Task 1

In [52]:
```
import pandas as pd
netflix_df = pd.read_csv(r'netflix_data.csv')
netflix_df_movies_only = netflix_df.query('type == "Movie" and country == "U
netflix_movies_col_subset = netflix_df_movies_only[['title', 'country', 'gen
netflix_movies_col_subset.head()
```

Out[52]:

	title	country	genre	release_year	duration
3	9	United States	Action	2009	80
4	21	United States	Dramas	2008	123
7	187	United States	Dramas	1997	119
10	1922	United States	Dramas	2017	103
14	3022	United States	Independent Movies	2019	91

Task 2

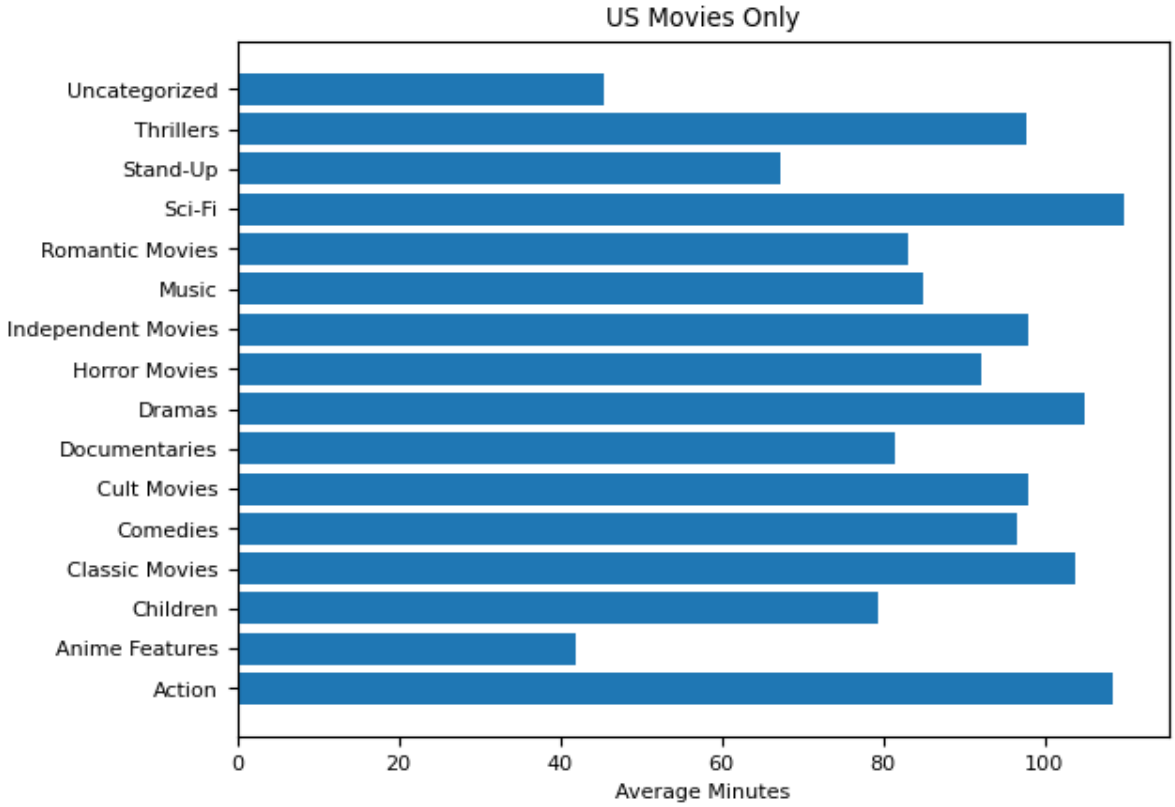
In [53]:
```
long_genre = netflix_movies_col_subset.groupby('genre').mean()
long_genre
```

Out[53]:

genre	release_year	duration
Action	2008.922449	108.428571
Anime Features	2019.000000	42.000000
Children	2011.917241	79.441379
Classic Movies	1968.404762	103.880952
Comedies	2012.445122	96.576220
Cult Movies	1990.111111	97.888889
Documentaries	2016.128463	81.372796
Dramas	2012.984085	104.965517
Horror Movies	2014.414414	92.117117
Independent Movies	2016.000000	98.000000
Music	2016.600000	85.000000
Romantic Movies	2017.500000	83.000000
Sci-Fi	2011.833333	109.833333
Stand-Up	2014.449275	67.256039
Thrillers	2013.300000	97.775000
Uncategorized	2012.818182	45.318182

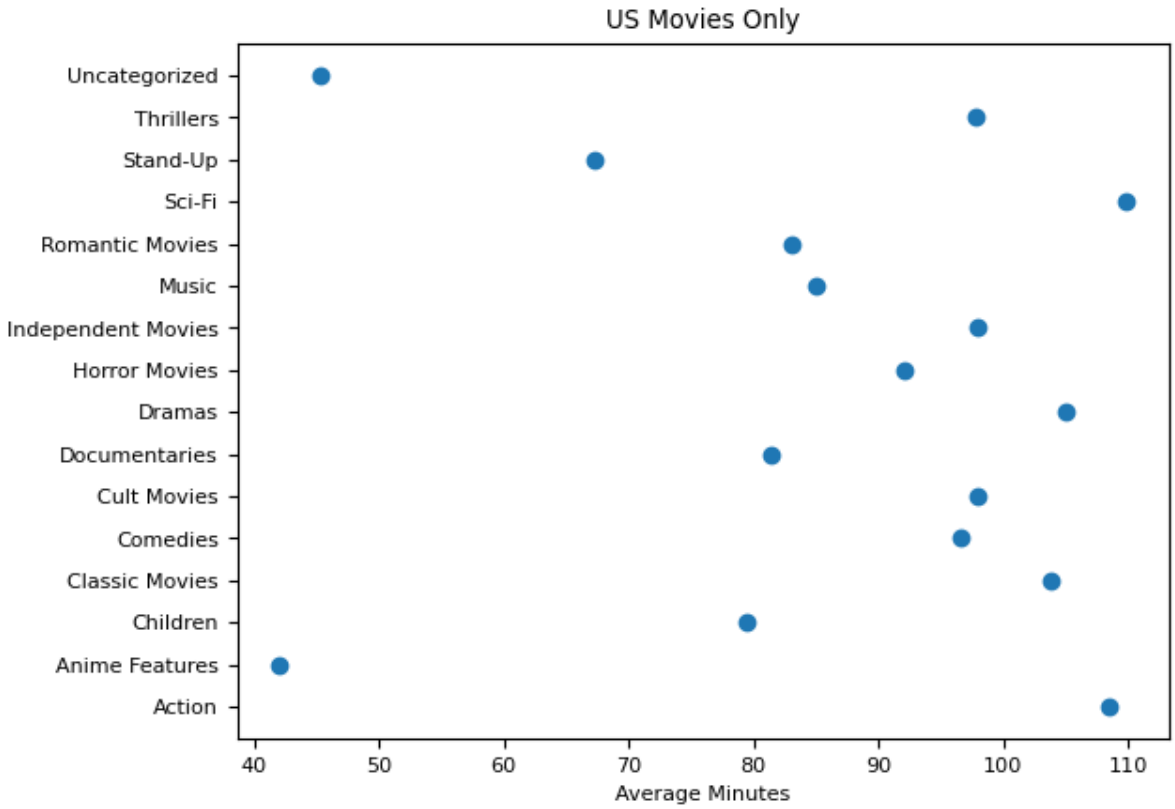
Task 3

In [54]:
```
import matplotlib.pyplot as plt
%matplotlib inline
plt.barh(long_genre.index, long_genre['duration'])
plt.title('US Movies Only')
plt.xlabel('Average Minutes')
plt.show()
```



Task 4

In [55]:
```
plt.scatter( long_genre['duration'] ,long_genre.index)
plt.title('US Movies Only')
plt.xlabel('Average Minutes')
plt.show()
```



Task 5

In [56]:
```
import pandas as pd
import numpy as np
netflix_df = pd.read_csv('netflix_data.csv')
netflix_subset = netflix_df.query('type != "Unknown" and 2016 <= release_yea
netflix_filtered = netflix_subset[['title', 'country', 'genre', 'release_yea
netflix_filtered
```

Out[56]:

	title	country	genre	release_year	duration
0	3%	Brazil	International TV	2020	4
1	7:19	Mexico	Dramas	2016	93
5	46	Turkey	International TV	2016	1
6	122	Egypt	Horror Movies	2019	95
8	706	India	Horror Movies	2019	118
...	...	...	...	...	...
7779	Zona Rosa	Mexico	International TV	2019	1
7780	Zoo	India	Dramas	2018	94
7784	Zulu Man in Japan	NaN	Documentaries	2019	44
7785	Zumbo's Just Desserts	Australia	International TV	2019	1
7786	ZZ TOP: THAT LITTLE OL' BAND FROM TEXAS	United Kingdom	Documentaries	2019	90

4879 rows × 5 columns

Task 6

In [57]:
```
country_counts = netflix_filtered.groupby('country').count()
country_counts
```

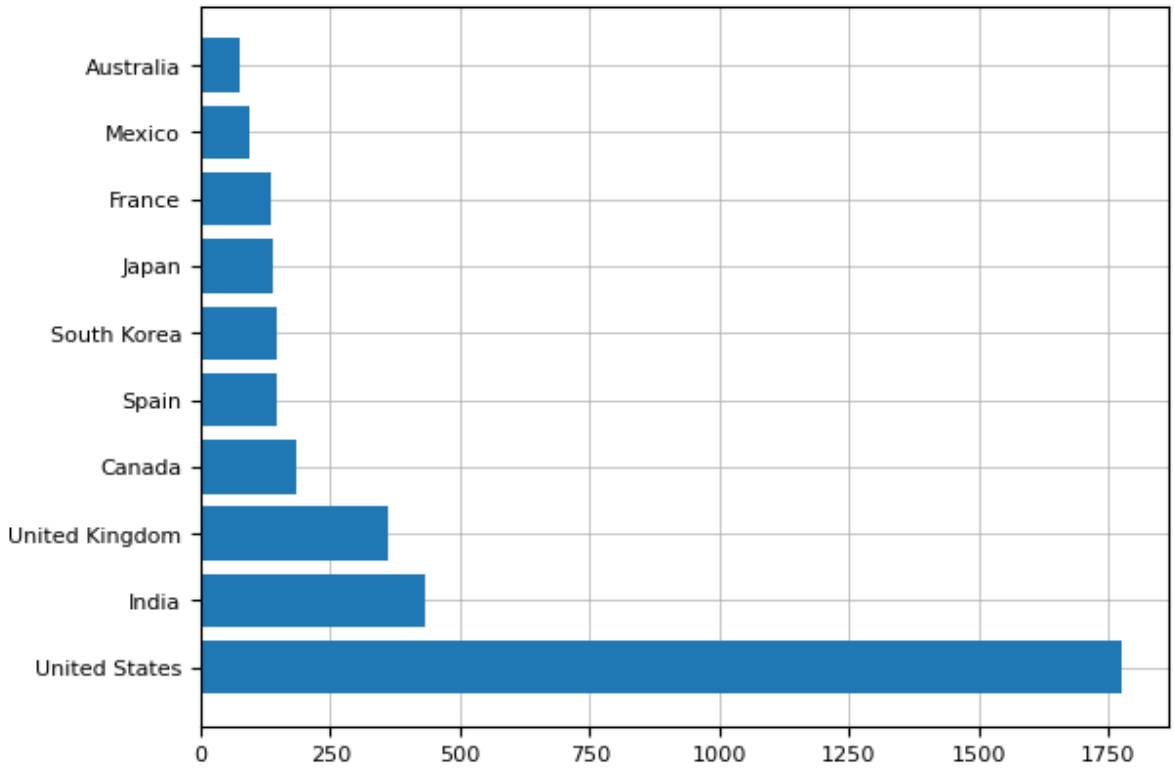
Out[57]:

country	title	genre	release_year	duration
Argentina	55	55	55	55
Australia	74	74	74	74
Austria	7	7	7	7
Bangladesh	2	2	2	2
Belarus	1	1	1	1
...	...	...	...	...
United States	1776	1776	1776	1776
Uruguay	8	8	8	8
Venezuela	1	1	1	1
Vietnam	3	3	3	3
Zimbabwe	1	1	1	1

78 rows × 4 columns

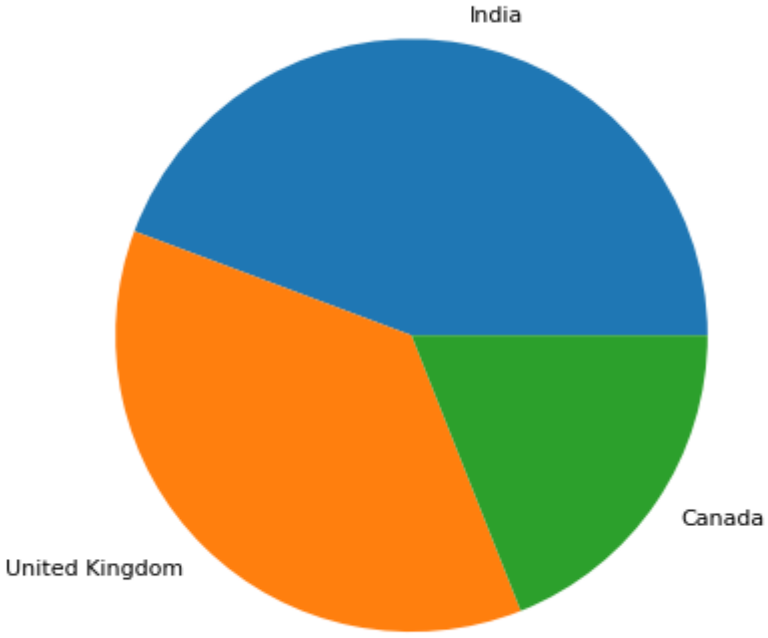
Task 7

In [42]:
```
sorted_countries = country_counts.sort_values('title', ascending=False).head

plt.barh(sorted_countries.index, sorted_countries['title'] , zorder=2)
plt.grid(True,linewidth=0.5, zorder=1)
plt.show()
```



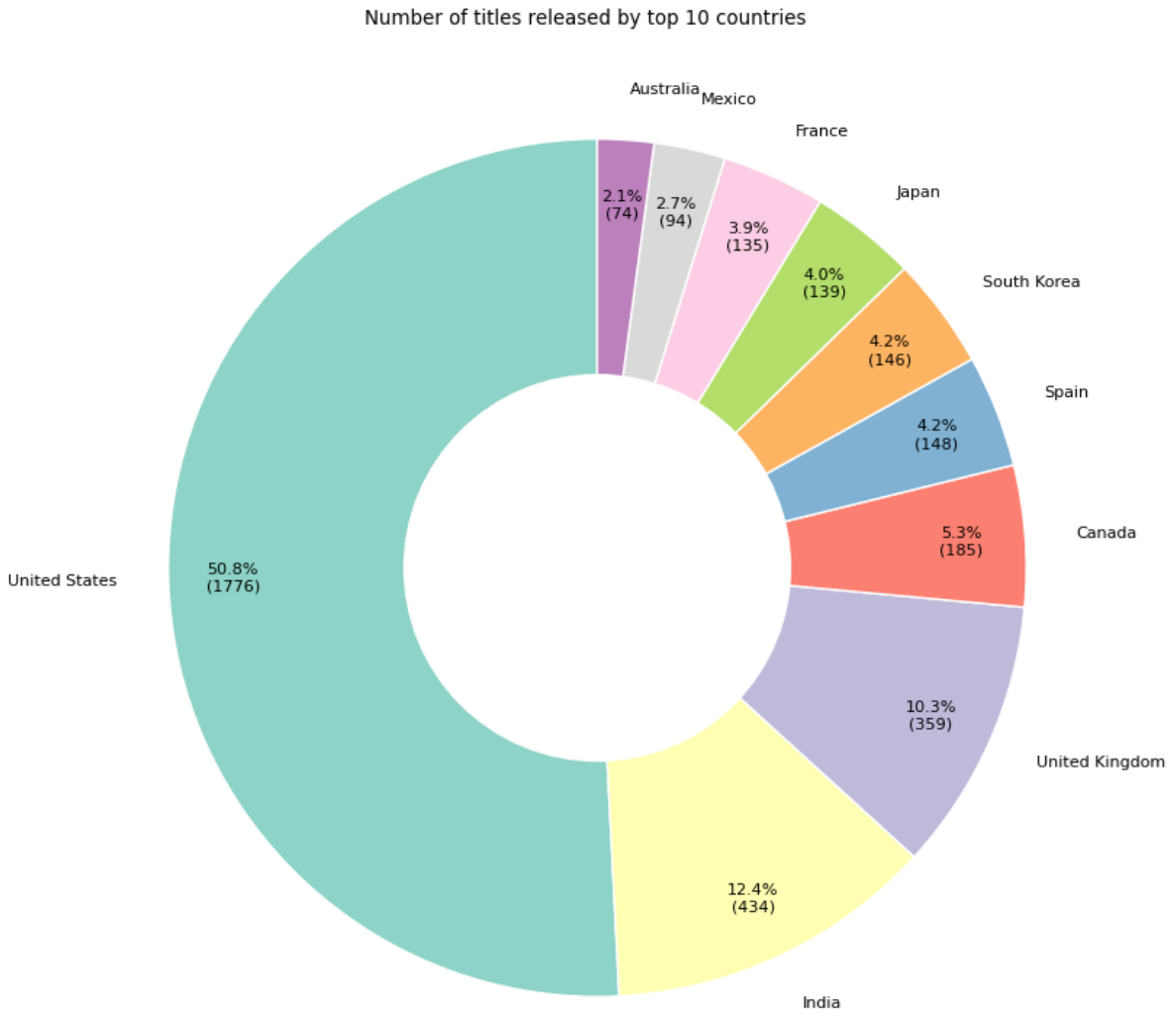
Task 8

In [43]:
```
selected_countries = sorted_countries.iloc[1:4]
plt.pie(selected_countries['title'], labels=selected_countries.index)
plt.show()
```



Task 9

In [51]:
```
import numpy as np
import matplotlib.pyplot as plt

my_circle = plt.Circle((0, 0), 0.9, color='white')

inner_labels = sorted_countries.index
inner_counts = sorted_countries['title']

inner_colors = plt.cm.Set3(np.arange(len(inner_labels)))

plt.rcParams['font.size'] = 8

def autopct_format(pct):
    count = int(pct/100 * np.sum(inner_counts))
    idx = np.where(sorted_countries.index == inner_labels)[0][0]
    label = '{:.1f}%\n({:d})'.format(pct, count)
    return label

plt.pie(inner_counts, labels=inner_labels, autopct=autopct_format,
        colors=inner_colors, wedgeprops={'edgecolor': 'white'}, radius=2, s

p = plt.gcf()
p.gca().add_artist(my_circle)
plt.suptitle("Number of titles released by top 10 countries", y=1.3)

plt.show()
```


Number of titles released by top 10 countries
Australia
Mexico
France
Japan
South Korea
Spain
Canada
United Kingdom
India
United States
2.1%
(74)
2.7%
(94)
3.9%
(135)
4.0%
(139)
4.2%
(146)
4.2%
(148)
5.3%
(185)
10.3%
(359)
12.4%
(434)
50.8%
(1776)

In [ ]: