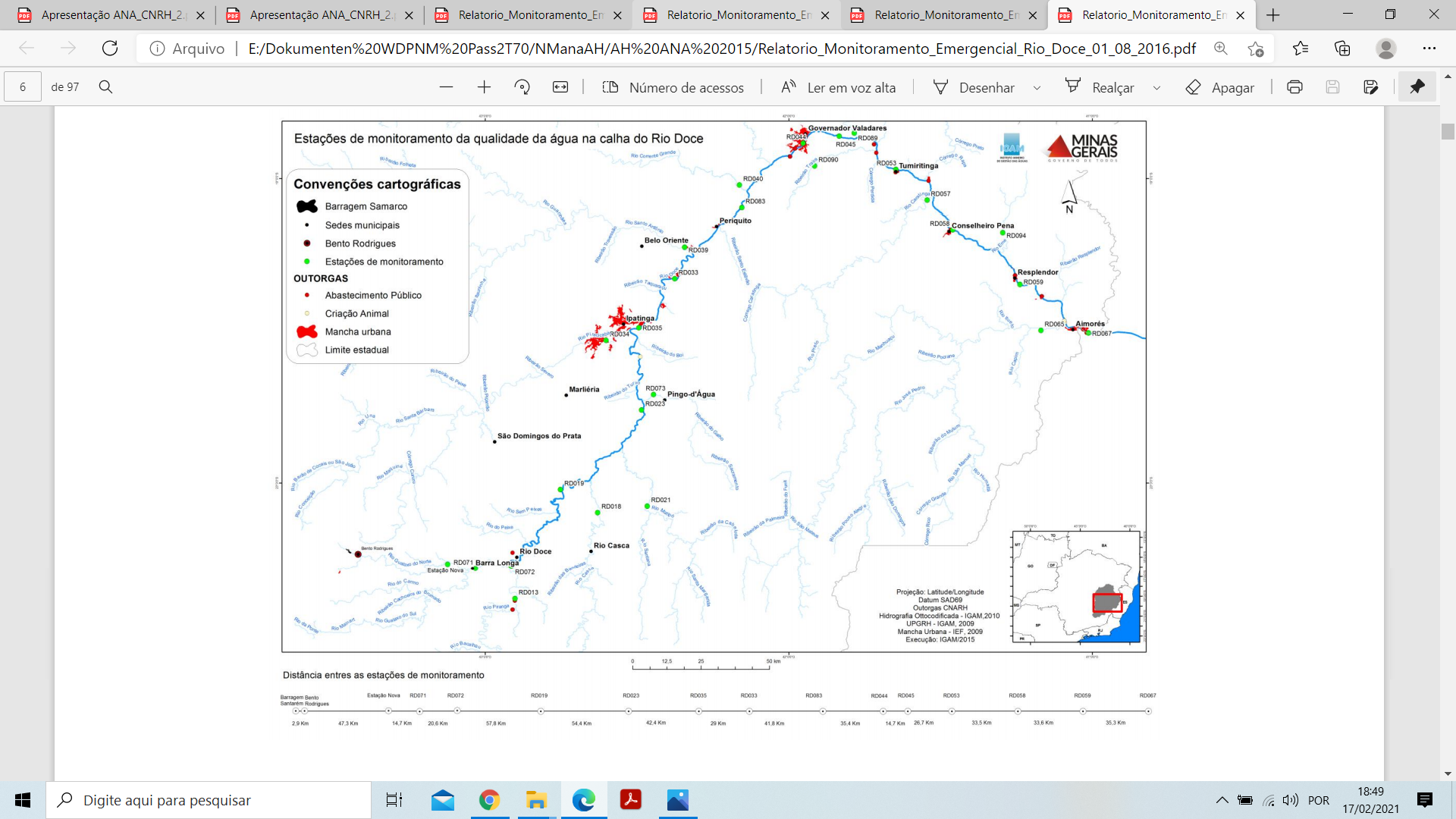
O Dataset escolhido para o trabalho foi o monitoramento da qualidade das águas na bacia hidrográfica do Rio Doce realizado pelo IGAM - Instituto Mineiro de Gestão das Águas, por meio do Programa Águas de Minas que deu início no ano de 1997.

Este monitoramento contempla atualmente 64 estações de amostragem de água, onde são realizadas coletas e análises laboratoriais com periodicidade trimestral e avaliação de aproximadamente 50 parâmetros físico-químicos e hidrobiológicos em 12 estações de monitoramento localizadas na calha do rio Doce, essas coletas e análises são realizadas mensalmente.





O objetivo deste trabalho é verificar se é possível fazer uma comparação pela qualidade da água dos períodos antes, durante e depois do acidente que ocorreu com o Rompimento da Barragem de Mariana em 05 de novembro de 2015, propriedade da SAMARCO, no distrito de Bento Rodrigues.

Da base de dados foi extraído a média do ano de 2014 para gerar os dados antes do acidente, a média de do ano de 2016 para gerar os dados durante o acidente (2014), e a média do ano de 2017 para gerar os dados durante o acidente (2017). Ao longo do trabalho os parâmetros físico-químicos e hidrobiológicos que foram encontrados no período do acidente não foram encontrados, sendo necessário reduzir o período de análise da base de dados para 6 meses desde o acidente, mas mesmo assim não apresentou resultados satisfatórios sendo necessário então encurtar o período para 3 meses do mês do acidente.

Foi a retirada de 3 pontos de medição do Rio Caratinga, pois já se sabia que as medições destas estações estavam erradas logo no início (Estação RD056 / RD057/ RD093).

Dos 50 parâmetros mencionados, foram selecionados apenas 13, os quais sofreram maior alteração e impactaram mais na qualidade da água, de acordo com a resolução do CONAMA 357 de 2005, tais como: Arsênio (mg/L), Chumbo (mg/L), Ferro (mg/L), Mercúrio (mg/L), Níquel (mg/L), Manganês (mg/L), Coliformes Totais, Demanda Bioquímica de Oxigênio (DBO), Demanda Química de Oxigênio (DQO), Oxigênio(mg/L), Sólidos Totais, Temperatura da água (oC), Turbidez. Abaixo apresento um resumo da tabela com as classes de qualidade de água.

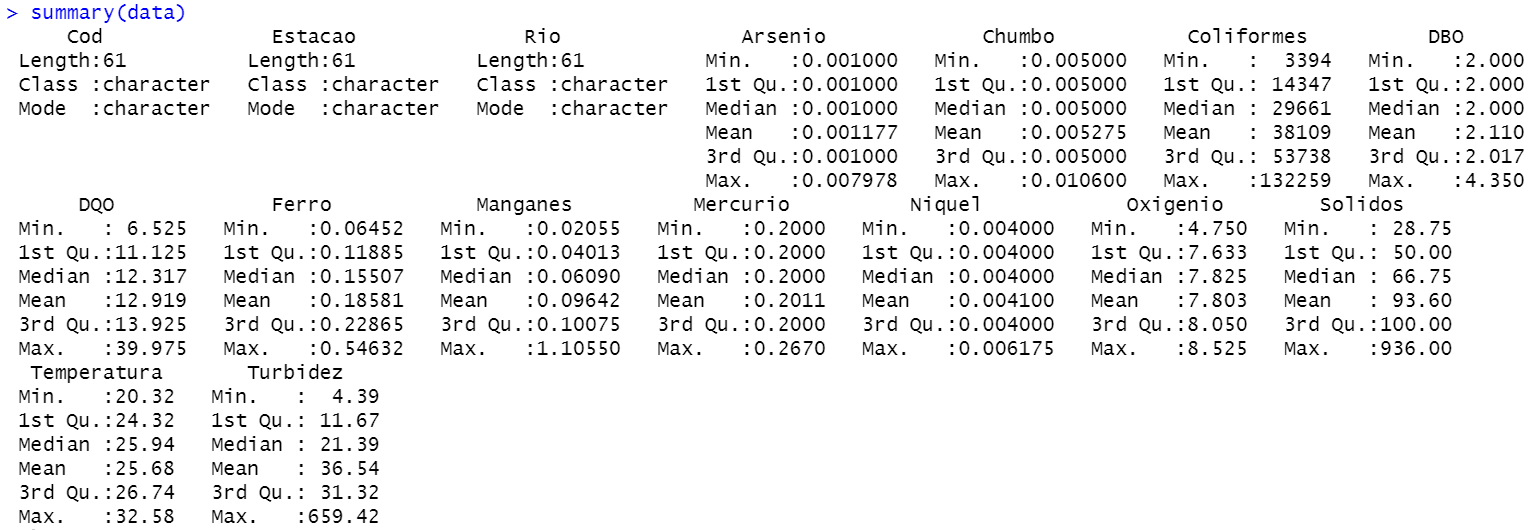


Figura - Resumo da Classificação da Água Doce (CONAMA 357)

