

pip install plotly

```
In [1]: 1 import pandas as pd
        2 import numpy as np
        3 import matplotlib.pyplot as plt
        4 import plotly as pl
        5 import plotly.graph_objs as go
        6 import plotly.offline as py
        7 # import plotly.express as px
        8 import plotly.figure_factory as ff
```

```
In [ ]: 1
```

```
In [2]: 1 df = pd.read_csv('athlete_events.csv')
```

```
In [3]: 1 df.head()
```

```
Out[3]:
```

	ID	Name	Sex	Age	Height	Weight	Team	NOC	Games	Year	Season	City	Sport	Event	Medal
0	1	A Dijiang	M	24.0	180.0	80.0	China	CHN	1992 Summer	1992	Summer	Barcelona	Basketball	Basketball Men's Basketball	NaN
1	2	A Lamusi	M	23.0	170.0	60.0	China	CHN	2012 Summer	2012	Summer	London	Judo	Judo Men's Extra-Lightweight	NaN
2	3	Gunnar Nielsen Aaby	M	24.0	NaN	NaN	Denmark	DEN	1920 Summer	1920	Summer	Antwerpen	Football	Football Men's Football	NaN
3	4	Edgar Lindenau Aabye	M	34.0	NaN	NaN	Denmark/Sweden	DEN	1900 Summer	1900	Summer	Paris	Tug-Of-War	Tug-Of-War Men's Tug-Of-War	Gold
4	5	Christine Jacoba Aaftink	F	21.0	185.0	82.0	Netherlands	NED	1988 Winter	1988	Winter	Calgary	Speed Skating	Speed Skating Women's 500 metres	NaN

```
In [4]: 1 df.columns
```

```
Out[4]: Index(['ID', 'Name', 'Sex', 'Age', 'Height', 'Weight', 'Team', 'NOC', 'Games',  
             'Year', 'Season', 'City', 'Sport', 'Event', 'Medal'],  
            dtype='object')
```

```
In [5]: 1 medal_counts = df['Medal'].value_counts()
```

```
In [6]: 1 medal_counts
```

```
Out[6]: Gold      13372  
       Bronze    13295  
       Silver    13116  
       Name: Medal, dtype: int64
```

```
In [7]: 1 medal_counts
```

```
Out[7]: Gold      13372  
       Bronze    13295  
       Silver    13116  
       Name: Medal, dtype: int64
```

```
In [8]: 1 label = medal_counts.index  
       2 value = medal_counts.values
```

```
In [9]: 1 label
```

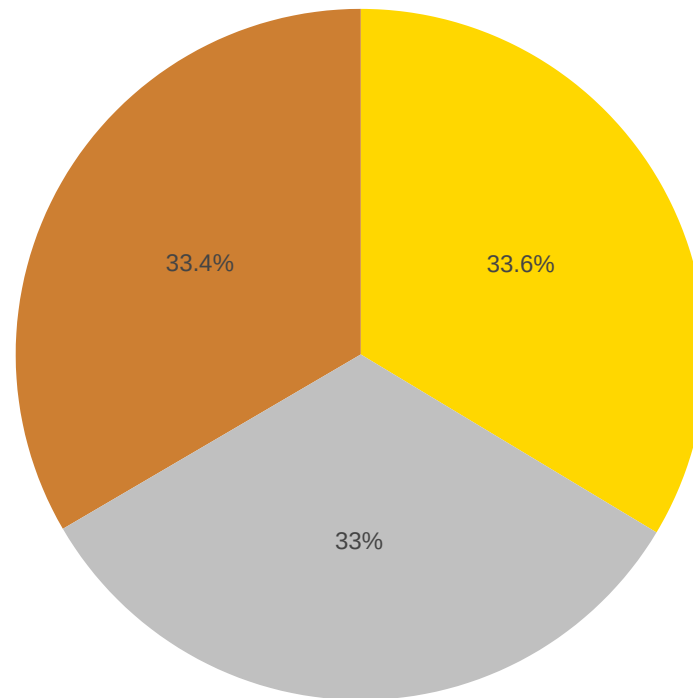
```
Out[9]: Index(['Gold', 'Bronze', 'Silver'], dtype='object')
```

```
In [10]: 1 value
```

```
Out[10]: array([13372, 13295, 13116])
```

```
In [11]: 1 color = ['#FFD700', '#CD7F32', '#C0C0C0']  
2 pie = go.Pie(labels=label, values=value, marker=dict(colors=color))  
3 layout = go.Layout(title='Medal Distribution')  
4 fig = go.Figure(data=[pie], layout=layout)  
5 py.iplot(fig)
```

Medal Distribution



In [12]: 1 df.head()

Out[12]:

	ID	Name	Sex	Age	Height	Weight	Team	NOC	Games	Year	Season	City	Sport	Event	Medal
0	1	A Dijiang	M	24.0	180.0	80.0	China	CHN	1992 Summer	1992	Summer	Barcelona	Basketball	Basketball Men's Basketball	NaN
1	2	A Lamusi	M	23.0	170.0	60.0	China	CHN	2012 Summer	2012	Summer	London	Judo	Judo Men's Extra-Lightweight	NaN
2	3	Gunnar Nielsen Aaby	M	24.0	NaN	NaN	Denmark	DEN	1920 Summer	1920	Summer	Antwerpen	Football	Football Men's Football	NaN
3	4	Edgar Lindenau Aabye	M	34.0	NaN	NaN	Denmark/Sweden	DEN	1900 Summer	1900	Summer	Paris	Tug-Of-War	Tug-Of-War Men's Tug-Of-War	Gold
4	5	Christine Jacoba Aaftink	F	21.0	185.0	82.0	Netherlands	NED	1988 Winter	1988	Winter	Calgary	Speed Skating	Speed Skating Women's 500 metres	NaN

Scatter Plot

In [13]: 1 df.columns

Out[13]: Index(['ID', 'Name', 'Sex', 'Age', 'Height', 'Weight', 'Team', 'NOC', 'Games', 'Year', 'Season', 'City', 'Sport', 'Event', 'Medal'], dtype='object')

In [14]: 1 df.head()

Out[14]:

	ID	Name	Sex	Age	Height	Weight	Team	NOC	Games	Year	Season	City	Sport	Event	Medal
0	1	A Dijiang	M	24.0	180.0	80.0	China	CHN	1992 Summer	1992	Summer	Barcelona	Basketball	Basketball Men's Basketball	NaN
1	2	A Lamusi	M	23.0	170.0	60.0	China	CHN	2012 Summer	2012	Summer	London	Judo	Judo Men's Extra-Lightweight	NaN
2	3	Gunnar Nielsen Aaby	M	24.0	NaN	NaN	Denmark	DEN	1920 Summer	1920	Summer	Antwerpen	Football	Football Men's Football	NaN
3	4	Edgar Lindenau Aabye	M	34.0	NaN	NaN	Denmark/Sweden	DEN	1900 Summer	1900	Summer	Paris	Tug-Of-War	Tug-Of-War Men's Tug-Of-War	Gold
4	5	Christine Jacoba Aaftink	F	21.0	185.0	82.0	Netherlands	NED	1988 Winter	1988	Winter	Calgary	Speed Skating	Speed Skating Women's 500 metres	NaN

In [15]: 1 tmp = df.groupby(['Year', 'City'])['Season'].value_counts()

In [16]: 1 type(tmp)

Out[16]: pandas.core.series.Series

In [17]: 1 tmp.values

Out[17]: array([[380, 1936, 1301, 1733, 3101, 4040, 4292, 460, 5233, 4992, 582, 352, 2969, 6506, 895, 6405, 1075, 8270, 1088, 1307, 4829, 298, 8119, 1116, 1778, 7702, 1891, 8588, 10304, 1655, 1861, 8641, 1746, 7191, 9454, 2134, 2639, 12037, 3436, 12977, 3160, 13780, 3605, 13821, 4109, 13443, 4382, 13602, 4402, 12920, 4891, 13688])

In [18]: 1 df_copy = pd.DataFrame(data={'Athlets':tmp.values},index=tmp.index).reset_index()

In [19]: 1 df_copy.head()

Out[19]:

	Year	City	Season	Athlets
0	1896	Athina	Summer	380
1	1900	Paris	Summer	1936
2	1904	St. Louis	Summer	1301
3	1906	Athina	Summer	1733
4	1908	London	Summer	3101

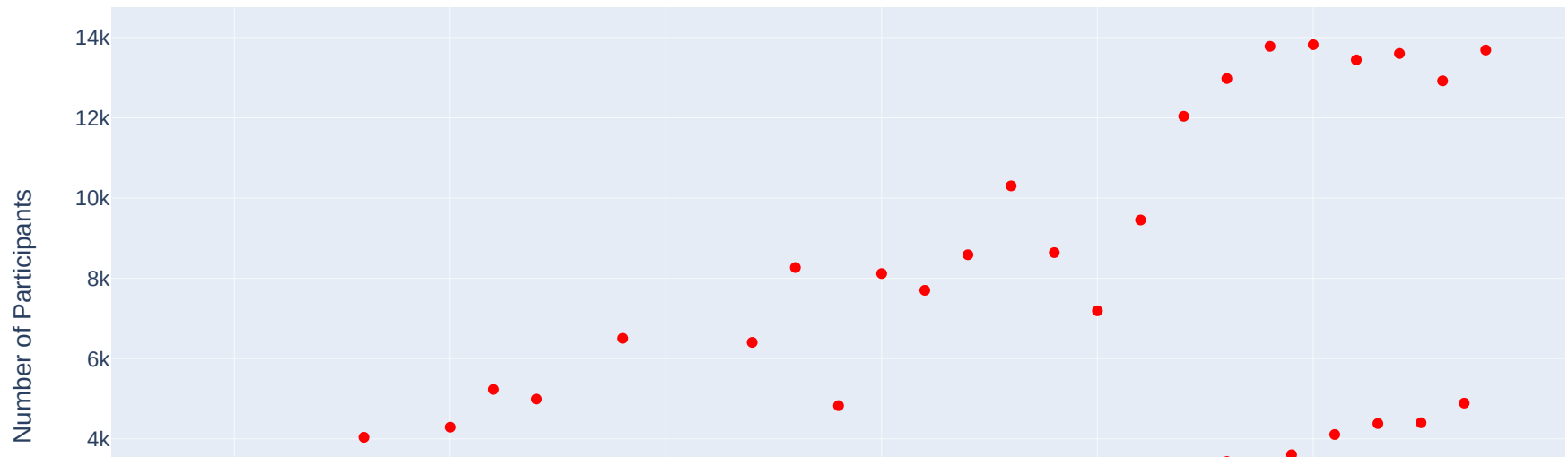
In []:

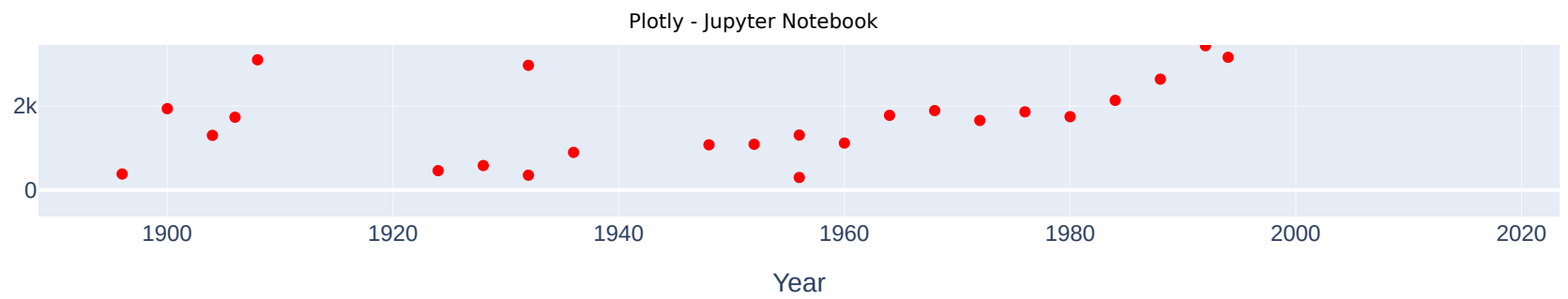
1

In [20]:

```
1 sct = go.Scatter(  
2     x = df_copy['Year'],  
3     y = df_copy['Athlets'],  
4     name="Athlets per Olympic games",  
5     marker = dict(color="Red"),mode="markers"  
6 )  
7  
8 data = [sct]  
9 # layout = dict(title='Athlets Per Olympic Games')  
10 # xaxis = dict(title="Year",showticklabels=True)  
11  
12  
13 layout = dict(  
14     title = 'Athlets Per Olympic Games',  
15     xaxis = dict(title="Year",showticklabels=True),  
16     yaxis = dict(title="Number of Participants",showticklabels=True),  
17 )  
18  
19 fig = dict(data=data,layout=layout)  
20 py.iplot(fig)
```

