Predicting Economic Loss from Natural Disasters

Prepared by: Andrew Davis | Victor Voskovsky | Bryan Lange

NLT New Light Technologies
1440 G Street Northwest
Washington, DC 20005
1-800-206-5994

Github Repository: Click here.
5/15/20

Project Overview

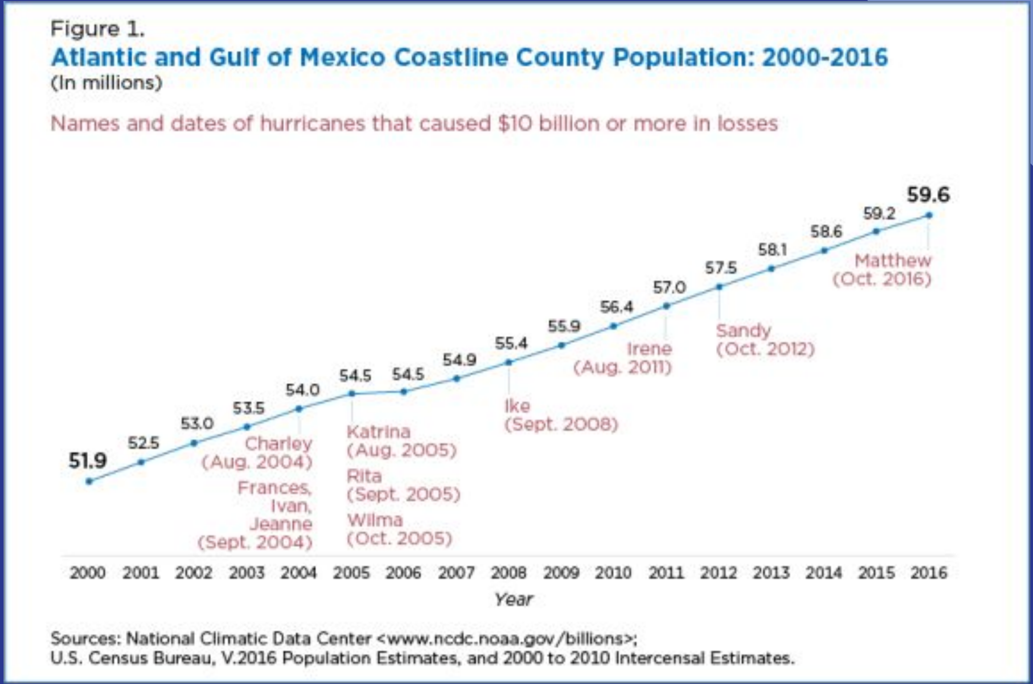
- Introduction and problem statement
- Data wrangling process
- EDA and visualizations
- Modeling the data
- Key observations from analysis
- Challenges and limitations
- Final insights and recommendations

Introduction

- According to the NOAA, since 1980 the United States has experienced 265 weather and climate disasters where the overall costs reached or exceeded one billion dollars

