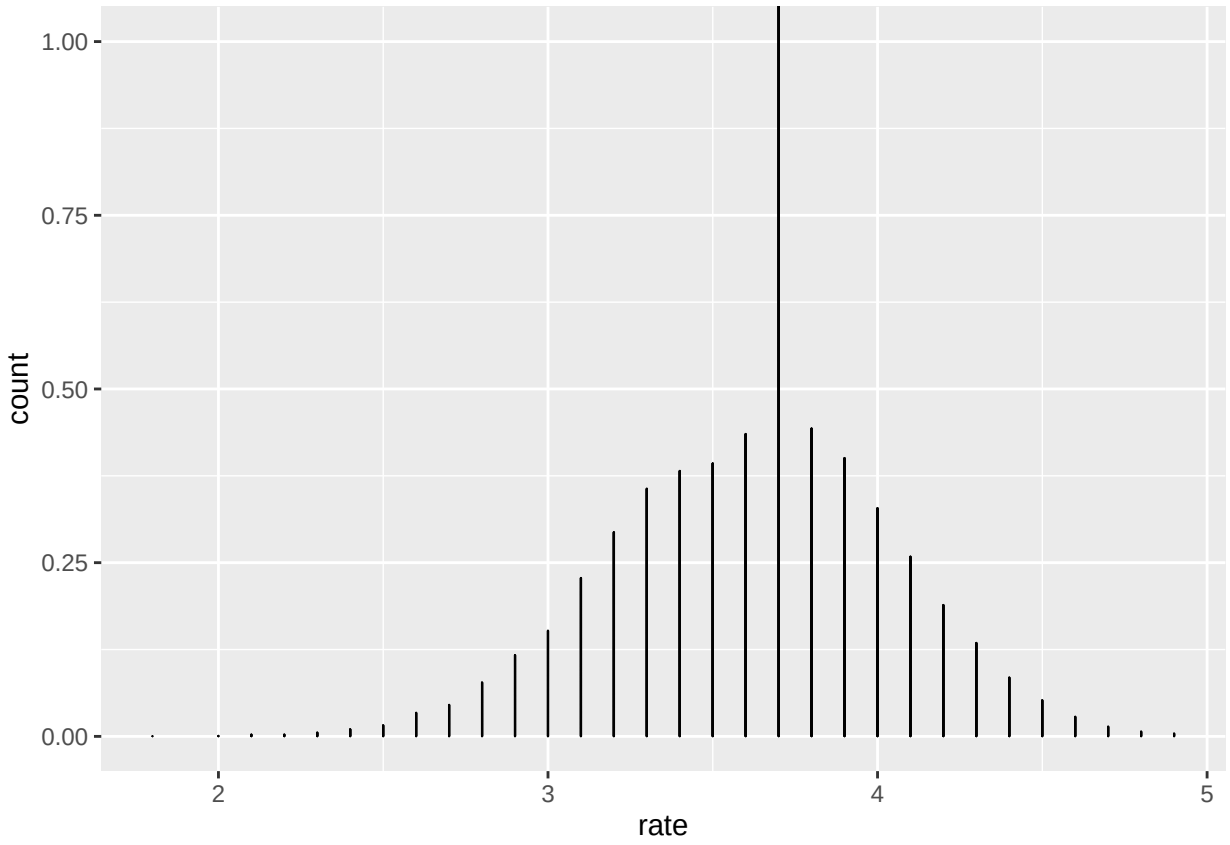
R Notebook

```
zomato <- read.csv("~/Downloads/export_dataframe.csv")
zomato_nocuisines <- zomato[-c(7)]
zomato_nocuisines$scaledvote <- scale(zomato_nocuisines$votes)
zomato_nocuisines$scaledcost <- scale(zomato_nocuisines$cost_for_two)
zomato_nocuisines <- zomato_nocuisines[-c(1)]
zomato_nocuisines <- zomato_nocuisines[-c(2:3)]
ggplot(zomato_nocuisines, aes(rate)) + geom_dotplot(binwidth = 0.001)
```

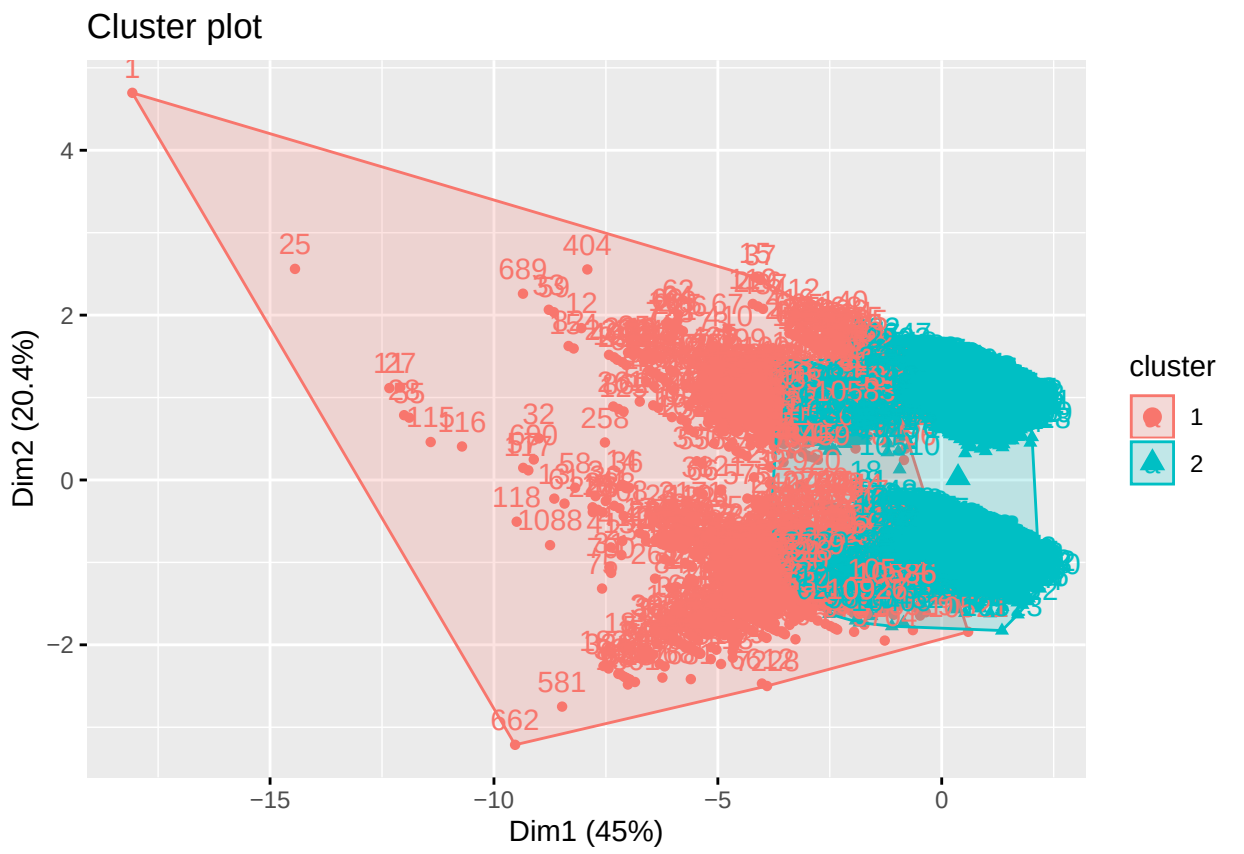



```
k2 <- kmeans(zomato_nocuisines, centers = 2, nstart = 25)
str(k2)
```

```
## List of 9
##  $ cluster     : int [1:12941] 1 1 1 1 2 1 2 1 2 1 ...
##  $ centers     : num [1:2, 1:5] 4.163 3.585 0.423 0.54 0.717 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:2] "1" "2"
##   .. ..$ : chr [1:5] "rate" "online_order" "book_table" "scaledvote" ...
##  $ totss       : num 32080
##  $ withinss    : num [1:2] 10673 9081
##  $ tot.withinss: num 19754
##  $ betweenss   : num 12326
##  $ size        : int [1:2] 1067 11874
```

```
##  $ iter        : int 1
##  $ ifault      : int 0
##  - attr(*, "class")= chr "kmeans"
```

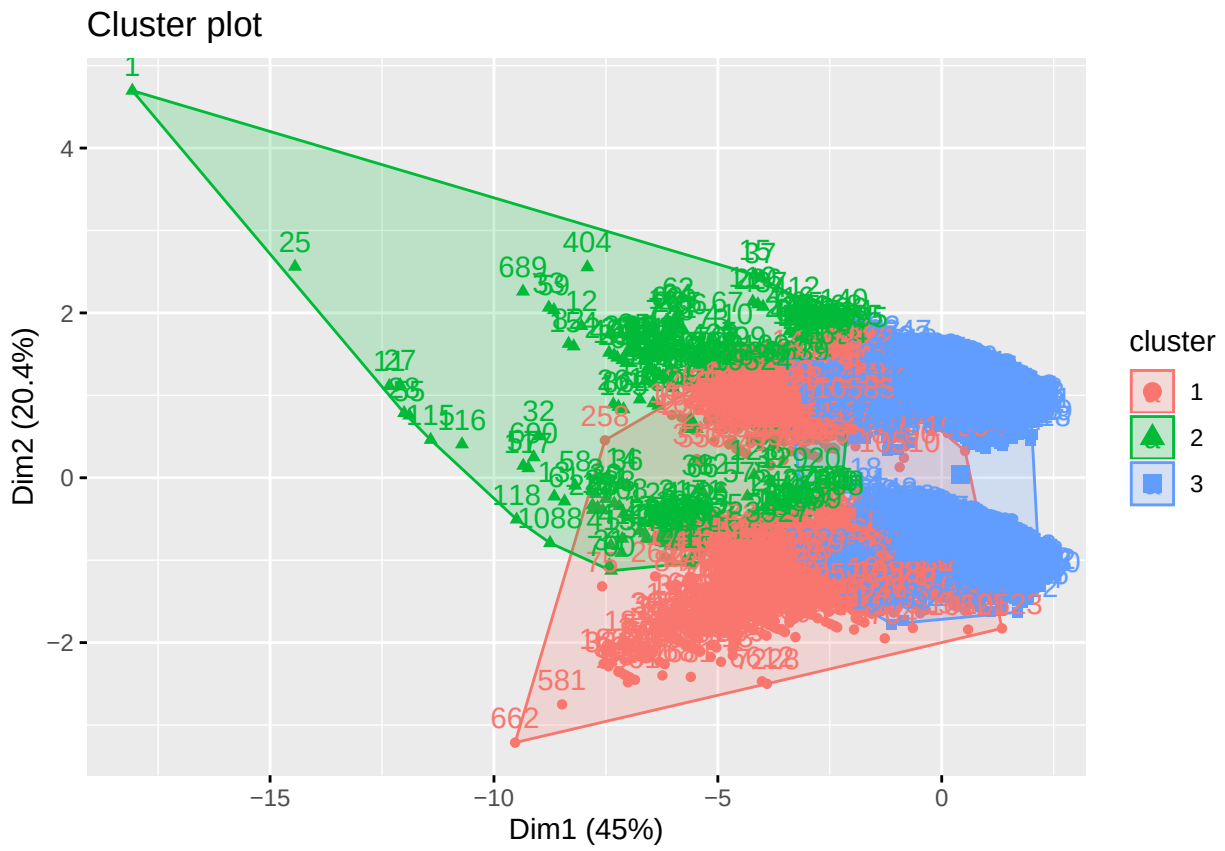
```
fviz_cluster(k2, data = zomato_nocuisines)
```




```
k3 <- kmeans(zomato_nocuisines, centers = 3, nstart = 25)
str(k3)
```

```
## List of 9
##  $ cluster     : int [1:12941] 2 2 2 2 3 1 3 1 1 2 ...
##  $ centers     : num [1:3, 1:5] 4.058 4.359 3.58 0.398 0.516 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:3] "1" "2" "3"
##   .. ..$ : chr [1:5] "rate" "online_order" "book_table" "scaledvote" ...
##  $ totss       : num 32080
##  $ withinss    : num [1:3] 4275 2691 8187
##  $ tot.withinss: num 15153
##  $ betweenss   : num 16927
##  $ size        : int [1:3] 1082 219 11640
##  $ iter        : int 4
##  $ ifault      : int 0
##  - attr(*, "class")= chr "kmeans"
```

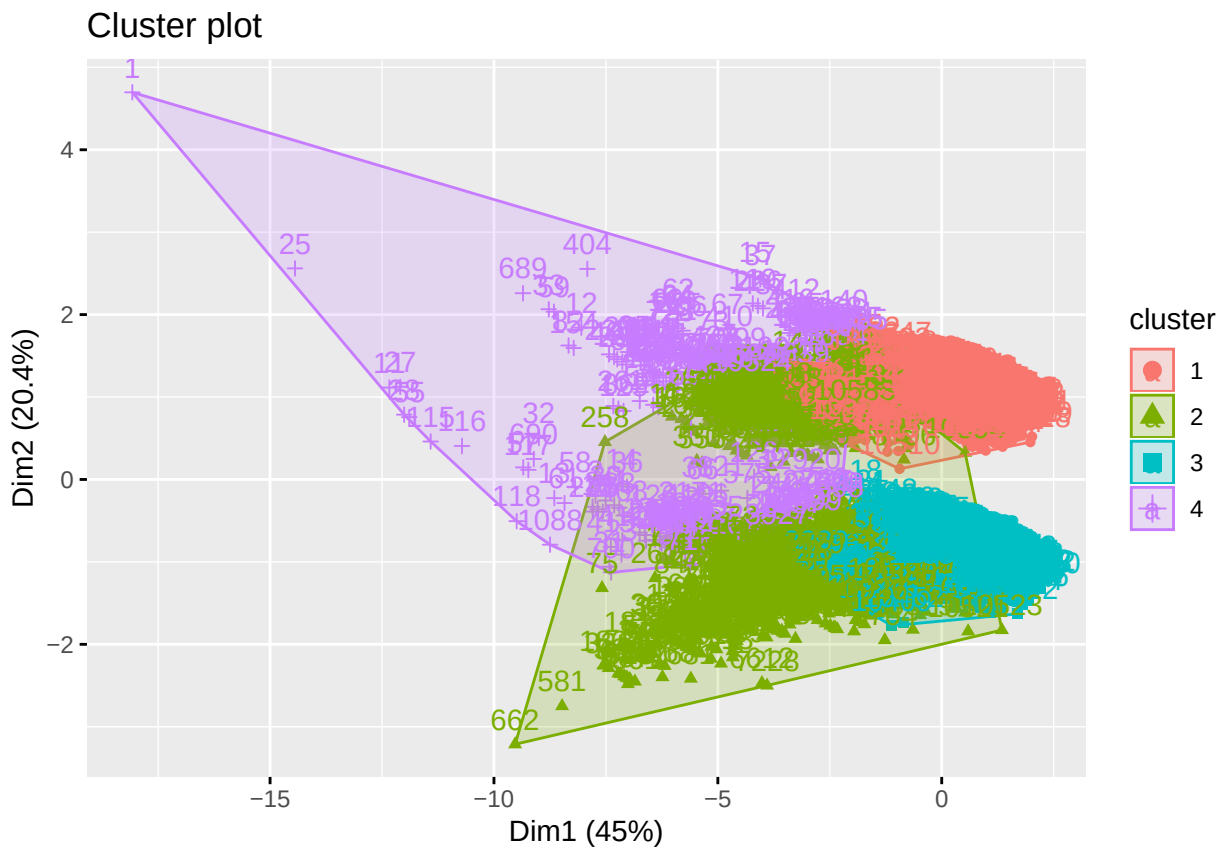
```
fviz_cluster(k3, data = zomato_nocuisines)
```




```
k4 <- kmeans(zomato_nocuisines, centers = 4, nstart = 25)
str(k4)
```

```
## List of 9
##  $ cluster     : int [1:12941] 4 4 4 4 3 2 1 2 2 4 ...
##  $ centers     : num [1:4, 1:5] 3.6 4.06 3.56 4.36 1 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:4] "1" "2" "3" "4"
##   .. ..$ : chr [1:5] "rate" "online_order" "book_table" "scaledvote" ...
##  $ totss       : num 32080
##  $ withinss    : num [1:4] 3161 4104 2216 2691
##  $ tot.withinss: num 12172
##  $ betweenss   : num 19909
##  $ size        : int [1:4] 6357 1034 5331 219
##  $ iter        : int 4
##  $ ifault      : int 0
##  - attr(*, "class")= chr "kmeans"
```

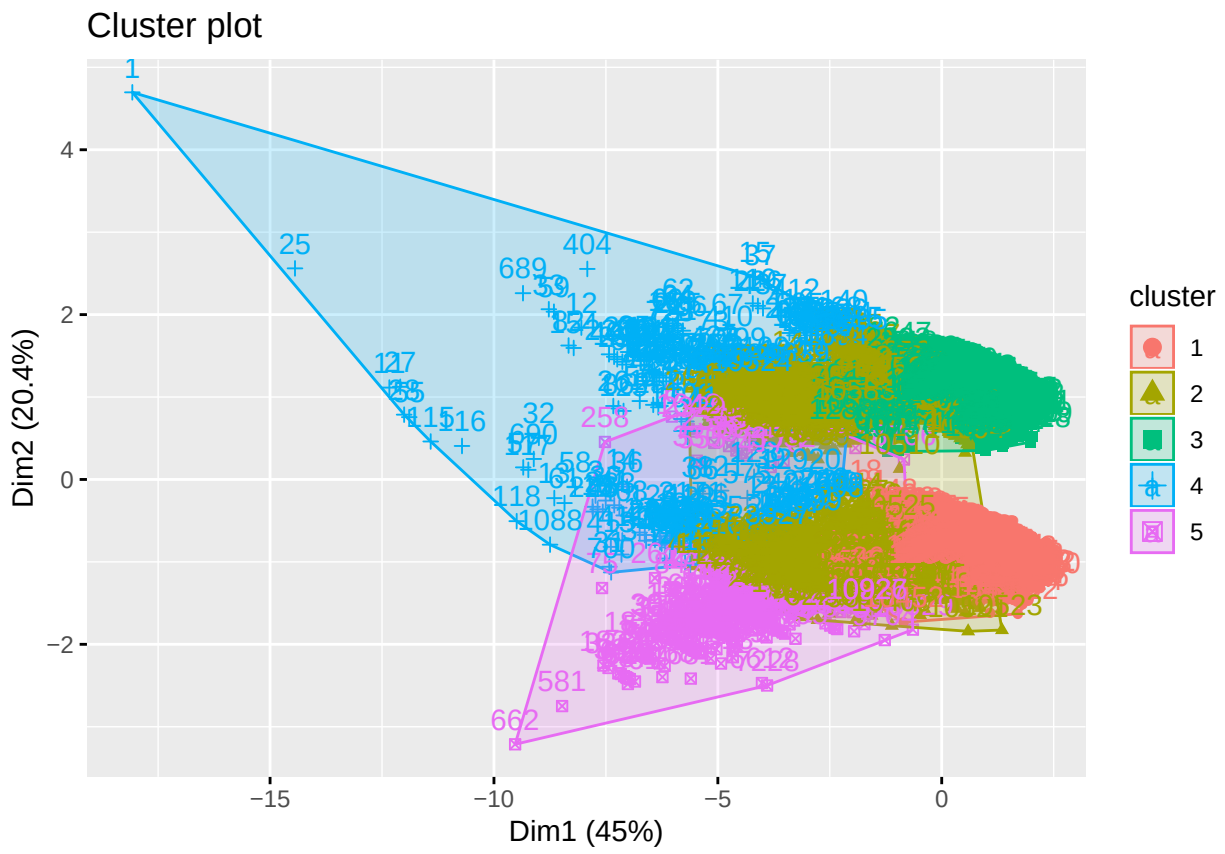
```
fviz_cluster(k4, data = zomato_nocuisines)
```




```
k5 <- kmeans(zomato_nocuisines, centers = 5, nstart = 25)
str(k5)
```

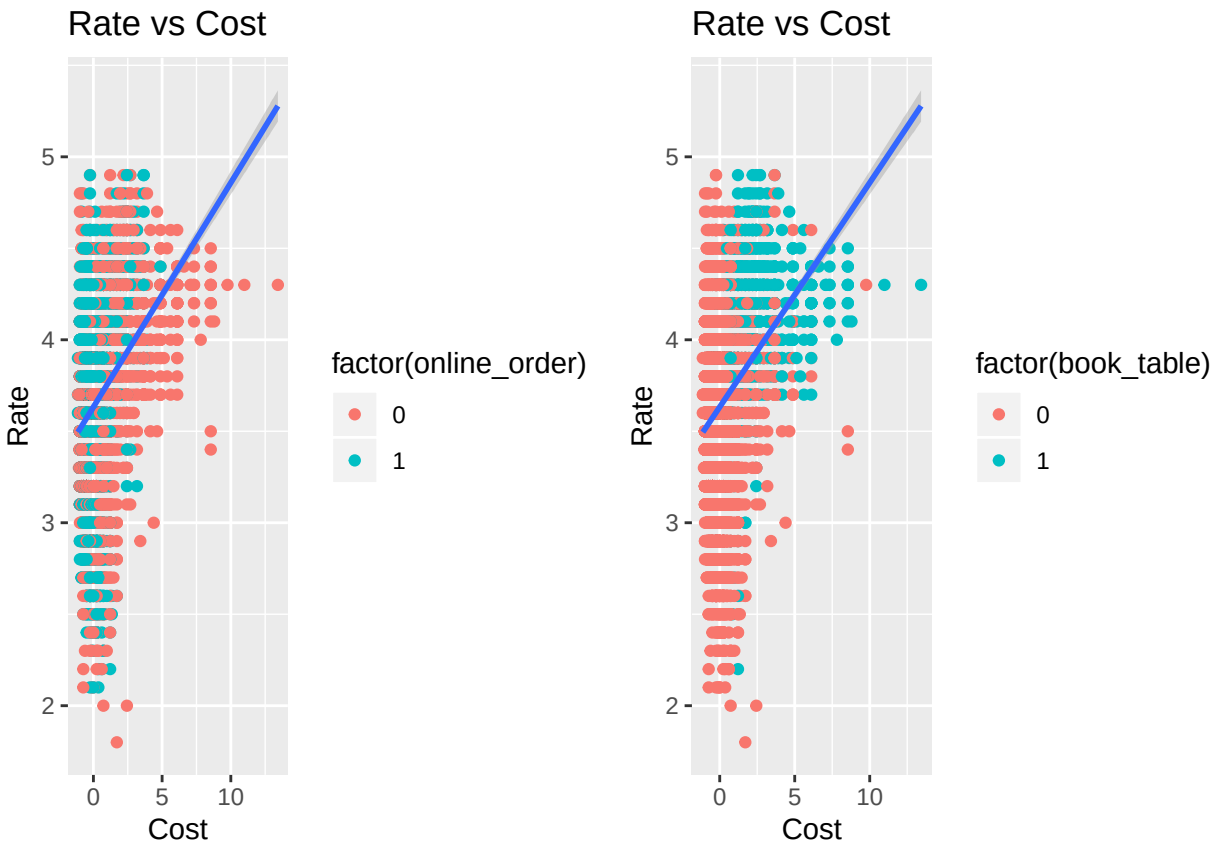
```
## List of 9
##  $ cluster     : int [1:12941] 4 4 4 4 1 5 3 5 2 4 ...
##  $ centers     : num [1:5, 1:5] 3.55 3.96 3.58 4.36 4.08 ...
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:5] "1" "2" "3" "4" ...
##   .. ..$ : chr [1:5] "rate" "online_order" "book_table" "scaledvote" ...
##  $ totss       : num 32080
##  $ withinss    : num [1:5] 1664 2651 2317 2684 836
##  $ tot.withinss: num 10152
##  $ betweenss   : num 21928
##  $ size        : int [1:5] 5129 1354 6009 219 230
##  $ iter        : int 5
##  $ ifault      : int 0
##  - attr(*, "class")= chr "kmeans"
```

```
fviz_cluster(k5, data = zomato_nocuisines)
```

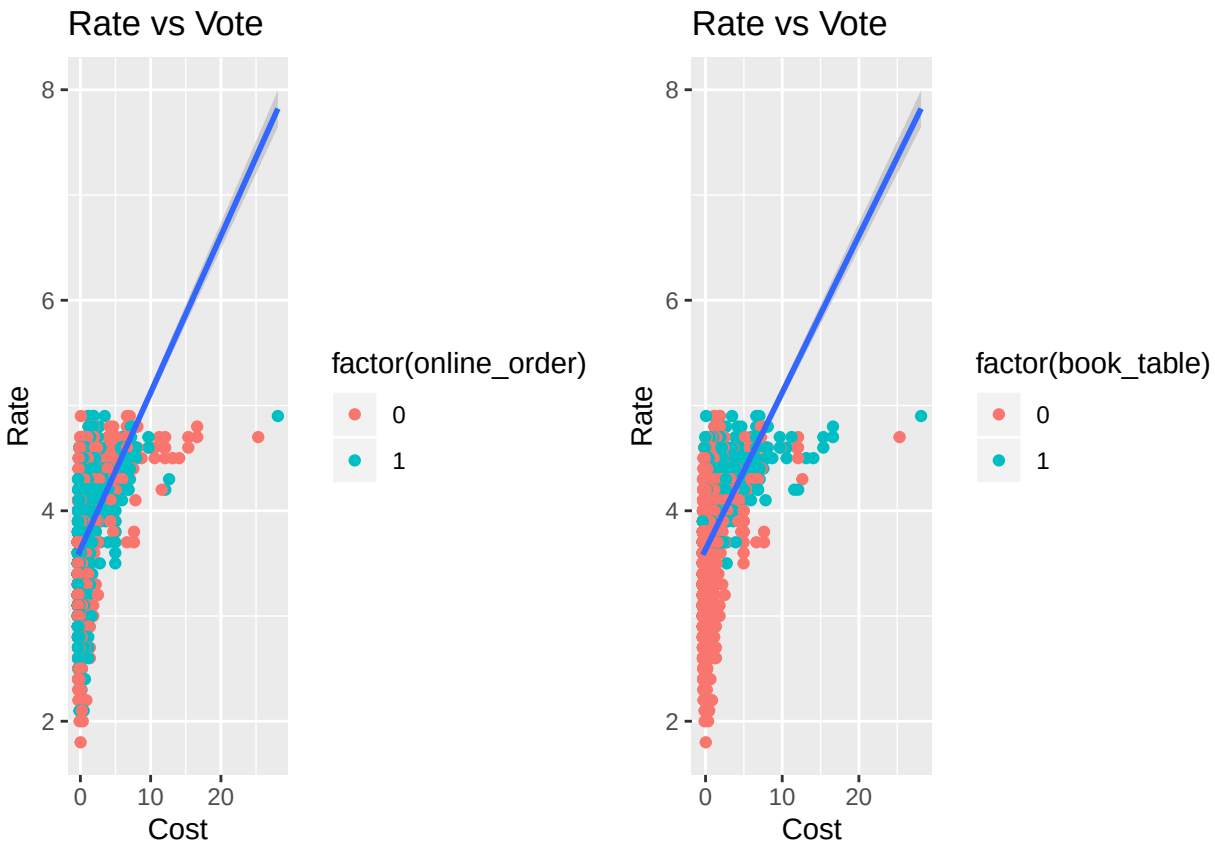



```
# plots to compare
p1 <- fviz_cluster(k2, geom = "point", data = zomato_nocuisines) +
    ggtitle("k = 2")
p2 <- fviz_cluster(k3, geom = "point", data = zomato_nocuisines) +
    ggtitle("k = 3")
p3 <- fviz_cluster(k4, geom = "point", data = zomato_nocuisines) +
    ggtitle("k = 4")
p4 <- fviz_cluster(k5, geom = "point", data = zomato_nocuisines) +
    ggtitle("k = 5")
library(gridExtra)
```

```
p1 <- ggplot(zomato_nocuisines, aes(y = rate, x = scaledcost)) +
    geom_point(aes(color = factor(online_order))) + labs(x = "Cost",
    y = "Rate", title = "Rate vs Cost") + geom_smooth(method = "lm")
p2 <- ggplot(zomato_nocuisines, aes(y = rate, x = scaledcost)) +
    geom_point(aes(color = factor(book_table))) + labs(x = "Cost",
    y = "Rate", title = "Rate vs Cost") + geom_smooth(method = "lm")
p3 <- ggplot(zomato_nocuisines, aes(y = rate, x = scaledvote)) +
    geom_point(aes(color = factor(online_order))) + labs(x = "Cost",
    y = "Rate", title = "Rate vs Vote") + geom_smooth(method = "lm")
p4 <- ggplot(zomato_nocuisines, aes(y = rate, x = scaledvote)) +
    geom_point(aes(color = factor(book_table))) + labs(x = "Cost",
    y = "Rate", title = "Rate vs Vote") + geom_smooth(method = "lm")
grid.arrange(p1, p2, nrow = 1)
```




```
grid.arrange(p3, p4, nrow = 1)
```


Rate vs Vote
Rate
Cost
factor(online_order)
0
1
Rate vs Vote
Rate
Cost
factor(book_table)
0
1

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	3.63	0.00	1122.91	0.00
scaledcost	0.12	0.00	37.88	0.00

Table 1: Table of model involving rate, and cost.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	3.63	0.00	1176.89	0.00
scaledvote	0.12	0.00	35.71	0.00
scaledcost	0.08	0.00	22.85	0.00

Table 2: Table of model involving rate, cost, and vote.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	3.60	0.00	784.68	0.00
online_order	0.01	0.01	2.04	0.04
book_table	0.29	0.01	20.73	0.00
scaledvote	0.10	0.00	29.94	0.00
scaledcost	0.03	0.00	8.58	0.00

Table 3: Table of model involving rate, online order, votes, cost, and book table.