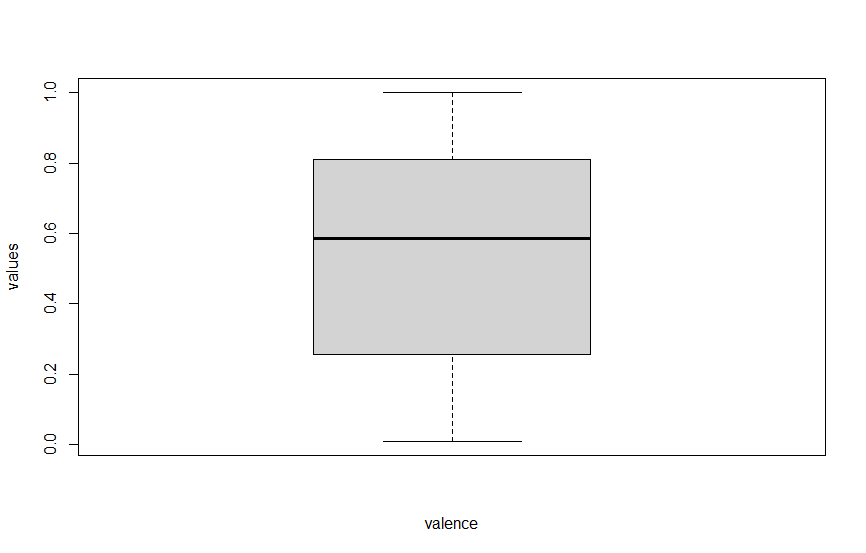
|  |
| --- |
| variant=29 #write down your variant number instead of N  data = read.csv("defence/spotify\_data.csv")  variant\_variable=variant%%10+1  x=data[((variant-1)\*100+1):(variant\*100),variant\_variable]  print(paste("This variant data column is", colnames(data)[variant\_variable]))  boxplot(x,xlab="valence",ylab="values")  ............................................................................................................................. |



* The lower quartile mark between 0.2 and 0.4 (25%)
* The median quartile (50%) mark is between 0.4 and 0.6
* The 75% quartile mark (group 3) is a bit below 0.8
* Half of the values are less than 0.6
* The middle 50% of the scores fall between 0.2 and 0.8
* 75% of the scores are below 0.8
* 25% of the scores are below 0.3
* The lower whisker is between 0.0 and 0.3
* The upper whisker is between 0.8 and 1.0

# Confidence Interval

|  |
| --- |
| variant=29 #write down your variant number instead of N  data = read.csv("defence/air\_quality.csv")  variant\_variable=variant%%12+1  x=data[((variant-1)\*100+1):(variant\*100),variant\_variable]  print(paste("This variant data column is", colnames(data)[variant\_variable]))  # 95% confidence interval for  #the population mean of values if the population variance is unknown  x\_bar=mean(x)  gamma=0.95  alpha=1-gamma  n=length(x)  s=sd(x)  tp=qt(1-alpha/2,n-1)  left=x\_bar-s\*tp/sqrt(n)  left  right=x\_bar+s\*tp/sqrt(n)  right  # 95% confidence interval for  #the population variance of values if the population mean is unknown.  n=length(x)  gamma=0.95  alpha=1-gamma  S2=var(x)  chi2p\_left=qchisq(1-alpha/2,n-1)  chi2p\_right=qchisq(alpha/2,n-1)  left=(n-1)\*S2/chi2p\_left  left  right=(n-1)\*S2/chi2p\_right  right |

* One can be 95% confident that the mean of values will be found in the interval (930.9527, 1025.889)
* One can be 95% confident that the variance of values will be found in the interval (44118.92, 77232.23)

# Hypothesis Testing

|  |
| --- |
| variant=29 #write down your variant number instead of N  data = read.csv("defence/air\_quality.csv")  variant\_variable=variant%%12+1  x=data[((variant-1)\*100+1):(variant\*100),variant\_variable]  print(paste("This variant data column is", colnames(data)[variant\_variable]))  t.test(x,mu=1000,conf.level=0.95,alternative='two.sided') |

H0: mu = 1000

H1: mu != 1000

data: x

t = -0.90202, df = 99, p-value = 0.3692

alternative hypothesis: true mean is not equal to 1000

95 percent confidence interval:

930.9527 1025.8893

sample estimates:

mean of x

978.421

p-value (0.3692) > Alpha(0.05), H0 is not rejected

|  |
| --- |
| variant=29 #write down your variant number instead of N  data = read.csv("defence/air\_quality.csv")  variant\_variable=variant%%12+1  x=data[((variant-1)\*100+1):(variant\*100),variant\_variable]  print(paste("This variant data column is", colnames(data)[variant\_variable]))  t.test(x,mu=950,conf.level=0.95,alternative='two.sided') |

H0: mu < 1000

H1: mu > 1000

data: x

data: x

t = 1.188, df = 99, p-value = 0.2377

alternative hypothesis: true mean is not equal to 950

95 percent confidence interval:

930.9527 1025.8893

sample estimates:

mean of x 978.421

p-value (0.2377) > Alpha(0.05), H0 is not rejected