

ΠΡΟΧΩΡΗΜΕΝΑ ΘΕΜΑΤΑ
ΤΕΧΝΟΛΟΓΙΑΣ ΚΑΙ ΕΦΑΡΜΟΓΩΝ
ΒΑΣΕΩΝ ΔΕΔΟΜΕΝΩΝ

ΠΡΟΓΡΑΜΜΑΤΙΣΤΙΚΗ ΕΡΓΑΣΙΑ
ΓΙΑ ΤΟ ΑΚΑΔΗΜΑΪΚΟ ΕΤΟΣ
2016-2017

ΟΜΑΔΑ 2

ΤΖΙΒΑΡΑΣ ΠΑΝΑΓΙΩΤΗΣ, 1931

ΤΖΙΒΑΡΑΣ ΒΑΣΙΛΗΣ, 1770

ΤΕΛΙΚΗ ΑΝΑΦΟΡΑ

ΜΑΪΟΣ **2017**

ΙΣΤΟΡΙΚΟ ΠΡΟΗΓΟΥΜΕΝΩΝ ΕΚΔΟΣΕΩΝ

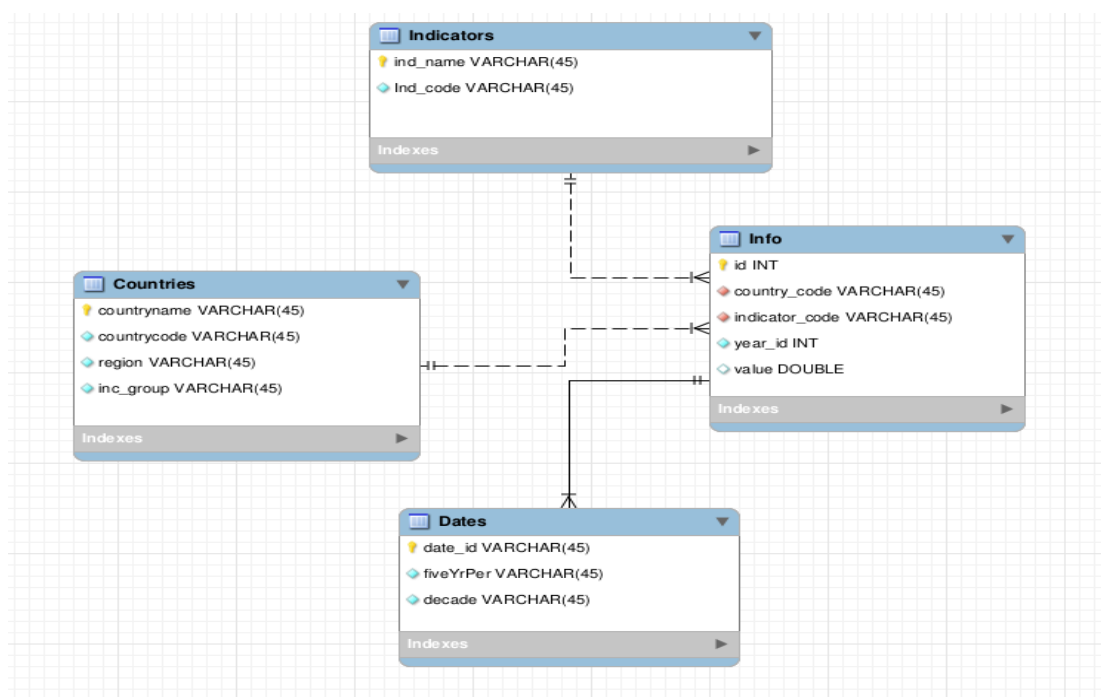
Ημερομηνία	Έκδοση	Περιγραφή	Συγγραφέας
2017/02/10	0.1	Σχεδίαση της βάσης, export schema	Τζιβάρας Παναγιώτης
2017/02/20	0.2	Data transformation scripts	Τζιβάρας Παναγιώτης
2017/03/05	0.3	Load data into the database	Τζιβάρας Παναγιώτης
2017/03/10	0.4	Server setup and first tests	Τζιβάρας Παναγιώτης
2017/03/20	0.5	Website design and development (HTML+CSS)	Τζιβάρας Βασίλης
2017/03/25	0.6	Εισαγωγή κώδικα PHP and MySQL code για αλληλεπίδραση με την βάση.	Τζιβάρας Βασίλης
Up today	0.7	Ολοκλήρωση του κώδικα και testing to software	Τζιβάρας Βασίλης

