Statistics 133 Project Report:

A Statistical Analysis of the Subprime Mortgage Crisis

Team Name: Data Friends

Members: Alex Wang, Sam Choi, Amanda Wu (Yanqian Wu)

[George Soros] noted, the financial crisis is beginning to have serious effects on the real economy, adding: "The extent of that is not, in my opinion, yet fully recognized."

Reuters (New York), April 9, 2008

1. Introduction

The subprime mortgage crisis that began in August 2007 has been considered by many to be the worst financial crisis since the Great Depression. The subprime mortgage crisis came to the public's attention when a steep rise in home foreclosures went out of control in 2007, triggering a nationwide financial crisis that went global within the year. The impact of the subprime crisis was worldwide, affecting millions of people, a multitude of domestic and foreign businesses, and the entire global economy. Consumer spending was down, the housing market had plummeted and the stock market was shaken. At the time of the crisis, most of us did not fully understand what was going on. Until now, with the statistical methods and analytical skills we learned in class, we hope to look back at the crisis that had such a huge impact on the real economy and on people's lives. We think it is important to understand the causes of the crisis, observe and find potential predictors of the crisis on data visualization techniques and explore the bigger effects that the crisis had and still has on the overall economy.

Specifically, in this report, we study how the subprime crisis may spill over from the financial sector to the real economy. In particular, we are interested in: 1). was it possible to predict the subprime mortgage crisis? If so, what were the identifiable signals that preceded the crisis? 2). What were the effects of the subprime mortgage crisis on the overall economy? Additionally, how did it affect industries outside of the housing sector? 3). What can we infer from the relationship between different securities and statistics?

In answering the questions proposed, we first collected data from a number of reliable resources. To get to the deeper effects on the economy, we decided to break down our analysis into three time segments that include the immediate effects from 2007 to mid 2008, and short term effects from mid 2008 to 2010, and lastly, long term effects from 2010 and onwards. Within our immediate effects from 2007 to mid 2008, we selected housing related securities such as mortgage back securities FNMA & FMCC and asset back security ABX Indexes and plotted them against time. Those securities played a central role in the financial crisis and wiped out trillions of dollars, and we want to use our graphs to show the exact time when they started to drop. In addition, we looked at the correlation between delinquencies and foreclosures as a direct chain effect after the crisis as people were unable to make payments on their mortgages.

Moving on to the short term effects from mid 2008 to 2010, we first looked at housing related industries, specifically construction spending. Not only construction spending is directly linked to the housing market, it can also be used to predict upcoming GDP numbers as construction investment is a factor in GDP calculations. Furthermore, we want to find correlations between the housing market in the U.S. and the overall financial market. Therefore, we chose SPY (SPDR S&P 500 Trust ETF), which is a proxy for S&P 500 and serves as one of the main bench marks of the U.S. equity market and indicates the financial health and stability of the economy. We will also be looking at VIX – CBOE Volatility Index, which measures uncertainty and fear in the market. VIX shows the market's expectation of 30-day volatility and is a widely used measure of market risk. Lastly, we have exchange-traded funds (ETF) that track an index, such as a stock index or bond index. Additionally, we want to look closely on the unemployment rate in the U.S. as it is a lagging indicator that generally rises or falls in the wake of changing economic conditions. We achieve this by comparing unemployment rate in 2006 and 2009 by state, by age and by ethnicity. We want to compare the time when these changes happened with the time of the subprime crisis.

As we notice some signs of world economy recovery, it is as important to recognize long term effects of the financial crisis from 2010 until now. we decided to look at the changes in interest rates overtime and we will measure the economic growth by looking at GDP, House Price Index and Layoffs vs. New Hires through time series.

It cannot be denied that the financial collapse had widespread effects on many sectors in the U.S. market. Through this report, we hope to reflect on the causes and discover the tight correlations between each sectors within the bigger economy.

2. Data Collection & Processing

In order to address the primary questions prompted by our topic, we looked at a diverse set of securities and statistics. A full list crediting the sources of the datasets used in this project is provided in the appendix of this report. The data on securities were obtained from: Yahoo Finance, Bloomberg, and the Chicago Board Options Exchange. The data on various statistics were obtained from: the U.S. Department of Housing and Urban Development, U.S. Bureau of Labor Statistics, and the Economic Research division of the Federal Reserve Bank of St. Louis.

The historical stock data obtained from Yahoo Finance were SPY (SPDR S&P 500 Trust ETF), FNMA (Federal National Mortgage Association), and FMCC (Federal Home Loan Mortgage Corporation). These datasets were formatted as csv files, and thus could be directly read into R. The data obtained using a Bloomberg terminal were DLQTDLQT (Delinquency Rates), FORLTOTL (Foreclosure Rates), ABX-HE-AAA-S6-2 (Index for AAA mortgage loan insurance derivatives), and ABX-HE-BBB-S6-2 (Index for BBB mortgage loan insurance derivatives). These datasets were formatted as xlsx files, and were first processed and converted to csv files. Any extraneous information and lines that did not contribute to forming the related R data frames were removed, and the csv file was read in as a data frame that could be manipulated in R. The data obtained from the Chicago Board Options Exchange was historical data on VIX (Volatility Index of S&P 500 Index Options). This dataset was formatted as a xslx file, and was similarly cleaned and converted to a csv file that was then read into R as a data frame.

The historical data obtained from the U.S. Department of Housing and Urban Development was a dataset of Mortgages Outstanding. This data was obtained as a csv file, and was subsequently read into R as a data frame. The data obtained from the U.S. Bureau of Labor Statistics were several datasets of unemployment rates in the U.S. by county in various years. These datasets were provided as csv files, and upon reading them all in as data frames, the data corresponding to 2006 and 2009 were chosen as the best indicators of the rise in Unemployment following the subprime mortgage crisis. The data obtained from the Economic Research division of the Federal Reserve Bank of St. Louis was historical GDP data, which was read into R as a csv, and plotted as a time series.

For all data in the form of securities, we processed the files obtained into csv files, which were then read into R as data frames. In order to visualize this financial data, we retained the information corresponding to Date and Closing Price (or Last Price), and formatted the dates to observe the data at regular intervals. A similar method was applied to clean the data for delinquency and foreclosure rates, and the Rate (percentage) was plotted against Date to form a time series. For the unemployment data, the dataset provided unemployment rates by year and county. Depending on the desired year, the data was first filtered based on the Date. In order to address the unemployment rate for each state, the group_by function was used to group all counties within the same state together. Thus, following these steps, the unemployment data for each year was represented by a dataframe with 50 cases to represent the 50 states mapped.

The code is provided in the appendix, which details all the specific methods and strategies used in R to clean the datasets, as described above. Much of these methods reflect what we've learned in this course, and a few additional methods were learned by exploring concepts that we touched on, but didn't fully utilize in this course. The code also shows how these plots were saved as image files to assist the visualization of our data throughout this report.

3. Analysis and Data Visualization:

3.1 Preface:

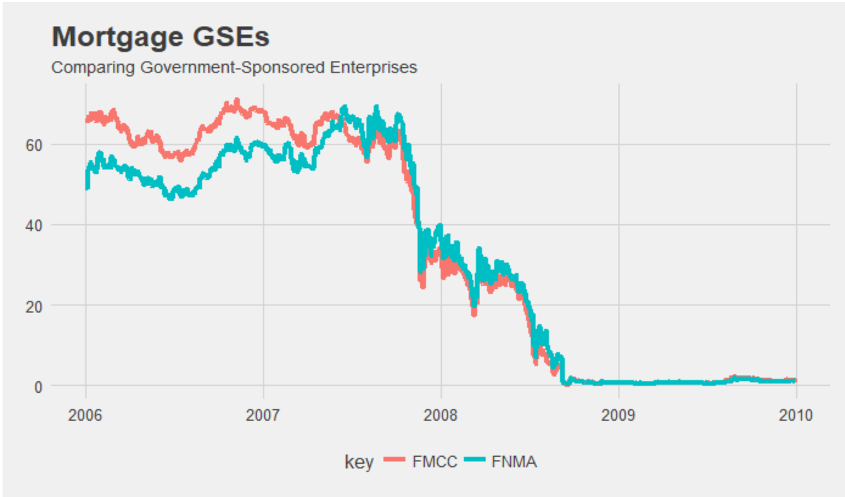
When analyzing our data and observing significant movements and potential relationships, we framed our search in three specific categories: immediate impacts (2007 to mid-2008), short-term effects (mid-2008-2010), and long-term effects (2010 and onwards). We wanted to observe the impact of subprime mortgage crisis in three stages because this allowed us to see whether events were correlated with one another (one event triggered the next event) as well as which sectors and industries were impacted at what times. As mentioned earlier, we framed our findings and observations in the context of three sequential questions:

1) Was it possible to predict the subprime mortgage crisis?
2) What were the effects of the subprime mortgage crisis on the overall economy? Additionally, how did it affect industries outside the housing sector?
3) What can we infer from the relationships between different securities and statistics?

