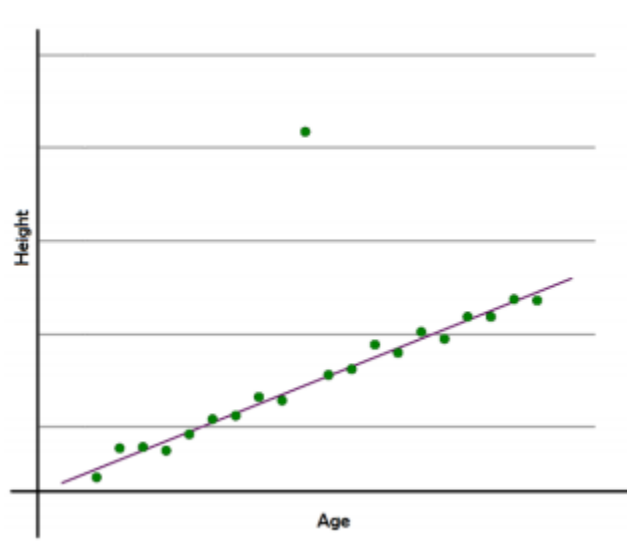**Analytics Worksheet 1**

1. Name the stages of Analytics that a computer scientist typically conducts?

2. Name two examples of data pre-processing.

3. An airport has a self service check in. What pre-processing problems might exist here that a computer scientist may need to account for?

4. Why would sorting data be useful for computer scientists?

5. Suggest (you may have to research this) what is a possible outlier value for a heart rate?

6. We had an example of an outlier being the age 200 entered. The easiest way to deal with this is to delete this piece of data. Suggest an alternative approach to removing this outlier

7. The visualisation below suggests an outlier. Identify this outlier and suggest a solution for dealing with this outlier in pre processing.

8. Name the extension of a text file and a CSV file.

9. **Give one reason why computer scientists often prefer CSV files.**

10. **When using the** open() **method, what is the first argument?**

11. **When using the** open() **method, what is the second argument? List and describe three options for the second argument.**

12. **What does CSV stand for?**

13. **What variable type is returned when reading a CSV file using the** open() **method?**