1 ΒΑΣΗ ΔΕΔΟΜΕΝΩΝ

Εδώ περιγράφονται τα σχήματα της βάσης (ή βάσεων, αν είναι παραπάνω από μία) δεδομένων που χρησιμοποιούνται στο project.

1.1 ΣΧΕΣΙΑΚΟ ΣΧΗΜΑ ΣΕ ΛΟΓΙΚΟ ΕΠΙΠΕΔΟ

Το σχήμα της βάσης μας αποτελείται από 4 πίνακες όπως φαίνεται και στην παρακάτω εικόνα. Ο πρώτος πίνακας είναι ο Indicators που εκεί αποθηκεύουμε όλους τους δείκτες από το αρχείο εισόδου, κρατώντας πληροφορία για τον κωδικό και το όνομα του κάθε δείκτη. Ο δεύτερος πίνακας είναι ο Info όπου εκεί κρατάμε πληροφορία για την τιμή της κάθε χώρας, κάθε χρονιά. Ο πίνακας αυτός είναι ο πιο μεγάλος στην βάση μας. Ο επόμενος είναι ο Dates όπου εκεί κρατάμε πληροφορία για τις ημερομηνίες για να μπορέσουμε αργότερα να groupάρουμε σε 5ετίες και 10ετίες. Τέλος, έχουμε τον πίνακα Countries όπου όπως είναι λογικό από



το όνομά του κρατά τις χώρες, δηλαδή τα ονόματα, τους κωδικούς καθώς και το region και το εισόδημα της κάθε μιας.

Όπως αναφέρουμε και στις πληροφορίες στο τέλος του doc, ίσως θα έπρεπε να έχουμε διαφορετικό σχεδιασμό στην βάση μας. Αυτό όμως φάνηκε στο τέλος του Project και δεν είχαμε χρόνο να αλλάξουμε το σχήμα και οτι άλλο αυτό συνεπάγεται.

1.2 ΣΧΕΣΙΑΚΟ ΣΧΗΜΑ ΣΕ ΦΥΣΙΚΟ ΕΠΙΠΕΔΟ





1.2.1 ΡΥΘΜΙΣΗ ΤΩΝ ΠΑΡΑΜΕΤΡΩΝ ΤΟΥ DBMS

Στην βάση χρησιμοποιήθηκε InnoDB ως μηχανή αποθήκευσης. Το memory pool του InnoDB είναι ρυθμισμένο στα 8388608 bytes. Τέλος τα max bytes του κάθε query είναι 16384. Επιπλέον πληροφορίες για το σύστημα μπορούν να βρεθούν παρακάτω:

