SETTING THE DATA WORLD ABLAZE

THE STORIES BEHIND WILDFIRES

Grace Cote
Hima Gharat
Joshua Heisler
Salita Santiago
Marcellis Valentin


Edmonton
Calgary
Vancouver
UNITED
STATES
Monterrey
MÉXICO

"Wildfires are **unplanned** fires that burn in **natural areas** like forests, grasslands or prairies. These dangerous fires spread quickly and can devastate not only wildfire and natural areas, but also communities."

—READY.GOV (FEMA)

ARE WILDFIRES INCREASING?

CONTEXTS OF RESEARCH

HISTORICAL

Has there been an increase or decrease in fires or acres over the years?

GEOGRAPHICAL

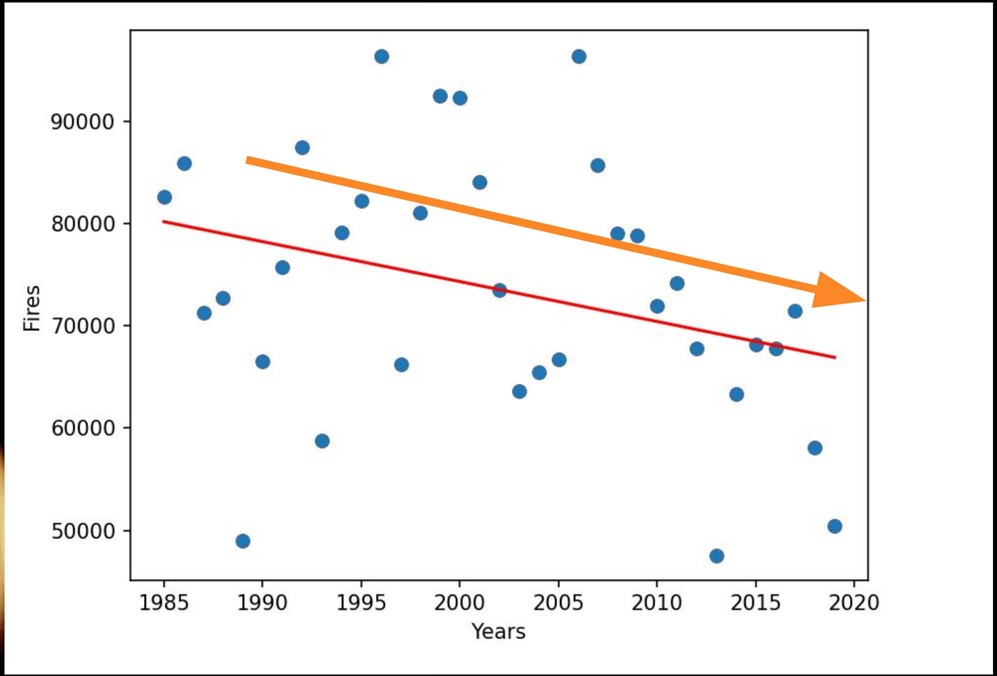
Does region or land mass information affect the number of wildfires or acres?

FINANCIAL

How have the financial damages of wildfires changed over the years?

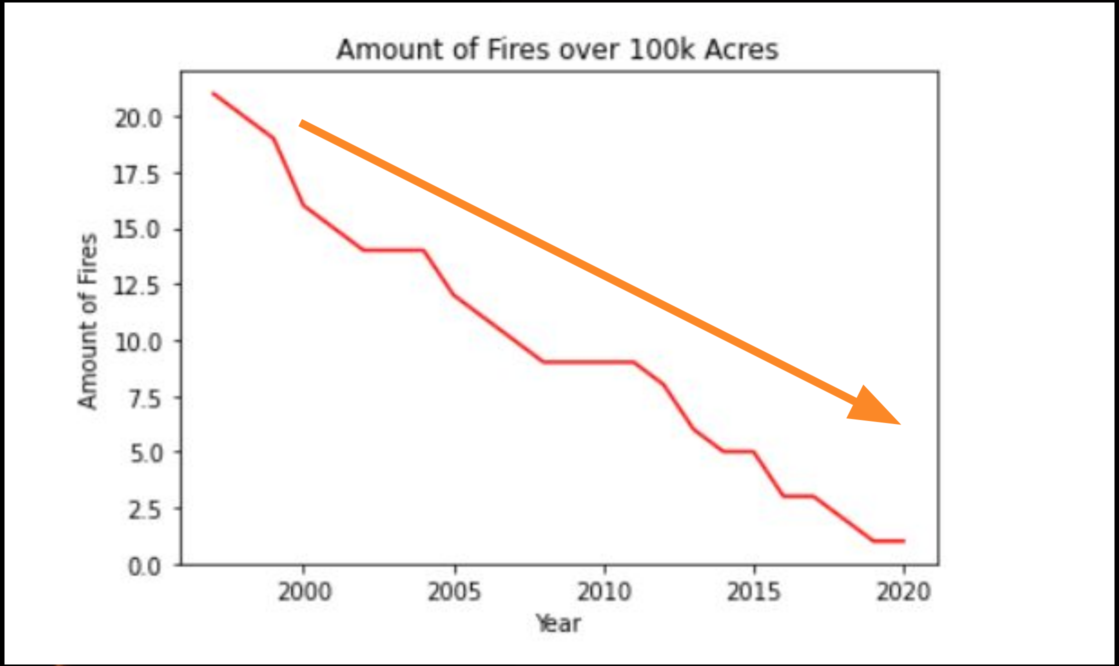
HISTORICAL

of Fires Since 1985



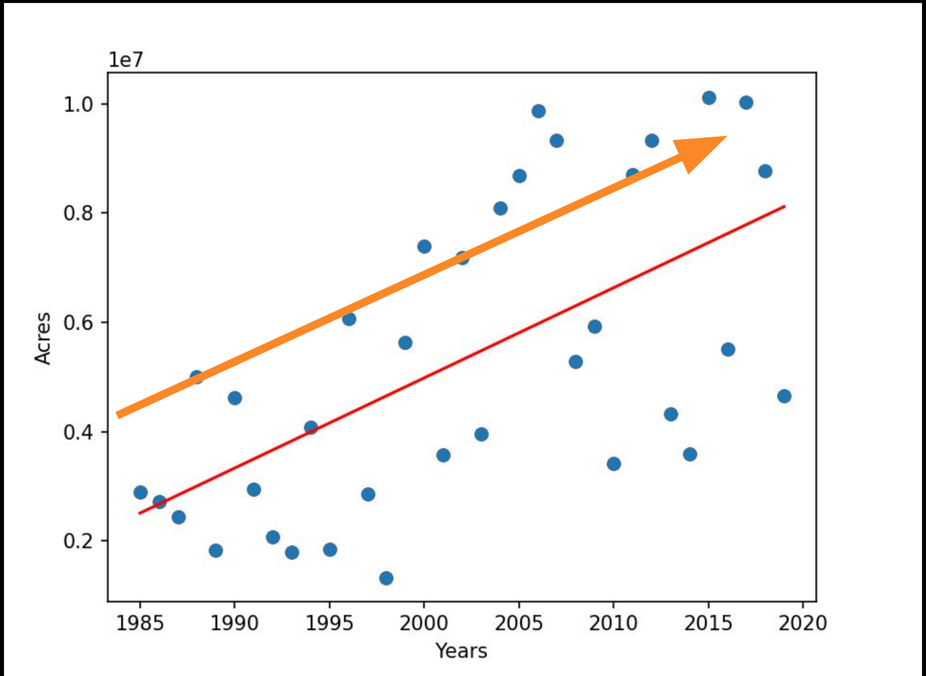
HISTORICAL

of Fires with Over 100K Acres Burned Since 1985



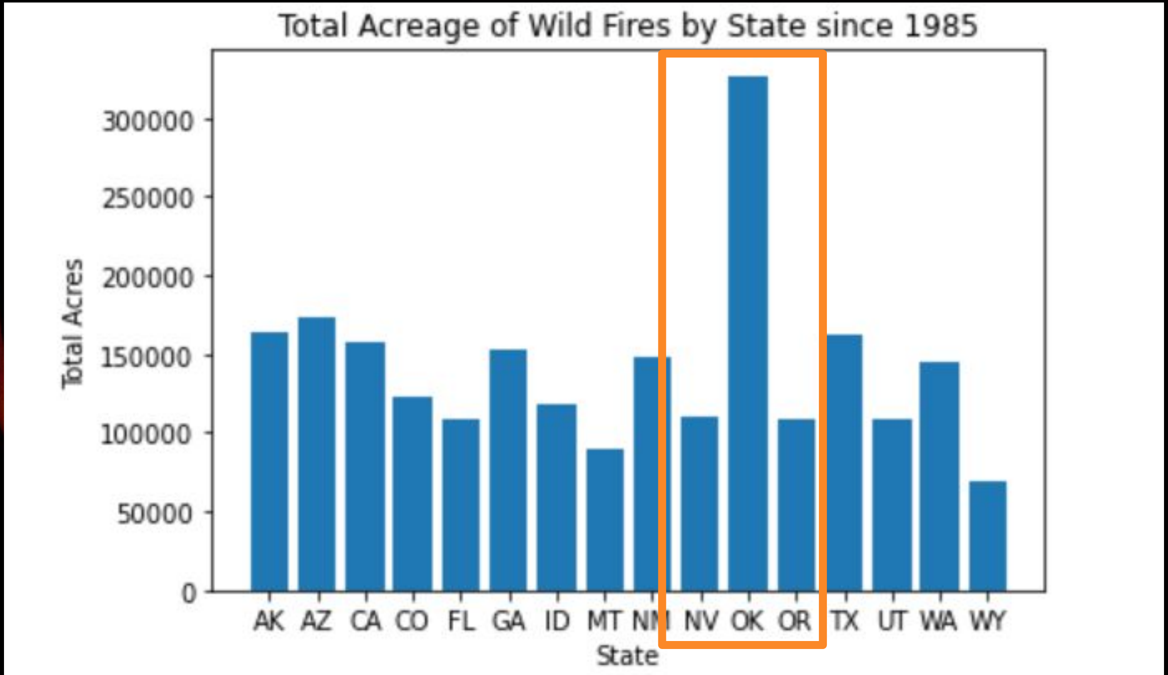
HISTORICAL

