Open Data

by [Santiago Mota](https://www.linkedin.com/in/santiagomota/)

2019-10-08

[Enlace Github](https://github.com/santiagomota/Open_Data)

## Fuentes de datos abiertos y APIs

* [CRAN Task View OpenData](https://github.com/ropensci/opendata)
* [Datos en paquetes de R](http://stat.ethz.ch/R-manual/R-patched/library/datasets/html/00Index.html)
* [Kaggle datasets](https://www.kaggle.com/datasets)
* [Revolution Analytics datasets (Microsoft R)](https://mran.revolutionanalytics.com/documents/data/)
* [UCI Machine Learning Repository](http://archive.ics.uci.edu/ml/)
* [DH Network](http://opendhn.dhnetwork.opendata.arcgis.com/)
* [9 Datasets para sistemas de recomendación](http://www.lab41.org/nine-datasets-for-investigating-recommender-systems/?utm_campaign=Data%2BElixir&utm_medium=email&utm_source=Data_Elixir_71)
* [Helsinki Open Data](http://www.hri.fi/en/)
* [Datasets de Quandl](https://www.quandl.com/search?query=)
* Amazon AWS: [este](http://aws.amazon.com/es/datasets/) y [este](https://aws.amazon.com/es/public-data-sets/)
* [Gobierno Estados Unidos](http://www.data.gov/)
* [Datos abiertos de la Unión Europea](http://open-data.europa.eu/es/data/)
* [Recopilatorio de datasets en Github](http://www.datasciencecentral.com/profiles/blogs/great-github-list-of-public-data-sets)
* [API de GitHub](https://developer.github.com/v3/)
* [API de Facebook](https://developers.facebook.com/docs/graph-api)
* [Blog. 100 recursos sobre Big Data y Data Science](http://todobi.blogspot.com.es/2015/02/mas-de-100-recursos-sobre-big-data-y.html)
* [NASDAQ](https://indexes.nasdaqomx.com/Index/History/NQASPA8600AUD)
* [Google finanzas](http://www.google.com/finance/)
* [CaixaBank Research](http://www.caixabankresearch.com/es/home)
* [Satélite Landsat](https://aws.amazon.com/public-data-sets/landsat/)
* [OCDE](https://data.oecd.org/)
* [19 Free Public Data Sets For Your First Data Science Project](https://www.mysliderule.com/blog/free-public-data-sets-data-science-project/?__s=atijywgwsusv7a2xfbee)
* [Open data EMT](http://opendata.emtmadrid.es/)
* [Datos abiertos del gobierno de España](http://datos.gob.es/)
* [Datos abiertos del Ayuntamiento de Madrid](http://datos.madrid.es/)
* [Datos abiertos de la Generalitat de Cataluña](http://dadesobertes.gencat.cat/es/)
* [Datos abiertos Junta de Andalucía](http://www.juntadeandalucia.es/datosabiertos/portal.html)
* [Datos abiertos de Santander](http://datos.santander.es/)
* [Natural Earth](http://www.naturalearthdata.com/)
* [Fuentes de datos espaciales](http://www.diva-gis.org/Data)
* [Opendata del CERN](http://opendata.cern.ch/)
* [Paquete de R ‘datasets’](http://stat.ethz.ch/R-manual/R-patched/library/datasets/html/00Index.html)
* [46 museos y bibliotecas que han digitalizado todo su conocimiento y lo ofrecen gratis en internet](http://www.xataka.com/otros/46-museos-y-bibliotecas-que-han-digitalizado-todo-su-conocimiento-humano)
* [Infraestructura de Datos Espaciales de España](http://idee.es/web/guest/centros-de-descarga?p_p_id=NewCentrosDescarga_WAR_NewCentrosDescargaportlet&p_p_lifecycle=1&p_p_state=normal&p_p_mode=view&p_p_col_id=column-1&p_p_col_count=1&_NewCentrosDescarga_WAR_NewCentrosDescargaportlet_tipoAmbito=LOCAL)
* [Infraestructura de Datos Espaciales de la Comunidad de Madrid](http://www.madrid.org/cartografia/idem/html/web/index.htm)
* [Microsoft Cognitive Services](https://www.microsoft.com/cognitive-services/)
* [Microsoft Project Oxford](https://www.projectoxford.ai/)
* [Google Cloud Vision API](https://cloud.google.com/vision/)
* [Análisis de 1.100 millones de trayectos de taxis y uber en NYC](https://github.com/toddwschneider/nyc-taxi-data)
* [European Data Portal](http://www.europeandataportal.eu/)
* [Propublica](https://www.propublica.org/data/)
* [NOAA. Agencia de meteo. USA.](http://www.nesdis.noaa.gov/index.html)
* [Datosclima. Base de datos meteo](http://datosclima.es/Aemet2013/DescargaDatos.html)
* [Dirección General de Tráfico (DGT)](https://sedeapl.dgt.gob.es/WEB_IEST_CONSULTA/inicio.faces)
* [National Historical Geographic Information System (NHGIS)](https://www.nhgis.org/)
* [Datos de todos los vuelos en USA entre 1987 y 2008 (datos originales)](http://stat-computing.org/dataexpo/2009/the-data.html)
* [Datos de todos los vuelos en USA entre 1987 y 2008 (otra fuente y ejemplos de uso en H2O). 120G](https://github.com/h2oai/h2o-2/wiki/Hacking-Airline-DataSet-with-H2O)
* [Conferencia internacional de datos abiertos en Madrid](http://opendatacon.org/)
* [Open Data Inception. 1.600 portales abiertos](http://wwwhatsnew.com/2016/03/19/open-data-inception-recopilacion-de-1600-portales-de-datos-abiertos/?utm_content=buffer4e4d4&utm_medium=social&utm_source=linkedin.com&utm_campaign=buffer)
* [Yahoo Flickr Creative Commons 100 Million (YFCC100m) dataset](http://yfcc100m.appspot.com/)
* [ImageNet database](http://www.image-net.org/)
* [Deep Learning datastsets](http://deeplearning.net/datasets/)
* [API TomTom. Tráfico en ciudades](http://developer.tomtom.com/products/onlinenavigation/onlinetraffic/onlinetrafficflow)
* [Mapas de Open Street Maps](http://download.geofabrik.de/)
* [European Data Portal](https://www.europeandataportal.eu/)
* [20 Awesome Websites For Collecting Big Data](https://datafloq.com/read/20-awesome-websites-for-collecting-big-data/2737?utm_source=Datafloq%20newsletter&utm_campaign=979b1fada5-EMAIL_CAMPAIGN_2017_03_13&utm_medium=email&utm_term=0_655692fdfd-979b1fada5-90449429)
* [Climate Data Online](https://www.ncdc.noaa.gov/cdo-web/)
* [Una recopilación de APIs públicas](https://github.com/toddmotto/public-apis)
* [Una recopilación de datasets públicos](https://github.com/caesar0301/awesome-public-datasets)
* [Recopilación de datasets de BigML](https://blog.bigml.com/list-of-public-data-sources-fit-for-machine-learning/)
* [Datasets de ejemplo de IBM Watson Analytics](https://www.ibm.com/communities/analytics/watson-analytics-blog/guide-to-sample-datasets/)
* [70 amazing and free data sources for data visualization](http://bigdata-madesimple.com/70-amazing-and-free-data-sources-for-data-visualization/?lipi=urn%3Ali%3Apage%3Ad_flagship3_feed%3BAJsccljrTlK%2BpMAMcPD6Yw%3D%3D)
* [Some datasets for teaching data science](https://simplystatistics.org/2018/01/22/the-dslabs-package-provides-datasets-for-teaching-data-science/?lipi=urn%3Ali%3Apage%3Ad_flagship3_feed%3BaLl4moGVT%2BS65o19ly%2F%2FaQ%3D%3D)
* [El planeta Tierra en AWS](https://aws.amazon.com/es/earth/)
* [Tráfico en el Reino Unido](https://www.dft.gov.uk/traffic-counts/)
* [European Banking Authority (EBA)](http://www.eba.europa.eu/risk-analysis-and-data;jsessionid=9D39E5DC46932D2C76C572A1D1B1B5BA)
* [World Bank Open Data](https://data.worldbank.org/)
* [Fondo Monetario Internacional](http://www.imf.org/en/data)