Athlets Per Olympic Games



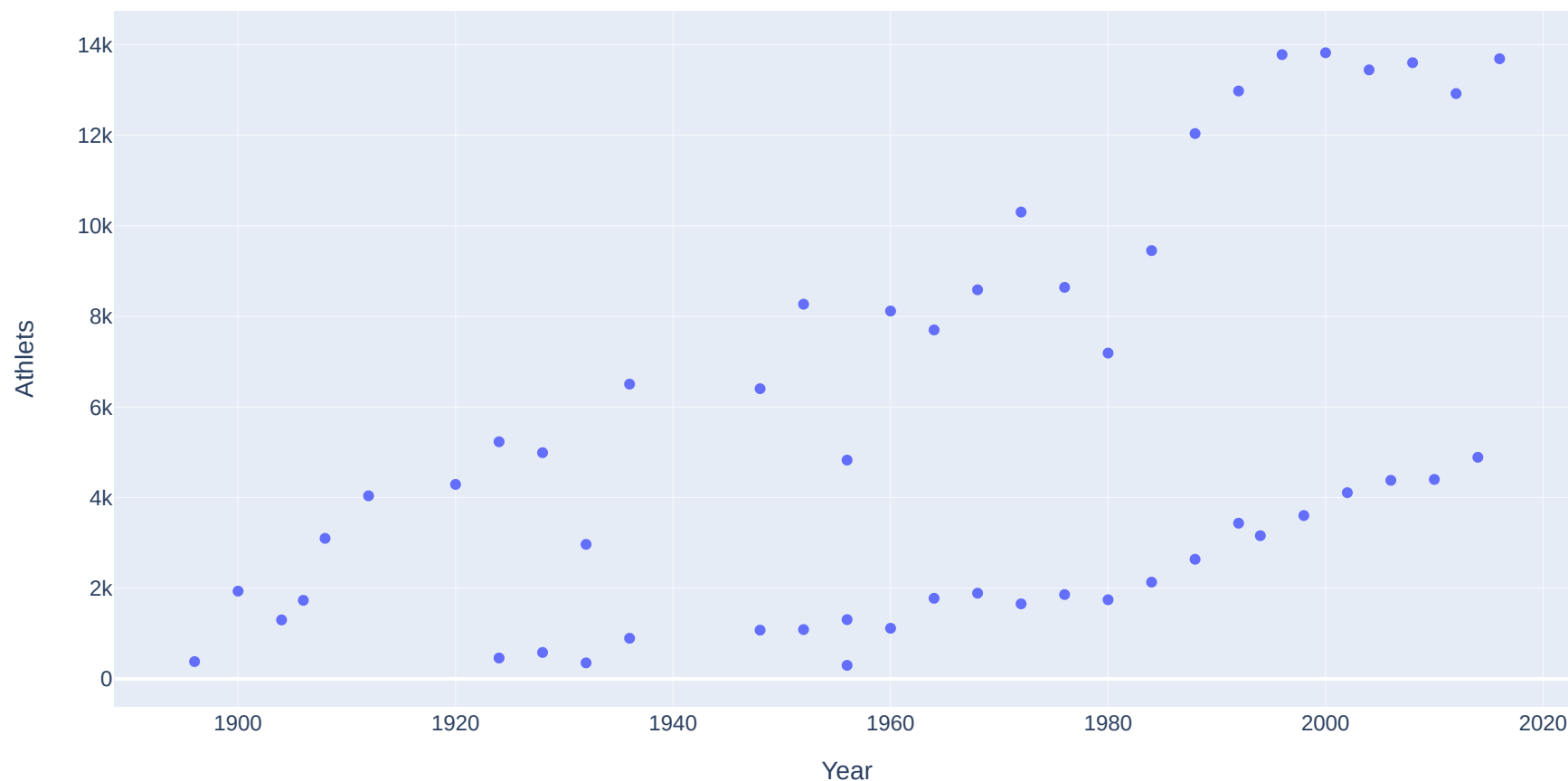


way 2

```
In [21]: 1 import plotly.express as px
```



```
In [22]: 1 px.scatter(df_copy,x=df_copy['Year'],y=df_copy['Athlets'])
```



```
In [ ]: 1
```

```
In [23]: 1 df_summer = df_copy[df_copy['Season'] == "Summer"]
```

```
In [24]: 1 df_winter = df_copy[df_copy['Season'] == "Winter"]
```

In []:

1

In []:

1

In [25]:

1 df_summer.head()

Out[25]:

	Year	City	Season	Athlets
0	1896	Athina	Summer	380
1	1900	Paris	Summer	1936
2	1904	St. Louis	Summer	1301
3	1906	Athina	Summer	1733
4	1908	London	Summer	3101

In [26]:

1 df_winter.head()

Out[26]:

	Year	City	Season	Athlets
7	1924	Chamonix	Winter	460
10	1928	Sankt Moritz	Winter	582
11	1932	Lake Placid	Winter	352
14	1936	Garmisch-Partenkirchen	Winter	895
16	1948	Sankt Moritz	Winter	1075

In []:

1

In []:

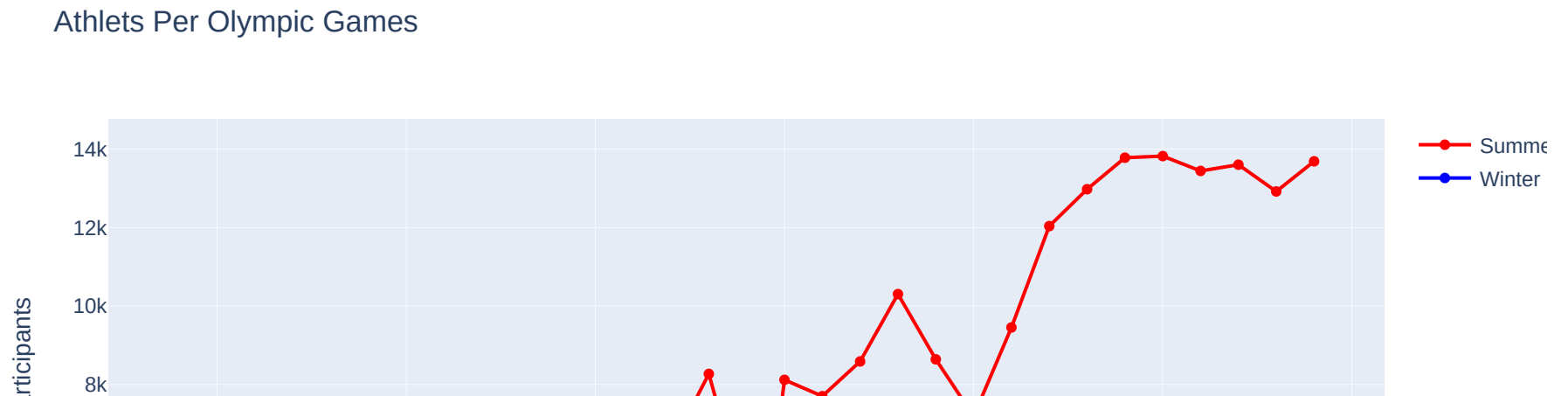
1

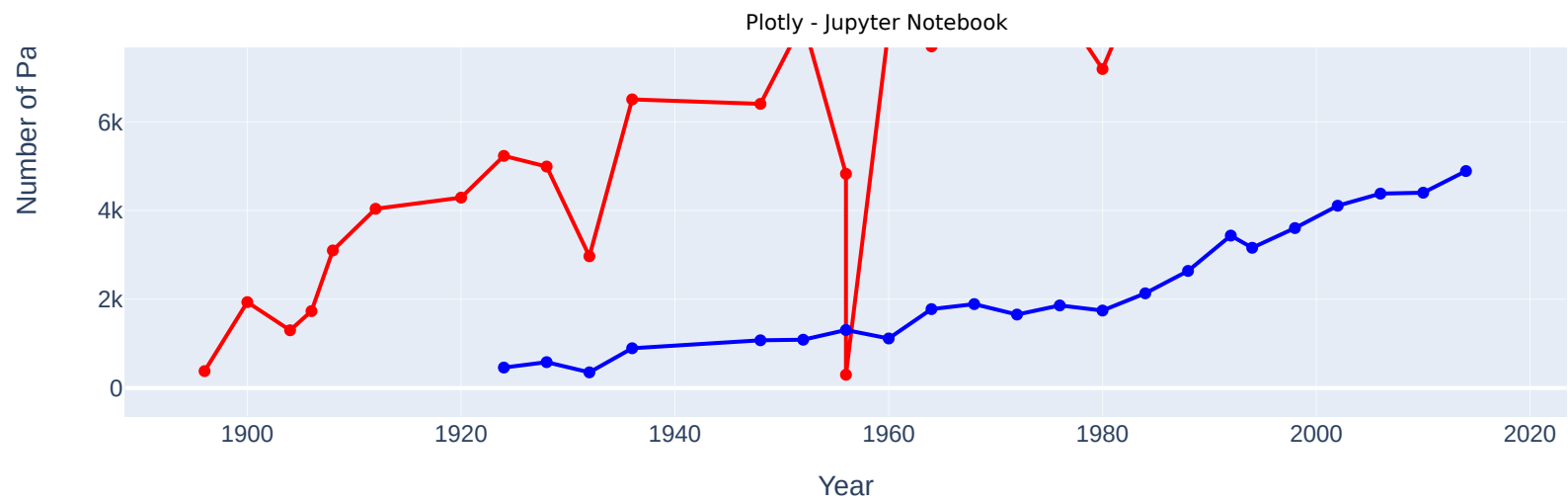
In [27]:

```

1 my_season_summer_sct = go.Scatter(
2     x = df_summer['Year'],
3     y = df_summer['Athlets'],
4     name="Summer Games",
5     marker = dict(color='red'),
6     mode = "markers+lines"
7 )
8
9 my_season_winter_sct = go.Scatter(
10     x = df_winter['Year'],
11     y = df_winter['Athlets'],
12     name="Winter Games",
13     marker = dict(color="Blue"),
14     mode = "markers+lines"
15 )
16
17 data = [my_season_summer_sct,my_season_winter_sct]
18 layout = dict(
19     title = 'Athlets Per Olympic Games',
20     xaxis = dict(title="Year",showticklabels=True),
21     yaxis = dict(title="Number of Participants",showticklabels=True),
22 )
23
24 fig = dict(data=data,layout=layout)
25 py.iplot(fig)

