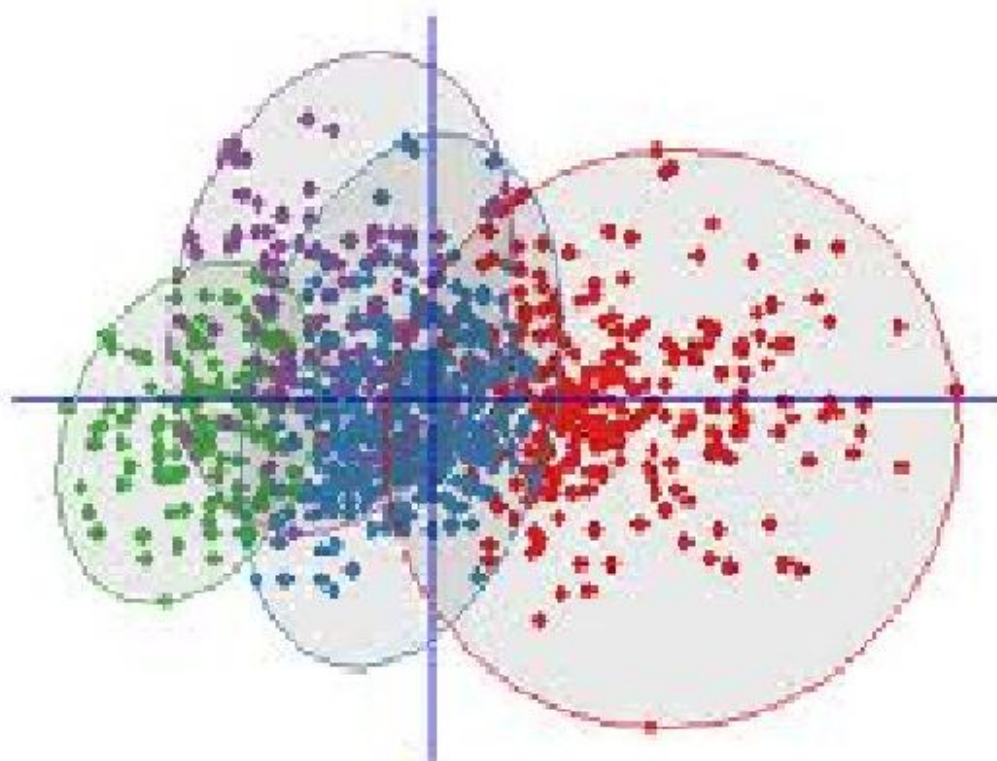




ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
Σχολή Αγρονόμων και Τοπογράφων Μηχανικών
Τομέας Γεωγραφίας και Περιφερειακού Σχεδιασμού

**ΠΟΛΥΜΕΤΑΒΛΗΤΗ ΤΑΞΙΝΟΜΗΣΗ ΔΕΔΟΜΕΝΩΝ:
ΠΑΡΑΔΕΙΓΜΑΤΑ ΕΦΑΡΜΟΓΩΝ ΣΕ ΓΣΠ**

Διπλωματική Εργασία



ΕΝΤΜΟΝΣ ΑΓΓΕΛΟΣ

Επιβλέπων: Ιωάννης Σαγιάς, *Επίκουρος Καθηγητής ΕΜΠ*

**ΑΘΗΝΑ
ΟΚΤΩΒΡΙΟΣ 2012**



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
Σχολή Αγρονόμων και Τοπογράφων Μηχανικών
Τομέας Γεωγραφίας και Περιφερειακού Σχεδιασμού

**ΠΟΛΥΜΕΤΑΒΛΗΤΗ ΤΑΞΙΝΟΜΗΣΗ ΔΕΔΟΜΕΝΩΝ:
ΠΑΡΑΔΕΙΓΜΑΤΑ ΕΦΑΡΜΟΓΩΝ ΣΕ ΓΣΠ**

Διπλωματική Εργασία

ΕΝΤΜΟΝΣ ΑΓΓΕΛΟΣ

Επιβλέπων: Ιωάννης Σαγιάς, *Επίκουρος Καθηγητής ΕΜΠ*

**ΑΘΗΝΑ
ΟΚΤΩΒΡΙΟΣ 2012**

ΠΡΟΛΟΓΟΣ

Με την ολοκλήρωση της εργασίας «Πολυμεταβλητή ταξινόμηση δεδομένων. Παραδείγματα εφαρμογών σε ΓΣΠ», θα ήθελα να ευχαριστήσω, καταρχήν, τον Καθηγητή κ. Κωστή Κουτσόπουλο για την ανάθεση της και αρχική επίβλεψή της καθώς και τα μέλη της τριμελούς επιτροπής Καθηγητή Α. Σιόλα, Καθηγητή Θ. Βλαστό και Επίκουρο Καθηγητή Ι. Σαγιά για την υποστήριξη και τις συμβουλές τους. Ιδιαίτερα θα ήθελα να ευχαριστήσω τον Δρ. Θωμά Χατζηχρήστο μέλος του Ειδικού και εργαστηριακού Διδακτικού Προσωπικού του Τομέα Γεωγραφίας και Περιφερειακού Σχεδιασμού για τη γενικότερη και πολυεπίπεδη συνεισφορά του στην εκπόνηση της Διπλωματικής μου Εργασίας.

Τέλος, θα ήθελα από καρδιάς να ευχαριστήσω τον Καθηγητή Γ.Φώτη για την πολύτιμη και ανιδιοτελή βοήθεια του καθώς και τους φίλους και την οικογένειά μου για την αμέριστη συμπαράσταση και ενθάρρυνση που μου παρείχαν καθ όλη τη διάρκεια των σπουδών μου και ειδικότερα σε όλες τις δύσκολες στιγμές που κατά καιρούς παρουσιάστηκαν.

Αθήνα 5 Οκτωβρίου 2012
Άγγελος Έντμονς

ΠΕΡΙΛΗΨΗ

Το αντικείμενο της διπλωματικής εργασίας με τίτλο «Πολυμεταβλητή ταξινόμηση δεδομένων: Παραδείγματα εφαρμογών σε ΓΣΠ» εντάσσεται στο επιστημονικό πεδίο των Γεωγραφικών Συστημάτων Πληροφοριών (ΓΣΠ) και Ανάλυσης Χώρου. Η εκπόνηση της πραγματοποιήθηκε στο πλαίσιο του προπτυχιακού προγράμματος σπουδών της σχολής Αγρονόμων και Τοπογράφων Μηχανικών του Εθνικού Μετσόβιου Πολυτεχνείου.

Θέμα της συγκεκριμένης διπλωματικής εργασίας, αποτελεί η ανάπτυξη ενός μεθοδολογικού πλαισίου εφαρμογής ανάλυσης, ταξινόμησης και χωροθέτησης γεωγραφικών δεδομένων σε περιβάλλον ΓΣΠ. Το εν λόγω πλαίσιο, αποτελείται από χωρικά δεδομένα και αναλυτικές οδηγίες εφαρμογής της μη επιβλεπόμενης ταξινόμησης τους και θα μπορούσε να αξιοποιηθεί κατά την αντιμετώπιση αντίστοιχων προβλημάτων καθώς και εκπαιδευτικών ασκήσεων εξοικείωσης σπουδαστών και φοιτητών με προβλήματα γεωγραφικού χαρακτήρα.

Παράλληλα, η συγκεκριμένη διπλωματική εξετάζει και αναλύει τις δυνατότητες που μπορεί να παρέχει η διαχείριση της γεωγραφικής πληροφορίας σε περιβάλλον Γεωγραφικών Συστημάτων Πληροφοριών, μέσα από τη σχεδίαση και ανάλυση γεωγραφικών βάσεων δεδομένων. Τέλος, παραθέτονται συμπεράσματα και προτάσεις τόσο για τα παραδείγματα που αναλύθηκαν στο πλαίσιο της παρούσας διπλωματικής όσο και για γενικότερες εφαρμογές στις οποίες θα μπορούσε να εφαρμοστεί και να αξιοποιηθεί η προτεινόμενη προσέγγιση.

ABSTRACT

The subject of the thesis entitled "Multivariate data classification: Examples of applications in GIS" is part of the scientific field of Geographic Information Systems (GIS) and Spatial Analysis. The preparation was carried out as part of the undergraduate curriculum of the School of Rural and Surveying Engineering of the National Technical University of Athens (NTUA).

The subject of this thesis is the development of a methodological framework for analysis application of geographic data, in GIS environment. This framework consists of spatial data and detailed instructions of unsupervised classification and could be exploited not only when addressing relevant problems but also as training examples for students facing geographical problems.

At the same time, this thesis examines and analyzes the opportunities that can be provided in management of geographic information in a GIS environment, through the design and analysis of geographical databases. Finally, conclusions and recommendations are given for both examples discussed in the context of this project.

ΠΕΡΙΕΧΟΜΕΝΑ

ΚΕΦΑΛΑΙΟ 1- ΤΑΞΙΝΟΜΗΣΗ.....	14
1.1. Εισαγωγή στην ταξινόμηση	14
1.2. Εφαρμογές ταξινόμησης	15
1.3. Στάδια ομαδοποίησης	16
1.3.1. Διαμεριστική ομαδοποίηση (Partitional clustering)	18
1.3.2. Αλγόριθμος K means.....	19
1.3.3. Βήματα αλγορίθμου k-means.....	20
ΚΕΦΑΛΑΙΟ 2 – ΓΕΩΓΡΑΦΙΚΑ ΣΥΣΤΗΜΑΤΑ ΠΛΗΡΟΦΟΡΙΩΝ	25
2.1. Γεωγραφικά Συστήματα Πληροφοριών ΓΣΠ	25
2.1.1. Καθορισμός του προβλήματος.....	26
2.1.2. Εισαγωγή Δεδομένων	26
2.1.3. Διαχείριση Δεδομένων	27
2.1.4. Χωρική Ανάλυση.....	28
2.1.5. Παρουσίαση Πληροφορίας.....	28
2.2. Τα Γεωγραφικά Συστήματα Πληροφοριών στην εκπαίδευση.....	29
ΚΕΦΑΛΑΙΟ 3 – ΕΦΑΡΜΟΓΗ ΤΑΞΙΝΟΜΗΣΗΣ ΓΕΩΓΡΑΦΙΚΩΝ ΔΙΑΜΕΡΙΣΜΑΤΩΝ	31
3.1 Προσδιορισμός του προβλήματος.....	31
3.2 Εισαγωγή και Διαχείριση δεδομένων.....	31
3.2.1 Σχεδιασμός Γεωγραφικής Βάσης Δεδομένων	31
3.2.2. Προσδιορισμός θεματικού επιπέδου.....	32
3.2.3 Ψηφιοποίηση θεματικού επιπέδου	32
3.2.4. Καταχώρηση πληροφορίας στα πολύγωνα	36
3.2.5 Παρουσίαση πληροφορίας	36
3.2.6 Απεικόνιση δεδομένων	37
3.3 Ταξινόμηση Γεωγραφικών διαμερισμάτων	44
3.3.1 Ο αλγόριθμος K-Means.....	44
3.4 Συμπεράσματα.....	Σφάλμα! Δεν έχει οριστεί σελιδοδείκτης.
ΚΕΦΑΛΑΙΟ 4- ΕΦΑΡΜΟΓΗ ΤΑΞΙΝΟΜΗΣΗΣ ΔΗΜΩΝ ΑΤΤΙΚΗΣ.....	51
4.1 Προσδιορισμός του προβλήματος.....	51
4.2 Εισαγωγή και διαχείριση δεδομένων.....	51
4.2.1 Κανονικοποίηση δεδομένων.....	52
4.2.2 Σχεδιασμός γεωγραφικής βάσης δεδομένων.....	53
4.2.3 Σύνδεση πληροφορίας και αντίστοιχης χωρικής οντότητας.....	53

4.2.4 Απεικόνιση δεδομένων	55
4.3 Ταξινόμηση.....	
Σφάλμα! Δεν έχει οριστεί σελιδοδείκτης.	
4.4 Οπτικοποίηση αποτελεσμάτων ταξινόμησης.....	61
4.5 Συμπεράσματα.....	64
ΚΕΦΑΛΑΙΟ 5 – ΣΥΜΠΕΡΑΣΜΑΤΑ.....	65
5.1. Ο ρόλος των ΓΣΠ.....	65
5.2. Ο ρόλος της ταξινόμησης.....	66
5.3. Αξιολόγηση των παραδειγμάτων εφαρμογής.....	66

ΠΕΡΙΟΧΟΜΕΝΑ ΕΙΚΟΝΩΝ

ΕΙΚΟΝΑ 2.1: Στάδια και διαδικασίες σε ένα ΓΣΠ	25
ΕΙΚΟΝΑ 3.1: Δημιουργία Sharpefile στο ArcCatalog.....	33
ΕΙΚΟΝΑ 3.2: Προσδιορισμός κατάλληλης μορφής στο νέο αρχείο.....	33
ΕΙΚΟΝΑ 3.3: Εισαγωγή αρχείων στο ArcMap.....	34
ΕΙΚΟΝΑ 3.4: Χάρτης γεωγραφικών διαμερισμάτων.....	34
ΕΙΚΟΝΑ 3.5: Αποτέλεσμα της ψηφιοποίησης στο ArcMap.....	36
ΕΙΚΟΝΑ 3.6:Εισαγωγή και επεξεργασία στο παράθυρο διαλόγου Layer Properties.....	37
ΕΙΚΟΝΑ 3.7: Διαδικασία απόδοσης συμβολισμού.....	38
ΕΙΚΟΝΑ 3.8: Οπτικοποίηση της πληροφορίας σε περιβάλλον ArcMap.....	39
ΕΙΚΟΝΑ 3.9: Εισαγωγή στοιχείων στο Minitab.....	46
ΕΙΚΟΝΑ 3.10: Διαδικασία εκτέλεσης αλγορίθμου K-Means.....	47
ΕΙΚΟΝΑ 3.11: Δήλωση μεταβλητών και αριθμού τάξεων.....	48
ΕΙΚΟΝΑ 3.12: Αποτελέσματα ταξινόμησης σε 3 τάξεις.....	
Σφάλμα! Δεν έχει οριστεί σελιδοδείκτης.	
ΕΙΚΟΝΑ 4.1: Αποτέλεσμα ψηφιοποίησης.....	53
ΕΙΚΟΝΑ 4.2: Επιλογή της εντολής Join.....	54
ΕΙΚΟΝΑ 4.3: Διαδικασία εκτέλεσης της εντολής Join.....	54
ΕΙΚΟΝΑ 4.5: Μετάβαση στον πίνακα χαρακτηριστικών ενός θεματικού επιπέδου.....	55
ΕΙΚΟΝΑ 4.6: Αποτελέσματα σύνδεσης	55
ΕΙΚΟΝΑ 4.7: Διαδικασία επιλογής πεδίου ονοματολογίας στο χάρτη	61
ΕΙΚΟΝΑ 4.8: Επιλογή συμβολισμού στο χάρτη.....	62

ΠΕΡΙΕΧΟΜΕΝΑ ΣΧΗΜΑΤΩΝ

ΣΧΗΜΑ 1.1: Ομαδοποίηση δεδομένων.....	15
ΣΧΗΜΑ 1.2: Κατηγοριοποίηση των τεχνικών clustering.....	18
ΣΧΗΜΑ 1.3: Εφαρμογή του αλγόριθμου των means.....	21
ΣΧΗΜΑ 1.4: Εφαρμογή του αλγόριθμου για τοποθέτηση 2 κέντρων σε δύο κλάσεις.....	22
ΣΧΗΜΑ 1.5: Εφαρμογή του αλγόριθμου για τοποθέτηση 4 κέντρων σε τρεις κλάσεις.....	23
ΣΧΗΜΑ 1.6: Βήματα K-Means.....	23

ΠΕΡΙΕΧΟΜΕΝΑ ΠΙΝΑΚΩΝ

ΠΙΝΑΚΑΣ 3.1: Διαπραχθέντα αδικήματα κατά κατηγορίες και γεωγραφικό διαμέρισμα τελέσεως αυτών. Έτος 2009.....	32
ΠΙΝΑΚΑΣ 3.2: Σειρά ψηφιοποίησης πολυγώνων.....	35
ΠΙΝΑΚΑΣ 4.1: Στατιστικά στοιχεία δήμων νότιας Αττικής.....	52
ΠΙΝΑΚΑΣ 4.2: Κανονικοποιημένα στοιχεία.....	52
ΠΙΝΑΚΑΣ 4.3: Στατιστικά στοιχεία δήμων νότιας Αττικής.....	59
ΠΙΝΑΚΑΣ 4.4: Κανονικοποιημένα στοιχεία-Αποτελέσματα ταξινόμησης.....	59
ΠΙΝΑΚΑΣ 4.5: Ταξινόμηση γεωγραφικών διαμερισμάτων.....	67
ΠΙΝΑΚΑΣ 4.6: Ταξινόμηση δήμων Αττικής.....	67

ΠΕΡΙΕΧΟΜΕΝΑ ΧΑΡΤΩΝ

ΧΑΡΤΗΣ 1: Γεωγραφικά διαμερίσματα Ελλαδικού χώρου.....	40
ΧΑΡΤΗΣ 2: Γεωγραφικά διαμερίσματα.....	41
ΧΑΡΤΗΣ 3: Χάρτης αδικημάτων.....	42
ΧΑΡΤΗΣ 4: Χάρτης εγκλημάτων ανά διαμέρισμα.....	43
ΧΑΡΤΗΣ 5: Χάρτης ταξινόμησης γεωγραφικών διαμερισμάτων.....	50
ΧΑΡΤΗΣ 6: Χάρτης μεταβλητών.....	57
ΧΑΡΤΗΣ 7: Χάρτης ποσοστού επί % σχετικού πληθυσμού.....	58

ΕΙΣΑΓΩΓΗ

Θέμα της παρούσας διπλωματικής εργασίας, αποτελεί η ανάπτυξη ενός μεθοδολογικού πλαισίου εφαρμογής ανάλυσης, ταξινόμησης και χωροθέτησης γεωγραφικών δεδομένων σε περιβάλλον ΓΣΠ. Το εν λόγω πλαίσιο, αποτελείται από χωρικά δεδομένα και αναλυτικές οδηγίες εφαρμογής της μη επιβλεπόμενης ταξινόμησης τους και θα μπορούσε να αξιοποιηθεί κατά την αντιμετώπιση αντίστοιχων προβλημάτων καθώς και εκπαιδευτικών ασκήσεων εξοικείωσης σπουδαστών και φοιτητών με προβλήματα γεωγραφικού χαρακτήρα. Βάση των παραπάνω, ο στόχος της συγκεκριμένης διπλωματικής είναι η υπόδειξη κατευθυντήριων γραμμών και χρήσιμων εργαλείων στους χρήστες, έτσι ώστε να είναι σε θέση να διαχειριστούν γεωγραφική πληροφορία και να δομήσουν το δικό τους γεωγραφικό μοντέλο, εισάγοντας κάθε φορά τα δεδομένα που είναι απαραίτητα για την επίλυση των προβλημάτων που καλούνται να αντιμετωπίσουν.

Πιο συγκεκριμένα, η συγκεκριμένη διπλωματική εργασία εξετάζει και αναλύει τις δυνατότητες που μπορεί να παρέχει η διαχείριση της γεωγραφικής πληροφορίας σε περιβάλλον Γεωγραφικών Συστημάτων Πληροφοριών, μέσα από τη σχεδίαση και ανάλυση γεωγραφικών βάσεων δεδομένων.

Πιο αναλυτικά, παρουσιάζονται δύο παραδείγματα εφαρμογών. Το πρώτο αποτελεί μία τυπική ταξινόμηση των γεωγραφικών διαμερισμάτων του Ελλαδικού χώρου, με βάση τον αριθμό των διαπραχθέντων αδικημάτων, του έτους 2009. Τα περιγραφικά δεδομένα προέρχονται από στοιχεία της Ε.Σ.Υ.Ε και αποτελούν τα πρωταρχικά δεδομένα για την ταξινόμηση.

Στο δεύτερο παράδειγμα, κατηγοριοποιούνται σε τρεις τάξεις (Clusters) δώδεκα γειτονικοί δήμοι της Νότιας Αττικής, αναλόγως με το ποσοστό εμφάνισης διαφόρων μεταβλητών, οι οποίες είναι τεχνητές και αφορούν στα δημογραφικά και οικονομικά στοιχεία του πληθυσμού. Επιπρόσθετα, με στόχο να δειχθεί η πρακτική σημασία της συγκεκριμένης εφαρμογής, με βάση τα δεδομένα που χρησιμοποιήθηκαν, πραγματοποιήθηκε η βέλτιστη επιλογή ζώνης χωροθέτησης ενός καταστήματος λιανικής πώλησης παιδικών προϊόντων.

Για το σκοπό αυτό, αρχικά πραγματοποιείται μία συλλογή πληροφοριών θέσης και των αντίστοιχων προς αυτές περιγραφικών χαρακτηριστικών, κατόπιν παρουσιάζεται η επεξεργασία και οι δυνατότητες χρήσης των Γεωγραφικών Συστημάτων Πληροφοριών, σε αυτές, και εν τέλει αναλύεται η διαδικασία ταξινόμησης σε ομάδες (Cluster Analysis) των χωρικών οντοτήτων, βάση των χαρακτηριστικών τους.

Η Ταξινόμηση Ομάδων (Cluster analysis) είναι μια τεχνική που εφαρμόζεται στην επίλυση προβλημάτων ομαδοποίησης και ταξινόμησης. Σκοπός της

μεθόδου είναι η ομαδοποίηση δεδομένων σε ομοειδή σύνολα ώστε ο βαθμός ομοιότητας και σχέσης να είναι ισχυρός μεταξύ των μελών μιας ομάδας και ασθενής ανάμεσα στα μέλη διαφορετικών κατηγοριών. Κάθε κατηγορία επικαλείται ένα στοιχείο του οποίου τα χαρακτηριστικά είναι μοναδικά και έτσι κάθε αντικείμενο που ανήκει σε αυτήν την ομάδα αποκτά τη δυνατότητα διάκρισής του από αντικείμενα που ανήκουν σε διαφορετικές κατηγορίες.

Και στις δύο περιπτώσεις, επιλέχθηκε ο πιο συχνά χρησιμοποιούμενος διαμεριστικός αλγόριθμος ταξινόμησης, “K-Means”, ο οποίος υλοποιήθηκε με χρήση του στατιστικού πακέτου “Minitab”, που φημίζεται ως ιδιαίτερα εύχρηστο για το προπτυχιακό επίπεδο. Εν συνεχεία, η χωρική πληροφορία μαζί με τα αποτελέσματα της ταξινόμησης εισήχθησαν εντός του λογισμικού GIS, διαμορφώνοντας τον πυρήνα μίας ψηφιακής βάσης δεδομένων, που παρέχει στον χρήστη εξαιρετικές δυνατότητες επεξεργασίας, όπως η οπτικοποίηση της πληροφορίας και η επίλυση προβλημάτων χωροθέτησης. Η επεξεργασία σε περιβάλλον GIS στοχεύει στην εξοικείωση των φοιτητών και ιδιαίτερα στην εκμάθηση παραγωγής χαρτών.

Τελικά και στις δύο περιπτώσεις τα αποτελέσματα της ταξινόμησης είναι τρεις τάξεις, οι οποίες περιγράφονται εκτενώς και προσδιορίζονται γεωγραφικά, με χρήση του λογισμικού ArcGis.

Για την όσο το δυνατό καλύτερη ανάπτυξη του θέματος, κρίθηκε κατάλληλος ο διαχωρισμός της εργασίας σε τέσσερα κεφάλαια.

Στο πρώτο κεφάλαιο περιγράφεται συνοπτικά το πρόβλημα της συσταδοποίησης που σχετίζεται με την τμηματοποίηση (partitioning, clustering) ενός συνόλου δεδομένων σε συστάδες έτσι ώστε τα στοιχεία του συνόλου των δεδομένων που ανήκουν σε μία συστάδα να είναι περισσότερο όμοια μεταξύ τους από ότι είναι με τα στοιχεία των άλλων συστάδων. Πιο συγκεκριμένα περιγράφεται ο αλγόριθμος k-means. και το στατιστικό πακέτο minitab με τη βοήθεια των οποίων ταξινομούνται τα στοιχεία που έχουμε στη διάθεσή μας .

Στο δεύτερο κεφάλαιο γίνεται μια προσπάθεια εισαγωγής στις βασικές έννοιες και τις δυνατότητες των Γεωγραφικών Συστημάτων Πληροφοριών, στις κύριες διαδικασίες τους καθώς και τον ρόλο τους στην εκπαίδευση. Επιδιωκόμενο στόχο αποτελεί η εννοιολογική αποσαφήνιση, η λειτουργική περιγραφή και οι δυνατότητες εφαρμογής των Γεωγραφικών Συστημάτων Πληροφοριών.

Όπως εύκολα διαπιστώνει κανείς στα δύο πρώτα κεφάλαια πραγματοποιείται μία καθαρά θεωρητική προσέγγιση που αφορά στη διαδικασία της ταξινόμησης και στα ΓΣΠ.

Στο τρίτο κεφάλαιο και στο τέταρτο κεφάλαιο η θεωρία μετατρέπεται σε πράξη μέσω των δύο παραδειγμάτων εφαρμογών που προαναφέρθηκαν. Το πρώτο αφορά στον αριθμό διαπραχθέντων αδικημάτων ανά γεωγραφικό διαμέρισμα και το δεύτερο αναφέρεται στην επεξεργασία δημογραφικών και οικονομικών δεδομένων δώδεκα δήμων της νότιας Αττικής. Και οι δύο περιπτώσεις, μέσα από την κλασική προσέγγιση ανάλυσης και με χρήση του λογισμικού ΓΣΠ,

καταλήγουν στην παραγωγή χαρτών, οι οποίοι προσφέρουν πλήρη εποπτεία και αμεσότητα στον παρατηρητή.

Η εργασία κλείνει με το τελευταίο κεφάλαιο, στο οποίο παραθέτονται συμπεράσματα και προτάσεις τόσο για τις γενικότερες εφαρμογές που παρουσιάστηκαν μέσα από την εργασία αυτή, όσο και για τα εργαλεία που χρησιμοποιήθηκαν πιο συγκεκριμένα στα πλαίσια της παρούσας διπλωματικής.

ΚΕΦΑΛΑΙΟ 1

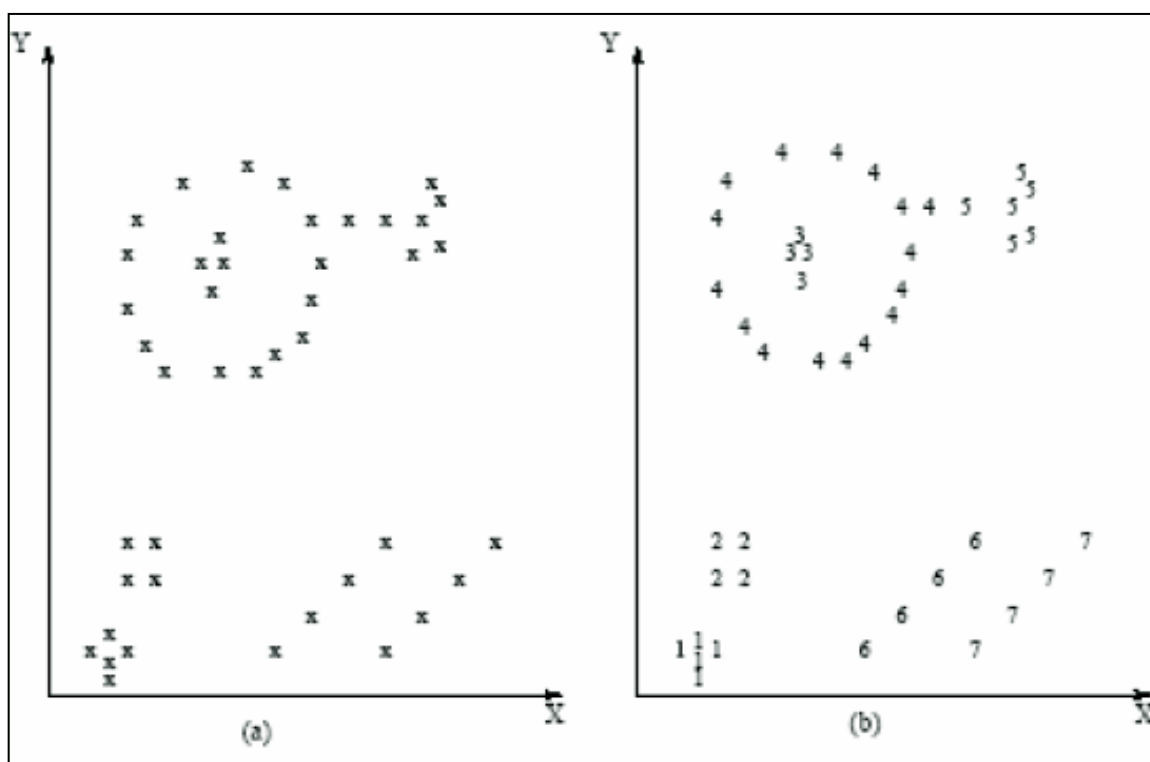
ΤΑΞΙΝΟΜΗΣΗ

Στο κεφάλαιο αυτό παρουσιάζεται συνοπτικά η τεχνική της ομαδοποίησης (Clustering/Cluster Analysis) και πιο συγκεκριμένα ο αλγόριθμος k-means. Για την επίλυση των δύο εκπαιδευτικών παραδειγμάτων που τέθηκαν χρησιμοποιήθηκαν οι παραπάνω τεχνικές με τη βοήθεια του στατιστικού πακέτου minitab.

1.1. Εισαγωγή στην ταξινόμηση

Ένας απλός ορισμός για την *ταξινόμηση ή ομαδοποίηση ή συσταδοποίηση (clustering)*: ομαδοποίηση ονομάζεται η διαδικασία που οργανώνει πρότυπα (παρατηρήσεις, δεδομένα ή διανύσματα χαρακτηριστικών) σε *ομάδες (συστάδες-clusters)*, όπου τα μέλη μιας ομάδας είναι παρόμοια μεταξύ τους σύμφωνα με κάποιο κριτήριο. Σκοπός είναι να προσδιοριστούν οι ομάδες που ανήκουν διάφορες ποσότητες δεδομένων, με βάση κάποια κριτήρια ομοιογένειας. Η τεχνική της ομαδοποίησης υπάγεται στην ευρύτερη κατηγορία των τεχνικών μάθησης χωρίς επίβλεψη. Η διαφορά της ομαδοποίησης δεδομένων (data clustering) από την ταξινόμηση δεδομένων (data classification) είναι ότι, στην ταξινόμηση οι ομάδες στις οποίες θα τοποθετηθούν τα δεδομένα είναι προκαθορισμένες. Αυτό σημαίνει, ότι είναι εκ των προτέρων γνωστός ο αριθμός των ομάδων, τα ονόματα και οι ταυτότητες τους. Είναι και αυτό ένα σύστημα μάθησης μιας και οι ετικέτες που δίνονται από τα διαθέσιμα πρότυπα χρησιμοποιούνται ώστε να μάθει το σύστημα ταξινόμησης την περιγραφή κάθε κλάσης και να είναι σε θέση να ταξινομήσει ένα νέο πρότυπο. Αντίθετα, στην ομαδοποίηση δεδομένων τονίζεται ιδιαίτερα ότι οι ομάδες δεν προϋπάρχουν αλλά αποφασίζονται από τον αλγόριθμο κατά δυναμικό τρόπο. Στην ομαδοποίηση δεδομένων δηλαδή, υπάρχει ένα σύνολο δεδομένων το οποίο πρέπει να διαχειριστεί ώστε από αυτό να προκύψουν δυναμικά οι ομάδες (είναι δηλαδή data driven). Σκοπός είναι να δημιουργηθούν ομάδες, που η καθεμία από αυτές θα συγκεντρώνει ομοιογενή στοιχεία. Κάθε μία από αυτές τις ομάδες διατηρεί ένα κέντρο, συνήθως το πιο κεντρικό στοιχείο της.

Ένα παράδειγμα ομαδοποίησης δίνεται παρακάτω(σχήμα 1.1), όπου αριστερά παρουσιάζεται το αρχικό σύνολο των στοιχείων πριν την ομαδοποίηση και δεξιά η καταχώρηση των στοιχείων σε τάξεις(clusters).



ΣΧΗΜΑ 1.1: Ομαδοποίηση δεδομένων

ΠΗΓΗ: <http://invenio.lib.auth.gr/record/114601/files/ptuxiaki.pdf?version=1>

1.2. Εφαρμογές ταξινόμησης

Η μέθοδος ταξινόμησης Ομάδων (*Cluster analysis*) επιτρέπει την εύκολη συμπύεση δεδομένων καθώς και την αναγνώριση χαρακτηριστικών υποσυνόλων δεδομένων (π.χ. για σκοπούς marketing όπου επιθυμούμε να εστιάσουμε σε συγκεκριμένες ομάδες). Αυτό πραγματοποιείται με την κατάτμηση βάσεων δεδομένων. Οι μέθοδοι clustering λοιπόν χρησιμοποιούνται για τεμαχισμό βάσεων δεδομένων σε ομογενή κομμάτια.

Μια επιπλέον δυνατότητα αποτελεί η οπτικοποίηση μεγάλων βάσεων δεδομένων. Τα clusters αποτελούν μεγαλύτερα σύνολα δεδομένων, τα οποία μέσω της οπτικοποίησης επιτρέπουν εύκολο εντοπισμό ομάδων και υποσυνόλων δεδομένων που έχουν παρόμοια χαρακτηριστικά.

Ενδεικτικά παραδείγματα εφαρμογής των μεθόδων ταξινόμησης :

- Επιχειρήσεις (Marketing)
- Βιολογία (Εξαγωγή συμπερασμάτων)
- Χωρική Ανάλυση Στοιχείων
- Εξόρυξη (Data Mining) στο Παγκόσμιο Ιστό
- Ψυχολογία και Ιατρική (Εξαγωγή συμπερασμάτων)

1.3. Στάδια ομαδοποίησης

Για την επίλυση ενός προβλήματος ομαδοποίησης δεδομένων συνήθως ακολουθούνται τα παρακάτω βήματα (Jain and Dubes, 1998):

1. Αναπαράσταση των προτύπων (επιλεκτικά μπορεί να περιέχει εξαγωγή χαρακτηριστικών, και/ή επιλογή)
2. Καθορισμός μιας μετρικής, ενδεικτικής της γειννίαςης των προτύπων, ανάλογα με τον τύπο δεδομένων
3. Τεχνική ομαδοποίησης των δεδομένων
4. Αφαίρεση δεδομένων (αν χρειαστεί)
5. Αξιολόγηση του τελικού αποτελέσματος

Είναι χαρακτηριστικό ότι η ομαδοποίηση είναι μια διαδικασία με επανατροφοδότηση: το αποτέλεσμα της διαδικασίας επανατροφοδοτείται στο σύστημα, το οποίο συνδυάζοντας το αποτέλεσμα αυτό με τις υπόλοιπες εισόδους, προχωράει στην εξαγωγή χαρακτηριστικών και στους υπολογισμούς των σχέσεων ομοιότητας, με στόχο την τελική εξαγωγή των ομάδων.

Η αναπαράσταση των προτύπων αναφέρεται στο πλήθος των κλάσεων, το πλήθος των διαθέσιμων προτύπων και το πλήθος, τύπο και κλίμακα των χαρακτηριστικών που είναι διαθέσιμα στον συγκεκριμένο αλγόριθμο ομαδοποίησης. Ωστόσο, μερικά από τα προηγούμενα δεν είναι πάντα άμεσα διαθέσιμα. Ενδιαφέρον παρουσιάζει η διαδικασία της επιλογής χαρακτηριστικών (features selection), κατά την οποία επιλέγονται τα πιο σημαντικά χαρακτηριστικά των στοιχείων τα οποία θα χρησιμοποιηθούν στο clustering. Επιπλέον, η διαδικασία της εξαγωγής χαρακτηριστικών (features extraction) χρησιμοποιεί έναν ή περισσότερους μετασχηματισμούς των χαρακτηριστικών εισόδου, για την παραγωγή άλλων νέων, τα οποία πιθανόν να είναι πιο ενδιαφέροντα. Οποιαδήποτε από τις τεχνικές αυτές, μπορεί να χρησιμοποιηθεί για τη δημιουργία ενός συνόλου με τα πιο κατάλληλα χαρακτηριστικά, που θα χρησιμοποιηθεί για την αναπαράσταση των στοιχείων που προορίζονται για ομαδοποίηση.

Η γειννίαση των προτύπων συνήθως μετριέται με βάση μία συνάρτηση απόστασης που ορίζεται για ζεύγη προτύπων. Η πιο απλή συνάρτηση απόστασης είναι η Ευκλείδεια. Η συνάρτηση απόστασης η οποία επιλέγεται, αποτελεί κάθε φορά το μέτρο της ομοιότητας μεταξύ των προτύπων. Με βάση αυτό το μέτρο γίνεται η καταχώρηση τους στην ίδια ή σε διαφορετικές ομάδες. Το πόσο επιτυχημένο θεωρείται το αποτέλεσμα της ομαδοποίησης δεδομένων, (Mirkin, 1996), (Kalgamuge, 2005), εξαρτάται από τα κριτήρια που θα χρησιμοποιηθούν για τον διαχωρισμό των στοιχείων σε ομάδες. Η σωστή επιλογή των κριτηρίων αυτών είναι ένα πολύ σημαντικό ζήτημα.

Το στάδιο της ομαδοποίησης μπορεί να πραγματοποιηθεί με πολλούς τρόπους. Υπάρχουν πολλοί αλγόριθμοι ομαδοποίησης (αναλύονται παρακάτω), καθένας μπορεί να έχει διαφορετικό αποτέλεσμα είτε αυστηρό, είτε ασαφές. Το στάδιο αυτό αποτελεί και το κυρίως μέρος της όλης διαδικασίας της ομαδοποίησης. Εδώ είναι το σημείο που πρέπει να επιλεγεί ο αλγόριθμος που θα χρησιμοποιηθεί για την επίλυση του προβλήματος. Φυσικά το μέτρο ομοιότητας

που επιλέχτηκε παραπάνω θα χρησιμοποιηθεί από τον αλγόριθμο που θα επιλεγεί. Οι αλγόριθμοι που επιλύουν προβλήματα ομαδοποίησης είναι πολλοί και στηρίζονται σε διαφορετικές τεχνικές. Η επιλογή εξαρτάται από τη μορφή των δεδομένων και από τον χρήστη. Οι κύριες κατηγορίες των αλγόριθμων ομαδοποίησης είναι δύο, οι διαμεριστικές και οι ιεραρχικές. (Jain and Dubes, 1998)

Οι ιεραρχικοί αλγόριθμοι προσπαθούν να δημιουργήσουν μια ιεραρχία μεταξύ των σημείων που προορίζονται για ομαδοποίηση. Δημιουργούν ένα δένδρογραμμα που υποδηλώνει το μέγεθος και τον αριθμό των ομάδων που δημιούργησαν. Κάθε κόμβος του δέντρου έχει παιδιά τα σημεία που συγχωνεύτηκαν στην ίδια ομάδα. Ανάλογα με το αν βρίσκονται κοντά ή μακριά από τη ρίζα προκύπτουν λίγες ομάδες με πολλά σημεία ή πολλές ομάδες με λίγα σημεία αντίστοιχα. Οι ιεραρχικοί αλγόριθμοι χωρίζονται στους συσσωρευτικούς και στους διαιρετικούς. Οι συσσωρευτικοί ξεκινούν θεωρώντας ότι κάθε σημείο είναι ένα από μόνο του μια ομάδα που περιέχει μόνο τον εαυτό του και στη συνέχεια πραγματοποιούν συγχωνεύσεις. Οι διαιρετικοί λειτουργούν αντίστροφα. Θεωρούν ότι αρχικά υπάρχει μία ομάδα με μέσα όλα τα σημεία και στη συνέχεια διαιρούν την ομάδα αυτή σε μικρότερες.

Οι διαμεριστικοί αλγόριθμοι χωρίζουν τα δεδομένα από την αρχή σε ένα συγκεκριμένο αριθμό από ομάδες και έπειτα βελτιστοποιούν το αποτέλεσμα. Και αυτοί χωρίζονται σε περεταίρω κατηγορίες (αναφέρονται παρακάτω). Οι αλγόριθμοι που είναι βασισμένοι στην πυκνότητα δημιουργούν ομάδες με βάση την πυκνότητα των αντικειμένων στο χώρο. Ένα σημείο το οποίο ανήκει σε κάποια ομάδα θα πρέπει να έχει στη γειτονιά του (ορίζεται η ακτίνα της γειτονιάς του σημείου) ένα συγκεκριμένο αριθμό από άλλα σημεία.

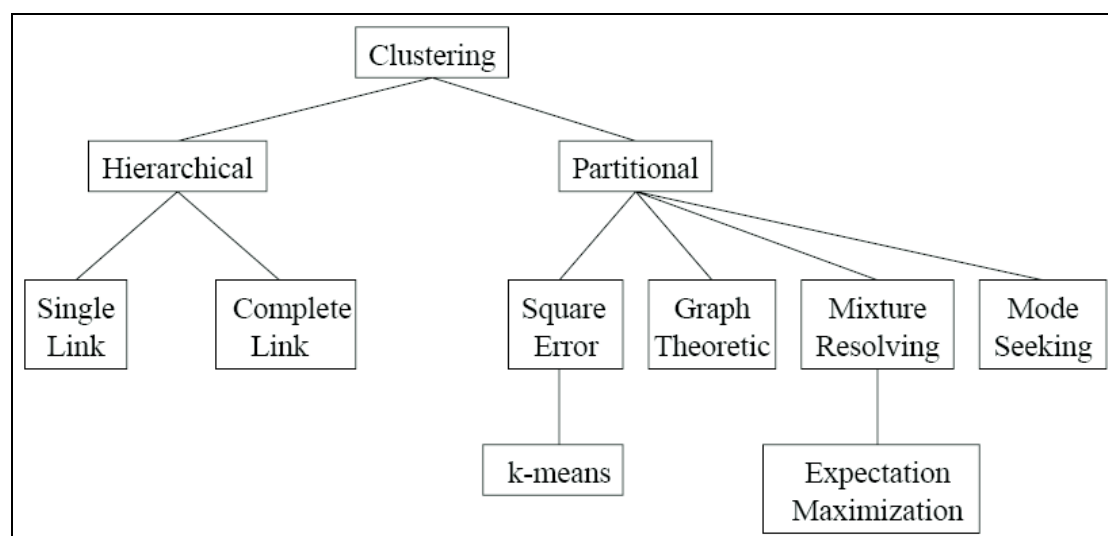
Δύο άλλες κατηγορίες αλγορίθμων που επιλύουν προβλήματα ομαδοποίησης είναι οι αυστηροί και οι ασαφείς αλγόριθμοι. Οι αυστηροί αλγόριθμοι θεωρούν ότι τα σημεία ανήκουν κατά απόλυτο τρόπο στις ομάδες τους. Αυτό πρακτικά σημαίνει ότι δεν μπορεί ένα σημείο να βρίσκεται ταυτόχρονα σε παραπάνω από μία ομάδα. Σε αντίθεση με αυτή τη λογική οι ασαφείς αλγόριθμοι θεωρούν ότι τα σημεία ανήκουν σε όλες τις ομάδες σε κάποιο βαθμό. Διαθέτουν μια συνάρτηση συμμετοχής η οποία μας δίνει το βαθμό συμμετοχής του κάθε σημείου σε κάποια ομάδα. Προφανώς από την ασαφή ομαδοποίηση μπορεί να προκύψει αυστηρή ομαδοποίηση.

Τέλος υπάρχουν οι αυξητικοί και μη αυξητικοί αλγόριθμοι. Στους αυξητικούς αλγόριθμους το σύνολο των δεδομένων που τίθεται προς ομαδοποίηση προσέρχεται σταδιακά. Αυτό σημαίνει ότι τα σημεία έρχονται ένα ένα ή κατά ομάδες. Εδώ γίνεται ομαδοποίηση από τον αλγόριθμο χωρίς να γνωρίζει εκ των προτέρων όλο το σύνολο των δεδομένων (on line προβλήματα ομαδοποίησης). Στους μη αυξητικούς αλγόριθμους είναι γνωστό ολόκληρο το σύνολο των δεδομένων εξ αρχής.

Η αφαίρεση δεδομένων, είναι η διαδικασία η οποία έχει σαν αποτέλεσμα μια απλή και συμπαγή αναπαράσταση του συνόλου των δεδομένων. Ο όρος απλή αναπαράσταση μπορεί να εξηγηθεί είτε από την οπτική γωνία της αυτοματοποιημένης ανάλυσης είτε από την οπτική γωνία του ανθρώπου. Στην

πρώτη περίπτωση, το επιθυμητό για τα δεδομένα είναι να αναπαρίστανται με τέτοιο σαφή και απλό τρόπο, ώστε μια περαιτέρω υπολογιστική επεξεργασία να είναι εξίσου εφικτή. Στη δεύτερη περίπτωση, η απλή αναπαράσταση των δεδομένων τα κάνει πιο κατανοητά στους ειδικούς που πρόκειται να τα επεξεργαστούν και να εξάγουν συμπεράσματα. Συνήθως, η αφαίρεση δεδομένων στο clustering είναι μια συνοπτική αναπαράσταση κάθε ομάδας μέσω κάποιου αντιπροσώπου-πρωτότυπου στοιχείου το οποίο καλείται centroid (κεντροειδής).

Τέλος, στο στάδιο της αξιολόγησης του αποτελέσματος ελέγχεται η εγκυρότητα των ομάδων. Εδώ εξετάζεται εάν το τελικό αποτέλεσμα του αλγορίθμου είναι επιτυχές οπότε ο αλγόριθμος θεωρείται αξιόπιστος. Πρακτικά εξετάζεται αν οι ομάδες είναι αντιπροσωπευτικές σε σχέση με τα σημεία που έπρεπε να ομαδοποιηθούν, αν τα σημεία τελικά τοποθετήθηκαν στις κατάλληλες ομάδες κ.ο.κ. Η αξιολόγηση συνήθως γίνεται συγκρίνοντας τη ληφθείσα δομή με μια δεδομένη εκ των προτέρων δομή. Για την ανάλυση που γίνεται σε αυτό το στάδιο, χρησιμοποιείται ένα συγκεκριμένο κριτήριο βελτιστοποίησης, ανάλογα με το πρόβλημα που αντιμετωπίζεται κάθε φορά.



ΣΧΗΜΑ 1.2: Κατηγοριοποίηση των τεχνικών clustering
 ΠΗΓΗ: <http://invenio.lib.auth.gr/record/114601/files/ptuxiaki.pdf?version=1>

1.3.1. Διαμεριστική ομαδοποίηση (Partitional clustering)

Στην διαμεριστική ομαδοποίηση, ορίζονται εξαρχής οι ομάδες και τοποθετούνται τα σημεία του συνόλου δεδομένων που είναι διαθέσιμα στις ομάδες αυτές. Στη συνέχεια επαναπροσδιορίζονται αυτές οι αναθέσεις έως ότου κάποιο κριτήριο τερματισμού εκπληρωθεί. Ουσιαστικά αν υπάρχει ένα σύνολο δεδομένων πλήθους N , κατασκευάζονται k ομάδες και στη συνέχεια βελτιστοποιούνται. Το κυρίως πρόβλημα αυτών των τεχνικών είναι η σωστή επιλογή των ομάδων και του αριθμού k . Όπως και προηγουμένως αφού έχουν πραγματοποιηθεί όλες οι αναθέσεις των σημείων στις ομάδες, επαναπροσδιορίζονται οι αναθέσεις αυτές.

Αυτό μπορεί να συμβεί αν σε κάθε ομάδα αποδοθεί μια τιμή και σιγά σιγά ελαχιστοποιείται. Η τιμή αυτή μπορεί να αποδοθεί με το άθροισμα των τετραγωνικών αποστάσεων από το μέσο όρο στην κάθε ομάδα. (Jain and Dubes, 1998)

Οι πιο δημοφιλείς αλγόριθμοι στην κατηγορία αυτή είναι οι Centroids αλγόριθμοι. Εδώ η κάθε ομάδα αναπαρίσταται με το κέντρο μάζας της. Οι πιο γνωστοί αλγόριθμοι στην κατηγορία αυτή είναι ο αλγόριθμος απλού περάσματος και οι αλγόριθμοι που χρησιμοποιούν το κριτήριο του τετραγωνικού λάθους.

- Αλγόριθμος απλού περάσματος

Στον αλγόριθμο αυτό ομαδοποιούνται όλα τα δεδομένα του συνόλου, με ένα μόνο πέρασμα. Η λογική του έχει ως εξής:

- Το πρώτο στοιχείο χρίζεται κέντρο μάζας της πρώτης ομάδας.
- Για τα επόμενα στοιχεία υπολογίζονται οι αποστάσεις τους από τα κέντρα μάζας των ομάδων που ήδη υπάρχουν και κρατιούνται οι μικρότερες, μία για κάθε στοιχείο.
- Αν η απόσταση d_i του στοιχείου i από την πιο κοντινή του ομάδα, είναι μικρότερη από μια τιμή κατωφλίου που έχει τεθεί, τότε: Ανατίθεται το στοιχείο i στην ομάδα.
- Διαφορετικά το στοιχείο i γίνεται κέντρο μάζας μιας νέας ομάδας.
- Όταν ομαδοποιηθούν όλα τα στοιχεία τερματίζει.

- Αλγόριθμοι ομαδοποίησης τετραγωνικού λάθους

Στους συγκεκριμένους αλγορίθμους χρησιμοποιείται το κριτήριο του τετραγωνικού λάθους. Τα συνηθέστερα βήματα που ακολουθούν οι αλγόριθμοι αυτοί είναι τα εξής:

1. Επιλέγουμε τα στοιχεία που θα εκπροσωπούν τις ομάδες.
 2. Αναθέτουμε τα υπόλοιπα στοιχεία στην πλησιέστερη ομάδα.
 3. Επαναπροσδιορίζουμε τα κέντρα των ομάδων.
 4. Επαναλαμβάνουμε το δεύτερο και τρίτο βήμα μέχρι να μην παρατηρούνται αλλαγές (δεν αλλάζουν τα κέντρα των ομάδων, δεν παρατηρούνται διαφορετικές αναθέσεις, το τετραγωνικό λάθος σταθεροποιείται).
 5. Συγχωνεύουμε τις ομάδες με βάση π.χ. κάποια ευρετική συνάρτηση.
- Ο πιο κλασικός αλγόριθμος της κατηγορίας αυτής είναι ο k-means (McQueen, 1994).

1.3.2. Αλγόριθμος K Means

Ο k-means είναι ένας από τους πιο γνωστούς και πιο απλούς αλγόριθμους που λύνουν το πρόβλημα του clustering. Δημοσιεύτηκε για πρώτη φορά από τον McQueen το 1994. Φυσικά ανήκει στην κατηγορία της μη επιβλεπόμενης ταξινόμησης κατά την οποία τα δεδομένα μας δεν έχουν καμία ετικέτα και δεν γνωρίζουμε τίποτα για αυτά. Ο αλγόριθμος ακολουθεί μια απλή και εύκολη διαδικασία για να κατηγοριοποιήσει τα δοσμένα δεδομένα σε έναν συγκεκριμένο αριθμό από κλάσεις (clusters). Η κύρια ιδέα του αλγορίθμου είναι να καθορίσουμε εμείς έναν συγκεκριμένο αριθμό από k κέντρα των κλάσεων

(centroids) όταν θα ξεκινάει ο αλγόριθμος, που θα συμβολίζουν φυσικά και τον αριθμό των τελικών clusters που θα έχουμε ως έξοδο του αλγορίθμου.

Το επόμενο βήμα που κάνει ο αλγόριθμος είναι να αναθέσει κάθε δεδομένο (datum) στο κοντινότερο του κεντροειδές (centroid). Όταν ανατεθούν όλα τα δεδομένα, ένα πρώιμο clustering έχει γίνει. Σε αυτό το σημείο επαναυπολογίζουμε τα κεντροειδή με βάση τα καινούργια clusters που έχουν δημιουργηθεί και τα τοποθετούμε έτσι ώστε να κατοπτρίζουν το κέντρο των δεδομένων που ανήκουν στο cluster τους.

Έτσι όταν έχουμε k καινούργια centroids επαναυπολογίζουμε τις θέσεις των δεδομένων και τα εναποθέτουμε στο κοντινότερο σε αυτά κέντρο. Με τον τρόπο αυτό βλέπουμε πως έχει δημιουργηθεί ένας βρόχος ο οποίος τερματίζεται όταν πλέον τα κέντρα δεν κουνιούνται από την θέση τους. Τέλος, ένας από τους βασικούς στόχους του αλγόριθμου είναι να καταφέρει να ελαχιστοποιήσει την παρακάτω συνάρτηση τετραγωνικού λάθους:

$$SSE = \sum_{i=1}^K \sum_{x \in C_i} \text{dist}(c_i, x)^2$$

όπου

- x : ένα στοιχείο των δεδομένων
- C_i : η κλάση i
- c_i : το κέντρο της κλάσης C_i
- K : ο αριθμός των κλάσεων
- m_i : ο αριθμός των στοιχείων στην κλάση i
- m : ο αριθμός όλων των στοιχείων

1.3.3. Βήματα αλγορίθμου k-means

Ο αλγόριθμος των k means clustering θα μπορούσε να περιγραφεί συντόμως από τα βήματα:

1. Τοποθετούμε K σημεία στο χώρο ο οποίος εκπροσωπείται από τα αντικείμενα τα οποία ομαδοποιούνται. Αυτά τα σημεία εκπροσωπούν τα αρχικά κέντρα των κλάσεων.