* [Lista de algunos datatsets dentro de paquetes de R](https://vincentarelbundock.github.io/Rdatasets/datasets.html)
* [10.000 airports, train stations and ferry terminals spanning the globe](https://openflights.org/data.html)
* [30 Amazing (And Free) Big Data And AI Public Data Sources For 2018](https://www.linkedin.com/pulse/30-amazing-free-big-data-ai-public-sources-2018-bernard-marr/?trackingId=nkTXcNLieYPDBqZuB3KIsw%3D%3D&lipi=urn%3Ali%3Apage%3Ad_flagship3_feed%3B9KuSD9KfQ6ie%2BALso3gwvw%3D%3D&licu=urn%3Ali%3Acontrol%3Ad_flagship3_feed-object)
* [Awesome Public Datasets 1](https://github.com/dipanjanS/awesome-public-datasets)
* [Awesome Public Datasets 2](https://github.com/awesomedata/awesome-public-datasets)
* [25 Open Datasets for Deep Learning Every Data Scientist Must Work With](https://www.analyticsvidhya.com/blog/2018/03/comprehensive-collection-deep-learning-datasets/)
* [NOAA Daily Global Historical Climatology Network - Kaggle dataset](https://www.kaggle.com/noaa/ghcn-d)
* [Crimen en UK](https://data.police.uk/)
* [Datos abiertos Ayuntamiento de Valencia](http://gobiernoabierto.valencia.es/es/data/)
* [Microsoft Research Open Data](https://msropendata.com/)
* [Kaggle Weekly Kernels Award Winner Announcements](https://www.kaggle.com/general/37924#post354114)
* [Open Data Renfe](http://data.renfe.com/)
* [Open Data Barometer](https://opendatabarometer.org/?_year=2017&indicator=ODB)
* [CIS. Centro de Investigaciones Sociológicas](http://www.cis.es/cis/opencms/ES/index.html)
* [Fivethirtyeight](https://data.fivethirtyeight.com/)
* [Reddit datasets](https://www.reddit.com/r/datasets/)
* [Data World](https://data.world/)
* [The world’s economic database](https://db.nomics.world/)
* [25 Open Datasets for Deep Learning Every Data Scientist Must Work With](https://www.analyticsvidhya.com/blog/2018/03/comprehensive-collection-deep-learning-datasets/?utm_source=linkedin.com&utm_medium=social)
* [Paquete pasra acceder al API del Instituto de Canario de Estadística](https://github.com/rOpenSpain/istacbaser)
* [Datos estadísticos DGT](https://sedeapl.dgt.gob.es/WEB_IEST_CONSULTA/)
* [Yelp Dataset](https://www.yelp.com/dataset)
* [Awesome Sentinel. Copernicus Sentinel Satellites resources](https://github.com/Fernerkundung/awesome-sentinel)
* [Datos Abiertos del Consorcio Regional de Transportes de Madrid](https://datos.crtm.es/)
* [Idealista labs](https://www.idealista.com/labs/)
* [Ayuntamiento de Madrid. Censo de locales, sus actividades y terrazas de hostelería y restauración](https://datos.gob.es/es/catalogo/l01280796-censo-de-locales-sus-actividades-y-terrazas-de-hosteleria-y-restauracion)
* [Agencia Tributaria. Estadísticas](https://www.agenciatributaria.es/AEAT.internet/Inicio/La_Agencia_Tributaria/Memorias_y_estadisticas_tributarias/Estadisticas/Estadisticas.shtml)
* [UK Office for National Statistics](https://www.ons.gov.uk/)
* [UK Open Geography Portal](https://geoportal.statistics.gov.uk/)
* [Open Trade Statistics](https://tradestatistics.io/)
* Facebook Neural-Code-Search-Evaluation-Dataset [dataset]](<https://github.com/facebookresearch/Neural-Code-Search-Evaluation-Dataset>) y [noticia](https://venturebeat.com/2019/10/03/facebook-open-sources-data-set-for-code-search-ai-benchmark/)
* Github semantic code search [page](https://experiments.github.com/semantic-code-search) y [noticia](https://venturebeat.com/2018/10/17/github-wants-ai-to-help-developers-code/)
