**Iaac** | Institute for advanced architecture of Catalonia | BARCELONA

**MASTER IN CITY & TECHNOLOGY**
**DIGITAL TOOLS AND BIG DATA**
**2020/2021**

**FACULTY** DIEGO PAJARITO

# Big Data

**Where is all that oil?**

**2017**

Regulating the internet giants

# The world's most valuable resource is no longer oil, but data

*The data economy demands a new approach to antitrust rules*

📖 Print edition | Leaders ›
May 6th 2017

A NEW commodity spawns a lucrative, fast-growing industry, prompting antitrust regulators to step in to restrain those who control its flow. A century ago, the resource in question was oil. Now similar concerns are being raised by the giants that deal in data, the oil of the digital era. These titans—Alphabet (Google's parent company), Amazon, Apple, Facebook and Microsoft—look unstoppable. They are the five most valuable listed firms in the world. Their profits are surging; they collectively racked up over

https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data

**2019**

# Here's Why Data Is Not The New Oil

**Bernard Marr** Contributor ⓘ
Enterprise & Cloud

It's a claim you've probably heard multiple times – "Data is the new oil!"

Now it's true, that in some ways, the analogy fits – it's easy to draw parallels due to the way information (data) is used to power much of the transformative technology we see today – artificial intelligence, automation and advanced, predictive analytics.
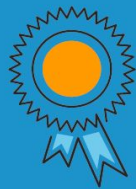
Adobe Stock   ADOBE STOCK

However, in many ways, it's also lazy and inaccurate – and while it's handy as a marketing shorthand (because it gets across the fact that data is a valuable commodity with many different uses across many applications) it's

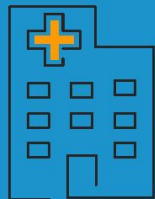https://www.forbes.com/sites/bernardmarr/2018/03/05/heres-why-data-is-not-the-new-oil/#5b2ecb5d3aa9

