“Objects” with Multidimensional Attributes/Features Increasingly Common

- Images
- “Customer” data
- Spatial data (especially important with location-aware devices)
- ...

Indexing critically important for efficiency!
Spatial Databases, with Geographic Data

- **Data:** Latitude-longitude pairs with associated attribute values:
  - <x1, y1, attributes>
  - <x2, y2, attributes>
  - ...

- **Queries:**
  What city is at <xi, yi>? What is within 5 miles of <xi, yi>?
  What is the nearest gas station to <xi, yi>?
  Also, “where am I” queries?

Toy Example, to Illustrate Issues

For a relation R(X, Y), find tuples where:
\[ X = 35 \text{ AND } Y > 50 \]

Can we use a B+-tree with concatenated search key \(<X||Y>\)?
Interesting Queries

- **Point queries**: equality conditions on all dimensions (e.g., \( X=35 \) AND \( Y=50 \))
- **Partial-match queries**: equality conditions on some but not all dimensions (e.g., \( Y=50 \))
- **Range queries**: inequality condition on at least one dimension (e.g., \( X=35 \) AND \( Y>50 \))
- **Nearest-neighbors queries**: given a distance function, find closest data point(s) to a given point \( <x, y> \)

Many Multidimensional Data Structures

- Grid files
- Partitioned hashing
- R-trees
- Quad trees
- ...
We moved beyond the Introduction to Databases course along several dimensions, including:

- **Types of data**: relational but also text, web pages, time series, …
  … with appropriate indexes (inverted files, indexes for OLAP and for spatial data)
- **Types of “queries”**: keyword queries, data mining, OLAP, information extraction
  … with appropriate “query”-processing strategies