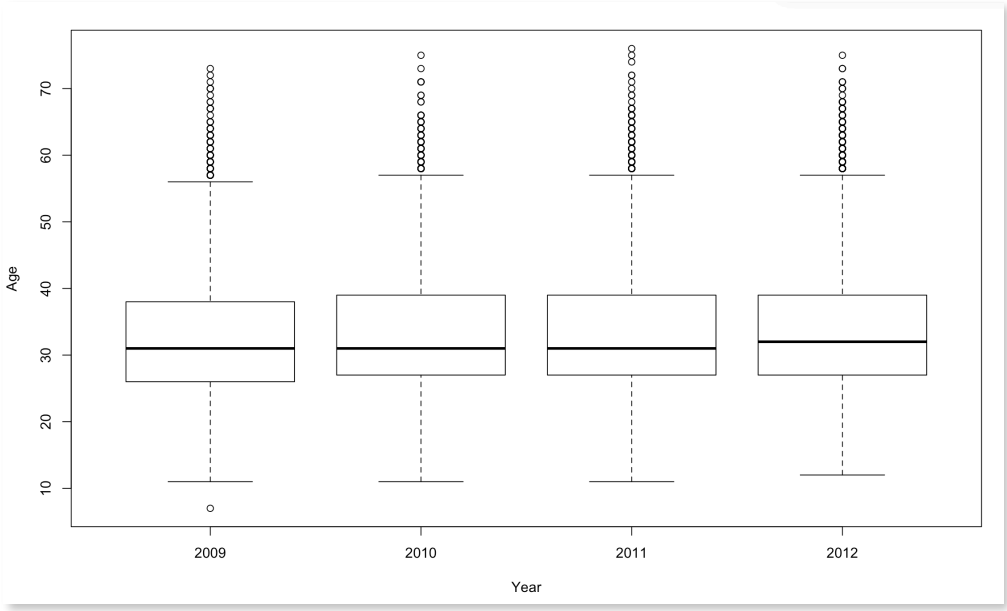
FE-582 – Assignment 2

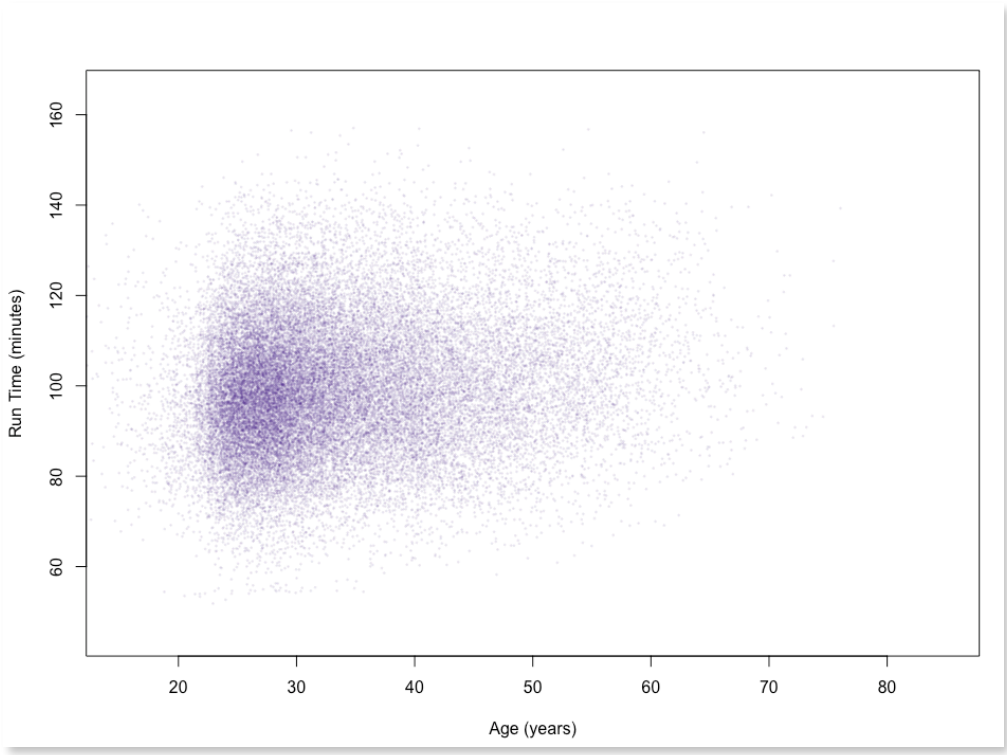
Shuo Jin

Follow the example in Lecture 3 on Modeling Runners' Times. Use the Race Results 1999-2012 link (http://cherryblossom.org/aboutus/results_list.php) to extract the Race Results for Order of Finish – Women for a couple of years. Extract and clean the data for as many years as you can. Please do the following