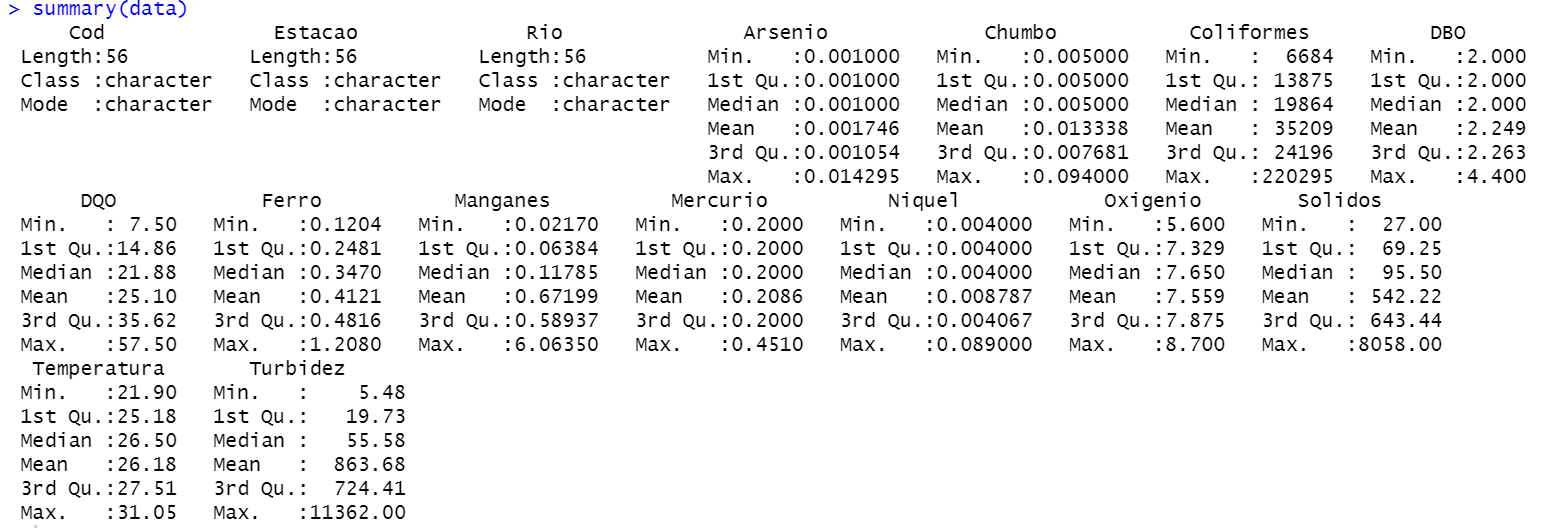
O TRABALHO

Resumo dos Dados obtidos pela base de dados

2014



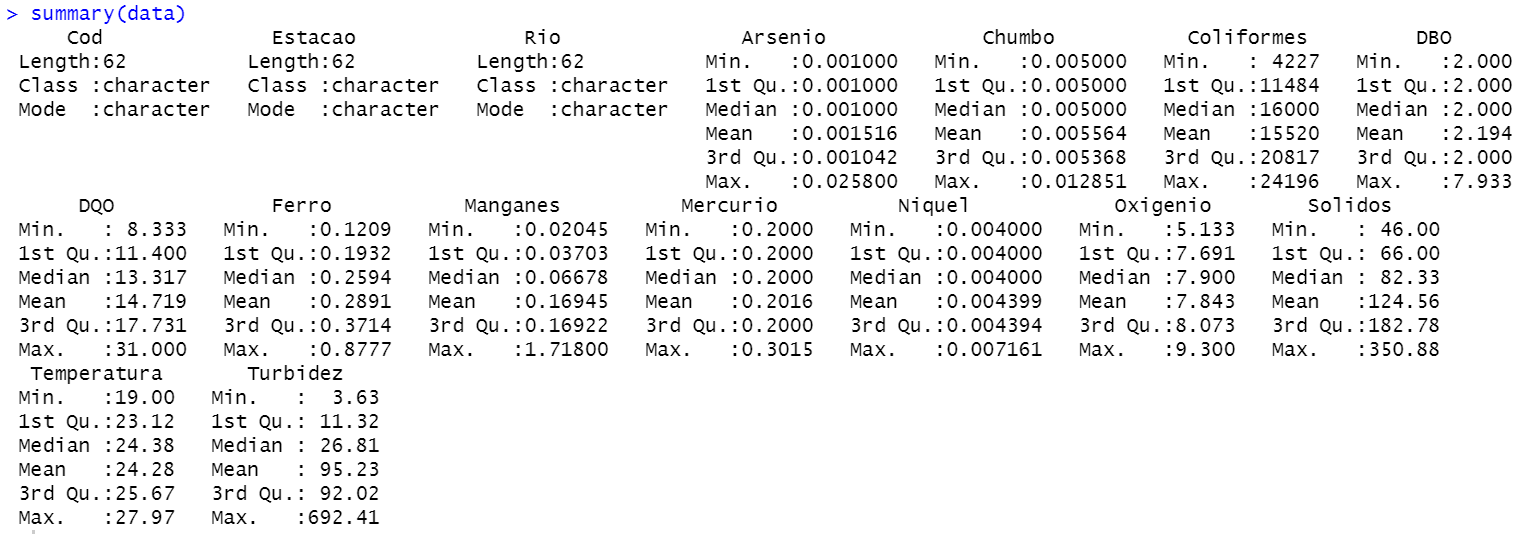
*Possuem outliers – optou-se por não remover pois é uma característica medir alterações marcantes em cada Estação.*

2015 – 2016 (3 meses) não 

*Possuem outliers– optou-se por não remover pois é uma característica medir alterações marcantes em cada Estação.*

Nos dados de 2016 foram omitidos alguns NAs com a função NA.OMIT (4 da variável Arsênio, 3 da variável Chumbo, 3 da variável Níquel, 5 da variável Mercúrio).

2017

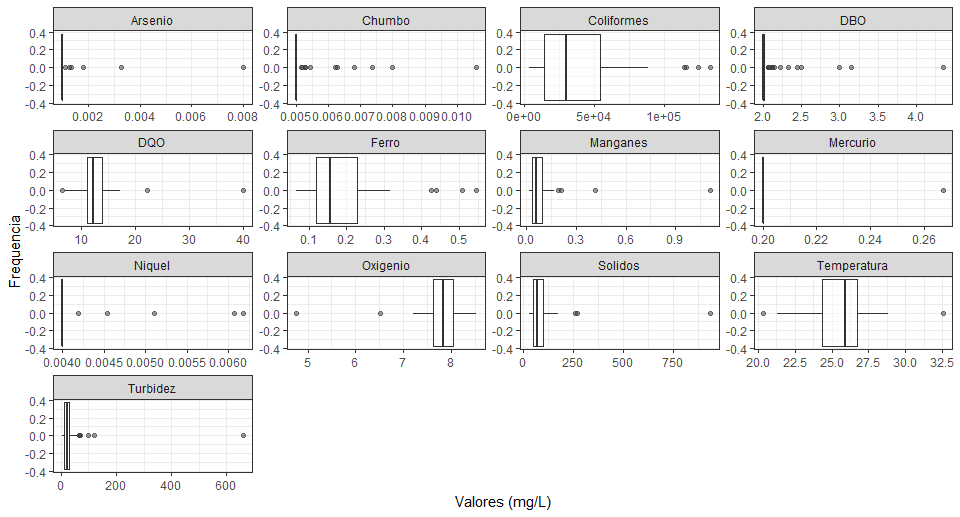


*Possuem outliers– optou-se por não remover pois é uma característica medir alterações marcantes em cada Estação.*

Boxplot – Mostra a variação de cada variável com seus respectivos outliers

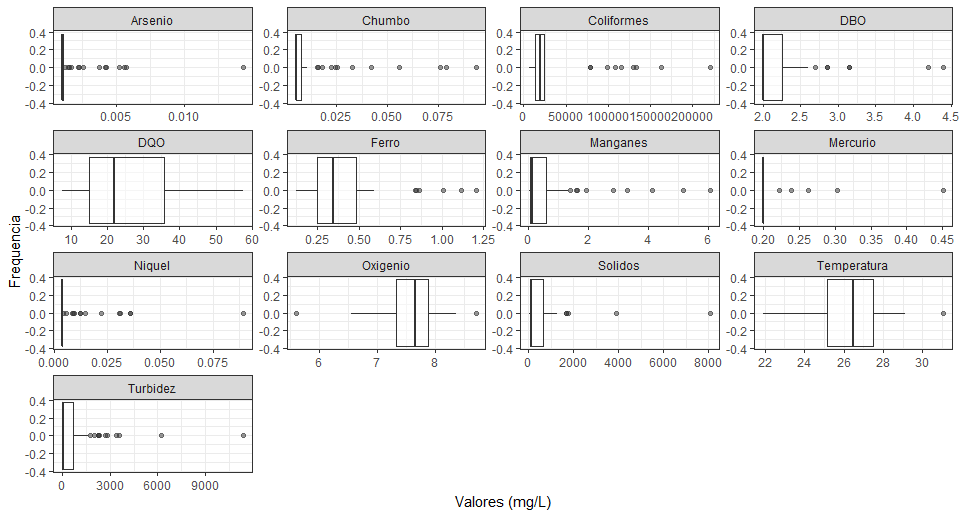
2014

Padrões de normalidade.



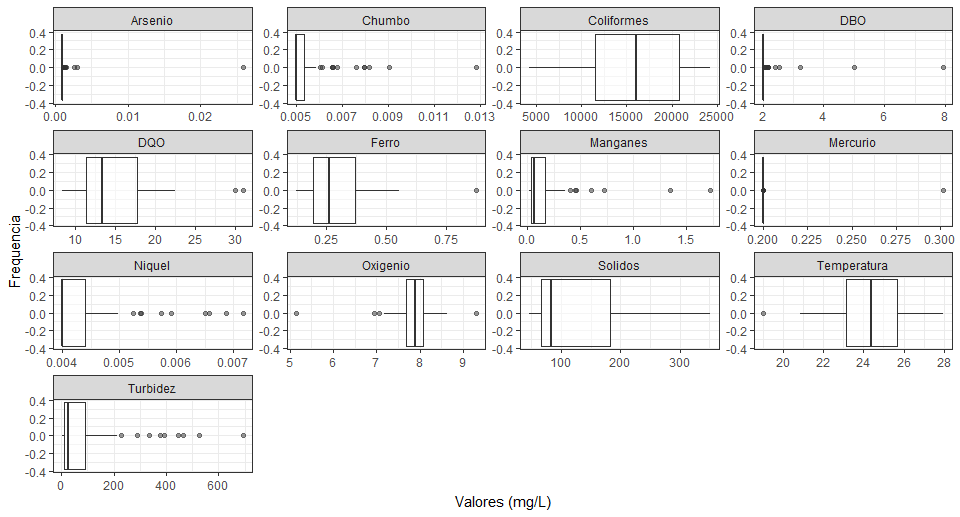
2015 – 2016 (3 meses)

Aumento da Turbidez, Solidos, Manganês, Ferro, Níquel – Bem acima da normalidade.



2017

Retornando os valores para a normalidade, embora a Turbidez, Ferro e Manganês ainda esteja acima do normal.



