ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ ΤΜΗΜΑ ΜΗΧ. Η/Υ & ΠΛΗΡΟΦΟΡΙΚΗΣ

ΠΡΟΧΩΡΗΜΕΝΑ ΘΕΜΑΤΑ ΤΕΧΝΟΛΟΓΙΑΣ ΚΑΙ ΕΦΑΡΜΟΓΩΝ ΒΑΣΕΩΝ ΔΕΔΟΜΕΝΩΝ

ΠΡΟΓΡΑΜΜΑΤΙΣΤΙΚΗ ΕΡΓΑΣΙΑ ΓΙΑ ΤΟ ΑΚΑΔΗΜΑΪΚΟ ΕΤΟΣ 2019-2020

ΟΜΑΔΑ

ΑΥΔΙΚΟΣ ΓΕΩΡΓΙΟΣ ΑΜ: 3184

BINH $\Sigma \Omega$ THPIO Σ AM: 2946

ΜΗΤΡΟΓΙΑΝΝΗΣ ΧΡΥΣΑΝΘΟΣ ΑΜ: 3029

ΤΕΛΙΚΗ ΑΝΑΦΟΡΑ

ΙΣΤΟΡΙΚΟ ΠΡΟΗΓΟΥΜΕΝΩΝ ΕΚΔΟΣΕΩΝ

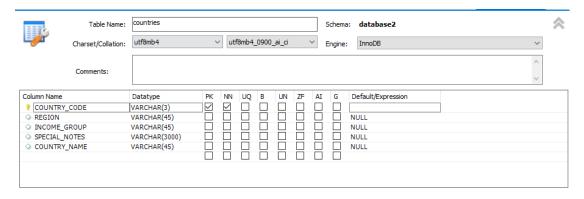
Ημερομηνία	Έκδοση	Περιγραφή	Συγγραφέας
2020/06/02	Τελική έκδοση	Η συγκεκριμένη έκδοση περιέχει ολοκληρωμένες όλες τις φάσεις υλοποίησης του project.	Αυδίκος Γ. Βίνης Σ. Μητρογιάννης Χ.

1 ΒΑΣΗ ΔΕΔΟΜΕΝΩΝ

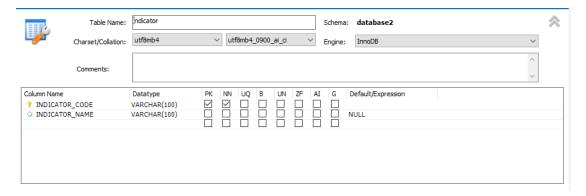
Στο συγκεκριμένο κεφάλαιο περιγράφονται τα σχήματα της βάσης δεδομένων που χρησιμοποιούνται στο project.

1.1 ΣΧΕΣΙΑΚΟ ΣΧΗΜΑ ΣΕ ΛΟΓΙΚΟ ΕΠΙΠΕΔΟ

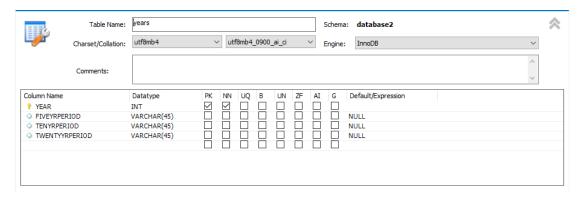
Σχήμα COUNTRIES της βάσης δεδομένων.



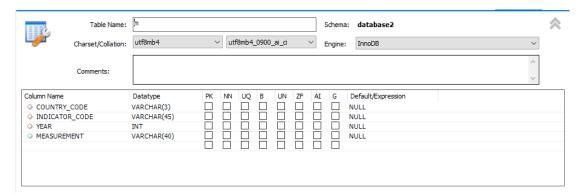
Σχήμα INDICATOR της βάσης δεδομένων.



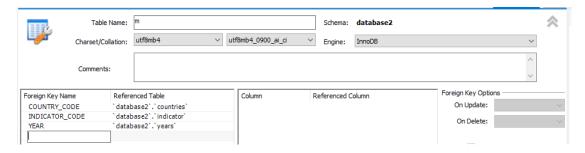
Σχήμα YEARS της βάσης δεδομένων.



