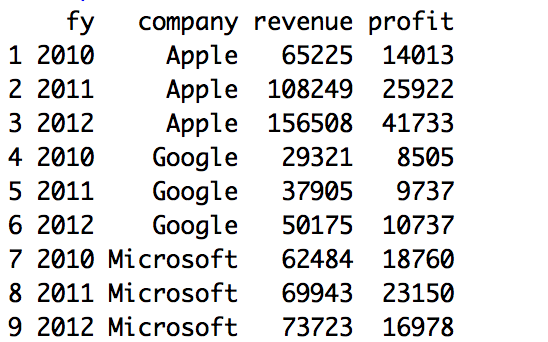
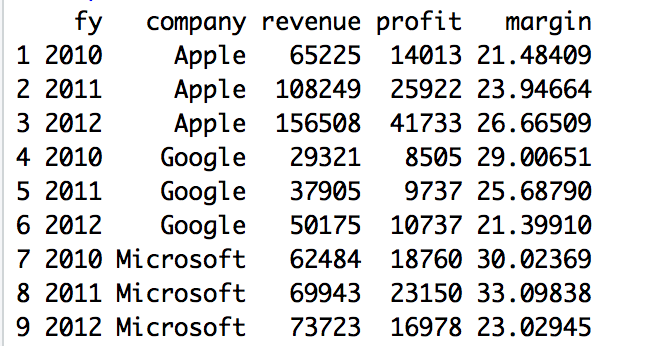
# INTRODUCTION TO DATA SCIENCE

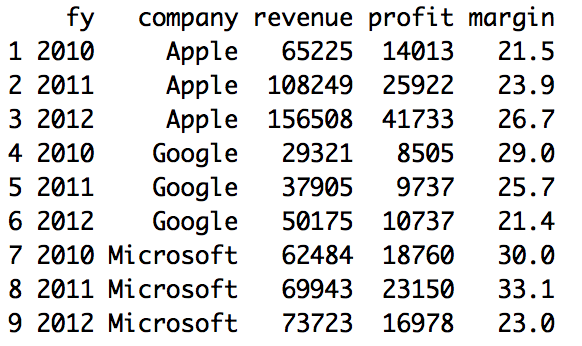
# Exercise.1: Create a data frame in R for data about IT companies. Use the information from the following table to create your data frame. Call your data frame as companiesData:



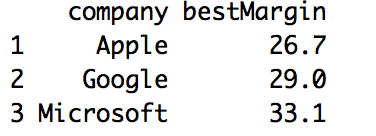
# Exercise.2: Add a column for marginal profit called >margin” by dividing profit by revenue and then multiplying by 100. The resulted new data frame looks like this:



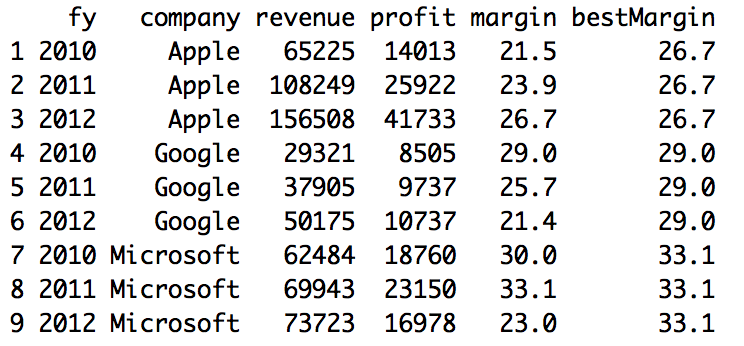
**Exercise.3:  Now round the margin column to one decimal place. The resulted data frame looks like this:**

****

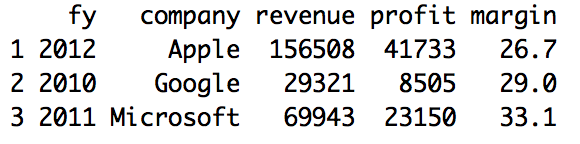
**Exercise.4:  Calculate the best profit margin for each company (i.e. max margin). The result is like this:**

****

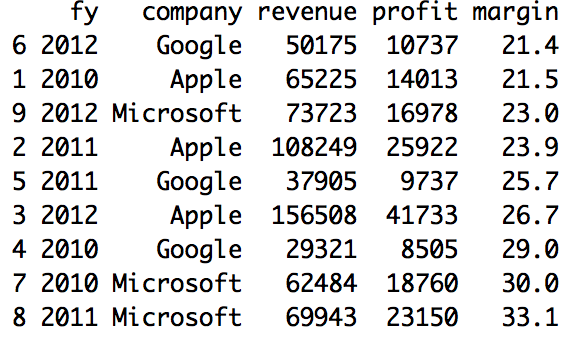
**You need to add this column to the original data frame (companiesData). The resulted new data frame looks like this:**

****

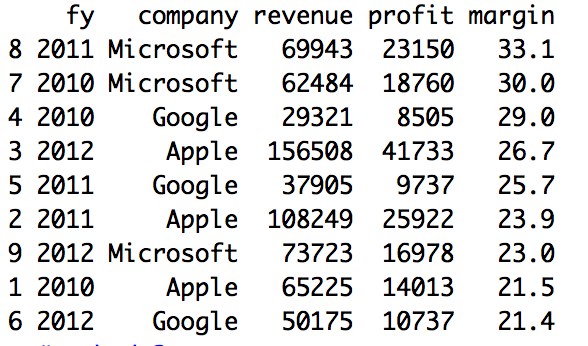
**Exercise.5: Exercise 4 list the best margin profit repeated for different years. What if we want to know which year each company has the maximum profit. The resulted new data frame looks like that:**

****

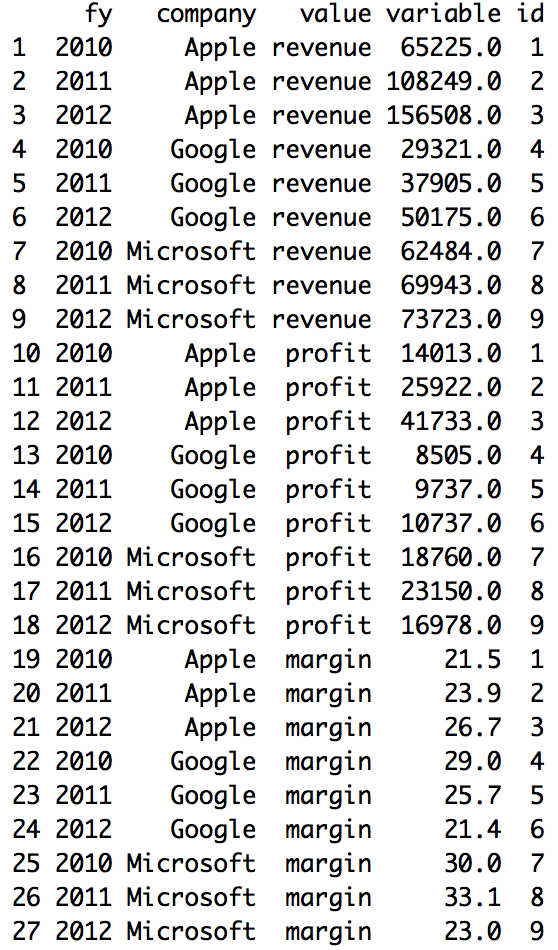
**Exercise.6: Sort companies according to their margin of profit in an ascending order. The resulted file looks like this:**



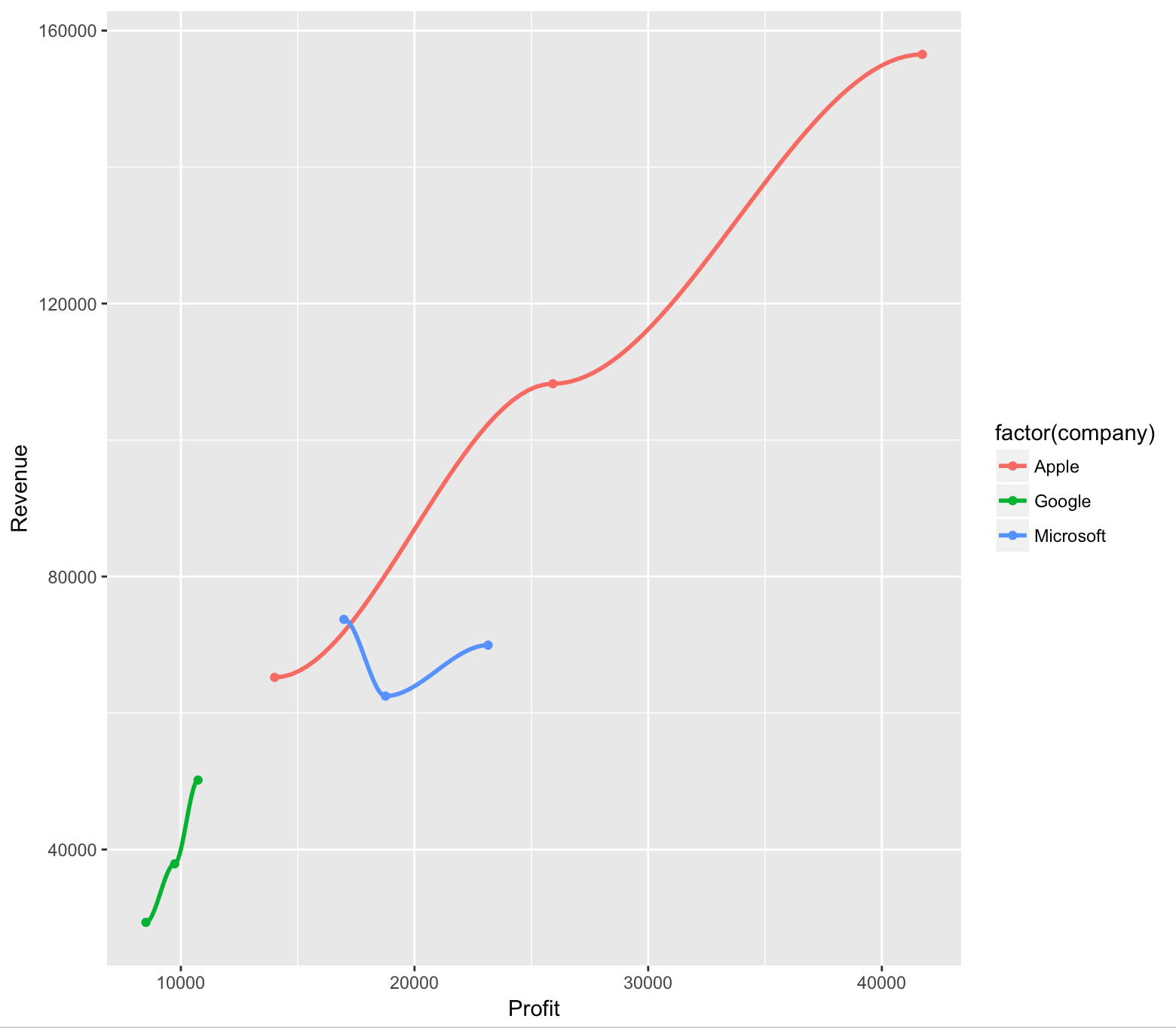
**Exercise.7: Sort companies according to their margin of profit in an descending order. The resulted file looks like this**

****

**Exercise.8: You are required to reshape the data frame so information appear like that:**

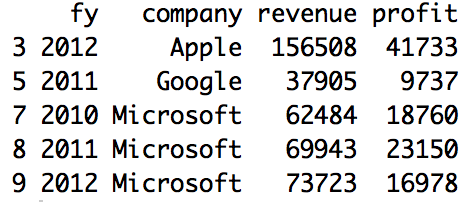
****

**Exercise.9: Use ggplot2 library and plot the revenue vs profit for each of the companies in the ordered ascending data frame. The resulted graph looks like this:**



**Exercise.10:**  **Draw another observation on the company’s data either by performing more transformations on the data frame or using graphs to show the observation.**

**1.Here I am using the rows\_to\_keep to remove some rows from the orginal data frame**

****

**2.Here I am using ddply function from the plyr package**

