ANALYSIS OF OLYMPIC SWIMMING DATA IN MEN'S 50-METER FREESTYLE

Table of Contents

Section I: Introduction

The Olympic Games are a series of sporting competitions that consist of both summer and winter athletics. They occur every four years, alternating every two years. These games draw the attention from thousands of athletes around the world, all uniting to compete in the world's biggest sporting competition with the viewership of millions around the globe. This iconic sporting event has a rich reputation and history as it has been bringing people together since the time of ancient Greece to compete, have fun, and share the Olympic spirit.

In this exploration, I will be focusing specifically on the men's 50-meter freestyle swimming event. The aim is to analyze the final swim times of competing athletes in this event from the 2020 Tokyo Summer Olympics. My rationale is to compare this data to previous Olympic Games to find any reoccurring trends. I will be using graphs to model and find potential factors that could correlate to an athlete's finish time along with any other trends or general observations that I find. Swimming has always been a fun activity that I practice in my spare time. I find it to be an excellent source of exercise and a refreshing one too. After taking lifeguard courses, and First Aid certification, I acquired a passion for competitive swimming. This led to me watching the Tokyo 2020 swimming competitions and to join my school's swim team which ultimately led to my decision of picking this topic.

Section II: Overview of Collected Data

To conduct this investigation, I gathered the swimming data from Entertainment and Sports Programming Network (ESPN) for the Tokyo 2020 men's 50-meter freestyle. To ensure a large enough sample size, I will only be using the overall data from the multiple heats as they contain the most competitors. Heats are used in swimming to organize swimmers into waves since typically there are more swimmers than available lanes in the limited capacity pool. The data for the Tokyo 2020 men's 50-meter freestyle can be found in *Table 1* in the Appendix. Despite the best of my abilities, I could not find the height and weight for some of the athletes. Due to this limitation, the graphs and information derived from this data table will have some inaccuracies which will be discussed in *Section VI* on page 20.

Section III: The Quadratic Regression Least Squares Method and Correlation

The quadratic regression least-squares method is the best way to find an equation for a curve of best fit manually without a quadratic regression calculator. The equation for the quadratic regression least-squares method is as follows:

$$a\sum x^4 + b\sum x^3 + c\sum x^2 = \sum x^2 y \quad ①$$

$$a\sum x^3 + b\sum x^2 + c\sum x = \sum xy \quad ②$$

$$a\sum x^2 + b\sum x + cn = \sum y \quad ③$$

$\sum x$ represents the sum of all x values.

$\sum y$ represents the sum of all y values.

n represent the total number of entries.

$\sum x^3$ represents the sum of all x^3 values.

$\sum x^2 y$ represents the sum of all $x^2 \cdot y$ values.

(Same concept for the other $\sum$ variables)

From this formula, I can find the parameters of a, b, and c, and then find the quadratic equation by solving this system of equations. To solve for this, I will be isolating the variables, a, b, and c, and then using the method of substitution. The value of correlation (r) can be calculated by using the Pearson correlation coefficient formula:

$$r = \frac{n\sum(xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

$\sum x$ represents the sum of all x values.

$\sum y$ represents the sum of all y values.

n represent the total number of entries.

The r value is between -1 and 1 and is used to represent how accurate the regression equation models with the points on the scatter plot. This can tell me whether this relationship has a strong or weak, positive or negative correlation. Additionally, the value of r^2 can be given easily by simply squaring the value of r. The value of r^2 is used to represent the relationship between the dependent variable and the independent variable(s) in a regression model. In this case, how close the data fits the quadratic regression and how much of the dependent variable is affected by the independent variable. This can be used to determine how statistically significant the correlation is, for the purposes of this investigation, $r^2 \geq 0.5$ would be deemed significant.

Section IV: Interpreting Collected Data and Analysis

After collecting the data for the men's 50-meter freestyle event at the Tokyo 2020 Summer Olympics, I can provide a visual representation for my data set by proceeding to graph a scatter plot to model the curve of best fit along with finding the correlation between the swimmer's finish time and the swimmer's height, weight, or age. This will determine whether a swimmer's finish time is affected by their human attributes or not. I hypothesize that there will be some correlation, but nothing drastically significant as swimming is an athletic sport after all.

To calculate the curve of best fit for my data set, I will be using the quadratic regression least-squares method explained in *Section III* on page 4. Instead of using a quintic function to get the highest possible correlation value (r), I will be using a quadratic function as a polynomial function to the degree of three or greater would not show any trends that I am looking for. This is because the function would have multiple local maximum and minimum points which would tell me nothing about the correlation of my independent and dependent variables. Additionally, I will not be using direct points from my data as I cannot guarantee that these points of data lie on the curve of best fit. If one of the substituted points selected was an outlier, the equation would be completely inaccurate. Instead, I will be using the summative of the values through the least-squares method to produce a more accurate equation. Once graphed, the quadratic equation should have a parabola opening either up or downwards with one absolute maximum or minimum point, this is what the final quadratic equation would look like:

$$f(x) = ax^2 + bx + c$$

x denotes the independent variable such as weight, height, and age. $\{x \mid x \geq 0,\ x \in R\}$	y denotes the dependent variable such as time. $\{y \mid y \geq 0,\ y \in R\}$

Section IV Part 1: Weight Correlation

To calculate the correlation between finish time and weight, let finish time represent the dependent variable (t) and weight represent the independent variable (w) since I am looking for the correlation that weight has on finish time. First, I will be calculating the quadratic equation for the curve of best fit for this set of data. Next, I will be calculating the value of r to find the correlation. Then, I will be plotting the data and curve of best fit onto a scatter plot. To use the quadratic regression least-squares method formula to find the quadratic equation for this data set, I will first need a chart with the values of w^2, w^3, w^4, and t^2.

Step 1, finding summative values:

Table 4: Summative Values Chart

w	t	w^2	w^3	w^4	wt	w^2t	t^2
190	21.32	36100	6859000	1303210000	4050.80	769652.00	454.54
218	21.65	47524	10360232	2258530576	4719.70	1028894.60	468.72
209	21.66	43681	9129329	1908029761	4526.94	946130.46	469.15
176	21.67	30976	5451776	959512576	3813.92	671249.92	469.58
194	21.85	37636	7301384	1416468496	4238.90	822346.60	477.42
190	21.88	36100	6859000	1303210000	4157.20	789868.00	478.73
205	21.89	42025	8615125	1766100625	4487.45	919927.25	479.17
187	21.92	34969	6539203	1222830961	4099.04	766520.48	480.48
203	21.93	41209	8365427	1698181681	4451.79	903713.37	480.92
220	21.96	48400	10648000	2342560000	4831.20	1062864.00	482.24
203	21.97	41209	8365427	1698181681	4459.91	905361.73	482.68
150	21.97	22500	3375000	506250000	3295.50	494325.00	482.68
176	22.01	30976	5451776	959512576	3873.76	681781.76	484.44
192	22.14	36864	7077888	1358954496	4250.88	816168.96	490.17
187	22.14	34969	6539203	1222830961	4140.18	774213.66	490.17
194	22.14	37636	7301384	1416468496	4295.16	833261.04	490.17
172	22.22	29584	5088448	875213056	3821.84	657356.48	493.72
190	22.22	36100	6859000	1303210000	4221.80	802142.00	493.72
203	22.25	41209	8365427	1698181681	4516.75	916900.25	495.06
207	22.25	42849	8869743	1836036801	4605.75	953390.25	495.06
187	22.31	34969	6539203	1222830961	4171.97	780158.39	497.73
192	22.33	36864	7077888	1358954496	4287.36	823173.12	498.62
183	22.46	33489	6128487	1121513121	4110.18	752162.94	504.45
181	22.46	32761	5929741	1073283121	4065.26	735812.06	504.45
187	22.52	34969	6539203	1222830961	4211.24	787501.88	507.15
172	22.56	29584	5088448	875213056	3880.32	667415.04	508.95
176	22.61	30976	5451776	959512576	3979.36	700367.36	511.21

Table 4: Summative Values Chart

	w	t	w^2	w^3	w^4	wt	w^2t	t^2
	181	22.67	32761	5929741	1073283121	4103.27	742691.87	513.92
	159	22.74	25281	4019679	639128961	3615.66	574889.94	517.10
	148	23.94	21904	3241792	479785216	3543.12	524381.76	573.12
	154	24.63	23716	3652264	562448656	3793.02	584125.08	606.63
	136	24.81	18496	2515456	342102016	3374.16	458885.76	615.53
	174	24.99	30276	5268024	916636176	4348.26	756597.24	624.50
	156	25.22	24336	3796416	592240896	3934.32	613753.92	636.04
	126	25.38	15876	2000376	252047376	3197.88	402932.88	644.14
	132	26.04	17424	2299968	303595776	3437.28	453720.96	678.08
	136	26.94	18496	2515456	342102016	3663.84	498282.24	725.76
	205	27.46	42025	8615125	1766100625	5629.30	1154006.50	754.05
$\sum$	**6851**	**871.11**	**1256719**	**234030815**	**44157083551**	**156204.27**	**28526926.75**	**20060.42**

Step 2, substituting summative values into equations ①, ②, and ③ from *Section III* on page 4:

Taking the values of the last row, which represents the sum of each column, I can substitute the values into the equations:

$$44157083551a + 234030815b + 1256719c = 28526926.75 \rightarrow ①$$

$$234030815a + 1256719b + 6851c = 156204.27 \rightarrow ②$$

$$1256719a + 6851b + 38c = 871.11 \rightarrow ③$$

Step 3, isolating variable a from equation ①:

$$44157083551a = 28526926.75 - 234030815b - 1256719c$$

$$a = \frac{28526926.75 - 234030815b - 1256719c}{44157083551} \rightarrow ④$$

Step 4, substitute $a = \frac{28526926.75 - 234030815b - 1256719c}{44157083551}$ back into equations ② and ③:

$$234030815\left(\frac{28526926.75 - 234030815b - 1256719c}{44157083551}\right) + 1256719b + 6851c = 156204.27$$

$$\frac{7.23 \cdot 10^{14}b + 8.41 \cdot 10^{13}c + 6.68 \cdot 10^{15}}{44157083551} = 156204.27 \rightarrow ⑤$$ ← simplify

$$1256719\left(\frac{28526926.75 - 234030815b - 1256719c}{44157083551}\right) + 6851b + 38c = 871.11$$

$$\frac{8.41 \cdot 10^{13}b + 9.86 \cdot 10^{11}c + 3.59 \cdot 10^{13}}{44157083551} = 871.11 \rightarrow ⑥$$ ← simplify

Step 5, isolating variable b from equation 5:

$$\frac{7.23 \cdot 10^{14} b + 8.41 \cdot 10^{13} c + 6.68 \cdot 10^{15}}{44157083551} = 156204.27$$

$$7.23 \cdot 10^{14} b + 8.41 \cdot 10^{13} c + 6.68 \cdot 10^{15} = 6.90 \cdot 10^{15}$$

$$7.23 \cdot 10^{14} b = -8.41 \cdot 10^{13} c + 2.21 \cdot 10^{14}$$

$$b = \frac{-8.41 \cdot 10^{13} c + 2.21 \cdot 10^{14}}{7.23 \cdot 10^{14}} \rightarrow 7$$

Step 6, substitute $b = \frac{-8.41 \cdot 10^{13} c + 2.21 \cdot 10^{14}}{7.23 \cdot 10^{14}}$ back into equation 6:

$$\frac{3.59 \cdot 10^{13} + 9.0 \cdot 10^{12} \left(\frac{-8.41 \cdot 10^{13} c + 2.21 \cdot 10^{14}}{7.23 \cdot 10^{14}}\right) + 1.0 \cdot 10^{11} c}{44157083551} = 871.11$$

$$\frac{1.54 \cdot 10^{22} c + 7.71 \cdot 10^{26}}{8.86 \cdot 10^{23}} = 871.11 \rightarrow 8$$ ← simplify

Step 7, isolating variable c from equation 8:

$$\frac{1.54 \cdot 10^{22} c + 7.71 \cdot 10^{26}}{8.86 \cdot 10^{23}} = 871.11$$

$$c = \frac{7.93 \cdot 10^{23}}{1.54 \cdot 10^{22}}$$

$$\underline{c = 51.47742}$$

Step 8, to find variable b, substitute $c = \frac{7.93 \cdot 10^{23}}{1.54 \cdot 10^{22}}$ into equation 7:

$$b = \frac{-8.41 \cdot 10^{13} c + 2.21 \cdot 10^{14}}{7.23 \cdot 10^{14}}$$

$$b = \frac{-8.41 \cdot 10^{13} \left(\frac{7.93 \cdot 10^{23}}{1.54 \cdot 10^{22}}\right) + 2.21 \cdot 10^{14}}{7.23 \cdot 10^{14}}$$

$$\underline{b = -0.2925}$$

Step 9, to find variable a, substitute $c = \frac{7.93 \cdot 10^{23}}{1.54 \cdot 10^{22}}$ and $b = -0.2925$ into equation 4:

$$a = \frac{28526926.75 - 234030815b - 1256719c}{44157083551}$$

$$a = \frac{28526926.75 - 234030815(-0.2925) - 1256719\left(\frac{7.93 \cdot 10^{23}}{1.54 \cdot 10^{22}}\right)}{44157083551}$$

$$\underline{a = 0.00073}$$

Step 10, determining the final equation:

$$\because a = 0.00073, b = -0.2925, c = 51.47742$$

$$\underline{\therefore f(w) = 0.00073w^2 - 0.2925w + 51.47742}$$

Now after finding the quadratic equation for the curve of best fit, I will calculate the correlation coefficient (r) value for this data set.

Step 1, substituting the variable values from *Table 4*:

$$\sum x = 6851 \qquad n = 38 \qquad \sum xy = 156204.27$$

$$\sum y = 871.11 \qquad \sum y^2 = 20060.42 \qquad \sum x^2 = 1256719$$

$$r = \frac{38(156204.27) - (6851)(871.11)}{\sqrt{[38(1256719) - (6851)^2][38(20060.42) - (871.11)^2]}}$$

$$r = \frac{5935762.26 - 5967974.61}{\sqrt{(47755322 - 46936201)(762295.96 - 758832.63)}}$$

$$r = \frac{-32212.35}{\sqrt{(819121)(3463.33)}}$$

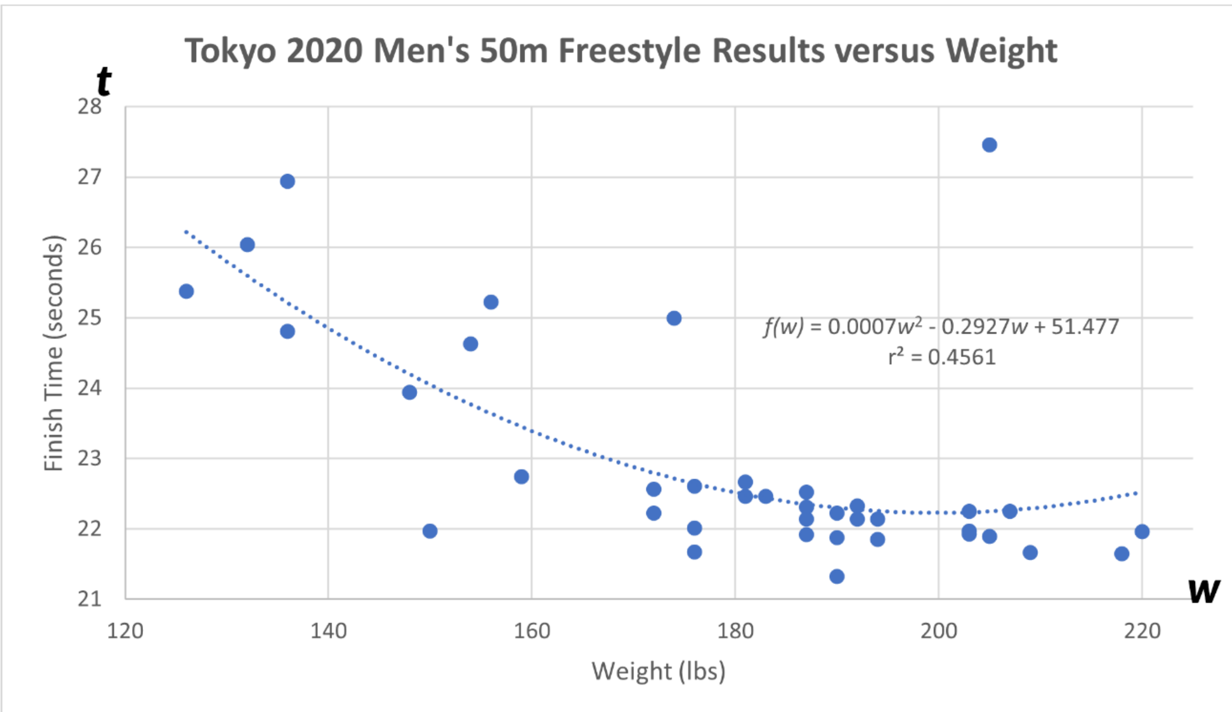
$$r = \frac{-32212.35}{\sqrt{2836884612.78}}$$

