TCC

Tiago de Almeida Silva

2023-01-06

Carregando os pacotes

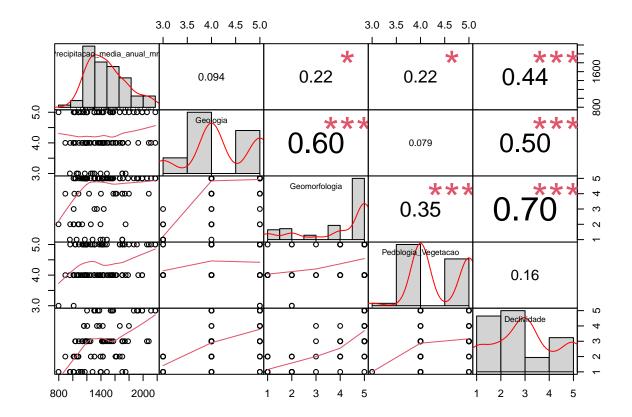
Carregando dataset

```
df_tcc <- readxl::read_xlsx("df_tcc.xlsx", col_names = T)</pre>
```

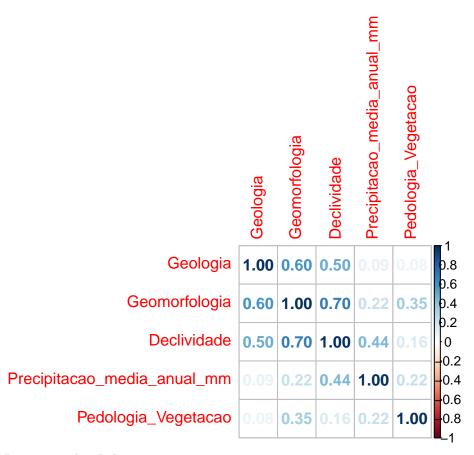
Correlação da base de Dados

Matriz de Correlação de Pearson com nível de significância

```
correl <- rcorr(as.matrix(df_tcc[ ,c(4,6:9)]))
chart.Correlation(df_tcc[ ,c(4,6:9)], histogram = T, pch = 19, method = "pearson")</pre>
```



corrplot(correl\$r, type = "full", order = "hclust", method = "number")



#Estatisticas Descritivas dos dados

```
summary(df_tcc[ ,c(4,6:9)])
```

```
Precipitacao_media_anual_mm
                                     Geologia
                                                   Geomorfologia
           : 797
                                                          :1.000
##
    Min.
                                 Min.
                                         :3.000
                                                   Min.
##
    1st Qu.:1174
                                  1st Qu.:4.000
                                                   1st Qu.:3.750
##
    Median:1366
                                  Median :4.000
                                                   Median :5.000
##
    Mean
           :1401
                                  Mean
                                         :4.228
                                                   Mean
                                                          :4.054
##
    3rd Qu.:1584
                                  3rd Qu.:5.000
                                                   3rd Qu.:5.000
##
           :2174
                                  Max.
                                         :5.000
                                                   Max.
                                                          :5.000
    Max.
##
    Pedologia_Vegetacao
                          Declividade
                                 :1.000
##
    Min.
           :3.000
                         Min.
    1st Qu.:4.000
                         1st Qu.:2.000
##
    Median :4.000
                         Median :3.000
##
##
    Mean
           :4.402
                         Mean
                                 :2.989
##
    3rd Qu.:5.000
                         3rd Qu.:4.000
    Max.
           :5.000
                                 :5.000
##
                         Max.
\#st(df\_tcc[,c(4,6:9)],\ title = "Estatísticas Descritivas")
```

