



iscte INSTITUTO
UNIVERSITÁRIO
DE LISBOA



Análise de clusters de estudantes da Irlanda

Métodos de Aprendizagem não Supervisionada

Trabalho realizado por:

- Allan Karden, 103380 CDB1
- André Plancha, 105289 CDB2
- Diogo Freitas, 104841 CDB1
- Ricardo Ângelo, 104826 CDB1





Contents

Introdução	3
Dados	3
Identificação das dimensões da análise.....	4
RC1 - Sucesso Académico (SA)	5
RC3 - Estatuto socioeconómico (ESE)	6
RC2 - Apoio por parte do professor (APP).....	6
RC4 - Alegria e bem-estar interior (ABEI).....	6
RC5 - Nível de educação e informação sobre carreiras (NEIC)	7
RC6 – Mudanças na educação (ME).....	7
Clusters	7
Cluster 1 -> Trabalhadores-Estudantes	9
Cluster 2 -> Estudantes “Despreocupados”	10
Cluster 3 -> Alunos “Estrangeiros”	11
Model Selection	12
Conclusão.....	15
Webgrafia.....	16
Anexos	17
Variáveis dos PCA’s.....	17
ICT	19
Organização e governo	19
Students’ Well-Being.....	20
Teachers	20
Socio-economic status	21



Introdução

Uma equipa de cientista de dados foi contratada pelo Ministério da Educação da Irlanda para identificar grupos de estudantes de 15 anos semelhantes, a fim de obter *insights* sobre as características que os distinguem, dados estes fornecidos pelo PISA, um programa de avaliação internacional criada pela OECD¹. Esta identificação de grupos e de padrões nos dados dos estudantes, através de *clusters* (técnica usada em métodos não supervisionados), poderá trazer vantagens importantes para a educação do país que, consequentemente, trarão benefícios para o futuro do país.

Será utilizada *PCA*², técnica que tem como objetivo reduzir a dimensionalidade de uma base de dados e, após esse processo, criar-se-ão *clusters* para, finalmente, criar as tais comunidades de alunos mais semelhantes entre si, agrupando-os através das suas características.

É importante referir que, para a realização desta análise, será necessário ter um conhecimento abrangente de diversas características da Irlanda e da sua população, estando estas associadas diretamente com os estudantes de 15 anos que realizaram os testes, tais como: Diferenças socioeconómicas das famílias, situação psicológica e social dos diferentes estudantes, uso e conhecimento das tecnologias de informação e comunicação por parte da população, etc... Somado a isto, também se terá em conta as notas do PISA, que também serão importantes para agrupar os estudantes.

Falando um pouco da estrutura deste trabalho, todos os tópicos estarão devidamente identificados, sendo necessário passar pelas seguintes fases: **Análise dos dados**: onde será comentado e explicado, não só a base de dados que nos foi fornecida, como também os processos que foram usados para tratar a base de dados a fim de poder estudar os dados que esta contém; **Identificação das dimensões da análise**: sendo esta a fase onde se usará *PCA*, explicando as características de cada uma das componentes principais; **Identificação da heterogeneidade na base de dados**: aqui serão aplicados os *clusters*, onde cada um destes será analisado e terá atribuído um nome; **Conclusão**: Nesta última fase, serão retiradas conclusões dos resultados e do que ocorreu no decorrer desta análise, concluindo que possíveis influências os resultados poderão ter na tomada de decisões por parte do Ministério da Educação da Irlanda. No final, será possível encontrar a **Bibliografia** e os **Anexos**.

Dados

Nesta fase serão comentadas as características da base de dados. A base de dados inicial da Irlanda possuía 5577 linhas e 104 colunas, possuindo 18 colunas totalmente omissas, fazendo com que se tivesse apenas 86 colunas para analisar e possivelmente utilizar.

Para se obter uma análise confiável, optou-se por eliminar, não só as colunas que possuíam mais de 10% de valores omissos, como também as linhas que possuíam mais de 20% de valores omissos. Tomou-se esta decisão pois foi realizou-se imputação por regressão, de forma a obter os dados mais próximos da realidade, a fim de não prejudicar o estudo das comunidades. Esta escolha fez com que se ficasse com 5456 linhas e 53 colunas.

¹ OECD - The Organization for Economic Cooperation and Development

² PCA - Análise de Componentes Principais



É de se notar que foi realizado o teste de *MCAR*³ que demonstrou que os dados eram totalmente aleatórios, dando a entender que os resultados com e sem imputação seriam muito parecidos, que foi o que aconteceu.

Por último, sobre os dados, foi necessário retirar as variáveis de perfil, isto é, variáveis que permitissem identificar características do indivíduo, e as escolhidas foram: a idade, se é imigrante, o género, o nível do ISCED e o grau de ensino dos pais.

Identificação das dimensões da análise

Na preparação da *PCA*'s foi realizado o teste de *KMO*⁴ que avalia se a variável é adequada para *PCA*'s. Foi escolhido usar as variáveis com um valor superior a 0.6, pois as variáveis com valores inferiores a este limite podem não estar suficientemente correlacionadas para justificar o seu uso na criação dos *PCA*'s.

Para a utilização da *PCA*'s, foi necessário realizar um breve estudo da Irlanda para ter em conta quais eram as variáveis menos úteis, ou seja, aquelas que praticamente eram iguais para todos os estudantes. Após esta escolha, procedeu-se então para a escolha do número de *PCA*'s.

Primeiramente, aplicou-se o critério de *Kaiser*⁵, tendo-se obtido uma recomendação de 9 *PCA*'s. Pareceu ser um valor bastante elevado de *PCA*'s e, portanto, recorreu ao gráfico de “*elbow*” para conseguir, não só uma melhor análise, como também um valor mais preciso de *PCA*'s a utilizar. Abaixo é possível visualizar o gráfico obtido:

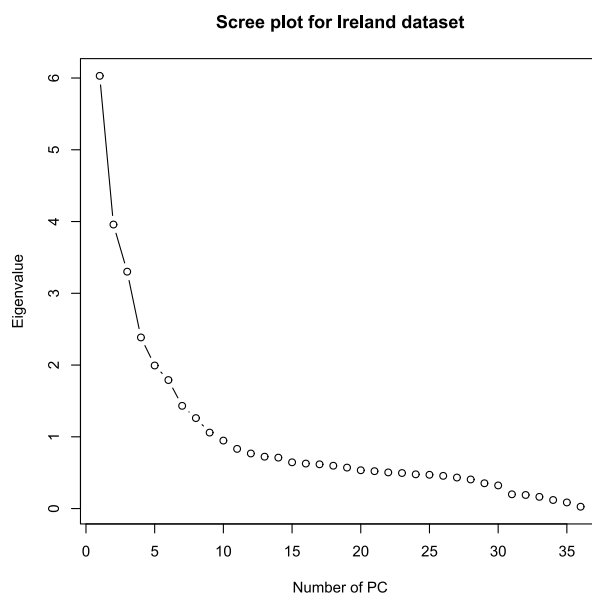


Figura 1 - "Elbow"

³ MCAR - Missing Completely at Random

⁴ KMO – “Kaiser-Meyer-Olkin”

⁵ Kaiser-Guttman - Regra utilizada na análise de componentes principais (*PCA*) para determinar o número de componentes principais a serem retidos. Os valores próprios devem ser acima de 1.



Analisando o gráfico com atenção e, também, após várias experiências, optou-se por utilizar 6 PCA's, pois era o que apresentava um maior valor de variáveis cumulativas.

Feita a análise, de seguida serão apresentadas as PCA's em si, com o respetivo nome e uma breve explicação da mesma, tendo, como extra, uma breve explicação de variáveis que poderiam ser úteis, mas não tinham influência nenhuma, ou variáveis interessantes de comentar. Em anexo, encontrar-se-ão as variáveis que as compõem,

RC1 - Sucesso Académico (SA)

Esta componente principal tem em conta os resultados da avaliação do estudante nos três campos de conhecimento (ciências, leitura e matemática) avaliados, bem como a dificuldade percebida pelo mesmo e, também, a sua percepção quer sobre dificuldade a do teste, quer sobre a sua competência.

