# Queimada

*Érika S. Machado*

*18 de dezembro de 2018*

# Bibliotecas utilizadas

```
#library(rvest)
library(tidyverse)
```

```
## -- Attaching packages -------------- tidyverse 1.2.1 --
```

```
## v ggplot2 3.1.0     v purrr   0.2.5
## v tibble  1.4.2     v dplyr   0.7.5
## v tidyr   0.8.1     v stringr 1.3.1
## v readr   1.1.1     v forcats 0.3.0
```

```
## Warning: package 'ggplot2' was built under R version 3.5.1
```

```
## -- Conflicts ----------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(knitr)
library(reshape2)
```

```
##
## Attaching package: 'reshape2'
```

```
## The following object is masked from 'package:tidyr':
##
##     smiths
```

```
library(rgdal)
```

```
## Warning: package 'rgdal' was built under R version 3.5.1
```

```
## Loading required package: sp
```

```
## rgdal: version: 1.3-6, (SVN revision 773)
##  Geospatial Data Abstraction Library extensions to R successfully loaded
##  Loaded GDAL runtime: GDAL 2.2.3, released 2017/11/20
##  Path to GDAL shared files: C:/Users/ErikaS/Documents/R/win-library/3.5/rgdal/gdal
##  GDAL binary built with GEOS: TRUE
##  Loaded PROJ.4 runtime: Rel. 4.9.3, 15 August 2016, [PJ_VERSION: 493]
##  Path to PROJ.4 shared files: C:/Users/ErikaS/Documents/R/win-library/3.5/rgdal/proj
##  Linking to sp version: 1.3-1
```

```
library(gdata)
```

```
## gdata: Unable to locate valid perl interpreter
## gdata:
## gdata: read.xls() will be unable to read Excel XLS and XLSX files
## gdata: unless the 'perl=' argument is used to specify the location
## gdata: of a valid perl intrpreter.
## gdata:
## gdata: (To avoid display of this message in the future, please
## gdata: ensure perl is installed and available on the executable
## gdata: search path.)
```

```
## gdata: Unable to load perl libaries needed by read.xls()
## gdata: to support 'XLX' (Excel 97-2004) files.
```

```
##
```

```
## gdata: Unable to load perl libaries needed by read.xls()
## gdata: to support 'XLSX' (Excel 2007+) files.
```

```
##
```

```
## gdata: Run the function 'installXLSXsupport()'
## gdata: to automatically download and install the perl
## gdata: libaries needed to support Excel XLS and XLSX formats.
```

```
##
## Attaching package: 'gdata'
```

```
## The following objects are masked from 'package:dplyr':
##
##     combine, first, last
```

```
## The following object is masked from 'package:purrr':
##
##     keep
```

```
## The following object is masked from 'package:stats':
##
##     nobs
```

```
## The following object is masked from 'package:utils':
##
##     object.size
```

```
## The following object is masked from 'package:base':
##
##     startsWith
```

```
library(factoextra)
```

```
## Warning: package 'factoextra' was built under R version 3.5.1
```

```
## Welcome! Related Books: `Practical Guide To Cluster Analysis in R` at https://goo.gl/13EFCZ
```

```
#library(leaflet)
#library(scales)
```

# O estudo

Nesse estudo faremos algumas análises da taxa PRODES, que são taxas anuais de desmatamento por região, que são usadas pelo governo brasileiro para estabelecimento de políticas públicas. Os dados são de 2004 a 2017, refelentes aos estados: AC, AM, AP, MA, MT, PA, RO, RR, TO, além da Amazônia Legal.