innodb_additional_mem_pool_size	8388608	Size of a memory pool InnoDB uses to store data dictionary information and other internal data structures
innodb_buffer_pool_size	16777216	Size of the memory buffer InnoDB uses to cache data and indexes of its tables
innodb_change_buffer_max_size	25	[rw] Maximum size for the InnoDB change buffer, as a percentage of the total size of buffer pool
innodb_ft_cache_size	8000000	Size of the cache that holds a parsed document in memory while creating an InnoDB FULLTEXT index
innodb_ft_max_token_size	84	Maximum length of words that are stored in an InnoDB FULLTEXT index
innodb_ft_min_token_size	3	Minimum length of words that are stored in an InnoDB FULLTEXT index
innodb_ft_total_cache_size	640000000	The total memory allocated for the InnoDB FULLTEXT search index cache.
innodb_log_block_size	512	
innodb_log_buffer_size	8388608	Size of buffer which InnoDB uses to write log to the log files on disk
innodb_log_file_size	5242880	Size of each log file in a log group
innodb_max_bitmap_file_size	104857600	
innodb_online_alter_log_max_size	134217728	[rw] Specifies an upper limit on size of the temporary log files used during online DDL operations for InnoDB tables.
innodb_page_size	16384	Specifies the page size for all InnoDB tablespaces in an instance
innodb_purge_batch_size	300	[rw] Specifies the number of InnoDB redo logs that trigger a purge operation.
innodb_sort_buffer_size	1048576	Specifies size of a buffer used for sorting data during creation of an InnoDB index
innodb_sync_array_size	1	Splits an internal data structure used to coordinate threads, for higher concurrency in workloads with large numbers of threads
join_buffer_size	262144	[rw] Size of buffer that is used for full joins

1.2.2 ΡΥΘΜΙΣΗ ΤΟΥ ΦΥΣΙΚΟΥ ΣΧΗΜΑΤΟΣ ΤΗΣ ΒΑΣΗΣ ΔΕΔΟΜΕΝΩΝ

Στην βάση δεδομένων μας έχουμε 4 Indexes και καθόλου views. Όπως φαίνεται και στην εικόνα είναι αποθηκευμένα σε B δέντρα και τα Indexes είναι Primary κλειδιά σε κάθε έναν από τους βασικούς πίνακες που αναφέραμε και πιο πανω.

Table	Name	Unique	Index Type	Index Comment	Column	Seq in Index	Packed	Collation	Cardinality
 Countries	PRIMARY	Yes	BTREE		country_code	1		A	58
 Dates	PRIMARY	Yes	BTREE		date_id	1		A	54
 Indicators	PRIMARY	Yes	BTREE		id	1		A	18378
 Info	PRIMARY	Yes	BTREE		id	1		A	268945

1.2.3 ΡΥΘΜΙΣΗ ΑΣΦΑΛΕΙΑΣ

Όσον αφορά την ασφάλεια του συστήματος στο λογισμικό δεν έχουμε προσθέσει κάποιο Login page. Το θεωρήσαμε μη απαραίτητο στα πλαίσια του μαθήματος. Σίγουρα κάτι που θα ήταν πολύ χρήσιμο είναι να χρησιμοποιηθούν sessions και φυσικά ένα αρχικό page για το authentication, ορισμός δικαιωμάτων καταχώρησης ή ανάκτησης δεδομένων σε διαφορετικούς ρόλους και χρήστες του συστήματος.

2 ΑΡΧΙΤΕΚΤΟΝΙΚΗ ΛΟΓΙΣΜΙΚΟΎ

2.1 ΑΡΧΙΤΕΚΤΟΝΙΚΗ ΚΑΙ ΔΟΜΗ ETL

Τα αρχικά δεδομένα που πήραμε ήταν ένα ενιαίο αρχείο. Έτσι λοιπόν βάση του σχήματος της βάσης φτιάξαμε 4 parser. Οι parser είναι οι `countries_script.py`, `info_script.py`, `dates_script.py`, `indicators_script.py`. Είναι γραμμένοι σε python και ο κάθε ένας κάνει parse το ένα από τα δυο η και τα δυο αρχεία εισόδου και δημιουργεί out αρχεία που στη συνέχεια τα φορτώνουμε στην βάση μας με το LOAD script. Για παράδειγμα ο `countries_script.py` παίρνει ως είσοδο τα αρχεία `input` και βγάζει έξοδο αρχείο με τα `countries names`, `codes`, `inc_group` και `region`.

Το μόνο αυτοματοποιημένο κομμάτι είναι η χρήση της βιβλιοθήκης `csv` στην python. Οι ρυθμίσεις αν μπορούν να θεωρηθούν ως μετρήσεις είναι

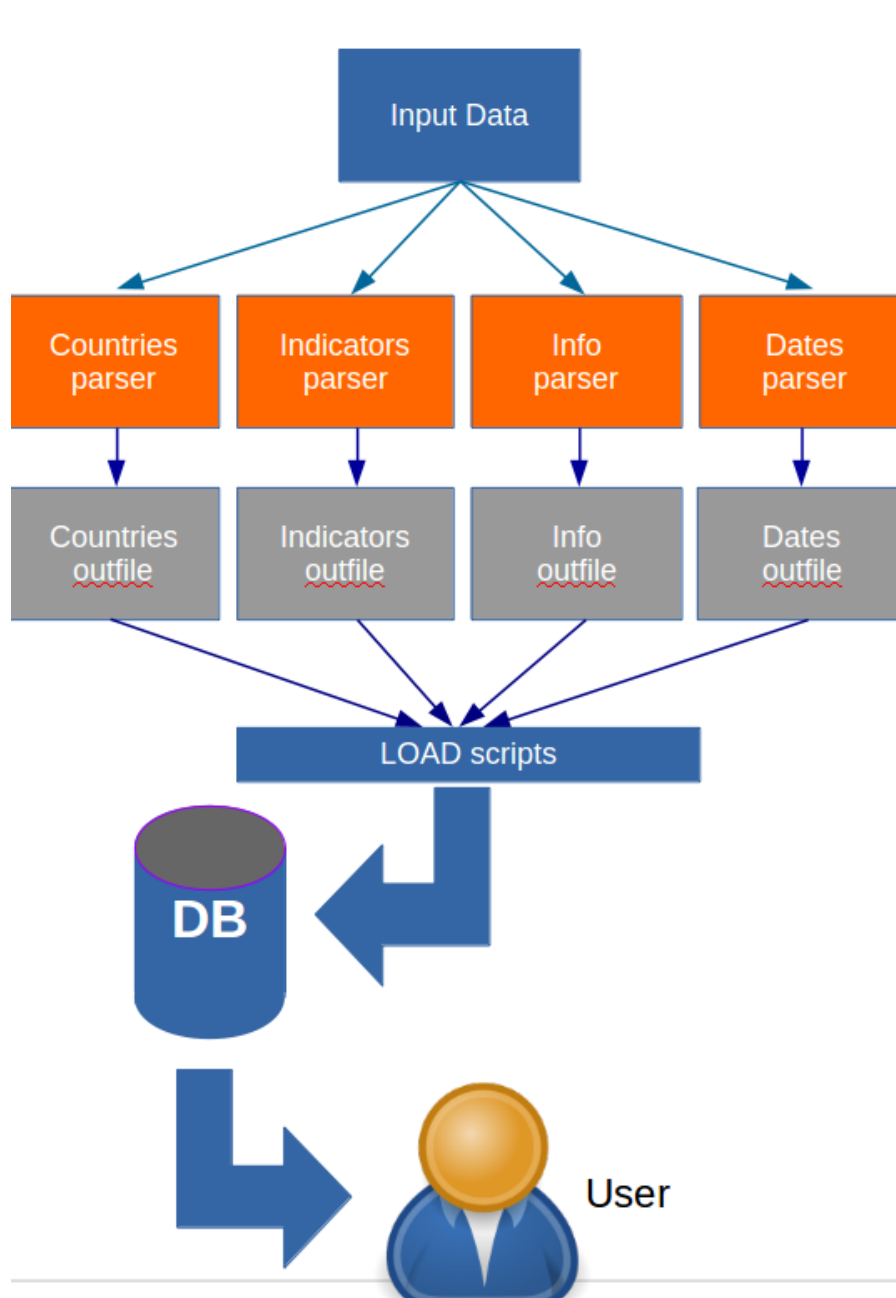
```
spamreader = csv.reader(csvfile, delimiter=',', quotechar='"')
```

