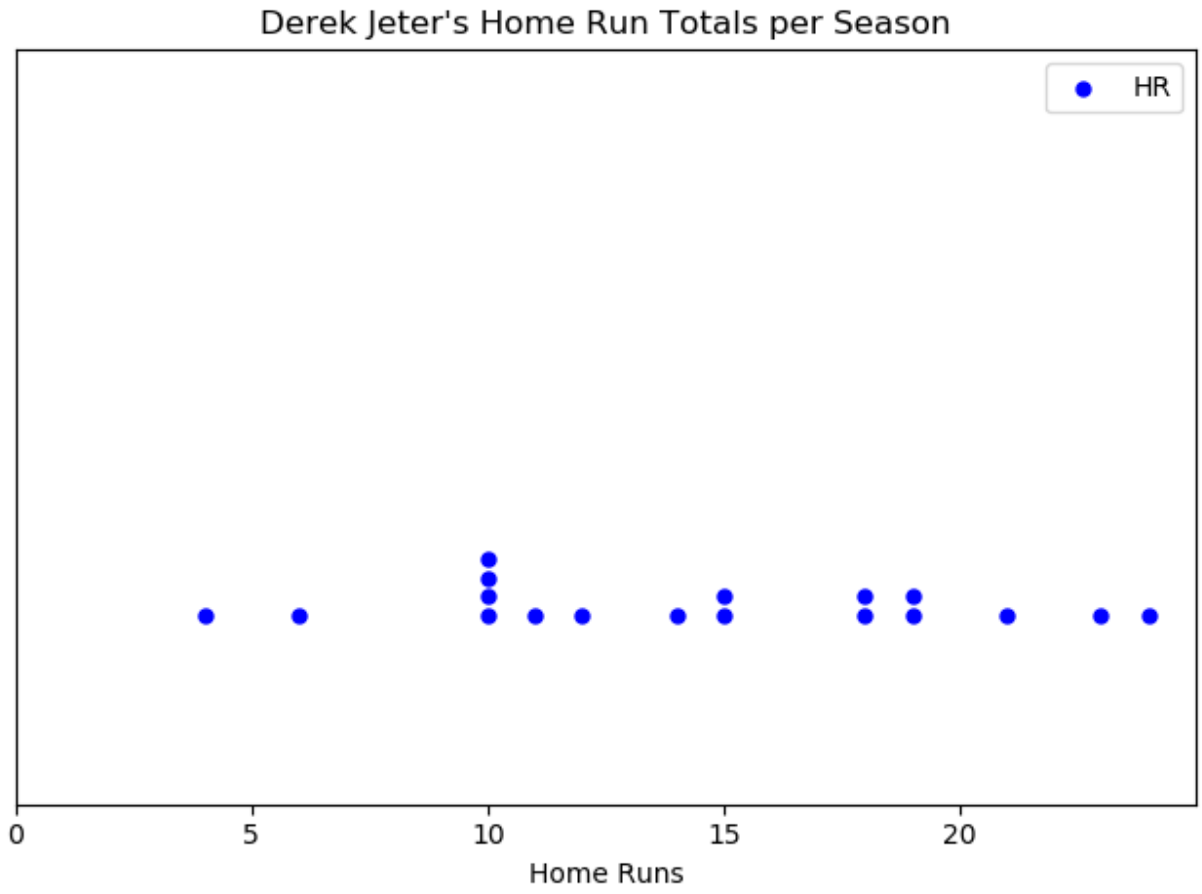
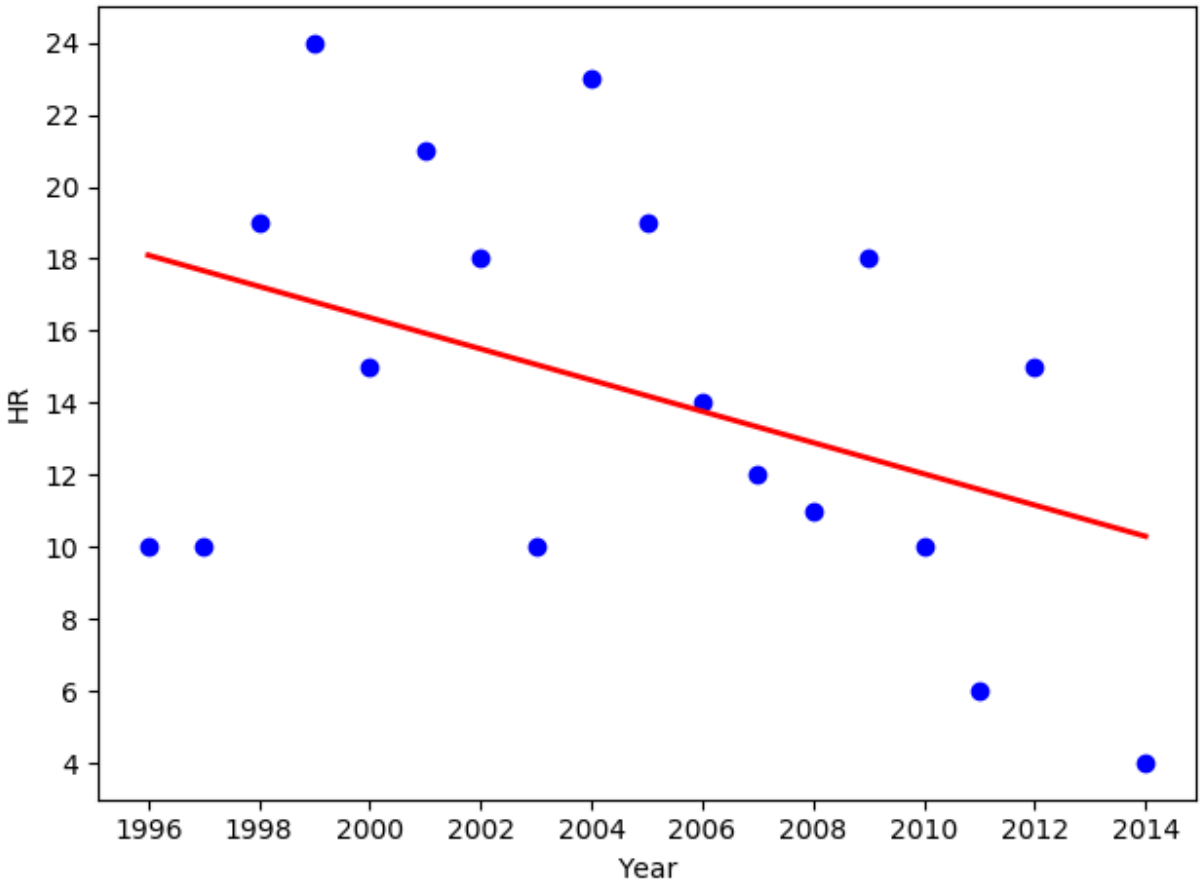
Analysis of Derek Jeter's Home Run Totals

Scanning over Jeter's hitting statistics, we see that he displayed good power over his career hitting 260 home runs. 260 home runs places Jeter 3rd all-time among shortstops, only trailing Cal Ripken Jr and Miguel Tejada who both moved over to third base the latter half of their careers.

We graph his season by season home runs using a dotplot. We can see from the graph below his home run numbers range from 4 to 24. There appears to be much variation in these home run numbers. There is a cluster of values in the 10 to 15 range. The median number of home runs hit is 14.5.



There appears to be much variation in these home run numbers. Maybe some of the variation of Jeter's home run numbers can be explained by the age which he hit them. A ball player generally improves in hitting ability in the early part of his career, peaking in his late 20's, and then declines in ability towards the end of his career. In the scatterplot below we graph the home run count against the year.



We can determine that Jeter's home run numbers were in the 15 to 20 range during seasons around 2000 in his prime years. It appears his season home run counts decreased over the later years of his career. We can summarize this by fitting a regression line in the scatter plot. The equation of this line is HR = 883.68 - .4337 x Year. Jeter's home run count tended to decrease about -.43 throughout his career. .43 is the slope of the linear regression line and 883 the intercept. There were come exceptions to this trend with 18 home runs in 2009 and 15 home runs in 2012.

Regression analysis helps us search for relationships among the variable's year and home runs. Year is the independent variable, input, or predictor. Home run is the dependent variable, output, or response. The coefficient of determination is a key output of regression analysis. It's a measure used in statistical analysis that assesses how well a model explains and predicts future outcomes. It ranges from 0 to 1 with higher translating to stronger predictions. We can calculate an R^2 of .17 from the Derek Jeter's home run totals by season which means 17% of the data is predictable. Typically, R^2 values above .5 are considered acceptable or strong and anything below is considered moderate or weak. A value of .17 is moderate to weak and we cannot use the data to confidently predict future outcomes.