3.2 Data Analysis:

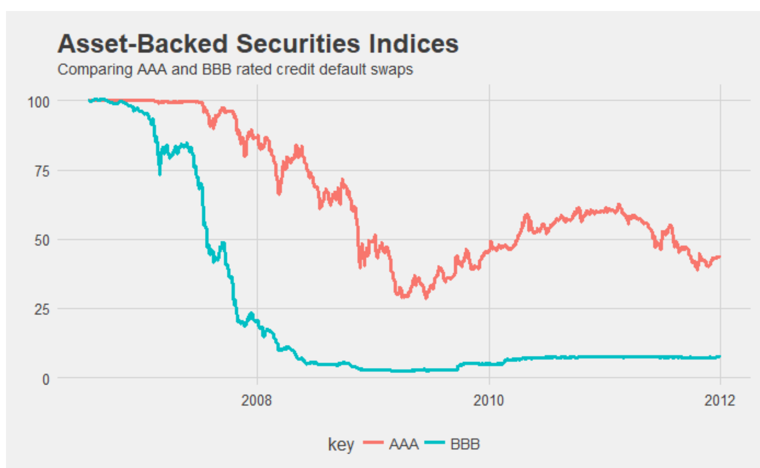
<u>Section A: Immediate Effects</u>

We identified key assets and securities that derived their value from the value of a mortgage or a home as well as performance indicators of said mortgage and home. We found



that there were two types of proxies available for valuing the housing market, securities-based assets such as stocks, Asset-Back Securities (ABS), and Mortgage-Backed Securities (MBS) as well as homeowner performance indicators such as delinquency rates and foreclosure rates.[1] To find the value of the MBS market, we tracked the financial performance of the Federal National Mortgage Association (FNMA or "Fannie Mae") and the Federal Home Loan Mortgage Corporation (FMCC or "Freddie Mac"), both of which are government sponsored enterprises (GSE). We obtained the daily prices of both securities over the course of 4 years from 2006 to 2010, where we proceeded to represent the data as a time series, where observed significant movement within the time period. Towards the end of 2008, there was sharp decline in price for both FNMA and FMCC which reflected the increase in delinquency and foreclosure rates. From external research, we found that around this time the adjustable rates were beginning to kick in and the attractive introductory rate that enticed people to take out a loan (sometimes multiple loans simultaneously) to buy a new home was replaced by a substantially higher interest rate. As a result, many of those who took out adjustable rate mortgages were now unable to pay back their debt, ultimately defaulting on their financial obligations. Because FNMA and FMCC both derive their value directly from the value of mortgages, the rising default rate coupled with the growing volume of adjustable rates starting adversely impacted the price of housing market securities. With debtors unable to make their payments the value of the mortgages dropped, accelerating as default rates increased.

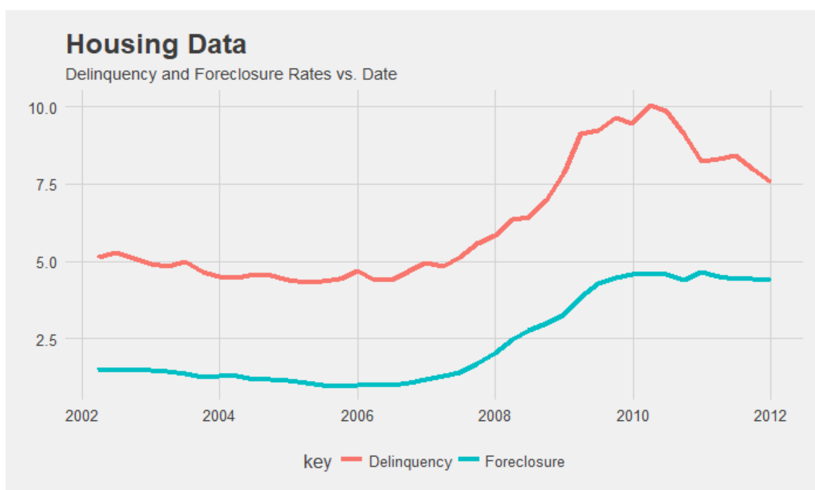
Further, we also we analyzed the prices over time of ABS Indices, specifically AAA and BBB rated credit default swaps, which served as a proxy for the valuation of Subprime Mortgages. Around the same time that FMNA and FMCC prices were facing sharping declines, these ABS products were also drastically falling, with BBB suffering the first largest drop before 2008 and AAA following suit shortly after. This ties back



[1] Delinquency is a situation where a borrower is late or overdue on a payment such as a mortgage, auto loan, etc. Foreclosure is a situation where the lender seizes the property as a result of the borrower not making adequate payments.

into the macroeconomic trend highlighted earlier that many of the adjustable mortgages were high-risk loans, thus warranting a BBB rating.[2] Because individuals were unable to make their mortgage payments, effectively defaulting on their loans, MBS products, which sustain their high value when a certain threshold of mortgages are paid for, were declining in valuing. As more and more homeowners were failing to make their payments, the value of the MBS products also declined since the pooled mortgages were paying out less while the number of defaulters was increasing (increasing risk and decreasing return). Consequently, the demise of BBB rated mortgages inevitably impacted AAA rated mortgages. Because MBS products bundle mortgages from AAA to BBB ratings as a means of diversifying risk, the devaluation of BBB mortgages lowered the value of the encompassing MBS, thus adversely impacting the value of AAA mortgages, albeit in a more delayed fashion. This is clear in the comparison of the AAA and BBB rates credit default swap time series graphs, where we identified a considerable time delay in value decline between the two securities.

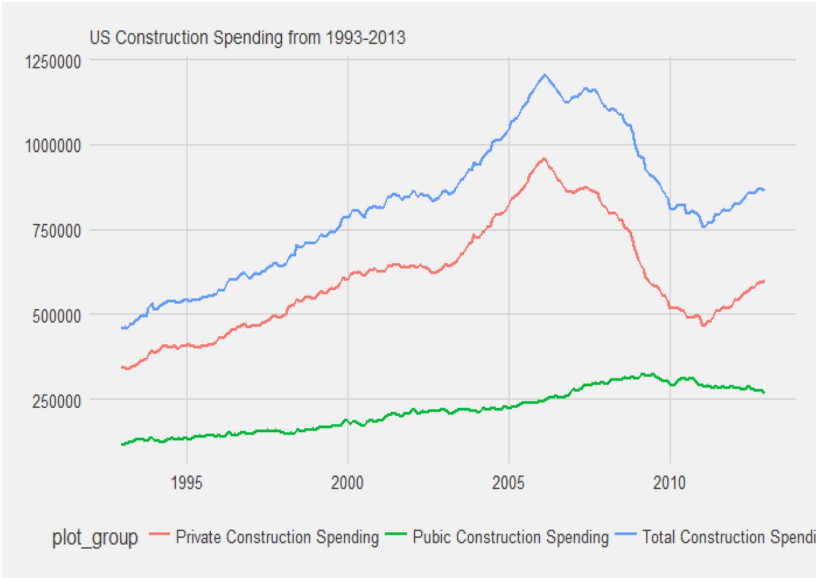
Finally, we hypothesized that homeowner performance indicators such as delinquency rates and foreclosure rates were among the immediately impacted areas. We found that from mid-2007 to mid-2008, these two rates began to climb significantly, substantiating our findings that the adjustable mortgages were causing homeowners default on their mortgages, thus causing a significant increase in delinquency rates and foreclosure rates. We also explored the correlation between the two performance indicators and found that foreclosure rates and delinquency rates were highly correlated with a Pearson's correlation coefficient of 0.7421 and a p-value of 2.2e-16, suggesting that the result is statistically significant. Intuitively, this makes sense because as delinquencies increase, foreclosures should also increase because foreclosures are triggered and initiated by delinquencies. Essentially, increase in mortgage defaults caused the deterioration of homeowner performance indicators, drastically increasing the rate in which homeowners were either delinquent with their payments or had their home foreclosed.

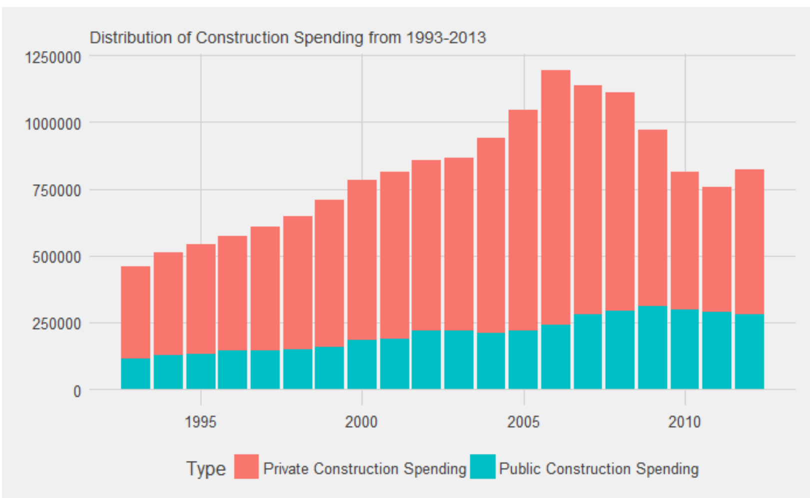


<u>Section B: Short-Term Effects</u>

Much of the economic turmoil that came as a result of the subprime mortgage crisis occurred in the following year (mid-2008-2010) as many of the immediate consequences had taken effect. Largely, the short-term impacts can be broken down into housing market related impacts, systemic market impacts, and affected consumer behavior. Due to the loss of value in

[2] For clarification, AAA is the highest rating any debt product can receive and, as such, is typically considered as the safest with regards to defaulting – or inability to pay back financial obligations. Consequently, BBB is the riskiest investment-grade fixed income security, which carries a high payout and high risk. Products rated below BBB are called as "junk" and have a large risk profile.



