CS 185C – BIG DATA – Team 1
Arpi Shah – Huy Vu – Seungbeom Ma

ANALYSIS OF SUMMER OLYMPIC GAMES

TABLE OF CONTENTS

CREDENTIALS

Username: root
Password: humsub1hai
Public IP: 173.1.75.98
Private IP: 10.113.121.2

PROPOSAL

The Olympic Games, originated in Ancient Greek, is a major international sporting event that held every four years. The modern Olympics exist in 2 forms: the Summer and Winter Games. In this project, we focus on analyzing exclusively the Summer version of this event, which first officially began in 1896. We will examine the growth of the Summer Olympic Games throughout its entire history with regard to its registered sports as well as participating country and athletes (except the most recent edition, London 2012 because we have no equivalent records for this edition). Also, we will look at several non-sport aspect of the events such as politics and cultures and how they affect the Olympics; we will as well analyze performance of several "major powers" of the game.

QUERIES & ANALYSIS

We will perform the following queries to analyze our Olympic Games dataset

1. Expansion of the Summer Olympics over its entire history
    a. Growth of participating nations
    b. Growth of athletes
    c. Growth of sports and disciplines
2. Olympic Games composition
    a. Analyze the distribution of sport events throughout history
    b. Number of appearance of each sport in the Olympic
3. Geographical distribution
    a. Distribution of the Olympic hosts over the continents (does that balance the IOC's* fairness policy)
    b. Distribution of medals won by continents
4. Effects of the home ground
    a. Compare performance of host country against the Olympic as a whole
    b. Determine which countries has hosted multiple times
5. Politics and the Olympics
    a. Analyze the performance of Eastern European countries before and after the collapse of the Iron Curtain in 1991 (Soviet and Russia, Czechoslovakia vs. Czech Republic and Slovakia, East and West Germany vs. Unified Germany)
    b. Germany (1936), Soviet Union (1980), and China (2008) are believed to integrate

propaganda into the Olympic Games to serve political purposes. Determine if these Olympic Games might have any special characteristics (medals, participants,...) compared to the general trends

6. Top 5 countries and continents in performance
    a. Before the Great War (before 1914)
    b. During the Cold War (1945-1990)
    c. In the 21st Century (after 2000)
7. Heroes of the People
    a. National heroes: those who won medals at 5 or more Olympics
    b. Superstars: those who have reached 10 medals
    c. Olympic Legends: the superstars with more than 5 gold medals
8. Sport Distribution on Gender
    a. Ratio of men and women and its adjustment over the entire history
    b. Number of events available for exclusively one gender
    c. Sport disciplines that designed exclusively for one gender
    d. Composition of mixed event in the Olympic
9. National strength in the modern Olympic (since 2000)
    a. Determine what is the leading countries for each sport in the modern Olympics
    b. Determine the super powers of the modern Olympics (top 3 countries)
10. Olympics result and FIFA correlation
    a. Winner of the Olympics and the corresponding FIFA World Cup (which happens regularly 2 years after)
    b. Frequency of teams finishing the podium of the Olympics and FIFA World Cup

**IOC: International Olympic Committee*
***NOC: National Olympic Committee (all athletes of a single nation or territory*
****Football: by definition of the IOC, football refers to association football (or soccer)*

TERMINOLOGY

Sport: a group of similar activities by their natures

Discipline: within a sport, activities are classified into disciplines

Event: within a discipline, events are content-specific

Gender: within an event, it can be classified into Men, Women, or Mixed (denoted as X in dataset, in which both men and women compete together)

NOC: In this project, NOC prefers to a country code

Example: Sport = Aquatics, Discipline=Swimming, Event=100m Freestyle, Gender = M

SUMMER OLYMPIC GAMES DATASET INFORMATION

Our dataset contains 11 files in comma-separated value format (csv) obtained from different sources on the Internet. Below are detailed descriptions of each files and the data they contain

1. **AtheleteMedal** table built from *AtheleteMedal.csv*
   This table is our main data, recording who gets what medal in what sporting discipline event in which year

Column Name	Column Description	Data Type
City	City that hosts the game	STRING
Edition	Year of the game	INT
Sport	Type of sport registered for the game	STRING
Discipline	Type of discipline within a particular sport	STRING
Athlete	Name of athlete	STRING
NOC	Nationality of athlete	STRING
Gender	Gender of athlete	STRING
Events	Sport event within a particular sport discipline	STRING
Event_gender	Gender of event	STRING
Medal	Medal obtained by athlete	STRING

2. **CountryCode** table built from *CountryCode.csv*
   This table map countries to their respective NOC code

Column Name	Column Description	Data Type
NOC	NOC code of a country	STRING
Country	Country name	STRING

3. **HostCountry** table built from *HostCountry.csv*
   This table shows which country hosts the game in which year

Column Name	Column Description	Data Type
Edition	Year of the game	INT
NOC	NOC code of a country	STRING

4. **MedalByYear** table built from *MedalByYear.csv*
   This table shows distribution of medal throughout the years

Column Name	Column Description	Data Type
Edition	Year of the game	INT
MenBronze	Number of bronze medal by men	INT
WomenBronze	Number of bronze medal by women	INT
TotalBronze	Number of bronze medal total	INT
MenSilver	Number of silver medal by men	INT

WomenSilver	Number of silver medal by women	INT
TotalSilver	Number of silver medal total	INT
MenGold	Number of gold medal by men	INT
WomenGold	Number of gold medal by women	INT
TotalGold	Number of gold medal total	INT
TotalMedal	Number of grand total medal	INT

5. **MedalByCountry** table built from *MedalByCountry.csv*
   This table shows total medals obtained by each nations through their athletes

Column Name	Column Description	Data Type
NOC	NOC code of a country	STRING
Country	Country name	STRING
Bronze	Number of bronze medal by a country	INT
Gold	Number of gold medal by a country	INT
Silver	Number of silver medal by a country	INT
TotalMedal	Number of grand total medal	INT

6. **FIFAWorldCupResult** table built from *FIFAWorldCupResult.csv*
   This table shows result of the FIFA World Cup, which is used to compared with football result from the Olympic Games

Column Name	Column Description	Data Type
Edition	Year of the World Cup	INT
Champ	Champion Team	STRING
Second	2nd place team	STRING
Third	3rd place team	STRING
Fourth	4th place team	STRING

7. **Events** table built from *Events.csv*
   This table shows list of sporting disciplines and events in Olympics, including those ceased to exist

Column Name	Column Description	Data Type
Sport	Type of sport registered for the game	STRING
Discipline	Type of discipline within a particular sport	STRING
Events	Sport event within a particular sport discipline	STRING
Gender	Gender of event	STRING

8. **TeamMedal** table built from *TeamMedal.csv*
   This table shows medals obtained through team sports by each country

Column Name	Column Description	Data Type
NOC	NOC code of a country	STRING
Country	Country name	STRING
TotalMedal	Number of grand total medal	INT
Gold	Number of gold medal by a country	INT
Silver	Number of silver medal by a country	INT
Bronze	Number of bronze medal by a country	INT

9. **MultipleMedalist** table built from *MultipleMedalist.csv*
   This table lists athletes who have earned multiple medals

Column Name	Column Description	Data Type
Athlete	Name of athlete	STRING
Nation	Country name	STRING
Sport	Type of sport registered for the game	STRING
Edition	Year of the game	STRING
Games	Type of Olympic Games	STRING
Gender	Gender of athlete	STRING
Gold	Number of gold medal by a country	INT
Silver	Number of silver medal by a country	INT
Bronze	Number of bronze medal by a country	INT
TotalMedal	Number of grand total medal	INT

10. **CountryContinent** table built from *CountryContinent.csv*
    This table shows countries and continents relationship

Column Name	Column Description	Data Type
Continent	Continent name	STRING
Country	Country name	STRING

11. **Participant** table built from *Participant.csv*
    This table shows number of participating nations and athletes and registered sports

Column Name	Column Description	Data Type
Edition	Year of the game	INT
Nation	Number of participating nation	INT
TotalCompetitor	Number of grand total competitors	INT
MenCompetitor	Number of men competitors	INT
WomenCompetitor	Number of women competitors	INT
Sport	Number of registered sports	INT
Discipline	Number of registered disciplines	INT
Event	Number of registered events	INT

PREPROCESSING

Most of the preprocessing works have been done in Excel

1. Incomplete and inconsistent data :
   Some countries have ceased to exist and thus their associated Olympic data was left incomplete (there are exceptions with bigger countries like Soviet Union and Czechoslovakia)

2. Column naming :
   Some columns have name same as reserved words in SQL, which may cause confusion and error when querying. They were therefore replaced with a different name
   Year → Edition
   First → Champ
   Event → Events

3. Embedded comma issue :
   a. Athlete Name
      Originally, the AthleteName column is shown as "**Lastname, Firstname**" and Hadoop does not handle embedded commas. This causes big problems as they parse into 2 columns instead of 1. Therefore, we change this column format to "**Lastnam Firstname**"
   b. Country:
      Similarly, some countries have their name formatted with embedded comma, ex: **Korea, South** or **Congo, Dem Rep**; we also eliminate these embedded comma to keep our data consistent.

LOADING DATASET ONTO SERVERS

1. To establish connection with server
   Mac users: Use terminal to connect
   Windows user: Download putty.exe and pscp.exe to establish connection with server. We copied pscp.exe into "C:\Windows\"
   Then we opened command prompt and connected to the server. For the purpose, go to step2

2. From our system, we copied files to the master node's local file system.
   We used the following commands for the purpose:

   a. In the master node, we logged in and created a blank directory(using putty) where we will store the files

```
[root@master ~]# mkdir ProjectInput3
[root@master ~]# ls ProjectInput3
[root@master ~]#
```

b. We copied the files using the following command (use command prompt)

```
C:\Users\User Account for SCS>pscp "C:\Users\User Account for SCS\Desktop\datase
t.zip" root@173.1.75.98:/root/ProjectInput3
root@173.1.75.98's password:
dataset.zip               | 434 kB | 434.1 kB/s | ETA: 00:00:00 | 100%

C:\Users\User Account for SCS>
```

c. We checked whether files are transferred on on master node by using putty

```
[root@master ~]# ls -l
total 1192
-rw-r--r--  1 root root 210357 Mar  1 04:45 22
-rw-------. 1 root root   1268 Aug 30  2011 anaconda-ks.cfg
drwxr-xr-x  2 root root   4096 Feb 18 10:28 Arpi_ip
-rwxr-xr-x  1 root root 501703 Feb 18 08:13 cloudera-manager-installer.bin
drwxr-xr-x  2 root root   4096 Feb 18 10:25 Example_IP
-rw-r--r--. 1 root root   8730 Aug 30  2011 install.log
-rw-r--r--. 1 root root   3094 Aug 30  2011 install.log.syslog
drwxr-xr-x  2 root root   4096 Mar  1 05:01 Ip
-rw-r--r--  1 root root 210357 Mar  1 04:53 \Ip
-rw-r--r--  1 root root 210357 Mar  1 04:55 one.pdf
drwxr-xr-x  4 root root   4096 Mar  1 05:11 ProjectInput
drwxr-xr-x  3 root root   4096 Mar  1 08:18 ProjectInput1
drwxr-xr-x  3 root root   4096 Mar  1 08:42 ProjectInput2
drwxr-xr-x  2 root root   4096 Mar  1 15:46 ProjectInput3
[root@master ~]# cd ProjectInput3
[root@master ProjectInput3]# ls -l
total 440
-rw-r--r-- 1 root root 444513 Mar  1 15:46 dataset.zip
[root@master ProjectInput3]#
```

d. We Unzipped the file

e. We went to the unzipped folder for further processing

```
[root@master ProjectInput3]# unzip dataset
Archive:  dataset.zip
   creating: dataset/
  inflating: dataset/AthleteMedal.csv
  inflating: dataset/CountryCode.csv
  inflating: dataset/CountryContinent.csv
  inflating: dataset/Event.csv
  inflating: dataset/FIFAWorldCupResult.csv
  inflating: dataset/HostCountry.csv
  inflating: dataset/MedalByCountry.csv
  inflating: dataset/MedalByYear.csv
  inflating: dataset/MultipleMedalist.csv
  inflating: dataset/Pariticipant.csv
  inflating: dataset/reference_info.xlsx
  inflating: dataset/TeamMedal.csv
  inflating: Olympicsproposal.docx
[root@master ProjectInput3]# ls -l
total 460
drwxr-xr-x 2 root root   4096 Mar  1 15:17 dataset
-rw-r--r-- 1 root root 444513 Mar  1 15:46 dataset.zip
-rw-r--r-- 1 root root  14502 Feb 26 11:15 Olympicsproposal.docx
[root@master ProjectInput3]# cd dataset
[root@master dataset]# ls -l
total 2564
-rw-r--r-- 1 root root 2510924 Mar  1 15:10 AthleteMedal.csv
-rw-r--r-- 1 root root    3091 Mar  1 15:11 CountryCode.csv
-rw-r--r-- 1 root root    3761 Mar  1 15:16 CountryContinent.csv
-rw-r--r-- 1 root root   32418 Mar  1 15:15 Event.csv
-rw-r--r-- 1 root root     743 Mar  1 15:14 FIFAWorldCupResult.csv
-rw-r--r-- 1 root root     260 Mar  1 15:11 HostCountry.csv
-rw-r--r-- 1 root root    3530 Mar  1 15:13 MedalByCountry.csv
-rw-r--r-- 1 root root    1140 Mar  1 15:13 MedalByYear.csv
-rw-r--r-- 1 root root   27578 Mar  1 15:16 MultipleMedalist.csv
-rw-r--r-- 1 root root     863 Mar  1 15:17 Pariticipant.csv
-rw-r--r-- 1 root root    8796 Mar  1 15:17 reference_info.xlsx
-rw-r--r-- 1 root root    3248 Mar  1 15:15 TeamMedal.csv
[root@master dataset]#
```

3. We copied files from master node's local file system to hadoop file system (HDFS) using the following steps.
   a. We created a directory where we can save the files (used putty)

```
[root@master ~]# hadoop fs -mkdir ProjectInput3
[root@master ~]# hadoop fs -ls ProjectInput3
[root@master ~]#
```

   b. We copied all the files from the unzipped folder in local file system to HDFS using the following command: (putty)