όπου ορίζουμε ως

- **csvfile** το αρχείο εισόδου (full path)
- **delimiter** το πως χωρίζονται τα πεδία μας
- **quotechar** οι χαρακτήρες μέσα στους οποίους καθορίζεται κάθε ένα πεδίο

η λογική πίσω από την οποία δουλεύουν αυτά τα script είναι πολύ απλή. Στην πραγματικότητα, απλά διαβάζουμε γραμμή γραμμή και κρατάμε σε λίστες όλα τα δεδομένα μας. Έπειτα κάνουμε αναζητήσεις και επιλέγουμε τα records και τα fields τα οποία θέλουμε. Επιπλέον, έχουμε προσθέσει στο script για τις χώρες έλεγχο για να πάρει μόνο τις Ευρωπαϊκές.

2.2 ΔΙΑΓΡΑΜΜΑΤΑ ΠΑΚΕΤΩΝ / ΥΠΟΣΥΣΤΗΜΑ ΤΩΝ ΚΕΝΤΡΙΚΗΣ ΕΦΑΡΜΟΓΗΣ



2.3 ΔΙΑΓΡΑΜΜΑ ΚΛΑΣΕΩΝ ΚΕΝΤΡΙΚΗΣ ΕΦΑΡΜΟΓΗΣ

Στην εφαρμογή μας δεν υπάρχουν κλάσεις. Η υλοποίηση δηλαδή δεν έγινε με αντικειμενοστρέφεια. Χρησιμοποιήθηκε Pure HTML, CSS and PHP σε συνδιασμό με MySQL για την βάση δεδομένων μας. Δεν χρησιμοποιήσαμε κανένα γνωστό framework. Σίγουρα θα μπορούσαμε να σπάσουμε τον κώδικα σε περισσότερα αρχεία για να τα αναλύσουμε εδώ ένα ένα αλλά προσπαθώντας να ολοκληρώσουμε το Project είχαμε πολλά σφάλματα. Στην αρχή έχουμε έναν κώδικα όπου συνδέεται στην βάση δεδομένων και

αν ο χρήστης έχει επιλέξει έναν δείκτη τραβάει απο τον πίνακα Indicators όλους τους δείκτες και τους εμφανίζει στην σχετική λίστα όπως φαίνεται και στο section 3. Έπειτα ο κώδικας HTML εμφανίζει τα στοιχεία αλληλεπίδρασης με τον χρήστη και στο τέλος υπάρχει ένα κομμάτι κώδικα όπου με ρηρ ελέγχουμε αν έχει γίνει κάποιο ερώτημα στην βάση και έχουν έρθει δεδομένα. Αν ήρθαν εμφανίζουμε με D3 την κατάλληλη γραφική παράσταση!

3 ΥΠΟΔΕΙΓΜΑΤΑ ΕΡΩΤΗΣΕΩΝ ΚΑΙ ΑΠΑΝΤΗΣΕΩΝ

Το λογισμικό μας αποτελείται κυρίως απο μια εννιαία διεπαφή (ιστοσελίδα). Εκεί έχουμε χωρίσει τα βήματα του χρήστη σε sections και έτσι ο χρήστης μπορεί να αλληλεπιδρά με το σύστημα. Πιο συγκεκριμένα Όταν ο χρήστης μπαίνει στο λογισμικό αρχικά πρέπει να επιλέξει έναν δείκτη. Αφού το επιλέξει έπειτα κάνει κλικ στο “Find Countries” και αν όλα πήγαν καλά εμφανίζεται ένα μήνυμα δίπλα απο το button όπως φαίνεται και παρακάτω.

