PABLO FIGUEROA

MINERÍA DE DATOS

PROGRAMA

PROGRAMA

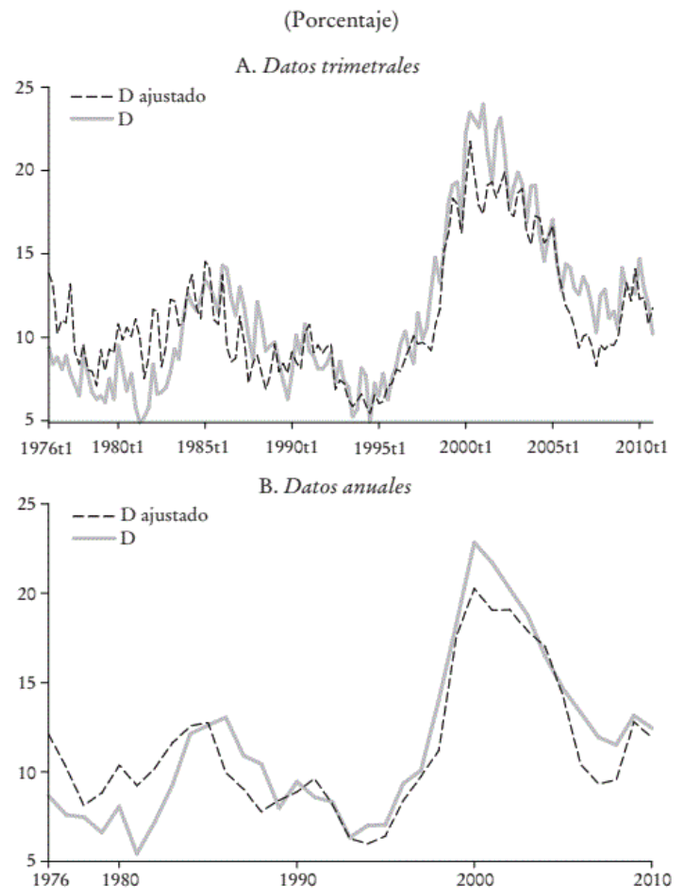
- Módulo I - Consideraciones iniciales
  - Introducción.
  - Diseño de la Investigación.
- Módulo II - Datos
  - Fuentes de datos.
  - Tipos de bases de datos.
  - Tipos de variables.

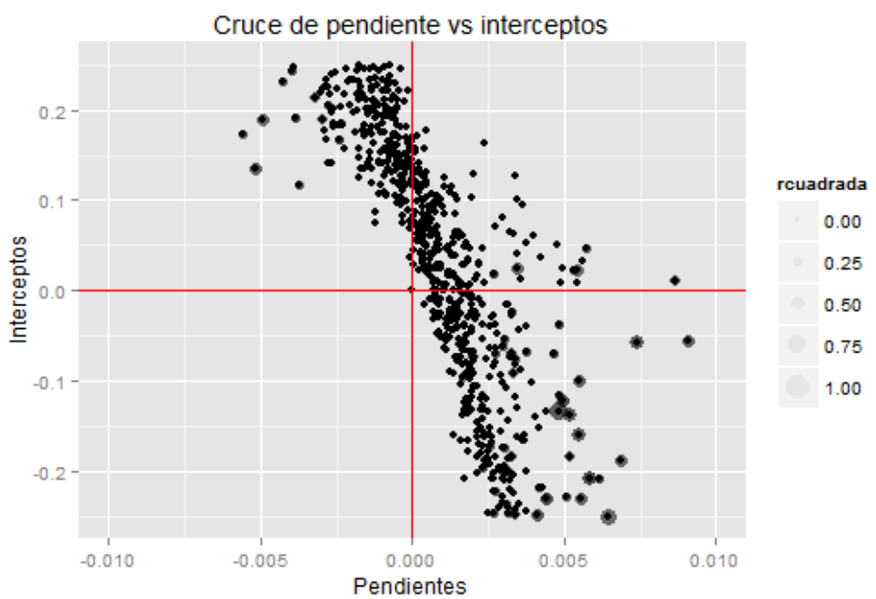
PROGRAMA

- Módulo III - Exploración
  - Estadística Descriptiva.
  - Reducción de Información.
  - Clustering.