Tabela 1 - Notas Escolares

Indicator	Sort	Ireland	OECD average
STUDENT PERFORMANCE IN READING			
Student performance in reading (mean score)	(2018) Download Indicator	518	487
STUDENT PERFORMANCE IN MATHEMATICS			
Student performance in maths (mean score)	(2018) Download Indicator	500	489
STUDENT PERFORMANCE IN SCIENCE			
Student performance in science (mean score)	(2018) Download Indicator	496	489

De forma geral, pode-se visualizar que os alunos da Irlanda se encontram mais preparados para a realização dos testes do PISA, apresentado valores melhores que a própria média da OECD.

Tabela 2 - ICT

Indicator	Sort	Ireland	OECD average
STUDENT'S ACTIVITIES OUTSIDE OF SCHOOL			
Average time per day students spend on the Internet outside of school during weekdays, in minutes	(2015) Download Indicator	144	146
Average time per day students spend on the Internet outside of school during weekends, in minutes	(2015) Download Indicator	185	184
ACCESS TO ICT			
Number of students per school computer	(2012) Download Indicator	2.59	4.66
Students browsing the Internet for schoolwork at school at least once a week (%)	(2012) Download Indicator	32.4	41.9
Students browsing the Internet for schoolwork outside of school at least once a week (%)	(2012) Download Indicator	45.4	54.9
Students who reported the use of computers in mathematics lessons during the month prior to the PISA test (%)	(2012) Download Indicator	17.6	31.6
Number of computers per student	(2015) Download Indicator	1	1

Visualizando os valores das componentes que se encontram em [anexo](#), é possível notar que existe uma diminuição por parte do uso das tecnologias à medida que as notas aumentam. A tabela acima demonstra uma possível relação, pois na Irlanda parece existir um menor incentivo associado ao uso de tecnologias nas escolas. Relativamente ao uso de tecnologias por lazer, o valor parece ser o mesmo que a média da OECD.



RC3 - Estatuto socioeconómico (ESE)

Esta componente principal tem em conta as posses do estudante em casa bem como o seu estatuto económico, social e cultural. Esta componente principal também tem em atenção os recursos que o estudante tem quer em casa, quer no que diz respeito a ICT por si só.

RC2 - Apoio por parte do professor (APP)

Esta componente principal tem em conta a perceção do estudante quer no que diz respeito ao interesse do professor, quer também no que diz respeito ao estímulo proveniente do mesmo para que os estudantes leiam. Esta componente principal também tem em conta a adaptabilidade do ensino do professor e apoio deste em aulas sobre línguas.

Tabela 3 - Número de aulas

Indicator	Sort	Ireland	OECD average
LEARNING TIME			
Learning time per week in foreign language lessons	(2018) Download Indicator	3	4
Learning time per week in regular language-of-instruction lessons	(2018) Download Indicator	3	4
Learning time per week in regular mathematics lessons	(2018) Download Indicator	3	4
Learning time per week in regular science lessons	(2018) Download Indicator	2	3

Acima é notório que na Irlanda existe um menor número de aulas quando comparado com a média da OECD.

RC4 - Alegria e bem-estar interior (ABEI)

Esta componente principal tem em conta a satisfação do estudante com a sua vida e o significado da mesma para si próprio. Esta componente principal também tem em conta os sentimentos positivos do estudante, sendo estes sentimentos como pertença, conforto com o seu corpo e a sua aparência, ausência de receio de falhar, resiliência e bom relacionamento com os pais.

Tabela 4 - Competição

Indicator	Sort	Ireland	OECD average
SCHOOL CLIMATE			
Index of student competition (average=0, std.deviation=1)	(2018) Download Indicator	0.20	-0.01
Index of student co-operation (average=0, std.deviation=1)	(2018) Download Indicator	-0.17	0.00
STUDENT WELL-BEING			
Index of students' positive feelings (average=0, std.deviation=1)	(2018) Download Indicator	-0.09	0.01
STUDENTS' GLOBAL COMPETENCE			
Index of students' cognitive adaptability	(2018) Download Indicator	0.11	-0.01

Por curiosidade, na Irlanda os alunos aparentam ser muito mais competitivos, e menos cooperativos, quando comparados com a média da OECD. Por consequência, apresentam-se com sentimentos menos positivos, mas com uma capacidade de adaptação superior.



RC5 - Nível de educação e informação sobre carreiras (NEIC)

Esta componente principal tem em conta o nível de educação do estudante e a informação que este tem sobre carreiras a seguir. No que diz respeito à informação que o estudante tem sobre carreiras a seguir, tem-se em conta a informação que o mesmo tem em geral e, também, a informação sobre este tema que lhe foi fornecida pela escola.

RC6 – Mudanças na educação (ME)

Esta componente principal tem em conta o total de vezes que o estudante mudou de escola, contando as possíveis consequências desta mudança.

Nota: Para a criação dos *PCA*'s, também se utilizaram outras técnicas de rotação, tal como o *varimax*, *quartimax*, *oblimin*, etc..., tendo-se chegado à conclusão de que a melhor foi o *varimax*.

Clusters

Nesta secção falar-se-á sobre os *clusters* criados e mostrar-se-ão os mesmos tendo em conta as *PCA*'s criadas e referidas acima.

Inicialmente, optou-se por usar o método hierárquico Ward D2 (linkagem de Ward) um método que se considera ser bastante sólido e eficaz, pois este método minimiza a soma dos quadrados das diferenças dentro dos *clusters*, possuindo uma tendência em formar *clusters* de tamanho similares, sendo sensível à forma dos *clusters*.

Após não só a análise do gráfico, como também dos resultados obtidos de vários testes diferentes, chegou-se à conclusão de que deveriam ser utilizados três *clusters*. A divisão feita está representada na Figura abaixo.

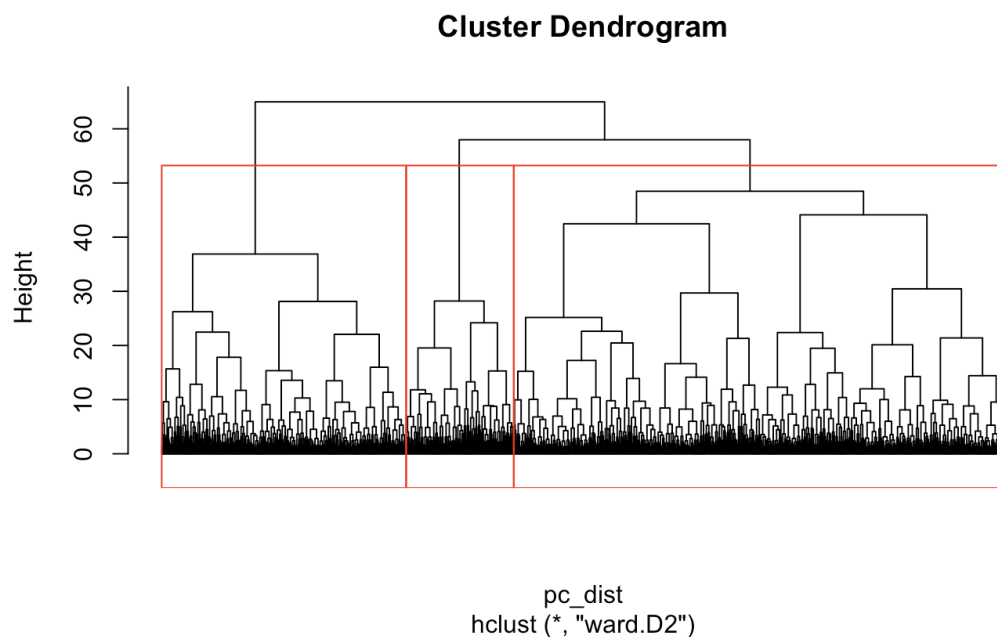


Figura 2 - Clusters Ward.D2



Fazendo uma análise crítica, é notável a existência de 3 grupos distintos, existindo um muito superior, em termos de volume, que os outros. Para demonstrar os resultados relativamente a estes *clusters*, é necessário visualizar a *silhouette* que irá relatar a qualidade de cada um dos 3 *clusters*.

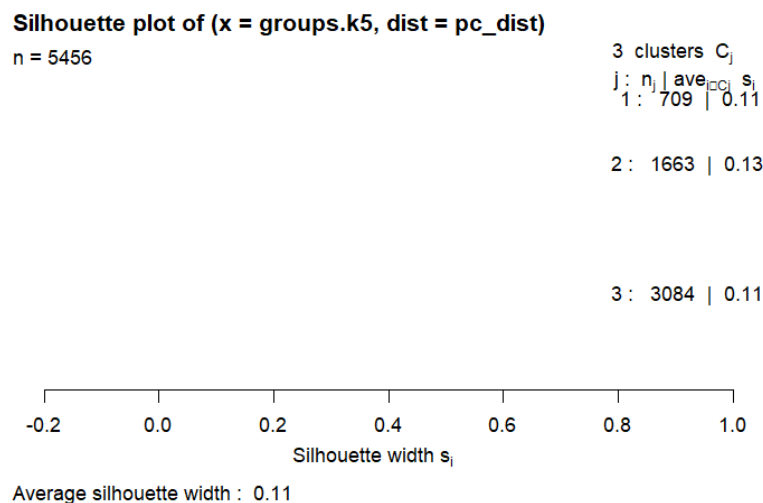


Figura 3 - Silhouette

Na figura acima, pode-se reparar que as qualidades dos *clusters* não foram muito positivas. Tendo uma amostra de 5456 estudantes: no *cluster* 1 encontram-se 709 indivíduos, tendo um ajuste de 0.11; no *cluster* 2 encontram-se 1663 indivíduos, tendo um ajuste de 0.13; no *cluster* 3 encontram-se 3084 indivíduos, tendo um ajuste de 0.11. Os resultados aqui apresentados não são muito agradáveis e, para comprovar isso, podemos interpretar a largura média da *silhueta* que apresenta um valor muito próximo de 0, ou seja, uma grande ambiguidade na separação dos *clusters*.

Devido a estes resultados tão baixos, optou-se por tentar visualizar qual seria o melhor número de clusters possível. De seguida, é possível visualizar os resultados obtidos:

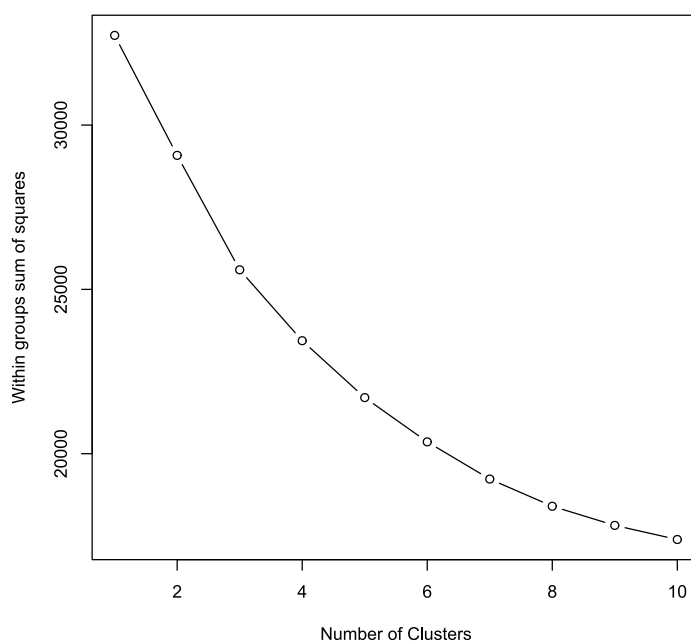


Figura 4 - Número de Clusters recomendado



Este gráfico corrobora a hipótese já apresentada de 3 *clusters*. Tendo isto em conta, continuar-se-á com a escolha de 3 *clusters*.

De seguida, foi feita uma análise utilizando o método “*complete linkage*”, método que calcula a distância entre dois grupos, como a maior distância entre quaisquer dois pontos dos grupos. Ele tende a formar *clusters* compactos e bem separados, mas pode ser sensível a *outliers*. Os resultados foram inferiores ao *Ward.D2* e, devido a isso, a explicação deste encontra-se em [anexo](#).

Feita a análise do número de *clusters* recomendados para K-means, criaram-se estes 3 *clusters* recomendados e fez-se uma comparação entre estes 3 *clusters* com os do *Ward.D2* e os resultados foram os seguintes:

Tabela 5 - Table Clusters

Group.k2	Cluster 1 – K.means	Cluster 2 – K.means	Cluster 3 – K.means
Cluster 1 - <i>Ward.D2</i>	33	85	591
Cluster 2 - <i>Ward.D2</i>	1453	130	80
Cluster 3 - <i>Ward.D2</i>	337	2528	219

O desejável seria visualizar os maiores valores na diagonal, mas não é o que acontece. Analisando a tabela, podemos visualizar que maior parte dos estudantes associados ao *Cluster 3 – Ward.D2* pertencem ao *Cluster 2* do K-means. Enquanto no *Cluster 2 – Ward.D2* maior parte dos estudantes estão associados ao *Cluster 1* do K-means.

Para então entender melhor como são os estudantes de cada um dos *clusters* criados, visualizar-se-ão as respetivas características de cada um deles, tendo em conta as *PCA*’s criadas anteriormente.

Cluster 1 → Trabalhadores-Estudantes

Cluster 1 - Average score in each principal component

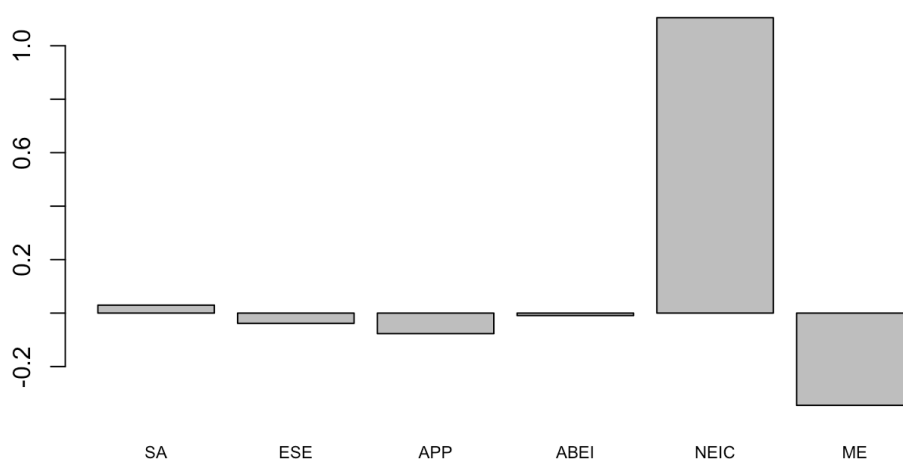


Figura 5 - Clusters dos Estudantes Trabalhadores

Analisando este *cluster* detalhadamente, é possível verificar que aqui encontram-se os estudantes que têm uma maior preocupação com o seu futuro, não tendo realizado muitas mudanças de escola.

De forma mais detalhada, são os alunos que possuem uma maior informação sobre as carreiras (INFOCAR), sobre o mercado de trabalho fornecidas pela própria escola (INFOJOB1), tendo um maior



conhecimento em cultura geral (ST001D01T). Por outro lado, nota-se falta de mudanças escolares que ocorreram na vida destes estudantes, ou seja, estes não realizaram muitas transições significativas relativamente à sua educação ao longo do tempo, como mudanças de escola, mudanças de nível educacional, mudanças de programa ou até de curso (CHANGE). Consequentemente. Também há a tendência de não mudarem de escola (SCCHANGE).

Pode-se então comentar que, possivelmente, se tratam de alunos-trabalhadores que têm uma maior noção da realidade e querem já ficar fixos num local para conseguir emprego e construir o seu futuro. Segundo o próprio site da OECD [Better Life Index](#), “cerca de 68% das pessoas com idades entre 15 a 64 anos na Irlanda têm emprego remunerado” demonstrando assim que existem estudantes de 15 que já trabalham. Algo que também pode intensificar esta variável é: “Os jovens com idades entre 15 e 24 anos na Irlanda, entretanto, enfrentam dificuldades com uma taxa de desemprego de 33,0% em comparação com a média da OCDE de 16,3%.”, podendo dar a entender que os alunos estão preocupados com o futuro das suas carreiras. O próprio site da [EU](#) comprova que é possível estudantes de 15 anos serem remunerados.

Não obstante, caso não sejam trabalhadores-estudantes, podem ser estudantes que realizaram cursos de verão ([Website de exemplo](#)) e, sendo estes remunerados, podem mudar os pontos de vista dos estudantes, fazendo com que estes se preocupem mais com o futuro.

Cluster 2 -> Estudantes “Despreocupados”

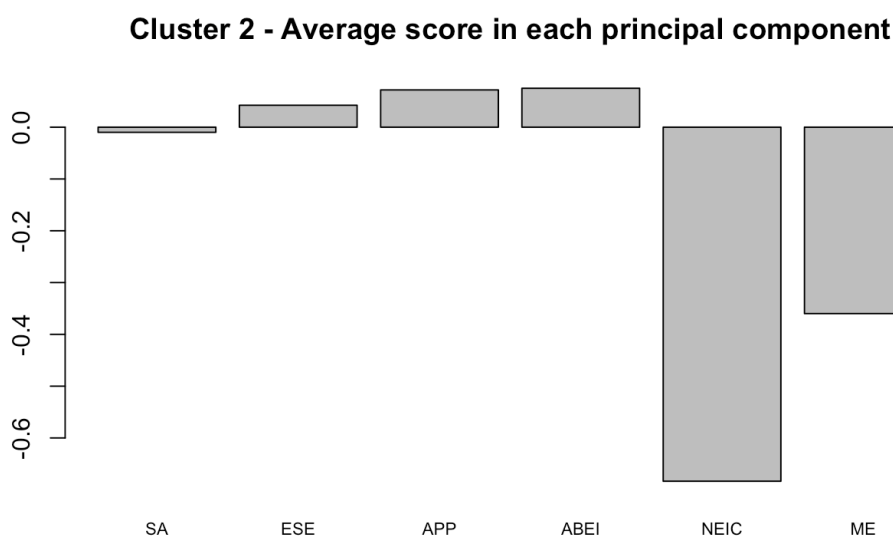


Figura 6 - Estudantes Despreocupados

Analisando este *cluster*, dá para notar que os estudantes que constituem este *cluster* são aqueles mais despreocupados com o futuro e que raramente ou nunca mudaram de escola. Inicialmente pensou-se que eram os alunos que não tinham como prioridade a escola, preferindo outras atividades, mas é importante relembrar que os resultados obtidos nos testes do PISA foram acima da média quando comparados com a média da OECD.

Tendo esta ideia em mente, surgiram várias hipóteses, como alunos cujos pais tinham posses e, devido a isso, não se preocupavam com o futuro, ou alunos que eram inteligentes, mas que preferiam realizar atividades extracurriculares.



Por curiosidade, a fim de se verificar se seria possível obter mais respostas, criou-se um gráfico a fim de visualizar como este *cluster* se comportava quando era comparada com o maior grau de escolaridade dos pais, e os resultados obtidos foram os seguintes:

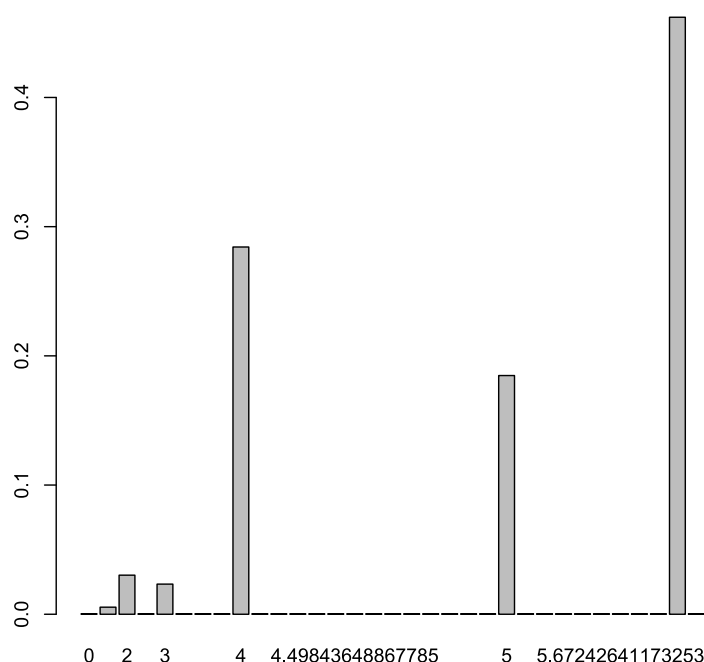


Figura 7 - Estudantes Despreocupados VS grau de ensino mais alto dos pais

O gráfico acima demonstra que existem vários casos diferentes, mas existe em maior número estudantes que possuem pais com um elevado grau de ensino. O grau de ensino dos pais pode ter alguma influência no seu rendimento, dando aos respetivos pais um maior prestígio. Este meio pode não contribuir positivamente para um entendimento da importância de estudar para os jovens estudantes, contribuindo para que estes não deem uma grande importância para o seu futuro.

Cluster 3 -> Alunos “Estrangeiros”

Cluster 3 - Average score in each principal component

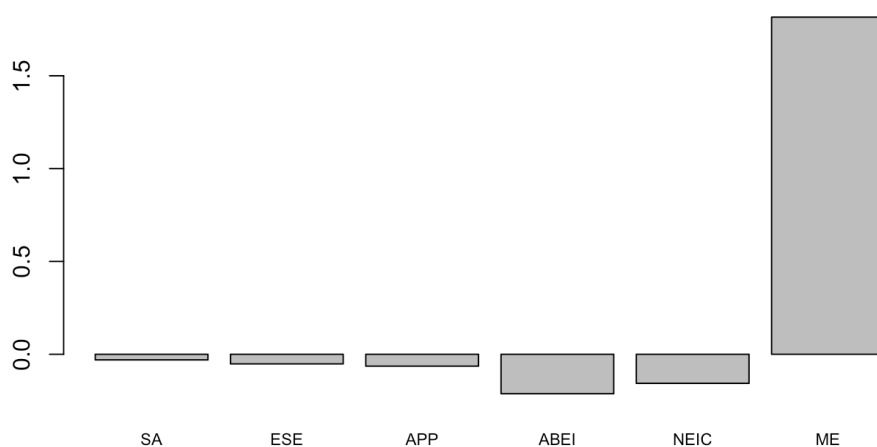


Figura 8 - Alunos que mudam muitas vezes de escola



Observando este gráfico, é notório que este *cluster* retrata os casos de estudantes que mudaram de escola e de programa escolar tantas vezes que muito provavelmente não se conseguem adaptar ao meio em que se encontram. Curiosamente, é de se reparar que estas mudanças excessivas têm alguma influência na felicidade do estudante.

De seguida, surgiu a dúvida se este *cluster* era apenas formado por emigrantes e, para responder a esta pergunta, foi criada um gráfico para visualizar as respetivas características dos estudantes associadas à emigração. O respetivo gráfico encontra-se abaixo:

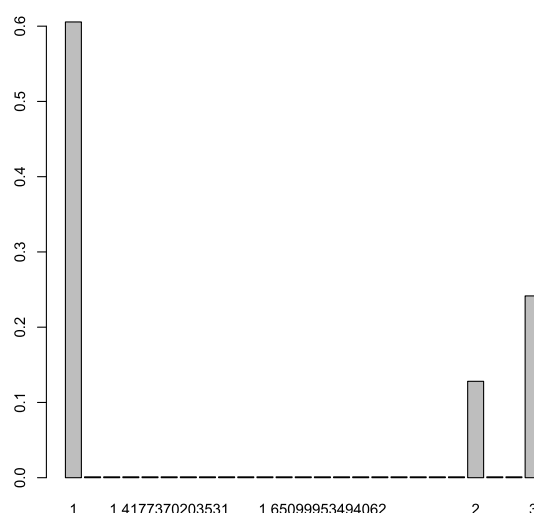


Figura 9 - Alunos que mudam muitas vezes de escola VS Imigrantes

Analisando o gráfico, é possível visualizar que a maioria realmente não é emigrante, mas também existem estudantes que mudaram de escola várias vezes que são emigrantes. Concluiu-se que, relativamente a este *cluster*, só eram realmente fornecidos dados sobre se os estudantes mudavam de escola muitas vezes, não tendo conseguido recolher nenhum conhecimento extrínseco ou intrínseco sobre tal.

Model Selection

Ao lado, é possível visualizar o PAM⁶ *clustering*, sendo muito difícil retirar conclusões dele. Primeiramente, é importante realçar que nele estão contidos 3 clusters, sendo muito difícil de identificar os diferentes medoides.

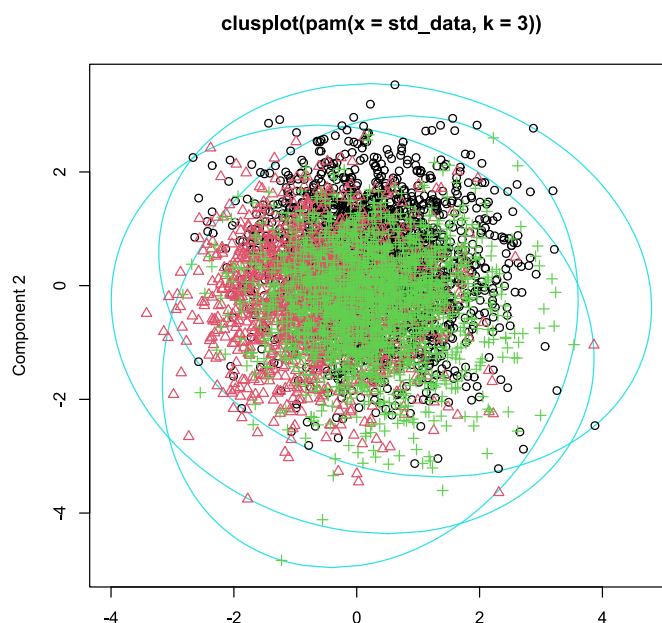


Figura 10 - PAM

⁶ PAM - Partitioning Around Medoids

De seguida, serão analisados os respetivos *clusters* de GMM

Antes de realizar o estudo relativamente aos *clusters* do GMM⁷, é necessário ter em atenção ao BIC⁸, sendo necessário testar diferentes números de componentes, a fim de identificar o número que proporciona o melhor equilíbrio entre ajuste e complexidade, ajudando a escolher a estrutura de cluster mais adequado. Após algumas análises e testes, chegou-se à conclusão de que seriam usados 2 componentes. Abaixo, é possível visualizar o resultado:

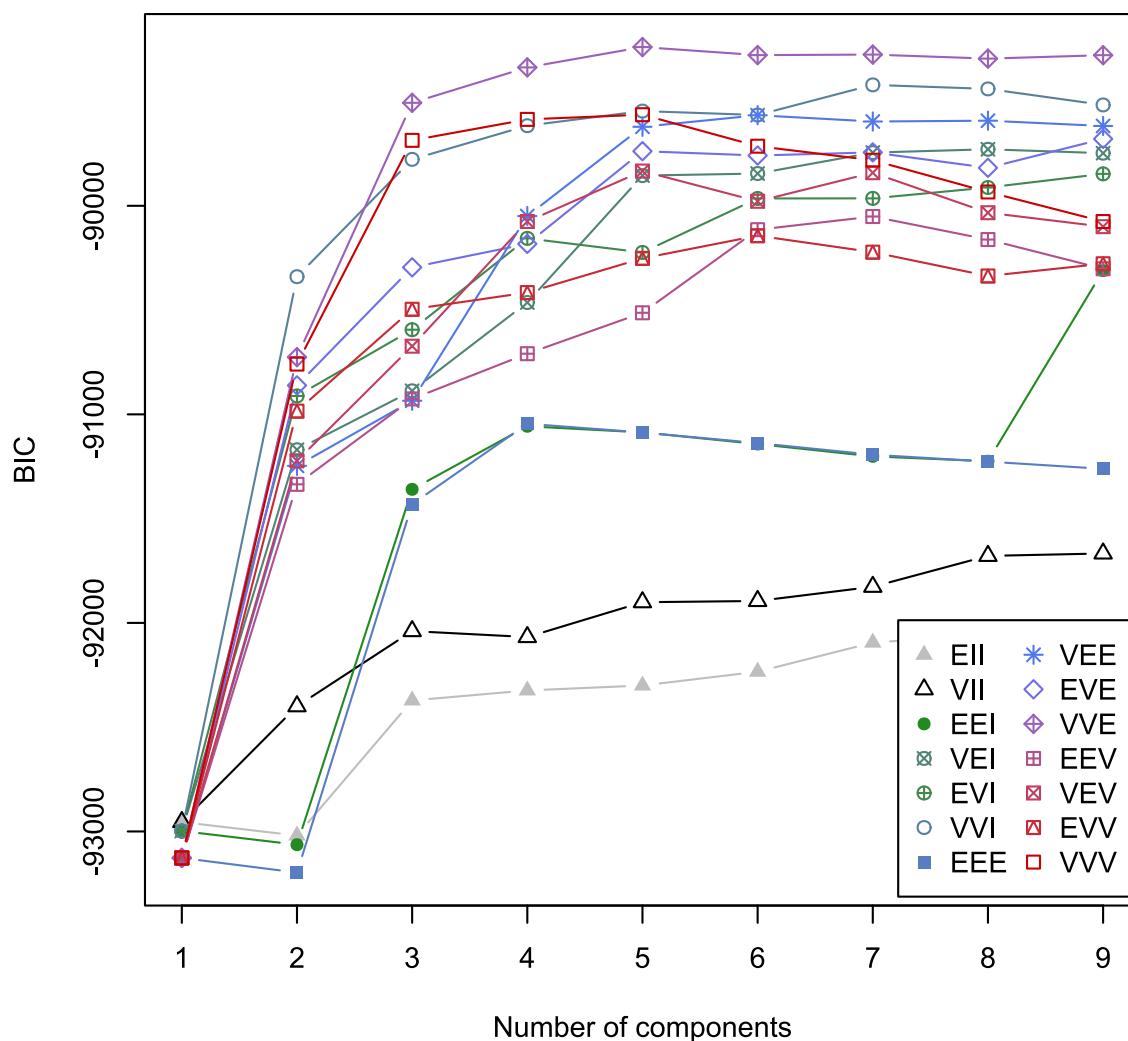


Figura 11 - Componentes

Tendo obtido um valor de BIC correspondente a -89505.96, tivemos em conta que o melhor modelo ótimo seria o VVE, isto é, (Volume variável, forma variável, orientação igual). Com isto já temos noção que os clusters terão diferentes formas e volumes, tendo as orientações semelhantes.

⁷ GMM - Gaussian Mixture Model

⁸ BIC - Bayesian Information Criterion



No que diz respeito aos *clusters* obtidos através de GMM, teve-se especial atenção no que diz respeito à representação gráfica quer da classificação quer da incerteza da classificação, visto que mostra, da melhor forma, a dificuldade de obtenção de *clusters*. Abaixo estão os gráficos de classificação e de incerteza para cada classificação, respetivamente.

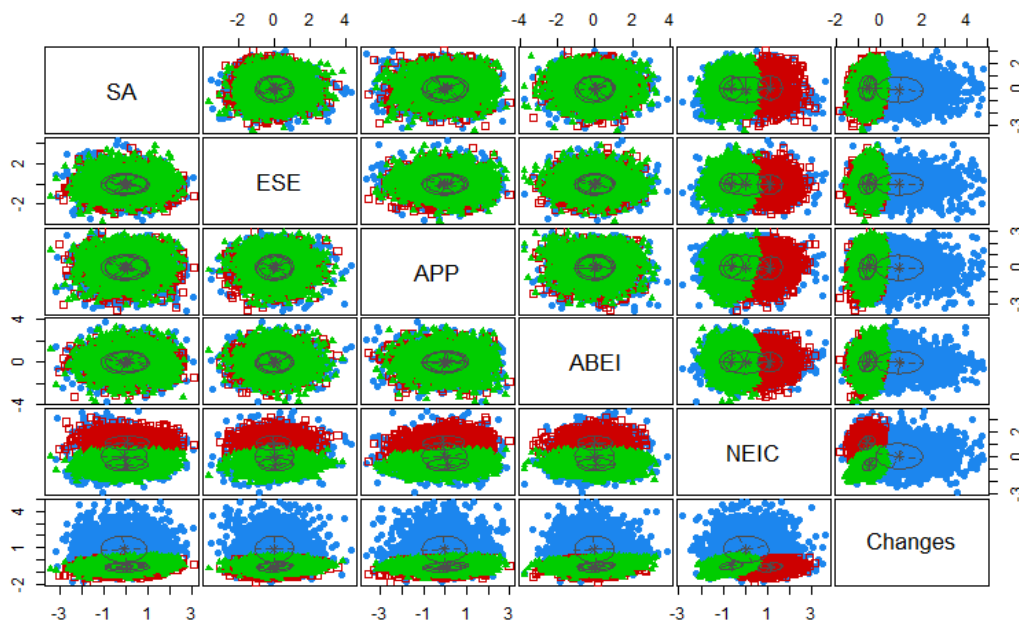


Figura 12 - GMM Classificação

Já tendo em conta que foram escolhidos 3 clusters, é notório que existe, de forma geral, muita sobreposição, sendo exceção os PCA's [ME](#) e [NEIC](#), existindo alguma consistência relativamente a estes, especialmente no PCA [ME](#).

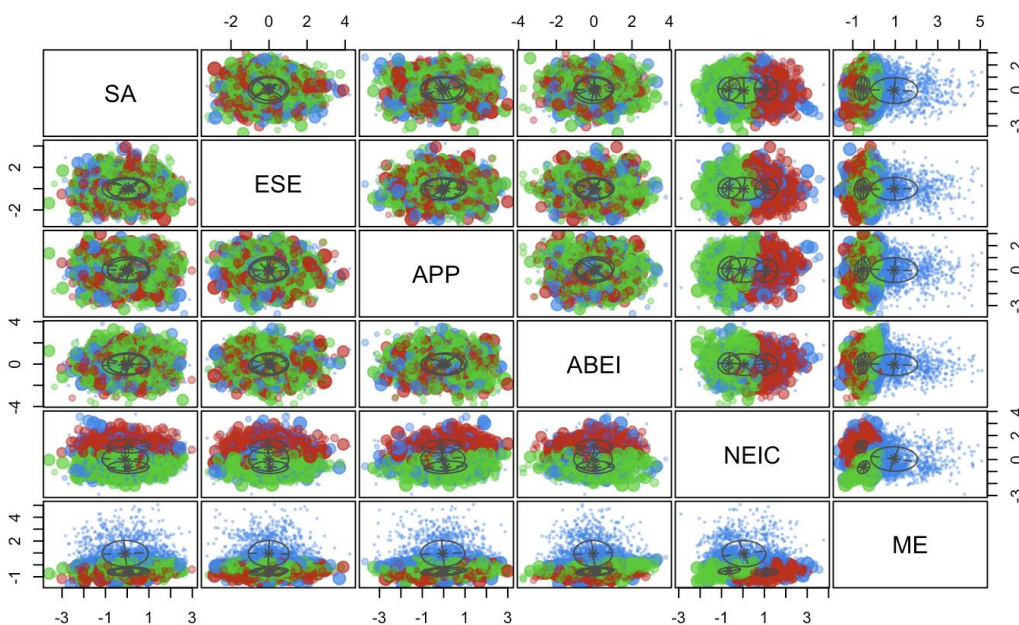


Figura 13 - GMM Incerteza



Relativamente à incerteza, é notório que esta é muito elevada. Como já se tinha referido anteriormente, os diferentes clusters encontram-se muito sobrepostos, mas agora com as suas respectivas bolas maiores, demonstrando que a incerteza desse estudante pertencer a esse cluster é muito elevado. Já era esperado, mas o *PCA ME* é único que possui algum cluster com uma incerteza baixa, correspondendo ao cluster de cor azul, não existindo uma incerteza muito elevada. De forma geral, concluímos que os clusters não foram muito bem definidos, possuindo sobreposições e incertezas muito fortes.

Conclusão

Concluindo, esta análise intensiva dos estudantes da Irlanda, que realizaram o teste do PISA, demonstrou que estes, de uma forma geral, são bastante homogêneos. Ao longo do trabalho foram realizadas várias decisões, consoante os resultados dos diferentes testes, que levaram a acreditar que 3 *clusters* seria a melhor escolha.

Os *PCA's* que aparentavam ter um maior peso para a classificação dos estudantes, correspondendo aqueles que possuíam um maior valor de *loading*, não foram muito produtivos em termos de conseguir diferenciar os diferentes *clusters*, sendo necessário referir o *PCA* associada às mudanças escolares dos estudantes que, aparentemente, foi aquele onde, não era só mais fácil de visualizar a separação dos diferentes clusters, como também era o que apresentava uma incerteza menor, dando a entender que os clusters tiveram muito em conta este *PCA*.

É necessário ressaltar que para esta análise de comunidades, foi usada os dados com imputação, pois o grupo acreditava que com um maior número de estudantes poderia aproximar-se mais à realidade dos estudantes da Irlanda, mesmo que a diferença fosse mínima. Mesmo assim, quando comparado os resultados entre com e sem imputação, os resultados foram muito parecidos, e até por vezes iguais.

Uma coisa muito importante a relatar é que, realmente, é a partir dos 15 anos que o interesse pelo seus respectivos futuros académicos aumenta e, prova disso, é perceber como a própria [educação da Irlanda](#) funciona, pois, “(...) preferem aceitar estudantes internacionais com 15 anos de idade para o programa do Ano de Transição. O Ano de Transição é um programa educacional de um ano que ajuda a criança a decidir sobre sua futura carreira. Não é um ano de exames, por isso é mais fácil para um estudante estrangeiro se adaptar.”

Indicator	Sort	Ireland	OECD average
PARTICIPATION IN EDUCATION			
Percentage of students enrolled in private institutions at pre-primary level (%)	(2020) Download Indicator	99.2	32.8

Um ponto a ressaltar, é que o as pré-primárias na Irlanda são privadas, não existindo creches, desta forma, fazendo com que os estudantes possam ter costumes muito parecidos. Também é importante comentar que na Irlanda a religião é um tema muito importante, tendo em conta que a maioria das escolas são católicas, chegando ao ponto de existir escolas que separam as meninas dos meninos.

