# AnalysisAndGraphs

# March 23, 2024

```
[1]: import pandas as pd
      import matplotlib.pyplot as plt
      import numpy as np
      import plotly.express as px
      import seaborn as sns
      import warnings
      warnings.filterwarnings("ignore")
      from plotnine.data import mpg
      from plotnine import ggplot, aes, geom_point, geom_bar, geom_boxplot, __
       Geom_hline, facet_grid, labs, geom_point, geom_col, stat_count, □

¬geom_histogram, stat_density, coord_flip
      from plotnine import scale_x_continuous, scale_y_continuous,_
       scale_colour_discrete, scale_x_reverse, scale_y_reverse, scale_color_manual,_
       scale_color_brewer, scale_x_discrete, scale_fill_manual
      from plotnine import theme_xkcd, theme_classic, theme_light, theme,_
       \hookrightarrowelement_blank
      from plotnine import geom_text, position_stack
 [3]: bucket_name = "bucketteam3"
      folder = "CSV"
      path="gs://{bucket_name}/{folder}/".format(bucket_name=bucket_name,_
       ⇔folder=folder)
 [3]: df_actors = pd.read_csv(path+"actors.csv")
 [9]: df_countries = pd.read_csv(path+"countries.csv")
 [5]: df_crew = pd.read_csv(path+"crew.csv")
[17]: df_genres = pd.read_csv(path+"genres.csv")
[36]: df_languages = pd.read_csv(path+"languages.csv")
 [4]: df_movies = pd.read_csv(path+"movies.csv")
 [5]: df_releases = pd.read_csv(path+"releases.csv")
```

#### 0.0.1 Distribución del número de películas por país

#### 0.0.2 Distribución del número de películas por país (excluyendo EE. UU.)

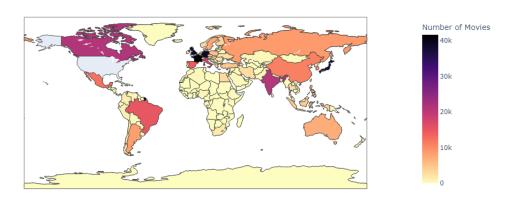
```
filtered_country_counts = country_counts[country_counts['country'] != 'USA']

fig = px.choropleth(
    filtered_country_counts,
    locations='country',
    locationmode='country names',
    color='movie_count',
    color_continuous_scale='Magma_r',
    title='Número de películas por país [excluyendo EE. UU.]'
)

fig.update_layout(
    coloraxis_colorbar=dict(title='Number of Movies')
)
fig.update_layout(
```

```
width=700,
height=500,
)
fig.show()
```