* [Ultimos datos de Open Street Map. Spain](https://download.geofabrik.de/europe/spain.html)
* [NLP Datasets](https://github.com/niderhoff/nlp-datasets/blob/master/README.md)

## Otras referencias interesantes

* [Data Science Blogs](https://github.com/rushter/data-science-blogs)
* [Chuleta general de R](https://cran.r-project.org/doc/contrib/Baggott-refcard-v2.pdf)
* [R Learning Path: From beginner to expert in R in 7 steps](http://www.kdnuggets.com/2016/03/datacamp-r-learning-path-7-steps.html)
* [Tutorial de R Markdown](http://www.sfs.uni-tuebingen.de/~jvanrij/Tutorial/tutorialMarkdown.html)
* [Rstudio cheatsheets](https://www.rstudio.com/resources/cheatsheets/?utm_content=buffer1b56a&utm_medium=social&utm_source=twitter.com&utm_campaign=buffer)
* [R Markdown cheatsheet](https://www.rstudio.com/wp-content/uploads/2015/02/rmarkdown-cheatsheet.pdf)
* [R Markdown referencia](https://www.rstudio.com/wp-content/uploads/2015/03/rmarkdown-reference.pdf)
* [A dive into R Markdown](http://cfss.uchicago.edu/program_rmarkdown.html)
* [Información de Rmarkdown en R Studio](http://rmarkdown.rstudio.com/)
* [Template para documentos científicos con Rmarkdown](http://www.petrkeil.com/?p=2401)
* [Formatos a medida para R Markdown](http://www.r-bloggers.com/r-markdown-custom-formats/)
* [blogdown: Creating Websites with R Markdown](https://bookdown.org/yihui/blogdown/)
* [Utilizando Sweave y Knitr](https://support.rstudio.com/hc/en-us/articles/200552056-Using-Sweave-and-knitr)
* [Pandoc User’s Guide](http://pandoc.org/MANUAL.html#templates)
* [Soporte técnico de RStudio](https://support.rstudio.com/hc/en-us)
* [100 Active Blogs on Analytics, Big Data, Data Mining, Data Science, Machine Learning](http://www.kdnuggets.com/2016/03/100-active-blogs-analytics-big-data-science-machine-learning.html#.VvqjkSV5Tio.linkedin)
* [Plataforma H2O](https://github.com/h2oai)
* [Computer vision](https://github.com/kjw0612/awesome-deep-vision)
* Pautas para dar formato al código programando en R: [Google](https://google.github.io/styleguide/Rguide.xml) y [Hadley Wickham (RStudio)](http://adv-r.had.co.nz/Style.html)
* [The State of Naming Conventions in R](https://journal.r-project.org/archive/2012-2/RJournal_2012-2_Baaaath.pdf)
* [R Code – Best practices](https://www.r-bloggers.com/r-code-best-practices/)
* [R Coding Style Guide](https://irudnyts.github.io//r-coding-style-guide/)
* [Naming files](https://speakerd.s3.amazonaws.com/presentations/5e4b07f0d9a94f8e9a29b902bad6ed0b/naming-slides.pdf)
* [Documentacion de R](https://www.rdocumentation.org/)
* [Chuleta de expresiones regulares](https://www.rstudio.com/wp-content/uploads/2016/09/RegExCheatsheet.pdf)
* [Regular Expressions Every R programmer Should Know](https://www.r-bloggers.com/regular-expressions-every-r-programmer-should-know/)
* [Regular Expression Language - Quick Reference](https://docs.microsoft.com/en-us/dotnet/standard/base-types/regular-expression-language-quick-reference)
* [Dealing with Regular Expressions](http://uc-r.github.io/regex)
* [Writing an R package from scratch](https://hilaryparker.com/2014/04/29/writing-an-r-package-from-scratch/)
* [16 Cursos](https://www.analyticsvidhya.com/blog/2016/10/16-new-must-watch-tutorials-courses-on-machine-learning/?utm_source=feedburner&utm_medium=email&utm_campaign=Feed%3A+AnalyticsVidhya+%28Analytics+Vidhya%29)
* [Galerias de graficos](http://www.r-graph-gallery.com/)