Matriz de Correlação de das variáveis

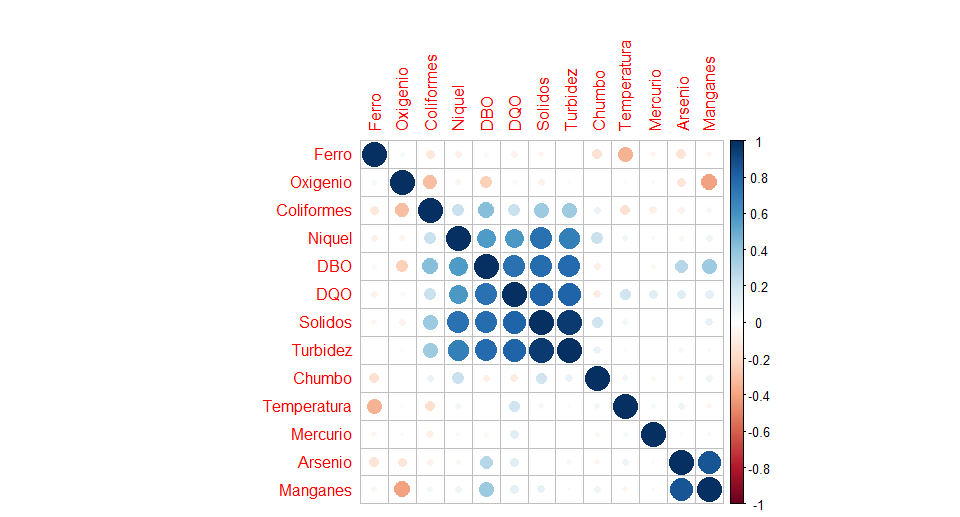
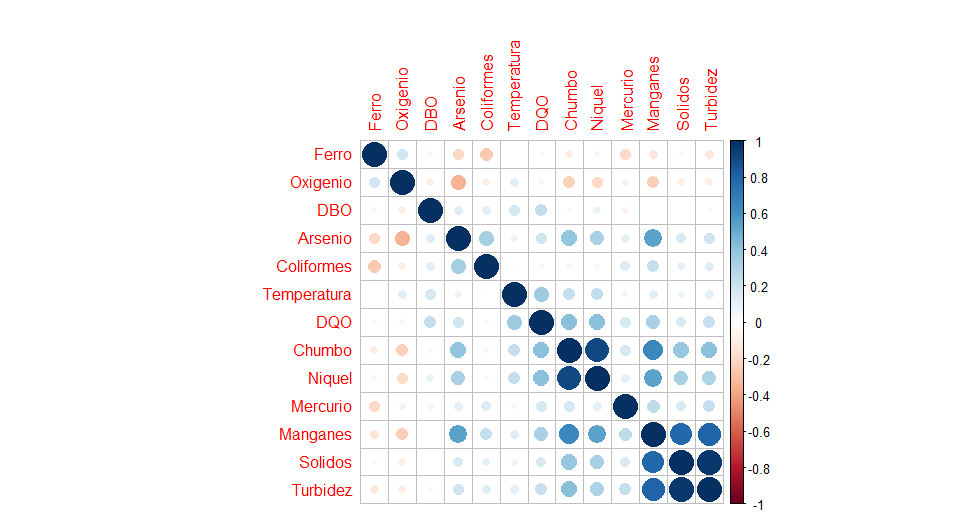
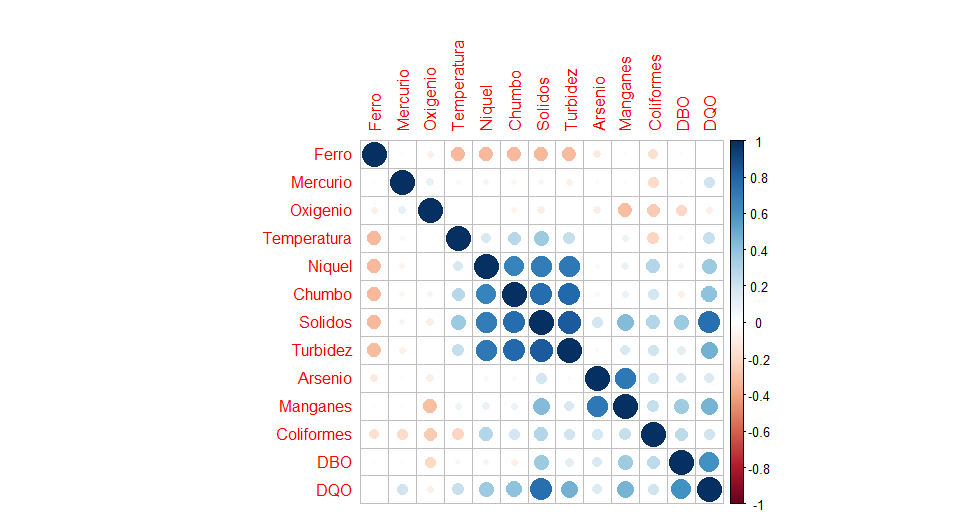
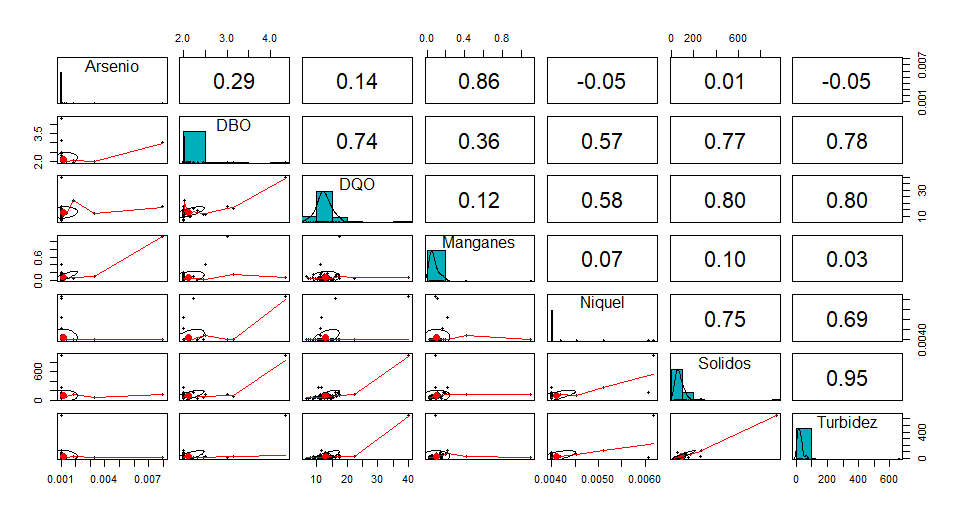
  

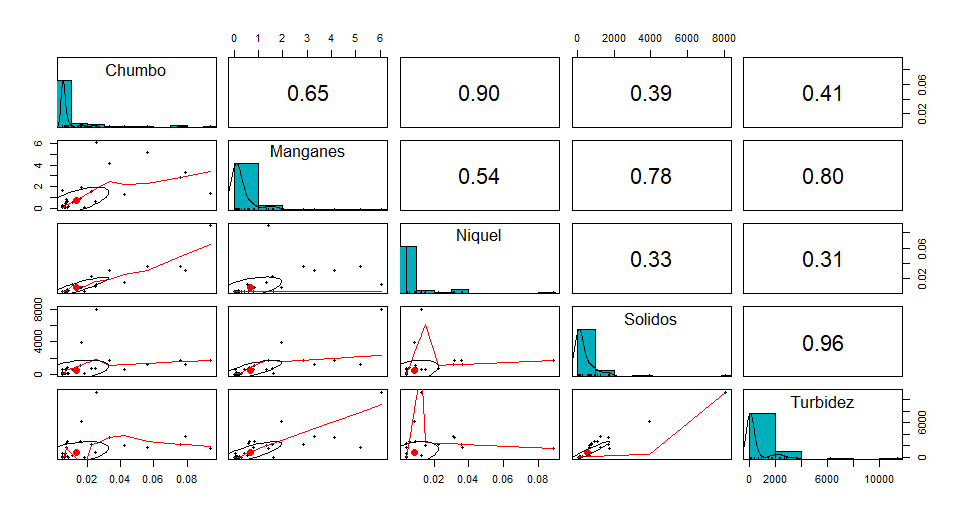
Figura - Matriz de Correlação 2014/ 2015-2016 (3 meses) / 2017

Foi separado as variáveis mais impactantes – valores acima de 0,60.

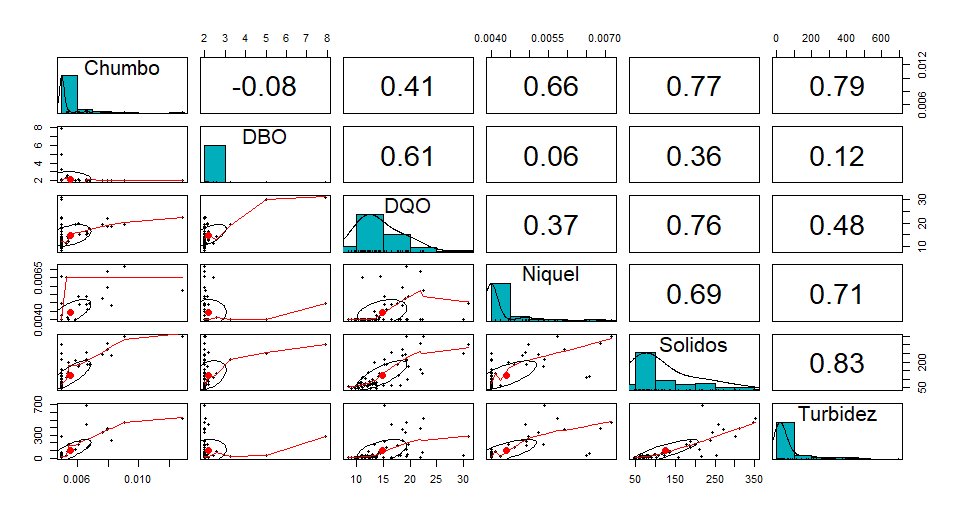
2014



2015 – 2016 (3 meses)

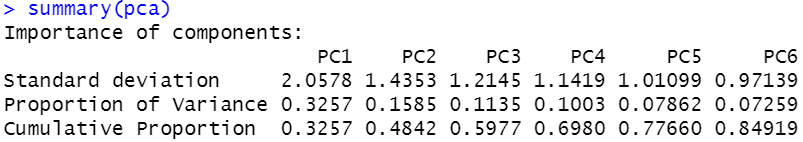


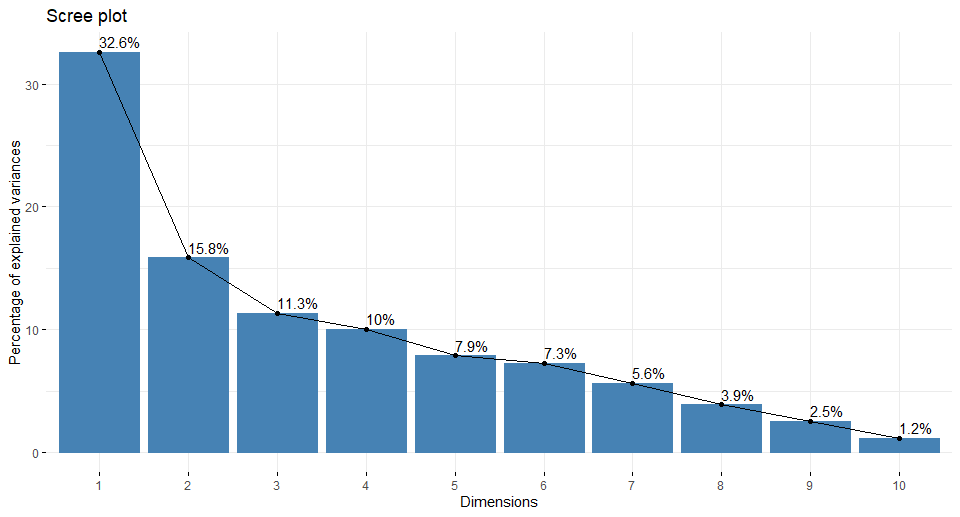
2017



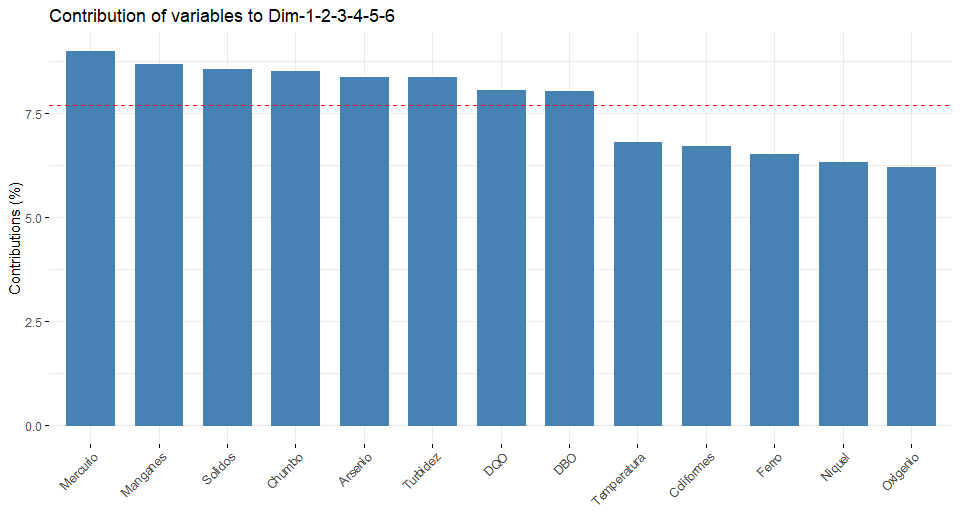
Os dados foram normalizados utilizando PCA para diminuir o ruído (redução de dimensionalidade - variáveis). Nos 3 bancos de dados em questão, a proporção acumulativa no valor de 80% corresponde a 6 variáveis.

2014

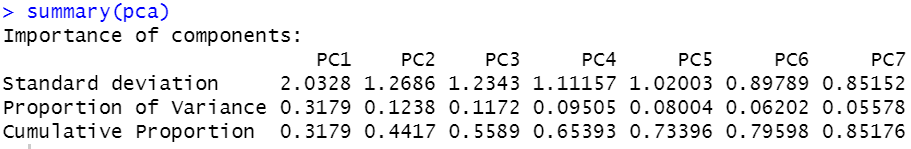


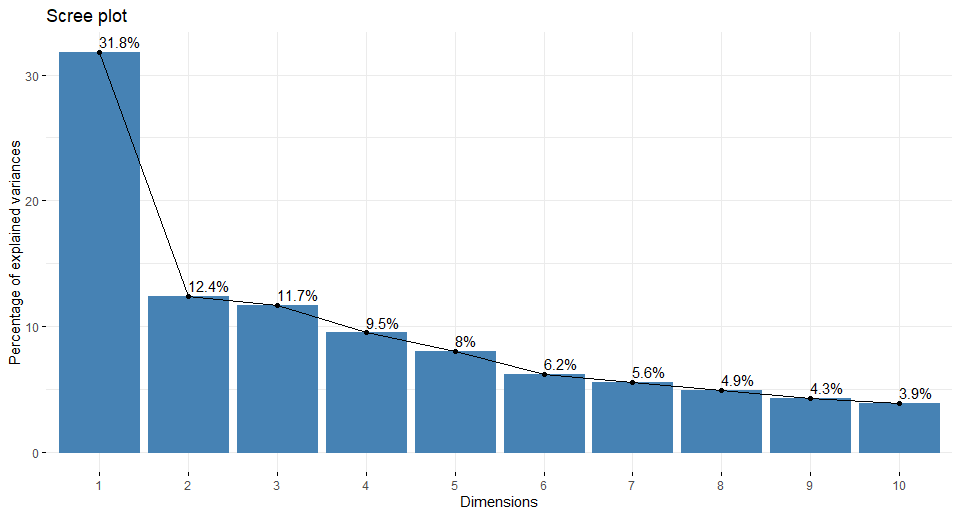


*2014 Soma = 1 ao 6= 84,92%*

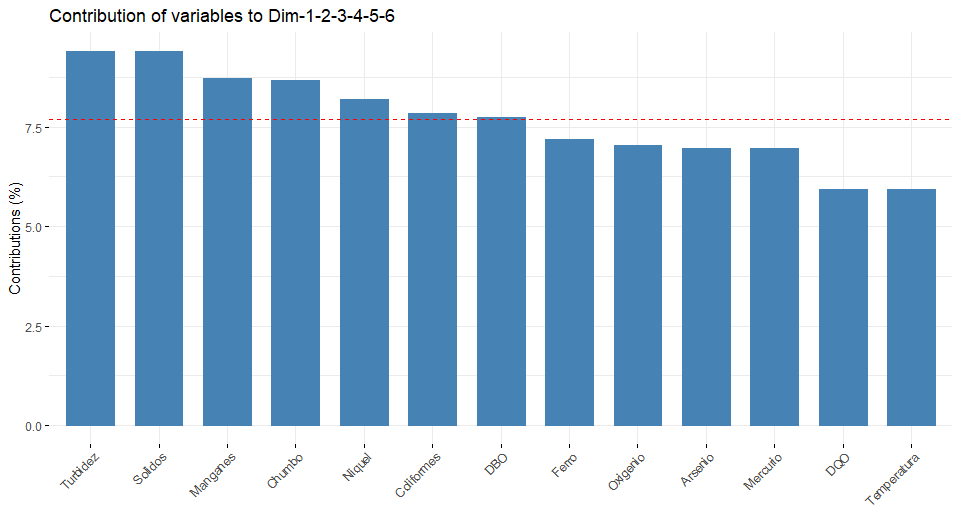


2015 – 2016 (3 meses)

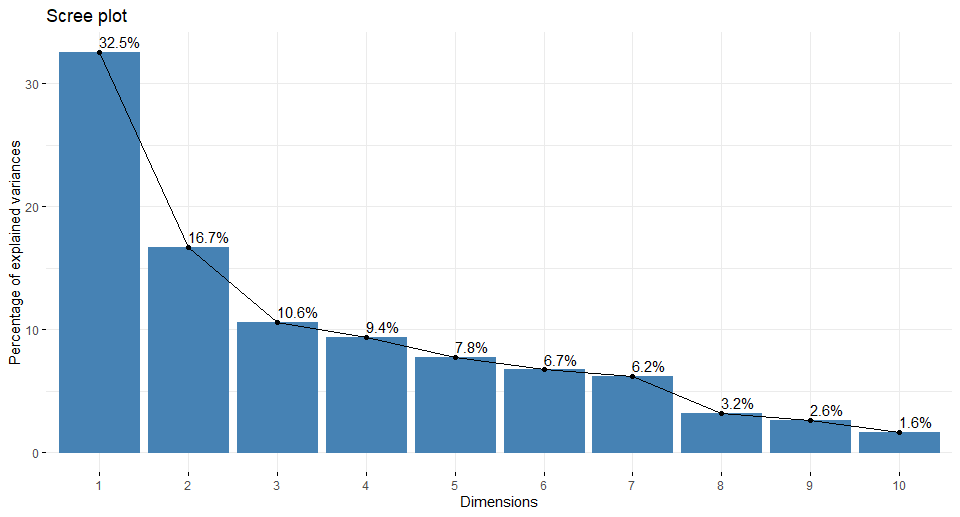
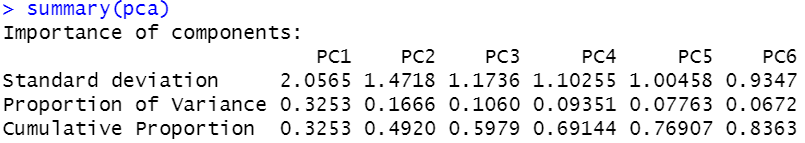