      From the directory where we kept all the files in the local file system, we executed commands in the following steps using putty.

```
[root@master ~]# cd ProjectInput3
[root@master ProjectInput3]# ls -l
total 460
drwxr-xr-x 2 root root   4096 Mar  1 15:17 dataset
-rw-r--r-- 1 root root 444513 Mar  1 15:46 dataset.zip
-rw-r--r-- 1 root root  14502 Feb 26 11:15 Olympicsproposal.docx
[root@master ProjectInput3]# cd dataset
[root@master dataset]# ls -l
total 2564
-rw-r--r-- 1 root root 2510924 Mar  1 15:10 AthleteMedal.csv
-rw-r--r-- 1 root root    3091 Mar  1 15:11 CountryCode.csv
-rw-r--r-- 1 root root    3761 Mar  1 15:16 CountryContinent.csv
-rw-r--r-- 1 root root   32418 Mar  1 15:15 Event.csv
-rw-r--r-- 1 root root     743 Mar  1 15:14 FIFAWorldCupResult.csv
-rw-r--r-- 1 root root     260 Mar  1 15:11 HostCountry.csv
-rw-r--r-- 1 root root    3530 Mar  1 15:13 MedalByCountry.csv
-rw-r--r-- 1 root root    1140 Mar  1 15:13 MedalByYear.csv
-rw-r--r-- 1 root root   27578 Mar  1 15:16 MultipleMedalist.csv
-rw-r--r-- 1 root root     863 Mar  1 15:17 Pariticipant.csv
-rw-r--r-- 1 root root    8796 Mar  1 15:17 reference_info.xlsx
-rw-r--r-- 1 root root    3248 Mar  1 15:15 TeamMedal.csv
[root@master dataset]# hadoop fs -put *.csv ProjectInput3
[root@master dataset]#
```

c. We checked whether files have come on hdfs.

```
[root@master ~]# hadoop fs -ls
Found 10 items
drwx------   - root root          0 2014-03-01 16:00 .Trash
drwx------   - root root          0 2014-03-01 08:30 .staging
drwxr-xr-x   - root output        0 2014-02-18 10:36 Arpi_op
drwxr-xr-x   - root root          0 2014-02-18 10:40 Arpi_op1
drwxr-xr-x   - root root          0 2014-03-01 07:25 ProjectInput
drwxr-xr-x   - root root          0 2014-03-01 08:29 ProjectInput1
drwxr-xr-x   - root root          0 2014-03-01 16:00 ProjectInput3
-rw-r--r--   3 root root       2497 2014-02-18 10:34 ip_wc.txt
drwxr-xr-x   - root root          0 2014-02-18 10:49 tip
drwxr-xr-x   - root root          0 2014-02-18 10:49 top
[root@master ~]# hadoop fs -ls ProjectInput3
Found 11 items
-rw-r--r--   3 root root    2510924 2014-03-01 16:00 ProjectInput3/AthleteMedal.csv
-rw-r--r--   3 root root       3091 2014-03-01 16:00 ProjectInput3/CountryCode.csv
-rw-r--r--   3 root root       3761 2014-03-01 16:00 ProjectInput3/CountryContinent.csv
-rw-r--r--   3 root root      32418 2014-03-01 16:00 ProjectInput3/Event.csv
-rw-r--r--   3 root root        743 2014-03-01 16:00 ProjectInput3/FIFAWorldCupResult.csv
-rw-r--r--   3 root root        260 2014-03-01 16:00 ProjectInput3/HostCountry.csv
-rw-r--r--   3 root root       3530 2014-03-01 16:00 ProjectInput3/MedalByCountry.csv
-rw-r--r--   3 root root       1140 2014-03-01 16:00 ProjectInput3/MedalByYear.csv
-rw-r--r--   3 root root      27578 2014-03-01 16:00 ProjectInput3/MultipleMedalist.csv
-rw-r--r--   3 root root        863 2014-03-01 16:00 ProjectInput3/Pariticipant.csv
-rw-r--r--   3 root root       3248 2014-03-01 16:00 ProjectInput3/TeamMedal.csv
[root@master ~]#
```

4. We used hive to create tables and load tables from the files imported into hdfs using: Using hive

```
[root@master ~]# hive
Logging initialized using configuration in jar:file:/usr/lib/hive/lib/hive-common-0.10.0-cdh4.5.0.jar!/hive-log4j.properties
Hive history file=/tmp/root/hive_job_log_140aa267-34c0-4e71-90b7-e77f641de617_132240879.txt
hive>
```

```
hive> CREATE TABLE IF NOT EXISTS CountryCode
    > (NOC STRING,
    > COUNTRY STRING)
    > COMMENT 'Country Code Table'
    > ROW FORMAT DELIMITED
    > FIELDS TERMINATED BY ','
    > STORED AS TEXTFILE;
OK
Time taken: 2.176 seconds
hive> LOAD DATA INPATH './ProjectInput3/CountryCode.csv' OVERWRITE INTO TABLE CountryCode;
```

This creates a subdirectory under the **/user/hive/warehouse directory** in HDFS

We checked if the subdirectory is created

```
[root@master /]# hadoop fs -ls ../hive/warehouse
Found 3 items
drwxrwxrwt   - root hive          0 2014-03-01 07:51 ../hive/warehouse/countrycode
```

DATA VERIFICATION

We checked if the file is properly loaded by running select * query on all of them. Below is the verification for the CountryCode table

```
hive> select count(*) from CountryCode
    > ;
Total MapReduce jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapred.reduce.tasks=<number>
Starting Job = job_201402180934_0006, Tracking URL = http://master.bigtest.com:50030/jobdetails.jsp?jobid=job_201402180934_0006
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_201402180934_0006
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2014-03-01 16:44:10,848 Stage-1 map = 0%,  reduce = 0%
2014-03-01 16:44:17,919 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.9 sec
2014-03-01 16:44:18,953 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.9 sec
2014-03-01 16:44:19,972 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.9 sec
2014-03-01 16:44:20,982 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.9 sec
2014-03-01 16:44:21,994 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.9 sec
2014-03-01 16:44:23,013 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.9 sec
2014-03-01 16:44:24,026 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.34 sec
2014-03-01 16:44:25,038 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.34 sec
2014-03-01 16:44:26,062 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.34 sec
2014-03-01 16:44:27,074 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 4.34 sec
MapReduce Total cumulative CPU time: 4 seconds 340 msec
Ended Job = job_201402180934_0006
MapReduce Jobs Launched:
Job 0: Map: 1  Reduce: 1   Cumulative CPU: 4.34 sec   HDFS Read: 3323 HDFS Write: 4 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 340 msec
OK
208
Time taken: 25.419 seconds
```

We did it for all the files in the data set. And verification from **HostCountry** table

```
hive> CREATE TABLE IF NOT EXISTS HostCountry
    > (YEAR INT,
    > NOC STRING)
    > COMMENT 'Host Country Table'
    > ROW FORMAT DELIMITED
    > FIELDS TERMINATED BY ','
    > STORED AS TEXTFILE;
OK
Time taken: 2.219 seconds
hive> LOAD DATA INPATH './ProjectInput1/HostCountry.csv' OVERWRITE INTO TABLE HostCountry
;
Loading data to table default.hostcountry
chgrp: changing ownership of '/user/hive/warehouse/hostcountry/HostCountry.csv': User doe
s not belong to hive
Table default.hostcountry stats: [num_partitions: 0, num_files: 1, num_rows: 0, total_siz
e: 260, raw_data_size: 0]
OK
Time taken: 2.778 seconds
hive> select * from hostCountry
    > ;
OK
1896    GRE
1900    FRA
1904    USA
1908    GBR
1912    SWE
1920    BEL
1924    FRA
1928    NED
1932    USA
1936    GER
1948    GBR
1952    FIN
1956    AUS
1960    ITA
1964    JPN
1968    MEX
1972    GER
1976    CAN
```

Verification of AtheleteMedal table by printing first 10 rows of athlete name (to see if preprocessing is correct)

```
00m freestyle   M       Gold
Time taken: 0.38 seconds
hive> select * from AthleteMedal limit 10;
OK
Athens  1896    Aquatics        Swimming        HAJOS Alfred    HUN     Men     100m freestyle  M       Gold
Athens  1896    Aquatics        Swimming        HERSCHMANN Otto AUT     Men     100m freestyle  M       Silver
Athens  1896    Aquatics        Swimming        DRIVAS Dimitrios        GRE     Men     100m freestyle for sailors      M       Bronze
Athens  1896    Aquatics        Swimming        MALOKINIS Ioannis       GRE     Men     100m freestyle for sailors      M       Gold
Athens  1896    Aquatics        Swimming        CHASAPIS Spiridon       GRE     Men     100m freestyle for sailors      M       Silver
Athens  1896    Aquatics        Swimming        CHOROPHAS Efstathios    GRE     Men     1200m freestyle M       Bronze
Athens  1896    Aquatics        Swimming        HAJOS Alfred    HUN     Men     1200m freestyle M       Gold
Athens  1896    Aquatics        Swimming        ANDREOU Joannis GRE     Men     1200m freestyle M       Silver
Athens  1896    Aquatics        Swimming        CHOROPHAS Efstathios    GRE     Men     400m freestyle  M       Bronze
Athens  1896    Aquatics        Swimming        NEUMANN Paul    AUT     Men     400m freestyle  M       Gold
Time taken: 0.202 seconds
hive> select Athlete from AthleteMedal limit 10;
Total MapReduce jobs = 1
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_201403010515_0024, Tracking URL = http://test.huydomain.com:50030/jobdetails.jsp?jobid=job_201403010515_0024
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_201403010515_0024
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 0
2014-03-05 02:26:00,188 Stage-1 map = 0%,  reduce = 0%
2014-03-05 02:26:07,256 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.38 se
2014-03-05 02:26:08,279 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.38 se
2014-03-05 02:26:09,326 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.38 se
2014-03-05 02:26:10,342 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.38 se
2014-03-05 02:26:11,354 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 1.38
MapReduce Total cumulative CPU time: 1 seconds 380 msec
Ended Job = job_201403010515_0024
MapReduce Jobs Launched:
Job 0: Map: 1   Cumulative CPU: 1.38 sec   HDFS Read: 65770 HDFS Write: 166 SUCC
Total MapReduce CPU Time Spent: 1 seconds 380 msec
OK
HAJOS Alfred
HERSCHMANN Otto
DRIVAS Dimitrios
MALOKINIS Ioannis
CHASAPIS Spiridon
CHOROPHAS Efstathios
```

Verification of Participant table by printing first 10 rows

```
HERSCHMANN Otto
ORIVAS Dimitrios
MALOKINIS Ioannis
CHASAPIS Spiridon
CHOROPHAS Efstathios
HAJOS Alfred
ANDREOU Joannis
CHOROPHAS Efstathios
NEUMANN Paul
Time taken: 18.513 seconds
hive> select * from Participant limit 10;
OK
1896    14      241     241     0       9       10      43
1900    24      997     975     22      19      20      85
1904    12      651     645     6       16      17      94
1908    22      2008    1971    37      22      25      110
1912    28      2407    2359    48      14      18      102
1920    29      2626    2561    65      22      29      156
1924    44      3089    2954    135     17      23      126
1928    46      2883    2606    277     14      20      109
1932    37      1332    1206    126     14      20      117
1936    49      3963    3632    331     19      25      129
Time taken: 0.28 seconds
hive>
```

QUERIES VISUALIZATION

In this section, all query results are visualized with charts, graphs and other formats. For raw result from Hive, please refer to the appendix with the provided links. In this project, following tools are used to create all visualization:

- Vizualize.me
- Infogr.am
- Microsoft Excel
- IBM Manyeyes
- Tagxedo

1/ Expansion of the Summer Olympics

This query has 3 parts: part (a) shows the increment in participating nations in the Olympics, part (b) shows total number of athletes and part (c) shows the growth of sports & disciplines

a/ Growth of participating nations (Number of participating nations)

Table: Participant, Cols: Nation, Total Competition

```
select Edition,Nation,TotalCompetitor from Participant;
```

Click here to go to raw query result

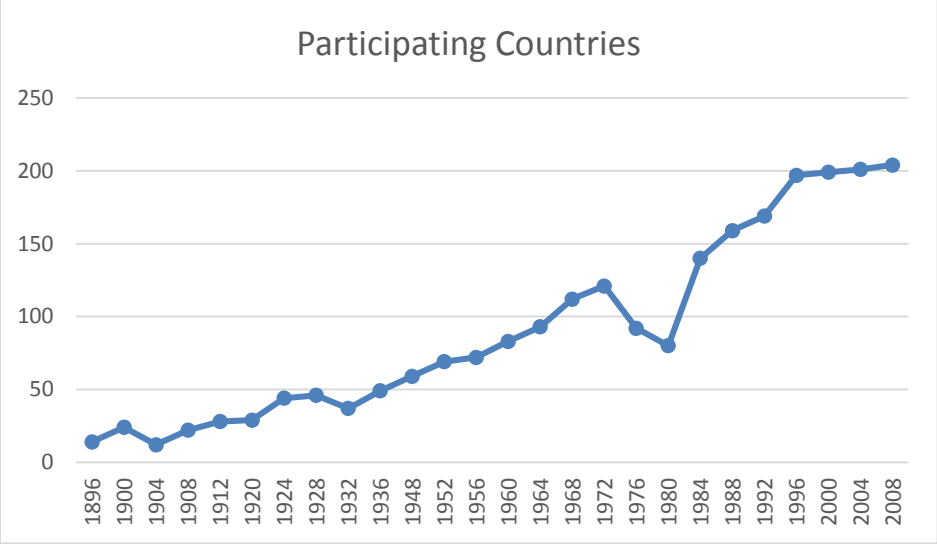


Figure 1.a: Growth of participating nations

b. Growth of athletes

Table: Participant, **Cols:** Nation, Total Competition

```
select Edition, TotalCompetitor, MenCompetitor, WomenCompetitor from Participant;
```

Click here to go to raw query result

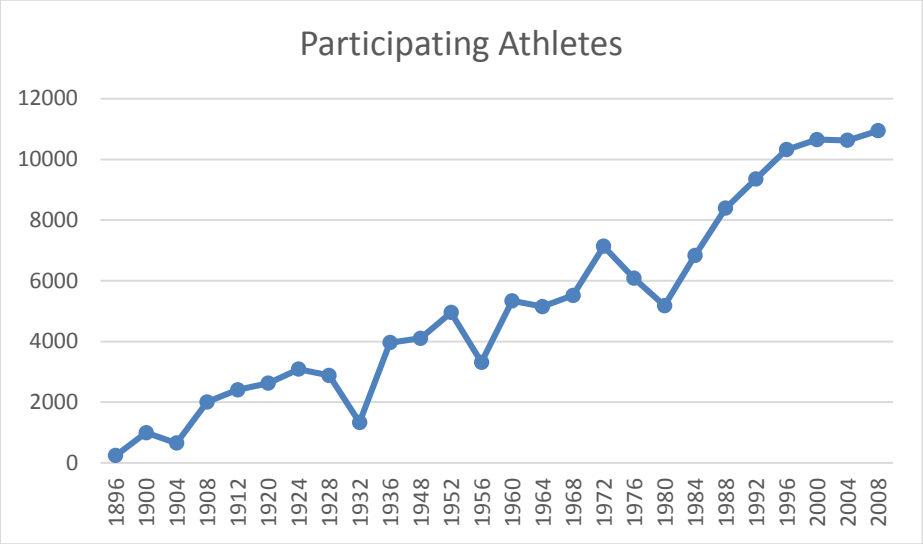


Figure 1.b: Growth of athletes participation

c. Growth **of sport, disciplines**

```
select Edition, Sport,Discipline from Participant;
```

Click here to go to raw query result

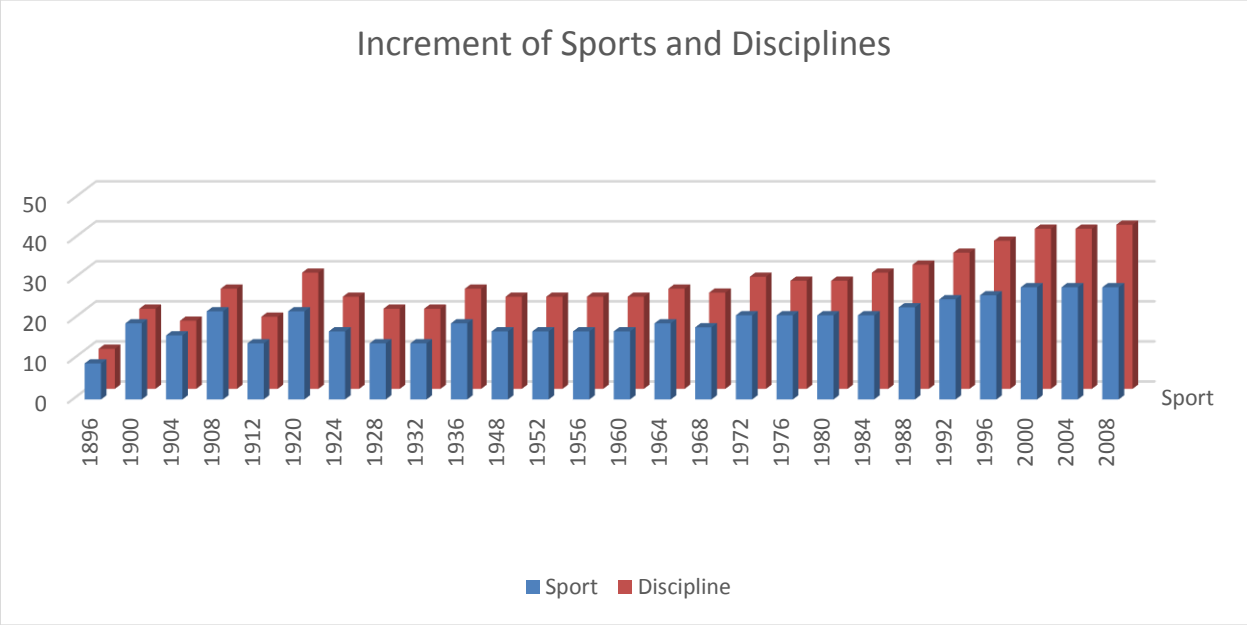


Figure 1.c: Growth of number of sports and disciplines

2. **Olympic Games composition and distribution**

In this section, we analyze the composition and distribution of all sports in the Olympic Games as well as their popularity throughout history. Both queries are represented in single figure.

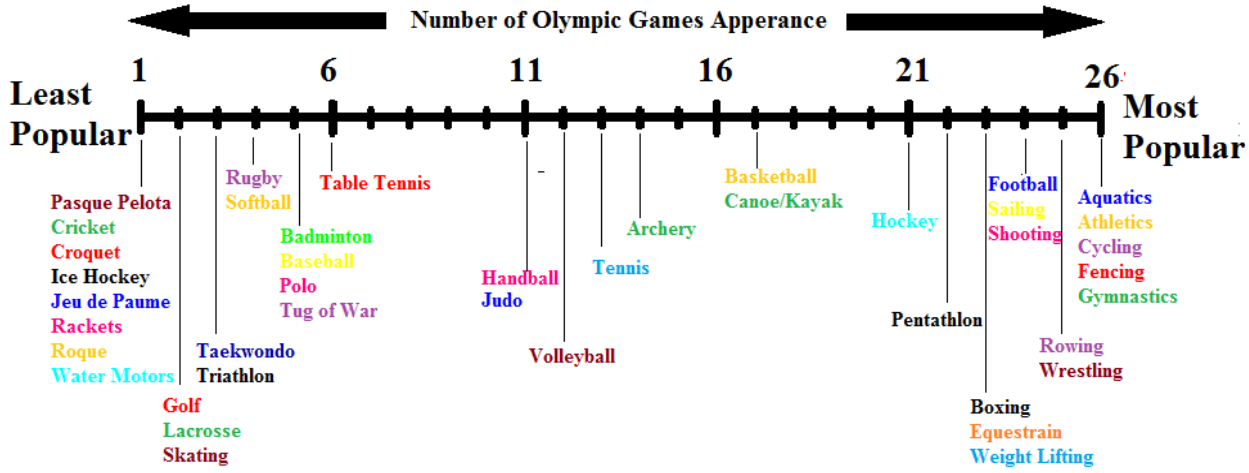


Figure 2: Composition and Distribution of Olympic Sport

a. Analyze the distribution of sport events throughout history

For each year, find the number of events each sport contributes to the Olympics Games. From this, it can be deduced which sport is more popular in certain period of time

```
select Edition, Sport, count(*) NumOfEvents from
(select distinct Edition, Sport, Discipline,Events,Event_gender from AtheleteMedal) T
group by Edition,Sport;
```

Click here to go to raw query result

b. Number of appearance of each sport in the Olympic

For each sport discipline, determine their frequency of appearance in the Olympics. From this information, it can be found whether a sport is contemporary (only in the Olympics for a short time), popular (appear in more than 10 Olympics) or native (appear in almost all edtions)

```
select Sport, Discipline, count(*) as YearsInOlympics from
(select distinct Edition, Sport, Discipline from AtheleteMedal) T
group by Sport, Discipline;
```

Click here to go to raw query result

3/Geographical distribution

a/Analyze the distribution of the Olympic hosts over the continents (does that balance the IOC's* fairness policy)

International Olympics Committee (IOC) addresses fairness policy, which rotates hosting privileges to countries around the world. This query determines the frequency of hosting in each continent to see if the IOC has been deciding the host consistently with this rule

```
select Continent, count(*) from HostCountry HC join CountryCode CC on HC.NOC=
CC.NOC join CountryContinent Cont on CC.Country=Cont.Country group by Continent;
```

Click here to go to raw query result

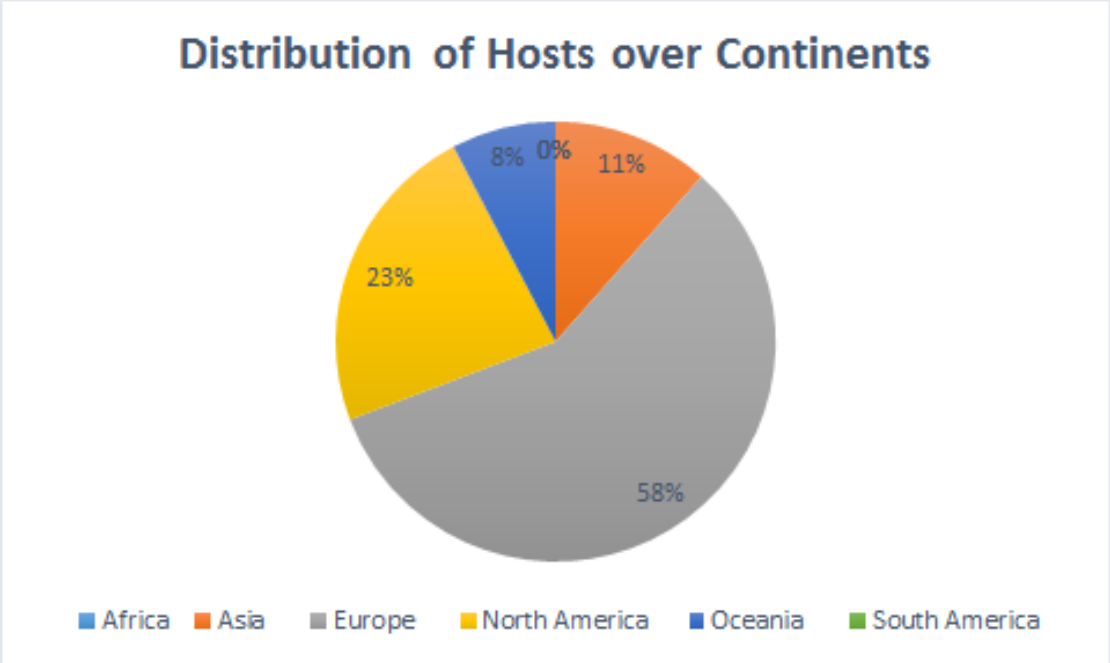


Figure 3.a: Distribution of host countries over continents

b/Analyze the distribution of the nations getting medals over the continents

This query shows how many country in each continent already obtained any medal in Olympics

```
select G.continent, count(*) from (select * from CountryContinent CC join MedalByCountry MC on CC.country = MC.country) G group by G.continent;
```

Click here to go to raw query result

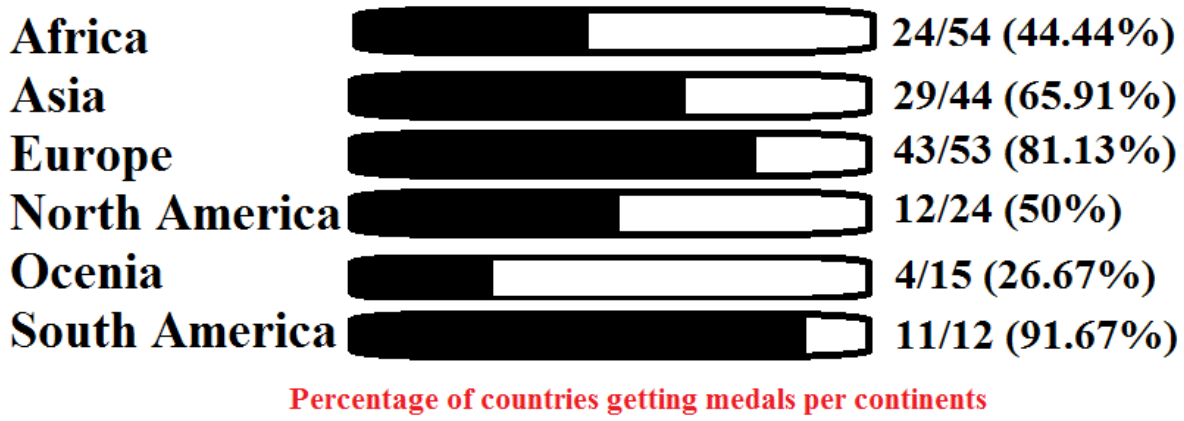


Figure 3.b: Percentage of countries getting medals per continents

4/ Effects of the home ground

The Olympic Games is hosted once every 4 year at one predetermined city. The hosting nation and its athletes are often said to gain some advantage due to their home ground.

a/ Performance of host compared to the whole Olympic

This query compare the medals obtained by host to the total medals in each edition

```
select T.Edition, T.Country, T.MedalCount, MBY.TotalMedal from (select AM.Edition,
CC.Country, count(*) as MedalCount
from AtheleteMedal AM join HostCountry HC on AM.Edition=HC.year and
AM.NOC=HC.NOC
join CountryCode CC on HC.NOC=CC.NOC group by AM.Edition, CC.Country) T join
(select Edition,TotalMedal from MedalByYear) MBY on T.Edition=MBY.Edition;
```

Click here to go to raw query result

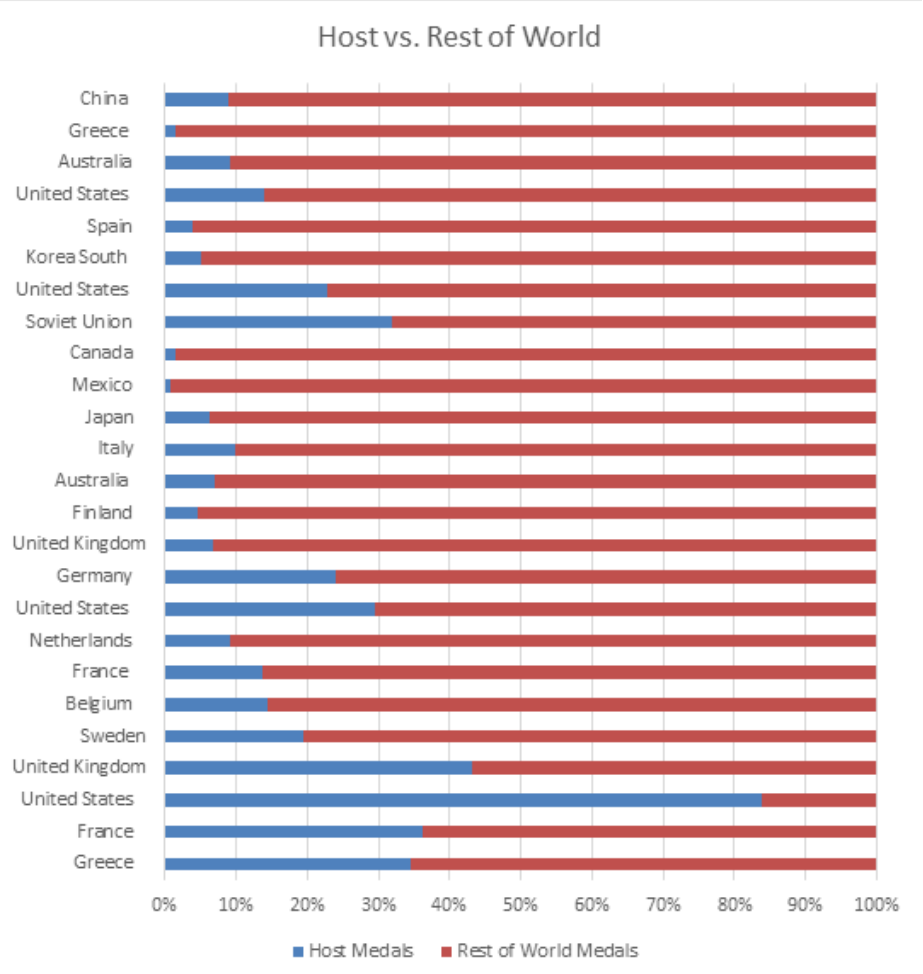


Figure 4.a: Percentage of countries getting medals per continents

b/ those who host more than once

List of countries that have hosted the Olympics multiple time

```
select CC.Country from HostCountry HC join CountryCode CC on HC.NOC=CC.NOC
group by CC.Country having count(*) >1;
```

Click here to go to raw query result

Figure 4.b: All host cities, with special indication on multiple hosts

5/ Politics and the Olympics

In this section, we analyze the performance of countries in the Olympics with regard to political changes. We observe how politics in a country play role in affecting the outcome of its athletes. For purpose of analysis, we will examine closely one of the biggest historical events in the 20th century: the fall of the Iron Curtain and the Soviet Bloc in Eastern Europe. We will also examine if the Olympics can be affected by propaganda in single party state countries.

a/ Analyze the performance of Soviet Bloc Eastern European countries before and after the collapse of the Iron Curtain in 1990

Before 1990, Eastern Europe was in control of communist regimes led by the Soviet Union until the collapse of the Soviet Empire. This part analyze performance of some of the former Soviet Bloc countries, before and after the end of the Cold War. In this section, we analyze performance of Soviet Union, Czechoslovakia, East Germany and West Germany (West Germany was not a Soviet Bloc, only search to compare to East Germany)

After 1990, Soviet Union collapsed into Russia, Czechoslovakia broke into Czech Republic and Germany unified East and West (East Germany annexed into West, thus their medal counts terminated while the West medal count continued)

```
-- Before the Iron Curtain collapsed 1990
select Country, count(*) as Medals from AtheleteMedal AM join CountryCode CC on
AM.NOC=CC.NOC where (AM.NOC='URS' or AM.NOC='TCH' or AM.NOC='GER' or
AM.NOC='GDR') and Edition <1990 group by Country;

-- After the Iron Curtain collapsed 1990
select Country, count(*) as Medals from AtheleteMedal AM join CountryCode CC on
AM.NOC=CC.NOC where (AM.NOC='RUS' or AM.NOC='CZE' or AM.NOC='GER') and
Edition >1990 group by Country;
```

