# ufo-exploration

October 28, 2018

In [1]:
```
%matplotlib inline

import warnings
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

In [2]:
```
df = pd.read_csv("../data/cleaned.csv")
df.head()
```

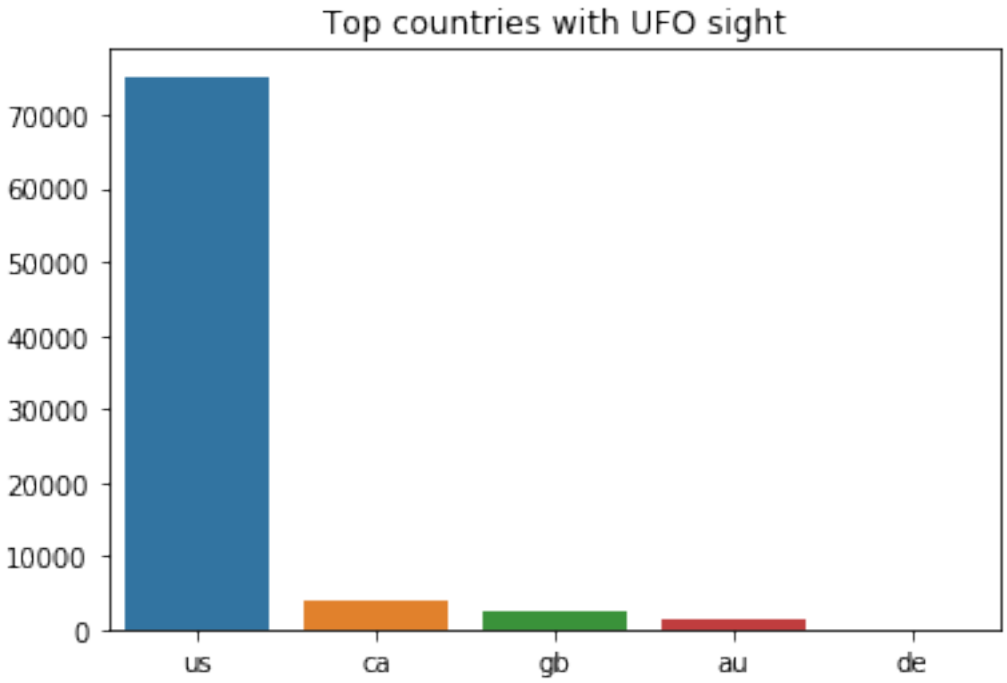
Out[2]:

| | datetime | city | state | country | latitude | longitude | duration | hour | day | month | year |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1949-10-10 20:30:00 | san marcos | tx | us | 29.883056 | -97.941111 | 2700.0 | 20 | 10 | 10 | 1949 |
| 1 | 1949-10-10 21:00:00 | lackland afb | tx | us | 29.384210 | -98.581082 | 7200.0 | 21 | 10 | 10 | 1949 |
| 2 | 1955-10-10 17:00:00 | chester (uk/england) | NaN | gb | 53.200000 | -2.916667 | 20.0 | 17 | 10 | 10 | 1955 |
| 3 | 1956-10-10 21:00:00 | edna | tx | us | 28.978333 | -96.645833 | 20.0 | 21 | 10 | 10 | 1956 |
| 4 | 1960-10-10 20:00:00 | kaneohe | hi | us | 21.418056 | -157.803611 | 900.0 | 20 | 10 | 10 | 1960 |

