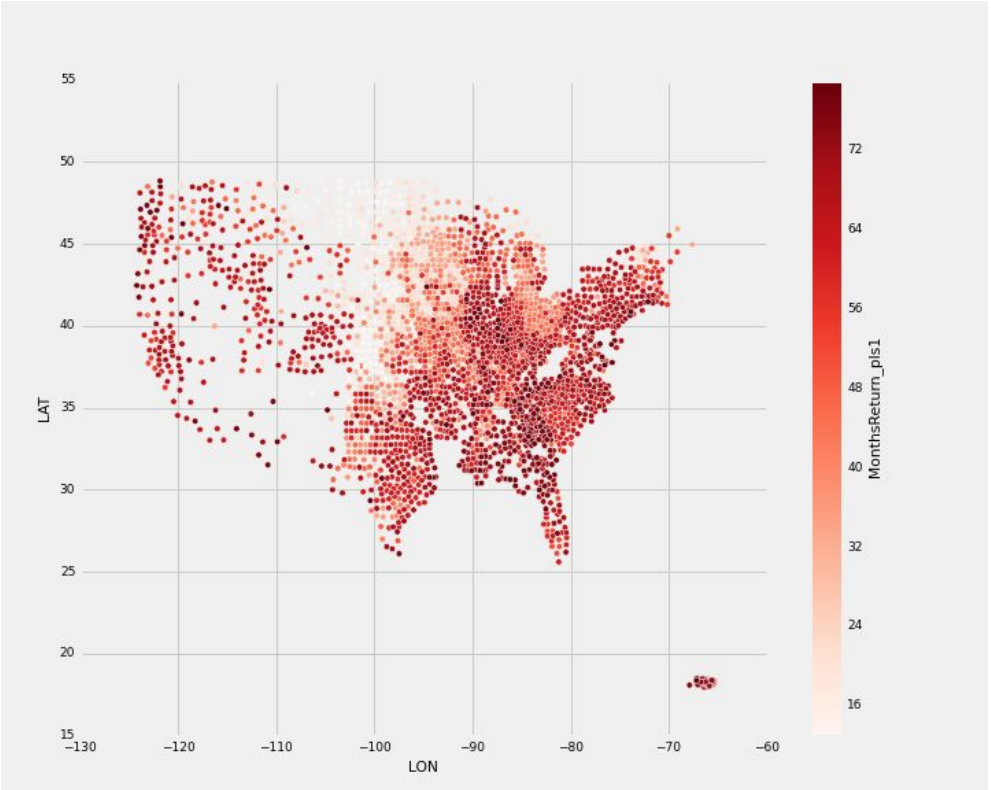
Predicting Local Experience after the Great Recession

Kyle Alden -- General Assembly DAT8, Fall 2015

The problem

Localities experienced the Great Recession in vastly different ways. Some economies quickly regained their pre-recession economic health, while others have yet to fully regain their footing 6 years after the official end of the recession.



The question

1. How many months will it take for any US county to return to within 1 percent of its pre-recession unemployment rate?

Calculate Response variable - the answer, if known

Step 1

Step 2

Step 3

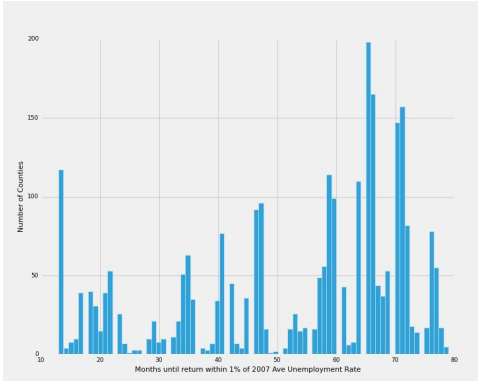
The Source

My unemployment data comes from the Bureau of Labor Statistics and includes the estimated unemployment rate for every county or county equivalent between 2006 and July 2015.

Calculate my response

I found the first month that the unemployment rate dropped below the average rate in 2007 plus one percent for each county. I used a 3 month moving average to help ignore anomalous months.

Explore



Hint: Its Seasonal!

