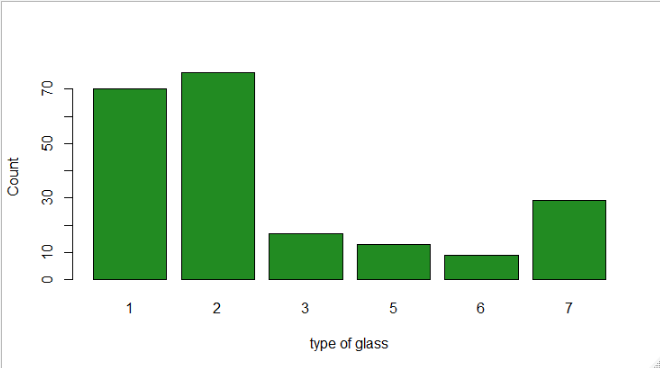
**KNN**

**Example- Model for Glass classification**

**Proportion of glass**



**1 2 3 5 6 7**

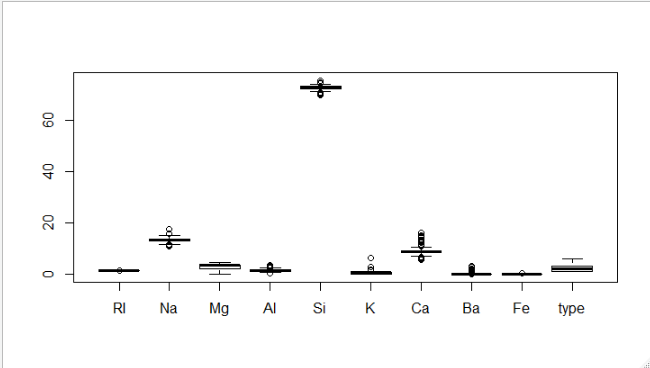
**32.71 35.51 7.94 6.07 4.21 13.55**

**Summary of glass data**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | RI | Na | Mg | Al | Si | K | Ca | Ba | Fe |
| Min | 1.511 | 10.73 | 0 | 0.290 | 69.81 | 0.0000 | 5.430 | 0.000 | 0.00000 |
| 1st Qu | 1.517 | 12.91 | 2.115 | 1.190 | 72.28 | 0.1225 | 8.240 | 0.000 | 0.00000 |
| Median | 1.518 | 13.3 | 3.480 | 1.360 | 71.79 | 0.5550 | 8.600 | 0.000 | 0.00000 |
| Mean | 1.518 | 13.41 | 2.685 | 1.445 | 72.65 | 0.4971 | 8.957 | 0.175 | 0.05701 |
| 3rd Qu | 1.519 | 13.82 | 3.600 | 1.630 | 73.09 | 0.6100 | 9.172 | 0.000 | 0.10000 |
| Max | 1.534 | 17.38 | 4.490 | 3.500 | 75.41 | 6.2100 | 16.190 | 3.150 | 0.51000 |

Mean and Median is approximately same for every variable so chances are very rare to get outlier from data.

**Box Plot for all variable**



**Model Prediction**

**Model 1 🡺 Using caret and K = 5**

**pred\_1**

**1 2 5 6 7**

**1 13 7 0 0 0**

**2 5 13 1 0 0**

**3 3 0 0 0 0**

**5 0 0 1 0 0**

**6 0 1 0 1 0**

**7 1 1 0 1 9**

**Accuracy 🡺 0.6491228**

**Model 2 🡺 Using class and K = 5**

**1 2 3 5 6 7**

**1 14 7 0 0 0 0**

**2 5 14 0 1 1 0**

**3 3 1 1 0 0 0**

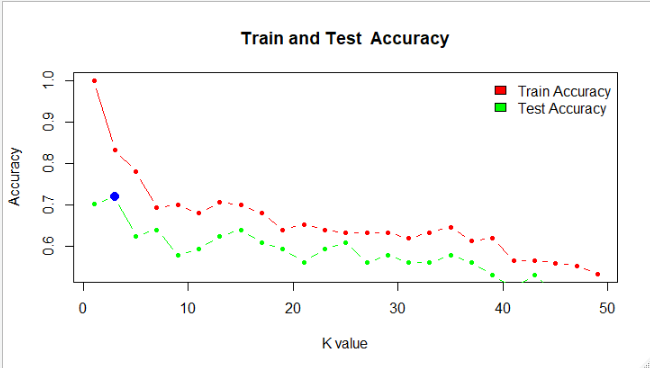
**5 0 2 0 1 0 0**

**6 0 1 0 0 1 0**

**7 1 1 0 0 1 9**

**Accuracy 🡺 0.625**

**Model 🡺 K= All odds from 1 to 50**



**From the above plot it is shown that train accuracy is high for value**

**k= 1 and test accuracy is high for k=3**

**Final Model 🡺 K= 3**

**model\_final**

**1 2 3 5 6 7**

**1 16 5 0 0 0 0**

**2 5 15 0 1 0 0**

**3 3 1 1 0 0 0**

**5 0 0 0 3 0 0**

**6 0 0 0 0 2 0**

**7 1 1 0 0 1 9**

**Accuracy 🡺 0.71875**

**From the above predictions this is best model with accuracy nearly 72% for classification.**