2. Κατανέμουμε κάθε αντικείμενο στο γκρουπ που έχει το κοντινότερο κέντρο σε αυτό.

3. Όταν όλα τα σημεία έχουν κατανεμηθεί επαναυπολογίζουμε τις θέσεις των K κέντρων.

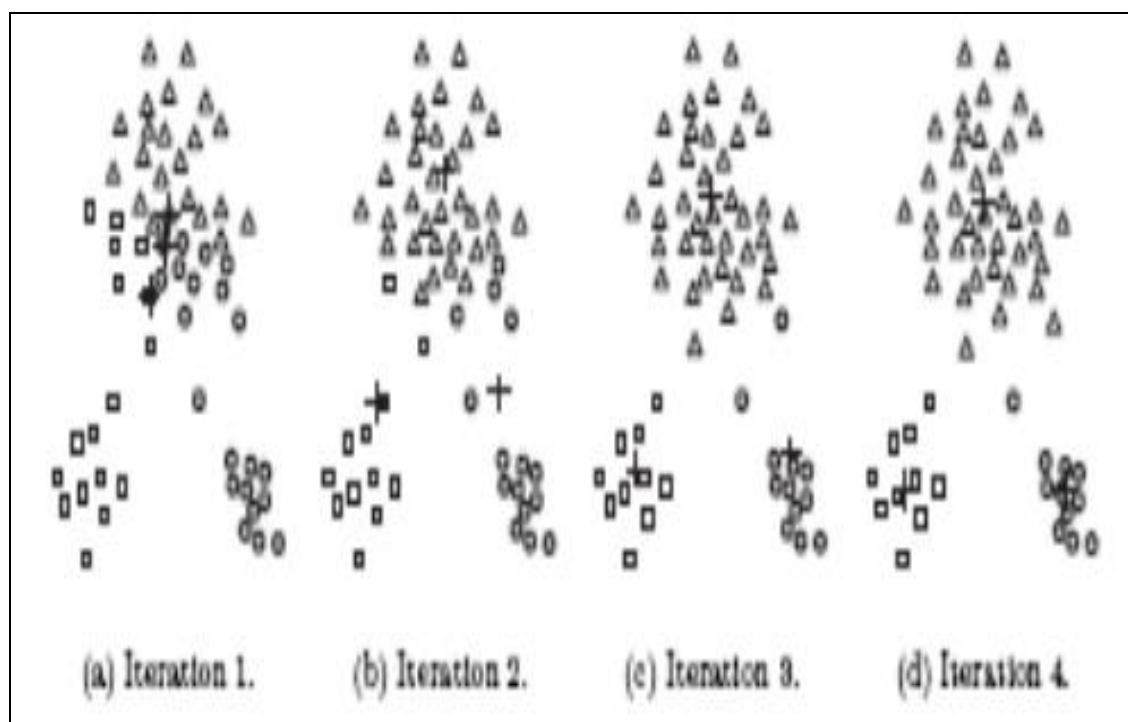
Επαναλαμβάνουμε τα βήματα 2 και 3 μέχρι τα κέντρα να μην μετακινούνται άλλο.

Αυτό παράγει διαχωρισμό των αντικειμένων σε γκρουπ από τα οποία μπορεί να υπολογιστεί η ελάχιστη τιμή της μέτρησης της μονάδος της απόστασης.

Αν και μπορεί να αποδειχθεί ότι η διαδικασία πάντα θα τερματίσει, ο αλγόριθμος k -means δεν βρίσκει πάντα την βέλτιστη λύση στο πρόβλημα του clustering

δηλαδή τα clusters που επιστρέφονται δεν είναι πάντοτε τα επιθυμητά. Σε αυτό το γεγονός συντελεί και η αρχικοποίηση του αλγορίθμου όπου πρέπει ο χρήστης να επιλέξει τον αρχικό αριθμό κέντρων με τον οποίο θα ξεκινήσει ο αλγόριθμος. Αυτό είναι και το σημαντικότερο πρόβλημα του αλγορίθμου k-means διότι μια αρχική επιλογή αριθμού κέντρων που δεν είναι και τόσο πετυχημένη μπορεί να οδηγήσει σε μια λάθος συσταδοποίηση.

Στη συνέχεια παρατίθεται ένα παράδειγμα εφαρμογής του αλγορίθμου των k means clustering στο απεικονιζόμενο σύνολο σημείων (σχήμα 1.3).

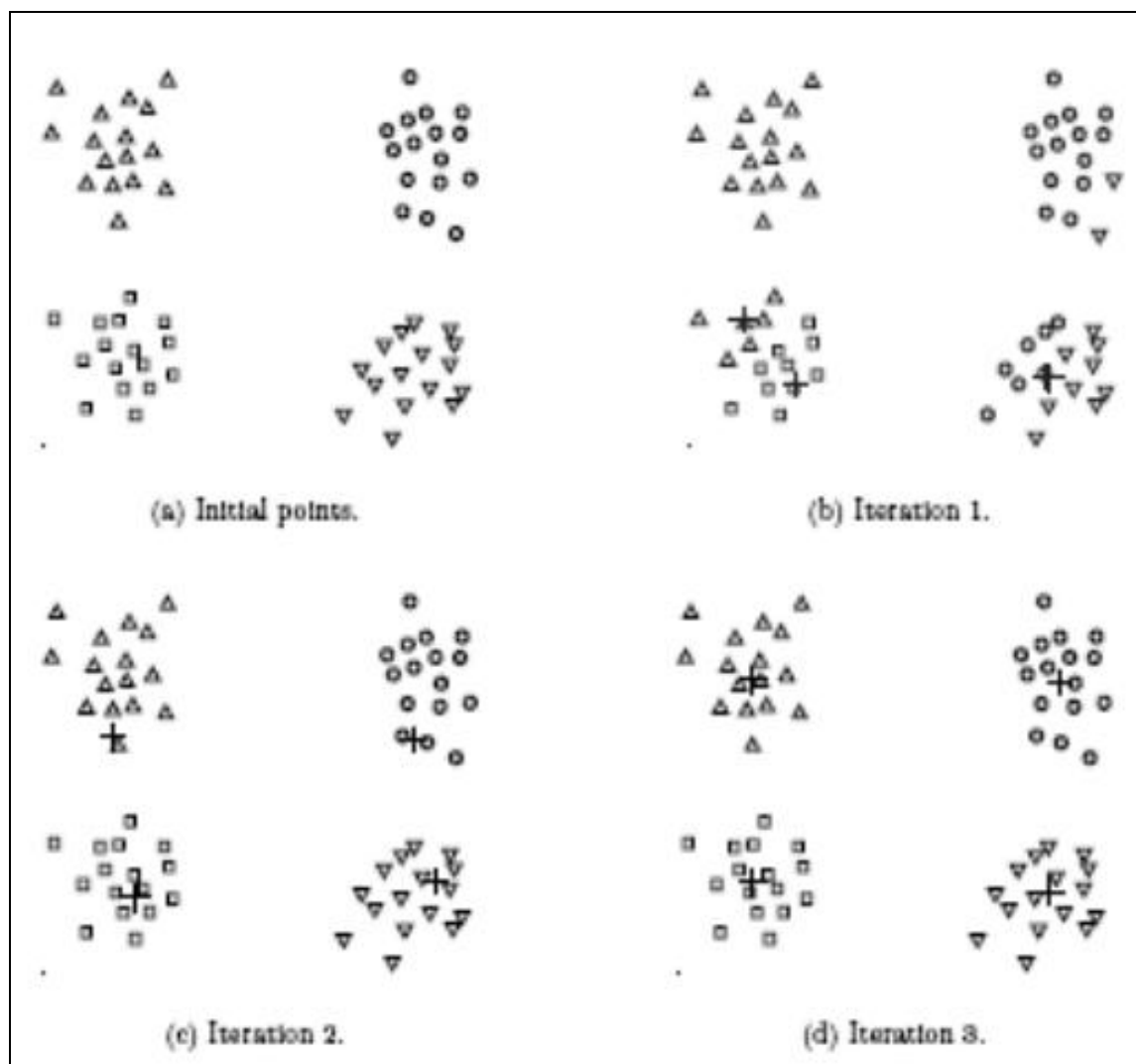


ΣΧΗΜΑ 1.3: Εφαρμογή του αλγορίθμου των k means
 ΠΗΓΗ: P.S. Bradley and U.M. Fayyad. Refining Initial Points for K-Means Clustering

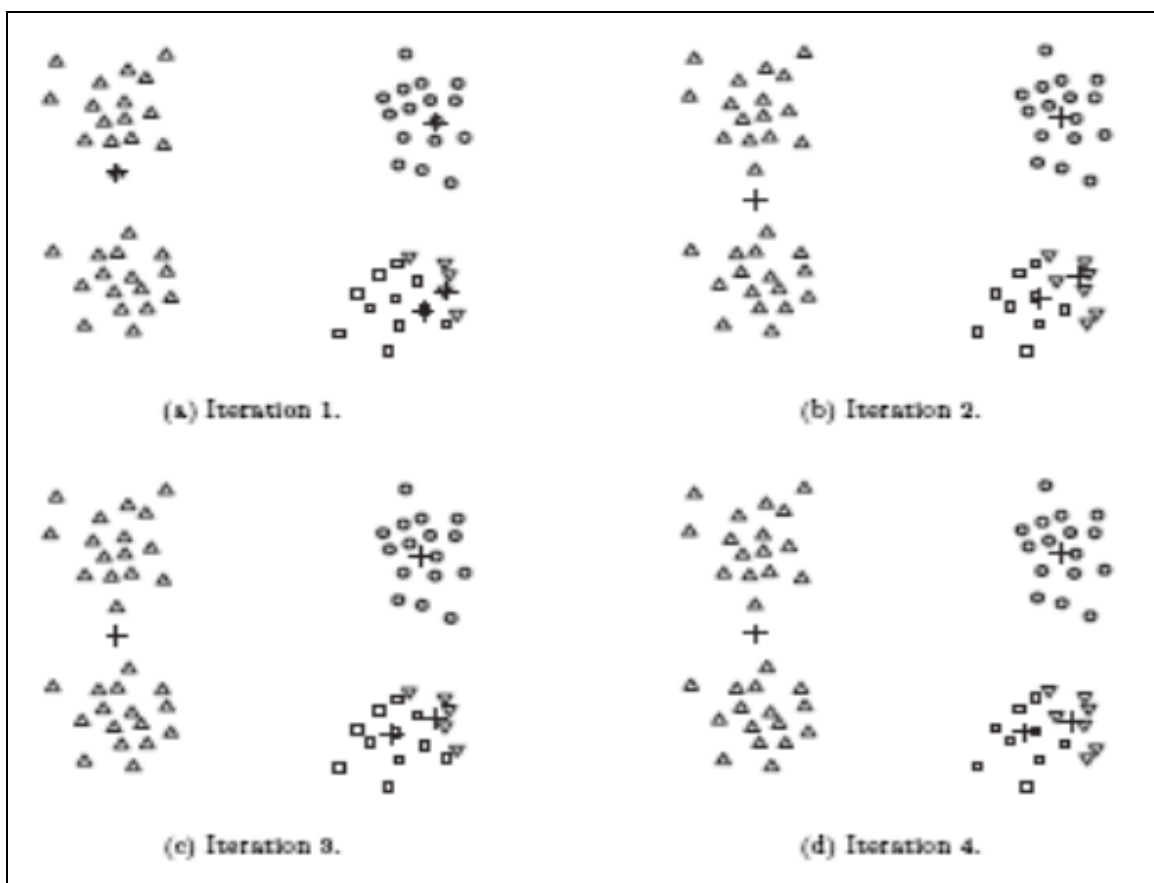
Σε ένα σχηματικό παράδειγμα ίδιου τύπου με το 1.3 μπορεί να δειχθεί η συμπεριφορά του αλγορίθμου για το διαχωρισμό του συνόλου στοιχείων που παρατίθενται σε τέσσερις κλάσεις (σχήμα 1.4).

Παρατηρείται ότι, ενώ αρχικά τοποθετούνται σε δύο κλάσεις δύο κέντρα, στη συνέχεια ο αλγόριθμος τα τοποθετεί στη σωστή τους θέση χωρίς να χάνει κάποια κλάση με την τρίτη επανάληψη.

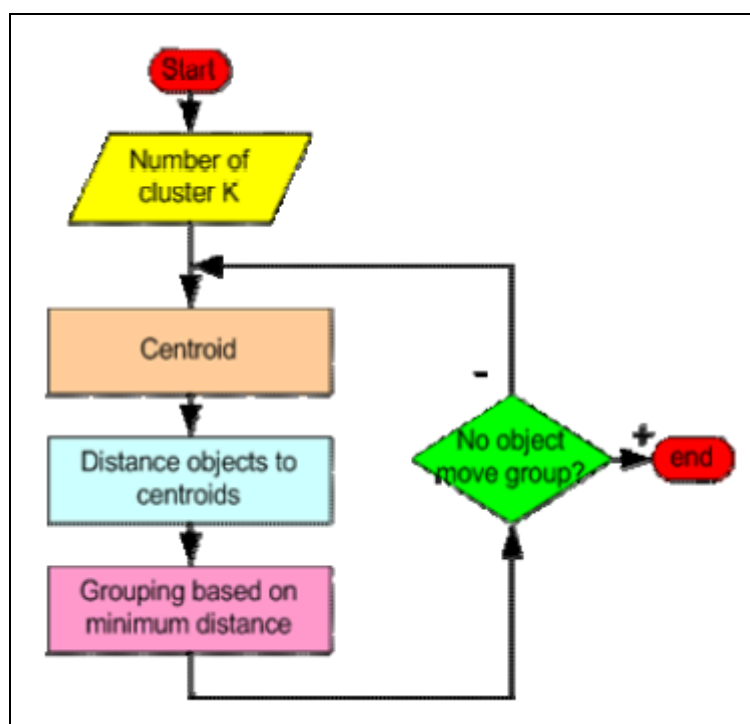
Ωστόσο δεν συμβαίνει το ίδιο με την περίπτωση διαχωρισμού των ίδιων δεδομένων με την τοποθέτηση δύο κέντρων κλάσεων σε δύο κλάσεις και άλλων δύο σε μία. Ειδικότερα όπως φαίνεται και από το σχήμα 1.5 ο αλγόριθμος χάνει μία κλάση τουλάχιστον στην τέταρτη επανάληψη.



ΣΧΗΜΑ 1.4: Εφαρμογή του αλγόριθμου για τοποθέτηση 2 κέντρων σε δύο κλάσεις
 ΠΗΓΗ: P.S. Bradley and U.M. Fayyad. Refining Initial Points for K-Means Clustering.



ΣΧΗΜΑ 1.5: Εφαρμογή του αλγόριθμου για τοποθέτηση 4 κέντρων σε τρεις κλάσεις
 ΠΗΓΗ : P.S. Bradley and U.M. Fayyad. Refining Initial Points for K-Means Clustering.



ΣΧΗΜΑ 1.6: Βήματα K-Means
 ΠΗΓΗ : P.S. Bradley and U.M. Fayyad. Refining Initial Points for K-Means Clustering.

Στο επόμενο κεφάλαιο γίνεται μια προσπάθεια εισαγωγής στις βασικές έννοιες και τις δυνατότητες των Γεωγραφικών Συστημάτων Πληροφοριών, στην εξέλιξη της αξιοποίησής τους, στην φύση και τους τύπους Συστημάτων Πληροφοριών και στις κύριες διαδικασίες τους. Επιδιωκόμενο στόχο αποτελεί η εννοιολογική αποσαφήνιση, η λειτουργική περιγραφή και οι δυνατότητες εφαρμογής των Γεωγραφικών Συστημάτων Πληροφοριών.

ΚΕΦΑΛΑΙΟ 2

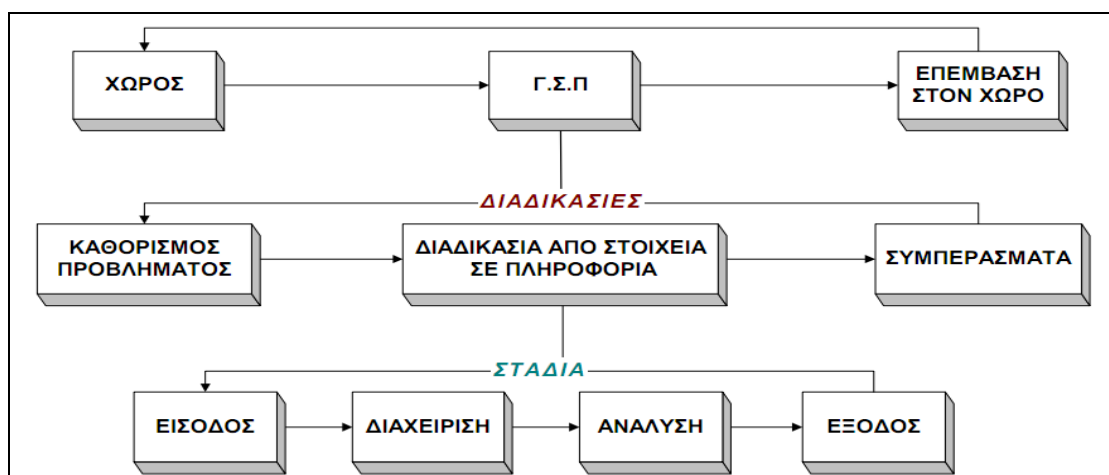
ΓΕΩΓΡΑΦΙΚΑ ΣΥΣΤΗΜΑΤΑ ΠΛΗΡΟΦΟΡΙΩΝ

2.1. Γεωγραφικά Συστήματα Πληροφοριών

Τα Γεωγραφικά Συστήματα Πληροφοριών (Geographic Information System - GIS) είναι πληροφοριακά συστήματα που επιτρέπουν την αποθήκευση και διαχείριση της γεωγραφικής πληροφορίας. Η αναγκαιότητα χρήσης των ΓΣΠ έγκειται στο γεγονός ότι όλα τα αντικείμενα εμφανίζουν γεωγραφική διάσταση. Τα ΓΣΠ επιτρέπουν, συνδυάζοντας τις δυνατότητες Βάσεων Δεδομένων και Σχεδιαστικών Προγραμμάτων, την επεξεργασία της χωρικής και περιγραφικής πληροφορίας, με τελικό αποτέλεσμα τη δημιουργία ενός δυναμικού χάρτη. Οι δυνατότητες που παρέχουν τα καθιστούν πολύτιμα εργαλεία για την επιστημονική έρευνα αλλά και σε πολλούς άλλους τομείς εφαρμογών.

Η κλασική προσέγγιση κατά την αντιμετώπιση ενός προβλήματος με τη βοήθεια των Γ.Σ.Π., αφού προηγηθεί ο καθορισμός του, περιλαμβάνει τέσσερα στάδια (Εικόνα 2.1):

- Την εισαγωγή των δεδομένων
- Την διαχείριση επεξεργασία τους
- Την ανάλυση και
- Την παρουσίασή τους



ΕΙΚΟΝΑ 2.1: Στάδια και διαδικασίες σε ένα ΓΣΠ
 ΠΗΓΗ :Κουτσόπουλος Κ. , 2002

2.1.1. Καθορισμός του προβλήματος

Στον καθορισμό του προβλήματος πρώτη ενέργεια είναι η οριοθέτηση του γενικού στόχου της μελέτης, αφού αποτελεί το βασικό άξονα προσανατολισμού της, καθώς περιέχει το σκοπό στον οποίο αποβλέπει η μελέτη και οριοθετεί το πρόβλημα προς επίλυση. (Κουτσόπουλος 2000) Πρέπει όμως να διατυπώνεται σωστά και αναλυτικά και να εστιάζεται σε υπαρκτά προβλήματα ώστε να υπάρχει η δυνατότητα εύρεσης της καταλληλότερης τεχνικής για την υλοποίησή του. Είναι βασικά το πρώτο βήμα που απαιτεί μια μελέτη με τη χρήση ΓΣΠ στην πορεία της, αλλά συγχρόνως είναι και το πιο καθοριστικό (Κουτσόπουλος 2002).

2.1.2. Εισαγωγή Δεδομένων

Στα πλαίσια της ενότητας αυτής περιγράφεται ο τρόπος με τον οποίο μπορούμε να εισάγουμε δεδομένα σε ένα ΓΣΠ. Τα δεδομένα, όπως είναι γνωστό, μπορούν να διακριθούν σε δύο βασικές κατηγορίες: τα Χωρικά δεδομένα και τα Μη Χωρικά δεδομένα. Τα χωρικά δεδομένα αναφέρονται σε πληροφορίες που αφορούν τη θέση και το σχήμα ενός φαινομένου στη γη και κατηγοριοποιούνται σε διανυσματικά και ψηφιδωτά μοντέλα. Η διανυσματική μορφή των δεδομένων δίνει έμφαση στην ύπαρξη διακριτών οντοτήτων, που μπορεί να είναι είτε σημειακές είτε γραμμικές είτε πολυγωνικές. Τα διανυσματικά δεδομένα που χρησιμοποιούνται σε ένα ΓΣΠ εισάγονται στο σύστημα μέσω της ψηφιοποίησης. Η πιο συχνή, εύχρηστη και αποτελεσματική μέθοδος ψηφιοποίησης είναι αυτή που πραγματοποιείται μέσα από τη οθόνη του υπολογιστή (heads up digitizing). Στην ψηφιδωτή μορφή, η συνεχής μορφή κατανέμεται σε ένα σύνολο απλών βασικών μονάδων (φατνία) με ακανόνιστο αλλά κυρίως κανονικό σχήμα. Τέλος, τα μη-χωρικά δεδομένα αφορούν πληροφορίες σχετικά με τα ποιοτικά και ποσοτικά χαρακτηριστικά των φαινομένων που συμβαίνουν στο χώρο και αποθηκεύονται σε μορφή πινάκων που ονομάζονται πίνακες περιγραφικών χαρακτηριστικών (Attribute Table) σε ένα αρχείο στη βάση δεδομένων. Τα ψηφιδωτά αρχεία, ως επί το πλείστον, δεν περιέχουν καμιά πληροφορία συντεταγμένων, παρά μόνο τον αριθμό και το μέγεθος των φατνίων στους δύο άξονες. Έτσι, είναι φανερό ότι σε όλες τις περιπτώσεις θα πρέπει να δημιουργηθεί γεωαναφορά στα ψηφιδωτά δεδομένα που διαθέτουμε, ορίζοντας την αντιστοιχία τους με τις γνωστές συντεταγμένες ενός άλλου χάρτη. Αφού εισαγάγουμε-καταχωρήσουμε τα δεδομένα μας σε ένα ΓΣΠ, θα πρέπει να τα οργανώσουμε με τέτοιο τρόπο ώστε να επιτρέπεται, με το μικρότερο δυνατό κόστος (σε χρήμα και χρόνο), η λήψη όλων των πληροφοριών που χρειάζονται οι χρήστες. Για να επιτευχθεί αυτό, ο τρόπος καταχώρησης των στοιχείων πρέπει να ακολουθεί τις βασικές αρχές που διέπουν την οργάνωση, διαχείριση και επεξεργασία των Βάσεων Δεδομένων.

Το επόμενο στάδιο κατά την εισαγωγή των δεδομένων αποτελεί η δόμηση της Τοπολογίας. Η τοπολογία αφορά στα χαρακτηριστικά του χώρου και συγκεκριμένα στις γεωμετρικές σχέσεις των αντικειμένων, που παραμένουν αμετάβλητες κατά την επιβολή των τοπολογικών μετασχηματισμών, δηλαδή της

μεταφοράς, στροφής και αλλαγής της κλίμακας. Με τη δόμηση της τοπολογίας ουσιαστικά δημιουργούνται οι χωρικές σχέσεις μεταξύ των γεωμετρικών οντοτήτων ενός θεματικού επιπέδου που μπορούν να αναγνωριστούν όταν κοιτάμε ένα χάρτη (συνέχεια, περιεκτικότητα, γεινίαση). Για τη δόμηση της τοπολογίας χρησιμοποιούνται τοπολογικοί κανόνες που ποικίλουν ανάλογα με τις χωρικές σχέσεις των γεωμετρικών οντοτήτων ενός γεωλογικού χάρτη ή οποιουδήποτε άλλου θεματικού επιπέδου. Αφού γίνει η δόμηση της τοπολογίας με την οποία δημιουργούνται οι χωρικές σχέσεις μεταξύ των γεωμετρικών χαρακτηριστικών του θεματικού επιπέδου, αναγνωρίζονται – εντοπίζονται τα λάθη που σχετίζονται με τη διαδικασία της ψηφιοποίησης.

Στα πλαίσια της εισαγωγής των δεδομένων εντάσσονται και τα μεταδεδομένα που συνοδεύουν τα γεωγραφικά δεδομένα και παρέχουν στους χρήστες επιπλέον πληροφορίες, σχετικά με τα ίδια τα δεδομένα. Συνήθως παρέχουν πληροφορίες σχετικές με το προβολικό σύστημα, τα όρια της περιοχής, το χρόνο και τον τρόπο δημιουργίας τους, τους περιορισμούς χρήσεως τους, την περιγραφή και τα όρια των τιμών των διαφόρων πεδίων, κλπ.

2.1.3. Διαχείριση Δεδομένων

Στη διαδικασία από στοιχεία σε πληροφορία, βασικός στόχος είναι η δημιουργία της βάσης δεδομένων (data base), που αποτελεί και την απαρχή της διαδικασίας ανάλυσης του αντικειμενικού στόχου του ΓΣΠ. Δηλαδή, η βάση δεδομένων αποτελεί τον ενδιάμεσο κρίκο μιας αλυσίδας ενεργειών, που αρχίζει από τον υπεύθυνο των αποφάσεων και καταλήγει στη διαμόρφωση των συμπερασμάτων από την ανάλυση των στοιχείων μέσα στο ΓΣΠ. Η έννοια διαχείρισης στα ΓΣΠ αφορά στον τρόπο με τον οποίο στοιχεία για τη θέση, την τοπολογία και τα χαρακτηριστικά των γεωγραφικών οντοτήτων δομούνται και οργανώνονται και επομένως, αντιστοιχεί στον όρο σύστημα διαχείρισης δεδομένων (database management system – DBMS) και αναφέρεται σε ένα λογισμικό σύστημα για τη διαχείριση (ενημέρωση, συντήρηση και ανάκτηση) των στοιχείων της βάσης δεδομένων. Κατά συνέπεια, το σύστημα διαχείρισης δεδομένων αποτελεί ένα αναπόσπαστο και ίσως το σημαντικότερο τμήμα ενός ΓΣΠ. Ιστορικά υπήρξαν δύο μορφές βάσης δεδομένων σε σχέση με χωρικά φαινόμενα και διαδικασίες. Η πρώτη μορφή αναφέρεται στην αποθήκευση πληροφορίας για κάθε ένα χαρακτηριστικό που θεωρείται αναγκαίο (single factor). Η δεύτερη μορφή είναι λιγότερο εξειδικευμένη και αναφέρεται στο τελικό αποτέλεσμα μια διαδικασίας που ορίζει ομογενείς χωρικές μονάδες (unit approach).

Σήμερα, με την εξέλιξη της πληροφορικής και του σχεδιασμού έχουν δημιουργηθεί τέσσερα βασικά μοντέλα Βάσης Δεδομένων, το ιεραρχικό, το δικτυακό, το σχεσιακό και το αντικειμενοστραφές.

2.1.4. Χωρική Ανάλυση

Η χωρική ανάλυση, εστιάζεται στην ανάπτυξη μεθόδων για την αξιολόγηση υπαρχόντων και προτεινόμενων προτύπων χωρικής οργάνωσης. Στόχος της χωρικής ανάλυσης είναι η σε βάθος γνώση της δομής του χώρου, των σχέσεων αλληλεξάρτησης και των διαδικασιών αλλαγής των διαφόρων διαστάσεων της φυσικής, κοινωνικής και οικονομικής τους διάστασης. Μέσα από τη χωρική ανάλυση, εντοπίζονται τα προβλήματα της περιοχής μελέτης, ταξινομούνται για καλύτερη διερεύνησή τους και αξιολογούνται. Από τη χωρική ανάλυση, προκύπτει η διαδικασία του χωρικού σχεδιασμού.

Σύμφωνα με τον Κουτσόπουλο (2000), είναι η 'διαδικασία από στοιχεία σε πληροφορία' ενώ όπως γράφει ο Haining (1994) η ανάλυση χώρου στοχεύει:

- Στην σωστή περιγραφή γεγονότων στο χώρο, που περιλαμβάνει κυρίως τη περιγραφή των χωρικών προτύπων.
- Στη συστηματική διερεύνηση των χωρικών προτύπων και χωρικών σχέσεων με σκοπό την καλύτερη κατανόηση των χωρικών διαδικασιών που ευθύνονται για τα χωρικά πρότυπα και τις σχέσεις που παρατηρούνται.
- Στην αύξηση της ικανότητας πρόβλεψης και ελέγχου γεγονότων που συμβαίνουν στο γεωγραφικό χώρο.
- Στη χρήση αυτών των τεχνικών και μεθόδων ως εργαλεία λήψης αποφάσεων για το χώρο

2.1.5. Παρουσίαση Πληροφορίας

Η έξοδος από τον Η/Υ και ο τρόπος που θα παρουσιαστεί η πληροφορία που η ανάλυση και γενικά το ΓΣΠ δημιούργησε, είναι καθοριστικά για την αποτελεσματικότητά του. Επομένως, η παρουσίαση της πληροφορίας είναι πρωταρχικής σημασίας για κάθε ΓΣΠ. Οι βασικές μορφές εξόδου της πληροφορίας είναι τρεις και περιλαμβάνουν:

- Μη σχεδιαστικές αποδόσεις όπως πίνακες, μαθηματικές συναρτήσεις, μέσοι όροι
- Μορφές γραφημάτων όπως τα ιστογράμματα, τα πολύγωνα συχνότητας
- Χάρτες

Το κύριο μέσο μετάδοσης της επεξεργασμένης πληροφορίας ενός ΓΣΠ είναι ο χάρτης, καθώς σε όλες του τις μορφές και τις διαστάσεις παρέχει μια άμεση εποπτεία στα χωρικά φαινόμενα, με ελεγμένη ακρίβεια και πληρότητα και παραμένει ένας βασικός τρόπος επικοινωνίας. Οι θεματικοί χάρτες έχουν ιδιαίτερη σημασία σαν μορφή εξόδου των ΓΣΠ, καθώς μπορούν να χρησιμοποιηθούν στην απεικόνιση τόσο των φυσικών φαινομένων όσο και των

φαινομένων που σχετίζονται με τις ανθρώπινες δραστηριότητες. Η μετάδοση και επικοινωνία συγκεκριμένων ιδεών μέσα από τους θεματικούς χάρτες είναι περισσότερο αποτελεσματική, γιατί η ανθρώπινη αντίληψη είναι πιο άμεση στις εικόνες παρά στα πινακοποιημένα στοιχεία. Πρέπει να σημειωθεί ότι τα ΓΣΠ παρέχουν τη δυνατότητα απεικόνισης όλων των στοιχείων που συνθέτουν έναν χάρτη, ήτοι στοιχεία εδάφους, ιδιότητες του γεωγραφικού χώρου, υπομνήματα, κλίμακες και μια αρκετά μεγάλη γκάμα θεματικού συμβολισμού, δηλαδή την κατασκευή χαρτογραφικών συνθέσεων.

2.2. Τα Γεωγραφικά Συστήματα Πληροφοριών στην εκπαίδευση

Αναμφισβήτητα τα τελευταία χρόνια οι εφαρμογές των ΓΣΠ έχουν σημειώσει αλματώδη ανάπτυξη και έχουν διεισδύσει σε όλο το φάσμα των καθημερινών μας δραστηριοτήτων. Η ανάπτυξη αυτή που συντελείται με γεωμετρική πρόοδο, οφείλεται στη δημιουργία αφενός ενός τεράστιου αριθμού Γεωγραφικών βάσεων δεδομένων σε όλο τον κόσμο και αφετέρου στην ανάπτυξη του διαδικτύου και των χαρτογραφικών κόμβων, γεγονός που δίνει τη δυνατότητα στον καθένα από εμάς, από τον υπολογιστή του στο σπίτι να βρει πολύ εύκολα την επιθυμητή γεωγραφική πληροφορία με τη χρησιμοποίηση μόνο κάποιου φυλλομετρητή διαδικτύου. (explorer, firefox κτλ.)

Η ανάπτυξη όμως αυτή των ΓΣΠ, δημιούργησε επιτακτικά την ανάγκη για εκπαίδευση στις βασικές αρχές των ΓΣΠ των προπτυχιακών και μεταπτυχιακών σπουδαστών των πανεπιστημίων και όχι μόνο.

Αν και η σημασία και συνεισφορά των Γεωγραφικών Συστημάτων Πληροφοριών σε πολλές εφαρμογές και κρίσιμες αποφάσεις έχει αναγνωριστεί σε παγκόσμιο επίπεδο, εντούτοις στη χώρα μας η χρησιμοποίησή τους περιορίζεται σε λίγα πανεπιστημιακά και ερευνητικά κέντρα, στις υπηρεσίες ορισμένων υπουργείων, σε κάποιους οργανισμούς τοπικής αυτοδιοίκησης και σε ορισμένες ιδιωτικές επιχειρήσεις. (Polidorides, 1992),(Assimacopoulos, 1992 and 1993), (Μανιάτης, 1993)

Ειδικότερα στον τομέα της εκπαίδευσης άρχισε να εμφανίζεται τα τελευταία χρόνια όλο και πιο συχνά στα προγράμματα σπουδών ορισμένων τμημάτων Α.Ε.Ι., όχι απαραίτητα τεχνικού περιεχομένου εκπαίδευσης, σε μεταπτυχιακά προγράμματα Γεωγραφίας, Γεωπληροφορικής, Περιβάλλοντος, αλλά και σε προγράμματα εξ αποστάσεως εκπαίδευσης, τροφοδοτώντας με εξειδικευμένα άτομα την παραγωγή και την έρευνα.

Πλέον η εκπαίδευση των GIS στα πανεπιστημιακά ιδρύματα που εφαρμόζεται κατέχει ενεργό ρόλο στην υποστήριξη της φοίτησης, καλλιεργώντας την κριτική σκέψη, αλλά και τις ικανότητες στην επίλυση διαφόρων προβλημάτων. Έτσι στα μαθήματα των φυσικών επιστημών στη εκπαίδευση, τα ΓΣΠ μπορούν να συνδυάσουν πληροφορίες από διάφορες πηγές, να τις οργανώσουν κατάλληλα και να τις αποτυπώσουν πάνω σε χάρτες προσφέροντας οπτική απεικόνιση και κάνοντας τα προσιτά και κατανοητά ακόμη και σε μη ειδικό ακροατήριο.

Ειδικότερα στον τομέα των μηχανικών, η αυξανόμενη ανάγκη για διαχείριση, ανάλυση, επεξεργασία και παρουσίαση των διαφόρων γεωγραφικών δεδομένων, έχει κάνει τη χρήση των ΓΣΠ απαραίτητη στις περισσότερες γεωγραφικές αλλά και χωρικές μελέτες.

Η χρήση ΓΣΠ στις συγκεκριμένες εφαρμογές, εκτός του ότι συμβάλλει στην καλύτερη κατανόηση και παρουσίαση των αποτελεσμάτων της ταξινόμησης μπορεί να θεωρηθεί ότι βοηθά:

- Στην εξοικείωση σπουδαστών και φοιτητών με ένα σύγχρονο λογισμικό πακέτο ΓΣΠ,
- Στην ανάπτυξη ικανότητας ερμηνείας χαρτών και στην επέμβαση σε αυτούς με τα διάφορα εργαλεία που προσφέρει το λογισμικό,
- Στη χρήση διαφόρων χαρτογραφικών συμβολισμών και στον πειραματισμό με αυτούς, ώστε να προκύψει ένα κατά το δυνατόν βέλτιστο οπτικά προϊόν,
- Στην εισαγωγή των σπουδαστών στην ανάλυση του χώρου μέσα από ρεαλιστικά παραδείγματα και στην αντιμετώπιση των διαφόρων προβλημάτων που προκύπτουν.
- Τέλος, με τη χρήση των ΓΣΠ μειώνεται στο ελάχιστο ο χρόνος εφαρμογής, επιτρέποντας στους φοιτητές να εμβαθύνουν περισσότερο στο θεωρητικό κομμάτι της διαδικασίας της εφαρμογής.

Στα δύο επόμενα κεφάλαια παρουσιάζονται δύο παραδείγματα εφαρμογών. Το πρώτο αποτελεί μία τυπική ταξινόμηση των γεωγραφικών διαμερισμάτων του Ελλαδικού χώρου, με βάση τον αριθμό των διαπραχθέντων αδικημάτων, του έτους 2009. Στο δεύτερο παράδειγμα, κατηγοριοποιούνται σε τρεις τάξεις δώδεκα γειτονικοί δήμοι της Νότιας Αττικής, αναλόγως με το ποσοστό εμφάνισης δημογραφικών μεταβλητών.

ΚΕΦΑΛΑΙΟ 3

ΕΦΑΡΜΟΓΗ ΤΑΞΙΝΟΜΗΣΗΣ ΓΕΩΓΡΑΦΙΚΩΝ ΔΙΑΜΕΡΙΣΜΑΤΩΝ

Στο παρόν κεφάλαιο αρχικά παρουσιάζονται δεδομένα που αφορούν στον αριθμό αδικημάτων, τα οποία διαπράχθηκαν το έτος 2009 στην Ελλάδα. Η χωρική μονάδα αναφοράς των δεδομένων είναι το γεωγραφικό διαμέρισμα. Εν συνεχεία τα δεδομένα επεξεργάζονται σε περιβάλλον ΓΣΠ και τελικά πραγματοποιείται η δημιουργία τριών τάξεων με χρήση στατιστικής ταξινόμησης, και συγκεκριμένα με τον αλγόριθμο K-Means.

3.1. Προσδιορισμός του προβλήματος

Αρχικά έγινε ο προσδιορισμός του προβλήματος το οποίο αφορά στην ταξινόμηση των γεωγραφικών διαμερισμάτων της Ελλάδας, με βάση τον αριθμό των διαπραχθέντων αδικημάτων που πραγματοποιήθηκαν το έτος 2009.

3.2. Εισαγωγή και Διαχείριση δεδομένων

Τα στοιχεία που αφορούν στα αδικήματα ανά γεωγραφικό διαμέρισμα αντλήθηκαν από την ιστοσελίδα της Ελληνικής Στατιστικής Αρχής (ΕΛ.ΣΤΑΤ.) και φαίνονται στον Πίνακα 3.1. Στη συνέχεια παρουσιάζεται αναλυτικά η διαδικασία κατά την οποία τα στατιστικά δεδομένα, εισήχθησαν και επεξεργάστηκαν σε περιβάλλον εργασίας ΓΣΠ.

3.2.1. Σχεδιασμός Γεωγραφικής Βάσης Δεδομένων

Το πρώτο βήμα που αποσκοπεί στην ανάπτυξη της γεωγραφικής βάσης δεδομένων είναι ο προσδιορισμός των απαραίτητων θεματικών επιπέδων. Ο σωστός σχεδιασμός της βάσης εξασφαλίζει την πληρότητα και τη διαθεσιμότητα των απαραίτητων γεωγραφικών και περιγραφικών χαρακτηριστικών πριν από το στάδιο της ανάλυσης.

	κατά περιουσιακών δικαιωμάτων	κατά της ιδιοκτησίας	περί τον γάμο και την οικογένεια	κατά της προσωπικής ελευθερίας	Σωματικές βλάβες	κατά της ζωής	περί την απονομή δικαιοσύνης	περί το νόμισμα
Αττική	1.075	60.824	145	3.030	11.029	527	1.237	731
Λοιπή Στερεά και Εύβοια	67	4.162	19	491	1.071	138	141	148
Πελοπόννησος	93	7.526	46	1.227	2.023	211	251	399
Νησιά Ιονίου	23	978	15	475	330	20	66	62
Ήπειρος	15	740	2	258	213	38	51	25
Θεσσαλία	41	1.647	164	393	730	74	238	207
Μακεδονία	307	20.916	217	1.065	3.188	213	505	1.247
Θράκη	22	820	6	131	409	33	53	135
Νησιά Αιγαίου	31	1.482	18	314	524	52	81	188
Κρήτη	62	3.640	75	616	577	103	153	208

ΠΙΝΑΚΑΣ 3.1: Διαπραχθέντα αδικήματα κατά κατηγορίες και γεωγραφικό διαμέρισμα τελέσεως αυτών. Έτος 2009, ΠΗΓΗ: ΕΛ.ΣΤΑΤ.

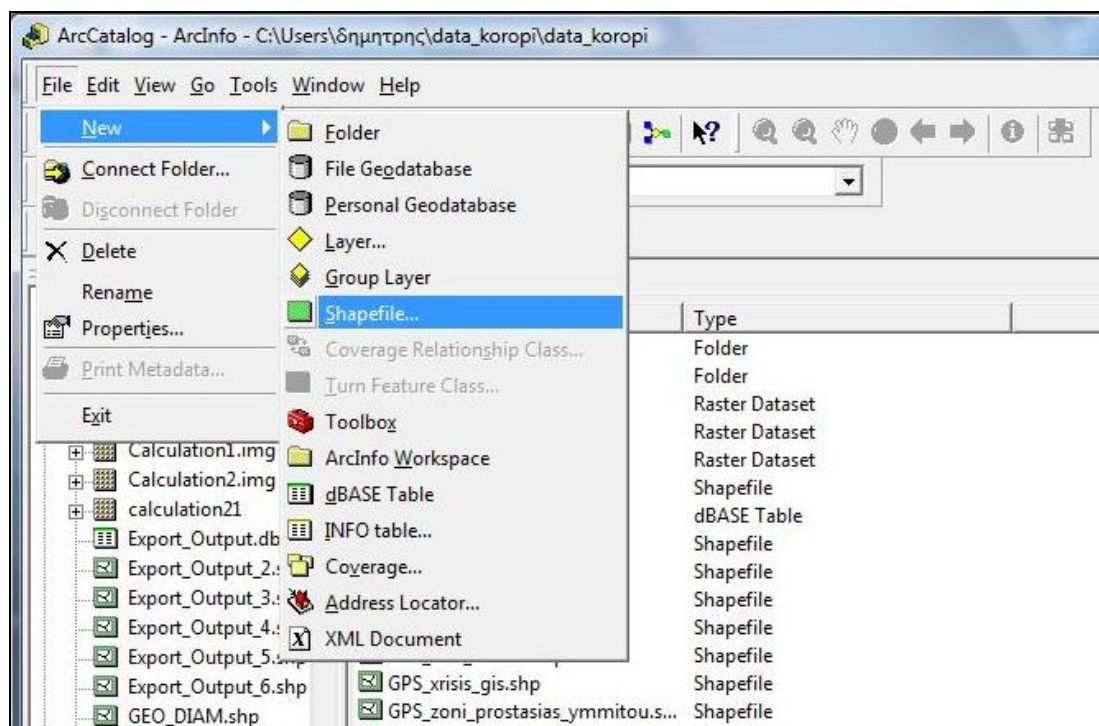
3.2.2. Προσδιορισμός θεματικού επιπέδου

Στη συγκεκριμένη εφαρμογή χρειάστηκε ένα επίπεδο πληροφορίας, το οποίο αφορά στα γεωγραφικά διαμερίσματα του Ελλαδικού χώρου.

Το θεματικό επίπεδο αποτελείται από ψηφιοποιημένα πολύγωνα που αντιστοιχούν στα γεωγραφικά διαμερίσματα. Με αυτόν τον τρόπο καθίσταται στη συνέχεια δυνατή η σύνδεση των πολυγώνων με τα χαρακτηριστικά τους και κατά επέκταση η γεωγραφική απεικόνισή τους.

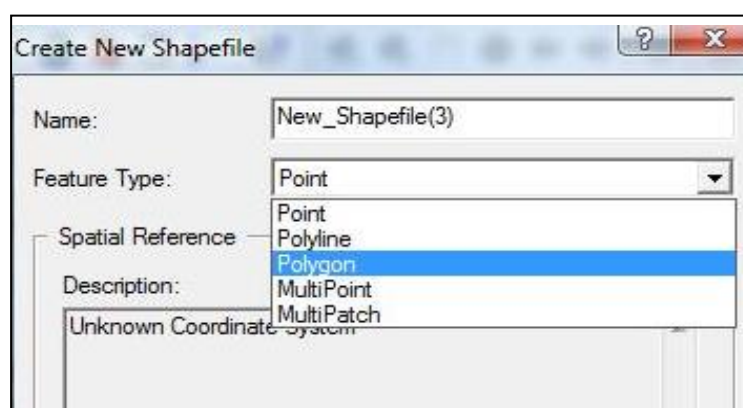
3.2.3. Ψηφιοποίηση θεματικού επιπέδου

Κατά τη διαδικασία της ψηφιοποίησης πρώτο βήμα αποτέλεσε η δημιουργία αρχείου Shapefile σε περιβάλλον ArcCatalog με το χαρακτηριστικό όνομα “Γεωγραφικά διαμερίσματα” (Εικόνα 3.1).




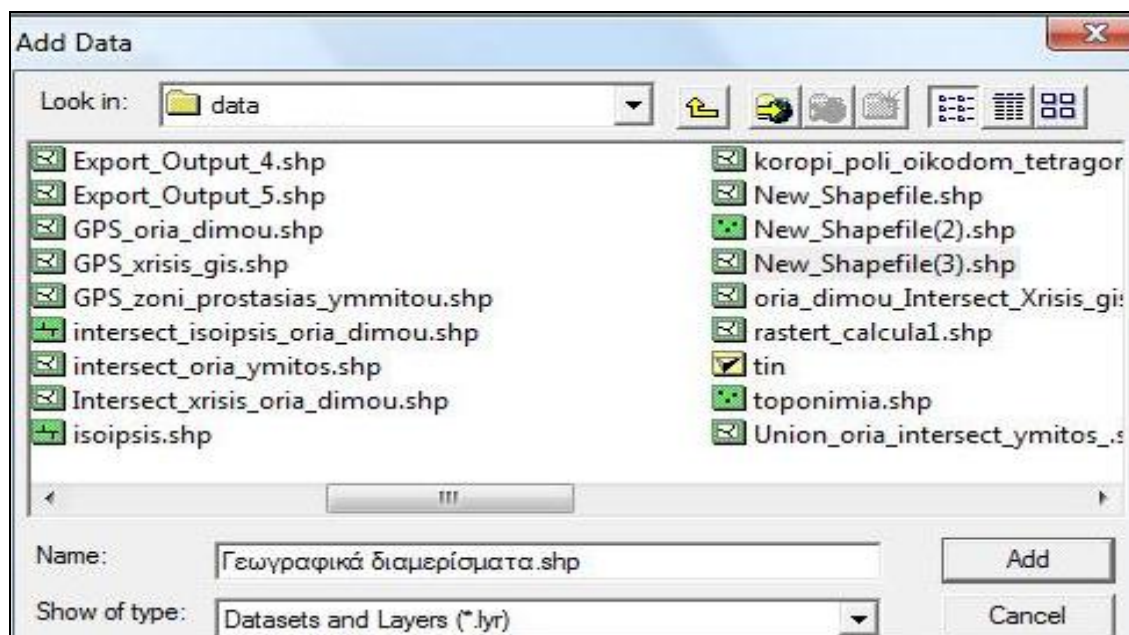
ΕΙΚΟΝΑ 3.1: Δημιουργία Shapefile στο ArcCatalog

Εν συνεχεία, δόθηκε στο νέο αρχείο η κατάλληλη μορφή και εφόσον επρόκειτο να ψηφιοποιηθούν γεωγραφικά διαμερίσματα, επιλέχθηκε ο πολυγωνικός τύπος αρχείου (Εικόνα 3.2).



ΕΙΚΟΝΑ 3.2: Προσδιορισμός κατάλληλης μορφής στο νέο αρχείο

Για να είναι δυνατή η ψηφιοποίηση, το νέο πολυγωνικό αρχείο εισήχθη μέσω της εντολής Add Data  στο περιβάλλον του ArcMap και αποτέλεσε το θεματικό επίπεδο “Γεωγραφικά διαμερίσματα”(Εικόνα 3.3).



ΕΙΚΟΝΑ 3.3:Εισαγωγή αρχείων στο ArcMap

Με τον ίδιο τρόπο εισήχθη ο χάρτης των γεωγραφικών διαμερισμάτων της Ελλάδας(Εικόνα 3.4), για να αποτελέσει το υπόβαθρο της ψηφιοποίησης.



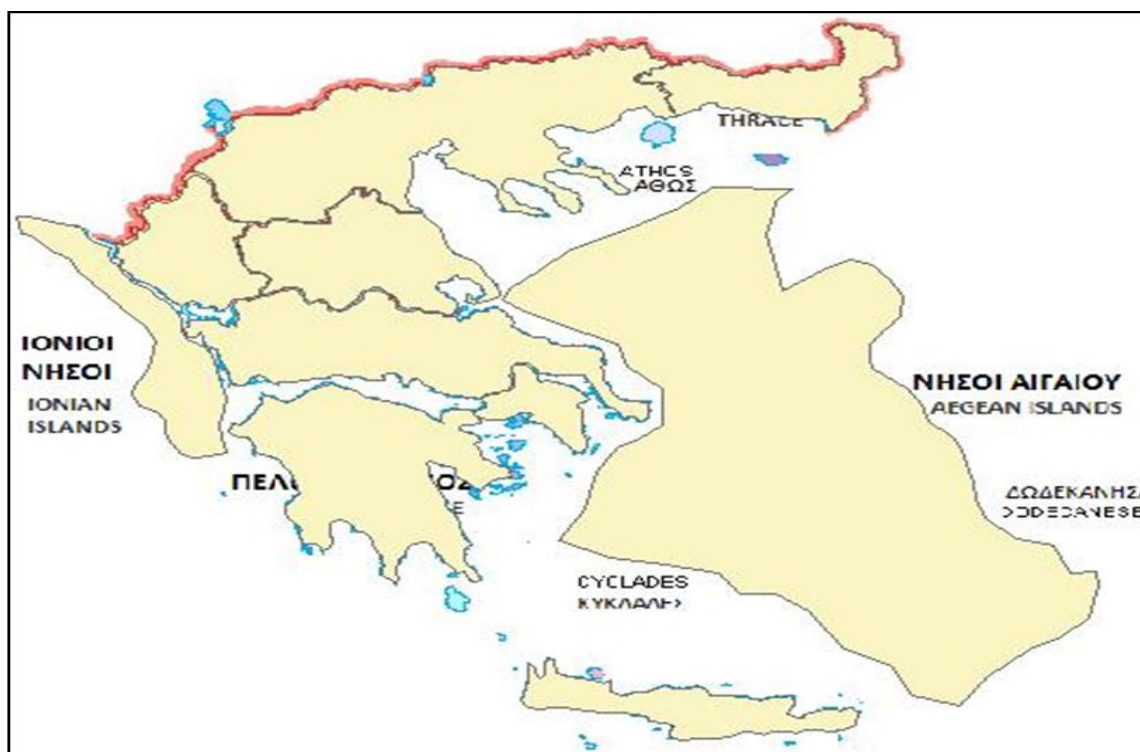
ΕΙΚΟΝΑ 3.4:Χάρτης γεωγραφικών διαμερισμάτων
 ΠΗΓΗ:http://stin-e-taxi.blogspot.gr/2011/05/blog-post_13.html

Στη συνέχεια με τη βοήθεια της εργαλειοθήκης του Editor και επιλέγοντας `start editing` πραγματοποιήθηκε η διαδικασία της ψηφιοποίησης με τη σειρά που φαίνεται στον πίνακα 3.2, έτσι ώστε να συμβαδίζει με τον πίνακα 3.1(σελ.32) των διαπραχθέντων αδικημάτων ανά γεωγραφικό διαμέρισμα.

ID ΠΟΛΥΓΩΝΟΥ	ΓΕΩΓΡΑΦΙΚΟ ΔΙΑΜΕΡΙΣΜΑ
Polygon 1	Αττική
Polygon 2	Λοιπή Στερεά και Εύβοια
Polygon 3	Πελοπόννησος
Polygon 4	Νησιά Ιονίου
Polygon 5	Ήπειρος
Polygon 6	Θεσσαλία
Polygon 7	Μακεδονία
Polygon 8	Θράκη
Polygon 9	Νησιά Αιγαίου
Polygon 10	Κρήτη

ΠΙΝΑΚΑΣ 3.2:Σειρά ψηφιοποίησης πολυγώνων

Τα πολύγωνα αποθηκεύτηκαν στο θεματικό επίπεδο με την εντολή `save edits` ενώ η διαδικασία της ψηφιοποίησης ολοκληρώθηκε με την εντολή `stop editing`. Τα αποτελέσματα της ψηφιοποίησης φαίνονται στην εικόνα 3.5.



ΕΙΚΟΝΑ 3.5:Αποτέλεσμα της ψηφιοποίησης στο ArcMap

3.2.4. Καταχώρηση πληροφορίας στα πολύγωνα

Στο συγκεκριμένο στάδιο πραγματοποιήθηκε η σύνδεση ανάμεσα στο ψηφιοποιημένο χάρτη και τον πίνακα, ο οποίος περιέχει τα διαπραχθέντα αδικήματα ανά γεωγραφικό διαμέρισμα. Η διαδικασία ολοκληρώθηκε στο περιβάλλον εργασίας ArcMap χρησιμοποιώντας την εντολή join. Με αυτόν τον τρόπο στο κάθε πολύγωνο-διαμέρισμα αντιστοιχήθηκαν τα χαρακτηριστικά του. Στο τέταρτο κεφάλαιο παρουσιάζεται αναλυτικά ο τρόπος που πραγματοποιήθηκε η σύνδεση.

3.2.5. Παρουσίαση πληροφορίας

Η έξοδος από τον ηλεκτρονικό υπολογιστή και ο τρόπος που θα παρουσιαστεί η πληροφορία που η ανάλυση στο Γεωγραφικό Σύστημα Πληροφοριών (ΓΣΠ) δημιούργησε, είναι καθοριστικά για την αποτελεσματικότητά του.

Το κύριο μέσο μετάδοσης της επεξεργασμένης πληροφορίας ενός ΓΣΠ είναι ο χάρτης, καθώς σε όλες τις μορφές και τις διαστάσεις του παρέχει μια άμεση εποπτεία στα χωρικά φαινόμενα, με ελεγμένη ακρίβεια και πληρότητα. Οι θεματικοί χάρτες έχουν ιδιαίτερη σημασία σαν μορφή εξόδου των ΓΣΠ, καθώς μπορούν να χρησιμοποιηθούν τόσο στην απεικόνιση των φυσικών φαινομένων όσο και των φαινομένων που σχετίζονται με τις ανθρώπινες δραστηριότητες. Η

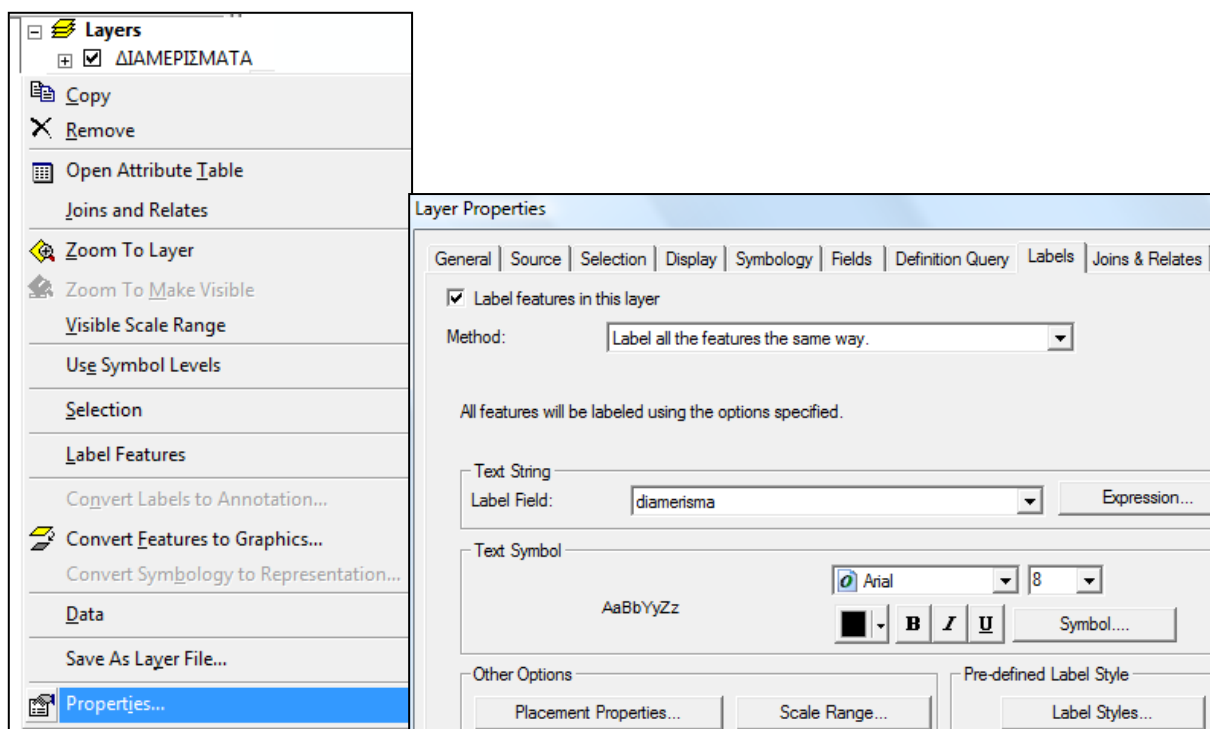
μετάδοση και επικοινωνία συγκεκριμένων ιδεών μέσα από τους θεματικούς χάρτες είναι περισσότερο αποτελεσματική, διότι η ανθρώπινη αντίληψη είναι πιο εξοικειωμένη στις εικόνες παρά στα πινακοποιημένα στοιχεία.

3.2.6 . Απεικόνιση δεδομένων

Στο συγκεκριμένο στάδιο, η βάση δεδομένων είναι ικανή να απεικονίσει την πληροφορία των πινάκων excel γραφικά με τη μορφή χάρτη.

Η προεργασία για να παραχθούν οι χάρτες πραγματοποιήθηκε στο περιβάλλον εργασίας “ArcMap” με τη βοήθεια των ιδιοτήτων του αντίστοιχου θεματικού επιπέδου (Layer Properties) και συγκεκριμένα με την ιδιότητα του συμβολισμού (Symprology) καθώς και αυτήν της ονομασίας (Labels).

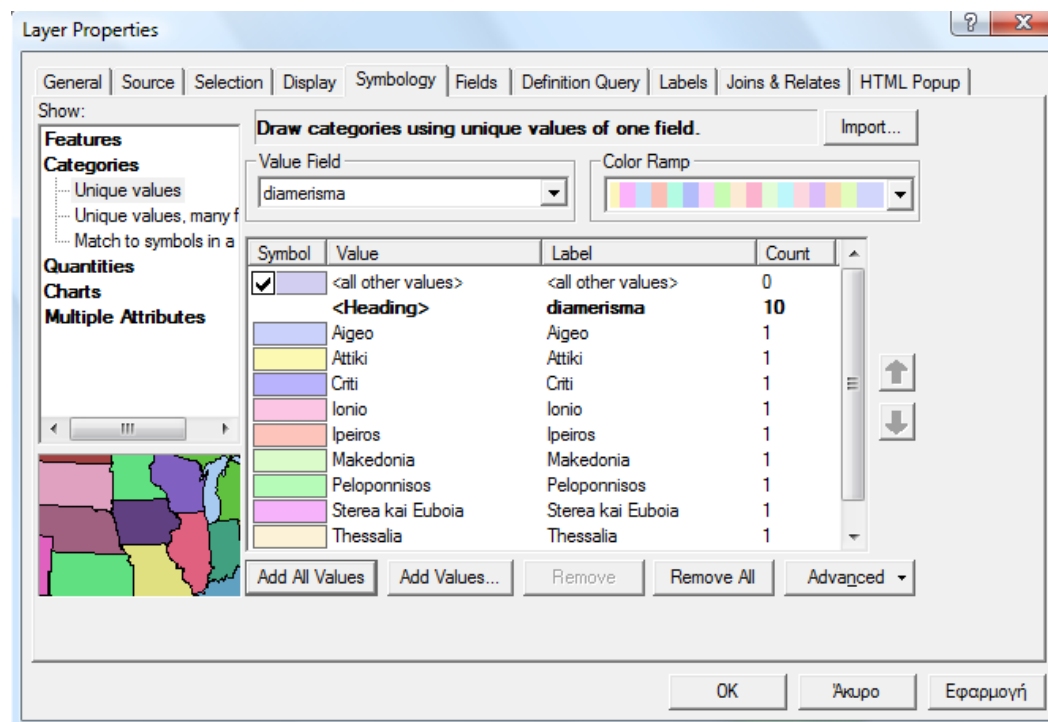
Παρακάτω, θα αναλυθεί η διαδικασία για τη σύνταξη ενός χάρτη που απεικονίζει τα γεωγραφικά διαμερίσματα βάσει της ονομασίας. Επιλέγοντας το θεματικό επίπεδο “Διαμερίσματα” και στη συνέχεια την επιλογή “Properties” αναπτύσσεται στην οθόνη το παράθυρο διαλόγου “Layer Properties”(Εικόνα 3.6):



ΕΙΚΟΝΑ 3.6:Εισαγωγή και επεξεργασία στο παράθυρο διαλόγου Layer Properties

Αφού επιλέχθηκε η καρτέλα “Labels” εμφανίστηκε στην οθόνη η εικόνα 3.6 και στο πλαίσιο “Label Field” επιλέχθηκε το πεδίο που αναφέρεται στην ονομασία ή την αρίθμηση που μας ενδιαφέρει, στην παρούσα περίπτωση, την ονομασία των γεωγραφικών διαμερισμάτων της Ελλάδας, “diamerisma”. Με την επιλογή Label features in this layer επιλέχθηκε ή όχι η αναγραφή της ονομασίας του πεδίου εντός των πολυγώνων.

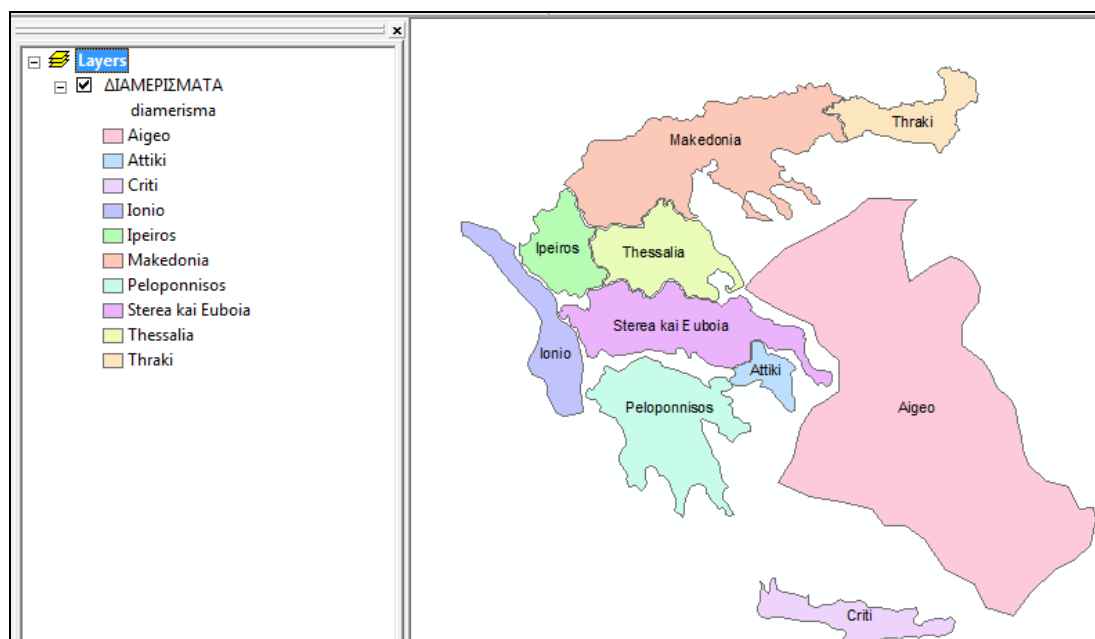
Στη συνέχεια, για την απόδοση ποιοτικού και ποσοτικού συμβολισμού στους χάρτες, στη φόρμα διαλόγου “Layer Properties” επιλέχθηκε η καρτέλα ιδιοτήτων συμβολισμού (Symbology).



ΕΙΚΟΝΑ 3.7: Διαδικασία απόδοσης συμβολισμού.

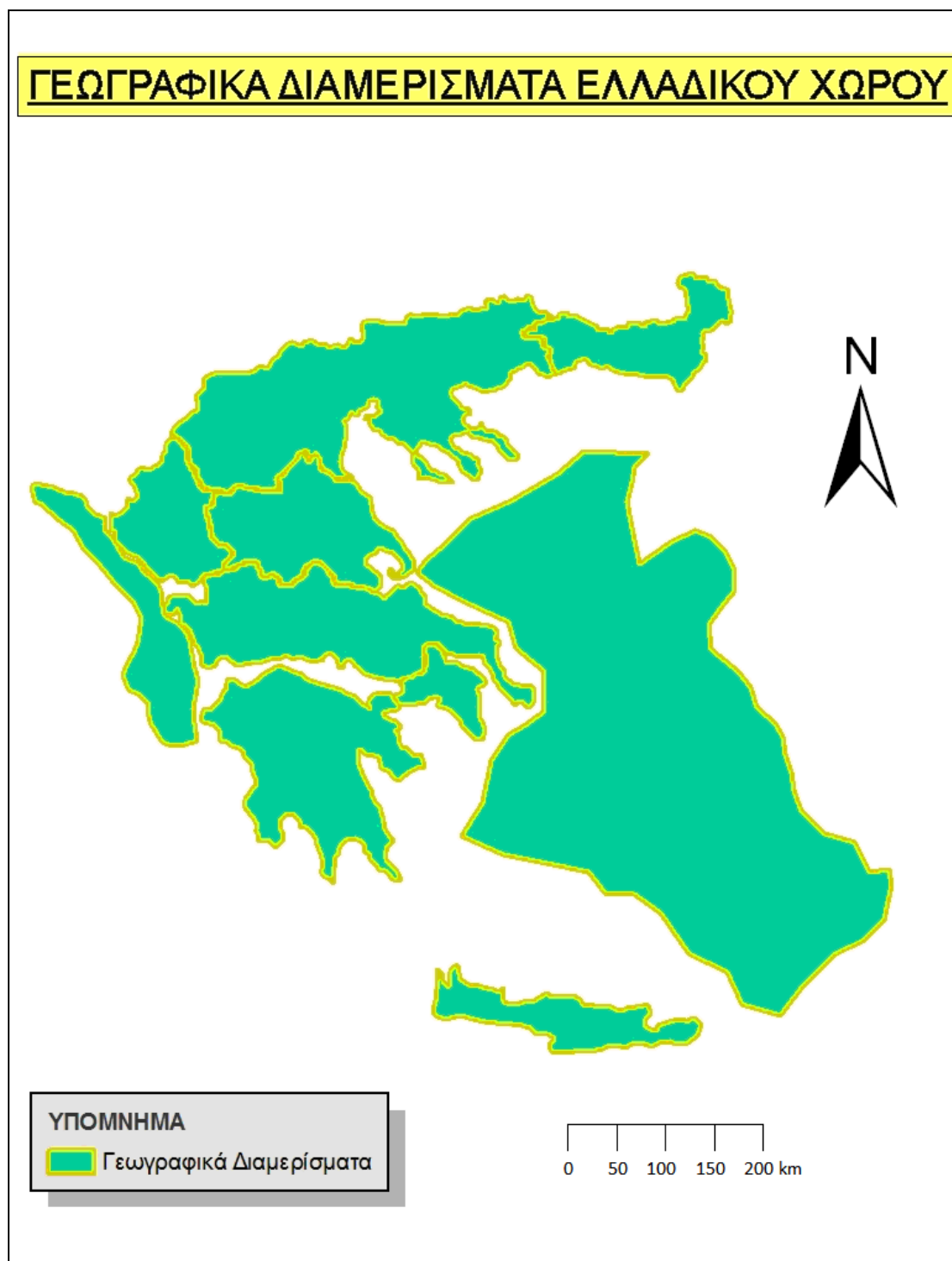
Στο πεδίο “Show” επιλέχτηκε Categories> Unique Values ενώ στο πεδίο “Value Field” “diamerisma”(Εικόνα 3.7). Το συγκεκριμένο πεδίο επιλέχθηκε διότι περιέχει την πληροφορία της ονομασίας των πολυγώνων, βάσει της οποίας έγινε η χρωματική κατηγοριοποίηση των πολυγώνων. Ο χρωματισμός προσδιορίστηκε με τη βοήθεια του πεδίου “Color Ramp”. Τέλος, πληκτρολογώντας “Add All Values” κατηγοριοποιήθηκαν όλα τα γεωγραφικά διαμερίσματα του θεματικού επιπέδου”(Εικόνα 3.7).

Ακολουθεί το αποτέλεσμα της διαδικασίας που περιγράφηκε (Εικόνα 3.8).

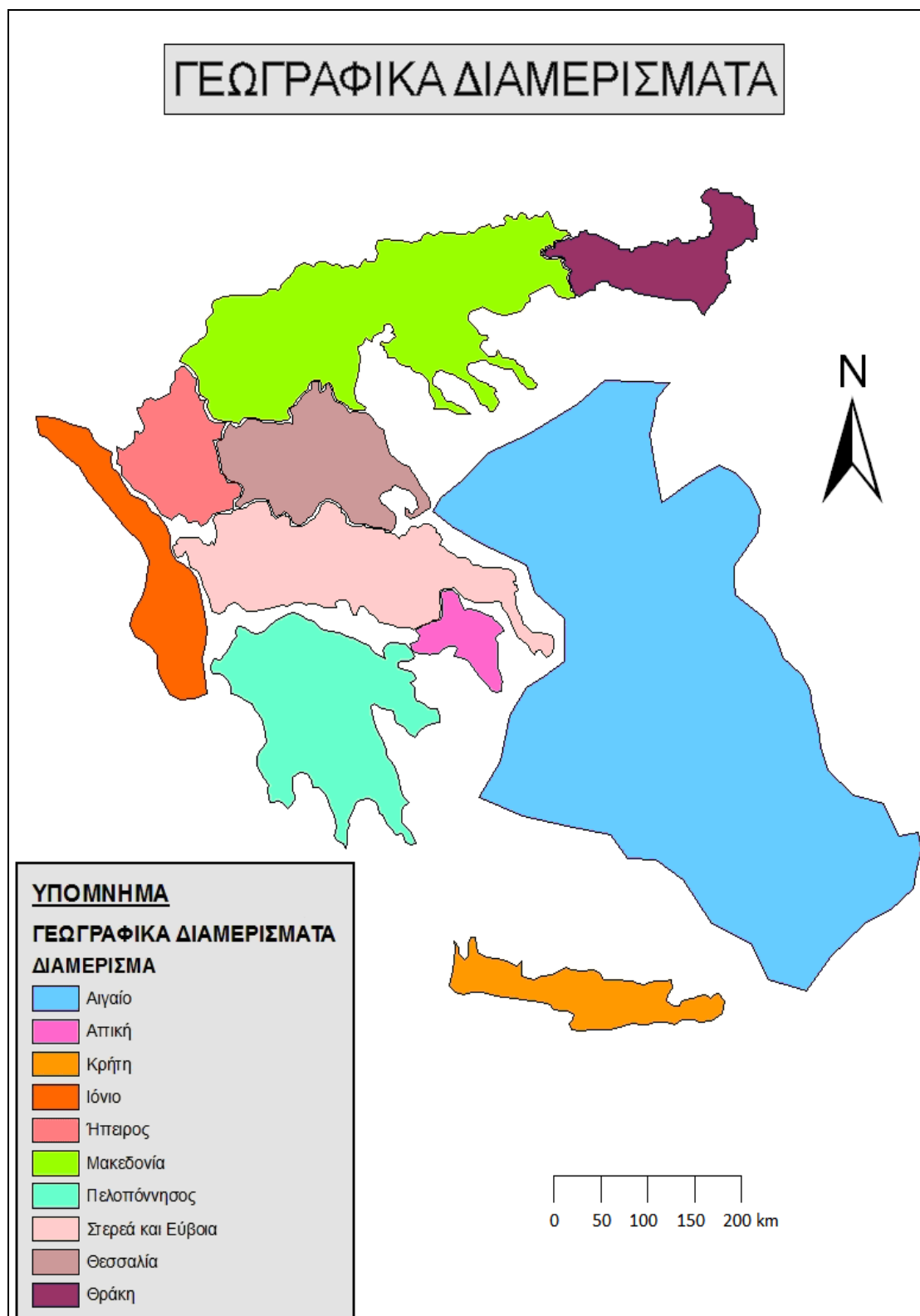


ΕΙΚΟΝΑ 3.8: Οπτικοποίηση της πληροφορίας σε περιβάλλον ArcMap

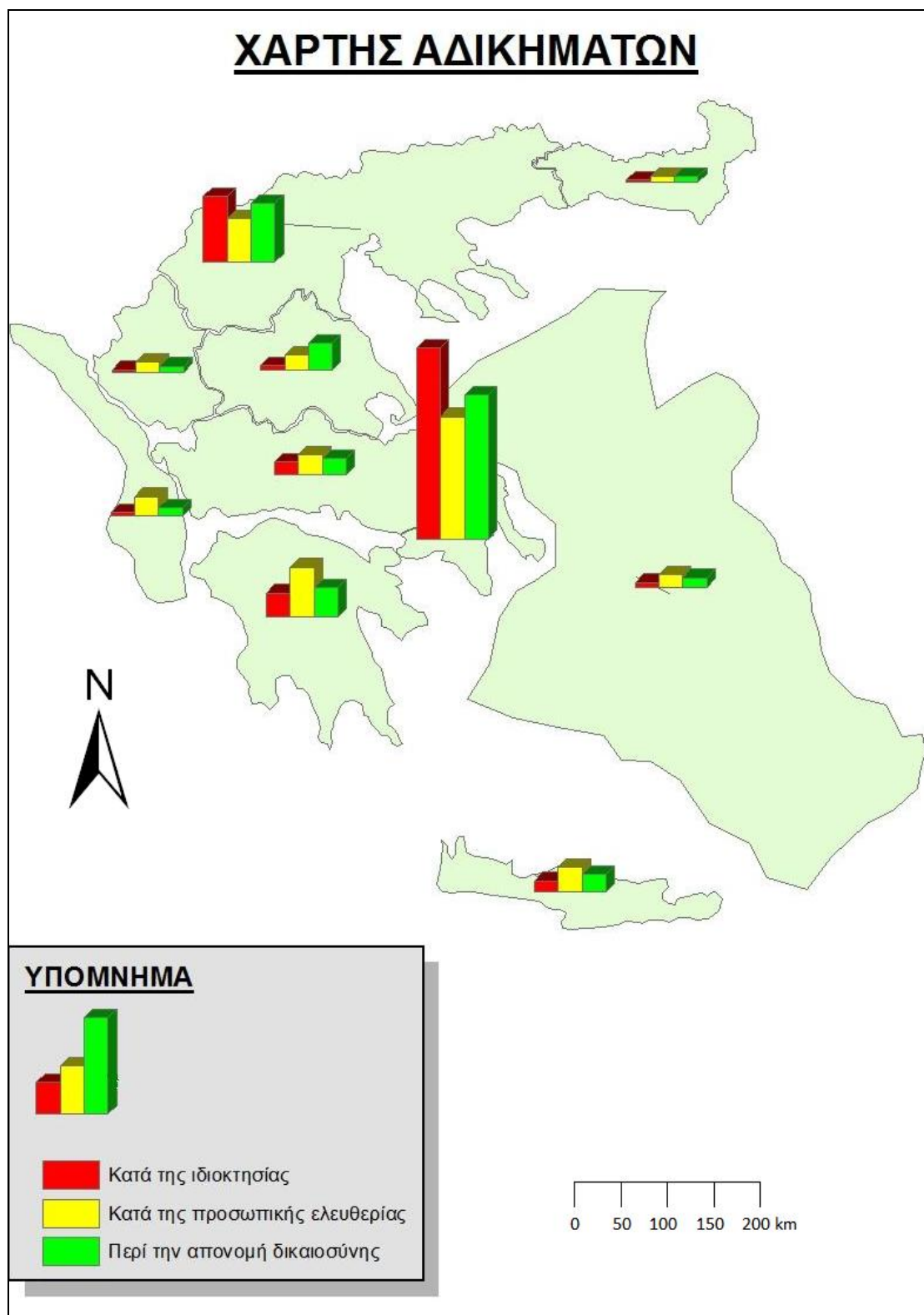
Είναι προφανές ότι η παραπάνω εικόνα δεν αποτελεί τον τελικό χάρτη αλλά απεικόνιση της πληροφορίας στο περιβάλλον ArcMap. Οι παράγωγοι χάρτες της εφαρμογής που αφορά στα διαπραχθέντα αδικήματα ανά γεωγραφικό διαμέρισμα, του έτους 2009 ,παρουσιάζονται στις επόμενες σελίδες.



