

```
In [1]: import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
```

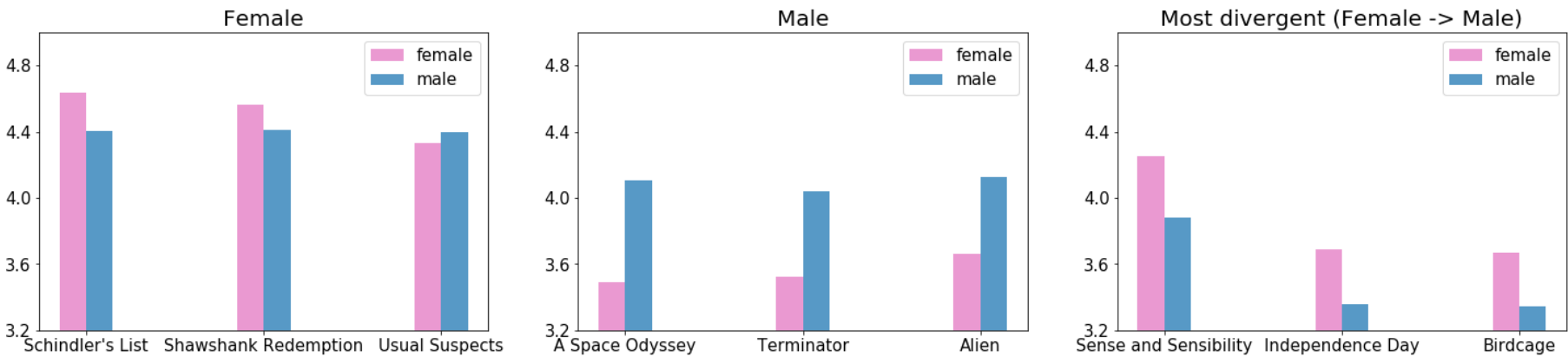
```
In [39]: #男女最受欢迎的电影 和男女最受争议的电影
fontSize = 15

plt.figure(figsize=(25,5))

# 100k
# fig 1
plt.subplot(1,3,1)
plt.title("Female",fontSize=20)
name_list = ['Schindler\'s List','Shawshank Redemption','Usual Suspects']
num_list = [4.633,4.563,4.333]
num_list1 = [4.406,4.411,4.400]
x =list(range(len(num_list)))
total_width, n = 0.3, 2
width = total_width / n
x_adjust_factor = -0.075
plt.ylim(3.2,5)
plt.yticks([3.2,3.6,4.0,4.4,4.8], fontSize=fontSize)
plt.xticks(x, name_list, fontSize=fontSize)
plt.bar(np.array(x)+x_adjust_factor, num_list, width=width, label='female',color='tab:pink',alpha=0.75)
for i in range(len(x)):
    x[i] = x[i] + width
plt.bar(np.array(x)+x_adjust_factor, num_list1, width=width, label='male',color='tab:blue',alpha=0.75)
plt.legend(fontSize=fontSize)
# plt.grid()
# fig 2
plt.subplot(1,3,2)
plt.title("Male",fontSize=20)
name_list = ['A Space Odyssey','Terminator','Alien']
num_list = [3.491228,3.524590,3.660714]
num_list1 = [4.103960,4.037500,4.123404]
x =list(range(len(num_list)))
total_width, n = 0.3, 2
width = total_width / n
x_adjust_factor = -0.075
plt.ylim(3.2,5)
plt.yticks([3.2,3.6,4.0,4.4,4.8], fontSize=fontSize)
plt.xticks(x, name_list, fontSize=fontSize)
plt.bar(np.array(x)+x_adjust_factor, num_list, width=width, label='female',color='tab:pink',alpha=0.75)
for i in range(len(x)):
    x[i] = x[i] + width
plt.bar(np.array(x)+x_adjust_factor, num_list1, width=width, label='male',color='tab:blue',alpha=0.75)
plt.legend(fontSize=fontSize)
# plt.grid()

# Fig 5
plt.subplot(1,3,3)
plt.title("Most divergent (Female -> Male)",fontSize=20)
name_list = ['Sense and Sensibility','Independence Day','Birdcage']
num_list = [4.252632,3.688679 ,3.670455]
num_list1 = [3.878613,3.356037,3.346341]
x = list(range(len(num_list)))
total_width, n = 0.3, 2
width = total_width / n
x_adjust_factor = -0.075
plt.ylim(3.2,5)
plt.yticks([3.2,3.6,4.0,4.4,4.8], fontSize=fontSize)
plt.xticks(x, name_list, fontSize=fontSize)
plt.bar(np.array(x)+x_adjust_factor, num_list, width=width, label='female',color='tab:pink',alpha=0.75)
for i in range(len(x)):
    x[i] = x[i] + width
plt.bar(np.array(x)+x_adjust_factor, num_list1, width=width, label='male',color='tab:blue',alpha=0.75)
plt.legend(fontSize=fontSize)

plt.savefig("gender_diff_100k.pdf")
```



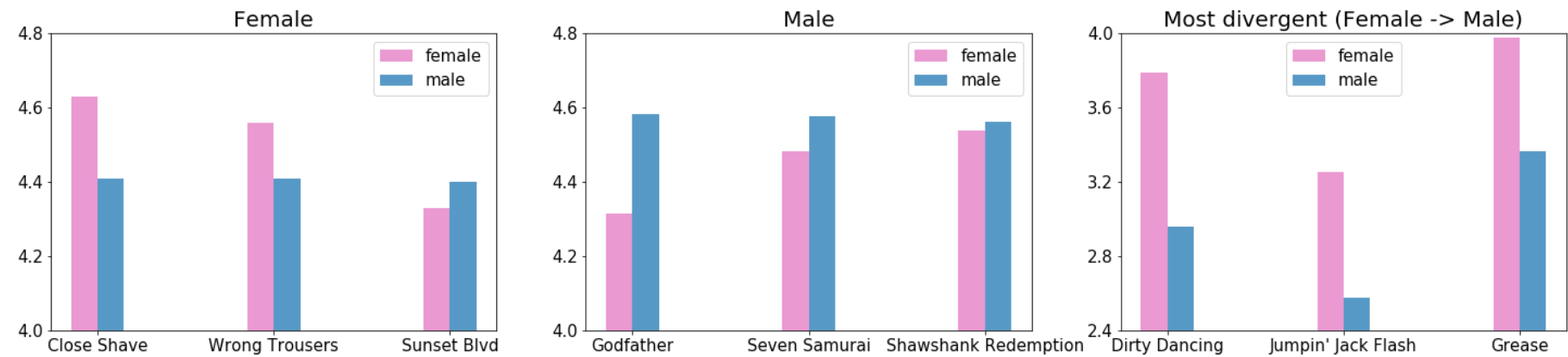
```
In [40]: # 1m

fontsize = 15

plt.figure(figsize=(25,5))
# fig 3
plt.subplot(1,3,1)
plt.title("Female",fontsize=20)
name_list = ['Close Shave','Wrong Trousers','Sunset Blvd']
num_list = [4.63,4.56,4.33]
num_list1 = [4.41,4.41,4.40]
x = list(range(len(num_list)))
total_width, n = 0.3, 2
width = total_width / n
x_adjust_factor = -0.075
plt.ylim(4,4.8)
plt.yticks([4, 4.2, 4.4, 4.6, 4.8], fontsize=fontsize)
plt.xticks(x, name_list, fontsize=fontsize)
plt.bar(np.array(x)+x_adjust_factor, num_list, width=width, label='female',color='tab:pink',alpha=0.75)
for i in range(len(x)):
    x[i] = x[i] + width
plt.bar(np.array(x)+x_adjust_factor, num_list1, width=width, label='male',color='tab:blue',alpha=0.75)
plt.legend(fontsize=fontsize)

# Fig 4
plt.subplot(1,3,2)
plt.title("Male",fontsize=20)
name_list = ['Godfather','Seven Samurai','Shawshank Redemption']
num_list = [4.314700,4.481132 ,4.539075]
num_list1 = [4.583333, 4.576628,4.560625]
x = list(range(len(num_list)))
total_width, n = 0.3, 2
width = total_width / n
x_adjust_factor = -0.075
plt.ylim(4,4.8)
plt.yticks([4, 4.2, 4.4, 4.6, 4.8], fontsize=fontsize)
plt.xticks(x, name_list, fontsize=fontsize)
plt.bar(np.array(x)+x_adjust_factor, num_list, width=width, label='female',color='tab:pink',alpha=0.75)
for i in range(len(x)):
    x[i] = x[i] + width
plt.bar(np.array(x)+x_adjust_factor, num_list1, width=width, label='male',color='tab:blue',alpha=0.75)
plt.legend(fontsize=fontsize)

# Fig 6
plt.subplot(1,3,3)
plt.title("Most divergent (Female -> Male)",fontsize=20)
name_list = ['Dirty Dancing',"Jumpin' Jack Flash",'Grease']
num_list = [3.790378,3.254717,3.975265]
num_list1 = [2.959596,2.578358,3.367041]
