

# ML Week

Bonjour !

Jeff Abrahamson

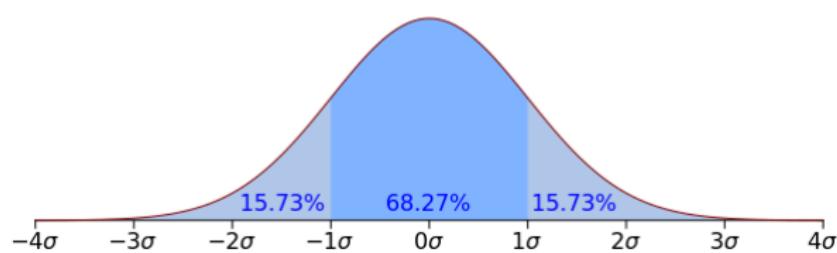
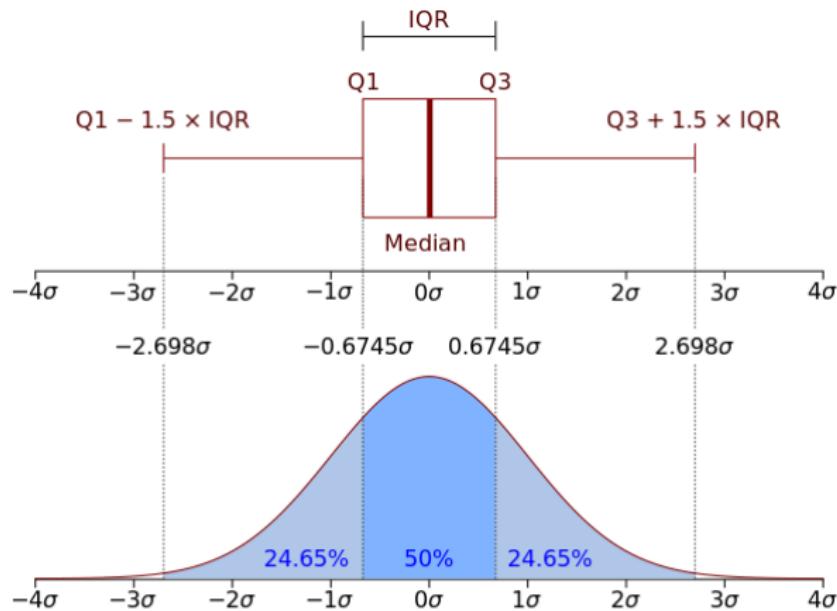
15 janvier 2016

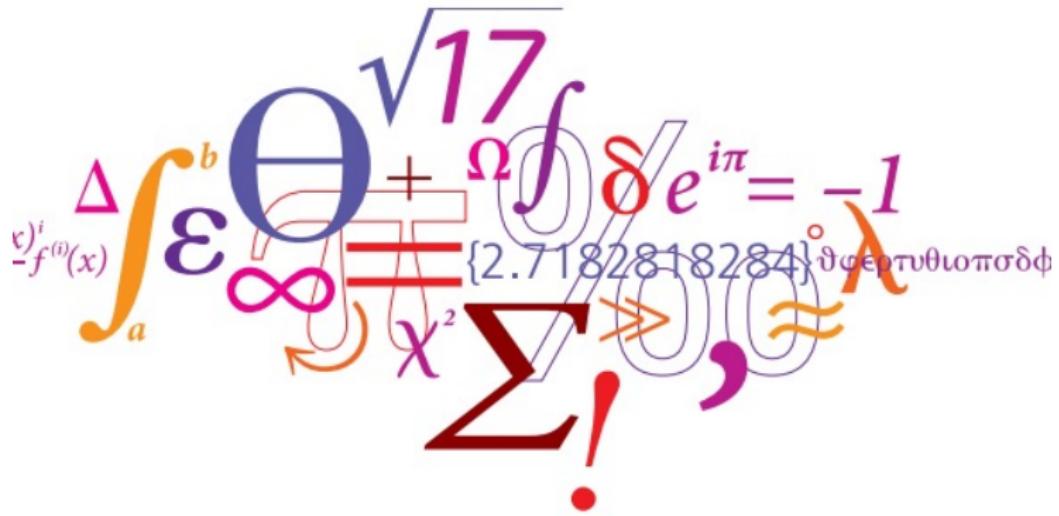
# **Machine Learning**

# Apprentissage automatique

# **Data Science**

# **Statistiques**

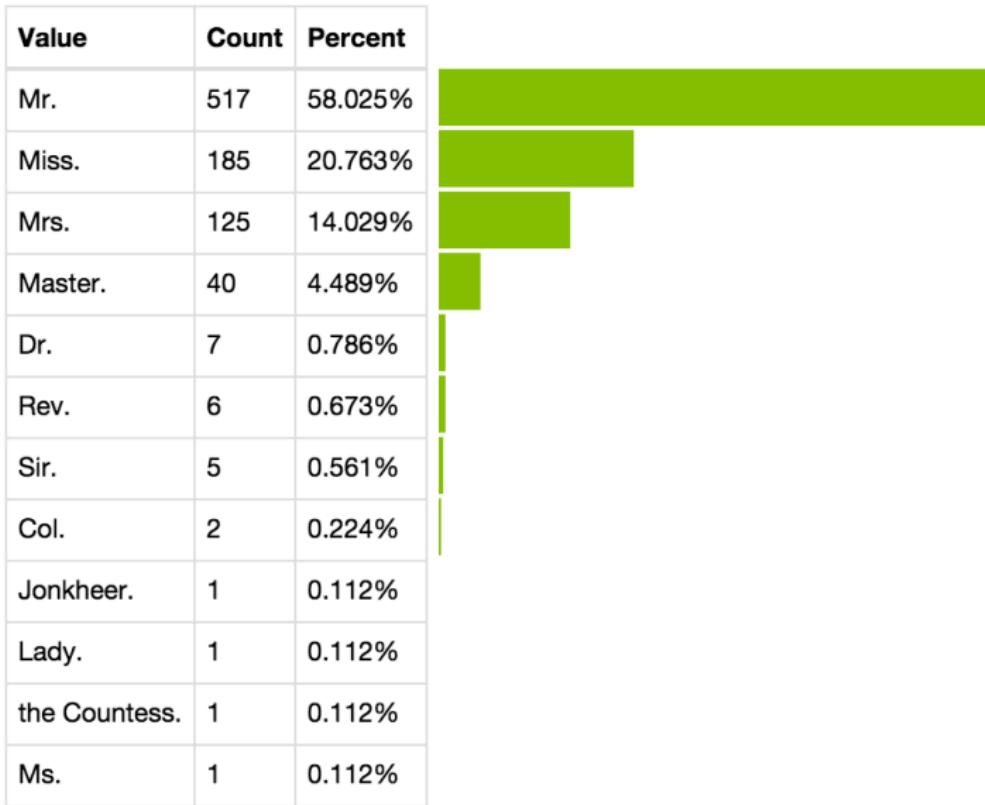




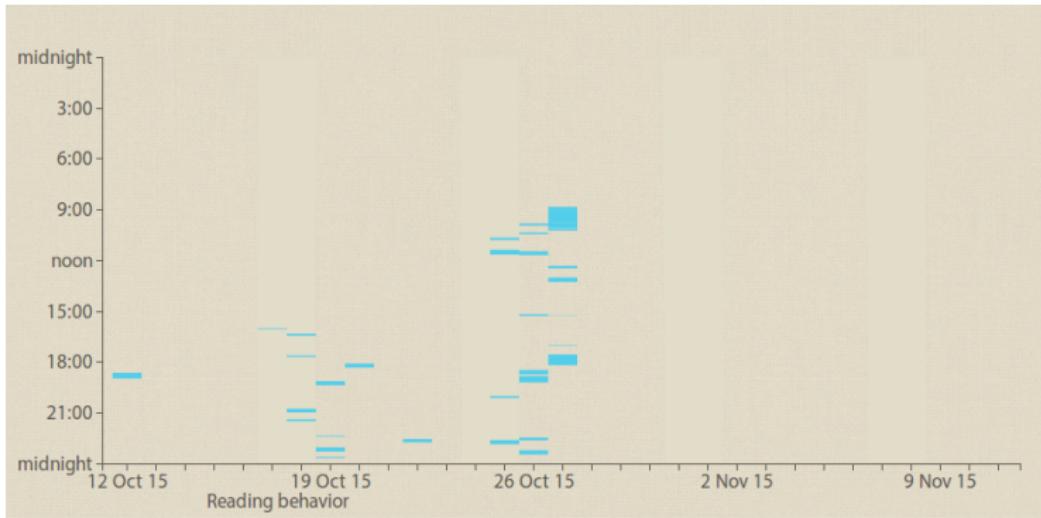
**Espace vectoriel**

**Critères**

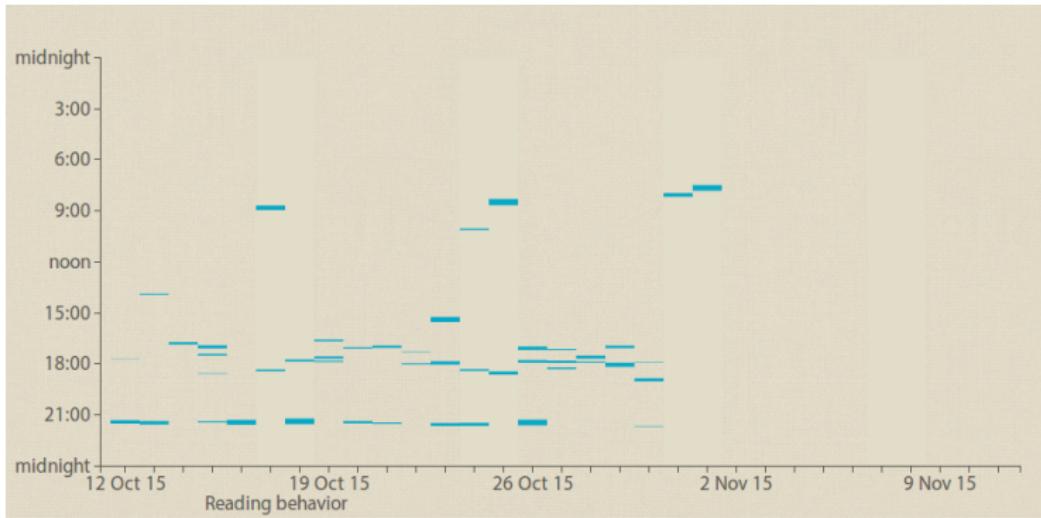
**Feature engineering**



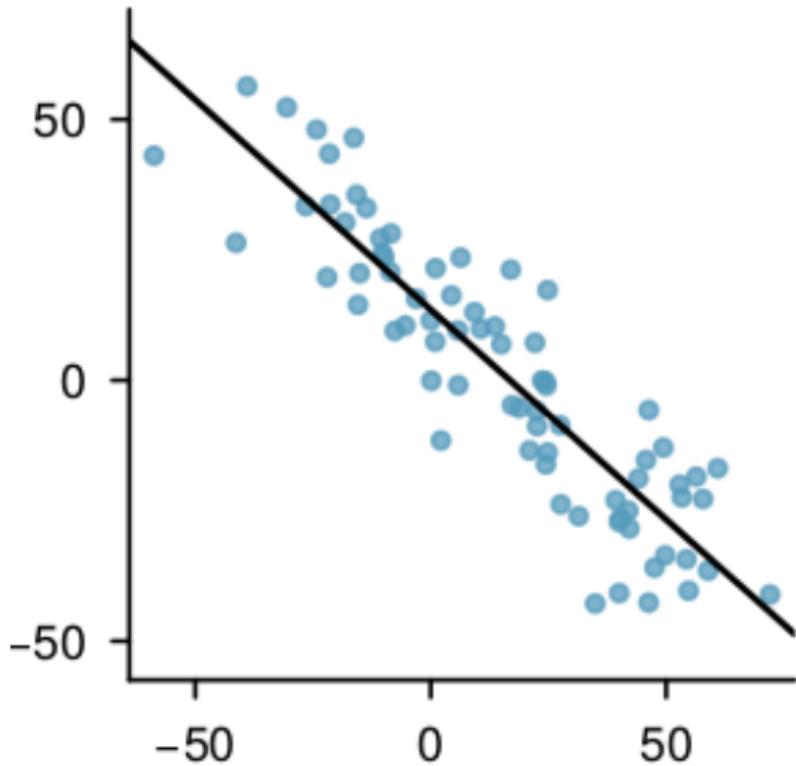
Kaggle

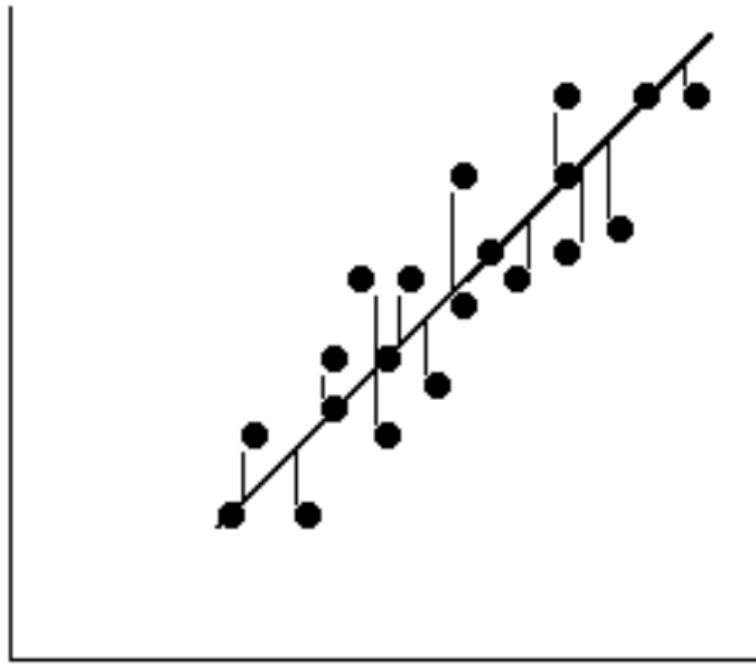


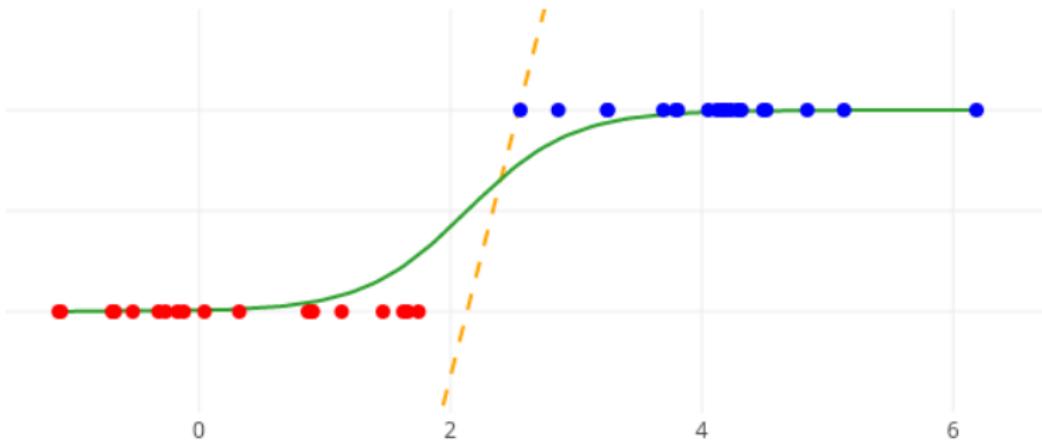
*Jellybooks*

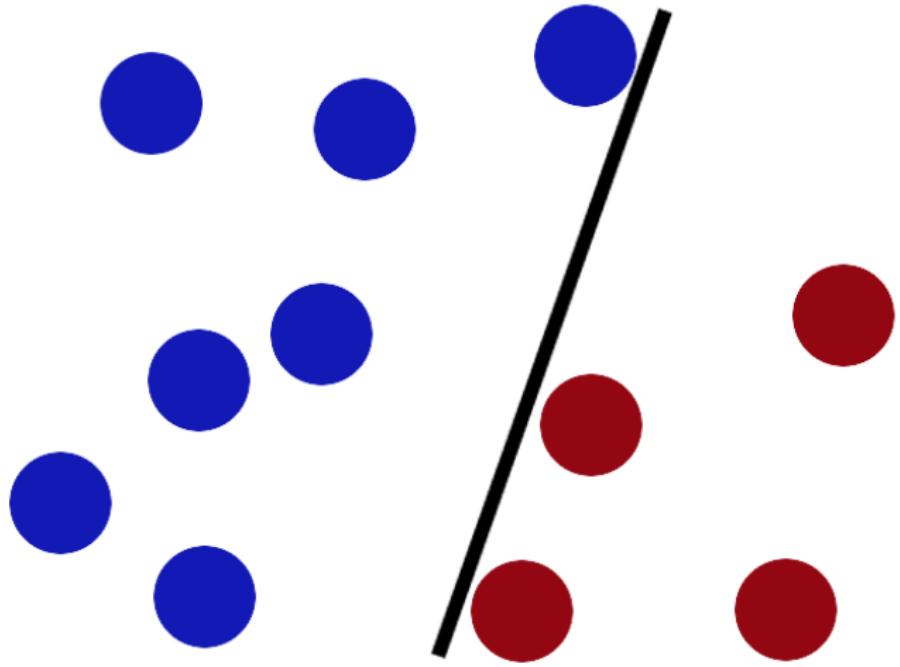


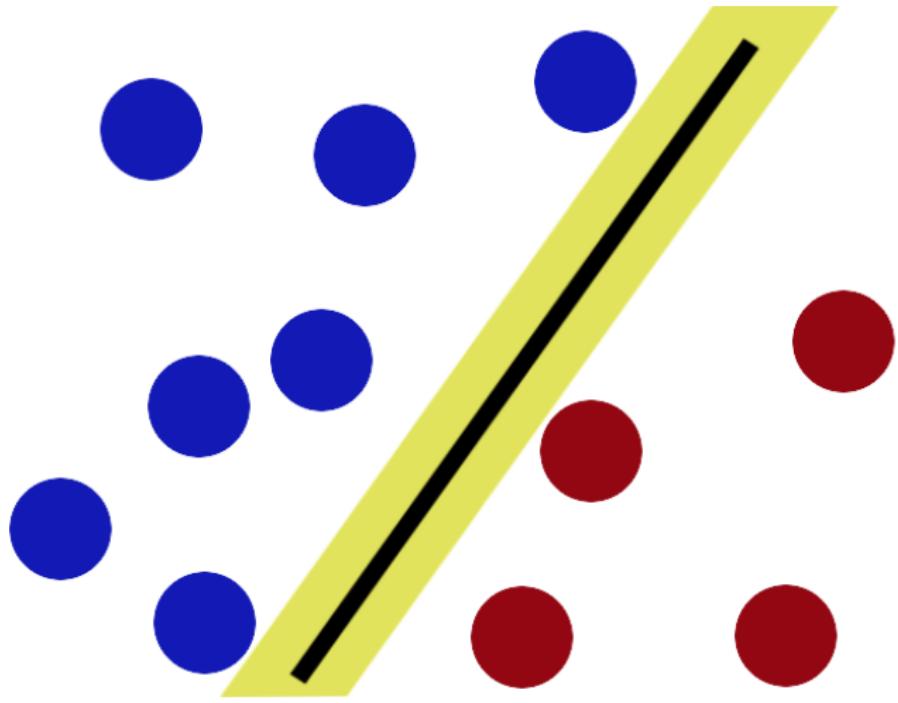
*Jellybooks*

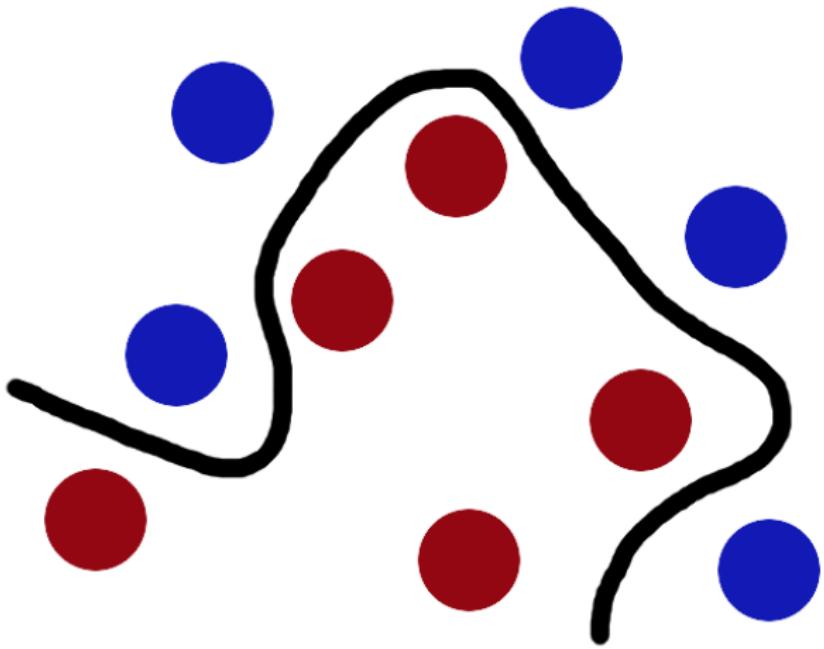


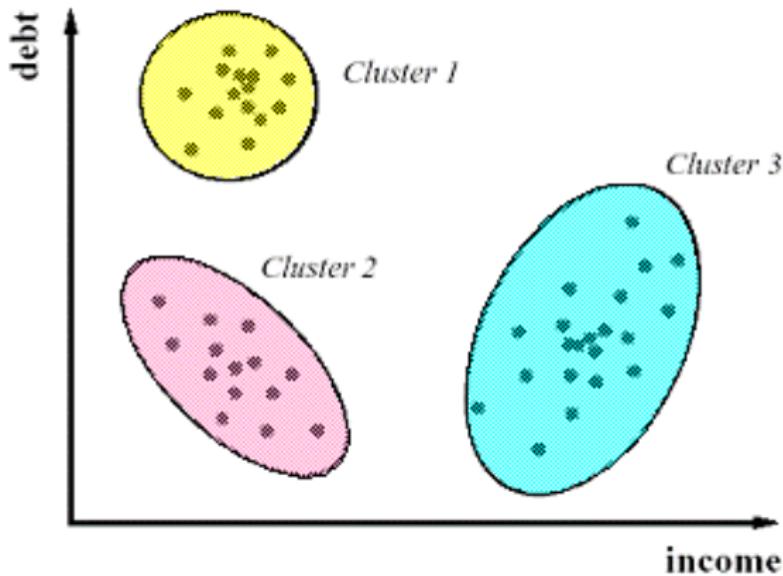


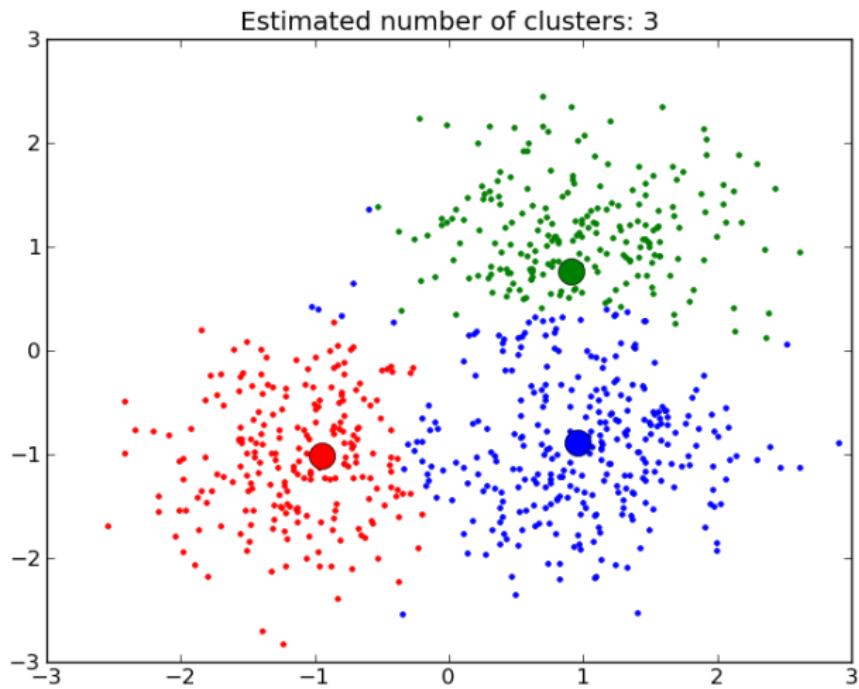


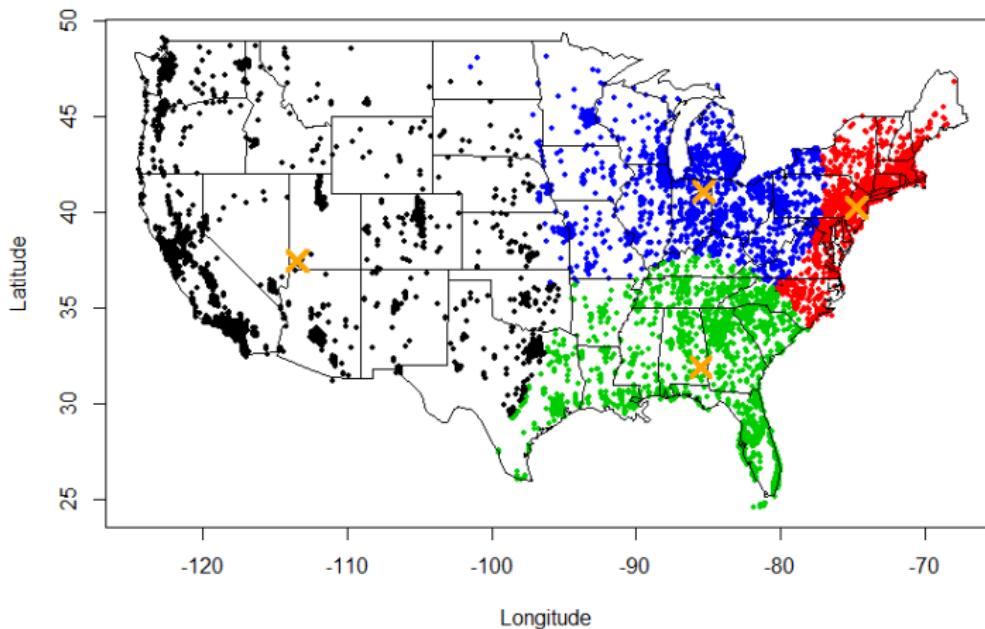




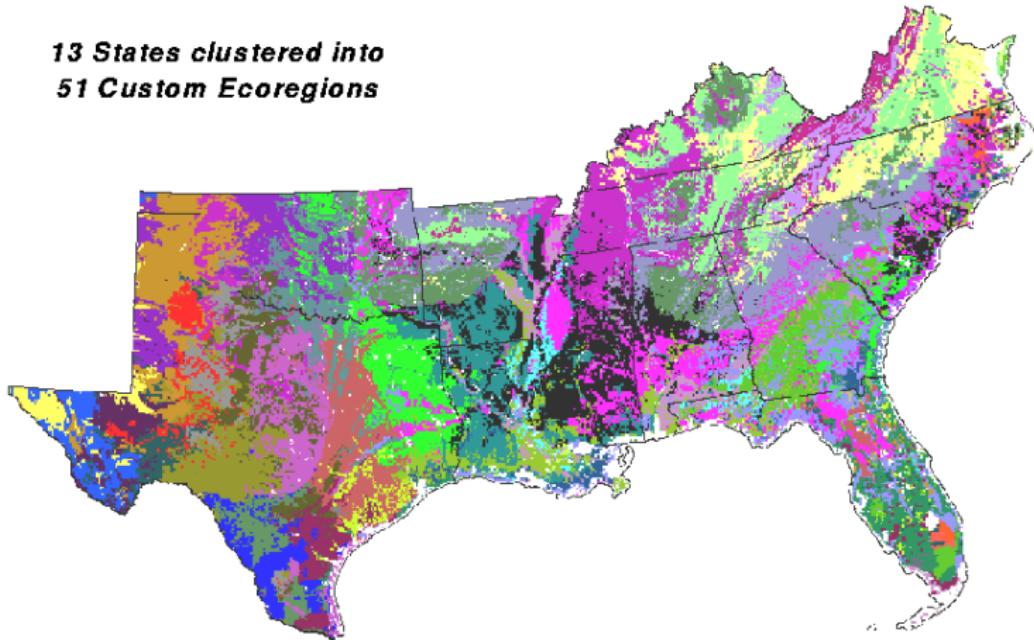








***13 States clustered into  
51 Custom Ecoregions***





*Eddie Bell @ Lyst*



*Eddie Bell @ Lyst*



*Eddie Bell @ Lyst*



*Eddie Bell @ Lyst*



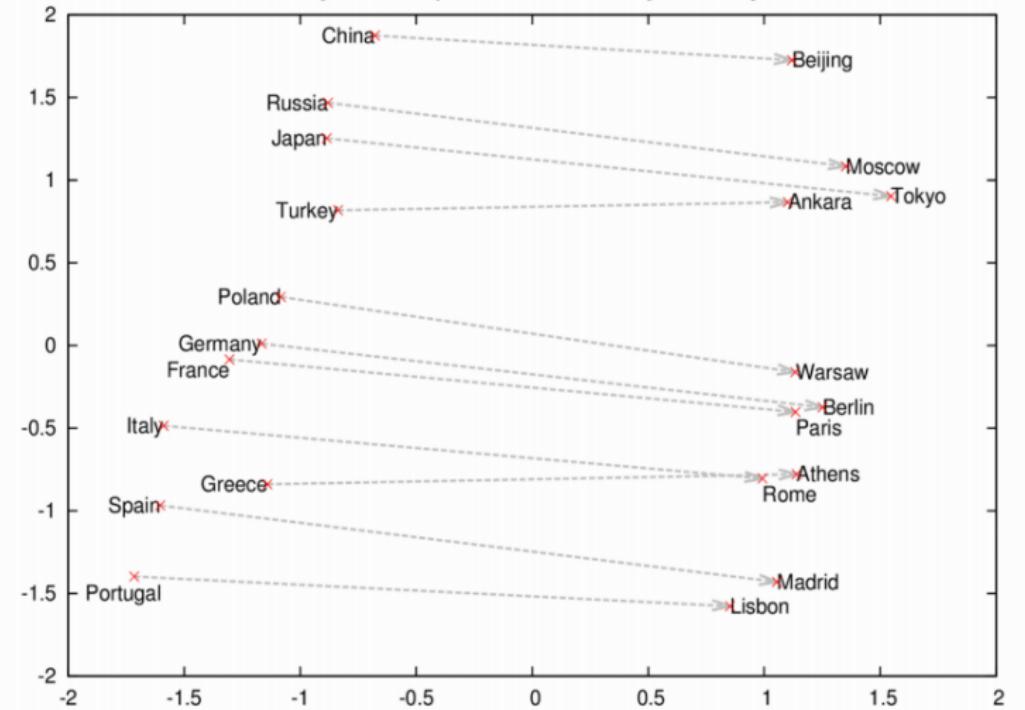
-1	0	+1
-2	0	+2
-1	0	+1



-1	-2	-1
0	0	0
+1	+2	+1



### Country and Capital Vectors Projected by PCA



Term	Similarity	
	"shift"	0.933104
	"gown"	0.887743
	"skirt"	0.881672
	"bandage"	0.880162
	"midi"	0.869786

**Similar to 'dress'**

*Eddie Bell @ Lyst*



a group of young girls standing next  
to each other on the beach



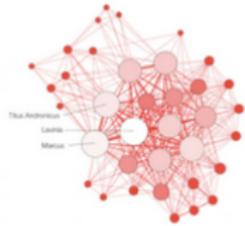
A clock tower with a clock on top of it

*Google?*



## A bunch of bananas hanging from a tree

*Google?*



**TITUS ANDRONICUS**  
Number of characters **36** | 50% Network density



**ROMEO AND JULIET**  
Number of characters **41** | 37% Network density



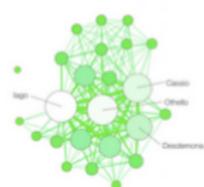
**JULIUS CAESAR**  
Number of characters **46** | 34% Network density



**HAMLET**  
Number of characters **37** | 39% Network density

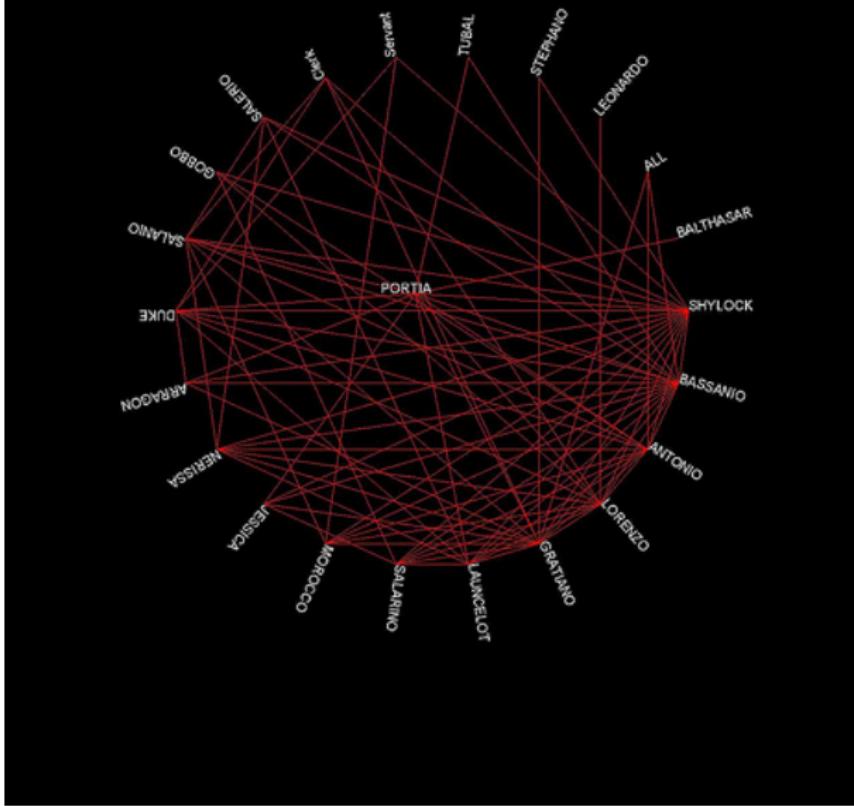


**TROILUS AND CRESSIDA**  
Number of characters **35** | 40% Network density



**OTHELLO**  
Number of characters **24** | 55% Network density

[https://flowingdata.com/2015/12/30/  
shakespeare-tragedies-as-network-graphs/](https://flowingdata.com/2015/12/30/shakespeare-tragedies-as-network-graphs/)