of Acres Since 1985



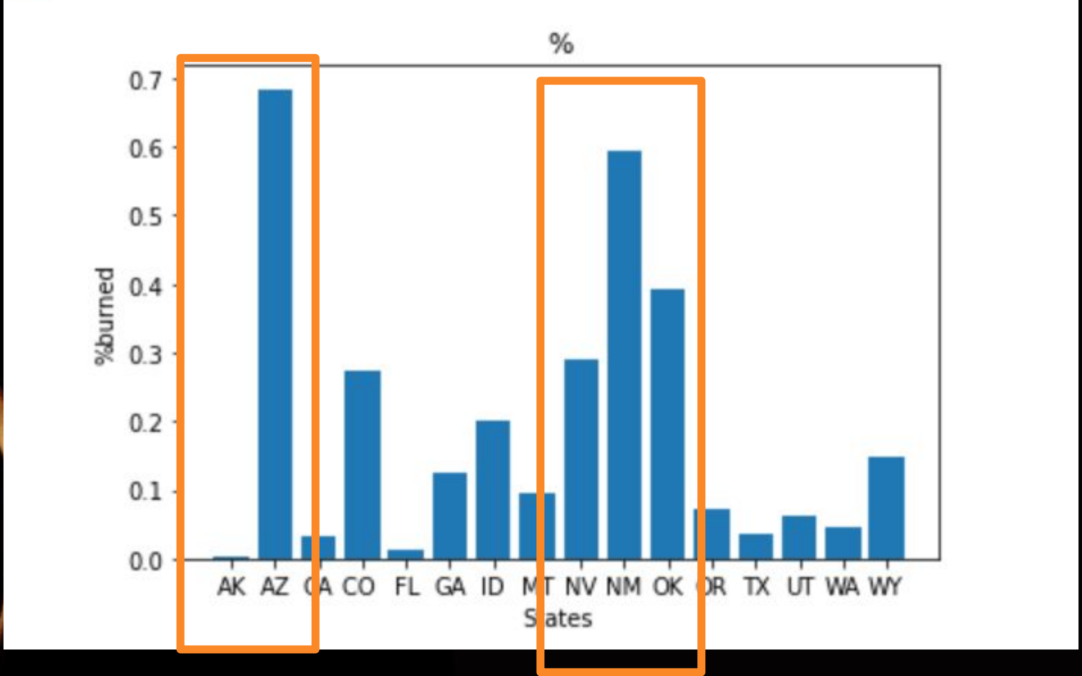
GEOGRAPHICAL

of Fires Over 100K Acres Since 1985 by State



GEOGRAPHICAL

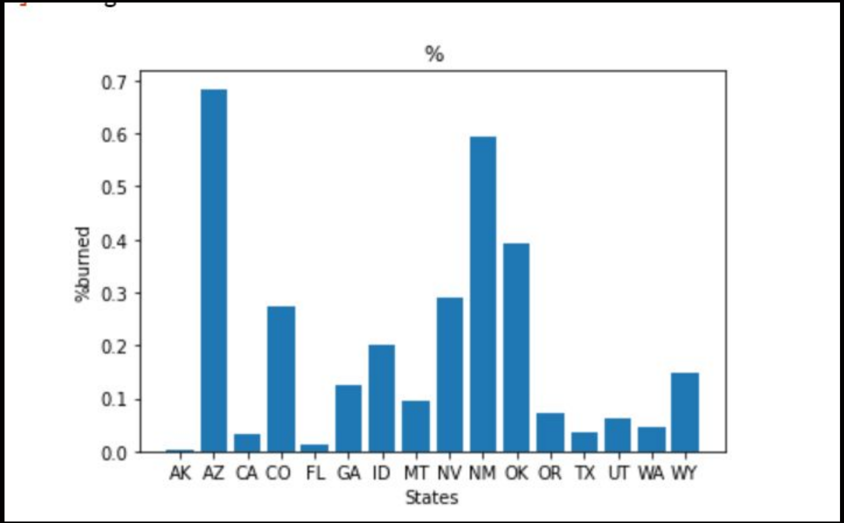
Percent Acres Burned Since 1985 by State



GEOGRAPHICAL

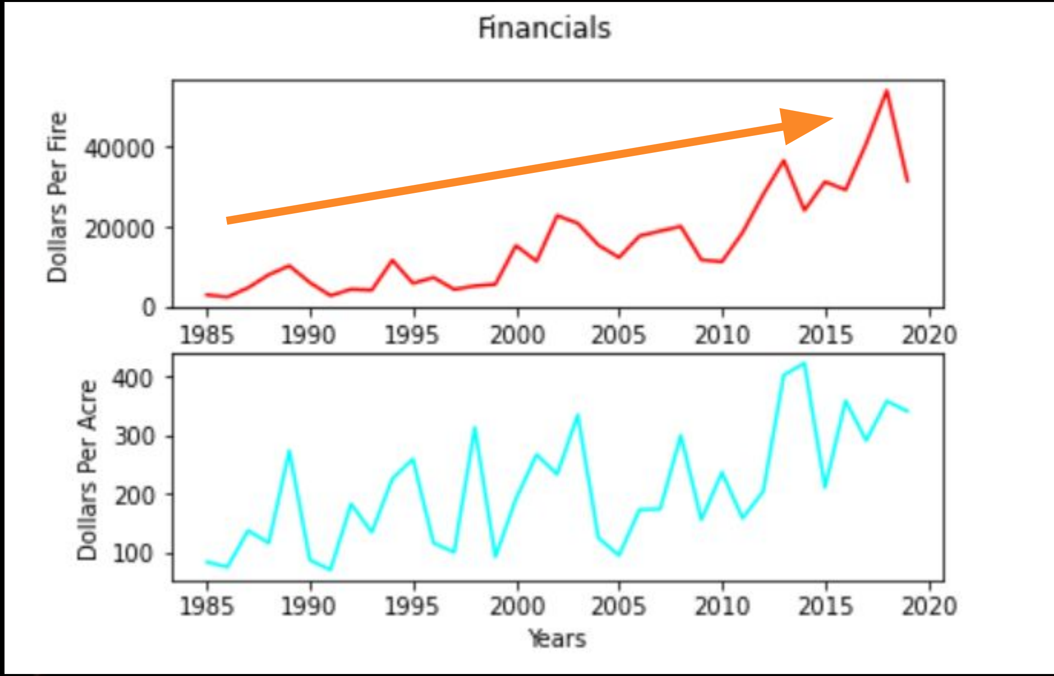
Percent Acres Burned Chi Square Test

	Percentage Burned	Expected Percentage
0	0.002714	0.19151
1	0.684243	0.19151
2	0.030981	0.19151
3	0.273945	0.19151
4	0.013982	0.19151
5	0.124636	0.19151
6	0.200042	0.19151
7	0.094127	0.19151
8	0.291501	0.19151
9	0.593038	0.19151
10	0.391785	0.19151
11	0.070678	0.19151
12	0.034501	0.19151
13	0.061793	0.19151
14	0.046882	0.19151
15	0.149306	0.19151



FINANCIAL

Suppression Spending Since 1985 by State



OF FIRES ARE DECREASING

General trends show fires as decreasing, but increasing in size and intensity

DON'T LIVE ON THE WEST COAST

The most number of wildfires occured on West Coast/4-Corners Region

COST IS INCREASING

The cost of extinguishing fires is increasing like the rate of inflation

SO NOW WHAT?

NEXT STEPS

LOOKING INTO CAUSES

- Natural v Human-error?

FURTHER GEOGRAPHICAL RESEARCH

- Were there states that didn't have a fire over 100K acres but have lots of smaller wildfires?

FINANCIAL DETAILS

- What was the suppression cost spent on?
- What were the other costs involved? How did those costs change over the years relative to the number of fires and burned?

THANK YOU!

ANY QUESTIONS?