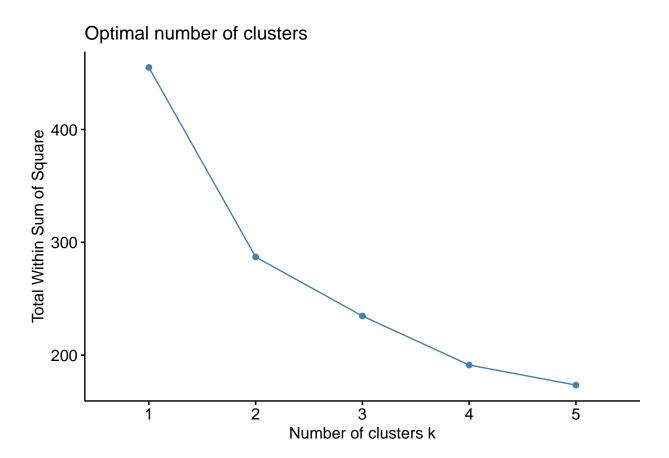
Padronização dos dados Dados (Z-score)

```
df_padronizado <- as.data.frame(scale(df_tcc[,c(4, 6:9)]))
rownames(df_padronizado) <- df_tcc$nome</pre>
```

#Esquema de aglomeração não hierárquico K-MEANS (Machine Learning) Elaboração da clusterização não hieráquica k-means

Método de Elbow para identificação do número ótimo de clusters

```
fviz_nbclust(df_padronizado[ ,1:5], kmeans, method = "wss", k.max = 5)
```



Criando variável categórica para indicação do cluster no banco de dados

```
df_padronizado$cluster_K <- factor(cluster_kmeans$cluster)</pre>
```

Análise de variância de um fator (ANOVA)

ANOVA da variável 'precipitação'

```
Df Sum Sq Mean Sq F value Pr(>F)
             4 43.71 10.928
                                   20.1 9.43e-12 ***
## cluster K
## Residuals 87 47.29
                         0.544
## ---
## Signif. codes: 0 '*** 0.001 '** 0.01 '* 0.05 '.' 0.1 ' 1
ANOVA da variável 'Geologia'
summary(anova_matematica <- aov(formula =</pre>
                             Geologia ~ cluster_K,data = df_padronizado))
              Df Sum Sq Mean Sq F value Pr(>F)
                          13.40
              4 53.59
                                 31.16 4.27e-16 ***
## cluster_K
## Residuals
              87 37.41
                           0.43
## Signif. codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1
ANOVA da variável 'Geomorfologia'
summary(anova_matematica <- aov(formula =</pre>
                             Geomorfologia ~ cluster_K,data = df_padronizado))
              Df Sum Sq Mean Sq F value Pr(>F)
                          19.49
               4 77.97
                                 130.2 <2e-16 ***
## cluster K
## Residuals
             87 13.03
                           0.15
## ---
## Signif. codes: 0 '*** 0.001 '** 0.01 '* 0.05 '.' 0.1 ' 1
ANOVA da variável 'Pedologia_Vegetacao'
summary(anova_matematica <- aov(formula =</pre>
                         Pedologia_Vegetacao ~ cluster_K,data = df_padronizado))
              Df Sum Sq Mean Sq F value
                                          Pr(>F)
               4 39.79
                          9.946
                                   16.9 2.77e-10 ***
## cluster_K
              87 51.21
                          0.589
## Residuals
## Signif. codes: 0 '*** 0.001 '** 0.01 '* 0.05 '.' 0.1 ' 1
ANOVA da variável 'Declividade'
summary(anova_matematica <- aov(formula =</pre>
                           Declividade ~ cluster_K,data = df_padronizado))
              Df Sum Sq Mean Sq F value Pr(>F)
##
               4 65.91 16.477
                                  57.12 <2e-16 ***
## cluster_K
## Residuals
              87 25.09
                          0.288
## Signif. codes: 0 '*** 0.001 '** 0.01 '* 0.05 '.' 0.1 ' ' 1
```

Cluster com pesos

```
df_padronizado$prec_p <- df_padronizado$Precipitacao_media_anual_mm * 0.2

df_padronizado$geo_p <- df_padronizado$Geologia * 0.15

df_padronizado$geomor_p <- df_padronizado$Geomorfologia * 0.2

df_padronizado$ped_p <- df_padronizado$Pedologia_Vegetacao * 0.1

df_padronizado$dec_p <- df_padronizado$Declividade * 0.35</pre>
```

