

A practical guide to scientific computing: Final Exam

NAME

Clone the git repository `world-cup.git`

```
git clone https://github.com/openfootball/world-cup.git
```

Now the directory `world-cup` contains one directory for each edition of the world cup. Inside each directory, you will find a `cup.txt` and `cup_finals.txt`. In these files, along with other information, you find all the matches that were played and the final scores. Note that you can find slightly different formats, for example

```
(28) 23 June   Italy      1-1   Cameroon  @ Balaidos, Vigo
(50) 8 July    West Germany 5-4pen 3-3aet 1-1 France @ Estadio Ramon Sanchez Pizjuan, Seville
(1) 4 June     Switzerland 1-1aet (1-1, 1-1) Germany @ Parc des Princes, Paris
(10) 19 June   Brazil      1-1 a.e.t. 1-1 Yugoslavia @ Stade Olympique de la Pontaise, Lausanne
(6) 17 June    England     4-4 a.e.t. 3-3 Belgium @ St. Jakob Stadium, Basel
```

and that the final score is the first that occurs (the others—which are optional—are referring to the status at the end of the regular time, and at the end of the overtime).

1. Four countries (five if we equate West Germany and Germany) hosted the world cup twice. Write a program that counts how many time each country hosted the cup, and list below those with two occurrences:

The countries that hosted the world cup twice are...

Brazil
France
Italy
Mexico

Suggestions:

- This can be done either in the terminal, in python or in R.
 - In python, `os.listdir('.')` lists all the directories in the current directory. In R, `dir()` lists all files and directories in the current directory.
2. The files `cup.yum` contain the list of the teams for each edition. These are coded, such that ita is Italy, bra is Brazil, fra is France and arg is Argentina. How many times did these teams participate?
 - Argentina: 16
 - Brazil: 20
 - France: 14
 - Italy: 18

Suggestions:

- `ls */cup.yum` lists all the files called `cup.yum` that are present in the sub-directories.
 - This is definitely a job for `grep`!
3. Tonight, Italy plays Costa Rica, Switzerland plays France, and Honduras plays Ecuador. How many times did these teams play each other in the past editions of the world cup?

Teams	Number of games played before
Costa Rica & Italy	0
Ecuador & Honduras	0
France & Switzerland	1

4. What is the most common match in the world cup? Count how many times each pair of countries appears, and write here the maximum along with the names of the two teams.

The most common match of the world cup is...
Brazil - Sweden (7 times)

5. What is the maximum number of goals ever reached (sum those scored by both teams)? Write here the teams and the total number of goals.

The match with the highest number of goals is...
Austria - Switzerland 7-5 (1954)

Suggestions for 3, 4, and 5:

- Concatenate all the files `cup.txt` and `cup_finals.txt` (using a mix of `cat` and `redirect`, or `find`, `grep` and `redirect`) into a single file.
- Use `python` with regular expressions to extract the matches from the file. Use `groups` or `split` the string in order to isolate (team a), (score team a), (score team b), (team b)
- Point 3 and 4 can be done together (while you search for these teams, you can count all the others as well). Sort the teams in alphabetical order to avoid finding that France-Switzerland is different from Switzerland-France