**Buzzword and common perception. Is it or not?**

more Open Data can help make
**better decisions**

**2,549 hours**
**wasted**
finding parking

**7,000 lives**
saved due to
quicker response

**629 million**
hours saved is
equivalent to
**€ 27.9 bn**

**5.5% less**
road fatalities

Congestion
**costs are**
**1% of GDP**

**16% less**
less energy used
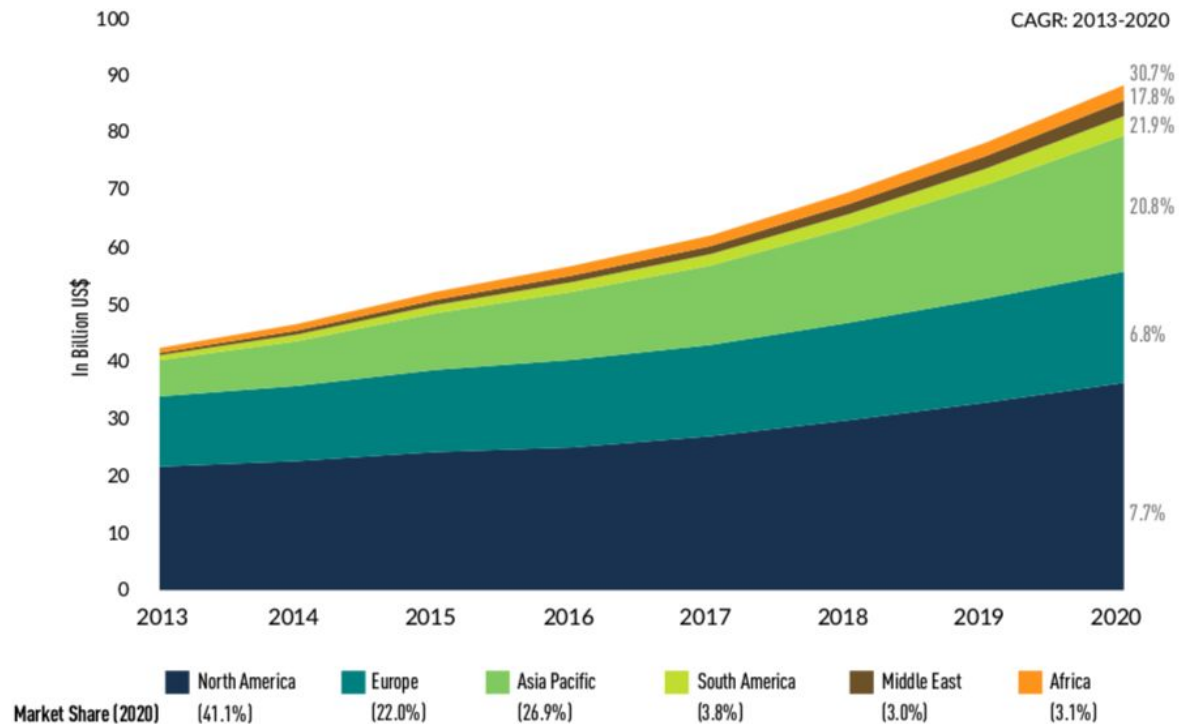
# Do we need private data?

Text describing open data

**Word cloud**

**Python library**

https://python-graph-gallery.com/260-basic-wordcloud/



Iaac | Institute for advanced architecture of Catalonia   **Do it yourself. Get some data and make it appealing**

**Global**
Global open data index index.okfn.org/
Natural earth www.naturalearthdata.com/
UN data (32 datamarts) data.un.org/
OECD (6k) data.oecd.org/
Global migration data analysis centre gmdac.iom.int/

**Regional**
EU portal (939k) www.europeandataportal.eu
Africa Geoportal www.africageoportal.com/
Inspire inspire-geoportal.ec.europa.eu/

**National**
USA (236k) www.data.gov/
Spain https://datos.gob.es/ www.idee.es/
UK (52k) data.gov.uk/
Ecuador (128) catalogo.datosabiertos.gob.ec

**Local**
Barcelona (452) opendata-ajuntament.barcelona.cat www.bcn.cat/geoportal
NY opendata.cityofnewyork.us/
Amsterdam data.amsterdam.nl/
Medellin www.medellin.gov.co/geomedellin/

Open Data Portals, a pretty small selection of them

Wikipedia / DBpedia wiki.dbpedia.org/
Wikidata www.wikidata.org/
OSM wiki.openstreetmap.org/
Geonames https://www.geonames.org/
Zenodo (40k) zenodo.org/

Uber movement.uber.com
Here student sample data
Waymo waymo.com/open
City Bike www.citibikenyc.com/system-data

# Technical

Connection, software, hardware, personal skills, general knowledge

# Legal / Licensing

License agreements, terms of use, national security or privacy restrictions.

# Pricing

From free to extremely expensive

# Availability

In a readable format, means and access point, offered by producer or custodian

# Scale

Spatial resolution (x,y), temporal (hms), value domains and coverage

# Time

Covering the time period needed for the analysis

Google places [developers.google.com/places](developers.google.com/places)

Twitter [developer.twitter.com/](developer.twitter.com/)

Yelp [www.yelp.com/developers/](www.yelp.com/developers/)

Open weather map [https://openweathermap.org/api](https://openweathermap.org/api)

Open Cage [https://opencagedata.com/](https://opencagedata.com/)

Tomtom https://www.tomtommaps.com/mapdata/

Bing Maps https://blogs.bing.com/maps

ESRI https://www.esri.com/

Census Europe https://ec.europa.eu/CensusHub2

Copernicus https://scihub.copernicus.eu/

US census data https://www.census.gov

Humanitarian OSM [tasks.hotosm.org/](tasks.hotosm.org/)

Wikipedia [www.wikipedia.org/](www.wikipedia.org/)

Open Sense Map [opensensemap.org/](opensensemap.org/)

Smart citizen kit [smartcitizen.me/kits/](smartcitizen.me/kits/)

Fablabs [fablabs.io/](fablabs.io/)

Moskito Alert [mosquitoalert.com](mosquitoalert.com)

# Big Data

**What do we mean by data?**

THIS IS DIEGO'S PREPARATION FOR THE PYTHON PROGRAMMING COURSE AT IAAC;

```
 000 00 00 0 00000000 00000000 00000000 0000 0 00000000000000000 0 000 0000000000000000000
 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80
 1111111111111111111 1 1111111111111111111111111 11111111111111 111 1 11111111111
 222 22 22 2222222 2 2222222222222222222222222222222222222222222 2 22222222222222222222
 3333333333333333333 33333333 33333 33333333333333333 3333333 3333 3 33333333
 44444444 4444444444444444444444444444444444444444444 44444 4 444444444444444444444444
 5555555555 55 5555 5555555 5555555 555555 5555555555 5555555 5 5555555555555555555555
 66666666666 66666666666 66 6 66666666666 6666 6666666666 66666666666 6666666666
 777777777777 7777 77 7777777777777777 7 7 777777 77 7 777777 7777777777777777777777777
 8 8888888888 888888888888888888888 888 8 88888888888888888888888888888888 8888888888
 99 99 999 9999999 999 99 99999 99999999999999 99 999 99999 999999 999999999999999
 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80
```

CDL  0815

**Punch card**

# Digital



# See it
## *List & Plot*

```
my_df.head()

my_df.plot()
```

# Read it
## *Pd.read*

```
pd.read_csv('path/to/file.csv')
```
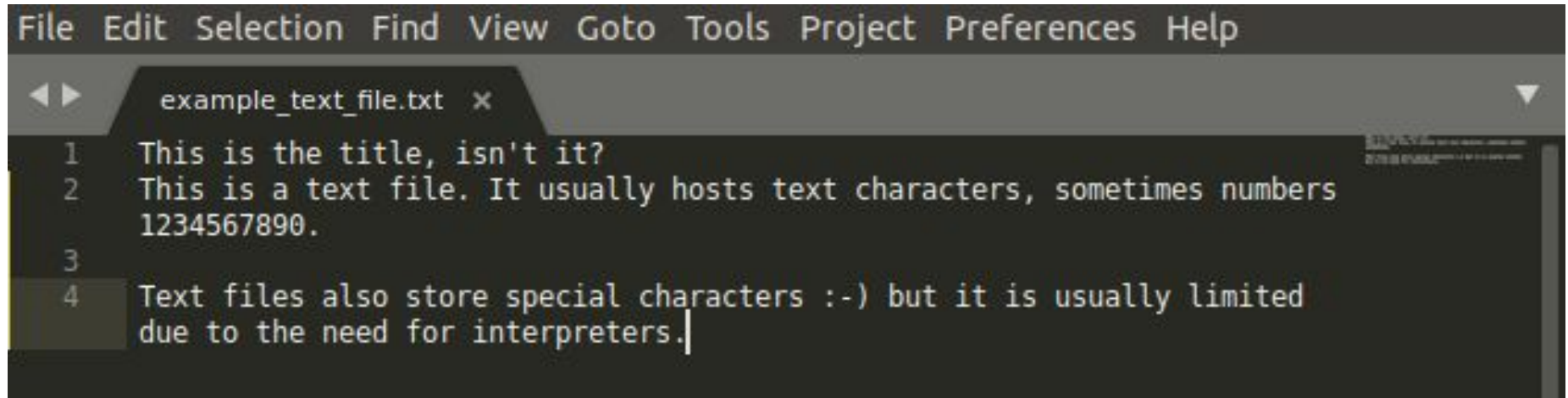
# Query it
## *a>3*

```
if a > 4:
    print('A is bigger than 3')
else:
    Print('No it is not')
```

Iaac | Institute for advanced architecture of Catalonia

# Open data defined right after the internet.
## In terms of Tim Berner Lee

https://5stardata.info/en/



Can we really access it?
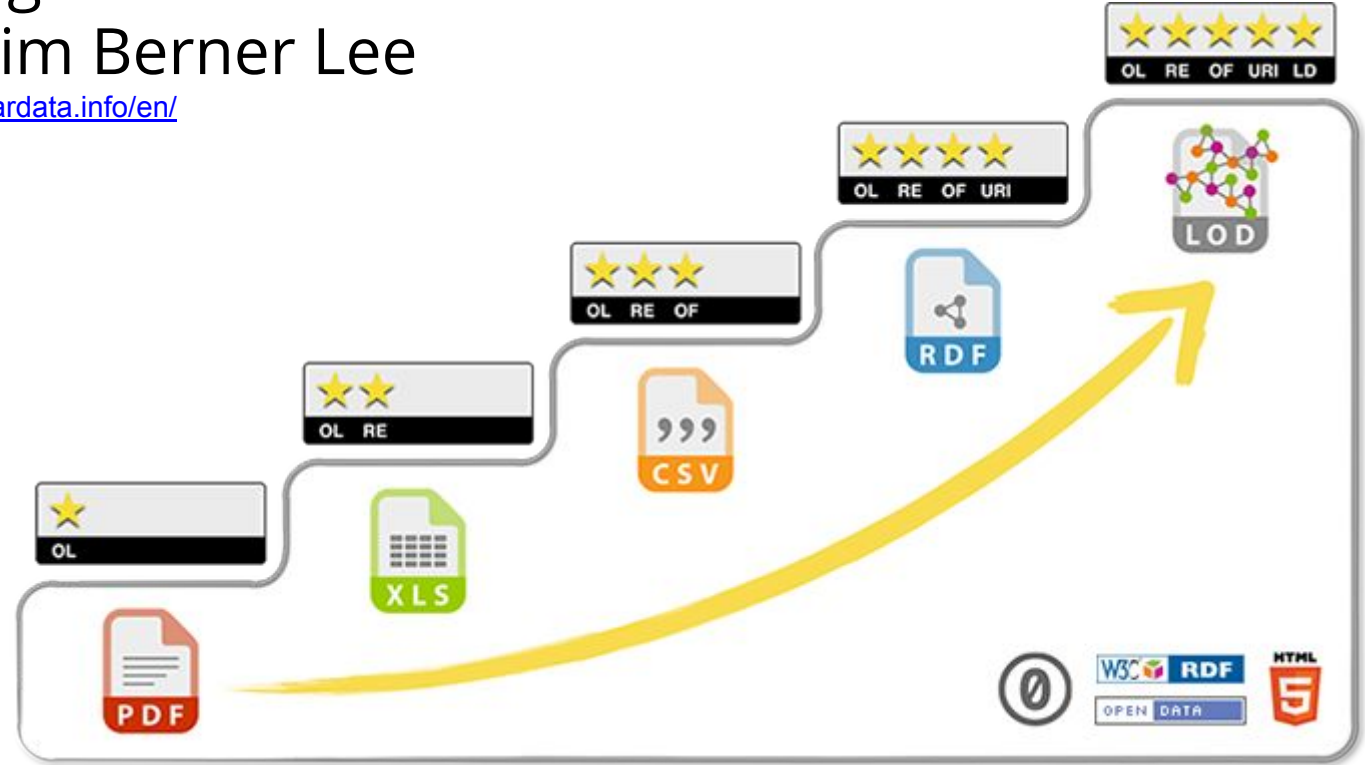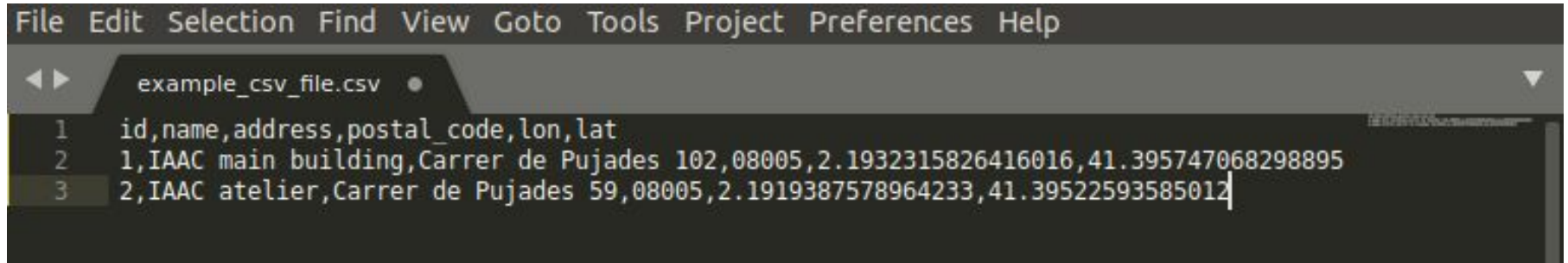
```
File  Edit  Selection  Find  View  Goto  Tools  Project  Preferences  Help

    ◄ ►      example_text_file.txt  ✕
    1       This is the title, isn't it?
    2       This is a text file. It usually hosts text characters, sometimes numbers
            1234567890.
    3
    4       Text files also store special characters :-) but it is usually limited
            due to the need for interpreters.
```

Text Files          -          Extension .txt

You can store almost everything here... they are quite heavy.