#Esquema de aglomeração não hierárquico K-MEANS

Elaboração da clusterização não hieráquica k-means

Estatística sobre o cluster

```
print(cluster_kmeans_p)
```

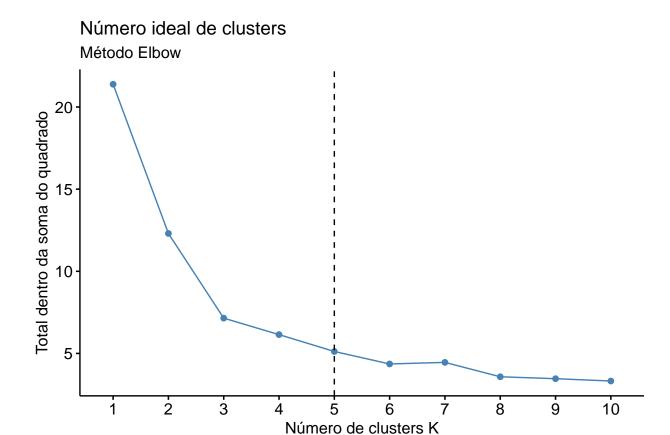
```
## K-means clustering with 5 clusters of sizes 12, 21, 18, 19, 22
##
## Cluster means:
##
          prec_p
                       geo_p
                                geomor_p
                                                 ped_p
                                                             dec_p
## 1 -0.08216134 0.15534891 0.11055211 0.049368679 0.13360400
## 2 0.16086091 0.08818679 0.12744712 -0.003961088 0.50098115
## 3 0.15568611 -0.00136271 0.06324609 0.018259648 -0.04074448
## 4 -0.15915970 -0.01587915 0.10432763 0.013347696 -0.09350784
## 5 -0.09865721 -0.15408531 -0.32380316 -0.049614600 -0.43699102
##
## Clustering vector:
##
                  Angra dos Reis
                                                        Aperibé
##
##
                        Araruama
                                                          Areal
##
##
              Armação dos Búzios
                                                Arraial do Cabo
##
##
                  Barra do Piraí
                                                    Barra Mansa
##
##
                    Belford Roxo
                                                     Bom Jardim
##
                                                              2
##
         Bom Jesus do Itabapoana
                                                      Cabo Frio
##
                                                              5
##
            Cachoeiras de Macacu
                                                        Cambuci
##
##
                                      Comendador Levy Gasparian
                       Carapebus
##
##
           Campos dos Goytacazes
                                                      Cantagalo
##
                                                              2
                               5
```

##	Cardoso Moreira	Carmo	
##	Cardoso Morerra 4	1	
##	Casimiro de Abreu	Conceição de Macabu	
##	3	3	
##	Cordeiro	Duas Barras	
##	2	2	
##	Duque de Caxias	Engenheiro Paulo de Frontin	
##	5	2	
##	Guapimirim	Iguaba Grande	
##	3	4	
##	Itaboraí	- Itaguaí	
##	5	5	
##	Italva	Itaocara	
##	1	1	
##	Itaperuna	Itatiaia	
##	4	3	
##	Japeri	Laje do Muriaé	
##	4	3	
##	Macaé	Macuco	
##	4	2	
##	Magé	Mangaratiba	
##	3	2	
##	Maricá	Mendes	
##	4	1	
##	Mesquita	Miguel Pereira	
##	5	2	
##	Miracema	Natividade	
##	4	1	
##	Nilópolis	Niterói	
##	5	5	
## ##	Nova Friburgo 2	Nova Iguaçu 5	
##	Paracambi	Paraíba do Sul	
##	raracambi	rararba do Sur	
##	Parati	Paty do Alferes	
##	2	1	
##	Petrópolis	Pinheiral	
##	2	4	
##	Piraí	Porciúncula	
##	1	3	
##	Porto Real	Quatis	
##	4	3	
##	Queimados	Quissamã	
##	4	5	
##	Resende	Rio Bonito	
##	3	3	
##	Rio Claro	Rio das Flores	
##	2	3	
##	Rio das Ostras	Rio de Janeiro	
##	Santa Mania Madalana	4	
##	Santa Maria Madalena 2	Santo Antônio de Pádua	
## ##	São Francisco de Itabapoana	4 São Fidelis	
##	5ao Francisco de Itabapoana	Sao Fidelis	
##	5	3	

```
##
                      São Gonçalo
                                               São João da Barra
##
##
              São João de Meriti
                                                 São José de Ubá
##
##
  São José do Vale do Rio Preto
                                             São Pedro da Aldeia
##
##
           São Sebastião do Alto
                                                         Sapucaia
##
##
                        Saquarema
                                                       Seropédica
##
                                 5
                                                                5
##
                     Silva Jardim
                                                        Sumidouro
##
##
                           Tanguá
                                                      Teresópolis
##
##
                Trajano de Moraes
                                                        Três Rios
##
##
                                                        Varre-Sai
                          Valença
##
                                3
                                                                 2
##
                        Vassouras
                                                    Volta Redonda
##
##
## Within cluster sum of squares by cluster:
   [1] 0.5133343 1.4384234 0.8561053 0.8863680 1.9062413
    (between_SS / total_SS = 73.8 %)
##
##
## Available components:
##
## [1] "cluster"
                       "centers"
                                       "totss"
                                                       "withinss"
                                                                       "tot.withinss"
## [6] "betweenss"
                       "size"
                                       "iter"
                                                       "ifault"
```

Método de Elbow para identificação do número ótimo de clusters

```
fviz_nbclust(df_padronizado[ ,7:11], kmeans, method = "wss", k.max = 10) +
  geom_vline(xintercept = 5, linetype = 2)+
  labs(title = "Número ideal de clusters",
        subtitle = "Método Elbow",
        x = "Número de clusters K",
        y = "Total dentro da soma do quadrado")
```



Criando variável categórica para indicação do cluster no banco de dados

```
df_padronizado$cluster_K_p <- factor(cluster_kmeans_p$cluster)</pre>
```

Análise de variância de um fator (ANOVA)

Análise de variância de um fator (ANOVA). Interpretação do output:

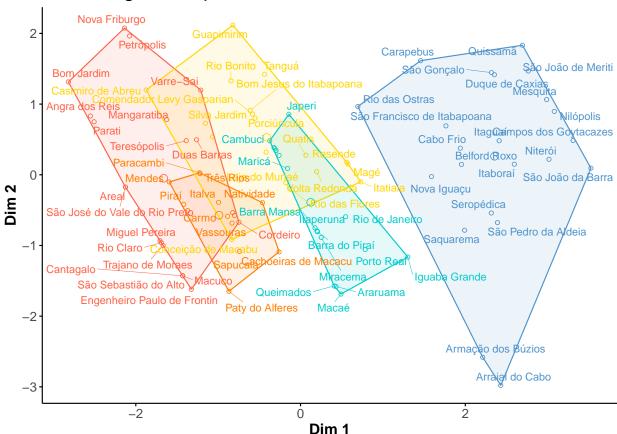
Mean Sq do cluster_K: indica a variabilidade entre grupos Mean Sq dos Residuals: indica a variabilidade dentro dos grupos F value: estatística de teste (Sum Sq do cluster_H / Sum Sq dos Residuals) Pr(>F): p-valor da estatística p-valor < 0.05: pelo menos um cluster apresenta média estatisticamente diferente dos demais

