# Section 1

## Question 1

### Scale

|  |  |
| --- | --- |
| **Attribute** | **Scale** |
| A1\_Score | 0 or 1 |
| A2\_Score | 0 or 1 |
| A3\_Score | 0 or 1 |
| A4\_Score | 0 or 1 |
| A5\_Score | 0 or 1 |
| A6\_Score | 0 or 1 |
| A7\_Score | 0 or 1 |
| A8\_Score | 0 or 1 |
| A9\_Score | 0 or 1 |
| A10\_Score | 0 or 1 |
| age | 46 Distinct Values (Beginning from 17 to 64) having an extreme value of 383 |
| gender | female or male |
| ethnicity | 10 Distinct values, without the duplicate value 'others' |
| jundice | Yes / No |
| autism | Yes / No |
| country\_of\_res | 67 Distinct Countries |
| used\_app\_before | Yes / No |
| result | Scales from 0 to 10 |
| age\_desc | 1 Distinct value |
| relation | 5 Distinct values [Self, Parent, Health care …, Relative, Others] |
| Class/ASD | Yes / No |

### Missing Values

Within the autism dataset of the 21 attributes it has, 3 contains missing values.

|  |  |  |
| --- | --- | --- |
| **Attribute** | **# Missing values** | **% of data** |
| Age | 2 | 0% |
| Ethnicity | 95 | 13% |
| Relation | 95 | 13% |

### Unusual Values

All attributes does not have any unusual values, with the exception of *Ethnicity* and *Age*.

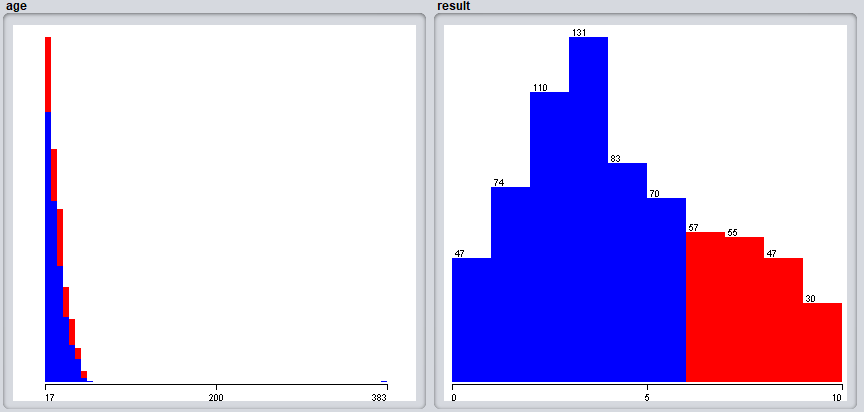
* *Ethnicity* has a double value for ‘***Others***’. One with an uppercase and a lowercase.
* *Age* has an extreme value of **383**.

|  |  |
| --- | --- |
| Others | Others |

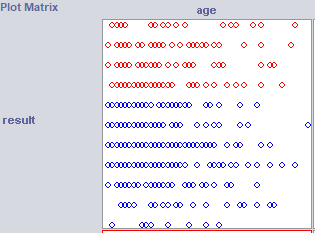
Furthermore, some of the attributes names have not been properly set, an example would be, ***austim*** and ***contry\_of\_res***.

### Distribution (Numeric attributes)

*Result* follows a right skewed distribution, and regarding the *Age*, it can be seen as having a right skewed distribution, but it is unclear since it has an extreme value which affected its distribution.



Upon checking the visualisation for any correlation between these two numeric attributes, I have found that these *Age* and *Result* have not correlation.



## Question 2

Following the instructions on one of the previous Weka labs, I have obtained the Normalised and Standardised datasets.

* Normalised – Using the unsupervised filter Normalize for all numeric attributes.
* Standardised - Using the unsupervised filter Standardised for all numeric attributes.