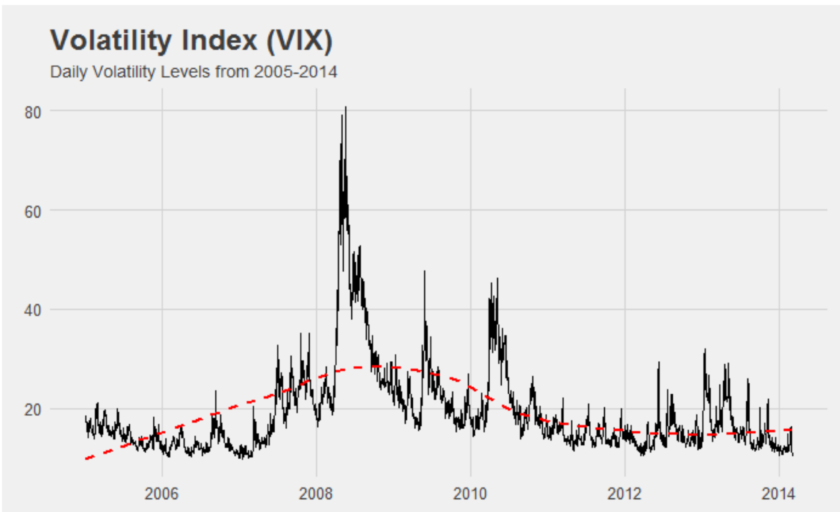


houses, construction projects began to dampen as many investors and land developers were anxious to enter into the economic fray. Collecting data on both private and public construction spending, we analyzed the data in two forms, one as a multi-variant time series across 2 decades and the other as stacked bar graph. We chose to analyze our data this way to observe the time-elapsed change in spending as well as the proportionality of each spending for within the time frame. Evidently, private construction spending enjoyed a long bull market and was largely insulated from past financial bubbles – most notably the bubbles in 1996 and 2001. However, because homes were rapidly losing values in the wake of the subprime mortgage crisis, there was a lack of liquidity in the housing market and not enough new investor who were interested in funding real estate projects or building new homes. As such, private construction spending plummeted as the interest in the house market dropped. Alternatively, public construction spending remained relatively flat in the same time period, slightly increasing during the financial crisis. We reasoned that this was the result of both the U.S. government intervening in the midst of the crisis to increase infrastructure spending and create employment opportunities in the public sector as well as high insulation of the public sector from economic crises.[3] Imperatively, the sharp declines in private construction spending first occurs in early 2008, and continues to accelerate into 2009, or the height of the following financial crisis. This confirms our claim that this is a short-term claim because of the considerable time delay from the initial effects of the subprime mortgage crisis. Further, the proportionality of public to private construction spending was significant in the subprime mortgage crisis time period (2007 – 2010). In this time frame, it is apparent that private construction spending declined and public

[3] Interesting observation that public construction spending increase in the height of the financial crisis. This finding is substantial because it gave us insight into how public goods can be necessarily shielded from economic turmoil. We explored to find that traditionally, government-funded projects are usually well insulated and only suffer from budget cuts.

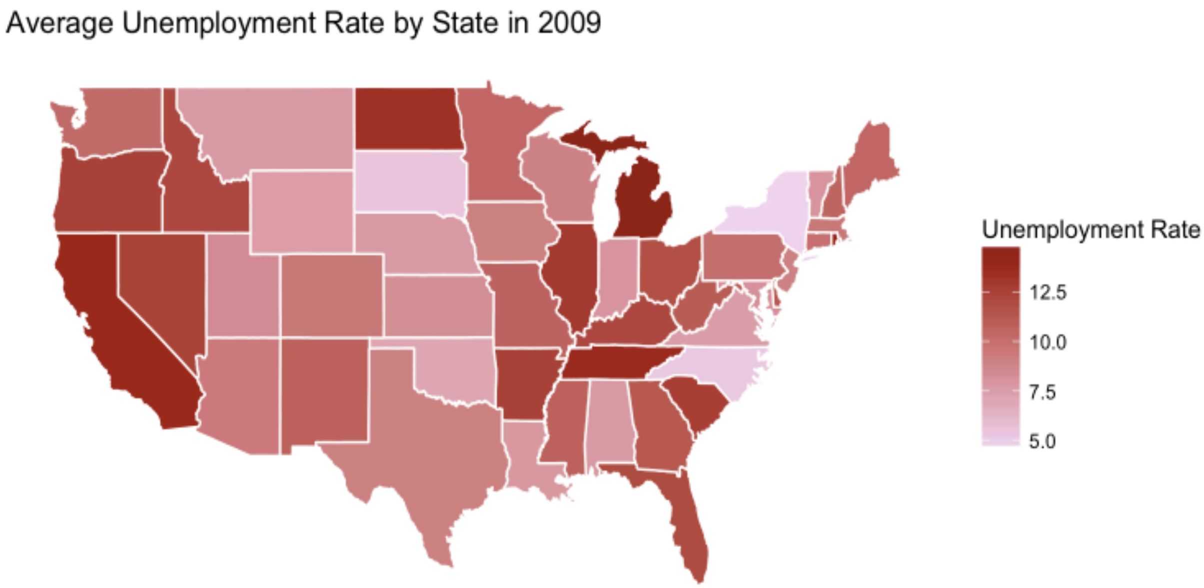
construction spending held its ground and marginally captured proportional gains. This is explained by the shielding of government-sponsored projects found in public construction as well as the drop in interest and funding in new construction projects.

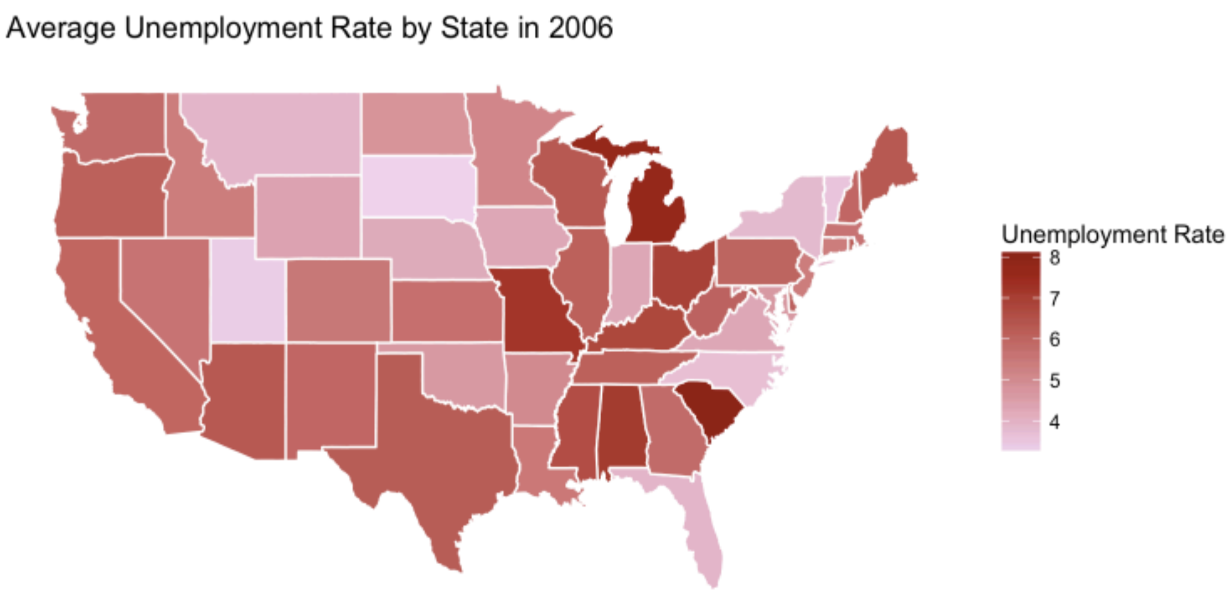
Looking at the overall market, we see that the subprime meltdown had a significant impact on the financial performance of the rest of the market. Tipping us off on the chaos in the market, the Volatility Index (VIX), begins to rise towards the end of 2007 and peaks from 2008 to 2009. This was surprising because we expected the VIX to expand earlier in the fiscal year of 2008. However, upon further investigation, we found that the significant delay was due to a sudden outburst in market uncertainty, largely attributed to the collapse of the global investment bank, Lehman Brothers as well as fear of financial crisis (ironically this would be a self-fulfilling prophecy). Considered "too big to fail," Lehman went bankrupt which had large financial repercussions on its related counterparties (entities that hold the other side of any given contract), which were typically other banks. At the time, MBS products were extremely popular and profitable for large banks because they were highly leveraged relatively low risk (no one thought the housing market could possibly collapse). As such, banks often held these products in their portfolios for extended periods of time that put the firms at an extremely risky position and traded it to other banks, increasing their credit exposure across multiple firms. As a result, the collapse of Lehman, which at the time of the failure was the 4th largest U.S. investment bank, generated large amounts of uncertainty in the rest of the financial market – and by extension the rest of the nation's and global economy. This is captured in the VIX time series, where we see heighten levels of volatility surrounding the explosion witnessed towards the middle-end of 2008.



Additionally, market sectors were hard hit by the fallout of the subprime mortgage crisis. Specifically, we looked SPDR S&P 500 Trust ETF (SPY), which is a proxy for the S&P 500 Market Index as well as a select class of sector ETFs that tracks the movement of various market sectors (i.e. consumer discretionary, financials, etc.). Analyzing the daily price of SPY as well as the class of sector ETFs, we found that all assets declined roughly around the same time as the spike in volatility seen in the VIX time series. Constructing a time series for SPY with a fitted regression curve as well as a multivariate time series for the group of sector ETFs, we were able to identify this event and confirm our expectation that the subprime mortgage crisis did indeed have a delayed impact on the rest of the market; because the meltdown was initially isolated to only the housing market, it made sense that the rest of the market was not significantly impacted immediately. However, as various counterparties defaulted and uncertainty and fear in the market skyrocketed, the overall market buckled under the stress and ultimately gave out in declined drastically in value, with some sectors not recovering their sustained losses. This is substantiated by the downward movements of both SPY and the sector ETFs during the later portion of 2008, which confirms our initial thoughts.

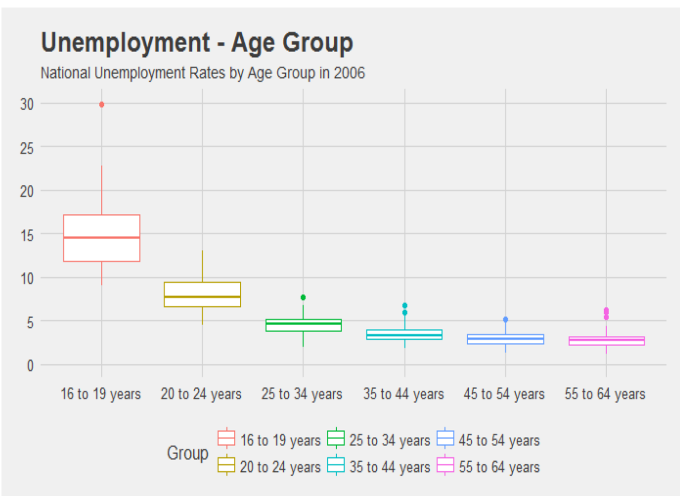
In the public sector, the impact of the subprime mortgage crisis was damaging to public, particularly consumers. To best capture the impact the crisis had on consumers, we explored the unemployment rate and how various groupings were impacted by the crisis, identifying and looking into various explanatory variables such as age group, ethnicity, and region.

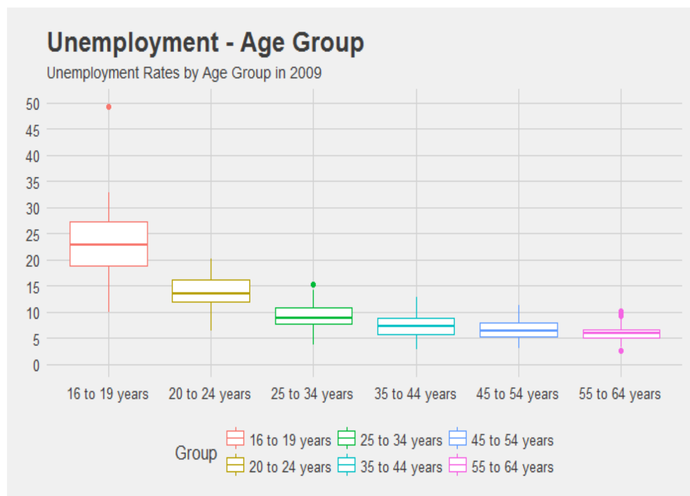


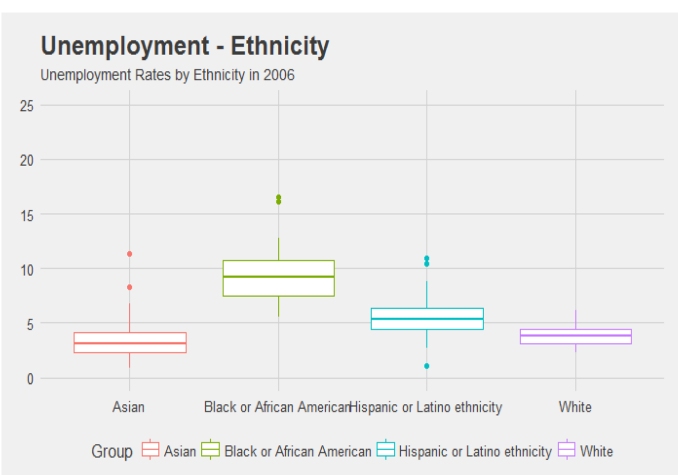


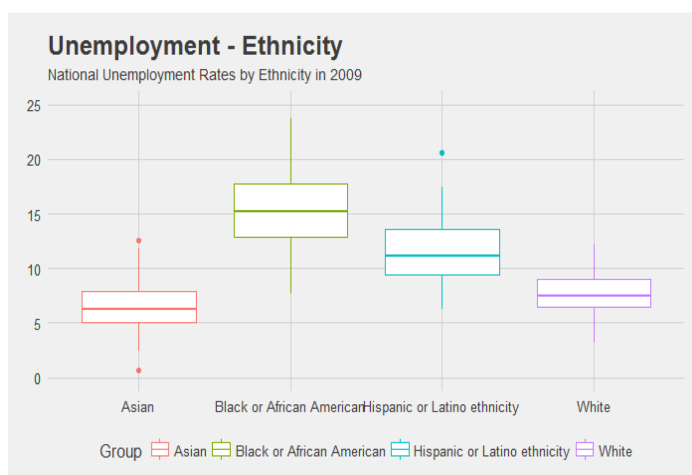
When constructing our argument, we suspected that unemployment increases were not equally distributed – that is, some groups will lose more than others. In essence, although all groups would be impacted by fluctuations in unemployment, some groups were inherently at a disadvantage, thus more probable to be harder hit by events such as the subprime mortgage crisis. To begin, we took a top-down approach and first looked into unemployment rates by state. as Michigan jumping from approximately 8% unemployment to more than 15%. This drastic change is not necessarily intuitively captured by looking at only the map only because the while the scale changes, the color distribution across the two graphs is generally the same.

Furthermore, we also looked into other potentially telling classifiers such as age and ethnicity. We used boxplots to capture the disparity across various groups. From there, we found that the financial crisis induced unemployment increases was not proportional over various groups, for instance, we found that while all groups increased by a considerable amount, age groups below 35 took on a larger share of the unemployment increase. Similarly, ethnic We found that while the distribution of unemployment rates across the U.S. was roughly the same pre-crisis and post-crisis, the actual unemployment rates increases overall with states such minorities, namely Hispanics and Blacks, were hit harder than the other ethnic groups in terms of unemployment rate increase, with the median of both groups increasing from approximately 6% and 9% to 12% and 16%. This confirmed our original suspicions that unemployment rates did not increase proportionally equally from pre-crisis to post-crisis. This is an important finding because it allowed us to observe how economic turmoil impacts various socioeconomic groups and, by extension, societal stakeholders. From our data visualization, we were able to pinpoint these discrepancies and better understand the societal impact of the subprime mortgage crisis.

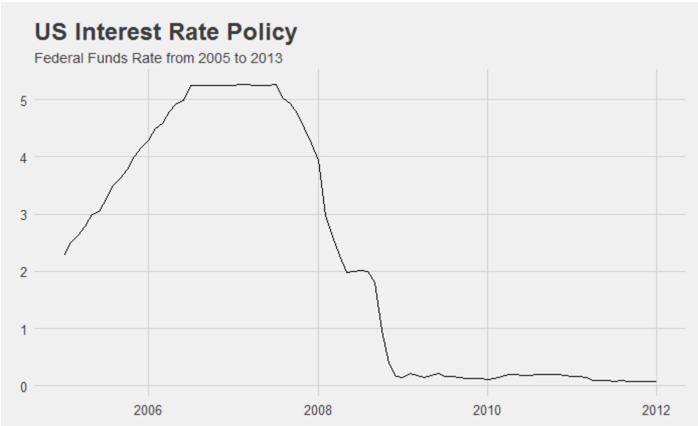




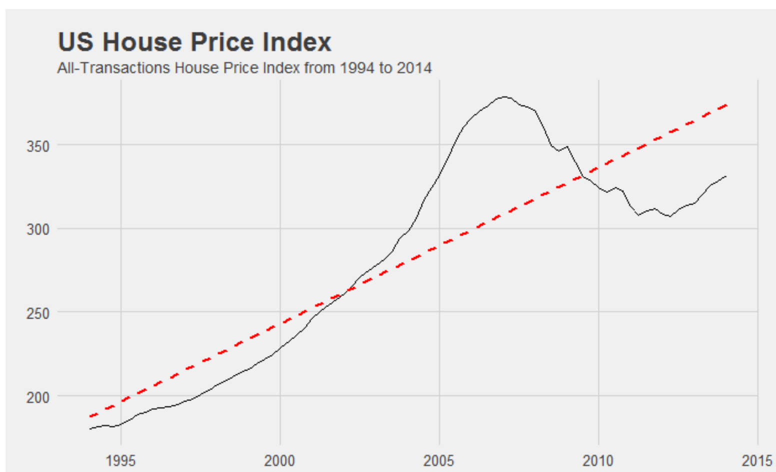




Section C: Long-Term Effects



The financial crisis had many significant long-term effects, largely in economic policy and growth in the U.S. To further explore the long-term effects of the subprime mortgage crisis, we categorized our findings into type classifiers: economic policy and economic growth. When analyzing data relevant to economic policy, we found that the most significant change from the financial crisis was the movement of the Federal Funds rate, which is a tool used by the Federal Reserve to dictate monetary policy.[4] Because the economy was now in a recession, the Fed worked to lower interest rates as a means of stimulating economic growth by having the banks loan out more money to other businesses and consumers. As such, we observed the interest rate drop from more than 5% pre-crisis to less than 0.25% from 2009 and onwards. This is significant because at this point, the Fed's monetary policy reached the zero-lower bound, which refers to the event when a country's central bank's interest rate reach a value of 0.25% or lower.[5] Further, we found that while the interest rate was at its lowest theoretical value, economic growth in the U.S. was still sluggish and not optimal. Ultimately, the subprime mortgage crisis was able to impact the country's long-term economic focus because the economy was now broken and recovering. From our findings, we noticed that the zero-lower bound interest rate persisted even after 2010 when the crisis had ended. This confirms our expectation that the crisis had far-reaching consequences and virtually damaged the economy beyond quick repair.



