

ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ  
ΤΜΗΜΑ ΜΗΧ. Η/Υ & ΠΛΗΡΟΦΟΡΙΚΗΣ

ΠΡΟΧΩΡΗΜΈΝΑ ΘΈΜΑΤΑ  
ΤΕΧΝΟΛΟΓΊΑΣ ΚΑΙ  
ΕΦΑΡΜΟΓΏΝ ΒΆΣΕΩΝ  
ΔΕΔΟΜΈΝΩΝ

---

ΠΡΟΓΡΑΜΜΑΤΙΣΤΙΚΗ  
ΕΡΓΑΣΙΑ ΓΙΑ ΤΟ  
ΑΚΑΔΗΜΑΪΚΟ ΈΤΟΣ 2018-  
2019

---

---

ΟΜΑΔΑ

ΌΘΩΝΑΣ ΓΚΑΒΑΡΔΙΝΑΣ, 2620

---

## ΤΕΛΙΚΗ ΑΝΑΦΟΡΑ

ΜΑΪΟΣ 2019

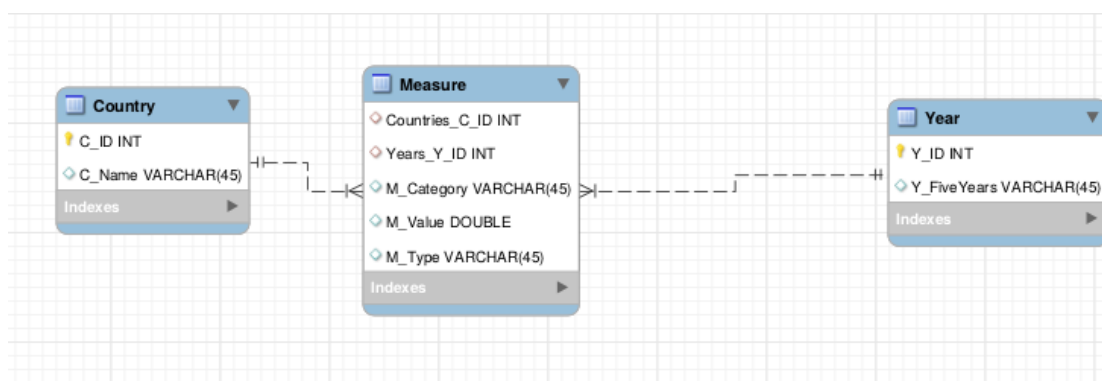
## ΙΣΤΟΡΙΚΟ ΠΡΟΗΓΟΥΜΕΝΩΝ ΕΚΔΟΣΕΩΝ

Ημερομηνία	Έκδοση	Περιγραφή	Συγγραφέας
2019/5/29	1	Ολόκληρη η εργασία	Όθωνας Γκαβαρδίνας

## 1 ΒΑΣΗ ΔΕΔΟΜΕΝΩΝ

### 1.1 ΣΧΕΣΙΑΚΟ ΣΧΗΜΑ ΣΕ ΛΟΓΙΚΟ ΕΠΙΠΕΔΟ

Το σχεσιακό σχήμα της βάσης δεδομένων (μπορείτε και αναλυτική λίστα των εντολών κατασκευής πινάκων, αλλά οπωσδήποτε αρχίστε με ένα workbench screenshot).



Η βάση διαθέτει τρείς πίνακες:

Country(C\_ID, C\_Name)

Για τις χώρες που χρησιμοποιούνται έχει δημιουργηθεί ένα μοναδικό Id, που τις χαρακτηρίζει, και επίσης διατηρείται το όνομα τους.

Year(Y\_ID, Y\_FiveYears)

Οι χρονιές που χρησιμοποιούνται έχουν ως Id, την ίδια τη χρονιά, και επίσης ένα πεδίο που κρατά την πενταετία στην οποία ανήκει.

Measure(M\_Category, M\_Value, M\_Type, Countries\_C\_ID, Years\_Y\_ID)

Εδώ βρίσκονται οι μετρικές που χρησιμοποιούνται οι οποίες έχουν ως ξένα κλειδιά, το κλειδί της χώρας, και το κλειδί της χρονιάς που αφορούν.

### 1.2 ΣΧΕΣΙΑΚΟ ΣΧΗΜΑ ΣΕ ΦΥΣΙΚΟ ΕΠΙΠΕΔΟ

#### 1.2.1 ΡΥΘΜΙΣΗ ΤΩΝ ΠΑΡΑΜΕΤΡΩΝ ΤΟΥ DBMS

innodb\_buffer\_pool\_size:

Για μικρές βάσεις δεδομένων, οι οποίες χωρούν ολίκληρες στην κύρια μνήμη είναι εφικτό να καθορίσουμε τους data buffers στο 10% παραπάνω από το μέγεθος της βάσης.