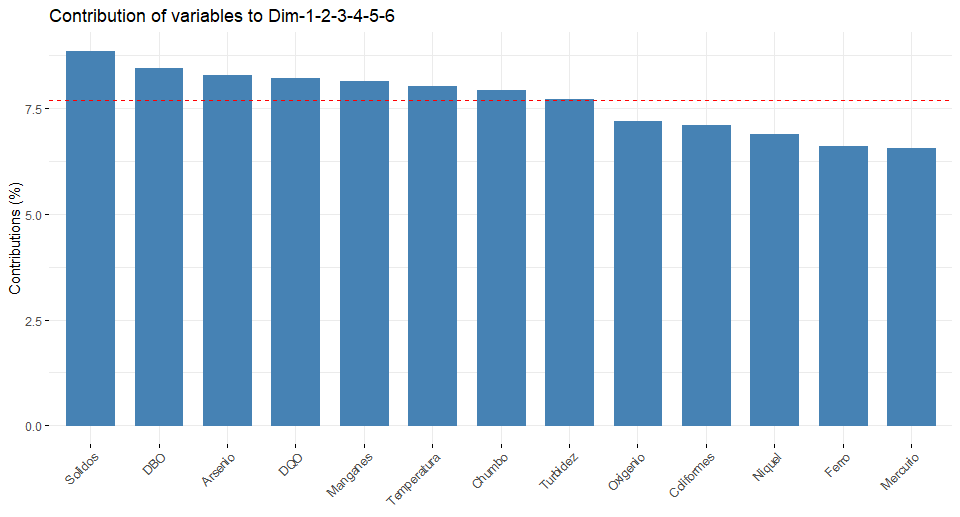
*2015 - 2016 (3 meses) - Soma 1 ao 6 = 85,18%*



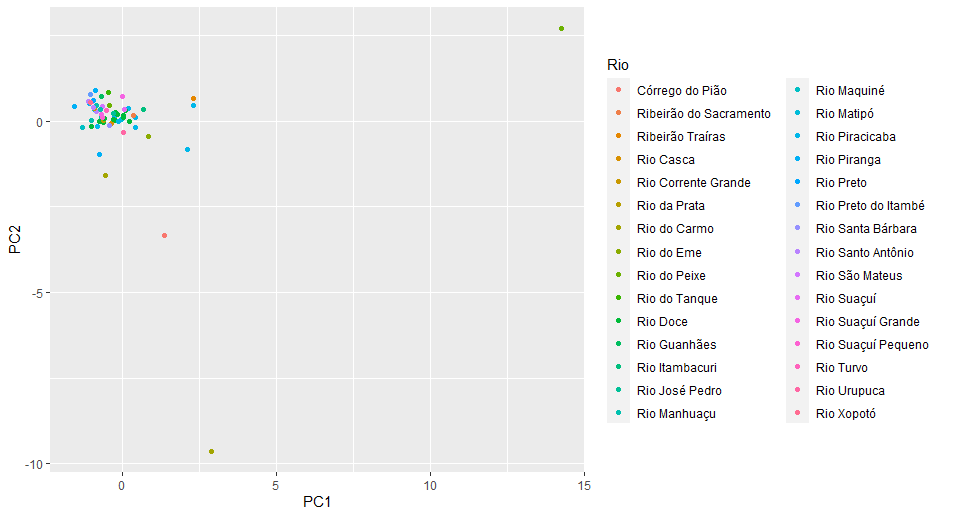
2017

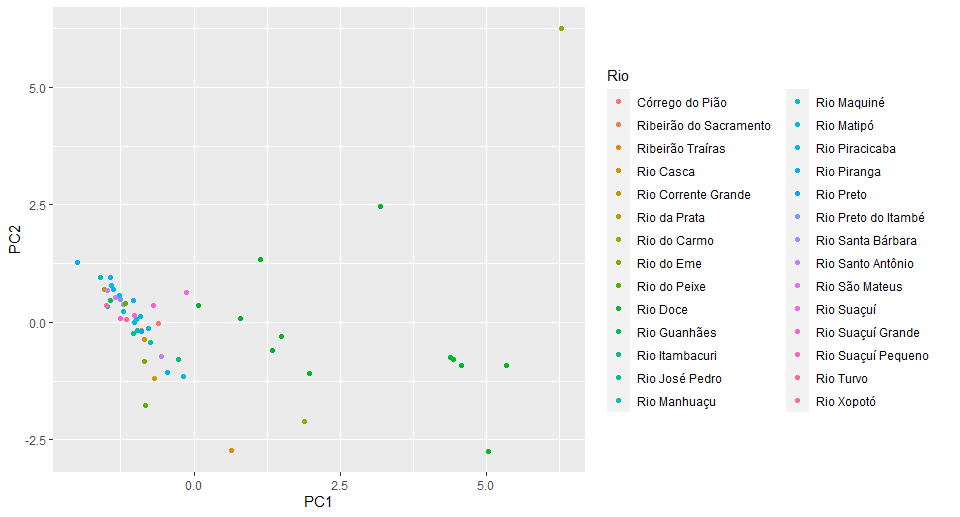


*2017 - Soma 1 ao 6 = 83,63%*



Dos 64 pontos de medição foram retirados 3 do Rio Caratinga, pois foi considerado um outlier logo no ínicio (fora d.





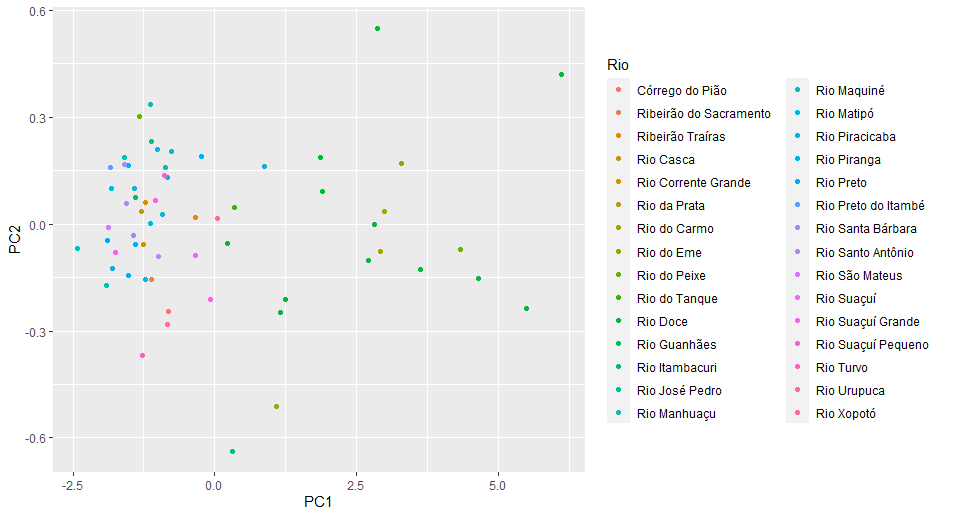
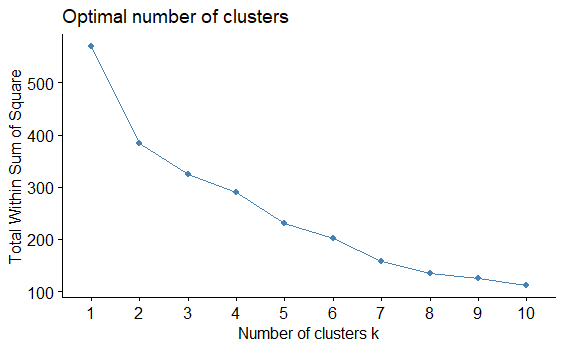
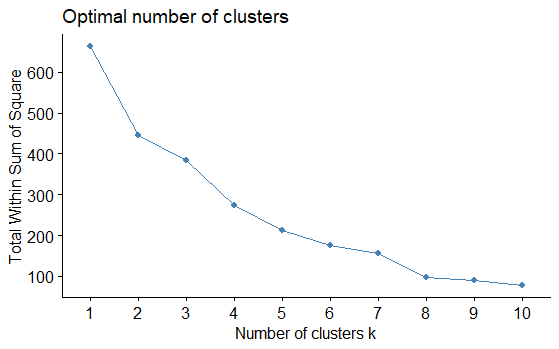


Figura - 2014 / 2016 / 2017 - por tipo de Classe



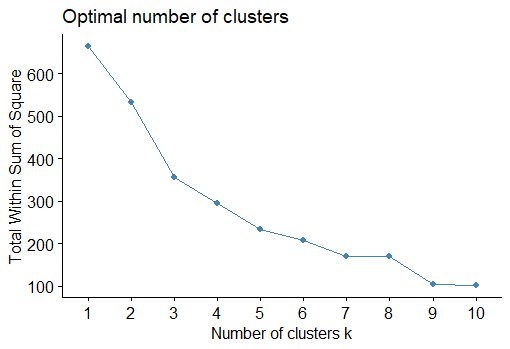


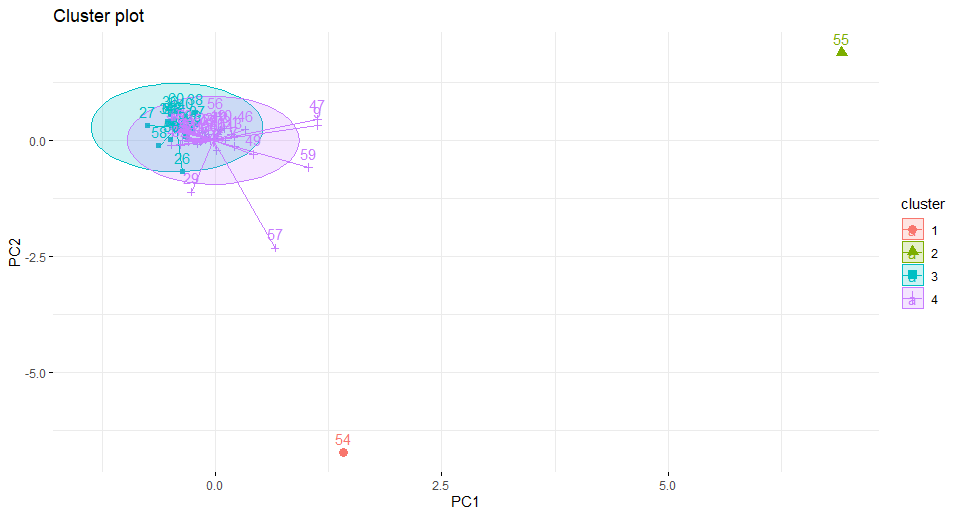
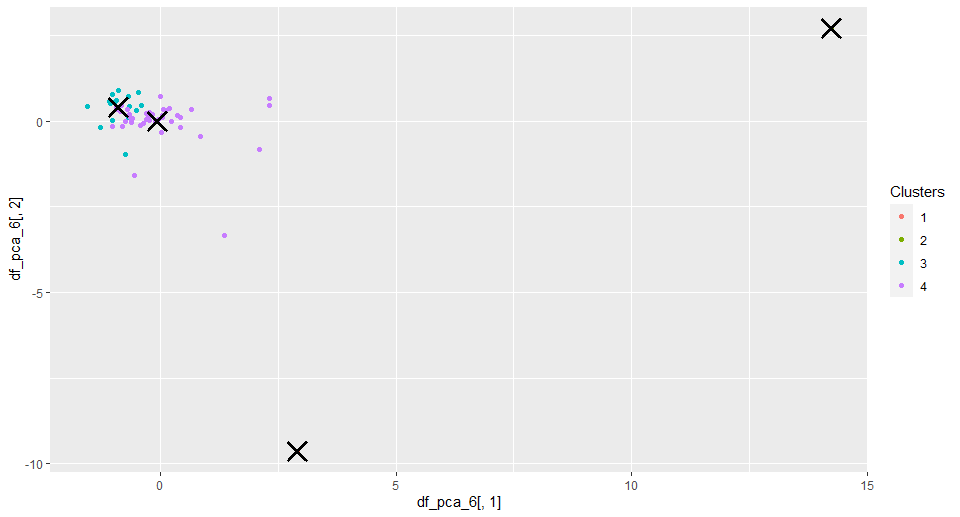
Figura - Número Ideal de Clusters 2014 / 2015-2016 (3 meses) / 2017

Todos indicam um Cluster de 6 pelo método de cotovelo.

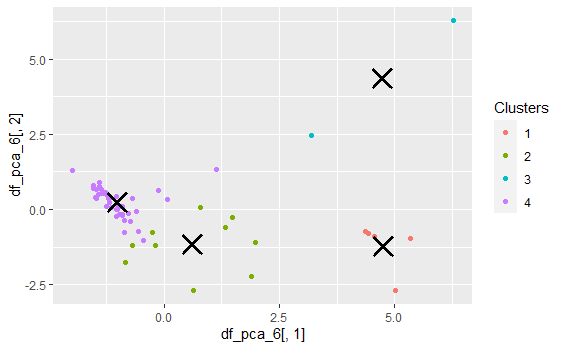
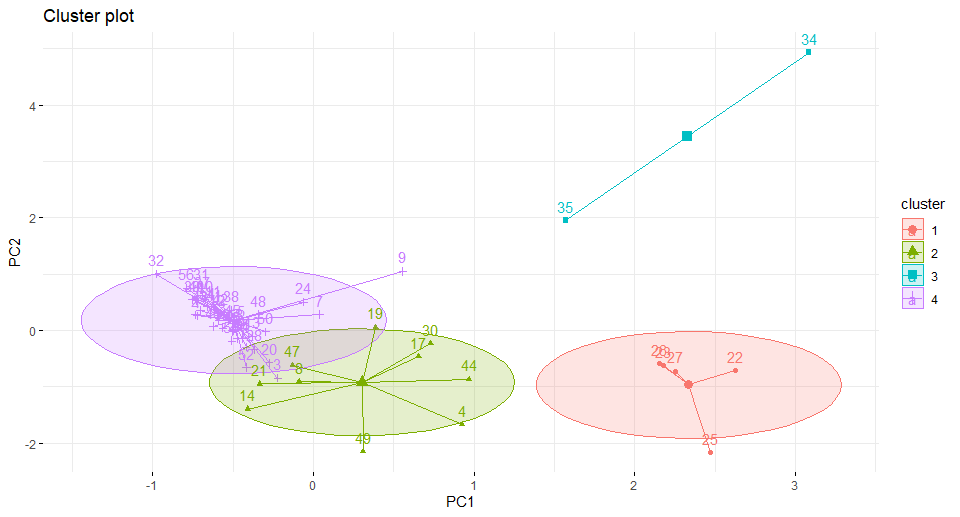
MÉTODOS DE AGRUPAMENTO

1. KMEANS – apesar do método cotovelo sugerir 6 agrupamentos utilizei 4 devido a classificação das águas se enquadrarem em 4 classes.

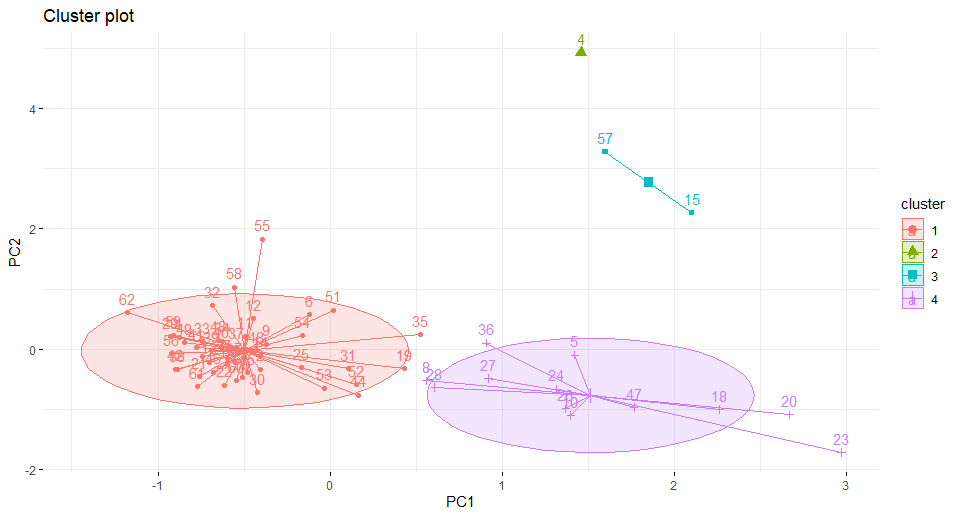
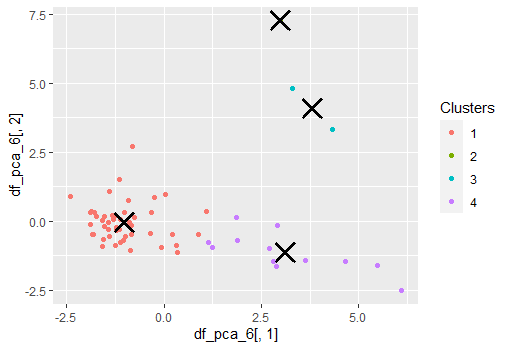
2014 (n=4)



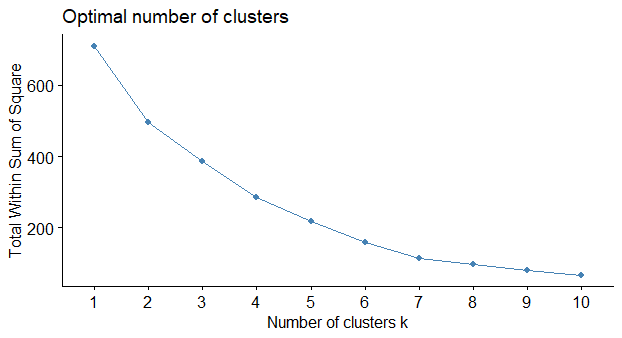
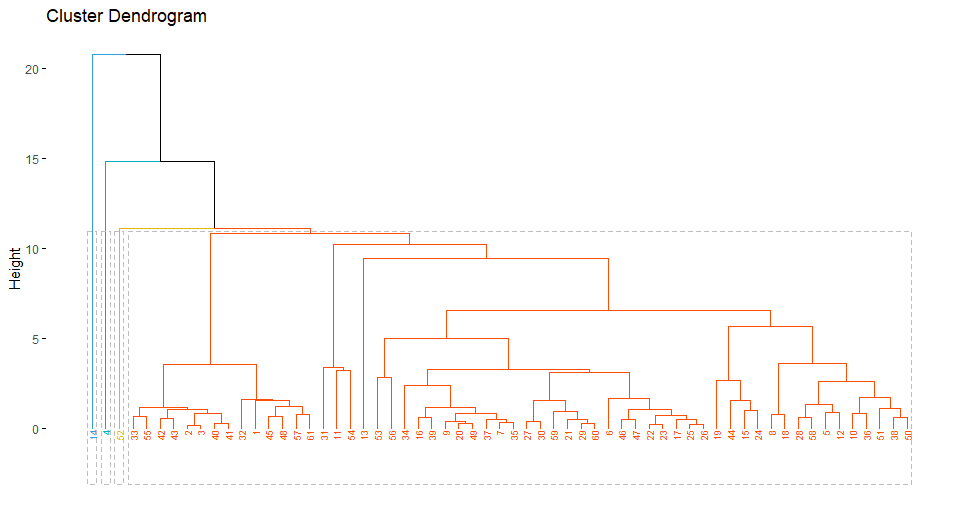
2015 – 2016 (3 meses) (n=4)

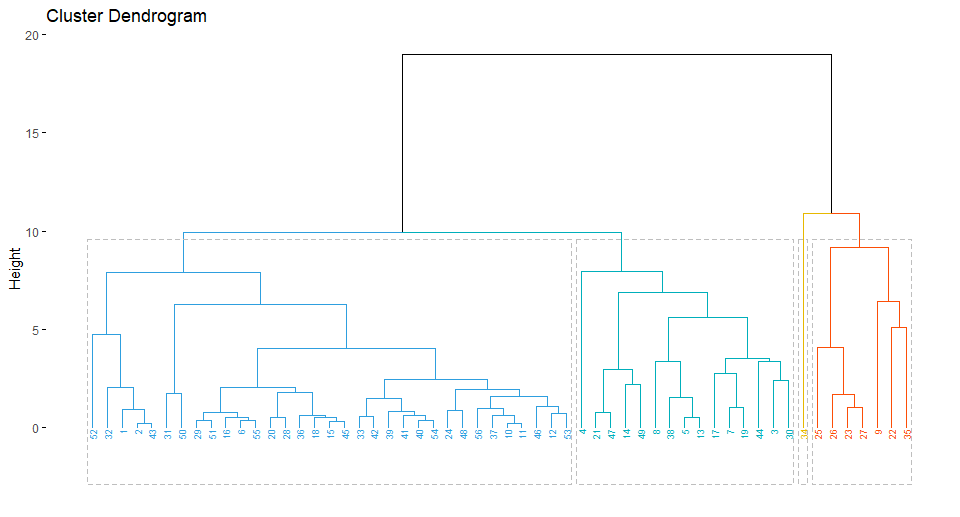
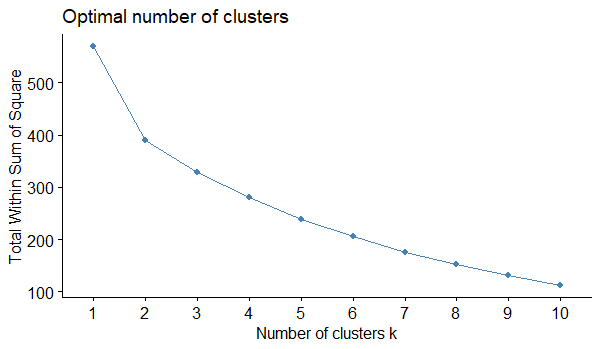
2017 (n=4)