x = list(range(len(num_list)))
total_width, n = 0.3, 2
width = total_width / n
x_adjust_factor = -0.075
plt.ylim(2.4,4.0)
plt.yticks([2.4,2.8,3.2,3.6,4.0], fontsize=fontsize)
plt.xticks(x, name_list, fontsize=fontsize)
plt.bar(np.array(x)+x_adjust_factor, num_list, width=width, label='female',color='tab:pink',alpha=0.75)
for i in range(len(x)):
    x[i] = x[i] + width
plt.bar(np.array(x)+x_adjust_factor, num_list1, width=width, label='male',color='tab:blue',alpha=0.75)
plt.legend(fontsize=fontsize)
plt.savefig("gender_diff_1M.pdf")
plt.show()
```



```

In [35]: # 100k
         fontsize = 15

         # plt.figure(figsize=(20,18))
         plt.figure(figsize=(25,10))

         # fig 1
         plt.subplot(2,3,1)
         plt.title("Age0-19",fontsize=20)
         name_list = ['Shawshank Redemption','Titanic','Empire Strikes Back']
         num_list = [4.666667,4.621622,4.620690]
         num_list1 = [4.430189 ,4.201278,4.168639 ]
         x =list(range(len(num_list)))
         total_width, n = 0.3, 2
         width = total_width / n
         x_adjust_factor = -0.075
         plt.ylim(4,4.8)
         plt.yticks([4, 4.2, 4.4, 4.6, 4.8], fontsize=fontsize)
         plt.xticks(x, name_list, fontsize=fontsize)
         plt.bar(np.array(x)+x_adjust_factor, num_list, width=width, label='Age0-19',color='tab:pink',alpha=0.75)
         for i in range(len(x)):
             x[i] = x[i] + width
         plt.bar(np.array(x)+x_adjust_factor, num_list1, width=width, label='Other age',color='tab:blue',alpha=0.75)
         plt.legend(fontsize=fontsize)

         # plt.grid()
         # fig 2
         plt.subplot(2,3,2)
         plt.title("Age20-29",fontsize=20)
         name_list = ['Shawshank Redemption',"Schindler's List",'Silence of the Lambs']
         num_list = [4.585366,4.467890,4.434211]
         num_list1 = [4.337500,4.465608,4.197479]
         x =list(range(len(num_list)))
         total_width, n = 0.3, 2
         width = total_width / n
         x_adjust_factor = -0.075
         plt.ylim(4,4.8)
         plt.yticks([4, 4.2, 4.4, 4.6, 4.8], fontsize=fontsize)
         plt.xticks(x, name_list, fontsize=fontsize)
         plt.bar(np.array(x)+x_adjust_factor, num_list, width=width, label='Age20-29',color='tab:pink',alpha=0.75)
         for i in range(len(x)):
             x[i] = x[i] + width
         plt.bar(np.array(x)+x_adjust_factor, num_list1, width=width, label='Other age',color='tab:blue',alpha=0.75)
         plt.legend(fontsize=fontsize)

         # plt.grid()

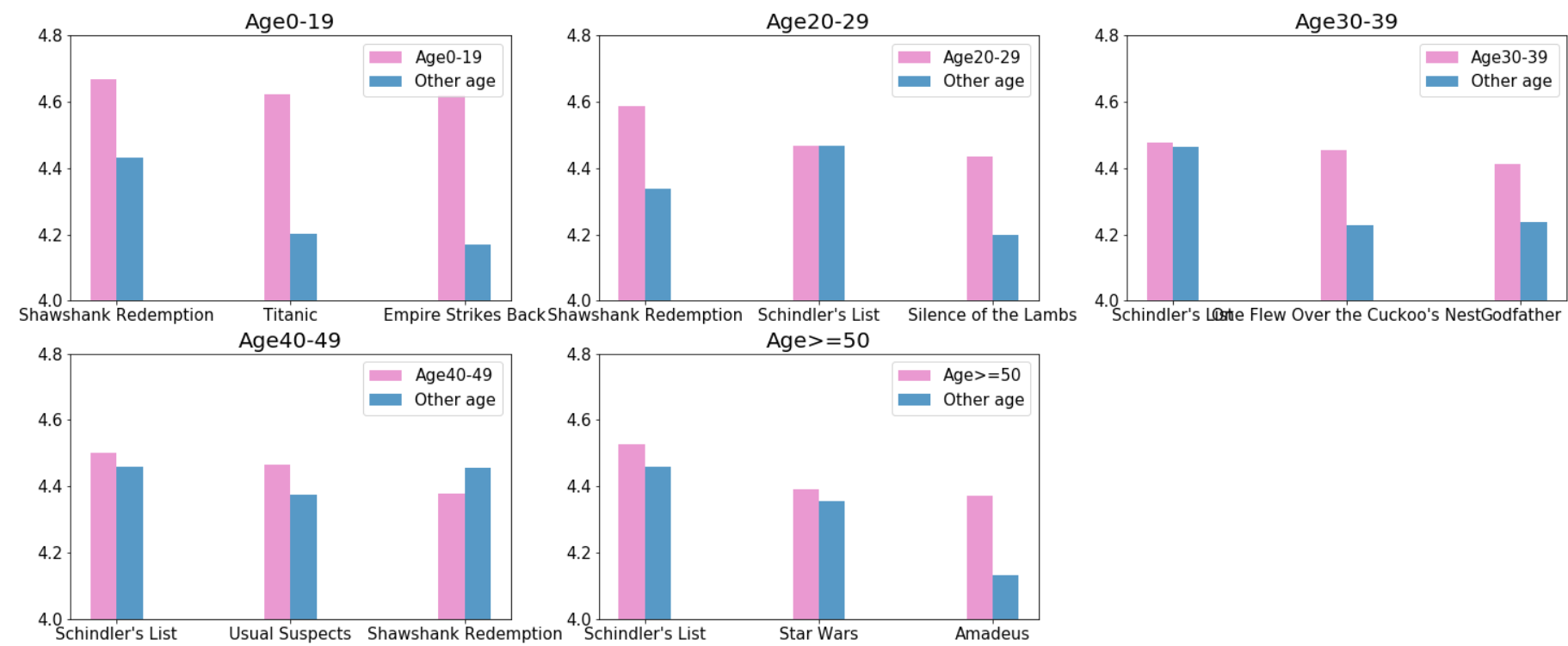
         # fig 3
         plt.subplot(2,3,3)
         plt.title("Age30-39",fontsize=20)
         name_list = ["Schindler's List","One Flew Over the Cuckoo's Nest" ,'Godfather']
         num_list = [4.475610,4.453333,4.412844]
         num_list1 = [4.462963,4.227513,4.236842]
         x = list(range(len(num_list)))
         total_width, n = 0.3, 2
         width = total_width / n
         x_adjust_factor = -0.075
         plt.ylim(4,4.8)
         plt.yticks([4, 4.2, 4.4, 4.6, 4.8], fontsize=fontsize)
         plt.xticks(x, name_list, fontsize=fontsize)
         plt.bar(np.array(x)+x_adjust_factor, num_list, width=width, label='Age30-39',color='tab:pink',alpha=0.75)
         for i in range(len(x)):
             x[i] = x[i] + width
         plt.bar(np.array(x)+x_adjust_factor, num_list1, width=width, label='Other age',color='tab:blue',alpha=0.75)
         plt.legend(fontsize=fontsize)

         # Fig 4
         plt.subplot(2,3,4)
         plt.title("Age40-49",fontsize=20)
         name_list = ["Schindler's List",'Usual Suspects','Shawshank Redemption']
         num_list = [4.500000,4.466667,4.378378]
         num_list1 = [4.460000, 4.375527,4.455285 ]
         x = list(range(len(num_list)))
         total_width, n = 0.3, 2
         width = total_width / n
         x_adjust_factor = -0.075
         plt.ylim(4,4.8)
         plt.yticks([4, 4.2, 4.4, 4.6, 4.8], fontsize=fontsize)
         plt.xticks(x, name_list, fontsize=fontsize)
         plt.bar(np.array(x)+x_adjust_factor, num_list, width=width, label='Age40-49',color='tab:pink',alpha=0.75)
         for i in range(len(x)):
             x[i] = x[i] + width
         plt.bar(np.array(x)+x_adjust_factor, num_list1, width=width, label='Other age',color='tab:blue',alpha=0.75)
         plt.legend(fontsize=fontsize)

         # Fig 5
         plt.subplot(2,3,5)
         plt.title("Age>=50",fontsize=20)

```

```
name_list = ["Schindler's List", 'Star Wars', 'Amadeus']
num_list = [4.527778,4.389831,4.371429]
num_list1 = [4.458015,4.354962,4.132780]
x = list(range(len(num_list)))
total_width, n = 0.3, 2
width = total_width / n
x_adjust_factor = -0.075
plt.ylim(4,4.8)
plt.yticks([4, 4.2, 4.4, 4.6, 4.8], fontsize=fontsize)
plt.xticks(x, name_list, fontsize=fontsize)
plt.bar(np.array(x)+x_adjust_factor, num_list, width=width, label='Age>=50',color='tab:blue',alpha=0.75)
for i in range(len(x)):
    x[i] = x[i] + width
plt.bar(np.array(x)+x_adjust_factor, num_list1, width=width, label='Other age',color='tab:red',alpha=0.75)
plt.legend(fontsize=fontsize)
# plt.grid()
plt.savefig("age_diff_100k.pdf")
plt.show()
```



```

In [38]: # 1m
# fig 6
fontsize = 15

plt.figure(figsize=(32,10))

plt.subplot(2,4,1)
plt.title("Age1",fontsize=20)
name_list = ['Shadowlands','Somewhere in Time','Night Shift']
num_list = [5,5,5]
num_list1 = [3.827465,3.570922,3.467577]
x =list(range(len(num_list)))
total_width, n = 0.3, 2
width = total_width / n
x_adjust_factor = -0.075
plt.ylim(3.2,5.2)
plt.yticks([3.2,3.6, 4.0, 4.4, 4.8], fontsize=fontsize)
plt.xticks(x, name_list, fontsize=fontsize)
plt.bar(np.array(x)+x_adjust_factor, num_list, width=width, label='Age0-19',color='tab:pink',alpha=0.75)
for i in range(len(x)):
    x[i] = x[i] + width
plt.bar(np.array(x)+x_adjust_factor, num_list1, width=width, label='Other age',color='tab:blue',alpha=0.75)
plt.legend(fontsize=fontsize)

# plt.grid()

# fig 7
plt.subplot(2,4,2)
plt.title("Age18",fontsize=20)
name_list = ["Usual Suspects","Shawshank Redemption" , 'Seven Samurai']
num_list = [4.680798,4.674569,4.604938]
num_list1 = [4.469609,4.522972,4.553931]
x = list(range(len(num_list)))
total_width, n = 0.3, 2
width = total_width / n
x_adjust_factor = -0.075
plt.ylim(3.6,5.2)
plt.yticks([3.6, 4.0, 4.4, 4.8], fontsize=fontsize)
plt.xticks(x, name_list, fontsize=fontsize)
plt.bar(np.array(x)+x_adjust_factor, num_list, width=width, label='Age18',color='tab:pink',alpha=0.75)
for i in range(len(x)):
    x[i] = x[i] + width
plt.bar(np.array(x)+x_adjust_factor, num_list1, width=width, label='Other age',color='tab:blue',alpha=0.75)
plt.legend(fontsize=fontsize)

# Fig 8
plt.subplot(2,4,3)
plt.title("Age25",fontsize=20)
name_list = ["Shawshank Redemption",'Raiders of the Lost Ark','Godfather']
num_list = [4.587699,4.578521,4.574051]
num_list1 = [4.532987,4.412574,4.496444]
x = list(range(len(num_list)))
total_width, n = 0.3, 2
width = total_width / n
x_adjust_factor = -0.075
plt.ylim(3.6,5.2)
plt.yticks([3.6, 4.0, 4.4, 4.8], fontsize=fontsize)
plt.xticks(x, name_list, fontsize=fontsize)
plt.bar(np.array(x)+x_adjust_factor, num_list, width=width, label='Age25',color='tab:pink',alpha=0.75)
for i in range(len(x)):
    x[i] = x[i] + width
plt.bar(np.array(x)+x_adjust_factor, num_list1, width=width, label='Other age',color='tab:blue',alpha=0.75)
plt.legend(fontsize=fontsize)

# Fig 9
plt.subplot(2,4,4)
plt.title("Age35",fontsize=20)
name_list = ["Sunset Blvd",'To Kill a Mockingbird','City Lights']
num_list = [4.611570,4.585859,4.581818]
num_list1 = [4.449857,4.382192,4.337963]
x = list(range(len(num_list)))
total_width, n = 0.3, 2
width = total_width / n
x_adjust_factor = -0.075
plt.ylim(3.6,5.2)
plt.yticks([3.6, 4.0, 4.4, 4.8], fontsize=fontsize)
plt.xticks(x, name_list, fontsize=fontsize)
plt.bar(np.array(x)+x_adjust_factor, num_list, width=width, label='Age35',color='tab:pink',alpha=0.75)
for i in range(len(x)):
    x[i] = x[i] + width
plt.bar(np.array(x)+x_adjust_factor, num_list1, width=width, label='Other age',color='tab:blue',alpha=0.75)
plt.legend(fontsize=fontsize)
# plt.grid()

# Fig 10
plt.subplot(2,4,5)

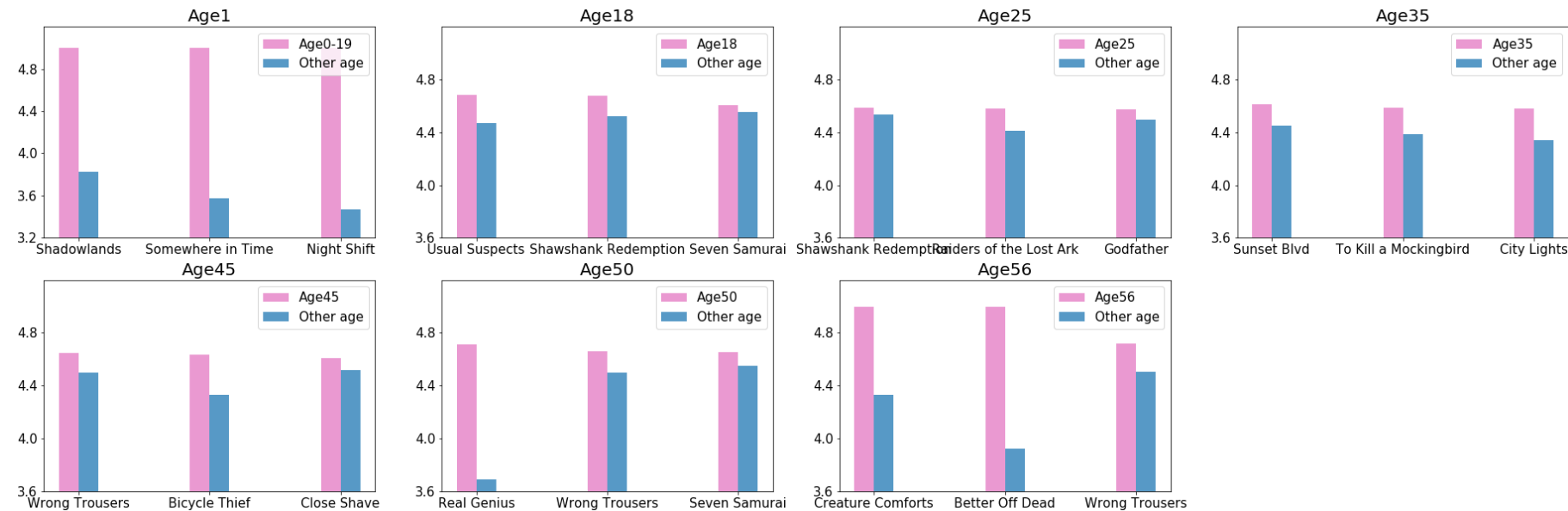
```



```
plt.title("Age45",fontsize=20)
name_list = ["Wrong Trousers",'Bicycle Thief','Close Shave']