Apesar disso, depois da creche o número de escolas privadas tende a diminuir, para mais informação poderemos analisar o seguinte [link](#).

Concluimos assim que a Irlanda é um país desenvolvido, possuindo uma educação forte e enriquecedora que contribui para a qualidade geral dos estudantes, não existindo estudantes “outliers”.



Webgrafia

Education GPS - Ireland - Student performance (PISA 2018). (n.d.). Gpseducation.oecd.org. Retrieved June 15, 2023, from <https://gpseducation.oecd.org/CountryProfile?primaryCountry=IRL&treshold=10&topic=PI>

Education GPS - Ireland - Student performance (PISA 2018). (n.d.). Gpseducation.oecd.org. <https://gpseducation.oecd.org/CountryProfile?primaryCountry=IRL&treshold=10&topic=PI>

FactsMaps. (2019, December 5). *PISA 2018 Worldwide Ranking - average score of mathematics, science and reading*. FactsMaps. <https://factsmaps.com/pisa-2018-worldwide-ranking-average-score-of-mathematics-science-reading/>

The Programme for International Student. (n.d.). https://www.oecd.org/pisa/publications/PISA2018_CN_IRL.pdf

OECD Better Life Index. (n.d.). Wwww.oecdbetterlifeindex.org. <https://www.oecdbetterlifeindex.org/pt/paises/ireland-pt/>

Cursos de verão de inglês para crianças e jovens em Irlanda. (n.d.). Wwww.languageinternational.pt. <https://www.languageinternational.pt/cursos-ingles-irlanda-junior>

Poradnuk. (2022, August 31). *Estudar na Irlanda*. Poradnuk.com.ua; Poradnuk. <https://poradnuk.com.ua/pt/educa%C3%A7%C3%A3o-na-irlanda.htm>

Sistema Escolar Irlandês - Kiwi - Irlanda. (2021, July 16). Kiwi - Irlanda - a Kiwi Education é Uma Agência Educacional Reconhecida Pelo Governo E Líder No Mercado Educacional Da Nova Zelândia. <https://kiwieducation.com.br/ie/study/sistema-escolar-irlandes/>

[PISA in Focus-n7 \(port\).indd \(oecd.org\)](#)



Anexos

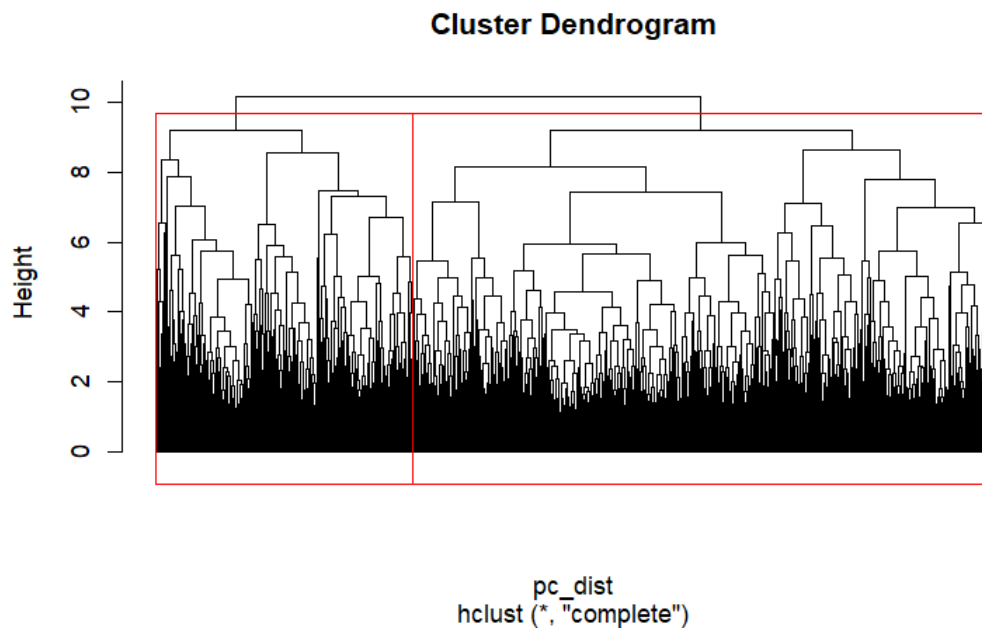
Aqui encontram-se dados extras sobre o trabalho, como curiosidades e experiências extras que não correram tão bem, mas valem a pena ser comentadas e demonstradas.

Variáveis dos PCA's

Loadings:							
	RC1	RC3	RC2	RC4	RC5	RC6	
PV1SCIE	0.833	0.123		-0.119		-0.183	
PV1READ	0.853	0.113		-0.114		-0.185	
PV1MATH	0.765	0.170				-0.213	
SOCONPA		0.128	0.136	0.483			
INFOJOB1	-0.101				0.766		
INFOCAR					0.792		
HOMESCH	-0.154	0.228	0.156		0.324	0.201	
BELONG				0.628			
MASTGOAL	0.291	0.156	0.237	0.370	0.119	0.137	
RESILIENCE	0.178		0.102	0.668			
SWBP				0.692			
EUDMO	-0.148		0.104	0.678		0.103	
GFOFAIL				-0.444	0.119		
WORKMAST	0.273	0.105	0.186	0.416	0.147	0.151	
PISADIFF	-0.699						
SCREADDIFF	-0.639			-0.162			
SCREADCOMP	0.723		0.121	0.112			
JOYREAD	0.601	0.101	0.129			0.159	
TEACHINT	0.141		0.734				
ADAPTIVITY			0.744	0.114			
STIMREAD	0.137		0.762				
PERFEED			0.669				
DIRINS			0.743	0.102			
TEACHSUP			0.770	0.113			
ICTRES		0.806					
WEALTH		0.808					
HEDRES	0.109	0.635	0.139	0.153			
CULTPOSS	0.285	0.564	0.109			0.116	
HOMEPOS	0.172	0.932					
ICTHOME		0.629					
ESCS	0.298	0.772					
CHANGE	-0.101					0.891	
SCCHANGE						0.878	
ST016Q01NA				0.750			
HISEI	0.276	0.452					
ST001D01T			-0.105		0.759		
	RC1	RC3	RC2	RC4	RC5	RC6	
SS loadings	4.423	4.308	3.548	3.277	2.006	1.894	
Proportion Var	0.123	0.120	0.099	0.091	0.056	0.053	
Cumulative Var	0.123	0.243	0.341	0.432	0.488	0.540	
PV1SCIE							
PV1READ							
PV1MATH							
SOCONPA							
INFOJOB1							
INFOCAR							
HOMESCH							
BELONG							
MASTGOAL							
RESILIENCE							
SWBP							
EUDMO							
GFOFAIL							
WORKMAST							
PISADIFF							
SCREADDIFF							
SCREADCOMP							
JOYREAD							
TEACHINT							
ADAPTIVITY							
STIMREAD							
PERFEED							
DIRINS							
TEACHSUP							
ICTRES							
WEALTH							
HEDRES							
CULTPOSS							
HOMEPOS							
ICTHOME							
ESCS							
CHANGE							
SCCHANGE							
ST016Q01NA							
HISEI							
ST001D01T							



Método “Complete”



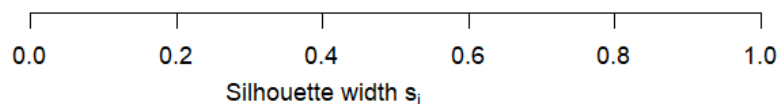
Fazendo uma análise não tão crítica, é possível visualizar que usando o método “Complete” era recomendado a criação de 2 clusters. Também é notável que os clusters são um pouco desequilibrados, dando a entender que o maior, possivelmente, possui quase $\frac{2}{3}$ da amostra. De seguida, será possível visualizar a respetiva *silhouette*.

