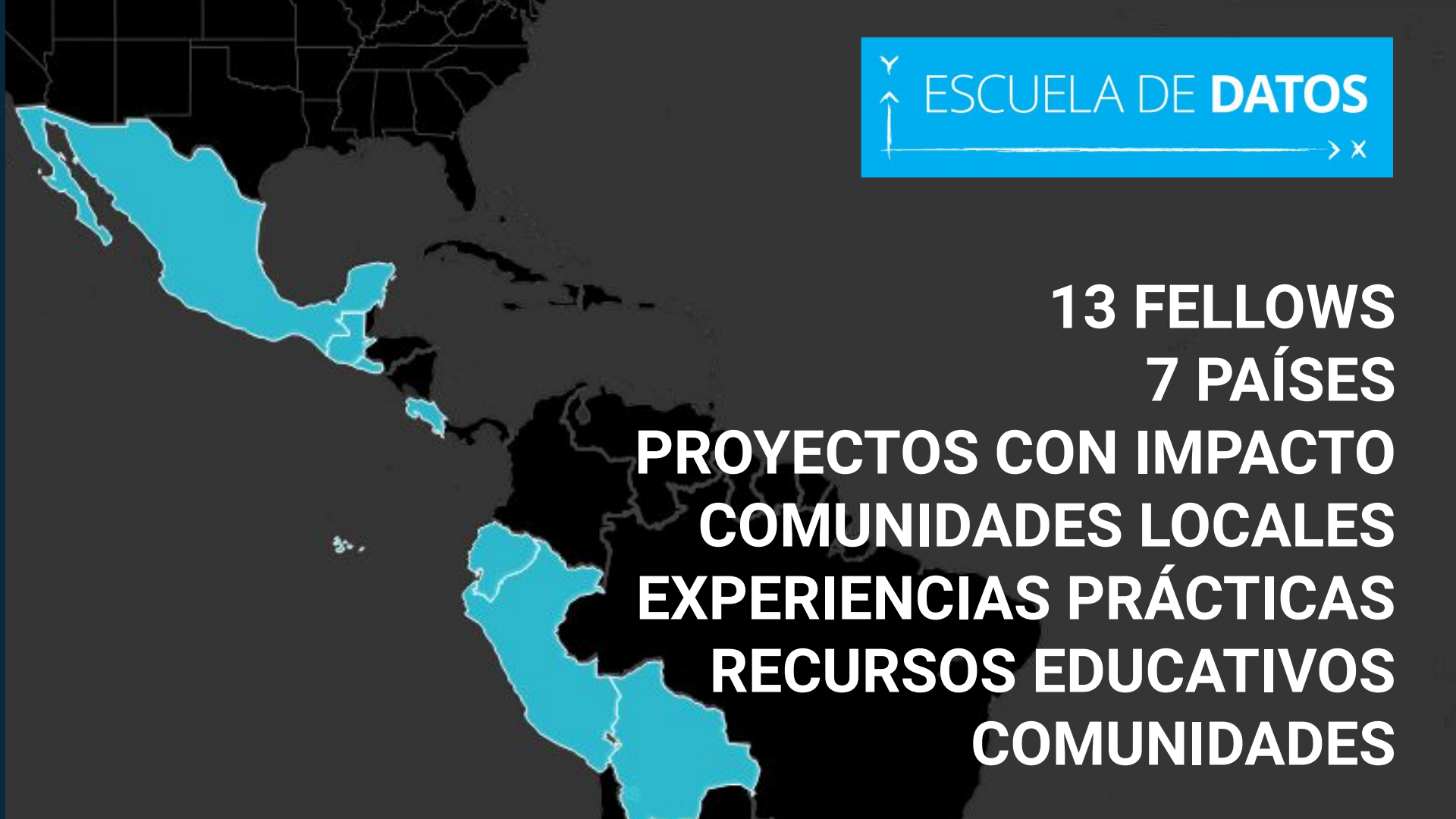
Aprovechamiento de los datos

APRENDIZAJE, HERRAMIENTAS Y EXPERIENCIAS PRÁCTICAS

ESCUELA DE **DATOS**

FORMACIÓN + INCIDENCIA = **IMPACTO**


ESCUELA DE DATOS
13 FELLOWS
7 PAÍSES
PROYECTOS CON IMPACTO
COMUNIDADES LOCALES
EXPERIENCIAS PRÁCTICAS
RECURSOS EDUCATIVOS
COMUNIDADES

Daniel Villatoro

Coordinador Escuela de Datos y Fellow 2016

@danyvillatoro

Pamela Gonzales

Fellow 2018
Bolivia

@10PAMELA20

Sofía Montenegro

Fellow 2018
Guatemala

@smontenegrom

APRENDIENDO A USAR DATOS A TRAVÉS DE

el DATA **PIPELINE**

LA METODOLOGÍA DE ESCUELA DE DATOS

El “DataPipeline”

- Proyectos basados en datos de principio a fin
- Trabajo en proceso enfocado hacia un impacto
- Capacidades básicas para usar datos en todo el proceso
- Adaptado a “nuestro” contexto Latinoamericano

¿Por qué una **tubería**?

Las tuberías, como conjunto tienen un propósito: llevar un flujo común. En este caso, ese flujo son los datos, pero a lo largo de la tubería y a través de diferentes procesos, estos datos se van transformando para llegar a un punto en el que no estaban antes.

Una tubería, además, es flexible. Es modular, lo permite adaptarla a

Definir.

Los proyectos guiados por datos deben empezar definiendo el problema que quieren resolver y sus acciones. Es en esta etapa te haces preguntas y llegas a los propósitos de tu proyecto. Definir tu problema implica pasar de un tema —contaminación ambiental, por ejemplo— a una o varias preguntas específicas —¿El uso de bicicletas ha reducido la contaminación del aire?

Ser específico te fuerza a formular tu pregunta de tal manera que provea pistas hacia los tipos de datos que necesitarás. Lo que te ayuda a definir la ambición de tu proyecto: ¿Los datos que necesito son fáciles de obtener? ¿O algunos datos principales serán difíciles de encontrar?

Datos con propósito

¿Qué información necesitas?

- estadísticas
- información histórica
- información pública
- testimonios
- análisis
- evidencia gráfica
- información geolocalizada
- ….

¿Qué harás con la información?

- informar
- sensibilizar
- visibilizar
- denunciar
- involucrar
- posicionar
- vigilar
- movilizar
- ….

¿Qué quieres lograr?

- participación ciudadana
- conciencia social
- acción social o gubernamental
- políticas públicas
- rendición de cuentas
- educación cívica
- ...

Ejemplo: Informar a consumidores

Organizaciones de la sociedad civil nos congratulamos ante la aprobación de la propuesta para un etiquetado de advertencia en productos ultraprocesados

25 julio, 2019 | : **Etiquetado de productos, En portada, Prensa, Salud nutricional**

https://elpoderdelconsumidor.org

Buscar.

Al tener una definición del problema, llegas a una idea de qué datos necesitarás, buscar estos datos es otro paso, con mucha o poca dificultad dependiendo del problema. Hay muchas herramientas y técnicas para hacer eso: desde una simple pregunta en tus redes sociales, hasta usar herramientas como un buscador, portales de datos abiertos o una solicitud de acceso a la información pidiendo datos que están disponibles en esa institución del gobierno. Esta fase puede ser definitiva para el éxito de tu proyecto, ya que si no encuentras los datos, no podrás hacer mayor cosa. Pero esta es la fase en la que también se necesita creatividad para actuar diferente.

¿DÓNDE BUSCAR?

Gobierno + Organizaciones + Academia

¿portales de datos abiertos?
vs
¿acceso restringido?

A googlear!

filetype:XLS
filetype:CSV
filetype:SHP
filetype:MDB
filetype:SQL
filetype:DB
filetype:PDF

inurl:donwloads
inurl:descargas
site:.gob
site:.gob.gt

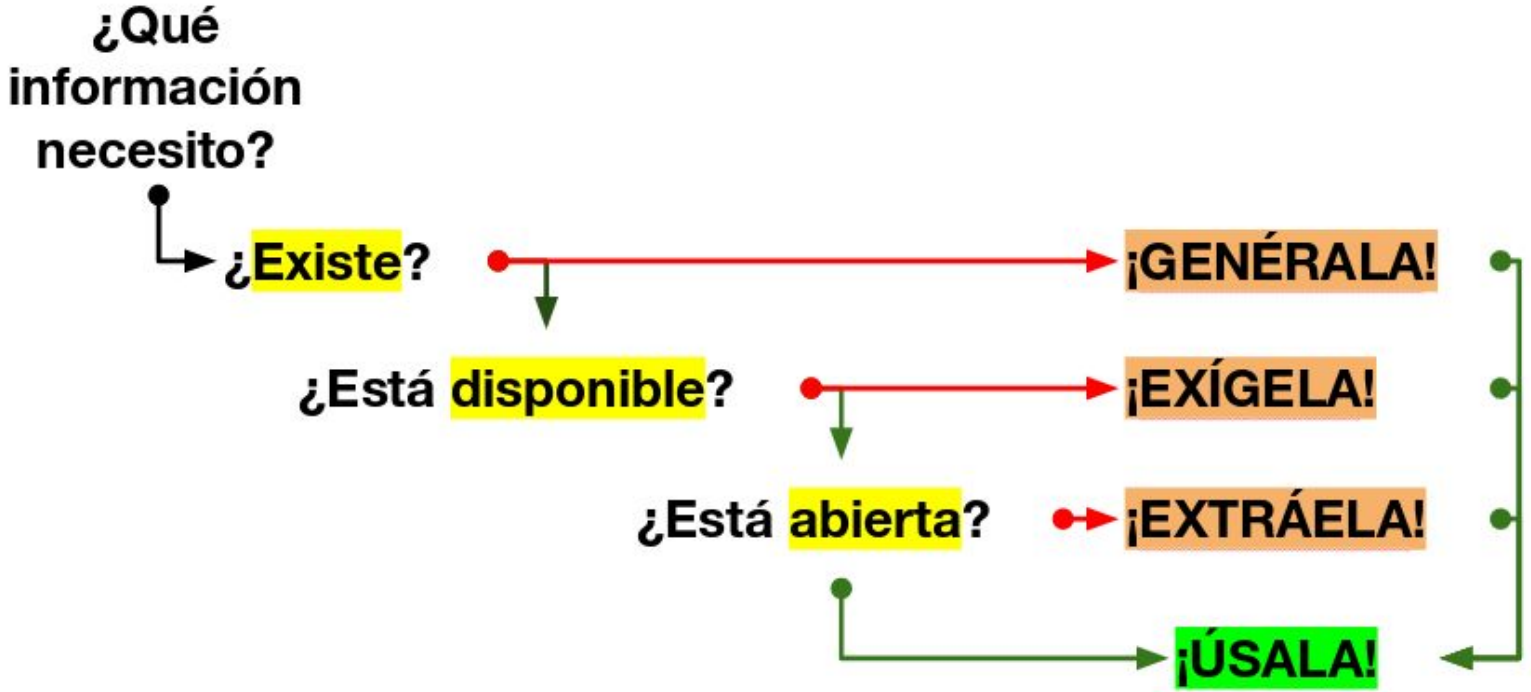
Ejemplo: Feminicidios en México

www.feminicidios.mx

Recolectar.

Hacer que los datos lleguen a tu computadora puede ser una tarea corta y fácil, o una larga y dolorosa. Hay muchas maneras de lograrlo. Se puede hacer crowdsourcing usando formularios en línea o encuestas,una recolección offline, scrapear una página web o simplemente descargar conjuntos de datos de los sitios del gobierno -usando portales de datos o a través de una solicitud de acceso a información-.

Ir por los datos



Generación de datos

Construir mis propios datos

Tú administras una encuesta, haces llamadas, vas llenando una base de datos, vas obteniendo datos manualmente

Puedes utilizar:

- Formularios online y offline

- Aplicaciones para levantar datos: preguntas, imágenes, audio, video y geolocalización

Ejemplo: Reconstrucción post-sismo

https://socialtic.org/blog/levantamiento-comunitario-de-datos-para-impacto-social/

Datos desde las expresiones en línea

Crowdsourcing

Las personas van dando sus datos en plataformas públicas, redes sociales o mecanismos de obtención de datos

Mecanismos comunes para crowdsourcing:

- Apps y formularios de reporte o denuncia

- Acceso a datos de redes sociales

- Extracción de datos de plataformas y espacios de comentarios y expresión ciudadana

Ejemplo: Reportes ciudadanos

http://www.cic.mx/

Solicitar datos públicos

Solicitudes de información pública

La Ley te faculta para pedirle información a entes públicos y cualquier privado que administre fondos públicos

Recomendaciones:

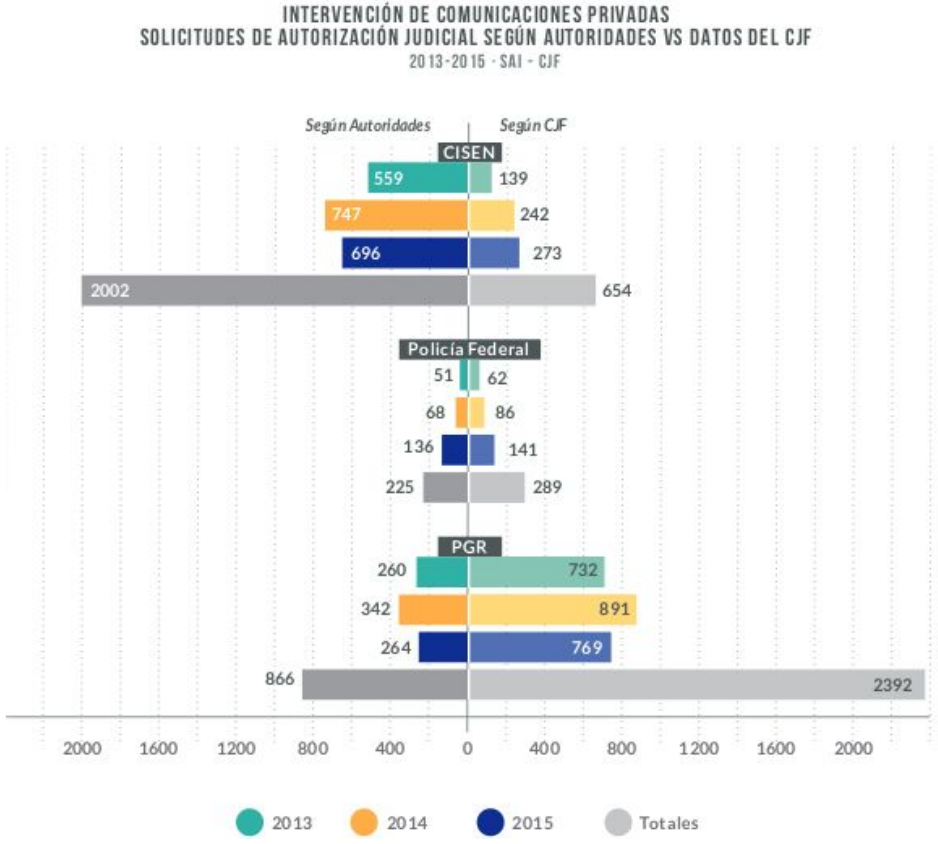
- Identificar área exacta de gobierno que la genera

