Olympic Games

Evidence and Facts

Milestone 4: Presenting Your Findings (Storytelling)!

Objective

Learn about Olympic Games results (medals).
Find evidences about the country, gender, sex and years to understand the Olympic Games results.

Audience

Journalists, coaches, countries (government) who are looking to understand the Olympic Games results. Journalists in order to publish the information. Coaches and government to improve the results of their team / country.

Questions

What has been the demographics of Olympics – what age group, gender(sex) and countries have more participants and better results?

Which country has been more regular in Olympics, who has been winning more medals? Which sports in which season Olympics is popular amongst the top countries?
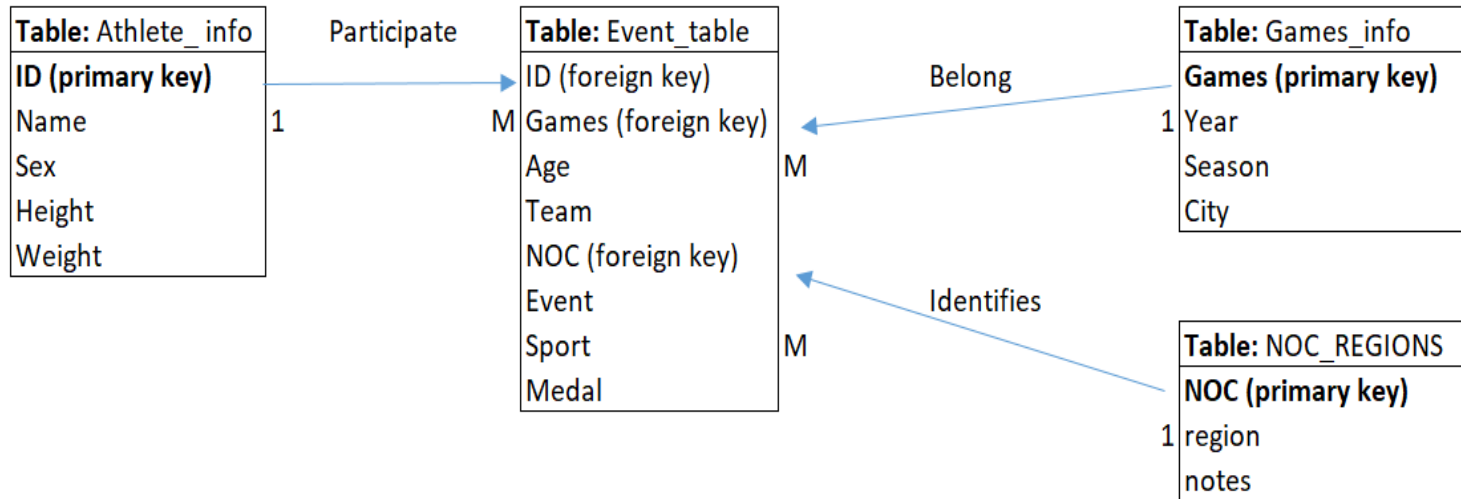
What effect does the host country have in the medals won at the Olympics?

Entity Relationship Diagram

Data Base

Chen Notation

| Table: Athlete_info		Participate	Table: Event_table		Belong	Table: Games_info
ID (primary key)			ID (foreign key)			**Games (primary key)**
Name	1		Games (foreign key)	M		1 Year
Sex			Age	M		Season
Height			Team			City
Weight			NOC (foreign key)		Identifies	
			Event			Table: NOC_REGIONS
			Sport	M		**NOC (primary key)**
			Medal			1 region
						notes

120 years of Olympic history: Athletes and Results

Basic bio data on athletes and medal results from Athens 1896 to Rio 2016

4

Demographics of Olympics

What has been the demographics of Olympics – what age group, gender (sex) and countries have more participants and better results?

Olympic Athletes by Age
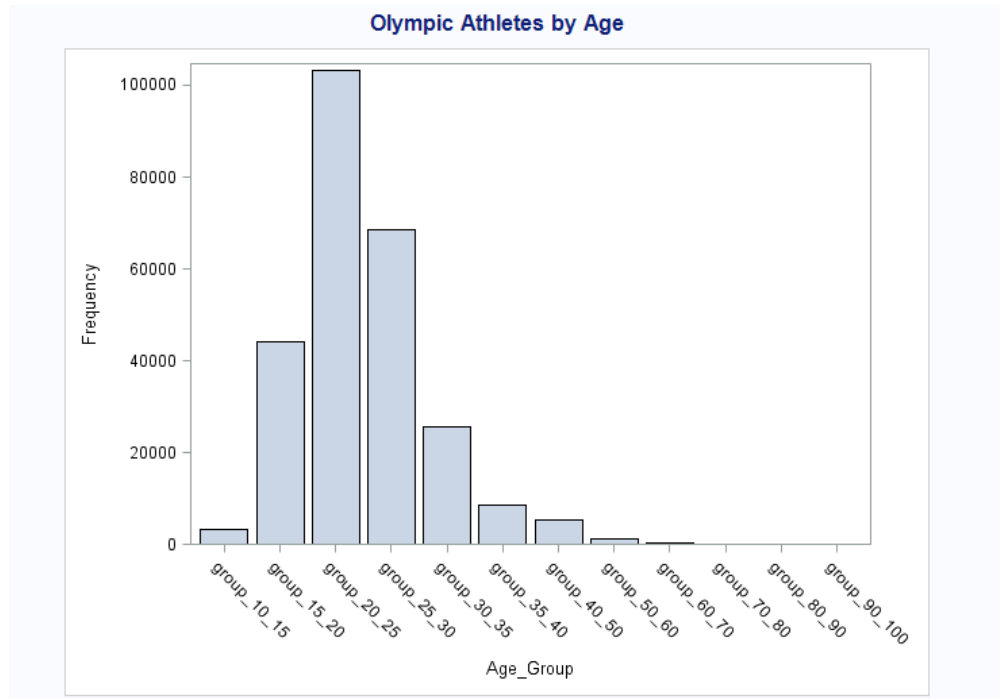
The age group between 20 and 25 is the most represented.

- Average: 25,5 years old
- Minimum: 10 years old
    Name: Dimitrios Loundras (Greece)
    Sport: Gymnastics
    Games: 1896 Summer
- Maximum: 97 years old
    Name: John Quincy Adams Ward (USA)
    Sport: Art Competitions
   Games: 1928 Summer

Demographics of Olympics

Average Age by Season Over Time

- The age of the participants decreased until 1980 (Summer Games) and 1984 (Winter Games) and then increased.
- Despite the increase registered in the last Olympics (since 1980) , the average age in the last years has never exceeded 26,5 years.
- Using the T-test, we concluded that the difference in the average age between Winter (25,1) and Summer (26,1) games is statistically significant (p-value<0,05).

Demographics of Olympics

Medals by Age Group
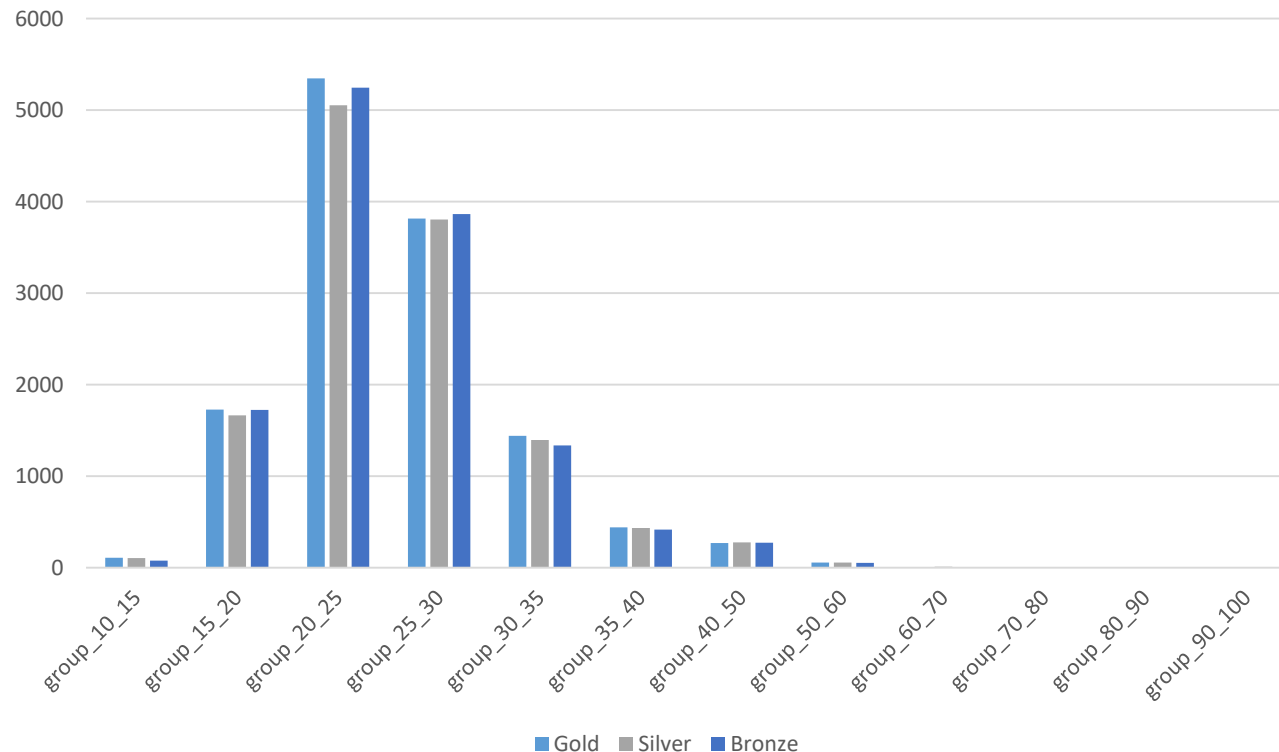
- The age group between 20 and 25 won more medals (5.348 gold, 5.055 silver and 5.247 bronze).