ANOVA da variável 'precipitação'

```
summary(anova_geom <- aov(formula =</pre>
                       Geologia ~ cluster_K_p,data = df_padronizado))
              Df Sum Sq Mean Sq F value Pr(>F)
## cluster_K_p 4 43.56 10.890 19.97 1.08e-11 ***
## Residuals 87 47.44
                         0.545
## ---
## Signif. codes: 0 '*** 0.001 '** 0.01 '* 0.05 '.' 0.1 ' ' 1
ANOVA da variável 'Geomorfologia'
summary(anova_geomor <- aov(formula =</pre>
                         Geomorfologia ~ cluster_K_p,data = df_padronizado))
              Df Sum Sq Mean Sq F value Pr(>F)
## cluster_K_p 4 76.83 19.208 117.9 <2e-16 ***
## Residuals 87 14.17 0.163
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1
ANOVA da variável 'Pedologia_Vegetacao'
summary(anova_ped <- aov(formula =</pre>
                   Pedologia_Vegetacao ~ cluster_K_p,data = df_padronizado))
##
              Df Sum Sq Mean Sq F value Pr(>F)
## cluster_K_p 4 9.31 2.3280 2.479 0.0498 *
## Residuals 87 81.69 0.9389
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1
ANOVA da variável 'Declividade'
summary(anova_decl <- aov(formula =</pre>
                       Declividade ~ cluster_K_p,data = df_padronizado))
              Df Sum Sq Mean Sq F value Pr(>F)
##
## cluster_K_p 4 80.67 20.167 169.8 <2e-16 ***
## Residuals 87 10.33
                         0.119
## Signif. codes: 0 '*** 0.001 '** 0.01 '* 0.05 '.' 0.1 ' ' 1
Scatter plot
fviz_cluster(cluster_kmeans_p, data = df_padronizado[ ,7:11],
            palette = c("darkorange1", "tomato", "gold1", "cyan3", "steelblue3"),
            geom = c("point", "text"),
            repel = T,
            ellipse.type = "convex",
            ggtheme = theme_classic(),
```

```
main = "Clusters gerados pelo 'K-means'",
    labelsize = 11,
    xlab = "Dim 1",
    ylab = "Dim 2",
    shape = 1,
    ellipse.alpha = 0.1,
    legend = "none"
    ) + theme(axis.text=element_text(size=12),
    axis.title=element_text(size=14,face="bold"),
    title = element_text(size = 16))
```

Clusters gerados pelo 'K-means'



Classificando as cidades com o grau de risco no dataset

```
#Escala númerica

df_padronizado$grau_risco_deslizamento[df_padronizado$cluster_K_p == 5] <- 1

df_padronizado$grau_risco_deslizamento[df_padronizado$cluster_K_p == 4] <- 2

df_padronizado$grau_risco_deslizamento[df_padronizado$cluster_K_p == 3] <- 3

df_padronizado$grau_risco_deslizamento[df_padronizado$cluster_K_p == 1] <- 4

df_padronizado$grau_risco_deslizamento[df_padronizado$cluster_K_p == 2] <- 5

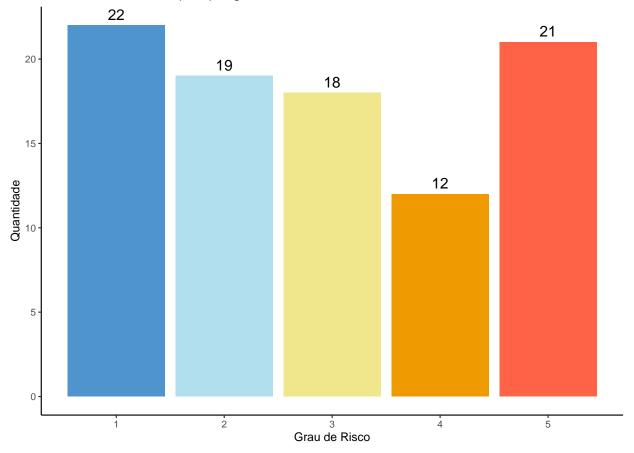
#Escala descritiva
```

```
df_padronizado$risco[df_padronizado$grau_risco_deslizamento == 1] <- "Muito Baixo"
df_padronizado$risco[df_padronizado$grau_risco_deslizamento == 2] <- "Baixo"
df_padronizado$risco[df_padronizado$grau_risco_deslizamento == 3] <- "Moderado"
df_padronizado$risco[df_padronizado$grau_risco_deslizamento == 4] <- "Alto"
df_padronizado$risco[df_padronizado$grau_risco_deslizamento == 5] <- "Muito Alto"
```

