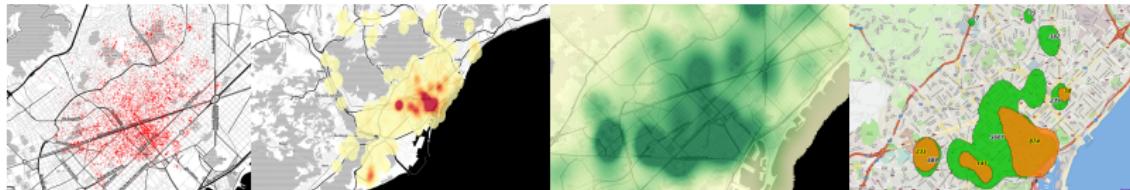


# Analysing GeoLocated Data from Social Media with QGIS

Joana Simões <sup>1</sup>

<sup>1</sup>Bdigital, CASA, CICS.NOVA

March 22, 2015



# Table of Contents

- 1 Introduction
- 2 Getting the Tweets Coordinates
- 3 Importing the Tweets into QGIS
- 4 Creating an Heatmap
- 5 Creating Clusters
- 6 3D Viz

## What do we want to do?

- Nowadays social media is a major source for information sharing. In some cases the user also shares some attributes, such as **geolocation**. By using this information as a proxy for human presence, and with the adequate methods, we are able to provide powerful representations of the distribution of social media users within the territory.

## What do we want to do?

- Nowadays social media is a major source for information sharing. In some cases the user also shares some attributes, such as **geolocation**. By using this information as a proxy for human presence, and with the adequate methods, we are able to provide powerful representations of the distribution of social media users within the territory.
- Due to its willingness in sharing data, Twitter has been a prime *playground*, for researchers and practitioners around the world.

## What do we want to do?

- Nowadays social media is a major source for information sharing. In some cases the user also shares some attributes, such as **geolocation**. By using this information as a proxy for human presence, and with the adequate methods, we are able to provide powerful representations of the distribution of social media users within the territory.
- Due to its willingness in sharing data, Twitter has been a prime *playground*, for researchers and practitioners around the world.
- The objective of this workshop is to provide a workflow for enhancing the visualization of geolocated Tweets.

Getting the Tweets Coordinates  
Importing the Tweets into QGIS  
Creating an Heatmap  
Creating Clusters  
3D Viz

## How are we going to do it?

## How are we going to do it?

- Pull data stored in a NoSQL database.

## How are we going to do it?

- Pull data stored in a NoSQL database.
- Create an Heatmap expressing Tweet density.

## How are we going to do it?

- Pull data stored in a NoSQL database.
- Create an Heatmap expressing Tweet density.
- Produce clusters from the heatmap.

## How are we going to do it?

- Pull data stored in a NoSQL database.
- Create an Heatmap expressing Tweet density.
- Produce clusters from the heatmap.
- Produce a 3D visualization, on a webpage.

Getting the Tweets Coordinates  
Importing the Tweets into QGIS  
Creating an Heatmap  
Creating Clusters  
3D Viz

# Objective

# Stack

- MongoDB
- QGIS + Python Plugins

# Stack

- MongoDB
- QGIS + Python Plugins
  - Heatmap

# Stack

- MongoDB
- QGIS + Python Plugins
  - Heatmap
  - OpenLayers

# Stack

- MongoDB
- QGIS + Python Plugins
  - Heatmap
  - OpenLayers
  - Qgis2threejs

# Approach

This will be a hands-on workshop.

## Approach

This will be a hands-on workshop.

- I will introduce each one of the micro-tasks, and will do it while you watch.

# Approach

This will be a hands-on workshop.

- I will introduce each one of the micro-tasks, and will do it while you watch.
- Then I will provide help to everyone who needs it to complete the task.

## Approach

This will be a hands-on workshop.

- I will introduce each one of the micro-tasks, and will do it while you watch.
- Then I will provide help to everyone who needs it to complete the task.
- **For the impatient:** If you don't want to wait for everyone to complete, you can attempt to complete the exercises on your own ;-) View the complete presentation here:

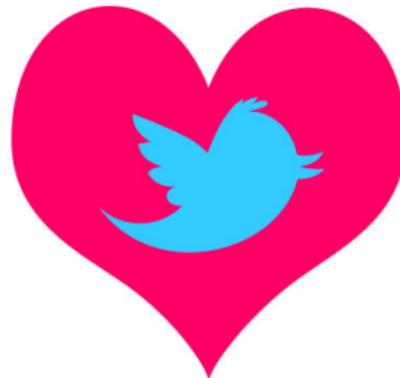
## Approach

This will be a hands-on workshop.