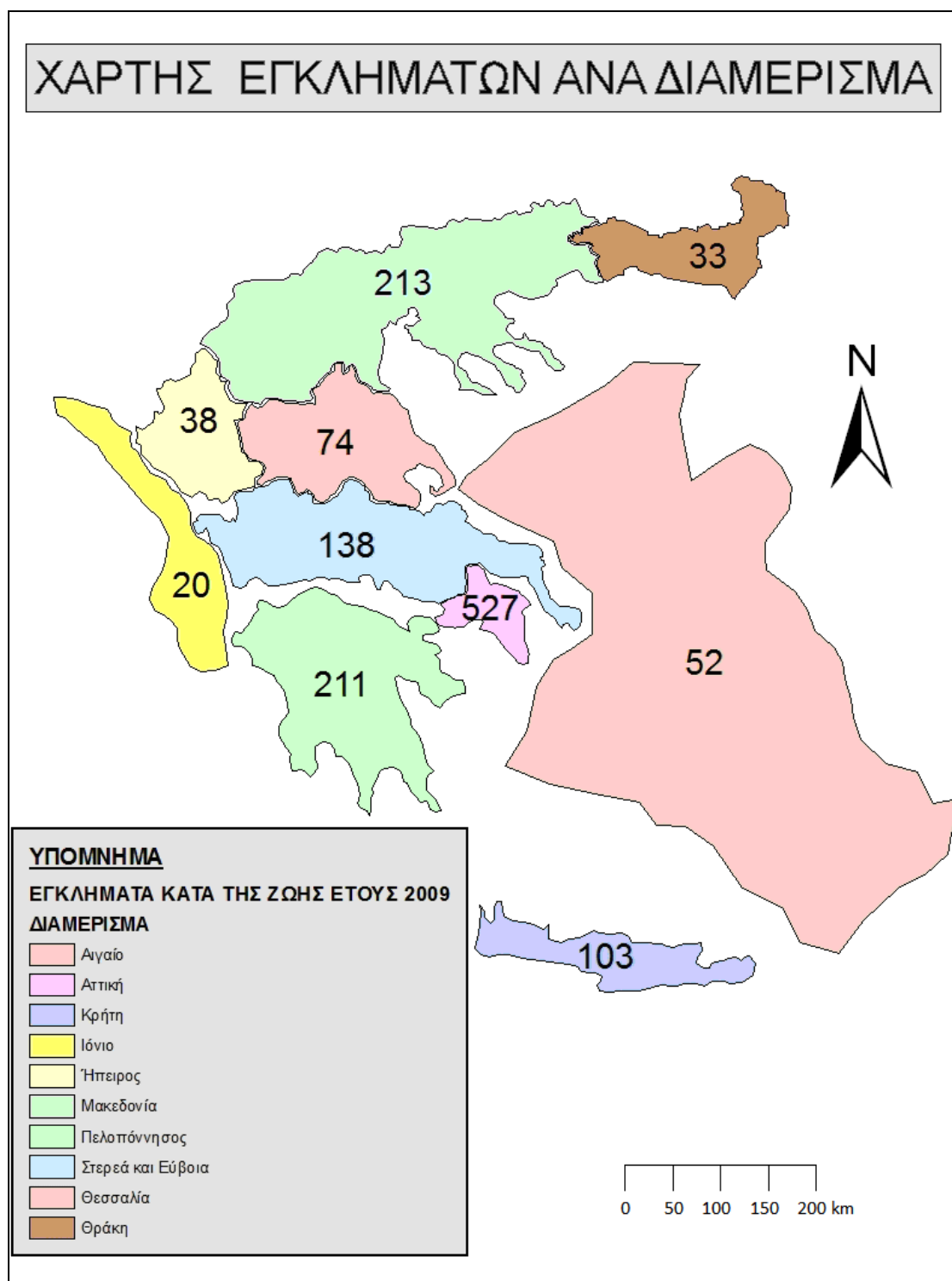
ΧΑΡΤΗΣ 1: Γεωγραφικά διαμερίσματα Ελλαδικού χώρου



ΧΑΡΤΗΣ 2: Γεωγραφικά διαμερίσματα



ΧΑΡΤΗΣ 3: Χάρτης αδικημάτων



ΧΑΡΤΗΣ 4: Χάρτης εγκλημάτων ανά διαμέρισμα

3.3. Ταξινόμηση Γεωγραφικών διαμερισμάτων

Στα πλαίσια του συγκεκριμένου κεφαλαίου πραγματοποιήθηκε η ταξινόμηση των γεωγραφικών διαμερισμάτων της Ελλάδας βάση του αριθμού των διαπραχθέντων αδικημάτων που πραγματοποιήθηκαν το έτος 2009. Η διαδικασία αυτή, πραγματοποιήθηκε με χρήση του στατιστικού πακέτου Minitab, το οποίο έχει αναπτυχθεί ειδικά για στατιστικές αναλύσεις σε εφαρμογές ελέγχου ποιότητας και προσφέρει γι' αυτόν το σκοπό μια αμέτρητη γκάμα στατιστικών μεθόδων ανάλυσης. Ουσιαστικά πρόκειται για ένα εξαιρετικά αξιόπιστο στατιστικό πακέτο.

Η νέα έκδοσή του, η 16 η οποία χρησιμοποιείται στην παρούσα διπλωματική, είναι ποιοτική και βελτιωμένη σε σχέση με τις προηγούμενες.

Στο σημείο αυτό πρέπει να αναφερθεί ότι το λογισμικό Minitab έχει ενσωματώσει αρκετές εκπαιδευτικές λειτουργίες, στην έκδοση αυτή, οι οποίες ενεργοποιούνται σε πραγματικό χρόνο με το Βοηθό(Assistant) του προγράμματος. Επιπλέον η αποστολή των αποτελεσμάτων απευθείας στο Word ή στο PowerPoint με δύο κλικ του ποντικιού επιτρέπει εύκολη ενσωμάτωση σε φοιτητικές εργασίες.

Αξιοσημείωτο επίσης είναι το γεγονός ότι το πρόγραμμα λειτουργεί αξιόπιστα και γρήγορα, ακόμη και σε ένα βασικό PC με λειτουργικό σύστημα 32-bit και 1GB μνήμη RAM, χωρίς να δεσμεύει αδηφάγα τους πόρους του συστήματος.

3.3.1. Ο αλγόριθμος K-Means

Η ταξινόμηση πραγματοποιήθηκε μέσω του στατιστικού πακέτου Minitab και συγκεκριμένα με χρήση του αλγόριθμου ομαδοποίησης τετραγωνικού λάθους, k-Means. Εντός του περιβάλλοντος του στατιστικού πακέτου εισήχθησαν τα στοιχεία του πίνακα 3.1 της σελίδας 32 που αποτελούν τις μεταβλητές του προβλήματος. Στη συνέχεια πληκτρολογήθηκε η κατάλληλη εντολή στο πρόγραμμα προκειμένου να ταξινομηθούν τα στοιχεία σε 3 τάξεις.

Τα αποτελέσματα της διαδικασίας παρουσιάζονται παρακάτω:

K-means Cluster Analysis:

Final Partition

Number of clusters: 3

	Number of observations	Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid
Cluster1	1	0,000	0,000	0,000
Cluster2	8	4,25565E+07	1933,144	5130,006
Cluster3	1	0,000	0,000	0,000

Cluster Centroids

Variable	Cluster1	Cluster2	Cluster3
kata twn periousiakwn dikaiomat	1075,0000	44,2500	307,0000
kata tis idioktisias	60824,0000	2624,3750	20916,0000
peri ton gamo kai tin oikogenei	145,0000	43,1250	217,0000
kata tis prosopikis eleutherias	3030,0000	488,1250	1065,0000
somatikes blabes	11029,0000	734,6250	3188,0000
kata tis zois	527,0000	83,6250	213,0000
peri tin aponomi tis dikaiosini	1237,0000	129,2500	505,0000
peri to nomisma	731,0000	171,5000	1247,0000

Variable	Grand centroid
kata twn periousiakwn dikaiomat	173,6000
kata tis idioktisias	10273,5000
peri ton gamo kai tin oikogenei	70,7000
kata tis prosopikis eleutherias	800,0000
somatikes blabes	2009,4000
kata tis zois	140,9000
peri tin aponomi tis dikaiosini	277,6000
peri to nomisma	335,0000

Distances Between Cluster Centroids

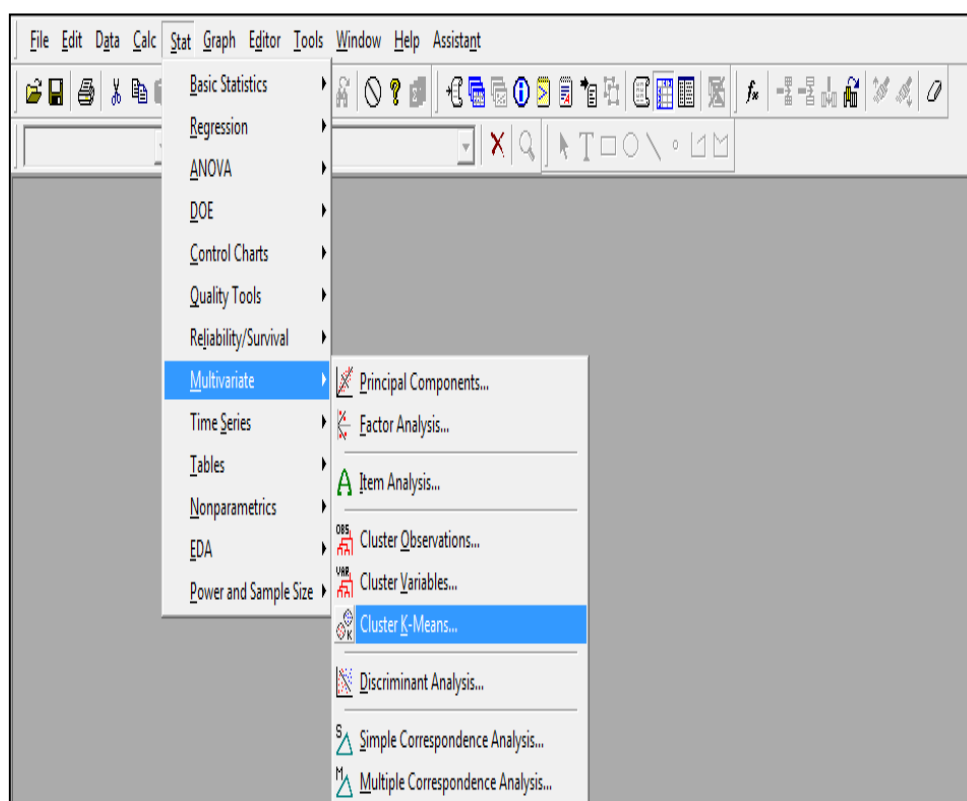
	Cluster1	Cluster2	Cluster3
Cluster1	0,0000	59181,4261	40736,7973
Cluster2	59181,4261	0,0000	18502,6829
Cluster3	40736,7973	18502,6829	0,0000

Πιο αναλυτικά, αφού έγινε η προετοιμασία και κωδικοποίηση των μεταβλητών στη σωστή γλώσσα, στη συνέχεια εισηχθη το λογιστικό φύλλο excel στο ειδικό πακέτο στατιστικής ανάλυσης όπως φαίνεται παρακάτω(Εικόνα 3.9).

+	C1-T	C2	C3	C4	C5	C6	C7	C8	C9	C10
		kata twn periousiakwn dikaiomat	kata tis idioiktisias	peri ton gamo kai tin oikogenei	kata tis prospikis eleutherias	somatikes blabes	kata tis zois	peri tin aponomi tis dikaiosini	peri to nomisma	
1	Attiki	1075	60824	145	3030	11029	527	1237	731	
2	Sterea kai Euboia	67	4162	19	491	1071	138	141	148	
3	Peloponnisos	93	7526	46	1227	2023	211	251	399	
4	Ionio	23	978	15	475	330	20	66	62	
5	Ipeiros	15	740	2	258	213	38	51	25	
6	Thessalia	41	1647	164	393	730	74	238	207	
7	Makedonia	307	20916	217	1065	3188	213	505	1247	
8	Thraki	22	820	6	131	409	33	53	135	
9	Aigeo	31	1482	18	314	524	52	81	188	
10	Criti	62	3640	75	616	577	103	153	208	
11										
12										

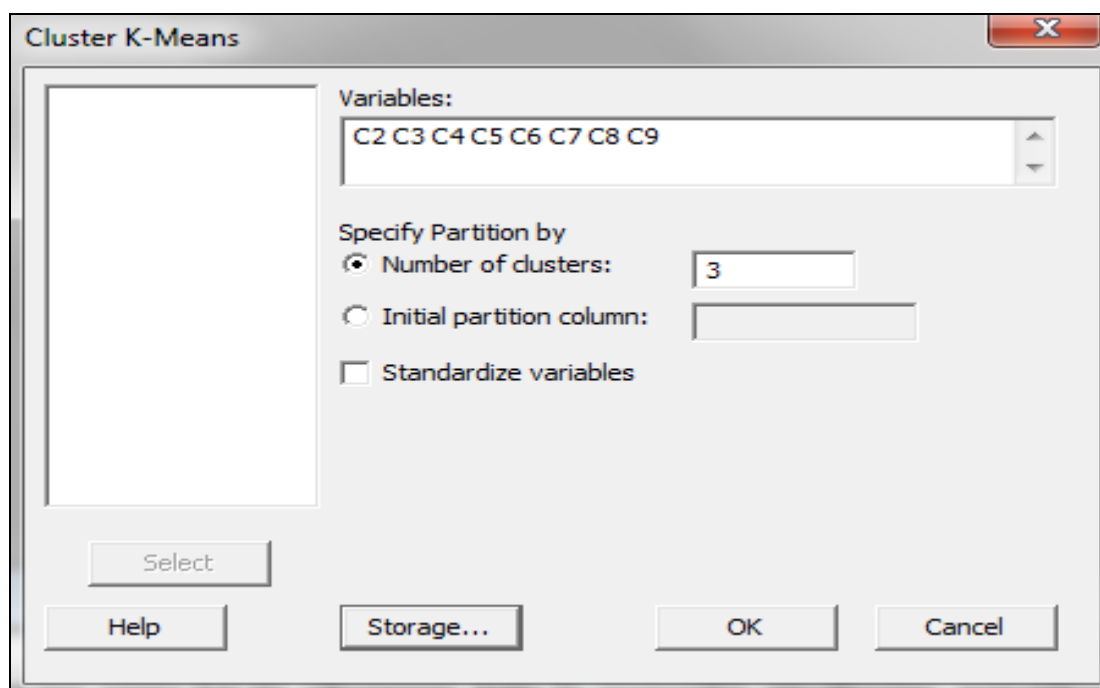
ΕΙΚΟΝΑ 3.9: Εισαγωγή στοιχείων στο Minitab

Στη συνέχεια και προκειμένου να εκτελεστεί ο αλγόριθμος K-Means επιλέχθηκαν διαδοχικά `stat>multivariate>cluster K-Means`. (Εικόνα3.10)



ΕΙΚΟΝΑ 3.10: Διαδικασία εκτέλεσης αλγορίθμου K-Means

Επόμενο βήμα αποτέλεσε η δήλωση των μεταβλητών προς ταξινόμηση καθώς και του αριθμού των επιθυμητών τάξεων. Η δήλωση των μεταβλητών πραγματοποιήθηκε με την εισαγωγή των αντίστοιχών στηλών του πίνακα C1,C2 κτλ. Όσον αφορά στον αριθμό των κλάσεων, αυτός επιλέχθηκε λόγω λιγοστού όγκου δεδομένων και ενδεικτικού χαρακτήρα της επίλυσης του προβλήματος, να είναι ίσος με (3) τρεις. (Εικόνα 3.11)



ΕΙΚΟΝΑ 3.11: Δήλωση μεταβλητών και αριθμού τάξεων

Τελικά, τα αποτελέσματα της ταξινόμησης καταχωρούνται στη νέα στήλη C10 με όνομα class. (Εικόνα 3.12)

C1-T	C2	C3	C4	C5	C6	C7	C8	C9	C10
	kata tvn periouasiakwn dikaiomat	kata tis idioktisiās	peri ton gamo kai tin oikogenei	kata tis prosopikis eleutherias	somatikes blabes	kata tis zois	peri tin aponomi tis dikaiosini	peri to nomisma	class
Attiki	1075	60824	145	3030	11029	527	1237	731	1
Stereā kai Euboia	67	4162	19	491	1071	138	141	148	2
Peloponnisos	93	7526	46	1227	2023	211	251	399	2
Ionio	23	978	15	475	330	20	66	62	2
Ipeiros	15	740	2	258	213	38	51	25	2
Thessalia	41	1647	164	393	730	74	238	207	2
Makedonia	307	20916	217	1065	3188	213	505	1247	3
Thraki	22	820	6	131	409	33	53	135	2
Aigeo	31	1482	18	314	524	52	81	188	2
Criti	62	3640	75	616	577	103	153	208	2

EIKONA 3.12: Αποτελέσματα ταξινόμησης σε 3 τάξεις

3.4. Συμπεράσματα

Όπως ήταν αναμενόμενο, τα γεωγραφικά διαμερίσματα της Αττικής και της Μακεδονίας, λόγω αυξημένου αριθμού αδικημάτων αποτέλεσαν τις τάξεις 1 και 3 αντίστοιχα, ενώ οι υπόλοιπες περιφέρειες κατατάχτηκαν στην τάξη που αντιστοιχεί στον αριθμό 2.

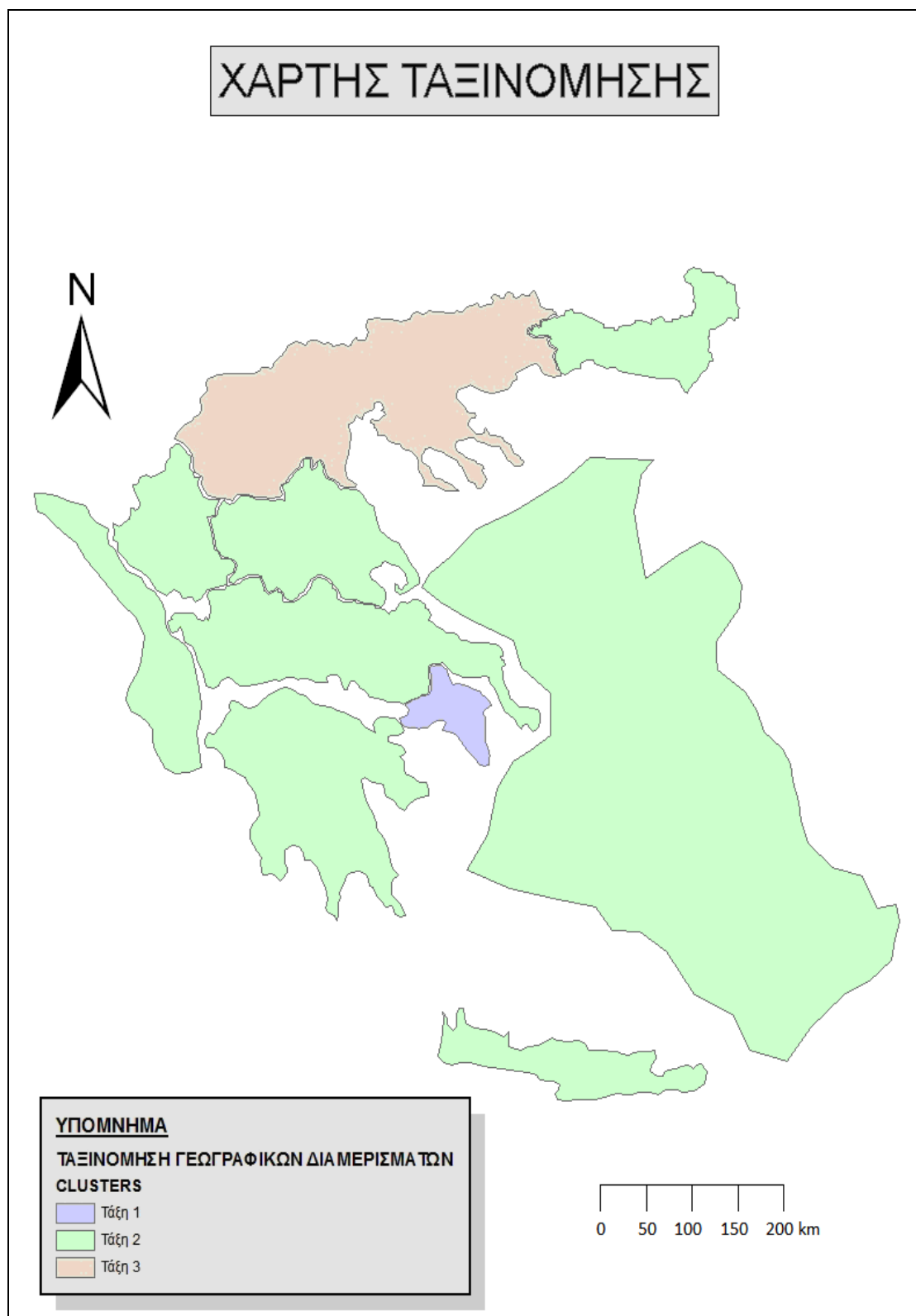
Στο σημείο αυτό, πρέπει να σημειωθεί πως οι αριθμοί που χαρακτηρίζουν οι τάξεις συμβολίζουν το βαθμό εγκληματικότητας ανά γεωγραφική περιφέρεια ως εξής:

1 \implies Έντονη εγκληματικότητα

2 \implies Ήπια εγκληματικότητα

3 \implies Μέση εγκληματικότητα

Για την οπτικοποίηση των αποτελεσμάτων της ταξινόμησης, παρατίθεται ο παρακάτω χάρτης, ο οποίος αποτελεί το τελικό προϊόν της όλης διαδικασίας και δημιουργήθηκε σε περιβάλλον ArcGis. Για τη παρουσίαση αυτού, πραγματοποιήθηκε σύνδεση των αποτελεσμάτων της ταξινόμησης και του θεματικού επιπέδου των Γεωγραφικών Διαμερισμάτων, με χρήση της εντολής Join, η οποία παρουσιάζεται αναλυτικά στο τέταρτο κεφάλαιο.



ΧΑΡΤΗΣ 5: Χάρτης ταξινόμησης γεωγραφικών διαμερισμάτων

ΚΕΦΑΛΑΙΟ 4

ΕΦΑΡΜΟΓΗ ΤΑΞΙΝΟΜΗΣΗΣ ΔΗΜΩΝ ΑΤΤΙΚΗΣ

Στο παρόν κεφάλαιο παρουσιάζονται δεδομένα που αφορούν τον πληθυσμό, την οικογενειακή κατάσταση, το μέσο εισόδημα και άλλα στοιχεία που σχετίζονται με τη χωροθέτηση ενός καταστήματος πώλησης παιδικών ειδών. Η χωρική μονάδα αναφοράς των δεδομένων είναι ο εκάστοτε δήμος της νότιας Αττικής. Εν συνεχεία τα δεδομένα επεξεργάζονται σε περιβάλλον ΓΣΠ και τελικά πραγματοποιείται η δημιουργία τριών τάξεων με χρήση στατιστικής ταξινόμησης, και συγκεκριμένα με τον αλγόριθμο K-Means. Επιπρόσθετα, με στόχο ναδειχθεί η πρακτική σημασία της συγκεκριμένης εφαρμογής, με βάση τα δεδομένα που χρησιμοποιήθηκαν, πραγματοποιήθηκε η βέλτιστη επιλογή ζώνης χωροθέτησης ενός καταστήματος λιανικής πώλησης παιδικών προϊόντων

4.1. Προσδιορισμός του προβλήματος

Στο πλαίσιο αυτό, αρχικά πραγματοποιείται η δημιουργία και διαχείριση τεχνητών μεταβλητών, που αφορούν σε δημογραφικά δεδομένα δώδεκα γειτονικών δήμων της Αττικής. Στη συνέχεια πραγματοποιείται η επεξεργασία των δεδομένων σε περιβάλλον ΓΣΠ και τελικά παρουσιάζεται η διαδικασία της ταξινόμησης και αναλύονται τα αποτελέσματα της, τα οποία οπτικοποιούνται με τη βοήθεια των γεωγραφικών συστημάτων πληροφοριών

4.2. Εισαγωγή και διαχείριση δεδομένων

Πρωταρχική πηγή χωρικών δεδομένων αποτελούν τα στοιχεία του Πίνακα 4.1(σελ52). Στη συνέχεια, παρουσιάζεται αναλυτικά η διαδικασία διαχείρισης των δεδομένων, κατά την οποία τα δεδομένα κανονικοποιήθηκαν, εισήχθησαν και επεξεργάστηκαν σε περιβάλλον εργασίας ΓΣΠ.

A	B	C	D	E	F	G	H	I	
DIMOS	PLITHISMOS	PANTREMENOI	APASXOLOYMENOI	PAIDIA EWS 10 ETWN	ANW TWN 65 ME EGGONIA	KATOXOI I.X	KONTA SE METRO-TRAM	MESO EISODIMA	
AG.DIMITRIOS	68719	25113	28632	12556		5047	13743	20615	1250
ALIMOS	39800	15124	16486	6805		2903	11940	1990	1700
ARGYROYPOLI	35076	12627	14794	6944		2333	7015	0	1345
GLYFADA	83665	31792	34785	15896		6898	29282	4183	2143
DAFNH	25058	9020	10188	4780		2083	5011	4011	1180
ELLHNIKO	16223	6164	6615	3205		1272	4055	162	1690
HLIOYPOLI	81024	30465	33579	12186		6732	20256	2430	1585
KALLITHEA	115150	40302	47357	14105		7254	11515	2303	1195
MOSXATO	24315	8874	9838	4614		1698	3647	729	1237
NEA SMYRNI	76508	28078	32278	8984		4966	19127	6120	1780
PALAIΟ FALHRO	67160	24916	26891	13703		5780	16790	2350	1986
BOYLA	25647	10130	10622	3039		1871	8976	456	2687

ΠΙΝΑΚΑΣ 4.1: Στατιστικά στοιχεία δήμων νότιας Αττικής

4.2.1. Κανονικοποίηση δεδομένων

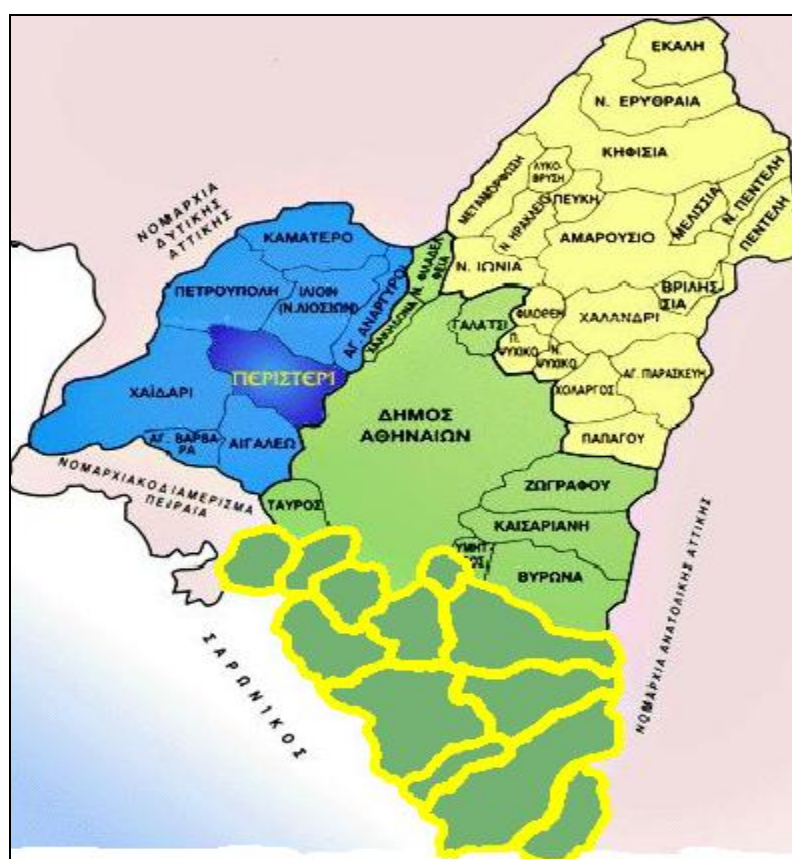
Τα δεδομένα αρχικά πρέπει να προετοιμαστούν για την ταξινόμηση. Οι αλγόριθμοι της ταξινόμησης είναι πολύ ευαίσθητοι σε δεδομένα με διαφορετική κλίμακα ή τύπο και γι αυτό το λόγο πριν από την ταξινόμηση ενδείκνυται η κανονικοποίηση των δεδομένων ώστε να υπάρχει ενιαία κλίμακα αναφοράς. Συνεπώς παρουσιάστηκε η ανάγκη αναγωγής των δεδομένων αυτών σε συγκρίσιμες μεταξύ τους μονάδες. Η αναγωγή αυτή έγινε με τη διαίρεση των δεδομένων με τον αντίστοιχο πληθυσμό του κάθε δήμου, για τα δεδομένα που αναφέρονταν στον πληθυσμό. Τα δεδομένα των μεταβλητών που αναφέρονταν στον πληθυσμό και το εισόδημα διαιρέθηκαν με το μέγιστο πληθυσμό και εισόδημα αντίστοιχα. Τα τελικά αποτελέσματα είναι ποσοστά επί τοις εκατό και αποτέλεσαν τη βάση δεδομένων για την εφαρμογή μας (Πίνακας 4.2).

C1-T	C2	C3	C4	C5	C6	C7	C8	C9	
DIMOS	PLITHISMOS	PANTREMENOI	APASXOLOYMENOI	PAIDIA EWS 10 ETWN	ANW TWN 65 ME EGGONIA	KATOXOI I.X	KONTA SE METRO-TRAM	MESO EISODIMA	
AG.DIMITRIOS	59,67	35,54	25,67	18,27		7,46	20,00	30,00	46,52
ALIMOS	34,56	44,88	41,42	23,66		13,84	30,87	5,90	63,26
ARGYROYPOLI	30,46	23,45	25,98	15,65		5,68	33,88	0,00	50,05
GLYFADA	72,60	29,57	21,65	27,78		18,77	52,78	7,34	79,07
DAFNH	21,76	35,97	31,98	12,92		16,90	25,80	16,01	43,91
ELLHNIKO	14,08	25,77	37,43	31,98		11,87	45,68	2,77	62,89
HLIOYPOLI	70,36	31,93	45,61	14,34		21,96	40,79	3,00	58,98
KALLITHEA	100,00	40,65	24,90	12,72		6,49	12,84	5,09	44,47
MOSXATO	21,11	38,48	19,64	16,06		15,80	15,00	10,63	46,03
NEA SMYRNI	66,40	22,67	36,78	11,56		9,68	20,66	18,94	66,24
PALAIΟ FALHRO	58,32	30,56	20,12	31,92		18,55	18,50	8,69	73,91
BOYLA	22,27	24,10	28,95	23,79		7,30	38,37	3,73	100,00

ΠΙΝΑΚΑΣ 4.2: Κανονικοποιημένα στοιχεία

4.2.2. Σχεδιασμός γεωγραφικής βάσης δεδομένων

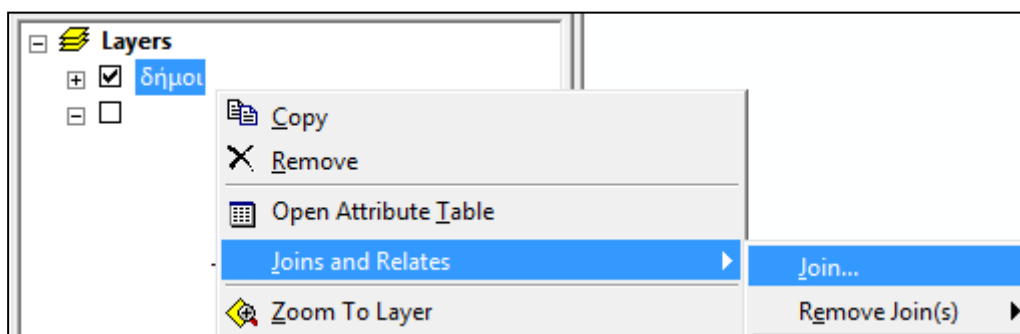
Στη συνέχεια, με τον ίδιο τρόπο όπως στην πρώτη εφαρμογή, δημιουργήθηκε στο ArcCatalog ένα νέο αρχείο Shapefile και προστέθηκε σε περιβάλλον ArcMap μαζί με τον χάρτη που αποτέλεσε το υπόβαθρο για τη δημιουργία πολυγώνων-δήμων. Το νέο αρχείο που ψηφιοποιήθηκε έλαβε τη χαρακτηριστική ονομασία “δήμοι” (Εικόνα 4.1).



ΕΙΚΟΝΑ 4.1: Αποτέλεσμα ψηφιοποίησης

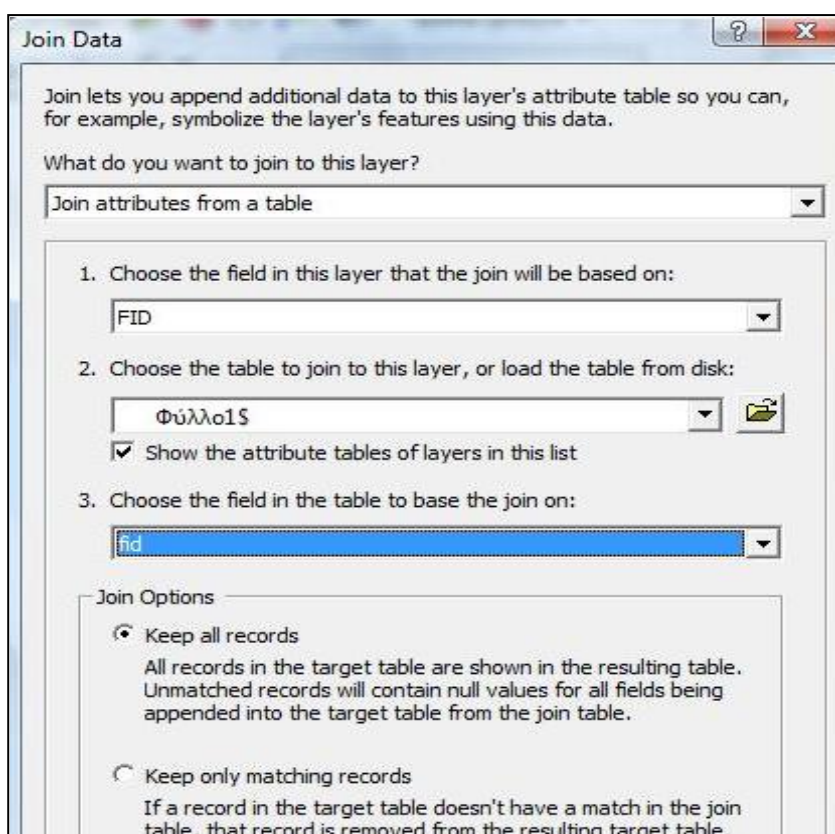
4.2.3. Σύνδεση πληροφορίας και αντίστοιχης χωρικής οντότητας

Επόμενο βήμα αποτέλεσε η σύνδεση ανάμεσα στο ψηφιοποιημένο αρχείο των 12 δήμων της Αττικής και στο αρχείο excel, το οποίο περιέχει τα στατιστικά στοιχεία για τους δήμους. Η καταχώρηση πληροφορίας στα πολύγωνα πραγματοποιήθηκε σε περιβάλλον ArcMap και αφού επιλέχθηκε το κατάλληλο θεματικό επίπεδο πραγματοποιήθηκε η εντολή join (Εικόνα 4.2).



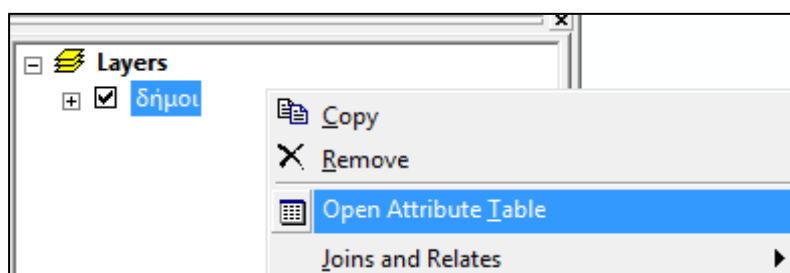
ΕΙΚΟΝΑ 4.2: Επιλογή της εντολής Join

Στη φόρμα διαλόγου που εμφανίζεται στην επόμενη σελίδα επιλέχθηκε στο πρώτο πλαίσιο, να γίνει σύνδεση με τα στοιχεία ενός πίνακα (join attributes from a table), στο δεύτερο πλαίσιο το πεδίο μέσα από το θεματικό επίπεδο στο οποίο βασίστηκε η σύνδεση με τον πίνακα (FID), στο τρίτο πλαίσιο το φύλλο excel που περιέχει τον πίνακα και τέλος, το πεδίο μέσα από τον πίνακα με το οποίο έγινε η σύνδεση (fid) (Εικόνα 4.3).



ΕΙΚΟΝΑ 4.3: Διαδικασία εκτέλεσης της εντολής Join

Τέλος, το αποτέλεσμα της σύνδεσης, φαίνεται εφόσον επιλεγθεί στο θεματικό επίπεδο των δήμων, το άνοιγμα του πίνακα χαρακτηριστικών “open attribute table” (Εικόνα 4.5).



ΕΙΚΟΝΑ 4.5: Μετάβαση στον πίνακα χαρακτηριστικών ενός θεματικού επιπέδου

Παρατηρείται πως το αποτέλεσμα αυτής της σύνδεσης είναι ότι ο πίνακας περιγραφικών χαρακτηριστικών εμφανίζει τα στοιχεία του λογιστικού φύλλου excel, παρέχοντας στον χρήστη τη δυνατότητα οπτικοποίησης των δεδομένων (Εικόνα 4.6).

	FID	Shape	Id	fid	DIMOS	PLITHISMOS	PANTREMENOI	APASXOLOYMENOI	PAI...
▶	0	Polygon	0	0	AG.DIMITRIOS	59,67	35,544	25,665	
	1	Polygon	0	1	ALIMOS	34,56	44,876	41,422	
	2	Polygon	0	2	ARGYROYPOLI	30,46	23,445	25,984	
	3	Polygon	0	3	GLYFADA	72,6	29,567	21,653	
	4	Polygon	0	4	DAFNH	21,76	35,966	31,978	
	5	Polygon	0	5	ELLHNIKO	14,08	25,768	37,431	
	6	Polygon	0	6	HLIOYPOLI	70,36	31,934	45,612	
	7	Polygon	0	7	KALLITHEA	100	40,652	24,897	
	8	Polygon	0	8	MOSXATO	21,11	38,476	19,643	
	9	Polygon	0	9	NEA SMYRNI	66,4	22,672	36,783	
	10	Polygon	0	10	PALAIΟ FALHRO	58,32	30,564	20,123	
	11	Polygon	0	11	BOYLA	22,27	24,098	28,945	

ΕΙΚΟΝΑ 4.6: Αποτελέσματα σύνδεσης

4.2.4. Απεικόνιση δεδομένων

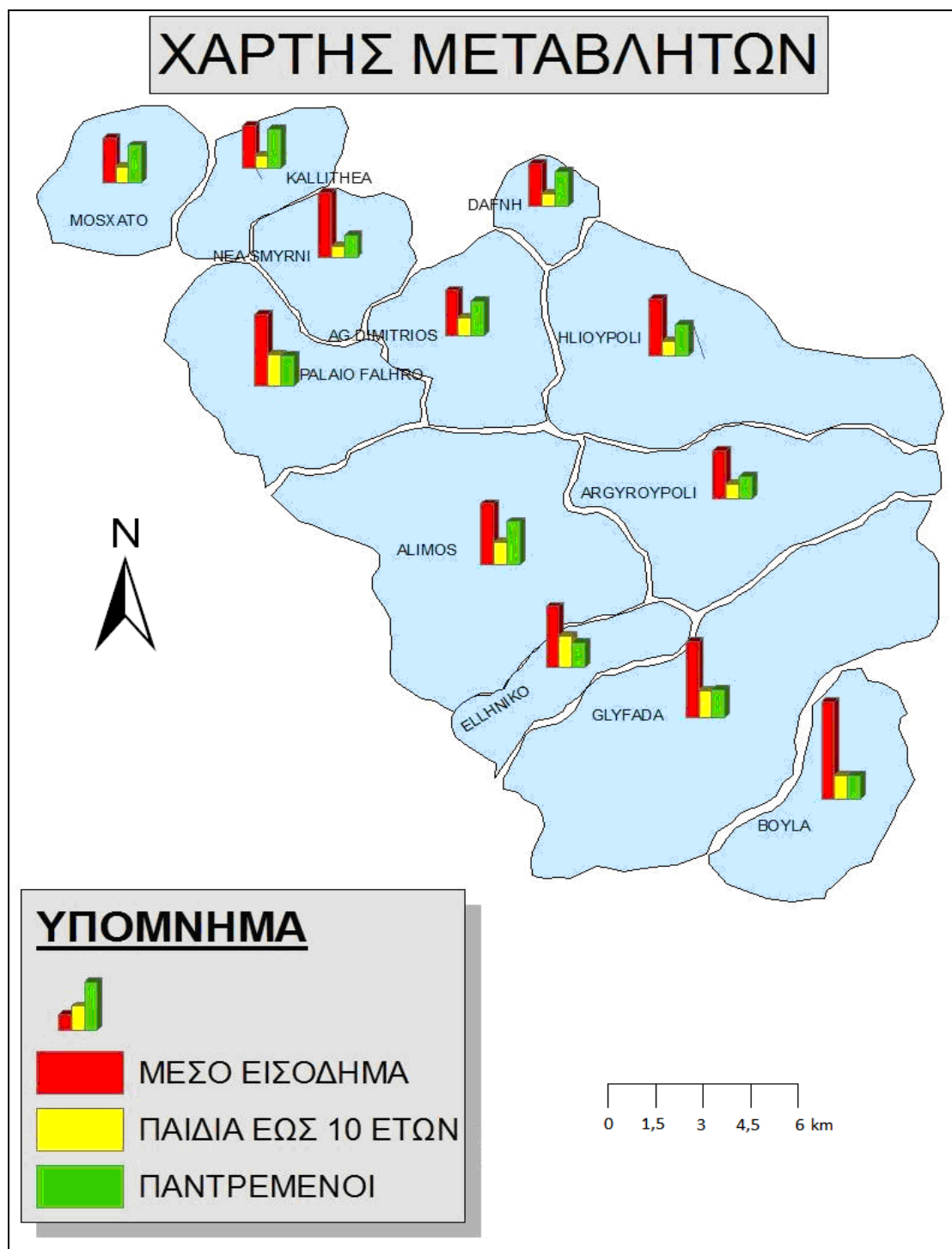
Στο συγκεκριμένο στάδιο, η βάση δεδομένων είναι ικανή να απεικονίσει την πληροφορία των πινάκων excel γραφικά με τη μορφή χάρτη. Η προεργασία για να παραχθούν οι χάρτες πραγματοποιήθηκε στο περιβάλλον εργασίας “ArcMap” με τη βοήθεια των ιδιοτήτων του αντίστοιχου θεματικού επιπέδου (Layer Properties) και συγκεκριμένα με την ιδιότητα του συμβολισμού (Symprology) καθώς και αυτήν της ονομασίας (Labels).

Στις επόμενες σελίδες παρουσιάζονται τα δεδομένα οπτικοποιημένα με χρήση των δυνατοτήτων που παρέχουν τα γεωγραφικά συστήματα πληροφοριών.

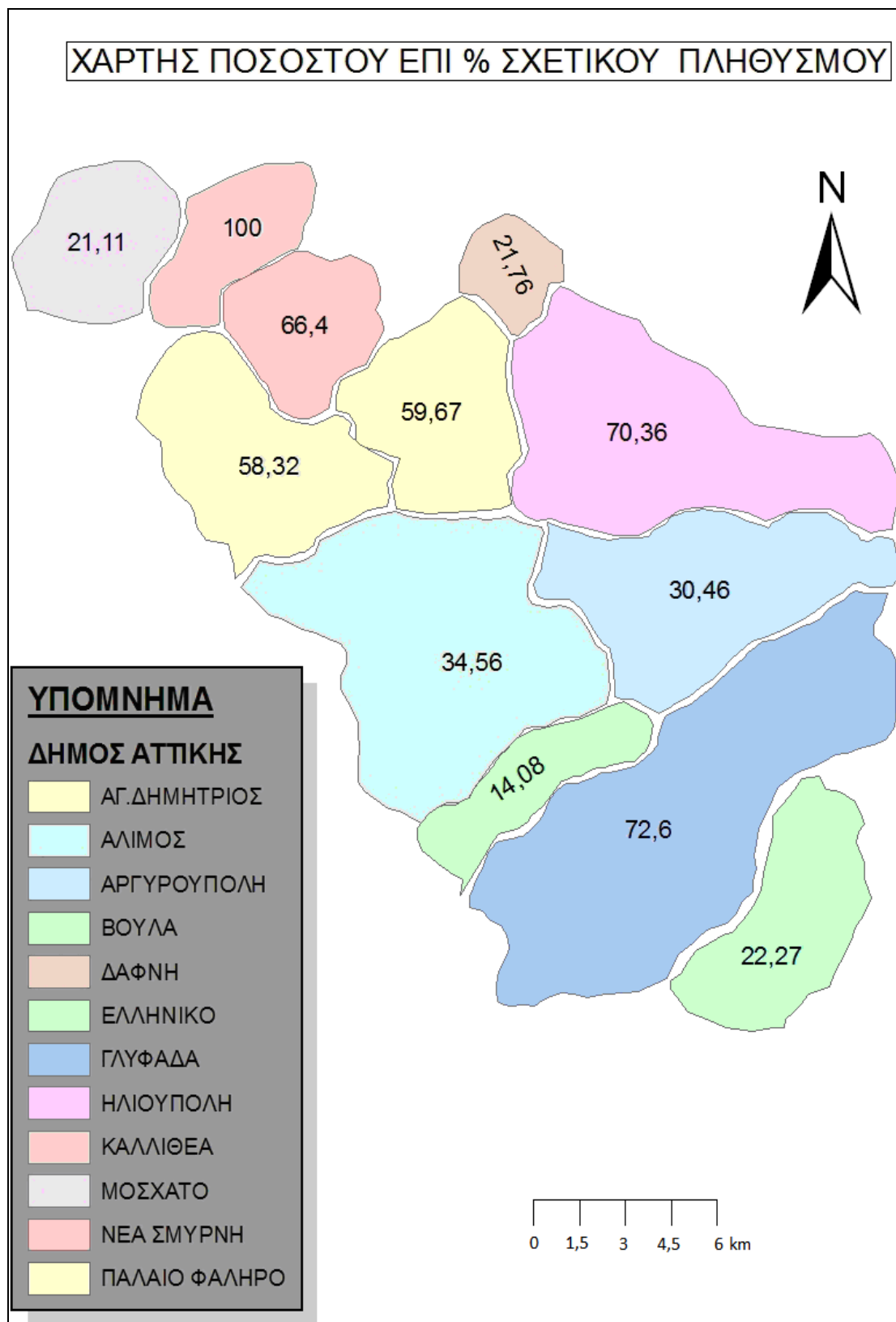
Προκειμένου οι χρήστες να κατανοήσουν εύκολα την έννοια του χάρτη, τα είδη των χαρτών, τις διαφορετικές πληροφορίες που μπορούν να αποτυπωθούν πάνω σε αυτούς, αλλά και τη βοήθεια που δύνανται να προσφέρουν στη σωστή λήψη αποφάσεων, με τη βοήθεια των ΓΣΠ τους δίνεται η δυνατότητα να πειραματιστούν πάνω στη δημιουργία τους.

Σε περίπτωση που ο εκάστοτε χρήστης θέλει να παρουσιάσει παραπάνω από ένα χαρακτηριστικά, τα ΓΣΠ παρέχουν αυτή τη δυνατότητα μέσω των ιδιοτήτων του συμβολισμού .

Στον επόμενο χάρτη οι πληροφορίες που παρουσιάζονται είναι η ονομασία του κάθε πολυγώνου, οι τιμές του μέσου εισοδήματος ανά δήμο, το ποσοστό των παντρεμένων, καθώς και το ποσοστό των οικογενειών με παιδιά μέχρι δέκα ετών. Η επιλογή εξαρτάται από το τι θέλει να τονίσει κάθε φορά ο χαρτογράφος και το σκοπό δημιουργίας του συγκεκριμένου χάρτη.



ΧΑΡΤΗΣ 6: Χάρτης μεταβλητών



ΧΑΡΤΗΣ 7: Χάρτης ποσοστού επί % σχετικού πληθυσμού

4.3. Ταξινόμηση

Στην παρούσα φάση, πραγματοποιήθηκε η ταξινόμηση δώδεκα δήμων της Αθήνας σε τρεις τάξεις με βάση τις μεταβλητές του πίνακα 4.3.

A	B	C	D	E	F	G	H	I	
DIMOS	PLITHISMOS	PANTREMENOI	APASKOLOYMENOI	PAIDIA EWS 10 ETWN	ANW TWN 65 ME EGGONIA	KATOXOI I.X	KONTA SE METRO-TRAM	MESO EISODIMA	
AG.DIMITRIOS	68719	25113	28632	12556		5047	13743	20615	1250
ALIMOS	39800	15124	16486	6805		2903	11940	1990	1700
ARGYROYPOLI	35076	12627	14794	6944		2333	7015	0	1345
GLYFADA	83665	31792	34785	15896		6898	29282	4183	2143
DAFNH	25058	9020	10188	4780		2083	5011	4011	1180
ELLHNIKO	16223	6164	6615	3205		1272	4055	162	1690
HLIOYPOLI	81024	30465	33579	12186		6732	20256	2430	1585
KALLITHEA	115150	40302	47357	14105		7254	11515	2303	1195
MOSXATO	24315	8874	9838	4614		1698	3647	729	1237
NEA SMYRNI	76508	28078	32278	8984		4966	19127	6120	1780
PALAIΟ FALHRO	67160	24916	26891	13703		5780	16790	2350	1986
BOYLA	25647	10130	10622	3039		1871	8976	456	2687

ΠΙΝΑΚΑΣ 4.3: Στατιστικά στοιχεία δήμων νότιας Αττικής

Στη συνέχεια εισήχθησαν τα δεδομένα στις στήλες (C1, C2, C3...C9) του φύλλου εργασίας του στατιστικού πακέτου Minitab και τελικά ταξινομήθηκαν σε τρεις τάξεις, μέσω του αλγόριθμου K-Means, με χρήση διαδοχικά των εντολών Stat>Multivariate>Cluster K-Means.

Τα αποτελέσματα της ταξινόμησης καταχωρήθηκαν στη τελευταία στήλη του πίνακα με όνομα class(Πίνακας 4.4).

C1-T	C2	C3	C4	C5	C6	C7	C8	C9	C10	
DIMOS	PLITHISMOS	PANTREMENOI	APASKOLOYMENOI	PAIDIA EWS 10 ETWN	ANW TWN 65 ME EGGONIA	KATOXOI I.X	KONTA SE METRO-TRAM	MESO EISODIMA	CLASS	
AG.DIMITRIOS	59,67	35,54	25,67	18,27		7,46	20,00	30,00	46,52	1
ALIMOS	34,56	44,88	41,42	23,66		13,84	30,87	5,90	63,26	2
ARGYROYPOLI	30,46	23,45	25,98	15,65		5,68	33,88	0,00	50,05	3
GLYFADA	72,60	29,57	21,65	27,78		18,77	52,78	7,34	79,07	2
DAFNH	21,76	35,97	31,98	12,92		16,90	25,80	16,01	43,91	3
ELLHNIKO	14,08	25,77	37,43	31,98		11,87	45,68	2,77	62,89	3
HLIOYPOLI	70,36	31,93	45,61	14,34		21,96	40,79	3,00	58,98	2
KALLITHEA	100,00	40,65	24,90	12,72		6,49	12,84	5,09	44,47	1
MOSXATO	21,11	38,48	19,64	16,06		15,80	15,00	10,63	46,03	3
NEA SMYRNI	66,40	22,67	36,78	11,56		9,68	20,66	18,94	66,24	1
PALAIΟ FALHRO	58,32	30,56	20,12	31,92		18,55	18,50	8,69	73,91	1
BOYLA	22,27	24,10	28,95	23,79		7,30	38,37	3,73	100,00	2

ΠΙΝΑΚΑΣ 4.4: Κανονικοποιημένα στοιχεία-Αποτελέσματα ταξινόμησης

Εν συνεχεία, παρουσιάζονται αναλυτικά τα αποτελέσματα της εφαρμογής του αλγόριθμου “K-means” στις μεταβλητές του συγκεκριμένου παραδείγματος.

K-means Cluster Analysis

Final Partition

Number of clusters: 3

	Number of observations	Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid
Cluster1	4	2879,133	26,054	35,712
Cluster2	4	4047,849	31,435	39,406
Cluster3	4	1660,051	19,826	27,190

Cluster Centroids

Variable	Cluster1	Cluster2	Cluster3	Grand centroid
PLITHISMOS	71,0975	49,9475	21,8525	47,6325
PANTREMENOI	32,3550	32,6200	30,9175	31,9642
APASXOLOYMENOI	26,8675	34,4075	28,7575	30,0108
PAIDIA EWS 10 ETWN	18,6175	22,3925	19,1525	20,0542
ANW TWN 65 ME EGGONIA	10,5450	15,4675	12,5625	12,8583
KATOXOI I.X	18,0000	40,7025	30,0900	29,5975
KONTA SE METRO-TRAM	15,6800	4,9925	7,3525	9,3417
MESO EISODIMA	57,7850	75,3275	50,7200	61,2775

Distances Between Cluster Centroids

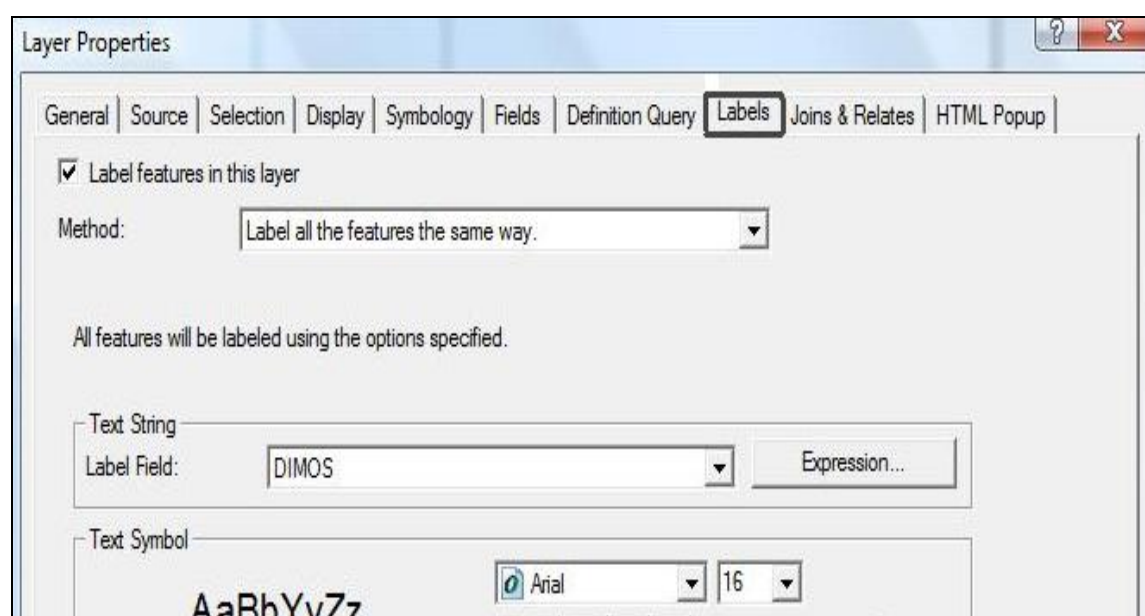
	Cluster1	Cluster2	Cluster3
Cluster1	0,0000	38,4720	51,9663
Cluster2	38,4720	0,0000	39,5830
Cluster3	51,9663	39,5830	0,0000

Όσον αφορά το πρακτικό κομμάτι της ταξινόμησης, αναλύοντας τα παραπάνω αποτελέσματα του αλγορίθμου και ειδικότερα το υπογραμμισμένο κομμάτι που αναφέρεται στα κέντρα των κλάσεων (Clusters Centroids), παρατηρούνται μεγαλύτεροι μέσοι όροι των μεταβλητών στους δήμους που απαρτίζουν την τάξη, η οποία αντιστοιχεί στον αριθμό 2. Το γεγονός αυτό μας οδήγησε στην εκπλήρωση του σκοπού της παρούσας εφαρμογής, δηλαδή της βέλτιστης επιλογής χωροθέτησης ενός καταστήματος λιανικής πώλησης παιδικών προϊόντων.

