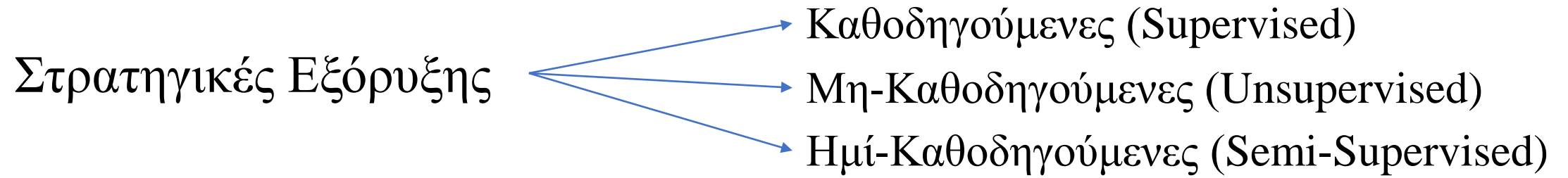


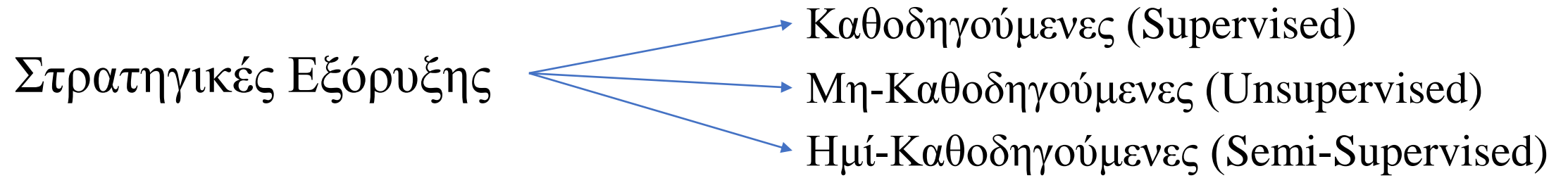
# **Μάθημα 3**

**ΕΞΟΥΣΙΑ ΔΕΔΟΜΕΝΩΝ ΑΠΟ ΨΗΦΙΑΚΟ  
ΚΑΙ ΔΙΑΔΙΚΤΥΑΚΟ ΠΕΡΙΕΧΟΜΕΝΟ**

# Στρατηγικές Εξόρυξης Πληροφορίας από Δεδομένα



# Στρατηγικές Εξόρυξης Πληροφορίας από Δεδομένα



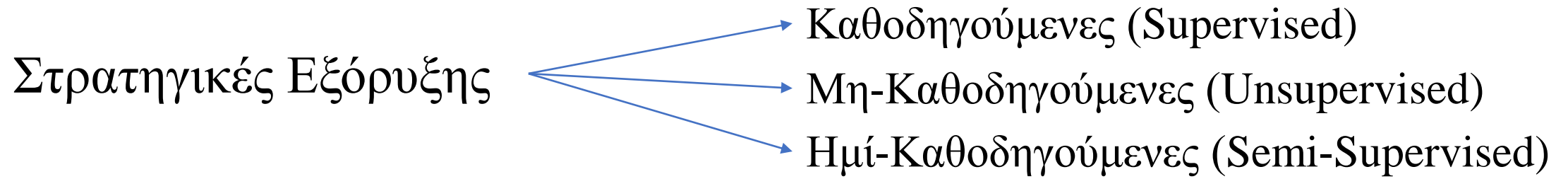
- **Καθοδηγούμενες (Supervised)**

- Χρησιμοποιείται επιπλέον πληροφορία που υπάρχει στα δεδομένα, όπως π.χ.

- ❖ Τα δεδομένα εξόδου σε ένα σύνολο δεδομένων εισόδου-εξόδου

- ❖ Η γνώση ενός ειδικού για τα εν λόγω δεδομένα

# Στρατηγικές Εξόρυξης Πληροφορίας από Δεδομένα



- **Καθοδηγούμενες (Supervised)**

- Χρησιμοποιείται επιπλέον πληροφορία που υπάρχει στα δεδομένα, όπως π.χ.
  - ❖ Τα δεδομένα εξόδου σε ένα σύνολο δεδομένων εισόδου-εξόδου
  - ❖ Η γνώση ενός ειδικού για τα εν λόγω δεδομένα
- Αντιπροσωπευτική μεθοδολογία είναι η Ανάλυση Παλινδρόμησης, η οποία υλοποιείται με Νευρωνικά Δίκτυα

# Στρατηγικές Εξόρυξης Πληροφορίας από Δεδομένα

- **Μη-Καθοδηγούμενες (Unsupervised)**
  - Η εξόρυξη της γνώσης γίνεται μόνο από την πληροφορία που περιέχουν τα δεδομένα
  - Αντιπροσωπευτική μέθοδος είναι η Συσταδοποίηση, η οποία υλοποιείται με εξειδικευμένους αλγόριθμους όπως ο αλγόριθμος k-Means και Fuzzy K-Means

# Στρατηγικές Εξόρυξης Πληροφορίας από Δεδομένα

- **Μη-Καθοδηγούμενες (Unsupervised)**
  - Η εξόρυξη της γνώσης γίνεται μόνο από την πληροφορία που περιέχουν τα δεδομένα
  - Αντιπροσωπευτική μέθοδος είναι η Συσταδοποίηση, η οποία υλοποιείται με εξειδικευμένους αλγόριθμους όπως ο αλγόριθμος k-Means και Fuzzy K-Means
- **Ημικαθοδηγούμενες (Semi-Supervised)**
  - Οι μέθοδοι αυτής της κατηγορίας είναι συνδυασμοί των άλλων δύο μεθόδων

# Βάσεις Δεδομένων

## Δεδομένα

- Οντότητες που συλλέγουμε για να περιγράψουμε μία πραγματική κατάσταση/πρόβλημα
- Περιέχουν συγκεκριμένη πληροφορία για μία κατάσταση/πρόβλημα, η οποία πληροφορία περιγράφεται πλήρως από έναν αριθμό ιδιοτήτων που χαρακτηρίζουν τα εν λόγω δεδομένα

# Βάσεις Δεδομένων

## Δεδομένα

- Οντότητες που συλλέγουμε για να περιγράψουμε μία πραγματική κατάσταση/πρόβλημα
- Περιέχουν συγκεκριμένη πληροφορία για μία κατάσταση/πρόβλημα, η οποία πληροφορία περιγράφεται πλήρως από έναν αριθμό ιδιοτήτων που χαρακτηρίζουν τα εν λόγω δεδομένα

## Βάση Δεδομένων

- Είναι ένας τρόπος αποθήκευσης των δεδομένων με βάση τις ιδιότητες που χαρακτηρίζουν τα δεδομένα
- Τυπική μορφή είναι η μορφή πίνακα, όπου οι ιδιότητες κωδικοποιούνται στις στήλες και τα δεδομένα στις γραμμές



# Βάσεις Δεδομένων: Σχήμα Βάσης

A/A	Ιδιότητα 1	Ιδιότητα 2	.....	Ιδιότητα M
1				
2				
3				
....	.....	.....	....	....
N				

# Βάσεις Δεδομένων: Σχήμα Βάσης

Ιδιότητες

Δεδομένα  
(Πλειάδες)

The diagram illustrates a database table structure. A central table has five columns and five rows. The columns are labeled 'A/A', 'Ιδιότητα 1', 'Ιδιότητα 2', '.....', and 'Ιδιότητα M'. The rows are labeled '1', '2', '3', '....', and 'N'. To the left of the table, the text 'Δεδομένα (Πλειάδες)' is connected to the rows by four blue arrows pointing to the first four rows. Above the table, the text 'Ιδιότητες' is connected to the columns by three blue arrows pointing to the first, second, and fifth columns.

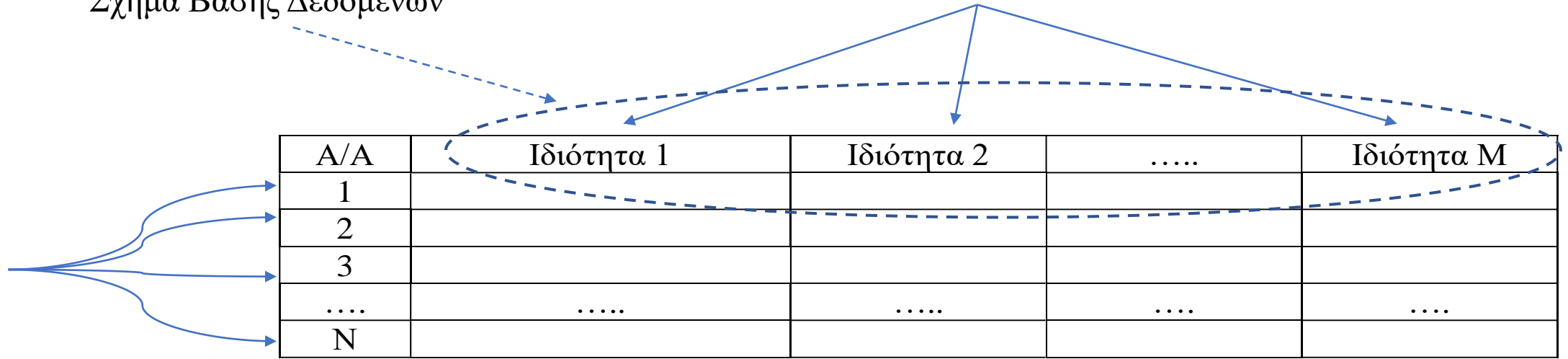
A/A	Ιδιότητα 1	Ιδιότητα 2	.....	Ιδιότητα M
1				
2				
3				
....	.....	.....	....	....
N				

# Βάσεις Δεδομένων: Σχήμα Βάσης

Σχήμα Βάσης Δεδομένων

Ιδιότητες

Δεδομένα  
(Πλειάδες)

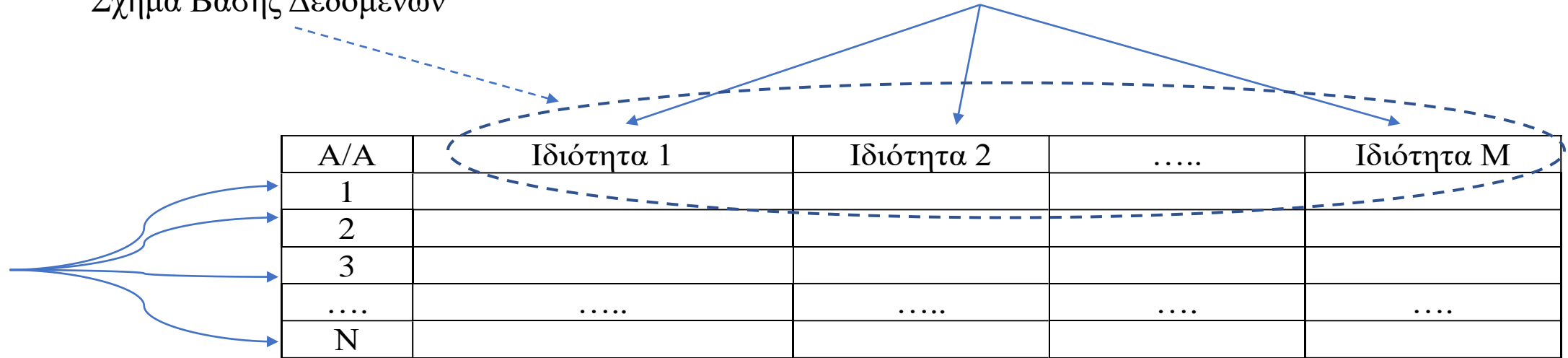


# Βάσεις Δεδομένων: Σχήμα Βάσης

Σχήμα Βάσης Δεδομένων

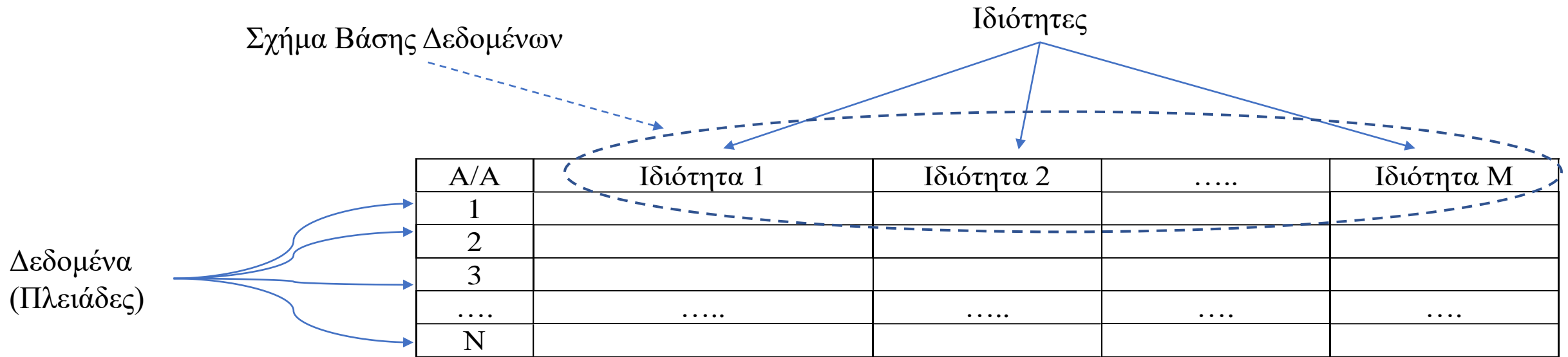
Ιδιότητες

Δεδομένα  
(Πλειάδες)



**Ιδιότητα:** Μπορεί να πάρει ένα σύνολο τιμών που ονομάζεται Πεδίο Ορισμού. Με βάση το τι εκφράζει/περιγράφει η Ιδιότητα το σύνολο τιμών της μπορεί να είναι

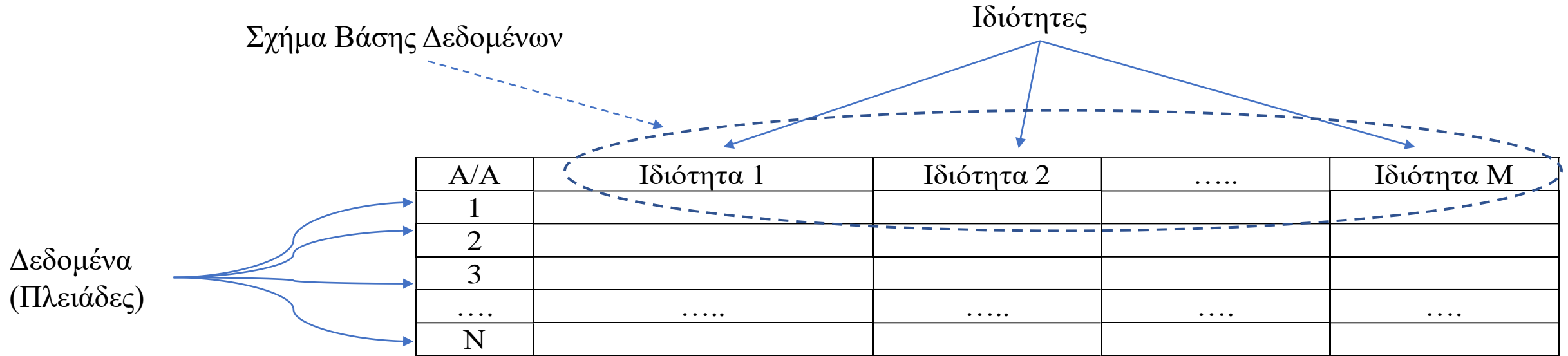
# Βάσεις Δεδομένων: Σχήμα Βάσης



**Ιδιότητα:** Μπορεί να πάρει ένα σύνολο τιμών που ονομάζεται Πεδίο Ορισμού. Με βάση το τι εκφράζει/περιγράφει η Ιδιότητα το σύνολο τιμών της μπορεί να είναι

- *Αριθμητικές τιμές:* Αριθμοί που ποσοτικοποιούν αυτό που περιγράφει η Ιδιότητα (π.χ. το ύψος ενός ανθρώπου είναι πραγματικοί αριθμοί σε μέτρα που ανήκουν στο διάστημα. Σε αυτή την περίπτωση τα δεδομένα λέγονται **Αριθμητικά Δεδομένα (Numerical Data)**)

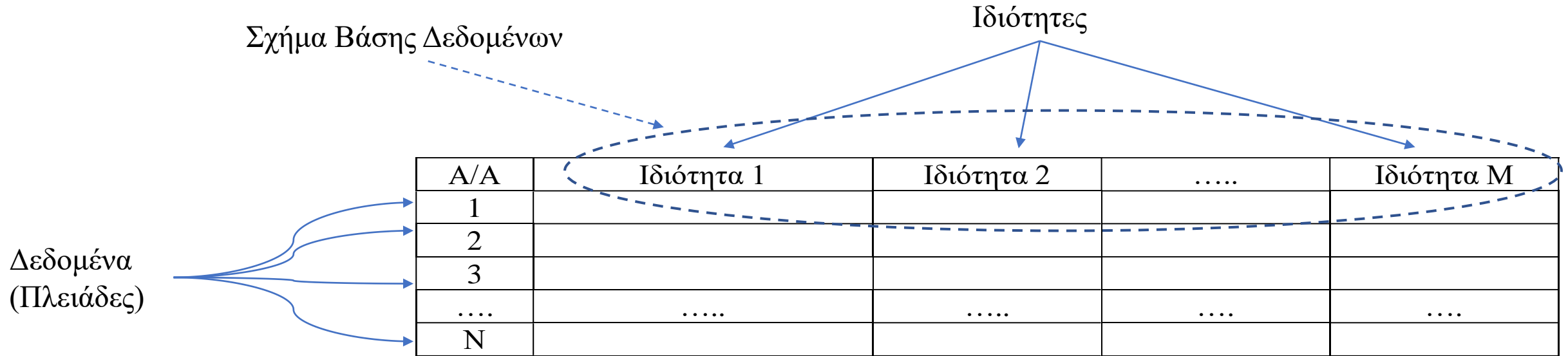
# Βάσεις Δεδομένων: Σχήμα Βάσης



**Ιδιότητα:** Μπορεί να πάρει ένα σύνολο τιμών που ονομάζεται Πεδίο Ορισμού. Με βάση το τι εκφράζει/περιγράφει η Ιδιότητα το σύνολο τιμών της μπορεί να είναι

- *Αριθμητικές τιμές:* Αριθμοί που ποσοτικοποιούν αυτό που περιγράφει η Ιδιότητα (π.χ. το ύψος ενός ανθρώπου είναι πραγματικοί αριθμοί σε μέτρα που ανήκουν στο διάστημα. Σε αυτή την περίπτωση τα δεδομένα λέγονται **Αριθμητικά Δεδομένα (Numerical Data)**)
- *Σύμβολα/Λέξεις:* Εξαρτάται από την φύση της Ιδιότητας

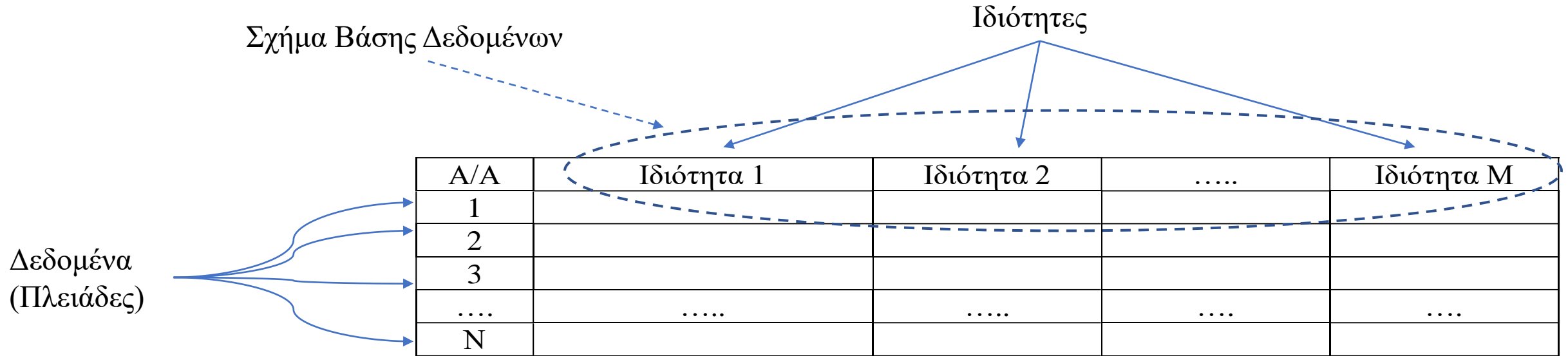
# Βάσεις Δεδομένων: Σχήμα Βάσης



**Ιδιότητα:** Μπορεί να πάρει ένα σύνολο τιμών που ονομάζεται Πεδίο Ορισμού. Με βάση το τι εκφράζει/περιγράφει η Ιδιότητα το σύνολο τιμών της μπορεί να είναι

- *Αριθμητικές τιμές:* Αριθμοί που ποσοτικοποιούν αυτό που περιγράφει η Ιδιότητα (π.χ. το ύψος ενός ανθρώπου είναι πραγματικοί αριθμοί σε μέτρα που ανήκουν στο διάστημα. Σε αυτή την περίπτωση τα δεδομένα λέγονται **Αριθμητικά Δεδομένα (Numerical Data)**)
- *Σύμβολα/Λέξεις:* Εξαρτάται από την φύση της Ιδιότητας
- *Κατηγορίες/Κλάσεις:* Οι τιμές υποδηλώνουν μία κατηγορία. Σε αυτή την περίπτωση τα δεδομένα ονομάζονται **Κατηγορικά Δεδομένα (Categorical Data)**

# Βάσεις Δεδομένων: Σχήμα Βάσης



**Ιδιότητα:** Μπορεί να πάρει ένα σύνολο τιμών που ονομάζεται Πεδίο Ορισμού. Με βάση το τι εκφράζει/περιγράφει η Ιδιότητα το σύνολο τιμών της μπορεί να είναι

- *Αριθμητικές τιμές:* Αριθμοί που ποσοτικοποιούν αυτό που περιγράφει η Ιδιότητα (π.χ. το ύψος ενός ανθρώπου είναι πραγματικοί αριθμοί σε μέτρα που ανήκουν στο διάστημα. Σε αυτή την περίπτωση τα δεδομένα λέγονται **Αριθμητικά Δεδομένα (Numerical Data)**)
- *Σύμβολα/Λέξεις:* Εξαρτάται από την φύση της Ιδιότητας
- *Κατηγορίες/Κλάσεις:* Οι τιμές υποδηλώνουν μία κατηγορία. Σε αυτή την περίπτωση τα δεδομένα ονομάζονται **Κατηγορικά Δεδομένα (Categorical Data)**
- *Πολυμέσα:* Εικόνες, Βίντεο, Γραφικά



# Βάσεις Δεδομένων: Παραδείγματα

## Βαθμολογία Μαθήματος Εξόρυξης Δεδομένων

A/A	Επίθετο	Όνομα	ΑΜ	Εξάμηνο	Βαθμός
1	Γεώργιος	Τσεκούρας	1312017043	7	6
2	Μαρία	Απτέρη	1312016020	9	8
3	Δήμητρα	Γκιώνη	1312017011	7	6
4	Γρηγόριος	Μηταράκης	1312017031	7	5
5	Αθηνά	Μήτσιου	1312015056	11	7
...	...	...	...	...	...

# Βάσεις Δεδομένων: Παραδείγματα

## Βαθμολογία Μαθήματος Εξόρυξης Δεδομένων

A/A	Επίθετο	Όνομα	ΑΜ	Εξάμηνο	Βαθμός
1	Γεώργιος	Τσεκούρας	1312017043	7	6
2	Μαρία	Απτέρη	1312016020	9	8
3	Δήμητρα	Γκιώνη	1312017011	7	6
4	Γρηγόριος	Μηταράκης	1312017031	7	5
5	Αθηνά	Μήτσιου	1312015056	11	7
...	...	...	...	...	...

Ιδιότητα 1- **Όνομα**: Οι τιμές που παίρνει είναι τα μικρά ονόματα (Σύμβολα/Λέξεις). Το Πεδίο Ορισμού είναι όλα τα μικρά ονόματα. Σε κάθε δεδομένο αντιστοιχεί μόνο ένα όνομα και άρα μόνο μία τιμή

# Βάσεις Δεδομένων: Παραδείγματα

## Βαθμολογία Μαθήματος Εξόρυξης Δεδομένων

A/A	Επίθετο	Όνομα	ΑΜ	Εξάμηνο	Βαθμός
1	Γεώργιος	Τσεκούρας	1312017043	7	6
2	Μαρία	Απτέρη	1312016020	9	8
3	Δήμητρα	Γκιώνη	1312017011	7	6
4	Γρηγόριος	Μηταράκης	1312017031	7	5
5	Αθηνά	Μήτσιου	1312015056	11	7
...	...	...	...	...	...

Ιδιότητα 1- **Όνομα**: Οι τιμές που παίρνει είναι τα μικρά ονόματα (Σύμβολα/Λέξεις). Το Πεδίο Ορισμού είναι όλα τα μικρά ονόματα. Σε κάθε δεδομένο αντιστοιχεί μόνο ένα όνομα και άρα μόνο μία τιμή

Ιδιότητα 2- **Επίθετο**: Οι τιμές που παίρνει είναι τα επίθετα (Σύμβολα/Λέξεις). Το Πεδίο Ορισμού είναι όλα τα επίθετα. Σε κάθε δεδομένο αντιστοιχεί μόνο ένα επίθετο και άρα μόνο μία τιμή

# Βάσεις Δεδομένων: Παραδείγματα

## Βαθμολογία Μαθήματος Εξόρυξης Δεδομένων

A/A	Επίθετο	Όνομα	ΑΜ	Εξάμηνο	Βαθμός
1	Γεώργιος	Τσεκούρας	1312017043	7	6
2	Μαρία	Απτέρη	1312016020	9	8
3	Δήμητρα	Γκιώνη	1312017011	7	6
4	Γρηγόριος	Μηταράκης	1312017031	7	5
5	Αθηνά	Μήτσιου	1312015056	11	7
...	...	...	...	...	...

