

Atelier P1-17

Journées APMEP de Marseille

20 octobre 2013

SFdS - Groupe Enseignement de la statistique

# Accès et usage de données statistiques réelles



Marthe-Aline JUTAND  
Philippe DUTARTE



# Le contexte

**Big data** : des moyens informatiques permettant une « récolte » abondante de données.

**Data mining** : des méthodes statistiques évoluant vers la recherche et l'exploitation de l'information pertinente.

**Open data** : une mise à disposition de données statistiques de plus en plus « ouverte ».

**Citoyenneté** : un impératif de formation fondamental pour le futur citoyen « éclairé ».

# Exploitation de vraies données statistiques dans les programmes du secondaire français

Programme de **seconde** générale et technologique (**2000**) : « l'objectif est de faire **réfléchir** les élèves sur la nature des données traitées, et de s'appuyer sur des représentations graphiques pour justifier des **choix** de résumé ».

« L'enseignant traitera des **données en nombre suffisant** pour que cela justifie une étude statistique ; il proposera des sujets d'étude et des simulations en fonction de l'intérêt des élèves, de l'**actualité** et de ses goûts ».

# Exploitation de vraies données statistiques dans les programmes du secondaire français

## Programme de **troisième** (2008) :

Pour les séries statistiques, l'étude des paramètres de position est poursuivie : médiane et quartiles. Une première approche de la dispersion est envisagée. L'éducation mathématique rejoint ici l'éducation du citoyen : prendre l'habitude de s'interroger sur la signification des nombres utilisés, sur l'information apportée par un résumé statistique. De même, c'est pour permettre au citoyen d'aborder l'incertitude et le hasard dans une perspective rationnelle que sont introduits les premiers éléments relatifs à la notion de probabilité.



# Exploitation de vraies données statistiques dans les programmes du secondaire français

## Programme de **troisième** (2008)

Le travail est conduit aussi souvent que possible en liaison avec les autres disciplines dans des situations où les données sont exploitables par les élèves. L'utilisation d'un tableur permet d'avoir accès à des situations plus riches que celles qui peuvent être traitées « à la main ».



# Exploitation de vraies données statistiques dans les programmes du secondaire français

Programme de **seconde** générale et technologique (2009) :

« L'objectif est de faire **réfléchir** les élèves sur des **données réelles**, riches et variées (issues, par exemple, d'un **fichier** mis à disposition par l'INSEE), synthétiser l'information et proposer des représentations pertinentes ».





# Exploitation de vraies données statistiques dans les programmes du secondaire français

Programme de première scientifique (2010) :  
« L'étude et la comparaison de séries statistiques menées en classe de seconde se poursuivent avec la mise en place de nouveaux outils dans l'analyse de données. L'objectif est de faire **réfléchir** les élèves sur des **données réelles**, riches et variées (issues, par exemple, de fichiers mis à disposition par l'Insee). »

# Exemples d'apprentissages orientés vers les données dans les curricula du secondaire à l'étranger

<http://publications-sfds.math.cnrs.fr/index.php/StatEns>

## Vol. 4, No 1 (2013): Le curriculum statistique dans le secondaire – Comparaisons internationales

NUMERO SPECIAL dirigé par Carmen BATANERO, Jeanne FINE et Jean-Pierre RAOULT

### Sommaire

#### Editorial

Le curriculum statistique dans le secondaire – Comparaisons internationales [PDF](#)  
*Carmen BATANERO, Jeanne FINE, Jean-Pierre RAOULT, Catherine VERMANDELE* 1-4

---

#### Recherches et perspectives

[Les idées statistiques fondamentales dans le curriculum scolaire](#) [PDF](#)  
*Gail BURRILL, Rolf BIEHLER* 5-24

[Le rapport GAISE \(U.S.\) – Cadre d'un curriculum statistique de la maternelle à la Terminale](#) [PDF](#)  
*Jeanne FINE* 25-54

[La statistique dans l'enseignement secondaire en France](#) [PDF](#)  
*Jean-Pierre RAOULT* 55-69

[Le curriculum statistique en Belgique francophone](#) [PDF](#)  
*Catherine VERMANDELE* 71-88

[La statistique dans le curriculum en Espagne](#) [PDF](#)  
*Carmen BATANERO, Juan Jesús ORTIZ, Rafael ROA, Luis SERRANO* 89-106

[L'enseignement de statistique et probabilités au Canada et en Italie](#) [PDF](#)  
*Linda GATTUSO, Sylvain VERMETTE* 107-129

[Analyse du curriculum statistique du collège au Mexique d'un point de vue de la formation à la statistique](#) [PDF](#)  
*Ernesto SÁNCHEZ, Verónica HOYOS* 131-142



# Exemples d'apprentissages orientés vers les données dans les curricula du secondaire à l'étranger

U.S.A. : le rapport GAISE (*Guideline for Assessment and Instruction in Statistics Education*) distingue quatre étapes dans la démarche statistique :

1. Formuler une question.
2. Collecter des données.
3. Analyser, explorer les données.
4. Interpréter les résultats.

# Exemples d'apprentissages orientés vers les données dans les curricula du secondaire à l'étranger

## *Espagne*

La statistique est la science des données. Dans les documents curriculaires espagnols, on reconnaît cette importance, depuis l'éducation **primaire** pour laquelle on suggère que les élèves recueillent et analysent leurs propres données en utilisant aussi bien des enquêtes que l'observation et la mesure. Dans les niveaux plus avancés on suggère un **recueil de données sur Internet ou dans les médias** et on met l'accent sur une **attitude critique** face aux données et aux sources des données.

# Des compétences à développer grâce à l'exploitation de vraies données

Des compétences transversales (du registre des attitudes), notamment favorisées par l'usage des outils informatiques (pour les trois premières) :

- **autonomie** ;
- **prise d'initiative** ;
- **créativité** ;
- **rigueur**.

L'usage de vraies données favorise le sens des notions et augmente la motivation des élèves.

# Des compétences à développer grâce à l'exploitation de vraies données

Compétences mathématiques au **collège** :

- **Rechercher, extraire et organiser l'information utile.**
- Mesurer, **calculer**, appliquer des consignes.
- **Modéliser**, conjecturer, **raisonner** et démontrer.
- **Argumenter et présenter les résultats dans un langage adapté.**

# Des compétences à développer grâce à l'exploitation de vraies données

Compétences mathématiques au *lycée* :

- 1. **Chercher** (extraire et traiter l'information, observer, expérimenter...).
- 2. **Modéliser** (« traduire » en langage mathématique, **simuler**, valider ou non un modèle...).
- 3. **Représenter** (choisir un **cadre adapté** – numérique, graphique, algébrique, géométrique..., changer de cadre...).



