

ΠΡΟΧΩΡΗΜΕΝΑ ΘΕΜΑΤΑ
ΤΕΧΝΟΛΟΓΙΑΣ ΚΑΙ ΕΦΑΡΜΟΓΩΝ
ΒΑΣΕΩΝ ΔΕΔΟΜΕΝΩΝ

ΠΡΟΓΡΑΜΜΑΤΙΣΤΙΚΗ ΕΡΓΑΣΙΑ ΓΙΑ
ΤΟ ΑΚΑΔΗΜΑΪΚΟ ΕΤΟΣ 2021-2022

ΟΜΑΔΑ HACK_THE_DB

ΧΡΗΣΤΟΣ ΧΡΙΣΤΙΔΗΣ, 4526

ΓΙΩΡΓΟΣ ΜΟΥΤΣΟΠΟΥΛΟΣ, 4115

ΤΕΛΙΚΗ ΑΝΑΦΟΡΑ

ΜΑΪΟΣ 2022

ΙΣΤΟΡΙΚΟ ΠΡΟΗΓΟΥΜΕΝΩΝ ΕΚΔΟΣΕΩΝ

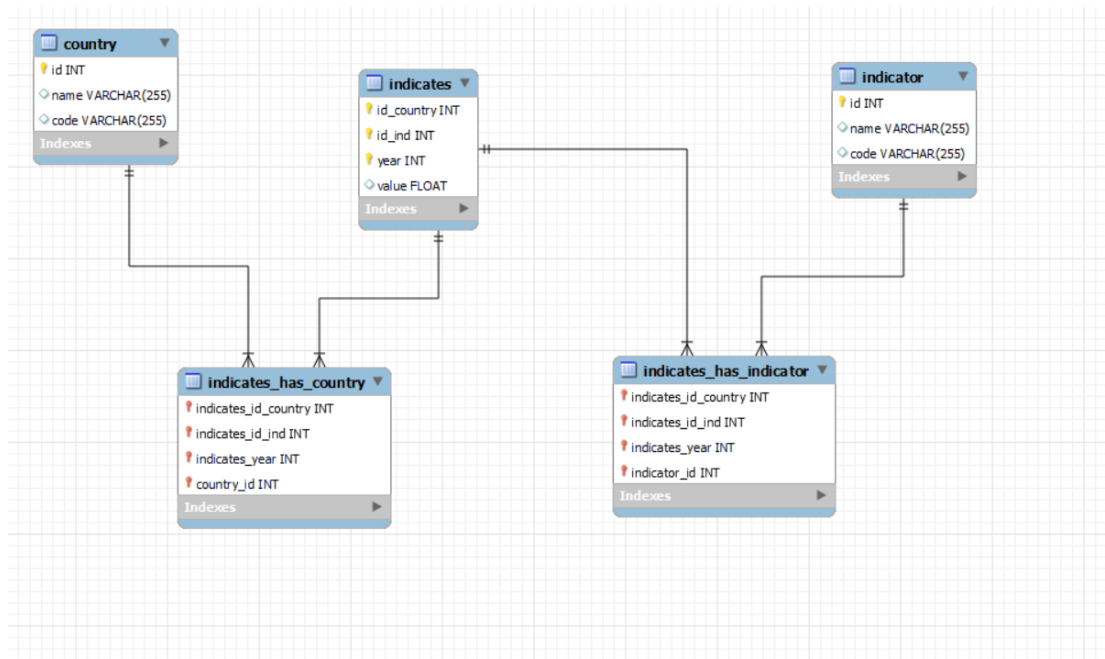
Ημερομηνία	Έκδοση	Περιγραφή	Συγγραφέας
2022/03/10	0.1	Download Data from Databank.	4526
2022/03/11	0.2	Data cleaning using pandas in notebook , make notebook python script.	4115
2022/03/12	0.3	Download MYSQL , insert data to database -Problem with MYSQL workbench.	4115
2022/03/13	0.3	Load Data from excel to database using LOAD DATA INSERT .	4526
2022/03/25	1.0	Create the simple user interface , create the connection between eclipse and MYSQL database .	4115
2022/04/02	1.1	Trying to create the scatterplot graph. Create dynamically constructed queries.	4526
2022/04/03	1.1	Experimenting with DAO design – Learning about the pattern.	4526
2022/04/04	1.2	Create the first timeline graph and changing again the user interface	4115
2022/04/20	1.3	Trying to create Bar chart – Problem with converting numerical variables to categorical.	4115
2022/05/03	1.5	Change the architecture of the app to be more object oriented .	4526
2022/05/11	1.6	Change in user interface and move some classes to other packages.	4115
2022/05/20	1.7	Create Template -design pattern	4526

We started downloading data and then cleaning it using python and pandas library , after that we load it in to the MYSQL database using MYSQL terminal. Then , we created our first version of application with some basic gui in order to test the query preprocessing and to start creating our first graph – timeline. After graphs , we decided to make some improvements in user interface and in application's architecture in order to make it much simpler for the software developer. Lastly , we created tests for code .

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

Key constraints:

- country has a non-composite primary key “id”.
- indicator has a non-composite primary key “id”.
- indicates table has a composite key (id_country, id_ind, year).

Referential integrity constraints:

- indicates table has a foreign key “id_country” derived from the primary key “id” of “country” table.
- indicates table has a foreign key “id_ind” derived from the primary key “id” of “indicator” table.

Structural constraints:

- Full participative, M-N relation: country-indicates.
- Full participative, M-N relation: indicator-indicates.