4.4. Οπτικοποίηση αποτελεσμάτων ταξινόμησης

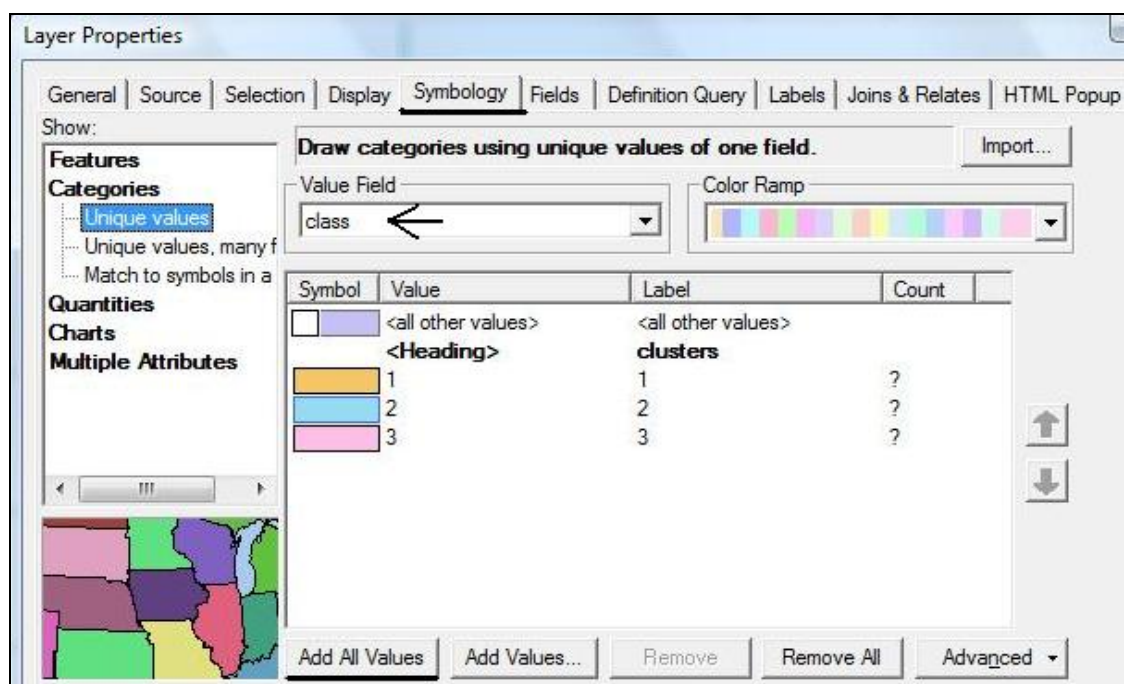
Στο στάδιο αυτό θα αναλυθεί η διαδικασία εξόδου της πληροφορίας δηλαδή της σύνταξης του χάρτη, ο οποίος παρουσιάζει τα αποτελέσματα της ταξινόμησης. Ομοίως με την πρώτη εφαρμογή, η προεργασία πραγματοποιήθηκε στο περιβάλλον εργασίας ArcMap και συγκεκριμένα στο παράθυρο διαλόγου Layer Properties.

Χρησιμοποιήθηκε το πεδίο που περιέχει την ονομασία των δήμων, “dimos” σε συνδυασμό με εκείνο των αποτελεσμάτων της ταξινόμησης(Εικόνα 4.7).



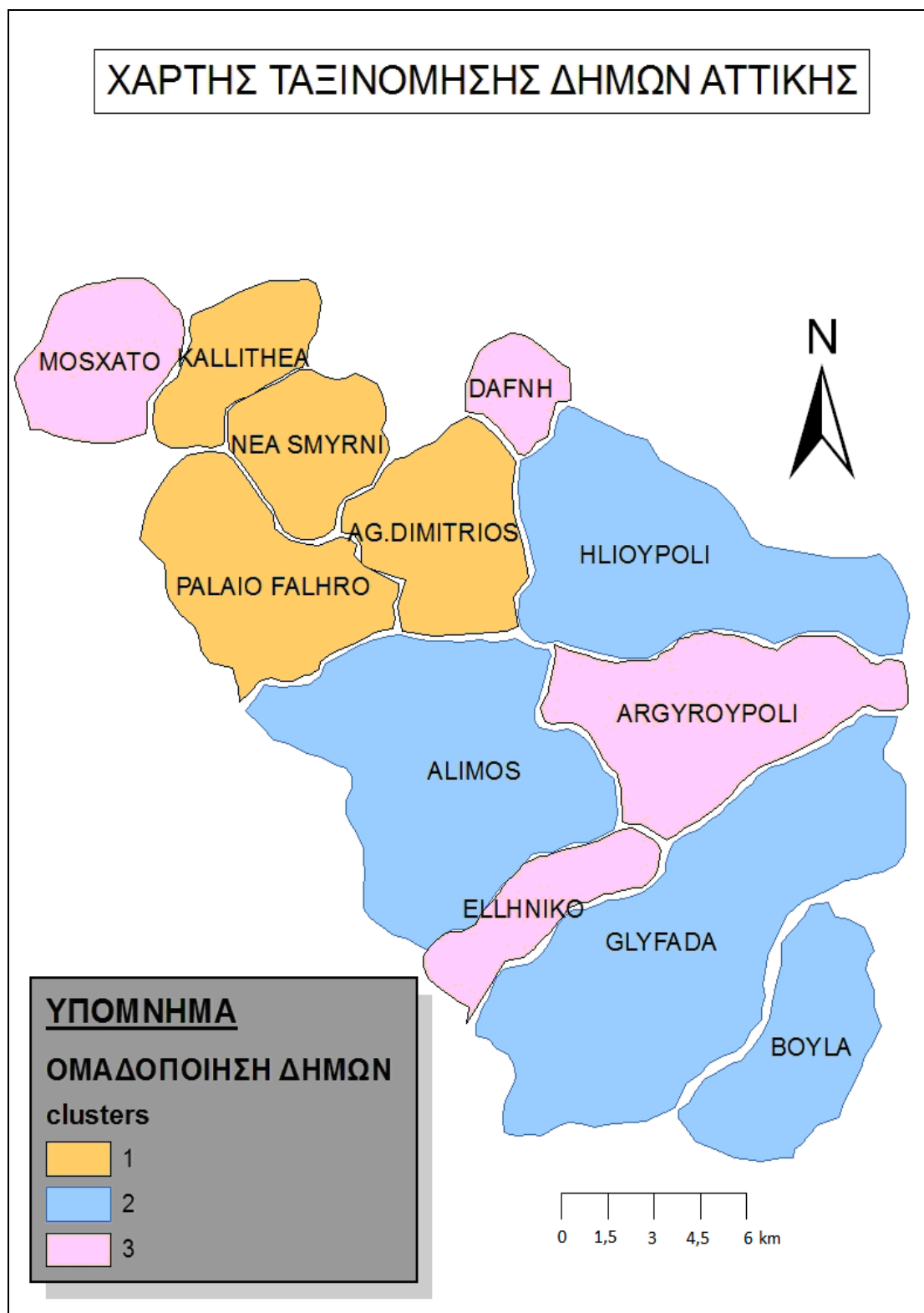
ΕΙΚΟΝΑ 4.7: Διαδικασία επιλογής πεδίου ονοματολογίας στο χάρτη

Στη συνέχεια με τη βοήθεια των ιδιοτήτων του συμβολισμού αποδόθηκε κοινό χρώμα στα πολύγωνα της ίδιας ομάδας(Εικόνα 4.8).



ΕΙΚΟΝΑ 4.8: Επιλογή συμβολισμού στο χάρτη

Τελικά, προκύπτει ο χάρτης ταξινόμησης των δήμων της Αττικής, με βάση τα δεδομένα που έχουν προαναφερθεί.



ΧΑΡΤΗΣ 8: Χάρτης ταξινόμησης δήμων Αττικής

4.5. Συμπεράσματα

Για τη δημιουργία του παραπάνω χάρτη, ο οποίος αποτελεί το τελικό προϊόν της συγκεκριμένης εφαρμογής πραγματοποιήθηκε συνδυαστική χρήση των Γεωγραφικών Συστημάτων Πληροφοριών και εφαρμογής του διαμεριστικού αλγόριθμου ταξινόμησης δεδομένων, K-Means.

Σε πρώτη φάση τα πρωταρχικά δεδομένα του πίνακα 4.1 εισήχθησαν και επεξεργάστηκαν σε περιβάλλον ΓΣΠ. Εν συνεχεία, παρουσιάστηκε αναλυτικά η διαδικασία της ταξινόμησης των δώδεκα δήμων του νομού Αττικής.

Επόμενο βήμα αποτέλεσε η εισαγωγή των αποτελεσμάτων της ταξινόμησης σε περιβάλλον ΓΣΠ με απώτερο σκοπό τη δημιουργία χάρτη ο οποίος παρέχει μια άμεση εμποπτεία στα χωρικά φαινόμενα και διευκολύνει τον χρήστη καθώς η ανθρώπινη αντίληψη είναι πιο εξοικειωμένη στις εικόνες παρά στα πινακοποιημένα στοιχεία.

Όσον αφορά το πρακτικό κομμάτι της ταξινόμησης, αναλύοντας τα αποτελέσματα του αλγορίθμου, παρατηρούνται μεγαλύτεροι μέσοι όροι των μεταβλητών στους δήμους που απαρτίζουν την τάξη, η οποία αντιστοιχεί στον αριθμό 2. Το γεγονός αυτό μας οδήγησε στην εκπλήρωση του σκοπού της παρούσας εφαρμογής, δηλαδή της βέλτιστης επιλογής χωροθέτησης ενός καταστήματος λιανικής πώλησης παιδικών προϊόντων.

Έτσι, ο ενδιαφερόμενος επιχειρηματίας, βάση των αποτελεσμάτων της ταξινόμησης των δήμων σε ομάδες θα έχει μεγαλύτερο κέρδος εφόσον τοποθετήσει το κατάστημα στην ευρεία ζώνη που έχει αντιστοιχηθεί στον χάρτη ταξινόμησης με τον αριθμό 2 (χάρτης ταξινόμησης δήμων Αττικής σελ.61).

Τελικά, η συγκεκριμένη εφαρμογή μας οδήγησε στο συμπέρασμα ότι η μέθοδος ταξινόμησης Ομάδων (*Cluster analysis*) επιτρέπει την εύκολη συμπίεση δεδομένων καθώς και την αναγνώριση χαρακτηριστικών υποσυνόλων δεδομένων (π.χ. για σκοπούς marketing όπου επιθυμούμε να εστιάσουμε σε συγκεκριμένες ομάδες).

ΚΕΦΑΛΑΙΟ 5

ΣΥΜΠΕΡΑΣΜΑΤΑ

Το κεφάλαιο αυτό αποτελεί αξιολόγηση των εργαλείων που χρησιμοποιήθηκαν καθώς και επισκόπηση των εφαρμογών που εκπονήθηκαν στην συγκεκριμένη εργασία.

Αντικείμενο της παρούσης εργασίας αποτέλεσε η ανάπτυξη ενός μεθοδολογικού πλαισίου εφαρμογής ανάλυσης, ταξινόμησης και χωροθέτησης γεωγραφικών δεδομένων σε περιβάλλον ΓΣΠ. Το εν λόγω πλαίσιο, αποτελείται από χωρικά δεδομένα και αναλυτικές οδηγίες εφαρμογής της μη επιβλεπόμενης ταξινόμησης τους και θα μπορούσε να αξιοποιηθεί κατά την αντιμετώπιση αντίστοιχων προβλημάτων καθώς και εκπαιδευτικών ασκήσεων εξοικείωσης σπουδαστών και φοιτητών με προβλήματα γεωγραφικού χαρακτήρα.

Το κείμενο διαμορφώθηκε στοχεύοντας στην υπόδειξη κατευθυντήριων γραμμών και χρήσιμων εργαλείων στους χρήστες, έτσι ώστε να είναι σε θέση να διαχειριστούν γεωγραφική πληροφορία και να δομήσουν το δικό τους γεωγραφικό μοντέλο, εισάγοντας κάθε φορά τα δεδομένα που είναι απαραίτητα για την επίλυση των προβλημάτων που καλούνται να αντιμετωπίσουν.

5.1. Ο ρόλος των ΓΣΠ

Τα δεδομένα καταχωρήθηκαν σε περιβάλλον ΓΣΠ ακολουθώντας τις βασικές αρχές που διέπουν την οργάνωση, διαχείριση και επεξεργασία των Βάσεων Δεδομένων.

Έτσι λοιπόν η Βάση Δεδομένων που δημιουργήθηκε αποτέλεσε μια κατάλληλα συνδυασμένη συγκέντρωση του συστήματος των πρωταρχικών δεδομένων με σκοπό τη δημιουργία ενός νέου, πιο αποτελεσματικού και λειτουργικά ολοκληρωμένου αρχείου με δυνατότητα εξυπηρέτησης πολλαπλών εφαρμογών και αναγκών .

Με την ολοκλήρωση της εργασίας επιβεβαιώνεται ότι τα γεωγραφικά Συστήματα Πληροφοριών αποτελούν κατάλληλα εργαλεία διαχείρισης των δεδομένων καθώς μπορούν να αντιμετωπίσουν ζητήματα ανάλυσης του χώρου επιτυχώς. Στις συγκεκριμένες εφαρμογές μέσω των ΓΣΠ πραγματοποιήθηκε η διαχείριση, χωρικών δεδομένων με συγκεκριμένη θέση ή γεωγραφική αναφορά και συνεργαζόμενων μη χωρικών δεδομένων.

Η χρήση ΓΣΠ στα συγκεκριμένα παραδείγματα εφαρμογών, εκτός του ότι συμβάλλει στην καλύτερη και πιο φιλική παρουσίαση με στόχο την κατανόηση των βασικών εννοιών της ταξινόμησης μπορεί να θεωρηθεί ότι βοηθά:

- Στην εξοικείωση του χρήστη με ένα σύγχρονο λογισμικό πακέτο ΓΣΠ.
- Στην ανάπτυξη ικανότητας ερμηνείας χαρτών και στην επέμβαση σε αυτούς με τα διάφορα εργαλεία που προσφέρει το λογισμικό.
- Στη χρήση διαφόρων χαρτογραφικών συμβολισμών και στον πειραματισμό με αυτούς, ώστε να προκύψει ένα κατά το δυνατόν βέλτιστο οπτικά προϊόν.
- Στην εισαγωγή του χρήστη στην ανάλυση του χώρου μέσα από ρεαλιστικά παραδείγματα και στην αντιμετώπιση των διαφόρων προβλημάτων που προκύπτουν.

Συμπερασματικά τα ΓΣΠ παρέχουν τη δυνατότητα στον χρήστη να αναλύει γεωγραφικές πληροφορίες για κάποιο συγκεκριμένο σκοπό και θεωρούνται ως το πλέον εύχρηστο εργαλείο για την ανάπτυξη εφαρμογών αντίστοιχων αυτών που αναλύθηκαν στην παρούσα διπλωματική.

5.2. Ο ρόλος της ταξινόμησης

Μέσα από τη συγκεκριμένη διπλωματική διαφαίνεται επίσης ο άμεσος συσχετισμός της γεωγραφικής πληροφορίας με τη διαδικασία της ταξινόμησης.

Η ταξινόμηση χρησιμοποιείται γενικά όταν υπάρχει μεγάλος όγκος δεδομένων, τα οποία είναι πιο εύχρηστα όταν είναι κατηγοριοποιημένα και αποτελεί μία πλέον ουσιαστική διαδικασία καθώς ομαδοποιεί τις χωρικές οντότητες με βάση τα κοινά τους χαρακτηριστικά. Συγκεκριμένα, στην παρούσα διπλωματική χρησιμοποιήθηκε ο αλγόριθμος *k-means*, καθώς είναι ένας από τους πιο γνωστούς και πιο απλούς αλγόριθμους, που λύνει το πρόβλημα της ταξινόμησης και φημίζεται για την αξιοπιστία και διαφάνεια του. Τα παραπάνω χαρακτηριστικά τον καθιστούν ιδανικό στην αντιμετώπιση αντίστοιχων προβλημάτων με των εφαρμογών που παρουσιάστηκαν.

5.3. Αξιολόγηση των παραδειγμάτων εφαρμογής

Όσον αφορά στο πρώτο παράδειγμα εφαρμογής, βασίστηκε σε δεδομένα της ΕΛ.ΣΤΑΤ. και ήταν συνολικά οκτώ μεταβλητές, οι οποίες με βάση την τυπική ταξινόμηση και τον αλγόριθμο *K-Means* ομαδοποίησαν τα γεωγραφικά διαμερίσματα σε τρεις τάξεις. Η απόδοση της χωρικής κατανομής των τάξεων έγινε με τη βοήθεια ενός ΓΣΠ και τα ιδιαίτερα χαρακτηριστικά των τάξεων οδήγησαν στον χαρακτηρισμό τους. Οι τάξεις που προέκυψαν από την ταξινόμηση ήταν τελικά οι εξής (Πίνακας 4.5):

Διαμέρισμα	Τάξη
Αττική	1
Στερεά και Εύβοια	2
Πελοπόννησος	2
Ιόνιο	2
Ήπειρος	2
Θεσσαλία	2
Μακεδονία	3
Θράκη	2
Αιγαίο	2
Κρήτη	2

Πίνακας 4.5: Ταξινόμηση γεωγραφικών διαμερισμάτων

Στο δεύτερο παράδειγμα εφαρμογής ο μη επιβλεπόμενος αλγόριθμος K-Means εφαρμόζεται σε δεδομένα του νομού Αττικής προκειμένου να δημιουργηθούν κοινωνικοοικονομικές ομάδες. Τα αποτελέσματα της ταξινόμησης χρησιμοποιήθηκαν για την επιλογή της κατάλληλης ζώνης χωροθέτησης καταστήματος που απευθύνεται και υπηρετεί ανάγκες των συγκεκριμένων ομάδων.

Οι τάξεις που προέκυψαν από την ταξινόμηση φαίνονται παρακάτω (Πίνακας 4.6):

Δήμος	Τάξη
Άγιος Δημήτριος	1
Άλιμος	2
Αργυρούπολη	3
Γλυφάδα	2
Δάφνη	3
Ελληνικό	3
Ηλιούπολη	2
Καλλιθέα	1
Μοσχάτο	3
Νέα Σμύρνη	1
Παλαιό Φάληρο	1
Βούλα	2

Πίνακας 4.6: Ταξινόμηση δήμων Αττικής

Η τελική απόδοση σε ζητήματα-προβλήματα ταξινομήσεων γεωγραφικής πληροφορίας ενισχύεται καταλυτικά με την παράλληλη χρήση των βέλτιστων σύγχρονων τεχνολογικών μέσων ΓΣΠ, τα οποία παρέχουν καλύτερη ανάλυση και εποπτεία των αποτελεσμάτων της, όπως φαίνεται από το σύνολο των χαρτών που παρουσιάστηκαν.

Καταλήγοντας, μέσα από την παρούσα διπλωματική εργασία διαφαίνεται ο άμεσος συσχετισμός της γεωγραφικής πληροφορίας με τη διαδικασία της ταξινόμησης και αναγνωρίζεται η αναγκαιότητα της εκπαίδευσης χρηστών στα ΓΣΠ. Επίσης συμπεραίνεται ότι, χάρη στη διαδικασία της ταξινόμησης ο χρόνος επίλυσης προβλημάτων που έχουν ως πυρήνα τη γεωγραφική πληροφορία είναι ανεξάρτητος του όγκου της. Επιπρόσθετα, η ραγδαία διάχυση της γεωγραφικής πληροφορίας ενισχύει την ανάγκη δημιουργίας ηλεκτρονικών βιβλιοθηκών και κατ επέκταση την ύπαρξη υπηρεσιών με σκοπό την εισαγωγή, ανάλυση και παρουσίαση της σε γεωγραφικές βάσεις δεδομένων. Ωστόσο, στην συγκεκριμένη εργασία παρουσιάστηκαν πολύ αναλυτικά τα βήματα ταξινόμησης και επεξεργασίας γεωγραφικών δεδομένων σε περιβάλλον ΓΣΠ, ούτως ώστε να μπορεί και ένας ανειδίκευτος χρήστης να εφαρμόσει την παραπάνω διαδικασία.

ΒΙΒΛΙΟΓΡΑΦΙΑ

ΞΕΝΗ ΒΙΒΛΙΟΓΡΑΦΙΑ

- Anderberg M. R. (1973). *Cluster analysis for Applications*. New York: Academic Press.
- Assimakopoulos D.G. (1992). The Impact of GIS in Greece: The Greek GIS community, *Mapping Awareness & GIS in Europe*, vol 6, no 7, pp.11-13.
- Assimakopoulos D.G. (1993). *The Greek GIS community*, EGIS '93 Genoa Conference.
- Bradley P. S. and Fayyad U.M. (1998). *Refining Initial Points for K-Means Clustering*. In Proc. Of the 15th intl. Conf. on Machine Learning, pages 91-99, Madison, WL, Morgan Kaufmann Publishers Inc.
- Freeman D. et.al. (1993). *Getting Started with GIS, Teaching Geography*, Vol. 18, No 2, pp. 57-60.
- Haining R. (1994). *Designing Spatial Data Analysis Modules for Geographical Information Systems*. Hong Kong: Taylor and Fransis
- Halgamuge S. K. and Lipo W. (2005). *Classification and Clustering for Knowledge Discovery*, Springer-Verlag Berlin and Heidelberg GmbH & Co.
- Hammerly G. and Eilkan C. (2002). *Alternatives to the k means algorithm that sind better clusterings*. In Proc. Of the 11th Intl. Conf. on information and knowledge management, pages 600-607, Mc Lean, Virginia: ACM Press.
- Jain A.K. and Dubes R.C. (1998). *Algorithms for Clustering Data, Prentice-Hall advanced reference series*, Prentice-Hall, Inc., Upper Saddle River, NJ.
- Korte G. (2001). *The GIS Book: How to implement, manage, assess the value of Geographic Information System*, 5th Edition, On word Press.
- McQueen J.B. (1994). *Some Methods for Classification and Analysis of Multivariate Observations*, Proc. of the 5th Berkley Symposium on Mathematical Statistics and Probability, pp. 281-297.
- Mirkin B. (1996). *Mathematical Classification and Clustering*, Kluwer Academic Publishers Group.
- Polidorides N. (1992). *GIS in Greece, Mapping Awareness & GIS in Europe*, vol 6, no 7, pp.16-18.
- Savaressi S. M. and Boley D. (2004). *A comparative analysis on the bisecting K means and the PDDP clustering algorithms*. *Intelligent Data Analysis*, 8(4): 345-362.

ΕΛΛΗΝΙΚΗ ΒΙΒΛΙΟΓΡΑΦΙΑ

Κιμιώνης Γ. (1995). *Τα Γεωγραφικά Πληροφοριακά Συστήματα ως διδακτικά εργαλεία στην Περιβαλλοντική Εκπαίδευση*, Μεταπτυχιακή Διατριβή, Πανεπιστήμιο Κρήτης, Ρέθυμνο

Κουτσόπουλος Κ. (2003). *Εφαρμογές Γεωγραφικών Συστημάτων Πληροφοριών με τη Χρήση του Λογισμικού ArcGIS*, Αθήνα: Εκδόσεις Παπασωτηρίου

Κουτσόπουλος Κ. (2005). *Εφαρμογές του λογισμικού ArcGIS 9x με απλά λόγια*, Αθήνα: Εκδόσεις Παπασωτηρίου

Μανιάτης Ι.(1993). *Γεωγραφικά Συστήματα Πληροφοριών*, Θεσσαλονίκη: εκδόσεις Ζήτη.

Στεφανάκης Ε. (2003). *Βάσεις Γεωγραφικών Δεδομένων και Συστήματα Γεωγραφικών Πληροφοριών*, Αθήνα: Εκδόσεις Παπασωτηρίου

Ελληνική Στατιστική Υπηρεσία, 2009, Τομέας Απογραφής

ΔΙΚΤΥΑΚΟΙ ΤΟΠΟΙ

<http://courses.arch.ntua.gr/113064.html>
http://dspace.lib.ntua.gr/bitstream/123456789/889/1/kalintzeosa_kmeans.pdf
http://en.wikipedia.org/wiki/Cluster_analysis
http://en.wikipedia.org/wiki/K-means_clustering
<http://en.wikipedia.org/wiki/Minitab>
<http://gis.aua.gr/EDU-PROPTYXIAKH.html>
http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/kmeans.html
<http://invenio.lib.auth.gr/record/114601/files/ptuxiaki.pdf?version=1>
http://stin-e-taxi.blogspot.gr/2011/05/blog-post_13.html
http://users.uom.gr/~it08105/Dimitriadoy_Kyriaki_Ergasia%20proxorimena%20themata%20basewn%20dedomenwn.pdf
<http://www.autonlab.org/tutorials/kmeans.html>
http://www.ece.ntua.gr/index.php?option=com_postgraduatecourses&task=course&id=61&Itemid=56&lang=el
http://www.env.aegean.gr/labs/Remote_sensing/publications/Hazard_full_v1.pdf
<http://www.geocomputation.gr/pages/greek/erevnitiki.htm>
<http://www.geocomputation.gr/pages/greek/synedria.htm>
<http://www.giseducation.gr/portal/index.php/el/>
<http://www.hellasgi.gr/>
<http://www.jstor.org/pss/2986157>
<http://www.rob.cs.tu-bs.de/content/04-teaching/06-interactive/Kmeans/Kmeans.html>
<http://www.srcosmos.gr/srcosmos/showpub.aspx?aa=6409>
<http://www.statistics.gr/portal/page/portal/ESYE>
<http://www.uic.edu/classes/idsc/ids472/labs/kmeans.htm>
www.esri.com
www.geoinformatics.gr
www.geoscience.wisc.edu
www.wikipedia.com