Graficos dos clusteres

Com peso

Quantidade de Municípios por grau de risco a deslizamentos de terra no RJ



Mapa

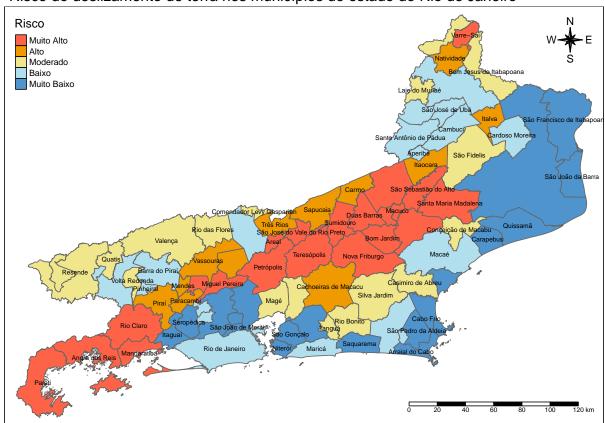
lendo o shapefile

Juntando os datasets

Plotando o mapa

```
tm_shape(shp = shp_rio) +
  tm_polygons("risco",
              palette = c("orange2", "lightblue2", "khaki", "tomato", "steelblue3"),
              legend.show = F) +
  tm layout(main.title = "Risco de deslizamento de terra nos municípios do estado do Rio de Janeiro",
           main.title.size = 1.2) +
  tm_text("MUNICN", size = 0.5,
         remove.overlap = TRUE,
          col = "black") +
  tm_compass(type = "8star",
            show.labels = 2,
             size = 2,
            position = c("right", "top")) +
  tm_add_legend("fill",
   labels = c("Muito Alto",'Alto','Moderado','Baixo', "Muito Baixo"),
   col = c("tomato", "orange2", "khaki", "lightblue2", "steelblue3"),
   title="Risco") +
  tm_scale_bar()
```

Risco de deslizamento de terra nos municípios do estado do Rio de Janeiro

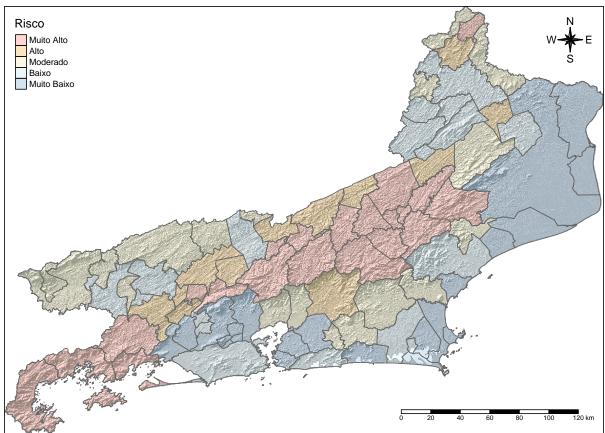


```
#labels = c("Muito Alto", "Alto", "Moderado", "Baixo", "Muito Baixo")
```

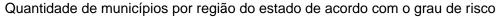
Mapa com relevo sombrado

```
#title = "Risco de deslizamento de terra",
            #style = "pretty",
            palette = c("orange2", "lightblue2", "khaki", "tomato", "steelblue3"),
            legend.show = F) +
tm_borders() +
tm_layout(main.title = "Risco de deslizamento de terra sobre o relevo sombreado do RJ",
          title.size = 0.5) +
tm_compass(type = "8star",
          show.labels = 2,
          size = 2,
          position = c("right", "top")) +
tm_add_legend("fill",
  labels = c("Muito Alto",'Alto','Moderado','Baixo', "Muito Baixo"),
  col = alpha(c("tomato", "orange2", "khaki", "lightblue2", "steelblue3"), 0.25),
  title="Risco") +
tm_scale_bar()
```