- I will introduce each one of the micro-tasks, and will do it while you watch.
- Then I will provide help to everyone who needs it to complete the task.
- **For the impatient:** If you don't want to wait for everyone to complete, you can attempt to complete the exercises on your own ;-) View the complete presentation here:
- **For the guru:** If you have finished everything and are bored, you may try to complete the tutorials or read the references, on the last page;-)

# Twitter Data

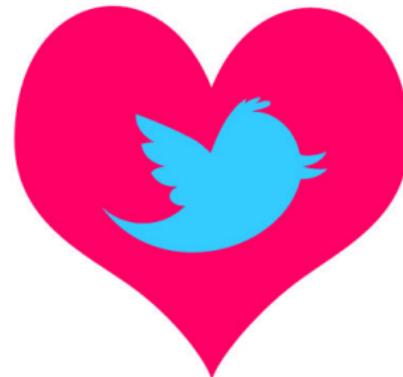
- Users on Twitter generate over 400 million tweets everyday.
- A proportion of these tweets is available to researchers and practitioners through public APIs, free of any charges.



Why we love Twitter

# Twitter Data

- Users on Twitter generate over 400 million tweets everyday.
- A proportion of these tweets is available to researchers and practitioners through public APIs, free of any charges.
- Approximately 1% of all Tweets published on Twitter are geolocated.



Why we love Twitter

# Twitter APIs

# Twitter APIs

- REST APIs.

# Twitter APIs

- REST APIs.
- Stream APIs.

# Twitter APIs

- REST APIs.
- Stream APIs.
  - User streams.

# Twitter APIs

- REST APIs.
- Stream APIs.
  - User streams.
  - Site streams.

# Twitter APIs

- REST APIs.
- Stream APIs.
  - User streams.
  - Site streams.
  - Public Streams.

# Twitter APIs

- REST APIs.
- Stream APIs.
  - User streams.
  - Site streams.
  - Public Streams.
- These APIs are accessed only via authenticated requests (OAuth).

# Twitter APIs

- REST APIs.
- Stream APIs.
  - User streams.
  - Site streams.
  - Public Streams.
- These APIs are accessed only via authenticated requests (OAuth).
- Access to APIs is also limited to a specific number of requests within a time window (rate limit).

# Twitter APIs

- REST APIs.
- Stream APIs.
  - User streams.
  - Site streams.
  - Public Streams.
- These APIs are accessed only via authenticated requests (OAuth).
- Access to APIs is also limited to a specific number of requests within a time window (rate limit).
- Responses from Twitter are in JSON format.

# JSON in a Nutshell



## JSON

```
{  
  "siblings": [  
    {"firstName": "Anna", "lastName": "Clayton"},  
    {"lastName": "Alex", "lastName": "Clayton"}  
  ]  
}
```

## XML

```
<siblings>  
  <sibling>  
    <firstName>Anna</firstName>  
    <lastName>Clayton</lastName>  
  </sibling>  
  <sibling>  
    <firstName>Alex</firstName>  
    <lastName>Clayton</lastName>  
  </sibling>  
</siblings>
```

# Mongo loves JSON

- NoSQL Database.
- Document-Oriented Storage.



## Hands-on

- Objective: Connect to the tweets database and view one record.

## Hands-on

- Objective: Connect to the tweets database and view one record.
- Tool: Mongodb command line client.
- Database properties:
  - host: 54.72.72.228 (on aws).
  - database: tweets\_workshop
  - username: workshop
  - password: geohipster
  - collection: tweets
- Reference: <http://docs.mongodb.org/manual/#>.

## Hands-on

- Objective: Connect to the tweets database and view one record.
- Tool: Mongodb command line client.
- Database properties:
  - host: 54.72.72.228 (on aws).
  - database: tweets\_workshop
  - username: workshop
  - password: geohipster
  - collection: tweets
- Reference: <http://docs.mongodb.org/manual/#>.
- `mongo -host 54.72.72.228 tweets_workshop -u workshop -p geohipster`
- `db.tweets.find().limit( 1 )`

## Hands-on (cont.)

- Objective: Export tweets coordinates into a csv file.

