789 Homework

This homework will require use of the dataset (esoph) from a case-control study of (o)esophageal cancer in Ile-et-Vilaine, France. This is an R dataset which can be loaded into a current R environment with the data() command. The data set can also be found on the class website in .csv form. The codebook can be found with the help() command.

1. Read in the Esophageal Cancer dataset (esoph) and attach the dataset.

2. Use an 'ifelse' statement to create a new variable where the six current categories of 'agegp' is condensed into three categories. The age groups of 25-34 and 35-44 compose the first category, 45-54 and 55-64 the second, 65-74 and 75+ the third.

3. Use an 'ifelse' statement to create a new variable where the categories of 'alcgp' are 1, 2, 3, and 4 instead of '0-39 gm/day', '40-79', etc. Repeat these steps for the tobacco variable.

4. Create a new data frame that combines your new age variable, 'ncases', and the original age group variable as column vectors

5. Create a matrix containing the sums of the number of cases summed by alcohol group then by the new age group. (*hint - use the 'tapply' function*). Repeat these steps for the tobacco variable.

6. Make your new alcohol group variable a vector containing only the unique values of your new alcohol group variable. Repeat these steps for the tobacco variable.

7. Plot the number of cases by alcohol group for each of the three new age groups (all three age groups should be on the same graph). Also, plot the number of cases by tobacco group for each of the three age groups. Each age group should be represented by a different color of your choice.

8. The tick marks of the x-axis on your original graphs will contain the new values for alcohol and tobacco group (1,2,3,4). Change each graph so that the tick marks on the x-axis contains the original labels for the alcohol and tobacco groups.

9. Use the 'par(mfrow=c(,))' command to place both graphs horizontally in the same window.

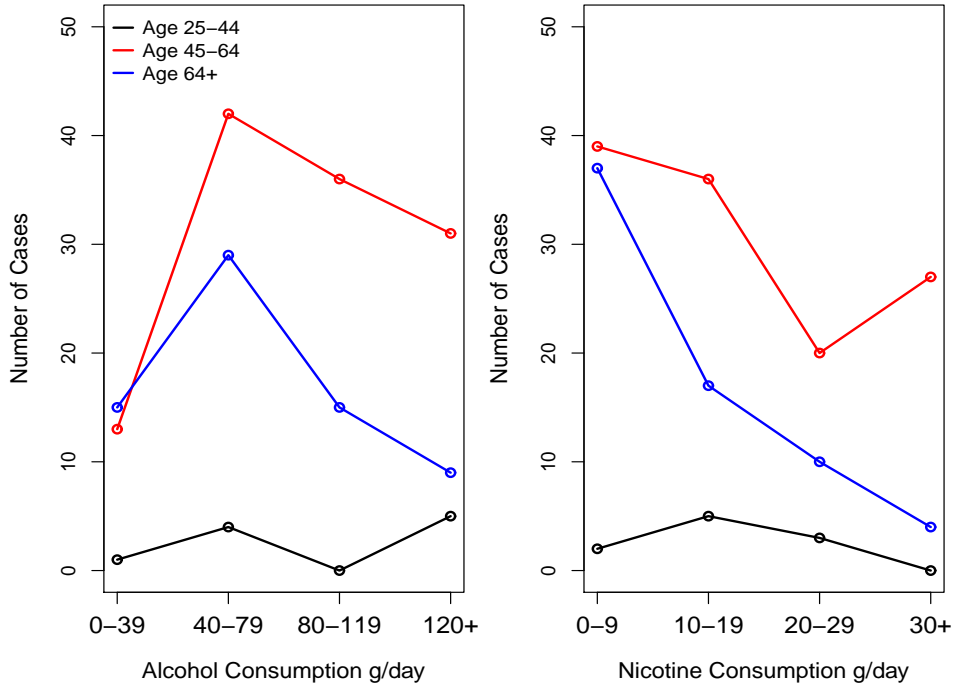10. The final graph should look like the graphs below. Save your graph.

Figure 1: