Is Working Good For Your Health?

Analytics Project: Is Working Good For Your Health

Chris Keitel
George Dagher
Chem Price

Background

- 2008 recession leads to 10% unemployment
- ACA passed in 2010
- Similar research has shown:
  1. Correlation between SES (income levels, occupation, education) and health (cause of death, subjectively/objectively reported quality).
  2. Different factors effecting utilization of health services.
  3. Unclear causal relationship.

Motivation

- Use publicly available inpatient data to objectively measure unemployment/health correlation
- Location dependent measurements that can be used by government to adjust medicare services as needed
- Reveal information that may enlighten people regarding their career choices.

Data

BLS - http://data.bls.gov
- Unemployment and other labor data for the entire U.S.

Census Bureau - http://factfinder2.census.gov/
- Employment, Occupation by class data

NY open data - http://data.ny.gov
- Inpatient data across 62 counties
- Zip code, Severity, Number of Stay Days, Cost

Obstacles

- Twitter geo-location search is very limited
- HPC cluster doesn't support new MapReduce API
- Data only goes so far back (2009)
- Data didn't support our initial intuition
  - But forced us to look closer

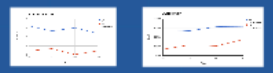
Abstract

- Employment statistics by county
- In-patient hospital data by zip code
- Correlation between unemployment and changes in health.

Design

Results

Results



Conclusions

Trends appear to be specific to different counties
- Expected statewide trends
- Counties differ greatly by population and demographics

Consider statewide budget allocations
- Albany, Franklin, Oneida counties tend to show most hospital utilization when unemployment increases

Future Work
- More years of data (this is coming)
- Process could easily scale nationwide

References

Analytics Project: Is Working Good For Your Health

Chris Keitel
George Dagher
Chen Price

Background

- 2008 recession leads to 10% unemployment
- ACA passed in 2010
- Similar research has shown:
    1. Correlation between SES (income levels, occupation, education) and health (cause of death, subjectively/ objectively reported quality).

    2. Different factors affecting utilization of health services.

    3. Unclear causal relationship.

Motivation

- Use publicly available inpatient data to objectively measure unemployment/health correlation.
- Location dependent measurements that can be used by government to adjust medicare services as needed.
- Reveal information that may enlighten people regarding their career choices.

Data

BLS - http://data.bls.gov

- Unemployment and other labor data for the entire U.S.

Census Bureau - http://factfinder2.census.gov/

- Employment, Occupation by class data

NY open data - http://data.ny.gov

- Inpatient data across 62 counties.
- Zip code, Severity, Number of Stay Days, Cost

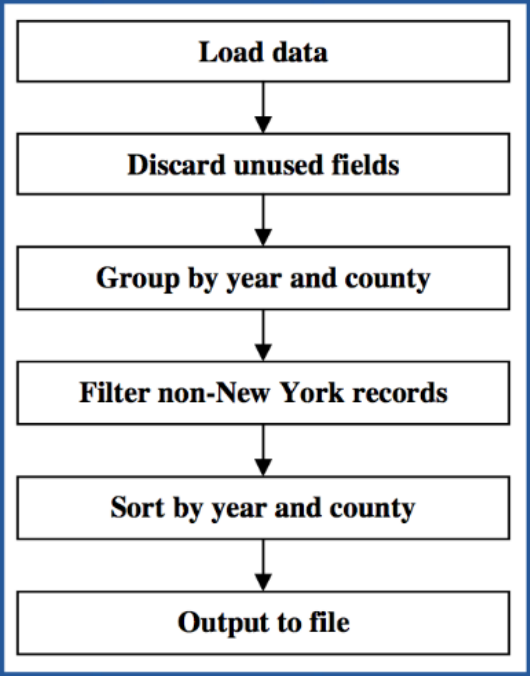
Obstacles

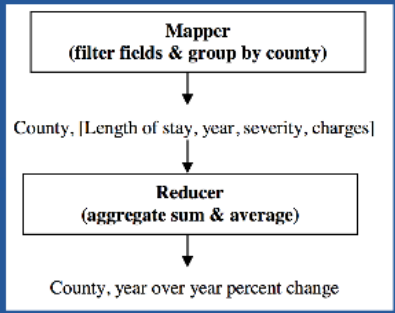
- Twitter geo-location search is very limited
- HPC cluster doesn't support new MapReduce API
- Data only goes so far back (2009)
- Data didn't support our initial intuition
    - But forced us to look closer

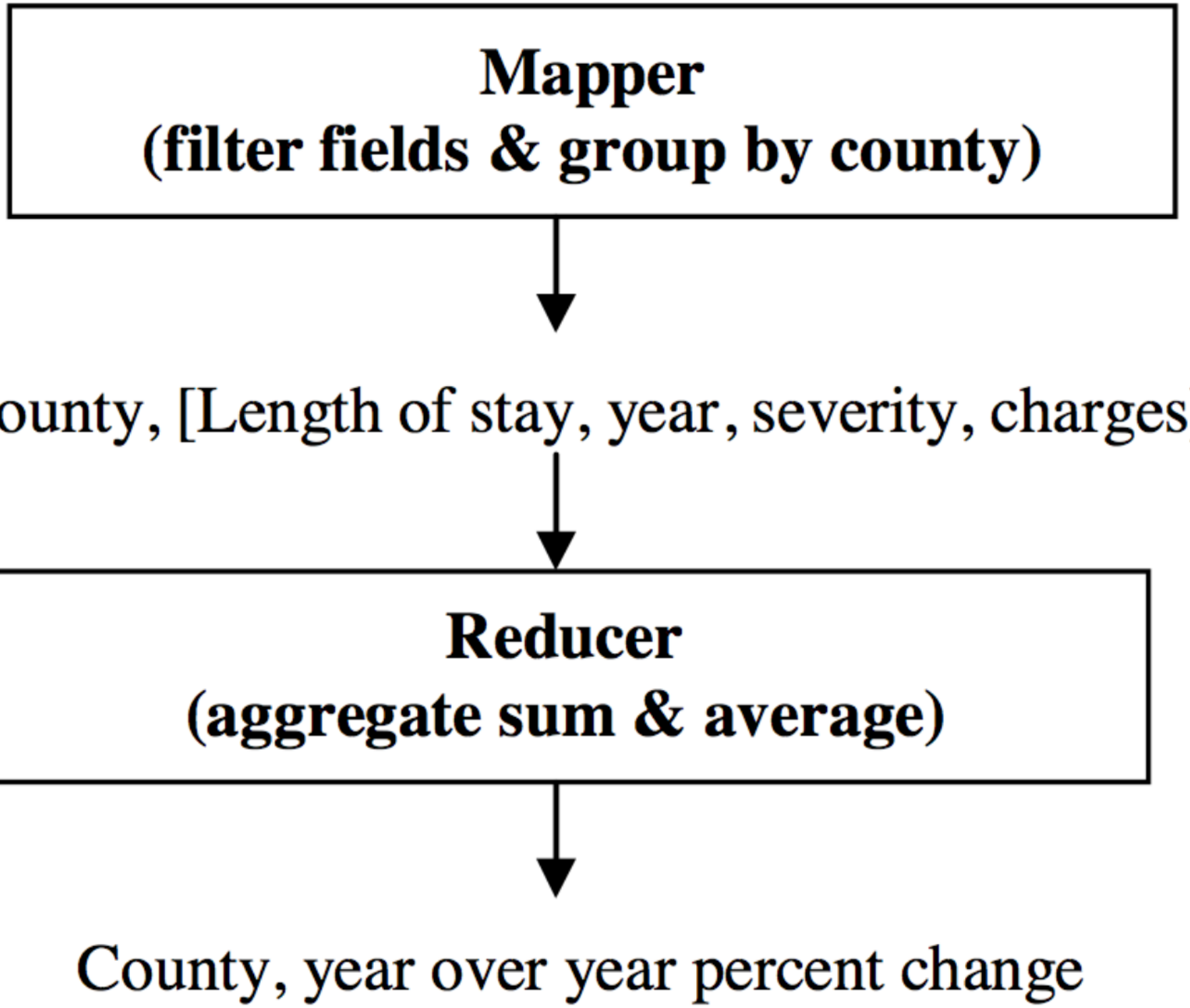
Abstract

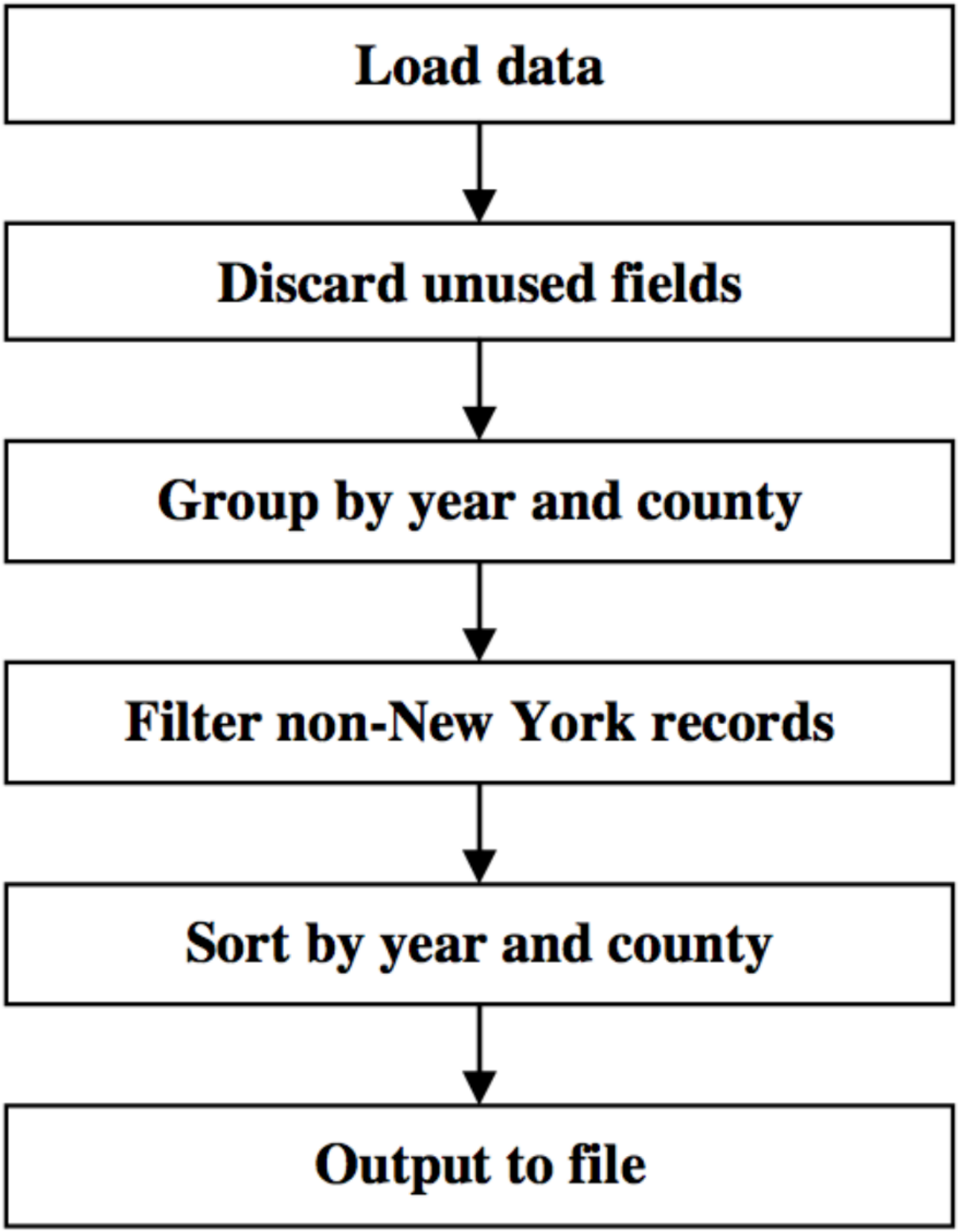
- Employment statistics by county
- In-patient hospital data by zip code
- Correlation between unemployment and changes in health.

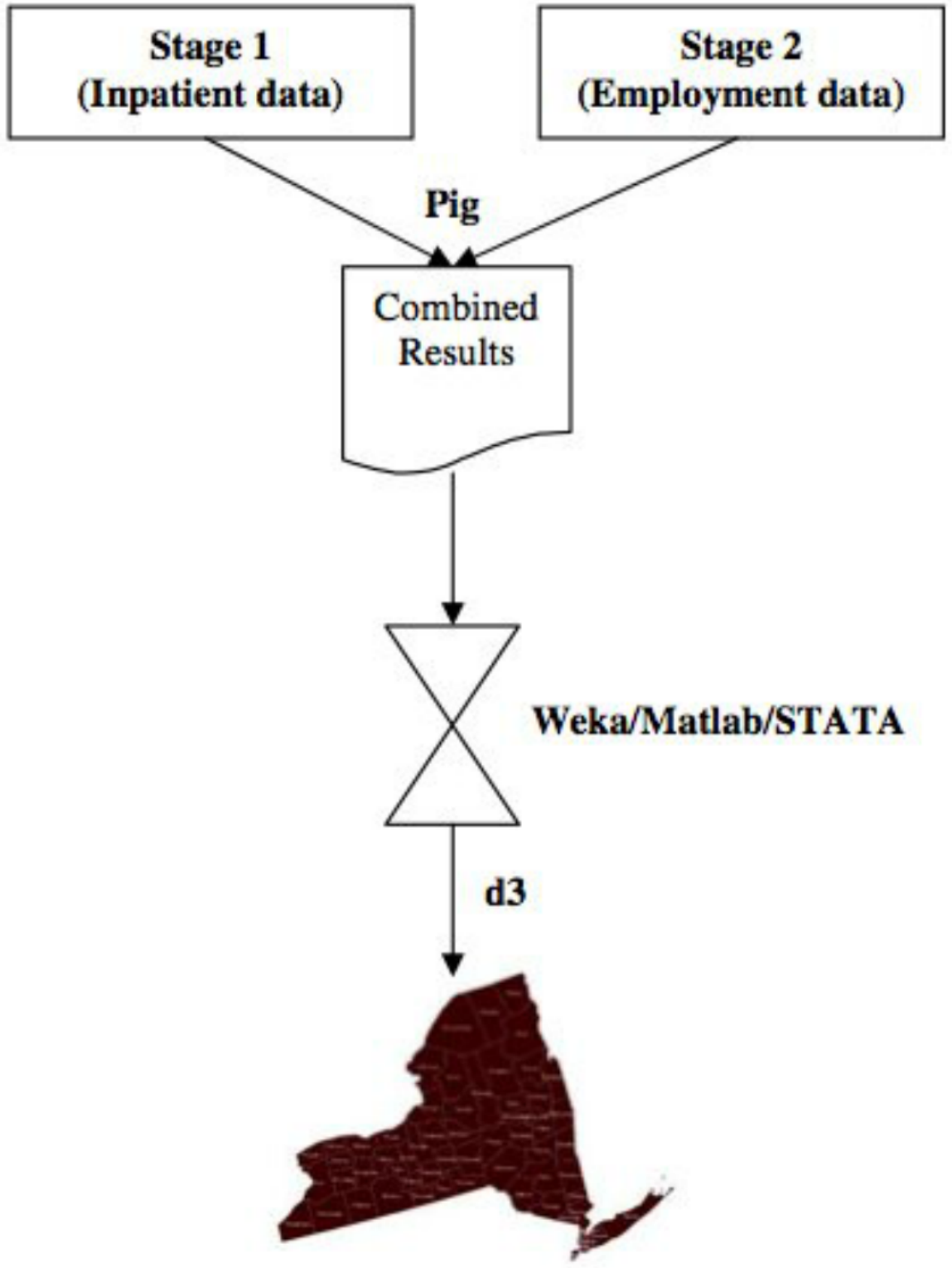
Design



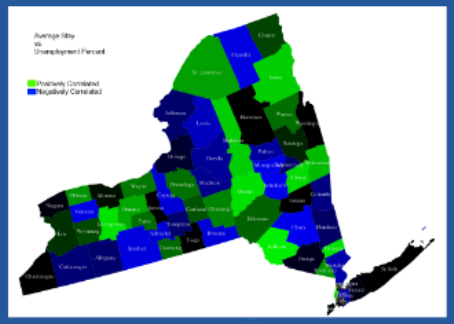


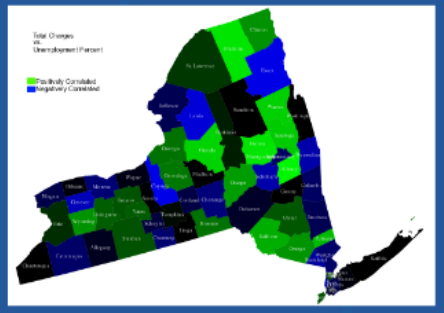

Mapper
(filter fields & group by county)
County, [Length of stay, year, severity, charges]
Reducer
(aggregate sum & average)
County, year over year percent change

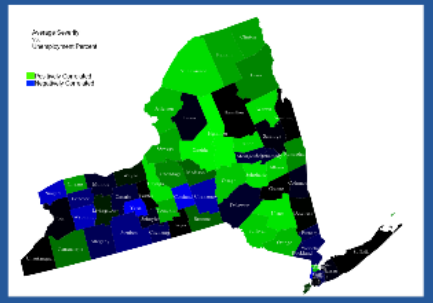

Load data
Discard unused fields
Group by year and county
Filter non-New York records
Sort by year and county
Output to file

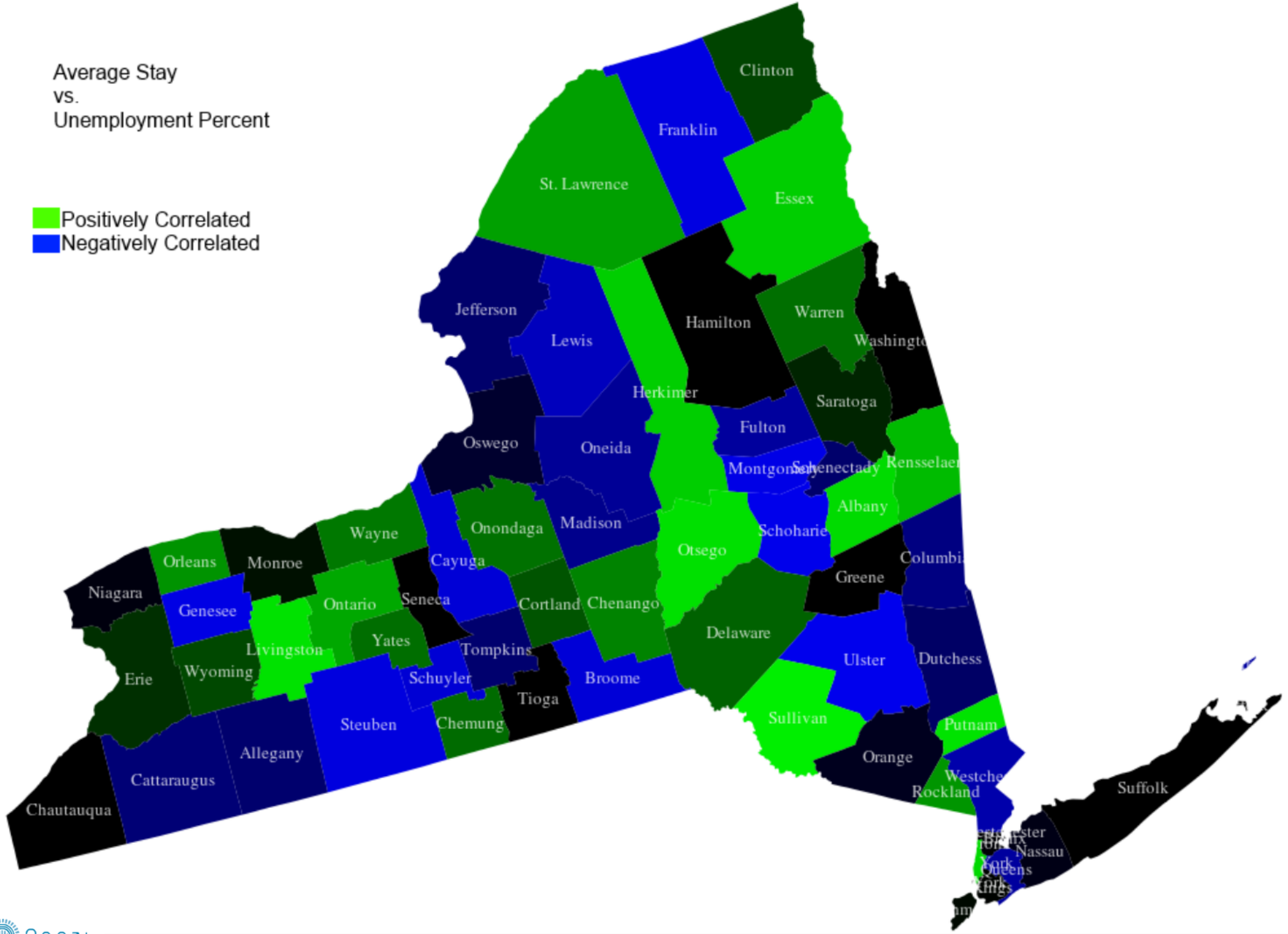

Stage 1
(Inpatient data)
Stage 2
(Employment data)
Pig
Combined
Results
Weka/Matlab/STATA
d3

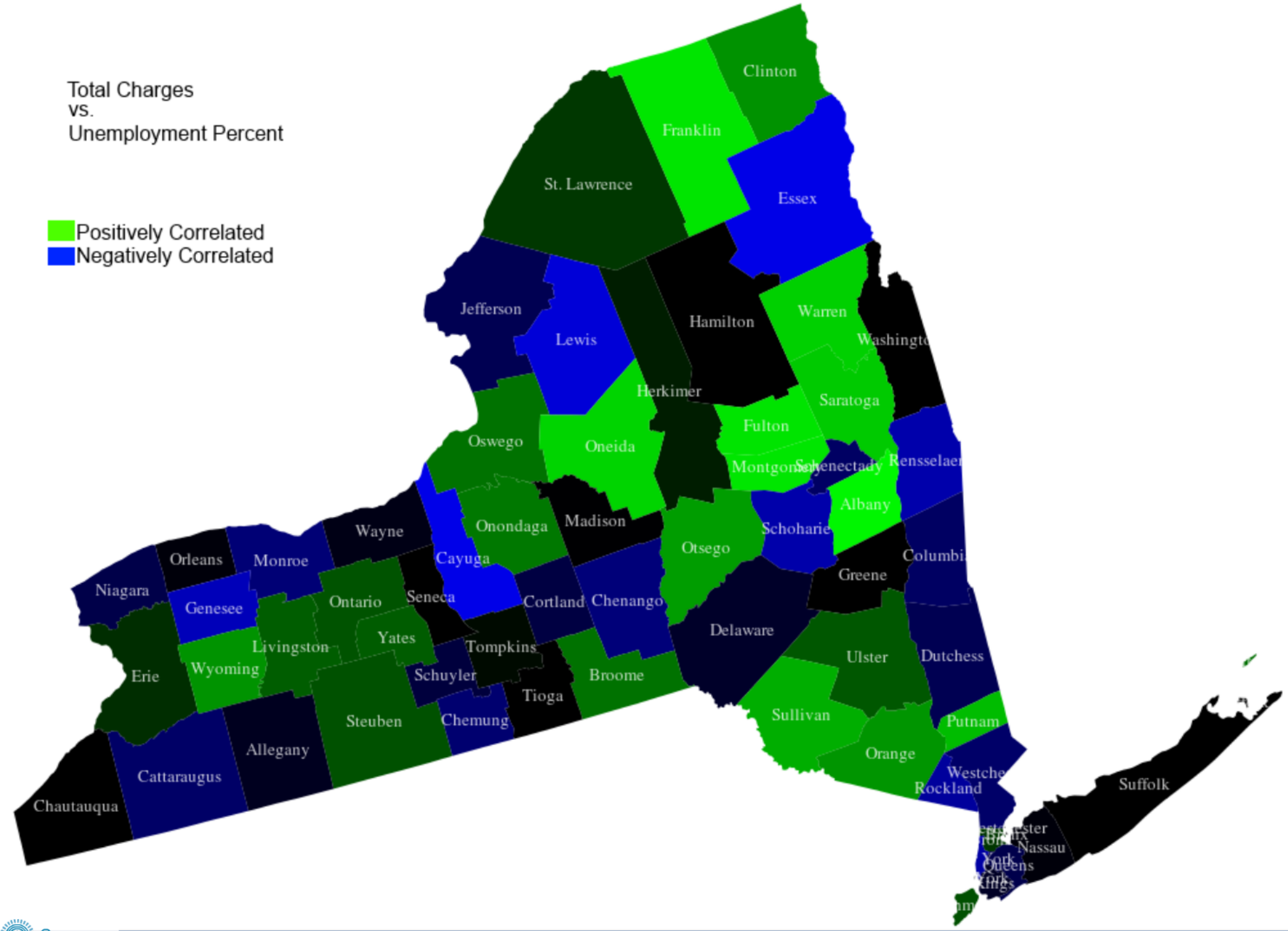
Results

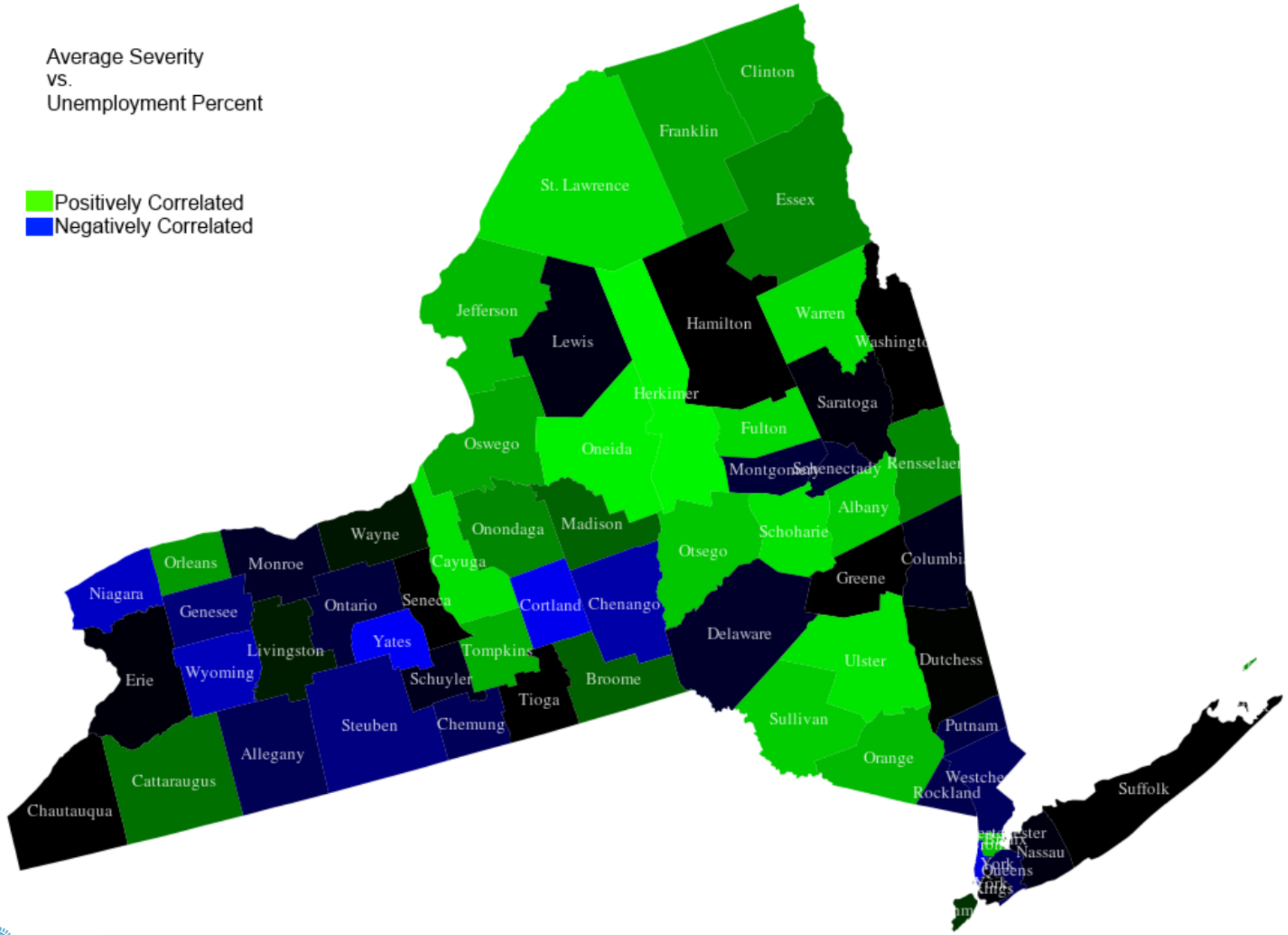


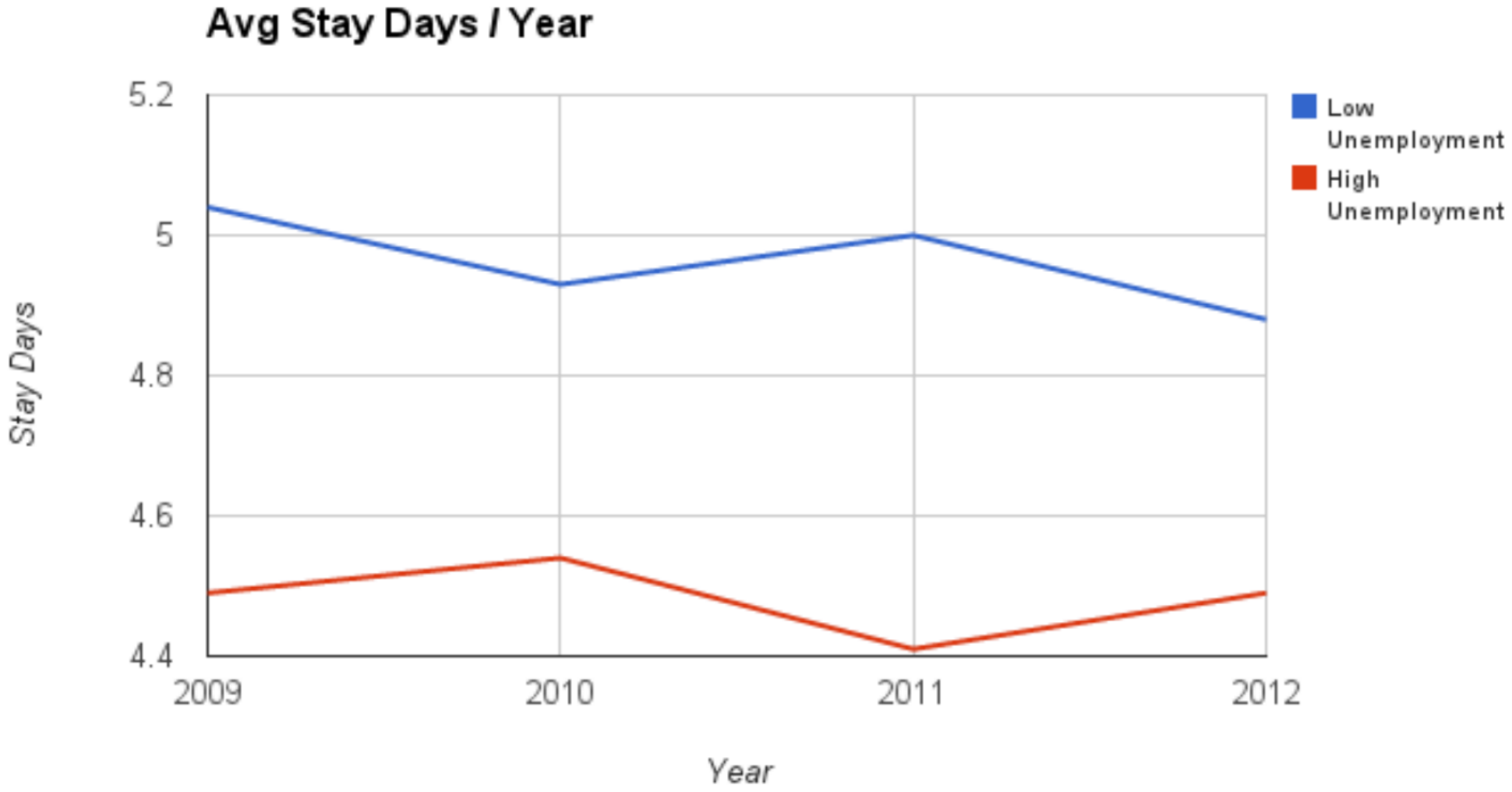


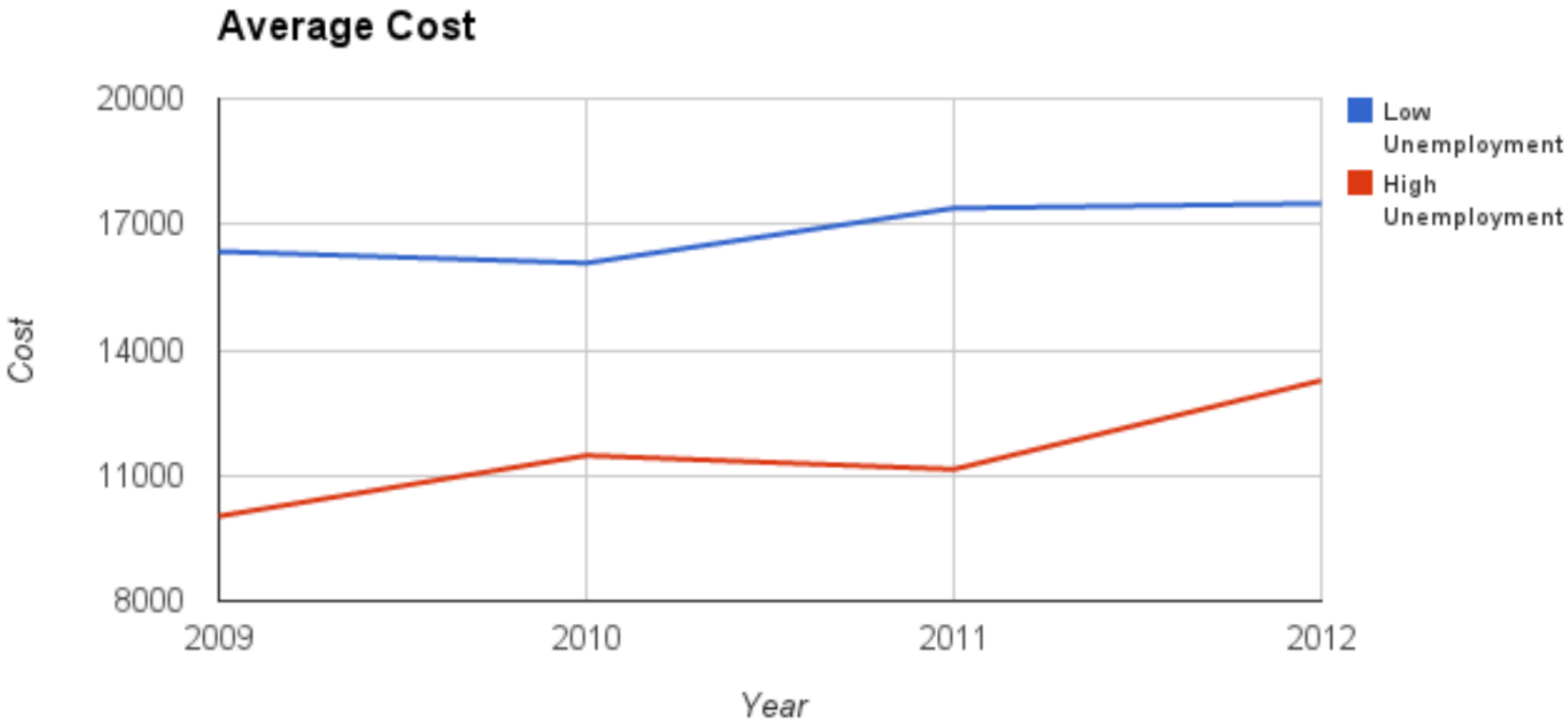



Average Stay
vs.
Unemployment Percent
Positively Correlated
Negatively Correlated
Clinton
Franklin
St. Lawrence
Essex
Jefferson
Lewis
Hamilton
Warren
Washingto
Herkimer
Saratoga
Fulton
Oswego
Oneida
Montgomery
Schenectady
Rensselaer
Albany
Onondaga
Madison
Wayne
Schoharie
Otsego
Columbi
Orleans
Monroe
Cayuga
Greene
Niagara
Genesee
Ontario
Seneca
Cortland
Chenango
Delaware
Livingston
Yates
Tompkins
Ulster
Dutchess
Erie
Wyoming
Schuyler
Broome
Tioga
Sullivan
Putnam
Steuben
Chemung
Orange
Allegany
Westche
Rockland
Cattaraugus
Suffolk
Chautauqua
Nassau
York
Queens


Total Charges
vs.
Unemployment Percent
Positively Correlated
Negatively Correlated
Clinton
Franklin
St. Lawrence
Essex
Jefferson
Lewis
Hamilton
Warren
Washingto
Herkimer
Saratoga
Fulton
Oswego
Oneida
Montgom
Schenectady
Rensselaer
Albany
Onondaga
Madison
Wayne
Otsego
Schoharie
Columbi
Orleans
Monroe
Cayuga
Greene
Niagara
Genesee
Ontario
Seneca
Cortland
Chenango
Delaware
Livingston
Yates
Tompkins
Ulster
Dutchess
Erie
Wyoming
Schuyler
Broome
Tioga
Sullivan
Putnam
Steuben
Chemung
Allegany
Orange
Westche
Rockland
Cattaraugus
Suffolk
Chautauqua
Nassau
Queens


Average Severity
vs.
Unemployment Percent
Positively Correlated
Negatively Correlated
Clinton
Franklin
St. Lawrence
Essex
Jefferson
Lewis
Hamilton
Warren
Washingto
Herkimer
Saratoga
Fulton
Oswego
Oneida
Montgomery
Schenectady
Rensselaer
Albany
Schoharie
Otsego
Onondaga
Madison
Wayne
Orleans
Monroe
Cayuga
Columbi
Greene
Niagara
Genesee
Ontario
Seneca
Cortland
Chenango
Delaware
Yates
Tompkins
Livingston
Ulster
Dutchess
Erie
Wyoming
Schuyler
Broome
Tioga
Sullivan
Putnam
Steuben
Chemung
Allegany
Orange
Westche
Rockland
Cattaraugus
Suffolk
Chautauqua
Nassau
Queens


Avg Stay Days / Year
5.2
5
4.8
4.6
4.4
Stay Days
2009
2010
2011
2012
Year
Low Unemployment
High Unemployment


Average Cost
20000
17000
14000
11000
8000
Cost
2009
2010
2011
2012
Year
Low Unemployment
High Unemployment

Conclusions

