Chapter 10 Problems 7

Kartheek Raj

12/26/2019

7. In the chapter, we mentioned the use of correlation-based distance and Euclidean distance as dissimilarity measures for hierarchical clustering. It turns out that these two measures are almost equivalent: if each observation has been centered to have mean zero and standard deviation one, and if we let rij denote the correlation between the ith and jth observations, then the quantity 1−rij is proportional to the squared Euclidean distance between the ith and jth observations. On the USArrests data, show that this proportionality holds.

Required packages : ISLR

Answer

Data pulling from ISLR Library

```
require(ISLR)
```

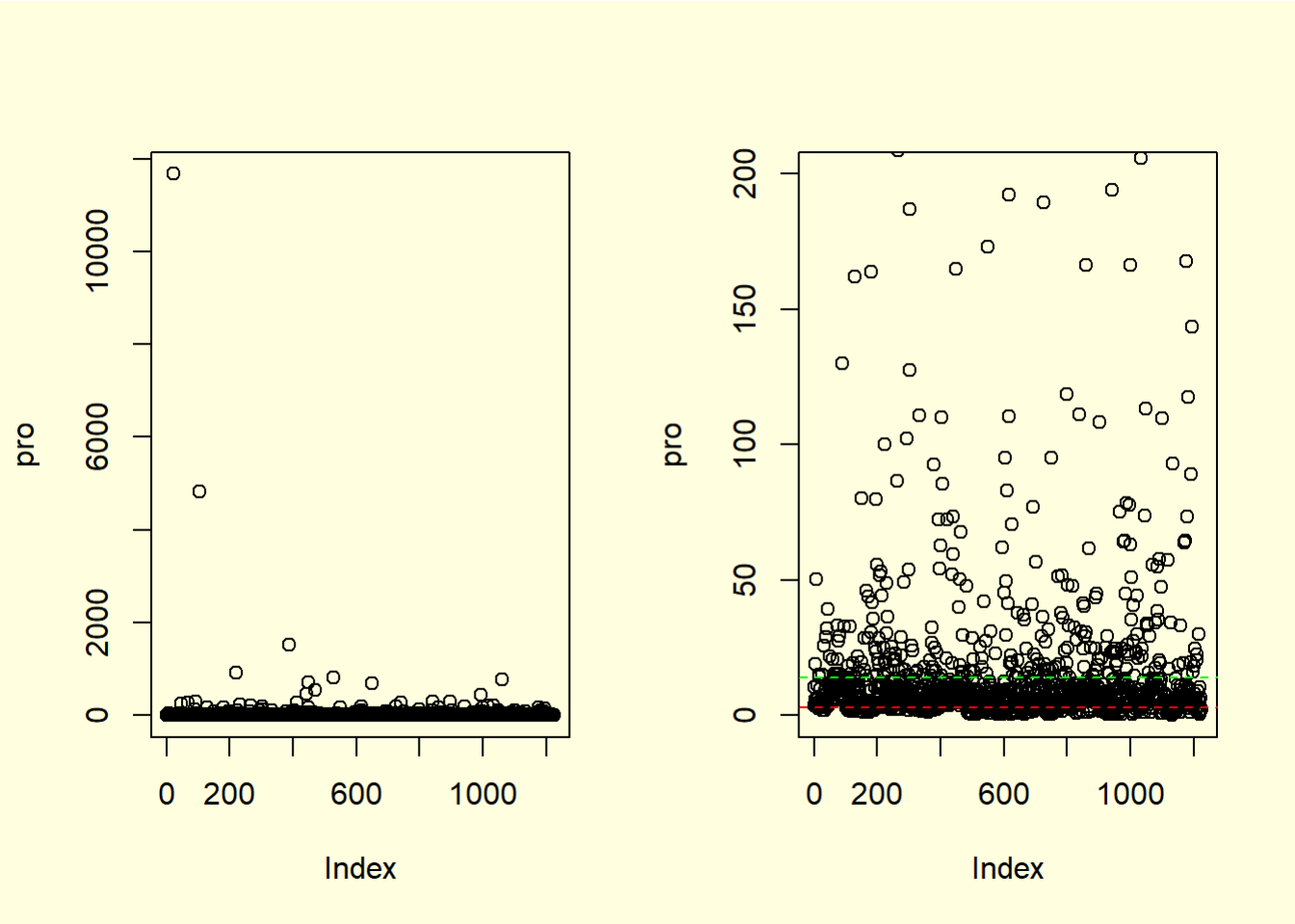
```
## Loading required package: ISLR
```

```
library(ISLR)
attach(USArrests)
```

Proportionality test Euclidean distance vs Correlation distance as mentioned in the question.

```
phtest<-function(x){
  sx<-scale(x)
  distdat<-dist(sx)
  sqrdistdat<-(distdat)^2
  tsx<-t(sx)
  cor11<-cor(tsx)
  cordata<-as.dist(1-cor11)
  pro<-(sqrdistdat/cordata)
  print(summary(pro))
  par(mfrow=c(1,2),bg="lightyellow")
  plot(pro)
  plot(pro,ylim = c(0,200))
  summary(pro)
  abline(h=3,col="red",lty="dashed")
  abline(h=14,col="green",lty="dashed")
}
paste0(phtest(USArrests))
```

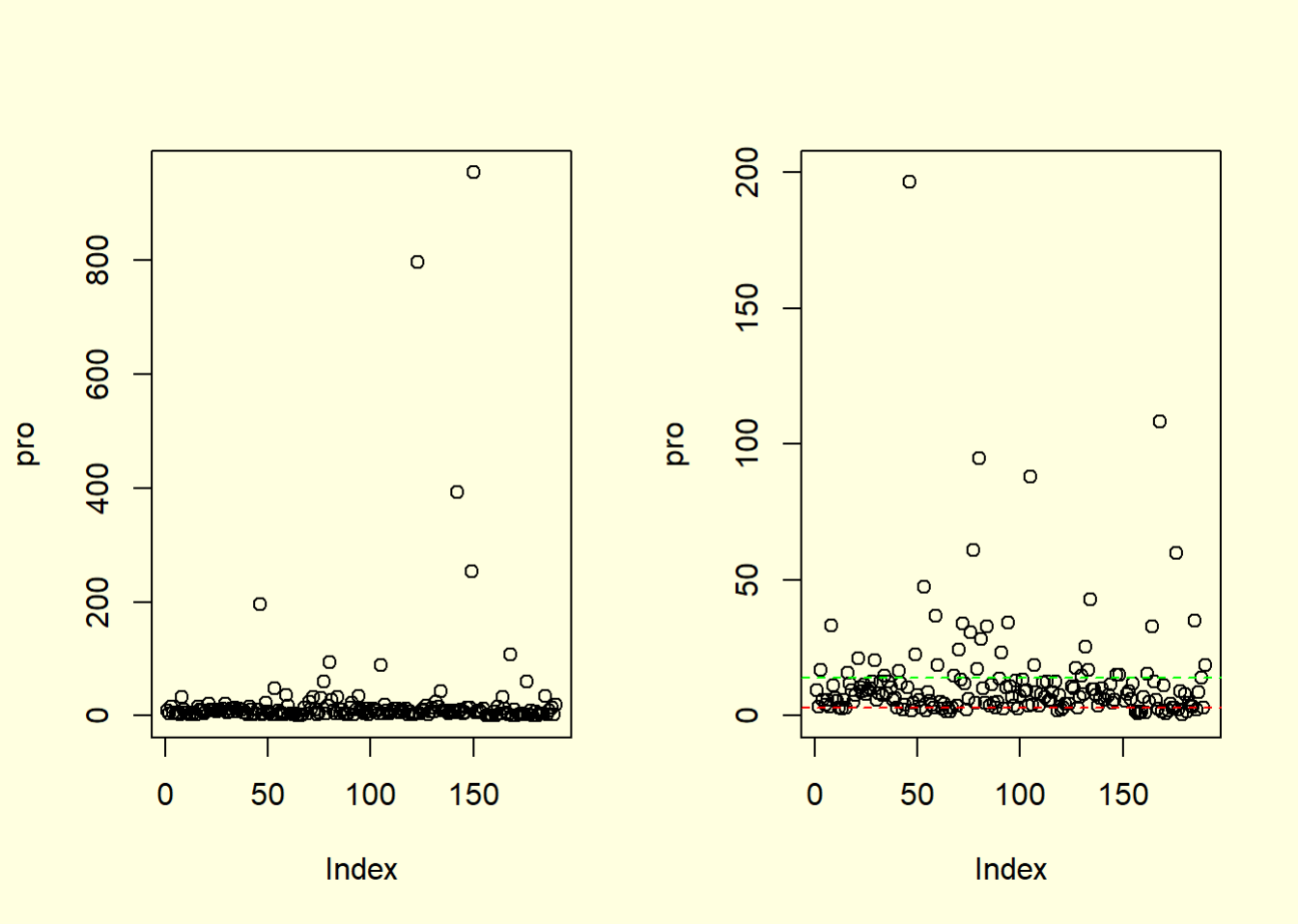
```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
##     0.205     3.808     7.466    35.722    14.464 11670.381
```




```
## character(0)
```

```
paste0(phtest(USArrests[1:20,]))
```

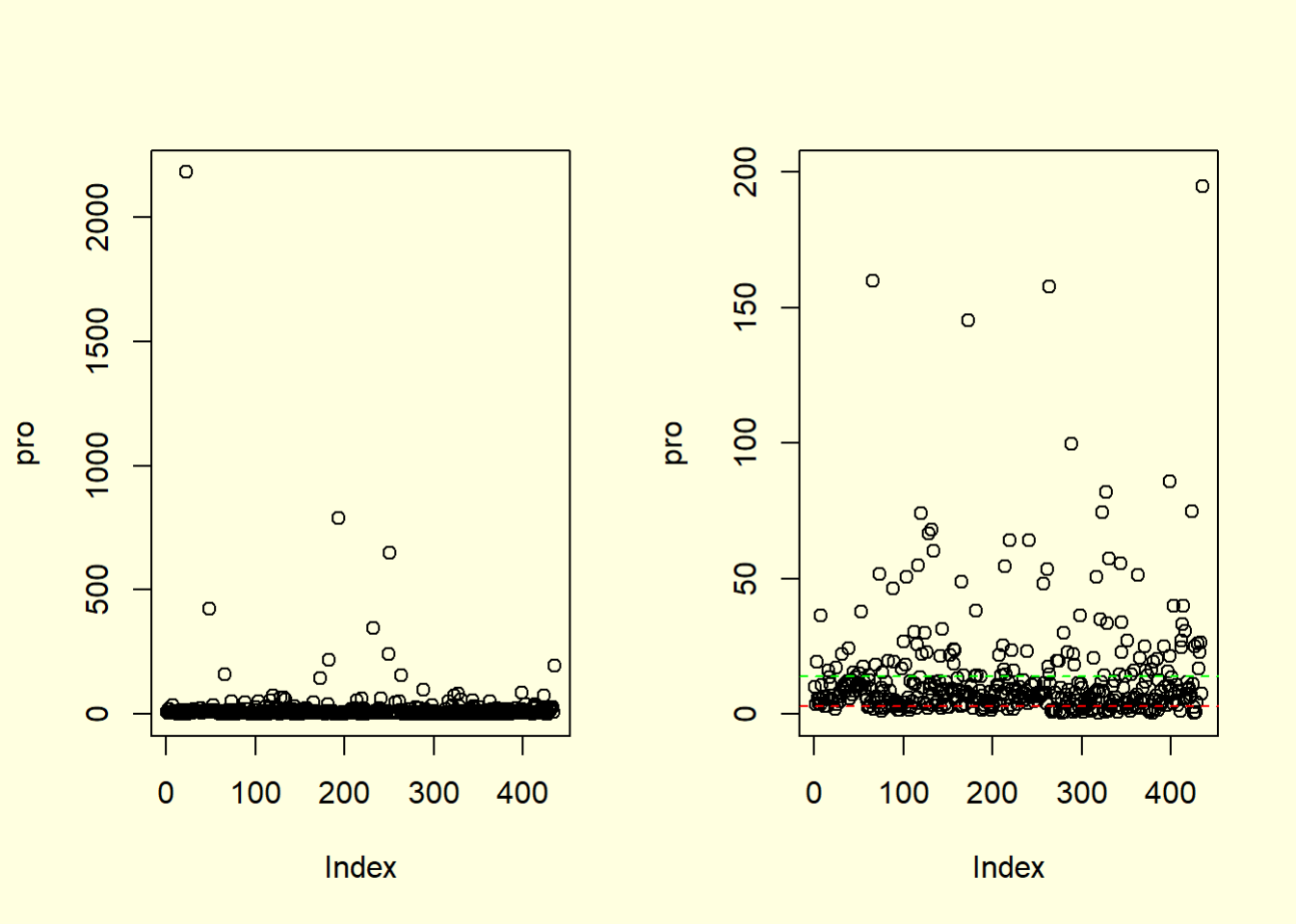
```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
##   0.5186   3.7826   7.6754  24.7658  12.5696 952.7748
```




```
## character(0)
```

```
paste0(phtest(USArrests[1:30,]))
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
##    0.3039    3.9568    7.8449   24.6119   14.3675 2180.9714
```




```
## character(0)
```

Insight 1 : As we plotted whole and samples its followed the pattern most of theobervation falls under green and blue dotted lines with some noise showing evidence proportionality holds.