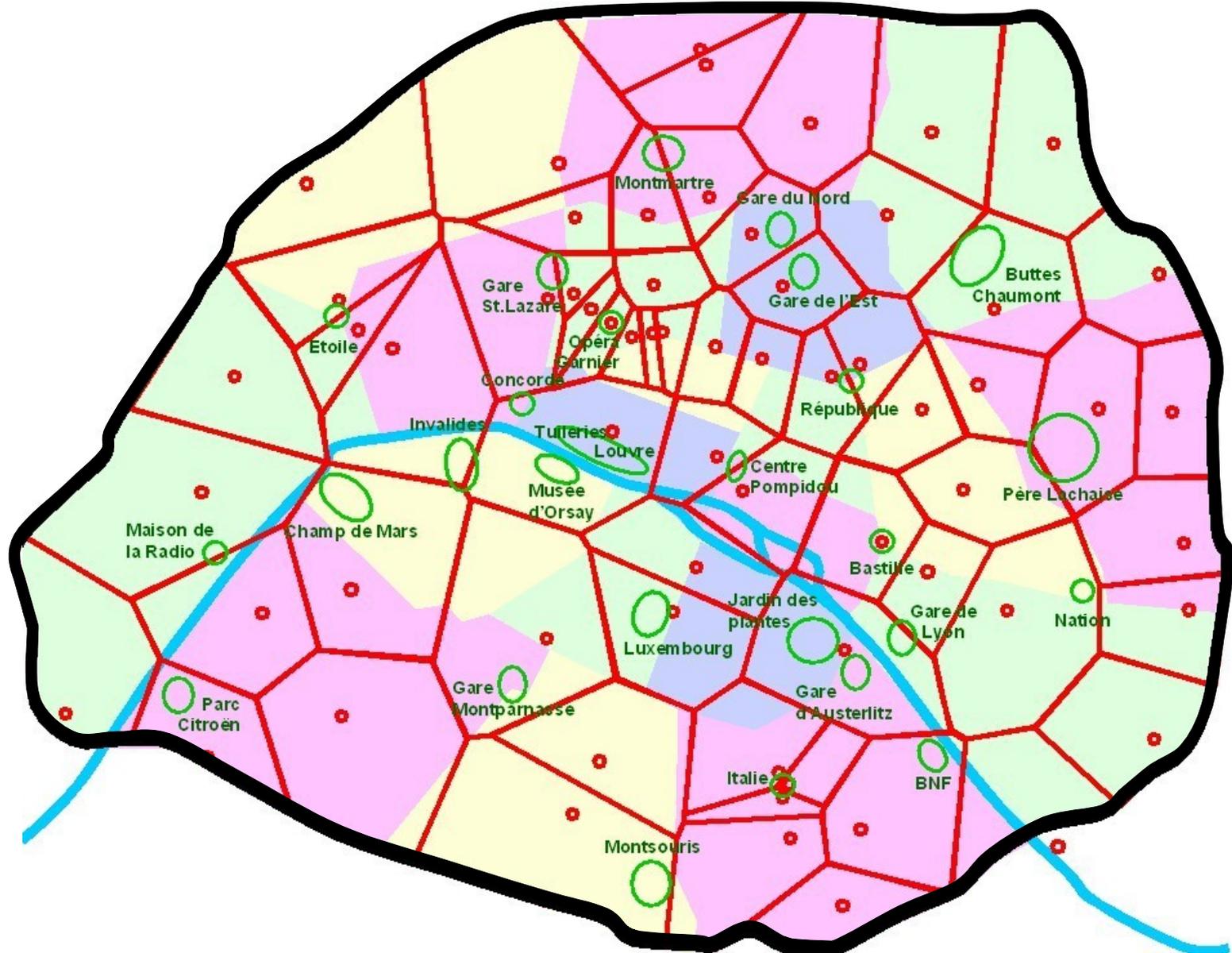
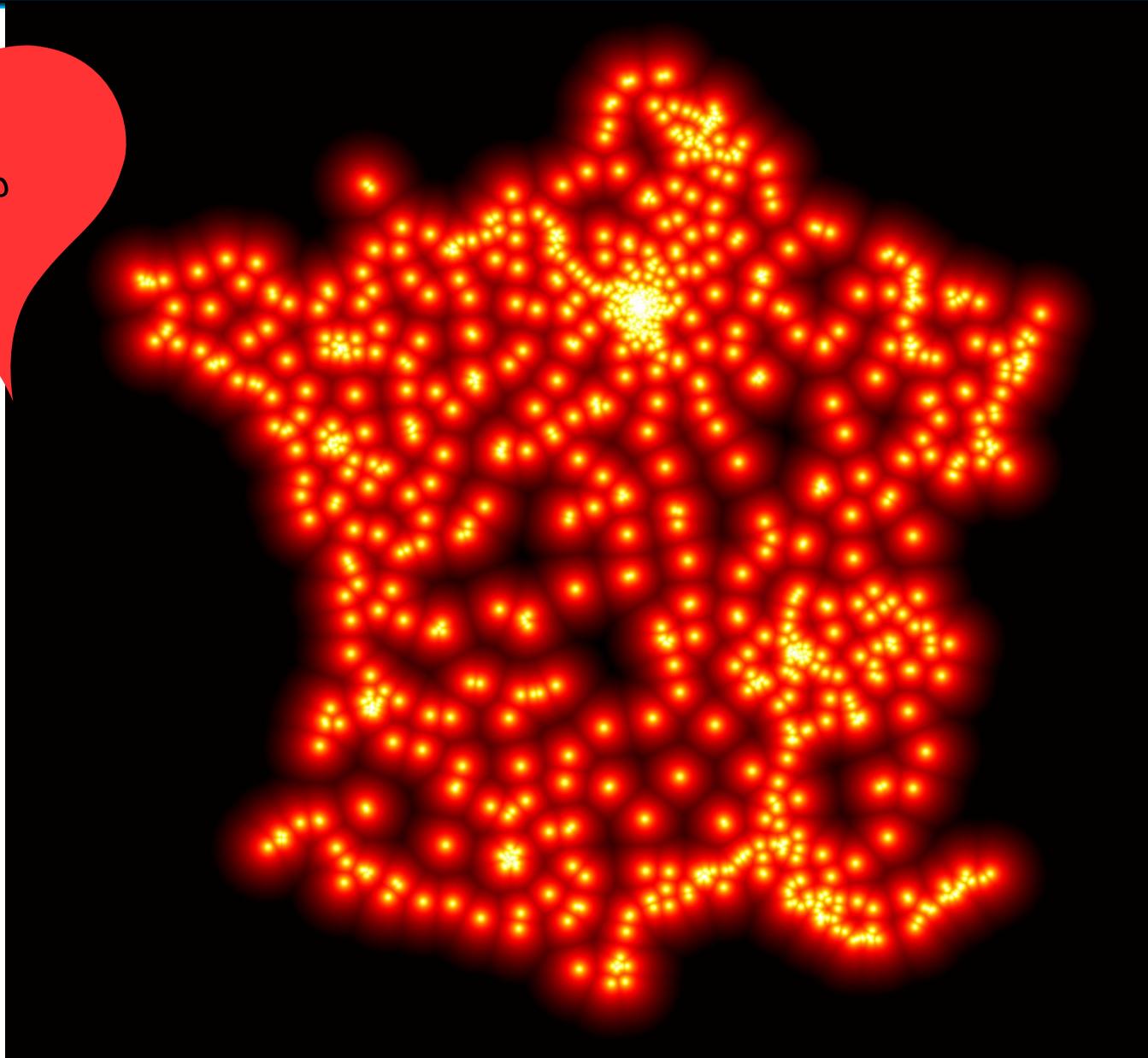
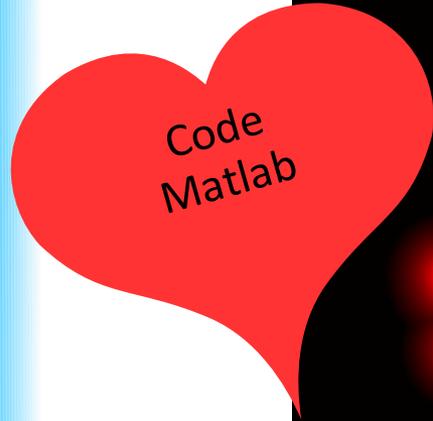
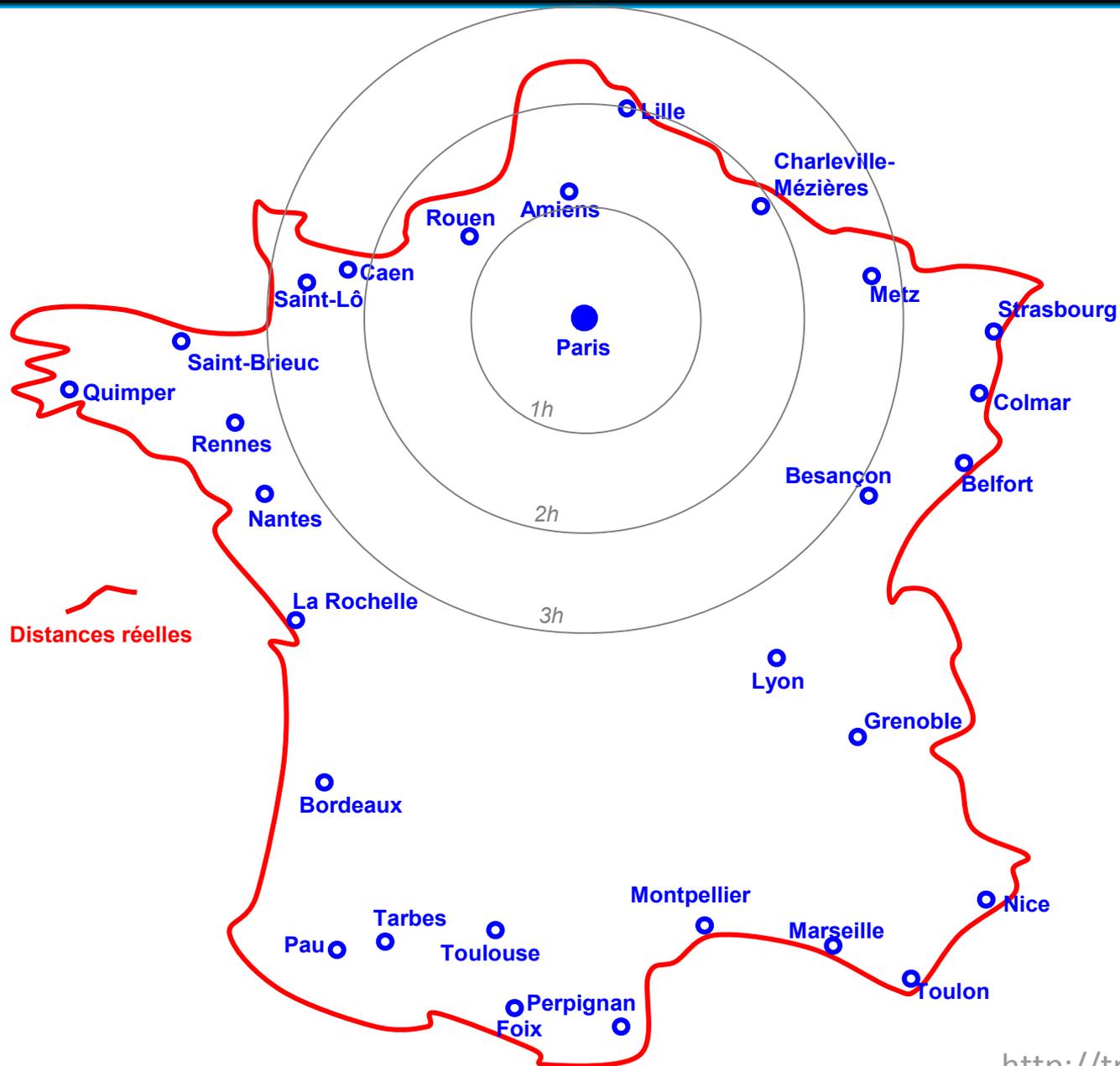


