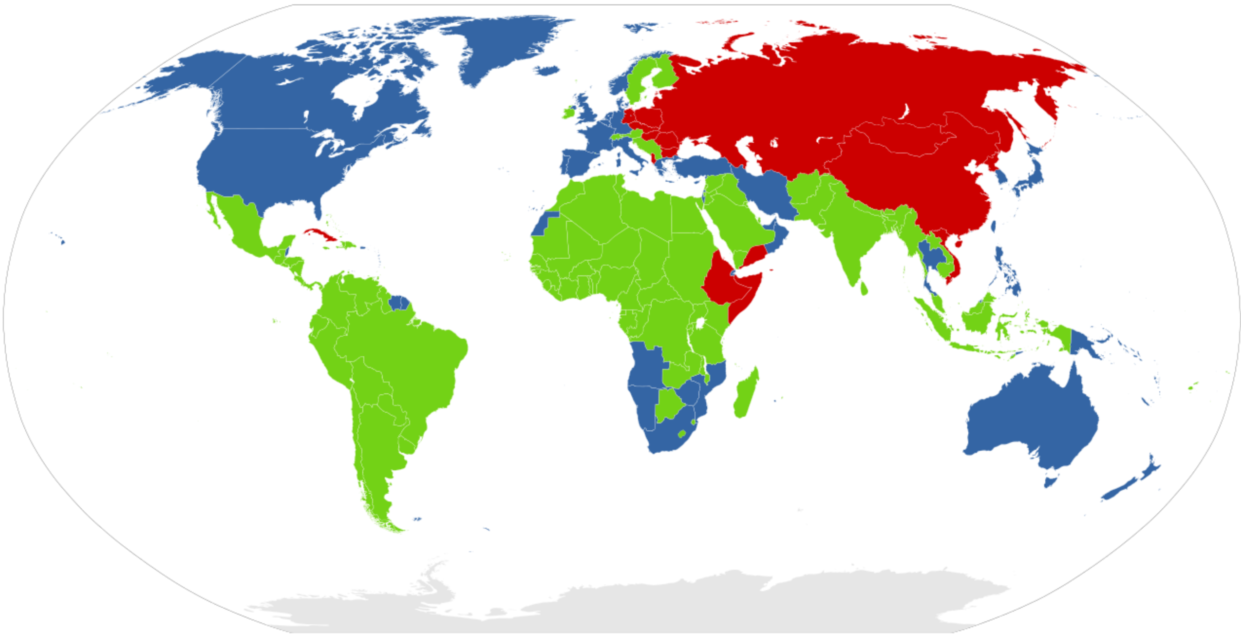
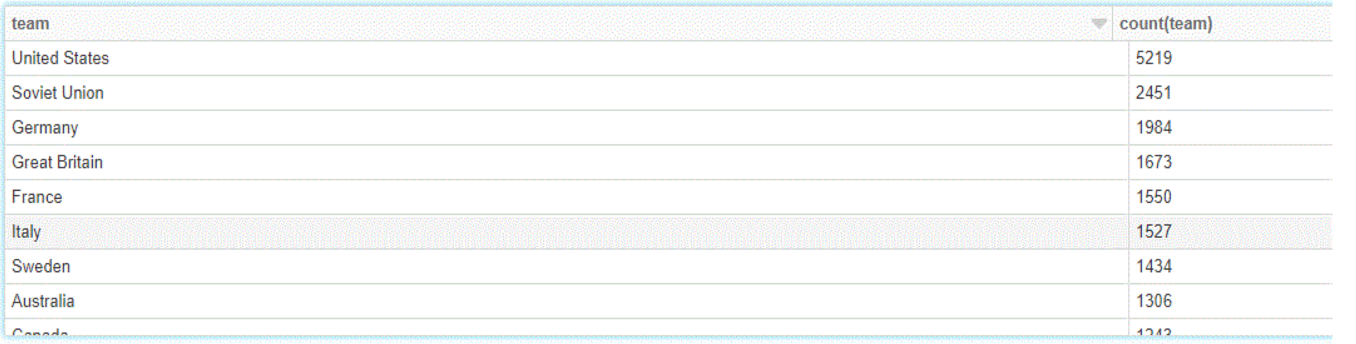
The state independency metric: how effective is state run sports system on the Olympic stage.



I have had a bit of experience of using Python for data analysis on academic projects and I wanted to do a simple straightforward project to combine Python and SQL to extract some data insights. I have come across an Olympics dataset – 120 years of data with every participant of the Olympic games since its beginning. Such analysis is interesting to perform and see, whether the system of state support in sport results in higher productivity of the athletes. As much s it is a method of propaganda, it also helps children find a purpose in life as well as helping people stay fit and healthy with mass sport culture. As trivial as that sounds, the success of the USSR`s athletes in the Olympics is a great example of that. While the country joined the IOC in 1954 and have been competing until 1992* the results are staggering and worth exploring a bit further.



team	count(team)
United States	5219
Soviet Union	2451
Germany	1984
Great Britain	1673
France	1550
Italy	1527
Sweden	1434
Australia	1306
Canada	1243

My hypothetical client:

I am working closely with SportsStats to find interesting insights for their partners. Upon reviewing the data provided, there can hopefully be made news worthy stories or health insights.

Hypotheses:

1) Outside of the USA, all of the state controlled sport system countries will perform better, than the amateur ones.
2) Each soviet country will have a time lag, in which the new system will be implemented. Therefore, the results of all will improve after 3 Olympic cycles from start of participation

Approach:

1. Import and clean up data; split the raw dataset into 1st and 2nd word countries tables

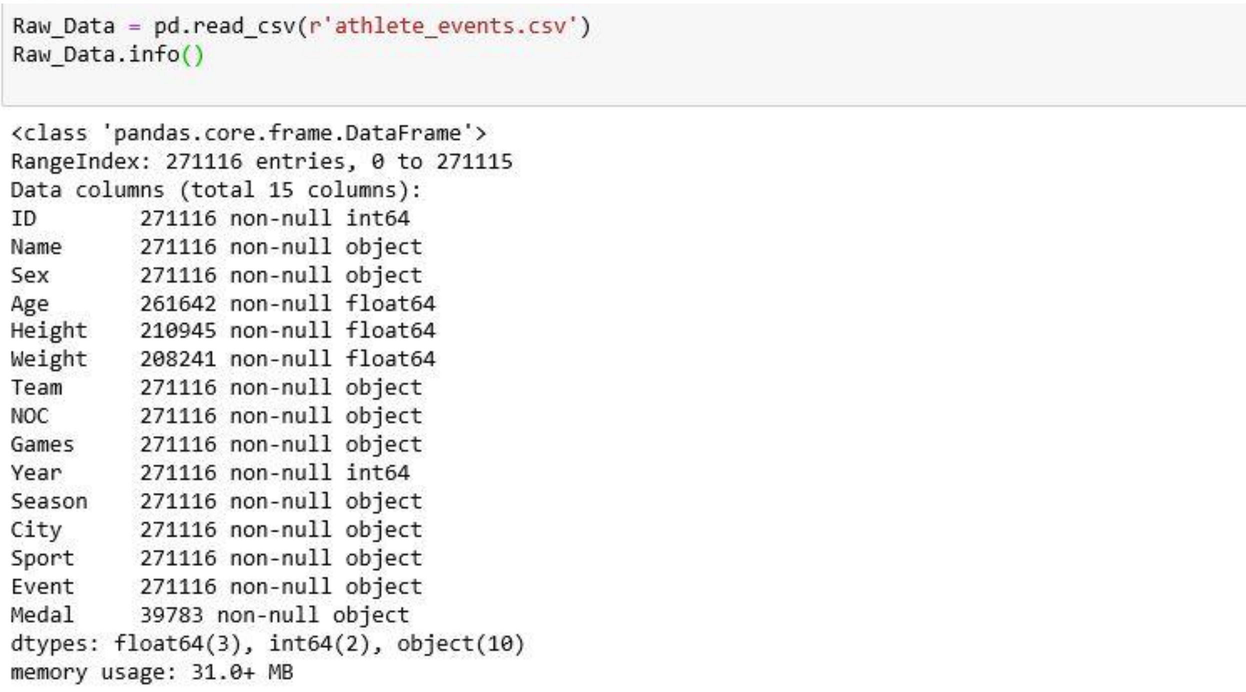
2. Filter against teams and years to see performance

3. Run statistical analysis via Pearson and P-value to find correlations

Step 1: Import and clean up data.