Click here to go to raw query result

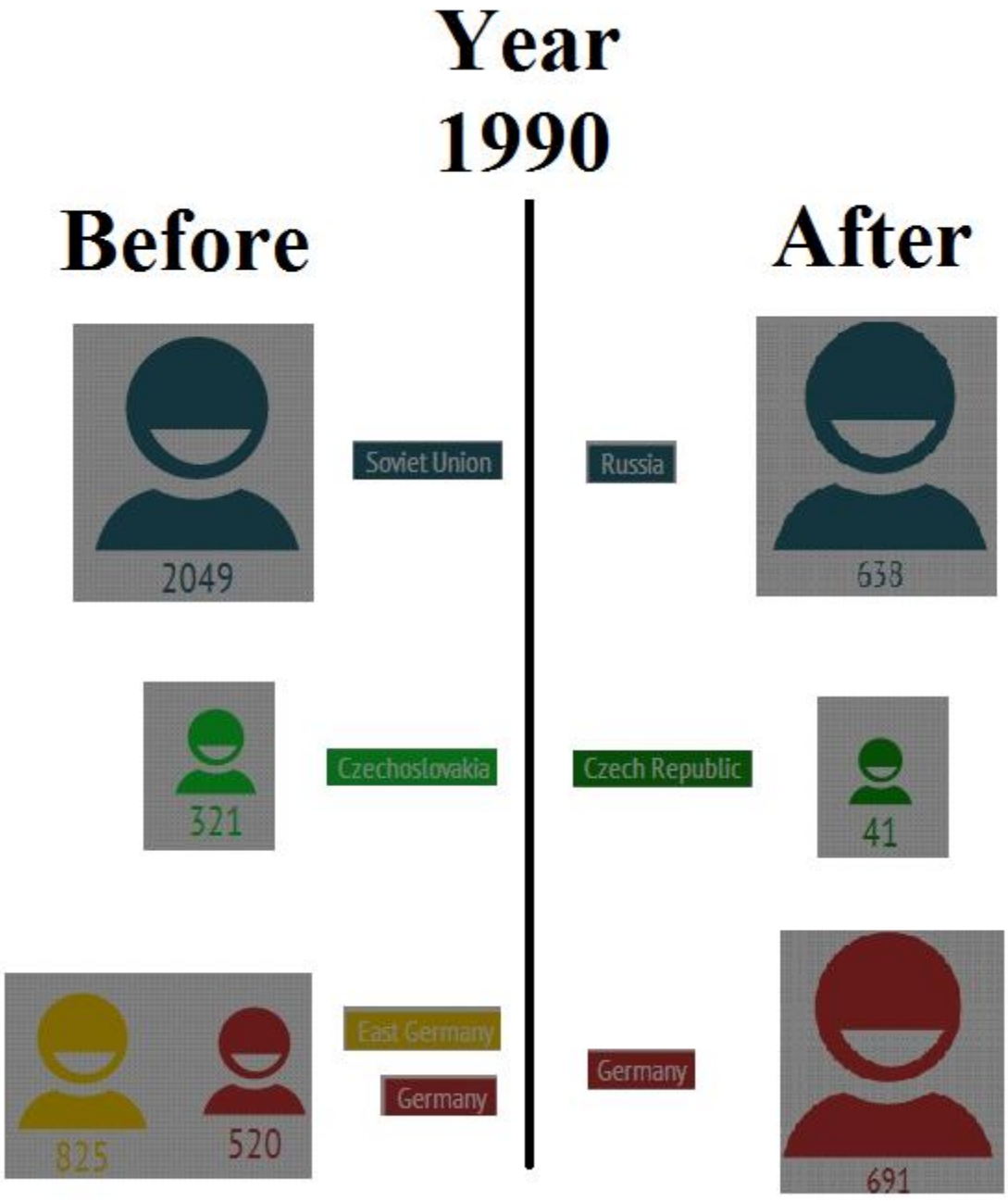


Figure 5.a: Performance of countries due to political changes in Eastern Europe

b/ Germany (1936), Soviet Union (1980), and China (2008) are believed to integrate propaganda into the Olympic Games to serve political purposes. Determine if these Olympic Games might have any special characteristics (medals, participants,...) compared to the general trends

There have been claimed that these 3 Olympics may have been fixed, to certain extent to serve propaganda purpose of the hosting nations. This query compares their medals and the overall medals in Olympics

```
select T.Edition, CC.Country, MBY.TotalGold, T.MedalCount, MBY.TotalMedal from
(select AM.Edition, AM.NOC, count(*) as MedalCount from AtheleteMedal AM
join HostCountry HC on AM.Edition=HC.Edition and AM.NOC=HC.NOC
group by AM.Edition, AM.NOC) T
join MedalByYear MBY on T.Edition=MBY.Edition
join CountryCode CC on T.NOC=CC.NOC
where T.Edition=1936 or T.Edition=1980 or T.Edition=2008;
```

Click here to go to raw query result

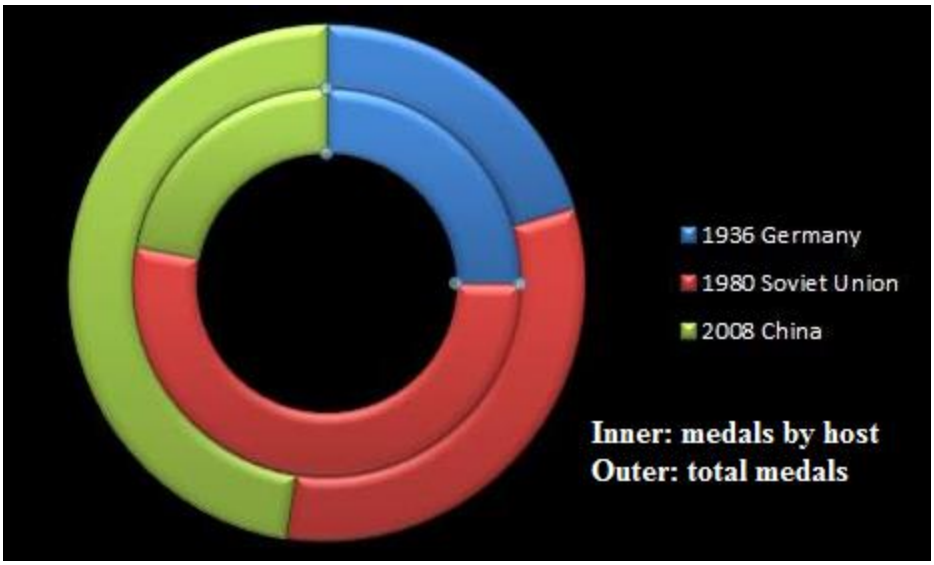


Figure 5.b: Medals of host and total in the 3 Olympics hosted by single party states countries

6/ Top performance in each historical period

In this section we determine the top 5 performance in the Olympics in 3 biggest timeline periods: before the Great War in 1914, during the Cold War (1945-1991) and in the 21st century (since 2000). With the dramatic changes in societies around the globe, we expect to see change in performance among countries and continents through the Olympics history

Click here to go to ALL raw query result for the entire section

a/ Before the Great War (1914)

In this period of time, European countries were so strong that they would dominate the rest of the world in almost everything. We expect to see that medals obtained by European countries would be largely dominating.

```
-- Top 5 countries
select CC.Country as Country, count(*)  as MedalCount from
(select distinct Edition, NOC, Sport, Discipline, Events, Event_gender, Medal
from AtheleteMedal where Edition<1914) T join CountryCode CC on T.NOC=CC.NOC
join CountryContinent Cont on CC.Country=Cont.Country
group by CC.Country sort by MedalCount desc limit 5;

-- By continents
select Continent, count(*) as MedalCount from
(select distinct Edition, NOC, Sport, Discipline, Events, Event_gender, Medal
from AtheleteMedal where Edition<1914) T join CountryCode CC on T.NOC=CC.NOC
join CountryContinent Cont on CC.Country=Cont.Country
group by Continent sort by MedalCount desc;
```

b/ 1945 – 1990 (Cold War)

In this period of time, the European colonization collapsed and replaced by the East-West rivalry between the two superpowers United States and Soviet Union along with their allies. We expect to see Olympics performance of Europe and North America to be more balanced and the two superpower countries to dominate the Olympics

```
-- Top 5 countries
select CC.Country as Country, count(*)  as MedalCount from
(select distinct Edition, NOC, Sport, Discipline, Events, Event_gender, Medal
from AtheleteMedal where Edition>1945 and Edition < 1990) T join CountryCode CC on
T.NOC=CC.NOC
join CountryContinent Cont on CC.Country=Cont.Country
group by CC.Country sort by MedalCount desc limit 5;
```

```
-- By continents
select Continent, count(*) as MedalCount from
(select distinct Edition, NOC, Sport, Discipline, Events, Event_gender, Medal
from AtheleteMedal where Edition>1945 and Edition < 1990) T join CountryCode CC on
T.NOC=CC.NOC
join CountryContinent Cont on CC.Country=Cont.Country
group by Continent sort by MedalCount desc;
```

c/ 21st Century

In the new century, Asian countries began to invest in sport. We expect to see performance of Asian countries surpass that of European or North American.

```
-- Top 5 countries
select CC.Country as Country, count(*)  as MedalCount from
(select distinct Edition, NOC, Sport, Discipline, Events, Event_gender, Medal
from AtheleteMedal where Edition>2000) T join CountryCode CC on T.NOC=CC.NOC
join CountryContinent Cont on CC.Country=Cont.Country
group by CC.Country sort by MedalCount desc limit 5;

-- By continents
select Continent, count(*) as MedalCount from
(select distinct Edition, NOC, Sport, Discipline, Events, Event_gender, Medal
from AtheleteMedal where Edition>2000) T join CountryCode CC on T.NOC=CC.NOC
join CountryContinent Cont on CC.Country=Cont.Country
group by Continent sort by MedalCount desc;
```

Figure 6.a: Top 5 countries in each period

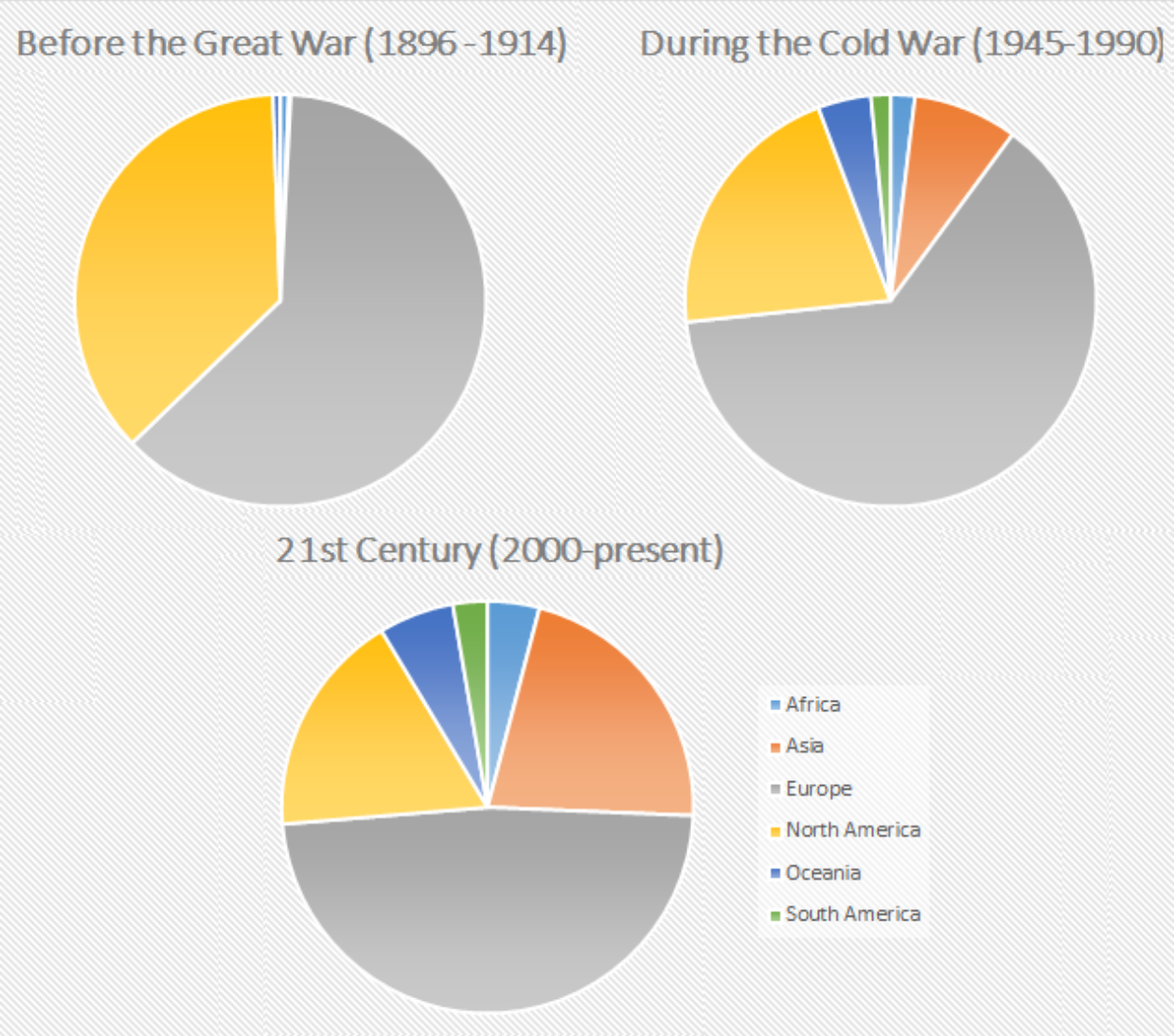


Figure 6.b: Distribution of medals by continents each period

7. Olympic Hero

Some people have become legends in their countries for great achievements in the Olympics

a. National Hero: those who won medals 5 or more Olympics

Out of hundred thousands of athletes, very few have participated in 5 Olympics (that's 20 year span). Out of those, some athletes left impressions by getting medals in all of the Olympics they attended and become heroes or heroines in their nations and cultures. This query looks for that kind of national hero who earned medals in at least 5 Olympics

```
select C.athlete, count(*) as count1  from (select distinct edition, athlete from
AtheleteMedal)C group by C.athlete, C.edition having count1>5;
```

