Rock you like a Hurricane : S&P Remix

Team: Little Data

Kim Campbell, Nancy Culley, Ashley Fay, Caroline Miller, Neil Patel

Financial Security in the face of climate change

Contents

- ❏ Questions & Data
- ❏ Data Cleanup & Exploration
- ❏ Data Analysis
- ❏ Discussion
- ❏ Post Mortem

Core Hypothesis

❖ If there is a natural disaster, such as a hurricane, the US equity market will show a greater loss in performance compared to the same time period with no hurricane present.

Questions

- ❖ Do hurricanes have an impact on US equity daily returns?
- ❖ How does the stock market typically perform ahead of a hurricane?
- ❖ How does the stock market typically perform following a hurricane?
- ❖ Is there a trend between rising/lowering returns and the scale of the hurricane?
- ❖ Is there a different reaction between sectors based on the hit of the hurricane?

Data Needed

Equity Index Returns

The S&P Composite 1500® - Combines three leading indices, the S&P 500®, the S&P MidCap 400®, and the S&P SmallCap 600® to cover approximately 90% of the U.S. market capitalization.

Source: S&P Dow Jones Indices – Data pulled from source through FactSet

Hurricane Data – 2014 through 2018

National Hurricane Center

https://www.nhc.noaa.gov/

Wikipedia

https://en.wikipedia.org/wiki/2018_Atlantic_hurricane_season
https://en.wikipedia.org/wiki/2017_Atlantic_hurricane_season
https://en.wikipedia.org/wiki/2016_Atlantic_hurricane_season
https://en.wikipedia.org/wiki/2014_Atlantic_hurricane_season

Data Sources

- S&P1500_Sector_All
- Energy
- Info_Technology
- Real_Estate
- Utilities
- Consumer_Staples
- Health_Care
- Consumer_Discretionary
- Financials
- Materials
- Industrials
- Communication_Services
- Hurricane_data_all

WIKIPEDIA
The Free Encyclopedia

S&P DOW JONES INDICES

FACTSET

NATIONAL WEATHER SERVICE

Stock Market Returns by Sector

	A	B	C	D	E	F	G	H	I	J	K	L
1	Date	(INDEX) S&	(INDEX) S&	(INDEX) S&	(INDEX) S&	(INDEX) S&	(INDEX) S&	(INDEX) S&	(INDEX) S&	(INDEX) S&	(INDEX) S&	(INDEX) S&P Composite 1500 / Utilities - SEC - Total Return
2	12/30/2013	0	0	0	0	0	0	0	0	0	0	0
3	12/31/2013	0.37	-0.08	-0.03	0.92	0.34	0	0.43	0.71	0.26	-0.49	0.23
4	1/1/2014	0.37	-0.08	-0.03	0.92	0.34	0	0.43	0.71	0.26	-0.49	0.23
5	1/2/2014	-0.16	-0.68	-1.12	-0.44	-0.23	-0.58	-0.87	-0.36	-0.48	-0.66	-1.34
6	1/3/2014	-0.35	-1.31	-1.27	-0.69	0.39	-0.35	-0.55	-0.71	-0.58	0.02	-1.57
7	1/6/2014	-0.98	-0.86	-1.68	-0.67	0.46	-0.75	-1.22	-1	-1.22	0.41	-1.55
8	1/7/2014	-0.45	-0.3	-1.12	0.16	0.61	0.32	-0.68	-0.05	-1.25	0.79	-0.61
9	1/8/2014	-0.6	-0.92	-1.81	-0.53	0.86	1.17	-0.74	0.06	-0.71	0.55	-1.13
10	1/9/2014	-0.77	-2.75	-1.41	-0.9	1.17	1.9	-0.33	-0.47	-1.03	0.58	-0.58
11	1/10/2014	-0.39	-2.39	-1.24	-0.71	1.11	2.34	-0.02	-0.26	-0.59	1.8	0.81
12	1/13/2014	-2.36	-3.45	-1.89	-2.63	-0.31	1.55	-1.16	-1.33	-1.89	1.06	-0.16
13	1/14/2014	-1.59	-3.04	-1.31	-1.5	0.48	2.86	-0.1	0.52	-0.56	1.82	-0.03
14	1/15/2014	-1.37	-1.62	-1.27	-1.78	1.58	2.8	0.65	1.7	0.05	2.32	-0.21
15	1/16/2014	-1.84	-1.23	-1.25	-1.7	1.03	3.05	0.37	1.63	0.15	2.55	0.39
16	1/17/2014	-2.34	-1.75	-2.04	-1.76	0.71	2.97	-0.14	1.06	-0.15	2.12	0.4
17	1/20/2014	-2.34	-1.75	-2.04	-1.76	0.71	2.97	-0.14	1.06	-0.15	2.12	0.4
18	1/21/2014	-2.55	-2.44	-1.68	-1.21	0.92	3.48	-0.06	1.55	0.36	2.9	1.55
19	1/22/2014	-2.23	-3.13	-1.67	-0.85	1.11	3.54	0.22	1.61	-0.36	3.22	1.7
20	1/23/2014	-2.93	-2.17	-2.56	-2	-0.45	2.81	-0.84	1.17	-1.64	2.72	1.41
21	1/24/2014	-4.83	-3.15	-3.5	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
22	1/27/2014	-5.23	-3.05	-3.79								
23	1/28/2014	-4.52	-3.04	-3.12								
24	1/29/2014	-6.09	-3.51	-4.89								

S&P1500_Sector_All

Hurricanes

	Name	Year	Active_date	End_date	Days	Main_state	Highest_Category	Central_pressure	Max_wind	Damage_million	Direct_deaths
2	Arthur	2014	7/1/2014	7/5/2014	5	NC	2	973	85	17	1
3	Hermine	2016	8/28/2016	9/3/2016	7	FL	1	981	80	550	4
4	Matthew	2016	9/28/2016	10/9/2016	12	FL	5	934	165	15,090	586
5	Harvey	2017	8/17/2017	9/1/2017	16	TX	4	937	130	125,000	68
6	Irma	2017	8/30/2017	9/12/2017	14	FL	5	914	180	64,760	52
7	Nate	2017	10/4/2017	10/9/2017	6	LA	1	981	90	787	46
8	Chris	2018	7/6/2018	7/12/2018	7	NC	2	969	105	0	1
9	Florence	2018	8/31/2018	9/17/2018	18	SC	4	939	140	17,900	30
10	Michael	2018	10/7/2018	10/12/2018	6	FL	4	919	155	14,600	60