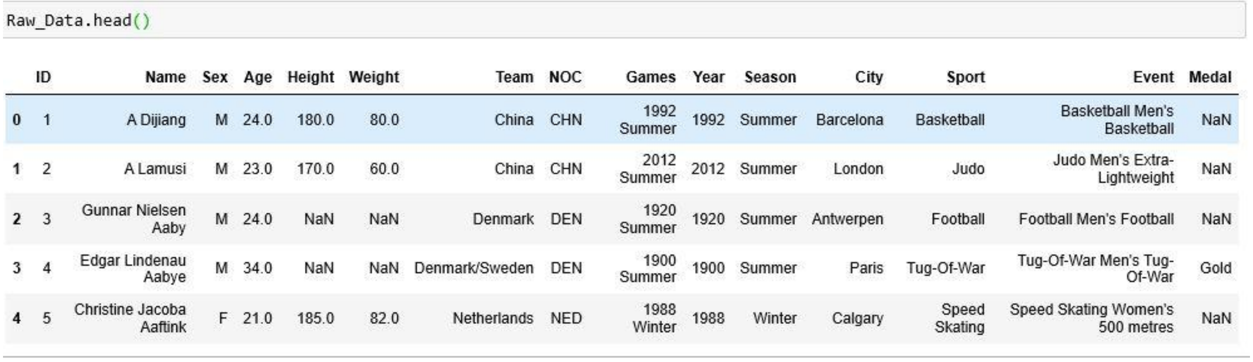
We begin by importing libraries and the csv of the dataset. Once imported, it is best to familiarize ourselves as to what data we have got.

```
import pandas as pd
import numpy as np
import matplotlib
import matplotlib.pyplot as plt
```



```
Raw_Data = pd.read_csv(r'athlete_events.csv')
Raw_Data.info()
```

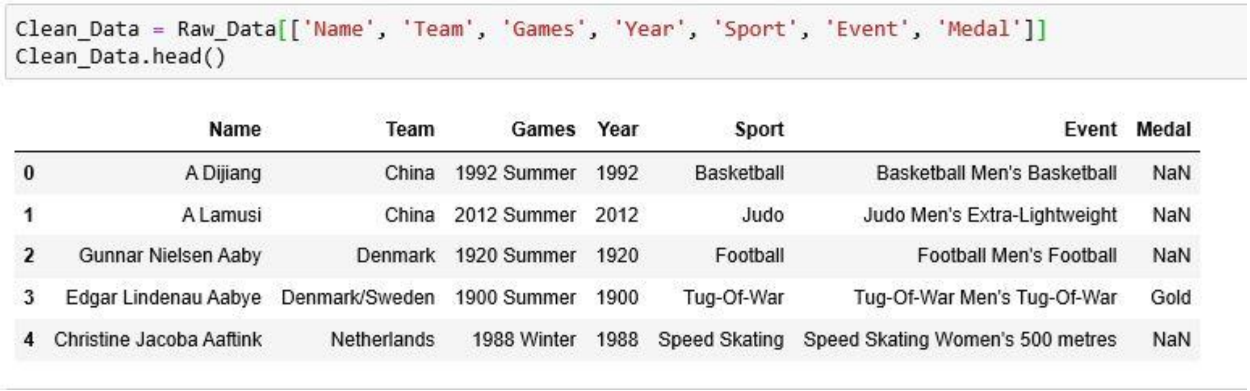
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 271116 entries, 0 to 271115
Data columns (total 15 columns):
ID        271116 non-null int64
Name      271116 non-null object
Sex       271116 non-null object
Age       261642 non-null float64
Height    210945 non-null float64
Weight    208241 non-null float64
Team      271116 non-null object
NOC       271116 non-null object
Games     271116 non-null object
Year      271116 non-null int64
Season    271116 non-null object
City      271116 non-null object
Sport     271116 non-null object
Event     271116 non-null object
Medal     39783 non-null object
dtypes: float64(3), int64(2), object(10)
memory usage: 31.0+ MB
```



```
Raw_Data.head()
```

	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

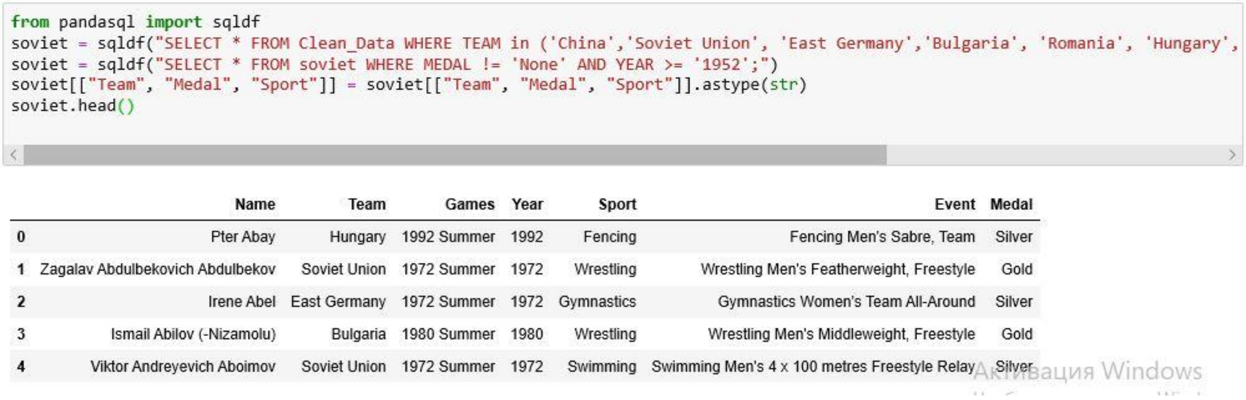
After looking through the columns, it is quite clear, that not all of the data provided will be necessary for the analysis. So we will drop some of them altogether.



```
Clean_Data = Raw_Data[['Name', 'Team', 'Games', 'Year', 'Sport', 'Event', 'Medal']]
Clean_Data.head()
```

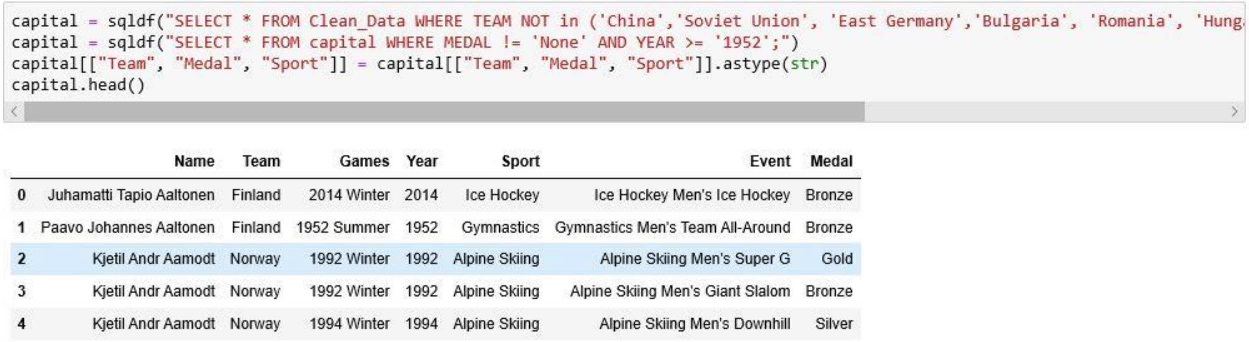
	Name	Team	Games	Year	Sport	Event	Medal
0	A Dijiang	China	1992 Summer	1992	Basketball	Basketball Men's Basketball	NaN
1	A Lamusi	China	2012 Summer	2012	Judo	Judo Men's Extra-Lightweight	NaN
2	Gunnar Nielsen Aaby	Denmark	1920 Summer	1920	Football	Football Men's Football	NaN
3	Edgar Lindenau Aabye	Denmark/Sweden	1900 Summer	1900	Tug-Of-War	Tug-Of-War Men's Tug-Of-War	Gold
4	Christine Jacoba Aaftink	Netherlands	1988 Winter	1988	Speed Skating	Speed Skating Women's 500 metres	NaN

Once we got the required columns filtered, it is time to split the dataset into soviet and capitalist. For the more accurate results, we will use the time constraints from 1954 -1992, since within that time the soviet system was implemented and active in the Warsaw pact countries.



```
from pandasql import sqldf
soviet = sqldf("SELECT * FROM Clean_Data WHERE TEAM in ('China','Soviet Union', 'East Germany','Bulgaria', 'Romania', 'Hungary',
soviet = sqldf("SELECT * FROM soviet WHERE MEDAL != 'None' AND YEAR >= '1952';")
soviet[["Team", "Medal", "Sport"]] = soviet[["Team", "Medal", "Sport"]].astype(str)
soviet.head()
```

	Name	Team	Games	Year	Sport	Event	Medal
0	Pter Abay	Hungary	1992 Summer	1992	Fencing	Fencing Men's Sabre, Team	Silver
1	Zagalav Abdulbekovich Abdulbekov	Soviet Union	1972 Summer	1972	Wrestling	Wrestling Men's Featherweight, Freestyle	Gold
2	Irene Abel	East Germany	1972 Summer	1972	Gymnastics	Gymnastics Women's Team All-Around	Silver
3	Ismail Abilov (-Nizamolu)	Bulgaria	1980 Summer	1980	Wrestling	Wrestling Men's Middleweight, Freestyle	Gold
4	Viktor Andreyevich Aboimov	Soviet Union	1972 Summer	1972	Swimming	Swimming Men's 4 x 100 metres Freestyle Relay	Silver

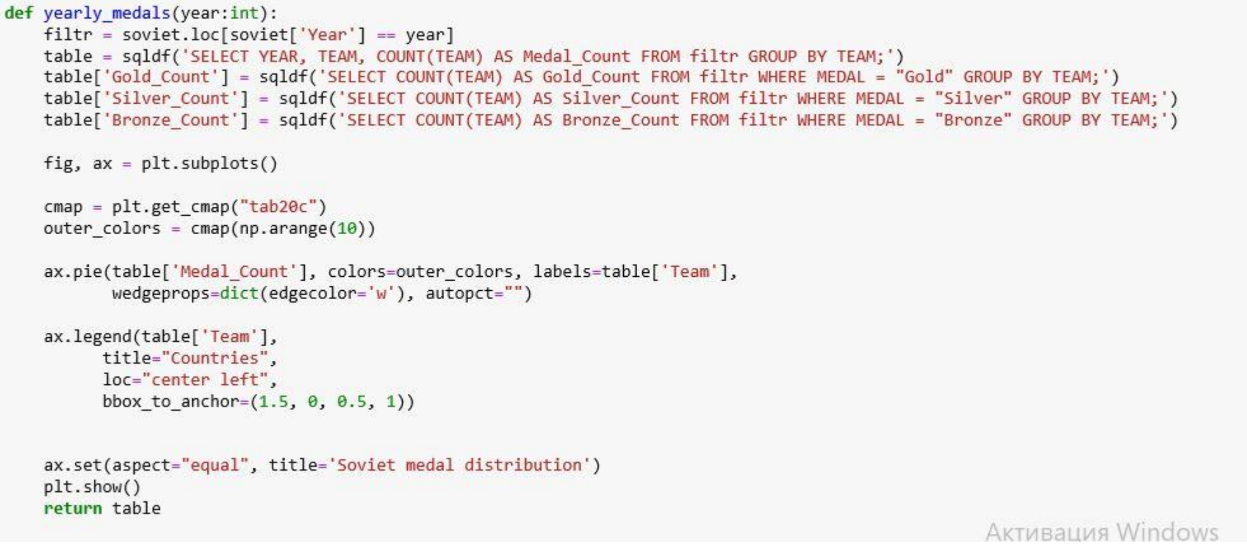


```
capital = sqldf("SELECT * FROM Clean_Data WHERE TEAM NOT in ('China','Soviet Union', 'East Germany','Bulgaria', 'Romania', 'Hung
capital = sqldf("SELECT * FROM capital WHERE MEDAL != 'None' AND YEAR >= '1952';")
capital[["Team", "Medal", "Sport"]] = capital[["Team", "Medal", "Sport"]].astype(str)
capital.head()
```

	Name	Team	Games	Year	Sport	Event	Medal
0	Juhamatti Tapio Aaltonen	Finland	2014 Winter	2014	Ice Hockey	Ice Hockey Men's Ice Hockey	Bronze
1	Paavo Johannes Aaltonen	Finland	1952 Summer	1952	Gymnastics	Gymnastics Men's Team All-Around	Bronze
2	Kjetil Andr Aamodt	Norway	1992 Winter	1992	Alpine Skiing	Alpine Skiing Men's Super G	Gold
3	Kjetil Andr Aamodt	Norway	1992 Winter	1992	Alpine Skiing	Alpine Skiing Men's Giant Slalom	Bronze
4	Kjetil Andr Aamodt	Norway	1994 Winter	1994	Alpine Skiing	Alpine Skiing Men's Downhill	Silver

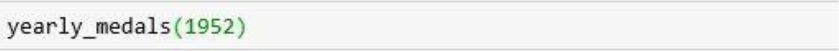
Step 2: Filter against teams and years to see performance

Once we have our countries split into tables, it is time to see, how they performed on the global scale! We will create a function to count the medals of each country, as well as the overall contribution to the soviet pool of medals won.

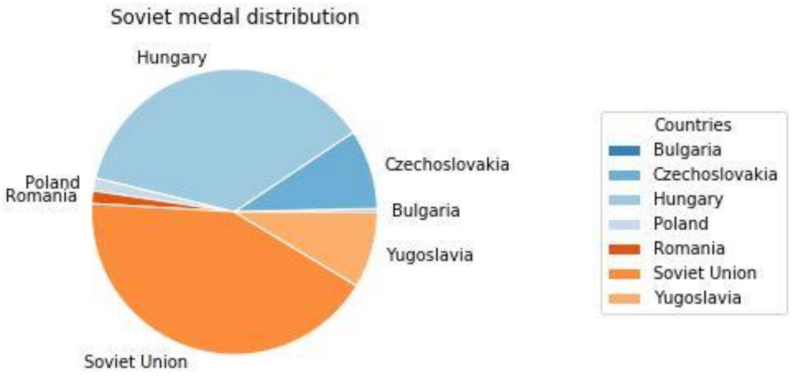


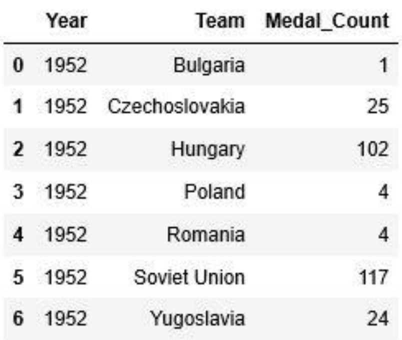
```
def yearly_medals(year:int):
    filtr = soviet.loc[soviet['Year'] == year]
    table = sqldf('SELECT YEAR, TEAM, COUNT(TEAM) AS Medal_Count FROM filtr GROUP BY TEAM;')
    table['Gold_Count'] = sqldf('SELECT COUNT(TEAM) AS Gold_Count FROM filtr WHERE MEDAL = "Gold" GROUP BY TEAM;')
    table['Silver_Count'] = sqldf('SELECT COUNT(TEAM) AS Silver_Count FROM filtr WHERE MEDAL = "Silver" GROUP BY TEAM;')
    table['Bronze_Count'] = sqldf('SELECT COUNT(TEAM) AS Bronze_Count FROM filtr WHERE MEDAL = "Bronze" GROUP BY TEAM;')

    fig, ax = plt.subplots()

    cmap = plt.get_cmap("tab20c")
    outer_colors = cmap(np.arange(10))

    ax.pie(table['Medal_Count'], colors=outer_colors, labels=table['Team'],
           wedgeprops=dict(edgecolor='w'), autopct="")

    ax.legend(table['Team'],
          title="Countries",
          loc="center left",
          bbox_to_anchor=(1.5, 0, 0.5, 1))


    ax.set(aspect="equal", title='Soviet medal distribution')
    plt.show()
    return table
```

When we run the function with the first 3 Olympic cycles, we et the following:

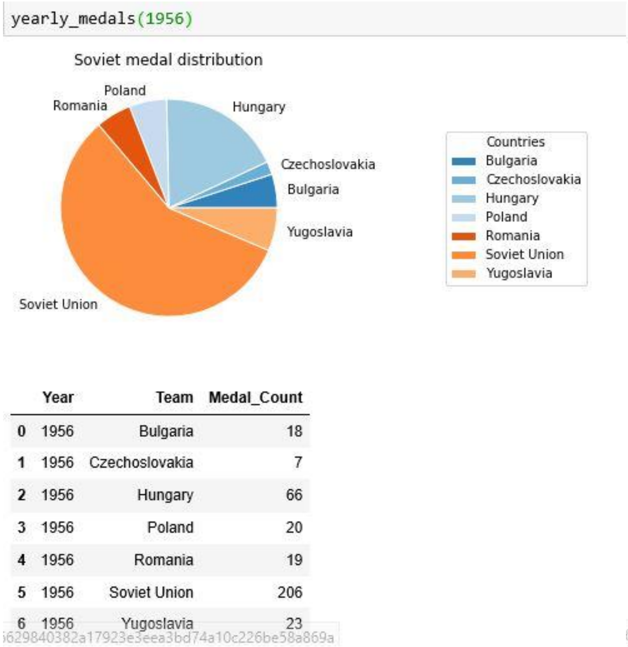


```
yearly_medals(1952)
```



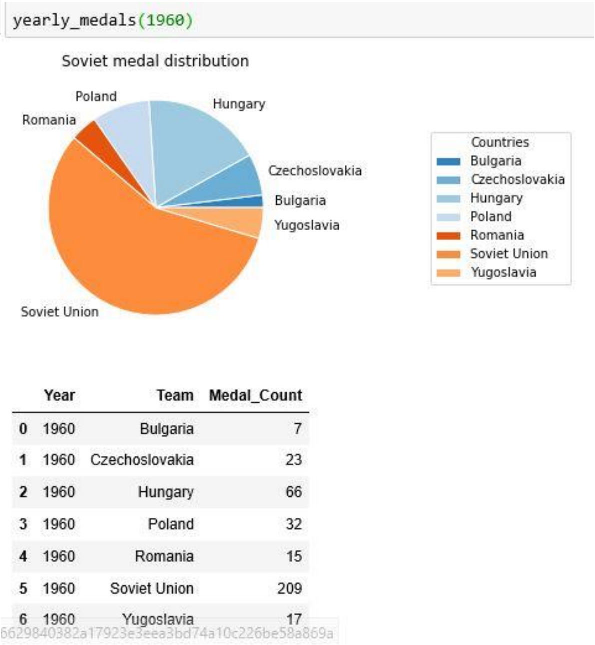


	Year	Team	Medal_Count
0	1952	Bulgaria	1
1	1952	Czechoslovakia	25
2	1952	Hungary	102
3	1952	Poland	4
4	1952	Romania	4
5	1952	Soviet Union	117
6	1952	Yugoslavia	24