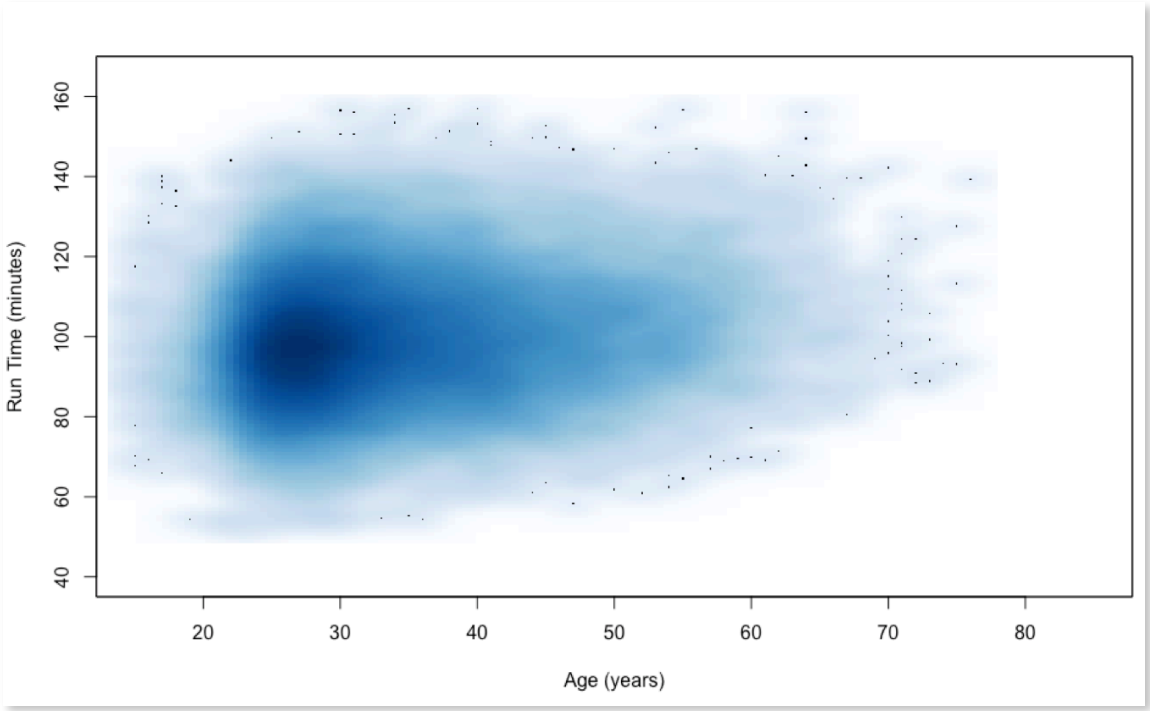
- Box Plot of Age by Year for Female Runners



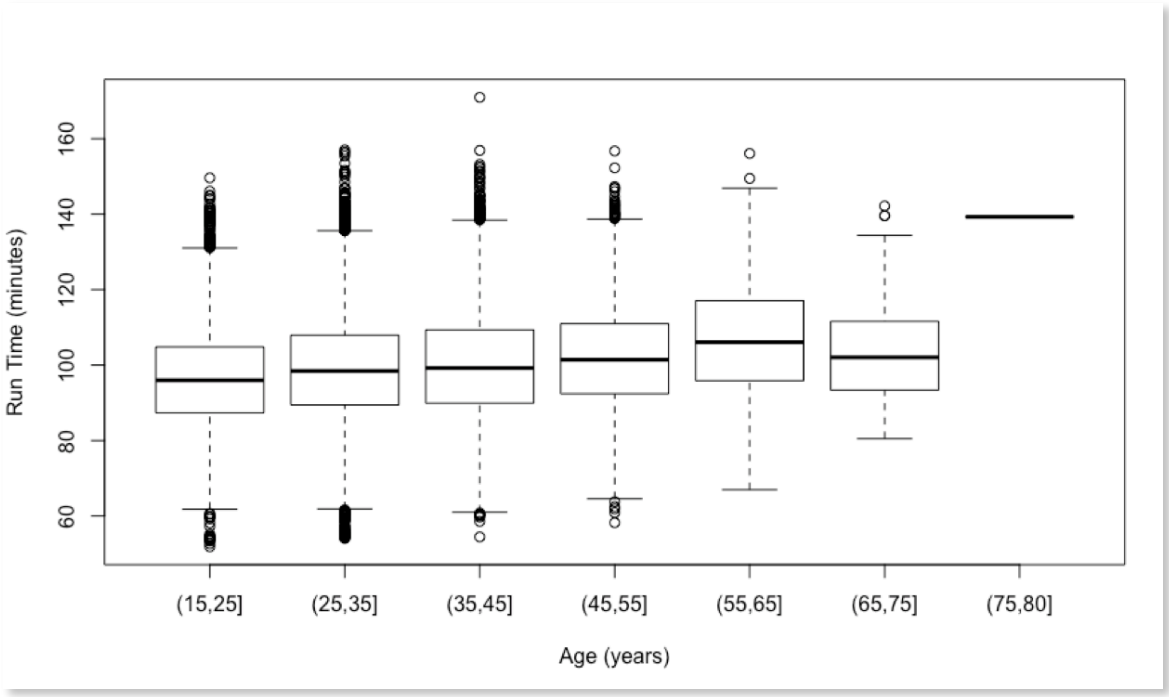
- Scatter Plot for Run Times vs. Age for Female Runners