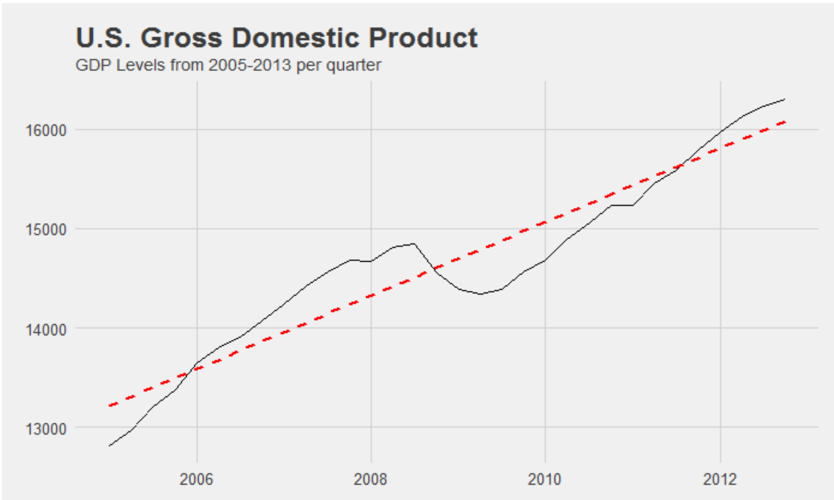
Furthermore, long-term economic growth was also largely

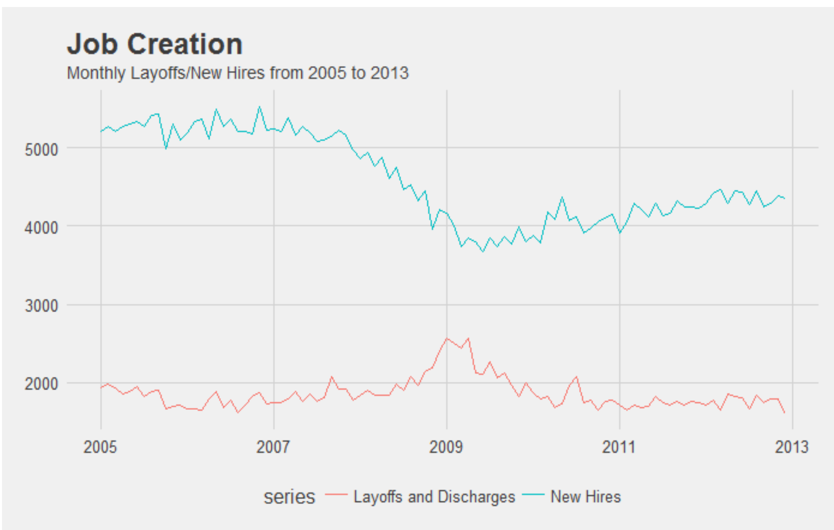
---

[4] The Federal Funds Rate is the bread and butter of the U.S. financial system. This interest rate is important because it is the rate at which the Federal Reserve charges banks that borrow money from the Fed. In essence, money is printed when a bank borrows extra money that it in turns loans out to its customers. In an apt fashion, a low interest rate policy (also known as an expansionary monetary policy) encourages economic growth at the risk of inflation growth where as a high interest rate policy (also known as a contractionary monetary policy) slows economic growth at the risk of unemployment growth.

[5] The zero-lower bound is an important concept in macroeconomics because this is the theoretical extent of an expansionary monetary policy. An interest rate that is below 0 is damaging to an economy because the implication of a negative interest rate is that the lender is actually paying for the borrower to borrow money.

impacted by the financial crisis. When we explored the effects of the crisis on economic growth, we identified three key indicators of economic growth: U.S. GDP, the U.S. House Price Index,





and job creation. Moving forward from economic policy, we suspected that economic growth would not only have a considerable time lag from the initial effects of the subprime mortgage crisis, but also have one of the longest impact time frames (highest duration of persistence). With regards to GDP, we found a considerable slump in the curve, illustrated in the graph below. We further investigated the GDP time series and ran a linear regression on the curvature of our GDP time series. We found that while GDP had generally grown at a constant rate, there slight decrease mid-2008 to mid-2009, which coincides with the height of the financial crisis. This is extremely significant because while GDP had been largely constant in growth, the subprime mortgage crisis caused a decrease in economic output for the U.S. Additionally, we also looked into the total U.S. housing market and found a similar behavior seen in our GDP time series. When analyzing the US House Price Index, we expanded our selected time frame to 2 decades, which gave us insight into how the index performed during other crisis periods such as the Asian financial crisis (1997) and the Dotcom bubble (2001). We found that the housing market was mostly shielded and insulated from these economic crises because mortgages were still considered safe bets.[6] However, for the subprime mortgage crisis, the cause was in faulty mortgages and soaring house valuations, which ultimately created a bubble in the housing market as seen around 2006-2007 where the index reaches its peak. During the meltdown of the subprime mortgage market, the index drastically dropped, eventually reaching its trough around 2012. This is eerily similarly to our GDP time series because it demonstrates the long-term impact on growth and recovery that subprime mortgage crisis had on the housing market. Finally, we explored employment in the U.S. since job growth is a robust indicator for economic

[6] This idea is still prevalent in today's financial markets. Because homeowners typically make their payments, mortgages are still considered safe bets, relative to other market instruments available.

growth and consumer spending power. We found that the level of new hires reached its lowest point early into 2009 when the market was facing the repercussions of a failing housing market, which coincides with when the level of discharges/layoff reach its highest point. This is consistent with our expectation that the financial crisis had a significant impact on the rate of job growth during this period of economic turmoil. Upon further investigation, we also found that from its trough, the level of new hires marginally picks up, suggesting that the long-term impact of the subprime mortgage crisis on job creation is considerably powerful. Essentially, the subprime mortgage crisis had a lasting impact on U.S. economics with regards to policy and growth.

4. Concluding remarks

When considering the scope of our analysis and the categorization of our observations, we re-framed our findings back into the previous three questions. Summarized, here is what our findings told us in the context of these questions:

1) Was it possible to predict the subprime mortgage crisis
Answer: While the subprime mortgage crisis and its subsequent effects occurred in quick succession, all these impacts were not all unforeseeable. If we divide the observations by their respective time frame of effect, we can see that there are telling signs of causation, with the effects of one stage impacting the effects in another. For instance, the increase in the number of adjustable rate mortgages ending their attractive introductory rate increased the risk profile of MBS products and decreased the quality of the overall mortgage pools. Thus, the subprime mortgage crisis could be predicted, provided that the right explanatory variables were observed and analyzed.

2) What were the effects of the subprime mortgage crisis on the overall economy? Additionally, how did it affect industries outside the housing sector?
Answer: The deteriorating state of the MBS products, signaled by the collapse of ABS related securities and the stock prices of FNMA and FMCC, triggered an initially slow but gradually accelerating market failure with the bankruptcy of notable companies that depended on the lack of delinquency and foreclosure rates. The subsequent decline in the market ultimately impacted the bottom lines of many business, thus forcing many enterprises to cut employees and spending, triggering an increase in unemployment across all groups. On a high level, the subprime mortgage crisis had an encompassing impact on the overall economy because many players in the housing market were counterparties to other financial dealings with other business entities across various industry sectors. We see that this impact occurred in two movements. The first movement was the direct decline of the housing market, impacting all stakeholders that had "skin" in the housing market. Further, these housing market stakeholders were most likely also counterparties to other transactions in different industries, which became a substantial issue when many parties were unable to pay back particular financial obligations, initiating a snowball effect on the rest of the economy. Conclusively, the subprime mortgage crisis caused the overall economy to recede and decline; the market suffered the loss of trillions of dollars in capitalization (equity value loss) whereas consumers suffered from an increase in unemployment.

3) What can we infer from the relationships between different securities and statistics?

Answer: As with any statistical challenge with regards to data analysis, we are concerned with identifying key correlations as well as finding robust and statistically significant relationships between various exogenous variables. Throughout our analysis and observations, we investigated a multitude of variables and relationships to pinpoint and explain events that may or may not be intuitive to the general audience. From this particular topic, we found that relationships have a distinct structure: there are strengths of relationships as well as scope of relationships. With strength, we looked at how directly correlated on event was to another event and sought to explore the degrees of connection, that is, how many intermediaries existed between the stimuli and the response. With scope, we looked at how many connections one event may have to other related events or if one event was the stimuli of another single event or the trigger of a group of potentially related events. Ultimately, we were able to infer this two-pronged idea of relationships and we applied it to our project, coming to the realization that all the events that happened throughout the subprime mortgage crisis as well as before and after were all connected in a considerable way.

