

ΠΡΟΧΩΡΗΜΕΝΑ ΘΕΜΑΤΑ ΤΕΧΝΟΛΟΓΙΑΣ
ΚΑΙ ΕΦΑΡΜΟΓΩΝ ΒΑΣΕΩΝ ΔΕΔΟΜΕΝΩΝ

ΠΡΟΓΡΑΜΜΑΤΙΣΤΙΚΗ ΕΡΓΑΣΙΑ ΓΙΑ
ΤΟ ΑΚΑΔΗΜΑΪΚΟ ΕΤΟΣ 2022-2023

ΟΜΑΔΑ

ΣΙΔΗΡΟΠΟΥΛΟΣ ΠΑΝΑΓΙΩΤΗΣ, 4489

ΖΕΖΟΣ ΑΡΓΥΡΙΟΣ, 4588

ΤΕΛΙΚΗ ΑΝΑΦΟΡΑ

ΜΑΪΟΣ 2023

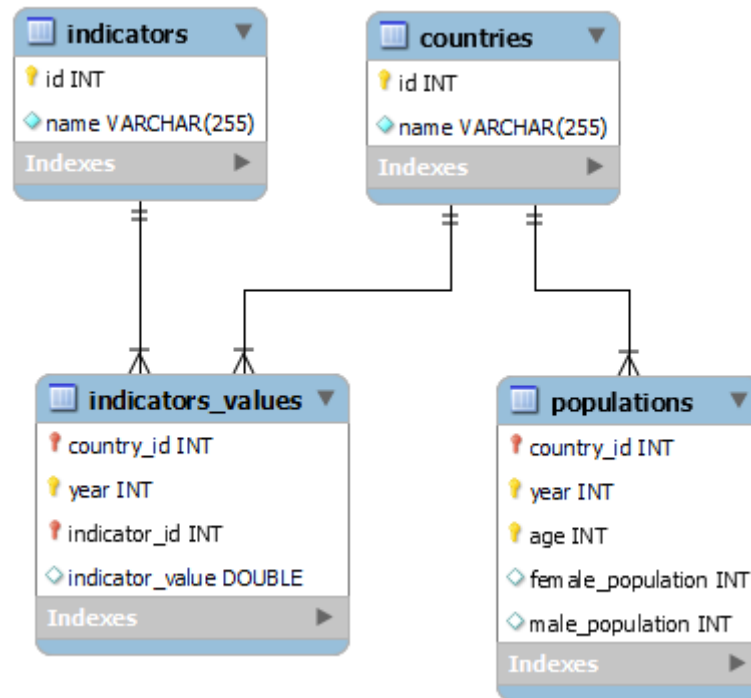
ΙΣΤΟΡΙΚΟ ΠΡΟΗΓΟΥΜΕΝΩΝ ΕΚΔΟΣΕΩΝ

Ημερομηνία	Έκδοση	Περιγραφή	Συγγραφέας
2023/05/23	1.0	Τελική έκδοση της εφαρμογής	Σιδηρόπουλος Παναγιώτης, Ζέζος Αργύριος

1 ΒΑΣΗ ΔΕΔΟΜΕΝΩΝ

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

1.1 ΣΧΕΣΙΑΚΟ ΣΧΗΜΑ ΣΕ ΛΟΓΙΚΟ ΕΠΙΠΕΔΟ



Σχήμα 1.1 Σχεσιακό σχήμα της βάσης δεδομένων του συστήματος

Επιλέξαμε το long format για τους πίνακες `indicators_values` και `populations` διότι το format αυτό διευκολύνει την προσθήκη νέων δεδομένων για νέες χρονιές και την προσθήκη νέων δεικτών χωρίς να χρειαστεί τροποποιηθεί ο ορισμός των πινάκων με `alter tables`.

