In [1]:

```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import warnings
warnings.filterwarnings('ignore')
```

In [2]:

```
olympic_data = pd.read_csv("winter.csv")
```

In [3]:

```
olympic_data.head()
```

Out[3]:

	Year	City	Sport	Discipline	Country	Gender	Event	Medal	Athlete
0	1924	Chamonix	Biathlon	Biathlon	FIN	Men	Military Patrol	Silver	FIN
1	1924	Chamonix	Biathlon	Biathlon	FRA	Men	Military Patrol	Bronze	FRA
2	1924	Chamonix	Biathlon	Biathlon	SUI	Men	Military Patrol	Gold	SUI
3	1924	Chamonix	Bobsleigh	Bobsleigh	BEL	Men	Four-Man	Bronze	BEL
4	1924	Chamonix	Bobsleigh	Bobsleigh	GBR	Men	Four-Man	Silver	GBR

In [4]:

```
olympic_data.tail()
```

Out[4]:

	Year	City	Sport	Discipline	Country	Gender	Event	Medal	Athlete
3269	2018	PyeongChang	Speed Skating	Speed Skating	KOR	Women	Mass Start	Silver	Kim Bo-Reum
3270	2018	PyeongChang	Speed Skating	Speed Skating	NED	Women	Mass Start	Bronze	Irene Schouten
3271	2018	PyeongChang	Speed Skating	Speed Skating	JPN	Women	Team Pursuit (6 laps)	Gold	Japan
3272	2018	PyeongChang	Speed Skating	Speed Skating	NED	Women	Team Pursuit (6 laps)	Silver	Netherlands
3273	2018	PyeongChang	Speed Skating	Speed Skating	USA	Women	Team Pursuit (6 laps)	Bronze	United States

In [5]:

```
olympic_data.shape
```

Out[5]:

```
(3274, 9)
```

In [6]:

```
olympic_data.columns
```

Out[6]:

```
Index(['Year', 'City', 'Sport', 'Discipline', 'Country', 'Gender', 'Even
t',
       'Medal', 'Athlete'],
      dtype='object')
```

In [7]:

```
olympic_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3274 entries, 0 to 3273
Data columns (total 9 columns):
 #   Column      Non-Null Count  Dtype 
---  ------      --------------  ----- 
 0   Year        3274 non-null   int64 
 1   City        3274 non-null   object
 2   Sport       3274 non-null   object
 3   Discipline  3274 non-null   object
 4   Country     3274 non-null   object
 5   Gender      3274 non-null   object
 6   Event       3274 non-null   object
 7   Medal       3274 non-null   object
 8   Athlete     3274 non-null   object
dtypes: int64(1), object(8)
memory usage: 230.3+ KB
```

In [8]:

```
olympic_data.describe()
```

Out[8]:

	Year
count	3274.000000
mean	1989.351252
std	24.275321
min	1924.000000
25%	1976.000000
50%	1994.000000
75%	2010.000000
max	2018.000000

In [9]:

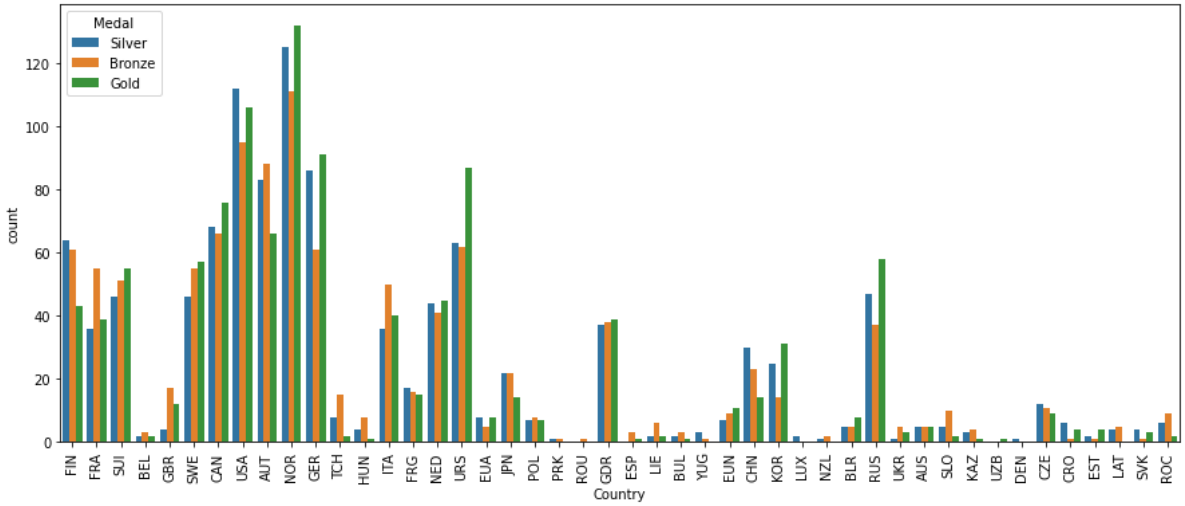
```
olympic_data.isnull().sum()
```

Out[9]:

```
Year          0
City          0
Sport         0
Discipline    0
Country       0
Gender        0
Event         0
Medal         0
Athlete       0
dtype: int64
```

In [11]:

```
plt.figure(figsize=(15,6))
sns.countplot(x = 'Country', hue = 'Medal', data = olympic_data)
plt.xticks(rotation = 90)
plt.show()
```



In [17]:

```
country_medals = olympic_data[['Country', 'Medal']] \
                .groupby('Country').count() \
                .sort_values('Medal', ascending=False) \
                .head(20)
```

In [19]:

```
country_medals
```

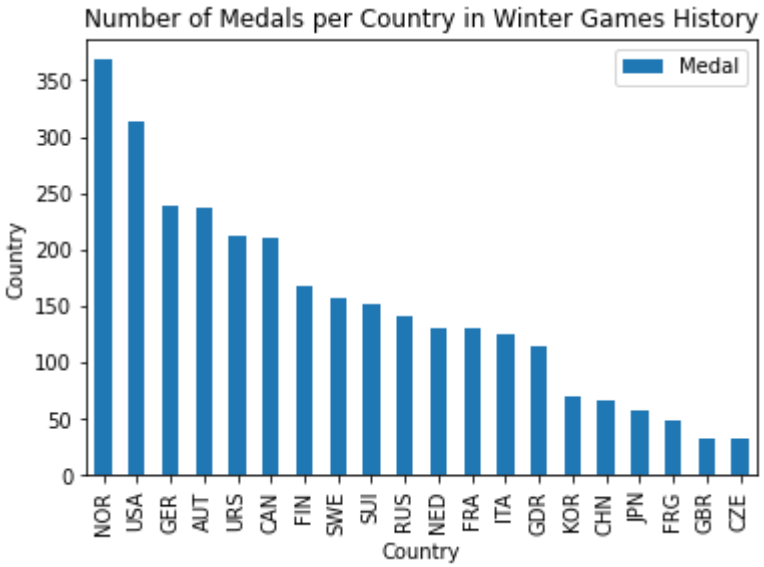
Out[19]:

Country	Medal
NOR	368
USA	313
GER	238
AUT	237
URS	212
CAN	210
FIN	168
SWE	158
SUI	152
RUS	142
NED	130
FRA	130
ITA	126
GDR	114
KOR	70
CHN	67
JPN	58
FRG	48
GBR	33
CZE	32

In [34]:

```
plt.figure(figsize=(10,6))
country_medals.plot(kind='bar')
plt.ylabel('Country')
plt.title('Number of Medals per Country in Winter Games History');
```

```
<Figure size 720x432 with 0 Axes>
```



In [23]:

```
medal_yearwise=olympic_data.groupby('Year')['Medal'].count()
```

In [24]:

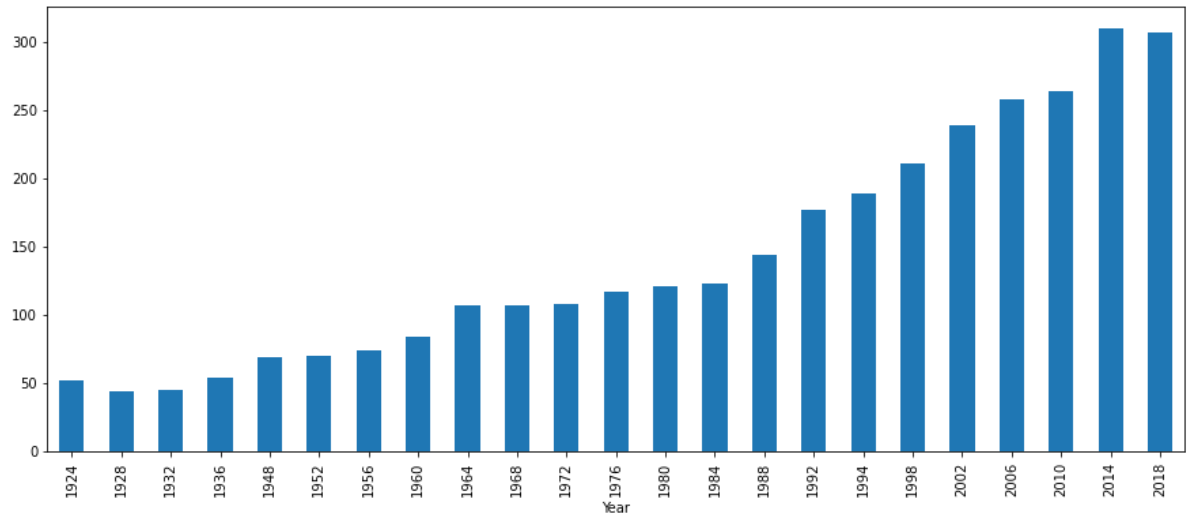
```
medal_yearwise
```

Out[24]:

```
Year
1924     52
1928     44
1932     45
1936     54
1948     69
1952     70
1956     74
1960     84
1964    107
1968    107
1972    108
1976    117
1980    121
1984    123
1988    144
1992    177
1994    189
1998    211
2002    239
2006    258
2010    264
2014    310
2018    307
Name: Medal, dtype: int64
```

In [32]:

```
plt.figure(figsize=(15,6))
medal_yearwise.plot(kind='bar')
plt.xticks(rotation = 90)
plt.show()
```

