The dataset **ironman1819.csv** contains data on female finishers of the Lake Placid Ironman Triathlon from 2002 to 2021. The motivation for this data analysis is to explore the relationships between swim times, bike times, and run times (in minutes) in order to gain insights into the performance patterns of the athletes. By analyzing these relationships, we can understand the interplay between different segments of the race and potentially identify areas of improvement for athletes. For this activity, we will specifically focus on times from finishers in the years 2018 and 2019.

1. Watch the Introductory Video
   1. <https://www.youtube.com/watch?v=zYG2qdJRolM>
2. Graph and describe the distribution for each leg of the Triathlon
   1. Swim Times
   2. Bike Times See Page 3
   3. Run Times
3. Graph the following relationships between the legs of the race.
   1. Swim Times vs. Bike Times
   2. Run Times vs. Bike Times See Page 4
   3. Swim Times vs. Run Times

|  |  |  |  |
| --- | --- | --- | --- |
| **Relationship** | **Correlation Appropriate?** | | **Guess for correlation** |
| Swim Times vs. Bike Times | Yes | No | r = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |
| Run Times vs. Bike Times | Yes | No | r = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |
| Swim Times vs. Run Times | Yes | No | r = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |

1. From Question 3, is a correlation an appropriate calculation for these data? If so, guess the correlation between the two legs of the race.

|  |  |
| --- | --- |
| **Relationship** | **Correlation**  0.5478 |
| Swim Times vs. Bike Times | r = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  0.6556 |
| Run Times vs. Bike Times | r = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  0.3943 |
| Swim Times vs. Run Times | r = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |

1. Now, using technology, calculate the correlations and compare them to your guesses

1. Which relationship has the largest correlation? What is a possible reason why we may see this?

r = 0.6556. A possible explanation for this would be that running and biking require similar cardio and leg strength. They also have similar motions.

1. Which relationship has the smallest correlation? What is a possible reason why we may see this?

r = 0.3843. A possible explanation for this would be that swimming and biking have very different motions. Swimming requires much more arm strength.

1. These data are for finishers of the triathlon. How might these relationships be different for athletes that did not finish or were disqualified?

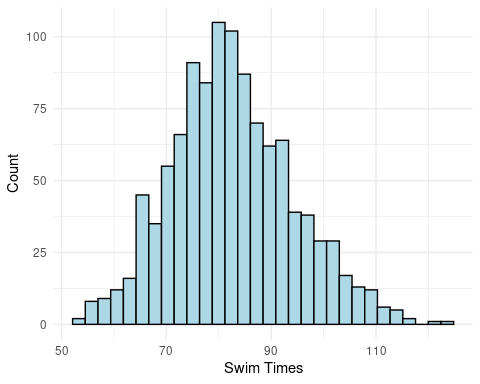
Answers may vary

Sample Response:

1. What do these correlations suggest for athletes about the aspects of their race they want to concentrate on?

Answers may vary

Sample Response: Focus on swimming and one of either biking or running.



2a)

Unimodal and Symmetric

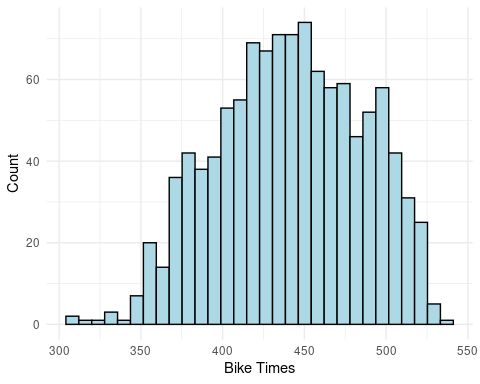
Center: around 80 minutes

Spread: Most values between 55 and 115 minutes

Possible outliers above 120 minutes

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Min | Q1 | Median | Q3 | Max |
| 53.36 | 74.68 | 81.7 | 90.02 | 123.65 |

Mean: 82.58  
SD: 11.625



Unimodal and Symmetric

Center: around 440 minute

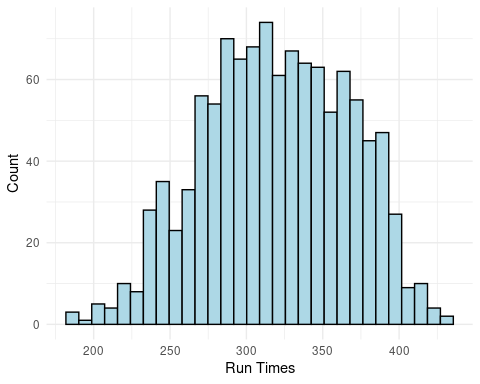
Spread: Most values between 350 and 525 minutes

Possible outliers below 350 minutes

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Min | Q1 | Median | Q3 | Max |
| 304.73 | 410.12 | 411.83 | 475.27 | 533.82 |

Mean: 441.30  
SD: 44.339

2b)



Unimodal and Symmetric

Center: around 300 minutes

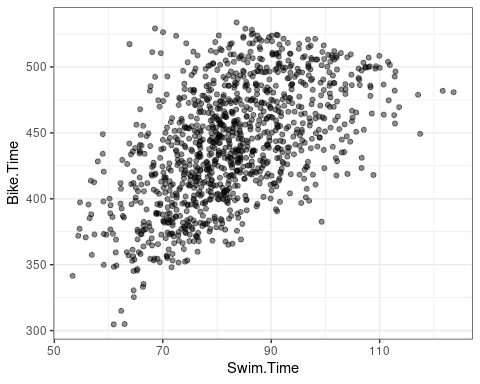
Spread: Most values between 225 and 400 minutes

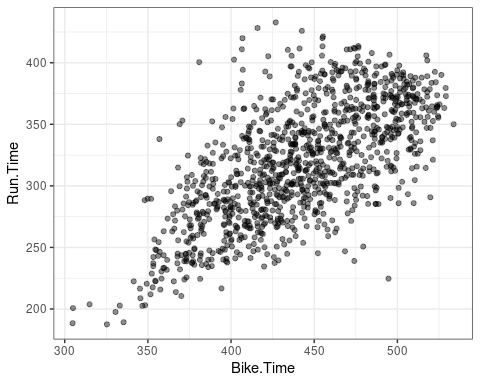
Possible outliers below 200 and above 400 minutes

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Min | Q1 | Median | Q3 | Max |
| 187.52 | 285.38 | 318.8 | 356.15 | 432.75 |

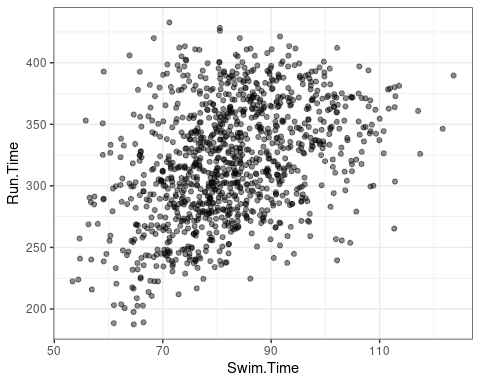
Mean: 319.30  
SD: 47.373

2c)

3a)



3b)

3c)