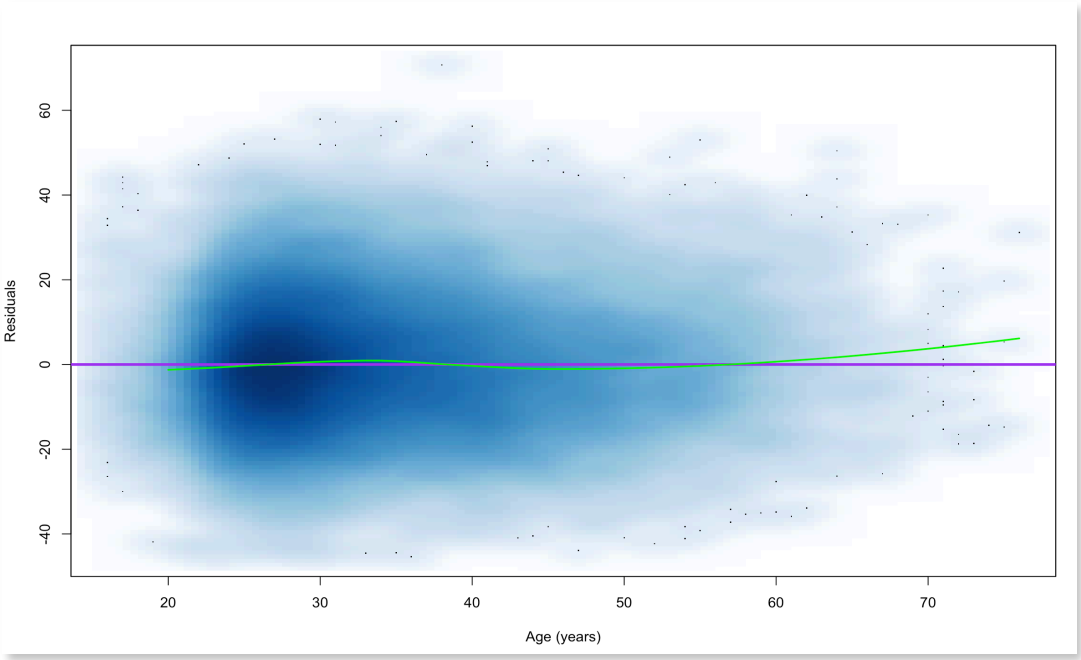
- Fit Models to Average Performance



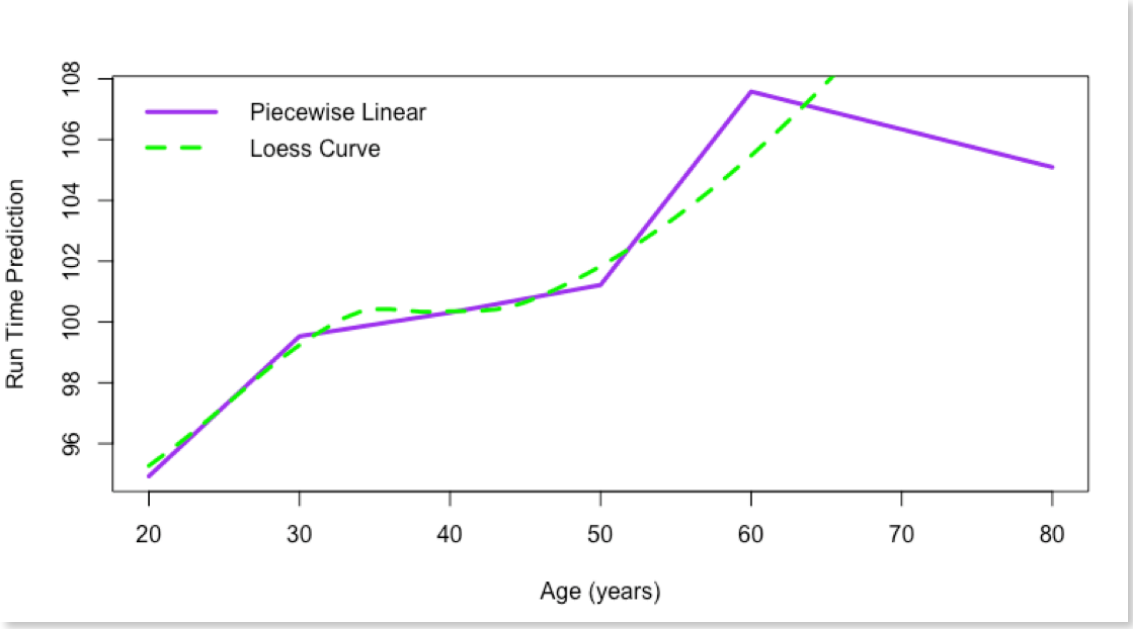
- Side-by-Side Boxplots of Female Runners' Run Time vs. Age



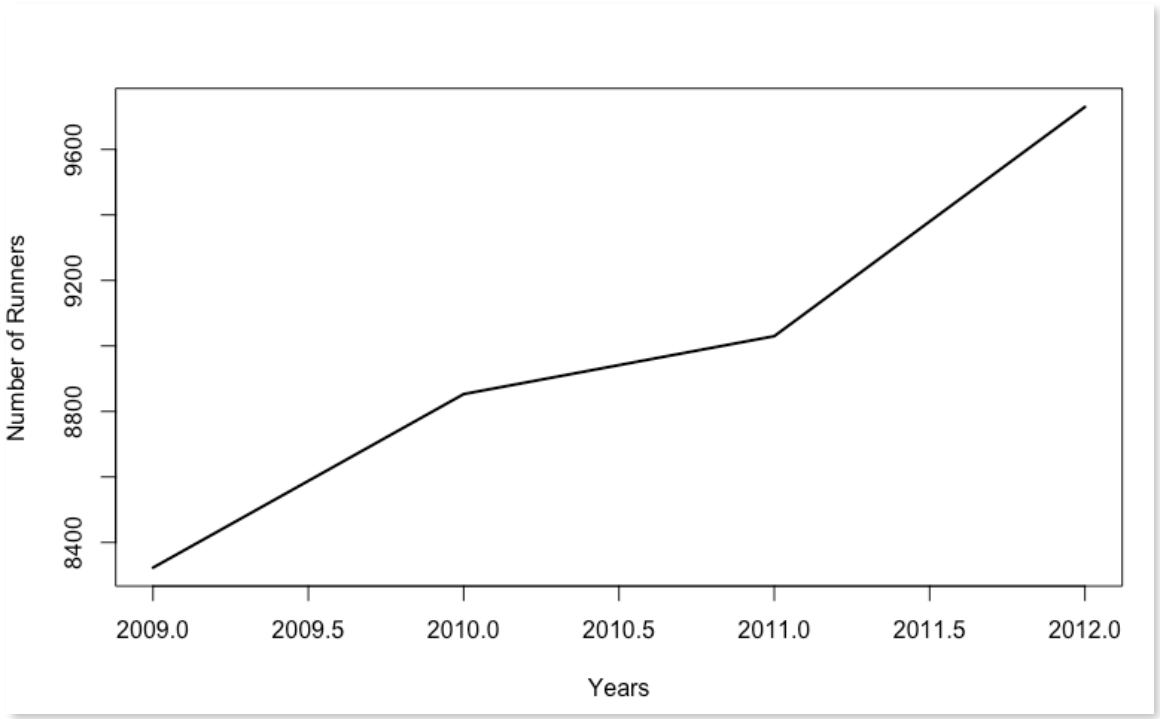
- Residual Plot from Fitting a Simple Linear Model of Performance to Age



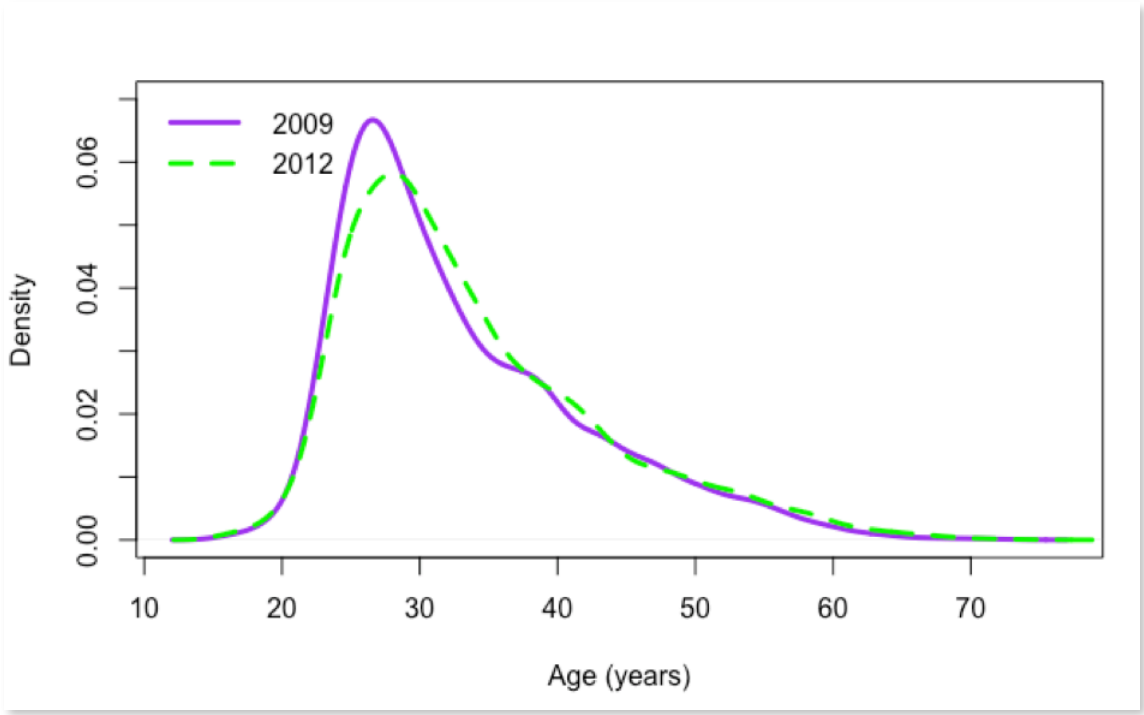
- Piecewise Linear and Loess Curves Fitted to Run Time vs. Age



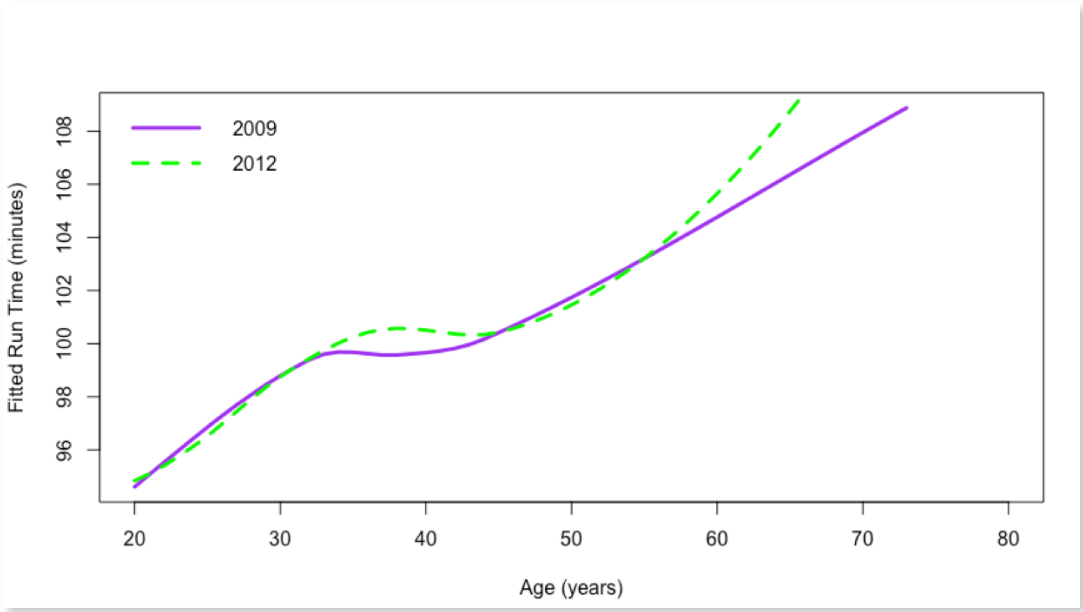
- Line Plot of the Number of Female Runners by Year



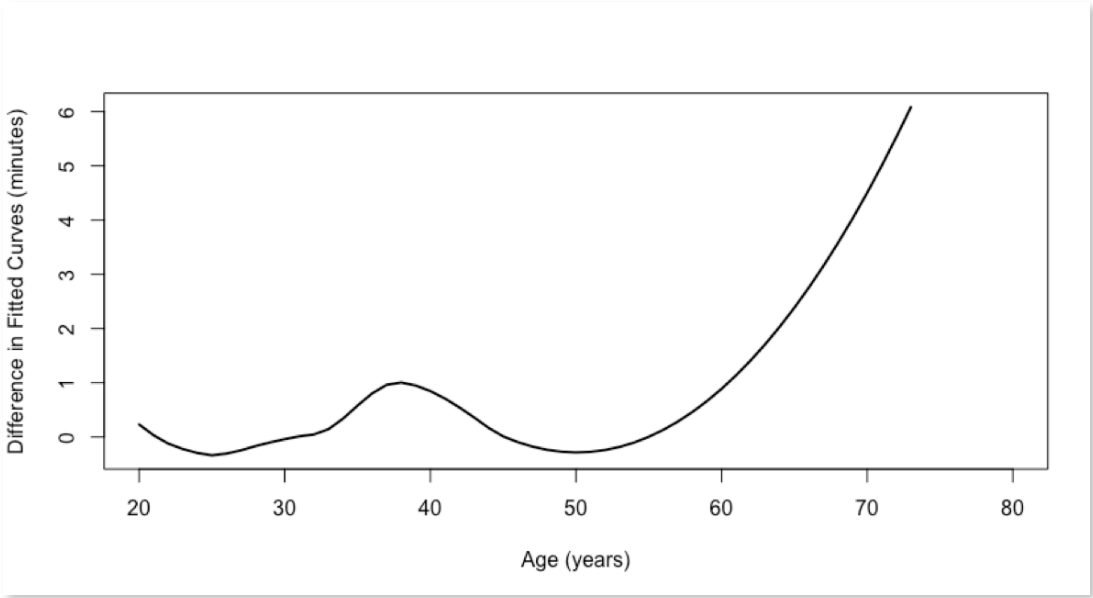
- Density Curves for the Age of Female Runners for 2 years (smallest and largest year that you analyzed)



- Loess Curves Fit to Performance for 2 years (smallest and largest year that you analyzed) Female Runners

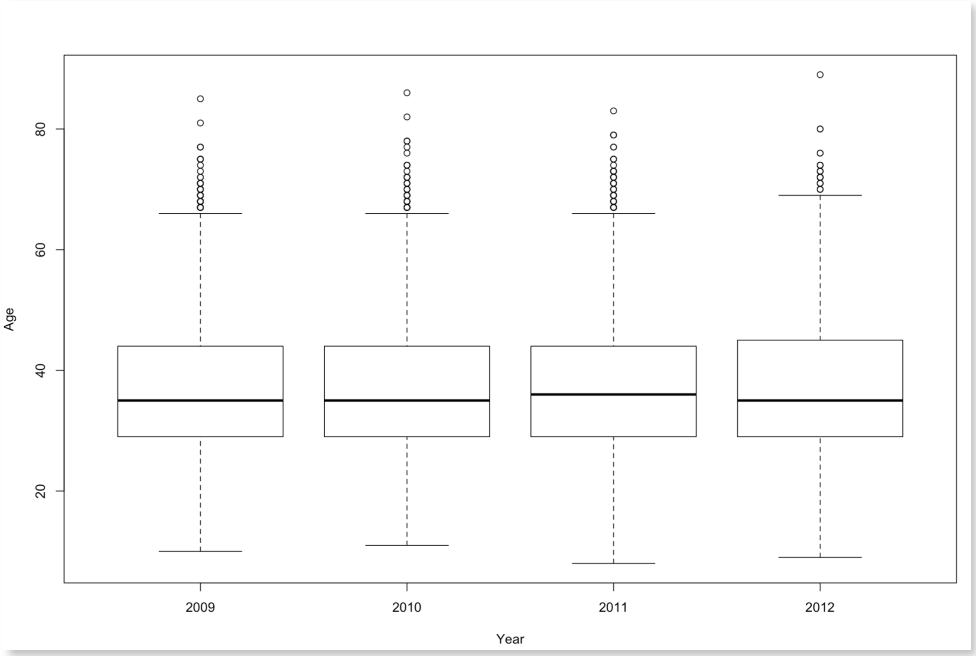


- Difference between Loess Curves of the predicted run time for 2 years (smallest and largest year that you analyzed)



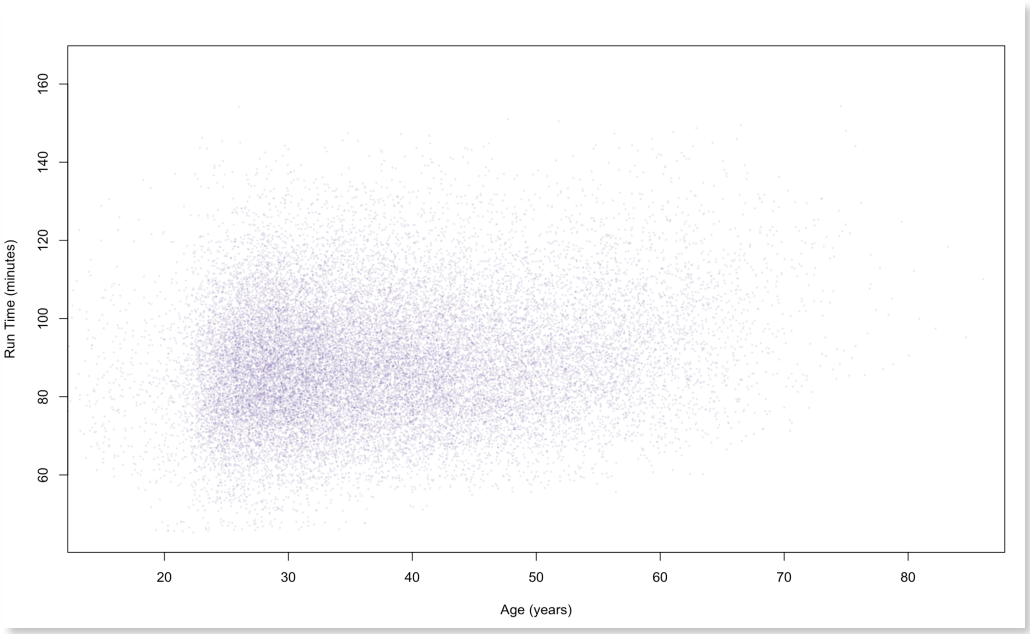
- Compare the results of the performance of the male runners (previously analyzed in class) and female runners for the yearly data that you selected to analyze

    - Box Plot of Age by Year for Male Runners

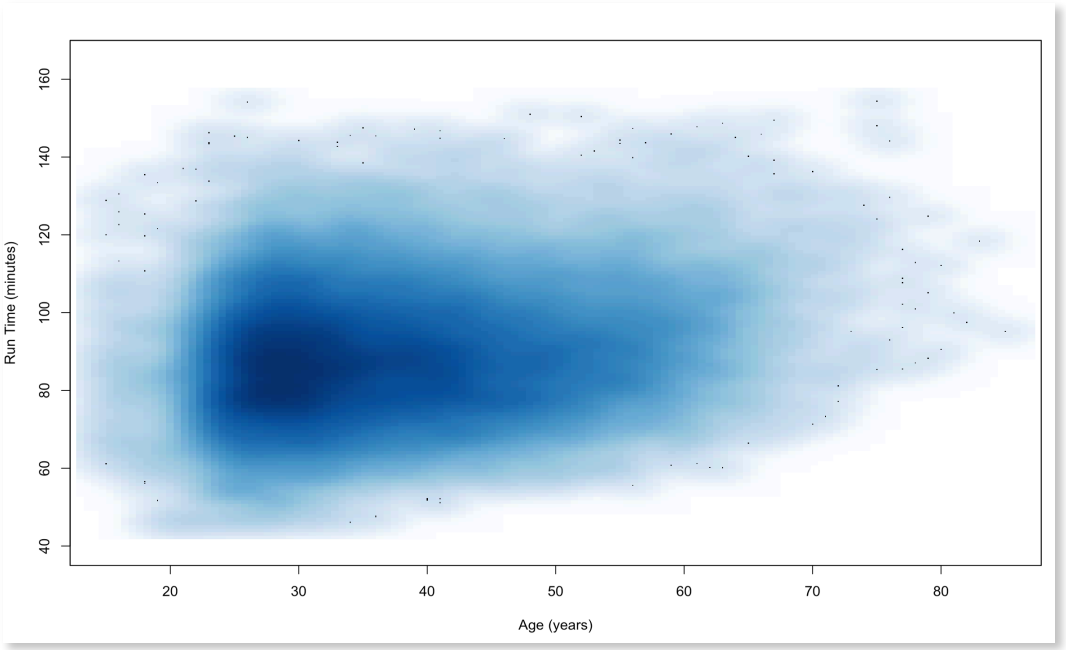


Compare to the female runner in 2009-2012. The average age of male is larger than the female. And the age range of the male is larger than the female.

    - Scatter Plot for Run Times vs. Age for Male Runners

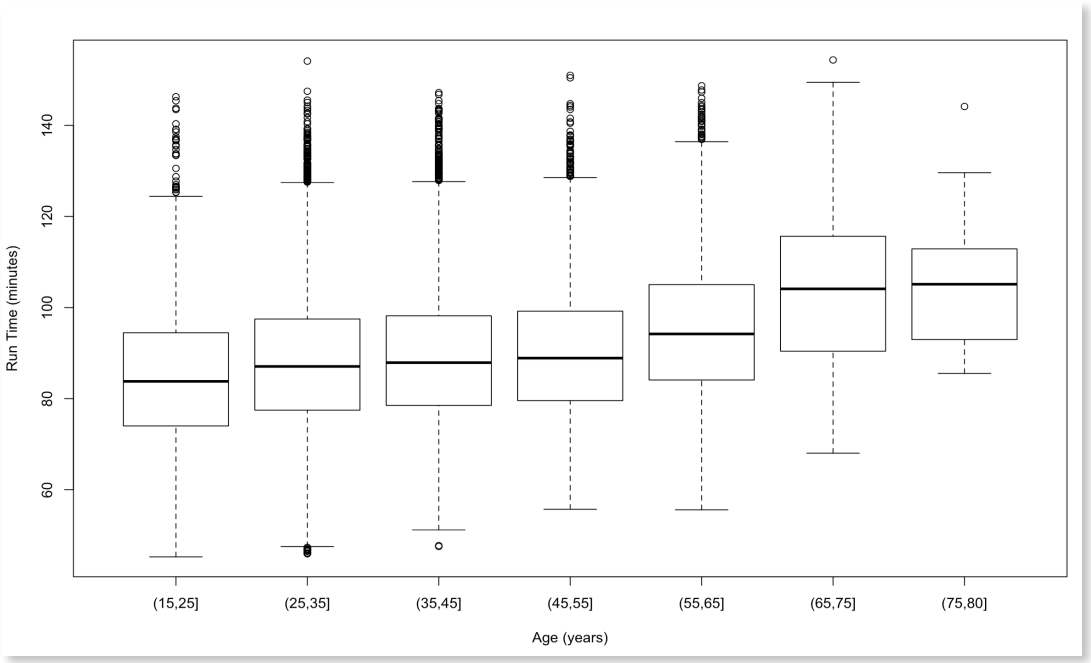


- Fit Models to Average Performance



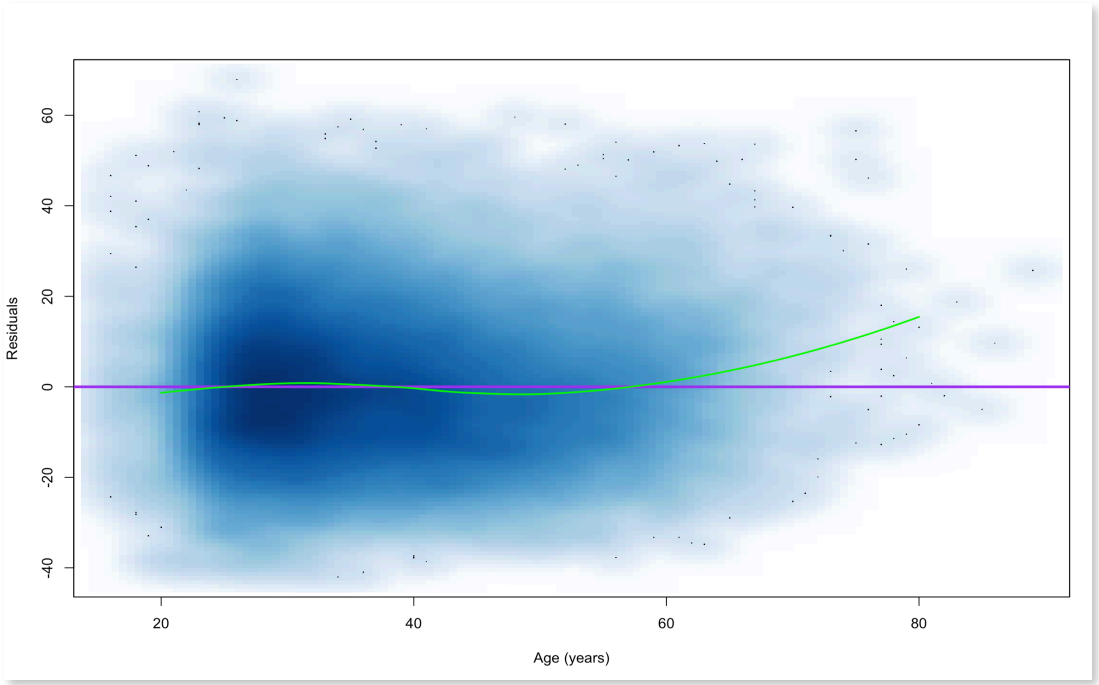
The female running time and age are more likely to have a centralize trends between age 20-30 and running time 80-120 minutes. However, the male has a centralize trends between age 20-40 when their running time is 60-120 minutes

- Side-by-Side Boxplots of Female Runners' Run Time vs. Age



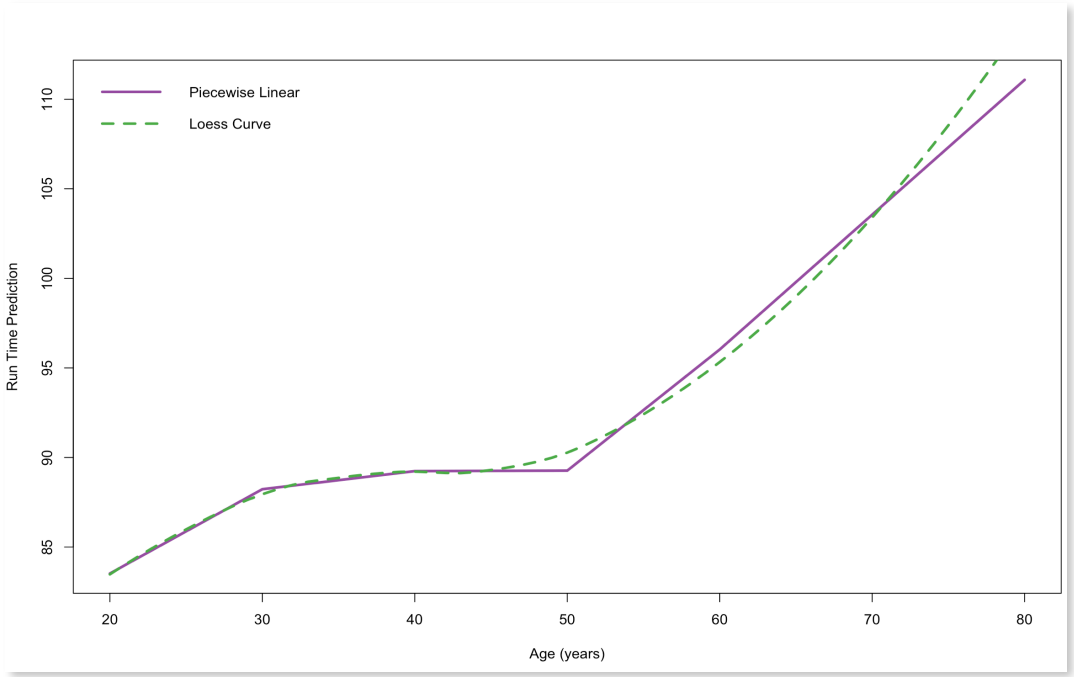
The female and male running time is increased by the age increase. However, the male running time is less than the female running time in general. And the age of 75-80 only 1 person in the female group but it has 7 people in the male group.

- Residual Plot from Fitting a Simple Linear Model of Performance to Age

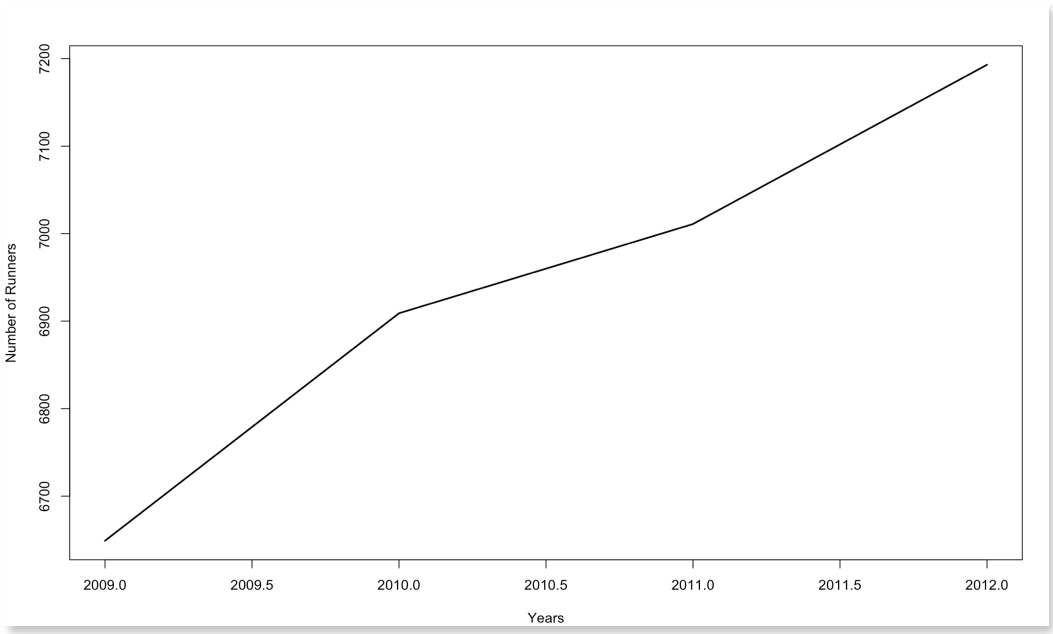


From the 20-60 the linear model is good fit for the dataset, However when the age above 60 there is a apparently trends in residual plot which mean the linear model not good predict the data when age above 60. So we consider the piecewise linear model.

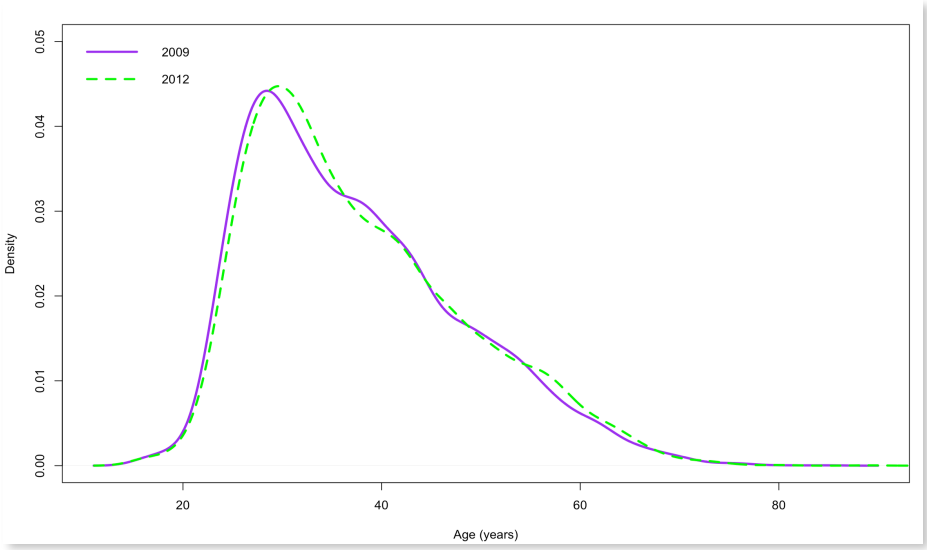
- Piecewise Linear and Loess Curves Fitted to Run Time vs. Age



- Line Plot of the Number of Male Runners by Year

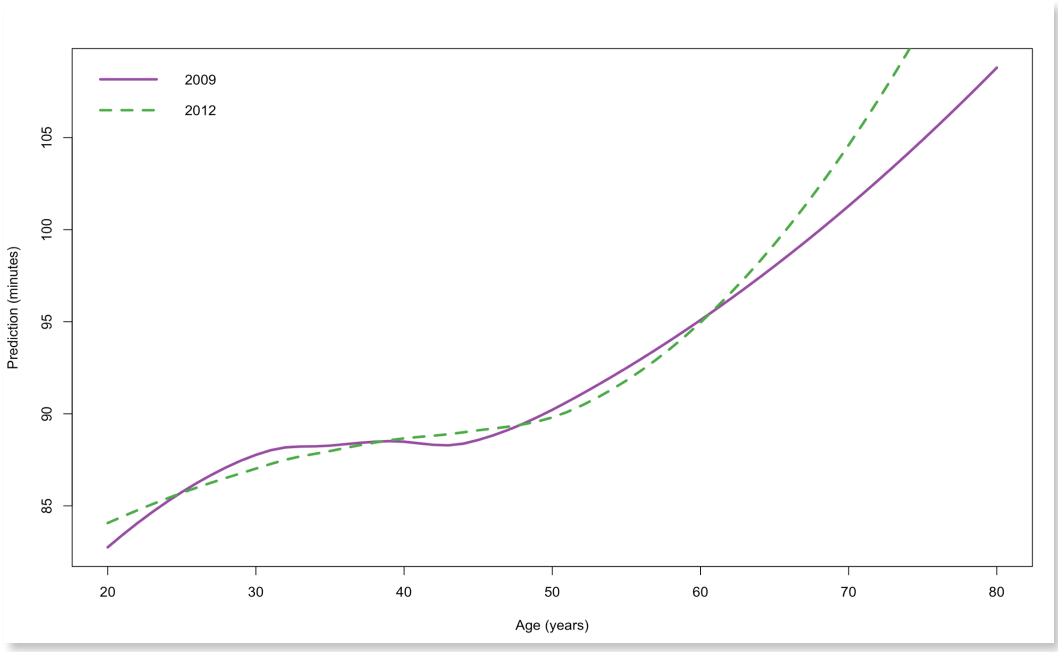


- Density Curves for the Age of Male Runners for 2 years (smallest and largest year that you analyzed)



The male age density between 20 and 40 is much smaller than the female age density.

- Loess Curves Fit to Performance for 2 years (smallest and largest year that you analyzed) Male Runners



- Difference between Loess Curves of the predicted run time for 2 years (smallest and largest year that you analyzed)