



#ANALYSE DE SENTIMENTS
APPLIQUÉE AUX TWEETS

R Ladies Meetup

Paris, le 24 janvier 2018

by Ahès Roulier & Tanguy Aurore

LE GROUPE ESTIA - HISTOIRE & CHIFFRES

8 ANNÉES
DE SUCCÈS

2

bureaux

Europe (Paris), Amérique
du Nord (Montréal)



100 collaborateurs
experts de la donnée
et de son exploitation à
forte valeur ajoutée



+ de 30 clients
grands comptes

**Un cœur de métier
orienté autour de
la donnée**

6 offres majeures

PILOTER

COMPRENDRE

COMMUNIQUER

PRÉVOIR

GÉRER

TRANSFORMER

1 centre de formation agréé
Estiaformation

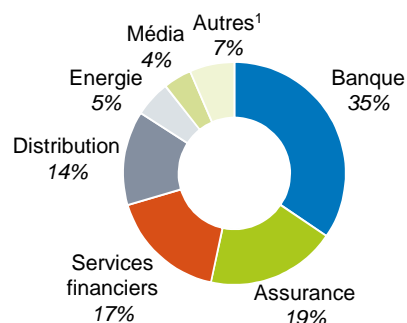
**1 parfaite maîtrise des outils de
management de la donnée**

1 laboratoire
de R&D



Estialab!

Turn data into business value



¹ Autres : transports, services, loisirs, construction, immobiliser, FMCG



Les Games of Data

- 5 animations dans l'année sous forme de compétition
- 4 binômes de consultants
- Sujets d'innovations sélectionnés par les consultants ou proposés par des clients
- 2 jours consacrés à la compétition
- Exemples de sujets déjà traités
 - Analyse de tweets
 - Participation à un Kaggle sur la consommation d'énergie
 - Création d'une application de test pour le recrutement
 - Analyse d'images avec du deep learning
 - Utilisation de D3JS pour la DataViz
 - ...

DATA SCIENCE IS COMING

Turn data into business value

INTRODUCTION

Accueil À propos

Recherchez sur Twitter

Vous avez déjà un compte ? [Se connecter](#)



Tweets **59,8 k** Abonnements **40,8 k** Abonnés **6,68 M** J'aime **8 440** Listes **4**

Suivre

Game Of Data

@GameOfData

Tweet what is yours. #GameOfData Watch episodes of #GoD on today: itsh.bo/PrepareForWint...

Groupe Estia

gameofdata.com

Inscrit en août 2010

Né le 26 octobre

38,7 k Photos et vidéos

Tweets Tweets & réponses Médias

Tweet épinglé
Game Of Data @GameOfData · 26 sept.

Exploiter les données des réseaux sociaux

430 2,3 k 7,2 k

Game Of Data @GameOfData · 9 oct.

Améliorer la relation client grâce à l'analyse de sentiment

430 2,3 k 7,2 k

Game Of Data @GameOfData · 9 oct.

Plus besoin d'enquête de satisfaction coûteuse en temps et en argent

430 2,3 k 7,2 k

Game Of Data @GameOfData · 9 oct.

Réagir en temps réel au besoin des clients

Nouveau sur Twitter ?

Inscrivez-vous maintenant pour obtenir votre fil d'actualités personnalisé !

S'inscrire

© 2017 Twitter À propos Centre d'assistance
Conditions Politique de confidentialité
Cookies Informations sur la publicité

Turn data into business value

ANALYSE DE SENTIMENTS APPLIQUÉE AUX TWEETS



2009

1 Twitter



2010

2 Implémentation



2012

3 Dataviz

Turn data into business value

ANALYSE DE SENTIMENTS APPLIQUÉE AUX TWEETS



2009

1

Twitter



2010

2

Implémentation



2012

3

Dataviz

Turn data into business value

TWITTER – FIL D'ACTUALITÉS



TWEETS

- Twitter : un réseau social sur lequel les utilisateurs peuvent publier des tweets, limités à 140 caractères, photos, vidéos, liens
- Twitter : un échange de messages rapides et fréquents, sous forme de conversations, envoyés depuis un compte





- **#** (« hashtag ») : mot-clé ou sujet
- **@** : précède les noms d'utilisateur
- **Home** : page d'accueil affichant le flux des tweets provenant des comptes suivis par l'utilisateur
- **RT** (« Retweet ») : partage d'un tweet existant à des followers
- **Follower** : personne abonnée à un fil d'actualités
- **Direct Messages** : messages privés échangés sur Twitter



ANALYSE DE SENTIMENTS APPLIQUÉE AUX TWEETS



2009

1

Twitter



2010

2

Implémentation



2012

3

Dataviz

Turn data into business value



- Liste des API Twitter (interface de programmation)

Tweets	Standard (free)	Premium	Enterprise
Publish and engage	✓		
Search Tweets	✓	✓	✓
Filter Tweets	✓		✓
Sample Tweets	✓		✓
Batch Tweets			✓
Direct Messages	✓		
Account and users	✓		✓
Metrics			✓
Ad campaigns	✓		
Publisher tools and SDKs	✓		

Turn data into business value





- **Twitter Search API**

	Standard	Premium	
Limitations	Gratuit	150\$	2 500\$
Tweets / requêtes	100	500	500
Requêtes / mois	250	500	10 000
Historique	30 jours	30 jours	30 jours

Ne renvoie que les tweets les plus récents et/ou populaires (certains tweets et utilisateurs peuvent ne pas apparaître)

Référence sur le site de twitter :

<https://developer.twitter.com/en/pricing/search-30day>

Turn data into business value



CRÉER UN COMPTE DEV SUR TWITTER



Create New App

- “Name” : doit être unique
- “Description” : doit contenir au moins 10 caractères
- “Website” : adresse réelle ou fictive

Application Management

<https://apps.twitter.com/>

Create an application

Application Details

Name *

Your application name. This is used to attribute the source of a tweet and in user-facing authorization screens. 32 characters max.

Description *

Your application description, which will be shown in user-facing authorization screens. Between 10 and 200 characters max.

Website *

Your application's publicly accessible home page, where users can go to download, make use of, or find out more information about your application. This fully-qualified URL is used in the source attribution for tweets created by your application and will be shown in user-facing authorization screens.
(If you don't have a URL yet, just put a placeholder here but remember to change it later.)

Callback URL

Cocher « Twitter Developer Agreement »

Create your Twitter Application

Keys and Access Tokens

Turn data into business value



COLLECTE DES DONNÉES

- Choix du fil d'actualités d'un tweetos
- Récupération des tweets
- Identification d'un tweet
- Récupération des réponses

NETTOYAGE ET PRÉ- TRAITEMENT

- Suppression de la ponctuation
- Suppression des « Stopwords »
- Lemmatisation (TreeTagger)

Turn data into business value



COLLECTE DES TWEETS



COLLECTE DES DONNÉES

- Choix du fil d'actualités d'un tweetos
- Récupération des tweets
- Identification d'un tweet
- Récupération des réponses

• Connexion à Twitter depuis R

```
library(twitteR)  
# si le package n'est pas déjà installé, il faut le faire au préalable,  
# avec la commande suivante: install.packages("twitteR", dependencies = TRUE)  
  
setup_twitter_oauth(consumer_key = "XXX",  
                    consumer_secret = "XXX",  
                    access_token = "XXX",  
                    access_secret = "XXX")
```

• Extraction des tweets

```
# récupération des tweets  
tweets <- searchTwitter("#Data", lang="fr", n=5000, resultType= "recent")  
  
# copier la liste dans une structure data.frame  
tweets_df <- twListToDF(tweets)
```



- Nuage de mots sans prétraitement

```
mots <- unlist(strsplit(x = tweets$Tweet, split = " "))  
mots <- as.data.frame(table(mots))  
mots <- mots[order(mots$Freq, decreasing = TRUE), ]  
  
# Création du nuage de mots  
library("wordcloud")  
wordcloud(words = mots$mots, freq = mots$Freq, max.words = 100)
```

- WordCloud sans prétraitement



NETTOYAGE ET PRÉTRAITEMENT



NETTOYAGE ET
PRÉ-
TRAITEMENT

- Suppression de la ponctuation
- Suppression des « Stopwords »
- Lemmatisation (TreeTagger)

Estia accueillera le meetup RLadies le 24 janvier ! #Data

Ponctuation et syntaxe twitter

Estia accueillera le meetup RLadies le 24 janvier ! #Data

StopWord

Estia accueillera le meetup RLadies le 24 janvier

Lemmatisation

Estia **accueillir meetup RLadies 24 janvier**

Turn data into business value

NETTOYAGE - Traitements spécifiques aux tweets



*# Supprimer les hahstags, mentions et URLs qui apparaissent
dans les tweets, pour cela on utilise des expressions régulières :*

```
nettoyage_tweets <- function(tweets){  
  # Suppression des hashtags #  
    tweets <- gsub("#.*?(\\s|$)", " ", tweets)  
  # Suppression des mentions @  
    tweets <- gsub("@.*?(\\s|$)", " ", tweets)  
  # Suppression des URLs http  
    tweets <- gsub("http.*?(\\s|$)", " ", tweets)  
  # Suppression des ReTweets  
    tweets <- gsub("^RT :", " ", tweets)  
  # Suppression de la ponctuation  
    tweets <- gsub("\\W", " ", tweets)  
  # Gestion des espaces  
    tweets <- trimws(tweets)  
    tweets <- gsub("\\s+", " ", tweets)  
  
  return (tweets)  
}
```

Turn data into business value

LEMMATISATION AVEC TREE TAGGER



- Université de Munich

- <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>

```
library("koRpus")
tagged.results <- treetag(
  file = enc2utf8("Estia accueillera le meetup RLadies le 24 janvier"),
  treetagger="manual", format="obj", TT.tknz=FALSE, lang="fr",
  encoding = "utf-8", stopwords = stopwords("fr"),
  TT.options=list(path="C:/TreeTagger", preset="fr")
)
# Sortie de TreeTagger
tagged.results@TT.res
```

	token	tag	desc	stop	lemma
1	Estia	NAM	proper name	FALSE	<unknown>
2	accueillera	VER:futu	verb futur	FALSE	accueillir
3	le	DET:ART	article	TRUE	le
4	meetup	NOM	noun	FALSE	<unknown>
5	RLadies	NOM	noun	FALSE	<unknown>
6	le	DET:ART	article	TRUE	le
7	24	NUM	numeral	FALSE	@card@
8	janvier	NOM	noun	FALSE	janvier



- Suppression des stopwords avec le package « tm »

```
library(tm)

# Corpus à partir des tweets lemmatisés
corpus <- Corpus(x = VectorSource(x = base$Tweet_lemma))

# Suppression des stopwords
corpus <- tm_map(x = corpus, FUN = removeWords,
                 words = c(stopwords("fr"), "être", "avoir", "déjà"))
```

