

## **Dinâmica do Desenvolvimento Africano no ano de 2018**

Autor: Lúcia Moreira

Data: Junho 2019

### **Resumo**

O presente trabalho analisou a economia e o desenvolvimento de vários países africanos referente ao ano de 2018. A análise baseou-se no estudo das Componentes Principais que explica a variância de variáveis relacionadas com indicadores de crescimento económico, financeiros, da força laboral, dados populacionais, perceção de corrupção e ainda indicadores de infância perdida (que agregam educação, saúde, proteção e trabalho infantil). A análise em Componentes Principais permitiu a redução do número de variáveis de 19 para uma interpretação em 5 dimensões que explicam 80 % da variância. As Componentes Principais foram então utilizadas em dois algoritmos de partição entre os vários países africanos. Obteve-se uma partição em dois *clusters* que dividem os países mais desenvolvidos dos menos desenvolvidos.

### **Dataset**

O presente conjunto de dados foi obtido em 3 fontes diferentes, sendo maior parte das variáveis provenientes do site da OCDE [1-3]. É relativo ao ano de 2018, e contempla dados sócio-demográficos e económicos de todos os países africanos. Da OCDE [1] obteve-se um conjunto de 26 variáveis que inclui: a) indicadores de crescimento, de emprego e de desigualdade (tais como: Produto Interno Bruto (PIB) nominal (em *purchasing-power-parity(PPP) dollars*), PIB (em *U.S. dollars*) e PIB (em *PPP dollars per capita*), b) indicadores sobre a repartição da economia (emprego na agricultura (% do total), emprego na indústria (% do total), emprego na manufatura (% do total) e emprego nos serviços (% do total)), c) dados sobre as finanças públicas (receitas gerais do governo (% do PIB), receitas com impostos (% do PIB), receitas com *grants* (% do PIB), receitas com *commodities* (% do PIB), despesas governamentais (% do PIB) e *deficit* governamental (% do PIB)), d) características da força laboral (emprego (% da população 15+), emprego feminino (% da população 15+), emprego masculino (% da população 15+), emprego precário (% da população empregada), e) dados populacionais (população total (em milhares), população urbana (em milhares), população rural (em milhares), dependência na velhice (razão  $65+/(15-64)*100$ ) e dependência infantil (razão  $<15/(15-64)*100$ )) e f) indicadores básicos de saúde (esperança de vida média, esperança de vida média feminina, esperança de vida média masculina e taxa de fertilidade (nados-vivos por mulher)).

Estas variáveis foram complementadas com dados sobre os recentes índices de corrupção por país (índices mais altos, numa escala de 0 a 100, estão relacionados com menor perceção de corrupção no país, *Corruption Perceptions Index 2018: African Union* [2]). E ainda com o indicador recente de “infância perdida” que agrega dados de educação, saúde, proteção e trabalho infantil (“*2019 End of Childhood Index* [3] *compares the latest data for 176 countries and assesses where the most and fewest children are missing out on childhood*”). Índices mais altos (numa escala de 0 a 1000, aqui transformados numa escala de 0 a 10) estão relacionados com um estatuto bastante elevado na saúde, educação e proteção infantil, isto é, poucas

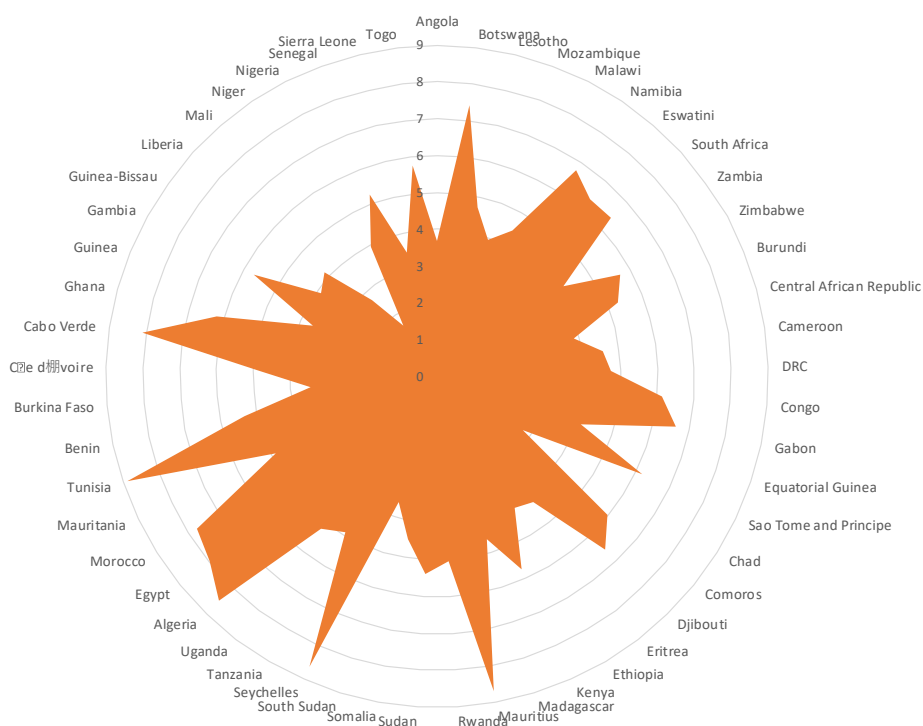
crianças com uma infância perdida [3]. O número total de variáveis quantitativas disponíveis foram então 28 variáveis.

O presente trabalho verteu-se sobre a análise destes dados numa perspetiva de Análise de Componentes Principais tendo em vista uma partição (*clustering*).

### Pré-processing

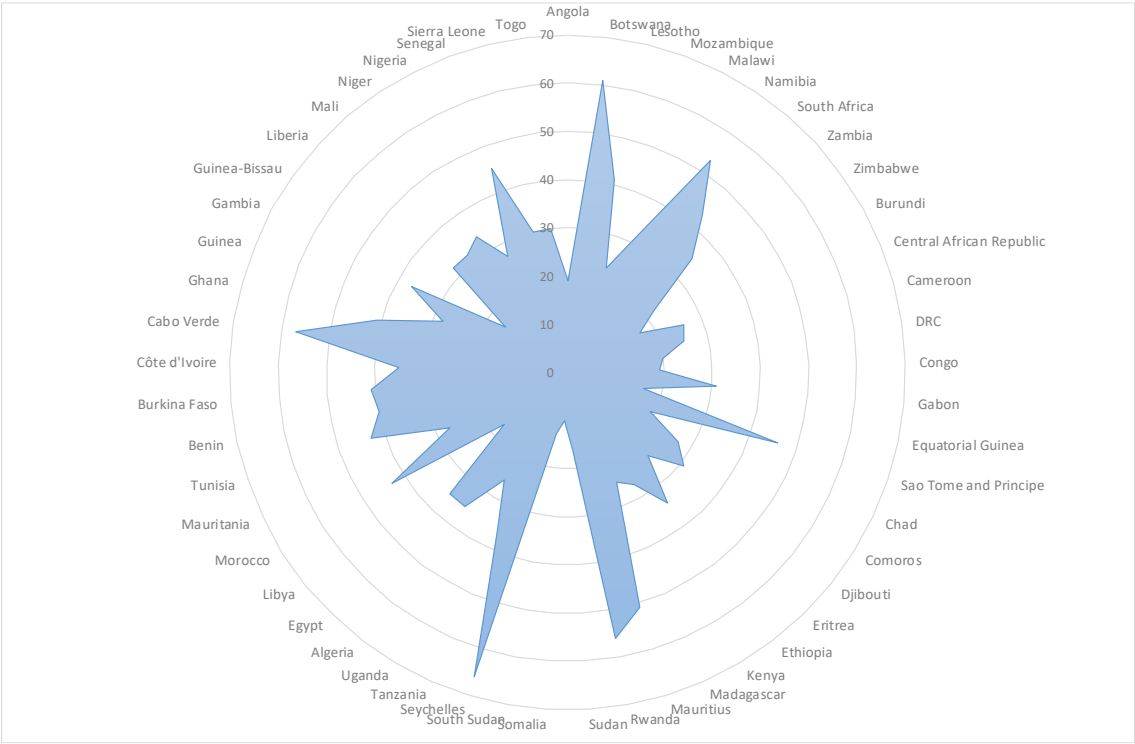
Algumas das variáveis são redundantes para a análise proposta. Deste modo, as variáveis Produto Interno Bruto (PIB) nominal (em *purchasing-power-parity(PPP) dollars*) e PIB (em *PPP dollars per capita*) foram removidas. Em relação aos dados de emprego as 4 variáveis disponíveis foram transformadas em apenas duas: emprego no sector primário e emprego no sector não-primário (que é a soma das percentagens de emprego na indústria, manufatura e serviços). Foram ainda removidos os dados sobre o emprego feminino e masculino, população total e esperança de vida média masculina e feminina. As variáveis PIB, população rural, população urbana foram ainda normalizadas pela população total. Ficaram então um total de 19 variáveis para analisar.

A Figura 1 mostra a distribuição dos índices de “infância perdida” por país. Países com a melhores condições para a infância são Cabo Verde, Tunísia, Marrocos, Egito, Seychelles, Ilhas Maurícias e Botswana (índice superior a 7). Somália, Chad e Níger mostram os piores indicadores para este índice.

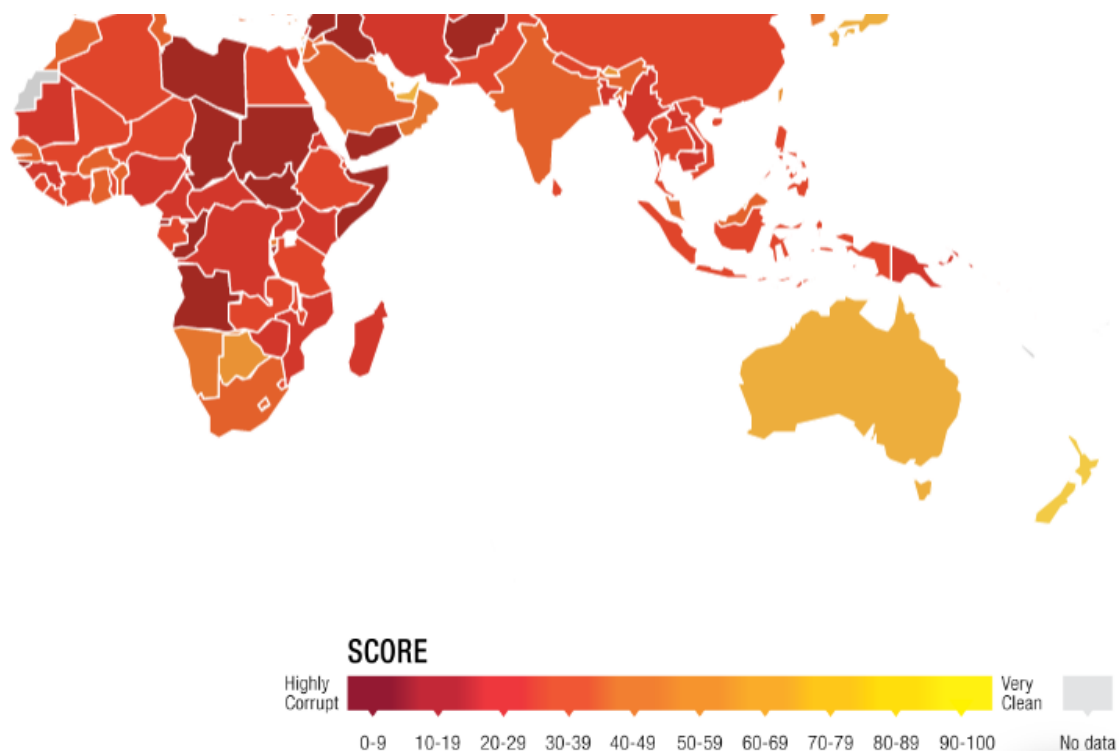


**Figura 1** - Distribuição dos índices de “infância perdida” por país [3].

As figuras 2 e 3 mostram o índice de percepção da corrupção em África. Os países com menor percepção de corrupção são as ilhas Seychelles e Botswana (índice maior ou igual a 60), mesmo assim com índices bastante baixos comparados com outros países mundiais com um ranking próximo de 100. Os países mais corruptos são Somália e Sudão Sul com índices inferiores a 15. A Figura 3 mostra os índices de percepção da corrupção no mapa geográfico. Na generalidade, os países africanos têm uma percepção elevada de corrupção.

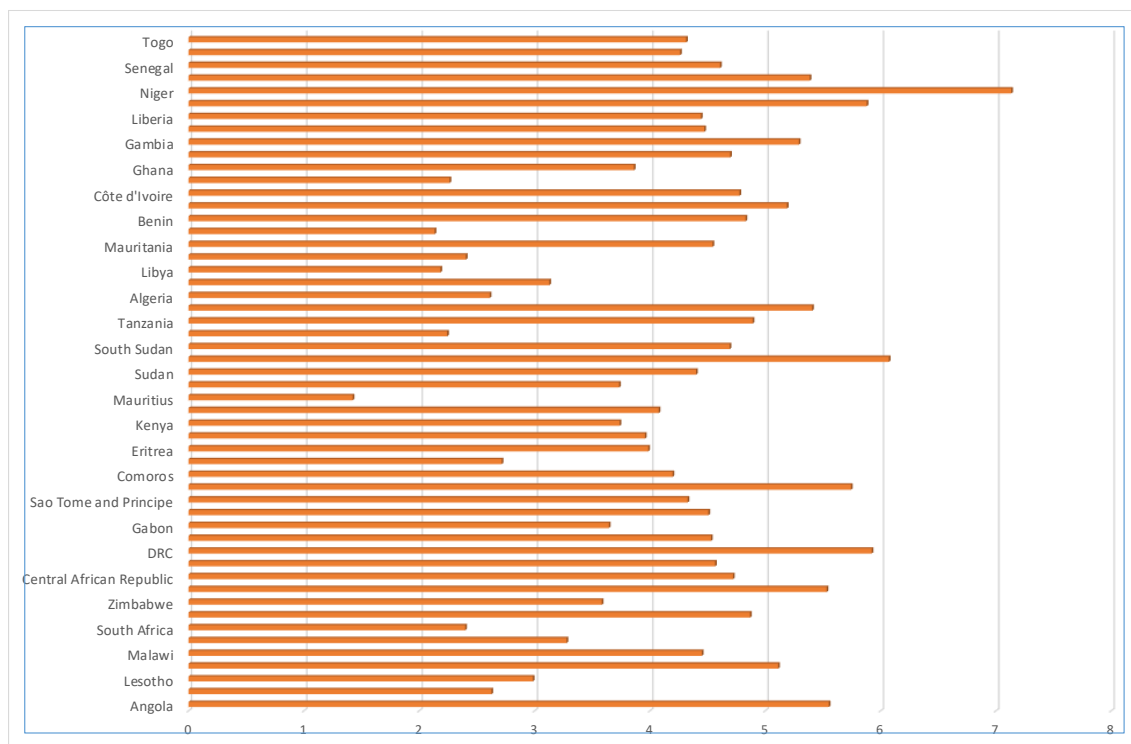


**Figura 2** - Índice de percepção da corrupção em África [2].



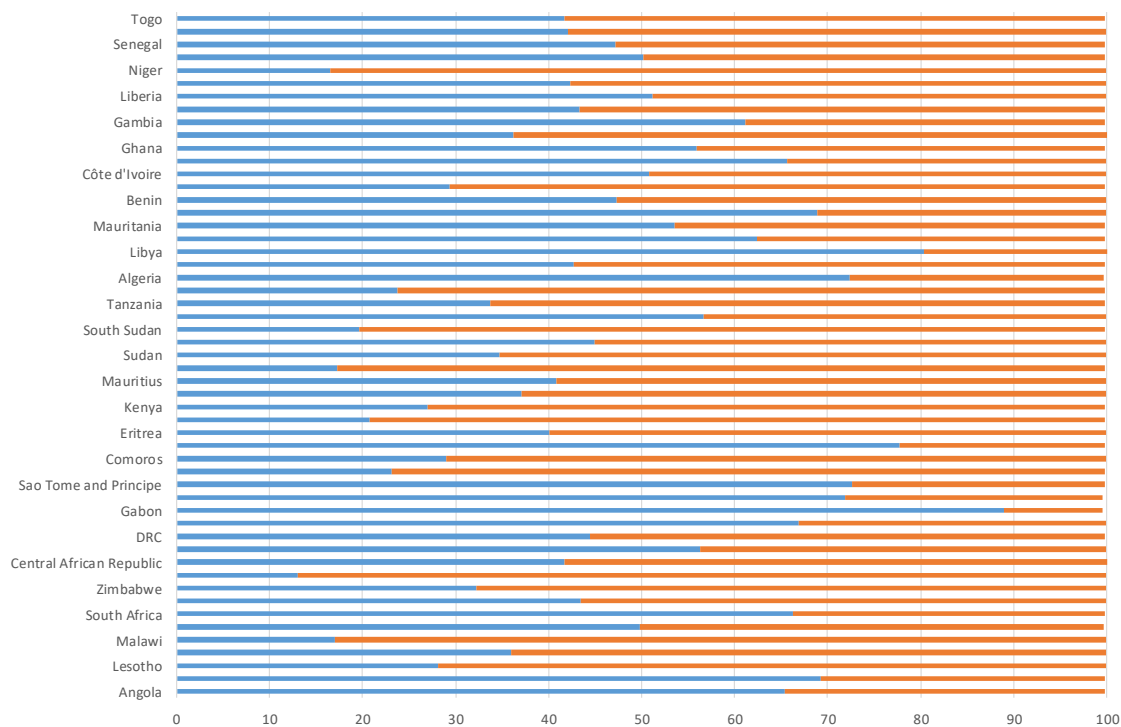
**Figura 3** - Índice de percepção da corrupção no mapa geográfico africano, adaptado de <https://www.transparency.org/cpi2018>

A Figura 4 mostra o índice de fecundidade por país. Vários países apresentam taxas superiores a 5 filhos por mulher, sendo o mais alto (mais de 7) no Níger e o mais baixo (1,4) nas Ilhas Maurícias.



**Figura 4** - Taxa de fertilidade (nados-vivos por mulher) em 2018.

A Figura 5 mostra a distribuição da população rural (azul) e urbana (laranja). A maioria dos países africanos tem ainda bastante população rural. O país com maior população rural é o Gabão (89 %) e o país com a maior população urbana é o Burundi (87 %).



**Figura 5** - Distribuição da população rural (azul) e população urbana (laranja) por país.

A Tabela 1 mostra a matriz de correlações das 19 variáveis utilizadas e a sua codificação, uma vez que o nome das variáveis é bastante comprido. A matriz das correlações indica que as variáveis estão todas substancialmente correlacionadas entre si, justificando assim uma análise ACP.