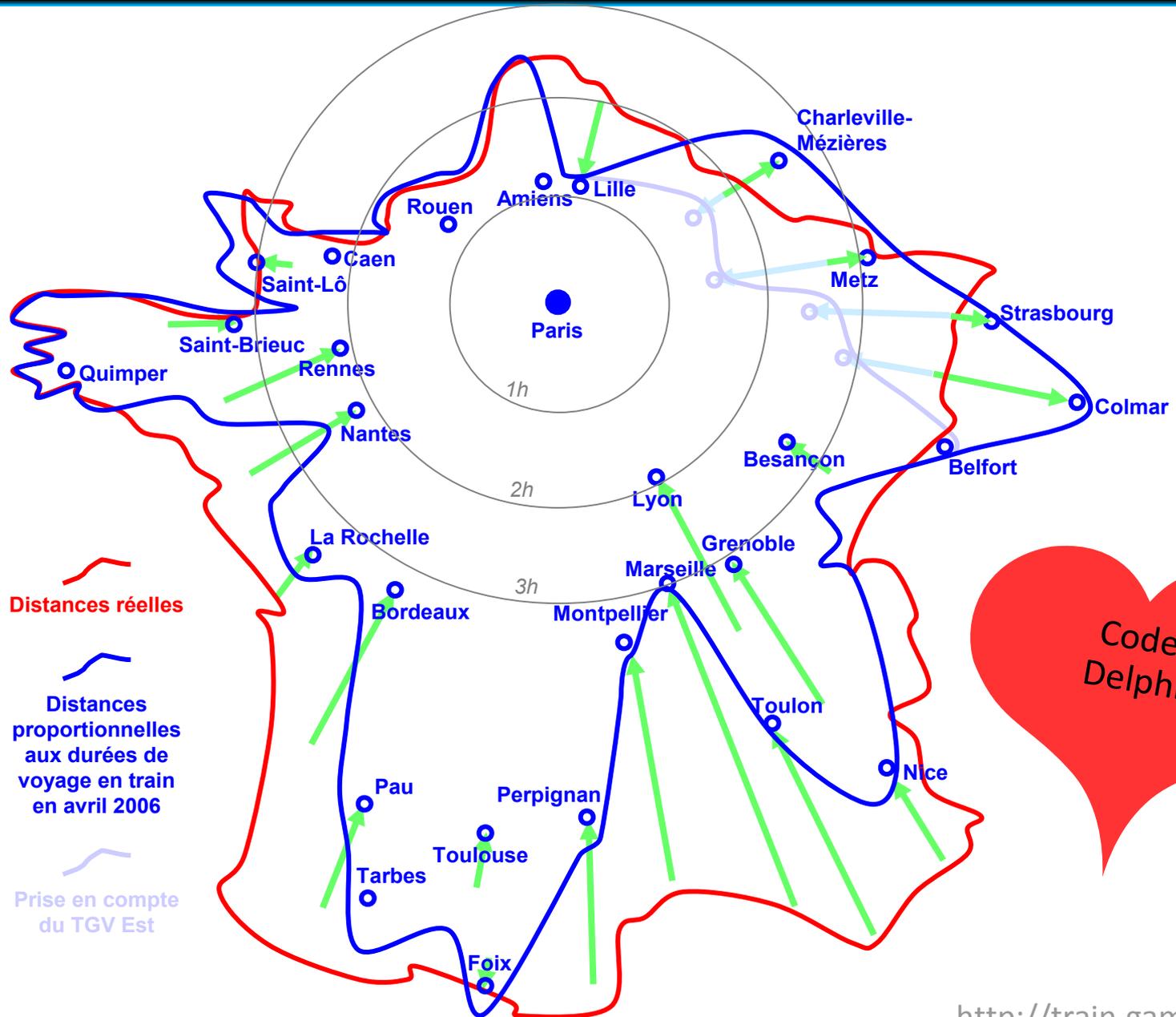
Diagramme de Voronoi des McDos français



La France en train depuis Paris

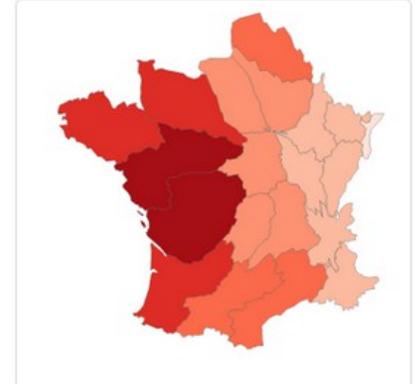
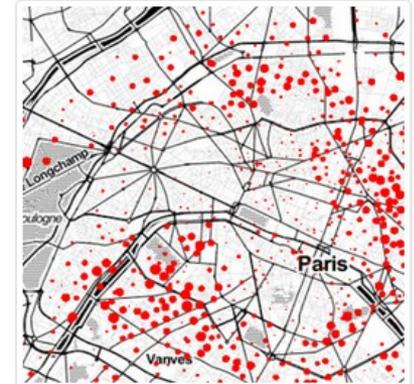
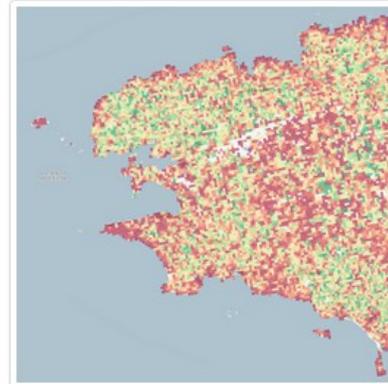
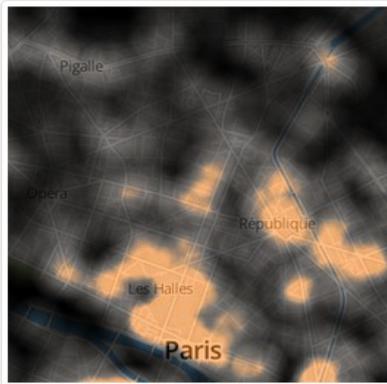


La France en train depuis Paris



D'autres traitements de données géographiques

Galerie.

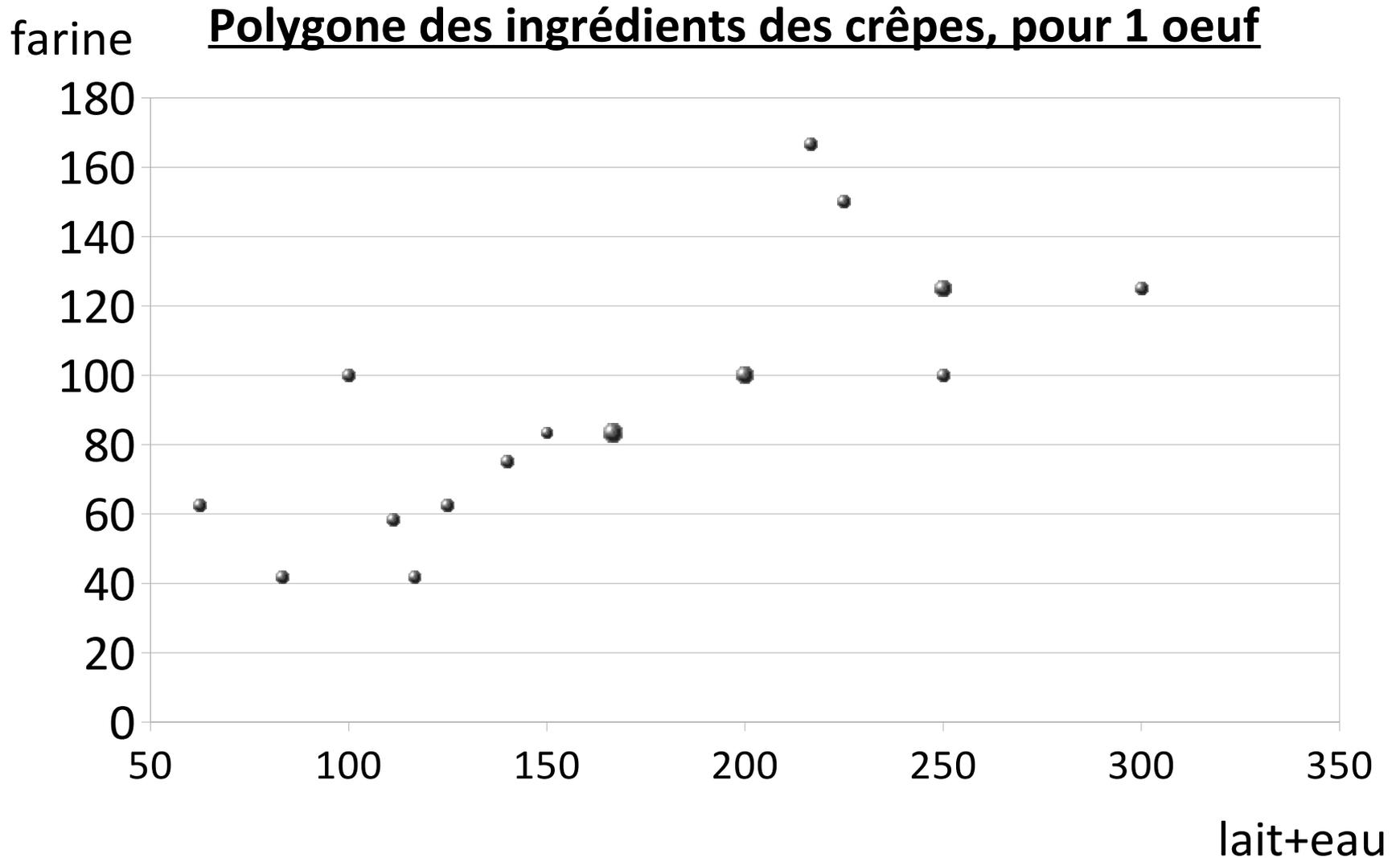


Données gastronomiques

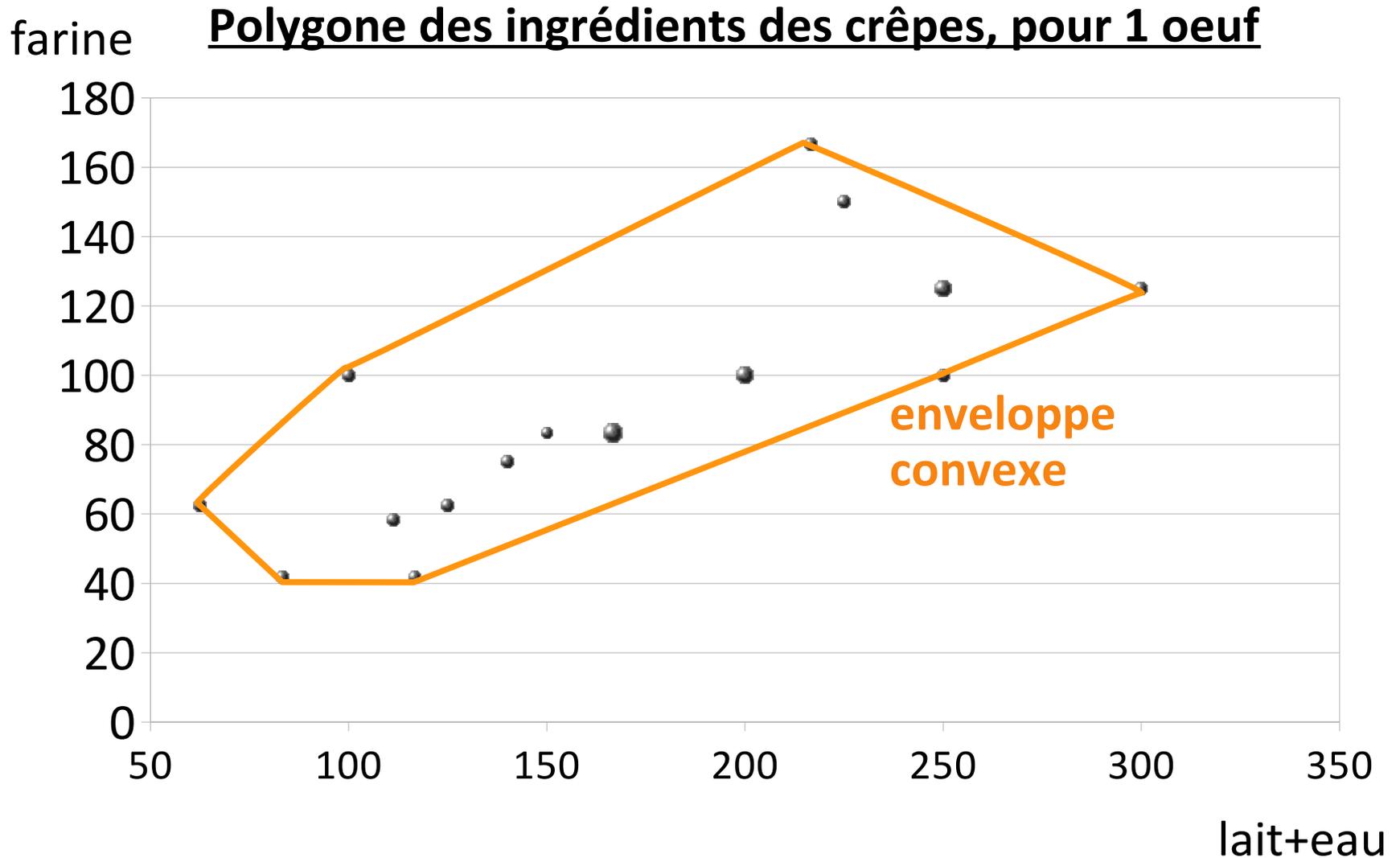
Visualisation de données de recettes de crêpes



