



Universidade de Brasília  
Departamento de Estatística  
EST0077 - 2023/1

## **SOLUÇÃO DOS EXERCÍCIOS PARA ENTREGA 3**

Igor de Oliveira Barros Faluhelyi  
Prof. George

Brasília  
2023

## Sumário

1 Exercício 27 da Lista 4 . . . . .	4
-------------------------------------	---

## 1 Exercício 27 da Lista 4

**Enunciado:** Este estudo se refere a dados sobre uma escala de felicidade em diferentes países do mundo. O nome do trabalho é World Happiness Report 2023. O trabalho é realizado por um grupo de pesquisadores independentes e, segundo os autores, não reflete a visão de qualquer organização, agência ou programa das Nações Unidas. A planilha *WolrdHappinessReport.xls* possui a seguinte informação:

**Country Name:** país avaliado. **Score:** escala de felicidade baseada na avaliação média de 2020-2022.

**Logged GDP per capita:** log do produto interno bruto (PIB), o quanto cada país produz, dividido pelo número de pessoas no país.

**Social support:** apoio social, ou se tem alguém com quem contar em tempos de dificuldade.

**Heathy Life Expectancy:** situação de saúde física e mental.

**Freedom to make life choices:** nível de satisfação com a liberdade de escolhas.

**Generosity:** senso de comunidade e engajamento positivo com a comunidade.

**Perceptions of corruption:** percepção sobre nível de corrupção no governo e nas empresas.

**Dystopia:** uma constante que representa um país imaginário com os piores indicadores de felicidade. Seria o país com a população mais infeliz no mundo.

Com base nos dados do arquivo,

- (a) Selecionar aleatoriamente 20 países entre os 69 melhor classificados (maior score) e 20 países entre os 69 com pior classificação.
- (b) Fazer uma análise de componentes principais para os 40 países selecionados. Inclua um biplot em sua análise.
- (c) Retirar, aleatoriamente, 25% da informação disponível, entre os dados selecionados em (a) e desconsiderando nome do país.
- (d) Fazer o preenchimento da matriz e repetir o estudo. Suas conclusões em (b) continuam as mesmas? Justifique?

**Resposta:**

(a) Dentre os 69 melhores e 69 piores com respeito ao *Score*, foram selecionados países, aleatoriamente, seguindo o arquivo *.py* (as variáveis apresentadas foram normalizadas):

Figura 1: Print dos 20 primeiros países selecionados

Country name	Score	Logged GDP per capita	Social support	...	Generosity	Perceptions of corruption	Country name
China	0.665899	0.686644	0.770998	...	0.271737	0.742187	China
Chile	0.752674	0.747934	0.854331	...	0.309608	0.864186	Chile
Greece	0.684905	0.776334	0.769832	...	0.018721	0.826473	Greece
Montenegro	0.649801	0.698795	0.855071	...	0.404448	0.891885	Montenegro
El Salvador	0.716982	0.580835	0.644740	...	0.186303	0.604745	El Salvador
Kazakhstan	0.720733	0.756428	0.900725	...	0.323631	0.734621	Kazakhstan
Switzerland	0.609117	0.919142	0.909159	...	0.357612	0.153562	Switzerland
Moldova	0.686117	0.647629	0.803650	...	0.221713	0.963942	Moldova
Jamaica	0.646522	0.593143	0.820305	...	0.188401	0.931904	Jamaica
Bahrain	0.725661	0.855840	0.783704	...	0.472597	0.754283	Bahrain
Uruguay	0.779604	0.735304	0.891882	...	0.241400	0.547228	Uruguay
Estonia	0.773094	0.817615	0.942440	...	0.364076	0.336350	Estonia
Kosovo	0.758360	0.624855	0.783621	...	0.652943	0.919581	Kosovo
Spain	0.769898	0.817445	0.920783	...	0.239732	0.720846	Spain
Costa Rica	0.798880	0.721454	0.827049	...	0.234797	0.794018	Costa Rica
Honduras	0.706313	0.586831	0.651625	...	0.447557	0.889922	Honduras
Mongolia	0.669640	0.627027	0.922478	...	0.566066	0.897356	Mongolia
Guatemala	0.721809	0.585171	0.733623	...	0.250764	0.881844	Guatemala
Netherlands	0.932517	0.882986	0.918848	...	0.594355	0.297348	Netherlands
Taiwan Province of China	0.786584	0.859374	0.846991	...	0.159507	0.683123	Taiwan Province of China

[20 rows x 8 columns]

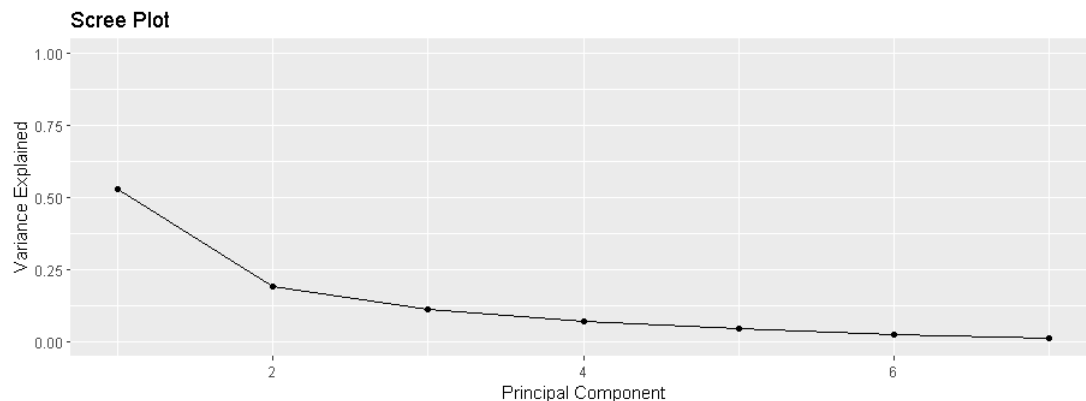
Figura 2: Print dos 20 últimos países selecionados

Country name	Score	Logged GDP per capita	Social support	Healthy life expectancy	Freedom to make life choices	Generosity	Perceptions of corruption	Country name
Botswana	0.265138	0.668918	0.642507	0.124071	0.621987	0.050474	0.873946	Botswana
Mauritania	0.481885	0.499605	0.471859	0.347376	0.414953	0.307354	0.657308	Mauritania
India	0.366144	0.927009	0.410251	0.352100	0.880423	0.414930	0.801391	India
Sierra Leone	0.215004	0.304455	0.333613	0.130862	0.480755	0.457201	0.908925	Sierra Leone
Bangladesh	0.409573	0.514926	0.316798	0.505570	0.799230	0.330074	0.706094	Bangladesh
Sri Lanka	0.434401	0.640362	0.735015	0.606040	0.030418	0.205476	0.845974	Sri Lanka
Paru	0.616703	0.631771	0.712145	0.711454	0.711888	0.172369	0.952262	Paru
Malawi	0.275214	0.209375	0.295474	0.209718	0.635418	0.330256	0.769929	Malawi
Lebanon	0.089686	0.644189	0.793048	0.567722	0.159131	0.144208	0.951486	Lebanon
Tunisia	0.443669	0.606018	0.605708	0.600722	0.336084	0.053041	0.571452	Tunisia
Nigeria	0.525002	0.404103	0.621454	0.130512	0.580859	0.416369	0.977110	Nigeria
Iraq	0.518385	0.582354	0.588142	0.461543	0.455579	0.316773	0.932533	Iraq
Senegal	0.509953	0.420578	0.445078	0.320888	0.672025	0.337380	0.893742	Senegal
Morocco	0.511976	0.561014	0.330274	0.408418	0.099663	0.020900	0.840956	Morocco
Georgia	0.546676	0.671676	0.504504	0.521168	0.698262	0.000000	0.642139	Georgia
Egypt	0.388801	0.626115	0.600067	0.464972	0.605045	0.090841	0.554466	Egypt
Tanzania	0.300619	0.300819	0.405207	0.305606	0.707015	0.555285	0.520923	Tanzania
Kenya	0.441970	0.477976	0.543774	0.770618	0.541509	0.630362	0.902080	Kenya
Colombia	0.634361	0.661594	0.749117	0.692030	0.728949	0.190873	0.877975	Colombia
Gabon	0.534162	0.653940	0.630330	0.261037	0.448685	0.085615	0.817652	Gabon

PS: C:\Users\lgon\Desktop>

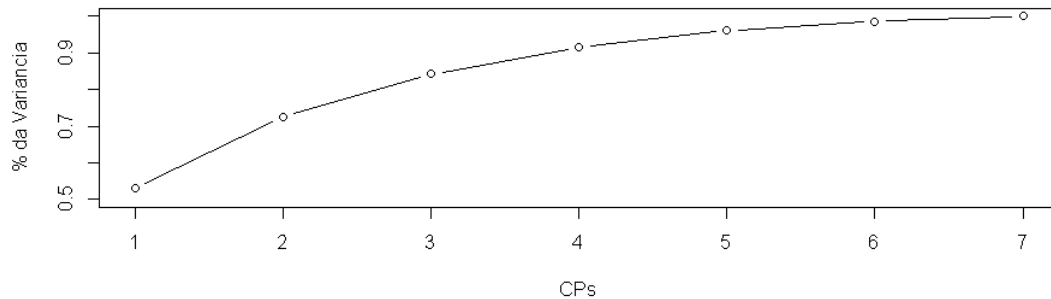
(b)

Figura 3: Screen plot



A primeira componente principal explica 53% da variância total compreendida pelos dados, a segunda explica 19.5%, a terceira 11.6%, a quarta 7.3%, quinta 4.5%, sexta 2.5% e a sétima 1.4%.

Figura 4: % Variância explicada



Somente as componentes 1, 2 e 3 juntas explicam aproximadamente 84% da variância total. Isso significa que podemos reduzir de 7 CPs para 3 (as três primeiras).

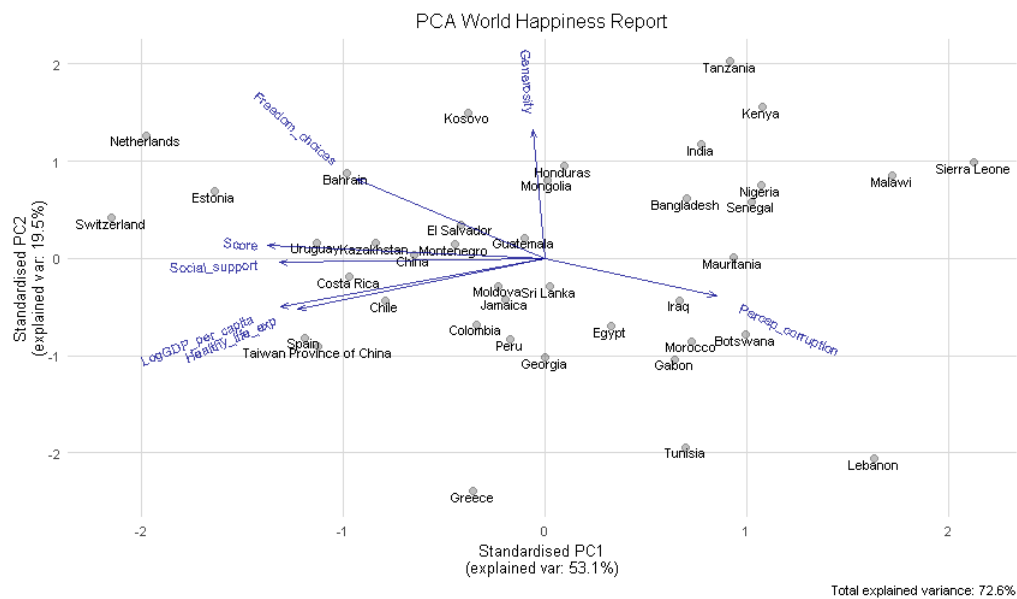
Figura 5: Rotation matrix

```
> cprecor$rotation
```

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Score	-0.47326371	0.07330535	-0.245163927	0.10394656	-0.05829869	0.681642583	-0.48135454
LogGDP_per_capita	-0.45109284	-0.28437091	-0.002265122	-0.24956980	-0.18958955	-0.610717601	-0.49440675
Social_support	-0.45292494	-0.02538115	-0.395062195	-0.10528165	-0.43642273	-0.008534832	0.66069568
Healthy_life_exp	-0.42151893	-0.29375882	0.035505596	-0.04632434	0.81744373	0.007808987	0.25366461
Freedom_choices	-0.32206404	0.46252964	0.152369498	0.74807133	0.03144971	-0.312965466	0.02403009
Generosity	-0.02188325	0.75217117	-0.305641704	-0.48778145	0.26832845	-0.159102184	-0.07140306
Percep_corruption	0.29242670	-0.21708506	-0.816040083	0.34079085	0.17026093	-0.197465667	-0.13160322

Através da *Rotation matrix* pode-se observar, dentre as 3 principais componentes (1, 2 e 3), qual variável original dos dados é importante dentro de cada componente. Por exemplo, a variável *Generosity* não é importante para a primeira componente principal. Ainda, é possível observar que a CP1 está relacionada negativamente com o *Score*, *Logged GDP per capita*, *Social support*, *Heathy Life Expectancy* e *Freedom to make life choices* o que significa, de acordo com a pesquisa, países mais felizes. A CP1 se relaciona positivamente com a *Perceptions of corruption* e também significa países com mais desconfiança com relação ao governo e empresas, o que sugere um *trade-off* entre felicidade e *Perceptions of corruption* nos países. Já a segunda CP está relacionada positivamente com *Freedom to make life choices* e *Generosity* e negativamente com *Logged GDP per capita*, *Heathy Life Expectancy* e *Perceptions of corruption*.