- The second age group with more medals is between 25 and 30 (3.817 gold, 3.806 silver and 3.865 bronze).

- Despite the difference in medals won, the probability of an athlete winning a medal participating in the Olympic Games between the ages of 20 and 50 is about 15% (lower for other ages).

Demographics of Olympics

Age Group by Medal - methodology:

Cross-tabulation using different age groups (group_10_15, group_15_20, group_20_25, group_25_30, group_30_35, group_35_40, group_40_50, group_50_60, group_60_70, group_70_80) and medals (gold, silver, bronze, NA).

The Pearson's chi-squared test was used to determine whether there is a statistically significant difference between the expected and observed frequencies.

Conclusion:

Age (specially the age group between 20 and 25) has an impact in the number of medals win by the athlete. The same test was performed separating male and female athletes and the result are the same. Reject the null hypothesis (p-value <0.01).

Statistics for Table of Age_Group by Medal

Statistic	DF	Value	Prob
Chi-Square	33	729.5813	<.0001
Likelihood Ratio Chi-Square	33	773.4427	<.0001
Mantel-Haenszel Chi-Square	1	39.9437	<.0001
Phi Coefficient		0.0529	
Contingency Coefficient		0.0529	
Cramer's V		0.0306	

Demographics of Olympics
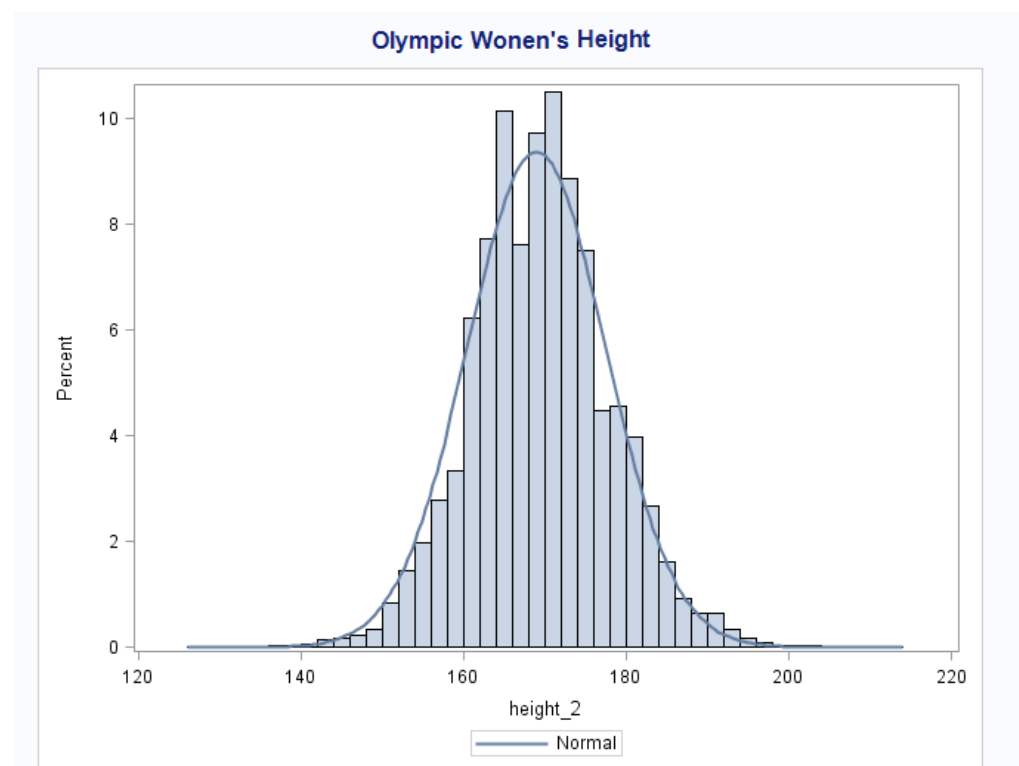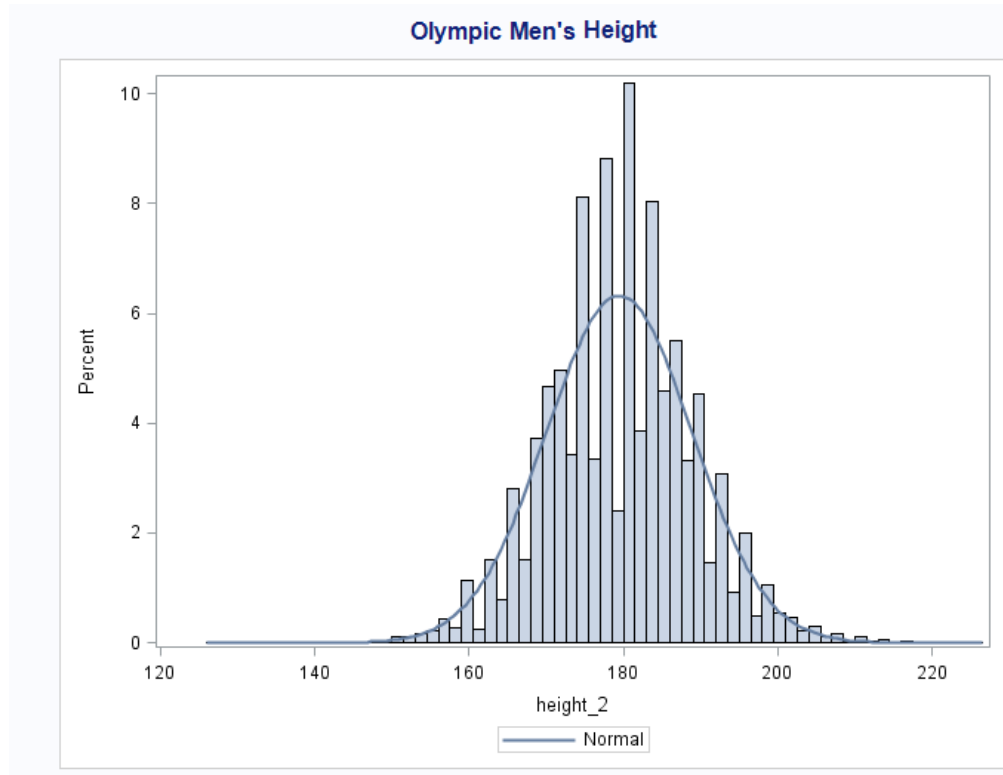
Male (cm) - Average: 179, Minimum: 127, Maximum 226, Standard Deviation 9,5.
Female (cm) - Average: 169, Minimum: 127, Maximum 213, Standard Deviation 8,5.
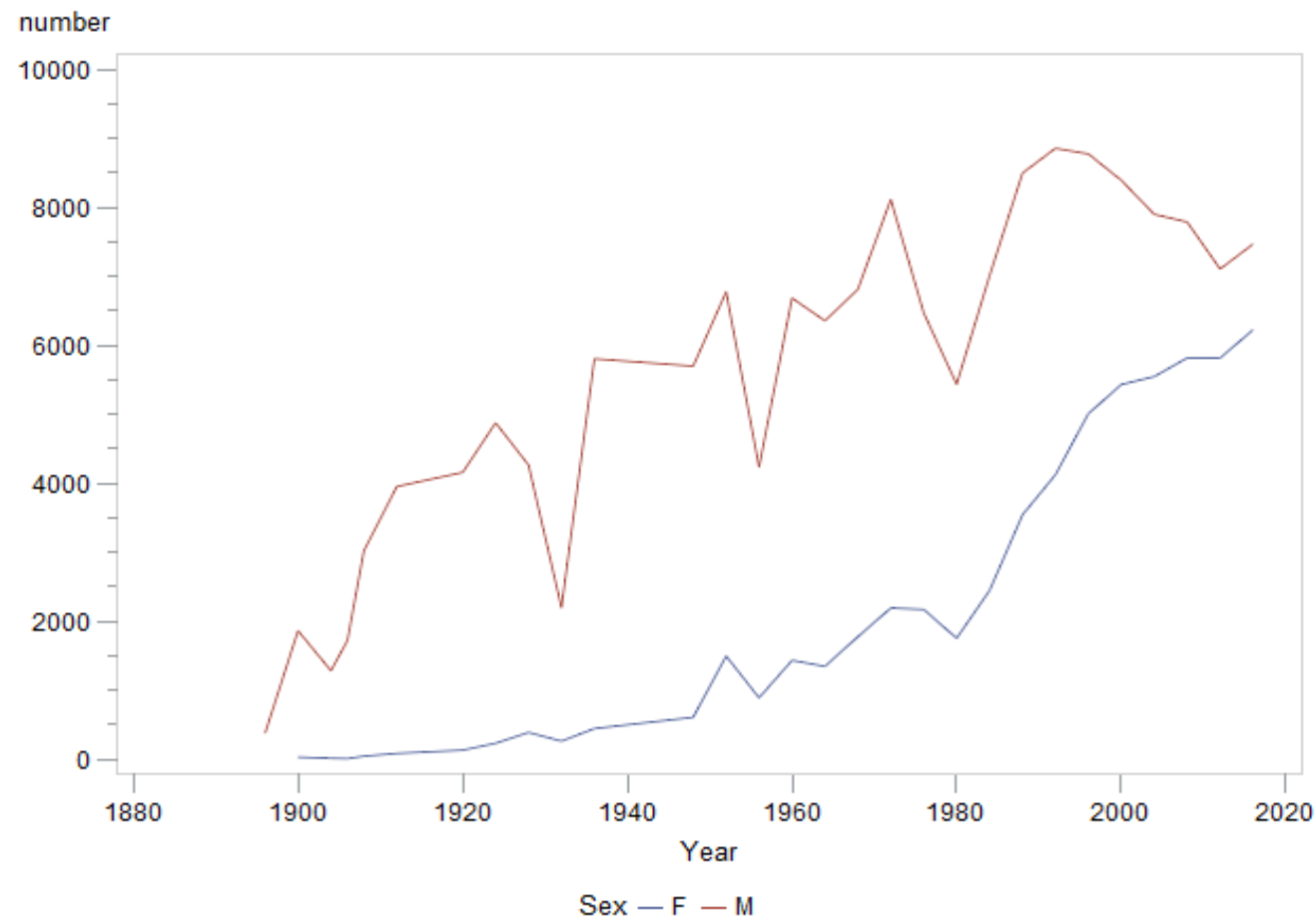
Demographics of Olympics
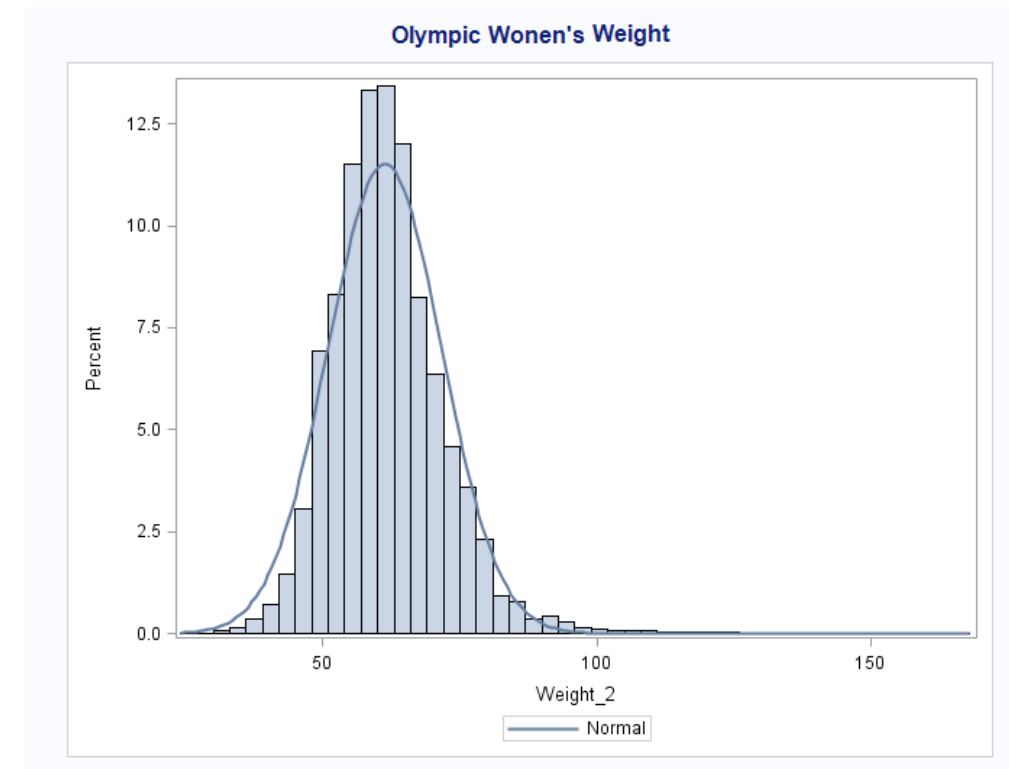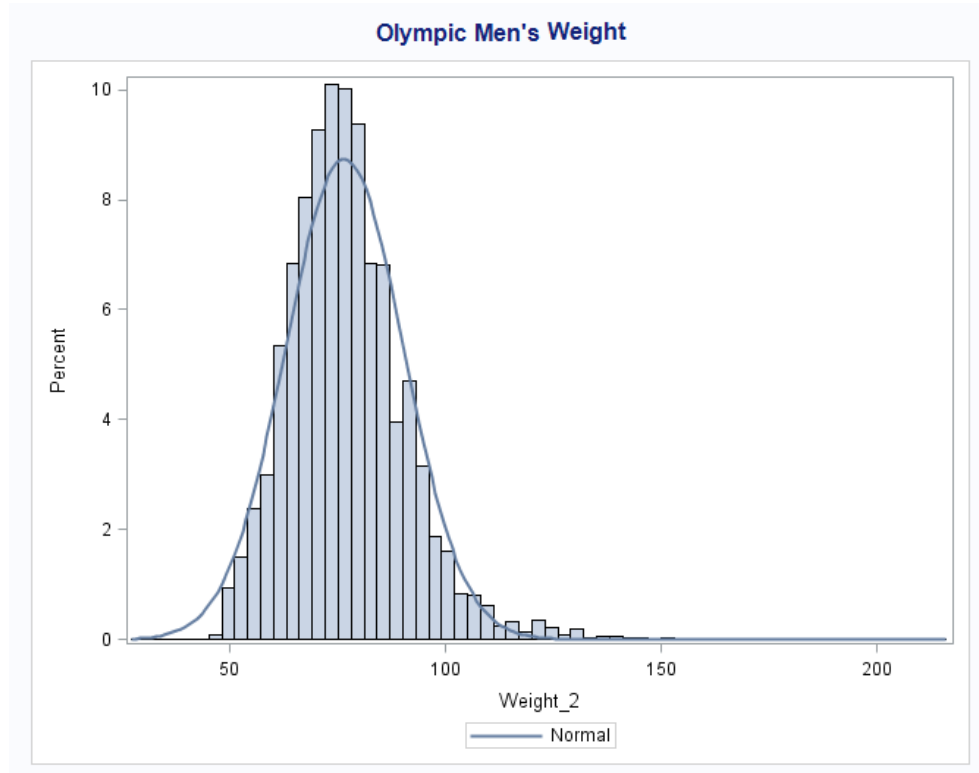
Male (kg) - Average: 76, Minimum: 28, Maximum: 214, Standard Deviation: 13,7.
Female (kg) - Average: 61; Minimum: 25, Maximum: 167, Standard Deviation: 10,4.
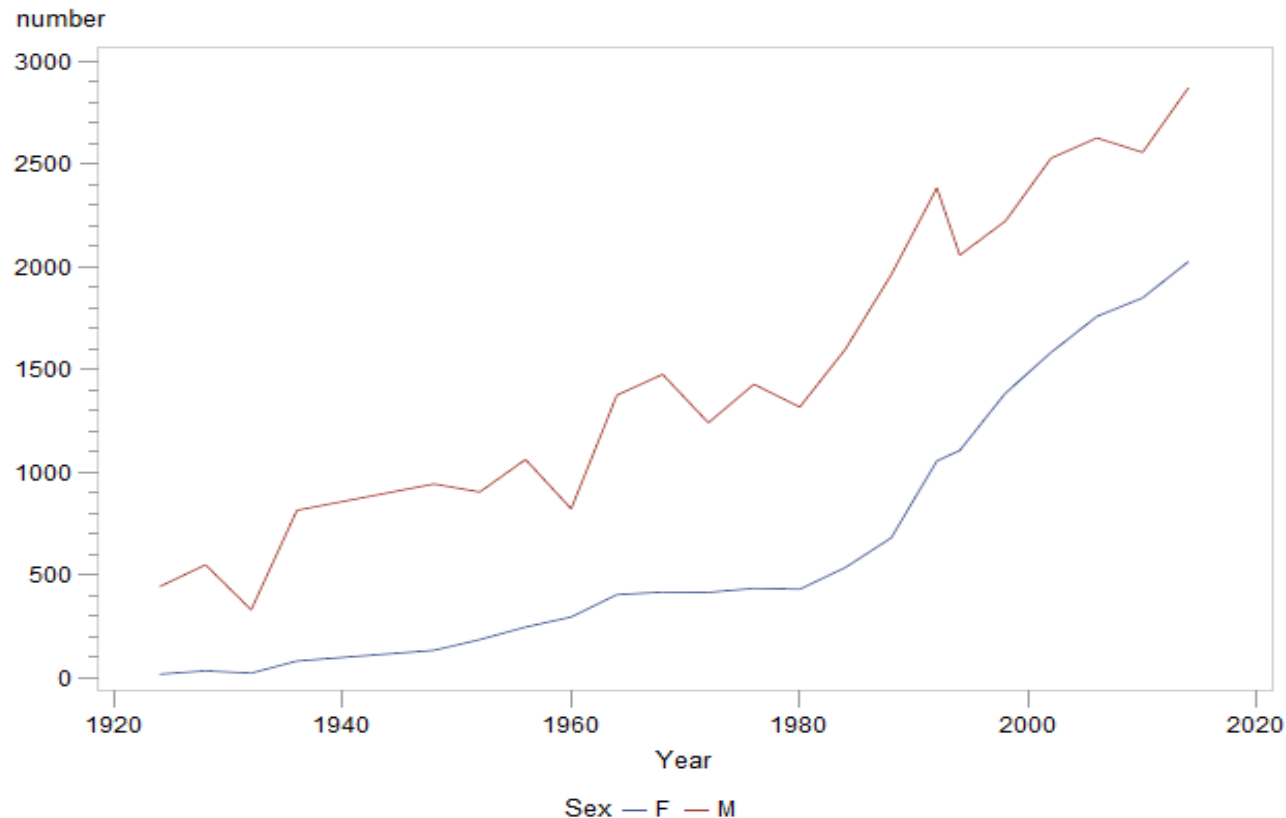
Demographics of Olympics

Summer Games by Gender Over Time

number

- The number of participants in Summer Olympics has increased over time.
- The gap between male and female participation in Summer Games is getting smaller in recent years.
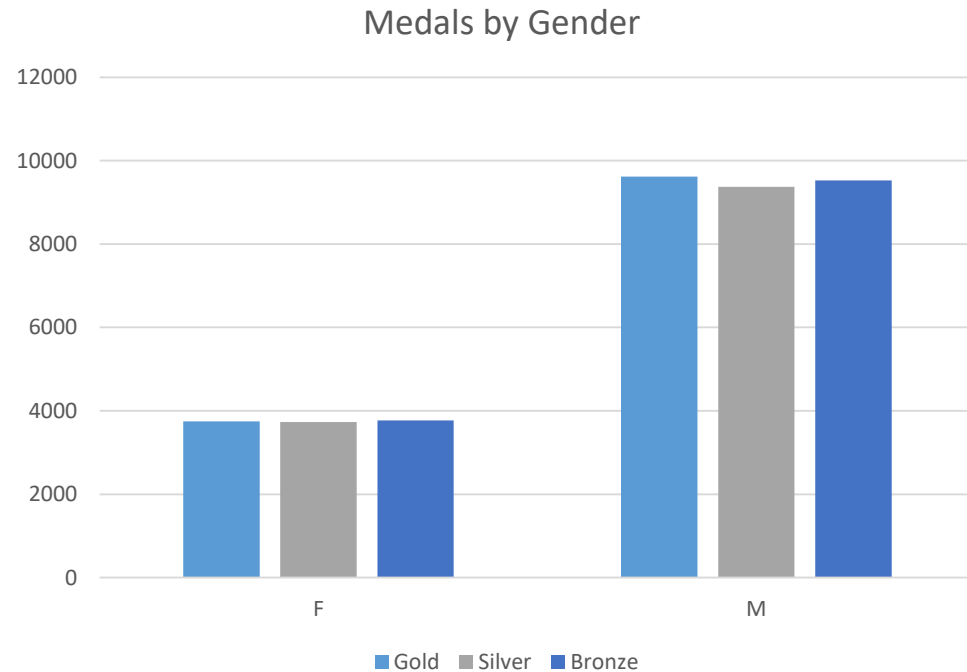
11

Demographics of Olympics

Winter Games by Gender Over Time

- The number of participants in Winter Olympics has increased over time.

- Contrary to what was registered in the Summer Games, the gap between male and female participation tend to remain constant over the years.

Demographics of Olympics

Medals by Gender

- Men participate more in Olympic Games, so as expected they get more medals.

- However, the probability of a woman who participate in the games win a medal is higher (15,1% for women and 14,6% for men).

Demographics of Olympics

Participants and Medals by country
(top 10 by medals)
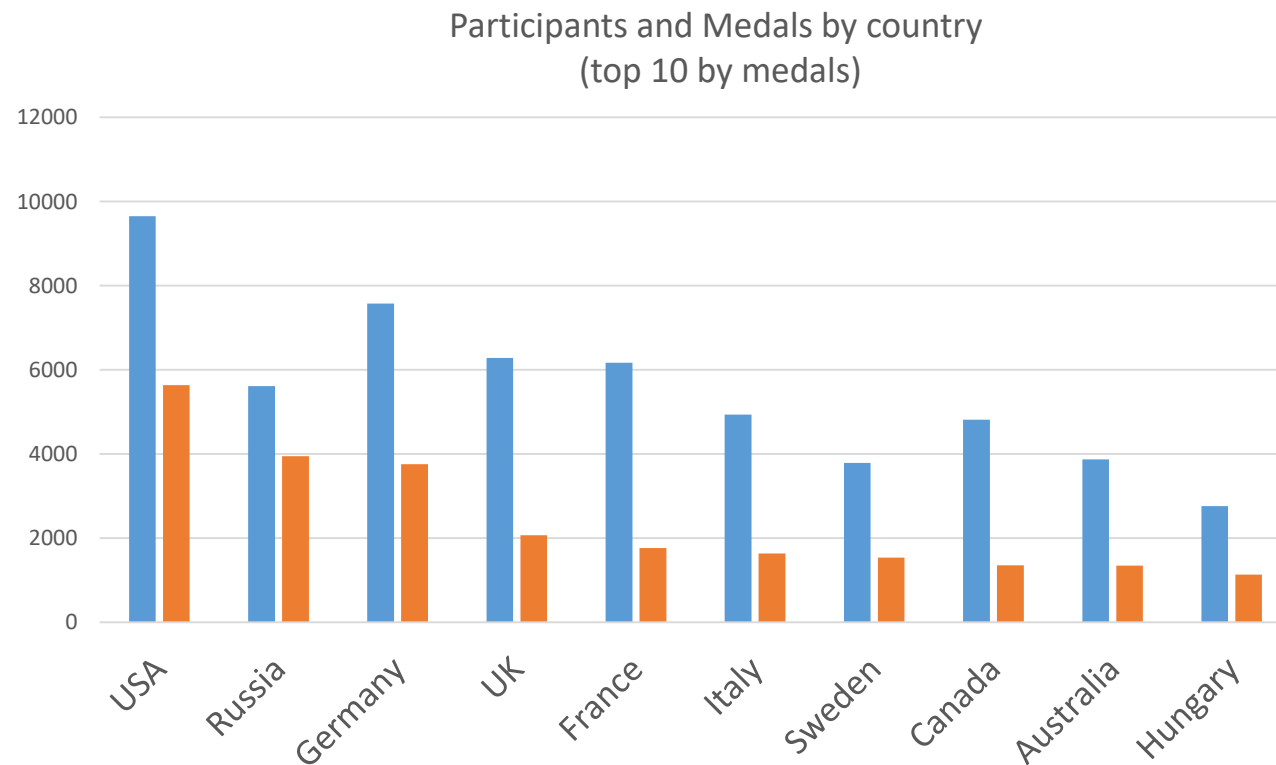
- USA, Russia and Germany are the 3 countries with more medals.

- In terms of ratio [number the medals / number of participants] the winner is Russia (3.947 medals in only 5.610 participants) – 70%.

- Among this 10 countries, France is the worst (1.767 medals in 6.170 participants) – 29%

14

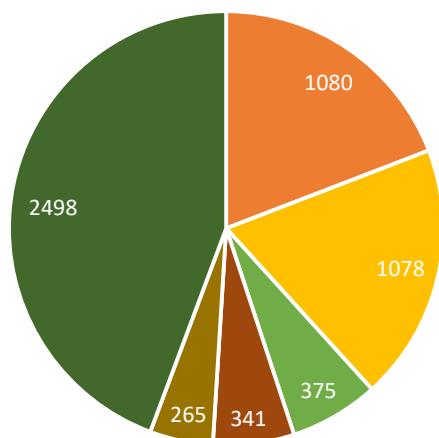# Regularities in the Olympics

Which country has been most regular in Olympics, who has been winning more medals? Which sports in which season Olympics is popular amongst the top countries?
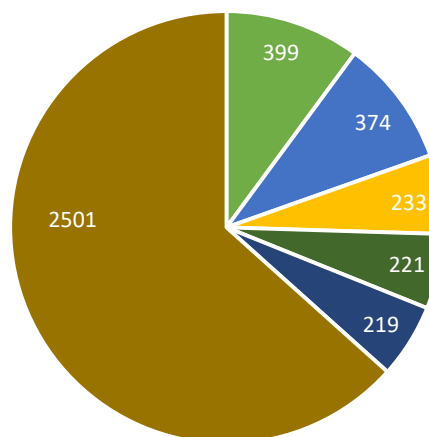
Times a Country Won the Olympics
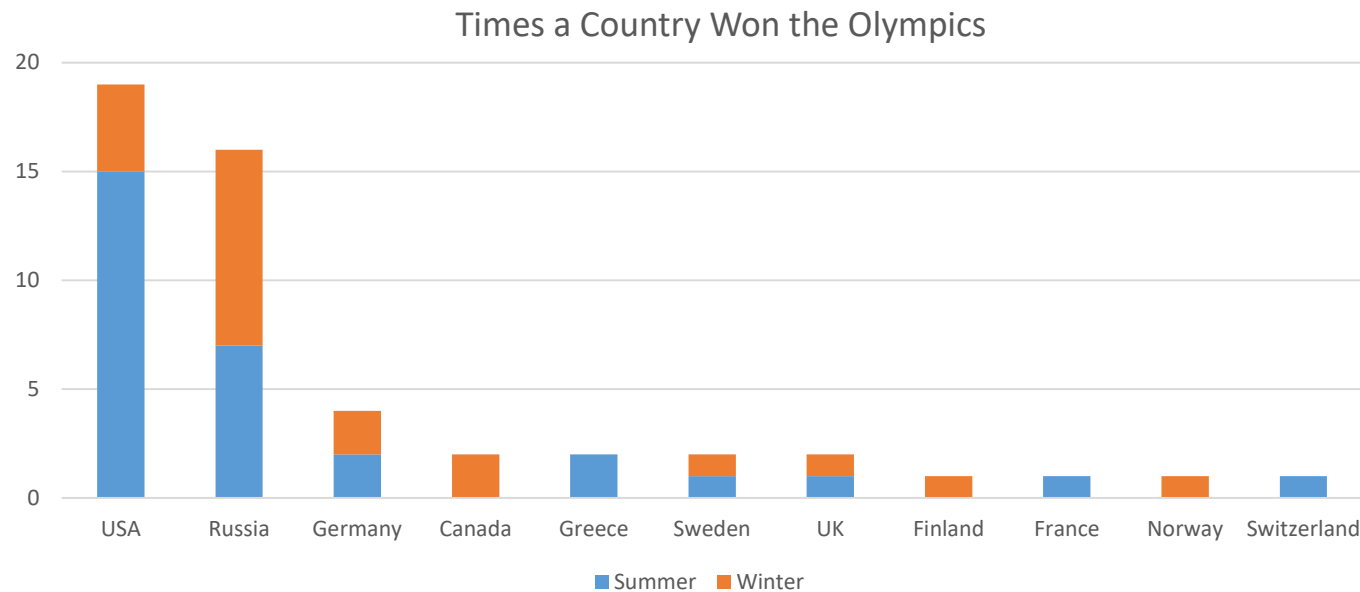
- USA (19) and Russia (16) are the countries that more times won the Olympics.

- USA (15) won more times Summer Olympics and Russia (9) won more times the Winter ones.

Regularities in the Olympics

USA Medals - Top 5 Sports + Others

Legend: Athletics, Swimming, Rowing, Basketball, Ice Hockey, Others

Values: 1080, 1078, 375, 341, 265, 2498

Russia Medals - Top 5 Sports + Others

Legend: Gymnastics, Athletics, Ice Hockey, Fencing, Volleyball, Others

Values: 399, 374, 233, 221, 219, 2501

Germany Medals- Top 5 Sports + Others

Legend: Rowing, Swimming, Athletics, Hockey, Canoeing, Others

Values: 471, 367, 365, 248, 229, 2076

- The sports where the 3 countries win more medals are different: USA athletics, Russia gymnastics and Germany Rowing (all Summer sports). Athletics is the only sport that is in top 5 in both countries.
- In terms of Winter sports, the USA and Russia have Ice Hokey in the top 5 and Germany does not have any sport.

Host Effect in Olympics

What effect does the host country have in the medals won at the Olympics?
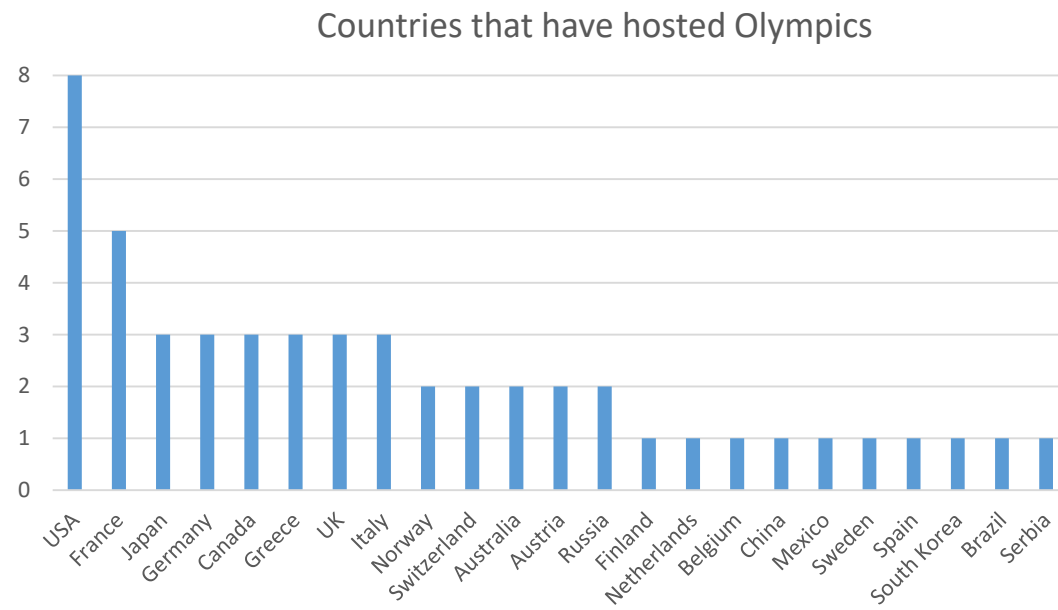
Countries that have hosted Olympics

The reasons why the host countries win more medals are due to home field advantage, crowd support and easier qualification criteria, hence more participation.

- The USA has hosted it the most, i.e. 8 times followed by France, which had hosted it 5 times.

17

Host Effect in Olympics
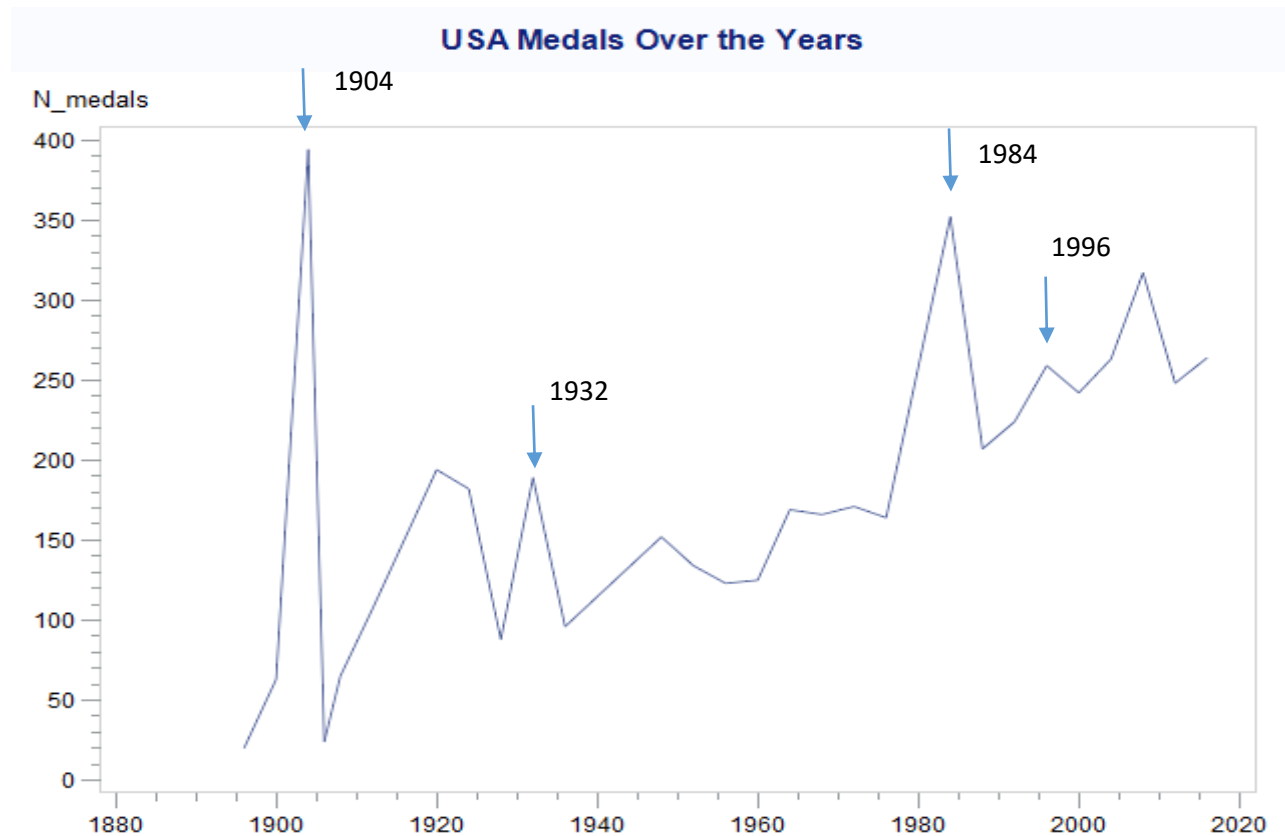
USA Summer Games Host Effect

- There seems to be a host effect in the USA. The country organized 4 Summer Olympics (1904, 1932, 1984 and 1996).

- In these years there has been an increase in the number of medals won by the country.

Host Effect in Summer Olympics

USA Medals Over the Years
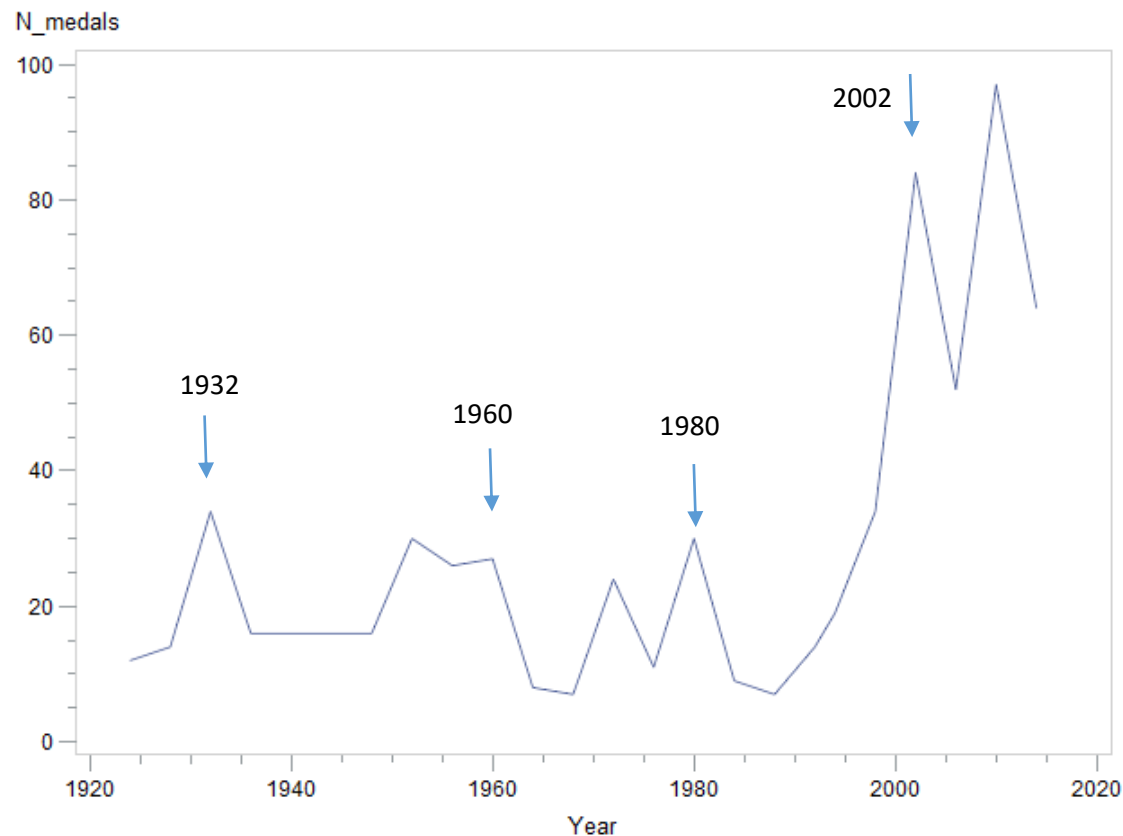
USA Winter Games Host Effect

- As seen before there seems to be a host effect in the USA. The country organized 4 Winter Olympics (1932, 1960, 1980 and 2002).

- In these years there has been an increase in the number of medals won by the country.

Host Effect in Summer Olympics

Before, Hosted and After by Country

Before, Hosted and After by Country

Host Effect in Summer Olympics

Before, Hosted and After by Country
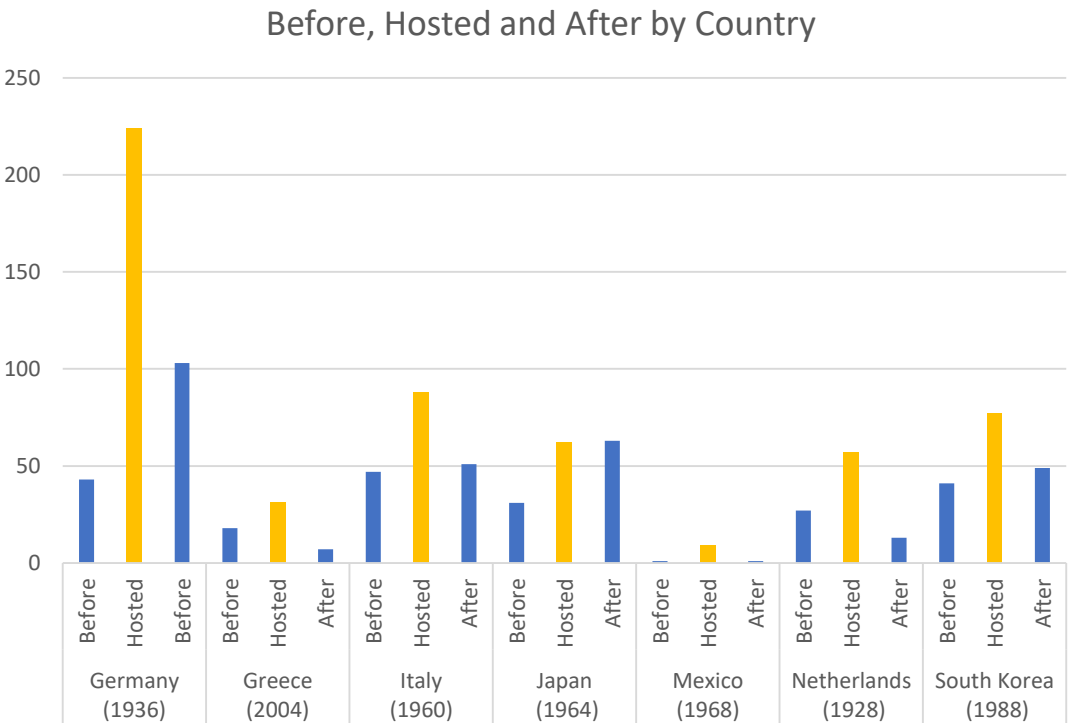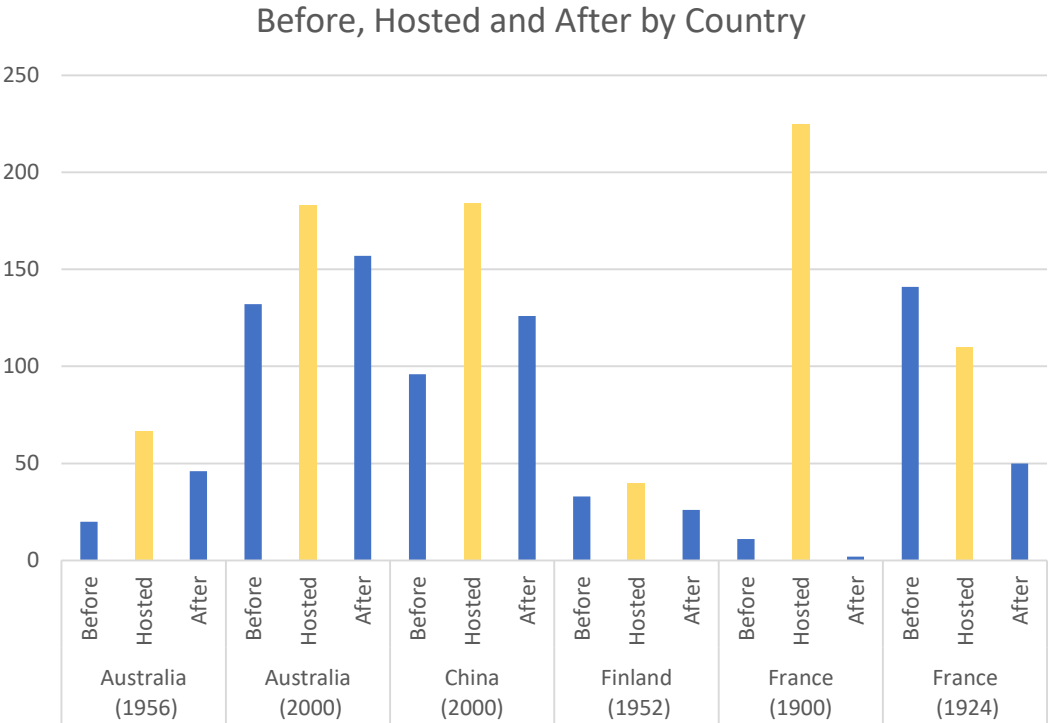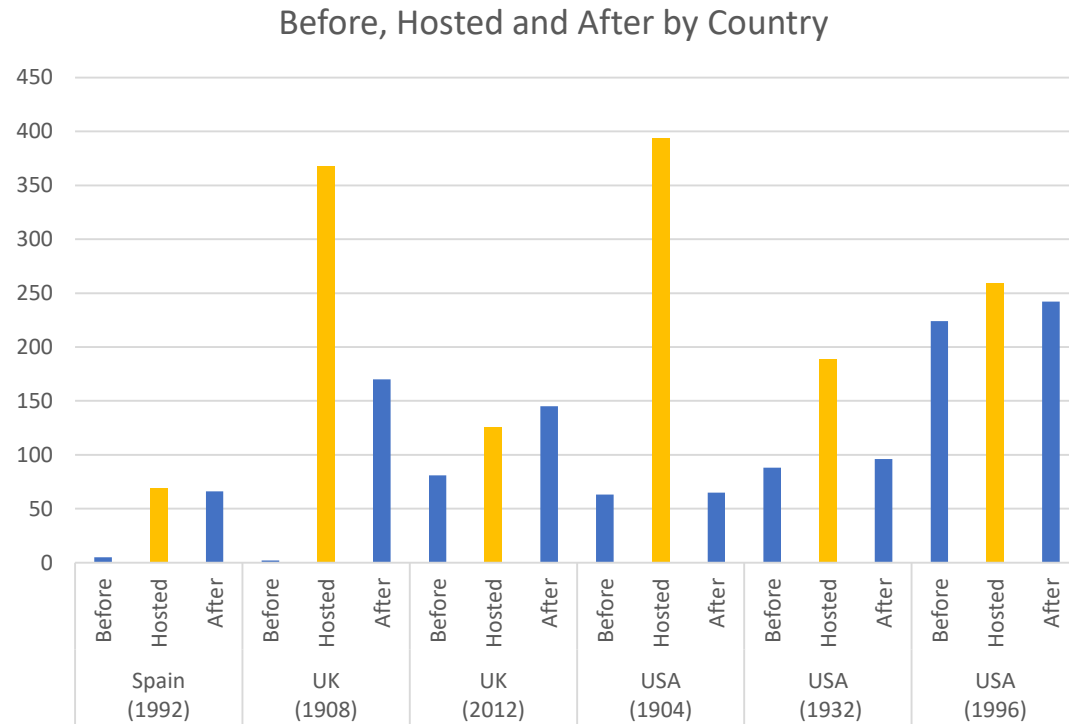
Summer Games Host Effect

- Only 3 times (France, 1924; Japan, 1964 and UK, 2012) the Host Effect has not been fulfilled.

- In the other 16 Summer Olympics we see the Host Effect.

Host Effect in Winter Olympics
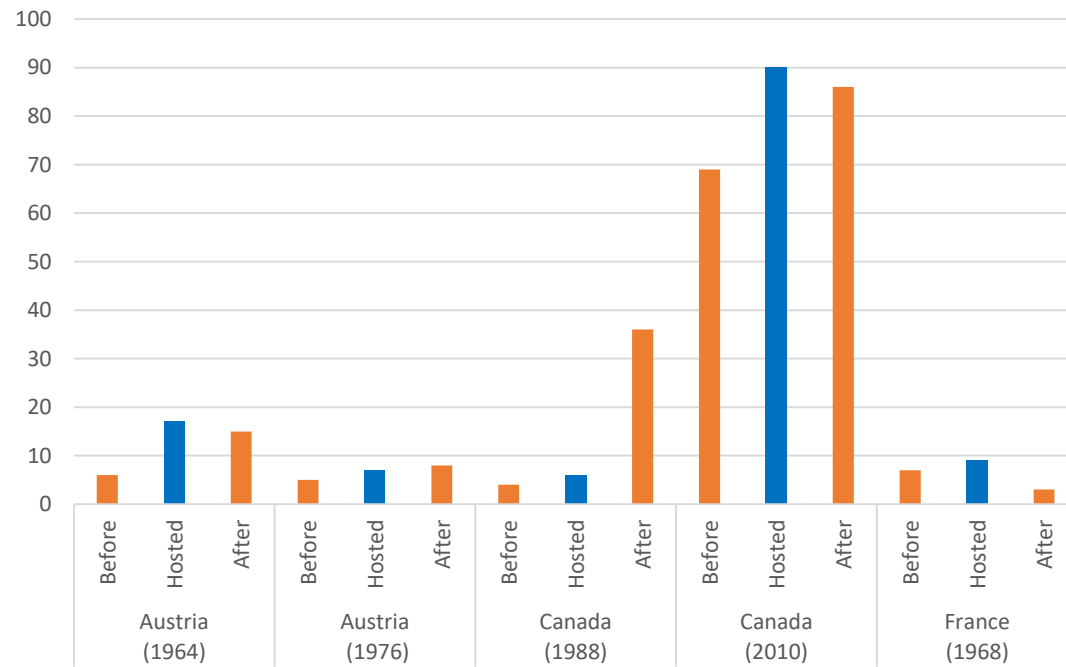
Before, Hosted and After by Country

Before, Hosted and After by Country

Host Effect in Winter Olympics

Before, Hosted and After by Country
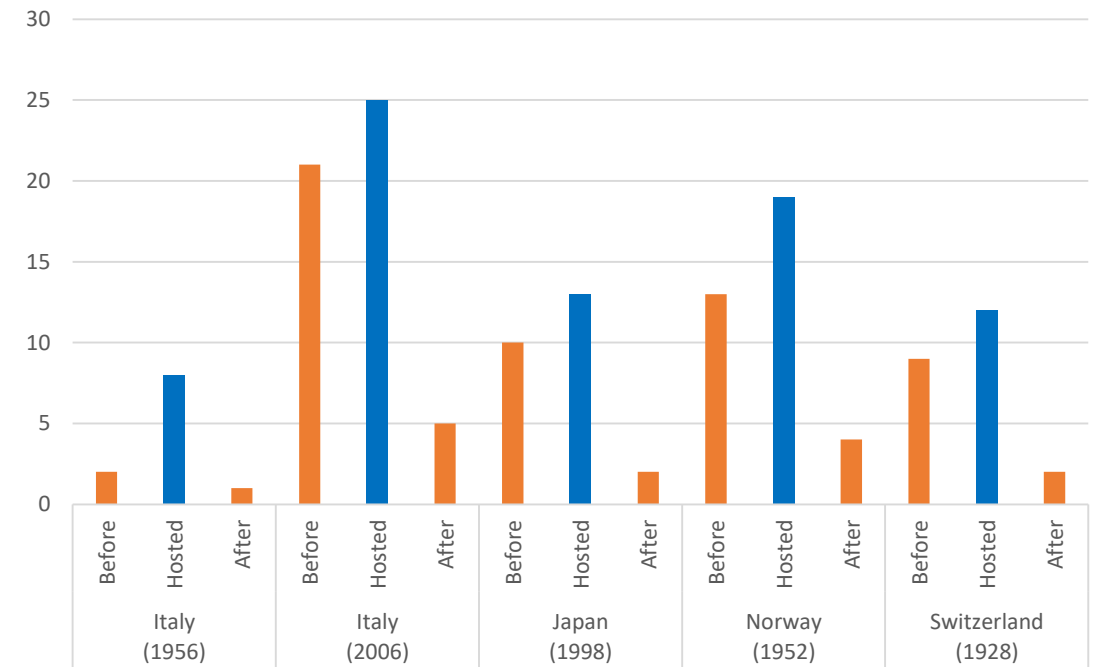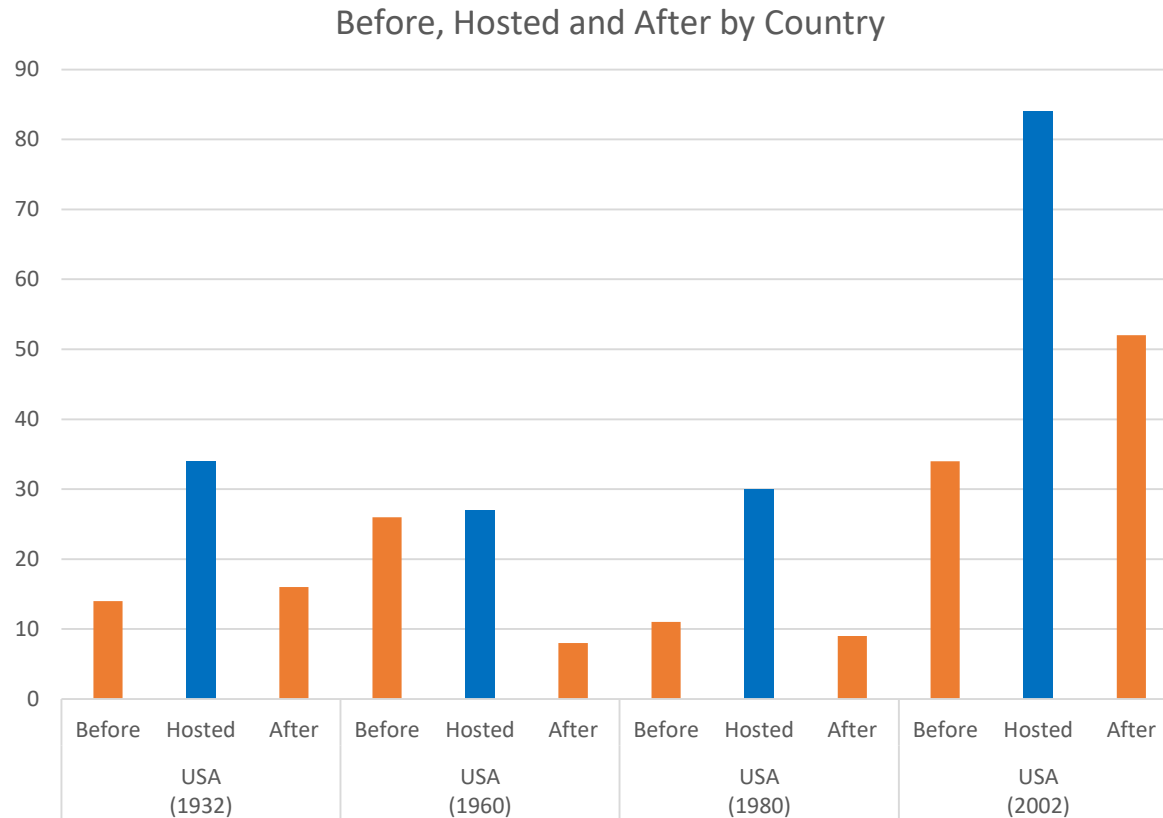
Winter Games Host Effect

- In the 14 Winter Olympics analyzed we see the Host Effect.

Host Effect - Liner Regression

Hypothesis 1: There is a host effect - a country that host the Olympics wins more medals (explanations: more support for the athletes, athletes invest more in preparation).

Hypothesis 2: In the following Olympic games (after 4 years), host countries still benefit from the host effect (explanations: mentality change, new forms of preparation – more knowledge, new infrastructures).

Model:

$$N\_Medals_{(c,t)} = \alpha_1 + \beta_1 * N\_Participants_{(c,t)} + \beta_2 * HE_{(c,t)} + \beta_3 * AHE_{(c,t)}$$

where:

$N\_Medals_{(c,t)}$ - Number of medals of country c in year t (dependent variable)

$N\_Participants_{(c,t)}$ - Number of participants of country c in year t (control variable)

$HE_{(c,t)}$ - *dummy variable* (1 if host country, else 0)

$AHE_{(c,t)}$ - *dummy variable* (1 if host country 4 years before, else 0)

Host Effect - Liner Regression

Data Base – 3. 806 observations

Year	Season	country	N_participants	N_medals	HE	AHE
1896	Summer	Australia	1	3	0	0
1896	Summer	Austria	3	5	0	0
1896	Summer	Denmark	3	6	0	0
1896	Summer	France	12	11	0	0
1896	Summer	Germany	19	32	0	0
1896	Summer	Greece	102	48	1	0
1896	Summer	Hungary	7	6	0	0
1896	Summer	Italy	1	0	0	0
1896	Summer	Sweden	1	0	0	0
1896	Summer	Switzerland	3	3	0	0
1896	Summer	UK	10	9	0	0
1896	Summer	USA	14	20	0	0
1900	Summer	Argentina	1	0	0	0
1900	Summer	Australia	2	6	0	0
1900	Summer	Austria	16	6	0	0
1900	Summer	Belgium	64	43	0	0

3 Linear Regressions:

- Entire database (Summer + Winter)
- Summer Olympics (2.789)
- Winter Olympics (1.017)

Host Effect - Liner Regression

Dependent Variable: N_medals

Number of Observations Read	3806
Number of Observations Used	3806

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	2625937	875312	3285.32	<.0001
Error	3802	1012973	266.43171		
Corrected Total	3805	3638910			

Root MSE	16.32274	R-Square	0.7216
Dependent Mean	10.44982	Adj R-Sq	0.7214
Coeff Var	156.20118		

Parameter Estimates						
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Standardized Estimate
Intercept	1	-4.83945	0.30777	-15.72	<.0001	0
N_participants	1	0.30603	0.00341	89.75	<.0001	0.82998
HE	1	12.49082	2.45601	5.09	<.0001	0.04645
AHE	1	4.46598	2.57133	1.74	0.0825	0.01509

Model 1 (3.086 observations)

- With a p-value of zero to four decimal places (F-test), the model is statistically significant.

- The R-squared is 0,7216, meaning that approximately 72% of the variability of 'Number of Medals' is explained by the variables in the model.

- All the variables are statistically significant, but AHE ('after host effect) is less than others (p-value=0,0825).

- The coefficient of HE ('host effect') is 12,49 – the host country get more 12,49 medals than the others.

Host Effect - Liner Regression

Dependent Variable: N_medals

Number of Observations Read	2789
Number of Observations Used	2789

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	2504031	834677	2474.51	<.0001
Error	2785	939409	337.31027		
Corrected Total	2788	3443440			

Root MSE	18.36601	R-Square	0.7272
Dependent Mean	12.21836	Adj R-Sq	0.7269
Coeff Var	150.31487		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Standardized Estimate
Intercept	1	-5.44503	0.41092	-13.25	<.0001	0
N_participants	1	0.30383	0.00417	72.85	<.0001	0.81382
HE	1	24.91551	3.79862	6.56	<.0001	0.07193
AHE	1	11.92796	3.93766	3.03	0.0025	0.03070

Model 2 (2.789 observations)

- Similar results to model 1
- The AHE ('after host effect) is more statistically significant than in the model 1. The 'after host effect' is responsible for more 11,9 medals.
- The coefficient of HE ('host effect') is 24,9 – to be the host country in Summer Olympics implies more 24,9 medals.

Host Effect - Liner Regression

Dependent Variable: N_medals

Number of Observations Read	1017
Number of Observations Used	1017

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	105887	35296	627.96	<.0001
Error	1013	56938	56.20691		
Corrected Total	1016	162824			

Root MSE	7.49713	R-Square	0.6503
Dependent Mean	5.59980	Adj R-Sq	0.6493
Coeff Var	133.88197		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Standardized Estimate
Intercept	1	-2.02780	0.29394	-6.90	<.0001	0
N_participants	1	0.27474	0.00666	41.24	<.0001	0.81714
HE	1	-1.42556	1.69459	-0.84	0.4004	-0.01639
AHE	1	-3.53113	1.77137	-1.99	0.0465	-0.03779

Model 3 (1.017 observations)

- The R-squared is only 0,6503, meaning that approximately 65% of the variability of 'Number of Medals' is explained by the variables in the model (less than the 72% of the others models).

- The HE ('host effect') AHE ('after host effect') are not statistically significant. These variables have no ability to explain the number of medals in the winter Olympics.

Demographics of Olympics - Conclusions

What has been the demographics of Olympics – what age group, gender(sex) and countries have more participants and better results?

- The average age of the participants is 25,5 years old. Over time, the average has decreased until 1980 (summer) and 1984 (winter) and in recent years it has increased slightly.

- If an athlete is between 20 and 25 years old the probability of winning a medal is higher than in other age (relation statistically significant) and works for both genders (male and female).

- Female and male participation has increased in recent years. In the summer games there is a decrease in the gap between men and women participation, but in the winter games the difference tends to be constant in more recent years. The chance of a woman winning a medal is slightly higher when participate in an Olympic game.

- USA, Russia and Germany are the 3 countries with the more medals. In terms of the ratio [number of medals / number of participants], the winner is Russia (3.947 medals out of just 5.610 participants) - 70%.

29

Regularities in the Olympics - Conclusions

Which country has been most regular in Olympics, who has been winning more medals? Which sports in which season Olympics is popular amongst the top countries?

- USA (19) and Russia (16) are the countries that more times won the Olympics. USA (15) won more times Summer Olympics and Russia (9) won more times the Winter ones.

- The sports where the 3 countries win more medals are different: USA athletics, Russia gymnastics and Germany Rowing (all Summer sports). Athletics is the only sport that is in top 5 in both countries.

- In terms of Winter sports, the USA and Russia have Ice Hokey in the top 5 and Germany does not have any sport.

Host Effect in Olympics - Conclusions

What effect does the host country have in the medals won at the Olympics?

- The countries that already hosted an Olympic competition have more medals.

- The reasons why the host countries win more medals are due to home field advantage, crowd support and easier qualification criteria, hence more participation.

- Using graphic visualization, in the year that a country receives the Olympics there is an increase in the medals won (true for 16 of the 19 countries analyzed in the Summer Olympics and for the 14 countries analyzed in the Winter Olympics).

- Linear regression analysis prove that the number of medals by country could be explained by the number of participants, the host country effect and the after host country effect. The host country effect and the after host country effect relation is only statistically significant for Summer games.

Olympic Games

Evidence and Facts

Milestone 4: Presenting Your Findings (Storytelling)!

Thank You !