The assignment is due on January 19 at 9 am. You need to show your work and answers to each problem as well as plots and **R** code and output to substantiate your answers. You don't have to typeset your solution; handwritten answers together with a printout of code and output are fine. Please combine your submission into one PDF file and submit it to your folder on canvas. You may discuss the homework with other students but you need to prepare your solution by yourself.

In this first assignment you will need to employ some basic **R** commands such as **plot, lm, qqnorm**. The **help(. . .)** command is very useful to see the usage and possible arguments that can be used with each command. Remember that you don't need to supply all the arguments that are available. At the end of the help file there is an example that shows how the command is used, which is often all you need. Besides the manual 'An Introduction to R' on CRAN, you are welcome to look at the lecture notes from last year at http://web.stanford.edu/class/stats191/. These lecture notes contain small examples with the appropriate **R** code as well as a link 'Some tips on **R**'.

1. Parts (a) to (g) of problem 2.12 in the book, see the accompanying file for a copy of the problem. The newspapers data are in the file 'NewspapersData.txt'. There is also a file 'NewspapersData2.txt' which only has the numbers. You can read the file into **R** using the **read.table** command.

2. One of the most startling and profound scientific revelations of the twentieth century was the discovery in 1929 by the American astronomer Edwin Hubble that the universe is expanding. Hubble's announcement shattered forever the ancient belief that the heavens are in a state of cosmic equilibrium. Actually, galaxies are receding from one another at fantastic velocities. If Y is a galaxy's recession velocity relative to any other galaxy and X is it's distance from that other galaxy, Hubble's law states that

$$Y = \beta_1 X$$

where β_1 is known as Hubble's constant. Hubble's constant is a very important number to cosmologists. When we take its reciprocal and convert its units to time, it becomes an estimate of the age of the universe. Hubble's law is consistent with the Big Bang theory, that the entire universe exploded into existence from a single point.

The table below gives distance and velocity measurements made on 11 galactic clusters (the data are in the file GalacticClusters):

Cluster	Distance, X_i (millions of light years)	Velocity, Y_i (thousands of miles/sec)
Virgo	22	0.75
Pegasus	68	2.4
Perseus	108	3.2
Coma Berenices	137	4.7
Ursa Major No.1	255	9.3
Leo	315	12.0
Corona Borealis	390	13.4
Gemini	405	14.4
Boötes	685	24.5
Ursa Major No.2	700	26.0
Hydra	1100	38.0

(a) The model suggested by Hubble's Law is different to the usual simple linear regression model since it imposes the condition that $\beta_0 = 0$, and in the fact that only one parameter β_1 is fit. *Derive* the least squares estimate of β_1 in this special model without intercept. (You should assume that the X's are *fixed.*)

(b) Fit the usual simple linear regression model and test the null hypothesis $H_0 : \beta_0 = 0$ against an appropriate alternative. Do you think Hubble's model is appropriate given the data? Explain by examing the assumptions of linearity of the data and of normality and homoscedasticity of the errors.

(c) **R** allows the fitting of the "no-intercept" model, see 'help(lm)' how to do this.
Fit this no-intercept model to the above data, and hence estimate Hubble's constant. Use your estimate to approximate the age of the universe.

(*Hint: In order to carry out the final calculation, use the fact that the speed of light is 186,000 miles/second; one light year is the distance that light travels in a year (365.25 days = 31,556,736 seconds); and the units of β_1 in this question are*

$$\frac{(thousands\ of\ miles)}{(seconds) \times (millions\ of\ light\ years)}.$$

Express your answer in years.)