

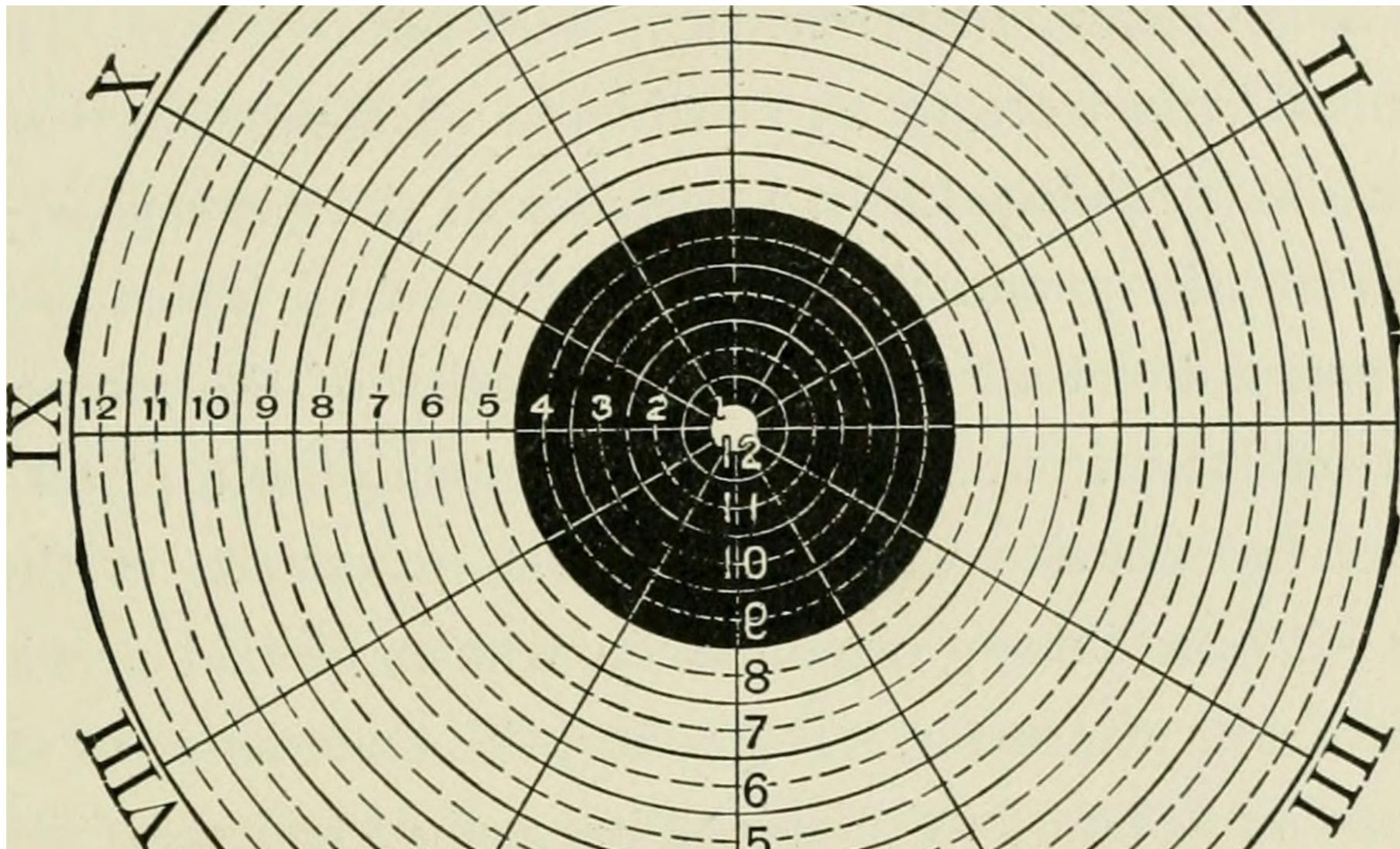
Raconter la Bretagne par ses données



<http://data-bzh.fr>

@DataBzh

contact@data-bzh.fr



Data Bzh - Qui ?

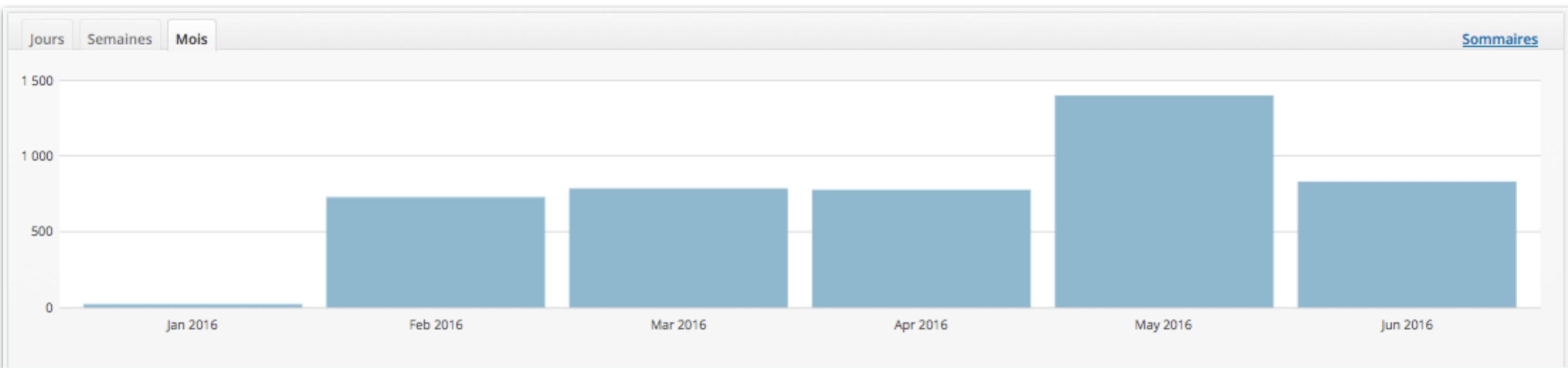
Data Bzh - Un peu d'histoire

- Crée en Janvier 2016
- Objectif : raconter la Bretagne à travers ses données
- Premier article :
#Rennes au mois de janvier



