

# CS 327E Lab 1

Shirley Cohen

September 19, 2016

# Plan for Today

- Reading Quiz

- Lab 1 in Class:

<http://www.cs.utexas.edu/~scohen/labs/lab1.pdf>

# Question 1

**Where can I find some government datasets to work with?**

- A. Data.gov
- B. Public Data EU
- C. World Health Organization
- D. Open Science Directory
- E. All of the above & many others

# Question 2

**When should you store your dataset in a relational database instead of a file system:**

- A. When the data resides in separate files and needs to be combined to support an analysis
- B. When the data is structured or semi-structured
- C. When the dataset is massive
- D. When the data is being updated frequently
- E. All of the above

# Question 3

**When should you store your dataset in a NoSQL database rather than a relational database:**

- A. When two records in the same file have different attributes and data types
- B. When the data is structured as <key,value> pairs and the value can be anything
- C. When there is a need for a flexible schema
- D. All of the above
- E. None of the above

# Question 4

**What can you do if you discover that a dataset is incomplete (e.g. missing records and/or attributes)?**

- A. Contact the organization that published the dataset to understand how the data was collected
- B. Try to look for another dataset that contains some of the missing data
- C. Make note of the issue and continue working with the dataset
- D. All of the above
- E. None of the above