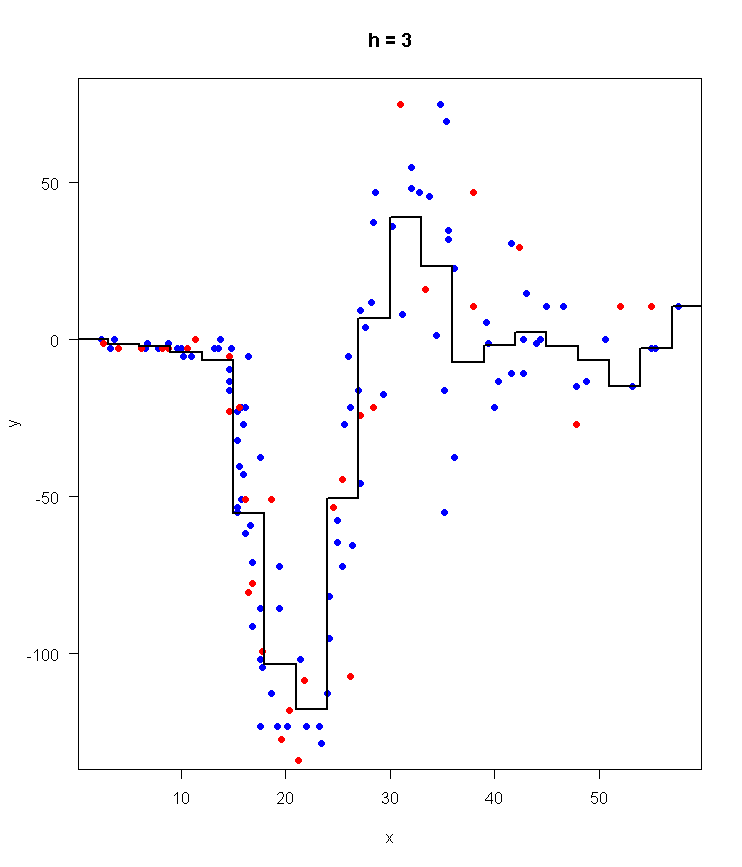
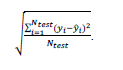
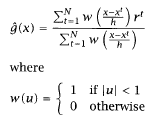
After reading the book section 8.8 which was different in the pdf edition 2 that I follow the course from, I imported the dataset for question 2. I read the data by csv function and split the dataset as x\_train,y\_train,x\_test,y\_test according to the pdf directions (100 for train set 33 for test set).

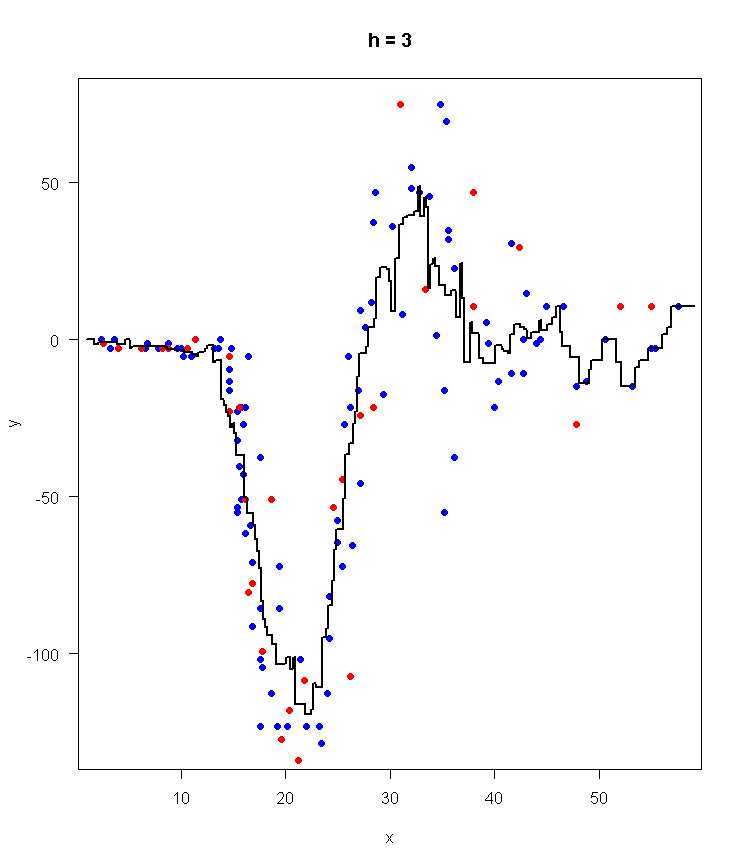
For Q3 I followed the lab code and book instructions to learn a regressogram with bin parameter= 3. I set the origin to 0 and maximum value to 60 as max(x) was around 57.6 I used 60 to make it simple. By simply deleting the denominator part of the lab code of histogram estimator and dividing it by the nominator we obtain g(x)/r. Since r is defined as y\_train in my code I multiplied the expression in the nominator which is in sum with y\_train. This gave me my g\_regressogram value. To plot the regressogram I used left\_boreds,g\_reg and right\_borders. To plot the points I used blue color as train points and red as test. Finally obtained the following graph

