



Licenciatura em Ciência de Dados - 2º ano

# **Contexto de políticas laborais para a reforma do mercado de trabalho com base na análise dos World Development Indicators**

Unidade Curricular de Métodos de Aprendizagem Não Supervisionada

12 de janeiro de 2024

**Discentes:** Samuel Ricardo nº 110884 / João Dias nº 110305 /  
Rafael Cerqueira nº 110860 / Felipe Licas nº 110861

## **Índice**

1. Introdução.....	2
2. Panorama Mundial dos Mercados de Trabalho.....	3
3. Objetivo de análise.....	5
4. Tratamento de dados.....	5
5. Adequabilidade de PCA.....	7
6. Caracterização dos PCs.....	8
7. Análise de PCAs.....	10
8. Identificação da heterogeneidade na base de dados.....	12
9. Conclusões obtidas com o clustering.....	13
10. Principais conclusões.....	15
11. Referências.....	16
12. Anexos.....	17

# **1. Introdução**

Os World Development Indicators (WDI), promovidos pelo World Bank, são um instrumento que procura medir os níveis de desenvolvimento económico e social em diferentes países de todo o mundo. Ao constituírem-se como uma das ferramentas mais reconhecidas a nível mundial, os WDI adotam uma abordagem abrangente para avaliar os indicadores económicos e fatores sociais e ambientais que impactam o desenvolvimento sustentável. Vão além de indicadores convencionais/tradicionais e capturam a complexidade dos fatores que moldam o crescimento económico e social a larga escala. [1]

A recolha de informações através dos WDI tem em consideração uma ampla gama de indicadores, incluindo dados económicos, sociais, laborais, ambientais e de saúde. Além disso, exploram as políticas de desenvolvimento, as práticas de governação e a eficiência das instituições em cada país participante. A análise é conduzida de maneira a compreender o desempenho recente, as tendências ao longo do tempo e as diferenças entre as nações. [3]

Os WDI desempenham um papel crucial no desenvolvimento de políticas económicas e sociais por formularem percepções essenciais para os órgãos legisladores tanto em contexto nacional como internacional. Tendo começado em 1960, os WDI têm sido realizados periodicamente desde então, levando em consideração uma avaliação contínua e um entendimento aprofundado das mudanças em diferentes componentes de desenvolvimento. [3]

Inicialmente envolvendo 28 países, o WDI expandiu a sua participação ao longo dos anos, contando agora com a adesão de 258 países/regiões. A diversidade de participantes mostra o interesse mundial na avaliação do desenvolvimento e na procura por estratégias eficazes de promoção do progresso económico e social. [2][3]

Um exemplo do uso dos WDI na área do ambiente é o estudo apresentado no Journal Heliyon que examina o impacto do financiamento agrícola na produção de alimentos na Nigéria. O período analisado é de 1981 a 2018, utilizando dados anuais provenientes dos Indicadores de Desenvolvimento Mundial (IDM) do Banco Mundial

e do Boletim Estatístico do Banco Central da Nigéria. Desse estudo foi indicado que deveria existir um financiamento agrícola mais amplo, com taxas de juros baixas, motivando os agricultores a adquirirem sementes, maquinaria e outros complementos agrícolas, que têm um impacto positivo no rendimento agrícola total, resultando em maior produção de alimentos. Consequentemente, o estudo recomenda mais fundos para o setor agrário, com condições de crédito menos restritivas, e a atribuição de mais terras aráveis para a prática da agricultura, entre outras medidas, visando impulsionar a produção de alimentos na Nigéria. [3]

Outro exemplo de estudo com base nos WDI é o estudo publicado no Global Journal of Management que fala sobre o papel da educação de qualidade no Desenvolvimento Sustentável do Paquistão. As conclusões tiradas neste estudo sugerem que investir em educação, representado pelo gasto em educação e outras métricas, está positivamente associado ao desenvolvimento sustentável do país. A relação de longo prazo destaca a importância contínua da qualidade da educação como um fator crucial para alcançar metas de desenvolvimento sustentável. [4]

Dada a ampla utilização destes indicadores para a análise de contexto político e para ambientes de tomada de decisão, como provam os exemplos anteriores, vamos concentrar os nossos esforços em através deles analisar o desempenho de 182 países, tendo em conta as variáveis que englobam e, que de alguma forma, permitem “medir” o estado dos diferentes mercados de trabalho um pouco por todo o mundo. [2]

## **2. Panorama Mundial dos Mercados de Trabalho**

O panorama mundial dos mercados de trabalho tem sofrido constantes transformações ao longo dos tempos. Desde a pandemia de COVID-19 tem-se verificado um crescimento no que respeita à adoção de modelos híbridos de trabalho, observando-se paralelamente um aumento de solicitações por parte de trabalhadores que já consideram este framework como fundamental para a sua produtividade e bem estar pessoal e profissional [5]. Devido às constantes mudanças

e consequentes expectativas que estas criam nos trabalhadores no que respeita à adoção de modelos de trabalho flexíveis, os seus empregadores tentam adaptar-se de forma a promover medidas que possam responder às novas exigências e que permitam, também, responder aos desafios que são para as suas empresas a retenção de talentos e a competição do próprio mercado. [6]

Para além desta grande mudança, o fenómeno dos nómadas digitais, fortalecido pela grande adesão ao trabalho remoto, tem expandido fronteiras no que toca ao recrutamento de trabalhadores de uma geografia ampla e diversificada. Os nómadas digitais ao contribuírem para a crescente digitalização e descentralização do trabalho trazem consigo grandes mudanças, tanto positivas como negativas, para um mercado de trabalho em constante mutação. [7]

Além dos aspectos já mencionados, importa destacar também a saturação do mercado com jovens licenciados, e assim certos setores enfrentam uma saturação de profissionais altamente qualificados, o que faz com que a concorrência por posições seja cada vez mais competitiva. Face a esta nova realidade, os jovens tentam assim distinguir-se da concorrência através das *soft skills*. Estas situações também levam à consideração de carreiras alternativas ligadas aos empreendedorismo jovem, tais como, o self employment, a mudança radical de setor e a procura de oportunidades melhores e mais flexíveis que se alinhem com os seus valores pessoais e profissionais. [8]

Devido a todas estas novas variantes presentes no mercado de trabalho, os legisladores estão em território inexplorado no que toca a matéria legislativa. Num cenário onde ainda existe pouca regulação que considera as mais recentes transformações, embora a produtividade das empresas pareça estar a beneficiar destas mudanças existem, também, preocupações no que toca à liderança dos trabalhadores e à sua percepção da cultura empresarial. Assim, é imperativo destacar a necessidade de políticas laborais e legislativas que tenham em conta as novas dinâmicas laborais, tendo como pilares a adaptabilidade e inovação que permitam o sucesso tanto dos trabalhadores, como da empresa. [9]

### **3. Objetivo de análise**

A Organização Internacional do Trabalho (OIT) é uma instituição vinculada à Organização das Nações Unidas, que incide principalmente sobre temáticas laborais, concentrando esforços em garantir o cumprimento de normas, convenções e recomendações internacionais e, nessa medida, solicitou à equipa encarregada pela análise dos World Development Indicators (WDI) uma investigação aprofundada sobre o contexto de políticas laborais com o objetivo de promover a reforma do mercado de trabalho à escala global, quer pela promoção de acordos de cooperação entre países, quer pela promoção de um quadro legislativo comum em matéria de regulação do mercado de trabalho, que como já mencionado tem sofrido inúmeras transformações. A intenção passa por compreender os indicadores relacionados com o mercado laboral nos vários países do mundo, de forma a compreender as assimetrias que se verificam entre os diferentes países.

Este estudo compila os resultados mais relevantes e os passos cruciais que foram tomados para extrair conclusões práticas. Assim, a análise detalhada dos WDI focar-se-á especificamente nos indicadores pertinentes que apresentam algum nível de relação com os mercados de trabalho globais, de forma a que seja possível destacar padrões e fatores determinantes que podem orientar normas e políticas laborais inovadoras e eficazes em diferentes contextos.

Ao concentrar-se em indicadores específicos relacionados com a força de trabalho, emprego e condições laborais, o objetivo é intensificar e aprimorar as políticas laborais um pouco por todo o mundo, estabelecendo um novo padrão para a reforma do mercado de trabalho global, apoiado pela análise abrangente dos WDI.

### **4. Tratamento de dados**

Com a análise de contexto concluída e os objetivos claramente definidos, procedemos ao tratamento do *dataset* do DataBank, que continha os World Development Indicators (WDI) referentes ao ano de 2021. A escolha deste ano

baseou-se essencialmente na procura dos dados mais recentes que estivessem disponíveis online. Depois de alguma experimentação, constatou-se que o dataset de 2022 (o mais recente disponível) apresentava muitas variáveis em falta. Por essa razão, decidimos testar com o dataset de 2021 e apesar de ainda existirem vários missings, optámos por avançar, já que o trade-off entre dados recentes e completos se encontrava dentro dos parâmetros que considerámos aceitáveis.

O dataset de 2021 utilizado continha inicialmente 392 887 linhas e 5 colunas, apresentando as variáveis associadas a cada país por linha, na coluna Series Name. Como tal, a nossa abordagem metodológica a este problema passou por transpor o dataset, logo após termos feito a seleção<sup>1</sup> de 25 variáveis a utilizar como variáveis de INPUT para a Principal Component Analysis e clustering (de onde se excluíram previamente as que apresentavam mais de 20% de missings por coluna). A justificação para a escolha de cada variável pode ver-se no dicionário de features (anexo A).

Também por linha foram excluídos 35 países que apresentavam mais de 20% de valores em falta e que não acrescentavam informação que pudesse ser utilizada como INPUT para a PCA. Além destes países, foram retiradas outras observações, de forma manual, que não eram efetivamente países. A título de exemplo, estamos a falar de observações que apareciam como “Least developed countries: UN classification” ou até “World”. Ainda na fase de pré-processamento ajustaram-se os data types das variáveis e de seguida procedeu-se ao tratamento dos outliers. A sua identificação foi feita através de boxplots, que permitiram descobrir, a título de exemplo, 13 outliers que correspondiam aos países poucos desenvolvidos da África na variável relativa ao acesso à eletricidade; 1 outlier (Japão) na dependência de população mais envelhecida; e 1 outlier no número de anos de educação da New Caledonia[10], que por se tratar de um erro de inserção foi ajustado para 9. Esta análise foi feita para todas as variáveis e pode ver-se no anexo B, e após verificar-se, que os outliers tratam-se de facto de indicadores que fazem sentido para os países e variáveis em questão, optou-se por recorrer à winsorização, que consiste em limitar os valores extremos de cada variável para reduzir o efeito dos outliers.

---

<sup>1</sup> O dicionário de features está disponível no anexo A.

Já com os outliers tratados, foi possível seguir para o tratamento dos missing values. Aqui, após analisarmos o padrão de missings, optámos por utilizar a técnica MICE (Multiple Imputation by Chained Equations), que envolve uma abordagem iterativa que considera cada variável com missings como uma equação condicional e imputa valores múltiplos para os dados em falta. O método "pmm" (Predictive Mean Matching), escolhido dentro do framework MICE é uma técnica que associa os valores em falta com os observados que estão mais próximos em termos de média preditiva, levando em conta a distribuição das variáveis. Nesta medida, a padronização prévia que fizemos nos dados através do processo de centralização e estandardização das variáveis foi essencial para garantir que tínhamos tudo na mesma escala (este ponto foi especialmente importante considerando que estamos a trabalhar com medidas de distância ou similaridade entre variáveis).

## 5. Adequabilidade de PCA

Ao considerar os resultados do teste de esfericidade de Bartlett, do índice de Kaiser-Meyer-Olkin (KMO) e da matriz de correlação, presente no anexo C, podemos justificar a adequabilidade da aplicação de PCA no contexto do seu estudo.

O teste de esfericidade de Bartlett avalia se as variáveis observadas estão suficientemente correlacionadas para ser aplicado ou não o PCA. Neste teste a hipótese nula: as variáveis não são correlacionadas na população, é rejeitada se o p-value for inferior a 0.05. O p-value obtido é 0, ou seja, H<sub>0</sub> é rejeitada e, consequentemente, é provável que existam correlações significativas entre as variáveis originais podendo ser aplicado o PCA.

O índice KMO é outra métrica essencial para avaliar a adequação dos dados para a realização deste tipo de análise. O valor de 0.77 está no intervalo 0.7-0.79, intervalo este que considera o uso da PCA razoável. Este índice mede a proporção da variância comum entre as variáveis em relação à variância total, sendo um indicador importante da coerência global dos seus dados para uma análise de componentes principais.

Ao examinar a metade diagonal inferior da matriz de correlação (Anexo C), é possível observar que a maioria dos seus valores é superior a 0.3, isso sugere a presença de relações significativas entre as variáveis. A forte correlação é um indício de que as variáveis não são completamente independentes, o que reforça a pertinência da utilização do PCA para capturar e representar a variabilidade dos dados por meio de componentes principais.

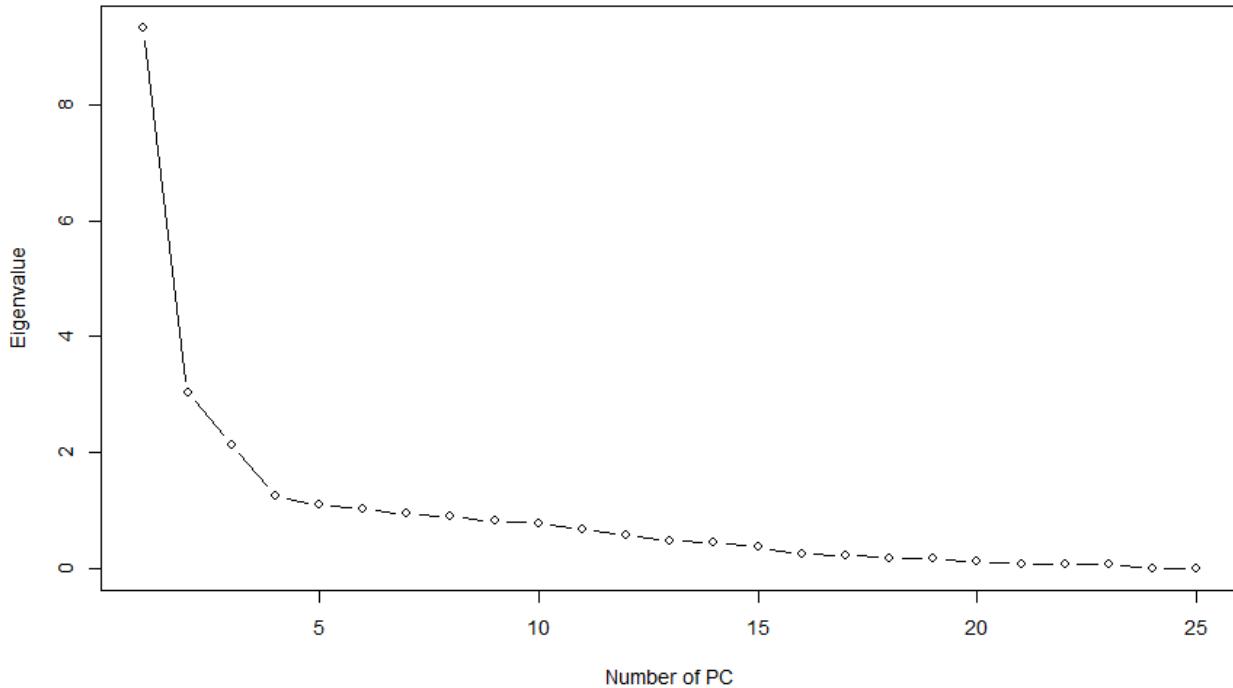
Portanto, com base nos resultados do teste de esfericidade de Bartlett, no índice KMO e na matriz de correlação, é plausível afirmar que as condições necessárias para a aplicação efetiva do PCA estão presentes.

## **6. Caracterização dos PCs**

A nossa análise foi iniciada com a aplicação dos PCs a um conjunto de 25 variáveis INPUT previamente selecionadas, assim, estas foram analisadas através dos seus valores próprios, do teste Kaiser-Meyer-Olkin (KMO) de 0.77 e do teste de Bartlett, que nos levou a rejeitar a hipótese nula, conforme descrito no ponto 5.

Depois de toda esta análise ter sido feita, decidiu-se testar para 3, 4 e 6 componentes principais tendo como base teórica o scree plot para 3 componentes, a análise da variância explicada para 4 componentes e o método de Kaiser para 6 componentes. Após algumas deliberações e testes, decidimos prosseguir, numa primeira fase, com 4 componentes baseados na variância explicada.

**Scree plot for WDI dataset**



*Gráfico 1- Scree Plot*

No entanto, surgiram dúvidas quanto à suficiência de 4 componentes, principalmente ao observar as communalidades das variáveis, já que existiam variáveis com 0.08 de communalidade, a título de exemplo. A nossa escolha inicial com 4 componentes revelou que algumas variáveis, como a densidade populacional, investimento estrangeiro e crescimento económico, não estavam adequadamente explicadas.

Desta forma, e após recomendação do professor, considerámos a inclusão de 6 componentes, conforme sugerido pelo critério de Kaiser. Com esta expansão, as communalidades das variáveis mencionadas melhoraram significativamente, sendo a communalidade mais baixa de 0.29, levando-nos a reconsiderar a nossa escolha inicial. Desta forma foi decidido fazer uma análise com 6 componentes uma vez que capta muito melhor a variabilidade do nosso dataset.

## 7. Análise de PCAs

Depois de calcular as componentes principais, analisámos os loadings resultantes e aplicámos o método de rotação VariMax. Este procedimento permitiu-nos examinar quais variáveis que estavam mais fortemente associadas a cada componente gerada. Desta maneira, conseguimos tirar conclusões específicas sobre cada componente, identificando as variáveis mais influentes em cada uma e sugerindo possíveis rótulos ou designações para essas componentes.

- **RC1 (Acesso a infraestruturas e dinâmicas de trabalho):**

**Principais variáveis retidas:** Access\_to\_electricity\_percentage\_of\_population, Age\_dependency\_ratio\_old\_percentage\_of\_working\_age\_population, Age\_dependency\_ratio\_youth\_percentage\_of\_working\_age\_population, Contributing\_family\_workers\_total\_percentage\_of\_total\_employment\_modeled\_ILO\_estimate, Employment\_in\_agriculture\_percentage\_of\_total\_employment\_modeled\_ILO\_estimate, Employment\_in\_services\_percentage\_of\_total\_employment\_modeled\_ILO\_estimate, GDP\_per\_capita\_current\_US, Individuals\_using\_the\_Internet\_percentage\_of\_population, Population\_growth\_annual\_percentage, Self-employed\_total\_percentage\_of\_total\_employment\_modeled\_ILO\_estimate, Vulnerable\_employment\_total\_percentage\_of\_total\_employment\_modeled\_ILO\_estimate.

- **RC2 (Precariedade nas condições laborais):**

**Principais variáveis retidas:**

Employment\_to\_population\_ratio\_15more\_total\_percentage\_modeled\_ILO\_estimate,

Labor\_force\_participation\_rate\_for\_ages\_15\_24\_total\_percentage\_modeled\_IL  
O\_estimate,

Ratio\_of\_female\_to\_male\_labor\_force\_participation\_rate\_percentage\_modeled  
\_ILO\_estimate,

Unemployment\_total\_percentage\_of\_total\_labor\_force\_modeled\_ILO\_estimat  
e.

- **RC3 (Formação e qualificação profissional):**

**Principais variáveis retidas:**

Primary\_education\_duration\_years, Secondary\_education\_duration\_years.

- **RC4 (Desenvolvimento económico):**

**Principais variáveis retidas:**

GDP\_growth\_annual\_percentage,

Preprimary\_education\_duration\_years.

- **RC5 (Evolução demográfica):**

**Principais variáveis retidas:**

Population\_density\_people\_per\_sq.\_km\_of\_land\_area,

Self\_employed\_total\_percentage\_of\_total\_employment\_modeled\_ILO\_estimat  
e,

Unemployment\_total\_percentage\_of\_total\_labor\_force\_modeled\_ILO\_estimat  
e.

- **RC6 (Investimento estrangeiro):**

**Principais variáveis retidas:**

Foreign\_direct\_investment\_net\_inflows\_BoP\_current\_US,

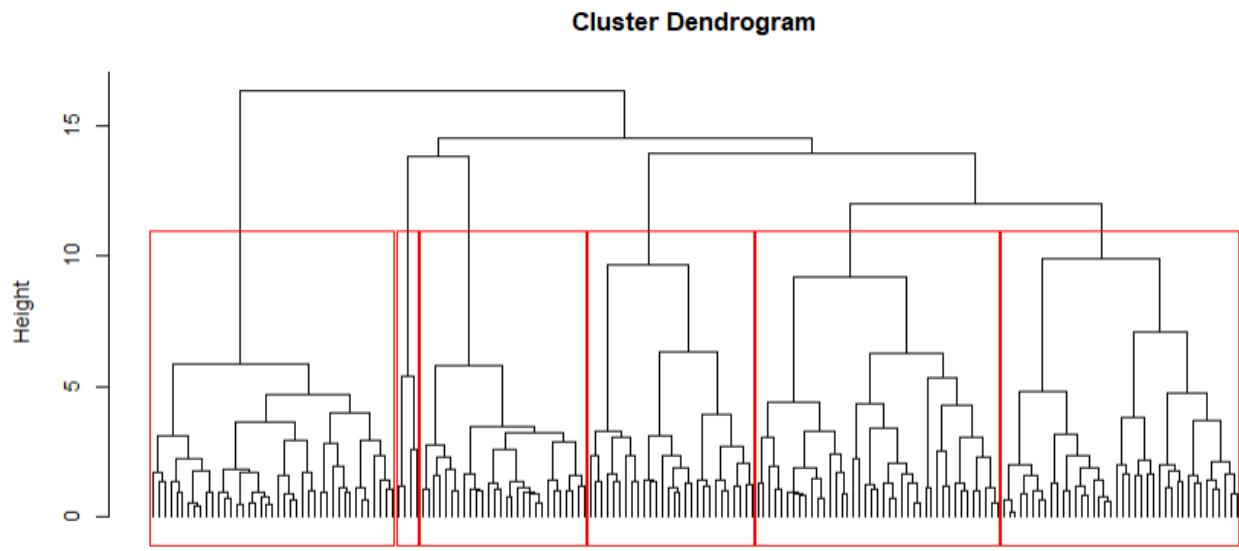
Compulsory\_education\_duration\_years

## **8. Identificação da heterogeneidade na base de dados**

De seguida, procedemos ao clustering das componentes criadas através de PCA para melhor compreensão da estrutura dos nossos dados onde foram testadas diferentes metodologias como o Clustering Hierárquico, K-means e posteriormente clustering probabilístico também.

A análise foi iniciada com o Clustering Hierárquico onde se aplicou o método "Ward.D2" aos dados, esta abordagem permitiu-nos visualizar as relações entre os clusters através de um dendrograma, onde se identificaram 5 e 6 clusters distintos, tendo-se feito algumas experiências para os dois casos. A imagem 1 apresenta aquele que considerámos o melhor cenário, com os cortes para o dendrograma em 6 clusters. Posteriormente, para validação da qualidade dos clusters procedeu-se à criação de uma silhueta onde a média da sua largura foi de 0.19 e 0.20, respetivamente.

Depois foi explorado o método "complete", onde se optou por segmentar os dados em 3 clusters. Foi novamente criada uma silhueta onde a média da sua largura foi inferior (0.14). No entanto, observou-se que, através deste método, não existia uma visualização tão clara para a determinação dos cortes apropriados, assim, acabou por se verificar que o método "Ward-D2" era mais indicado para o propósito em questão.



*Imagen 1 - Dendrogram com corte em 6 clusters (ward)*

Conhecendo o valor do nosso  $k$  através do clustering hierárquico ( $k=6$ ), avançámos para o k-means. Para validar ou encontrar um  $k$  alternativo, recorremos ao scree plot (anexo D), onde se observou que o ponto ideal de corte seria 7 (embora o elbow não seja tão direto de se observar). De todas as formas, optou-se por se utilizar o número de clusters sugerido com o clustering hierárquico com  $k=6$  para o K-means, onde depois os clusters foram armazenados numa nova coluna do dataframe final.

## 9. Conclusões obtidas com o clustering

As conclusões retiradas da criação dos clusters e a respetiva caracterização feita com base no anexo E pode ver-se abaixo:

- **Cluster 1 - Economias emergentes com mercados de trabalho precários -**  
Tem um nível mais elevado no que respeita ao acesso a infraestruturas e desenvolvimento económico, enquanto regista paralelamente um nível alto de precariedade. Em termos de educação e formação profissional tem valores mais baixos e acaba por não receber tanta atenção em termos de

investimento estrangeiro.

- **Cluster 2 - Mercados estáveis e envelhecidos** - Baixo nível de precariedade e educação e fraca evolução demográfica.
- **Cluster 3 - Hubs de emprego impulsionados pelo investimento** - Caracteriza-se pelo elevado investimento estrangeiro.
- **Cluster 4 - Mercados em crescimento populacional** - Caracteriza-se pelo baixo desenvolvimento económico e elevada evolução demográfica.
- **Cluster 5 - Mercados desenvolvidos com elevado grau de capacitação profissional** - Apresenta elevados níveis de educação e formação, acesso a infraestruturas e desenvolvimento económico.
- **Cluster 6 - Economias de trabalho precárias com infraestruturas limitadas** - Tem fracas infraestruturas e um elevado nível de precariedade.

Como foi possível fazer a caracterização dos clusters com os dados disponibilizados através das variáveis de INPUT, optámos por não utilizar variáveis de PROFILE. Ainda, assim, para complementar a análise que conduzimos ao nível do clustering, além do K-means optámos por fazer também o PAM, com k=6. Ao olhar para o anexo F, do clusplot, vemos que há algum overlap entre clusters, o que sugere a possibilidade dos mercados de trabalho não apresentarem muitos fatores distintivos entre eles.

Alternativamente a estas técnicas de clustering ditas “tradicionais”, foram realizadas também técnicas de clustering probabilístico, nomeadamente análises recorrendo a BIC e GMM para complementar o estudo conduzido. Os resultados da aplicação destas técnicas de clustering probabilístico podem ver-se nos anexos G e H.

## **10. Principais conclusões**

Em primeiro lugar, foi feito um mapeamento do panorama atual do mercado de trabalho, de forma a que o objetivo do trabalho pudesse ser adequado tendo em conta aqueles que são os desafios da actualidade. De seguida, e após a definição do objetivo do trabalho, foi feita toda a limpeza e tratamento dos dados, onde se eliminaram os outliers e se tratou dos valores omissos através da técnica MICE.

Ao considerar os resultados do teste de esfericidade de Bartlett, do índice de Kaiser-Meyer-Olkin (KMO) e da matriz de correlação podemos concluir que o uso do PCA era adequado a este caso. Foi testado o uso de 3, 4 e 6 componentes, tendo sido escolhido 6 componentes tendo em conta o critério de Kaiser com os seguintes nomes: acesso a infraestruturas e dinâmicas de trabalho, precariedade nas condições laborais, formação e qualificação profissional, desenvolvimento económico, evolução demográfica e investimento estrangeiro.

De seguida, procedeu-se ao clustering de forma a proceder à melhor compreensão da estrutura dos nossos dados. Assim foram testadas diferentes metodologias como o Clustering Hierárquico e o K-means. Tendo em conta estas metodologias foi possível concluir que o melhor resultado seria a criação de 6 clusters. Para validação da qualidade dos clusters, foi criada uma silhueta onde a média da sua largura foi de 0.19 e 0.20, respetivamente.

Por fim, foram definidos e caracterizados os clusters com os seguintes nomes: economias emergentes com mercados de trabalho precários, mercados estáveis e envelhecidos, hubs de emprego impulsionados pelo investimento, mercados em crescimento populacional, mercados desenvolvidos com elevado grau de capacitação profissional e economias de trabalho precárias com infraestruturas limitadas. Esta análise foi ainda complementada com uma camada de estudo que envolveu o clustering probabilístico e que pode ser visto nos últimos anexos disponibilizados.

## 11. Referências

- [1] United Nations Convention to Combat Desertification (UNCCD). (s.d.). World Development Indicators (WDI). Disponível em: <https://www.unccd.int/resources/knowledge-sharing-system/world-development-indicators-wdi>  
Consultado a 10/01/2024
- [2] The World Bank. (s.d.). World Development Indicators: The Story. Disponível em: <https://datatopics.worldbank.org/world-development-indicators/stories/world-development-indicators-the-story.html>  
Consultado a 10/01/2024
- [3] Huang, H., Chi, M., Song, Y., & Jin, H. (15/03/2022). Mining World Indicators for Analyzing and Modeling the Development of Countries. ACM/IMS Transactions on Data Science, 2(4), Artigo No: 30. Disponível em: <https://doi.org/10.1145/3488059>  
Consultado a 10/01/2024
- [4] CELL. Disponível em: [https://www.cell.com/heliyon/pdf/S2405-8440\(20\)30845-8.pdf](https://www.cell.com/heliyon/pdf/S2405-8440(20)30845-8.pdf).  
Consultado em 11/01/2024
- [5] Splashtop. (22/09/2023). O que é trabalho híbrido? Splashtop Inc. Disponível em: <https://www.splashtop.com/pt/blog/what-is-hybrid-work>  
Consultado a 10/01/2024
- [6] Randstad Portugal. (28/11/2023). O que aprendemos sobre o modelo de trabalho híbrido? Randstad. Disponível em: <https://www.randstad.pt/tendencias-360/mercado-do-trabalho/o-que-aprendemos-sobre-o-modelo-de-trabalho-hibrido/>  
Consultado a 10/01/2024
- [7] Notícias ao Minuto. (14/06/2023). Nómadas digitais em Portugal: Conheça a nova tendência profissional. Notícias ao Minuto. Disponível em: <https://www.noticiasaoiminuto.com/casa/2340988/nomadas-digitais-em-portugal-conheca-a-nova-tendencia-profissional>  
Consultado a 10/01/2024
- [8] Público. (29/09/2021). Aumentou o risco de os recém-licenciados ficarem desempregados: estudo. Público. Disponível em:

<https://www.publico.pt/2021/09/29/p3/noticia/aumentou-risco-recemlicenciados-ficar-em-desempregados-estudo-1979290>

Consultado a 10/01/2024

[9] Observador. (02/04/2022). Workplace digital: os desafios do ambiente de trabalho híbrido. Observador. Disponível em:

<https://observador.pt/explicadores/workplace-digital-os-desafios-do-ambiente-de-trabalho-hibrido/>

Consultado a 10/01/2024

[10] <https://www.scholaro.com/db/Countries/New-Caledonia/Education-System>

## 12. Anexos

### Anexo A - Dicionário de features

- **Country\_Name e Country\_Code:** Variáveis categóricas referentes ao nome e código do país, respetivamente. Ambos com 182 observações.
- **Access\_to\_electricity\_percentage\_of\_population:** Percentagem da população com acesso à eletricidade, variando entre 7.748% e 100%, com média de 85.699%.
  - **Relevância:** Reflete a infraestrutura básica e é um fator crítico para países onde a população ativa esteja mais alocada a atividades tecnológicas ou atividades onde o acesso à energia seja essencial para a execução de determinadas tarefas.
- **Age\_dependency\_ratio\_old\_percentage\_of\_working\_age\_population:** Percentagem de pessoas idosas em relação à população em idade ativa, variando entre 1.685% e 50.971%, com média de 14.583%.

- **Relevância:** Crucial para entender o nível de dependência da população idosa face à classe trabalhadora; ponderação de reformas; etc.
- **Age\_dependency\_ratio\_young\_percentage\_of\_working\_age\_population:** Percentagem de pessoas jovens em relação à população em idade ativa, variando entre 16.23% e 100.47%, com média de 44.43%.
  - **Relevância:** Essencial para compreender o nível de dependência da população jovem face à classe trabalhadora; incentivos de entrada no mercado laboral para jovens; potenciais restrições de idades; etc..
- **Compulsory\_education\_duration\_years:** Duração obrigatória da educação em anos, variando entre 0 e 17 anos, com 13 valores em falta, com média de 9.769.
  - **Relevância:** Indica os anos de “investimento” na formação e capacitação de capital humano, com impacto direto na capacidade técnica e de inovação dos profissionais no mercado de trabalho.
- **Contributing\_family\_workers\_total\_percentage\_of\_total\_employment\_modeled\_ILO\_estimate:** Percentagem de trabalhadores familiares em relação ao emprego total, variando entre 0.00247% e 49.70619%, com média de 8.07828%.
  - **Relevância:** Insights sobre a estrutura do mercado de trabalho, destacando trabalhadores familiares e empregadores.
- **Employers\_total\_percentage\_of\_total\_employment\_modeled\_ILO\_estimate** : Percentagem de empregadores em relação ao emprego total, variando entre 0.04993% e 12.32320%, com média de 3.16618%.
  - **Relevância:** Mostra a participação dos empregadores no tecido económico e a sua capacidade de influência nos mercados laborais.
- **Employment\_in\_agriculture\_percentage\_of\_total\_employment\_modeled\_ILO\_estimate:** Percentagem de empregados no setor agrícola em relação ao emprego total, variando entre 0.1864% e 85.8503%, com média de 22.6025%.

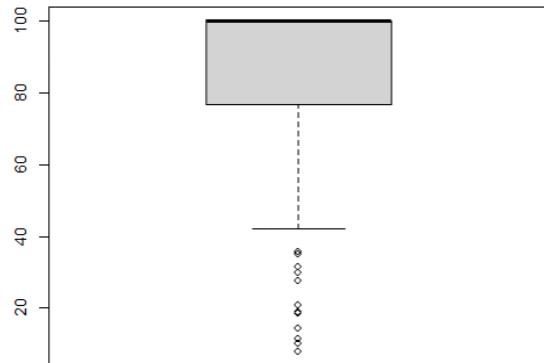
- **Relevância:** Destaca as especificidades dos empregos no setor agrícola.
- **Employment\_in\_industry\_percentage\_of\_total\_employment\_modeled\_ILO\_estimate:** Percentagem de empregados no setor industrial em relação ao emprego total, variando entre 3.619% e 53.999%, com média de 20.041%.
  - **Relevância:** Destaca as especificidades dos empregos no setor industrial.
- **Employment\_in\_services\_percentage\_of\_total\_employment\_modeled\_ILO\_estimate:** Percentagem de empregados no setor de serviços em relação ao emprego total, variando entre 10.53% e 92.70%, com média de 57.36%.
  - **Relevância:** Destaca as especificidades dos empregos no setor de serviços..
- **Employment\_to\_population\_ratio\_15more\_total\_percentage\_modeled\_ILO\_estimate:** Proporção de emprego para a população com 15 anos ou mais, variando entre 22.44% e 87.74%, com média de 55.57%.
  - **Relevância:** Fornece a visão da participação na força de trabalho, com base na idade.
- **Foreign\_direct\_investment\_net\_inflows\_BoP\_current\_US:** Fluxos líquidos de investimento direto estrangeiro em dólares americanos correntes, variando entre -1.343e+11 e 4.931e+11.
  - **Relevância:** Reflete confiança na economia, impactando crescimento, criação de empregos e modernização.
- **GDP\_growth\_annual\_percentage:** Crescimento anual do Produto Interno Bruto (PIB), variando entre -7.085% e 41.745%, com média de 5.584%.
  - **Relevância:** Indicador-chave para avaliar desempenho económico e capacidade de criar riqueza.

- **GDP\_per\_capita\_current\_US:** Produto Interno Bruto per capita em dólares americanos atuais, variando entre 221.2 e 133590.1.
  - **Relevância:** Avalia padrão de vida médio, fornecendo insights sobre distribuição de riqueza e desenvolvimento
- **Individuals\_using\_the\_Internet\_percentage\_of\_population:** Percentagem da população que utiliza a internet, variando entre 5.805% e 100%.
  - **Relevância:** Indica conectividade e acesso à informação, influenciando educação e oportunidades de emprego
- **Inflation\_consumer\_prices\_annual\_percentage:** Taxa anual de inflação de preços ao consumidor, variando entre -0.7728% e 359.0930%.
  - **Relevância:** Indicador crítico para avaliar estabilidade económica, influenciando investimentos, poder de compra e ambiente de negócios
- **Labor\_force\_participation\_rate\_for\_ages\_15\_24\_total\_percentage\_modeled\_ILO\_estimate:** Taxa de participação da força de trabalho na faixa etária de 15 a 24 anos, variando entre 13.62% e 74.55%, com média de 40.11%.
  - **Relevância:** Crucial para entender o impacto do trabalho em idades muito jovens.
- **Population\_density\_people\_per\_sq.\_km\_of\_land\_area:** Densidade populacional por quilômetro quadrado de área terrestre, variando entre 2.149 e 20806.273 pessoas.
  - **Relevância:** Fundamental para compreender a saturação do mercado de trabalho.
- **Population\_growth\_annual\_percentage:** Crescimento populacional anual, variando entre -4.2566% e 3.7074%, com média de 0.9439%.
  - **Relevância:** Determinante para projeções de procura por serviços, educação e emprego.

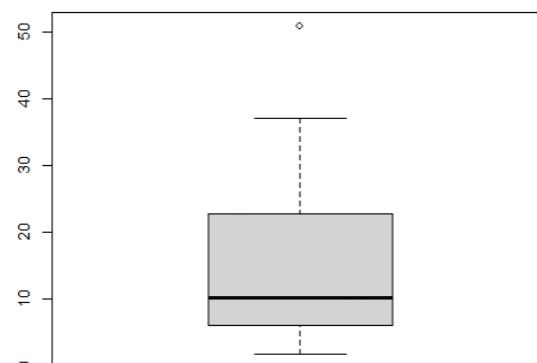
- **Preprimary\_education\_duration\_years:** Duração da educação pré-primária em anos, variando entre 1 e 7 anos, com 9 valores ausentes, com média de 4.087.
  - **Relevância:** Reflete o investimento em capital humano e preparação da população para o mercado de trabalho.
- **Primary\_education\_duration\_years:** Duração da educação primária em anos, variando entre 4 e 8 anos, com 1 valor ausente, com média de 5.707.
  - **Relevância:** Fornece insights sobre os investimentos em capital humano e a qualidade do sistema de ensino.
- **Ratio\_of\_female\_to\_male\_labor\_force\_participation\_rate\_percentage\_modeled\_ILO\_estimate:** Proporção entre a taxa de participação na força de trabalho de mulheres e homens, variando entre 8.479% e 105.571%, com média de 72.378%.
  - **Relevância:** Igualdade no acesso a oportunidades de emprego, influenciando a participação das mulheres nos mercados laborais.
- **Secondary\_education\_duration\_years:** Duração da educação secundária em anos, variando entre 4 e 9 anos, com 2 valores ausentes, com média de 6.472.
  - **Relevância:** Reflete a duração da educação secundária, essencial para avaliar a preparação da população para desafios mais avançados no mercado de trabalho
- **Self\_employed\_total\_percentage\_of\_total\_employment\_modeled\_ILO\_estimate:** Percentagem de trabalhadores autónomos em relação ao emprego total, variando entre 0.3774% e 95.1025%, com média de 39.8718%.
  - **Relevância:** Fornecem insights sobre estrutura do emprego e empreendedorismo na economia.

- **Unemployment\_total\_percentage\_of\_total\_labor\_force\_modeled\_ILO\_estimated**  
**ate:** Taxa de desemprego em relação à força de trabalho total, variando entre 0.170% e 27.954%, com média de 7.898%.
  - **Relevância:** Indicador crítico para avaliar eficiência do mercado de trabalho e condições económicas gerais.
- **Vulnerable\_employment\_total\_percentage\_of\_total\_employment\_modeled\_ILO\_estimate:** Percentagem de emprego vulnerável em relação ao emprego total, variando entre 0.04968% e 94.31692%, com média de 36.70561%.
  - **Relevância:** Indica segurança do emprego, refletindo condições de trabalho e qualidade do emprego na economia.

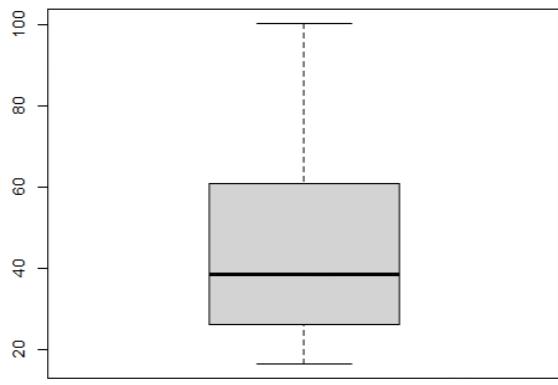
## Anexo B - Boxplots e análise de outliers



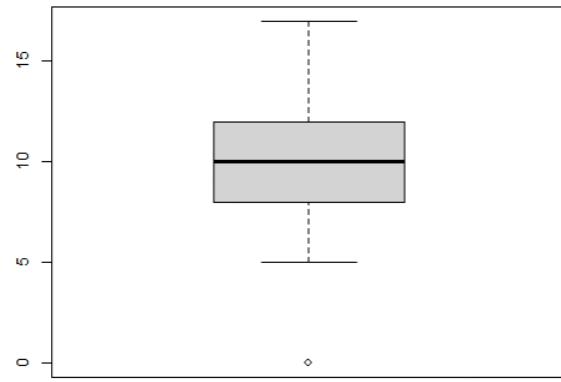
Boxplot de  
Access\_to\_electricity\_percentage\_of\_population



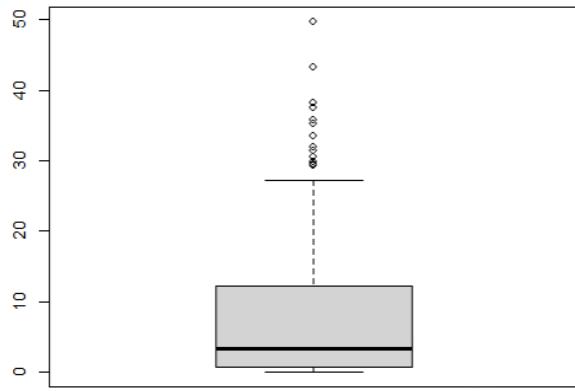
Boxplot de  
Age\_dependency\_ratio\_old\_percentage\_of\_working\_age\_population



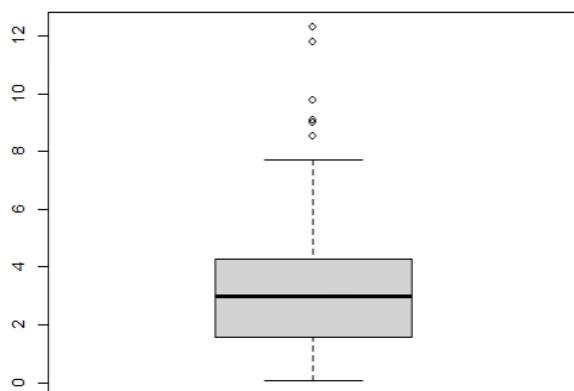
*Boxplot de  
Age\_dependency\_ratio\_youth\_percentage\_of\_w  
orking\_age\_population*



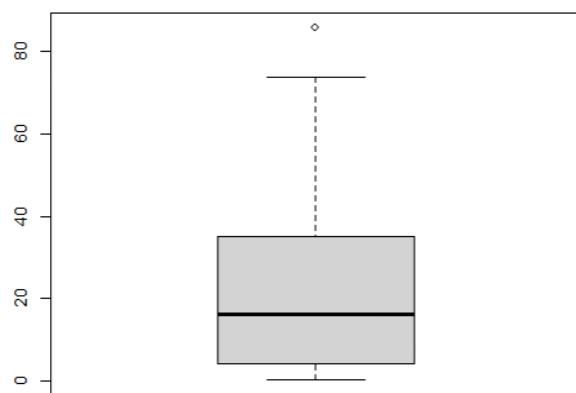
*Boxplot de  
Compulsory\_education\_duration\_years*



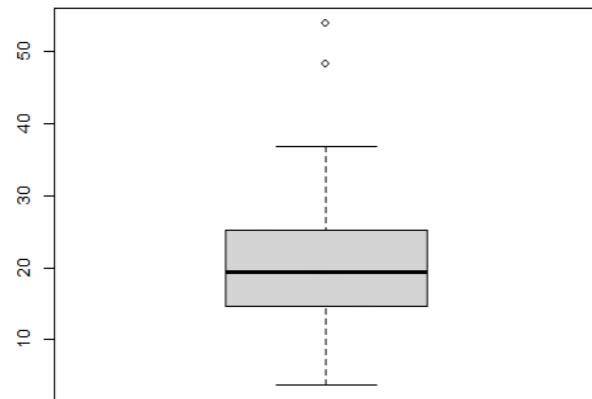
*Boxplot  
Contributing\_family\_workers\_total\_percentage\_o  
f\_total\_employment\_modeled\_ILO\_estimate*



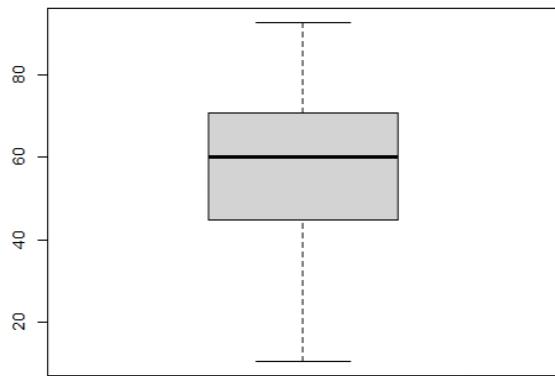
*Boxplot de  
Employers\_total\_percentage\_of\_total\_employme  
nt\_modeled\_ILO\_estimate*



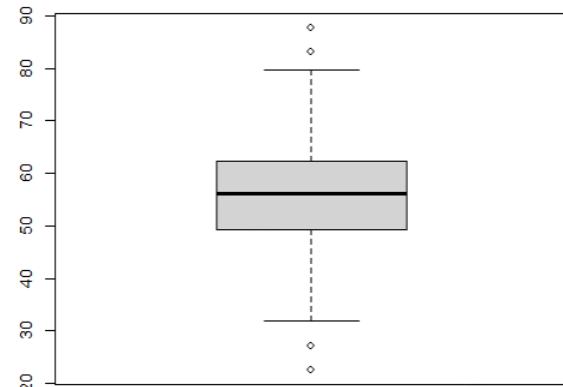
*Boxplot de  
Employment\_in\_agriculture\_percentage\_of\_total\_employment\_modeled\_ILO\_estimate*



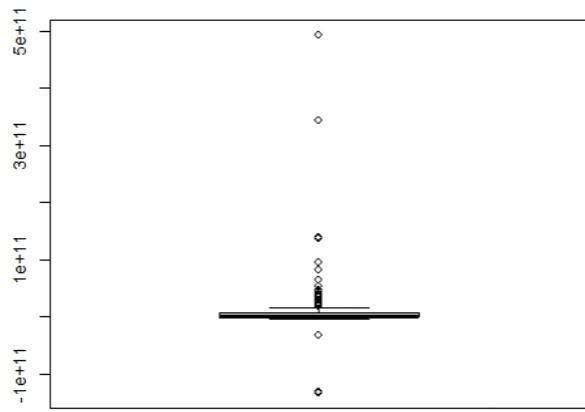
*Boxplot de  
Employment\_in\_industry\_percentage\_of\_total\_employment\_modeled\_ILO\_estimate*



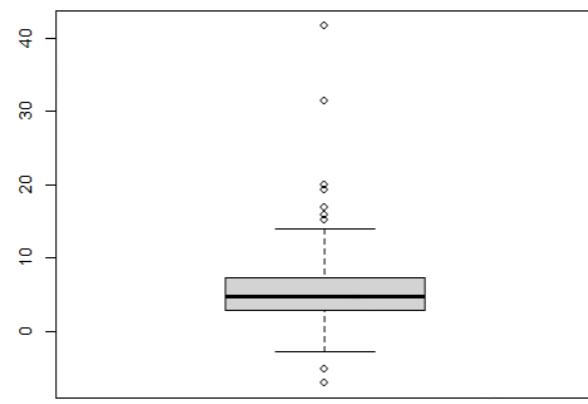
*Boxplot de  
Employment\_in\_services\_percentage\_of\_total\_employment\_modeled\_ILO\_estimate*



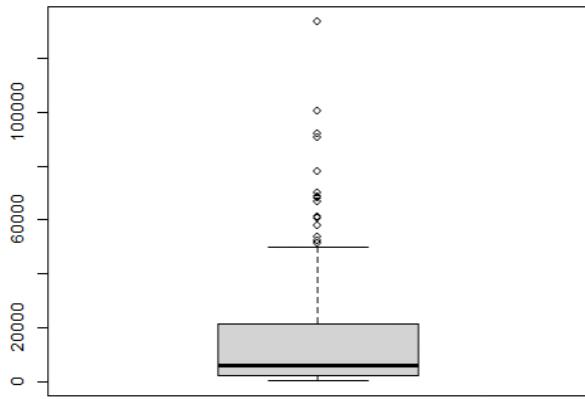
*Boxplot de  
Employment\_to\_population\_ratio\_15more\_total\_percentage\_modeled\_ILO\_estimate*



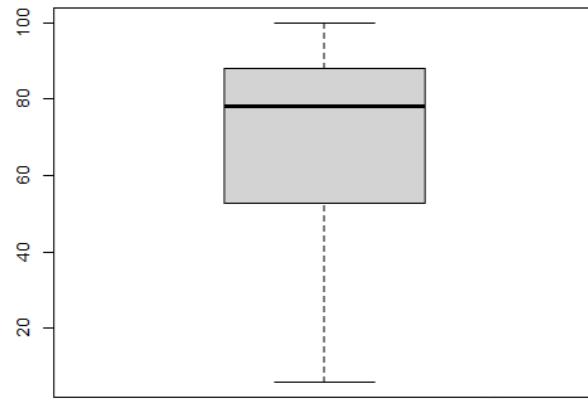
Boxplot de  
Foreign\_direct\_investment\_net\_inflows\_BoP\_cur  
rent\_US



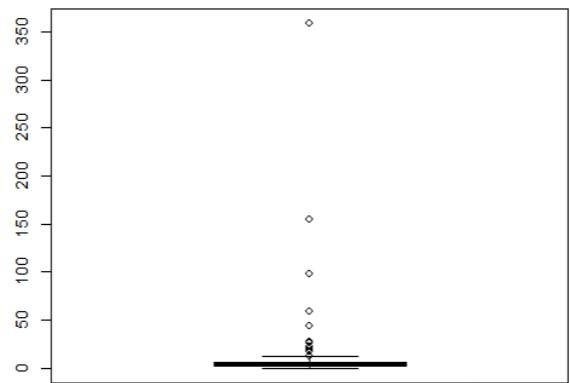
Boxplot de GDP\_growth\_annual\_percentage



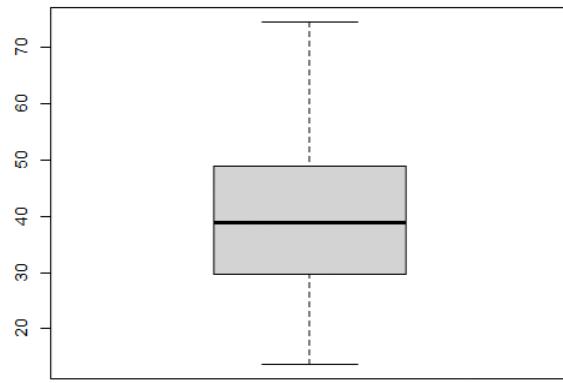
Boxplot de GDP\_per\_capita\_current\_US



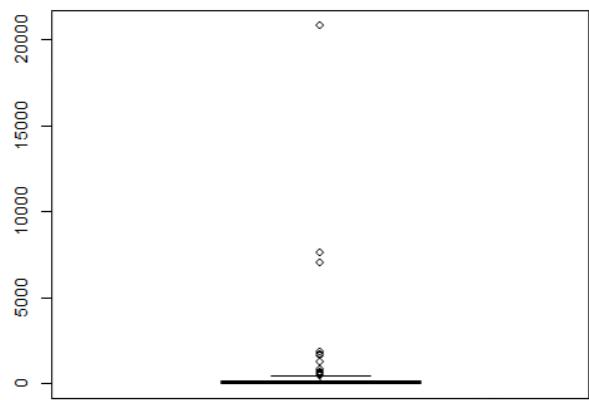
Boxplot de  
Individuals\_using\_the\_Internet\_percentage\_of\_p  
opulation



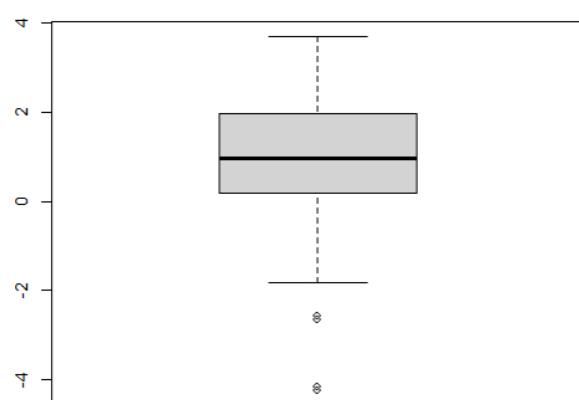
Boxplot de  
*Inflation\_consumer\_prices\_annual\_percentage*



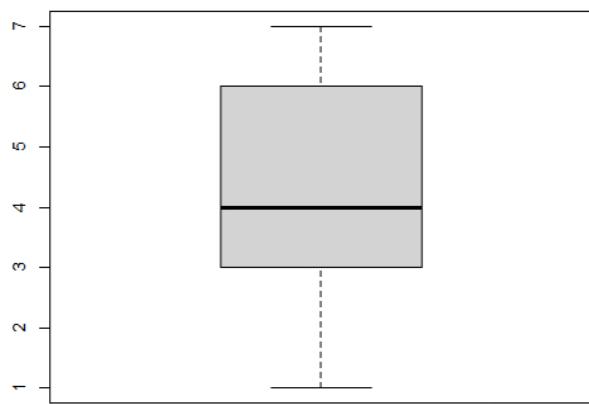
Boxplot de  
*Labor\_force\_participation\_rate\_for\_ages\_15\_24\_total\_percentage\_modeled\_ILO\_estimate*



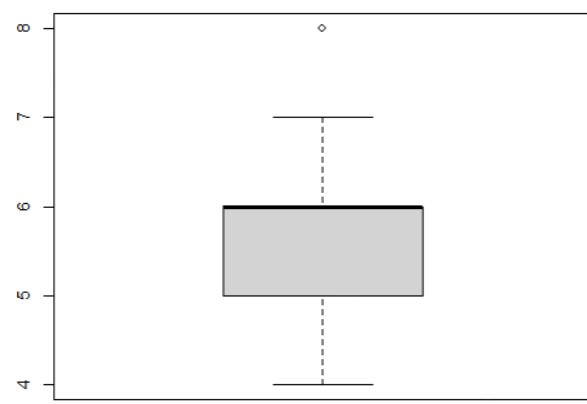
Boxplot de  
*Population\_density\_people\_per\_sq\_km\_of\_land\_area*



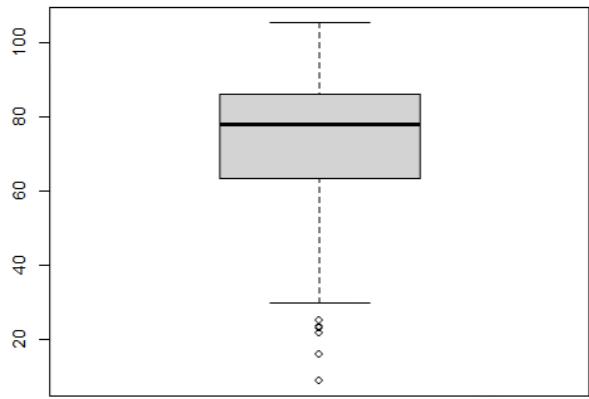
Boxplot de  
*Population\_growth\_annual\_percentage*



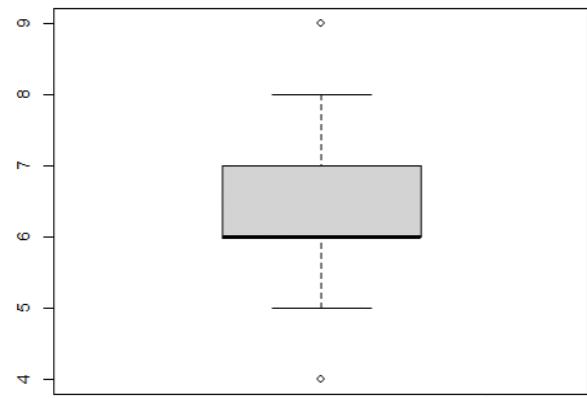
*Boxplot de  
Preprimary\_education\_duration\_years*



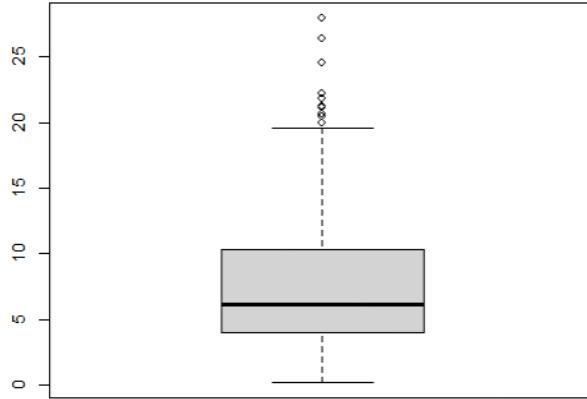
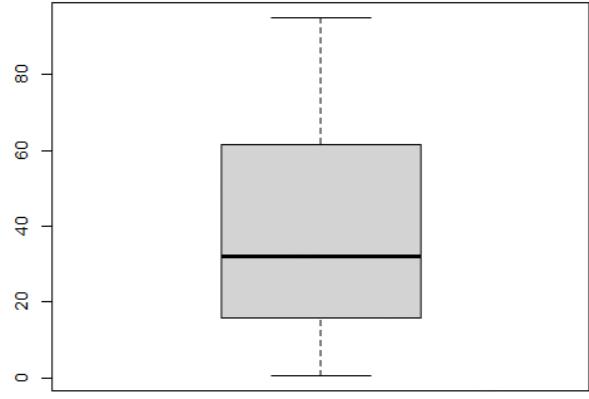
*Boxplot de Primary\_education\_duration\_years*



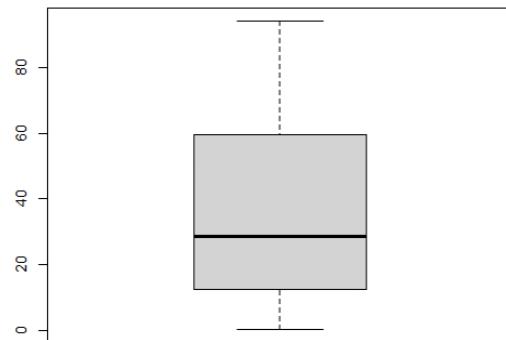
*Boxplot de  
Ratio\_of\_female\_to\_male\_labor\_force\_participation\_rate\_percentage\_modeled\_ILO\_estimate*



*Boxplot de Secondary\_education\_duration\_years*

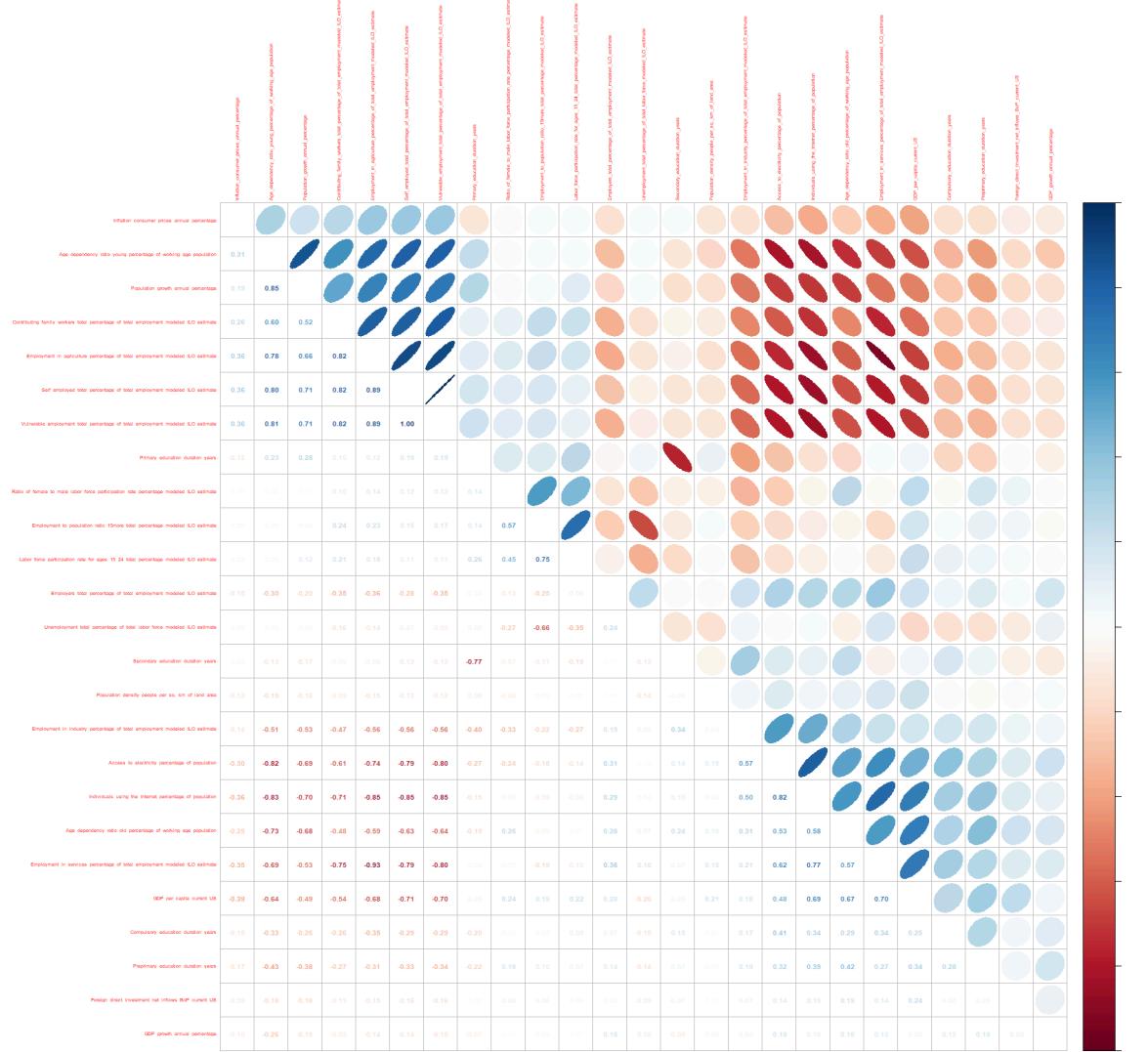


*Boxplot de*  
*Self\_employed\_total\_percentage\_of\_total\_employment\_modeled\_ILO\_estimate*      *Unemployment\_total\_percentage\_of\_total\_labor\_force\_modeled\_ILO\_estimate*



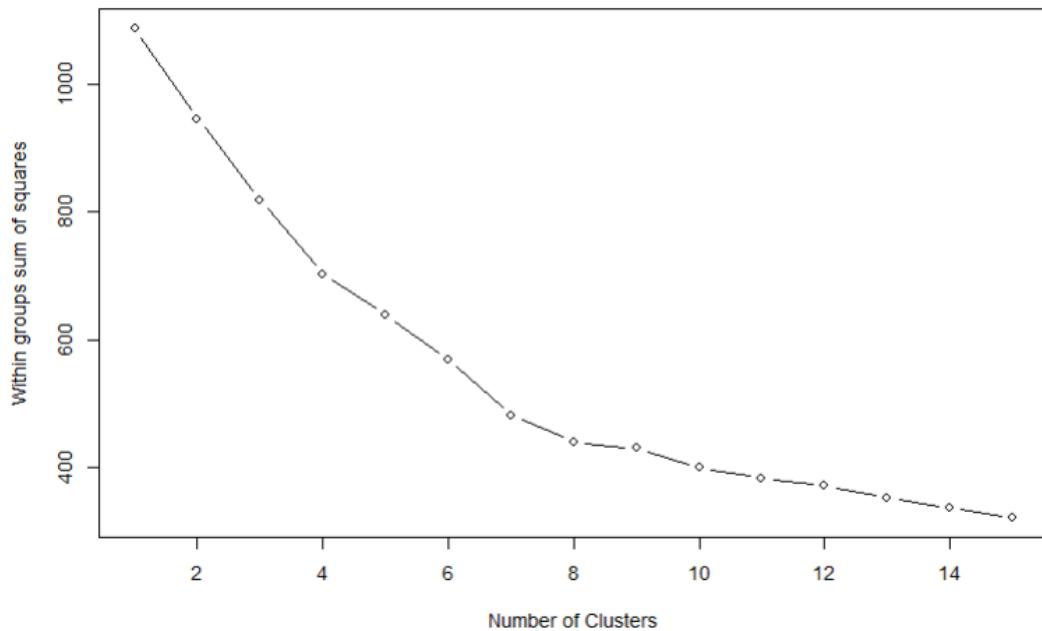
*Boxplot de*  
*Vulnerable\_employment\_total\_percentage\_of\_total\_employment\_modeled\_ILO\_estimate*

## Anexo C - Matriz de correlação

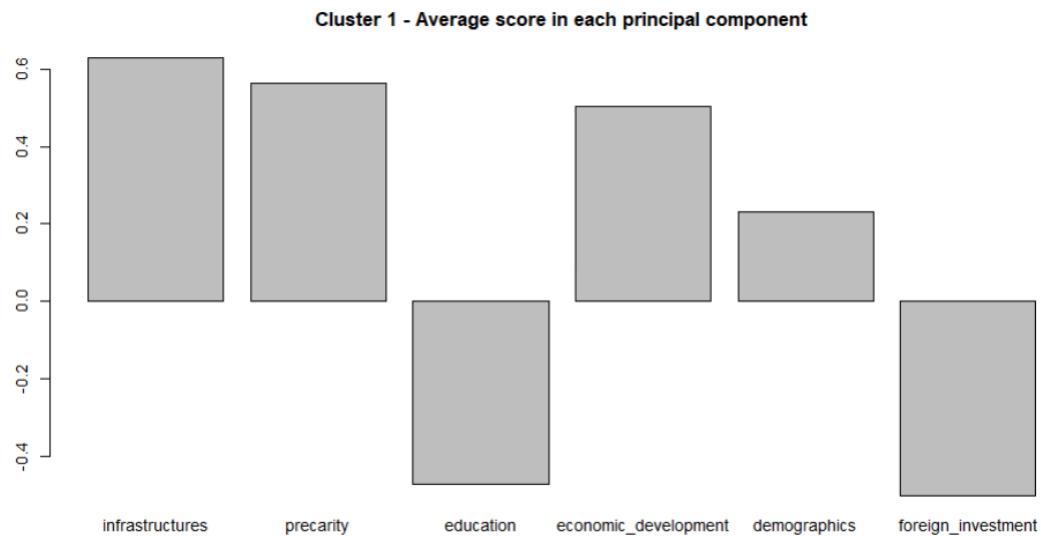


**Nota:** A figura tem resolução de 1800x1800, pelo que fazendo zoom em cada um dos componentes é possível ver os valores de correlação e respectivas variáveis.

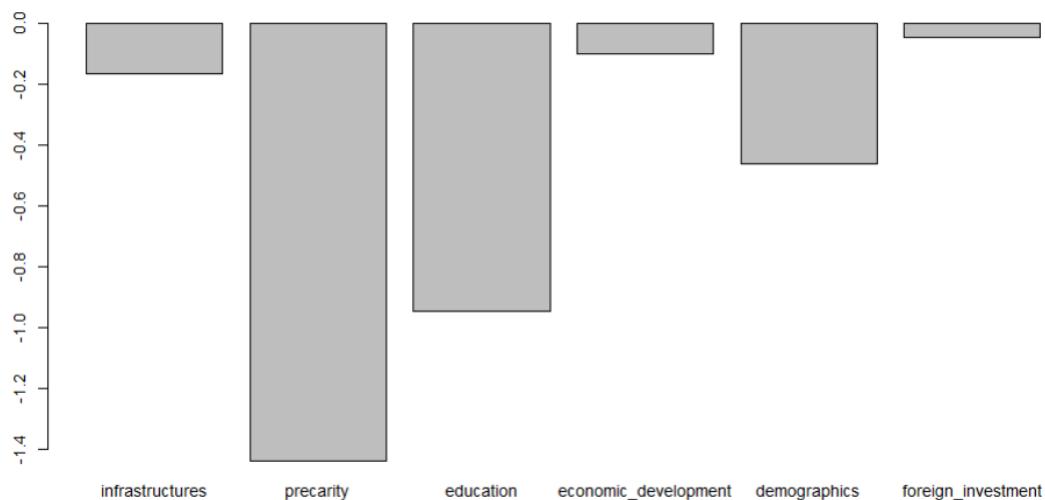
#### Anexo D - Scree-plot para encontrar o “cotovelo” para o k-means



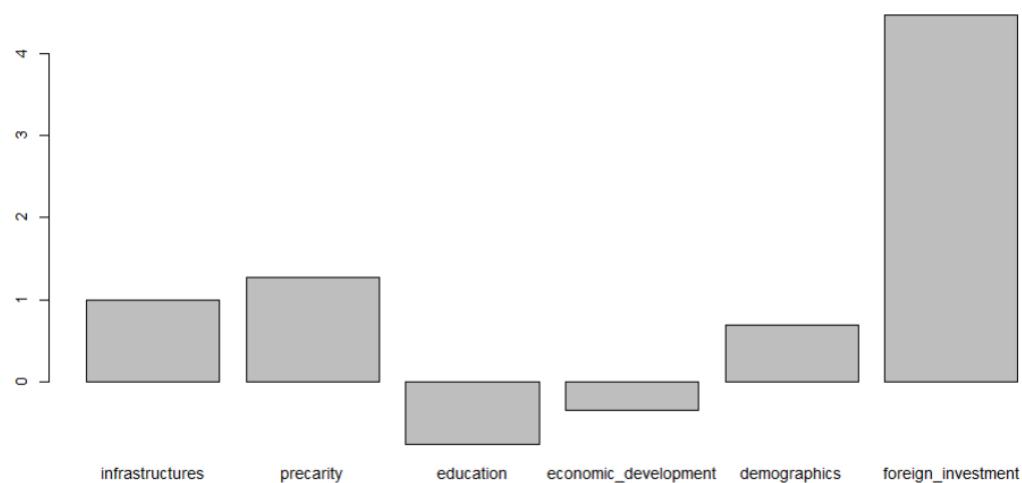
#### Anexo E - Barplots dos average scores em cada PC



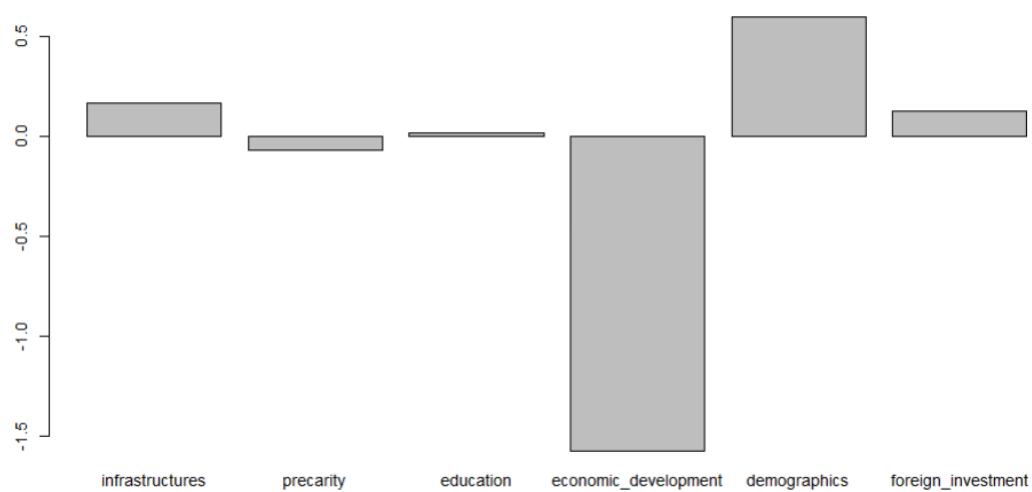
**Cluster 2 - Average score in each principal component**



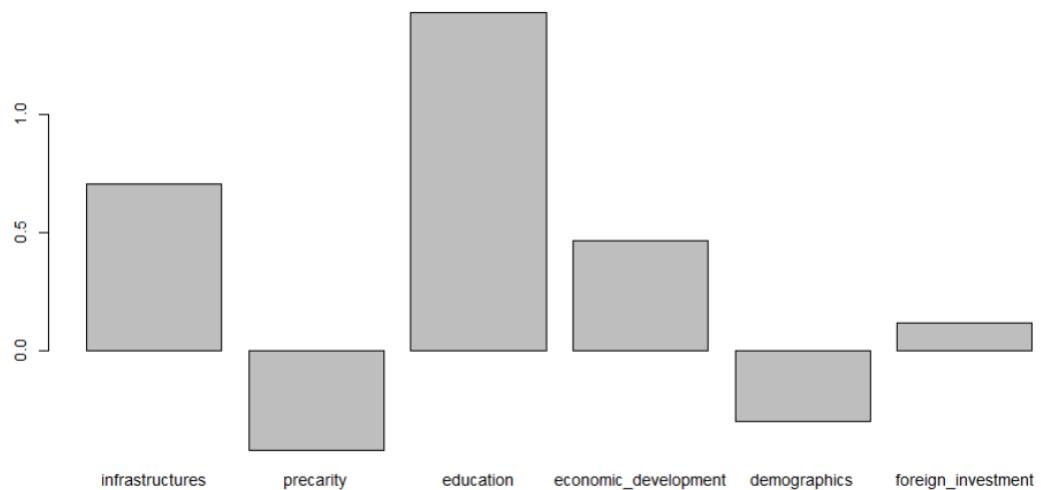
**Cluster 3 - Average score in each principal component**



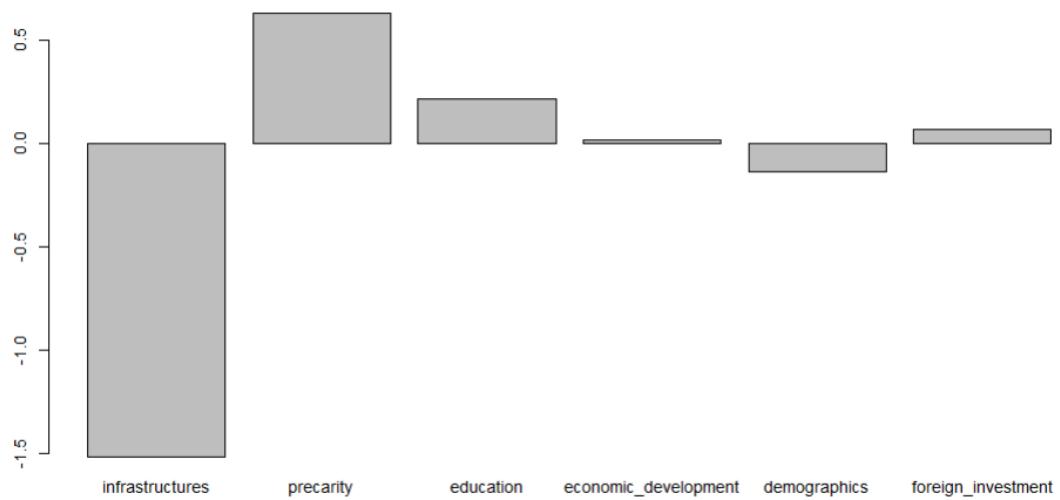
**Cluster 4 - Average score in each principal component**



**Cluster 5 - Average score in each principal component**

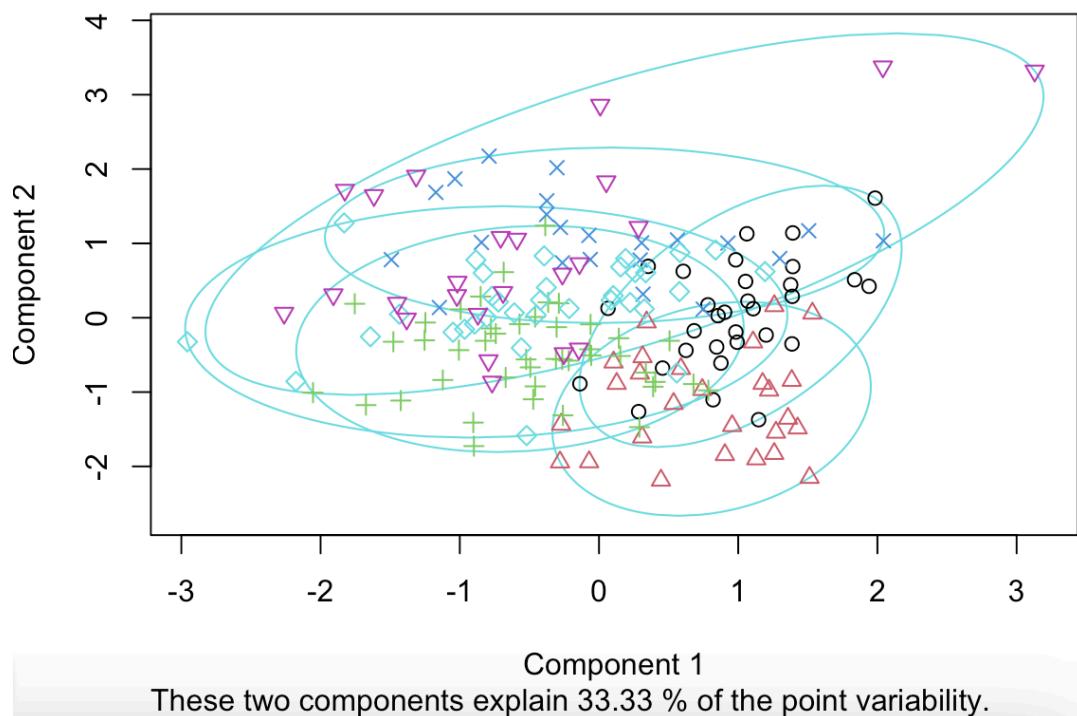


**Cluster 6 - Average score in each principal component**

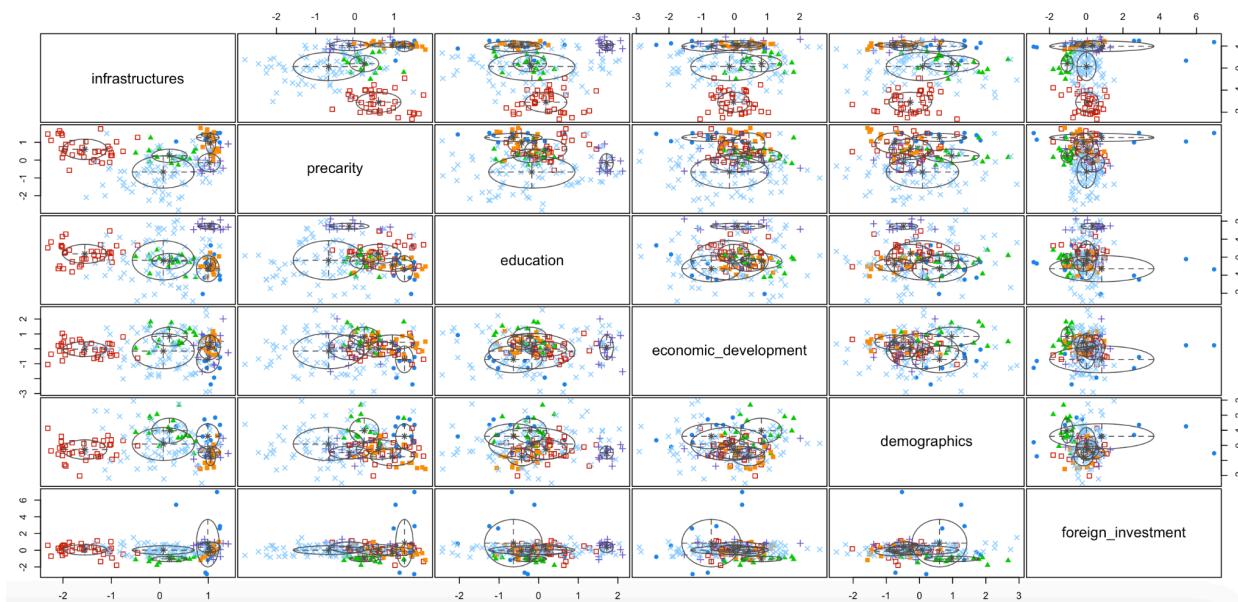


## Anexo F - PAM clustering

`clusplot(pam(x = std_data, k = 6))`



## Anexo G - Gráficos de classificação



## Anexo H - Gráficos de incerteza