Section-break (continuous)

```
-- Table structure for table `country`
--

DROP TABLE IF EXISTS `country`;
/*!40101 SET @saved_cs_client      = @@character_set_client */;
/*!50503 SET character_set_client = utf8mb4 */;
CREATE TABLE `country` (
  `id` int NOT NULL,
  `name` varchar(255) DEFAULT NULL,
  `code` varchar(255) DEFAULT NULL,
  PRIMARY KEY (`id`)
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4_0900_ai_ci;
/*!40101 SET character_set_client = @saved_cs_client */;

--
-- Dumping data for table `country`
--

LOCK TABLES `country` WRITE;
/*!40000 ALTER TABLE `country` DISABLE KEYS */;
INSERT INTO `country` VALUES
(1,'Albania','ALB\r'), (2,'Andorra','AND\r'), (3,'Austria','AUT\r'), (4,
'Belgium','BEL\r'), (5,'Bulgaria','BGR\r'), (6,'Bosnia and
Herzegovina','BIH\r'), (7,'Cyprus','CYP\r'), (8,'Czech
Republic','CZE\r'), (9,'Germany','DEU\r'), (10,'Denmark','DNK\r'), (11,'
Estonia','EST\r'), (12,'Finland','FIN\r'), (13,'France','FRA\r'), (14,'F
aroe
Islands','FRO\r'), (15,'Georgia','GEO\r'), (16,'Gibraltar','GIB\r'), (17
,'Greece','GRC\r'), (18,'Greenland','GRL\r'), (19,'Croatia','HRV\r'), (2
0,'Hungary','HUN\r'), (21,'Isle of
Man','IMN\r'), (22,'Ireland','IRL\r'), (23,'Iceland','ISL\r'), (24,'Ital
y','ITA\r'), (25,'Kosovo','XKX\r');
/*!40000 ALTER TABLE `country` ENABLE KEYS */;
UNLOCK TABLES;

--
-- Table structure for table `indicates`
--

DROP TABLE IF EXISTS `indicates`;
/*!40101 SET @saved_cs_client      = @@character_set_client */;
/*!50503 SET character_set_client = utf8mb4 */;
CREATE TABLE `indicates` (
  `id_country` int NOT NULL,
  `id_ind` int NOT NULL,
  `year` int NOT NULL,
  `value` float DEFAULT NULL,
  PRIMARY KEY (`id_country`,`id_ind`,`year`),
```

```
    KEY `ind_id` (`id_ind`),
    CONSTRAINT `country_id` FOREIGN KEY (`id_country`) REFERENCES
`country` (`id`),
    CONSTRAINT `ind_id` FOREIGN KEY (`id_ind`) REFERENCES `indicator`
(`id`)
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4_0900_ai_ci;
/*!40101 SET character_set_client = @saved_cs_client */;

--
--
-- Dumping data for table `indicates`
--

LOCK TABLES `indicates` WRITE;
/*!40000 ALTER TABLE `indicates` DISABLE KEYS */;
INSERT INTO `indicates` VALUES
(1,1,1960,0),(1,1,1961,0),(1,1,1962,0),(1,1,1963,0),(1,1,1964,0),(1,1,
1965,0),(1,1,1966,0),(1,1,1967,0),(1,1,1968,0),(1,1,1969,0),(1,1,197
0,0),(1,1,1971,14.5788),(1,1,1972,0),(1,1,1973,0),(1,1,1974,0),(1,1,1
975,0),(1,1,1976,0),(1,1,1977,0),(1,1,1978,19.7044),(1,1,1979,0),(1,1
,1980,0),(1,1,1981,19.9456),(1,1,1982,0),(1,1,1983,22.5807),(1,1,1984
,22.6471),(1,1,1985,22.1704),(1,1,1986,23.6376),(1,1,1987,24.7846),(1
,1,1988,25.2308),...
(233,'Adjusted net enrollment rate, primary, male (% of primary
school age children)','SE.PRM.TENR.MA\r'),(234,'Persistence to grade
5, female (% of cohort)','SE.PRM.PRS5.FE.ZS\r'),(235,'School
enrollment, primary (% gross)','SE.PRM.ENRR\r'),(236,'School
enrollment, preprimary, female (%
gross)','SE.PRE.ENRR.FE\r'),(237,'GNI per capita, PPP (current
international $)','NY.GNP.PCAP.PP.CD\r'),(238,'Gross domestic savings
(% of GDP)','NY.GDS.TOTL.ZS\r'),(239,'GDP growth (annual
%)','NY.GDP.MKTP.KD.ZG\r'),(240,'Inflation, GDP deflator (annual
%)','NY.GDP.DEFL.KD.ZG\r'),(241,'Adjusted savings: natural resources
depletion (% of GNI)','NY.ADJ.DRES.GN.ZS\r');
/*!40000 ALTER TABLE `indicator` ENABLE KEYS */;
UNLOCK TABLES;

-- -----
--
-- Table structure for table `test`
--

DROP TABLE IF EXISTS `test`;
/*!40101 SET @saved_cs_client      = @@character_set_client */;
/*!50503 SET character_set_client = utf8mb4 */;
CREATE TABLE `test` (
  `t1` int NOT NULL,
  `t2` int NOT NULL,
  `t3` int DEFAULT NULL,
  PRIMARY KEY (`t1`,`t2`)
) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4_0900_ai_ci;
/*!40101 SET character_set_client = @saved_cs_client */;

--
-- Dumping data for table `test`
--

LOCK TABLES `test` WRITE;
/*!40000 ALTER TABLE `test` DISABLE KEYS */;
/*!40000 ALTER TABLE `test` ENABLE KEYS */;
```

```
UNLOCK TABLES;  
SET @@SESSION.SQL_LOG_BIN = @MYSQLDUMP_TEMP_LOG_BIN;  
/*!40103 SET TIME_ZONE=@OLD_TIME_ZONE */;  
  
/*!40101 SET SQL_MODE=@OLD_SQL_MODE */;  
/*!40014 SET FOREIGN_KEY_CHECKS=@OLD_FOREIGN_KEY_CHECKS */;  
/*!40014 SET UNIQUE_CHECKS=@OLD_UNIQUE_CHECKS */;  
/*!40101 SET CHARACTER_SET_CLIENT=@OLD_CHARACTER_SET_CLIENT */;  
/*!40101 SET CHARACTER_SET_RESULTS=@OLD_CHARACTER_SET_RESULTS */;  
/*!40101 SET COLLATION_CONNECTION=@OLD_COLLATION_CONNECTION */;  
/*!40111 SET SQL_NOTES=@OLD_SQL_NOTES */;
```

1.2 ΣΧΕΣΙΑΚΟ ΣΧΗΜΑ ΣΕ ΦΥΣΙΚΟ ΕΠΙΠΕΔΟ

Όταν θα έχετε στήσει και ρυθμίσει τη βάση δεδομένων σας, εδώ καταγράφονται και οι ρυθμίσεις σε φυσικό επίπεδο. Ενδεικτικά:

1.2.1 ΡΥΘΜΙΣΗ ΤΩΝ ΠΑΡΑΜΕΤΡΩΝ ΤΟΥ DBMS

- Storage engine = InnoDB
- To check how much memory our database uses, we used the following query:

```
SELECT mye030,
```

```
-> ROUND(SUM(data_length + index_length) / 1024 / 1024, 2) AS "Size (MB)
```

```
-> FROM information_schema.TABLES
```

```
-> GROUP BY table_schema;
```

- And the return value was 29.16 (MB).
- We didn't proceed to allocate further memory for the database, as we thought that the already allocated amount was enough for the size of the dataset.