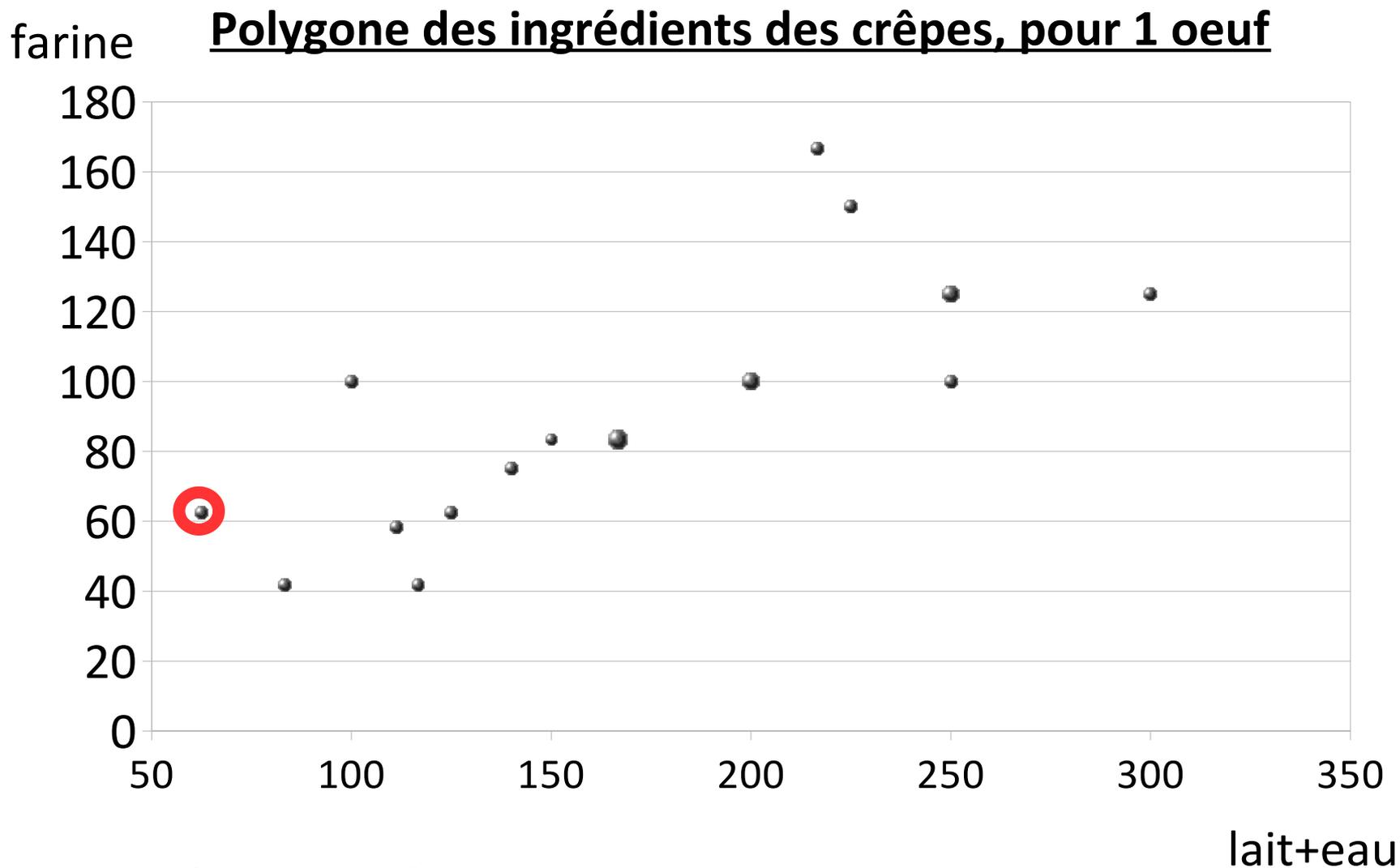
Visualisation de données de recettes de crêpes