# Des compétences à développer grâce à l'exploitation de vraies données

Compétences mathématiques au *lycée* :

- 4. **Calculer** (à la main ou **instrumenté**, mettre en œuvre un algorithme, **contrôler** les calculs...).
- 5. **Raisonner** (utiliser la logique, **différents types de raisonnements**, démontrer...).
- 6. **Communiquer** (convertir langage naturel et langage formel, **critiquer** une démarche, **s'exprimer avec clarté** à l'oral et à l'écrit...).

# Sources d'informations possibles



- Documents ressources
- Ouvrages scolaires
  - Avec pour la plupart de nombreux exercices et abondant des domaines variés
  - De plus ou moins bonnes qualités
  - Parfois des erreurs d'interprétations concernant IF et IC
  - ...

# Sources d'informations possibles – pour aller plus loin

– Bibliographie proposées sur site le portail des IREM

- Manuels universitaires et monographies sur la Statistique et les Probabilités

<http://www.univ-irem.fr/spip.php?article507>

Cependant attention car la plupart du temps pas de normalisation de vocabulaire

+ Intervalle de fluctuation : notion pas toujours présente

+ toujours très axés sur la notion de tests statistiques

+ plus souvent directement théorème centrale limite et non Moivre Laplace



# Sources d'informations possibles – pour aller plus loin

- Via le site de la SFdS
  - Polycopiés de cours



- [http://www.sfds.asso.fr/190-Polys\\_denseignement](http://www.sfds.asso.fr/190-Polys_denseignement)
  - Attention des cours souvent très complets, voir trop
  - Des pages qui ne sont plus actives
  - dépassant largement le programme de terminale,
  - développant peu les informations IF et IC
  - Pas de normes dans les présentations
  - Cependant des exemples de données
- Pour fouiller !!



# Plate-forme de cours et exos

- Lyon « enseignement de la statistique en biologie
  - » <http://pbil.univ-lyon1.fr/R/>
  - » Exemple de cours : « des électeurs, des boules, des cercles, des étudiants satisfaits : AB Dufour, JR Lobry et D Chessel
  - » [http://pbil.univ-lyon1.fr/R\\_svn/pdf/bem5.pdf](http://pbil.univ-lyon1.fr/R_svn/pdf/bem5.pdf)
- Site ISPED, un enseignement statistique descriptive
  - » [http://campus.isped.u-bordeaux2.fr/PLEIADE/PV3/ASPX/PV2\\_VPS.aspx?16971](http://campus.isped.u-bordeaux2.fr/PLEIADE/PV3/ASPX/PV2_VPS.aspx?16971)
  - » Un fichier + un texte de questions, ne couvrant pas exactement le programme de terminale mais facile à convertir
- Toulouse : INSA et Institut de mathématique
  - » <http://wikistat.fr/>





Melange de lois normales

Recherche Google



Rechercher

Allez-y !

AdCh...

**Accueil**

**Modèles de la variabilité**

- Page d'entrée
- Page de liens
- English Section
- Espace invités
- Maintenance

Lois de probabilités classiques : approche pratique.

[ <http://pbll.univ-lyon1.fr/R/html/tar2.html> Version : 18.09.12 ]

**Cours**

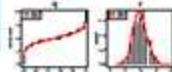
- Introductions
- Biologie et modélisation (L)
- Tests d'hypothèse
- Analyse des données
- Fiches de stage
- Probabilité & Statistique
- Écologie et Statistique
- Évolution moléculaire
- Modélisation
- Divers



Ref : tdr21 Taille : 790 ko Version : 20.09.12 (77 jours)

**Lois de Probabilités**

Aucun sommaire



**Annales**

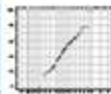
- Biostatistiques (L)
- Biologie et modélisation (L)
- Analyse des données (M)
- aMIG (M)
- Algèbre & Statistique
- Statistiques & logiciels
- Autres disciplines et \Rlogo



Ref : tdr22 Taille : 428 ko Version : 20.09.12 (77 jours)

**Graphes quantiles-quantiles**

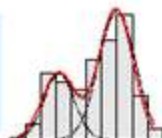
Histogrammes, fonctions de répartition, droite de Henri (ou de Henry) et `texttt{qqnorm()}`.  
Mélanges de distributions



Ref : tdr221 Taille : 209 ko Version : 20.09.12 (77 jours)

**Estimer un mélange de lois normales**

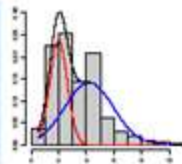
Maximum de vraisemblance, solution numérique, simulations et estimations.



Ref : tdr222 Taille : 378 ko Version : 20.09.12 (77 jours)

**Mélange de lois normales et tailles de génomes bactériens**

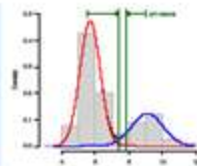
Étude de la taille de 279 génomes de bactéries exprimée en kilobases (données de 2002). Mise à jour avec des données plus récentes à partir de GOLD (Genomes Online Database).



Ref : tdr223 Taille : 266 ko Version : 20.09.12 (77 jours)

**Mélanges et point isoélectrique des protéines**

Étude de la distribution du point isoélectrique dans un protéome. Une simulation pour invalider une interprétation trop jolie pour être vraie. D'après un article de G.F. Weiller, G. Caraux et N. Sylvester (2004) *Proteomics*, 4:943-949.



**Fiches de TD**

- Le logiciel R
- Biologie et modélisation (L)
- Statistique descriptive
- La variabilité
- Les tests d'hypothèses
- Courbes de réponse
- Analyse multivariée (1)
- Analyse multivariée (2)
- Structures spatiales
- Sciences Financières et Assurances



Ref : tdr23 Taille : 873 ko Version : 20.09.12 (77 jours)



# Plate-forme de cours et exos attention non mis à jour

- St@ternet – Toulouse
  - <http://www2.toulouse.iufm.fr/mathematiques/stat2/presentation/indexp.htm> (*site non mis à jour IUFM Midi-Pyrénées*)
  - *Support de cours intéressant, quelques données mais exemple peu approprié*
- SMEL, "*Statistique Médicale En Ligne*", cours de probabilités, statistiques descriptive et inférentielle, simulations  
[http://mistis.inrialpes.fr/software/SMEL/articles/cadre\\_articles.html](http://mistis.inrialpes.fr/software/SMEL/articles/cadre_articles.html)

