Оглавление

Исследование компьютерных игр

Вы работаете в интернет-магазине «Стримчик», который продаёт по всему миру компьютерные игры.
Из открытых источников доступны исторические данные о продажах игр, оценки пользователей и экспертов,
жанры и платформы (например, Xbox или PlayStation). Вам нужно выявить определяющие успешность игры закономерности.
Это позволит сделать ставку на потенциально популярный продукт и спланировать рекламные кампании.
Перед вами данные до 2016 года. Представим, что сейчас декабрь 2016 г., и вы планируете кампанию на 2017-й.
Нужно отработать принцип работы с данными. Неважно, прогнозируете ли вы продажи на 2017 год по данным
2016-го или же 2027-й — по данным 2026 года.

В наборе данных попадается аббревиатура ESRB (Entertainment Software Rating Board) — это ассоциация,
определяющая возрастной рейтинг компьютерных игр. ESRB оценивает игровой контент и присваивает ему
подходящую возрастную категорию, например, «Для взрослых», «Для детей младшего возраста» или
«Для подростков».

Данные за 2016 год могут быть неполными.

1 Задачи проекта

1. Открою файл с данными и изучу общую информацию
2. Проведу предобработку данных:
   - Заменю названия столбцов (приведу к нижнему регистру).

- Преобразую данные в нужные типы. Опишу, в каких столбцах заменил тип данных и почему.
- Обработаю пропуски при необходимости. Объясню, почему заполнил пропуски определённым образом или почему не стал это делать. Опишу причины, которые могли привести к пропускам. Внимание на аббревиатуру 'tbd' в столбце с оценкой пользователей. Отдельно разберу это значение и опишу, как его обработать.
- Посчитаю суммарные продажи во всех регионах и запишу их в отдельный столбец.

3. Проведу исследовательский анализ данных:
   - Посмотрю, сколько игр выпускалось в разные годы. Важны ли данные за все периоды?
   - Посмотрю, как менялись продажи по платформам. Выберу платформы с наибольшими суммарными продажами и построю распределение по годам. За какой характерный срок появляются новые и исчезают старые платформы?
   - Возьму данные за соответствующий актуальный период. Актуальный период определю самостоятельно в результате исследования предыдущих вопросов. Основной фактор — эти данные помогут построить прогноз на 2017 год.
   - Не буду учитывать в работе данные за предыдущие годы.
   - Какие платформы лидируют по продажам, растут или падают? Выберу несколько потенциально прибыльных платформ.
   - Построю график «ящик с усами» по глобальным продажам игр в разбивке по платформам. Опишу результат.
   - Посмотрю, как влияют на продажи внутри одной популярной платформы отзывы пользователей и критиков. Построю диаграмму рассеяния и посчитаю корреляцию между отзывами и продажами. Сформулирую выводы.
   - Соотнесу выводы с продажами игр на других платформах.
   - Посмотрю на общее распределение игр по жанрам. Что можно сказать о самых прибыльных жанрах? Выделяются ли жанры с высокими и низкими продажами?
4. Составлю портрет пользователя каждого региона:
   - Определю для пользователя каждого региона (NA, EU, JP)
     - - Самые популярные платформы (топ-5). Опишу различия в долях продаж.
     - - Самые популярные жанры (топ-5). Поясню разницу.
     - - Влияет ли рейтинг ESRB на продажи в отдельном регионе?
5. Проверю гипотезы:
   - Средние пользовательские рейтинги платформ Xbox One и PC одинаковые
   - Средние пользовательские рейтинги жанров Action (англ. «действие», экшен-игры) и Sports (англ. «спортивные соревнования») разные.
     - - Задам самостоятельно пороговое значение alpha. Опишу как сформулировал нулевую и альтернативную гипотезы. Какой критерий применил для проверки гипотез и почему.
6. Напишу общий вывод.
7. Дополнительные исследования:
   - Объём продаж производителей.
   - Старые и новые платформы.

In [1]:
```
import pandas as pd
#импортирую библиотеку pandas
import numpy as np
#импортирую библиотеку numpy
import plotly.express as px
#импортирую библиотеку plotly.express
import plotly.graph_objects as go
#импортирую библиотеку plotly.graph_objects
from plotly.subplots import make_subplots
#импортирую библиотеку lotly.subplots
import seaborn as sns
#импортирую библиотеку seaborn
import matplotlib.pyplot as plt
#импортирую библиотеку matplotlib.pyplot
from matplotlib.colors import LogNorm
#импортирую библиотеку matplotlib.colors
from scipy import stats as st
#импортирую библиотеку scipy
import random as rn
#импортирую библиотеку random
from IPython.display import display
#импортирую display для красивого выведения датафремов в среде JB
import cowsay as dino
#импортирую cowsay для крайней строки
```

In [2]:
```
!pip install cowsay
#установка библиотеки cowsay
```

```
Requirement already satisfied: cowsay in c:\p\anaconda\envs\da_practicum_env\lib\site-packages (5.0)
```

In [3]:
```
try:
    data = pd.read_csv('/datasets/games.csv')
    #указываю путь csv-файла для ревьюера на linux
except:
    data = pd.read_csv(r'C:\P\Anaconda\envs\da_practicum_env\and_my_files\20221130_games.csv')
    #локально читаю csv-файл
```

1.1 Описание данных

- Name — название игры
- Platform — платформа
- Year_of_Release — год выпуска
- Genre — жанр игры
- NA_sales — продажи в Северной Америке (миллионы проданных копий)
- EU_sales — продажи в Европе (миллионы проданных копий)
- JP_sales — продажи в Японии (миллионы проданных копий)
- Other_sales — продажи в других странах (миллионы проданных копий)
- Critic_Score — оценка критиков (максимум 100)
- User_Score — оценка пользователей (максимум 10)
- Rating — рейтинг от организации ESRB (англ. Entertainment Software Rating Board). Эта ассоциация определяет рейтинг компьютерных игр и присваивает им подходящую возрастную категорию.

In [4]:
```
data.head()
#вывожу первые 5 строк датафрейма
```

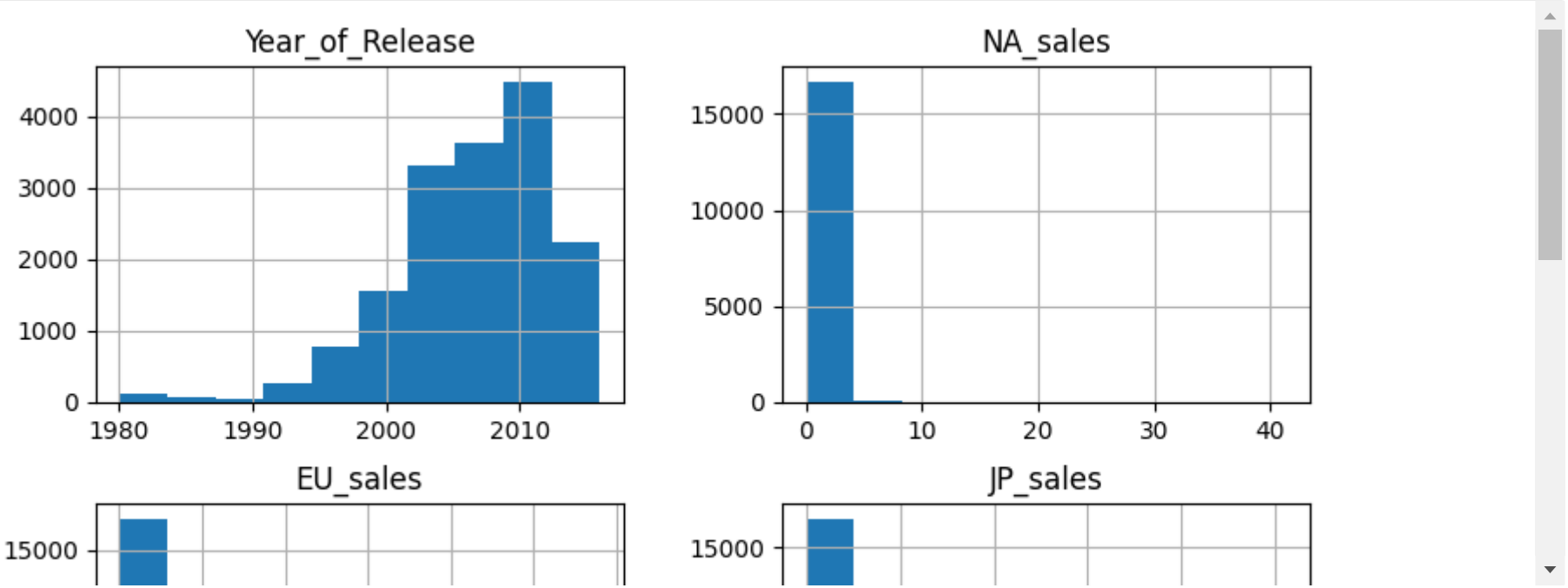
Out[4]:

	Name	Platform	Year_of_Release	Genre	NA_sales	EU_sales	JP_sales	Other_sales	Critic_Score	User_Score	Rating
0	Wii Sports	Wii	2006.0	Sports	41.36	28.96	3.77	8.45	76.0	8	E
1	Super Mario Bros.	NES	1985.0	Platform	29.08	3.58	6.81	0.77	NaN	NaN	NaN
2	Mario Kart Wii	Wii	2008.0	Racing	15.68	12.76	3.79	3.29	82.0	8.3	E
3	Wii Sports Resort	Wii	2009.0	Sports	15.61	10.93	3.28	2.95	80.0	8	E
4	Pokemon Red/Pokemon Blue	GB	1996.0	Role-Playing	11.27	8.89	10.22	1.00	NaN	NaN	NaN

In [5]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 16715 entries, 0 to 16714
Data columns (total 11 columns):
 #   Column           Non-Null Count  Dtype  
---  ------           --------------  -----  
 0   Name             16713 non-null  object 
 1   Platform         16715 non-null  object 
 2   Year_of_Release  16446 non-null  float64
 3   Genre            16713 non-null  object 
 4   NA_sales         16715 non-null  float64
 5   EU_sales         16715 non-null  float64
 6   JP_sales         16715 non-null  float64
 7   Other_sales      16715 non-null  float64
 8   Critic_Score     8137 non-null   float64
 9   User_Score       10014 non-null  object 
 10  Rating           9949 non-null   object 
dtypes: float64(6), object(5)
memory usage: 1.4+ MB
```

In [6]:
```
data.hist(figsize=(9,9));#общие гистограммы для числовых столбцов датафрейма
```



- В годах популярны значения после 2000 года.
- Данные по продажам NA могут превышать 5, но в основном сконцентрированы до этого значения.
- Данные по продажам EU могут превышать 2.5, но в основном сконцентрированы до этого значения.
- Данные по продажам JP могут превышать 1, но в основном сконцентрированы до этого значения.
- Данные по продажам в невыделенных регионах могут превышать 1, но в основном сконцентрированы до этого значения.
- Оценки критиков стремятся к нормальному распределению со смещением влево.
- Оценки пользователей имеют не числовой тип данных.

1.2 Функции, списки и полезные колонки для проекта

In [7]:
```
def pass_value_barh(df):
    try:
        (
            (df.isna().sum())
            .to_frame()
            .rename(columns = {0:'space'})
            .query('space > 0')
            .sort_values(by = 'space', ascending = True)
            .plot(kind = 'barh', figsize = (19,6), legend = False, fontsize = 16, color='black')
            .set_title('Пропуски' + '\n', fontsize = 22, color = 'Navy')
        );
    except:
        print('пропусков не осталось :) ')
#функция для визуализации пропущенных значений в барах
```

In [8]:
```
def cat_plfrm (platform):
    old_plfrm = ['2600', '3DO', 'DC', 'DS', 'GB','GBA', 'GC', 'GEN', 'GG', 'N64', 'NES', 'NG',
                 'PCFX', 'PS', 'PS2', 'PS3', 'PSP', 'SAT', 'SCD',
                 'SNES', 'TG16', 'WS', 'Wii', 'X360', 'XB']
    actual_plfm = ['WiiU', '3DS', 'XOne', 'PS4', 'PC', 'PSV']
    if platform in actual_plfm:
        return platform
    if platform in old_plfrm:
        return 'OLD_PLTFM'
    else:
        return 'ISSUE'
#функция для категоризации платформ
```

In [9]:
```
def s_matrix_plt (index):
    platform = index
    index = good_data.query('flag_tbd != "1" and platform == @platform')\
                             .pivot_table(index= ['ads_plt'],
                              values=['total_cop', 'user_score','critic_score'],
                              aggfunc='median')#медиана для снижения влияния выбросов
                    #query('year_of_release == 2015 and flag_tbd != "1" and platform == @platform')
    fig = px.scatter_matrix(index,
                     title='Матрица диаграмм рассеяния для ' + platform,
                     labels={col:col.replace('_', ' ') for col in index.columns})\
                     .update_traces(diagonal_visible=False)
    fig.show()
#функция для построения матриц корреляции
```

In [10]:
```
def plt_corr (index):
    platform = index
    index_plt = good_data.query('flag_tbd != "1" and platform == @platform')\
                               .pivot_table(index= ['ads_plt'],
                                values=['total_cop', 'user_score','critic_score'],
                                aggfunc='median')#медиана для снижения влияния выбросов
                          #query('year_of_release == 2015 and flag_tbd != "1" and platform == @platform')
    return index_plt.corr()
#функция для вывода Коэффициента Пирсона
```

In [11]:
```
def pie_top5 (df):
    for i in ['Европа', 'Япония', 'Северная Америка', 'Другие регионы']:#добавил перевод, для отоббражения при построе
        df_t = df.sort_values(by=i, ascending = False)[:5]
        tig = make_subplots(rows=1, cols=1)
        tig.add_trace(go.Pie(labels=df_t[:5].index, values=df_t[i], name='выбранный период'))

        tig.update_traces(textinfo='percent+label', hole=.05, hoverinfo='label+percent+name',
                          textposition='inside',legendgrouptitle_text='жанр')
        tig.update_layout(height=450, width=450,
                          title_text=i, title_x=0.5)
        tig.update_annotations(align='right', yshift=10)
        tig.show()
#функция для построения круговых диаграмм по топ5
```

In [12]:
```
def cat_brand (platform):
    Sony = ['PS3', 'PS2', 'PS', 'PS4', 'PSP', 'PSV', ]
    Nintendo = ['Wii', 'WiiU', '3DS', 'GB', 'NES', 'SNES', 'GBA', 'N64', 'GC', 'DS',]
    Microsoft = ['X360', 'XOne', 'XB', ]
    PC = ['PC', ]
    Atari = ['2600', ]
    Sega = ['GEN', 'DC', 'SAT', 'SCD', 'GG', ]
    Bandai = ['WS', ]
    SNK = ['NG', ]
    NEC = ['TG16', 'PCFX']
    The3DOC = ['3DO', ]#разработчик The 3DO Company, а производитель Panasonic, Sanyo, Creative и Goldstar (ныне LG)
    if platform in Sony:
        return 'Sony'
    if platform in Nintendo:
        return 'Nintendo'
    if platform in Microsoft:
        return 'Microsoft'
    if platform in PC:
        return 'PC'
    if platform in Atari:
        return 'Atari'
    if platform in Sega:
        return 'Sega'
    if platform in Bandai:
        return 'Bandai'
    if platform in SNK:
        return 'SNK'
    if platform in NEC:
        return 'NEC'
    if platform in The3DOC:
        return 'The3Doc'
    else:
        return 'ISSUE'
# функцию для категоризации брендов
```

In [13]:
```
data['ads_plt'] = np.arange(len(data))
#добавляю уникальное значение для каждой строки датасета
```

In [14]:
```
cool_pltf = ['PS2', 'DS', 'PC', 'GBA', 'PS3', 'X360', 'XOne',
             'PSV', '3DS', 'PSP', 'Wii', 'WiiU', 'PS4']
#список платформ из топов по количеству релизов в разные периоды
old_pltf = ['NES', 'GB', 'SNES', 'N64', 'PS', 'XB',
            '2600', 'GC', 'GEN', 'DC', 'SAT', 'SCD',
            'WS', 'NG', 'TG16', '3DO', 'GG', 'PCFX']
#список старых платформ
actual_plfm = ['WiiU', '3DS', 'XOne', 'PS4', 'PC', 'PSV']
#актуальные платформы
genre_list = ['Shooter', 'Role-Playing', 'Action', 'Sports', 'Fighting',
       'Racing', 'Simulation', 'Platform', 'Misc', 'Strategy',
       'Adventure', 'Puzzle']
#список жанров
```

2 Проведу предобработку данных

2.1 Заменю названия столбцов

In [15]:
```
data.columns
#проверяю названия колонок на соответствие snake_case
```

Out[15]:
```
Index(['Name', 'Platform', 'Year_of_Release', 'Genre', 'NA_sales', 'EU_sales',
       'JP_sales', 'Other_sales', 'Critic_Score', 'User_Score', 'Rating',
       'ads_plt'],
      dtype='object')
```

In [16]:
```
data = data.rename(columns={'Name':'name', 'Platform':'platform', 'Year_of_Release':'year_of_release',
                            'Genre':'genre', 'NA_sales':'na_sales',
                            'EU_sales':'eu_sales', 'JP_sales':'jp_sales',
                            'Other_sales':'other_sales', 'Critic_Score':'critic_score',
                            'User_Score':'user_score', 'Rating':'rating' })
                           #привел все названия к snake_case
```

2.2 Преобразую данные в нужные типы

2.2.1 в float

In [17]:
```
data['user_score'].sort_values(ascending = False).unique()
#смотрю уникальные значения в столбце
#особенность данных: рейтинги привести к числовому типу, при этом 0 нельзя
```

Out[17]:
```
array(['tbd', '9.7', '9.6', '9.5', '9.4', '9.3', '9.2', '9.1', '9', '8.9',
       '8.8', '8.7', '8.6', '8.5', '8.4', '8.3', '8.2', '8.1', '8', '7.9',
       '7.8', '7.7', '7.6', '7.5', '7.4', '7.3', '7.2', '7.1', '7', '6.9',
       '6.8', '6.7', '6.6', '6.5', '6.4', '6.3', '6.2', '6.1', '6', '5.9',
       '5.8', '5.7', '5.6', '5.5', '5.4', '5.3', '5.2', '5.1', '5', '4.9',
       '4.8', '4.7', '4.6', '4.5', '4.4', '4.3', '4.2', '4.1', '4', '3.9',
       '3.8', '3.7', '3.6', '3.5', '3.4', '3.3', '3.2', '3.1', '3', '2.9',
       '2.8', '2.7', '2.6', '2.5', '2.4', '2.3', '2.2', '2.1', '2', '1.9',
       '1.8', '1.7', '1.6', '1.5', '1.4', '1.3', '1.2', '1.1', '1', '0.9',
       '0.7', '0.6', '0.5', '0.3', '0.2', '0', nan], dtype=object)
```

> В столбце 'user_score' встречается значение tbd (To Be Determined, то есть "Будет определено"), обычно перед релизом (так как на момент релиза игра еще не доступна пользователям, но может быть доступна критикам) игры такая аббревиатура указывается у значений которые еще на этапе сбора. Вынесу значение tbd в отдельную колонку флагом. Возможно пригодится в исследовании.

In [18]:
```
data.loc[(data['user_score'] == 'tbd'), 'flag_tbd'] = '1'
#создам параметр для сохранения флага tbd
```

In [19]:
```
data['user_score'] = pd.to_numeric(data['user_score'], errors='coerce')
#перевожу колонку user_score принудильно в float с потерей tbd
```

> Для параметра user_score задал тип float, чтобы можно было работать с его свойствами числовыми методами.

2.2.2 шкала оценки

Шкала 'user_score' 10-бальная, а шкала 'critic_score' 100-бальная. Предполажил, что оценки критиков нужно разделить на 10, чтобы шкалы стали одинаковыми. После окончания работы над проектом, заметил, что это не влияет на проверку гипотиз, так как там сравниваются Z-параметры. Еще рейтинги использовал в опредлении корреляции, но приведение к одной шкале тоже не дало эффекта изменений. В итоге решил закоментировать этот этап.

In [20]:
```
#data['user_score'].sort_values(ascending = False).unique()
```

In [21]:
```
#data['critic_score'].sort_values(ascending = False).unique()
```

In [22]:
```
#data['critic_score'] = data['critic_score'] /10
#заменяю значения в том же параметре
```

2.2.3 в datatime

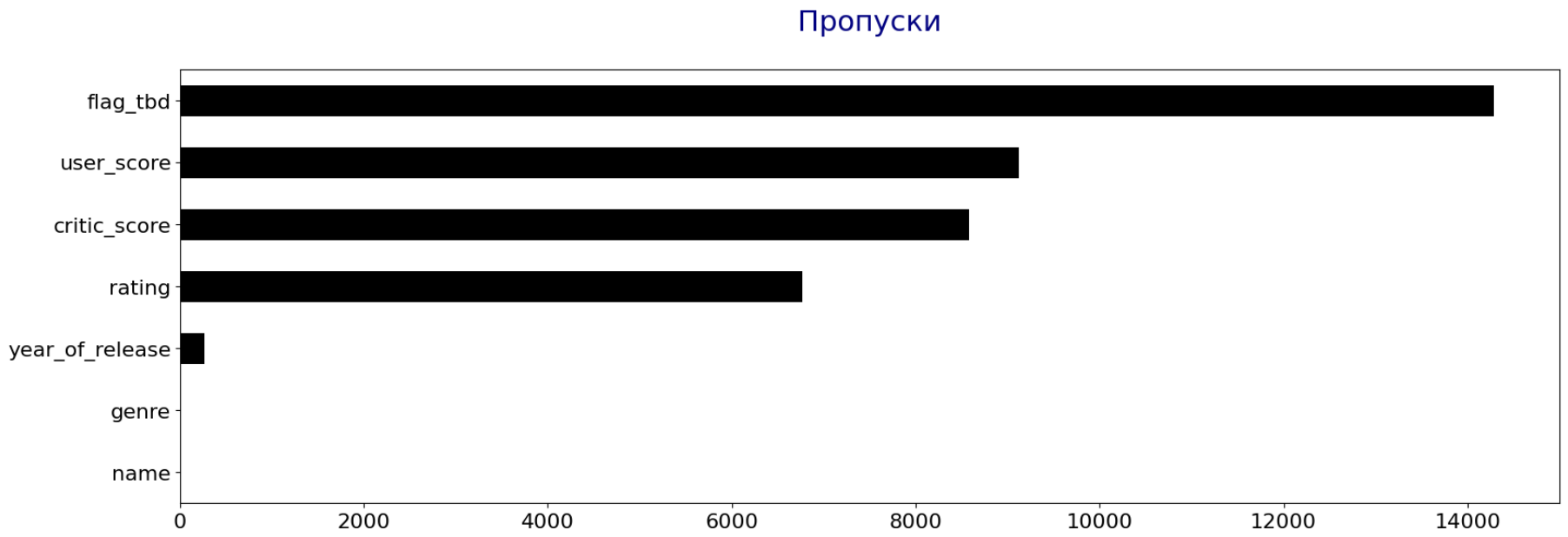
Преобразования параметра 'year_of_release' в datatime делать не имеет смысла, так как исследование делаю по годам, без учета месяца и времени года.

2.3 Обработаю пропуски

In [23]:
```
data.isna().sum().sum()
#считаю количество пропусков в столбцах
```

Out[23]: 39033

In [24]:
```
pass_value_barh(data)
#визуализирую количество пропусков по параметрам
```



In [25]:
```
data.loc[(data['name'].isna()) | (data['genre'].isna()) & (data['year_of_release'].isna())]
#проверю совпадение минимальных значений пропусков по параметрам
```

Out[25]:

	name	platform	year_of_release	genre	na_sales	eu_sales	jp_sales	other_sales	critic_score	user_score	rating	ads_plt	flag_tbd
659	NaN	GEN	1993.0	NaN	1.78	0.53	0.00	0.08	NaN	NaN	NaN	659	NaN
14244	NaN	GEN	1993.0	NaN	0.00	0.00	0.03	0.00	NaN	NaN	NaN	14244	NaN

In [26]:
```
data = data.dropna(subset=['genre'])
#удаляю два пропуска по колонке genre
```

In [27]:
```
pd.options.mode.chained_assignment = None
#убираю предупреждение об ошибке
data['rating'] = data['rating'].fillna('UKW')
#заполняю пропуски ESRB на unknown, так как тип данных в этом параметре object
```

Пропуски в 'rating' заполнил заглушкой - UKW, для удобства выведения в диаграммы. Пропуски в 'year_of_release' обработаю после добавление колонки с общими продажами.

In [28]:
```
data.shape
#проверяю количество оставшихся строк
```

Out[28]:
```
(16713, 13)
```

In [29]:
```
data.duplicated().sum()
#проверяю явные дубликаты
```

Out[29]:
```
0
```

Пропуски 'user_score', 'critic_score' не заполнял синтетическими значениями, так как они сформированы по механизму MNAR (Missing Not At Random) и я не могу их явно предсказать из информации датасета. А пропуски в 'flag_tbd' созданы мной.

2.4 Посчитаю суммарные продажи во всех регионах и запишу их в отдельный столбец

In [30]:
```
data['total_cop'] = data['na_sales'] + data['eu_sales'] + data['jp_sales'] + data['other_sales']
data['total_cop'] = data['total_cop'].round(2)
```

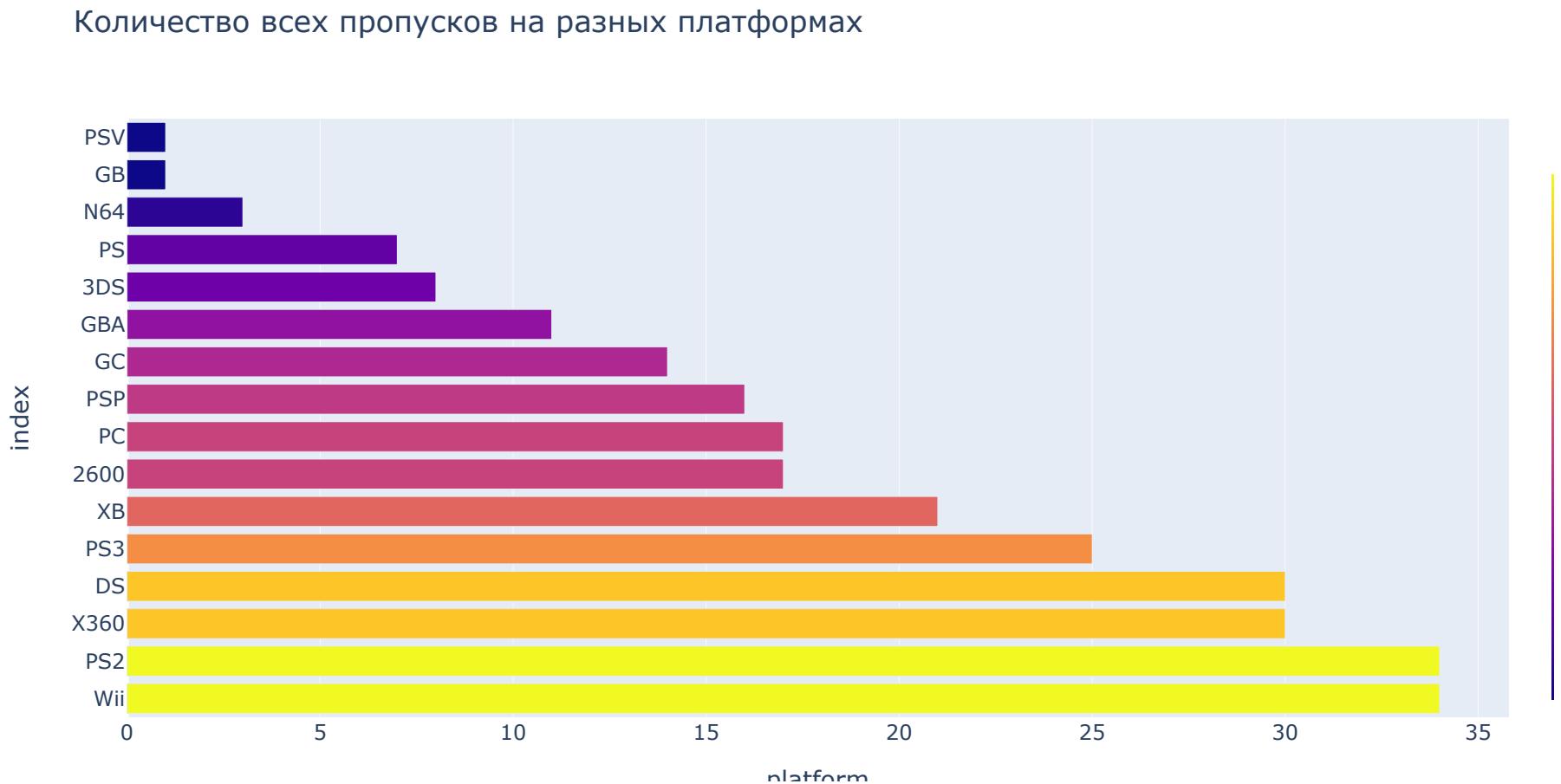
Посмотрю, какие объёмы продаж с пропусками в годах, сортировку сделаю по количеству милиоонов проданных копий.

In [31]:
```
display(data.loc[data['year_of_release'].isna()].sort_values(by='total_cop', ascending = False).head(11))
```

	name	platform	year_of_release	genre	na_sales	eu_sales	jp_sales	other_sales	critic_score	user_score	rating	ads_plt	flag_tb
183	Madden NFL 2004	PS2	NaN	Sports	4.26	0.26	0.01	0.71	94.0	8.5	E	183	Nal
377	FIFA Soccer 2004	PS2	NaN	Sports	0.59	2.36	0.04	0.51	84.0	6.4	E	377	Nal
456	LEGO Batman: The Videogame	Wii	NaN	Action	1.80	0.97	0.00	0.29	74.0	7.9	E10+	456	Nal
475	wwe Smackdown vs. Raw 2006	PS2	NaN	Fighting	1.57	1.02	0.00	0.41	NaN	NaN	UKW	475	Nal
609	Space Invaders	2600	NaN	Shooter	2.36	0.14	0.00	0.03	NaN	NaN	UKW	609	Nal
627	Rock Band	X360	NaN	Misc	1.93	0.33	0.00	0.21	92.0	8.2	T	627	Nal
657	Frogger's Adventures: Temple of the Frog	GBA	NaN	Adventure	2.15	0.18	0.00	0.07	73.0	NaN	E	657	
678	LEGO Indiana Jones: The Original Adventures	Wii	NaN	Action	1.51	0.61	0.00	0.21	78.0	6.6	E10+	678	Nal
719	Call of Duty 3	Wii	NaN	Shooter	1.17	0.84	0.00	0.23	69.0	6.7	T	719	Nal
805	Rock Band	Wii	NaN	Misc	1.33	0.56	0.00	0.20	80.0	6.3	T	805	Nal
1131	Call of Duty: Black Ops	PC	NaN	Shooter	0.58	0.81	0.00	0.23	81.0	5.2	M	1131	Nal

И посмотрю распределение пропусков по платформам.

In [32]:
```
plt_isna = data.loc[data['year_of_release'].isna(), 'platform'].value_counts()
px.bar(plt_isna, color='platform', hover_name=plt_isna,
       x = 'platform', title = 'Количество всех пропусков на разных платформах')
```



В основном пропуски в устаревших платформах, кроме 17-ти игр на PC (в реальном исследовании, можно заполнить эти 17 пропусков руками, через поиск в www.ya.ru), но только одна игра входит в топ по продажам (заметил в ходе дальнейшего исследования). Удаляю эти пропуски, так как это не сильно повлияет на исследование, а с параметром year_of_release удобней будет работать.

In [33]:
```
data['year_of_release'].isna().sum()
```

Out[33]:
```
269
```

In [34]:
```
data = data.dropna(subset=['year_of_release'])
```

'year_of_release' переведу в тип int, ведь это целые числа.

In [35]:
```
data['year_of_release'].astype(int).head()
```

Out[35]:
```
0    2006
1    1985
2    2008
3    2009
4    1996
Name: year_of_release, dtype: int32
```

In [36]:
```
data.shape
#посмотрю количество оставшихся строк после всех правок
```

Out[36]:
```
(16444, 14)
```

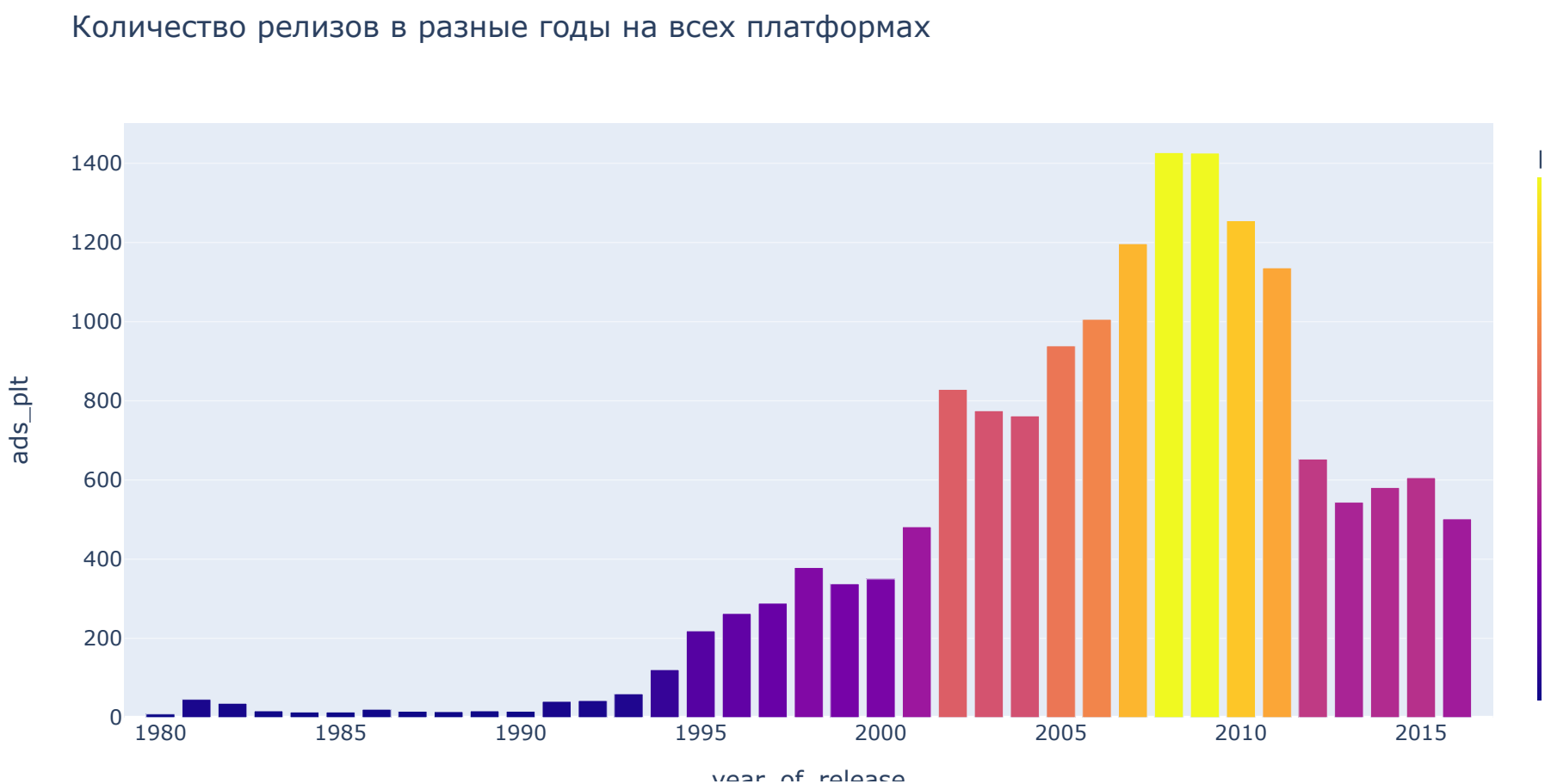
Потери, после обработки не всех пропусков, составили менее 2%.

3 Проведу исследовательский анализ данных

3.1 Посмотрю, сколько игр выпускалось в разные годы на всех платформах

In [37]:
```
all_game = data.groupby(['year_of_release']).agg('count')
```

In [38]:
```
px.bar(all_game,
       y ='ads_plt',
       color='platform',
       range_x=[1979,2017],
       title = 'Количество релизов в разные годы на всех платформах')
```

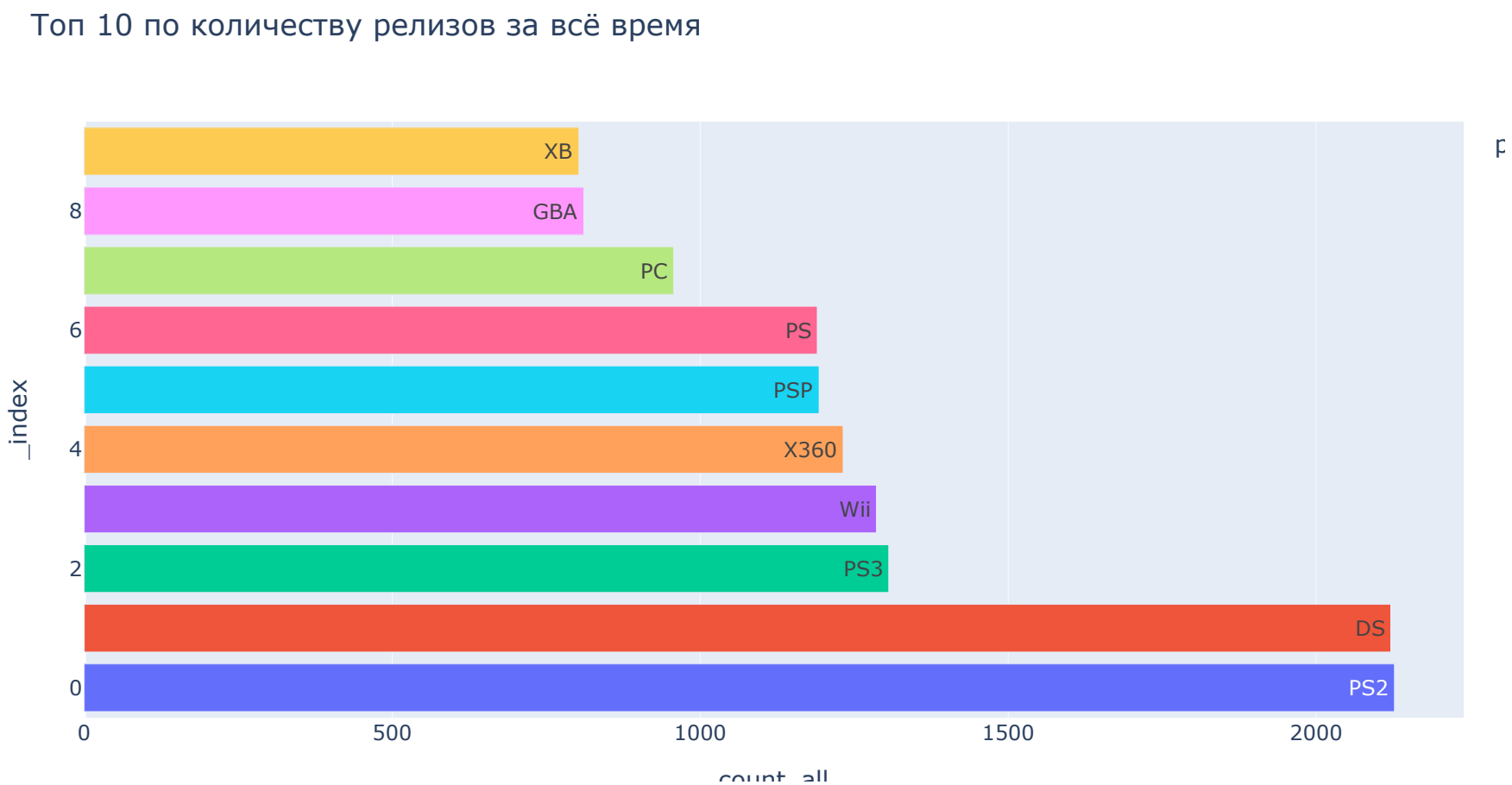


В датасете есть данные по выпуску игр с 1980 года. С 1994 года начинается тенденция увеличения количества релизов видеоиг. В 2002 году происходит резкое увеличение количества релизов. В 2008-2009 году количество релизов в год достигает своего пика. С 2010 можно наблюдать падение релизов. Резкое снижение происходит в 2012 году, то есть между резкими изменениями прошло 10 лет. За 2016 год данные не полные, но можно судить о нисходящем тренде. Гистограмма стримиться к нормальному распределению, а значит с 2016 по 2020 будет продолжаться нисходящий тренд - значит, можно исследовать подробно крайний, полный по данным год. Ситуацию может изменить появление какой-то инновации в игровой индустрии, которая изменит линию тренда на восходящий.

3.2 Посмотрю, сколько игр выпускалось в разные периоды по платформам

Посмотрю ТОП-10 по количеству вышедшех игр за все годы датасета.

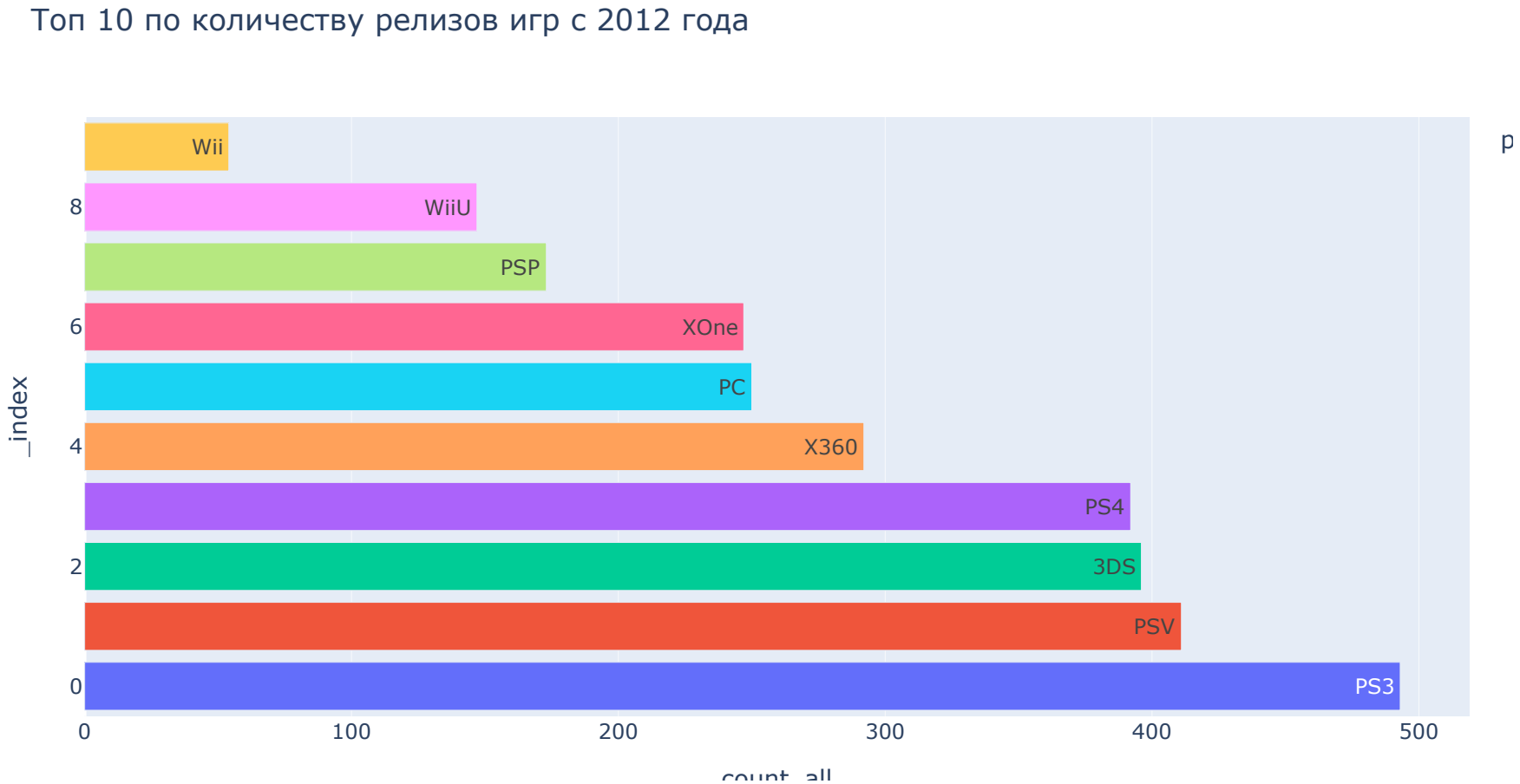
In [39]:
```
count_all_game = data.pivot_table(index=['platform'],
                                  values=['ads_plt'],aggfunc=['count']).reset_index()
count_all_game.columns = ['platform', 'count_all']
px.bar(count_all_game.sort_values(by='count_all',ascending=False).reset_index().head(10),
       x = 'count_all',
       color='platform', text='platform',
       title = 'Топ 10 по количеству релизов за всё время')
```



Первая тройка по количеству вышедших игр за всё время: PS2, DS, PS3. Интересно, что Nintendo DS это портативная консоль, на смену которой уже пришло её следующее поколение 3DS. Можно считать, что Nintendo (DS и Wii) и Sony (PS2 и PS3) в общем зачете оставляют Microsoft с X360 на третьем месте (а реально пятом). PC попал в топ, но только на 8 место. Три места из десяти занимают портативные консоли DS, PSP, GBA.

Посмотрю ТОП-10 по количеству вышедших с 2012 года, когда происходит падение количества релизов.

In [40]:
```
count_2012_game = data.query('year_of_release >= 2012').pivot_table(index=['platform'],
                                                                    values=['ads_plt'],aggfunc=['count']).reset_index(
count_2012_game.columns = ['platform', 'count_all']
px.bar(count_2012_game.sort_values(by='count_all',ascending=False).reset_index().head(10),
       x = 'count_all',
       color='platform', text='platform',
       title = 'Топ 10 по количеству релизов игр с 2012 года')
```



Если рассмотреть количество игр с 2012 года, в этот год резко снизилось количество релизов, то первая тройка почти не изменилась, два места у Sony (PS3 и PSV) и одно место у Nintendo (3DS). А сразу на четвертом месте Sony PS4, которая вышла в 2013 году. Microsoft с X360 всё так же на пятом месте, при этом их новое поколение XOne на седьмом месте, с выходом платформы в 2013 году. PC улучшил свои позиции и теперь располагается на 6 месте. Портативных консолей стало больше, теперь их четыре и лидирует PSV (вес 219 грамм), на втором 3DS (вес 226 грамм) и за ним PSP (от 280 грамм до 158 грамм). На смену Wii (10 место) приходит WiiU (9 место), которая позиционирует себя как портативная консоль (джойстик с экраном в одном корпусе), но с возможностью стационарного использования (вес 1600 грамм).

3.3 Посмотрю, как менялись продажи по платформам

In [41]:
```
pivot_game_platform = data.pivot_table(index = ['platform', 'year_of_release'],
                                       values='total_cop', aggfunc = ['sum']).reset_index()
pivot_game_platform.columns = ['platform', 'year_of_release', 'total_cop']
```

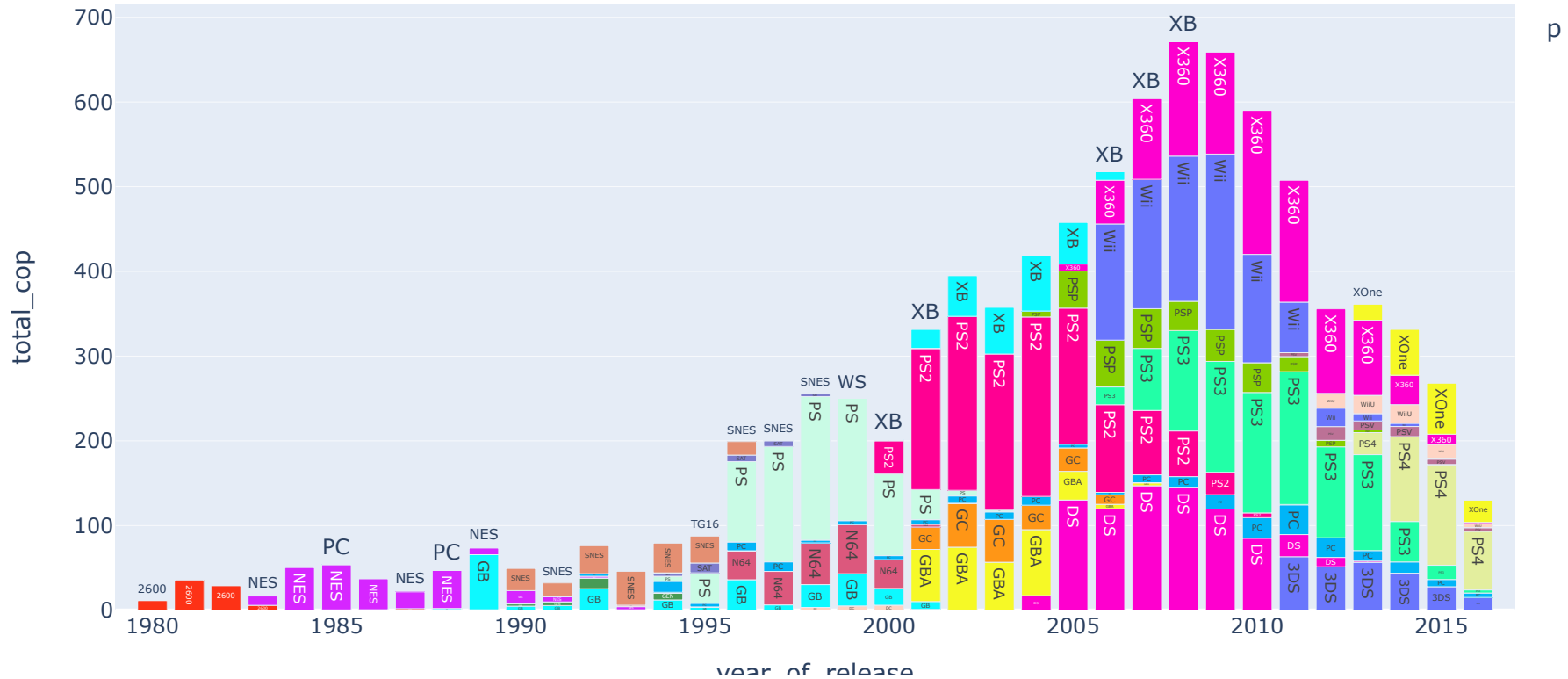
In [42]:
```
pivot_game_platform
```

Out[42]:

	platform	year_of_release	total_cop
0	2600	1980.0	11.38
1	2600	1981.0	35.68
2	2600	1982.0	28.88
3	2600	1983.0	5.84
4	2600	1984.0	0.27
...	...	...	...
233	XB	2008.0	0.18
234	XOne	2013.0	18.96
235	XOne	2014.0	54.07
236	XOne	2015.0	60.14
237	XOne	2016.0	26.15

238 rows × 3 columns

In [43]:
```
px.bar(pivot_game_platform,
       x = 'year_of_release',
       y='total_cop',
       color='platform',
       title = 'Миллионы копий проданных игр на разных платформах за всё время',
       range_y=[0,715],
       text='platform',
       range_x=[1979,2017],
       hover_name='year_of_release',
       color_discrete_sequence=px.colors.qualitative.Light24)
```

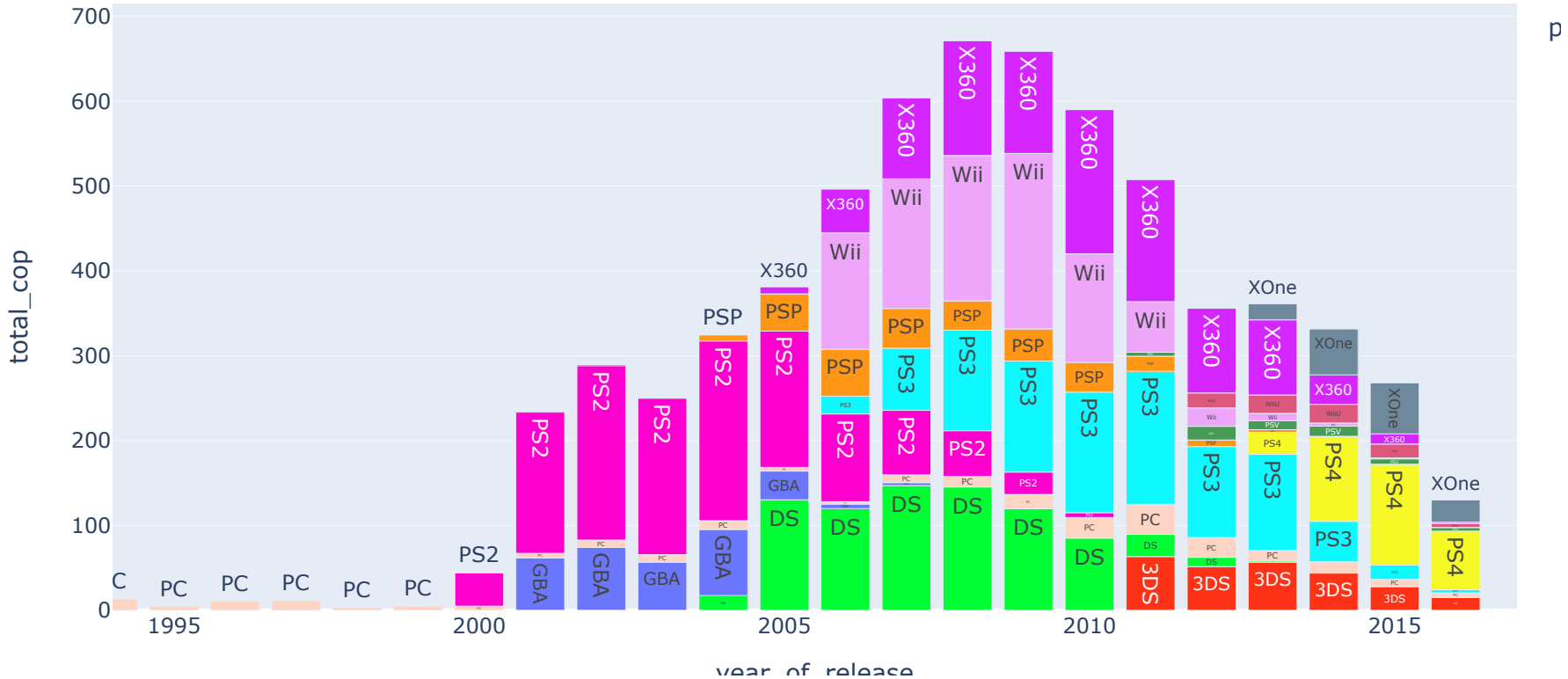


In [44]:
```
data.groupby('platform').agg({'total_cop': 'sum'}).sort_values(by='total_cop').tail(6).index.to_list()
```

Out[44]:
```
['PS', 'DS', 'Wii', 'PS3', 'X360', 'PS2']
```

In [45]:
```
px.bar(pivot_game_platform.query('platform == @cool_pltf'),
       x = 'year_of_release',
       y='total_cop',
       color='platform',
       title = 'Миллионы копий проданных игр на разных платформах за всё время',
       range_y=[0,715],
       text='platform',
       range_x=[1994,2017],
       hover_name='year_of_release',
       color_discrete_sequence=px.colors.qualitative.Light24)
```

Миллионы копий проданных игр на разных платформах за всё время



Ярко выделяется платформа компании Sony PlayStation (PS, PS2, PS3, PS4), с 1994 года цифровые копии проданных игр занимают большой объём в каждом году. С 2001 года в конкурентах у Sony появляется Microsoft XB. В 2006 году хорошую конкуренцию по продажам начинает Nintendo Wii, но и Sony с Microsoft не стоят на месте, они выпускают новое поколение платформ. PC занимает незначительный объём продаж во всех годах.

До 2004 года Nintendo лидирует в сегменте портативных консолей. С 2004 в явных конкурентах у портативных Nintendo появляется Sony с PSP.

При близком рассмотрении 1994 года, можно заметить, что как раз в этот год выходит первая PlayStation (в Японии) и начинается тенденция к увеличению продаж цифровых копий и выпуска игр. Пики в 2002, 2008, 2009, 2012 годы данной диаграммы очень схожи с пиками диаграммы 'Количество игр вышедших в разные годы'.

