Description du jeu de données sur les évènements et les courses à pied à l'internationale pour l'année 2017

Mathilde Plard, Violaine Guichet

▶ **To cite this version:**

Mathilde Plard, Violaine Guichet. Description du jeu de données sur les évènements et les courses à pied à l'internationale pour l'année 2017. 2018. hal-02011955v2

HAL Id: hal-02011955
https://hal.science/hal-02011955v2

Submitted on 29 Nov 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

DESCRIPTION DU JEU DE DONNÉES 2017

LES ÉVÉNEMENTS ET LES COURSES À PIED À L'INTERNATIONAL EN 2017

Le Running DataLab

Le Running DataLab

Le Running DataLab (RDL) récolte, produit et gère les informations running. En plus de cette partie géomatique, le RDL est aussi le lieu de test.

Missions du Running DataLab

En effet, la seconde mission du DataLab est de mettre au point des méthodologies innovantes de récoltes et d'analyses de données. Le but de ces tests sont : (1) de réaliser des méthodologies qui soient opérationnelles et interopérables sur d'autres projets de sciences humaines et sociales (2) trouver des moyens de représentations graphiques qui permettent une lecture simple de jeu de données complexes et riches, telles celles du Big Data.

Acteurs du Running DataLab

Le Running DataLab rassemble chercheurs et professionnels. Précisément, ce document a été dirigé par Mathilde Plard, chercheuse CNRS et directrice scientifique du RDL et Violaine Guichet, Urbaniste-Géomaticienne à l'Atelier de l'espace, directrice technique et graphique du RDL.

Pour plus d'informations :

running-datalab.com
latelier-de-lespace.com

cnrs
L'atelier de l'espace

Les jeux de données

Présentation du document

Le document présente les jeux de données mis à disposition sur la plateforme Nakala.

Ces derniers regroupent des informations sur les événements de courses à pied et les courses à pied à l'échelle internationale en 2017. Pour faciliter le partage, ces documents sont transmis en format .xlsx. La longitude et la latitude de chaque entité sont indiquées, ceci permet de les spatialiser avec un Système d'Information Géographie (SIG).

La création du jeu de données n'est pas détaillée ici puisqu'elle a fait l'objet de différents papiers. Les références sont indiquées ci-dessous pour connaître la méthodologie.

Liens pérennes vers les fichiers présentés

Les deux fichiers présentés disposent de handle – identifiant pérenne (handle) permettant en plus de la sécurisation des données, un accès permanent.

Pour le fichier concernant les événements de courses à pied internationaux en 2017, voici le lien :

https://www.nakala.fr/data/11280/4a43ea4c

Le lien suivant dirige vers le fichier sur les courses à pied internationales en 2017 :

https://www.nakala.fr/data/11280/edb930a8.

Références sur la méthodologie de création de jeu de données

– Violaine Guichet, Mathilde Plard. Captation, extraction et restructuration de données à partir de sources numériques hétérogènes. 2018. <hal-01939647>

– Violaine Guichet, Mathilde Plard. Construction de l'objet running à partir de sources numériques hétérogènes. 2018. <hal-01943614>

Précautions.

Précédemment, il a été spécifié complexité de la manipulation des données du web (3 V). Ainsi, le jeu de données créé a des limites.

Toutes les informations ne sont pas données pour l'entièreté des événements ou des courses. Afin que l'utilisation des données soit précise, il faut mentionner pour chaque carte produite la part du jeu de données mobilisée pour avertir le lecteur.

Malgré un contrôle des doublons, la variété des langues, des orthographes ou encore de règle de nommages ne permet pas de réaliser une vérification exhaustive. Ainsi, il existe encore des doublons. Pour pallier ce biais, une marge d'erreur a été calculée.

Pour des raisons de temps, il est impossible de vérifier la pertinence sémantique des données. Cependant, les sites sources sont des sites fiables (institutionnels ou semi-institutionnels). Les données qui y sont déposées sont vérifiées et validées par des administrateurs. Ainsi peut être imaginé que la marge d'erreur est faible.

