HW2.bees.R

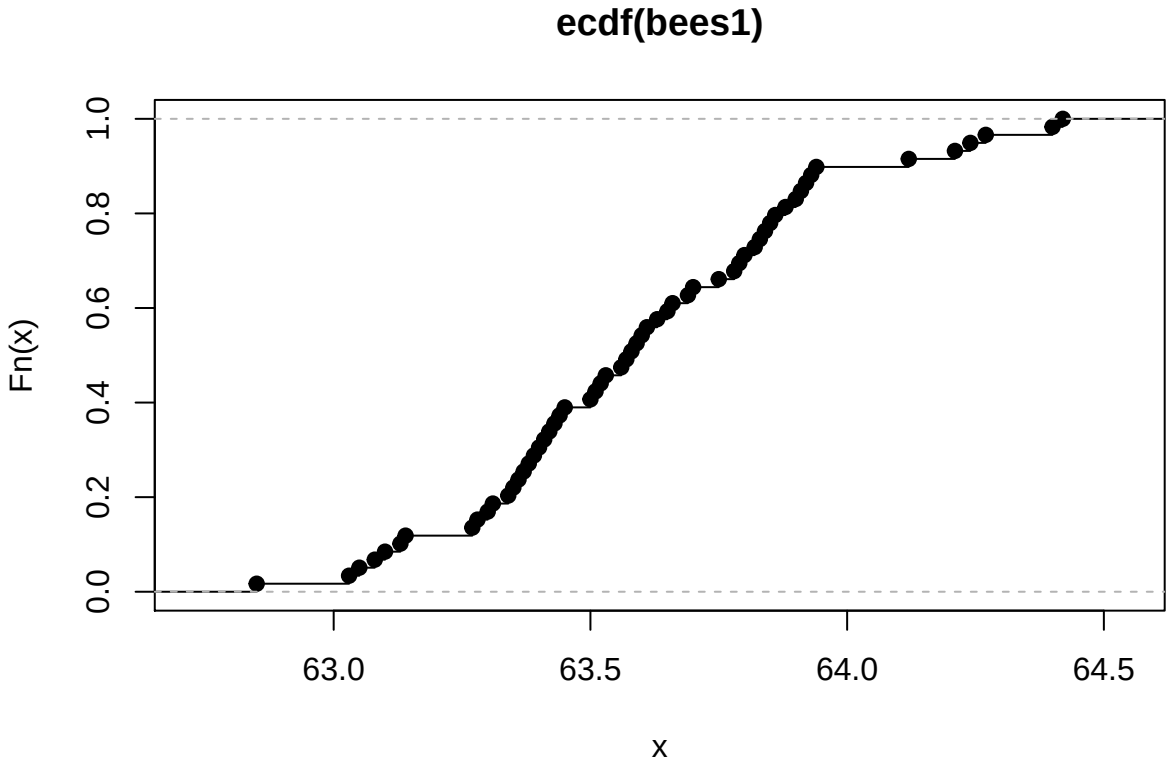
jiayuan

Sun Sep 20 23:43:43 2015

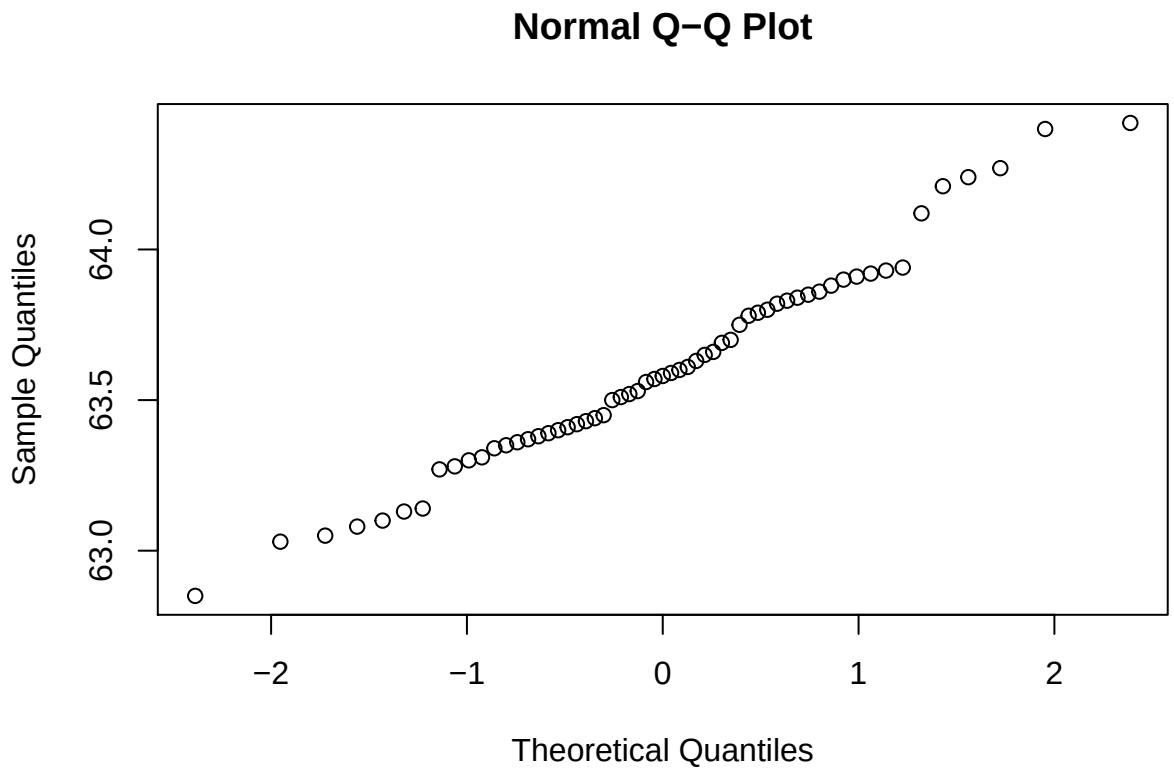
```
library(ggplot2)
bees <- read.csv("/users/jiayuan/Documents/MA681/bees.csv",header = FALSE) #read data
bees <- data.frame(bees)
bees1 <- sort(bees$V1, decreasing=FALSE) #sort the increasing data