$$r = \frac{-32212.35}{53262.41}$$

$$\underline{r \cong -0.60}$$

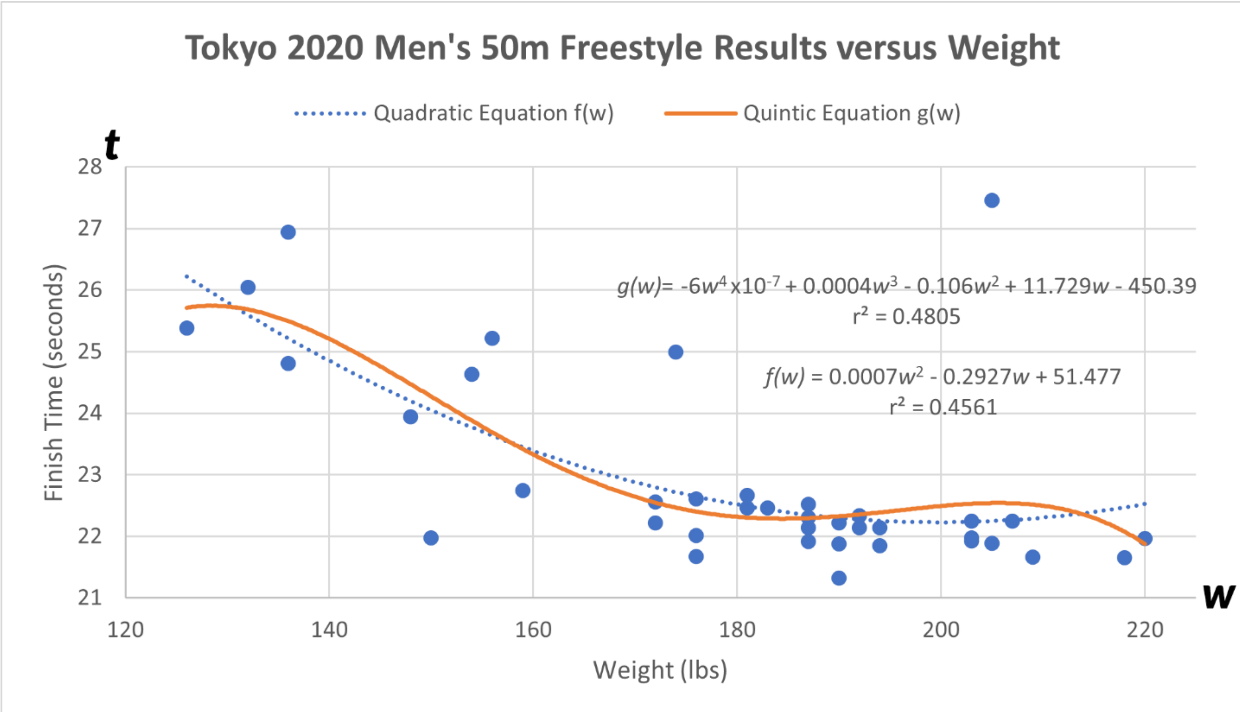
Now knowing the equation for the quadratic function and the correlation coefficient value, I can compare this with the quadratic function and r value given by using technology. From Microsoft Excel, the quadratic equation provided is $f(w) = 0.0007w^2 - 0.2927w + 51.477$ with a correlation of $r = -0.68$ ($r^2 = 0.46$). From the results, I can conclude that the quadratic regression least-squares method is indeed accurate as it gave the same quadratic equation with a slight difference due to rounding and the number of significant digits. Additionally, the correlation coefficient was only slightly off due to rounding. With this, I can graph this model in the form of a scatter plot.

Graph 1: Tokyo 2020 Results versus Weight



From *Graph 1*, it can be observed that the lighter someone weighs, the worse they perform in terms of their finish time. This could be due to the skin friction drag that someone endures when pushing through the water as someone lighter would experience greater resistance. From this, I can also observe that the ideal weight is around 180 to 200 pounds as those swimmers seem to perform the best, excluding the one outlier. This correlation is moderately negative.

Graph 2: Tokyo 2020 Results versus Weight Quintic

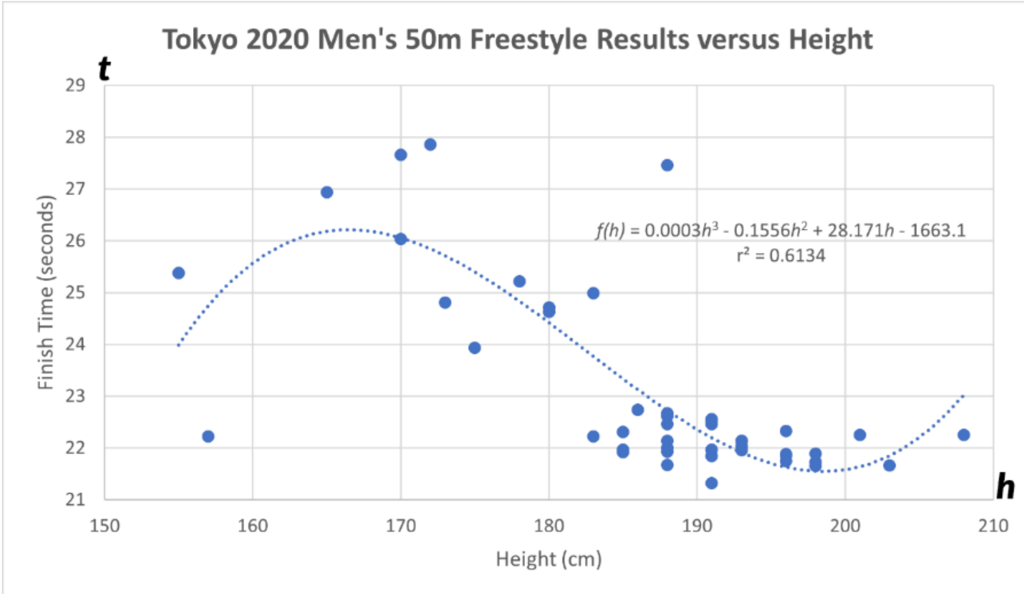


Comparing this to a quintic equation given by Microsoft Excel for the same data set, displayed in *Graph 2*, it can be observed that the quintic function goes up slightly around 205 pounds on the w axis. This is because the quintic function is trying to include the outlier, which would make the curve of best fit less accurate shall I had not used the quadratic regression least-squares method.

Section IV Part 2: Height Correlation

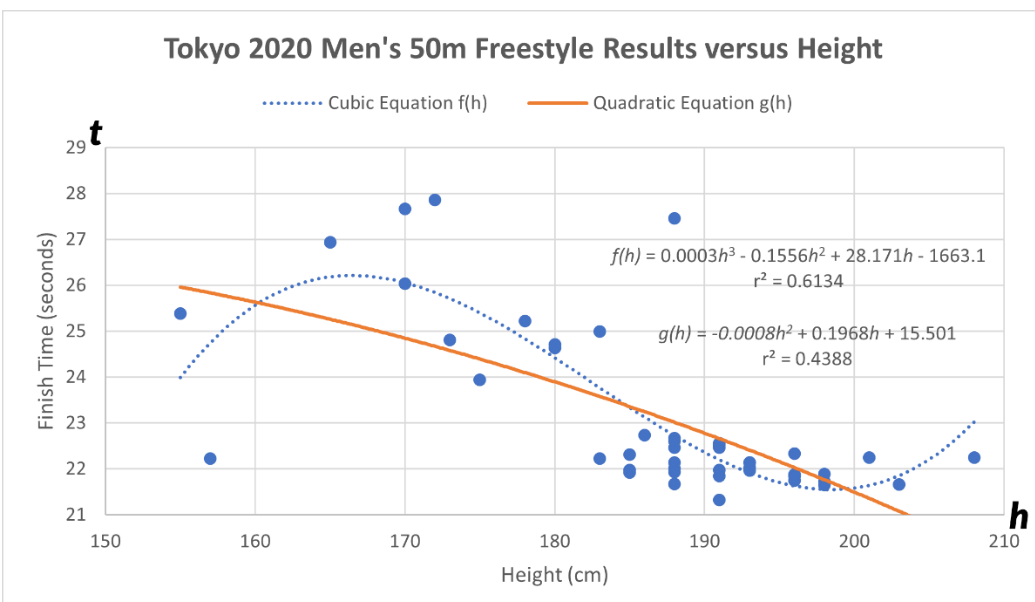
Now I will be finding the correlation between a swimmer's height (h), in centimeters, and their finish time (t) in seconds. For this scatter plot, a cubic polynomial equation best fits this set of data. Thus, I will be using Microsoft Excel to find the cubic equation and justifying why it models better than a quadratic equation. From Microsoft Excel, $f(h) = 0.003h^3 - 0.1556h^2 + 28.171h - 1663.1$ with $r = -0.7832$ ($r^2 = 0.6134$). Below is a scatter plot displaying the correlation of height versus swim finish time.