```





In []:

1

In []:

1

Bar Plot

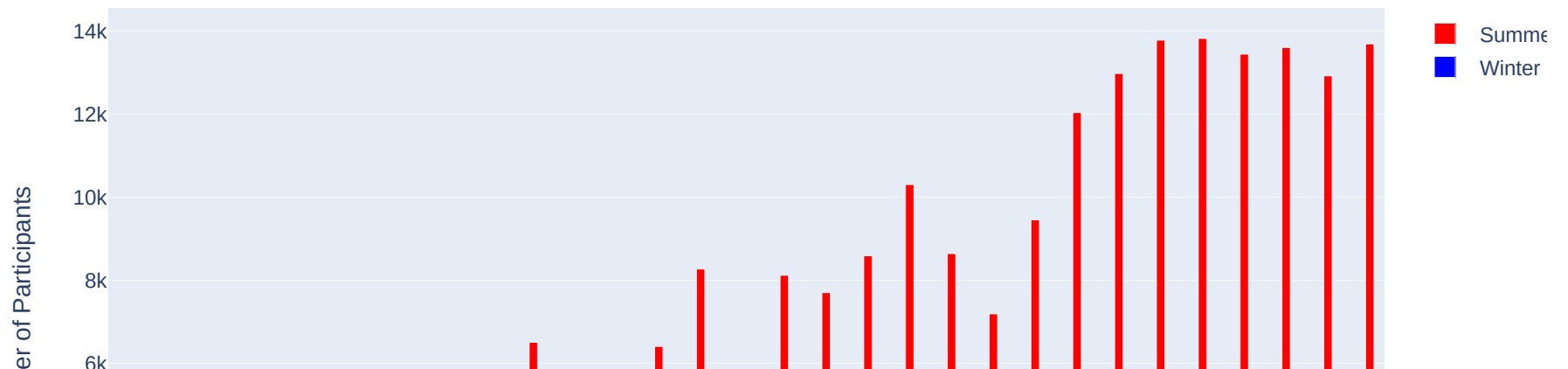
In [28]:

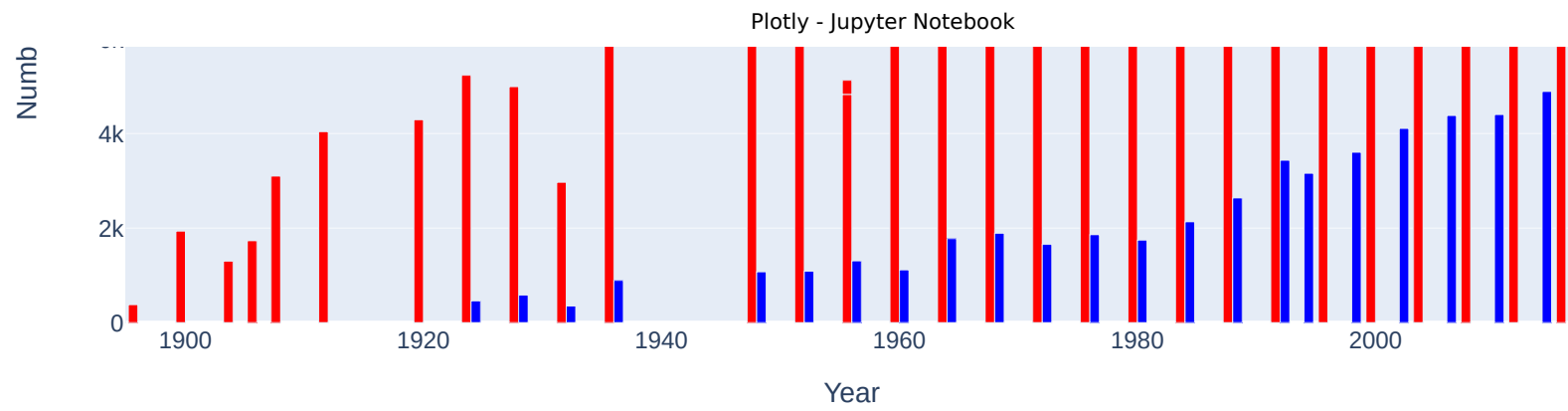
```

1 my_season_summer_bar = go.Bar(
2     x = df_summer['Year'],
3     y = df_summer['Athlets'],
4     name="Summer Games",
5     marker = dict(color='red')
6 )
7
8 my_season_winter_bar = go.Bar(
9     x = df_winter['Year'],
10    y = df_winter['Athlets'],
11    name="Winter Games",
12    marker = dict(color="Blue")
13 )
14
15 data = [my_season_summer_bar,my_season_winter_bar]
16 layout = dict(
17     title = 'Athlets Per Olympic Games',
18     xaxis = dict(title="Year",showticklabels=True),
19     yaxis = dict(title="Number of Participants",showticklabels=True),
20 )
21 fig = dict(data=data,layout=layout)
22 py.iplot(fig)

