Contents

1 Setup

```
using Pkg;
Pkg.activate("~/PP/MonitoriaEstatistica/")
Pkg.add("PlotlyJS")
Pkg.add("CSV")
Pkg.add("DataFrames")
Pkg.add("MLJ")
Pkg.add("MLJMultivariateStatsInterface")
```

2 PCA

```
using PlotlyJS, CSV, DataFrames, MLJ
```

(Faça o download do conjunto)

```
df = DataFrame(CSV.File("../data/csv/cigarro.csv"))
```

```
features = names(df)
```

```
# load and fit PCA
PCA = @load PCA pkg="MultivariateStats"
mach = machine(PCA(pratio=1), df[!, features])
fit!(mach)
```

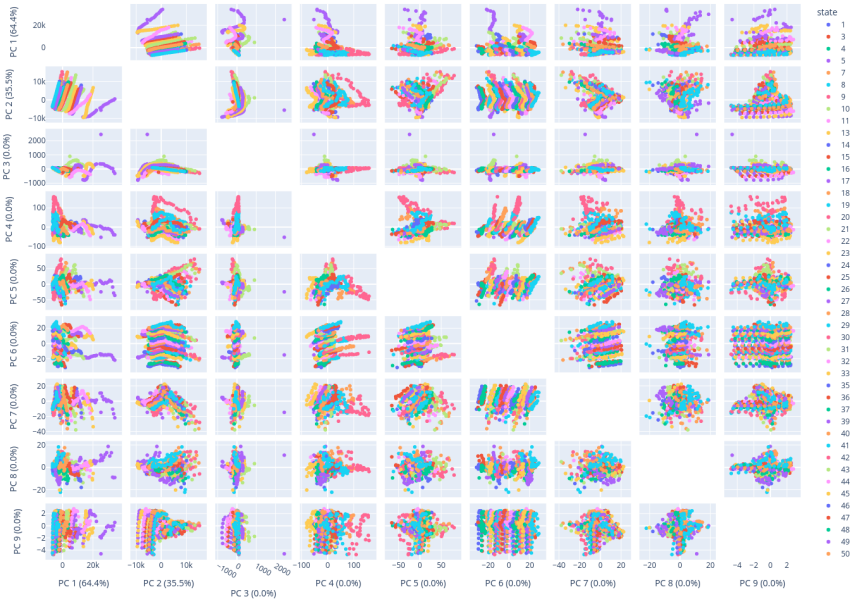
```
mach.report
```

```
# compute explained variance for each dimension
explained_variance = report(mach).principalvars
explained_variance ./= sum(explained_variance)
explained_variance .*= 100

# transform data to get components
components = MLJ.transform(mach, df[!, features])
dimensions = Symbol.(names(components))
components.state = df.state

labels = attr(;
              [
                  dimensions[i] => "PC $i ($v%)"
                  for (i, v) in enumerate(round.(explained_variance,
                  digits=1))
                      ]...
                          )

# plot
plot(components, dimensions=dimensions, labels=labels,
     color=:state, kind="splom")
```



3 Fontes:

https://plotly.com/julia/pca-visualization/ https://r-data.pmagunia.com/dataset/r-dataset-package-plm-cigar