#### Número de películas por país [excluyendo EE. UU.]



# Mapeamos Pais con Continente

```
[13]: country_to_continent = {
          'Afghanistan': 'Asia',
          'Albania': 'Europe',
          'Algeria': 'Africa',
          'American Samoa': 'Oceania',
          'Andorra': 'Europe',
          'Angola': 'Africa',
          'Anguilla': 'North America',
          'Antarctica': 'Antarctica',
          'Antigua and Barbuda': 'North America',
          'Argentina': 'South America',
          'Armenia': 'Asia',
          'Aruba': 'North America',
          'Australia': 'Oceania',
          'Austria': 'Europe',
          'Azerbaijan': 'Asia',
          'Bahamas': 'North America',
          'Bahrain': 'Asia',
          'Bangladesh': 'Asia',
          'Barbados': 'North America',
          'Belarus': 'Europe',
          'Belgium': 'Europe',
```

```
'Belize': 'North America',
'Benin': 'Africa',
'Bermuda': 'North America',
'Bhutan': 'Asia',
'Bolivarian Republic of Venezuela': 'South America',
'Bolivia': 'South America',
'Bosnia and Herzegovina': 'Europe',
'Botswana': 'Africa',
'Bouvet Island': 'Antarctica',
'Brazil': 'South America',
'British Indian Ocean Territory': 'Asia',
'British Virgin Islands': 'North America',
'Brunei Darussalam': 'Asia',
'Bulgaria': 'Europe',
'Burkina Faso': 'Africa',
'Burundi': 'Africa',
'Cambodia': 'Asia',
'Cameroon': 'Africa',
'Canada': 'North America',
'Cape Verde': 'Africa',
'Cayman Islands': 'North America',
'Central African Republic': 'Africa',
'Chad': 'Africa',
'Chile': 'South America',
'China': 'Asia',
'Christmas Island': 'Asia',
'Cocos (Keeling) Islands': 'Asia',
'Colombia': 'South America',
'Comoros': 'Africa',
'Congo': 'Africa',
'Cook Islands': 'Oceania',
'Costa Rica': 'North America',
'Croatia': 'Europe',
'Cuba': 'North America',
'Cyprus': 'Europe',
'Czechia': 'Europe',
'Czechoslovakia': 'Europe',
'Democratic Republic of Congo': 'Africa',
'Denmark': 'Europe',
'Djibouti': 'Africa',
'Dominica': 'North America',
'Dominican Republic': 'North America',
'East Germany': 'Europe',
'Ecuador': 'South America',
'Egypt': 'Africa',
'El Salvador': 'North America',
'Equatorial Guinea': 'Africa',
```

```
'Eritrea': 'Africa',
'Estonia': 'Europe',
'Eswatini': 'Africa',
'Ethiopia': 'Africa',
'Falkland Islands': 'South America',
'Faroe Islands': 'Europe',
'Federated States of Micronesia': 'Oceania',
'Fiji': 'Oceania',
'Finland': 'Europe',
'France': 'Europe',
'French Guiana': 'South America',
'French Polynesia': 'Oceania',
'French Southern Territories': 'Antarctica',
'Gabon': 'Africa',
'Gambia': 'Africa',
'Georgia': 'Asia',
'Germany': 'Europe',
'Ghana': 'Africa',
'Gibraltar': 'Europe',
'Greece': 'Europe',
'Greenland': 'North America',
'Grenada': 'North America',
'Guadeloupe': 'North America',
'Guam': 'Oceania',
'Guatemala': 'North America',
'Guinea': 'Africa',
'Guinea-Bissau': 'Africa',
'Guyana': 'South America',
'Haiti': 'North America',
'Heard Island and McDonald Islands': 'Antarctica',
'Honduras': 'North America',
'Hong Kong': 'Asia',
'Hungary': 'Europe',
'Iceland': 'Europe',
'India': 'Asia',
'Indonesia': 'Asia',
'Iran': 'Asia',
'Iraq': 'Asia',
'Ireland': 'Europe',
'Israel': 'Asia',
'Italy': 'Europe',
'Ivory Coast': 'Africa',
'Jamaica': 'North America',
'Japan': 'Asia',
'Jordan': 'Asia',
'Kazakhstan': 'Asia',
'Kenya': 'Africa',
```

```
'Kiribati': 'Oceania',
'Kosovo': 'Europe',
'Kuwait': 'Asia',
'Kyrgyzstan': 'Asia',
"Lao People's Democratic Republic": 'Asia',
'Latvia': 'Europe',
'Lebanon': 'Asia',
'Lesotho': 'Africa',
'Liberia': 'Africa',
'Libya': 'Africa',
'Liechtenstein': 'Europe',
'Lithuania': 'Europe',
'Luxembourg': 'Europe',
'Macao': 'Asia',
'Madagascar': 'Africa',
'Malawi': 'Africa',
'Malaysia': 'Asia',
'Maldives': 'Asia',
'Mali': 'Africa',
'Malta': 'Europe',
'Marshall Islands': 'Oceania',
'Martinique': 'North America',
'Mauritania': 'Africa',
'Mauritius': 'Africa',
'Mayotte': 'Africa',
'Mexico': 'North America',
'Monaco': 'Europe',
'Mongolia': 'Asia',
'Montenegro': 'Europe',
'Montserrat': 'North America',
'Morocco': 'Africa',
'Mozambique': 'Africa',
'Myanmar': 'Asia',
'Namibia': 'Africa',
'Nauru': 'Oceania',
'Nepal': 'Asia',
'Netherlands': 'Europe',
'Netherlands Antilles': 'North America',
'New Caledonia': 'Oceania',
'New Zealand': 'Oceania',
'Nicaragua': 'North America',
'Niger': 'Africa',
'Nigeria': 'Africa',
'Niue': 'Oceania',
'Norfolk Island': 'Oceania',
'North Korea': 'Asia',
'North Macedonia': 'Europe',
```

```
'Northern Mariana Islands': 'Oceania',
'Norway': 'Europe',
'Oman': 'Asia',
'Pakistan': 'Asia',
'Palau': 'Oceania',
'Panama': 'North America',
'Papua New Guinea': 'Oceania',
'Paraguay': 'South America',
'Peru': 'South America',
'Philippines': 'Asia',
'Pitcairn': 'Oceania',
'Poland': 'Europe',
'Portugal': 'Europe',
'Puerto Rico': 'North America',
'Qatar': 'Asia',
'Republic of Moldova': 'Europe',
'Romania': 'Europe',
'Russian Federation': 'Europe',
'Rwanda': 'Africa',
'Réunion': 'Africa',
'Saint Helena, Ascension and Tristan da Cunha': 'Africa',
'Saint Kitts and Nevis': 'North America',
'Saint Lucia': 'North America',
'Saint Pierre and Miquelon': 'North America',
'Saint Vincent and the Grenadines': 'North America',
'Samoa': 'Oceania',
'San Marino': 'Europe',
'Sao Tome and Principe': 'Africa',
'Saudi Arabia': 'Asia',
'Senegal': 'Africa',
'Serbia': 'Europe',
'Serbia and Montenegro': 'Europe',
'Seychelles': 'Africa',
'Sierra Leone': 'Africa',
'Singapore': 'Asia',
'Slovakia': 'Europe',
'Slovenia': 'Europe',
'Solomon Islands': 'Oceania',
'Somalia': 'Africa',
'South Africa': 'Africa',
'South Georgia and the South Sandwich Islands': 'Antarctica',
'South Korea': 'Asia',
'South Sudan': 'Africa',
'Spain': 'Europe',
'Sri Lanka': 'Asia',
'State of Palestine': 'Asia',
'Sudan': 'Africa',
```

```
'Suriname': 'South America',
    'Svalbard and Jan Mayen': 'Europe',
    'Sweden': 'Europe',
    'Switzerland': 'Europe',
    'Syrian Arab Republic': 'Asia',
    'Taiwan': 'Asia',
    'Tajikistan': 'Asia',
    'Thailand': 'Asia',
    'Timor-Leste': 'Asia',
    'Togo': 'Africa',
    'Tokelau': 'Oceania',
    'Tonga': 'Oceania',
    'Trinidad and Tobago': 'North America',
    'Tunisia': 'Africa',
    'Turkey': 'Asia',
    'Turkmenistan': 'Asia',
    'Turks and Caicos Islands': 'North America',
    'Tuvalu': 'Oceania',
    'UK': 'Europe',
    'US Virgin Islands': 'North America',
    'USA': 'North America',
    'USSR': 'Europe',
    'Uganda': 'Africa',
    'Ukraine': 'Europe',
    'United Arab Emirates': 'Asia',
    'United Republic of Tanzania': 'Africa',
    'United States Minor Outlying Islands': 'North America',
    'Uruguay': 'South America',
    'Uzbekistan': 'Asia',
    'Vanuatu': 'Oceania',
    'Vatican City': 'Europe',
    'Vietnam': 'Asia',
    'Wallis and Futuna': 'Oceania',
    'Western Sahara': 'Africa',
    'Yemen': 'Asia',
    'Yugoslavia': 'Europe',
    'Zambia': 'Africa',
    'Zimbabwe': 'Africa'
}
```

```
[14]: total_movies = country_counts['movie_count'].sum()
country_counts['percentage'] = (country_counts['movie_count'] / total_movies) *_\cup \leftrightarrow 100
country_counts['continent'] = country_counts['country'].

\( \times \text{map}(country_to_continent) \)
```

### 0.0.3 Distribución de Películas por Continente

```
fig = px.sunburst(country_counts, path=['continent', 'country'], walues='percentage',

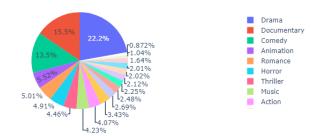
title='Porcentage' de películas producidas por continente yu país', hover_data={'percentage': ':.3'})
fig.show()
```

