Tall swimmers, fast swimmers?

Suppose you are a swim coach and notice that you for swimmers in the same age range and of the same gender, the taller swimmers tend to be faster than the shorter swimmers. You wonder if this is coincidental, or if a correlation exists between the two variables. In order to investigate, you collect the height of 11 of your swimmers as well as their 50 yard freestyle time. The data you collected is below.

|  |  |
| --- | --- |
| Height (in centimeters) | 50 yard freestyle time (in seconds) |
| 161.6 | 35.95 |
| 155.8 | 35.32 |
| 149.2 | 36.09 |
| 168.8 | 29.05 |
| 165.7 | 29.63 |
| 157.5 | 38.67 |
| 161.5 | 32.15 |
| 172.0 | 28.68 |
| 165.9 | 34.92 |

Answer the following questions to explore the relationship between height and 50 yard freestyle times among swimmers.

1. Enter the data set into the regression applet found at <http://www.shodor.org/interactivate/activities/Regression/> Adjust the window on the applet if you'd like to. What do you notice about the data? Does there appear to be a correlation between the two variables? If so, which direction is the correlation and what does that mean in the context of the activity?
2. Check the "Fit your own line" box and estimate what you think the line of best fit should be. Make sure to record the equation for this regression model. Now hit the "Show Residuals" button. Seeing the residual plot, would you change your line of best fit? Why or why not.
3. Now check the "Display line of best fit" box to compare the line of best fit with your estimate. Make sure to record this equation as well as the r value. Is the r value positive or negative? Given the value of r, would you say that there is a strong or weak correlation between height and 50 yard freestyle time?
4. Interpret the slope of the line of best fit as it relates to the context of this activity. Interpret the y-intercept as it relates to this activity. Does the y-intercept make sense in this context?
5. If you were the coach, would you decide that there is or is not a linear relationship between height of a swimmer and their 50 yard freestyle time? Explain your answer.
6. Now uncheck "Fit your own line." Pretend that on the day the swim coach collected data, one swimmer missed practice. The next time they showed up, the coach measured their height and recorded their 50 yard freestyle time. The coach found that the swimmer was 160.0cm tall and had a 50 yard freestyle time of 25.47 seconds. Add this new data point to the scatter plot.
7. Would your answer to number 5 change? Why or why not?
8. Now remove the point that was most recently added. Add a new point that would represent someone who is very short and yet very fast. How did this change your regression model? Your r value? Do you think this point you added could be possible data given the context of this activity? (For example, if you added the point (0,0) this does not make sense because you wouldn't be gathering data from someone who is 0cm tall.)
9. Move this point around and observe how the line of best fit changes. Why do you think one point has the potential to change the line of best fit so significantly?