In [2]:
```python
from pymongo import MongoClient
import matplotlib.pyplot as plt
import pandas as pd
client = MongoClient('localhost', 27017)
db = client.hotels
#reviews_collection= db.ta_review_details_hotels
users_collection= db.tripadvisor_users
```

In [3]:
```python
def memsince_date_trim():
    memsince = users_collection.aggregate(
    [
        {
            "$project" : {
                "_id" : "$_id",
                "name": "$user_name_list",
                "memdate": {"$split": ["$member_since_list", " "]}
            },
        },
        { "$unwind" : "$memdate" },
        {
            "$group" : {
                "_id" : "$memdate",
                "count" : {
                    "$sum" : 1.0
                }
            }
        },
        { "$sort" : { "_id" : -1 } }
    ],
    );
    return memsince
```

In [4]:
```python
memsince_dates = pd.DataFrame(list(memsince_date_trim()))[4:]
print(memsince_dates)
```

```
      _id   count
4    2018   605.0
5    2017  1531.0
6    2016  1601.0
7    2015  1710.0
8    2014  1590.0
9    2013  1418.0
10   2012  1499.0
11   2011   961.0
12   2010   976.0
13   2009   922.0
14   2008   856.0
15   2007   627.0
16   2006   295.0
17   2005   237.0
18   2004    66.0
19   2003    41.0
20   2002     3.0
```
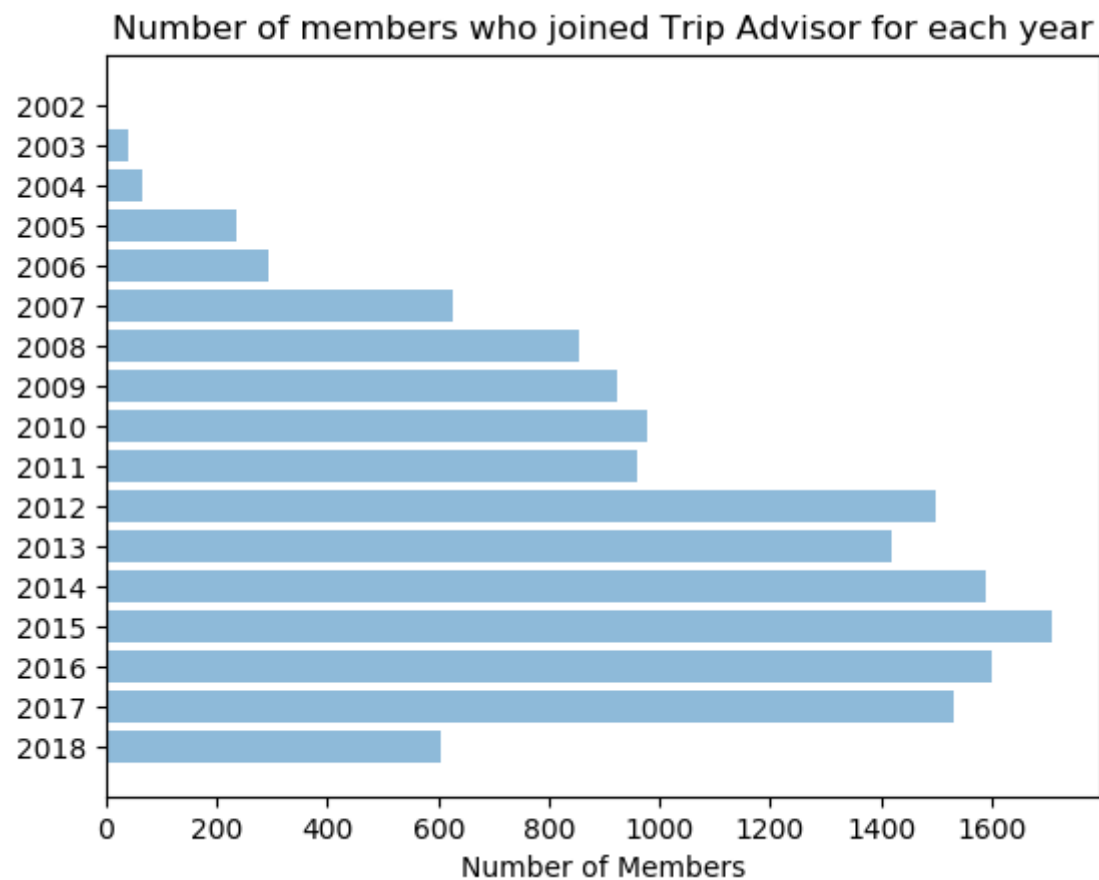
In [5]:
```python
import matplotlib.pyplot as plt; plt.rcdefaults()
import numpy as np
import matplotlib.pyplot as plt

members = memsince_dates['_id']
y_pos = np.arange(len(members))
counts = memsince_dates['count']

plt.barh(y_pos, counts, align='center', alpha=0.5)
plt.yticks(y_pos, members)
plt.xlabel('Number of Members')
plt.title('Number of members who joined Trip Advisor for each year')

plt.show()
```

Number of members who joined Trip Advisor for each year

In [1]:
```python
from pymongo import MongoClient
import matplotlib.pyplot as plt
import pandas as pd

client = MongoClient('localhost', 27017)
db = client.hotels
reviews_collection= db.tripadvisor_reviews
#users_collection= db.ta_user_details
```

In [2]:
```python
#df = pd.DataFrame(list(reviews_collection.find()))
#print(df.head())
#df = pd.DataFrame(list(users_collection.find()))
```

Number of reviews for each hotel from TripAdvisor

In [3]:
```python
def review_count_for_each_hotel():
    reviews = reviews_collection.aggregate(
    [
        {
            "$group" : {
                "_id" : "$Hotel_name",
                "count" : {
                    "$sum" : 1.0
                }
            }
        },
        {
            "$sort" : {
                "count" : -1
            }
        }
    ],
    );
    return reviews
```

In [4]:
```python
df = pd.DataFrame(list(review_count_for_each_hotel()))
print(df)
```
```
                        _id    count
0  Crowne Plaza Changi Airport  3060.0
1                  Shangri La  3005.0
2                  Park Hotel  3005.0
3                 Oasia Hotel  3005.0
4               Village Hotel  3005.0
5            Marina Bay Sands  3005.0
```

In [5]:
```python
fig1, ax1 = plt.subplots()
ax1.pie(df['count'], labels=df['_id'], autopct='%1.1f%%',
        shadow=True, startangle=90)
ax1.axis('equal')
plt.show()
```
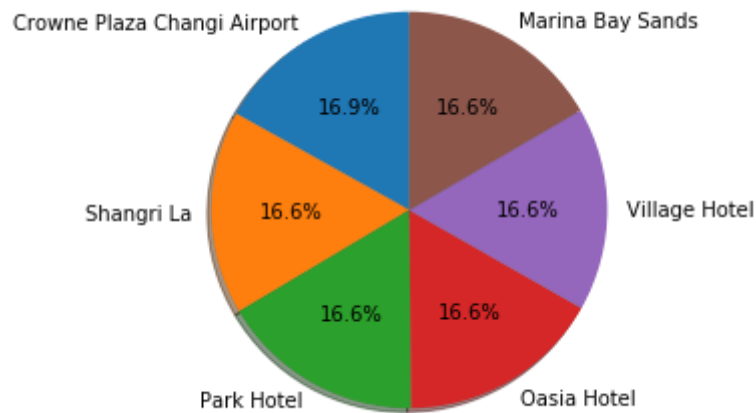

Number of reviews per month for each Hotel

```python
In [6]: def review_date_trim():
            reviews_date_trim = reviews_collection.aggregate(
            [
                {
                    "$project" : {
                        "_id" : "$_id",
                        "Hotel_name": "$Hotel_name",
                        "date": {"$split": ["$review_date_list", " "]}
                    },
                },
                { "$unwind" : "$date" },
                {
                    "$group" : {
                        "_id" : "$date",
                        "count" : {
                            "$sum" : 1.0
                        }
                    }
                },
                { "$sort" : { "_id" : -1 } }
            ],
            );
            return reviews_date_trim
```

In [7]:
```python
df_review_dates = pd.DataFrame(list(review_date_trim()))[0:12].sort_values(['count'], ascending=False)
print(df_review_dates)
#Top 3 months to visit Singapore are July, April and May
```
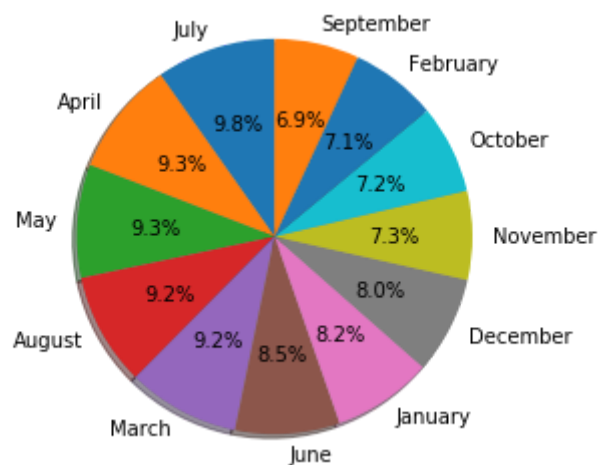
```
          _id    count
6        July   1767.0
11      April   1689.0
3         May   1675.0
10     August   1669.0
4       March   1661.0
5        June   1546.0
7     January   1478.0
9    December   1439.0
2    November   1317.0
1     October   1311.0
8    February   1284.0
0   September   1249.0
```

In [8]:
```python
fig2, ax2 = plt.subplots()
ax2.pie(df_review_dates['count'], labels=df_review_dates['_id'], autopct='%1.1f%%',
        shadow=True, startangle=90)
ax2.axis('equal')
plt.show()
```


Review Mode Counts

Review Mode Counts

```
In [9]:  def review_mode_counts():
             reviews = reviews_collection.aggregate(
             [
                 {
                     "$group" : {
                         "_id" : "$review_mode_list",
                         "numModes" : {
                             "$sum" : "1"
                         }
                     }
                 },
             ],
             );
             return reviews
```