## Act 4, Scene 1

**Perseus** And I am nothing else to stink his hand.  
**Medea** Creon is the chough, I like it not.

**Perseus** And therefore have I little talked of love.

**Perseus** Look, see, how comes the belly towards my self.

**Perseus** Needs me not let us go on.

**Perseus** That we do, we do, we do, we do.

**Perseus** What is he that comes? Thee, Perseus, mad?

**Perseus** What is he that comes?

**Perseus** Come to me, make no noise.

**Perseus** No noise, that's all I want you to do.

**Perseus** No noise, no noise, no noise, no noise.

**Perseus** I will speak to you of your loves.

**Perseus** So what? I will speak to you too.

**Perseus** Being spylid before your family, than to your face.

**Perseus** How now! why is he not here and there?

**Perseus** The trees have given me nothing by that.

**Perseus** This aysong is noches song, and that report.

**Perseus** And what is aysong? I update it to thy bark.

**Perseus** My bark is tree, and that report is wood.

**Perseus** Or shall I come to you at evening mass.

**Medea** My bane serves me, peruse daughter, now.

**Perseus** Juliet, on Thursday early will I noise ye.

**Perseus** just hope, just cure, just help.

**Medea** I hear thou must, and nothing may prologue it.

Tell me not, friar, that thou  
hear'st of this.

Thou hast the strength of will to slay thyself.

Act 5, Scene 1

My dreams presage some joyful news at hand.

**Author:** Her body sleeps in Capels' monument.

**ANSWER** I will hence to-night.

*[Author:] I do beseech you, sir, have patience.*

Sustaining the Environment

get thee gone.

300 Hypotheses

That the life-weary taker may fall dead.

is my theory. But Martha's law

famine is in thy cheeks.

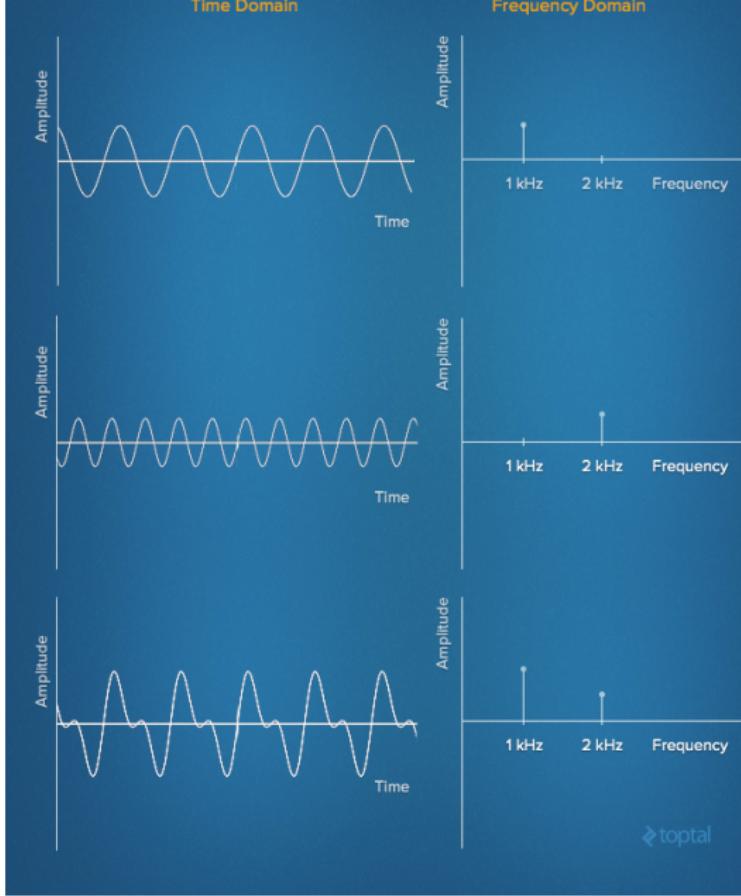
#### **How the world works**

Put this in any liquid thing you will.

— There is thy gold, worse poison to men's souls.

*http:*

*//www.nand.io/visualisation/understanding-shakespeare*



toptal

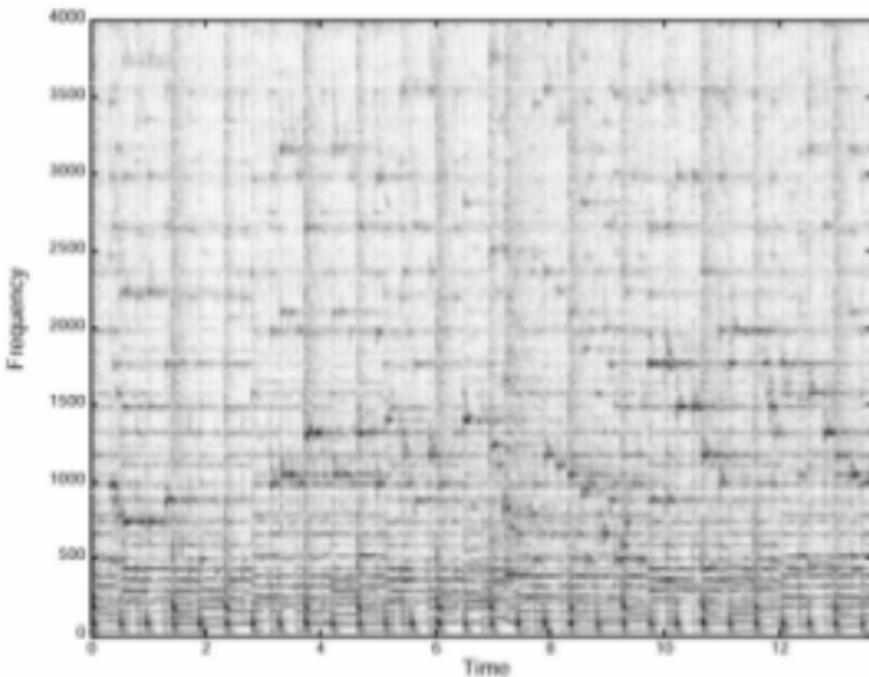


Fig. 1A - Spectrogram

*https:*

//www.ee.columbia.edu/~dpwe/papers/Wang03-shazam.pdf

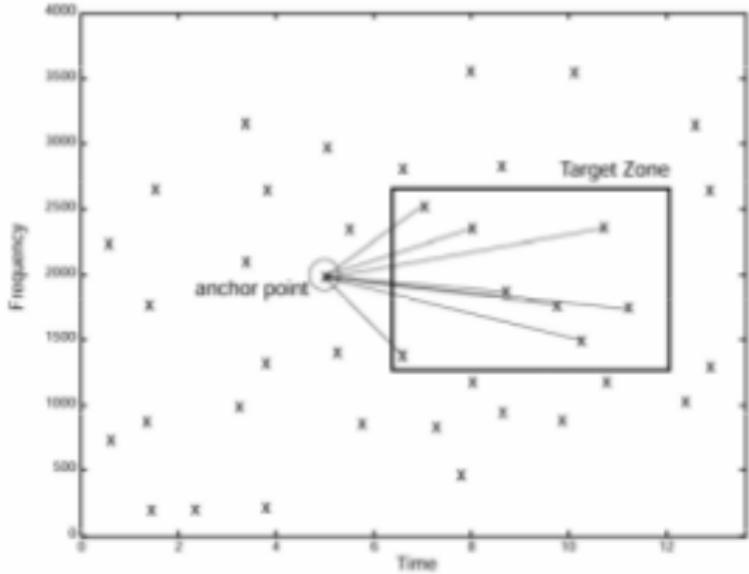
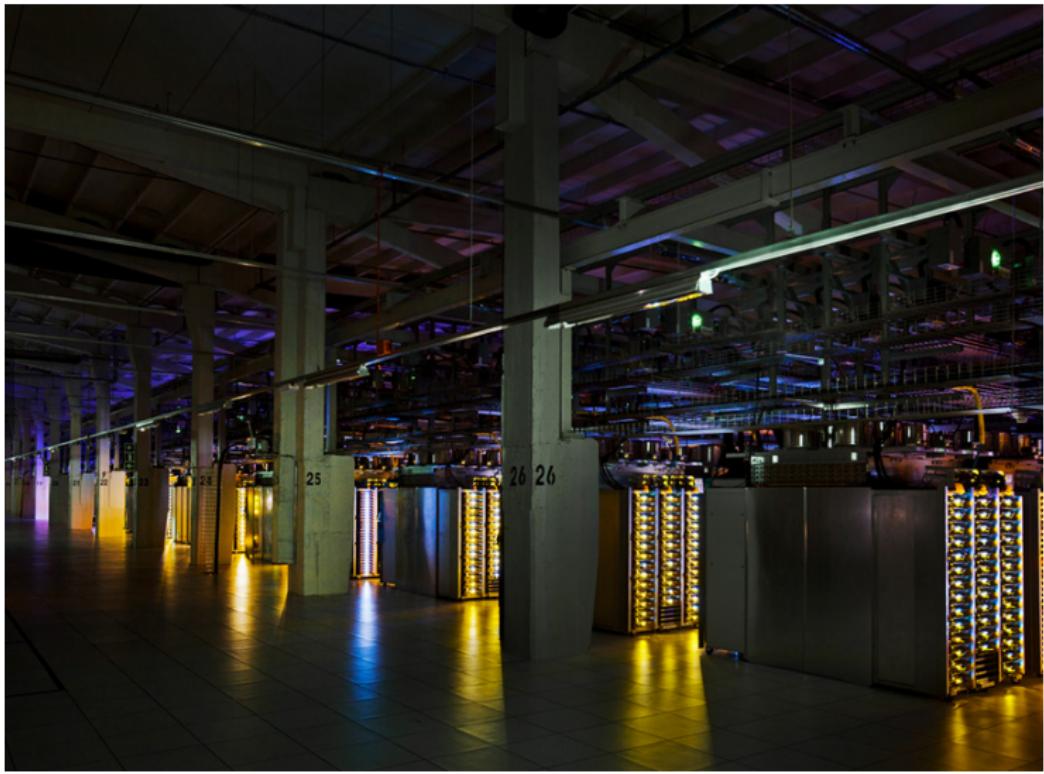


Fig. 1C - Combinatorial Hash Generation

[https:](https://www.ee.columbia.edu/~dpwe/papers/Wang03-shazam.pdf)

//[www.ee.columbia.edu/~dpwe/papers/Wang03-shazam.pdf](https://www.ee.columbia.edu/~dpwe/papers/Wang03-shazam.pdf)



*Google*

## **Big data?**

*http://aadrake.com/*

*command-line-tools-can-be-235x-faster-than-your-hadoop-cluster.html*

# Un peu de maths...

- 100 critères
- int (*4 octets par critère*)
- 16 GB RAM
- $O(n \log n)$  en mémoire

$$400N + 400N \log(400N) < \frac{16 \cdot 10^9}{400}$$

# Un peu de maths...

- 100 critères
- int (*4 octets par critère*)
- 16 GB RAM
- $O(n \log n)$  en mémoire

$$400N + 400N \log(400N) < \frac{16 \cdot 10^9}{400}$$

$$N \approx 2 \cdot 10^6$$

# Questions?

[purple.com/talk-feedback](http://purple.com/talk-feedback)