Graph 3: Tokyo 2020 Results versus Height



From *Graph 3*, it can be observed that most of the swimmers have a height between 180 and 200 centimeters. This shows that being taller does give someone an advantage against shorter competitors as they will have longer arm strokes. Nevertheless, someone short can still perform well. This correlation is moderately negative, just like *Graph 1*. Comparing this cubic function to a quadratic it is apparent that a cubic function models this much better than the quadratic as shown in *Graph 4*.

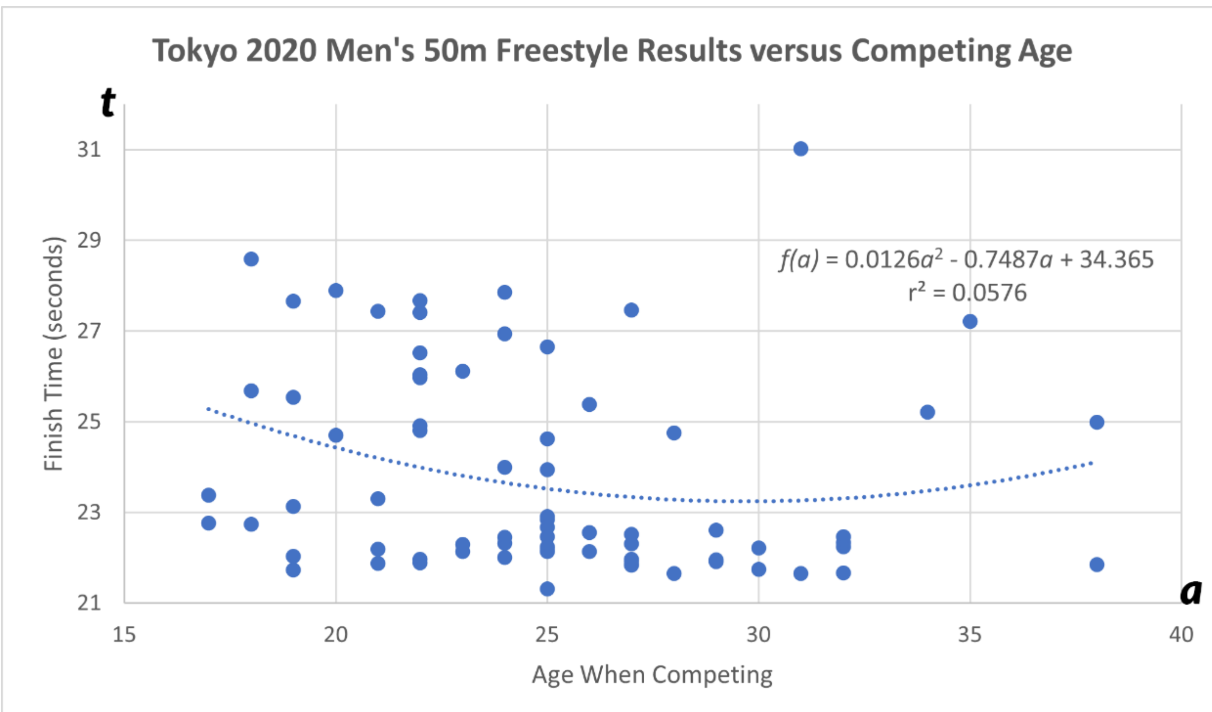
Graph 4: Tokyo 2020 Results versus Height Quadratic



Section IV Part 3: Age Correlation

To find the correlation between a swimmer's age (a), in years, and their finish time (t), in seconds, I will be using the same steps from before to plot a curve of best fit. Through Microsoft Excel, it was determined that a quadratic equation best fits this data set and is $f(a) = 0.0126a^2 - 0.7487a + 34.365$ with $r = -0.24$ ($r^2 = 0.0576$). Below is a scatter plot displaying the correlation of age versus swim finish time.

Graph 5: Tokyo 2020 Results versus Age



From *Graph 5*, it can be observed that there is practically no correlation, hence the weak correlation coefficient of -0.24 and r^2 being close to zero. From this, I can conclude that a swimmer's age has little to no impact on their swimming capabilities and their performance. It is generally perceived that as someone's age increases, the weaker they get thus affecting their swimming abilities; however, this finding would disprove that. It is worth noting that past the age of 60, body muscle mass declines hence why there are no competitors past the age of 38 as they retire.

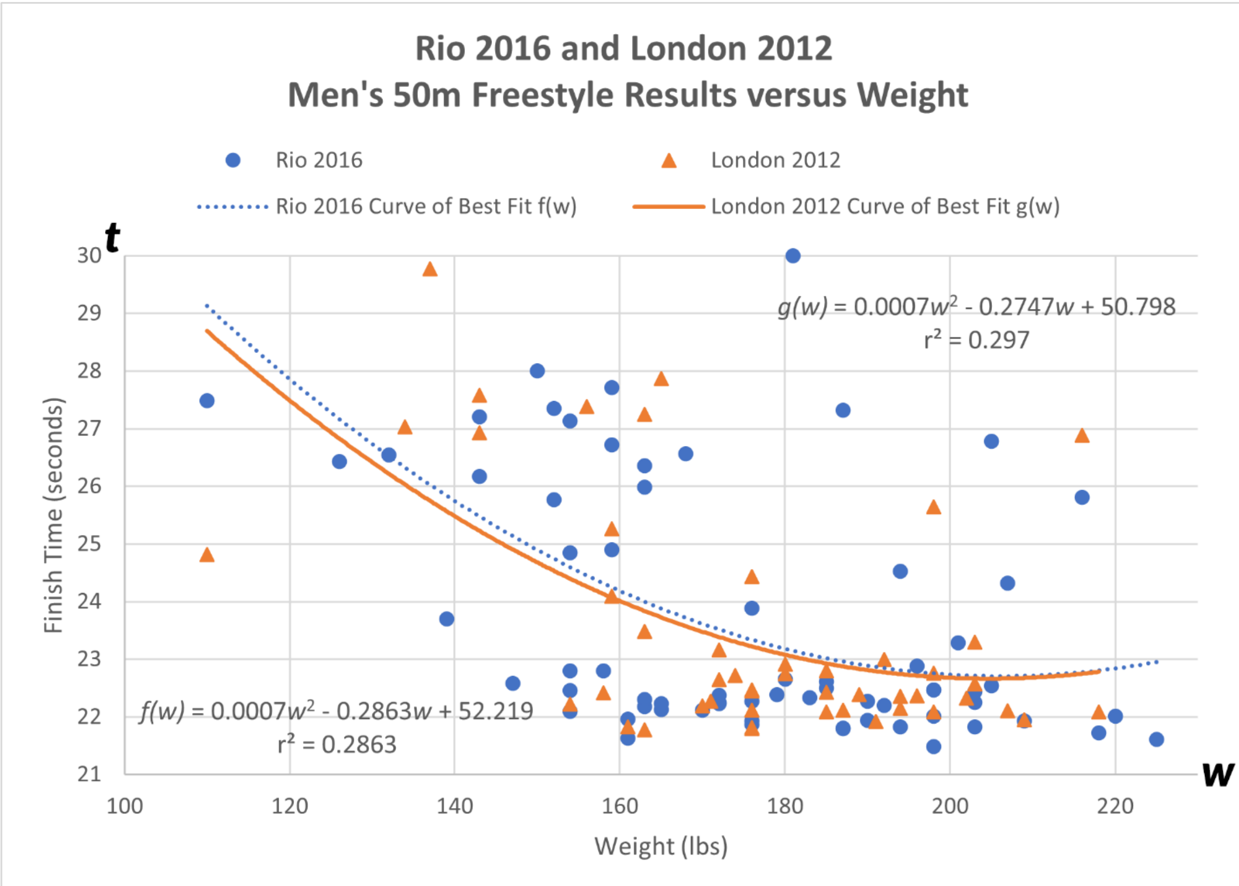
Section V: Extended Analysis of Data with Previous Olympic Games

To extend on the analysis from *Section IV* on page 5, I can use the same method to find the correlation that weight, height, and age have on a swimmer's finish time from previous Olympic Games. I can then compare these correlations to look for any reoccurring trends between the games. Once again, I will be using data collected from ESPN for Rio 2016 and London 2012 for the same swimming event, men's 50-meter freestyle. This data is shown in *Table 2* and *Table 3* from the Appendix. In addition to finding the correlation that weight, height, and age have on a swimmer's finish time, I can find the mode, mean, median, and inter-quartile range (IQR) of each data set to make additional observations. Here are the following formulas:

Mode is the value in the data set that occurs the most frequently.	
Mean is the average of the terms.	$Mean\ (\bar{x}) = \frac{sum\ of\ all\ x\ terms}{total\ number\ of\ x\ terms}$
Median is the middle term in the range.	$Median = \frac{total\ number\ of\ terms\ +\ 1}{2}\ term$
Q_1 is the median of the lower half of the range.	$Q_1 = \frac{number\ of\ terms\ in\ lower\ half + 1}{2}\ term$
Q_3 is the median of the upper half of the range.	$Q_3 = \frac{number\ of\ terms\ in\ upper\ half + 1}{2}\ term$
IQR represents the spread of the middle half in the data.	$IQR = Q_3 - Q_1$

Section V Part 1: Weight Analysis

Graph 6: Rio 2016 and London 2012 Results versus Weight



Comparing *Graph 6* with *Graph 1* it is noticeable that the models are directed in a negative correlation. I do notice that many swimmers finish between 21 and 23 seconds but with a weight ranging from 150 to 220 pounds consistently between the three models. Thus, I can conclude that weight does have a factor, however, it is not of great significance, proven by the r^2 value being below 0.5.

Utilizing the formulas above, I can calculate the mode, mean, median, Q_1, Q_3 and IQR for athlete weight in Tokyo 2020 using the data from *Table 1*. Below are the terms ordered from least to greatest.

126, 132, 136, 136, 148, 150, 154, 156, 159, 172, 172, 174, 176, 176, 176, 181, 181, 183, 187, 187, 187, 187, 190, 190, 190, 192, 192, 194, 194, 203, 203, 203, 205, 205, 207, 209, 218, 220

Finding mode: Mode = 187	Finding mean: $Mean\ (\bar{x}) = \frac{6851}{38}$ $\underline{Mean\ (\bar{x}) \cong 180.3}$

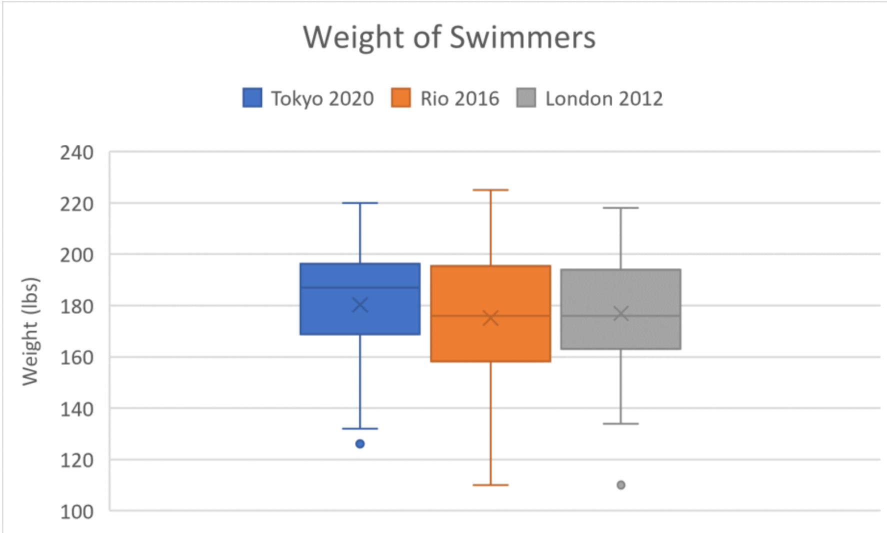
Finding median: $Median = \frac{38\ +\ 1}{2}\ term$ $Median = 19.5\ term$	← the average of the 19th and 20th terms	$Median = \frac{187 + 187}{2}$ $\underline{Median = 187}$
Finding Q_1: $Q_1 = \frac{19.5 + 1}{2}\ term\ of\ lower\ half$ $Q_1 = 10.25\ term\ of\ lower\ half$	the average of the 10th and 11th ← terms of lower half	$Q_1 = \frac{172 + 172}{2}$ $\underline{Q_1 = 172}$
Finding Q_3: $Q_3 = \frac{19.5 + 1}{2}\ term\ of\ upper\ half$ $Q_3 = 10.25\ term\ of\ upper\ half$	the average of the 10th and ← 11th terms of upper half	$Q_3 = \frac{194 + 194}{2}$ $\underline{Q_3 = 194}$
Finding IQR: $IQR = 194 - 172$		$\underline{IQR = 22}$

Using the same process, I can find the mode, mean, median, and IQR for Rio 2016 and London 2012 by using the data from *Table 2* and *Table 3*, below is a table to organize this information. With this, I can make a box whisker plot to display this information and compare it side by side.

Table 5: Summarizing Weight Data

	Mode	Mean	Median	Q_1	Q_3	IQR
Rio 2016	154	175.1	176	159	195	36
London 2012	176	176.9	176	163	194	31

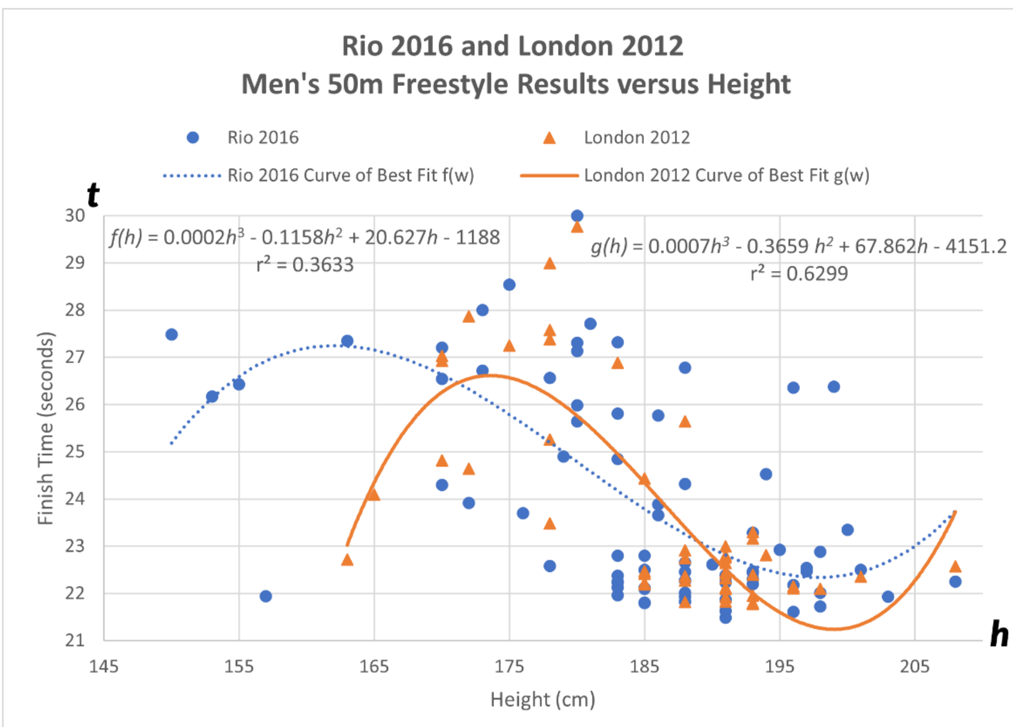
Figure 1: Weight of Swimmers



From the box whisker plot in *Figure 1*, I can observe that the IQR ranges of Tokyo 2020, Rio 2016, and London 2012 are all generally between 160 and 200 pounds. This means that most of the swimmers weigh 160 to 200 pounds, with a few outliers which was expected.

Section V Part 2: Height Analysis

Graph 7: Rio 2016 and London 2012 Results versus Height

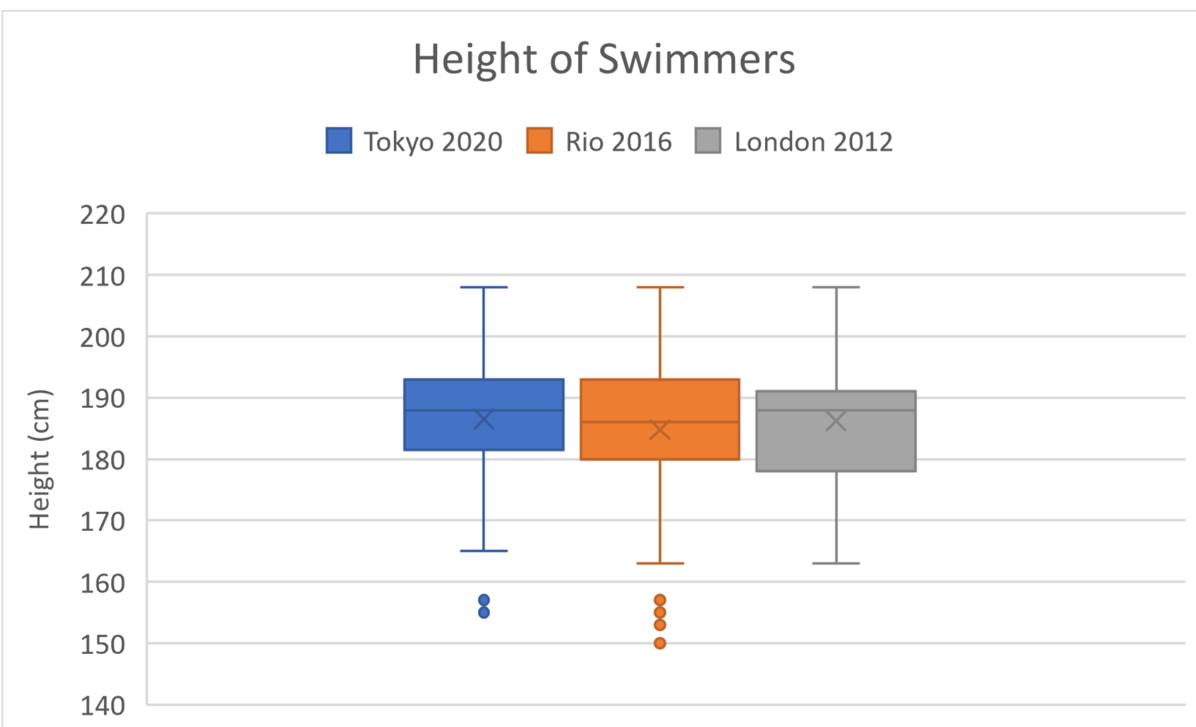


From *Graph 3* and *Graph 7*, I see that the general trend is that the taller the swimmer is, the better they tend to perform. Therefore, apart from a few outliers, due to the r^2 value from Tokyo 2020 and London 2012 being around 0.6, I can conclude that height does have a slight significance on your swimming performance, which logically makes sense as the taller someone is the longer their arm strokes are. This gives them a slight advantage as they can theoretically move more water and touch the end of the pool faster. By using the same steps to calculate the center of data and measure of spread, I will now do the same process with height for Tokyo 2020, Rio 2016, and London 2012 by using the data from *Table 1*, *Table 2*, and *Table 3*.

Table 6: Summarizing Height Data

	Mode	Mean	Median	Q1	Q3	IQR
Tokyo 2020	188	186.5	188	183	193	10
Rio 2016	188	184.8	186	180	193	13
London 2012	191	186.2	188	179	191	12

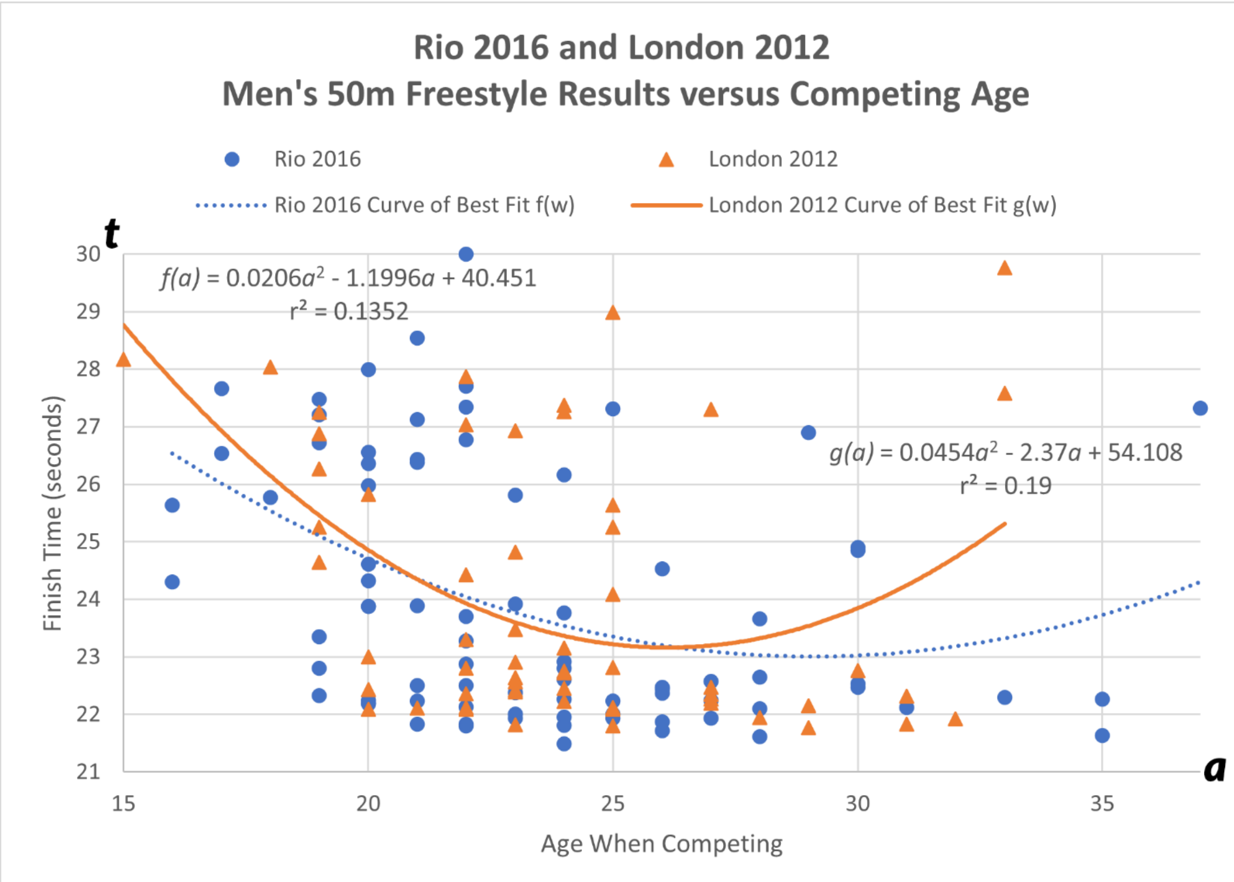
Figure 2: Height of Swimmers



From *Figure 2*, it is observed that the IQR ranges for all three Olympic Games are generally between 180 and 195 centimeters. This means that most of the swimmers have a height between 180 to 195 centimeters. Again, with a couple of outliers since they fall outside of the "whiskers" of the box plot.

Section V Part 3: Age Analysis

Graph 8: Rio 2016 and London 2012 Results versus Age

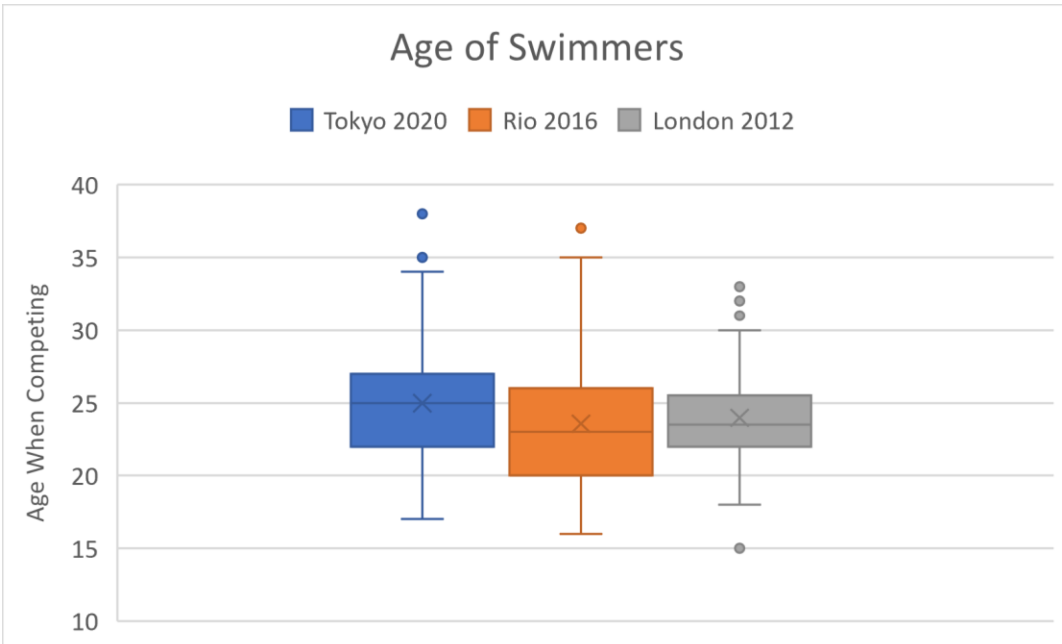


From *Graph 5* and *Graph 8* it is evident that there is no correlation whatsoever, as shown by the r^2 value being below 0.2 which is weak. From the r^2 value, at most 20% of swimmers would have their swimming time affected by their age. This goes to show that someone's age does not determine their swimming capabilities as previously stated in *Section IV Part 3* on page 13. Now I will perform the same process to generate a box whisker plot for the age of the swimmers at the time they were competing.