Appendix

<u>Data Sources:</u>

- Yahoo Finance (https://finance.yahoo.com/lookup)
    - SPY (SPDR S&P 500 Trust ETF)
    - FNMA (Federal National Mortgage Association; *Fannie Mae*)
    - FMCC (Federal Home Loan Mortgage Corporation; *Freddie Mac*)
    - XLY (Consumer Discretionary Select Sector SPDR ETF)
    - XLP (Consumer Staples Select Sector SPDR ETF)
    - XLE (Energy Select Sector SPDR ETF)
    - XLF (Financial Select Sector SPDR ETF)
    - XLB (Materials Select Sector SPDR ETF)
    - XLK (Technology Select Sector SPDR ETF)
    - IYR (iShares US Real Estate ETF)
    - IDU (iShares US Utilities ETF)
    - IYJ (iShares US Industrials ETF)
    - IYH (iShares Dow Jones US Healthcare ETF)
- Bloomberg (*accessed via Bloomberg's private software application*)
    - DLQTDLQT (Delinquency Rates)
    - FORLTOTL (Foreclosure Rates)
    - ABX-HE-AAA-S6-2 (Index for AAA mortgage loan insurance derivatives)
    - ABX-HE-BBB-S6-2 (Index for BBB mortgage loan insurance derivatives)
- U.S. Department of Housing and Urban Development (https://www.huduser.gov/portal/pdrdatas_landing.html)
    - Mortgages Outstanding
- U.S. Bureau of Labor Statistics (https://www.bls.gov/data/)
    - Unemployment Rates by County (Measured per Year)
- Chicago Board Options Exchange (http://www.cboe.com/data/historical-options-data)
    - VIX (Volatility Index)
- Economic Research, Federal Reserve Bank of St. Louis (https://fred.stlouisfed.org/)
    - GDP (U.S. Gross Domestic Product)
    - Federal Funds Rate
    - Hiring/Layoff Economic Data
    - All-Transactions House Price Index for the United States