- [Courbes de croissance](#) (M. Perreau-Guimaraes et B. Ycart)
- [Entre De Moivre et Laplace](#) (B. Ycart)
- [Abrogeons la loi des moyennes](#) (I. Stewart)
- [Les coïncidences sont parfois plus probables qu'il n'y paraît](#) (I. Stewart)
- [La taille de l'empereur de Chine](#) (C. Robert)
- [Le vieux fidèle](#) (C. Robert)
- [Délit d'initié](#) (C. Robert)
- [Mariages et produit intérieur brut](#) (C. Robert)
- [Les dix suspects](#) (C. Robert)
- [Le premier chiffre à gauche](#) (M. Roman et C. Robert)
- [Une série à la roulette](#) (C. Robert)
- [Le rapido](#) (B. Ycart)
- [Un problème de sondage simplifié](#) (C. Robert)
- [Estimation du risque de trisomie 21](#) (T. Faraut *et al.*)
- [Embolie pulmonaire](#) (Ch. Pison et J.L. Bosson)
- [Analyse automatique d'audiogrammes](#) (A. Job *et al.*)
- [Maladie thromboembolique veineuse et cancer](#) (J.L. Bosson *et al.*)
- [Activité chirurgicale du cancer du sein](#) (H.Q. Cong *et al.*)
- [Quantification échographique des thromboses veineuses](#) (J.L. Bosson *et al.*)
- [Hémiparésie ataxique par infarctus lacunaire](#) (G. Besson *et al.*)
- [Pathologies pulmonaires et menuiserie industrielle](#) (A. Marcuccilli *et al.*)
- [Vieillesse dans la profession des remontées mécaniques](#) (R de Gaudemar *et al.*)
- [Programmes d'échanges de seringues en France](#) (J. Emmanuelli *et al.*)
- [Les infections à méningocoques en France en 1995](#) (F. Fourquet et B. Hubert)
- [Le sida en France en 1995](#) (S. Tchakamian *et al.*)
- [Le tétanos en France en 1995](#) (I. Rebière)

# Plate-forme de cours et exos

- Statnet
  - <http://www.agro-montpellier.fr/cnam-lr/statnet/index.htm>
    - Des exemple avec certains fichiers
    - Des animations avec exercices
    - Mais fonction de répartition et théorème central limite

# Des sites de commentaires, mais aussi d'exemples

- Statistix

- Centre de ressources, lieu de partage et de mutualisation pour l'enseignement de la statistique.

- Pour qui ? Les enseignants des écoles, des collèges et des lycées de toutes les disciplines.

- <http://statistix.fr/>

- Commentaires sur document ressource première

- <http://statistix.fr/spip.php?article89>

- Journée mondiale de la statistique

- <http://statistix.fr/spip.php?article86>

- Exemple pour Loi normale en terminale S/ES/L

- <http://statistix.fr/spip.php?article93>



# Usage de données réelles



- Regroupées ou non
- Selon un ou plusieurs critères
- Pour utiliser un logiciel ou des calculs simples avec calculatrice
- Pour manipuler uniquement ou pour répondre à une question pas uniquement dans le cadre du cours de mathématiques?

# Exemples de sources

- Le réseau Quételet
    - coordonne les activités d'archivage, de documentation et de diffusion des données en sciences humaines et sociales du Centre Maurice Halbwachs (ADISP), du Centre de données socio-politiques et du service des enquêtes de l'Ined. L'action du réseau est organisée par Roxane Silberman, secrétaire générale du CCDSHS
- [http://www.reseau-quetelet.cnrs.fr/spip/rubrique.php3?id\\_rubrique=3](http://www.reseau-quetelet.cnrs.fr/spip/rubrique.php3?id_rubrique=3)
- Présentation de la source des données + l'outil de recueil
  - Nécessité de demander les fichiers

- Le CDSP
- Enquêtes françaises
- Enquêtes internationales
- beQuali
- Enquêtes en ligne
- **Les élections en France**
- VizLab
- Les élections dans les 27
- Dépôt de données
- Projets en cours
- Equipement DIME-SHS

## Les élections en France

Le CDSP diffuse les **résultats des élections politiques en France depuis 1965**.

### Élections en ligne

- **Européennes** (depuis 1979)
- **Législatives** (depuis 1997)
- **Présidentielles** (depuis 1965)
- **Referendum** (depuis 1992)
- **Régionales** (depuis 1998)

Et par le service de consultation interactive des résultats électoraux :

- **Municipales** (depuis 2001 pour les communes de plus de 30 000 habitants)

>> Consulter la [liste des résultats disponibles](#) au format Excel et [commander les fichiers de données sur le portail Quetelet](#).

### Consultation interactive des résultats électoraux



>> **Résultats des élections présidentielles, législatives, régionales et européennes par circonscription, département et région**

>> **Résultats des élections présidentielles et municipales par commune** (en maintenance)

### Actualités du CDSP

**Recrutements - Concours ITRF 2013**  
Le CDSP recrute un ingénieur de recherche BAP D et un ingénieur d'étude BAP F par voie de (...)

**Le CDSP a déménagé**  
Les locaux du Centre de (...)

**Élections présidentielle et législatives 2012**  
Tous les résultats de l'élection présidentielle et des élections législatives sont disponibles par circonscription, département et région sur (...)

Newsletter  
e-mail  ✓

Rechercher  
 ✓

# Exemples de sources : INSEE

- Données issues de l'INSEE : <http://www.insee.fr/fr/>
- L'Institut national de la statistique et des études économiques collecte, produit, analyse et diffuse des informations sur l'économie et la société françaises.
- Six missions pour y contribuer
  - Collecter et produire
  - Analyser
  - Diffuser
  - Coordonner
  - Enseigner et développer la recherche
  - Contribuer à l'édification d'un espace statistique international



Institut national de la statistique  
et des études économiques

Mesurer pour comprendre

Mobile | Actualités | Agendas | Contactez-nous | Aide | FAQ | Liens | English | Home page

Chercher sur le site



Accueil

Thèmes

Bases de données

Publications et services

Régions

Définitions et méthodes

Accès par public

L'Insee et la statistique publique

## Bases de données

- Alisse, statistiques d'entreprises
- Indices et séries chronologiques
- Les résultats des recensements de la population
- Données locales
- Les fichiers détail
- Répertoire SIRENE

Accueil Bases de données

## Bases de données

L'Insee propose plusieurs bases de données statistiques sur des domaines variés.

- Statistiques structurelles d'entreprises** : toutes les statistiques d'entreprises issues d'Esane et de la statistique publique.
- Indices et séries chronologiques** : tous les indices et indicateurs de l'Insee, sur longue période, documentés et accompagnés de graphiques.
- Les résultats des recensements de la population** : les résultats détaillés, les cartes et toutes les informations utiles sur les recensements.
- Données locales** : un éventail de données localisées du quartier à la région.
- Les fichiers détail** : les fichiers détail proposés sur ce site sont des bases de données électroniques comportant les enregistrements individuels, anonymisés, afférents à des enquêtes ou opérations statistiques réalisées par l'Insee.
- Répertoire Sirene** : présentation des services et des usages du répertoire Sirene.

### Pour en savoir plus

Les bases de données vous permettent d'accéder à des données plus complètes, plus fines que celles disponibles dans les chiffres clés ou les données détaillées des accès par thèmes ou régions.

Elles sont souvent présentées à des niveaux géographiques plus localisés : zones d'emploi, communes ou infra-communal.





Accueil

Thèmes

Bases de données

Publications et services

Régions

Définitions et méthodes

Accès par public

L'Insee et la statistique publique

Thèmes

- ▣ Agriculture
- ▣ Commerce
- ▣ Comptes nationaux - Finances publiques
- ▣ Conditions de vie - Société
- ▣ Conjoncture
- ▣ Économie
- > Enseignement - Éducation
  - Élèves, établissements et enseignants
  - > Diplômes - Formation
  - Dépenses d'éducation
- ▣ Entreprises
- ▣ Industrie - IAA - Construction
- ▣ Population
- ▣ Revenus - Salaires
- ▣ Santé
- ▣ Services - Tourisme - Transports
- ▣ Territoire
- ▣ Travail - Emploi

Accueil Thèmes Enseignemen... Diplômes - ... Commune de ... Diplômes - ... FOR1 - Population de 2 ans ou plus par scolarisation, sexe, âge et lieu d'études (Âge détaillé)

FOR1 - Population de 2 ans ou plus par scolarisation, sexe, âge et lieu d'études (Âge détaillé)

Commune de Bordeaux (33063)

■ Téléchargement / Niveaux d'affichage / Documentation / Liens / Changer de territoire

Imprimer la page

Sexe : Ensemble

	Commune de résidence	Autre commune du département de résidence	Autre département de la région de résidence	Autre région en France métropolitaine	Autre (Dom, Com, étranger)	Ensemble
2 ans	319	2	0	0	0	2 32
3 ans	1 840	78	0	1	0	2 25
4 ans	2 008	96	0	3	0	2 18
5 ans	1 975	126	0	5	0	2 18
6 ans	1 956	127	0	4	0	2 11
7 ans	1 846	106	0	2	0	1 98
8 ans	1 792	132	0	1	0	1 93
9 ans	1 962	143	3	3	0	2 12
10 ans	1 978	142	0	0	0	2 13
11 ans	1 931	136	4	1	0	2 08
12 ans	1 994	114	9	6	0	2 13

# Portail de la statistique publique

- [www.statistique-publique.fr/](http://www.statistique-publique.fr/)
- Éditeur l'INSEE
- créé à la demande du Conseil national de l'Information statistique ([Cnis](#))
  - permettre au public d'accéder plus facilement à l'ensemble des informations statistiques produites par la statistique publique.
- Le Portail procure un accès unifié à ces éléments statistiques parfois dispersés sur plusieurs sites. Il est alimenté par les sites internet de la statistique publique (Insee, Services statistiques ministériels et organismes publics conventionnés).
- liste et adresses des sites accessibles dans la rubrique [Sites nationaux](#)



**Recherche**

» Recherche libre :  **ok**

» Recherche par mots-clés

- Recherche par domaine**
- o Agriculture - Pêche- Sylviculture
  - o Construction - Logement
  - o Culture-Sport
  - o Délinquance - Justice
  - o Économie générale
  - o Éducation-Formation-Recherche
  - o Emploi et Travail
  - o Énergie
  - o Entreprise
  - o Environnement
  - o Finances publiques
  - o Mode de vie des ménages
  - o Population et démographie
  - o Revenu(s)
  - o Santé et social
  - o Territoire(s)
  - o Tourisme
  - o Transport



## LES ACTUALITÉS STATISTIQUES

### Conjoncture - Tableau de bord, analyse mensuelle et indicateurs

10/10/13 Hausse des prix des logements au 2e trimestre 2013 (+0,7 %)  
 10/10/13 En août 2013, la production manufacturière progresse légèrement (+0,3 %)

[» voir toutes les actualités](#)

## DOSSIERS

### Revenus, salaires et patrimoine

[» Voir les dossiers](#)

## ÉVÉNEMENTS

### - Ensaï, Bruz - Journées de statistique : travaux académiques ou appliqués dans des domaines en lien avec le secteur de l'assurance

[» Voir tous les événements](#)

# Exemples de sources : Ined

- Données Ined : <http://www.ined.fr/>
- « L'INED est le plus important institut de recherche démographique au monde. Il étudie les populations de la France et des pays étrangers avec les outils du démographe et les apports des autres disciplines : histoire, géographie, sociologie, anthropologie, économie, biologie, épidémiologie. Il conduit des recherches sur les phénomènes universels que sont les naissances, les unions, les migrations, la mort, etc. Ses chercheurs travaillent dans des domaines aussi divers que la contraception et l'avortement, les migrations, les populations en marge, l'allongement de la durée de vie. »



Accueil > Ressources et documentation

### Mortalité

#### Les causes de décès

Depuis 1968, l'INSERM est responsable de la publication de la statistique des causes des décès (SC8). Pour les années 1978 à 1997, cette statistique est connue sous le nom de Service d'information sur les causes de décès (SC8). Ce service présente des données de mortalité par sexe, par groupe décennal d'âges et lieu de décès (ville). Il présente aussi la manière d'

- Portail de la bibliothèque
- Revue de presse documentaire
- Publications de l'ined >>
- Sélection de liens >>
- Enquêtes de l'ined >>
- Données détaillées >>
- Pages thématiques >>
- Rendez-vous de l'ined
- Focus sur >>
- Lire sur la toile >>

de la publication de la NSEE. Pour les années 1978 à 1997, cette statistique est connue sous le nom de Service d'information sur les causes de décès (SC8). Ce service présente des données de mortalité par sexe, par groupe décennal d'âges et lieu de décès (ville). Il présente aussi la manière d'

Pour travailler sur des séries plus longues, il faut résoudre les problèmes posés par les révisions de la nomenclature et les variations des pratiques de déclaration et d'enregistrement. Deux équipes de chercheurs de l'ined travaillent à l'élaboration de séries temporelles de décès par causes à définition constante, avec des approches distinctes pour répondre à des besoins différents.

**Données détaillées**

- Causes de décès de 1925 à 1999
- Causes de décès de 1950 à 1999
- Mortalité par département au XIXè s.



