

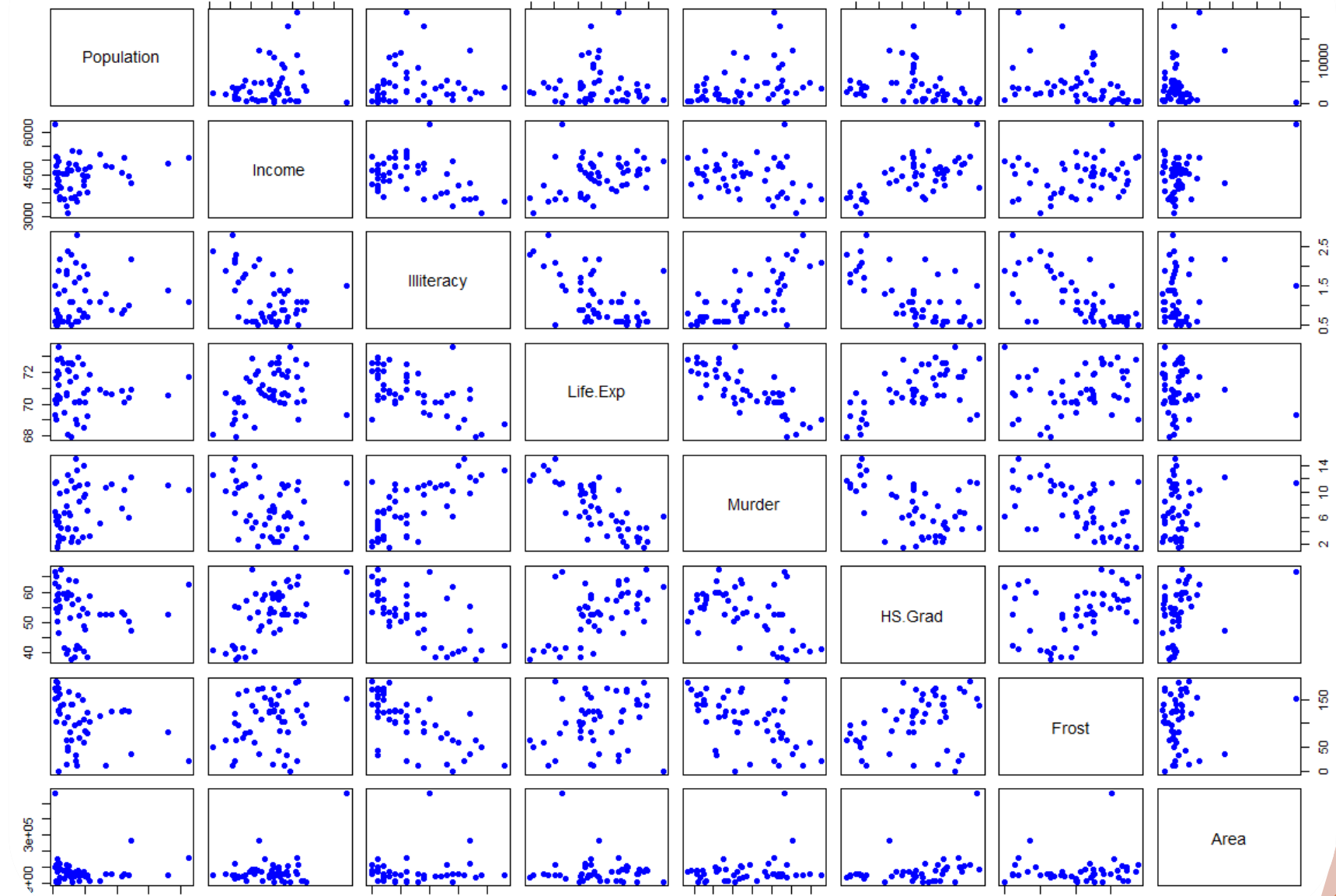


Ejemplo

Ejemplo

Consideremos las siguientes variables univariantes medidas en 50 estados de USA:

- ▶ x_1 : estimación de población (en miles) a fecha 1ro Julio, 1975
- ▶ x_2 : ingresos per cápita (en dólares) en el año 1974
- ▶ x_3 : analfabetismo (porcentaje) en el año 1970
- ▶ x_4 : esperanza de vida entre los años 1969-1971
- ▶ x_5 : tasa de homicidios y homicidios no negligentes (por 100000) 1976
- ▶ x_6 : porcentaje de graduados de secundaria 1970
- ▶ x_7 : número medio de días con temperatura mínima bajo cero (1931-1960) en la capital o gran ciudad
- ▶ x_8 : área de terreno en millas cuadradas



Vector de media muestral $\bar{\mathbf{x}}$

Population	4246.4200
Income	4435.8000
Illiteracy	1.1700
Life.Exp	70.8786
Murder	7.3780
HS.Grad	53.1080
Frost	104.4600
Area	70735.8800

Matriz de covarianza muestral S_x

Population	19931683.7588	571229.7796	292.8679592	-4.078425e+02	5663.523714	-3551.509551	-77081.97265	8.587917e+06
Income	571229.7796	377573.3061	-163.7020408	2.806632e+02	-521.894286	3076.768980	7227.60408	1.904901e+07
Illiteracy	292.8680	-163.7020	0.3715306	-4.815122e-01	1.581776	-3.235469	-21.29000	4.018337e+03
Life.Exp	-407.8425	280.6632	-0.4815122	1.802020e+00	-3.869480	6.312685	18.28678	-1.229410e+04
Murder	5663.5237	-521.8943	1.5817755	-3.869480e+00	13.627465	-14.549616	-103.40600	7.194043e+04
HS.Grad	-3551.5096	3076.7690	-3.2354694	6.312685e+00	-14.549616	65.237894	153.99216	2.298732e+05
Frost	-77081.9727	7227.6041	-21.2900000	1.828678e+01	-103.406000	153.992163	2702.00857	2.627039e+05
Area	8587916.9494	19049013.7510	4018.3371429	-1.229410e+04	71940.429959	229873.192816	262703.89306	7.280748e+09

Matriz de auto-vectores $V_8^{S_x}$

	V1	V2	V3	V4	V5	V6	V7	V8
1	-1.182966e-03	9.996005e-01	0.0278490777	-4.671254e-03	3.349393e-04	1.386661e-04	-5.183454e-05	2.191850e-05
2	-2.616550e-03	2.796866e-02	-0.9991766328	2.821732e-02	-7.792882e-03	-1.119562e-04	3.850506e-05	6.290403e-05
3	-5.518945e-07	1.420515e-05	0.0005844687	7.100747e-03	-4.054743e-02	-3.091522e-02	2.550656e-02	9.983480e-01
4	1.688521e-06	-1.928393e-05	-0.0010367078	-3.875966e-03	1.193295e-01	2.855357e-01	9.508427e-01	-1.057617e-02
5	-9.881522e-06	2.787128e-04	0.0027764911	2.816092e-02	-2.386638e-01	-9.200852e-01	3.058552e-01	-4.620107e-02
6	-3.157288e-05	-1.882545e-04	-0.0082661337	-2.784545e-02	9.622385e-01	-2.656351e-01	-4.075556e-02	3.209925e-02
7	-3.607163e-05	-3.871630e-03	-0.0280421226	-9.987733e-01	-3.452920e-02	-1.986814e-02	6.252701e-03	4.942864e-03
8	-9.999959e-01	-1.255538e-03	0.0025827049	-3.168841e-05	-6.558672e-06	1.882356e-05	-4.090819e-07	-1.494594e-06

Los auto-valores de S_x

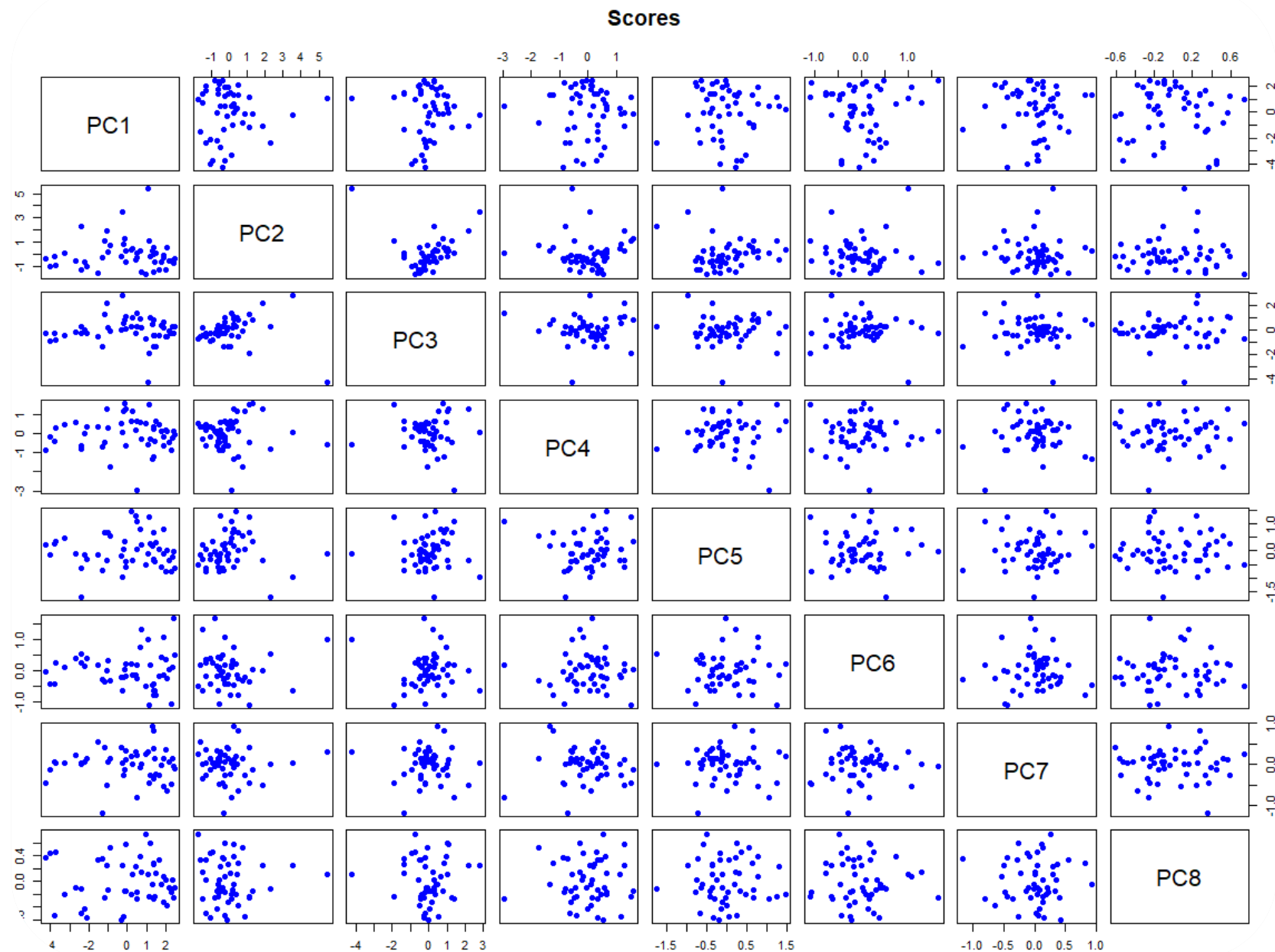
i	λ_i	proporción de variabilidad	proporción de variabilidad acumulada
1	7.28×10^9	0.997	0.997
2	1.99×10^7	2.73×10^{-3}	0.9999569
3	3.12×10^5	4.28×10^{-5}	0.9999997
4	2.15×10^3	2.94×10^{-7}	1
5	36.5	5×10^{-9}	1
6	6.05	8.29×10^{-10}	1
7	0.43	5.93×10^{-11}	1
8	0.08	1.15×10^{-11}	1

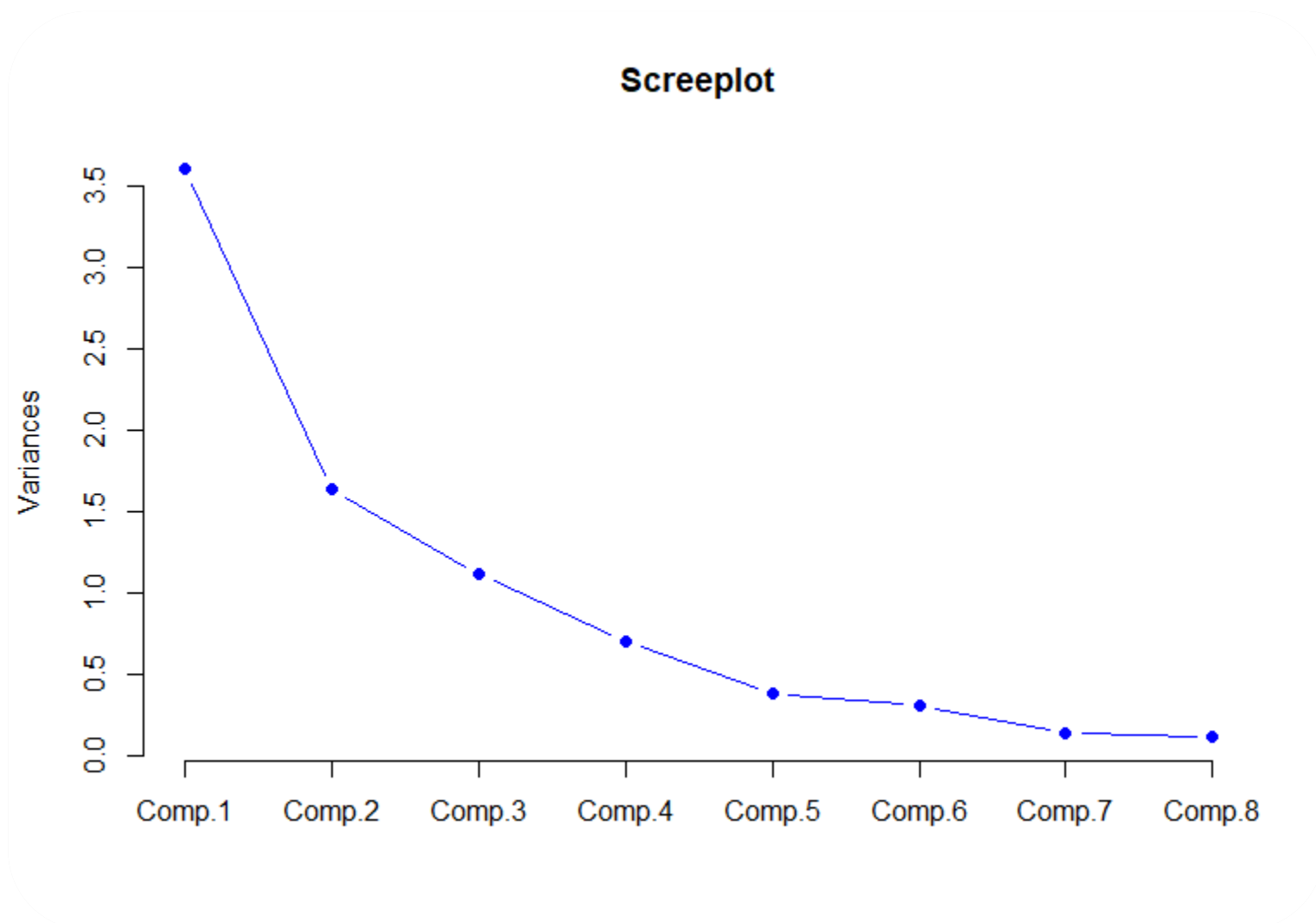
Matriz de correlaciones muestrales R_x

Population	1	0.2082276	0.10762237	-0.06805195	0.3436428	-0.09848975	-0.3321525	0.02254384
Income	0.20822756	1	-0.43707519	0.34025534	-0.2300776	0.61993232	0.2262822	0.36331544
Illiteracy	0.10762237	-0.4370752	1	-0.58847793	0.7029752	-0.65718861	-0.6719470	0.07726113
Life.Exp	-0.06805195	0.3402553	-0.58847793	1	-0.7808458	0.58221620	0.2620680	-0.10733194
Murder	0.34364275	-0.2300776	0.70297520	-0.78084575	1	-0.48797102	-0.5388834	0.22839021
HS.Grad	-0.09848975	0.6199323	-0.65718861	0.58221620	-0.4879710	1	0.3667797	0.33354187
Frost	-0.33215245	0.2262822	-0.67194697	0.26206801	-0.5388834	0.36677970	1	0.05922910
Area	0.02254384	0.3633154	0.07726113	-0.10733194	0.2283902	0.33354187	0.0592291	1

Los auto-valores de R_x

i	λ_i	proporción de variabilidad	proporción de variabilidad acumulada
1	3.59	0.449	0.449
2	1.63	0.203	0.653
3	1.11	0.138	0.792
4	0.71	0.088	0.881
5	0.38	0.048	0.929
6	0.31	0.038	0.967
7	0.14	0.018	0.985
8	0.11	0.014	1





Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8

Ejemplo

- ▶ En este caso podemos seleccionar los primeros tres componentes principales ya que explican el 79% de la variabilidad total.
- ▶ Además, la media de los autovalores es igual a 1 y $\lambda_3^{R_x} = 1.11$ que está por arriba de la media, y $\lambda_4^{R_x} = 0.7$ que está por debajo.
- ▶ El 1er componente es combinación lineal con las variables originales centradas:

$$z_1 = -0.12\tilde{x}_1 + 0.29\tilde{x}_2 - 0.46\tilde{x}_3 + 0.41\tilde{x}_4 - 0.44\tilde{x}_5 + 0.42\tilde{x}_6 + 0.35\tilde{x}_7 + 0.03\tilde{x}_8$$

- ▶ El 1er componente distingue entre estados: con poblaciones ricas, longevas, con buena educación, temperatura fría y más grandes, de estados: pobres, con menor esperanza de vida, peor educación, más cálidos y más pequeños.

Ejemplo

- El 2do componente es combinación lineal con las variables originales centradas:

$$z_2 = 0.41\tilde{x}_1 + 0.51\tilde{x}_2 + 0.05\tilde{x}_3 - 0.08\tilde{x}_4 + 0.30\tilde{x}_5 + 0.29 - 0.15\tilde{x}_7 + 0.58\tilde{x}_8$$

- El 2do componente distingue entre estados: con poblaciones densas y de gran área, con buena educación, y más riqueza, pero a la vez más violentos, de estados: pequeños y menos poblados, con pobreza y peor educación.

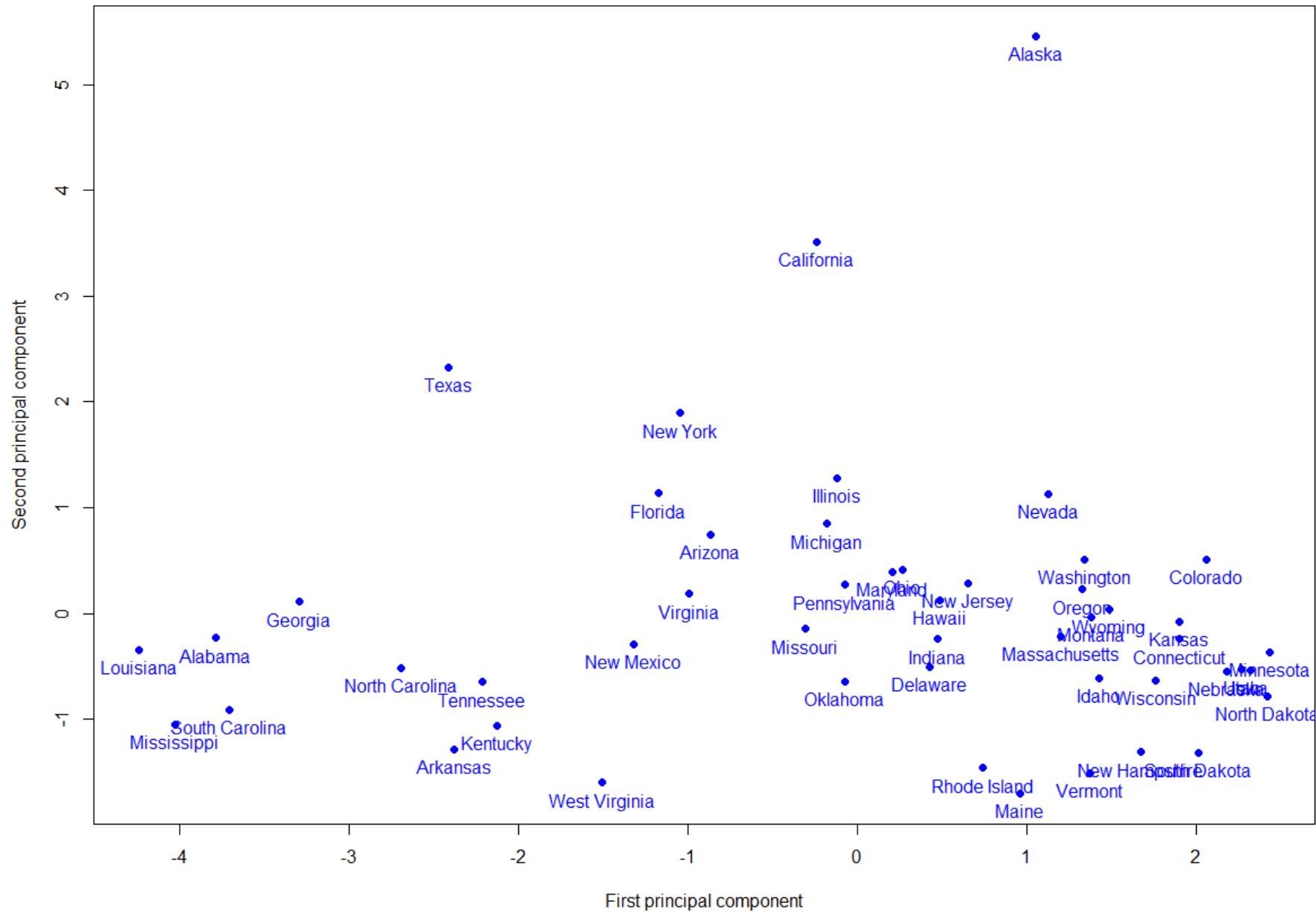
Ejemplo

- ▶ El 3er componente es combinación lineal con las variables originales centradas:

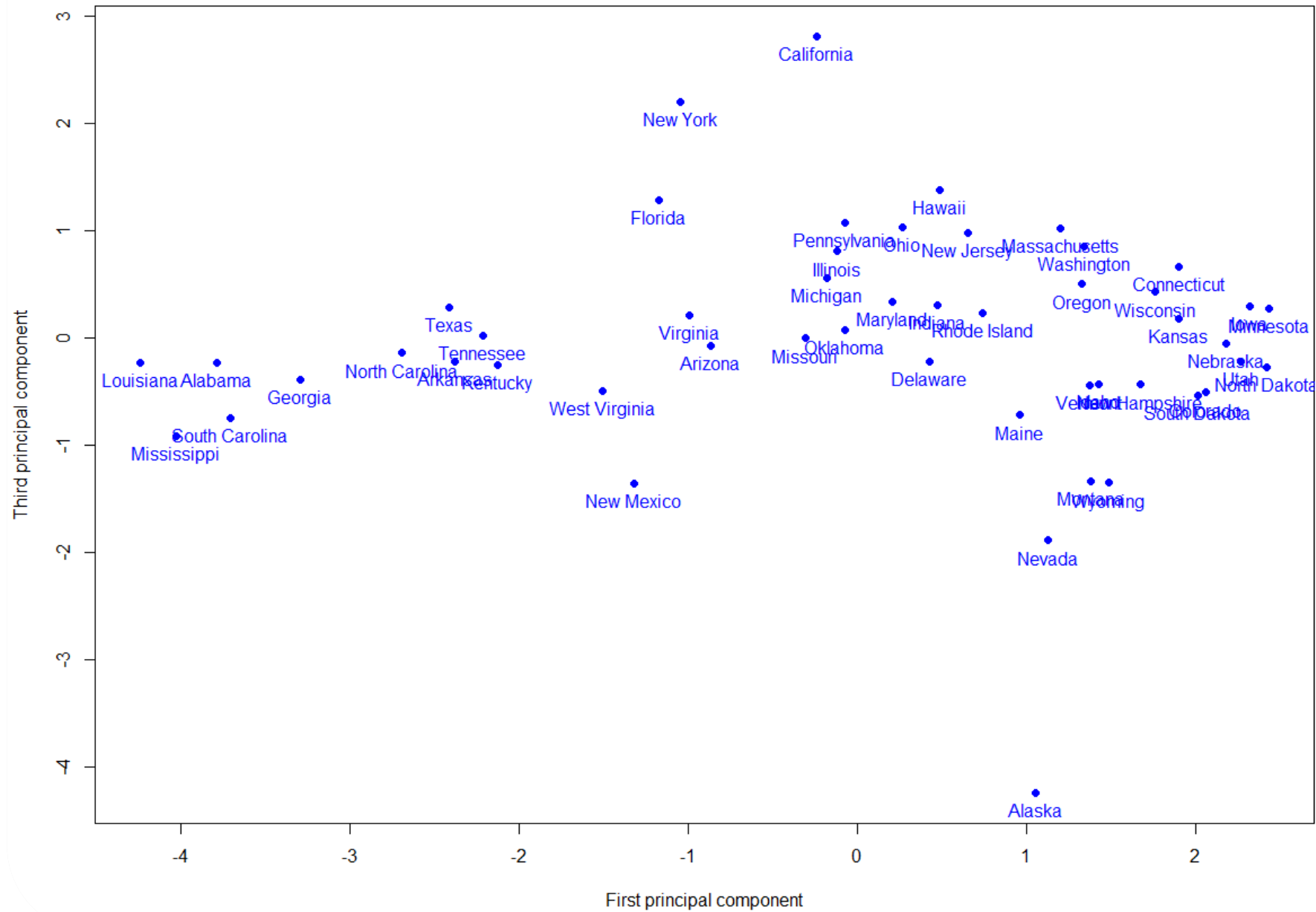
$$z_3 = 0.65\tilde{x}_1 + 0.10\tilde{x}_2 - 0.07\tilde{x}_3 + 0.35\tilde{x}_4 - 0.10\tilde{x}_5 - 0.04\tilde{x}_6 - 0.38\tilde{x}_7 - 0.51\tilde{x}_8$$

- ▶ El 3er componente distingue entre estados: muy poblados, ricos y longevos, de estados: más cálidos y más grandes que tienden a tener peor educación y a tener más violencia.

First and second principal component



First and third principal component



Second and third principal component