Ιδιότητα 1- **Όνομα**: Οι τιμές που παίρνει είναι τα μικρά ονόματα (Σύμβολα/Λέξεις). Το Πεδίο Ορισμού είναι όλα τα μικρά ονόματα. Σε κάθε δεδομένο αντιστοιχεί μόνο ένα όνομα και άρα μόνο μία τιμή

Ιδιότητα 2- **Επίθετο**: Οι τιμές που παίρνει είναι τα επίθετα (Σύμβολα/Λέξεις). Το Πεδίο Ορισμού είναι όλα τα επίθετα. Σε κάθε δεδομένο αντιστοιχεί μόνο ένα επίθετο και άρα μόνο μία τιμή

Ιδιότητα 3- **ΑΜ**: Οι τιμές που παίρνει είναι ακέραιοι αριθμοί (Αριθμητικά Δεδομένα). Το Πεδίο Ορισμού είναι όλοι οι ακέραιοι αριθμοί. Σε κάθε δεδομένο αντιστοιχεί μόνο ένας αριθμός και άρα μόνο μία τιμή

# Βάσεις Δεδομένων: Παραδείγματα

## Βαθμολογία Μαθήματος Εξόρυξης Δεδομένων

A/A	Επίθετο	Όνομα	ΑΜ	Εξάμηνο	Βαθμός
1	Γεώργιος	Τσεκούρας	1312017043	7	6
2	Μαρία	Απτέρη	1312016020	9	8
3	Δήμητρα	Γκιώνη	1312017011	7	6
4	Γρηγόριος	Μηταράκης	1312017031	7	5
5	Αθηνά	Μήτσιου	1312015056	11	7
...	...	...	...	...	...

Ιδιότητα 1- **Όνομα**: Οι τιμές που παίρνει είναι τα μικρά ονόματα (Σύμβολα/Λέξεις). Το Πεδίο Ορισμού είναι όλα τα μικρά ονόματα. Σε κάθε δεδομένο αντιστοιχεί μόνο ένα όνομα και άρα μόνο μία τιμή

Ιδιότητα 2- **Επίθετο**: Οι τιμές που παίρνει είναι τα επίθετα (Σύμβολα/Λέξεις). Το Πεδίο Ορισμού είναι όλα τα επίθετα. Σε κάθε δεδομένο αντιστοιχεί μόνο ένα επίθετο και άρα μόνο μία τιμή

Ιδιότητα 3- **ΑΜ**: Οι τιμές που παίρνει είναι ακέραιοι αριθμοί (Αριθμητικά Δεδομένα). Το Πεδίο Ορισμού είναι όλοι οι ακέραιοι αριθμοί. Σε κάθε δεδομένο αντιστοιχεί μόνο ένας αριθμός και άρα μόνο μία τιμή

Ιδιότητα 4- **Εξάμηνο**: Οι τιμές που παίρνει είναι ακέραιοι αριθμοί (Αριθμητικά Δεδομένα). Το Πεδίο Ορισμού είναι όλοι οι ακέραιοι αριθμοί (???). Σε κάθε δεδομένο αντιστοιχεί μόνο ένας αριθμός και άρα μόνο μία τιμή

# Βάσεις Δεδομένων: Παραδείγματα

## Βαθμολογία Μαθήματος Εξόρυξης Δεδομένων

A/A	Επίθετο	Όνομα	ΑΜ	Εξάμηνο	Βαθμός
1	Γεώργιος	Τσεκούρας	1312017043	7	6
2	Μαρία	Απτέρη	1312016020	9	8
3	Δήμητρα	Γκιώνη	1312017011	7	6
4	Γρηγόριος	Μηταράκης	1312017031	7	5
5	Αθηνά	Μήτσιου	1312015056	11	7
...	...	...	...	...	...

Ιδιότητα 1- **Όνομα**: Οι τιμές που παίρνει είναι τα μικρά ονόματα (Σύμβολα/Λέξεις). Το Πεδίο Ορισμού είναι όλα τα μικρά ονόματα. Σε κάθε δεδομένο αντιστοιχεί μόνο ένα όνομα και άρα μόνο μία τιμή

Ιδιότητα 2- **Επίθετο**: Οι τιμές που παίρνει είναι τα επίθετα (Σύμβολα/Λέξεις). Το Πεδίο Ορισμού είναι όλα τα επίθετα. Σε κάθε δεδομένο αντιστοιχεί μόνο ένα επίθετο και άρα μόνο μία τιμή

Ιδιότητα 3- **ΑΜ**: Οι τιμές που παίρνει είναι ακέραιοι αριθμοί (Αριθμητικά Δεδομένα). Το Πεδίο Ορισμού είναι όλοι οι ακέραιοι αριθμοί. Σε κάθε δεδομένο αντιστοιχεί μόνο ένας αριθμός και άρα μόνο μία τιμή

Ιδιότητα 4- **Εξάμηνο**: Οι τιμές που παίρνει είναι ακέραιοι αριθμοί (Αριθμητικά Δεδομένα). Το Πεδίο Ορισμού είναι όλοι οι ακέραιοι αριθμοί (???). Σε κάθε δεδομένο αντιστοιχεί μόνο ένας αριθμός και άρα μόνο μία τιμή

Ιδιότητα 5- **Βαθμός**: Οι τιμές που παίρνει είναι πραγματικοί αριθμοί (Αριθμητικά Δεδομένα). Το Πεδίο Ορισμού είναι όλοι οι πραγματικοί αριθμοί στο διάστημα  $[0, 10]$ . Σε κάθε δεδομένο αντιστοιχεί μόνο ένας αριθμός και άρα μόνο μία τιμή

# Βάσεις Δεδομένων: Παραδείγματα

## Βαθμολογία Μαθήματος Εξόρυξης Δεδομένων

A/A	Επίθετο	Όνομα	ΑΜ	Εξάμηνο	Βαθμός
1	Γεώργιος	Τσεκούρας	1312017043	7	6
2	Μαρία	Απτέρη	1312016020	9	8
3	Δήμητρα	Γκιώνη	1312017011	7	6
4	Γρηγόριος	Μηταράκης	1312017031	7	5
5	Αθηνά	Μήτσιου	1312015056	11	7
...	...	...	...	...	...

Ιδιότητα 1- **Όνομα**: Οι τιμές που παίρνει είναι τα μικρά ονόματα (Σύμβολα/Λέξεις). Το Πεδίο Ορισμού είναι όλα τα μικρά ονόματα. Σε κάθε δεδομένο αντιστοιχεί μόνο ένα όνομα και άρα μόνο μία τιμή

Ιδιότητα 2- **Επίθετο**: Οι τιμές που παίρνει είναι τα επίθετα (Σύμβολα/Λέξεις). Το Πεδίο Ορισμού είναι όλα τα επίθετα. Σε κάθε δεδομένο αντιστοιχεί μόνο ένα επίθετο και άρα μόνο μία τιμή

Ιδιότητα 3- **ΑΜ**: Οι τιμές που παίρνει είναι ακέραιοι αριθμοί (Αριθμητικά Δεδομένα). Το Πεδίο Ορισμού είναι όλοι οι ακέραιοι αριθμοί. Σε κάθε δεδομένο αντιστοιχεί μόνο ένας αριθμός και άρα μόνο μία τιμή

Ιδιότητα 4- **Εξάμηνο**: Οι τιμές που παίρνει είναι ακέραιοι αριθμοί (Αριθμητικά Δεδομένα). Το Πεδίο Ορισμού είναι όλοι οι ακέραιοι αριθμοί (???). Σε κάθε δεδομένο αντιστοιχεί μόνο ένας αριθμός και άρα μόνο μία τιμή

Ιδιότητα 5- **Βαθμός**: Οι τιμές που παίρνει είναι πραγματικοί αριθμοί (Αριθμητικά Δεδομένα). Το Πεδίο Ορισμού είναι όλοι οι πραγματικοί αριθμοί στο διάστημα  $[0, 10]$ . Σε κάθε δεδομένο αντιστοιχεί μόνο ένας αριθμός και άρα μόνο μία τιμή

**Παρατήρηση**: Δύο ή περισσότερες ιδιότητες μπορεί να έχουν το ίδιο Πεδίο Ορισμού

# Βάσεις Δεδομένων: Παραδείγματα

## Αξιολόγηση Αυτοκινήτου

A/A	Μοντέλο	Πωλήσεις	Κόστος Συντήρησης	Αριθμός Πορτών	Αριθμός Επιβατών	Πορτμπαγκάζ	Ασφάλεια
1	Toyota Auris	Πάρα Πολλές	Μέτριο	5	4	Μικρό	Καλή
2	Smart	Πολλές	Μεγάλο	2	2	Πολύ Μικρό	Καλή
3	Mercedes	Μέτριες	Πολύ Μεγάλο	4	4	Πολύ Μεγάλο	Πολύ Καλή
4	Fiat Panda	Πολλές	Μικρό	5	4	Μέτριο	Καλή
...							



# Βάσεις Δεδομένων: Παραδείγματα

## Αξιολόγηση Αυτοκινήτου

A/A	Μοντέλο	Πωλήσεις	Κόστος Συντήρησης	Αριθμός Πορτών	Αριθμός Επιβατών	Πορτμπαγκάζ	Ασφάλεια
1	Toyota Auris	Πάρα Πολλές	Μέτριο	5	4	Μικρό	Καλή
2	Smart	Πολλές	Μεγάλο	2	2	Πολύ Μικρό	Καλή
3	Mercedes	Μέτριες	Πολύ Μεγάλο	4	4	Πολύ Μεγάλο	Πολύ Καλή
4	Fiat Panta	Πολλές	Μικρό	5	4	Μέτριο	Καλή
....							

Ιδιότητα 1- **Μοντέλο**: Το Πεδίο Ορισμού είναι όλα τα ονόματα μοντέλων αυτοκινήτων (Σύμβολα/Λέξεις)

# Βάσεις Δεδομένων: Παραδείγματα

## Αξιολόγηση Αυτοκινήτου

A/A	Μοντέλο	Πωλήσεις	Κόστος Συντήρησης	Αριθμός Πορτών	Αριθμός Επιβατών	Πορτμπαγκάζ	Ασφάλεια
1	Toyota Auris	Πάρα Πολλές	Μέτριο	5	4	Μικρό	Καλή
2	Smart	Πολλές	Μεγάλο	2	2	Πολύ Μικρό	Καλή
3	Mercedes	Μέτριες	Πολύ Μεγάλο	4	4	Πολύ Μεγάλο	Πολύ Καλή
4	Fiat Panta	Πολλές	Μικρό	5	4	Μέτριο	Καλή
....							

Ιδιότητα 1- **Μοντέλο**: Το Πεδίο Ορισμού είναι όλα τα ονόματα μοντέλων αυτοκινήτων (Σύμβολα/Λέξεις)

Ιδιότητα 2- **Πωλήσεις**: Πεδίο Ορισμού={Πολύ Λίγες, Λίγες, Μέτριες, Πολλές, Πάρα Πολλές (είναι κατηγορίες και άρα είναι Κατηγορικά Δεδομένα)}

# Βάσεις Δεδομένων: Παραδείγματα

## Αξιολόγηση Αυτοκινήτου

A/A	Μοντέλο	Πωλήσεις	Κόστος Συντήρησης	Αριθμός Πορτών	Αριθμός Επιβατών	Πορτμπαγκάζ	Ασφάλεια
1	Toyota Auris	Πάρα Πολλές	Μέτριο	5	4	Μικρό	Καλή
2	Smart	Πολλές	Μεγάλο	2	2	Πολύ Μικρό	Καλή
3	Mercedes	Μέτριες	Πολύ Μεγάλο	4	4	Πολύ Μεγάλο	Πολύ Καλή
4	Fiat Panta	Πολλές	Μικρό	5	4	Μέτριο	Καλή
....							

Ιδιότητα 1- **Μοντέλο**: Το Πεδίο Ορισμού είναι όλα τα ονόματα μοντέλων αυτοκινήτων (Σύμβολα/Λέξεις)

Ιδιότητα 2- **Πωλήσεις**: Πεδίο Ορισμού={Πολύ Λίγες, Λίγες, Μέτριες, Πολλές, Πάρα Πολλές (είναι κατηγορίες και άρα είναι Κατηγορικά Δεδομένα)}

Ιδιότητα 3- **Κόστος Συντήρησης**: Πεδίο Ορισμού={Πολύ Λίγες, Λίγες, Μέτριες, Πολλές, Πάρα Πολλές} (είναι κατηγορίες και άρα είναι Κατηγορικά Δεδομένα)

# Βάσεις Δεδομένων: Παραδείγματα

## Αξιολόγηση Αυτοκινήτου

A/A	Μοντέλο	Πωλήσεις	Κόστος Συντήρησης	Αριθμός Πορτών	Αριθμός Επιβατών	Πορτμπαγκάζ	Ασφάλεια
1	Toyota Auris	Πάρα Πολλές	Μέτριο	5	4	Μικρό	Καλή
2	Smart	Πολλές	Μεγάλο	2	2	Πολύ Μικρό	Καλή
3	Mercedes	Μέτριες	Πολύ Μεγάλο	4	4	Πολύ Μεγάλο	Πολύ Καλή
4	Fiat Panta	Πολλές	Μικρό	5	4	Μέτριο	Καλή
....							

Ιδιότητα 1- **Μοντέλο**: Το Πεδίο Ορισμού είναι όλα τα ονόματα μοντέλων αυτοκινήτων (Σύμβολα/Λέξεις)

Ιδιότητα 2- **Πωλήσεις**: Πεδίο Ορισμού={Πολύ Λίγες, Λίγες, Μέτριες, Πολλές, Πάρα Πολλές (είναι κατηγορίες και άρα είναι Κατηγορικά Δεδομένα)}

Ιδιότητα 3- **Κόστος Συντήρησης**: Πεδίο Ορισμού={Πολύ Λίγες, Λίγες, Μέτριες, Πολλές, Πάρα Πολλές} (είναι κατηγορίες και άρα είναι Κατηγορικά Δεδομένα)

Ιδιότητα 4- **Αριθμός Πορτών**: Πεδίο Ορισμού={2, 4, 5} (Ακέραιοι αριθμοί και άρα Αριθμητικά Δεδομένα.... Η ιδιότητα αυτή μπορεί όμως να λογισθεί και ως κατηγορική ιδιότητα)

# Βάσεις Δεδομένων: Παραδείγματα

## Αξιολόγηση Αυτοκινήτου

A/A	Μοντέλο	Πωλήσεις	Κόστος Συντήρησης	Αριθμός Πορτών	Αριθμός Επιβατών	Πορτμπαγκάζ	Ασφάλεια
1	Toyota Auris	Πάρα Πολλές	Μέτριο	5	4	Μικρό	Καλή
2	Smart	Πολλές	Μεγάλο	2	2	Πολύ Μικρό	Καλή
3	Mercedes	Μέτριες	Πολύ Μεγάλο	4	4	Πολύ Μεγάλο	Πολύ Καλή
4	Fiat Panta	Πολλές	Μικρό	5	4	Μέτριο	Καλή
....							

Ιδιότητα 1- **Μοντέλο**: Το Πεδίο Ορισμού είναι όλα τα ονόματα μοντέλων αυτοκινήτων (Σύμβολα/Λέξεις)

Ιδιότητα 2- **Πωλήσεις**: Πεδίο Ορισμού={Πολύ Λίγες, Λίγες, Μέτριες, Πολλές, Πάρα Πολλές (είναι κατηγορίες και άρα είναι Κατηγορικά Δεδομένα)}

Ιδιότητα 3- **Κόστος Συντήρησης**: Πεδίο Ορισμού={Πολύ Λίγες, Λίγες, Μέτριες, Πολλές, Πάρα Πολλές} (είναι κατηγορίες και άρα είναι Κατηγορικά Δεδομένα)

Ιδιότητα 4- **Αριθμός Πορτών**: Πεδίο Ορισμού={2, 4, 5} (Ακέραιοι αριθμοί και άρα Αριθμητικά Δεδομένα.... Η ιδιότητα αυτή μπορεί όμως να λογισθεί και ως κατηγορική ιδιότητα)

Ιδιότητα 5- **Αριθμός Ατόμων**: Πεδίο Ορισμού={2, 4, 5} (Ακέραιοι αριθμοί και άρα Αριθμητικά Δεδομένα.... Η ιδιότητα αυτή μπορεί όμως να λογισθεί και ως κατηγορική ιδιότητα)

# Βάσεις Δεδομένων: Παραδείγματα

## Αξιολόγηση Αυτοκινήτου

A/A	Μοντέλο	Πωλήσεις	Κόστος Συντήρησης	Αριθμός Πορτών	Αριθμός Επιβατών	Πορτμπαγκάζ	Ασφάλεια
1	Toyota Auris	Πάρα Πολλές	Μέτριο	5	4	Μικρό	Καλή
2	Smart	Πολλές	Μεγάλο	2	2	Πολύ Μικρό	Καλή
3	Mercedes	Μέτριες	Πολύ Μεγάλο	4	4	Πολύ Μεγάλο	Πολύ Καλή
4	Fiat Panta	Πολλές	Μικρό	5	4	Μέτριο	Καλή
....							