Visualisation de données de recettes de crêpes

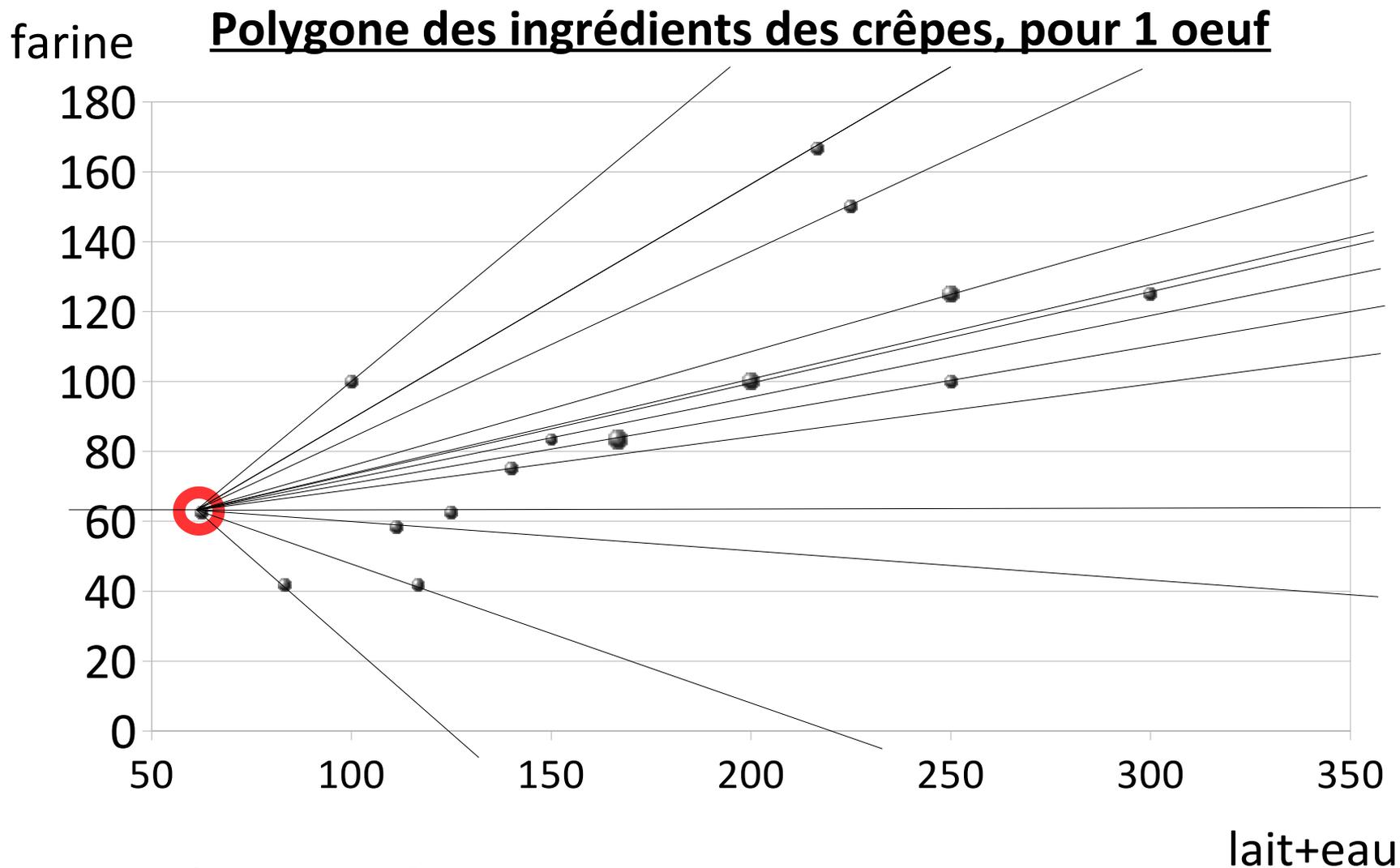


Visualisation de données de recettes de crêpes



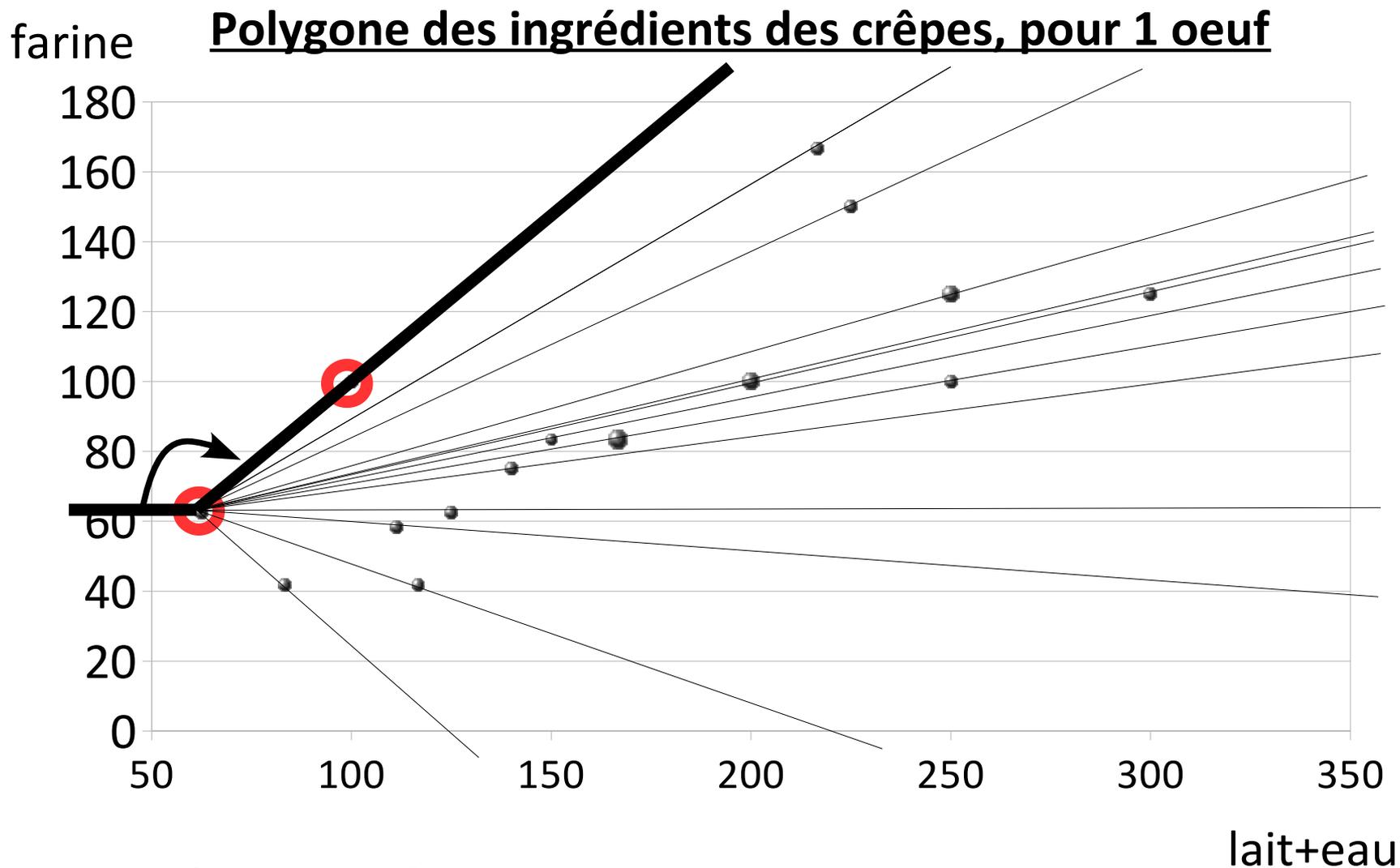
Marche de Jarvis

Visualisation de données de recettes de crêpes



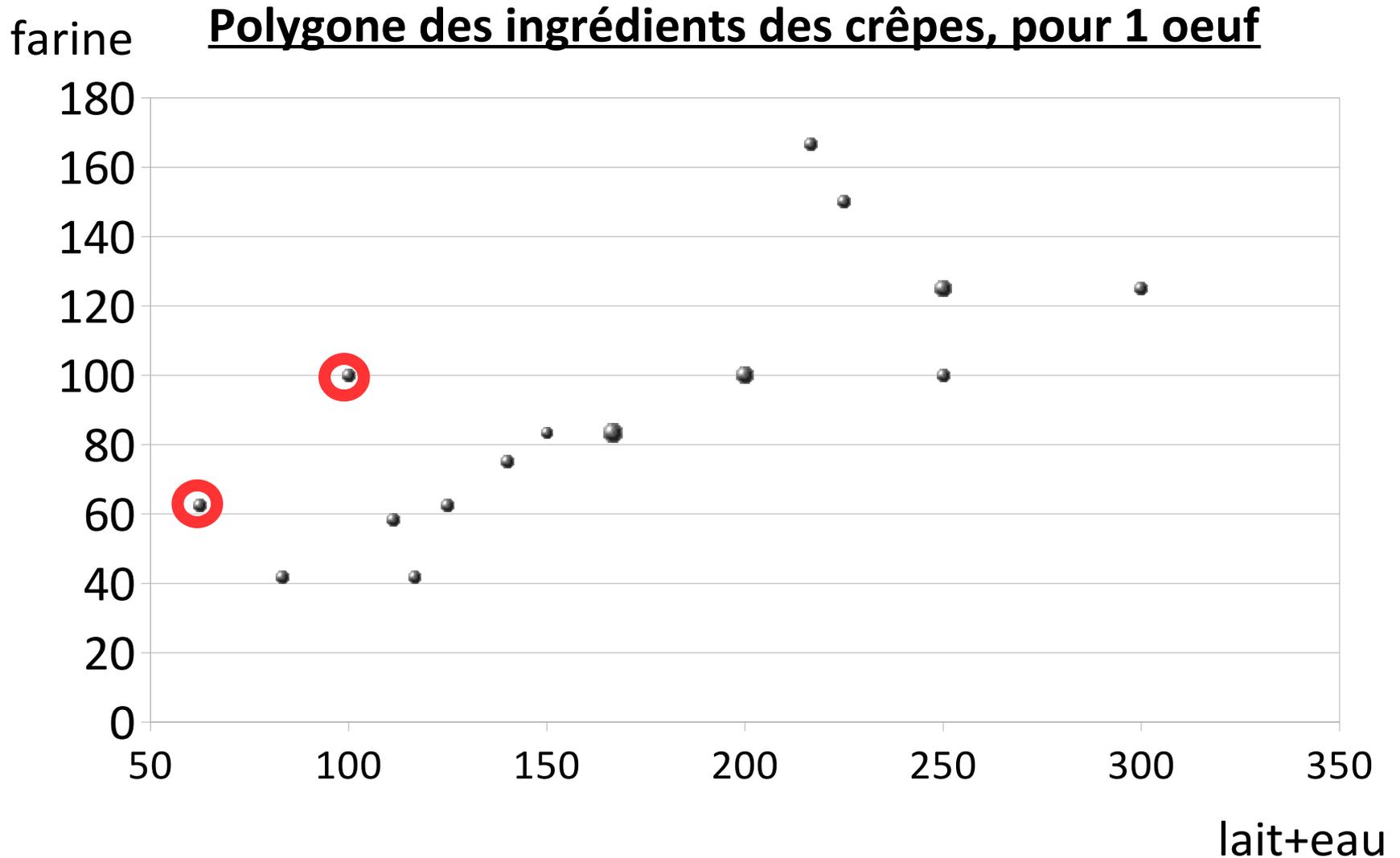
Marche de Jarvis

Visualisation de données de recettes de crêpes



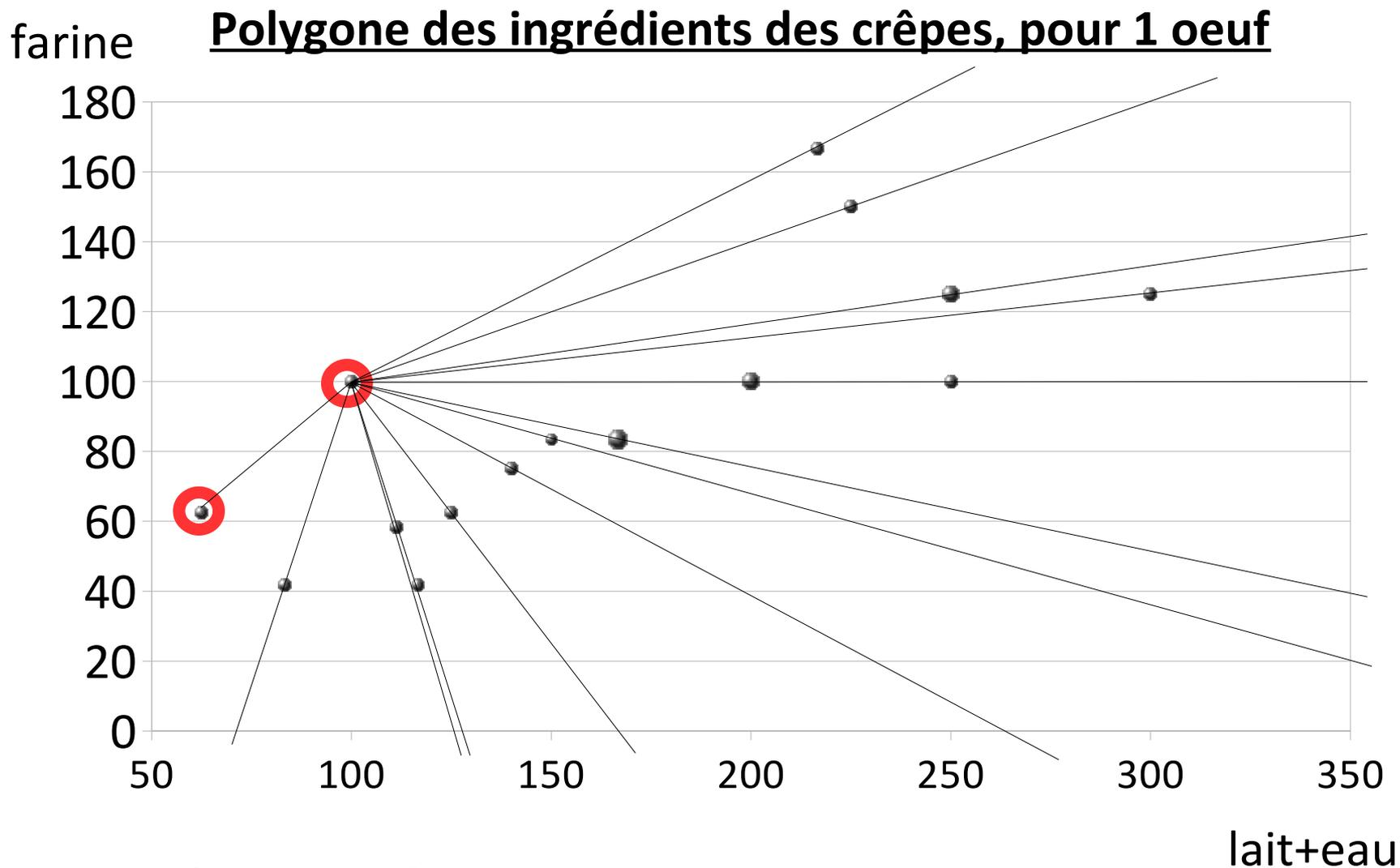
Marche de Jarvis

Visualisation de données de recettes de crêpes



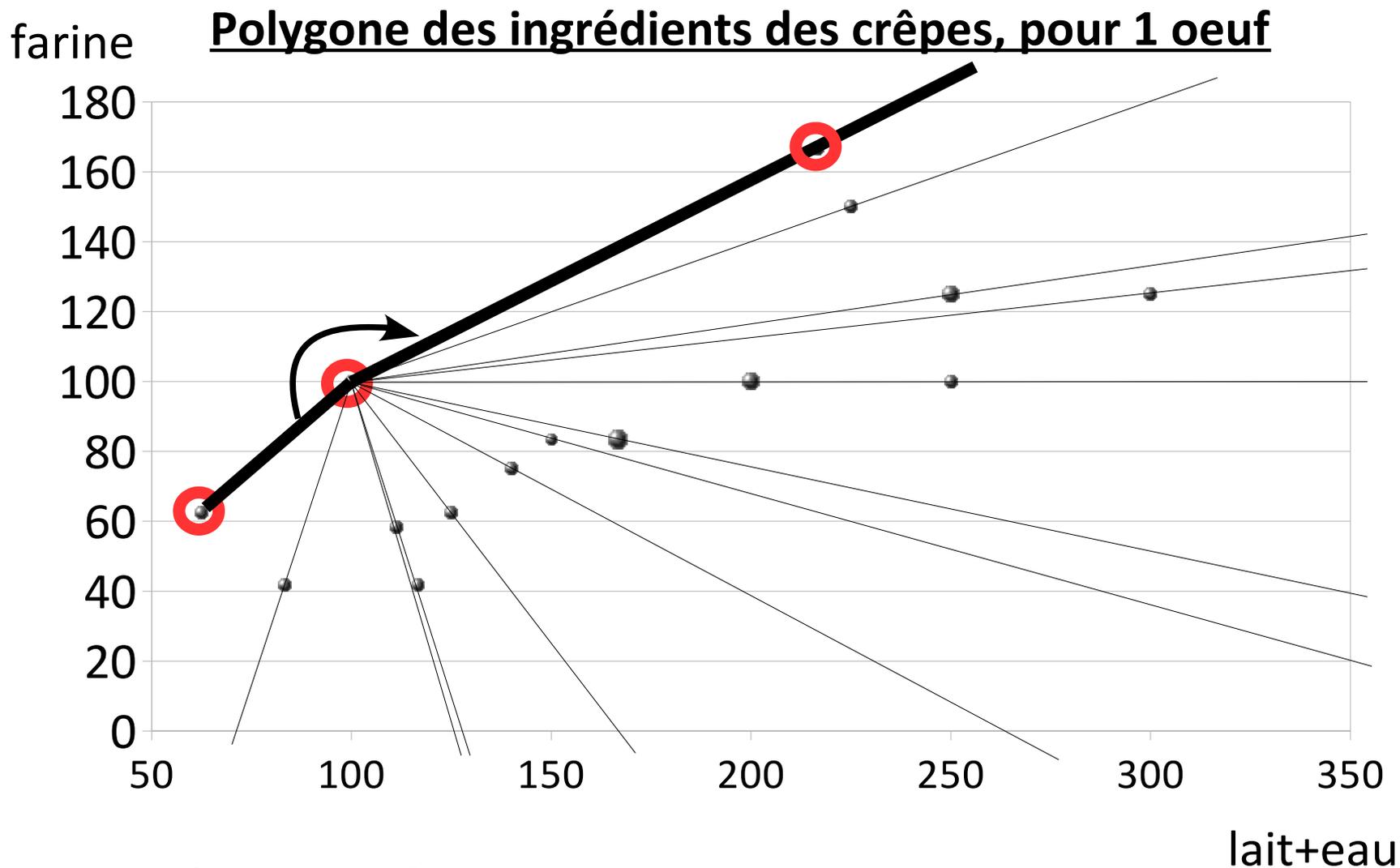
Marche de Jarvis

Visualisation de données de recettes de crêpes



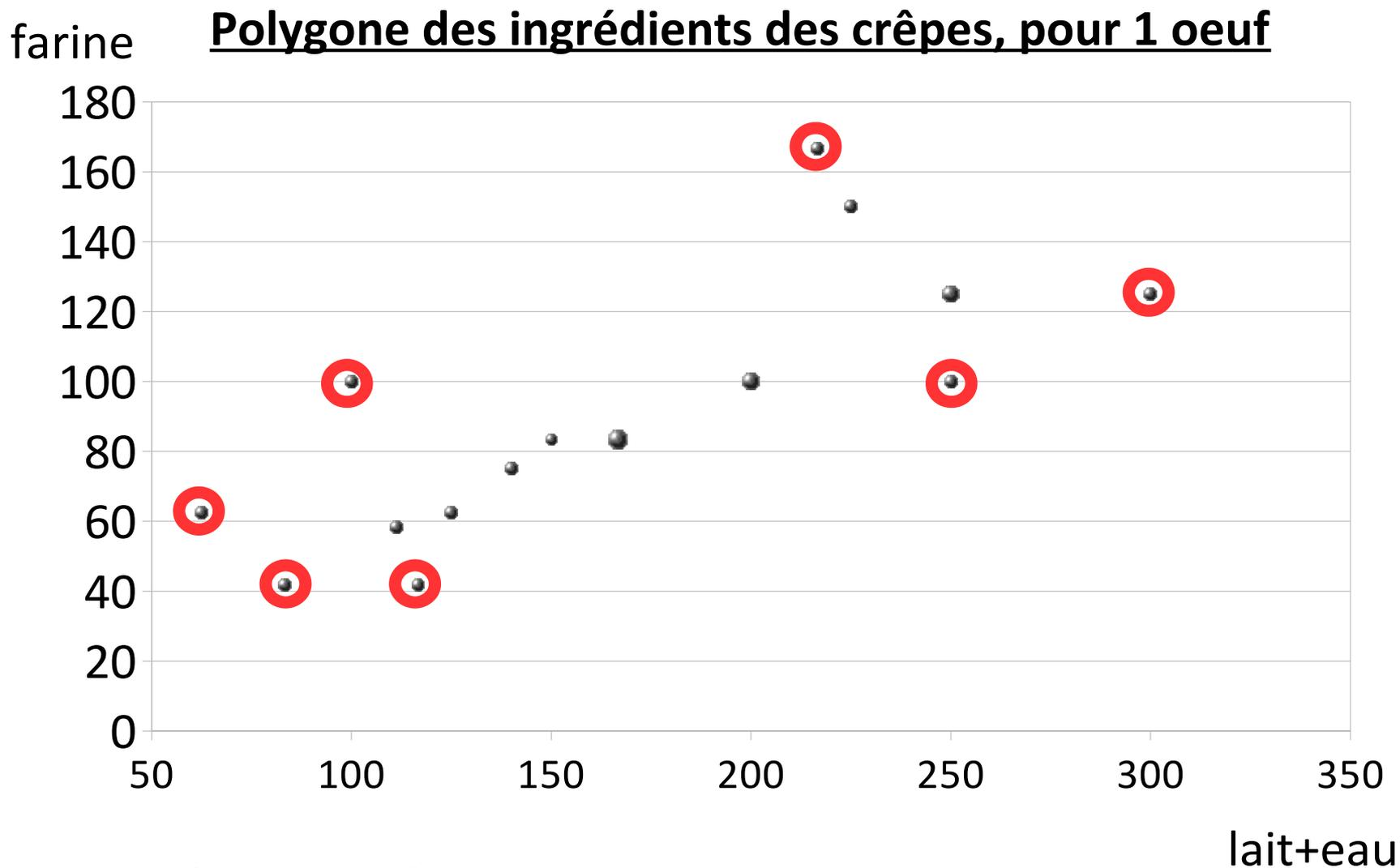
Marche de Jarvis

Visualisation de données de recettes de crêpes



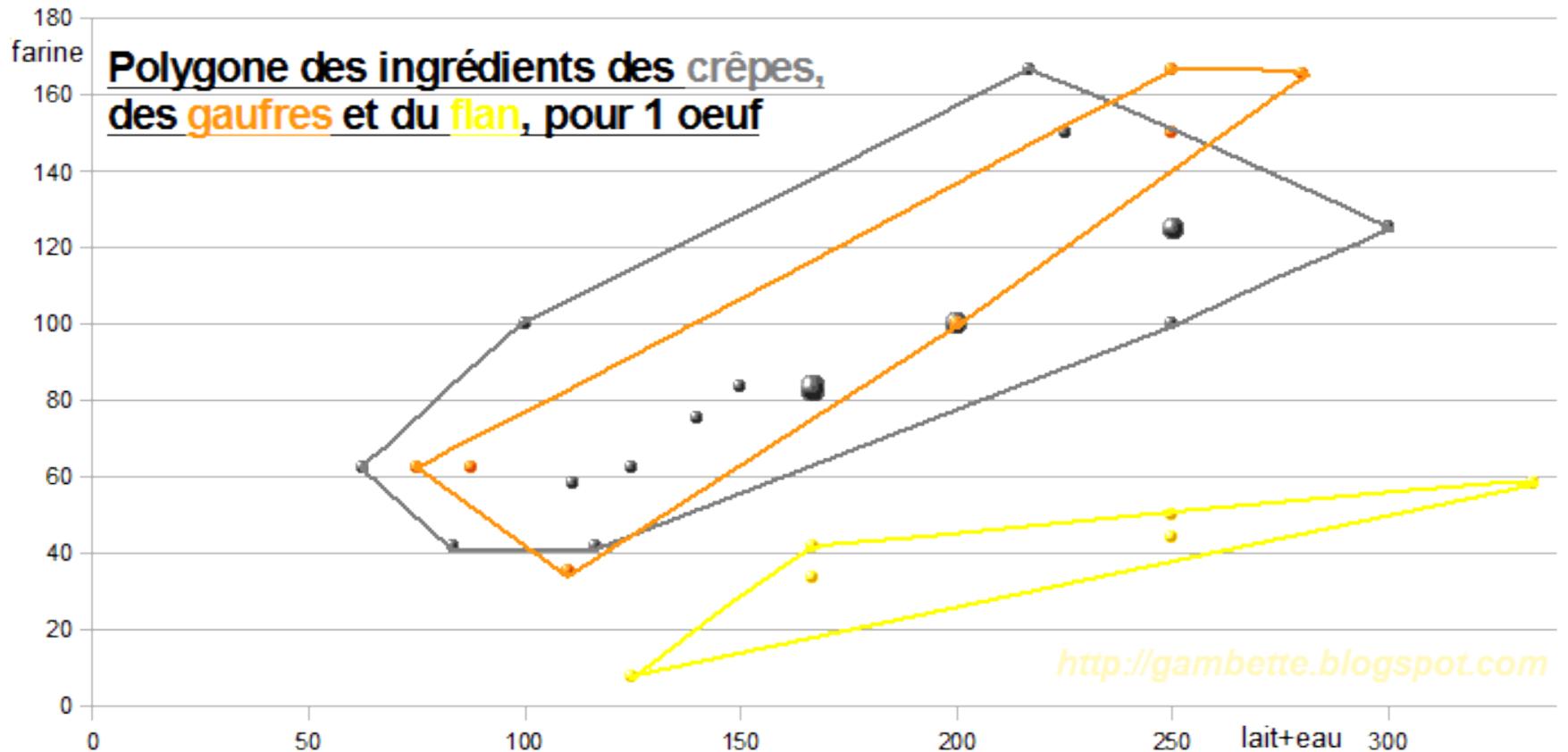
Marche de Jarvis

Visualisation de données de recettes de crêpes



Marche de Jarvis

Visualisation de données de recettes de crêpes



A vertical blue bar with a gradient is located on the left side of the slide.

Données de moteurs de recherche

Google Fight!



Suggestion de combats

Les 20 derniers combats



Select your version  

