

Visualization of a hashtag's activity in real time

BS.C Thesis

Davvetas Athanasios

Technological Educational Institute of Athens

cs121018@teiath.gr

Tools used

During this project we used open source tools and libraries like:

- Netbeans IDE 8.0.2
- Glassfish server 4.1
- D3.js
- Apache commons for REST API calls
- JSON simple 1.1
- Debian based OS(Crunchbang plus plus)

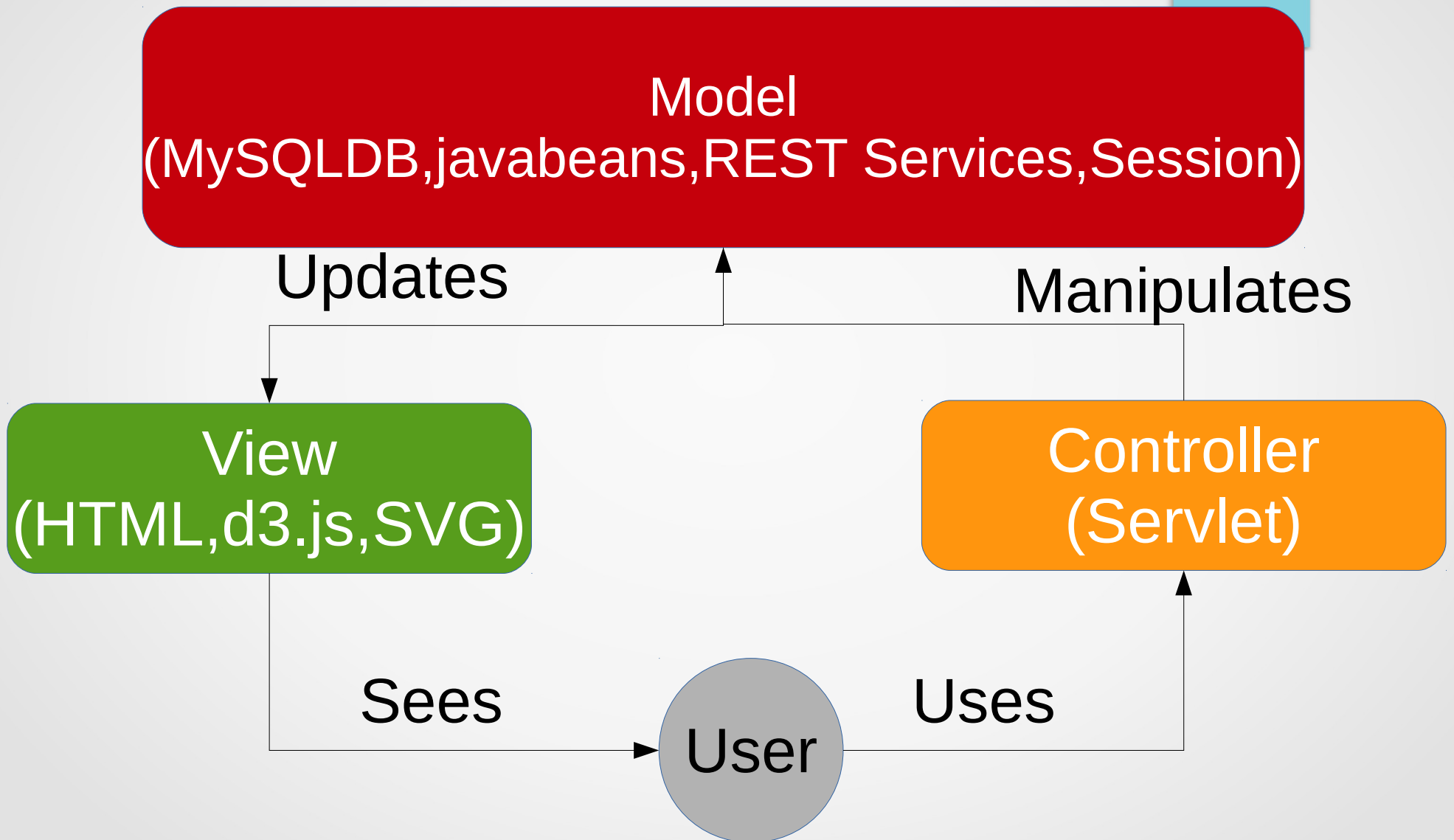
Thesis Objective

- Visualizing hashtag activity one week past today using choropleth map and globe
- Visualizing hashtag activity in real time using bar chart
- Predicting upcoming hashtags
- Identifying emotion and opinions about a hashtag through text analysis

Implementation

- MVC model
- REST services in order to make twitter api calls and process data
- User actions implemented with java beans
- Data visualization with d3.js

Thesis basic components



Twitter & hashtags

- Micro blog service allowing users to post short messages about 140 characters long
- These posts are called “tweets”
- Twitter also allows users to label their posts with a keyword in order to categorize them
- This keyword is called “hashtag” and its syntax is #<word or phrase with no spaces>
- For example : #fosscom2015,#fosscomm,#opensource
κλπ

Activity visualization procedure

- User inputs a keyword or hashtag that wants to observe
- Our application calls twitter API in order to get posts that contain said hashtag or keyword
- Find user's country of origin
- Visualize them using a choropleth map

Problems during implementation

- Converting hashtag or keyword into twitter api query
- Finding user's location
- Data Visualization

Solution(1/5)

Input verification:

- Using `java.net.URLEncoder` in order to convert other languages than English to UTF-8
- Servlet configuration in order to encode response and request to UTF-8 (`setCharacterEncoding("UTF-8")`)
- Using `java.net.URLDecoder` to convert back

Solution(2/5)

- Twitter api response field “location/geo” usually null
- Instead parse users profile location
- Users profile location is user input meaning it is not always easy to identify the country of origin
- Example: ~|Worldwide|~ , Anywhere , Neverland κλπ
- Abbreviations: NY,ABQ , AUS κλπ

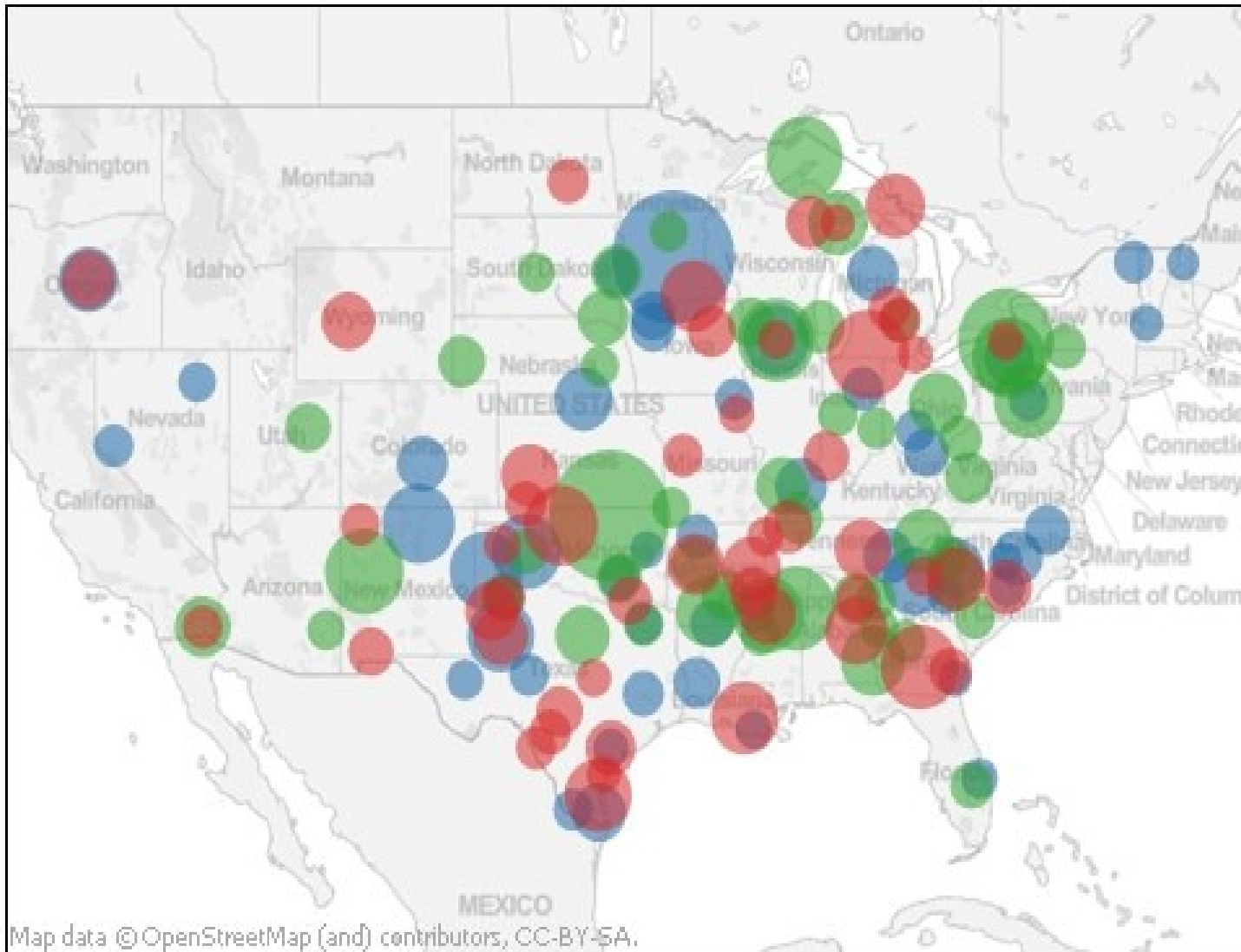
Solution(3/5)

- Solve user input problems by cross examining location with GeoNames.org API
- We cross examine the location with GeoNames.org API to get the country of origin after input sanitation
- Build JSON response

Solution(4/5)

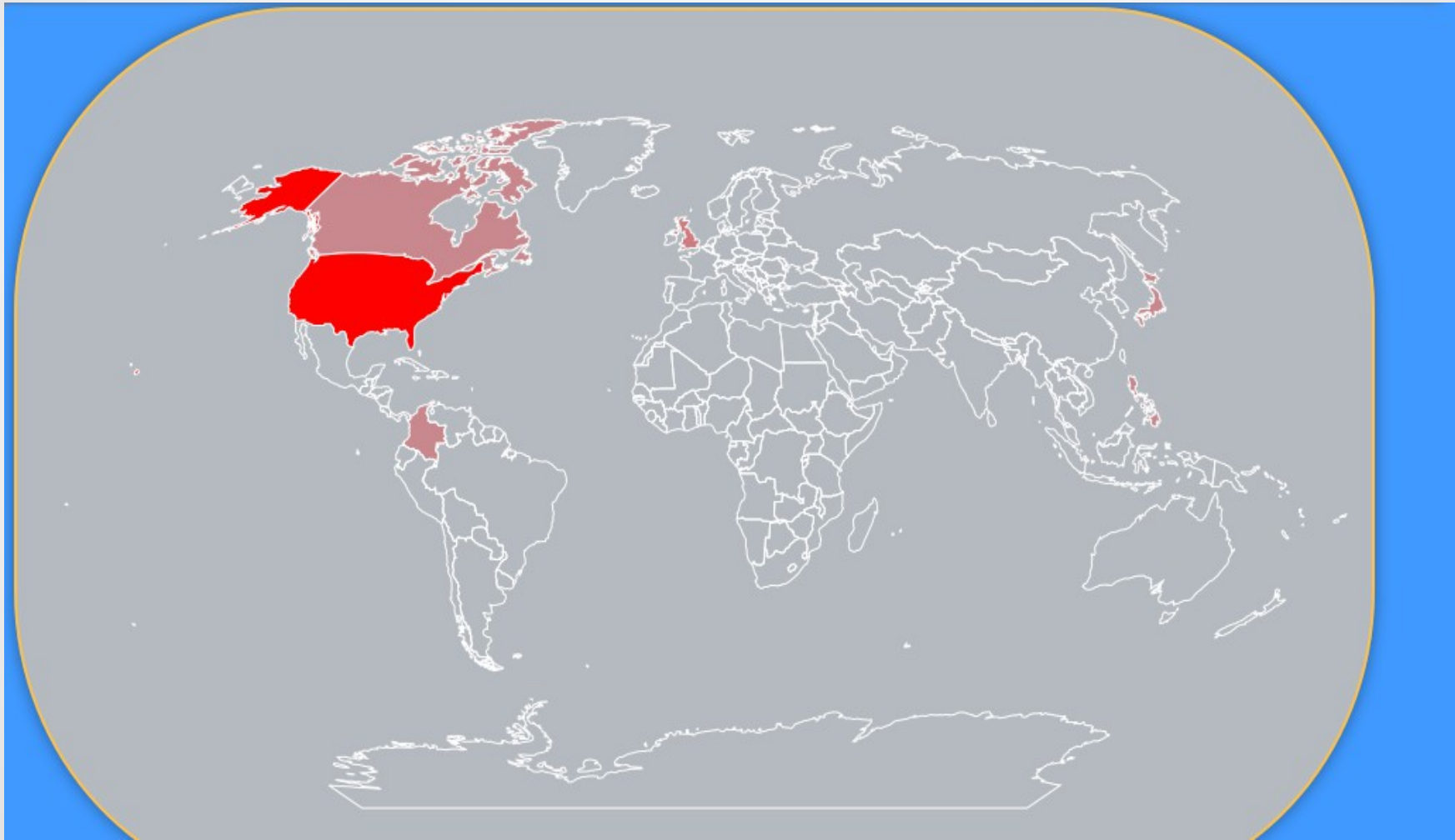
- What type of map is the best to use while visualizing the activity of a hashtag?
- Easy solution: visualize every tweet as single point
- Problematic because as the number of tweets is increasing so does the points
- Better solution: using choropleth map

Visualizing with points

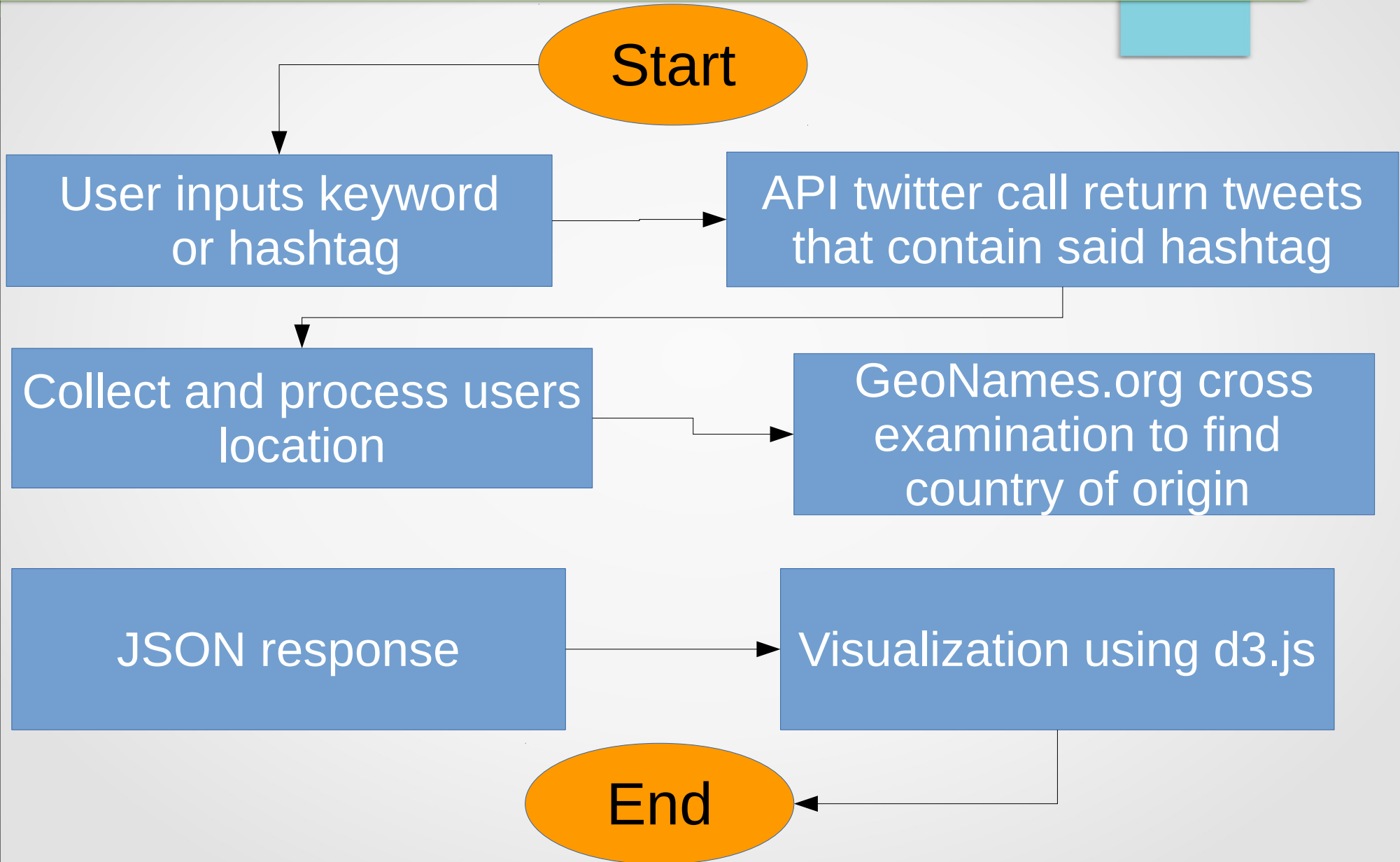


<https://nucloud.com/blog/muscle-car-hot-spots-interactive-data-map/>

Solution choropleth(5/5)



Visualization procedure overview



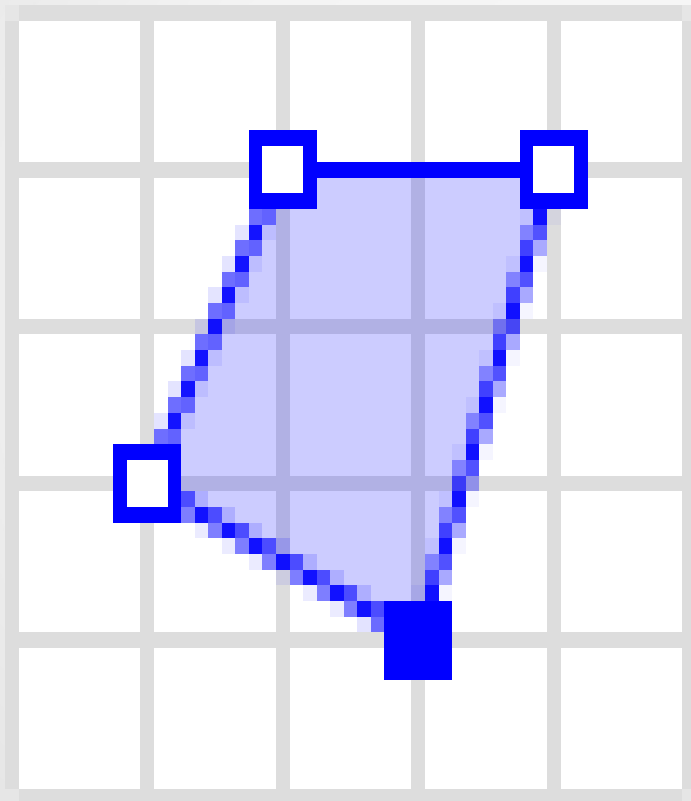
D3.js

- Javascript library used in order to visualize data in the best way possible
- Use of HTML, CSS and SVG
- Data driven transformation
- <https://github.com/mbostock/d3>

GeoJSON

- Open standard format used to represent geographical properties
- Uses points, lines , polygons and combination of the above
- Example: { "type": "Polygon",
 "coordinates": [
 [[30, 10], [40, 40], [20, 40], [10, 20], [30, 10]]
]
}

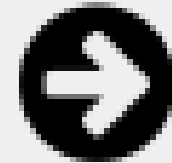
GeoJSON



<https://en.wikipedia.org/wiki/GeoJSON>

Case study: #capitalcontrols (1/3)