```

Athlets Per Olympic Games

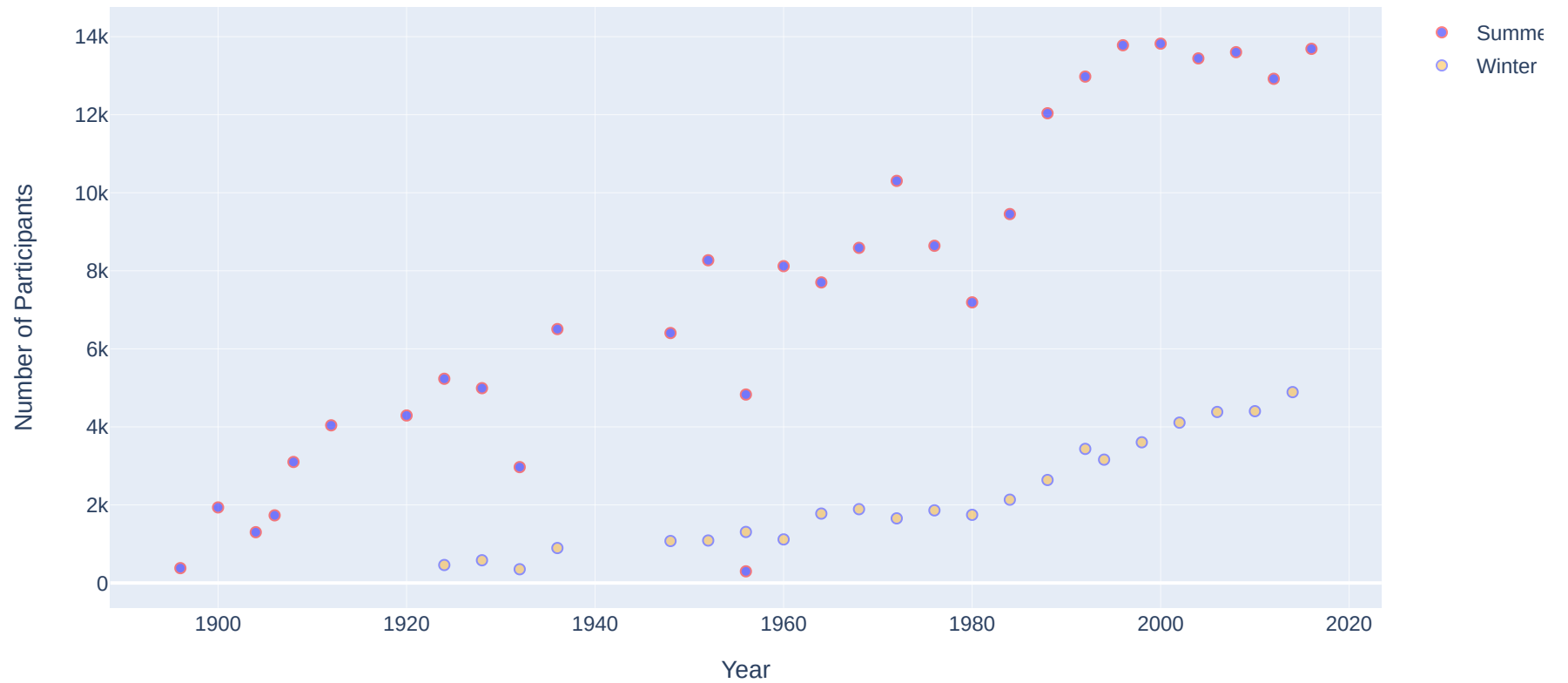




In [29]:

```
1 my_season_summer_sct = go.Scatter(  
2     x = df_summer['Year'],  
3     y = df_summer['Athlets'],  
4     name="Summer Games",  
5     marker = dict(color='Blue',  
6                   line=dict(  
7                       color='red',  
8                       width=1  
9                   ),  
10    opacity=0.5  
11    ),  
12    mode="markers"  
13 )  
14  
15 my_season_winter_sct = go.Scatter(  
16     x = df_winter['Year'],  
17     y = df_winter['Athlets'],  
18     name="Winter Games",  
19     marker = dict(color='orange',  
20                   line=dict(  
21                       color='blue',  
22                       width=1  
23                   ),  
24    opacity=0.4  
25    ),  
26    mode="markers"  
27 )  
28  
29 data = [my_season_summer_sct,my_season_winter_sct]  
30 layout = dict(  
31     title = 'Athlets Per Olympic Games',  
32     xaxis = dict(title="Year",showticklabels=True),  
33     yaxis = dict(title="Number of Participants",showticklabels=True),  
34 )  
35 fig = dict(data=data,layout=layout)  
36 py.iplot(fig)
```

Athlets Per Olympic Games



In []:

1

In [30]: 1 df.head()

Out[30]:

	ID	Name	Sex	Age	Height	Weight	Team	NOC	Games	Year	Season	City	Sport	Event	Medal
0	1	A Dijiang	M	24.0	180.0	80.0	China	CHN	1992 Summer	1992	Summer	Barcelona	Basketball	Basketball Men's Basketball	NaN
1	2	A Lamusi	M	23.0	170.0	60.0	China	CHN	2012 Summer	2012	Summer	London	Judo	Judo Men's Extra-Lightweight	NaN
2	3	Gunnar Nielsen Aaby	M	24.0	NaN	NaN	Denmark	DEN	1920 Summer	1920	Summer	Antwerpen	Football	Football Men's Football	NaN
3	4	Edgar Lindenau Aabye	M	34.0	NaN	NaN	Denmark/Sweden	DEN	1900 Summer	1900	Summer	Paris	Tug-Of-War	Tug-Of-War Men's Tug-Of-War	Gold
4	5	Christine Jacoba Aaftink	F	21.0	185.0	82.0	Netherlands	NED	1988 Winter	1988	Winter	Calgary	Speed Skating	Speed Skating Women's 500 metres	NaN

In [31]: 1 df.columns

Out[31]: Index(['ID', 'Name', 'Sex', 'Age', 'Height', 'Weight', 'Team', 'NOC', 'Games', 'Year', 'Season', 'City', 'Sport', 'Event', 'Medal'], dtype='object')

In []: 1

In [32]: 1 country = df.groupby(['Team'])['Year'].nunique()

```
In [33]: 1 country
```

```
Out[33]: Team
30. Februar      1
A North American Team  1
Acipactli        1
Acturus          1
Afghanistan      14
..
Zambia           13
Zefyros          1
Zimbabwe         14
Zut              1
rn-2             1
Name: Year, Length: 1184, dtype: int64
```

```
In [34]: 1 country_df = pd.DataFrame(data={'Editions':country.values},index=country.index).reset_index()
```

In [35]: 1 country_df

Out[35]:

	Team	Editions
0	30. Februar	1
1	A North American Team	1
2	Acipactli	1
3	Acturus	1
4	Afghanistan	14
...
1179	Zambia	13
1180	Zefyros	1
1181	Zimbabwe	14
1182	Zut	1
1183	rn-2	1

1184 rows × 2 columns

In [36]: 1 country_df.columns = ['Country', 'Editions']

In [37]: 1 country_df

Out[37]:

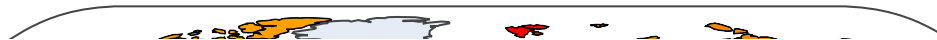
	Country	Editions
0	30. Februar	1
1	A North American Team	1
2	Acipactli	1
3	Acturus	1
4	Afghanistan	14
...
1179	Zambia	13
1180	Zefyros	1
1181	Zimbabwe	14
1182	Zut	1
1183	rn-2	1