Tapez 2 mots clés et cliquez sur le bouton 'Fight !'. Le gagnant est celui qui a la meilleure visibilité sur Google. 

mathématiques

VS

informatique

FIGHT !

 Tweet 45

 Like 57

 +1 28

 Share 7

 Pin it

COMBAT DU JOUR :

FRANCE

FIGHT

IRLANDE

Google Fight!



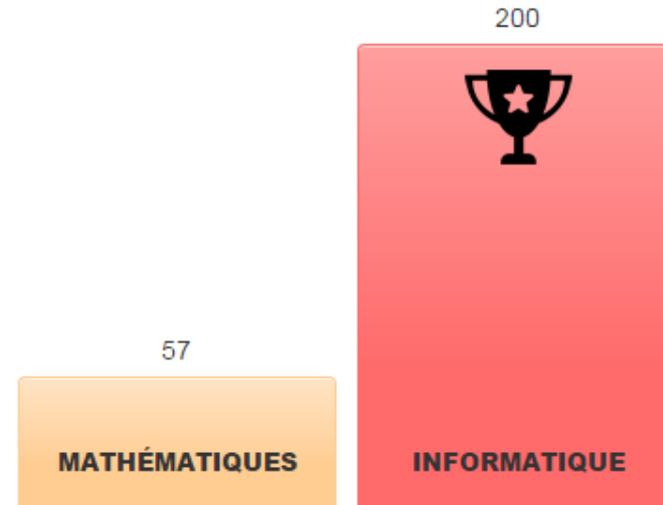
Suggestion de combats

Les 20 derniers combats



Select your version 

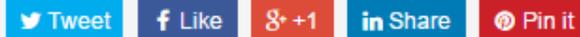
MATHÉMATIQUES vs INFORMATIQUE



Mode de calcul 

Powered by  SEMFLUSH
COMPETITIVE INTELLIGENCE

Partagez ce combat :



ESSAYEZ AUSSI CES COMBATS

Roger Federer

FIGHT

Raphal nadal

Argent

FIGHT

Bonheur

Novak Djokovic

FIGHT

Raphael Nadal

Google Fight!



Suggestion de combats

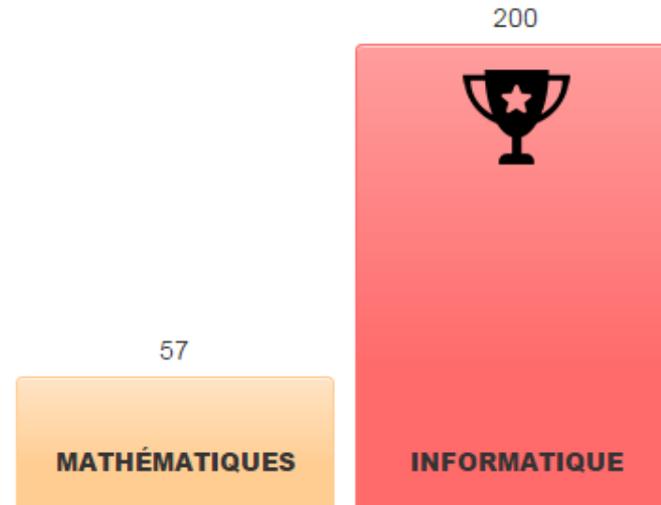
Les 20 derniers combats



Select your version  

Attention à la fiabilité !
<http://blog.veronis.fr/2005/01/web-comptes-bidons-chez-google.html?m=0>

MATHÉMATIQUES vs INFORMATIQUE



Mode de calcul 

Powered by  SEMFLUSH
COMPETITIVE INTELLIGENCE

Partagez ce combat :



ESSAYEZ AUSSI CES COMBATS

Roger Federer

FIGHT

Raphal nadal

Argent

FIGHT

Bonheur

Novak Djokovic

FIGHT

Raphael Nadal

Google Fight pour l'orthographe ?