Hurricane_data_all

Contents

- ❏ Questions & Data
- ❏ Data Cleanup & Exploration
- ❏ Data Analysis
- ❏ Discussion
- ❏ Post Mortem

Bringing S&P1500 csv into jupyter notebook

After importing the S&P1500 csv file, the intent was to parse the current headers into a more user friendly naming convention. Using .head() to view the headers and top 5 rows, we noticed two columns were formatted slightly different from the rest.

```
# Load S&P file
sp1500 = "resources/S&P1500_Sector_All.csv"
sp1500 = pd.read_csv(sp1500)

# Display the data table for preview
sp1500.head()
```

	Date	(INDEX) S&P Composite 1500 / Consumer Discretionary - SEC - Total Return	(INDEX) S&P Composite 1500 / Communication Services -SEC - Total Return	(INDEX) S&P Composite 1500 / Consumer Staples - SEC - Total Return	(INDEX) S&P Composite 1500 / Energy - SEC - Total Return	(INDEX) S&P Composite 1500 / Financials - SEC - Total Return	(INDEX) S&P Composite 1500 / Health Care - SEC - Total Return	(INDEX) S&P Composite 1500 / Industrials - SEC - Total Return	(INDEX) S&P Composite 1500 / Information Technology - SEC - Total Return	(INDEX) S&P Composite 1500 / Materials - SEC - Total Return	(INDEX) S&P Composite 1500 Real Estate - SEC - Total Return	(INDEX) S&P Composite 1500 / Utilities - SEC - Total Return
0	12/30/2013	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
1	12/31/2013	0.37	-0.08	-0.03	0.92	0.34	0.00	0.43	0.71	0.26		
2	1/1/2014	0.37	-0.08	-0.03	0.92	0.34	0.00	0.43	0.71	0.26		
3	1/2/2014	-0.16	-0.68	-1.12	-0.44	-0.23	-0.58	-0.87	-0.36	-0.48		

The Communications services column was missing a space between the dash and “SEC” and the Real Estate column was missing a slash after 1500.

Bringing S&P1500 csv into jupyter notebook

For the missing space, we replaced all instances of “-SEC” with “- SEC” (only 1), using the replace command. For the missing “/”, we used the same replace command, but instead of adding a “/” to the Real Estate column, we deleted it from all other columns so we wouldn’t have to deal with it later (first line of code below).

```
# Fix columns with slightly different formatting, to make all uniform (i.e., Communication Services and Real Estate cols)
sp1500.columns = sp1500.columns.str.replace('-SEC','- SEC').str.replace('1500 /','1500')

# Delete all special characters by replacing with empty string
sp1500.columns = sp1500.columns.str.replace(r'(','').str.replace(r')','')
```

After some research, we discovered special characters can be tricky when trying to parse a field, so we added the second line of code above which replaced the open and close parentheses with blanks.

Bringing S&P1500 csv into jupyter notebook

To remove all remaining unwanted text from the column headers (keeping sector only), we again used the replace command to delete the extra words and spaces (first two lines of code below). We also decided that it was easier to work with column headers that don't contain any spaces, so in the third line of code below we inserted an underscore to replace any spaces.

```
# Remove remaining text we don't want
sp1500.columns = sp1500.columns.str.replace('INDEX S&P Composite 1500 ','').str.replace(' - SEC - Total Return ','')
sp1500.columns = sp1500.columns.str.replace('   ','') # replaces 3 back to back spaces left over after removing other text

# Replace space with underscore for ease of referencing column later on
sp1500.columns = sp1500.columns.str.replace(' ','_')

# Reformat date column, so we can filter on it later
sp1500['Date'] = pd.to_datetime(sp1500['Date'])

# Display the data table to see new column headers
sp1500.head()
```

Date	Consumer_Discretionary	Communication_Services	Consumer_Staples	Energy	Financials	Health_Care	Industrials	Information_Technology	Materials

Finally, we reformatted the date column so that we could use it later on for charting and it would be recognized as a date in formulas, by using the pandas library command to_datetime (4th line of code above)

Bringing hurricane csv into jupyter notebook

We created a CSV file from scratch, so we formatted the columns before bringing it in. Once we imported the data, we again formatted the two date columns using pandas command to_datetime, so we could use them later on in our analysis

```
# Load Hurricane data
hurricanes = "resources/Hurricane_data_all.csv"
hurricanes = pd.read_csv(hurricanes)

# Reformat date columns, so we can use them later as part of formula
hurricanes['Active_date'] = pd.to_datetime(hurricanes['Active_date'])
hurricanes['End_date'] = pd.to_datetime(hurricanes['End_date'])

# Display the data table for preview
hurricanes.head(20)
```

	Name	Year	Active_date	End_date	Days	Main_state	Highest_Category	Central_pressure	Max_wind	Damage_million	Direct_deaths
0	Arthur	2014	2014-07-01	2014-07-05	5	NC	2	973	85	17	1
1	Hermine	2016	2016-08-28	2016-09-03	7	FL	1	981	80	550	4

Using timedelta and strftime to manipulate dates

```
# Hermine
#calculate 21 days before Active date and 21 days after End date to set as x-axis range
hurricanes_her= hurricanes.set_index('Name')
start_date_hermine = hurricanes_her.loc["Hermine", "Active_date"]
end_date_hermine = hurricanes_her.loc["Hermine", "End_date"]

#set 3 week delta for x min and max
x_min_herm= start_date_hermine - timedelta(days=21)
x_min_herm = x_min_herm.strftime('%m/%d/%Y')

x_max_herm= end_date_hermine + timedelta(days=21)
x_max_herm = x_max_herm.strftime('%m/%d/%Y')
```

Timedelta automatically found the date 21 days before and after the date the hurricane hit.