* [Aprender ciencia de datos. Fuentes para Python y R](https://www.datacamp.com/community/tutorials/learn-data-science-resources-for-python-r)
* [Curso Caltech. Learning from data](https://work.caltech.edu/telecourse.html)
* [Usar git](https://try.github.io/levels/1/challenges/1)
* [Blogs con github](http://jmcglone.com/guides/github-pages/) y [Blogs con github y RStudio](http://andysouth.github.io/blog-setup/)
* [Ejemplos de Shiny](http://zevross.com/blog/2016/04/19/r-powered-web-applications-with-shiny-a-tutorial-and-cheat-sheet-with-40-example-apps/)
* [UK government using R to modernize reporting of official statistics](https://www.r-bloggers.com/uk-government-using-r-to-modernize-reporting-of-official-statistics/)
* [Great R packages for data import, wrangling and visualization](http://www.computerworld.com/article/2921176/business-intelligence/great-r-packages-for-data-import-wrangling-visualization.html?idg_eid=96342b7ab1a902d39d1237b81a92554b&email_SHA1_lc=&cid=cw_nlt_computerworld_enterprise_apps_2017-06-23&utm_source=Sailthru&utm_medium=email&utm_campaign=Computerworld%20Enterprise%20Apps%202017-06-23&utm_term=computerworld_enterprise_apps)
* [Ggplot](http://socviz.co/)
* Sistemas de Coordenadas. [Aqui](http://rspatial.org/spatial/rst/6-crs.html) y [aqui](https://www.nceas.ucsb.edu/~frazier/RSpatialGuides/OverviewCoordinateReferenceSystems.pdf)
* [Codificación de caracteres](https://www.joelonsoftware.com/2003/10/08/the-absolute-minimum-every-software-developer-absolutely-positively-must-know-about-unicode-and-character-sets-no-excuses/)
* [Tutorials for learning R](https://www.r-bloggers.com/how-to-learn-r-2/)
* [Awesome R](https://github.com/qinwf/awesome-R)
* [R Data Science Tutorials](https://github.com/ujjwalkarn/DataScienceR)
* [useR! Machine Learning Tutorial](https://github.com/ledell/useR-machine-learning-tutorial)
* [Paquetes de R interesantes](https://www.computerworld.com/article/2921176/business-intelligence/great-r-packages-for-data-import-wrangling-visualization.html?utm_content=bufferf197d&utm_medium=social&utm_source=linkedin.com&utm_campaign=buffer)
* [Otra lista de recursos variados en Github](https://github.com/Shujian2015/FreeML)
* [Tipos de licencias de software](https://choosealicense.com/licenses/)
* [Glosario de Machine Learning de Google](https://developers.google.com/machine-learning/glossary/)
* [Google Rules of Machine Learning: Best Practices for ML Engineering](http://martin.zinkevich.org/rules_of_ml/rules_of_ml.pdf)
* Statistical Learning de Stanford [Curso](https://lagunita.stanford.edu/courses/HumanitiesSciences/StatLearning/Winter2016/about), [Libro](https://web.stanford.edu/~hastie/ElemStatLearn/), [Código](https://github.com/khanhnamle1994/statistical-learning) y [Transparencias](https://github.com/khanhnamle1994/statistical-learning/tree/master/Lecture-Slides)
* [100 Free Tutorials for Learning R](https://www.listendata.com/p/r-programming-tutorials.html)
* [Google’s best practices in machine learning](https://developers.google.com/machine-learning/guides/rules-of-ml/)
* [Web Scraping TripAdvisor, Text Mining and Sentiment Analysis for Hotel Reviews](https://towardsdatascience.com/scraping-tripadvisor-text-mining-and-sentiment-analysis-for-hotel-reviews-cc4e20aef333)