**Tabela 1** - Correlações entre as variáveis.

		B	D	E	H	I	J	K	L	M	N	Q	S	T	U	V	W	Z	child	corrupt
PIB per capita	B	1																		
Emprego primário (%)	D	-0,4	1																	
Emprego não-primário(%)	E	0,2	-0,9	1																
Receitas (% do PIB)	H	0,2	-0,2	0,3	1															
Impostos (% do PIB)	I	0,3	-0,4	0,4	0,4	1														
Grants (% do PIB)	J	-0,3	-0,2	0,1	0,1	-0,2	1													
Commodities (% do PIB)	K	0,1	-0,1	0,2	0,7	0,1	-0,3	1												
Despesa pública (% do PIB)	L	0,2	-0,3	0,3	0,7	0,5	0,2	0,2	1											
Déficit (% do PIB)	M	-0,1	0,2	-0,2	-0,1	0,0	-0,1	-0,1	-0,1	1										
Emprego (% população 15+)	N	0,0	-0,3	0,3	-0,1	0,1	0,0	0,1	0,0	-0,2	1									
Emprego precário (%)	Q	-0,7	0,3	-0,2	-0,4	-0,2	0,2	-0,2	-0,4	-0,3	0,2	1								
População Urbana (%)	S	0,7	-0,6	0,4	0,1	0,5	0,0	0,0	0,2	0,1	0,1	-0,6	1							
População rural (%)	T	-0,7	0,6	-0,4	-0,1	-0,5	0,0	0,0	-0,2	-0,1	-0,1	0,6	-1,0	1						
razão 65+/(15-64)	U	0,5	-0,2	0,3	0,3	0,4	-0,3	0,3	0,4	0,1	-0,4	-0,6	0,4	-0,4	1					
razão <15/(15-64)	V	-0,6	0,4	-0,3	-0,3	-0,4	0,3	-0,1	-0,4	0,0	0,2	0,7	-0,6	0,6	-0,8	1				
Esperança de vida (anos)	W	0,5	-0,3	0,4	0,3	0,5	0,0	0,0	0,5	0,2	-0,3	-0,7	0,5	-0,5	0,7	-0,7	1			
Fertilidade	Z	-0,6	0,3	-0,2	-0,3	-0,3	0,2	0,0	-0,4	-0,2	0,3	0,7	-0,6	0,6	-0,7	0,9	-0,7	1		
Childhood missed	child	0,6	-0,5	0,5	0,3	0,5	-0,1	0,1	0,4	0,2	-0,2	-0,8	0,8	-0,8	0,7	-0,9	0,8	-0,9	1	
Corruption index	corrupt	0,4	-0,5	0,3	0,2	0,4	0,2	-0,3	0,5	-0,2	0,1	-0,4	0,4	-0,4	0,1	-0,4	0,5	-0,4	0,5	1

## Análise em Componentes Principais

A ACP foi feita considerando todas as variáveis quantitativas e normalizadas (ACP normada). Os muito poucos valores em falta foram interpolados com a função *imputePCA()* do package *missMDA* (ver o script em R em Apêndice).

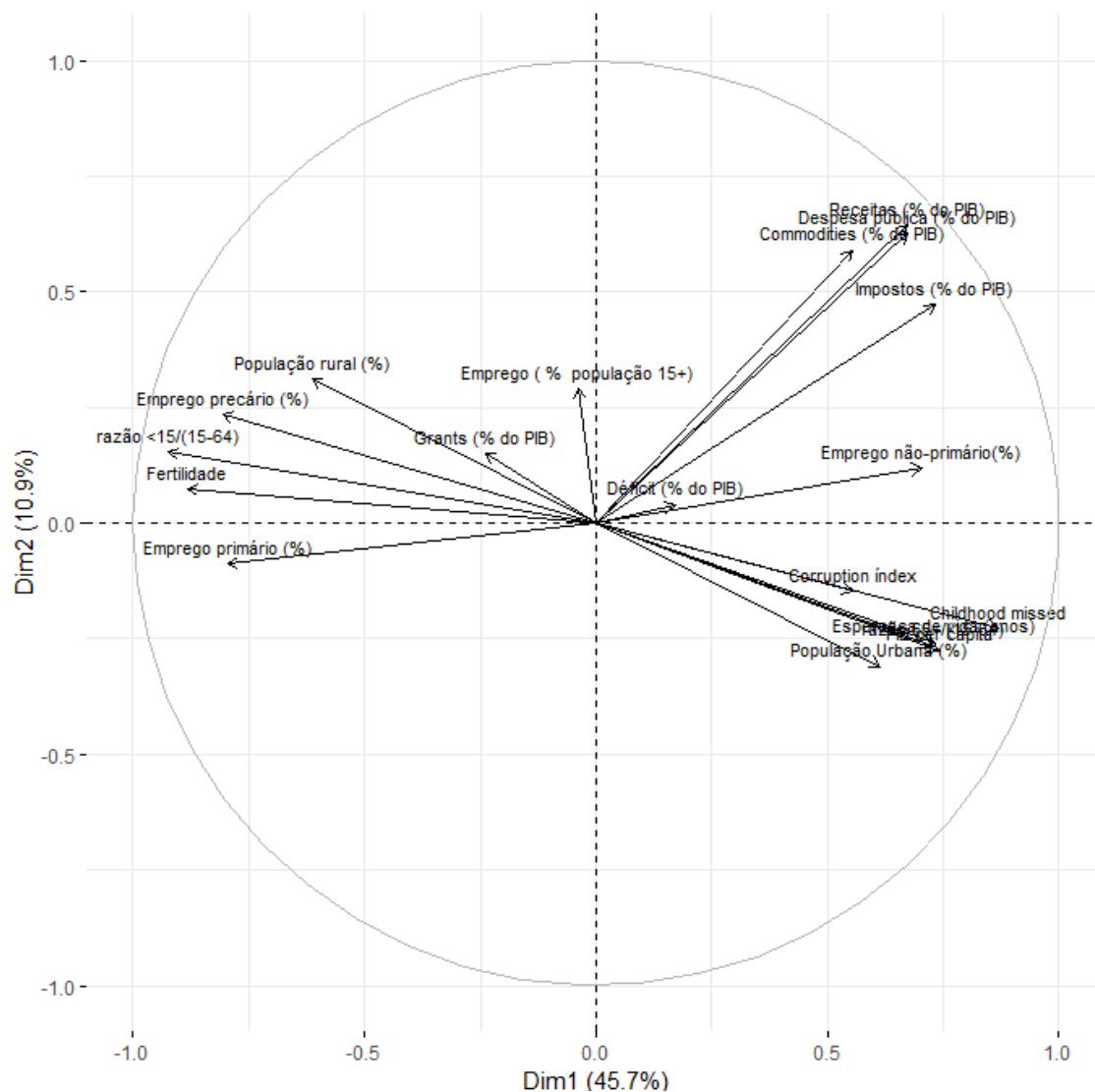
A Figura 6 mostra a “*print*” dos valores próprios e contribuições de cada componente principal obtido após ACP.

	eigenvalue	percentage of variance	cumulative percentage of variance
comp 1	8.682124e+00	4.569539e+01	45.69539
comp 2	2.071841e+00	1.090443e+01	56.59982
comp 3	1.638507e+00	8.623721e+00	65.22354
comp 4	1.528240e+00	8.043367e+00	73.26691
comp 5	1.139803e+00	5.998963e+00	79.26587
comp 6	9.468541e-01	4.983443e+00	84.24931
comp 7	7.084865e-01	3.728876e+00	87.97819
comp 8	6.128811e-01	3.225690e+00	91.20388
comp 9	4.217004e-01	2.219476e+00	93.42335
comp 10	3.305518e-01	1.739746e+00	95.16310
comp 11	2.865549e-01	1.508183e+00	96.67128
comp 12	1.958402e-01	1.030738e+00	97.70202
comp 13	1.484504e-01	7.813179e-01	98.48334
comp 14	1.247882e-01	6.567799e-01	99.14012
comp 15	7.777137e-02	4.093230e-01	99.54944
comp 16	4.141952e-02	2.179975e-01	99.76744
comp 17	2.873305e-02	1.512266e-01	99.91867
comp 18	1.543888e-02	8.125726e-02	99.99992
comp 19	1.469423e-05	7.733805e-05	100.00000

Figura 6 – Print do output obtido após ACP efetuada.

Pode-se concluir que a componente 1 explica 45,7 % da variância e a componente 2 explica 10,9 %. As primeiras 5 componentes explicam praticamente 80 % da variância.

A Figura 7 mostra graficamente a análise ACP para as 19 variáveis consideradas. Pode-se observar que a 1ª componente, que explica quase 50 % da variância, opõe os países com melhores indicadores financeiros (receitas e despesas públicas), com uma grande percentagem de emprego nos sectores secundários e terciários, uma esperança de vida média elevada, com baixos índices de corrupção, com uma elevada taxa de população urbana, altas taxas de escolaridade, elevadas taxas de dependência na velhice e PIB per capita elevado a países com elevadas taxas de emprego precário, elevadas taxas de fertilidade, elevada taxa de emprego no sector primário, elevadas taxas de dependência infantil e elevada população rural. Ou seja, a primeira componente separa os países mais desenvolvidos dos menos desenvolvidos em África. A segunda componente explica ca. de 11 % da variância e está mais relacionada com o sistema financeiro.



**Figura 7** - Representação gráfica das variáveis nos dois primeiros eixos.

A Figura 8 mostra o *print* do output das contribuições e correlações de cada variável nas primeiras 5 componentes principais. Pode-se observar que na primeira componente as contribuições estão relativamente distribuídas (entre 9 % e 5 %) para 12 variáveis, sendo as mais contribuições mais significativas a taxa de dependência infantil (ca. 10%), a taxa de fertilidade (ca. 9%) e o índice de infância perdida (ca. 9%). A segunda componente explica os indicadores financeiros (os vários tipos de receitas e despesas governamentais) com contribuições entre 10 % e 20 %. A terceira componente explica as receitas com *grants* e a taxa e tipo de emprego (entre 10 % a 20 %). A quarta componente explica a distribuição da população e o índice de corrupção (ca. 23 % cada um) enquanto que a quinta dimensão explica maioritariamente o deficit (52 %) mas ainda a taxa de emprego e esperança de vida (10 % cada um).

\$contrib	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
PIB per capita	6.33369063	3.6465612	1.46441312	0.034756311	7.71064753
Emprego primário (%)	7.30361156	0.3748290	10.14542179	0.005748354	1.12598982
Emprego não-primário(%)	5.70702086	0.6914002	11.98369653	0.050126298	1.43144220
Receitas (% do PIB)	5.23302586	20.0788594	0.09068628	0.020569701	0.65189615
Impostos (% do PIB)	6.18953436	10.7533274	0.05377428	0.226063429	1.25480301
Grants (% do PIB)	0.65965126	1.0994796	16.49841361	3.657032898	0.43504712
Commodities (% do PIB)	3.52901788	16.8151176	5.11787561	4.615390835	8.18815179
Despesa pública (% do PIB)	5.20519212	18.9160068	0.29016249	0.014541608	1.49236844
Déficit (% do PIB)	0.33886043	0.0752460	5.24782006	9.957376947	51.70735617
Emprego ( % população 15+)	0.01714833	4.0449035	19.74104164	0.150874526	10.26709515
Emprego precário (%)	7.48679626	2.6535592	2.95922238	0.041819369	1.61116325
População Urbana (%)	4.33540051	4.6346955	7.97492296	23.167247816	0.05777888
População rural (%)	4.33710651	4.6748478	7.95671284	23.138778056	0.06933287
razão 65+/(15-64)	6.10021713	3.3922935	1.31973145	5.532172069	0.10794293
razão <15/(15-64)	9.81373509	1.1283799	1.14029858	0.321930164	0.02650318
Esperança de vida (anos)	6.19548405	3.2018727	0.37168705	3.314938511	10.73789750
Fertilidade	8.95761475	0.2655589	2.11788542	1.770893730	0.31344164
Childhood missed	8.69592979	2.5178663	0.17423797	0.676031912	2.67416371
Corruption index	3.56096262	1.0351956	5.35199595	23.303707466	0.13697868

\$cor	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
PIB per capita	0.74155168	-0.27486532	-0.15490161	-0.02304690	-0.29645605
Emprego primário (%)	-0.79630939	-0.08812412	-0.40771735	0.00937276	0.11328753
Emprego não-primário(%)	0.70391097	0.11968588	0.44311815	0.02767761	-0.12773261
Receitas (% do PIB)	0.67404586	0.64498218	-0.03854739	-0.01773004	-0.08619937
Impostos (% do PIB)	0.73306417	0.47200830	-0.02968325	0.05877747	0.11959215
Grants (% do PIB)	-0.23931515	0.15092868	0.51993042	0.23640692	0.07041790
Commodities (% do PIB)	0.55352843	0.59023934	-0.28958029	-0.26558283	-0.30549762
Despesa pública (% do PIB)	0.67225089	0.62602681	-0.06895167	-0.01490740	0.13042262
Déficit (% do PIB)	0.17152342	0.03948389	-0.29323351	-0.39009306	0.76769915
Emprego ( % população 15+)	-0.03858548	0.28948914	0.56873397	0.04801796	0.34208867
Emprego precário (%)	-0.80623382	0.23447287	0.22019778	0.02528043	0.13551416
População Urbana (%)	0.61351843	-0.30987661	0.36148259	-0.59502193	-0.02566253
População rural (%)	-0.61363913	0.31121601	-0.36106965	0.59465621	0.02811153
razão 65+/(15-64)	0.72775576	-0.26510927	-0.14705064	0.29076598	0.03507616
razão <15/(15-64)	-0.92306050	0.15289943	0.13668896	-0.07014175	0.01738057
Esperança de vida (anos)	0.73341641	-0.25756108	-0.07803921	0.22507823	0.34984407
Fertilidade	-0.88187939	0.07417518	0.18628392	-0.16450989	-0.05977137
Childhood missed	0.86890244	-0.22839917	-0.05343127	0.10164344	0.17458579
Corruption index	0.55602806	-0.14645001	0.29612974	0.59677176	0.03951312

**Figura 8** - Contribuições (acima) e correlações (abaixo) da análise ACP normada.

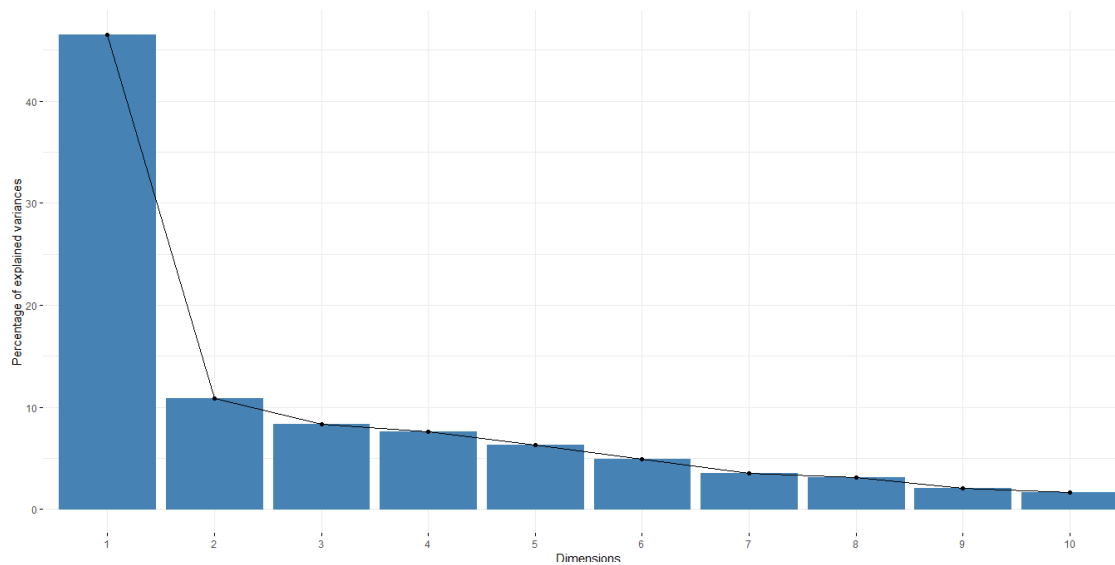
Em relação às correlações, e como já mencionado, pode-se observar que a primeira componente se correlaciona bastante com a taxa de dependência infantil seguida da taxa de fertilidade e do índice de “infância perdida”, esta última numa correlação contrária em relação às duas primeiras variáveis. Ou seja, as crianças “perdem” mais a sua infância quando são em número elevado nas famílias, provavelmente relacionado com o custo elevado de uma família em criar e educar uma criança devidamente. Daí a elevada importância de informar as famílias sobre os diferentes métodos disponíveis de controlo de natalidade. A quinta dimensão diferencia os países pela taxa de deficit.

### Análises de partição

As primeiras 5 componentes foram escolhidas como as novas variáveis em dimensão reduzida para fazer uma análise de *clustering*. Obteve-se uma redução de 19 variáveis para 5, portanto uma redução de quase 75 %. Os critérios usados foram valores próprios superiores (critério de



Kaiser) à unidade e uma percentagem de inércia explicada pelo menos de 80 % (critério de Pearson). O critério de Catrell não se mostrou muito relevante neste caso (Figura 9).



**Figura 9** – Critério de Catrell para as primeiras 10 componentes principais.

As 5 novas variáveis foram calculadas com os primeiros 5 vetores próprios obtidos e as variáveis devidamente normadas (cada variável foi centrada e dividida pelo respetivo desvio padrão). Com estas novas variáveis foi realizada uma partição utilizando *k-means* e *mclust* (“Gaussian finite mixture model fitted by EM algorithm”), - ver o script em apêndice.

O algoritmo *mclust* dividiu os países em dois *clusters* (Figura 10): os países com cores mais escuras são os países mais bem desenvolvidos enquanto que a cor mais clara indica os países menos desenvolvidos. Escolheu-se a componente número 5 para a dimensão vertical do eixo devido à sua elevada correlação com o deficit, assim separando os países por valor do deficit (*size factor*) e ajudando numa melhor visualização. A quinta componente também foi escolhida porque neste caso a sua contribuição não é muito diferente da contribuição da segunda componente. Observa-se que o algoritmo *mclust* não apresenta uma partição com bordas bem definidas dos dois grupos.

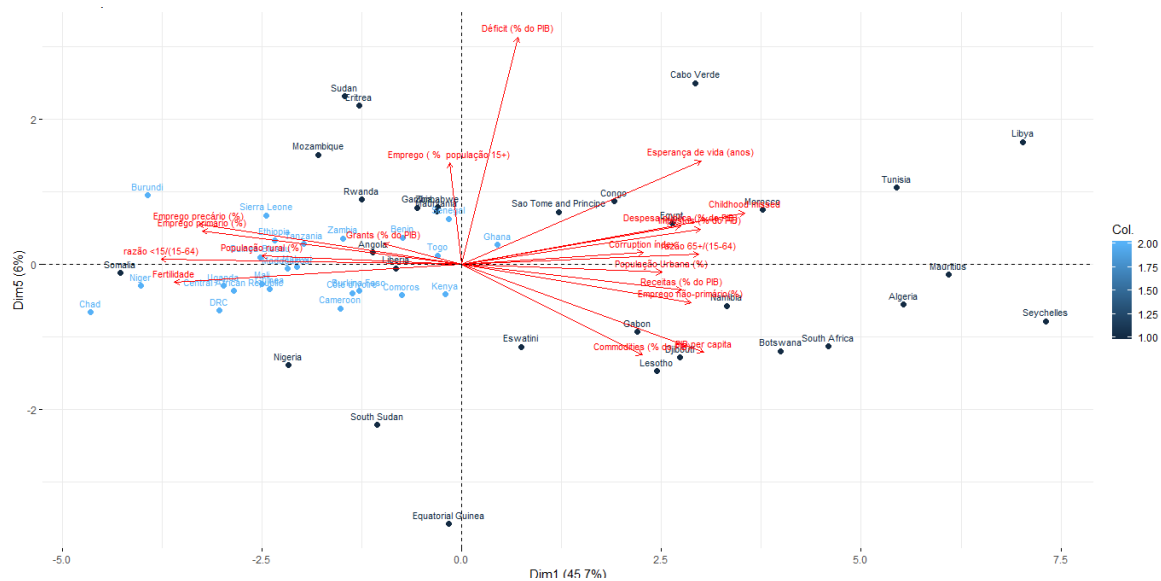


Figura 10 – Partição com *mclust*.

Utilizando agora o algoritmo *k-means* e 2 centros (como descoberto pelo *mclust*) obteve-se a separação abaixo (cores mais escuras são os países mais desenvolvidos) (Figura 11). Este algoritmo fez uma melhor partição que o *mclust*. Isto pode ocorrer porque as novas variáveis em componentes principais não seguem uma distribuição estritamente normal (Figura 12). Pode-se observar que o algoritmo separa bem os países do norte de África (países islâmicos) e os países em torno da África do Sul, mais desenvolvida, assim como as ilhas Seychelles e as ilhas Maurícias dos restantes países africanos.

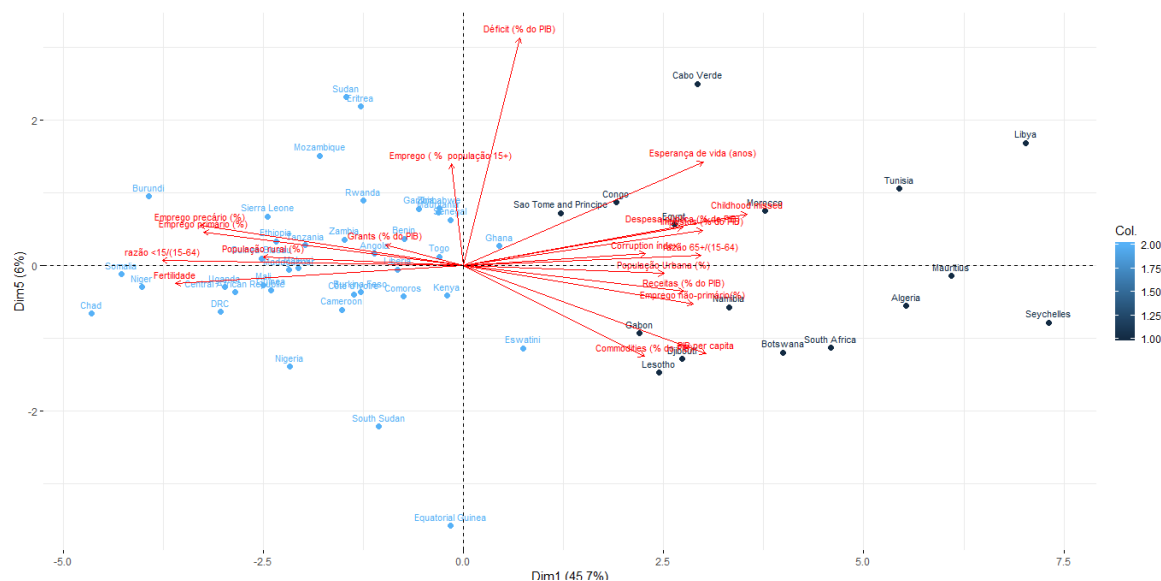
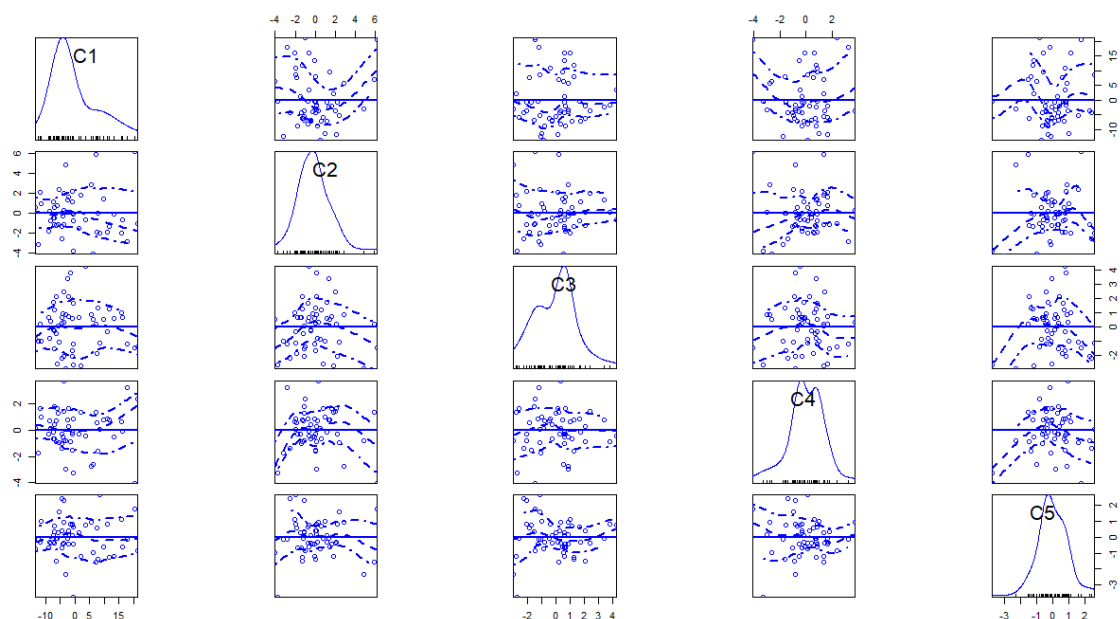


Figura 11 – Partição com *k-means*.



**Figura 12** - Distribuição das variáveis dos componentes principais.

## Conclusões

Neste trabalho analisou-se o desenvolvimento de vários países africanos referente ao ano de 2018. A análise das Componentes Principais foi utilizada para explicar a variância de vários indicadores quantitativos relacionados com crescimento económico, finanças públicas, força laboral, dados populacionais, perceção de corrupção e infância perdida. A Análise em Componentes Principais permitiu uma interpretação em 5 dimensões que explicam 80 % da variância. As 5 componentes principais obtidas foram utilizadas em dois algoritmos de partição entre os vários países africanos. O algoritmo *k-means* separa bem em duas partições: os países do norte de África (países islâmicos) e os países em torno da África do Sul, mais desenvolvida, assim como as ilhas Seychelles e as ilhas Maurícias dos restantes países africanos. Ou seja, faz uma partição entre países mais desenvolvidos e em desenvolvimento.

## Referências

- [1] - Africa's Development Dynamics - Statistical Annex, OCDE,  
[https://stats.oecd.org/Index.aspx?DataSetCode=AEO11\\_OVERVIEW\\_CHAPTER1\\_FIG1\\_PT](https://stats.oecd.org/Index.aspx?DataSetCode=AEO11_OVERVIEW_CHAPTER1_FIG1_PT),  
 acedido em Maio 2019.
- [2] - Corruption Perceptions Index 2018: African Union  
[https://www.transparency.org/files/content/pages/2018\\_CPI\\_Executive\\_Summary.pdf](https://www.transparency.org/files/content/pages/2018_CPI_Executive_Summary.pdf),  
 acedido em Maio 2019.
- [3] - 2019 End of Childhood Index  
<https://reliefweb.int/sites/reliefweb.int/files/resources/global-childhood-report-2019-pdf%20%281%29.pdf>,  
 acedido em Maio 2019.

## Apêndice- Script em R

```
library("FactoMineR")
library(reader)
library("factoextra")
library(missMDA)

Data_Africa <- read.delim2("completo3.csv", sep=";", header = FALSE, dec=".",
stringsAsFactors=FALSE)
row.names(Data_Africa) <- Data_Africa[,1]

#add childhood missed index as numerical
child <- read.delim2("child3.csv", header = FALSE, row.names=1, sep=";", dec="." )

childMiss <- rep("", nrow(Data_Africa))

for(i in 1:nrow(Data_Africa)){
  for(j in 1:nrow(child)){
    if (row.names(Data_Africa)[i] == row.names(child)[j]) childMiss[i]=child[,1][j]
  }
}
Data_Africa$child <- as.numeric(childMiss)

#add corruption index as quantitative

corrup <- read.delim2("corrup_2.csv", header = FALSE, row.names=1, sep=";", dec="." )

corrupIn <- rep("", nrow(Data_Africa))
for(i in 1:nrow(Data_Africa)){
  for(j in 1:nrow(corrup)){
    if (row.names(Data_Africa)[i] == row.names(corrup)[j]) corrupIn[i]=corrup[,1][j]
  }
}
Data_Africa$corrup <- as.numeric(corrupIn)

colnames(Data_Africa) <- c("país", "A", "B", "C", "D", "E", "F", "G", "H", "I", "J", "K", "L",
"M", "N", "O", "P", "Q", "R", "S", "T", "U", "V", "W", "X", "Y", "Z", "child", "corrupt")

#normaliza GDP pela população
Data_Africa$B <-Data_Africa$B/Data_Africa$R

#calcula percentagem de população rural e urbana
Data_Africa$S <-Data_Africa$S/Data_Africa$R*100
Data_Africa$T <-Data_Africa$T/Data_Africa$R*100

Data_Africa$E <- Data_Africa$E + Data_Africa$F + Data_Africa$G #junta emprego dos sectores
secundario /terciario
#remove statistics from males and females and only consider both
Data_Africa <- Data_Africa[, -c(2, 4, 7, 8, 16,17,19, 25,26)]#remove variaveis redundantes (A,
C, F, G, O, P, R, X, Y)
colnames(Data_Africa) <- c("países", "PIB per capita", "Emprego primário (%)", "Emprego não-
primário(%)", "Receitas (% do PIB)", "Impostos (% do PIB)", "Grants (% do PIB)", "Commodities (%
do PIB)", "Despesa pública (% do PIB)", "Déficit (% do PIB)", "Emprego ( % população 15+)",
"Emprego precário (%)", "População Urbana (%)", "População rural (%)", "razão 65+/(15-64)",
"razão <15/(15-64)", "Esperança de vida (anos)", "Fertilidade", "Childhood missed", "Corruption
index")

#write.csv (Data_Africa, "dados1.csv")

corr1 <- cor(Data_Africa[,-1], use = "na.or.complete")
write.csv(corr1, "correl.csv" )
PCA_Af <-imputePCA(Data_Africa[,-1], ncp=8)
PCA_Af_2 <-PCA(PCA_Af$completeObs, graph = TRUE, axes = c(1,2)) #faz PCA normada
PCA_Af_2$eig
PCA_Af_2$var

fviz_screplot(PCA_Af_2)

fviz_pca_var(PCA_Af_2, axes = c(1, 2), labelsize = 3, col.var = "black")
fviz_pca_var(PCA_Af_2, axes = c(1, 5), labelsize = 3, col.var = "black")

#clustering
#novas variaveis:
library(matrixStats)
data2 <- as.matrix(PCA_Af$completeObs)
average <- colMeans(data2)
```

```

data3 <- apply(data2,1,'-',average)
data3 <- t(data3)

var1 <- colVars(data3)
var1 <- sqrt(var1)
data4 <- apply(data3,1,'/',var1)
data4 <- t(data4)

C1 <- data4 %*% PCA_Af_2$var$coord[,1]
C2 <- data4 %*% PCA_Af_2$var$coord[,2]
C3 <- data4 %*% PCA_Af_2$var$coord[,3]
C4 <- data4 %*% PCA_Af_2$var$coord[,4]
C5 <- data4 %*% PCA_Af_2$var$coord[,5]

Data_Africa$C1 <- C1
Data_Africa$C2 <- C2
Data_Africa$C3 <- C3
Data_Africa$C4 <- C4
Data_Africa$C5 <- C5

library(mclust)
library(stats)

clust3 <- Mclust(Data_Africa[,21:25])
clust2 <- kmeans(Data_Africa[,21:25], centers = 2, iter.max = 20)

fviz_pca_biplot(PCA_Af_2, axes = c(1,3), labelsize = 3, pointsize = 2,
col.ind=clust3$classification, col.var="red")
fviz_pca_biplot(PCA_Af_2, axes = c(1,5), labelsize = 4, pointsize = 2, col.ind=clust2$cluster,
col.var="red")

library(car)
scatterplotMatrix(Data_Africa[,21:25])

```