Интересно, что в 2008 и 2009 году (высший пик для этой диаграммы) платформы компании Nintendo Wii и Nintendo DS занимают больший объем в сравнении с конкурентами. С 2012 года самый уверенным брендом по продажам выглядит Sony с PS3, PS4 и PSV.

3.4 Выберу платформы с наибольшими суммарными продажами и построю распределение по годам

In [46]:
```
data['platform'].unique()
```

Out[46]:
```
array(['Wii', 'NES', 'GB', 'DS', 'X360', 'PS3', 'PS2', 'SNES', 'GBA',
       'PS4', '3DS', 'N64', 'PS', 'XB', 'PC', '2600', 'PSP', 'XOne',
       'WiiU', 'GC', 'GEN', 'DC', 'PSV', 'SAT', 'SCD', 'WS', 'NG', 'TG16',
       '3DO', 'GG', 'PCFX'], dtype=object)
```

Создам два списка:

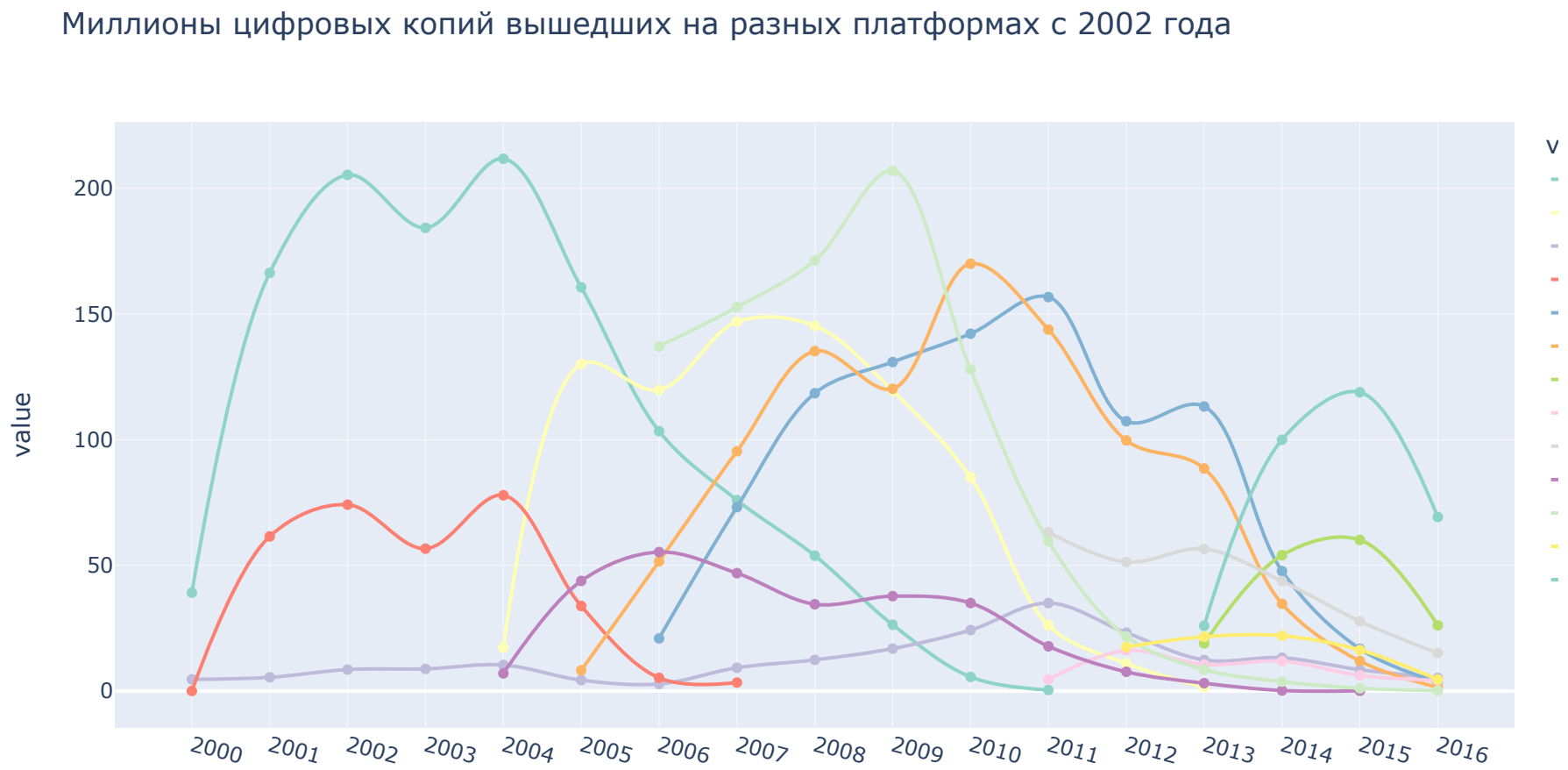
- список платформ из топа с 2012 года и топ-3 с 2002 года + Game Boy Advance (GBA)
- все остальные платформы

In [47]:
```
#список cool_pltf добавлен в пункте 1.2
#список actual_plfm добавлен в пункте 1.2
```

Рассмотрю период с 2002 года, когда начался резкий взлёт релизов и количества цифровых копий, для платформ попавших в топы релизов за разные годы.

In [48]:
```
pltf_year = data.pivot_table(index='year_of_release', columns='platform', values='total_cop',
                             aggfunc='sum').reset_index()
```

In [49]:
```
px.line(pltf_year.query('year_of_release >= 2000'),#указал 2000 год, чтобы видеть старты PS2 и GBA
        x='year_of_release',
        y=cool_pltf,
        markers=True,
        hover_name='year_of_release',
        line_shape = 'spline',
        title = 'Миллионы цифровых копий вышедших на разных платформах с 2002 года',
        color_discrete_sequence=px.colors.qualitative.Set3).update_xaxes(dtick=1,tickangle=15)
```



По общему количеству цифровых копий самые заметные пики на общем фоне у PS2 (2, 4 годы существования платформы) и Wii (на 3 год). Далее в этом плане выделяются PS3 и X360, первый пик у них на 3 году и второй на 5 году.
У Wii сразу на старте в 2006 году было много цифровых копий. Заметно, что у Sony PS3 и PS4 больше цифровых копий на старте, чем у Microsoft X360 и XOne. Если на старте много цифровых копий, а по и их количеству я оцениваю популярность платформы у пользователей, то можно предположить, что пользователи очень сильно ожидали платформу Nintendo Wii, на втором месте платформы Sony PS2-PS3 и на третьем Microsoft X360-XOne. Портативная Nintendo DS тоже на выходе имела большое количество цифровых копий, больше чем в том же году у PSP.

Тенденция построения линий у PS3 и X360 схожа, но до 2010 лидирует X360, а с 2011 лидерство переходит к PS3. Рассвет DS пришелся на угасание PS2, но совпал с более успешным стартом и подъемом Wii. С 2012 года многие платформы показывают нисходящий тренд, только PS3, 3DS, WiiU удерживают позиции и немного стремятся вверх. В 2013 выходит новое поколение приставок Sony PS4 и Microsoft XOne с резким восходящим трендом, и к 2015 году Sony имеет в два раза больше цифровых релизов, чем XOne. С 2013 года остальные платформы показывают падение количества цифровых релизов. PC, PSV и WiiU в 2014 показывает небольшой рост, но к 2015 приходят с нисходящим трендом. На 2015 год топ-3: PS4, XOne, 3DS.

Если оставить линии только Sony: то PS4 выходит на 7 год существования PS3, а PS3 выходит на 6 год существования PS2.
Если оставить линии только Microsoft: то XOne выходит на 8 год существования X360, а X360 выходит на 5 год существования XB (данные по XB из пункта 7.2).
Можно сделать вывод, что после выхода нового поколения консоли, поддержка старого продолжается 2-3 года.

Много пересечений линий (конкурентная борьба) в годы максимального количества релизов - 2008-2009.

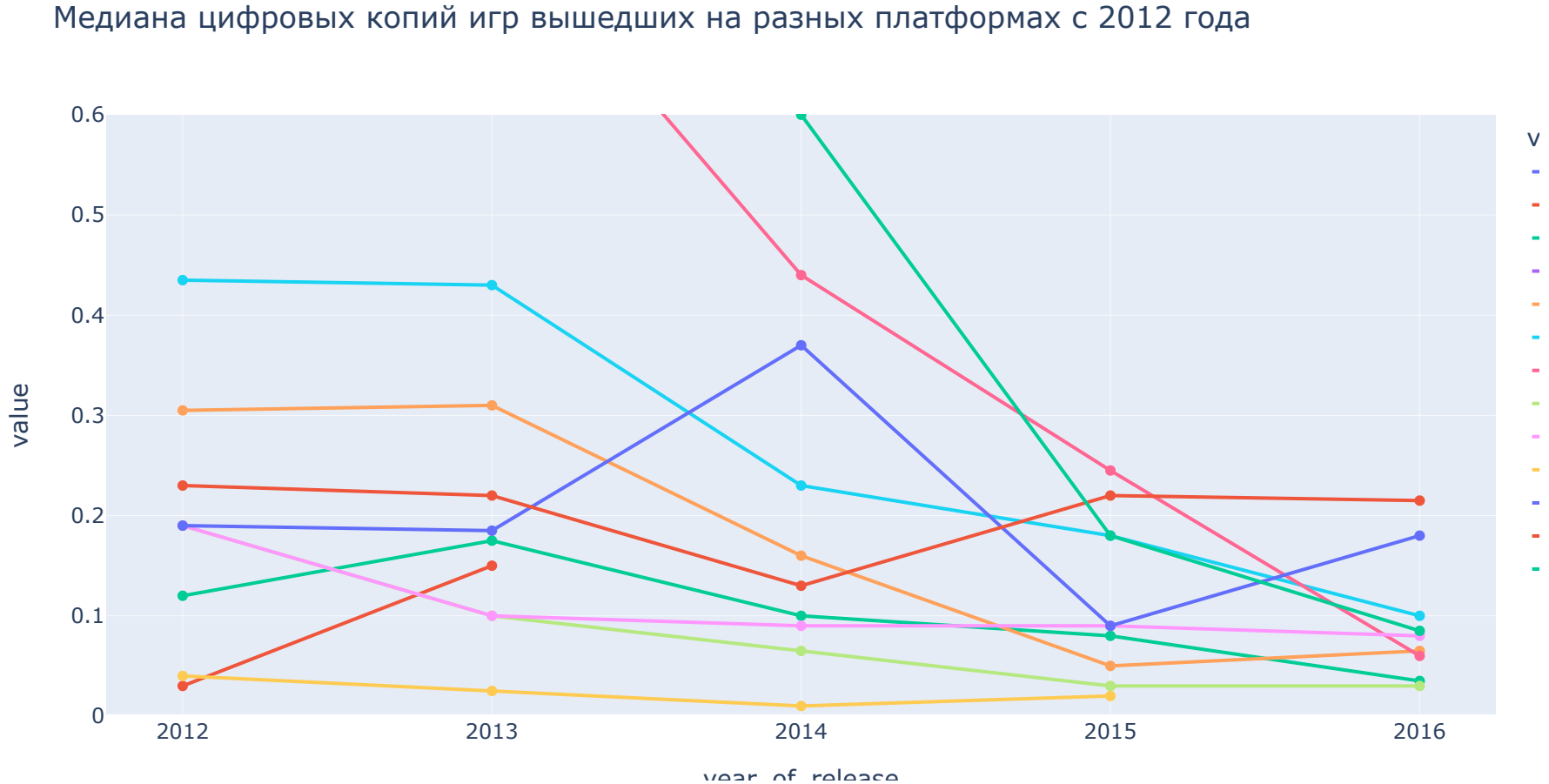
По высоким пикам PS2 и Wii можно судить о релизе на этих платформах очень успешных игр, которые привели к высокому выпуску цифровых копий. Использую медиану и попробую снизить это влияние на исследование более актуального периода. Рассмотрю медиану цифровых копий на платформах попавших в топы. В период с 2012 года, когда происходит резкое снижение релизов и количества цифровых копий.

Для снижения влияния выбросов применять дальше median (из выводов выше), но если нужно посмотреть объём занимаемый в

доле, то скорей всего лучше использовать sum при построении pivot table.

In [50]:
```
pltf_year_median = data.pivot_table(index='year_of_release',
                                    columns='platform', values='total_cop',
                                    aggfunc='median').reset_index()
```

In [51]:
```
px.line(pltf_year_median.query('year_of_release >= 2012'),
        x='year_of_release',
        y=cool_pltf,
        markers=True,
        hover_name='year_of_release',
        range_y=[0,0.6],
        title = 'Медиана цифровых копий игр вышедших на разных платформах с 2012 года').update_xaxes(dtick=1)
```



PS4 и XOne в нисходящем тренде по медиане цифровых копий с 2013 года, не смотря на то, что на выходе этих платформ было много цифровых копий. За год до выхода PS4 и XOne практически все платформы не изменили количества цифровых копий, кроме PC и DS, при том для DS этот 2013 год последний. В 2014 году все показывают падение, кроме портивной Wii. В 2015 годе все платформы в нисходящем тренде, кроме нового поколения Nintendo WiiU. Лидероом по медиане цифровых копий в 2015 году становится XOne, за ним идёт WiiU, а третье место делят между собой X360 и PS4.

Посмотрю медиану цифровых релизов на платформах, которые актуальны на 2015-2016 годы. За 2016 год данные не полные.

In [52]:
```
pivot_more2015_platform = data.pivot_table(index=['year_of_release', 'platform'],
                                           values=['total_cop'],
                                           aggfunc=['median']).reset_index()
pivot_more2015_platform.columns = ['year_of_release', 'platform', 'total_cop']
```

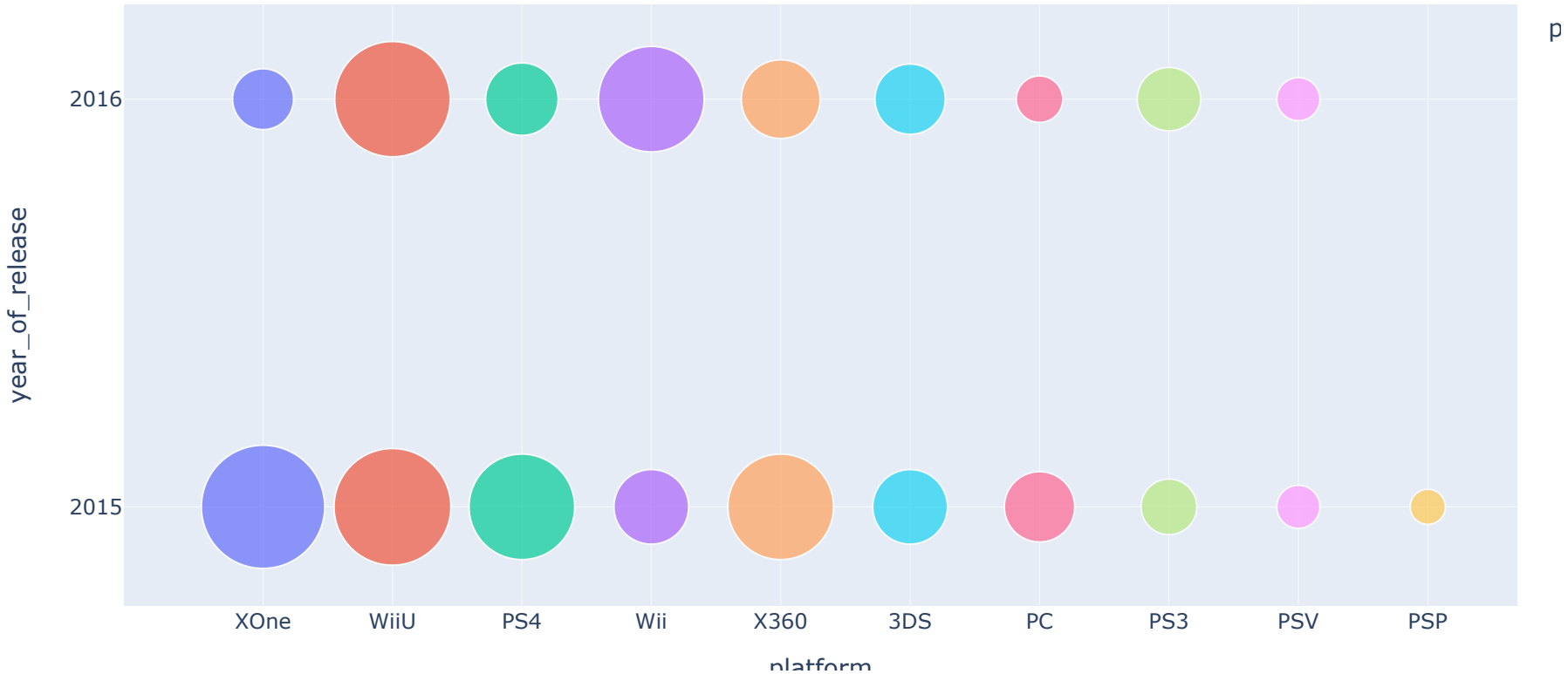
In [53]:
```
pivot_more2015_platform
```

Out[53]:

	year_of_release	platform	total_cop
0	1980.0	2600	0.770
1	1981.0	2600	0.465
2	1982.0	2600	0.540
3	1983.0	2600	0.460
4	1983.0	NES	1.635
...	...	...	...
233	2016.0	PSV	0.030
234	2016.0	Wii	0.180
235	2016.0	WiiU	0.215
236	2016.0	X360	0.100
237	2016.0	XOne	0.060

238 rows × 3 columns

In [54]:
```
px.scatter(pivot_more2015_platform.query('year_of_release >= 2015').sort_values(by='total_cop',ascending=False),
       x = 'platform',
       y='year_of_release',
       color='platform',
       title = 'Медиана цифровых копий в 2015-2016 годы',
       size='total_cop',
       log_y=True,
       hover_name='total_cop',
       size_max=50).update_yaxes(tickvals=[2015, 2016])
#всё, что осталось от презентации Гансом Рослингом анимации Gapminder на TED
```



В 2016 Выпуск цифровых копий продолжаются на всех платформах, кроме PSP:

- Nintendo - 3DS, Wii, WiiU
- Sony - PS3, PS4, PSV
- Microsoft - X360, XOne
- Персональный компьютер - PC

Тройка лидеров в 2015 году: XOne, WiiU, PS4.

Выделю самое актуальное поколение платформ по каждому бренду для list actual_plfm:

- портативные: 3DS, PSV (пришла на замену PSP);
- стационарные: PS4, XOne (пришел на замену X360), WiiU (пришла на замену Wii), PC (долгожитель).

In [55]:
```
data['platform'].sort_values().unique()
#список всех платформ для создания функции cat_plfrm
```

Out[55]:
```
array(['2600', '3DO', '3DS', 'DC', 'DS', 'GB', 'GBA', 'GC', 'GEN', 'GG',
       'N64', 'NES', 'NG', 'PC', 'PCFX', 'PS', 'PS2', 'PS3', 'PS4', 'PSP',
       'PSV', 'SAT', 'SCD', 'SNES', 'TG16', 'WS', 'Wii', 'WiiU', 'X360',
       'XB', 'XOne'], dtype=object)
```

In [56]:
```
len(data['platform'].sort_values().unique())
#количество платформ в исследовании
```

Out[56]:
```
31
```

In [57]:
```
#список actual_plfm добавлен в пункте 1.2
```

Создам новый параметр в датасете с ранжировкой на старые платформы и все остальные.

In [58]:
```
data['ctgize_platform'] = data['platform'].apply(cat_plfrm)
```

3.5 Характерный срок появления новых и исчезнавения старых платформ

Построю анимацию по годам выпуска игр для платформ попавших в топы, но за исключением актуальных платформ, на которых продолжается выпуск цифровых копий игр. Включу в следующую диаграмму платформы Wii, PSP и X360, так как им на смену

> пришло следующее поколение консолей WiiU, PSV и XOne.

In [59]:
```
life_span = data.pivot_table(index = ['platform', 'year_of_release'],
                             values='ads_plt',
                             aggfunc = ['count']).reset_index()
life_span.columns = ['platform', 'year_of_release', 'ads_plt']
```

In [60]:
```
anim_slow_life_span = px.bar(life_span.query('platform == @cool_pltf and platform not in @actual_plfm')\
                             .sort_values(by='ads_plt',ascending=False),
       x = 'year_of_release',
       y= 'ads_plt',
       color='platform',
       title = 'Количество релизов вышедших на разных устаревших платформах по годам',
       range_y=[0,500],
       text='platform',
       animation_frame='platform',
       animation_group='year_of_release',
       range_x=[1999,2017]).update_xaxes(dtick=1)
anim_slow_life_span.layout.updatemenus[0].buttons[0].args[1]['frame']['duration'] = 1400
anim_slow_life_span.layout.updatemenus[0].buttons[0].args[1]['transition']['duration'] = 2800
anim_slow_life_span
```

Из диаграммы видно, что вподряд проходят релизы следующее количество лет:

- DS - 10 лет, с пиком в пятом году платформы(есть игра в 1985, но в этом году не было такой приставки, похоже на аномалию)
- GBA - 8 лет, с пиком на третьем и пятом году платформы
- PS2 - 12 лет, с пиком на третьем году платформы
- PS3 - 11 лет, с пиком на шестом году платформы
- PSP - 12 лет, с пиком на третьем и седьмом году платформы

- Wii - 11 лет, с пиком на четвертом году платформы
- X360 - 12 лет, с пиком на седьмом году платформы.

Посмотрим, сколько в среднем живут платформы, за исключением чемпиона PC:

In [61]:
```
year_life = pd.Series([10,8,12,11,12,11,12])
#Создам объект Series, без PC и уходящих в 2017 год.
year_life.mean()
```

Out[61]:
```
10.857142857142858
```

В среднем, основные лидеры рынка живут 10 лет 8 месяцев. Для полноты исследования, платформу PC рассмотрю отдельно.

In [62]:
```
px.bar(life_span.query('platform == "PC"'),
       x = 'year_of_release',
       y= 'ads_plt',
       color='platform',
       title = 'Количество релизов вышедших в разные годы на PC',
       text='platform',
       range_x=[1993,2017]).update_xaxes(dtick=2)
```

Главный долгожитель PC. В данных до 1994 года есть пропуски (возможно нет данных или они потеряны), но с этого года PC получает новые игры 23 года и будет получать дальше. Пики в 2003, 2009 и 2011 годы. С 2013 года восходящий тренд. Осталось посмотреть сколько уже на рынке находятся актуальные платформы, за исключением исследованного PC.

In [63]:
```
actual_plfm_slow_life_span = px.bar(life_span.query('platform == @actual_plfm and platform not in "PC"')\
                                    .sort_values(by='ads_plt',ascending=False),
       x = 'year_of_release',
       y= 'ads_plt',
       color='platform',
       title = 'Количество релизов вышедших на актуальных поколениях платформ по годам',
       text='platform',
       animation_frame='platform',
       animation_group='year_of_release')
actual_plfm_slow_life_span.layout.updatemenus[0].buttons[0].args[1]['frame']['duration'] = 1000
actual_plfm_slow_life_span.layout.updatemenus[0].buttons[0].args[1]['transition']['duration'] = 1000
actual_plfm_slow_life_span
```

Следующие платформы на рынке представлены:

- 3DS - 6 лет, нисходящий тренд
- PS4 - 4 года, восходящий тренд
- PSV - 6 лет, нисходящий тренд
- WiiU - 5 лет, нисходящий тренд
- XOne - 4 года, восходящий тренд

Можно сделать вывод, что у PS4 и XOne есть потенциал для роста.

3.6 Определю период исследования для прогнозирования 2017 года

Задача проекта заключается в описании потенциально популярного продукта и планировании его рекламного бюджета на 2017 год. Данные за 2016 у меня не полные. Значит будет интересным рассмотреть актуальные данные за пару лет, соответственно возьму объемы цифровых копий за 2014-2015 годы, но платформы выберу из самых актуальных в 2016 году.

3.7 Возьму данные за соответствующий актуальный период

Создам новую переменную с фильтром от 2013 года, чтобы видеть на графиках как начинался 2014 год. И выберу данные только по актуальным платформам.

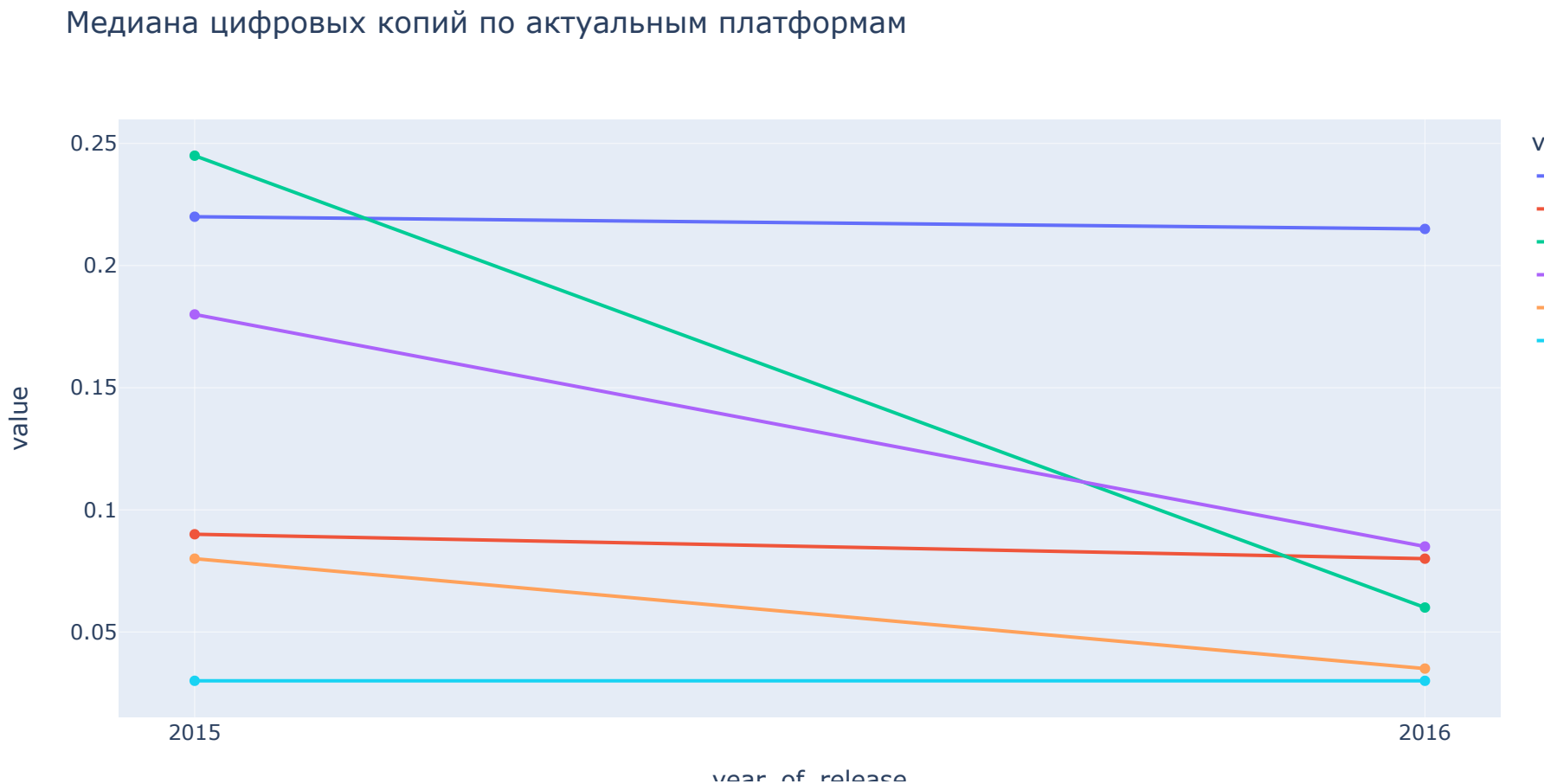
In [64]:
```
#good_data = data.query('year_of_release >= 2013 and platform == @actual_plfm')
```

In [65]:
```
good_data = data.query('year_of_release >= 2015')
```

3.8 Какие платформы лидируют по продажам, растут или падают, потенциально прибыльные

In [66]:
```
median_pvt_gd = good_data.pivot_table(index='year_of_release',
                                      columns='platform', values='total_cop',
                                      aggfunc='median').reset_index()
```

In [67]:
```
px.line(median_pvt_gd,
        x='year_of_release',
        y=actual_plfm,
        markers=True,
        title = 'Медиана цифровых копий по актуальным платформам').update_xaxes(dtick=1)
```



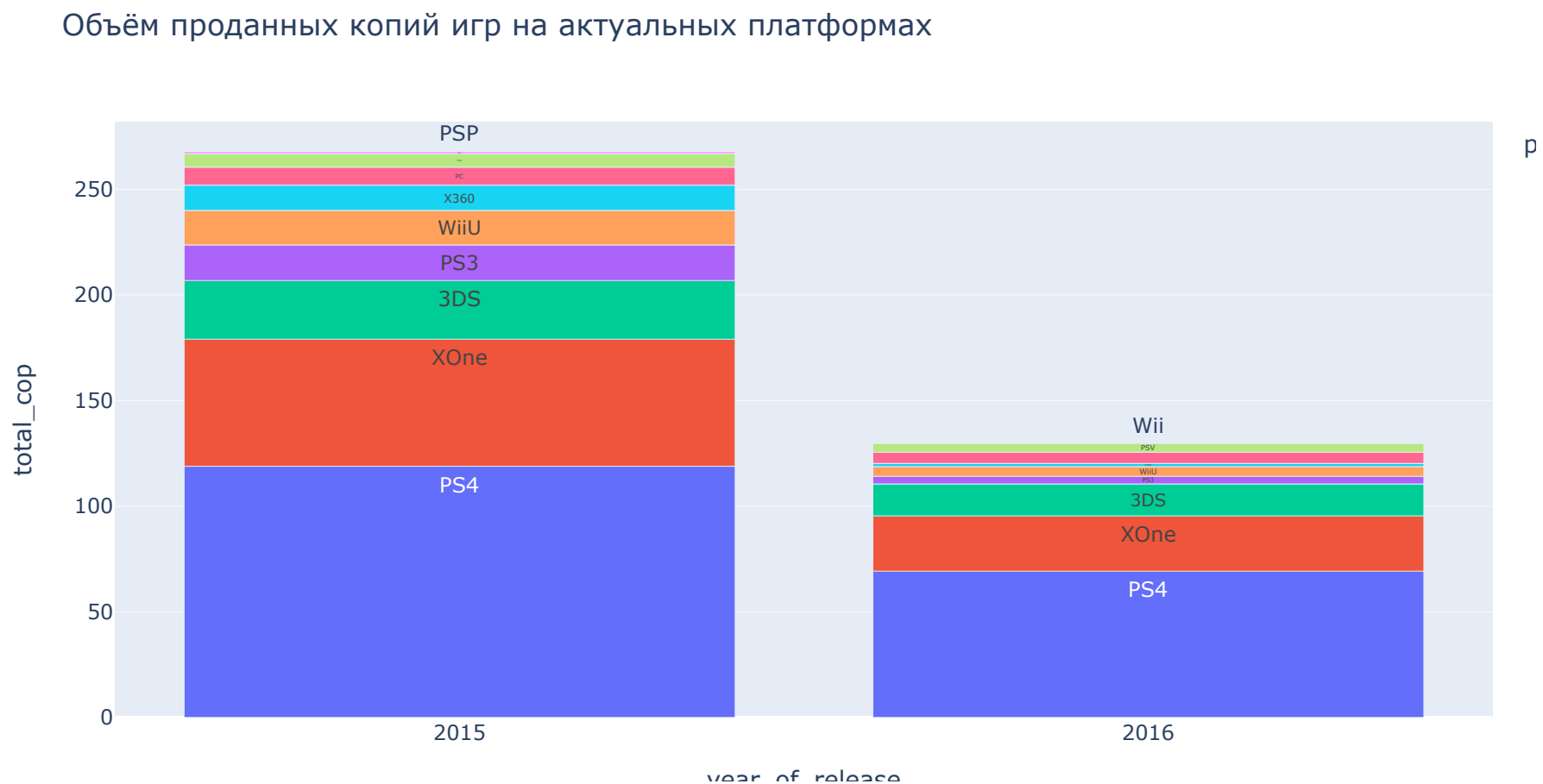
Лучше всего медиана продаж у WiiU, на втором месте PS4 с нисходящим трендом и на третьем 3DS с краткосрочным спокойным трендом. На четвертом XOne с сильным падением от 2015.

2014 год, первое место у PS4, второе у XOne, а на третьем WiiU. PC сразу за ними.
2015 год, первое место у XOne, второе у WiiU, а на третьем PS4.

Можно рассмотреть сумму всех цифровых копий, чтобы понимать объём с учетом удачных релизов.

In [68]:
```
sum_pvt_gd = good_data.pivot_table(index = ['platform', 'year_of_release'],
                                   values='total_cop',
                                   aggfunc = 'sum').reset_index().sort_values(by=['total_cop'],ascending=[False])
sum_pvt_gd.columns = ['platform', 'year_of_release', 'total_cop']
```

In [69]:
```
px.bar(sum_pvt_gd,
       x = 'year_of_release',
       y='total_cop',
       color='platform',
       title = 'Объём проданных копий игр на актуальных платформах',
       text='platform').update_xaxes(dtick=1)
```



На 2016 объёмы продаж рапределяются так же, как в 2015.

По общему объёму выпуска цифровых копий в 2014-2015 годы лидирует PS4, XOne - второе место, 3DS - третье место. PC занимает незначительный объём.

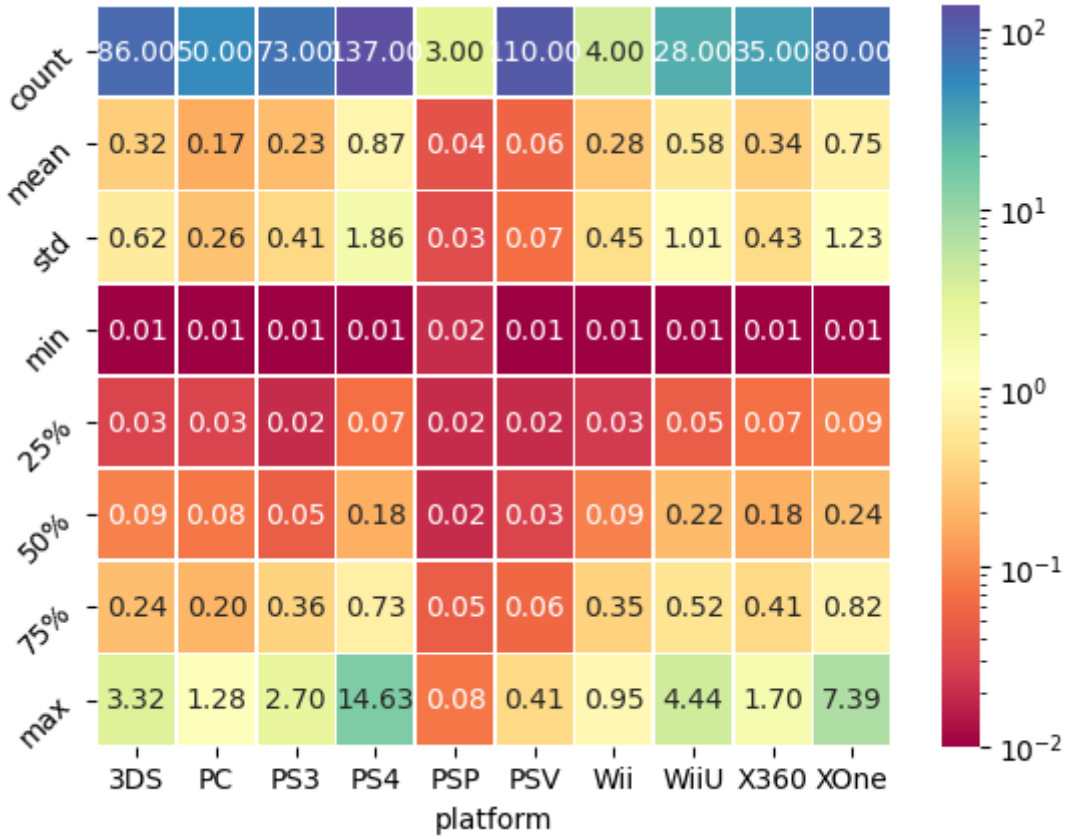
3.9 Построю график "ящик с усами" по глобальным продажам игр в разбивке по платформам, опишу результат

In [70]:
```
px.box(good_data#.query('2016 > year_of_release >= 2014')\
       .sort_values(by=['total_cop', 'year_of_release'], ascending=[False, True]),
       y='platform',
       x='total_cop',
       hover_name = 'name',
       title = 'Распределение объёмов цифровых релизов на актуальных платформах',
       color='year_of_release')
```

В 2016 поле 50% объёмов продаж (всё, что в цветной области это основная масса продаж копий игр) лучше всех у WiiU (+ самая высокая медиана), на втором месте PS4, а третье у XOne. На четвертом месте 3DS.

Ящики с усами сделал только по 2014-2015 году, так как за эты годы есть полные данные в выбранном мной периоде. Если есть выбросы, значит на платформе выходят игры, которые скупают большим тиражом. На PS4 такие высокотиражные игры интересней всего, самую высокую точку занимает Call of Duty: Black Ops 3 в 2014 году и Grand Theft Auto V в 2015 году. У XOne есть релизы этих же игр, но по суммарным выпускам цифровых копий в такие же годы, они занимают значительно меньшую позицию. Из портативных консолей много больших тиражей у 3DS. На PC есть такие успехи, но их мало. Медиана у всех от года к году падает, кроме WiiU у которой есть рост. Скорей всего в 2016 году произойдет падение количества выпуска цифровых копий.

In [71]:
```
describe_v = good_data.query('year_of_release == 2015')\
                      .pivot_table(index='ads_plt', columns='platform',
                                   values='total_cop').describe()
sns.heatmap(describe_v, vmax=1, annot=True,
                        cmap="Spectral",
                        norm=LogNorm(),
                        annot_kws={'size':10}, fmt='.2f', linewidths=.5)
plt.yticks(rotation=45);
```



