Data Visualization of Olympic Games Tokyo 2020

In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
import warnings
warnings.filterwarnings('ignore')
```

Data Visualization

Athletes

List of Name of Athletes playing in more than one discipline are

In [2]:
```
df_a=pd.read_excel('Athletes.xlsx')
a=df_a[['Name','Discipline']].groupby(df_a['Name']).agg('count')
a=a[a['Name']==2]
a.drop(columns=['Discipline'],axis=1,inplace=True)
a.rename(columns={"Name":'No. of Discipline'})
```

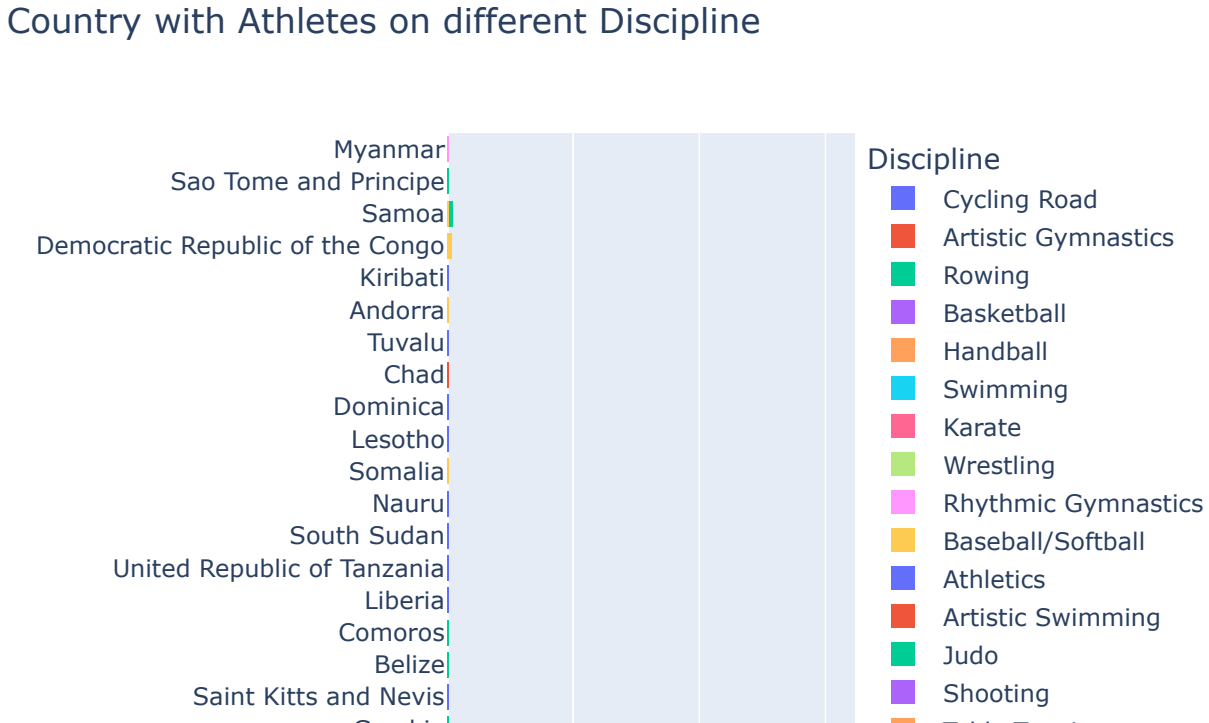
Out[2]:

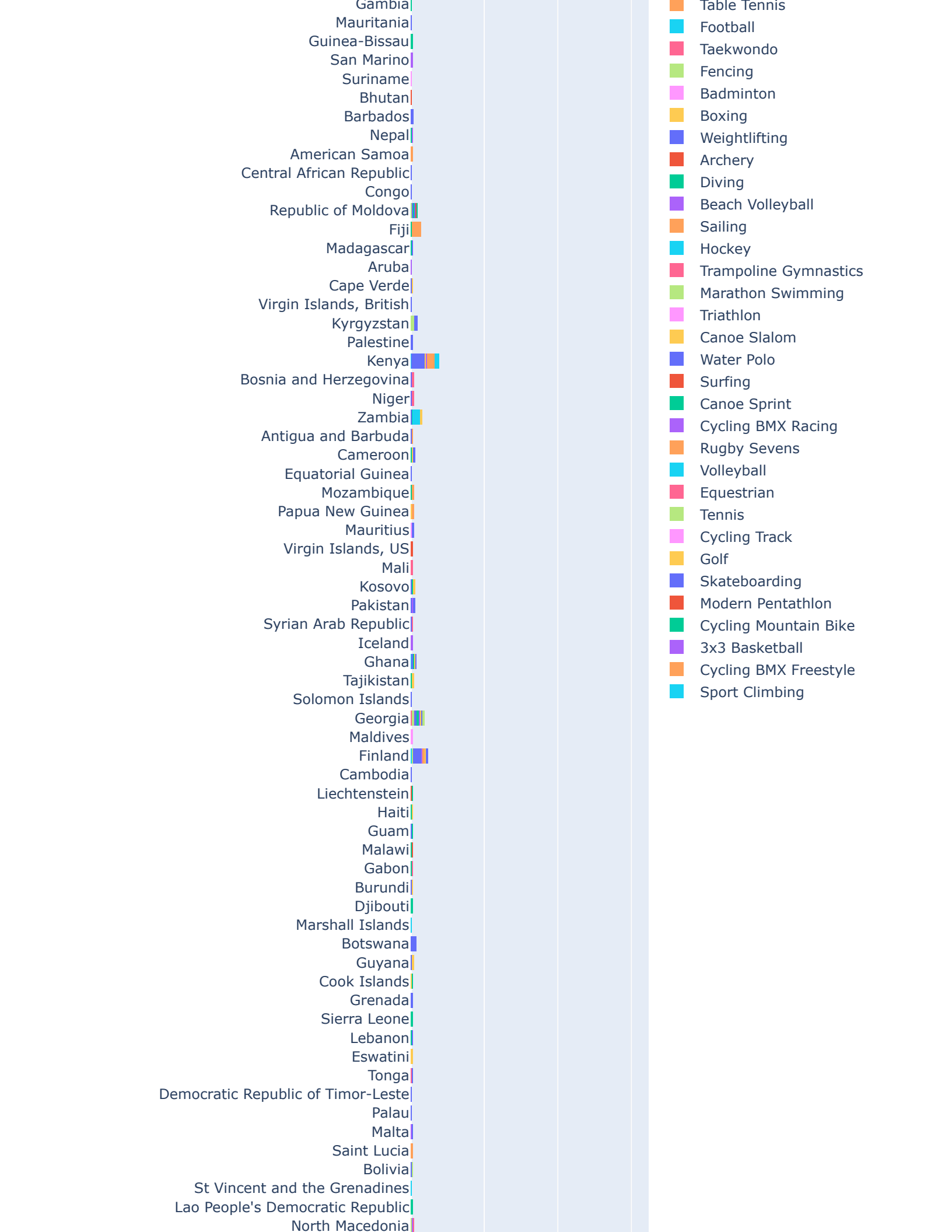