### Importando os dados

```
Dados = read.table("taxaprodes.txt", header = T)
kable(Dados)
```

| Ano.Estados | AC | AM | AP | MA | MT | PA | RO | RR | TO | AMZLEGAL |
|---:|---:|---:|---:|---:|---:|---:|---:|---:|---:|---:|
| 2004 | 728 | 1232 | 46 | 755 | 11814 | 8870 | 3858 | 311 | 158 | 27772 |
| 2005 | 592 | 775 | 33 | 922 | 7145 | 5899 | 3244 | 133 | 271 | 19014 |
| 2006 | 398 | 788 | 30 | 674 | 4333 | 5659 | 2049 | 231 | 124 | 14286 |
| 2007 | 184 | 610 | 39 | 631 | 2678 | 5526 | 1611 | 309 | 63 | 11651 |
| 2008 | 254 | 604 | 100 | 1271 | 3258 | 5607 | 1136 | 574 | 107 | 12911 |
| 2009 | 167 | 405 | 70 | 828 | 1049 | 4281 | 482 | 121 | 61 | 7464 |
| 2010 | 259 | 595 | 53 | 712 | 871 | 3770 | 435 | 256 | 49 | 7000 |
| 2011 | 280 | 502 | 66 | 396 | 1120 | 3008 | 865 | 141 | 40 | 6418 |
| 2012 | 305 | 523 | 27 | 269 | 757 | 1741 | 773 | 124 | 52 | 4571 |
| 2013 | 221 | 583 | 23 | 403 | 1139 | 2346 | 932 | 170 | 74 | 5891 |
| 2014 | 309 | 500 | 31 | 257 | 1075 | 1887 | 684 | 219 | 50 | 5012 |
| 2015 | 264 | 712 | 25 | 209 | 1601 | 2153 | 1030 | 156 | 57 | 6207 |
| 2016 | 372 | 1129 | 17 | 258 | 1489 | 2992 | 1376 | 202 | 58 | 7893 |

| Ano.Estados | AC | AM | AP | MA | MT | PA | RO | RR | TO | AMZLEGAL |
|---|---|---|---|---|---|---|---|---|---|---|
| 2017 | 257 | 1001 | 24 | 265 | 1561 | 2433 | 1243 | 132 | 31 | 6947 |

## Estatística descritiva

```
summary(Dados[,2:11])
```

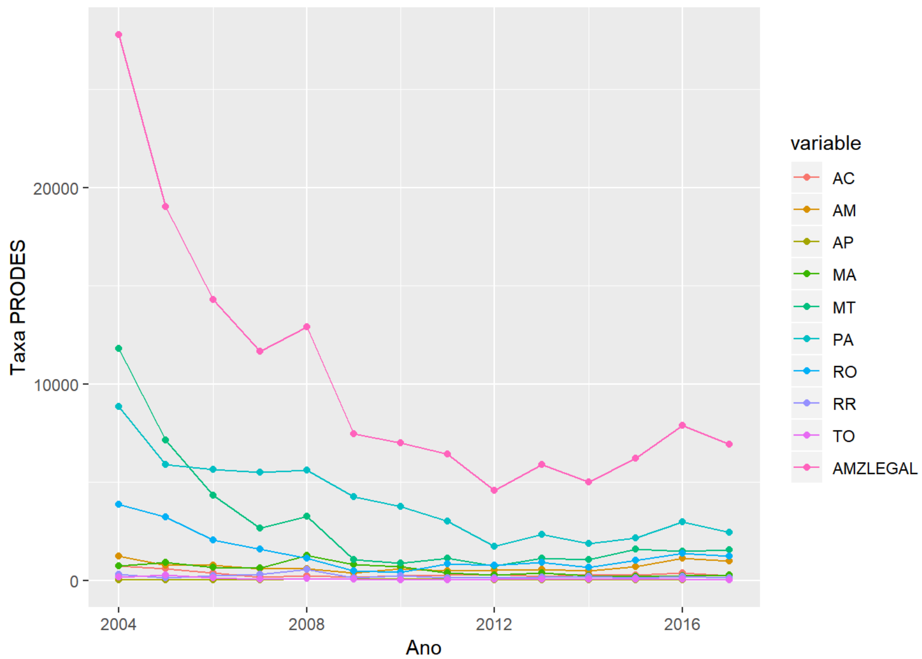
```
##        AC              AM              AP               MA        
##  Min.   :167.0   Min.   : 405.0   Min.   : 17.00   Min.   : 209.0  
##  1st Qu.:254.8   1st Qu.: 538.0   1st Qu.: 25.50   1st Qu.: 266.0  
##  Median :272.0   Median : 607.0   Median : 32.00   Median : 517.0  
##  Mean   :327.9   Mean   : 711.4   Mean   : 41.71   Mean   : 560.7  
##  3rd Qu.:356.2   3rd Qu.: 784.8   3rd Qu.: 51.25   3rd Qu.: 744.2  
##  Max.   :728.0   Max.   :1232.0   Max.   :100.00   Max.   :1271.0  
##        MT              PA             RO              RR       
##  Min.   :  757   Min.   :1741   Min.   : 435   Min.   :121.0  
##  1st Qu.: 1086   1st Qu.:2368   1st Qu.: 796   1st Qu.:135.0  
##  Median : 1525   Median :3389   Median :1083   Median :186.0  
##  Mean   : 2849   Mean   :4012   Mean   :1408   Mean   :219.9  
##  3rd Qu.: 3113   3rd Qu.:5587   3rd Qu.:1552   3rd Qu.:249.8  
##  Max.   :11814   Max.   :8870   Max.   :3858   Max.   :574.0  
##        TO            AMZLEGAL    
##  Min.   : 31.00   Min.   : 4571  
##  1st Qu.: 50.50   1st Qu.: 6260  
##  Median : 59.50   Median : 7232  
##  Mean   : 85.36   Mean   :10217  
##  3rd Qu.: 98.75   3rd Qu.:12596  
##  Max.   :271.00   Max.   :27772
```

Vemos que os maiores índices de ocorrência de queimadas são na Amazônia Legal e Mato Grosso.

```
a = melt(Dados[,2:11]) %>% mutate(Ano = rep(2004:2017,10))
```

```
## No id variables; using all as measure variables
```

```
ggplot(a, aes(x = Ano, y = value, fill = variable,  colour=variable))  +
  geom_line(aes(group=variable))+
  geom_point(aes(group=variable)) +
  xlab("Ano") +
  ylab("Taxa PRODES") +
  scale_colour_discrete()
```




Outra informação importante que devemos perceber é que houve uma significativa diminutição das queimadas na Amazônia Legal no decorrer dos anos. Provavelmente por algum tipo de política de prevenção, fiscalização e cuidados.

```
kable(cov(Dados[,2:11]), caption = "Covariância")
```

Covariância

| | AC | AM | AP | MA | MT | PA | RO | RR | TO | AMZL |
|---|---|---|---|---|---|---|---|---|---|---|
| AC | 24400.4396 | 26170.593 | -618.65934 | 10178.264 | 438763.736 | 207636.04 | 142811.527 | 735.8352 | 7547.82418 | 8576 |
| AM | 26170.5934 | 62227.786 | -2176.27473 | -6805.736 | 478333.429 | 216389.27 | 171776.066 | 2307.9505 | 4750.32418 | 9529 |
| AP | -618.6593 | -2176.275 | 541.45055 | 5533.066 | 3981.088 | 17388.24 | -3364.099 | 1733.1319 | 42.87912 | 230 |
| MA | 10178.2637 | -6805.736 | 5533.06593 | 101308.681 | 453473.088 | 480947.24 | 107643.593 | 25181.8242 | 11647.10989 | 11891 |
| MT | 438763.7363 | 478333.429 | 3981.08791 | 453473.088 | 9760094.681 | 5641311.37 | 3026289.868 | 102872.2527 | 154423.12088 | 200595 |
| PA | 207636.0440 | 216389.275 | 17388.24176 | 480947.242 | 5641311.374 | 4252834.37 | 1661298.560 | 124546.4835 | 88582.12088 | 126909 |
| RO | 142811.5275 | 171776.066 | -3364.09890 | 107643.593 | 3026289.868 | 1661298.56 | 1024894.725 | 16643.3407 | 52959.14286 | 62009 |
| RR | 735.8352 | 2307.951 | 1733.13187 | 25181.824 | 102872.253 | 124546.48 | 16643.341 | 14648.2253 | 933.71978 | 2896 |
| TO | 7547.8242 | 4750.324 | 42.87912 | 11647.110 | 154423.121 | 88582.12 | 52959.143 | 933.7198 | 4090.24725 | 3249 |
| AMZLEGAL | 857625.6044 | 952973.412 | 23060.82418 | 1189107.132 | 20059542.637 | 12690933.71 | 6200952.725 | 289602.7637 | 324976.48901 | 425887 |

```
kable(cor(Dados[,2:11]), caption = "Correlação")
```