Охарактеризую данные за 2015 год. Среднее у всех больше медианы, значит в распределении количества цифровых релизов есть длинных хвосты с высокими значениями. Стандартное отклонение большое, значит данные в признаках не однородны. Первый квартиль (25%) у всех примерно на одном уровне, минимальное количество цифровых копий практически одинаковое.

Широкие межквартильные размахи у PS4, XOne и WiiU, в которых находится 50% значений, высока вероятность попасть в это значение. Медиана у всех находится в нижней части МКР (IQR), значит в основном происходят большие выпуски цифровых копий.

Ориентируясь на верхние значения 'усов' PS4, XOne и PC за 2015 год, можно посмотреть ТОП5 продаваемых 2015 по каждой платформе. Возьму дополнительно 2014 год, так как в этому году по всем трём платформам крайние значения верхних 'усов' выше 2015 года.

In [72]:
```
#пятёрка супер тиражи на ps4 за два года
good_data.query('2016 > year_of_release >= 2014 and total_cop >= 1.72 and platform == "PS4"')\
                .sort_values(by='total_cop', ascending=False).head(5);
```

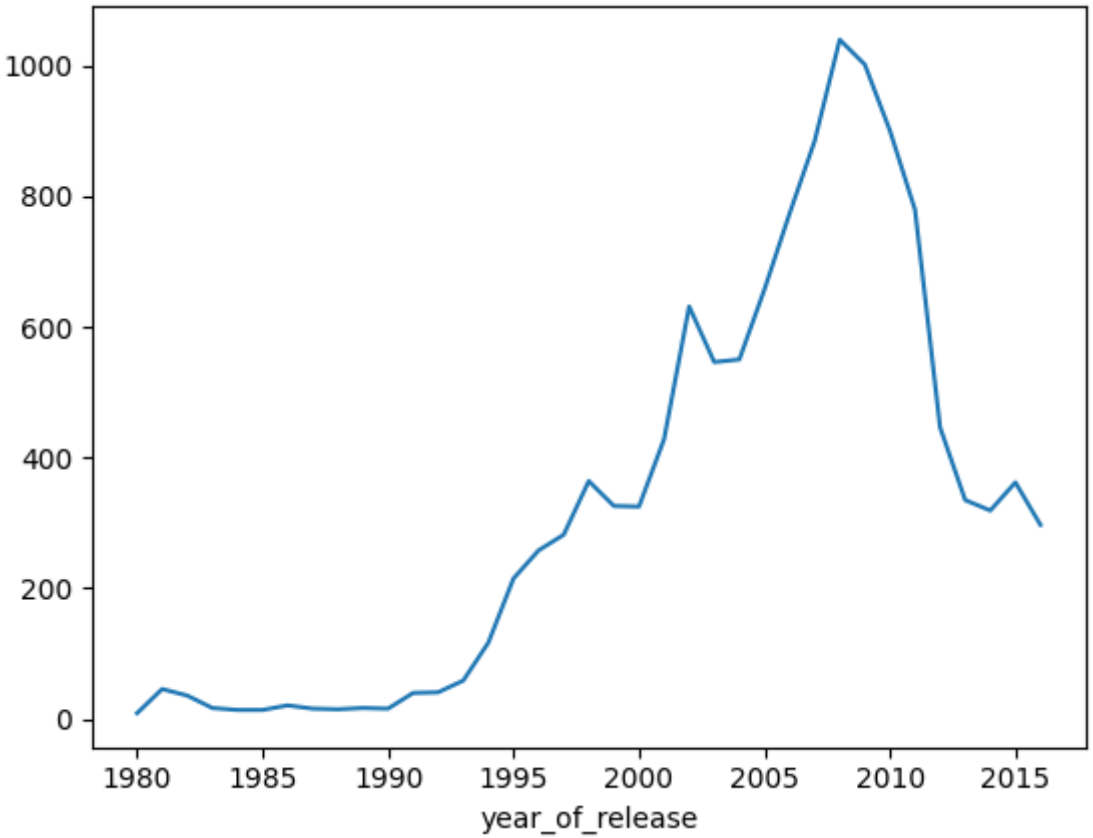
In [73]:
```
#пятёрка супер тиражи на XOne за два года
good_data.query('2016 > year_of_release >= 2014 and total_cop >= 1.8 and platform == "XOne"')\
                .sort_values(by='total_cop', ascending=False).head(5);
```

In [74]:
```
#пятёрка супер тиражи на PC за два года
good_data.query('2016 > year_of_release >= 2014 and total_cop >= 0.35 and platform == "PC"')\
                .sort_values(by='total_cop', ascending=False).head(5);
```

Список супер-релизов по объёмам продаж по каждой платформе с 2014 года:

- **PS4** - Call of Duty: Black Ops 3, Grand Theft Auto V, FIFA 16, Star Wars Battlefront (2015), Call of Duty: Advanced Warfare.
- **XOne** - Call of Duty: Black Ops 3, Grand Theft Auto V, Call of Duty: Advanced Warfare, Halo 5: Guardians, Fallout 4.
- **PC** - The Sims 4, Fallout 4, Farming Simulator 2015, Grand Theft Auto V, The Elder Scrolls Online.

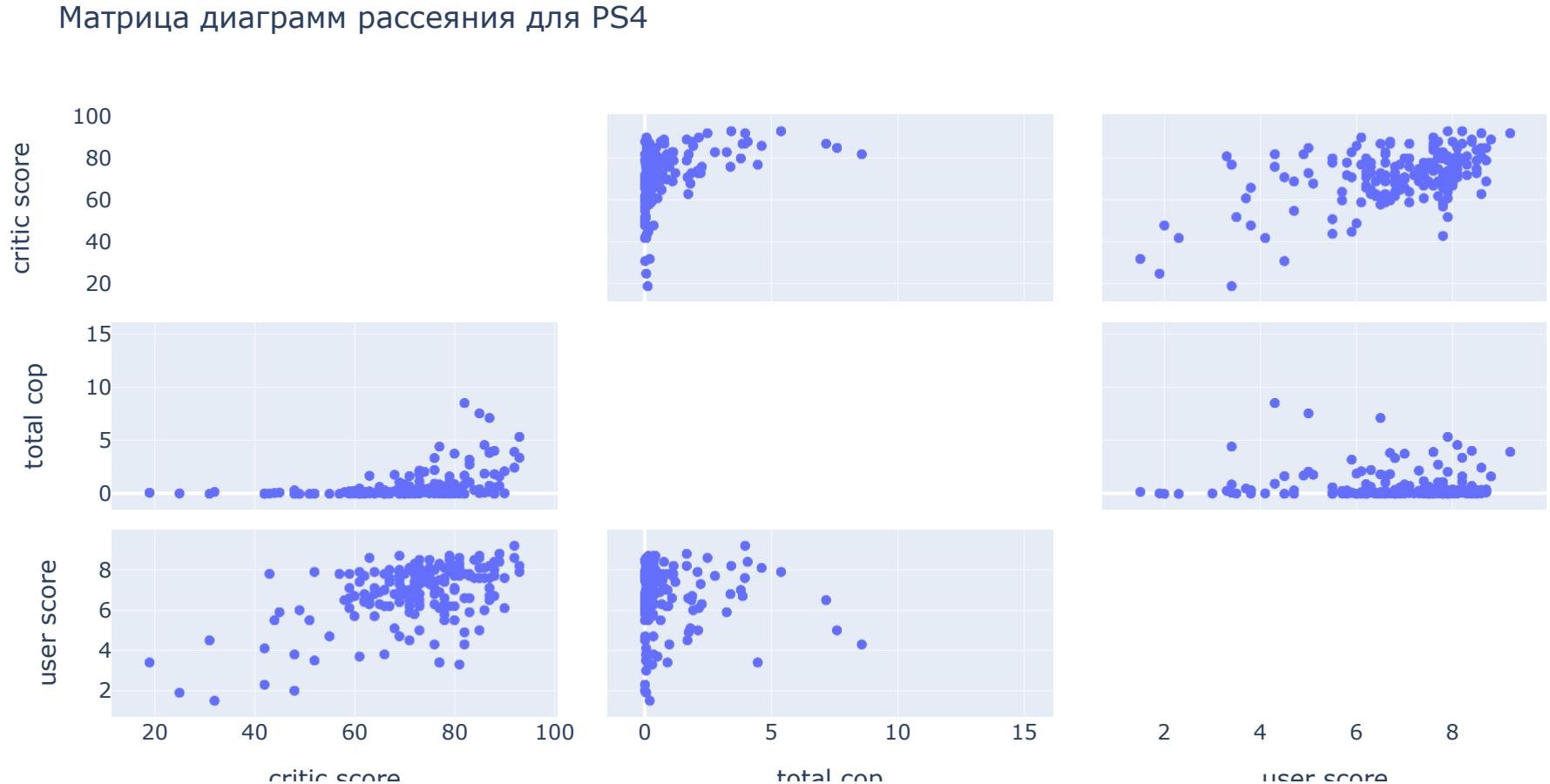
In [75]:
```
#для будущего исследования прогрессии продаж игры в разные годы нужно что-то такое
data.groupby(by='year_of_release')['name'].nunique().plot();
```



3.10 Посмотрю, как влияют на продажи внутри одной популярной платформы отзывы пользователей и критиков

Рассмотрю данные за полный 2015 год, за исключением ожидающих рейтинга (исключу флаг tbd). Выберу платформу PS4, так как она актуальна и часто в топе.

In [76]:
```
s_matrix_plt('PS4')
```



In [77]:
```
plt_corr('PS4')
```

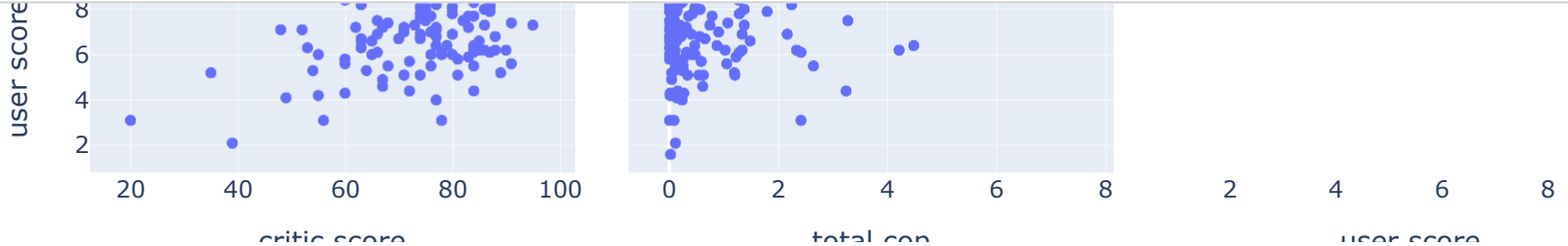
Out[77]:

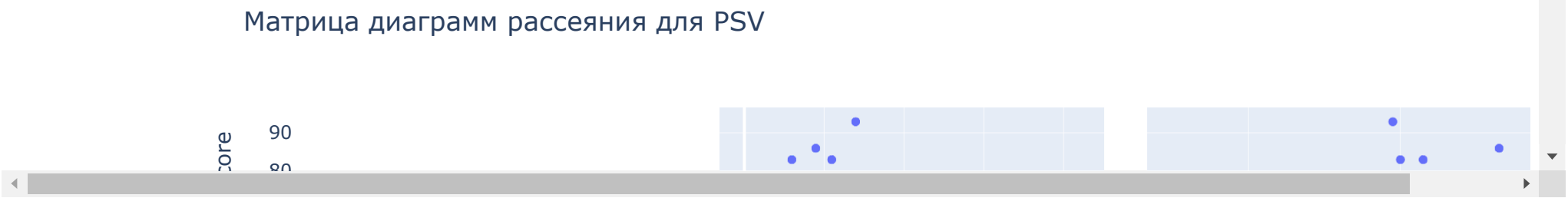
	critic_score	total_cop	user_score
critic_score	1.000000	0.392738	0.533330
total_cop	0.392738	1.000000	-0.059738
user_score	0.533330	-0.059738	1.000000

У общего объёма продаж у PS4 сильная связь с отзывами критиков, а отзывы пользователей не имеют связи или её характер более сложный.

3.11 Соотнесу выводы с продажами игр на других платформах

In [78]:
```
list_plt_matrix = ['PC', 'XOne', 'PSV', 'WiiU', '3DS']#выбрал другие платформами, которые будут в 2017
for i in list_plt_matrix:
    s_matrix_plt (i)
```





In [79]:
```
#Посмотрю корреляцию для PC.
plt_corr('PC')
```

Out[79]:

	critic_score	total_cop	user_score
critic_score	1.000000	0.377332	0.482804
total_cop	0.377332	1.000000	0.150381
user_score	0.482804	0.150381	1.000000

У PC связь общего объёма продаж с отзывами критиков сильней, чем с отзывами пользователей. Отзывы пользователей влияют на общий объем продаж, но не сильно.

In [80]:
```
#Посмотрю корреляцию для XOne.
plt_corr('XOne')
```

Out[80]:

	critic_score	total_cop	user_score
critic_score	1.000000	0.425068	0.459443
total_cop	0.425068	1.000000	-0.041467
user_score	0.459443	-0.041467	1.000000

У XOne влияние на объёмы продаж отзывов критиков сильное, а у отзывы пользователей не имеют связи или её характер более сложный.

In [81]:
```
#Посмотрю корреляцию для PSV.
plt_corr('PSV')
```

Out[81]:

	critic_score	total_cop	user_score
critic_score	1.000000	0.002749	0.454487
total_cop	0.002749	1.000000	0.078335
user_score	0.454487	0.078335	1.000000

У PSV критики слабо влияют на объёмы продаж, с отзывами юзеров больше связи, но и низкие значения.

In [82]:
```
#Посмотрю корреляцию для WiiU
plt_corr('WiiU')
```

Out[82]:

	critic_score	total_cop	user_score
critic_score	1.000000	0.325674	0.678858
total_cop	0.325674	1.000000	0.363519
user_score	0.678858	0.363519	1.000000

У WiiU отзывы критиков и пользователей влияют на объём продаж, при этом отзывы пользователей больше.

In [83]:
```
#Посмотрю корреляцию для 3DS
plt_corr('3DS')
```

Out[83]:

	critic_score	total_cop	user_score
critic_score	1.000000	0.177575	0.791853
total_cop	0.177575	1.000000	0.198796
user_score	0.791853	0.198796	1.000000

У 3DS отзывы критиков и пользователей почти одинаково влияют на объём продаж, но не сильно.

После исследования корреляции за 2015-2016, выводы изменились, но тенденция, что отзывы критиков слабо влияют на объемы продаж осталась. У PC и WiiU слабая связь объемов продаж с user_score. У 3DS очень слабая связь с отзывами пользователей.

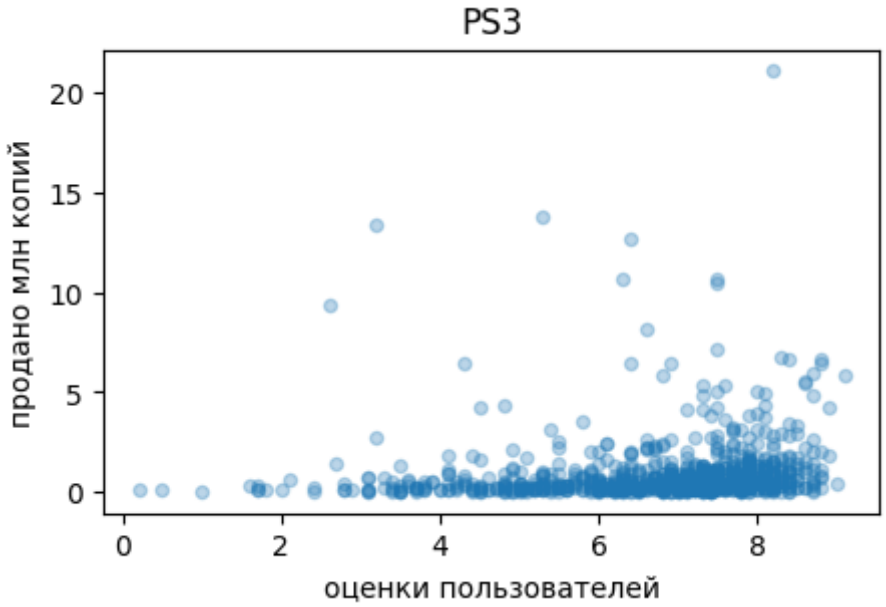
Общий вывод по всем актуальным платформам (список actual_plfm) для 2015 года представлю в виде таблицы:

In [84]:
```
conclusion_corr = [['PS4', '++', '-'], ['PC', '++', '+'], ['XOne', '++', '+'],
                   ['PSV', '-', '+'], ['WiiU', '+', '++'], ['3DS', '+', '+']]
conclusion_corr_name = pd.DataFrame(conclusion_corr,
                                    columns=['platform', 'critic_score', 'user_score'])
display(conclusion_corr_name)
##'++' явная связь
##'+' менее явная связь
##'-' нет связи
```

	platform	critic_score	user_score
0	PS4	++	-
1	PC	++	+
2	XOne	++	+
3	PSV	-	+
4	WiiU	+	++
5	3DS	+	+

На объём продаж для топовых платформ PS4, PC, XOne более влиятельны отзывы критиков, а для портативных консолей большее влияние имеют отзывы пользователей. Стоит учитывать, что по некоторым отзывам пользователей рейтинг ожидается (tbd).

In [85]:
```
top_2_data = data.query('platform in ["3DS", "WiiU", "PS3", "PC"]')
for platform in top_2_data['platform'].unique():
    data\
    .loc[data['platform'] == platform]\
    .plot(kind='scatter', x='user_score', y='total_cop',
          alpha=0.3, figsize=(5,3),
          xlabel='оценки пользователей', ylabel='продано млн копий', title=platform)
```



3.12 Посмотрю на общее распределение игр по жанрам, какие прибыльные, выделяются ли жанры с высокими и низкими продажами

Краткое описание жанров:
Shooter - стрелялка, как правило с линейным прохождением.
Role-Playing (RPG) - игра, в которой игроку предоставляется выбор роли: хороший, злой, маг, воин и т.д. С дальнейшей возможностью усиления выбранных ветвей развития.
Action - боевик, в котором важна реальная скорость реакции игрока или владение компьютерной мышью (элементами управления).

Sports - симулятор какого-то вида спорта.
Fighting - драки, с наличием разных персонажей, у которых есть ультимативная способность (супер удар).
Racing - гонки в любой среде (космос, подводой, на трассе и т.п.), физика перемещения может быть реальной или выдуманной.
Simulation - игра-имитация реального жизненного процесса с течением времени.
Platform - основа жанра - горизонтальное и вертикальное перемещение игрока, много прыжков по платформам.
Misc - казуальная игра, с простыми правилами, предназначенная для широкой аудитории, в большинстве своём подходит для вечеринок.
Strategy - игра, в которой для победы игроку нужно принимать стратегические решения. В управление игрока предоставляется группа юнитов.
Adventure - игра с упором на приключение, решение загадок, исследование пространства.
Puzzle - логическая игра, игроку нужно не стандартно мыслить, владеть интуицией.

In [86]:
```
ganre_count = good_data.pivot_table(index=['genre', 'year_of_release'],
                                    values=['ads_plt'],
                                    aggfunc=['count']).reset_index()
ganre_count.columns = ['genre', 'year_of_release', 'ads_plt']
```

In [87]:
```
ganre_count
```

Out[87]:

	genre	year_of_release	ads_plt
0	Action	2015.0	253
1	Action	2016.0	178
2	Adventure	2015.0	54
3	Adventure	2016.0	56
4	Fighting	2015.0	21
5	Fighting	2016.0	16
6	Misc	2015.0	39
7	Misc	2016.0	32
8	Platform	2015.0	13
9	Platform	2016.0	15
10	Puzzle	2015.0	6
11	Puzzle	2016.0	1
12	Racing	2015.0	18
13	Racing	2016.0	24
14	Role-Playing	2015.0	78
15	Role-Playing	2016.0	54
16	Shooter	2015.0	34
17	Shooter	2016.0	47
18	Simulation	2015.0	15
19	Simulation	2016.0	18
20	Sports	2015.0	59
21	Sports	2016.0	48
22	Strategy	2015.0	16
23	Strategy	2016.0	13

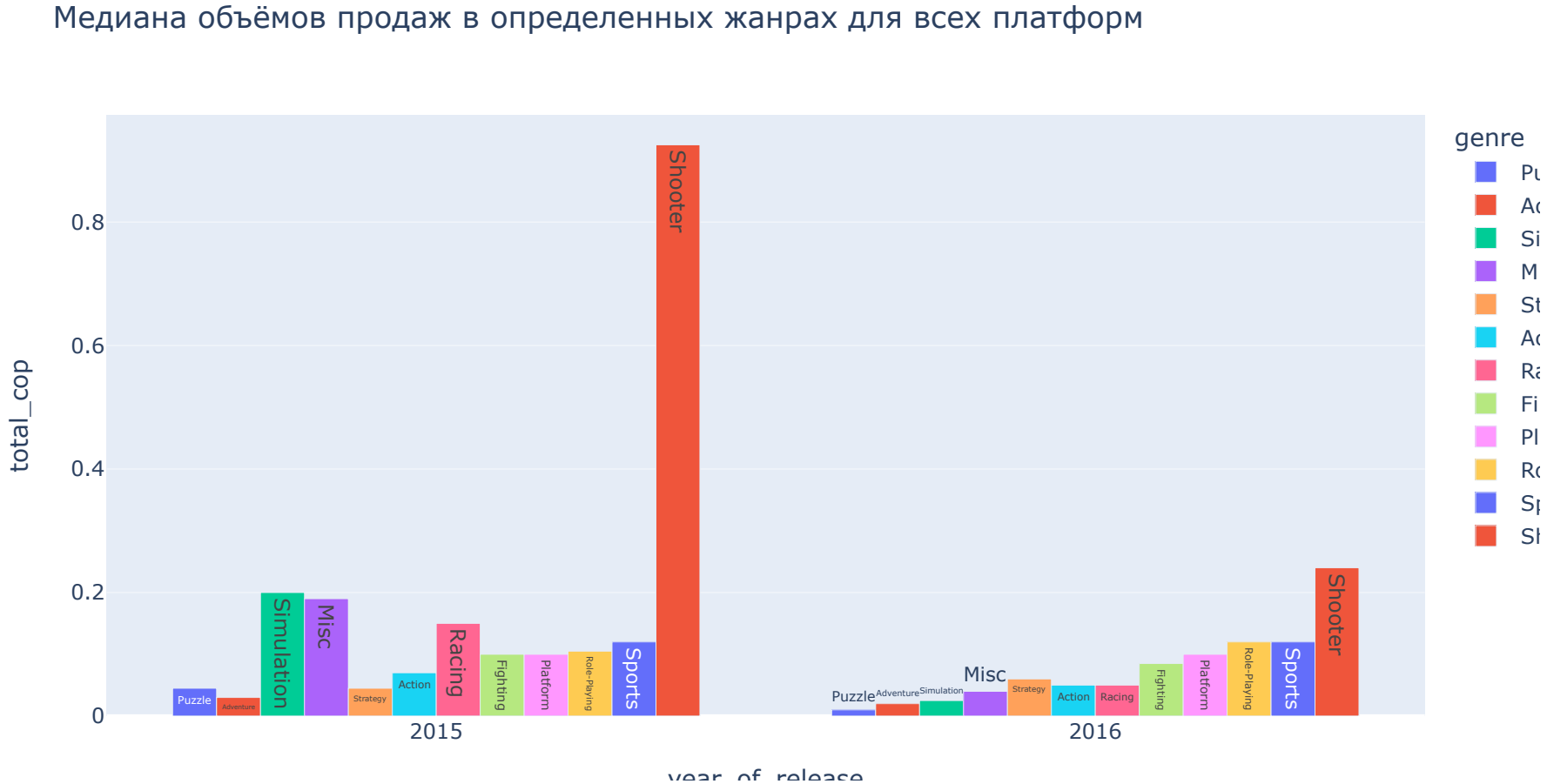
In [88]:
```
px.bar(ganre_count.sort_values(by=['ads_plt'], ascending=[False]),
       x = 'year_of_release',
       y='ads_plt',
       color='genre',
       text='genre',
       barmode = 'group',
       title = 'Количество релизов в определенных жанрах для всех платформ')
```

Количество релизов Action на первом месте в 2016, второе место у Adventure, на третьем RPG. Sports на четвертом, а Shooter на пятом месте.

По количеству релизов Action занимает первое место, в 2014, 2015 году заметен рост. Второе место занимают RPG (по тексту далее такая аббревиатура для Role-Playing), с небольшим ростом. На третьем месте Adventure, у которых в 2015 году снизилось количество релизов. Все остальные жанры в 2015 году немного подросли по количеству релизов (кроме Puzzle, количество не изменилось), самый заметный рост у Strategy и Fighting. Меньше всего игр выпускают в жанре: Puzzle, Platform, Simulation.

In [89]:
```
p_ganre_median = good_data.pivot_table(index=['genre', 'year_of_release'],
                                       values=['total_cop'],
                                       aggfunc=['median']).reset_index()
                                       #смотрю медиану, чтобы снизить влияние выбросов
                           #query('2016 > year_of_release >= 2014')\
p_ganre_median.columns = ['genre', 'year_of_release', 'total_cop']
```

In [90]:
```
px.bar(p_ganre_median.sort_values(by=['total_cop']),
       x = 'year_of_release',
       y='total_cop',
       color='genre',
       text='genre',
       barmode = 'group',
       title = 'Медиана объёмов продаж в определенных жанрах для всех платформ')
```



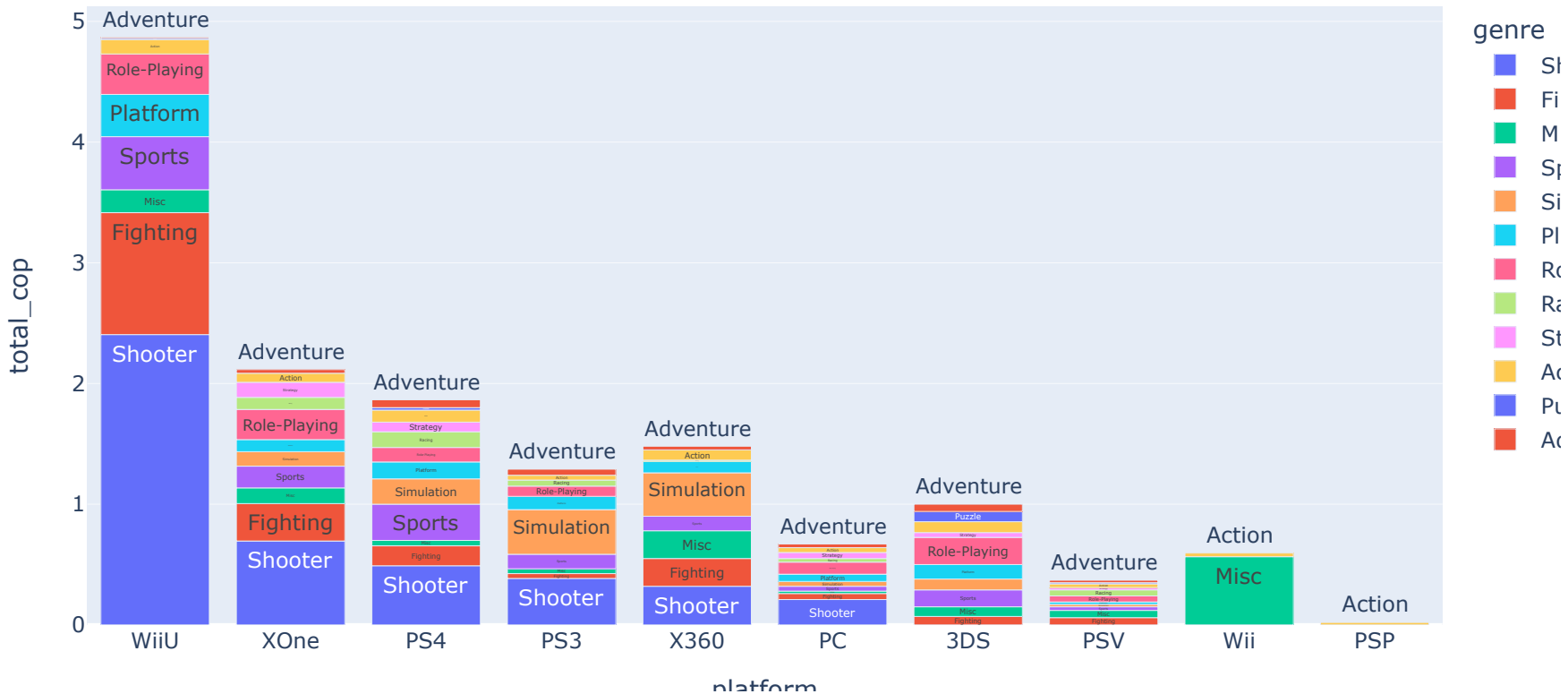
Медиана продаж 2016 на первом месте у Shooter, второе Sports, третье RPG. Сильные падения у Simulation, Misc, Racing. Adventure не в топ-7. Action на 7 месте. Fighting на 5 месте.

В 2015 году самый продаваемый жанр Shooter, показывает восходящий тренд (почти 1.5 раза). На втором месте Racing, с очень не большим ростом. На третьем месте Simulation, без изменения медианы количества проданных копий. Сильное падение в 2015 году у Platform (в 9.5 раз), у Sports (в 3.8 раза) и у Action (в 2.4 раза). У всех остальных жанров небольшое падание или небольшой рост, как у Fighting и RPG. Меньше всего медиана в 2015 году у Adventure, Puzzle, Strategy.

Посмотрю на популярность жанров на разных платформах.

In [91]:
```
pltf_genre = good_data.pivot_table(index=['platform', 'genre'],
                                   values=['total_cop'],
                                   aggfunc=['median']).reset_index()
                                   #смотрю медиану, чтобы снизить влияние выбросов
pltf_genre.columns = ['platform', 'genre', 'total_cop']
```

In [92]:
```
px.bar(pltf_genre.sort_values(by = 'total_cop', ascending = False),
       x = 'platform',
       y='total_cop',
       color='genre',
       text='genre',
       title = 'Сумма объёмов продаж по жанрам для актуальных платформ')
```

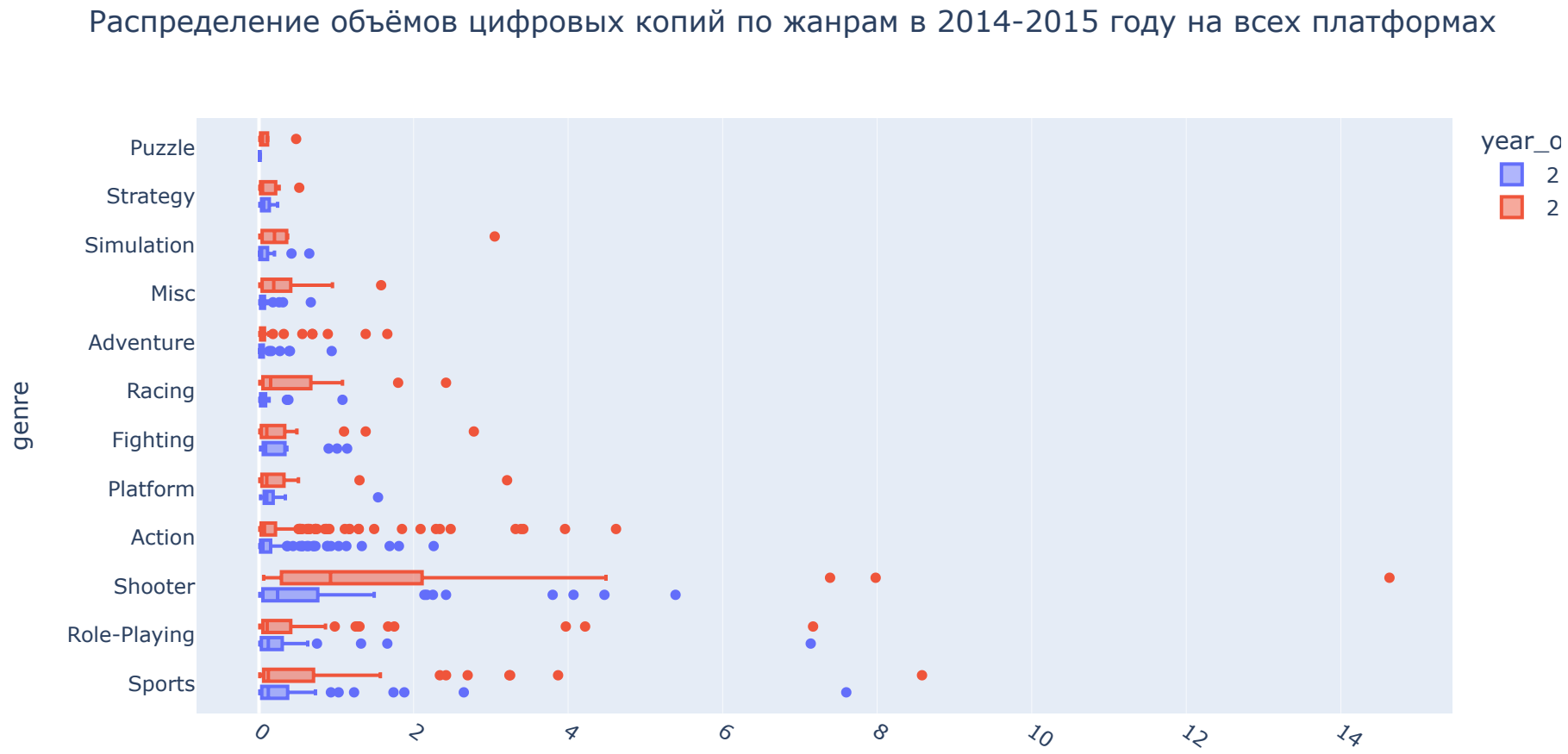


По популярности, в первую тройку вошли по объёмам продаж:

- На PS4 - Action, Shooter, Sports
- На 3DS - Action, RPG, Simulation
- На XOne - Shooter, Action, Sports

Если убрать популярные жанры Action, Shooter, Sports, то интересно выглядит RPG, так как на всех платформах этот жанр занимает средний объём.

In [93]:
```
px.box(good_data.sort_values(by=['year_of_release', 'total_cop'], ascending=[False, False]),
                y='genre',
                x='total_cop',
                hover_name = 'name',
                title = 'Распределение объёмов цифровых копий по жанрам в 2014-2015 году на всех платформах',
                color='year_of_release').update_xaxes(tickangle=35)
```



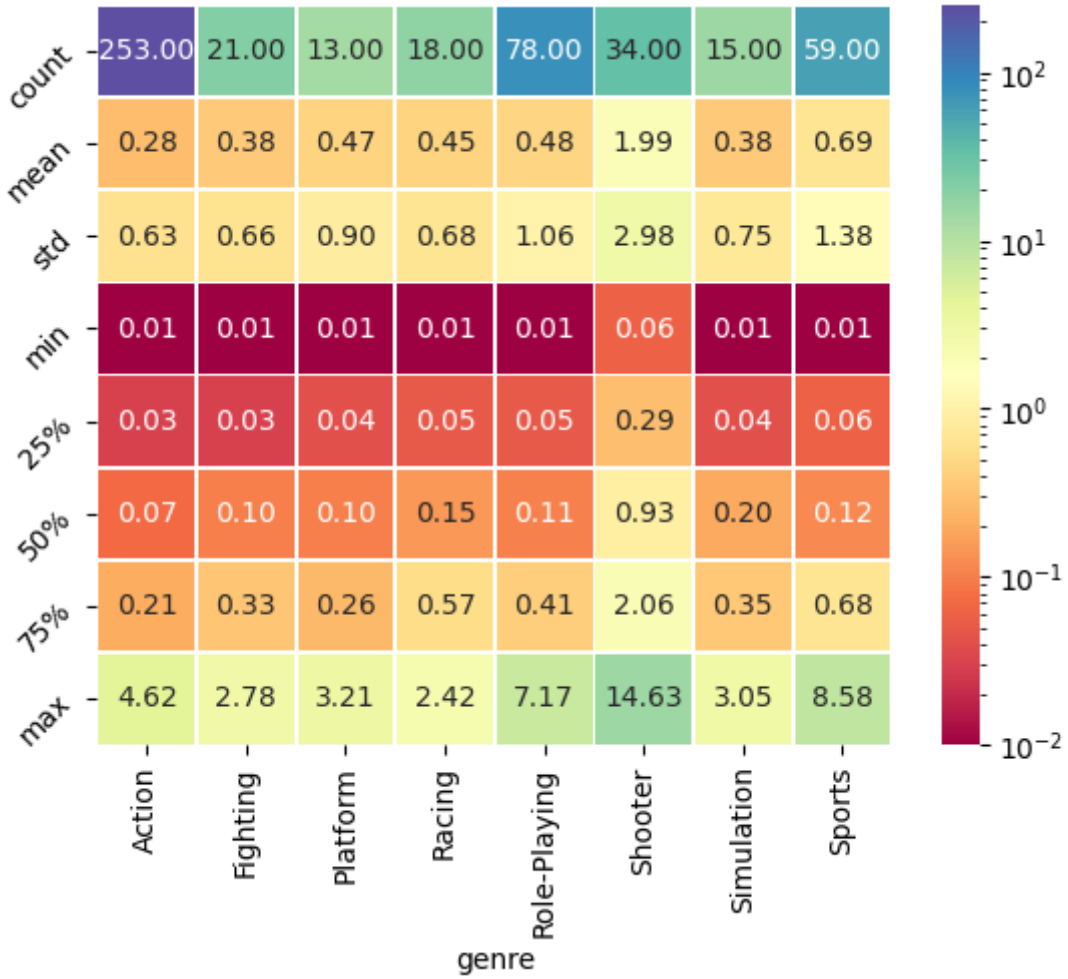
В 2016 поле 50% объёмов продаж шире всех у Shooter (+ самая высокая медиана), второе место у Sports, третье у Fighting. RPG на четвертом месте. У Action шестое, но лучше чем у Adventure.

Самые высокие выбросы за 2014-2015 годы у Shooter, Action, RPG, Sports - такими выбросами являются игры с высоким количеством выпущенных цифровых копий. Больше всего выбросов у Action, особенно в 2015 году. По количеству выбросов еще лидируют Sports и RPG, но у них количество меньше. Медиана у всех жанров от года к году снижается и только у Shooter сильно растёт, а RPG, Fighting и Racing немного растёт. Я заметил, что Platform в 2014 без выбросов, получается, все продажи попали в поле нормальных значений. У Fighting нет верхнего уса и две игры в поле высоких выбросов.

Охарактеризую данные по выделяющимся жанрам за 2015 год.

In [94]:
```
highlighted_genre = ['Shooter', 'Sports', 'Role-Playing', 'Action', 'Platform', 'Simulation', 'Fighting', 'Racing']
```

In [95]:
```
describe_g = good_data.query('year_of_release == 2015 and genre == @highlighted_genre')\
                      .pivot_table(index='ads_plt',
                                   columns='genre', values='total_cop').describe()
sns.heatmap(describe_g, vmax=1,
            cmap="Spectral",
            norm=LogNorm(),
            annot=True, annot_kws={'size':10}, fmt='.2f', linewidths=.5)
plt.yticks(rotation=45);
```



In [96]:
```
describe_g
```

Out[96]:

genre	Action	Fighting	Platform	Racing	Role-Playing	Shooter	Simulation	Sports
count	253.000000	21.000000	13.000000	18.000000	78.000000	34.000000	15.000000	59.000000
mean	0.284664	0.376190	0.465385	0.448333	0.482564	1.985588	0.377333	0.692203
std	0.627174	0.656563	0.895252	0.678253	1.055524	2.975623	0.752410	1.380827
min	0.010000	0.010000	0.010000	0.010000	0.010000	0.060000	0.010000	0.010000
25%	0.030000	0.030000	0.040000	0.050000	0.050000	0.290000	0.040000	0.060000
50%	0.070000	0.100000	0.100000	0.150000	0.105000	0.925000	0.200000	0.120000
75%	0.210000	0.330000	0.260000	0.567500	0.407500	2.062500	0.355000	0.680000
max	4.620000	2.780000	3.210000	2.420000	7.170000	14.630000	3.050000	8.580000

Самый лучший первый квартиль (25%) у Shooter, на втором месте Racing. Самый высокий третий квартиль у Shooter, Racing и Sports. Получается, что наилучший МКР (IQR) у Shooter и Racing, значит 50% цифровых копий продают между этим значениями. Выше всех медиана у Shooter, на втором месте Racing, на третьем Fighting. Наименьшее стандартное отклонение у Action, значит не смотря на выбросы, данные плотно сконцентрированы вокруг среднего значения. У Fighting и Racing тоже низкое значение стандартного отклонения от среднего.

4 Составлю портрет пользователя каждого региона, NA, EU, JP

4.1 Самые популярные платформы (топ-6), опишу различия в долях продаж

Изучу данные только за 2015 год, так как с 2014 года заметил снижение количества релизов и объёмов продаж. По платформам сделаю топ-6 (все актуальные платформы для 2017 года), а по жанрам топ-5. Добавил данные по другим регионам для полноты картины.

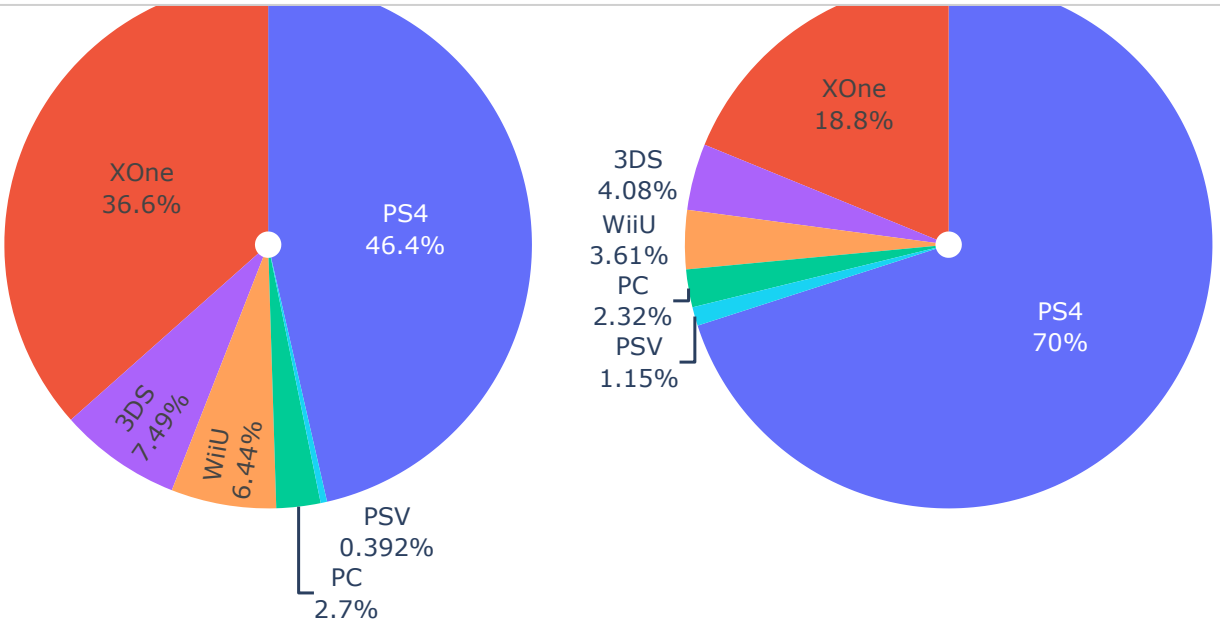
In [97]:
```
good_data_pivot = good_data.query('platform == @actual_plfm').pivot_table(index = ['platform'],
                                        values=['na_sales', 'eu_sales', 'jp_sales', 'other_sales'],
                                        aggfunc = ['sum']).reset_index()
good_data_pivot.columns = ['platform', 'eu_sales', 'jp_sales', 'na_sales', 'other_sales']
```

До ревью у меня была логика такая, что с 2012 года падение релизов и продаж, медианы в 2014 лучше, чем в 2015, а значит в 2016 всё будет похуже. Далее я убирал 2016, так как думал опасно строить предположения на 2017 из предположений 2016. Но сейчас в переменной отфильтрованы 2015-2016 и я просмотрю смотрю тенденции.

In [98]:
```
labels = good_data_pivot['platform']
specs = [[{'type':'domain'}, {'type':'domain'}], [{'type':'domain'}, {'type':'domain'}]]
name_p_h = 'выбранный период'