num_list = [4.647059,4.636364,4.607143]
num_list1 = [4.499398,4.330435,4.516693]
x = list(range(len(num_list)))
total_width, n = 0.3, 2
width = total_width / n
x_adjust_factor = -0.075
plt.ylim(3.6,5.2)
plt.yticks([3.6, 4.0, 4.4, 4.8], fontsize=fontsize)
plt.xticks(x, name_list, fontsize=fontsize)
plt.bar(np.array(x)+x_adjust_factor, num_list, width=width, label='Age45',color='tab:pink',alpha=0.75)
for i in range(len(x)):
    x[i] = x[i] + width
plt.bar(np.array(x)+x_adjust_factor, num_list1, width=width, label='Other age',color='tab:blue',alpha=0.75)
plt.legend(fontsize=fontsize)

# Fig 11
plt.subplot(2,4,6)
plt.title("Age50",fontsize=20)
name_list = ["Real Genius",'Wrong Trousers','Seven Samurai']
num_list = [4.714286,4.660000,4.657143]
num_list1 = [3.691954,4.498798,4.548387]
x = list(range(len(num_list)))
total_width, n = 0.3, 2
width = total_width / n
x_adjust_factor = -0.075
plt.ylim(3.6,5.2)
plt.yticks([3.6, 4.0, 4.4, 4.8], fontsize=fontsize)
plt.xticks(x, name_list, fontsize=fontsize)
plt.bar(np.array(x)+x_adjust_factor, num_list, width=width, label='Age50',color='tab:pink',alpha=0.75)
for i in range(len(x)):
    x[i] = x[i] + width
plt.bar(np.array(x)+x_adjust_factor, num_list1, width=width, label='Other age',color='tab:blue',alpha=0.75)
plt.legend(fontsize=fontsize)

# Fig 12
plt.subplot(2,4,7)
plt.title("Age56",fontsize=20)
name_list = ["Creature Comforts",'Better Off Dead','Wrong Trousers']
num_list = [5,5,4.72]
num_list1 = [4.330882,3.919321,4.503472]
x = list(range(len(num_list)))
total_width, n = 0.3, 2
width = total_width / n
x_adjust_factor = -0.075
plt.ylim(3.6,5.2)
plt.yticks([3.6, 4.0, 4.4, 4.8], fontsize=fontsize)
plt.xticks(x, name_list, fontsize=fontsize)
plt.bar(np.array(x)+x_adjust_factor, num_list, width=width, label='Age56',color='tab:pink',alpha=0.75)
for i in range(len(x)):
    x[i] = x[i] + width
plt.bar(np.array(x)+x_adjust_factor, num_list1, width=width, label='Other age',color='tab:blue',alpha=0.75)
plt.legend(fontsize=fontsize)
plt.savefig("age_diff_1M.pdf")
plt.show()
```



In []: