



# Twitter: Εισαγωγή και Γεωγραφική Απεικόνιση Δεδομένων σε Χάρτη



Διπλωματική Εργασία της Τουμπόγλου Αικατερίνης

Επιβλέπων Καθηγητής: Χατζηγεωργίου Αλέξανδρος

# Σκοπός

Στόχο της διπλωματικής εργασίας αποτελεί η οπτική αναπαράσταση του τρόπου εξάπλωσης των αναρτήσεων στο Twitter επί γεωγραφικής απεικόνισης των δεδομένων που θα συλλεχθούν στον χάρτη.

Η εργασία συμπεριλαμβάνει την ανάπτυξη λογισμικού για την υλοποίηση των τεχνικών και την μελέτη του τρόπου εξάπλωσης της πληροφορίας δια μέσου γράφου που δημιουργείται.

# Ερωτήματα - Υποθέσεις

Υπάρχουν  
ισχυροί  
δεσμοί  
μεταξύ των  
χωρών στο  
Twitter;

Πόσο  
σημαντικό  
ρόλο παίζει η  
γεωγραφία  
στο Twitter;

Πόσο επηρεάζει η  
γλώσσα τη  
γεωγραφική  
μετάδοση των  
δεδομένων στο  
Twitter;



Πώς  
διαφοροποιείται  
για συγκεκριμένα  
θέματα η  
εξάπλωση της  
πληροφορίας  
γεωγραφικά;

Ποιες χώρες  
συμμετέχουν  
περισσότερο  
(συγκριτικά με τον  
πληθυσμό τους στο  
διαδίκτυο) στο  
κοινωνικό δίκτυο  
του Twitter;

# Βιβλιογραφική Επισκόπηση

Διαπιστώνεται πως η γεωγραφία διαδραματίζει σημαντικό ρόλο στη διαμόρφωση της κοινωνίας. Το αποτέλεσμα έρευνας έδειξε πως η συντριπτική πλειονότητα των κοντινότερων χωρών είναι γεωγραφικοί ή γλωσσικοί γείτονες (Kulshrestha, Kooti, Nikraves, & Gummadi, 2012).

Έχει διαπιστωθεί πως τα σύνορα μιας χώρας οριοθετούν σε πολύ μεγάλο βαθμό τις γλώσσες που χρησιμοποιούνται στο Twitter στην χώρα αυτή, με τις περισσότερες να έχουν ως κυρίαρχη γλώσσα τα αγγλικά (Kalev H., Shaowen, Guofeng, Anand, & Eric, 2013).

# Θεωρητικό Υπόβαθρο

Το διαδίκτυο, έχει επιτρέψει νέες μορφές επικοινωνίας, συλλογικής δράσης και συνεργασίας όπως τα κοινωνικά δίκτυα.

Το Twitter, δημιουργήθηκε στις 21 Μαρτίου του 2006 και έγινε άμεσα δημοφιλές ως κοινωνικό δίκτυο.

Σε ένα κοινωνικό δίκτυο, η γεωγραφία παίζει ιδιαίτερα σημαντικό ρόλο.

Η θεωρία των γράφων αναλύει την δομή ενός κοινωνικού δικτύου.

# Γλώσσες Προγραμματισμού, API's, Βιβλιοθήκες

## Γλώσσες Προγραμματισμού:

- HTML,
- CSS,
- JavaScript,
- PHP

## API's (Application Programming Interface):

- Twitter REST API,
- Twitter Streaming API,
- Google Maps API

Οι βασικότερες  
Βιβλιοθήκες που  
χρησιμοποιήθηκαν:

TwitterOAuth

Phirehose

Vis

MarkerClusterer

Gini

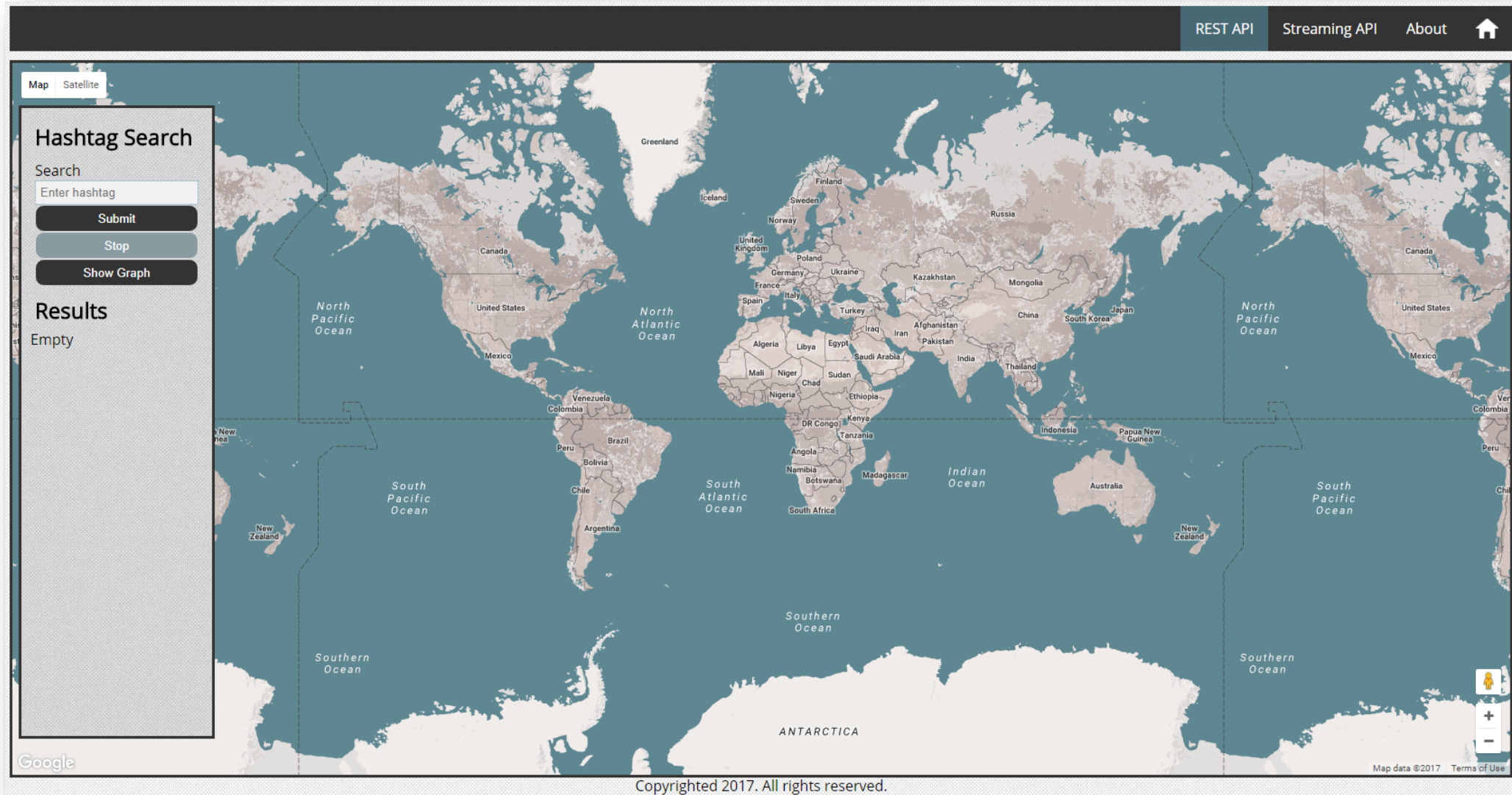
ειν!

