**Assignment 7.2**

**Associated Data Files**

This Data set is about Olympics. You can download the data set from the below link:

[https://drive.google.com/open?id=0ByJLBTmJojjzV1czX3Nha0R3bTQ](https://drive.google.com/open?id=0ByJLBTmJojjzV1czX3Nha0R3bTQ%20)

olympix\_data.csv

**Michael Phelps 23 United States 2008 08-24-08 Swimming 8 0 0 8**

**Michael Phelps 19 United States 2004 08-29-04 Swimming 6 0 2 8**

**Michael Phelps 27 United States 2012 08-12-12 Swimming 4 2 0 6**

**Natalie Coughlin 25 United States 2008 08-24-08 Swimming 1 2 3 6**

**Aleksey Nemov 24 Russia 2000 10-01-00 Gymnastics 2 1 3 6**

**Alicia Coutts 24 Australia 2012 08-12-12 Swimming 1 3 1 5**

**Missy Franklin 17 United States 2012 08-12-12 Swimming 4 0 1 5**

**Ryan Lochte 27 United States 2012 08-12-12 Swimming 2 2 1 5**

**Allison Schmitt 22 United States 2012 08-12-12 Swimming 3 1 1 5**

**Natalie Coughlin 21 United States 2004 08-29-04 Swimming 2 2 1 5**

DATE SET DESCRIPTION

The data set consists of the following fields.

**Athlete**: This field consists of the athlete name

**Age**: This field consists of athlete ages

**Country**: This fields consists of the country names which participated in Olympics

**Year**: This field consists of the year

**Closing Date**: This field consists of the closing date of ceremony

**Sport**: Consists of the sports name

**Gold Medals**: No. of Gold medals

**Silver Medals**: No. of Silver medals

**Bronze Medals**: No. of Bronze medals

**Total Medals**: Consists of total no. of medals

**Problem Statement:**

**1. Write a Hive program to find the number of medals won by each country in swimming.**

**2. Write a Hive program to find the number of medals that India won year wise.**

**3. Write a Hive Program to find the total number of medals each country won.**

**4. Write a Hive program to find the number of gold medals each country won.**

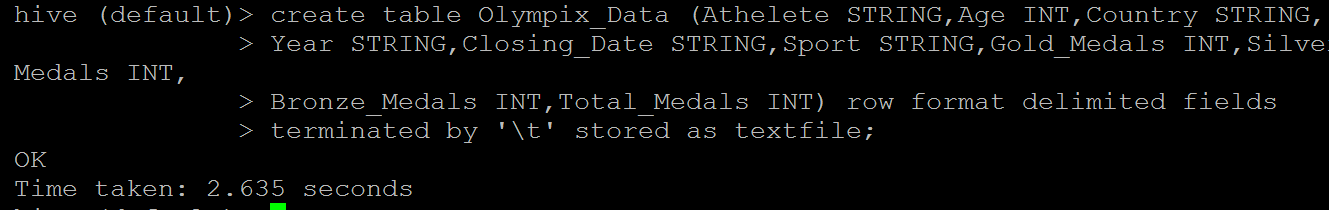
**Table creation and data loading**

create table Olympix\_Data (Athelete STRING,Age INT,Country STRING,

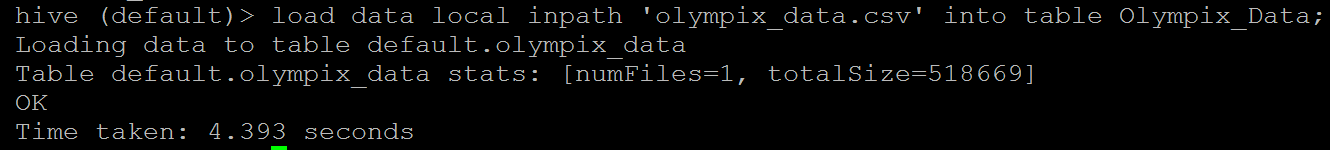
Year STRING,Closing\_Date STRING,Sport STRING,Gold\_Medals INT,Silver\_Medals INT,