PROGRAMA

- Módulo IV - Análisis
  - Conceptos minería de datos predictiva.
  - Regresiones.
  - Maquina Soporte Vectorial.
  - Árboles de Decisión.
  - Métodos de Consenso y Potenciación
  - Inferencia Bayesiana.
  - Redes Neuronales.
- Módulo V - Evaluación de modelos





ESTRUCTURA DE CLASES

18:00 - 18:45	18:45 - 19:30	19:30 - 20:15	20:15 - 21:00
Clase Teórica	Clase Teórica	Aplicación Teoría	Aplicación Proyecto
7		8	

PLAN DE HITOS

Fecha	Horario	Hito	Módulos	%
06 de Octubre	18:00-19:30	Prueba I	I - II - III	20%
06 de Octubre	19:30-21:00	Presentación de avances	I - II - III	10%
24 de Noviembre	18:00-19:30	Prueba II	IV - V	20%
24 de Noviembre	19:30-21:00	Presentación de avances	IV - V	10%
15 de Diciembre	18:00-21:00	Presentación de Proyectos y entrega Informes	I - II - III - IV - IV	40%
22 de Diciembre	18:00-19:30	Exámen	I - II - III - IV - IV	Suficiencia

TECNOLOGÍAS – BASE

python

jupyter

scikit learn

R

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$

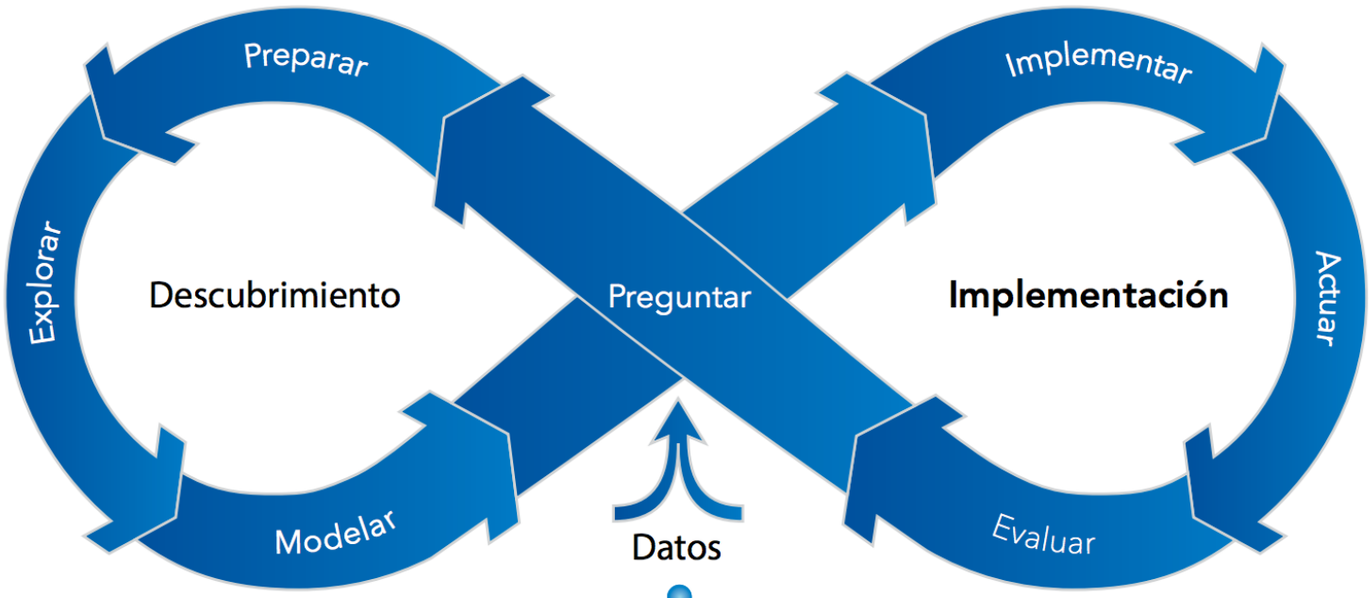
NumPy

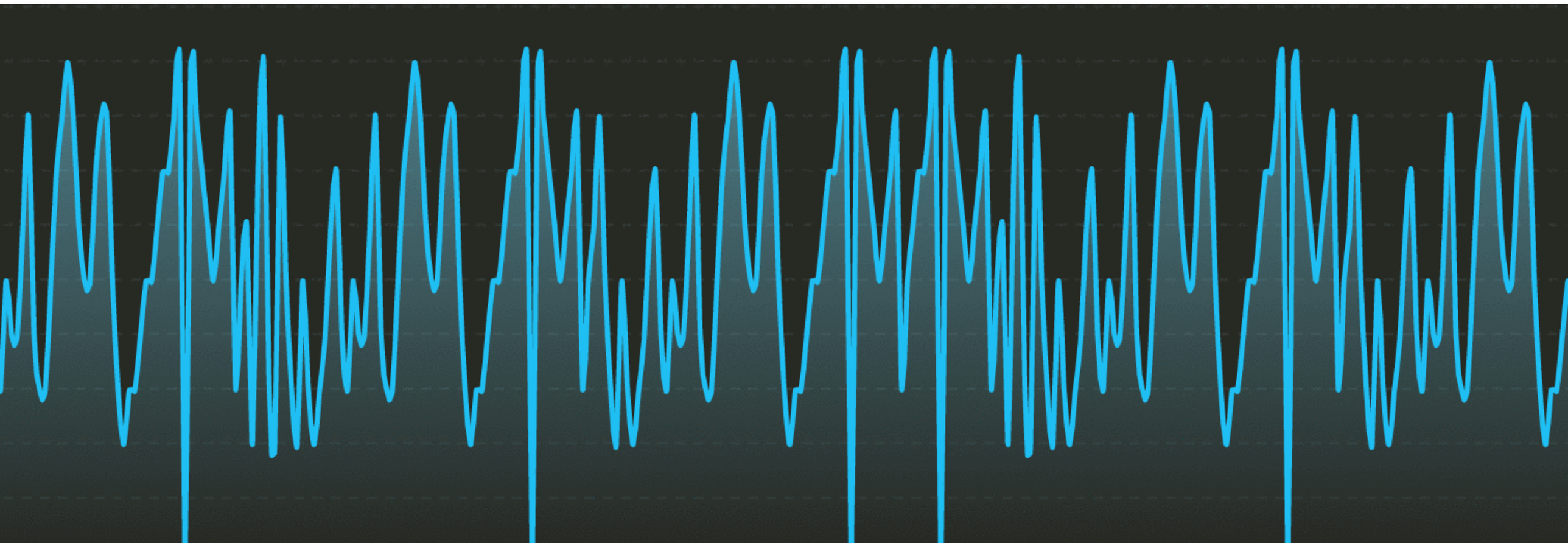
K + TensorFlow

matplotlib

INTRODUCCIÓN

MINERÍA DE DATOS





PROCESO NO TRIVIAL DE IDENTIFICAR PATRONES VÁLIDOS, NOVEDOSOS, POTENCIALMENTE ÚTILES Y EN ÚLTIMA INSTANCIA COMPRENSIBLES A PARTIR DE LOS DATOS.

Fayyad

MINERÍA DE DATOS

COMPRENSIÓN DEL NEGOCIO

COMPRENSIÓN DE LOS DATOS

MINERÍA DE DATOS

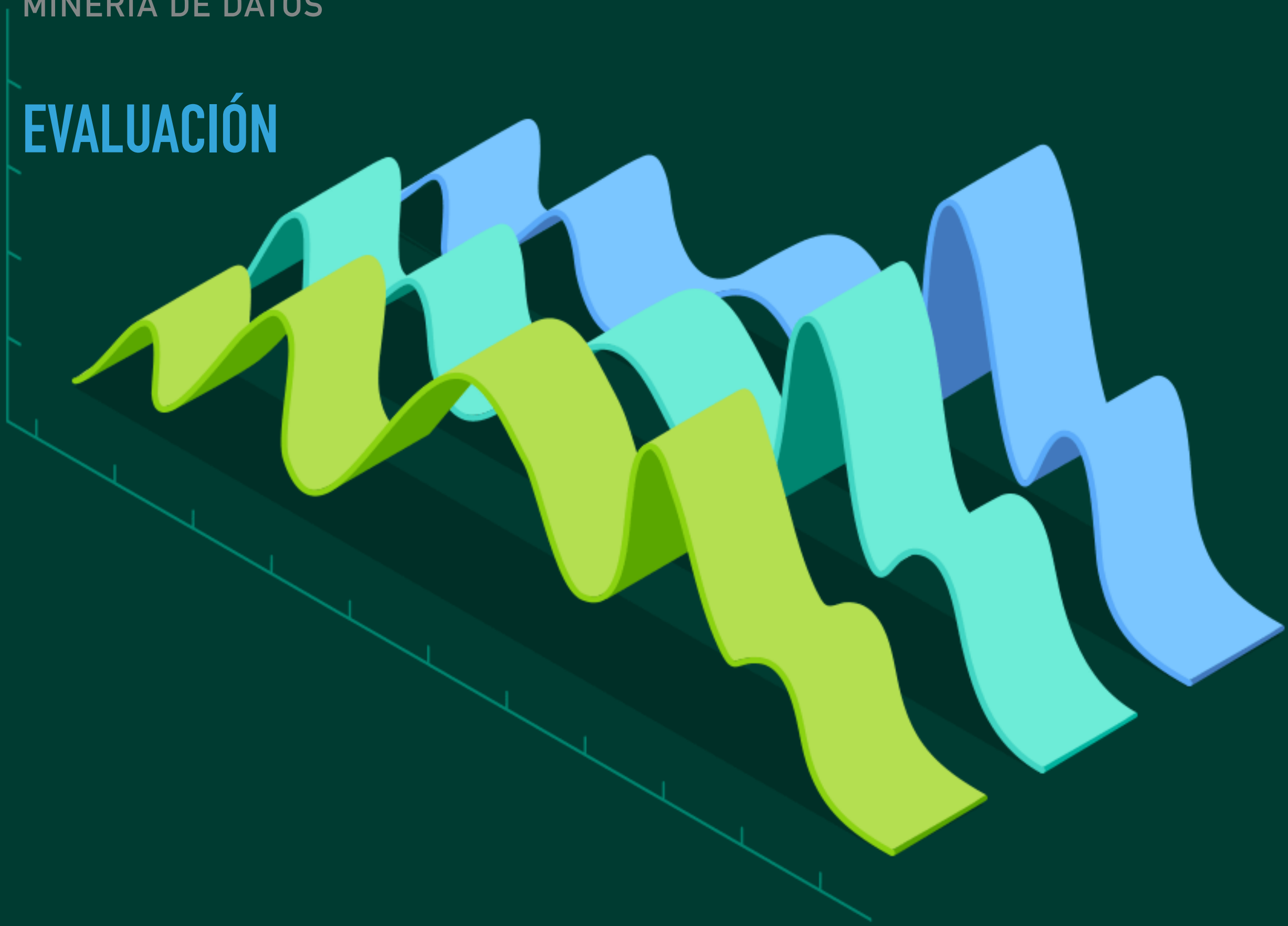
PREPARACIÓN DE LOS DATOS

MINERÍA DE DATOS

MODELADO

MINERÍA DE DATOS

EVALUACIÓN



DESPLIEGUE

PROYECTOS

KAGGLE DATASETS

- https://www.kaggle.com/datasets

Welcome to Kaggle Datasets

The best place to discover and seamlessly analyze open data

Discover

Use the search box to find open datasets on everything from government, health, and science to popular games and dating trends.

Explore

Execute, share, and comment on code for any open dataset with our in-browser analytics tool, Kaggle Kernels. You can also download datasets in an easy-to-read format.

Create a Dataset

Contribute to the open data movement and connect with other data enthusiasts by clicking "New Dataset" to publish an open dataset of your own.

Learn More | New Dataset

API SBIF

- http://api.sbif.cl/

API SBIF .v3

La Interfaz de Programación de Aplicaciones de la Superintendencia de Bancos e Instituciones Financieras de Chile (www.sbif.cl), en adelante API de SBIF, permite obtener información de manera directa desde la base de datos del sitio web utilizando los servicios web provistos en esta plataforma.

Saber más »

Qué es la API

La API SBIF permite el acceso a diversos recursos de información con datos actuales e históricos de los diferentes tipos de reportes que se ofrecen en el sitio web de la Superintendencia.

Revisar »

Ejemplos

Los datos de la API SBIF pueden ser consumidos para diversas funcionalidades; una de ellas es la de hacer gráficos con los datos obtenidos. Aquí mostramos algunos ejemplos.

Probar »

Documentación

Para facilitar el uso de la API, en esta sección del sitio web se entrega la documentación que reseña las características y forma de uso de cada recurso que se puede acceder a través de la API.

Ingresar »

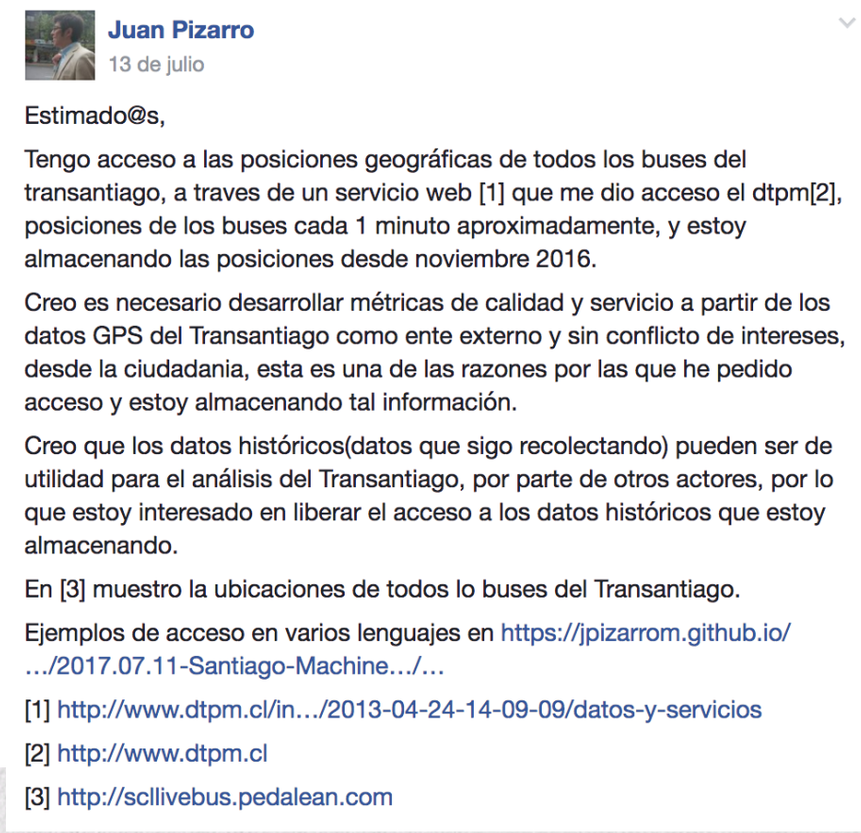
TRANSPARENCIA – GOBIERNO DE CHILE

http://datos.gob.cl/
https://datosabiertos.cl/
http://datos.bcn.cl/es/
http://datosabiertos.chilecompra.cl/home
http://opendata.congreso.cl/
http://www.consejotransparencia.cl/.../2012.../130034.html
https://www.camara.cl/camara/opendata.aspx
http://www.startupchile.org/industries/open-data/
http://energiaabierta.cl/

DATOS TRANSPORTE PÚBLICO

- http://www.dtpm.cl/index.php/2013-04-24-14-09-09/datos-y-servicios

https://jpizarrom.github.io/…/2017.07.11-Santiago-Machine…/…
http://www.dtpm.cl/in…/2013-04-24-14-09-09/datos-y-servicios
http://www.dtpm.cl
http://scllivebus.pedalean.com



Juan Pizarro
13 de julio

Estimado@s,

Tengo acceso a las posiciones geográficas de todos los buses del transantiago, a traves de un servicio web [1] que me dio acceso el dtpm[2], posiciones de los buses cada 1 minuto aproximadamente, y estoy almacenando las posiciones desde noviembre 2016.

Creo es necesario desarrollar métricas de calidad y servicio a partir de los datos GPS del Transantiago como ente externo y sin conflicto de intereses, desde la ciudadania, esta es una de las razones por las que he pedido acceso y estoy almacenando tal información.

Creo que los datos históricos(datos que sigo recolectando) pueden ser de utilidad para el análisis del Transantiago, por parte de otros actores, por lo que estoy interesado en liberar el acceso a los datos históricos que estoy almacenando.