# Επίδειξη Λειτουργίας Εφαρμογής

**Streaming API (Live Tweets)**

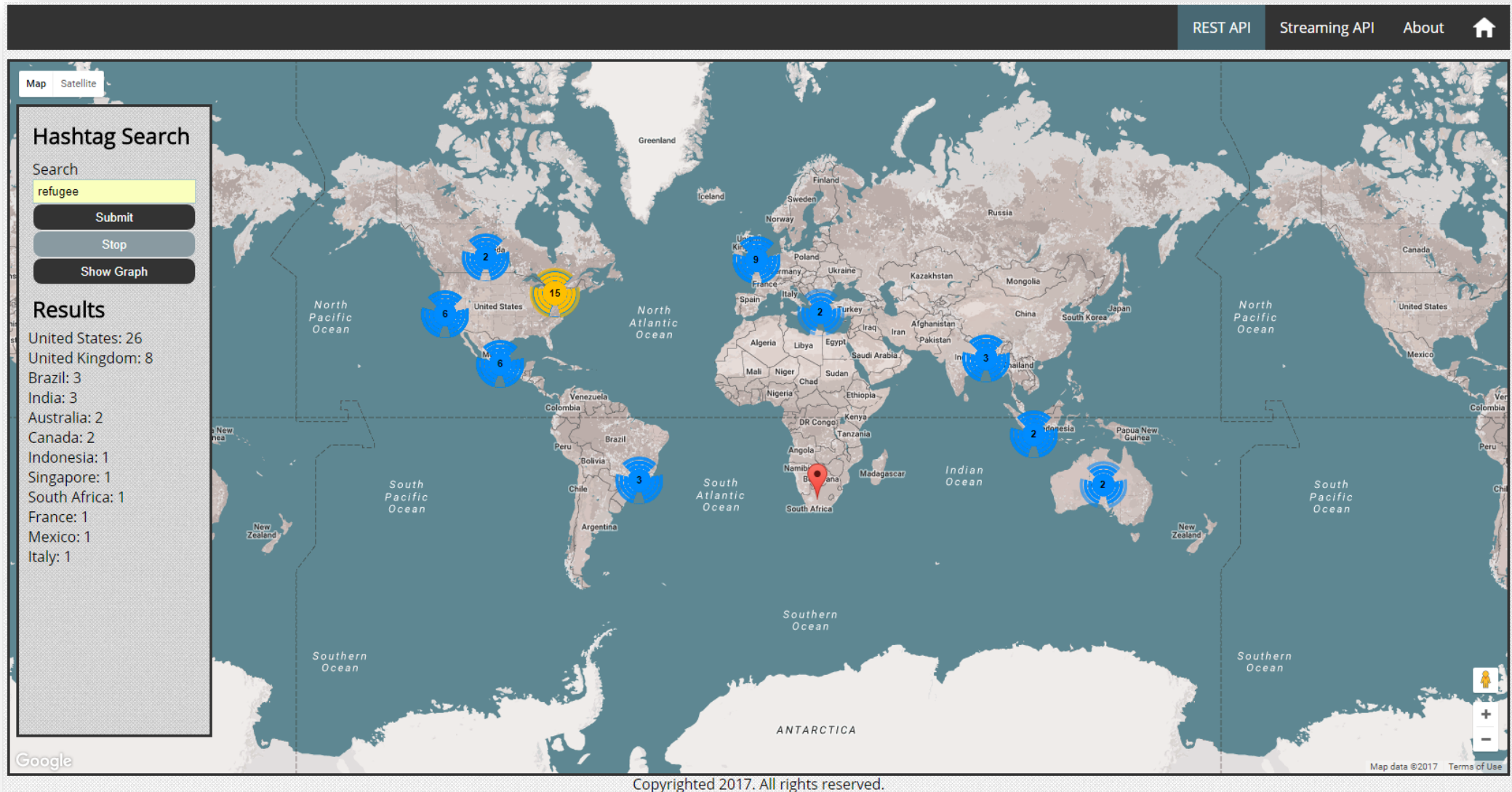
**REST API (Recent Tweets)**

# REST API





# Αναζήτηση



# Γράφος

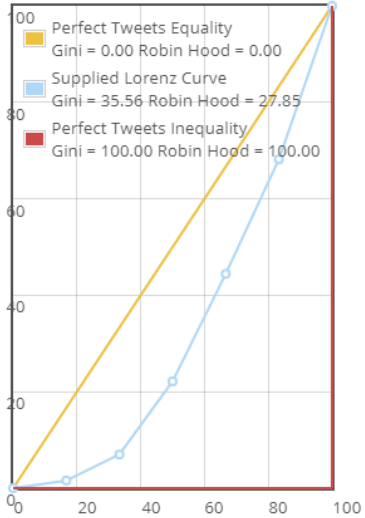
Tweets per  
population

0.091  
0.13  
0.022  
0.0065  
0.097  
0.062

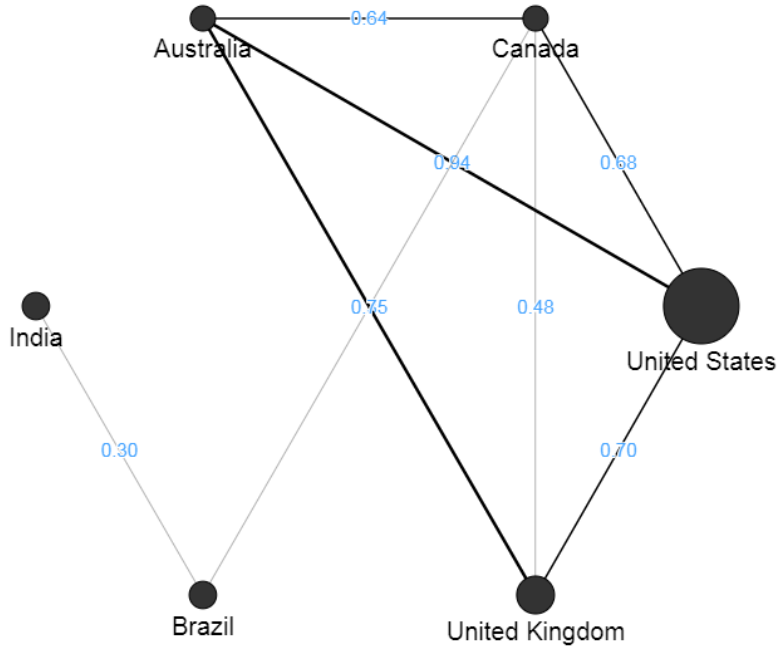
% of tweets per country population

Go

Gini Coefficient  
% of population



Searching for: refugee



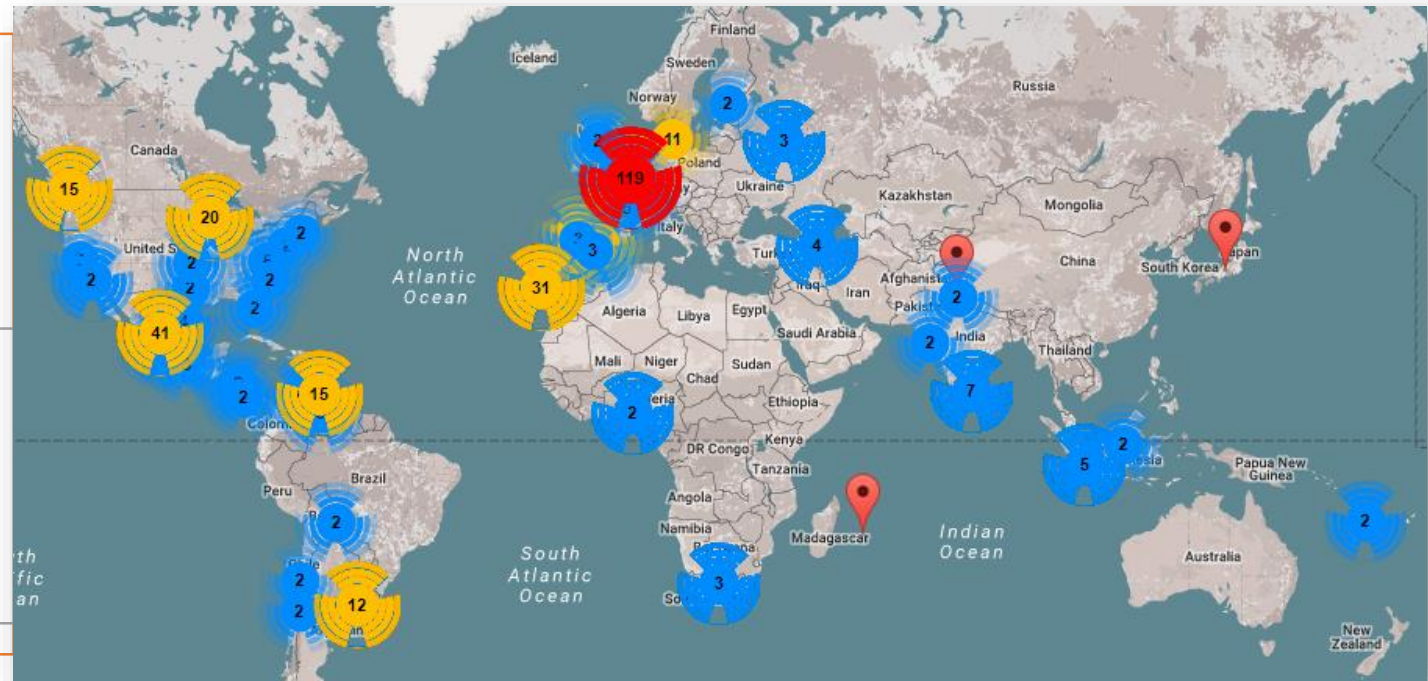
Graph Results

- Australia 0.64 Canada
- Australia 0.75 United Kingdom
- Australia 0.94 United States
- Brazil 0.30 India
- Brazil 0.35 Canada
- Canada 0.35 Brazil
- Canada 0.48 United Kingdom
- Canada 0.64 Australia
- Canada 0.68 United States
- India 0.30 Brazil
- United Kingdom 0.48 Canada
- United Kingdom 0.70 United States
- United Kingdom 0.75 Australia
- United States 0.68 Canada
- United States 0.70 United Kingdom
- United States 0.94 Australia