Επίσης επιλέξαμε να αποθηκεύσουμε τα δεδομένα για τους πληθυσμούς των χωρών ανά ηλικία στον πίνακα `populations` έτσι ώστε να έχουμε τα λεπτομερείς δεδομένα για του πληθυσμούς συγκεντρωμένα σε έναν πίνακα και μετά μέσω διαφορετικών ερωτήσεων στον ίδιο πίνακα να πάρουμε ότι δεδομένα χρειαζόμαστε (πχ συνολικούς πληθυσμούς, ομαδοποιήσεις χρονιών κ.α.).

```
CREATE DATABASE IF NOT EXISTS world_data_visualizer;

USE world_data_visualizer;

SET GLOBAL local_infile = TRUE;
```

```
CREATE TABLE `countries` (  
  `id` int NOT NULL,  
  `name` varchar(255) NOT NULL,  
  PRIMARY KEY (`id`),  
  UNIQUE KEY `id_UNIQUE` (`id`),  
  UNIQUE KEY `official_name_UNIQUE` (`name`)  
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4_0900_ai_ci;  
  
CREATE TABLE `indicators` (  
  `id` int NOT NULL,  
  `name` varchar(255) NOT NULL,  
  PRIMARY KEY (`id`),  
  UNIQUE KEY `name_UNIQUE` (`name`)  
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4_0900_ai_ci;  
  
CREATE TABLE `indicators_values` (  
  `country_id` int NOT NULL,  
  `year` int NOT NULL,  
  `indicator_id` int NOT NULL,  
  `indicator_value` double DEFAULT NULL,  
  PRIMARY KEY (`country_id`,`year`,`indicator_id`),  
  KEY `fk_indicator_id_idx` (`indicator_id`),  
  CONSTRAINT `fk_c_id` FOREIGN KEY (`country_id`) REFERENCES `countries` (`id`)  
  ON DELETE CASCADE ON UPDATE CASCADE,  
  CONSTRAINT `fk_indicator_id` FOREIGN KEY (`indicator_id`) REFERENCES  
  `indicators` (`id`) ON DELETE CASCADE ON UPDATE CASCADE  
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4_0900_ai_ci;  
  
CREATE TABLE `populations` (  
  `country_id` int NOT NULL,  
  `year` int NOT NULL,  
  `age` int NOT NULL,  
  `female_population` int DEFAULT NULL,  
  `male_population` int DEFAULT NULL,
```

```
PRIMARY KEY (`country_id`,`year`,`age`),  
KEY `idx_populations` (`country_id`,`year`,`age`),  
CONSTRAINT `fk_country_id` FOREIGN KEY (`country_id`) REFERENCES `countries`  
(`id`) ON DELETE CASCADE ON UPDATE CASCADE  
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4_0900_ai_ci;
```

1.2 ΣΧΕΣΙΑΚΟ ΣΧΗΜΑ ΣΕ ΦΥΣΙΚΟ ΕΠΙΠΕΔΟ

1.2.1 ΡΥΘΜΙΣΗ ΤΩΝ ΠΑΡΑΜΕΤΡΩΝ ΤΟΥ DBMS

Name	Value
default_tmp_storage_engine	InnoDB

Name	Value
innodb_buffer_pool_chunk_size	8388608
innodb_buffer_pool_dump_at_shutdown	ON
innodb_buffer_pool_dump_now	OFF
innodb_buffer_pool_filename	ib_buffer_pool
innodb_buffer_pool_instances	1
innodb_buffer_pool_load_abort	OFF
innodb_buffer_pool_load_at_startup	ON
innodb_buffer_pool_load_now	OFF
innodb_buffer_pool_size	8388608