## Hands-on (cont.)

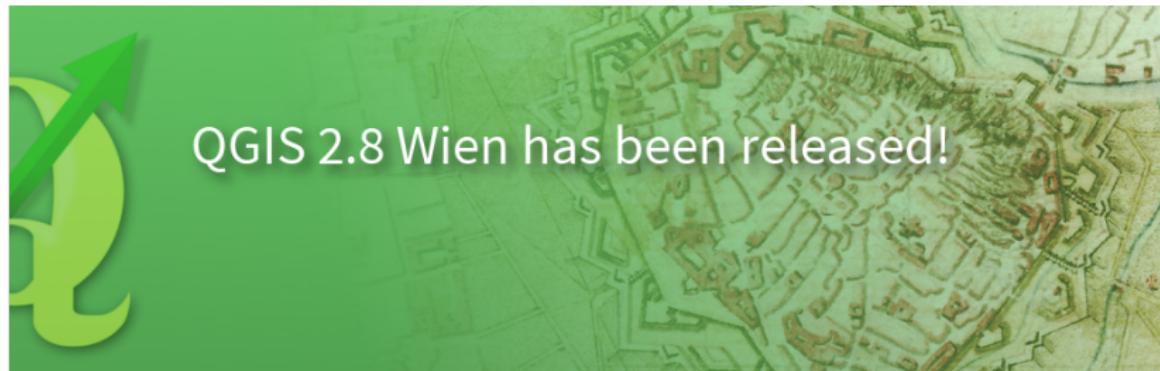
- Objective: Export tweets coordinates into a csv file.
- Take notice of the long, lat fields.
- Exit the client.
- Use mongoexport to write the values on a text file.
- Reference: <http://docs.mongodb.org/manual/core/import-export/>.

## Hands-on (cont.)

- Objective: Export tweets coordinates into a csv file.
- Take notice of the long, lat fields.
- Exit the client.
- Use mongoexport to write the values on a text file.
- Reference: <http://docs.mongodb.org/manual/core/import-export/>.
- ```
mongoexport -host 54.72.72.228 -u workshop -p geohipster
-db tweets_workshop -collection tweets -csv -out
out_tweets.csv -fields
geoLocation.longitude,geoLocation.latitude -query
'geoLocation: $ne: null'
```
- View the exported file:e.g.: `cat /tmp/out/out_tweets.csv`

# QGIS

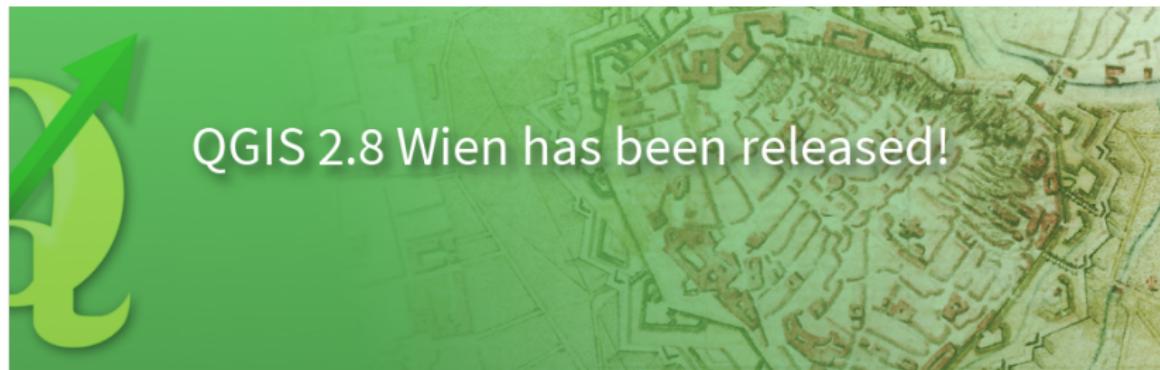
QGIS is a Free and Open-Source GIS System:



# QGIS

QGIS is a Free and Open-Source GIS System:

- Cross-platform (Windows, Mac, Linux and Android).



# QGIS

QGIS is a Free and Open-Source GIS System:

- Cross-platform (Windows, Mac, Linux and Android).
- Easy to use.



# QGIS

QGIS is a Free and Open-Source GIS System:

- Cross-platform (Windows, Mac, Linux and Android).
- Easy to use.
- Great community support.



# QGIS

QGIS is a Free and Open-Source GIS System:

- Cross-platform (Windows, Mac, Linux and Android).
- Easy to use.
- Great community support.
- Latest version has been translated into 46 languages.



# Bug Affecting the Heatmap Plugin

Problem with QGIS Heatmap Plugin - Geographic Information Systems Stack Exchange - Mozilla Firefox <2>

gis.stackexchange.com/questions/139829/problem-with-qgis-heatmap-plugin

Most Visited Getting Started Spatial Framework Fo...

## Geographic Information Systems

Questions Tags Users Badges Unanswered Ask Question

### Problem with QGIS Heatmap Plugin

I am experiencing some issues with the Heatmap Plugin (0.2) on QGIS 2.8.1, and I wanted to double-check if I am doing something wrong, or if there is some sort of bug lying around. As a background, I have to mention I was able to produce heatmaps before.

I have a Shapefile with tweet locations.

Asked Today  
Viewed 7 Times

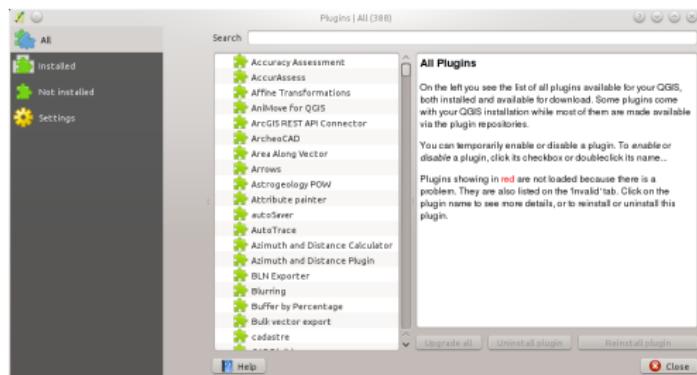
Featured on Meta  
 Let's improve our site navigation

Linked  
7 How to use Heatmap plugin and its output?

Related  
3 heatmap plugin slow/unresponsive in QGIS 1.8  
1 Stats problem: QGIS heatmap  
3 Which parameter values should I use with QGIS Heatmap plugin?  
4 How to create a Human

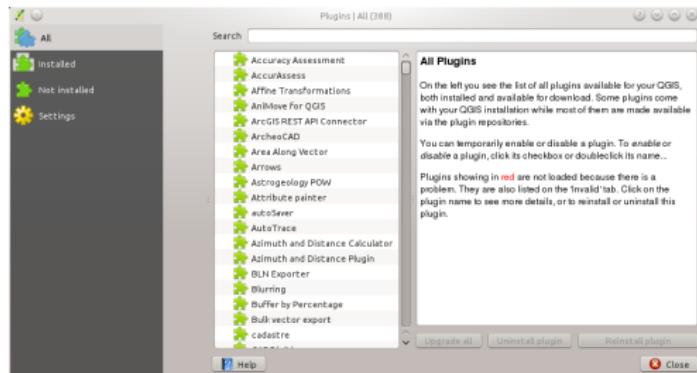
# There is a Plugin for Everything

- The core of QGIS is extended through user-submited Python plugins.



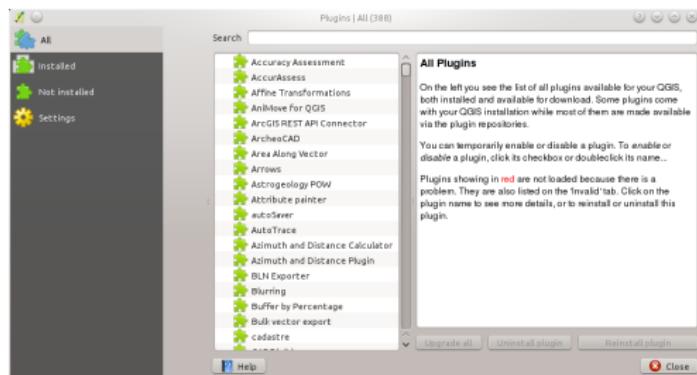
# There is a Plugin for Everything

- The core of QGIS is extended through user-submited Python plugins.



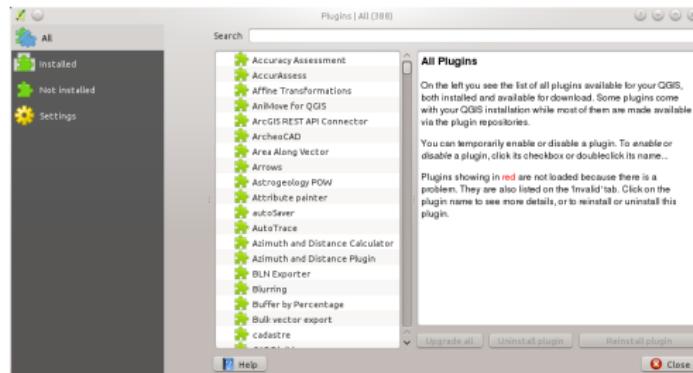
# There is a Plugin for Everything

- The core of QGIS is extended through user-submited Python plugins.
- Plugins can be easily installed through the plugin manager.



# There is a Plugin for Everything

- The core of QGIS is extended through user-submited Python plugins.
- Plugins can be easily installed through the plugin manager.
- In this workshop we will use three plugins (one is already installed).



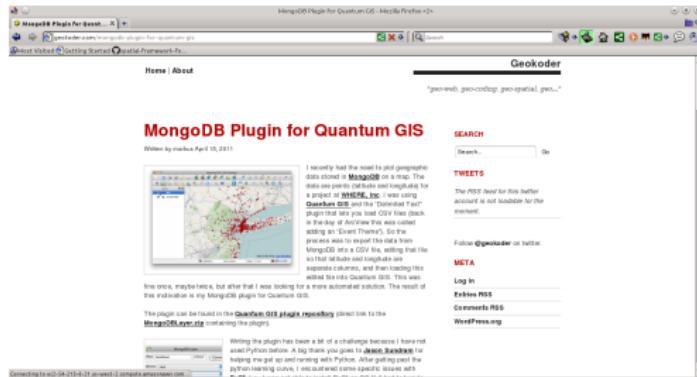
# Isn't there a Plugin to connect to MongoDB?

- There was a plugin written in 2011.



# Isn't there a Plugin to connect to MongoDB?

- There was a plugin written in 2011.



# Isn't there a Plugin to connect to MongoDB?

- There was a plugin written in 2011.
- There seemed to be some issues with it, and sadly it hasn't been ported to the latest versions of QGIS.



# Isn't there a Plugin to connect to MongoDB?

- There was a plugin written in 2011.
- There seemed to be some issues with it, and sadly it hasn't been ported to the latest versions of QGIS.
- The project appears to be now dead.



# Isn't there a Plugin to connect to MongoDB?

- There was a plugin written in 2011.
- There seemed to be some issues with it, and sadly it hasn't been ported to the latest versions of QGIS.
- The project appears to be now dead.
- The community would really welcome a new attempt for a MongoDB connector on QGIS.



## Hands-on

- Objective: Display the tweets coordinates in QGIS, with background map.
- Tool: QGIS + OpenLayers plugin.
- Steps:



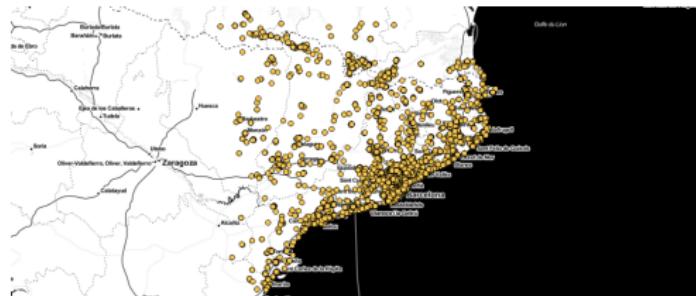
## Hands-on

- Objective: Display the tweets coordinates in QGIS, with background map.
- Tool: QGIS + OpenLayers plugin.
- Steps:
  - Start QGIS and create a project.
  - Download and enable the *OpenLayers* plugin (Plugins – > Manage and Install Plugins).
  - Choose a background map (e.g.: OSM, Bing Maps, etc).
  - Import the json file, using the Text Layer Importer (Layer – > Add Layer – > Add Delimited Text Layer).



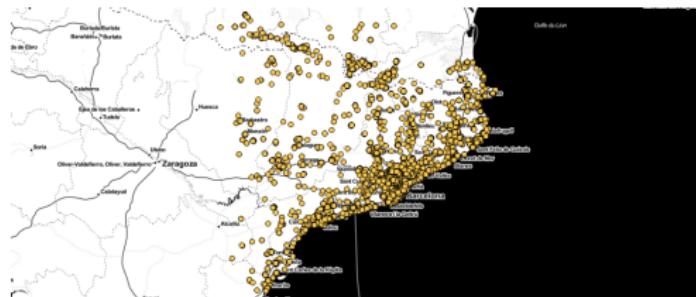
## Hands-on (cont.)

- Objective: Save the tweets from the region of Catalunya, in a Shapefile.
- Steps:



## Hands-on (cont.)

- Objective: Save the tweets from the region of Catalunya, in a Shapefile.
- Steps:
  - Zoom into the desired location using the zoom tool (View – > Zoom in) .
  - Select the visible points using the select tool.
  - Save the selected features in a Shapefile.
    - Pay attention to the crs.
    - Make sure you save **only** the selected features.