```
yearly_medals(1956)
```

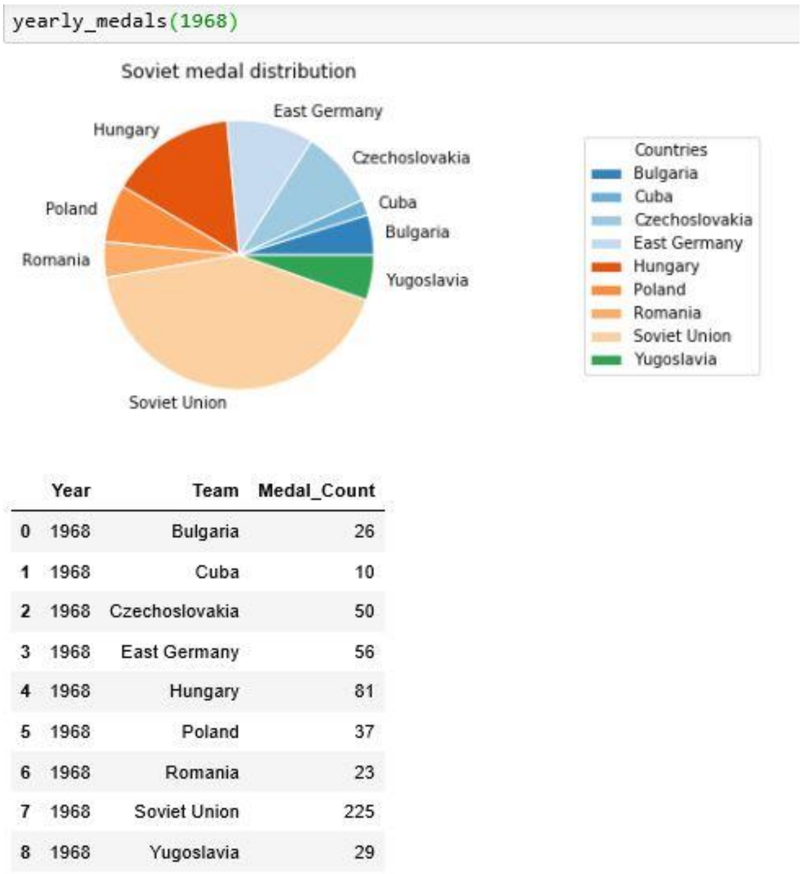
	Year	Team	Medal_Count
0	1956	Bulgaria	18
1	1956	Czechoslovakia	7
2	1956	Hungary	66
3	1956	Poland	20
4	1956	Romania	19
5	1956	Soviet Union	206
6	1956	Yugoslavia	23



```
yearly_medals(1960)
```

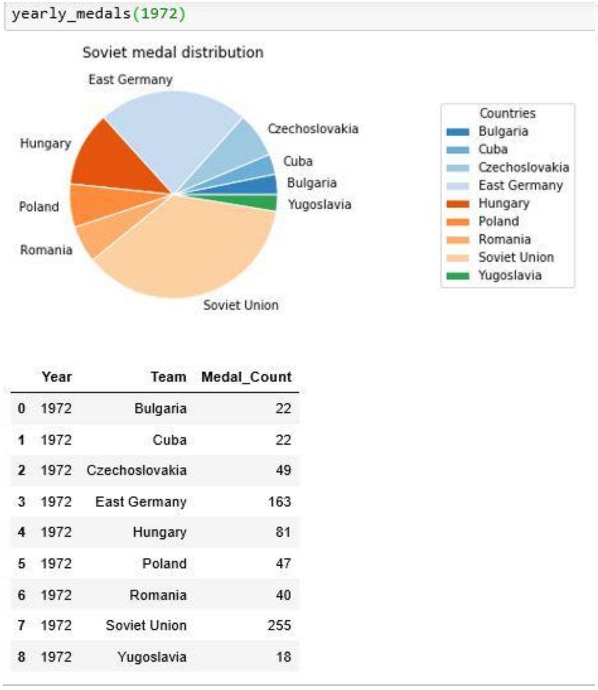
	Year	Team	Medal_Count
0	1960	Bulgaria	7
1	1960	Czechoslovakia	23
2	1960	Hungary	66
3	1960	Poland	32
4	1960	Romania	15
5	1960	Soviet Union	209
6	1960	Yugoslavia	17

Judging by the numbers presented, the trend is not consistent across the board. USSR, Bulgaria, Czechoslovakia, Poland and Romania show increase in medals as the time passes. However, Hungary (who finished 3rd overall in 154 Olympics) and Yugoslavia show downward trend. It is also worth noting, that until 1964, Germany participated as a unified country, so it will be a good idea to look into the medal situation from 1968 games.



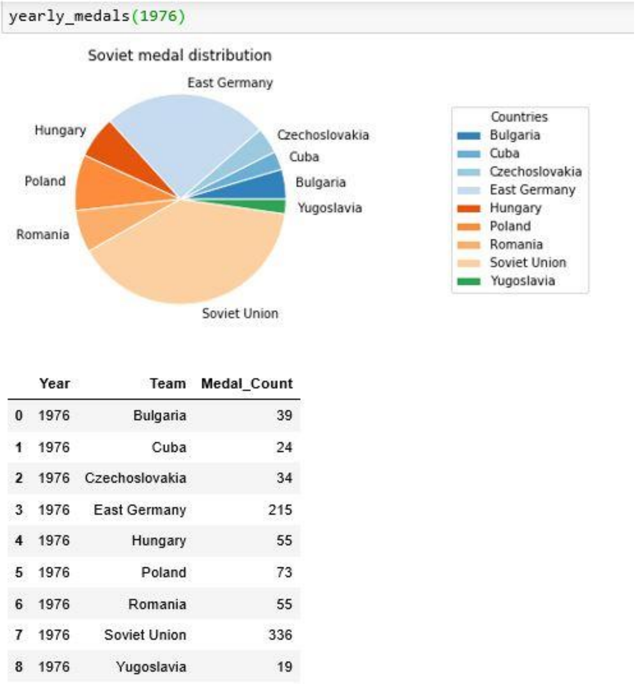
```
yearly_medals(1968)
```

	Year	Team	Medal_Count
0	1968	Bulgaria	26
1	1968	Cuba	10
2	1968	Czechoslovakia	50
3	1968	East Germany	56
4	1968	Hungary	81
5	1968	Poland	37
6	1968	Romania	23
7	1968	Soviet Union	225
8	1968	Yugoslavia	29



```
yearly_medals(1972)
```

	Year	Team	Medal_Count
0	1972	Bulgaria	22
1	1972	Cuba	22
2	1972	Czechoslovakia	49
3	1972	East Germany	163
4	1972	Hungary	81
5	1972	Poland	47
6	1972	Romania	40
7	1972	Soviet Union	255
8	1972	Yugoslavia	18

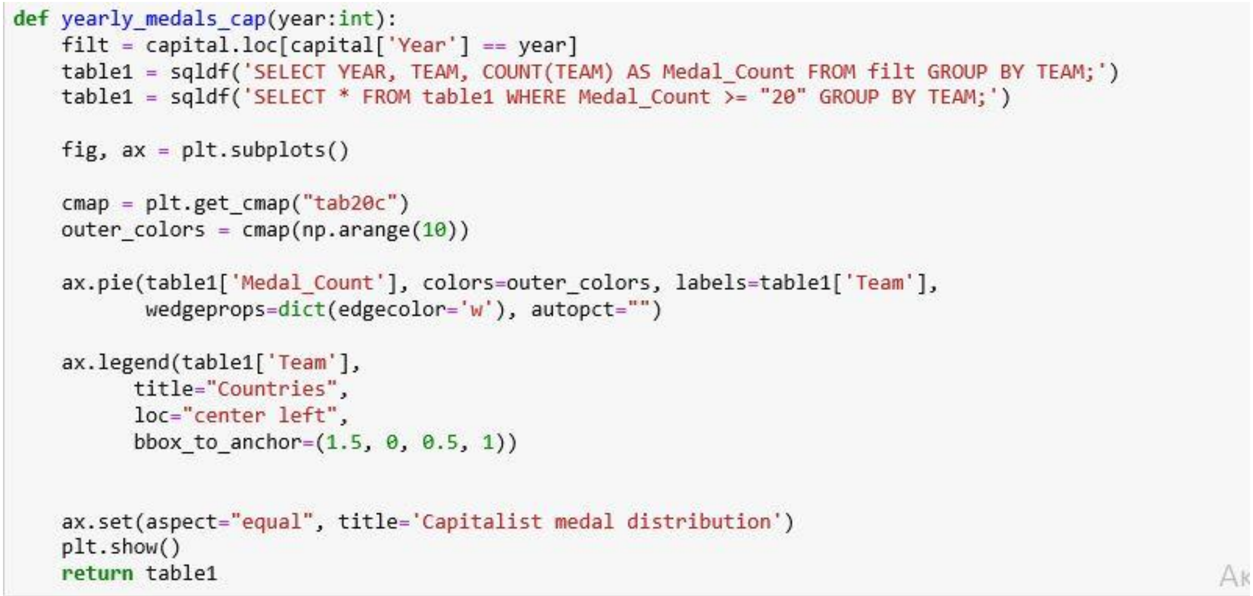