Bronze\_Medals INT,Total\_Medals INT) row format delimited fields

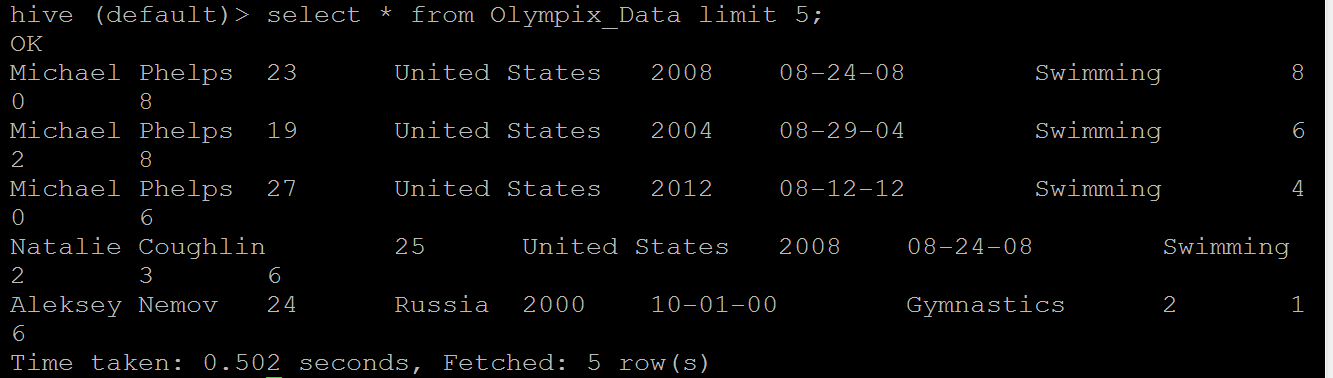
terminated by '\t' stored as textfile;



load data local inpath 'olympix\_data.csv' into table Olympix\_Data;



select \* from Olympix\_Data limit 5;



**1. Write a Hive program to find the number of medals won by each country in swimming.**

select Country,SUM(Total\_Medals) from Olympix\_Data where Sport = 'Swimming' GROUP BY Country;

Argentina 1

Australia 163

Austria 3

Belarus 2

Brazil 8

Canada 5

China 35

Costa Rica 2

Croatia 1

Denmark 1

France 39

Germany 32

Great Britain 11

Hungary 9

Italy 16

Japan 43

Lithuania 1

Netherlands 46

Norway 2

Poland 3

Romania 6

Russia 20

Serbia 1

Slovakia 2

Slovenia 1

South Africa 11

South Korea 4

Spain 3

Sweden 9

Trinidad and Tobago 1

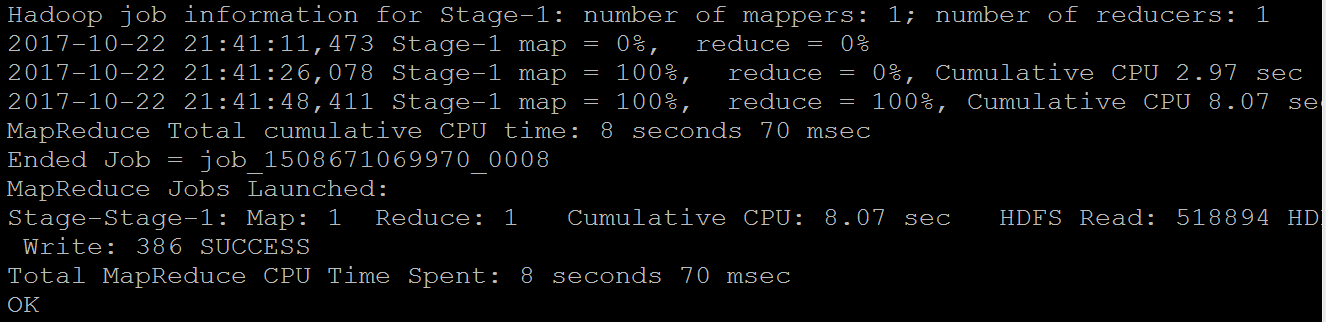
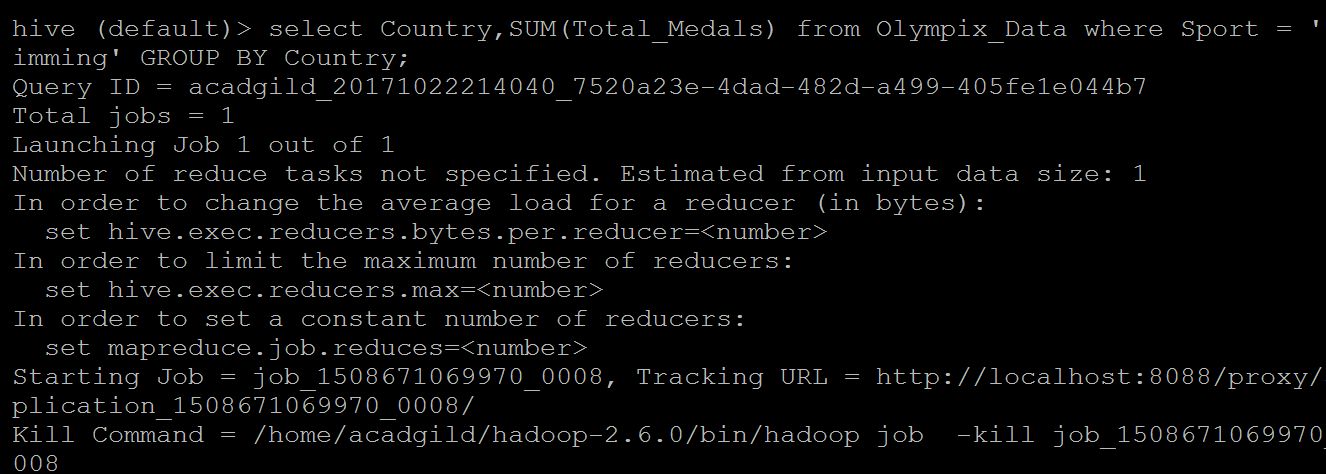
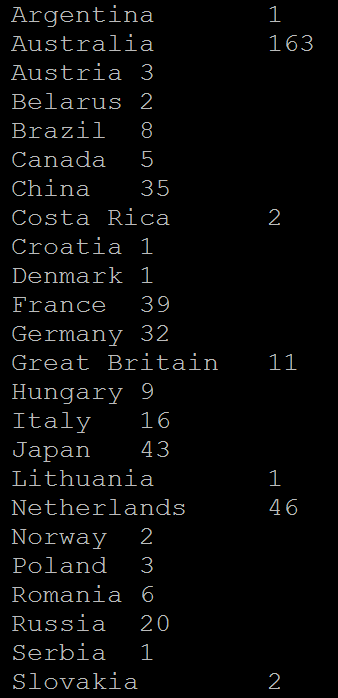
Tunisia 3

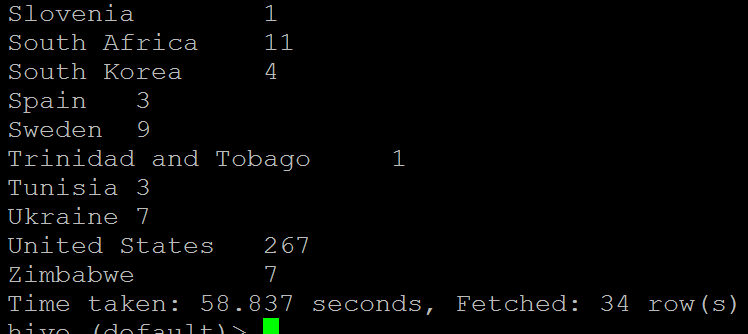
Ukraine 7

United States 267

Zimbabwe 7

Time taken: 58.837 seconds, Fetched: 34 row(s)



