



METROPOLITAN COLLEGE

C E N T R E O F E X C E L L E N C E

Spatial Data Analysis Project

Student ID: 2018506

Μάθημα: Spatial Data Analysis

MSc Data Science

Σχολή Πληροφορικής

Μητροπολιτικό Κολλέγιο

Αθήνα, Ελλάδα

Ημερομηνία:

14/05/2021

Κατάλογος περιεχομένων

Κατάλογος εικόνων	3
Άσκηση 1	4
Απάντηση	4
Άσκηση 2	6
Απάντηση	6
Άσκηση 3	8
Απάντηση	8
Τοπικές (Local) λειτουργίες	9
Εστιακές / κεντρικές λειτουργίες (Focal).....	10
Λειτουργίες ζώνης (Zonal)	12
Γενικευμένες (Global) λειτουργίες.....	12
Άσκηση 4	13
Απάντηση	13
Άσκηση 5 (Project).....	16
Ορισμός και ιστορία των θεματικών χαρτών.....	17
Τεχνικές δημιουργίας θεματικών χαρτών (είδη)	19
Υλοποίηση Project - Δημιουργία θεματικού χάρτη από πρωτογενή δεδομένα	21
Σκοπός	21
Περιγραφική ανάλυση πρωτογενών δεδομένων	22
Δημιουργία θεματικού χάρτη σε βήματα	23
Επιπλέον αριθμητική ανάλυση και δημιουργία θηκογράμματος	32
Συμπεράσματα	33
Επόμενα βήματα	35
Αναφορές	37

Κατάλογος εικόνων

Εικόνα 1 - Ψηφιδωτό μοντέλο χωρικών δεδομένων	9
Εικόνα 2 - Τοπική λειτουργία (αριθμητική πρόσθεση επιπέδων)	10
Εικόνα 3 - Κεντρική λειτουργία άθροισης τιμών (FocalSUM)	11
Εικόνα 4 - Κεντρικές λειτουργίες στατιστικής ανάλυσης	12
Εικόνα 5 - Τύποι παραθύρων γειτονίας κεντρικών λειτουργιών : α) ορθογώνιο 3 X 3, β) κύκλος, γ) δακτύλιος δ) κυκλικός τομέας ε) ακανόνιστο.	12
Εικόνα 6 - Λειτουργία ζώνης (ZonalMAX)	13
Εικόνα 7 - Γενικευμένη λειτουργία εύρους τιμών (Global Range)	14
Εικόνα 8 - Γενικός τύπος ευκλείδειας απόστασης για n διαστάσεις	14
Εικόνα 9 - Ευκλείδεια απόσταση για μια διάσταση (n=1)	15
Εικόνα 10 - Ευκλείδεια απόσταση για δύο διαστάσεις (n=2)	15
Εικόνα 11 - Γενικευμένη λειτουργία ευκλείδειας απόστασης (distance).	15
Εικόνα 12 - Γενικευμένη λειτουργία ευθείας απόστασης από τις πρωτεύουσες των νομών της Πελοποννήσου.	17
Εικόνα 13 - Θεματικός Χάρτης που δείχνει την θνησιμότητα του πληθυσμού (ανά 100.000) από κάποιον τραυματισμό.	18
Εικόνα 14 - Χαρακτηριστικό παράδειγμα θεματικού χάρτη 2 μεταβλητών (Choropleth & Proportional point symbol)	20
Εικόνα 15 - Χωροπληθής θεματικός χάρτης	20
Εικόνα 16 - Θεματικός χάρτης που χρησιμοποιεί τη μέθοδο αναλογικών σημείων	21
Εικόνα 17 - Χαρτόγραμμα που δείχνει την βιολογική γεωργία σε στρέμματα	21
Εικόνα 18 - Χάρτης κατανομής σημείων που δείχνει τις περιπτώσεις ελονοσίας στην Αφρική. Μια κουκίδα ισούται με 6400 κρούσματα.	22
Εικόνα 19 - Νεπάλ shapefile 75 συνολικών πολυγώνων, χωρίζοντας ουσιαστικά τη χώρα σε 75 γεωγραφικά διαμερίσματα	23
Εικόνα 20 - Μορφή δεδομένων CSV αρχείου με δείκτες ευημερίας του Νεπάλ.	24
Εικόνα 21 - geom_path() (αριστερά) και geom_polygon() (δεξιά)	25
Εικόνα 22 - Παραγόμενο dataframe 2 στηλών και 75 γραμμών (hpi.data)	26
Εικόνα 23 - shp_df μεταβλητή μετά την συγχώνευση	27
Εικόνα 24 - Αρχικός χάρτης χωρίς επιπλέον παραμέτρους	28
Εικόνα 25 - Βελτίωση στα χρώματα	29
Εικόνα 26 - Επιπλέον βελτιωτικές ενέργειες στον παραγόμενο θεματικό χάρτη	30
Εικόνα 27 - Προσθήκη ονομάτων στα γεωγραφικά διαμερίσματα	31
Εικόνα 28 - Χωροπληθής θεματικός χάρτης του Νεπάλ, που απεικονίζει ονομαστικά τα γεωγραφικά διαμερίσματα των οποίων ο HPI ξεπερνάει την τιμή 40.	32
Εικόνα 29 - Nepal HPI Boxplots	34
Εικόνα 30 - Ομαδοποιημένες περιοχές Νεπάλ με υψηλό και χαμηλό HPI.	35

Άσκηση 1

Δώστε τα τρία βασικά μέρη από τα οποία αποτελείται ένα Γεωγραφικό Σύστημα Πληροφοριών (ΓΣΠ). Αναλύστε το καθένα από αυτά.

Απάντηση

Ένα Γεωγραφικό Σύστημα Πληροφοριών (ΓΣΠ) είναι ένα υπολογιστικό σύστημα το οποίο απαθανατίζει, ελέγχει, αποθηκεύει και παρουσιάζει δεδομένα που σχετίζονται με συντεταγμένες πάνω στην επιφάνεια της Γης (NationalGeographic, n.d.). Τα δεδομένα αυτά μπορούν να εκφραστούν με διαφορετικούς τρόπους, όπως ακριβείς συντεταγμένες (latitude, longitude), διευθύνσεις, ταχυδρομικοί κώδικες κ.ά. Σε ένα σύνολο δεδομένων μπορούν να περιέχονται επίσης και δημογραφικά στοιχεία, όπως εισόδημα και πληθυσμός, αλλά και πληροφορίες σχετικά με την βλάστηση του εδάφους, θερμοκρασίες, υψόμετρο κ.ά. Σε γενικές γραμμές, ένα πλήρες ΓΣΠ δίνει τη δυνατότητα να εξεταστεί μια περιοχή υπό πολλά διαφορετικά πρίσματα, και έπειτα από την κατάλληλη ανάλυση να εξαχθούν πολλά και χρήσιμα συμπεράσματα.

Οι περισσότερες σύγχρονες πηγές αναφέρουν ότι ένα ΓΣΠ αποτελείται από πέντε βασικά μέρη: Υλικό (Hardware), Λογισμικό (Software), Δεδομένα (Data), Ανθρώπους/Χρήστες (People) και Μεθόδους (Methods) (Ali, 2020):

- **Υλικό (Hardware):** Είναι τα υπολογιστικά συστήματα πάνω στα οποία λειτουργεί ένα ΓΣΠ. Υπάρχει μεγάλο εύρος εφαρμογής τους, από απλούς υπολογιστές σπιτιού μέχρι κεντρικούς servers με τεράστια επεξεργαστική και υπολογιστική ισχύ. Ένα ΓΣΠ μπορεί να λειτουργήσει σε επεξεργαστή 300MHz, αλλά και σε υπερυπολογιστές των οποίων οι επεξεργαστές αξιολογούνται με TeraFLOPS¹. Σαφώς, τα αποτελέσματα (ακρίβεια, ταχύτητα, αποτελεσματικότητα) συνδέονται άρρητα με το hardware που είναι διαθέσιμο κάθε φορά. Η λέξη υλικό όμως αφορά και όλες τις συσκευές που χρησιμοποιούνται για την εισαγωγή δεδομένων σε ένα ΓΣΠ:
 - **Scanners:** Μετατρέπουν μια εικόνα σε ψηφιακή για περαιτέρω ανάλυση. Το αποτέλεσμα (output) ενός scanner μπορεί να αποθηκευτεί σε πολλές μορφές, όπως JPG, TIFF, BMP.
 - **Ψηφιοποιητής (digitizer board):** Είναι μια επίπεδη οθόνη που χρησιμοποιείται για την διανυσματοποίηση (vectorization) χωρικών δεδομένων.
- **Λογισμικό (Software):** Το λογισμικό ενός ΓΣΠ παρέχει όλες τις λειτουργίες και εργαλεία που χρειάζονται για την αποθήκευση, ανάλυση και απεικόνιση χωρικών δεδομένων. Αποτελείται και αυτό από κάποια κύρια μέρη:
 - Εργαλεία για εισαγωγή και επεξεργασία των χωρικών δεδομένων

¹ FLOPS = **F**loating **P**oint operations per **S**econd. 1 TeraFLOP είναι ένα τρισεκατομμύριο υπολογισμοί ανά δευτερόλεπτο.

- Βάση δεδομένων
- Εργαλεία ανάλυσης και οπτικοποίησης δεδομένων
- Περιβάλλον χρήστη (Graphical User Interface, GUI): Συχνά συνδυασμός των παραπάνω που προσδίδουν ένα εύχρηστο πρόγραμμα στον τελικό χρήστη έτοιμο να τρέξει ένα ΓΣΠ. Τέτοια περιβάλλοντα απαιτούν χρόνια ανάπτυξης και συντήρησης κώδικα, και έρχονται είτε ως ανοιχτό λογισμικό (open source) είτε με την αγορά κάποιου license. Πολύ σπάνια κάποιος που θέλει να εφαρμόσει χωρική ανάλυση με χρήση ενός ΓΣΠ, θα θέλει να αφιερώσει και τον χρόνο να υλοποιήσει κάτι από την αρχή, με τόσες έτοιμες λύσεις διαθέσιμες. Μερικά από τα πιο γνωστά περιβάλλοντα είναι τα ArcGIS, MapInfo, Global Mapper, AutoCAD Map κ.ά.
- **Δεδομένα (data):** Πιθανότατα το πιο σημαντικό κομμάτι ενός ΓΣΠ είναι τα δεδομένα. Τα χωρικά δεδομένα και όλα τα συμπληρωματικά (υπό μορφή πίνακα) δεδομένα, μπορούν να προέλθουν είτε από προσωπική ενασχόληση-έρευνα, είτε από δημόσιες πηγές στο διαδίκτυο, αλλά και να αγοραστούν από έναν Data Provider. Υπάρχει πληθώρα πηγών στο διαδίκτυο, με datasets σε όλες τις μορφές (GIS file formats):
 - **Raster:** Το συγκεκριμένο format αποτελείται από ψηφιακές εικόνες οι οποίες αναπαριστούνται υπό την μορφή αναγώγιμων και μεγεθύνσιμων (enlargeable) πινάκων. Οι ανωτέρω εικόνες μπορούν να έχουν την μορφή JPG, TIFF κλπ αλλά και δυαδικά αρχεία (binary files) μεγάλου μεγέθους (BLOB) έτοιμα να εισαχθούν απευθείας σε μια βάση δεδομένων.
 - **Vector:** Τα χωρικά δεδομένα εκφράζονται επίσης και ως διανύσματα, αποτελούμενα από διάφορα γεωμετρικά σχήματα. Τα κυριότερα εξ' αυτών είναι τα σημεία (points), οι γραμμές (lines or polylines) και τα 2D πολύγωνα (2D polygons). Ο συνδυασμός και η ένωση σημείων, γραμμών και πολυγώνων παράγει ένα τελικό αποτέλεσμα το οποίο είναι text-based (δεν είναι binary file δηλαδή, μπορεί να διαβαστεί από άνθρωπο και από κάποιον text editor). Τα πιο συνηθισμένα vector formats για χωρικά δεδομένα είναι GeoJSON (based on JSON), GML (based on XML) και Shapefiles.
- **Άνθρωποι/Χρήστες (People):** Ο χώρος της χωρικής ανάλυσης και των ΓΣΠ είναι περιορισμένης γνώσης, οπότε οι μηχανικοί που εργάζονται για να φτιάξουν τόσο τις συσκευές όσο και τα κατάλληλα λογισμικά για να μπορέσει να λειτουργήσει απροβλημάτιστα ένα ΓΣΠ εκτιμώνται ιδιαίτερω από την επιστημονική (και όχι μόνο) κοινότητα. Ο όρος χρήστης όμως αναφέρεται και σε κάποιον ο οποίος δεν έχει τις τεχνικές γνώσεις για να δημιουργήσει κάτι από τα προαναφερθέντα, αλλά είναι άρτια εκπαιδευμένος ως τελικός αποδέκτης, ώστε να εφαρμόζει τις επιστημονικές του γνώσεις πάνω στα κατάλληλα διαμορφωμένα προγράμματα, με στόχο εν τέλει την ανάλυση των χωρικών δεδομένων και την εξαγωγή συμπερασμάτων που προκύπτουν

από την έρευνα. Ο συνδυασμός λοιπόν των μηχανικών (που φτιάχνουν τα κατάλληλα εργαλεία) και των χρηστών (που τα χρησιμοποιούν) αποτελεί τους ανθρώπους οι οποίοι είναι ένα αναπόσπαστο και πολύ σημαντικό μέρος ενός ΓΣΠ.

- **Μέθοδοι (Methods):** Ένα ΓΣΠ λειτουργεί αποτελεσματικά και με επιτυχία, βασισμένο σε ένα μεθοδικό και μελετημένο πλάνο και εμπορικούς κανόνες, τα οποία είναι μοναδικά στοιχεία και πρακτικές για κάθε οργανισμό που ασχολείται με τη χωρική ανάλυση. Το πλάνο, οι κανόνες και οι μέθοδοι που χρησιμοποιούνται σε ένα ΓΣΠ είναι ένα σύνολο από τα παρακάτω:
 - Υπολογισμός της διάστασης, του πρίσματος κάτω από το οποίο ένα σύνολο δεδομένων θα υποστεί ανάλυση.
 - Απεικόνιση και αποθήκευση των υπολογισμών και των δεδομένων, συνήθως σε μια σχεσιακή βάση δεδομένων.
 - Ο κατάλληλος χειρισμός και μετασχηματισμός των δεδομένων, προκειμένου να εξαχθούν συμπεράσματα αλλά και να προκύψουν νέες σχέσεις, διασυνδέοντας δεδομένα από διαφορετικές πηγές.
 - Υλοποιήσεις και τεχνικές δημιουργίας ενός χάρτη για ερευνητικούς σκοπούς, ο οποίος μπορεί να προκύψει είτε από αυτόματη μετατροπή raster to vector, πολλές φορές όμως χρειάζεται χειροκίνητη δουλειά από προγραμματιστές.

Άσκηση 2

Από τι προσδιορίζονται τα γεωγραφικά δεδομένα (δεδομένα που έχουν ορισθεί χωρικά).

Απάντηση

Γεωγραφικά δεδομένα είναι τα δεδομένα τα οποία μπορούν να χαρτογραφηθούν, δεδομένα που μπορούν να εντοπισθούν στο χώρο (Wikipedia, Wikipedia, 2021). Ενσωματώνουν τέσσερα διακριτά στοιχεία:

- Θέση στο χώρο
- Περιγραφή
- Χωρική Σχέση
- Χρόνο

Όπως έχει αναλυθεί και στο πρώτο ερώτημα, τα γεωγραφικά δεδομένα μπορούν να αναπαρίστανται με δύο βασικές μορφές: τα δεδομένα διανυσματικής μορφής (vector) και τα δεδομένα κανονικοποιημένης ψηφιδωτής μορφής (raster). Σε όλα τα ΓΣΠ οι δύο αυτές μορφές

αποδίδονται ταυτόχρονα σε κοινές απεικονίσεις, ενώ πολλά λογισμικά GIS² προσφέρουν τη δυνατότητα εύκολης μετάβασης από τη μία μορφή στην άλλη.

1. Διανυσματικά δεδομένα (vector data): Τα διανυσματικά δεδομένα διακρίνονται σε

- a. **Χωρικά δεδομένα.** Τα χωρικά δεδομένα μπορούν να αναπαρασταθούν με σημεία, γραμμές και πολύγωνα. Για παράδειγμα η απόδοση της θέσης μια πόλης σε ένα χάρτη μπορεί να χρησιμοποιηθεί με ένα σημείο, η αποτύπωση του οδικού δικτύου με μια γραμμή αποτελούμενη από πολλές κορυφές και η αποτύπωση μιας ιδιοκτησίας με ένα πολύγωνο. Η γεωμετρία που θα υιοθετηθεί για το συμβολισμό ενός αντικειμένου εξαρτάται από την κλίμακα απεικόνισης και το σκοπό της εφαρμογής που αναπτύσσεται. Έτσι π.χ. σε μια πολύ μεγάλη χαρτογραφική κλίμακα (1:1000) τα κτίσματα αποτυπώνονται ως πολύγωνα, ενώ σε μικρότερες χαρτογραφικές κλίμακες (1:10.000) είναι ορθότερο να χρησιμοποιηθεί η γεωμετρία του σημείου. Τέλος κάθε γεωμετρία συνδέεται με μια σχέση 1-1 με μια εγγραφή σε ένα πίνακα περιγραφικών χαρακτηριστικών.
- b. **Περιγραφικά δεδομένα.** Τα περιγραφικά δεδομένα έχουν να κάνουν με τους πίνακες ιδιοτήτων μίας γεωγραφικής οντότητας πάνω σε ένα επίπεδο. Κάθε γεωγραφικό δεδομένο έχει πάντα έναν πίνακα πληροφορίας.

2. Κανονικοποιημένης ψηφιδωτής μορφής δεδομένα (raster data): Τα raster αποτελούνται από κελιά τα οποία διαθέτουν είτε δεκαδικούς αριθμούς, είτε ακέραιους. Χρησιμοποιούνται σε περιπτώσεις που το χωρικό φαινόμενο που αποτυπώνεται, χαρακτηρίζεται ως συνεχής μεταβλητή (π.χ. το υψόμετρο του εδάφους, η εξέλιξη μίας πυρκαγιάς, κλπ.) ή σε περιπτώσεις που στο ΓΣΠ θέλουμε να ενσωματώσουμε μια δορυφορική εικόνα ή μια σαρωμένη αεροφωτογραφία. Τα κανονικοποιημένης ψηφιδωτής μορφής δεδομένα έχουν περιορισμένες δυνατότητες σύνδεσης με περιγραφικά χαρακτηριστικά.

Οι δομές των δεδομένων διακρίνονται σε επιμέρους κατηγορίες:

- **Μη Τοπολογική Δομή-Σπαγγέτι (απλή σειριακή δομή).** Σε αυτή τη κατηγορία δεν αναπτύσσονται χωρικές σχέσεις μεταξύ των γεωγραφικών δεδομένων στα επίπεδα (layers), αλλά πραγματοποιείται καταγραφή των γεωγραφικών συντεταγμένων αυτών των οντοτήτων.
- **Τοπολογία (χωρικές σχέσεις).** Σε αυτή τη κατηγορία αναπτύσσονται οι σχέσεις μεταξύ των γεωγραφικών δεδομένων (είτε vector, είτε raster), τις οποίες αποθηκεύει ο υπολογιστής σε μορφή τόξου-κόμβου και πολυγώνου και χρησιμοποιεί η χωρική ανάλυση.

² Geographic Information System ή αλλιώς Γεωγραφικό Πληροφοριακό Σύστημα (ΓΣΠ)

Οι σχέσεις μεταξύ των γεωγραφικών δεδομένων είναι οι εξής:

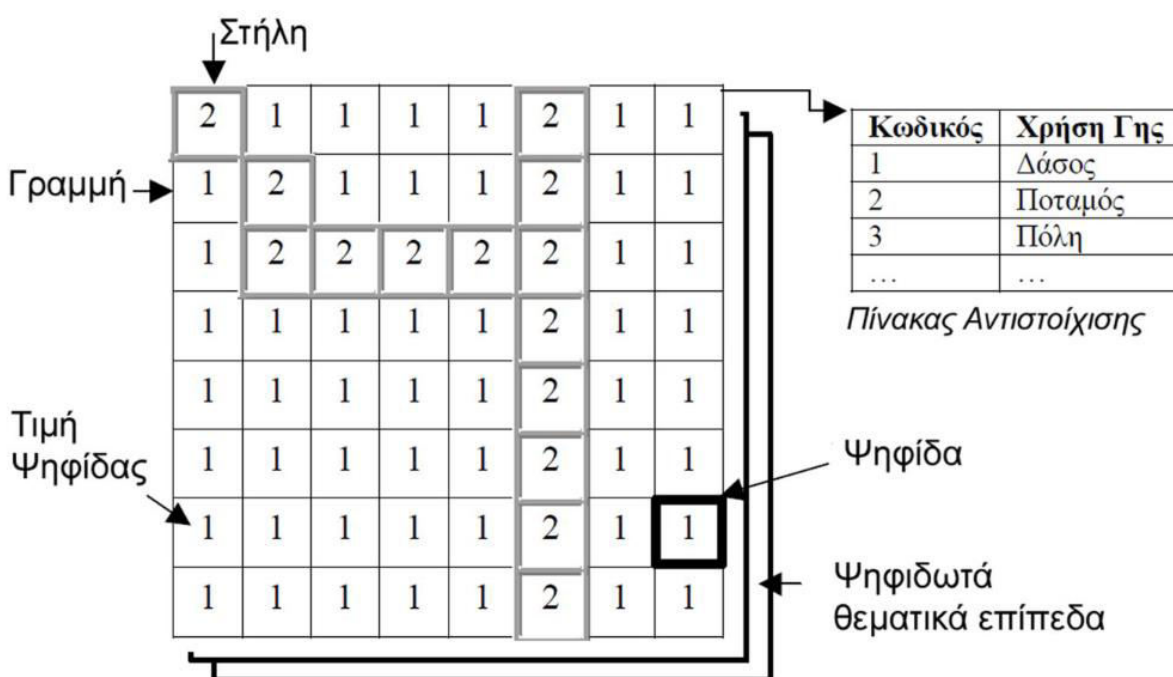
1. Συνδεσιμότητα
2. Περιεκτικότητα
3. Γειτνίαση

Άσκηση 3

Δώστε τον ορισμό και από ένα παράδειγμα τοπικής, εστιακής, ζώνης και γενικευμένης λειτουργίας σε ψηφιδωτά δεδομένα.

Απάντηση

Στο ακαδημαϊκό σύγγραμμα του Αν. Καθηγητή Χρίστου Χαλκιά (Χαροκόπειο Πανεπιστήμιο, Τμήμα Γεωγραφίας) περιγράφονται εις βάθος οι διαφορετικές λειτουργίες αναφορικά με την ανάλυση ψηφιδωτών δεδομένων (Χαλκιάς, 2015). Το ψηφιδωτό μοντέλο χωρικών δεδομένων έχει ως βασικό χαρακτηριστικό την εξέταση τμημάτων του χώρου, τα οποία συνήθως έχουν τη μορφή περιοχών τετραγωνικού σχήματος οι οποίες καλούνται ψηφίδες ή κελιά. Τα ψηφιδωτά δεδομένα οργανώνονται σε επίπεδα (raster layers ή grids) ορθογώνιου σχήματος αφού αποτελούνται από ένα συγκεκριμένο αριθμό στηλών και γραμμών οι οποίες περιέχουν τετραγωνικές ψηφίδες.



Εικόνα 1 - Ψηφιδωτό μοντέλο χωρικών δεδομένων

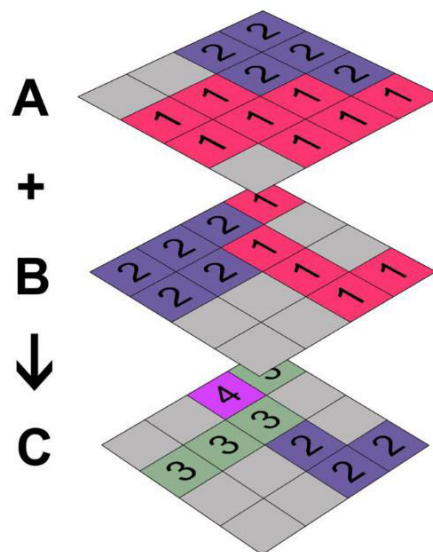
Η ανάλυση και επεξεργασία των ψηφιδωτών δεδομένων αποτελείται από τέσσερις βασικές κατηγορίες:

- Τοπικές (**local**)
- Κεντρικές ή λειτουργίες γειτονίας (**focal**)
- Λειτουργίες ζωνών (**zonal**)
- Γενικευμένες (**global**)

Στις παραπάνω κατηγορίες χρησιμοποιείται η θεωρία της άλγεβρας χαρτών η οποία τεκμηριώθηκε στις αρχές τις δεκαετίας του 1990 (Tomlin, 1990). Μέσω αυτής της θεωρίας, η οποία στην ουσία είναι μια υπολογιστική γλώσσα, η επιφάνεια της γης μοντελοποιείται με τη χρήση επιπέδων γεωγραφικών δεδομένων τα οποία είναι θεματικά ανεξάρτητα μεταξύ τους, αλλά έχουν χωρική ταύτιση.

Τοπικές (Local) λειτουργίες

Οι τοπικές λειτουργίες μπορούν να εφαρμοστούν είτε σ' ένα μεμονωμένο είτε συνδυαστικά σε δύο ή περισσότερα ψηφιδωτά θεματικά επίπεδα. Η εφαρμογή μιας τοπικής λειτουργίας σ' ένα θεματικό επίπεδο συνήθως εξαρτάται από την εφαρμογή μιας μαθηματικής ή και λογικής πράξης για κάθε ψηφίδα του επιπέδου. Η τιμή για κάθε ψηφίδα του παραγόμενου επιπέδου μετά από μια τοπική λειτουργία εξαρτάται αποκλειστικά και μόνο από την τιμή της συγκεκριμένης ψηφίδας του επιπέδου (ή των επιπέδων) εισόδου. Οι τοπικές λειτουργίες μπορεί να έχουν τη μορφή απλών αριθμητικών πράξεων (πρόσθεση, αφαίρεση, πολλαπλασιασμός, διαίρεση), μαθηματικών λειτουργιών (λογαριθμικές, εκθετικές, τριγωνομετρικές κ.λ.π), λογικών τελεστών και συνθηκών (π.χ. AND, OR, NOT, IF, THEN, ELSE), επαναταξινομήσεων ή επιλογών σ' ένα ψηφιδωτό θεματικό επίπεδο, καθώς και άλλων σύνθετων λειτουργιών (π.χ. συνδυασμοί / COMBINE επιπέδων, έλεγχος συνθηκών / εντολή CON).

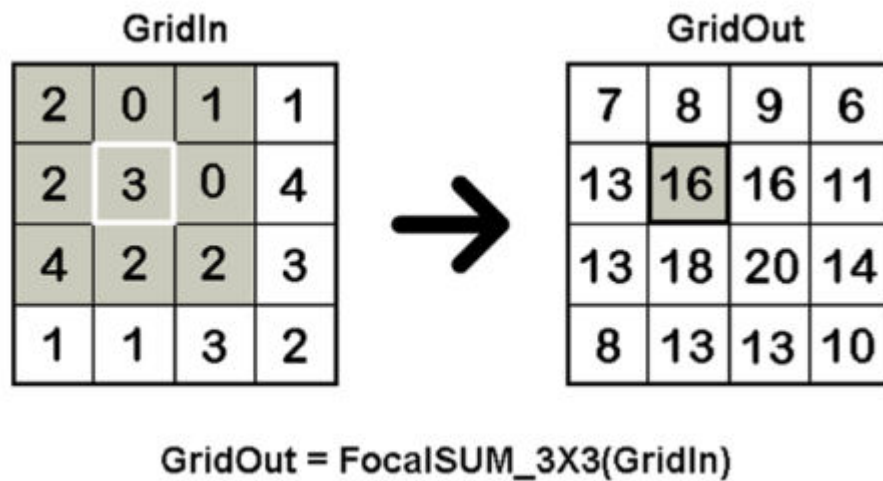


Εικόνα 2 - Τοπική λειτουργία (αριθμητική πρόσθεση επιπέδων)

Η εκτέλεση μιας τοπικής λειτουργίας σε δύο ή περισσότερα θεματικά επίπεδα είναι ιδιαίτερα συνηθισμένη λειτουργία σε πολλές εφαρμογές γεωγραφικής ανάλυσης. Σε μοντέλα σταθμισμένης χαρτογραφικής υπέρθεσης συχνά η τελική σύνθεση των κριτηρίων/μεταβλητών του μοντέλου γίνεται με την εκτέλεση τοπικής λειτουργίας της παρακάτω μορφής: $Output = Layer_1 * w1 + Layer_2 * W2 + \dots + Layer_n * wn$, όπου $Layer_1, 2, \dots, n$ είναι τα θεματικά επίπεδα των κριτηρίων/μεταβλητών τα οποία συνδυάζονται και $w1, 2, \dots, n$ είναι συντελεστές βαρύτητας κάθε κριτηρίου.

Εστιακές / κεντρικές λειτουργίες (Focal)

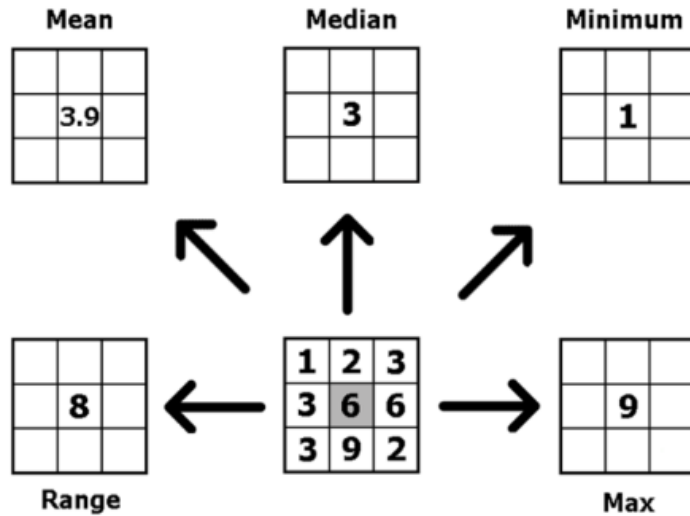
Οι λειτουργίες αυτές οι οποίες καλούνται και λειτουργίες γειτονίας, σχετίζονται άμεσα με τον 1^ο νόμο της Γεωγραφίας (Tobler, 1970) σύμφωνα με τον οποίο κάθε αντικείμενο του χώρου σχετίζεται με όλα τα άλλα αντικείμενα, αλλά τα κοντινά μεταξύ τους αντικείμενα σχετίζονται περισσότερο από τα απομακρυσμένα. Οι λειτουργίες αυτές μπορούν να εφαρμοστούν είτε σ' ένα μεμονωμένο είτε συνδυαστικά σε δύο ή περισσότερα ψηφιδωτά θεματικά επίπεδα και για την εκτέλεσή τους χρειάζεται να οριστεί η περιοχή (ή παράθυρο) γειτονίας (kernel ή moving window). Σ' αυτές τις λειτουργίες, η τιμή για μια συγκεκριμένη ψηφίδα στο παραγόμενο θεματικό επίπεδο είναι μια συνάρτηση της τιμής των ψηφίδων σε περιοχή γειτονίας περιμετρικά αυτής της ψηφίδας. Για παράδειγμα, αν οριστεί ως περιοχή γειτονίας ένα παράθυρο μεγέθους 3X3 μπορεί να εκτελεστεί η κεντρική λειτουργία άθροισης τιμών.



Εικόνα 3 - Κεντρική λειτουργία άθροισης τιμών (FocalSUM)

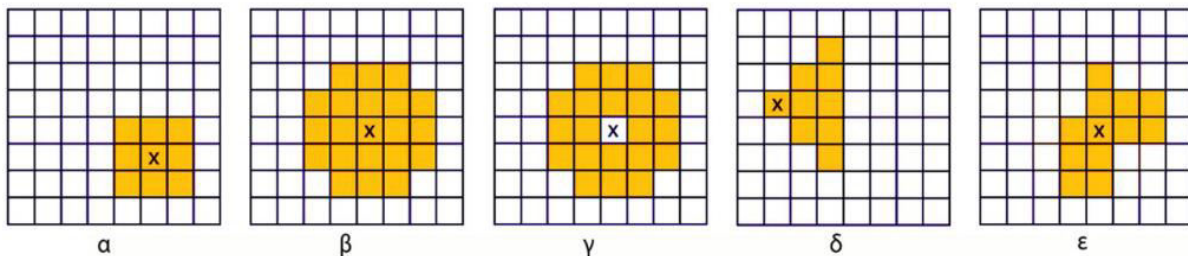
Συνεπώς, για τη σημειωμένη ψηφίδα με αρχική τιμή 3, υπολογίζεται το άθροισμα των τιμών των ψηφίδων γύρω από αυτήν σε παράθυρο 3X3 (άθροισμα = 2+0+1+2+3+0+4+2+2) και η τιμή αυτή (16) είναι η τιμή της ψηφίδας στο παραγόμενο επίπεδο. Θα πρέπει να σημειωθεί ότι, στα όρια του επιπέδου το παράθυρο 3X3 δεν περιέχει 9 άλλα 6 (στις πλευρές) ή 4 (στις

γωνίες) ψηφίδες. Έτσι, οι υπολογισμοί των κεντρικών λειτουργιών σε αυτές τις ψηφίδες προκύπτουν από τις αρχικές τιμές και δε λαμβάνουν υπόψη τις γειτονικές τιμές, εκτός των ορίων του επιπέδου. Αν αυτό δεν είναι επιθυμητό, μια πιθανή λύση δίνει η επέκταση της περιοχής μελέτης. Οι κεντρικές λειτουργίες στατιστικής ανάλυσης (Focal_mean, Focal_StDev, Focal_sum, Focal_Min, Focal_Max, Focal_Range, Focal_Majority, Focal_Frequency κλπ) σε ψηφιδωτό θεματικό επίπεδο αποτελούν χαρακτηριστικά παραδείγματα λειτουργιών αυτού του τύπου.



Εικόνα 4 - Κεντρικές λειτουργίες στατιστικής ανάλυσης

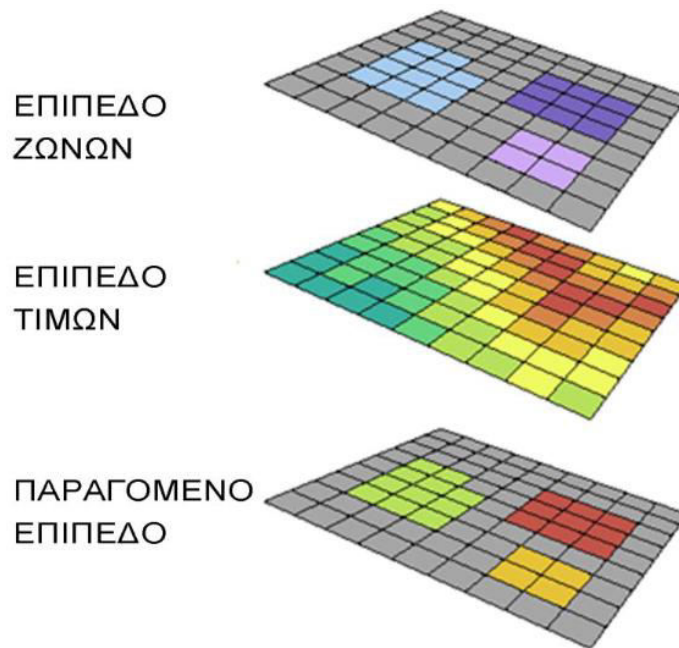
Η εκτέλεση των κεντρικών λειτουργιών, βασίζεται στην ιδέα του «κινητού παραθύρου» το οποίο διατρέχει το επίπεδο εισόδου και το οποίο ορίζει την περιοχή επεξεργασίας τιμών για κάθε ψηφίδα. Η επεξεργασία αρχίζει να εκτελείται από την πάνω αριστερή ψηφίδα, καθώς το παράθυρο κινείται από αριστερά προς τα δεξιά και από πάνω προς τα κάτω. Το σχήμα του κινητού παράθυρου γειτονίας μπορεί να είναι ορθογώνιο, κύκλος, δακτύλιος, κυκλικός τομέας ή ακανόνιστο.



Εικόνα 5 - Τύποι παραθύρων γειτονίας κεντρικών λειτουργιών : α) ορθογώνιο 3 X 3, β) κύκλος, γ) δακτύλιος δ) κυκλικός τομέας ε) ακανόνιστο.

Λειτουργίες ζώνης (Zonal)

Στις λειτουργίες ζώνης, μονάδα ανάλυσης αποτελούν οι περιοχές με εσωτερική ομοιογένεια στο επίπεδο ζωνών. Σ' αυτές τις λειτουργίες η τιμή για κάθε ψηφίδα του παραγόμενου επιπέδου υπολογίζεται με βάση τις τιμές όλων των ψηφίδων της ζώνης στην οποία ανήκει η συγκεκριμένη ψηφίδα στο επίπεδο εισόδου. Θα πρέπει να σημειωθεί ότι, οι ζώνες αυτές αποτελούνται από ψηφίδες οι οποίες δεν είναι απαραίτητα γειτονικές. Για το λόγο αυτό, στο παραγόμενο επίπεδο όλες οι ψηφίδες εντός συγκεκριμένης ζώνης έχουν την ίδια τιμή. Από τα παραπάνω, γίνεται φανερό ότι το αποτέλεσμα μιας λειτουργίας ζώνης μπορεί να αποδοθεί και σε πίνακα με αριθμό εγγραφών ίσο με τον αριθμό των ζωνών. Για παράδειγμα, αν επιχειρείται η εύρεση της μέγιστης κλίσης του εδάφους, για κάθε τύπο κάλυψης γης σε μια περιοχή ενδιαφέροντος θα πρέπει να εκτελεστεί η λειτουργία ZonalMax στο επίπεδο Κλίση_εδάφους με τον ορισμό του επιπέδου Εδαφική_κάλυψη ως επίπεδο ζωνών.



Εικόνα 6 - Λειτουργία ζώνης (ZonalMAX)

Γενικευμένες (Global) λειτουργίες

Στις γενικευμένες λειτουργίες οι τιμές σε κάθε ψηφίδα του παραγόμενου επιπέδου υπολογίζονται δυνητικά συναρτήσεως όλων των ψηφίδων του επιπέδου εισόδου. Έτσι, οι γενικευμένες λειτουργίες είναι ανάλογες με τις λειτουργίες ζώνης αν θεωρήσουμε ότι το σύνολο του επιπέδου εισόδου αντιπροσωπεύει μια ζώνη. Παραδείγματα αυτού του τύπου αποτελούν οι λειτουργίες εύρεσης απόστασης, οι λειτουργίες παρεμβολής, οι γενικευμένες στατιστικές λειτουργίες (π.χ. λειτουργία globalrange), οι λειτουργίες υδρολογικής ανάλυσης, οι λειτουργίες ανάλυσης ορατότητας κλπ.

ΑΡΧΙΚΟ ΕΠΙΠΕΔΟ					ΠΑΡΑΓΟΜΕΝΟ ΕΠΙΠΕΔΟ				
456	416	364	326	243	451	451	451	451	451
448	364	315	276	218	451	451	451	451	451
359	325	268	234	164	451	451	451	451	451
306	296	201	133	44	451	451	451	451	451
274	231	184	65	5	451	451	451	451	451

Global_Range
→

Εικόνα 7 - Γενικευμένη λειτουργία εύρους τιμών (Global Range)

Άσκηση 4

Τι γνωρίζετε για την ευκλείδεια απόσταση; Ποια τα βήματα κατασκευής χάρτη ευκλείδειας απόστασης από γεωγραφικές οντότητες

Απάντηση

Στα μαθηματικά, η Ευκλείδεια απόσταση 2 σημείων ορίζεται ως το μήκος του ευθύγραμμου τμήματος που ενώνει τα 2 σημεία. Το μήκος αυτό μπορεί να υπολογιστεί από τις καρτεριανές συντεταγμένες των σημείων, χρησιμοποιώντας το Πυθαγόρειο Θεώρημα³, και γι' αυτό το λόγο ονομάζεται και Πυθαγόρεια απόσταση.

Ο γενικός τύπος για τον υπολογισμό της ευκλείδειας απόστασης ανάμεσα σε 2 σημεία, σε χώρο n διαστάσεων είναι ο παρακάτω:

$$d(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_i - q_i)^2 + \dots + (p_n - q_n)^2}.$$

Εικόνα 8 - Γενικός τύπος ευκλείδειας απόστασης για n διαστάσεις

Ο παραπάνω τύπος μπορεί να χρησιμοποιηθεί για όλες τις διαστάσεις, με τις κατάλληλες τροποποιήσεις. Για παράδειγμα, σε ένα μονοδιάστατο χώρο, απλοποιείται ως εξής

³ Το Πυθαγόρειο Θεώρημα είναι η σχέση της ευκλείδειας γεωμετρίας ανάμεσα στις πλευρές ενός ορθογωνίου τριγώνου. Ορίζεται ως «το τετράγωνο της υποτεινούς ενός ορθογωνίου τριγώνου ισούται με το άθροισμα των τετραγώνων των δύο κάθετων πλευρών».

$$d(p, q) = \sqrt{(p - q)^2}.$$

Εικόνα 9 - Ευκλείδεια απόσταση για μια διάσταση ($n=1$)

Και αντιστοίχως σε ένα δισδιάστατο χώρο (που έχει συνήθη εφαρμογή και στη χωρική ανάλυση)

$$d(p, q) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2}.$$

Εικόνα 10 - Ευκλείδεια απόσταση για δύο διαστάσεις ($n=2$)

Η ευκλείδεια απόσταση αποτελεί θεμελιώδη έννοια στα μαθηματικά και τη γεωμετρία, και έχει και τις εξής ιδιότητες:

- Είναι **συμμετρική**, που σημαίνει ότι για όλα τα σημεία x, y ισχύει $d(x, y) = d(y, x)$. Αυτό με απλά λόγια σημαίνει ότι η απόσταση είναι ίδια για 2 σημεία, ανεξαρτήτως το ποιο σημείο θα αποτελέσει την αρχή και ποιο τον προορισμό.
- Είναι **θετική**, που σημαίνει ότι η ευκλείδεια απόσταση ανάμεσα σε 2 σημεία είναι πάντα θετικός αριθμός, και η απόσταση ενός σημείου από τον εαυτό του είναι 0.
- Υπακούει στην **τριγωνική ανισότητα**. Αυτό σημαίνει πως για τρία σημεία x, y, z ισχύει η ανισότητα $d(x, y) + d(y, z) \geq d(x, z)$. Με λόγια, η απόσταση για να ταξιδέψει κάποιος από το σημείο x στο σημείο z , μέσω του σημείου y , είναι πάντα μεγαλύτερη από την απόσταση για να ταξιδέψει κάποιος από το x στο z απευθείας.

Όσον αφορά τα χωρικά δεδομένα, η ευκλείδεια απόσταση θεωρείται ως μια γενικευμένη λειτουργία μέτρησης απόστασης και χρησιμοποιείται ευρύτατα σε εφαρμογές γεωγραφικής ανάλυσης. Στο παρακάτω σχήμα φαίνεται η εκτέλεση της γενικευμένης λειτουργίας ευκλείδειας απόστασης από τις ψηφίδες του αρχικού επιπέδου με τιμές 1 και 2. Στο παραγόμενο θεματικό επίπεδο υπολογίζεται για κάθε ψηφίδα η απόσταση προς την πλησιέστερη ψηφίδα με τιμή 1 ή 2. Στο αποτέλεσμα η απόσταση αυτή καταγράφεται σε μονάδα μήκους ίση με την πλευρά κάθε ψηφίδας.

		1	1	Ευκλείδεια απόσταση από 1 και 2	2	1	0	0
			1		1.4	1	1	0
	2				1	0	1	1
					1.4	1	1.4	2

Εικόνα 11 - Γενικευμένη λειτουργία ευκλείδειας απόστασης (distance).

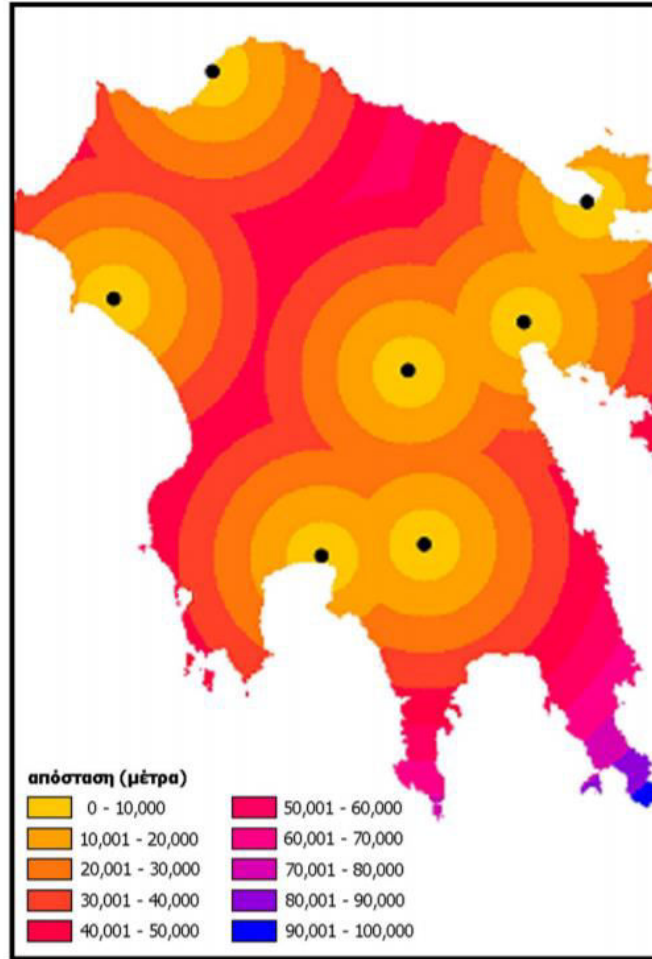
Σε πιο πολύπλοκες περιπτώσεις, όπως είναι αυτές του πραγματικού κόσμου με πραγματικά δεδομένα, θα μπορούσαμε να ορίσουμε τα εξής βήματα για την ορθή κατασκευή ενός χάρτη ευκλείδειας απόστασης:

- 1. Εισαγωγή του χάρτη αναφοράς.** Προκειμένου να φτιαχτεί ένας χάρτης ευκλείδειας απόστασης, θα πρέπει να γίνει εισαγωγή στο πρόγραμμα ενός αρχικού χάρτη (σε μορφή raster data, shapefile κλπ) πάνω στον οποίο θα γίνει όλο το υπόλοιπο χτίσιμο. Στο παράδειγμα της εικόνας 12, έχει δοθεί ο χάρτης της Πελοποννήσου ως χάρτης αναφοράς.
- 2. Ορισμός των γεωγραφικών οντοτήτων.** Σαν δεύτερο βήμα θα πρέπει να καθοριστεί από τον ερευνητή τι θεωρείται γεωγραφική οντότητα, και να δώσει ως input στο λογισμικό τις συντεταγμένες των οντοτήτων για τις οποίες θέλει να υπολογίσει ευκλείδειες αποστάσεις. Για παράδειγμα (προς καλύτερη κατανόηση) στην εικόνα 12 φαίνεται ότι οι γεωγραφικές οντότητες του συγκεκριμένου χάρτη είναι οι πρωτεύουσες των νομών της Πελοποννήσου.
- 3. Καθορισμός της μονάδας μέτρησης απόστασης.** Η ευκλείδεια απόσταση υπολογίζεται είτε με μετρικές (μέτρα, εκατοστά, χιλιόμετρα) είτε με βρετανικές (ίντσες, μίλια, ναυτικά μίλια) μονάδες μέτρησης. Θα πρέπει να καθοριστεί σαφώς στο πρόγραμμα ποια μονάδα να χρησιμοποιηθεί για την απεικόνιση των αποτελεσμάτων στο χάρτη. Στο παράδειγμα έχει χρησιμοποιηθεί η μονάδα του μέτρου.
- 4. Ρύθμιση των χρωματικών ζωνών.** Για να είναι πιο ευπαρουσίαστος ο χάρτης, και πιο ευδιάκριτα τα αποτελέσματα, θα πρέπει να δωθούν οι κατάλληλες προγραμματιστικές οδηγίες στο λογισμικό, ώστε να απεικονίσει με διαφορετικά χρώματα τις διάφορες χρωματικές ζώνες, ανάλογα με την απόσταση. Κάτι τέτοιο φαίνεται χαρακτηριστικά και στην εικόνα 12.
- 5. Εκτέλεση των τελικών εντολών για τον υπολογισμό και την απεικόνιση του χάρτη.** Αφού έχουν προηγηθεί όλα τα παραπάνω, τότε θα πρέπει να εκτελεστούν οι τελικές εντολές (υπολογισμός αποστάσεων και κατασκευή χάρτη) προκειμένου να υπάρχει το οπτικό αποτέλεσμα. Αυτές οι εντολές σαφώς ποικίλλουν από πρόγραμμα σε πρόγραμμα. Για παράδειγμα, στο ArcGIS, ένα από τα πιο διαδεδομένα λογισμικά χωρικής ανάλυσης, κάτι τέτοιο γίνεται με την εντολή EucDistance, η οποία δέχεται αρκετές παραμέτρους⁴.

Σε κάθε ψηφίδα του παραγόμενου GRID καταγράφεται η ευθεία απόστασή της από το πλησιέστερο αντικείμενο αναφοράς. Το αποτέλεσμα αυτής της λειτουργίας είναι ένα ψηφιδωτό επίπεδο το οποίο περιέχει τα δεδομένα υπολογισμού ευθείας απόστασης απ' όλα τα αντικείμενα αναφοράς. Η λειτουργία αυτή χρησιμοποιείται είτε αυτόνομα είτε σε

⁴ Input data (δεδομένα εισαγωγής), maximum_distance (μέγιστη απόσταση), cell_size (μέγεθος κελιού) κ.ά.

εφαρμογές κατασκευής χαρτών καταλληλότητας, όπου ως ένα από τα κριτήρια καταλληλότητας εμφανίζεται και η ευθεία απόσταση από συγκεκριμένες γεωγραφικές οντότητες (Χαλκιάς, 2015).



Εικόνα 12 - Γενικευμένη λειτουργία ευθείας απόστασης από τις πρωτεύουσες των νομών της Πελοποννήσου.

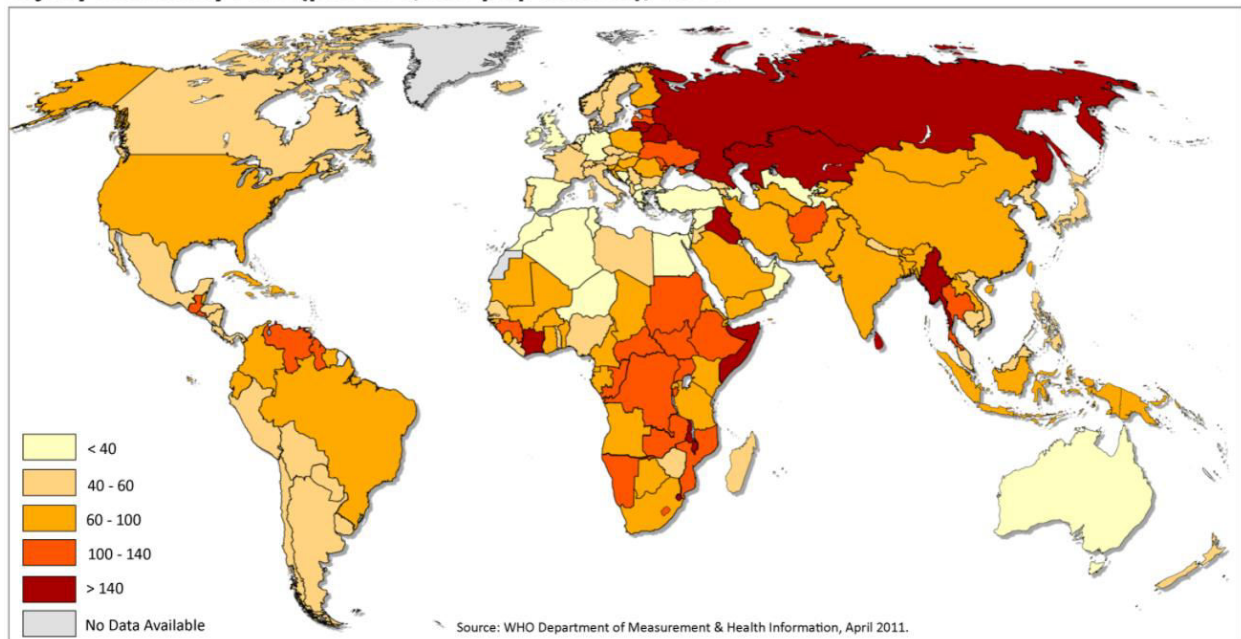
Άσκηση 5 (Project)

Παρουσιάστε τη δημιουργία θεματικού χάρτη της επιλογής σας από δεδομένα μορφής *shapefile* που είναι δημόσια διαθέσιμα στο διαδίκτυο. Τα πρωτογενή δεδομένα θα πρέπει να χαρτογραφηθούν και να αναλυθούν στατιστικά και γραφικά σύμφωνα με τις μεθόδους που έχουν παρουσιαστεί στο μάθημα.

Ορισμός και ιστορία των θεματικών χαρτών

Θεματικός χάρτης είναι ένας τύπος χάρτη ο οποίος απεικονίζει τη διάταξη που ακολουθεί ένα συγκεκριμένο θέμα υπό έρευνα, σε μια προκαθορισμένη γεωγραφική περιοχή. Κάτι τέτοιο συνήθως συμπεριλαμβάνει τη χρήση συμβόλων και χρωμάτων, προκειμένου να οπτικοποιηθούν επιλεγμένα γεωγραφικά χαρακτηριστικά τα οποία δεν είναι φυσικώς ορατά. Παραδείγματα τέτοιων χαρακτηριστικών είναι η θερμοκρασία, η γλώσσα ή ο πληθυσμός (Petchenik, 1979). Με αυτόν τον τρόπο τα υπό έρευνα χαρακτηριστικά ξεχωρίζουν από τις οντότητες που έχει ένας χάρτης γενικής αναφοράς, όπως δρόμοι, κτίρια, ποτάμια κλπ. Οι θεματικοί χάρτες έχουν και εναλλακτικές ονομασίες, όπως χάρτες ειδικού σκοπού, στατιστικοί χάρτες ή χάρτες κατανομής (distribution maps), αλλά αυτές οι ονομασίες δεν είναι ευρέως χρησιμοποιούμενες (Dent, Torguson, & Hodler, 2009).

Injury mortality rate (per 100,000 population), 2008



Εικόνα 13 - Θεματικός Χάρτης που δείχνει την θνησιμότητα του πληθυσμού (ανά 100.000) από κάποιον τραυματισμό.

Οι πρώτοι αξιόλογοι θεματικοί χάρτες έχουν καταγραφεί στην ιστορία από το 18^ο και 19^ο αιώνα, την εποχή που ξεκίνησε η μαζική συλλογή χωρικών δεδομένων με σκοπό την εξαγωγή χρήσιμων συμπερασμάτων που απορρέουν από την κατανομή διαφόρων γεωγραφικών χαρακτηριστικών (Robinson, 1982). Οι χάρτες αυτοί αποτέλεσαν μια ακόμη καινοτομία της Βιομηχανικής Επανάστασης⁵, υιοθετώντας και κάποια στοιχεία από την εποχή

⁵ Περίοδος της ιστορίας (1760-1840) κατά την οποία σημειώθηκαν ραγδαίες αλλαγές στην οικονομία και την κοινωνία, οι οποίες οφείλονται στην ανακάλυψη και χρήση νέων εργαλείων στη βιομηχανία τα οποία αντικαθιστούσαν τη χειρωνακτική εργασία.

του Διαφωτισμού⁶, αλλά λόγω της εποχής (μέχρι και τα μέσα του 17^{ου} αιώνα) ήταν χαμηλής ποιότητας, όπως και οι απλοί χάρτες οι οποίοι απλά απεικόνιζαν προσεγγιστικά και με μικρή ακρίβεια τα βασικά σύνορα πόλεων, χωρών κλπ. Παρ' όλα αυτά, οι τεχνικές που αναπτύχθηκαν σε αυτή την εποχή (και θα αναλυθούν παρακάτω) για τη δημιουργία των θεματικών χαρτών, χρησιμοποιούνται ως βάση μέχρι και σήμερα όπου η τεχνολογία, τα λογισμικά και οι συσκευές που χρησιμοποιούνται για χωρική ανάλυση είναι σαφώς πολύ ανώτερα από τα εφόδια που είχαν στη διάθεση τους οι τότε ερευνητές.

Ο σκοπός ενός θεματικού χάρτη είναι να απεικονίσει τη γεωγραφική κατανομή ενός ή περισσότερων (φυσικών, δημοκρατικών, οικονομικών) φαινομένων. Στις περισσότερες περιπτώσεις ο χαρτογράφος⁷ είναι εξοικειωμένος με τα υπό διερεύνηση χαρακτηριστικά, και δημιουργεί το χάρτη με σκοπό την παρουσίαση του σε ένα ευρύτερο κοινό. Άλλες φορές ο χάρτης δημιουργείται υπό το πρίσμα της εύρεσης νέων αγνώστων μοτίβων, τα οποία ενδεχομένως να οδηγήσουν σε νέες έρευνες ή πληροφορίες/συμπεράσματα τα οποία δεν είχαν έρθει στην επιφάνεια μέχρι τη δεδομένη χρονική στιγμή. Θα πρέπει να σημειωθεί ωστόσο πως δεν είναι όλοι οι θεματικοί χάρτες επιτυχημένοι, και χρειάζεται αρκετή τριβή και γνώση προκειμένου να χρησιμοποιηθούν τα σωστά attributes ενός dataset έτσι ώστε το αποτέλεσμα να είναι χρήσιμο. Σε κάθε περίπτωση, μπορεί να θεωρηθεί ότι οι θεματικοί χάρτες έχουν 3 ειδών στόχους (Slocum, McMaster, Kessler, & Howard, 2009):

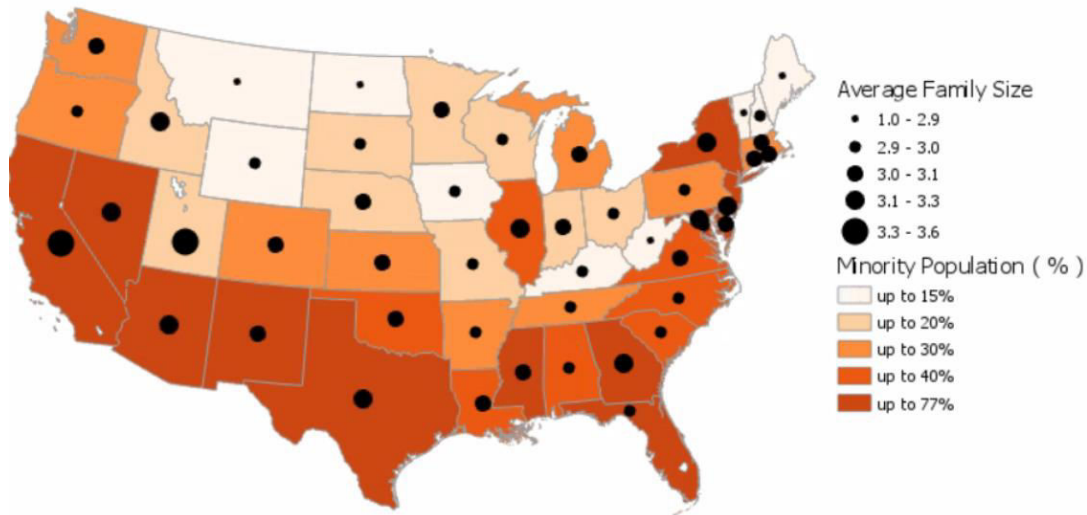
- Παροχή εξειδικευμένης πληροφορίας για κάποια συγκεκριμένη τοποθεσία. *Παράδειγμα: «Ποιο είναι το ποσοστό αλλοδαπών στην Αθήνα;»*
- Παροχή γενικευμένης πληροφορίας για εύρεση γεωγραφικών μοτίβων. *Παράδειγμα: «Σε ποιες περιοχές καλλιεργείται το καλαμπόκι;»*
- Σύγκριση σχημάτων και χαρτών ίδιου είδους με σκοπό να απεικονιστεί η μεταβολή κάποιου στοιχείου, στο πέρασμα του χρόνου. *Παράδειγμα: «Ποια είναι η διαφορά σε ψήφους για το πολιτικό κόμμα Α, στις εκλογικές περιόδους 2008 και 2012;»*

Οι χαρτογράφοι χρησιμοποιούν διάφορες δοκιμασμένες μεθόδους για την αποτελεσματική δημιουργία θεματικών χαρτών. Συχνά γίνεται αναφορά για διαφορετικά είδη θεματικών χαρτών, αλλά αυτό που ουσιαστικά αλλάζει είναι η μέθοδος κατασκευής του χάρτη και όχι το είδος του. Σε ορισμένες περιπτώσεις πιο πολύπλοκης ανάλυσης, χρησιμοποιούνται και συνδυάζονται κατάλληλα δύο ή περισσότερες μέθοδοι, παράγοντας χάρτες δύο ή περισσότερων μεταβλητών (bivariate/multivariate thematic maps).

⁶ Διανοούμενο/Φιλοσοφικό κίνημα (1715-1789) το οποίο άλλαξε ριζικά τις ιδέες και τον τρόπο σκέψης στην Ευρώπη και κατ' επέκταση στον υπόλοιπο κόσμο.

⁷ Επιστήμονας από τον κλάδο της Γεωγραφίας που ασχολείται με την κατασκευή χαρτών.

Minorities and Family Size*

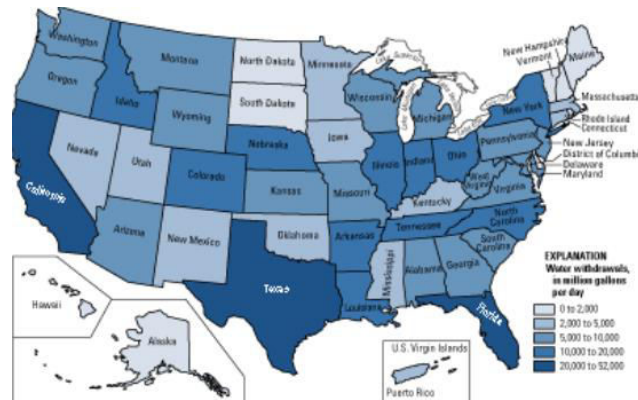


Εικόνα 14 - Χαρακτηριστικό παράδειγμα θεματικού χάρτη 2 μεταβλητών (Choropleth & Proportional point symbol)

Τεχνικές δημιουργίας θεματικών χαρτών (είδη)

Παρακάτω θα παρουσιαστούν ταχέως οι σημαντικότερες τεχνικές δημιουργίας θεματικών χαρτών, συνοδευόμενες και από αντιπροσωπευτικές εικόνες προς καλύτερη κατανόηση:

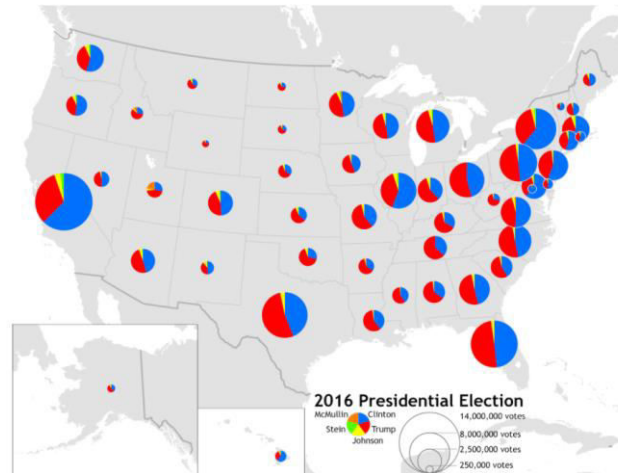
- **Choropleth:** Οι χωροπληθείς χάρτες δείχνουν στατιστικά δεδομένα σε προκαθορισμένες γεωγραφικές περιοχές (χώρες, πολιτείες), χρωματίζοντας τες. Αποτελεί ίσως την πιο διαδεδομένη τεχνική θεματικού χάρτη.



Εικόνα 15 - Χωροπληθείς θεματικός χάρτης

- **Proportional point symbol:** Η συγκεκριμένη μέθοδος χρησιμοποιεί συμβολικά σχήματα πάνω στο χάρτη (συνήθως κύκλους) διαφορετικών διαστάσεων. Η διαφορά στις διαστάσεις χρησιμοποιείται για να αναδείξει και τυχόν ποσοτικές διαφορές που

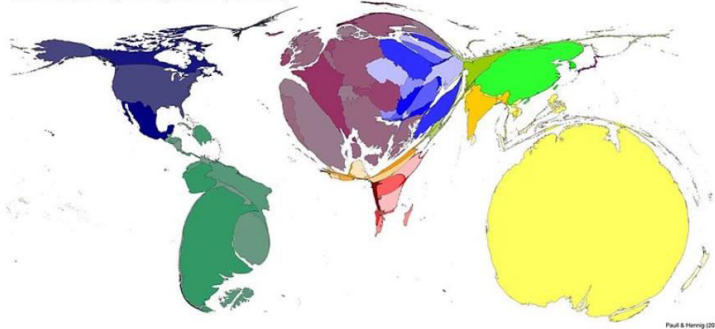
υπάρχουν στα χαρακτηριστικά που απεικονίζει ο χάρτης. Στο παρακάτω παράδειγμα, το μέγεθος ενός κύκλου αντικατοπτρίζει και τον συνολικό αριθμό ψήφων ανά πολιτεία.



Εικόνα 16 - Θεματικός χάρτης που χρησιμοποιεί τη μέθοδο αναλογικών σημείων

- **Cartogram:** Το χαρτόγραμμα είναι ένας χάρτης ο οποίος εσκεμμένα παραμορφώνει τον γεωγραφικό χώρο και τις γεωγραφικές οντότητες, με σκοπό να αναδείξει την ποσοτική διαφορά που υπάρχει σε ένα συγκεκριμένο χαρακτηριστικό. Ο πιο συνηθισμένος τρόπος ανάπτυξης ενός χαρτογράμματος είναι η αυξομείωση των διαστάσεων των οντοτήτων, η οποία δείχνει ευδιάκριτα τις μεγαλύτερες παρατηρήσεις του συνόλου δεδομένων.

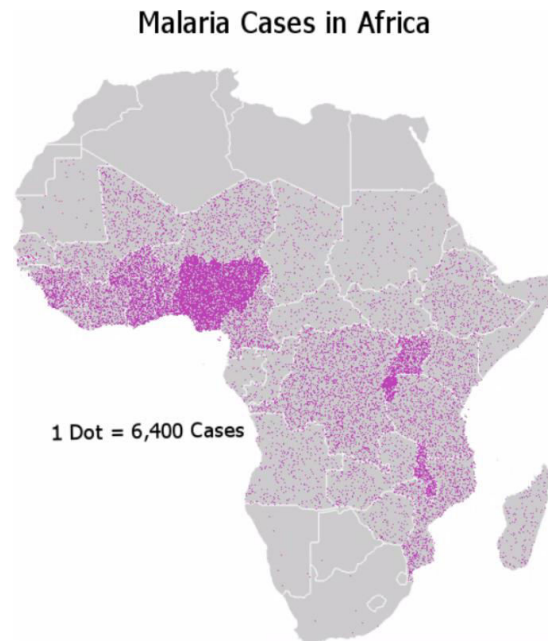
World Map of Organic Agriculture (hectares)



Εικόνα 17 - Χαρτόγραμμα που δείχνει την βιολογική γεωργία σε στρέμματα

- **Dot:** Ένας χάρτης κατανομής σημείων (dot distribution map) τοποθετεί μικρά σημεία πάνω στην επιφάνεια για να δώσει έμφαση στην γεωγραφική κατανομή ενός φαινομένου υπό έρευνα. Τέτοιοι χάρτες δημιουργούνται πολλές φορές σε ιατρικές ή δημογραφικές έρευνες, για να αναδειχθεί η κατανομή των θανάτων σε μια περιοχή, η έξαρση και ο τρόπος διασποράς ενός ιού, η αναλογία ανθρώπων διαφορετικών

εθνικοτήτων κλπ. Μέσω υπομνήματος αναγράφεται και η έννοια της μιας κουκίδας πάνω στο χάρτη.



Εικόνα 18 - Χάρτης κατανομής σημείων που δείχνει τις περιπτώσεις ελονοσίας στην Αφρική. Μια κουκίδα ισούται με 6400 κρούσματα.

Υλοποίηση Project - Δημιουργία θεματικού χάρτη από πρωτογενή δεδομένα

Σκοπός

Στο παρόν project θα υλοποιηθεί χωροπληθής θεματικός χάρτης από πρωτογενή δεδομένα, με σκοπό την οπτικοποίηση πληροφορίας και την εξαγωγή συμπερασμάτων βασισμένων στο παραγόμενο οπτικό αποτέλεσμα, το οποίο σαφώς θα προέρχεται από την αριθμητική και χωρική ανάλυση των δεδομένων που θα δοθούν ως input στο πρόγραμμα. Εκτός από το τελικό αποτέλεσμα που θα είναι ένας άρτιος θεματικός χάρτης, θα περιγραφεί και ολόκληρη η διαδικασία της έρευνας, καθώς όπως έχει προαναφερθεί απαιτείται περισσότερη δουλειά από τις 2 τελικές εντολές που δημιουργούν τον χάρτη, προκειμένου το αποτέλεσμα να είναι αξιοποιήσιμο και χρήσιμο. Πιο συγκεκριμένα θα χρησιμοποιηθούν τα παρακάτω δεδομένα, τα οποία θα αναλυθούν και περαιτέρω σε ξεχωριστή ενότητα:

1. Shapefile⁸ της χώρας του Nepal, αρχείο δημόσια διαθέσιμο από την ιστοσελίδα <https://gadm.org>
2. Αρχείο CSV⁹ που περιέχει δεδομένα με τους δείκτες ευημερίας των ανθρώπων ανά γεωγραφικό διαμέρισμα του Νεπάλ. Και το συγκεκριμένο dataset είναι δημόσια διαθέσιμο από την ιστοσελίδα [opennepal](http://data.opennepal.net/content/human-poverty-index-value-districts-2011).

⁸ Download URL: https://gadm.org/download_country_v3.html

⁹ Download URL: <http://data.opennepal.net/content/human-poverty-index-value-districts-2011>

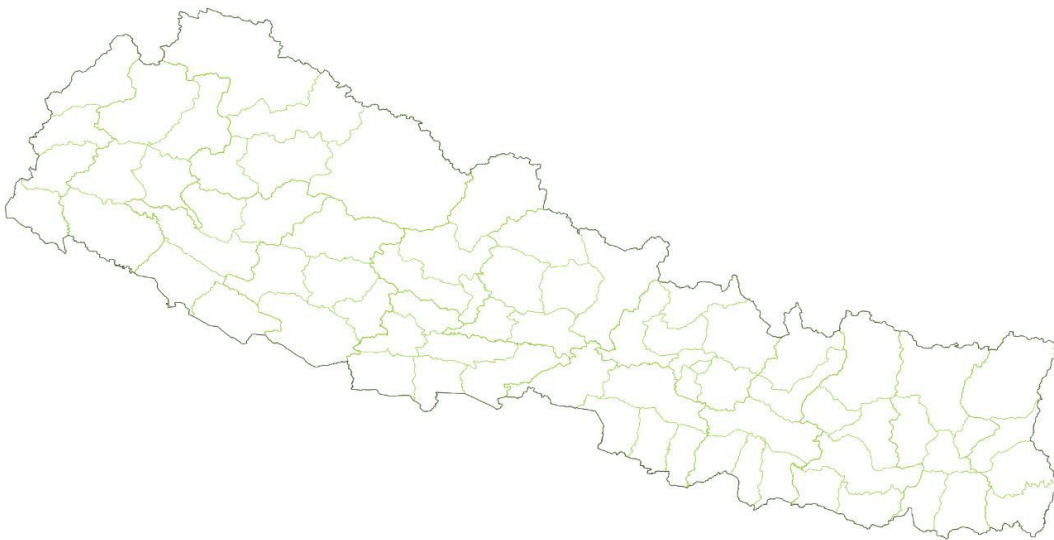
Περιγραφική ανάλυση πρωτογενών δεδομένων

Το Shapefile, είναι ένα απλό μη τοπολογικό format χωρικών δεδομένων, για την αποθήκευση γεωμετρικής και περιγραφικής πληροφορίας. Οι γεωγραφικές οντότητες σε ένα shapefile αποθηκεύονται ως γραμμές ή σημεία ή πολύγωνα και οι περιγραφικές πληροφορίες μπορούν να είναι αποθηκευμένες σε πίνακες που να συνδέονται με το shapefile (kallipros, 2015). Τα αρχεία αυτά μπορούν να βρίσκονται αποθηκευμένα σε έναν απλό φάκελο του συστήματος ή/και σε ένα workspace. Επίσης μπορούν να εξαχθούν και σε άλλες δομές δεδομένων (CAD, Coverage, Geodatabase) αλλά και να αποτελέσουν πηγή για εισαγωγή δεδομένων (import) σε μια γεωβάση (geodatabase).

Χρησιμοποιώντας τον online shapefile viewer <https://mapshaper.org/> , παρατηρείται ότι το shapefile του Νεπάλ αποτελείται από τις εξής ζώνες:

1. Gadm_36_NPL_0: 1 πολύγωνο
2. Gadm_36_NPL_1: 5 πολύγωνα
3. Gadm_36_NPL_2: 14 πολύγωνα
4. Gadm_36_NPL_3: 75 πολύγωνα
5. Gadm_36_NPL_4: 3.983 πολύγωνα

Συμπεραίνεται πως όσο αυξάνεται ο αριθμός της εκάστοτε ζώνης, τόσα περισσότερα πολύγωνα προστίθενται στον χάρτη, συνεπώς και περισσότερη λεπτομέρεια. Στο παρόν project η χωρική ανάλυση θα πραγματοποιηθεί χρησιμοποιώντας την τέταρτη ζώνη (Gadm_36_NPL_3). Το shapefile στο συγκεκριμένο επίπεδο αυτής της ζώνης θα έχει την παρακάτω μορφή:



Εικόνα 19 - Νεπάλ shapefile 75 συνολικών πολυγώνων, χωρίζοντας ουσιαστικά τη χώρα σε 75 γεωγραφικά διαμερίσματα

Το δεύτερο σύνολο δεδομένων που θα χρησιμοποιηθεί ως input για τη χωρική ανάλυση και την δημιουργία του θεματικού χάρτη, είναι σε μορφή CSV. Το συγκεκριμένο format (Comma Separated Values) είναι ευρέως διαδεδομένο για αποθήκευση δεδομένων, και πρόκειται για αρχείο κειμένου το οποίο χρησιμοποιεί το κόμμα (ή κάποιο άλλο διαχωριστικό όπως space, tab) για να ξεχωρίσει τις στήλες μεταξύ τους. Κάθε γραμμή του μπορεί να θεωρηθεί ως μια ξεχωριστή εγγραφή, με την πρώτη γραμμή του αρχείου να περιέχει συνήθως τα ονόματα των στηλών. Τα CSV αρχεία συνεπώς μπορούν εύκολα να φορτωθούν σε κατάλληλες γλώσσες προγραμματισμού (εν προκειμένω R) και να μετατραπούν σε διαχειρίσιμα dataframes. Τα δεδομένα έχουν την παρακάτω μορφή:

District ▲	Zone ▼	Geographical Region	Development Region	Sub Group ▼	Value ▼
Achham	Seti	Hill	Far-Western	Percentage of People not expected to survive age 40	9.5
Achham	Seti	Hill	Far-Western	Adult illiteracy rate	54.92
Achham	Seti	Hill	Far-Western	Percentage without safe water	44.31
Achham	Seti	Hill	Far-Western	Percentage of children under age five who are malnourished	59.2
Achham	Seti	Hill	Far-Western	Deprivation in economic provisioning	51.76
Achham	Seti	Hill	Far-Western	HPI	46.68
Arghakhanchi	Lumbini	Hill	Western	Percentage of People not expected to survive age 40	7.71
Arghakhanchi	Lumbini	Hill	Western	Adult illiteracy rate	34.43
Arghakhanchi	Lumbini	Hill	Western	Percentage without safe water	23.48
Arghakhanchi	Lumbini	Hill	Western	Percentage of children under age five who are malnourished	31
Arghakhanchi	Lumbini	Hill	Western	Deprivation in economic provisioning	27.24
Arghakhanchi	Lumbini	Hill	Western	HPI	27.37
Baglung	Dhawalagiri	Hill	Western	Percentage of People not expected to survive age 40	8.52
Baglung	Dhawalagiri	Hill	Western	Adult illiteracy rate	34.46

Εικόνα 20 - Μορφή δεδομένων CSV αρχείου με δείκτες ευημερίας του Νεπάλ.

Με κόκκινο έχουν σημειωθεί τα 2 σημεία κλειδιά βάσει των οποίων θα γίνει η χωρική ανάλυση. Πιο συγκεκριμένα, παρατηρούμε ότι για κάθε γεωγραφικό διαμέρισμα (district) υπάρχουν 6 καταχωρήσεις, αλλά θα χρησιμοποιηθούν μόνο οι παρατηρήσεις που αφορούν τον δείκτη (Sub Group) HPI. Το αρχείο έχει συνολικά **450 εγγραφές**, αλλά βάσει του προαναφερθέντος συλλογισμού θα χρησιμοποιηθούν οι **75** (450/6) εξ' αυτών, όσα είναι δηλαδή και τα γεωγραφικά διαμερίσματα του Νεπάλ. Το dataset συνολικά έχει **6 στήλες** (District, Zone, Geographical Region, Development Region, Sub Group, Value), αλλά ουσιαστικά θα χρησιμοποιηθούν οι **3**.

Δημιουργία θεματικού χάρτη σε βήματα

Ξεκινάμε φορτώνοντας τις απαραίτητες βιβλιοθήκες, και εν συνεχεία το κατάλληλο shapefile (της τέταρτης ζώνης όπως προαναφέρθηκε), και δημιουργούμε κάποια αρχικά plots με την βιβλιοθήκη ggplot για να δούμε τι σχήμα έχει ο χάρτης. Θα χρησιμοποιήσουμε 2 εντολές (path & polygon) της ggplot για να δούμε με 2 διαφορετικούς τρόπους τον χάρτη:

```

library(rgdal)
library(ggplot2)
library(dplyr)

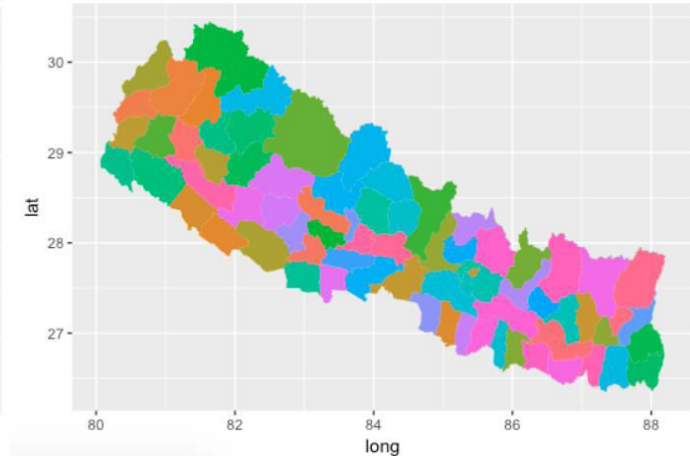
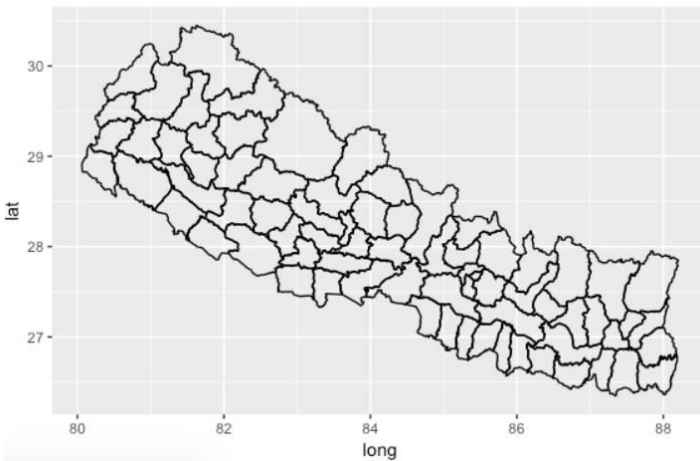
nepal_shp <-
  readOGR(dsn="data/gadm36_NPL_shp", layer="gadm36_NPL_3", stringsAsFactors =
  FALSE)
shp_df <- fortify(nepal_shp, region = "NAME_3")

map <- ggplot(data = shp_df, aes(x = long, y = lat, group = group))

map + geom_path()

map +
  geom_polygon(aes(fill = id)) +
  coord_fixed(1.3) +
  guides(fill = FALSE)

```



Εικόνα 21 - `geom_path()` (αριστερά) και `geom_polygon()` (δεξιά)

Αφού φορτώσαμε τον χάρτη του Nepal και τον μετατρέψαμε σε ένα διαχειρίσιμο dataframe (μέσω της εντολής `fortify`¹⁰), το οποίο θα δοθεί ως input στην `ggplot`, θα φορτώσουμε και τα Nepal HPI δεδομένα μας τα οποία είναι αποθηκευμένα στο CSV αρχείο. Έπειτα, θα τα καθαρίσουμε και θα κρατήσουμε μόνο τις στήλες που μας ενδιαφέρουν.

```

# load initial csv data file
nepal_HPI_data <-
  read.csv("data/HPI_Nepal_2011.csv", stringsAsFactors = FALSE)

```

¹⁰ Περισσότερες λεπτομέρειες για το πως λειτουργεί το `fortify` <https://github.com/tidyverse/ggplot2/wiki/plotting-polygon-shapefiles#preparation>


```
# clean data and filter out needless data
hpi.data <- nepal_HPI_data %>%
  filter(Sub.Group == "HPI") %>%
  select(id = District, HPI = Value)
```

hpi.data

Με τις παραπάνω εντολές φορτώθηκε το csv αρχείο ως μεταβλητή (Nepal_HPI_data) στη μνήμη του προγράμματος, και καθαρίστηκε κρατώντας μόνο τις παρατηρήσεις οι οποίες έχουν Sub Group = HPI. Στο τέλος δημιουργήθηκαν 2 στήλες στο νέο dataframe (hpi.data) οποίες είναι το όνομα του γεωγραφικού διαμερίσματος (id) και ο δείκτης ευημερίας (HPI). Το παραγόμενο dataframe θα έχει την εξής μορφή:

	id	HPI
1	Achham	46.68
2	Arghakhanchi	27.37
3	Banke	32.10
4	Baglung	27.33
5	Baitadi	39.58
6	Bajhang	45.32
7	Bajura	43.28
8	Bara	40.09
9	Bardiya	32.30
10	Bhaktapur	19.43

Εικόνα 22 - Παραγόμενο dataframe 2 στηλών και 75 γραμμών (hpi.data)

Συνεχίζοντας, θα χρησιμοποιήσουμε 2 εντολές μέσω των οποίων θα εντοπιστούν τυχόν διαφορές ανάμεσα στα δυο σύνολα δεδομένων (hpi.data & shp_df) που έχουμε μέχρι στιγμής στην μνήμη. Πιο συγκεκριμένα, με την πρώτη εντολή θα βρούμε όλα τα γεωγραφικά διαμερίσματα τα οποία δεν εμπεριέχονται στο shapefile, και με τη δεύτερη εντολή θα βρούμε το αντίστροφο. Θεωρητικά, οι 2 αυτές εντολές θα πρέπει να παράγουν ίδιο αριθμό αποτελεσμάτων, και στην πράξη παρατηρούμε πως κάτι τέτοιο συμβαίνει.

```
# identify the mismatched districts
# first line finds all ids from csv file that are not included in shapefile
unique(hpi.data$id[!hpi.data$id %in% unique(shp_df$id)])
# second line finds all ids from shp dataframe that do not exist in csv
unique(unique(shp_df$id)[!unique(shp_df$id) %in% hpi.data$id])
```

Οι 2 εντολές παράγουν τα εξής αποτελέσματα:

```

> # identify the mismatched districts
> # first line finds all ids from csv file that are not included in shapefile
> unique(hpi.data$id[!hpi.data$id %in% unique(shp_df$id)])
[1] "Darchaula" "Kavre"
> # second line finds all ids from shp dataframe that do not exist in csv
> unique(unique(shp_df$id)[!unique(shp_df$id) %in% hpi.data$id])
[1] "Darchula" "Kavrepalanchok"

```

Παρατηρούμε ότι 2 γεωγραφικά διαμερίσματα είναι καταγεγραμμένα με διαφορετικά ονόματα στις δυο μεταβλητές μας, οπότε θα πρέπει αυτή η διαφορά να εξαλειφθεί, και ακολούθως να γίνει συγχώνευση των 2 dataframes σε ένα τελικό dataframe, βάσει του οποίου θα φτιαχτεί και ο χωροπληθής θεματικός χάρτης. Η συγχώνευση θα γίνει με left join¹¹.

```

# fix the mismatched districts
hpi.data$id[hpi.data$id == "Darchaula"] <- "Darchula"
hpi.data$id[hpi.data$id == "Kavre"] <- "Kavrepalanchok"

# merge hpi data with shapefile dataframe
shp_df <- shp_df %>%
  left_join(hpi.data, by="id")

```

Η διαδικασία της συγχώνευσης και η μορφή του τελικού πλαισίου δεδομένων απεικονίζεται με ακρίβεια στην ακόλουθη εικόνα:

nepal.adm3.shp.df						
id	long	lat	order	hole	piece	group
Achham	81.17178	29.38659	1	FALSE	1	Achham.1
Achham	81.17886	29.38498	2	FALSE	1	Achham.1
Achham	81.18110	29.38050	3	FALSE	1	Achham.1
Achham	81.18681	29.37690	4	FALSE	1	Achham.1
Achham	81.18793	29.37479	5	FALSE	1	Achham.1
Achham	81.19377	29.37168	6	FALSE	1	Achham.1

hpi.data	
id	HPI
Achham	46.68
Arghakhanchi	27.37
Baglung	27.33
Baitadi	39.58

merge(nepal.adm3.shp.df, hpi.data, by="id")

id	long	lat	order	hole	piece	group	HPI
Achham	81.17178	29.38659	1	FALSE	1	Achham.1	46.68
Achham	81.17886	29.38498	2	FALSE	1	Achham.1	46.68
Achham	81.18110	29.38050	3	FALSE	1	Achham.1	46.68
Achham	81.18681	29.37690	4	FALSE	1	Achham.1	46.68
Achham	81.18793	29.37479	5	FALSE	1	Achham.1	46.68

Εικόνα 23 - shp_df μεταβλητή μετά την συγχώνευση

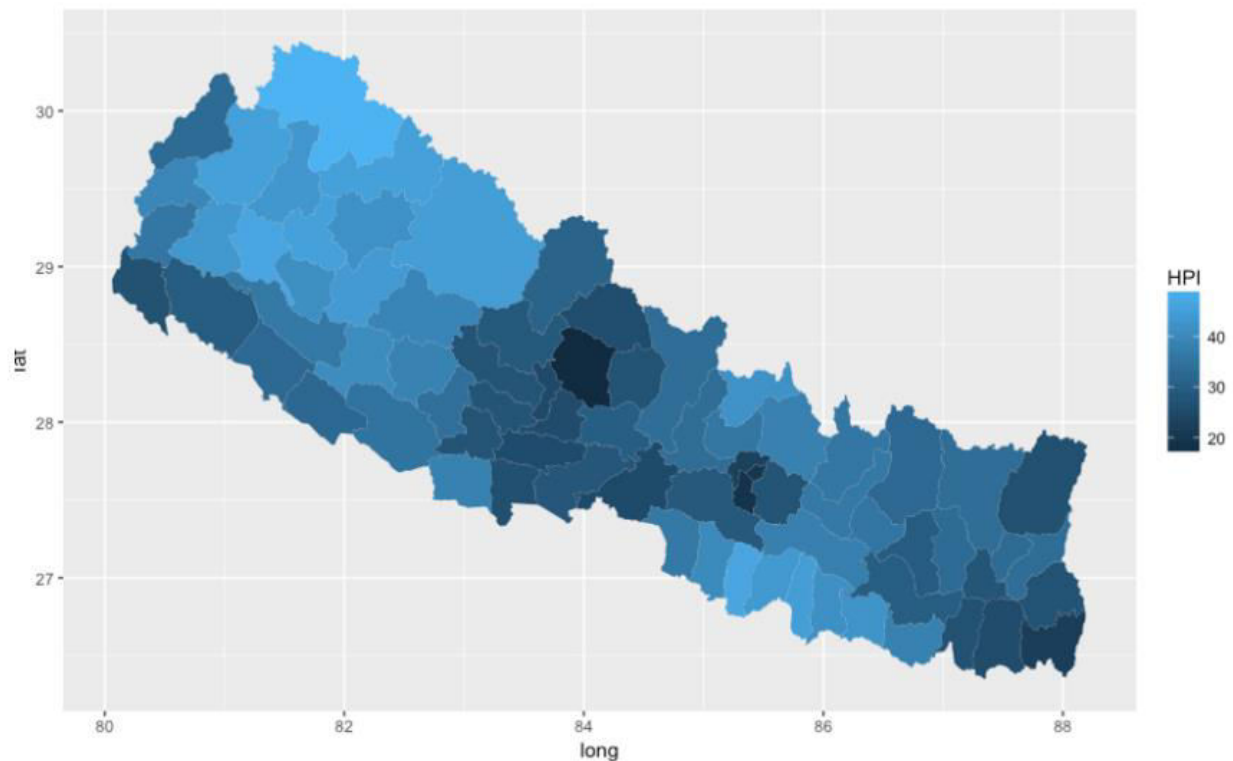
Έχοντας πλέον το τελικό dataframe, θα ξεκινήσει η διαδικασία δημιουργίας του θεματικού χάρτη, προοδευτικά, χρησιμοποιώντας τη μέθοδο trial and error και προσθέτοντας σε κάθε βήμα και μια επιπλέον βελτίωση. Επιλέχθηκε αυτός ο τρόπος παρουσίασης του project, διότι

¹¹ Λεπτομέρειες για το πώς λειτουργεί το left_join https://www.w3schools.com/sql/sql_join_left.asp

είναι πιο κοντά στην πραγματική δουλειά ενός data scientist, όπου δεν γίνεται να παραχθεί εξ'αρχής ένα άρτιο αποτέλεσμα χωρίς να ξεκινήσει η έρευνα από κάτι μικρό το οποίο θα βελτιώνεται συνεχώς.

```
map <- ggplot(data = shp_df, aes(x = long, y = lat, group = group))  
  
map +  
  geom_polygon(aes(fill = HPI), color = 'gray', size = 0.1) +  
  coord_fixed(1.3)
```

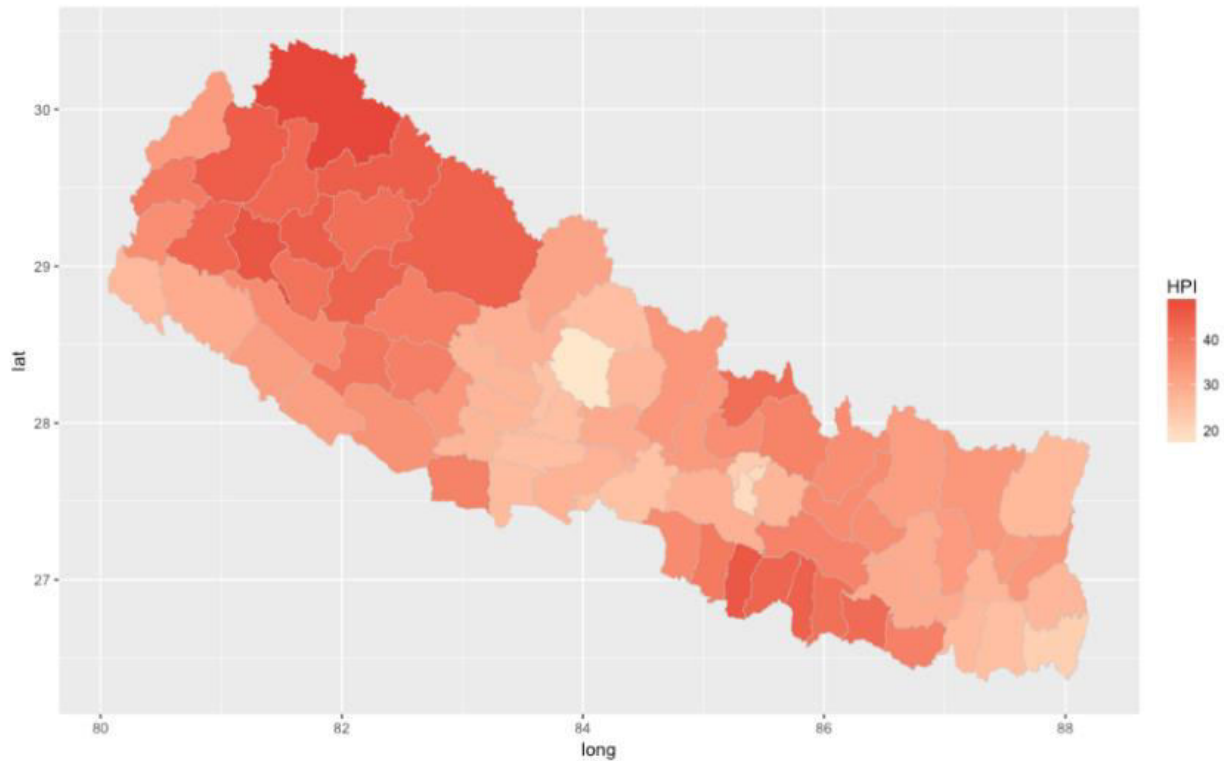
Οι παραπάνω γραμμές παράγουν τον χάρτη:



Εικόνα 24 - Αρχικός χάρτης χωρίς επιπλέον παραμέτρους

Η πρώτη βελτίωση που θα κάνουμε θα είναι να βελτιώσουμε τα χρώματα, χρησιμοποιώντας σκούρα χρώματα για τις υψηλές τιμές και ανοιχτά για τις χαμηλές τιμές, με την εντολή `scale_fill_gradient`

```
map +  
  geom_polygon(aes(fill = HPI), color = 'gray', size = 0.1) +  
  scale_fill_gradient(high = "#e34a33", low = "#fee8c8", guide = "colorbar")  
+  
  coord_fixed(1.3)
```



Εικόνα 25 - Βελτίωση στα χρώματα

Επιπλέον βελτιωτικές ενέργειες, προκειμένου να προστεθεί τίτλος στον χάρτη και να μεταφερθεί το υπόμνημα εντός της εικόνας:

```
map +
  geom_polygon(aes(fill = HPI), color = 'gray', size = 0.1) +
  scale_fill_gradient(high = "#e34a33", low = "#fee8c8", guide = "colorbar")
+
  coord_fixed(1.3) +
  guides(fill=guide_colorbar(title="HP Index")) +
  theme(legend.justification=c(0,0), legend.position=c(0,0))
```

Στη συνέχεια, θα δημιουργήσουμε ένα νέο theme για τον χάρτη, στο οποίο θα απουσιάζει το γκρι background, για πιο ευπαρουσίαστο οπτικό αποτέλεσμα:

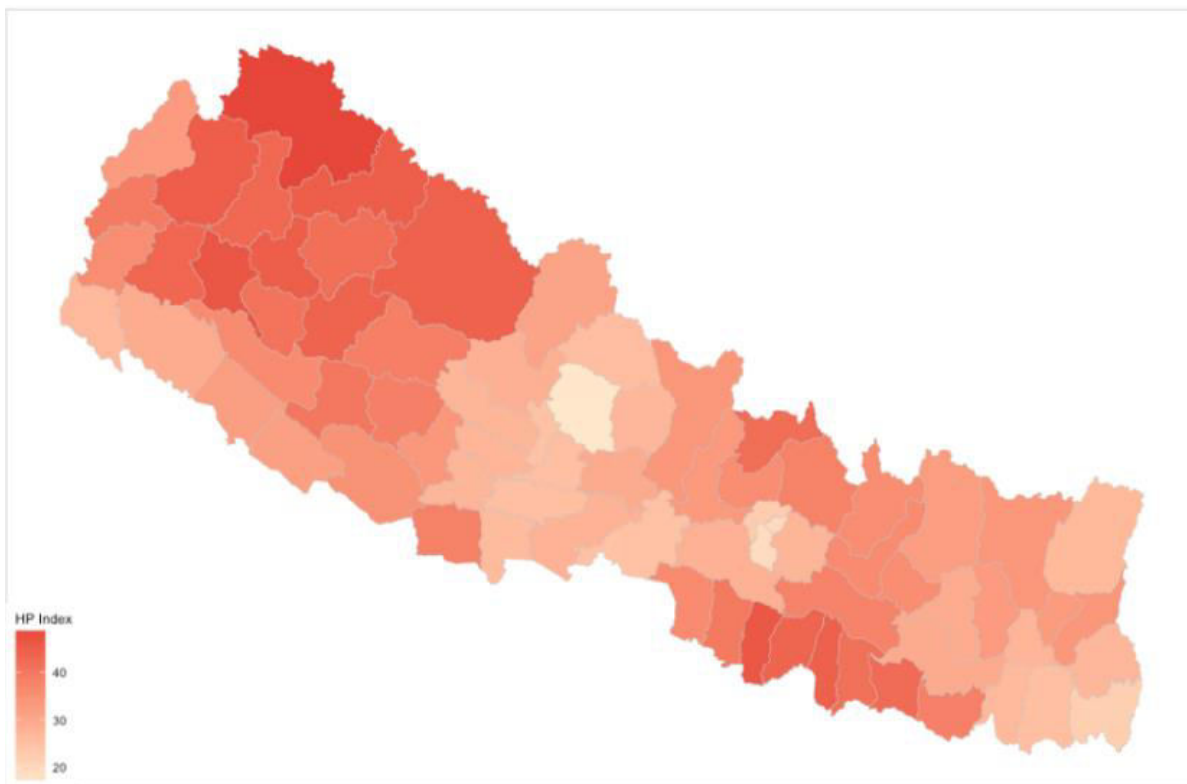
```
theme_bare <- theme(
  axis.line = element_blank(),
  axis.text.x = element_blank(),
  axis.text.y = element_blank(),
  axis.ticks = element_blank(),
  axis.title.x = element_blank(),
  axis.title.y = element_blank(),
  legend.text=element_text(size=7),
```

```

legend.title=element_text(size=8),
panel.background = element_blank(),
panel.border = element_rect(colour = "gray", fill=NA, size=0.5)
)
map +
  geom_polygon(aes(fill = HPI), color = 'gray', size = 0.1) +
  guides(fill=guide_colorbar(title="HP Index")) +
  scale_fill_gradient(high = "#e34a33", low = "#fee8c8", guide = "colorbar")
+
  coord_fixed(1.3) +
  theme(legend.justification=c(0,0), legend.position=c(0,0)) +
  theme_bare

```

Μέχρι στιγμής έχουμε φτάσει στο αρκετά καλό παρακάτω αποτέλεσμα:

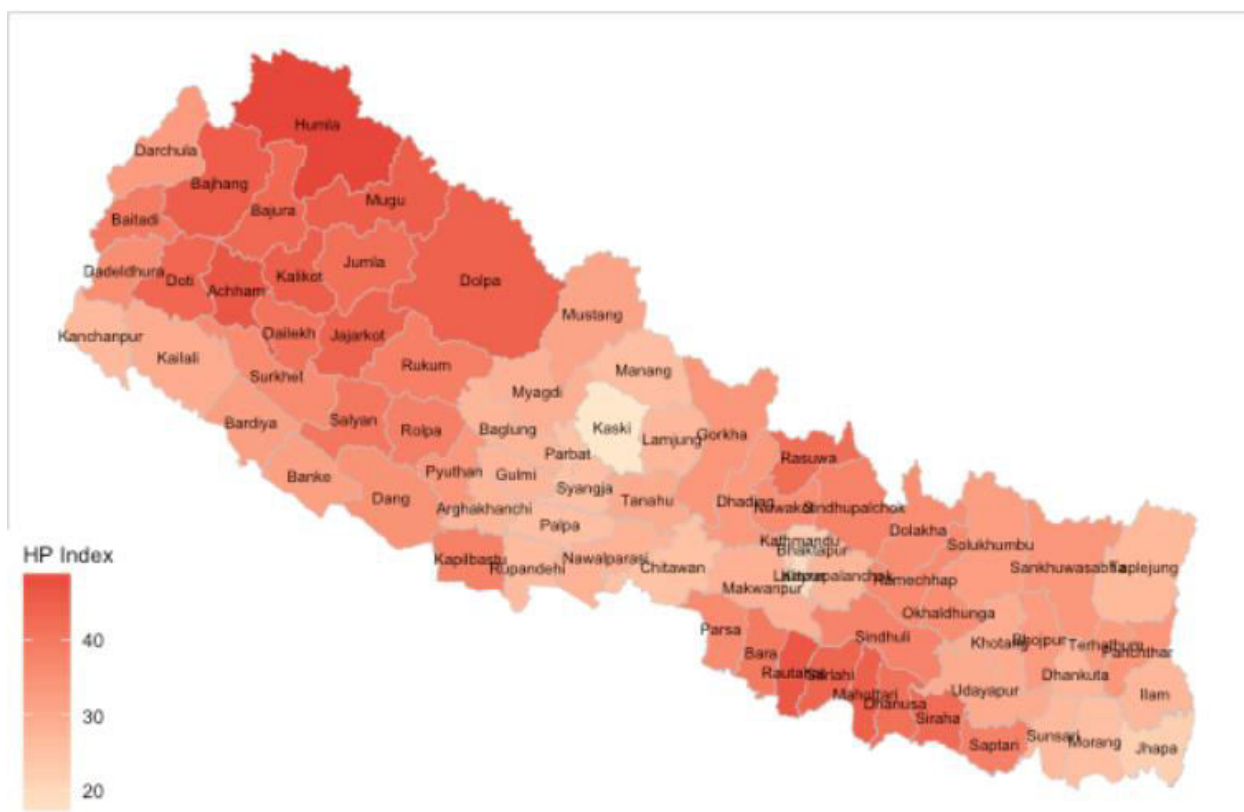


Εικόνα 26 - Επιπλέον βελτιωτικές ενέργειες στον παραγόμενο θεματικό χάρτη

Όμως, για να μπορεί κάποιος παρατηρητής ο οποίος δε γνωρίζει γεωγραφία να δει ποια γεωγραφικά διαμερίσματα έχουν υψηλό ΗΠΙ, θα πρέπει να προστεθούν και οι κατάλληλες ταμπέλες (labels) σε κάθε πολύγωνο, με το όνομα του εκάστοτε γεωγραφικού διαμερίσματος.

Κάτι τέτοιο σαφώς δεν πρόκειται για τόσο απλή διαδικασία, καθώς θα πρέπει να υπολογιστεί το κεντροειδές¹² (centroid) για κάθε district.

```
# finding the centers of all the district polygons.
centroids <-
  rgeos::gCentroid(nepal_shp, byid = TRUE, id = unique(nepal_shp$NAME_3))
centroids_df <- centroids %>%
  as.data.frame() %>%
  mutate(label = row.names(.))
```



Εικόνα 27 - Προσθήκη ονομάτων στα γεωγραφικά διαμερίσματα

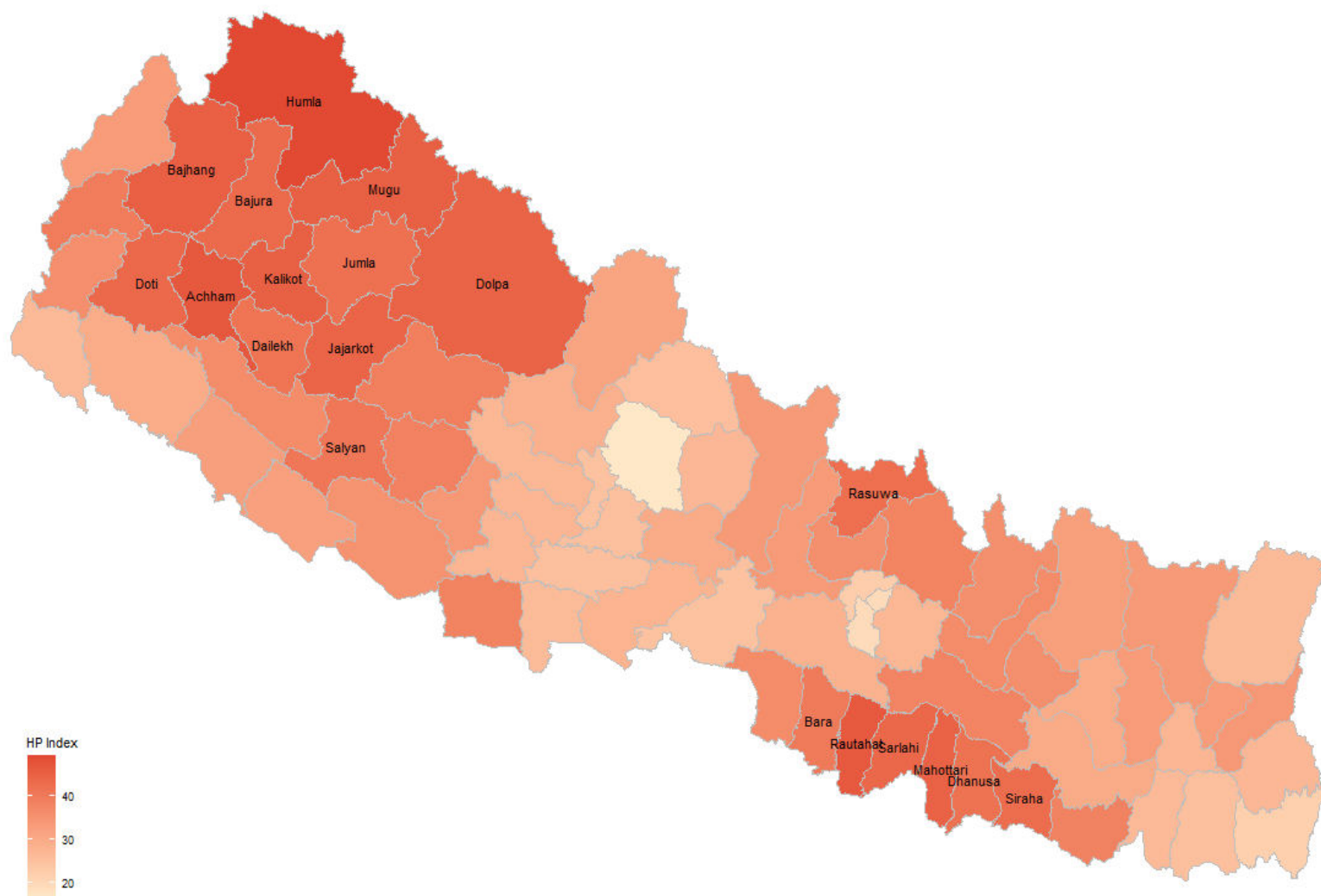
Η δημιουργία του θεματικού χάρτη έχει φτάσει σχεδόν στο τέλος, όμως με την προσθήκη όλων των ονομάτων δημιουργείται μια μικρή σύγχυση και ο χάρτης δε διαβάζεται εύκολα. Η τελευταία βελτιωτική ενέργεια που θα γίνει, θα είναι να εμφανίζονται μόνο τα ονόματα των γεωγραφικών διαμερισμάτων τα οποία έχουν HPI τιμή μεγαλύτερη του 40:

```
# showing district names for which HPI>40
centroids.selected <-
  centroids_df[centroids_df$label %in% (hpi.data[hpi.data$HPI>40,]$id),]
```

¹² Περισσότερες λεπτομέρειες για τον υπολογισμό centroids και την προσθήκη labels μέσω ggplot <https://stackoverflow.com/questions/28962453/how-can-i-add-labels-to-a-choropleth-map-created-using-ggplot2>

Ο χωροπληθής θεματικός χάρτης έχει φτάσει στο τέλος του και έχει την παρακάτω μορφή:

Nepal Human Poverty Index Map



Εικόνα 28 - Χωροπληθής θεματικός χάρτης του Νεπάλ, που απεικονίζει ονομαστικά τα γεωγραφικά διαμερίσματα των οποίων ο ΗΠΙ ξεπερνάει την τιμή 40.