Name	No. of Discipline
ALI Mohamed	2
ALVAREZ Jorge	2
CHEN Yang	2

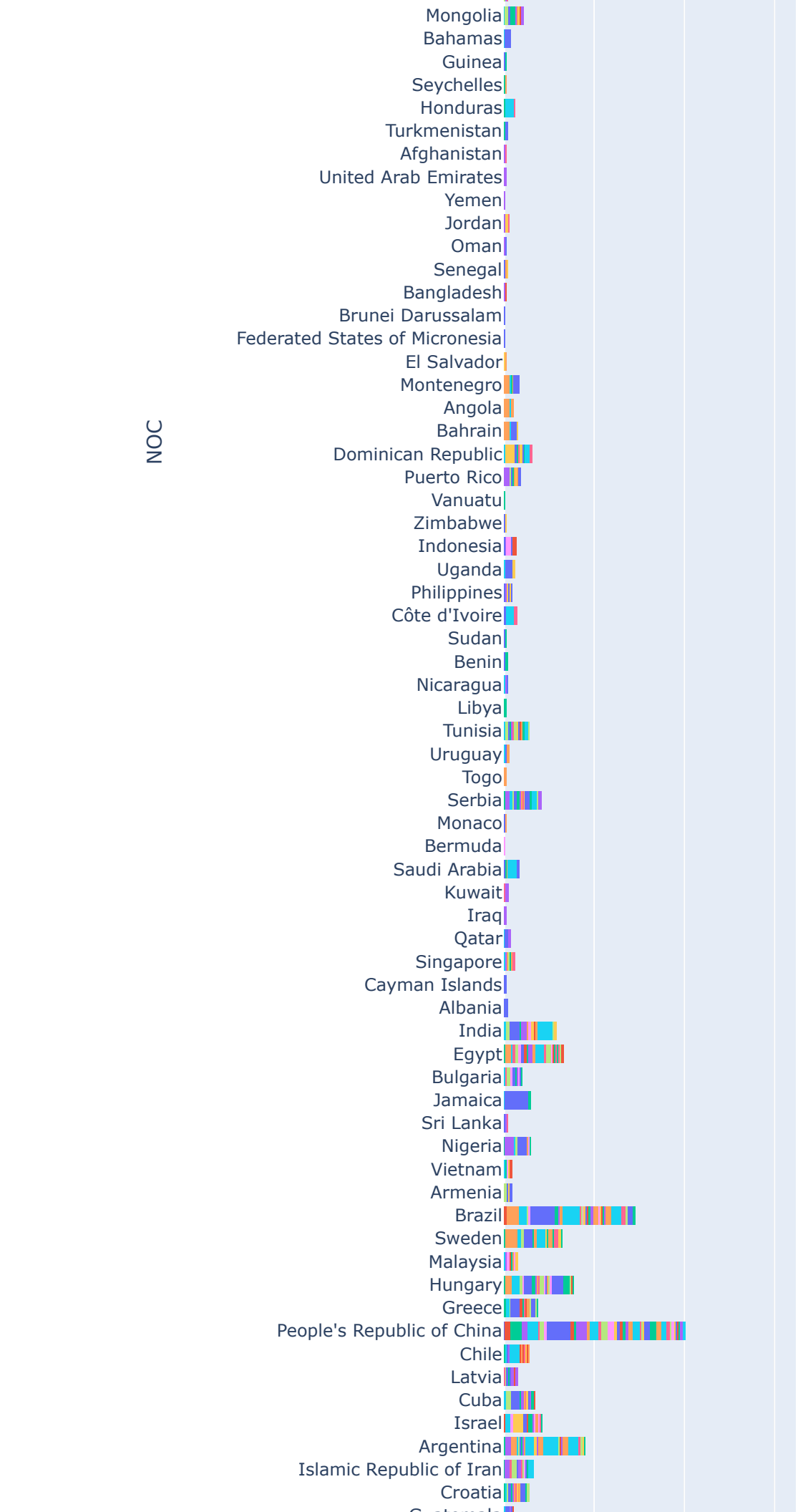
DYGERT Chloe	2
GANNA Filippo	2
HALL James	2
HAVIK Yoeri	2
KIM Hyunsoo	2
KOPECKY Lotte	2
KOVACS Zsofia	2
KURBANOV Ruslan	2
LI Qian	2
MARTIN Daniel	2
PALTRINIERI Gregorio	2
PEREZ Maria	2
PEREZ Paola	2
PORTELA Teresa	2
SUN Jiajun	2
WANG Yang	2
WATANABE Yuta	2
WELLBROCK Florian	2
ZHANG Xin	2
van ROUWENDAAL Sharon	2

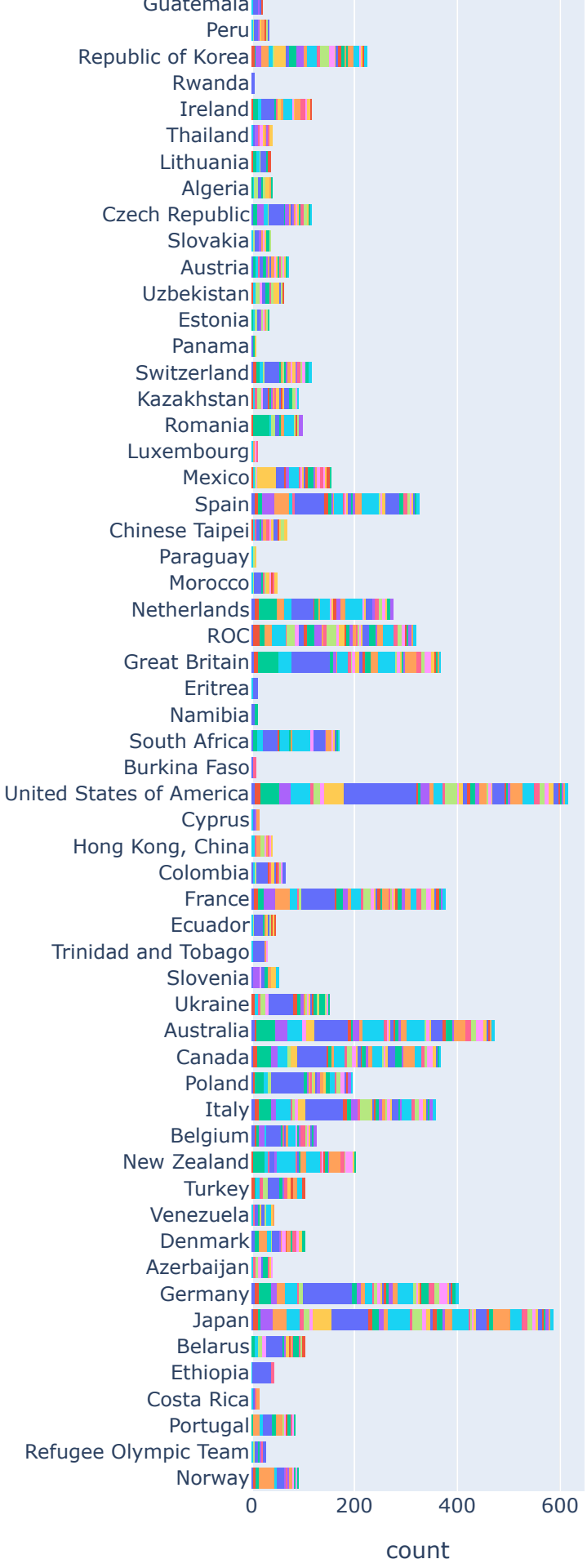
There are `23` Athletes from different country taking participate in two different `Discipline`.

In [3]:
```
px.histogram(df_a,y='NOC',color='Discipline',height=3500,title='Country with Athletes on
```



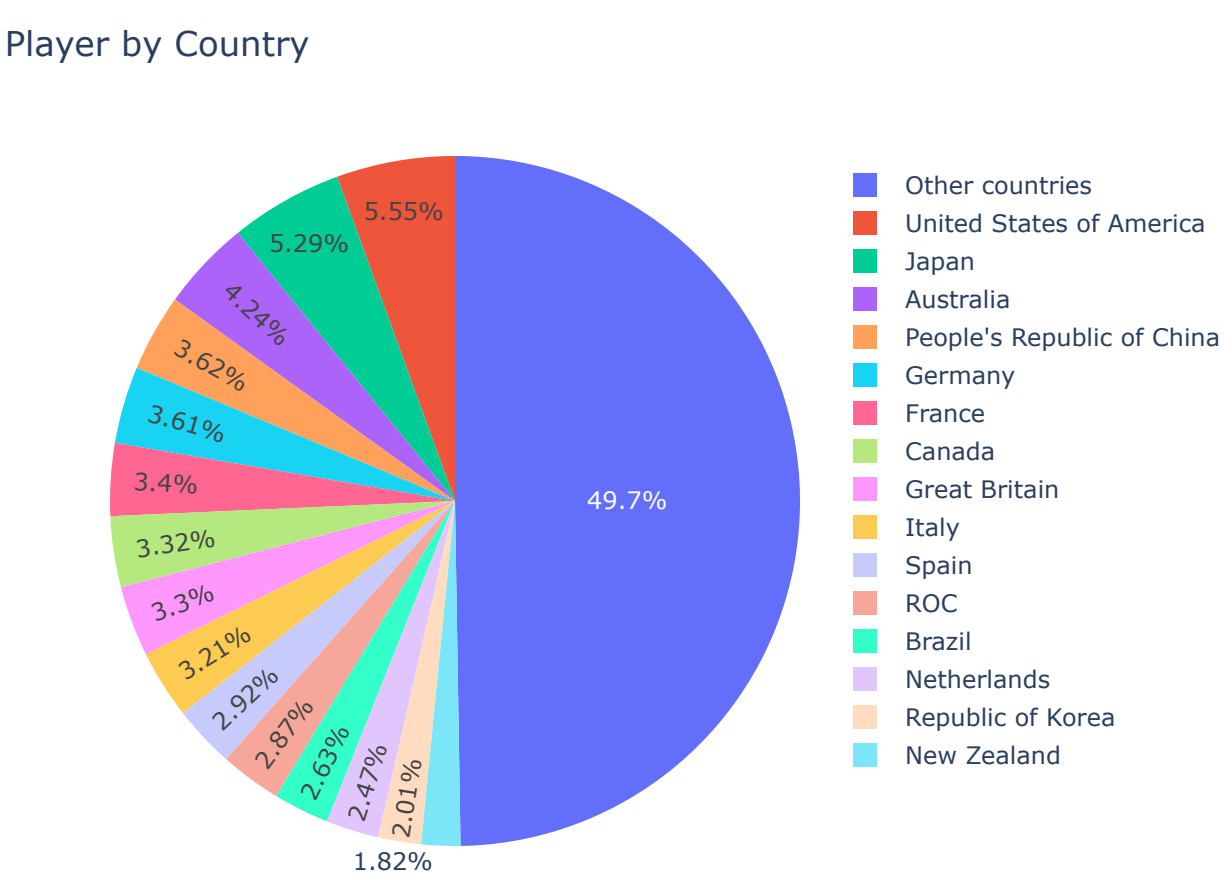

Gambia
Mauritania
Guinea-Bissau
San Marino
Suriname
Bhutan
Barbados
Nepal
American Samoa
Central African Republic
Congo
Republic of Moldova
Fiji
Madagascar
Aruba
Cape Verde
Virgin Islands, British
Kyrgyzstan
Palestine
Kenya
Bosnia and Herzegovina
Niger
Zambia
Antigua and Barbuda
Cameroon
Equatorial Guinea
Mozambique
Papua New Guinea
Mauritius
Virgin Islands, US
Mali
Kosovo
Pakistan
Syrian Arab Republic
Iceland
Ghana
Tajikistan
Solomon Islands
Georgia
Maldives
Finland
Cambodia
Liechtenstein
Haiti
Guam
Malawi
Gabon
Burundi
Djibouti
Marshall Islands
Botswana
Guyana
Cook Islands
Grenada
Sierra Leone
Lebanon
Eswatini
Tonga
Democratic Republic of Timor-Leste
Palau
Malta
Saint Lucia
Bolivia
St Vincent and the Grenadines
Lao People's Democratic Republic
North Macedonia
Table Tennis
Football
Taekwondo
Fencing
Badminton
Boxing
Weightlifting
Archery
Diving
Beach Volleyball
Sailing
Hockey
Trampoline Gymnastics
Marathon Swimming
Triathlon
Canoe Slalom
Water Polo
Surfing
Canoe Sprint
Cycling BMX Racing
Rugby Sevens
Volleyball
Equestrian
Tennis
Cycling Track
Golf
Skateboarding
Modern Pentathlon
Cycling Mountain Bike
3x3 Basketball
Cycling BMX Freestyle
Sport Climbing


NOC
Mongolia
Bahamas
Guinea
Seychelles
Honduras
Turkmenistan
Afghanistan
United Arab Emirates
Yemen
Jordan
Oman
Senegal
Bangladesh
Brunei Darussalam
Federated States of Micronesia
El Salvador
Montenegro
Angola
Bahrain
Dominican Republic
Puerto Rico
Vanuatu
Zimbabwe
Indonesia
Uganda
Philippines
Côte d'Ivoire
Sudan
Benin
Nicaragua
Libya
Tunisia
Uruguay
Togo
Serbia
Monaco
Bermuda
Saudi Arabia
Kuwait
Iraq
Qatar
Singapore
Cayman Islands
Albania
India
Egypt
Bulgaria
Jamaica
Sri Lanka
Nigeria
Vietnam
Armenia
Brazil
Sweden
Malaysia
Hungary
Greece
People's Republic of China
Chile
Latvia
Cuba
Israel
Argentina
Islamic Republic of Iran
Croatia



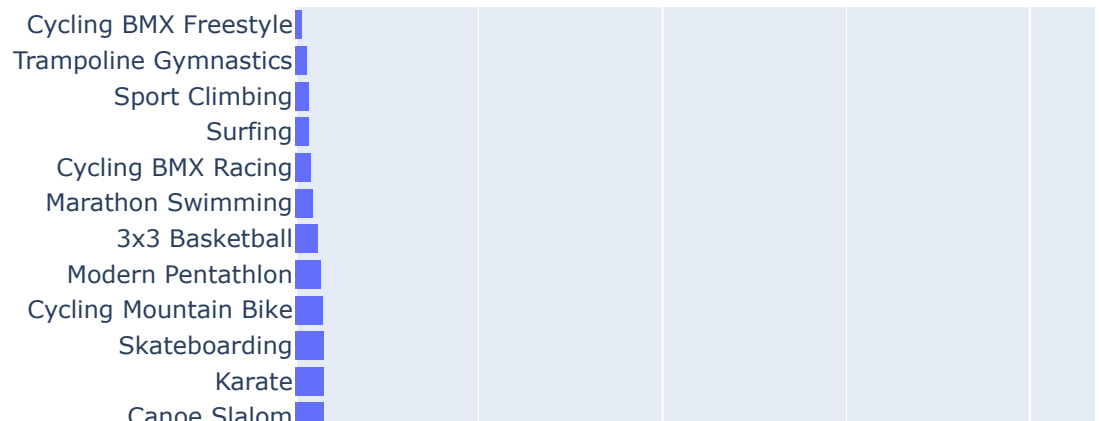
From above histogram, it is clear that `USA` , `Japan` and so on are having more player in Olympic and they all are participating in different Discipline.

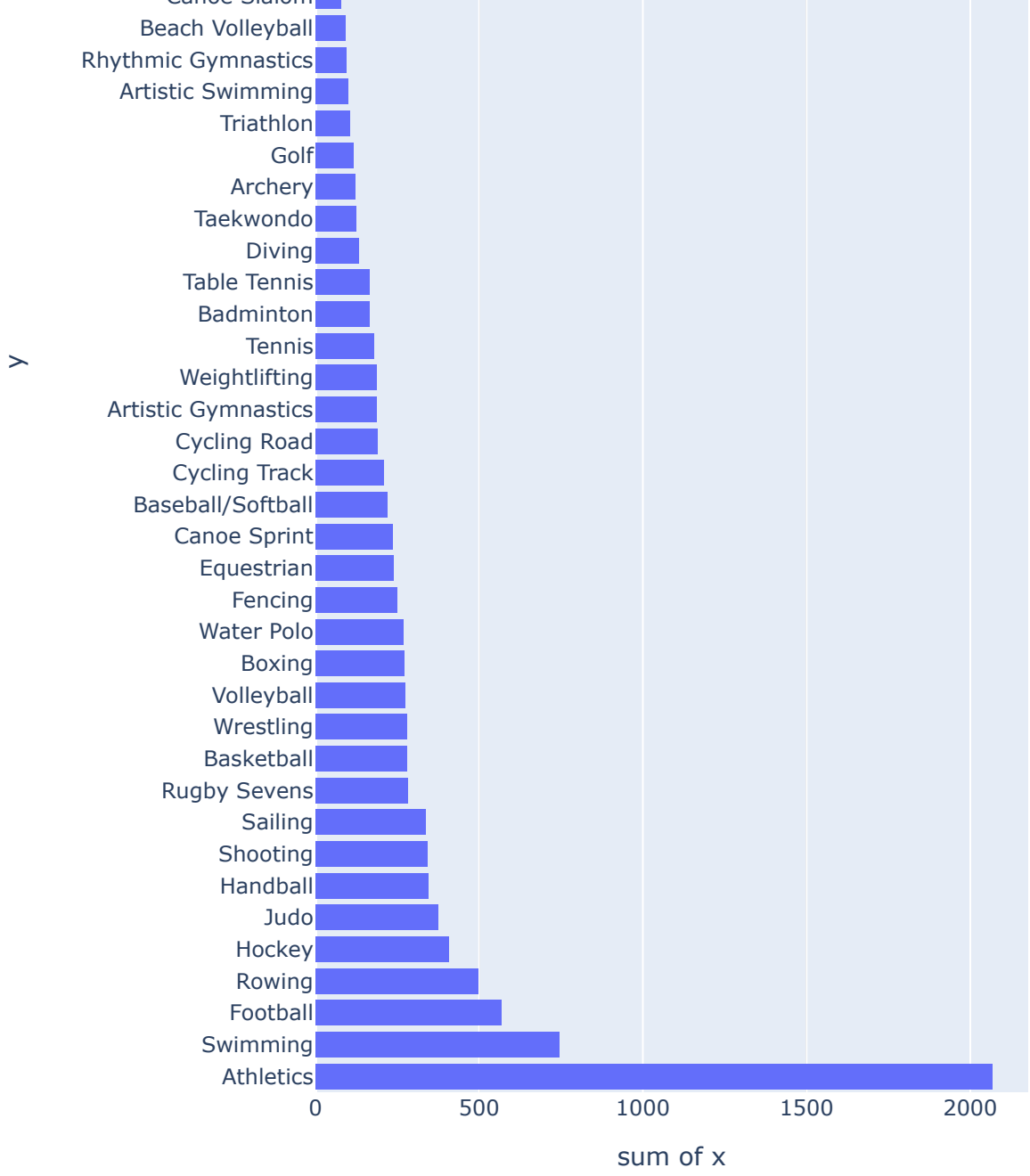
In [4]:
```
a=df_a['NOC'].value_counts()
df_a1=pd.DataFrame({'NOC':a.keys(),'Player':a.values})
df_a1.loc[df_a1['Player']<=200,'NOC']='Other countries'
px.pie(df_a1,values='Player',names='NOC',title='Player by Country')
```



From the above bar diagram, It is crystal and clear that most of the players are from `USA` , `Japan` , `Australia` and so on.