Porcentaje de películas producidas por continente y país



#### 0.0.4 Generos

```
[18]: df_genres.head()
[18]:
              id
                      genre
     0 1000001
                     Comedy
      1 1000001
                    Fantasy
      2 1000001 Adventure
      3 1000002
                     Comedy
      4 1000002
                   Thriller
[19]: df_countries.head()
[19]:
              id
                      country
      0 1000001
                           UK
      1 1000001
                          USA
      2 1000002 South Korea
      3 1000003
                          USA
      4 1000004
                     Germany
[20]: | genre_counts = df_genres['genre'].value_counts().reset_index()
      genre_counts.columns = ['Genre', 'Count']
      fig = px.pie(genre_counts, values='Count', names='Genre', title='Distribución_
       ⇔de géneros de películas')
      fig.show()
```



```
[21]: merged_df = pd.merge(df_genres, df_countries, on='id')
      grouped_df = merged_df.groupby(['country', 'genre']).size().
       ⇔reset_index(name='count')
      idx = grouped_df.groupby(['country'])['count'].transform(max) ==__
       ⇒grouped_df['count']
      highest_count_genres = grouped_df[idx]
      total_movies_per_country = grouped_df.groupby('country')['count'].sum()
      highest_count_genres['percentage'] = highest_count_genres.apply(
          lambda row: (row['count'] / total_movies_per_country[row['country']]) * 100,
          axis=1
      highest_count_genres['percentage'] = highest_count_genres['percentage'].round(2)
[23]: merged_df.head()
[23]:
              id
                      genre country
      0 1000001
                     Comedy
                                 UK
      1 1000001
                     Comedy
                                USA
      2 1000001
                    Fantasy
                                 UK
```

```
3 1000001
             Fantasy
                         USA
4 1000001 Adventure
                          UK
```

# [22]: highest\_count\_genres.head()

[22]:		country	genre	count	percentage
	5	Afghanistan	Documentary	93	47.45
	20	Albania	Drama	248	43.59
	39	Algeria	Drama	208	38.52
	51	American Samoa	Comedy	2	40.00
	58	Andorra	Drama	9	28.12

#### 0.0.5 Género más popular por país

```
[24]: fig = px.choropleth(
    highest_count_genres,
    locations='country',
    locationmode='country names',
    color='genre',
    title='Género con mayor recuento por país',
    hover_data=['count', 'percentage'],
    color_discrete_sequence=px.colors.qualitative.Set2
)

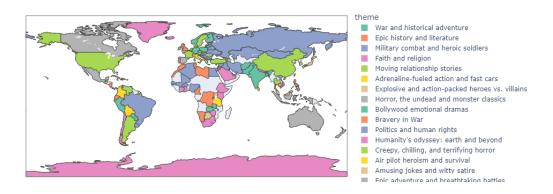
fig.update_layout(
    coloraxis_colorbar=dict(title='Genre'),
    width=700,
    height=500,
)
fig.show()
```

Género con mayor recuento por país



```
axis=1
[26]: highest_count_themes.head()
[26]:
              country
                                                      theme count percentage
                              War and historical adventure
      17
          Afghanistan
                                                                 3
                                                                     13.043478
      18
              Albania
                               Epic history and literature
                                                                 2
                                                                      8.333333
      22
              Albania Historical battles and epic heroism
                                                                 2
                                                                      8.333333
              Albania Military combat and heroic soldiers
      27
                                                                 2
                                                                      8.333333
      30
              Albania
                                 Politics and human rights
                                                                      8.333333
[27]: # Elimina los duplicados manteniendo el que tiene el recuento más alto y luegou
      ⇒el que tiene el nombre del tema en orden alfabético más alto.
      highest_count_themes = highest_count_themes.sort_values(['country', 'count', __
       →'theme'], ascending=[True, False, True]) \
          .drop duplicates(subset=['country'])
      highest_count_themes['percentage'] = highest_count_themes['percentage'].round(2)
[28]: highest_count_themes.head()
[28]:
               country
                                                       theme count
                                                                    percentage
      17
           Afghanistan
                               War and historical adventure
                                                                  3
                                                                          13.04
      18
               Albania
                                Epic history and literature
                                                                  2
                                                                           8.33
      43
               Algeria Military combat and heroic soldiers
                                                                  3
                                                                          12.50
      49
                                                                  1
            Antarctica
                                         Faith and religion
                                                                          12.50
      120
             Argentina
                                Moving relationship stories
                                                                 20
                                                                           4.89
[29]: fig = px.choropleth(
          highest_count_themes,
          locations='country',
          locationmode='country names',
          color='theme',
          title='Temas con mayor recuento por país',
          hover_data=['count', 'percentage'],
          color_discrete_sequence=px.colors.qualitative.Set2
      )
      fig.update_layout(
          coloraxis_colorbar=dict(title='Genre'),
          width=700,
          height=500,
      fig.show()
```

Temas con mayor recuento por país



#### 0.0.6 Releases

# [30]: df\_releases.head()

[30]:		id	country	date	type	rating
	0	1000001	Andorra	2023-07-21	Theatrical	NaN
	1	1000001	Singapore	2023-07-20	Theatrical	PG13
	2	1000001	Saudi Arabia	2023-08-10	Theatrical	NaN
	3	1000001	Romania	2023-07-21	Theatrical	NaN
	4	1000001	Puerto Rico	2023-12-16	TV	PG-13

### 0.0.7 Distribución de tipos de lanzamiento.

```
[31]: type_counts = df_releases['type'].value_counts()
fig = px.pie(type_counts, values=type_counts.values, names=type_counts.index,
title='Distribución de tipos de lanzamiento')
fig.show()
```

Distribución de tipos de lanzamiento



#### 0.0.8 Tipos de lanzamientos populares en todo el mundo

[32]: grouped\_df = df\_releases.groupby(['country', 'type']).size().

```
⇔reset_index(name='count')
      idx = grouped_df.groupby(['country'])['count'].transform(max) ==__
       ⇒grouped_df['count']
      highest_count_releases = grouped_df[idx]
      total_movies_per_country = grouped_df.groupby('country')['count'].sum()
      highest_count_releases['percentage'] = highest_count_releases.apply(
          lambda row: (row['count'] / total_movies_per_country[row['country']]) * 100,
          axis=1
      )
      highest_count_releases['percentage'] = highest_count_releases['percentage'].
       ⇒round(2)
[33]: grouped_df.head()
[33]:
             country
                                    type count
      0 Afghanistan
                                 Digital
                                             30
      1 Afghanistan
                                Physical
                                             14
      2 Afghanistan
                                              8
                                Premiere
      3 Afghanistan
                              Theatrical
                                             90
      4 Afghanistan Theatrical limited
                                              3
[34]: fig = px.choropleth(
          highest_count_releases,
          locations='country',
          locationmode='country names',
          color='type',
          title='Tipos de lanzamiento más populares en todo el mundo',
          hover_data=['count', 'percentage'],
          color_discrete_sequence=px.colors.qualitative.Set2
      )
      fig.update_layout(
          coloraxis_colorbar=dict(title='Release Type'),
          width=700,
          height=500,
      fig.show()
```

Tipos de lanzamiento más populares en todo el mundo



Parece que el tipo de estreno más popular es el cine en todo el mundo; sin embargo, ciertos países como Tayikistán (Asia Central) y ciertos países africanos: Guinea-Bissau, Uganda, Mozambique y Botswana tienen estrenos digitales como más populares.

#### 0.0.9 Idiomas

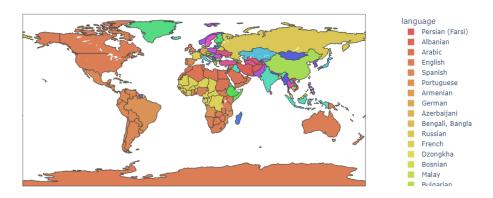
```
[37]: df_languages.head()
[37]:
              id
                              type language
         1000001
                          Language
                                    English
      1 1000002 Primary language
                                     Korean
                   Spoken language English
      2 1000002
        1000002
                   Spoken language
                                     German
      3
        1000002
                   Spoken language
                                     Korean
[38]: #Hay 188 idiomas unicos en el dataset
      merged_df = pd.merge(df_themes, df_countries, on='id')
      grouped_df = merged_df.groupby(['country', 'theme']).size().
       ⇔reset_index(name='count')
      df_languages["language"].nunique()
[38]: 188
[39]:
     merged_df.head()
[39]:
              id
                                   theme country
      0
         1000001
                     Relationship comedy
                                               UK
       1000001
                     Relationship comedy
      1
                                              USA
        1000001
                          Song and dance
                                              UK
```

```
3 1000001
                          Song and dance
                                             USA
      4 1000001 Crude humor and satire
                                              UK
[40]: merged_df = pd.merge(df_languages, df_countries, on='id')
      grouped_df = merged_df.groupby(['country', 'language']).size().
       ⇔reset_index(name='count')
      idx = grouped_df.groupby(['country'])['count'].transform(max) ==__

¬grouped_df['count']

      highest_count_languages = grouped_df[idx]
      total_movies_per_country = grouped_df.groupby('country')['count'].sum()
      highest_count_languages['percentage'] = highest_count_languages.apply(
          lambda row: (row['count'] / total movies_per_country[row['country']]) * 100,
      highest_count_languages['percentage'] = highest_count_languages['percentage'].
       ⇒round(2)
[41]: merged_df.head()
[41]:
                              type language
              id
                                                 country
      0 1000001
                          Language English
                                                      UK
      1 1000001
                          Language English
                                                     USA
      2 1000002 Primary language
                                    Korean South Korea
      3 1000002
                   Spoken language English South Korea
      4 1000002
                   Spoken language
                                     German South Korea
[42]: #Funcion para codigo de colores para luego usar en figura
      languages = list(highest_count_languages["language"].unique())
      colors = sns.color_palette("hls", len(languages))
      language_colors = dict(zip(languages, colors))
      def rgb_to_hex(rgb):
          """Converts an RGB tuple to hexadecimal."""
          return \frac{\pi}{02x}\frac{02x}{02x} tuple(int(255 * x) for x in rgb)
      language_colors_hex = {language: rgb_to_hex(color) for language, color in_
       ⇔language colors.items()}
[43]: print(language_colors_hex)
     {'Persian\xa0(Farsi)': '#db5e56', 'Albanian': '#db6956', 'Arabic': '#db7356',
     'English': '#db7e56', 'Spanish': '#db8856', 'Portuguese': '#db9256', 'Armenian':
     '#db9d56', 'German': '#dba756', 'Azerbaijani': '#dbb256', 'Bengali, Bangla':
     '#dbbc56', 'Russian': '#dbc756', 'French': '#dbd156', 'Dzongkha': '#dadb56',
     'Bosnian': '#cfdb56', 'Malay': '#c5db56', 'Bulgarian': '#badb56', 'Khmer':
     '#b0db56', 'Chinese': '#a5db56', 'Croatian': '#9bdb56', 'Greek\xa0(modern)':
     '#91db56', 'Czech': '#86db56', 'Danish': '#7cdb56', 'Somali': '#71db56',
     'Tigrinya': '#67db56', 'Estonian': '#5cdb56', 'Amharic': '#56db5b', 'Faroese':
```

```
'#56db65', 'Finnish': '#56db70', 'Georgian': '#56db7a', 'Kalaallisut,
     Greenlandic': '#56db85', 'Haitian, Haitian Creole': '#56db8f', 'Cantonese':
     '#56db99', 'Hungarian': '#56dba4', 'Icelandic': '#56dbae', 'Hindi': '#56dbb9',
     'Indonesian': '#56dbc3', 'Kurdish': '#56dbce', 'Hebrew\xa0(modern)': '#56dbd8',
     'Italian': '#56d3db', 'Japanese': '#56c8db', 'Kazakh': '#56bedb', 'Kyrgyz':
     '#56b3db', 'Lao': '#56a9db', 'Latvian': '#569fdb', 'Southern Sotho': '#5694db',
     'Lithuanian': '#568adb', 'Malagasy': '#567fdb', 'Divehi, Dhivehi, Maldivian':
     '#5675db', 'Marshallese': '#566adb', 'Mongolian': '#5660db', 'Serbian':
     '#5756db', 'Burmese': '#6256db', 'Nepali': '#6c56db', 'Dutch': '#7756db',
     'Māori': '#8156db', 'Korean': '#8c56db', 'Macedonian': '#9656db', 'Norwegian':
     '#a056db', 'Urdu': '#ab56db', 'Tagalog': '#b556db', 'Polish': '#c056db',
     'Romanian': '#ca56db', 'Slovak': '#d556db', 'Slovene': '#db56d6', 'Sinhalese,
     Sinhala': '#db56cc', 'Swedish': '#db56c1', 'Tajik': '#db56b7', 'Thai':
     '#db56ac', 'Tonga\xa0(Tonga Islands)': '#db56a2', 'Turkish': '#db5698',
     'Turkmen': '#db568d', 'Ukrainian': '#db5683', 'Uzbek': '#db5678', 'Bislama':
     '#db566e', 'Vietnamese': '#db5663', 'Serbo-Croatian': '#db5659'}
[44]: fig = px.choropleth(
          highest count languages,
          locations='country',
          locationmode='country names',
          color='language',
          title='Idiomas de películas más populares en todo el mundo',
          hover data=['count', 'percentage'],
          color discrete map=language colors hex #Resultado de funcion
      )
      fig.update_layout(
          coloraxis_colorbar=dict(title='Release Type'),
          width=700,
          height=500,
      fig.show()
```



