R Notebook

# PCA in Action

#download.file("https://raw.githubusercontent.com/mauriciogtec/PropedeuticoDataScience2017/master/Datos/DatosINEGI.csv", "inegi.csv")  
inegi = read.csv("inegi.csv")

Para determinar las variables per cápita para poder análisis son Secundarias, DefuncionesGenerales, Nacimientos, Divorcios y Matrimonios.

for(variable in c("Secundarias", "DefuncionesGenerales", "Nacimientos", "Divorcios", "Matrimonios")) {  
 inegi[, variable] = inegi[, variable] / inegi[, "Poblacion"]  
}  
#View(inegi)

Es importante limpiar los datos... guardando en una matriz x las variables numéricas a analizar (menos Estado y Población), y vamos a guardar como row.names.

x = inegi[, -(1:3)]  
print(x)

## Secundarias IndiceAprovechamientoSecundaria PorcentajeAnalfabetas  
## 1 0.0002936719 85.3 3.26  
## 2 0.0001885854 86.0 2.57  
## 3 0.0002448880 87.9 3.21  
## 4 0.0003866539 78.5 8.31  
## 5 0.0002012086 75.9 2.63  
## 6 0.0002597782 81.8 5.13  
## 7 0.0004125856 89.9 17.80  
## 8 0.0002266279 80.1 3.66  
## 9 0.0001588507 85.9 2.09  
## 10 0.0005683022 80.6 3.82  
## 11 0.0003082183 89.2 8.18  
## 12 0.0005178873 86.7 16.68  
## 13 0.0004472765 90.8 10.23  
## 14 0.0002607921 91.8 4.36  
## 15 0.0002386685 83.5 4.38  
## 16 0.0003668091 76.3 10.18  
## 17 0.0002655823 84.1 6.42  
## 18 0.0005170607 89.9 6.31  
## 19 0.0002073727 82.8 2.20  
## 20 0.0005818049 88.8 16.27  
## 21 0.0003749246 87.2 10.38  
## 22 0.0002691559 79.1 6.31  
## 23 0.0002866674 85.2 4.77  
## 24 0.0006331420 85.1 7.91  
## 25 0.0003226435 89.7 4.97  
## 26 0.0002569033 87.0 3.04  
## 27 0.0003314567 82.1 7.06  
## 28 0.0002276236 80.7 3.61  
## 29 0.0003034354 80.4 5.19  
## 30 0.0004188040 82.7 11.44  
## 31 0.0003047694 79.0 9.23  
## 32 0.0007728079 85.1 5.55  
## DefuncionesGenerales Nacimientos Divorcios Matrimonios  
## 1 0.004444741 0.02272835 0.0011957846 0.005240524  
## 2 0.004676917 0.01992697 0.0010440339 0.005557721  
## 3 0.004299668 0.02019384 0.0011679272 0.004483333  
## 4 0.004819799 0.02230312 0.0013836859 0.006404107  
## 5 0.005530509 0.02181858 0.0013345990 0.005597457  
## 6 0.005713583 0.02063161 0.0010114441 0.005207861  
## 7 0.004672287 0.03010958 0.0004419816 0.005014823  
## 8 0.006513204 0.02047137 0.0018209493 0.005497781  
## 9 0.006326347 0.01735732 0.0012546492 0.004014877  
## 10 0.005440514 0.02466052 0.0009877925 0.005924306  
## 11 0.005004035 0.02231201 0.0008759887 0.005706868  
## 12 0.005352093 0.02900317 0.0004966407 0.006667615  
## 13 0.004970323 0.02248540 0.0007189445 0.003805978  
## 14 0.005517175 0.02193225 0.0005826670 0.005322907  
## 15 0.004744442 0.02150863 0.0008493751 0.004775017  
## 16 0.005587174 0.02410575 0.0008241713 0.006167036  
## 17 0.005805111 0.02106990 0.0007826800 0.004430498  
## 18 0.005275678 0.02371567 0.0009954110 0.005633289  
## 19 0.005149934 0.02029459 0.0016196557 0.005789673  
## 20 0.005661551 0.02511992 0.0002056833 0.005502948  
## 21 0.005509333 0.02548449 0.0005003608 0.003637997  
## 22 0.004592609 0.02273984 0.0008878862 0.005061991  
## 23 0.003514693 0.02167809 0.0007747564 0.006925281  
## 24 0.005167630 0.02159993 0.0006841956 0.005202826  
## 25 0.005182528 0.02079045 0.0013588601 0.006173943  
## 26 0.005634221 0.02019471 0.0011335296 0.005447928  
## 27 0.004839625 0.02180199 0.0008107735 0.004957109  
## 28 0.005462966 0.01935504 0.0007205021 0.005582897  
## 29 0.004837017 0.02410645 0.0003863459 0.004914799  
## 30 0.006225800 0.02147976 0.0006024968 0.005410827  
## 31 0.005601416 0.01862724 0.0012661225 0.006115842  
## 32 0.005644449 0.02389868 0.0012088540 0.006415245  
## PorcentajePartosHospitales PorcentajeAguaPotable PorcentajeAguaEntubada  
## 1 97.1 98.0 98.9  
## 2 65.7 93.3 95.9  
## 3 95.2 86.7 92.4  
## 4 87.0 89.5 90.3  
## 5 90.3 97.9 98.2  
## 6 98.5 97.9 98.5  
## 7 33.2 74.5 78.1  
## 8 78.0 93.9 95.2  
## 9 94.8 97.8 97.8  
## 10 81.3 93.0 94.7  
## 11 85.5 93.5 94.7  
## 12 60.7 70.6 71.3  
## 13 78.4 88.9 91.2  
## 14 88.5 93.0 96.0  
## 15 91.1 92.4 94.4  
## 16 85.6 90.7 92.1  
## 17 86.9 91.6 91.8  
## 18 82.0 93.1 93.0  
## 19 96.4 95.5 96.2  
## 20 59.7 76.7 77.2  
## 21 80.0 85.3 88.1  
## 22 94.2 93.1 95.0  
## 23 89.9 92.8 92.2  
## 24 86.9 84.9 86.4  
## 25 94.1 94.4 94.9  
## 26 95.4 95.3 96.5  
## 27 80.4 76.9 82.2  
## 28 96.5 96.4 95.5  
## 29 93.6 98.2 98.2  
## 30 75.3 76.8 81.1  
## 31 89.9 96.0 97.0  
## 32 91.4 94.1 94.4  
## PorcentajeElectricidad PorcentajeParedesSolidas PorcentajePisoTierra  
## 1 99.2 92.3 1.7  
## 2 98.5 77.0 3.3  
## 3 96.7 90.3 5.8  
## 4 96.8 80.7 4.7  
## 5 99.1 84.8 1.6  
## 6 99.0 94.7 4.5  
## 7 95.9 67.4 14.7  
## 8 96.3 77.2 3.2  
## 9 99.5 97.9 1.0  
## 10 96.1 66.7 6.3  
## 11 98.2 93.2 4.1  
## 12 95.3 63.0 18.4  
## 13 96.9 91.3 7.1  
## 14 99.0 91.9 3.0  
## 15 98.9 93.9 3.8  
## 16 98.0 80.3 10.3  
## 17 98.8 90.6 7.2  
## 18 96.9 90.0 4.0  
## 19 98.3 94.0 2.0  
## 20 94.3 66.4 18.7  
## 21 97.7 87.3 9.5  
## 22 97.7 96.6 3.7  
## 23 96.2 87.2 3.7  
## 24 95.6 80.9 8.7  
## 25 98.4 94.7 6.1  
## 26 97.9 88.2 5.3  
## 27 98.4 87.5 6.4  
## 28 97.0 85.5 3.3  
## 29 98.5 90.3 3.9  
## 30 96.6 81.5 11.7  
## 31 97.4 93.2 2.8  
## 32 98.3 63.4 3.4

library(FactoMineR)  
model = PCA(x)