Figura 6: Biplot



(c) Foi retirada 25% da informação disponível, vide arquivo .py.

Figura 7: Dados em (a)

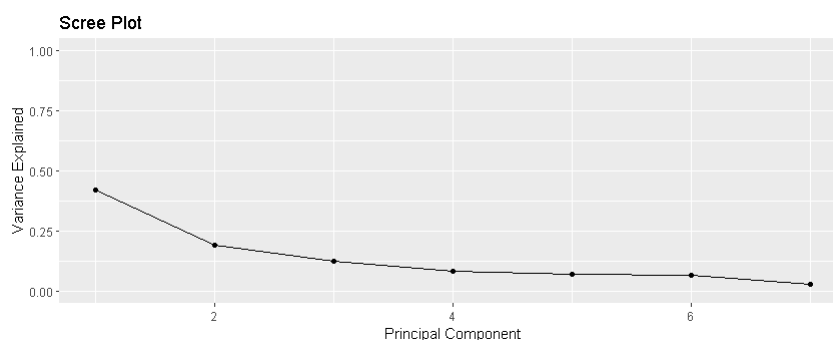
	A	Y	B	Y	C	Y	D	Y	E	Y	F	Y	G	Y	H	Y
1	Score		Logged GDP per capita		Social support		Healthy life expectancy		Freedom to make life choices		Generosity		Perceptions of corruption		Country name	
2	0.67		0.69	0.71	0.67		0.66	0.67	0.65		0.67		0.64		China	
3	0.75		0.75	0.65	0.73		0.73	0.71	0.31		0.66		0.66		China	
4	0.68		0.76	0.77	0.76		0.76	0.32	0.02		0.63		0.63		Greece	
5	0.65		0.70	0.66	0.60		0.73	0.40			0.69		0.69		Montenegro	
6	0.72		0.58	0.64	0.55		0.62	0.19			0.60		0.60		El Salvador	
7	0.72		0.76	0.82	0.55		0.81	0.32			0.73		0.73		Kazakhstan	
8	0.91		0.92	0.90	0.83		0.88	0.36			0.15		0.15		Switzerland	
9	0.67		0.65	0.80	0.53		0.79	0.22			0.66		0.66		Moldova	
10	0.65		0.59	0.82	0.59		0.76	0.19			0.93		0.93		Jamaica	
11	0.73		0.86	0.78	0.56		0.97	0.47			0.75		0.75		Bahrain	
12	0.78		0.74	0.89	0.62		0.89	0.24			0.55		0.55		Uruguay	
13	0.77		0.82	0.94	0.70		0.94	0.36			0.34		0.34		Estonia	
14	0.76		0.62	0.78	0.53		0.83	0.65			0.62		0.62		Kosovo	
15	0.77		0.82	0.92	0.81		0.69	0.24			0.72		0.72		Spain	
16	0.80		0.72	0.83	0.72		0.89	0.23			0.79		0.79		Costa Rica	
17	0.70		0.51	0.66	0.69		0.80	0.45			0.69		0.69		Honduras	
18	0.67		0.63	0.92	0.35		0.55	0.57			0.90		0.90		Mongolia	
19	0.72		0.57	0.73	0.64		0.82	0.25			0.88		0.88		Guatemala	
20	0.51		0.88	0.62	0.70		0.87	0.59			0.30		0.30		Netherlands	
21	0.79		0.86	0.85	0.70		0.73	0.16			0.69		0.69		Taiwan Province of China	
22	0.27		0.67	0.64	0.12		0.62	0.05			0.07		0.07		Botswana	
23	0.46		0.50	0.47	0.35		0.41	0.31			0.65		0.65		Mauritania	
24	0.37		0.53	0.42	0.36		0.89	0.41			0.80		0.80		India	
25	0.22		0.30	0.33	0.13		0.48	0.46			0.91		0.91		Sierra Leone	
26	0.41		0.51	0.32	0.51		0.80	0.33			0.71		0.71		Bangladesh	
27	0.43		0.65	0.76	0.61		0.70	0.29			0.85		0.85		Sri Lanka	
28	0.62		0.63	0.71	0.71		0.71	0.17			0.95		0.95		Peru	
29	0.28		0.29	0.30	0.27		0.64	0.33			0.77		0.77		Malawi	
30	0.09		0.64	0.29	0.57		0.16	0.14			0.95		0.95		Lebanon	
31	0.44		0.61	0.61	0.60		0.34	0.05			0.97		0.97		Tunisia	
32	0.53		0.48	0.62	0.13		0.58	0.42			0.96		0.96		Nigeria	
33	0.52		0.58	0.59	0.45		0.45	0.32			0.93		0.93		Iraq	
34	0.30		0.43	0.45	0.33		0.67	0.34			0.89		0.89		Senegal	
35	0.51		0.56	0.33	0.69		0.70	0.03			0.85		0.85		Morocco	
36	0.55		0.67	0.38	0.52		0.70	0			0.64		0.64		Georgia	
37	0.39		0.63	0.60	0.66		0.61	0.09			0.55		0.55		Egypt	
38	0.31		0.36	0.49	0.31		0.79	0.56			0.52		0.52		Tanzania	
39	0.44		0.48	0.54	0.27		0.54	0.69			0.90		0.90		Kenya	
40	0.63		0.66	0.75	0.69		0.73	0.19			0.88		0.88		Colombia	
41	0.53		0.65	0.63	0.26		0.45	0.09			0.82		0.82		Gabon	

Figura 8: Dados em (a) com missing

	A	T	B	T	C	T	D	T	E	T	F	T	G	T	H	T
1	Score		Logged GDP per capita		Social support		Healthy life expectancy		Freedom to make life choices		Generosity		Perceptions of corruption		Country name	
2	0.67		0.69		0.85		0.67		0.27		0.31		0.74		China	
3	0.75		0.75		0.85		0.73		0.31		0.31		0.86		Chile	
4	0.68		0.78		0.77		0.76		0.32		0.02		0.83		Greece	
5	0.65		0.70		0.86		0.60		0.73		0.40		0.89		Montenegro	
6	0.72		0.58		0.64				0.92		0.19		0.60		El Salvador	
7			0.76		0.92		0.55				0.32		0.73		Kazakhstan	
8			0.92		0.90				0.88		0.36		0.15		Switzerland	
9	0.67		0.65		0.80		0.53				0.22		0.22		Moldova	
10					0.82				0.76		0.19				Armenia	
11	0.73		0.86				0.86		0.97				0.75		Bahrain	
12	0.78		0.74		0.89		0.82		0.89		0.24		0.55		Uganda	
13			0.82				0.70		0.94		0.36		0.34		Estonia	
14	0.76				0.78		0.53				0.65		0.92		Kosovo	
15			0.82		0.92						0.24		0.72		Spain	
16							0.72		0.89		0.23		0.79		Costa Rica	
17	0.70		0.51		0.66				0.80		0.45		0.89		Honduras	
18	0.67		0.63		0.92		0.35		0.55		0.57		0.90		Mongolia	
19			0.59		0.73		0.44		0.82				0.88		Ghana	
20	0.93		0.88								0.59		0.39		Netherlands	
21	0.79		0.86		0.85				0.73				0.60		Taiwan Province	
22	0.27		0.67		0.64		0.12		0.62		0.05				Bolivia	
23	0.48						0.35		0.41		0.31		0.05		Mauritania	
24	0.37		0.53		0.42		0.36		0.89		0.41		0.80		India	
25	0.22		0.30		0.33		0.13		0.48				0.91		Sierra Leone	
26	0.41				0.32		0.51		0.80		0.33				Bangladesh	
27	0.43		0.65		0.76		0.61		0.70		0.29				Sri Lanka	
28	0.62						0.71		0.71		0.17		0.95		Peru	
29	0.28		0.29		0.30				0.64				0.77		Malawi	
30									0.16						Lebanon	
31			0.61		0.61		0.80		0.34						Turkey	
32	0.53		0.48		0.62		0.13		0.89		0.42		0.98		Nigeria	
33	0.52		0.88		0.59		0.45		0.46		0.32				Iraq	
34	0.50		0.43		0.45		0.33		0.67		0.34		0.89		Senegal	
35	0.51		0.96		0.33				0.70		0.03				Mexico	
36	0.55				0.88		0.52		0.70		0		0.64		Georgia	
37			0.63		0.60		0.46		0.61						Egypt	
38	0.31				0.49		0.31				0.56		0.52		Tanzania	
39	0.44		0.48		0.54		0.27						0.90		Kenya	
40	0.63		0.66				0.69		0.73		0.19		0.88		Colombia	
41	0.53						0.26						0.82		Gabon	

(d)

Figura 9: Screen plot



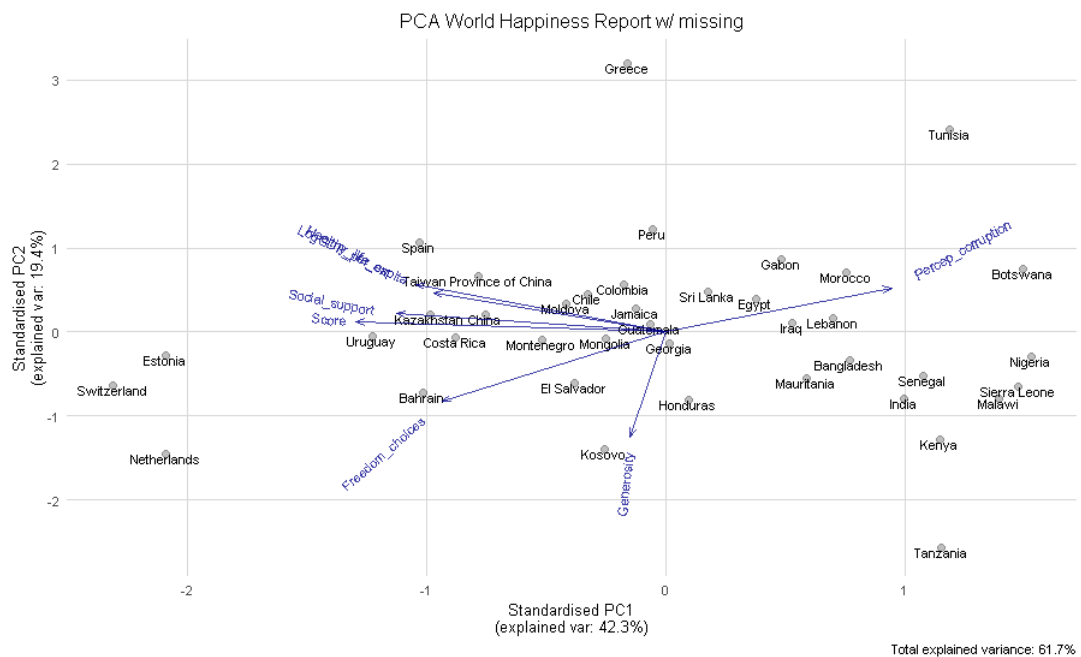
A primeira componente principal explica 42% da variância total compreendida pelos dados, a segunda explica 19%, a terceira 12.5%, a quarta 8%, quinta 7%, sexta 6.7% e a sétima 3%.

Figura 10: Rotation matrix

```
> cprecor$rotation
      PC1      PC2      PC3      PC4      PC5      PC6      PC7
Score -0.49809165 0.07107134 -0.3424729 0.05433883 -0.20905656 0.05784689 0.76128989
LogGDP_per_capita -0.37264733 0.26519433 0.5027234 0.02157632 0.62847265 0.36391725 0.10097507
Social_support -0.43268959 0.12773426 -0.2651630 -0.73038153 -0.09688138 0.15183925 -0.40031728
Healthy_life_exp -0.40171302 0.31820852 -0.3295403 0.52741523 0.16965912 -0.40089752 -0.40137681
Freedom_choices -0.35955562 -0.46472887 0.1647672 0.38533675 -0.39702839 0.48648074 -0.29123726
Generosity -0.05872011 -0.70814023 -0.3751163 -0.05705072 0.58982102 -0.04908288 0.02871265
Percep_corruption 0.36292803 0.29935368 -0.5321775 0.18225959 0.13276629 0.66435785 -0.05692910
> |
```

A CP1 permaneceu com a interpretação equivalente, com intensidades ligeiramente. Já a CP2, a interpretação foi equivalente com mudança de sinal e intensidades ligeiramente diferentes.

Figura 11: Biplot w/ labels



A conclusão é que o processo de imputação de dados faltantes foi bem sucedido, a análise ficou equivalentemente oposta com respeito à CP2.