# Μελέτη Περίπτωσης 1 και 2: Streaming API, #Portugal

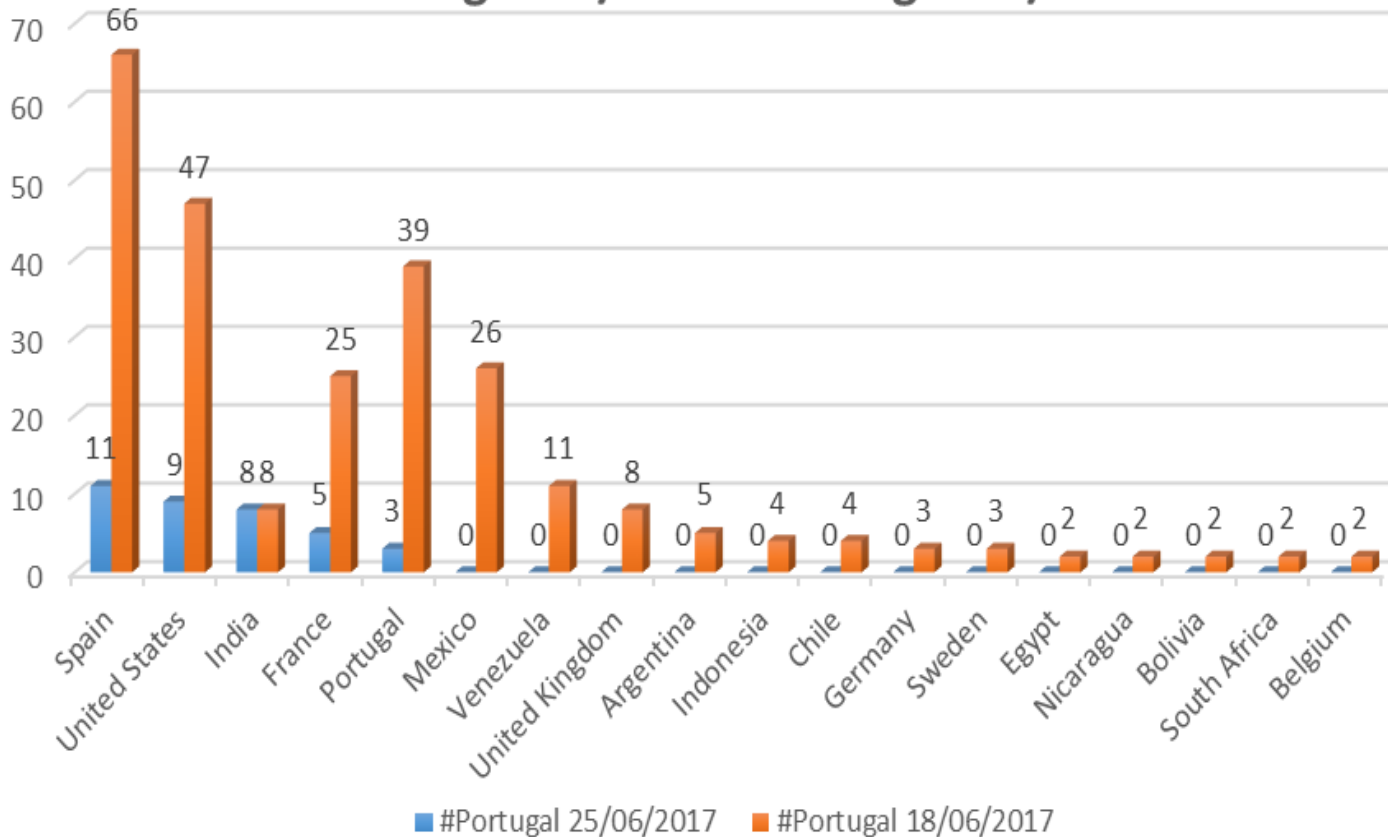
Στις 18/06: Η Ισπανία, η Πορτογαλία και η Γαλλία είναι οι χώρες με τα περισσότερα tweets.

Στις 25/06: Το σύνολο των tweets είναι μόλις 48 σε αντίθεση με την ημέρα που έγινε το συμβάν όπου ο αριθμός ήταν 282 tweets.



# Σύγκριση Μελέτης Περίπτωσης 1 και Μελέτης Περίπτωσης 2

#Portugal 18/06 & #Portugal 25/06



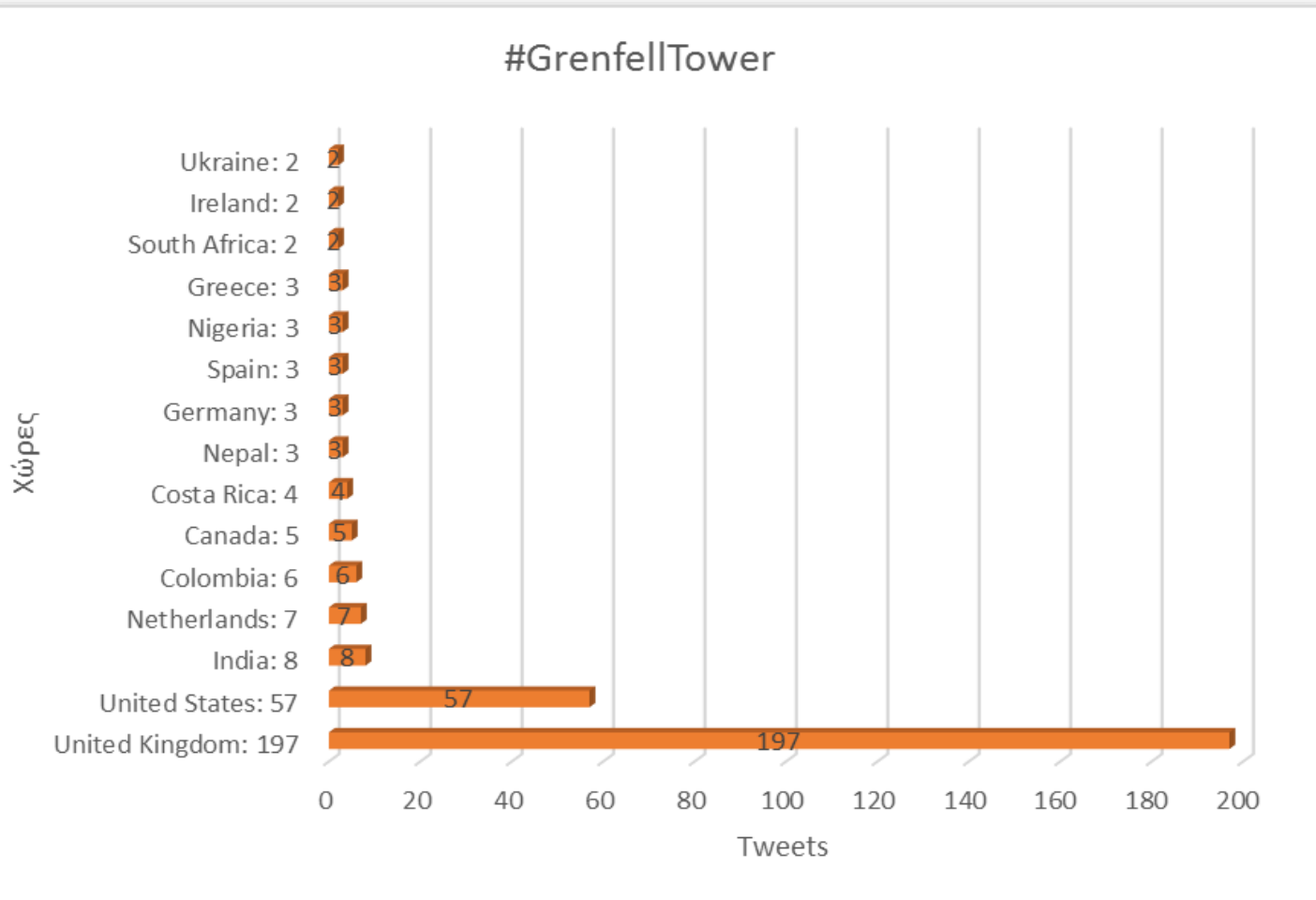
Ισπανία: 6 φορές  
περισσότερα  
tweets.

ΗΠΑ: 5.2 φορές  
περισσότερα  
tweets.

Γαλλία: 5 φορές  
περισσότερα  
tweets.

Πορτογαλία: 13  
φορές  
περισσότερα  
tweets.

# Μελέτη Περίπτωσης 3: REST API, #GrenfellTower, 18/06/2017

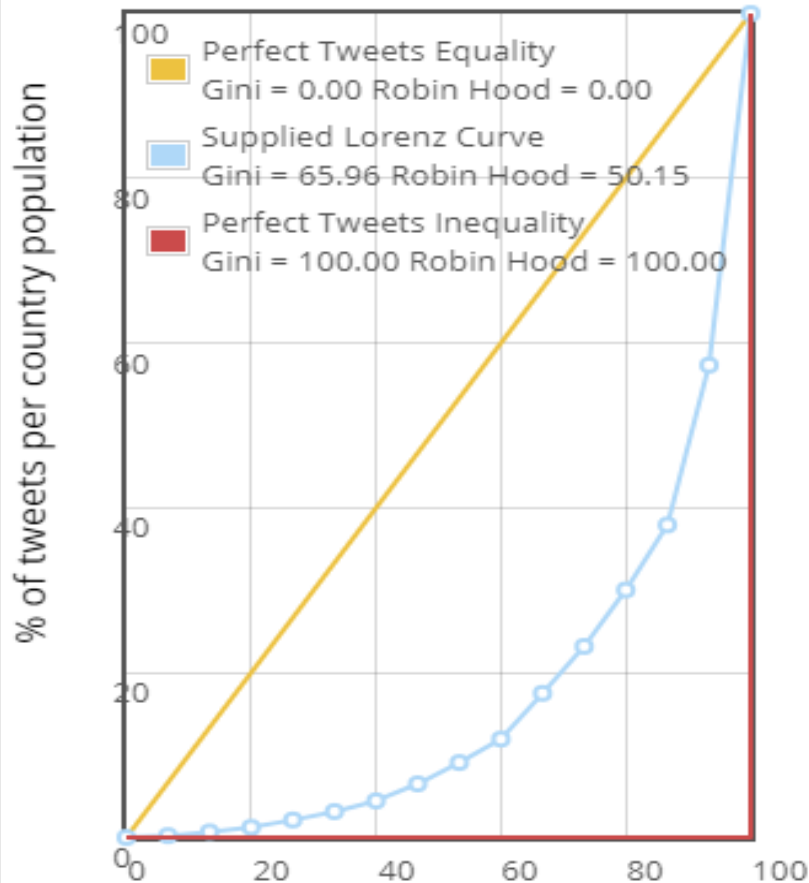


Το θέμα φαίνεται να συζητείται σχεδόν τετραπλάσιες φορές στην περιοχή που έγινε το συμβάν (Ηνωμένο Βασίλειο) από την δεύτερη χώρα (ΗΠΑ) η οποία εμφανίζεται στα αποτελέσματα.