Il est difficile de connaître le nombre de courses à pied organisées qui se déroulent autour du monde. Il est alors impossible de connaître le pourcentage d'exhaustivité de cette base de données constituée. Selon une étude de 2014 publiée par Kantar Media et Uniteam Sport, en 2013 ont été organisés 5 971 événements de courses à pied en France[1]. Le journal le parisien estime entre « 5 000 à 6 000 épreuves annuelles françaises »[2]. Dans la base de données créée par le DataLab, la France comptabilise en 20 176 314 événements de courses à pied.

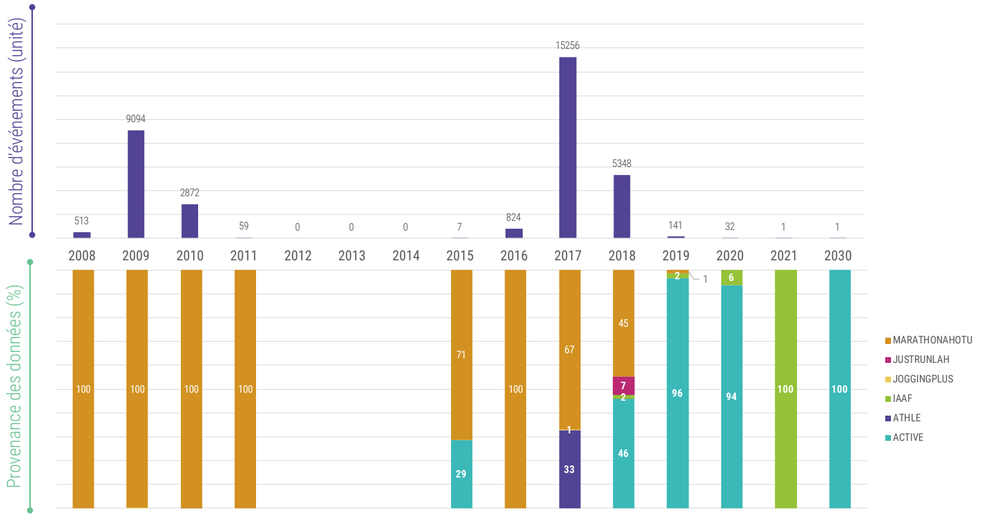
La base de données spatiale du DataLab regroupe 34 984 événements de courses à pied.

1 : http://www.sportbuzzbusiness.fr/caisse-depargne-kantar-media-uni-team-sport-decryptent-pratique-du-running-en-france.html, site consulté le 13/04/2018
2 : http://www.leparisien.fr/magazine/grand-angle/le-parisien-magazine-le-running-un-business-en-or-16-12-2016-6455259.php, site consulté le 13/04/2018

Ces événements s'étalent de 2008 à 2030. Le site Marathon Ahotu représente les trois quarts des événements extraits. Athlé englobe environ 15 %, Active 9 % JustRunLah 1 %, IAAF 0,3 % et JoggingPlus 0,3 %.

Ces pourcentages s'expliquent de différentes manières. Si marathon Ahotu représente 75 % c'est non seulement parce que le site dispose d'une offre importante, mais aussi parce que toutes les données (de 2006 à 2019) ont été extraites. Les sites ATHLE et IAAF n'englobent que les courses officielles (classantes et qualifiantes), ainsi les événements non officiels ne sont pas compatibilités. De plus, pour ATHLE et JoggingPlus n'ont été extraits que les événements de 2017 et pour JustRunLah, ceux de 2018. Pour JoggingPlus, seuls les événements des Pays de la Loire ont été extraits. Aussi, le site JustRunLah ne publie que les courses d'Asie et d'Australie.

Le jeu de données « événement ».