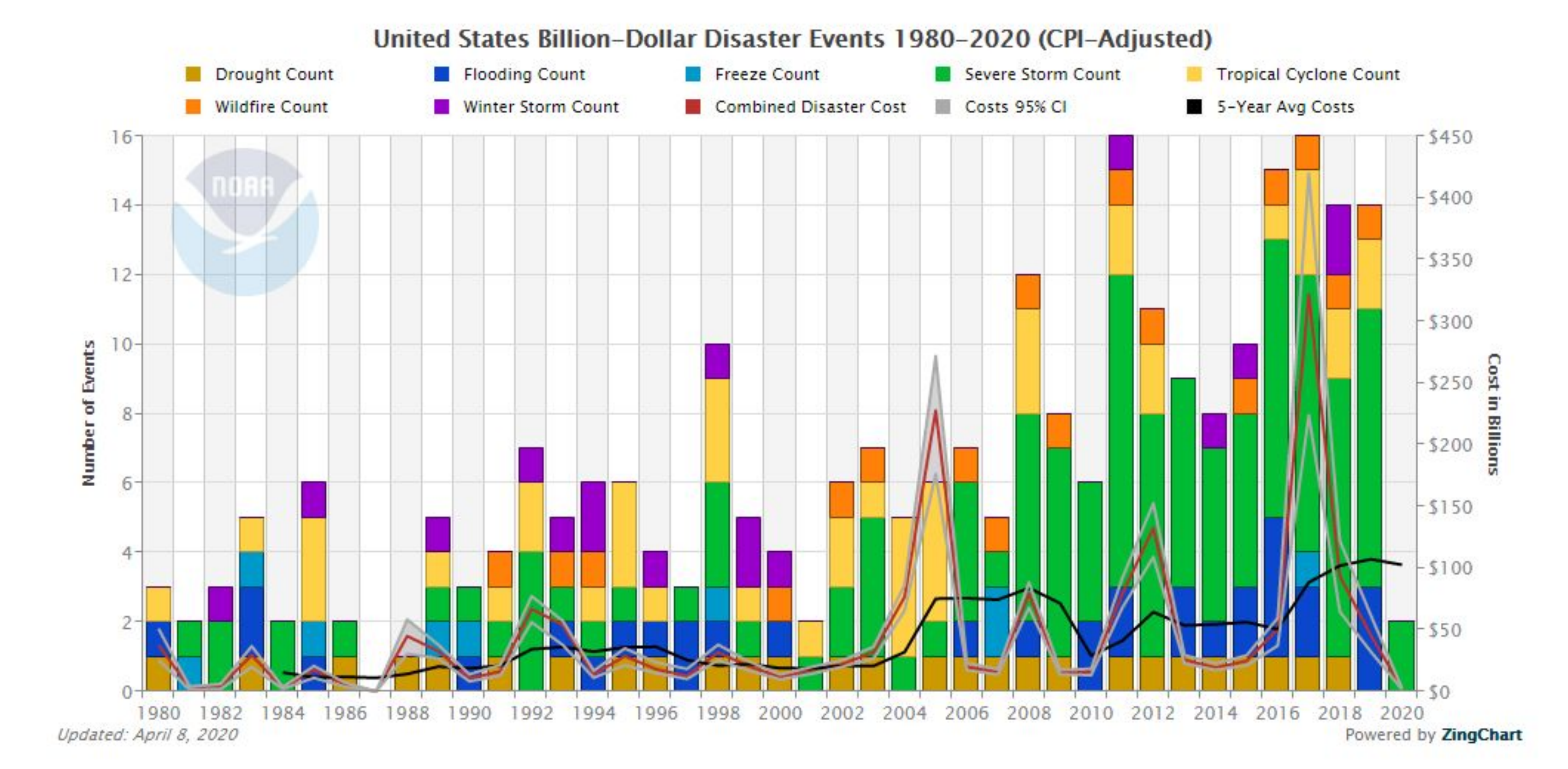
- Given the overall increase in disasters in the recent past combined with population growth in coastal regions, planning effective response and relief efforts has never been more critical

- The ability to predict potential wage loss from natural disasters will help communities focus their rebuilding efforts on the industries that need it most and help to mitigate economic impact



https://www.ncdc.noaa.gov/billions/

Focus on Hurricanes



https://www.ncdc.noaa.gov/billions/time-series/US

Problem Statement

- New Light Technologies is a DC based organization that works closely with FEMA, USAID and other agencies to help provide solutions in the aftermath of emergencies and natural disasters

- Using major hurricane information in combination with data from the Bureau of Labor Statistics, our goal is to use employment and wage figures to help cluster data and predict economic impact on local industries after a hurricane

Primary Objective

- Given known industry and geographic info, how can we help devastated communities allocate resources to bounce back as quickly as possible?

Chart 5. Over-the-year changes in employment in the construction sector, New Orleans, January 2004 to June 2006

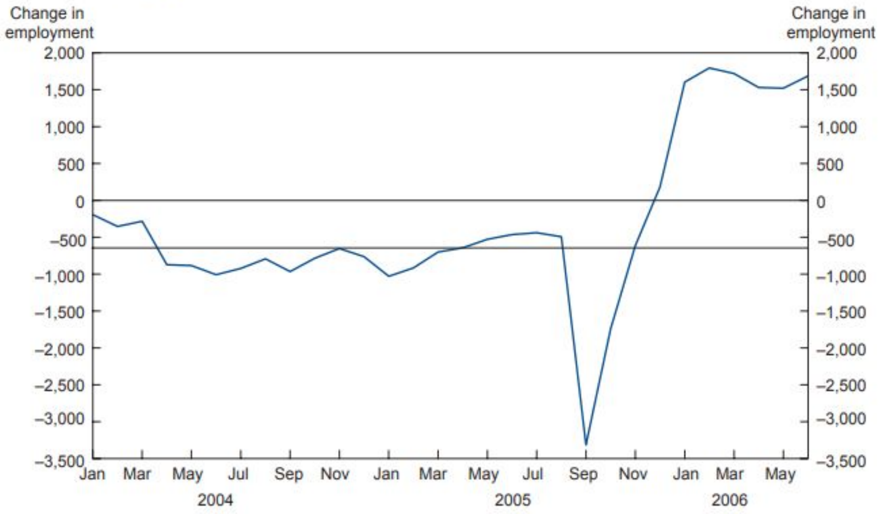


Table 6. Fourth-quarter employment and wages, Orleans Parish, 2005

Industry	Average monthly employment (thousands)	Percent of Orleans Parish employment	Percent change in employment, 2000–05	Total wages (millions)	Percent of Orleans Parish total wages	Average weekly wage	Over-the-year percent change in wage
All industries	144.2	100.0	–41.7	$1,838.8	100.0	$981	29.4
Private	110.1	76.4	–43.2	1,429.1	77.7	998	34.0
Agriculture, forestry, fishing, and hunting	(1)	(2)	–27.2	.4	(2)	708	58.7
Mining	4.1	2.9	.3	104.0	5.7	1,943	5.6
Utilities	1.0	.7	–.9	18.3	1.0	1,428	–.9
Construction	5.3	3.7	–12.0	74.4	4.0	1,076	25.8
Manufacturing	6.1	4.2	–20.2	84.0	4.6	1,059	6.8
Wholesale trade	4.5	3.1	–27.8	67.9	3.7	1,170	14.3
Retail trade	7.2	5.0	–62.8	52.5	2.9	562	20.9
Transportation and warehousing	7.4	5.1	–31.7	99.4	5.4	1,035	22.6
Information	3.9	2.7	–34.6	47.5	2.6	942	31.2
Finance and insurance	6.1	4.2	–35.0	116.8	6.4	1,481	35.6
Real estate and rental leasing	1.9	1.3	–46.8	18.6	1.0	758	25.3
Professional and technical services	11.8	8.2	–15.7	226.5	12.3	1,481	2.9
Management of companies and enterprises	2.7	1.9	–41.1	51.4	2.8	1,462	20.3
Administrative and waste services	9.0	6.2	–40.1	109.3	5.9	938	77.0
Educational services	6.0	4.2	–35.6	87.6	4.8	1,117	36.6
Health care and social assistance	11.1	7.7	–56.4	111.5	6.1	774	3.3
Arts, entertainment, and ecreation	3.7	2.6	–51.3	43.3	2.4	893	41.1
Accommodation and food services	14.5	10.1	–59.3	81.1	4.4	430	25.4
Other services, except public administration	3.6	2.5	–53.8	30.9	1.7	663	31.5
Port operations	11.5	8.0	–22.9	203.4	11.1	1,118	21.6
Tourism	18.2	16.6	–57.9	124.4	8.7	394	33.0
Federal government	11.4	7.9	–9.4	195.6	10.6	1,319	17.3
State government	16.3	11.3	–17.2	158.3	8.6	745	1.9
Local government	6.4	4.4	–70.3	55.8	3.0	675	–2.0

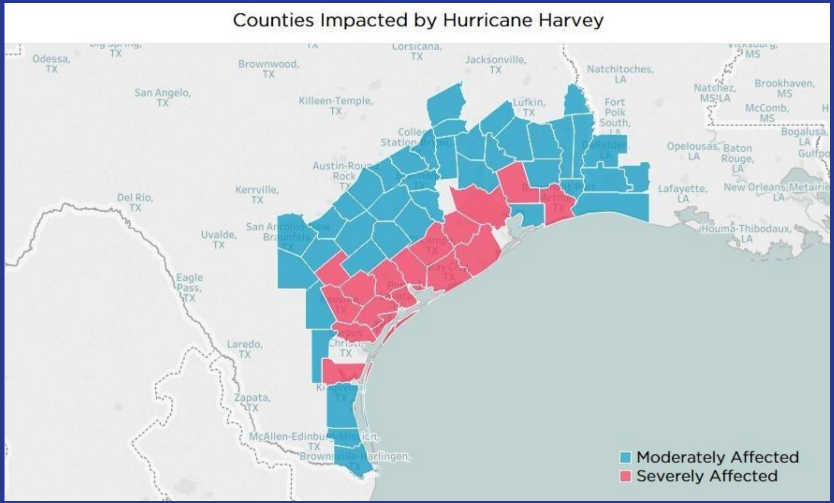
[1] Fewer than 500 employees.
[2] Less than 0.1 percent.