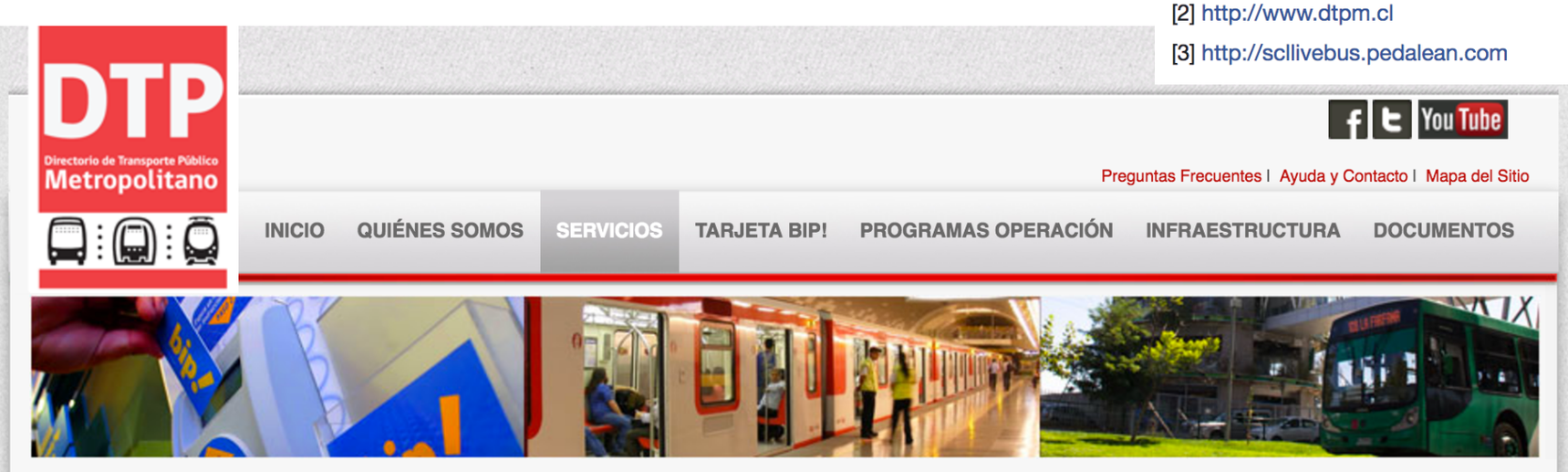
En [3] muestro la ubicaciones de todos lo buses del Transantiago.

Ejemplos de acceso en varios lenguajes en https://jpizarrom.github.io/…/2017.07.11-Santiago-Machine…/…

[1] http://www.dtpm.cl/in…/2013-04-24-14-09-09/datos-y-servicios

[2] http://www.dtpm.cl

[3] http://scllivebus.pedalean.com



ESTUDIO DE DESIGUALDADES

- https://www.desiguales.org/base-de-datos/

Encuesta

BASE DE DATOS Y METODOLOGÍA

BASE DE DATOS

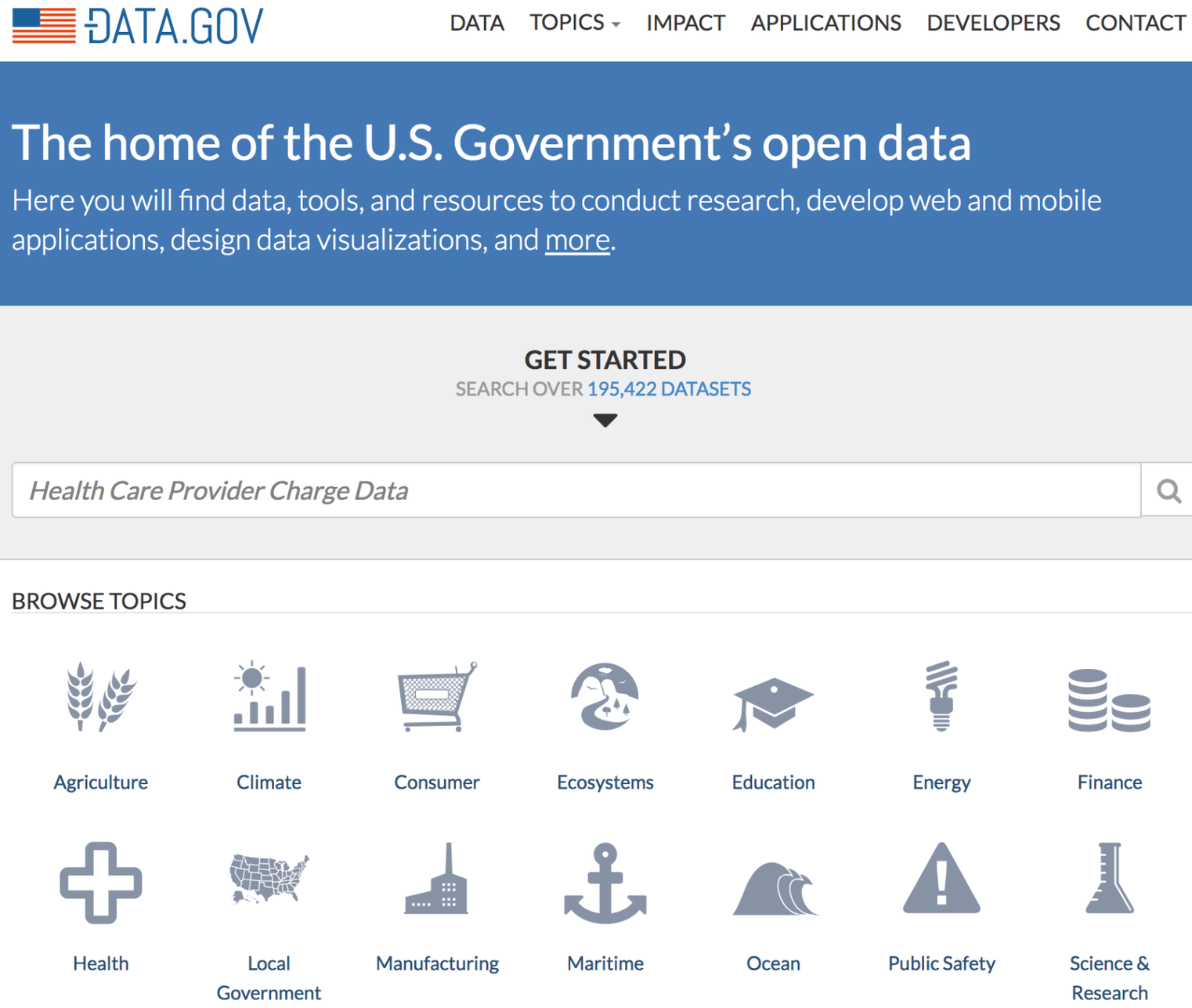
DESCARGAR DATOS (STATA, 7.6 MB)

DESCARGAR DATOS (SPSS, 1.8 MB)

DESCARGAR DATOS (CSV, 2.2 MB)

INFORMACIÓN GOBIERNO USA

- https://www.data.gov/



DATOS SOBRE CONFLICTOS ARMADOS

- https://www.prio.org/Data/Armed-Conflict/?id=348

PRIO

About PRIO | How To Find | Careers | Library | Contact | Intranet

News | Events | Research | Publications | People | Data | Education | Blogs | www.prio.org

Home > Data > Data on Armed Conflict

Data on Armed Conflict

CSCW and Uppsala Conflict Data Program (UCDP) at the Department of Peace and Conflict Research, Uppsala University, have collaborated in the production of a dataset of armed conflicts, both internal and external, in the period 1946 to the present. The Armed Conflict Dataset is primarily intended for academic use in statistical and macro-level research. It complements the annual compendium of ongoing armed conflicts published in the Journal of Peace Research, as well as the UCDP online database. CSCW houses the academic conflict dataset and continues to work closely with UCDP to provide more and better data.

UCDP/PRIO Armed Conflict Dataset
Download 1946–2008 armed conflict data, structured for quantitative analysis.