34 984 événements de courses à pied

6 sites sources

De **2008** à **2030**

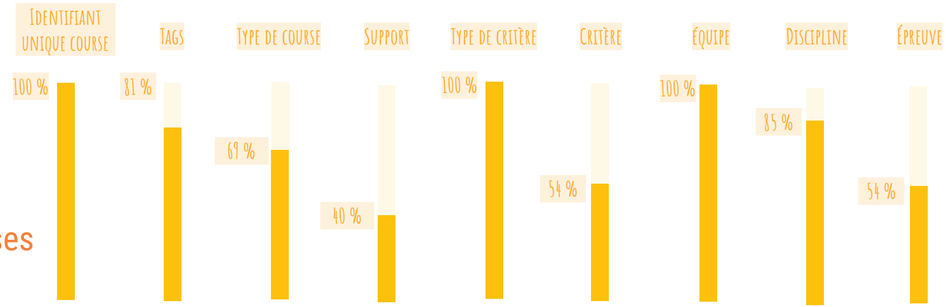
Events datas.

Ici sont présentées les informations disponibles sur les événements ainsi que le pourcentage d'objets qui détiennent l'information.

Races datas.

87 880 courses à pied

Environ **2,5 courses** par **événement**

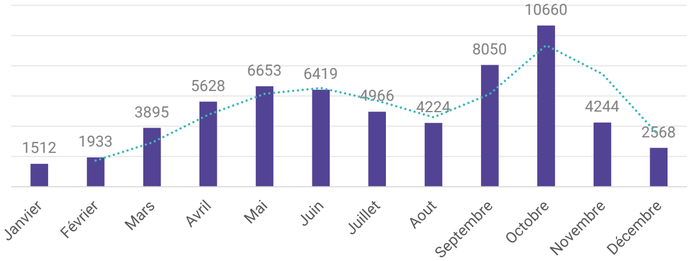


Ici sont présentées les informations disponibles sur les courses à pied ainsi que le pourcentage d'objets qui détiennent l'information.

Les données « événement » 2017.

Origines numériques des événements de courses à pied internationaux (unités)

Temporalités des événements de courses à pied internationaux (unités)



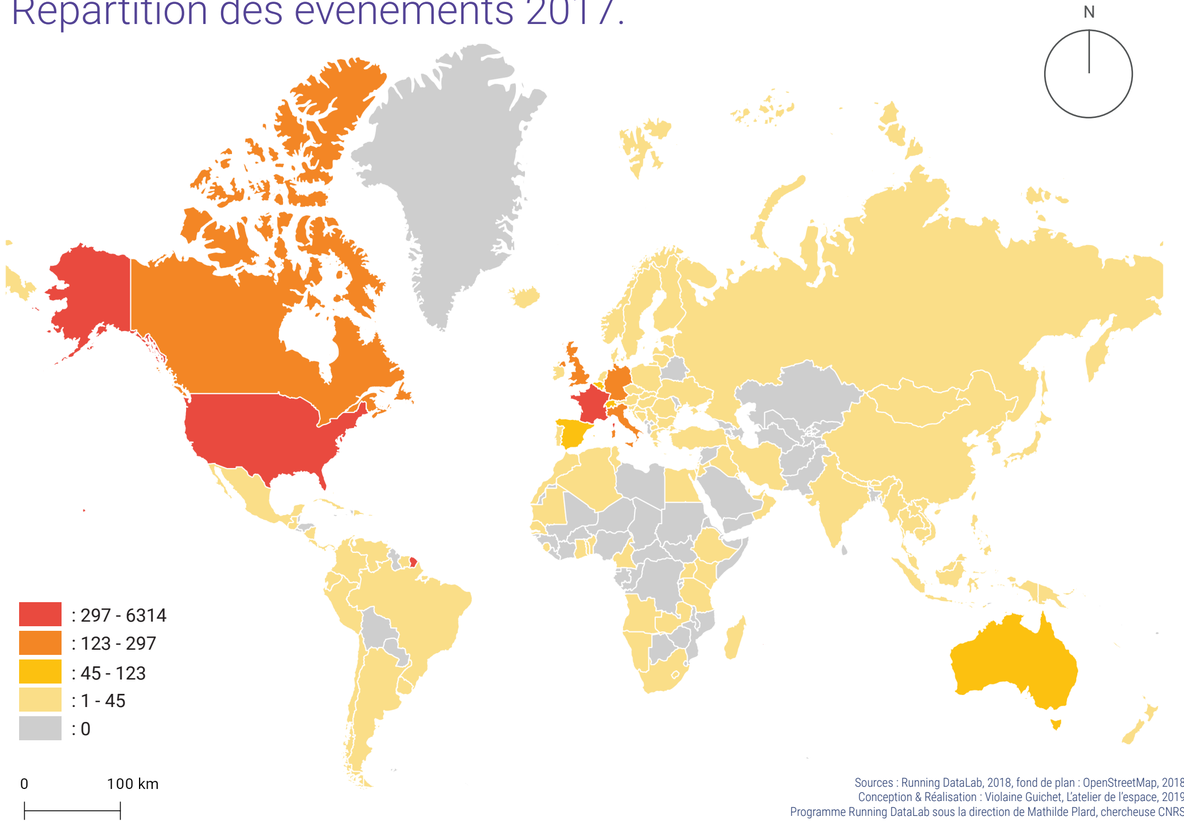
Le jeu de données recense 15256 événements internationaux de courses à pied pour l'année 2017.

