DaigleInClassLabWk9D3.R

daiglechris

Fri Mar 23 20:05:33 2018

```
# Christopher Daigle
# Week9D3 In Class Lab - Principal Component Analysis

# Exercise ##########################################
# Perform a visualisation similar to that in class from a data set of your choice

# I have chosen to create a dataset from Kaggle related to breweries and two datasets from
# the CDC on impaired/drunk driving

# setwd("/Users/2011home/Library/Mobile Documents/com~apple~CloudDocs/Education/UConn/Spring 2018/R/Dat
setwd("/Users/daiglechris/Library/Mobile Documents/com~apple~CloudDocs/Education/UConn/Spring 2018/R/Da

# Loading
breweries <- read.csv("breweries.csv", stringsAsFactors = FALSE)
  #breweries2 <- read.csv("breweries2.csv")
duiDeaths <- read.csv("occupantAndAlcohol-ImpairedDrivingDeaths2005_2014.csv", stringsAsFactors = FALSE
reported <- read.csv("percentAdultsReport2014.csv", stringsAsFactors = FALSE)

head(breweries)
```

```
##                  address
## 1           407 Radam Ln
## 2 1135 N W Galveston Ave
## 3       830 W Bannock St
## 4         1022 Texan Trl
## 5             Po Box 135
## 6      31111 Via Colinas
##
## 1
## 2
## 3                           brewery, Restaurant, Pub, Brewery, pub, Brewery and Pub Downtown Boise
## 4                        Wine Tours & Tastings, Food & Drink, Tours, Brewery Tours, transportation,
## 5 American Restaurants, Bar & Grill Restaurants, Banquet Facilities, Breweries & Brew Pubs, Cocktail
## 6
##               city country                                        key
## 1           Austin      US                   us/tx/austin/407radamln
## 2             Bend      US              us/or/bend/1135nwgalvestonave
## 3            Boise      US                  us/id/boise/830wbannockst
## 4        Grapevine      US                us/tx/grapevine/1022texantrl
## 5          Villard      US                      us/mn/villard/pobox135
## 6 Westlake Village      US us/ca/westlakevillage/31111viacolinas/202
##        lat       long                                        name
## 1       NA         NA                                (512) Brewing Co
## 2       NA         NA 10 Barrel Brewing Company, 10 Barrel Brewing Co
## 3 43.61771 -116.20288        10 Barrel Brewing, 10 Barrel Brewing Co.
## 4 32.93838  -97.06434                                 10 Gallon Tours
## 5       NA         NA                                  10 Mile Tavern
```

```
## 6       NA         NA                                 101 Cider House
##                                    phones postalCode province
## 1                                5127072337      78745       TX
## 2               (541) 585-1007, 5415851007      97703       OR
## 3 (208) 344-5870, 2083445870, 2.08344587E9      83702       ID
## 4                18174031832, 817 403-1832      76051       TX
## 5                                3205542939      56385       MN
## 6                                8188519057      91362       CA
##                        websites
## 1
## 2                  10barrel.com
## 3                  10barrel.com
## 4 http://www.10gallontours.com/
## 5   http://www.10miletavern.com
## 6
```

```r
colnames(breweries)[11] <- "State"
table(breweries[,11])
```

```
##
##          AB          AK          AL          AR       AU-WA          AZ
##           1          27          67          56          13         155
##          CA          CO          CT          DC          DE      Denver
##         899         306          53          14          24           2
##          FL          GA          HI          IA          ID          IL
##         271          77          24         106          46         234
##          IN          KS          KY          LA Los Angeles          MA
##         144          33          51          36           1         107
##          MD          ME          MI          MN          MO          MS
##         125         325         270         111         124          13
##          MT          NC          ND          NE          NH          NJ
##          73         184           5          49          50         152
##          NL          NM          NV          NY          OH          OK
##           1          49          59         461         376          25
##          ON          OR          PA  Pittsburgh          RI  Sacramento
##           1         298         401           1          20           1
##   San Diego          SC          SD   St. Louis       Tampa          TN
##           1          60          14           1           1          95
##          TX          UT          VA          VT          WA          WI
##         314          33         175          57         329         326
##          WV          WY
##          32          16
```

```r
# Cleaning ####
breweries <- breweries[!(breweries$State == 'AB'),] # These observations of Alberta are not in the US
breweries$State[breweries$State == 'AU-WA'] <- 'WA'
breweries$State[breweries$State == 'Denver'] <- 'CO'
breweries$State[breweries$State == 'Los Angeles'] <- 'CA'
breweries <- breweries[!(breweries$State == 'NL'),] # This observation of Ontario is not in the US
breweries <- breweries[!(breweries$State == 'ON'),] # This observation of Ontario is not in the US
breweries$State[breweries$State == 'Pittsburgh'] <- 'PA'
breweries$State[breweries$State == 'Sacramento'] <- 'CA'
breweries$State[breweries$State == 'San Diego'] <- 'CA'
breweries$State[breweries$State == 'St. Louis'] <- 'MO'
breweries$State[breweries$State == 'Tampa'] <- 'FL'
```

```r
duiDeaths <- duiDeaths[,1:3]

reported <- reported[,1:3]
reported[,1] <- state.abb[match(reported[,1],state.name)]

# Structuring ####
brewByState <- as.data.frame(table(breweries[,11]))
colnames(brewByState) <- c('State', 'numOfBreweries')

colnames(duiDeaths) <- c('State', 'AlcoholImparedDrivingDeaths', 'OccupantDeaths')

colnames(reported) <- c('State', 'PrevalenceIn2012', 'PrevalenceIn2014')

struc1 <- merge(brewByState,reported, all = TRUE)
comData <- merge(struc1, duiDeaths, all = TRUE)
comData[,1] <- as.character(comData[,1])
comData <- comData[-c(8, 52, 53), ]


# Analyzing ####

head(comData)
```

```
##   State numOfBreweries PrevalenceIn2012 PrevalenceIn2014
## 1    AK             27              1.2              1.6
## 2    AL             67              1.7              1.3
## 3    AR             56              1.0              1.4
## 4    AZ            155              1.7              1.6
## 5    CA            902              1.8              1.9
## 6    CO            308              1.9              1.9
##   AlcoholImparedDrivingDeaths OccupantDeaths
## 1                         205            439
## 2                        2997           7754
## 3                        1636           4526
## 4                        2656           5712
## 5                        9791          20733
## 6                        1599           3521
```

```r
princomp(c(comData$numOfBreweries, comData$AlcoholImparedDrivingDeaths))
```

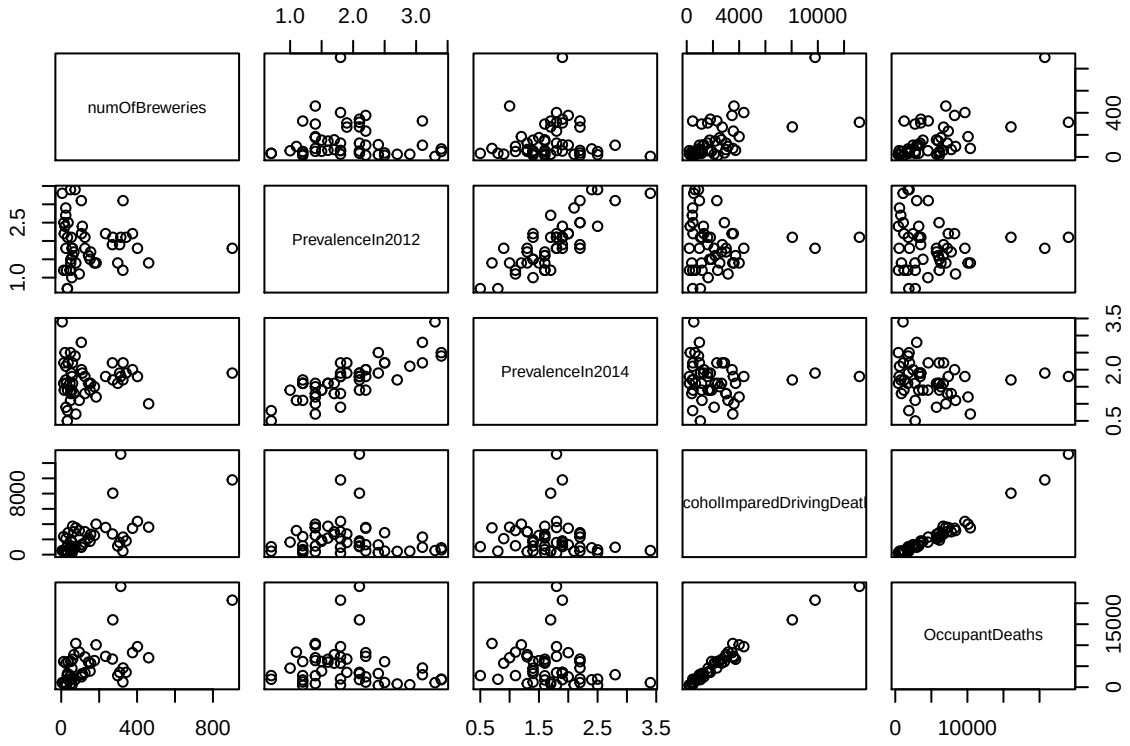
```
## Call:
## princomp(x = c(comData$numOfBreweries, comData$AlcoholImparedDrivingDeaths))
##
## Standard deviations:
##  Comp.1
## 2000.23
##
##  1  variables and  100 observations.
```

```r
head(comData)
```

```
##   State numOfBreweries PrevalenceIn2012 PrevalenceIn2014
## 1    AK             27              1.2              1.6
## 2    AL             67              1.7              1.3
## 3    AR             56              1.0              1.4
```

```
## 4    AZ            155              1.7              1.6
## 5    CA            902              1.8              1.9
## 6    CO            308              1.9              1.9
##   AlcoholImparedDrivingDeaths OccupantDeaths
## 1                         205            439
## 2                        2997           7754
## 3                        1636           4526
## 4                        2656           5712
## 5                        9791          20733
## 6                        1599           3521
```

```
dataInterest <- comData[,2:6]
pairs(dataInterest)
```



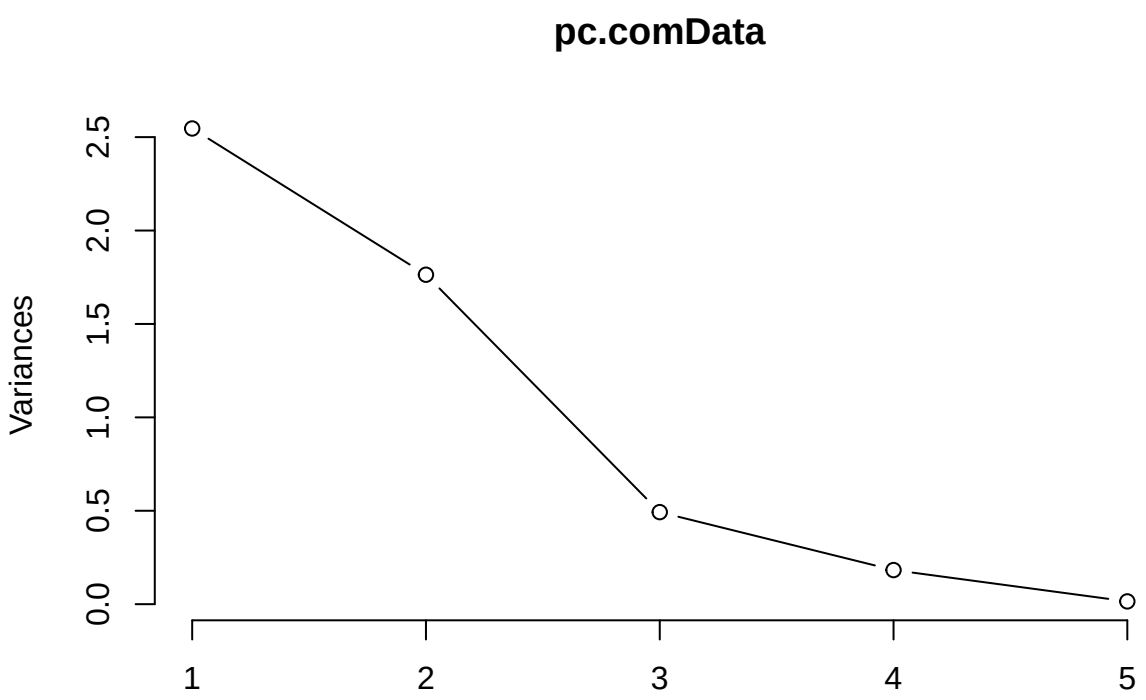

```
rownames(dataInterest) <- comData[,1]
head(dataInterest)
```

```
##    numOfBreweries PrevalenceIn2012 PrevalenceIn2014
## AK             27              1.2              1.6
## AL             67              1.7              1.3
## AR             56              1.0              1.4
## AZ            155              1.7              1.6
## CA            902              1.8              1.9
## CO            308              1.9              1.9
##    AlcoholImparedDrivingDeaths OccupantDeaths
## AK                         205            439
## AL                        2997           7754
## AR                        1636           4526
## AZ                        2656           5712
## CA                        9791          20733
## CO                        1599           3521
```

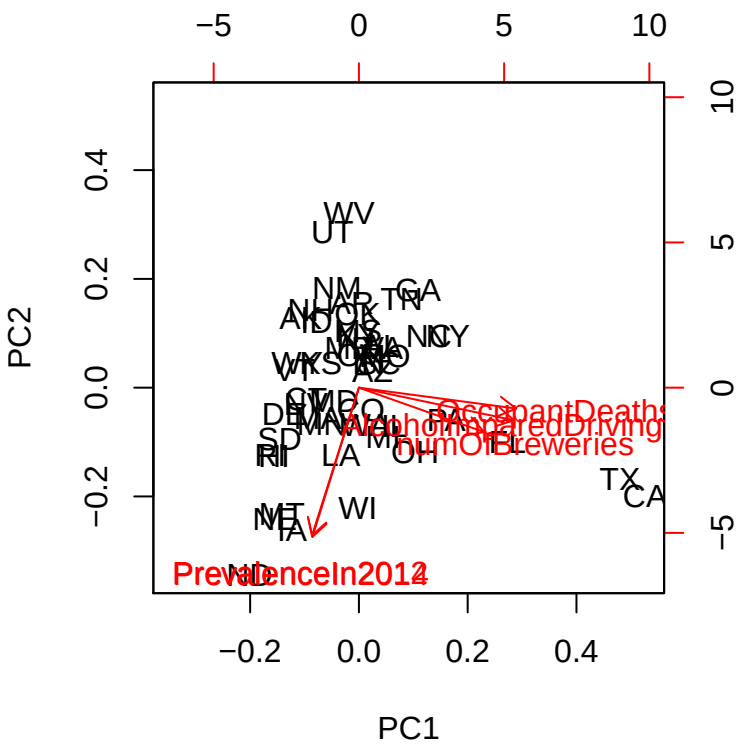
```
pc.comData <- prcomp(dataInterest, scale.=TRUE)
pc.comData$rotation
```

```
##                                    PC1         PC2        PC3         PC4
## numOfBreweries                0.4770741 -0.20317093 -0.8310709  0.20103545
## PrevalenceIn2012             -0.1789099 -0.67758144  0.2265219  0.67639908
## PrevalenceIn2014             -0.1773565 -0.68420893 -0.1037132 -0.69807258
## AlcoholImparedDrivingDeaths   0.5909780 -0.14937084  0.3547977 -0.10504874
## OccupantDeaths                0.5997343 -0.09566325  0.3483835 -0.06106123
##                                    PC5
## numOfBreweries                0.005266879
## PrevalenceIn2012             -0.006837468
## PrevalenceIn2014             -0.048384463
## AlcoholImparedDrivingDeaths   0.701082546
## OccupantDeaths               -0.711384366
```

```
plot(pc.comData, type="l")
```




```
biplot(pc.comData)
```


−5
0
5
10
0.4
0.2
0.0
−0.2
PC2
PC1
WV
UT
NM
GA
TN
NY
WI
LA
OH
TX
CA
PrevalenceIn2014