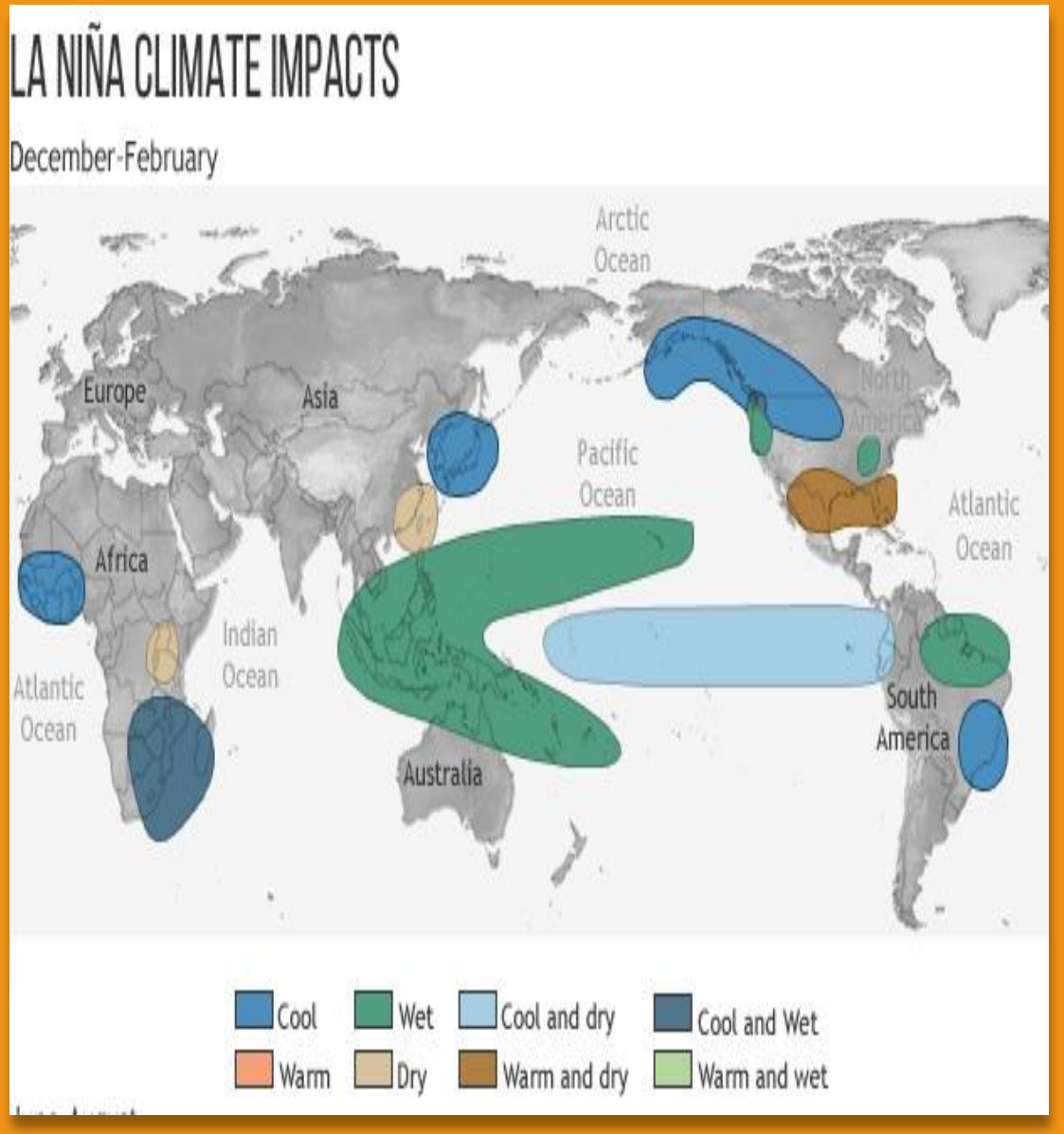
Weather Phenomenon and Fire Outbreaks

DSC 530 Data Exploration and Analysis

Gabriel Valenzuela

Statistical Question / Hypothesis

- The weather phenomenon known as La Nina that occurs in the Pacific Ocean will cause more fire outbreaks in the Amazon due to the wind direction causing a dryer season in the area