Ces événements proviennent de trois sites sources. Environ douze pour cent des événements de courses à pied sont présents dans plusieurs sites (deux à trois sites. Concernant le reste des événements, 57 % proviennent du site Marathon Ahotu, 31 % du site Athlé et 0,3 % de Jogging-plus.

Les deux « pics » événementiels se déroulent avant et après les mois de juillet et aout. Une première hausse se déroule d'avril à juin. La seconde a lieu en septembre et octobre.

Les plus fortes densités événementielles se situent en Europe, Amérique du Nord et Australie.

Répartition des événements 2017.



Sources : Running DataLab, 2018, fond de plan : OpenStreetMap, 2018
Conception & Réalisation : Violaine Guichet, L'atelier de l'espace, 2019
Programme Running DataLab sous la direction de Mathilde Plard, chercheuse CNRS

Events datas 2017.

Ici sont présentées les variables disponibles sur les événements internationaux de courses à pied pour l'année 2017. Chaque fois est indiquée la part des événements qui sont renseignés par cette variable.

Détails des variables.

Identifiant unique événement

Chaque événement possède un identifiant unique.

Sites sources

Sites internet d'où proviennent les informations concernant l'événement.

Nom

Nom de l'événement.

Date de début

Date de début de l'événement.

Date de fin

Date de fin de l'événement.

Localisation

Localisation de l'événement (longitude, latitude).

Edition

Édition de l'événement.

Nombre de finishers

Nombre de finishers de l'événement - attention ne pas confondre avec le nombre de participants.

Organisateur

Organisateurs de l'événement.

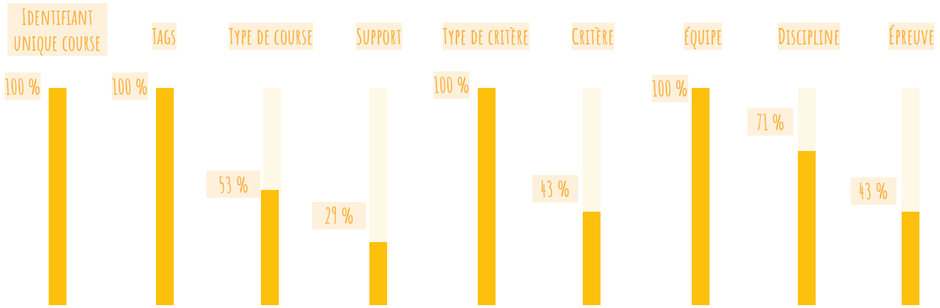
Types FFA

Dans ce jeu de données sont présents les types de compétition suivants : « Meet », « Cham », « G-E, Meet, France », « Coup », « Matc ». Cette catégorisation est celle de la Fédération Française d'Athlétisme.