# Exemples de sources : Institutions

<http://www.data.gouv.fr/>

Circulaire du 26 mai 2011 relative à la création du portail unique des informations publiques de l'Etat « data.gouv.fr » par la mission « Etalab » et l'application des dispositions régissant le droit de réutilisation des informations publiques





PLATEFORME FRANÇAISE D'OUVERTURE DES DONNÉES PUBLIQUES (OPEN DATA)

Rechercher une donnée RECHERCHER

RECHERCHE AVANCÉE



MÉTHODE ET ORGANISATION POUR FACILITER LE RECENSEMENT DES DONNÉES PUBLIQUES

SUGGESTION DE RECHERCHE LES PLUS RECHERCHÉS

- Principales infractions en vigueur
- Résultats des élections européennes
- Statistiques du commerce extérieur
- Financement et dépenses de la sécurité sociale
- Statistiques du commerce extérieur
- Liste des associations subventionnées
- Effort financier de l'Etat en faveur des PME
- Pôles de compétitivité
- Avis de rappel de produits



REJOIGNEZ LA COMMUNAUTÉ

REPÈRES

353 226 jeux de données publiques et plus sur data.gouv.fr

ETALAB

etalab.gouv.fr Suivez notre actualité sur le blog

# Licence Ouverte / Open Licence

- Cette licence, élaborée en concertation avec l'ensemble des acteurs concernés, facilite et encourage la réutilisation des données publiques mises à disposition gratuitement sur cette plateforme.
- présente les caractéristiques suivantes :
  - grande liberté de réutilisation des informations :
    - ouverte, libre et gratuite, qui apporte la sécurité juridique nécessaire aux producteurs et aux réutilisateurs des données publiques ;
    - promeut la réutilisation la plus large en autorisant la reproduction, la redistribution, l'adaptation et l'exploitation commerciale des données ;
    - s'inscrit dans un contexte international en étant compatible avec les standards des licences Open Data développées à l'étranger et notamment celles du gouvernement britannique (Open Government Licence) ainsi que les autres standards internationaux (ODC-BY, CC-BY 2.0).



LICENCE OUVERTE  
OPEN LICENCE

# Licence Ouverte / Open Licence

- présente les caractéristiques suivantes :
  - Une exigence forte de transparence de la donnée et de qualité des sources en rendant obligatoire la mention de la paternité.
  - Une opportunité de mutualisation pour les autres données publiques en mettant en place un standard réutilisable par les collectivités territoriales qui souhaiteraient se lancer dans l'ouverture des données publiques.



LICENCE OUVERTE  
OPEN LICENCE

# Licence Ouverte / Open Licence

- présente les caractéristiques suivantes :
  - Une exigence forte de transparence de la donnée et de qualité des sources en rendant obligatoire la mention de la paternité.
  - Une opportunité de mutualisation pour les autres données publiques en mettant en place un standard réutilisable par les collectivités territoriales qui souhaiteraient se lancer dans l'ouverture des données publiques.



LICENCE OUVERTE  
OPEN LICENCE



PLATEFORME FRANÇAISE D'OUVERTURE DES DONNÉES PUBLIQUES (OPEN DATA)

Search bar with magnifying glass icon, RECHERCHER button, and RECHERCHE AVANCÉE button

INDICATEURS DE VALEUR AJOUTÉE DES LYCÉES D'ENSEIGNEMENT GÉNÉRAL ET TECHNOLOGIQUE - ACTUALISATION 2011

Donnée D

Date publiée le : 06-04-2012

Auteur : Ministère de l'Education nationale, de la Jeunesse et de la Vie associative

Description des données statistiques : les indicateurs de valeur ajoutée des lycées permettent d'évaluer l'action propre de chaque lycée. Ils sont établis à partir des résultats des élèves au baccalauréat et de leur parcours scolaire dans l'établissement. Les lycées d'enseignement général et technologique, publics et privés sous contrat, sont concernés. Il ne s'agit aucunement d'un classement mais d'un regard croisé sur les indicateurs et les « valeurs ajoutées » correspondantes.

Période : du 01/09/2010 au 31/08/2011

Date de production des données : 31/03/2012

Fréquence de mise à jour : annuelle

Licence : Licence Ouverte/Open Licence

Accéder aux données

FORMAT

CSV Télécharger (498.1 Ko)

Checksum : bb5075d771426b9611671793f9747769

Licence

LICENCE OUVERTE / OPEN LICENCE




LICENCE OUVERTE OPEN LICENCE



# **OPEN DATA** hauts-de-seine

La plateforme des données publiques du département

**Connexion** **S'inscrire**

-  **LES DONNÉES** **COMPRENDRE** **PARTICIPER** **DÉVELOPPER**



**Tweets**  **Follow @OpenData92**

 **Fondapol** @Fondapol 6 Oct

l'#opendata permet au Danemark d'économiser \$45M par an, mais de générer 85m\$ de revenus dans le privé.  
[geospatial.blogsp.com/geospatial/201...](https://geospatial.blogsp.com/geospatial/201...)


Retweeted by OpenDataHautsdeSeine

Expand



# **OPEN DATA** hauts-de-seine

La plateforme des données publiques du département

Recherche dans les jeux de données 

**Connexion** **S'inscrire**

-  **LES DONNÉES**
- COMPRENDRE
- PARTICIPER
- DÉVELOPPER

► Passer à la thématique suivante

## Les données

### Toutes les données

- Patrimoine
- Culture
- Politique de la ville
- Subventions
- Habitat

### Les plus téléchargées

- Données budgétaires
- Solidarité
- Environnement
- Aménagement du territoire

### Licence et utilisation

La "Licence Ouverte / Open Licence"

176 km<sup>2</sup>

8 805 habitants au km<sup>2</sup>



OpenDataHautsdeSeine  
Expand

# Exemples de sources : INSERM

- Données françaises : Présentation de l'INSERM et de son site WWW.INSERM.fr
  - Réseau sentinelles
    - <http://websenti.u707.jussieu.fr/sentiweb/?page=presentation>
    - système d'information basé sur un réseau de médecins généralistes en France métropolitaine
    - **surveillance continue de 9 indicateurs de santé**



### Tableaux de données

#### Génération automatique de tableaux de données

- Maladie** Syndromes Grippaux [1984]
- Semaine de début** Acte suicidaire [1999]  
Crises d'asthme [2002 - 2011]  
Dianhée aiguë [1990]
- Semaine de fin** Hépatite A aiguë [2000 - 2008]  
Hépatite B aiguë [2000 - 2008]
- Estimation :** Hépatite C [2000 - 2008]  
Hospitalisation [1997 - 2011]  
Maladie de Lyme [2009]  
Oreillons [1985]  
Rougeole [1984 - 2008]  
Sérologies VHC prescrites [2000 - 2008]
- Type de sortie** **Syndromes Grippaux [1984]**
- Code géographique (niveau région)** Test VIH [1987 - 2002]  
Urétrite masculine [1984]  
Varicelle [1990]  
Zona [2004]