STEP 1 - Select an indicator

Net migration
Vulnerable employment, total (% of total employment)
Poverty headcount ratio at \$1.90 a day (2011 PPP) (% of population)
Income share held by lowest 20%
Incidence of tuberculosis (per 100,000 people)
Prevalence of stunting, height for age (% of children under 5)
Maternal mortality ratio (modeled estimate, per 100,000 live births)
Pregnant women receiving prenatal care (%)
Improved sanitation facilities (% of population with access)
Mortality rate, under-5 (per 1,000 live births)

Find countries

Countries updated!

Ταυτόχρονα μέσα απο την διαδικασία αυτή έχουν φορτωθεί όλες οι χώρες που σχετίζονται με τον συγκεκριμένο δείκτη. Αυτό έχει ως αποτέλεσμα να εμφανιστούν οι παρακάτω λίστες.

STEP 2 - Select a country and the year mode

Azerbaijan
Kazakhstan
Macedonia, FYR
Romania
Uzbekistan

Single year mode
Five year mode
Ten year mode

* keep Ctrl key for multiple selection

STEP 3 - Pick your chart type

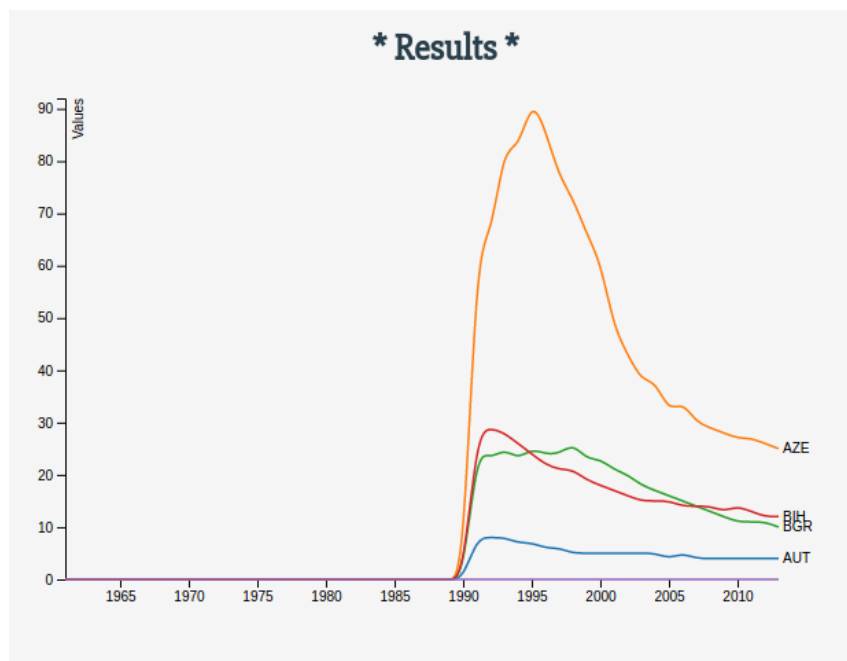
Line chart
Bar Chart
Scatter Plot

Create chart

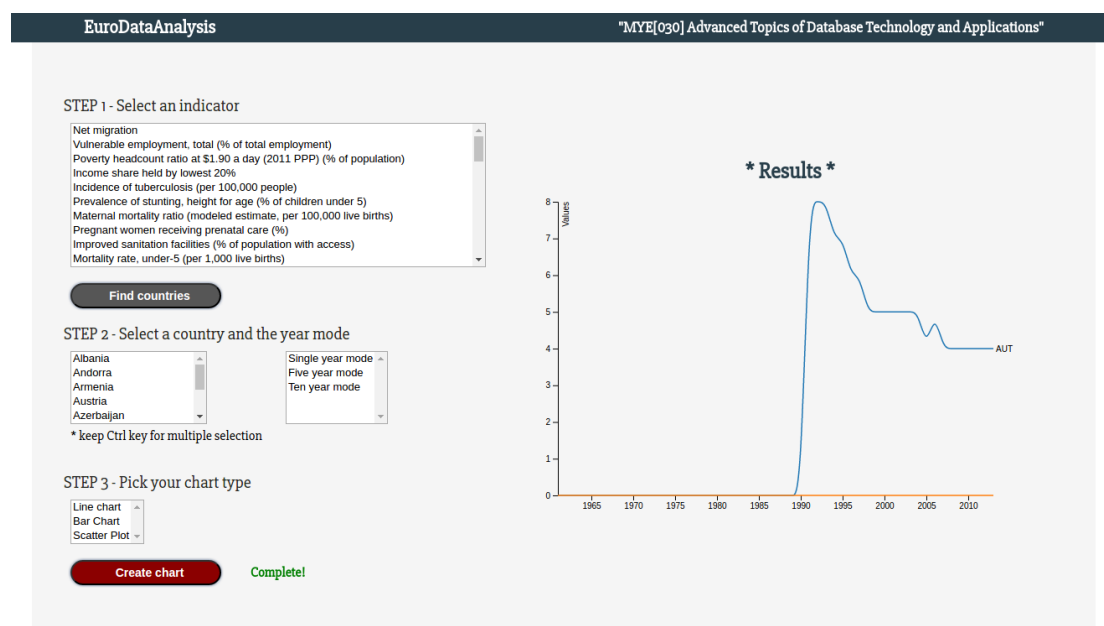
Complete!

Εδώ ο χρήστης επιλέγει μια ή παραπάνω χώρες, το year mode αν θέλει δηλαδή να εκτυπωθούν σε 5ετίες ή 10ετίες τα δεδομένα του και τέλος τον τύπο του διαγράμματος που θέλει.

Επιλέγοντας όλα τα παραπάνω και κάνοντας κλικ στο “Create Chart” εμφανίζεται στα δεξιά το γράφημα.



