# Exploration

This XML tags(Normal and Abnormal) includes 8 attributes as shown in the table below.

|  |  |  |  |
| --- | --- | --- | --- |
| Abnormal Concentration | Normal Concentration | Data Type | Explanation |
| biospecimen | biospecimen | Category | Data consist of CSF , Blood , AF , Feces , Urine , etc  In this case , the row will be filtered into blood only |
| concentration\_value | concentration\_value | Range | The concentration value measure in unit (column 3) |
| concentration\_units | concentration\_unit | Category | In case of blood biospecimen , all units are measured in Um |
| patient\_age | subject\_age | Category | Adult , Children , Elderly , Not specified |
| patient\_sex | subject\_sex | Category | Male , Female , Both |
| patient\_information | subject\_information | String | Detail information of the patient/subject |
| comment | comment | String | Detail explanation of the disease, might exist or not |
| references | references | String | Contains reference text and pubmed id |

## **Abnormal Concentration**

### **Patient sex and age distribution**

The patient age and sex are mostly the category of children between 1-13 years old and both gender, as shown in figure 1. Several errors in data input were found (red square) such as typo and space between dash line

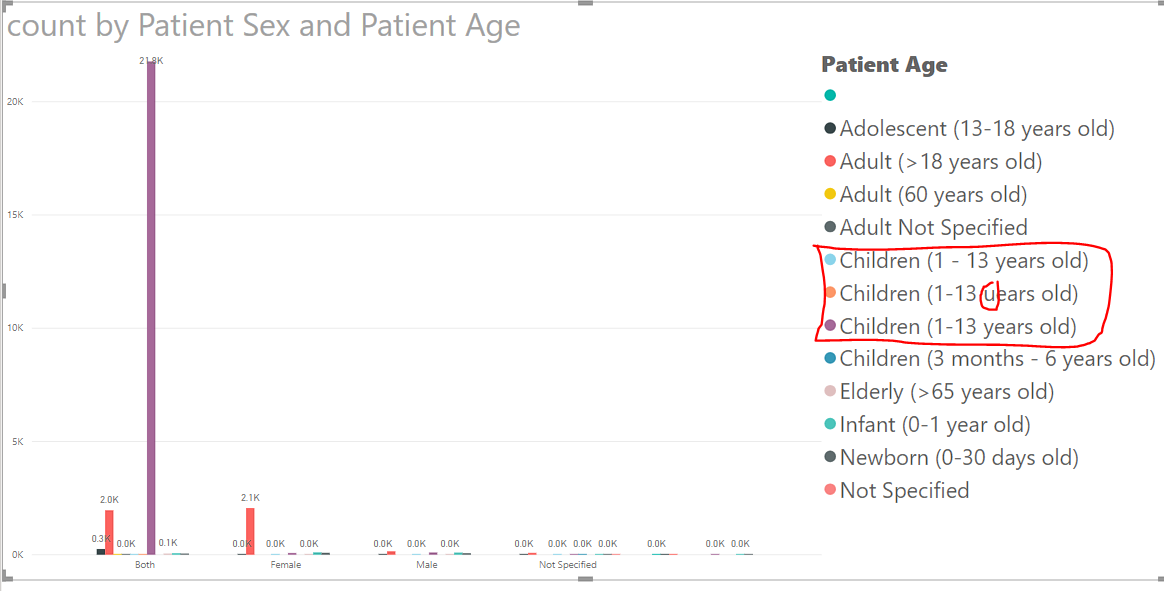


Figure 1 Patient Sex and Age Distribution

## **Normal Concentration**

### **Subject sex and age distribution**

The majority of the subjects are adult(>18 years old) with null and both data value .The categorical value of subject sex is broad but having a little rowcount (as shown in figure 3) .thus the histogram shown below are filtered to row count greater than 100000.

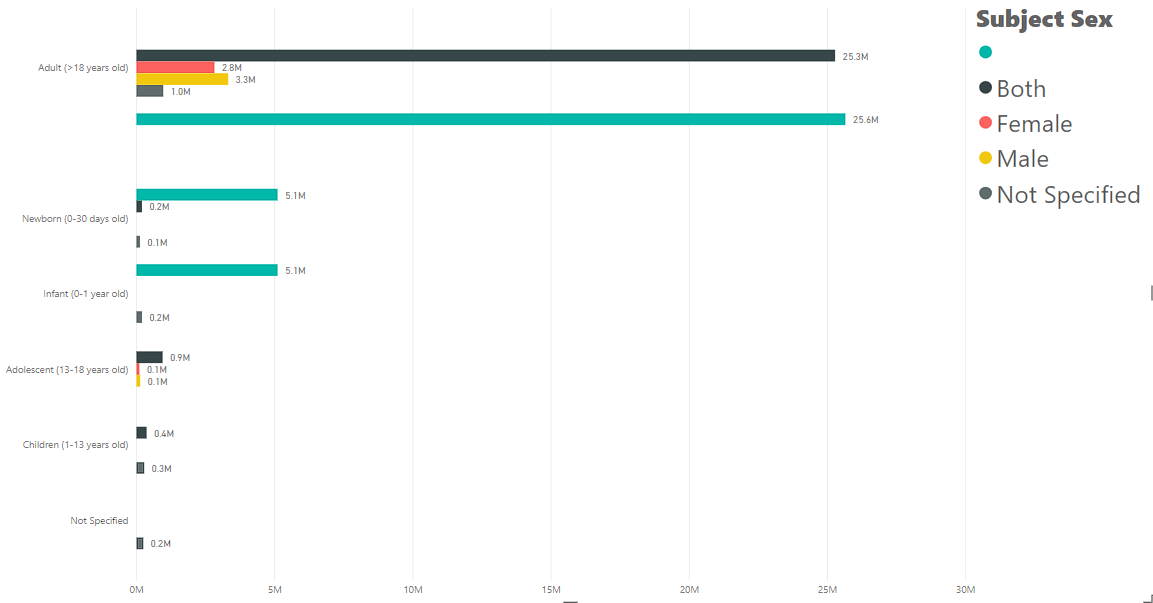


Figure 2 Subject Sex and Age Distribution

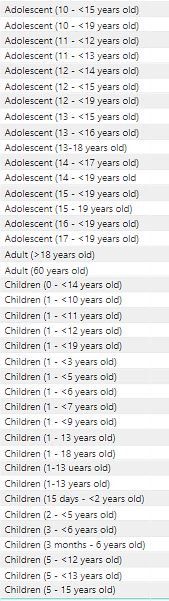


Figure 3 Subject Age