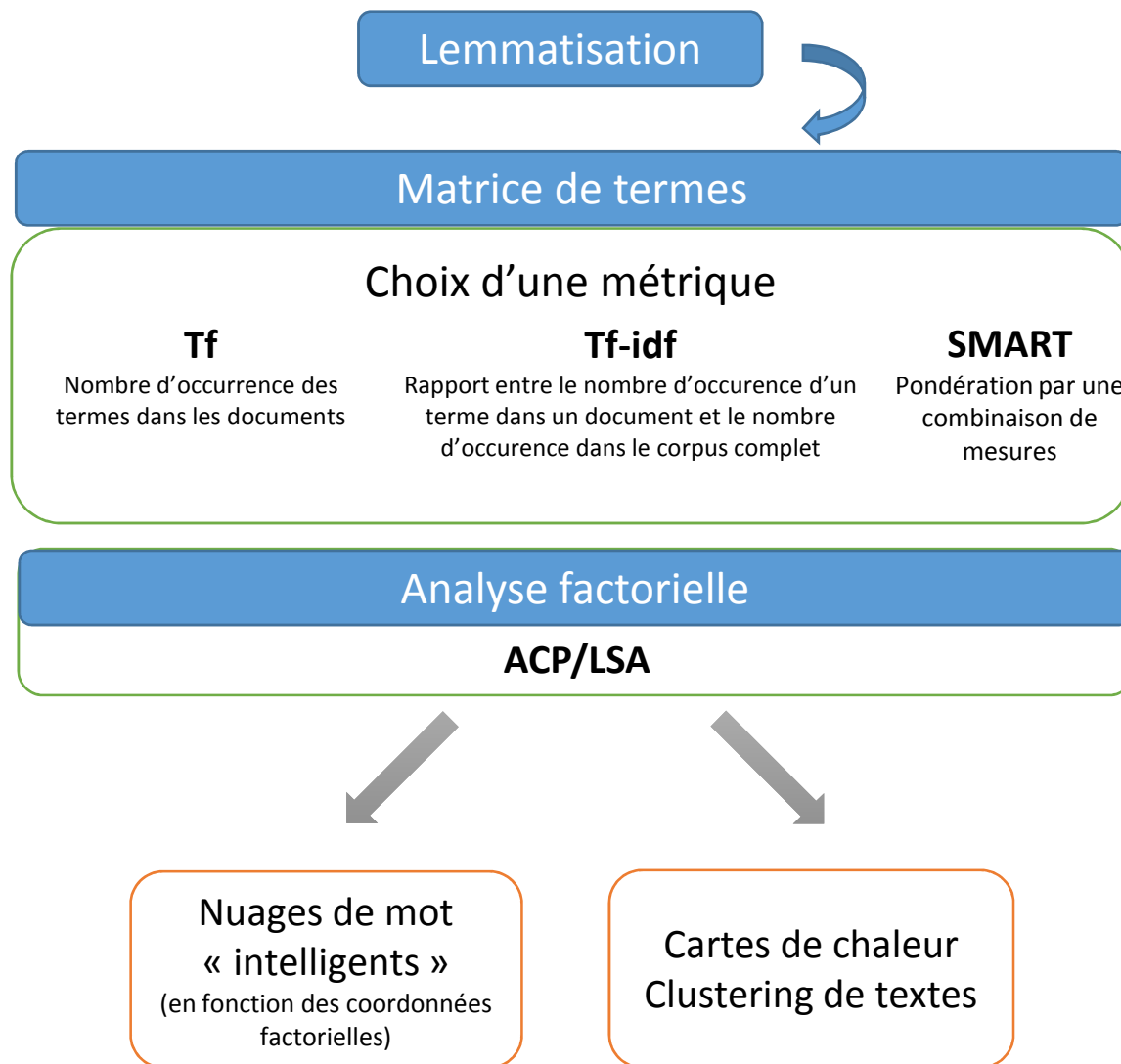
RÉSULTATS



WordCloud après traitement



Turn data into business value



ANALYSES STATISTIQUES PLUS POUSSÉES



	Document 1	Document 2	Document 3	Document 4	Document 5	Document 6	Document 7	Document 8
Term(s) 1	10	0	1	0	0	0	0	2
Term(s) 2	0	2	0	0	0	18	0	2
Term(s) 3	0	0	0	0	0	0	0	2
Term(s) 4	6	0	0	4	6	0	0	0
Term(s) 5	0	0	0	0	0	0	0	2
Term(s) 6	0	0	1	0	0	1	0	0
Term(s) 7	0	1	8	0	0	0	0	0
Term(s) 8	0	0	0	0	0	3	0	0