Με την παρακάτω εντολή πρόβαλα το μέγεθος του buffer pool:

```
show variables like 'innodb_buffer_pool_size';
```

Απ: 134217728 bytes

Με την παρακάτω εντολή έλαβα ως αποτέλεσμα για τη βάση μου που ονομάζεται mydb, ότι έχει μέγεθος 20529152 bytes.

Στη συνέχεια υπολόγισα το 10% της βάσης:

$$20529152 * 0.1 = 2052915,2 \text{ (~ } 2052915,2 \text{ bytes)}$$

Άθροισμα των 2 παραπάνω:

$$20529152 + 2052915 = 22582067 \text{ bytes.}$$

Οπότε έθεσα την παράμετρο:

```
set global innodb_buffer_pool_size=22582067;
```

Με την παρακάτω εντολή και το μέγεθος του buffer pool:

```
show variables like 'innodb_buffer_pool_size';
```

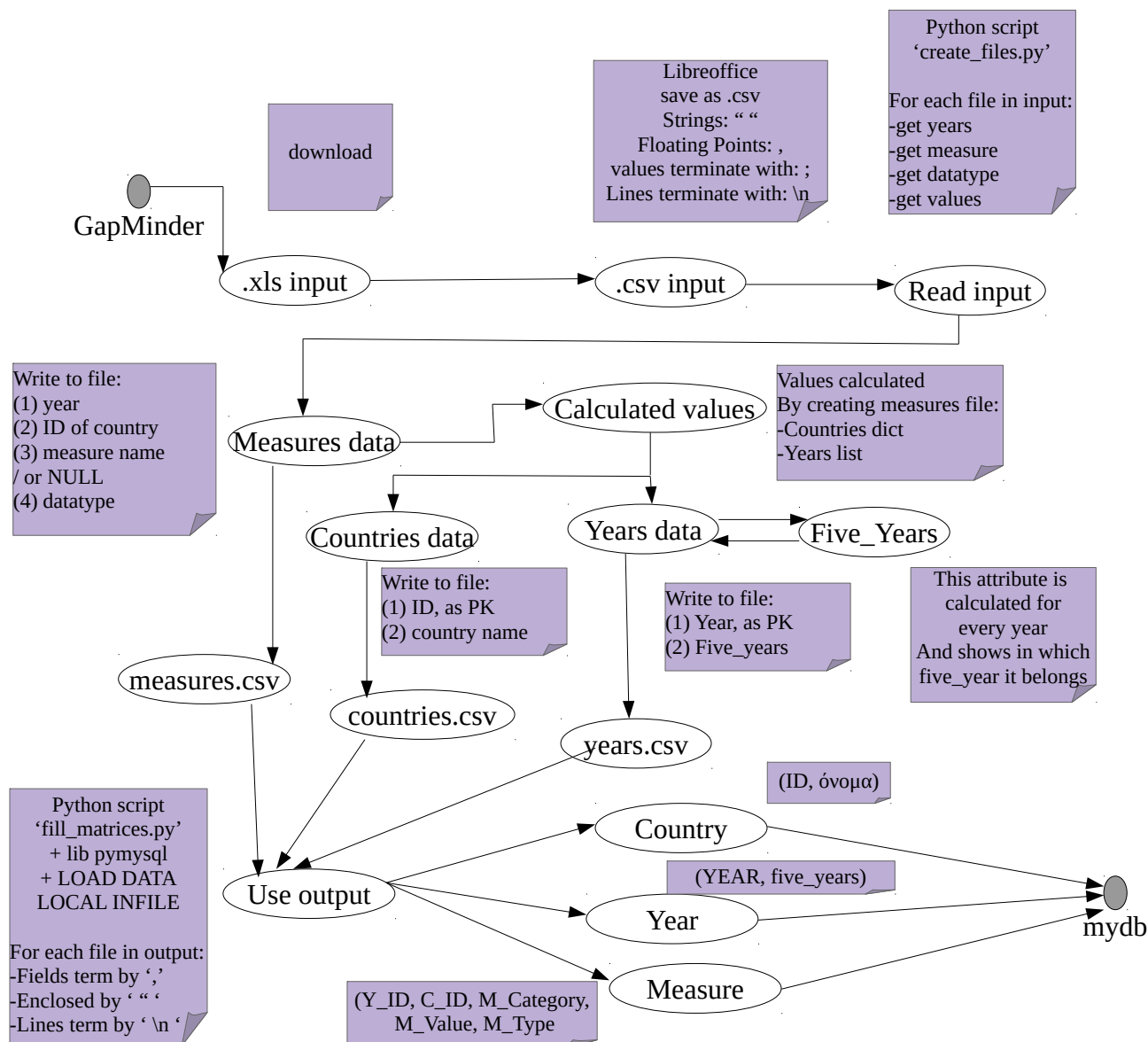
Απ: 134217728 bytes

→ Διαπίστωσα πως η βάση μου είναι πολύ μικρή και επομένως ο buffer pool διατηρεί το αρχικό μέγεθος.