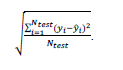


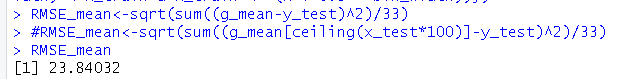
For Q4 I used the formula given in the pdf  my y was the new g\_reg that I calculated by passing floor(x\_test/3)+1 to sapply instead of iterating over length of left\_borders. Than I subtracted it by y\_test value and followed the formula. I obtained the following result

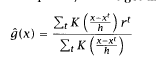


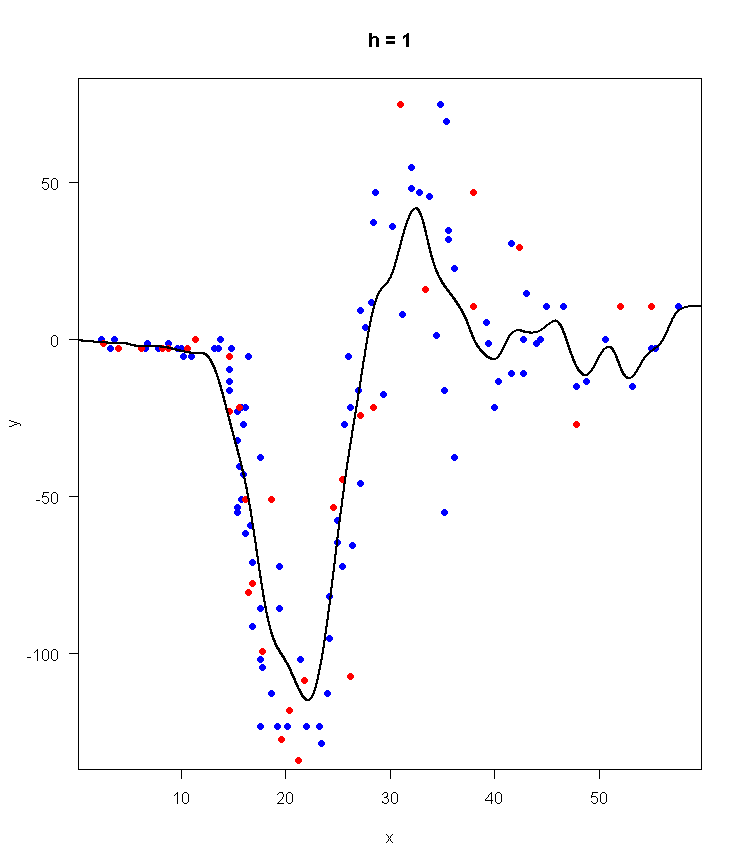
For Q5 I again set the bin width to 3 and followed the formula  given in the book but as the professor warned us I changed the condition in w(u) and followed the professor’s approach. I modified the lab 6 ‘s naïve estimator code and obtained the following output Followed a similar way to plot the graph as the regressogram.

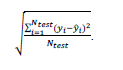


For Q6 instead of using data interval in sapply I used x\_test this gave me my new g\_mean which is y in the formula and again I used the formula as I did in Q4. I obtained the following result



For Q7 I defined bin width as 1 and modified lab6’s kernel estimator to use the following formula  and I obtained the following output. Followed a similar way to plot the graph as the regressogram



For Q8 to get my new g\_kern is used x\_test instead of data\_interval this gave me the  y in the formula. After following the formula I obtained the result.



The following graphs show the plotted points of the dataset. Which is used for every step of the homework