fig = make_subplots(rows=2, cols=2, specs=specs, vertical_spacing=0.15,
                    subplot_titles=['Европа', 'Япония', 'Северная Америка', 'Другие регионы'])
fig.add_trace(go.Pie(labels=labels, values=good_data_pivot['eu_sales'], name=name_p_h),1, 1)
fig.add_trace(go.Pie(labels=labels, values=good_data_pivot['jp_sales'], name=name_p_h),1, 2)
fig.add_trace(go.Pie(labels=labels, values=good_data_pivot['na_sales'], name=name_p_h),2, 1)
fig.add_trace(go.Pie(labels=labels, values=good_data_pivot['other_sales'], name=name_p_h),2, 2)


fig.update_traces(textinfo='percent+label', hole=.05, hoverinfo='label+percent+name',
                  textposition='auto',legendgrouptitle_text='platform')
fig.update_layout(height=800, width=800,
                  title_text='Использование платформ в разных регионах', title_x=0.5)
fig.update_annotations(align='right', yshift=10)
fig.show()
```



- Европа - PS4 на первом месте, с большим отрывом XOne, третье место с большим отрывом у PC.
- Япония - 3DS на первом месте, на втором с большим отрывом PS4, третье место с небольшим отрывом у PSV.
- Северная Америка - PS4 на первом месте, с небольшим отрывом на втором месте XOne, третье место с большим отрывом у 3DS.

Опишу результат для первой тройки платформ в каждом регионе:

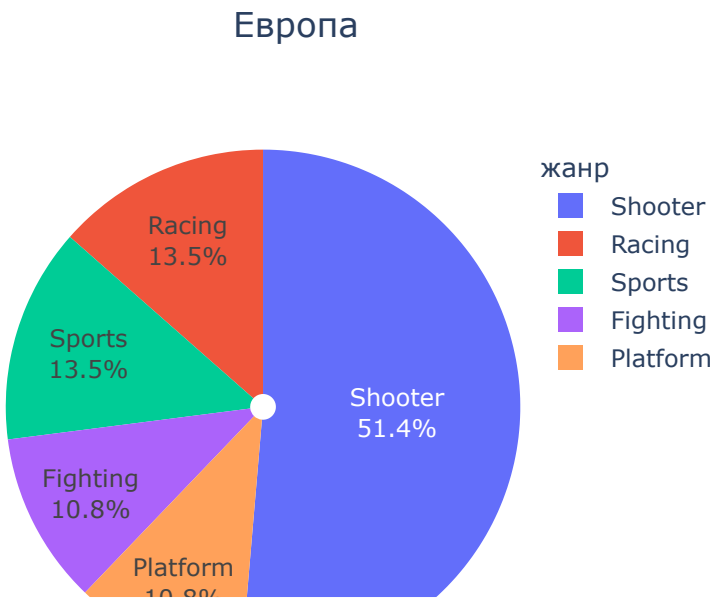
- Европа - PS4 на первом месте, с большим отрывом XOne, третье место с большим отрывом у PC.
- Япония - 3DS на первом месте, на втором с большим отрывом PS4, третье место с небольшим отрывом у PSV.
- Северная Америка - PS4 на первом месте, с небольшим отрывом на втором месте XOne, третье место с большим отрывом у WiiU.
- Другие регионы - первое место у PS4, на втором с очень большим отрывом XOne, на третьем месте с большим отрывом WiiU.

Примечание: по Японии нет данных в good_data_2015 об объёмах продаж PC. При этом можно сделать вывод, что в Японии более предпочитают портативные консоли своего производителя Nintendo, но и поддерживают своего производителя Sony, но в значительно меньшей доле.

4.2 Самые популярные жанры (топ-5), поясню разницу

In [99]:
```
g_no_res_index = good_data.query('platform == @actual_plfm').pivot_table(index = ['genre'],
                                    values=['eu_sales', 'jp_sales', 'na_sales', 'other_sales'],
                                    aggfunc = ['median'])#использую медиану, чтобы снизить влияние выбросов
g_no_res_index.columns = ['Европа', 'Япония', 'Северная Америка', 'Другие регионы']
```

In [100]:
```
pie_top5 (g_no_res_index)
```



- Европа - первое место с большим отрывом у Shooter, на втором месте Racing и Sports, третье место делят Fighting и Simulation.
- Япония - первое место с небольшим отрывом у RPG, на втором месте Fighting, на третьем Misc.
- Северная Америка - первое место с большим отрывом у Shooter, на втором месте Platform, на третьем Fighting.

Опишу результат по медиане для первой тройки из топ-5 жанров в каждом регионе на 2015 год:

- Европа - первое место с большим отрывом у Shooter, на втором месте Racing, третье место делят Fighting и Simulation.
- Япония - первое место с небольшим отрывом у RPG, на втором месте Puzzle, на третьем Misc.
- Северная Америка - первое место с большим отрывом у Shooter, на втором месте Platform, на третьем Fighting.
- Другие регионы - первое место с большим отрывом у Shooter, на втором месте Platform, на третьем Fighting.

> Примечание: по Японии нет данных в good_data_2015 об объёмах продаж PC. Если жанр занимает первое место, то можно считать этот жанр в регионе самым окупаемым, так как на него приходится большинство цифровых копий.

4.3 Влияет ли рейтинг ESRB на продажи в отдельном регионе

- Описание рейтингов:

- UKW - нет информации в датасете

- Актуальное на 2014-2016 год:

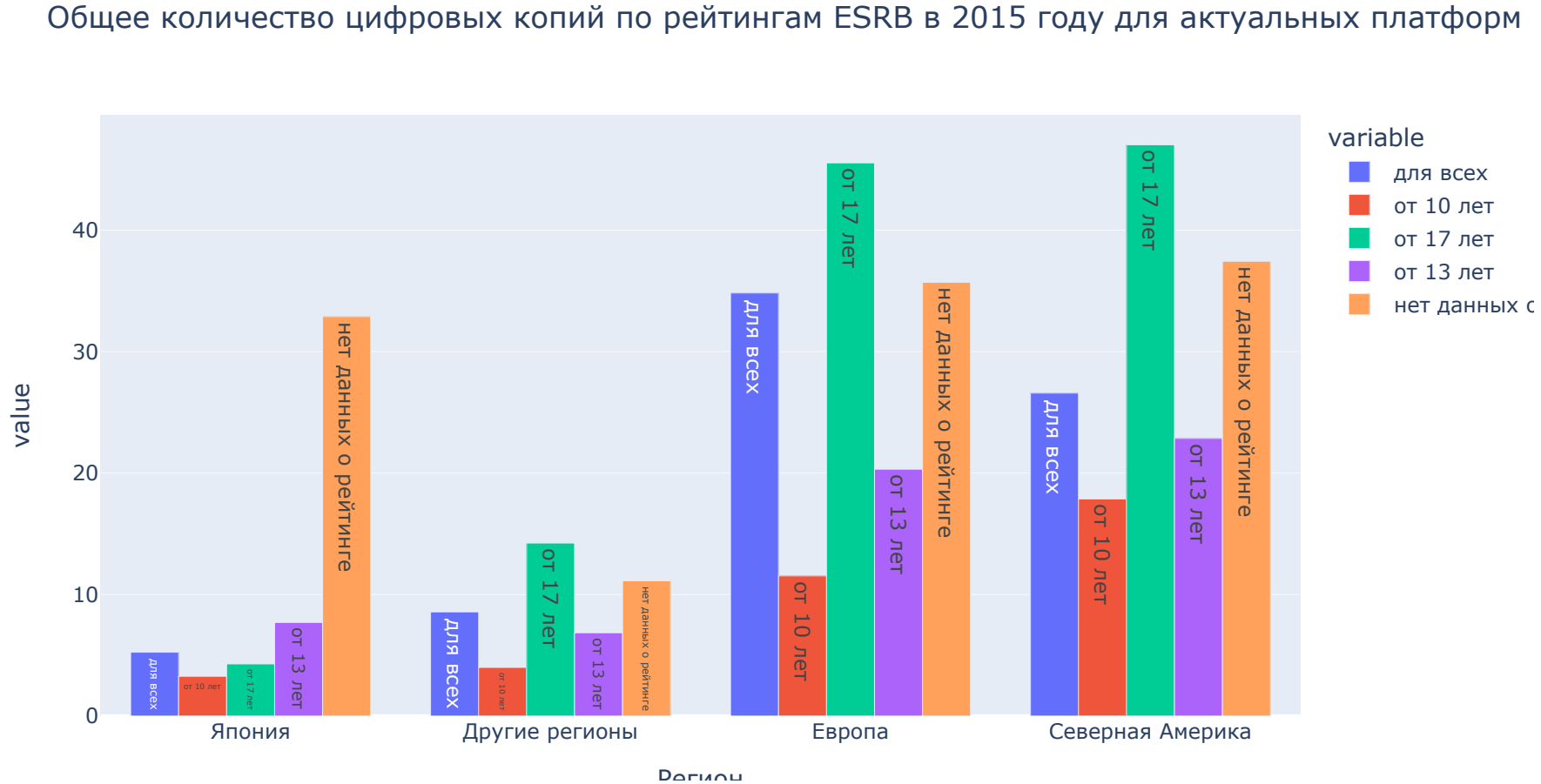
- M - Для взрослых от 17 лет
- T - Подросткам от 13 лет
- E - Для всех
- E10+ - Для всех от 10 лет и старше

- В исследуемом периоде не встречаются:

- K-A - Для всех, рейтинг 1994-1996 года
- AO - Только для взрослых от 18 лет
- EC - Для детей младшего возраста от 3 лет
- RP - Рейтинг ожидается

In [101]:
```
rating_region = good_data.pivot_table(columns = ['rating'],
                                      values=['na_sales', 'eu_sales', 'jp_sales', 'other_sales'],
                                      aggfunc = ['sum']).sort_values(by=('sum', 'M')).reset_index()
#['E', 'E10+', 'M', 'T', 'UKW'] - порядок рейтингов для перевода
rating_region.columns = ['Регион', 'для всех', 'от 10 лет', 'от 17 лет', 'от 13 лет', 'нет данных о рейтинге']
rating_region['Регион'] = rating_region['Регион'].map({'jp_sales':'Япония','other_sales':'Другие регионы',
                                                       'eu_sales':'Европа', 'na_sales':'Северная Америка'})
```

In [102]:
```
list_region = ['для всех', 'от 10 лет', 'от 17 лет', 'от 13 лет', 'нет данных о рейтинге']
px.bar(rating_region,
       x = 'Регион',
       y=list_region,
       barmode = 'group',
       text = 'variable',
       title = 'Общее количество цифровых копий по рейтингам ESRB в 2015 году для актуальных платформ')
```



- Европа - больше всего продаж у рейтинга от 17 лет, на втором месте без ограничения возраста, на третьем от 13 лет.
- Япония - на первом месте от 13 лет, на втором без ограничения возраста, а на третьем от 17 лет.
- Северная Америка - первое место от 17 лет, второе место без ограничения возраста, третье место от 13 лет.

Опишу результат по общим продажам цифровых копий для первой тройки из топ-5 жанров в каждом регионе на 2015 год:

- Для всех регионов - рейтинг UKW (нет данных о рейтинге возраста) не учитываю (исследование ниже), так как с этой отметкой много свойств и это может искажать картину оценки.
- Европа - больше всего продаж у рейтинга от 17 лет, на втором месте без ограничения возраста, на третьем от 13 лет.
- Япония - на первом месте от 13 лет, на втором без ограничения возраста, а на третьем от 10 лет.
- Северная Америка - первое место от 17 лет, второе место без ограничения возраста, третье место от 13 лет.
- Другие регионы - первое место от 17 лет, второе место без ограничения возраста, третье место от 13 лет.

4.4 Исследование UKW

In [103]:
```
good_data.shape
#всего строк в датасете
```

Out[103]:
```
(1108, 15)
```

In [104]:
```
good_data.query('rating == "UKW" and year_of_release == 2015')\
         .pivot_table(index=['rating'],
                      values=['na_sales', 'eu_sales', 'jp_sales', 'other_sales'],
                      aggfunc = ['count'])
#на каждую колонку региона влияет пропуск UKW
```

Out[104]:

	count			
	eu_sales	jp_sales	na_sales	other_sales
rating				
UKW	291	291	291	291

In [105]:
```
good_data.query('rating == "UKW" and year_of_release == 2015')\
        .pivot_table(index=['genre'],
                     values=['na_sales', 'eu_sales', 'jp_sales', 'other_sales'],
                     aggfunc = ['sum'])
#сумма продаж по жанрам в колонках с пропусками
```

Out[105]:

	sum			
	eu_sales	jp_sales	na_sales	other_sales
genre				
Action	4.79	11.51	4.23	1.34
Adventure	1.37	0.87	1.47	0.40
Fighting	0.30	0.23	0.22	0.09
Misc	0.73	1.55	0.03	0.04
Platform	0.08	0.09	0.09	0.02
Puzzle	0.10	0.45	0.06	0.01
Racing	2.02	0.14	0.80	0.46
Role-Playing	2.87	2.74	3.60	1.03
Shooter	15.02	0.68	17.70	5.24
Simulation	0.59	0.04	0.53	0.17
Sports	0.18	0.14	0.04	0.03
Strategy	0.19	0.13	0.14	0.02

In [106]:
```
len(good_data.query('rating == "UKW" and eu_sales==0 and na_sales > 0 \
                    and other_sales==0 and jp_sales==0 and year_of_release == 2015'))
#количество строк с продажами в na_sales регионе и отсутствием продаж в других регионах
```

Out[106]: 4

In [107]:
```
len(good_data.query('rating == "UKW" and eu_sales > 0 and na_sales==0 \
                     and other_sales==0 and jp_sales==0 and year_of_release == 2015'))
#количество строк с продажами в eu_sales регионе и отсутствием продаж в других регионах
```

Out[107]: 22

In [108]:
```
len(good_data.query('rating == "UKW" and eu_sales==0 and na_sales==0 \
                     and other_sales==0 and jp_sales > 0 and year_of_release == 2015'))
#количество строк с продажами в jp_sales и отсутствием продаж в других регионах
```

Out[108]: 160

In [109]:
```
len(good_data.query('rating == "UKW" and eu_sales==0 and na_sales==0 \
                     and other_sales > 0 and jp_sales==0 and year_of_release == 2015'))
#количество строк с продажами в other_sales регионе и отсутствием продаж в других регионах
```

Out[109]: 0

Всего в диаграмме 'цифровых копий по рейтингам ESRB в 2015' визуализированы 1604 строки, из которых UKW составляют 727 строк. Так как ESRB действует в регионе США и Канады, то можно предположить, что рейтинг можно применять только к параметру na_sales, но при проверке продаж на уровне жанров везде есть продажи, а значит с таким рейтингом игры продаются и в других регионах (или запись данных аномальна). Попробую рассмотреть другие регионы. Если только в одном регионе есть продажи игры, а в других нет, то возможно для этой игры существует свой региональный рейтинг. Для jp_sales получилось 420 строк, для eu_sales 60 строк, для na_sales 15 строк, для other_sales 0 строк в которых есть продажи для указанного региона и нет продаж для других регионов. Возможно эта часть строк должна быть исключена из анализа по рейтингам, но стоит проверить с представителем данных, что отсутствие продаж заполнено верно.

In [110]:
```
print ('В любоме случае', 2 + 19 + 130 + 0, 'строка с UKW. А эта сумма меньше общего количества UKW в 2015 году.')
```

```
В любоме случае 151 строка с UKW. А эта сумма меньше общего количества UKW в 2015 году.
```

Система рейтингов существует в игровой индустрии уже давно, к примеру, в 39 странах Европы с 2003 используют PEGI (Pan European Game Information). А в США и Канаде в 1994 ввели рейтинг ESRB (Entertainment Software Rating Board). В Японии CERO с 2002.

5 Проверка гипотез

alpha = 0.05

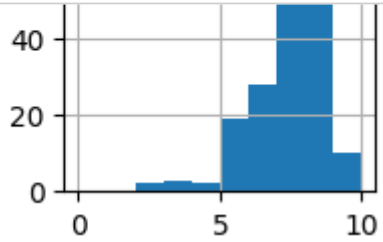
Я выбрал такое значение, так как это общепринятое значение. Но как я попытался разобраться почему общепринято:

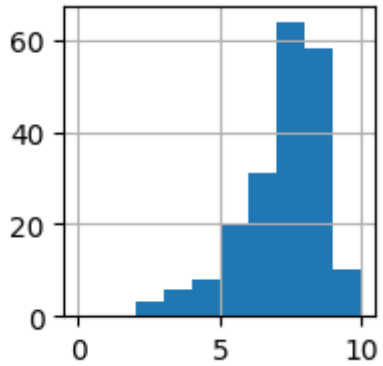
Альфа означает, насколько невероятными должны быть наблюдаемые результаты, чтобы отклонить нулевую гипотезу. Для итогов с 95—процентным уровнем вероятности значение Альфа равно 1 - 0,95 = 0,05. Я заметил, что величина p-value больше 0,05, ведёт в пределы двух сигм (двух стандартных отклонений), то есть в поле нормальных значений и высокой вероятности. А если я попал p-value в alpha или ниже, то я рискую найти изменения там, где их не было – это ошибка первого рода (тут я не уверен и возможно такая ошибка, только если я не верно поставил H0 гипотезу).

Нулевая гипотеза всегда 'равно' или 'одинаковые'. Всегда! - тоже пока запомнил так.

Выбрал ttest_ind по формулировке: специальный метод для проверки гипотезы о равенстве среднего двух генеральных совокупностей по взятым из них выборкам. Так как у меня данные ко всем значениям совокупности, то я посчитал её генеральной. Условие для использования ttest_ind - выборочные средние, которые получаются, если взять выборки одинакового размера из этой генеральной совокупности, должны быть нормально распределены.

In [111]:
```
for i in range (2):#две выборки
    data.loc[(data['user_score'].notna())]['user_score']\
                                .sample(200).hist(bins=10, #размер выборки 200
                                                  figsize = (2,2), range = (0,10));#range 10 по шкале user score
    plt.show()
    #проверю две случайные выборки, они нормально распределены
```





5.1 гипотиза 1: Средние пользовательские рейтинги платформ Xbox One и PC одинаковые.

H_0: Средние рейтинги (user_score) пользователей XOne = средним рейтингам (user_score) пользователей PC
H_a: Средние рейтинги (user_score) пользователей XOne ≠ средним рейтингам (user_score) пользователей PC

In [112]:
```
#xone_g1 = data[(data['platform'] == 'XOne') & (data['user_score'].notna())
#                                            & (data['year_of_release'] == 2015)]['user_score']
#pc_g1 = data[(data['platform'] == 'PC') & (data['user_score'].notna())
#                                        & (data['year_of_release'] == 2015)]['user_score']
```

In [113]:
```
xone_g1 = good_data[(good_data['platform'] == 'XOne') & (good_data['user_score'].notna())]['user_score']
pc_g1 = good_data[(good_data['platform'] == 'PC') & (good_data['user_score'].notna())]['user_score']
```

In [114]:
```
# results = вызов метода для проверки гипотезы
results = st.ttest_ind(
          xone_g1,
          pc_g1,
          equal_var = False#нет оснований предполагать, что дисперсии одинаковы, так как в выборках разное количество
)

alpha = .05#значение уровня значимости

print('p-value:', results.pvalue)#вывод значения p-value на экран
```

```
p-value: 0.2946308864003345
```

In [115]:
```
if results.pvalue < alpha:
    print('Отвергаем нулевую гипотезу')
else:
    print('Не получилось отвергнуть нулевую гипотезу')
```

```
Не получилось отвергнуть нулевую гипотезу
```

В изначальном условии задачи не было указано, за какой период брать выборки. Я решил, что для актуальности исследования будет достаточно взять крайний полный год - 2015. P-value 0.6 (или 6% - процентных пункта) больше Alpha 0.05, значит высока, вероятность случайно получить наблюдаемое в нулевой гипотезе событие.

Вероятность случайно получить такие же или бОльшие отличия между значениями при верной нулевой гипотезе равна почти 30%, поэтому нулевую гипотезу я не могу отвергнуть. Получается, что средние рейтинги пользователей XOne и PC равны с высокой долей вероятности.

5.2 гипотиза 2: Средние пользовательские рейтинги жанров Action (англ. «действие», экшен-игры) и Sports (англ. «спортивные соревнования») разные.

H_0: Средние рейтинги (user_score) жанра Action = средним рейтингам (user_score) жанра Sports
H_a: Средние рейтинги (user_score) жанра Action ≠ средним рейтингам (user_score) пользователей Sports

In [116]:
```
xone_g2 = good_data[(good_data['genre'] == 'Action') & (good_data['user_score'].notna())]['user_score']
pc_g2 = good_data[(good_data['genre'] == 'Sports') & (good_data['user_score'].notna())]['user_score']
```

In [117]:
```
# results = вызов метода для проверки гипотезы
results = st.ttest_ind(
          xone_g2,
          pc_g2,
          equal_var = False#нет оснований предполагать, что дисперсии одинаковы, так как в выборках разное количество
)

alpha = .05#значение уровня значимости

print('p-value:', results.pvalue)#вывод значения p-value на экран
```

```
p-value: 5.97163549920592e-10
```

In [118]:
```
if results.pvalue < alpha:
    print('Отвергаем нулевую гипотезу')
else:
    print('Не получилось отвергнуть нулевую гипотезу')
```