Niveau FFA

Dans ce jeu de données sont présents les niveaux de compétition événementiels de courses à pied suivants : « Int », « Dép », « Nat », « Rég », « Mon », « Meet », « Eur ». Cette catégorisation est celle de la Fédération Française d'Athlétisme.

Les données « courses » 2017.



37 480 courses à pied en 2017.

Ici sont présentées les informations disponibles sur les courses à pied internationales pour l'année 2107 ainsi que le pourcentage d'objets qui détiennent l'information.

Détails des variables.

Identifiant unique course

Chaque course à pied possède un identifiant unique. Cet identifiant unique est composé d'un numéro de course singulier concaténé à l'identifiant unique de l'événement afin de connaître l'événement hôte de la course à pied.

Tags

Les tags sont des mots-clés renseignés par les organisateurs pour définir la course à pied. Les attributs étant distincts d'un site à l'autre, leur nombre est grand.

Type de course

Le type de course à pied classe les courses selon ces attributs : « sprint », « demi-fond », « fond », « ultrafond », « course sans distance », « multisports », « course verticale ».

Support

Le support des courses à pied est distinct par : « Course sur sentier » et « Course sur route ».

Type de critère

Le type de critère renvoie les attributs suivants : « Distance », « Dénivelé », « Temps ».

critère

En fonction du type de critère, les critères sont précisés. Par exemple, pour « Distance » il y a comme critères tels que : 10 K, 60 K, 21 097 K, etc., ou encore pour « Temps » : 12 h, 24 h, etc.

Équipe

Ici, deux attributs sont renseignés : « Solo » ou « Relais »

Discipline

La discipline catégorise les courses selon les attributs suivants : « Course à pied », « Course verticale », « Course à obstacles », Course à étapes », « Course sans distance », « Triathlon », « Relais », « Duathlon »

Épreuve

L'épreuve reprend les critères afin de les regrouper. Exemple : « Semi-marathon », « 10 K », «] 50 K ; 100K[“, ‘100 K’, ‘50 K’,” Entre 100 K et 200 K », “] 10 K ; 21 097 K [(hors 20 K)”, etc.

Métadonnées

INT_RUNNING_EVENT_2107

Intitulé de la ressource : Jeu de données des événements de courses à pied internationaux en 2017

Résumé de la ressource : le jeu de données recense les événements internationaux de courses à pied se déroulant en 2107. Il a été créé à partir de sources hétérogènes dans le cadre du projet de recherche Running DataLab (running-datalab.com).

Emprise géographique : internationale

Étendue temporelle : du 1er janvier 2017 au 31 décembre 2017

Format :. xlsx

Encodage : UTF-8

Date de création du jeu de données : 2018

Productrice : Violaine Guichte, Ingénieur d'études à l'université de Nantes

Contact : violaine.guichet@latelierdelespace.com

INT_RUNNING_RACE_2017

Intitulé de la ressource : Jeu de données des courses à pied internationales en 2017

Résumé de la ressource : le jeu de données recense les courses à pied internationales se déroulant en 2107. Il a été créé à partir de sources hétérogènes dans le cadre du projet de recherche Running DataLad (running-datalab.com).

Emprise géographique : internationale

Étendue temporelle : du 1er janvier 2017 au 31 décembre 2017

Format :. xlsx

Encodage : UTF-8

Date de création du jeu de données : 2018

Productrice : Violaine Guichet, Ingénieur d'études à l'université de Nantes

Contact : violaine.guichet@latelierdelespace.com

Contacts :

Responsable scientifique

mathilde.plard@univ-nantes.fr

Responsable technique et graphique

violaine.guichet@latelierdelespace.com

Sites :

https://running-datalab.com
https://www.latelier-de-lespace.com/

Conception et réalisation du document :

V. Guichet, l'atelier de l'espace 2019

cnrs

L'atelier de l'espace