In [1]:
```
import pandas as pd #install pandas using pip
import os
import numpy as np #numpy module is used to perform sum on rows/columns
import matplotlib.pyplot as plt #to plot graphs
import seaborn as sns
```

In [2]:
```
pwd  #current directory
```

Out[2]:
```
'C:\\Users\\susan\\Desktop\\Data anlaysis on forest fires'
```

In [3]:
```
df = pd.read_csv(r"D:\Data anlaysis on forest fires\archive\amazon.csv")  #importing the csv files
```

In [4]:
```
df.head()  #checking if the csv file is imported using head()-returns first 5 rows of the dataset
```

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]:
```
# headers
print(df.columns)
```

```
Index(['year', 'state', 'month', 'number', 'date'], dtype='object')
```

In [6]:
```
#translating the month names
months_rename = {'Janeiro': 'January', 'Fevereiro': 'February', 'Março': 'March', 'Abril': 'April', 'Maio': 'May',
                 'Junho': 'June', 'Julho': 'July', 'Agosto': 'August', 'Setembro': 'September', 'Outubro': 'October',
                 'Novembro': 'November', 'Dezembro': 'December'}
#mmapping the month names
df['month'] = df['month'].map(months_rename)
```

In [7]:
```
df.month.unique()
```

Out[7]:
```
array(['January', 'February', 'March', 'April', 'May', 'June', 'July',
       'August', 'September', 'October', 'November', 'December'],
      dtype=object)
```

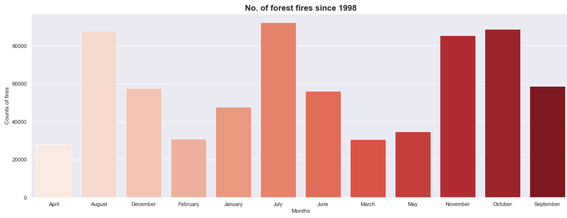
In [8]:
```
#total no.fires by month
pivot2 = pd.pivot_table(df,values = "number",index = ["month"], aggfunc=np.sum)
pivot2
```

Out[8]:

month	number
April	28188.770
August	88050.435
December	57535.480
February	30848.050
January	47747.844
July	92326.113
June	56010.675
March	30717.405
May	34731.363
November	85508.054
October	88681.579
September	58578.305

In [9]:
```
#plotting a graph
sns.set()
plt.figure(figsize=(20,7))
ax = sns.barplot(x=pivot2.index, y='number', data=pivot2,palette='Reds')
ax.set_xlabel('Months')
ax.set_ylabel('Counts of fires')
ax.set_title('No. of forest fires since 1998',fontdict={'fontsize': '17', 'fontweight' : 'bold'})
```

Out[9]:
```
Text(0.5, 1.0, 'No. of forest fires since 1998')
```



Conclusions:

1. In Februaru,March,April and May the lowest no. of forest fires occured
2. In the four months(July,august,October and November) the maximum no. forest fires occured

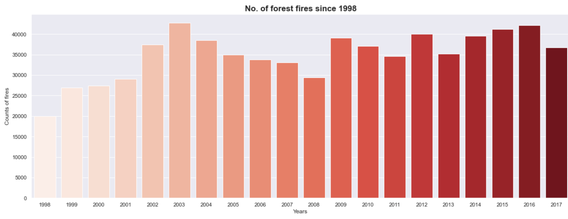
In [10]:
```
# total fires reported by year

pivot1 = pd.pivot_table(df,values = "number",index = ["year"], aggfunc=np.sum)
pivot1
```

Out[10]:

year	number
1998	20013.971
1999	26882.821
2000	27351.251
2001	29071.612
2002	37390.600
2003	42760.674
2004	38453.163
2005	35004.965
2006	33832.161
2007	33037.413
2008	29378.964
2009	39117.178
2010	37037.449
2011	34633.545
2012	40084.860
2013	35146.118
2014	39621.183
2015	41208.292
2016	42212.229
2017	36685.624

In [11]:
```
#plotting a graph
plt.figure(figsize=(20,7))
ax = sns.barplot(x=pivot1.index, y='number', data=pivot1,palette='Reds')
ax.set_xlabel('Years')
ax.set_ylabel('Counts of fires')
ax.set_title('No. of forest fires since 1998',fontdict={'fontsize': '17', 'fontweight' : 'bold'})
```

Out[11]:
```
Text(0.5, 1.0, 'No. of forest fires since 1998')
```



Conclusions:

1. The forest fires have been increasing since 1998 with a large spike in 2002
2. After the year 2003, there has been a decline in the no of forest fires
3. Again in the year 2009, there was an in the no. of forest fires and continuing going up and down b/w 350k-450k

In [12]:
```
#total no.fires by state
pivot3 = pd.pivot_table(df,values = "number",index = ["state"], aggfunc=np.sum)
pivot3
```

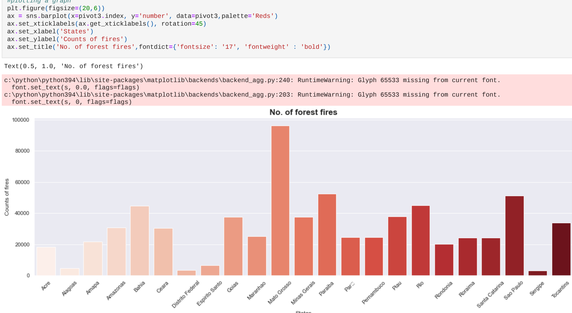
Out[12]:

state	number
Acre	18464.030
Alagoas	4644.000
Amapa	21831.575
Amazonas	30650.129
Bahia	44746.226
Ceara	30428.063
Distrito Federal	3561.000
Espirito Santo	6546.000
Goias	37095.520
Maranhao	25129.131
Mato Grosso	96246.028
Minas Gerais	37475.258
Paraiba	52435.910
Par	24512.144
Pernambuco	24498.000
Piau	37803.747
Rio	45160.865
Rondonia	20285.429
Roraima	24385.074
Santa Catarina	24359.852
Sao Paulo	51121.198
Sergipe	3237.000
Tocantins	33707.889

In [13]:
```
#plotting a graph
plt.figure(figsize=(20,6))
ax = sns.barplot(x=pivot3.index, y='number', data=pivot3,palette='Reds')
ax.set_xticklabels(ax.get_xticklabels(), rotation=45)
ax.set_xlabel('States')
ax.set_ylabel('Counts of fires')
ax.set_title('No. of forest fires',fontdict={'fontsize': '17', 'fontweight' : 'bold'})
```

Out[13]:
```
Text(0.5, 1.0, 'No. of forest fires')
```

```
c:\python\python394\lib\site-packages\matplotlib\backends\backend_agg.py:240: RuntimeWarning: Glyph 65533 missing from current font.
  font.set_text(s, 0.0, flags=flags)
c:\python\python394\lib\site-packages\matplotlib\backends\backend_agg.py:203: RuntimeWarning: Glyph 65533 missing from current font.
  font.set_text(s, 0, flags=flags)
```



Conclusions:

1. Large no. of forest fires occur in Manto Grosso
2. Alagaos, Distrito Federal,Espirito Santo and Sergipe see a small no. of fires