# Gini Coefficient

Gini Coefficient  
% of population



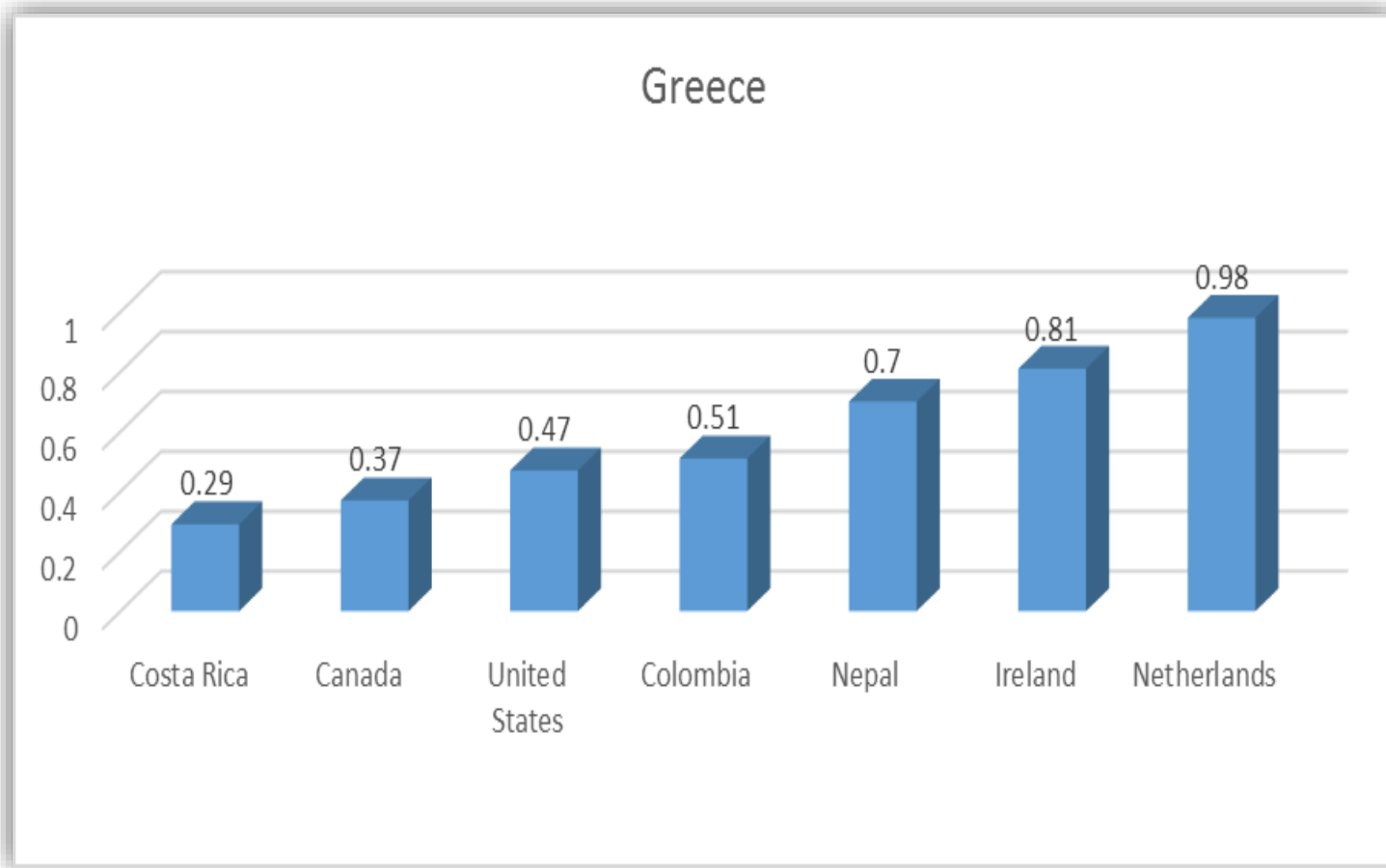
Ο συντελεστής Gini Coefficient εκφράζει το βαθμό κατά τον οποίο μια κατανομή είναι ομοιόμορφη. Εάν ο συντελεστής Gini είναι μηδέν τότε αυτό υποδεικνύει την τέλεια ισότητα ενώ εάν είναι ένα εκφράζει την τέλεια ανισότητα

Θα μπορούσαμε να ισχυριστούμε πως οι Ηνωμένες Πολιτείες Αμερικής, εκτός από το ότι έχουν τον υψηλότερο πληθυσμό συγκριτικά με τις υπόλοιπες χώρες που εμφανίζονται, κατέχουν την δεύτερη θέση και λόγω της γλώσσας.