Σχήμα Μ της βάσης δεδομένων.

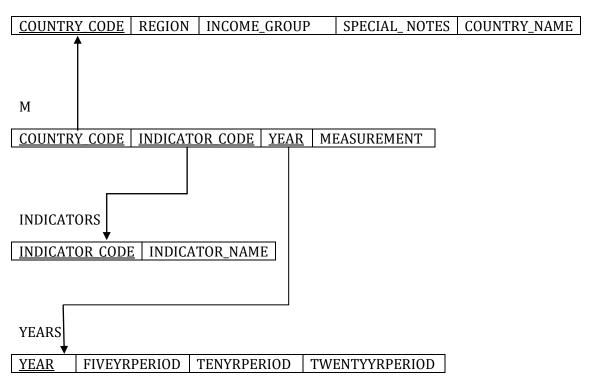


Εδώ φαίνονται τα foreign keys του σχήματος Μ.



Παρακάτω φαίνεται ο σχεδιασμός του σχεσιακού σχήματος της βάσης δεδομένων:

COUNTRIES



1.2 ΣΧΕΣΙΑΚΟ ΣΧΗΜΑ ΣΕ ΦΥΣΙΚΟ ΕΠΙΠΕΔΟ

Σε αυτό το κεφάλαιο θα αναλύσουμε κάποιες ρυθμίσεις που αφορούν τις παραμέτρους του DBMS καθώς και τις ρυθμίσεις του φυσικού σχήματος ης βάσης δεδομένων.

1.2.1 ΡΥΘΜΙΣΗ ΤΩΝ ΠΑΡΑΜΕΤΡΩΝ ΤΟΥ DBMS

Με την εντολή SET GLOBAL innodb_buffer_pool_size = 3435973837, δίνουμε στην mySQL το 80% της κύριας μνήμης του υπολογιστή. Χρησιμοποιήσαμε InnoDB τύπο αποθήκευσης, όπως άλλωστε φαίνεται και στα ανωτέρω screenshot. Όλες τις υπόλοιπες παραμέτρους τις αφήσαμε όπως ήταν εξ' αρχής.

1.2.2 ΡΥΘΜΙΣΗ ΤΟΥ ΦΥΣΙΚΟΥ ΣΧΗΜΑΤΟΣ ΤΗΣ ΒΑΣΗΣ ΔΕΔΟΜΕΝΩΝ

Στο project μας έχουμε υλοποιήσει συνολικά τρία σχήματα όψεων για να αποφεύγουμε τις εμφωλευμένες ερωτήσεις, καθώς και για λόγους απόδοσης.

Το πρώτο σχήμα αφορά την τεχνική ScatterPlot ενώ τα άλλα δύο αφορούν την τεχνική BarPlot. Επειδή τα δεδομένα των όψεων προσαρμόζονται ανάλογα με την ερώτηση που θέτει ο χρήστης, δημιουργούνται δυναμικά και στη συνέχεια, καταστρέφονται δυναμικά. Πιο συγκεκριμένα, για την τεχνική ScatterPlot δημιουργήσαμε μία όψη, η οποία περιέχει τρεις στήλες. Η πρώτη στήλη περιέχει το πεδίο YEAR(πχ. TENYEAR, αναλόγως τι έχει επιλέξει ο χρήστης) και οι επόμενες δύο στήλες περιέχουν, η μεν πρώτη τις μετρήσεις εναλλάξ ξεκινώντας για την πρώτη μετρική που έδωσε ο χρήστης και η μεν δεύτερη, εναλλάξ πάλι, ξεκινώντας από τη δεύτερη μετρική που έδωσε ο χρήστης. Για το λόγο αυτό, η ερώτηση που πραγματοποιήσαμε πάνω στην όψη απαλείφει τη μία μετρική από τη μία στήλη και την άλλη μετρική από την άλλη στήλη. Σαν αποτέλεσμα, προκύπτει ένας πίνακας, όπου στην πρώτη στήλη έχουμε την χρονολογία, στην δεύτερη στήλη τις μετρήσεις για την πρώτη μετρική και στην τελευταία στήλη τις μετρήσεις για την δεύτερη μετρική. Όσον αφορά τις άλλες δύο όψεις, έχουμε να αναφέρουμε τα εξής:

Στη συγκεκριμένη τεχνική αντιμετωπίσαμε δύο προβλήματα. Το πρώτο πρόβλημα αφορά το γεγονός ότι υπολογιζόταν και τα κενά κελιά στους μέσους όρους. Για παράδειγμα, αν ο χρήστης έδινε χρονολογία ανά πενταετίες, μπορεί μια πενταετία να είχε μόνο τρεις μετρήσεις και οι άλλες δύο να ήταν κενές. Σε αυτή την περίπτωση, η ερώτηση υπολόγιζε το μέσο όρο ως άθροισμα των μετρήσεων δια πέντε, ενώ εμείς θέλαμε το άθροισμα δια τρία. Το δεύτερο πρόβλημα είναι ότι όσα group χρονολογιών δεν είχαν τιμές, η ερώτηση δεν τις επέστρεφε καθόλου ενώ εμείς, για λόγους οπτικοποίησης, θέλαμε να επιστρέφεται το μηδέν.

Για να λύσουμε τα δύο παραπάνω προβλήματα κατασκευάσαμε από μία όψη, αντίστοιχα. Στην ερώτηση της εφαρμογής, κάναμε μία αριστερή συνένωση μεταξύ των δύο όψεων, έτσι ώστε να κρατήσουμε τους σωστούς μέσους όρους για τα group χρονολογιών που είχαν κενές γραμμές, όπως αναφέρθηκε και παραπάνω. Όσα group είχαν όλες τις μετρήσεις τους κενές λάμβαναν την τιμή μηδέν. Έτσι, λύσαμε τα δύο παραπάνω προβλήματα.

2 ΑΡΧΙΤΕΚΤΟΝΙΚΗ ΛΟΓΙΣΜΙΚΟΥ

Αρχικά, θα κάνουμε μία περιγραφή προεπεξεργασίας των δεδομένων μας, ενώ στην συνέχεια, θα παρουσιάσουμε τα διαγράμματα πακέτων και κλάσεων.

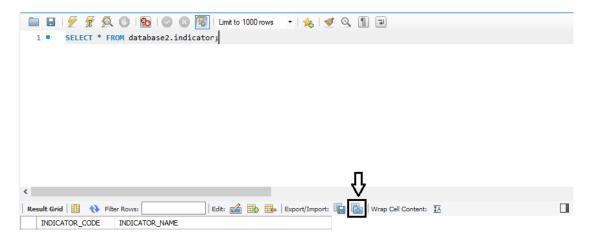
2.1 APXITEKTONIKH KAI ΔOMH ETL

Για την προεπεξεργασία των δεδομένων μας, δημιουργήσαμε ένα script¹ σε μορφή .ipynb και κώδικα python. Αρχικά, για το αρχείο data_per_country.csv μεταφέραμε χειρωνακτικά τα δεδομένα για τις χώρες και τις μετρικές. Πιο συγκεκριμένα, μεταφέραμε δέκα εγγραφές για κάθε χώρα, όπου η κάθε μία αντιστοιχίζεται σε κάθε μία μετρική. Τα υπόλοιπα τρία αρχεία τα δημιουργήσαμε μέσω του script που υλοποιήσαμε.