* [Common Probability Distributions: The Data Scientist’s Crib Sheet](https://blog.cloudera.com/blog/2015/12/common-probability-distributions-the-data-scientists-crib-sheet/?utm_content=buffer49e9f&utm_medium=social&utm_source=facebook.com&utm_campaign=buffer)
* [RDocumentation](https://www.rdocumentation.org/)
* [ArcGIS to R spatial cheat sheet](http://www.seascapemodels.org/data/ArcGIS_to_R_Spatial_CheatSheet.pdf)
* [Hand-curated, high quality resources for doing data journalism with R](https://rddj.info/)
* [RMarkdown Driven Development (RmdDD)](https://emilyriederer.netlify.com/post/rmarkdown-driven-development/)

## Libros

* [R intro](https://cran.r-project.org/doc/manuals/R-intro.pdf)
* [R for everyone](https://www.jaredlander.com/r-for-everyone/)
* [R in action](https://www.manning.com/books/r-in-action-second-edition)
* [R Programming for Data Science. Roger D. Peng.](https://leanpub.com/rprogramming)
* [R para principiantes](https://cran.r-project.org/doc/contrib/rdebuts_es.pdf)
* [Introducción a R](https://cran.r-project.org/doc/contrib/R-intro-1.1.0-espanol.1.pdf)
* [R para profesionales de los datos: una introducción](https://www.datanalytics.com/libro_r/)
* [10 great books about R](https://www.datasciencecentral.com/profiles/blogs/10-great-books-about-r-1)
* [10 Free Must-Read Books for Machine Learning and Data Science](https://www.kdnuggets.com/2017/04/10-free-must-read-books-machine-learning-data-science.html?utm_content=bufferc386f&utm_medium=social&utm_source=twitter.com&utm_campaign=buffer)
* Introduction to Data Science [Libro](https://rafalab.github.io/dsbook/) y [Código](https://github.com/rafalab/dsbook)
* Fundamentals of Data Visualization [Libro](http://serialmentor.com/dataviz/?utm_content=buffer7a991&utm_medium=social&utm_source=linkedin.com&utm_campaign=buffer) y [Código](https://github.com/clauswilke/dataviz)
* Data Science Live Book [Libro](https://livebook.datascienceheroes.com/) y [Código](https://github.com/pablo14/data-science-live-book)
* R for Statistical Learning [Libro](https://daviddalpiaz.github.io/r4sl/) y [Código](https://github.com/daviddalpiaz/r4sl)
* Applied Statistics with R [Libro](https://daviddalpiaz.github.io/appliedstats/) y [Código](https://github.com/daviddalpiaz/appliedstats)
* Geocomputation with R [Libro](https://geocompr.robinlovelace.net/) y [Código](https://github.com/Robinlovelace/geocompr/)
* [Handling Strings with R](http://www.gastonsanchez.com/r4strings/)
* [Text Mining with R](https://www.tidytextmining.com/)
* [Efficient R programming](https://csgillespie.github.io/efficientR/)
* [BBC Visual and Data Journalism cookbook for R graphics](https://bbc.github.io/rcookbook/)
* [Databases using R by RStudio](https://db.rstudio.com/getting-started/)
* [Interpretable Machine Learning](https://christophm.github.io/interpretable-ml-book/)
* [Forecasting: Principles and Practice](https://otexts.com/fpp2/)
* [Bioinformática Estadística. Análisis estadístico de datos Ómicos](https://www.uv.es/ayala/docencia/tami/tami13.pdf)
* [Estadística básica](https://www.uv.es/ayala/docencia/nmr/nmr13.pdf)
* [Análisis de datos con R](https://www.uv.es/ayala/docencia/ad/ad13.pdf)
* [An Introduction to Spatial Data Analysis and Visualisation in R](https://www.spatialanalysisonline.com/An%20Introduction%20to%20Spatial%20Data%20Analysis%20in%20R.pdf)
* [Exploratory Data Analysis with R - Roger D. Peng](https://bookdown.org/rdpeng/exdata/)
* [What They Forgot to Teach You About R](https://whattheyforgot.org/)