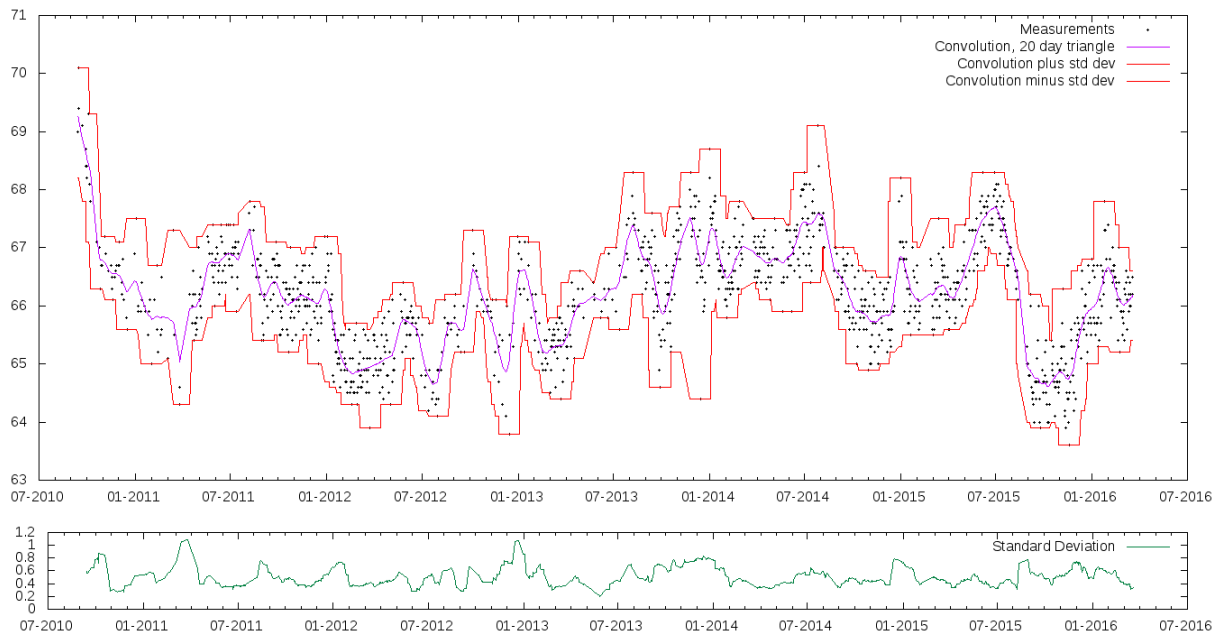


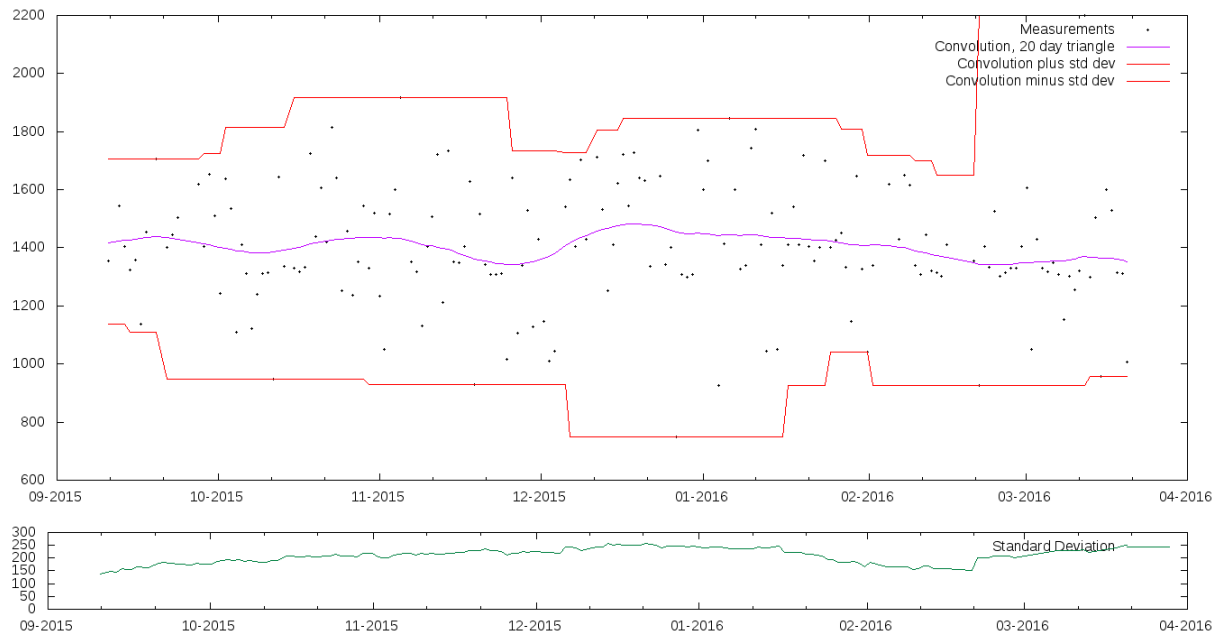
# Mais qu'est-ce que je fous?

comment glander et apprendre les maths en même temps

Jeff Abrahamson

23 March 2016





# Le problème

Avant j'utilisais ratpoison :

```
task=$(ratpoison -c 'windows %s%t' | egrep '^\\*' | perl -pwe 's/^\\*//; s/ /_/g;')  
echo $(date +%s) $task >> $task_file
```

Maintenant j'utilise i3 :

```
id=$(xprop -root | awk '/_NET_ACTIVE_WINDOW\ (WINDOW\)/{print $NF}')  
task=$(xprop -id $id | awk '/_NET_WM_NAME/{$1=$2="";print}' | cut -d'"' -f2)  
echo $(date +%s) $task >> $task_file
```

Et ça donne :

```
1458399349 emacs@starshine - talk-header.tex : /home/jeff/src/jma/talks/2016-03__breizhcamp-que-je-fous/talk-header.tex
1458399409 emacs@starshine - talk.tex : /home/jeff/src/jma/talks/2016-03__breizhcamp-que-je-fous/talk.tex
1458399469 listings - How to put C++ source code into beamer slides - TeX - LaTeX Stack Exchange - Mozilla Firefox
1458399529 emacs@starshine - talk.tex : /home/jeff/src/jma/talks/2016-03__breizhcamp-que-je-fous/talk.tex
1458399589 talk.pdf -- Mais qu'est-ce que je fous? - comment glander et apprendre les maths en même temps
1458399649 emacs@starshine - talk.tex : /home/jeff/src/jma/talks/2016-03__breizhcamp-que-je-fous/talk.tex
1458399709 emacs@starshine - talk.tex : /home/jeff/src/jma/talks/2016-03__breizhcamp-que-je-fous/talk.tex
1458399769 emacs@starshine - macros.tex : /home/jeff/src/jma/talks/2016-03__breizhcamp-que-je-fous/macros.tex
1458399829 emacs@starshine - talk.tex : /home/jeff/src/jma/talks/2016-03__breizhcamp-que-je-fous/talk.tex
1458399890 emacs@starshine - talk.tex : /home/jeff/src/jma/talks/2016-03__breizhcamp-que-je-fous/talk.tex
1458399950 talk.pdf -- Mais qu'est-ce que je fous? - comment glander et apprendre les maths en même temps
1458400010 jeff@starshine:~/src/jma/talks/2016-03__breizhcamp-que-je-fous
```

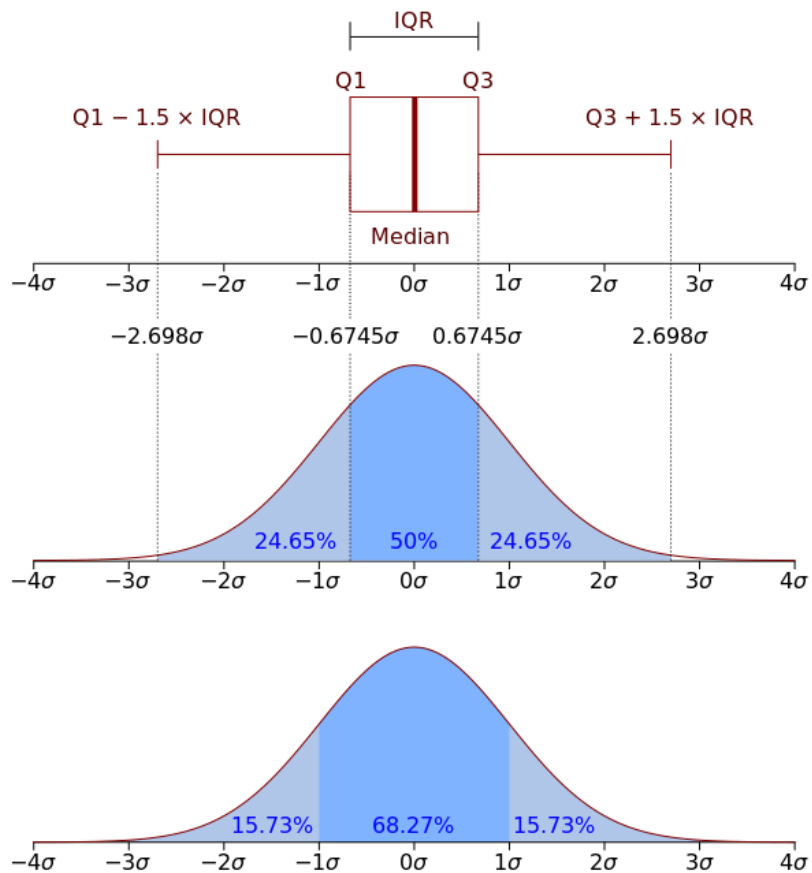
**Quelles questions puis-je poser?**

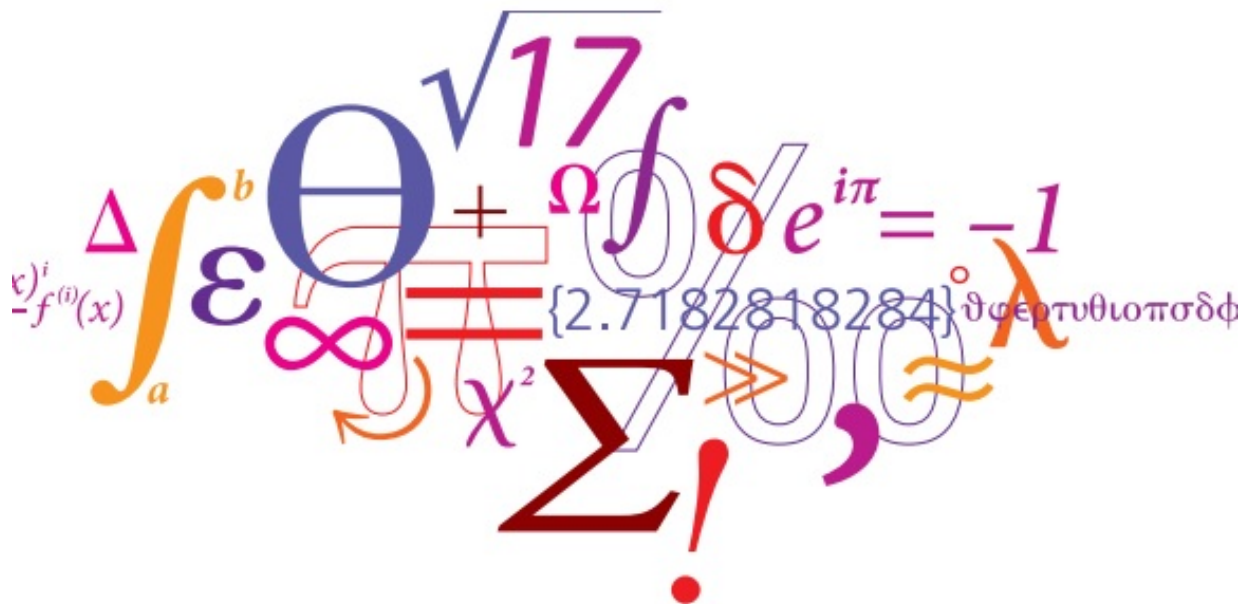


# Machine Learning

# Data Science

# Statistics

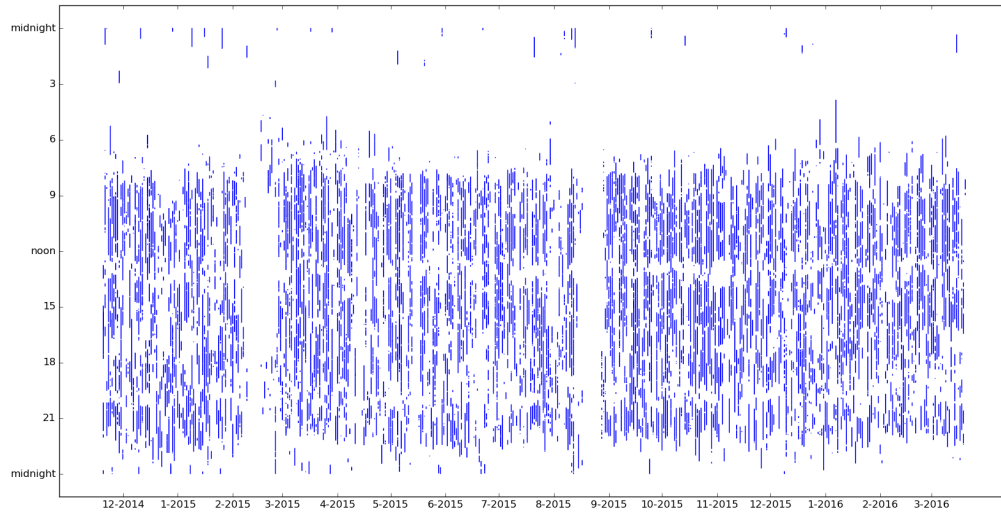


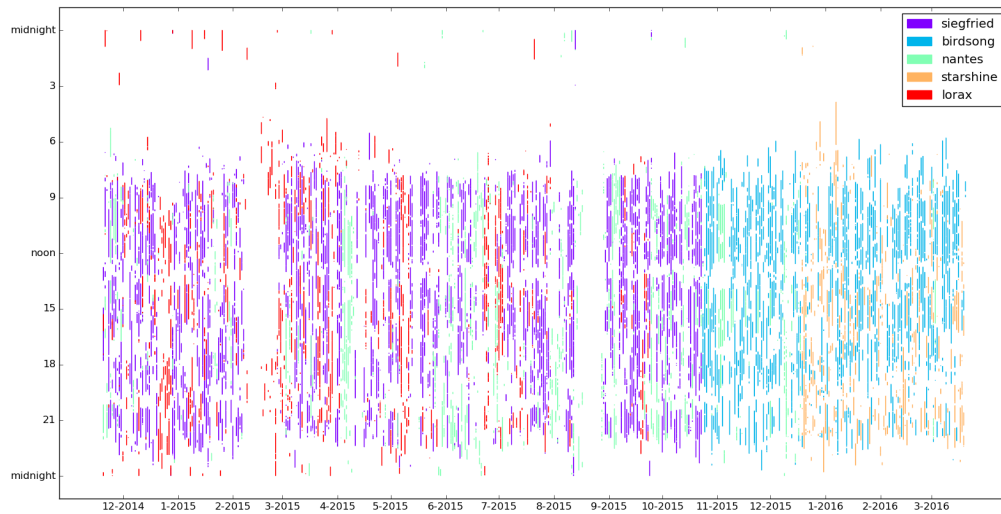


**Vector Space**

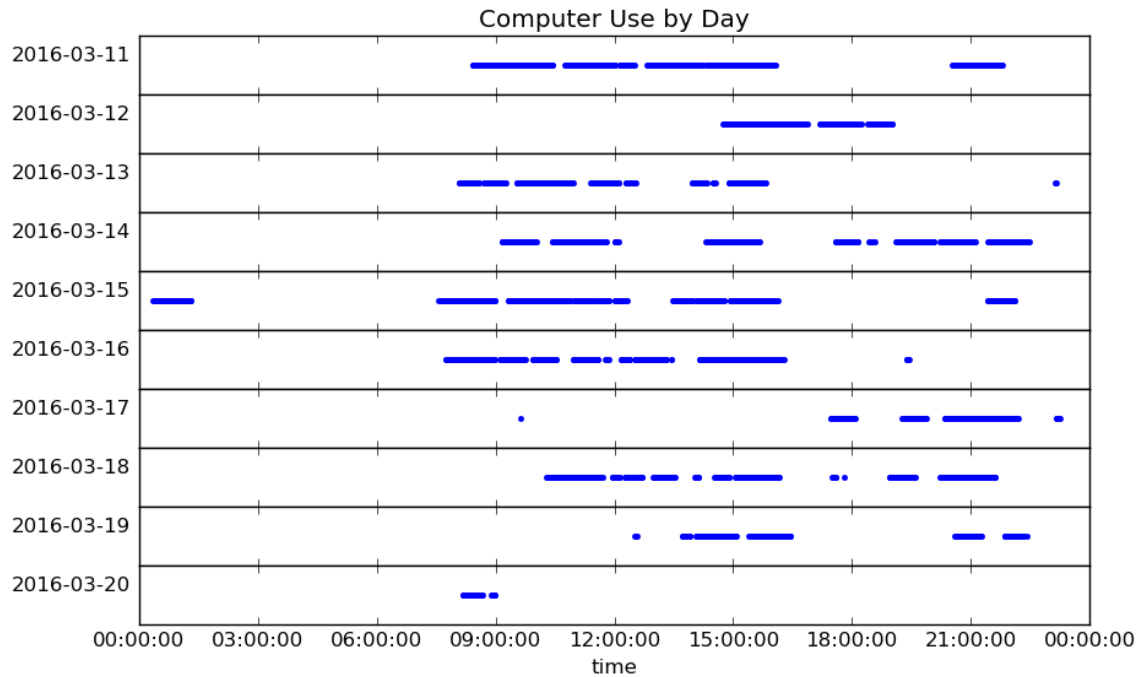
**Features**

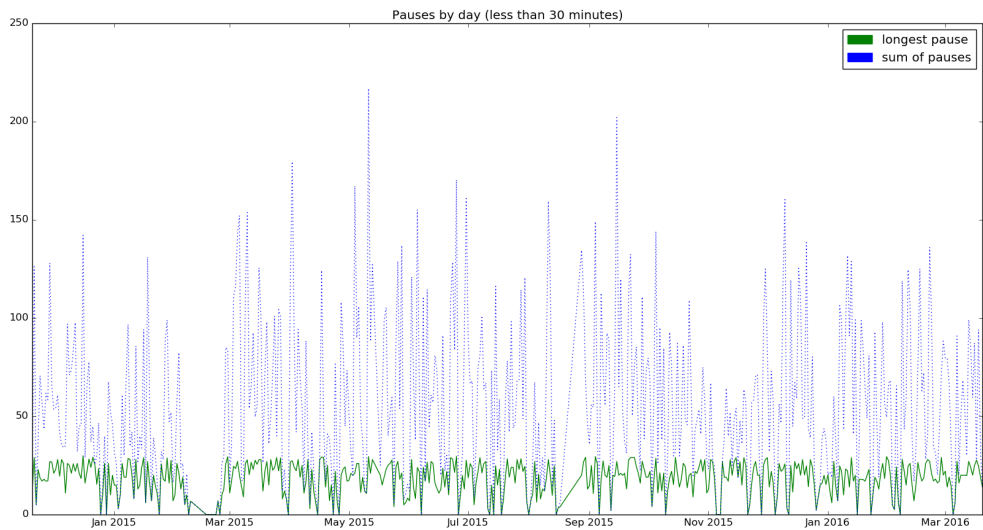
**Feature engineering**

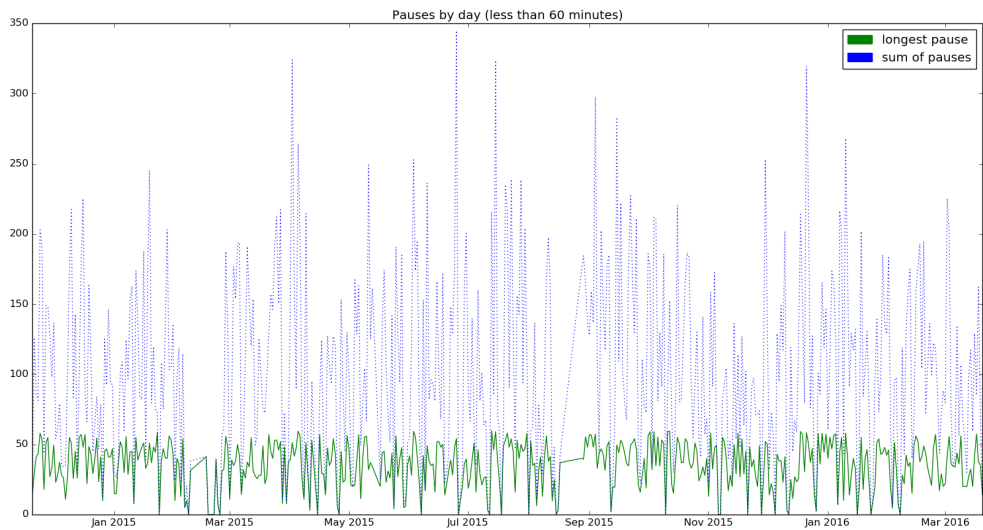


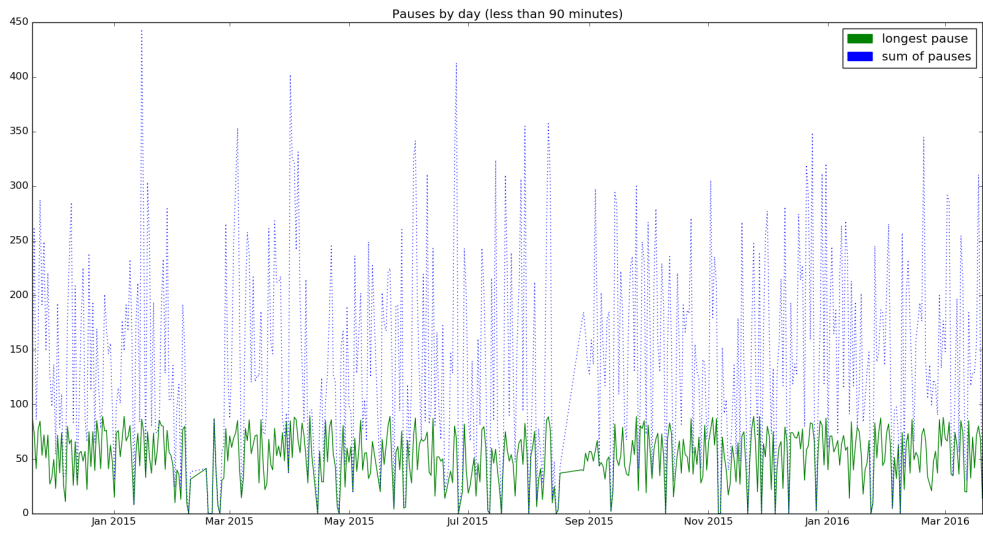


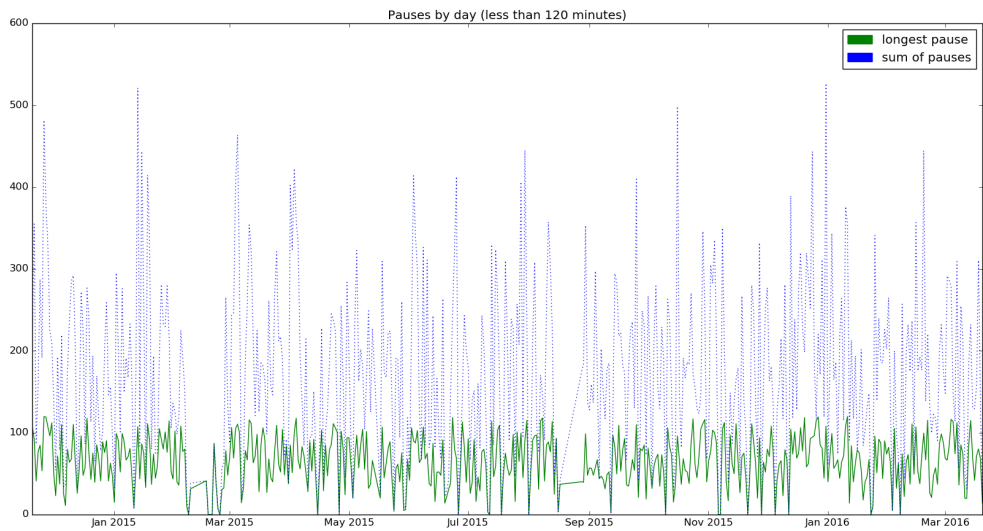












**sac de mots**

Exemple :

*Il est nuit. La cabane est pauvre, mais bien close.  
Le logis est plein d'ombre et l'on sent quelque chose  
Qui rayonne à travers ce crépuscule obscur.  
Des filets de pêcheur sont accrochés au mur.  
Au fond, dans l'encoignure où quelque humble vaisselle  
Aux planches d'un bahut vaguement étincelle,  
On distingue un grand lit aux longs rideaux tombants.  
Tout près, un matelas s'étend sur de vieux bancs,  
Et cinq petits enfants, nid d'âmes, y sommeillent  
La haute cheminée où quelques flammes veillent  
Rougit le plafond sombre, et, le front sur le lit,  
Une femme à genoux prie, et songe, et pâlit.  
C'est la mère. Elle est seule. Et dehors, blanc d'écume,  
Au ciel, aux vents, aux rocs, à la nuit, à la brume,  
Le sinistre océan jette son noir sanglot.*

Exemple (plus simple) :

*Il est nuit. La cabane est pauvre, mais bien close.*

*Le logis est plein d'ombre et l'on sent quelque chose*



Exemple (plus simple) :

*Il est nuit. La cabane est pauvre, mais bien close.*

*Le logis est plein d'ombre et l'on sent quelque chose*

```
vectorizer = CountVectorizer(analyzer='word')  
ft = vectorizer.fit_transform(pauvres_gens)  
ft.todense()
```

```
[1, 1, 0, 1, 2, 0, 1, 1, 0, 0, 1, 1, 0, 0, 1, 0, 0, 0]  
[0, 0, 1, 0, 1, 1, 0, 0, 1, 1, 0, 0, 1, 1, 0, 1, 1, 1]
```

```
vectorizer = CountVectorizer(analyzer='word')
ft = vectorizer.fit_transform(les_pauvres_gens)
cosine_distance = 1 - cosine_similarity(ft)

for i in range(cosine_distance.shape[0]):
    for j in range(cosine_distance.shape[1]):
        if i < j:
            if cosine_distance[i, j] < .43:
                print('{i:3}, {j:3}: d={d:.2}\n    {t1}\n    {t2}'.format(
                    i=i, j=j, d=cosine_distance[i, j],
                    t1=les_pauvres_gens[i], t2=les_pauvres_gens[j]))
```

Pluie ou bourrasque, il faut qu'il sorte, il faut qu'il aille,  
Il n'avait pas assez de peine ; il faut que j'aille

Pluie ou bourrasque, il faut qu'il sorte, il faut qu'il aille,  
Quand il verra qu'il faut nourrir avec les nôtres

I s'en va dans l'abîme et s'en va dans la nuit.  
Or, la nuit, dans l'ondée et la brume, en décembre,

Comme il faut calculer la marée et le vent !  
Et l'onde et la marée et le vent en colère.

C'est l'heure où, gai danseur, minuit rit et folâtre  
Et c'est l'heure où minuit, brigand mystérieux,

Sous sa cape aux longs plis qu'est-ce donc qu'elle emporte ?  
Qu'est-ce donc qu'elle cache avec un air troublé

```
bigram_vectorizer = CountVectorizer(ngram_range=(1,2))
bigram_ft = bigram_vectorizer.fit_transform(les_pauvres_gens)
bigram_analyze = bigram_vectorizer.build_analyzer()
bigram_analyze(phrase_1)
bigram_cosine_distance = 1 - cosine_similarity(bigram_ft)

for i in range(bigram_cosine_distance.shape[0]):
    for j in range(bigram_cosine_distance.shape[1]):
        if i < j:
            if bigram_cosine_distance[i, j] < .43:
                print(' {i:3}, {j:3}: d={d:.2}\n    {t1}\n    {t2}'.format(
                    i=i, j=j, d=bigram_cosine_distance[i, j],
                    t1=les_pauvres_gens[i], t2=les_pauvres_gens[j]))
```

'Comme il faut calculer la marée et le vent !'

['comme', 'il', 'faut', 'calculer', 'la', 'marée', 'et', 'le', 'vent',  
'comme il', 'il faut', 'faut calculer', 'calculer la', 'la marée', 'marée  
et', 'et le', 'le vent']

Pluie ou bourrasque, il faut qu'il sorte, il faut qu'il aille,  
Quand il verra qu'il faut nourrir avec les nôtres

Comme il faut calculer la marée et le vent !  
Et l'onde et la marée et le vent en colère.

Qu'est-ce donc que Jeannie a fait chez cette morte ?  
Qu'est-ce donc que Jeannie emporte en s'en allant ?

Sous sa cape aux longs plis qu'est-ce donc qu'elle emporte ?  
Qu'est-ce donc qu'elle cache avec un air troublé

```
emacs@siegfried - analytics.py : /home/jeff/src/jellybooks/analytics/visualisation/jellybooks/analytics.py
0.02 emacs@siegfried - analytics.py : /home/jeff/src/jellybooks/analytics/jellybooks/analytics.py
0.04 emacs@siegfried - analytics.py : /home/jeff/src/jellybooks/analytics_backtrack/visualisation/jellybooks/analytics.py
0.06 emacs@lorax - analytics.py : /home/jeff/src/jellybooks/analytics/jellybooks/analytics.py
0.06 emacs@nantes - analytics.py : /home/jeff/src/jellybooks/analytics/jellybooks/analytics.py
0.06 emacs@birdsong - analytics.py : /home/jeff/src/jellybooks/analytics/jellybooks/analytics.py
0.14 emacs@siegfried - analytics.py : /home/jeff/src/jellybooks/analytics/visualisation/jellybooks/analytics.py
0.17 emacs@siegfried - analytics.py : /home/jeff/src/jellybooks/analytics/jellybooks/analytics.py
0.19 emacs@siegfried - baldur.py : /home/jeff/src/jellybooks/analytics/visualisation/jellybooks/baldur.py
0.19 emacs@siegfried - plot_local.py : /home/jeff/src/jellybooks/analytics/visualisation/jellybooks/plot_local.py
0.19 emacs@siegfried - actions.py : /home/jeff/src/jellybooks/analytics/visualisation/jellybooks/actions.py
```

```
emacs@siegfried - analytics.py : /home/jeff/src/jellybooks/analytics/visualisation/jellybooks/analytics.py

0.28 emacs@siegfried - Makefile<analytics> : /home/jeff/src/jellybooks/analytics/Makefile
0.28 emacs@siegfried - demo.py : /home/jeff/src/jellybooks/analytics/visualisation/demo.py
0.28 emacs@siegfried - demo_balduz.py : /home/jeff/src/jellybooks/analytics/visualisation/demo_balduz.py
0.28 emacs@birdsong - registration_progress_pb2.py :
/home/jeff/src/jellybooks/analytics/mongo_links/jellybooks/registration_progress_pb2.py
0.29 jeff@siegfried: /src/jellybooks/analytics/visualisation/d3
0.3 emacs@birdsong - fetch_sync_reminders.py : /home/jeff/src/jellybooks/analytics/fetch_sync_reminders.py
0.3 emacs@nantes - demographic_decay_plots.py : /home/jeff/src/jellybooks/analytics/demographic_decay_plots.py
0.3 emacs@birdsong - reading_speed.py : /home/jeff/src/jellybooks/analytics/reading_speed.py
0.3 emacs@birdsong - manual_decay_plots.py : /home/jeff/src/jellybooks/analytics/manual_decay_plots.py
0.3 emacs@birdsong - demographic_decay_plots.py : /home/jeff/src/jellybooks/analytics/demographic_decay_plots.py
```



# Matches 0–10, bigram(1, 2)

```
emacs@siegfried - analytics.py : /home/jeff/src/jellybooks/analytics/visualisation/jellybooks/analytics.py
0.052 emacs@siegfried - analytics.py : /home/jeff/src/jellybooks/analytics/jellybooks/analytics.py
0.08 emacs@siegfried - analytics.py : /home/jeff/src/jellybooks/analytics_backtrack/visualisation/jellybooks/analytics.py
0.13 emacs@lorax - analytics.py : /home/jeff/src/jellybooks/analytics/jellybooks/analytics.py
0.13 emacs@nantes - analytics.py : /home/jeff/src/jellybooks/analytics/jellybooks/analytics.py
0.13 emacs@birdsong - analytics.py : /home/jeff/src/jellybooks/analytics/jellybooks/analytics.py
0.19 emacs@siegfried - analytics.py : /home/jeff/src/jellybooks/analytics/visualisation/jellybooks/analytics.py
0.26 emacs@siegfried - analytics.py : /home/jeff/src/jellybooks/analytics/jellybooks/analytics.py
0.28 emacs@siegfried - baldur.py : /home/jeff/src/jellybooks/analytics/visualisation/jellybooks/baldur.py
0.28 emacs@siegfried - plot_local.py : /home/jeff/src/jellybooks/analytics/visualisation/jellybooks/plot_local.py
0.28 emacs@siegfried - actions.py : /home/jeff/src/jellybooks/analytics/visualisation/jellybooks/actions.py
```

# Matches 90–100, bigram(1, 2)

```
emacs@siegfried - analytics.py : /home/jeff/src/jellybooks/analytics/visualisation/jellybooks/analytics.py
0.41 emacs@siegfried - .gitignore : /home/jeff/src/jellybooks/analytics/.gitignore
0.41 emacs@siegfried - Makefile : /home/jeff/src/jellybooks/analytics/Makefile
0.41 jeff@siegfried: /src/jellybooks/analytics
0.41 emacs@birdsong - fetch_sync_reminders.py : /home/jeff/src/jellybooks/analytics/fetch_sync_reminders.py
0.41 emacs@nantes - demographic_decay_plots.py : /home/jeff/src/jellybooks/analytics/demographic_decay_plots.py
0.41 emacs@birdsong - reading_speed.py : /home/jeff/src/jellybooks/analytics/reading_speed.py
0.41 emacs@birdsong - manual_decay_plots.py : /home/jeff/src/jellybooks/analytics/manual_decay_plots.py
0.41 emacs@birdsong - demographic_decay_plots.py : /home/jeff/src/jellybooks/analytics/demographic_decay_plots.py
0.41 emacs@birdsong - fetch_email_events.py : /home/jeff/src/jellybooks/analytics/fetch_email_events.py
0.41 emacs@nantes - manual_decay_plots.py : /home/jeff/src/jellybooks/analytics/manual_decay_plots.py
```

# Matches 0–10, Tf-Idf

```
emacs@siegfried - analytics.py : /home/jeff/src/jellybooks/analytics/visualisation/jellybooks/analytics.py
0.057 emacs@siegfried - analytics.py : /home/jeff/src/jellybooks/analytics/jellybooks/analytics.py
0.097 emacs@nantes - analytics.py : /home/jeff/src/jellybooks/analytics/jellybooks/analytics.py
0.1 emacs@birdsong - analytics.py : /home/jeff/src/jellybooks/analytics/jellybooks/analytics.py
0.1 emacs@siegfried - analytics.py : /home/jeff/src/jellybooks/analytics_backtrack/visualisation/jellybooks/analytics.py
0.11 emacs@lorax - analytics.py : /home/jeff/src/jellybooks/analytics/jellybooks/analytics.py
0.21 analytics.py
0.23 emacs@siegfried - html.py : /home/jeff/src/jellybooks/analytics/visualisation/jellybooks/html.py
0.26 jeff@siegfried: /src/jellybooks/analytics/visualisation
0.27 jeff@siegfried: /src/jellybooks/analytics/visualisation/jellybooks
0.27 emacs@nantes - html.py : /home/jeff/src/jellybooks/analytics/visualisation/jellybooks/html.py
```

```
emacs@siegfried - analytics.py : /home/jeff/src/jellybooks/analytics/visualisation/jellybooks/analytics.py
0.48 emacs@siegfried - baldur.py : /home/jeff/src/jellybooks/analytics_backtrack/visualisation/jellybooks/baldur.py
0.48 jeff@candy: /src/jellybooks/analytics
0.48 jeff@siegfried: /src/jellybooks/analytics/visualisation/nvd3-examples
0.49 emacs@nantes - plot_d3.py : /home/jeff/src/jellybooks/analytics/jellybooks/plot_d3.py
0.49 emacs@siegfried - demo.py<2> : /home/jeff/src/jellybooks/analytics_backtrack/visualisation/jellybooks/demo.py
0.49 emacs@birdsong - plot_d3.py : /home/jeff/src/jellybooks/analytics/jellybooks/plot_d3.py
0.49 emacs@siegfried - rex.py : /home/jeff/src/jellybooks/analytics/src/rex.py
0.49 emacs@siegfried - sessions_pb2.py : /home/jeff/src/jellybooks/analytics/jellybooks/sessions_pb2.py
0.49 emacs@siegfried - png_test.py : /home/jeff/src/jellybooks/analytics/jellybooks/png_test.py
0.49 emacs@siegfried - im_png_test.py : /home/jeff/src/jellybooks/analytics/jellybooks/im_png_test.py
```

$$TF_{td} = \frac{f_{td}}{\max_k f_{kd}} \quad IDF_t = \log_2 \left( \frac{N}{n_t} \right)$$

$$TF\text{-}IDF_{td} = TF_{td} \cdot IDF_t$$

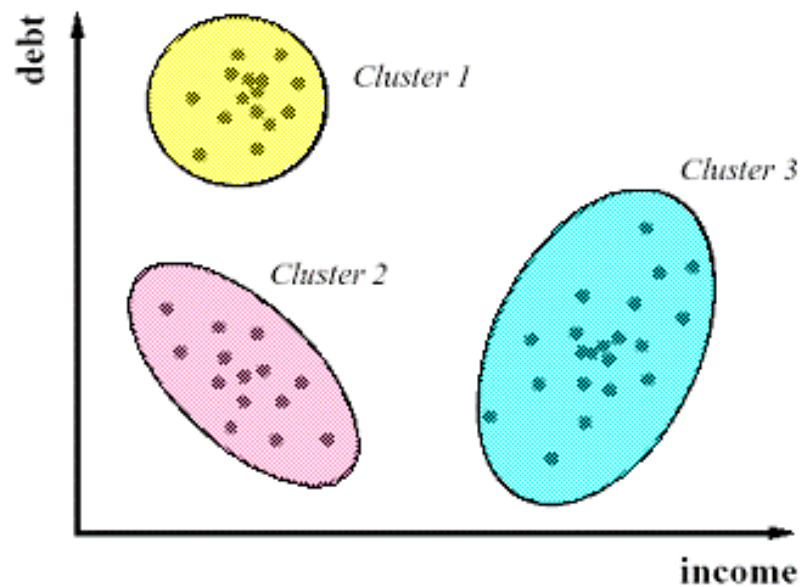
with

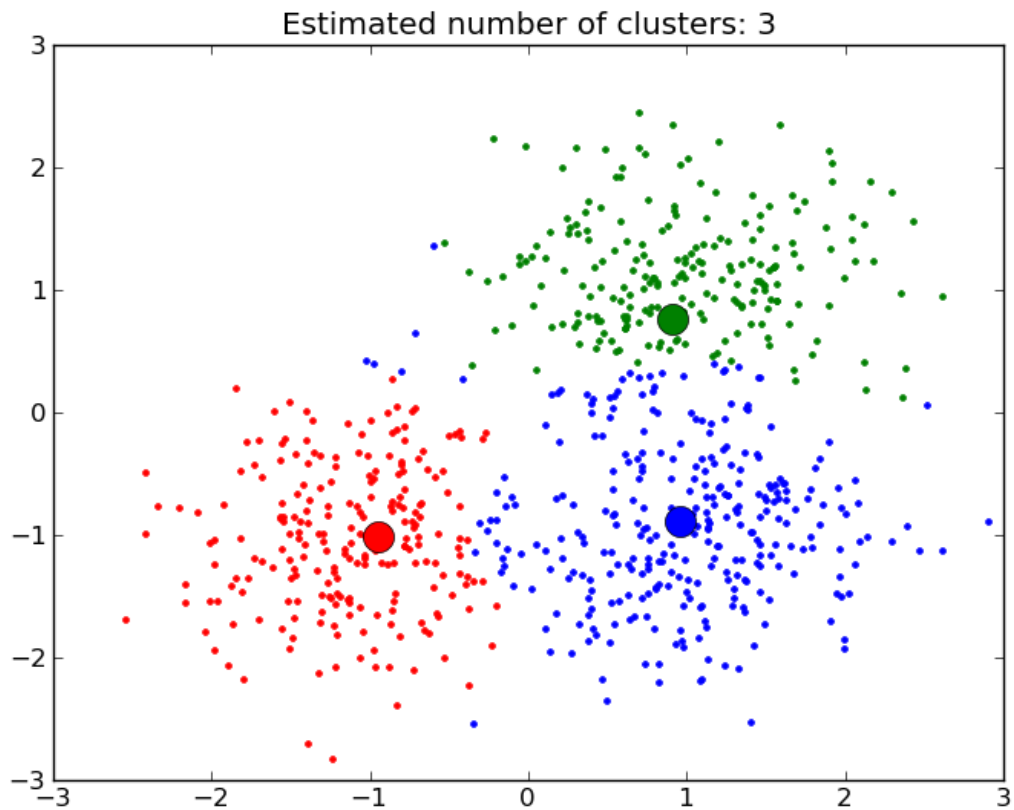
$f_{td}$  = frequency of word (term)  $t$  in document  $d$