Αναλυτικότερα, για το αρχείο YEAR δημιουργήσαμε πίνακες με τις χρονολογίες, τις οποίες στη συνέχεια, τις μετατρέψαμε σε dataframe. Μέσα από αυτό το dataframe και με τις εντολές που μας παρέχουν οι βιβλιοθήκες της python, το κάναμε export σε ένα .csv αρχείο. Έπειτα, προχωρήσαμε με τη δημιουργία του αρχείου INDICATOR.csv. Το αρχείο αυτό περιέχει το indicator code για κάθε μία από τις δέκα μετρικές που χρησιμοποιήσαμε. Για να το δημιουργήσαμε διαβάσαμε τα αντίστοιχα στοιχεία από το αρχείο data_per_country.csv και ακολουθήσαμε την ίδια διαδικασία για τη δημιουργία του csv αρχείου. Τέλος, για το αρχείο M.csv ακολουθήσαμε παρόμοια διαδικασία με παραπάνω.

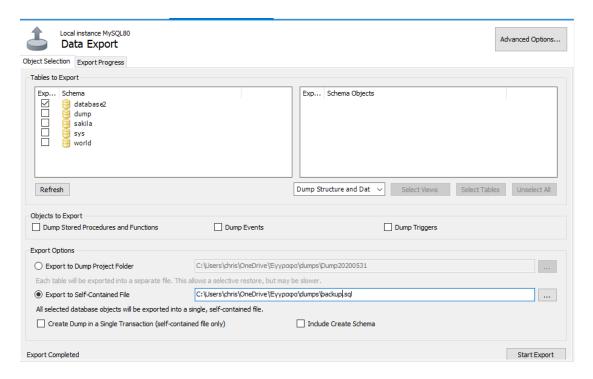
Συμπερασματικά, τα αρχεία μας ήταν πλέον έτοιμα να φορτωθούν στη βάση δεδομένων της mySQL.

Για να εισαχθούν τα δεδομένα στη βάση μας, αρχικά δημιουργήσαμε τέσσερα σχήματα, όπως αναφέρθηκαν στην ενότητα 1.1. Έπειτα, για κάθε ένα από αυτά, φορτώσαμε τα δεδομένα του κάθε αρχείου στον αντίστοιχο πίνακα. Συγκεκριμένα, για κάθε σχήμα πατήσαμε την επιλογή που φαίνεται στην παρακάτω εικόνα, επιλέξαμε το κατάλληλο αρχείο κάνοντας browse στο folder system και ακολουθήσαμε τα βήματα που προβάλλονται στο παράθυρο της εφαρμογής.

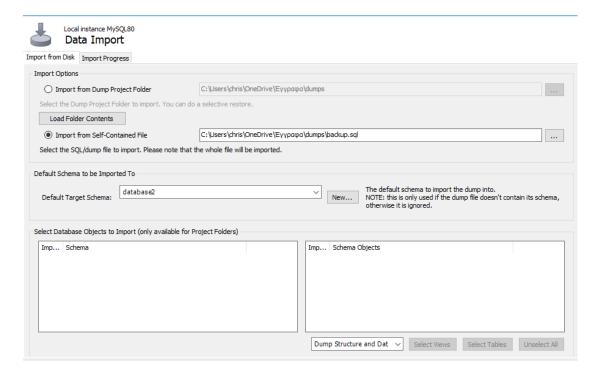


Μέσω του workbench της mySQL, και συγκεκριμένα στην επιλογή Server>Data export εμφανίζεται μία καρτέλα που περιέχει πληροφορίες για την φόρτωση αρχείων, όπως φαίνεται στην παρακάτω εικόνα.

¹ Επισυνάπτεται το αρχείο **preprocessing_script.ipynb**.



Επιλέξαμε τη βάση δεδομένων, το choiceBox "Export to Self-Contained File" και την επιλογή "Dump Structure and Data", προκειμένου να πάρουμε **backup** της δομής και των δεδομένων των σχημάτων. Σε περίπτωση που χαθεί η βάση δεδομένων, αυτό που κάνουμε είναι η αντίστροφη διαδικασία, δηλαδή επιλέγουμε Server>Data import, στο αντίστοιχο choiceBox την επιλογή "Import from Self-Contained File" και την επιλογή "Dump Structure and Data". Τέλος, μεταβαίνουμε στην καρτέλα Import Progress και πατάμε το κουμπί "Start Import", προκειμένου να φορτωθεί το backup μας.



2.2 ΔΙΑΓΡΑΜΜΑΤΑ ΠΑΚΕΤΩΝ / ΥΠΟΣΥΣΤΗΜΑΤΩΝ ΚΕΝΤΡΙΚΗΣ ΕΦΑΡΜΟΓΗΣ

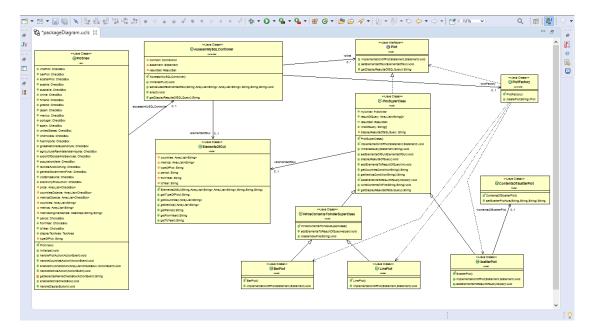
Στην εικόνα που ακολουθεί, φαίνεται το διάγραμμα πακέτων του project μας. Όπως θα εξηγηθεί και παρακάτω έχουμε υλοποιήσει το **μοντέλο MVC**.



Όπως παρατηρούμε, υπάρχουν τρία πακέτα. Στο πακέτο view πραγματοποιείται η διαχείριση του γραφικού περιβάλλοντος της εφαρμογής. Στο πακέτο model εκτελούνται οι ερωτήσεις του project ενώ τέλος το πακέτο controller λειτουργεί ως συνδετικός κρίκος μεταξύ των δύο πακέτων. Στόχος μας ήταν να μην υπάρχει άμεση επικοινωνία μεταξύ του γραφικού περιβάλλοντος και της εσωτερικής υλοποίησης, έτσι ώστε να μειώσουμε τις εξαρτήσεις και να κάνουμε τον κώδικα απλό, κατανοητό και επεκτάσιμο.

2.3 ΔΙΑΓΡΑΜΜΑ(ΤΑ) ΚΛΑΣΕΩΝ ΚΕΝΤΡΙΚΗΣ ΕΦΑΡΜΟΓΗΣ

Στην εικόνα που ακολουθεί, φαίνεται το διάγραμμα κλάσεων του project μας. Όπως θα εξηγηθεί και παρακάτω, έχουμε εφαρμόσει διάφορα patterns για την αποφυγή διπλότυπου κώδικα και για να καταστήσουμε τον κώδικα επεκτάσιμο.



Αρχικά, για την υλοποίηση του γραφικού περιβάλλοντος χρησιμοποιήσαμε τη βιβλιοθήκη javaFX και το εργαλείο SceneBuilder. Το εργαλείο SceneBuilder παράγει αυτόματα ένα αρχείο τύπου .fxml. Το αρχείο αυτό περιέχει όλα τα components που βρίσκονται πάνω στο γραφικό περιβάλλον. Η εκκίνηση της εφαρμογής πραγματοποιείται από την κλάση MainMenuGUI, η οποία συνδέει το .fxml αρχείο με τη κλάση PlotView και απεικονίζει το γραφικό περιβάλλον στο χρήστη. Τη διαχείριση του γραφικού περιβάλλοντος την αναλαμβάνει ο PlotView, όπως φαίνεται και στο UML diagram. Η πληροφορία που δίνει ο χρήστης μεταφέρεται από την κλάση PlotView στην

κλάση AccessInMySQLController, η οποία αναλαμβάνει να αποθηκεύσει την πληροφορία που έδωσε ο χρήστης στην κλάση ElementsOfGUI.