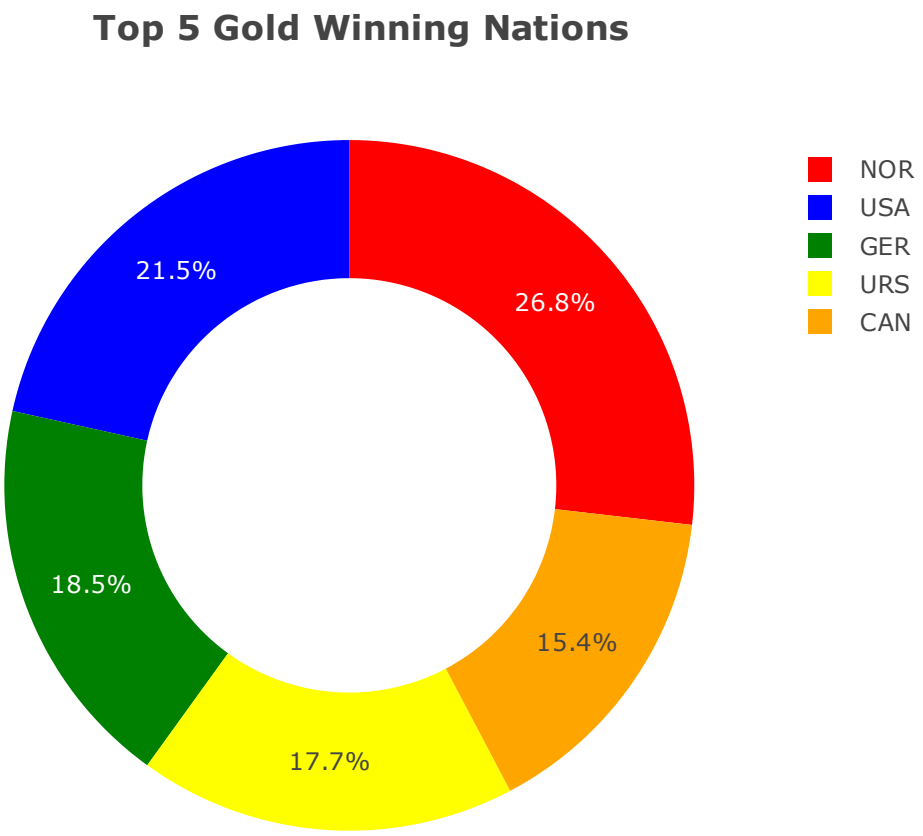


In [26]:

```
import plotly.express as px
```

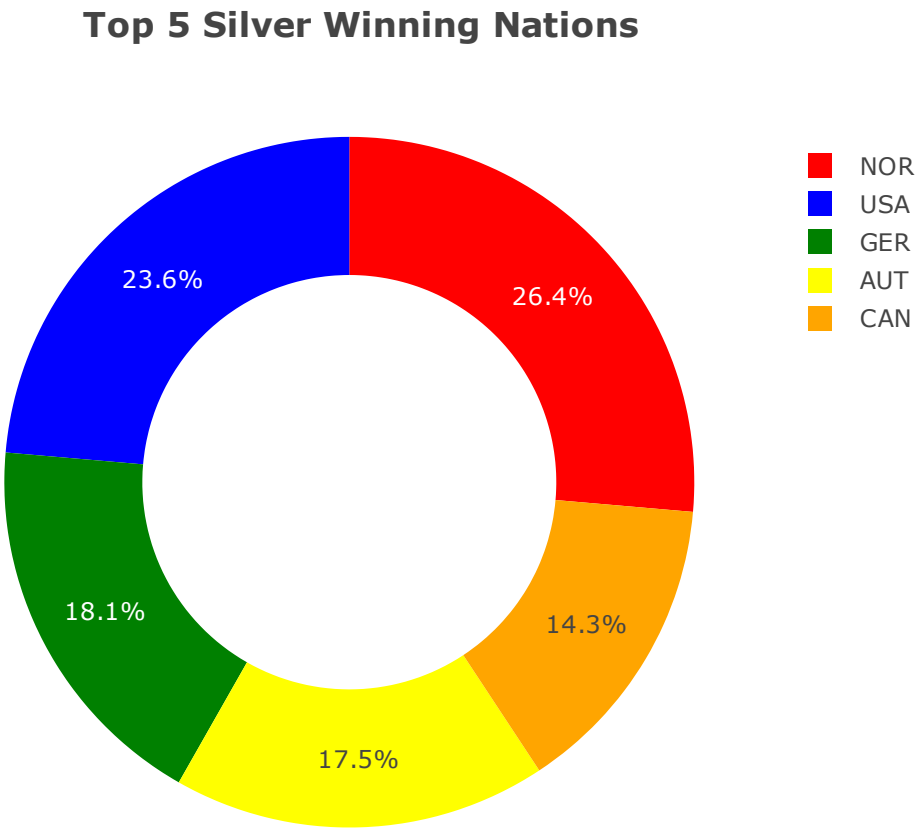
In [27]:

```
gold_medal = olympic_data[olympic_data['Medal']=="Gold"] \
           .groupby(["Country"])\
           .count()\
           .sort_values(by='Medal', ascending=False).reset_index()
line_colors = ["red", "blue","green","yellow","orange"]
gold_medal_plot= px.pie(gold_medal.head(5), values='Medal',hole=0.6, names='Country', ti
gold_medal_plot.show()
```



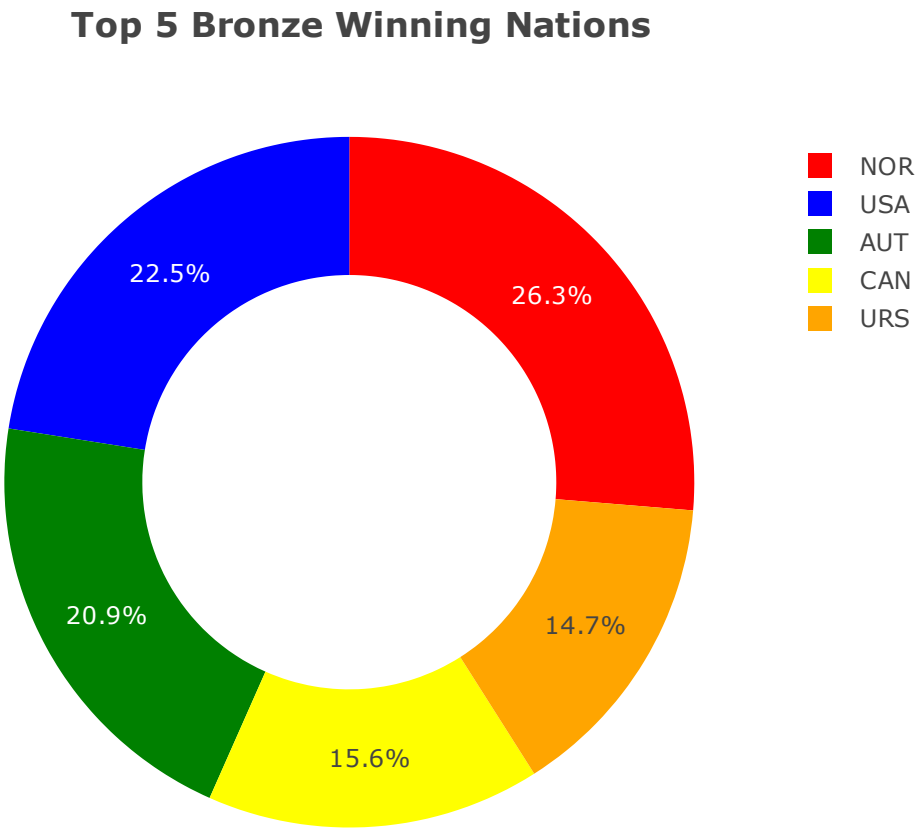
In [29]:

```
silver_medal = olympic_data[olympic_data['Medal']=="Silver"] \
           .groupby(["Country"])\
           .count()\
           .sort_values(by='Medal', ascending=False).reset_index()
line_colors = ["red", "blue","green","yellow","orange"]
silver_medal_plot= px.pie(silver_medal.head(5), values='Medal',hole=0.6, names='Country
silver_medal_plot.show()
```



In [30]:

```
bronze_medal = olympic_data[olympic_data['Medal']=="Bronze"] \
           .groupby(["Country"])\
           .count()\
           .sort_values(by='Medal', ascending=False).reset_index()
line_colors = ["red", "blue","green","yellow","orange"]
bronze_medal_plot= px.pie(bronze_medal.head(5), values='Medal',hole=0.6, names='Country
bronze_medal_plot.show()
```



In [35]:

```
yearly_medal=olympic_data.groupby('Year')['Medal'].value_counts()
```

In [36]:

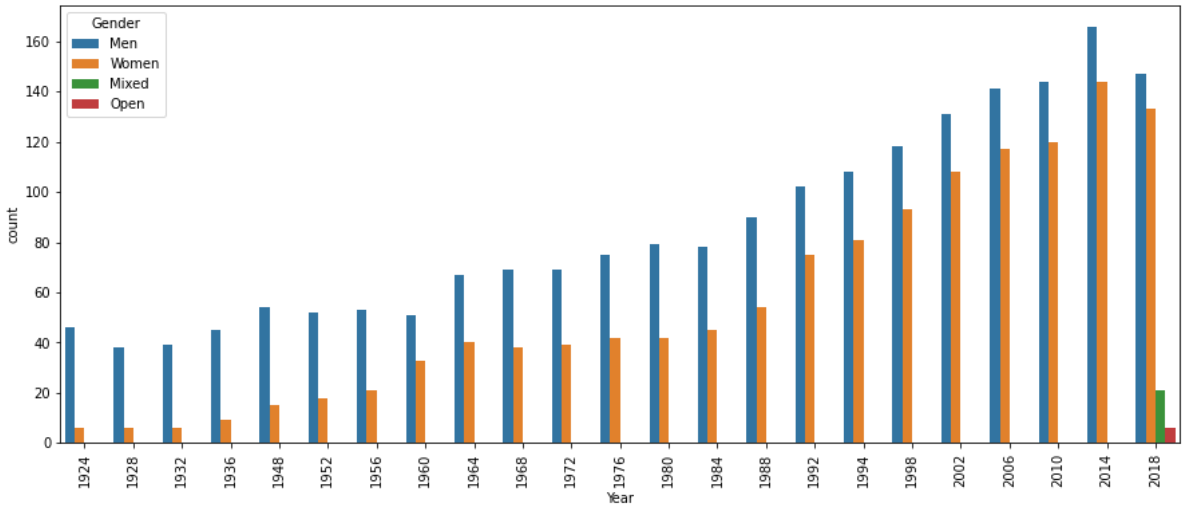
```
yearly_medal
```

Out[36]:

```
Year  Medal
1924  Bronze     18
      Gold       17
      Silver     17
1928  Bronze     16
      Gold       15
               ...
2014  Gold      104
      Silver    102
2018  Gold      103
      Bronze    102
      Silver    102
Name: Medal, Length: 69, dtype: int64
```