1.2.2 ΡΥΘΜΙΣΗ ΤΟΥ ΦΥΣΙΚΟΥ ΣΧΗΜΑΤΟΣ ΤΗΣ ΒΑΣΗΣ ΔΕΔΟΜΕΝΩΝ

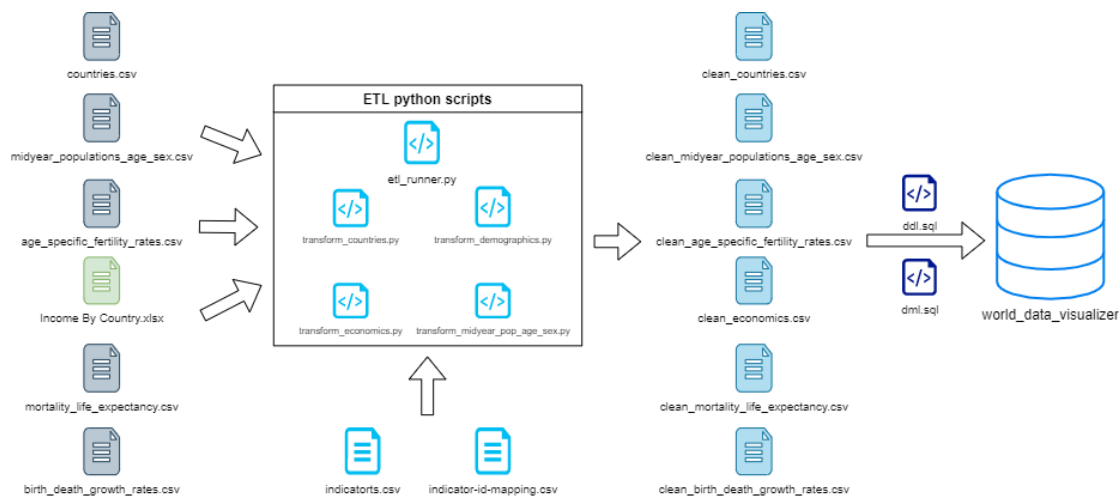
Δεν έχουν δημιουργηθεί όψεις και ευρετήρια πέραν αυτών που δημιουργούνται αυτόματα για τα primary και foreign keys από το DBMS.

1.2.3 ΡΥΘΜΙΣΗ ΑΣΦΑΛΕΙΑΣ

Δεν έχουν οριστεί διαφορετικοί ρόλοι και χρήστες του συστήματος με δικαιώματα καταχώρησης ή ανάκτησης δεδομένων.

2 ΑΡΧΙΤΕΚΤΟΝΙΚΗ ΛΟΓΙΣΜΙΚΟΥ

2.1 ΑΡΧΙΤΕΚΤΟΝΙΚΗ ΚΑΙ ΔΟΜΗ ETL



Σχήμα 2.1 ETL diagram

Η ETL είναι αυτοματοποιημένη διαδικασία χωρίς να χρειαστεί να παρέμβουμε manually στα δεδομένα:

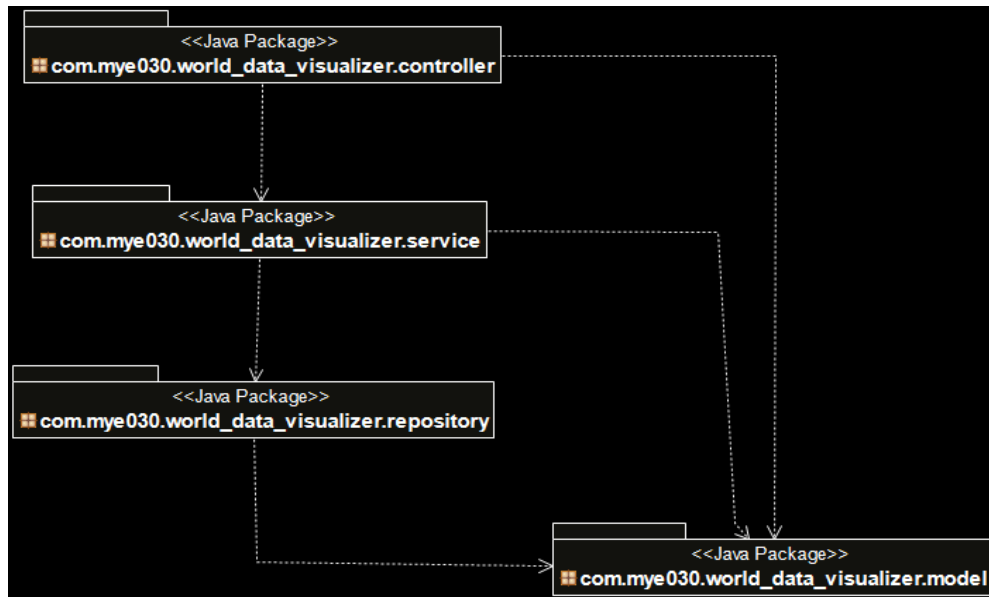
Οδηγίες για την εκτέλεση ETL διαδικασίας.

1. Μετασχηματίζουμε τα extracted δεδομένα εκτελώντας το python script: `data\etl-process\etl_runner.py`
2. Δημιουργούμε το σχήμα της βάσης εκτελώντας το sql script: `data\etl-process\ddl.sql`
3. Φορτώνουμε τα μετασχηματισμένα δεδομένα του βήματος 1, εκτελώντας το sql script: `data\etl-process\dml.sql`

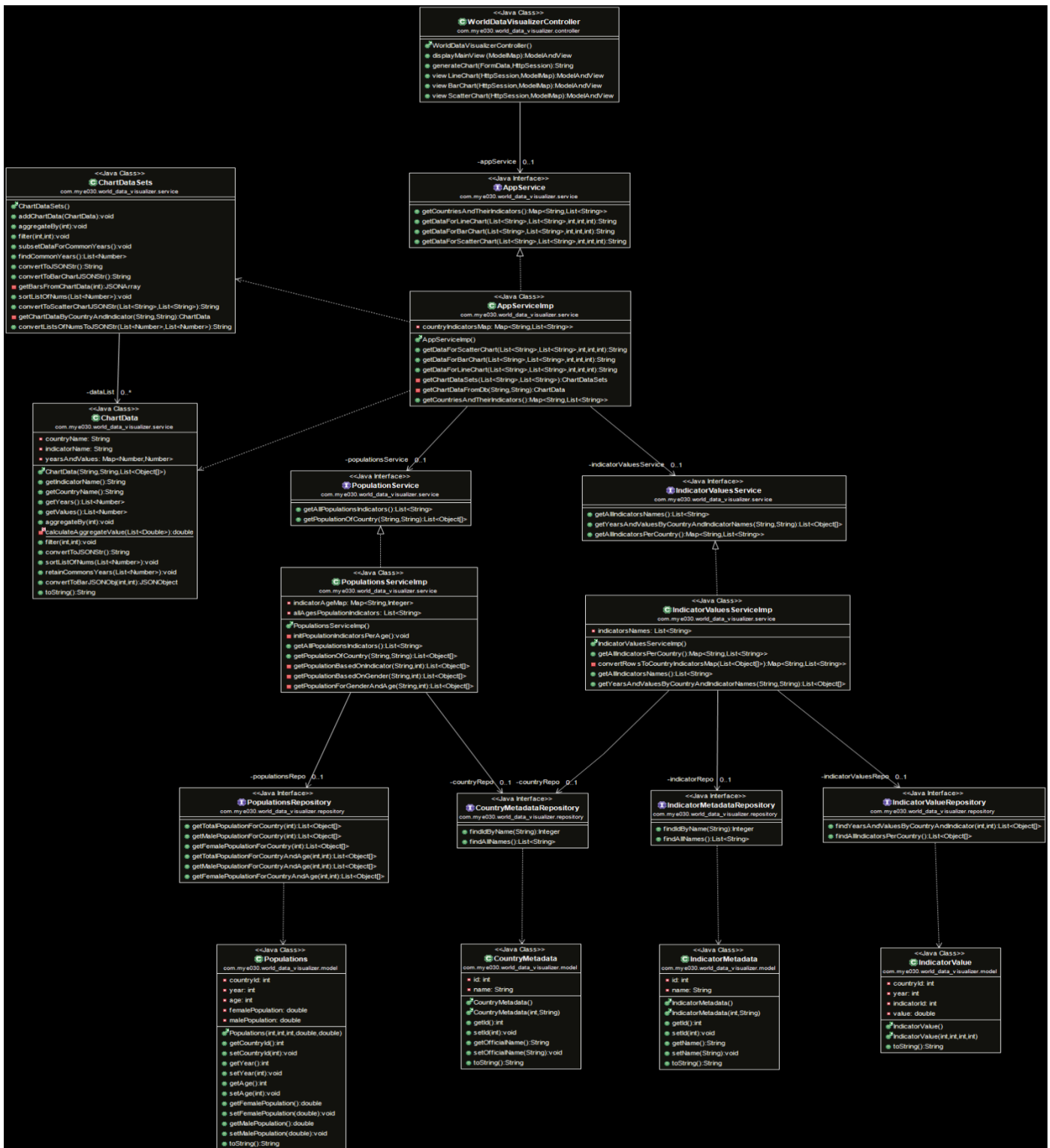
Παρατηρήσεις:

- Στο “transform_countries.py” script φαίνονται διάφορες τροποποιήσεις δεδομένων που δεν βρίσκονταν στη σωστή μορφή. Για παράδειγμα υπήρχαν εγγραφές που δεν είχαν Official όνομα, άλλες είχαν εισαγωγικά στο Official ονόμα τους ή κόμματα στο ονόμα τους κτλ.
- Για το “Income By Country.xlsx” αρχείο στο transform_economic.py script για κάθε sheet του excel αρχείου διαγράφουμε τις εγγραφές που περιέχουν null τιμές (δηλαδή για εγγραφές που έχουν τιμές τις δύο τελείες “..”) και ομαδοποιούμε τις στήλες ανά 5 χρόνια και το προσθέτουμε σε ένα dataframe που τελικά θα μετατραπεί στο αρχείο clean_economics.csv.
- Στο “data\etl-process\dml.sql” αρχείο μετά το φόρτωμα των δεδομένων εκτελούμε και ένα delete statement το οποίο διαγράφει όλες τις εγγραφές του πίνακα countries που το id της χώρας δεν βρίσκεται σε κανέναν από του πίνακες με τα δεδομένα των δεικτών.

2.2 ΔΙΑΓΡΑΜΜΑΤΑ ΠΑΚΕΤΩΝ / ΥΠΟΣΥΣΤΗΜΑΤΩΝ ΚΕΝΤΡΙΚΗΣ ΕΦΑΡΜΟΓΗΣ



Σχήμα 2.2.1 Διάγραμμα πακέτων εφαρμογής



Σχήμα 2.2.2 Διάγραμμα κλάσεων του συστήματος

3 ΥΠΟΔΕΙΓΜΑΤΑ ΕΡΩΤΗΣΕΩΝ ΚΑΙ ΑΠΑΝΤΗΣΕΩΝ

Index σελίδα:

World Data Visualizer

Make a Chart

Chart Generator σελίδα:

Chart Generator

Chart Type:

Bar Chart

Country Indicator

Add Pair

Filter:

Min year Max year

Aggregate:

No aggregation

Generate Chart

Συμπλήρωση πεδίων της φόρμας χρησιμοποιώντας dropdown λίστες και πεδία κειμένου για το φιλτράρισμα των ετών:

Chart Generator

Chart Type:
Bar Chart

Greece Birth rate (births per 1.000 population)

Cyprus Birth rate (births per 1.000 population)

Add Pair

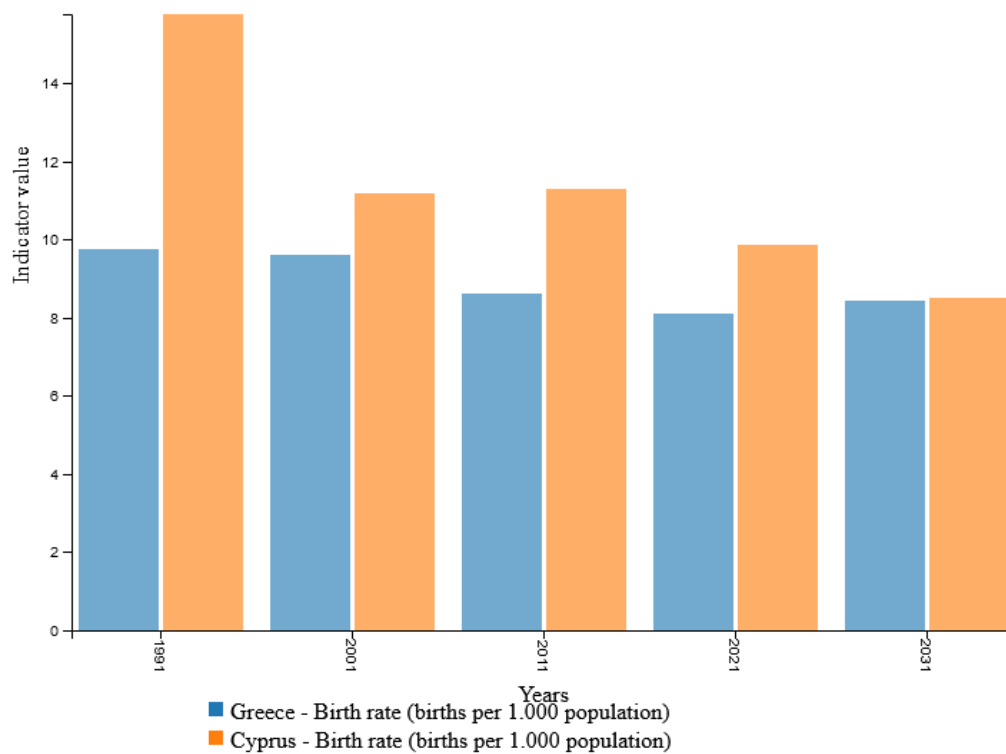
Filter:
Min year 2040

Aggregate:
10 years

Generate Chart

Το φιλτράρισμα των ετών μπορεί να γίνει είτε καθορίζοντας μόνο το ένα από τα δυο όρια, είτε καθορίζοντας και τα δυο όρια.

Το γράφημα που αντιστοιχεί στην συμπληρωμένη φόρμα:



Για την δημιουργία ενός scatter chart η επιλογή των χωρών και δεικτών θα πρέπει να γίνει κατά αυτόν τον τρόπο:

Chart Generator

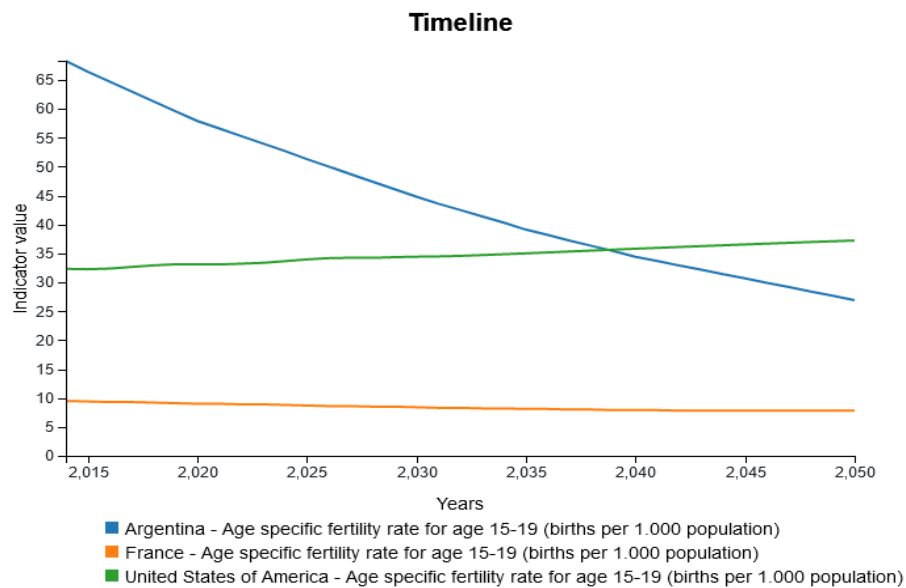
Chart Type:

Scatter Chart		
Greece	Female population all ages	x
Greece	Male population all ages	x
Cyprus	Female population all ages	x
Cyprus	Male population all ages	x
Add Pair		

δηλαδή θα πρέπει να επιλεγθούν ακριβώς δυο δείκτες οι οποίοι είναι κοινοί για κάθε χώρα.

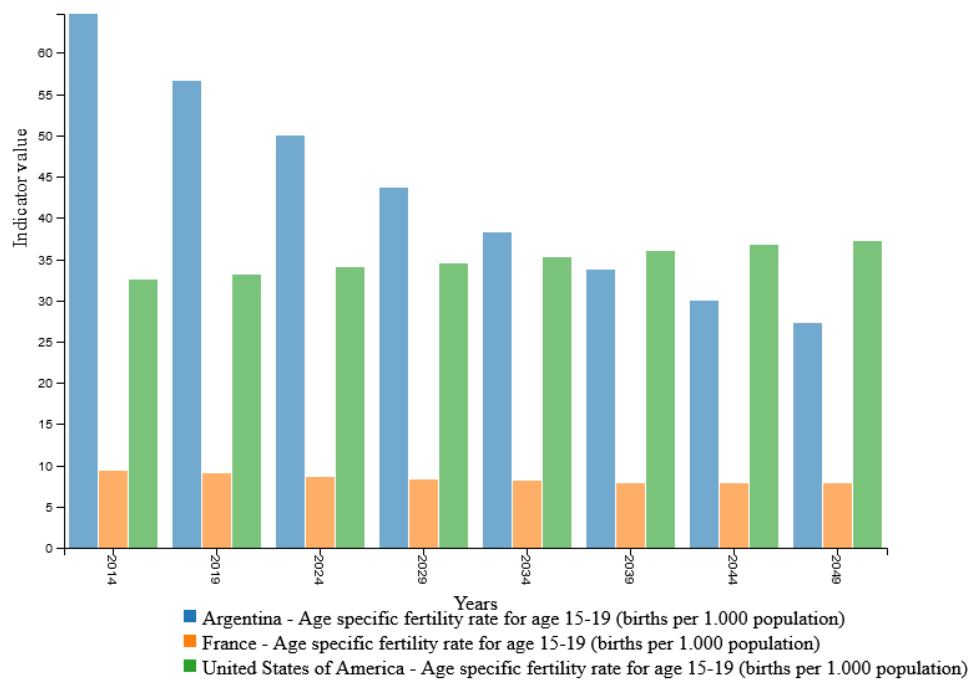
4 ΜΕΛΕΤΗ ΔΕΔΟΜΕΝΩΝ

a)



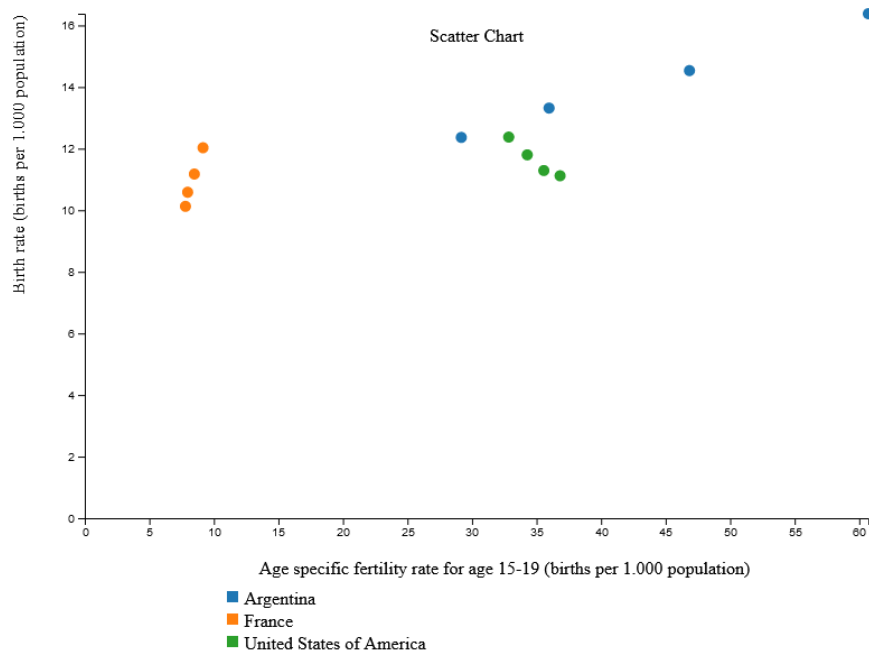
(aggregated by 5 years)

b)



(aggregated by 5 years)

c)

*((aggregated by 10 years))*

5 ΛΟΙΠΑ ΣΧΟΛΙΑ

-