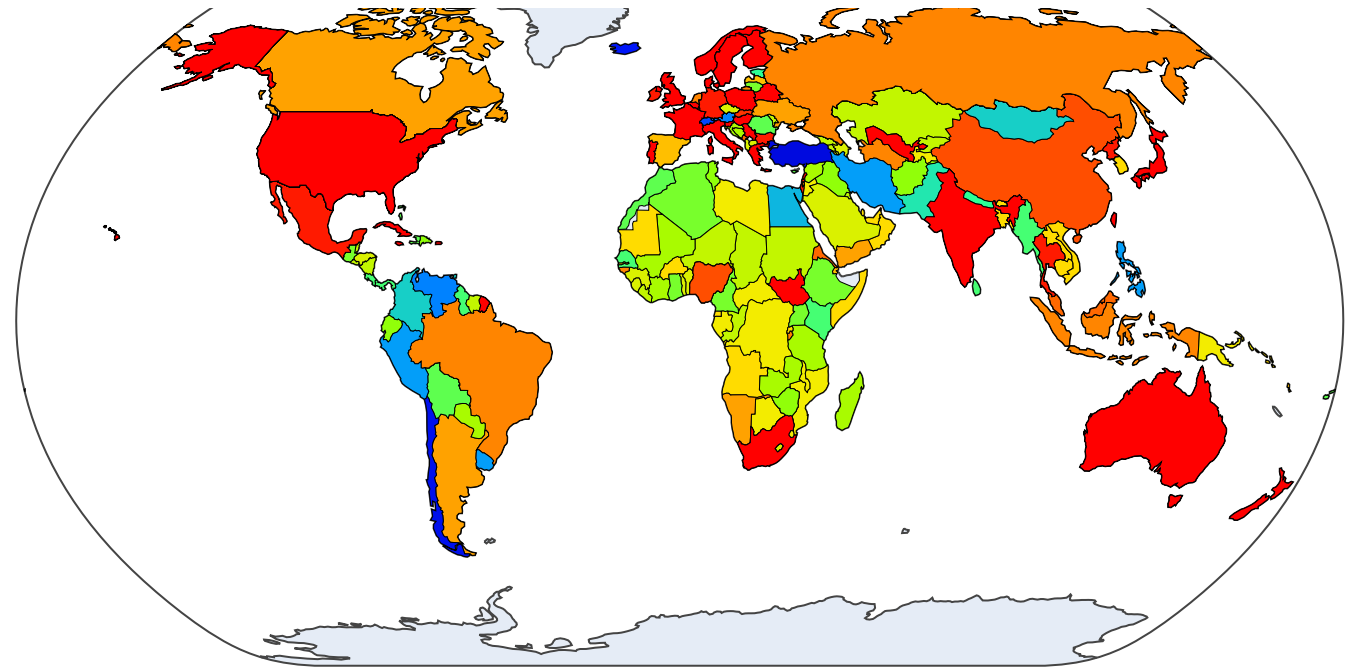
1184 rows × 2 columns

In []: 1

```
In [38]: 1 country_map = go.Choropleth(  
2         locations= country_df['Country'],  
3         locationmode='country names',  
4         z = country_df['Editions'],  
5         text = country_df['Country'],  
6         autocolorscale = False,  
7         reversescale = True,  
8         colorscale = 'rainbow',  
9         marker = dict(  
10            line =dict(  
11                color = 'black',  
12                width=0.5)  
13            ),  
14            colorbar = dict(  
15                title = 'Editions',  
16                tickprefix = '')  
17        )  
18  
19 data = [country_map]  
20 layout = go.Layout(  
21     title = 'Olympic Countries Participating',  
22     geo= dict(  
23         showframe = True,  
24         showlakes = False,  
25         showcoastlines = True,  
26         projection = dict(  
27             type = 'natural earth'  
28         )  
29     )  
30 )  
31  
32 fig = dict(data=data,layout=layout)  
33 py.iplot(fig)
```

Olympic Countries Participating

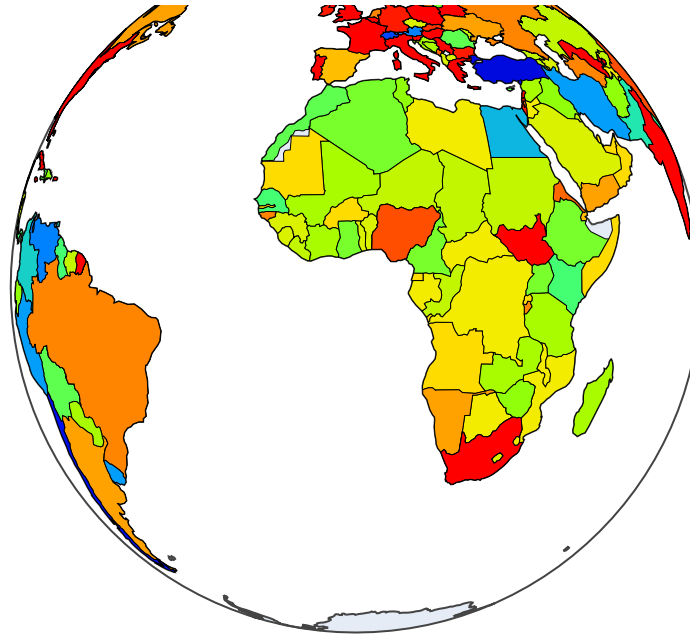




```
In [39]: 1 country_map = go.Choropleth(  
2         locations= country_df['Country'],  
3         locationmode='country names',  
4         z = country_df['Editions'],  
5         text = country_df['Country'],  
6         autocolorscale = False,  
7         reversescale = True,  
8         colorscale = 'rainbow',  
9         marker = dict(  
10            line =dict(  
11                color = 'black',  
12                width=0.5)  
13            ),  
14            colorbar = dict(  
15                title = 'Editions',  
16                tickprefix = '')  
17        )  
18  
19 data = [country_map]  
20 layout = go.Layout(  
21     title = 'Olympic Countries Participating',  
22     geo= dict(  
23         showframe = True,  
24         showlakes = False,  
25         showcoastlines = True,  
26         projection = dict(  
27             type = 'orthographic'  
28         )  
29     )  
30 )  
31  
32 fig = dict(data=data,layout=layout)  
33 py.iplot(fig)
```

Olympic Countries Participating





In [40]: 1 df_winter.head()

Out[40]:

	Year	City	Season	Athlets
7	1924	Chamonix	Winter	460
10	1928	Sankt Moritz	Winter	582
11	1932	Lake Placid	Winter	352
14	1936	Garmisch-Partenkirchen	Winter	895
16	1948	Sankt Moritz	Winter	1075

In []:

1

create dist plot

In [41]: 1 df.head()

Out[41]:

	ID	Name	Sex	Age	Height	Weight	Team	NOC	Games	Year	Season	City	Sport	Event	Medal
0	1	A Dijiang	M	24.0	180.0	80.0	China	CHN	1992 Summer	1992	Summer	Barcelona	Basketball	Basketball Men's Basketball	NaN
1	2	A Lamusi	M	23.0	170.0	60.0	China	CHN	2012 Summer	2012	Summer	London	Judo	Judo Men's Extra-Lightweight	NaN
2	3	Gunnar Nielsen Aaby	M	24.0	NaN	NaN	Denmark	DEN	1920 Summer	1920	Summer	Antwerpen	Football	Football Men's Football	NaN
3	4	Edgar Lindenau Aabye	M	34.0	NaN	NaN	Denmark/Sweden	DEN	1900 Summer	1900	Summer	Paris	Tug-Of-War	Tug-Of-War Men's Tug-Of-War	Gold
4	5	Christine Jacoba Aaftink	F	21.0	185.0	82.0	Netherlands	NED	1988 Winter	1988	Winter	Calgary	Speed Skating	Speed Skating Women's 500 metres	NaN

i have to show the height distribution for athlets grouped by gender

In [42]: 1 df[df['Sex'] == 'F']['Height'].dropna()

Out[42]:

```

4      185.0
5      185.0
6      185.0
7      185.0
8      185.0
...
271080  168.0
271099  171.0
271102  171.0
271103  171.0
271110  171.0
Name: Height, Length: 67378, dtype: float64

```

```
In [43]: 1 df[df['Sex'] == 'M']['Height'].dropna()
```

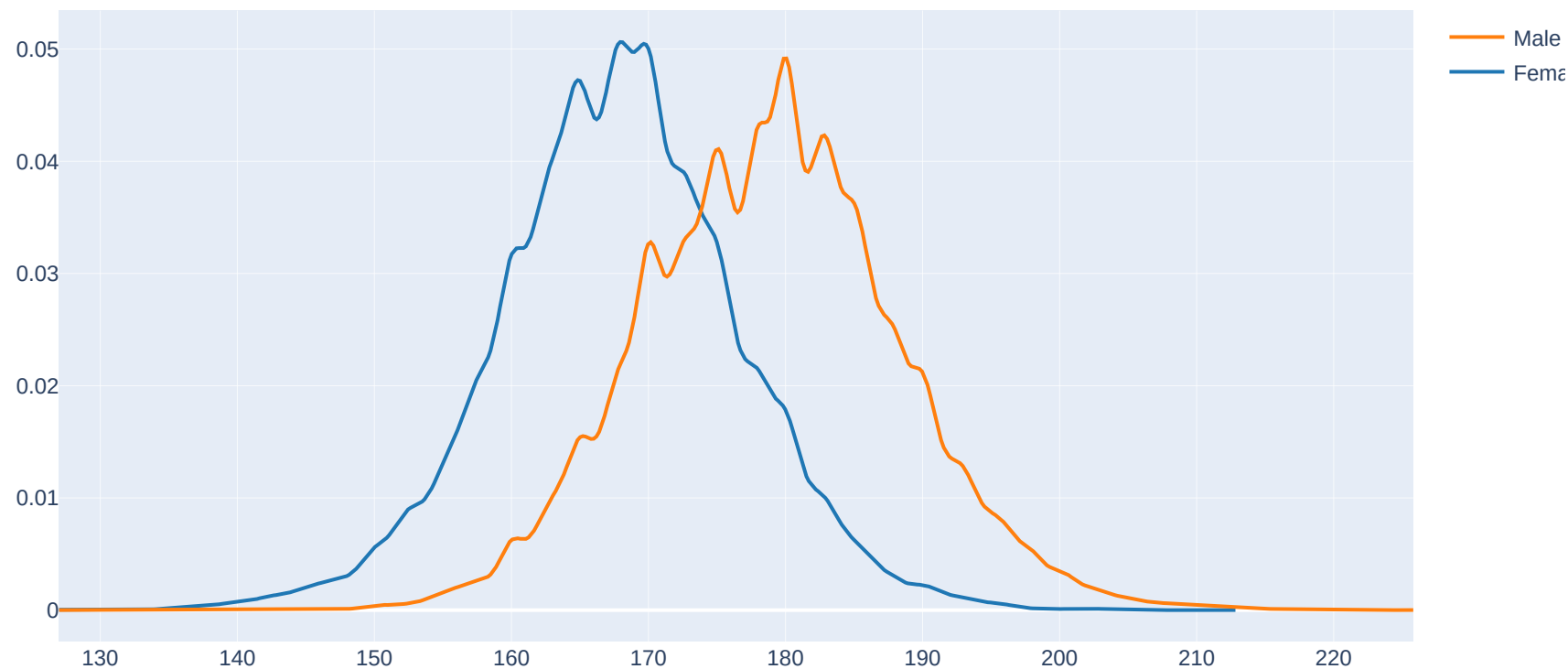
```
Out[43]: 0      180.0  
         1      170.0  
        10      188.0  
        11      188.0  
        12      188.0  
         ...  
       271111    179.0  
       271112    176.0  
       271113    176.0  
       271114    185.0  
       271115    185.0  
Name: Height, Length: 143567, dtype: float64
```

```
In [44]: 1 female_data = df[df['Sex'] == 'F']['Height'].dropna()
```

```
In [45]: 1 male_data = df[df['Sex'] == 'M']['Height'].dropna()
```

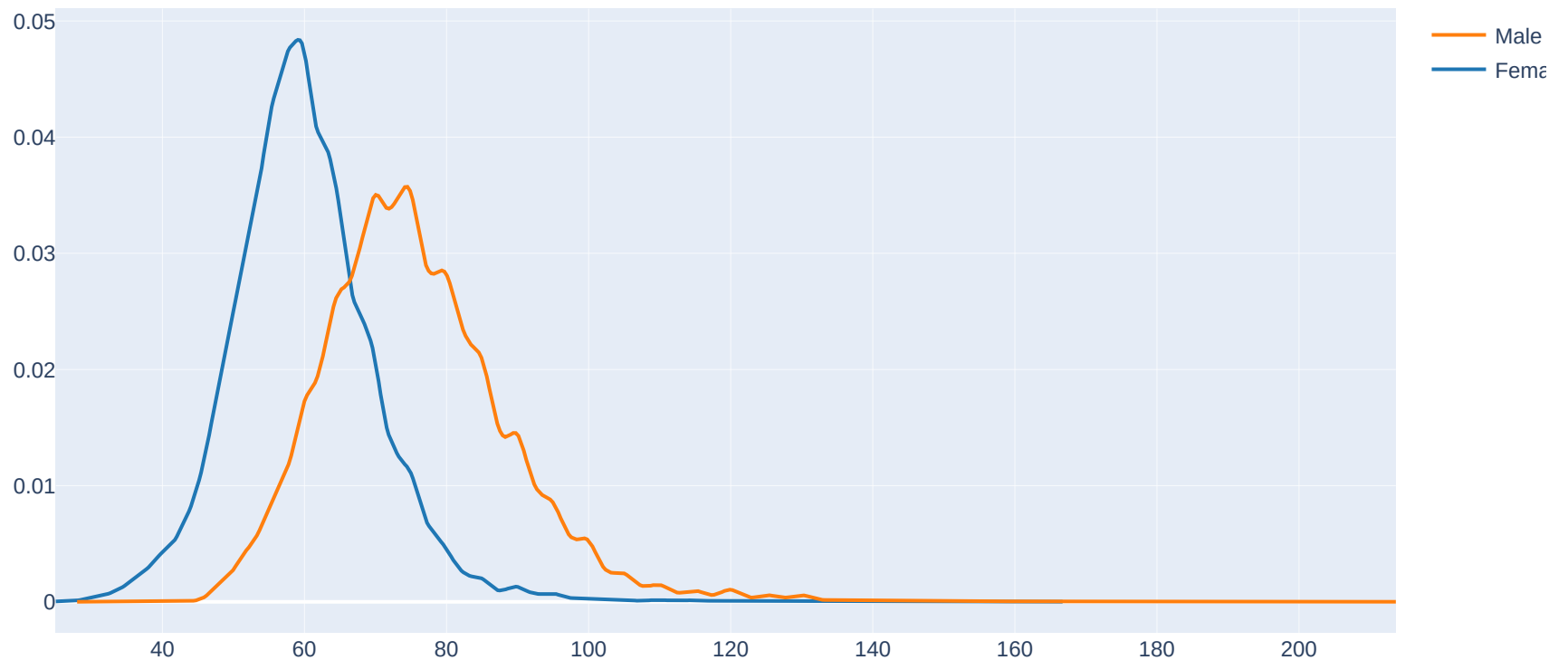
```
In [46]: 1 mydata = [female_data,male_data]
2 group_labels = ['Female Height','Male Height']
3 fig = ff.create_distplot(mydata,group_labels,show_hist=False,show_rug=False)
4 fig['layout'].update(title='Athlets Height distribution plot')
5 py.iplot(fig)
```

Athlets Height distribution plot



```
In [47]: 1 female_data = df[df['Sex']== 'F']['Weight'].dropna()
2 male_data = df[df['Sex'] == 'M']['Weight'].dropna()
3 mydata = [female_data,male_data]
4 group_labels = ['Female Height', 'Male Height']
5 fig = ff.create_distplot(mydata,group_labels,show_hist=False,show_rug=False)
6 fig['layout'].update(title='Athlets Weight distribution plot')
7 py.iplot(fig)
```

Athlets Weight distribution plot



In [56]: 1 df_summer.head()

Out[56]:

	Year	City	Season	Athlets
0	1896	Athina	Summer	380
1	1900	Paris	Summer	1936
2	1904	St. Louis	Summer	1301
3	1906	Athina	Summer	1733
4	1908	London	Summer	3101

box plot

In [64]:

```

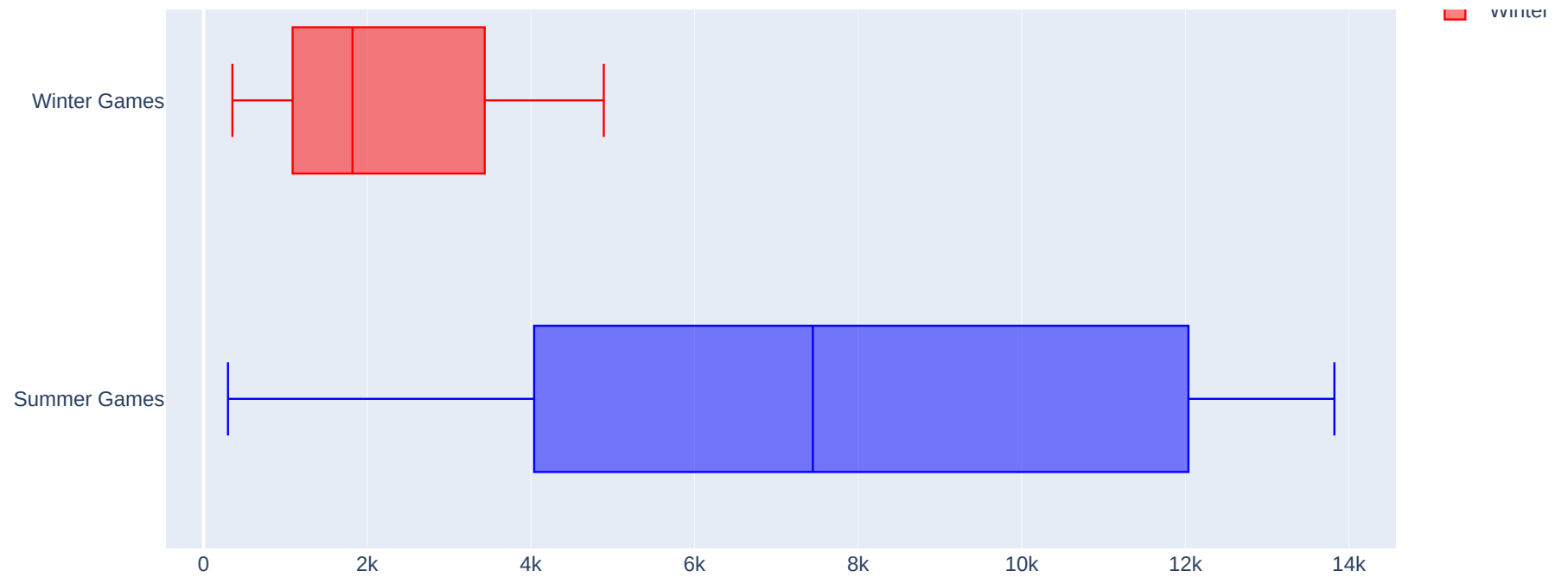
1 summer_data = go.Box(
2     x = df_summer['Athlets'],
3     name='Summer Games',
4     marker = dict(
5         color='blue'
6     ),
7
8     line = dict(
9         width=1.2,
10        #         color='orange'
11    ),
12    orientation = 'h'
13 )
14
15 winter_data = go.Box(
16     x = df_winter['Athlets'],
17     name='Winter Games',
18     marker = dict(
19         color= 'red',
20     ),
21
22     line = dict(
23         width=1.2
24     ),
25     orientation = 'h'
26 )
27
28
29 data = [summer_data,winter_data]
30 layout = dict(title = 'Athlets Per Olympic Games')
31 fig = dict(data=data,layout=layout)
32 py.iplot(fig)

```

Athlets Per Olympic Games

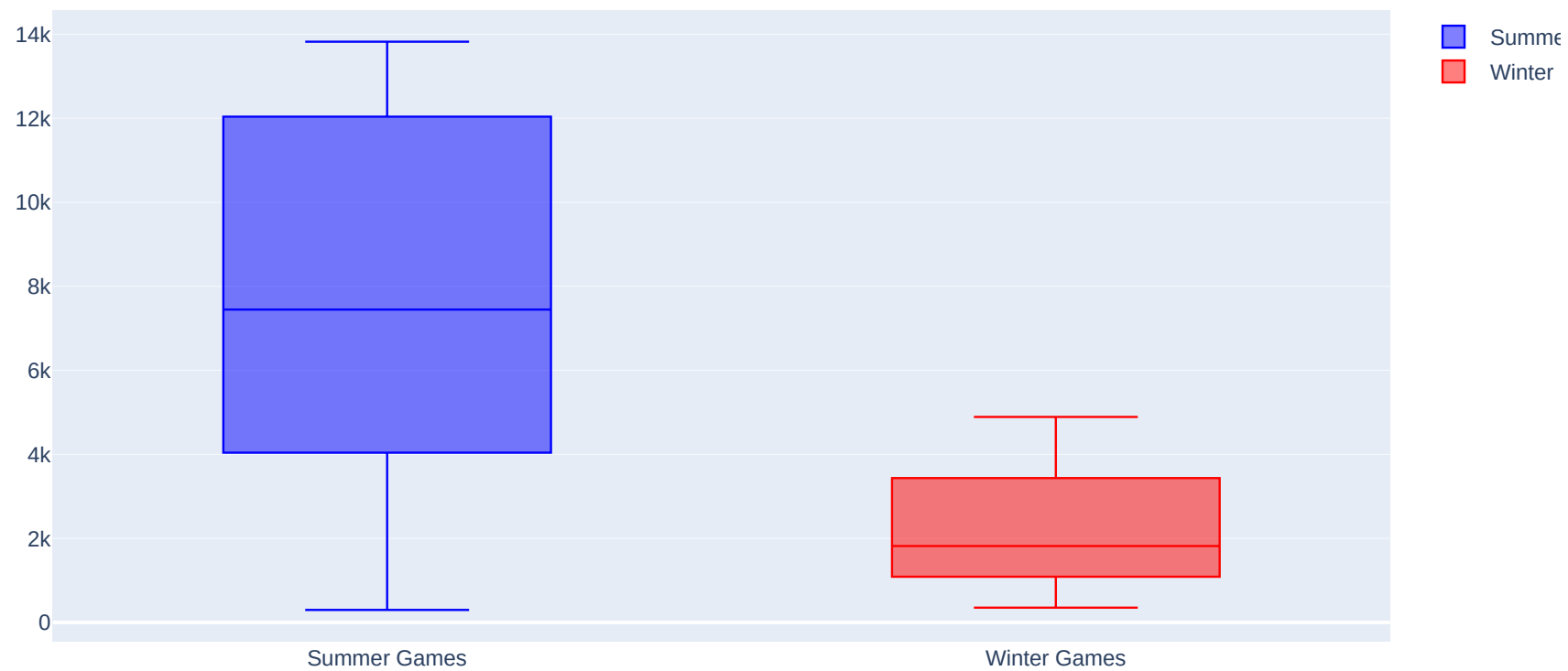


■ Summer
 ■ Winter



```
In [65]: 1 summer_data = go.Box(  
2     y = df_summer['Athlets'],  
3     name='Summer Games',  
4     marker = dict(  
5         color='blue'  
6     ),  
7  
8     line = dict(  
9         width=1.2,  
10    #         color='orange'  
11    ),  
12  
13     orientation = 'v'  
14 )  
15  
16 winter_data = go.Box(  
17     y = df_winter['Athlets'],  
18     name='Winter Games',  
19     marker = dict(  
20         color= 'red',  
21     ),  
22  
23     line = dict(  
24         width=1.2  
25     ),  
26  
27  
28     orientation = 'v'  
29 )  
30  
31  
32 data = [summer_data,winter_data]  
33 layout = dict(title = 'Athlets Per Olympic Games')  
34 fig = dict(data=data,layout=layout)  
35 py.iplot(fig)
```

Athlets Per Olympic Games



In []:

1