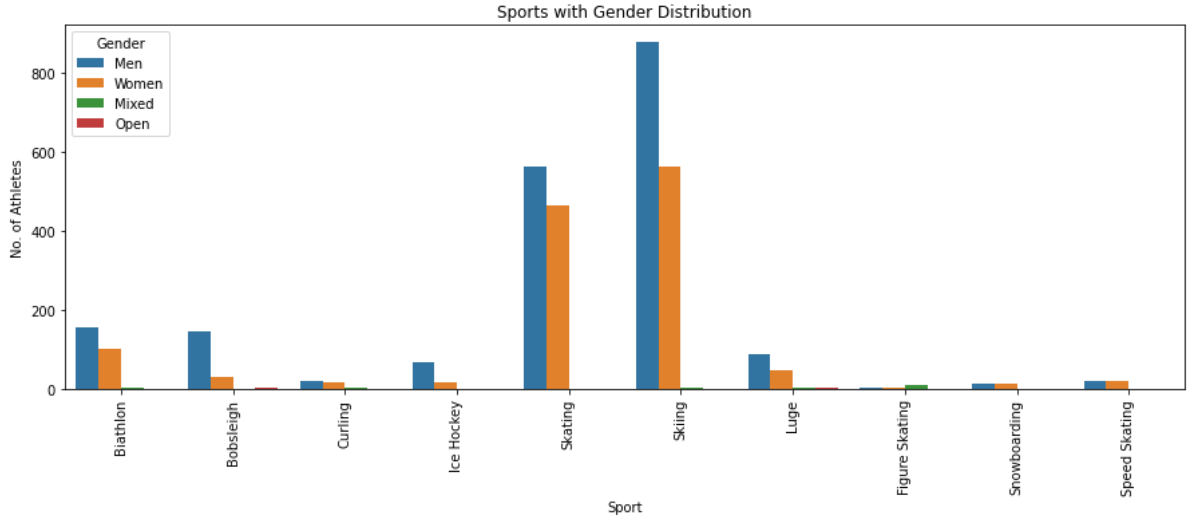
In [37]:

```
plt.figure(figsize=(15,6))
sns.countplot(x = 'Year', hue = 'Gender', data = olympic_data)
plt.xticks(rotation = 90)
plt.show()
```



In [39]:

```
plt.figure(figsize=(15, 5))
sport_by_gender = olympic_data['Sport'].value_counts().index
sns.countplot(x='Sport', hue = 'Gender', data = olympic_data)
plt.xticks(rotation=90)
plt.title('Sports with Gender Distribution')
plt.xlabel('Sport')
plt.ylabel('No. of Athletes');
```

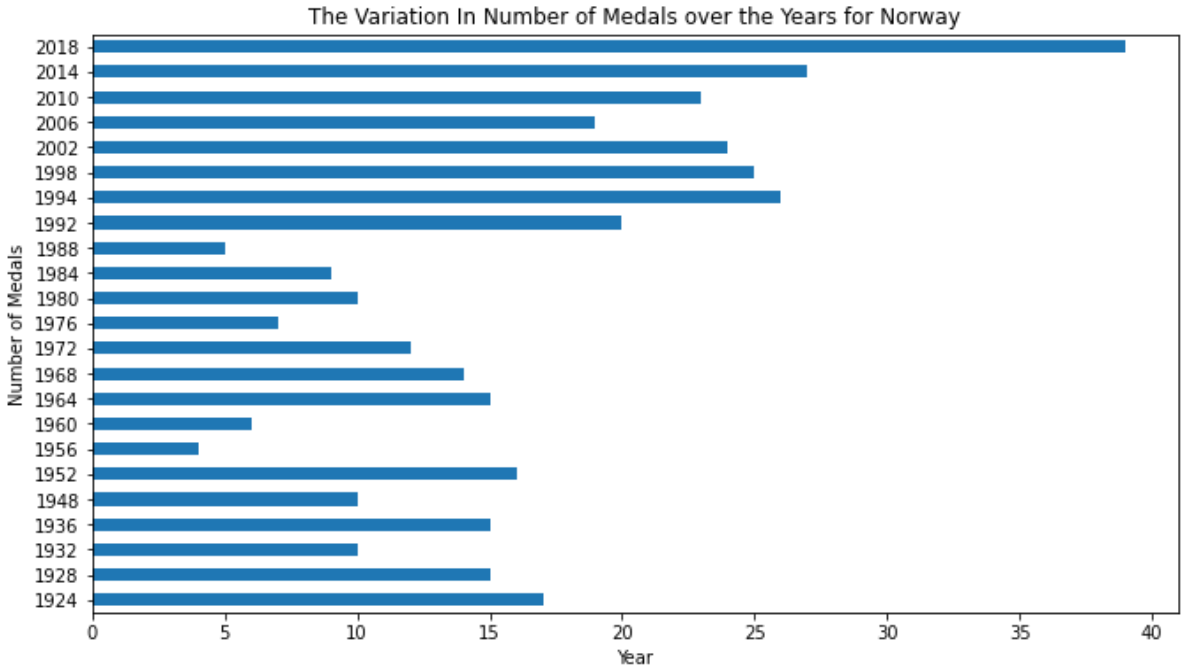


In [40]:

```
norway_country = olympic_data[olympic_data['Country']=='NOR']
norway_medals_data = norway_country.groupby('Year')['Medal'].count()
norway_medals_data.plot(kind='barh', figsize=(11,6))
plt.ylabel('Number of Medals')
plt.xlabel('Year')
plt.title('The Variation In Number of Medals over the Years for Norway')
```