Strfttime formatted the date in a more readable format for the graphs

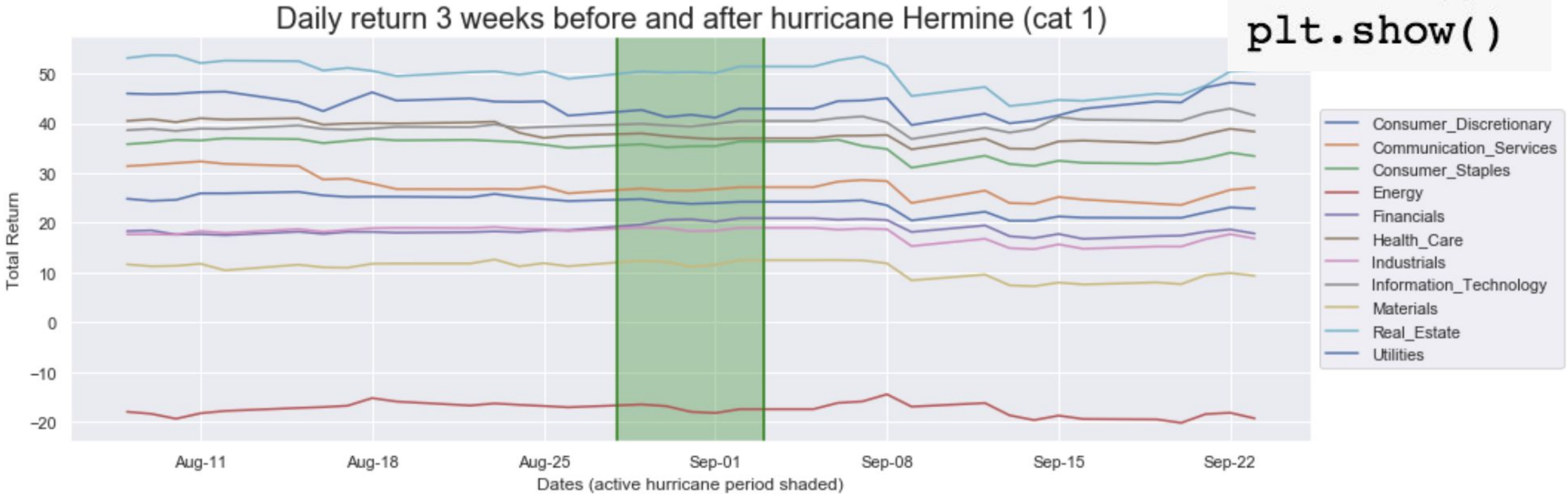
Shading the hurricane time period

```
# Plot 11 sectors
fig = plt.figure()
ax = plt.axes()
for x in range(1, 12):
    plt.plot(hurricane_hermine['Date'], hurricane_hermine.iloc[:,x])

# Shade area representing Hurricane Hermine
x_Fmt = DateFormatter("%b-%d")
plt.axvline(start_date_hermine, color='green')
plt.axvline(end_date_hermine, color='green')

ax.xaxis.set_major_formatter(x_Fmt)
ax.axvspan(start_date_hermine, end_date_hermine, alpha=0.3, color='green')

plt.ylabel('Total Return')
plt.xlabel('Dates (active hurricane period shaded)')
plt.title('Daily return 3 weeks before and after hurricane Hermine (cat 1)', fontsize = 20)
plt.legend(loc='center left', bbox_to_anchor=(1.0, 0.5))
plt.rcParams["figure.figsize"] = (15,5)
```

Setting plt.axvline for each hurricane's start and end date visualized when the hurricane occured within the 21 day period

Using Seaborne to style the graph

```
#Finesse
sns.set()
plt.show()
```



Contents

- ❑ Questions & Data
- ❑ Data Cleanup & Exploration
- ❑ Data Analysis
- ❑ Discussion
- ❑ Post Mortem

Data Analysis - T-testing jupyter notebook sample code

To analyze if there was an effect on the stock market when a hurricane occurs, we ran a t-test for each hurricane comparing the means of the same timeframe (one from the actual hurricane, and the other mean from the same period in 2015, as a control since no 2015 hurricanes). Then a 2 independent sample t-test was conducted to determine if the two means were statistically significant from each other.

Hurricane Arthur, category 2

```
# Arthur ttest data
hurricane_arthur_ttest = sp1500[(sp1500['Date']>= x_min_art)& (sp1500['Date'] <= x_max_art)]

# Add avg all sectors to 2015 control for arthur
sp1500_arthur['Avg All Sectors'] = sp1500_arthur.mean(axis=1)

# run t-test
(ttest, p) = scipy.stats.ttest_ind(hurricane_arthur_ttest['Avg All Sectors'], sp1500_arthur['Avg A
ll Sectors'])

if p < 0.05:
    print("The differences are significant.", p)
else:
    print("The differences between are due to chance.", p)
```

```
The differences are significant. 1.6636427012809893e-37
```