Ιδιότητα 1- **Μοντέλο**: Το Πεδίο Ορισμού είναι όλα τα ονόματα μοντέλων αυτοκινήτων (Σύμβολα/Λέξεις)

Ιδιότητα 2- **Πωλήσεις**: Πεδίο Ορισμού={Πολύ Λίγες, Λίγες, Μέτριες, Πολλές, Πάρα Πολλές (είναι κατηγορίες και άρα είναι Κατηγορικά Δεδομένα)}

Ιδιότητα 3- **Κόστος Συντήρησης**: Πεδίο Ορισμού={Πολύ Λίγες, Λίγες, Μέτριες, Πολλές, Πάρα Πολλές} (είναι κατηγορίες και άρα είναι Κατηγορικά Δεδομένα)

Ιδιότητα 4- **Αριθμός Πορτών**: Πεδίο Ορισμού={2, 4, 5} (Ακέραιοι αριθμοί και άρα Αριθμητικά Δεδομένα.... Η ιδιότητα αυτή μπορεί όμως να λογισθεί και ως κατηγορική ιδιότητα)

Ιδιότητα 5- **Αριθμός Ατόμων**: Πεδίο Ορισμού={2, 4, 5} (Ακέραιοι αριθμοί και άρα Αριθμητικά Δεδομένα.... Η ιδιότητα αυτή μπορεί όμως να λογισθεί και ως κατηγορική ιδιότητα)

Ιδιότητα 6- **Πορτμπαγκάζ**: Πεδίο Ορισμού={Πολύ Μικρό, Μικρό, Μέτριο, Μεγάλο, Πολύ Μεγάλο} (Κατηγορικά Δεδομένα)

# Βάσεις Δεδομένων: Παραδείγματα

## Αξιολόγηση Αυτοκινήτου

A/A	Μοντέλο	Πωλήσεις	Κόστος Συντήρησης	Αριθμός Πορτών	Αριθμός Επιβατών	Πορτμπαγκάζ	Ασφάλεια
1	Toyota Auris	Πάρα Πολλές	Μέτριο	5	4	Μικρό	Καλή
2	Smart	Πολλές	Μεγάλο	2	2	Πολύ Μικρό	Καλή
3	Mercedes	Μέτριες	Πολύ Μεγάλο	4	4	Πολύ Μεγάλο	Πολύ Καλή
4	Fiat Panta	Πολλές	Μικρό	5	4	Μέτριο	Καλή
....							

Ιδιότητα 1- **Μοντέλο**: Το Πεδίο Ορισμού είναι όλα τα ονόματα μοντέλων αυτοκινήτων (Σύμβολα/Λέξεις)

Ιδιότητα 2- **Πωλήσεις**: Πεδίο Ορισμού={Πολύ Λίγες, Λίγες, Μέτριες, Πολλές, Πάρα Πολλές (είναι κατηγορίες και άρα είναι Κατηγορικά Δεδομένα)}

Ιδιότητα 3- **Κόστος Συντήρησης**: Πεδίο Ορισμού={Πολύ Λίγες, Λίγες, Μέτριες, Πολλές, Πάρα Πολλές} (είναι κατηγορίες και άρα είναι Κατηγορικά Δεδομένα)

Ιδιότητα 4- **Αριθμός Πορτών**: Πεδίο Ορισμού={2, 4, 5} (Ακέραιοι αριθμοί και άρα Αριθμητικά Δεδομένα.... Η ιδιότητα αυτή μπορεί όμως να λογισθεί και ως κατηγορική ιδιότητα)

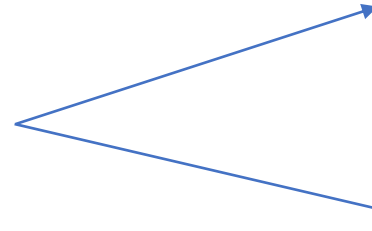
Ιδιότητα 5- **Αριθμός Ατόμων**: Πεδίο Ορισμού={2, 4, 5} (Ακέραιοι αριθμοί και άρα Αριθμητικά Δεδομένα.... Η ιδιότητα αυτή μπορεί όμως να λογισθεί και ως κατηγορική ιδιότητα)

Ιδιότητα 6- **Πορτμπαγκάζ**: Πεδίο Ορισμού={Πολύ Μικρό, Μικρό, Μέτριο, Μεγάλο, Πολύ Μεγάλο} (Κατηγορικά Δεδομένα)

Ιδιότητα 7- **Ασφάλεια**: Πεδίο Ορισμού={Πολύ κακή, Κακή, Καλή, Πολύ Καλή} (Κατηγορικά Δεδομένα)

# Είδη Κατηγορικών Δεδομένων

Είδη Δεδομένων με βάση τις τιμές  
που παίρνουν



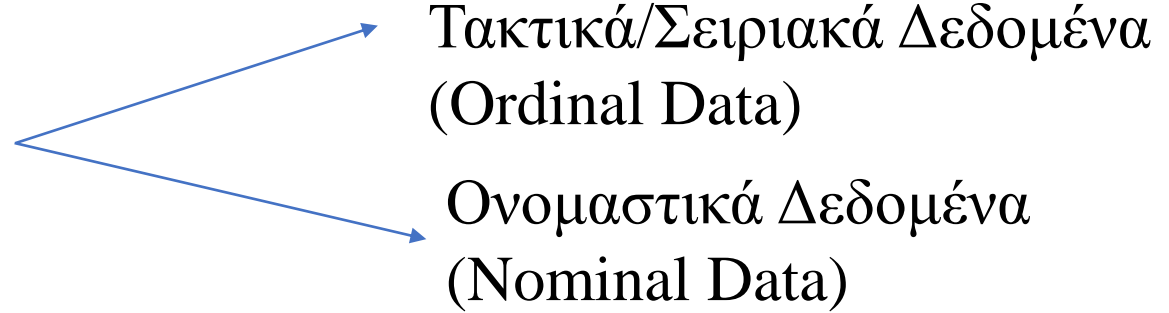
Τακτικά/Σειριακά Δεδομένα  
(Ordinal Data)

Ονομαστικά Δεδομένα  
(Nominal Data)



# Είδη Κατηγορικών Δεδομένων

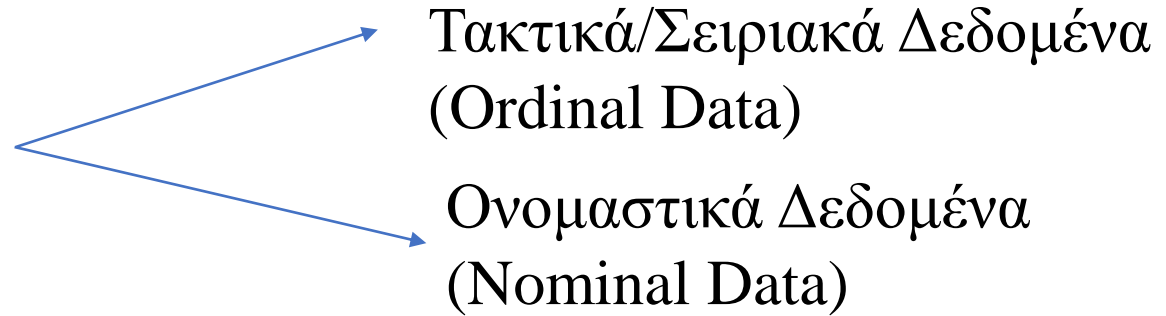
Είδη Δεδομένων με βάση τις τιμές που παίρνουν



- **Ordinal Categorical Data:** Είναι δεδομένα των οποίων οι τιμές που περιέχονται στο Πεδίο Ορισμού διατηρούν μία τάξη με την έννοια ότι δύο τιμές μπορούν να συγκριθούν μεταξύ τους και να βγάλουμε το συμπέρασμα ότι η μία είναι η μεγαλύτερη και άρα η άλλη είναι η μικρότερη. Άρα οι τιμές μπαίνουν σε μία αύξουσα σειρά από την μικρότερη στην μεγαλύτερη ή σε μία φθίνουσα σειρά από την μεγαλύτερη στην μικρότερη.