```
Отвергаем нулевую гипотезу
```

Вторую гипотезу проверил так же для 2015 года. Так как число 4.160552540931437 x (10 в степени -8) - это очень маленькое число и оно меньше Alpha 0.05, то я отвергаю нулевую гипотезу, так как вероятность получить наблюдаемое событие очень мала.

Вероятность случайно получить такие же или бОльшие отличия между значениями при верной нулевой гипотезе меньше уровня значимости в Alpha, поэтому нулевую гипотезу я отвергаю. Получается, что средние рейтинги пользователей Action и Sports не равны.

6 Напишу общий вывод

Исследование по задаче проекта
Магазин продаёт игры по всему миру.
Выбрать потенциально популярный продукт на 2017

По общему объёму продаж в 2015 самый перспективный рынок Северная Америка. В 2016 возможно на первое место (или будет очень рядом) выйдет Европейский. Возможно есть смысл отказаться от рекламы в Японии и использовать этот бюджет на освоении 36.6 доли XOne на рынке Северной Америки. Или можно 80% бюджета направить на основные группы, а 20% выделить на экспериментальные группы:

Основные группы

- Европа – ESRB от 17, жанр Shooter, PS4, важны мнения критиков

- Япония – ESRB от 13 лет, RPG, 3DS, рейтинги пользователей влиятельней,
- Северная Америка – ESRB от 17 лет, Shooter, PS4, важны мнения критиков

Эксперимент

- Европа – от 17, Sports (2 место медиана продаж), XOne
- Япония – ESRB от 13 лет, Fighting, PS4
- Северная Америка – ESRB от 17, Sports (с 4 места (с долей продаж 9.84%), но поле нормальных значений у него шире, чем у файтингов с 3 места(с долей продаж 13.1%)), XOne
  #на втором Platform, но они показывали резкое падение

Проверка гипотез
Описание выбора типа теста, альфы и теория в пункте 5. Выдвинутые гипотезы:
Средние пользовательские рейтинги платформ Xbox One и PC одинаковые

- – верно.

Средние пользовательские рейтинги жанров Action (англ. «действие», экшен-игры) и
Sports (англ. «спортивные соревнования») разные.

- – верно (тут я не менял гипотизу, а подвел значение под логику такого вопроса).

6.1 полный вывод

Предобработка
В ходе проведенной работы я привел все названия параметров к snake_case. Далее я преобразовал данные в некоторых колонках в другой тип, чтобы с ними было проще работать в построении диаграмм. В столбце user_score значение tbd (перед релизом игры такая аббревиатура указывается у данных которые еще на этапе сбора у юзеров, на момент релиза игра еще не доступна им, но может быть доступна критикам) вывел в отдельный параметр, чтобы при анализе корреляции исключить эти строки. Преобразование типа данных на datime в year_of_release не делал, так как не указан месяц, день. Шкалы оценки в user_score и critic_score не менял, так как в ходе исследования удостоверился, что это не влияет на результаты. Пропуски в rating заполнил временным значением - UKW (сокр.англ: неизвестно), для удобства выведения диаграммы. Пропуски user_score, critic_score не заполнял синтетическими значениями, так как они сформированы по механизму MNAR (Missing Not At Random) и я не могу их явно предсказать из информации датасета. В 'flag_tbd' оставил пропуски. Посчитал суммарные продажи во всех регионах и записал их в параметр total_cop с двумя знаками после запятой. Изучил 269 пропусков в параметре year_of_release (из них 2 пропуска совпадали с пропусками в genre и name) и удалил их, так как

анализ объёма цифровых копий по ним не влияет на общий ход исследования, данные по старым платформам (интересными были 17 пропусков по платформе PC, которые можно было заполнить руками, используя информацию из открытых источников). В боевой ситуации я лучше бы провел работу по исключению пропусков на уровне получения данных, возможно требуется изучение системы сбора/записи/выгрузки в/из БД данных по играм, для составления рекомендаций повышающих стабильность работы. Пропуски могут связаны и с неправильным объединением таблиц. После удаления пропусков в year_of_release, изменил тип данных на удобный для работы с диаграммами. Потери, после обработки не всех пропусков, составили менее 2%.

Исследование общее:
В датасете есть данные по релизу игр на разных платформах с 1980 г по 2015 (+ неполные данные за 2016). Рост количества релизов начинается с 1994, а в 2002 происходит быстрый рост, пик релизов в 2008, 2009. В 2012 резкое снижение количества релизов, то есть между изменениями прошло 10 лет. С 2016 по 2018 продолжается нисходящий тренд, исследую подробно крайний полный по данным год. Ситуацию может изменить появление инновации в игровой индустрии, которая изменит линию тренда на восходящий. За весь период лидеры по релизам: PS2, DS, PS3 – эти платформы принадлежат бренду Sony и Nintendo, а PC (персональный компьютер) на восьмом месте. Если посмотреть лидеров по релизам с 2012 (когда происходит резкое снижение), то первая тройка (два места из трёх у портативных консолей): PS3, PSV (самая лёгкая по весу), 3DS - бренды Sony и Nintendo сохраняют лидерство, а PC улучшает позиции и занимает шестое место. В этом топе на девятой позиции инновация от Nintendo – WiiU, которая позиционирует себя как портативная консоль (джойстик с экраном в одном корпусе), но с возможностью стационарного использования (вес 1600 грамм).

С 1994 по общим продажам лидируют платформы Sony. С 2001 в борьбу за рынок консолей вступает Microsoft, но и Nintendo удерживает хорошие позиции. PC занимает незначительный объём продаж во всем исследуемом периоде. Пики продаж у всех платформ в 2002, 2008, 2009, 2012. В 2008,2009 первое место по объёмам у Nintendo. С 2012 самым уверенным брендом по продажам выглядит Sony с: PS3, PS4 и PSV. При исследовании продаж, заметно, что чем больше продаж на старте платформы, тем больше её потенциал в достижении максимального пика, к примеру у PS4 больше продаж в первый год существования, чем у XOne. И далее каждый год игры PS4 продаются лучше, чем у XOne (но тут может играть фактор ожидания аудитории, пользователи больше хотят играть на PS4, чем на XOne). С 2012 многие платформы показывают нисходящий тренд, только PS3, 3DS, WiiU удерживают позиции и немного стремятся вверх. В 2013 выходит новое поколение приставок Sony PS4 и Microsoft XOne с резким восходящим трендом, и к 2015 году PS4 получает в два раза больше продаж, чем XOne. С 2013 остальные платформы показывают падение количества продаж. PC, PSV и WiiU в 2014 показывает небольшой рост, но к 2015 приходят с нисходящим трендом. На 2015 год топ-3 по объёмам продаж: PS4, XOne, 3DS.

Новое поколение платформы XOne вышло на 8 год существования X360, а PS4 выходит на 7 год существования PS3. После выхода нового поколения, поддержка старого продолжается 2-3 года. Продажи PS4 и XOne в нисходящем тренде по медиане продаж с 2013, не смотря на то, что на выходе этих платформ было много продаж. В 2015 все платформы в нисходящем тренде, кроме нового поколения Nintendo WiiU. Лидером по медиане цифровых копий в 2015 становится XOne, за ним идёт WiiU, а третье место делят между собой X360 и PS4. Можно предположить, что нисходящий тренд сохранится и в 2016.

В 2016 выпуск игр продолжаются на:

Nintendo - 3DS, Wii, WiiU
Sony - PS3, PS4, PSV
Microsoft - X360, XOne
Персональный компьютер – PC.
В среднем, основные лидеры рынка живут 10 лет 8 месяцев. Исключение PC. В данных до 1994 есть пропуски (возможно нет данных или они потеряны), но с этого года PC получает новые игры 23 года и будет получать дальше. Пики в 2003, 2009 и 2011 годы. С 2013 восходящий тренд.

Следующие новые поколения платформ представлены на рынке:
3DS - 6 лет, нисходящий тренд;
PS4 - 4 года, восходящий тренд;
PSV - 6 лет, нисходящий тренд;
WiiU - 5 лет, нисходящий тренд;
XOne - 4 года, восходящий тренд;
Можно сделать вывод, что у PS4 и XOne есть потенциал для роста и на рынке они будут представлены в 2017. WiiU в 2017 будет получать игры. Портативный PSV и 3DS тоже переходят в 2017.

Фокусировка на задаче проекта:
Задача проекта заключается в описании потенциально популярного продукта и планировании его рекламного бюджета на 2017. Датасет с неполными данными за 2016 год, лучше обновить данные, так как популярные игры AAA-класса и большинство интересных релизов происходят в конце года – ноябрь, декабрь или начале следующего года – январь, февраль. Я ориентировался на то, что 2016 в своей медиане по релизам и продажам будет ниже, чем 2015. По 2014,2015 полные данные, можно построить стратегию на 2017 по ним, а 25 декабря 2016 обогатить датасет и внести в стратегию корректировки, спрогнозированная картина будет не сильно измениться.

По медиане продаж 2014, первое место у PS4, второе у XOne, а на третьем WiiU. PC сразу за ними.
По медиане продаж 2015, первое место у XOne, второе у WiiU, а на третьем PS4.
По общему объёму продаж в 2014, 2015 лидирует PS4 - первое место, XOne - второе место, 3DS - третье место. PC занимает незначительный объём.

Посмотрел распределение объёмов продаж по релизам. Заметил релизы, которые продавались большим тиражом. На PS4 такие высокотиражные игры интересней всего, самую высокую точку занимает Call of Duty: Black Ops 3 в 2014 и Grand Theft Auto V в 2015. У XOne есть релизы этих же игр, но по объёму продаж в такие же годы, они занимают значительно меньшую позицию. Из портативных консолей много продаж у 3DS. На PC есть такие успехи, но их мало. Медиана продаж у всех от года к году падает, кроме WiiU у которой есть рост. Распределение объёмов продаж лучше остальных у PS4, XOne и WiiU, в которых 50% значений находятся на большем расстоянии от медианы, чем у других платформ.

Список ААА-релизов по объёмам продаж по каждой платформе на 2014, 2015:
PS4 - Call of Duty: Black Ops 3, Grand Theft Auto V, FIFA 16, Star Wars Battlefront (2015), Call of Duty: Advanced Warfare;
XOne - Call of Duty: Black Ops 3, Grand Theft Auto V, Call of Duty: Advanced Warfare, Halo 5: Guardians, Fallout 4;
PC - The Sims 4, Fallout 4, Farming Simulator 2015, Grand Theft Auto V, The Elder Scrolls Online.
Исследовал, как влияют на продажи внутри одной популярной платформы отзывы пользователей и критиков. Рассмотрел данные за полный 2015, за исключением ожидающих рейтинга (исключил строки с флагом tbd). Выбрал платформу PS4, так как она актуальна и часто в топе. У общего объёма продаж PS4 есть связь с отзывами критиков, а отзывы пользователей не имеют связи или её характер более сложный. Соотнес выводы с остальными платформами, на объём продаж для топовых платформ PC, XOne более влиятельны отзывы критиков, а для портативных консолей PSV, WiiU, 3DS большее влияние имеют отзывы пользователей.

Жанры:
Кратко описал особенности жанров в пункте исследования 3.12. По количеству релизов Action занимает первое место, в 2014, 2015 заметен рост. Второе место в 2014 у RPG (по тексту далее такая аббревиатура для Role-Playing), с небольшим ростом. На третьем месте Adventure, у которых в 2015 снизилось количество релизов. Все остальные жанры в 2015 немного подросли по количеству релизов (кроме Puzzle, количество не изменилось), самый заметный рост у Strategy и Fighting. Меньше всего игр выпускают в жанре: Puzzle, Platform, Simulation. В 2015 самый продаваемый жанр Shooter, показывает восходящий тренд (почти 1.5 раза). На втором месте Racing, с не большим ростом. На третьем месте Simulation, без изменения медианы продаж. Сильное падение в 2015 у Platform (в 9.5 раз), у Sports (в 3.8 раза) и у Action (в 2.4 раза). У всех остальных жанров небольшое падание или небольшой рост, как у Fighting и RPG. Меньше всего медиана продаж в 2015 у Adventure, Puzzle, Strategy.

По популярности на топовых платформах, в первую тройку вошли по объёмам продаж:
На PS4 - Action, Shooter, Sports
На 3DS - Action, RPG, Simulation
На XOne - Shooter, Action, Sports
Если убрать популярные жанры Action, Shooter, Sports, то перспективно выглядит RPG, так как на всех платформах этот жанр занимает средний объём.

Посмотрел распределение объёмов продаж по жанрам. Хорошие тиражи за 2014-2015 у Shooter, Action, RPG, Sports. Больше всего таких тиражей у Action, особенно в 2015. Лидируют так же Sports и RPG, у них количество успешных продаж меньше. Медиана у всех жанров от года к году снижается и только у Shooter сильно растёт, а у RPG, Fighting и Racing немного растёт.

По распределению продаж у Shooter и Racing 50% значений находятся на большем расстоянии от медианы. Выше всех медиана у Shooter, на втором месте Racing, на третьем Fighting. Наименьшее стандартное отклонение у Action, значит не смотря на выбросы, данные плотно сконцентрированы вокруг среднего значения. У Fighting и Racing тоже низкое значение стандартного отклонения от среднего.

Составил портрет пользователя за 2015 по регионам NA, EU, JP (другие регионы дополнительно описаны в 4 пункте исследования).

Результат для первой тройки платформ в каждом регионе:
Европа - PS4 59.8%, XOne 21.3%, 6.43% PC.
Япония - 3DS 51.2%, PS4 20.8%, 16% PSV.
Северная Америка - PS4 45%, 38.3% XOne, 7.19% WiiU.

Самые популярные жанры:
Европа - 50.7% Shooter, 18.8% Racing, третье Fighting 11.6% и Simulation 11.6%.
Япония - 31.3% RPG, 21.9% Puzzle, 18.8% Misc.
Северная Америка - 70.5% Shooter, 9.85% Platform, 7.58% Fighting.
Влияние рейтинга ESRB на продажи в отдельном регионе: Европа - много от 17 лет, на втором месте без ограничения возраста, на третьем от 13 лет. Япония - много 13 лет, на втором без ограничения возраста, на третьем от 10 лет. Северная Америка -много 17 лет, второе место без ограничения возраста, третье место от 13 лет. примечание: Для всех регионов выше - рейтинг UKW (нет данных о рейтинге возраста) не учитываю (исследование ниже), так как с этой отметкой много свойств.

Проверка гипотез:
Описание выбора типа теста, альфы и теория в пункте 5. Выдвинутые гипотезы:
Средние пользовательские рейтинги платформ Xbox One и PC одинаковые – не отвергаем.
Средние пользовательские рейтинги жанров Action (англ. «действие», экшен-игры) и Sports (англ. «спортивные соревнования») разные. – отвергаем

Дополнительно посмотрел, как бренды представлены за период по всем данным:
С 1983 Nintendo активно занимает объемы выпусков цифровых копий. С 1994 в борьбу с Nintendo вступает Sony и до 2005 показывает более хорошие результаты. С 2006 по 2010 Nintendo вновь обгоняет Sony. Но с 2011 Sony опять вырывается вперед Nintendo. С 2000 на третьем месте Microsoft с своими платформами. Эти три бренда активней всего себя проявили. В 2014, 2015 лидер Sony, на втором месте Microsoft, на третьем Nintendo. PC тоже присутствует, но занимает маленький объем. Если смотреть на пики, то у Sony пики бывают часто, у Microsoft пару пиков (самый заметный в 2010) и после нисходящий тренд, а у Nintendo 4 пика в 2006 - 2009 и они приходятся на момент рассвета релизов на всех платформах, а после нисходящий тренд. По суммарным продажам игр на старых платформах самые заметные пики по объёмам продаж у PS, XB, GC, N64.

Продукт для рекламы - игра. В разных регионах игры и платформы представлены разными групами, для каждой лучше составить свой таргет: описан в первый вывод

Дополнительно: можно взять какую-то часть рекламного бюджета (20-30%) и использовать её для эксперемента, так как описанные выше таргеты ведут на высококонкурентный рынок. А эксперемент провести на перспективном варианте игры или платформы, выход которой можно предположить по заявкам студий разработчиков. Возможен выход популярной многопользовательской игры (мультиплеерной/командной). Хорошо дополнительно учитывать конкуретное развитие VR / AR рынка.

Техническое:
1) platform : DS - есть игра в 1985, но в этом году не было такой приставки, похоже на аномалию.
2) platform : GB - 1993 год без игр - а игры были, возможно аномалия.
2) По jp_sales нет данных в good_data_2015 об объёмах продаж platform : PC.
3) Исследование UKW в user_score. Всего в диаграмме 'цифровых копий по рейтингам ESRB в 2015' визуализированы 1604 строки, из которых UKW составляют 727 строк. Так как ESRB действует в регионе США и Канады, то можно предположить, что рейтинг можно применять только к параметру na_sales, но при проверке продаж на уровне жанров везде есть продажи, а значит с таким рейтингом игры продаются и в других регионах (или запись данных аномальна). Попробую рассмотреть другие регионы. Если только в одном регионе есть продажи игры, а в других нет, то возможно для этой игры существует свой региональный рейтинг. Для jp_sales получилось 420

Мысли после зачета
После WiiU пришло следующее её поколение Nintendo Switch. То есть, аудитория WiiU не пропала, она купила следующее поколение, но кто-то мог уйти и к другому бренду. Наблюдение, что WiiU и 3DS (и Switch) - это Nintendo , которое популярно в Азии и эта аудитория не охотно отказывается от Nintendo. Игра должна создавать свой уникальный мир - можно предположить, что нужно комбинировать жанры для успешности следующего проекта. Что в данных нахватает класса проектов 3А и т.д (параметров: названия студий, срок производства и нахождения на рынке, объёмы продаж игры в каждой год и т.п). Правильным было бы почитать новости, после подведения выводов исследования, так как может оказаться, что на следующий год большой объем ААА или какая-то платформа уходит/снимается с рынка (может быть планируется релиз новой консоли/поколения в следующем году). Использование защиты типа Denuvo может повысить продажи на PC, но её часто взламывают и у нее плохая оптимизация. В идеале выделить портативных в отдельный класс, ПС рассматривать как-то отдельно, а все актуальные поколения платформ отдельно.

Выбор даты релиза
Каждый год повторяется одно и то же: сперва идут спокойные девять месяцев, в течение которых выходит несколько крупных игр, а затем наступает перенасыщенный период с огромным количеством релизов ААА-игр. Конечно, время перед Рождеством и Новым годом бывает крайне удачным для компаний — для некоторых из них на этот период приходится 60% от годовой

прибыли. Но в эти месяцы крайне сильна конкуренция, из-за чего многие проекты не дают ожидаемых результатов. День - самое главное — это когда продвигать тайтл, а не когда его выпускать. Семейные игры обычно выходит без большого шума, но её начинают активно продвигать перед праздниками, на которых могут собираться компании людей — Хэллоуин, День Благодарения, Рождество. Насколько специфическая аудитория у вашей игры? По сути, календарное окно со сниженной конкуренцией — это потенциальная возможность привлечь новых покупателей. Оцените свой маркетинговый бюджет - в начале года затраты на маркетинг меньше, а также в это время ниже конкуренция. Поэтому при небольшом маркетинговом бюджете имеет смысл рассмотреть этот период. Оцените возможность продвижения с помощью СМИ - журналисты с большей охотой возьмутся писать про популярные игры, потому что это привлечёт больше внимания со стороны аудитории. Учитывайте время проведения крупных распродаж в ЧП или цифровых прилавках (но можно и сыграть на том, что в распродажи всё скупают и выходить под них). Не бойтесь откладывать релиз, но это не всегда верный вариант - некоторые издатели решают выпустить свою игру близко к релизу тайтла-конкурента, потому что она якобы «привлекает другую аудиторию» (но часто это не срабатывает, а вся аудитория уходит в одну сторону). Оцените абсолютно всех соперников - планировании релиза нужно учитывать все игры, выходящие примерно в то же время. Сейчас людям всегда могут найти себе занятие, поэтому недостаточно учитывать только прямых конкурентов. Например, Ubisoft ничего не могла поделать с тем, что Starlink была рассчитана на ту же аудиторию, что и Fortnite. Кроме того, в анализе важно выходить за рамки одних лишь игр. Например, в 2017 году выдалась напряжённая неделя когда выходят игры и крутые сериалы (сериалы поедают аудиторию игры, сериалы могут быть такие «Очень странных дел» и «Top:Рагнарёк»). Следите за окном между серединой февраля и началом марта - примерно через шесть-восемь недель после 25 декабря люди заканчивают прохождение своих рождественских игр, а затем начинают искать что-то новое. Обсудите дату релиза с ритейлером - например, как много детских игр выходит перед Рождеством или Пасхой? Или сколько инди-метроидваний выйдет летом?

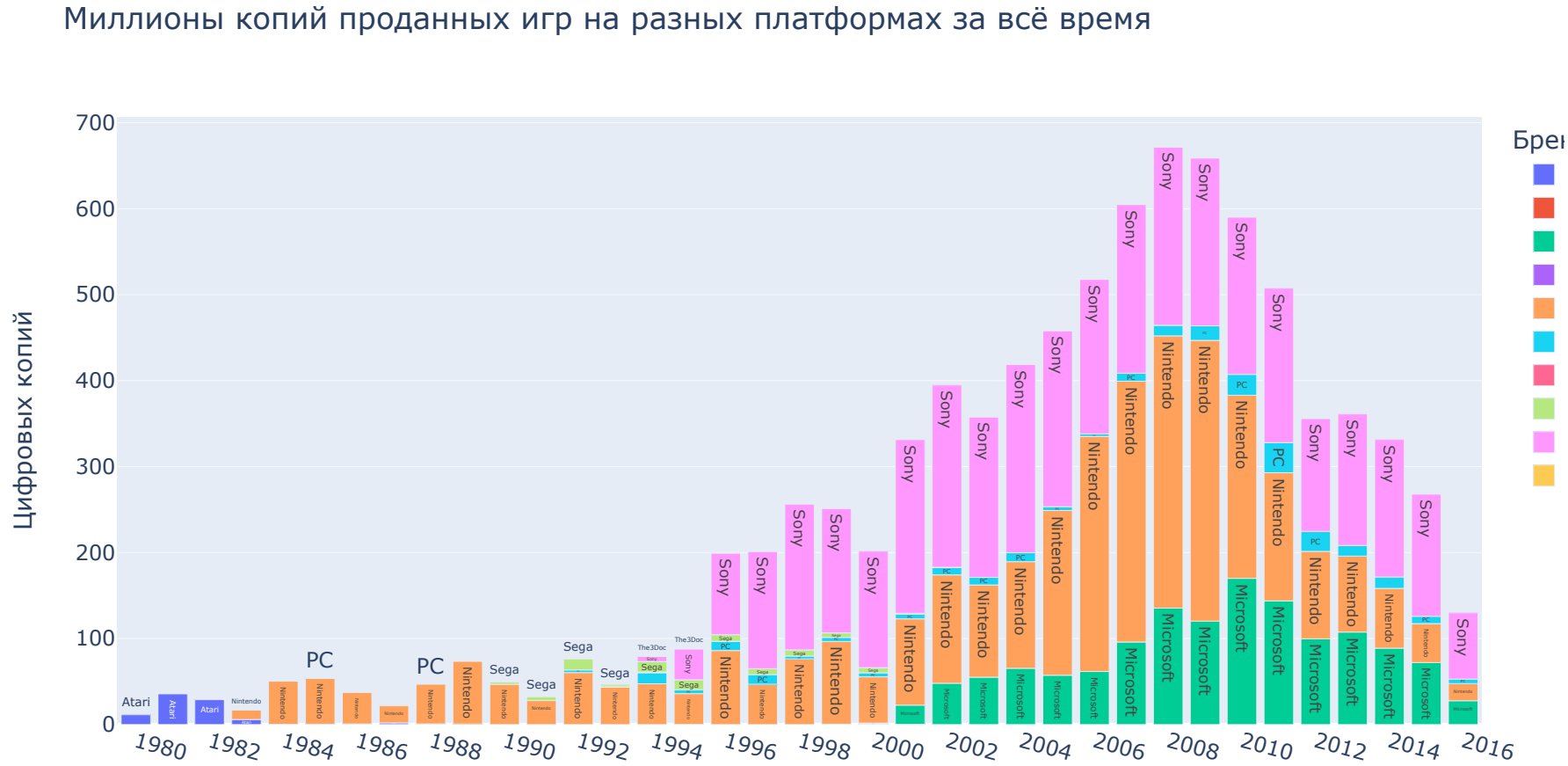
7 Дополнительные исследования

7.1 объём продаж брендов

In [119]:
```
data['brand'] = data['platform'].apply(cat_brand)
```

In [120]:
```
cat_brand = data.pivot_table(index = ['brand', 'year_of_release'],
                             values='total_cop',
                             aggfunc = ['sum']).reset_index()
cat_brand.columns = ['Бренд', 'Год', 'Цифровых копий']
```

In [121]:
```
px.bar(cat_brand,
       x = 'Год',
       y='Цифровых копий',
       color='Бренд',
       title = 'Миллионы копий проданных игр на разных платформах за всё время',
       text='Бренд').update_xaxes(dtick=2, tickangle=15)
```



С 1983 Nintendo активно занимает объемы выпусков цифровых копий. С 1994 года в борьбу с Nintendo вступает Sony и до 2005 года показывает более хорошие результаты. С 2006 года по 2010 Nintendo вновь обгоняет Sony. Но с 2011 года Sony опять вырывается вперед Nintendo. С 2000 года на третьем месте Microsoft с своими платформами. Эти три бренда активней всего себя проявили. В 2014-2015 году лидер Sony, на втором месте Microsoft, на третьем Nintendo. PC тоже присутствует, но занимает маленький объем. Если смотреть на пики, то у Sony пики бывают часто, у Microsoft пару пиков (самый заметный в 2010) и после нисходящий тренд, а у Nintendo 4 пика в 2006 - 2009 и они приходятся на момент рассвета количества релизов на всех платформах, а после нисходящий тренд.

7.2 суммарные продажи игр на старых платформах

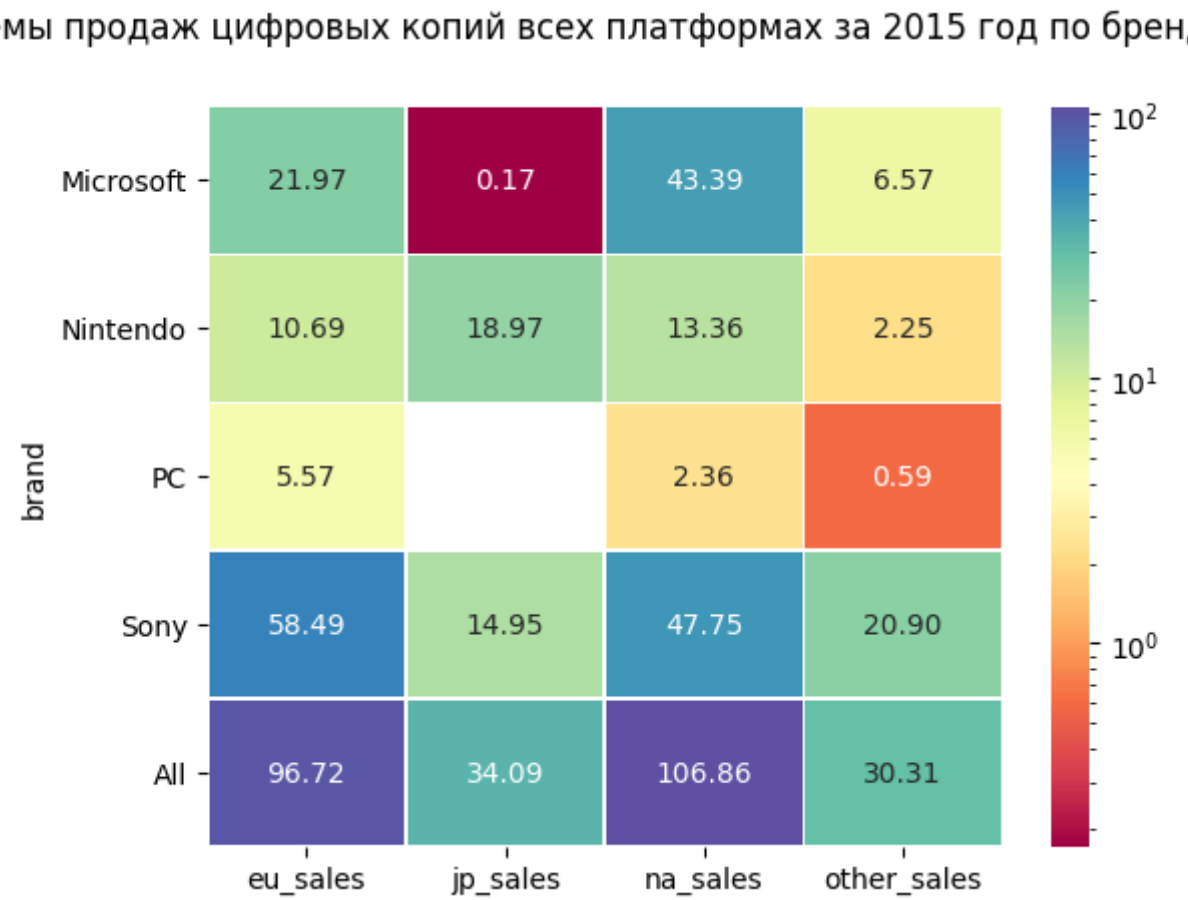
In [122]:
```
cat_pltf_rel = data.pivot_table(index='year_of_release',
                                columns='platform', values='total_cop',
                                aggfunc='sum').reset_index()
```

In [123]:
```
px.line(cat_pltf_rel,
        x='year_of_release',
        y=old_pltf,
        markers=True,
        hover_name='year_of_release',
        line_shape = 'spline',
        range_x=[1979,2009],
        title = 'Миллионы цифровых копий вышедших на разных платформах до 2012 года',
        color_discrete_sequence=px.colors.qualitative.Set3).update_xaxes(dtick=2, tickangle=45)
```

Самые заметные пики по объёмам продаж у PS, XB, GC, N64.

7.3 объёмы продаж цифровых копий на всех платформах за 2015 год по брендам и регионам

In [124]:
```
perspective_market = data.query('year_of_release == 2015')\
                         .pivot_table(index='brand', values=['na_sales', 'eu_sales', 'jp_sales', 'other_sales'],
                                      aggfunc='sum', margins='True')
sns.heatmap(perspective_market, vmax=1, annot=True, cmap="Spectral",
            annot_kws={'size':10}, fmt='.2f', linewidths=.5, norm=LogNorm())
plt.yticks(rotation=0);
plt.suptitle('Объёмы продаж цифровых копий всех платформах за 2015 год по брендам и регионам');
```



По общим объёмам продаж за 2015 год топ3: Северная Америка, Европа, Япония.

In [125]:
```
print(dino.get_output_string('trex', 'Game Wi!! Continue'));
```

```
 __________________
| Game Wi!! Continue |
  ==================
                     \
                      \
                       \
                        \
                            .-=-==--==--.
                       ..-=="  ,'o`)      `.
                     ,'         `"'         \
                    :  (                    `.__....._
                    |                 )    /          `-=-.
                    :       ,vv.-._   /    /              `---==-._
                     \/\/\/VV ^ d88`;'    /                        `.
                         ``  ^/d88P!'    /                 ,          `._
                               ^/    !'   ,.      ,      /                 "-,,__,,--'"""-.
                              ^/    !'  ,'  \ . .(      (          _          )  ) ) ) ))_,-.\
                             ^(__ ,!',"'   ;:+.:%:a.     \:.. . ,'          )  )  ) ) ,"'    '
                             ',,,'',' '     /o:::":%:%a.    \:.:.:         .    )  ) _,'
                              """'       ;':::'' `+%%%a._  \%:%|         ;.). _,-""
                                       ,-='_.-'      ``:%::)  )%:|       /:._,"
                                      ((/("           ," ,'_,'%%%:      (_,'
                                                     (  (//(`.___;       \
                                                      \     \    `        `
                                                       `.     `.   `.       :
                                                         \. . .\    : . . . :
                                                          \. . .:    `.. . .:
                                                           `..:.:\     \:...\
                                                            ;:.:.;      ::....:
                                                            ):%::       ::::::;
                                                         __,::%:(       :::::
                                                        ,;:%%%%%%:       ;:%::
                                                          ;,--""-.`\  ,=--':%:%:\
                                                         /"       "| /-".:%%%%%%\
                                                                    ;,-"'`)%%)
                                                                   /"     "|
```