**2. Write a Hive program to find the number of medals that India won year wise**

select Year,SUM(Total\_Medals) from Olympix\_Data where Country = 'India' GROUP BY Year;

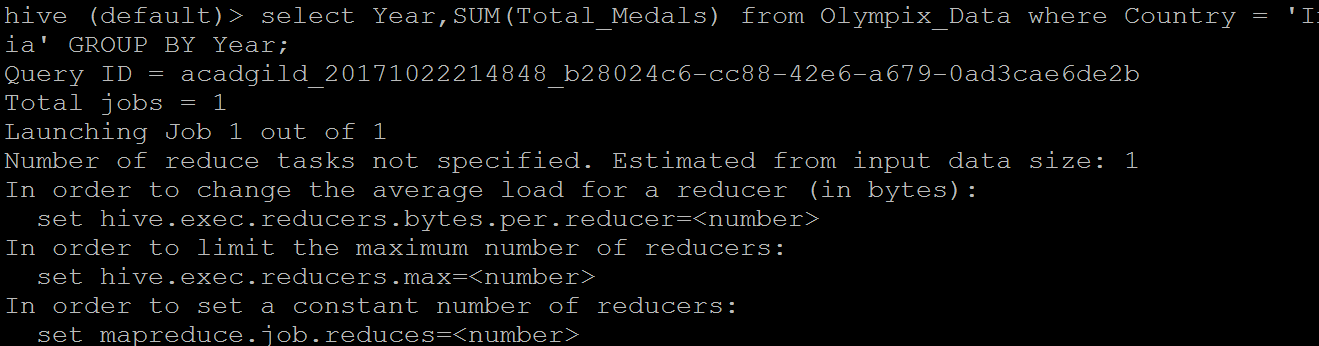
2000 1

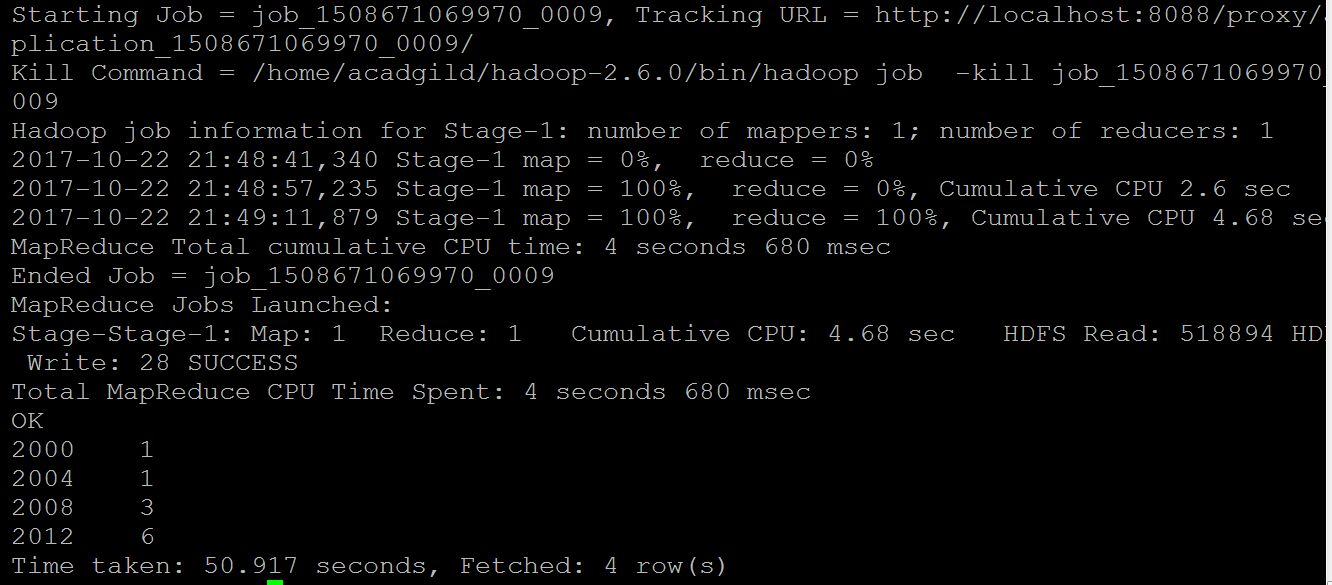
2004 1

2008 3

2012 6

Time taken: 50.917 seconds, Fetched: 4 row(s)





**3. Write a Hive Program to find the total number of medals each country won.**

select Country,SUM(Total\_Medals) from Olympix\_Data

GROUP BY Country;

Afghanistan 2

Algeria 8

Argentina 141

Armenia 10

Australia 609

Austria 91

Azerbaijan 25

Bahamas 24

Bahrain 1

Barbados 1

Belarus 97

Belgium 18

Botswana 1

Brazil 221

Bulgaria 41

Cameroon 20

Canada 370

Chile 22

China 530

Chinese Taipei 20

Colombia 13

Costa Rica 2

Croatia 81

Cuba 188

Cyprus 1

Czech Republic 81

Denmark 89

Dominican Republic 5

Ecuador 1

Egypt 8

Eritrea 1

Estonia 18

Ethiopia 29

Finland 118

France 318

Gabon 1

Georgia 23

Germany 629

Great Britain 322

Greece 59

Grenada 1

Guatemala 1

Hong Kong 3

Hungary 145

Iceland 15

India 11

Indonesia 22

Iran 24

Ireland 9

Israel 4

Italy 331

Jamaica 80

Japan 282

Kazakhstan 42

Kenya 39

Kuwait 2

Kyrgyzstan 3

Latvia 17

Lithuania 30

Macedonia 1

Malaysia 3

Mauritius 1

Mexico 38

Moldova 5

Mongolia 10

Montenegro 14

Morocco 11

Mozambique 1

Netherlands 318

New Zealand 52

Nigeria 39

North Korea 21

Norway 192

Panama 1

Paraguay 17

Poland 80

Portugal 9

Puerto Rico 2

Qatar 3

Romania 123

Russia 768

Saudi Arabia 6

Serbia 31

Serbia and Montenegro 38

Singapore 7

Slovakia 35

Slovenia 25

South Africa 25

South Korea 308

Spain 205

Sri Lanka 1

Sudan 1

Sweden 181

Switzerland 93

Syria 1

Tajikistan 3

Thailand 18

Togo 1

Trinidad and Tobago 19

Tunisia 4

Turkey 28

Uganda 1

Ukraine 143

United Arab Emirates 1

United States 1312

Uruguay 1

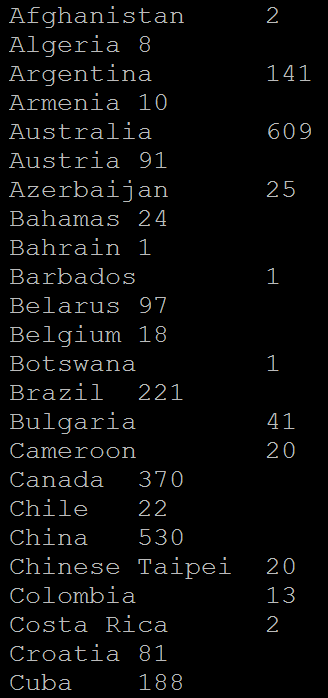
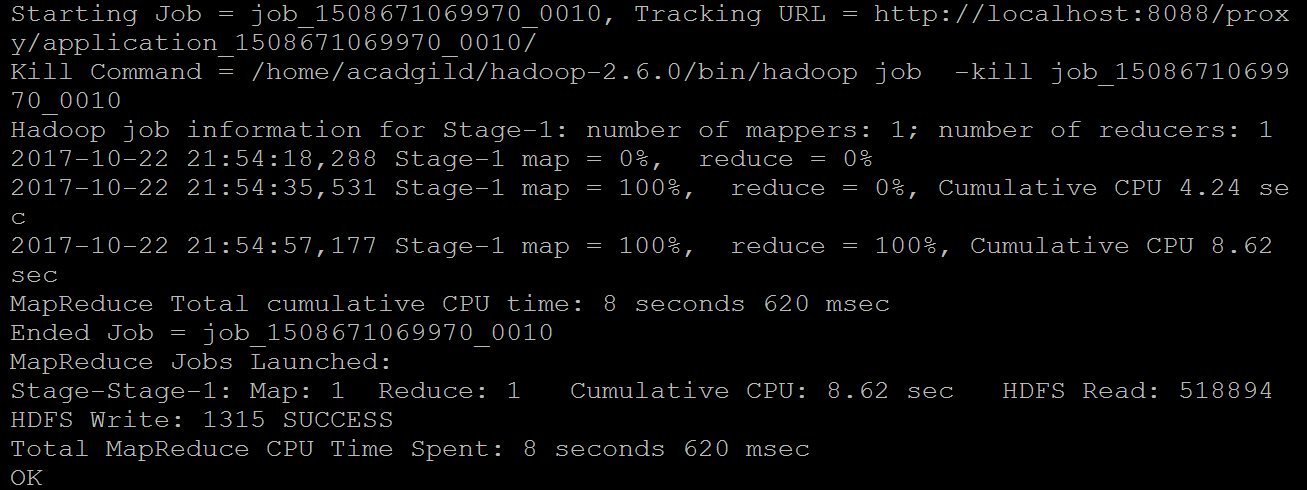
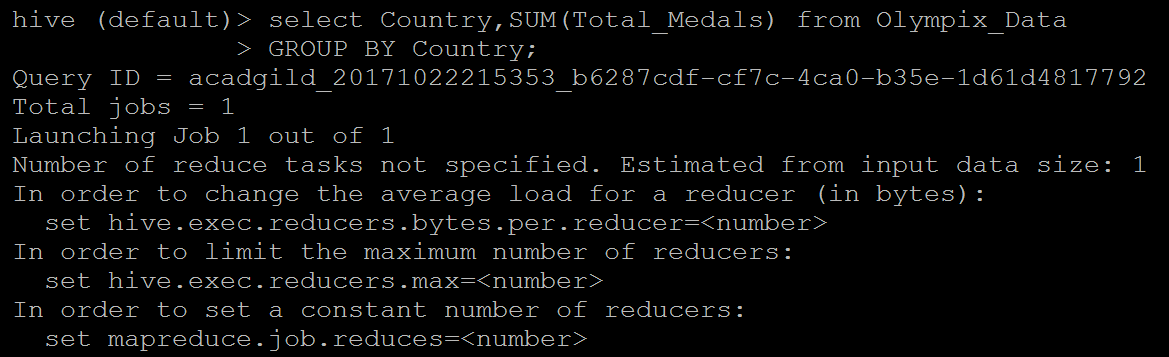
Uzbekistan 19

