

Ejercicio Obligatorio 2

- Patrick Dey
- Matias Lombardi
- Santos Rosati

Variables:

Country, Area, GDP, Inflation, Life.expect, Military, Pop.growth, Unemployment

Librerías:

- pca
- numpy

```
with open('europe.csv', newline='') as csvfile:
    # create np array with csv rows
    rows = np.array(list(csv.reader(csvfile, delimiter=',')))
    headers = np.array(rows[0][1:])
    rows = rows[1:] # ignoro headers
    countries = [row[0] for row in rows] # me quedo con los paises para poner los labels
    rows = np.array([row[1:] for row in rows]) # elimino pais

    csvfile.close()
```

```
# Initialize to reduce the data up to the number of componentes that explains 95% of the variance.
model = pca(n_components=0.95, normalize=True)

print(headers)
# Fit transform
results = model.fit_transform(rows, col_labels=headers, row_labels=countries)

# Plot explained variance
fig, ax = model.plot()

# Scatter first 2 PCs
fig, ax = model.scatter()

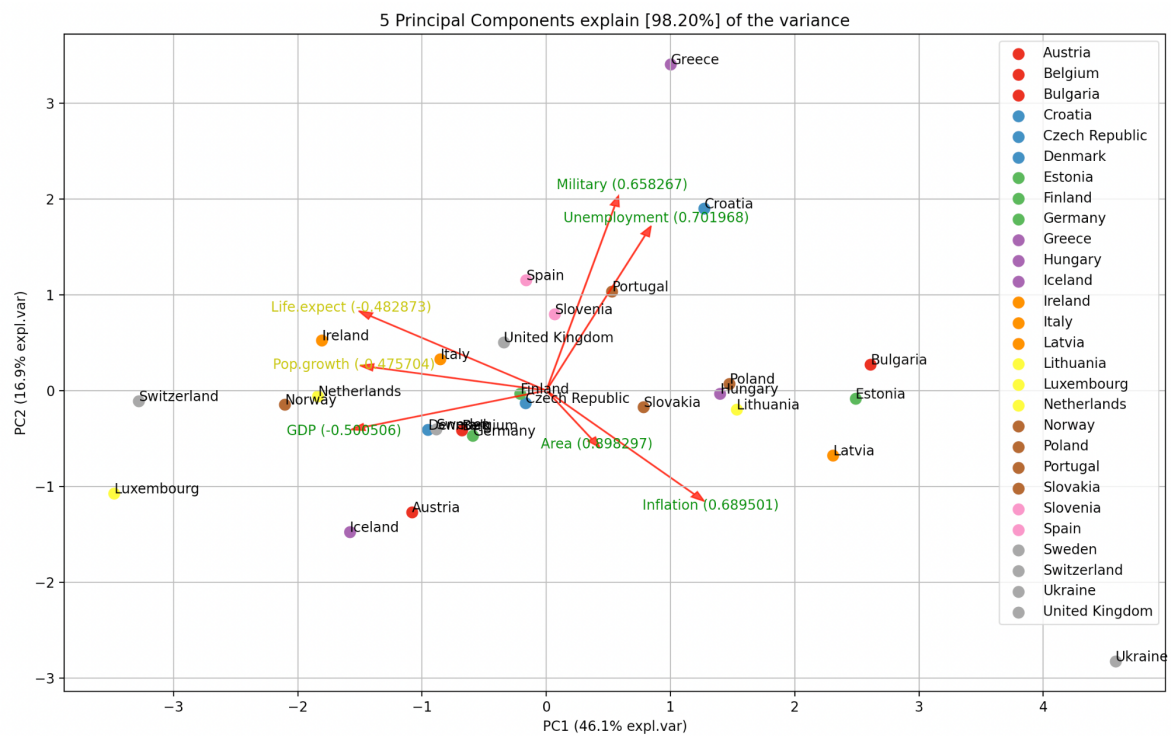
# Make biplot with the number of features
fig, ax = model.biplot()
```

$$y = 0.124874 * x_1 - 0.500506 * x_2 + 0.406518 * x_3 - 0.482873 * x_4 + 0.188112 * x_5 - 0.475704 * x_6 + 0.271656 * x_7$$

PC1	0.124874	-0.500506	0.406518	-0.482873	0.188112	-0.475704	0.271656
PC2	-0.172872	-0.130140	-0.369657	0.265248	0.658267	0.082622	0.553204
PC3	0.898297	0.083956	0.198195	0.246082	0.243679	0.163697	0.000500
PC4	0.044850	-0.084255	0.164686	0.026771	-0.562375	0.392463	0.701968

PC5 -0.324017 0.390632 0.689501 -0.101787 0.368148 0.347868 0.010159

La variable que más influencia tiene sobre este “índice” es el GPD, seguido por life expectancy



Variabilidad de las cargas

x1	x2	x3	x4	x5	x6
0.46102367	0.16958906	0.15188436	0.11005085	0.06540695	0.02409627
0.01794884					