# Heatmaps

Heatmaps enhance the visualization of the density distribution of a phenomena (e.g.: crime, accidents, tweets).



# Heatmaps

Heatmaps enhance the visualization of the density distribution of a phenomena (e.g.: crime, accidents, tweets).

- The "heat" in the term refers to the concentration of the geographic entity within any given spot.



# Heatmaps

Heatmaps enhance the visualization of the density distribution of a phenomena (e.g.: crime, accidents, tweets).

- The "heat" in the term refers to the concentration of the geographic entity within any given spot.
- Sometimes they are referred as hot spot mapping or clustering.



# Heatmaps

Heatmaps enhance the visualization of the density distribution of a phenomena (e.g.: crime, accidents, tweets).

- The "heat" in the term refers to the concentration of the geographic entity within any given spot.
- Sometimes they are referred as hot spot mapping or clustering.
- They are great tools to detect spatial patterns.



## Heatmaps (cont.)

- Creating heatmaps involves interpolating discrete points to create a continuous surface known as a **density surface**.

## Heatmaps (cont.)

- Creating heatmaps involves interpolating discrete points to create a continuous surface known as a **density surface**.
- There are three key parameters involved in this calculation:

## Heatmaps (cont.)

- Creating heatmaps involves interpolating discrete points to create a continuous surface known as a **density surface**.
- There are three key parameters involved in this calculation:
  - Cell size of the output raster file.
  - Bandwidth or search radius.
  - Interpolation algorithm.

## Hands-on

- Objective: Create heatmap depicting tweet density.
- Tool: QGIS + Heatmap plugin.
- Steps:

## Hands-on

- Objective: Create heatmap depicting tweet density.
- Tool: QGIS + Heatmap plugin.
- Steps:
  - Ensure the heatmap plugin is enabled.
  - Create the heatmap using the plugin (Raster – > Heatmap).
  - Pay special attention to the following parameters:

## Hands-on

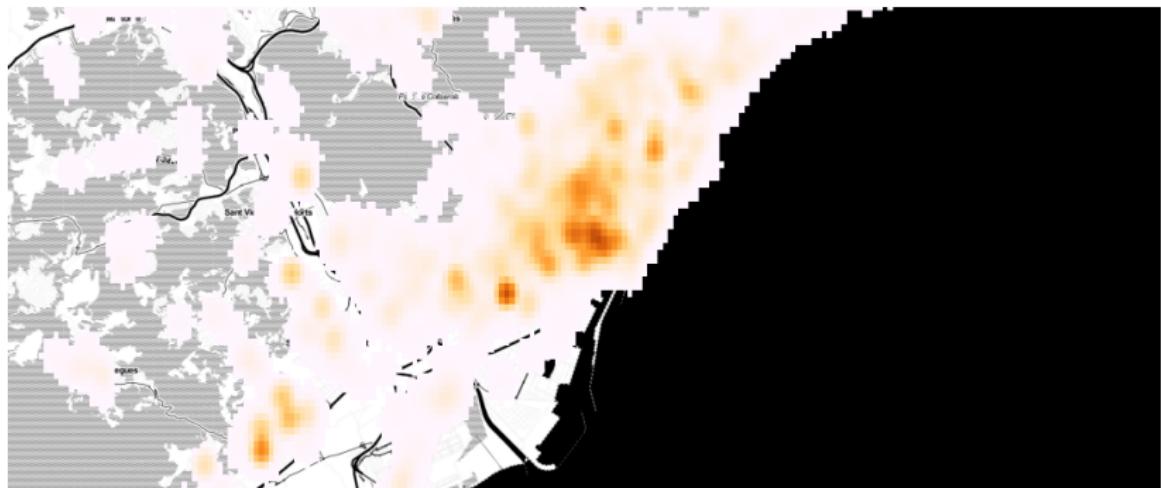
- Objective: Create heatmap depicting tweet density.
- Tool: QGIS + Heatmap plugin.
- Steps:
  - Ensure the heatmap plugin is enabled.
  - Create the heatmap using the plugin (Raster –> Heatmap).
  - Pay special attention to the following parameters:
    - Radius: search radius in meters.
    - Rows and columns: cell size of the output raster.
    - Rows and columns: cell size of the output raster.
    - Kernel shape: QGIS uses one of the following decay algorithms.