Στην κλάση AccessInMySQLController υπάρχει ένα HashMap που διατηρεί τα αντικείμενα για κάθε plot. Αυτό δημιουργήθηκε για να απαλείψουμε τον διπλότυπο κώδικα και σε περίπτωση που επιθυμούμε μελλοντικά να προσθέσουμε κάποιο επιπλέον plot, αυτό να πραγματοποιείται απλά προσθέτοντας μία γραμμή στο αρχείο nameOfPlots.txt. Με τον τρόπο αυτό το πρόγραμμά μας γίνεται επεκτάσιμο και κατανοητό. Επίσης, έχουμε πιο καθαρό κώδικα με λιγότερες εξαρτήσεις. Τέλος, η κλάση AccessInMySQLController ελέγχει ποια από τις τρεις επιλογές (BarPlot, LinePlot, ScatterPlot) έχει επιλέξει ο χρήστης και καλεί τη κατάλληλη μέθοδο της αντίστοιχης κλάσης-παιδί.

Συνεχίζουμε την ανάλυσή μας με το πακέτο model, το οποίο περιέχει και την εσωτερική υλοποίηση. Εδώ έχουμε υλοποιήσει ένα συνδυασμό των **Template** και **Strategy patterns**, καθώς υπάρχει ένα γενικό interface ονόματι Plot. Για κάθε ένα plot, υπάρχει μία κλάση που κάνει **έμμεσα** implements την κλάση Plot. Τονίζουμε τη λέξη έμμεσα, διότι ενδιάμεσα υπάρχουν δύο abstract κλάσεις που μας βοηθούν να απαλείψουμε διπλότυπο κώδικα χρησιμοποιώντας το **SuperClass refactory**. Πιο συγκεκριμένα, η κλάση WriteContentsToIndexSuperClass περιέχει κοινό κώδικα μεταξύ των κλάσεων BarPlot και LinePlot (την κληρονομούν) και κάνει extends την κλάση PlotSuperClass. Η κλάση PlotSuperClass περιέχει κοινό κώδικα των κλάσεων WriteContentsToIndexSuperClass και του ScatterPlot και κάνει implements το interface Plot. Τέλος, η κλάση PlotFactory πραγματοποιεί τη δημιουργία των αντικειμένων τύπου Plot, ώστε να απομονώσουμε τη δημιουργία τους. Αυτό επιτυγχάνεται χρησιμοποιώντας το **Factory method pattern**.

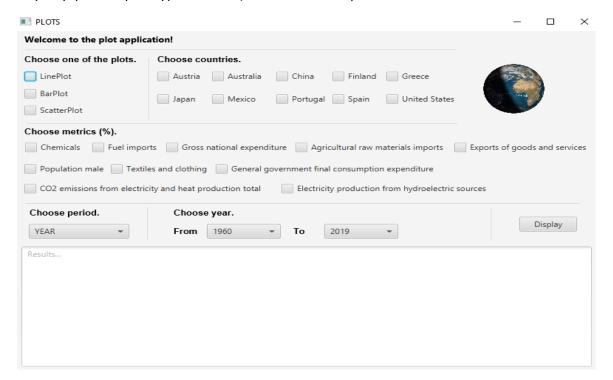
Στόχος όλων των παραπάνω είναι η απαλοιφή του διπλότυπου κώδικα και η επίτευξη της επεκτασιμότητας αυτού.

Η εφαρμογή μας μέσω της κλάσης AccessInMySQLController συνδέεται με τη βάση δεδομένων της mySQL. Επιπρόσθετα, θα θέλαμε να σημειώσουμε ότι η βάση δεδομένων της mySQL ανταλλάσσει πληροφορίες με το πακέτο model μέσω της σύνδεσης που επιτεύχθηκε νωρίτερα. Τέλος, για την οπτικοποίηση των αποτελεσμάτων χρησιμοποιήσαμε τον **Web Server for Chrome** 2 και εμφανίζεται στο browser.

 $^{^2}$ Για την εκκίνηση του server αυτού απλά τον ενεργοποιούμε μέσω του Google Chrome.