NOTE: Percentage bases include the approximately 0.3 percent of private employment with nonclassifiable industries.

https://www.bls.gov/opub/mlr/2007/06/art1full.pdf

Preliminary research

- Bureau of Labor Statistics (BLS) public database:
  - The Quarterly Census of Employment and Wages (QCEW)
- National Hurricane Center database (NOAA):
  - Location, wind, and pressure of tropical cyclones in Atlantic and Pacific Oceans
- Hurricane trajectory maps.
  - Visualizing major and minor impacts for localities.
- Picking target hurricanes.
  - Considerations:
    - Impact extent
    - Locations
    - Year
    - Quarters



Resources used for this project

What did we use to collect the data?

- Google BigQuery - Public Bureau of Labor Statistics DB
- Google project and key.json
- SQL
- GBQ library in Python
- Google service_account authentication in Python
- Other:
  - Pandas
  - Numpy
  - Scikit Learn

Data EDA

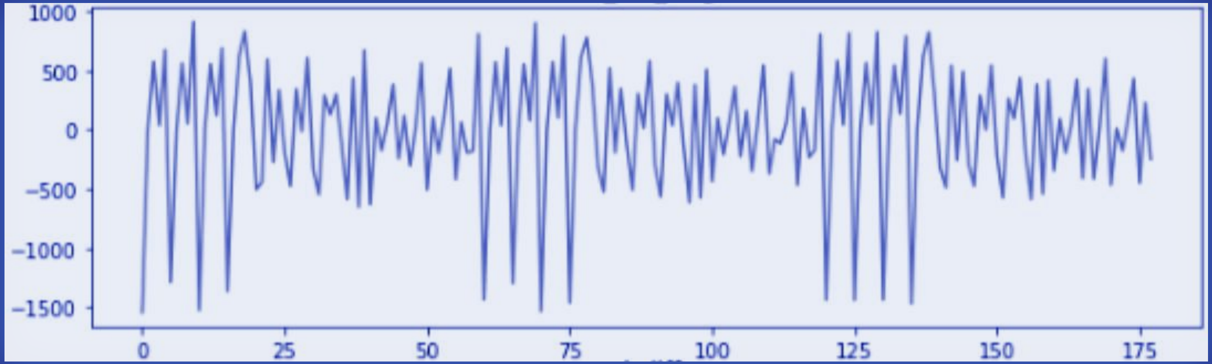
Our hurricane data is broken down by 3 main categories for 30 distinct industries:

- Quarterly establishments. (single physical location at which business is conducted)
- Average weekly wages by quarter.
- Employment over-all for county for month 3 of the quarter. (raw employment, total number of emps)

Our industries are split between real regular values and Location Quotient. The LQ was the key for our analysis.

"Location quotient (LQ) is basically a way of quantifying how concentrated a particular industry, cluster, occupation, or demographic group is in a region as compared to the nation. It can reveal what makes a particular region "unique" in comparison to the national average."
Source: EMSI Understanding Location Quotient

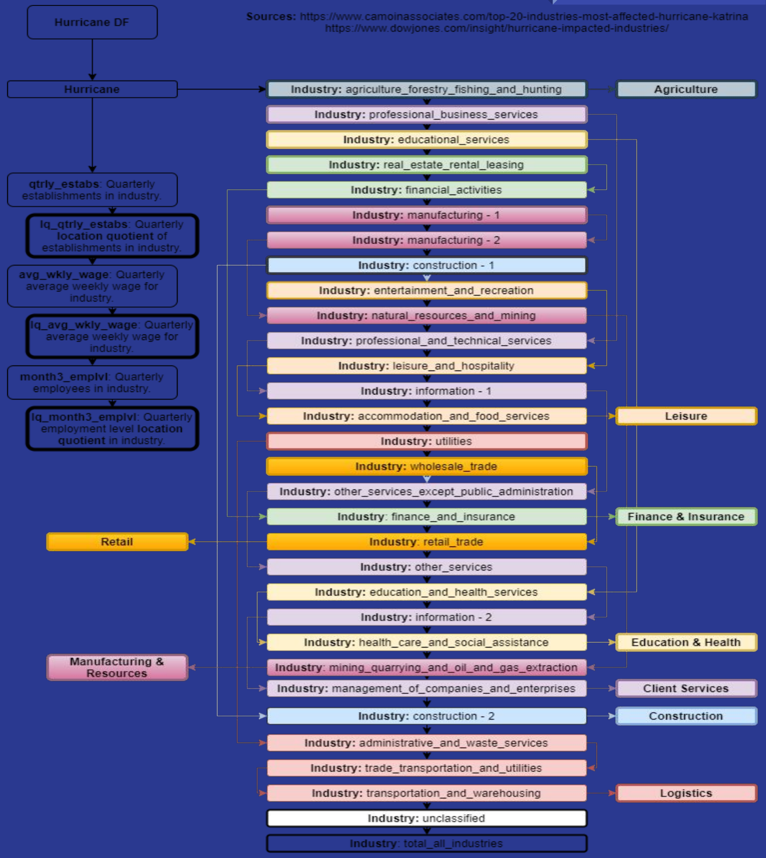
- We were also able to account our model for seasonality after discovering this trend using some time series graphing, for example:



Categorizing our Features

- We decided to categorize our industries as shown on the right. It allowed us to cut down on the amount of features our model needs to work through and in general cleaned up our dataset.
- “Unclassified” was omitted because we didn’t have a good understanding of what gets sorted into that category exactly.
- “Total Industries” was excluded as well because its jsut a summation of all the other features.

- Our final categories were:
  - Agriculture
  - Leisure
  - Finance & Insurance
  - Retail
  - Education & Health
  - Manufacturing & Resources
  - Client Services
  - Construction
  - Logistics

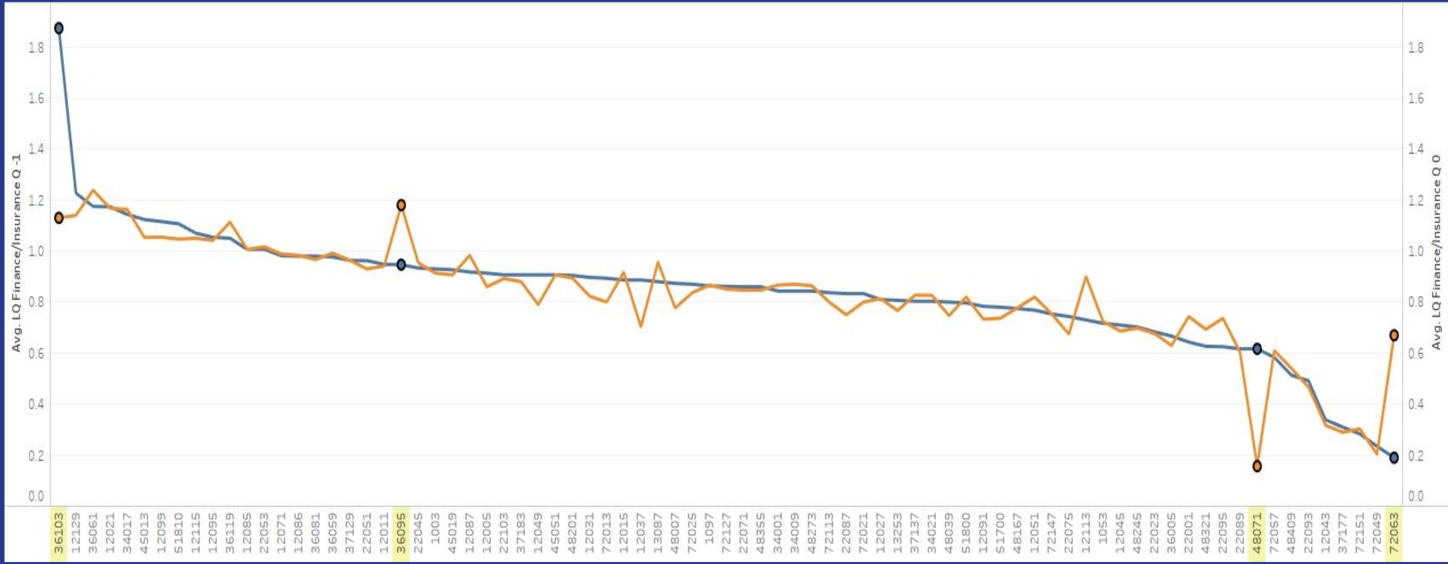


Finance/Insurance LQ

For example: Let's take the location quotients for 3 industries.

- Real Estate and Rental Leasing
- Financial Activities
- Finance and Insurance

- We group these under "Finance/Insurance" and we plot the wages location quotient.
- We then compare Q -1, which is 1 quarter before the Hurricane and Q 0, quarter of the hurricane.



Top points of interest:

Location | Hurricane | Impact

- 36103: Suffolk, NY (Sandy) - Major

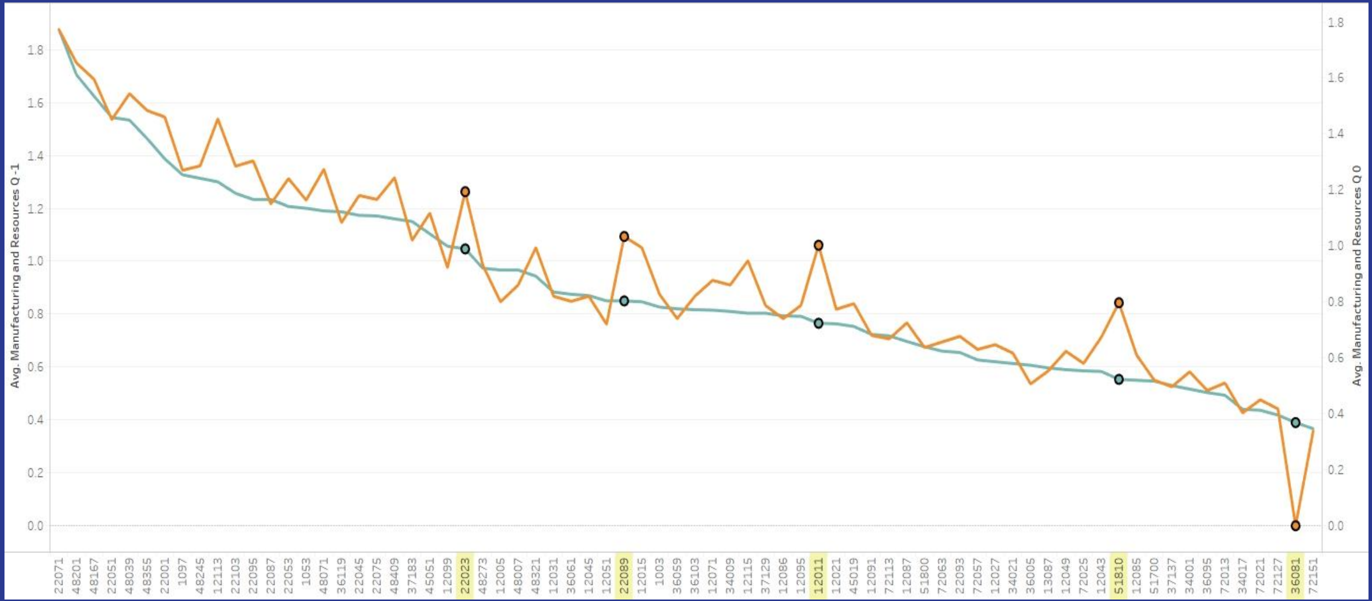
- 36095: Schoharie, NY (Sandy) - Major

- 48071: Chambers, TX (Ike) - Major

- 72063: Gurabo, P.R. (Irene) - Major

Manuf. & Resources LQ

- Graphing the mining & resources industry Q -1 to Q 0, has shown some interesting results also.
- We could see the impact of a hurricane on the wages industry by comparing two quarters.
- With this fluctuation of abnormal results could mean that a viable model could be built with this data.
- Mining includes:
    - Manufacturing
    - Oil and Gas
    - Mining and quarrying



Top points of interest:

Location | Hurricane | Impact

- 51810: Virginia Beach, FL (Irene) - Major

- 12011: Broward, FL (Irma) - Major

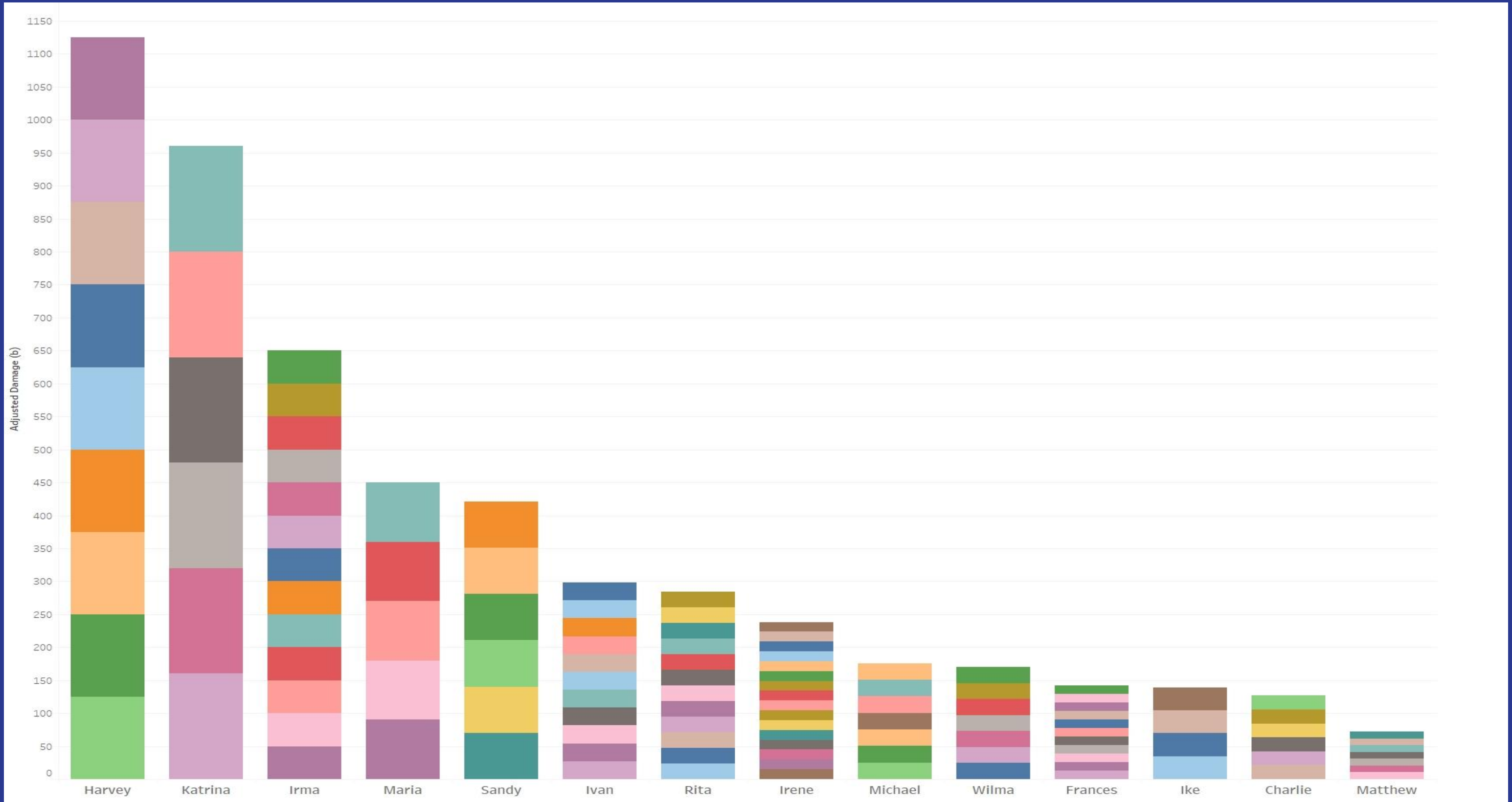
- 22089: St. Charles Parish, LA (Katrina) - Major

- 22023: Cameron, LA (Rita) - Major

- 36081: Queens, NY (Sandy) - Major

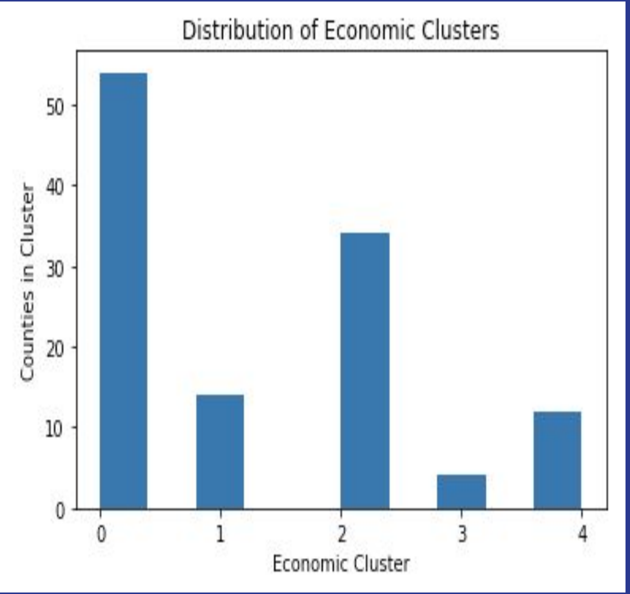
14 Hurricanes Included in Project:

Adjusted Total Damage by Hurricane ($)



Clustering Counties

- Using industries, averaged LQ quotients over quarters preceding hurricane
- K-Means Clustering to group together counties with similar economic snapshots
- 5 total clusters
- Realize that the same county used in different hurricanes assessments should be in same cluster



	geoid	lq_cluster	area_fips	year	county	emp_chg
8	Frances	1	72113	2004	Ponce	0.070335
9	Frances	1	72057	2004	Guayama	0.095100
11	Frances	1	72025	2004	Caguas	0.095684
30	Irene	1	72151	2011	Yabucao	0.016069
32	Irene	1	72063	2011	Gurabo	0.058637
47	Irma	1	72113	2017	Ponce	0.022784
48	Irma	1	72025	2017	Caguas	-0.003478
49	Irma	1	72021	2017	Bayamon	-0.010113
50	Irma	1	72013	2017	Arecibo	0.002989
77	Maria	1	72113	2017	Ponce	0.022784
78	Maria	1	72025	2017	Caguas	-0.003478
79	Maria	1	72021	2017	Bayamon	-0.010113
80	Maria	1	72013	2017	Arecibo	0.002989

Modeling the Data

Aggregating LQ data → **Clustering** → **Linear regression w/Lasso Regularization**

- Economic damage by county (in $100m)
- 2nd order model
- Clustering, Category, 3 trailing economic metrics
- r^2 scores all below .2
- Random forest regressor .22 best score

Constant	505.84
Category	4.104
Cluster_4	1.5
Category * Cluster_1	1.86
County Employer Establishments*Cluster_3	0.37
Cluster_4^2	1.5

Modeling the Data

- Attempting to model Change in Overall Employment
- Model still overfit, test scores around 0
- Some industries still very predictive, why?
- Much of the data is clustered around 0
- Fitting this while also including Katrina disaster counties is tough for MLR
- How can we make this more useful?

Positive for Employment Rebounding from Hurricane	Model Beta(%)	Negative for Employment Rebounding from Hurricane	Model Beta(%)
Finance	1.3	Mining/Energy	-2.8
Agriculture	1.6	Leisure/Hospitality	-3.2
Public Construction	1.0	Education/Health	-1.0

Using Clusters for Classification

- Addressing the 0 problem
- Use clusters to predict counties that will suffer a >5% drop in employment
- Random Forest Classifier using same features as regression
- A useful model!
- Null model: .833
- Our scores: .97 train , .933 test
- Caveats: data is fairly clustered by hurricane
- Need economic data to give this context

Test Set	Actual 5% Drop	Actual Less Affected
Predicted 5% Drop	3	0
Predicted Less Affected	2	25

Limitations

- Economic effects mixed in with hurricanes
- Limited data (only 20 hurricanes in recent times that caused a large blip)
- Only 118 counties, any classifier will have SSS issues
- Null values for LQ codes or quarters missing
- Economic snapshots changing over time may be misclassified
- Could improve by using better hurricane data other than ‘Category’
- Economic data acquisition extremely time intensive

Conclusions

- Nothing in any of the models groundbreaking
- Hurricanes disproportionately affect real estate, tourism, mining, and oil/gas industries
- Areas that rely on those industries likely to see significant economic problems
- Expanding geographically or periodically might help expand data which is badly needed
- Economic data has many important exogenous factors
- Predictions are tough
- Eventually, economic recovery possible if not probable