#### 0.0.10 Studios

```
[45]: df studios.head()
[45]:
              id
                                   studio
      0 1000001 LuckyChap Entertainment
      1 1000001
                             Heyday Films
      2 1000001
                           NB/GG Pictures
      3 1000001
                                   Mattel
      4 1000001
                    Warner Bros. Pictures
[46]: merged df = pd.merge(df studios, df countries, on='id')
      grouped_df = merged_df.groupby(['country', 'studio']).size().
      →reset index(name='count')
      idx = grouped_df.groupby(['country'])['count'].transform(max) ==__
       ⇒grouped_df['count']
      highest_count_studios = grouped_df[idx]
      total_movies_per_country = grouped_df.groupby('country')['count'].sum()
      highest_count_studios['percentage'] = highest_count_studios.apply(
          lambda row: (row['count'] / total_movies_per_country[row['country']]) * 100,
          axis=1
      highest_count_studios['percentage'] = highest_count_studios['percentage'].
       →round(2)
[47]: merged_df.head()
[47]:
              id
                                   studio country
      0 1000001 LuckyChap Entertainment
                                               UK
```

```
3 1000001 Heyday Films USA
4 1000001 NB/GG Pictures UK

[48]: #Reutilizamos la estructura de la funcion de arriba para los estudios.
studios = list(highest_count_studios["studio"].unique())
colors = sns.color_palette("hls", len(studios))
studio_colors = dict(zip(studios, colors))

def rgb_to_hex(rgb):
    """Converts an RGB tuple to hexadecimal."""
    return '#%02x%02x%02x' % tuple(int(255 * x) for x in rgb)
studio_colors_hex = {studio: rgb_to_hex(color) for studio, color in_u
studio_colors.items()}
```

Heyday Films

USA

UK

## [49]: print(studio\_colors\_hex)

2 1000001

1 1000001 LuckyChap Entertainment

{'Shaheen Film': '#db5e56', 'Shqipëria e Re': '#db6056', 'Algerian Television': '#db6156', 'Navigator Island Pictures': '#db6356', 'Arlong Productions': '#db6456', 'Imminent Produccions': '#db6656', 'Geração 80': '#db6756', 'Continental Producciones': '#db6956', 'Department of Youth and Culture': '#db6a56', 'Ensueño Films': '#db6c56', 'HandMade Films': '#db6d56', 'Paragon Entertainment Corporation': '#db6f56', 'Tornasol Films': '#db7156', 'BBC': '#db7256', 'Burning Boat Productions': '#db7456', 'Disneynature': '#db7556', 'F11 Studio': '#db7756', 'Film Victoria': '#db7856', 'Getzels Gordon Productions': '#db7a56', 'Gumption Entertainment': '#db7b56', 'John Downer Productions Ltd.': '#db7d56', 'Management Entertainment': '#db7e56', 'Seek Discomfort Films': '#db8056', 'Seymour Entertainment': '#db8156', 'The Cousteau Society': '#db8356', 'Timeline Production': '#db8456', 'Yes Theory': '#db8656', 'Hama Production': '#db8756', 'INCAA': '#db8956', 'Armenia Studio': '#db8b56', 'Millennium Media': '#db8c56', 'Screen Australia': '#db8e56', 'ORF': '#db8f56', 'Azerbaijanfilm': '#db9156', 'TinRes Entertainment': '#db9256', 'Vilper Creative Agency': '#db9456', 'Impress Telefilm': '#db9556', 'Step by Step Productions': '#db9756', 'Belarusfilm': '#db9856', 'RTBF': '#db9a56', 'Bennett Marine Video': '#db9b56', 'Third Coast Productions': '#db9d56', 'Koffi Productions': '#db9e56', 'Triventure Films': '#dba056', 'HanWay Films': '#dba156', 'Centro Nacional Autónomo de Cinematografía (CNAC)': '#dba356', 'Grupo Ukamau': '#dba556', 'SCCA / pro.ba': '#dba656', 'National Geographic': '#dba856', 'Globo Filmes': '#dba956', 'Pidax Film': '#dbab56', 'British Lion Films': '#dbac56', 'Next Door Films': '#dbae56', 'Marhain Entertainment': '#dbaf56', 'Origin Artistic': '#dbb156', 'Boyana Film': '#dbb256', 'CNC': '#dbb456', 'Direction de la Cinematographie Nationale': '#dbb556', 'Les Productions de la Lanterne': '#dbb756', 'Sahélis Productions': '#dbb856', 'Burundi Film Center': '#dbba56', 'Studio MALEMBE Maa asbl': '#dbbb56', 'Khemara Pictures': '#dbbd56', 'Les Films du Raphia': '#dbbf56', 'ONF | NFB': '#dbc056', 'RTP': '#dbc256', 'Compass Light Productions': '#dbc356', 'MAESterpiece Productions': '#dbc556', 'Revolution

Films': '#dbc656', 'Strike Twice Creative': '#dbc856', 'Alliance française de Bangui': '#dbc956', 'Göi-Göi Productions': '#dbcb56', 'Fabula': '#dbcc56', 'Changchun Film Studio': '#dbce56', 'Channel 4 Television': '#dbcf56', 'Department of Defense': '#dbd156', 'UNESCO': '#dbd256', 'BOGOSHORTS Film Agency': '#dbd456', 'Al Jazeera': '#dbd556', 'Aurora Films': '#dbd756', 'En Quête Prod': '#dbd956', 'HMZ Productions': '#dbda56', 'Lacoupure': '#dadb56', 'Les Films façon façon': '#d8db56', 'Pica 269': '#d7db56', 'Allison Productions': '#d5db56', 'Claude Antoine Films': '#d4db56', 'Fixer Congo': '#d2db56', 'IMAX': '#d1db56', 'Isabelle Films': '#cfdb56', "L'Office National du Cinema": '#cedb56', "Les Films d'Ici": '#ccdb56', 'Les Films de la Passerelle': '#cbdb56', 'Lohmann Medien GmbH': '#c9db56', 'Mapa Filmes': '#c8db56', 'Max Scenars': '#c6db56', 'Polifilm': '#c5db56', 'Ryva': '#c3db56', 'Suka! Productions': '#c2db56', 'Field Theory': '#c0db56', 'Filmwork': '#bedb56', 'Imagezone (Dean Thomas)': '#bddb56', 'Imagezone (Fiona Thomas)': '#bbdb56', 'Multinesia Productions': '#badb56', 'Ngapuhi Filmworks': '#b8db56', 'The University of Hawai i at Mānoa': '#b7db56', 'Centro Costarricense de Producción Cinematográfica': '#b5db56', 'Hrvatska Radiotelevizija': '#b4db56', 'ICAIC': '#b2db56', 'A.B. Seahorse Films': '#b1db56', 'AMP Filmworks': '#afdb56', 'Česká televize': '#aedb56', 'Filmové studio Barrandov': '#acdb56', 'Nordisk Film Denmark': '#abdb56', 'A.S.A.P': '#a9db56', 'ARTE': '#a7db56', 'Canal+': '#a6db56', "Centre du Cinéma et de l'Audiovisuel de la FWB": '#a4db56', 'Conti Films': '#a3db56', 'Coopération Belge au Développement': '#a1db56', 'Eurimages': '#a0db56', 'Les films d'en Face': '#9edb56', "Man's Films": '#9ddb56', 'Maïa Cinéma': '#9bdb56', 'Media Programme of the European Community': '#9adb56', 'Moustapha Elmi': '#98db56', 'Radiodiffusion Television Djiboutienne': '#97db56', 'Samawada Films': '#95db56', 'Vlaams Audiovisueel Fonds': '#94db56', 'ZDF': '#92db56', 'Banyan Production': '#91db56', 'Bardouille Productions': '#8fdb56', 'Deschamps Corporation': '#8ddb56', 'Documentree Films': '#8cdb56', 'Page 2 Productions': '#8adb56', 'Studio Anansi': '#89db56', 'The University of the West Indies': '#87db56', 'goPOP Films': '#86db56', 'Joendry Actor Films': '#84db56', 'DEFA': '#83db56', 'Retrogusto Films': '#81db56', 'Film Clinic': '#80db56', 'Meridiano 89': '#7edb56', 'Acige': '#7ddb56', 'BonyPictures': '#7bdb56', 'CCEM': '#7adb56', 'Centro Cultural de España en Malabo': '#78db56', 'Cotton Tree Productions': '#77db56', 'Amir Grafics': '#75db56', 'Baltic Film and Media School': '#73db56', 'Abramorama': '#72db56', 'BIP Films': '#70db56', 'Cinehollywood S.r.l.': '#6fdb56', 'DFI': '#6ddb56', 'Discovery Channel': '#6cdb56', 'Doc Society': '#6adb56', 'Fork Films': '#69db56', 'Independent Film Producers (IFP)': '#67db56', 'Intaba Creative': '#66db56', 'Loma Nasha': '#64db56', 'Original Creative Breed': '#63db56', 'Reeleyes Film': '#61db56', 'Scion Films': '#60db56', 'Shine Global': '#5edb56', 'The Bertha Foundation': '#5ddb56', 'Wah Film Productions Ltd.': '#5bdb56', 'Ethiopian Film Corporation': '#59db56', 'FITV': '#58db56', 'Santana Brothers': '#56db56', 'Super16': '#56db58', 'BE-FILMS': '#56db59', 'Big Band Story': '#56db5b', 'Palikū Documentary Films': '#56db5c', 'Angoa-Agicoa': '#56db5e', 'ITVS': '#56db5f', 'Metro-Goldwyn-Mayer': '#56db61', 'New World Disorder': '#56db62', 'PIC': '#56db64', 'PROCIREP': '#56db65', 'Pasifika Transmissions': '#56db67', 'Pictanovo': '#56db68', 'Reddy Productions': '#56db6a', 'Sacrebleu Productions': '#56db6b', 'Te Maka Productions Inc.': '#56db6d', 'Tim Bonython Productions':

```
'#56db6e', 'Unifocus Pictures': '#56db70', 'Wan Smolbag Theatre': '#56db72',
'qWaves.com': '#56db73', 'YLE': '#56db75', 'Palaviré Productions': '#56db76',
'In Vivo Prod': '#56db78', 'Fulmar West': '#56db79', 'Adélaïde Productions':
'#56db7b', 'Ce.Na.Ci.': '#56db7c', 'Centre National du Cinema du Gabon':
'#56db7e', 'IGIS': '#56db7f', "Société Logovéenne du Film et de l'Image Pictures
(SLOGF)": '#56db81', 'Blue Monday Productions': '#56db82', 'Downsideup Film':
'#56db84', 'Film Project Gambia': '#56db85', 'Gunjur Film and Drama Production':
'#56db87', 'Gunjur Productions': '#56db88', 'Koollife Entertainment': '#56db8a',
'Lotus-Film': '#56db8c', 'Nomad Earth Media': '#56db8d', 'Nowness': '#56db8f',
'Young Stars Movie Production': '#56db90', 'dfm Fiktion': '#56db92',
'Georgianfilm': '#56db93', 'Obibini Pictures': '#56db95', 'Aderyn Productions':
'#56db96', 'Audio Visual Enterprises': '#56db98', 'Craig T. Rumar Productions':
'#56db99', 'Mulloway Limited': '#56db9b', 'Serra & Becerra Productions':
'#56db9c', 'Warner Bros. Pictures': '#56db9e', 'Greek Film Centre': '#56db9f',
'NukkiNukki Productions': '#56dba1', 'Mirror and Hammer Films': '#56dba2',
'Quarless Video Concepts': '#56dba4', 'Caraibe Films Compagnie': '#56dba6',
'Project: Inspire': '#56dba7', 'Casa Comal': '#56dba9', 'Cinematografica Tikal':
'#56dbaa', 'Arco Iris': '#56dbac', 'LX Filmes': '#56dbad', "Les Films de
l'Après-Midi": '#56dbaf', 'Spectre Productions': '#56dbb0', 'Guyana
Entertainment': '#56dbb2', 'Mahadeo Shivraj Productions': '#56dbb3', 'Nivelli
International': '#56dbb5', 'The Victor Jara Collective': '#56dbb6', 'Whitehorse
Films': '#56dbb8', 'Astrolabio': '#56dbb9', 'Cinemateca Universitaria Enrique
Ponce Garay': '#56dbbb', 'Shaw Brothers': '#56dbbd', 'Mafilm': '#56dbbe', 'RÚV':
'#56dbc0', 'SVF Entertainment': '#56dbc1', 'Starvision Plus': '#56dbc3',
'Kanoon': '#56dbc4', 'Iraqi Cinema and Theater Foundation': '#56dbc6', 'Fis
Éireann/Screen Ireland': '#56dbc7', 'United King Films': '#56dbc9', 'RAI':
'#56dbca', 'Abyssa Films': '#56dbcc', 'Toei Company': '#56dbcd', 'Baalbek
studio': '#56dbcf', 'Sooraj': '#56dbd0', 'Kazakhfilm Studios': '#56dbd2', 'One
Fine Day Films': '#56dbd3', 'Ironbound Films': '#56dbd5', 'OneMinutesJr':
'#56dbd7', 'Phoenix Islands Protected Area': '#56dbd8', 'Television Trust for
the Environment': '#56dbda', 'Qendra Kinematografike e Kosovës': '#56dadb',
                     ': '#56d9db', 'Kyrgyzfilm': '#56d7db', 'Lao
New Wave Cinema Productions': '#56d6db', 'Latvijas Kultūras akadēmija':
'#56d4db', 'Abbout Productions': '#56d3db', 'Day Zero Film and Video':
'#56d1db', 'Mokoari Street': '#56d0db', 'A Tarr woryonwon company': '#56cedb',
'Appian Way': '#56cddb', 'Big World Cinema': '#56cbdb', 'Bold Films': '#56cadb',
'Docster': '#56c8db', 'DoubleH Entertainment': '#56c7db', 'Fox Tale Films':
<code>'#56c5db'</code>, <code>'Fusion Productions'</code>: <code>'#56c4db'</code>, <code>'Gabriel Films'</code>: <code>'#56c2db'</code>, <code>'Ink & </code>
Pepper': '#56c0db', 'Just Media': '#56bfdb', 'Lunar Films': '#56bddb', 'Mami
Wata Pekin Productions': '#56bcdb', 'Nine Stories Productions': '#56badb',
'Obrigado Productions': '#56b9db', 'RYOT Films': '#56b7db', 'SVT': '#56b6db',
'Unity Film Productions': '#56b4db', 'Vision Film Company': '#56b3db', 'Vulcan
Productions': '#56b1db', 'Filmco International Productions': '#56b0db', 'Michael
Arthur Films': '#56aedb', 'Prodif Ets.': '#56addb', 'Lithuanian Academy of Music
and Theatre': '#56abdb', 'Samsa Film': '#56aadb', 'Pontus Maximus Productions':
'#56a8db', 'Laterit Productions': '#56a6db', 'BAM Studios': '#56a5db', 'BBC
Film': '#56a3db', 'BFI': '#56a2db', 'First Dawn Arts': '#56a0db', 'Freedom From
Fistula Foundation': '#569fdb', 'Group Theory': '#569ddb', 'Ken Anderson Films':
```

'#569cdb', 'Makropol': '#569adb', 'Participant': '#5699db', 'Potboiler Productions': '#5697db', 'Purple Field Productions': '#5696db', 'Satellite Entertainment': '#5694db', 'Satellite multimedia': '#5693db', 'Vertical Ascent': '#5691db', 'Astro Shaw': '#5690db', 'Dark Rain Entertainment': '#568edb', 'Les Films Cissé': '#568cdb', 'Latina Pictures': '#568bdb', 'Microwave Films': '#5689db', 'R.F.O.': '#5688db', 'Les Films Soleil O': '#5686db', 'Caméléon Production': '#5685db', 'Vice-Rectorat de Mayotte': '#5683db', 'Instituto Mexicano de Cinematografía': '#5682db', 'Association pour la Création Cinématographique de Monaco': '#5680db', 'Check Productions': '#567fdb', 'France 2 Cinéma': '#567ddb', 'Mongolkino': '#567cdb', 'Radiotelevizija Crne Gore': '#567adb', 'Living Letters Productions': '#5679db', 'University of Pittsburgh': '#5677db', 'Centre Cinématographique Marocain': '#5676db', 'Ébano Multimédia': '#5674db', 'Triple Six Movie Production': '#5672db', 'Collective Productions Namibia': '#5671db', 'The Commonwealth Film Unit': '#566fdb', 'Shital shijan Digital Pvt Ltd': '#566edb', 'VPRO': '#566cdb', 'A-Film Distribution': '#566bdb', 'A-Film Home Entertainment': '#5669db', 'Blaq Out': '#5668db', 'CCCP': '#5666db', 'CG Entertainment': '#5665db', 'CZAR Belgium': '#5663db', 'Camera Obscura': '#5662db', 'Cosmic Illusion Productions': '#5660db', 'FishEye Feature Films': '#565fdb', 'Independent Theatre Foundation St. Martin': '#565ddb', 'Inspire Pictures': '#565cdb', "L'Altrofilm": '#565adb', 'Legend Films International': '#5658db', 'Manteka Studios': '#5657db', 'Mercurio Cinematografica': '#5756db', 'Metro Tartan Distribution Ltd.': '#5956db', 'Palisades Tartan': '#5a56db', 'Pietra Production Films': '#5c56db', 'Provider Films': '#5d56db', 'Siren Visual Entertainment': '#5f56db', 'Storytellers Film & TV': '#6056db', 'The Netherlands Information Bureau': '#6256db', 'United Artists': '#6356db', 'symbol productions': '#6556db', 'New Zealand Film Commission': '#6656db', 'Instituto Nicaraguense de Cine (INCINE)': '#6856db', 'Cabas Film': '#6956db', 'Iroko Films': '#6b56db', 'University of Waikato': '#6c56db', 'Eye Design Films': '#6e56db', 'Korean Film': '#6f56db', 'Makedonska Televizija': '#7156db', 'D&R Visuals': '#7356db', 'Kitty Films': '#7456db', 'Open Boat Film': '#7656db', 'Sisbro Studio': '#7756db', 'Norsk Film': '#7956db', 'AVRO Television': '#7a56db', 'Cassette for Timescapes': '#7c56db', 'Film Solutions': '#7d56db', 'IDTV Film': '#7f56db', 'Nebula Productions': '#8056db', 'Oman Art Production Company': '#8256db', 'Oman Film & Theatre Society': '#8356db', 'Oman Film Society': '#8556db', 'Required Reading Productions': '#8656db', 'Sharjah Art Foundation': '#8856db', '7th Sky Entertainment': '#8a56db', 'Momina Duraid Productions': '#8b56db', 'Jaguar Films': '#8d56db', 'Institute of Papua New Guinea Studios': '#8e56db', 'Sabaté Films': '#9056db', 'Tondero': '#9156db', 'Viva Films': '#9356db', 'Kinomachina': '#9456db', 'PWSFTviT': '#9656db', 'TVP': '#9756db', 'Divedco': '#9956db', 'Film Unit of the Division of Community Education, Puerto Rico': '#9a56db', 'Simpals': '#9c56db', 'Studioul Cinematografic București': '#9d56db', 'CTB Film Company': '#9f56db', 'Almond Tree Films': '#a056db', 'A Private View': '#a256db', 'Bobi Lux': '#a456db', 'C.L.A.P.': '#a556db', 'Calren Studios': '#a756db', 'Film i Väst': '#a856db', 'La Cerise Sur Le Gâteau': '#aa56db', 'Latzel film- & tvprduction': '#ab56db', 'Le Fresnoy': '#ad56db', 'Lemming Film': '#ae56db', "Les Films du Haut d'un Arbre": '#b056db', 'Lithops Films': '#b156db', 'Ludwig Kameraverleih': '#b356db', 'Plattform Produktion': '#b456db', 'Région Réunion':

'#b656db', 'Tala Media': '#b756db', 'Tiktak Production': '#b956db', 'VMP Films': '#ba56db', 'We Film': '#bc56db', 'trust your ears': '#be56db', 'Granada Television': '#bf56db', 'Warner Sisters Productions': '#c156db', 'Wendover Productions': '#c256db', 'MSR Media': '#c456db', 'Allbiz': '#c556db', 'Iland Boy Films': '#c756db', 'Natural Groove Pictures': '#c856db', 'Nexus Networks': '#ca56db', 'Untold Stories': '#cb56db', 'Video Associates': '#cd56db', 'Warwick Productions': '#ce56db', 'Wiseman Media': '#d056db', 'Cinémaginaire': '#d156db', 'Epithète films': '#d356db', 'France 3 Cinéma': '#d456db', 'Pierre Films': '#d656db', 'Red Ochre Productions': '#d856db', 'SODEC': '#d956db', 'Société Radio-Canada': '#db56db', 'Téléfilm Canada': '#db56d9', 'Wavelight Productions': '#db56d8', 'Shadowhand Productions': '#db56d6', 'Blueskin Films': '#db56d5', 'm2s1 films': '#db56d3', 'CINEPHIL San Marino': '#db56d2', "Cooperativa Mmezzogiorno Nuovo D'Italia": '#db56d0', 'Instituzione della Cinema di San Marino': '#db56cf', 'BNK Best Movies': '#db56cd', 'Telfaz11': '#db56cc', 'Almadies Films': '#db56ca', 'Cinekap': '#db56c9', 'Fakultet Dramskih Umetnosti Beograd (FDU)': '#db56c7', 'Radiotelevizija Beograd': '#db56c6', 'Puzzle Piece Productions': '#db56c4', 'Radio Television Seychelles': '#db56c3', 'SBC': '#db56c1', 'Braddock Films': '#db56bf', 'Do Global Productions': '#db56be', 'Jimmy Bangura Group': '#db56bc', 'King J Media Production': '#db56bb', 'Paradise Films': '#db56b9', 'Photo Planet': '#db56b8', 'The Freetown Media Center': '#db56b6', 'Try Hard Films': '#db56b5', 'WLIW New York': '#db56b3', 'WeOwnTV': '#db56b2', 'Malay Film Productions Ltd.': '#db56b0', 'RTVS': '#db56af', 'RTV Slovenija': '#db56ad', "Day's Edge Productions": '#db56ac', 'Dreamcast Theater Solomon Islands': '#db56aa', 'Jackhead': '#db56a9', 'OMS Film Group': '#db56a7', 'Sukwadi Media': '#db56a5', 'Wantok Stori': '#db56a4', 'Distant Horizon': '#db56a2', 'Untitled Film Works': '#db56a1', 'CJ Entertainment': '#db569f', 'Weltfilm': '#db569e', 'TVE': '#db569c', 'Cine-FilmsLanka': '#db569b', 'Maharaja Entertainments (Pvt.) Ltd': '#db5699', 'Philistine Films': '#db5698', 'The Back Lot Foundation': '#db5696', 'Aurora Filmlyd': '#db5695', 'C.R.I.M. Produções': '#db5693', 'SRF': '#db5692', 'Komîna Fîlm a Rojava': '#db5690', 'Central Motion Picture Corporation': '#db568f', 'Tajikfilm': '#db568d', 'Sahamongkolfilm': '#db568b', 'FairTrade Films': '#db568a', 'Amanou Production': '#db5688', 'New Zealand National Film Unit': '#db5687', 'Malosi Pictures': '#db5685', 'Escuela Internacional de Cine y Televisión (EICTV)': '#db5684', 'Fundación del Nuevo Cine Latinoamericano': '#db5682', 'Cinétéléfilms': '#db5681', 'Erler Film': '#db567f', 'Turkmenfilm': '#db567e', 'Regent Entertainment': '#db567c', 'HesseGreutert Film': '#db567b', 'Mare Imbrium': '#db5679', 'Savannah College of Art and Design': '#db5678', 'American Film Institute (AFI)': '#db5676', 'Garza Films': '#db5674', 'U.C.E. Productions': '#db5673', 'Varna Films': '#db5671', 'Mosfilm': '#db5670', 'Ramon Film Productions': '#db566e', 'Ukranimafilm': '#db566d', 'Imagenation Abu Dhabi FZ': '#db566b', 'Kijiweni Productions': '#db566a', 'Cedarlark': '#db5668', 'RTL': '#db5667', 'Imágenes (Uy)': '#db5665', 'Uzbekfilm': '#db5664', 'Wan Smolbag Films': '#db5662', 'Centro Televisivo Vaticano': '#db5661', 'Hãng Phim Truyện Việt Nam': '#db565f', 'France 3': '#db565e', 'France Télévisions': '#db565c', 'Mérapi Productions': '#db565a', 'Wallis et Futuna 1ere': '#db5659', 'Prominent Television': '#db5657', 'Adenium Productions': '#db5756', 'Avala Film': '#db5856', 'EK Films': '#db5a56', 'JBA Production': '#db5b56', 'Nyerai

#### Films': '#db5d56'}

```
fig = px.choropleth(
    highest_count_studios,
    locations='country',
    locationmode='country names',
    color='studio',
    title='Los estudios más populares del mundo',
    hover_data=['count', 'percentage'],
    color_discrete_map=studio_colors_hex
)

fig.update_layout(
    coloraxis_colorbar=dict(title='Studio'),
    width=700,
    height=500,
)
fig.show()
```

Los estudios más populares del mundo