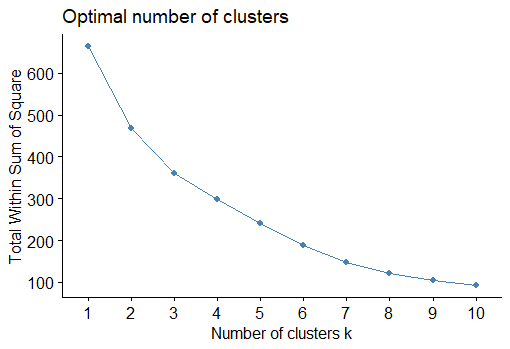
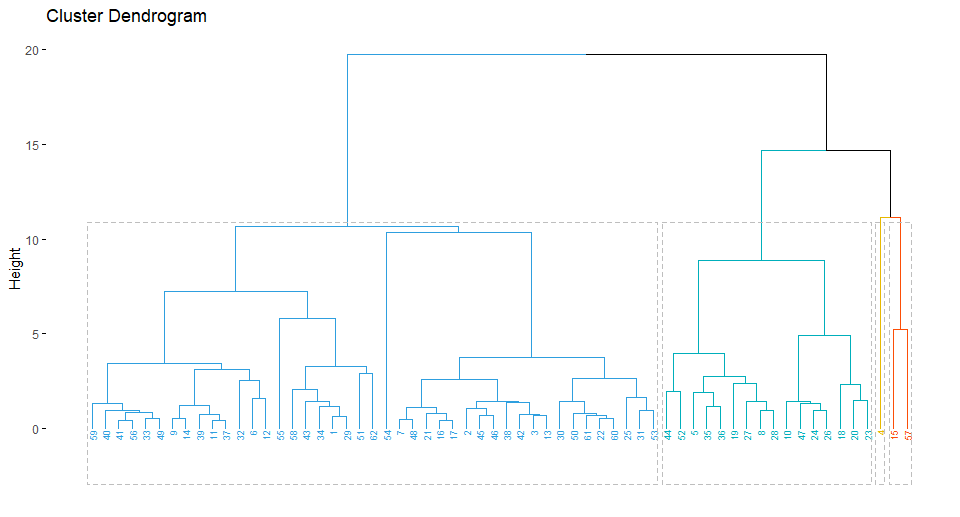
1. HIERÁRQUICO (N = 4)



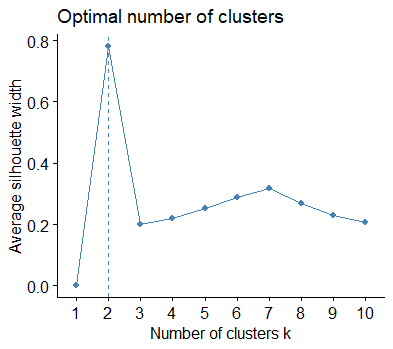
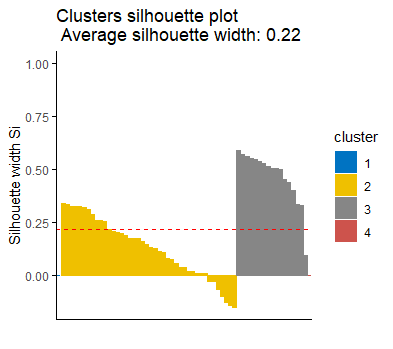
2015 – 2016 (3 meses)

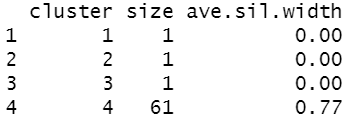
 

2017

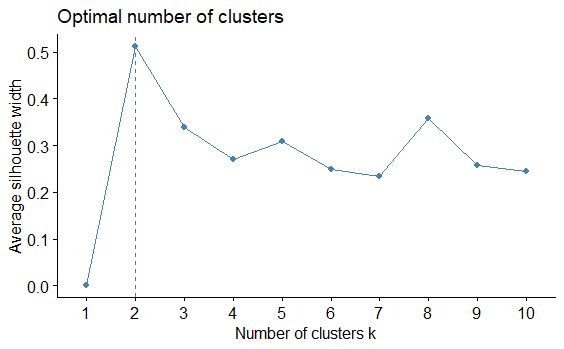
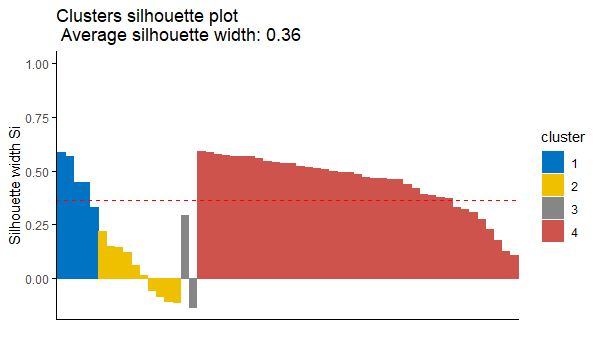


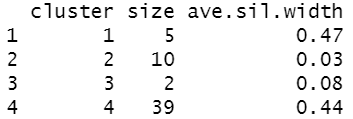
ANÁLISE DA SILHUETA



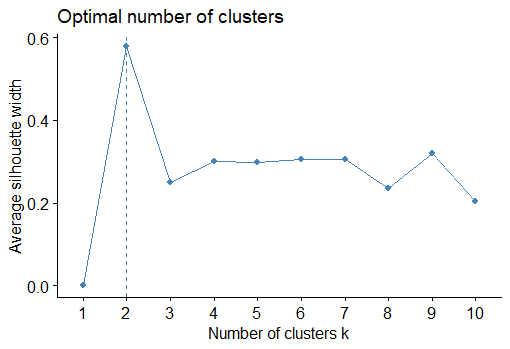
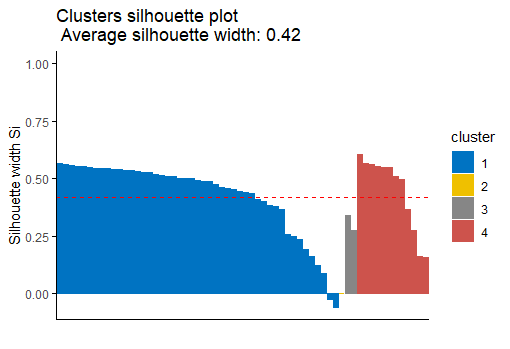


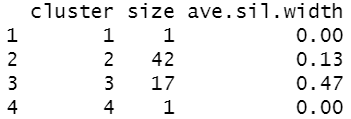
2015 – 2016 (3 meses)



