

In [1]:

```
1 import numpy as np
2 import pandas as pd
3 import matplotlib.pyplot as plt
4 import squarify
5 import seaborn as sns
6 %matplotlib inline
```

In [2]:

```
1 gds = pd.read_csv('dataset.csv', encoding='utf-8')
```

In [3]:

```
1 gds.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1034 entries, 0 to 1033
Data columns (total 9 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Game                  1034 non-null   object
1   Year                  825 non-null    float64
2   Genre                 1034 non-null   object
3   Publisher             825 non-null    object
4   North America         1034 non-null   float64
5   Europe                1034 non-null   float64
6   Japan                 1034 non-null   float64
7   Rest of World         1034 non-null   float64
8   Global                1034 non-null   float64
dtypes: float64(6), object(3)
memory usage: 60.6+ KB
```

Таблица PS4 Games Sales

Датасет можно найти [тут \(https://www.kaggle.com/sidtwr/videogames-sales-dataset?select=PS4_GamesSales.csv\)](https://www.kaggle.com/sidtwr/videogames-sales-dataset?select=PS4_GamesSales.csv)

In [4]:

```
1 gds.head(10)
```

Out[4]:

	Game	Year	Genre	Publisher	North America	Europe	Japan	Rest of World	Global
0	Grand Theft Auto V	2014.0	Action	Rockstar Games	6.06	9.71	0.60	3.02	19.39
1	Call of Duty: Black Ops 3	2015.0	Shooter	Activision	6.18	6.05	0.41	2.44	15.09
2	Red Dead Redemption 2	2018.0	Action-Adventure	Rockstar Games	5.26	6.21	0.21	2.26	13.94
3	Call of Duty: WWII	2017.0	Shooter	Activision	4.67	6.21	0.40	2.12	13.40
4	FIFA 18	2017.0	Sports	EA Sports	1.27	8.64	0.15	1.73	11.80
5	FIFA 17	2016.0	Sports	Electronic Arts	1.26	7.95	0.12	1.61	10.94
6	Uncharted (PS4)	2016.0	Action	Sony Interactive Entertainment	4.49	3.93	0.21	1.70	10.33
7	Spider-Man (PS4)	2018.0	Action-Adventure	Sony Interactive Entertainment	3.64	3.39	0.32	1.41	8.76
8	Call of Duty: Infinite Warfare	2016.0	Shooter	Activision	3.11	3.83	0.19	1.36	8.48
9	Fallout 4	2015.0	Role-Playing	Bethesda Softworks	2.91	3.97	0.27	1.34	8.48

In [5]:

```
1 gds.Genre.unique()
```

Out[5]:

```
array(['Action', 'Shooter', 'Action-Adventure', 'Sports', 'Role-Playing',
      'Misc', 'Platform', 'Racing', 'Fighting', 'Adventure', 'MMO',
      'Simulation', 'Music', 'Party', 'Strategy', 'Puzzle',
      'Visual Novel'], dtype=object)
```

Гипотезы:

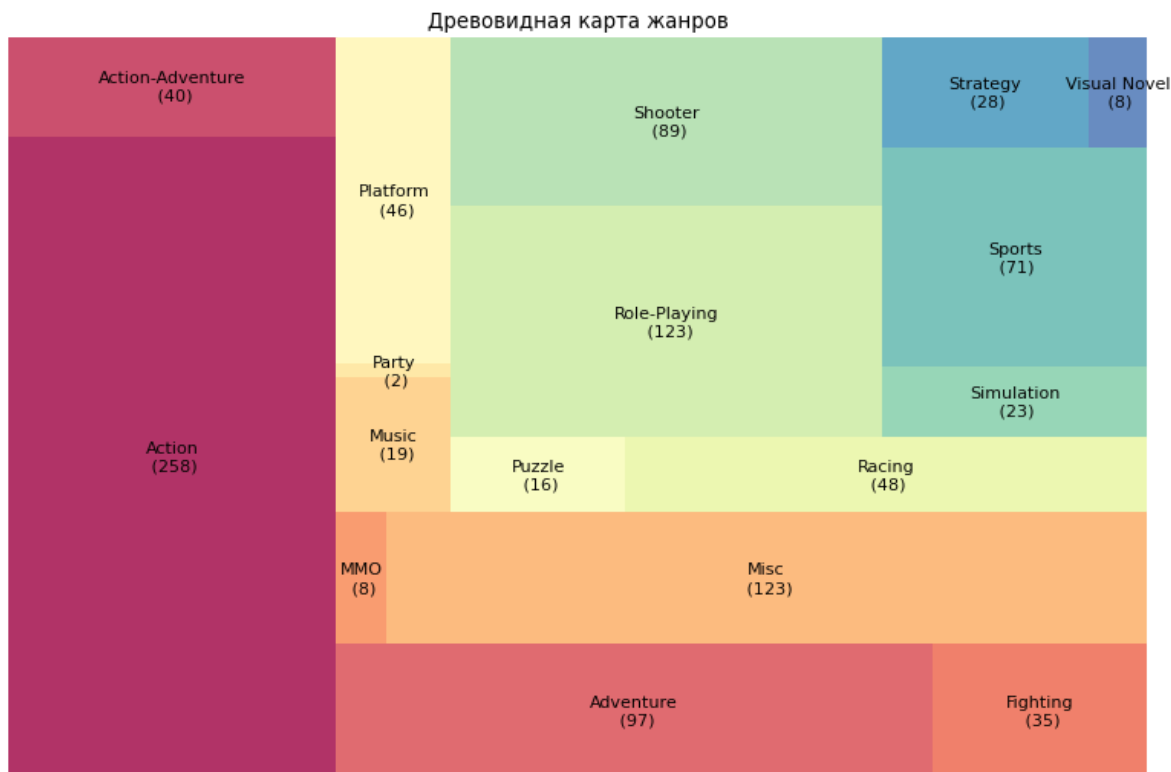
- Самый многочисленный жанр - Action
- Игры от Activision продаются лучше игр от Sony Interactive Entertainment
- Больше всего шутеров было выпущено в 2016 году

In [6]:

```

1 tree = gds.groupby('Genre').size().reset_index(name='counts')
2 labels = tree.apply(lambda x: str(x[0]) + "\n (" + str(x[1]) + ")", axis=1)
3 sizes = tree['counts'].values.tolist()
4 colors = [plt.cm.Spectral(i/float(len(labels))) for i in range(len(labels))]
5
6 plt.figure(figsize=(12,8), dpi= 80)
7 squarify.plot(sizes=sizes, label=labels, color=colors, alpha=.8)
8
9 plt.title('Древовидная карта жанров')
10 plt.axis('off')
11 plt.show()

```



Больше всего было выпущено игр в жанре Action, затем РПГ и Разное

In [7]:

```
1 gds.isnull().sum()
```

Out[7]:

```
Game          0
Year          209
Genre         0
Publisher     209
North America 0
Europe        0
Japan         0
Rest of World 0
Global        0
dtype: int64
```

У 209 игр не указаны год выпуска и издатель

In [8]:

```
1 gds.groupby('Publisher')['Global'].sum().reset_index().sort_values(by='Global', ascending=False)
```

Out[8]:

	Publisher	Global
8	Activision	72.44
142	Ubisoft	59.16
42	Electronic Arts	54.96
115	Sony Interactive Entertainment	54.85
41	EA Sports	47.55

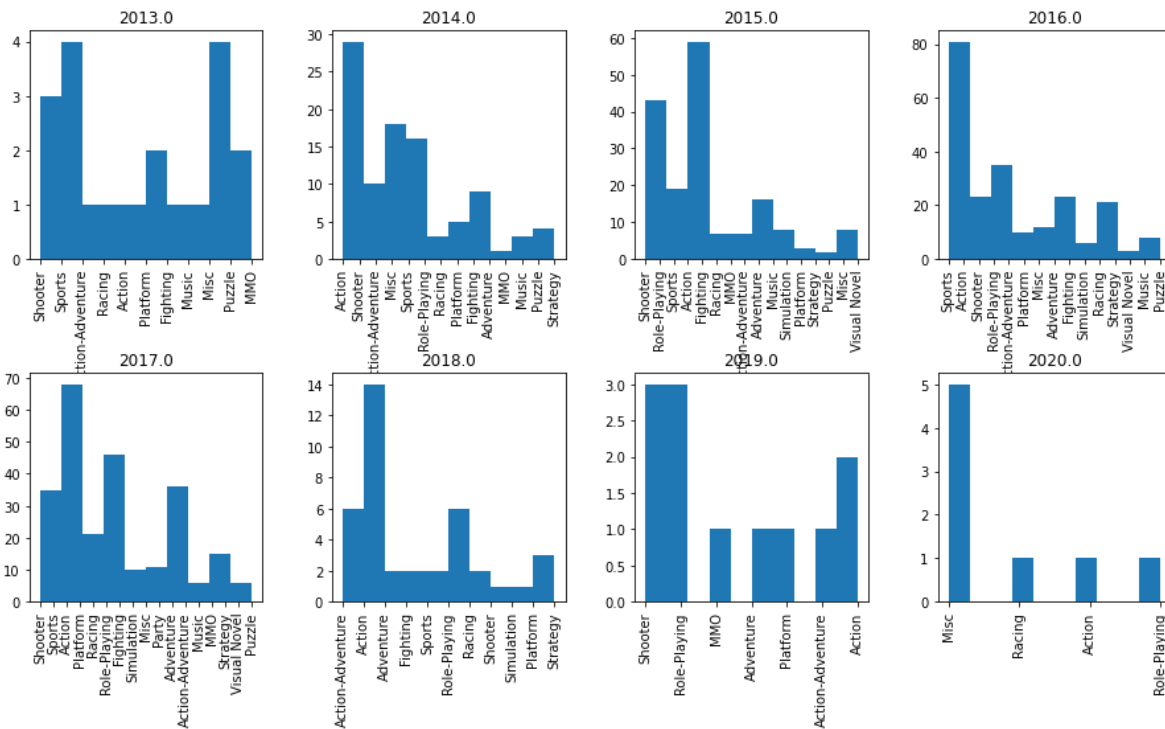
Activision продала больше всего копий своих игр

In [9]:

```
1 gds.hist('Genre', bins=10, by = ['Year'], layout=[2,4],figsize = [15, 8])
```

Out[9]:

```
array([[<matplotlib.axes._subplots.AxesSubplot object at 0x1C8435C8>,
      <matplotlib.axes._subplots.AxesSubplot object at 0x1B7BBE38>,
      <matplotlib.axes._subplots.AxesSubplot object at 0x1CAEB2C8>,
      <matplotlib.axes._subplots.AxesSubplot object at 0x1C86ECA0>],
      [<matplotlib.axes._subplots.AxesSubplot object at 0x1C88D6A0>,
      <matplotlib.axes._subplots.AxesSubplot object at 0x1C8AD058>,
      <matplotlib.axes._subplots.AxesSubplot object at 0x1C8AD0D0>,
      <matplotlib.axes._subplots.AxesSubplot object at 0x1C8C0AD8>]],
      dtype=object)
```



Больше всего шутеров было выпущено в 2015 году

Далее мы проанализируем характерные жанры для нескольких известных издателей, тем самым нормализовав данные

In [10]:

```

1 def PubInf(gdsPub, lim, publisher):
2     x_var = 'Genre'
3     groupby_var = 'Year'
4     gds_agg = gdsPub.loc[:, [x_var, groupby_var]].groupby(groupby_var)
5     vals = [gdsPub[x_var].values.tolist() for i, gdsPub in gds_agg]
6
7     plt.figure(figsize=(16,9), dpi= 80)
8     colors = [plt.cm.Spectral(i/float(len(vals)-1)) for i in range(len(vals))]
9     n, bins, patches = plt.hist(vals, gdsPub[x_var].unique().__len__(), stacked=True, c
10
11     plt.legend({group:col for group, col in zip(np.unique(gdsPub[groupby_var]).tolist(),
12     plt.title(f"Гистограмма по жанрам издателя ${publisher}$", fontsize=22)
13     plt.xlabel(x_var)
14     plt.ylabel("Количество тайтлов")
15     plt.ylim(0, lim)
16     plt.xticks(ticks=bins, labels=np.unique(gdsPub[x_var]).tolist(), rotation=90, horiz
17     plt.show()

