



Big Data: Revelando los Secretos de Twitter en México

22 de Octubre 2014

Presentado por:
Abel Alejandro Coronado Iruegas
@abxda



abel.coronado@inegi.org.mx

Objetivo

Inspirarlos para que le entren
al mundo de Big Data.

Big Data

Temas

[Suscribirse](#)



Big Data

Término de búsqueda

+ Agregar término

Interés a lo largo del tiempo



Titulares de noticias



Previsión



Diciembre de 2011

big data: 15

2005

2007

2009

2011

2013

<https://www.google.com.mx/trends/>

<https://twitter.com/abxda>

¿Qué es Big Data?

¿Qué es Big Data?

Ryan Swanson

Data Science Blogger, Data Science 101 ↗

Twitter: @swgoof ↗

“*Big data used to mean data that a single machine was unable to handle. Now big data has become a buzzword to mean anything related to data analytics or visualization.*

¿Qué es Big Data?

Anna Smith

Analytics Engineer, Rent the Runway ↗

Twitter: @OMGannaks ↗

“ ***Big data is when data grows to the point that the technology supporting the data has to change. It also encompasses a variety of topics relating to how disparate data can be combined, processed into insights, and/or reworked into smart products.***

¿Qué es Big Data?

David Leonhardt

Editor, The Upshot , The New York Times

Twitter: @DLeonhardt 

“

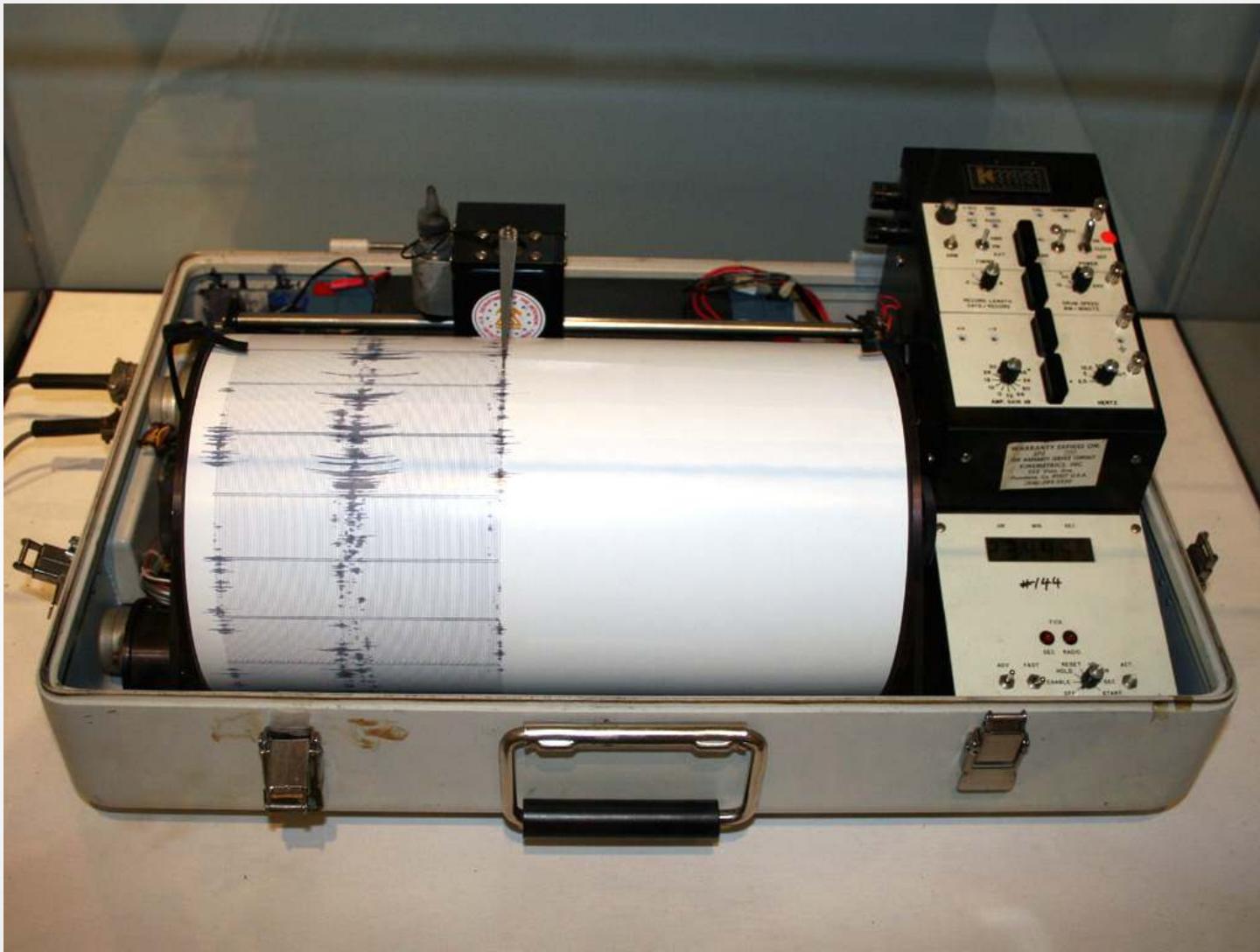
Big Data is nothing more than a tool for capturing reality — just as newspaper reporting, photography and long-form journalism are. But it's an exciting tool, because it holds the potential of capturing reality in some clearer and more accurate ways than we have been able to do in the past.

”

Volumen



Velocidad



http://upload.wikimedia.org/wikipedia/commons/0/0f/Kinematics_seismograph.jpg

<https://twitter.com/abxda>

Variedad



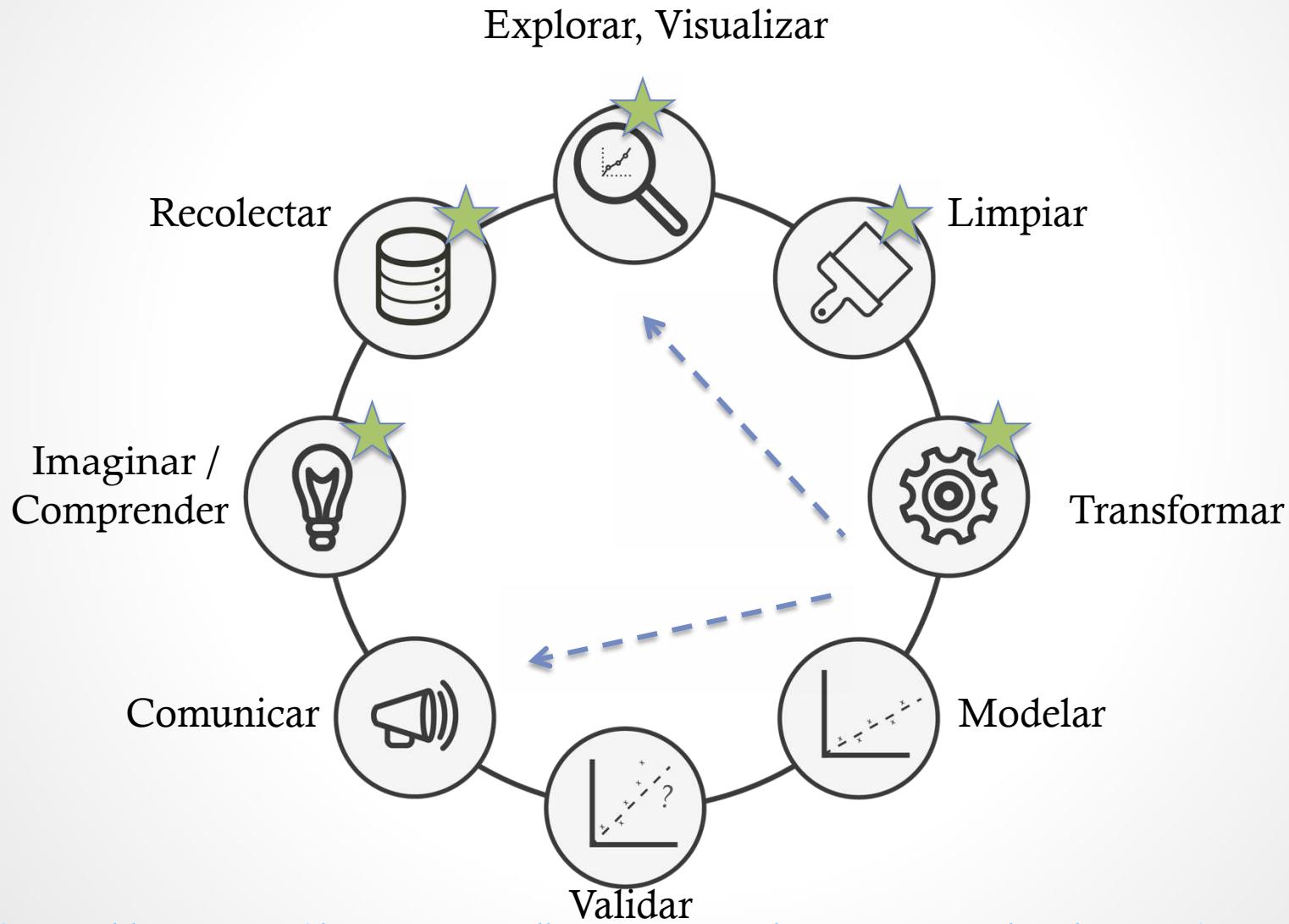
Tomar decisiones, actuar y crear valor



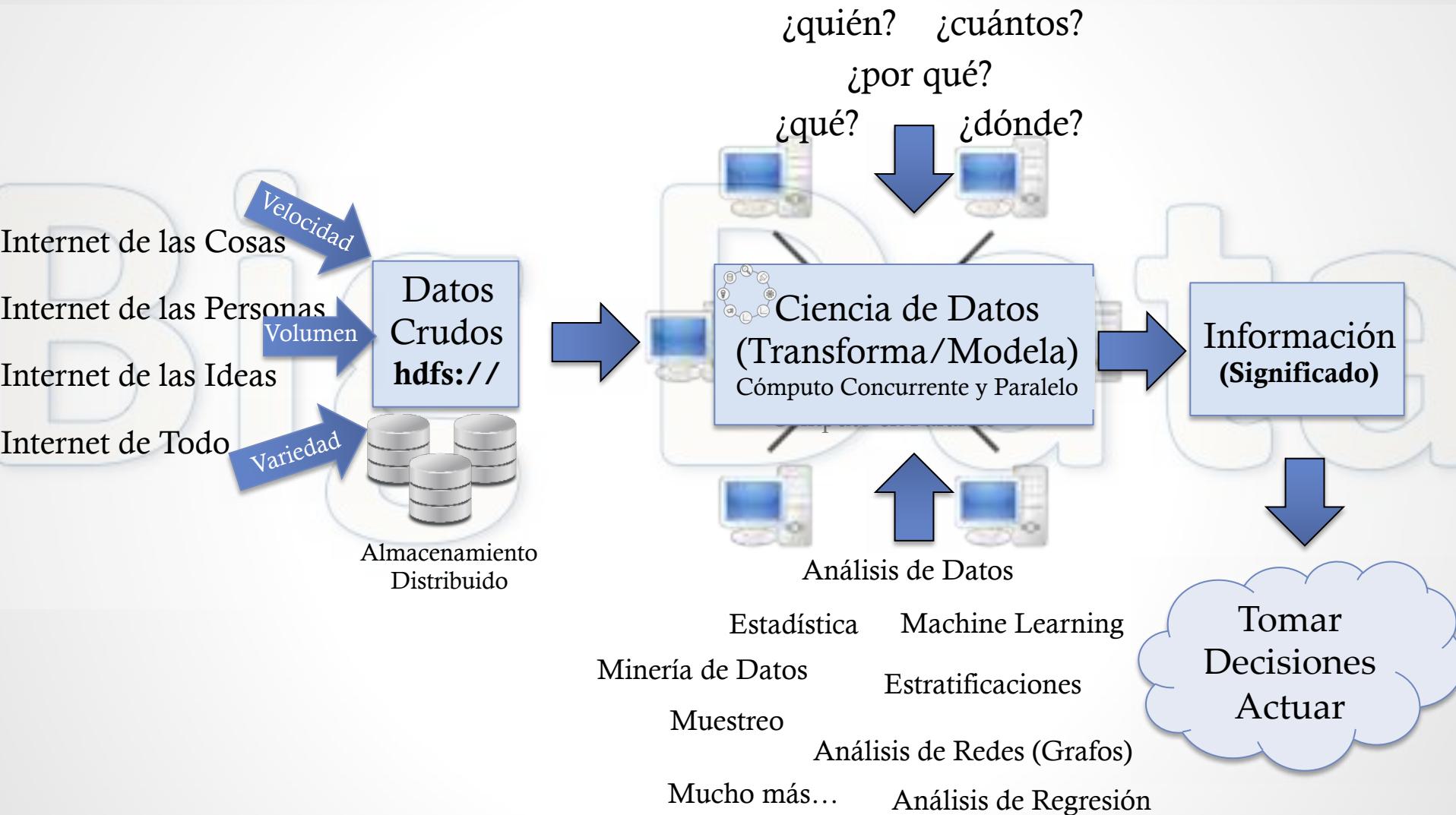
Ciencia de Datos



Ciencia de Datos



Ciencia de Datos y Big Data



Big Data en las Oficinas Nacionales de Estadística

UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS

10 March 2013

WHAT DOES “BIG DATA” MEAN FOR OFFICIAL STATISTICS?