APPENDIX A

Jupyter Notebook
Screenshots

In [1]:
```
%matplotlib notebook
```

In [12]:
```
from matplotlib import pyplot as plt
from scipy.stats import linregress
import scipy.stats as stats
import numpy as np
import pandas as pd
```

In [13]:
```
Acre_Fire = "Resources/total_wildland_fires_acres.csv"
```

In [14]:
```
Acre_Fire_df = pd.read_csv(Acre_Fire)
```

In [15]:
```
Acre_Fire_df
```

Out[15]:

	Year	Fires	Acres
0	2019	50477	4664364
1	2018	58083	8767492
2	2017	71499	10026086
3	2016	67743	5509995
4	2015	68151	10125149
5	2014	63312	3595613
6	2013	47579	4319546
7	2012	67774	9326238
8	2011	74126	8711367
9	2010	71971	3422724
10	2009	78792	5921786
11	2008	78979	5292468

In [16]:
```
Acre_plot=plt.scatter(Acre_Fire_df['Year'],Acre_Fire_df['Acres'], color='red')
plt.xlabel("years")
plt.ylabel("Acres")
```

In [17]:
```python
x_values = Acre_Fire_df['Year']
y_values = Acre_Fire_df['Acres']
```

In [18]:
```python
(slope, intercept, rvalue, pvalue, stderr) = linregress(x_values, y_values)
regress_values = x_values * slope + intercept
line_eq = "y = " + str(round(slope,2)) + "x + " + str(round(intercept,2))
plt.scatter(x_values,y_values)
plt.plot(x_values,regress_values,"r-")
plt.annotate(line_eq,(6,10),fontsize=15,color="red")
plt.xlabel('Years')
plt.ylabel('Acres')
print(f"The r-squared is: {rvalue**2}")
print (line_eq)
plt.show()
```

```
The r-squared is: 0.37368463133392654
y = 165024.71x + -325067569.77
```

In [1]:
```
%matplotlib notebook
```

In [12]:
```
from matplotlib import pyplot as plt
from scipy.stats import linregress
import scipy.stats as stats
import numpy as np
import pandas as pd
```

In [13]:
```
Acre_Fire = "Resources/total_wildland_fires_acres.csv"
```

In [14]:
```
Acre_Fire_df = pd.read_csv(Acre_Fire)
```

In [15]:
```
Acre_Fire_df
```

0	2019	50477	4664364
1	2018	58083	8767492
2	2017	71499	10026086
3	2016	67743	5509995
4	2015	68151	10125149
5	2014	63312	3595613
6	2013	47579	4319546
7	2012	67774	9326238
8	2011	74126	8711367
9	2010	71971	3422724
10	2009	78792	5921786
11	2008	78979	5292468
12	2007	85705	9328045

In [24]:
```
Acre_plot=plt.scatter(Acre_Fire_df['Year'],Acre_Fire_df['Fires'], color='red')
plt.xlabel("years")
plt.ylabel("Fires")
```

```
In [25]: x_values = Acre_Fire_df['Year']
         y_values = Acre_Fire_df['Fires']
```

```
In [26]: (slope, intercept, rvalue, pvalue, stderr) = linregress(x_values, y_values)
         regress_values = x_values * slope + intercept
         line_eq = "y = " + str(round(slope,2)) + "x + " + str(round(intercept,2))
         plt.scatter(x_values,y_values)
         plt.plot(x_values,regress_values,"r-")
         plt.annotate(line_eq,(6,10),fontsize=15,color="red")
         plt.xlabel('Years')
         plt.ylabel('Fires')
         print(f"The r-squared is: {rvalue**2}")
         print (line_eq)
         plt.show()

         The r-squared is: 0.09959109721536608
         y = -390.12x + 854546.69
```

In [3]:
```python
from matplotlib import pyplot as plt
from scipy.stats import linregress
import scipy.stats as stats
import numpy as np
import pandas as pd
```

In [4]:
```python
Acre_Fire = "Resources/100k_acre_fires.csv"
```

In [5]:
```python
Acre_Fire_df = pd.read_csv(Acre_Fire)
```

In [8]:
```python
Acre_Fire_df
```

Out[8]:

	Year	Fire Name	# of Fires (if applicable	State	Total Acres	Avg Acres
0	1997	Inowak	NaN	AK	610000.0	610000.00000
1	1999	Big Bar Complex	2.0	CA	140948.0	70474.00000
2	1999	Mule Butte	NaN	ID	138220.0	138220.00000
3	1999	Dun Glen Complex	9.0	NV	361658.0	40184.22222
4	1999	Sadler Complex	NaN	NV	297000.0	297000.00000
...	...	...	...	...	...	...
211	2019	Black River	NaN	AK	107078.0	107078.00000
212	2019	North River	NaN	AK	101451.0	101451.00000
213	2019	Woodbury	NaN	AZ	123875.0	123875.00000
214	2019	Sheep	NaN	ID	112106.0	112106.00000
215	2020	Frozen Calf	NaN	AK	240543.0	240543.00000

216 rows × 6 columns

In [19]:
```python
Acre_Fire_df['Year'].value_counts()
```

Out[19]:
```
2004    21
2015    19
2007    16
2006    15
2017    14
2012    14
2011    14
2018    12
2009    11
2002    10
2005     9
2000     9
1999     9
2013     9
2019     8
2016     6
2008     5
2014     5
2003     3
2010     3
2001     2
```

