Q1:

a) Read in the data airquality.csv from the st1050 folder and name it 'air'.

Make the following variables from the dataset:

Ozone = air$Ozone

Solar.R = air$Solar.R

Wind = air$Wind

Temp = air$Temp

Month = air$Month

b) Since 'airquality' is an R dataset you could also use it from R without reading the csv file. Try the command names(airquality)- this is using the built-in R dataset. You can change the name by setting airR=airquality, or you can leave the name as airquality. Set up the above variables (Ozone etc) using the airquality dataframe.

Use the following command to set up R to send graphics files to a pdf called Rplotxxx where xxx is the number of the plot you’re putting out. The ‘onefile=F’ argument sends each graphics event to its own file (if onefile is set to ‘T’ they all go into one large pdf.) The pdf will be written to your ST1050 folder.

pdf(file ="Rplot%03d.pdf",onefile=F)

Q2:

Make a new temperature variable TempC which is the Temp variable in degrees Celsius:

TempC=round((5/9)\*(Temp-32),1)

a) We are going to make a ‘categorical’ variable called tempcat from TempC- a categorical variable has values that are labels instead of numbers. In this case we want to make two groups named ‘<25’ and ‘>=25’. There are (at least) 2 ways to make this categorical variable:

i) tempcat=TempC; tempcat[TempC<25]=’<25’; tempcat[TempC>=25]=’>=25’;

ii) use the ifelse command- which is a shortcut for making a variable with two values:

tempcat= ifelse(TempC<25,'<25','>=25')). This command often ends up being used- so go to ‘help’ and read how it works. Then use this method to make the variable tempcat.

b) Convert the variable tempcat to a factor.

c) Use the sort() function to sort Ozone values for each of the two levels of tempcat.

Q3:

Plot a histogram of Ozone for the days that are <25; add a histogram to that plot of Ozone for days that are >=25 (using 'add=T'). Use the following histogram commands; you will need to:

a) fill in a,b,c and d (you will need to find out what they are- an easy way is to just make 2 histograms and look at what you will need for the min and maxes of x and y.)

b) set up a variable named ‘bins’ to specify the Histogram 'bin widths'. (The 'breaks' argument in the histogram function will be set to these bin widths).

Make bins= a sequence from the min to the max of xlim with jumps of 20: i.e. (20,40,60…max(xlim)). You’ll need to round the max(xlim) up so that it is divisible by 20.

a=

b=

c=

d=

bins=

hist(Ozone[tempcat=='<25'],xlim=c(a,b),ylim=c(c,d),breaks=bins,col=5,main='Ozone by Temperature',xlab='Ozone')

hist(Ozone[tempcat=='>=25'],add=T,xlim=c(a,b),ylim=c(c,d),breaks=bins,density=10)

Q4:

Using the par command set font.main=4 and font.axis=3.

Do a boxplot of Ozone level by tempcat, colour the cool temperatures blue and the warmer temperatures red. Add the title 'Ozone by Temperature (Celsius)'.

Q5:

Do a t-test to see if there is a significant difference between Ozone levels for the two temperature groups. Comment on it.

Q6:

Use the ifelse command to create a variable called windcat which is Wind divided into 2 groups- ‘<10’ or ‘>=10’. Make it a factor and then do the histogram, boxplot and t-test as in Q4. Comment (1 sentence) on each plot and on the t-test results.

To output the final pdf type in:

dev.off()

(Each plot command will probably output the previous plot to a pdf- but that may be system dependent. Note that the name of the plot will appear, but if it hasn't been 'output' by R the file will not open.)