Risco de deslizamento de terra sobre o relevo sombreado do RJ



Quantidade de municípios em cada grau de risco por região do estado



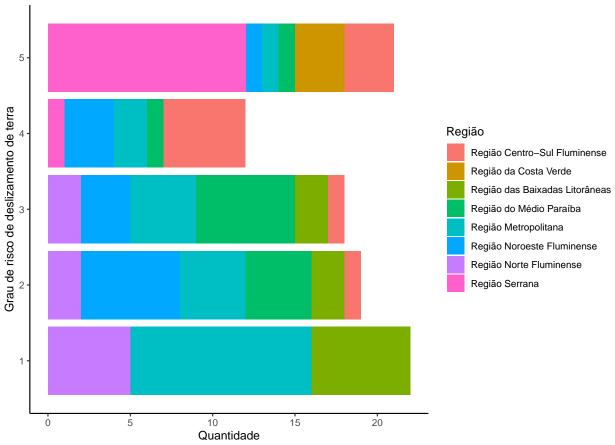
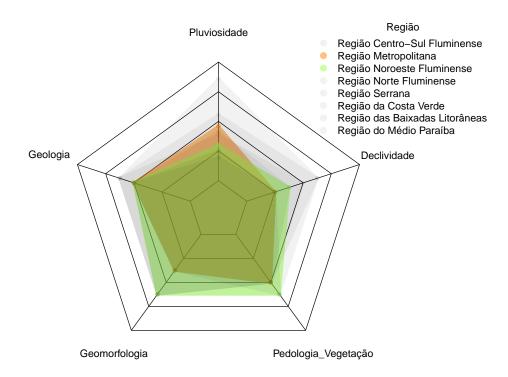


Gráfico de radar com regiões do estado

```
set.seed(99)
#criando o dataset para o grafico
df_reg <- left_join(x = df_tcc,</pre>
                    y = shp_rio@data,
                    by =c("nome" = "MUNICN"))
df_reg <- df_reg %>%
  group_by(REGGOV) %>%
  summarise(Pluviosidade = round(mean(Precipitacao_media_anual_mm.x)),
            Geologia = round(mean(Geologia.x)),
            Geomorfologia = round(mean(Geomorfologia.x)),
            Pedologia_Vegetação = round(mean(Pedologia_Vegetacao.x)),
            Declividade = round(mean(Declividade.x)))
df_reg[9, 1] <- "Max."
df_reg[9, 2] <- 2000
df_reg[10, 1] <- "Min."
df_reg[10, 2] <- 1000
```

```
df_reg[9, 3:6] <- 8
df_reg[10, 3:6] <- 0
df_{reg2} \leftarrow df_{reg[c(9,10,1:8),]}
rownames(df_reg2) <- df_reg2$REGGOV</pre>
areas \leftarrow c(rgb(0, 0.02, 0, 0.05),
           rgb(1, 0.5, 0, 0.5),
           rgb(0.6, 1, 0.3, 0.5),
           rgb(0, 0.02, 0, 0.05),
           rgb(0, 0.02, 0, 0.05),
           rgb(0, 0.02, 0, 0.05),
           rgb(0, 0.02, 0, 0.05),
           rgb(0, 0.02, 0, 0.05))
radarchart(df_reg2[ ,2:6],
                        # Tipo de línea del grid
           cglty = 1,
           cglcol = "black", # Color del grid
           pcol = areas, # Color para cada línea
           plwd = 4,
                          # Ancho para cada línea
           plty = 1,
           pfcol = areas,
           #title = "Região com maior e menor risco de deslizamento
                  #de terra explicitada por cada variável"
legend("topright",
       legend = df_reg2$REGGOV[3:10],
       bty = "n", pch = 20,
       text.col = "black", pt.cex = 2, col = areas, title = "Região")
```