In [17]:
```python
Acre_Fire_df['Year'].unique()
```

Out[17]:
```
array([1997, 1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008,
       2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019,
       2020], dtype=int64)
```

In [21]:
```python
plt.plot(Acre_Fire_df['Year'].unique(),Acre_Fire_df['Year'].value_counts(), color='red')
```

In [1]:
```
%matplotlib notebook
%matplotlib inline
```

In [2]:
```
import matplotlib.pyplot as plt
from scipy.stats import linregress
import scipy.stats as stats
from scipy.stats import chisquare
import numpy as np
import pandas as pd
```

In [3]:
```
Geo_Data = "C:\\UPenn\\Wildfire Project\\Wildfire Project\\100K_acre_fires.csv"
```

In [4]:
```
Geo_Data_df = pd.read_csv(Geo_Data)
```

In [5]:
```
Geo_Data_df
```

Out[5]:

	Year	Fire Name	# of Fires (if applicable	State	Total Acres	Avg Acres
0	1997	Inowak	1	AK	610000.0	610000.00000
1	1999	Big Bar Complex	2	CA	140948.0	70474.00000
2	1999	Mule Butte	1	ID	138220.0	138220.00000
3	1999	Dun Glen Complex	9	NV	361658.0	40184.22222
4	1999	Sadler Complex	1	NV	297000.0	297000.00000
...	...	...	...	...	...	...
211	2019	Black River	1	AK	107078.0	107078.00000
212	2019	North River	1	AK	101451.0	101451.00000
213	2019	Woodbury	1	AZ	123875.0	123875.00000
214	2019	Sheep	1	ID	112106.0	112106.00000
215	2020	Frozen Calf	1	AK	240543.0	240543.00000

216 rows × 6 columns

In [6]:

```python
Grouped_USA_df = Geo_Data_df.groupby(['State'])

print (Grouped_USA_df)

Grouped_USA_df.count()
```

```
<pandas.core.groupby.generic.DataFrameGroupBy object at 0x0000029CBD922208>
```

Out[6]:

State	Year	Fire Name	# of Fires (if applicable	Total Acres	Avg Acres
AK	65	65	65	65	65
AZ	6	6	6	6	6
CA	24	24	24	24	24
CO	2	2	2	2	2
FL	1	1	1	1	1
GA	5	5	5	5	5
ID	28	28	28	28	28
MT	11	11	11	11	11
NM	4	4	4	4	4
NV	25	25	25	25	25
OK	4	4	4	4	4
OR	17	17	17	17	17
TX	11	11	11	11	11
UT	3	3	3	3	3
WA	9	9	9	9	9
WY	1	1	1	1	1

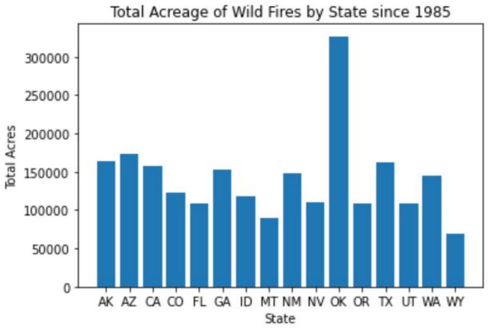
In [7]:

```python
Grouped_USA_df['Avg Acres'].apply(pd.to_numeric)
Acerage_Fire = Grouped_USA_df['Avg Acres'].median()
Acerage_Fire
```

Out[7]:

```
State
AK    164542.0
AZ    173414.5
```

In [8]:
```python
plt.bar(Acerage_Fire.index.values,Acerage_Fire.values)
plt.xlabel("State")
plt.ylabel("Total Acres")
plt.title("Total Acreage of Wild Fires by State since 1985")
plt.show()
```



In [9]:
```python
Area_data = "C:\\UPenn\\Wildfire Project\\Wildfire Project\\State Area Measurements.csv"
```

In [10]:
```python
Area_data_df = pd.read_csv(Area_data)
```

In [20]:
```python
Area_data_df
```

Out[20]:

	States	Sq. Mi.	Acres	Acres Burned	Percentage Burned	Expected Percentage
0	AK	94743	60635520	164542.0	0.002714	0.19151
1	AZ	396	253440	173414.5	0.684243	0.19151
2	CA	7916	5066240	156956.5	0.030981	0.19151
3	CO	701	448640	122902.5	0.273945	0.19151
4	FL	12133	7765120	108574.0	0.013982	0.19151
5	GA	1912	1223680	152515.0	0.124636	0.19151
6	ID	926	592640	118553.0	0.200042	0.19151
7	MT	1494	956160	90000.0	0.094127	0.19151