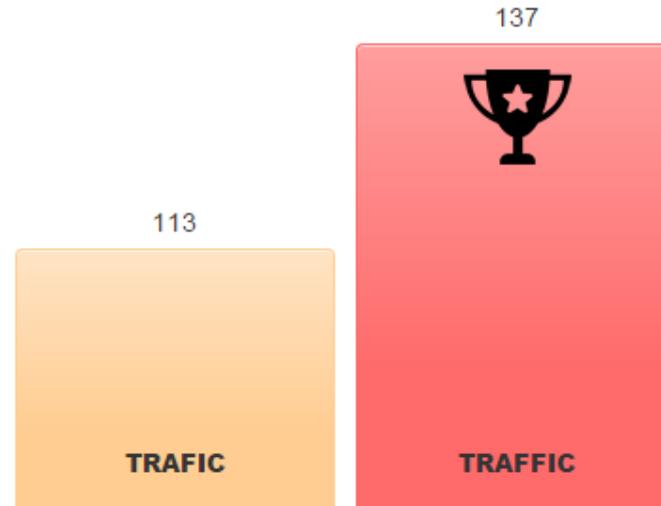
Suggestion de combats

Les 20 derniers combats



Select your version  

TRAFIC vs TRAFFIC



Mode de calcul 

Powered by 

Partagez ce combat :



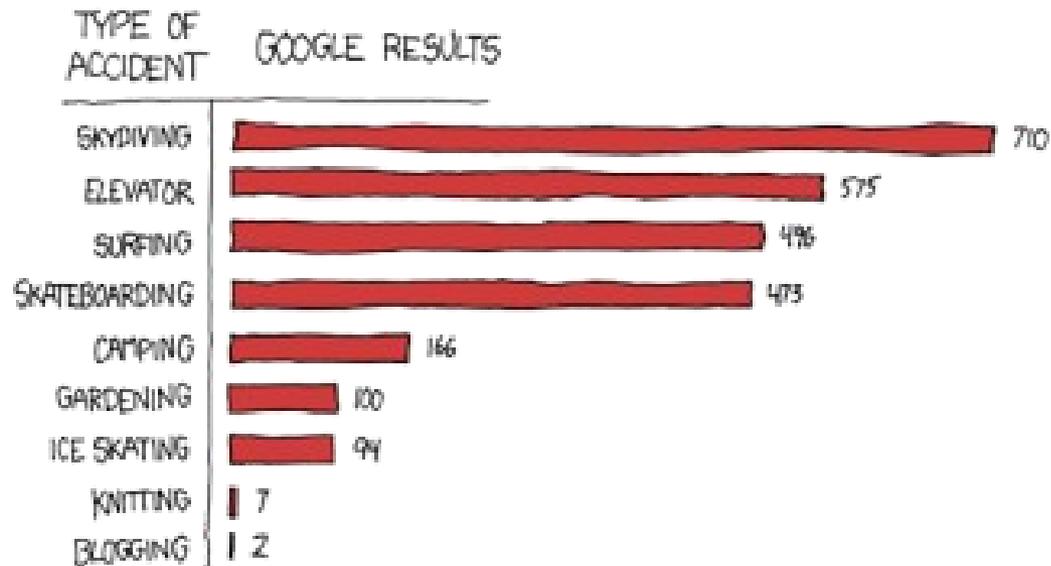
ESSAYEZ AUSSI CES COMBATS

Beyonce	FIGHT	Rihanna
Mac OS	FIGHT	Windows
Alexandrie	FIGHT	Alexandra

Google Fights : dangers !

DANGERS

INDEXED BY THE NUMBER OF GOOGLE RESULTS FOR
"DIED IN A _____ ACCIDENT"

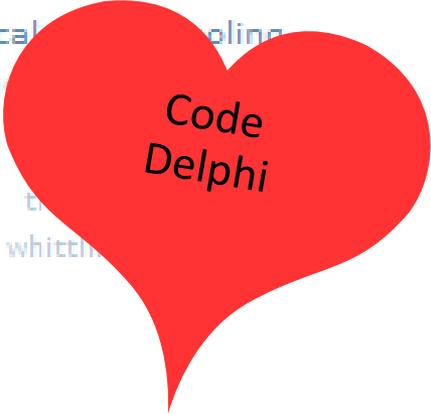


Google Fights : dangers !

acting amusing appalling **baking** ballooning bathing **biking** blasting bleaching boiling
bombing bowling breathing breathing breeding brewing **building** burning **camping** caving
cheering **choking** **climbing** cloning **coaching** coasting computing cooking covering
crafting crippling crossing crushing **cycling** dancing debugging devastating digging
disturbing **diving** drifting **drilling** drinking **driving** drumming dying embarrassing
engineering ensuing entertaining exercising **falling** fanning felling filming filtering firing
fishing flaming flaying floating **flooding** **flying** framing frightening frying gambling
gardening **gliding** guiding hanging harvesting healing **hunting** hurling interesting
jogging **jousting** juggling jumping landing lightening loading **logging** log-rolling
longing lumbering manufacturing **marching** mating melting milling **mining** moving
mowing mustering painting playing plowing potting practicing prancing praying printing
programming qualifying **racing** rallying rambling raping recycling refueling rhyming
riding **riding** rigging routing rowing running **sailing** sampling scalding schooling
seeming sharing shaving shearing shelving **shipping** shocking shopping singing sinking
skidding **skiing** skipping sliding smoking soaring speeding spinning sporting stalling
starting startling **switching** tanning terrifying **testing** thrashing touring trading tramping
trampling transplanting traveling vacationing vaulting **walking** welding whittling working
wrestling writing

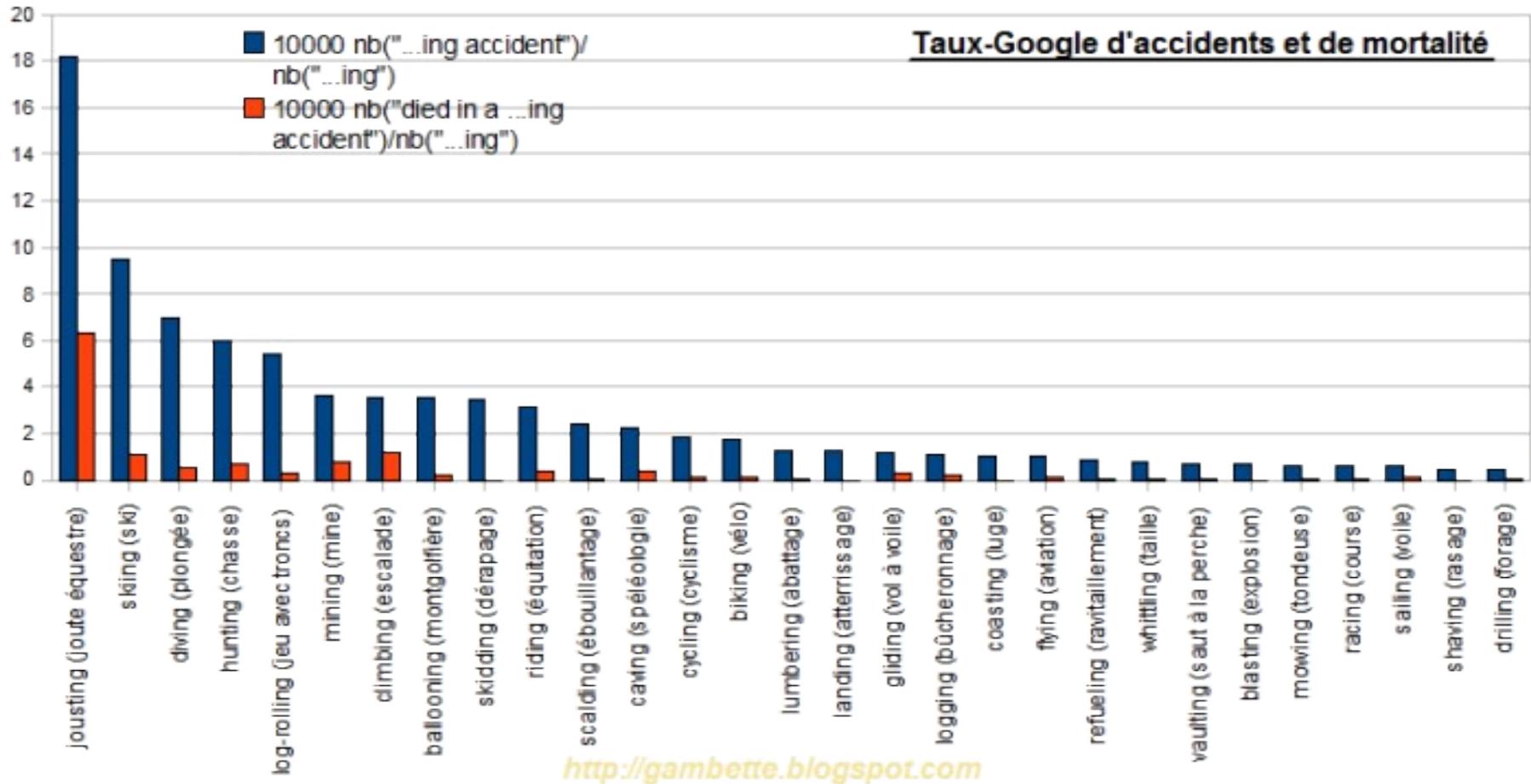
Google Fights : dangers !

acting amusing appalling **baking** ballooning bathing **biking** blasting bleaching boiling
bombing bowling breathing breathing breeding brewing **building** burning **camping** caving
cheering **choking** **climbing** cloning **coaching** coasting computing cooking covering
crafting crippling crossing crushing **cycling** dancing debugging devastating digging
disturbing **diving** drifting **drilling** drinking **driving** drumming dying embarrassing
engineering ensuing entertaining exercising **falling** fanning felling filming filtering firing
fishing flaming flaying floating **flooding** **flying** framing frightening frying gambling
gardening gliding guiding hanging harvesting healing **hunting** hurling interesting
jogging **jousting** juggling jumping landing lightening loading **logging** log-rolling
longing lumbering manufacturing **marching** mating melting milling **mining** moving
mowing mustering painting playing plowing potting practicing prancing praying printing
programming qualifying **racing** rallying rambling raping recycling refueling rhyming
riding **riding** rigging routing rowing running **sailing** sampling scaling sailing
seeming sharing shaving shearing shelving **shipping** shocking shopping
skidding **skiing** skipping sliding smoking soaring speeding spinning
starting startling **switching** tanning terrifying **testing** thrashing touring treading
trampling transplanting traveling vacationing vaulting **walking** welding whittling
wrestling writing



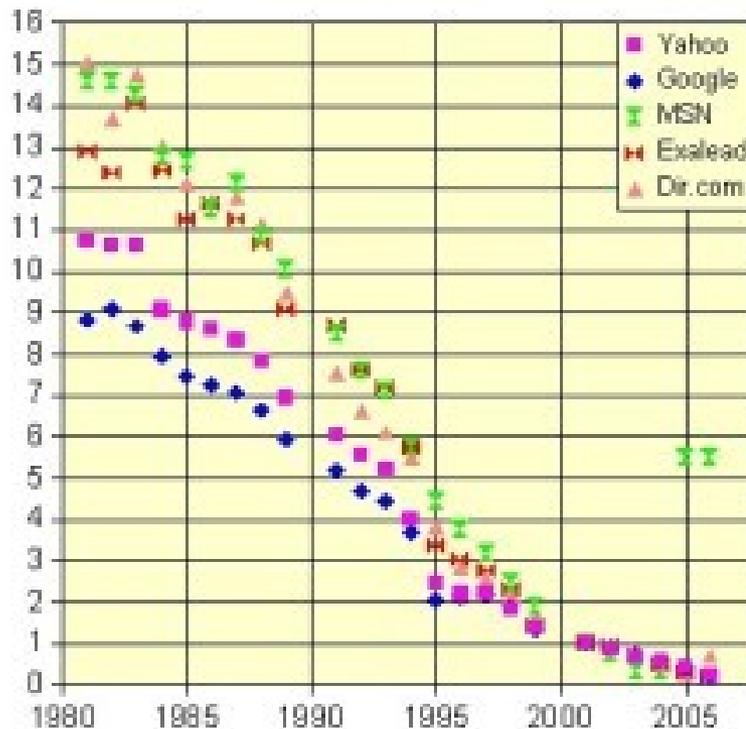
Code
Delphi

Google Fights : dangers !