- Solicitar información puntual de acuerdo a la nomenclatura identificada en contexto de gobierno

- Solicitar formatos abiertos para entrega

- Apalancar en información otorgada o abierta

Ejemplo: Vigilancia Estatal en México



https://r3d.mx

Extracción (“scraping”) de datos

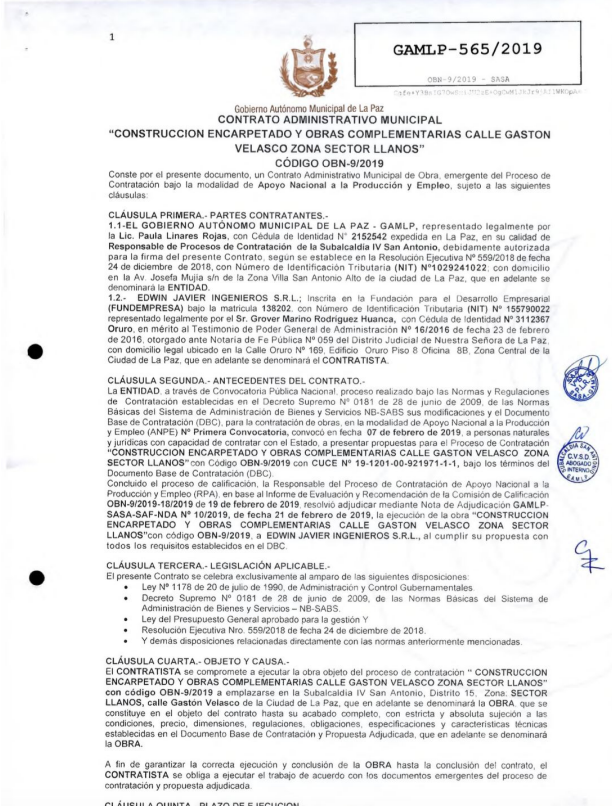
Scraping, rascado o extracción automática

La información está en la web, pero no está junta ni sistematizada. Con herramientas y código la extraes para tu uso.

Puedes extraer datos desde:

- Papel … con personas voluntarias :S
- Imágenes utilizando OCR
- PDFs
- Sitios web / HTML

Ejemplo: Extracción desde imagen



GAMLP-565/2019

OBN-9/2019 - SASA

Gobierno Autónomo Municipal de La Paz

CONTRATO ADMINISTRATIVO MUNICIPAL

"CONSTRUCCION ENCARPETADO Y OBRAS COMPLEMENTARIAS CALLE GASTON VELASCO ZONA SECTOR LLANOS"

CÓDIGO OBN-9/2019

Conste por el presente documento, un Contrato Administrativo Municipal de Obra, emergente del Proceso de Contratación bajo la modalidad de **Apoyo Nacional a la Producción y Empleo**, sujeto a las siguientes cláusulas:

CLÁUSULA PRIMERA.- PARTES CONTRATANTES.-

1.1-EL GOBIERNO AUTÓNOMO MUNICIPAL DE LA PAZ - GAMLP, representado legalmente por la **Lic. Paula Linares Rojas**, con Cédula de Identidad N° **2152542** expedida en La Paz, en su calidad de **Responsable de Procesos de Contratación de la Subalcaldia IV San Antonio**, debidamente autorizada para la firma del presente Contrato, según se establece en la Resolución Ejecutiva Nº 559/2018 de fecha 24 de diciembre de 2018, con Número de Identificación Tributaria **(NIT) Nº1029241022**, con domicilio en la Av. Josefa Mujia s/n de la Zona Villa San Antonio Alto de la ciudad de La Paz, que en adelante se denominará la **ENTIDAD.**

1.2.- EDWIN JAVIER INGENIEROS S.R.L.; inscrita en la Fundación para el Desarrollo Empresarial **(FUNDEMPRESA)** bajo la matricula **138202**, con Número de Identificación Tributaria **(NIT) Nº 155790022** representado legalmente por el **Sr. Grover Marino Rodriguez Huanca,** con Cédula de Identidad **Nº 3112367 Oruro**, en mérito al Testimonio de Poder General de Administración **Nº 16/2016** de fecha 23 de febrero de 2016, otorgado ante Notaria de Fe Pública Nº 059 del Distrito Judicial de Nuestra Señora de La Paz, con domicilio legal ubicado en la Calle Oruro Nº 169, Edificio Oruro Piso 8 Oficina 8B, Zona Central de la Ciudad de La Paz, que en adelante se denominará el **CONTRATISTA.**

CLÁUSULA SEGUNDA.- ANTECEDENTES DEL CONTRATO.-

La **ENTIDAD**, a través de Convocatoria Pública Nacional, proceso realizado bajo las Normas y Regulaciones de Contratación establecidas en el Decreto Supremo Nº 0181 de 28 de junio de 2009, de las Normas Básicas del Sistema de Administración de Bienes y Servicios NB-SABS sus modificaciones y el Documento Base de Contratación (DBC), para la contratación de obras, en la modalidad de Apoyo Nacional a la Producción y Empleo (ANPE) **Nº Primera Convocatoria**, convocó en fecha **07 de febrero de 2019**, a personas naturales y jurídicas con capacidad de contratar con el Estado, a presentar propuestas para el Proceso de Contratación **"CONSTRUCCION ENCARPETADO Y OBRAS COMPLEMENTARIAS CALLE GASTON VELASCO ZONA SECTOR LLANOS"** con Código **OBN-9/2019** con **CUCE Nº 19-1201-00-921971-1-1**, bajo los términos del Documento Base de Contratación (DBC).

Concluido el proceso de calificación, la Responsable del Proceso de Contratación de Apoyo Nacional a la Producción y Empleo (RPA), en base al Informe de Evaluación y Recomendación de la Comisión de Calificación **OBN-9/2019-18/2019** de **19 de febrero de 2019**, resolvió adjudicar mediante Nota de Adjudicación **GAMLP-SASA-SAF-NDA Nº 10/2019, de fecha 21 de febrero de 2019**, la ejecución de la obra **"CONSTRUCCION ENCARPETADO Y OBRAS COMPLEMENTARIAS CALLE GASTON VELASCO ZONA SECTOR LLANOS"**con código **OBN-9/2019**, a **EDWIN JAVIER INGENIEROS S.R.L.**, al cumplir su propuesta con todos los requisitos establecidos en el DBC.

CLÁUSULA TERCERA.- LEGISLACIÓN APLICABLE.-

El presente Contrato se celebra exclusivamente al amparo de las siguientes disposiciones:

- Ley Nº 1178 de 20 de julio de 1990, de Administración y Control Gubernamentales.
- Decreto Supremo Nº 0181 de 28 de junio de 2009, de las Normas Básicas del Sistema de Administración de Bienes y Servicios – NB-SABS.
- Ley del Presupuesto General aprobado para la gestión Y
- Resolución Ejecutiva Nro. 559/2018 de fecha 24 de diciembre de 2018.
- Y demás disposiciones relacionadas directamente con las normas anteriormente mencionadas.

CLÁUSULA CUARTA.- OBJETO Y CAUSA.-

El **CONTRATISTA** se compromete a ejecutar la obra objeto del proceso de contratación " **CONSTRUCCION ENCARPETADO Y OBRAS COMPLEMENTARIAS CALLE GASTON VELASCO ZONA SECTOR LLANOS" con código OBN-9/2019** a emplazarse en la Subalcaldia IV San Antonio, Distrito 15, Zona: **SECTOR LLANOS, calle Gaston Velasco** de la Ciudad de La Paz, que en adelante se denominará la **OBRA**, que se constituye en el objeto del contrato hasta su acabado completo, con estricta y absoluta sujeción a las condiciones, precio, dimensiones, regulaciones, obligaciones, especificaciones y características técnicas establecidas en el Documento Base de Contratación y Propuesta Adjudicada, que en adelante se denominará la **OBRA.**

A fin de garantizar la correcta ejecución y conclusión de la **OBRA** hasta la conclusión del contrato, el **CONTRATISTA** se obliga a ejecutar el trabajo de acuerdo con los documentos emergentes del proceso de contratación y propuesta adjudicada.

Result

Images | Text | Fonts | Metadata

Only the first 100 lines of extracted text are shown. Use the down
computer.

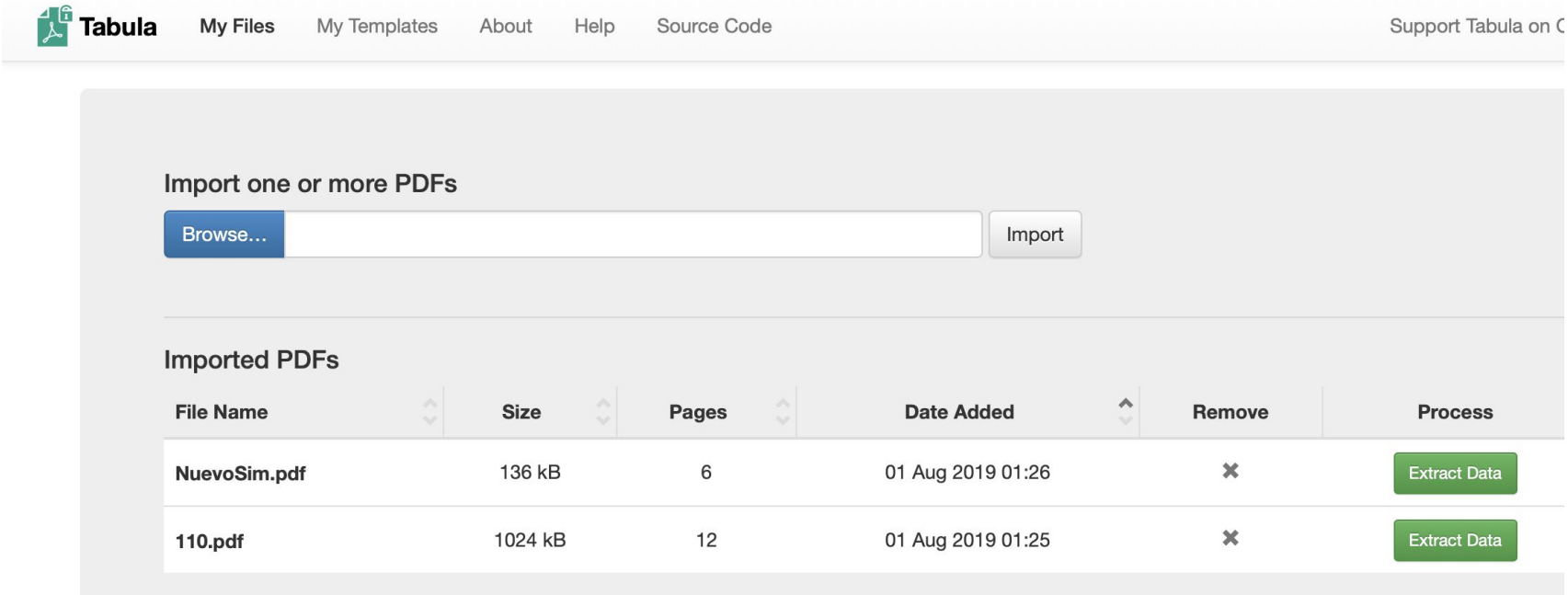
Download result as a file

to d o s los requisitos establecidos en el DBC.
CLÁU S U LA TER C ER A .- LEGISLACIÓN A P LIC A B LE .El presente Contrato se celebra exclusivar
•
Ley N° 1178 de 20 de julio de 1990, de Administración y Control Gubernamentales.
•
Decreto Supremo N° 0181 de 28 de junio de 2009, de las Normas Básicas de! Sistema de
Adm inistración de Bienes y Servicios - NB-SABS.
•
Ley del Presupuesto General aprobado para la gestión Y
•
Resolución Ejecutiva Nro. 559/2018 de fecha 24 de diciembre de 2018.
® Y dem ás disposiciones relacionadas directamente con las normas anteriorm ente mencionadas.
C LAU SULA C U A R TA .- O BJETO Y CAUSA.El C O N TR ATISTA se com prom ete a ejecutar la obr
ENCARPETADO Y O BRAS COMPLEMENTARIAS C ALLE GASTON V E LA S C O ZO N A SECTOR LLAI
con c ó d ig o O B N -9 /2 0 1 9 a e m p la za rse en la Subalcaldía IV San A n to n io , D istrito 15, Zon

Back to start

https://www.extractpdf.com/

Ejemplo: Extracción desde PDF



Tabula My Files My Templates About Help Source Code

judging-rubric.pdf | Templates | Clear All Selections | Autodetect Tables

1

2

5 pts	Each component of the team submission is well thought out.
5 pts	The team's strong dedication and work ethic is clear, even if the submission is not complete.
5 pts	The way the team approaches and solves the problem is unique.
5 pts	The submission stands out from others.
	Overall Impression Total Score:

Total Score	
Ideation Total Score	
Technical Total Score	
Entrepreneurship Total Score*	
Pitch Total Score	
Overall Impression Total Score	
Team Submission Total Score	

*Senior Division only

Is the extracted data incorrect?

You can revise your selected cells or try an alternate extraction method.

Revise Selected Cells

Data has been extracted from the cells you selected in the previous step. You can revise your selection(s) to add or remove cells.

← Revise selection(s)

Choose Alternate Extraction Method

The current preview uses the **Stream** extraction method. If the data is not mapped to the correct cells, try the **Lattice** method instead.

judging-rubric.pdf **Export Forma** ✓ CSV TSV JSON (dimensions) JSON (data) zip of CSVs Script Export Copy to Clipboard

Preview of Extr ar Data

Total S
Ideation Total Score
Technical Total Score
Entrepreneurship Total Score*
18 Pitch Total Score
Overall Impression Total Score
Team Submission Total Score

Verificar.

Obtener los datos no significa que el problema está resuelto. Es necesario verificar si su información es válida, así como revisar los metadatos y la metodología con la que se recolectó este conjunto de información. Es importante también conocer quién organizó este conjunto de datos y si es una fuente con credibilidad en el tema y en la técnica de recolección.

¿Esta información es la que necesito?

¿Tengo la información completa?

¿En qué contexto están estos datos?

¿Están en un formato que me permita analizarlos?

¿Estos datos responden a mis preguntas iniciales?

¿Son datos oficiales?

¿La fuente es confiable?

¿Cuál fue la metodología para producirlos?

Ejemplo: Verificación de noticias

https://es.schoolofdata.org/tag/fact-checking-la-fiebre-que-se-contagia-por-toda-latinoamerica/

chequeado

MÉTODO DE VERIFICACIÓN DEL DEBATE PÚBLICO

Ocho pasos para un buen chequeo:

1- Seleccionar una frase del ámbito público

2- Ponderar su relevancia

3- Consultar a la fuente original

4- Consultar a la fuente oficial

5- Consultar a fuentes alternativas

6- Ubicar en contexto

7- Confirmar, relativizar o desmentir la afirmación

8- Calificar

ConPruebas es un ejercicio periodístico de verificación del discurso público

En efecto, los datos respaldan el enunciado.

Todos los datos e información indican que el enunciado no es real.

Algunos datos son verdaderos, pero no todos.

No hay documentos que prueben que el enunciado es correcto.

Limpiar.

Es muy común que los datos que se obtienen y validan estén en desorden y tengan problemas de formato: filas duplicadas, nombres de columna que no combinan con los registros, valores que contienen caracteres raros o que impiden el procesamiento de la computadora y otros más. En este paso, necesitamos habilidades y herramientas que nos permitan tener los datos en un formato legible para analizarlo por computadora. Herramientas como OpenRefine, LibreOffice Calc o Excel y conceptos como bases de datos relacionales son útiles en esta etapa.

¿Porqué hay que limpiar datos?

País			
Estados Unidos	122	394	281
Estados Unidos de America	500	200	300
Estados Unidos,	400	200	124
Unión Soviética	280	39	76
Union Soviética	110	230	140
Union Sovietica	5	50	80

País			
Estados Unidos	1022	794	705
Unión Soviética	395	319	296

¿Cuándo hay que limpiar datos?

La ortografía es inconsistente

Los dedazos o errores “de dedo” en palabras o números

Espacios o caracteres inconsistentes

Formatos de fecha inconsistentes

Texto que fue convertido a números

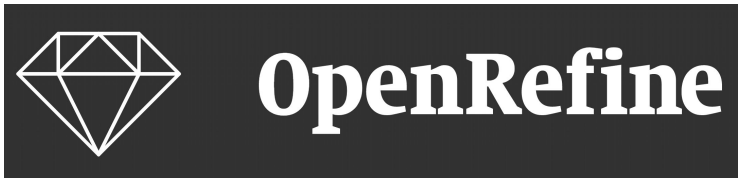
Números que fueron guardados como texto

Extracción (“scraping”) de datos

Limpieza sencilla desde hojas de cálculo

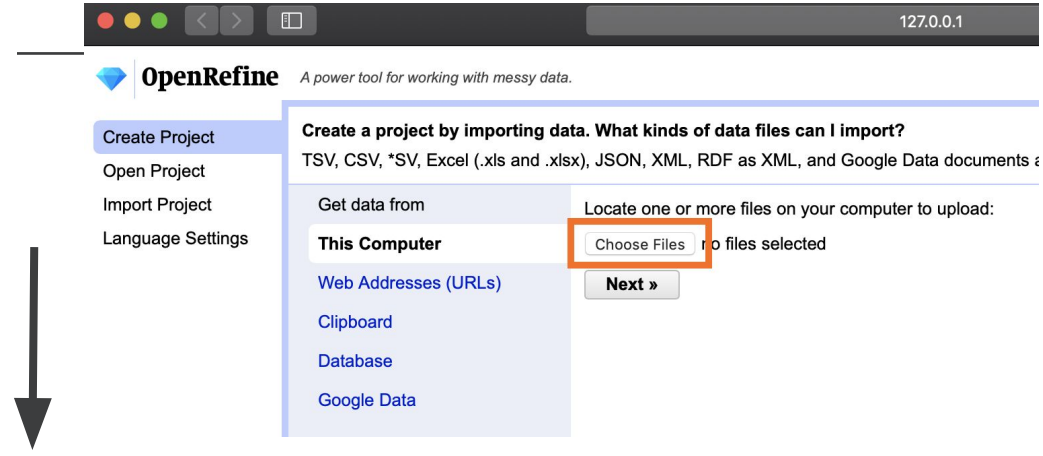
> Funciones básica tales como buscar, reemplazar, filtros y formatos de celdas

Limpieza de grandes volúmenes de datos



DataWrangler alpha

Ejemplo:



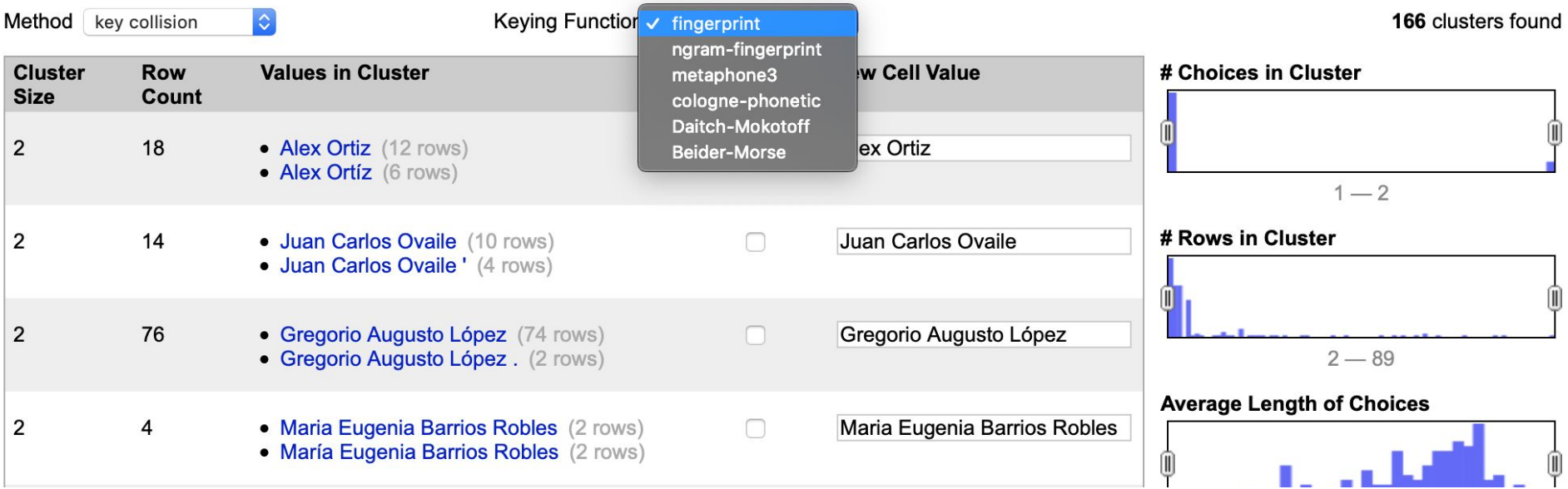
1236 records Extensions: Wikidata

Show as: rows **records** Show: 5 **10** 25 50 records « first ‹ previous **1 - 10** next › last »