Data on Armed Conflict

- Data on Armed Conflict
- Urban Social Disorder v2
- GEO-SVAC Dataset
- Data on religious cleavages and civil war
- Conflict Site Dataset
- Onset and Duration of Intrastate Conflict
- ACLED - Armed Conflict Location and Event Data
- Battle Deaths Data
- UCDP/PRIO Armed Conflict Dataset

70 BASES DE DATOS EN LÍNEA QUE DEFINEN NUESTRO PLANETA

- https://www.technologyreview.com/s/421886/the-70-online-databases-that-define-our-planet/

A View from **Emerging Technology from the arXiv**

The 70 Online Databases that Define Our Planet

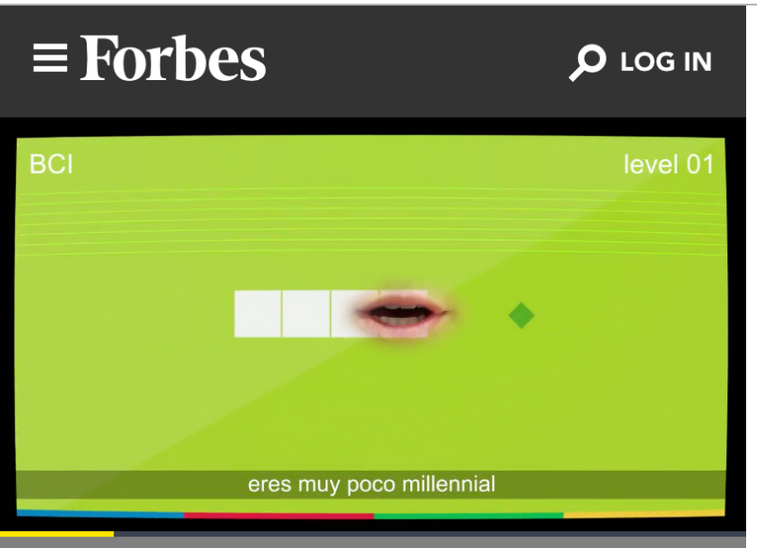
If you want to simulate the Earth, you'll need data on the climate, health, finance, economics, traffic and lots more. Here's where to find it.

December 3, 2010

Back in April, we looked at an ambitious European plan to simulate the entire planet. The idea is to exploit the huge amounts of data generated by financial markets, health records, social media and climate monitoring to model the planet's climate, societies and economy. The vision is that a system like this can help to understand and predict crises before they occur so that governments can take appropriate measures in advance.

33 FUENTES DE DATOS GRATUITAS

- https://www.forbes.com/sites/bernardmarr/2016/02/12/big-data-35-brilliant-and-free-data-sources-for-2016/#18d610aab54d



Tech / #BigData

FEB 12, 2016 @ 02:42 AM 213,595

Big Data: 33 Brilliant And Free Data Sources For

LISTADO DE BASES DE DATOS ABIERTAS/PÚBLICAS

- https://news.ycombinator.com/item?id=1493768

Hacker News new | comments | show | ask | jobs | submit login

▲ Ask HN: List of open/public databases

64 points by DanielBMarkham 2605 days ago | hide | past | web | 26 comments | favorite

A year or two ago somebody posted a list of open/free/accessible datasources for hackers to download and play around with. I thought this was a great resource, so I saved it, but heck if I can find it now.

Does anybody have such a list? Things like zip-codes for the US, locations of Starbucks stores, current weather forecasts, list of major newspapers, list of publicly-traded stocks, etc. I know there are tons of open/free databases waiting for us to mashup, just can't seem to find a list of them.

EDIT: The goal is a downloadable chunk of data to mashup, reformat, and use. That means CSV/XML/etc format and a public/anonymous FTP or something.

▲ randomtask 2605 days ago [-]

You might find the datasets/opendata sub-reddits interesting.

http://www.reddit.com/r/datasets http://www.reddit.com/r/opendata

▲ sidmitra 2605 days ago [-]

Delicious is a good place for the same. Eg.

http://delicious.com/sidmitra/datasets

▲ robin_reala 2605 days ago [-]

Freebase is a big one: http://www.freebase.com/ .

Also, if you're in the UK the government datasets might interest you: http://data.gov.uk/

▲ IgorPartola 2605 days ago [-]

http://themoviedb.org, http://thetvdb.com/ and my own API on top of these: http://igorpartola.com/projects/disciddb

▲ brown9-2 2605 days ago [-]

http://flowingdata.com/2009/10/01/30-resources-to-find-the-d...

US Govt data on the stimulus: http://www.recovery.gov/FAQ/Pages/DownloadCenter.aspx

ES MEJOR TENER UNA RESPUESTA APROXIMADA A LA PREGUNTA CORRECTA QUE UNA RESPUESTA EXACTA A LA PREGUNTA EQUIVOCADA

John W. Tukey