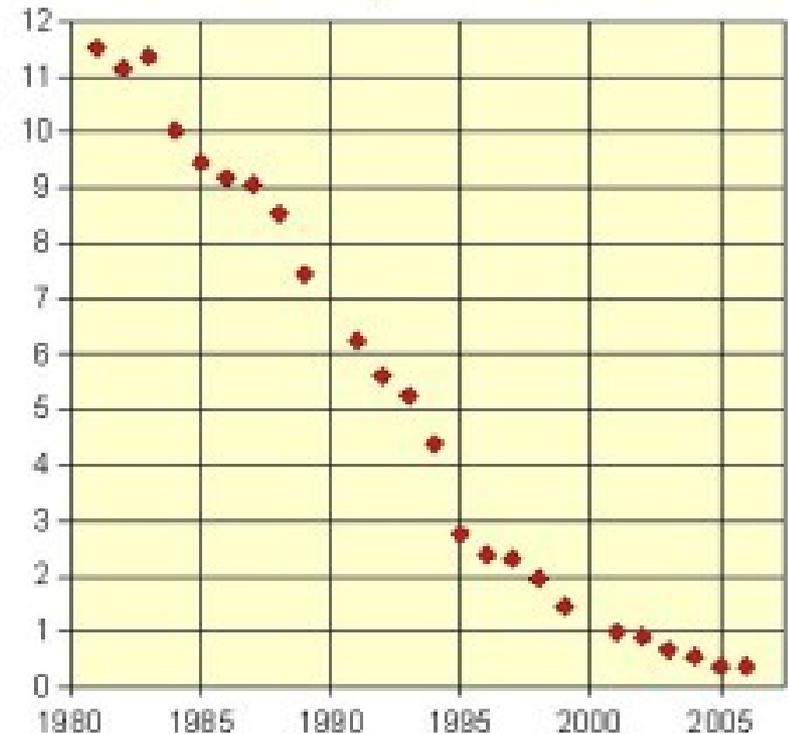


Google Fights : années

Années sur les moteurs de recherche

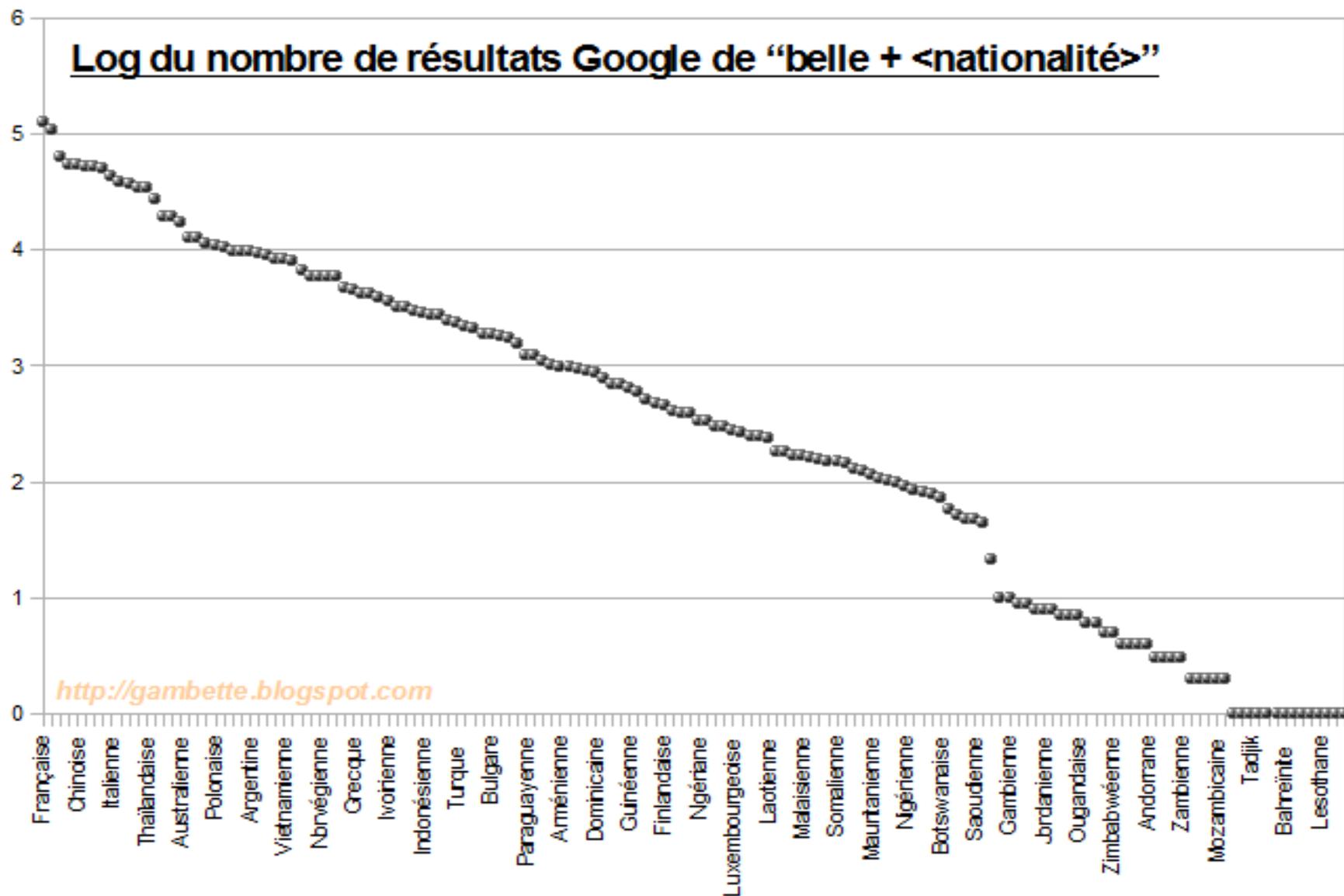


Moyenne

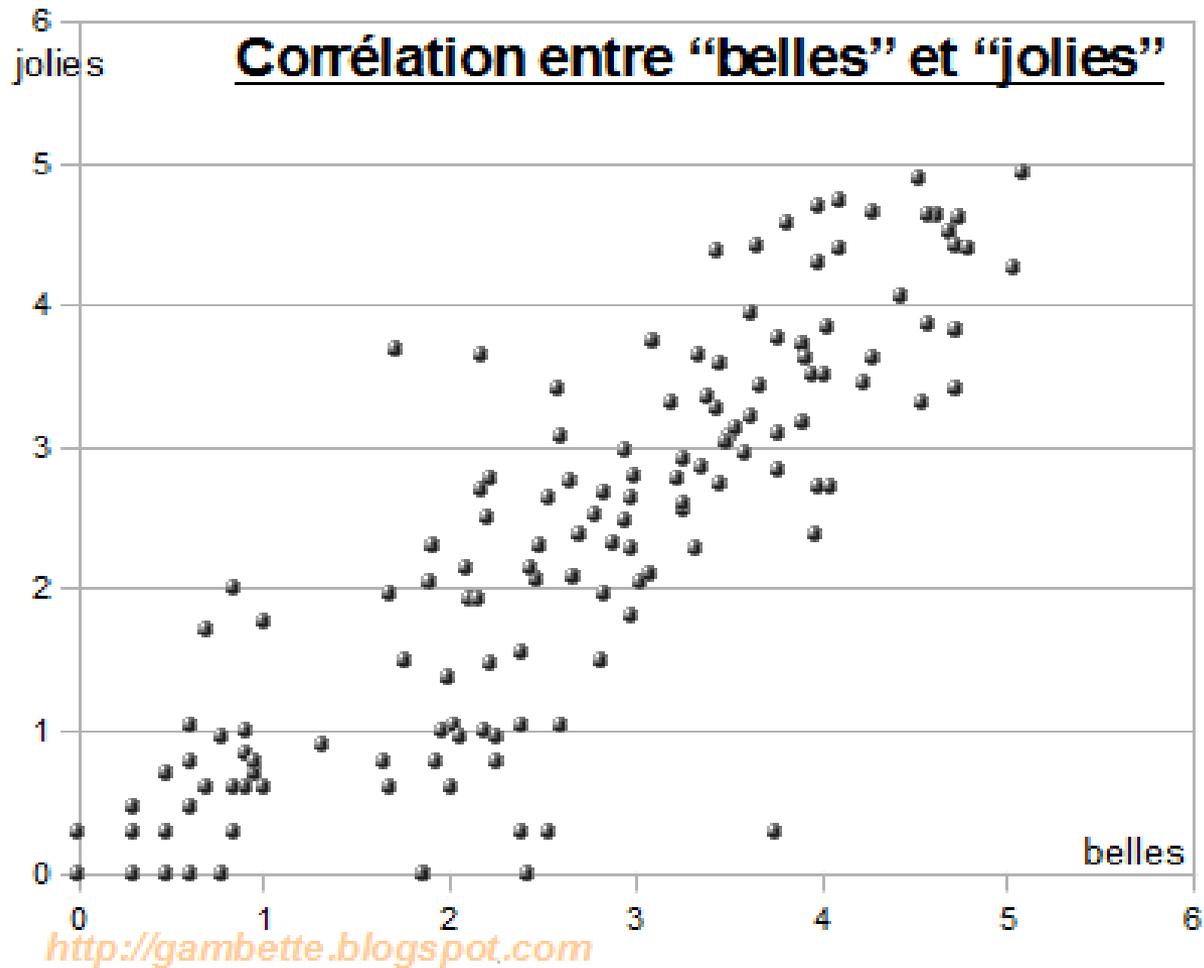


$1\ 000\ 000\ 000/n(x)$, où $n(x)$ est le nombre de résultats pour l'année x

Google Fights : Miss Google 2010

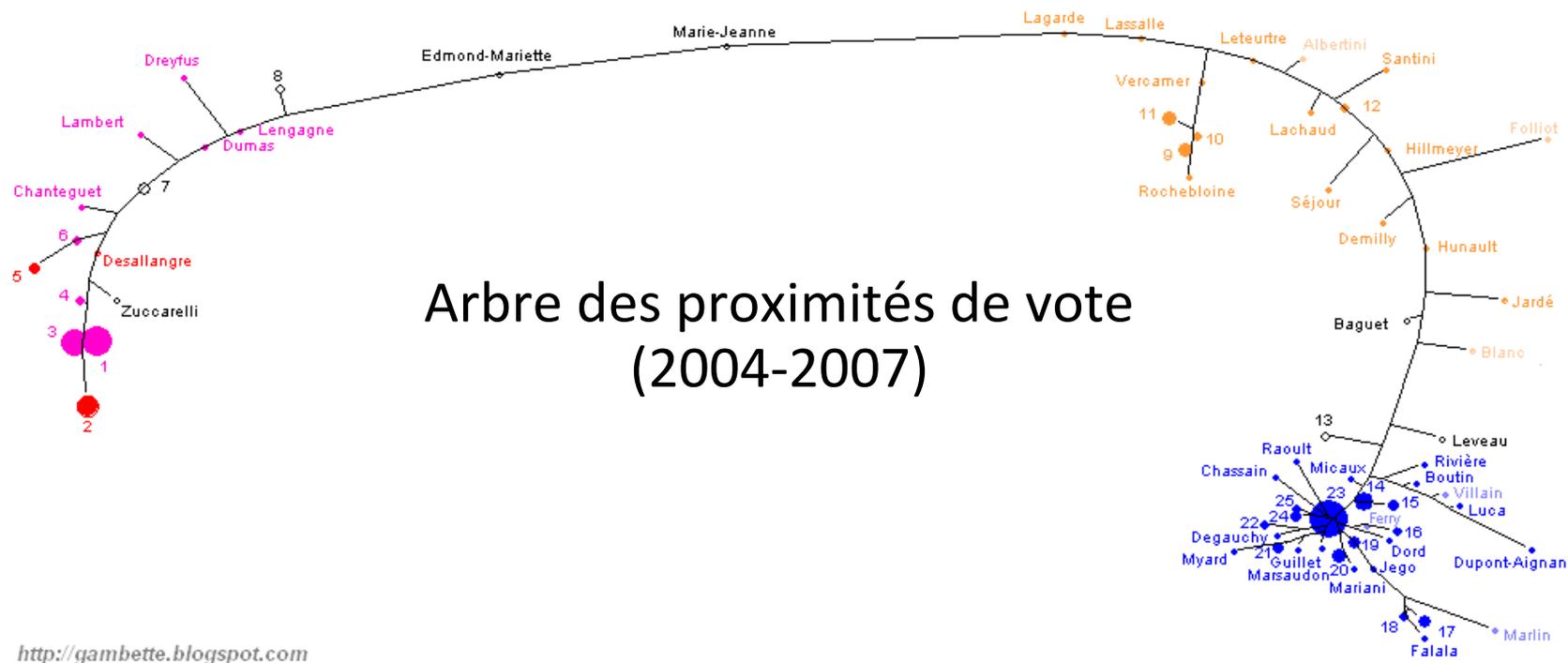


Google Fights : Miss Google 2010

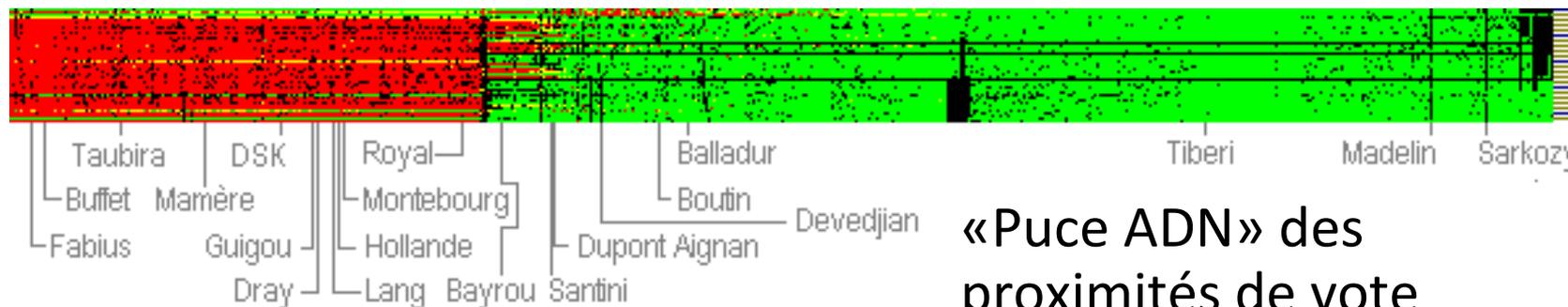


Données en arbres

Vote des députés



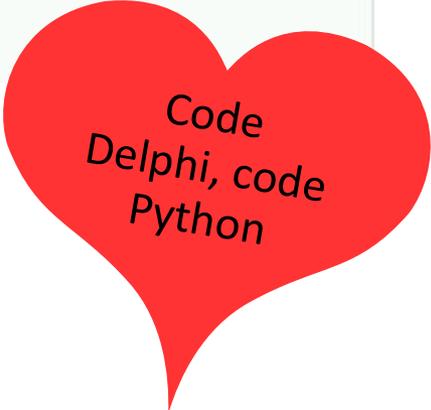
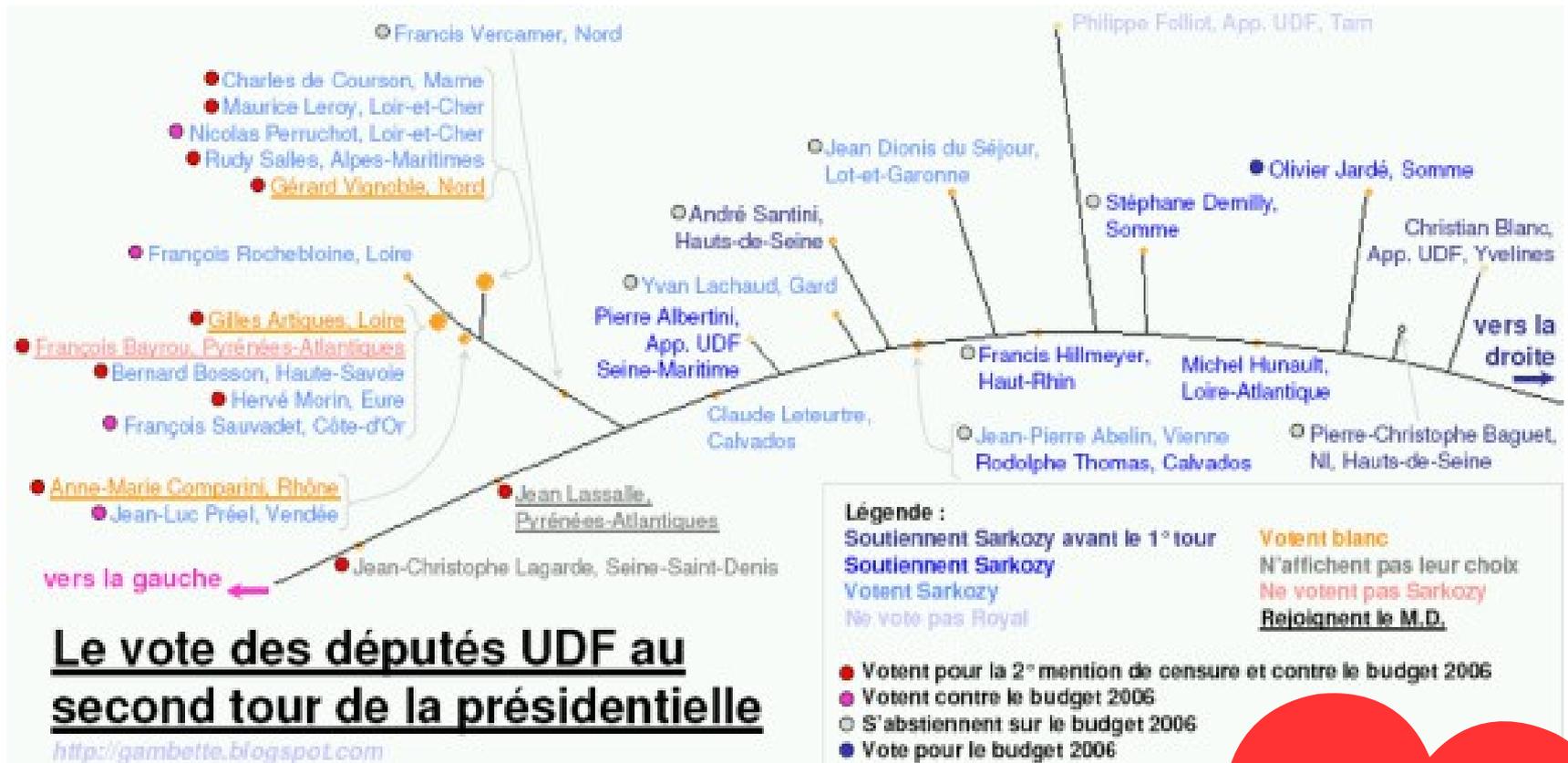
<http://gambette.blogspot.com>



<http://gambette.blogspot.fr/2007/01/arbre-phylogntique-des-dputs.html>

<http://gambette.blogspot.fr/2007/02/la-puce-adn-des-dputs.html>

Vote des députés



Arbres phylogénétiques et arbres de mots

ESPÈCES

Séquences ADN

Données sur
les feuilles

MOTS

Position des mots

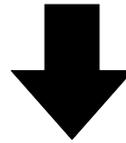
Arbres phylogénétiques et arbres de mots

ESPÈCES

Séquences ADN

Distances fondées sur la différence entre les deux séquences (mutations, insertions, délétions)

Données sur les feuilles



Distances entre les feuilles

	A	B	C	D
A	0	2	5	6
B	2	0	5	6
C	5	5	0	3
D	6	6	3	0

MOTS

Position des mots

Distances fondées sur la cooccurrence entre les deux mots

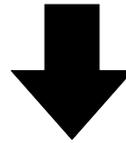
Arbres phylogénétiques et arbres de mots

ESPÈCES

Séquences ADN

Distances fondées sur la différence entre les deux séquences (mutations, insertions, délétions)

Données sur les feuilles



Distances entre les feuilles

	A	B	C	D
A	0	2	5	6
B	2	0	5	6
C	5	5	0	3
D	6	6	3	0



classification hiérarchique ascendante

Arbre



MOTS

Position des mots

Distances fondées sur la cooccurrence entre les deux mots

Arbres phylogénétiques et arbres de mots

ESPÈCES

Séquences ADN

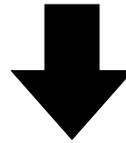
Distances fondées sur la différence entre les deux séquences (mutations, insertions, délétions)

MOTS

Position des mots

Distances fondées sur la cooccurrence entre les deux mots

Données sur les feuilles



Distances entre les feuilles

	A+B	C	D
A+B	0	5	6
C	5	0	3
D	6	3	0



classification hiérarchique ascendante

Arbre



Arbres phylogénétiques et arbres de mots

ESPÈCES

Séquences ADN

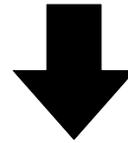
Distances fondées sur la différence entre les deux séquences (mutations, insertions, délétions)

MOTS

Position des mots

Distances fondées sur la cooccurrence entre les deux mots

Données sur les feuilles



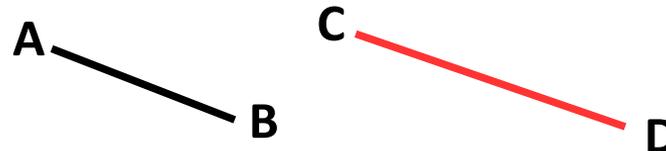
Distances entre les feuilles

	A+B	C	D
A+B	0	5	6
C	5	0	3
D	6	3	0



classification hiérarchique ascendante

Arbre



Arbres phylogénétiques et arbres de mots

ESPÈCES

Séquences ADN

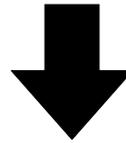
Distances fondées sur la différence entre les deux séquences (mutations, insertions, délétions)

MOTS

Position des mots

Distances fondées sur la cooccurrence entre les deux mots

Données sur les feuilles



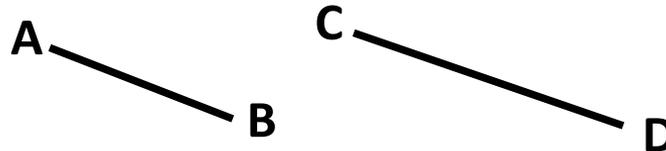
Distances entre les feuilles

	A+B	C+D
A+B	0	5,5
C+D	5,5	0



classification hiérarchique ascendante

Arbre



Arbres phylogénétiques et arbres de mots

ESPÈCES

Séquences ADN

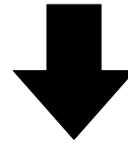
Distances fondées sur la différence entre les deux séquences (mutations, insertions, délétions)

MOTS

Position des mots

Distances fondées sur la cooccurrence entre les deux mots

Données sur les feuilles



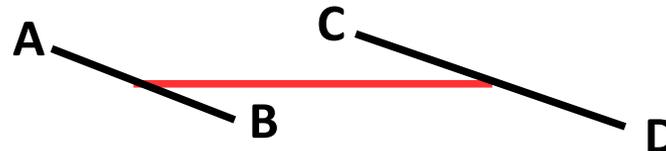
Distances entre les feuilles

	A+B	C+D
A+B	0	5,5
C+D	5,5	0



classification hiérarchique ascendante

Arbre



Arbres phylogénétiques et arbres de mots

ESPÈCES

Séquences ADN

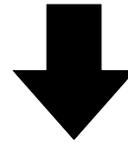
Distances fondées sur la différence entre les deux séquences (mutations, insertions, délétions)

MOTS

Position des mots

Distances fondées sur la cooccurrence entre les deux mots

Données sur les feuilles



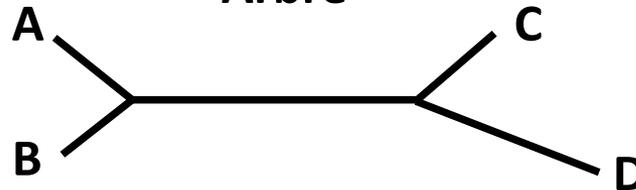
Distances entre les feuilles

	A	B	C	D
A	0	2	5	6
B	2	0	5	6
C	5	5	0	3
D	6	6	3	0



classification hiérarchique ascendante

Arbre



Outils pratiques

Quelques outils pratiques

- plugin **iMacros** de Firefox

Pour récupérer un ensemble de pages web

- **expressions régulières**

Pour extraire de l'information ou la changer de format

Dans la fonction rechercher/remplacer d'un éditeur de texte ou dans un script Python

- **bibliothèques Javascript** D3.js, Google Charts, Charts.js, etc.

Pour visualiser les données de manière interactive sur le web

<http://www.sitepoint.com/15-best-javascript-charting-libraries/>

Les expressions régulières selon xkcd

A CHAQUE FOIS QUE JE DÉCOUVRE UNE NOUVELLE TECHNIQUE, J'IMAGINE DES SITUATIONS COMPLEXES OU ÇA ME PERMETTRAIT D'ÊTRE UN HÉROS.

OH NON ! LE TUEUR A DÛ LA SUIVRE SUR SON LIEU DE VACANCES !



POUR LES RETROUVER, IL FAUDRAIT RECHERCHER CE QUI RESSEMBLE À UNE ADRESSE PARMIS SES 200 Mo DE MAILS !



C'EST SANS ESPOIR !

LAISSEZ-MOI FAIRE.



JE CONNAIS LES EXPRESSIONS RÉGULIÈRES.





Data-Driven Documents



Pour continuer à jouer avec les données

- Data Job 2015 – 26 novembre 2015 à Paris :

<http://datajob.fr/>

(entrée gratuite pour étudiants moins de 28 ans)

- Blog *Je véronise* :

<http://gambette.blogspot.com/>

- Les interventions d'Henri Verdier sur l'open data :

https://www.youtube.com/results?search_query=Henri+Verdier

- Actualités de la révolution des données

<http://radar.oreilly.com/data>