Big Data en las Oficinas Nacionales de Estadística

- It is clear that during the next two years there is a need to identify a few pilot projects that will serve as proof of concept.
- Statistical organisations are, therefore, encouraged to address formally Big data issues in their annual and multi-annual work programmes by undertaking research and pilot projects in selected areas and by allocating appropriate resources for that purpose.



Big Data en las Oficinas Nacionales de Estadística

- '**new**' exploration and analysis methods are required: *Visualization methods*, *Text mining*, and *High Performance Computing*.
- To use Big data, **statisticians are needed with a different mind-set and new skills**. The processing of more and more data for official statistics requires statistically aware people with an analytical mind-set, an affinity for IT (e.g. programming skills)



Twitter como fuente de BigData



¿Cuántos caracteres?

The screenshot shows a Twitter post from user Abel Coronado (@abxda). The post contains the following text:

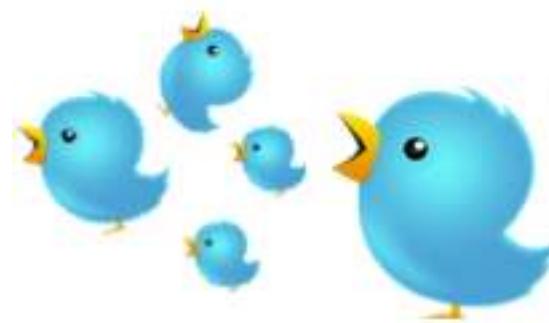
Mañana 22 de Oct !!! Big Data: Revelando los Secretos de Twitter en México | SG: sg.com.mx/sgvirtual/7/se... vía @RevistaSG

Below the text is a map showing the location of the event, which is marked with a red pin in the Jardines de las Bambas area of Mexico City.



<https://twitter.com/abxda>

140 ??



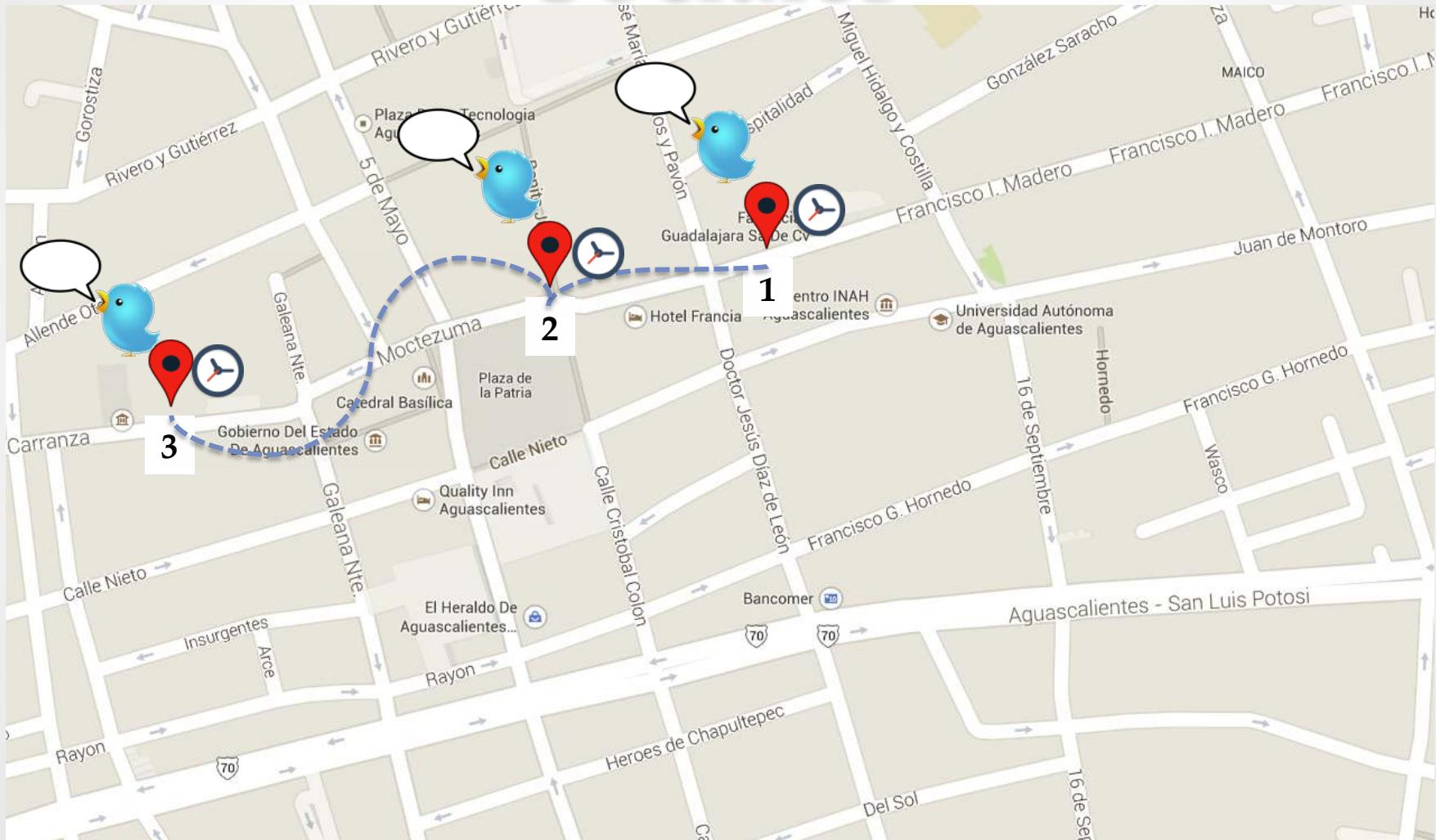
"text": Mañana 22 de Oct !!! Big Data: Revelando los efectos de Twitter en México | Segundo Informe
"created_at": "2014-02-21T17:00:50.000Z",
"source": "Android",
"truncated": false,
"mention": [],
"retweet_count": 0,
"hashtag": [],
"location": {
 "lat": 19.39617897,
 "lon": -99.22636055
},
"place": {
 "id": "3ad512d283f67a11",
 "name": "Aguascalientes",
 "type": "city",
 "full_name": "Aguascalientes, Aguascalientes",
 "street_address": null,
 "country": "México",
 "country_code": "MX",
 "url": "https://api.twitter.com/1.1/geo/id/3ad512d283f67a11.json"
},
"link": [
 {
 "url": "http://t.co/AUNXLVSimQ",
 "display_url": "4sq.com/lplLyUL",
 "expand_url": "http://4sq.com/lplLyUL",
 "start": 28,
 "end": 50
 }
],
"user": {
 "id": 205760874,
 "name": "Abel Coronado",
 "screen_name": "abxda",
 "location": "",
 "description": "Filósofo, Desarrollador de Software, M.C. en Estadística Oficial by CIMAT, Emprendedor,
 "profile_image_url": "http://pbs.twimg.com/profile_images/378800000635411988/5c5b7a5754d65e3d1a2895",
 "profile_image_url_https": "https://pbs.twimg.com/profile_images/378800000635411988/5c5b7a5754d65e3"
}

1482

{ JSON }

Json: Formato de Intercambio

Nuestra huella en las Redes Sociales



Todos los tuits están disponibles para su recolección en tiempo real.

The screenshot shows the Twitter Developers documentation page for the Streaming API. At the top, there's a navigation bar with links for Developers, API Health, Blog, Discussions, Documentation, and a Search bar. Below the navigation bar, a breadcrumb trail indicates the current location: Home → Documentation. The main title is "The Streaming APIs". There are two buttons below the title: "View" and "What links here". At the bottom of the page, there's a note about the last update ("Updated on Mon, 2012-09-24 14:47") and two API version buttons: "API version 1" and "API version 1.1".

The Streaming APIs

[View](#)

[What links here](#)

Updated on Mon, 2012-09-24 14:47

[API version 1](#)

[API version 1.1](#)

Overview

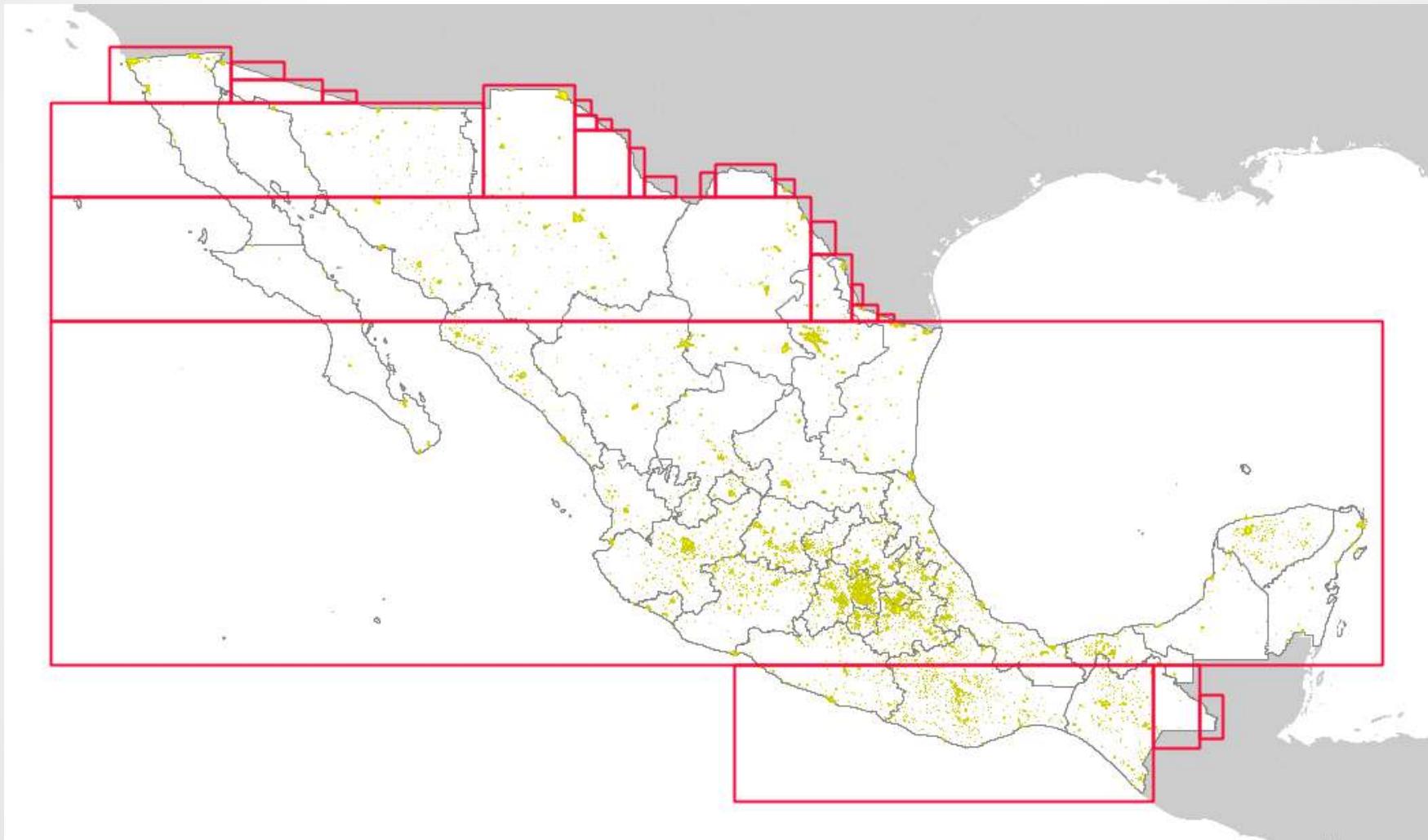
The set of streaming APIs offered by Twitter give developers low latency access to Twitter's global stream of Tweet data. A proper implementation of a streaming client will be pushed messages indicating Tweets and other events have occurred, without any of the overhead associated with polling a REST endpoint.

Twitter offers several streaming endpoints, each customized to certain use cases.

[Public streams](#)

Streams of the public data flowing through Twitter. Suitable for following specific users or topics, and data mining.

Incluso permite consultas geográficas



¿Dónde recolectar?

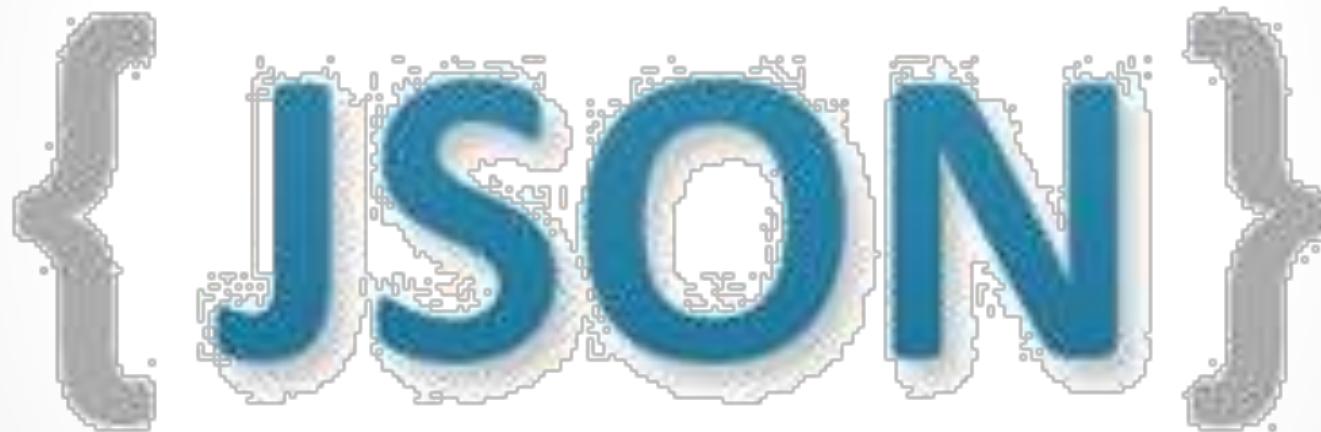




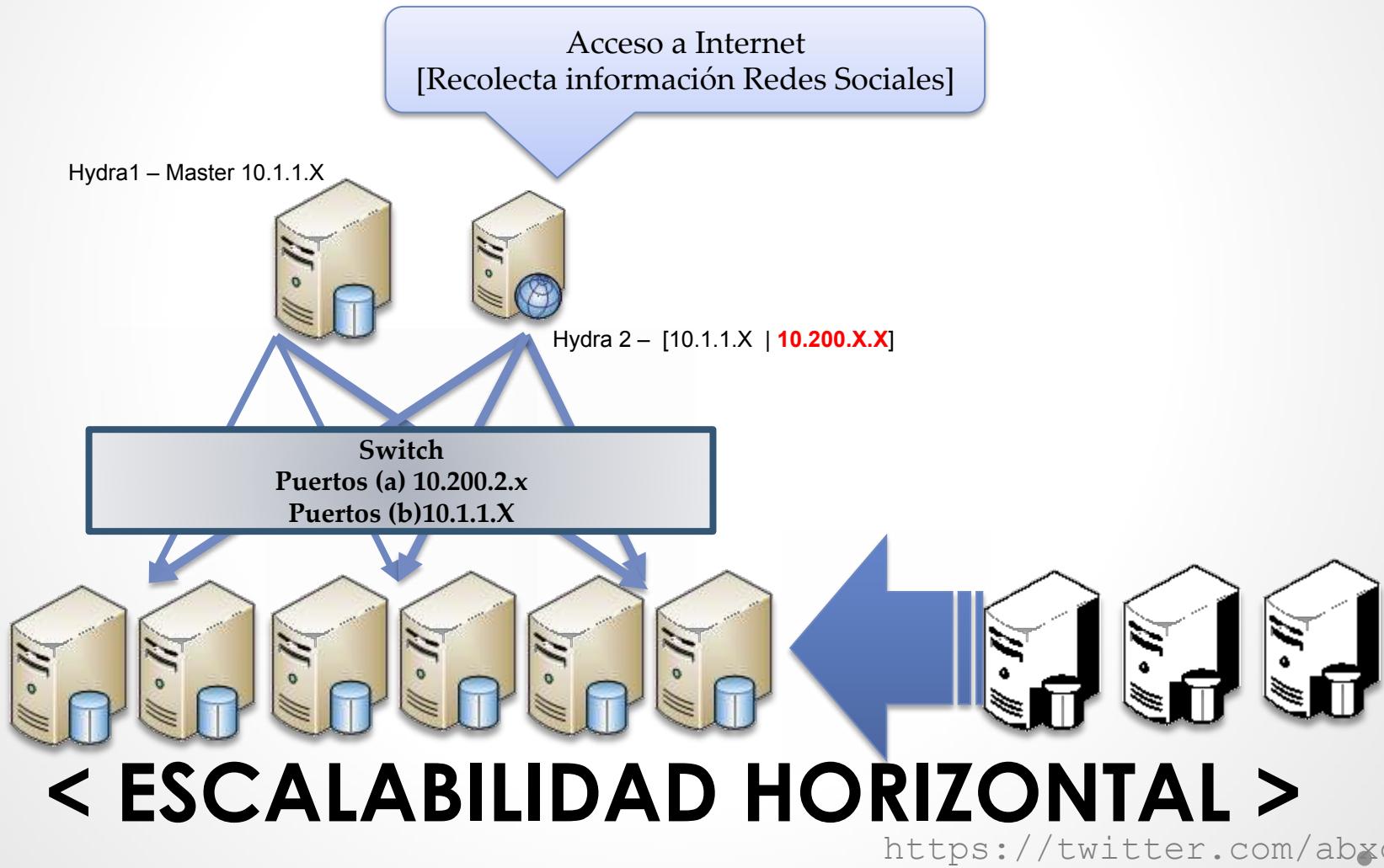
elasticsearch

<http://www.elasticsearch.org/>

¿Por qué ElasticSearch?



¿Por qué Elasticsearch?



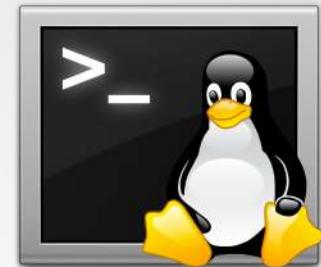
Hydra



Hydra



Twitter River



<https://github.com/elasticsearch/elasticsearch-river-twitter>

```
curl -XPUT localhost:9200/_river/my_twitter_river/_meta -d'
{
  "type" : "twitter",
    "twitter" : {
      "oauth" : {
        "consumer_key" :"XXXXXXXXXXxXxX",
        "consumer_secret" : "XXXXXXXXXXxXXXXXXXxXXXXxXXXXXXxXXXXXX",
        "access_token" : "XXXXXXXXXXxXxXXXXXXXxXXxXxXXXXXXxXXXXXxXxX",
        "access_token_secret" : "XXXXXXXXXXXXXXxXXXXXXxXXXXXX"
      },
      "filter" : {
        "locations" :" -118.40764955,14.53209836,-86.71040527,32.71865357"
      }
    }
  '
```

La recolección 2014

elasticsearch

Indices Query Mappings </> REST

Node Diagnostics Monitor Nodes

10:02:09 Cluster Overview

Cluster Statistics

8 Nodes	12 Total Shards	12 Successful Shards	2 Indices	91,808,849 Documents	72.0GB Size
------------	--------------------	-------------------------	--------------	-------------------------	----------------



elasticsearch

Extractor

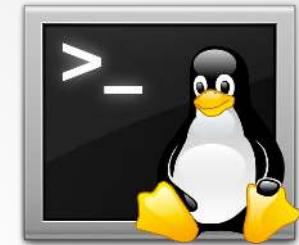


```
es = Elasticsearch(['10.200.2.41:9200'])
rs = es.search(index=['my_twitter_river'],
scroll=duracion, search_type='scan', size=int(noTuits),
body={
    "query": {
        "range" : {
            "created_at" : {
                "gte": fechaInicio,
                "lte": fechaFin
            }
        }
    }
})
```

CSV

```
MacBook-Pro-de-Abel:DataBase abxda$ ls -alh tweets_f_s.csv  
-rw-r--r-- 1 abxda staff 18G Sep 25 11:48 tweets_f_s.csv  
MacBook-Pro-de-Abel:DataBase abxda$ head tweets_f_s.csv
```

Se extraen los puntos del CSV



```
$cat tweets_feb_sep_ord_loc.csv | awk -F',' '{print $3 "," $4}'
```

```
20.281523,-100.809407  
20.281523,-100.809407  
20.281667,-100.809311  
20.281479,-100.809394  
20.281526,-100.809377  
20.281422,-100.809428  
20.281478,-100.809406  
20.281495,-100.809371  
20.281521,-100.80937  
25.767972,-103.274890  
25.768021,-103.274900  
25.768059,-103.274955  
25.768019,-103.274900  
25.768098,-103.274992
```

Quantum GIS

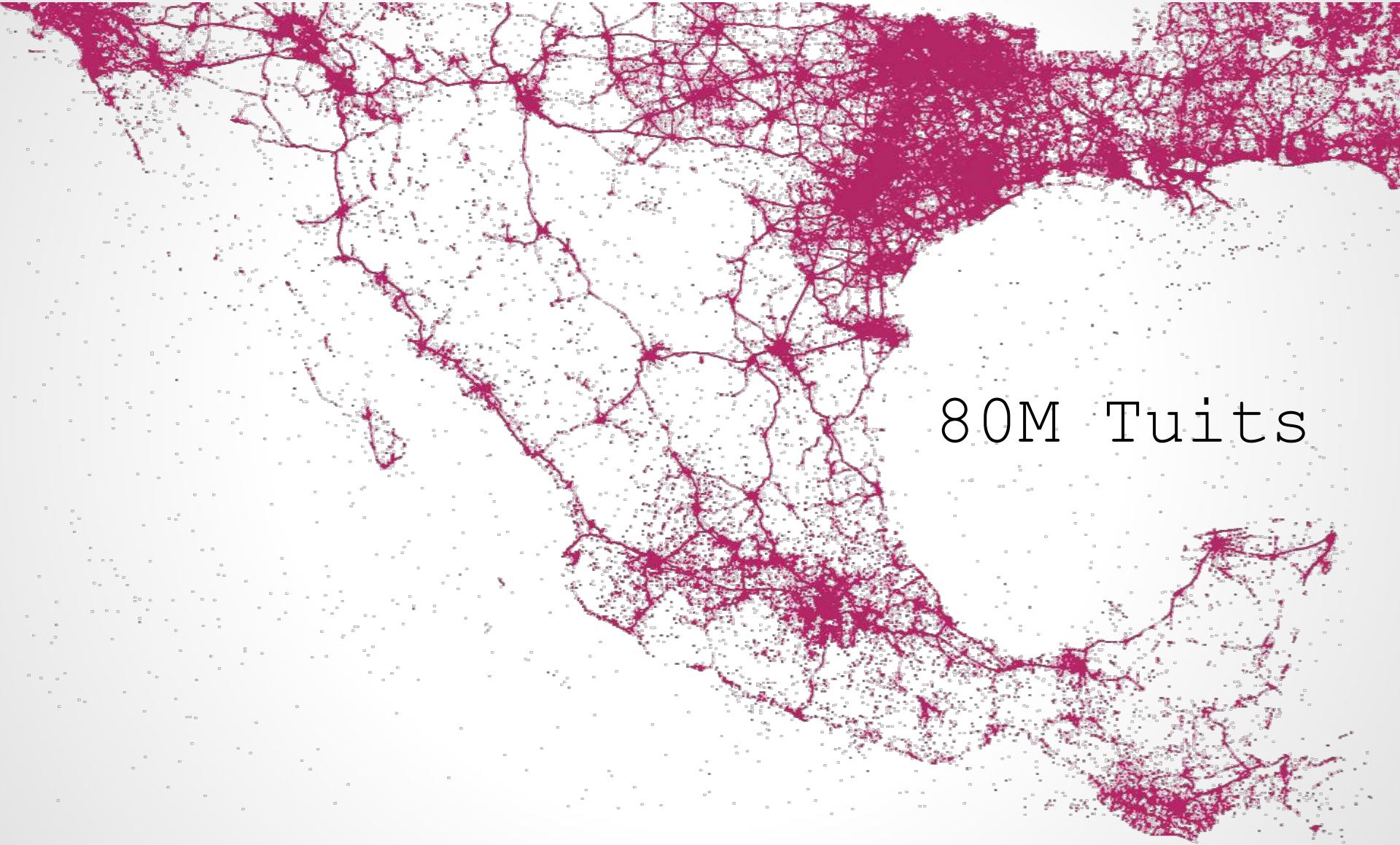


The screenshot shows the Quantum GIS application window. The menu bar at the top includes 'QGIS', 'Proyecto', 'Edición', 'Ver', 'Capa' (which is highlighted in blue), 'Configuración', 'Complementos', and 'Vectorial'. Below the menu bar is a toolbar with various icons for file operations and editing. To the left is a panel titled 'Capas' (Layers) containing a list of layers. The main workspace is currently empty. A context menu is open on the right side, listing options for adding new layers:

- Nueva
- Empotrar capas y grupos...
- Añadir capa vectorial...
- Añadir capa ráster...
- Añadir capas PostGIS...
- Añadir capa SpatialLite...
- Añadir capa espacial de MSSQL...
- Añadir capa WMS/WMTS...
- Agregar capa WCS...
- Añadir capa WFS...
- Añadir capa de texto delimitado...** (this option is highlighted in blue)

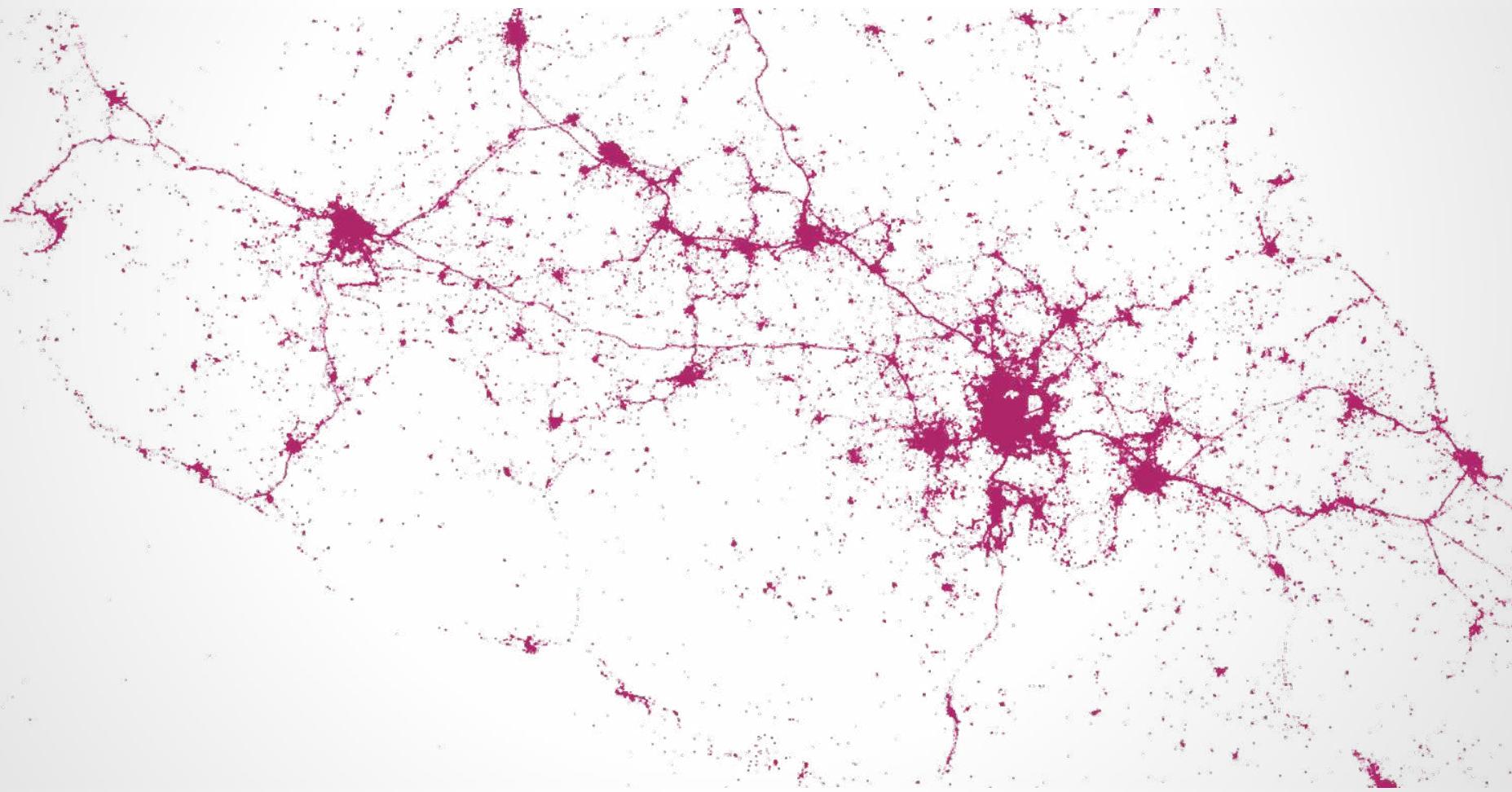
<http://www.qgis.org/>

Resultado de la recolección



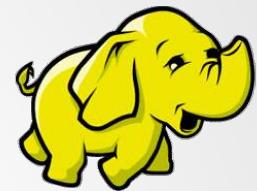
80M Tuits

Un acercamiento



Hadoop Distributed File System

hdfs://



Type	Name	Size	User	Group	Permissions	Date
Folder	.		acoronado	acoronado	drwxr-xr-x	October 06, 2014 02:53 PM
Folder	..		acoronado	acoronado	drwxr-xr-x	September 25, 2014 06:18 AM
File	2014-02_al_2014-09-23.csv	18.2 GB	acoronado	acoronado	-rw-r--r--	September 25, 2014 08:57 AM

Hadoop / Apache Spark

Punto de Partida

[Hadoop]:

48 Cores >3 Ghz

128 Gb RAM

4 TB Almacenamiento Permanente

Punto de Partida

[Spark]:

24 Cores >3 Ghz

128 Gb RAM

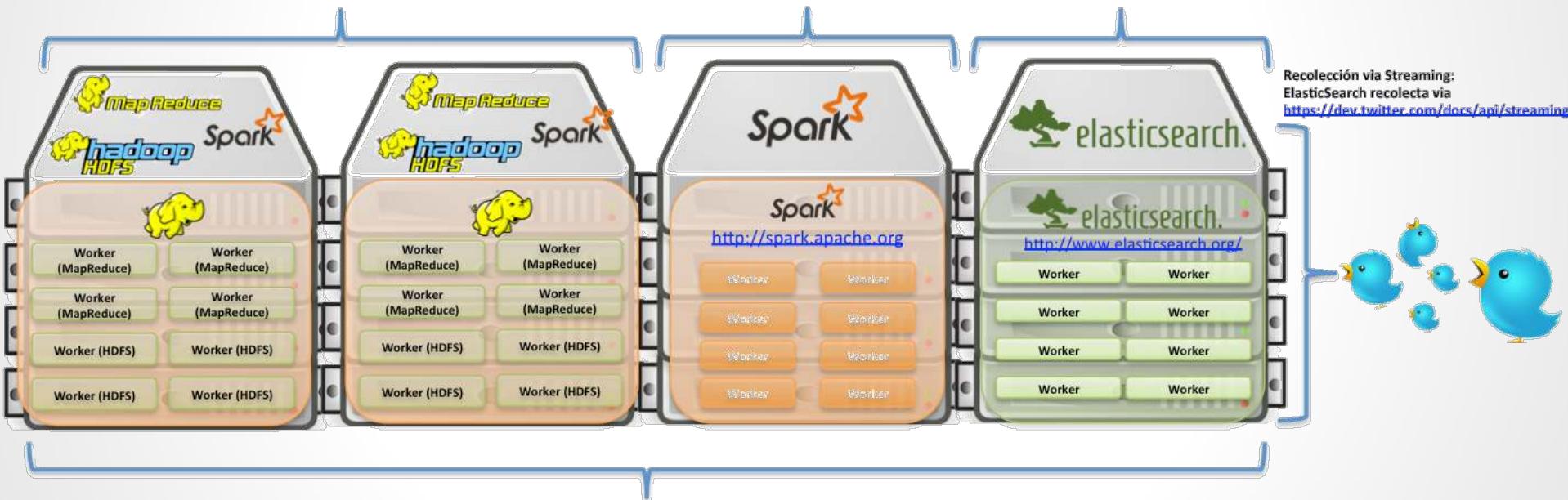
1 TB Almacenamiento Volátil

Actualmente [ES]:

18 Cores 2.5 Ghz

68 Gb RAM

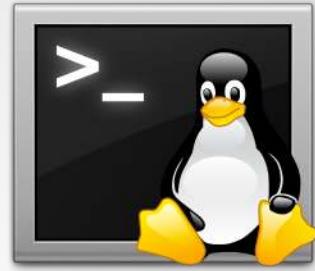
900 Gb Almacenamiento



Recorte Geográfico



```
object SimpleApp {  
def main(args: Array[String]) {  
    ...  
    val csvPath = "hdfs://m01/user/acoronado/mov/2014-02_al_2014-09-23.csv"  
    val csv = sc.textFile(csvPath)  
    csv.cache()  
    val clipPoints = csv.map({line: String =>  
        val Array(usuario, lat, lon, date) = line.split(",").map(_.trim)  
        val geometryFactory = JTSFactoryFinder.getGeometryFactory();  
        val reader = new WKTReader(geometryFactory);  
        val point = reader.read("POINT (" + lon + " " + lat + ")")  
        val envelope = point.getEnvelopeInternal  
        val internal = geoDataMun.get(envelope)  
        val (cve_est, cve_mun) = internal match {  
            case l => {  
                val existe = l.find(f => f match {  
                    case (g: Geometry, e: String, m: String) => g.intersects(point)  
                    case _ => false}  
                existe match {  
                    case Some(t) => t match {  
                        case (g: Geometry, e: String, m: String) => (e, m)  
                        case _ => ("0", "0")  
                    }  
                    case None => ("0", "0")  
                }  
            }  
            case _ => ("0", "0")  
        }  
        val time = ...  
        line + "," + time + "," + cve_est + "," + cve_mun  
    })  
    clipPoints.coalesce(5, true).saveAsTextFile("hdfs://m01/user/acoronado/mov/resultados_movilidad_parts.csv")  
}  
}
```



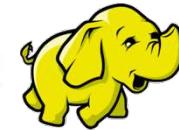
```
cat tweets_feb_sep.csv | awk -F',' '{print $1}'|sort| uniq | wc -l
```

Más de 700,000 tuiteros
dentro del territorio
Mexicano.

Calcular total de tuits por Hora

```
val csvPath ="hdfs://master/user/acoronado/tweets_feb_sep.csv"  
csv.cache  
val csv = sc.textFile(csvPath)  
  
val hours =  
csv.map({line:String =>  
    val campos = line.split(",").map(_.trim)  
    val d1 = new Date(campos(8).toLong)  
    val format = new SimpleDateFormat("dd-MM-yyyy,HH")  
    (format.format(d1),1)}).reduceByKey((a,b) => a+b)  
  
hours.coalesce(1).saveAsTextFile("hdfs://.../days_hours_string.csv")
```





ACTIONS

[View as binary](#)[Edit file](#)[Download](#)[View file
location](#)[Refresh](#)

INFO

Last modified

Sept. 29, 2014
7:41 a.m.

User

acoronado

Group

acoronado

Size[First Block](#)[Previous Block](#)[Next Block](#)[Last Block](#)

(28-08-2014,07,3883)
(05-05-2014,23,12930)
(25-08-2014,08,5085)
(09-06-2014,22,14460)
(06-06-2014,23,11730)
(14-02-2014,20,10515)
(01-07-2014,21,9643)
(22-08-2014,05,788)
(04-04-2014,23,10204)
(03-06-2014,20,12069)
(11-02-2014,21,13744)
(05-08-2014,20,10271)
(21-07-2014,09,5644)
(30-07-2014,03,3516)
(31-05-2014,06,923)
(29-06-2014,06,1170)

Generar la Gráfica



Screenshot of Microsoft Excel showing a table of data and the ribbon menu.

The ribbon menu includes:

- Excel (Apple logo)
- Archivo
- Edición
- Ver
- Insertar
- Formato
- Herramientas

The Quick Access Toolbar includes:

- Save
- Open
- New
- Print
- Cut
- Paste
- Find
- Replace
- Sort
- Filter
- Σ
- Z
- Filter

The Home tab ribbon includes:

- Inicio
- Diseño
- Tablas
- Gráficos
- SmartArt
- Fórmulas

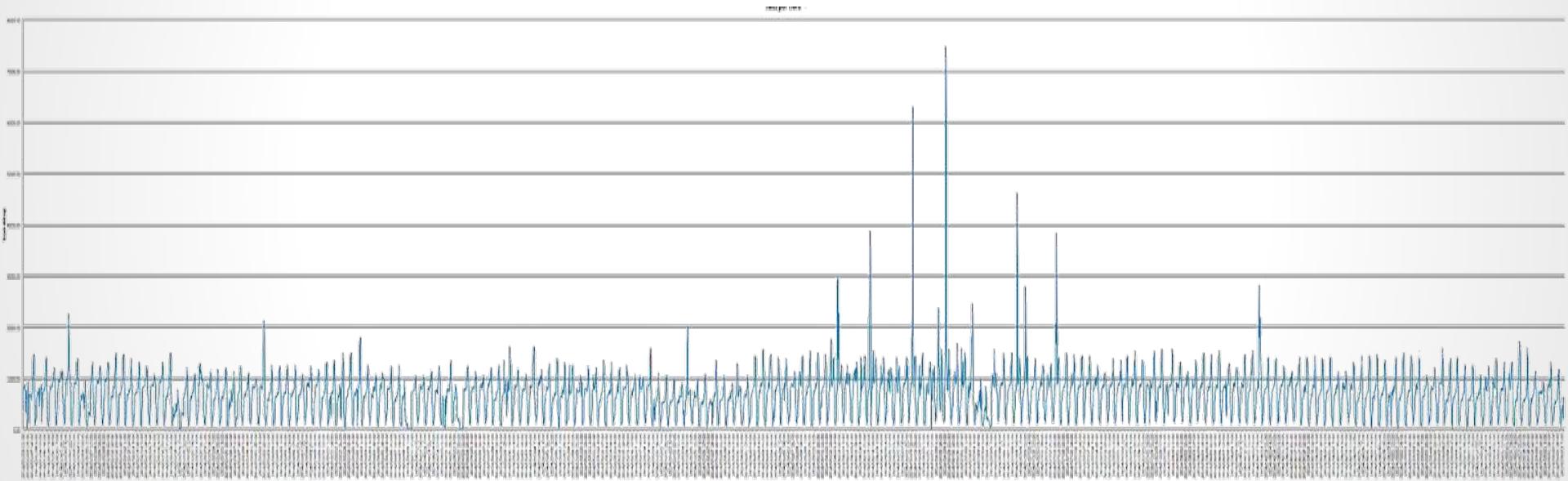
The Font ribbon includes:

- Editar
- Rellenar
- Fuente (Font): Calibri (Cuerpo) 12pt
- Text styles: A, A, A
- Font color: N, K, S
- Font size: 12
- Font style: Bold, Italic, Underline
- Font orientation: 90°, 270°, 45°, 315°
- Font color: Blue, Yellow, Red

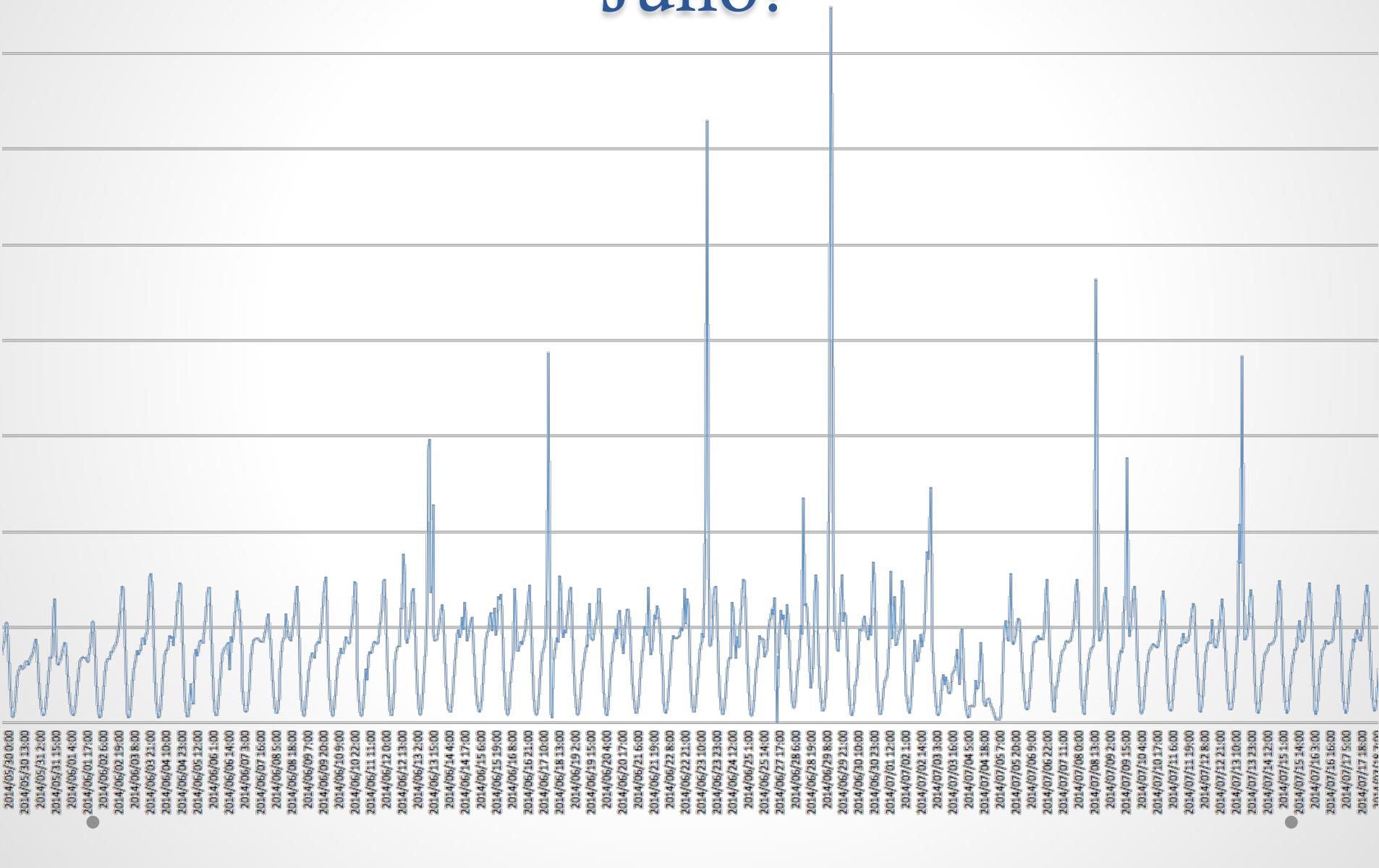
The table data is as follows:

	O	P	Q	R	S	T
1	Fecha - Hora	Total Tuits				
2	23/01/14 - 11 hrs.	3727.00				
3	23/01/14 - 12 hrs.	7342.00				
4	23/01/14 - 13 hrs.	7412.00				
5	23/01/14 - 14 hrs.	8318.00				
6	23/01/14 - 15 hrs.	8777.00				
7	23/01/14 - 16 hrs.	8198.00				
8	23/01/14 - 17 hrs.	8149.00				
9	23/01/14 - 18 hrs.	7145.00				
10	24/01/14 - 8 hrs.	3287.00				
11	24/01/14 - 9 hrs.	7331.00				
12	24/01/14 - 10 hrs.	7707.00				

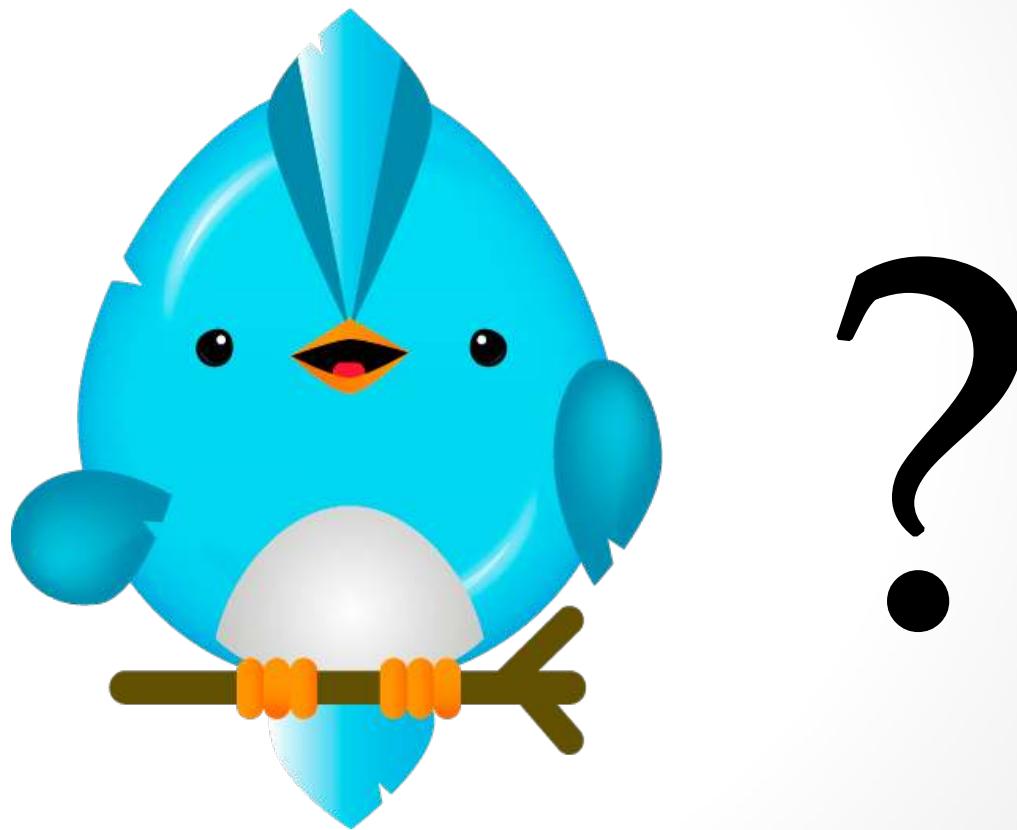
A lo largo del tiempo



¿Qué pasó entre el 12 de Junio y el 13 de Julio?



Pregúntale a Twitter



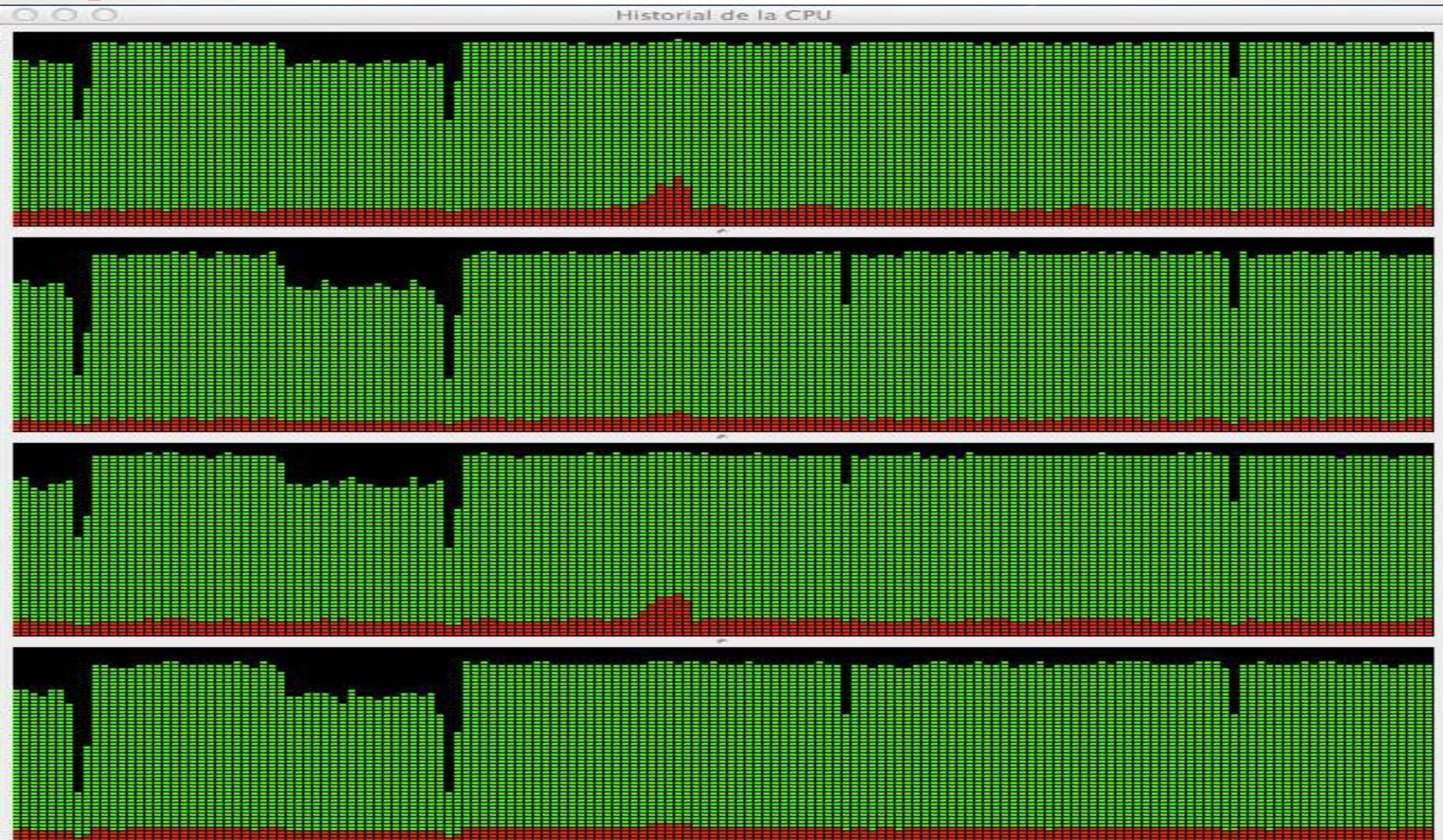
Busca tuits en la fecha específica



```
object Main extends App {  
  val fech1 = new SimpleDateFormat("yyyy-MM-dd'T'HH:mm:ss").parse("2014-06-12T00:00:00")  
  val fech2 = new SimpleDateFormat("yyyy-MM-dd'T'HH:mm:ss").parse("2014-07-13T23:59:59")  
  scala.io.Source.fromFile("/abxda/BigData/tweets_feb_sep_ord_loc.csv")  
    .getLines()  
    .grouped(250000)  
    .flatMap { y=>  
      y.par.filter({line: String =>  
        val campos = line.split(",").map(_.trim)  
        val time = new Date(campos(8).toLong)  
        time.after(fech1) && time.before(fech2)  
      })  
    }.foreach({ x: String =>  
      println(x.toString)  
    })  
}
```

Cómputo paralelo

y.**par**.filter



Encuentra Hashtags

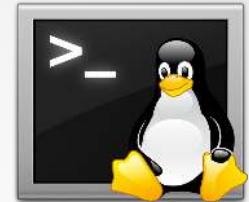


```
# coding=utf-8
import codecs
import re
cnt = 0
with codecs.open('/abxda/BigData/Periodo.csv', 'r', 'utf-8') as f:
    for line in f:
        try:
            csv = line.split(',')
            text = csv[7]
            hashtags=re.findall(u"#([áéíóúÁÉÍÓÚñÑA-Za-z0-9_]+)",text,re.U)
            for ht in hashtags:
                print '#'+ht
        except Exception:
            pass
```



Prepara archivo para Wordle

<http://www.wordle.net/>



```
cat hashtagsMundial.txt | sort | uniq -c | sort -n | awk -F' ' '{print $2 ":" $1}' > wordleMun.txt
```

```
#NED:8313
#MundialBrasil2014:8777
#VamosMexico:8947
#BRA:10098
#CallMeCam:14531
#ARG:15663
#Brasil2014:16428
#GER:18030
#MEX:34035
```



¿Qué pasó entre el 12 de junio y el 13 de julio?



¿Qué pasó el 23 de junio?

#LosChavosHablaronEnLaCancha

#Mex

#VamosMexico

#GanaroGanar

#CroaciaTeVasDeBrasil

#CROvsMEX
#MEXICOvsCROACIA
#Mexico

#MundialBrasil2014

#Brasil2014

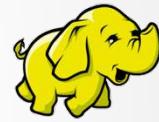
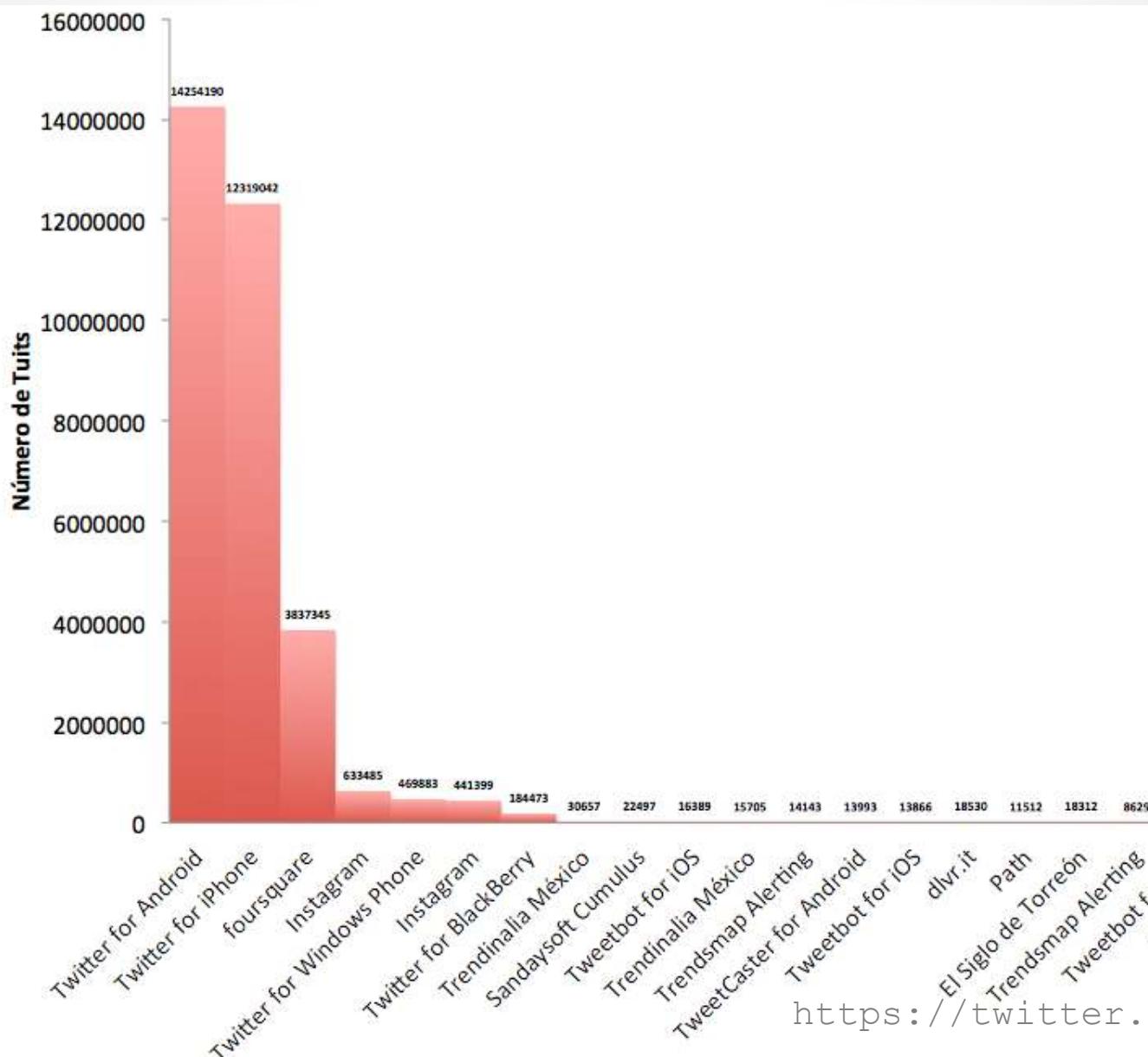
#CRO
#NED

MEX

¿Qué pasó el 29 de junio?



¿Con qué tuiteamos?





¿A qué hora tuiteamos?



¿Qué tuiteamos?

A large word cloud in Spanish with various words and their meanings. The words are in different colors and sizes, representing frequency or importance. Some words have small boxes with symbols or numbers next to them.

Key words and their meanings include:

- Todo** (All)
- noches** (nights)
- Bueno** (Good)
- Alguien** (Someone)
- donde** (where)
- jajaja** (jajaja)
- amigo** (friend)
- Mañana** (Tomorrow)
- Internacional** (International)
- tengo** (I have)
- excelente** (Excellent)
- Siempre** (Always)
- novio** (Boyfriend)
- soy** (I am)
- feliz** (Happy)
- DE** (Degree)
- <3** (Under 3)
- Buenos** (Good)
- saber** (To know)
- nuevo** (New)
- #RetoTelehit** (Twitter challenge)
- corazón** (Heart)
- Tu** (Your)
- saludos** (Greetings)
- Yo** (I)
- Jajaja** (jajaja)
- mal** (Bad)
- decir** (Say)
- bueno** (Good)
- gente** (People)
- contigo** (With you)
- feliz** (Happy)
- Quiero** (I want)
- salir** (Go out)
- puede** (Can)
- dice** (Says)
- Cinépolis** (Cinépolis)
- comer** (Eat)
- aunque** (Although)
- viendo** (Watching)
- Mexico** (Mexico)
- hermosa** (Beautiful)
- hola** (Hello)
- encanta** (Enchanting)
- eres** (You are)
- Hoy** (Today)
- #KCAMexico** (#KCAMexico)
- #MarioBautista** (#MarioBautista)
- extraño** (Strange)
- mañana** (Tomorrow)
- voto** (Vote)
- sí** (Yes)
- mañana** (Morning)
- veo** (I see)
- gusta** (It likes)
- amor** (Love)
- quiero** (I want)
- pinche** (damn)
- amiga** (Friend)
- bonito** (Nice)
- amor** (Love)
- quiero** (I want)
- quieres** (You want)
- ahora** (Now)
- espero** (I expect)
- like** (Like)
- Universidad** (University)
- bueno** (Good)
- sueño** (Dream)
- Eres** (You are)
- Parque** (Park)
- qui** (qui)
- de** (of)
- México** (Mexico)
- gracias** (Thank you)
- fin** (End)
- foto** (Photo)
- jajaja** (jajaja)
- México;** (Mexico;)
- love** (Love)
- Jajaja** (jajaja)
- poder** (Power)

•

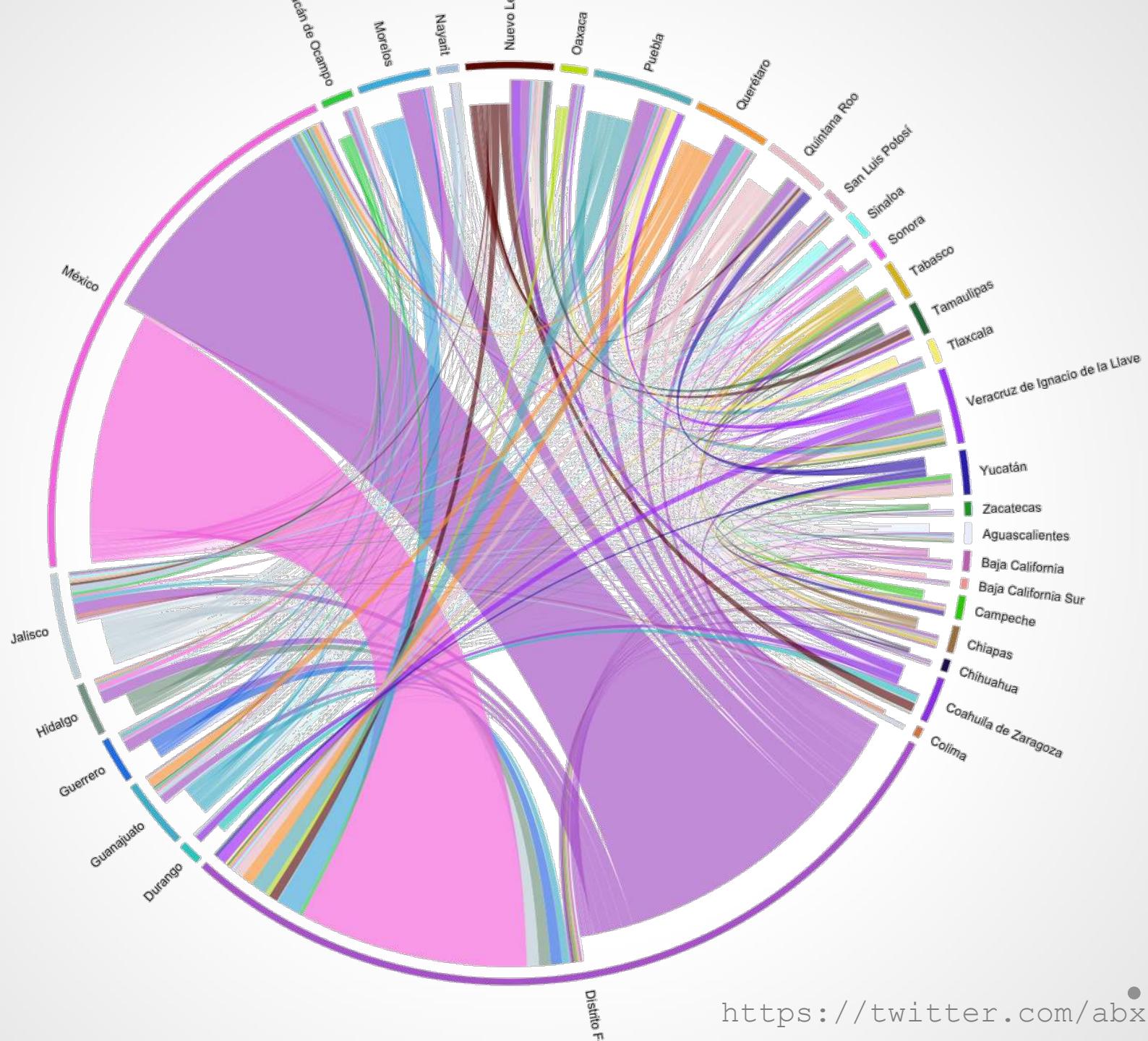
¿Cómo nos desplazamos mientras tuiteamos?

Gráfica de Movilidad

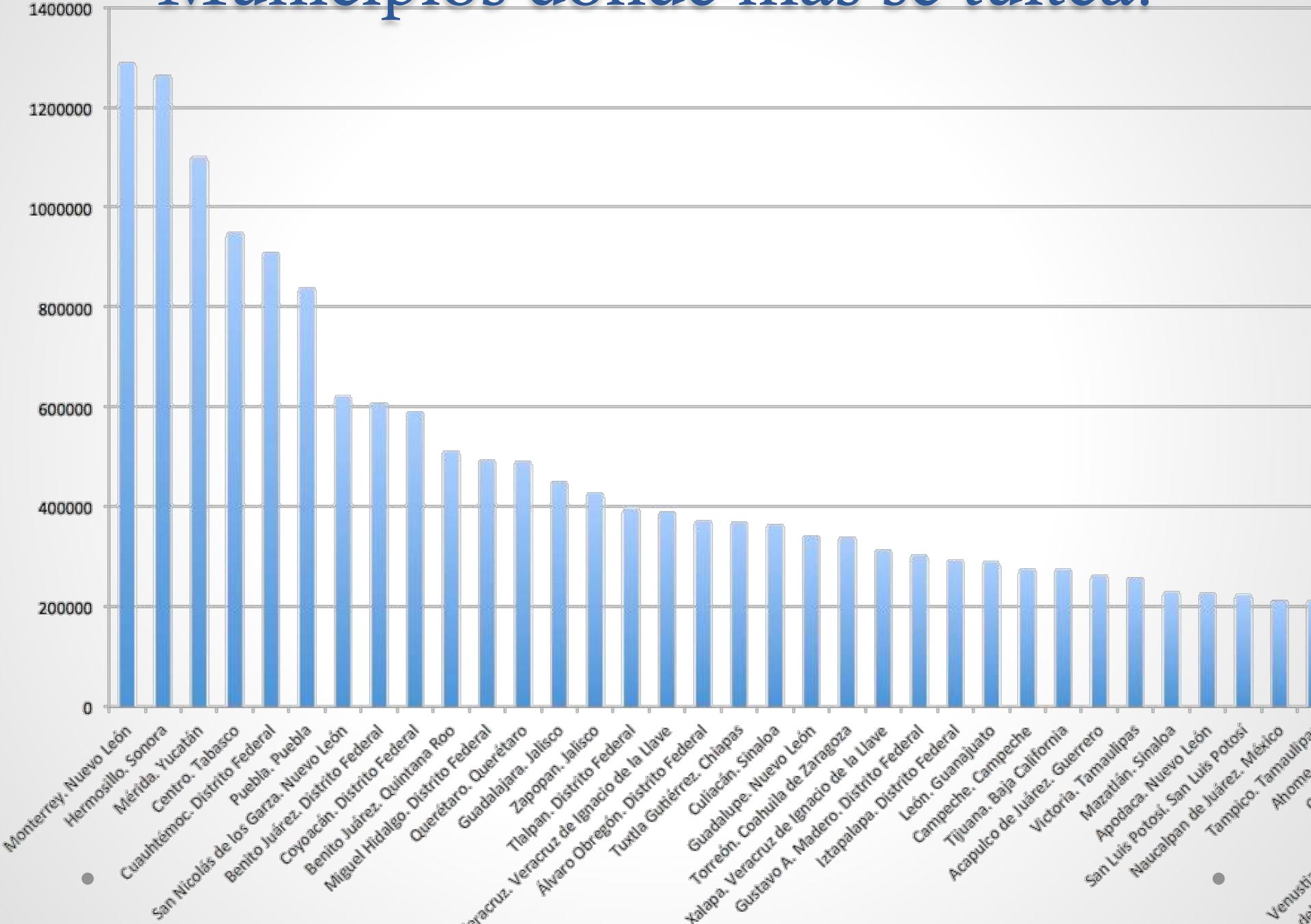


```
library(circlize)
testados = read.table("/abxda/TransladosConDFMexMUNICIPAL.csv", sep=";",
header=TRUE, stringsAsFactors = FALSE, quote = " ")
m = table(testados$estadoorigen, testados$estadodestino)
states = union(rownames(m), colnames(m))
circos.clear()
par(mar = c(1, 1, 1, 1))
chordDiagram(m, directional = TRUE, transparency = 0.3, annotationTrack = "grid",
annotationTrackHeight = 0.01,
preAllocateTracks = 1)

for(si in get.all.sector.index()) {
  xlim = get.cell.meta.data("xlim", sector.index = si, track.index = 1)
  ylim = get.cell.meta.data("ylim", sector.index = si, track.index = 1)
  circos.text(mean(xlim), ylim[1], si, facing = "clockwise", adj = c(0, 0.5),
             niceFacing = TRUE, cex = 0.9, col = "black", sector.index = si,
track.index = 1)
}
```



Municipios donde más se tuitea.



Twitter-Bienestar Subjetivo.

Investigación para la utilización de Twitter: Fuente de datos



- Estructura del tuit
- Disponibilidad
- aleatorización
- filtros georreferenciados

Estudio en otros países

“Análisis de sentimiento” Universidad de Pensilvania

“Mood of the Nation” de los Británicos

“Big Data and Official Statistics” de los Holandeses

“Taller de Análisis de Sentimiento 2013” de la SEPLN

Métodos de clasificadores

Naive Bayes, Support Vector Machines (SVM)

KNN

Word Count

Listas de Palabras y diccionarios utilizados en los ejercicios de análisis de sentimientos

Spanish Emotion Lexicon (SEL)KNN

AFINN

WordNet

ANEW <https://twitter.com/abxda>

Twitter-Bienestar Subjetivo.

Proceso de análisis de tuits seleccionado en INEGI:

- No se utilizará el contabilizar y calificar palabras sueltas y tokenizadas (2 o 3 palabras juntas)
- Se probaran métodos supervisados de aprendizaje
- Manualmente se califica el sentimiento y se clasifica el tema de un conjunto de tuits (conjunto de entrenamiento)
- El conjunto de entrenamiento se utiliza para “enseñarle” al sistema a reconocerlos y a utilizarlos por similitudes para calificar y clasificar el resto de los tuits.



Twitter-Bienestar Subjetivo.

Para generar nuestro conjunto de entrenamiento se desarrolló una aplicación para calificar el sentimiento de los tuits en positivo, negativo o neutro, y clasificarlos en varios temas.



<http://cienciadedatos.inegi.org.mx/pioanalisis>



<https://twitter.com/abxda>



Acerca del proyecto

Bienvenido

Ayudanos a clasificar tuits 

Completa el siguiente formulario para continuar...

Aguascalientes

Masculino

36

Maestría



881



Comenzar

<http://cienciadedatos.inegi.org.mx/pioanalisis>



0 de 20 - nivel 0

Y ahí... entre todos tus gustos raros estaba yo.

¿El tema del tuit  es?

Política

Cultural / Entretenimiento

Deporte

Escolar / Laboral

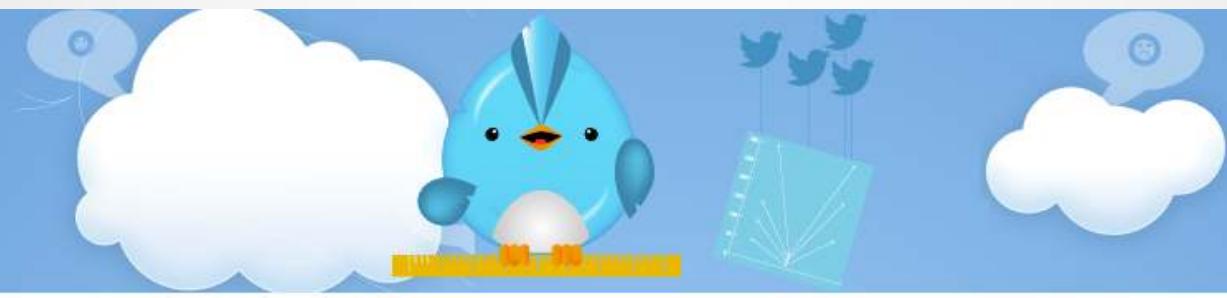
Personal

Ni idea

¿El tuitero se sentía?



<https://twitter.com/abxda>



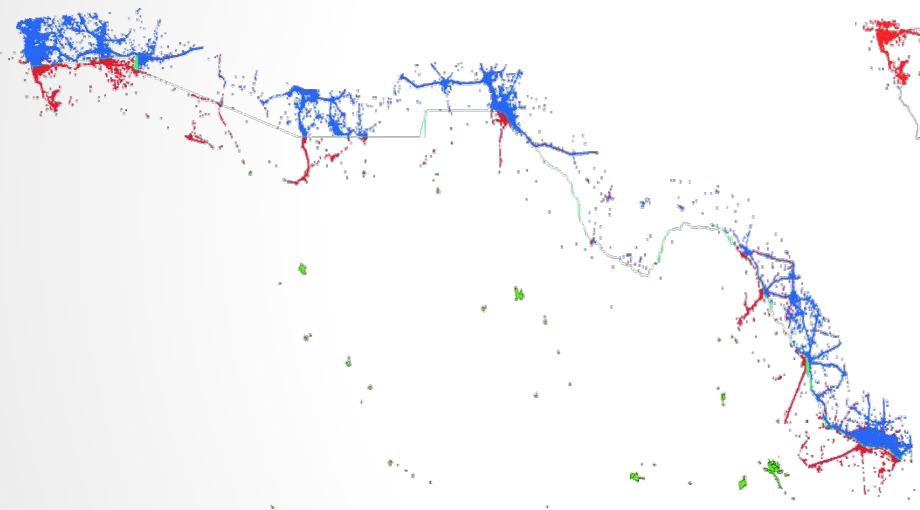
¿Cuántas veces han entrado a Pio Análisis, por estado?



Estado	Entradas
Son.	310
Mex.	129
Sin.	95
Nl.	91
Dgo.	85
Jal.	81
Mich.	78
Ags.	75
Chih.	73
Pue.	66
Tamps.	63
Mor.	52
Df.	41
Tab.	40
Ver.	36
Qroo.	26
Yuc.	24
Slp.	19
Qro.	19
Coah.	18
Gro.	17
Hgo.	16
Bc.	15
Nay.	14
Bcs.	13
Gto.	12
Tlax.	11
Chis.	10

Estudios de movilidad.

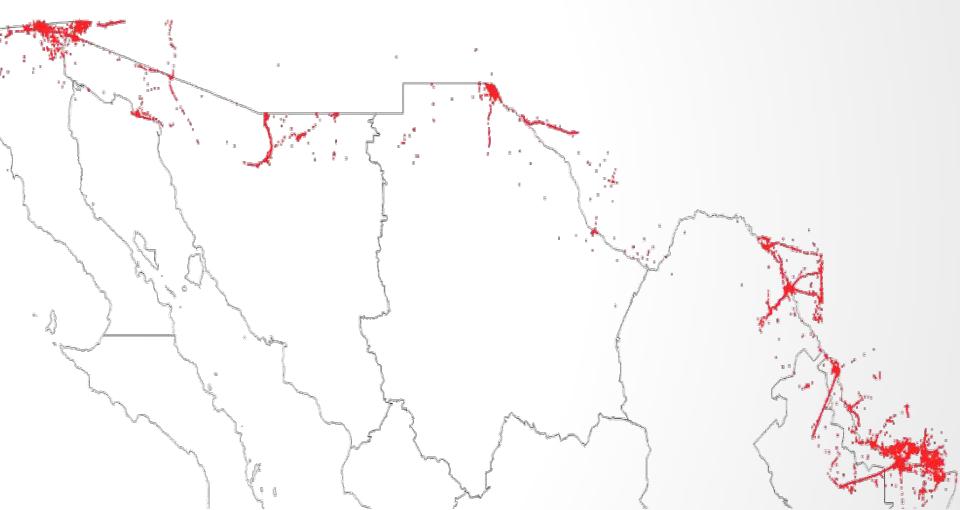
Exploración para el desarrollo de una metodología de análisis para medir la movilidad transfronteriza con los tuits georreferenciados.



Actividad de los tuiteros en
la frontera

Azul =tuiteros de origen EUA

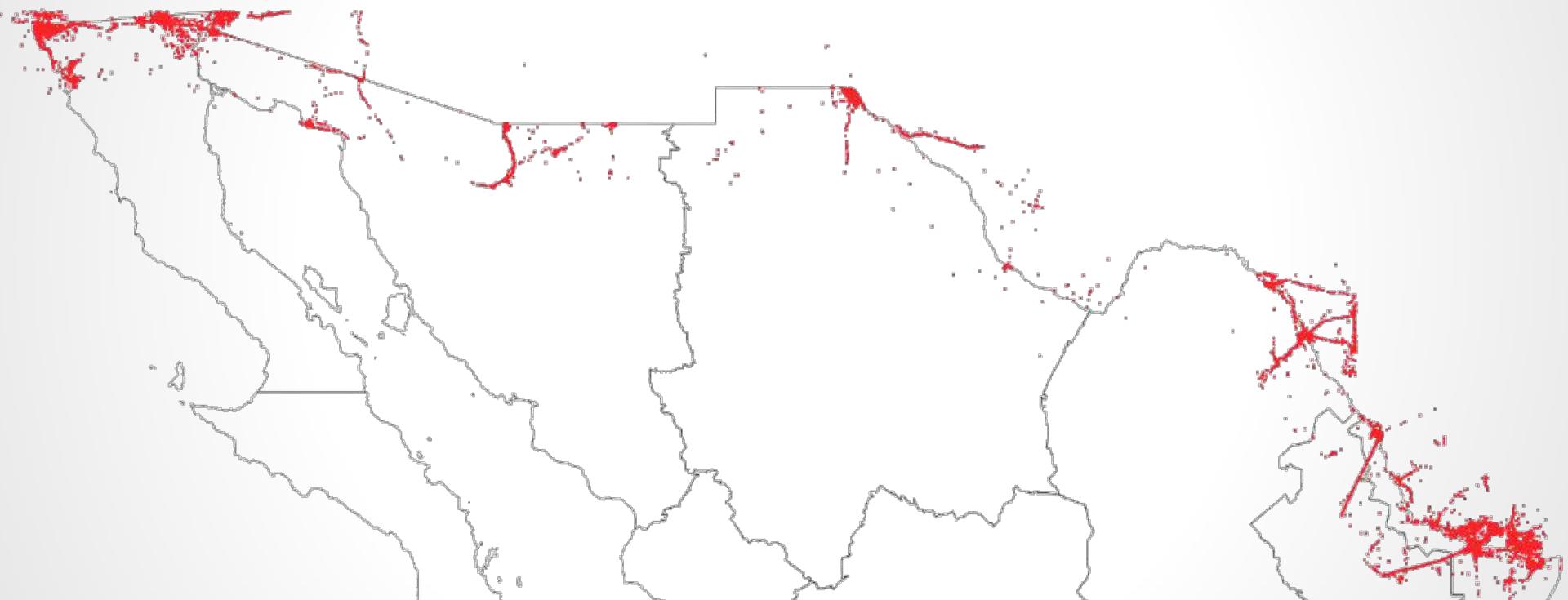
Rojo=tuiteros de origen MX.

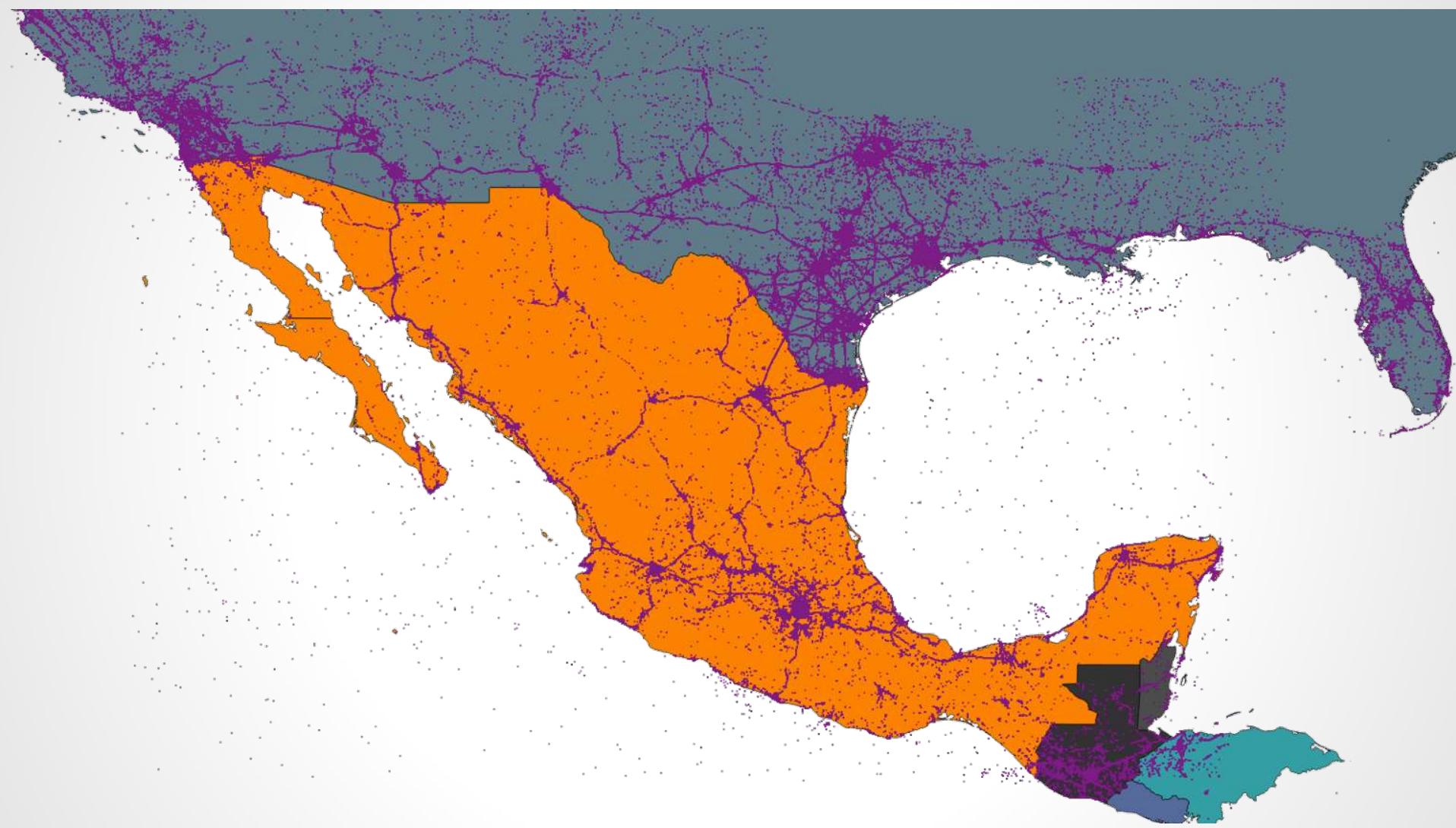


Actividad solamente de
tuiteros MX

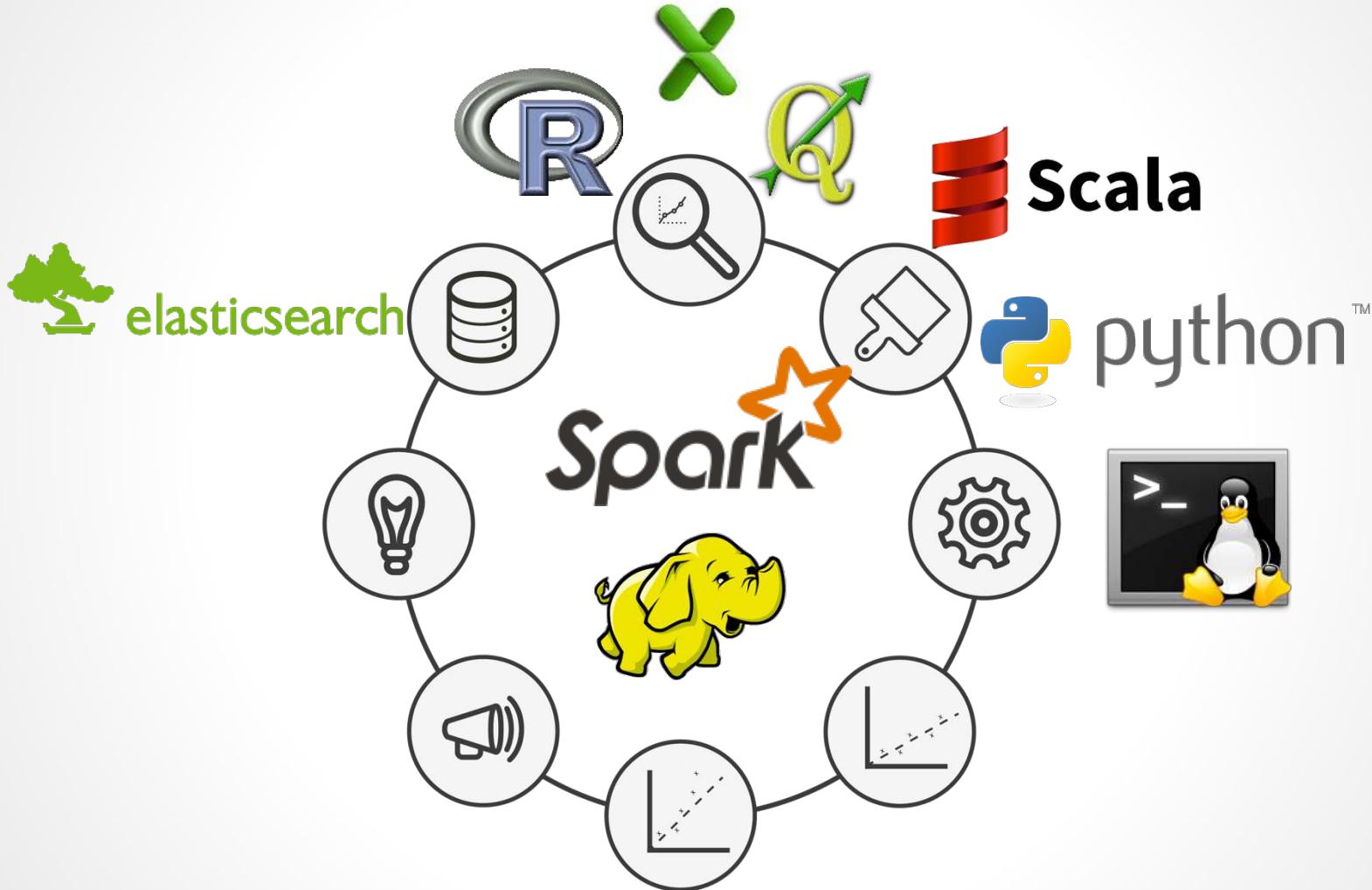
<https://twitter.com/abxda>

Actividad solamente de tuiteros MX

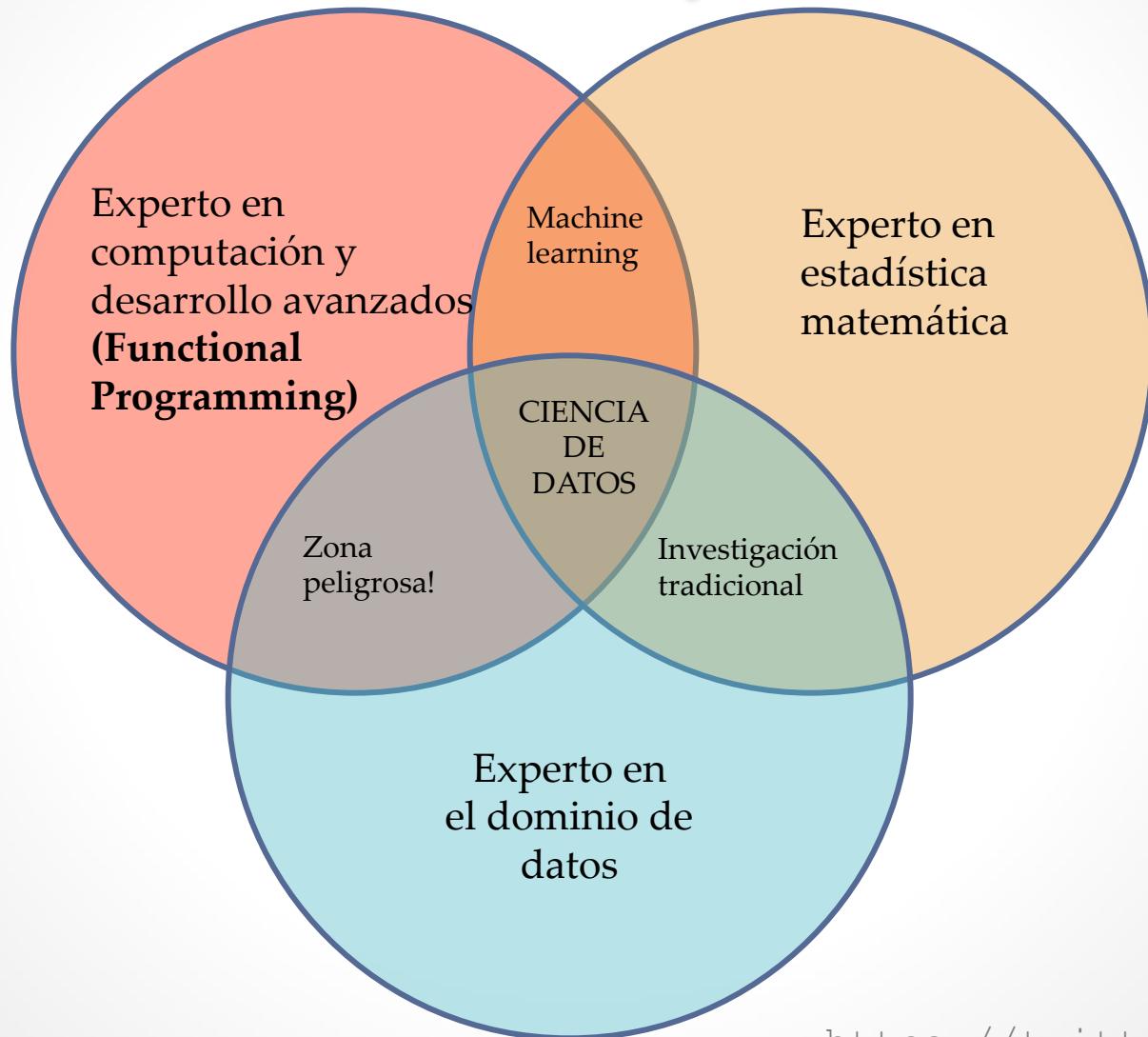




Herramientas

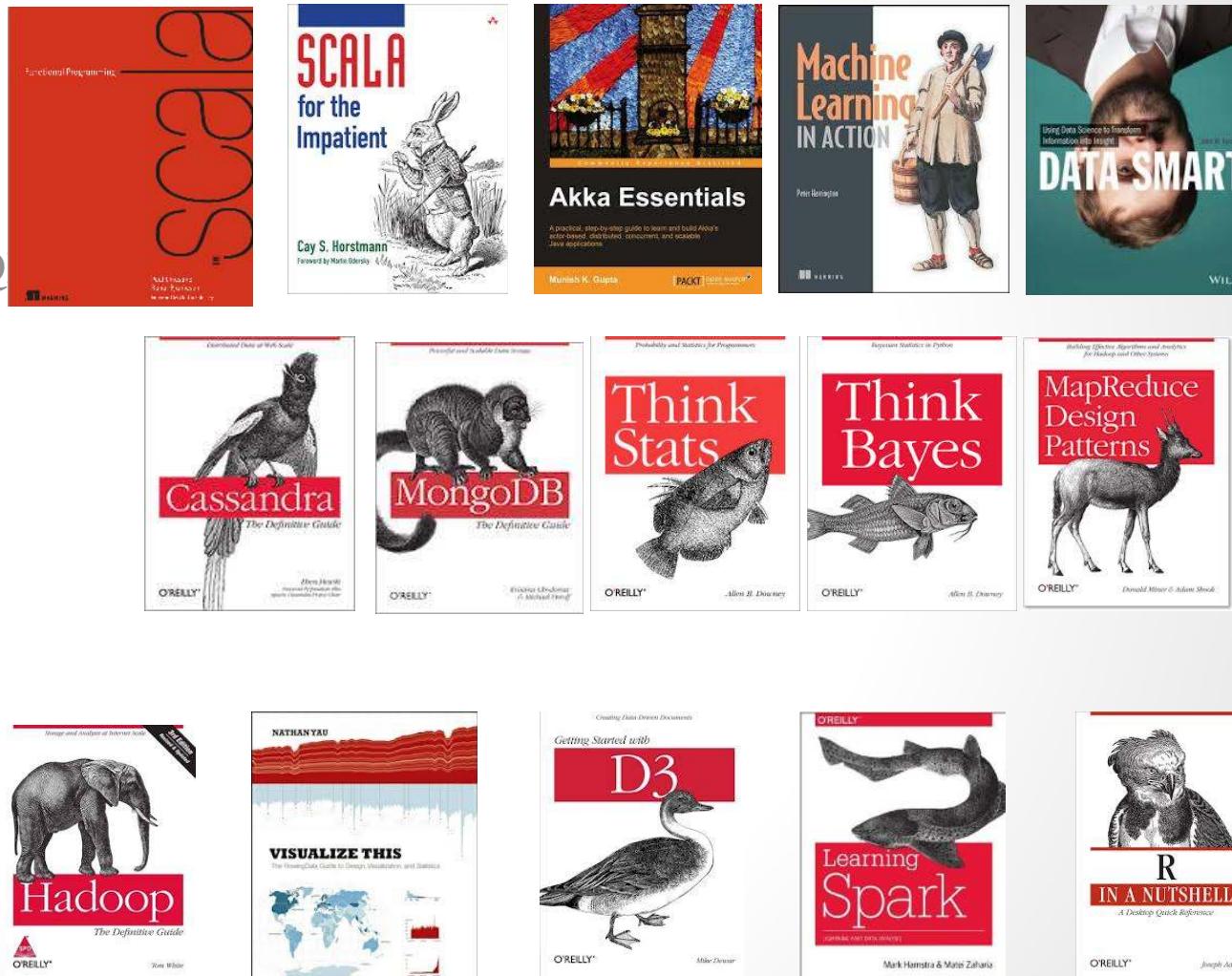


Los Retos: Infraestructura y Personal



- Programación funcional
 - Scala
 - Akka
- Estadística
 - Probabilidad y Estadística
 - Muestreo
 - Machine Learning
 - R
- Almacenes de Datos NoSQL
 - Cassandra
 - MongoDB
 - Hbase
 - ElasticSearch
- Plataformas Big Data
 - Hadoop
 - Spark
- Visualización de Datos
 - D3.js

La tarea



<https://twitter.com/abxda>



Abel Alejandro Coronado Iruegas



@abxda