1.2.2 ΡΥΘΜΙΣΗ ΤΟΥ ΦΥΣΙΚΟΥ ΣΧΗΜΑΤΟΣ ΤΗΣ ΒΑΣΗΣ ΔΕΔΟΜΕΝΩΝ

--> Για την ευρετηριοποίηση των tables ορίσαμε id ως primary key στους πίνακες country και indicator , ενώ για τον πίνακα indicates το primary key βγαίνει από (id_country, id_ind, year).

--> Ένας ορισμός όψεως που θα μπορούσαμε να χρησιμοποιήσουμε:

```
CREATE VIEW IND AS SELECT ID , VALUE FROM INDICATES WHERE
ID_COUNTRY="FRANCE"
```

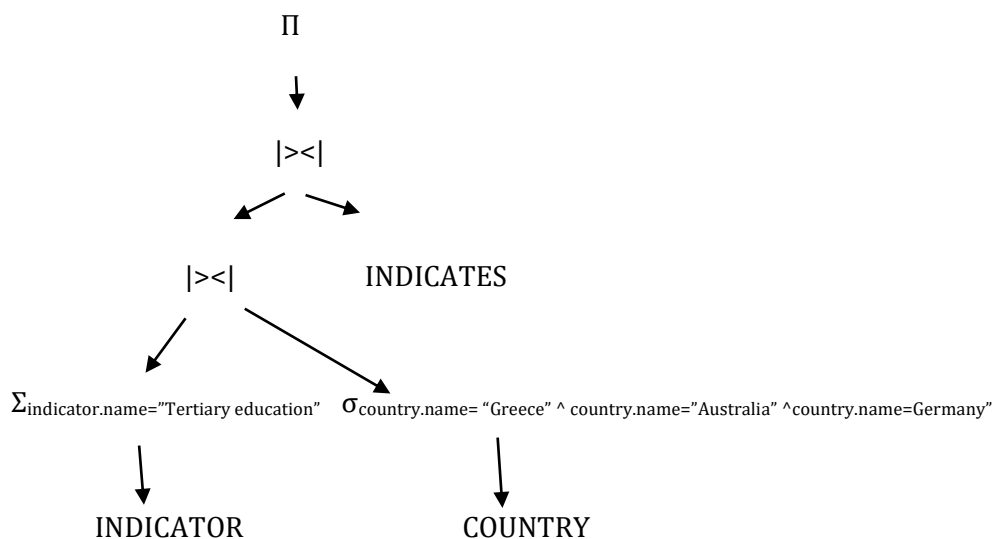
--> Παράδειγμα πλάνου που καθυστερεί :

Επιλέξτε όλα τα values από των πίνακα indicates όπου η χώρα είναι η Ελλάδα, η Αυστραλία και η Γερμανία και η μετρική είναι το "Tertiary education":

```
SELECT INDICATES.value
```

```
FROM INDICATOR, COUNTRY, INDICATES
```

```
WHERE INDICATOR.id = INDICATES.id AND INDICATOR.id=COUNTRY.id AND
INDICATOR.name = "Tertiary education" AND COUNTRY.name ="Greece" AND
COUNTRY.name="Australia" AND COUNTRY.name="Germany"
```

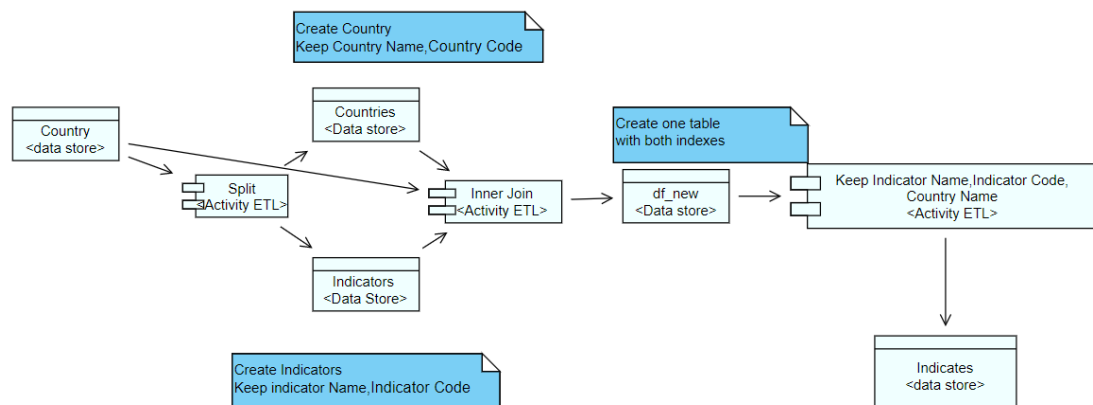


1.2.3 ΡΥΘΜΙΣΕΙΣ ΑΣΦΑΛΕΙΑΣ

Η δομή των ερωτημάτων που θα περαστούν στην βάση είναι στατική και ο μόνος τρόπος χειραγώγησης τους είναι η εισαγωγή επιλογών από το αρχικό παράθυρο (βλ. εικόνα 1). Επιπλέον, η εμφάνιση των αποτελεσμάτων γίνεται αμιγώς σε οπτικό επίπεδο, συνεπώς δεν προτρέπει κίνδυνος εκμείευσης των περιεχομένων της βάσης. Η είσοδος στη βάση γίνεται με ημι-αυτόματο τρόπο μέσω της μεθόδου `connectWithDB` (κλάση-> `Connector`, package -> `dao`), όπου είναι αποθηκευμένα τα `credentials` εισόδου στην `mysql` βάση.