Variables within Analysis

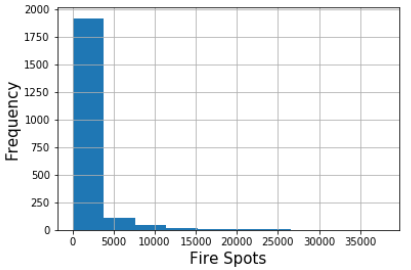
- def_area_2004_2019.csv
  - PA
  - AMZ LEGAL
- el_nino_la_nina_1999_2019.csv
  - Start year
  - End year
  - Severity
- inpe_brazilian_amazon_fires_1999_2019.csv
  - Year
  - Fire spots
  - Latitude
  - Longitude

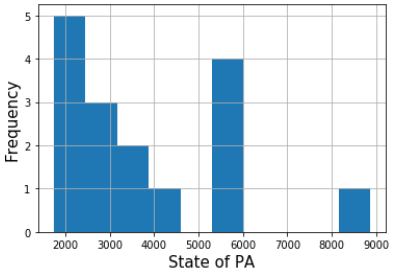
Description of Variables

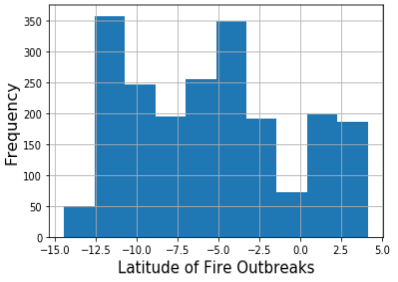
- def_area_2004_2019.csv
  - PA – deforested area in Para
  - AMZ LEGAL – sum of deforested area in Brazil
- el_nino_la_nina_1999_2019.csv
  - Start year – year of the start of the phenomenon
  - End year – year of the end of the phenomenon
  - Severity – weak / moderate / strong / very strong
- inpe_brazilian_amazon_fires_1999_2019.csv
  - Year – year of occurrence
  - Fire spots – number of forest fire outbreaks
  - Latitude – average latitude of all occurrences
  - Longitude – average longitude of all occurrences

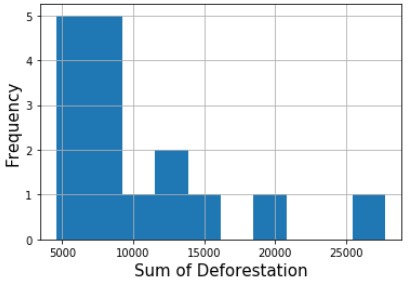
Histogram & Outliers

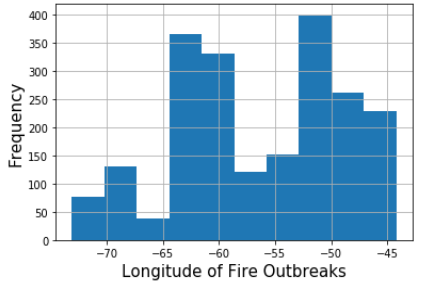
- Outliers
  - Fire Spots in Amazon Fire Outbreak Data Set
    - Values begin to taper off as the number of fire outbreaks increase
    - Removal: It is important to understand the years that had an enormous amounts of fire out breaks but should not be included within the general data since it does not occur often enough
  - Sum of Deforestation and State of PA in Deforestation Data Set
    - Values begin to taper off in the higher numbers for amount of desforestation in the state and overall
    - Removal: Like before, with keeping all the information together it should be exlcuded, but it is important to note that it does occur











Descriptive Characteristics

- Summary of descriptive characteristics including median, mode, and spread

```
Longitude of Fire Outbreaks Descriptive Characteristics
count    2104.000000
mean      -56.510314
std         7.382619
min       -73.085000
25%       -62.376004
50%       -55.876480
75%       -50.396154
max       -44.188000
Name: longitude, dtype: float64

Median: -55.876479764783525
Mode: ModeResult(mode=array([-48.649]), count=array([2]))
Spread: 28.89699999999999
```

```
State of PA Descriptive Characteristics
count      16.000000
mean     3923.625000
std      1945.802487
min      1741.000000
25%      2411.250000
50%      3389.000000
75%      5546.250000
max      8870.000000
Name: PA, dtype: float64

Median: 3389.0
Mode: ModeResult(mode=array([1741], dtype=int64), count=array([1]))
Spread: 7129
```

```
Sum of Deforestation Descriptive Characteristics
count       16.000000
mean     10020.937500
std       6112.467134
min       4571.000000
25%       6365.250000
50%       7500.000000
75%      11966.000000
max      27772.000000
Name: AMZ LEGAL, dtype: float64

Median: 7500.0
Mode: ModeResult(mode=array([4571], dtype=int64), count=array([1]))
Spread: 23201
```

```
Fire Spots Descriptive Characteristics
count     2104.000000
mean      1167.417776
std       2959.558714
min          1.000000
25%         16.000000
50%        121.000000
75%        755.000000
max      37926.000000
Name: firespots, dtype: float64

Median: 121.0
Mode: ModeResult(mode=array([1], dtype=int64), count=array([93]))
Spread: 37925
```

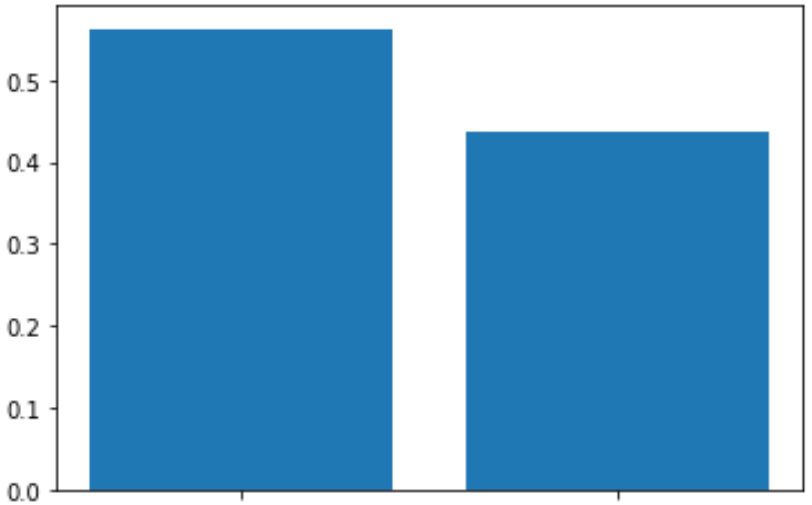
```
Latitude of Fire Outbreaks Descriptive Characteristics
count    2104.000000
mean       -5.439282
std         4.852439
min       -14.431908
25%        -9.946974
50%        -5.808292
75%        -2.595169
max         4.151000
Name: latitude, dtype: float64

Median: -5.808291666666666
Mode: ModeResult(mode=array([-7.632]), count=array([2]))
Spread: 18.582908
```

