

# Effects of Employment on Health

## A Big Data Analytic

Authors: Christian Keitel, George Dagher, Khen Price



# Motivation

To determine if there exists a relationship between being unemployed and experiencing negative health symptoms. We hope to capture this correlation by analyzing hospital stay information and cross-referencing that to employment statistics from New York State.



# The Data

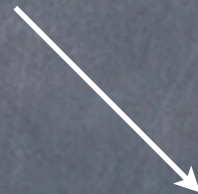
**Measuring Health:** Our main source for measuring health comes from a dataset that lists every patient accepted into the hospital and their length of stay along with the “severity” of their illness for all zip codes in New York.

**Measuring Employment:** The dataset we are using provides statistics on the percentage of people who are employed in various county’s verses those unemployed.



# Data Flow

Inpatient Data Broken up by zip code.



Grouping overall health and professions  
by zip code will give us an indicator as  
to whether a correlation exists  
between the two.

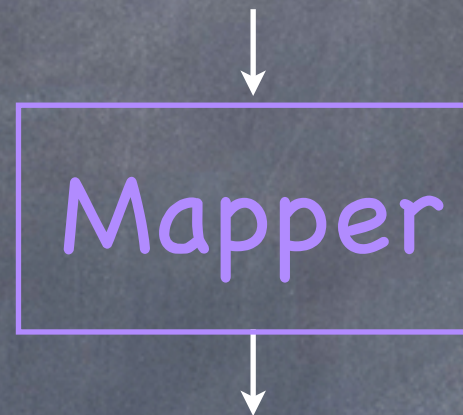
Occupation Broken Up by zip Code



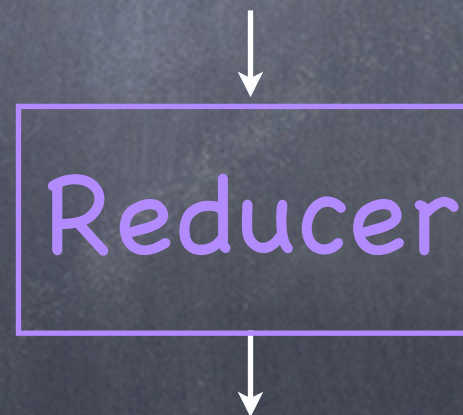


# Hadoop Process

Hospital County, Zip Code Length of Stay, APR Severity of Illness Code, Total Charges



Zip, [Length of Stay, Year, Severity, Charges]



Zip, Year, Total/Avg Stays, Total/Avg Severity, Total/Avg Charges



# Pig Process

**Table A:** \*LAUS\_Code, \*State\_FIPS\_Code, County\_FIPS\_Code, County\_State, Year, EMPTY, Force, Employed, Unemployed, UnemploymentPercentage

Discard Unused Fields

**Table B:** County\_State, Year, EMPTY, Force, Employed, Unemployed, UnemploymentPercentage

Group by Year and County

**Table C:** GROUP B by (Year, County\_State)

Filter Non-New York Records

**RAWFILTERED:** County\_NY, Year, EMPTY, Force, Employed, Unemployed, UnemploymentPercentage

Sort by Year and County

**SORTED:** County\_State, Year, EMPTY, Force, Employed, Unemployed, UnemploymentPercentage

Output to File

unemployment.tsv

\*LAUS: Local Area Unemployment Statistics

\*FIPS: Federal Information Processing Standards