## Hands-on

- Objective: Create heatmap depicting tweet density.
- Tool: QGIS + Heatmap plugin.
- Steps:
  - Ensure the heatmap plugin is enabled.
  - Create the heatmap using the plugin (Raster –> Heatmap).
  - Pay special attention to the following parameters:
    - Radius: search radius in meters.
    - Rows and columns: cell size of the output raster.
    - Rows and columns: cell size of the output raster.
    - Kernel shape: QGIS uses one of the following decay algorithms.
  - Customize a color scheme, by right clicking the layer and choosing properties.

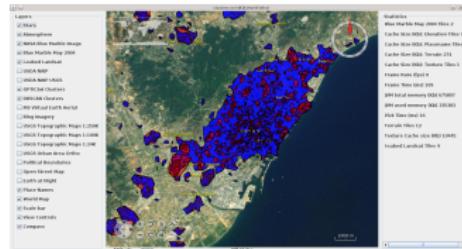
## Hands-on (cont.)

- Suggestion: try different combinations of parameters and different visualizations.
- Tip: Use the identify tool to ensure the maximum is being picked up.



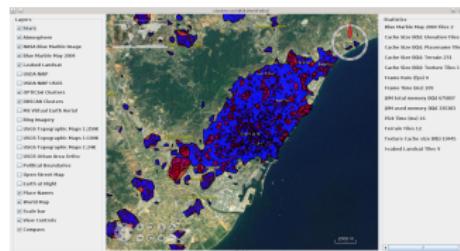
# Clustering

- It is a descriptive data mining technique, often used for dimensionality reduction.



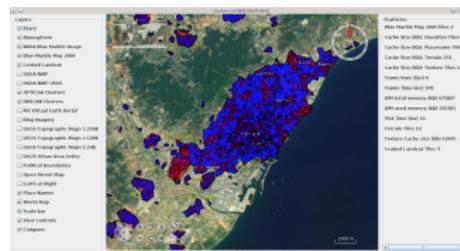
# Clustering

- It is a descriptive data mining technique, often used for dimensionality reduction.
- It corresponds to a family of unsupervised machine learning algorithms.



# Clustering

- It is a descriptive data mining technique, often used for dimensionality reduction.
- It corresponds to a family of unsupervised machine learning algorithms.
- It groups a set of objects in such a way that objects in the same group are more similar to each other, than to object in other groups.



# Clustering

- It is a descriptive data mining technique, often used for dimensionality reduction.
  - It corresponds to a family of unsupervised machine learning algorithms.
  - It groups a set of objects in such a way that objects in the same group are more similar to each other, than to object in other groups.
  - Density based clusters are separated from each other by contiguous regions of low density of objects.



## Clustering (cont.)

- In June 2014, it was released (the first?) QGIS plugin for *spatially constrained clustering*: clusterpy.
  - [https://plugins.qgis.org/plugins/clusterpy\\_qgis\\_plugin/](https://plugins.qgis.org/plugins/clusterpy_qgis_plugin/)
- In this workshop we will use an alternative method to define areas with high density.
  - [http://www.qgistutorials.com/en/docs/creating\\_heatmaps.html](http://www.qgistutorials.com/en/docs/creating_heatmaps.html)

## Hands-on

- Objective: Detect areas with higher density of tweets.
- Tool: QGIS.
- Steps:

## Hands-on

- Objective: Detect areas with higher density of tweets.
- Tool: QGIS.
- Steps:
  - Use the identify tool to establish a threshold: from that value on, pixels will be considered part of an high-density cluster.
  - Use the raster calculator, to select pixels above that threshold (Raster –> Raster Calculator).
  - Example of a select expression: "*heatmap@1*" > 150

## Hands-on

- Objective: Detect areas with higher density of tweets.
- Tool: QGIS.
- Steps:
  - Use the identify tool to establish a threshold: from that value on, pixels will be considered part of an high-density cluster.
  - Use the raster calculator, to select pixels above that threshold (Raster –> Raster Calculator).
  - Example of a select expression: "*heatmap@1*" > 150
  - The new map only contains binary values (1: above the threshold; 0: bellow the threshold).
  - Tip: ensure the legend picks up correctly the values on the dataset.

## Hands-on (cont.)

- Vectorize this raster file (Raster – > Conversion – > Polygonize (Raster to Vector)).

## Hands-on (cont.)

- Vectorize this raster file (Raster – > Conversion – > Polygonize (Raster to Vector)).
- Check out the attribute table of the Shapefile: the polygons have a value of 0, or 1 (right click + Open attribute table).

## Hands-on (cont.)

- Vectorize this raster file (Raster – > Conversion – > Polygonize (Raster to Vector)).
- Check out the attribute table of the Shapefile: the polygons have a value of 0, or 1 (right click + Open attribute table).
- If we filter out the polygons with value 0, we will have the areas with high density.

