Q1. How many genes are in this dataset?
38694
Q2. How many 'control' cell lines do we have?
4
Q3. How would you make the above code in either approach more robust?

I would increase the sample size in order to make the results more accurate.

Q4. Follow the same procedure for the treated samples (i.e. calculate the mean per gene across drug treated samples and assign to a labeled vector called treated.mean)

658.00 0.00 546.00 316.50 78.75 0.00
Q5 (b). You could also use the **ggplot2** package to make this figure producing the plot below. What **geom_?()** function would you use for this plot?
Geom_point
Q6. Try plotting both axes on a log scale. What is the argument to **plot()** that allows you to do this?
log = "xy"
Q7. What is the purpose of the arr.ind argument in the **which()** function call above? Why would we then take the first column of the output and need to call the **unique()** function?

It will give us both the row and column values where there are "true" values.
Q8. Using the up.ind vector above can you determine how many up regulated genes we have at the greater than 2 fc level?
250
Q9. Using the down.ind vector above can you determine how many down regulated genes we have at the greater than 2 fc level?
367
Q10. Do you trust these results? Why or why not?

I don't fully trust these results because we don't know if they are statistically significant.

Q11. Run the **mapIds()** function two more times to add the Entrez ID and UniProt accession and GENENAME as new columns called res$entrez, res$uniprot and res$genename.

(In code)