

---

# Πείραμα 2 και ημερολόγιο συναντήσεων μετά τις 28 Φεβρουαρίου 2023

---

Κώστας Παπαδόπουλος

Πειραιάς, 2030

## Contents

1	Περιγραφή	3
2	Δεδομένα	3
2.1	Πηγές δεδομένων . . . . .	3
2.2	Επιλογή δεδομένων . . . . .	5
3	Πειράματα	5
4	Αποτελέσματα	5
4.1	Πείραμα 1 . . . . .	6
4.2	Πείραμα 2 . . . . .	7
4.3	Πείραμα 3 . . . . .	7
4.4	Πείραμα 4 . . . . .	9
5	Προτάσεις	10
6	Ένθετο	10
7	Επανεκίνηση μετά τα Χριστούγεννα	11
8	Δοκιμές για διαφορετικά βάρη στις παραμέτρους	13
9	Για τη 31η Γενάρη	14
10	Για την 28η Φεβρουαρίου	19
10.1	Αλλαγή 1 . . . . .	19
10.2	Μεσαία χώρα ανά τα έτη . . . . .	19
10.3	Όλες οι χώρες για όλα τα έτη . . . . .	20
11	Για τη 14η Μαρτίου	23
11.1	Πίνακας $r^2$ με βέλτιστα επιλεγμένα βάρη . . . . .	23

11.2	Οπτικοποίηση δεδομένων . . . . .	25
11.2.1	Πληθυσμός . . . . .	25
11.2.2	Πληθωρισμός . . . . .	27
11.2.3	Total energy Supply . . . . .	29
11.2.4	GDP per capita . . . . .	29
11.2.5	Verified emissions . . . . .	30
11.2.6	Agriculture . . . . .	33
11.2.7	Industry . . . . .	35
11.2.8	Manufacturing . . . . .	35
<b>12</b>	<b>Για την 21η Μαρτίου</b>	<b>37</b>
12.1	Γιατί η Πολωνία και η Γαλλία αποτυγχάνουν να εξηγήσουν τις άλλες πριν το 2012; . . . . .	37
12.2	Τα βέλτιστα βάρη για όλα τα σενάρια . . . . .	38
<b>13</b>	<b>Για την 28η Μαρτίου</b>	<b>39</b>
13.1	Τι όντως έκανα . . . . .	39
13.2	PPS (Purchasing Power Standards) . . . . .	39
13.2.1	Χρήση PPS . . . . .	39
13.2.2	Υπολογισμός . . . . .	39
13.3	Power Intensity . . . . .	40
13.4	Για την επόμενη φορά . . . . .	40
13.4.1	Σε δεύτερο χρόνο . . . . .	40
13.4.2	Άμεσα, για το τετρασέλιδο . . . . .	41
<b>14</b>	<b>Για τις 4 Απριλίου</b>	<b>41</b>
14.1	Συσταδοποίηση (έπρεπε...) . . . . .	41
14.1.1	Πλήθος συστάδων . . . . .	41
14.1.2	Βέλτιστες συστάδες . . . . .	43
14.1.3	Συστάδες, αν όλα τα δεδομένα ήταν κατά κεφαλήν και κανονικοποιημένα ύστερα . . . . .	43
14.2	Ικανότητα εξήγησης υπολοίπων μέσα στην ίδια συστάδα . . . . .	44
14.3	Υπάρχει κάποια εύκολη σχέση μεταξύ των δωρεάν αδειών και κάποιου feature; . . . . .	45

# 1 Περιγραφή

Σε αυτό το σημείο θα γίνει η περιγραφή του τι προσπαθήσαμε να κάνουμε σε αυτό το βήμα. Η υπόθεση που θα προσπαθήσουμε να δούμε αν ισχύει είναι πολύ απλή και λογική ταυτόχρονα. **"Αν μοιάζουμε, τότε πρέπει να μας μεταχειρίζονται με παρόμοιο τρόπο."** Ισχύει κάτι τέτοιο στο ETS;

Για να απαντήσουμε σε αυτό το ερώτημα, θα κάνουμε μία πολύ απλουστευτική παραδοχή. Το ότι χώρες με παρόμοιο κατά κεφαλήν ακαθάριστο προϊόν, πληθωρισμό, πληθυσμό, παροχή ενέργειας και αναλογίες στο ακαθάριστο προϊόν "μοιάζουν". Επομένως, θα ορίσουμε 2 διάνυσματα για κάθε χώρα  $i$ :

$$\chi\omega\rho\alpha_i = \langle GDP\_PC_i, inflation_i, population_i, total\_energy\_supply_i \rangle$$

$$\chi\omega\rho\alpha1_i = \langle GDP\_PC_i, inflation_i, population_i, total\_energy\_supply_i \rangle$$

$$\chi\omega\rho\alpha2_i = \langle GDP\_PC_i, inflation_i, population_i, total\_energy\_supply_i, \\ verified\_emissions_i, GDP\_Agricultural_i, GDP\_Industrial_i, GDP\_Manufacturing_i \rangle$$

Από εκεί, θα υπολογίσουμε τη νόρμα 2 (Ευκλείδεια απόσταση των σημείων) από κάθε χώρα προς κάθε άλλη χώρα.

$$distance\_simulated_{i,j} = \|\chi\omega\rho\alpha2_i - \chi\omega\rho\alpha2_j\|_2$$

Στη συνέχεια, υπολογίζουμε τις αποστάσεις οι οποίες προκύπτουν κάπως έτσι για το 2015, για το διάνυσμα χώρα2 (όπως φαίνεται στους πίνακες 1,2,3 στο ένθετο (οι πίνακες χρησιμοποιούνται ως διαγώνιοι πίνακες, όμως εδώ δεν είχε τόσο νόημα αυτό, καθώς δεν χωρά σε μία σελίδα)).

Στη συνέχεια υπολογίζουμε τον αριθμό των δωρεάν αδειών που έχουν λάβει αυτές οι χώρες. Με σκοπό να υπολογίσουμε τη δεύτερη απόστασή τους.

$$distance2_{i,j} = \|free_i - free_j\|_2$$

Τέλος, προσπαθούμε να δούμε αν υπάρχει κάποια σχέση μεταξύ αυτών των αποστάσεων.

## 2 Δεδομένα

### 2.1 Πηγές δεδομένων

Ας δούμε λίγο πιο αναλυτικά όλα τα δεδομένα:

- Total energy supply.
  - Source: [https://ec.europa.eu/eurostat/databrowser/view/nrg\\_bal\\_s\\_1/default/table?lang=en](https://ec.europa.eu/eurostat/databrowser/view/nrg_bal_s_1/default/table?lang=en)
  - Year: 2011 - 2020
  - Unit: Thousand tonnes of oil equivalent
- Inflation.
  - Source: <https://data.worldbank.org/indicator/FP.CPI.TOTL.ZG>
  - Year: 1960-2021
  - Unit: %

Σημαντική παρατήρηση: Αυτή η τιμή δε λογαριθμήθηκε όταν λογαριθμήθηκαν οι άλλες.

- GDP per capita.
  - Source: <https://data.worldbank.org/indicator/NY.GDP.PCAP.CD>
  - Year: 1960 - 2021
  - Unit: US\$
- Population.
  - Source: <https://data.worldbank.org/indicator/SP.POP.TOTL>
  - Year: 2011 - 2021
  - Unit: Persons
- Verified emissions.
  - Source: EU ETS Database
  - Table: eutl\_compliance
  - Column: verified
- Free Allowances.
  - Source: EU ETS Database
  - Table: eutl\_compliance
  - Column: freeAlloc
- Agriculture.
  - Source: <http://wdi.worldbank.org/table/4.2>
  - Year: 2020
  - Unit: Billions USD \* Agricultural percentage.

Σημαντική παρατήρηση: Για αυτό το δεδομένο χρησιμοποιήθηκε η τιμή του 2020, όχι του 2015.

- Industry.
  - Source: <http://wdi.worldbank.org/table/4.2>
  - Year: 2020
  - Unit: Billions USD \* Industry percentage.

Σημαντική παρατήρηση: Για αυτό το δεδομένο χρησιμοποιήθηκε η τιμή του 2020, όχι του 2015.

- Manufacturing.
  - Source: <http://wdi.worldbank.org/table/4.2>
  - Year: 2020
  - Unit: Billions USD \* Manufacturing percentage.

Σημαντική παρατήρηση: Για αυτό το δεδομένο χρησιμοποιήθηκε η τιμή του 2020, όχι του 2015.

## 2.2 Επιλογή δεδομένων

Το πρώτο διάνυσμα αποτελείται από

- Τον πληθυσμό.
- Τον πληθωρισμό.
- Το κατά κεφαλήν ακαθάριστο προϊόν.
- Το σύνολο της διαθέσιμης ενέργειας στη χώρα.

Παράλληλα το δεύτερο διάνυσμα έχει όλα τα παραπάνω και επιπλέον:

- Verified Emissions.
- Ακαθάριστο προϊόν το οποίο σχετίζεται με Γεωργικές διαδικασίες.
- Ακαθάριστο προϊόν το οποίο σχετίζεται με Βιομηχανίες.
- Ακαθάριστο προϊόν το οποίο σχετίζεται με Κατασκευαστικά έργα.

Μερικά ενδιαφέροντα στοιχεία είναι τα παρακάτω.

Αρχικά, δε χρησιμοποιήσαμε δεδομένα τα οποία να σχετίζονται με τις υπηρεσίες ως εκατομμύρια του ακαθάριστου προϊόντος της κάθε χώρας, καθώς αυτή η παράμετρος πρέπει να έχει πολύ μικρή συσχέτιση με τους ρύπους του διοξειδίου.

Στο δεύτερο διάνυσμα μπήκαν πολλές τιμές οι οποίες έχουν να κάνουν με το μέγεθος της παραγωγής, αλλά και με πολλές απόλυτες τιμές του ακαθάριστου προϊόντος. Αυτό έχει σαν αποτέλεσμα το να τονίζεται ακόμα περισσότερο η διαφορά της Γερμανίας αλλά και των υπόλοιπων πιο ανεπτυγμένων χωρών με τις υπόλοιπες.

## 3 Πειράματα

Ουσιαστικά κάναμε 3 πειράματα με αυτά τα δεδομένα.

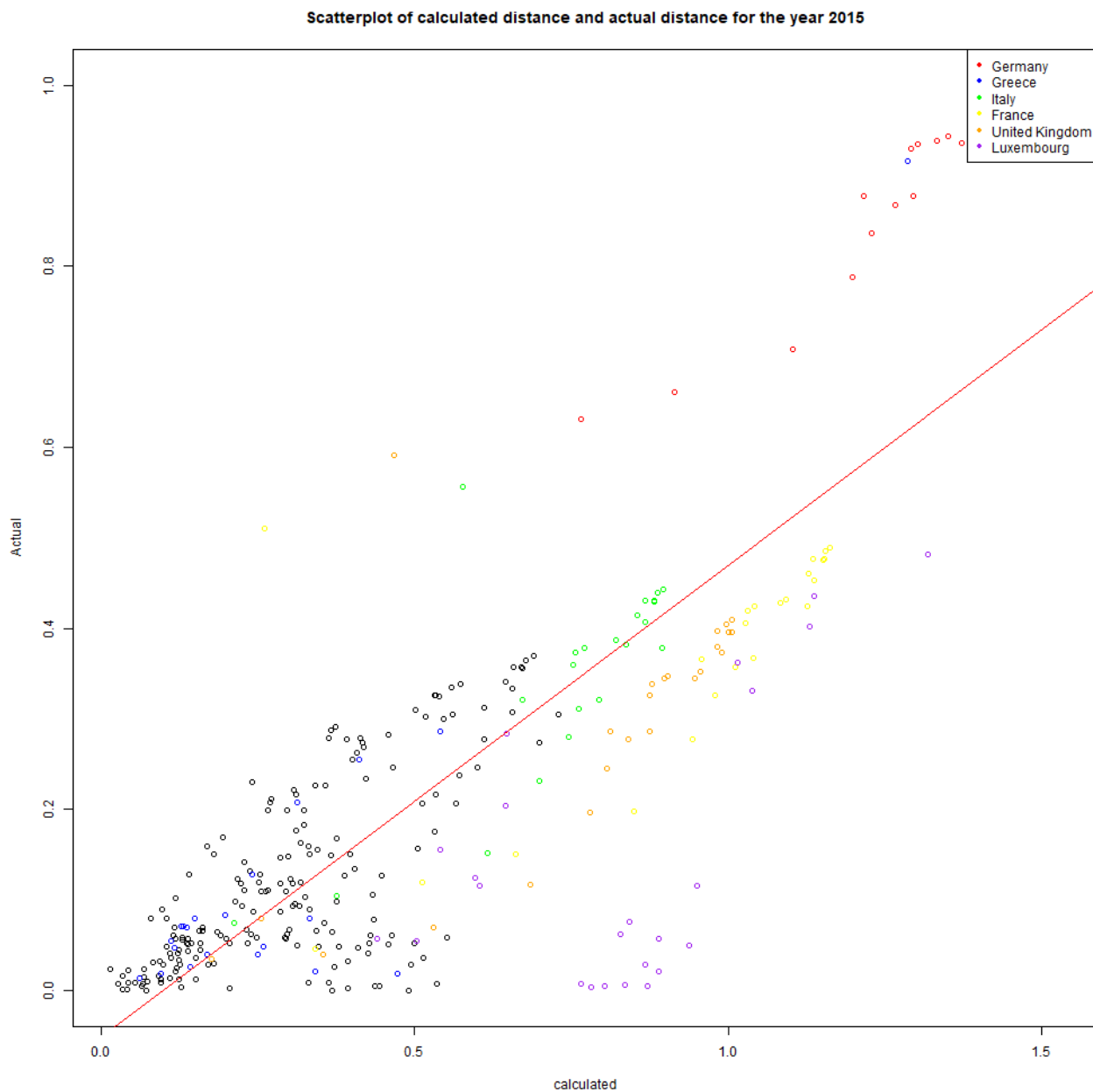
1. Χώρες που μοιάζουν είναι όντως κοντά σε δωρεάν άδειες; (spoiler ναι.)
2. Εάν επαναλάβουμε τους υπολογισμούς για διαφορετικές χρονιές, τα αποτελέσματα παραμένουν ίδια ή υπάρχει διακύμανση; (spoiler παραμένουν ίδια.) Σε αυτήν την ερώτηση περιμέναμε από πριν πως η απάντηση θα είναι "παραμένουν ίδια" καθώς καμία από τις τιμές που έχουν στη διάθεσή μας δεν αλλάζει σημαντικά και το μέγεθος με το οποίο συγκρίνουμε τα αποτελέσματά μας παραμένει επίσης αρκετά σταθερό.
3. Τι θα συμβεί αν βάλουμε περισσότερες πληροφορίες οι οποίες να περιγράφουν καλύτερα τις διαφορές χωρών με μεγαλύτερη οικονομική δύναμη; Θα μειωθεί η απόκλιση; (spoiler ναι.)
4. Εφόσον Η Γερμανία είναι σαν να παίζει άλλο παιχνίδι, αντί να συμμετέχει στην ίδια διαδικασία, θα είχε νόημα να λογαριθμίσουμε όλα τα δεδομένα, προκειμένου να αποκτήσει περισσότερη ανάλυση στις χαμηλές τιμές και να έρθει σε πιο "γήινες" τιμές η Γερμανία. Αν το κάνουμε αυτό, η γραμμική σχέση στέκει; (spoiler Πλέον είναι σαφές πως η σχέση δεν είναι τόσο γραμμική όσο φαινόταν.)

## 4 Αποτελέσματα

Σε όλα τα ακόλουθα διαγράμματα κάθε σημείο απεικονίζει ένα ζευγάρι χωρών. Ο οριζόντιος άξονας δείχνει το πώς μετρήθηκε η απόσταση αυτών των δύο χωρών, ενώ ο κατακόρυφος άξονας δείχνει την ίδια απόσταση στο πόσες δωρεάν άδειες έλαβαν οι χώρες αυτές. Οι χώρες οι οποίες εμφανώς έχουν διαφορετική συμπεριφορά από το γενικό σύνολο έχουν τα δικά τους χρώματα.

## 4.1 Πείραμα 1

Το πρώτο πείραμα έχει το παρακάτω αποτέλεσμα για το 2015:



Παρατηρήσεις:

- Η πρώτη παρατήρηση είναι πως τα δεδομένα δείχνουν να έχουν μία σχεδόν γραμμική σχέση, με μερικές εξαιρέσεις, όπως το Λουξεμβούργο, τη Γερμανία και τη Γαλλία.
- Ακόμα και έτσι, οι χώρες οι οποίες παρεκκλίνουν μοιάζουν σαν να υπακούν μία δική τους γραμμική σχέση με παρόμοια κλίση, αλλά διαφορετική dc συνιστώσα. Το οποίο σημαίνει πως δεν έχουν καταφέρει με τα δεδομένα

μας να μαζέψουμε όση πληροφορία χρειάζεται ή πως υπάρχει εγγενώς μία αδικία και η Γερμανία είναι μονίμως ευνοημένη ενώ το Λουξεμβούργο είναι συνεχώς αδικημένο.

## 4.2 Πείραμα 2

Εδώ δοκιμάζουμε για να δούμε αν υπάρχει κάποια σημαντική αλλαγή μέσα στα χρόνια από το 2014 μέχρι το 2019.

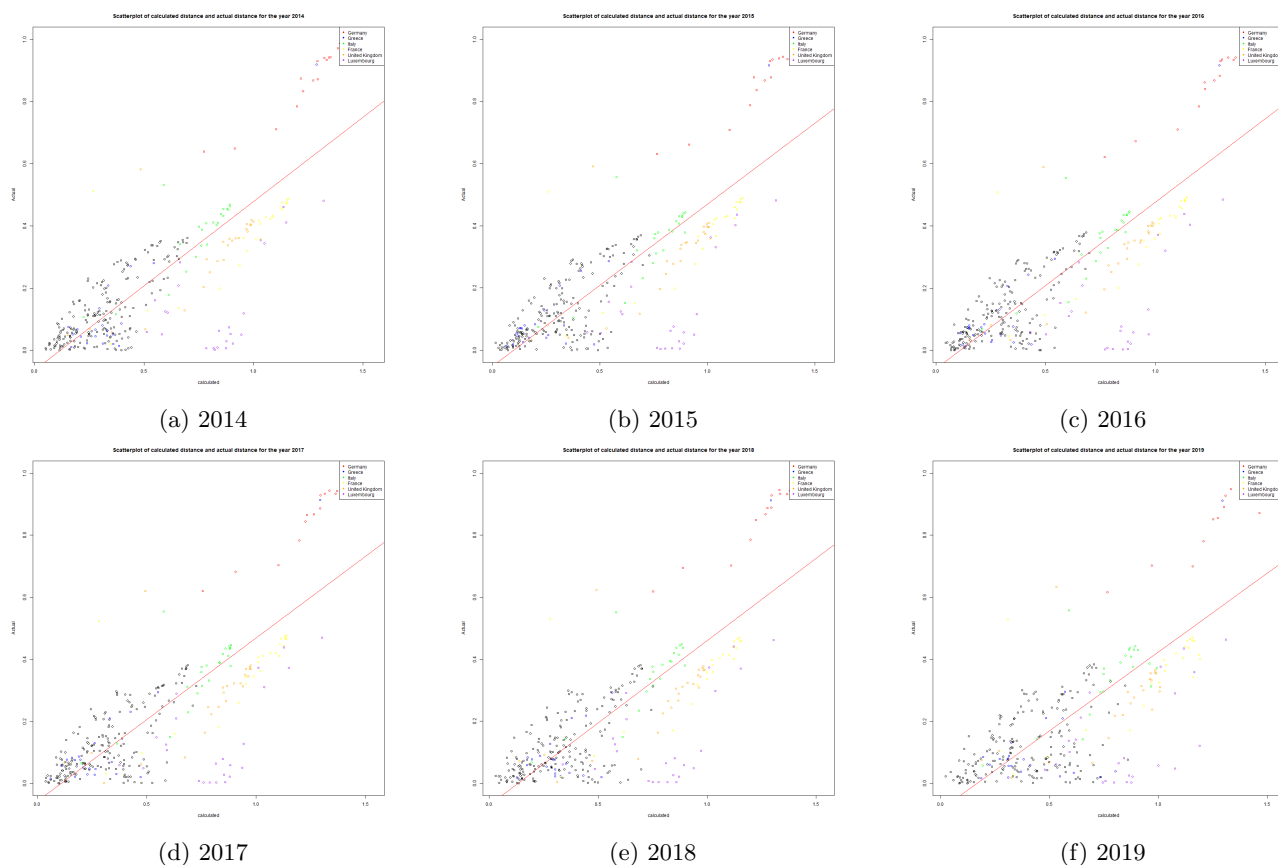
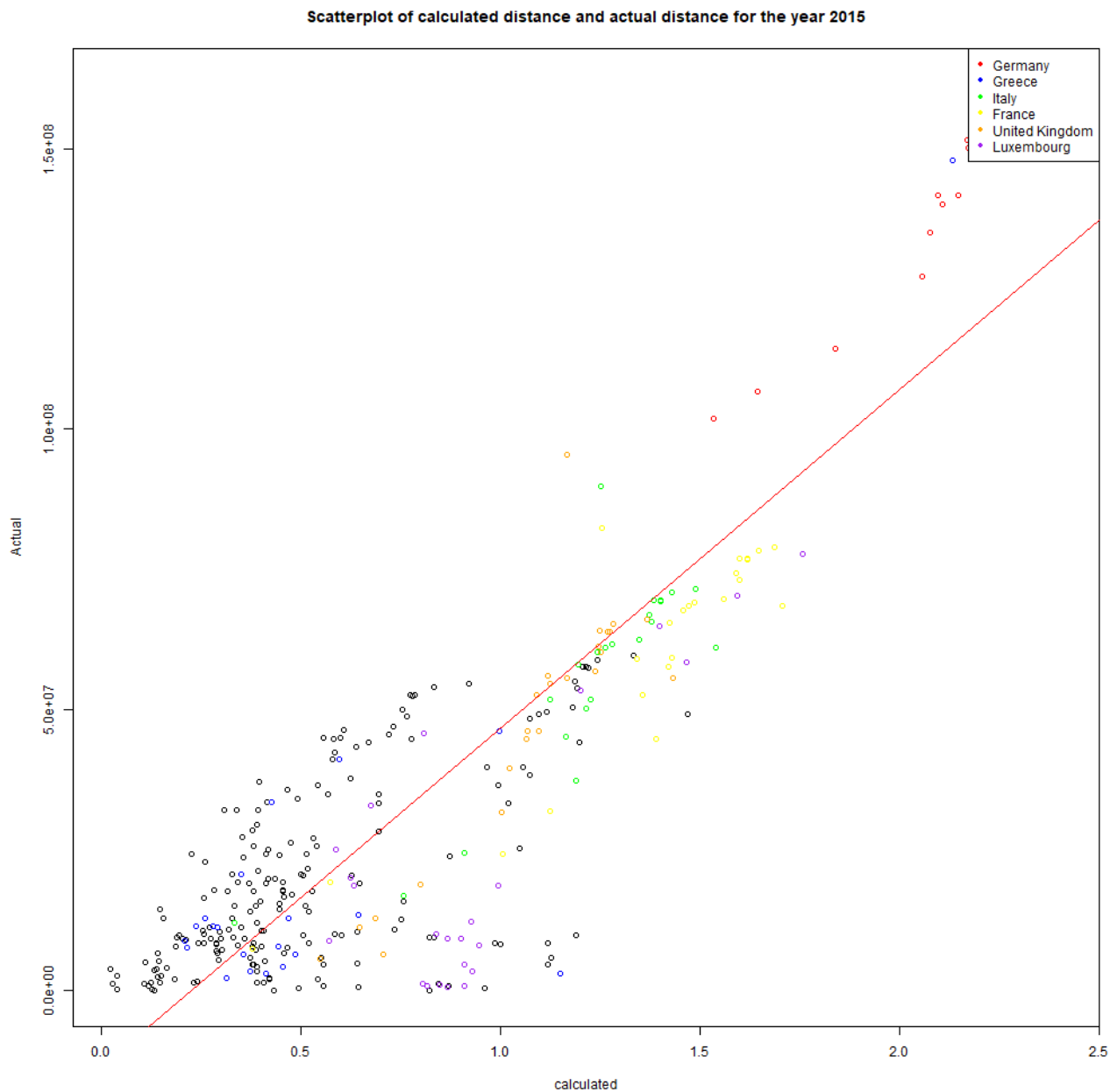


Figure 1: Συγκριτικό διάγραμμα για το πείραμα 2

## 4.3 Πείραμα 3

Όπως είδαμε ήδη στο πείραμα 1, ή υπάρχει κάποια εγγενής αδικία ή χρειαζόμαστε περισσότερα δεδομένα, τα οποία να κάνουν τις αποστάσεις των χωρών πιο αντιπροσωπευτικές. Για να το πετύχουμε αυτό, συμπεριλάβαμε περισσότερα δεδομένα, τα οποία να μπορούν να δείξουν τη βασική διαφορά μεταξύ του λουξεμβούργου και της Γερμανίας. Τη διαφορά στην κλίμακα της οικονομίας για την οποία μιλάμε. Έτσι βάλαμε 3 μεγέθη τα οποία έχουν να κάνουν και με την ανάπτυξη της οικονομίας αλλά και με τον καταμερισμό της οικονομίας σε διαφορετικούς κλάδους. Φροντίσαμε επίσης, όλα τα νέα στοιχεία να σχετίζονται άμεσα με την παραγωγή διοξειδίου, οπότε για παράδειγμα οι υπηρεσίες της κάθε χώρας αφείθηκαν εκτός. Επίσης, γνωρίζουμε από πριν πως αυτό το μείγμα χαρακτηριστικών δίνει μεγαλύτερη βαρύτητα στο μέγεθος της οικονομίας. Έτσι προκύπτουν τα παρακάτω:



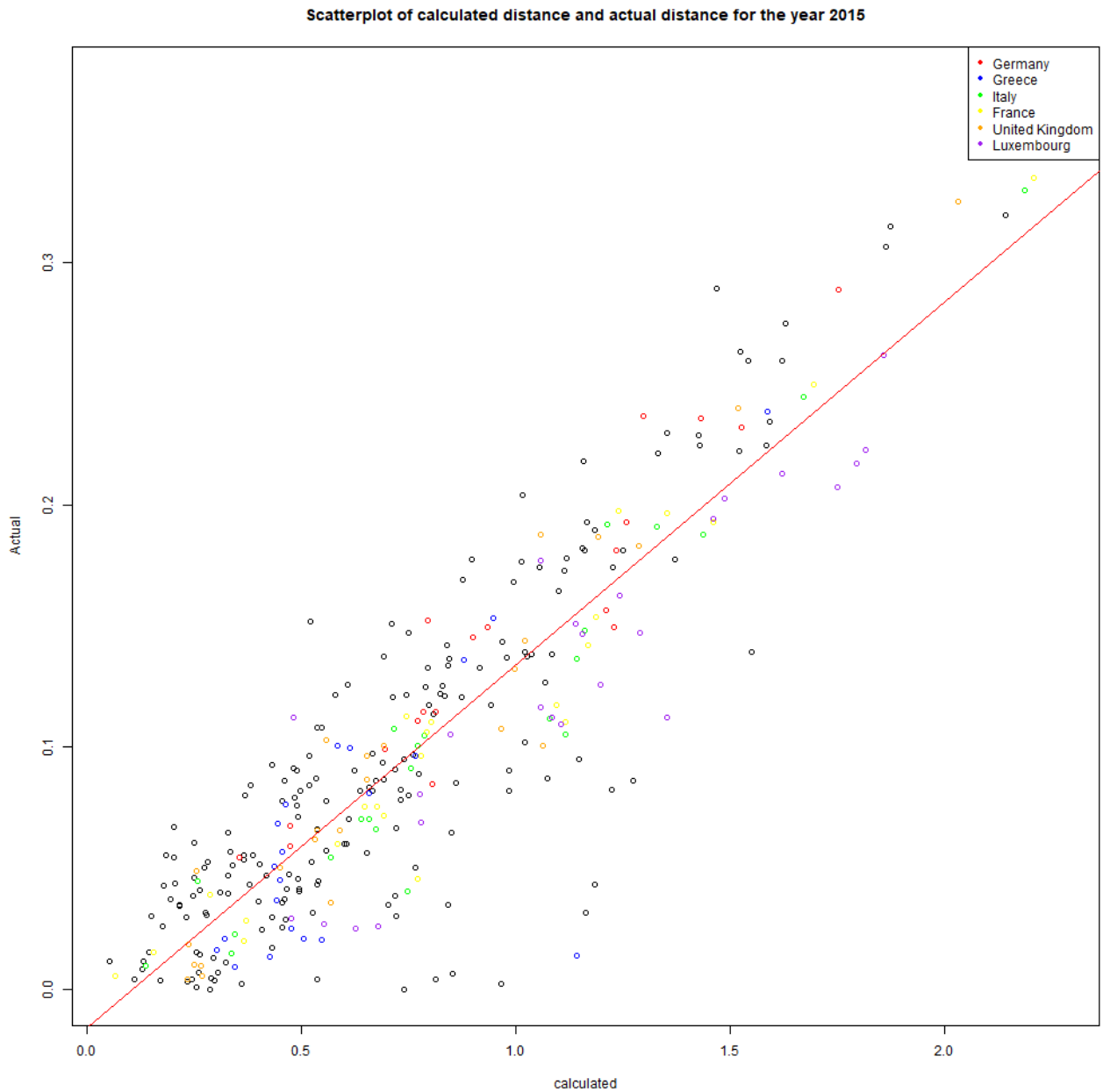
Παρατηρήσεις:

- Παρατηρούμε πως όλες οι χώρες έχουν έρθει πιο κοντά στη γραμμή. Όμως αυτό έχει να κάνει περισσότερο με το ότι η Γερμανία βρέθηκε πιο μακριά από τις υπόλοιπες, με αποτέλεσμα να "μειώνεται η ανάλυση" στις μικρότερες χώρες. Συνεπώς, παρόλο που όλα δείχνουν να είναι πιο κοντά στη γραμμή, στην πραγματικότητα απλώς έχουμε "ξεζουμάρει" και το γράφημα είναι παρόμοιο με πριν.
- Η προηγούμενη παρατήρηση βέβαια δεν αναιρεί το γεγονός ότι αυτό το μοντέλο όντως καταφέρνει να περιγράψει καλύτερα την Γερμανία. Απλώς χάνει σε όλες τις υπόλοιπες χώρες.



#### 4.4 Πείραμα 4

Βλέποντας την αποτυχία του προηγούμενου μοντέλου, θα προσπαθήσουμε να λογαριθμίσουμε τα πάντα, ώστε να έχουμε περισσότερη ανάλυση στις χαμηλές τιμές και να φέρουμε τη Γερμανία σε πιο χαμηλές τιμές. Έτσι προκύπτει το παρακάτω:



Παρατηρήσεις:

- Black Magic.
- Πέρα από την πλάκα, είναι σαφές πως εδώ όλα έχουν έρθει αρκετά πιο κοντά όλες οι τιμές. Επομένως, οι

ίδιες αποκλίσεις με πριν πλέον σηματοδοτούν ακόμα μεγαλύτερη πραγματική απόσταση. Όμως το αποτέλεσμα δείχνει πως κάποια μορφή τέτοιες "δικαιοσύνης" υπάρχει όντως.

## 5 Προτάσεις

Αρχικά για να επαναλάβει κάποιος τα πειράματα αυτά θα χρειαστεί την βάση EU ETS και αυτό εδώ το github: [https://github.com/kwpap/Diplomatiki\\_kwpap\\_step\\_1](https://github.com/kwpap/Diplomatiki_kwpap_step_1)

- Καλύτερη επιλογή διανυσμάτων.
- Βάρη στην κάθε τιμή του κάθε διανύσματος. Εδώ πρέπει βέβαια να δοθεί προσοχή για να αποφύγουμε το overfitting.
- Συζήτηση για μελλοντικά ενδιαφέροντα πράγματα που μπορούν να γίνουν.
- Θα ήταν ενδιαφέρον αν σε ένα δεύτερο επίπεδο δε μιλούσαμε πλέον για αποστάσεις, αλλά για απόλυτα νούμερα. Ανεξαρτήτως χώρας και πλέον η αναζήτηση είχε ως σκοπό να βρει τα χαρακτηριστικά εκείνα τα οποία μπορούν να αυξήσουν τις δωρεάν άδειες μίας χώρας.

## 6 Ένθετο

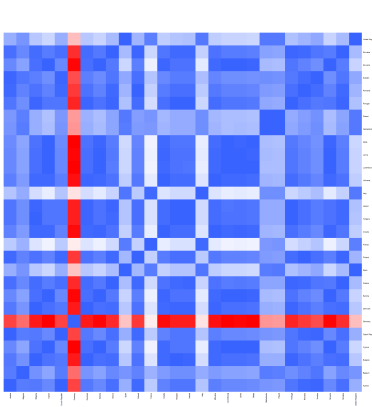
Παρακάτω φαίνονται τα χαρακτηριστικά των γραμμικών παλινδρομήσεων:

Test	Slope	intercept	Multiple $R^2$	Adjusted $R^2$	p-value
2014	0.54185	-0.06402	0.6968	0.6958	< 2.2e-16
2015	0.52049	-0.05067	0.6823	0.6813	< 2.2e-16
2016	0.53501	-0.05890	0.6916	0.6906	< 2.2e-16
2017	0.52720	-0.05817	0.6752	0.6742	< 2.2e-16
2018	0.52966	-0.06977	0.6625	0.6615	< 2.2e-16
2019	0.50753	-0.08467	0.6008	0.5996	< 2.2e-16
2015 all data	60298846	-13563147	0.7886	0.788	< 2.2e-16
2015 all data + log	0.149767	-0.015708	0.7977	0.7971	< 2.2e-16

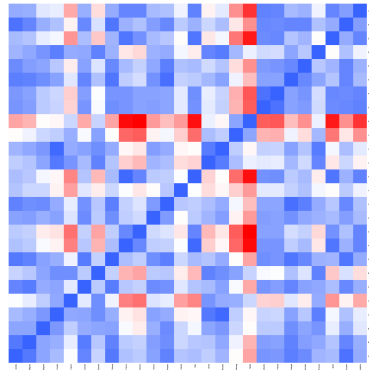
Στα γραφήματα 2 φαίνονται κάποιες από τις διαφορές των χωρών σε μία κλίμακα στην οποία το μπλε είναι η μικρότερη απόσταση και το κόκκινο είναι η μεγαλύτερη απόσταση.

	Austria	Belgium	Bulgaria	Croatia	Cyprus	Denmark	Estonia	Finland	France	Germany
Austria	0.00	0.18	0.49	0.64	1.12	0.26	0.83	0.17	0.69	0.77
Belgium	0.18	0.00	0.49	0.55	1.02	0.20	0.75	0.23	0.77	0.81
Bulgaria	0.49	0.49	0.00	0.40	0.79	0.54	0.49	0.47	1.09	1.21
Croatia	0.64	0.55	0.40	0.00	0.52	0.50	0.33	0.56	1.19	1.26
Cyprus	1.12	1.02	0.79	0.52	0.00	0.98	0.33	1.06	1.70	1.75
Denmark	0.26	0.20	0.54	0.50	0.98	0.00	0.75	0.19	0.74	0.80
Estonia	0.83	0.75	0.49	0.33	0.33	0.75	0.00	0.80	1.46	1.53
Finland	0.17	0.23	0.47	0.56	1.06	0.19	0.80	0.00	0.68	0.78
France	0.69	0.77	1.09	1.19	1.70	0.74	1.46	0.68	0.00	0.29
Germany	0.77	0.81	1.21	1.26	1.75	0.80	1.53	0.78	0.29	0.00
Greece	0.48	0.44	0.55	0.46	0.95	0.30	0.77	0.32	0.78	0.88
Hungary	0.25	0.33	0.30	0.54	1.02	0.36	0.73	0.22	0.80	0.93
Ireland	0.72	0.85	0.85	1.18	1.55	0.97	1.22	0.84	1.12	1.23
Italy	0.67	0.75	1.08	1.16	1.67	0.72	1.44	0.66	0.07	0.26
Latvia	0.79	0.71	0.49	0.18	0.37	0.66	0.23	0.71	1.35	1.43
Lithuania	0.61	0.52	0.41	0.05	0.54	0.46	0.34	0.54	1.17	1.23
Luxembourg	1.14	1.06	0.85	0.78	0.55	1.11	0.48	1.15	1.81	1.86
Malta	1.52	1.47	1.16	1.16	0.86	1.52	0.84	1.54	2.21	2.27
Netherlands	0.38	0.43	0.77	0.83	1.33	0.38	1.10	0.33	0.37	0.47
Poland	0.40	0.46	0.76	0.84	1.35	0.43	1.11	0.36	0.37	0.47
Portugal	0.21	0.25	0.32	0.47	0.97	0.26	0.69	0.15	0.79	0.90
Romania	0.30	0.43	0.49	0.73	1.23	0.46	0.94	0.29	0.65	0.81
Slovenia	0.61	0.52	0.37	0.21	0.52	0.52	0.24	0.58	1.24	1.30
Spain	0.65	0.72	1.02	1.08	1.59	0.66	1.37	0.60	0.15	0.36
Sweden	0.13	0.26	0.56	0.69	1.18	0.28	0.92	0.14	0.58	0.69
United Kingdom	0.53	0.57	0.97	1.02	1.52	0.56	1.29	0.54	0.27	0.25

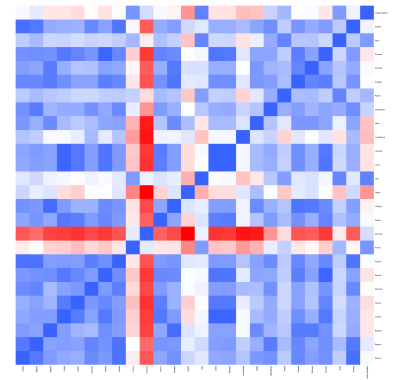
Table 1: Distance between countries in 2015, part 1



(a) Με βάση τα free.



(b) Με βάση το διάνυσμα χώρα<sub>2</sub> και λογαριθμημένα δεδομένα.



(c) Με βάση το διάνυσμα χώρα<sub>1</sub>.

Figure 2: Αποστάσεις χωρών σε heatmap

## 7 Επανεκίνηση μετά τα Χριστούγεννα

Έγινε μία καλύτερη επιλογή των πηγών για τα δεδομένα, το οποίο οδήγησε, χωρίς να μπορώ να το εξηγήσω σε κάτι πολύ περίεργες συμπεριφορές. Τα καινούρια δεδομένα ήταν τα:

1. population ίδιο

	Greece	Hungary	Ireland	Italy	Latvia	Lithuania	Luxembourg	Malta	Netherlands	Poland
Austria	0.48	0.25	0.72	0.67	0.79	0.61	1.14	1.52	0.38	0.40
Belgium	0.44	0.33	0.85	0.75	0.71	0.52	1.06	1.47	0.43	0.46
Bulgaria	0.55	0.30	0.85	1.08	0.49	0.41	0.85	1.16	0.77	0.76
Croatia	0.46	0.54	1.18	1.16	0.18	0.05	0.78	1.16	0.83	0.84
Cyprus	0.95	1.02	1.55	1.67	0.37	0.54	0.55	0.86	1.33	1.35
Denmark	0.30	0.36	0.97	0.72	0.66	0.46	1.11	1.52	0.38	0.43
Estonia	0.77	0.73	1.22	1.44	0.23	0.34	0.48	0.84	1.10	1.11
Finland	0.32	0.22	0.84	0.66	0.71	0.54	1.15	1.54	0.33	0.36
France	0.78	0.80	1.12	0.07	1.35	1.17	1.81	2.21	0.37	0.37
Germany	0.88	0.93	1.23	0.26	1.43	1.23	1.86	2.27	0.47	0.47
Greece	0.00	0.43	1.14	0.75	0.61	0.45	1.20	1.59	0.45	0.46
Hungary	0.43	0.00	0.74	0.79	0.67	0.53	1.08	1.43	0.50	0.49
Ireland	1.14	0.74	0.00	1.12	1.27	1.16	1.35	1.58	0.98	0.98
Italy	0.75	0.79	1.12	0.00	1.33	1.14	1.79	2.19	0.34	0.34
Latvia	0.61	0.67	1.27	1.33	0.00	0.20	0.68	1.04	1.00	1.01
Lithuania	0.45	0.53	1.16	1.14	0.20	0.00	0.78	1.17	0.81	0.82
Luxembourg	1.20	1.08	1.35	1.79	0.68	0.78	0.00	0.48	1.46	1.49
Malta	1.59	1.43	1.58	2.19	1.04	1.17	0.48	0.00	1.86	1.87
Netherlands	0.45	0.50	0.98	0.34	1.00	0.81	1.46	1.86	0.00	0.13
Poland	0.46	0.49	0.98	0.34	1.01	0.82	1.49	1.87	0.13	0.00
Portugal	0.34	0.11	0.81	0.77	0.62	0.46	1.06	1.43	0.46	0.46
Romania	0.51	0.21	0.70	0.64	0.87	0.72	1.29	1.62	0.42	0.39
Slovenia	0.58	0.53	1.07	1.21	0.25	0.18	0.63	1.03	0.88	0.90
Spain	0.66	0.74	1.15	0.14	1.25	1.07	1.75	2.14	0.30	0.29
Sweden	0.44	0.27	0.77	0.57	0.84	0.67	1.24	1.63	0.28	0.31
United Kingdom	0.65	0.69	1.06	0.23	1.19	1.00	1.62	2.03	0.24	0.25

Table 2: Distance between countries in 2015, part 2

2. GDP per capita ίδιο

3. Total energy supply ίδιο

4. inflation ίδιο

5. verified emissions ίδιο

6. manufacturing σε δισεκατομύρια δολλάρια, άλλαξε και πλέον δεν έχουμε δεδομένα μόνο για 2010 και 2020, αλλά έχουμε ξεχωριστά για κάθε χρονιά

7. industry σε δισεκατομύρια δολλάρια, ομοίως

8. Agriculture σε δισεκατομύρια δολλάρια, ομοίως

Ένα πιθανό πρόβλημα το οποίο ίσως προκύπτει είναι το ότι  $(6 + 7 + 8) \approx 1 * 2$ . Όμως ρεαλιστικά:  $1 * 2 - (6 + 7 + 8) \approx \text{services}$ . Οπότε ελπίζω πως αυτό δεν είναι σοβαρό πρόβλημα.

Το δεύτερο που παρατηρώ είναι πως για κάποιο λόγο μετά τις βελτιώσεις στον κώδικα, πλέον αν βάλω τα δεδομένα απλώς να λογαριθμιστούν, τα αποτελέσματα είναι λογικά. Αν βάλω τα αποτελέσματα να κανονικοποιηθούν, πάλι είναι λογικά, αλλά προσεγγίζουν λιγότερο την ευθεία γραμμή. Αν όμως τα βάλω και τα δύο, τότε τα δεδομένα δείχνουν να μην έχουν κάποια γραμμική συσχέτιση, το οποίο δεν το καταλαβαίνω.

Επίσης, λόγω έλλειψης δεδομένων για την παραγωγή της Βουλγαρίας, τα δεδομένα υπολογίστηκαν ως GDP - Agriculture - services - industry.

	Portugal	Romania	Slovenia	Spain	Sweden	United Kingdom
Austria	0.21	0.30	0.61	0.65	0.13	0.53
Belgium	0.25	0.43	0.52	0.72	0.26	0.57
Bulgaria	0.32	0.49	0.37	1.02	0.56	0.97
Croatia	0.47	0.73	0.21	1.08	0.69	1.02
Cyprus	0.97	1.23	0.52	1.59	1.18	1.52
Denmark	0.26	0.46	0.52	0.66	0.28	0.56
Estonia	0.69	0.94	0.24	1.37	0.92	1.29
Finland	0.15	0.29	0.58	0.60	0.14	0.54
France	0.79	0.65	1.24	0.15	0.58	0.27
Germany	0.90	0.81	1.30	0.36	0.69	0.25
Greece	0.34	0.51	0.58	0.66	0.44	0.65
Hungary	0.11	0.21	0.53	0.74	0.27	0.69
Ireland	0.81	0.70	1.07	1.15	0.77	1.06
Italy	0.77	0.64	1.21	0.14	0.57	0.23
Latvia	0.62	0.87	0.25	1.25	0.84	1.19
Lithuania	0.46	0.72	0.18	1.07	0.67	1.00
Luxembourg	1.06	1.29	0.63	1.75	1.24	1.62
Malta	1.43	1.62	1.03	2.14	1.63	2.03
Netherlands	0.46	0.42	0.88	0.30	0.28	0.24
Poland	0.46	0.39	0.90	0.29	0.31	0.25
Portugal	0.00	0.28	0.48	0.72	0.25	0.65
Romania	0.28	0.00	0.74	0.60	0.25	0.59
Slovenia	0.48	0.74	0.00	1.15	0.69	1.06
Spain	0.72	0.60	1.15	0.00	0.54	0.27
Sweden	0.25	0.25	0.69	0.54	0.00	0.45
United Kingdom	0.65	0.59	1.06	0.27	0.45	0.00

Table 3: Distance between countries in 2015, part 3

## 8 Δοκιμές για διαφορετικά βάρη στις παραμέτρους

Ξέρω πως δεν είπαμε να κάνω αυτό, αλλά όταν ξεκίνησα να το κάνω, δεν μπορούσα να σταματήσω τις δοκιμές. Αρχικά. Όλα όσα θα παρουσιαστούν παρακάτω είναι έχοντας ενεργοποιημένη μόνο την κανονικοποίηση και όχι την λογαρίθμηση, γιατί για αυτήν είχα απορία.

Για να κρίνω το πόσο σωστά είναι τα βάρη, διάλεξα να προσπαθώ να μεγιστοποιήσω το  $r^2$  της παλινδρόμησης. Εγιναν λοιπόν οι παρακάτω δοκιμές:

Population	GDP per Capita	Inflation	Agriculture GDP	Industry GDP	Manufacturing GDP	Total Energy Supply	Verified emissions	$r^2$
1	1	1	1	1	1	1	1	0.76
4	0	0	0	10	0	3	10	0.93
0	0	0	0	6.8	0	3.2	10	0.9496
100	0	0	50	626	0	400	1000	0.9483
0	0	0	60	570	0	360	981	0.953

Εδώ είναι πολύ εμφανής η τεράστια εξάρτηση από τα verified emissions, το οποίο είναι εντελώς αναμενόμενο. Στην επόμενη δοκιμή, τα verified emissions είχαν αυστηρά τιμή 0.

Population	GDP perCapita	Inflation	Agriculture GDP	Industry GDP	Manufacturing GDP	Total Energy Supply	Verified emissions	$r^2$
1	1	1	1	1	1	1	0	0.70
0	0	0	0	1000	50	500	0	0,913
0	0	0	500	10000	0	2700	0	0,92

Αντίστοιχα, εδώ φαίνεται μία πολύ μεγάλη εξάρτηση από την παράμετρο του ακαθάριστου προϊόντος η οποία αφορά στην βιομηχανία. Αν την αφαιρέσουμε και αυτήν από το παιχνίδι, προσπαθώντας να βρούμε ένα λίγο διαφορετικό mix βλέπουμε πως:

Population	GDP perCapita	Inflation	Agriculture GDP	Industry GDP	Manufacturing GDP	Total Energy Supply	Verified emissions	$r^2$
1	1	1	1	0	1	1	0	0,63
0.5	0.5	0.5	0.5	0	1	10	0	0.84
0	0	0	0	0	50	1000	0	0.86
0	0	0	0	0	4300	99500	0	0.8645

Η τελευταία δοκιμή για να δούμε αν τα δεδομένα έστω και λίγο βγάζουν κάποιο νόημα είναι αυτήν την οποία βγάζουμε και το total energy supply το οποίο κυριαρχεί ξανά:

Population	GDP perCapita	Inflation	Agriculture GDP	Industry GDP	Manufacturing GDP	Total Energy Supply	Verified emissions	$r^2$
1	1	1	1	0	1	0	0	0.549
200	0	0	0	0	100	0	0	0.767
6640	30	10	350	0	3720	0	0	0.768

## 9 Για τη 31η Γενάρη

Δεν ξέρω αν έχει νόημα αυτό το pdf να γίνει λίγο σαν ημερολόγιο, αλλά ελπίζω να μην είναι μεγάλο πρόβλημα. Την τελευταία φορά είπαμε:

1. Η σύγκριση να γίνει μόνο με μία χώρα.
2. Είπε ο κ. Φωτάκης δύο εικασίες. Πρώτη πως το allocation είναι το αποτέλεσμα ενός optimization αλγόριθμου. Δεύτερη, πως είναι ένα πολυκριτηριακό optimization. Αν αυτά ισχύουν θα ήταν έξυπνο / προφανές να βρούμε, συγκρίνοντάς το με άλλα αντίστοιχα, ποιο πρόβλημα προσπαθεί να λύσει. Στη συνέχεια θα είναι αιτιολογημένο και πιο λογικό να προσθέσουμε πάνω σε αυτό οποιοδήποτε επιπλέον constraint θέλουμε.
3. "Policy options to improve the effectiveness of the EU emissions trading system: A multi-criteria analysis, stefan clo 2013" - Είναι αρκετά γενικό αλλά ίσως είναι χρήσιμο για να δούμε την στοχοθεσία της ΕΕ.
4. Για να κρίνουμε μία παλινδρόμηση, καλό είναι να ελέγχονται 3 κριτήρια:  $r^2$ , MSE, MAE
5. Είναι λογικό ο πληθωρισμός για αυτά τα δεδομένα να παίζει ελάχιστο ρόλο.

Λοιπόν, το πρώτο βήμα είναι να βάλουμε μία χώρα με την οποία θα συγκρίνουμε όλες τις υπόλοιπες. Επομένως, πλέον δεν κοιτάμε αν για να γίνει η σύγκριση με την μεσαία χώρα πρέπει πρώτα να οριστεί η μεσαία χώρα. Αυτή

προκύπτει ως εξής: Δίνουμε σε κάθε χώρα πόντους ίσους με το άθροισμα των θέσεών της ως προς κάθε feature. Στη συνέχεια διαλέγουμε αυτή με τη μεσαία βαθμολογία. Κάναμε όμως και δοκιμές για άλλες χώρες για σιγουριά. Μία πρώτη εντύπωση είναι πως όσο πιο ακραία μικρή ή ακραία μεγάλη είναι η οικονομία της χώρας, τόσο πιο γραμμική είναι η σχέση της απόστασης.

2017	$r^2$
Hungary	0.75
Germany	0.9147
Greece	0.6694
Malta	0.8331

Στη συνέχεια, αν συγκρίνουμε όλες τις χώρες με την Ουγγαρία, τότε προκύπτει αυτό το διάγραμμα. Συγκριμένα, βλέπουμε:

- Στον x άξονα: Την απόσταση της εκάστοτε χώρας με την Ουγγαρία ως προς το feature set που έχουμε επιλέξει. (GDP per capita, inflation, Population, manufacturing, industry, verified emissions, agriculture, total energy supply)
- Στον y άξονα: Την απόσταση της εκάστοτε χώρας με την Ουγγαρία ως προς το σύνολο των δωρεάν αδειών που έλαβαν μέσα στο 2017 εταιρείες με έδρα την Ουγγαρία.
- Τα δεδομένα κανονικοποιήθηκαν στο  $[0,1]$  πριν τον υπολογισμό των αποστάσεων, αλλά δε λογαριθμίστηκαν.