All		«iOJ	NOMBRE	NIT / DPI	AFILIADO	SIMPATIZANTE	CUENTA BANCA	No. DE RECIBO	fecha1	fecha2
1.										
2.	1	Ileana Guadalupe Calles Domínguez		X	■ í:		327	0.00	50.00	
3.	2	Milton Francisco Sánchez Cuellar		X			320/325	200,00	200.00	
4.	3	Verónica Liseth del Alba Crispin M,		X			328	100,00	0.00	
5.	4	Maria Eugenia Sarrios Robles		x			329	: 0,00	100.00	
6.	1	Egar Justino Ovalle Maído nado		x	1 323		800.00	0,00	0.00	
7.	2	Jaime Antonio de la Peña		X	336		0.00	100.00	0.00	
8.	3	Sergio Raúl Franco Sagastume		X	321		100.00	0.00	0.00	
9.	4	Gregorio Augusto López		X	326		100.00	0.00	0.00	
10.	5	Milton Francisco Sánchez Cuellar j		X	335		0.00	0.00	200.00	

Fingerprint

Quita todos los espacios en blanco, cambia todos los caracteres a minúsculas, remueve toda la puntuación y normaliza cualquier carácter especial a una versión estándar. Luego, parte el texto y aplica espacios en blanco. Así encuentra las coincidencias.



https://es.schoolofdata.org/2018/05/03/algoritmos-y-clusters-encuentra-errores-y-limpialos-de-manera-facil-con-openrefine/

Analizar.

Esta es la parte en la que obtenemos conocimiento sobre el problema que definimos al principio. Al poner en práctica nuestras habilidades estadísticas y matemáticas, podemos entrevistar un conjunto de datos como cualquier periodista entrevista a sus fuentes. Solo que en vez de usar una grabadora y una libreta, analizamos con muchas herramientas y habilidades. Podemos generar visualizaciones que nos muestren la distribución de diferentes variables o podemos usar paquetes de lenguajes de programación como Pandas (Python) o R. También podemos usar hojas de cálculo como LibreOfficeCalc y Excel, o programas estadísticos como SPSS.

Tipos de análisis de datos

Análisis descriptivo

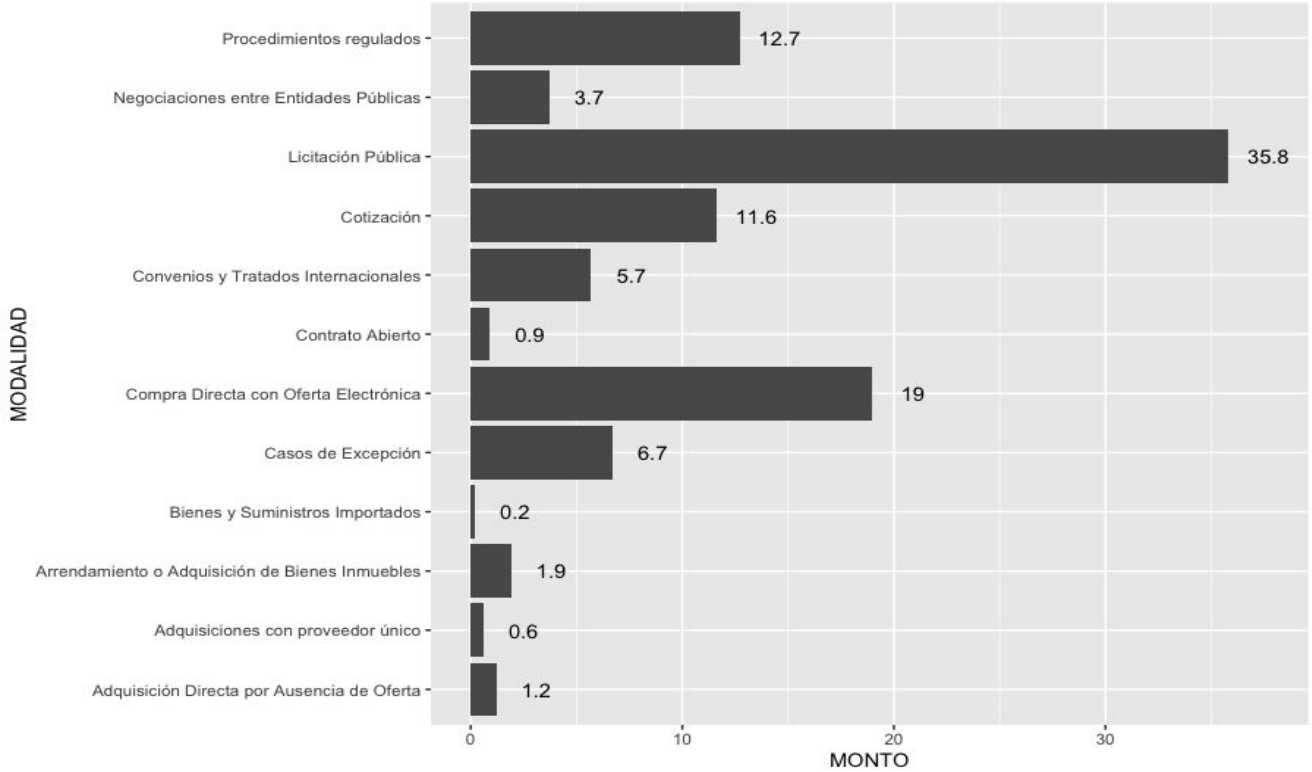
Medidas de tendencia central

- Promedio
- Moda
- Mediana
- Desviación estándar

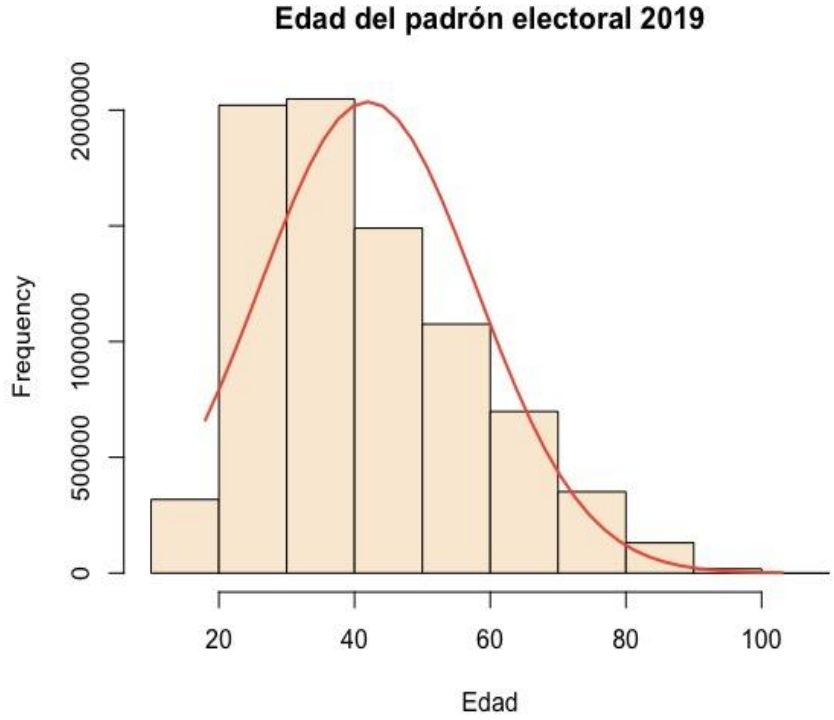
Análisis inferencial

- Prueba de hipótesis
- Intervalos de confianza
- Correlación y causalidad
- Intro a probabilidad
- Outlayers