# Συσχετίσεις χωρών γράφου #GrenfellTower

Greece	0.98	Netherlands		Costa Rica	0.45	United Kingdom
Colombia	0.91	United States		Netherlands	0.45	United States
South Africa	0.89	Spain		Canada	0.44	South Africa
Ireland	0.87	Nepal		Nigeria	0.44	Spain
Germany	0.83	Nigeria		Colombia	0.42	Ireland
Ireland	0.83	Netherlands		Germany	0.42	Ukraine
Greece	0.81	Ireland		Costa Rica	0.41	Nepal
Canada	0.8	United States		Germany	0.4	India
Spain	0.79	Ukraine		Spain	0.39	United States
Canada	0.73	Colombia		Ireland	0.38	United States
Nepal	0.72	Netherlands		Canada	0.37	Greece
Greece	0.7	Nepal		Canada	0.36	Netherlands
South Africa	0.7	Ukraine		Colombia	0.36	Nepal
Canada	0.63	Ukraine		Colombia	0.36	Spain
Germany	0.6	South Africa		Costa Rica	0.35	Ireland
Germany	0.53	Spain		Nigeria	0.35	Ukraine
Colombia	0.51	Greece		South Africa	0.35	United States
Colombia	0.5	Netherlands		Nepal	0.33	United States
Nigeria	0.5	South Africa		Colombia	0.32	South Africa
Ukraine	0.5	United States		Canada	0.3	Ireland
Canada	0.49	Spain		Costa Rica	0.29	Greece
India	0.49	Nigeria		Costa Rica	0.29	Netherlands
Greece	0.47	United States		Canada	0.26	Germany
Colombia	0.45	Ukraine		Canada	0.26	Nepal

# Συσχετισμός χωρών με Ελλάδα για #GrenfellTower



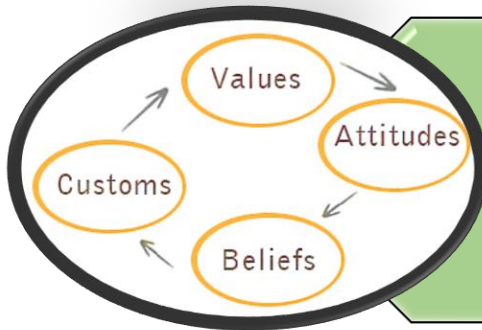
Όπως φαίνεται, υπάρχει αρκετά μεγάλη ταύτιση με την Ολλανδία και έπειτα με την Ιρλανδία.



# Συμπεράσματα



Το κύριο συμπέρασμα που προκύπτει από την έρευνα που έγινε είναι πως η γεωγραφία παίζει σημαντικό ρόλο στις αναρτήσεις των χρηστών στο κοινωνικό δίκτυο του Twitter.



Θα μπορούσαμε να ισχυριστούμε πως άνθρωποι με κοντινότερη κουλτούρα επηρεάζονται για τα ίδια ζητήματα και σχολιάζουν εξίσου στα κοινωνικά δίκτυα.



Η γλώσσα παίζει μεγάλο ρόλο, αφού οι ΗΠΑ συμμετέχουν σχεδόν πάντα με υψηλά ποσοστά ενώ αντίθετα για παράδειγμα η Κίνα και η Τουρκία σπάνια εμφανίζονται στον χάρτη και αν εμφανιστούν κατέχουν χαμηλά ποσοστά συμμετοχής.

# Όρια και περιορισμοί της έρευνας

Google Maps JavaScript API: To Reverse Geocoding (δηλαδή η μετατροπής μίας διεύθυνσης σε γεωγραφικό μήκος και γεωγραφικό πλάτος) είναι υπεύθυνο για την εισαγωγή μιας διεύθυνσης σε ένα σημείο επάνω στον χάρτη. Οι περιορισμοί που τίθενται με το βασικό (δωρεάν) πακέτο του API είναι έως 2500 αιτήματα την ημέρα και έως 50 αιτήματα ανά δευτερόλεπτο.

Μία προσωρινή λύση που δόθηκε, ήταν η προσπάθεια επανατοποθέτησης του tweet που ξεπερνούσε το όριο του Google Maps στον χάρτη έπειτα από μικρό χρονικό διάστημα, για να μη χαθεί κάποιο tweet.

# Μελλοντικές Επεκτάσεις

Όσον αφορά το Streaming API, τα tweets αυτά αντλούνται αλλά δεν χρησιμοποιούνται όλα λόγω των περιορισμών του Google Maps. Θα μπορούσε να γίνει παράλληλη έρευνα στο σύνολο των δεδομένων χωρίς να απεικονιστούν στον χάρτη.

Μελλοντικά, μέσω ενός μεγαλύτερου και πιο ολοκληρωμένου συνόλου στοιχείων, θα μπορούσε να γίνει έρευνα σχετικά με το πόσο επηρεάζει η γλώσσα ένα θέμα στο κοινωνικό δίκτυο Twitter και ποιες χώρες -ίσως- χρησιμοποιούν μόνο τη δική τους γλώσσα για να αναφερθούν σε ένα ζήτημα.

Ευχαριστώ!