In [5]:
```
a=df_a['Discipline'].value_counts()
px.histogram(y=a.keys(),x=a.values,height=1000,title='Players participate in different D
```

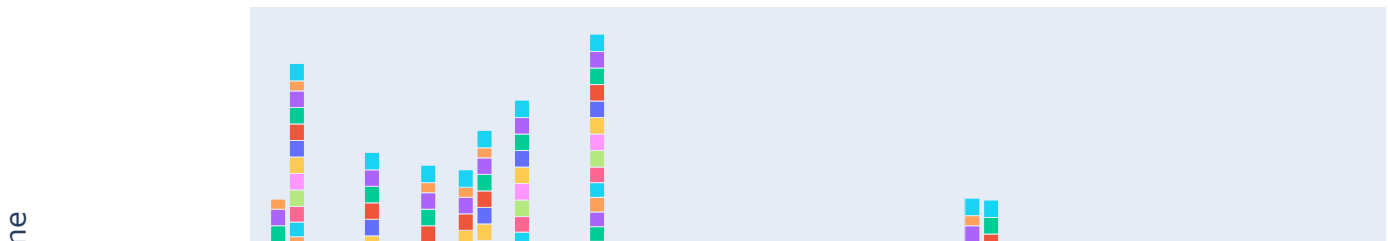


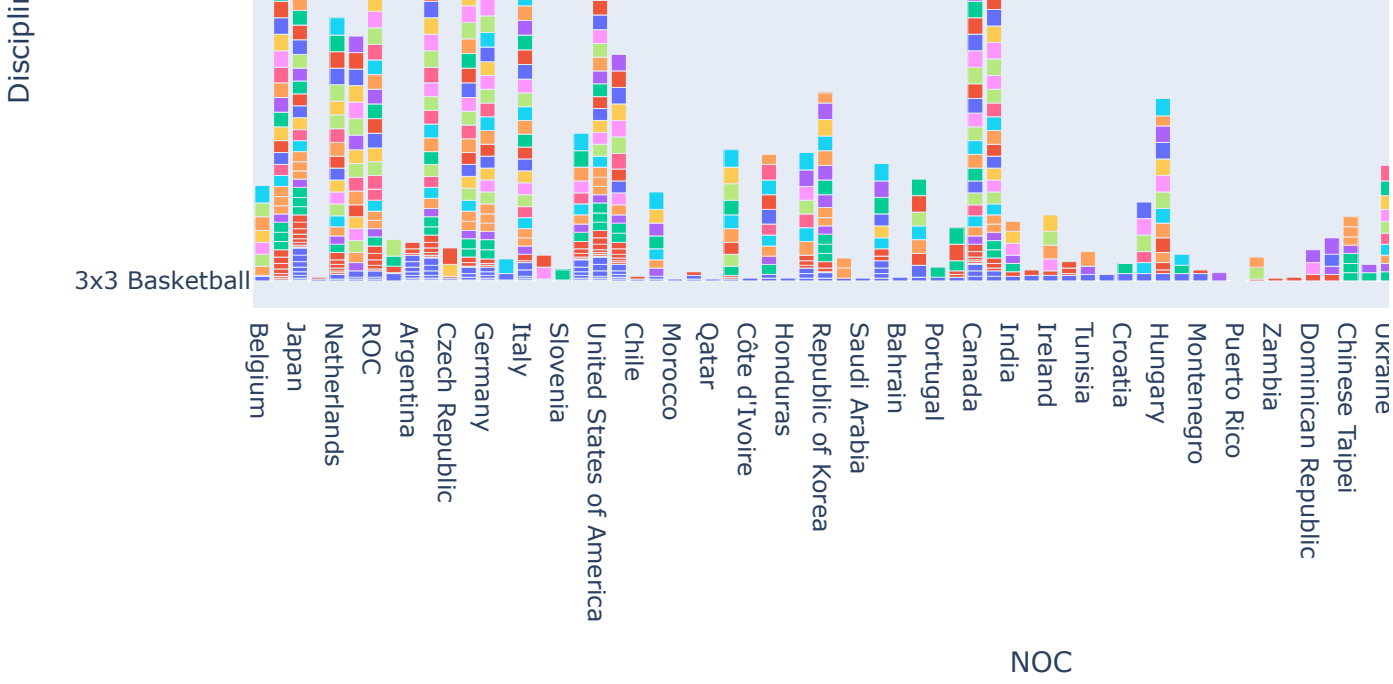


From the above visjalization, It is clear that maximum athletes are participate in `Athletics`, `Swimming`, `Football`, `Rowing` and so on.

Team

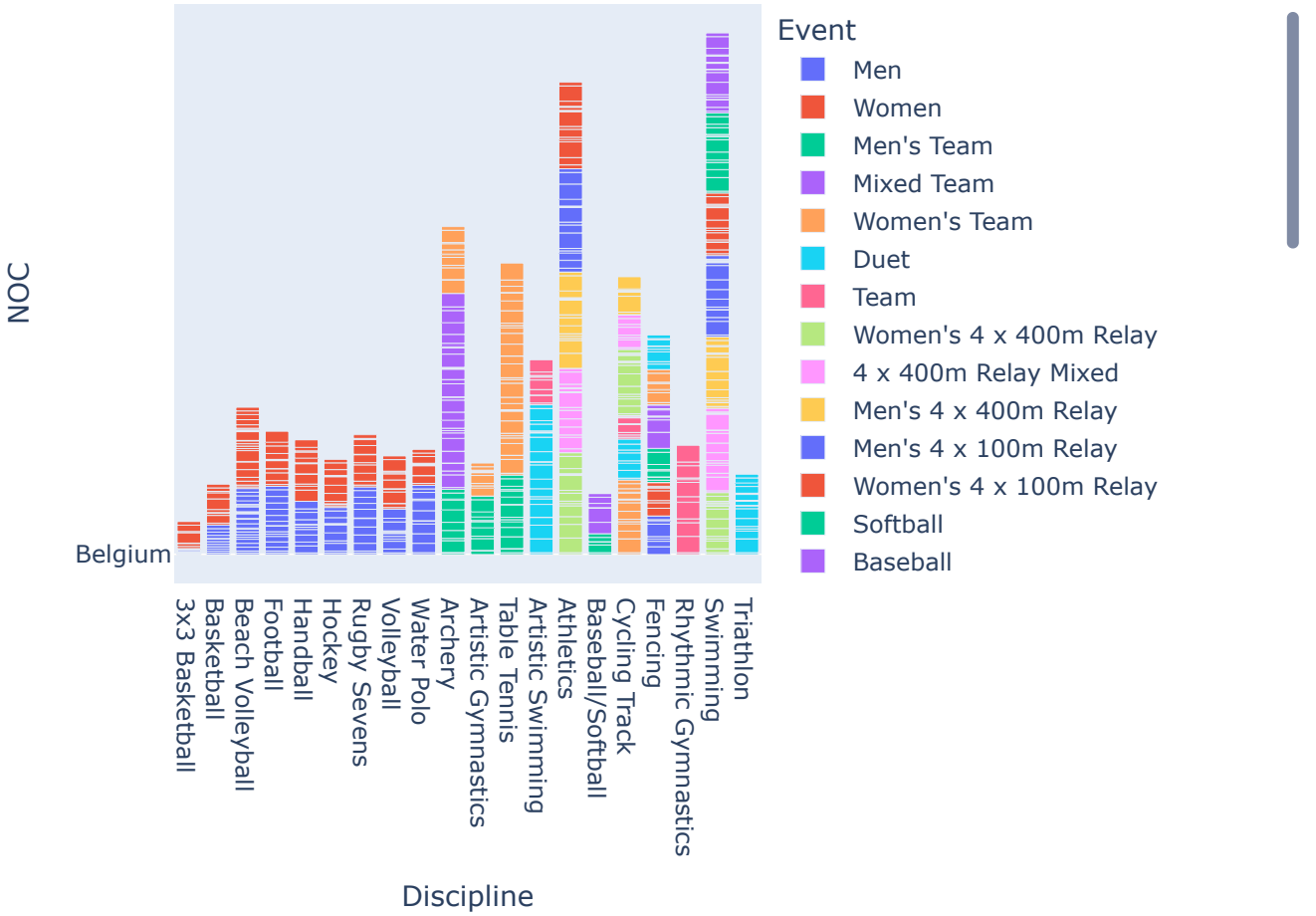
In [6]:
```
df_t=pd.read_excel('Teams.xlsx')
px.bar(df_t,x='NOC',y='Discipline',color='Event',width=1200)
```





From above bar diagram, It is clear that the `highest` number of team having with `USA`.

```
In [7]: px.bar(df_t,x='Discipline',y='NOC',color='Event')
```



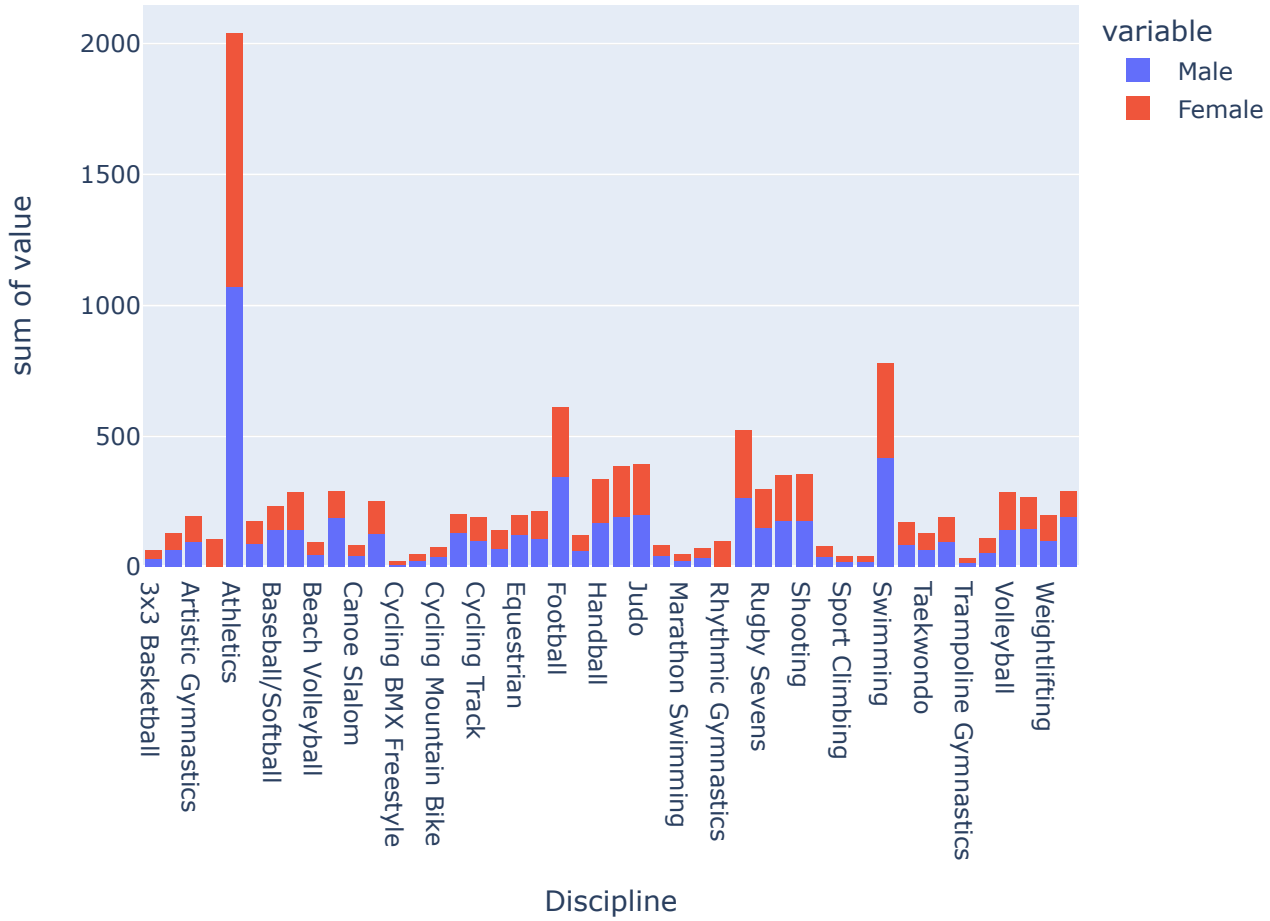
From above bar diagram, It is clear that `Swimming`, `Athletics`, `Archery` and so on having highest number of team in Olympic 2020.

Gender

In [8]:
```
df_g=pd.read_excel('EntriesGender.xlsx')
df_g.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 46 entries, 0 to 45
Data columns (total 4 columns):
 #   Column      Non-Null Count  Dtype 
---  ------      --------------  ----- 
 0   Discipline  46 non-null     object
 1   Female      46 non-null     int64 
 2   Male        46 non-null     int64 
 3   Total       46 non-null     int64 
dtypes: int64(3), object(1)
memory usage: 1.6+ KB
```

In [9]:
```
px.histogram(df_g,x="Discipline",y=['Male','Female'])
```



From the above histogram, it is clear that maximum male are participated in `Athletics`, `Swimming`, `Football` and so on where as maximum female are participated in `Athletics`, `Swimming`, `Football`, `Rowing` and so on.
Overall, the highest athletes are in `Athletics` discipline.

Coaches

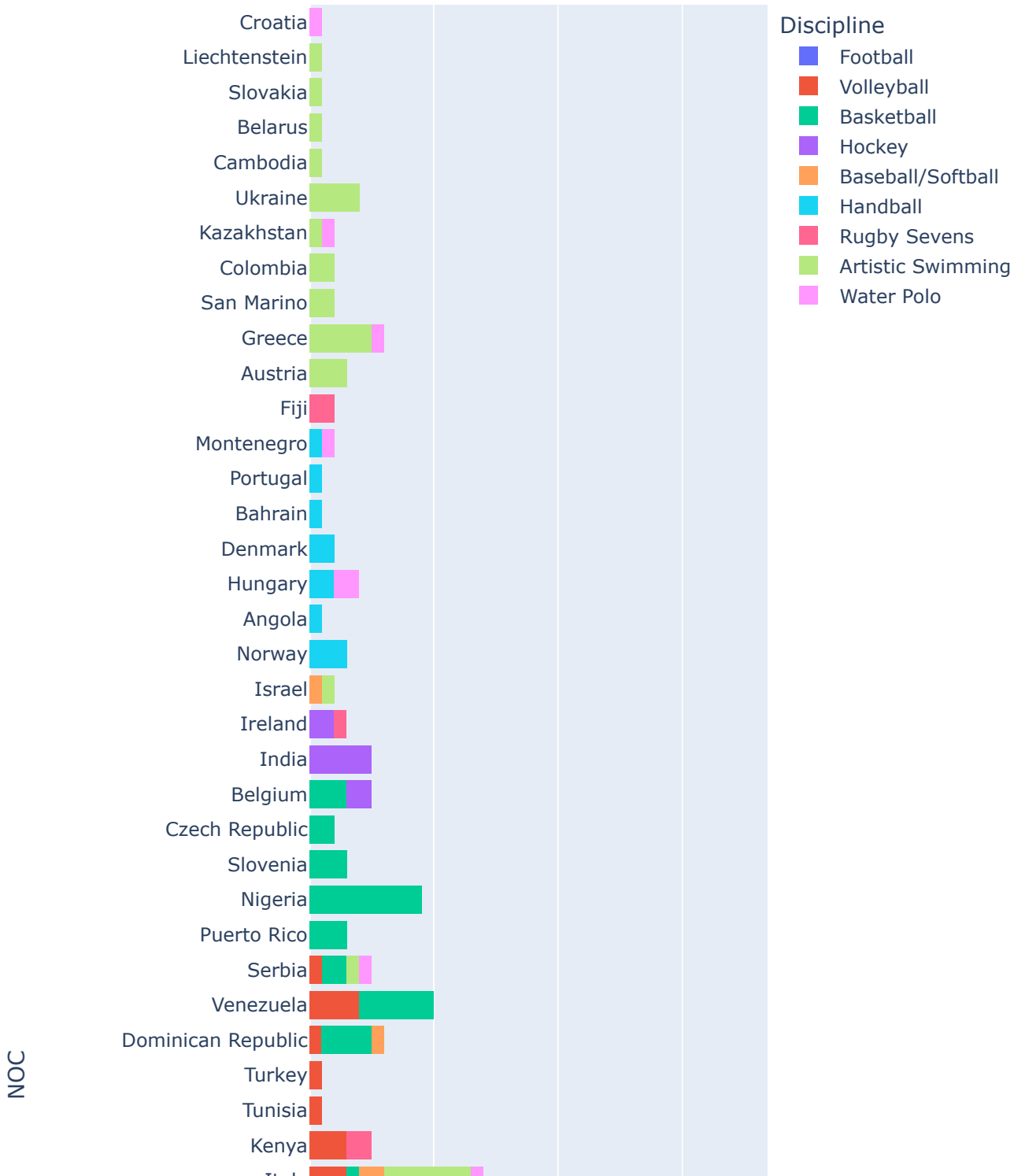
In [10]:
```
df_c=pd.read_excel('Coaches.xlsx')
```

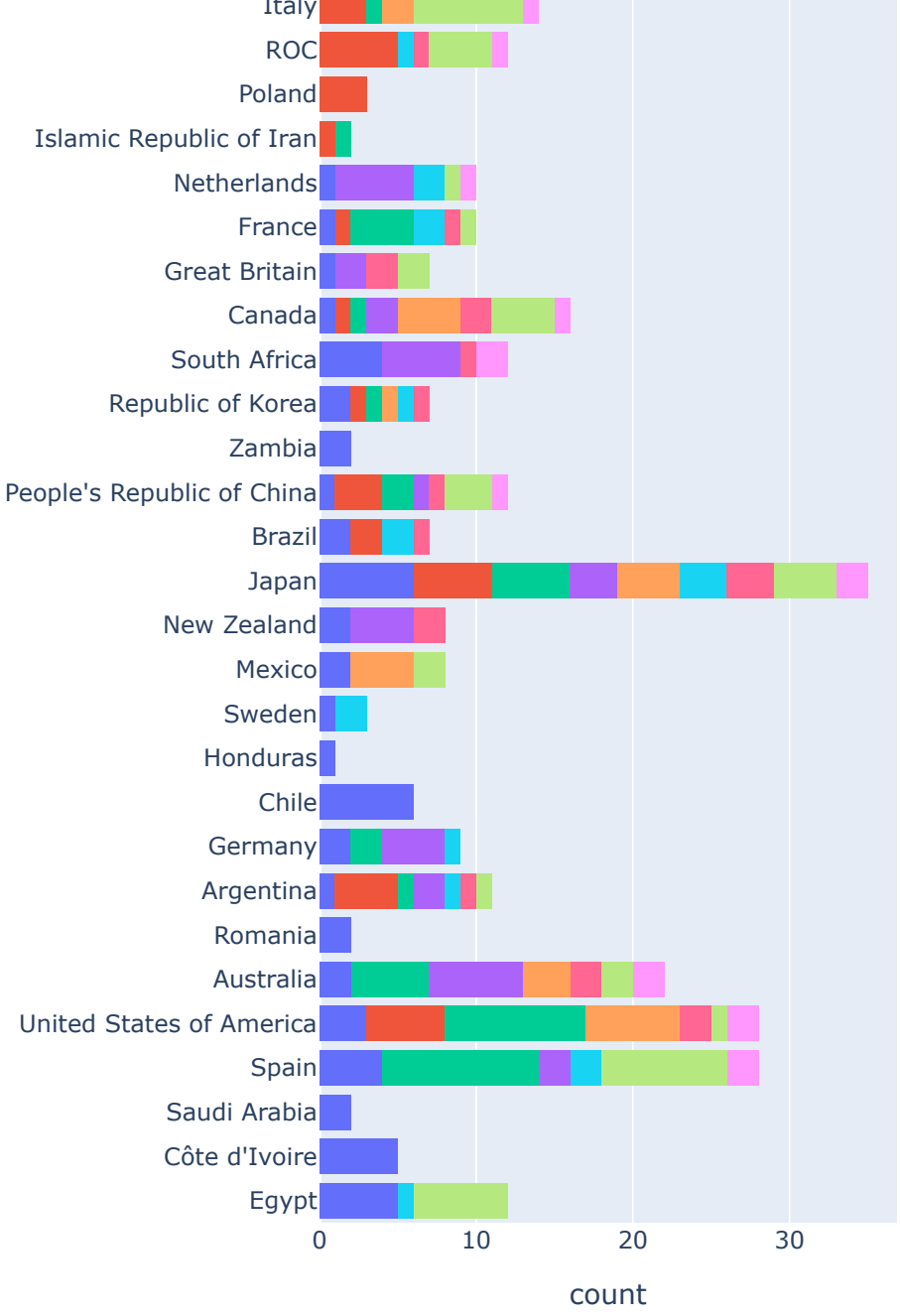
```
df_c.head()
```

Out[10]:

	Name	NOC	Discipline	Event
0	ABDELMAGID Wael	Egypt	Football	NaN
1	ABE Junya	Japan	Volleyball	NaN
2	ABE Katsuhiko	Japan	Basketball	NaN
3	ADAMA Cherif	Côte d'Ivoire	Football	NaN
4	AGEBA Yuya	Japan	Volleyball	NaN

In [11]:
```
px.histogram(df_c,y='NOC',color='Discipline',height=1500)
```





From the above histogram, highest number of coaches are with 'Japan' in Olympic 2020 whereas `USA` and `Spain` are in the Second highest position and `Australia` is at third position having maximum number of coaches.

Top 50 Higest Number of Coaches

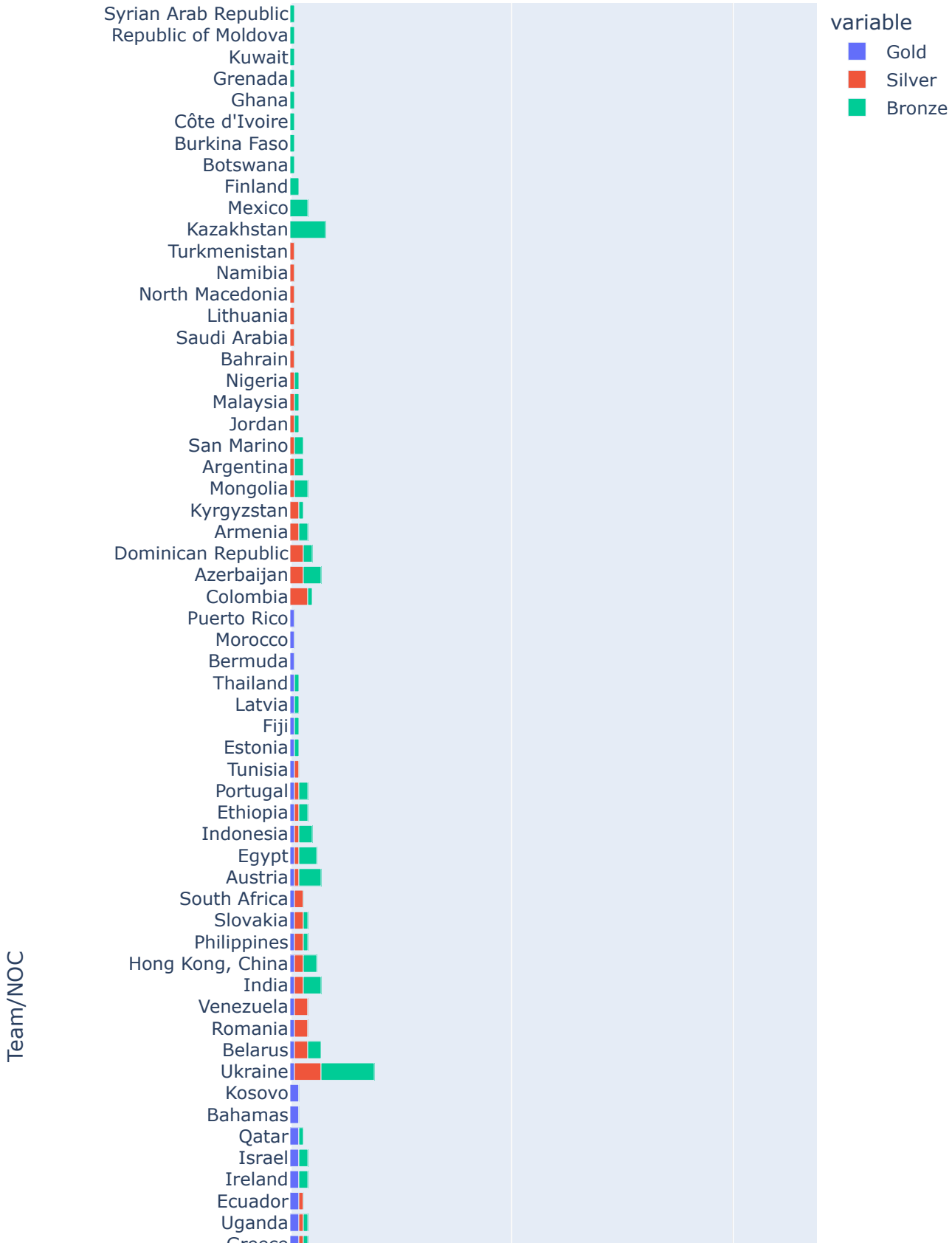
Medals

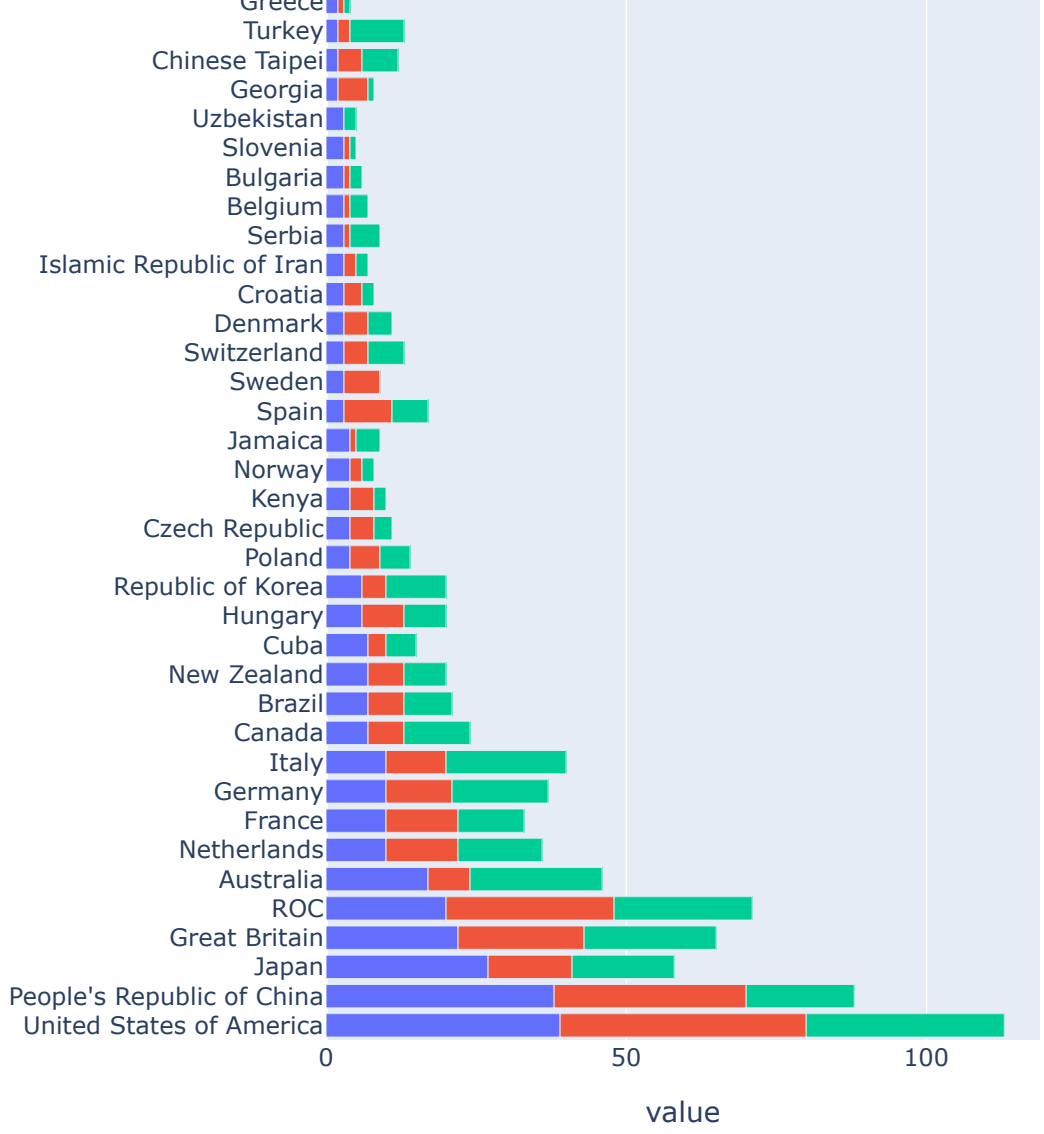
In [12]:
```
df_m=pd.read_excel('Medals.xlsx')
df_m.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 93 entries, 0 to 92
Data columns (total 7 columns):
 #   Column          Non-Null Count  Dtype 
---  ------          --------------  ----- 
 0   Rank            93 non-null     int64 
 1   Team/NOC        93 non-null     object
```

```
 2   Gold            93 non-null     int64
 3   Silver          93 non-null     int64
 4   Bronze          93 non-null     int64
 5   Total           93 non-null     int64
 6   Rank by Total   93 non-null     int64
dtypes: int64(6), object(1)
memory usage: 5.2+ KB
```

In [13]:
```
px.bar(df_m,y='Team/NOC',x=['Gold','Silver','Bronze'],height=1500)
```

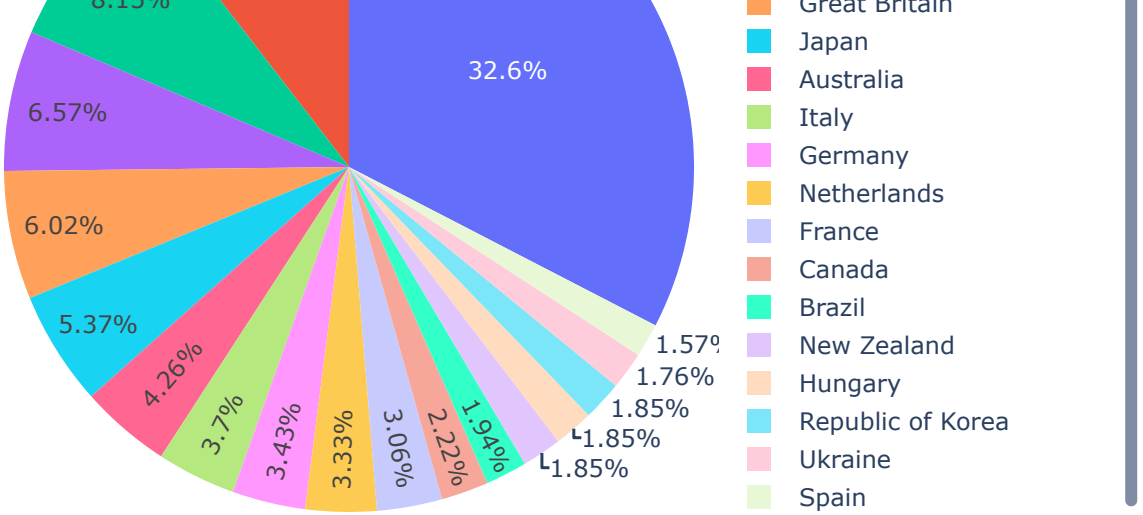




From the above histogram,

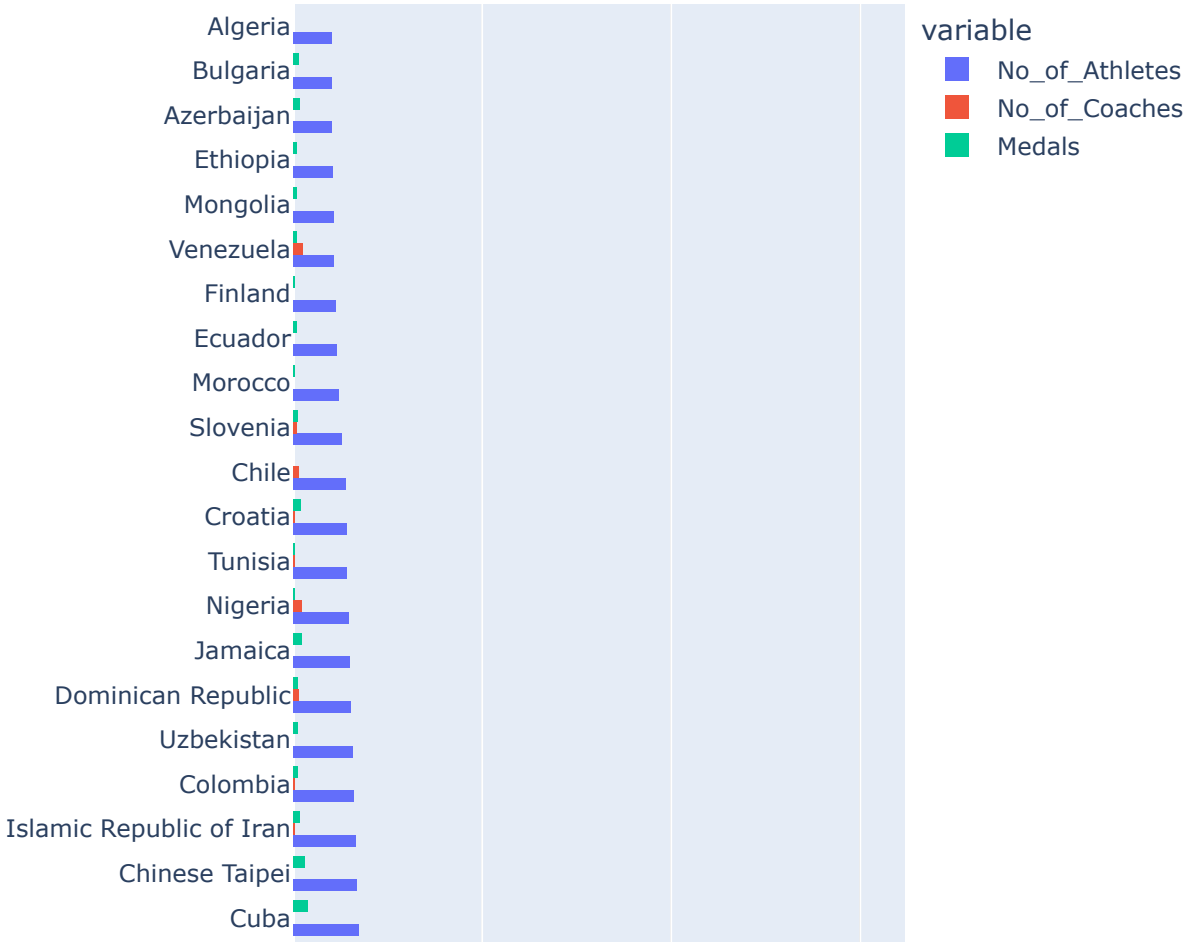
- It is clear that the highest number of Gold is won by `America`, second highest is `China` and third highest is `Japan`.
- The highest number of Silver is won by `America`, `China` the second highest and `Russia` the third highest.
- The highest number of Bronze is won by `America`, `Russia` the second highest and `Great Britain` and `Australia` is the third.

In [14]:
```
df_m1=df_m.copy()
df_m1.loc[df_m['Total']<=15,'Team/NOC']='Other countries'
px.pie(df_m1, values='Total', names='Team/NOC', title='Medal won by Country')
```



From above bar diagram, it is clear that the highest medals is won by USA athletes, second highest is `China` and third highest is `Rassia` and so on.

In [15]:
```
df_combine_a=pd.DataFrame({'NOC':df_a.NOC.value_counts().keys(),'No_of_Athletes':df_a.NO
df_combine_c=pd.DataFrame({'NOC':df_c.NOC.value_counts().keys(),'No_of_Coaches':df_c.NOC
df_combine_m=pd.DataFrame({'NOC':df_m['Team/NOC'],'Medals':df_m['Total'],'Gold':df_m['Go
df_combine=pd.merge(left=df_combine_a,right=df_combine_c,how='outer',on='NOC')
df_combine=pd.merge(left=df_combine,right=df_combine_m,how='outer',on='NOC')
px.histogram(df_combine[:61],y='NOC',x=['No_of_Athletes','No_of_Coaches','Medals'],barmo
```



From the above histogram,

- `USA`
  China took `first` poition by wining 133 medals where 39 are gold,41 are Silver and 33 are Bronze.
  The `highest` number of Athletes took participated.
  There are 28 coaches for Athletes training from USA which is second highest.

- `China`
  China took `second` poition by wining 88 medals where 38 are gold,32 are Silver and 18 are Bronze.
  The `forth` highest number of players participated.
  The `sixth` highest number of coaches are there for Athletes in different Discipline training i.e 12.

- `Japan`
  Japan took `third` poition by wining 58 medals where 27 are gold,14 are Silver and 17 are Bronze.
  The `second` highest number of Athletes took participated.
  The `highest` number of coaches are there for Athletes in different Discipline training i.e 35.
  6th

- `Great Britain`
  Great Britain took `forth` poition by wining 65 medals where 22 are gold,21 are Silver and 22 are Bronze.
  The `eighth` highest number of players participated.
  There are only 7 coaches for Athletes in different Discipline training.

- `Russia`
  Russia took `fifth` poition by wining 71 medals where 20 are gold,28 are Silver and 23 are Bronze.
  The `eleventh` highest number of players participated.
  The `sixth` highest number of coaches are there for Athletes in different Discipline training i.e 12.

- `Australia`
  Australia took `sixth` poition by wining 46 medals where 17 are gold,7 are Silver and 22 are Bronze.
  The `third` highest number of players participated.
  The `third` highest number of coaches are there for Athletes in different Discipline training i.e 22

- `Netherlands`
  Australia took `seventh` poition by wining 36 medals where 10 are gold,12 are Silver and 14 are Bronze.
  The `thirteenth` highest number of players participated.
  The number of coaches are there for Athletes in different Discipline training i.e 10

- `France`
  France took `eighth` poition by wining 33 medals where 10 are gold,12 are Silver and 11 are Bronze.
  The `sixth` highest number of players participated.
  There are only 10 coaches for Athletes in different Discipline training.

- `Germany`
  Germany took `ninth` poition by wining 37 medals where 10 are gold,11 are Silver and 16 are Bronze.
  The `fifth` highest number of players participated.
  There are only 9 coaches for Athletes in different Discipline training.

- `Italy`
  Italy took `tenth` poition by wining 40 medals where 10 are gold,10 are Silver and 20 are Bronze.

The `ninth` highest number of players participated.
The `forth` highest number of coaches are there for Athletes in different Discipline training i.e 16.

- `Canada`
  Canada took `eleventh` poition by wining 24 medals where 7 are gold,6 are Silver and 11 are Bronze.
  The `seventh` highest number of players participated.
  The `forth` highest number of coaches are there for Athletes in different Discipline training i.e 16.