$N$  = number of documents

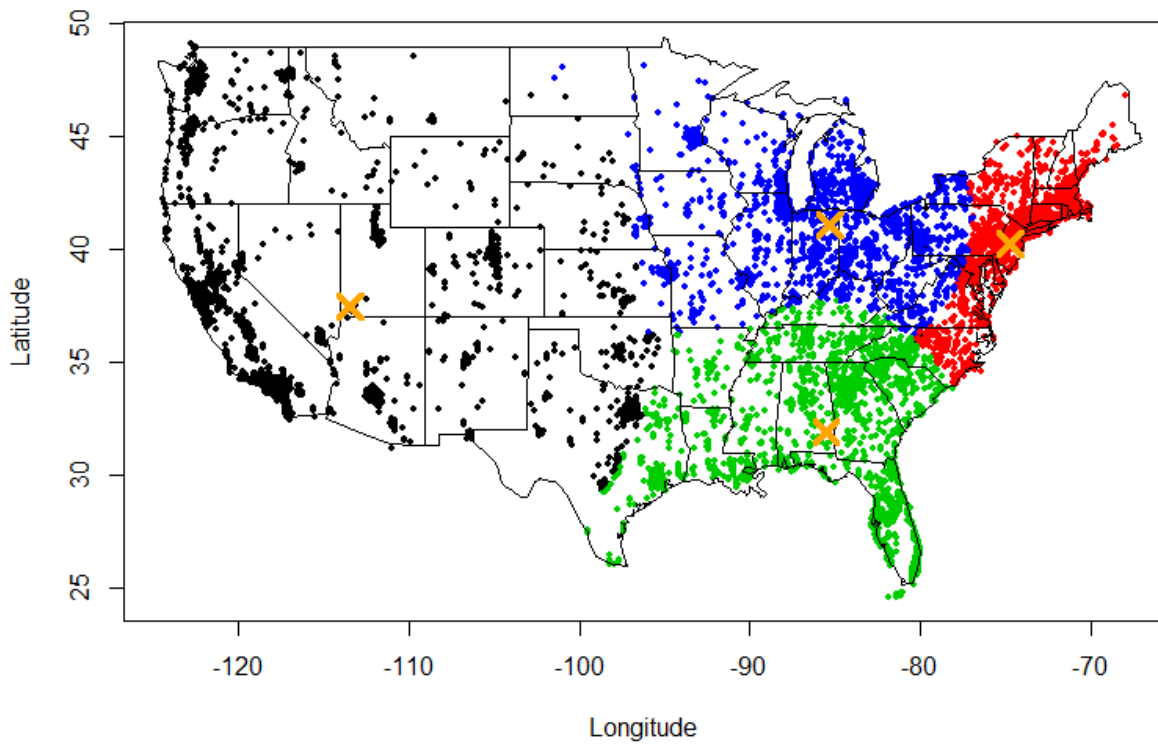
$n_t$  = number of documents containing term  $t$

# *k*-means

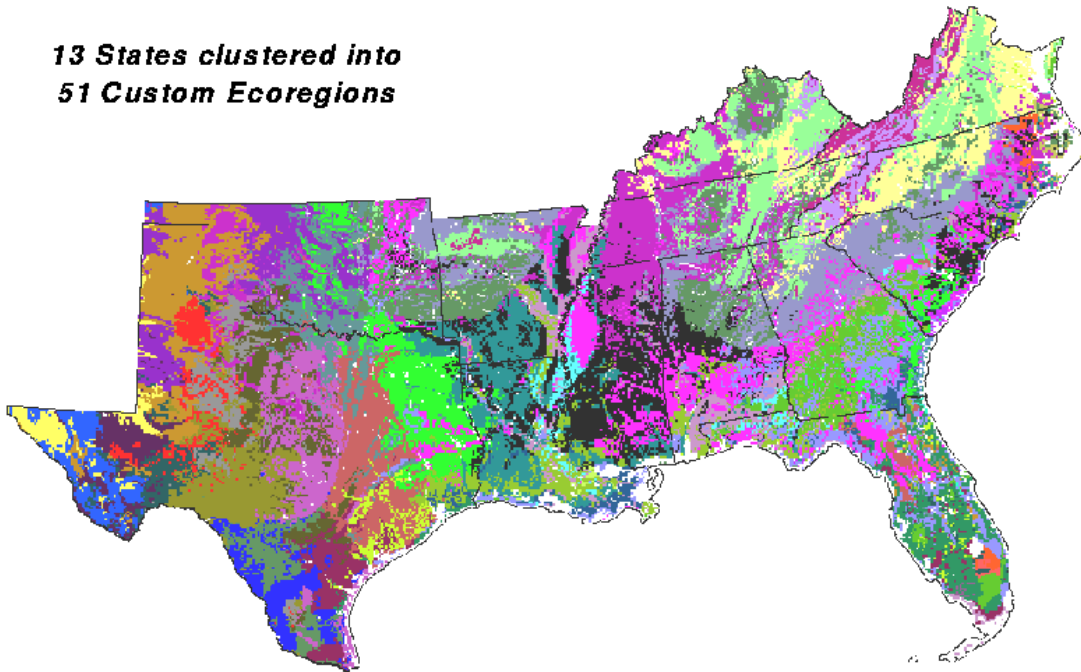


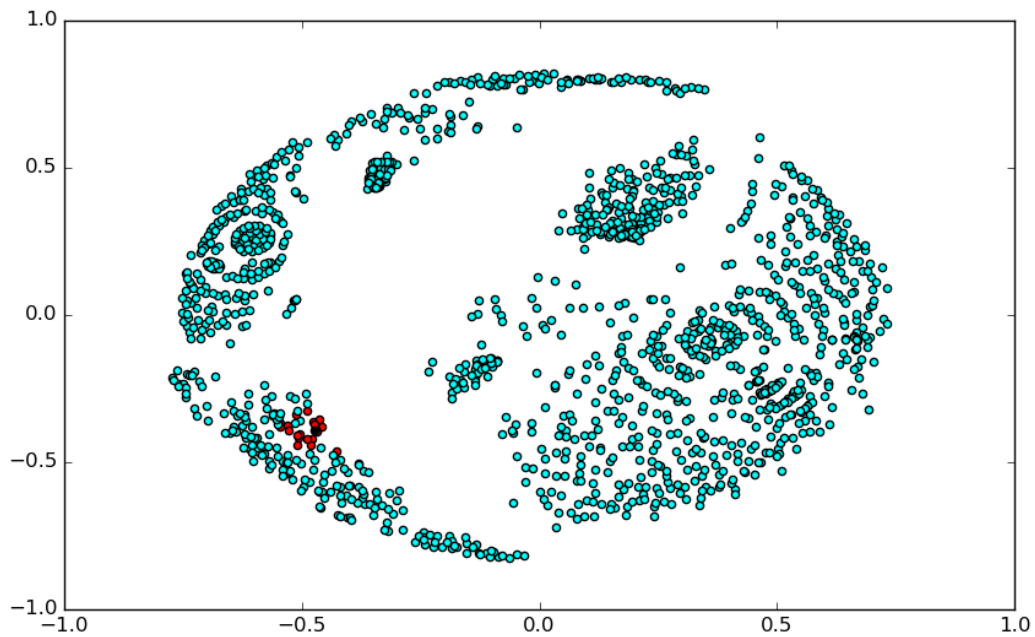


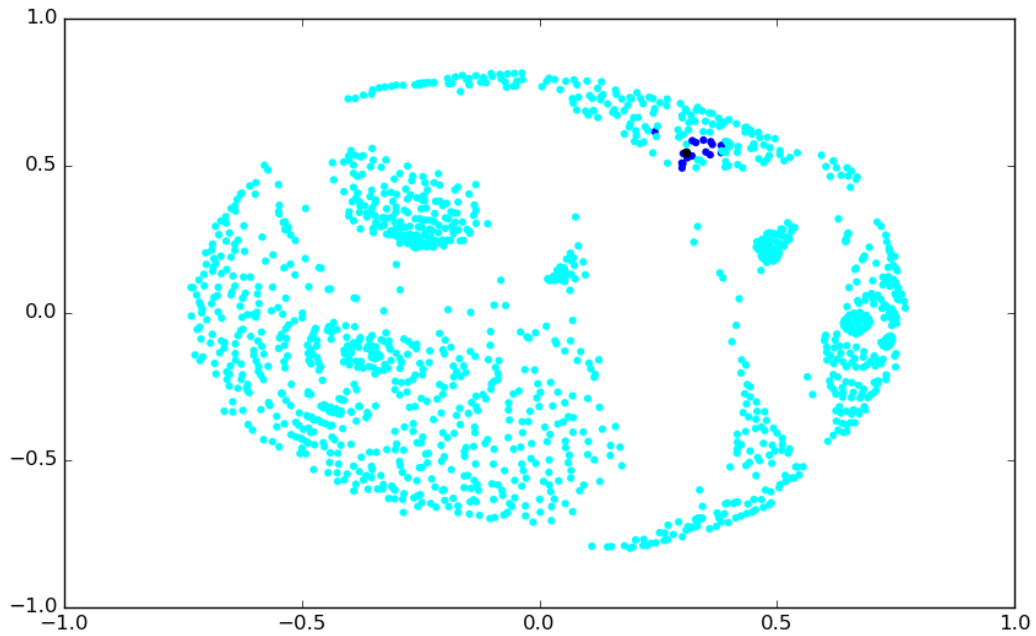


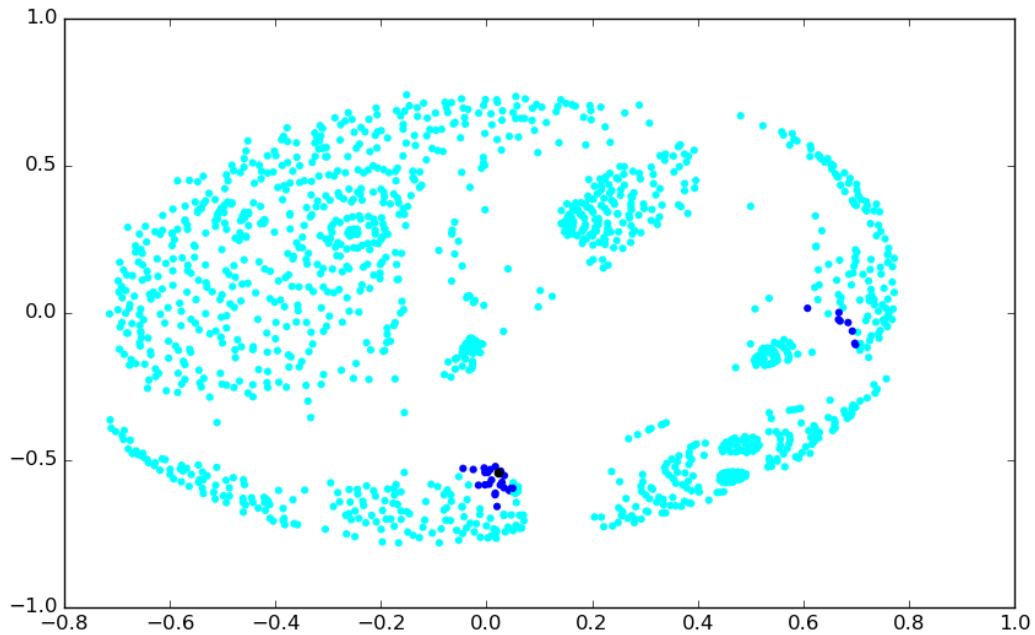


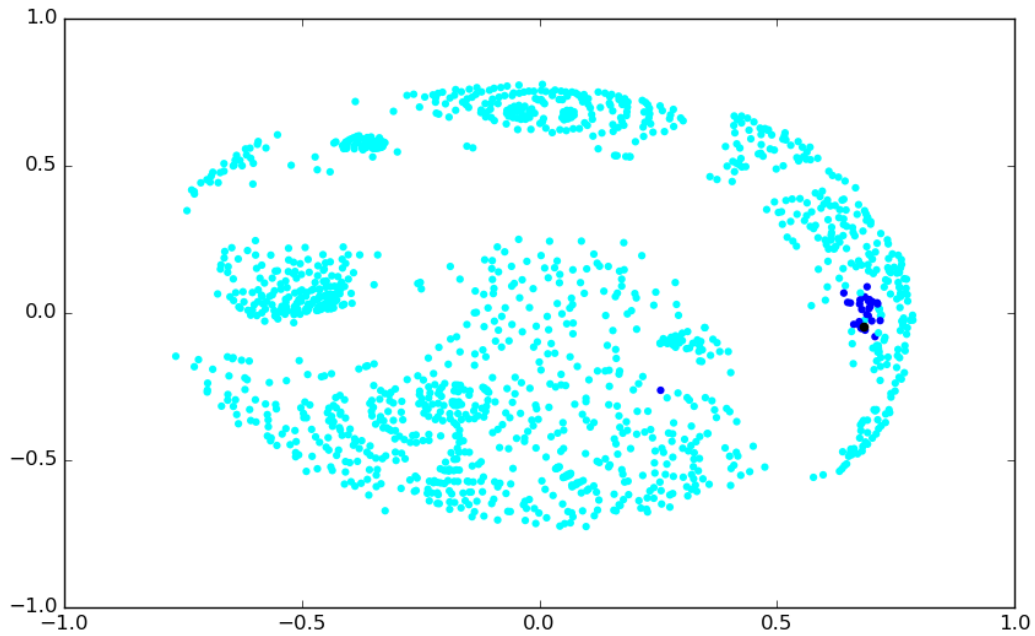
***13 States clustered into  
51 Custom Ecoregions***

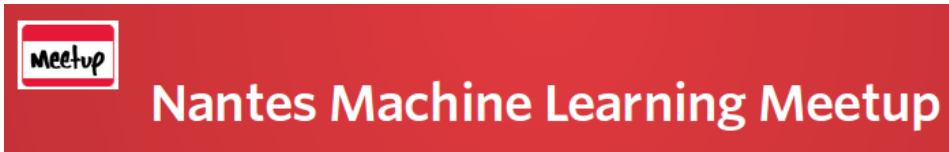












<http://www.meetup.com/Nantes-Machine-Learning-Meetup/>



<http://www.ml-week.com/>



**Questions?**