3 ΥΠΟΔΕΙΓΜΑΤΑ ΕΡΩΤΗΣΕΩΝ ΚΑΙ ΑΠΑΝΤΗΣΕΩΝ

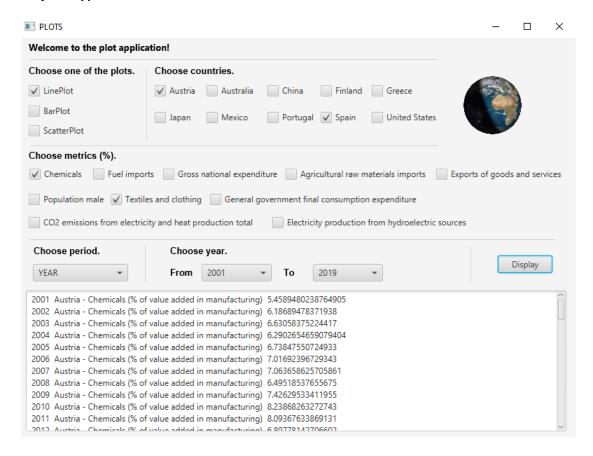
Σε αυτό το κεφάλαιο θα κάνουμε μία παρουσίαση της εφαρμογής, παραθέτοντας συγκεκριμένα παραδείγματα καθώς και τα αποτελέσματα αυτών.



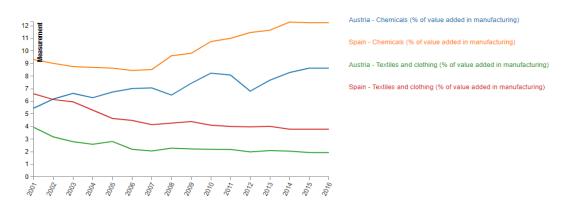
Στην παραπάνω εικόνα φαίνεται η αρχική διεπαφή που εμφανίζεται στο χρήστη κατά την εκκίνηση του προγράμματος. Όπως παρατηρούμε ο χρήστης μπορεί, αρχικά, να επιλέξει ένα από τα τρία plots. Στη συνέχεια, μπορεί να επιλέξει ένα συνδυασμό από χώρες και μετρικές, να καθορίσει χρονική περίοδο (5ετία, κλπ) καθώς και το χρονικό διάστημα που επιθυμεί. Αφού επιλέξει τις επιλογές που επιθυμεί και πατήσει το κουμπί Display, εμφανίζονται τα αποτελέσματα της ερώτησης που εκτελέστηκε στο λευκό πλαίσιο, ενώ σχεδόν ταυτόχρονα ανοίγει ένας browser με την οπτικοποίηση του αποτελέσματος. Στη συνέχεια, θα παραθέσουμε κάποια σχετικά παραδείγματα της εκτέλεσης του προγράμματος.

Κάθε plot περιλαμβάνει και animation.

Παράδειγμα 1

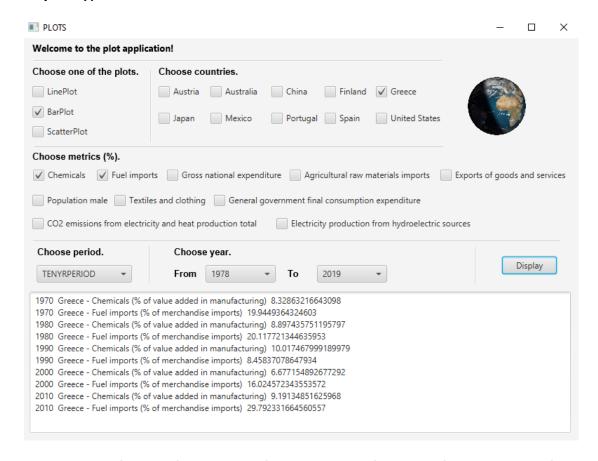


Το παραπάνω παράδειγμα αφορά το LinePlot και όπως αναλύσαμε προηγουμένως απεικονίζονται τα αποτελέσματα στο λευκό πλαίσιο, ενώ στην παρακάτω φωτογραφία φαίνεται η οπτικοποίηση του αποτελέσματος.