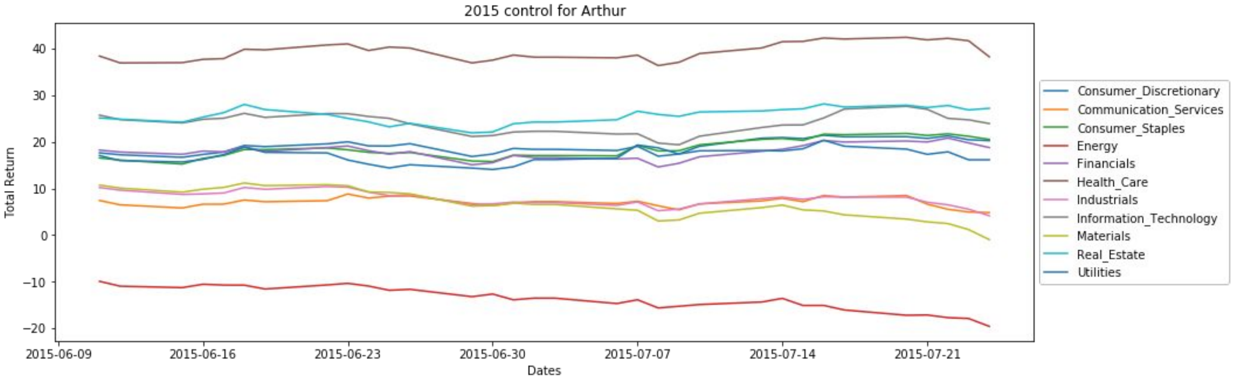
T-test results

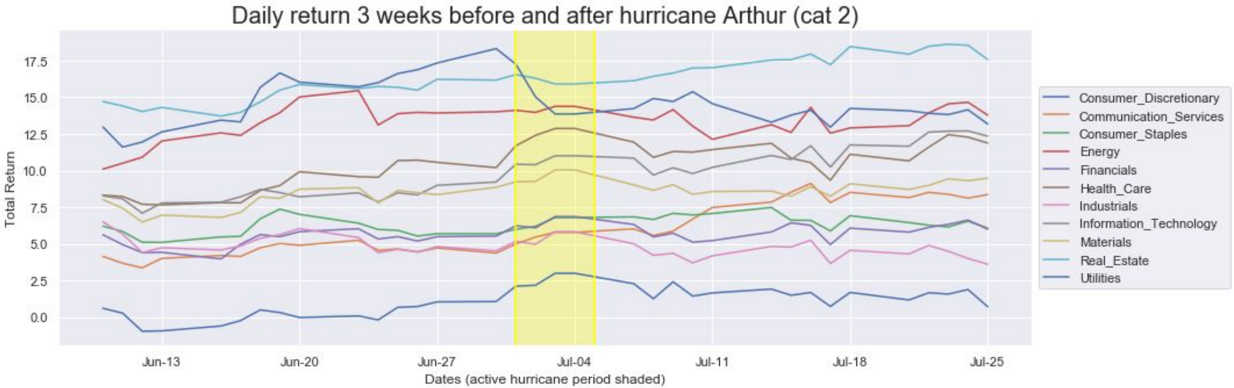
Measuring hurricanes against the same time period in 2015, our t-test results found all were significant!

The following slides help visualize the data used for each t-test. Shade areas represent the actual hurricane (green - cat 1, yellow cat 2, orange cat 4, red cat 5)

Hurricane	Category	P value
Matthew (2016)	5	5.74E-39
Irma (2017)	5	2.79E-62
Harvey (2017)	4	2.02E-61
Florence (2018)	4	2.09E-86
Michael (2018)	4	9.65E-51
Arthur (2014)	2	1.66E-37
Chris (2018)	2	1.09E-76
Hermine (2016)	1	1.86E-24
Nate (2017)	1	3.99E-59

2014 Hurricane Arthur



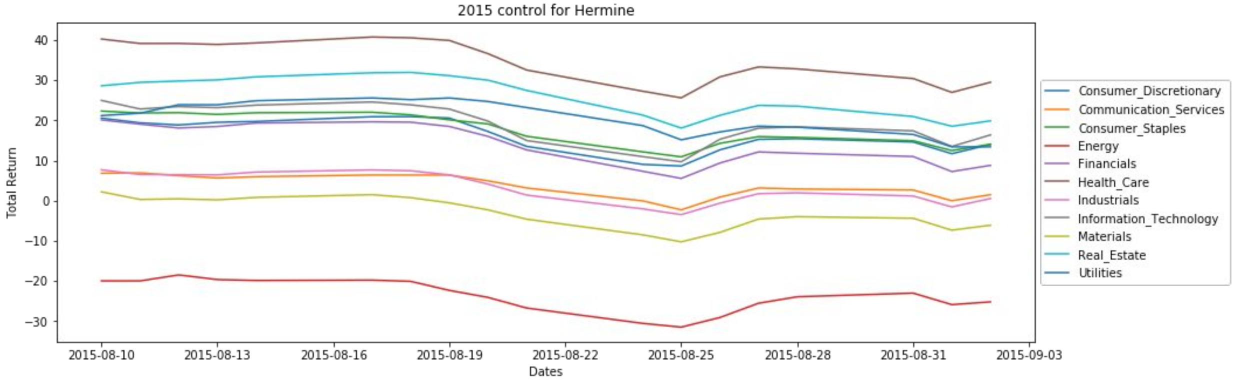


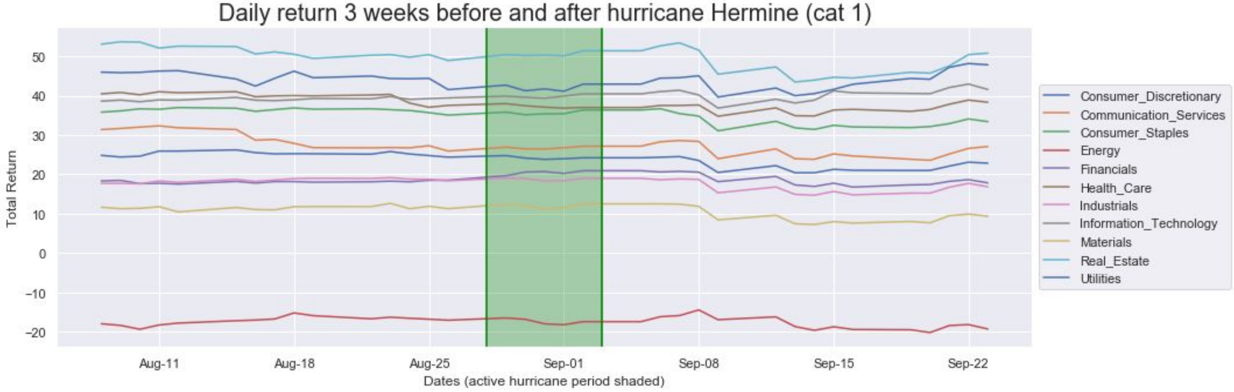
RESULTS

T-Test p-value= 1.6636427012809893e-37

The differences are significant

2016 Hurricane Hermine



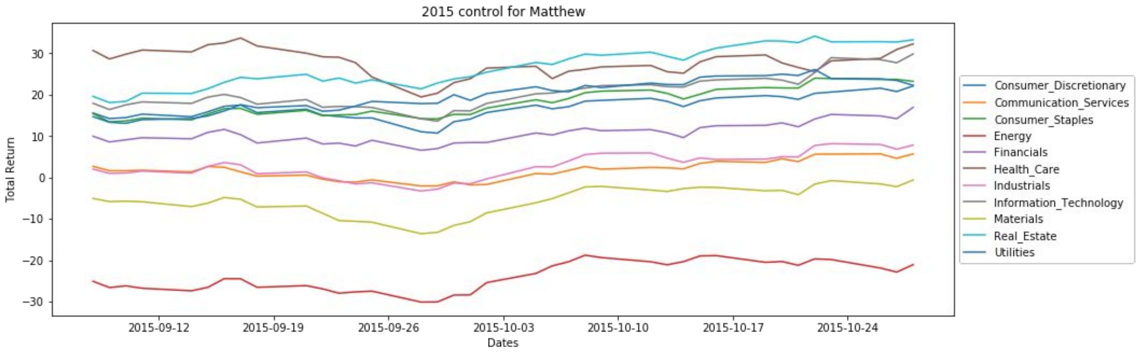


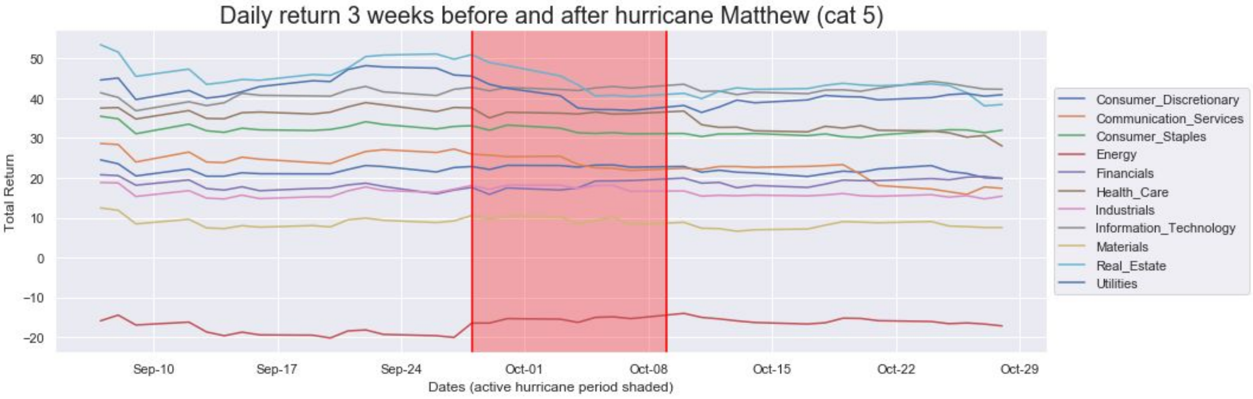
RESULTS

T-Test p-value= 1.8580317551628005e-24

The differences are significant

2016 Hurricane Matthew



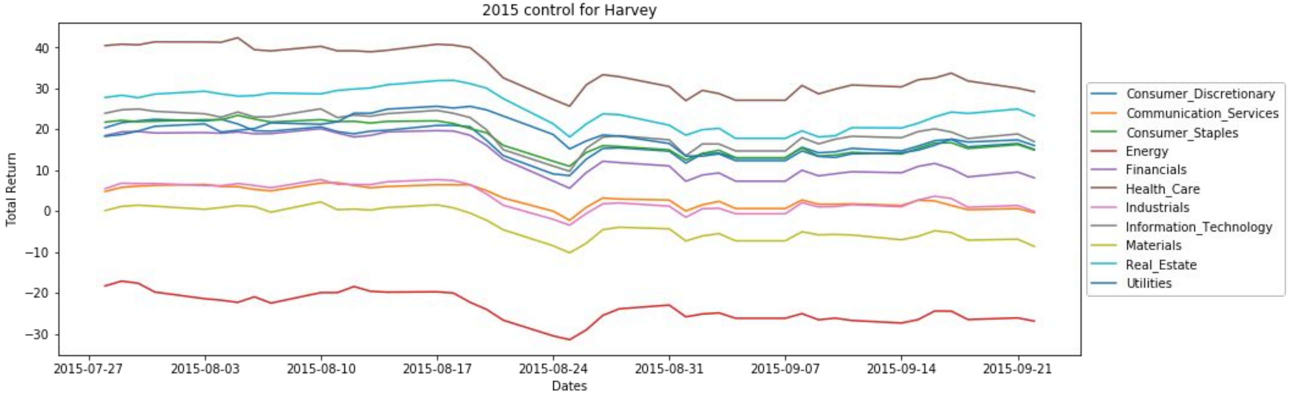


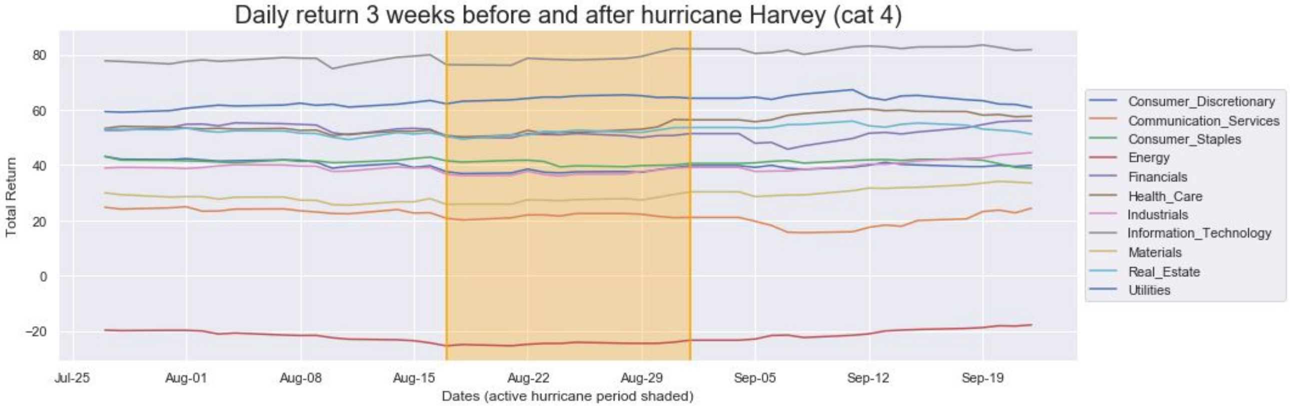
RESULTS

T-Test p-value= 5.73613043667607e-39

The differences are significant

2017 Hurricane Harvey



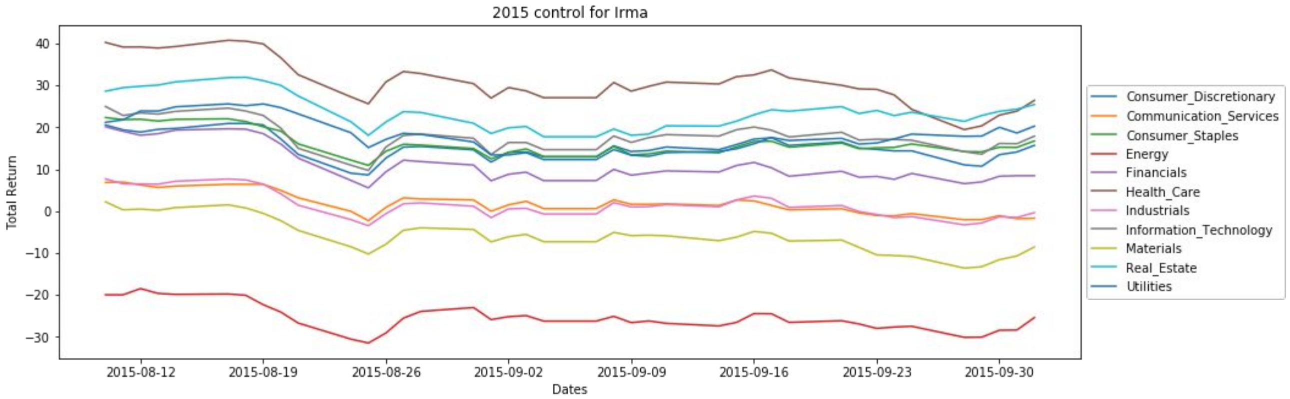


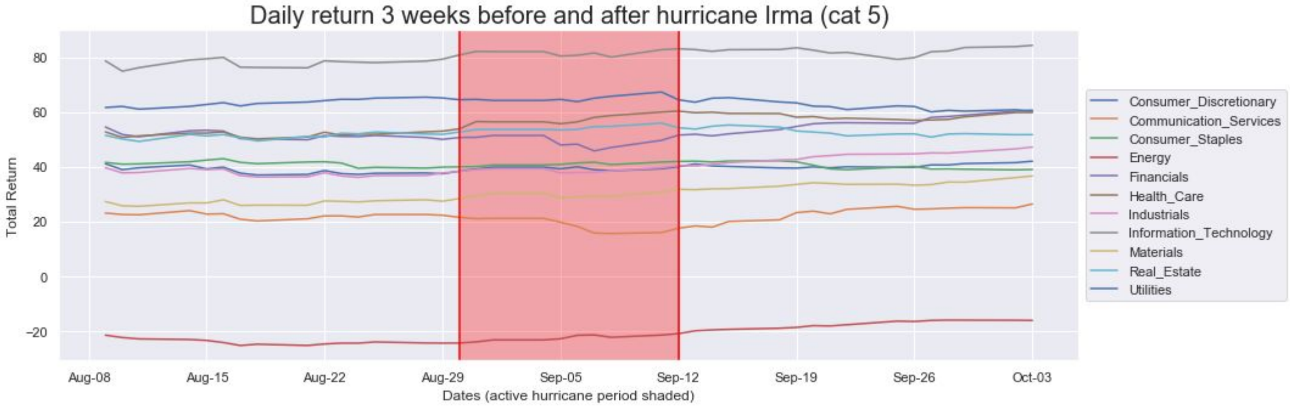
RESULTS

T-Test p-value= 2.0237412514233564e-61

The differences are significant

2017 Hurricane Irma



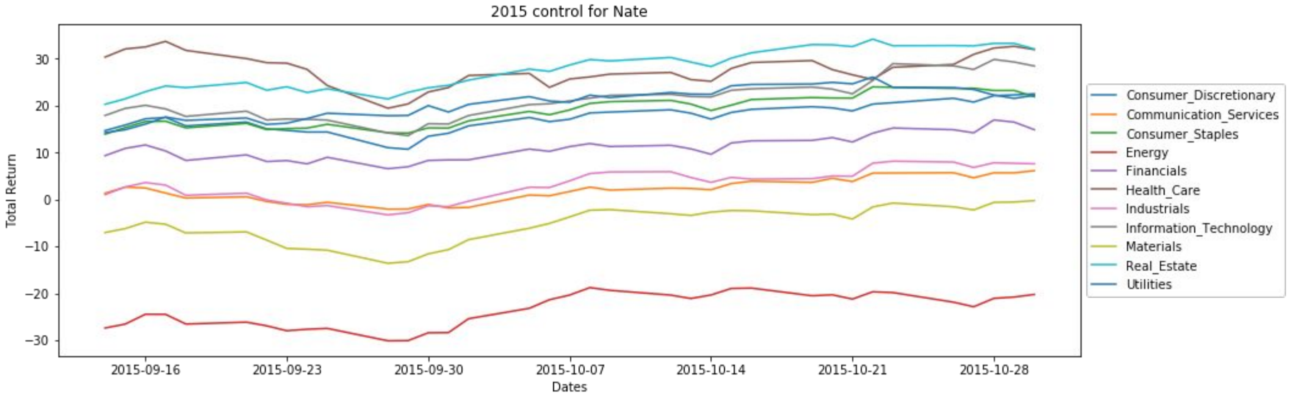


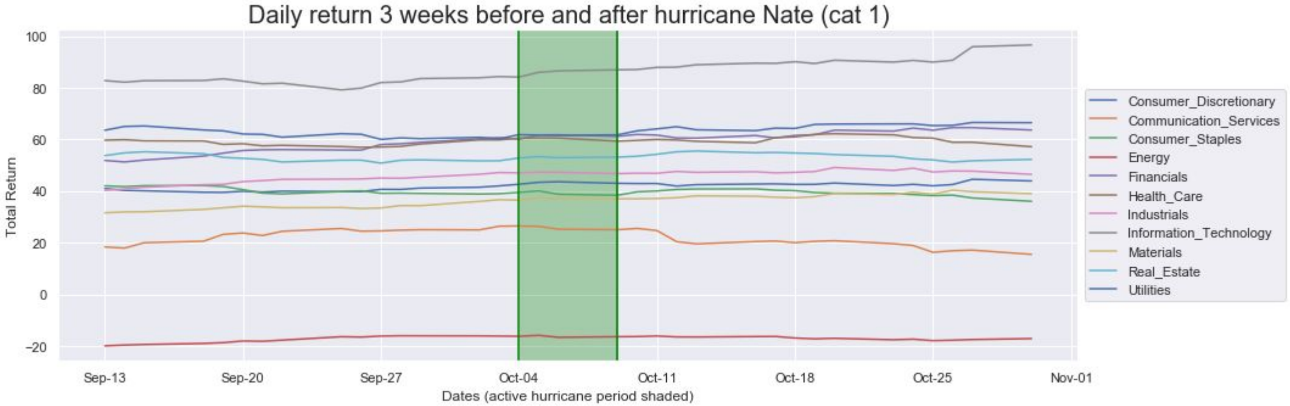
RESULTS

T-Test p-value= 2.7867516695780473e-62

The differences are significant

2017 Hurricane Nate



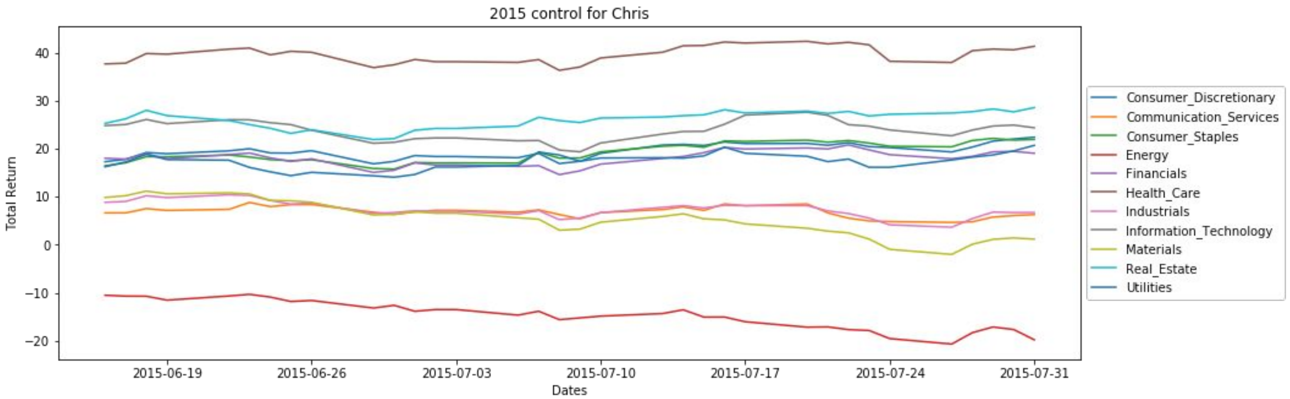


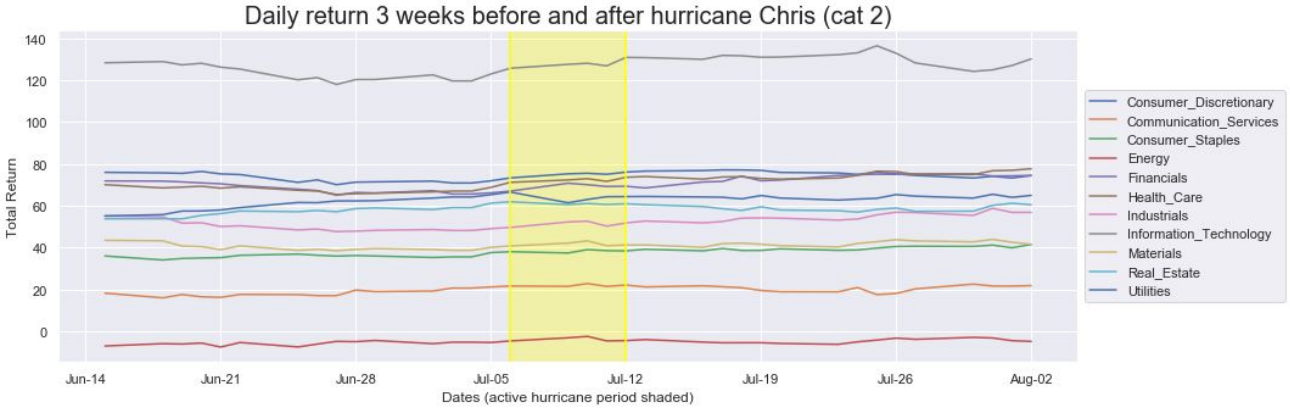
RESULTS

T-Test p-value= 3.987174173771928e-59

The differences are significant

2018 Hurricane Chris



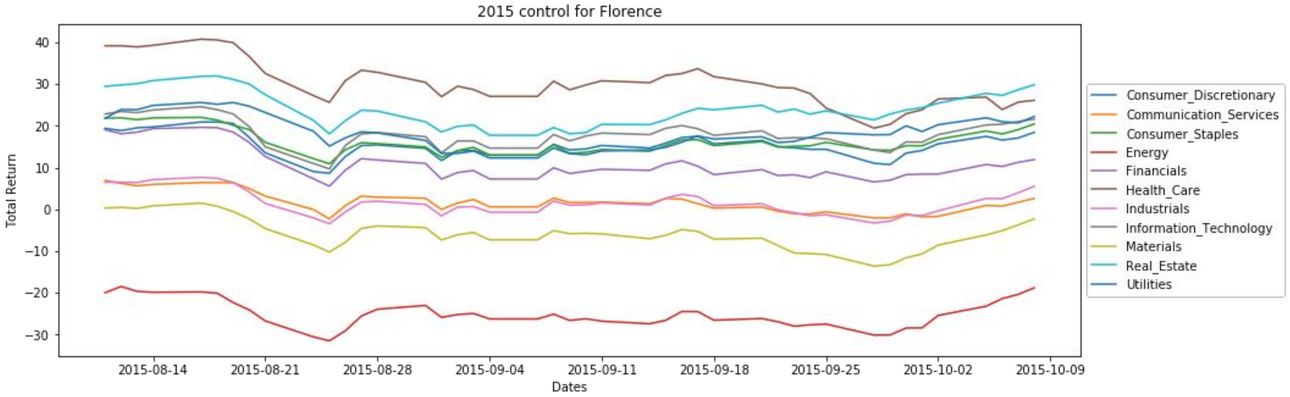


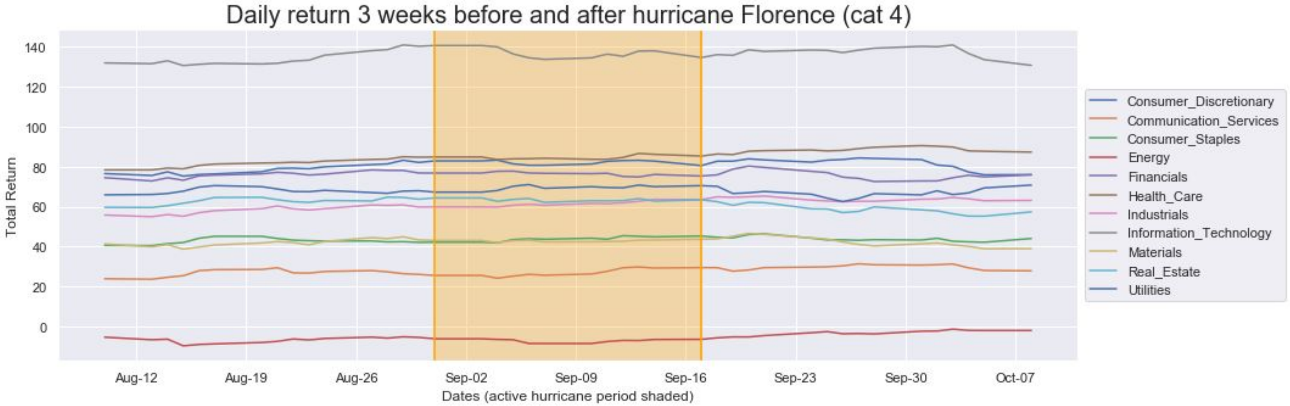
RESULTS

T-Test p-value= 1.0945510935425884e-76

The differences are significant

2018 Hurricane Florence



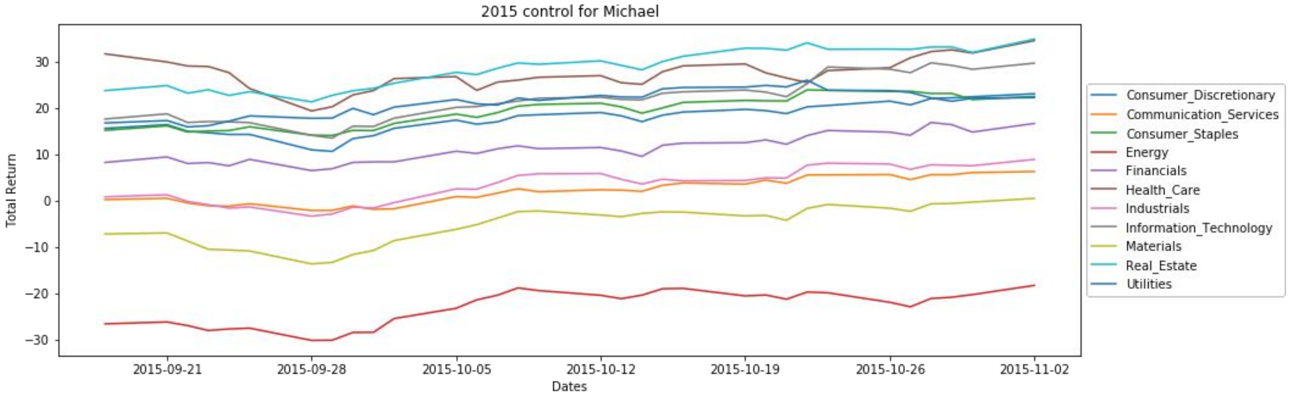


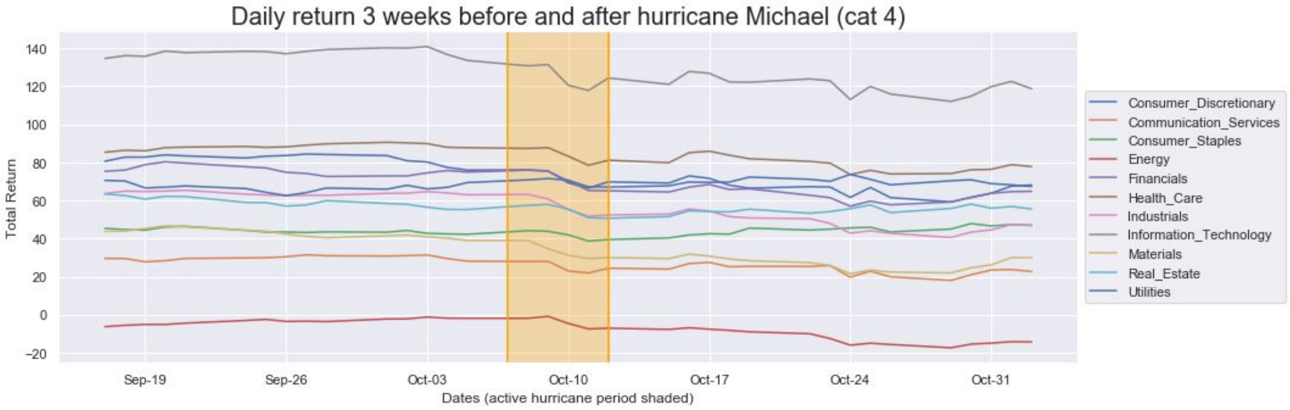
RESULTS

TT-Test p-value= 2.09232889660908e-86

The differences are significant

2018 Hurricane Michael





RESULTS

T-Test p-value= 9.653905778241775e-51

The differences are significant

Contents

- ❏ Questions & Data
- ❏ Data Cleanup & Exploration
- ❏ Data Analysis
- ❏ Discussion
- ❏ Post Mortem

Discussion

- We expected to find that there would be a statistically significant difference between the control year and the hurricane year.

- We were predicting that the more extreme hurricanes would be more likely to have a statistically significant difference however just the fear of an impending hurricane has an affect on the market.

- The changes in the daily returns for duration of all 9 hurricanes we analyzed were found to be statistically significant.

- The S&P 1500 had a statistically significant dip in its daily returns for the specified duration when the hurricanes were active, therefore we fail to reject the null hypothesis.

Contents

- ❏ Questions & Data
- ❏ Data Cleanup & Exploration
- ❏ Data Analysis
- ❏ Discussion
- ❏ Post Mortem

Post Mortem

Difficulties encountered:

1. with the government shut down weather data had to be sourced from private resources.
2. determining which, large/mid/small cap, index would be appropriate to focus.
3. limited number of hurricanes every year.
4. The multi variability of factors affecting the markets makes it difficult to pin point hurricanes as being the sole cause for the drop

If we had two more weeks:

1. we would evaluate additional avenues such as catastrophe bonds and drill down on sectors such as insurance to determine calculate the effect hurricanes have on them.
2. we could generate a model which would take into account other markets in order to better predict the market loss based on additional hurricane data and other weather anomalies.