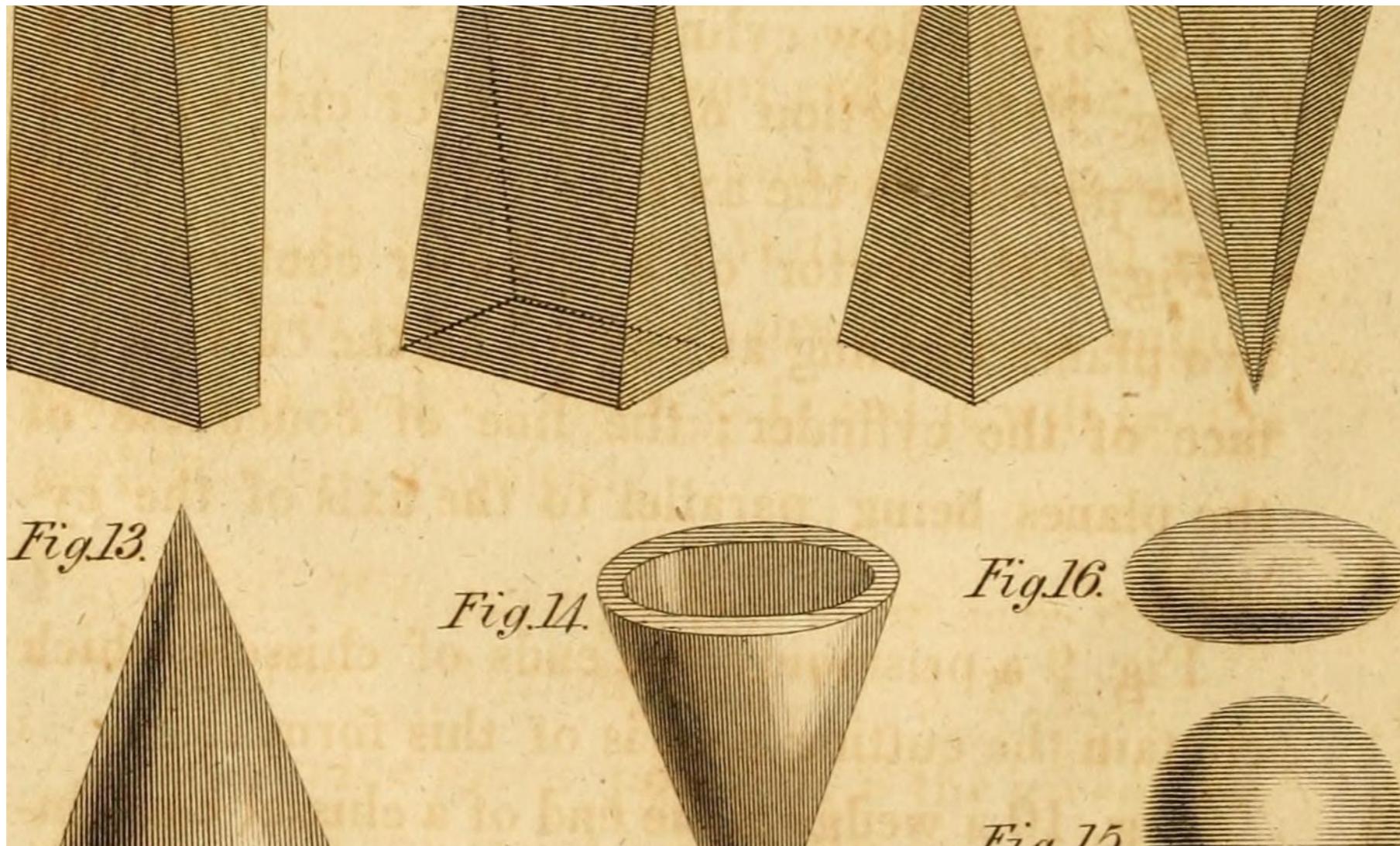
Data Bzh - 6 mois plus tard

- 1000 + followers sur Twitter
- Entre 700 et 1300 visites / mois



Data Bzh - 6 mois plus tard

- 1 rédacteur
- Plusieurs propositions de contributions
- Demain : une équipe qui s'agrandit ?



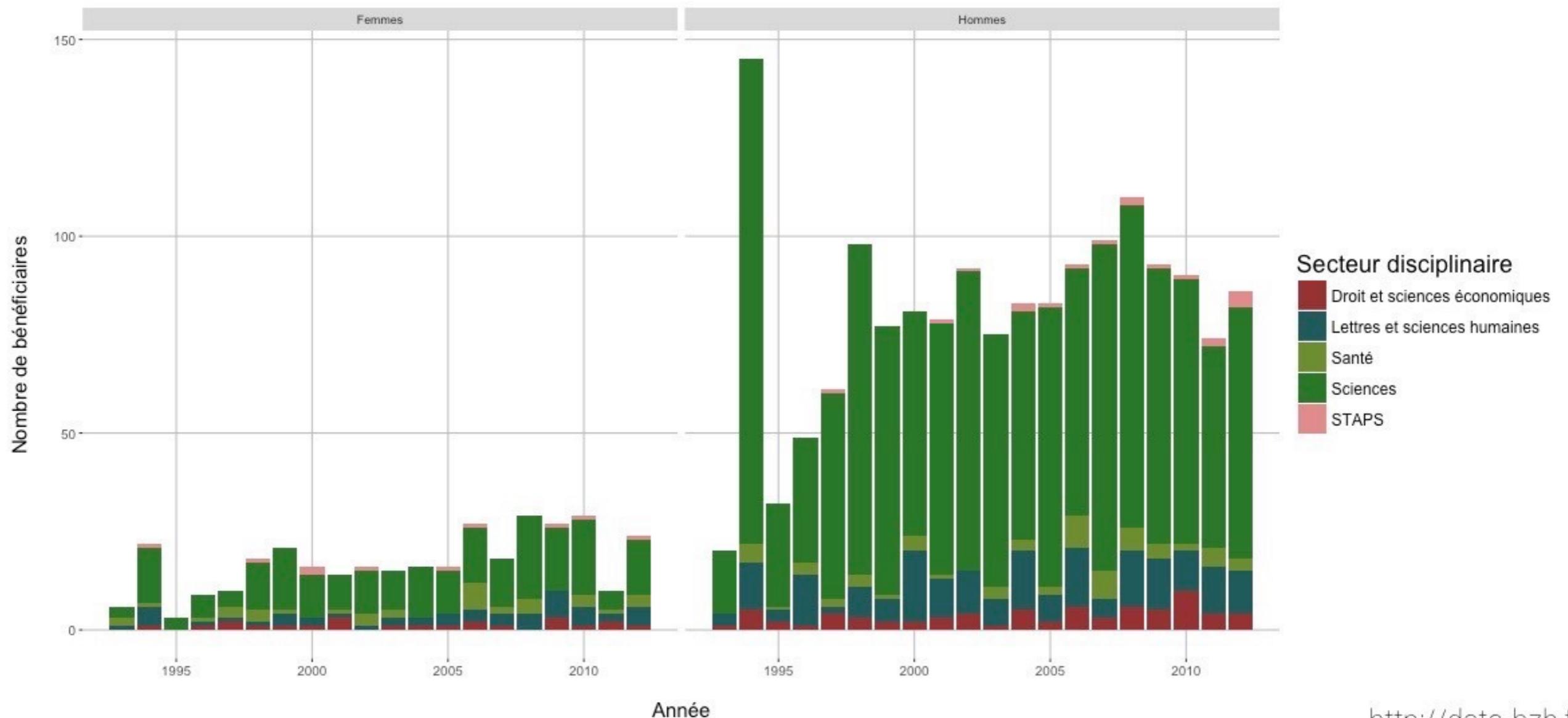
Data Bzh - Pourquoi ?

« Fundamentally learning about the world
through data is really, really cool. »

- Hadley Wickham

Raconter la Bretagne par l'Open data

Effectifs des bénéficiaires bretons de la prime d'excellence scientifique par discipline et par sexe



Raconter la Bretagne par l'Open data

MINISTÈRE
DE L'ÉDUCATION NATIONALE,
DE L'ENSEIGNEMENT SUPÉRIEUR
ET DE LA RECHERCHE

ENSEIGNEMENT SUPÉRIEUR ET
RECHERCHE

OPENDATA

JEUX DE DONNÉES LA DÉMARCHE RÉUTILISATIONS CARTOGRAPHIE API

11 474 enregistrements Aucun filtre actif.

Filtres

Rechercher... 

Année

2012	706
2009	701
2011	700
2010	689
2008	688
2006	664
> Plus	

Sexe

Hommes	7 222
Femmes	4 252

Les bénéficiaires de la prime d'excellence scientifique

Informations Tableau Carte Export API

Ce jeu de données présente les effectifs des bénéficiaires de la prime d'excellence scientifique (ex prime d'encadrement doctoral et de recherche) par sexe, secteur CNU, corps et établissement.

Identifiant du jeu de données: [fr-esr-pes-pedr-beneficiaires](#)

Téléchargements: 3 073

Début de la période couverte: 1 septembre 1993

Fin de la période couverte: 1 septembre 2012

Thème: Education, Formation, Recherche, Enseignement, Administration, Gouvernement, Finances publiques, Citoyenneté

Mots clés: recherche, enseignant-chercheur, enseignement supérieur, région, académie, pedr, pes, université

Licence: Licence Ouverte (Etalab)

Langue: Français

Modifié: 17 juin 2014 10:50

Producteur: Ministère de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche

Références: SIREDO

Suivre les mises à jour  Suivre les mises à jour

Un tableau de 11 474 lignes et de 19 colonnes, pour toute la France...

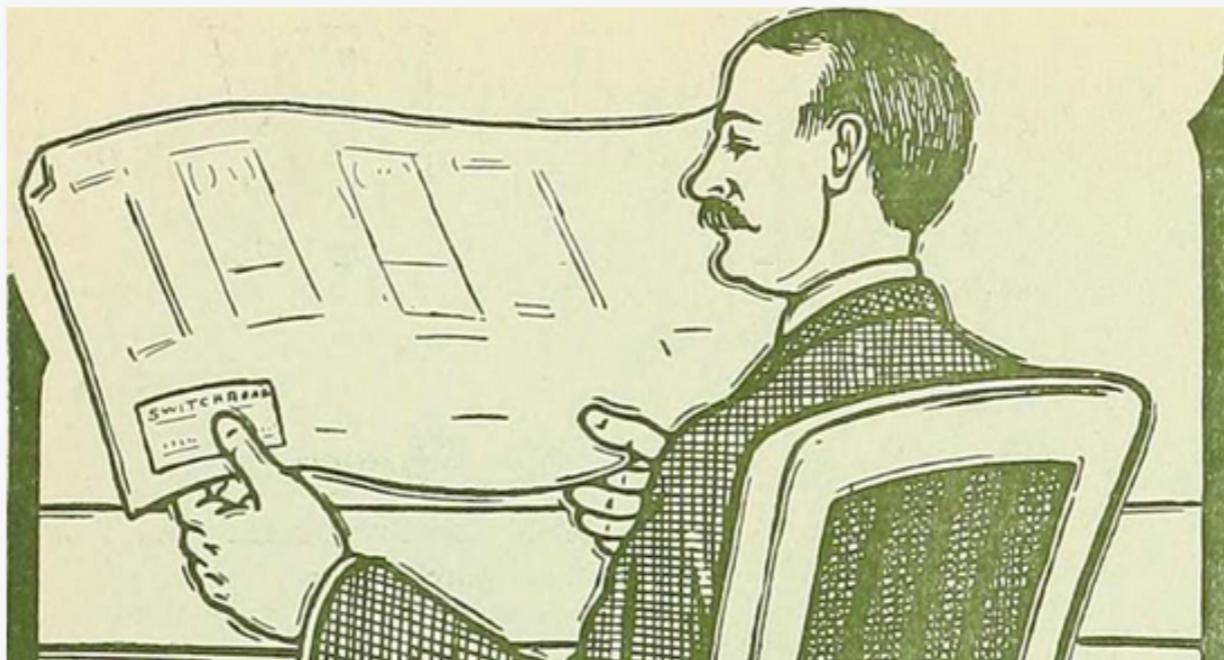
Raconter la Bretagne par l'Open data

The screenshot shows the RStudio interface with the following components:

- Script Editor:** The file "beneficiaire-prime-scientifique.R" is open, containing R code to analyze data from Bretagne. The code includes library imports (dplyr, ggplot2, data.table), data loading (fread), and data manipulation (arrange) followed by ggplot2 visualization.
- Environment Browser:** Shows the global environment with objects like "acad" (11474 obs. of 19 variables), "acad_bret" (522 obs. of 19 variables), and variables "bret", "noms", and "palette".
- Console:** Displays the R command history, including the execution of the R code from the script editor.

... et une trentaine de lignes de code.

30 JOURS SUR OUEST FRANCE EMPLOI



À PROPOS



30 Mai 30 jours sur Ouest France Emploi

Posted at 18:00h in Territoire by Colin

Analyse des données collectées sur l'alerte "Bretagne" du flux RSS de Ouest France Emploi, entre le 29 février et le 29 mars.

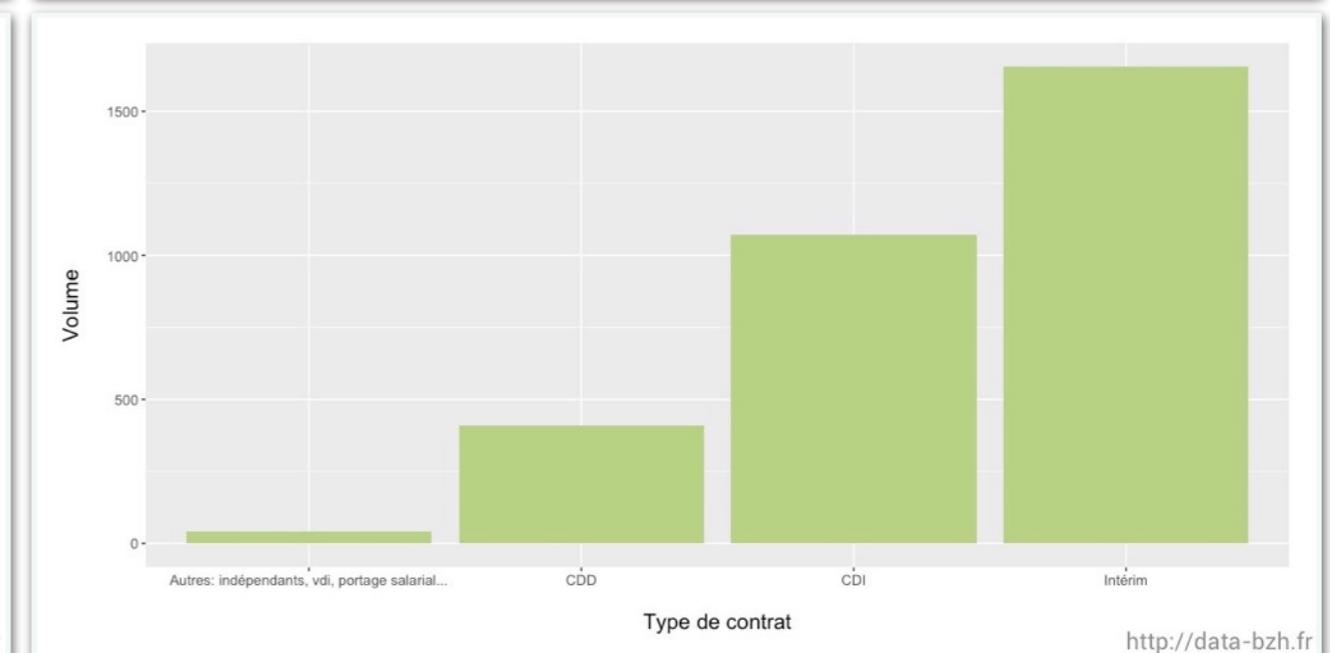
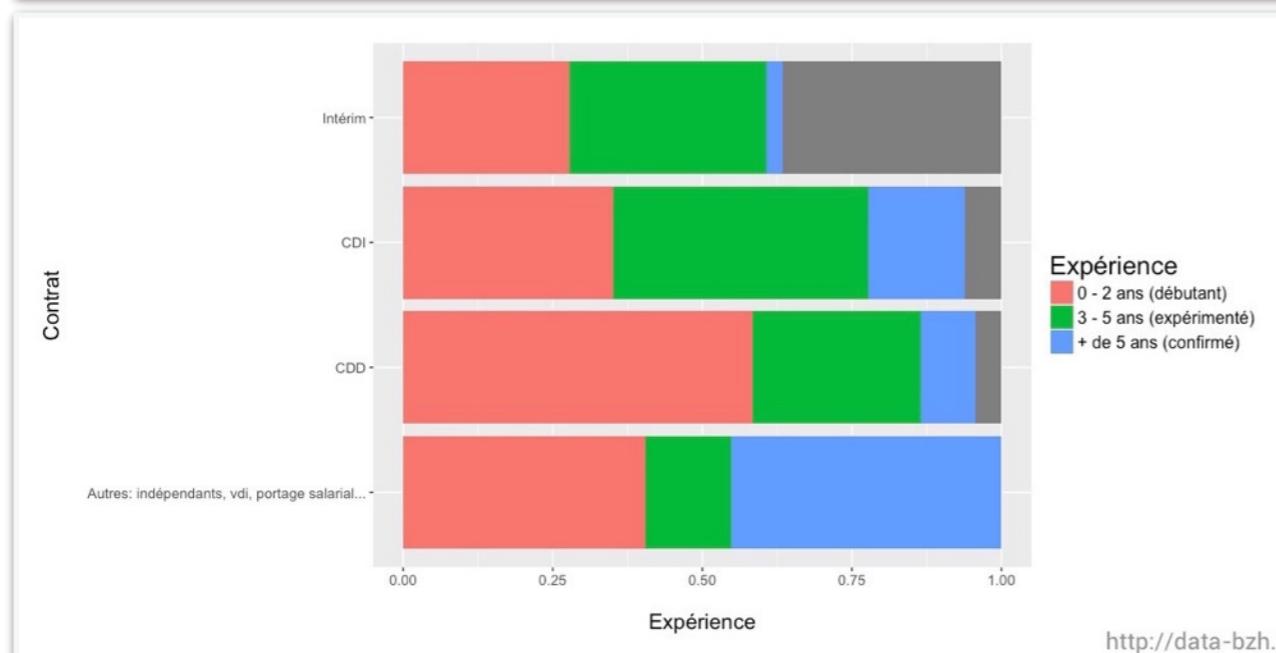
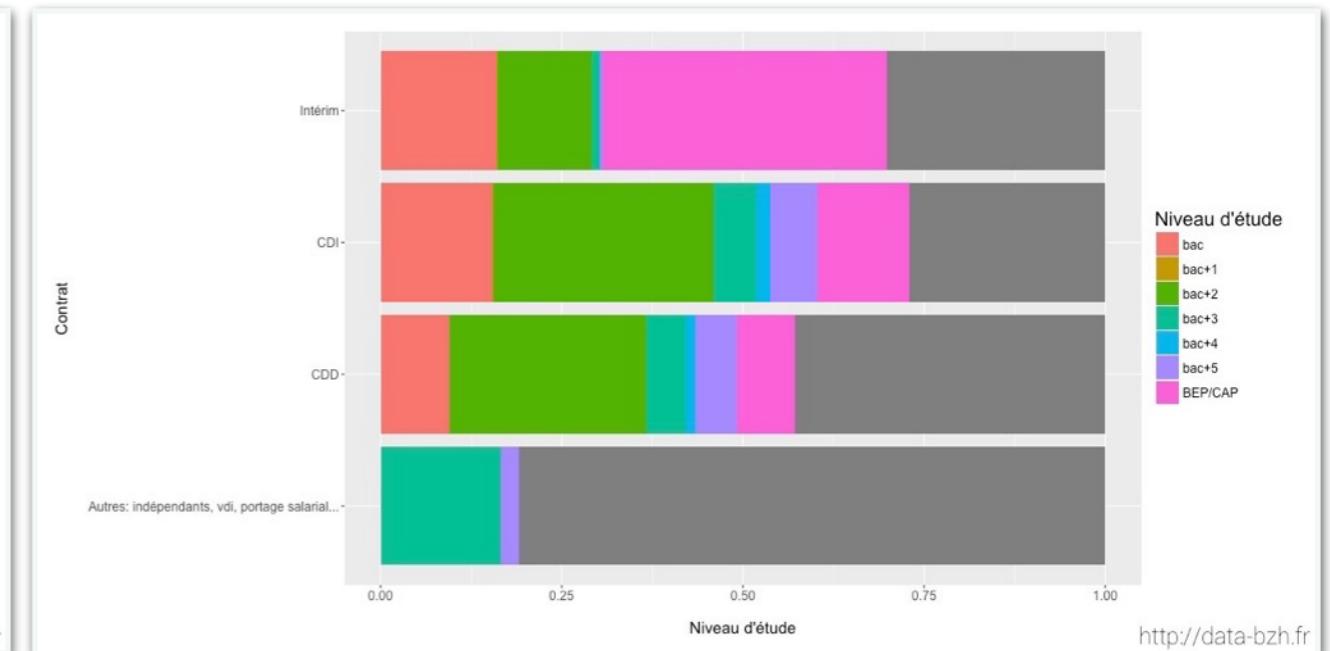
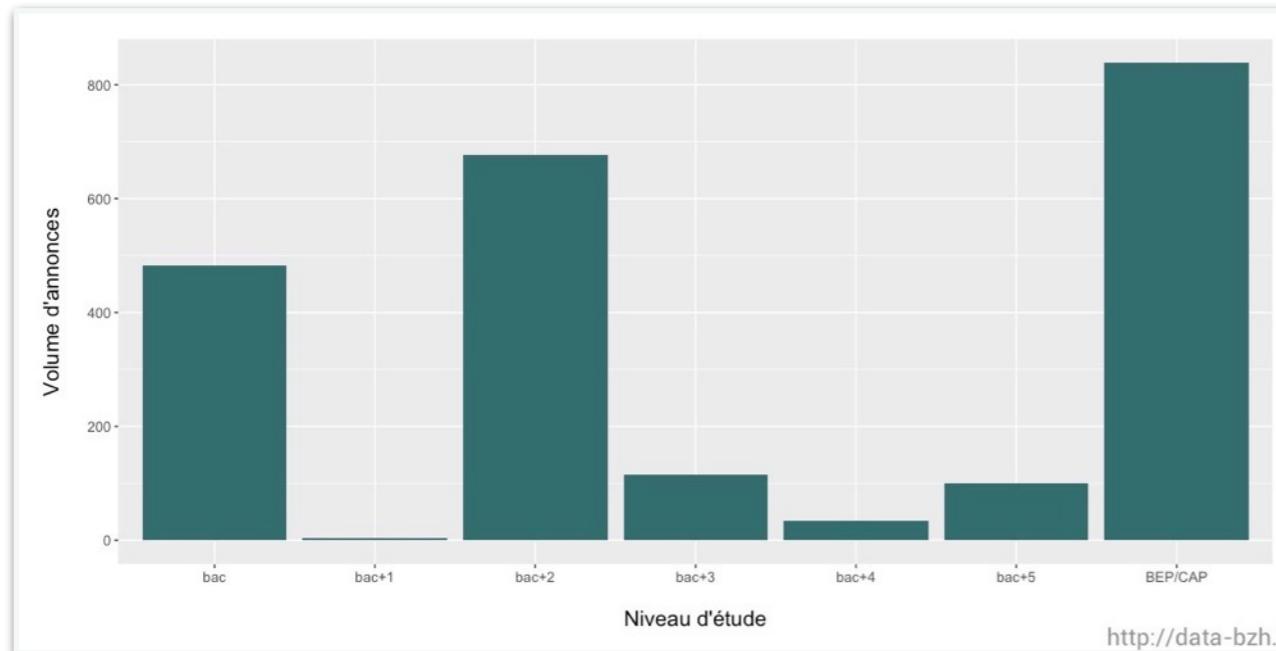
À PROPOS DE CES DONNÉES

Ces données ont été collectées sur le flux RSS de Ouest France Emploi, réglé sur une alerte

Data Bzh souhaite mettre en avant la richesse du territoire breton en collectant, triant et analysant différents jeux de données produits dans la région.

Vous trouverez les résultats de nos expérimentations dans l'onglet Dataviz. Sur l'onglet Blog, des articles d'introduction et d'approfondissement sur les thématiques du site.

30 jours sur Ouest France Emploi



Un tableau de 3334 lignes sur 4 colonnes...

	Date	Annonce	Lien	Entreprise
1	February 29, 2016 at 12:02AM	Collaborateur d'assurances (H/F) – Temporis consult...	http://www.uestfrance-emploi.com/offre-d-emploi...	Assurance. Temporis consulting Brest L'agence ...
2	February 29, 2016 at 12:02AM	Dégraisseur de jambons H/F – SAMSIC RH – Lampaul...	http://www.uestfrance-emploi.com/offre-d-emploi...	Agro/Agri (Métiers du Secteur Agroalimentaire). SAMS...
3	February 29, 2016 at 12:02AM	MENUISIER BOIS H/F – Job-Box – Côtes-d'Armor	http://www.uestfrance-emploi.com/offre-d-emploi...	Menuiserie Opérateurs de production/Conduite de lig...
4	February 29, 2016 at 12:02AM	ASSISTANT ADV (H/F) – Adecco DZ Bretagne Est – Mo...	http://www.uestfrance-emploi.com/offre-d-emploi...	Commerce/Vente Accueil/Secrétariat/Gestion admini...
5	February 29, 2016 at 12:02AM	TECHNICIEN CONSEIL INFORMATIQUE / HOTLINER (H/...	http://www.uestfrance-emploi.com/offre-d-emploi...	Maintenance informatique, Dépannage Qualité. Winph...
6	February 29, 2016 at 12:02AM	PEINTRE EN BATIMENT (H/F) – INTERIM AVENUE – Ren...	http://www.uestfrance-emploi.com/offre-d-emploi...	Décoration, Peinture, Revêtement. INTERIM AVENUE In...
7	February 29, 2016 at 12:02AM	Régleur CN (H/F) – Temporis Ploermel – Ploërmel	http://www.uestfrance-emploi.com/offre-d-emploi...	Métallurg./Mécanique/Aéron. (Mécanique, Aéronautiq...
8	February 29, 2016 at 12:02AM	Conducteur d'installations (H/F) – Manpower Recrute...	http://www.uestfrance-emploi.com/offre-d-emploi...	Agro/Agri (Métiers du Secteur Agroalimentaire). Manp...
9	February 29, 2016 at 12:02AM	Assistant(e) Ressources Humaines (H/F) – SAMSIC EM...	http://www.uestfrance-emploi.com/offre-d-emploi...	Ressources humaines. SAMSIC EMPLOI SAMSIC EMPLOI...
10	February 29, 2016 at 12:02AM	RESPONSABLE D'AGENCE (H/F) – Temporis Redon – G...	http://www.uestfrance-emploi.com/offre-d-emploi...	Direction générale – Direction centre de profits Mana...
11	February 29, 2016 at 08:41AM	Chauffeur de pelle / Manoeuvre (H/F) – CRIT – AGENC...	http://www.uestfrance-emploi.com/offre-d-emploi...	Terrassement. CRIT – AGENCE VITRE Nous recherchon...
12	February 29, 2016 at 08:42AM	Attaché de recrutement (H/F) – CRIT – AGENCE VITRE ...	http://www.uestfrance-emploi.com/offre-d-emploi...	Ressources humaines Assistant polyvalent. CRIT – A...
13	February 29, 2016 at 08:41AM	Opérateur de finition électronique (H/F) – CRIT – AGE...	http://www.uestfrance-emploi.com/offre-d-emploi...	Opérateurs de production/Conduite de ligne. CRIT – ...
14	February 29, 2016 at 08:42AM	Gestionnaire Supply Chain Transport (H/F) – CRIT – A...	http://www.uestfrance-emploi.com/offre-d-emploi...	Transport / Livraison / Déménagement. CRIT – AGEN...
15	February 29, 2016 at 08:41AM	ELECTROMECANICIEN (H/F) – CRIT – AGENCE VITRE – ...	http://www.uestfrance-emploi.com/offre-d-emploi...	Agro/Agri (Métiers du Secteur Agroalimentaire). CRIT ...

... et plus de 400 lignes de code !

```

53 messagep <- paste(message, collapse=" ")
54 messagesm <- gsub("[\t:\n\r]", " ", messagep)
55 messages <- gsub("http.*\\*", "", messagesm)
56 messageVS <- VectorSource(messages)
57 corpus <- Corpus(messageVS)
58 corpus <- tm_map(corpus, content_transformer(tolower))
59 corpus <- tm_map(corpus, stripWhitespace)
60 corpus <- tm_map(corpus, removeNumbers)
61 corpus <- tm_map(corpus, removePunctuation)
62 corpus <- tm_map(corpus, removeWords, stopwords("french"))
63 corpus <- tm_map(corpus, removeWords, c("www", "http", "https", "utm", "bit", "rennes", "brest", "quimper",
64 TDM <- TermDocumentMatrix(corpus)
65 matrixFrenchCourt <- as.matrix(TDM)
66 sortedmatrix <- sort(rowSums(matrixFrenchCourt),decreasing=TRUE)
67 frequenc <- data.frame(word = names(sortedmatrix),freq=sortedmatrix)
68 tete <- head(frequenc, 25)
69 tete$word <- factor(tete$word, levels = tete$word[order(tete$freq)])
70 barplot(height=tete$freq, names.arg=tete$word, xlab="Mots", ylab="Fréquence", col="#973232", main="Text-min"
71 library(wordcloud)
72 wordcloud(frequenc$word, frequenc$freq, max.words=300, rot.per=FALSE, colors= palette)
73
74 tete$word <- factor(tete$word, levels = tete$word[order(tete$freq)])
75
76 LBC5 <- as.data.frame(sort(table(x = LBC4$fin), decreasing = TRUE))
77 LBC5$names <- rownames(LBC5)
78 names(LBC5) <- c("freq", "names")
79 LBC6 <- head(LBC5, 50)
80 LBC6$names <- factor(LBC6$names, levels = LBC6$names[order(LBC6$freq)])
81
82 library(ggplot2)
83 ggplot(tete, aes(x=word, y=freq)) +
84   geom_bar(stat="identity") +
85   #coord_flip()

```

```

324 theme(axis.text=element_text(size=12),
325   axis.title=element_text(size=15),
326   title=element_text(size=18),
327   axis.title.x=element_text(margin=margin(20,0,0,0)),
328   axis.title.y=element_text(margin=margin(0,20,0,0)),
329   legend.text=element_text(size = 12),
330   plot.margin=margin(20,20,20,20))
331
332 write.csv2(plop, "Ouest-France-Emploi-mars.csv"),
333
334 mean(plop$`Nombre de postes`, na.rm = TRUE),
335
336 adresses <- as.vector(LBC[2501:3335, 3]),
337
338 ann3 <- data.frame(),
339 for(i in 1:length(adresses)) {
340   page <- adresses[i]
341   page <- readLines(page)
342   page <- as.data.frame(page)
343   debut <- grep(pattern = "<dt>Entreprise</dt>", x = page[,1])
344   fin <- grep("</dl>", page[,1])
345   page5 <- as.data.frame(page[debut:fin,:])
346   names(page5) <- 'url'
347   page5[,1] <- gsub("\t","", page5[,1])
348   page5[,1] <- gsub("<dt>","", page5[,1])
349   page5[,1] <- gsub("<dd>","", page5[,1])
350   page5[,1] <- gsub("</dd>","", page5[,1])
351   page5[,1] <- gsub("</dt>","", page5[,1])
352   page5[,1] <- gsub("</dl>","", page5[,1])
353   page5[,1] <- gsub("</a>","", page5[,1])
354   page5[,1] <- gsub("</li>","", page5[,1])
355   ann3 <- cbind.fill(ann3, page5, fill = NA)
356 }

```

```

137 debut <- grep(pattern = "<dt>Référence</dt>", x = page4[,1]),
138 fin <- grep("</dl>", page4[,1]),
139 page6 <- as.data.frame(page4[debut:fin,:],
140 page6[,1] <- gsub("\t","", page6[,1]),
141 page6[,1] <- gsub("<dt>","", page6[,1]),
142 page6[,1] <- gsub("<dd>","", page6[,1]),
143 page6[,1] <- gsub("</dd>","", page6[,1]),
144 page6[,1] <- gsub("</dt>","", page6[,1]),
145 page6[,1] <- gsub("</dl>","", page6[,1]),
146 page6[,1] <- gsub("</a>","", page6[,1]),
147 page6[,1] <- gsub("</li>","", page6[,1]),
148 page6[,1] <- gsub("</ul>","", page6[,1]),
149
150 for(i in 2:length(ann5)) {
151   debut <- grep(poke, ann5[,i])
152   if(is.null(debut)) {
153     ann5[5:10,i] <- NA
154   } else {
155     next
156   }
157 }
158 library(rwrf)
159
160 ann5 <- ann4
161 for(i in 1:length(ann5)) {
162   ann5[,i] <- gsub('<a data-categorie="detail-ann"', "", ann5[,i])
163   ann5[,i] <- gsub('data-action="clic"', "", ann5[,i])
164   libelle <- grep("^ data-libelle", ann5[,i])
165   ann5[libelle, i] <- ""
166   href <- grep("^ href=", ann5[,i])
167   ann5[href, i] <- ""
168   title <- grep("^ title", ann5[,i])
169   ann5[title, i] <- ""
170   href <- grep("en href=", ann5[,i])

```

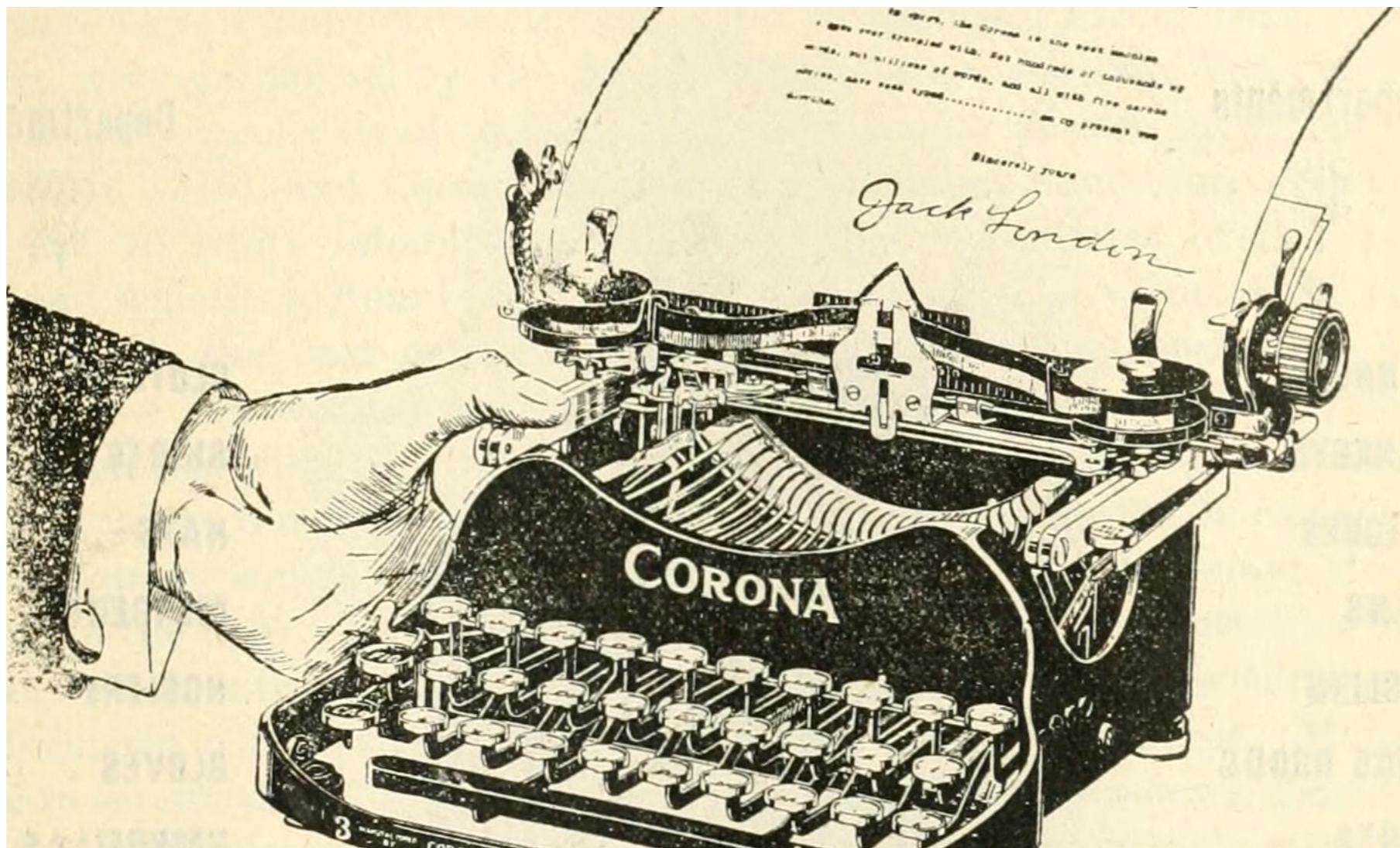
```

393 if(length(loc1) == 0){
394   loc1 <- NA
395 } else if(loc1 == "Contrat"){
396   loc1 <- NA
397 }
398 cont1 <- pouet[(cont+1)]
399 if(length(cont1) == 0){
400   cont1 <- NA
401 } else if(cont1 == "Expérience"){
402   cont1 <- NA
403 }
404 exp1 <- pouet[(exp+1)]
405 if(length(exp) == 0){
406   exp1 <- NA
407 } else {
408   exp1 <- pouet[(exp+1)]
409 }
410 et1 <- pouet[(et+1)]
411 if(length(et) == 0){
412   et1 <- NA
413 }
414 plop2 <- c(cent1, ref1, nbr1, loc1, cont1, exp1, et1)
415 names(plop2) <- c("Entreprise", "Référence", "Nombre de postes", "Localisation", "Contrat", "Expérience")
416 plop <- rbind(plop, plop2)
417 }
418 }
419
420 plop <- plop[-1,]
421 plop[,1] <- gsub(" - ", "-", plop[,1]),
422 plop[,1]
423 write.csv2(plop, "Ouest-France-Emploi-mars.csv")
424
425

```

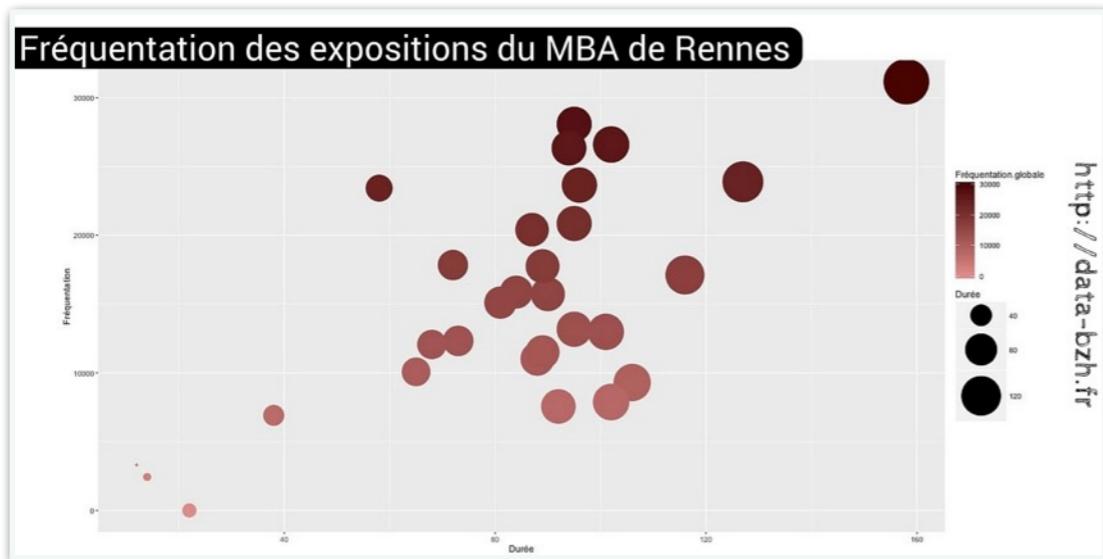
L'Open data

- L'ouverture des données : un enjeu démocratique fondamental
- Beaucoup de valeur ajoutée informative
- Demande des compétences d'analyse ...
- ... et un besoin de simplifier / vulgariser

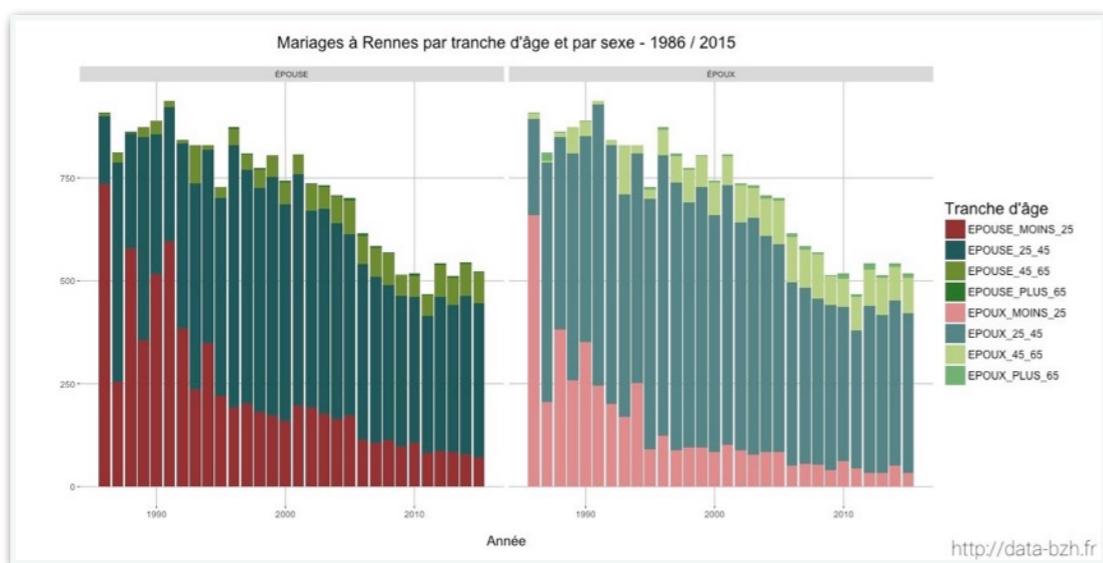


Data Bzh - Contenus

Des datavisualisations « sur le pouce »



- Accessibles
- À partager
- Tirées de données ouvertes



L'histoire de la Bretagne par l'Open Data

LES TRANSMUSICALES DE RENNES EN DATA

rencontres trans-musicales

JEUDI 14 JUIN

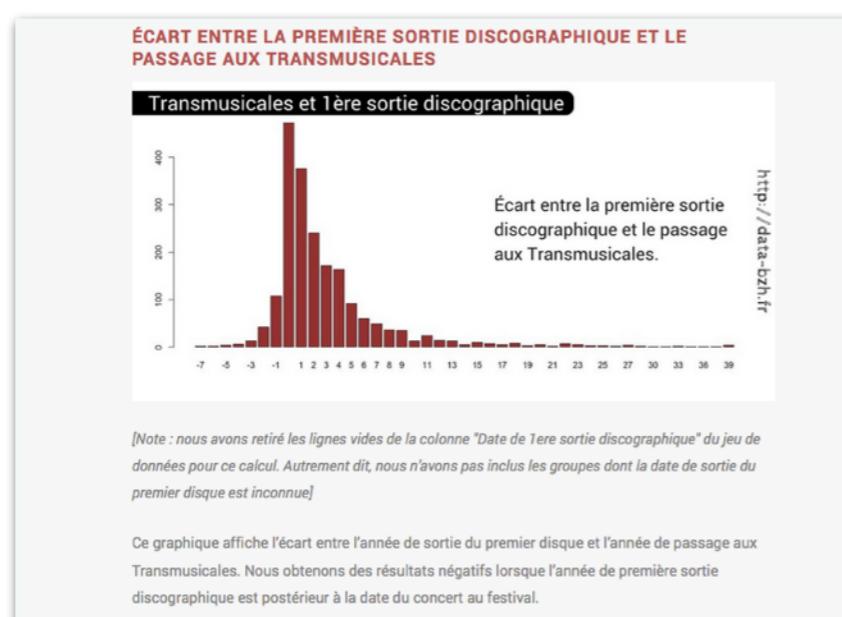
VEND. 15 JUIN

zero P.brunel groupe anche too doo cool duo oniris tvc 19 exces de zèle projectile Librement aux frai DE 19 A 24h Koukou maman

24 Avr Les Transmusicales de Rennes en data

Posted at 18:00h in Territoire by Colin

Exploration en 5 visualisations de la programmation des



- Des données ouvertes plus volumineuses
- Dossiers plus en profondeur, qui entrent dans les détails
- Présentant la Bretagne

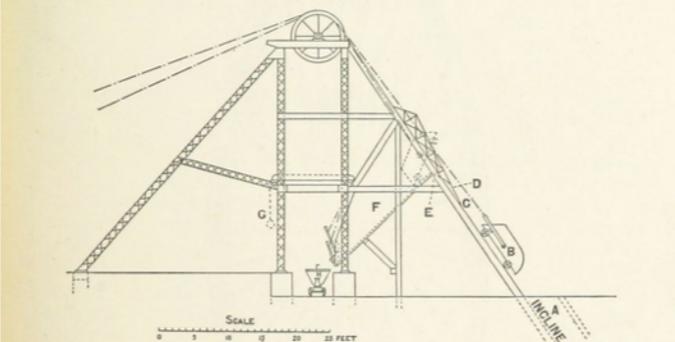
Web et social media mining



- Des données ouvertes plus techniques
- De l'information « souterraine »
- La Bretagne vue par le prisme des réseaux sociaux

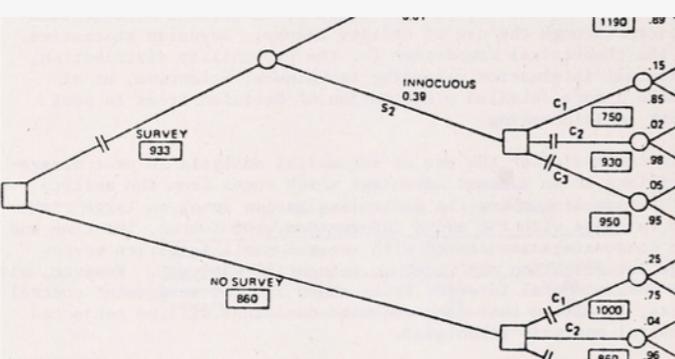
Blog : tutorials et billets de vulgarisation

PREMIERS PAS EN TEXT-MINING AVEC R – PARTIE 1



20 Mar Premiers pas en text-mining avec R – Partie 1
Posted at 18:00h in Blog by Colin
Annelé "Fouille de textes" dans la langue de Molière. le

UN ALGORITHME, QU'EST-CE QUE C'EST ?



14 Fév Un algorithme, qu'est-ce que c'est ?
Posted at 18:00h in Blog by Colin
Aujourd'hui, les algorithmes sont partout : dans nos fils

- Rendre le traitement de données accessible
- Apprendre
- Vulgariser

L'open data made in Data Bzh

Posted at 10:57h in Open Social Data by Colin

Jeu de données sur l'animation du compte Twitter de @DataBzh en mai 2016.

FICHE DÉTAILLÉE

Editeur : Data Bzh
Catégorie : Social Data
Date de mise à disposition : 04 juin 2016
Dernière mise à jour : 04 juin 2016
Description : Données téléchargeables depuis le backoffice du

FORMATS DISPONIBLES

CSV

Posted at 22:37h in Territoire by Colin

Données territoriales issues de Géo API pour l'Ille-et-Vilaine, collectées via rgeoapi.

FICHE DÉTAILLÉE

Editeur : Data Bzh
Catégorie : Territoire
Date de mise à disposition : 04 juin 2016
Dernière mise à jour : 04 juin 2016
Description : Données accessibles sur Géo API, collectées avec

FORMATS DISPONIBLES

CSV

(Cliquez sur le logo pour télécharger)

- Rendre les données accessibles
- Expérimenter
- Partager

Contact

- Colin FAY
- contact@data-bzh.fr
- [@DataBzh sur Twitter](https://twitter.com/DataBzh)