```
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
                                   0.0 2000-01-01
          2 2000 Acre Janeiro
          3 2001 Acre Janeiro
                                   0.0 2001-01-01
                                   0.0 2002-01-01
          4 2002 Acre Janeiro
           def new_month(old, new_m):
               df.month.replace(old, new_m,inplace=True)
           new_month('janeiro', 'January')
           new_month('Fevererio', 'February')
           new_month('Marco','March')
           new_month('Abril','April')
new_month('Maio','May')
           new_month('Junho','June')
new_month('Juhlo','July')
           new_month('Agosto','August')
           new_month('Setembro','September')
           new_month('Outubro','October')
           new_month('Novembro', 'November')
           new_month('Dezembro', 'December')
           df.head()
 In [4]:
             year state month number
 Out[4]:
                                             date
          0 1998 Acre Janeiro
                                   0.0 1998-01-01
          1 1999 Acre Janeiro
                                   0.0 1999-01-01
                                   0.0 2000-01-01
          2 2000 Acre Janeiro
          3 2001 Acre Janeiro
                                   0.0 2001-01-01
          4 2002 Acre Janeiro
                                   0.0 2002-01-01
           df.describe()
 Out[5]:
                                number
                       year
          count 6454.000000 6454.000000
          mean 2007.461729 108.293163
                   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
           by_state =df.groupby(['year'], as_index=False).sum()
In [13]:
           sns.catplot(x='state', y='number', data=df[['state', 'number']], kind='bar', aspect=4, estimator=sum);
           worst_hit = 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])
          Worst states hit:
            100000
             80000
             60000
             40000
             20000
                                                     Bahia
                                                              Ceara Distrito Fede Espirito Santo Goias
                                                                                             Maranhao Mato Grossdinas Gerais Pará
                                                                                                                                                                       Roraima Santa CatarinaSao Paulo
                                                                                                                                                                                                Sergipe
                     Acre
                            Alagoas
                                     Amapa
                                           Amazonas
                                                                                                                               Paraiba Pernambuco Piau
                                                                                                                                                         Rio
                                                                                                                                                               Rondonia
          by_year = df.groupby(['year'], as_index=False).sum()
In [16]:
           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);
                                                            Sum of number of fires from 1998-2017
            40000
            35000
       30000 2
            25000
            20000
                                                                                                         2012.5
                                                                                                                         2015.0
                            2000.0
                                           2002.5
                                                           2005.0
                                                                          2007.5
                                                                                          2010.0
                                                                           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);
                                                            Sum of number of fires from 1998-2017
            40000
            35000
       30000 V
            25000
            20000
                            2000.0
                                           2002.5
                                                           2005.0
                                                                          2007.5
                                                                                          2010.0
                                                                                                         2012.5
                                                                                                                         2015.0
                                                                           year
           by_year = by_year[by_year['year']>2004]
In [32]:
           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()
                                                      Fitted curve
           50000
                                                                                         --- Fitted curve
           45000
           40000
          35000
           30000
           25000
           20000
                                         2008
                                                      2010
                                                                   2012
                                                                                 2014
              2004
                            2006
                                                                                              2016
```