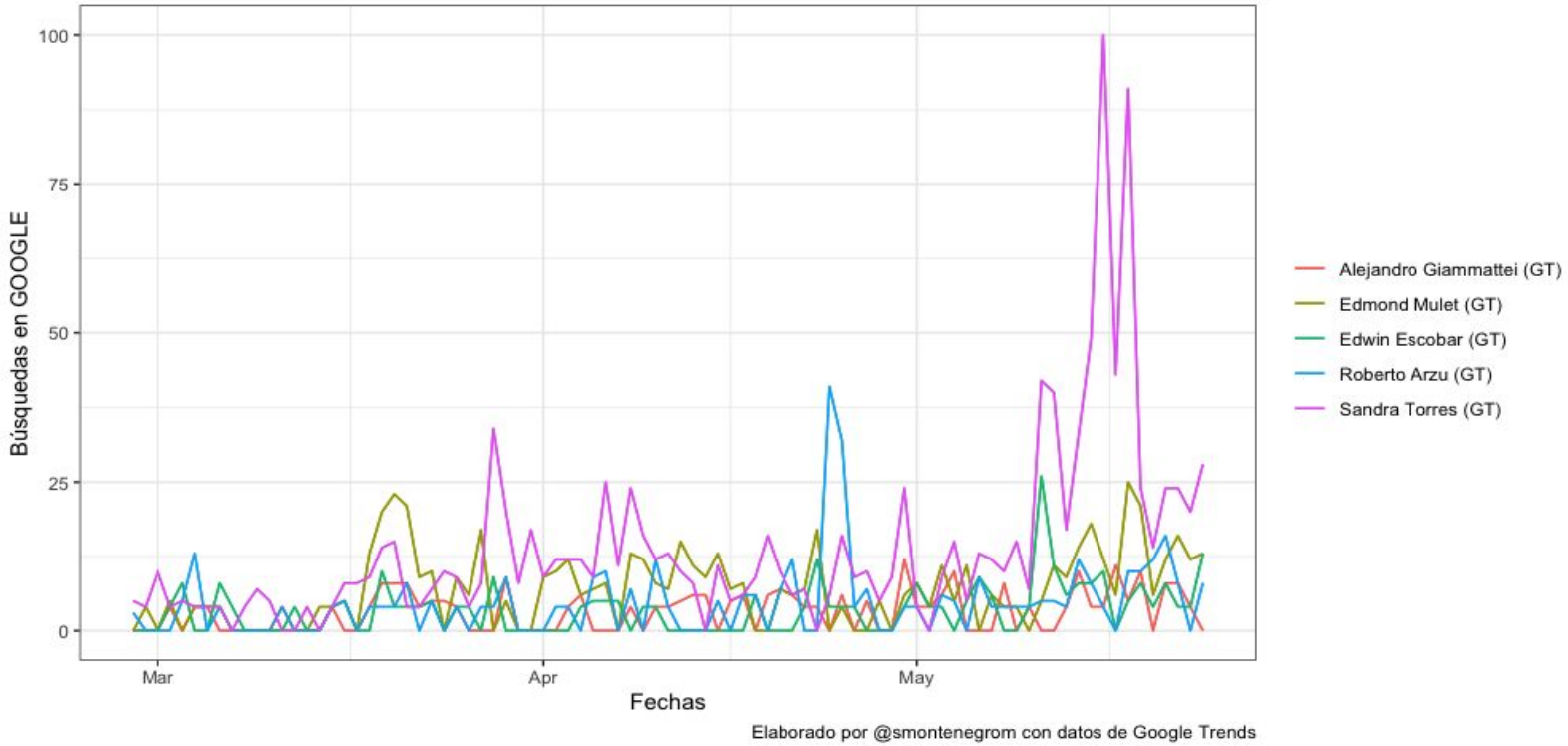
¿Cómo se ve el análisis de datos?



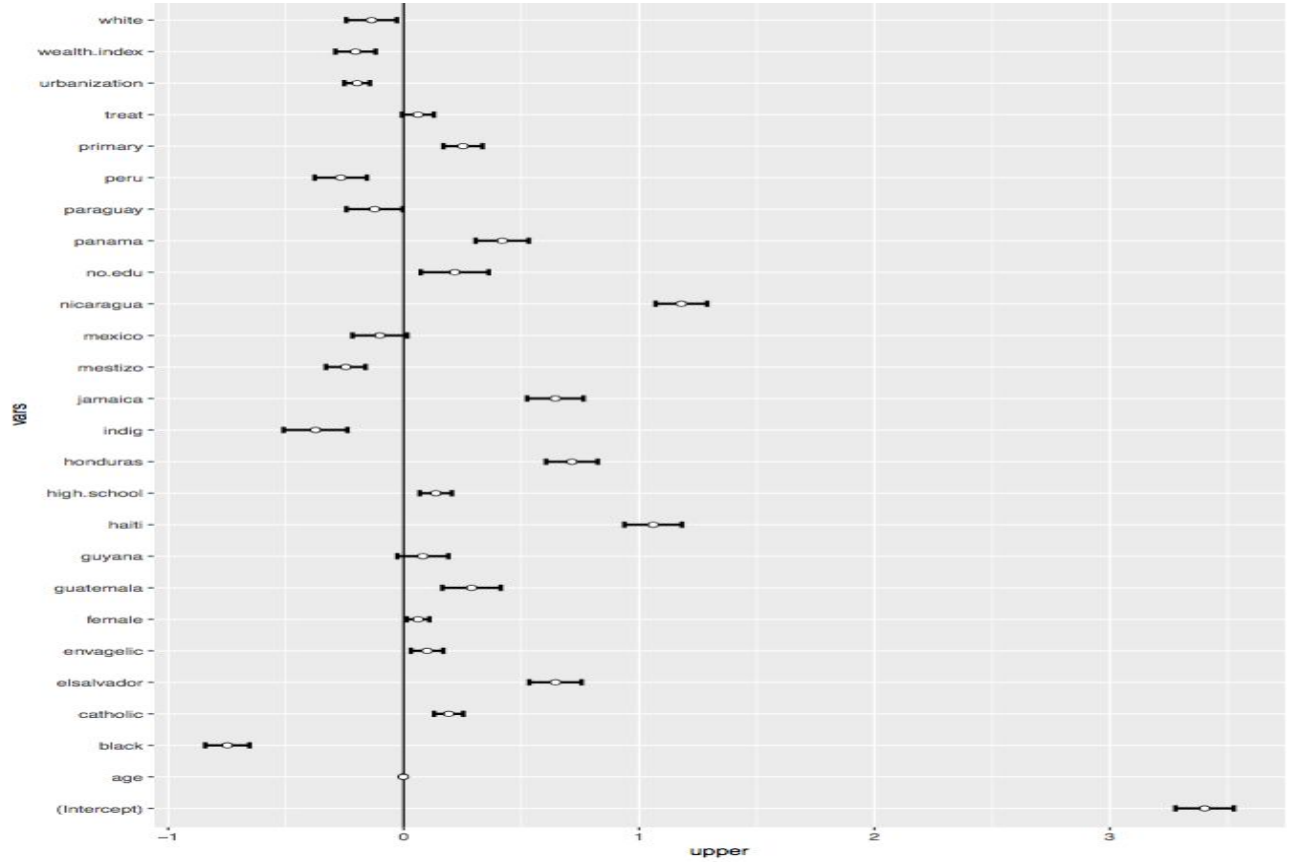
¿Cómo se ve el análisis de datos?