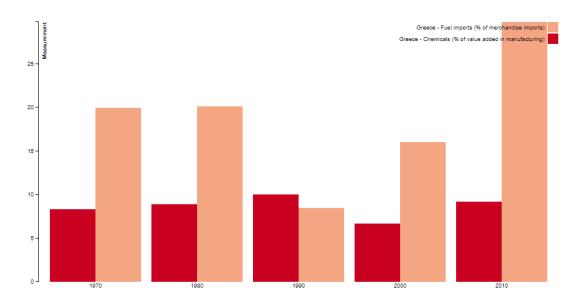


Μία επιπλέον δυνατότητα που παρέχεται στον χρήστη, είναι ότι μπορεί κάνοντας click στο όνομα της αντίστοιχης γραμμής, να εξαφανίσει τη γραμμή, ενώ μπορεί με τον ίδιο τρόπο να την εμφανίσει. Με αυτόν τον τρόπο μπορεί να συγκρίνει οποιοδήποτε συνδυασμό από τις διαθέσιμες επιλογές επιθυμεί.

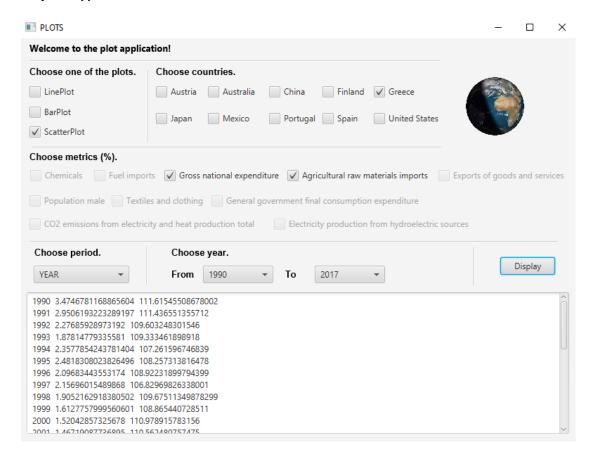
Παράδειγμα 2



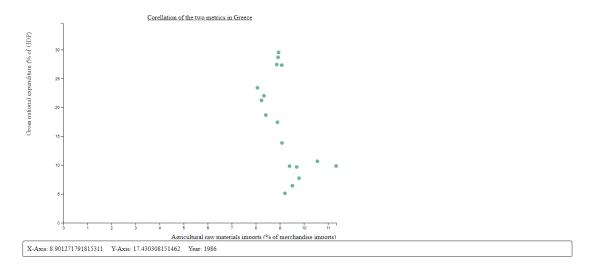
Το παραπάνω παράδειγμα αφορά το BarPlot και όπως αναλύσαμε προηγουμένως απεικονίζονται τα αποτελέσματα στο λευκό πλαίσιο, ενώ στην παρακάτω φωτογραφία φαίνεται η οπτικοποίηση του αποτελέσματος.



Παράδειγμα 3



Το παραπάνω παράδειγμα αφορά το ScatterPlot και όπως αναλύσαμε προηγουμένως απεικονίζονται τα αποτελέσματα στο λευκό πλαίσιο, ενώ στην παρακάτω φωτογραφία φαίνεται η οπτικοποίηση του αποτελέσματος.



Μία επιπλέον δυνατότητα που παρέχεται στον χρήστη, είναι ότι μπορεί να δει τις πληροφορίες για κάθε σημείο ακουμπώντας το ποντίκι πάνω στο αντίστοιχο σημείο, όπως φαίνεται στο παραπάνω σχήμα.

4 ΤΕΚΜΗΡΙΩΣΗ ΚΑΙ ΛΟΙΠΑ ΣΧΟΛΙΑ

Με σκοπό να επιβεβαιώσουμε την ορθότητα του κώδικα, δημιουργήσαμε κάποια test. Πιο συγκεκριμένα, δημιουργήσαμε τρία test. Κάθε ένα από αυτά αντιστοιχίζεται σε ένα plot. Σε αυτά τα test συγκρίναμε τα πραγματικά αποτελέσματα που επιστρέφει η mySQL με τα αποτελέσματα που επιστρέφει η εφαρμογή μας, με σκοπό να επιβεβαιώσουμε ότι είναι ίδια. Όπως φαίνεται και στην παρακάτω εικόνα, τα test έχουν επιτυχία στο σύνολό τους.