## Hands-on (cont.)

- Vectorize this raster file (Raster –> Conversion –> Polygonize (Raster to Vector)).
- Check out the attribute table of the Shapefile: the polygons have a value of 0, or 1 (right click + Open attribute table).
- If we filter out the polygons with value 0, we will have the areas with high density.
  - Select by expression: " $DN = 1$ ".

## Hands-on (cont.)

- Vectorize this raster file (Raster –> Conversion –> Polygonize (Raster to Vector)).
- Check out the attribute table of the Shapefile: the polygons have a value of 0, or 1 (right click + Open attribute table).
- If we filter out the polygons with value 0, we will have the areas with high density.
  - Select by expression: " $DN = 1$ ".
- Save the selection as a new Shapefile.

## Hands-on (cont.)

- Vectorize this raster file (Raster –> Conversion –> Polygonize (Raster to Vector)).
- Check out the attribute table of the Shapefile: the polygons have a value of 0, or 1 (right click + Open attribute table).
- If we filter out the polygons with value 0, we will have the areas with high density.
  - Select by expression: " $DN = 1$ ".
- Save the selection as a new Shapefile.
- Enhancement: count the number of tweets inside each polygon (Vector –> Analysis Tools –> Count Points in Polygon).

Introduction

Getting the Tweets Coordinates

Importing the Tweets into QGIS

Creating an Heatmap

Creating Clusters

# TODO

## Hands-on

- Objective: create webpage with 3d visualization of the clusters.
- Tool: QGIS + Qgis2threejs plugin.
- Steps:

## Hands-on

- Objective: create webpage with 3d visualization of the clusters.
- Tool: QGIS + Qgis2threejs plugin.
- Steps:
  - Ensure that Qgis2threejs is installed and enabled.

## Hands-on

- Objective: create webpage with 3d visualization of the clusters.
- Tool: QGIS + Qgis2threejs plugin.
- Steps:
  - Ensure that Qgis2threejs is installed and enabled.
  - Prepare the scene you want to render (e.g.: visible layers, styles, etc).

## Hands-on

- Objective: create webpage with 3d visualization of the clusters.
- Tool: QGIS + Qgis2threejs plugin.
- Steps:
  - Ensure that Qgis2threejs is installed and enabled.
  - Prepare the scene you want to render (e.g.: visible layers, styles, etc).
  - Run Qgis2threejs (Web – > Qgis2threejs – > Qgis2threejs)

## Hands-on

- Objective: create webpage with 3d visualization of the clusters.
- Tool: QGIS + Qgis2threejs plugin.
- Steps:
  - Ensure that Qgis2threejs is installed and enabled.
  - Prepare the scene you want to render (e.g.: visible layers, styles, etc).
  - Run Qgis2threejs (Web – > Qgis2threejs – > Qgis2threejs)
  - Costumize the visualization. Some things to pay attention to:

## Hands-on

- Objective: create webpage with 3d visualization of the clusters.
- Tool: QGIS + Qgis2threejs plugin.
- Steps:
  - Ensure that Qgis2threejs is installed and enabled.
  - Prepare the scene you want to render (e.g.: visible layers, styles, etc).
  - Run Qgis2threejs (Web – > Qgis2threejs – > Qgis2threejs)
  - Costumize the visualization. Some things to pay attention to:
    - Enable the layers you want to extrude (in this case, the clusters).
    - The height value should be read fromthe count field (PNTCNT).
    - You may want to exagerate the vertical scale by a factor.
    - The output HTML file path defines where you will save the result webpage (and associated files).

## References

- Aji, A., Wang, F., Vo, H., Lee, R., Liu, Q., Zhang, X. and Saltz, J." Hadoop-GIS: A High Performance Spatial Data Warehousing System over MapReduce". Proceedings of the VLDB Endowment International Conference on Very Large Data Bases, 6(11), p1009. (2013)
- Cuesta, H. "Practical Data Analysis". Packt Publishing (2013)
- Dunning, T. and Friedman, E. "Time Series Databases: New Ways to Store and Access Data". O'Reilly Media; 1 edition (October, 2014)
- Myatt, G and Johnson, W. "Making Sense Of Data I: A Practical Guide to Exploratory Data Analysis and Data Mining". O'Reilly : 2 edition (2014).

Introduction

Getting the Tweets Coordinates

Importing the Tweets into QGIS

Creating an Heatmap

Creating Clusters

# Tutorials