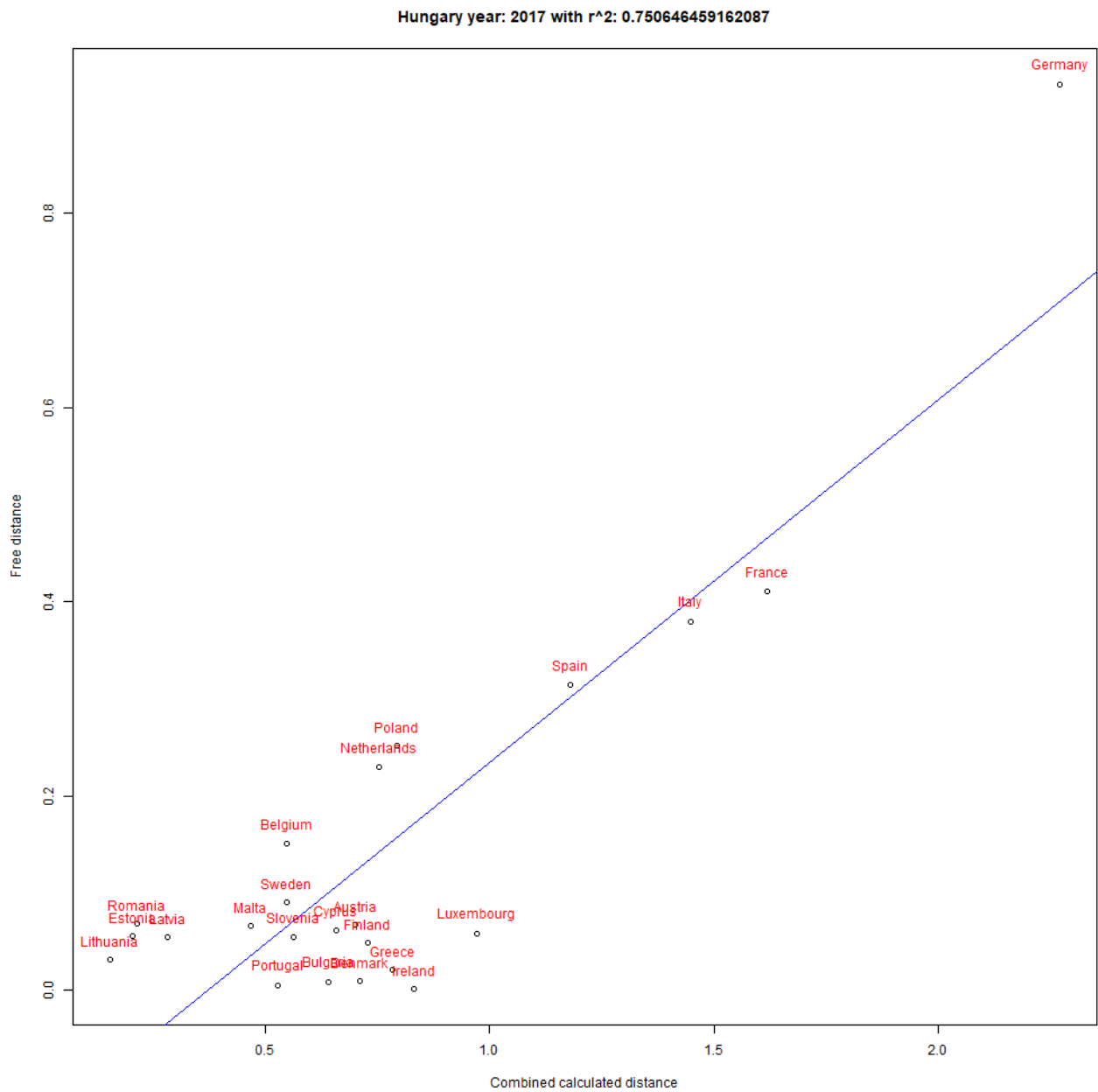
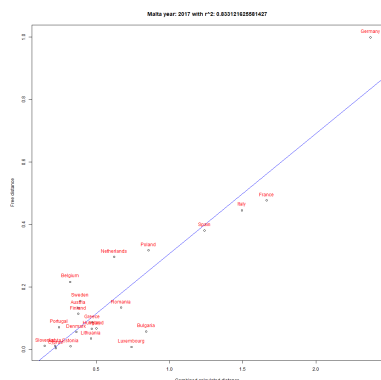


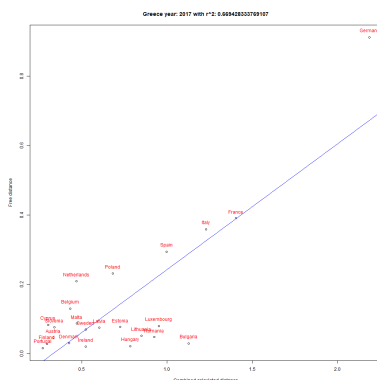
Figure 3: Hungary 2017



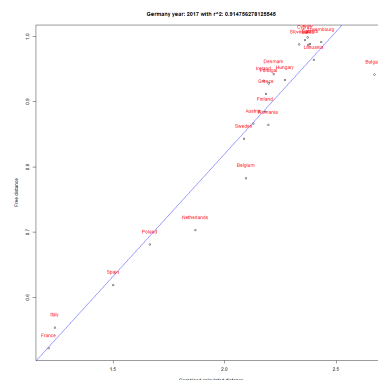
Το οποίο μπορούμε να το αντιπαράθεσουμε με τις 3 άλλες χώρες που είδαμε και πριν.



(a) Malta 2017



(b) Greece 2017



(c) Germany 2017

Παράλληλα μπορούμε να δούμε μέσα στις χρονιές επαναλαμβάνοντας το ίδιο πείραμα για όλες τις χώρες, τι θα προκύψει.

year	slope	mean r_squared	min_r_squared	max_r_squared
2008	0.34	0.65	0.00	0.79
2009	0.32	0.60	0.02	0.78
2010	0.33	0.58	0.04	0.79
2011	0.33	0.62	0.04	0.78
2012	0.34	0.64	0.07	0.82
2013	0.37	0.72	0.51	0.92
2014	0.36	0.71	0.46	0.93
2015	0.37	0.75	0.48	0.94
2016	0.35	0.70	0.50	0.91
2017	0.36	0.73	0.47	0.91
2018	0.37	0.74	0.43	0.92

Συμπεράσματα:

- Η κλίση αυτής της ευθείας δείχνει να παραμένει πολύ σταθερή
- Μέχρι το 2012 (δηλαδή και το τέλος της φάσης II) βλέπουμε πως υπάρχουν χώρες για τις οποίες δεν ισχύει το συμπέρασμα το οποίο είχαμε βγάλει, περί δικαιοσύνης του συστήματος.
- Μετά τη φάση III ακόμα υπάρχουν χώρες για τις οποίες η μέθοδος αυτή δε δίνει σημαντικά καλές τιμές.

Ένα ερώτημα που νομίζω πως είναι λογικό να προκύψει ξαφνικά είναι το "ποιο κριτήριο κάνει μία χώρα να είναι κατάλληλη για αυτή τη δουλειά; Πότε μπορούμε να διαλέξουμε μία χώρα για να είναι χρήσιμες οι αποστάσεις της;". Οπότε τα βάζουμε όλα μαζί στο ίδιο γράφημα και έχουμε:

Figure 5: 2017  $r^2$  vs GDP

Δοκιμάζοντας και άλλα χαρακτηριστικά δεν κατέληξα σε κάτι πιο ενδιαφέρον από το παραπάνω διάγραμμα. Η Γερμανία δείχνει να είναι πολύ καλή για να κάνουμε αυτή τη δουλειά, μάλλον γιατί στον "χώρο των αποστάσεων"

είναι τόσο μακριά από όλους τους άλλους που οι υπόλοιποι καταλήγουν να είναι συγκριτικά κοντά. Αυτό το λέω σαν διαισθητική παρατήρηση, δεν ξέρω αν στηρίζεται μαθηματικά.

Το επόμενο βήμα είναι να δούμε το προηγούμενο με τα βέλτιστα βάρη.

## 10 Για την 28η Φεβρουαρίου

Την τελευταία φορά προτείναμε να κάνουμε/δούμε/διαβάσουμε, χωρίς κάποια σειρά:

1. Να διαβάσω τον αλγόριθμο του benchmark, πώς προκύπτει ακριβώς.
2. Να εξετάσουμε αν τα αποτελέσματα σχετικά με το "ποιες χώρες εξηγούν καλύτερα τα δεδομένα είναι consistent μέσα στον χρόνο.
3. Υπάρχει κάποιο clustering μεταξύ των χωρών που εξηγούν καλά ή κακά τις άλλες χώρες; Έχουν κάποια κοινά χαρακτηριστικά όλες αυτές;
4. Γιατί είναι η Ολλανδία τόσο διαφορετική;
5. Να επαναληφθούν οι δοκιμές της προηγούμενης φοράς, αλλά για τις φάσεις 1,2 (πιθανώς θα προκύψει πρόβλημα με τα δεδομένα για αυτές τις περιόδους, αλλά θα το δούμε)
6. Κάνε δοκιμές και με μία fictional χώρα η οποία να έχει ως τιμές της μόνο τα medians όλων των χωρών.
7. Να γίνει καλύτερα η σύγκριση μεταξύ κάθε δύο παλινδρομήσεων, αξιολογώντας και το  $r^2$ , αλλά και άλλες παραμέτρους, όπως το p-value και MSE/MAE.
8. Τι συμβαίνει στα πρώτα χρόνια με την Πολωνία και την Γαλλία;
9. Να χρησιμοποιηθεί ggplot για τα διαγράμματα και να έχουν πάντα grid.
10. Για διάβασμα : A multi-criteria decision analysis model for carbon emission quota allocation in China's east coastal areas: Efficiency and Equity.

### 10.1 Αλλαγή 1

Το πρώτο βήμα που έγινε είναι το να ξαναγίνουν πολλά από τα πειράματα, μόνο που αυτή τη φορά, για να θεωρηθεί μία γραμμική παλινδρόμηση καλύτερη από μία άλλη, θα πρέπει να είναι αληθή τα παρακάτω κριτήρια:

- Έχει μεγαλύτερο  $r^2$ .
- Έχει p-value μικρότερο του 0.05.
- Το Mean Square Error δεν είναι πολύ μεγαλύτερο από την προηγούμενη μέγιστη τιμή (1.5 φορά).

### 10.2 Μεσαία χώρα ανά τα έτη

Εδώ βλέπουμε την πιο μεσαία χώρα κάθε χρονιάς και το πόσο καλά αποδίδει στο να εξηγεί τις άλλες χώρες. Οι τιμές στο p-value βγήκαν όντως 0, η μεγαλύτερη είχε την τιμή 0.00002581...

Year	R <sup>2</sup>	p-value	MSE	Country
2008	0.77	0.00	0.01	Denmark
2009	0.69	0.00	0.01	Denmark
2010	0.57	0.00	0.01	Greece
2011	0.65	0.00	0.02	Denmark
2012	0.75	0.00	0.01	Ireland
2013	0.71	0.00	0.01	Portugal
2014	0.58	0.00	0.02	Hungary
2015	0.77	0.00	0.01	Portugal
2016	0.70	0.00	0.01	Portugal
2017	0.75	0.00	0.01	Hungary
2018	0.74	0.00	0.01	Hungary

### 10.3 Όλες οι χώρες για όλα τα έτη

	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	max p-value	max MSE
Austria	0.70	0.62	0.66	0.68	0.71	0.74	0.73	0.82	0.72	0.68	0.74	0.00	0.02
Belgium	0.62	0.62	0.62	0.64	0.65	0.67	0.67	0.65	0.57	0.60	0.59	0.00	0.01
Bulgaria	0.75	0.67	0.73	0.67	0.78	0.84	0.85	0.86	0.69	0.81	0.85	0.00	0.01
Cyprus	0.79	0.71	0.69	0.71	0.78	0.67	0.58	0.73	0.72	0.75	0.78	0.00	0.02
Denmark	0.77	0.69	0.61	0.65	0.74	0.71	0.72	0.73	0.72	0.68	0.69	0.00	0.02
Estonia	0.79	0.74	0.67	0.69	0.66	0.64	0.68	0.78	0.68	0.82	0.81	0.00	0.02
Finland	0.71	0.64	0.66	0.73	0.68	0.73	0.77	0.80	0.72	0.69	0.76	0.00	0.01
France	0.08	0.08	0.20	0.19	0.25	0.78	0.79	0.78	0.78	0.73	0.76	0.19	0.01
Germany	0.76	0.75	0.79	0.76	0.80	0.92	0.93	0.94	0.91	0.91	0.92	0.00	0.01
Greece	0.62	0.57	0.57	0.60	0.57	0.51	0.46	0.69	0.65	0.67	0.71	0.00	0.02
Hungary	0.74	0.52	0.58	0.69	0.69	0.67	0.58	0.78	0.73	0.75	0.74	0.00	0.02
Ireland	0.70	0.76	0.47	0.70	0.75	0.73	0.69	0.77	0.73	0.63	0.64	0.00	0.02
Italy	0.63	0.61	0.65	0.64	0.60	0.86	0.82	0.75	0.74	0.72	0.78	0.00	0.01
Latvia	0.76	0.70	0.62	0.66	0.68	0.77	0.74	0.75	0.78	0.84	0.81	0.00	0.02
Lithuania	0.77	0.78	0.63	0.70	0.74	0.79	0.76	0.74	0.74	0.79	0.81	0.00	0.02
Luxembourg	0.79	0.67	0.59	0.77	0.73	0.68	0.65	0.75	0.60	0.74	0.70	0.00	0.02
Malta	0.76	0.65	0.62	0.68	0.78	0.75	0.72	0.82	0.71	0.83	0.82	0.00	0.02
Netherlands	0.51	0.53	0.53	0.44	0.54	0.60	0.60	0.51	0.50	0.47	0.43	0.00	0.01
Poland	0.00	0.02	0.04	0.04	0.07	0.60	0.60	0.48	0.54	0.69	0.69	0.92	0.02
Portugal	0.71	0.66	0.66	0.60	0.59	0.71	0.73	0.77	0.70	0.75	0.77	0.00	0.02
Romania	0.53	0.46	0.42	0.58	0.45	0.60	0.75	0.80	0.64	0.66	0.67	0.00	0.01
Slovenia	0.78	0.59	0.59	0.67	0.70	0.76	0.73	0.77	0.78	0.78	0.81	0.00	0.02
Spain	0.63	0.61	0.63	0.57	0.60	0.84	0.74	0.76	0.72	0.77	0.81	0.00	0.01
Sweden	0.70	0.56	0.61	0.59	0.67	0.74	0.75	0.81	0.75	0.77	0.77	0.00	0.02
United Kingdom	0.75	0.71	0.77	0.78	0.82	0.63	0.69	0.66	0.66	0.72	0.71	0.00	0.01

Το οποίο γραφικά φαίνεται και ως εξής:

**R<sup>2</sup> values for each country and year**

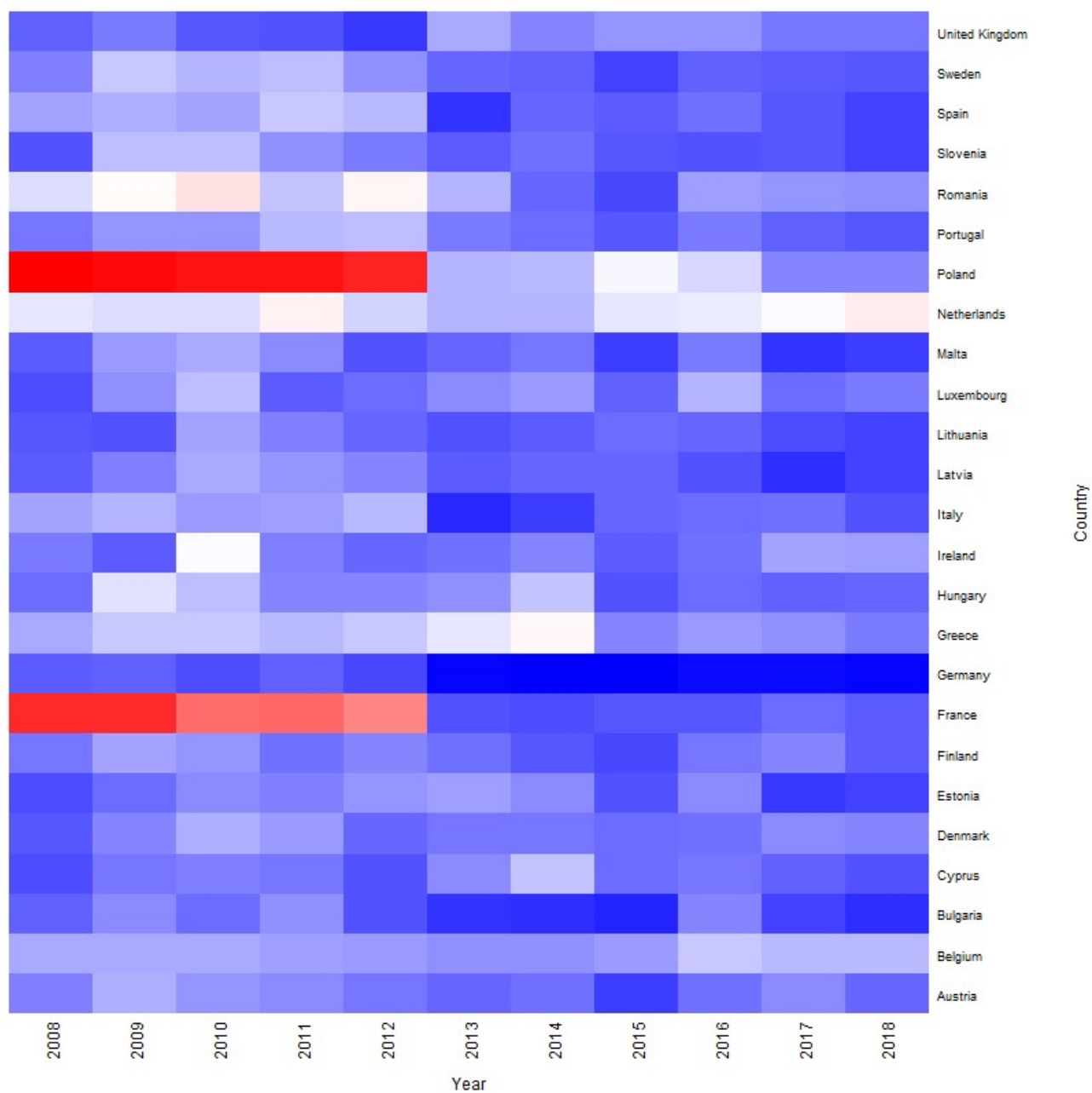


Figure 6:  $r^2$

Το πιο λογικό λοιπόν βήμα είναι το να ελέγξουμε το γιατί αυτές οι δύο χώρες αλλάζουν τόσο πολύ το 2013. Οπότε ας δούμε τα δεδομένα που χρησιμοποιούμε για αυτές:

	2012	2013	2012	2013
GEO	Poland	Poland	France	France
Total_energy_supply	0.3083121	0.3038433	0.8215192	0.8080548
GDPpc	0.1163326	0.1141371	0.3630366	0.3550416
Population	0.4732704	0.4716958	0.8164021	0.8183792
Inflation	0.57990986	0.07563843	0.28530815	0.19301719
Verified_emissions	0.4212666	0.4212431	0.2421093	0.2422275
Agriculture	0.3402106	0.3644831	1.0000000	0.8928814
Industry	0.1558893	0.1482026	0.4982965	0.5053396
Manufacturing	0.1166064	0.1091561	0.3911272	0.3913494

Όπως είναι εμφανές, στα δεδομένα των χωρών δεν υπάρχει κάτι προφανές το οποίο να δικαιολογεί αυτήν την τόσο απότομη αλλαγή. Τα δεδομένα τα βλέπουμε κανονικοποιημένα.

Επίσης, για το ποιες χώρες τα πηγαίνουν καλύτερα, μπορούμε να αφτιάξουμε αυτήν την κατάταξη, η οποία είναι αρκετά αντιπροσωπευτική.

	Country	Average $R^2$
1	Germany	0.85
2	Bulgaria	0.77
3	Lithuania	0.75
4	Malta	0.74
5	Latvia	0.74
6	Estonia	0.72
7	Slovenia	0.72
8	Cyprus	0.72
9	United Kingdom	0.72
10	Finland	0.72
11	Italy	0.71
12	Austria	0.71
13	Sweden	0.70
14	Denmark	0.70
15	Spain	0.70
16	Luxembourg	0.70
17	Portugal	0.70
18	Ireland	0.69
19	Hungary	0.68
20	Belgium	0.63
21	Greece	0.60
22	Romania	0.60
23	Netherlands	0.52
24	France	0.49
25	Poland	0.34

## 11 Για τη 14η Μαρτίου

Είπαμε να διορθωθούν τα παρακάτω:

- Οι πίνακες με δεδομένα να έχουν 2 δεκαδικά ψηφία ή όσα είναι αναγκαία, όχι παραπάνω.
- Να προστίθενται γραμμές στις αναφορές ανά 5 στοιχεία, ώστε να είναι ευδιάκριτα τα νούμερα.
- Να φτιαχτεί ένα πινακάκι με descriptive statistics για τα δεδομένα τα οποία έχουμε, να περιλαμβάνει min, max, quantiles, median και boxplot.
- Να γίνει η παραπάνω παρουσίαση για τα αρχικά δεδομένα, αλλά και για τα κανονικοποιημένα δεδομένα.
- Πιθανώς να δούμε την διαφορά που θα είχε η χρήση κάποιου διαφορετικού τρόπου κανονικοποίησης. (Μέχρι τώρα η κανονικοποίηση γίνεται διαιρώντας με την μέγιστη τιμή. Ένας άλλος τρόπος θα ήταν ο  $\frac{X - X_{min}}{X_{max} - X_{min}}$ )

### 11.1 Πίνακας $r^2$ με βέλτιστα επιλεγμένα βάρη

Εδώ φαίνεται ο πίνακας της προηγούμενης εβδομάδας, με την διαφορά ότι κάθε χώρα μπορούσε να επιλέξει το "mix" δεδομένων που θα χρησιμοποιήσει ώστε να είναι βέλτιστο το  $r^2$ . Εδώ βλέπουμε πως η Γαλλία έχει πολύ πιο μικρό πρόβλημα.

Ο αλγόριθμος ο οποίος χρησιμοποιήθηκε για να βγει αυτός ο πίνακας είναι πολύ κακός. Θεωρεί πως υπάρχει κάποια γραμμική συσχέτιση της ικανότητας μίας χώρας να περιγράψει τις άλλες και του βάρους κάθε διαφορετικού δεδομένου (για παράδειγμα, τα verified emissions έχουν μία γεννήτως αύξουσα σχέση με την ικανότητα της χώρας να περιγράψει τις άλλες.) Αυτό δεν ισχύει και ίσως είναι ενδιαφέρον το να δούμε πώς μοιάζει η συνάρτηση αυτή.

Πάντως ο αλγόριθμος αυτός κάνει τα παρακάτω:

1. Διαλέγει ένα από τα βάρη των διαφορετικών δεδομένων.
2. Δοκιμαστικά "ανεβάζει" και "κατεβάζει" το αντίστοιχο βάρος και βλέπει ποιο οδηγεί σε καλύτερη παλινδρόμηση.
3. Επαναλαμβάνει το προηγούμενο βήμα μέχρι να χτυπήσει κάποιο όριο ή να είναι χειρότερες και οι δύο επιλογές.
4. Προχωρά στο επόμενο βάρος το οποίο μπορεί να βελτιστοποιήσει.
5. Οι παράμετροι ήταν:
  - Ελάχιστο βάρος: 0
  - Μέγιστο βάρος: 1000
  - Βήμα: 10
  - Τιμή εκκίνησης: 50
  - Υπάρχουν 8 βάρη, τυχαία διαλέγει το βάρος το οποίο θα προσπαθήσει να βελτιστοποιήσει 100 φορές

	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	max p-value	max MSE
Austria	0.99	0.98	0.98	0.98	0.99	0.98	0.98	0.98	0.98	0.97	0.97	0.00	0.00
Belgium	0.98	0.98	0.98	0.99	0.98	0.91	0.90	0.91	0.89	0.88	0.89	0.00	0.00
Bulgaria	0.98	0.98	0.98	0.98	0.99	0.97	0.97	0.97	0.97	0.97	0.97	0.00	0.00
Cyprus	0.99	0.98	0.98	0.98	0.99	0.97	0.96	0.97	0.97	0.97	0.97	0.00	0.00
Denmark	0.99	0.98	0.97	0.98	0.99	0.97	0.97	0.97	0.97	0.97	0.96	0.00	0.00
Estonia	0.99	0.98	0.98	0.98	0.98	0.96	0.97	0.97	0.97	0.97	0.97	0.00	0.00
Finland	0.99	0.98	0.98	0.98	0.98	0.97	0.98	0.97	0.98	0.97	0.97	0.00	0.00
France	0.08	0.08	0.97	0.96	0.98	0.98	0.98	0.96	0.98	0.98	0.98	0.19	0.01
Germany	1.00	0.99	0.99	0.99	0.99	0.98	0.98	0.98	0.98	0.97	0.98	0.00	0.00
Greece	0.98	0.98	0.98	0.98	0.98	0.96	0.96	0.97	0.98	0.97	0.97	0.00	0.00
Hungary	0.99	0.97	0.97	0.98	0.98	0.97	0.96	0.97	0.97	0.97	0.97	0.00	0.00
Ireland	0.99	0.98	0.97	0.99	0.99	0.97	0.97	0.97	0.97	0.96	0.96	0.00	0.00
Italy	0.96	0.93	0.97	0.98	0.99	0.96	0.96	0.97	0.97	0.97	0.96	0.00	0.00
Latvia	0.98	0.98	0.97	0.98	0.98	0.97	0.97	0.97	0.98	0.98	0.97	0.00	0.00
Lithuania	0.99	0.98	0.97	0.98	0.99	0.98	0.98	0.98	0.98	0.98	0.97	0.00	0.00
Luxembourg	0.98	0.97	0.97	0.98	0.98	0.96	0.96	0.97	0.97	0.97	0.97	0.00	0.00
Malta	0.99	0.98	0.97	0.98	0.99	0.97	0.97	0.97	0.97	0.98	0.97	0.00	0.00
Netherlands	0.97	0.97	0.97	0.97	0.98	0.81	0.81	0.80	0.79	0.73	0.69	0.00	0.00
Poland	0.00	0.02	0.04	0.04	0.07	0.94	0.94	0.94	0.95	0.95	0.94	0.92	0.02
Portugal	0.99	0.98	0.98	0.98	0.98	0.97	0.98	0.97	0.98	0.97	0.97	0.00	0.00
Romania	0.95	0.95	0.93	0.96	0.96	0.98	0.98	0.98	0.97	0.97	0.97	0.00	0.00
Slovenia	0.99	0.97	0.97	0.98	0.98	0.97	0.97	0.97	0.97	0.97	0.97	0.00	0.00
Spain	0.98	0.97	0.98	0.97	0.98	0.95	0.94	0.95	0.95	0.95	0.95	0.00	0.00
Sweden	0.99	0.98	0.97	0.98	0.99	0.97	0.98	0.98	0.97	0.97	0.97	0.00	0.00
United Kingdom	0.99	0.99	0.99	0.99	0.99	0.92	0.94	0.97	0.98	0.98	0.98	0.00	0.00

Σε αυτόν τον πίνακα τα μέσα βάρη ήταν:

- Population: 59.63
- GDP per capita: 19.85
- Inflation: 15.12
- Agriculture: 69.09
- Industry: 407.60
- Manufacturing: 12.69
- Total energy supply: 170.25
- Verified emissions: 849.81



## 11.2 Οπτικοποίηση δεδομένων

### 11.2.1 Πληθυσμός

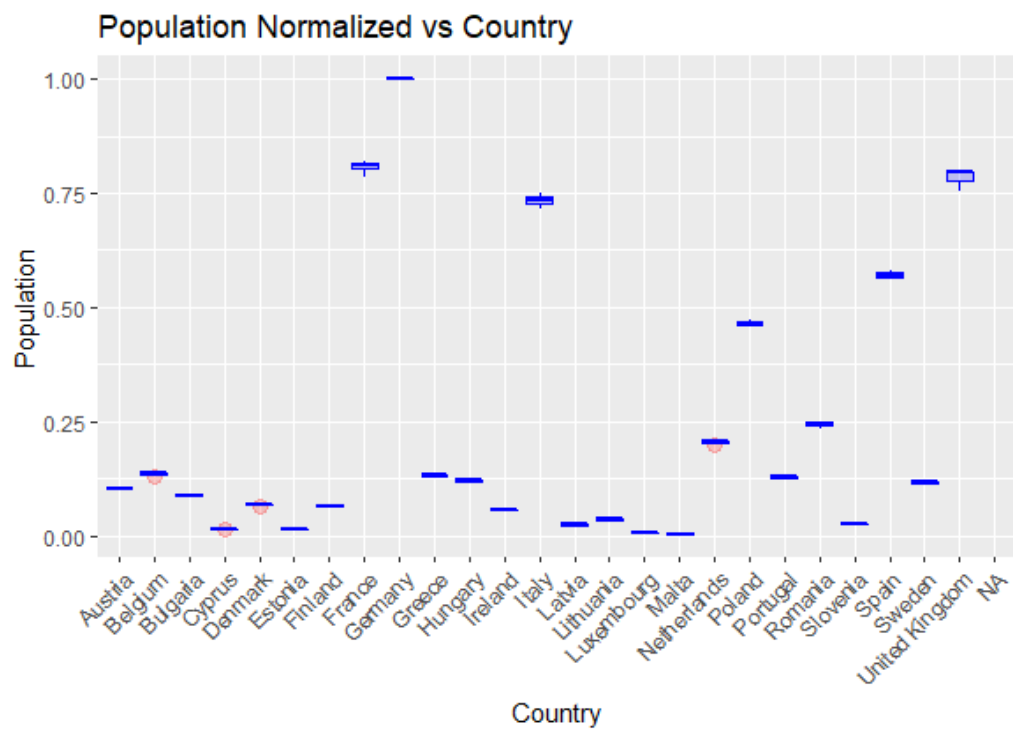
- Source: <https://data.worldbank.org/indicator/SP.POP.TOTL>
- Year: 2011 - 2021
- Unit: Persons

Για τον πληθυσμό δεν έχω τίποτα εντυπωσιακό να πω. Ίσως είναι κάπως εύκολο να δούμε όλα τα δεδομένα ταυτόχρονα και να παρατηρήσουμε τι αλλαγή επιφέρει η κανονικοποίηση στα δεδομένα. Όπως φαίνεται στα figure 7 και 8. Τα σχήματα είναι ολόιδια. Η διαφορά είναι πως το ένα έχει διαρεθεί με το 1000000 ενώ το άλλο έχει διαρεθεί με το μέγιστο κάθε χρονιάς, οπότε είναι ελαφρώς διαφορετικές οι σχέσεις μεταξύ των αριθμών κάθε χώρας.

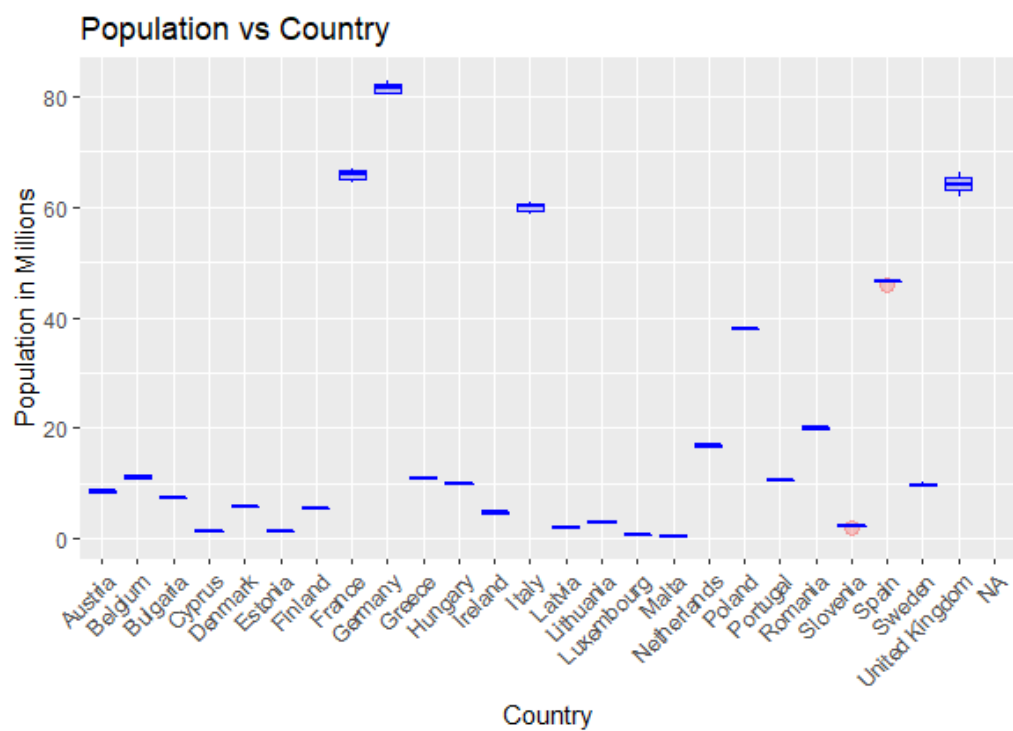
Population	min	25-quantile	median	75-quantile	max	Std
Austria	8.32	8.38	8.48	8.69	8.84	0.19
Belgium	10.71	10.97	11.16	11.30	11.43	0.24
Bulgaria	7.03	7.15	7.27	7.37	7.49	0.15
Cyprus	1.08	1.12	1.14	1.17	1.19	0.03
Denmark	5.49	5.56	5.61	5.71	5.79	0.10
Estonia	1.31	1.32	1.32	1.33	1.34	0.01
Finland	5.31	5.38	5.44	5.49	5.52	0.07
France	64.37	65.19	66.00	66.64	67.10	0.93
Germany	80.27	80.81	81.78	82.23	82.91	0.91
Greece	10.73	10.80	10.97	11.09	11.12	0.15
Hungary	9.78	9.83	9.89	9.99	10.04	0.09
Ireland	4.49	4.57	4.62	4.73	4.87	0.12
Italy	58.83	59.33	60.23	60.58	60.79	0.73
Latvia	1.93	1.97	2.01	2.08	2.18	0.08
Lithuania	2.80	2.89	2.96	3.06	3.20	0.13
Luxembourg	0.49	0.51	0.54	0.58	0.61	0.04
Malta	0.41	0.42	0.43	0.45	0.48	0.03
Netherlands	16.45	16.65	16.80	16.99	17.23	0.25
Poland	37.97	37.98	38.04	38.06	38.15	0.06
Portugal	10.28	10.34	10.46	10.56	10.57	0.12
Romania	19.47	19.76	19.98	20.20	20.54	0.33
Slovenia	2.02	2.05	2.06	2.06	2.07	0.01
Spain	45.95	46.46	46.58	46.68	46.80	0.24
Sweden	9.22	9.41	9.60	9.86	10.18	0.31
United Kingdom	61.81	63.01	64.13	65.36	66.46	1.55

Table 4: Population in millions 2008-2018

Παράλληλα, μπορούμε να παρατηρήσουμε τα δεδομένα κάπως καλύτερα, όπου φαίνεται πως η κανονικοποίηση δεν επηρεάζει καθόλου την σχέση μεταξύ των τιμών:



(a)



(b)

### 11.2.2 Πληθωρισμός

- Source: <https://data.worldbank.org/indicator/FP.CPI.TOTL.ZG>

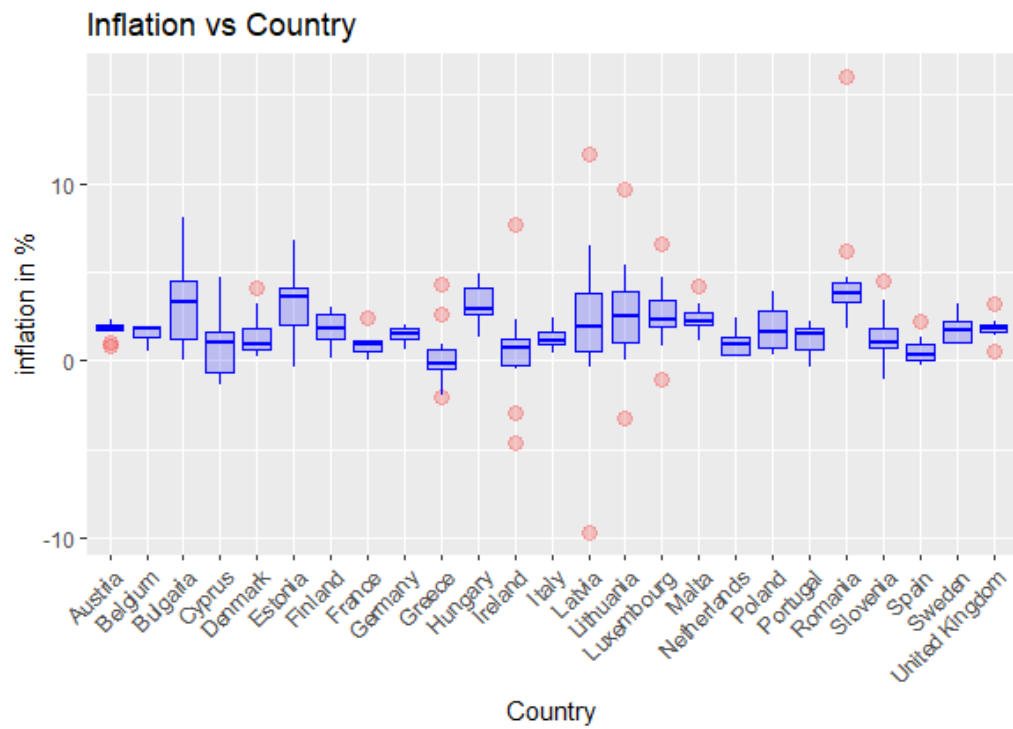
- Year: 1960-2021

- Unit: %

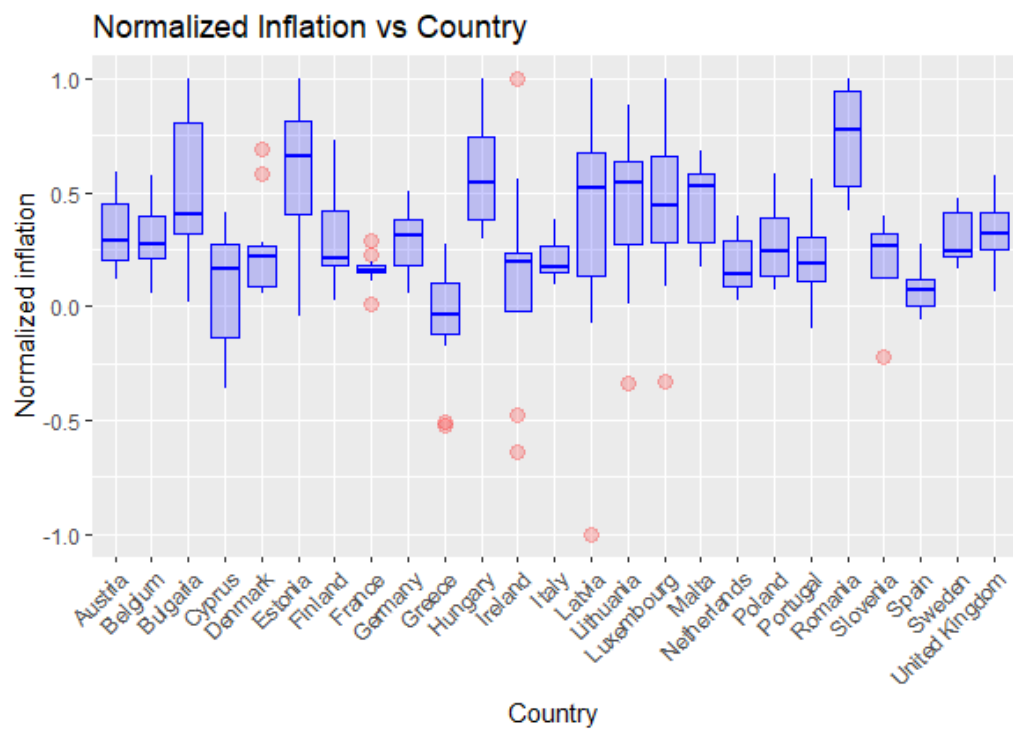
Inflation	min	25-quantile	median	75-quantile	max	Std
Austria	0.87	1.71	1.85	2.01	2.30	0.45
Belgium	0.53	1.30	1.81	1.90	1.96	0.47
Bulgaria	0.07	1.23	3.32	4.52	8.10	2.45
Cyprus	-1.33	-0.66	1.00	1.65	4.73	1.74
Denmark	0.25	0.58	0.89	1.78	4.13	1.27
Estonia	-0.39	2.00	3.59	4.05	6.80	2.02
Finland	0.09	1.22	1.77	2.59	3.04	1.02
France	0.07	0.55	0.95	1.10	2.37	0.58
Germany	0.65	1.20	1.50	1.87	1.97	0.46
Greece	-2.05	-0.44	-0.18	0.62	4.34	1.85
Hungary	1.32	2.66	2.89	4.11	4.85	1.15
Ireland	-4.62	-0.28	0.71	1.22	7.70	3.10
Italy	0.44	0.92	1.13	1.58	2.40	0.54
Latvia	-9.67	0.49	1.92	3.77	11.65	5.17
Lithuania	-3.30	1.06	2.53	3.88	9.71	3.30
Luxembourg	-1.11	1.94	2.28	3.36	6.61	1.99
Malta	1.12	2.06	2.22	2.75	4.22	0.82
Netherlands	0.19	0.35	0.94	1.36	2.44	0.79
Poland	0.30	0.75	1.65	2.82	3.89	1.34
Portugal	-0.39	0.67	1.51	1.78	2.25	0.90
Romania	1.80	3.33	3.77	4.38	16.02	3.88
Slovenia	-1.03	0.68	1.04	1.86	4.47	1.49
Spain	-0.22	0.06	0.32	0.90	2.25	0.76
Sweden	0.93	1.04	1.74	2.25	3.24	0.75
United Kingdom	0.51	1.60	1.82	2.03	3.23	0.65

Table 5: Πληθωρισμός μεταξύ 2008-2018

Εδώ βέβαια ίσως έχει σημασία να δούμε λίγο τι συμβαίνει στα δεδομένα με την κανονικοποίηση. Ούτε εδώ αλλάζουν σημαντικά οι σχέσεις μεταξύ τους.



(a)



(b)

### 11.2.3 Total energy Supply

- Source: [https://ec.europa.eu/eurostat/databrowser/view/nrg\\_bal\\_s\\_1/default/table?lang=en](https://ec.europa.eu/eurostat/databrowser/view/nrg_bal_s_1/default/table?lang=en)
- Year: 2011 - 2020
- Unit: Thousand tonnes of oil equivalent

	min	25-quantile	median	75-quantile	max	Std
Austria	32011.29	33031.17	33177.75	33562.30	34166.28	660.54
Belgium	52238.38	53015.79	55039.87	55272.03	59313.06	2255.37
Bulgaria	16923.38	17726.43	18234.79	18722.35	19823.90	813.92
Cyprus	1955.96	2121.56	2262.44	2440.09	2616.93	221.87
Denmark	16374.16	16838.22	17383.65	18304.45	19558.31	1128.72
Estonia	4399.62	5378.58	5648.48	5846.32	5978.41	496.84
Finland	32022.99	33165.71	33582.86	34469.06	36251.74	1195.84
France	248383.41	250072.60	256292.91	259845.13	266394.54	6206.33
Germany	305036.83	310746.75	313107.91	318940.66	335474.27	9403.81
Greece	22748.62	23240.71	23407.97	27231.17	30404.91	2854.85
Hungary	23652.51	24816.51	25609.63	26395.50	26900.70	1096.99
Ireland	12776.63	13293.33	13625.33	14225.86	15022.14	727.58
Italy	146769.88	152859.01	156093.49	168758.58	181736.20	10969.89
Latvia	4259.46	4307.46	4407.65	4465.42	4640.00	129.86
Lithuania	6946.51	7067.95	7290.53	7647.47	9553.50	819.46
Luxembourg	3682.05	3785.50	3948.40	4127.44	4212.94	195.48
Malta	594.33	687.68	780.25	834.21	881.48	97.20
Netherlands	71379.09	73882.75	75619.51	77123.41	82743.78	3132.34
Poland	93773.11	96236.75	97971.41	101131.11	108970.23	4555.35
Portugal	21439.29	22060.46	22651.14	23433.93	24716.20	1092.37
Romania	31378.66	31668.07	33454.78	34845.41	39485.27	2432.37
Slovenia	6473.51	6722.90	6875.85	7127.16	7982.69	419.35
Spain	114522.76	119335.11	125486.61	126434.02	138166.04	6411.76
Sweden	44092.81	48432.69	49103.98	49874.22	50118.89	1857.82
United Kingdom	174024.39	177000.66	186386.17	191789.70	208268.88	11384.30

Table 6: Total energy supply μεταξύ 2008-2018

### 11.2.4 GDP per capita

- Source: <https://data.worldbank.org/indicator/NY.GDP.PCAP.CD>
- Year: 1960 - 2021
- Unit: US\$

	min	25-quantile	median	75-quantile	max	Std
Austria	44.20	47.17	48.56	51.46	51.92	2.75
Belgium	41.01	44.19	44.76	47.48	48.30	2.43
Bulgaria	6.85	7.17	7.57	7.88	9.45	0.74
Cyprus	23.41	26.89	28.91	31.57	35.40	3.57
Denmark	53.25	57.83	58.51	61.67	64.32	3.39
Estonia	14.66	17.40	18.20	19.66	23.06	2.45
Finland	42.80	46.46	47.71	50.16	53.77	3.25
France	36.65	39.73	41.59	42.84	45.52	2.75
Germany	41.10	41.89	44.65	46.50	48.02	2.61
Greece	17.92	19.17	21.79	26.10	32.13	4.84
Hungary	12.72	13.09	13.72	14.46	16.43	1.21
Ireland	48.66	51.83	55.60	62.44	79.11	9.57
Italy	30.24	33.51	35.56	36.63	40.94	3.17
Latvia	11.42	13.56	14.33	15.72	17.87	1.87
Lithuania	11.82	14.32	14.94	16.14	19.19	2.11
Luxembourg	105.46	109.81	112.58	119.51	123.68	6.14
Malta	21.08	22.37	24.77	26.19	31.57	3.23
Netherlands	45.19	49.37	52.20	52.97	57.88	3.66
Poland	11.53	12.60	13.70	13.94	15.47	1.08
Portugal	19.25	21.03	22.10	23.18	24.95	1.68
Romania	8.21	8.76	9.55	10.24	12.40	1.23
Slovenia	20.89	23.07	23.53	24.96	27.60	1.92
Spain	25.74	28.25	29.50	31.11	35.51	2.74
Sweden	46.95	52.42	54.59	59.03	61.13	4.45
United Kingdom	38.95	41.18	42.69	44.56	47.79	2.92

Table 7: GDP per capita in thousands USD

#### 11.2.5 Verified emissions

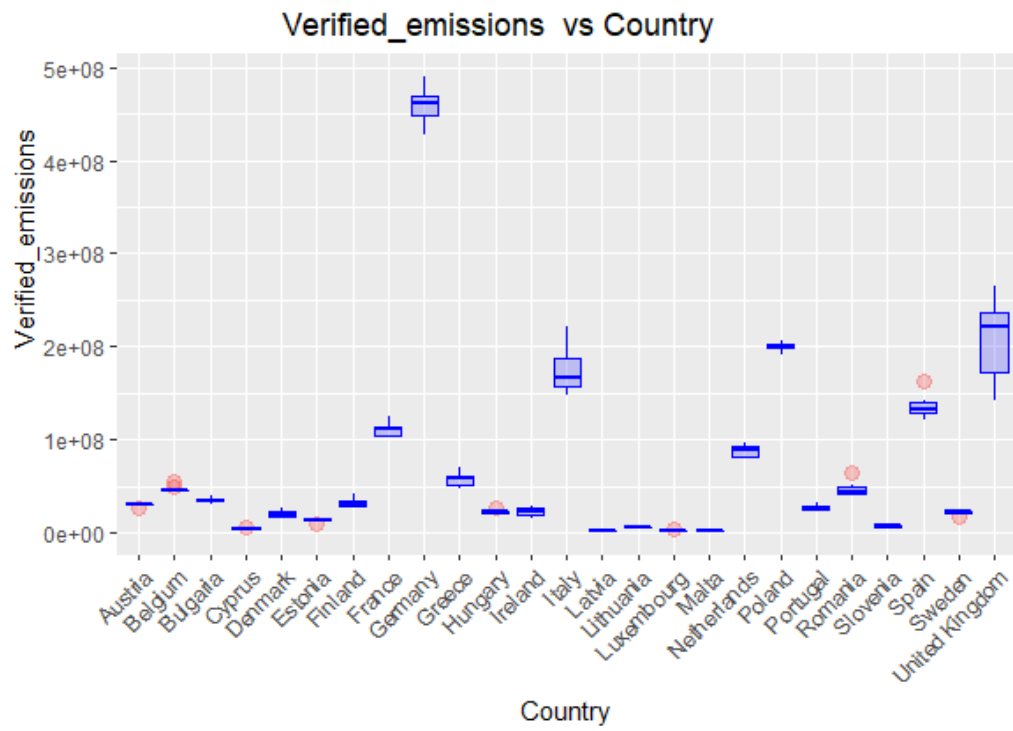
- Source: EU ETS Database
- Table: eutl.compliance
- Column: verified

Από όποια μονάδα και να ήταν, εδώ πάνε 1.000.000 φορές κάτω:

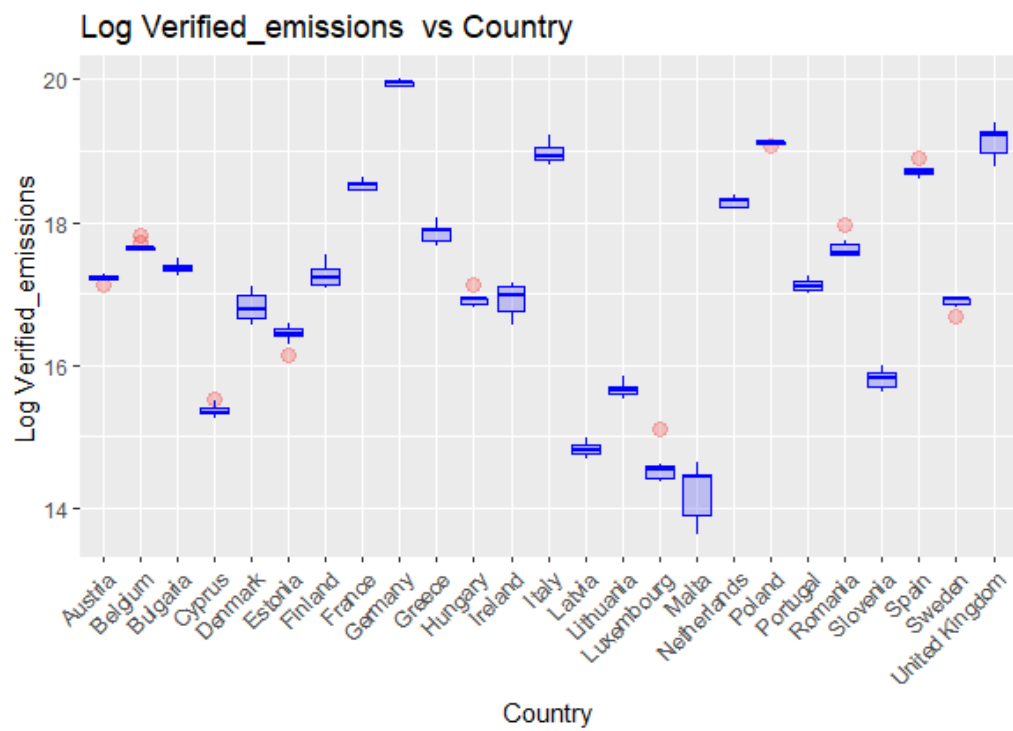
	min	25-quantile	median	75-quantile	max	Std
Austria	27.36	29.61	30.51	30.87	32.08	1.27
Belgium	45.03	45.17	45.99	46.29	55.46	3.16
Bulgaria	31.28	33.24	34.57	35.95	40.00	2.63
Cyprus	4.19	4.58	4.61	4.92	5.58	0.41
Denmark	15.50	17.05	19.46	23.71	26.55	4.05
Estonia	10.38	13.50	13.89	14.76	16.00	1.55
Finland	26.18	27.81	30.68	34.72	41.30	4.74
France	101.40	104.86	110.90	114.46	124.13	7.10
Germany	428.29	448.66	462.35	469.31	489.86	18.24
Greece	47.34	50.74	58.84	61.09	69.85	7.10
Hungary	20.08	21.23	22.40	22.61	27.24	1.91
Ireland	15.77	18.88	23.63	27.08	28.53	4.72
Italy	148.37	157.09	166.78	187.42	220.68	21.98
Latvia	2.43	2.59	2.74	2.96	3.24	0.26
Lithuania	5.61	5.92	6.23	6.65	7.56	0.59
Luxembourg	1.73	1.83	2.06	2.17	3.62	0.52
Malta	0.84	1.09	1.89	1.92	2.28	0.51
Netherlands	79.97	82.27	89.14	92.79	96.47	6.31
Poland	191.17	198.28	199.73	203.07	206.35	4.12
Portugal	24.17	25.64	26.99	28.75	31.42	2.23
Romania	40.53	42.21	43.07	48.72	63.82	6.83
Slovenia	6.18	6.60	7.45	8.03	8.86	0.91
Spain	121.48	128.26	132.69	140.52	163.46	11.31
Sweden	17.49	21.05	22.51	22.63	22.86	1.71
United Kingdom	141.76	172.90	220.88	236.57	265.06	42.22

Table 8: Verified emissions

Εδώ επειδή δεν φαίνεται τίποτα, θα βάλω log απλά για να μπορούμε να διακρίνουμε κάτι, χωρίς κάποια κανονικοποίηση.



(a)



(b)



### 11.2.6 Agriculture

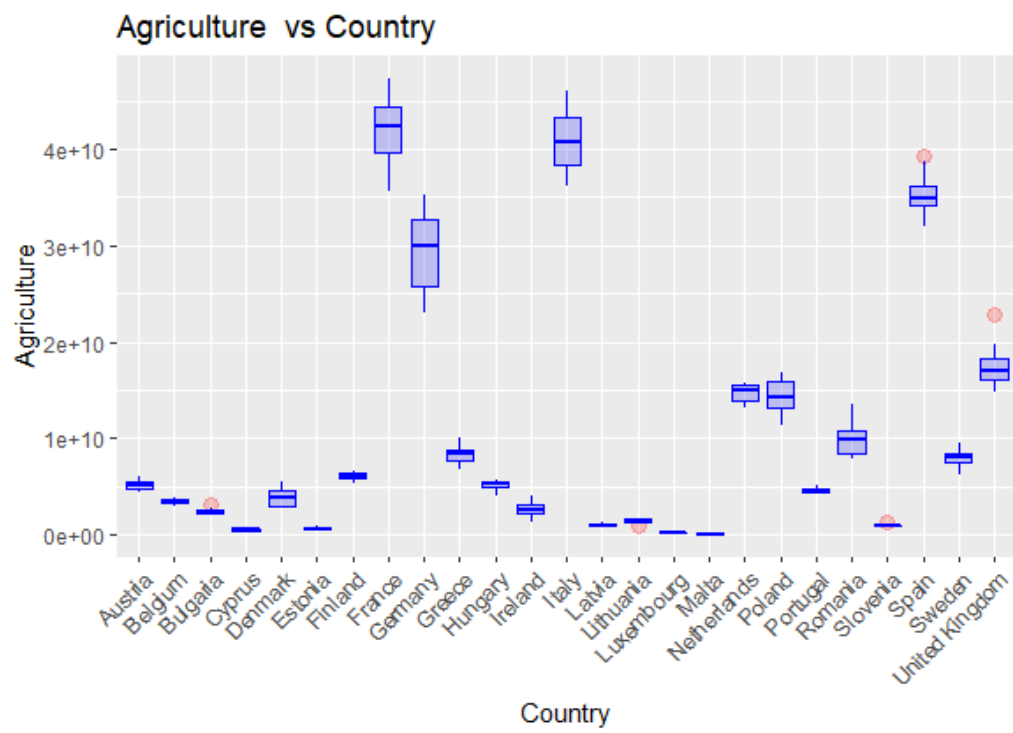
- Source: <http://wdi.worldbank.org/table/4.2>

- Year: 2020

- Unit: Agricultural in Billions USD.

	min	25-quantile	median	75-quantile	max	Std
Austria	4.32	4.79	5.14	5.46	6.05	0.55
Belgium	2.99	3.22	3.39	3.62	3.88	0.27
Bulgaria	2.04	2.20	2.39	2.56	3.20	0.33
Cyprus	0.37	0.44	0.49	0.52	0.59	0.07
Denmark	2.68	3.02	3.87	4.49	5.41	0.91
Estonia	0.49	0.63	0.65	0.80	0.85	0.13
Finland	5.29	5.82	6.03	6.41	6.59	0.43
France	35.54	39.60	42.43	44.42	47.28	3.91
Germany	22.99	25.78	29.91	32.78	35.20	4.54
Greece	6.79	7.71	8.33	8.85	9.99	0.90
Hungary	4.01	4.87	5.31	5.51	5.67	0.60
Ireland	1.32	2.28	2.61	3.06	3.96	0.71
Italy	36.20	38.32	40.70	43.21	45.94	3.17
Latvia	0.85	0.96	0.97	1.07	1.24	0.11
Lithuania	0.95	1.38	1.55	1.66	1.71	0.24
Luxembourg	0.13	0.14	0.15	0.17	0.20	0.02
Malta	0.09	0.10	0.11	0.11	0.13	0.01
Netherlands	13.20	13.89	15.07	15.49	15.71	0.92
Poland	11.25	13.13	14.22	15.96	16.76	1.91
Portugal	4.16	4.38	4.64	4.77	5.17	0.32
Romania	7.87	8.47	9.88	10.88	13.51	1.74
Slovenia	0.89	0.91	0.92	1.01	1.22	0.10
Spain	31.90	34.18	34.84	36.07	39.19	2.23
Sweden	6.27	7.53	8.11	8.47	9.59	0.87
United Kingdom	14.74	16.11	17.05	18.33	22.84	2.28

Table 9: Agriculture



(a)

### 11.2.7 Industry

- Source: <http://wdi.worldbank.org/table/4.2>
- Year: 2020
- Unit: Industry in Billions USD

	min	25-quantile	median	75-quantile	max	Std
Austria	96.16	102.24	106.03	111.06	116.48	6.50
Belgium	90.94	98.13	100.59	104.53	111.72	6.42
Bulgaria	12.04	13.16	13.62	14.01	14.87	0.87
Cyprus	2.03	2.45	2.95	3.63	4.98	0.90
Denmark	60.50	64.40	68.89	69.76	79.84	5.40
Estonia	4.61	5.63	5.97	6.42	7.36	0.81
Finland	54.66	61.38	63.49	65.70	84.47	7.78
France	431.14	459.74	479.72	506.08	551.30	36.97
Germany	843.80	934.68	1000.01	1014.25	1085.27	69.42
Greece	27.44	28.64	35.83	42.48	55.73	9.66
Hungary	32.23	32.95	33.99	36.13	40.79	2.98
Ireland	51.79	57.18	62.99	110.85	141.75	32.57
Italy	383.06	431.67	449.66	474.26	568.48	50.60
Latvia	4.90	5.37	5.81	6.01	7.81	0.78
Lithuania	9.38	11.12	12.22	13.05	13.97	1.52
Luxembourg	5.99	6.42	6.93	7.48	7.80	0.62
Malta	1.36	1.48	1.54	1.59	1.84	0.14
Netherlands	138.29	155.19	166.89	173.29	204.75	19.28
Poland	129.76	146.41	150.02	156.35	169.32	10.65
Portugal	38.81	41.82	43.54	47.46	53.70	4.42
Romania	54.81	61.02	63.42	67.76	78.81	7.80
Slovenia	12.08	12.77	13.77	13.95	16.62	1.31
Spain	240.11	268.98	278.66	328.16	429.02	57.54
Sweden	97.80	114.87	122.21	126.94	135.67	10.52
United Kingdom	448.57	473.78	494.10	522.75	583.44	40.55

Table 10: Industry

### 11.2.8 Manufacturing

- Source: <http://wdi.worldbank.org/table/4.2>
- Year: 2020
- Unit: Manufacturing in Billions USD.

Εδώ φαίνεται εμφανώς πως έχω κάνει ένα τεράστιο λάθος. Λείπουν οι τιμές για το Manufacturing στην Βουλγαρία και νόμιζα πως το είχα βάλει να παίρνει έναν μέσο όρο και πως αυτό δεν θα άλλαζε σημαντικά τα αποτελέσματα. Όμως είναι εμφανές πως η Βουλγαρία όχι μόνο δεν έπαιρνε τις τιμές που ήθελα, αλλά και να συνέβαινε αυτό είναι πολύ πιθανό και πάλι να μην ήταν αρκετά λογικό το αποτέλεσμα. Η Βουλγαρία μάλλον θα πρέπει να αφαιρεθεί από τις δοκιμές, μαζί με χώρες όπως η Σλοβακία.

	min	25-quantile	median	75-quantile	max	Std
Austria	63.75	66.61	70.28	72.47	76.57	4.24
Belgium	58.69	62.46	64.21	66.70	72.39	4.02
Bulgaria	-470.74	-453.31	-432.37	-420.07	-399.33	22.79
Cyprus	0.85	0.95	1.12	1.33	1.52	0.23
Denmark	35.21	37.47	40.47	41.64	46.70	3.45
Estonia	2.41	3.20	3.35	3.56	4.13	0.47
Finland	34.44	38.11	39.73	42.29	59.40	6.76
France	254.30	268.11	278.31	292.38	325.40	20.87
Germany	603.23	697.07	743.97	755.59	796.43	56.32
Greece	15.64	16.78	18.19	22.95	30.27	4.63
Hungary	22.45	24.56	25.39	27.39	29.71	2.23
Ireland	43.26	47.00	49.10	100.06	126.39	31.76
Italy	264.39	290.81	301.69	308.97	372.63	28.24
Latvia	2.57	2.85	3.20	3.32	3.63	0.32
Lithuania	5.65	7.20	8.00	8.15	8.91	0.96
Luxembourg	2.53	2.88	3.14	3.41	3.93	0.42
Malta	0.83	0.96	1.00	1.06	1.23	0.11
Netherlands	82.70	89.39	91.54	95.12	109.07	7.43
Poland	71.81	82.05	85.70	88.28	98.64	7.14
Portugal	24.20	25.62	27.24	27.71	31.35	2.12
Romania	35.95	37.70	39.72	44.53	49.91	4.70
Slovenia	8.43	8.69	9.28	9.90	11.00	0.87
Spain	135.09	148.07	155.06	166.17	207.17	19.84
Sweden	60.02	69.65	72.99	77.49	83.80	6.77
United Kingdom	218.33	242.14	251.42	268.16	286.56	20.19

Table 11: Manufacturing

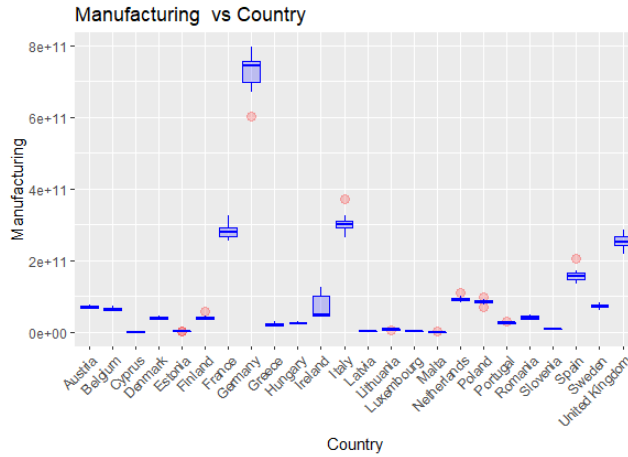


Figure 11

## 12 Για την 21η Μαρτίου

Είχαμε πει, να βάλουμε κάποια πρόταση για δουλειά, ώστε να αρχίσει να ετοιμάζεται το paper. Αυτά που έπρεπε να κάνω εγώ είναι:

- Να δούμε αν τα free ήταν το πρόβλημα για την πολωνία και την Γαλλία για το 2012 και το 2013.
- Να βρεθούν βάρη τα οποία να λειτουργούν για όλους καλά. Χωρίς να αλλάζουν κάθε φορά
- Διάβασμα για Nass Welfare
- Να αναφερθούν ελλείψεις στα data

### 12.1 Γιατί η Πολωνία και η Γαλλία αποτυγχάνουν να εξηγήσουν τις άλλες πριν το 2012;

Τελικά η απάντηση ήταν απίστευτα απλή. Κάτι πάει λάθος με τις δωρεάν άδειες, ή απλά έπαιρναν υπέρβολικά πολλές στην προηγούμενη φάση. όλες οι χώρες έχασαν πολλές δωρεάν άδειες από το 2012 στο 2013. Αλλά ας δούμε τι ποσοστό έχασαν και όλα θα βγάλουν νόημα. Spoiler, δεν βγάζει. Όχι μόνο δεν είναι παό τις πιο "ακραίες", αλλά μάλιστα είναι και σχετικά στην μέση.:

Country	2012 free in Millions	2013 free in Millions	Percentage drop in %
Malta	2.37	0.18	-92.52
Cyprus	6.70	1.15	-82.89
Estonia	14.30	3.11	-78.25
Greece	65.92	14.50	-78.01
Bulgaria	43.08	10.68	-75.22
United Kingdom	283.33	72.30	-74.48
Slovenia	8.31	2.32	-72.04
<b>Poland</b>	<b>213.47</b>	<b>63.63</b>	<b>-70.20</b>
Luxembourg	4.80	1.45	-69.79
Romania	75.61	24.06	-68.18
Portugal	35.04	12.37	-64.69
Ireland	28.76	10.54	-63.36
Germany	467.32	173.23	-62.93
Spain	163.59	67.11	-58.98
Italy	197.62	87.81	-55.56
Hungary	25.79	12.24	-52.54
Denmark	25.17	12.65	-49.72
Netherlands	99.37	50.24	-49.44
Latvia	5.30	2.77	-47.83
<b>France</b>	<b>159.96</b>	<b>84.38</b>	<b>-47.25</b>
Finland	40.28	23.05	-42.77
Belgium	61.61	37.56	-39.04
Austria	35.38	22.88	-35.33
Lithuania	8.43	6.55	-22.33
Sweden	25.72	30.21	17.47

Άρα καμία απάντηση ακόμα...

## 12.2 Τα βέλτιστα βάρη για όλα τα σενάρια

Το πρώτο βήμα είναι λογικά να δοκιμαστούν τα πρώτα τα βάρη που προέκυψαν ως μέσος όρος όλων των βέλτιστων βαρών. Σε δεύτερο βήμα, πιθανότατα έχει νόημα να δοκιμάσουμε να μεγιστοποιήσουμε το γινόμενο όλων.

- Population: 59.63
- GDP per capita: 19.85
- Inflation: 15.12
- Agriculture: 69.09
- Industry: 407.60
- Manufacturing: 12.69
- Total energy supply: 170.25
- Verified emissions: 849.81

	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	max p-value	max MSE
Austria	0.94	0.93	0.94	0.94	0.95	0.97	0.97	0.97	0.96	0.95	0.95	0.00	0.00
Belgium	0.91	0.90	0.93	0.93	0.93	0.86	0.83	0.82	0.80	0.78	0.78	0.00	0.00
Bulgaria	0.92	0.91	0.93	0.92	0.94	0.96	0.96	0.97	0.95	0.96	0.96	0.00	0.00
Cyprus	0.95	0.94	0.94	0.95	0.96	0.95	0.94	0.96	0.96	0.96	0.96	0.00	0.00
Denmark	0.95	0.94	0.93	0.94	0.96	0.96	0.96	0.96	0.96	0.95	0.95	0.00	0.00
Estonia	0.94	0.94	0.94	0.94	0.95	0.94	0.95	0.96	0.95	0.96	0.96	0.00	0.00
Finland	0.94	0.93	0.94	0.94	0.95	0.96	0.97	0.96	0.97	0.96	0.95	0.00	0.00
France	0.58	0.53	0.73	0.77	0.83	0.90	0.86	0.85	0.85	0.90	0.89	0.00	0.01
Germany	0.92	0.92	0.93	0.94	0.95	0.97	0.96	0.95	0.95	0.94	0.93	0.00	0.00
Greece	0.87	0.88	0.89	0.89	0.91	0.94	0.94	0.96	0.97	0.96	0.96	0.00	0.00
Hungary	0.94	0.92	0.93	0.94	0.95	0.96	0.95	0.97	0.96	0.96	0.96	0.00	0.00
Ireland	0.95	0.94	0.93	0.95	0.95	0.96	0.96	0.95	0.95	0.94	0.94	0.00	0.00
Italy	0.83	0.80	0.87	0.91	0.93	0.88	0.90	0.94	0.94	0.94	0.93	0.00	0.00
Latvia	0.94	0.94	0.94	0.94	0.95	0.96	0.96	0.96	0.96	0.97	0.96	0.00	0.00
Lithuania	0.94	0.94	0.94	0.94	0.95	0.97	0.97	0.97	0.97	0.97	0.96	0.00	0.00
Luxembourg	0.94	0.93	0.93	0.94	0.95	0.94	0.94	0.95	0.94	0.95	0.95	0.00	0.00
Malta	0.94	0.93	0.94	0.94	0.96	0.96	0.96	0.96	0.96	0.97	0.96	0.00	0.00
Netherlands	0.88	0.87	0.88	0.86	0.88	0.76	0.74	0.71	0.72	0.67	0.62	0.00	0.01
Poland	0.19	0.21	0.28	0.35	0.45	0.81	0.86	0.89	0.93	0.93	0.92	0.04	0.01
Portugal	0.94	0.93	0.94	0.94	0.95	0.96	0.97	0.97	0.96	0.96	0.96	0.00	0.00
Romania	0.84	0.82	0.83	0.86	0.87	0.96	0.98	0.98	0.96	0.95	0.96	0.00	0.00
Slovenia	0.95	0.93	0.93	0.94	0.95	0.96	0.96	0.96	0.96	0.96	0.96	0.00	0.00
Spain	0.93	0.92	0.89	0.88	0.90	0.80	0.87	0.90	0.88	0.92	0.90	0.00	0.00
Sweden	0.94	0.93	0.93	0.93	0.94	0.94	0.94	0.94	0.94	0.93	0.93	0.00	0.00
United Kingdom	0.92	0.94	0.95	0.94	0.94	0.84	0.89	0.89	0.96	0.94	0.94	0.00	0.00

Εδώ πλέον γίνεται πολύ αμφισβητήσιμο το αν χρειάζεται να ψάξουμε για κάτι καλύτερο. Με εξαίρεση την Πολωνία και την Γαλλία, οι υπόλοιποι δείχνουν να τα πηγαίνουν πολύ καλά.

## 13 Για την 28η Μαρτίου

Είπαμε να γίνουν τα παρακάτω:

- Να δούμε αν η Πολωνία είναι τόσο ανώμαλη, εξαιτίας του κανόνα 10c.
- Να προστεθούν στα features ξεχωριστά η πράσινη ενέργεια και ξεχωριστά η υπόλοιπη ενέργεια.
- Να οριστεί και να προσομοιωθεί το παρακάτω πρόβλημα:

$$- \max \sum u_i(x_i) * GDP_i$$

$$- \sum x_i \leq \text{Cap.}$$

- Όπου  $x_i$  είναι το allocation στην χώρα i.

- Η  $u_i$  είναι ξεχωριστή για κάθε χώρα και δηλώνει την ικανοποίηση που λαμβάνει η αντίστοιχη χώρα με το αντίστοιχο allocation.

- Πιθανότατα η  $u_i$  θα μπορούσε να είναι το energy intensity.
- Μετά μπορούν να προστεθούν επιπλέον constraints
- Να προστεθεί το energy Intensity στα features της κάθε χώρας.

### 13.1 Τι όντως έκανα

- Δεν υλοποίησα κανένα κομμάτι του κώδικά.
- Το power intensity δεν είναι απλώς μία τιμή.

### 13.2 PPS (Purchasing Power Standards)

Το PPS είναι ένα εικονικό νόμισμα. Το οποίο προσπαθεί να πετύχει να είναι αποπληθωρισμένο, ενώ ταυτόχρονα προσπαθεί να αφαιρέσει διαφορές λόγω άλλης αγοραστικής δύναμης.

#### 13.2.1 Χρήση PPS

Το PPS χρησιμοποιείται προκειμένου να κάνουμε συγκρίσεις μεταξύ διαφορετικών χωρών. Το βρίσκουμε συνήθως στα dataset της eurostat ως MPPS (Million Purchasing Power Standarts).

#### 13.2.2 Υπολογισμός

Ένα PPS ισοδυναμεί με ένα καλάθι στην εκάστοτε χώρα. Το καλάθι αυτό έχει ένα μείγμα από διάφορα προϊόντα και υπηρεσίες. Το κόστος αυτών συμπεριλαμβάνεται σε έναν σταθμισμένο μέσο, εκ του οποίου προκύπτει μία τιμή, η οποία είναι το κόστος του καλαθιού. Στην συνέχεια, μπορούμε να χρησιμοποιήσουμε αυτήν την τιμή για να κανονικοποιήσουμε διαφορετικές "αξίες" μεταξύ διαφορετικών χωρών ώστε να μπορούμε να κάνουμε συγκρίσεις με βάση την αγοραστική δύναμη και όχι απλά ευρώ.

### 13.3 Power Intensity

Energy intensity can be considered as an approximation of the energy efficiency of a country's economy, and shows the amount of energy needed to produce a unit of GDP. There are various reasons for observed improvements in energy intensity: a general shift from industry towards a service-based economy in Europe, a shift within industry to less energy-intensive activities and production methods, the closure of inefficient units, and more energy-efficient appliances.

To Power Intensity υπολογίζεται ως

$$\text{Power Intensity} = \frac{\text{Units of Energy}}{\text{Units of GDP}}$$

Αυτό λοιπόν μπορούμε να το υπολογίσουμε με δύο διαφορετικούς τρόπους.

- Chain Linked Volumes. Χρησιμοποιείται για να κάνουμε συγκρίσεις για μία χώρα σε βάθος χρόνου. Εδώ είναι σημαντικό το να σημειωθεί πως είναι πιθανό να μην αρκεί μία τιμή για να κάνουμε την δουλειά μας. Οι τιμές προκύπτουν σε επίπεδο έτους. Όμως οι οικονομικές αλλαγές στις χώρες (όπως αλλαγές σε φορολογίες) δημιουργούν μεγάλες ανωμαλίες στα δεδομένα, επομένως είναι καλύτερο να χρησιμοποιούνται περισσότερα από 4 data points.
- Η άλλη επιλογή είναι το να χρησιμοποιήσουμε τα δεδομένα σε PPS, αυτή η επιλογή μας επιτρέπει να βγάλουμε συμπεράσματα μεταξύ χωρών.

### 13.4 Για την επόμενη φορά

#### 13.4.1 Σε δεύτερο χρόνο

- Ορίζουμε ένα πρόβλημα βελτιστοποίησης. Χρησιμοποιούμε το power intensity ως το εργαλείο το οποίο θα μας εξηγήσει ποια χώρα είναι σε θέση να διαχειριστεί με πιο αποτελεσματικό τρόπο τους ρίπους. Μας ενδιαφέρει το power intensity όπως αυτό βρίσκεται στην eurostat στο NRG\_IND\_EI, από όπου θα αναζητήσουμε την έτοιμη τιμή PPS.

Στην συνέχεια θα υλοποιηθούν τα παρακάτω:

- Vanilla version. Θα είναι λογικά ίδιο με αυτό που είχαμε κάνει στο παρελθόν, μόνο που εκεί αντί για το power intensity είχαμε τον λόγο  $\frac{GDP}{verified}$ .
  - Διάφορα constraints και συνδυασμοί αυτών.
  - Να πλησιάσουμε να βρούμε συνδυασμό από περιορισμούς οι οποίοι να προσεγγίζει τις τιμές που προκύπτουν από το EU ETS.
- Διάβασμα το "Cap and Trade and Emission Clustering : A spatial-temporal analysis of the EU ETScheme".

Ορίζουμε ένα πρόβλημα βελτιστοποίησης. Χρησιμοποιούμε το power intensity ως το εργαλείο το οποίο θα μας εξηγήσει ποια χώρα είναι σε θέση να διαχειριστεί με πιο αποτελεσματικό τρόπο τους ρίπους. Μας ενδιαφέρει το power intensity όπως αυτό βρίσκεται στην eurostat στο NRG\_IND\_EI, από όπου θα αναζητήσουμε την έτοιμη τιμή PPS.

Στην συνέχεια θα υλοποιηθούν τα παρακάτω:

- Vanilla version. Θα είναι λογικά ίδιο με αυτό που είχαμε κάνει στο παρελθόν, μόνο που εκεί αντί για το power intensity είχαμε τον λόγο  $\frac{GDP}{verified}$ .
- Διάφορα constraints και συνδυασμοί αυτών.
- Να πλησιάσουμε να βρούμε συνδυασμό από περιορισμούς οι οποίοι να προσεγγίζει τις τιμές που προκύπτουν από το EU ETS.



### 13.4.2 Άμεσα, για το τετρασέλιδο

Για κάθε χώρα έχουμε έναν συνδυασμό από features. Με βάση αυτά τα δεδομένα μπορούμε να τις ομαδοποιήσουμε σε 3 ή 4 κατηγορίες οι οποίες θα έχουν κοινά χαρακτηριστικά. Έστερα, να απαντηθούν τα παρακάτω ερωτήματα:

- Να προστεθούν και άλλα features για τις χώρες αυτές, όπως είναι το power intensity, το ποσοστό της πράσινης ενέργειας που παράγουν.
- Οι χώρες οι οποίες είμαι μέσα στο ίδιο cluster μπορούν να εξηγήσουν καλά τις υπόλοιπες του cluster τους;
- Μήπως χώρες διαφορετικών χωρών μπορούν να εξηγήσουν καλύτερη η μία την άλλη; (Υπόθεση Κώστα, μπορεί να ελεγχθεί τρέχοντας το ίδιο πρόβλημα, αλλά αφαιρώντας όλες τις άλλες χώρες του ίδιου cluster)
- Αν πάρω την χώρα που δρα ως κέντρο σε κάθε cluster τι προκύπτει;
- Μήπως μέσα στο ίδιο cluster παρατηρείται κάποια απλή σχέση των free με κάποιον από τα features ή με κάποιο συνδυασμό αυτών;

## 14 Για τις 4 Απριλίου

Καλό μήνα!

### 14.1 Συσταδοποίηση (έπρεπε...)

#### 14.1.1 Πλήθος συστάδων

Στην R υπάρχει ένα πολύ όμορφο εργαλείο το οποίο προτείνει τον αριθμό των βέλτιστων clusters με 30 διαφορετικούς δείκτες (elbow, silhouette κλπ). Με βάση τα δεδομένα μας, νομίζω πως οι 3 συστάδες είναι η βέλτιστη λύση.

\* Among all indices:

- \* 6 proposed 2 as the best number of clusters
- \* 9 proposed 3 as the best number of clusters
- \* 1 proposed 4 as the best number of clusters
- \* 1 proposed 5 as the best number of clusters
- \* 5 proposed 9 as the best number of clusters
- \* 2 proposed 10 as the best number of clusters

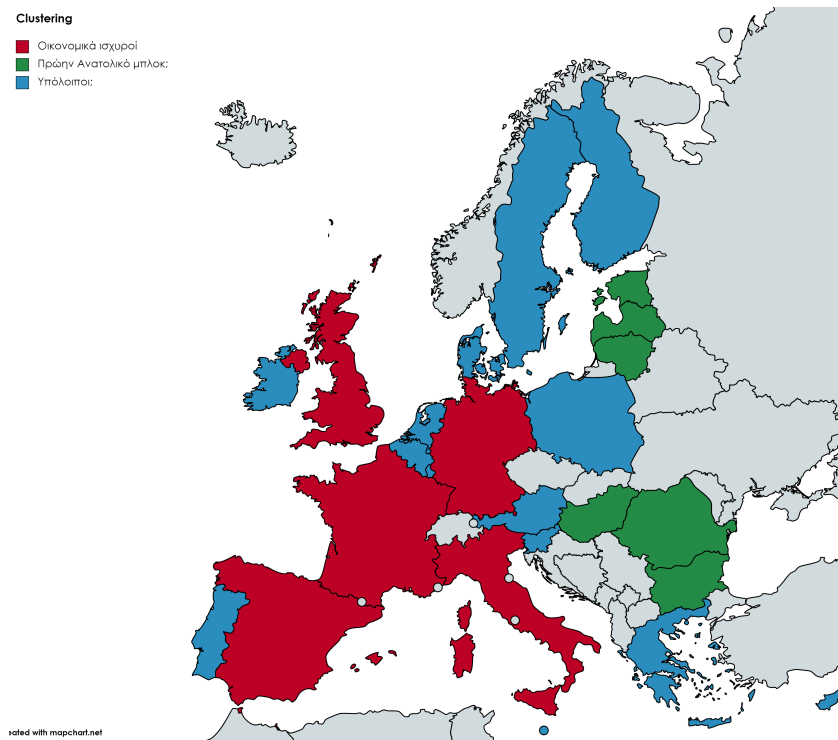
\*\*\*\*\* Conclusion \*\*\*\*\*

\* According to the majority rule, the best number of clusters is 3 Συγκεκριμένα, για όλες τις διαφορετικές μεθόδους προκύπτουν τα παρακάτω:

Clusters	KL	CH	Hartigan	CCC	Scott	Marriot	TrCovW	TraceW	Friedman	Rubin	Cindex	DB	Silhouette
2	4.16	30.55	10.06	5.49	140.15	0.00	0.91	5.56	221.35	4.54	0.35	0.76	0.56
3	6.73	25.81	2.74	5.51	191.04	0.00	0.30	3.87	252.61	6.53	0.27	0.90	0.41
4	0.13	19.34	13.76	4.39	219.59	0.00	0.25	3.44	276.56	7.34	0.26	1.27	0.31
5	4.75	26.15	3.77	6.65	303.07	0.00	0.12	2.08	590.66	12.16	0.33	0.83	0.35
6	3.09	24.33	1.65	6.22	370.12	0.00	0.09	1.75	1,054.45	14.45	0.30	0.70	0.40
7	0.16	21.14	7.77	5.22	383.36	0.00	0.08	1.61	1,044.11	15.70	0.30	0.79	0.34
8	1.56	25.54	6.22	6.54	459.86	0.00	0.03	1.12	1,440.67	22.48	0.24	0.82	0.32
9	3.17	29.46	2.31	7.46	587.08	0.00	0.02	0.82	3,137.27	30.71	0.20	0.70	0.42
10	5.57	28.34	0.68	6.99	613.29	0.00	0.02	0.72	3,743.78	35.14	0.19	0.71	0.38

Clusters	Duda	Pseudot2	Beale	Ratkowsky	Ball	Ptbiserial	Frey	McClain	Dunn	Hubert	SDindex	Dindex	SDbw
2	0.56	14.33	3.99	0.47	2.78	0.76	2.96	0.20	0.40	0.16	8.57	0.42	0.64
3	0.74	4.29	1.74	0.46	1.29	0.59	6.15	0.66	0.14	0.16	8.25	0.33	0.45
4	0.63	2.99	2.37	0.42	0.86	0.43	-0.07	1.53	0.14	0.16	10.58	0.31	0.54
5	2.03	-3.05	-2.30	0.41	0.42	0.46	0.36	1.46	0.21	0.18	8.82	0.25	0.21
6	1.03	-0.17	-0.13	0.38	0.29	0.46	14.61	1.58	0.21	0.19	9.19	0.22	0.15
7	0.35	14.73	7.30	0.36	0.23	0.41	0.10	2.04	0.14	0.19	9.54	0.21	0.15
8	3.69	-3.64	-2.57	0.34	0.14	0.41	0.35	2.04	0.21	0.20	9.06	0.18	0.13
9	5.57	-3.28	-0.00	0.32	0.09	0.39	1.45	2.27	0.32	0.21	11.70	0.15	0.08
10	5.93	-1.66	-2.93	0.31	0.07	0.36	-1.40	2.77	0.25	0.21	11.91	0.14	0.07

### 14.1.2 Βέλτιστες συστάδες



Χώρα	partition
France	1
Germany	1
Italy	1
Spain	1
United Kingdom	1
Bulgaria	2
Estonia	2
Hungary	2
Latvia	2
Lithuania	2
Romania	2
Austria	3
Belgium	3
Cyprus	3
Denmark	3
Finland	3
Greece	3
Ireland	3
Luxembourg	3
Malta	3
Netherlands	3
Poland	3
Portugal	3
Slovenia	3
Sweden	3
EUROZONE	

Figure 12: Συστάδες

Είναι λοιπόν σαν να έχουμε 3 βασικές κατηγορίες και σε αυτό ίσως φτάνει τα δεδομένα τα οποία έχω χρησιμοποιήσει. Μοιάζει σαν να είναι:

- Εκβιομηχανισμένες χώρες
- Μέρος του ανατολικού μπλοκ
- Υπόλοιποι

### 14.1.3 Συστάδες, αν όλα τα δεδομένα ήταν κατά κεφαλήν και κανονικοποιημένα ύστερα

Αυτό δεν είχαμε πει να γίνει, όμως ένοιωσα την ανάγκη να το δοκιμάσω.

Εδώ λοιπόν φαίνεται πως κάτι πήγε πολύ λάθος. Δεν ξέρω τι κοινό έχει η πρώτη συστάδα. Μπορώ να δεχτώ πως η Βουλγαρία είναι μία κατηγορία από μόνης της, αλλά τα μπλε δεν ξέρω πώς σχετίζονται. Το περίεργο είναι πως δεν είναι ο πληθυσμός η ειδοποιός διαφορά, καθώς το Λουξεμβούργο, η Μάλτα και η Βουλγαρία είναι σε 3 διαφορετικές ομάδες. Σημείωση, αυτή τη φορά 15 από τα 23 κριτήρια πρότειναν τις 3 συστάδες.

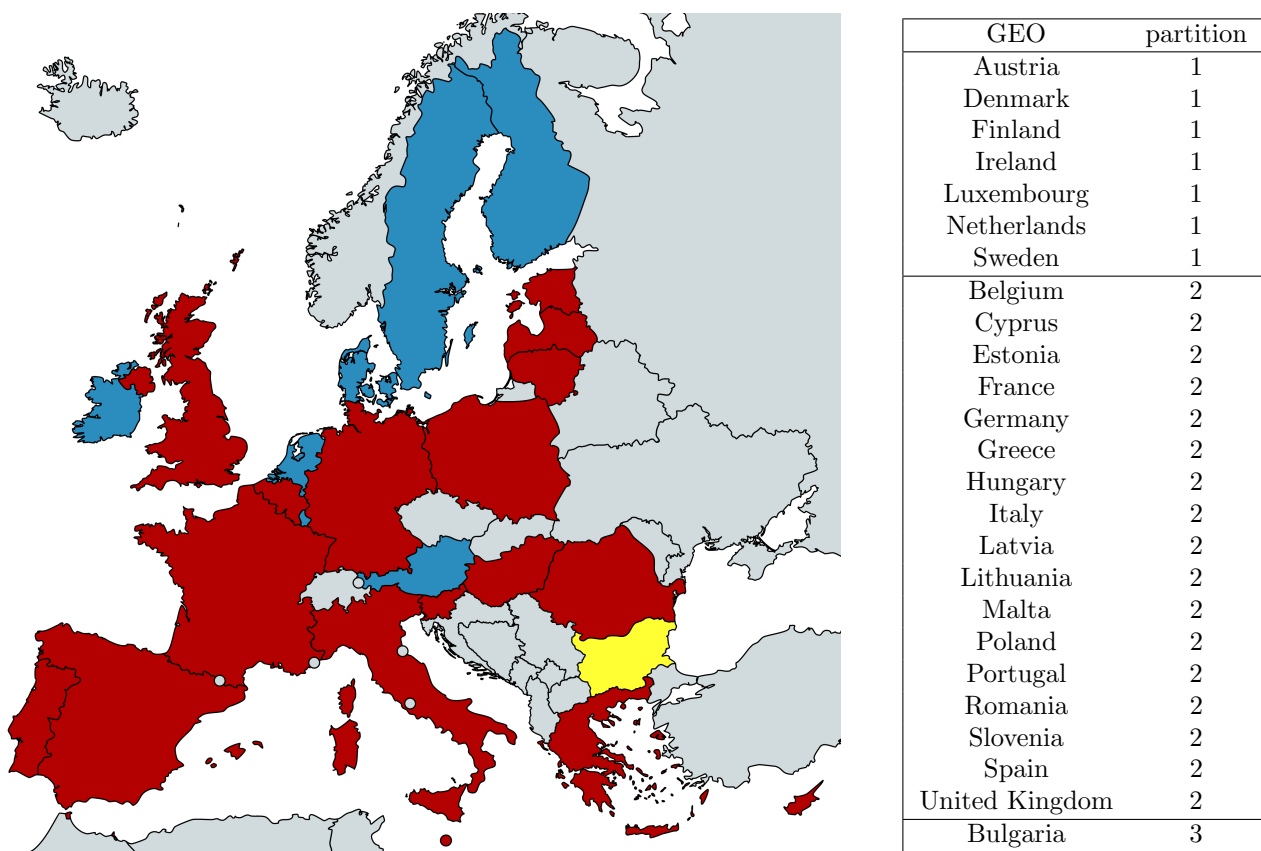


Figure 13: Συστάδες αλλά όλα τα δεδομένα είναι κατά κεφαλήν.

## 14.2 Ικανότητα εξήγησης υπολοίπων μέσα στην ίδια συστάδα

Θα διατηρήσουμε τις συστάδες έτσι όπως διαμορφώθηκαν από την πρώτη διαδικασία:

Εκβιομηχανισμένες	Μισό ανατολικό μπλοκ	Λοιποί
Γαλλία	Βουλγαρία	Αυστρία
Γερμανία	Εσθονία	Βέλγιο
Ιταλία	Ουγγαρία	Κύπρος
Ισπανία	Λετονία	Δανία
Ηνωμένο Βασίλειο	Λιθουανία	Φινλανδία
	Ρουμανία	Ελλάδα
		Ιρλανδία
		Λουξεμβούργο
		Μάλτα
		Ολλανδία
		Πολωνία
		Πορτογαλία
		Σλοβενία
		Σουιδία

Εδώ λοιπόν αφήνουμε τον αλγόριθμο μόνο του να ψάξει για τη μεσαία χώρα και στην συνέχεια να προσπαθήσει μέσω αυτής να εξηγήσει τις υπόλοιπες της ίδιας συστάδας. Έχουμε λοιπόν:

#### 1. Εκβιομηχανισμένες

- Μεσαία χώρα:
- Ικανότητα εξήγησης των άλλων  $r^2$ :

#### 2. Ανατολικό Μπλοκ

- Μεσαία χώρα:
- Ικανότητα εξήγησης των άλλων  $r^2$ :

#### 3. Λοιποί

- Μεσαία χώρα:
- Ικανότητα εξήγησης των άλλων  $r^2$ :

### 14.3 Υπάρχει κάποια εύκολη σχέση μεταξύ των δωρεάν αδειών και κάποιου feature;

Εδώ ένας εύκολος τρόπος να προβλέψουμε την σχέση αυτή είναι να αναζητήσουμε τα καταλληλότερα βάρη ώστε να γίνεται βέλτιστη η γραμμική παλινδρόμηση. Οπότε τρέχοντάς το προκύπτει:

Τιμή	Εκβιομηχανισμένες	Ανατολικό Μπλοκ	Λοιποί
$r^2$	0.9999999	0.976670	0.9463548
Βάρη			
Population	9	0	0
GDPpc	50	1	1
Inflation	50	0	0
Agriculture	59	100	5
Industry	50	100	0
Manufacturing	50	0	0
Total Energy Supply	50	0	100
Verified Emissions	50	1	1
Πιθανή αναλογία free	Οτιδήποτε	Industry + Agriculture	Total energy supply