ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ ΤΜΗΜΑ ΜΗΧ. Η/Υ & ΠΛΗΡΟΦΟΡΙΚΗΣ

ΠΡΟΧΩΡΗΜΈΝΑ ΘΈΜΑΤΑ ΤΕΧΝΟΛΟΓΊΑΣ ΚΑΙ ΕΦΑΡΜΟΓΏΝ ΒΆΣΕΩΝ ΔΕΛΟΜΈΝΩΝ

ΠΡΟΓΡΑΜΜΑΤΙΣΤΙΚΗ ΕΡΓΑΣΙΑ ΓΙΑ ΤΟ ΑΚΑΔΗΜΑΪΚΟ ΈΤΟΣ 2019-2020

ΟΜΑΔΑ ΒΚΜ

BOYPONIKOΣ ΒΑΣΙΛΗΣ, 3152

ΚΟΥΓΙΑ ΙΩΑΝΝΑ, 2731

ΜΗΤΣΗΣ ΠΑΝΤΕΛΗΣ, 2298

ΤΕΛΙΚΗ ΑΝΑΦΟΡΑ

IΟΥΝΙΟΣ 2020

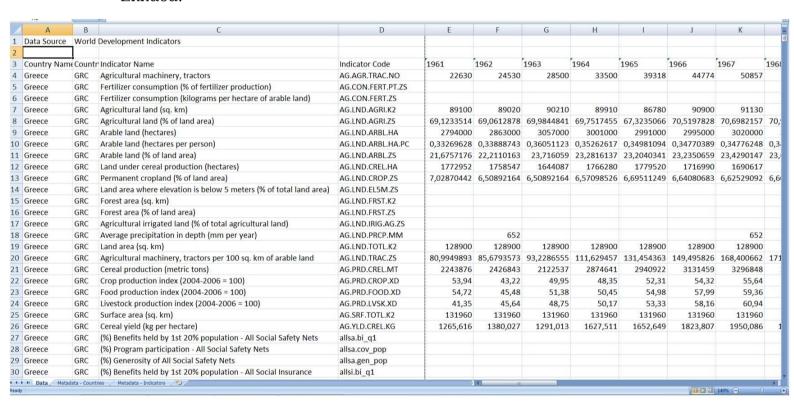
1 ΒΑΣΗ ΔΕΔΟΜΕΝΩΝ

1.1 ΣΧΕΣΙΑΚΌ ΣΧΉΜΑ ΣΕ ΛΟΓΙΚΌ ΕΠΊΠΕΔΟ

Για τη δημιουργία της βάσης χρησιμοποιήσαμε δεδομένα της WorldBank. Ενδιαφερόμαστε για τα δεδομένα που προσφέρονται ανά χώρα (http://data.worldbank.org/country).

Η δομή των csv αρχείων είναι η εξης: Έχουμε ένα αρχείο για κάθε χώρα. Κάθε στήλη αντιστοιχεί σε μια χρονολογία, και κάθε γραμμή σε έναν δείκτη. Συνδιάζοντας αυτά τα δύο παίρνουμε την τιμή (value) που έχει ένας δείκτης για κάποιο έτος.

Στον παρακάτω πίνακα βλέπουμε ενδεικτικά ένα κομμάτι του csv για την Ελλάδα.



Διαλέξαμε δώδεκα χώρες, δώδεκα δείκτες, για όλα τα έτη (απο 1960 μέχρι και σήμερα).

Μέσα σε αυτόν τον πίνακα υπάρχει πολλή επαναλλαμβανόμενη - άρχηστη- πληροφορία. Για να μειώσουμε τον πλεονασμό και την εξάρτηση των δεδομένων αποφασίσαμε να φέρουμε τη βάση μας σε 3NF κανονική μορφή χωρίζοντας τα δεδομένα σε πολλαπλούς πίνακες - αντί για έναν που θα κρατάει τη δομή του csv.

Βέβαια ο πίνακας Years δεν είναι σε 3NF, αλλά άν τον κάναμε θα καταλήγαμε με πολλούς και πολυ μικρούς πίνακες που θα έκαναν την δουλειά μας αγρείαστα δύσκολη.