```

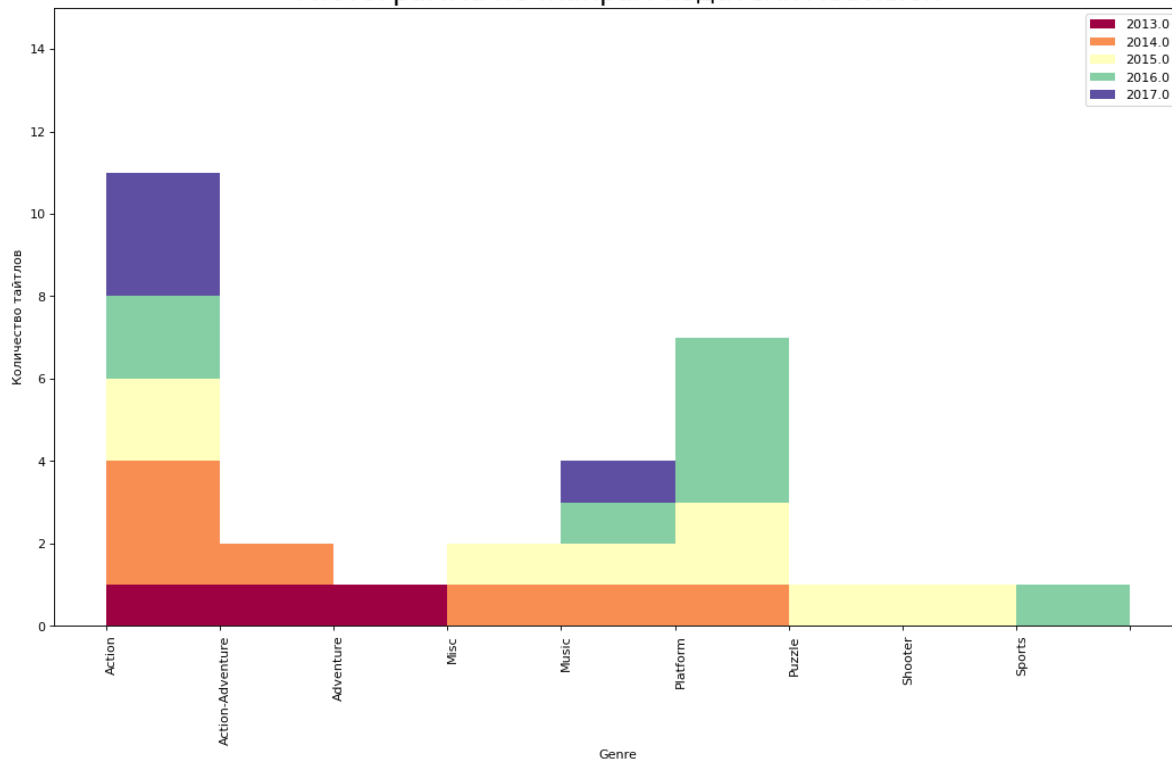
In [11]:

```

1 act = gds[gds.Publisher == 'Activision']
2 PubInf(act, 15, 'Activision')

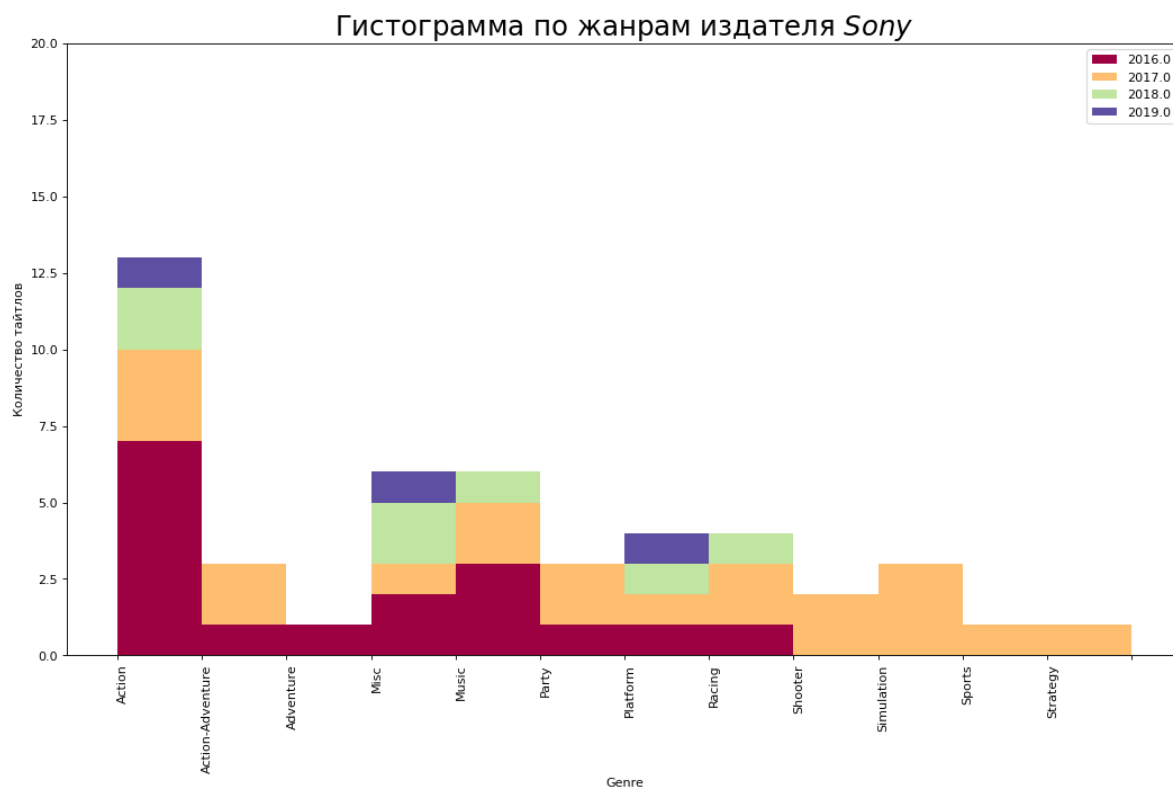
```

Гистограмма по жанрам издателя Activision



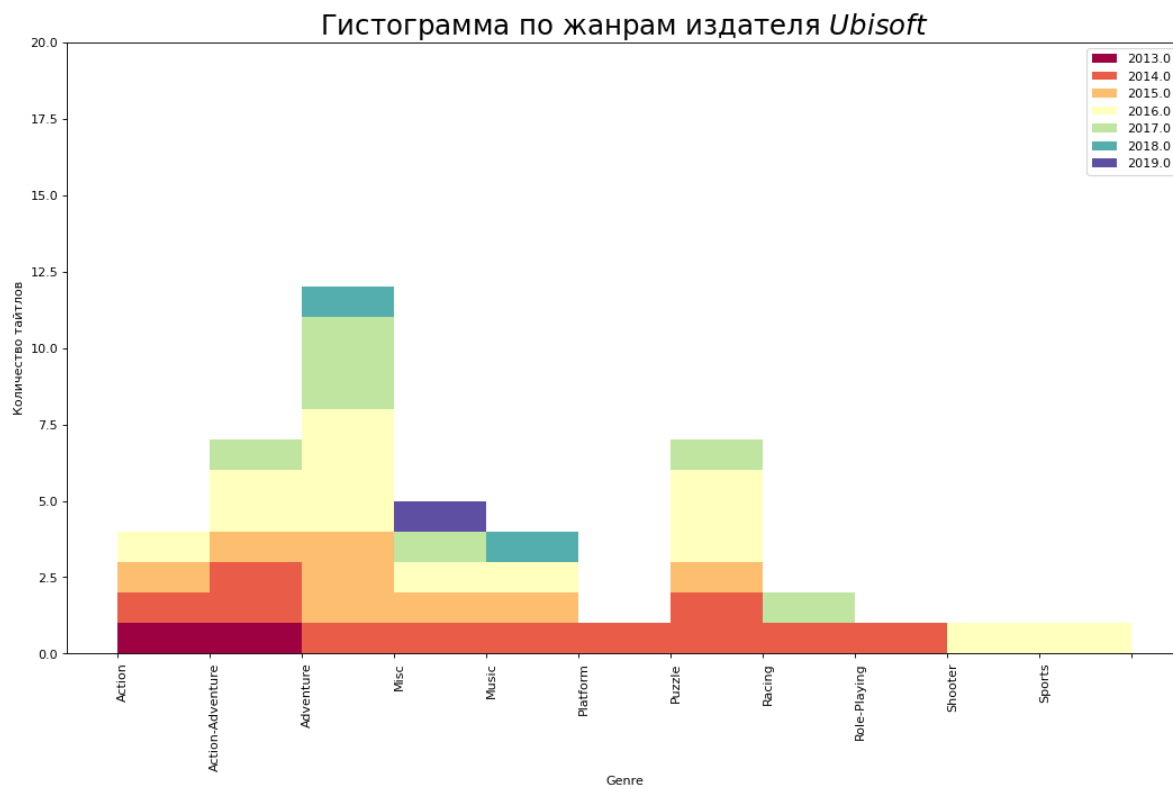
In [12]:

```
1 sony = gds[gds.Publisher == 'Sony Interactive Entertainment']  
2 PubInf(sony, 20, 'Sony')
```



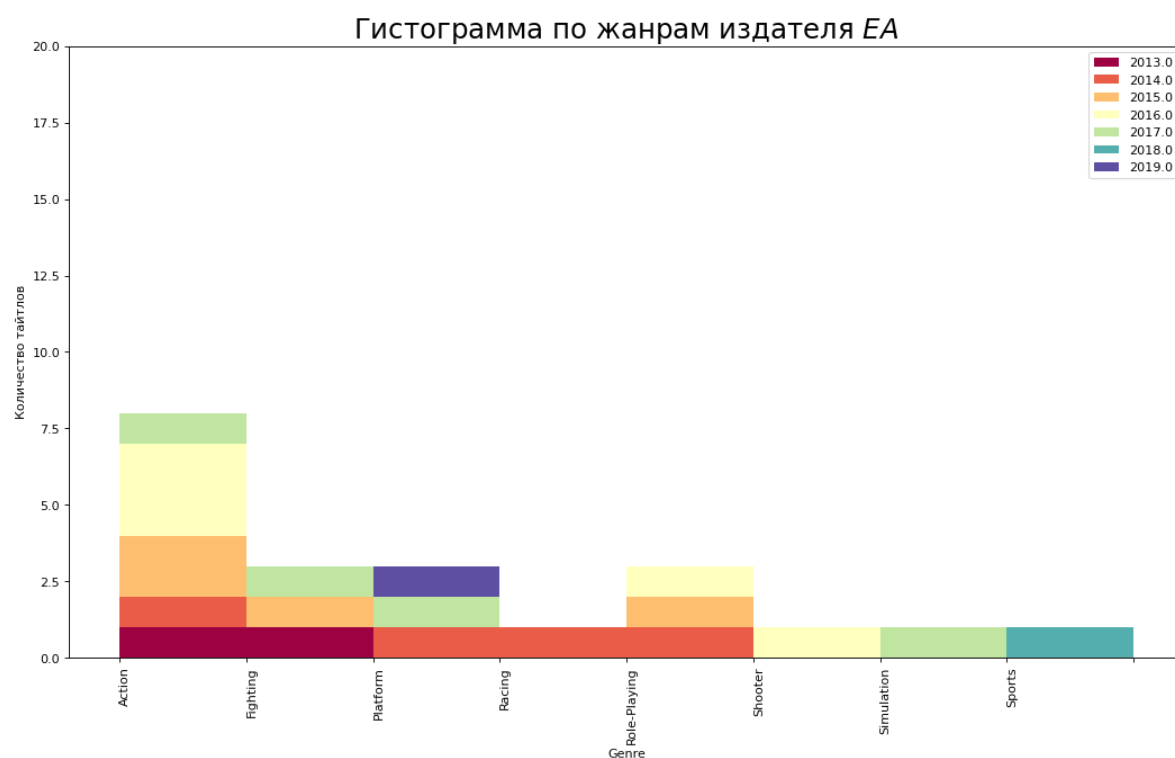
In [13]:

```
1 ubi = gds[gds.Publisher == 'Ubisoft']  
2 PubInf(ubi, 20, 'Ubisoft')
```



In [14]:

```
1 ea = gds[gds.Publisher == 'Electronic Arts']  
2 PubInf(ea, 20, 'EA')
```



В качестве дополнительного материала использовалась статья "50 оттенков matplotlib" - <https://habr.com/ru/post/468295/> (<https://habr.com/ru/post/468295/>).

In []:

1