## 2 ΑΡΧΙΤΕΚΤΟΝΙΚΗ ΛΟΓΙΣΜΙΚΟΎ

### 2.1 ΑΡΧΙΤΕΚΤΟΝΙΚΗ ΚΑΙ ΔΟΜΗ ETL

Γίνεται χρήση UML-based ETL, και το διάγραμμα είναι το εξής:



#### Περιγραφή της διαδικασίας:

1. Αρχικά τα δεδομένα κατεβαίνουν από το site της GapMinder, και βρίσκονται σε φάκελο, έχοντας τη μορφή xls αρχείων.
2. Γίνεται η μετατροπή των αρχείων σε αρχεία με μορφή csv, από το πρόγραμμα Libreoffice Calc, με την επιλογή Save As, και στη συνέχεια

αποθήκευση των αρχείων σε ξεχωριστό φάκελο. Στοιχεία των csv αρχείων:

- Τα Strings χωρίζονται με " ".
- Η κινητή υποδιαστολή χωρίζεται με ","
- Οι τιμές μίας γραμμής χωρίζονται με ";"
- Η κάθε γραμμή περιέχει μία χώρα και τις τιμές γι' αυτή.
- Η πρώτη γραμμή περιέχει μια μετρική, και τον τύπο της σε παρένθεση, και τις διάφορες χρονολογίες.

3. Με το script 'create\_files.py' γίνεται το διάβασμα των αρχείων του φακέλου με csv αρχεία.

4. Δημιουργείται το αρχείο 'measures.csv'.

Για κάθε αρχείο:

Από την πρώτη γραμμή:

- Κρατώνται οι διάφορες χρονιές, η μετρική, και ο τύπος της μετρικής.
- Κάθε χρονιά αν δεν την έχω προσθέσει την προσθέτω σε μια δομή.

Διαβάζω τις υπόλοιπες γραμμές του αρχείου και από τα δεδομένα αυτά ξεκινώ να δημιουργώ τις γραμμές του 'measures.csv'.

Μαζεύω τις χώρες που έχω συναντήσει σε δομή λεξικού και δίνω σε κάθε νέα χώρα ένα μοναδικό ID.

Παράγω γραμμή του αρχείου εξόδου η οποία περιλαμβάνει:

- Τη χρονιά που αντιστοιχεί σε μια στήλη.
- Το ID της χώρας που έχω κρατήσει πιο πάνω.
- Το όνομα της μετρικής.
- Τις διάφορες τιμές, και αν δεν υπάρχουν \N.  
(Το \N, στη διαδικασία load, αναγνωρίζεται ως NULL)
- Τον τύπο της μετρικής.

Επιστρέφονται οι δομές των countries, και years.

5. Δημιουργείται το αρχείο 'years.csv'.

Λαμβάνω το years list και ανα πέντε χρόνια σημειώνω μια πενταετία, κρατώντας, την αρχή της και το τέλος της.

Δημιουργώ τις γραμμές του αρχείου με τα εξής δεδομένα:

- Χρονιά (Primary Key)
- Πενταετία (που αποτελείται από αρχή και τέλος χωρισμένα με "-").

6. Δημιουργείται το αρχείο 'countries.csv'.

Λαμβάνω το countries dict, το κάνω sort κατά την τιμή του που είναι τα διάφορα IDs.

Σε κάθε γραμμή του αρχείου θα υπάρχουν:

- ID χώρας (Primary Key)
- Όνομα χώρας.

7. Με το script 'fill\_matrices.py', τη βιβλιοθήκη pymysql, καθώς και με τη χρήση της εντολής LOAD DATA LOCAL INFILE, φορτώνω τα δεδομένα μου στους αντίστοιχους πίνακες που βρίσκονται στην βάση μου.

Για να επιτευχθεί αυτό αρχικά εισέρχομαι στη βάση μου και πληκτρολογώ την εντολή 'SET GLOBAL local\_infile = 1;', έτσι ώστε να

επιτρέψω τη μεταφορά των δεδομένων στους πίνακες.

Στην εντολή περιγράφεται ότι τα δεδομένα μιας γραμμής είναι διαχωρισμένα με ',', περιβαλλόμενα από '"', και οι γραμμές είναι διαχωρισμένες με τον ειδικό χαρακτήρα '\n'.

Τα δεδομένα ακόμη, προστίθενται με τη σειρά που υπάρχουν στους πίνακες του σχήματός μου, δηλαδή:

- Country(ID, όνομα)
- Year(Year, πενταετία)
- Measure(Y\_ID, C\_ID, M\_Category, M\_Value, M\_Type)

## 2.2 ΔΙΑΓΡΑΜΜΑΤΑ ΠΑΚΕΤΩΝ / ΥΠΟΣΥΣΤΗΜΑΤΩΝ ΚΕΝΤΡΙΚΗΣ ΕΦΑΡΜΟΓΗΣ



Η εφαρμογή έχει τη μορφή φόρμας, επομένως για την αναπαράσταση ακολουθείται μια διαδικασία ροής.

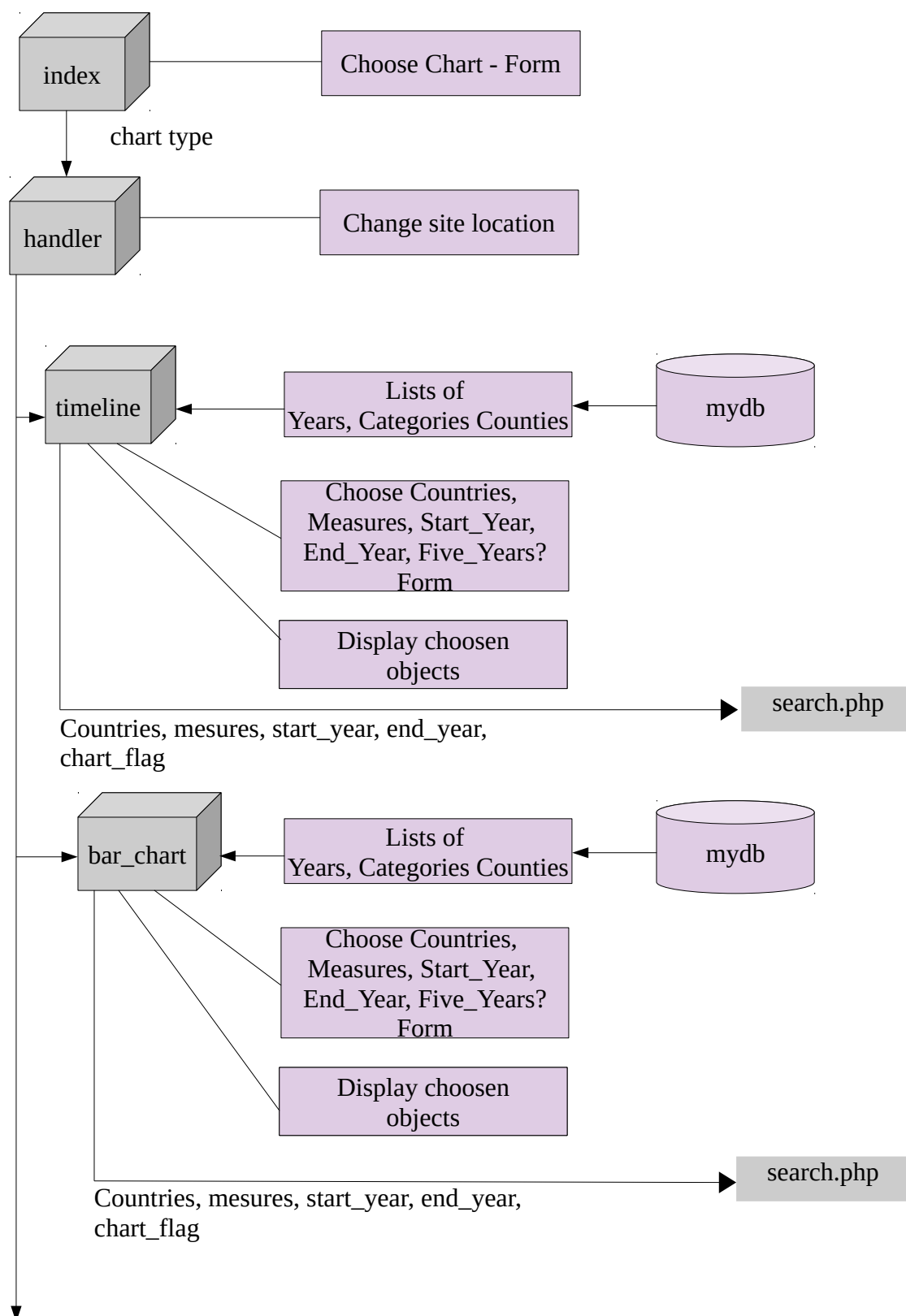
Index Menu: Εδώ ο χρήστης επιλέγει τύπο διαγράμματος, και μεταβιβάζεται στη φόρμα επιλογής δεδομένων.