Silhouette plot of (x = groups.k5_c, dist = pc_dist)

n = 5456

2 clusters C_j
 $j : n_j \mid \text{ave}_{i \in C_j} s_i$
 1 : 1678 | 0.004

2 : 3778 | 0.11



Average silhouette width : 0.08

Observando atentamente os resultados, é notório que estes não são muito positivos, sendo estes inferiores àqueles obtidos tendo usado o método *Ward.D2*. De seguida, iremos então comparar os clusters destes 2:



Groups.k5/Groups.k5_c	Cluster 1 - Complete	Cluster 2 - Complete
Cluster 1 – Ward.D2	293	218
Cluster 2 – Ward.D2	602	1032
Cluster 3 – Ward.D2	782	2528

Observando a tabela, esta não nos apresenta quase nenhuma informação extra, sendo apenas notório que existe um maior número de estudantes do cluster 3 do Ward.D2 em cada um dos clusters do método Complete.

Breve Estudo da Irlanda

Relativamente a este tópico do trabalho, aqui encontra-se uma breve pesquisa realizada pelo grupo para conseguir ter umas boas bases para começar a trabalhar com os dados da Irlanda, sendo necessário visualizar vários dados que a OECD continha.

De seguida, poderá ser lido os dados recolhidos, cada um destes separados pelos seus respetivos tópicos:

ICT

[Education GPS - Computers, education & skills \(oecd.org\)](https://www.oecd.org/education/gps/computers-education-skills/)

- Os alunos em leitura, tanto digital como física, ficaram acima da média da OECD
- Existe muito menos estudantes, quando comparado com a média da OECD, que esperam trabalhar como ICT profissionais
- De forma geral, existe um menor uso de ICT na educação da Irlanda, estando abaixo dos valores médios da OECD, mas, apesar disso, de forma geral, todos os alunos têm acesso a um computador

Organização e governo

[Education GPS - Organisation & Governance \(oecd.org\)](https://www.oecd.org/education/gps/organisation-governance/)

- Na Irlanda, o número de chumbos é muito menor que a média da OECD e, somado a isso, estes possuem mais apoio e suporte, facilitando a vida académica dos alunos repetentes;
- A Irlanda possui uma característica muito interessante, quando comparada com os outros países da OECD, quase 100% dos alunos tiveram em uma pré-primária privada. A educação da Irlanda encontra-se em sexto lugar mundialmente, sendo muito prestigiada ([LINK](#));
- Na Irlanda, quando um aluno frequenta uma escola que utiliza agrupamento por habilidades para criar as turmas, levando em consideração o perfil socioeconómico dos alunos e das escolas, houve um aumento de 2 pontos na pontuação da leitura, enquanto na média da OECD diminuiu 3 pontos;
- Relativamente à mudança de pontuação na leitura nos alunos repetentes diminuiu, tanto na Irlanda, como na média da OECD, quer tendo em atenção o perfil socioeconómico do aluno, mas esta diminuição é muito menos impactante na Irlanda.
- De forma muito interessante, os alunos possuem menos horas de estudo nas três áreas de teste do PISA quando comparado com a média dos países da OECD (Menos 1 hora em todas).



Students' Well-Being

[Education GPS - Students' Well-Being \(oecd.org\)](https://www.oecd.org/education/gps/students-well-being/)

- Na Irlanda, os estudantes têm um nível de pertença abaixo da média da OECD (mesmo este sendo muito baixo, -0.15)
- Existe uma menor percentagem de alunos que referem estar tristes (5%), sendo este valor 1.5 pontos percentuais menores que a média da OECD e, somado a isto, em média os alunos da Irlanda sentem-se mais insatisfeitos com a vida quando comparado com a média da OECD. Consequentemente, estes possuem uma maior ansiedade a cometer erros, por outras palavras, um maior medo de fracassar quando comparado com a média da OECD.
- As raparigas possuem um maior interesse em aprender matéria relacionada com ciências, sendo este oposto do valor da média do OECD. De forma geral. Na Irlanda existe um maior interesse para aprender matéria sobre ciência
- Na Irlanda, existe um impacto superior em relação à interação professor-aluno para aprender matemática.
- Tendo já em conta que os estudantes da Irlanda possuem uma maior insatisfação da vida, quando comparado com a média da OECD, não existe uma grande diferença na nota de Ciências destes;
- Curiosamente, estudantes irlandeses têm um nível razoável de motivação (0.39 à média da OECD) intrínseca para não só para atingir objetivos académicos, como também para se esforçarem a fim de obterem um bom desempenho.
- Os estudantes com um maior nível de ansiedade tendem a ter uma menor satisfação com a vida na Irlanda (sendo esta afirmação mais notável na Irlanda que na média da OECD)
- De forma geral, na Irlanda os alunos tendem a ter uma maior ansiedade que a média da OECD. Maior parte dos alunos continuam a ter muita ansiedade mesmo que bem preparados para um teste (63.2%, superior à média da OECD). Interessantemente, existe uma maior ansiedade por parte dos rapazes;
- De forma geral, existe mais alunos na Irlanda que sentem que não pertencem ao clima da escola. (51.4%)
- Relativamente ao Bullying, na Irlanda existe uma exposição menor quando comparado com a média da OECD, mesmo quando comparado com escolas privilegiadas ou não, não havendo quase impacto nenhum nas notas quando comparado as escolas com maioríssimas incidências de bullying e menor.
- Existe uma diferença significativa na satisfação de vida entre estudantes do quartil inferior e superior de riqueza. Apesar disso, esta insatisfação é menor que a média da OECD.
- Os estudantes irlandeses praticam mais desporto antes/depois da escola que a média da OECD.
- Existe uma maior insatisfação de vida nos estudantes que usam mais internet.

Teachers

[Education GPS - Teachers \(oecd.org\)](https://www.oecd.org/education/gps/teachers/)

- Em média, há cerca de 10.43 alunos para cada professor nas escolas socioeconomicamente desfavorecidas, sendo este valor inferior à média da OECD (11.22);
- Em média, há cerca de 13.421 alunos para cada professor nas escolas socioeconomicamente favorecidas, sendo este valor superior à média da OECD (12.43);



Socio-economic status

[Education GPS - Socio-economic status \(oecd.org\)](https://www.oecd.org/education/gps/socio-economic-status/)

- Na Irlanda, 9% dos alunos socioeconomicamente desfavorecidos repetiram um ano letivo, uma percentagem muito menor quando comparado com a média da OECD, de 20.3%.
- Na Irlanda, 4.5% dos alunos socioeconomicamente favorecidos repetiram um ano letivo, uma percentagem menor quando comparado com a média da OECD, de 5%
- A Irlanda tem uma disparidade socioeconômica, comparativamente aos alunos do quartil superior e inferior do índice ESCS do PISA, na satisfação com a vida abaixo da média da OECD, sendo esta $7.93 - 5.89 = 2.04$
- Os alunos irlandeses do quartil inferior do índice ESCS do PISA, possuem um senso de pertencimento inferior aos do quartil superior, sendo esta diferença menos gritante quando comparada com a média da OECD.
- Os alunos não imigrantes têm uma maior propensão a se sentirem pertencentes à escola em comparação com os alunos imigrantes de primeira geração. Esta diferença é superior à média da OECD.
- Os estudantes irlandeses do quartil superior do índice PISA ESCS geralmente correspondem a estudantes com maior vantagem socioeconômica, tendo uma maior probabilidade de relatar que os seus pais estão interessados nas suas atividades escolares, em comparação com os estudantes do quartil inferior, que geralmente correspondem a estudantes com menor vantagem socioeconômica, ocorrendo o mesmo quando os estudantes passam por dificuldades. Mesmo assim é importante referir que esta diferença é muito inferior quando compara com a média da OECD