Αυτός ο τρόπος θα κάνει τα queries πιο γρήγορα, και την πρόσθεση δεδομένων στη βάση πιο έυκολη. Το μεγαλυτερο τίμημα που πληρώνουμε εδώ είναι η πολυπλοκότητα των ερωτήσεων της σχεσιακής άλγεβρας.

Μέσω του script export_normalized_csv.py φέραμε τα αρχεία στην μορφη που θέλουμε να έχει η βάση μας (βλέπε 2.1). Ακολουθούν κομμάτια των πινακών Data και Years, καθως και οι πίνακες Indicators και Countries

	А	В	С	D	E
170	id	Country Code	Indicator Code	Year	Indicator Value
2		BRA	TX.VAL.MRCH.WL.CD	1960	
3	2	BRA	TX.VAL.MRCH.WL.CD	1961	1402500000
4	3	BRA	TX.VAL.MRCH.WL.CD	1962	1213500000
5	4	BRA	TX.VAL.MRCH.WL.CD	1963	1406900000
6	5	BRA	TX.VAL.MRCH.WL.CD	1964	1430900000
7	6	BRA	TX.VAL.MRCH.WL.CD	1965	1595300000
8	7	BRA	TX.VAL.MRCH.WL.CD	1966	1741000000
9	8	BRA	TX.VAL.MRCH.WL.CD	1967	1654600000
10	9	BRA	TX.VAL.MRCH.WL.CD	1968	1881600000
11	10	BRA	TX.VAL.MRCH.WL.CD	1969	2310420000
12	11	BRA	TX.VAL.MRCH.WL.CD	1970	2738370000
13	12	BRA	TX.VAL.MRCH.WL.CD	1971	2902400000
14	13	BRA	TX.VAL.MRCH.WL.CD	1972	4003540000
15	14	BRA	TX.VAL.MRCH.WL.CD	1973	6198800000
16	15	BRA	TX.VAL.MRCH.WL.CD	1974	7949290000
17	16	BRA	TX.VAL.MRCH.WL.CD	1975	8663000000
18	17	BRA	TX.VAL.MRCH.WL.CD	1976	10097380000
19	18	BRA	TX.VAL.MRCH.WL.CD	1977	12137800000
20	19	BRA	TX.VAL.MRCH.WL.CD	1978	12649110000

TABLE 1: PORTION OF DATA TABLE

	A	В	С
(3)	Year	5 Years	10 Years
2	1960	1960-1964	1960-1969
3	1961	1960-1964	1960-1969
4	1962	1960-1964	1960-1969
5	1963	1960-1964	1960-1969
6	1964	1960-1964	1960-1969
7	1965	1965-1969	1960-1969
8	1966	1965-1969	1960-1969
9	1967	1965-1969	1960-1969
10	1968	1965-1969	1960-1969
11	1969	1965-1969	1960-1969
12	1970	1970-1974	1970-1979
13	1971	1970-1974	1970-1979
14	1972	1970-1974	1970-1979
15	1973	1970-1974	1970-1979
16	1974	1970-1974	1970-1979
17	1975	1975-1979	1970-1979
18	1976	1975-1979	1970-1979
19	1977	1975-1979	1970-1979

TABLE 2: PORTION OF YEARS TABLE

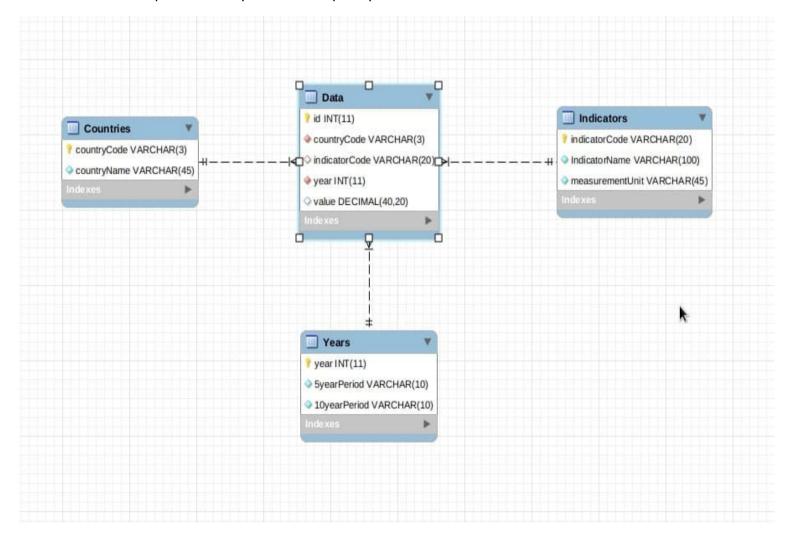
	A	В	
1	Country Code	Country Name	
2	BRA	Brazil	
3	CAN	Canada	
4	CHE	Switzerland	
5	DEU	Germany	
6	EGY	Egypt Arab Rep.	
7	GRC	Greece	
8	IND	India	
9	JPN	Japan	
10	KEN	Kenya	
11	MEX	Mexico	
12	RUS	Russian Federation	
13	TUN	Tunisia	
14	100000000000000000000000000000000000000	100000000000000000000000000000000000000	

TABLE 3: COUNTRIES TABLE

	A	В	C
0	Indicator Code	Indicator Name	Measurement unit
2	TX.VAL.MRCH.WL.CD	Merchandise exports by the reporting economy	current US\$
3	TM.VAL.MRCH.WL.CD	Merchandise imports by the reporting economy	current US\$
4	TG.VAL.TOTL.GD.ZS	Merchandise trade	% of GDP
5	NY.GDS.TOTL.ZS	Gross domestic savings	% of GDP
6	NY.GDP.TOTL.RT.ZS	Total natural resources rents	% of GDP
7	NY.GDP.PCAP.KD.ZG	GDP per capita growth	annual %
8	EG.IMP.CONS.ZS	Energy imports net	% of energy use
9	NY.GDP.DEFL.ZS	GDP deflator	base year varies by country
10	NY.EXP.CAPM.KN	Exports as a capacity to import	constant LCU
11	NE.TRD.GNFS.ZS	Trade	% of GDP
12	NE.GDI.TOTL.ZS	Gross capital formation	% of GDP
13	NE.EXP.GNFS.ZS	Exports of goods and services	% of GDP
14		1 - 1 - 1 - 1 - 1 - 1 - 1 - 1 - 1 - 1 -	
15			

TABLE 4: INDICATORS TABLE

Μέσω του MySQL Workbench φτιάξαμε το σχεσιακό σχήμα της βάσης δεδομένων που φαίνεται στην παρακάτω εικόνα.



Οι εντολές για την δημιουργία των πινάκων είναι οι εξής:

- CREATE TABLE IF NOT EXISTS Countries (
 countryCode VARCHAR(3) NOT NULL UNIQUE,
 countryName VARCHAR(45) NOT NULL UNIQUE,
 PRIMARY KEY (countryCode)) ENGINE = InnoDB"
- CREATE TABLE IF NOT EXISTS Years (
 year INT(11) NOT NULL UNIQUE,
 5yearPeriod VARCHAR(11) NOT NULL,
 10yearPeriod VARCHAR(11) NOT NULL,
 PRIMARY KEY (year)) ENGINE = InnoDB
- CREATE TABLE IF NOT EXISTS Indicators (
 indicatorCode VARCHAR(20) NOT NULL UNIQUE,
 indicatorName VARCHAR(100) NOT NULL UNIQUE,
 measurementUnit VARCHAR(45) NOT NULL,
 PRIMARY KEY (indicatorCode)) ENGINE = InnoDB
- CREATE TABLE IF NOT EXISTS Data (
 id INT(11) NOT NULL UNIQUE,
 countryCode VARCHAR(3) NOT NULL,
 indicatorCode VARCHAR(20),
 year INT(11) NOT NULL,
 value DECIMAL(40,20),
 PRIMARY KEY (id),
 FOREIGN KEY (countryCode) REFERENCES
 Countries(countryCode) ON DELETE CASCADE ON UPDATE CASCADE,

FOREIGN KEY (indicatorCode) REFERENCES Indicators(indicatorCode) ON DELETE CASCADE ON UPDATE CASCADE,

FOREIGN KEY (year) REFERENCES Years(year) ON DELETE CASCADE ON UPDATE CASCADE) ENGINE = InnoDB

1.2 ΣΧΕΣΙΑΚΌ ΣΧΉΜΑ ΣΕ ΦΥΣΙΚΟ ΕΠΊΠΕΔΟ

1.2.1 ΡΎΘΜΙΣΗ ΤΩΝ ΠΑΡΑΜΈΤΡΩΝ ΤΟΥ DBMS

Δεν έχουμε μεταβάλει τις παραμέτρους του DBMS από τις default τιμες τους.

1.2.2 ΡΎΘΜΙΣΗ ΤΟΥ ΦΥΣΙΚΟΎ ΣΧΉΜΑΤΟΣ ΤΗΣ ΒΆΣΗΣ $\Delta \text{E}\Delta \text{OM}\texttt{E} \text{N} \Omega \text{N}$

Δεν έχουμε ορίσει ευρετήρια ούτε όψεις.

1.2.3 ΡΎΘΜΙΣΗ ΑΣΦΆΛΕΙΑΣ

Δικαίωματα καταχώρησης ή διαγραφής δεδομένων έχουν μόνο οι admins μιας και η βάση προστατεύεται με μυστικό κωδικό.

2 ΑΡΧΙΤΕΚΤΟΝΙΚΗ ΛΟΓΙΣΜΙΚΟΎ

2.1 APXITEKTONIKH KAI ΔOMH ETL

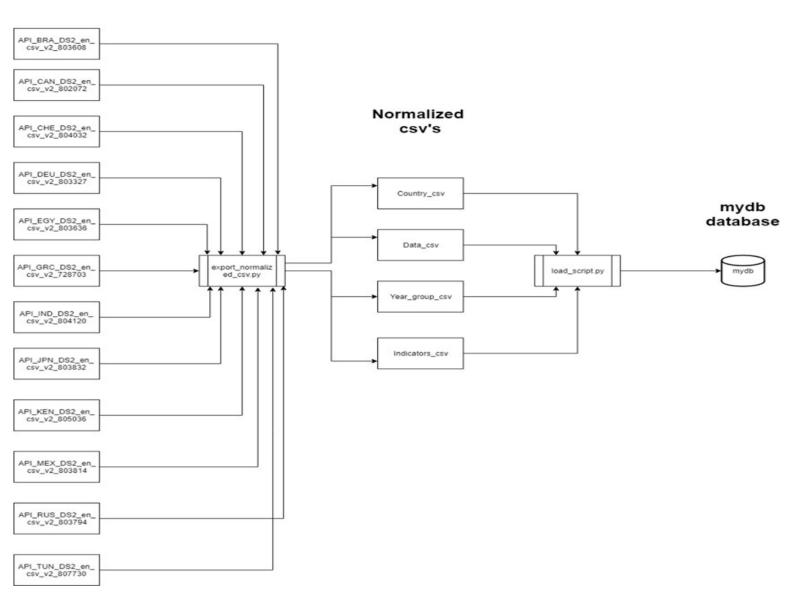
Για την ETL διαδικασία δημιουργήσαμε δύο scripts, και δεν κάναμε καμία παρέμβαση manually.

Το script export_normalized_csv.py παίρνει ώς είσοδο τα αρχεία των χωρών που επιλέξαμε αυτούσια, όπως τα κατεβάσαμε από τη WorlBank, και φτιάχνει τα norminalized csv's. Για να τρέξουμε το script, δίνουμε σαν όρισμα το όνομα του φακέλου στον οποίο βρίσκονται τα αρχεία. Η λίστα με τους δείκτες που επιλέξαμε ορίζεται μέσα στο script. Αρχικά για κάθε αρχείο, κρατάμε μόνο τις γραμμές με τους επιλεγμένους δείκτες. Έπειτα δημιουργούμε τέσσερα αρχεία, ένα για κάθε πίνακα, και τα συμπληρώνουμε αναλόγως.