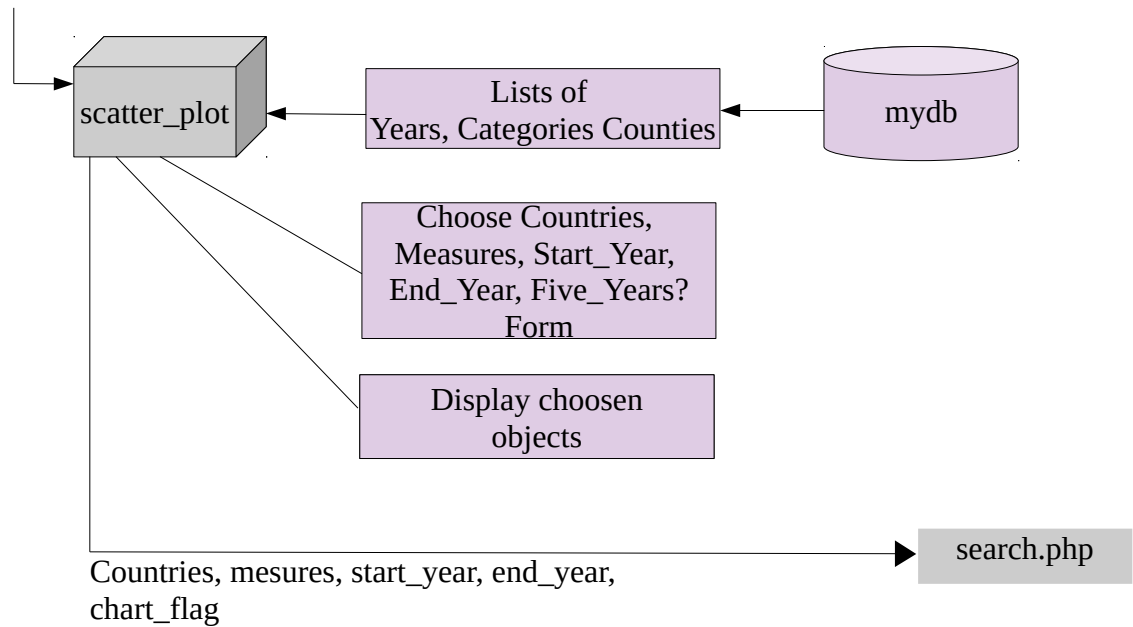
Specific Menu: Εδώ ο χρήστης επιλέγει δεδομένα για το διάγραμμά του.

Chart: Οπτική αναπαράσταση των δεδομένων στο επιλεγμένο γράφημα.