¿Cómo se ve el análisis de datos?



¿Cómo se ve el análisis de datos?



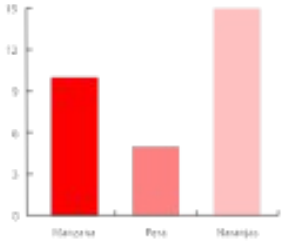
Presentar.

Es necesario presentar los datos: hablar con tu audiencia para que conozca las preguntas que buscabas responder y el medio que te ha permitido llegar a ciertas conclusiones o iniciar una conversación. En esta etapa debemos enfocarnos en entender buenas prácticas para presentar los datos de manera visual y sabernos dar a entender.

Visualización vs. Narrativas

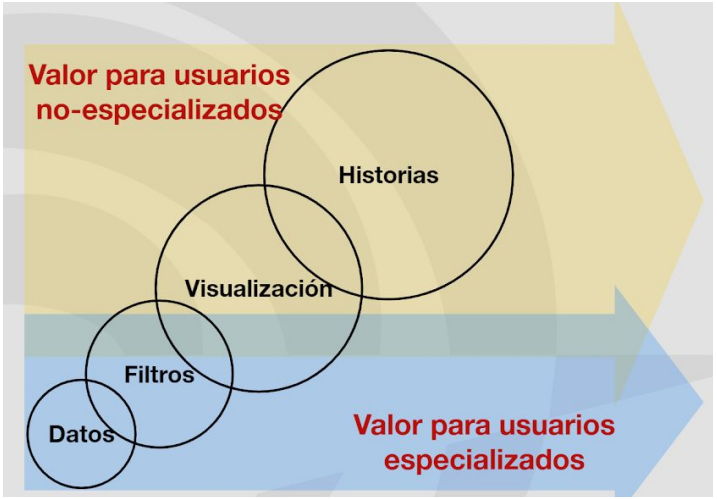
Visualización

Una gráfica, un diagrama

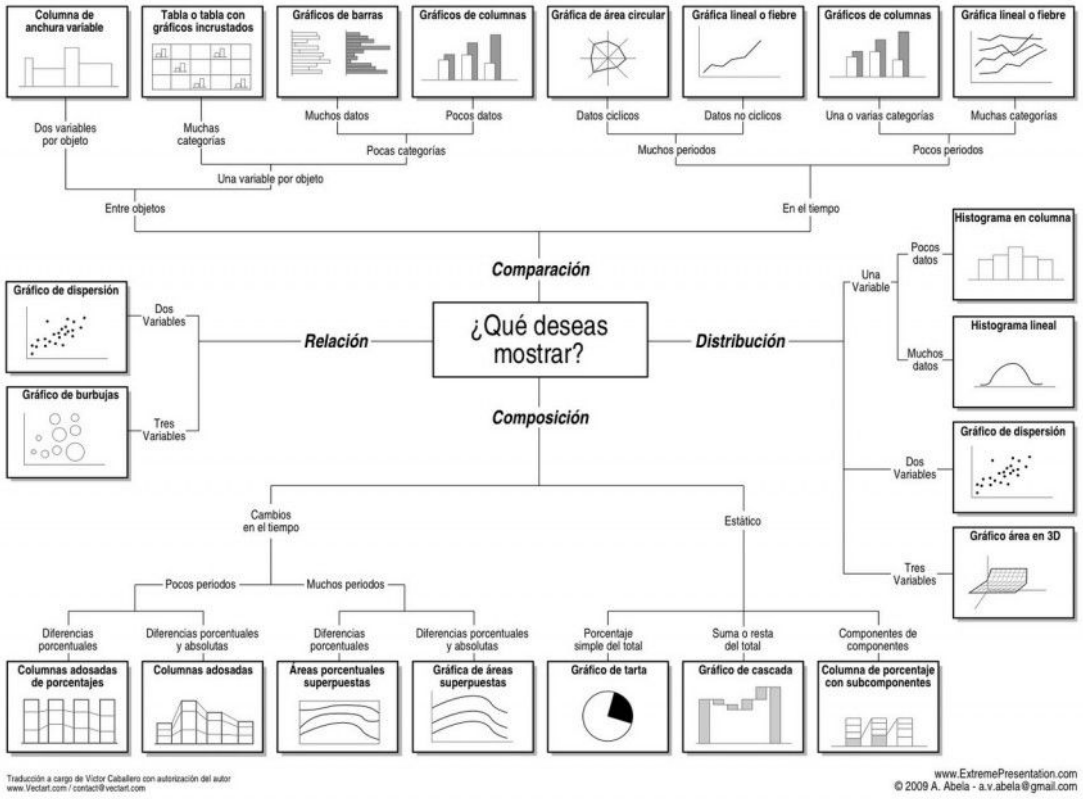


Narrativas

Contar una historia
Hacer evidente un mensaje
Contextualizar
Llevar de lo general a lo particular



¿Qué gráfica usar?



¿Qué se desea mostrar?

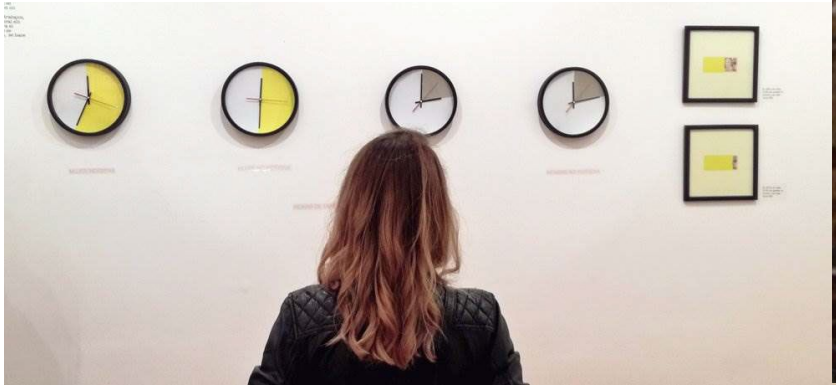
Ejemplo: Contextualización del dato

www.pictoline.com

Ejemplo: Data Art

brecha

Un proyecto para visibilizar la desigualdad entre sexos



https://www.behance.net/gallery/55950059/brecha-instalacion-y-visualizacion-de-datos

Herramientas para visualizar datos

Tableau - https://www.tableau.com/es-es
Flourish - https://flourish.studio
Infogr.am - https://infogram.com
Piktochart - https://piktochart.com
Plot.ly - https://plot.ly
KnightLab - https://knightlab.northwestern.edu/projects
Cómo elegir la mejor gráfica - https://datavizcatalogue.com/

Elementos a considerar

Que lo que digo esté fundamentado

Que las personas puedan leerlo

Que conecte con ellos en una dimension emocional

¡Gracias!

@**Escuela**De**Datos**