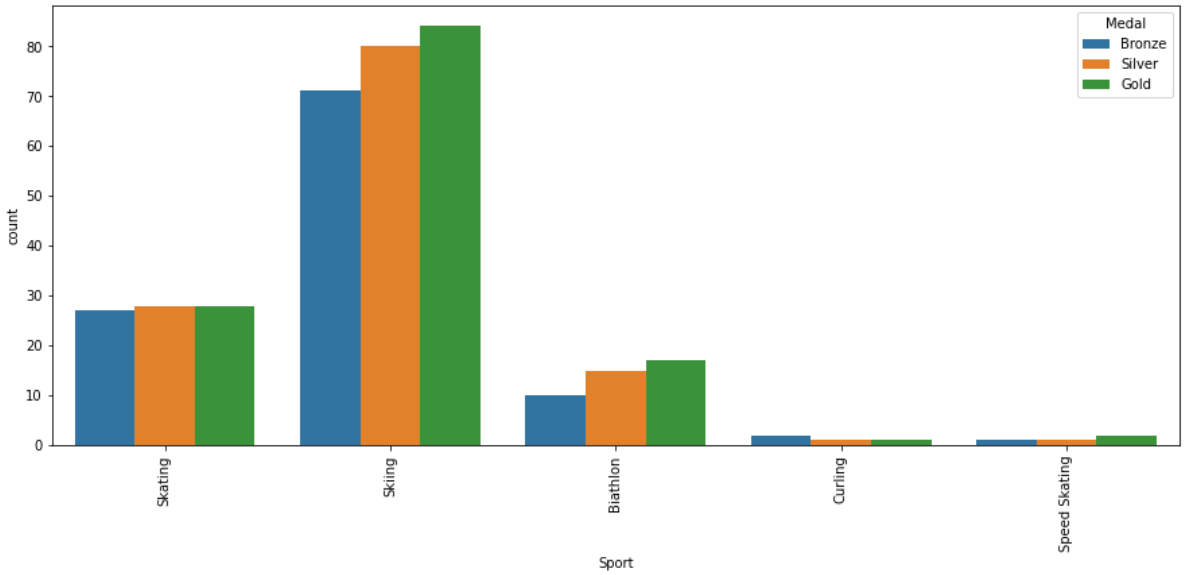
Out[40]:

```
Text(0.5, 1.0, 'The Variation In Number of Medals over the Years for Norwa
y')
```



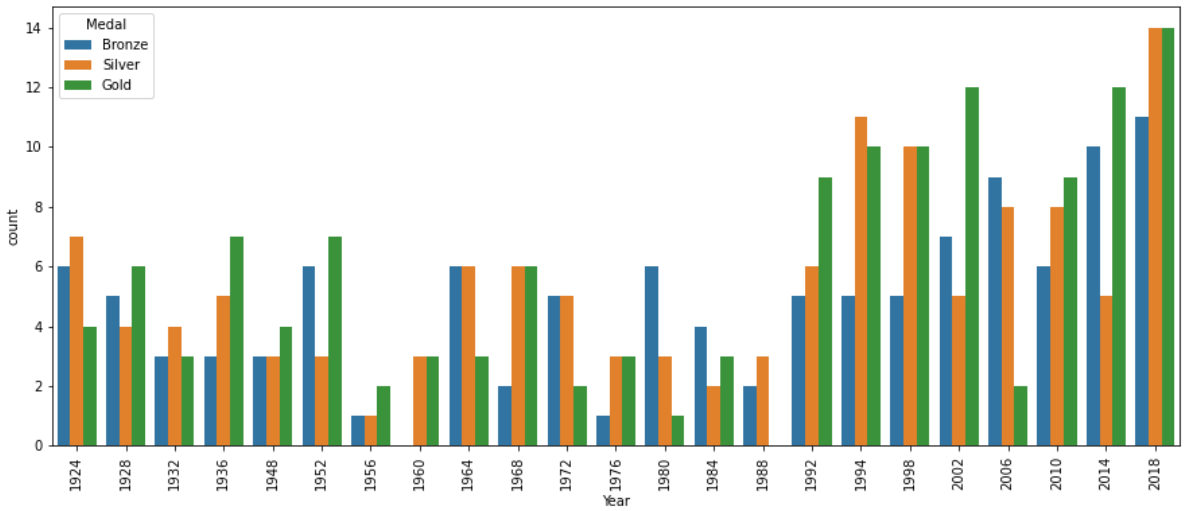
In [41]:

```
plt.figure(figsize=(15,6))
sns.countplot(x = 'Sport', hue = 'Medal', data = norway_country)
plt.xticks(rotation = 90)
plt.show()
```