← Word Vector
(Passage Vector)

Document Vector

Turn data into business value

POLARITÉ



Récupération des #Data

ExitFDP	2018-01-22 17:26:44	Facebook redouble de nouveautés pour vous espionner... et ça fait peur #Facebook #Intrusion #Privacy #Data #Spying □ https://t.co/ne1pP9iHYC
PViotCoster	2018-01-22 17:25:10	@NozhaBoujemaa Les algorithmes donnent du sens aux données #optictalks #data

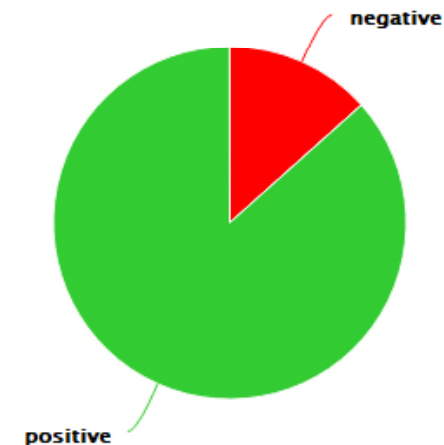


Dictionnaire de polarité

	mot	polarité
144	Bête	NEGATIF
145	Beuglement	NEGATIF
146	Beugler	NEGATIF
147	Bien	POSITIF
148	Bienfaiteur	POSITIF
149	Bienheureux	POSITIF

Polarité des tweets

POLARITY



Turn data into business value

RADAR DE SENTIMENTS



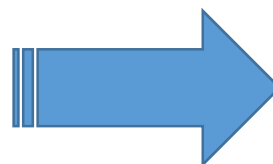
Récupération des #Data

ExitFDP	2018-01-22 17:26:44	Facebook redouble de nouveautés pour vous espionner... et ça fait peur #Facebook #Intrusion #Privacy #Data #Spying □ https://t.co/ne1pP9iHYC
PViotCoster	2018-01-22 17:25:10	@NozhaBoujemaa Les algorithmes donnent du sens aux données #optictalks #data

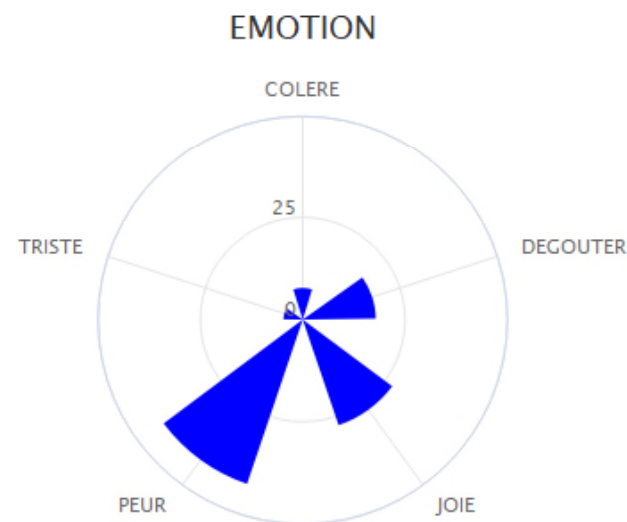


Dictionnaire des sentiments

	mot	polarité
144	Bête	COLERE
145	Beuglement	PEUR
146	Beugler	PEUR
147	Bien	JOIE
148	Bienfaiteur	AUCUN
149	Bienheureux	JOIE



Radar de sentiments



Turn data into business value



ANALYSE DE SENTIMENTS APPLIQUÉE AUX TWEETS



2009

1

Twitter



2010

2

Implémentation



2012

3

Dataviz

Turn data into business value



- **Qu'est-ce que c'est ?**
 - package R (Rstudio)
 - création de pages web interactives
 - pas besoin de HTML, ni CSS, ni JavaScript
 - tout est programmé uniquement en R !
- **Comment ça fonctionne ?**
 - un fichier "ui.R"
 - un fichier "server.R"
- **A quoi ça ressemble ?**

Turn data into business value

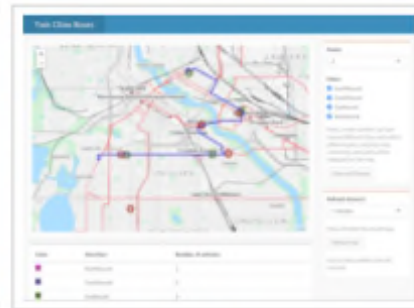


Interactive visualizations

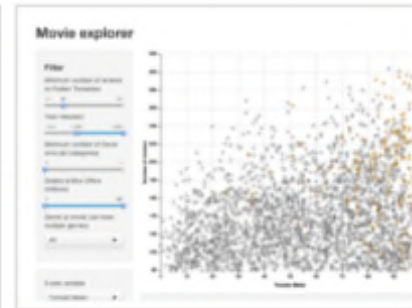
Shiny is designed for fully interactive visualization, using JavaScript libraries like [d3](#), [Leaflet](#), and [Google Charts](#).



SuperZip example



Bus dashboard



Movie explorer



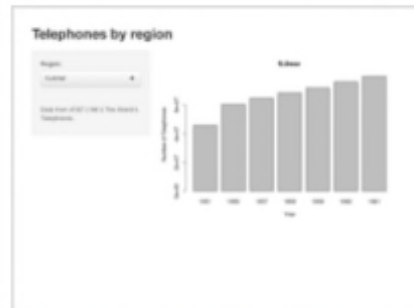
Google Charts

Start simple

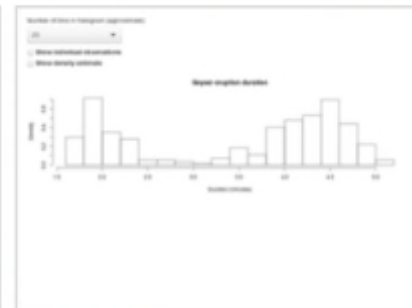
If you're new to Shiny, these simple but complete applications are designed for you to learn from.



Kmeans example



Telephones by region



Faithful



Word cloud

DÉMONSTRATION





Critères d'extraction

Recherche par
☒ Utilisateur ☐ Mots clés

Recherche

Nombre de tweets :

Rechercher

Menu

- Demiers tweets
- Pré-traitement
- Nuages
- Carte de chaleur
- Sentiment

Nom	Date	Tweet	Nb favori	Nb RT
FABERNOVEL_DM	2018-01-22 18:19:26	La #web #analyse 'Data' face #3 ! Apprendre à faire un #audit #data avec nos experts ! #DigitalMarketing #Google https://t.co/3NdG9ly5e	1	0
CartoVista	2018-01-22 18:14:10	The latest version of CartoVista 5.7 will be launched in a few days! Follow us to discover what's new with 5.7! La https://t.co/9x3uExrYon	0	0
led_laboratory	2018-01-22 18:10:05	La data a-t-elle une opinion ? https://t.co/lF1obaj9SQ via @lareclame #data #actu #IA	0	0
Traders_QSL	2018-01-22 18:05:13	Comment passer d'une infrastructure de production obsolète à une infrastructure hébergée de dernière génération? https://t.co/rqmjRqTqy	0	0
josephinekt	2018-01-22 18:00:15	LES TENDANCES À SUIVRE EN 2018 https://t.co/95nPLtkACz #prospective #TENDANCE #Millennials #marketing #publicite https://t.co/QB8XPmqQnY	0	0

Turn data into business value



CONCLUSION

- Simplicité pour réaliser la dataviz
- Traitements avancés avec usage de R
- Détection de la polarité (générer un lexique en français)
- Limitations de l'API



Turn data into business value