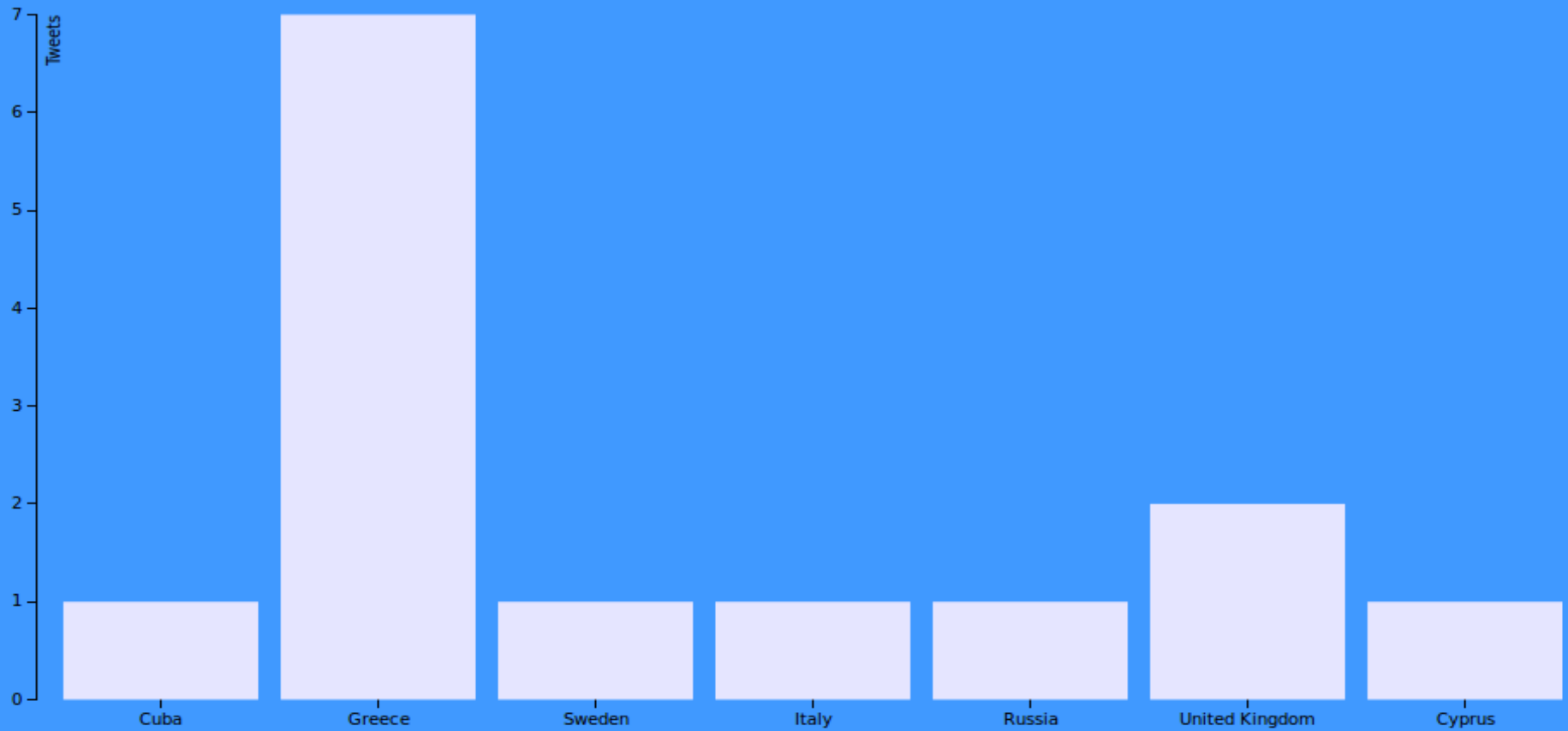
Trending in Greece 7/7:
#capitalcontrols



Case study: #capitalcontrols (2/3)



Case study: #capitalcontrols (3/3)



Recommendation system (1/2)

Latest Search: capitalcontrols

Recommend for you: 30/39
#Kourema



Recommendation system (2/2)

- Use of user's latest search
- Get tweets that contain the latest search
- Find hashtags that users post together with latest search
- Display hashtags that users post together with latest search

Conclusion

- Our application makes the monitoring of a hashtag's popularity easy by comparing his real time activity with older activity
- We get more information by using the recommendation system about the popularity outburst
- Choosing method of visualization is important in order to assimilate the displayed informations

Thank you

Questions?