Trends appear to be specific to different counties

- Expected statewide trends
- Counties differ greatly by population and demographics

Consider statewide budget allocations

- Albany, Franklin, Oneida counties tend to show most hospital utilization when unemployment increases

Future Work

- More years of data (this is coming)
- Process could easily scale nationwide

References

[1] T. White. Hadoop: The Definitive Guide. O'Reilly Media Inc., Sebastopol, CA, May 2012.

[2] A. Gates. Programming Pig. O'Reilly Media Inc.,Sebastopol, CA, October 2011.

[3] J. Dean and S. Ghemawat. MapReduce: Simplified data processing on large clusters. In proceedings of 6th Symposium on Operating Systems Design and Implemenation, 2004.

[4] S. Ghemawat, H. Gobioff, S. T. Leung. The Google File System. In Proceedings of the nineteenth ACM Symposium on Operating Systems Principles – SOSP '03, 2003.

[5] Bureau of Labor Statistics. http://data.bls.gov/timeseries/LNS14000000

[6] Bureau of Labor Statistics. http://data.bls.gov/timeseries/IIU00000000061100

[7] B. Hallerod and J. Justafsson. A longitudinal analysis of the realtionship between changes in socio-economic status and changes in health. In Social Science & Medicine, 2010, Vol. 72 No. 1 pp 116-23.

[8] G. Brant Morefield, David C. Ribar, and Christopher J. Ruhm. Occupational and Health Transitions. National Bureau of Economic Research, Cambridge, MA, February 2011.

[9] C. Ross and J.Mirowsky. Does Employment Affect Health? Journal of Health and Social Behavior 1995, Vol. 36 (September) pp 230-43.
[10] R. L. Jin, C. P. Shah, and T. J. Svoboda. The Impact of Unemployment on Health: A Review of the Evidence. Journal of Public Health Policy, 1993, Vol. 18 No. 3, pp 275-301.

[11] Bernstein AB, Hing E, Moss AJ, Allen KF, Siller AB, Tiggle RB. Health care in America: Trends in utilization. Hyattsville, Maryland: National Center for Health Statistics. 2003.

[12] N. Adler and K. Newman. Socioeconomic Disparities In Health: Pathways and Policies. Health Affairs, March, 2002, Vol. 21 No. 2 pp 60-76.

[13] Weka 3: Data Mining Software in Java. http://www.cs.waikato.ac.nz/ml/weka/