Συνοπτικά, επισυνάπτεται ολόκληρος ο κώδικας R χωρίς τις περιττές (ενδιάμεσες) εντολές, ο οποίος παράγει το παραπάνω αποτέλεσμα μετά την εκτέλεση του. Προαπαιτούμενο σαφώς για να τρέξει είναι να υπάρχει στον ίδιο φάκελο με το directory “data”, στο οποίο υπάρχουν τα πρωτογενή δεδομένα του Νεπάλ.



nepal_HPI_Thematic
_Map.r

Επιπλέον αριθμητική ανάλυση και δημιουργία θηκογράμματος

Αναφορικά με το CSV αρχείο που περιέχει τους δείκτες HPI του Νεπάλ, πραγματοποιήθηκε περαιτέρω αριθμητική ανάλυση και δημιουργήθηκε και το αντίστοιχο θηκόγραμμα, ενέργειες που θα βοηθήσουν στην εξαγωγή συμπερασμάτων παρακάτω:

```
# load initial csv data file
nepal_HPI_data <-
  read.csv("data/HPI_Nepal_2011.csv", stringsAsFactors = FALSE)

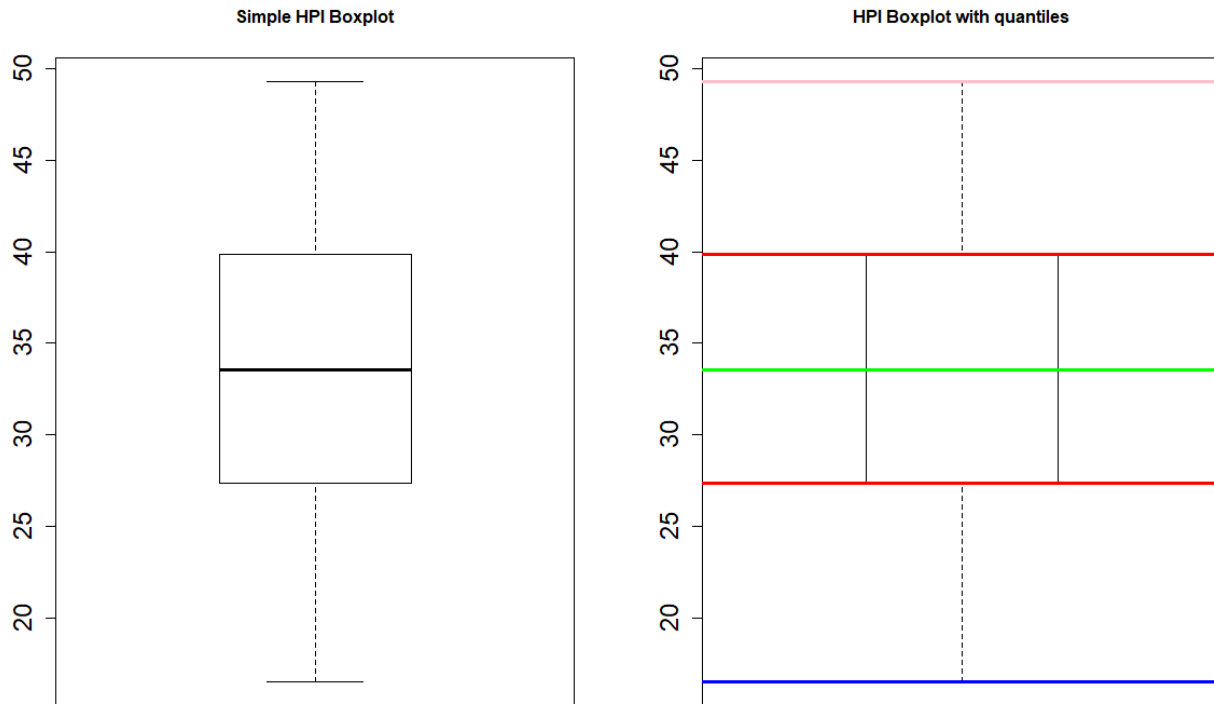
# clean data and filter out needless data
hpi.data <- nepal_HPI_data %>%
  filter(Sub.Group == "HPI") %>%
  select(HPI = Value)

# View quantiles and create boxplots
summary(hpi.data$HPI)
par(mfrow = c(1, 2), cex.main=1)
boxplot(hpi.data$HPI, cex.axis= 1.5)
title(main = "Simple HPI Boxplot")
boxplot(hpi.data$HPI, cex.axis= 1.5)
title(main = "HPI Boxplot with quantiles")
abline(h = min(hpi.data$HPI), col = "Blue", lwd=3)
abline(h = max(hpi.data$HPI), col = "Pink", lwd=3)
abline(h = median(hpi.data$HPI), col = "Green", lwd=3)
abline(h = quantile(hpi.data$HPI, c(0.25, 0.75)), col = "Red", lwd=3)
```

Η συγκεκριμένη υλοποίηση δεν κρύβει κάποια πολύπλοκη λογική, αρκεί η μετατροπή των τιμών HPI σε ένα vector και η χρήση των κατάλληλων εντολών για την δημιουργία των boxplots. Η εντολή `summary(hpi.data$HPI)` δίνει το παρακάτω output:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
16.50	27.36	33.58	33.80	39.84	49.26

Και τα παραγόμενα boxplots:



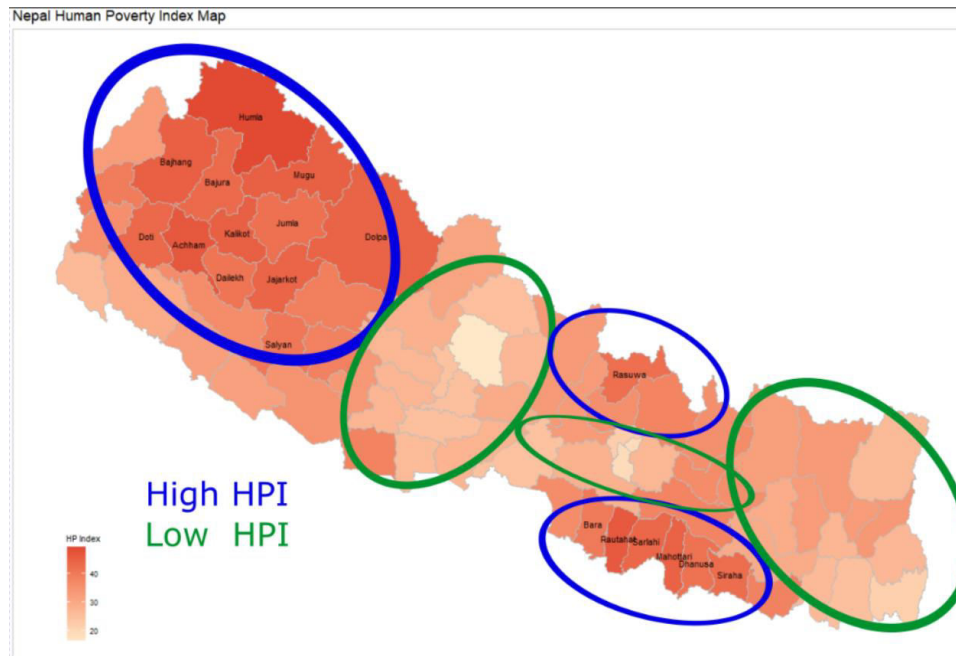
Εικόνα 29 - Nepal HPI Boxplots

Η ανάλυση τους γίνεται στην ενότητα [Συμπεράσματα](#).

Συμπεράσματα

Το HPI (Human Poverty Index) είναι μια ενδεικτική τιμή φτώχειας (ή αλλιώς, κακών συνθηκών επιβίωσης). Δημιουργήθηκε από τα Ενωμένα Έθνη ως συμπληρωματικό του HDI (Human Development Index) το 1997, καθώς απεικονίζει με μεγαλύτερη ακρίβεια τις περιπτώσεις όπου οι συνθήκες στις οποίες ζει μια πόλη/κοινότητα/χώρα είναι χαμηλές (Wikipedia, 2021). Ο δείκτης αυτός υπολογίζεται λαμβάνοντας υπόψιν ορισμένους παράγοντες, όπως είναι η πιθανότητα να ξεπεράσει κάποιος το προσδόκιμο ζωής των 60 ετών, βασική μόρφωση και εκπαίδευση, μακροχρόνια ανεργία και ποσοστό πληθυσμού που αμείβεται με χαμηλότερο μισθό από τον μέσο όρο. Ενδεικτικά, οι κορυφαίες χώρες του κόσμου έχουν HPI που κινείται στα επίπεδα 6-8, ενώ αντιθέτως οι χειρότερες χώρες ξεπερνάνε το 40.

Κατόπιν οπτικής ανάλυσης του παραχθέντος θεματικού χάρτη, ο οποίος φτιάχτηκε στην [αντίστοιχη ενότητα](#), σχηματίζονται οι εξής περιοχές:



Εικόνα 30 - Ομαδοποιημένες περιοχές Νεπάλ με υψηλό και χαμηλό HPI.

Παρατηρείται ότι έχει δημιουργηθεί ένα αρκετά μεγάλο cluster (περίπου 25% της συνολικής έκτασης και ~18 παρατηρήσεις) με γειτονικά γεωγραφικά διαμερίσματα, τα οποία παρουσιάζουν ιδιαίτερα υψηλό HPI, στο βορειοδυτικό μέρος της χώρας του Νεπάλ. Εκτός αυτού, μια μικρότερη ομάδα περιοχών (6-8 παρατηρήσεις) υπάρχει νοτιοανατολικά που βρίσκεται στα ίδια επίπεδα HPI (>40), καθώς και η περιοχή “Rasuwa” λίγο πιο βόρεια η οποία μπορεί να θεωρηθεί και ως έκτροπη παρατήρηση στο σύνολο των δεδομένων μας.

Αντιθέτως, όλη η ανατολική πλευρά της χώρας όπως και το κέντρο της βρίσκεται σε καλύτερη κατάσταση, με τον HPI να έχει σαφώς χαμηλότερες τιμές στα περισσότερα districts και τις συνθήκες διαβίωσης των κατοίκων να είναι πιο ευνοϊκές.

Το άμεσο συμπέρασμα από το παραπάνω είναι ότι έχουν σχηματιστεί ισχυρές συσταδοποιήσεις, με 4 διακριτές συστάδες γεωγραφικών διαμερισμάτων, εκ των οποίων οι 2 παρουσιάζουν υψηλό HPI και οι άλλες 2 χαμηλό. Το φαινόμενο αυτό μπορεί να χαρακτηριστεί εν μέρει ως λογικό, καθώς δεν είναι σύνηθες να παρατηρείται απότομη εναλλαγή των συνθηκών διαβίωσης σε περιοχές που γειτνιάζουν.

Επιπροσθέτως, από την αριθμητική ανάλυση που πραγματοποιήθηκε στα δεδομένα, καθώς και από τα θηκογραφήματα, προκύπτει ότι ο μέσος όρος βρίσκεται στο ~33, ενώ το πρώτο τεταρτημόριο (κάτω κόκκινη γραμμή) είναι στο 27.36, που σημαίνει ότι το 75% των

παρατηρήσεων βρίσκονται πάνω από αυτήν την τιμή. Οι τιμές αυτές χαρακτηρίζονται δυσοίωνες για το Νεπάλ και το γενικότερο επίπεδο ζωής της χώρας, κάτι το οποίο επιβεβαιώνεται και από τον χάρτη, στον οποίο μόλις 2 από τα 75 διαμερίσματα (Kathmandu, Kaski, βλ. εικόνα 27) έχουν το χαρακτηριστικό ανοιχτό ερυθρό χρώμα, που υποδεικνύει σχετικά χαμηλό HPI.

Επόμενα βήματα

Σαν επόμενα βήματα, τα οποία θα μπορούσαν να υλοποιηθούν στα πλαίσια επέκτασης του παρόντος project, θα μπορούσαμε να διευρύνουμε την χωρική ανάλυση των δεδομένων ευημερίας του Νεπάλ, υπό διάφορα πρίσματα. Πιο συγκεκριμένα, θα μπορούσαν να χρησιμοποιηθούν διαφορετικές τεχνικές υλοποίησης και δημιουργίας θεματικών χαρτών, οι οποίες αναλύθηκαν στη [σχετική ενότητα](#).

Ακόμη, παρατηρώντας τη μορφή του CSV αρχείου θα μπορούσε να γίνει ενδελεχής ανάλυση και στις υπόλοιπες τιμές (Percentage of People not expected to survive age 40, Adult illiteracy rate, Percentage without safe water κλπ.) και να εξαχθεί ένα γενικότερο συμπέρασμα το οποίο να συσχετίζει αυτές τις τιμές με τον δείκτη HPI. Τέλος, μια σημαντική βελτίωση η οποία θα ανήγαγε το project σε επίπεδο διπλωματικής εργασίας, θα ήταν να φτιαχτεί κατάλληλο Graphical User Interface (GUI) μέσω του οποίου ο χρήστης να επιλέγει ποιο χαρακτηριστικό επιθυμεί να αναλυθεί από το CSV, και να παράγεται ένας διαδραστικός θεματικός χάρτης ο οποίος θα διαμορφώνεται «εν θερμώ» (on the fly) και θα δείχνει τα επιθυμητά αποτελέσματα.

Αναφορές

Ali, E. (2020). *ResearchGate*. Retrieved from

https://www.researchgate.net/publication/340182760_Geographic_Information_System_GIS_Definition_Development_Applications_Components

Dent, B. D., Torguson, J. S., & Hodler, T. W. (2009). *Cartography: Thematic Map Design*. McGraw-Hill ISBN 978-0-07-294382-5.

NationalGeographic. (n.d.). *nationalgeographic.org*. Retrieved from

<https://www.nationalgeographic.org/encyclopedia/geographic-information-system-gis>

Petchenik, B. B. (1979). *From Place to Space: The Psychological Achievement of Thematic Mapping, The American Cartographer*. DOI: 10.1559/152304079784022763.

Robinson, A. H. (1982). *Early Thematic Mapping in the History of Cartography*. . University of Chicago Press.

Slocum, T. A., McMaster, R. B., Kessler, F. C., & Howard, H. H. (2009). *Thematic Cartography and Visualization (3rd ed.)*. Pearson/Prentice Hall.

Tobler. (1970). *A Computer Movie Simulating Urban Growth in the Detroit Region*. *Economic Geography*,46, doi:10.2307/143141.

Tomlin, D. (1990). *GIS and Cartographic Modeling*. ESRI Press.

Wikipedia. (2021). *Wikipedia*. Retrieved from

https://el.wikipedia.org/wiki/%CE%A3%CF%8D%CF%83%CF%84%CE%B7%CE%BC%CE%B1_%CE%93%CE%B5%CF%89%CE%B3%CF%81%CE%B1%CF%86%CE%B9%CE%BA%CF%8E%CE%BD_%CE%A0%CE%BB%CE%B7%CF%81%CE%BF%CF%86%CE%BF%CF%81%CE%B9%CF%8E%CE%BD

Χαλκιάς, Χ. (2015). Γεωγραφική Ανάλυση με την αξιοποίηση της Γεωπληροφορικής. In Χ. Χαλκιάς, *Γεωγραφική Ανάλυση με την αξιοποίηση της Γεωπληροφορικής* (pp. 46-53). kallipos.