Joaquin Díaz Nicolás López

Answers

1. You keep the city attribute because it's the main identifier of each record, but you don't use it because it won't have any correlation with any other attribute because it's unique.

2. If we look at the chart we can find that the related variable are:

- Pop.1980 with Pop.1990
- Pop.1990 with Pop.2000
- Growth with Food
- Food with Phones
- Vehicles with Food

3. We calculated the correlation value of all the attributes and decided that a number above 0,65 shows a correlation and a number less than -0,65 shows a negative correlation. In this case we got this for the negative correlation:

```
subset(na.omit(data.frame(expand.grid(dimnames(CM)), value = c(CM))), value < -0.65)
```

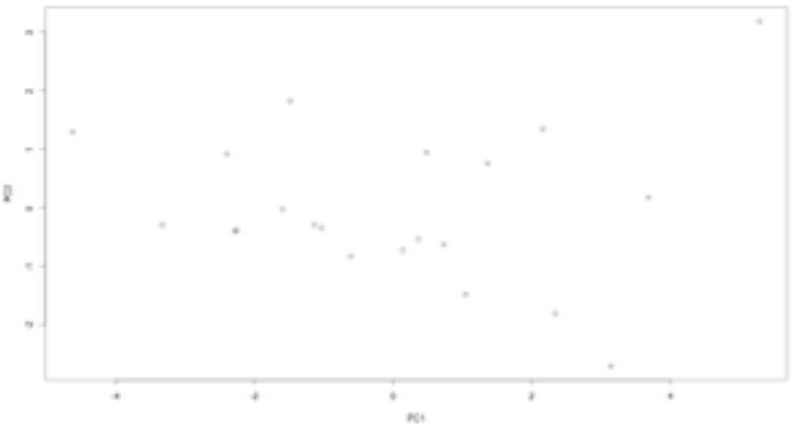
- Growth Pop.1980 -0.6935458
- Phones Food -0.8429700
- Vehicles Food -0.7610807
- Water PersRoom -0.6643779
- Elec PersRoom -0.7934353

subset(na.omit(data.frame(expand.grid(dimnames(CM)), value = c(CM))), value > 0.65)

- Pop.1990 Pop.1980 0.9559486
- Pop.2000 Pop.1980 0.7919490
- Phones Pop.1980 0.6734610
- Pop.2000 Pop.1990 0.9233506
- Growth PersRoom 0.6711674
- Water Elec 0.8304116

4.

```
withOutCity <- data[, -1] # take city out because its non numeric
withOutCityScaled <- scale(withOutCity) # scale variables
out.pca<- prcomp(withOutCityScaled) # apply PCA
summary(out.pca) # print the summary
plot(out.pca$x[,1:2]) # create a chart
```



5. Now we have a bi dimensional data model which its a lot easier to work with.