The Yelp Database, used in hw2

- Yelp is a database-backed web app
- Users write reviews on local shops, restaurants, services, etc.
- Yelp tries to filter out spam, garbage reviews, saves qualified reviews in its growing database, along with photos, etc.
- Businesses are categorized. For example a pizza shop could have categories Pizza, Restaurant, and Food.
- Every year, Yelp releases an “academic subset” of its database and runs a contest for new ideas.
Yelp_db core tables

- **Review table**: the big table in the middle
  - 4.5M rows in both DBs, but different storage of review text (the actual texts of the submitted reviews, up to 64KB in length)
  - Oracle: 6.7GB data
  - Mysql: 3.7GB data, smaller because all the review texts are stored separately (not just the bigger review texts as in Oracle)
  - Each review is written by a certain user about a particular business

- **Business table**
  - 150K rows, 22MB data (1400 16KB pgs), so 140 bytes/row
  - Typical business has 30 reviews by simple division: 4.5M/150K = 30

- **Yuser table**
  - Avoiding name “user”, a reserved term in Oracle
  - 1M rows, 150MB data (9400 pgs), so 150 bytes/row on ave.
  - Typical user writes 4.5 reviews by simple division
Hw2 Delivery by Tues. 10/9 midnight

- Most important homework—need to master SQL!
- Work on homework must be your own
- Lowest (10-point) homework dropped for final grade
  So try your best, turn in what you can…
- Our grader is experienced in detecting copying
- First offence: 0 on the assignment (or exam)
- Second offence: report plagiarism to department