Probability Mass Function

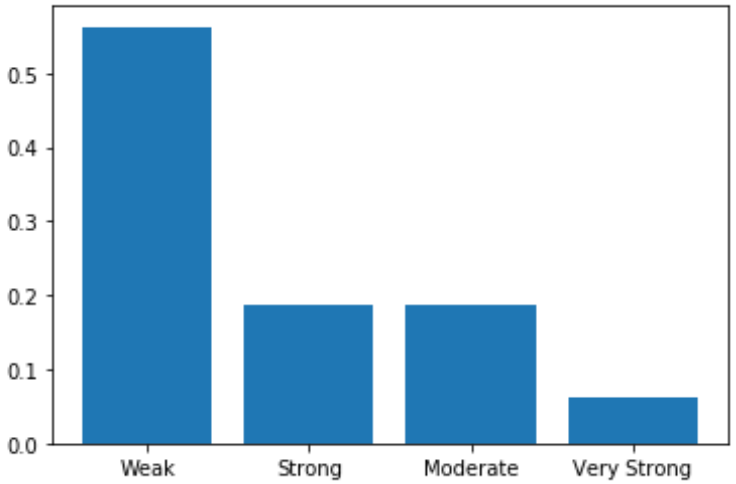
Weather Phenomenon El Nino and La Nina

	item_counts	item	probability
La Nina	9	La Nina	0.5625
El Nino	7	El Nino	0.4375



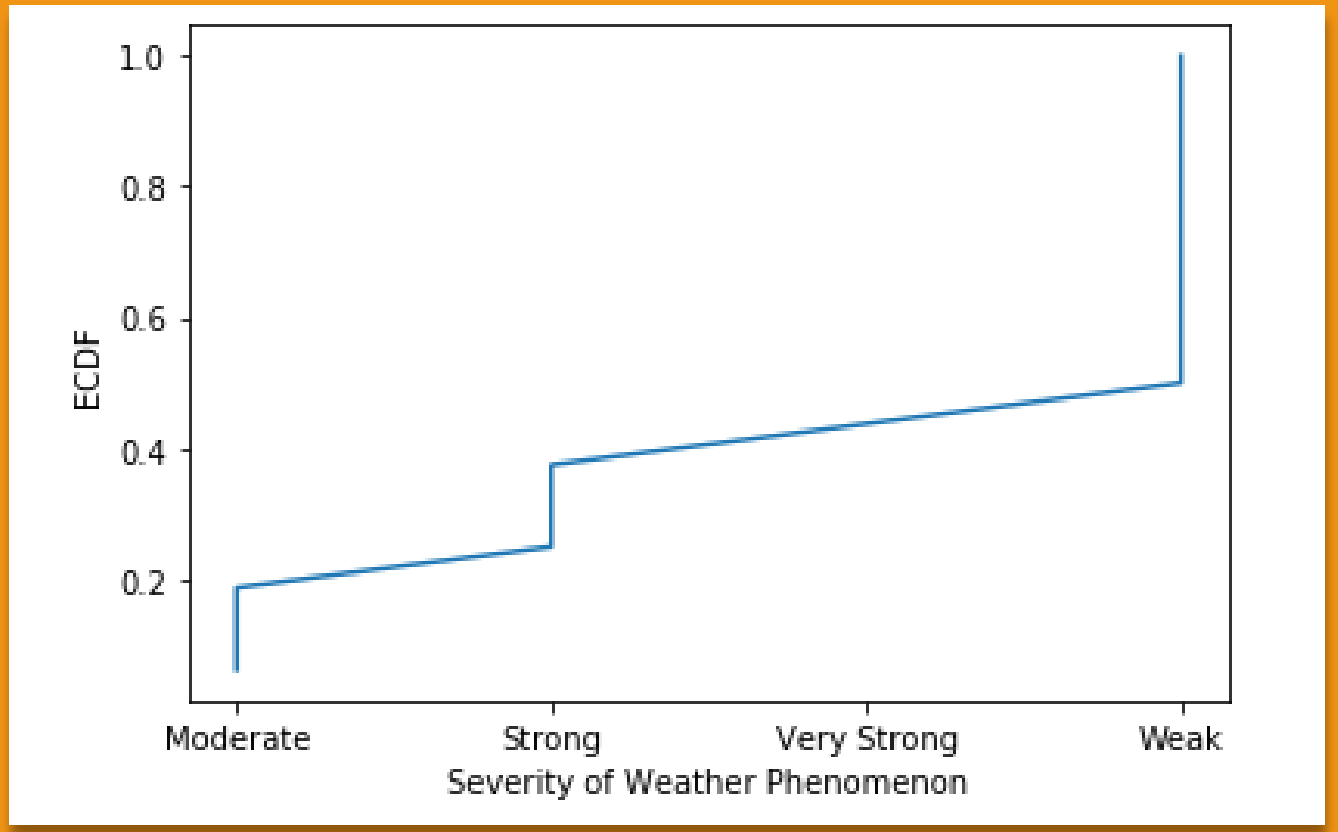
Weather Phenomenon Severity

	item_counts	item	probability
Weak	9	Weak	0.5625
Strong	3	Strong	0.1875
Moderate	3	Moderate	0.1875
Very Strong	1	Very Strong	0.0625



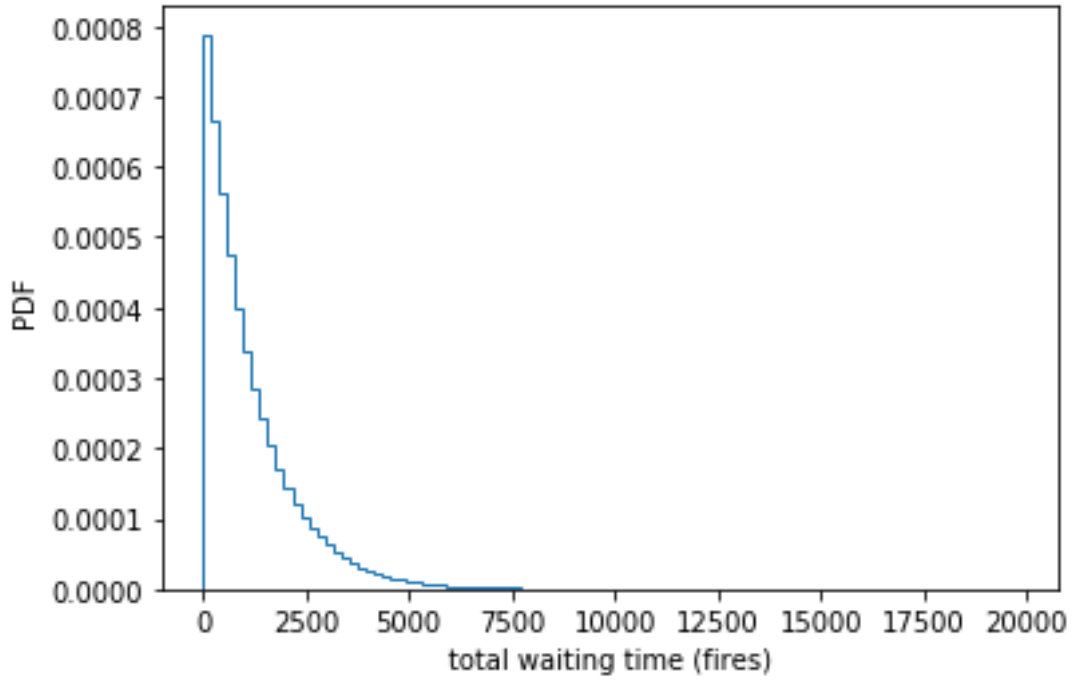
Cumulative Distribution Function

- ECDF of Severity of the Weather Phenomenon Data Set
  - 50% of occurrences are a weak storm
  - 45% of occurrences are a very strong storm
  - 35% of occurrences are a strong storm
  - 20% of occurrences are a moderate storm
- By determining the likelihood of the severity of the storms, I can then look back towards the years of the fire outbreaks and the storms to determine if there were more outbreaks and which phenomenon it was that year



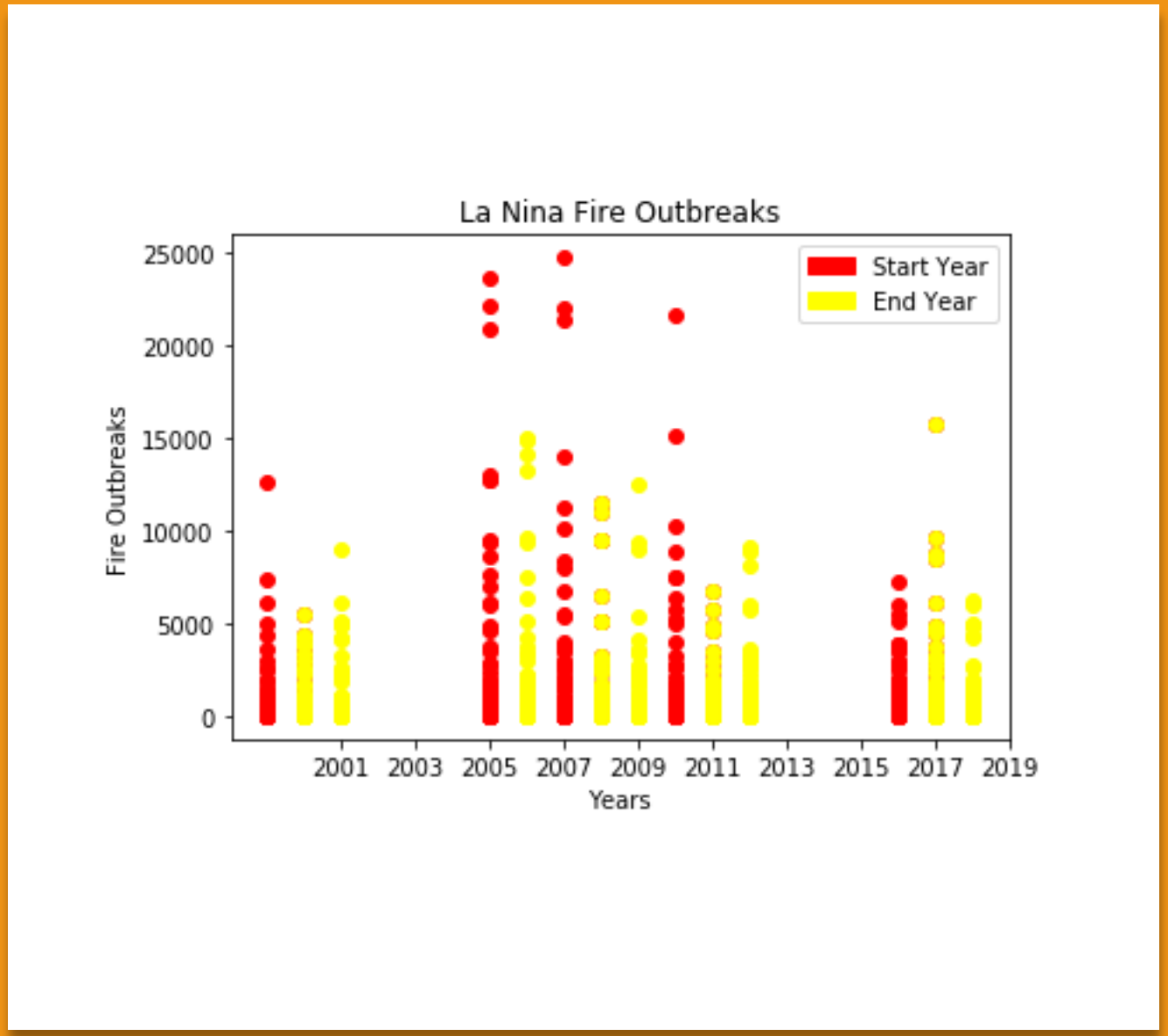
Exponential Distribution

- With fire outbreaks, we can see the time it takes for a fire spot to occur in any of the several states is not long and occur at a rate that is almost continuous



Correlation and Causation

Scatter Plot of fire Spots the start year and the end year



Correlation and Causation

- Correlation and Covariance of fire spots for the start year of La Nina
  - Weak negative relationship

Correlation

	Start Year	Fire Spots
Start Year	1.	-0.00739929
Fire Spots	-0.00739929	1.

Covariance

	Start Year	Fire Spots
Start Year	3.16201352e+01	-3.60253454e+03
Fire Spots	-3.60253454e+03	1.26639303e+07

Correlation and Causation

- Correlation and Covariance of fire spots for the end year of La Nina
  - Weak negative relationship

Correlation

	End Year	Fire Spots
End Year	1.	-0.00383514
Fire Spots	-0.00383514	1.

Covariance

	End Year	Fire Spots
End Year	3.34075319e+01	-4.43724198e+01
Fire Spots	-4.43724198e+01	4.00701066e+06

Hypothesis Test – Difference of Means

- T-Value: 0.333
  - Indicates that the two groups(fire spots in start year & fire spots in end year) are similar
- P-Value: 0.741
  - Indicates that these occurrences happened by chance and could not easily be reproduced

```
t = 0.3333437465572071
p = 0.7407071998450 82
```

Linear Regression Analysis

- Dependent variable: Years that La Nina started
- Explanatory variable: Fire out breaks during the years that La Nina started

- Prediction for years 2020 – 2025 of fire outbreaks based on model:
  - 2020 - 1083
  - 2021 - 1079
  - 2022 - 1075
  - 2023 - 1072
  - 2024 - 1068
  - 2025 - 1064

```
[1083.13151203 1079.47981548 1075.82811893 1072.17642238 1068.52472583
 1064.87302928]
```

Conclusion

- La Nina does have a strong correlation for fire outbreaks in the amazon
  - Even when the severity of the storm was the highest, the fire outbreaks were much on the higher side
- Based on the model:
  - Fire outbreaks from the storms will continue with the possibility of increasing