In [9]:
```python
Area_data = "C:\\UPenn\\Wildfire Project\\Wildfire Project\\State Area Measurements.csv"
```

In [10]:
```python
Area_data_df = pd.read_csv(Area_data)
```

In [20]:
```python
Area_data_df
```

Out[20]:

	States	Sq. Mi.	Acres	Acres Burned	Percentage Burned	Expected Percentage
0	AK	94743	60635520	164542.0	0.002714	0.19151
1	AZ	396	253440	173414.5	0.684243	0.19151
2	CA	7916	5066240	156956.5	0.030981	0.19151
3	CO	701	448640	122902.5	0.273945	0.19151
4	FL	12133	7765120	108574.0	0.013982	0.19151
5	GA	1912	1223680	152515.0	0.124636	0.19151
6	ID	926	592640	118553.0	0.200042	0.19151
7	MT	1494	956160	90000.0	0.094127	0.19151
8	NV	791	506240	147569.5	0.291501	0.19151
9	NM	292	186880	110827.0	0.593038	0.19151
10	OK	1304	834560	326968.0	0.391785	0.19151

In [12]:
```python
Area_data_df["Acres Burned"]= Acerage_Fire.values
```

In [37]:
```python
((Area_data_df["Percentage Burned"]- Area_data_df["Expected Percentage"])**2/Area_data_df["Percentage Burned"]).sum()
```

Out[37]: 18.797603491891895

In [19]:
```python
Area_data_df["Expected Percentage"] = Area_data_df["Percentage Burned"].mean()
```

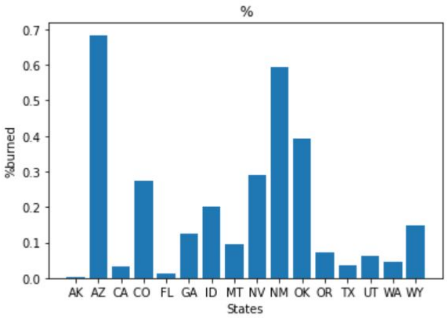
In [21]:
```python
Area_data_df["Percentage Burned"]=Area_data_df["Acres Burned"]/Area_data_df["Acres"]
Area_data_df
```

Out[21]:

	States	Sq. Mi.	Acres	Acres Burned	Percentage Burned	Expected Percentage
0	AK	94743	60635520	164542.0	0.002714	0.19151
1	AZ	396	253440	173414.5	0.684243	0.19151

In [40]:
```
plt.bar(Area_data_df["States"],Area_data_df["Percentage Burned"])
plt.xlabel("States")
plt.ylabel("%burned")
plt.title("%")
plt.figure()
```

Out[40]: <Figure size 432x288 with 0 Axes>



<Figure size 432x288 with 0 Axes>

In [22]:
```
chi_square_df.dtypes
```

Out[22]:
```
States                  object
Percentage Burned      float64
dtype: object
```

In [39]:
```
chi_square_df = Area_data_df [["Percentage Burned", "Expected Percentage"]]
critical_value = stats.chi2.ppf(q = 0.95, df=15)
stats.chisquare(chi_square_df["Percentage Burned"], chi_square_df["Expected Percentage"])
```

Out[39]: Power_divergenceResult(statistic=3.3765918705166422, pvalue=0.9991700155518832)

In [33]:
```
critical_value
```

Out[33]: 1.3233036969314664

In [39]:
```
chi_square_df = Area_data_df [["Percentage Burned", "Expected Percentage"]]
critical_value = stats.chi2.ppf(q = 0.95, df=15)
stats.chisquare(chi_square_df["Percentage Burned"], chi_square_df["Expected Percentage"])
```

Out[39]:
```
Power_divergenceResult(statistic=3.3765918705166422, pvalue=0.9991700155518832)
```

In [33]:
```
critical_value
```

Out[33]:
```
1.3233036969314664
```

In [31]:
```
chi_square_df
```

Out[31]:

	Percentage Burned	Expected Percentage
0	0.002714	0.19151
1	0.684243	0.19151
2	0.030981	0.19151
3	0.273945	0.19151
4	0.013982	0.19151
5	0.124636	0.19151
6	0.200042	0.19151
7	0.094127	0.19151
8	0.291501	0.19151
9	0.593038	0.19151
10	0.391785	0.19151
11	0.070678	0.19151
12	0.034501	0.19151
13	0.061793	0.19151
14	0.046882	0.19151
15	0.149306	0.19151

In [1]:
```
%matplotlib notebook
%matplotlib inline
```

In [2]:
```
from matplotlib import pyplot as plt
from scipy.stats import linregress
import numpy as np
import pandas as pd
```