- Territoire**
- France Métropolitaine
  - Région (niveaux NUTS2)
- Alsace
  - Aquitaine
  - Auvergne
  - Basse-Normandie
  - Bourgogne
  - Bretagne
  - Centre
  - Champagne-Ardenne
  - Corse
  - Franche-Comté
  - Haute-Normandie

# Exemples de sources : INSERM

- CEPIDC Le CépiDc, Centre d'épidémiologie sur les causes médicales de décès,
  - missions essentielles :
    - production annuelle de la statistique des causes médicales de décès en France (540 000 décès par an),
    - diffusion des données et les études et recherches sur les causes médicales de décès.
    - <http://www.cepidc.inserm.fr/>
      - (attention les données avant 2010)



Rechercher



Partager

Autres >>

Get more Add-ons

Sites suggérés

données détaillées

## Interrogation des données détaillées

### 1 Type d'indicateur :

Effectifs de décès

Taux bruts de décès

### 2 Sélection Géographique :

Niveau :

Zone :

### 3 Année(s) de décès :

Depuis :

Jusqu'à :

### 4 Détail de la cause de décès :

Liste abrégée

Liste détaillée CIM (4 caractères)

### 5 Sélection de la cause de décès :

Chapitre :

[Retour au menu](#)

[Lancer la demande](#)

**CépiDc** Interrogation des données sur les causes de décès de 1979 à 2010

**Effectifs de décès**

Année : 2006  
 Chapitre : Tous détaillés  
 Zone : France métropolitaine

[Liste des codes CIM](#)

[Retour au menu](#) [Export CSV](#)

Code CIM	Libellé	Sexe	Total	<1	1-4	5-14	15-24	25-34	35-44	45-64
15-A19, B90	Tuberculose	M	390	0	0	0	1	3	9	
		F	336	0	0	0	1	2	2	
		T	726	0	0	0	2	5	11	
39	Infection à méningocoques	M	14	2	3	3	3	0	1	
		F	17	2	3	2	5	0	0	
		T	31	4	6	5	8	0	1	
20 -B24	SIDA et maladies à VIH	M	608	0	0	0	4	27	229	1
		F	182	0	0	1	5	13	82	
		T	790	0	0	1	9	40	311	2
15 -B19	Hépatites virales	M	394	0	0	0	0	7	61	
		F	328	0	0	0	0	2	14	



# Ressources disponibles : DREES

<http://www.drees.sante.gouv.fr/>



La DREES (Direction de la recherche, des études, de l'évaluation et des statistiques, du ministère des affaires sociales et de la santé) propose des extractions en format Excel des données utilisées pour ses études. On choisit un thème dans la colonne de gauche, puis on a en général deux « onglets » : Articles/Documents et données. L'onglet « Documents et données » donne généralement accès à des fichiers Excel.

# Ressources disponibles : Eco-Santé

<http://www.ecosante.fr/>

Le site « Eco-Santé » permet d'obtenir des données en table, avec gestion directe sur le site ou exportation.

**Les bases Eco-Santé en ligne**

Eco-Santé c'est plus de 5 millions de données en accès gratuit et immédiat à travers un environnement convivial.  
Eco-Santé c'est aussi des données actualisées, séries statistiques actualisées sur des périodes longues, souvent depuis 1990.  
Eco-Santé c'est aussi des données regroupées par grands thèmes de l'économie de la santé : Dépenses de santé, État de santé, Professions de santé, Hospital, Secteur pharmaceutique, Pollution visuelle, Démographie, Indicateurs économiques.  
Eco-Santé c'est aussi des données présentées à plusieurs niveaux géographiques : national, régional, départemental.  
Eco-Santé c'est enfin des bases de données mises à jour régulièrement.

France	Régions & Départements	Québec
Mise à jour : décembre 2012 GRATUIT	Mise à jour : décembre 2012 GRATUIT	Mise à jour : décembre 2012 GRATUIT
		
430 000 données de 1990 à 2012 1 pays Les séries disponibles Disponible en 	4 700 000 données de 1990 à 2011 124 séries géographiques Les séries disponibles Disponible en 	800 000 données de 1991 à 2010 75 régions Les séries disponibles Disponible en 

Une collaboration :

# Accès fichiers de données

- Sites perso de [gilles.hunault@univ-angers.fr](mailto:gilles.hunault@univ-angers.fr)
  - **Fichiers de données réelles pour études statistiques + sites de stockage**
    - <http://www.info.univ-angers.fr/pub/gh/Datasets/datasets.htm>

# Un site qui recense les sites!

- <http://lib.stat.cmu.edu/datasets/>
- Attention dernière mise à jour 2005

# Des exemples d'activités de classe



Nous présentons ici dix exemples d'activités de classe à partir de « vraies données » obtenues sur le net dans les domaines suivants :

1. Santé.
2. Economie et social.
3. Climat et développement durable.
4. Société.



# Des exemples d'activités de classe :

## 1. Consommation d'antibiotiques

[antibiotiques.xls](#)

The screenshot shows the ANSM (Agence nationale de sécurité du médicament et des produits de santé) website. The header includes the ANSM logo, navigation links (Glossaire, Abonnement, Rechercher..., Valider), and user role buttons (Grand public, Professionnels de santé, Industriels, Experts, Presse). The main navigation bar contains 'L'ANSM', 'S'informer', 'Activités', 'Dossiers thématiques' (highlighted), 'Publications', and 'Services'. The breadcrumb trail reads: Accueil > Dossiers thémat... > Antibiotiques > Bien utiliser les antibiotiques. The 'Dossiers thématiques' sidebar lists various topics, with 'Antibiotiques' selected. The main content area for 'Antibiotiques' includes a list of sub-topics: 'Bien utiliser les antibiotiques', 'Plan national d'alerte sur les antibiotiques - Contribution de l'ANSM', 'Dermatologie', 'Gastro - Entérologie', 'Gynécologie - Urologie', 'Odonto - Stomatologie', 'Ophtalmologie', 'Pneumologie - ORL', 'Autres recommandations impliquant des antibiotiques', and 'Infos médicaments'. A pagination control shows '← 1 2 3 4 5 6 7 ... 10 →'. Below the main content, the 'Actualité' section features the article 'Evolution des consommations d'antibiotiques en France entre 2000 et 2012 - Rapport d'analyse (17/06/2013) (1095 ko)'. On the right, a 'Produits de santé' section displays a grid of icons for various categories: PIC, MTG, PSL, PIA, THA, Med, MDS, SP, Vac, DM, DIV, Cos, PT, and Aul.

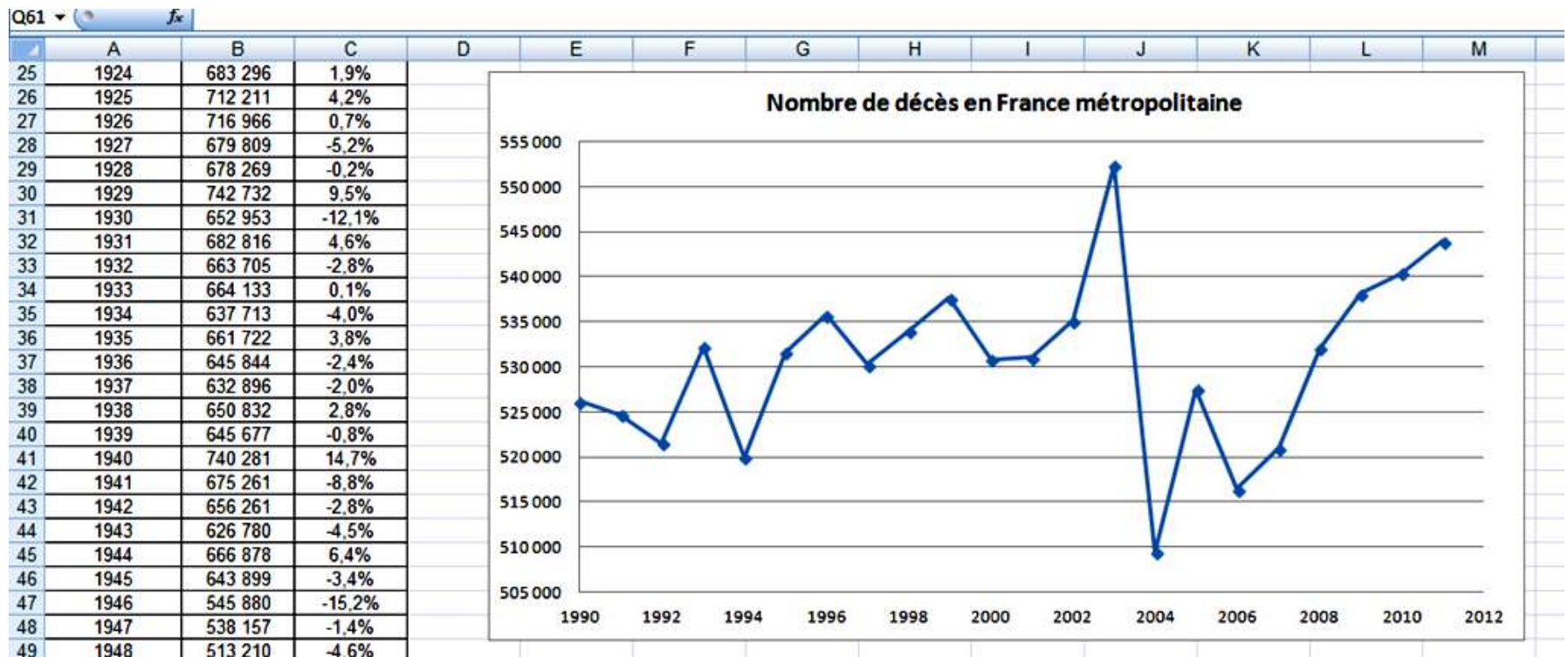


# Des exemples d'activités de classe :

## 2. Mortalité en France et ailleurs

[mortalite\\_corrige.xls](#)

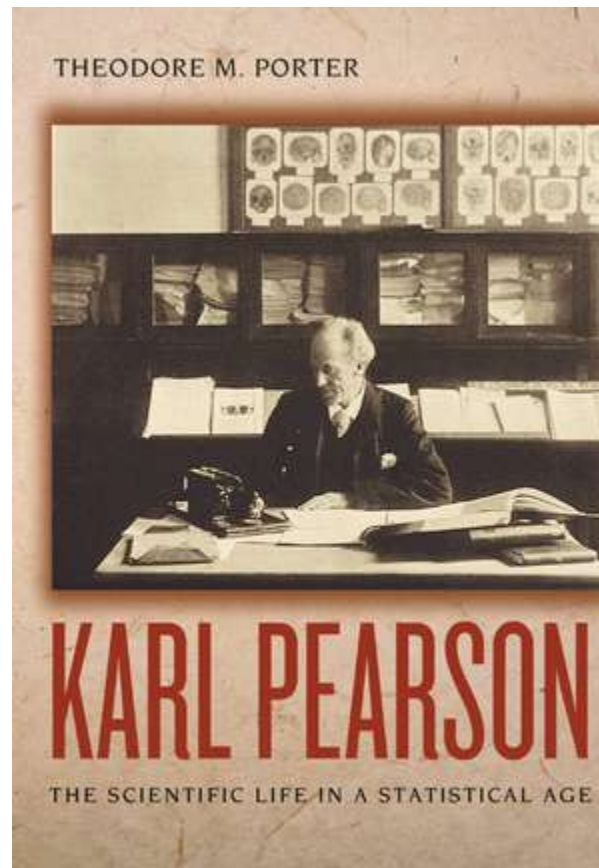
Sources : INSEE INED



# Des exemples d'activités de classe :

## 3. Tailles père / fils

[tailles\\_pere\\_fils\\_Pearson\\_corrige.xls](#)



# Des exemples d'activités de classe :

## 4. Indicateurs de pauvreté

[pauvrete\\_France.xls](#)

Source : INSEE

B10		fx =(1+B9)^(1/5)-1					
	A	B	C	D	E	F	G
1	<b>Pauvreté en France</b>						
2		2005	2006	2007	2008	2009	2010
3	Niveau de vie médian (euros 2010/mois)	1 527 €	1 548 €	1 582 €	1 608 €	1 613 €	1 606 €
4	Seuil de pauvreté (60% médiane)	916 €	929 €	949 €	965 €	968 €	964 €
5	Niveau de vie médian des personnes pauvres (euros 2010/mois)	744 €	762 €	776 €	786 €	784 €	781 €
6	Intensité de la pauvreté (%)	18,8%	18,0%	18,2%	18,5%	19,0%	18,9%
7	Nombre de personnes pauvres (en milliers)	7 766	7 828	8 035	7 836	8 173	8 617
8	Taux d'évolution annuel du nombre de personnes pauvres		0,80%	2,64%	-2,48%	4,30%	5,43%
9	Taux d'évolution du nombre de personnes pauvres sur la période	11%					
10	Taux d'évolution moyen annuel du nombre de personnes pauvres sur la période	2,101%					
11							



# Des exemples d'activités de classe :

## 5. Evolution du chômage

[chomage corrige.xls](#)

Source : INSEE



**Insee** Institut national de la statistique et des études économiques  
Mesurer pour comprendre

Mobile | Actualités | Agendas | Contactez-nous | Aide | FAQ | Liens | English | Home page

Chercher sur le site  

Accueil | **Thèmes** | Bases de données | Publications et services | Régions | Définitions et méthodes | Accès par public | L'Insee et la statistique publique

**Thèmes**

- o Agriculture
- o Commerce
- o Comptes nationaux - Finances publiques
- o Conditions de vie - Société
- > Conjoncture
  - > Indicateurs de conjoncture
  - + Tableau de bord de la conjoncture en France
  - + Analyse de la conjoncture
  - + Données économiques et financières
- o Économie

Accueil > Thèmes > Conjoncture > Indicateurs... > Chômage au sens du BIT et indicateurs sur le marché du travail (résultats de l'enquête emploi)

**Chômage au sens du BIT et indicateurs sur le marché du travail (résultats de l'enquête emploi)**

Document : [1-18] sur 18  
Page : [1]

- o **Le taux de chômage augmente de 0,1 point au deuxième trimestre 2013**  
En moyenne sur le deuxième trimestre 2013, le taux de chômage au sens du BIT, une fois neutralisé l'effet lié à la rénovation du questionnaire 2013, s'élève à 10,9 % de la population active en France ...  
Résultats du 2e trimestre 2013 - paru le 5 septembre 2013 à 07h30
- o Estimation du taux de chômage au sens du BIT pour le 1er trimestre 2013

**Téléchargement**

 La note méthodologique simplifiée (pdf, 144 Ko)

**Archives**

Voir les archives de l'indicateur

**Pour en savoir plus**

Retrouvez les séries longues dans la BDM : G688, G699, G842, G701, G705, G982.

# Des exemples d'activités de classe :

## 6. Normales de saison

[normales saison corrige.xls](#)

D'après IREM de Grenoble



  
ECA&D

European Climate Assessment & Dataset

Home    FAQ    Daily data    Indices of extremes    Return values    Extreme events    Project info

See also:    [KNMI Climate Explorer](#)    [ICA&D](#)    [EURO4M project](#)

Indices of extremes

**Indices of extremes**

For every ECA&D station, a total of 75 indices have been calculated. Indices data are freely available

# Des exemples d'activités de classe :

## 7. Production éolienne

[prod\\_eolienne\\_corrige.xls](#)

The screenshot shows the homepage of 'THE WINDPOWER' website, which is a database for wind turbines and wind farms. The page features a navigation menu with categories like Home, Country, Wind farms, Turbines, Players, Statistics, Media, and Reports. A central section titled 'Our products' lists various databases, including wind farms, offshore projects, developers, operators, owners, manufacturers, turbines, and countries. There are also sections for 'The best selling reports' and 'Actors'. The website includes a newsletter sign-up form, a 'Click Here for more details' button, and a sidebar with links to 'Wind Energy Cost' and 'Wind Power News'. The top right corner has a banner for 'GlobalData' and a search bar with flags for the UK, Germany, and Spain.

**THE WINDPOWER**  
Wind turbines and wind farms database

Unlock Your Investment Potential with **GlobalData**»  
[Click Here](#)

CONTACT PRICES/SERVICES SEARCH

Home Country Wind farms Turbines Players Statistics Media Reports

**Newsletter**  
Email address  
  
[Register](#)

**WIND PROSPECT**  
  
[Click Here](#)  
for more details

**The Wind Power** is a worldwide database about wind turbines and wind farms. It contains data related to wind farms, turbines, manufacturers, developers and operators.

**Our products**

- + Database of 14,157 wind farms (281.1 GW)
- + Database of 947 offshore projects (282.6 GW)
- + Database of 1,442 developers
- + Database of 2,117 operators
- + Database of 1,215 owners
- + Database of 146 manufacturers
- + Database of 959 turbines
- + Database of 98 countries
- + Strategic maps
- + Our booklet (PDF)

**The best selling reports**

- + Wind Operations & Maintenance Market, 2013 Update
- + Global Market Size, Turbine Market Share, Installation Prices, Regulations and Investment Analysis to 2020
- + The Importance of Ports in the UK's Booming Offshore Wind Industry
- + Other GlobalData reports

**Country**

- List
- World map
- Africa
- America
- Asia
- Europe
- Oceania

**Wind farms**

- List
- Offshore
- Africa
- America
- Asia
- Europe
- Oceania

**Turbines and manufacturers**

- Manufacturers
- Turbines
- Search

**Actors**

- Developers
- Operators
- Owners
- WEA

**Wind Energy Cost**

**Wind Power News**



# Des exemples d'activités de classe :

## 8. Old Faithful

[Old Faithful corrige.xls](#)

Un grand nombre de données : modélisation.



# Des exemples d'activités de classe :

## 9. Coupes du monde de football

[coupes du monde corrige.xls](http://www.fotforever.com)



The screenshot shows a website interface for the 2014 FIFA World Cup. At the top, there is a yellow banner with the URL [www.fotforever.com](http://www.fotforever.com) and a small globe icon. Below this is a black circular logo. The main content area features a yellow banner with the text "Coupe du Monde de Football - Brésil/Brazil 2014". Underneath are three icons: the Brazilian flag, the official 2014 FIFA World Cup logo, and a map of South America with Brazil highlighted in black. Below these icons is another yellow banner with the text "Tirage au sort dans 54 jour(s)". At the bottom of the page, a dark green banner contains the text "Coupe du monde de Football, Copa America et Euro par édition et par pays" and "Statistiques - Historique - Tous les matches - Tous les joueurs - Tous les buteurs".

[www.fotforever.com](http://www.fotforever.com)

Coupe du Monde de Football - Brésil/Brazil 2014

Tirage au sort dans 54 jour(s)

Coupe du monde de Football, Copa America et Euro par édition et par pays  
Statistiques - Historique - Tous les matches - Tous les joueurs - Tous les buteurs

# Des exemples d'activités de classe :

## 10. Sondages politiques

[sondages\\_Gallup.xls](#)



The screenshot shows the Gallup website's header with the logo and navigation menu. Below the header is a promotional banner for Gallup Analytics, featuring an image of a hand interacting with a tablet displaying data charts. To the right of the image is a sign-up form with the following fields:

**GALLUP**  
Search Gallup.com Search

HOME POLITICS ECONOMY WELLBEING WORLD GALLUP ANALYTICS STRATEGIC CONSULTING

**Gallup Analytics is coming soon.** Sign up now for more information and to receive the latest updates.

**Organization \***

**Name \***

First Last

**Email \***

Phone

**Submit**

**GALLUP ANALYTICS**  
Interact with, explore, and analyze data from the U.S. and around the world.

Atelier P1-17  
Journées APMEP de Marseille  
20 octobre 2013  
SFdS - Groupe Enseignement de la statistique

**Merci de votre attention**



Marthe-Aline JUTAND  
Philippe DUTARTE