Dataset 2

```
df2 <- left_join(df_padronizado,</pre>
                 df_tcc,
                 by =c("mun"="nome"))
group_by(df2, grau_risco_deslizamento) %>%
  summarise(
    mean = mean(Precipitacao_media_anual_mm.y, na.rm = TRUE),
    sd = sd(Precipitacao_media_anual_mm.y, na.rm = TRUE),
    min = min(Precipitacao_media_anual_mm.y, na.rm = TRUE),
    max = max(Precipitacao_media_anual_mm.y, na.rm = TRUE))
## # A tibble: 5 x 5
##
     grau_risco_deslizamento mean
                                      sd
                                           min
                                                 max
##
                       <dbl> <dbl> <dbl> <dbl> <dbl> <
## 1
                           1 1244. 265. 797 1751
## 2
                           2 1149. 112. 893. 1427.
                           3 1647. 188. 1426 2081
## 3
## 4
                           4 1271. 147. 1033 1436.
## 5
                           5 1655. 330. 1202. 2174
```

```
group_by(df2, grau_risco_deslizamento) %>%
  summarise(
   mean = mean(Geologia.y, na.rm = TRUE),
   sd = sd(Geologia.y, na.rm = TRUE),
   min = min(Geologia.y, na.rm = TRUE),
   max = max(Geologia.y, na.rm = TRUE))
## # A tibble: 5 x 5
    grau_risco_deslizamento mean
                                      sd
                                           min
##
                       <dbl> <dbl> <dbl> <dbl> <dbl> <
## 1
                           1 3.55 0.671
## 2
                                                   5
                           2 4.16 0.375
## 3
                           3 4.22 0.428
## 4
                           4 4.92 0.289
                                                   5
## 5
                           5 4.62 0.498
group_by(df2, grau_risco_deslizamento) %>%
  summarise(
   mean = mean(Geomorfologia.y, na.rm = TRUE),
   sd = sd(Geomorfologia.y, na.rm = TRUE),
   min = min(Geomorfologia.y, na.rm = TRUE),
   max = max(Geomorfologia.y, na.rm = TRUE))
## # A tibble: 5 x 5
   grau_risco_deslizamento mean
                                      sd
                                           min
##
                       <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1
                           1 1.77 0.813
                                             1
                                                   4
## 2
                           2 4.79 0.419
                                             4
                                                   5
## 3
                           3 4.5 0.618
                                             3
                                                   5
## 4
                           4 4.83 0.577
                                             3
                                                   5
## 5
                           5 4.95 0.218
group_by(df2, grau_risco_deslizamento) %>%
  summarise(
   mean = mean(Pedologia_Vegetacao.y, na.rm = TRUE),
   sd = sd(Pedologia_Vegetacao.y, na.rm = TRUE),
   min = min(Pedologia_Vegetacao.y, na.rm = TRUE),
   max = max(Pedologia_Vegetacao.y, na.rm = TRUE))
## # A tibble: 5 x 5
##
    grau_risco_deslizamento mean
                                      sd
                                           min
##
                       <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1
                           1 4.14 0.560
                                             3
                                                   5
## 2
                           2 4.47 0.513
                                             4
                                                   5
## 3
                           3 4.5 0.514
                                             4
                                                   5
## 4
                           4 4.67 0.492
                                             4
                                                   5
## 5
                           5 4.38 0.498
group_by(df2, grau_risco_deslizamento) %>%
 summarise(
   mean = mean(Declividade.y, na.rm = TRUE),
```

```
sd = sd(Declividade.y, na.rm = TRUE),
min = min(Declividade.y, na.rm = TRUE),
max = max(Declividade.y, na.rm = TRUE))
```

```
## # A tibble: 5 x 5
## grau_risco_deslizamento mean sd min max
##
                   <dbl> <dbl> <dbl> <dbl> <dbl> <
## 1
                     1 1.32 0.477
                                  1 2
## 2
                      2 2.63 0.597
## 3
                     3 2.83 0.383
                                    2
                                         3
                                   3
                      4 3.5 0.522
## 4
                                         4
## 5
                      5 4.90 0.301 4 5
```

Radar 2

```
radar2 <- readxl::read_xlsx("radar2.xlsx", col_names = T)</pre>
rownames(radar2) <- radar2$Cluster</pre>
class(radar2$Declividade) <- "numeric"</pre>
areas2 <- c("steelblue3", "lightblue2", "khaki", "orange2", "tomato")</pre>
radarchart(radar2[ ,2:6],
                           # Tipo de línea del grid
           cglty = 1,
           cglcol = "gray", # Color del grid
           pcol = areas2,  # Color para cada linea
                      # Ancho para cada línea
           plwd = 3,
           plty = 1)
legend("topright",
       legend = radar2$Cluster[3:7],
       bty = "n", pch = 20,
      text.col = "black", pt.cex = 2, col = areas2,
      title = "Cluster")
```