Correlação

| | AC | AM | AP | MA | MT | PA | RO | RR | TO | AMZLEGAL |
|---|---|---|---|---|---|---|---|---|---|---|
| AC | 1.0000000 | 0.6716184 | -0.1702054 | 0.2047159 | 0.8990936 | 0.6445620 | 0.9030771 | 0.0389215 | 0.7555238 | 0.8413009 |
| AM | 0.6716184 | 1.0000000 | -0.3749231 | -0.0857155 | 0.6137783 | 0.4206340 | 0.6801910 | 0.0764437 | 0.2977531 | 0.5853843 |
| AP | -0.1702054 | -0.3749231 | 1.0000000 | 0.7470731 | 0.0547640 | 0.3623573 | -0.1428071 | 0.6154031 | 0.0288132 | 0.1518616 |
| MA | 0.2047159 | -0.0857155 | 0.7470731 | 1.0000000 | 0.4560381 | 0.7327146 | 0.3340606 | 0.6536893 | 0.5721634 | 0.5724665 |
| MT | 0.8990936 | 0.6137783 | 0.0547640 | 0.4560381 | 1.0000000 | 0.8756163 | 0.9568500 | 0.2720688 | 0.7728773 | 0.9838897 |
| PA | 0.6445620 | 0.4206340 | 0.3623573 | 0.7327146 | 0.8756163 | 1.0000000 | 0.7957357 | 0.4989991 | 0.6716330 | 0.9429893 |
| RO | 0.9030771 | 0.6801910 | -0.1428071 | 0.3340606 | 0.9568500 | 0.7957357 | 1.0000000 | 0.1358340 | 0.8179496 | 0.9385798 |
| RR | 0.0389215 | 0.0764437 | 0.6154031 | 0.6536893 | 0.2720688 | 0.4989991 | 0.1358340 | 1.0000000 | 0.1206283 | 0.3666592 |
| TO | 0.7555238 | 0.2977531 | 0.0288132 | 0.5721634 | 0.7728773 | 0.6716330 | 0.8179496 | 0.1206283 | 1.0000000 | 0.7786272 |
| AMZLEGAL | 0.8413009 | 0.5853843 | 0.1518616 | 0.5724665 | 0.9838897 | 0.9429893 | 0.9385798 | 0.3666592 | 0.7786272 | 1.0000000 |

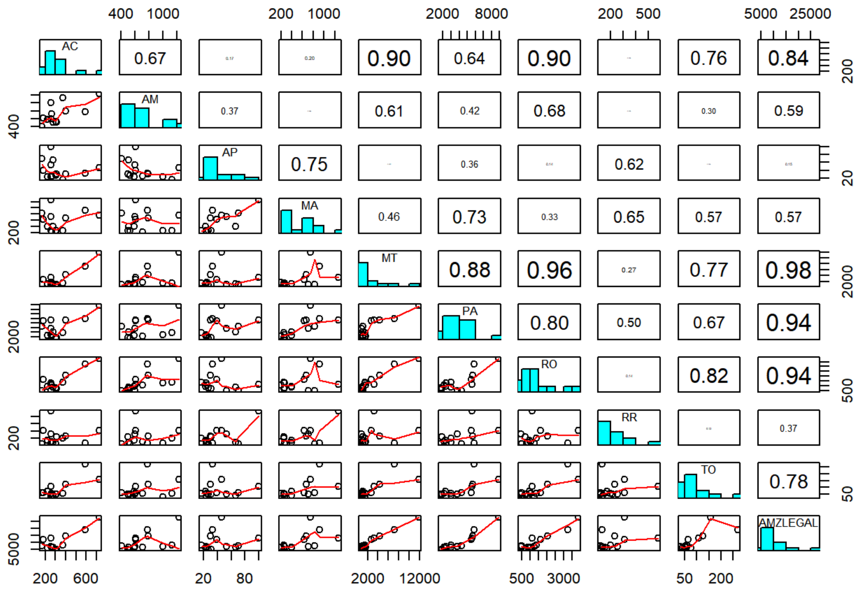
As correlações são maiores para os estados que são vizinhos uns dos outros.

```
#função retirada do help(pairs)
panel.hist <- function(x, ...)
{
  usr <- par("usr"); on.exit(par(usr))
  par(usr = c(usr[1:2], 0, 1.5) )
  h <- hist(x, plot = FALSE)
  breaks <- h$breaks; nB <- length(breaks)
  y <- h$counts; y <- y/max(y)
  rect(breaks[-nB], 0, breaks[-1], y, col = "cyan", ...)
}

#função retirada do help(pairs)
panel.cor <- function(x, y, digits = 2, prefix = "", cex.cor, ...)
{
  usr <- par("usr"); on.exit(par(usr))
  par(usr = c(0, 1, 0, 1))
  r <- abs(cor(x, y))
  txt <- format(c(r, 0.123456789), digits = digits)[1]
  txt <- paste0(prefix, txt)
  if(missing(cex.cor)) cex.cor <- 0.8/strwidth(txt)
  text(0.5, 0.5, txt, cex = cex.cor * r)
}

panel.lm <- function (x, y, col = par("col"), bg = NA, pch = par("pch"),
    cex = 1, col.line="red") {
    points(x, y, pch = pch, col = col, bg = bg, cex = cex)
    ok <- is.finite(x) & is.finite(y)
    if (any(ok)) {
       abline(lm(y[ok]~x[ok]), col = col.line)
    }
}

pairs(Dados[,2:11], diag.panel = panel.hist, upper.panel = panel.cor,
      lower.panel = panel.smooth)
```




Nesse conjunto de gráficos e correlações dá pra ver melhor o que está acontecendo e que muitas variáveis são correlacionadas.

# Análise de Componentes principais

Como as variáveis possuem correlações fortes, iremos padronizar os dados e fazer análise de componentes principais para buscar a combinação linear que mais representa a variabilidade dos dados.

## Padronizando os valores

```
pdados = data.frame(Dados$Ano.Estados,scale(Dados[,-1]))
kable(pdados, caption = "Dados padronizados")
```

Dados padronizados

| Dados.Ano.Estados | AC | AM | AP | MA | MT | PA | RO | RR | TO | AMZLEGAL |
|---:|---:|---:|---:|---:|---:|---:|---:|---:|---:|---:|
| 2004 | 2.5616291 | 2.0871215 | 0.1841806 | 0.6104042 | 2.8695212 | 2.3555522 | 2.4196385 | 0.7524708 | 1.1358426 | 2.6900170 |
| 2005 | 1.6909861 | 0.2551276 | -0.3745005 | 1.1350826 | 1.3750180 | 0.9148858 | 1.8131413 | -0.7182407 | 2.9027088 | 1.3480020 |
| 2006 | 0.4490396 | 0.3072412 | -0.5034269 | 0.3559195 | 0.4749231 | 0.7985075 | 0.6327439 | 0.0914768 | 0.6042191 | 0.6235162 |
| 2007 | -0.9209427 | -0.4063144 | -0.1166477 | 0.2208227 | -0.0548270 | 0.7340146 | 0.2000961 | 0.7359459 | -0.3495759 | 0.2197471 |
| 2008 | -0.4728176 | -0.4303668 | 2.5048558 | 2.2315661 | 0.1308256 | 0.7732922 | -0.2690996 | 2.9254884 | 0.3384074 | 0.4128208 |
| 2009 | -1.0297731 | -1.2281060 | 1.2155918 | 0.8397546 | -0.5762546 | 0.1303021 | -0.9151079 | -0.8173898 | -0.3808479 | -0.4218396 |
| 2010 | -0.4408087 | -0.4664455 | 0.4850088 | 0.4753074 | -0.6332308 | -0.1174867 | -0.9615336 | 0.2980374 | -0.5684797 | -0.4929398 |
| 2011 | -0.3063712 | -0.8392582 | 1.0436899 | -0.5174971 | -0.5535282 | -0.4869878 | -0.5367881 | -0.6521413 | -0.7092036 | -0.5821214 |
| 2012 | -0.1463265 | -0.7550747 | -0.6323533 | -0.9165040 | -0.6697211 | -1.1013682 | -0.6276638 | -0.7926026 | -0.5215718 | -0.8651428 |
| 2013 | -0.6840766 | -0.5145503 | -0.8042552 | -0.4955046 | -0.5474465 | -0.8079979 | -0.4706068 | -0.4125310 | -0.1775801 | -0.6628752 |
| 2014 | -0.1207194 | -0.8472757 | -0.4604514 | -0.9542054 | -0.5679323 | -1.0305714 | -0.7155763 | -0.0076723 | -0.5528437 | -0.7975670 |
| 2015 | -0.4087998 | 0.0025770 | -0.7183042 | -1.1050112 | -0.3995646 | -0.9015855 | -0.3738043 | -0.5282050 | -0.4433918 | -0.6144536 |
| 2016 | 0.2825931 | 1.6742214 | -1.0621080 | -0.9510637 | -0.4354147 | -0.4947463 | -0.0320323 | -0.1481335 | -0.4277558 | -0.3561026 |
| 2017 | -0.4536123 | 1.1611027 | -0.7612797 | -0.9290711 | -0.4123682 | -0.7658108 | -0.1634071 | -0.7265032 | -0.8499274 | -0.5010611 |