# Είδη Κατηγορικών Δεδομένων

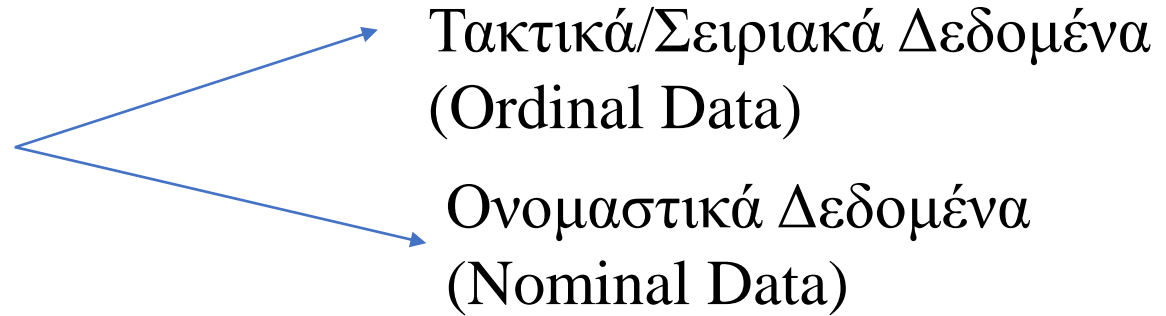
Είδη Δεδομένων με βάση τις τιμές που παίρνουν



- **Ordinal Categorical Data:** Είναι δεδομένα των οποίων οι τιμές που περιέχονται στο Πεδίο Ορισμού διατηρούν μία τάξη με την έννοια ότι δύο τιμές μπορούν να συγκριθούν μεταξύ τους και να βγάλουμε το συμπέρασμα ότι η μία είναι η μεγαλύτερη και άρα η άλλη είναι η μικρότερη. Άρα οι τιμές μπαίνουν σε μία αύξουσα σειρά από την μικρότερη στην μεγαλύτερη ή σε μία φθίνουσα σειρά από την μεγαλύτερη στην μικρότερη. Παραδείγματα:
  - Τα κατηγορικά δεδομένα που έχουν ως πεδία ορισμού παρόμοια με τα παρακάτω:
    - ✓ {Πολύ Λίγο, Λίγο, Πολύ, Πάρα Πολύ}
    - ✓ {Πολύ Νέος, Νέος, Μεσήλικας, Ηλικιωμένος, Υπερήλικας}
    - ✓ Πιθανή απάντηση σε ερώτηση {0-10, 10-20, 20-40, 40-50} κ.λ.π.

# Είδη Κατηγορικών Δεδομένων

Είδη Δεδομένων με βάση τις τιμές που παίρνουν



- **Ordinal Categorical Data:** Είναι δεδομένα των οποίων οι τιμές που περιέχονται στο Πεδίο Ορισμού διατηρούν μία τάξη με την έννοια ότι δύο τιμές μπορούν να συγκριθούν μεταξύ τους και να βγάλουμε το συμπέρασμα ότι η μία είναι η μεγαλύτερη και άρα η άλλη είναι η μικρότερη. Άρα οι τιμές μπαίνουν σε μία αύξουσα σειρά από την μικρότερη στην μεγαλύτερη ή σε μία φθίνουσα σειρά από την μεγαλύτερη στην μικρότερη. Παραδείγματα:
  - Τα κατηγορικά δεδομένα που έχουν ως πεδία ορισμού παρόμοια με τα παρακάτω:
    - ✓ {Πολύ Λίγο, Λίγο, Πολύ, Πάρα Πολύ}
    - ✓ {Πολύ Νέος, Νέος, Μεσήλικας, Ηλικιωμένος, Υπερήλικας}
    - ✓ Πιθανή απάντηση σε ερώτηση {0-10, 10-20, 20-40, 40-50} κ.λ.π.
- **Nominal Categorical Data:** Είναι δεδομένα των οποίων οι τιμές που περιέχονται στο Πεδίο Ορισμού δεν διατηρούν μία τάξη με την έννοια ότι δύο τιμές δεν μπορούν να συγκριθούν μεταξύ τους και να βγάλουμε το συμπέρασμα ποια είναι η μεγαλύτερη και ποια η μικρότερη. Άρα οι τιμές δεν μπαίνουν σε μία αύξουσα σειρά ή φθίνουσα σειρά.

# Είδη Κατηγορικών Δεδομένων

Είδη Δεδομένων με βάση τις τιμές που παίρνουν

Τακτικά/Σειριακά Δεδομένα  
(Ordinal Data)

Ονομαστικά Δεδομένα  
(Nominal Data)

- **Ordinal Categorical Data:** Είναι δεδομένα των οποίων οι τιμές που περιέχονται στο Πεδίο Ορισμού διατηρούν μία τάξη με την έννοια ότι δύο τιμές μπορούν να συγκριθούν μεταξύ τους και να βγάλουμε το συμπέρασμα ότι η μία είναι η μεγαλύτερη και άρα η άλλη είναι η μικρότερη. Άρα οι τιμές μπαίνουν σε μία αύξουσα σειρά από την μικρότερη στην μεγαλύτερη ή σε μία φθίνουσα σειρά από την μεγαλύτερη στην μικρότερη. Παραδείγματα:
  - Τα κατηγορικά δεδομένα που έχουν ως πεδία ορισμού παρόμοια με τα παρακάτω:
    - ✓ {Πολύ Λίγο, Λίγο, Πολύ, Πάρα Πολύ}
    - ✓ {Πολύ Νέος, Νέος, Μεσήλικας, Ηλικιωμένος, Υπερήλικας}
    - ✓ Πιθανή απάντηση σε ερώτηση {0-10, 10-20, 20-40, 40-50} κ.λ.π.
- **Nominal Categorical Data:** Είναι δεδομένα των οποίων οι τιμές που περιέχονται στο Πεδίο Ορισμού δεν διατηρούν μία τάξη με την έννοια ότι δύο τιμές δεν μπορούν να συγκριθούν μεταξύ τους και να βγάλουμε το συμπέρασμα ποια είναι η μεγαλύτερη και ποια η μικρότερη. Άρα οι τιμές δεν μπαίνουν σε μία αύξουσα σειρά ή φθίνουσα σειρά. Παραδείγματα:
  - Τα κατηγορικά δεδομένα που έχουν ως πεδίο ορισμού παρόμοια με τα παρακάτω:
    - ✓ {Άνδρας, Γυναίκα}: Απάντηση στην ερώτηση του Φύλου
    - ✓ {O-, O+, B-, B+, A-, A+, AB-, AB+}: Απάντηση στην ερώτηση τι ομάδα αίματος έχεις
    - ✓ {Αγρίνιο, Αθήνα, Μυτιλήνη, Θεσσαλονίκη, Βόλος, Άλλο}: Απάντηση την ερώτηση σε ποια πόλη ζεις
    - ✓ {Μαύρα, Ξανθά, Άσπρα, Καστανά, Άλλο}: Απάντηση στην ερώτηση τι χρώμα μαλλιών έχεις
    - ✓ Πολυμεσικά δεδομένα που περιέχουν οπτικοακουστικό υλικό (Εικόνες, Βίντεο)

# Τα Σύμβολα και οι Λέξεις ως Δεδομένα

Το πρόβλημα με τα Σύμβολα και τις Λέξεις είναι να βρεθεί ένας τρόπος ποσοτικοποίησης τους. Αυτό μπορεί να γίνει ως εξής:

# Τα Σύμβολα και οι Λέξεις ως Δεδομένα

Το πρόβλημα με τα Σύμβολα και τις Λέξεις είναι να βρεθεί ένας τρόπος ποσοτικοποίησης τους. Αυτό μπορεί να γίνει ως εξής:

## Παράδειγμα

Έστω ότι θέλουμε να περιγράψουμε ποσοτικά την λέξη “Γιώργος”

Χρησιμοποιούμε μία array υπό την έννοια ενός μονοδιάστατου πίνακα με τύπο string

Γ	ι	ώ	ρ	γ	ο	ς
---	---	---	---	---	---	---

# Τα Σύμβολα και οι Λέξεις ως Δεδομένα

Το πρόβλημα με τα Σύμβολα και τις Λέξεις είναι να βρεθεί ένας τρόπος ποσοτικοποίησης τους. Αυτό μπορεί να γίνει ως εξής:

## Παράδειγμα

Έστω ότι θέλουμε να περιγράψουμε ποσοτικά την λέξη “Γιώργος”

Χρησιμοποιούμε μία array υπό την έννοια ενός μονοδιάστατου πίνακα με τύπο string

Γ	ι	ώ	ρ	γ	ο	ς
---	---	---	---	---	---	---

Κάθε κελί του πίνακα είναι μία **ordinal** κατηγορική μεταβλητή (ιδιότητα) που έχει ως Πεδίο Ορισμού το σύνολο των χαρακτήρων του Αλφάβητου.

# Τα Σύμβολα και οι Λέξεις ως Δεδομένα

Το πρόβλημα με τα Σύμβολα και τις Λέξεις είναι να βρεθεί ένας τρόπος ποσοτικοποίησης τους. Αυτό μπορεί να γίνει ως εξής:

## Παράδειγμα

Έστω ότι θέλουμε να περιγράψουμε ποσοτικά την λέξη “Γιώργος”

Χρησιμοποιούμε μία array υπό την έννοια ενός μονοδιάστατου πίνακα με τύπο string

Γ	ι	ώ	ρ	γ	ο	ς
---	---	---	---	---	---	---

Κάθε κελί του πίνακα είναι μία **ordinal** κατηγορική μεταβλητή (ιδιότητα) που έχει ως Πεδίο Ορισμού το σύνολο των χαρακτήρων του Αλφάβητου.

## Συμπέρασμα

- *Τα Σύμβολα και οι Λέξεις Μπορούν να περιγραφούν/ποσοτικοποιηθούν ως Σύνθετα Ordinal Κατηγορικά Δεδομένα*



# Τα Σύμβολα και οι Λέξεις ως Δεδομένα

Το πρόβλημα με τα Σύμβολα και τις Λέξεις είναι να βρεθεί ένας τρόπος ποσοτικοποίησης τους. Αυτό μπορεί να γίνει ως εξής:

## Παράδειγμα

Έστω ότι θέλουμε να περιγράψουμε ποσοτικά την λέξη “Γιώργος”

Χρησιμοποιούμε μία array υπό την έννοια ενός μονοδιάστατου πίνακα με τύπο string

Γ	ι	ώ	ρ	γ	ο	ς
---	---	---	---	---	---	---

Κάθε κελί του πίνακα είναι μία **ordinal** κατηγορική μεταβλητή (ιδιότητα) που έχει ως Πεδίο Ορισμού το σύνολο των χαρακτήρων του Αλφάβητου.

## Συμπέρασμα

- *Τα Σύμβολα και οι Λέξεις Μπορούν να περιγραφούν/ποσοτικοποιηθούν ως Σύνθετα Ordinal Κατηγορικά Δεδομένα*
- *Το ίδιο ισχύει και για Πολυμεσικά Δεδομένα (είναι αρχεία που έχουν Μεταδεδομένα)*

# Διαφορές Μεταξύ Αριθμητικών και Κατηγορικών Δεδομένων

- Τα Αριθμητικά Δεδομένα είναι πραγματικοί αριθμοί και άρα συνεχή, ενώ τα κατηγορικά δεδομένα είναι διακριτά (συγκεκριμένες κατηγορίες στις οποίες αντιστοιχίζονται κυρίως ακέραιοι αριθμοί)

# Διαφορές Μεταξύ Αριθμητικών και Κατηγορικών Δεδομένων

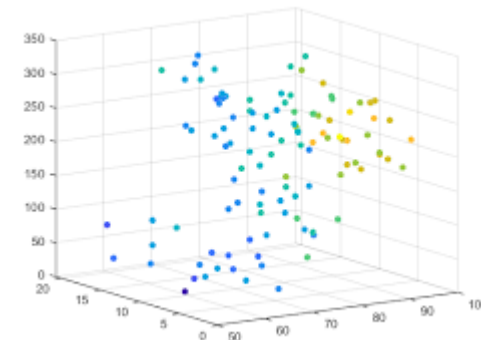
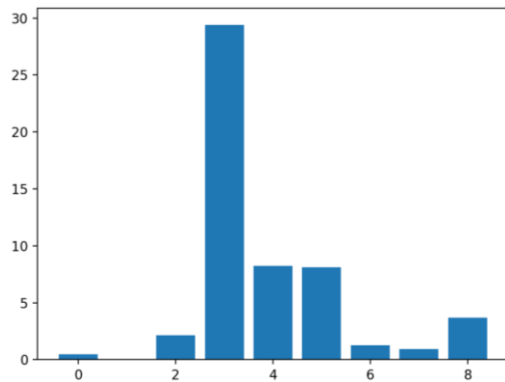
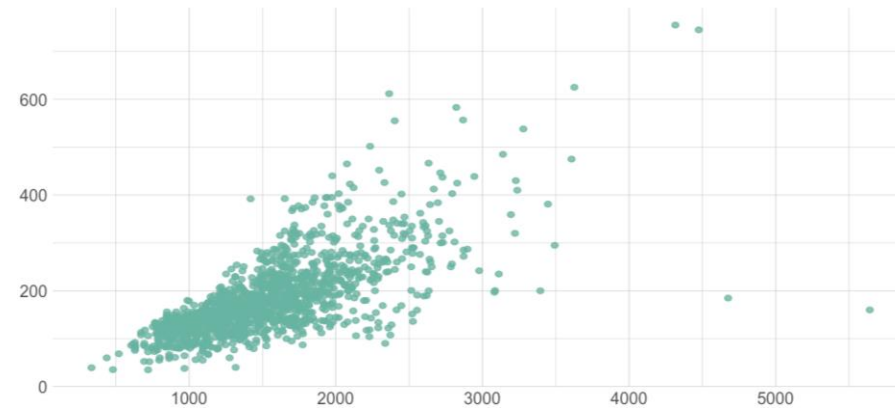
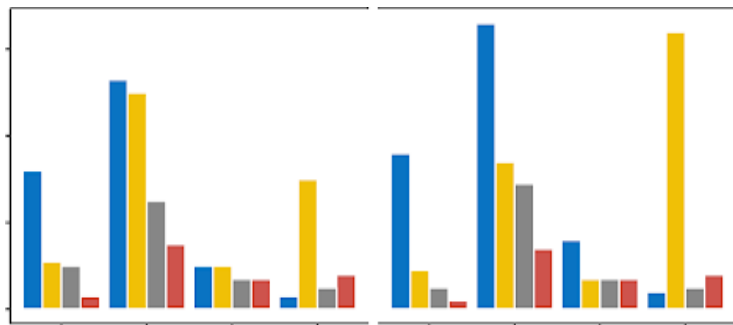
- Τα Αριθμητικά Δεδομένα είναι πραγματικοί αριθμοί και άρα συνεχή, ενώ τα κατηγορικά δεδομένα είναι διακριτά (συγκεκριμένες κατηγορίες στις οποίες αντιστοιχίζονται κυρίως ακέραιοι αριθμοί)
- Επειδή τα Αριθμητικά Δεδομένα είναι συνεχή μπορούμε να υπολογίζουμε αποστάσεις μεταξύ των δεδομένων με την χρήση της Ευκλείδειας απόστασης που είναι πολύ εύκολη και βολική. Για τα Κατηγορικά Δεδομένα η πιο κατάλληλη είναι η Hamming απόσταση. Όμως, δεν αντικατοπτρίζει την πραγματική απόσταση μεταξύ δύο κατηγορικών δεδομένων.

# Διαφορές Μεταξύ Αριθμητικών και Κατηγορικών Δεδομένων

- Τα Αριθμητικά Δεδομένα είναι πραγματικοί αριθμοί και άρα συνεχή, ενώ τα κατηγορικά δεδομένα είναι διακριτά (συγκεκριμένες κατηγορίες στις οποίες αντιστοιχίζονται κυρίως ακέραιοι αριθμοί)
- Επειδή τα Αριθμητικά Δεδομένα είναι συνεχή μπορούμε να υπολογίζουμε αποστάσεις μεταξύ των δεδομένων με την χρήση της Ευκλείδειας απόστασης που είναι πολύ εύκολη και βολική. Για τα Κατηγορικά Δεδομένα η πιο κατάλληλη είναι η Hamming απόσταση. Όμως, δεν αντικατοπτρίζει την πραγματική απόσταση μεταξύ δύο κατηγορικών δεδομένων.
- Τα Κατηγορικά Δεδομένα είναι ένας τύπος δεδομένων που χρησιμοποιείται για την ομαδοποίηση (γκρουπάρισμα) πληροφορίας σε ομάδες με κοινά χαρακτηριστικά (κατηγορίες) ενώ τα Αριθμητικά Δεδομένα είναι τύπος δεδομένων που χρησιμοποιεί αριθμούς για την περιγραφή της πληροφορίας

# Διαφορές Μεταξύ Αριθμητικών και Κατηγορικών Δεδομένων

- Τα Αριθμητικά Δεδομένα μπορούν να οπτικοποιηθούν με πολλούς τρόπους, π.χ. διαγράμματα καμπυλών, διαγράμματα διασποράς, στήλες, ράβδους, πίνακες, τρισδιάστατα γραφικά, κλπ. Σε αντίθεση, τα Κατηγορικά Δεδομένα μπορούν να οπτικοποιηθούν κυρίως μόνο με ράβδους και στήλες



# Διαφορές Μεταξύ Αριθμητικών και Κατηγορικών Δεδομένων

- Τα Αριθμητικά Δεδομένα είναι συμβατά με την στατιστική θεωρία και τις μεθόδους της και για αυτό είναι τα πλέον χρησιμοποιήσιμα δεδομένα. Τα Κατηγορικά Δεδομένα είναι πολύ δύσκολο να τα αναλύσουμε και για αυτό τα αποτελέσματά τους δεν είναι τόσο ακριβή όσο των Αριθμητικών Δεδομένων

# Διαφορές Μεταξύ Αριθμητικών και Κατηγορικών Δεδομένων

- Τα Αριθμητικά Δεδομένα είναι συμβατά με την στατιστική θεωρία και τις μεθόδους της και για αυτό είναι τα πλέον χρησιμοποιήσιμα δεδομένα. Τα Κατηγορικά Δεδομένα είναι πολύ δύσκολο να τα αναλύσουμε και για αυτό τα αποτελέσματά τους δεν είναι τόσο ακριβή όσο των Αριθμητικών Δεδομένων
- Με βάση τα παραπάνω, η επεξεργασία των Αριθμητικών Δεδομένων είναι πολύ πιο αποδοτική από τους Η/Υ σε σχέση με τα Κατηγορικά Δεδομένα

ΚΑΛΟ ΑΠΟΓΕΥΜΑ