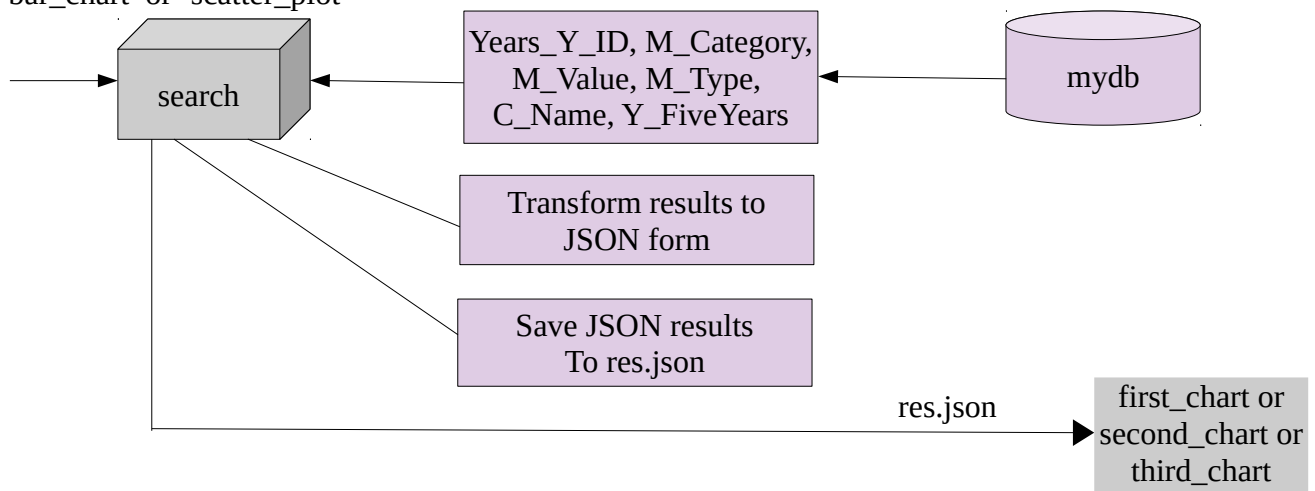


## 2.3 ΔΙΑΓΡΑΜΜΑ(ΤΑ) ΚΛΑΣΕΩΝ ΚΕΝΤΡΙΚΗΣ ΕΦΑΡΜΟΓΗΣ

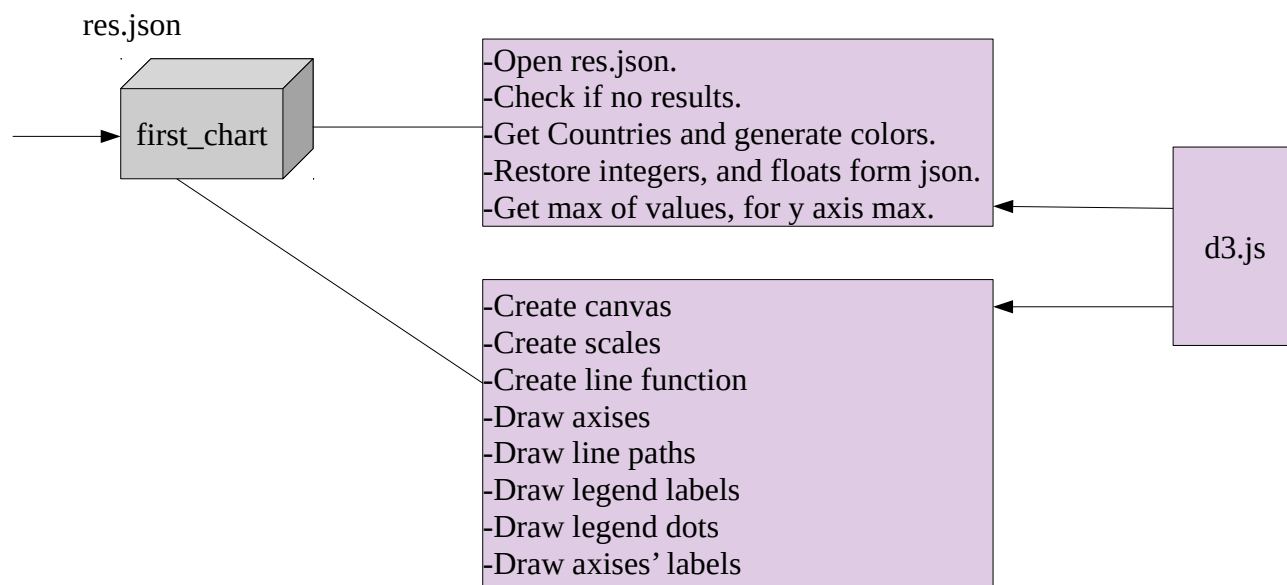
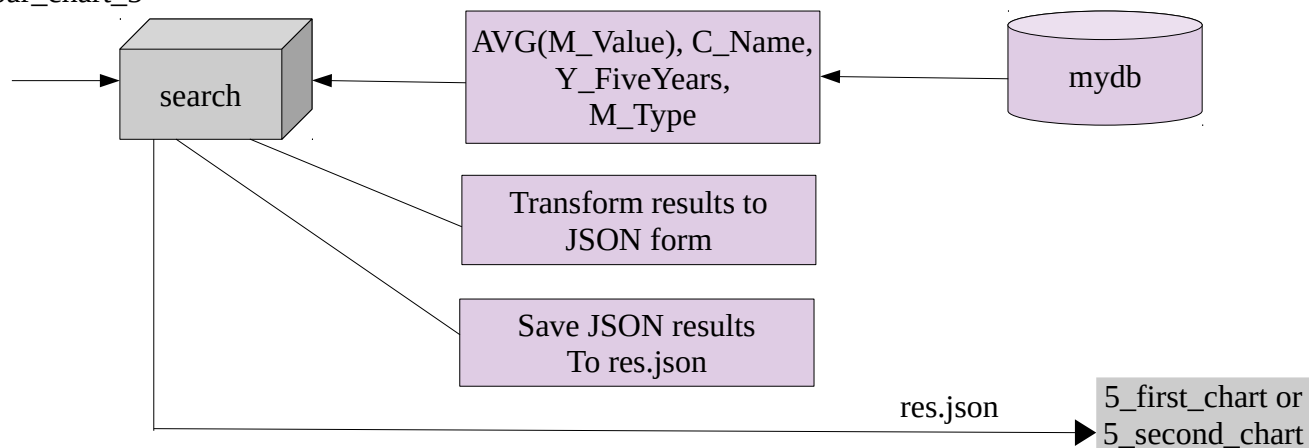


