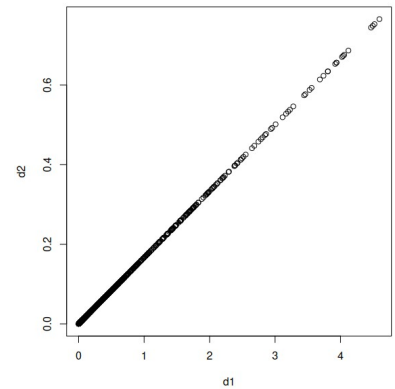


## Question 7

After using the matrix transpose function and creating both the data's distance and correlation to the other points, I plotted them. If they are the same,  $d1 = d2$ , so the line should fit  $y=x$ , which it does as shown by the plot



## Question 9

a) I set the seed and performed the Hierarchical clustering

b) After cutting the tree at a height = 3, I got 3 distinct clusters.

c) I did the same clustering after scaling the variables

```
# C)
hc2 <- hclust(dist(scale(USArrests)), method = "complete")
```

```
[[1]]
[1] "Alabama"      "Alaska"      "Arizona"
[4] "California"   "Delaware"    "Florida"
[7] "Illinois"     "Louisiana"   "Maryland"
[10] "Michigan"     "Mississippi" "Nevada"
[13] "New Mexico"   "New York"    "North Carolina"
[16] "South Carolina"

[[2]]
[1] "Arkansas"     "Colorado"    "Georgia"
[4] "Massachusetts" "Missouri"    "New Jersey"
[7] "Oklahoma"     "Oregon"      "Rhode Island"
[10] "Tennessee"    "Texas"       "Virginia"
[13] "Washington"   "Wyoming"

[[3]]
[1] "Connecticut"  "Hawaii"      "Idaho"
[4] "Indiana"      "Iowa"        "Kansas"
[7] "Kentucky"     "Maine"       "Minnesota"
[10] "Montana"      "Nebraska"    "New Hampshire"
[13] "North Dakota" "Ohio"        "Pennsylvania"
[16] "South Dakota" "Utah"        "Vermont"
[19] "West Virginia" "Wisconsin"
```

d) The result is the same. We should scale the variables, because if we are clustering based on multiple predictors, they could have different scales, so the clustering dissimilarity method results will vary depending on if the variables are scaled or not.

## Question 13

Each of the clustering methods all produced 2 separate distinct groupings. The complete cluster method seemed to produce a result that is more balanced and tree-like, the single is very lopsided, and the average method just seems all over the place, but it identified 2 (ish, kinda) groups of points (more like 4, but we could call it a stretch)

