

1.

ΕΙΣΑΓΩΓΗ ΣΤΑ ΔΕΔΟΜΕΝΑ

1.1 Σύνοψη κεφαλαίου

ΜΕΧΡΙ ΤΟ ΤΕΛΟΣ του κεφαλαίου θα μπορείτε:

- Να περιγράψετε με τη βοήθεια ιστορικών παραδειγμάτων την επίδραση και τη διαχρονική επιρροή των δεδομένων.
- Να αποκτήσετε προσωπική άποψη για το αντικείμενο της ανάλυσης δεδομένων, κάνοντας τις δικές σας προβλέψεις και συγκρίνοντας τις με δεδομένα της πραγματικής ζωής.
- Να ορίσετε τον όρο «δεδομένα».
- Να περιγράψετε το φαινόμενο των «τεράστιων δεδομένων» με τα ενδεχόμενα πλεονεκτήματά τους και τους πιθανούς κινδύνους.

1.2 Τα δεδομένα μάς κατακλύζουν

Τα δεδομένα βρίσκονται παντού. Από τις τιμές των μετοχών στις πρωινές ειδήσεις μέχρι τις λίστες με τις θερμίδες που καίμε στα μηχανήματα γυμναστικής, όλοι μας συναντάμε καθημερινά τεράστιες ποσότητες δεδομένων, ποσοτικές ή ποιοτικές πληροφορίες για τον εαυτό μας, την κοινωνία ή το σύμπαν. Θα ήταν αδύνατο να αναλύσουμε όλα αυτά τα δεδομένα. Ωστόσο, αν αναλύσουμε κάποια από αυτά, μπορούμε να μάθουμε πολλά για τον εαυτό μας, την κοινωνία και τον κόσμο στον οποίο ζούμε.

Τα δεδομένα μάς επιτρέπουν να ανακαλύπτουμε ευρήματα τα οποία το ένστικτο ή η κοινή λογική δεν μπορούν να αποκαλύψουν. Τα δεδομένα δρομολογούν αλλαγές πολιτικής των κυβερνήσεων. Τα δεδο-

μένα διαμορφώνουν τη συμπεριφορά των εταιρειών και των οργανισμών. Τα δεδομένα αλλάζουν μέχρι και τον τρόπο με τον οποίο σκέφτονται οι άνθρωποι. Με δεδομένη την ισχύ των δεδομένων, η ικανότητα να αναλύουμε δεδομένα είναι μια ιδιαίτερη δεξιότητα την οποία η κοινωνία εκτιμά όλο και περισσότερο. Ωστόσο, εφόσον η δεξιότητα αυτή είναι τόσο σημαντική, μπορεί να χρησιμοποιηθεί, όπως και κάθε μορφή δύναμης/εξουσίας, είτε για καλό είτε για κακό. Παρόλο που πολλοί αναλυτές δεδομένων πασχίζουν να κάνουν το καλό παρουσιάζοντας τα ευρήματά τους με ακρίβεια και σαφήνεια, άλλοι διαστρεβλώνουν εσκεμμένα τα δεδομένα για προσωπική τους ωφέλεια. Ο στόχος του βιβλίου αυτού είναι να σας βοηθήσει να αναπτύξετε τις δικές σας δεξιότητες ανάλυσης δεδομένων ώστε να κάνετε καλά πράγματα στον κόσμο και να αναγνωρίζετε τότε κάποιοι αναλυτές δεδομένων διαστρεβλώνουν εσκεμμένα τα ευρήματά τους.

1.3 Η δύναμη των δεδομένων: Ομίχλη, ρύπανση και καταστροφή στο Λονδίνο

Δεν έχετε πειστεί ακόμα για τη δύναμη των δεδομένων; Ας δούμε με ένα παράδειγμα πώς τα δεδομένα, που συλλέχθηκαν επιμελώς και παρουσιάστηκαν στο κοινό με σαφήνεια, έσωσαν αμέτρητες ζωές και άλλαξαν για πάντα τη ζωή σε μία από τις μεγαλύτερες πόλεις του κόσμου.

Φανταστείτε ότι ζείτε στο Λονδίνο τον Δεκέμβριο του 1952. Οι καμινάδες της πόλης και τα τζάκια που καίνε κάρβουνο εκπέμπουν καυσάερια, τα οποία αναμειγνύονται με τη χειμωνιάτικη ομίχλη, και δημιουργούν μια πυκνή, αποπνικτική αιθαλομίχλη που σε παραλύει (MacNee & Donaldson, 2008, σ. 121). Η αιθαλομίχλη διαρκεί για πέντε ημέρες. Η ορατότητα μειώνεται τόσο πολύ που σταματά η κυκλοφορία των λεωφορείων και ακυρώνονται πολλές θεατρικές παραστάσεις ακόμα και σε εσωτερικούς χώρους, καθώς οι θεατές δεν μπορούν να δουν τη σκηνή (BBC News, 2008· Wallace & Hobbs, 2006, σ. 179). Ακόμα πιο τρομακτική είναι η επίδραση της ομίχλης στην ανθρώπινη υγεία. Ανάμεσα στις 5 και τις 9 Δεκεμβρίου του 1952, οι συνολικές εισαγωγές στα νοσοκομεία αυξάνονται κατά 50% και πολλές από τις εισαγωγές αυτές οφείλονται σε αναπνευστικά προβλήματα των ασθενών (MacNee & Donaldson, 2008, σ. 121).

Παρόλο που η ομίχλη του Δεκεμβρίου του 1952 ήταν εξαιρετικά πυκνή, τα περιστατικά ατμοσφαιρικής ρύπανσης αυτής της έκτασης εμφανίζονταν με ανησυχητική συχνότητα στο Λονδίνο του 19ου και των αρχών του 20ού αιώνα. Σε αρκετές προηγούμενες περιπτώσεις, η αιθαλομίχλη είχε σκοτώσει εκατοντάδες ανθρώπους στα τέλη του 19ου αιώνα (Brimblecombe, 1987, σ. 124). Εντούτοις, αυτές οι τραγωδίες του παρελθόντος δεν δρομολόγησαν πολιτικές αλλαγές. Παρόλο που κάποιοι παρατηρητές εξέφρασαν ανησυχίες για τις πιθανές συνέπειες στη δημόσια υγεία, για παράδειγμα, στα τέλη του 19ου αιώνα ένας μετεωρολόγος υποστήριξε ότι τα αυξημένα ποσοστά βρογχίτιδας στο Λονδίνο οφείλονταν εν μέρει στην αιθαλομίχλη (MacNee & Donaldson, 2008, σ. 121· Russell, 1889), δεν υπήρχαν ακόμη δεδομένα διαθέσιμα ώστε να αναλυθεί το ζήτημα. Ωστόσο, μέχρι το 1952 αυτό είχε αλλάξει. Η δυνατότητα να συλλέγονται δεδομένα για τις συνέπειες της αιθαλομίχλης, να αναλύονται με ακρίβεια και να ανακοινώνονται τα ευρήματα στο κοινό άλλαξε τα πάντα.

Για την ομίχλη του 1952, το πρώτο σύνολο σημαντικών δεδομένων δημοσιεύθηκε από το Υπουργείο Υγείας αρκετές εβδομάδες μετά την τραγωδία, όπου αποκαλύφθηκε ότι στο Λονδίνο καταμετρήθηκαν 2.800 επιπλέον θάνατοι κατά τη διάρκεια της ομίχλης σε σχέση με την ίδια εβδομάδα του 1951 (Thorsheim, 2006, σ. 165).¹ Αυτό το τραγικό στατιστικό στοιχείο έλαβε μεγάλη δημοσιότητα από τα μέσα ενημέρωσης τόσο στο Ηνωμένο Βασίλειο όσο και στον υπόλοιπο κόσμο (“Week of London fog...”, 1952). Παρακίνησε τον κόσμο να αντιμετωπίσει την ομίχλη του 1952 ως μια τραγωδία που θα μπορούσε να αποτραπεί μέσω πολιτικών δράσεων για τη ρύπανση (Thorsheim, 2004, σ. 166). Η δημόσια ανησυχία για την ομίχλη του 1952 άνοιξε τον δρόμο ώστε να υιοθετηθεί ο Νόμος για Καθαρό Αέρα το 1956, ο οποίος ρύθμιζε τις βιομηχανικές εκπομπές αερίων και απαγόρευε στα περισσότερα σπίτια και τις επιχειρήσεις να χρησιμοποιούν κάρβουνο για θέρμανση – μια τεράστια κοινωνική αλλαγή, που δεν είχε προκύψει όμως ως αντίδραση στις παλαιότερες ομίχλες (Thorsheim, 2006, σσ. 173-174).

1. Άλλες αναφορές υποστηρίζουν ότι πέθαναν περίπου 4.000 άνθρωποι από προβλήματα στο αναπνευστικό λόγω της αιθαλομίχλης (Wallace & Hobbs, 2006, σ. 179), ενώ σε μια πιο περιεκτική, σύγχρονη ανάλυση βρέθηκε ότι σχετίζονται με την αιθαλομίχλη μέχρι και 12.000 επιπλέον θάνατοι ανάμεσα στον Δεκέμβριο του 1952 και τον Φεβρουάριο του 1953 (Bell & Davis, 2001).

Το παράδειγμα αυτό απεικονίζει τη δύναμη που έχουν τα δεδομένα ώστε να τροποποιούν άμεσα την αντίληψη των ανθρώπων – ακόμα και να επηρεάζουν τις πολιτικές αποφάσεις. Μια τέτοια πολιτική αλλαγή δρομολογήθηκε μόνο αφότου δόθηκε στο κοινό μια ξεκάθαρη απόδειξη για τον αριθμό των ατόμων που πέθαναν από την αιθαλομίχλη. Με άλλα λόγια, χρειάστηκε να συλλεχθούν δεδομένα με ακρίβεια και να παρουσιαστούν με σαφήνεια ώστε να προκληθεί αυτός ο μετασχηματισμός.

1.4 Ο διαχρονικός αντίκτυπος των δεδομένων:

Το έργο του Alexis de Tocqueville

Τα δεδομένα που παρακίνησαν στη λήψη μέτρων για την ατμοσφαιρική ρύπανση ήταν ποσοτικά δεδομένα – δηλαδή αποτελούνταν από αριθμούς (στην περίπτωση αυτή, ένα απλό, ανησυχητικό στατιστικό στοιχείο για τα άτομα που πέθαναν). Όταν ακούμε τη λέξη «δεδομένα», συνήθως μας έρχονται στον νου στατιστικά στοιχεία όπως αυτό. Δεν είναι απαραίτητο όμως τα δεδομένα να αποτελούνται από αριθμούς. Αντίθετα, μπορεί να αποτελούνται από λέξεις, ενέργειες, συμπεριφορές, εικόνες, αντικείμενα και αμέτρητα άλλα χαρακτηριστικά της ιδιωτικής ζωής ή της δημόσιας σφαίρας. Τα δεδομένα που δεν είναι αριθμητικά ονομάζονται ποιοτικά δεδομένα και, όπως και στα ποσοτικά δεδομένα, η ανάλυση ποιοτικών δεδομένων έχει επηρεάσει με αντίστοιχο τρόπο την πολιτική και την κοινωνία με άπειρους τρόπους. Στην πραγματικότητα, ο απόηχος της επίδρασης αυτής θα διαρκεί για αιώνες στο μέλλον.

Θα δούμε λοιπόν ένα παράδειγμα του διαχρονικού αντίκτυπου των δεδομένων: την ποιοτική έρευνα ορόσημο του Alexis de Tocqueville. Το 1831 ο de Tocqueville και ο συνεργάτης του Gustave Beaumont ανέλαβαν ένα εξαντλητικό και προσεκτικά οργανωμένο ταξίδι διάρκειας εννέα μηνών στις ανατολικές, τις μεσοδυτικές και τις νότιες περιοχές των Ηνωμένων Πολιτειών. Παρόλο που η γαλλική κυβέρνηση του ανέθεσε να μελετήσει το σωφρονιστικό σύστημα των ΗΠΑ, ο de Tocqueville διεύρυνε τον σκοπό της ανάλυσής του ώστε να συμπεριλάβει όλα τα στοιχεία της κοινωνικής και πολιτικής ζωής του νεοϊδρυθέντος κράτους. Στην εκτενή και πρωτοποριακή εργασία που εκπόνησε ο de Tocqueville, *Η δημοκρατία στην Αμερική* (1835),² αναλύει με πολυεπίπεδο τρόπο τη

2. [Σ.τ.Μ.] *Η δημοκρατία στην Αμερική*, μετ. Μπάμπης Λυκούδης, εκδ. Στοχαστής, 2008.

δημοκρατία (και τη σοκαριστική αντίφαση της δουλοκτημοσύνης με τη δημοκρατία), τις κοινωνικές τάξεις, το φύλο, τις πολιτισμικές επιλογές και τις πολιτικές αξίες (Kurweil, 1999).

Παρόλο που το έργο του de Tocqueville θεωρείται συχνά ότι αφορά την πολιτική φιλοσοφία, στην πραγματικότητα αποτελεί ένα «παράδειγμα δημιουργικής ποιοτικής έρευνας», εφόσον, για να ολοκληρώσει ο de Tocqueville τη μελέτη του, έκανε συστηματική συλλογή και ανάλυση ποιοτικών δεδομένων για τη συμπεριφορά των Αμερικανών πολιτών και τους μηχανισμούς της δημοκρατικής κυβέρνησής τους (Lingenfelter, 2016, Κεφάλαιο 5). Πιο συγκεκριμένα, το έργο του de Tocqueville αποτελεί ένα πρώιμο παράδειγμα «συμμετοχικής παρατήρησης» (Whitley, 2008, σ. 98· δείτε επίσης τα Handler, 2005, σ. 22 και Kurweil, 1999, σ. 153), δηλαδή μιας μεθόδου ποιοτικής έρευνας στην οποία ένας ερευνητής διεξάγει εκτεταμένη έρευνα πεδίου και συμμετέχει στις καθημερινές δραστηριότητες και το πρόγραμμα μιας συγκεκριμένης περιοχής, ενός οργανισμού, μιας κοινωνίας ή μιας ομάδας που τον ενδιαφέρει. Η συμμετοχική παρατήρηση επιτρέπει στον ερευνητή να αποκτήσει μια προσωπική αίσθηση για το τι σημαίνει να είσαι μέλος μιας συγκεκριμένης ομάδας και να συμμετέχεις σε συγκεκριμένες δραστηριότητες. Για να αποκτήσει ο de Tocqueville αυτή τη γνώση, ταξίδεψε σε πολλές περιοχές της χώρας, κουβέντιασε με αμέτρητους Αμερικανούς και συμμετείχε ευρέως στην κοινωνική ζωή. Όντας ξένος, ο de Tocqueville κατάφερε να παρατηρήσει την κοινωνία που μελετούσε με ένα μοναδικό, οξυδερκές βλέμμα, και έχει πλέον αναγνωριστεί από επιστήμονες της εποχής μας ως «ο πρώτος σύγχρονος κοινωνικός επιστήμονας» (Kurweil, 1999, σ. 153).

Πιθανόν η πιο ξεκάθαρη απόδειξη της αξίας του έργου του de Tocqueville είναι ο διαχρονικός αντίκτυπός του που φτάνει μέχρι σήμερα. Ακόμα και τώρα, το βιβλίο *Η δημοκρατία στην Αμερική* χρησιμοποιείται τακτικά ως παραπομπή από τους πολιτισμικούς σχολιαστές και ανατίθεται συχνά ως εργασία στους φοιτητές των κοινωνικών επιστημών στα πανεπιστήμια των ΗΠΑ. Για παράδειγμα, σε ένα άρθρο του 2015 στην εφημερίδα *Washington Post* το έργο του de Tocqueville περιγράφεται ως «το βιβλίο που πρέπει να διαβάσει κάθε νέος Αμερικανός πολίτης – αλλά και οι μεγαλύτεροι» (Lozada, 2015). Και το 2017, σε ένα άρθρο των *BBC News* με τίτλο «Θα καταφέρει η δημοκρατία να επιβιώσει από την επίδραση του Facebook;» (Can democracy survive Facebook?) αναφέρεται το έργο του de Tocqueville

(Rajan, 2017). Το γεγονός ότι τα ευρήματα του de Tocqueville θεωρούνται ακόμα επίκαιρα τον 21ο αιώνα στον δημόσιο διάλογο για τα μέσα κοινωνικής δικτύωσης και τη δημοκρατία δείχνει ότι, μια ποιοτική έρευνα –όπως και μια ποσοτική έρευνα– που πραγματοποιείται εμπειριστατωμένα, μπορεί να έχει μακροχρόνιο αντίκτυπο. Αν ενδιαφέρεστε να μάθετε περισσότερα για τη συμμετοχική παρατήρηση και για άλλες μορφές ποιοτικής έρευνας, μπορείτε να περιμένετε μέχρι το Κεφάλαιο 6, όπου θα εξετάσουμε λεπτομερώς την ανάλυση ποιοτικών δεδομένων.

1.5 Τα ερωτήματα που προάγουν την ανάλυση δεδομένων: Το έργο του Adolphe Quetelet

Η ανάλυση δεδομένων προάγεται από ερωτήματα. Το πρώτο βήμα για να διεξαγάγετε μια ενδιαφέρουσα και εμπειριστατωμένη κοινωνική έρευνα είναι να μάθετε να θέτετε ενδιαφέροντα και βαθυστόχαστα ερωτήματα. Ένας από τους στόχους του βιβλίου αυτού είναι να σας εμπνεύσει να διατυπώνετε τέτοια ερωτήματα σχετικά με τον κόσμο. Σε όλα τα κεφάλαια θα μελετήσουμε παραδείγματα από τον πραγματικό κόσμο, δηλαδή κάποια από τα πλέον συναρπαστικά ερωτήματα που έχουν θέσει οι αναλυτές δεδομένων. Στην πραγματικότητα, θα ξεκινήσουμε αμέσως τώρα με ένα τέτοιο παράδειγμα: το έργο του Lambert Adolphe Jacques (ή πιο απλά Adolphe) Quetelet (1796-1874). Η πρωτοποριακή ποσοτική έρευνα του Quetelet δείχνει με ποιον τρόπο τα βαθυστόχαστα ερωτήματα μπορούν να παρακινήσουν την καινοτόμο ανάλυση δεδομένων και να οδηγήσουν σε μη αναμενόμενες αλλά θεμελιώδεις γνώσεις για τον κοινωνικό κόσμο. Ο Quetelet, ένας εξαιρετικός πρωτοπόρος αναλυτής δεδομένων από το Βέλγιο, παρουσίασε ρηξικέλευθα ευρήματα σε ένα μεγάλο εύρος επιστημονικών κλάδων, συμπεριλαμβανομένων της ιατρικής και της εγκληματολογίας. Θα επιλέξουμε από το έργο του Quetelet δύο επίμαχα θέματα: τη σχέση ανάμεσα στο ανθρώπινο ύψος και βάρος, και τη σχέση ανάμεσα στον καιρό και την εγκληματικότητα.

1.5.1 Ύψος και βάρος

Πώς μεταβάλλεται το ύψος και το βάρος ενός ατόμου κατά τη διάρκεια της ζωής του; Επηρεάζουν τις όποιες αλλαγές το φύλο, η κοινωνική τάξη και η γεωγραφική περιοχή; Τα ερωτήματα αυτά, που εμπεριέχονται στις σύγχρονες συζητήσεις για τα αυξανόμενα επίπεδα παχυσαρκίας, είχαν συναρπάσει και τον Quetelet τη δεκαετία του 1840. Ωστόσο, σε αντίθεση με άλλους αναλυτές που έκαναν εικασίες, ο Quetelet ασχολήθηκε με την ανάλυση ποσοτικών δεδομένων για να δώσει ενδελεχείς απαντήσεις στα ερωτήματα αυτά. Τα δεδομένα του Quetelet (1842) προέρχονται από ένα ευρύ φάσμα πηγών και περιλαμβάνουν κρατικά αρχεία στο Βέλγιο, μετρήσεις σε βρέφη από το Νοσοκομείο Έκθετων Βρεφών στο Βέλγιο, μετρήσεις σε παιδιά που εργάζονταν σε εργοστάσια στο Μάντσεστερ και το Στόκπορτ (στην Αγγλία), και μετρήσεις σε προπτυχιακούς φοιτητές στο Πανεπιστήμιο του Κέιμπριτζ.

Ο Quetelet (1842) παρατήρησε ότι τα πιο υγιή άτομα είναι γενικά πιο ψηλά από τον μέσο όρο, και ότι η ανάπτυξη των πιο φτωχών ατόμων καθυστερεί λόγω της φτώχειας και των στερήσεων.³ Επίσης παρατήρησε ότι γενικά το σωματικό βάρος (σε χιλιόγραμμα) ενός μέσου ατόμου είναι ανάλογο με το τετράγωνο του ύψους του (σε μέτρα) – μια παρατήρηση που η επίδρασή της διαρκεί μέχρι σήμερα (Eknoyan, 2008, σ. 48) καθώς οδήγησε στον σχηματισμό του Δείκτη Μάζας Σώματος (ΔΜΣ) (BMI, Keys, Karvonen, Kimura, & Taylor, 1972). Σήμερα, ο ΔΜΣ αποτελεί έναν από τους πιο ευρέως χρησιμοποιούμενους δείκτες για να αξιολογηθεί ένα άτομο ως λιποβαρές, υπέρβαρο, παχύσαρκο ή κανονικού βάρους και, καθώς η παχυσαρκία αποτελεί πλέον ένα σημαντικό ζήτημα για την παγκόσμια υγεία, οι επαγγελματίες της υγείας και τα μέσα μαζικής ενημέρωσης δίνουν μεγάλη σημασία στον ΔΜΣ. Για παράδειγμα, τα Κέντρα Ελέγχου και Πρόληψης Νοσημάτων στις

3. Παρόλο που συνολικά ο Quetelet (1842) βάσισε το έργο του σε ένα εύρος πηγών όπως αυτά που αναφέραμε νωρίτερα, για τα συγκεκριμένα ευρήματα χρησιμοποίησε ένα πιο περιορισμένο σύνολο δεδομένων, το οποίο περιλάμβανε μόνο δεδομένα από τις Βρυξέλλες και τη βελγική επαρχία της Βραβάντης. Γι' αυτό και ο Quetelet επισήμανε ότι τα ευρήματά του μπορούν να εφαρμοστούν μόνο στις περιοχές αυτές, και δεν εγγυόταν ότι ισχύουν και για άλλες περιοχές. Αυτή η ρητή αναγνώριση των περιορισμών που διέπουν τα ευρήματά μας είναι θεμελιώδης για όλους τους αναλυτές δεδομένων.

ΗΠΑ (CDC, 2012) τονίζουν ότι: «Για τους ενήλικες, χαρακτηρίζουμε ένα άτομο ως υπέρβαρο ή παχύσαρκο χρησιμοποιώντας το βάρος και το ύψος του ώστε να υπολογίσουμε έναν αριθμό που ονομάζεται “δείκτης μάζας σώματος”». Ο ΔΜΣ χρησιμοποιείται επίσης για να καταγράφεται παγκοσμίως η τάση για παχυσαρκία, κάτι που έχει αποκαλύψει ότι από το 1980 μέχρι το 2013 το ποσοστό των παχύσαρκων ανδρών σε ολόκληρο τον κόσμο αυξήθηκε από το 19% στο 37%, ενώ το ποσοστό των γυναικών αυξήθηκε από το 30% στο 38% (Ng κ.ά., 2014).

Ο ΔΜΣ έχει πολλούς περιορισμούς. Εφόσον δεν μας βοηθά να κάνουμε τη διάκριση ανάμεσα στο λίπος και τους μυς, μπορεί να μην είναι χρήσιμος για όλα τα άτομα (όπως π.χ. τους αθλητές, που έχουν μεγάλη μυϊκή μάζα), αλλά να είναι χρήσιμος σε επίπεδο πληθυσμού, όπως δείχνει και το παράδειγμα με την τάση για παχυσαρκία σε ολόκληρο τον κόσμο (δείτε περισσότερα στο Stephenson, 2013). Παρόλο που υπάρχουν πλέον πιο εξελιγμένες μέθοδοι για τη μέτρηση της παχυσαρκίας, είναι πιθανόν η απλότητα και η ευκολία με την οποία υπολογίζεται ο ΔΜΣ να ενισχύουν την προτίμηση προς αυτόν. Επιπλέον, όπως περιγράφει ο Wells (2014), δεν είναι ξεκάθαρο αν κάποιος εναλλακτικοί τρόποι μέτρησης, όπως π.χ. η περιφέρεια της μέσης, αποτελούν καλύτερους δείκτες για τον κίνδυνο να εμφανιστούν χρόνιες ασθένειες από τον ΔΜΣ.

Η ανθεκτικότητα και η σταθερή αξία του ΔΜΣ δείχνουν την επίδραση που μπορεί να έχει μια εμπεριστατωμένη και καλά σχεδιασμένη ανάλυση δεδομένων. Η παρατήρηση του Quetelet σχετικά με το ανθρώπινο ύψος και βάρος που ενέπνευσε τον ΔΜΣ δεν βασίστηκε στη διαίσθηση ή σε υποθέσεις. Όπως είδαμε, η παρατήρηση του Quetelet (1842) βασίστηκε στην εμπεριστατωμένη του ανάλυση των δεδομένων για το ύψος και το βάρος πολλών διαφορετικών ατόμων. Αν η παρατήρησή του βασιζόταν περισσότερο σε εικασίες πιθανόν να μην ήταν τόσο σωστή ή να μην επιτύγχανε την απαιτούμενη ακρίβεια ώστε να επιβιώσει μέχρι τον 21ο αιώνα, όπως κατάφερε ο ΔΜΣ. Επιπλέον, η ανθεκτικότητα του ΔΜΣ τονίζει επίσης τη σπουδαιότητα του τρόπου παρουσίασης των δεδομένων. Η αρχική παρατήρηση του Quetelet (1842) είναι ένας απλός τύπος, κατανοητός σε όλους, και μάλλον είναι αυτή η απλότητα που βοήθησε να αντέξει ο ΔΜΣ, παρόλο που σήμερα διαθέτουμε πλέον πολύ πιο εξελιγμένες τεχνικές για να μετράμε την παχυσαρκία.

1.5.2 Καιρός και εγκληματικότητα

Πέρα από το ύψος και το βάρος, ο Quetelet ερευνήσε το αμφιλεγόμενο ερώτημα αν ο καιρός επηρεάζει τους δείκτες της εγκληματικότητας. Το ερώτημα αυτό εξακολουθεί να απασχολεί τους επιστήμονες, και οι παρατηρήσεις του Quetelet, που βασίζονταν σε δεδομένα, αποδείχθηκαν σημαντικές και στον τομέα αυτό (δείτε, για παράδειγμα, Rotton & Frey, 1985· Yan, 2004).

Για να μελετήσει το ερώτημα αυτό, ο Quetelet (1842) εξέτασε δεδομένα από τη Γαλλία της δεκαετίας του 1820 και διαπίστωσε ότι συνήθως τους πιο κρύους μήνες του έτους σημειώνονταν περισσότερα εγκλήματα κατά της ιδιοκτησίας, ενώ τους πιο ζεστούς μήνες του έτους σημειώνονταν περισσότερα εγκλήματα βίας (δείτε επίσης Bray, 1994, σ. 68). Ο Quetelet εξήγησε το μοτίβο αυτό υποστηρίζοντας ότι η υπερβολική ζέστη αυξάνει την επιθετικότητα των ανθρώπων, η οποία ωθεί σε περισσότερα εγκλήματα βίας. Αντίστοιχα, τον χειμώνα η έλλειψη τροφής και άλλων αναγκαίων μέσων της καθημερινότητας οδηγεί σε περισσότερα εγκλήματα κατά της ιδιοκτησίας (Hipp, Bauer, Curran, & Bollen, 2004).

Παρόλο που η επίδραση του καιρού στην εγκληματικότητα δεν έχει μελετηθεί τόσο σχολαστικά όσο η επίδραση των κοινωνικοοικονομικών και δημογραφικών παραγόντων, πολλοί ερευνητές έχουν ενδιαφερθεί για το ζήτημα (Cohn, 1990). Όπως και ο Quetelet, διάφοροι επιστήμονες έχουν διατυπώσει θεωρίες για μια πιθανή σύνδεση ανάμεσα στις υψηλές θερμοκρασίες και την επιθετικότητα ή την εχθρικότητα (δείτε περισσότερα στο Anderson, 2001). Επιπλέον, όταν επικρατούν υψηλότερες θερμοκρασίες, οι άνθρωποι περνούν περισσότερο χρόνο έξω από το σπίτι τους και έρχονται σε επαφή με περισσότερους ανθρώπους, κάτι που είναι πιθανό να αυξήσει τις πιθανότητες για βία και εγκληματικότητα (Moser, 2012).

ΔΙΑΛΕΙΜΜΑ ΓΙΑ ΣΚΕΨΗ

Θεωρείτε ότι ο καιρός επηρεάζει τους δείκτες εγκληματικότητας; Αν ναι, συμφωνείτε με την ιδέα του Quetelet ότι τα εγκλήματα βίας είναι συχνότερα τις πιο ζεστές περιόδους του έτους, ενώ τις πιο κρύες περιόδους είναι συχνότερα τα εγκλήματα κατά της περιουσίας; Γιατί ναι ή γιατί όχι; Αν πραγματοποιούσατε μια ανάλυση δεδομένων για να απαντήσετε στα ερωτήματα αυτά, τι δεδομένα θα συλλέγατε;

Η άποψη του Quetelet, ότι σημειώνονται περισσότερα εγκλήματα βίας όταν επικρατούν υψηλότερες θερμοκρασίες, στηρίζεται και από πρόσφατες αναλύσεις δεδομένων. Για παράδειγμα, οι Butke και Sheridan (2010) εξέτασαν τα αρχεία της αστυνομίας του Κλίβελαντ στις ΗΠΑ για τα έτη 1999-2004 και εστίασαν στα ακόλουθα είδη αδικημάτων: ενδοοικογενειακή βία, βιαιοπραγίες, ληστείες, βιασμοί και δολοφονίες. Η ανάλυση που έκαναν στα δεδομένα αποκάλυψε ότι η εμφάνιση τέτοιου είδους αδικημάτων κορυφώνεται τους θερινούς μήνες και μειώνεται κατακόρυφα τους χειμερινούς μήνες. Έδειξε επίσης ότι τα αδικήματα αυτά αυξάνονται παράλληλα με την αύξηση της θερμοκρασίας.

Οι McDowall, Loftin και Pate (2012) διερεύνησαν παρόμοια ερωτήματα χρησιμοποιώντας μεγαλύτερο γεωγραφικό εύρος και χρονικό διάστημα. Πιο συγκεκριμένα, εξέτασαν τα μηνιαία δεδομένα για τα έτη 1977-2000 από όλες τις αμερικανικές πόλεις με πληθυσμό μεγαλύτερο από 200.000 κατοίκους το 2000. Χρησιμοποίησαν δεδομένα από τις Ενιαίες Αναφορές Εγκλήματος (Uniform Crime Reports) του FBI, όπου καταχωρίζονται τα αδικήματα που δηλώνονται στις αστυνομικές αρχές, ώστε να πάρουν πληροφορίες για τους δείκτες δολοφονιών, βιασμών, κλοπών, βιαιοπραγιών, διαρρήξεων, ληστειών και κλοπών οχημάτων. Ανακάλυψαν ότι οι δείκτες όλων αυτών των αδικημάτων, εκτός από τις ληστείες, κορυφώνονται τον Ιούλιο ή τον Αύγουστο και ελαχιστοποιούνται τον Φεβρουάριο. Παρόλο που τα ευρήματα αυτά ταιριάζουν εν μέρει με την υπόθεση του Quetelet, παραδοσιακά οι κλοπές και οι ληστείες θεωρούνται εγκλήματα κατά της ιδιοκτησίας και, σε αντίθεση με αυτό που πρότεινε ο Quetelet (1842), τα αδικήματα αυτά επίσης αυξάνονται κατά τους θερινούς μήνες.

Μια άλλη σημαντική ανάλυση δεδομένων για τη σχέση ανάμεσα στον καιρό και την εγκληματικότητα πραγματοποιήθηκε από τους Cohn και Rotton (2000), οι οποίοι ανέλυσαν τα δεδομένα της αστυνομίας για την εγκληματικότητα κατά τη διάρκεια δύο ετών στην αμερικανική πόλη της Μινεάπολης. Και αυτοί βρήκαν επίσης ότι, σε αντίθεση με αυτό που πρότεινε ο Quetelet, κατά τη διάρκεια των πιο ζεστών περιόδων του έτους δηλώνονται στην αστυνομία περισσότερα εγκλήματα κατά της ιδιοκτησίας απ' ό,τι στις πιο κρύες περιόδους. Για να ερμηνεύσουν τα ευρήματα αυτά, πρότειναν ότι, όταν είναι καλύτερος ο καιρός, εμφανίζονται περισσότερες ευκαιρίες για τη διάπραξη συγκεκριμένων εγκλημάτων κατά της ιδιοκτησίας, εφόσον είναι πιο πιθανό τα

δυνάμει θύματα να αφήσουν ανοιχτά τα παράθυρά τους ή να αφήσουν διάφορα αγαθά έξω από το σπίτι τους.