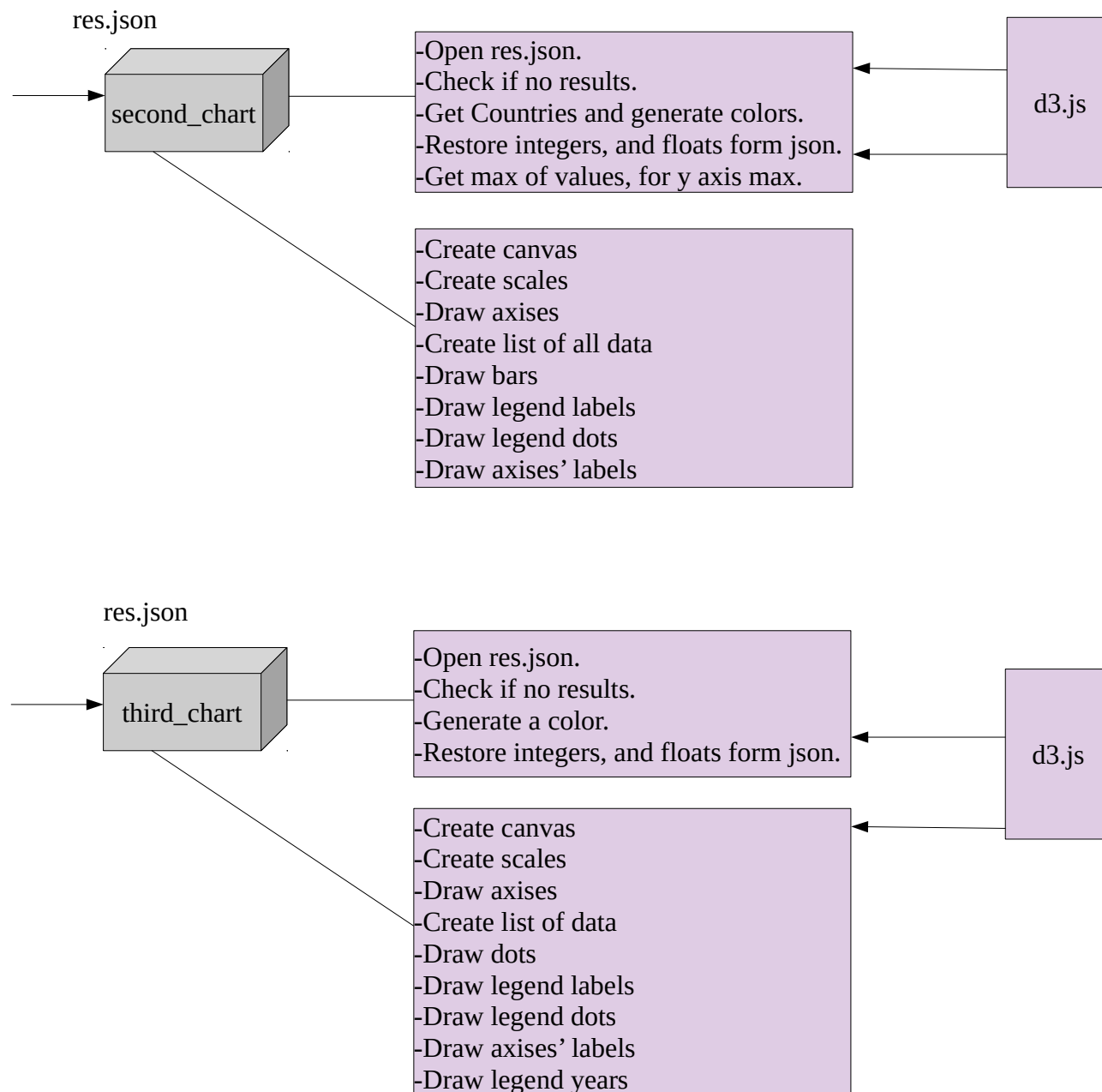


flag = 'timeline' or  
'bar\_chart' or 'scatter\_plot'



flag = 'timeline\_5' or  
'bar\_chart\_5'





Τα **5\_first\_chart** και **5\_second\_chart** είναι όπως τα **first\_chart** και **second\_chart** με τη διαφορά ότι το scaling για τον x axis, πλέον γίνεται με βάση τις πενταετίες. Επίσης, κάποιες μεταβλητές τροποποιήθηκαν έτσι ώστε να αρμώζουν στην aggregation ερώτηση που γίνεται.

### 3 ΥΠΟΔΕΙΓΜΑΤΑ ΕΡΩΤΗΣΕΩΝ ΚΑΙ ΑΠΑΝΤΗΣΕΩΝ

## Data Optimization

Please select the type of graph:

- ☒ Timeline  
☐ Bar Chart  
☐ Scatter Plot

Submit Query

### Παράδειγμα χρήσης:

-Αρχικά επιλέγω διάγραμμα timeline.

-Στη συνέχεια δίνω ως στοιχεία Countries=Greece, Germany, France. Measures=Electricity consumption total. Start year: 1978, End year: 1998.