### 0.0.1 What areas of the country are most likely to have UFO sightings?

In [3]:
```
temp = df.country.value_counts()
sns.barplot(temp.index, temp.values).set_title('Top countries with UFO sight')
```

Out[3]: <matplotlib.text.Text at 0x10d3e26d8>



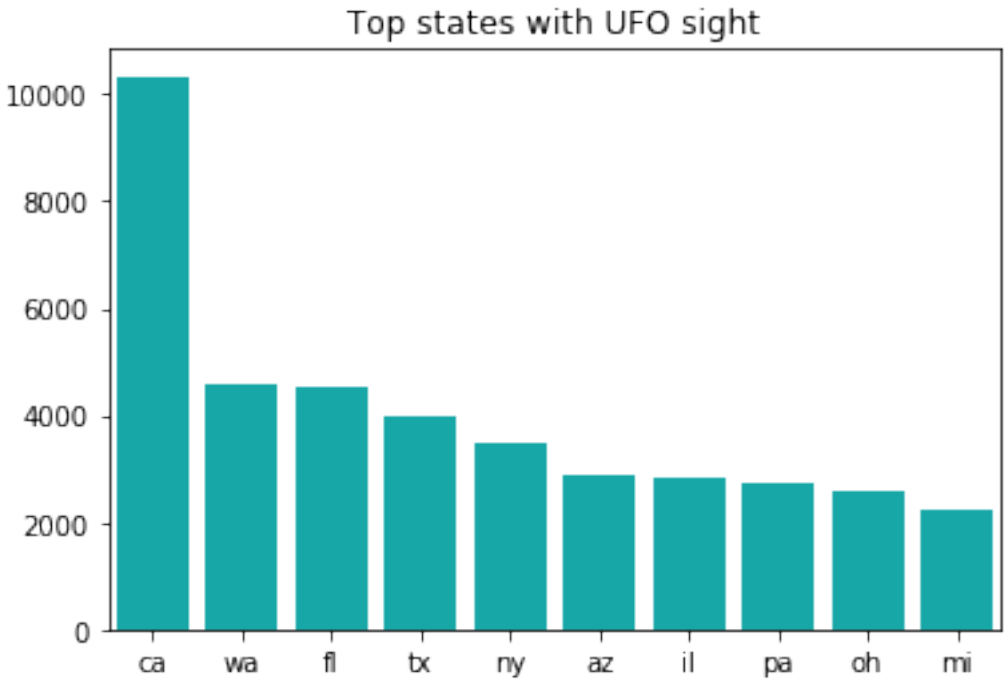

In [4]:
```
# Seem like US data dominates, for now we analyse US data only
df_us = df[df.country == "us"]
df_us.head()
```

Out[4]:

| | datetime | city | state | country | latitude | longitude | duration | hour | day | month | year |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1949-10-10 20:30:00 | san marcos | tx | us | 29.883056 | -97.941111 | 2700.0 | 20 | 10 | 10 | 1949 |
| 1 | 1949-10-10 21:00:00 | lackland afb | tx | us | 29.384210 | -98.581082 | 7200.0 | 21 | 10 | 10 | 1949 |
| 3 | 1956-10-10 21:00:00 | edna | tx | us | 28.978333 | -96.645833 | 20.0 | 21 | 10 | 10 | 1956 |
| 4 | 1960-10-10 20:00:00 | kaneohe | hi | us | 21.418056 | -157.803611 | 900.0 | 20 | 10 | 10 | 1960 |
| 5 | 1961-10-10 19:00:00 | bristol | tn | us | 36.595000 | -82.188889 | 300.0 | 19 | 10 | 10 | 1961 |

In [5]:
```
temp = df.state.value_counts()[:10]
sns.barplot(temp.index, temp.values, color="c").set_title('Top states with UFO sight')
```

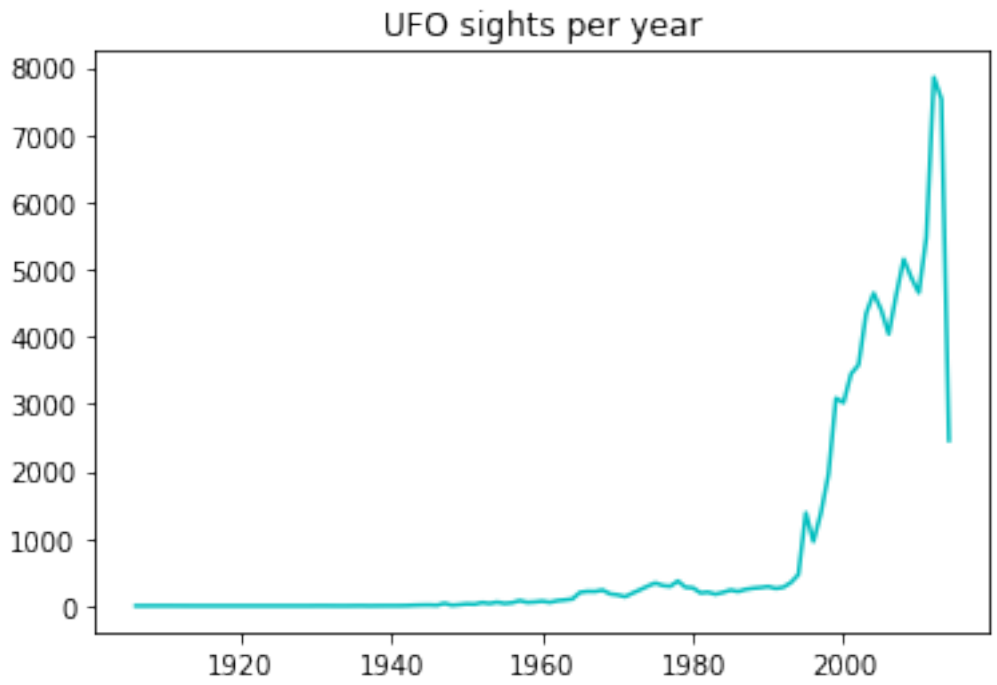
Out[5]: <matplotlib.text.Text at 0x11b55ef60>




### 0.0.2 Are there any trends in UFO sightings over time? Do they tend to be clustered or seasonal?

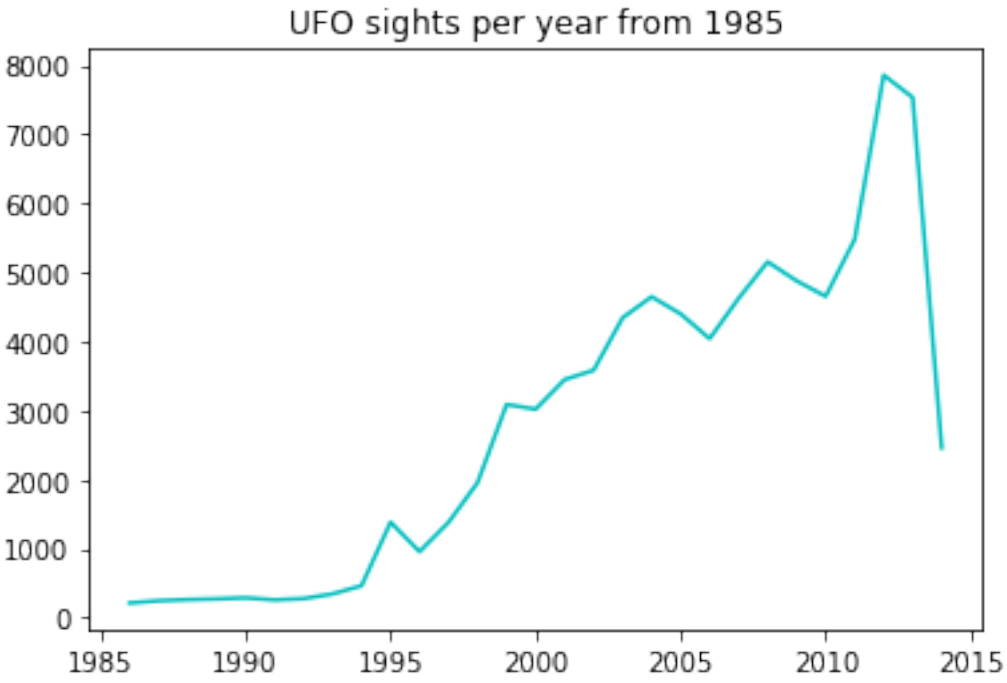
In [6]:
```
temp = df.year.value_counts()
sns.lineplot(temp.index, temp.values, color="c").set_title('UFO sights per year')
```

Out[6]: <matplotlib.text.Text at 0x11b71abe0>



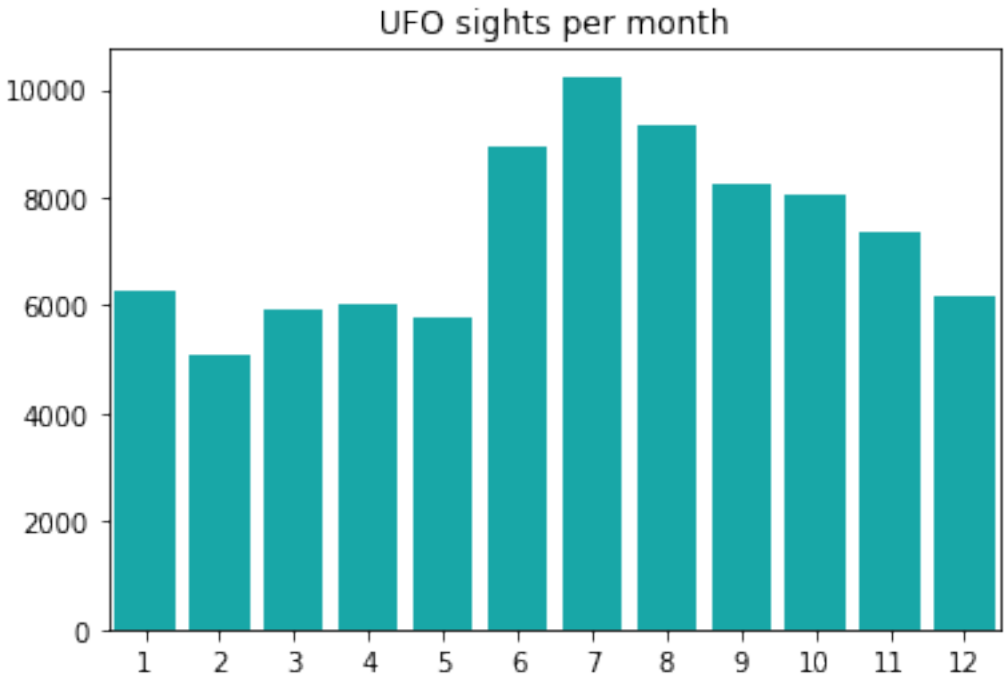

```
In [7]: temp1 = temp[temp.index > 1985]
        sns.lineplot(temp1.index, temp1.values, color="c").set_title('UFO sights per year from
```

```
Out[7]: <matplotlib.text.Text at 0x11bc4e5c0>
```



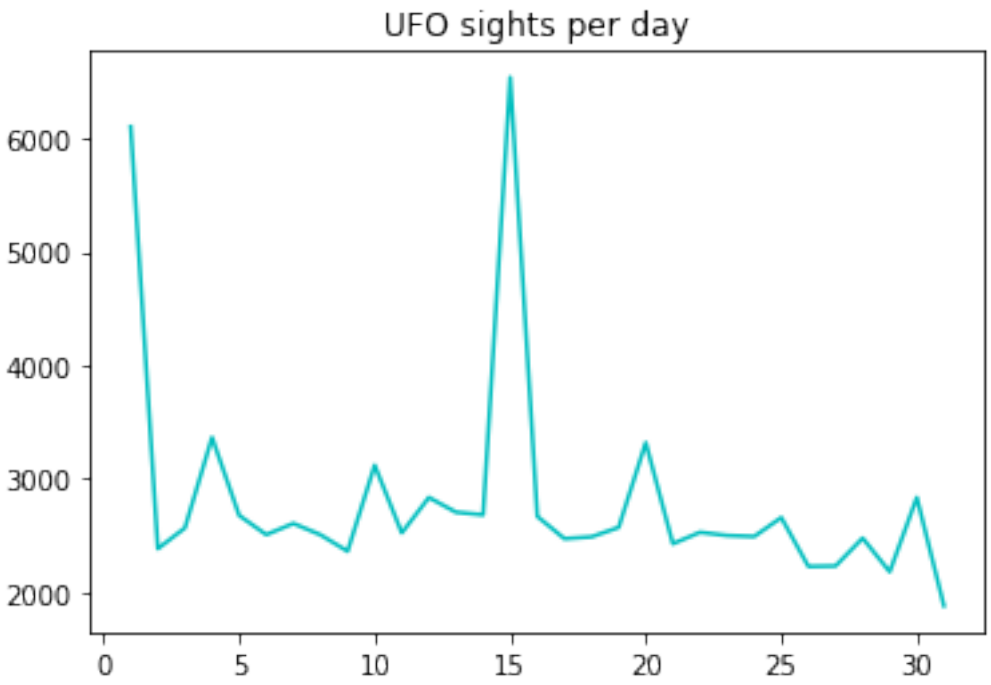
In [8]:
```
temp = df.month.value_counts()
sns.barplot(temp.index, temp.values, color="c").set_title('UFO sights per month')
```

Out[8]: <matplotlib.text.Text at 0x11bcf1d68>




In [9]:
```
temp = df.day.value_counts()
sns.lineplot(temp.index, temp.values, color="c").set_title('UFO sights per day')
```

Out[9]: <matplotlib.text.Text at 0x11be07ef0>




```
In [10]: temp = df.hour.value_counts()
         sns.lineplot(temp.index, temp.values, color="c").set_title('UFO sights per hour')

Out[10]: <matplotlib.text.Text at 0x11be98fd0>
```