Τέλος βλέπουμε το ολοκληρωμένο σύστημα στην παρακάτω εικόνα.



4 ΤΕΚΜΗΡΙΩΣΗ ΚΑΙ ΛΟΙΠΑ ΣΧΟΛΙΑ

Έχουμε αντιμετωπίσει κάποια προβλήματα που δυστυχώς με τον τρόπο που είναι σχεδιασμένη η βάση και το software η επίλυση τους κάνει τον κώδικα χαοτικό. Το version που παραδίδουμε δεν έχει το ολοκληρωμένο software και αποφασίσαμε να παραδώσουμε κάτι που να δουλεύει χωρίς σοβαρά προβλήματα (τουλάχιστον απο όσα testing κάναμε). Παρακάτω αναφέρουμε τα issues που έχουμε για να ολοκληρωθεί το software.

Το κυρίως πρόβλημα που αντιμετωπίσαμε και έφαγε πολυ χρόνο είναι στη δεύτερη φάση των data transformation όταν δηλαδή φέρνουμε απο την βάση το αποτέλεσμα του query και πρέπει το result set να το φτιάξουμε στην μορφή που πρέπει να γραφτούν σε ένα αρχείο για να τα φορτώσουμε στη D3. (η D3 θέλει συγκεκριμένα αρχεία).

4.1.1 SOFTWARE ISSUES

Issue #1

Τα SQL queries στην βάση για παραπάνω απο μια χώρα μάλλον δεν επιστρέφουν σωστά αποτελέσματα. Είναι πολύ λίγα data που επιστρέφονται. Ίσως είναι θέμα των αρχικών δεδομένων η μπορεί και του SQL query που χρησιμοποιούμε.

Issue #2

Τα barchart και scatterplot διαγράμματα για παραπάνω απο μια χώρα επιστρέφουν κάποιο σφάλμα και δεν δέχονται τα data όπως τα δίνουμε.

Issue #3

Τα mode ανα 5ετίες και 10ετίες δεν δουλεύουν νομίζω όπως θα έπρεπε. Ίσως είναι θέμα του SQL query μας ή των δεδομένων στην βάση.