Can you guess why?

```
File  Edit  Selection  Find  View  Goto  Tools  Project  Preferences  Help

◀ ▶        example_csv_file.csv  ●                                                          ▼
    1     id,name,address,postal_code,lon,lat
    2     1,IAAC main building,Carrer de Pujades 102,08005,2.1932315826416016,41.395747068298895
    3     2,IAAC atelier,Carrer de Pujades 59,08005,2.1919387578964233,41.39522593585012
```

Comma separated Files        -        Extension .csv

This is a way to offer a more structured version of a text file.

It only works with simple structures, why?

| E6 | | ▼ | $f_x$ $\Sigma$ = | 4337032 | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | A | B | C | D | E | F | G | H | I | J | K | L |
| 1 | # | Country (or dependency) | Population (2019) | Yearly Change | Net Change | Density (P/Km²) | Land Area (Km²) | Migrants (net) | Fert. Rate | Med. Age | Urban Pop % | World Share |
| 2 | 1 | China | 1433783686 | 0.43 % | 6135900 | 153 | 9388211 | -348399 | 1.7 | 38 | 60 % | 18.59 % |
| 3 | 2 | India | 1366417754 | 1.02 % | 13775474 | 460 | 2973190 | -532687 | 2.2 | 28 | 35 % | 17.71 % |
| 4 | 3 | United States | 329064917 | 0.60 % | 1968652 | 36 | 9147420 | 954806 | 1.8 | 38 | 82 % | 4.27 % |
| 5 | 4 | Indonesia | 270625568 | 1.10 % | 2955025 | 149 | 1811570 | -98955 | 2.3 | 30 | 56 % | 3.51 % |
| 6 | 5 | Pakistan | 216565318 | 2.04 % | 4337032 | 281 | 770880 | -233379 | 3.6 | 23 | 35 % | 2.81 % |
| 7 | 6 | Brazil | 211049527 | 0.75 % | 1580204 | 25 | 8358140 | 21200 | 1.7 | 33 | 87 % | 2.74 % |
| 8 | 7 | Nigeria | 200963599 | 2.60 % | 5088916 | 221 | 910770 | -60000 | 5.4 | 18 | 51 % | 2.61 % |
| 9 | 8 | Bangladesh | 163046161 | 1.03 % | 1669453 | 1253 | 130170 | -369501 | 2.1 | 28 | 39 % | 2.11 % |
| 10 | 9 | Russia | 145872256 | 0.09 % | 138218 | 9 | 16376870 | 182456 | 1.8 | 40 | 74 % | 1.89 % |
| 11 | 10 | Mexico | 127575529 | 1.10 % | 1384741 | 66 | 1943950 | -60000 | 2.1 | 29 | 83 % | 1.65 % |
| 12 | 11 | Japan | 126860301 | -0.27 % | -341891 | 348 | 364555 | 71560 | 1.4 | 48 | 92 % | 1.64 % |
| 13 | 12 | Ethiopia | 112078730 | 2.61 % | 2854316 | 112 | 1000000 | 30000 | 4.3 | 19 | 21 % | 1.45 % |
| 14 | 13 | Philippines | 108116615 | 1.37 % | 1465221 | 363 | 298170 | -67152 | 2.6 | 26 | 47 % | 1.40 % |
| 15 | 14 | Egypt | 100388073 | 2.00 % | 1964475 | 101 | 995450 | -38033 | 3.3 | 25 | 43 % | 1.30 % |
| 16 | 15 | Vietnam | 96462106 | 0.96 % | 916144 | 311 | 310070 | -80000 | 2.1 | 32 | 37 % | 1.25 % |
| 17 | 16 | DR Congo | 86790567 | 3.24 % | 2722476 | 38 | 2267050 | 23861 | 6 | 17 | 45 % | 1.13 % |
| 18 | 17 | Germany | 83517045 | 0.47 % | 392627 | 240 | 348560 | 543822 | 1.6 | 46 | 76 % | 1.08 % |
| 19 | 18 | Turkey | 83429615 | 1.32 % | 1089527 | 108 | 769630 | 283922 | 2.1 | 32 | 75 % | 1.08 % |
| 20 | 19 | Iran | 82913906 | 1.36 % | 1113718 | 51 | 1628550 | -55000 | 2.2 | 32 | 75 % | 1.07 % |
| 21 | 20 | Thailand | 69625582 | 0.28 % | 197129 | 136 | 510890 | 19444 | 1.5 | 40 | 50 % | 0.90 % |

Sheet1

Microsoft Excel Files     -     Extension .xlsx .xls .ods

Open Structure based on grids, multiple data formats and functions. It is built on top rows, columns and sheets.

Iaac | Institute for advanced architecture of Catalonia     Data formats - XLSX Files

Javascript object notation Files
Extension .json

It is a text file with a more complex structure optimised for data exchange through the web

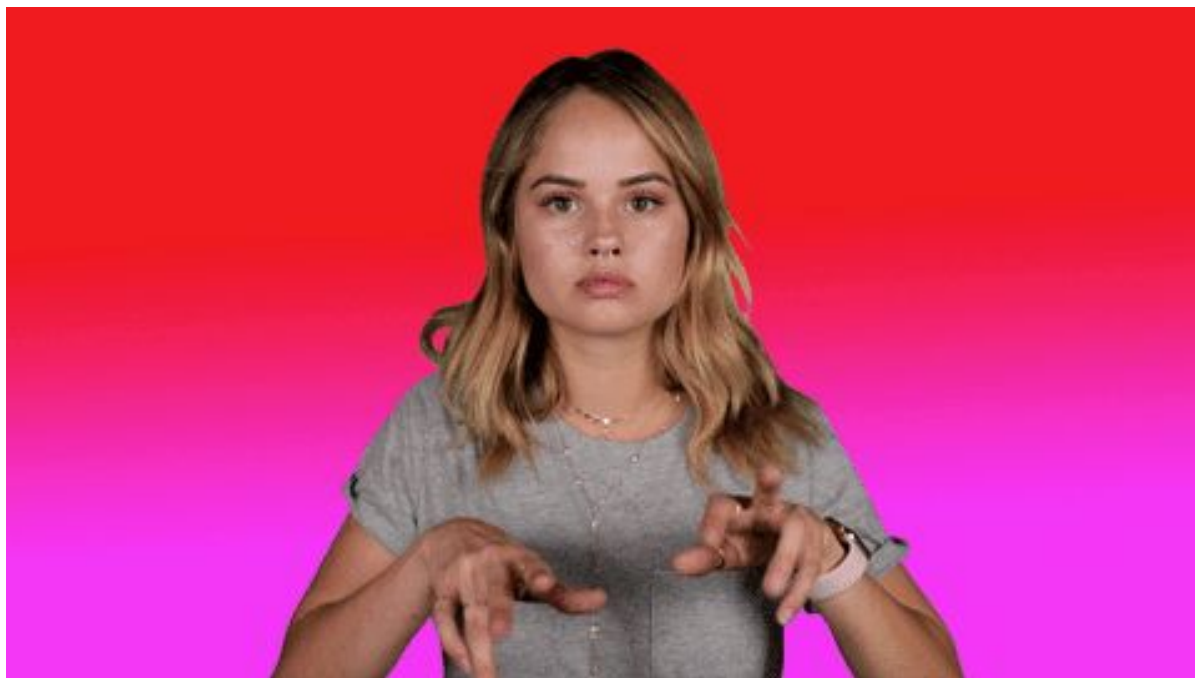Say HI to our friends "dictionaries" and "lists"?

Data formats - JSON Files

eXtended markup language Files
Extension .xml / .{}ml

It is a text file with arrays of markups designed for machines to understand web data

This is the weirdest so far, isn't it?

# Big Data

**Data compilation**

Go back to the compilation of data sources

Identify two or more cities/places in which you think an interesting problem deserves to be analysed

Download some data files and try to open it with the software you have in your computer

Create two or more repositories for cities

Barcelona, New York and others

What questions can we answer this data sets?

Can we create plots, maps, or other geospatial visualisation

**Share some examples with your peers.**

Iaac | Institute for advanced architecture of Catalonia | BARCELONA

**MASTER IN CITY & TECHNOLOGY**
**DIGITAL TOOLS AND BIG DATA**
**2020/2021**

**FACULTY** DIEGO PAJARITO