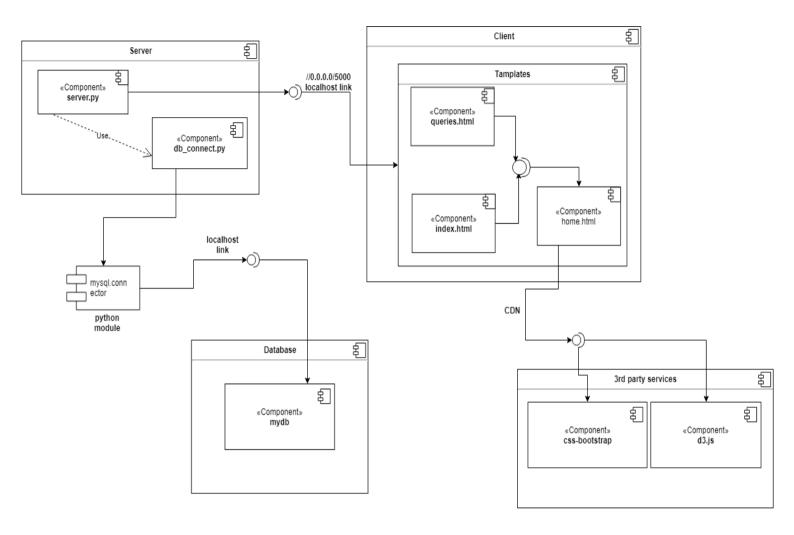
Το load_script.py φορτώνει τα δεδομένα στη βάση αφού πρώτα συνδεθεί με αυτήν. Παίρνει ως είσοδο τα norminalized csv's. Ένα σημείο άξιο να σημειωθεί ειναι η εντολή "SET FOREIGN_KEY_CHECKS = 0;" που πρέπει να εκτελεστεί πρωτού φορτωθούν τα αρχεία, και αφού ολοκληρωθεί η φόρτωση εκτελείται η "SET FOREIGN_KEY_CHECKS = 1;". Εξ'ορισμού η μεταβλητή αυτή είναι στο 1. Το σχήμα της βάσης και τα ξένα κλειδιά έχουν ήδη οριστεί. Όμως έτσι η mysql βγάζει error στα ξένα κλειδιά γιατί υποτίθεται ότι πρέπει πρώτα να γεμίσεις τους πίνακες με δεδομένα και μετά με alter table να ορίσεις τα ξένα κλειδιά. Επομένως με αυτές τις δύο εντολές λύνεται εύκολα το θέμα αυτό. Στη συνέχεια φορτώνει κάθε ένα πίνακα χρησιμοποιώντας την εντολή load data infile.

Η εξαγωγή του backup έγινε αφού φορτώθηκε η βάση, μέσω του WorkBench.

Raw csv's



2.2 ΔΙΑΓΡΆΜΜΑ(ΤΑ) ΚΛΆΣΕΩΝ ΚΕΝΤΡΙΚΗΣ ΕΦΑΡΜΟΓΗΣ



Μετά τη δημιουργία της βάσης και αφού τα δεδομένα έχουν φορτωθεί σε αυτή, η εφαρμογή χωρίζεται σε τρία μέρη. Το πρώτο είναι η ίδια η βάση δεδομένων (mydb) με την οποία συνδέεται ο server και κάνει ερωτήσεις μέσω του mysgl.connector.

Το δεύτερο μέρος είναι ο Server που έχει δυο βασικές αρμοδιότητες. Να δημιουργεί μια σύνδεση με την βάση mydb και να της στέλνει ερωτήσεις σύμφωνα με τις επιλογές του χρήστη (db connect.py).

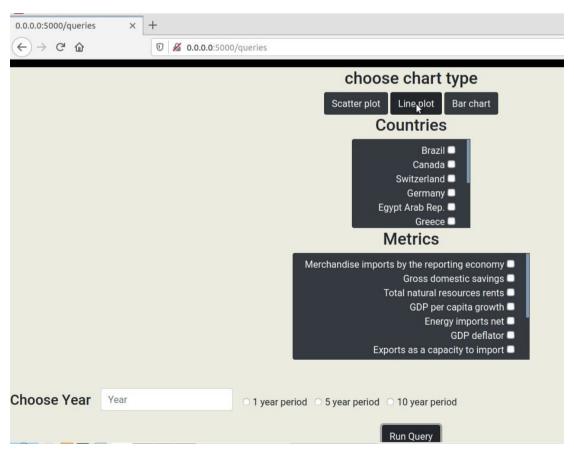
Το τρίτο και τελευταίο μέρος είναι ο Client. Σε αυτό το μέρος έχουμε το GUI της εφαρμογής. Εδώ ο χρήστης επιλέγει τι ερωτήσεις θέλει να κάνει στη βάση μέσα από τη φόρμα επιλογών που του παρουσιάζουμε. Επίσης μπορεί να επιλέξει και τα διαγράμματα που επιθυμεί να δει. Σε αυτό το μέρος χρησιμοποιούνται και τα εργαλεία bootstrap και d3.js για το Style της σελίδας και τη δημιουργία των διαγραμμάτων αντίστοιχα.

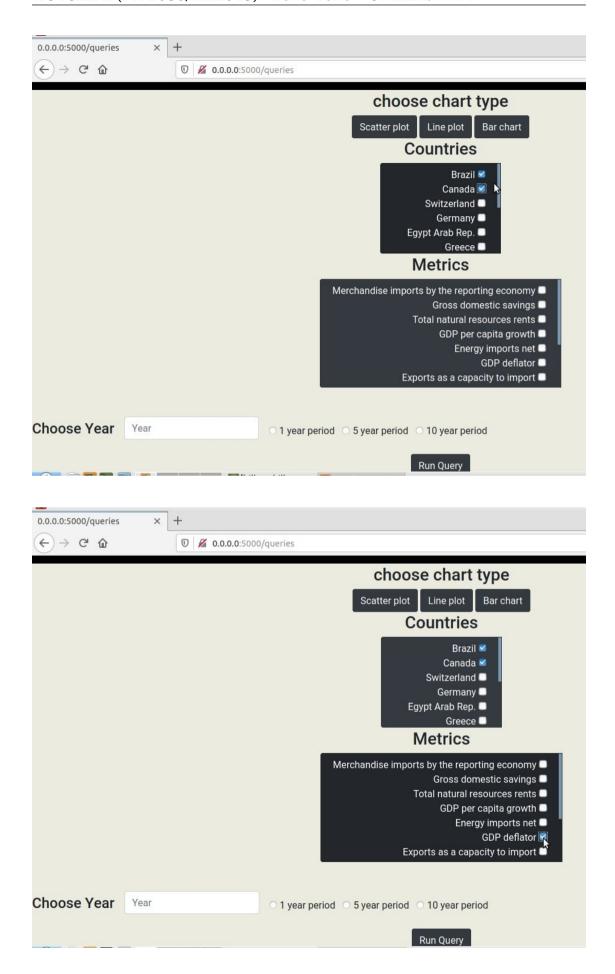
3 ΥΠΟΔΕΊΓΜΑΤΑ ΕΡΩΤΉΣΕΩΝ ΚΑΙ ΑΠΑΝΤΉΣΕΩΝ

Δεν καταφέραμε να ολοκληρώσουμε την υλοποίηση ακριβώς οπως τη σκεφτόμασταν. Υπάρχει πρόβλημα με την οπτικοποίηση μέσω της d3. Συγκεκριμένα, ενώ ακολουθήσαμε τα βήματα από τα tutorials που είδαμε, τα plots, όπως θα δείτε βγαίνουν ανάποδα, και χωρίς τιμές τον άξονα y, παρόλο που τα αποτελέσματα μπαίνουν σωστά στη συνάρτηση. Γι'αυτό παραθέτουμε και τα αποτελέσματα του κάθε query, δηλαδή τις παραμέτρους εισόδου της d3. Επίσης υλοποιούνται μόνο το line plot.

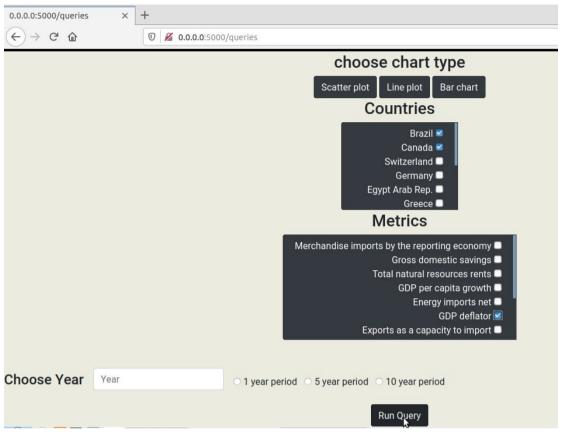
Αυτές είναι οι ερωτήσεις που μπορούν να απαντηθούν απο το πρόγραμμά μας για όλα τα έτη (χωρίς δηλαδή να επιλέξουμε έτος), ή για περιόδους 5ετίας ή 10ετίας.

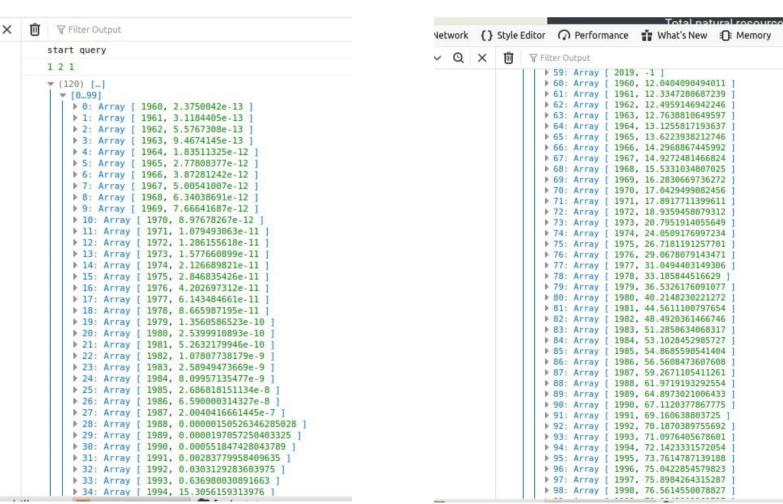
- επιλογή μιας χώρας και ενός δείκτη
- επιλογή μιας γώρας και δύο ή περισσότερων δεικτών
- επιλογή δύο ή περισσοτέρων χωρών, και ενός δείκτη
- 1. Ακολουθεί ένα ολοκληρωμένο παράδειγμα (βήμα-βήμα) με επιλογή δύο χωρών και ενός δείκτη χωρίς επιλογή έτους

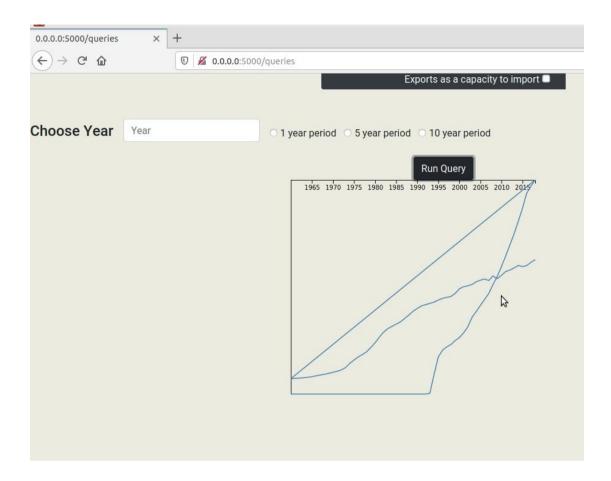




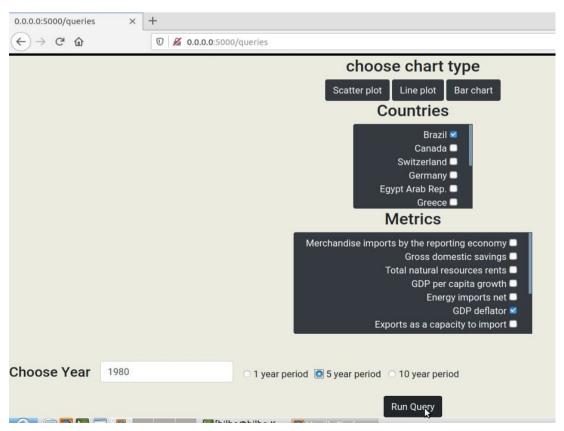
ΠΘΤ&ΕΒΔ (MYE030/ΠΛΕ045) 2019-2020 OMAΔA: BKM

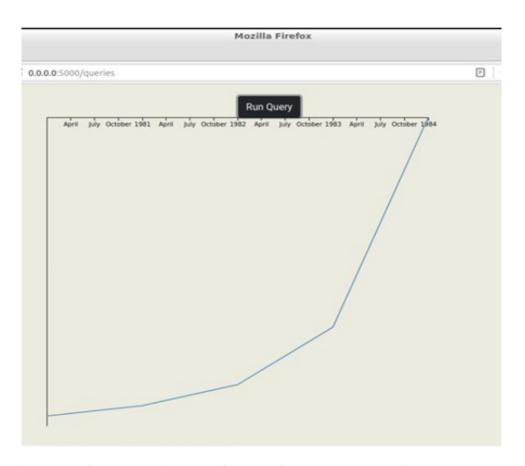






2. Ακολουθεί ένα παράδειγμα με επιλογή μίας χώρας, ενός δείκτη, για μια 5ετία.

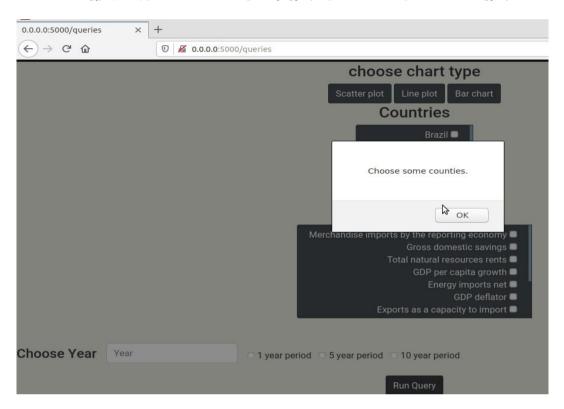




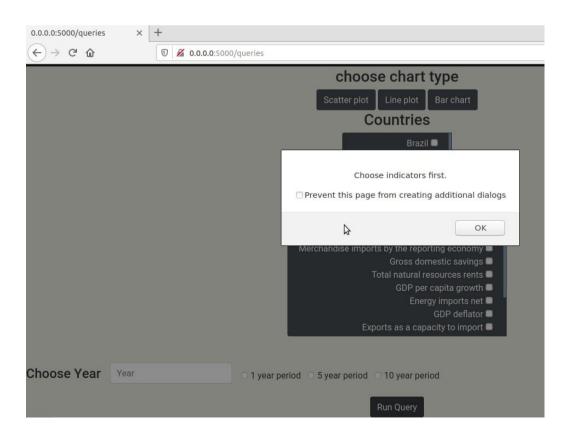
Εύκολα εδώ διαπιστώνεται, ότι ο τρόπος που υπολογίζονται τα plots, πέρα απο το παρουσιαστικό της υπόθεσης, είναι λάθος.

Ακολουθούν κάποια screenshots από το error-handling όσον αφορά τις επιλογές του χρήστη.

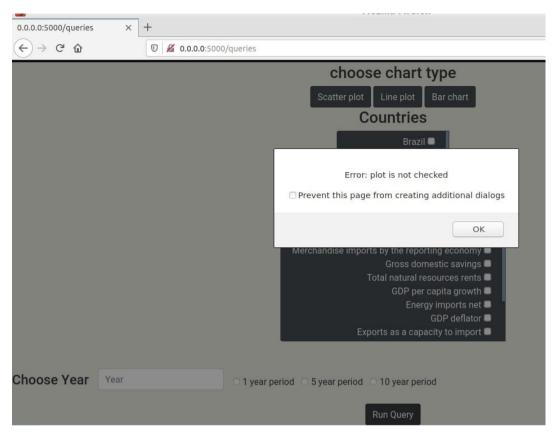
ο χρήστης πατάει run query χωρίς να επιλέξει κάποια χώρα



• ο χρήστης πατάει run query χωρίς να επιλέξει κάποιον δείκτη



• ο χρήστης πατάει run query χωρίς να επιλέξει κάποιο plot



• ο χρήστης επιλέγει έτος για το οποίο δεν έχει γίνει καταγραφή