Ποιες είναι οι επιπτώσεις των ευρημάτων αυτών στα επιχειρήματα του Quetelet; Φαίνεται ότι οι ισχυρισμοί του για τα εγκλήματα βίας υποστηρίζονται εν μέρει και από σύγχρονες αναλύσεις δεδομένων, παρόλο που η σύνδεση ανάμεσα στη ζέστη και την επιθετικότητα, στην οποία εστίασε ο Quetelet (1842), μάλλον δεν είναι η μοναδική εξήγηση. Η άποψη του Quetelet για τα εγκλήματα κατά της ιδιοκτησίας δεν υποστηρίζονται από τις σύγχρονες μελέτες που αναφέρουμε εδώ. Η απόκλιση αυτή δείχνει ότι τα κοινωνικά φαινόμενα μεταβάλλονται με το πέρασμα του χρόνου και απαιτούνται επικαιροποιημένες αναλύσεις για την αξιολόγηση των νέων δεδομένων. Όταν τα ευρήματα μιας επικαιροποιημένης ανάλυσης δεδομένων αποκλίνουν από τα ευρήματα μιας παλαιότερης ανάλυσης, είναι σημαντικό να διαθέσουμε λίγο χρόνο ώστε να αναρωτηθούμε για ποιον λόγο μπορεί να συμβαίνει αυτό. Υπάρχει πραγματικά κάποια μεταβολή στο φαινόμενο που μας ενδιαφέρει (δηλαδή στο συγκεκριμένο παράδειγμα η σχέση μεταξύ καιρού και εγκληματικότητας) ή οι διαφορετικές μέθοδοι ανάλυσης των δεδομένων μάς προσφέρουν νέες δυνατότητες κατανόησης του φαινομένου (ενώ το ίδιο δεν έχει μεταβληθεί);

Παρόλο που ο Quetelet εστίασε στη θερμοκρασία, ίσως υπάρχουν και άλλα χαρακτηριστικά του καιρού που επηρεάζουν επίσης τους δείκτες εγκληματικότητας. Πριν συνεχίσουμε παρακάτω, έχει ενδιαφέρον να εξετάσουμε ένα από τα χαρακτηριστικά αυτά: τη βροχή. Μολονότι έχει μελετηθεί λιγότερο από τη θερμοκρασία, σε μια ανάλυση της εφημερίδας *New York Times* εξετάζονται δεδομένα για τις βροχοπτώσεις και τις δολοφονίες στη Νέα Υόρκη κατά τα έτη 2003-2008 (Lehren & Hauser, 2009). Κατά την ανάλυση βρέθηκε ότι σημειώθηκαν λιγότερες δολοφονίες κατά τη διάρκεια των περισσότερων βροχερών καλοκαιριών. Οι πιθανές ερμηνείες για τα ευρήματα αυτά σχετίζονται με τη θερμοκρασία (εφόσον οι βροχερές ημέρες είναι γενικά πιο ψυχρές, θα περιμέναμε να βρούμε λιγότερα εγκλήματα βίας, σύμφωνα με τα ευρήματα των μελετών που αναφέραμε νωρίτερα) ή με το ότι οι βροχερές ημέρες προσφέρουν λιγότερες ευκαιρίες για δολοφονίες αγνώστων, εφόσον είναι πιο πιθανό τα θύματα να παραμένουν στο σπίτι τους και να μη συναντούν αγνώστους. [Ωστόσο, οι συγγραφείς αναφέρουν ότι ένα από τα μειονεκτήματα της δεύτερης ερμηνείας είναι πως οι περισσότερες δολοφονίες δεν αφορούν δολοφονίες αγνώστων (Lehren & Hauser, 2009).]

Παραδόξως, στοιχεία από άλλες πόλεις υποδεικνύουν ένα παρόμοιο μοτίβο. Για παράδειγμα, δεδομένα από τη Γλασκόβη στη Σκωτία αποκαλύπτουν ότι τα εγκλήματα βίας είναι κατά σχεδόν 45% λιγότερα τις βροχερές ημέρες, ενώ τα αδικήματα που δηλώνονται συνολικά είναι κατά 25% λιγότερα (“Rainy days help...”, 2011). Ωστόσο, δεν παρατηρείται το ίδιο μοτίβο για τους βιασμούς (“Rainy days help...”, 2011). Στο Λος Άντζελες, κατά τη διάρκεια μιας ασυνήθιστα βροχερής εβδομάδας το 2005, τα εγκλήματα βίας ήταν κατά 23% λιγότερα, ενώ τα εγκλήματα κατά της ιδιοκτησίας ήταν κατά 25% λιγότερα σε σχέση με την αντίστοιχη εβδομάδα του προηγούμενου έτους. Ωστόσο, η παιδική κακοποίηση και η ενδοοικογενειακή βία ήταν κατά 25% υψηλότερη σε σχέση με την αντίστοιχη εβδομάδα του προηγούμενου έτους (Winton & Lee, 2005). (Τα ευρήματα αυτά βγάζουν νόημα αν αναλογιστούμε ότι είναι πιο πιθανό τα συγκεκριμένα αδικήματα να σημειωθούν στο σπίτι, και λόγω της βροχής είναι πιο πιθανό οι άνθρωποι να βγαίνουν λιγότερο από το σπίτι.) Ουσιαστικά η έρευνα αποκάλυψε ότι, όταν ο καιρός είναι πιο ευχάριστος, η εγκληματικότητα συνήθως αυξάνεται, ενώ σε αντίξοες συνθήκες ή σε ακραία υψηλές ή χαμηλές θερμοκρασίες η εγκληματικότητα μάλλον μειώνεται (Winton & Lee, 2005).

ΔΙΑΛΕΙΜΜΑ ΓΙΑ ΣΚΕΨΗ

Τώρα που έχετε εξετάσει πλέον διάφορα ευρήματα από αναλυτές δεδομένων αναφορικά με τη σχέση ανάμεσα στον καιρό και την εγκληματικότητα, θεωρείτε ότι τα ευρήματα αυτά υποστηρίζουν ή έρχονται σε αντίθεση με τις απαντήσεις που δώσατε στα ερωτήματα του προηγούμενου «διαλείμματος για σκέψη»; Με απλά λόγια, ταιριάζουν τα ευρήματα αυτά με τις προβλέψεις σας; Ποια επιπλέον ερωτήματα πρέπει να απαντηθούν σε μελλοντικές έρευνες;

Είτε το πιστεύετε είτε όχι, με αυτή την απλή άσκηση «διαλείμματος για σκέψη» έχετε ήδη κάνει το πρώτο βήμα ώστε να κατανοήσετε μία από τις πιο θεμελιώδεις διαδικασίες της έρευνας: τον έλεγχο υποθέσεων. Υπόθεση ονομάζεται μια αυστηρά διατυπωμένη πρόταση ή ένα επιχείρημα που εντοπίζει μια σχέση ή εξηγεί ένα φαινόμενο του πραγματικού κόσμου. Όπως θα βλέπουμε συχνά στο παρόν βιβλίο, οι ερευ-

νητές μπορούν να ελέγξουν την εγκυρότητα μιας συγκεκριμένης υπόθεσης ελέγχοντας και αναλύοντας δεδομένα. Ο έλεγχος υποθέσεων είναι πολύ θεμελιώδης στην έρευνα, οπότε θα επιστρέψουμε σε αυτόν με περισσότερες λεπτομέρειες στο Κεφάλαιο 2.

1.6 Ορίζοντας τον όρο «δεδομένα»

Στις προηγούμενες σελίδες συζητήσαμε αναλυτικά για τα δεδομένα, ήρθε όμως η στιγμή να αφιερώσουμε λίγο χρόνο και να αναλογιστούμε με πιο συστηματικό τρόπο για την έννοια αυτή. Τι είναι τα δεδομένα; Στην αρχή του παρόντος κεφαλαίου ορίσαμε τα δεδομένα ως «ποσοτικές ή ποιοτικές πληροφορίες σχετικά με τον εαυτό μας, την κοινωνία και το σύμπαν». Αυτός είναι ο ορισμός που θα χρησιμοποιήσουμε σε ολόκληρο το βιβλίο, καθώς τονίζει την ποικιλομορφία και την ευρύτητα των δεδομένων. Ωστόσο, ο ορισμός μου δεν είναι ο μοναδικός που έχουν προτείνει οι ειδικοί ως προς το θέμα επιστήμονες. Για παράδειγμα, ο Van Beveren (2002, σ. 19) υποστηρίζει ότι τα δεδομένα είναι «ακατέργαστες πληροφορίες». Για τους Meaden και Chi (1996, σ. 12) τα δεδομένα είναι «επιλεγμένες παρατηρήσεις που έχουν καταγραφεί και αποθηκευτεί». Από την άλλη πλευρά, οι Polit και Hungler (1999, σ. 267) υποστηρίζουν ότι μπορούμε να χρησιμοποιήσουμε τον όρο «δεδομένα» για οποιαδήποτε πληροφορία αποκτούμε κατά τη διαδικασία μιας μελέτης ή μιας έρευνας. Οι απόψεις των ειδικών ως προς το θέμα επιστημόνων διίστανται σχετικά με το αν πρέπει να χρησιμοποιούμε τη λέξη «δεδομένα» (data) ως ουσιαστικό ενικού ή πληθυντικού. Μιλώντας κυριολεκτικά, η λέξη βρίσκεται στον πληθυντικό (που σημαίνει ότι πρέπει να γράφουμε «Τα δεδομένα είναι...»), συχνά όμως χρησιμοποιείται σαν να είναι στον ενικό («Το δεδομένο είναι...»). Στην πραγματικότητα, στο παρόν βιβλίο χρησιμοποιούμε συχνά τη μορφή του ενικού.⁴ Αυτή η γραμματικής φύσης διαφωνία υπογραμμίζει ένα πολύ μεγαλύτερο θέμα φιλοσοφικής φύσης: Τα δεδομένα αποτελούν μια ετερογενή έννοια, και δεν υπάρχει κάποιος συγκεκριμένος ορισμός που να τον αποδέχονται όλοι οι αναλυτές (δείτε περισσότερες πληροφορίες σχετικές με τις αμέτρητες διαφορετικές προσεγγίσεις που έχουν

4. [Σ.τ.Μ.] Στην παρούσα έκδοση μεταφράζουμε τον όρο data μόνο στον πληθυντικό: δεδομένα.

οι σχετικοί με το θέμα επιστήμονες για να δώσουν έναν ορισμό στα δεδομένα στο Fox, Levitin, & Redman, 1994, σσ. 11-12).

Ανεξάρτητα από τον ορισμό που προτιμάτε για τα δεδομένα, είναι σημαντικό να κατανοήσετε ότι πρέπει να διαχειρίζεστε και να οργανώνετε τα δεδομένα με τέτοιον τρόπο ώστε να προσφέρουν γνώση που έχει αξία για τα άτομα που τα χρησιμοποιούν (Hoffer, Prescott, & McFadden, 2002· Van Beveren, 2002, σ. 19). Εφόσον έχουμε οργανώσει τα δεδομένα σε μια αξιοποιήσιμη μορφή, μπορούμε να τα αναλύσουμε ώστε να μας παρέχουν πληροφορίες (Meaden & Chi, 1996, σ. 12). Θα μελετήσουμε σε βάθος το ζήτημα της πλέον αποτελεσματικής οργάνωσης των δεδομένων ώστε να προετοιμαστούμε για την πιο λεπτομερή ανάλυσή του σε επόμενα κεφάλαια του παρόντος βιβλίου.

1.6.1 Σκεφτόμαστε για τα δεδομένα με δημιουργικό τρόπο

Σε προηγούμενες ενότητες του κεφαλαίου μελετήσαμε αρκετά παραδείγματα ερευνητών που σκέφτονταν με δημιουργικό τρόπο για τα δεδομένα. Από τις βαθυστόχαστες, ποιοτικές παρατηρήσεις του de Tocqueville για τις κοινωνικές δυναμικές μέχρι τις καινοτόμες, ποσοτικές παρατηρήσεις του Quetelet για την επίδραση του καιρού, είναι προφανές ότι τα δεδομένα μπορούν να αναλυθούν με μια μεγάλη ποικιλία διαφορετικών και πρωτότυπων τρόπων. Καθώς ξεκινάτε το ταξίδι σας στην ανάλυση δεδομένων, κρατήστε αυτό στο μυαλό σας: Δεν υπάρχει ένας και μοναδικός «σωστός» τρόπος να αναλύετε δεδομένα. Ναι, υπάρχουν βέλτιστες πρακτικές και κατευθυντήριες γραμμές περί δεοντολογίας που όλοι οι υπεύθυνοι αναλυτές δεδομένων οφείλουν να ακολουθούν (θα τις συζητήσουμε στο Κεφάλαιο 3), πέρα από τους κανόνες αυτούς, όμως, οι αναλυτές δεδομένων έχουν μεγάλη ελευθερία να επιλέγουν ανάμεσα σε διαφορετικές προσεγγίσεις ανάλυσης. Για να γίνετε «καλοί» αναλυτές, πρέπει να μάθετε να χρησιμοποιείτε με σύνεση την ελευθερία αυτή και να σκέφτεστε δημιουργικά για τον τρόπο με τον οποίο θα δώσετε τις καλύτερες απαντήσεις στα ερευνητικά σας ερωτήματα. Συχνά, το να σκεφτόμαστε δημιουργικά για τα δεδομένα σημαίνει να συνδυάζουμε διαφορετικές μεθόδους ανάλυσης. Για παράδειγμα, αντί να στηριζόμαστε αποκλειστικά είτε στις ποσοτικές είτε στις ποιοτικές μεθόδους, μπορούμε να επιλέξουμε να συνδυάσουμε τα αντίστοιχα πλεονεκτήματά τους.

Ένα παράδειγμα από τον πραγματικό κόσμο που αποτελεί ορόσημο για αυτές τις δημιουργικές προσεγγίσεις αποτελεί η μελέτη «Mid-

dletown», την οποία διεύθυναν οι κοινωνιολόγοι Robert S. Lynd και Helen Merrell Lynd τη δεκαετία του 1920. Οι Lynd ήθελαν να κατανοήσουν τον τρόπο με τον οποίο η οικονομική ανάπτυξη και η εκβιομηχάνιση στα τέλη του 19ου και τις αρχές του 20ού αιώνα επηρέασαν τη ζωή σε μια τυπική, μεσαίου μεγέθους αμερικανική πόλη. Για να απαντήσουν τα ερευνητικά τους ερωτήματα, πραγματοποίησαν έρευνα πεδίου σε ένα διάστημα τριών ετών στην πόλη Μάντσι της Ιντιάνα, την οποία αναφέρουν με το ψευδώνυμο Middletown στα ευρήματα που δημοσιοποίησαν (Geelhoed, 2000· Emmel, 2013). Οι Lynd και Lynd (1929) χρησιμοποίησαν ποσοτικά και ποιοτικά δεδομένα για να διερευνήσουν τα ερευνητικά τους ερωτήματα. Πιο συγκεκριμένα, πήραν συνεντεύξεις, ανέλυσαν σχετικά τεκμήρια όπως οι τοπικές εφημερίδες και συνέλεξαν κομβικά στατιστικά στοιχεία. Η χρήση διαφορετικών ειδών δεδομένων ήταν μια εκπληκτική στρατηγική και τους επέτρεψε να σχηματίσουν ένα ζωντανό πορτρέτο της κοινότητας αυτής. Παρόλο που την εποχή των Lynd η ταυτόχρονη χρήση ποσοτικών και ποιοτικών δεδομένων στην ίδια μελέτη ήταν καινοτόμος, σήμερα είναι πλέον πιο δημοφιλής. Η προσέγγιση αυτή ονομάζεται έρευνα μεικτών μεθόδων, και θα τη μελετήσουμε λεπτομερώς στο Κεφάλαιο 8.

Οι Lynd εισήγαγαν επιπλέον μια δημιουργική οπτική στο έργο τους προσπαθώντας να μελετήσουν το Μάντσι ως «ξένοι». Με άλλα λόγια, ερεύνησαν τη βιομηχανία, την κοινωνία, τις κοινωνικές τάξεις, τη θρησκεία και την οικογενειακή ζωή σαν να προέρχονταν από έναν ξένο πολιτισμό (Peters & Simonson, 2004, σ. 58). (Αν θυμάστε, η προσέγγιση αυτού του είδους βοήθησε και τον de Tocqueville.) Ο σκοπός των Lynd ήταν να χρησιμοποιήσουν τη μοναδική, λεπτομερή μελέτη περίπτωσης που έκαναν στο Middletown ώστε να διατυπώσουν ευρύτερα συμπεράσματα για τον τρόπο που η εκβιομηχάνιση επηρέασε τις ζωές των ανθρώπων, και όχι μόνο στο Μάντσι της Ιντιάνα, αλλά στο σύνολο των «μέσων Αμερικανών» (Schreier, 2018). Το βιβλίο όπου αναφέρουν τα ευρήματά τους, *Middletown: A Study in Modern American Culture* (Lynd & Lynd, 1929), βρήκε εξαιρετική απήχηση στο κοινό, και μάλιστα «τα βιβλιοπωλεία το τοποθετούσαν δίπλα στα δημοφιλή μυθιστορήματα εκείνης της εποχής» (Horowitz, 2004, Κεφάλαιο 32) – ένδειξη της επινοητικότητας και της σπουδαιότητας του έργου τους.

Το βιβλίο ενέπνευσε αμέτρητες μελέτες παρακολούθησης (follow up studies), ξεκινώντας με την προσπάθεια των ίδιων των Lynd τη δεκαετία του 1930 να αξιολογήσουν τον τρόπο με τον οποίο η Παγκό-

σμια Οικονομική Ύφεση του 1929 και η παρεπόμενη αύξηση της ανεργίας επηρέασε την κοινωνική και οικογενειακή ζωή στο Middletown (Lynd & Lynd, 1937). Τη δεκαετία του 1970, μια νέα γενιά κοινωνιολόγων επέστρεψε στο Middletown και επιστράτευσαν πολλές από τις ερευνητικές μεθόδους που είχαν χρησιμοποιήσει αρχικά οι Lynd, ώστε να αξιολογήσουν τις αλλαγές που είχαν συμβεί στην πόλη στα πενήντα χρόνια από το αρχικό έργο των Lynd (Κέντρο Σπουδών για το Middletown, 2018). Ως μέρος αυτής της επικαιροποιημένης έρευνας, οι Caplow, Bahr, Chadwick, Hill και Williamson (1982) πραγματοποίησαν μελέτες στα τέλη της δεκαετίας του 1970 και συνέλεξαν εκτεταμένα ποσοτικά δεδομένα για μυριάδες λεπτομέρειες της κοινωνικής ζωής, όπως το πλήθος των ωρών ανά εβδομάδα που περνούν οι μητέρες και οι πατεράδες με τα παιδιά τους (ταξινομημένα ανά κοινωνική τάξη) και τη συχνότητα εκκλησιασμού, σύμφωνα με τις αναφορές των ενηλίκων της πόλης. Τις δεκαετίες του 1990 και του 2000, ανέδειξαν την αδυναμία της αρχικής μελέτης να εξηγήσει πλήρως την ποικιλομορφία του Middletown και έδωσαν μεγαλύτερη σημασία στις κοινότητες των Αφροαμερικανών και των Εβραίων της πόλης (Emmel, 2013· δείτε επίσης τα Lassiter, Goodall, Campbell, & Johnson, 2004· Rottenberg, 1997). Οι ερευνητές συνέχισαν να μελετούν το Middletown, και το Κέντρο Σπουδών για το Middletown (2018) λειτουργεί στο Πανεπιστήμιο Μπολ Στέιτ (BSU) στο Μάντσι της Ιντιάνα (δηλαδή στο πραγματικό Middletown). Πολύ πρόσφατα μάλιστα, το 2015, δημοσιεύθηκε μια νέα μεγάλη μελέτη στην οποία ερευνούνται τα βιβλία και γενικά τα αναγνώσματα που είναι δημοφιλή στο Middletown, ώστε να υπάρξει βαθύτερη κατανόηση των αξιών και της κουλτούρας της κοινότητας αυτής (Felsenstein & Connolly, 2015).

Αυτό το μακροχρόνιο ενδιαφέρον για την έρευνα του Middletown δείχνει την τεράστια και συνεχή δύναμη των δεδομένων να δημιουργεί νέες και ενδιαφέρουσες γνώσεις για τον κόσμο. Φέρνοντας νέες οπτικές, θέτοντας δημιουργικά ερευνητικά ερωτήματα και χρησιμοποιώντας εμπεριστατωμένες ερευνητικές μεθόδους, πολλές γενιές κοινωνικών επιστημόνων μπόρεσαν να αποκαλύψουν συναρπαστικές λεπτομέρειες για ένα «συνηθισμένο» μέρος. Με άλλα λόγια, οι αναλυτές δεδομένων μπορούν να μας βοηθήσουν να δούμε το «συνηθισμένο» με νέους και συναρπαστικούς τρόπους.

1.7 Από τα «δεδομένα» στα «τεράστια δεδομένα» (*big data*)

Εκτός από τη λέξη «δεδομένα», μπορεί να έχετε συναντήσει και τη φράση *τεράστια δεδομένα*. Η φράση αυτή χρησιμοποιείται με αυξανόμενη συχνότητα τα τελευταία χρόνια και αναφέρεται στα διευρυμένου εύρους και μεγαλύτερης κλίμακας δεδομένα που είναι διαθέσιμα λόγω της τεχνολογικής προόδου. Το διαδίκτυο, τα ψηφιακά αρχεία, τα έξυπνα κινητά τηλέφωνα και τα κοινωνικά δίκτυα είναι λίγες μόνο από τις τεχνολογικές εξελίξεις που έχουν συμβάλει στην αυξημένη διαθεσιμότητα δεδομένων. Τα τεράστια δεδομένα δεν επηρεάζουν απλώς κάποιες από τις βιομηχανίες. Η επίδρασή τους βρίσκεται ήδη σε μια μεγάλη ποικιλία τομέων (δείτε, για παράδειγμα, το Liebowitz, 2013). Έχουν επηρεαστεί ακόμα και φαινομενικά παραδοσιακοί κλάδοι της οικονομίας, όπως η γεωργία, όπου οι γεωργοί χρησιμοποιούν τα τεράστια δεδομένα για να κάνουν πιο αποδοτική τη χρήση της γης τους (Doering, 2014). Η ύπαρξη τέτοιων μετασχηματισμών δείχνει ότι τώρα είναι η κατάλληλη χρονική περίοδος να μελετήσουμε σε βάθος την ανάλυση δεδομένων και ότι οι ευκαιρίες και οι ανάγκες για την απόκτηση ικανοτήτων στον τομέα αυτό θα συνεχίσουν να αυξάνονται (Mayer-Schönberger & Cukier, 2013).

Παρά τις σημαντικές προοπτικές των τεράστιων δεδομένων, κάποιοι ειδικοί με το θέμα επιστήμονες συνιστούν προσοχή και προειδοποιούν για την ανάγκη να εντοπίσουμε με πιο ουσιαστικό τρόπο τους περιορισμούς τους. Για παράδειγμα, ο Taleb (2013) προειδοποιεί ότι τα τεράστια δεδομένα απλώς δημιουργούν μια μεγαλύτερη στοίβα άχυρα όπου πρέπει να φάξει ένας ερευνητής για να βρει τη βελόνα της σημαντικής πληροφορίας. Με βάση την ιδέα αυτή, θα μπορούσαμε να υποστηρίξουμε ότι τα τεράστια δεδομένα κάνουν την ανάλυση πιο περίπλοκη και κουραστική αντί να τη βελτιστοποιούν. Επιπλέον, παρόλο που τα τεράστια δεδομένα μπορούν να εντοπίσουν σχέσεις ή μοτίβα που δεν μπορούν να βρεθούν σε μικρότερα σύνολα δεδομένων, δεν μπορούν να αναγνωρίσουν ποιες από αυτές τις σχέσεις και τα μοτίβα είναι πραγματικά έγκυρα και ουσιαστικά (Marcus & Davis, 2014). Τέλος, εφόσον δίνουν τη δυνατότητα σε έναν αναλυτή να εντοπίζει περισσότερες πιθανές σχέσεις, τα μεγαλύτερα σύνολα δεδομένων αυξάνουν την πιθανότητα να εντοπιστούν ανύπαρκτες σχέσεις, δηλαδή μοτίβα που είναι περισσότερο αποτέλεσμα της τύχης και όχι πραγματικά χαρακτηριστικά των δεδομένων.

Παράλληλα με αυτές τις σκέψεις για την επίδραση των τεράστιων δεδομένων στην ανάλυση δεδομένων, η εξάπλωσή τους έχει δημιουργήσει ευρέως διαδεδομένες ανησυχίες για την ιδιωτικότητα αναφορικά με τον τρόπο που οι εταιρείες και οι κυβερνήσεις συλλέγουν, αποθηκεύουν και χρησιμοποιούν τα προσωπικά δεδομένα. Οι ανησυχίες αυτές για την ιδιωτικότητα και την τεχνολογία δεν είναι καινούργιες. Τη δεκαετία του 1960 είχαν εκφραστεί παρόμοιοι φόβοι όταν η ομοσπονδιακή κυβέρνηση των ΗΠΑ είχε αρχίσει να χρησιμοποιεί μεγάλους κεντρικούς υπολογιστές για να επεξεργάζεται συγκεκριμένα δεδομένα σχετικά με τις επιστροφές φόρου (Lohr, 2013). Ωστόσο, σήμερα η χρήση των μέσων κοινωνικής δικτύωσης, τα έξυπνα κινητά τηλέφωνα, η τεχνολογία GPS και άλλες εφαρμογές που σχετίζονται με τα τεράστια δεδομένα έχουν αυξήσει την ποσότητα των προσωπικών δεδομένων που μοιραζόμαστε οικειοθελώς ή όχι με τους παρόχους τεχνολογίας και άλλους φορείς. Οι εφαρμογές αυτές έχουν προκαλέσει συζητήσεις και συστάσεις για κανόνες που να αφορούν την ιδιωτικότητα των δεδομένων.

Στις ΗΠΑ, για παράδειγμα, τον Ιανουάριο του 2014 ο πρώην πρόεδρος Μπαράκ Ομπάμα προώθησε μια επανεξέταση του ζητήματος των τεράστιων δεδομένων και των ζητημάτων περί ιδιωτικότητας. Στα μέλη της επιτροπής επανεξέτασης περιλαμβάνονταν ο Υπουργός Εμπορίου, ο Υπουργός Ενέργειας και ο επιστημονικός σύμβουλος του Προέδρου (Podesta, 2014). Οι προτάσεις της επιτροπής περιλάμβαναν την άποψη ότι τα άτομα «δικαιούνται ξεκάθαρες, κατανοητές, εύλογες αρχές για το πώς χρησιμοποιούνται οι προσωπικές τους πληροφορίες την εποχή των τεράστιων δεδομένων», όπως και μια πρόταση να υιοθετηθούν «συγκεκριμένες εθνικές αρχές για την παραβίαση των δεδομένων» ώστε να ελέγχονται οι μεγάλης κλίμακας παραβιάσεις των προσωπικών δεδομένων (Podesta, 2014).

ΔΙΑΛΕΙΜΜΑ ΓΙΑ ΣΚΕΨΗ

Είχατε ακούσει τη φράση «τεράστια δεδομένα» πριν διαβάσετε το παρόν βιβλίο; Θεωρείτε ότι η εξάπλωση των τεράστιων δεδομένων είναι συναρπαστική, τρομακτική ή τίποτα από τα δύο; Πέρα από τις ανησυχίες που μόλις περιγράψαμε, προβλέπετε ότι θα προκύψουν και άλλα πιθανά προβλήματα αναφορικά με τα τεράστια δεδομένα;

1.8 Συμπερασματικές σκέψεις

Παρόλο που ο τίτλος του παρόντος κεφαλαίου είναι «Εισαγωγή στα δεδομένα», στην πραγματικότητα τα δεδομένα δεν χρειάζονται καμία εισαγωγή. Στην καθημερινότητά μας, τα δεδομένα μάς κατακλύζουν ανά πάσα στιγμή. Επίσης, δημιουργούμε διαρκώς νέα δεδομένα όταν πλοηγούμαστε στο διαδίκτυο και επιλέγουμε υπερσυνδέσμους. Στην πραγματικότητα, σε έναν κόσμο όπου η τεχνολογία βρίσκεται παντού, είναι σχεδόν αδύνατο να αποφύγουμε την επαφή με δεδομένα.

Με δεδομένη την πανταχού παρουσία των δεδομένων, η ικανότητά μας να τα αντιλαμβανόμαστε, και συνεπώς να καταλαβαίνουμε τι πραγματικά σημαίνουν τα δεδομένα αυτά, αποτελεί μια σημαντική δεξιότητα. Ο στόχος του παρόντος εγχειριδίου είναι να σας βοηθήσει να αναπτύξετε τέτοιες δεξιότητες. Στα επόμενα κεφάλαια θα ερευνήσουμε πολλά διαφορετικά είδη δεδομένων, από αριθμούς και στατιστικά στοιχεία μέχρι λέξεις και ανθρώπινες αλληλεπιδράσεις, και θα εξετάσουμε τις τεχνικές που μπορείτε να χρησιμοποιήσετε για να αναλύετε τις διάφορες μορφές δεδομένων. Θα διερευνήσουμε τα οφέλη της ανάλυσης δεδομένων, αλλά και τα μειονεκτήματα και τους περιορισμούς της. Όπως είδαμε μόλις κατά την εξέταση του φαινομένου των τεράστιων δεδομένων, η ανάλυση δεδομένων μπορεί να θέσει αρκετές προκλήσεις σε ηθικό επίπεδο, τις οποίες καλείται να αντιμετωπίσει ένας νεόκοπος αναλυτής δεδομένων.

1.9 Ανακεφαλαίωση

- Τα δεδομένα μάς κατακλύζουν και μας επιτρέπουν να κάνουμε ανακαλύψεις στις οποίες δεν θα μπορούσαμε να οδηγηθούμε μόνο με τη διαίσθησή μας ή την κοινή λογική. Ο στόχος του παρόντος βιβλίου είναι να αναπτύξετε τις ικανότητές σας στην ανάλυση δεδομένων και να καταφέρετε να τιθασεύσετε τη δύναμη των δεδομένων.
- Μπορούμε να εντοπίσουμε τη δύναμη των δεδομένων στον πολιτικό αντίκτυπο που είχαν τα δεδομένα τα οποία συλλέχθηκαν για το συμβάν της ατμοσφαιρικής ρύπανσης στο Λονδίνο του 1952, στη διαχρονική δημοφιλή των ποιοτικών παρατηρήσεων του Alexis de Tocqueville για τις ΗΠΑ, αλλά και στον επίκαιρο μέχρι σήμερα Δείκτη Μάζας Σώματος (ΔΜΣ), έναν στατιστικό δείκτη που προέκυψε από το έργο του αναλυτή δεδομένων Adolphe Quetelet τον 19ο αιώνα.

- Οι ειδικοί ως προς το θέμα επιστήμονες έχουν προτείνει μια ποικιλία ορισμών για τον όρο «δεδομένα».
- Μια από τις πιο συναρπαστικές πρόσφατες εφαρμογές στην ανάλυση δεδομένων αφορά την ανάδυση των τεράστιων δεδομένων. Ωστόσο, το φαινόμενο αυτό εγείρει ανησυχίες τόσο για την πιθανή επίδρασή του στην ανάλυση δεδομένων όσο και για την πιθανή επίδρασή του στην ιδιωτικότητα.

1.10 Προτάσεις για διάβασμα

Θέλετε να μάθετε περισσότερα για την αιθαλομίχλη του Λονδίνου και τον ρόλο των δεδομένων στην αντιμετώπιση της ατμοσφαιρικής μόλυνσης; Ρίξτε μια ματιά στην παρακάτω συναρπαστική καταγραφή:

Thorsheim, P. (2006). *Inventing pollution: Coal, smoke, and culture in Britain since 1800*. Athens, OH: Ohio University Press.

Τα εποχικά μοτίβα στους δείκτες εγκληματικότητας είναι συναρπαστικά και δείχνουν την ικανότητα της ανάλυσης δεδομένων να εντοπίζει σχέσεις μεγάλης κλίμακας. Δείτε ένα παράδειγμα τέτοιας ενδιαφέρουσας και εξελιγμένης ανάλυσης στο εξής:

McDowall, D., Loftin, C., & Pate, M. (2012). Seasonal cycles in crime, and their variability. *Journal of Quantitative Criminology*, 28, 389-410.

Στο κεφάλαιο αυτό δείξαμε γιατί θεωρείται ο Adolphe Quetelet ένας από τους πιο καινοτόμους και πρωτοπόρους στοχαστές στο πεδίο της ανάλυσης δεδομένων. Μπορείτε να κάνετε τη δική σας έρευνα για το έργο του. Παρόλο που κάποιες από τις ιδέες του είναι ξεπερασμένες, παραμένει ενδιαφέρον να παρακολουθήσετε το έργο ενός αναλυτή δεδομένων ο οποίος καταπιάνεται με κοινωνικά φαινόμενα που δεν είχαν μελετηθεί μέχρι τότε με συστηματικό τρόπο. Μπορείτε να βρείτε την αρχική μετάφραση στα αγγλικά του έργου ορόσημο του Quetelet για τη σχέση ανάμεσα στο ύψος και το βάρος στο εξής:

Quetelet, L. A. J. (1842). *A treatise on man and the development of his faculties*. Edinburgh: William and Robert Chambers. (Reprinted in: Hawkins, F. B. (1973). *Comparative statistics in the 19th century*. Farnborough: Gregg International Publishers.)