

## VARIABLES AFFECTING QUALITIES OF BANANAS

**Summary.** This paper investigates the relationship between numeric characteristics of different attributes of bananas and the qualities of bananas. The various strategy used to achieve the goal includes finding measures of spreads and locations of numerical data of different attributes. Graphical representation of data for example normal QQ plot and side-by-side boxplots between attributes are also included in this paper. It is found that ripeness of bananas is the most relevant attributes associated to the quality of bananas.

### 1. INTRODUCTION

The results from this paper is vital as B.Pathare (2023, Volume 134) had mentioned that consumer interest in high fruit quality had increased due to the international food trade. The quality of banana is predicted and classified by optical properties such as absorption, suggested by Hashim (2016, Volume 212). On the other hand, the ripeness of banana can be identified by new feature such as ripening factor as proposed by Mazen (2019). This paper will first analyse the numerical characteristics of a banana, this includes various attributes, including Size, Weight, Sweetness, Softness, Ripeness, Acidity, and Harvest Time. We will then discuss the association between numerical characteristic and the quality of banana.

### 2. METHOD

The data collected contain negative entries as the data are demeaned measures, which means it is already subtracted with mean of each variable from its original value. We find the summary statistics for all variables to summarise the data. A matrix of normal QQ plot and scatterplots are also included, the scatterplot is used to show the pairwise relationship between variables. We also have 2 histograms of Harvest Time, one for “Good” and one for “Bad” bananas. 6 pair of side-by-side boxplots corresponding to the variables are also included.