ecdf(bees1) #compute an empirical cumulative distribution function
```

```
## Empirical CDF
## Call: ecdf(bees1)
##  x[1:59] =  62.85,  63.03,  63.05,  ...,   64.4,  64.42
```

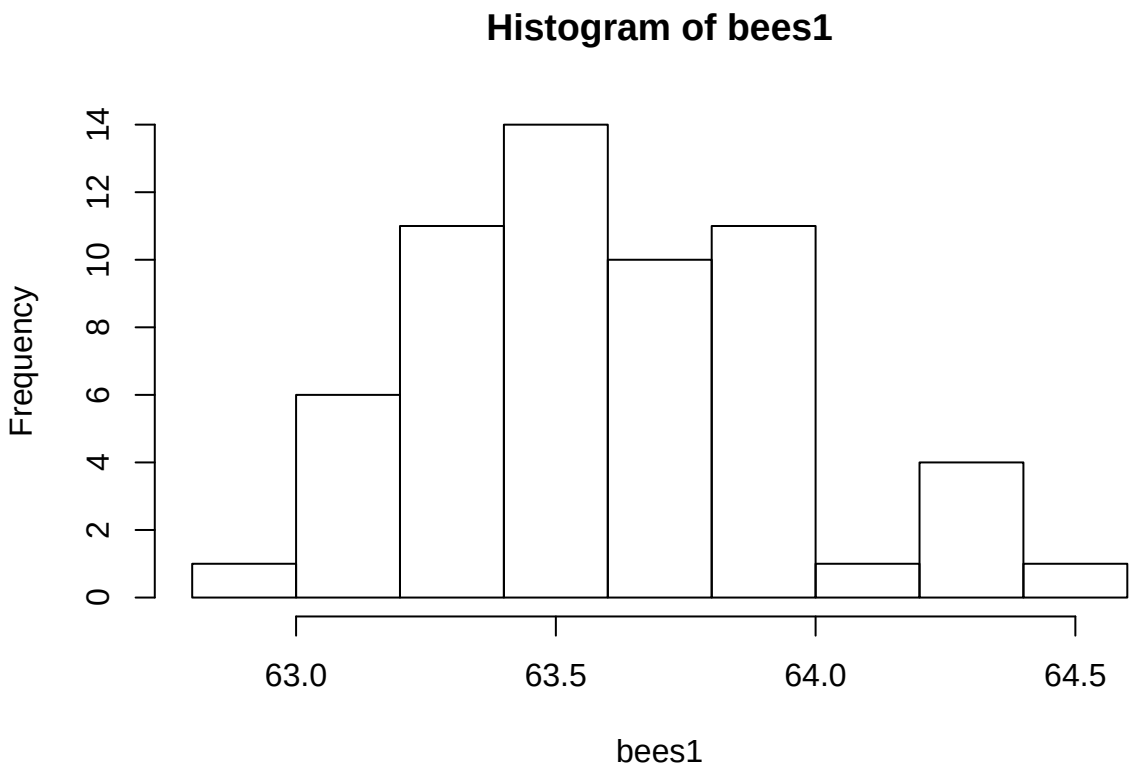
```
plot(ecdf(bees1)) #plot the an empirical cumulative distribution function
```



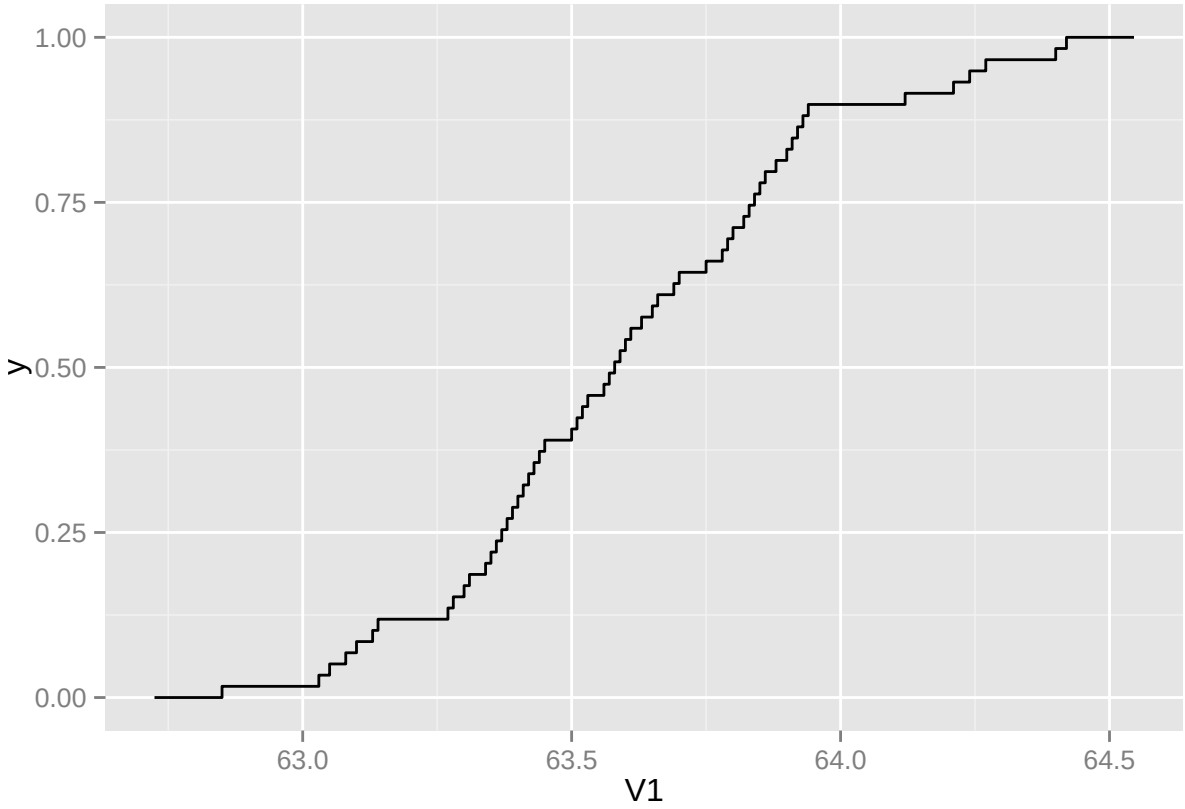
```r
qqnorm(bees1) #almost standard normal distribution
```




```r
hist(bees1) #also seems to be a standard normal distribution
```



```
qplot(data=bees, x=V1, stat = "ecdf", geom = "step")
```




```
ks.test(bees1,y="pnorm")
```

```
##
##  One-sample Kolmogorov-Smirnov test
##
## data:  bees1
## D = 1, p-value = 7.772e-16
## alternative hypothesis: two-sided
```

```
#Do the KS test:
#H0: F(x)=F*(x) for -inf<x<inf, H1: H0 is not true.
#The statistics D*n=sup(abs(Fn(x)-F*(x)))=1, and p-value=7.772e-16<0.01.
#Therefore, there is enough evidence to reject H0, and we can say that
#the cdf of the melting tempereture of beeswax as observed with a variety of beeswax samples
#is not actually a particular normal distribution cdf F*(x).
#This result is different from the above we get.
bees1
```

```
##  [1] 62.85 63.03 63.05 63.08 63.10 63.13 63.14 63.27 63.28 63.30 63.31
## [12] 63.34 63.35 63.36 63.37 63.38 63.39 63.40 63.41 63.42 63.43 63.44
## [23] 63.45 63.50 63.51 63.52 63.53 63.56 63.57 63.58 63.59 63.60 63.61
## [34] 63.63 63.65 63.66 63.69 63.70 63.75 63.78 63.79 63.80 63.82 63.83
## [45] 63.84 63.85 63.86 63.88 63.90 63.91 63.92 63.93 63.94 64.12 64.21
## [56] 64.24 64.27 64.40 64.42
```