

CSE 6331, Spring 2017

Hadoop MapReduce, Visualization

Data Sets of Interest:

<https://www-odi.nhtsa.dot.gov/downloads/> (Vehicle recalls)

<https://archive.ics.uci.edu/ml/datasets/Census-Income+%28KDD%29> (Census data)

<http://www.nber.org/data/> (economy data)

<https://catalog.data.gov/dataset/health-care-provider-credential-data> (Health care)

Your data set must have at least 100 thousand tuples (rows) and less than 10 million tuples.

Hadoop:

(downloads: <http://www.apache.org/dyn/closer.cgi/hadoop/common/>,

more at: <http://hadoop.apache.org/>)

Try various numbers of mappers/reducers/file system parameters

Visualizing:

1. Get D3.js from <http://d3js.org/> Using a cloud service provider:

3. Plot (understandable) scatter charts from pairs of attributes in the dataset.

4. Plot bar graphs (at least two)