In [42]:

```
plt.figure(figsize=(15,6))
sns.countplot(x = 'Year', hue = 'Medal', data = norway_country)
plt.xticks(rotation = 90)
plt.show()
```



In [45]:

```
country_joined = olympic_data.groupby('Year')['Country'].nunique() \
                .reset_index()
```

In [46]:

```
country_joined
```

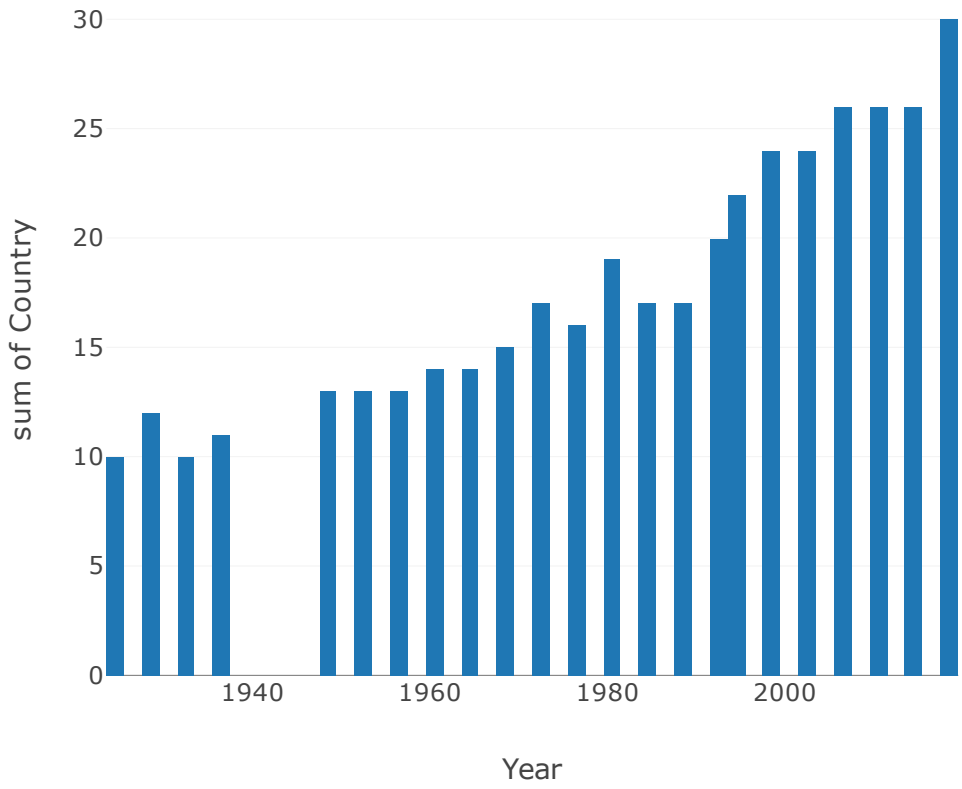
Out[46]:

	Year	Country
0	1924	10
1	1928	12
2	1932	10
3	1936	11
4	1948	13
5	1952	13
6	1956	13
7	1960	14
8	1964	14
9	1968	15
10	1972	17
11	1976	16
12	1980	19
13	1984	17
14	1988	17
15	1992	20
16	1994	22
17	1998	24
18	2002	24
19	2006	26
20	2010	26
21	2014	26
22	2018	30

In [48]:

```
fig=px.histogram(country_joined,x='Year',y='Country',title="<b>The Variation In Number o
fig.show()
```

e Variation In Number of Participating Countries over the Yea



In [49]:

```
discipline_games = olympic_data.groupby('Year')['Discipline'].nunique() \
.reset_index()
```

In [50]:

```
discipline_games
```

Out[50]:

	Year	Discipline
0	1924	9
1	1928	8
2	1932	7
3	1936	8
4	1948	9
5	1952	8
6	1956	8
7	1960	8
8	1964	10
9	1968	10
10	1972	10
11	1976	10
12	1980	10
13	1984	10
14	1988	10
15	1992	12
16	1994	12
17	1998	14
18	2002	15
19	2006	15
20	2010	15
21	2014	15
22	2018	15

In [51]:

```
fig=px.histogram(discipline_games,x='Year',y='Discipline',title="<b>The Variation In Nur
fig.show()
```