Table 7: Summarizing Age Data

	Mode	Mean	Median	Q1	Q3	IQR
Tokyo 2020	25	25	25	22	27	5
Rio 2016	22	23.6	23	20	26	6
London 2012	23	24	23.5	22	25	3

Figure 3: Age of Swimmers



From *Figure 3*, it is observed that the general IQR range for the three Olympic Games is between 22 and 26 years old with some outliers older or younger than that. This shows that for most swimmers their peak swimming performance is between the ages of 22 to 26 years old and typically retire before the age of 40, hence the oldest swimmer being 38 years old. Another observation would be that the IQR box for Tokyo 2020 is higher than both Rio 2016 and London 2012. This is most likely because the Tokyo 2020 Olympics took place in 2021 instead of 2020 due to the COVID-19 global pandemic. This means that all the swimmers were an extra year older than from 2020.

Section VI: Limitations and Uncertainties

Throughout the conduction of my investigation, many limitations and uncertainties were encountered that would have altered my results and affected the accuracy of my analyses and conclusions. Foremost, when collecting my data some of the data points for several athletes could not be found despite my best attempt at searching on Google. Without some of these data points, some of my tables, such as *Table 1: Tokyo 2020 Heats Results*, were limited as they had fewer data points than the others. This led to a smaller sample size, potentially altering the models and analyses. Additionally, there would be uncertainties in the scales used to weigh the athletes, the measuring tapes to get the height of the athletes, and in the stopwatches used when timing the swimmers, especially if these measurements were performed by humans as humans will make mistakes and have a reaction time that cannot be mitigated. Secondly, when doing my calculations, I would sometimes round my answer which would lower the accuracy of the calculation. Furthermore, there are added uncertainties when it comes to swimming itself. Skin friction drag from the water, the angle that the swimmer dived into the water at, the temperature of the water, etc.

Section VII: Conclusion

From my investigation, I have concluded that age does not affect someone's swimming capabilities at a rather young age. Many people believe that as someone's age increases, so does their strength and athletic ability. The peak swimming performance is between the ages of 22 to 26 years old with no swimmers competing past the age of 38 due to their decreasing muscle mass and athletic abilities. Furthermore, weight and height do have a slight significance but nothing drastic. This proves my hypothesis correct and is logically reasonable as the taller someone is the longer their arm strokes are which would give them an advantage. Most of the athletes weigh between 160 to 200 pounds, are between 180 to 195 centimeters tall, and are between 22 to 26 years old. This has been consistent throughout the previous three Olympic Games at Tokyo 2020, Rio 2016, and London 2012. The information discovered in my investigation can be used to help train future athletes who wish to compete in the men's 50-meter freestyle by better optimizing their training routine, diet, and workouts. Moreover, the type of math used in this investigation could also be applied to other sports and sporting events, not limited to just swimming. For example, finding the correlation between height and the number of goals scored in soccer. As a swimmer myself, this investigation was extremely interesting and fun as I learned many new things that I can apply to my swim sessions and future competitions.

Section VIII: References

BBC. (2021, July 30). *Tokyo Olympics: Swimming - Men's 50m Freestyle Results - BBC Sport*. BBC News. Retrieved October 21, 2021, from https://www.bbc.com/sport/olympics/58022287.

ESPN. (2012, August 2). *2012 Summer Olympics Results - Swimming*. ESPN. Retrieved October 21, 2021, from https://www.espn.com/olympics/summer/2012/results/_/event/145/discipline/39.

ESPN. (2016, August 11). *2016 Summer Olympics Results - Swimming*. ESPN. Retrieved October 21, 2021, from https://www.espn.com/olympics/summer/2016/results/_/event/145/discipline/39.

ESPN. (2021, July 30). *2020 Summer Olympics Results - Swimming*. ESPN. Retrieved October 21, 2021, from https://www.espn.com/olympics/summer/2020/results/_/event/145/discipline/39.

London Olympic Committee. (2012, August 10). *Official results book Swimming 28 July - 10 August*. 2012 Olympic Games Results - USA Swimming. Retrieved October 21, 2021, from https://www.usaswimming.org/docs/default-source/eventsdocuments/meet-results/international-event-results/olympic-games/2012-olympic-games.pdf.

Section IX: Appendix

Table 1: Tokyo 2020 Heats Results

Name	Competition Finish Place	Age When Competing	Height (cm)	Weight (lbs)	Finish Time (seconds)
Caeleb Dressel	1	25	191	190	21.32
Florent Manaudou	2	31	198	218	21.65
Kristian Gkolomeev	3	28	203	209	21.66
Bruno Fratus	4	32	188	176	21.67
Vladyslav Bukhov	5	19	198		21.73
Thom De Boer	6	30	196		21.75
Jesse Puts	7	27	191		21.84
Brent Hayden	8	38	196	194	21.85
Lorenzo Zazzeri	9	27			21.86
Kliment Kolesnikov	9	21	196	190	21.88
Michael Andrew	11	22	198	205	21.89
Vladimir Morozov	12	29	185	187	21.92
Benjamin Proud	12	27	188	203	21.93
Alberto Mestre	14	29	193	220	21.96
Maxime Grousset	15	22	191	203	21.97
Pawel Juraszek	15	27	185	150	21.97
Meiron Cheruti	17	24	188	176	22.01
Joshua Liendo	18	19	193		22.03
Yu Hexin	19	25	193	192	22.14
Nikola Miljenic	19	23	193	187	22.14
Santo Condorelli	19	26	188	194	22.14
Heiko Gigler	22	25			22.17
Björn Seeliger	23	21			22.19
Ali Khalafalla	24	25	183	172	22.22
Brad Tandy	24	30	157	190	22.22
Ari-Pekka Liukkonen	26	32	208	203	22.25
Maxim Lobanovszkij	26	25	201	207	22.25
Andrej Barna	28	23			22.29
Cameron McEvoy	29	27	185	187	22.31
Gabriel Castaño	30	24			22.32
Konrad Czerniak	31	32	196	192	22.33
Ian Ho	32	24			22.45
Brett Fraser	33	32	188	183	22.46
Dylan Carter	33	25	191	181	22.46

Table 1: Tokyo 2020 Heats Results

Name	Competition Finish Place	Age When Competing	Height (cm)	Weight (lbs)	Finish Time (seconds)
Enzo Martínez	35	27	191	187	22.52
Renzo Tjon A Joe	36	26	191	172	22.56
Sahnoune Oussama	37	29	188	176	22.61
Santiago Grassi	38	25	188	181	22.67
Hwang Sun-Woo	39	18	186	159	22.74
David Popovici	40	17			22.77
Luke Gebbie	41	25			22.84
Emir Muratovic	42	25			22.91
Artur Barseghyan	43	19			23.14
Alaa Maso	44	21			23.30
Nikolas Antoniou	45	17			23.38
Ghirmai Efrem	46	25	175	148	23.94
Filipe Gomes	47	24			24.00
Delgerkhuu Myagmar	48	25	180	154	24.63
Shane Cadogan	49	20	180		24.71
Alassane Seydou Lancina	50	28			24.75
Md Shariful Islam	51	22	173	136	24.81
Puch Hem	52	22			24.91
Marc Dansou	53	38	183	174	24.99
Adama Ouedraogo	54	34	178	156	25.22
Eloi Imaniraguha	55	26	155	126	25.38
Shaquille Moosa	56	19			25.54
Mawupemon Otogbe	57	18			25.68
Troy Pina	58	22			25.97
Santisouk Inthavong	59	22	170	132	26.04
Olimjon Ishanov	60	23			26.12
Mamadou Bah	61	22			26.52
Abdelmalik Muktar	62	25			26.65
Simanga Dlamini	63	24	165	136	26.94
Charly Ndjoume	64	35			27.22
Houssein Gaber Ibrahim	65	22			27.41
Ebrima Buaro	66	21			27.44
Shawn Wallace	67	27	188	205	27.46
Adam Mpali	68	19	170		27.66
Fahim Anwari	69	22			27.67
Phillip Kinono	70	24	172		27.86
Joshua Wyse	71	20			27.90