## Componentes principais

```
cp = prcomp(pdados[,2:11])
cp
```

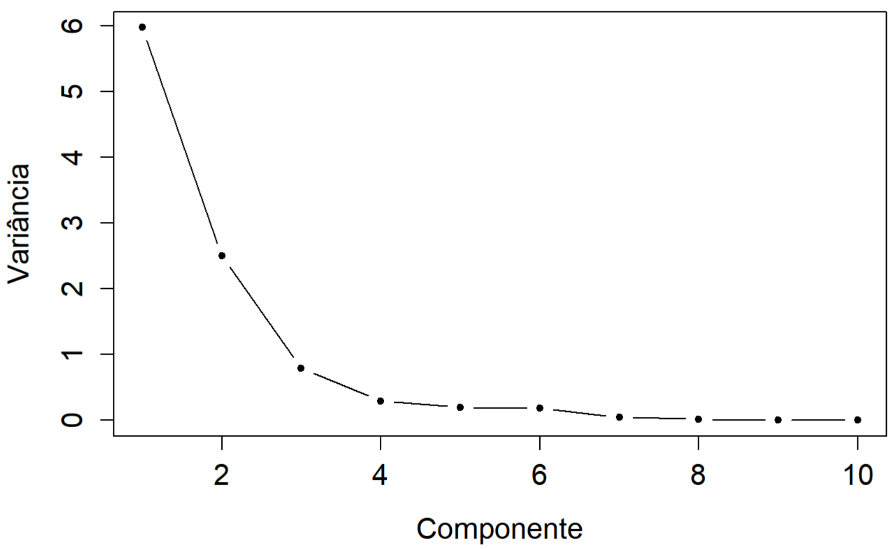
```
## Standard deviations (1, .., p=10):
##  [1] 2.444338e+00 1.581076e+00 8.885131e-01 5.419573e-01 4.435656e-01
##  [6] 4.236622e-01 2.144976e-01 1.296960e-01 5.627185e-02 7.344045e-17
##
## Rotation (n x k) = (10 x 10):
##                 PC1           PC2          PC3         PC4         PC5
## AC       -0.35305759  0.2434359809  0.066571640 -0.10741351  0.65944222
## AM       -0.24312872  0.3353050103 -0.601302052 -0.01367151 -0.02807591
## AP       -0.06032671 -0.5828562053  0.006308125 -0.52607799  0.39916865
## MA       -0.24335367 -0.4730112572  0.195739103  0.16209929 -0.21514964
## MT       -0.39870432  0.0767053683 -0.006916095 -0.19439443  0.05678123
## PA       -0.37664294 -0.1634991260 -0.054402917 -0.25401397 -0.51318361
## RO       -0.38552015  0.1869736774  0.047985494  0.03140582 -0.11585128
## RR       -0.15419334 -0.4431406658 -0.561012786  0.55681256  0.20476505
## TO       -0.33992645  0.0299974589  0.524114564  0.49931149  0.12866098
## AMZLEGAL -0.40571177 -0.0003698691 -0.030148480 -0.15030100 -0.14225793
##                 PC6         PC7         PC8         PC9         PC10
## AC       -0.07349938  0.52800415 -0.21045285 -0.19066444 -0.020540353
## AM        0.66712219 -0.06220882  0.02699434  0.12326000 -0.032802070
## AP        0.23433871 -0.36412306 -0.14542624  0.10884618 -0.003059768
## MA        0.42308360  0.34338139  0.30408013 -0.46371656 -0.041853560
## MT       -0.28542130 -0.12053084  0.70417697  0.18295016 -0.410805350
## PA       -0.15156728  0.30871749 -0.46895041  0.29805856 -0.271174164
## RO       -0.12955679 -0.55700365 -0.28710833 -0.61033925 -0.133121654
## RR       -0.32669855 -0.02879998 -0.06339053  0.03624156 -0.015914815
## TO        0.24716333 -0.21302127 -0.12630528  0.46957103 -0.008409765
## AMZLEGAL -0.16305489 -0.02346379  0.15226660  0.07028108  0.858137047
```

```
summary(cp)
```

```
## Importance of components:
##                           PC1    PC2     PC3     PC4     PC5     PC6
## Standard deviation     2.4443 1.5811 0.88851 0.54196 0.44357 0.42366
## Proportion of Variance 0.5975 0.2500 0.07895 0.02937 0.01968 0.01795
## Cumulative Proportion  0.5975 0.8475 0.92640 0.95578 0.97545 0.99340
##                           PC7     PC8     PC9      PC10
## Standard deviation     0.2145 0.12970 0.05627 7.344e-17
## Proportion of Variance 0.0046 0.00168 0.00032 0.000e+00
## Cumulative Proportion  0.9980 0.99968 1.00000 1.000e+00
```

A primeira componente principal responde por cerca de 59,75% da variância total dos dados padronizados. Se pegarmos as três primeiras componentes principais, conseguimos atingir 92,64% da variância total.

```
plot(1:10, cp$sdev^2, type = "b", xlab = "Componente",
 ylab = "Variância", pch = 20, cex.axis = 1.3, cex.lab = 1.3)
```




A seguir estão as correlações entre os dois primeiros componentes principais e as variáveis:

```
cp$sdev[1:2]*t(cp$rotation[,1:2])
```

```
##             AC         AM         AP         MA        MT         PA
## PC1 -0.8629920 -0.5942887 -0.1474589 -0.5948386 -0.974568 -0.9206426
## PC2  0.3848909  0.5301428 -0.9215402 -0.7478669  0.121277 -0.2585046
##             RO         RR          TO      AMZLEGAL
## PC1 -0.9423415 -0.3769006 -0.83089506 -0.9916966044
## PC2  0.2956197 -0.7006393  0.04742827 -0.0005847914
```

# Análise de agrupamentos

Em seguida faremos agrupamentos dos anos da amostra para verificar os anos que apresentam características comuns. Esse grupo de ser tratado como momentos importantes de transição no histórico de queimadas.