## Timeline Graph

### Insert Countries

French Polynesia  
 Gabon  
 Gambia  
 Georgia  
 Germany  
 Ghana  
 Gibraltar  
 Greece

Submit Country

### Insert Measures

Hydro production total  
 Natural Gas Production total  
 Natural Gas Proved reserves total  
 Nuclear production total  
 Oil Consumption total  
 Oil proved reserves total  
 Total GDP  
 Total GNI PPP

Submit Measure

### Choose start year

1977  
 1978  
 1979  
 1980

Submit

### Choose end year

1996  
 1997  
 1998  
 1999

Submit

## Per five years?

YES  
 NO

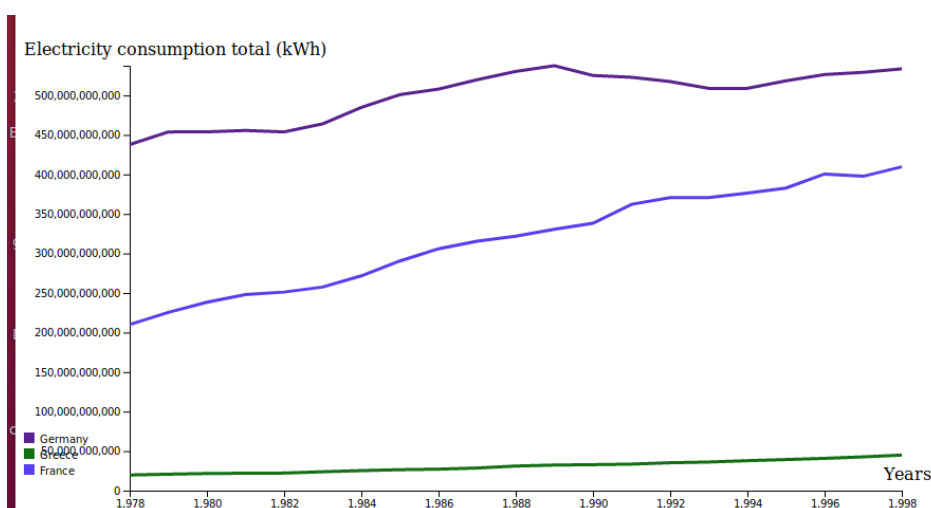
Submit

Reset

Countries: France,Greece,Germany

Measures: Electricity consumption total

from 1978 to 1998 per year .



### Άλλες λειτουργίες:

-Αν θέλω να διαγράψω κάποια επιλογή μου, δίνω τον αριθμό επιλογής (από το 0) της στο κουτάκι κάτω από τον τίτλο 'Remove element by index'.

-Με αυτό τον τρόπο μπορώ να διαγράψω χώρες και μετρικές που έχω επιλέξει.

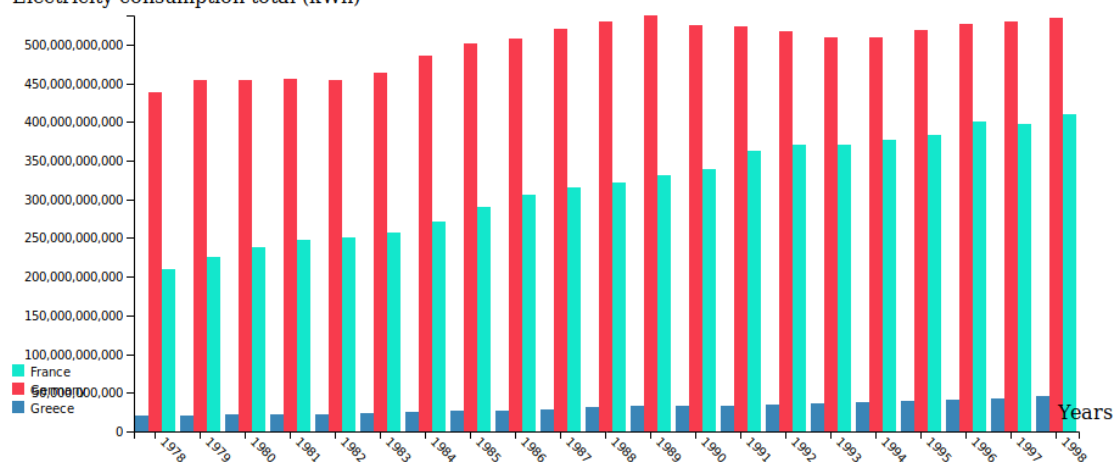
-Στο παράδειγμα, αν ήθελα να διαγράψω από τις χώρες την επιλογή 'Greece', θα έδινα τον αριθμό 1.

-Επίσης, αν δε δώσω αριθμό και πατήσω Remove Country θα διαγράψει την πρώτη χώρα στη λίστα επιλογής, δηλαδή στο παράδειγμα την 'France'.

-Υπάρχει επίσης το κουμπί 'Reset', με το οποίο καθαρίζουν όλες οι επιλογές, και το εύρος των ετών ορίζεται από 1800 έως 2013.

### Τα υπόλοιπα διαγράμματα:

Electricity consumption total (kWh)



Electricity consumption total (kWh)