In [4]:
```
Fire_cost = "C:\\UPenn\\Wildfire Project\\Wildfire Project\\Firecost.csv"
```

In [5]:
```
Fire_cost_df = pd.read_csv(Fire_cost)
```

In [6]:
```
Fire_cost_df
```

Out[6]:

	Year	Fires	Acres	Forest Service	DOI Agencies	Total
0	1985	82591	2896147	161505000	78438000	239943000
1	1986	85907	2719162	111625000	91153000	202778000
2	1987	71300	2447296	253657000	81452000	335109000
3	1988	72750	5009290	429609000	149317000	578926000
4	1989	48949	1827310	331672000	168115000	499787000
5	1990	66481	4621621	253700000	144252000	397952000
6	1991	75754	2953578	132300000	73820000	206120000
7	1992	87394	2069929	290300000	87166000	377466000
8	1993	58810	1797574	184000000	56436000	240436000
9	1994	79107	4073579	757200000	161135000	918335000
10	1995	82234	1840546	367000000	110126000	477126000
11	1996	96363	6065998	547500000	153683000	701183000
12	1997	66196	2856959	179100000	105048000	284148000
13	1998	81043	1329704	306800000	109904000	416704000
14	1999	92487	5626093	361100000	154416000	515516000
15	2000	92250	7383493	1076000000	334802000	1410802000
16	2001	84079	3570911	683122000	269574000	952696000

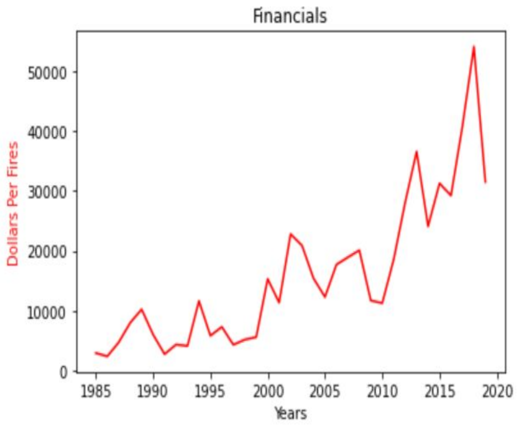
In [7]:

```
Fire_cost_df["Dollars Per Fire"]= Fire_cost_df["Total"]/ Fire_cost_df["Fires"]
Fire_cost_df["Dollars Per Acre"]= Fire_cost_df["Total"]/ Fire_cost_df["Acres"]
```

In [8]:

```
plt.plot(Fire_cost_df['Year'], Fire_cost_df['Dollars Per Fire'], color ='red')
plt.xlabel("Years")
plt.ylabel("Dollars Per Fires", color ='red')
plt.title("Financials")
plt.figure()
```

Out[8]: <Figure size 432x288 with 0 Axes>



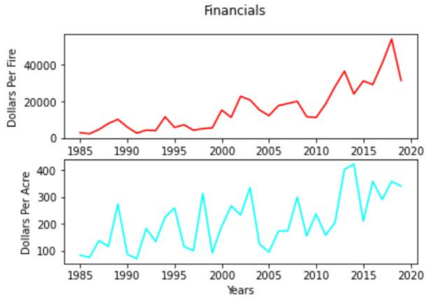
<Figure size 432x288 with 0 Axes>

In [9]:

```
fig, (ax1, ax2) = plt.subplots(2)
fig.suptitle("Financials")
ax1.plot(Fire_cost_df['Year'], Fire_cost_df['Dollars Per Fire'], color ='red')
ax2.plot(Fire_cost_df['Year'], Fire_cost_df['Dollars Per Acre'], color ='cyan')
plt.xlabel("Years")
ax1.set(ylabel="Dollars Per Fire")
ax2.set(ylabel="Dollars Per Acre")
```

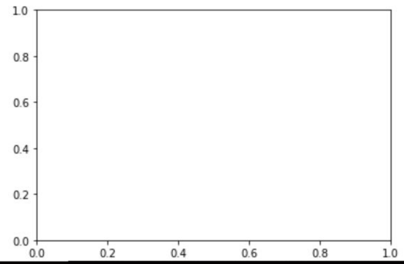
Out[9]:

```
[Text(0, 0.5, 'Dollars Per Acre')]
```



In [10]:

```
fig, ax = plt.subplots()
```



APPENDIX B

Data Source Descriptions

NATIONAL INTERAGENCY FIRE CENTER (NFIC)

ABOUT SOURCE

- Government training organization supporting emergency response training
- Organizes and allocates resources to different federal, state, and local government agencies
  - Ex: National Park Service, Departments of Agriculture and Interior, etc.

DATA

- Total Wildland Fires and Acres (1926-2019)
- Wildfires larger than 100,000+ acres (1997-2019)
- Suppression Costs (1985-2019)