Table 1: Tokyo 2020 Heats Results

Name	Competition Finish Place	Age When Competing	Height (cm)	Weight (lbs)	Finish Time (seconds)
José Viegas	72	18			28.59
Diosdado Miko	73	31			31.03

Table 2: Rio 2016 Heats Results

Name	Competition Finish Place	Age When Competing	Height (cm)	Weight (lbs)	Finish Time (seconds)
Andrii Govorov	1	24	191	198	21.49
Nathan Adrian	2	28	196	225	21.61
Anthony Ervin	3	35	191	161	21.63
Florent Manaudou	4	26	198	218	21.72
Cameron McEvoy	5	22	185	187	21.80
Vladimir Morozov	6	24	185	187	21.81
Santo Condorelli	7	21	188	194	21.83
Benjamin Proud	7	22	188	203	21.83
Luca Dotto	9	26	191	176	21.87
Bruno Fratus	10	27	188	176	21.93
Kristian Gkolomeev	10	23	203	209	21.93
Brad Tandy	12	25	157	190	21.94
Italo Duarte	13	24	183	161	21.96
Shinri Shioura	14	25	188	198	22.01
Simonas Bilis	14	23	198	220	22.01
Norbert Trandafir	16	28	185	154	22.10
Krisztián Takács	17	31	185	170	22.12
Katsumi Nakamura	18	22	183	165	22.13
Damian Wierling	19	20	196	163	22.18
Yu Hexin	20	20	193	192	22.20
Filip Wypych	21	25	183	165	22.23
Renzo Tjon A Joe	21	21	191	172	22.23
Ali Khalafalla	23	20	183	172	22.25
Ari-Pekka Liukkonen	23	27	208	203	22.25
Sahnoune Oussama	25	24	188	176	22.27
Frédérick Bousquet	25	35	188	190	22.27
George Richard Bovell	27	33	193	163	22.30
Aleksei Bryanskiy	28	19	191	183	22.33
Douglas Erasmus	29	26	183	172	22.37
Ning Zetao	30	23	191	179	22.38
Federico Grabich	31	26	193	203	22.44

Table 2: Rio 2016 Heats Results

Name	Competition Finish Place	Age When Competing	Height (cm)	Weight (lbs)	Finish Time (seconds)
Geoggrey Cheah	32	26	188	154	22.46
Matthew Abood	33	30	197	198	22.47
Odyssefs Meladinis	33	26	193	198	22.47
Yuri Kisil	35	21	201	185	22.50
Pawel Juraszek	35	22	185		22.50
Federico Bocchia	37	30	197	205	22.54
François Heersbrandt	38	27	178	147	22.58
Jasper Aerents	39	24	190	185	22.61
Mario Todorovic	40	28	188	180	22.65
Sidni Hoxha	41	24	185	158	22.80
Ziv Kalontarov	41	19	183	154	22.80
Shane Ryan	43	22	198	196	22.88
Christian Quintero	44	24	195		22.92
Jordan Augier	45	22	193	201	23.28
Jose Quintanilla	46	19	200		23.35
Nahan Mkhitaryan	47				23.50
Abdou Khadre Niane	48	28	186		23.66
Hilal Hemed Hilal	49	22	176	139	23.70
Anthony Barbar	50	24			23.77
Meli Malani	51	20			23.88
Maksim Inic	51	20	186	176	23.88
Ahmad Attellesey	53	21			23.89
Mohmmad Mahfizur Rahman	54	23	172		23.92
Abeiku Jackson	55	16	170		24.30
Adam Viktora	56	20	188	207	24.32
Lum Zhaveli	57	26	194	194	24.53
Faraj Farhan	58	20			24.61
Samson Opuakpo	59	30	183	154	24.85
Dulguun Batsaikhan	60	30	179	159	24.90
Nikolas Sylvester	61	16	180		25.64
Olim Kurbanov	62	18	186	152	25.77
Giordan Harris	63	23	183	216	25.81
Joshua Tibatemwa	64	20	180	163	25.98
Dionisio Augustine II	65	24	153	143	26.17
Billy Scott Irakoze	66	20	196	163	26.36
Tindwendé Thierry Sawadogo	67	21	199		26.38

Table 2: Rio 2016 Heats Results

Name	Competition Finish Place	Age When Competing	Height (cm)	Weight (lbs)	Finish Time (seconds)
Eloi Imaniraguha	68	21	155	126	26.43
Santisouk Inthavong	69	17	170	132	26.54
Moctar Albachir	70	20	178	168	26.56
Ibrahim Nishwan	71	19	173	159	26.72
Shawn Wallace	72	22	188	205	26.78
Osman Kamara	73	29			26.90
Borhane Ahmed Abro	74	21	180	154	27.13
Maël Ambonguilat	75	19	170	143	27.21
Soule Soilihi Athoumane	76	25	180		27.31
Jules Bessan	77	37	183	187	27.32
Amadou Camara	78	22	163	152	27.35
Pap Jonga	79	19	150	110	27.48
Eméric Kpegba	80	17			27.67
Abdelaziz Ahmed	81	22	181	159	27.71
Dienov Koka	82	20	173	150	28.00
Brave Lifa	83	21	175		28.54
Christian Nassif	84	22	180	181	30.00
Frantz Dorsainvil	85	25	170	130	30.86

Table 3: London 2012 Heats Results

Name	Competition Finish Place	Age When Competing	Height (cm)	Weight (lbs)	Finish Time (seconds)
George Richard Bovell	1	29	193	163	21.77
Cesar Cielo Filho	2	25	193	176	21.80
Bruno Fratus	3	23	188	176	21.82
Anthony Ervin	4	31	191	161	21.83
Roland Schoeman	5	32	191	191	21.92
Cullen Jones	6	28	193	209	21.95
Florent Manaudou	7	22	198	218	22.09
Andrey Grechin	7	25	191	185	22.09
Andrii Govorov	7	20	191	198	22.09
James Magnussen	10	21	196	207	22.11
Luca Dotto	11	22	191	176	22.12
Gideon Louw	11	25	191	187	22.12
Brent Hayden	13	29	196	194	22.15
Krisztián Takács	14	27	185	170	22.19
Norbert Trandafir	15	24	185	154	22.22

Table 3: London 2012 Heats Results

Name	Competition Finish Place	Age When Competing	Height (cm)	Weight (lbs)	Finish Time (seconds)
Eamon Sullivan	16	27	188	171	22.27
Stefan Nystrand	17	31	191	202	22.32
Amaury Leveux	18	27	201	194	22.35
Marco Orsi	19	22	188	196	22.36
Adam Brown	20	23	193	189	22.39
Sergey Fesikov	21	23	185	158	22.42
Jasper Aerents	22	20	191	185	22.43
Hanser Garcia	23	24	191	176	22.45
Roy-Allan Saul Burch	24	27	185	176	22.47
Ari-Pekka Liukkonen	25	23	208	203	22.57
Shi Yang	26	23	191	172	22.64
David Dunford	27	24	163	174	22.72
Mario Todorovic	28	24	188	180	22.75
Barry Murphy	29	30	191	198	22.76
Ioannis Kalargaris	30	22	191	185	22.80
Mar Arni Arnason	31	25	194		22.81
Brett Fraser	32	23	188	180	22.91
Kacper Majchrzak	33	20	191	192	23.00
Shehab Younis	34	24	193	172	23.16
Federico Grabich	35	22	193	203	23.30
Luke Hall	36	23	178	163	23.48
Kareem Ennab	37	25	165	159	24.09
Chkyl Camal	38	22	185	176	24.43
Mohammad Mahfizur Rahman	39	19	172		24.64
Kerson Hadley	40	23	170	110	24.82
Adama Ouedraogo	41	25	178	159	25.26
Tepaia Payne	41	19			25.26
Emile Bakale	43	25	188	198	25.64
Franck Olivier Brou Kouassi	44	20			25.82
Tolga Akcayli	45	19			26.27
Giordan Harris	46	19	183	216	26.88
Prasiddha Jung Shah	47	23	170	143	26.93
Ponlpeu Hemthon	48	22	170	134	27.03
Osman Abdourahman Mohamed	49	19	175	163	27.25
Mohamed Elkhedr	50	24			27.26
Ching Maou Wei	51	27			27.30
Jackson Niyomugabo	52	24	178	156	27.38

Table 3: London 2012 Heats Results

Name	Competition Finish Place	Age When Competing	Height (cm)	Weight (lbs)	Finish Time (seconds)
Ganzi Mugula	53	33	178	143	27.58
Edingue Ekane	54	22	172	165	27.87
Christian Nassif	55	18			28.04
Pathana Inthavong	56	15			28.17
Mulualem Girma	57	25	178		28.99
Godonou Wilfrid Tevoedjre	58	33	180	137	29.77