Exercise Sheet 5

Problem 1. Data aggregation
Consider the `WM_teams_2014.csv` data set that contains information on each player of the World Cup 2014, e.g., age, club, country, caps (number of plays for the national team).

(a) Use the `aggregate` function to compute the mean and median age of the players stratified on country and position

(b) Create a data set that contains the clubs and number of players in each club that participate in the world cup. The data set should be ordered from highest to lowest

Problem 2. Simulation of confidence intervals
The interpretation of a confidence interval at the 95% level is that, if an experiment were repeated a hundred times on multiple samples, the calculated confidence interval for each sample would encompass the true parameter 95% of the time. We can check how this works in practice in a simulation study.

(a) The `t.test()` function returns a list-like result. use `str()` and extract the one list component with the `conf.int`. To see how to do this, simply look at t.test(X) for a simple sample such as X <- 1:20. For the interval of this one sample, you should get a numeric vector of length 2 (of the two interval end points)

(b) Write a function that takes as argument a loop counter `i` and the sample size n and

- create a vector of *n* random variables normally distributed with mean 0 and standard deviation 1
- Compute a t-test
- returns whether the confidence interval contains the true mean value (in this case 0).

You can take inspiration in the `check_level` function defined in the lecture at p. 78

(c) Run the simulation 1000 times and check the assumption