<u>Code:</u>

```
library(DataComputing)
library(ggplot2)
library(tidyr)
library(dplyr)
library(lubridate)
library(ggthemes)
library(stringr)
library(gdata)
library(maps)
library(mapproj)
# SPY
```

```
### SPDR S&P 500 Trust ETF
SPY <- read.csv("yahoo-finance-data-stat133/SPY-06-09.csv")
SPY_TS <- SPY %>% select(Date, Close) %>% arrange(Date)
SPY_TS$Date <- ymd(SPY_TS$Date)

SPY_plot <- SPY_TS %>% ggplot(aes(x=Date, y=Close, group=1)) + geom_point(size=0.1,
alpha=0.15) + geom_line(size=1.4, color = ggthemes_data$fivethirtyeight["blue"]) + labs(title =
"SPY", subtitle = "SPDR S&P 500 Trust ETF", x = "Date", y = "Price") +
theme_fivethirtyeight()
# + geom_smooth(method="auto", size=0.8)
SPY_plot



#FNMA
### Federal National Mortgage Association
FNMA <- read.csv("yahoo-finance-data-stat133/FNMA-06-09.csv")
FNMA_TS <- FNMA %>% select(Date, Close) %>% arrange(Date)
FNMA_TS$Date <- ymd(FNMA_TS$Date)
FNMA_plot <- FNMA_TS %>% arrange(Date) %>% ggplot(aes(x=Date, y=Close, group = 1))
+ geom_point(size=0.1, alpha=0.15) + geom_line(size=0.7) + geom_smooth(method="auto") +
labs(title = "FNMA", subtitle = "Federal National Mortgage Association", x = "Date", y =
"Price") +
  theme_fivethirtyeight() + theme_fivethirtyeight()
FNMA_plot

#FMCC
### Federal Home Loan Mortgage Corporation
FMCC <- read.csv("yahoo-finance-data-stat133/FMCC-06-09.csv")
FMCC_TS <- FMCC %>% select(Date, Close) %>% arrange(Date)
FMCC_TS$Date <- ymd(SPY_TS$Date)
FMCC_plot <- FMCC_TS %>% arrange(Date) %>% ggplot(aes(x=Date, y=Close, group = 1)) +
geom_point(size=0.1) + geom_line(size=0.7) + geom_smooth(method="auto") + labs(title =
"FMCC", subtitle = "Federal Home Loan Mortgage Corporation", x = "Date", y = "Price") +
theme_fivethirtyeight()
FMCC_plot


#DLQTDLQT Index
### Delinquency Rate (as % of Total Loans, Seasonally Adjusted)
DLQT <- read.csv("bloomberg-data-stat133/DLQTDLQT.csv")
DLQT_TS <- DLQT %>% select(Date, PX_LAST) %>% arrange(Date)
DLQT_TS$Date <- mdy(DLQT_TS$Date)
DLQT_TS <- DLQT_TS %>% filter(Date >= as.Date("2002-01-01"), Date <= as.Date("2012-
01-01"))
```

```
DLQT_plot <- DLQT_TS %>% arrange(Date) %>% ggplot(aes(x=Date, y=PX_LAST,
group=1)) + geom_point(size=0.1) + geom_line(size=1) + geom_smooth(method="auto",
size=0.7) + labs(title = "DLQTDLQT Index", subtitle = "Delinquency Rate (as % of Total Loans,
Seasonally Adjusted)", x = "Date", y = "Percent (%)") + theme_fivethirtyeight()
DLQT_plot


#FORLTOTL Index
### Foreclosures (as % of Total Loans, Not Seasonally Adjusted)
FORLTOTL <- read.csv("bloomberg-data-stat133/FORLTOTL.csv")
FORLTOTL_TS <- FORLTOTL %>% select(Date, PX_LAST) %>% arrange(Date)
FORLTOTL_TS$Date <- mdy(FORLTOTL_TS$Date)
FORLTOTL_TS <- FORLTOTL_TS %>% filter(Date >= as.Date("2002-01-01"), Date <=
as.Date("2012-01-01"))
FORLTOTL_plot <- FORLTOTL_TS %>% arrange(Date) %>% ggplot(aes(x=Date,
y=PX_LAST, group = 1)) + geom_point(size=0.1) + geom_line(size=1) +
geom_smooth(method="auto", size=0.7) + labs(title = "FORLTOTL Index", subtitle =
"Foreclosures (as % of Total Loans, Not Seasonally Adjusted)", x = "Date", y = "Percent (%)") +
theme_fivethirtyeight()
FORLTOTL_plot

#ABX AAA CDSI S6-2
### Asset Backed Security for AAA rated subprime mortgages
ABX_AAA <- read.csv("bloomberg-data-stat133/ABX-AAA-S6-2.csv")
ABX_AAA_TS <- ABX_AAA %>% select(Date, PX_LAST) %>% arrange(Date)
ABX_AAA_TS$Date <- mdy(ABX_AAA_TS$Date)
ABX_AAA_TS <- ABX_AAA_TS %>% filter(Date >= as.Date("2002-01-01"), Date <=
as.Date("2012-01-01"))
ABX_AAA_plot <- ABX_AAA_TS %>% arrange(Date) %>% ggplot(aes(x=Date,
y=PX_LAST, group=1)) + geom_point(size=0.1) + geom_line(size=1) +
geom_smooth(method="auto", size=0.7) + labs(title = "ABX-AAA-S6-2", subtitle = "Asset
Backed Security for AAA rated subprime mortgages", x = "Date", y = "Price") +
theme_fivethirtyeight()
ABX_AAA_plot


#ABX BBB CDSI S6-2
### Asset Backed Security for BBB rated subprime mortgages
ABX_BBB <- read.csv("bloomberg-data-stat133/ABX-BBB-S6-2.csv")
ABX_BBB_TS <- ABX_BBB %>% select(Date, PX_LAST) %>% arrange(Date)
ABX_BBB_TS$Date <- mdy(ABX_BBB_TS$Date)
ABX_BBB_TS <- ABX_BBB_TS %>% filter(Date >= as.Date("2002-01-01"), Date <=
as.Date("2012-01-01"))
ABX_BBB_plot <- ABX_BBB_TS %>% arrange(Date) %>% ggplot(aes(x=Date, y=PX_LAST,
group=1)) + geom_point(size=0.1) + geom_line(size=1) + geom_smooth(method="auto",
```

```
size=0.7) + labs(title = "ABX-BBB-S6-2", subtitle = "Asset Backed Security for BBB rated
subprime mortgages", x = "Date", y = "Price") + theme_fivethirtyeight()
ABX_BBB_plot

### Combined Visualizations
#Foreclosure
Foreclosure <- FORLTOTL_TS
Foreclosure <- rename.vars(Foreclosure, from = "PX_LAST", to = "Foreclosure") %>%
arrange(Date)
Delinquency <- DLQT_TS
Delinquency <- rename.vars(Delinquency, from = "PX_LAST", to = "Delinquency") %>%
arrange(Date)
housing_df <- merge(Foreclosure, Delinquency, by="Date") %>% arrange(Date)
housing_df_narrow <- housing_df %>% gather(key, PX, Foreclosure, Delinquency)

housing_plot <- housing_df_narrow %>% ggplot(aes(x=Date, y=PX, col = key)) +
geom_point(size = 0.1) + geom_line(size=1.6) + labs(title = "Housing Data", subtitle =
"Delinquency and Foreclosure Rates vs. Date", x = "Date", y = "Percent (%)") +
theme_fivethirtyeight()
housing_plot

Fannie <- FNMA_TS
Fannie <- rename.vars(Fannie, from = "Close", to = "FNMA") %>% arrange(Date)
Freddie <- FMCC_TS
Freddie <- rename.vars(Freddie, from = "Close", to = "FMCC") %>% arrange(Date)
mortgage_df <- merge(Fannie, Freddie, by="Date") %>% arrange(Date)
mortgage_df_narrow <- mortgage_df %>% gather(key, Price, FNMA, FMCC)
mortgage_plot <- mortgage_df_narrow %>% ggplot(aes(x=Date, y=Price, col = key)) +
geom_point(size = 0.1) + geom_line(size=1.5) + labs(title = "Mortgage GSEs", subtitle =
"Comparing Government-Sponsored Enterprises", x = "Date", y = "Price ($)") +
theme_fivethirtyeight()
mortgage_plot

AAA <- ABX_AAA_TS
AAA <- rename.vars(AAA, from = "PX_LAST", to = "AAA") %>% arrange(Date)
BBB <- ABX_BBB_TS
BBB <- rename.vars(BBB, from = "PX_LAST", to = "BBB") %>% arrange(Date)
ABX_df <- merge(AAA, BBB, by="Date") %>% arrange(Date)
ABX_df_narrow <- ABX_df %>% gather(key, Price, AAA, BBB)
ABX_plot <- ABX_df_narrow %>% ggplot(aes(x=Date, y=Price, col = key)) + geom_point(size
= 0.1) + geom_line(size=1.4) + labs(title = "Asset-Backed Securities Indices", subtitle =
"Comparing AAA and BBB rated credit default swaps", x = "Date", y = "Price ($)") +
theme_fivethirtyeight()
ABX_plot

#Volatility Time Series
```

```
#VIX
VIX2 <- read.csv("C:/Users/darzo/Downloads/vixcurrent.csv")
#VIX2$Date <- as.Date(VIX2$Date)
open <- ts(VIX2$VIX.Open, freq=365.25, start=c(2005, 1))
close <- ts(VIX2$VIX.Close, freq=365.25, start=c(2005, 1))
Volatility_Open_Price <- autoplot(open) +
  geom_smooth(aes(y=VIX2$VIX.Open), method = "loess", linetype = 'dashed', col = "red",
se=FALSE) +
  labs(title = "Volatility Index (VIX)", subtitle = "Daily Volatility Levels from 2005-2014") +
  ylab("Actual Volatility Levels") +
  xlab("Years") +
  theme_fivethirtyeight()
Volatility_Open_Price

#Time series of GDP from 2005
###GDP
GDP <- read.csv("C:/Users/darzo/Downloads/gdp.csv")
GDP$DATE <- ymd(GDP$DATE)
GDP <- filter(GDP, year(GDP$DATE) >= 2005 & year(GDP$DATE) < 2013)
GDPts <- ts(GDP$GDP, start=c(2005, 1), freq=4)
GDPts <- autoplot(GDPts, ts.size=1) +
  geom_smooth(aes(y=GDP$GDP), method="lm", linetype= "dashed", col="red", se=FALSE) +
  labs(title = "U.S. Gross Domestic Product", subtitle = "GDP Levels from 2005-2013 per
quarter") +
  ylab("GDP (in $, MN)") +
  xlab("Years") +
  theme_fivethirtyeight()

GDPts
#Time series of Public, Private, and Total Construction Spending
###Construction Spending

#construction_data
construction_data <- read.csv("C:/Users/darzo/Downloads/fredgraph.csv")

#get years from 1993 to 2013
construction_data$DATE <- dmy(gsub("/", "-", construction_data$DATE))
construction_data <- filter(construction_data, year(construction_data$DATE) >=1993 &
year(construction_data$DATE) <= 2012)

#create multivariate time series
construction <- ts(data.frame(construction_data$Public.Construction.Spending,
construction_data$Private.Construction.Spending,
construction_data$Total.Construction.Spending), freq=12, start=c(1993,1), names = c("Pubic
Construction Spending", "Private Construction Spending", "Total Construction Spending"))
```

```
#create graph
construction_linegraph <- autoplot(construction, facet=FALSE, ts.size = 1) +
  labs(subtitle = "US Construction Spending from 1993-2013") +
  ylab("Spending (in Millions USD)") +
  xlab("Years") +
  theme_fivethirtyeight()
construction_linegraph

#Bar Chart of Public vs Private Construction Spending - Change of Dist. over time
###Construction Spending

#construction_data
construction_data2 <- read.csv("C:/Users/darzo/Downloads/fredgraph2.csv")
construction_data2$DATE <- dmy(gsub("/", "-", construction_data2$DATE))
construction_data2 <- filter(construction_data2, year(construction_data2$DATE) >=1993 &
year(construction_data2$DATE) <= 2012)
construction_data2 <- filter(construction_data2, day(construction_data2$DATE) == 1)
construction_bargraph <- ggplot(construction_data2, aes(x=DATE, y=Values, fill=Type)) +
  geom_bar(stat="identity") +
  labs(subtitle = "Distribution of Construction Spending from 1993-2013") +
  ylab("Spending (in Millions USD)") +
  xlab("Years") +
  theme_fivethirtyeight()
 construction_bargraph

#Performance of Industry Sectors
###Sector ETFs

#Sector ETF Chart - Performance of US economy
sectoretf <- read.csv("C:/Users/darzo/OneDrive/Documents/SectorETFs.csv")
sector_etf_ts <- ts(data.frame(sectoretf$XLY, sectoretf$XLP, sectoretf$XLE, sectoretf$XLF,
sectoretf$XLB, sectoretf$XLK, sectoretf$IYR, sectoretf$IDU, sectoretf$IYJ, sectoretf$IYH),
freq=12, start=c(2005,1), names = c("Consumer Discretionary", "Consumer Staples", "Energy",
"Financials", "Materials", "Technology", "Real Estate", "Utilities", "Industrials", "Healthcare"))
sector_etf_ts <- autoplot(sector_etf_ts, facet=FALSE, ts.alpha=1, ts.size=1) +
  labs(title = "Industry Sectors", subtitle = "Monthly Performance of US Industry Sectors from
2005-2013") +
  ylab("Performance as %") +
  xlab("Years") +
  theme_fivethirtyeight()
sector_etf_ts

#Unemployment 2006 Age Boxplot
### Unemployment
Unemployment2006byAge <- read.csv("C:/Users/darzo/Downloads/unemployment-06.csv")
```

```
Unemployment2006byAge <- filter(Unemployment2006byAge, Group.Code >= 19 &
Group.Code <= 24)
Unemployment2006byAge$Unemployed.Percent.of.Population <-
as.numeric(as.character(Unemployment2006byAge$Unemployed.Percent.of.Population))


Unemployment2006_Age_Graph <- ggplot(Unemployment2006byAge, aes(x= Group, y =
Unemployed.Percent.of.Population, group = Group, col = Group)) +
  geom_boxplot() +
  labs(title = "Unemployment - Age Group", subtitle = "National Unemployment Rates by Age
Group in 2006") +
  scale_y_continuous(name = "Unemployment Rate",
                     breaks = seq(0,30, 5),
                     limits=c(0, 30)) +
  scale_x_discrete(name = "Age Bracket (excluding 65+)") +
  theme_fivethirtyeight()

Unemployment2006_Age_Graph
#Unemployment 2009 Age Boxplot
### Unemployment
Unemployment2009byAge <- read.csv("C:/Users/darzo/Downloads/unemployment-09.csv")
Unemployment2009byAge <- filter(Unemployment2009byAge, Group.Code >= 19 &
Group.Code <= 24)
Unemployment2009byAge$Unemployed.Percent.of.Population <-
as.numeric(as.character(Unemployment2009byAge$Unemployed.Percent.of.Population))
Unemployment2009_Age_Graph <- ggplot(Unemployment2009byAge, aes(x= Group, y =
Unemployed.Percent.of.Population, group = Group, col = Group)) +
  geom_boxplot() +
  labs(title = "Unemployment - Age Group", subtitle = "National Unemployment Rates by Age
Group in 2009") +
  scale_y_continuous(name = "Unemployment Rate",
                     breaks = seq(0,50, 5),
                     limits=c(0, 50)) +
  scale_x_discrete(name = "Age Bracket (excluding 95+)") +
  theme_fivethirtyeight()
Unemployment2009_Age_Graph

#Unemployment 2006 Ethnicity Boxplot
### Unemployment
Unemployment2006byEthnicity <- read.csv("C:/Users/darzo/Downloads/unemployment-
06.csv")
Unemployment2006byEthnicity <- filter(Unemployment2006byEthnicity, Group.Code == 4 |
Group.Code == 7 | Group.Code == 10 | Group.Code == 13)
Unemployment2006byEthnicity$Unemployed.Percent.of.Population <-
as.numeric(as.character(Unemployment2006byEthnicity$Unemployed.Percent.of.Population))
```

```
Unemployment2006_Ethnicity_Graph <- ggplot(Unemployment2006byEthnicity, aes(x= Group,
y = Unemployed.Percent.of.Population, group = Group, col = Group)) +
  geom_boxplot(na.rm = TRUE) +
  labs(title = "Unemployment - Ethnicity", subtitle = "National Unemployment Rates by
Ethnicity in 2006") +
  scale_y_continuous(name = "Unemployment Rate",
                     breaks = seq(0,25, 5),
                     limits=c(0, 25)) +
  scale_x_discrete(name = "Ethnicity") +
  theme_fivethirtyeight()
Unemployment2006_Ethnicity_Graph

#Unemployment 2009 Ethnicity Boxplot
### Unemployment
Unemployment2009byEthnicity <- read.csv("C:/Users/darzo/Downloads/unemployment-
09.csv")

Unemployment2009byEthnicity <- filter(Unemployment2009byEthnicity, Group.Code == 4 |
Group.Code == 7 | Group.Code == 10 | Group.Code == 13)
Unemployment2009byEthnicity$Unemployed.Percent.of.Population <-
as.numeric(as.character(Unemployment2009byEthnicity$Unemployed.Percent.of.Population))
Unemployment2009_Ethnicity_Graph <- ggplot(Unemployment2009byEthnicity, aes(x= Group,
y = Unemployed.Percent.of.Population, group = Group, col = Group)) +
  geom_boxplot(na.rm = TRUE) +
  labs(title = "Unemployment - Ethnicity", subtitle = "National Unemployment Rates by
Ethnicity in 2009") +
  scale_y_continuous(name = "Unemployment Rate",
                     breaks = seq(0,25, 5),
                     limits=c(0, 25)) +
  scale_x_discrete(name = "Ethnicity") +
  theme_fivethirtyeight()
Unemployment2009_Ethnicity_Graph

#Job Creation Time Series
###Hiring/Layoff Economic Data

#hiring/layoff data
job_creation_data <- read.csv("C:/Users/darzo/Downloads/hiring-and-layoff-economic-
data.csv")

#get years from 2005 to 2013
job_creation_data$DATE <- dmy(gsub("/", "-", job_creation_data$DATE))
job_creation_data <- filter(job_creation_data, year(job_creation_data$DATE) >=2005 &
year(job_creation_data$DATE) <= 2012)

#create multivariate time series
```

```
job_creation <- ts(data.frame(job_creation_data$Layoffs.and.Discharges,
job_creation_data$New.Hires), freq=12, start=c(2005,1), names = c("Layoffs and Discharges",
"New Hires"))

#create graph
job_creation <- autoplot(job_creation, facet=FALSE, ts.size = 1) +
  labs(title = "Job Creation", subtitle = "Monthly Layoffs/New Hires from 2005 to 2013") +
  ylab("Level in Thousands") +
  xlab("Years") +
  theme_fivethirtyeight()
job_creation

#US Interest Rate Policy
###Federal Funds Rate
fed_funds_data <- read.csv("C:/Users/darzo/Downloads/FEDFUNDS.csv")
fed_funds_ts <- ts(fed_funds_data$FEDFUNDS, freq=12, start=c(2005,1))
fed_funds_ts <- autoplot(fed_funds_ts, ts.size = 1) +
  labs(title = "US Interest Rate Policy", subtitle = "Federal Funds Rate from 2005 to 2013") +
  ylab("Rate in %") +
  xlab("Years") +
  theme_fivethirtyeight()
fed_funds_ts

#House Price Index
###All-Transactions House Price Index for the United States
house_price_data <- read.csv("C:/Users/darzo/Downloads/US-total-house-price-index.csv")
house_price_ts <- ts(house_price_data$USSTHPI, freq=4, start=c(1994,1))
house_price_ts <- autoplot(house_price_ts, ts.size = 1) +
  labs(title = "US House Price Index", subtitle = "All-Transactions House Price Index from 1994
to 2014") +
  geom_smooth(aes(y=house_price_data$USSTHPI), method = "lm", linetype = 'dashed', col =
"red", se=FALSE) +
  ylab("Index Level") +
  xlab("Years") +
  theme_fivethirtyeight()
house_price_ts

#Linear Model for Relationship between D and F
#Delinquencies and Foreclosures
Foreclosure <- read.csv("C:/Users/darzo/Downloads/FORLTOTL.csv")
Delinquency <- read.csv("C:/Users/darzo/Downloads/DLQTDLQT.csv")
df <- data.frame(Foreclosure$PX_LAST, Delinquency$PX_LAST)
summary(lm(Delinquency.PX_LAST ~ Foreclosure.PX_LAST, data=df))
```