2 ΑΡΧΙΤΕΚΤΟΝΙΚΗ ΛΟΓΙΣΜΙΚΟΥ

2.1 ΑΡΧΙΤΕΚΤΟΝΙΚΗ ΚΑΙ ΔΟΜΗ ETL



Σχήμα 2.1

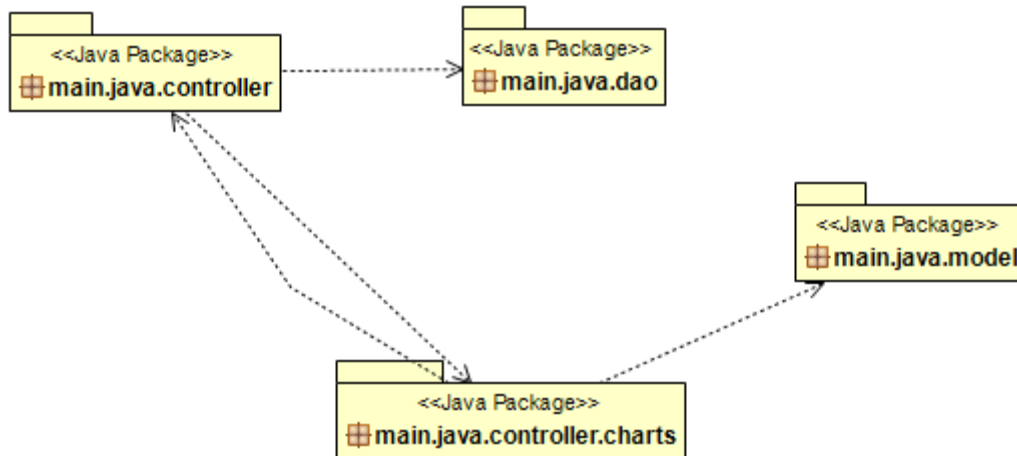
Ξεκινήσαμε κατεβάζοντας τον πίνακα με όλα τα δεδομένα από το databank, ύστερα τον χωρίσαμε σε 2 πίνακες: τον Country με τα columns 'Country name', 'Country Code', 'index Country' και τον πίνακα Indicators με τα columns 'Indicator Name', 'Indicator Code', 'index_indicat'.

Μετά αφού είχαμε διαφορετικά indexes για τους δυο πίνακες, τους ενώσαμε (inner join) και με τον αρχικό πίνακα όπου είχε τα columns 'year', 'value' κλπ. Ύστερα από τον μεγάλο πίνακα που προέκυψε κρατήσαμε τα columns: 'Index_country', 'index_indicat', 'year', 'value'. Όπου μετά τα ονομάσαμε 'id_country', 'id_ind', 'year', 'value' και τα φορτώσαμε στο MYSQL database

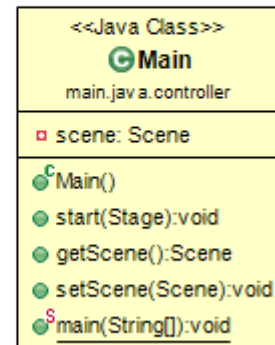
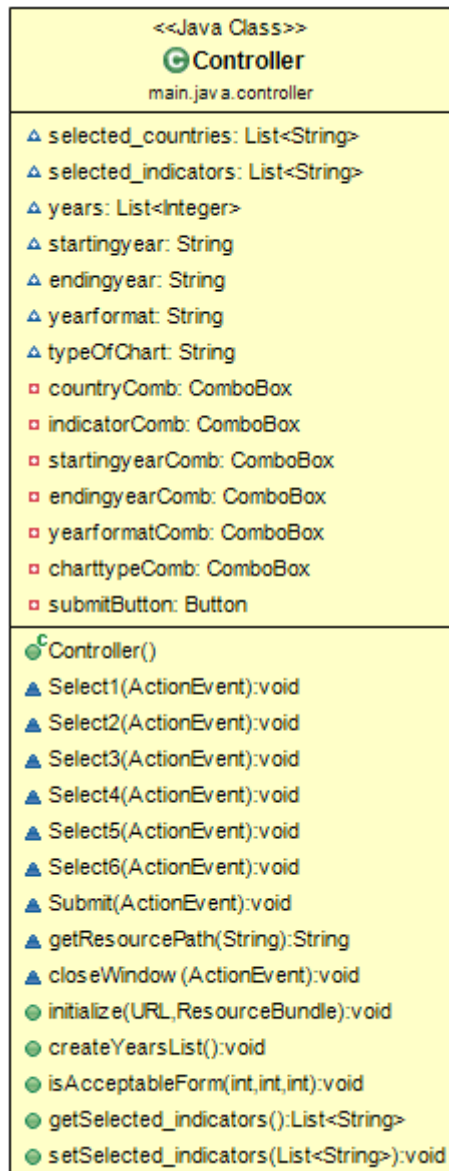
2.2 ΔΙΑΓΡΑΜΜΑ(ΤΑ) ΚΛΑΣΕΩΝ ΚΕΝΤΡΙΚΗΣ ΕΦΑΡΜΟΓΗΣ

Αν η ανάπτυξη γίνει αντικειμενοστρεφώς, εδώ μπαίνουν τα διαγράμματα κλάσεων + ο σχολιασμός της κεντρικής εφαρμογής. Αλλιώς μπαίνουν διαγράμματα που διευκολύνουν την κατανόηση της εσωτερικής αρχιτεκτονικής του λογισμικού (π.χ., component/ deployment diagrams / ...)

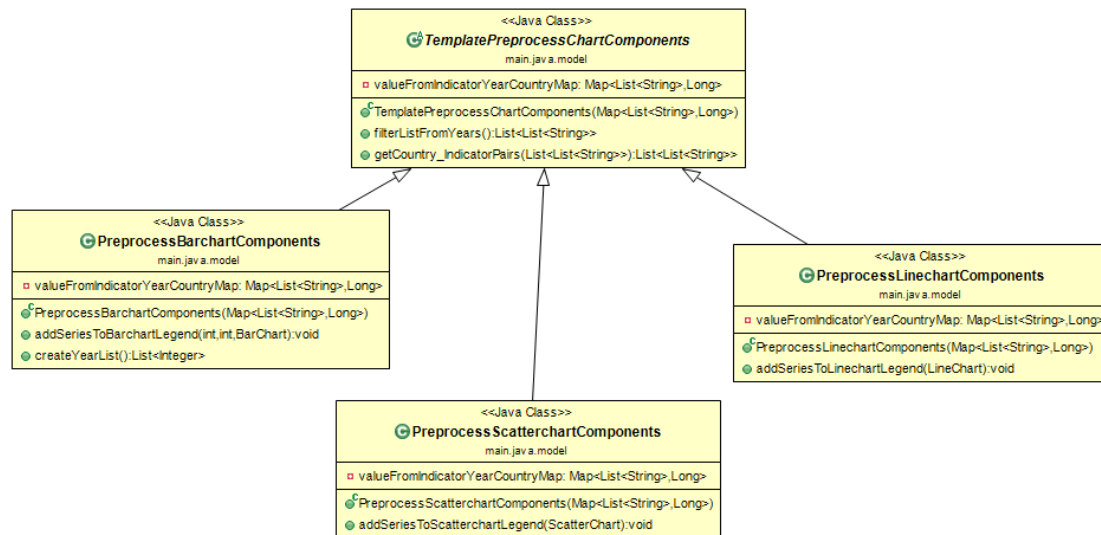
main.java packages UML:



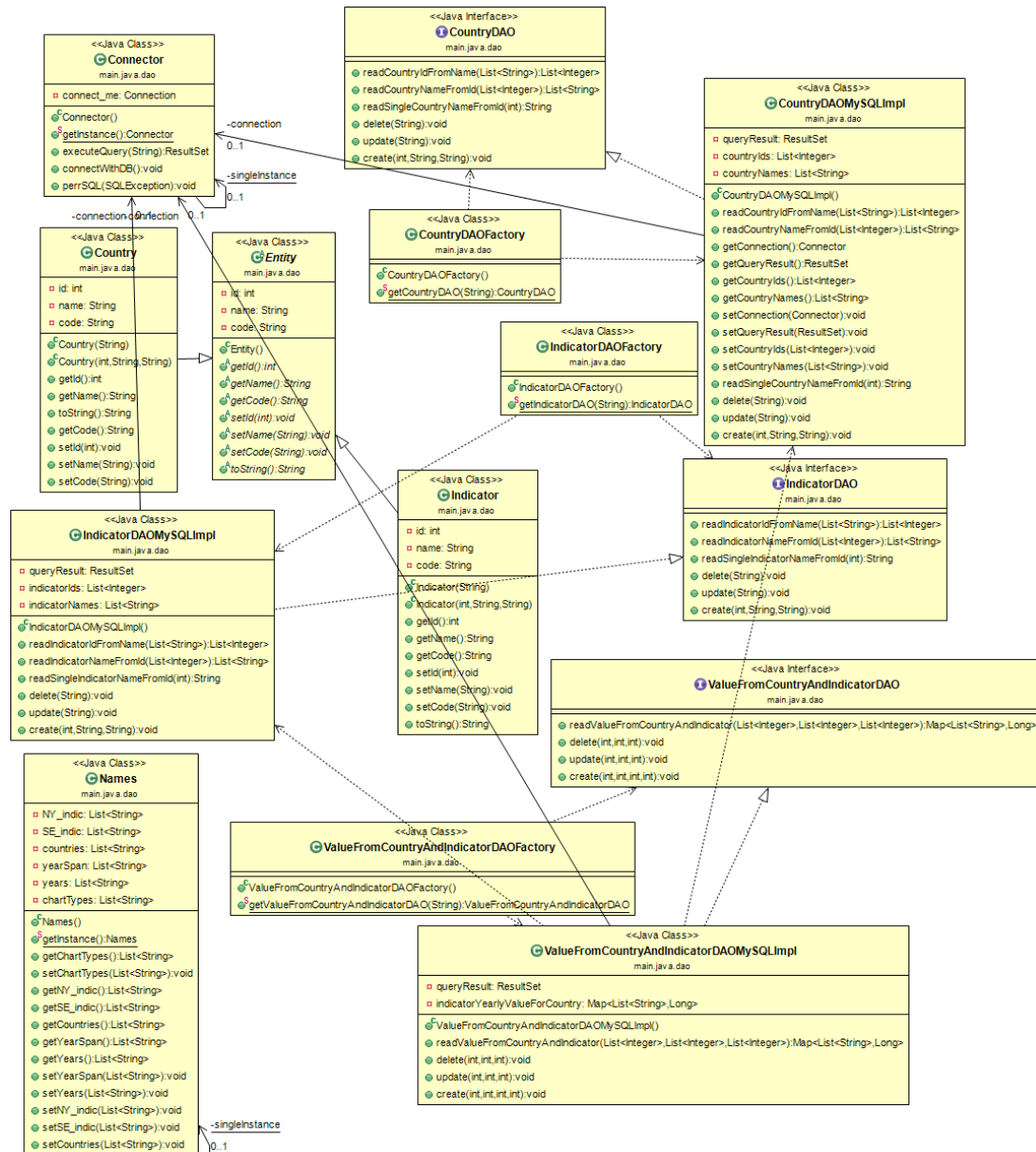
controller UML:



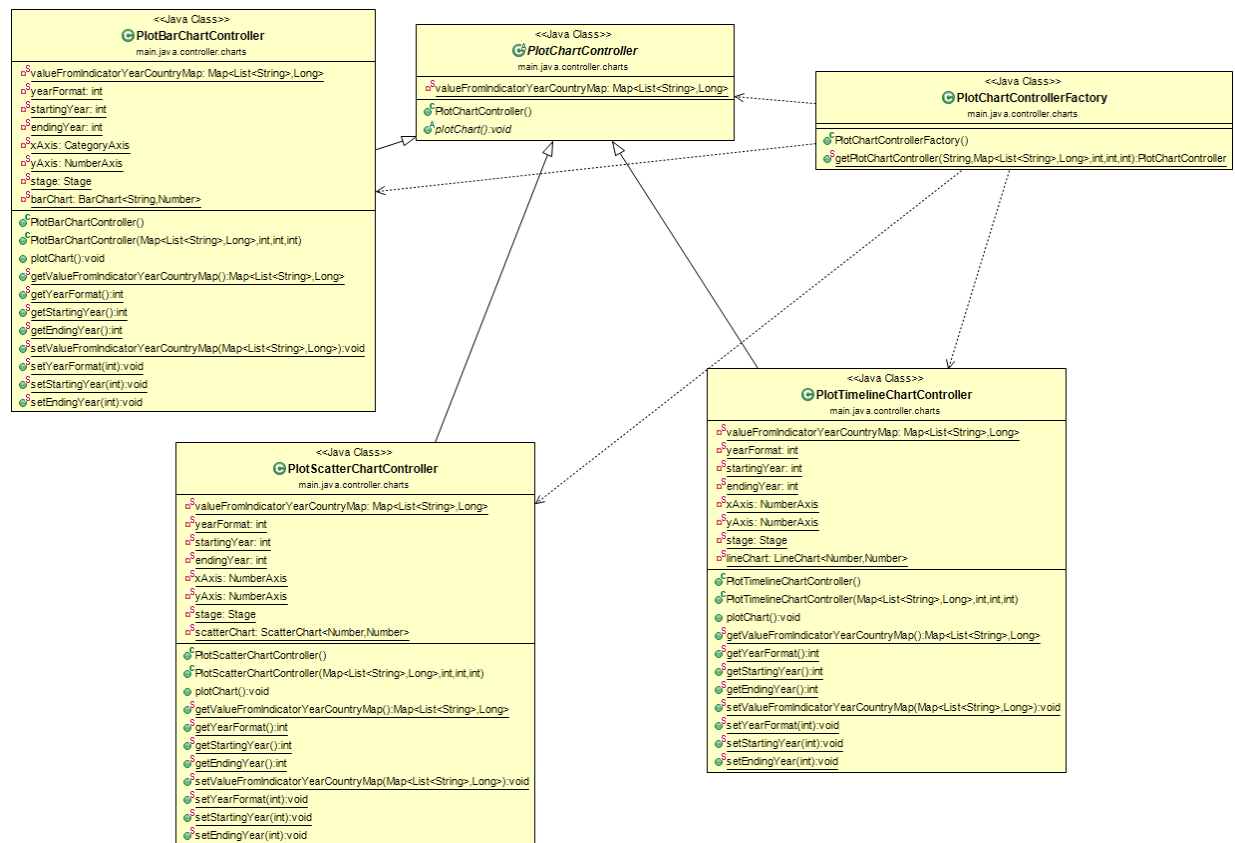
model UML:



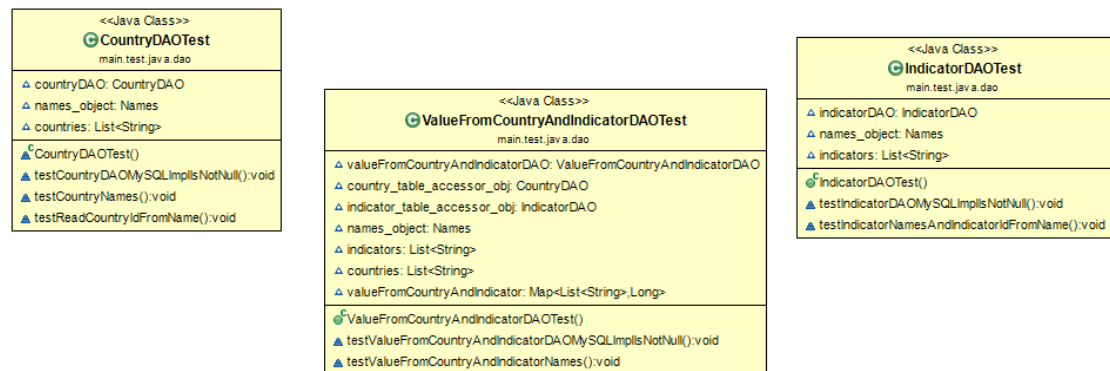
dao UML:



controller.charts UML:

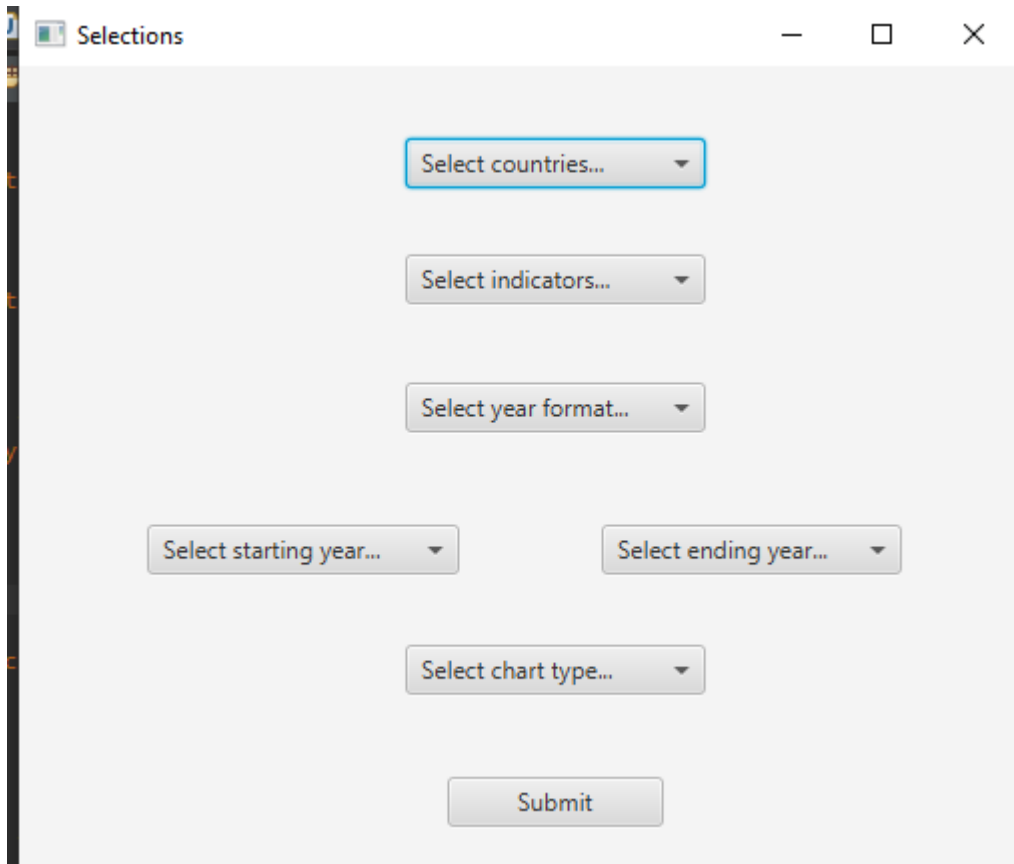


test.java.dao UML:



3 ΥΠΟΔΕΙΓΜΑΤΑ ΕΡΩΤΗΣΕΩΝ ΚΑΙ ΑΠΑΝΤΗΣΕΩΝ

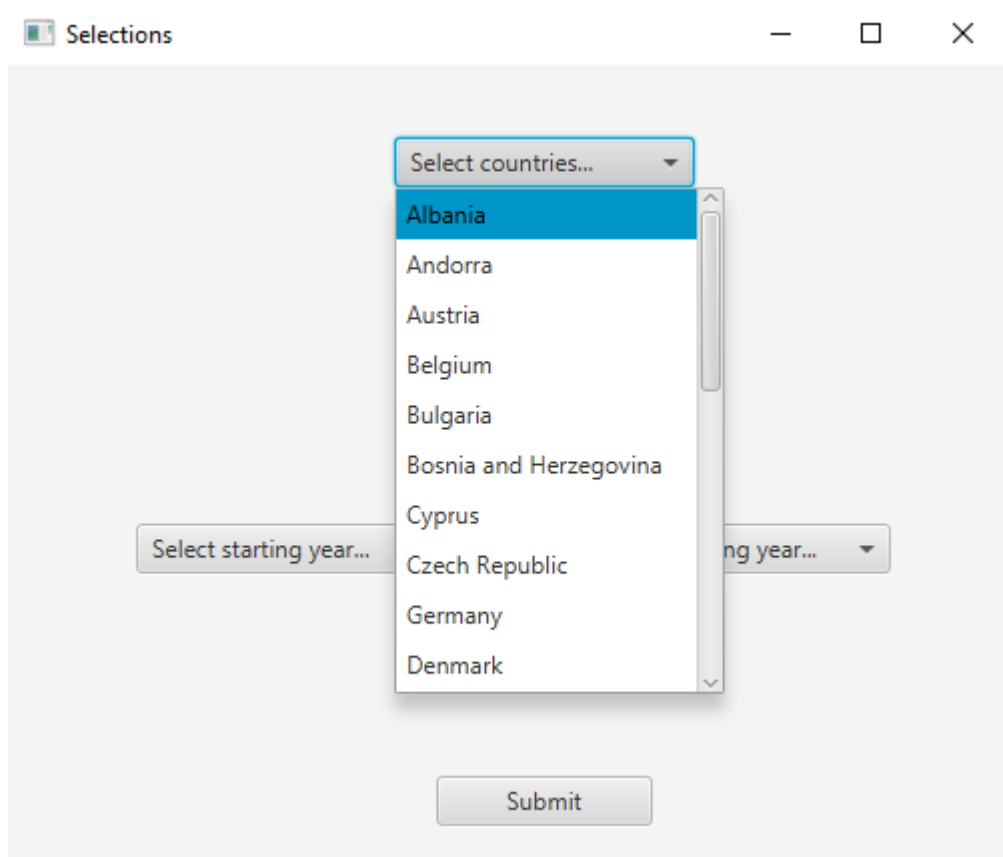
Με την εκκίνηση της εφαρμογής, στον χρήστη εμφανίζεται το εξής παράθυρο:



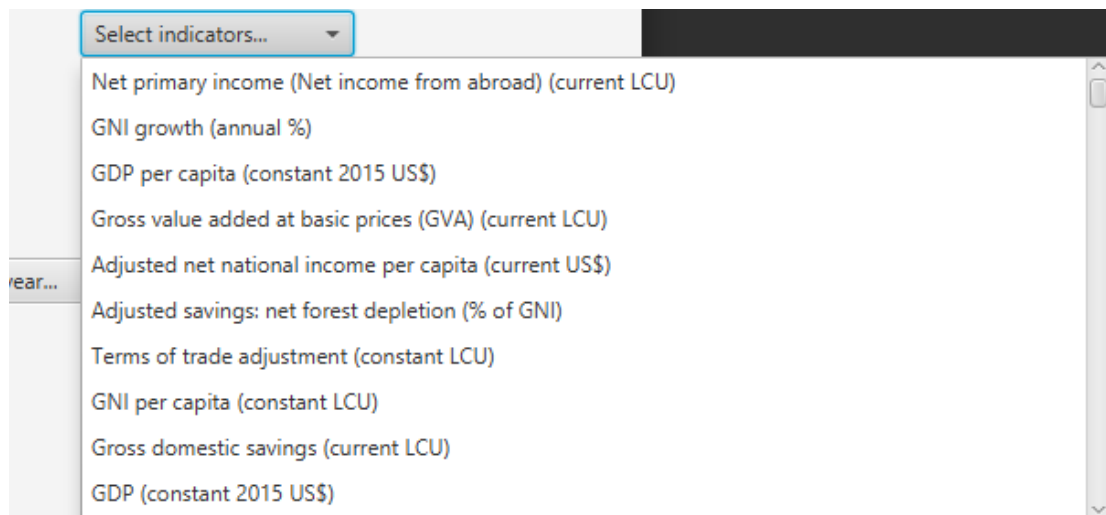
The image shows a web browser window with a dialog box titled "Selections". The dialog box has a light gray background and a title bar with a close button (X). Inside the dialog, there are five dropdown menus arranged vertically: "Select countries...", "Select indicators...", "Select year format...", "Select starting year...", and "Select ending year...". Below these, there is a "Select chart type..." dropdown. At the bottom of the dialog is a "Submit" button.

Εικόνα 1: αρχικό παράθυρο επιλογών χρήστη.

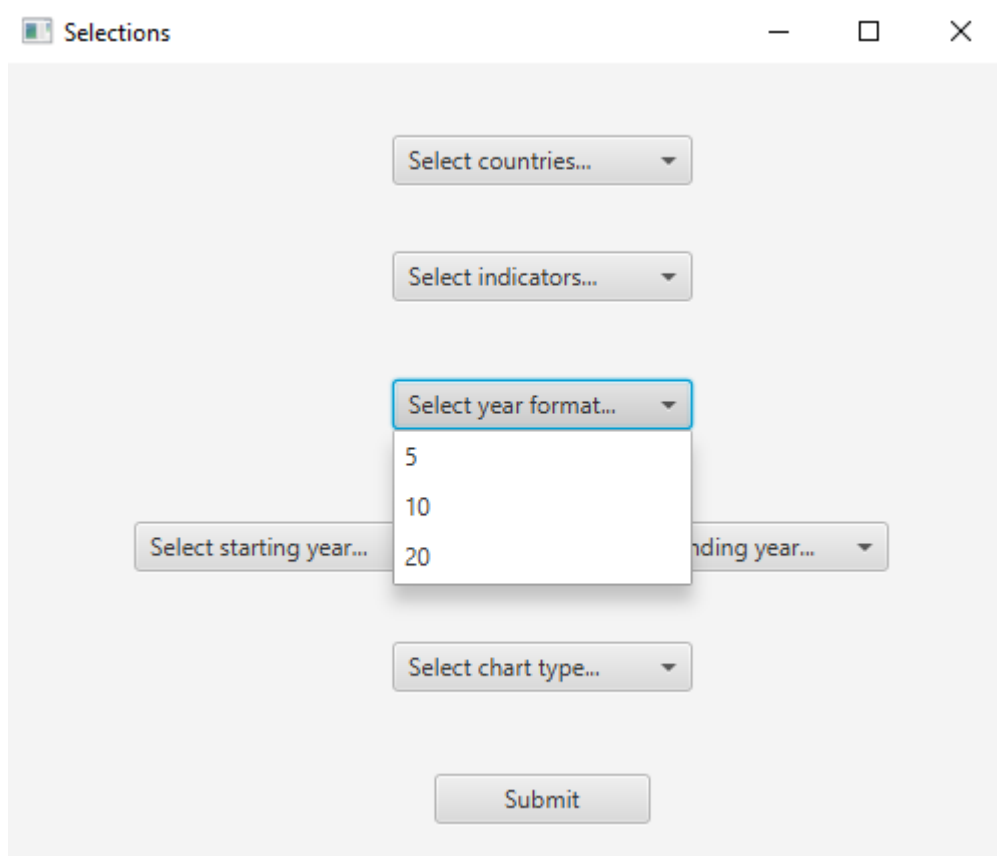
Παροτρύνοντας τον χρήστη να εισάγει τις επιθυμητές επιλογές για οπτικοποίηση σε μία εκ των 3 επιλογών γραφήματος.