### 3. RESULTS AND DISCUSSION

```
> summary(Size)
  Min. 1st Qu.  Median    Mean 3rd Qu.
-7.9981 -2.2777 -0.8975 -0.7478  0.6542
  Max.
  7.9708

> summary(Ripeness)
  Min. 1st Qu.  Median    Mean 3rd Qu.
-7.4232 -0.5742  0.9650  0.7811  2.2616
  Max.
  7.2490

> summary(HarvestTime)
  Min. 1st Qu.  Median    Mean 3rd Qu.
-7.5700 -2.1207 -0.9342 -0.7513  0.5073
  Max.
  6.2933

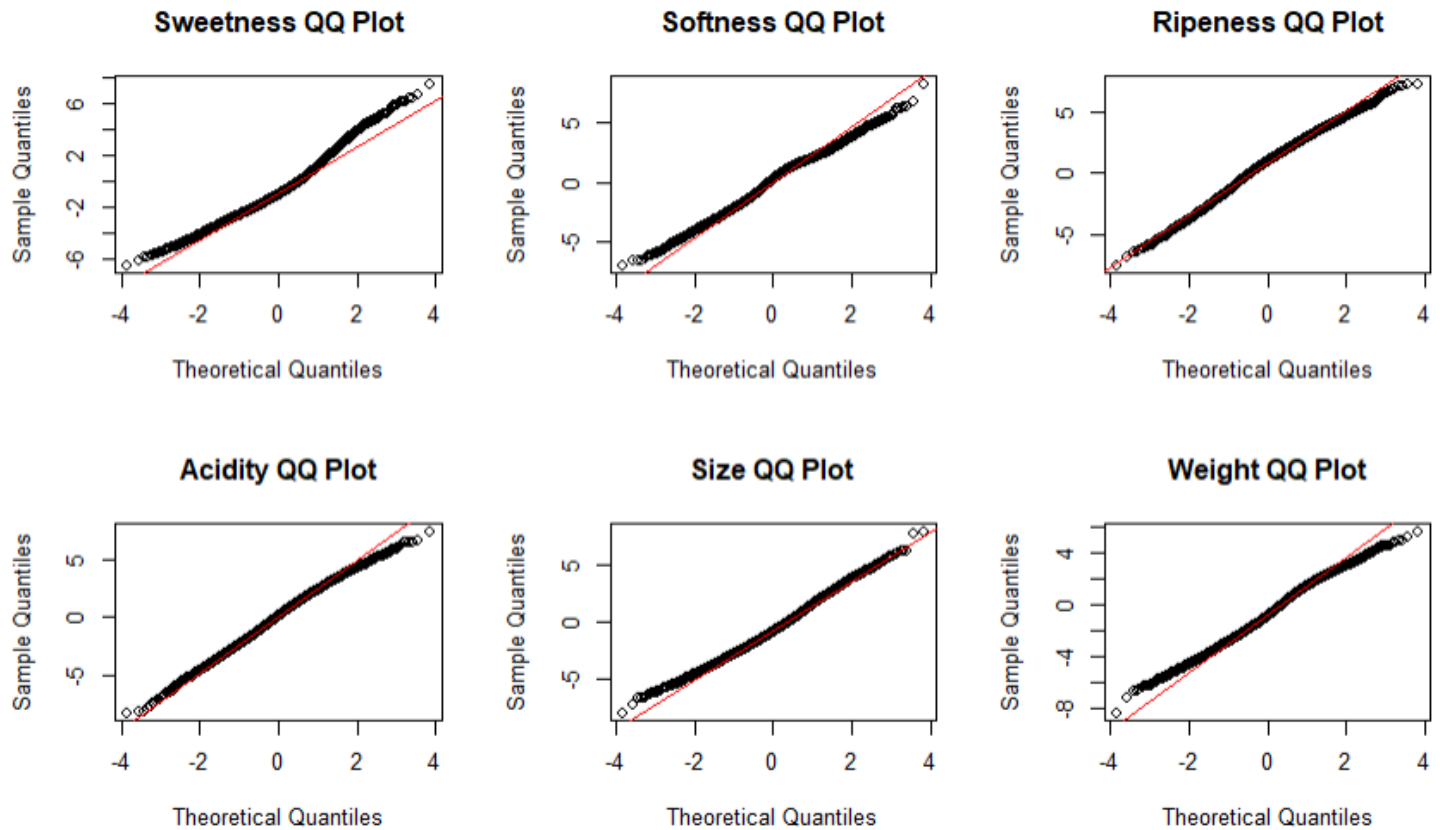
> summary(Weight)
  Min. 1st Qu.  Median    Mean 3rd Qu.
-8.2830 -2.2236 -0.8687 -0.7610  0.7755
  Max.
  5.6797

> summary(Softness)
  Min. 1st Qu.  Median    Mean 3rd Qu.
-6.95932 -1.59046  0.20264 -0.01444  1.54712
  Max.
  8.24155

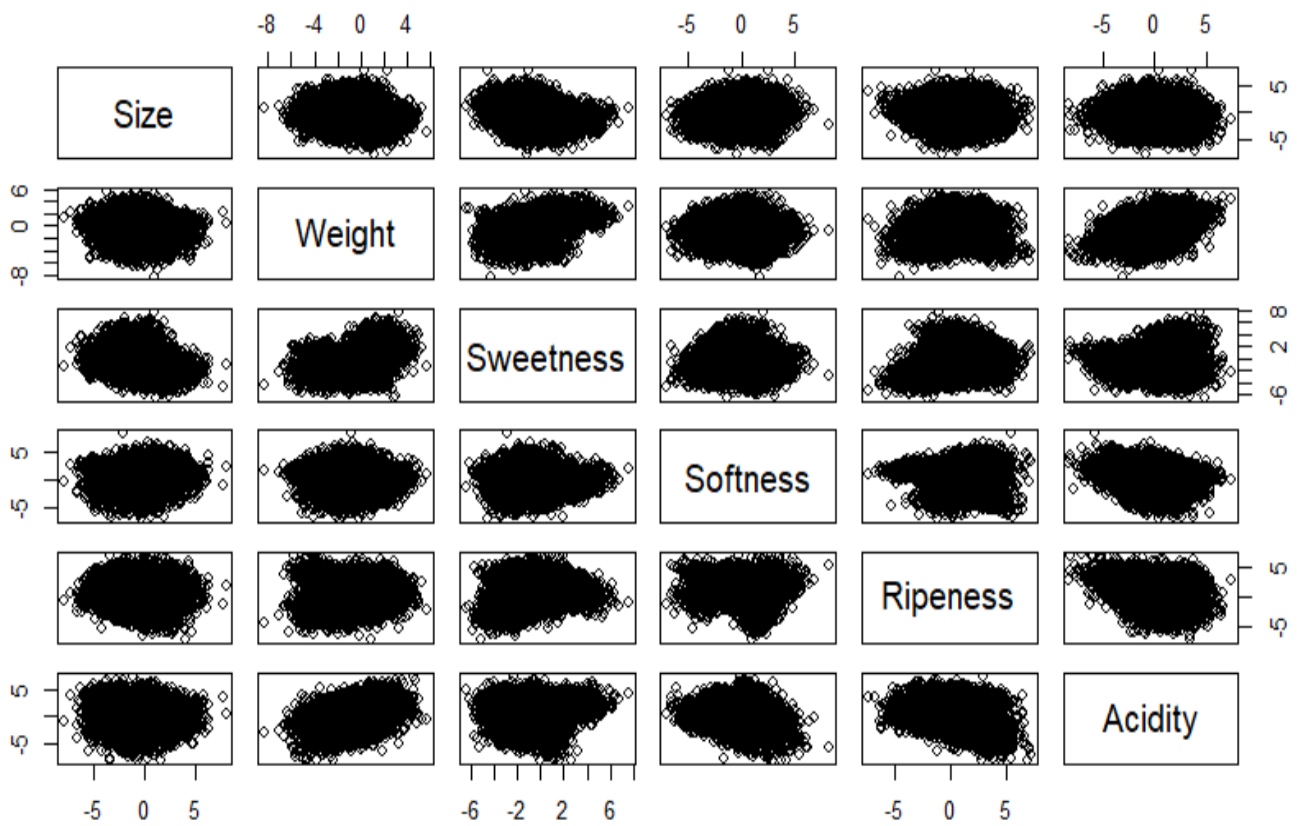
> summary(Acidity)
  Min. 1st Qu.  Median    Mean
-8.226977 -1.629450  0.098735  0.008725
 3rd Qu.    Max.
 1.682063  7.411633

> summary(Sweetness)
  Min. 1st Qu.  Median    Mean 3rd Qu.
-6.4340 -2.1073 -1.0207 -0.7702  0.3110
  Max.
  7.5394
```

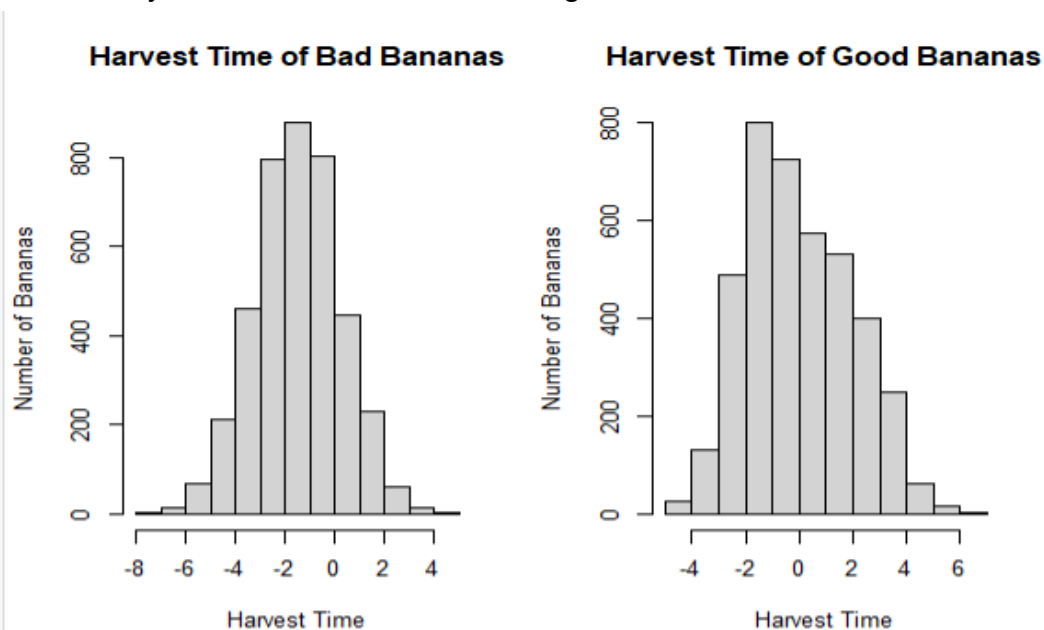
From the data we can see that the value of median and mean of all 7 attributes are almost the same, this shows that the data may have symmetric distribution around the centre. Even if the distribution is not perfectly symmetric, it still suggests that the skewness of the data is relatively mild.



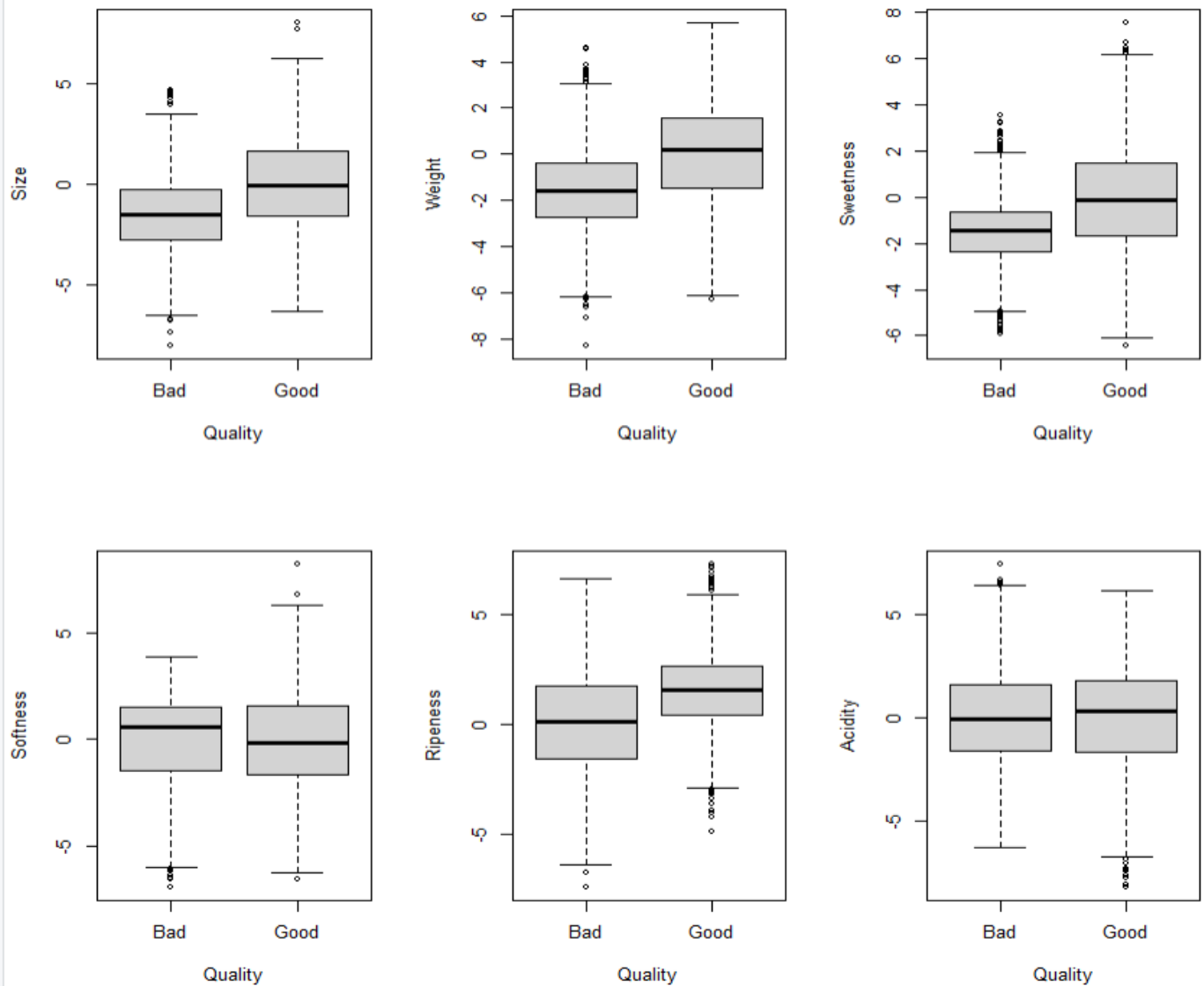
We can observe from the normal QQ plot that the points for sweetness, softness and weight slightly deviate from the diagonal reference line, especially in both tails of the distribution. This indicates that the distribution may be skewed or non-normal. For points in ripeness, acidity, and size QQ plot, the points follow more closely to the reference line, suggesting the distribution to be more likely symmetrically distributed around the mean. One interesting common feature we can observe is that all 6 normal QQ plots have a few outliers on both tails.



Most of the pairwise scatterplots did not show any clear trend, indicating no correlation between the attributes. Exceptionally, the scatterplot between weight and acidity shows a moderate positive correlation whereas scatterplot between softness and acidity shows evidence of weak negative correlation.



There is a slight difference in the harvest time of bananas as shown above, the harvest time of good bananas tends to be earlier.



A clear difference between good and bad bananas can be seen in attributes of size, weight, sweetness, and ripeness. The interquartile range of these attributes tend to be higher for good bananas. As for the aspect of softness and acidity, they have similar interquartile range, though there is a slight difference in the median. One interesting feature found is that the range of data for specific attributes appeared to have large difference, for example range of sweetness for bad bananas is much smaller than the range of good bananas. Outliers can also clearly be seen from the diagram above, which refers to data points that fall beyond the whiskers of the boxplot. From the boxplot, we can conclude that good bananas have bigger size, heavier weight, higher rating of sweetness and higher rating of ripeness.

### Summary of Good Bananas

Size	Weight	Sweetness	Softness
Min. :-6.268382	Min. :-6.26452	Min. :-6.43402	Min. :-6.57027
1st Qu.:-1.599698	1st Qu.:-1.48422	1st Qu.:-1.68648	1st Qu.:-1.66698
Median :-0.073752	Median : 0.21186	Median :-0.14796	Median :-0.13926
Mean : 0.005001	Mean : 0.02153	Mean :-0.03638	Mean :-0.01782
3rd Qu.: 1.637941	3rd Qu.: 1.58320	3rd Qu.: 1.45745	3rd Qu.: 1.56076
Max. : 7.970800	Max. : 5.67969	Max. : 7.53937	Max. : 8.24155
HarvestTime	Ripeness	Acidity	Quality
Min. :-4.952645	Min. :-4.908	Min. :-8.226977	Length:4006
1st Qu.:-1.531200	1st Qu.: 0.411	1st Qu.:-1.669554	Class :character
Median :-0.226019	Median : 1.515	Median : 0.289259	Mode :character
Mean :-0.000519	Mean : 1.521	Mean : 0.006757	
3rd Qu.: 1.501951	3rd Qu.: 2.629	3rd Qu.: 1.758702	
Max. : 6.293280	Max. : 7.249	Max. : 6.159333	

### Summary of Bad Bananas

Size	Weight	Sweetness	Softness	HarvestTime
Min. :-7.9981	Min. :-8.2830	Min. :-5.8931	Min. :-6.95932	Min. :-7.5700
1st Qu.:-2.7776	1st Qu.:-2.7016	1st Qu.:-2.3720	1st Qu.:-1.49176	1st Qu.:-2.6920
Median :-1.5168	Median :-1.5673	Median :-1.4763	Median : 0.59832	Median :-1.4588
Mean :-1.5029	Mean :-1.5459	Mean :-1.5063	Mean :-0.01105	Mean :-1.5043
3rd Qu.:-0.2668	3rd Qu.:-0.3882	3rd Qu.:-0.6506	3rd Qu.: 1.54421	3rd Qu.:-0.3504
Max. : 4.6313	Max. : 4.6281	Max. : 3.5134	Max. : 3.87153	Max. : 4.0969
Ripeness	Acidity	Quality		
Min. :-7.42316	Min. :-6.29234	Length:3994		
1st Qu.:-1.60508	1st Qu.:-1.60854	Class :character		
Median : 0.06912	Median :-0.08343	Mode :character		
Mean : 0.03897	Mean : 0.01070			
3rd Qu.: 1.70940	3rd Qu.: 1.59247			
Max. : 6.62530	Max. : 7.41163			

The summary supports the statement that “good bananas have bigger size, heavier weight, higher rating of sweetness and higher rating of ripeness.” For example, good bananas have a mean size of 0.005001 whereas bad bananas only have a mean size of -1.5029. The difference is quite significant.

The study has also further investigated and found that when the ripeness ratings are above the mean, there is a higher occurrence of good qualities bananas. Among the 4006 good qualities bananas, 2712 bananas have their ripeness rating above the mean.

## 4. CONCLUSION

We can conclude that the ripeness rating has a direct association with the qualities of the bananas. However, the data is limited as there may be other factors which affects the qualities of the bananas, for example geographical location and weather conditions. In order to improve the accuracy of this study, comparison of data collected from different regions can be made and analyse how other factors may impact bananas qualities across different regions.

(806 words)

## REFERENCES

- Al-Dairi, M., Pathare, P. B., Al-Yahyai, R., Jayasuriya, H. P. W., & Al-Attabi, Z. (2023). Postharvest quality, technologies, and strategies to reduce losses along the supply chain of banana: A review. *Trends in Food Science and Technology*, 134, 177–191. <https://doi.org/10.1016/j.tifs.2023.03.003>
- Mazen, F. M. A., & Nashat, A. A. (2019). Ripeness classification of bananas using an artificial neural network. *Arabian Journal for Science and Engineering*, 44(8), 6901–6910. <https://doi.org/10.1007/s13369-018-03695-5>
- Adebayo, S. E., Hashim, N., Abdan, K., Hanafi, M., & Mollazade, K. (2016). Prediction of quality attributes and ripeness classification of bananas using optical properties. *Scientia Horticulturae*, 212, 171–182. <https://doi.org/10.1016/j.scienta.2016.09.045>

## APPENDIX

```
banana.data = read.csv("banana.csv", header = TRUE)
```

```
attach(banana.data)
```

```
summary(Size)
```

```
summary(Weight)
```

```
summary(Sweetness)
```

```
summary(Softness)
```

```
summary(Ripeness)
```

```
summary(Acidity)
```

```
summary(HarvestTime)
```

```
par(mfrow=c(2,3))
```

```
qqnorm(banana.data$Sweetness, main="Sweetness QQ Plot")
```

```
qqline(banana.data$Sweetness, col="red")
```

```
qqnorm(banana.data$Softness, main="Softness QQ Plot")
```

```
qqline(banana.data$Softness, col="red")
```

```
qqnorm(banana.data$Ripeness, main="Ripeness QQ Plot")
```

```
qqline(banana.data$Ripeness, col="red")
```

```
qqnorm(banana.data$Acidity, main="Acidity QQ Plot")
```

```
qqline(banana.data$Acidity, col="red")
```

```
qqnorm(banana.data$Size, main="Size QQ Plot")
```

```
qqline(banana.data$Size, col="red")
```

```
qqnorm(banana.data$Weight, main="Weight QQ Plot")
```

```
qqline(banana.data$Weight, col="red")
```

```
c = data.frame(Size, Weight, Sweetness, Softness,Ripeness, Acidity)
```

```
pairs(c)
```

```
good = subset(banana.data,Quality == "Good")
```

```
hist(good$HarvestTime,main="Harvest Time of Good Bananas",xlab="Harvest Time",ylab  
      = "Number of Bananas" )
```

```
bad = subset(banana.data,Quality == "Bad")
```

```
hist(bad$HarvestTime,main="Harvest Time of Bad Bananas",xlab="Harvest Time",ylab =  
      "Number of Bananas" )
```

```
par(mfrow=c(2,3))
```

```
boxplot(Size~Quality,data = banana.data)
```

```
boxplot(Weight~Quality,data = banana.data)
```

```
boxplot(Sweetness~Quality,data = banana.data)
```

```
boxplot(Softness~Quality,data = banana.data)
```

```
boxplot(Ripeness~Quality,data = banana.data)
```

```
boxplot(Acidity~Quality,data = banana.data)
```

```
summary(good)
```

```
summary(bad)
```

```
x = subset(banana.data,Quality == "Good" & Ripeness >= 0.7811)
```