Maps

==

```{r}
# load data for unemployment rate in 2006 before the crisis
unem_06 <- read.csv("unemployment-06.csv")
unem_06 <- unem_06 %>% select(state, unemployment_rate)

# group the table by state
unem_06 <- unem_06 %>%
  group_by(state) %>%
  summarise(mean_06 = mean(unemployment_rate, na.rm=TRUE))
summarise(unem_06)

# merge by states
region =
c("alaska","alabama","arkansas","arizona","california","colorado","connecticut","district of
columbia","delaware","florida","georgia","hawaii","iowa","idaho","illinois","indiana","kansas","
kentucky","louisiana","massachusetts","maryland","maine","michigan","minnesota","missouri","
mississippi","montana","north carolina","north dakota", "nebraska","new hampshire", "new
jersey","new mexico", "nevada", "new york",
"ohio","oklahoma","oregon","pennsylvania","rhode island", "south carolina","south
dakota","tennessee","texas","utah","virginia","vermont", "washington", "wisconsin", "west
virginia","wyoming")
unem_06 <- unem_06 %>% mutate(region)
unem_total_06 <- merge(all_states, unem_06, by = "region")

# plot all states with ggplot
unem_map1_06 <- ggplot() + geom_polygon(data = unem_total_06, aes(x = long, y = lat, group
= group, fill = unem_total_06$mean_06), col = "white") + scale_fill_continuous(low =
"thistle2", high = "darkred", guide="colorbar")
unem_map2_06 <- unem_map1_06 + theme_bw() + labs(fill = "Unemployment Rate", title =
"State Average Unemployment Rate in 2006")
unem_map3_06 <- unem_map2_06 + scale_y_continuous(breaks = c()) +
scale_x_continuous(breaks = c()) + theme(panel.border = element_blank()) +
theme_fivethirtyeight()


g1 <- ggplot() + geom_map(data=all_states, map=all_states, aes(x = long, y = lat,
map_id=region, group = group),fill="#ffffff", color="#ffffff", size=0.15)
g2 <- g1 + geom_map(data=unem_total_06, map=all_states, aes(fill=mean_06, map_id=region),
color="#ffffff", size=0.10)
gg <- g2 + scale_fill_continuous(low='thistle2', high='darkred',
                                 guide='colorbar')
gg <- gg + labs(x=NULL, y=NULL)
gg <- gg + coord_map("albers", lat0 = 39, lat1 = 45)
```

```
gg <- gg + theme(panel.border = element_blank())
gg <- gg + theme(panel.background = element_blank())
gg <- gg + theme(axis.ticks = element_blank())
gg <- gg + theme(axis.text = element_blank())
gg <- gg + labs(fill = "Unemployment Rate", title = "Average Unemployment Rate by State in
2006")

unem_map3_06 <- gg + theme_fivethirtyeight()
unem_map3_06

# load data for unemployment rate in 2009 after the crisis
unem_09 <- read.csv("unemployment-09.csv")
unem_09 <- unem_09 %>% select(state, unemployment_rate)

# group the table by state
unem_09 <- unem_09 %>%
  group_by(state) %>%
  summarise(mean_09 = mean(unemployment_rate, na.rm=TRUE))
summarise(unem_09)


# merge by states
unem_09 <- unem_09 %>% mutate(region)
unem_total_09 <- merge(all_states, unem_09, by = "region")
g1 <- ggplot() + geom_map(data=all_states, map=all_states, aes(x = long, y = lat,
map_id=region, group = group),fill="#ffffff", color="#ffffff", size=0.15)
g2 <- g1 + geom_map(data=unem_total_09, map=all_states, aes(fill=mean_09, map_id=region),
color="#ffffff", size=0.10)
gg <- g2 + scale_fill_continuous(low='thistle2', high='darkred',
                                 guide='colorbar')
gg <- gg + labs(x=NULL, y=NULL)
gg <- gg + coord_map("albers", lat0 = 39, lat1 = 45)
gg <- gg + theme(panel.border = element_blank())
gg <- gg + theme(panel.background = element_blank())
gg <- gg + theme(axis.ticks = element_blank())
gg <- gg + theme(axis.text = element_blank())
gg <- gg + labs(fill = "Unemployment Rate", title = "Average Unemployment Rate by State in
2009")
unem_map3_09 <- gg + theme_fivethirtyeight()
unem_map3_09
```

Tools
==

```
savePlot <- function(myPlot, name) {
```

```
  plot_name = str_c(name, ".pdf")
  pdf(plot_name, width=10, height=5)
  print(myPlot)
  dev.off()

### Configure to Save Plots
savePlot(SPY_plot, "SPY")
# savePlot(FNMA_plot, "FNMA")
# savePlot(FMCC_plot, "FMCC")
# savePlot(DLQT_plot, "DLQT")
# savePlot(FORLTOTL_plot, "FORLTOTL")
# savePlot(ABX_AAA_plot, "ABX_AAA")
# savePlot(ABX_BBB_plot, "ABX_BBB")
savePlot(housing_plot, "housing-stats-compare")
savePlot(mortgage_plot, "GSE-compare")
savePlot(ABX_plot, "ABX-compare")
savePlot(VIX_plot, "VIX")
savePlot(GDP_plot, "GDP")
# savePlot(, "")
savePlot(unem_map3_06, "unemployment-2006")
savePlot(unem_map3_09, "unemployment-2009")
# savePlot(, "")
savePlot(gg1, "us-unemployment-2008")
```