Click here to go to raw query result

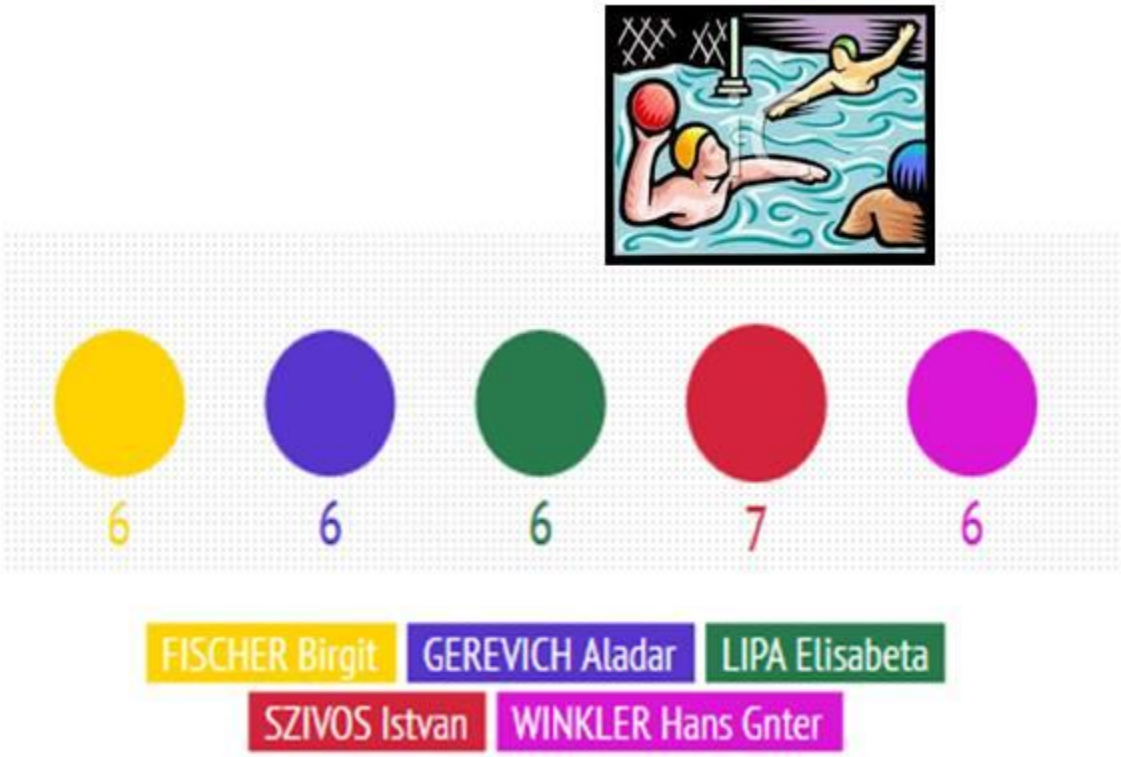


Figure 7.a: National heroes who earned medals in more than 5 different Olympic Games

b. Superstars: who have reached 10 medals

A superstar of the Olympics is those who have achieved 10 medals during their career. This query finds athletes with such honors and the country he or she served for.

```
select Athlete, Nation, Total from MultipleMedalist where Total > 10;
```

Click here to go to raw query result

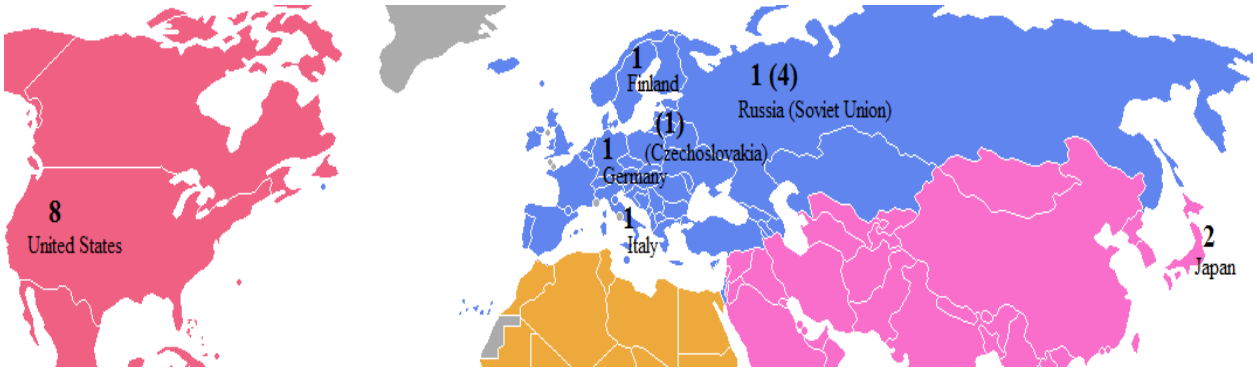


Figure 7.b: Countries with athletes who have more than 10 medals

c. Olympic Legends: superstars with more than 5 gold medals

Out of the superstars, some have reached legends of the oldest international sporting event by having more than 5 gold medals. In the Olympics, the count of gold medal is superior than the total number of medals

```
select Athlete, Nation, Gold from MultipleMedalist where Gold > 5;
```

Click here to go to raw query result

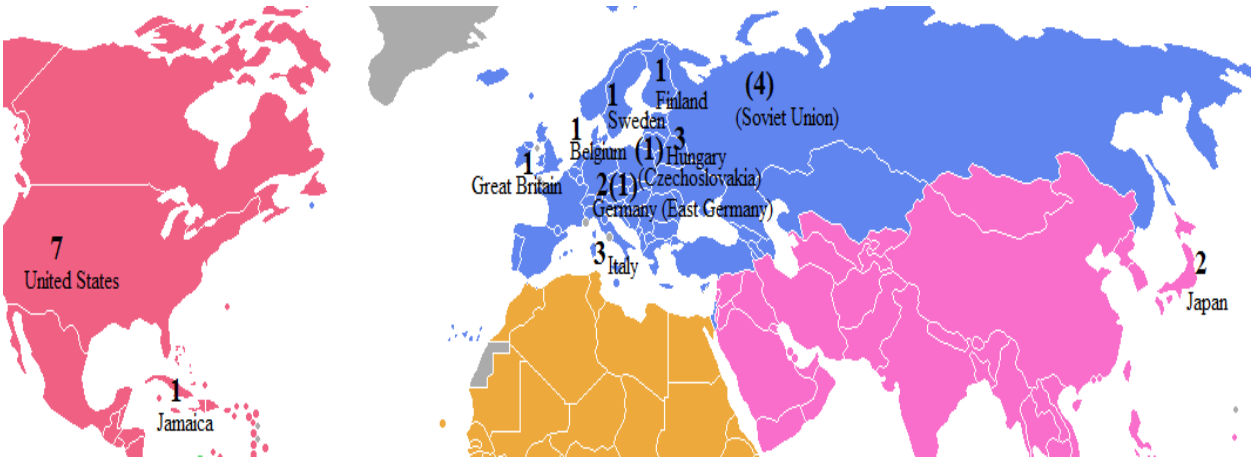


Figure 7.c: Countries with athletes who have more than 5 gold medals

8. Gender Distribution in Sport

a. Ratio of men and women and its adjustment over the entire history

In the early days, women were not allowed to participate in men activities, including sport. In this section, we analyze the distribution of men and women in the Olympics

```
select Edition,Nation,(MenCompetitor*100/TotalCompetitor) Ratio_Men,
(WomenCompetitor*100/TotalCompetitor) Ratio_Women, Sport from Participant;
```

Click here to go to raw query result

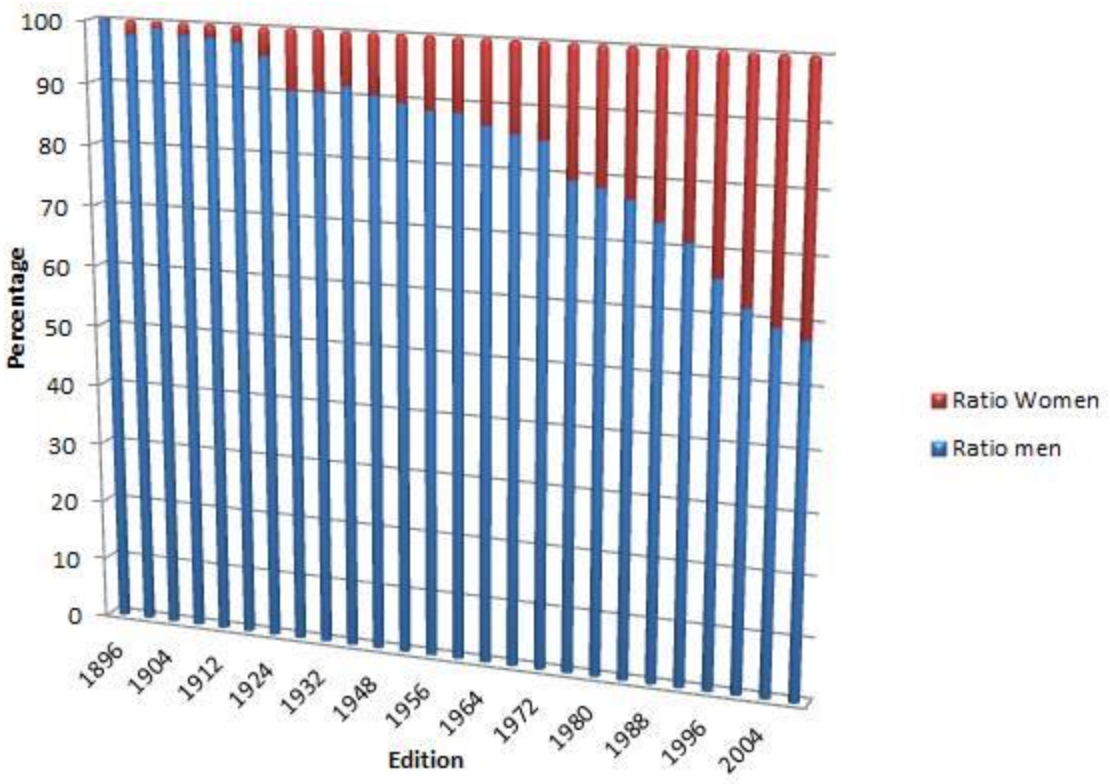


Figure 8.a: Change in ratio of men and women throughout history

b. Analyze distribution of sport event based on genders

This query show total number of Olympic event allocated for men, women, and mixed gender (denoted by M, W, X in Event_gender property respectively)

```
select Gender, count(*) from Events group by Gender;
```

Click here to go to raw query result

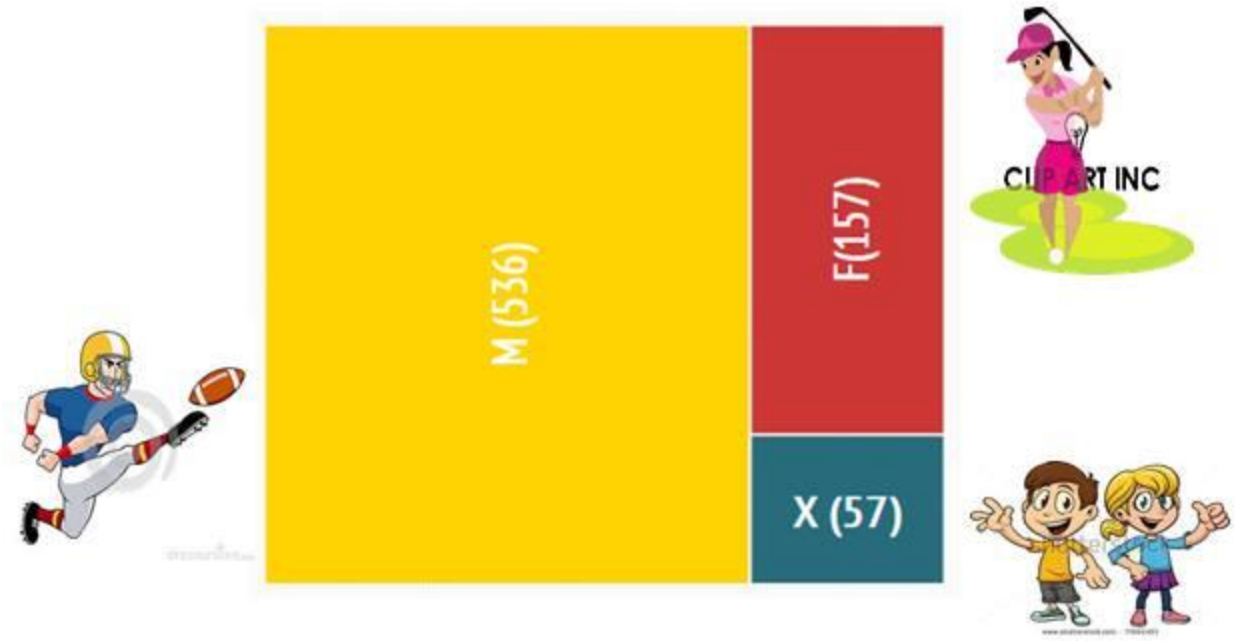
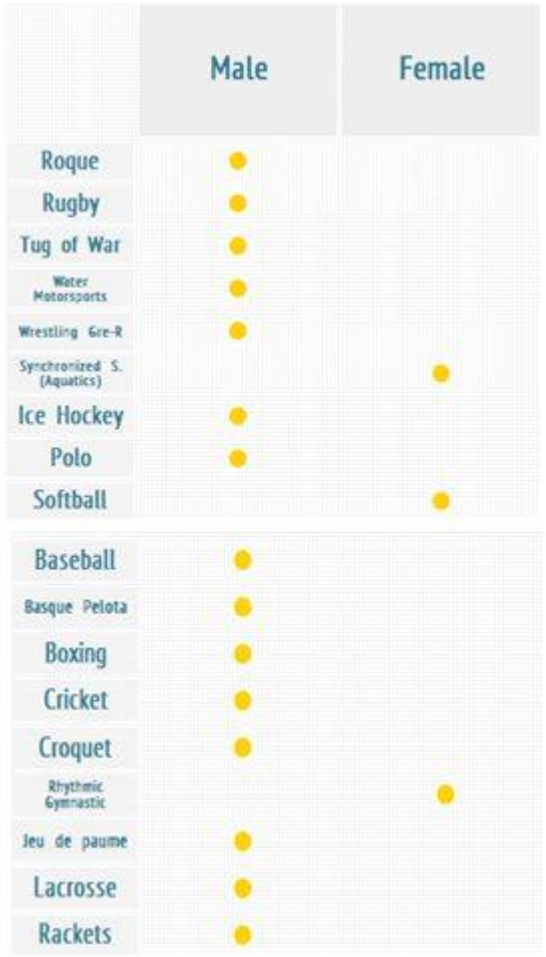


Figure 8.b: Sport events composition by gender

c. Sport disciplines and events that is exclusive for either Men or Women



List of Sport disciplines that is particular for only one gender

-- Sport disciplines that only available for men but not women

select M.Sport, M.Discipline
from (select distinct Sport, Discipline from Events where Gender='M') M
left outer join (select distinct Sport, Discipline from Events where Gender='W') W
on M.Sport = W.Sport and M.Discipline=W.Discipline where W.Sport is null;

-- Sport disciplines that only available for women but not men

select W.Sport, W.Discipline
from (select distinct Sport, Discipline from Events where Gender='M') M
right outer join (select distinct Sport, Discipline from Events where Gender='W') W
on M.Sport = W.Sport and M.Discipline=W.Discipline where M.Sport is null;

Click here to go to raw query result

Figure 8.c: Sport events composition by gender

d. Composition of mixed event (Event_gender='X') in the Olympic

Number of Olympics appearance of each Mixed sport

```
select Sport, Discipline, count(*) from
(select distinct Edition, Sport, Discipline from AtheleteMedal where Event_gender='X') T
group by Sport, Discipline;
```

Click here to go to raw query result

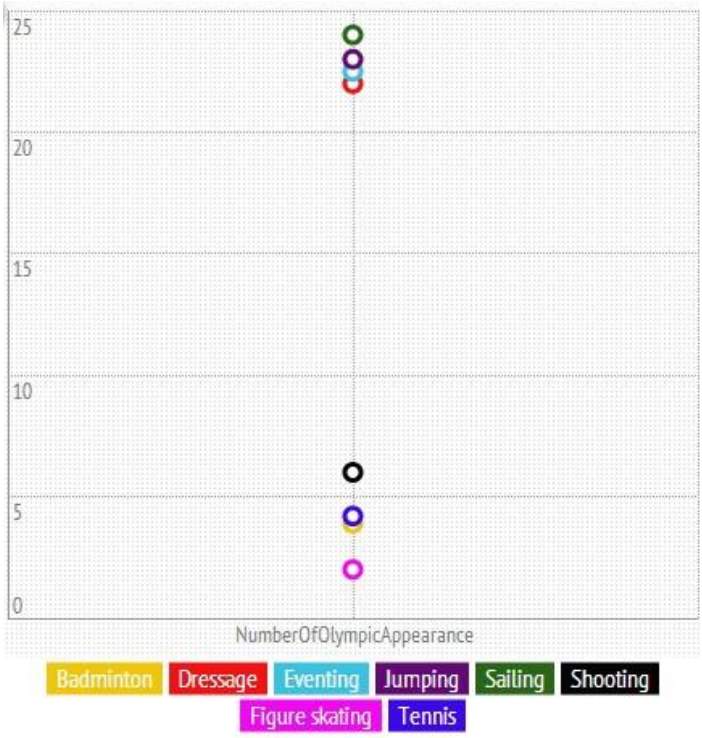


Figure 8.d: Mixed events appearance in the Olympics

9. **National strengths: each nation has their own strength and tends to perform better at some particular disciplines or events, this query show which country is dominating a particular sport in the modern Olympic Games (ever since 2000)**

This query show the dominating countries for each sport in the Olympics since 2000. The result is obtained by combination of queries: first, find the maximum medals for each sport, and find the medal allocation in each sport and match result from 2 queries to obtain the result

```
select T2.Sport, T2.Medals, CC.Country from (select Sport, max(Medals) as Medals from
(select Sport, NOC, count(*) as Medals from
(select distinct Edition, NOC,Sport,Discipline,Event,Event_gender,Medal from
AtheleteMedal where Edition>=2000) T group by Sport, NOC) T1 group by Sport) T2
join (select Sport, NOC, count(*) as Medals from(select distinct Edition,
NOC,Sport,Discipline,Event,Event_gender,Medal from AtheleteMedal
where Edition>=2000) T4 group by Sport, NOC) T3 on T3.Sport=T2.Sport and
T3.Medals=T2.Medals
join CountryCode CC on T3.NOC=CC.NOC order by T2.Sport
```

Click here to go to raw query result

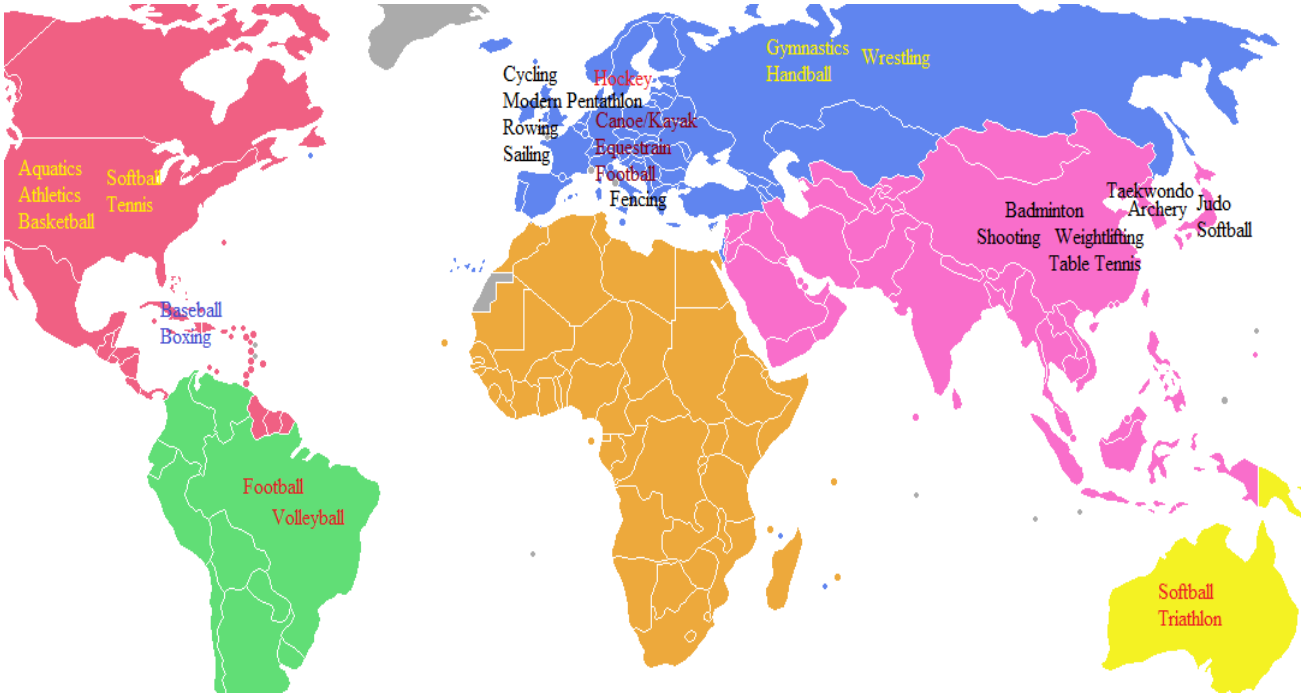


Figure 9.a: Nation strength in Olympics. Sport name is indicated on the country that dominate it.

b. Major powers of the new Olympics (top 3 rankings in Olympics since 2000)

This query gives the top 3 performance since 2000

```
select L.NOC,L.co from (select M.NOC,count(*) co from (select Edition,NOC,Medal from AtheleteMedal where edition>=2000)M group by M.NOC)L where L.co>400;
```

Click here to go to raw query result

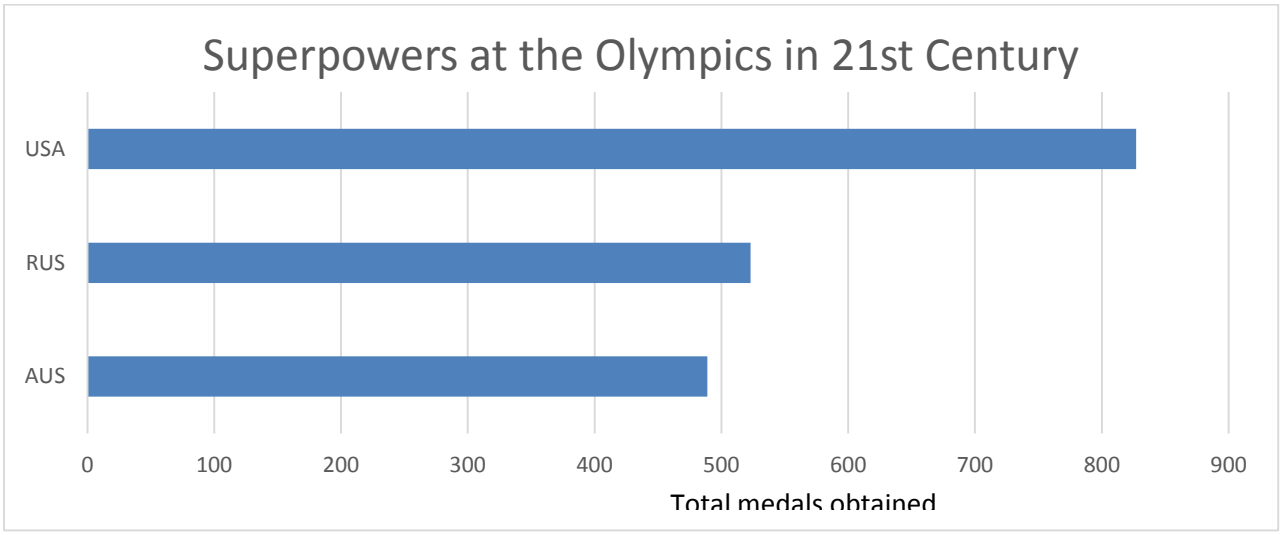


Figure 9.b: Superpowers of the Olympics in 21st century.

10/ Olympics and FIFA World Cup

The FIFA World Cup, the most globally watched, is an international men football tournament that held every 4 years (similar to the Olympics, but with 2 years offset. Ex: Olympics in 2000, World Cup in 2002, then Olympics in 2004, World Cup in 2006… and so on)

In this section, we focus on analyzing "Men Football" to see if there are any correlation between the Olympic Games (in which football teams are composed of players who are under 23) and the FIFA World Cup (no limitation on national team)

a/ The FIFA World Cup Champion

This query show the champion of football in the Olympics and the corresponding World Cup (which happens 2 years later) to see that if performance in the Olympics can be correlated to result in the World Cup

```
select AM.Edition as OlymYear, CC.Country as OlympicChamp, FF.Edition as
WorldCupChamp, FF.Champ as WorldCupChamp
from AtheleteMedal AM join CountryCode CC on AM.NOC=CC.NOC
join FIFAWorldCupResult FF on AM.Edition+2=FF.Edition
where Sport='Football' and AM.Edition >= 1928 and Gender='Men' and Medal='Gold'
group by AM.Edition,CC.Country,FF.Edition,FF.Champ;
```

Click here to go to raw query result

OlymYear	OlympicChamp	WCYear	WorldCupChamp
1928	Uruguay	1930	Uruguay
1936	Italy	1938	Italy
1948	Sweden	1950	Uruguay
1952	Hungary	1954	Germany
1956	Soviet Union	1958	Brazil
1960	Yugoslavia	1962	Brazil
1964	Hungary	1966	England
1968	Hungary	1970	Brazil
1972	Poland	1974	Germany

1976	East Germany	1978	Argentina
1980	Czechoslovakia	1982	Italy
1984	France	1986	Argentina
1988	Soviet Union	1990	Germany
1992	Spain	1994	Brazil
1996	Nigeria	1998	France
2000	Cameroon	2002	Brazil
2004	Argentina	2006	Italy
2008	Argentina	2010	Spain

Figure 10.a: Men football champion in Olympics and World Cup that happens 2 years later.

b/ Olympics and FIFA Podium

Olympic podium defined as Gold, Silver, Medal (Top 3 of competition)
FIFA podium defined as Semi-finalists (Top 4 of competition)

This query shows countries that have been in the podium (by above definitions) of the Olympics and World Cup and their frequencies. This is to see if there are some patterns between the two most prestige international sport championship

```
select coalesce(OLP.Country,WCP.Country),
coalesce(OLP.Podium,0) as OLPodium, coalesce(WCP.Podium,0) as WCPodium from
(select CC.Country as Country, count(*) as Podium from
(select distinct Edition, NOC, Medal from AtheleteMedal
where Sport='Football' and Edition >= 1928 and Gender='Men') T
join CountryCode CC on T.NOC=CC.NOC
group by CC.Country) OLP full outer join
(select coalesce(C.Country,coalesce(S.Country,coalesce(T.Country,F.Country))) as Country,
coalesce(FST,0) + coalesce(SND,0) + coalesce(TRD,0) + coalesce(FTH,0) as Podium from
(select Champ as Country, count(*) as FST from FIFAWorldCupResult group by Champ) C
full outer join (select Second as Country, count(*) as SND from FIFAWorldCupResult
group by Second) S on C.Country=S.Country
full outer join (select Third as Country, count(*) as TRD from FIFAWorldCupResult group
by Third) T on C.Country=T.Country
full outer join (select Fourth as Country, count(*) as FTH from FIFAWorldCupResult
group by Fourth) F on C.Country=F.Country) WCP
on OLP.Country=WCP.Country;
```

Click here to go to raw query result

Country	OLPodium	WCPodium	Country	OLPodium	WCPodium
Argentina	4	4	Italy	3	8
Austria	2	2	Japan	1	0
Belgium	0	1	KoreaSou	0	1
Brazil	4	10	Netherlar	0	4
Bulgaria	2	1	Nigeria	2	0
Cameroor	1	0	Norway	1	0
Chile	1	1	Paraguay	1	0
Croatia	0	1	Poland	3	2
Czechoslo	2	2	Portugal	0	2
Denmark	2	0	SovietUni	5	1
EastGerma	3	0	Spain	2	2
England	0	2	Sweden	6	4
France	1	5	Turkey	0	1
Germany	0	12	UnitedSta	0	1
Ghana	1	0	Uruguay	1	5
Hungary	5	2	Yugoslavia	5	2

Figure 10.b: Number of time a country reaches podiums in Olympics and World Cup.

APPENDIX: QUERY RESULTS & VERIFICATION

Query 1.a: Growth of Nation

Verification: look at the result and check it against the Participant table, all data match

1896	14	241
1900	24	997
1904	12	651
1908	22	2008
1912	28	2407
1920	29	2626
1924	44	3089
1928	46	2883
1932	37	1332
1936	49	3963
1948	59	4104

Result:

```
edition    nation    totalcompetitor
1896    14    241
1900    24    997
1904    12    651
1908    22    2008
1912    28    2407
1920    29    2626
1924    44    3089
1928    46    2883
1932    37    1332
1936    49    3963
1948    59    4104
1952    69    4955
1956    72    3314
1960    83    5338
1964    93    5151
1968    112    5516
1972    121    7134
1976    92    6084
1980    80    5179
1984    140    6829
1988    159    8391
1992    169    9356
1996    197    10318
2000    199    10651
2004    201    10625
2008    204    10942
Time taken: 18.091 seconds
```

Query 1.b: Growth of Athletes

Verification: look at the result and check it against the Participant table, all data match

1896	241	241	0
1900	997	975	22
1904	651	645	6
1908	2008	1971	37
1912	2407	2359	48
1920	2626	2561	65
1924	3089	2954	135
1928	2883	2606	277
1932	1332	1206	126
1936	3963	3632	331
1948	4104	3714	390

```
Total MapReduce CPU Time Spent: 1 seconds 380 msec
OK
edition     totalcompetitor mencompetitor   womencompetitor
1896        241     241     0
1900        997     975     22
1904        651     645     6
1908        2008    1971    37
1912        2407    2359    48
1920        2626    2561    65
1924        3089    2954    135
1928        2883    2606    277
1932        1332    1206    126
1936        3963    3632    331
1948        4104    3714    390
1952        4955    4436    519
1956        3314    2938    376
1960        5338    4727    611
1964        5151    4473    678
1968        5516    4735    781
1972        7134    6075    1059
1976        6084    4824    1260
1980        5179    4064    1115
1984        6829    5263    1566
1988        8391    6197    2194
1992        9356    6652    2704
1996        10318   6806    3512
2000        10651   6582    4069
2004        10625   6296    4329
2008        10942   6305    4637
Time taken: 19.075 seconds
```

Query 1.c: Growth of Sports and Disciplines

Verification: look at the result and check it against the Participant table, all data match

1896	9	10
1900	19	20
1904	16	17
1908	22	25
1912	14	18
1920	22	29
1924	17	23
1928	14	20
1932	14	20
1936	19	25
1948	17	23
1952	17	23

```
Total MapReduce CPU Time Spent: 1 seconds 460 msec
OK


edition	sport	discipline
1896	9	10
1900	19	20
1904	16	17
1908	22	25
1912	14	18
1920	22	29
1924	17	23
1928	14	20
1932	14	20
1936	19	25
1948	17	23
1952	17	23
1956	17	23
1960	17	23
1964	19	25
1968	18	24
1972	21	28
1976	21	27
1980	21	27
1984	21	29
1988	23	31
1992	25	34
1996	26	37
2000	28	40
2004	28	40
2008	28	41
```

Query 2.a: Sport Composition

Verification: check the result against the AtheleteMedal table, in the year 1896, only 5 sports took place, in which Aquatics has 4 events, Athletics has 12 events and so on, so query is correct

```
Result:
edition	sport	numofevents
1896	Aquatics	4
1896	Athletics	12
1896	Cycling	6
1896	Fencing	3
1896	Wrestling	1
1900	Aquatics	8
-------- Skip (Too many row)-------
-------- Skip (Too many row)-------
-------- Skip (Too many row)-------
```

```
2008    Taekwondo    8
2008    Tennis    4
2008    Triathlon    2
2008    Volleyball    4
2008    Weightlifting    15
2008    Wrestling    18
Time taken: 47.143 seconds
```

Query 2.b: Sport Appearance in Olympics

Verification: using Excel to filter with (year,sport) and count how many times a sport occurs in Olympics, which is exactly as the number shown next to each sport in the result (screenshot show partial filter result in Excel, and we can see that it matches our result query)

Aquatics	Diving	24
Aquatics	Swimming	26
Aquatics	Synchroni	7
Aquatics	Water pol	24
Archery	Archery	14
Athletics	Athletics	26
Badminto	Badminto	5
Baseball	Baseball	5

```
Result:
Total MapReduce CPU Time Spent: 9 seconds 140 msec
OK
sport    discipline    AppearanceInOlympics
Aquatics    Diving    24
Aquatics    Swimming    26
Aquatics    Synchronized S.    7
Aquatics    Water polo    24
Archery    Archery    14
Athletics    Athletics    26
Badminton    Badminton    5
Baseball    Baseball    5
Basketball    Basketball    17
Basque Pelota    Basque Pelota    1
Boxing    Boxing    23
Canoe / Kayak    Canoe / Kayak F    17
Canoe / Kayak    Canoe / Kayak S    6
Cricket    Cricket    1
Croquet    Croquet    1
Cycling    BMX    1
Cycling    Cycling Road    23
Cycling    Cycling Track    25
Cycling    Mountain Bike    4
Equestrian    Dressage    22
Equestrian    Eventing    22
Equestrian    Jumping    23
Equestrian    Vaulting    1
Fencing    Fencing    26
Football    Football    24
Golf    Golf    2
Gymnastics    Artistic G.    26
Gymnastics    Rhythmic G.    7
Gymnastics    Trampoline    3
Handball    Handball    11
Hockey    Hockey    21
Ice Hockey    Ice Hockey    1
Jeu de paume    Jeu de Paume    1
Judo    Judo    11
Lacrosse    Lacrosse    2
Modern Pentathlon    Modern Pentath.    22
```

```
Polo    Polo    5
Rackets    Rackets    1
Roque    Roque    1
Rowing    Rowing    25
Rugby    Rugby    4
Sailing    Sailing    24
Shooting    Shooting    24
Skating    Figure skating    2
Softball    Softball    4
Table Tennis    Table Tennis    6
Taekwondo    Taekwondo    3
Tennis    Tennis    13
Triathlon    Triathlon    3
Tug of War    Tug of War    5
Volleyball    Beach volley.    4
Volleyball    Volleyball    12
Water Motorsports    Water Motorspor    1
Weightlifting    Weightlifting    23
Wrestling    Wrestling Free.    23
Wrestling    Wrestling Gre-R    24
Time taken: 41.452 seconds
```

Query 3.a: Host distribution over continents

Verification: check the HostCountry table with Excel and verify the frequency of continents occur in this table (screenshot show part of the demonstration in Excel of host distribution, which can be counted by continent to verify)

1896	GRE	Europe
1900	FRA	Europe
1904	USA	North America
1908	GBR	Europe
1912	SWE	Europe
1920	BEL	Europe
1924	FRA	Europe
1928	NED	Europe
1932	USA	North America
1936	GER	Europe
1948	GBR	Europe
1952	FIN	Europe
1956	AUS	Ocenia
1960	ITA	Europe
1964	JPN	Asia
1968	MEX	North America

```
Result:
continent               HostingChance
Africa                  0
Asia                    3
Europe                  15
North America           6
Oceania                 2
South America           0
Time taken: 67.501 seconds
```

Query 3.b: Number of countries with medals in each continents

Verification: Look at MedalByCountry and ContryContinent table to verify that Africa has 24 countries with medals, Asia has 29, and so on

Result:
Total MapReduce CPU Time Spent: 8 seconds 140 msec
OK

continent	CountryWithMedal
Africa	24
Asia	29
Europe	43
North America	12
Oceania	4
South America	11

Time taken: 42.792 seconds

Query 4.a: Performance of host compared to the whole

Verification: check MedalByYear table to see that total medals are correctly found, and also count the medals of respective host to verify that results are correct

edition	country	medals	EntireOlympicMedal
1896	Greece	52	151
1900	France	185	512
1904	United States	394	470
1908	United Kingdom	347	804
1912	Sweden	173	885
1920	Belgium	188	1298
1924	France	122	884
1928	Netherlands	65	710
1932	United States	181	615
1936	Germany	210	875
1948	United Kingdom	56	814
1952	Finland	40	889
1956	Australia	61	885
1960	Italy	88	882
1964	Japan	64	1010
1968	Mexico	9	1031
1976	Canada	20	1305
1980	Soviet Union	442	1387
1984	United States	333	1459
1988	Korea South	77	1546
1992	Spain	66	1705
1996	United States	260	1859
2000	Australia	183	2015
2004	Greece	31	1998
2008	China	184	2042

Time taken: 68.577 seconds

Query 4.b: Multiple hosts

Verification: look at HostCountry table to see these following countries appear more than once, which is consistent to the result (screenshot shows ALL countries which host more than once,

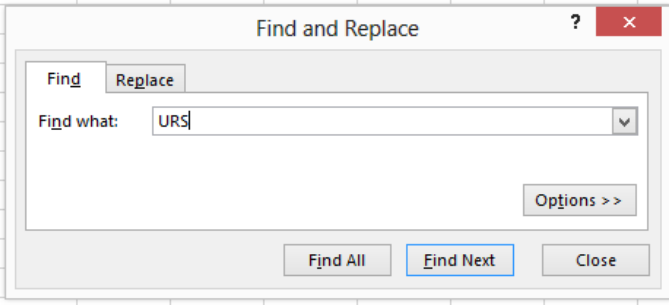
they can be counted to verify the query)

1956	Australia
2000	Australia
1900	France
1924	France
1908	Great Britain
1948	Great Britain
1936	Germany
1972	Germany
1896	Greece
2004	Greece
1904	United States
1932	United States
1984	United States
1996	United States

country
Australia
France
Germany
Greece
United Kingdom
United States
Time taken: 33.728 seconds

Query 5.a: Performance of Eastern European Soviet Bloc before and after Iron Curtain 1991
Verification: check MedalByCountry and AtheleteMedal to determine the number of medals obtained by these country before and after 1990 (screenshot showing verification of medal by Soviet Union, which is consistent with the result)

USA	United Sta	1052	2088	1195	4335
URS	Soviet Uni	584	838	627	2049
GBR	United Kir	505	498	591	1594
FRA	France	475	378	461	1314
AUS	Australia	413	293	369	1075
ITA	Italy	374	460	394	1228
HUN	Hungary	345	400	308	1053
SWE	Sweden	325	347	349	1021
NED	Netherlan	320	212	250	782
ROU	Romania	282	155	187	624
JPN	Japan	270	206	228	704
RUS	Russia	240	192	206	638
CAN	Canada	227	154	211	592
GDR	East Germ	225	329	271	825



-- Before the fall of Iron Curtain in 1990

country medals
Czechoslovakia 321
East Germany 825
Germany 520
Soviet Union 2049
Time taken: 52.356 seconds

-- After the fall of Iron Curtain in 1990

country medals
Czech Republic 41
Germany 691
Russia 638
Time taken: 36.981 seconds

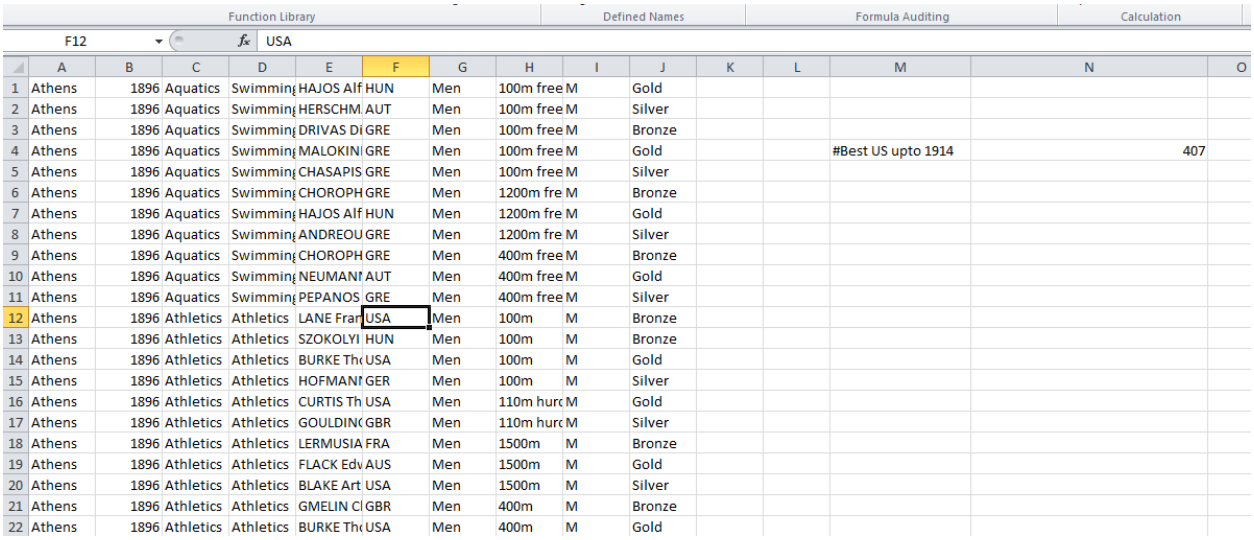
Query 5.b: Performance of questionable Olympic Games

Verification: similar to query 4a, but only look for Germany 1936, Soviet 1980 and China 2008

Edition	country	Medals	EntireOlympicMedals
1936	Germany	210	875
1980	Soviet Union	442	1387
2008	China	184	2042

Query 6.abc: Performance throughout history

Verification: look at AtheleteMedal table to verify the top 5 countries and all continents during each period of time by counting the occurrence of top 5 countries for each range of year. (screenshot below show example of checking US medal up to 1914)



Function Library | Defined Names | Formula Auditing | Calculation

F12 | USA

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Athens	1896	Aquatics	Swimming	HAJOS Alf	HUN	Men	100m free	M	Gold					
2	Athens	1896	Aquatics	Swimming	HERSCHM	AUT	Men	100m free	M	Silver					
3	Athens	1896	Aquatics	Swimming	DRIVAS Di	GRE	Men	100m free	M	Bronze					
4	Athens	1896	Aquatics	Swimming	MALOKIN	GRE	Men	100m free	M	Gold			#Best US upto 1914	407	
5	Athens	1896	Aquatics	Swimming	CHASAPIS	GRE	Men	100m free	M	Silver					
6	Athens	1896	Aquatics	Swimming	CHOROPH	GRE	Men	1200m fre	M	Bronze					
7	Athens	1896	Aquatics	Swimming	HAJOS Alf	HUN	Men	1200m fre	M	Gold					
8	Athens	1896	Aquatics	Swimming	ANDREOU	GRE	Men	1200m fre	M	Silver					
9	Athens	1896	Aquatics	Swimming	CHOROPH	GRE	Men	400m free	M	Bronze					
10	Athens	1896	Aquatics	Swimming	NEUMANI	AUT	Men	400m free	M	Gold					
11	Athens	1896	Aquatics	Swimming	PEPANOS	GRE	Men	400m free	M	Silver					
12	Athens	1896	Athletics	Athletics	LANE Fran	USA	Men	100m	M	Bronze					
13	Athens	1896	Athletics	Athletics	SZOKOLYI	HUN	Men	100m	M	Bronze					
14	Athens	1896	Athletics	Athletics	BURKE Th	USA	Men	100m	M	Gold					
15	Athens	1896	Athletics	Athletics	HOFMANI	GER	Men	100m	M	Silver					
16	Athens	1896	Athletics	Athletics	CURTIS Th	USA	Men	110m hurc	M	Gold					
17	Athens	1896	Athletics	Athletics	GOULDIN	GBR	Men	110m hurc	M	Silver					
18	Athens	1896	Athletics	Athletics	LERMUSIA	FRA	Men	1500m	M	Bronze					
19	Athens	1896	Athletics	Athletics	FLACK Edv	AUS	Men	1500m	M	Gold					
20	Athens	1896	Athletics	Athletics	BLAKE Art	USA	Men	1500m	M	Silver					
21	Athens	1896	Athletics	Athletics	GMELIN Cl	GBR	Men	400m	M	Bronze					
22	Athens	1896	Athletics	Athletics	BURKE Th	USA	Men	400m	M	Gold					

	Before the Great War (1896-1914)	During the Cold War (1945-1990)	21st Century (2000-present)
Top 5 Countries	country medalcount United States 407 United Kingdom 220 France 131 Sweden 90 Germany 71 Time taken: 109.027 seconds	country medalcount Soviet Union 1005 United States 957 East Germany 409 Hungary 281 Italy 244 Time taken: 157.868 seconds	country medalcount United States 212 Russia 164 China 163 Australia 95 Germany 90 Time taken: 132.454 seconds
Continents Medals	continent medalcount Europe 759 North America 448 Africa 8 Oceania 7 Asia 2 Time taken: 109.041 seconds	continent medalcount Europe 3513 North America 1164 Asia 456 Oceania 232 Africa 105 South America 86 Time taken: 97.973 seconds	continent medalcount Europe 886 Asia 398 North America 326 Oceania 109 Africa 75 South America 50 Time taken: 117.568 seconds

Query 7.a: National heroes, who got medals in more than 5 Olympics

Verification: Check for the listing of distinct years where an athlete won . Check forall the athelete in the result.

```
hive> select count(*) from AtheleteMedal where Athlete='FISCHER Birgit';
Total MapReduce jobs = 1
Job 0: Map: 1  Reduce: 1   Cumulative CPU: 4.32 sec   HDFS Read: 2419679 HDFS Write: 30 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 320 msec
OK
1980
1988
1992
1996
2000
2004
Time taken: 19.854 seconds
```

We can see that count=6.

athlete	count1
FISCHER Birgit	6
GEREVICH Aladar	6
LIPA Elisabeta	6
SZIVOS Istvan	7
WINKLER Hans Gnter	6

Time taken: 49.263 seconds

Query 7.b: Superstars (>10 medals)

Verification: using excel to filter AtheleteMedal to see only those whose name appear 10 times

Athlete	Nation	Total
Alexei Nemov	Russia	12
Birgit Fischer	Germany	12
Boris Shakhlin	Soviet Union	13
Carl Osburn	United States	11
Dara Torres	United States	12
Edoardo Mangiarotti	Italy	13
Jenny Thompson	United States	12
Larisa Latynina	Soviet Union	18
Mark Spitz	United States	11
Matt Biondi	United States	11
Michael Phelps	United States	22
Natalie Coughlin	United States	12
Nikolai Andrianov	Soviet Union	15
Paavo Nurmi	Finland	12
Ryan Lochte	United States	11
Sawao Kato	Japan	12
Takashi Ono	Japan	13
V?ra ?�slavsk�	Czechoslovakia	11
Viktor Chukarin	Soviet Union	11

Time taken: 11.978 seconds

Query 7.c: Legends (>5 Gold medals)

Verification: using excel to filter AtheleteMedal to see only those whose name appear 5 times with Gold medals

athlete	nation	gold
Akinori Nakayama	Japan	6
Alad�r Gerevich	Hungary	7
Amy Van Dyken	United States	6
Birgit Fischer	Germany	8
Boris Shakhlin	Soviet Union	7
Carl Lewis	United States	9
Chris Hoy	Great Britain	6
Edoardo Mangiarotti	Italy	6

Gert Fredriksson	Sweden	6
Hubert Van Innis	Belgium	6
Jenny Thompson	United States	8
Kristin Otto	East Germany	6
Larisa Latynina	Soviet Union	9
Mark Spitz	United States	9
Matt Biondi	United States	8
Michael Phelps	United States	18
Nedo Nadi	Italy	6
Nikolai Andrianov	Soviet Union	7
Paavo Nurmi	Finland	9
P�l Kov�cs	Hungary	6
Ray Ewry	United States	8
Reiner Klimke	Germany	6
Rudolf K�rp�ti	Hungary	6
Sawao Kato	Japan	8
Usain Bolt	Jamaica	6
Valentina Vezzali	Italy	6
V?ra ?�slavsk�	Czechoslovakia	7
Viktor Chukarin	Soviet Union	7
Vitaly Scherbo	Belarus	6

TimTime taken: 12.81 seconds

Query 8.a: Ratio between men and women

Verification: Choose 3 distinct years(1896,1952,2008) to check for values ratio men

```
hive> select edition,(MenCompetitor*100/TotalCompetitor) Ration_Men  from Partic
ipant where Edition=1896 or Edition=1952 or Edition=2008;
Total MapReduce jobs = 1
Launching Job 1 out of 1
Total MapReduce CPU Time Spent: 3 seconds 890 msec
OK
1896    100.0
1952    89.52573158425832
2008    57.622006945713764
Time taken: 18.295 seconds
```

1896	100
1952	89.52573158
2008	57.62200695

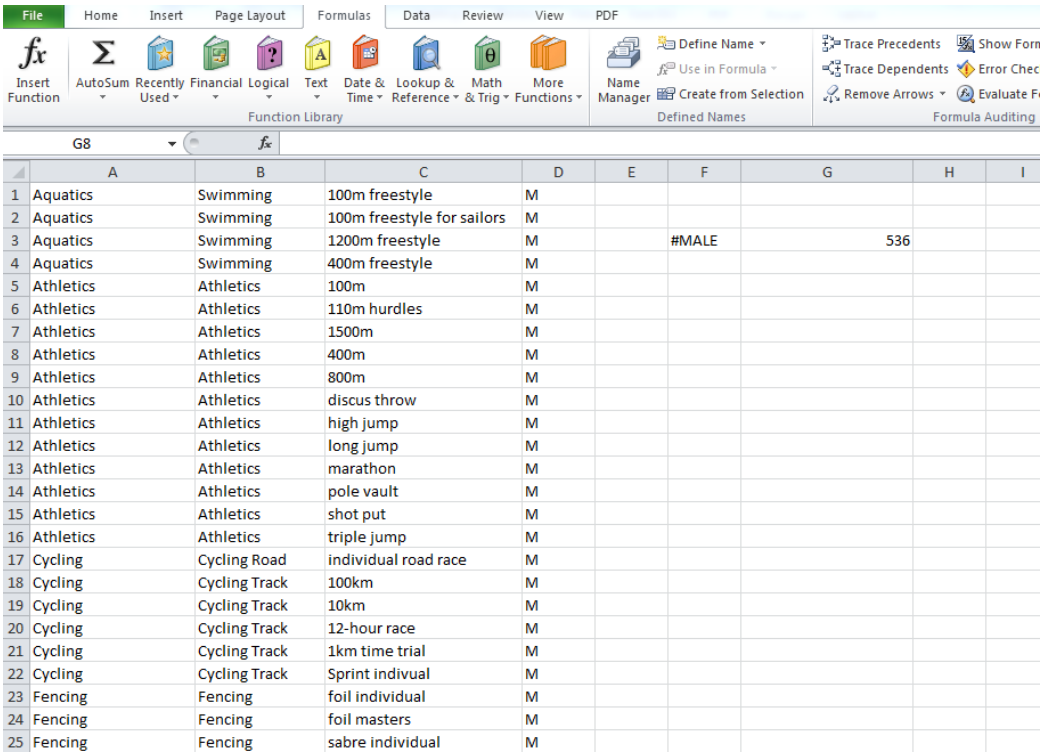
Total MapReduce CPU Time Spent: 2 seconds 470 msec
OK

edition	nation	ratio_men	ratio_women	sport
1896	14	100.0	0.0	9
1900	24	97.79338014042126	2.206619859578736	19
1904	12	99.07834101382488	0.9216589861751152	16
1908	22	98.15737051792829	1.842629482071713	22
1912	28	98.00581636892397	1.9941836310760281	14
1920	29	97.52475247524752	2.4752475247524752	22
1924	44	95.62965360958239	4.370346390417611	17
1928	46	90.39195282691641	9.608047173083593	14
1932	37	90.54054054054055	9.45945945945946	14
1936	49	91.6477416098915	8.352258390108503	19
1948	59	90.49707602339181	9.502923976608187	17
1952	69	89.52573158425832	10.474268415741674	17
1956	72	88.65419432709716	11.345805672902836	17
1960	83	88.55376545522668	11.446234544773324	17
1964	93	86.83750728013977	13.162492719860222	19
1968	112	85.84118926758521	14.158810732414793	18
1972	121	85.1555929352397	14.844407064760302	21
1976	92	79.28994082840237	20.71005917159763	21

1980	80	78.47074724850357	21.52925275149643	21
1984	140	77.06838482940401	22.931615170595986	21
1988	159	73.8529376713145	26.147062328685497	23
1992	169	71.09876015391193	28.90123984608807	25
1996	197	65.96239581314208	34.03760418685792	26
2000	199	61.79701436484837	38.20298563515163	28
2004	201	59.256470588235295	40.743529411764705	28
2008	204	57.622006945713764	42.377993054286236	28

Query 8.b: Distribution of events by gender

Verification: Counting the no of males using EXCEL COUNTIF function on Event table



	A	B	C	D	E	F	G	H	I
1	Aquatics	Swimming	100m freestyle	M					
2	Aquatics	Swimming	100m freestyle for sailors	M					
3	Aquatics	Swimming	1200m freestyle	M		#MALE	536		
4	Aquatics	Swimming	400m freestyle	M					
5	Athletics	Athletics	100m	M					
6	Athletics	Athletics	110m hurdles	M					
7	Athletics	Athletics	1500m	M					
8	Athletics	Athletics	400m	M					
9	Athletics	Athletics	800m	M					
10	Athletics	Athletics	discus throw	M					
11	Athletics	Athletics	high jump	M					
12	Athletics	Athletics	long jump	M					
13	Athletics	Athletics	marathon	M					
14	Athletics	Athletics	pole vault	M					
15	Athletics	Athletics	shot put	M					
16	Athletics	Athletics	triple jump	M					
17	Cycling	Cycling Road	individual road race	M					
18	Cycling	Cycling Track	100km	M					
19	Cycling	Cycling Track	10km	M					
20	Cycling	Cycling Track	12-hour race	M					
21	Cycling	Cycling Track	1km time trial	M					
22	Cycling	Cycling Track	Sprint indivual	M					
23	Fencing	Fencing	foil individual	M					
24	Fencing	Fencing	foil masters	M					
25	Fencing	Fencing	sabre individual	M					

Total MapReduce CPU Time Spent: 4 seconds 130 msec
OK
Gender NumberOfEvent
M 536
W 157
X 57
Time taken: 23.423 seconds

Query 8.c: Exclusive sports for different genders

Verification: using excel to filter AthleteMedal to eliminate sport events that appears in both gender (only keep gender exclusive sports) to verify the query is correct

-- Sport disciplines that only available for men but not women

Total MapReduce CPU Time Spent: 12 seconds 590 msec
OK
Sports Disciplines
Baseball Baseball
Basque Pelota Basque Pelota
Boxing Boxing
Cricket Cricket
Croquet Croquet

Ice Hockey	Ice Hockey
Jeu de paume	Jeu de Paume
Lacrosse	Lacrosse
Polo	Polo
Rackets	Rackets
Roque	Roque
Rugby	Rugby
Tug of War	Tug of War
Water Motorsports	Water Motorspor
Wrestling	Wrestling Gre-R

Time taken: 69.311 seconds

-- Sport disciplines that only available for women but not men

Total MapReduce CPU Time Spent: 11 seconds 690 msec
OK

Sports	Disciplines
Aquatics	Synchronized S.
Gymnastics	Rhythmic G.
Softball	Softball

Time taken: 62.563 seconds

Query 8.d: Appearance of mixed events

Verification: using excel to filter AtheleteMedal to verify the frequency of mixed gender event

sport	discipline	NumberOfOlympicAppearance
Badminton	Badminton	4
Equestrian	Dressage	22
Equestrian	Eventing	22
Equestrian	Jumping	23
Sailing	Sailing	24
Shooting	Shooting	6
Skating	Figure skating	2
Tennis	Tennis	4

Time taken: 48.691 seconds

Query 9.a: Dominators of each sport

Verification: using excel to filter AtheleteMedal and check against CountryCode table to see which country (NOC) has highest frequency of medals in each sport

Total MapReduce CPU Time Spent: 32 seconds 290 msec
OK

Sport	MedalMax	LeadingCountry
Aquatics	97	United States
Archery	14	Korea South
Athletics	64	United States
Badminton	21	China
Baseball	3	Cuba
Basketball	6	United States
Boxing	22	Cuba
Canoe / Kayak	25	Germany
Cycling	22	United Kingdom
Equestrian	13	Germany
Fencing	19	Italy
Football	3	Germany
Football	3	Brazil
Football	3	United States
Gymnastics	31	Russia
Handball	3	Russia
Hockey	5	Netherlands
Judo	25	Japan
Modern Pentathlon	4	United Kingdom
Rowing	13	United Kingdom

Sailing	16	United Kingdom
Shooting	25	China
Softball	3	Japan
Softball	3	Australia
Softball	3	United States
Table Tennis	22	China
Taekwondo	12	Korea South
Tennis	6	United States
Triathlon	4	Australia
Volleyball	11	Brazil
Weightlifting	24	China
Wrestling	30	Russia

Time taken: 144.786 seconds

Query 9.b: Superpowers of the 21st century (top 3)

Verification: using excel to filter AtheleteMedal and find the top 3 (which turns out to be AUS, RUS and USA) in the 21st century

Total MapReduce CPU Time Spent: 6 seconds 60 msec
OK

noc	count
AUS	489
RUS	523
USA	827

Query 10.a: Champion of Olympic and World Cup men football

Verification: check AtheleteMedal and FIFAWorldCupResult tables to see the country that win the tournament for men football in respective year, which is same as the result in this table (this screenshot shows partial FIFAWorldCup table for champion, which verify part of query result

1930	Uruguay
1934	Italy
1938	Italy
1950	Uruguay
1954	Germany
1958	Brazil
1962	Brazil
1966	England
1970	Brazil
1974	Germany

Total MapReduce CPU Time Spent: 19 seconds 80 msec
OK

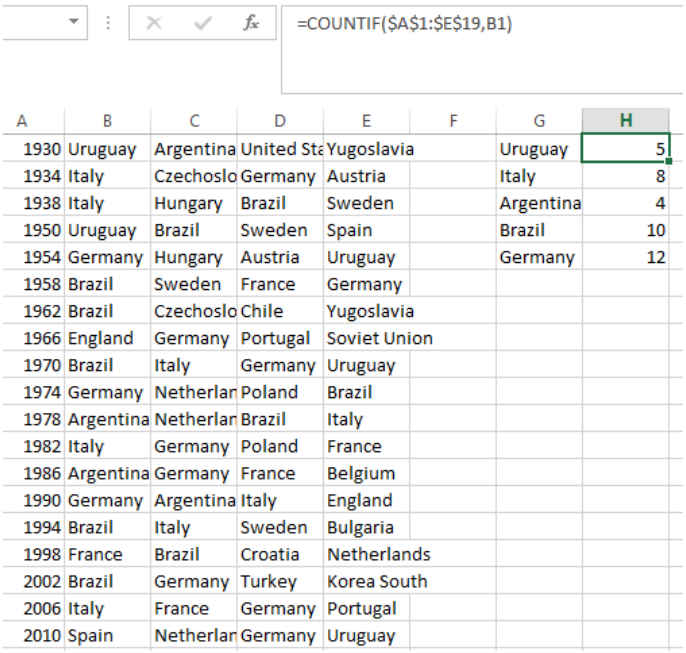
OlymYear	OlympicChamp	WCYear	WorldCupChamp
1928	Uruguay	1930	Uruguay
1936	Italy	1938	Italy
1948	Sweden	1950	Uruguay
1952	Hungary	1954	Germany
1956	Soviet Union	1958	Brazil
1960	Yugoslavia	1962	Brazil
1964	Hungary	1966	England
1968	Hungary	1970	Brazil
1972	Poland	1974	Germany
1976	East Germany	1978	Argentina
1980	Czechoslovakia	1982	Italy
1984	France	1986	Argentina
1988	Soviet Union	1990	Germany

1992	Spain	1994	Brazil
1996	Nigeria	1998	France
2000	Cameroon	2002	Brazil
2004	Argentina	2006	Italy
2008	Argentina	2010	Spain

Time taken: 64.735 seconds

Query 10.b: Frequency of Podium in Olympics and World Cup for men football

Verification: using excel to filter AtheleteMedal and FIFAWorldCupResult to verify the frequency (for men football) of that country in the table (this screenshot partially show the FIFAWorldCup podium result)



A	B	C	D	E	F	G	H
1930	Uruguay	Argentina	United Sta	Yugoslavia		Uruguay	5
1934	Italy	Czechoslo	Germany	Austria		Italy	8
1938	Italy	Hungary	Brazil	Sweden		Argentina	4
1950	Uruguay	Brazil	Sweden	Spain		Brazil	10
1954	Germany	Hungary	Austria	Uruguay		Germany	12
1958	Brazil	Sweden	France	Germany			
1962	Brazil	Czechoslo	Chile	Yugoslavia			
1966	England	Germany	Portugal	Soviet Union			
1970	Brazil	Italy	Germany	Uruguay			
1974	Germany	Netherlan	Poland	Brazil			
1978	Argentina	Netherlan	Brazil	Italy			
1982	Italy	Germany	Poland	France			
1986	Argentina	Germany	France	Belgium			
1990	Germany	Argentina	Italy	England			
1994	Brazil	Italy	Sweden	Bulgaria			
1998	France	Brazil	Croatia	Netherlands			
2002	Brazil	Germany	Turkey	Korea South			
2006	Italy	France	Germany	Portugal			
2010	Spain	Netherlan	Germany	Uruguay			

Total MapReduce CPU Time Spent: 44 seconds 800 msec
OK

Country	OLPodium	WCPodium
Argentina	4	4
Austria	1	1
Austria	1	1
Belgium	0	1
Brazil	4	10
Bulgaria	2	1
Cameroon	1	0
Chile	1	1
Croatia	0	1
Czechoslovakia	2	2
Denmark	2	0
East Germany	3	0
England	0	2
France	1	5
Germany	0	12
Ghana	1	0
Hungary	5	2
Italy	3	8
Japan	1	0
Korea South	0	1
Netherlands	0	3
Netherlands	0	1
Nigeria	2	0
Norway	1	0

Paraguay	1	0
Poland	3	2
Portugal	0	1
Portugal	0	1
Soviet Union	5	1
Spain	2	2
Sweden	2	1
Sweden	2	1
Sweden	2	2
Turkey	0	1
United States	0	1
Uruguay	1	5
Yugoslavia	5	2

Time taken: 176.582 seconds