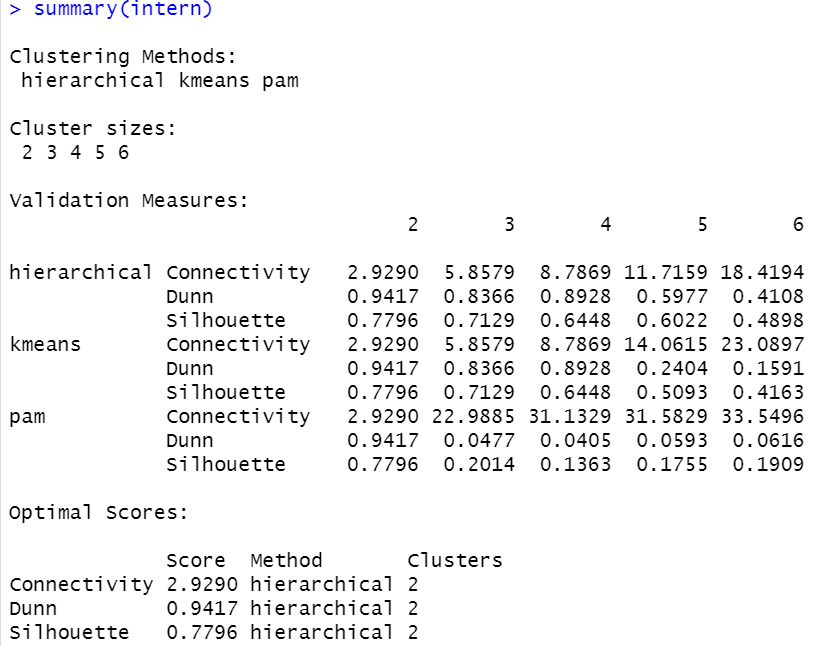


2017





2014

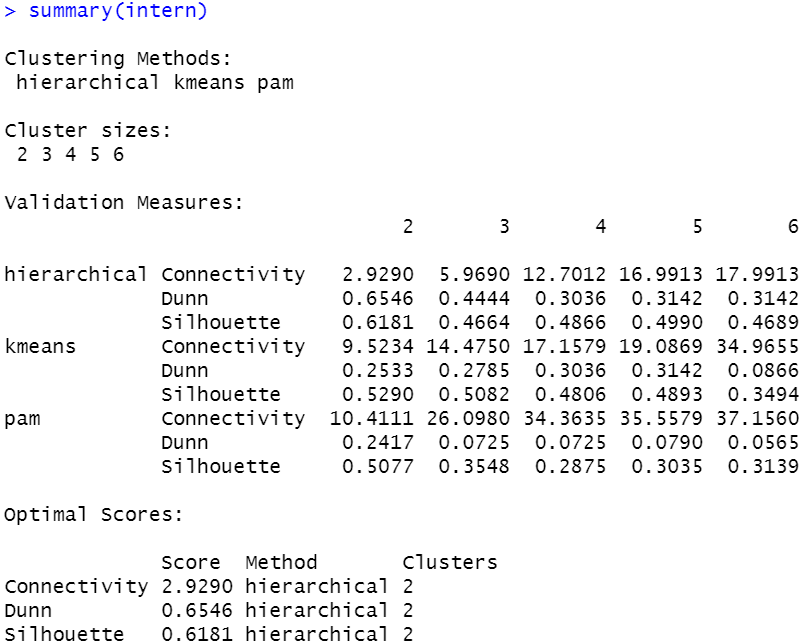


*- Vizinhos mais próximos*

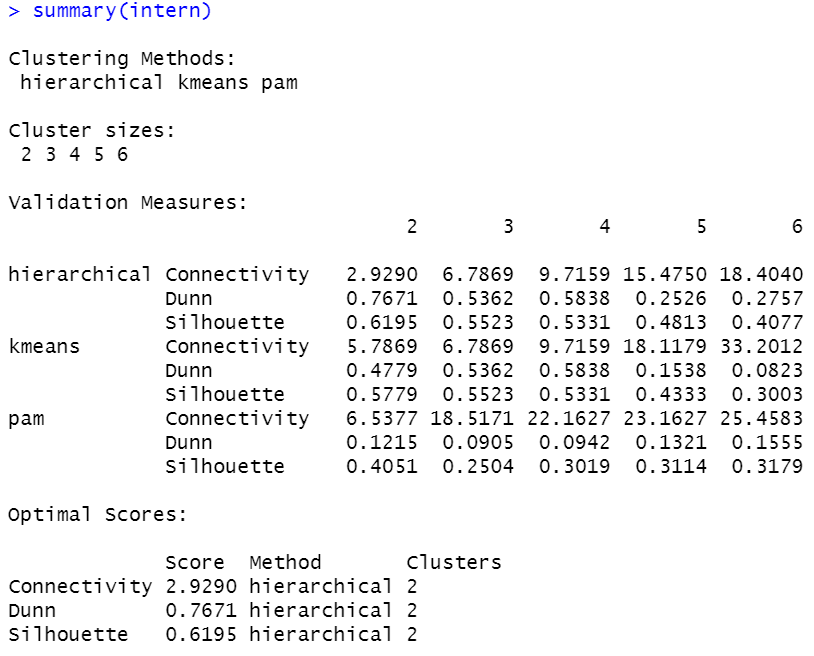
*- Índice do pior caso*

*- Como melhor se encontra cada objeto em seu conjunto (Mais próximo de 1 – melhor classificado) ideal >0,7*

2015 – 2016 (3 meses)



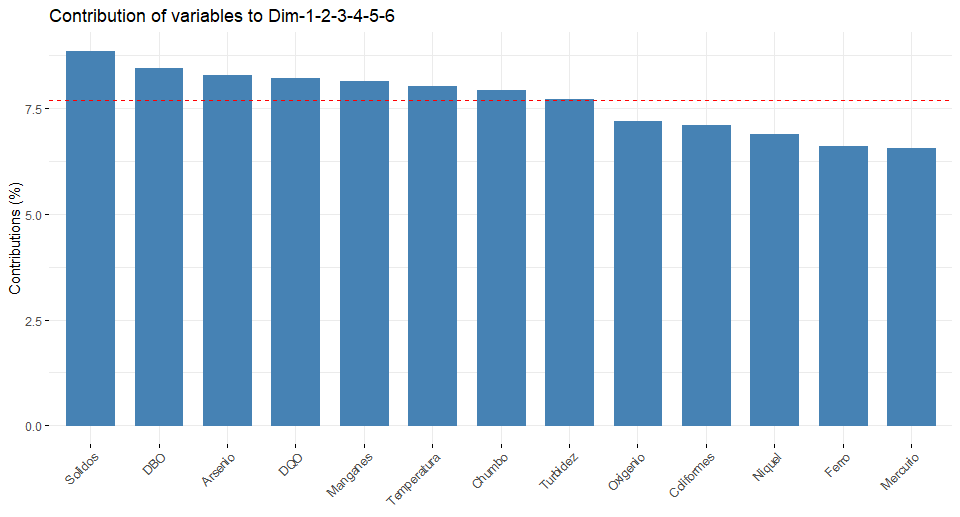
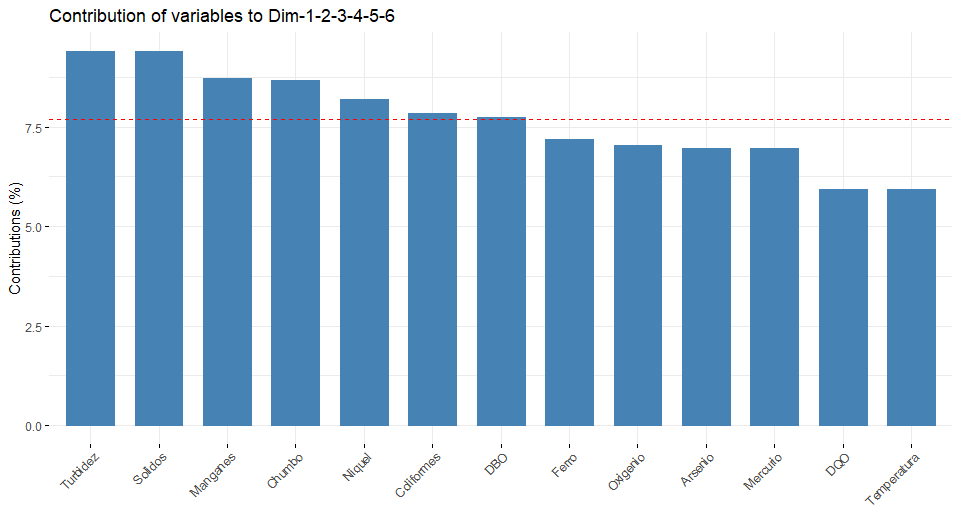
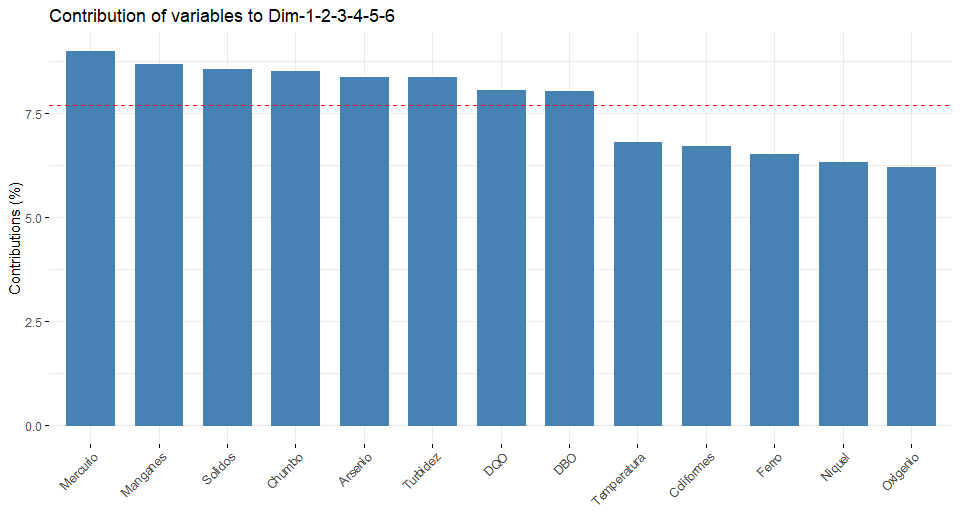
2017



CONCLUSÃO

Devido ao rompimento da barragem os parâmetros que comprovadamente sofreram maior alteração foram: Aumento dos índices de Turbidez, Ferro, Manganês, Níquel e Sólidos em suspensão e Redução dos índices de Oxigênio e DBO. Com relação aos índices de Mercúrio e Arsênio, estes não sofreram alteração, já que os mesmos existiam no sedimento do rio.

Nas bases de dados selecionadas pudemos observar esta mudança de comportamento entre as variáveis. (parâmetros). Sendo observado durante o período do acidente estes parâmetros citados acima como mais importantes a Turbidez, Sólidos, Manganês, Níquel. Confirmando a adequação da base de dados e a algoritmo à realidade.



2015-2016 (3 meses)

2017 (1 ano)

2014 (1 ano)

Conforme mencionado acima, apesar do método cotovelo sugerir agrupamento de 6 foi utilizado agrupamento de 4, observando o critério do CONAMA que classifica em 4 classes a qualidade de águas.

* Antes do rompimento da barragem (2014), o método Kmeans agrupa a maioria das estações em 2 grupos bem próximos, significando estas possuem características semelhantes bem próximas, ficando algumas estações como outliers. Embora o índice silhueta permite criar até 3 grupos, apresentou melhor resultado para 2 grupos, ambos acima de 0,7. Foi sugerido método cluster Hierárquico para 2 grupos ou Kmeans (apresentou o mesmo índice).
* Logo após o acidente (2015 - 2016), o agrupamento Kmeans criou 3 grupos sendo que 2 deles mais próximos e o terceiro um pouco mais distante. Este distante é o que apresentou mais alterações na qualidade de água, referente a algumas Estações do Rio Doce. O índice silhueta apresenta melhor resultado para apenas 2 grupos, porém abaixo de 0,7. Foi sugerido cluster Hierárquico de 2 grupos.
* 1 anos após o acidente (2017), o agrupamento gerado foi em 2 grupos, porém as distâncias entre estes ficaram mais próximas. O índice silhueta não apresentou um resultado satisfatório para 2 grupos. Sugerido cluster Hierárquico de 2 grupos.

Índice abaixo de 0,7 pode indicar que a base de dados não foi bem trabalhada, isto é, possíveis outliers poderiam ter sido retirados para não atrapalhar na classificação, já que o método Kmeans é sensível a outliers. O que foi possível detectar pela presença de clusters com poucos pontos. Porém não foram retirados pois o intuito era analisar todo o trecho da Bacia do Rio Doce.