Εικόνα 2: ο χρήστης δύναται να επιλέξει ανάμεσα σε 25 χώρες.

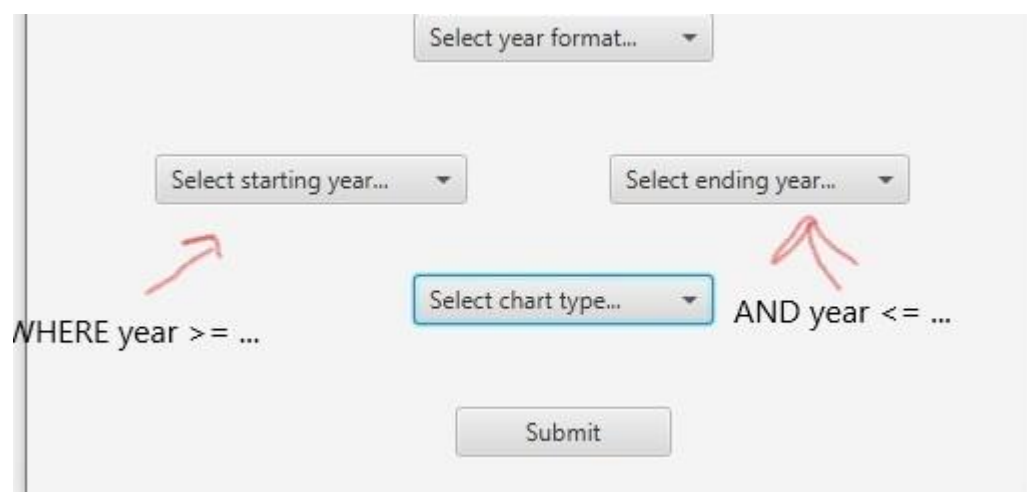


Εικόνα 3: ο χρήστης μπορεί να επιλέξει ανάμεσα σε 241 δείκτες.



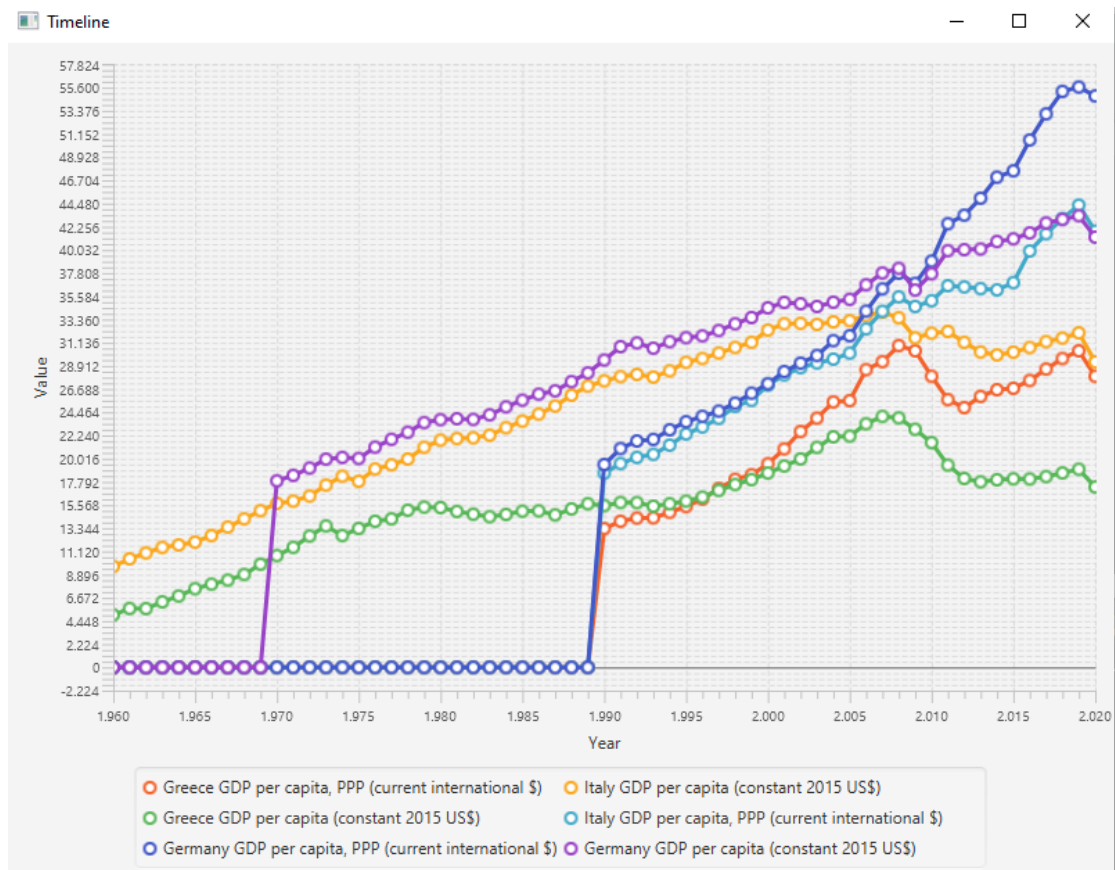
The screenshot shows a window titled "Selections" with a light gray background. It contains several dropdown menus and a "Submit" button. The dropdown menus are labeled "Select countries...", "Select indicators...", "Select year format...", "Select starting year...", and "Select ending year...". The "Select year format..." dropdown is currently open, showing a list of options: "5", "10", and "20". The "Submit" button is located at the bottom center of the window.

Εικόνα 4: επιλογή time aggregation.



This screenshot shows a portion of the "Selections" dialog box, focusing on the "Select chart type..." dropdown menu, which is open and displays the options "5", "10", and "20". Red arrows point from the "Select starting year..." and "Select ending year..." dropdowns to the text "WHERE year >= ..." and "AND year <= ..." respectively, indicating how the selected years are used in the generated SQL query. The "Submit" button is also visible at the bottom.

Εικόνα 5: επιλογή αρχικής και καταληκτικής χρονίας οπτικοποίησης.



Εικόνα 6: παράδειγμα οπτικοποίησης σε Timeline, με time aggregation = 5.

4 ΛΟΙΠΑ ΣΧΟΛΙΑ

TODO LIST:

- Encrypted login window
- Update graph button