Feature Variables

Step 1 → Step 2 → Step 3

Types of Sources

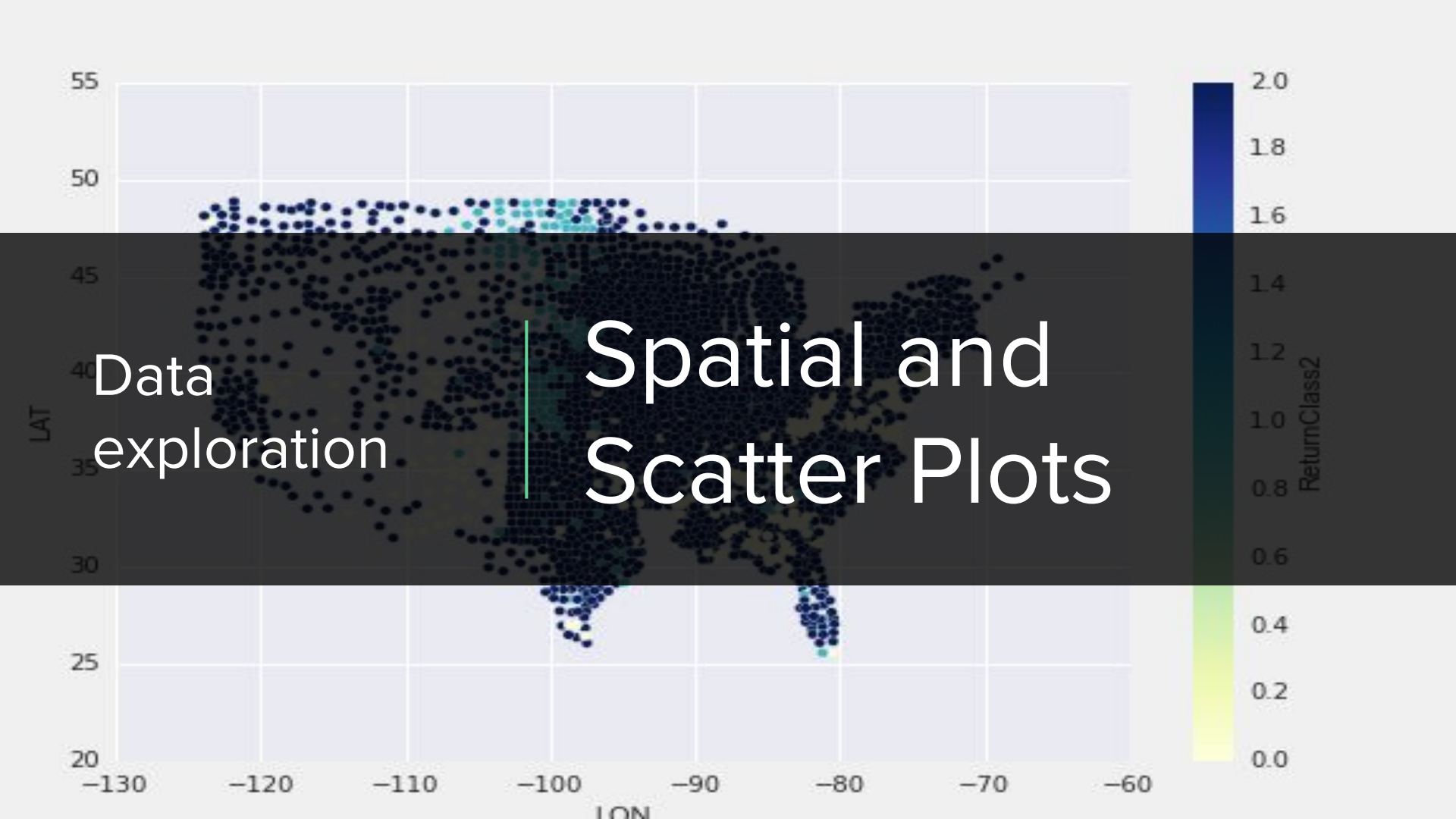
- Demographic (Census)
- Education (Census)
- Industrial Typlology (USDA)
- Economic (HH income, stimulus received)
- More to come!

Feature Management

Merges were used (as well as some old school spreadsheet work) to join my feature data to the response dataset. Each county has a unique FIPS code or I used concatenated county, state names.

Feature Issues + Engineering

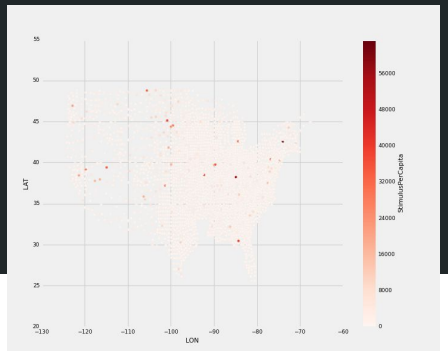
In several cases data was not available for every county.



Data exploration

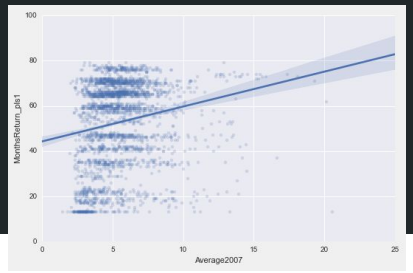
Spatial and Scatter Plots

Explore Trends



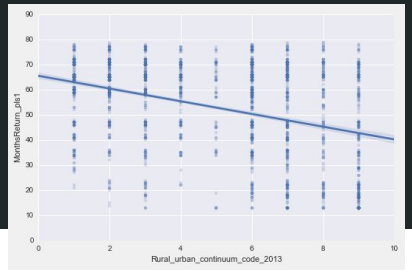
Stimulus Per Capita

There are significant outliers in our data for example Albany, NY received over $62,000 per resident, while most counties got less than $1,000



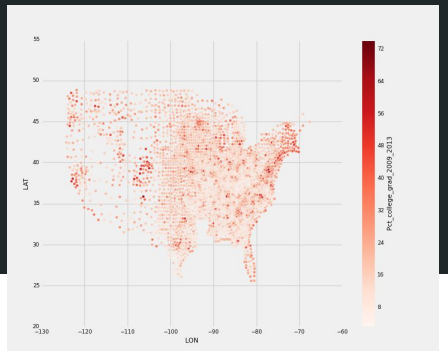
2007 Rate

Counties with higher unemployment rates in 2007 have taken longer to return to pre-recession employment than ones with lower rates.



Urban-Rural Continuum

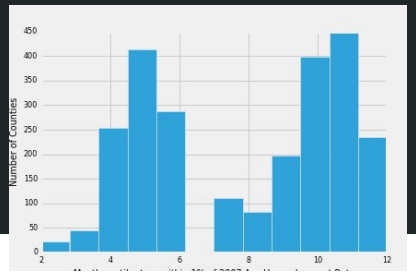
Rural counties seem to have (surprisingly) taken less time to return than urban counties, I doubt this relationship is linear though.



% College Grad

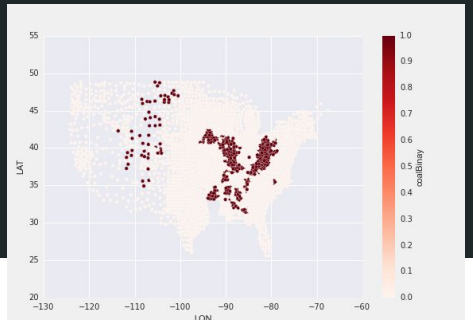
There doesn't seem to be a simple relationship between counties with many college grads and their comeback time.

Explore Trends Part 2



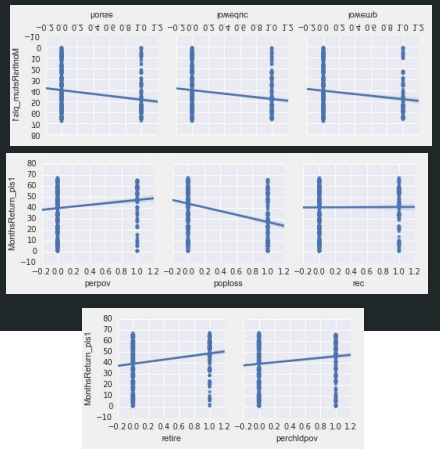
Month Returned

There is a clear monthly trend with two yearly peaks -- especially when I removed the first month in the dataset.



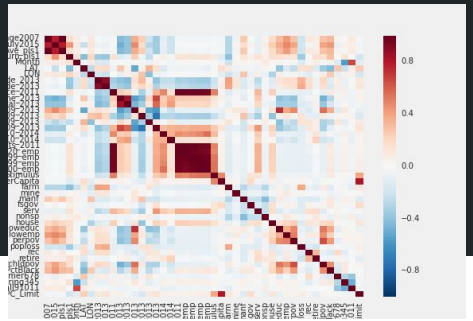
Coal Counties

The price of coal has dropped as regulation has increased over the previous several years, probably indicating a slow return to pre-recession employment



USDA Policy Codes

Some of the many USDA policy code types (persistent poverty, population loss, etc) were useful in determining when a county would return.



Correlation Matrix

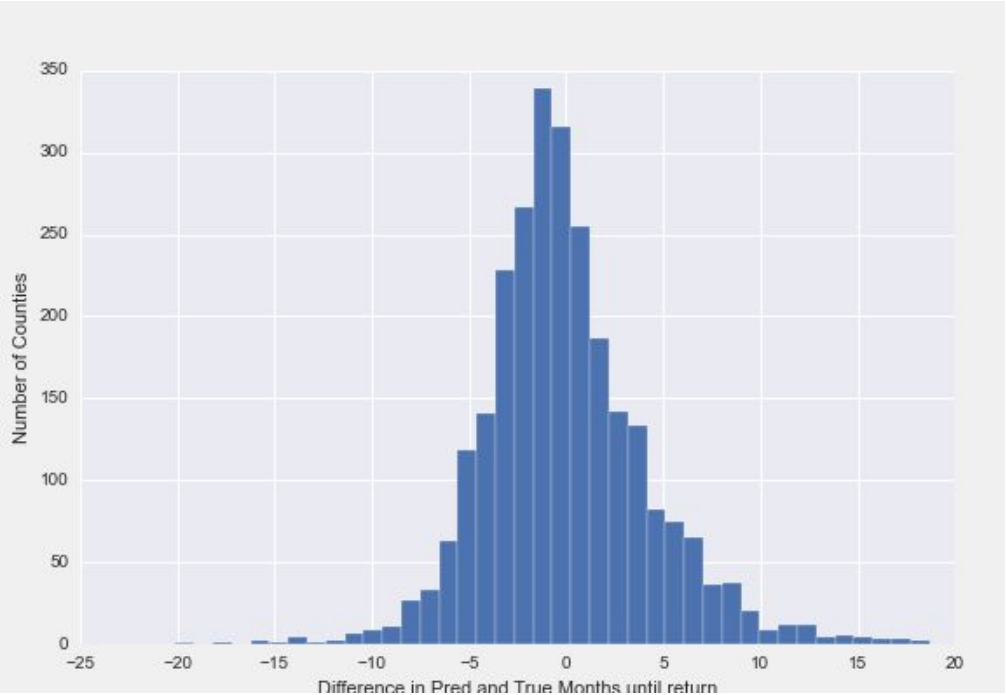
My most useful graphic was the above correlation matrix.

Results: How long will it take for a county to return to within 1% of pre-recession unemployment?

My best model correctly predicted the number of months it took, within 6 months, for 88% of counties to return to pre-recession unemployment.

After trying several regression methods, including linear regression and regression trees, with several feature sets my best model was a random forests regressor using all of my features where the best RMSE (root mean square error) of residuals was 11.6.

- 23% correct within 1 Month
- 57% correct within 3 Months
- 88% correct within 6 Months



What next?

Residuals don’t seem to be spatially clustered, which is a sign that there may not be many additional spatial variables that are out there that could explain the outstanding differences.

Research on individual outliers might help explain their unique circumstances and could be incorporated into further modeling.