```
yearly_medals(1976)
```

	Year	Team	Medal_Count
0	1976	Bulgaria	39
1	1976	Cuba	24
2	1976	Czechoslovakia	34
3	1976	East Germany	215
4	1976	Hungary	55
5	1976	Poland	73
6	1976	Romania	55
7	1976	Soviet Union	336
8	1976	Yugoslavia	19

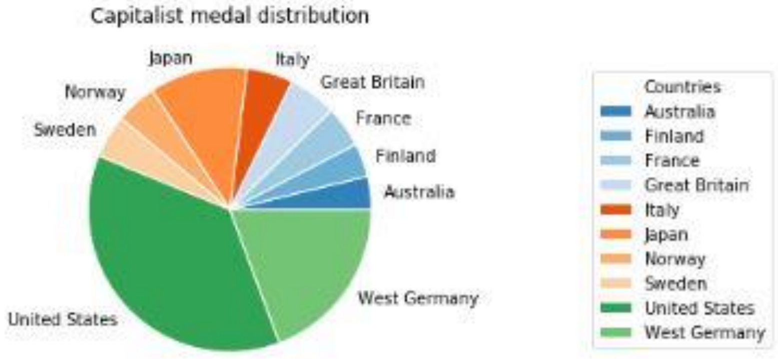
With the introduction of East Germany, most teams show the similar trend. USSR, East Germany, Bulgaria, Cuba, Poland, and Romania show increase in medals as the time passes. And Hungary, Yugoslavia (after a slight increase) and Czechoslovakia show downward trend. Which makes our second hypothesis a bit of a mixed batch – most countries start showing better results with time, but there are some of the countries, that show opposite trend.

We will use the same build of the function to establish, which countries in the capitalist block contribute most to overall medal performance.



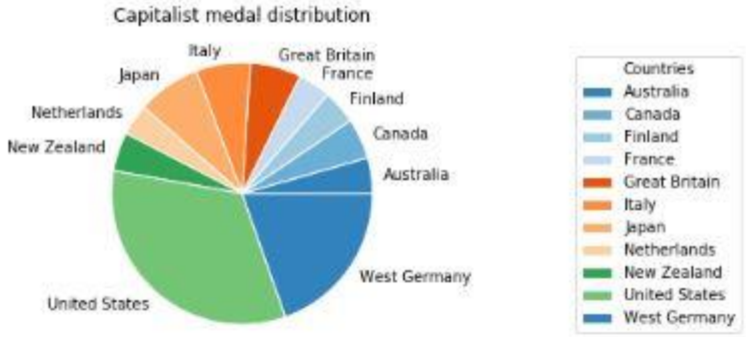
```
def yearly_medals_cap(year:int):
    filt = capital.loc[capital['Year'] == year]
    table1 = sqldf('SELECT YEAR, TEAM, COUNT(TEAM) AS Medal_Count FROM filt GROUP BY TEAM;')
    table1 = sqldf('SELECT * FROM table1 WHERE Medal_Count >= "20" GROUP BY TEAM;')

    fig, ax = plt.subplots()

    cmap = plt.get_cmap("tab20c")
    outer_colors = cmap(np.arange(10))

    ax.pie(table1['Medal_Count'], colors=outer_colors, labels=table1['Team'],
           wedgeprops=dict(edgecolor='w'), autopct="")

    ax.legend(table1['Team'],
          title="Countries",
          loc="center left",
          bbox_to_anchor=(1.5, 0, 0.5, 1))


    ax.set(aspect="equal", title='Capitalist medal distribution')
    plt.show()
    return table1
```

```
yearly_medals_cap(1972)
```



	Year	Team	Medal_Count
0	1972	Australia	20
1	1972	Finland	20
2	1972	France	25
3	1972	Great Britain	29
4	1972	Italy	28
5	1972	Japan	59
6	1972	Norway	25
7	1972	Sweden	26
8	1972	United States	195
9	1972	West Germany	102

```
yearly_medals_cap(1976)
```

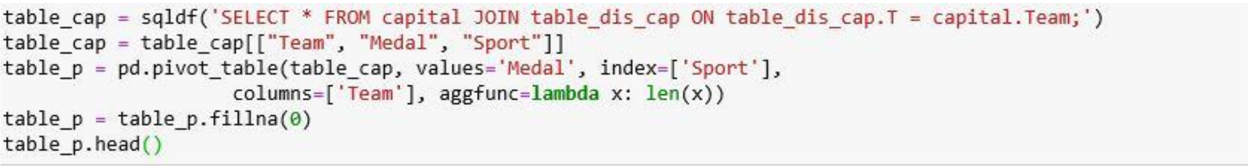


	Year	Team	Medal_Count
0	1976	Australia	23
1	1976	Canada	26
2	1976	Finland	22
3	1976	France	21
4	1976	Great Britain	33
5	1976	Italy	35
6	1976	Japan	41
7	1976	Netherlands	20
8	1976	New Zealand	25
9	1976	United States	173
10	1976	West Germany	102

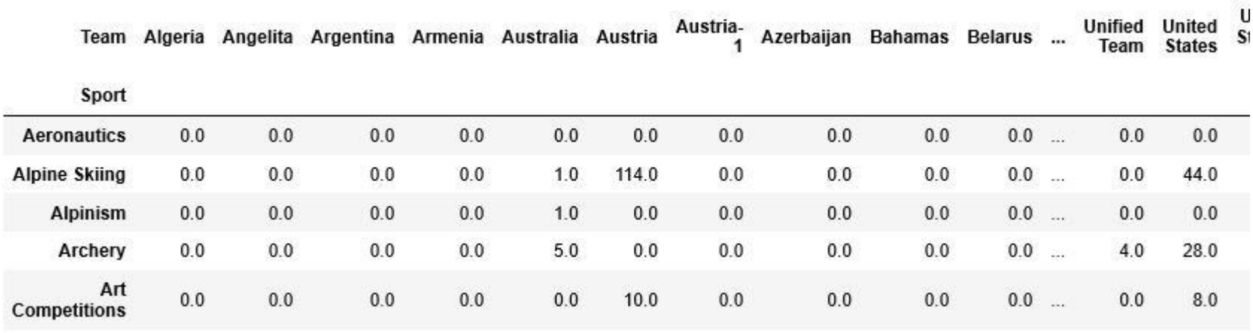
The two games summaries provided allow us to compare the performance of the soviet against capitalist countries. Once the metric was calculated, it turned out that highest performing capitalist countries were more productive until 1968. However, outside of USA, these countries performed betterm than their soviet counterparts until the rise of East Germany and its contributions, tipping the scales from 1972 onward.

(*NOTE: we excluded 1980 & 1984 Olympics, due to the boycott from capitalist countries in 1980 and soviets in 1984 respectively.)

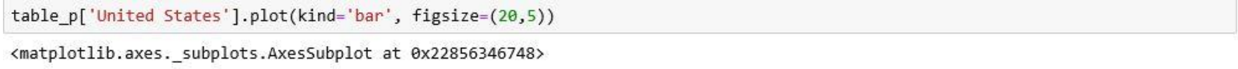
After running the function on all of the games, we managed to isolate the best performing countries. Now we can create a table of their most productive disciplines, so we can determine, which system was best at a given sport and what field was proving most fruitful for each block.



```
table_cap = sqldf('SELECT * FROM capital JOIN table_dis_cap ON table_dis_cap.T = capital.Team;')
table_cap = table_cap[["Team", "Medal", "Sport"]]
table_p = pd.pivot_table(table_cap, values='Medal', index=['Sport'],
                    columns=['Team'], aggfunc=lambda x: len(x))
table_p = table_p.fillna(0)
table_p.head()
```

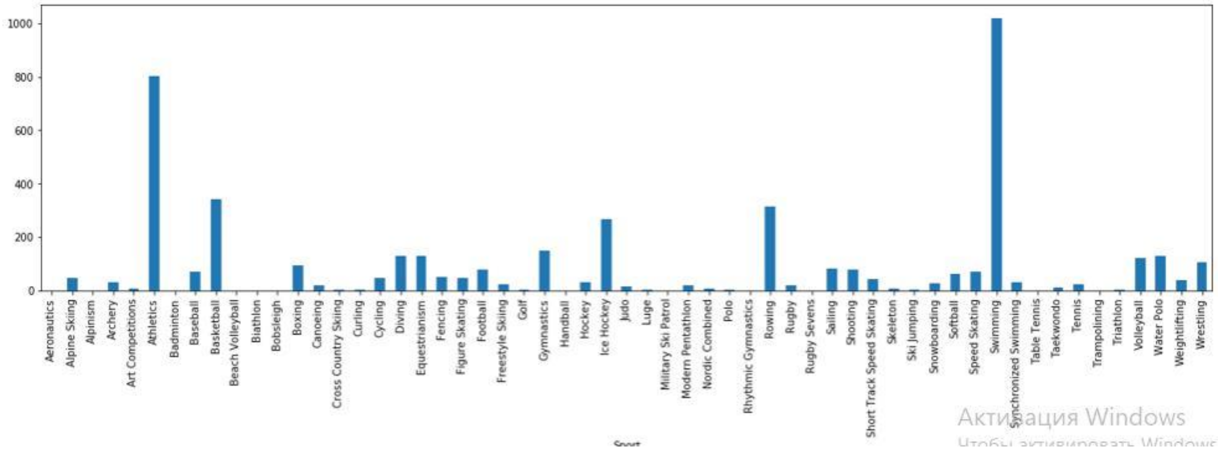


Team	Algeria	Angelita	Argentina	Armenia	Australia	Austria	Austria-1	Azerbaijan	Bahamas	Belarus	...	Unified Team	United States	U St
Sport														
Aeronautics	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	
Alpine Skiing	0.0	0.0	0.0	0.0	1.0	114.0	0.0	0.0	0.0	0.0	...	0.0	44.0	
Alpinism	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	
Archery	0.0	0.0	0.0	0.0	5.0	0.0	0.0	0.0	0.0	0.0	...	4.0	28.0	
Art Competitions	0.0	0.0	0.0	0.0	0.0	10.0	0.0	0.0	0.0	0.0	...	0.0	8.0	

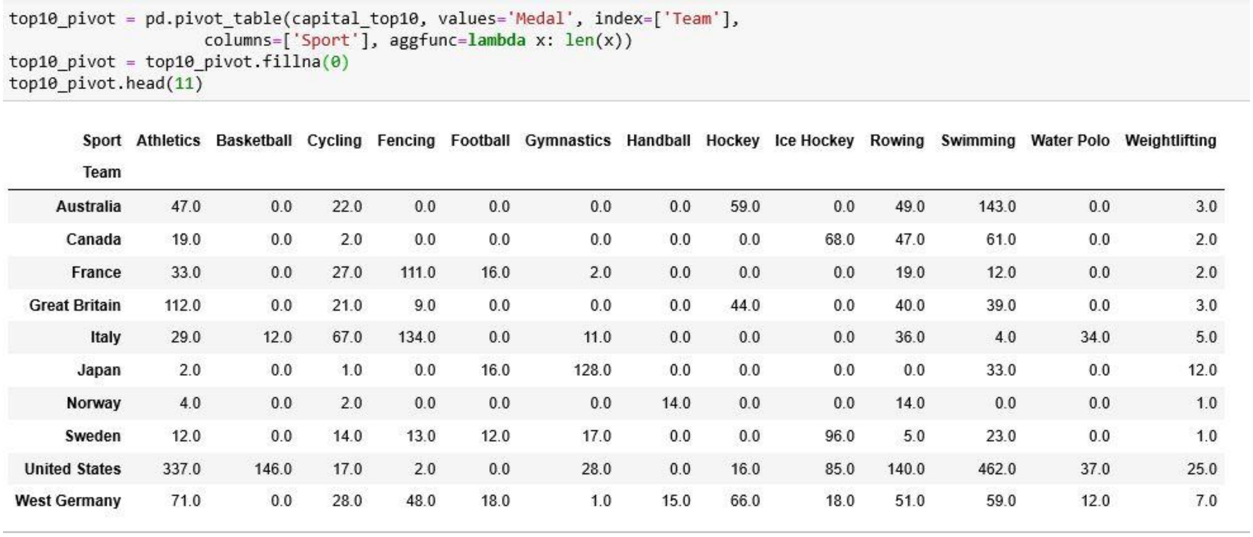


```
table_p['United States'].plot(kind='bar', figsize=(20,5))
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x22856346748>
```

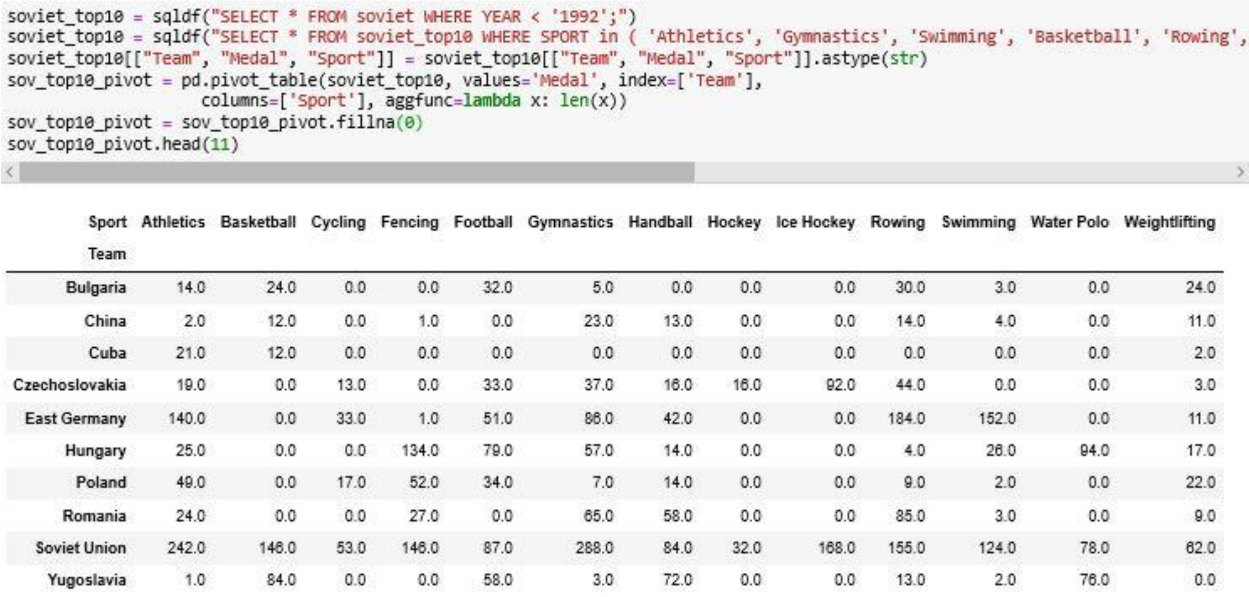


We can clearly see on the bar chart above which sports were most accountable for the USA success on the Olympics. Running these charts for every team on our lists, we will create the performance table for both blocks.



```
top10_pivot = pd.pivot_table(capital_top10, values='Medal', index=['Team'],
                    columns=['Sport'], aggfunc=lambda x: len(x))
top10_pivot = top10_pivot.fillna(0)
top10_pivot.head(11)
```

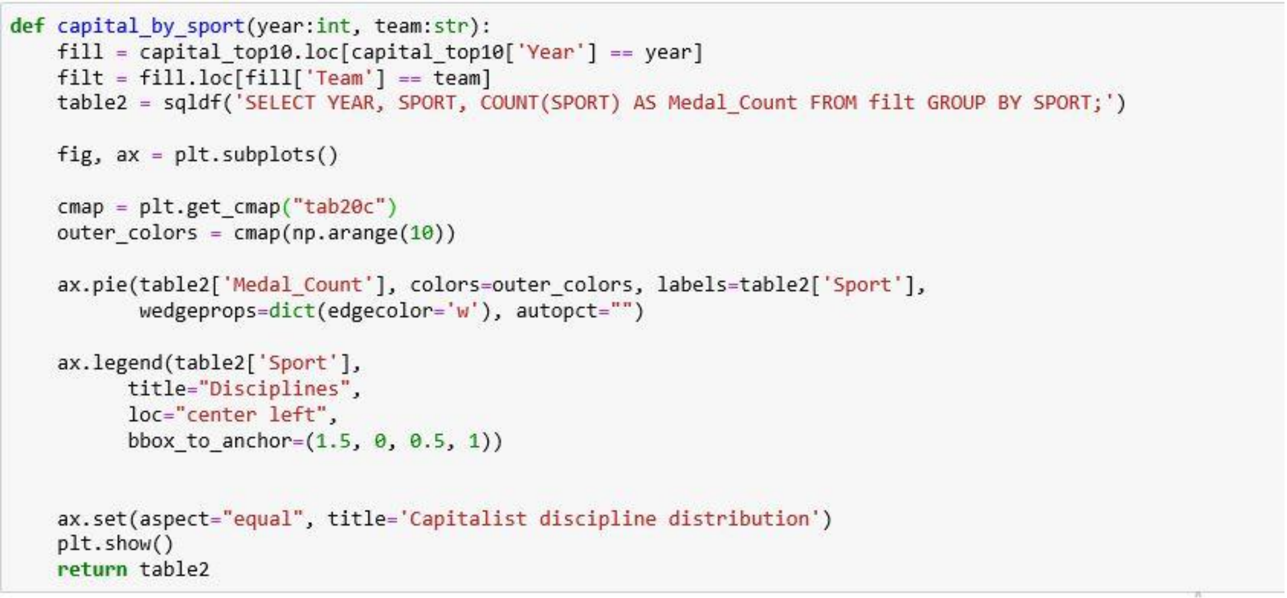
Sport	Athletics	Basketball	Cycling	Fencing	Football	Gymnastics	Handball	Hockey	Ice Hockey	Rowing	Swimming	Water Polo	Weightlifting
Team													
Australia	47.0	0.0	22.0	0.0	0.0	0.0	0.0	59.0	0.0	49.0	143.0	0.0	3.0
Canada	19.0	0.0	2.0	0.0	0.0	0.0	0.0	0.0	68.0	47.0	61.0	0.0	2.0
France	33.0	0.0	27.0	111.0	16.0	2.0	0.0	0.0	0.0	19.0	12.0	0.0	2.0
Great Britain	112.0	0.0	21.0	9.0	0.0	0.0	0.0	44.0	0.0	40.0	39.0	0.0	3.0
Italy	29.0	12.0	67.0	134.0	0.0	11.0	0.0	0.0	0.0	36.0	4.0	34.0	5.0
Japan	2.0	0.0	1.0	0.0	16.0	128.0	0.0	0.0	0.0	0.0	33.0	0.0	12.0
Norway	4.0	0.0	2.0	0.0	0.0	0.0	14.0	0.0	0.0	14.0	0.0	0.0	1.0
Sweden	12.0	0.0	14.0	13.0	12.0	17.0	0.0	0.0	96.0	5.0	23.0	0.0	1.0
United States	337.0	146.0	17.0	2.0	0.0	28.0	0.0	16.0	85.0	140.0	462.0	37.0	25.0
West Germany	71.0	0.0	28.0	48.0	18.0	1.0	15.0	66.0	18.0	51.0	59.0	12.0	7.0



```
soviet_top10 = sqldf("SELECT * FROM soviet WHERE YEAR < '1992';")
soviet_top10 = sqldf("SELECT * FROM soviet_top10 WHERE SPORT in ( 'Athletics', 'Gymnastics', 'Swimming', 'Basketball', 'Rowing',
soviet_top10[["Team", "Medal", "Sport"]] = soviet_top10[["Team", "Medal", "Sport"]].astype(str)
sov_top10_pivot = pd.pivot_table(soviet_top10, values='Medal', index=['Team'],
                    columns=['Sport'], aggfunc=lambda x: len(x))
sov_top10_pivot = sov_top10_pivot.fillna(0)
sov_top10_pivot.head(11)
```

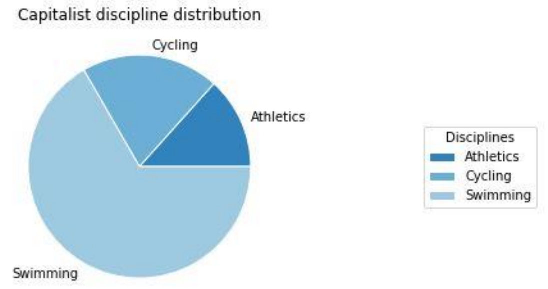
Sport	Athletics	Basketball	Cycling	Fencing	Football	Gymnastics	Handball	Hockey	Ice Hockey	Rowing	Swimming	Water Polo	Weightlifting
Team													
Bulgaria	14.0	24.0	0.0	0.0	32.0	5.0	0.0	0.0	0.0	30.0	3.0	0.0	24.0
China	2.0	12.0	0.0	1.0	0.0	23.0	13.0	0.0	0.0	14.0	4.0	0.0	11.0
Cuba	21.0	12.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.0
Czechoslovakia	19.0	0.0	13.0	0.0	33.0	37.0	16.0	16.0	92.0	44.0	0.0	0.0	3.0
East Germany	140.0	0.0	33.0	1.0	51.0	86.0	42.0	0.0	0.0	184.0	152.0	0.0	11.0
Hungary	25.0	0.0	0.0	134.0	79.0	57.0	14.0	0.0	0.0	4.0	26.0	94.0	17.0
Poland	49.0	0.0	17.0	52.0	34.0	7.0	14.0	0.0	0.0	9.0	2.0	0.0	22.0
Romania	24.0	0.0	0.0	27.0	0.0	65.0	58.0	0.0	0.0	85.0	3.0	0.0	9.0
Soviet Union	242.0	146.0	53.0	146.0	87.0	288.0	84.0	32.0	168.0	155.0	124.0	78.0	62.0
Yugoslavia	1.0	84.0	0.0	0.0	58.0	3.0	72.0	0.0	0.0	13.0	2.0	76.0	0.0

All of the countries in question have several sports excellence to thank for their high performance. Upon filtering the country via the year of the games, we get the following results.



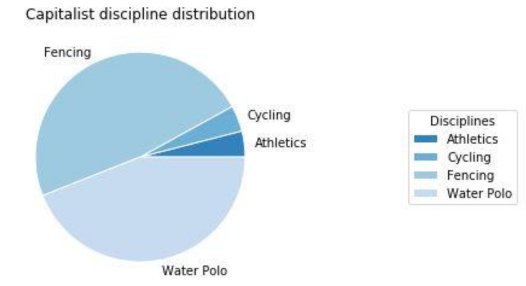
```
def capital_by_sport(year:int, team:str):
    fill = capital_top10.loc[capital_top10['Year'] == year]
    filt = fill.loc[fill['Team'] == team]
    table2 = sqldf('SELECT YEAR, SPORT, COUNT(SPORT) AS Medal_Count FROM filt GROUP BY SPORT;')

    fig, ax = plt.subplots()

    cmap = plt.get_cmap("tab20c")
    outer_colors = cmap(np.arange(10))

    ax.pie(table2['Medal_Count'], colors=outer_colors, labels=table2['Sport'],
           wedgeprops=dict(edgecolor='w'), autopct="")

    ax.legend(table2['Sport'],
          title="Disciplines",
          loc="center left",
          bbox_to_anchor=(1.5, 0, 0.5, 1))


    ax.set(aspect="equal", title='Capitalist discipline distribution')
    plt.show()
    return table2
```

```
capital_by_sport(1972, 'Australia')
```



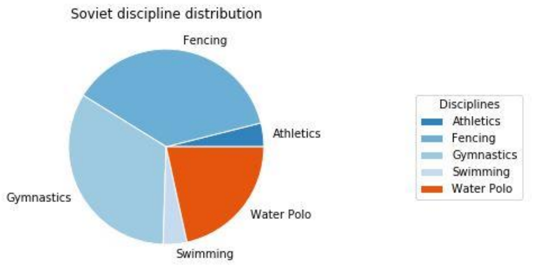
	Year	Sport	Medal_Count
0	1972	Athletics	2
1	1972	Cycling	3
2	1972	Swimming	10

```
capital_by_sport(1976, 'Italy')
```



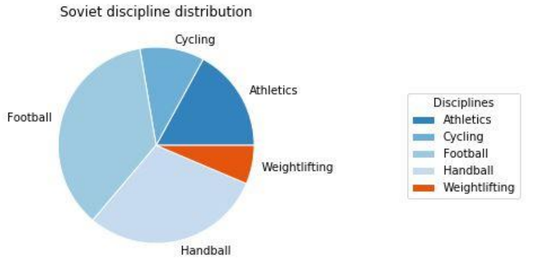
	Year	Sport	Medal_Count
0	1976	Athletics	1
1	1976	Cycling	1
2	1976	Fencing	12
3	1976	Water Polo	11

```
soviet_by_sport(1956, 'Hungary')
```



	Year	Sport	Medal_Count
0	1956	Athletics	2
1	1956	Fencing	19
2	1956	Gymnastics	17
3	1956	Swimming	2
4	1956	Water Polo	11

```
soviet_by_sport(1976, 'Poland')
```



	Year	Sport	Medal_Count
0	1976	Athletics	8
1	1976	Cycling	5
2	1976	Football	17
3	1976	Handball	14
4	1976	Weightlifting	3

These graphs indicate, that the soviet countries favor more team sports, where they went on to be a formidable force all throughout the measured period. Individual disciplines were more adherent to what sporting school was strong for the individual country (*skiing in the USSR). It is also visible, that the results improved after the soviet reform of the sporting system in the country, as well as finding the federations from scratch (*Boxing in Cuba, which was not popularized until the revolution)

Capitalist countries, on the other hand, were better at individual disciplines – athletics, swimming, cycling. As later findings show, the sport most popularized or widespread in the country ends up taking charge in the medal hopes – like athletics for the USA, cycling for the UK or hockey for the UK and Australia.

Step 3. Run statistical analysis via Pearson and P-value to find correlations

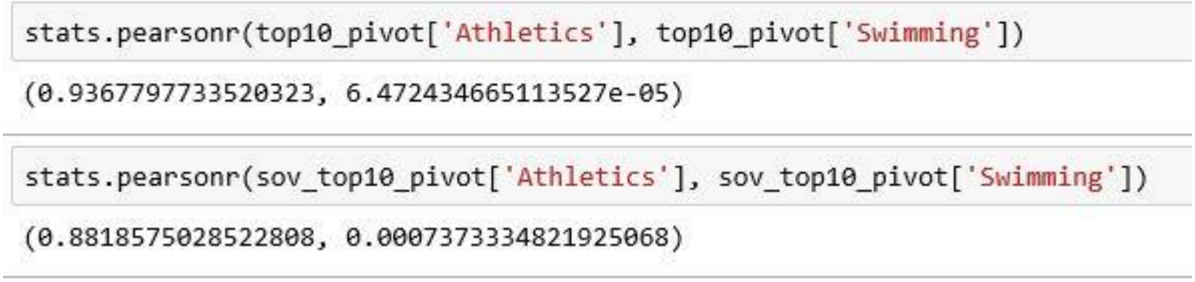
***

Pearson correlation (the bivariate correlation) is a statistic that measures linear correlation between two variables X and Y. It has a value between +1 and −1, where 1 is total positive linear correlation, 0 is no linear correlation, and −1 is total negative linear correlation

P-value is the probability that you would have found the current result if the correlation coefficient were in fact zero (null hypothesis). If this probability is lower than the conventional 5% (P<0.05) the correlation coefficient is called statistically significant.

***

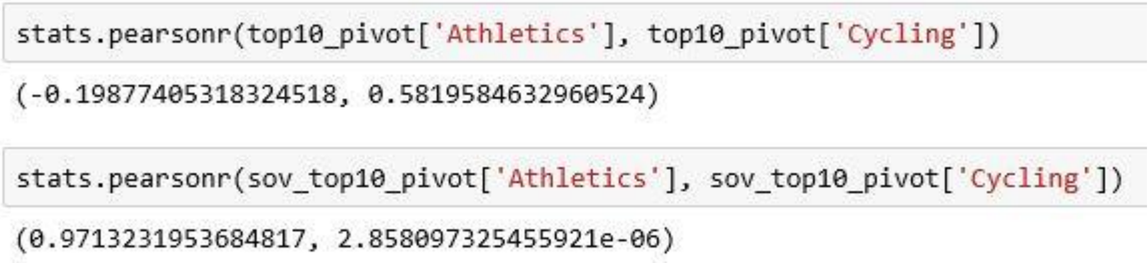
We are going to find correlations between the performance of sports in each table. Since the biggest contributors to the medal pool are athletics and swimming, we will compare them first and then see, how well the other disciplines perform.



```
stats.pearsonr(top10_pivot['Athletics'], top10_pivot['Swimming'])

(0.9367797733520323, 6.472434665113527e-05)
```

```
stats.pearsonr(sov_top10_pivot['Athletics'], sov_top10_pivot['Swimming'])

(0.8818575028522808, 0.0007373334821925068)
```

Looks like the correlation is quite strong for both tables. This comes as no surprise, since we are looking at the most successful teams across the games. Let`s look at the correlation between athletics and cycling.



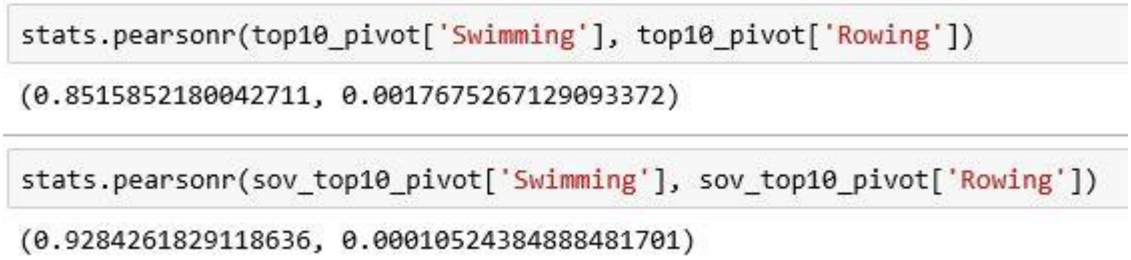
```
stats.pearsonr(top10_pivot['Athletics'], top10_pivot['Cycling'])
```

```
(-0.19877405318324518, 0.5819584632960524)
```

```
stats.pearsonr(sov_top10_pivot['Athletics'], sov_top10_pivot['Cycling'])
```

```
(0.9713231953684817, 2.858097325455921e-06)
```

This shows even stronger correlation, but our capitalist table gave a weak Pearson`s with a very high P-value. Hence, the correct result will be a non-linear correlation between these two. Situation is with the water sports bears close resemblance to the athletics-swimming pair: good performers in one usually perform well in the other.

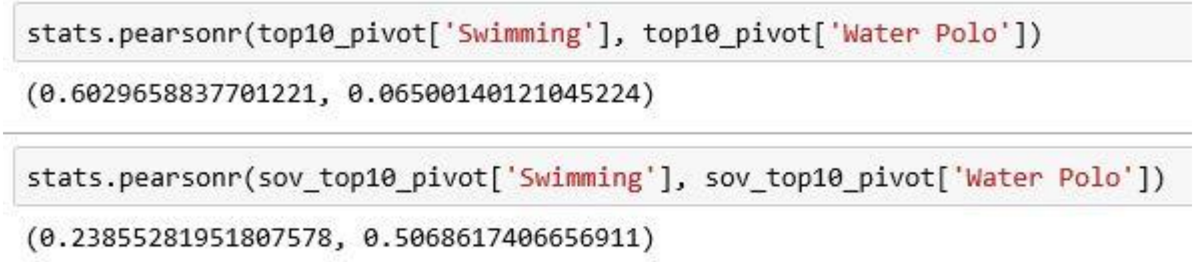


```
stats.pearsonr(top10_pivot['Swimming'], top10_pivot['Rowing'])
```

```
(0.8515852180042711, 0.0017675267129093372)
```

```
stats.pearsonr(sov_top10_pivot['Swimming'], sov_top10_pivot['Rowing'])
```

```
(0.9284261829118636, 0.00010524384888481701)
```



```
stats.pearsonr(top10_pivot['Swimming'], top10_pivot['Water Polo'])
```

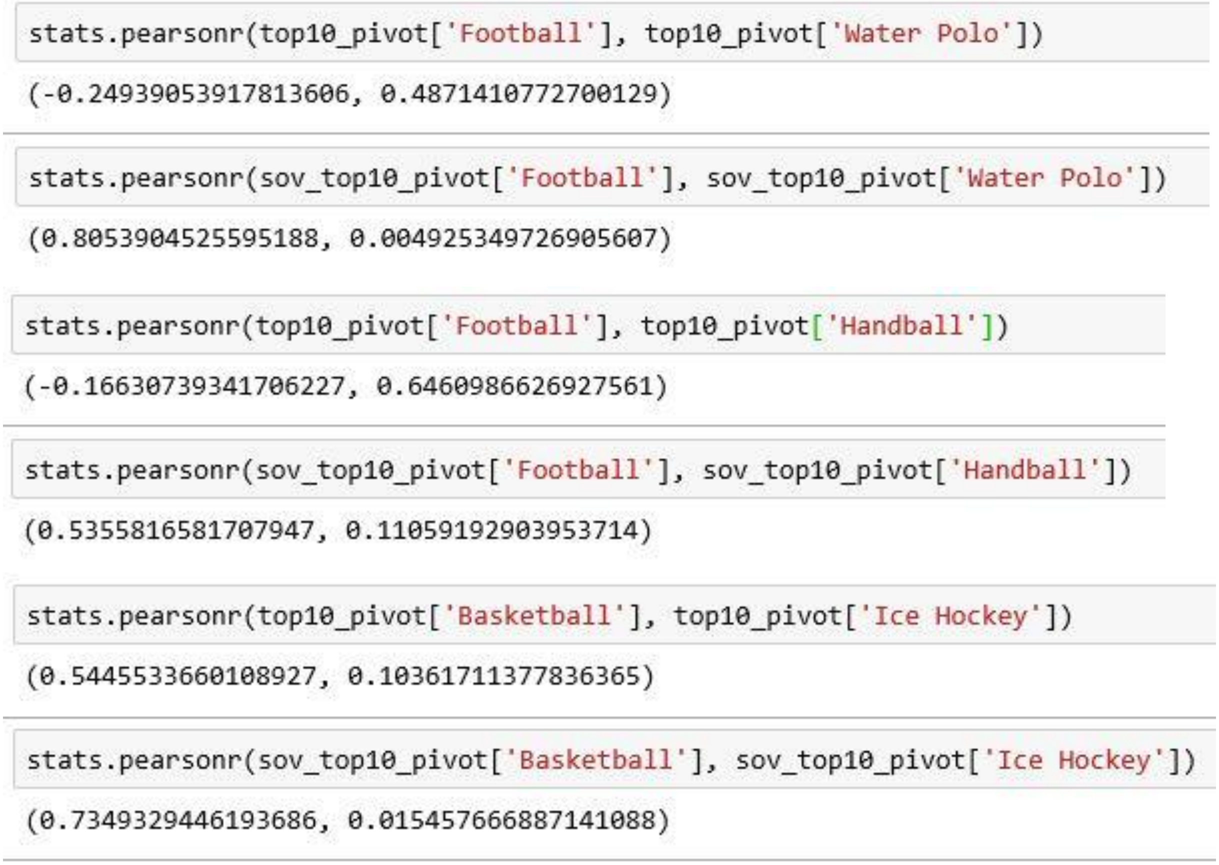
```
(0.6029658837701221, 0.06500140121045224)
```

```
stats.pearsonr(sov_top10_pivot['Swimming'], sov_top10_pivot['Water Polo'])
```

```
(0.23855281951807578, 0.5068617406656911)
```

However, the capitalist table shows more of a linear correlation, whereas soviet one looks like a non-linear relationship.

But what about team sports? Since the soviet teams should perform “so much better”? Well, the results are less than straightforward:



```
stats.pearsonr(top10_pivot['Football'], top10_pivot['Water Polo'])

(-0.24939053917813606, 0.4871410772700129)

stats.pearsonr(sov_top10_pivot['Football'], sov_top10_pivot['Water Polo'])

(0.8053904525595188, 0.004925349726905607)

stats.pearsonr(top10_pivot['Football'], top10_pivot['Handball'])

(-0.16630739341706227, 0.6460986626927561)

stats.pearsonr(sov_top10_pivot['Football'], sov_top10_pivot['Handball'])

(0.5355816581707947, 0.11059192903953714)

stats.pearsonr(top10_pivot['Basketball'], top10_pivot['Ice Hockey'])

(0.5445533660108927, 0.10361711377836365)

stats.pearsonr(sov_top10_pivot['Basketball'], sov_top10_pivot['Ice Hockey'])

(0.7349329446193686, 0.015457666887141088)
```

The relationships in the capitalist table shows hardly any correlation, whereas the soviet performance is more linear, although not entirely.

Conclusions

1. Adoption of the soviet sport system improves the performance of every country with every consecutive games, even though the biggest benefactor, judging by the results was the USSR

2. Capitalist countries with independent systems of preparation show more consistent results in their strongest disciplines.

3. Eastern European school of gymnastics gave a huge advantage to the soviet countries. From early 1950s, gymnastics popularization became widespread for the "improvement of physical wellbeing of soviet people". Later that decree transformed into a way to earn more medals at the highest level, which resulted in soviet dominance on the Olympic stage. Apart from Japan, no other capitalist country was decisively better performing at the Olympics.

4. A special mention needs to be attributed to fencing. As a competitive sport, it originated in Italy, who long has been the most formidable force in the sport. Much praise can also be said about the Hungarian school of fencing, considered second best in Europe. After the European division, the high standards of Hungarian fencing were taught all over the soviet block, resulting in much improvement of the soviet athletes.

5. The strong schools of each country, joining the soviet block, contributed to the other countries` training methods. Soviet system took the experience of one nation and adopted the practice. Good examples – fencing, gymnastics and water polo

Hypotheses: Verdict

1. **Outside of the USA, all of the state controlled sport system countries will perform better, than the amateur ones.**
   Partially correct – from the 1972 games onward

2. **Each soviet country will have a time lag, in which the new system will be implemented. Therefore, the results of all will improve after 3 Olympic cycles from start of participation.**
   Partially correct – some countries showed downward trend. After the 1968 inclusion of East Germany, all countries showed a steady growth.

Discussion.

As much as sports were a method of political propaganda for the soviet regime, it was a mass phenomenon, that drew from the best practices of the countries it harbored and implemented them across the board. The same approach was taken by China in preparation for 2008 games – and it propelled them to become one of the major forces at the games. There is much room in international cooperation, when it comes to best training methods to make the games more competitive and the pool of victors wider and more inclusive of other nations.

Thank you for reading, you can find the code on my GitHub, any feedback, comments, suggestions for this would be much appreciated!