Venezuela 4

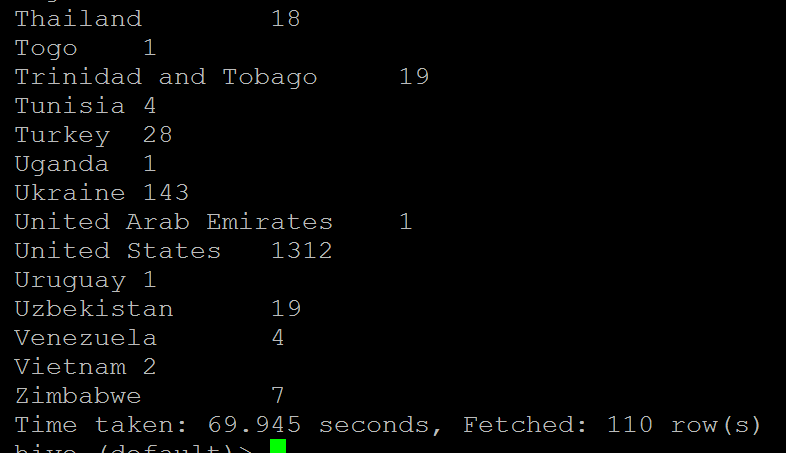
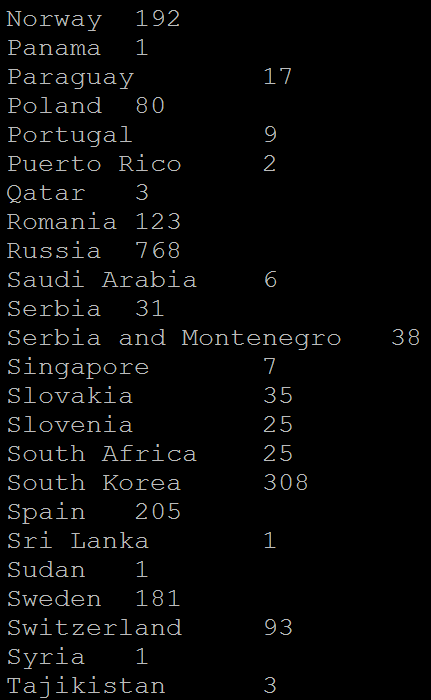
Vietnam 2

Zimbabwe 7

Time taken: 69.945 seconds, Fetched: 110 row(s)

****





**4. Write a Hive program to find the number of gold medals each country won.**

select Country,SUM(Gold\_Medals) from Olympix\_Data GROUP BY Country;

Afghanistan 0

Algeria 2

Argentina 49

Armenia 0

Australia 163

Austria 36

Azerbaijan 6

Bahamas 11

Bahrain 0

Barbados 0

Belarus 17

Belgium 2

Botswana 0

Brazil 46

Bulgaria 8

Cameroon 20

Canada 168

Chile 3

China 234

Chinese Taipei 2

Colombia 2

Costa Rica 0

Croatia 35

Cuba 57

Cyprus 0

Czech Republic 14

Denmark 46

Dominican Republic 3

Ecuador 0

Egypt 1

Eritrea 0

Estonia 6

Ethiopia 13

Finland 11

France 108

Gabon 0

Georgia 6

Germany 223

Great Britain 124

Greece 12

Grenada 1

Guatemala 0

Hong Kong 0

Hungary 77

Iceland 0

India 1

Indonesia 5

Iran 10

Ireland 1

Israel 1

Italy 86

Jamaica 24

Japan 57

Kazakhstan 13

Kenya 11

Kuwait 0

Kyrgyzstan 0

Latvia 3

Lithuania 5

Macedonia 0

Malaysia 0

Mauritius 0

Mexico 19

Moldova 0

Mongolia 2

Montenegro 0

Morocco 2

Mozambique 1

Netherlands 101

New Zealand 18

Nigeria 6

North Korea 6

Norway 97

Panama 1

Paraguay 0

Poland 20

Portugal 1

Puerto Rico 0

Qatar 0

Romania 57

Russia 234

Saudi Arabia 0

Serbia 1

Serbia and Montenegro 11

Singapore 0

Slovakia 10

Slovenia 5

South Africa 10

South Korea 110

Spain 19

Sri Lanka 0

Sudan 0

Sweden 57

Switzerland 21

Syria 0

Tajikistan 0

Thailand 6

Togo 0

Trinidad and Tobago 1

Tunisia 2

Turkey 9

Uganda 1

Ukraine 31

United Arab Emirates 1

United States 552

Uruguay 0

Uzbekistan 5

Venezuela 1

Vietnam 0

Zimbabwe 2

Time taken: 50.962 seconds, Fetched: 110 row(s)

