

```
In [1]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import math
```

```
In [2]: df = pd.read_csv('amazon.csv', encoding='latin1')
df.head()
```

Out[2]:

	year	state	month	number	date
0	1998	Acre	Janeiro	0.0	1998-01-01
1	1999	Acre	Janeiro	0.0	1999-01-01
2	2000	Acre	Janeiro	0.0	2000-01-01
3	2001	Acre	Janeiro	0.0	2001-01-01
4	2002	Acre	Janeiro	0.0	2002-01-01

```
In [3]: def new_month(old, new_m):
df.month.replace(old, new_m,inplace=True)

new_month('janeiro', 'January')
new_month('Fevereiro', 'February')
new_month('Marco', 'March')
new_month('Abril', 'April')
new_month('Maio', 'May')
new_month('Junho', 'June')
new_month('Juhlio', 'July')
new_month('Agosto', 'August')
new_month('Setembro', 'September')
new_month('Outubro', 'October')
new_month('Novembro', 'November')
new_month('Dezembro', 'December')
```

```
In [4]: df.head()
```

Out[4]:

	year	state	month	number	date
0	1998	Acre	Janeiro	0.0	1998-01-01
1	1999	Acre	Janeiro	0.0	1999-01-01
2	2000	Acre	Janeiro	0.0	2000-01-01
3	2001	Acre	Janeiro	0.0	2001-01-01
4	2002	Acre	Janeiro	0.0	2002-01-01

```
In [5]: df.describe()
```

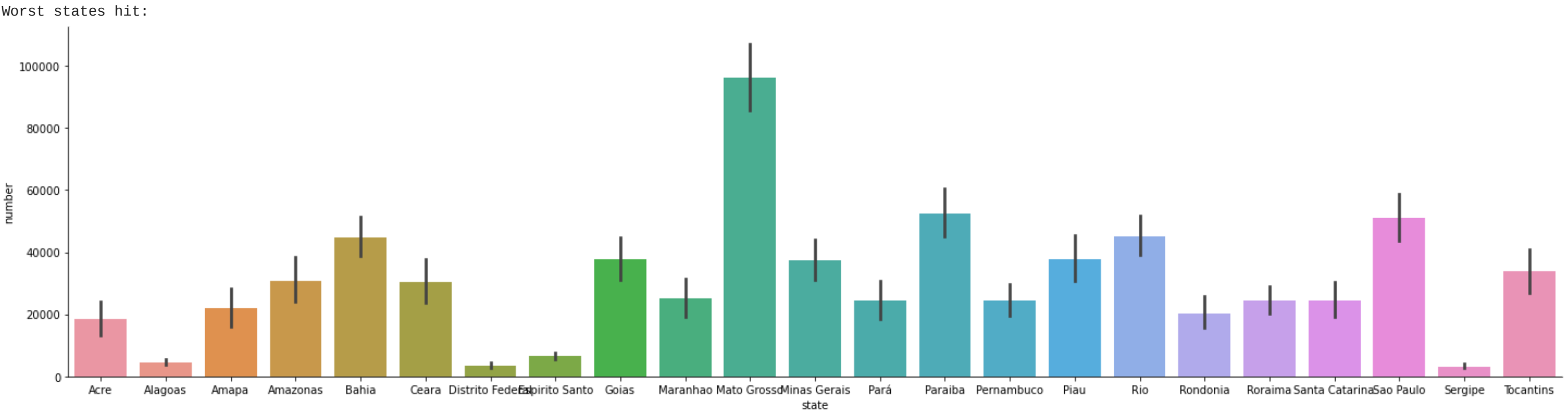
Out[5]:

	year	number
count	6454.000000	6454.000000
mean	2007.461729	108.293163
std	5.746654	190.812242
min	1998.000000	0.000000
25%	2002.000000	3.000000
50%	2007.000000	24.000000
75%	2012.000000	113.000000
max	2017.000000	998.000000

```
In [13]: by_state = df.groupby(['year'], as_index=False).sum()

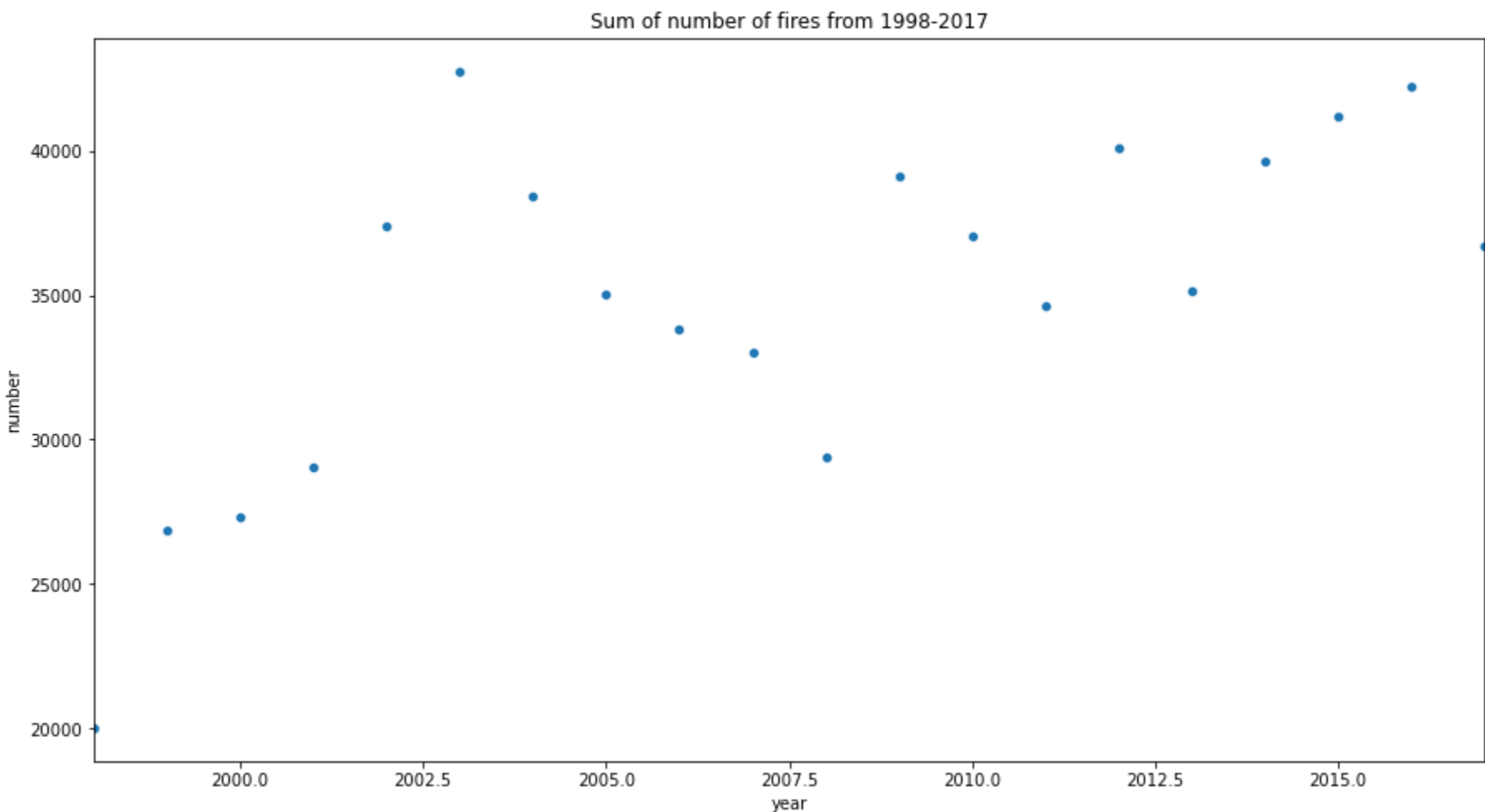
sns.catplot(x='state', y='number',data=df[['state','number']], kind='bar', aspect=4, estimator=sum);

worst_hit = by_state[by_state['number']> by_state['number'].mean() +by_state['number'].std()]
print('Worst states hit: ')
#for i in range(len(worst_hit)):
#    print(worst_hit['state'].values[i])
```



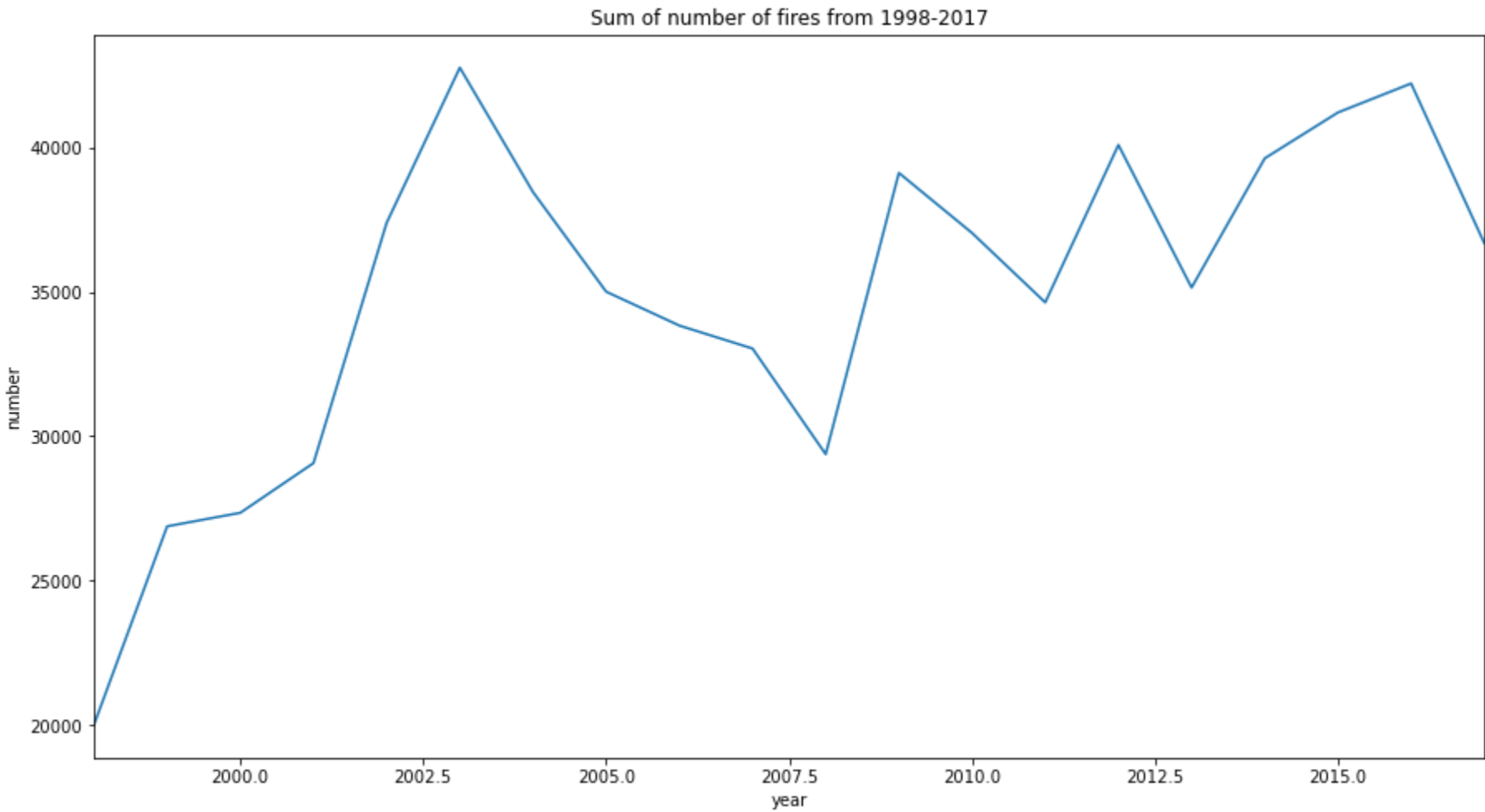
```
In [16]: by_year = df.groupby(['year'], as_index=False).sum()

plt.figure(figsize=[15,8])
plt.xlim([1998, 2017])
plt.title('Sum of number of fires from 1998-2017')
sns.scatterplot(x='year',y='number',data=by_year);
```



```
In [17]: by_year = df.groupby(['year'], as_index=False).sum()

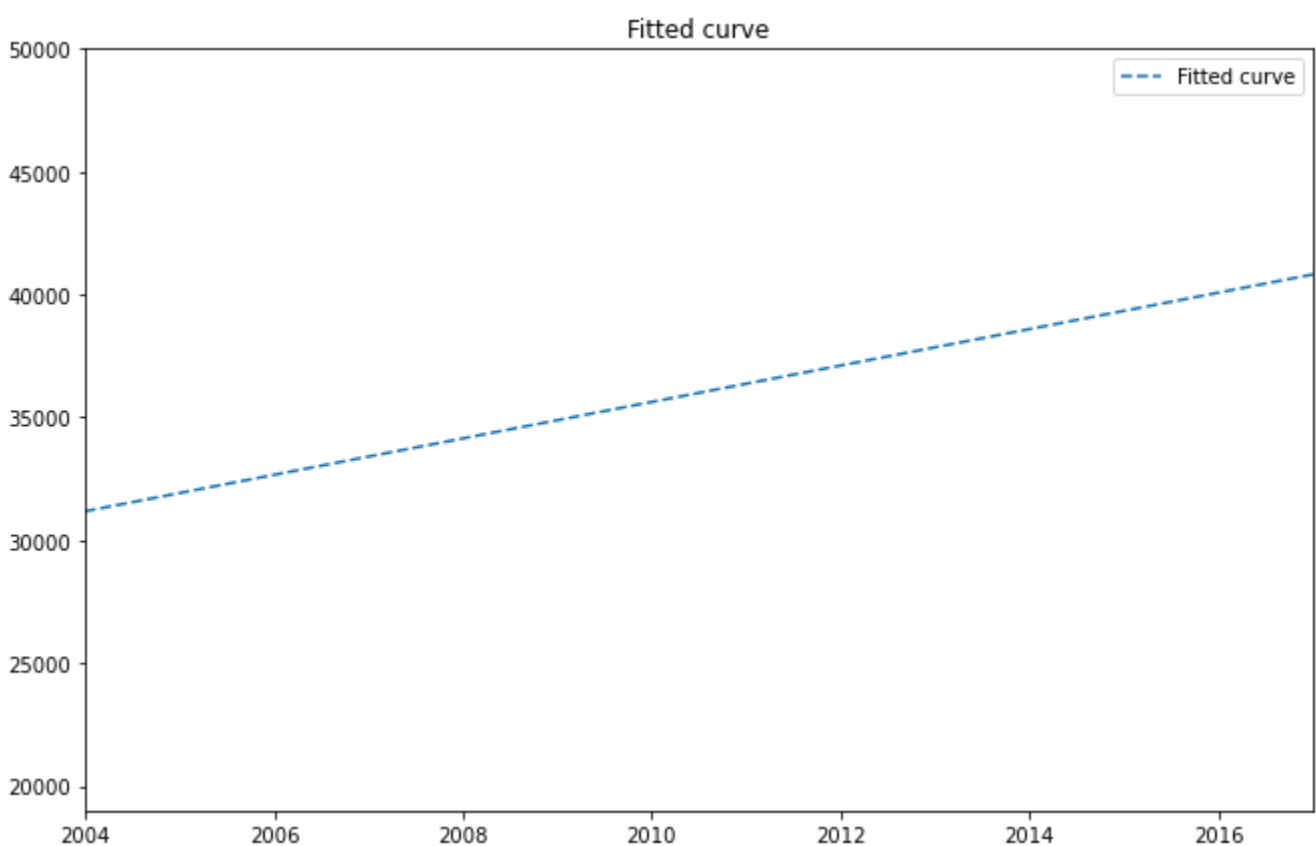
plt.figure(figsize=[15,8])
plt.xlim([1998, 2017])
plt.title('Sum of number of fires from 1998-2017')
sns.lineplot(x='year',y='number',data=by_year);
```



```
In [32]: by_year = by_year[by_year['year']>2004]
z = np.poly1d(np.polyfit(by_year['year'],by_year['number'],1))

years = np.linspace(1998, 2017, 13)

plt.figure(figsize=[11,7])
#plt.plot(years, by_year['number'], '-', label='Raw data')
plt.plot(years,z(years), '---', label='Fitted curve')
plt.xlim([2004, 2017])
plt.ylim([19000, 50000])
plt.title('Fitted curve')
plt.legend()
plt.show()
```



```
In [ ]:
```