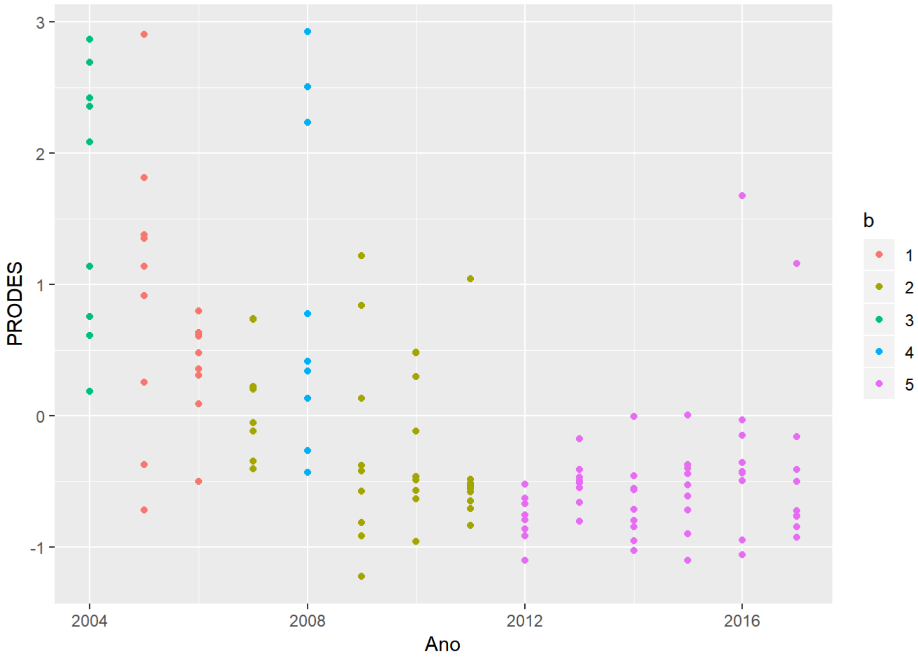
```
set.seed(20)
#tpdados = t(pdados[,2:11])
#colnames(tpdados) = 2004:2017
rownames(pdados) = 2004:2017
pdados = pdados[,-1]
kdados = kmeans(pdados,5)
d = reshape2::melt(pdados)
```

```
## No id variables; using all as measure variables
```

```
kdados
```

```
## K-means clustering with 5 clusters of sizes 2, 4, 1, 1, 6
##
## Cluster means:
##           AC         AM         AP         MA         MT          PA
## 1  1.0700129  0.2811844 -0.4389637  0.7455011  0.9249705  0.85669668
## 2 -0.6744739 -0.7350310  0.6569107  0.2545969 -0.4544601  0.06496057
## 3  2.5616291  2.0871215  0.1841806  0.6104042  2.8695212  2.35555220
## 4 -0.4728176 -0.4303668  2.5048558  2.2315661  0.1308256  0.77329224
## 5 -0.2551569  0.1201667 -0.7397920 -0.8918933 -0.5054079 -0.85034668
##           RO         RR         TO   AMZLEGAL
## 1  1.2229426 -0.3133820  1.7534639  0.9857591
## 2 -0.5533334 -0.1088869 -0.5020268 -0.3192884
## 3  2.4196385  0.7524708  1.1358426  2.6900170
## 4 -0.2690996  2.9254884  0.3384074  0.4128208
## 5 -0.3971818 -0.4359412 -0.4955118 -0.6328671
##
## Clustering vector:
## 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017
##    3    1    1    2    4    2    2    2    5    5    5    5    5    5
##
## Within cluster sum of squares by cluster:
## [1] 5.424747 6.922917 0.000000 0.000000 8.134388
##  (between_SS / total_SS =  84.2 %)
##
## Available components:
##
## [1] "cluster"      "centers"      "totss"        "withinss"
## [5] "tot.withinss" "betweenss"    "size"         "iter"
## [9] "ifault"
```

```
d = d[-c(1:14),]
Ano = rep(2004:2017, 9)
e = data.frame(d, Ano)
names(e) = c("UF", "PRODES", "Ano")
kdados$cluster <- as.factor(kdados$cluster)
b = kdados$cluster
Ano = factor(names(b))
b = data.frame(Ano, b)
tabelafinal = merge(e,b, by.x = "Ano", by.y = "Ano")
ggplot(tabelafinal, aes(Ano, PRODES, color = b)) + geom_point()
```




```
fviz_cluster(kdados, data = pdados)#1/sq
```