



Universidade de Brasília

DEPARTAMENTO DE ESTATÍSTICA

14 maio 2023

Exercício 29 da Lista 3 - Extra

Prof. Dr. George von Borries

Análise Multivariada 1

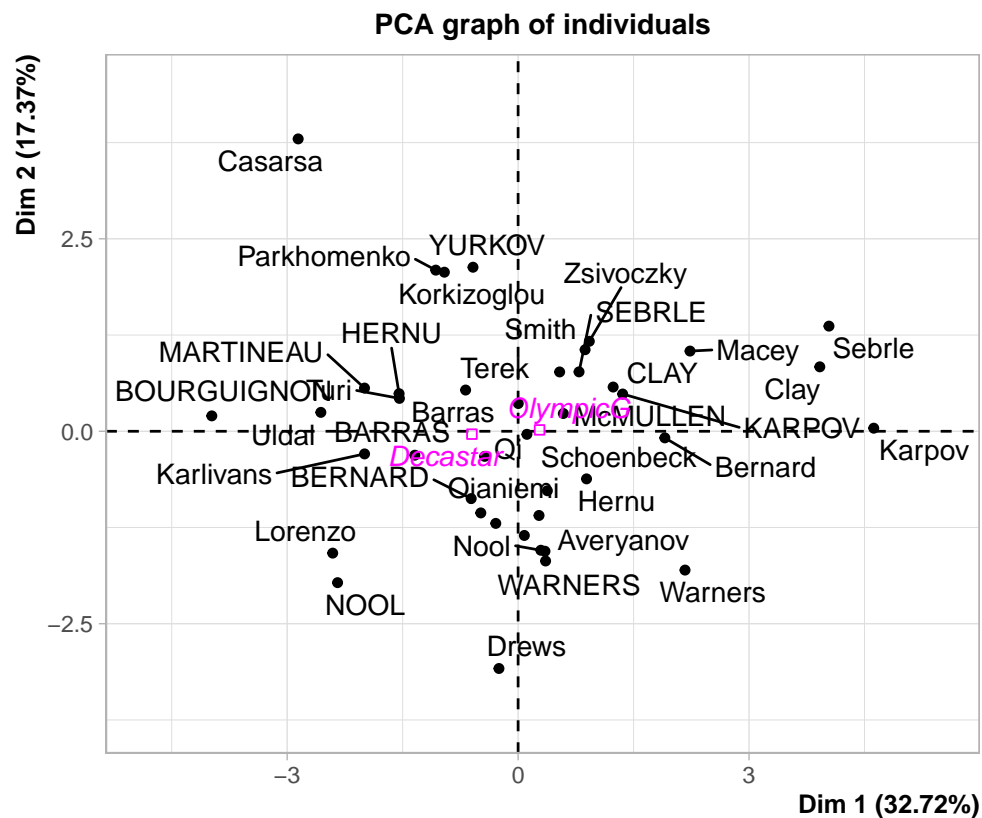
Aluno: Bruno Gondim Toledo | Matrícula: 15/0167636

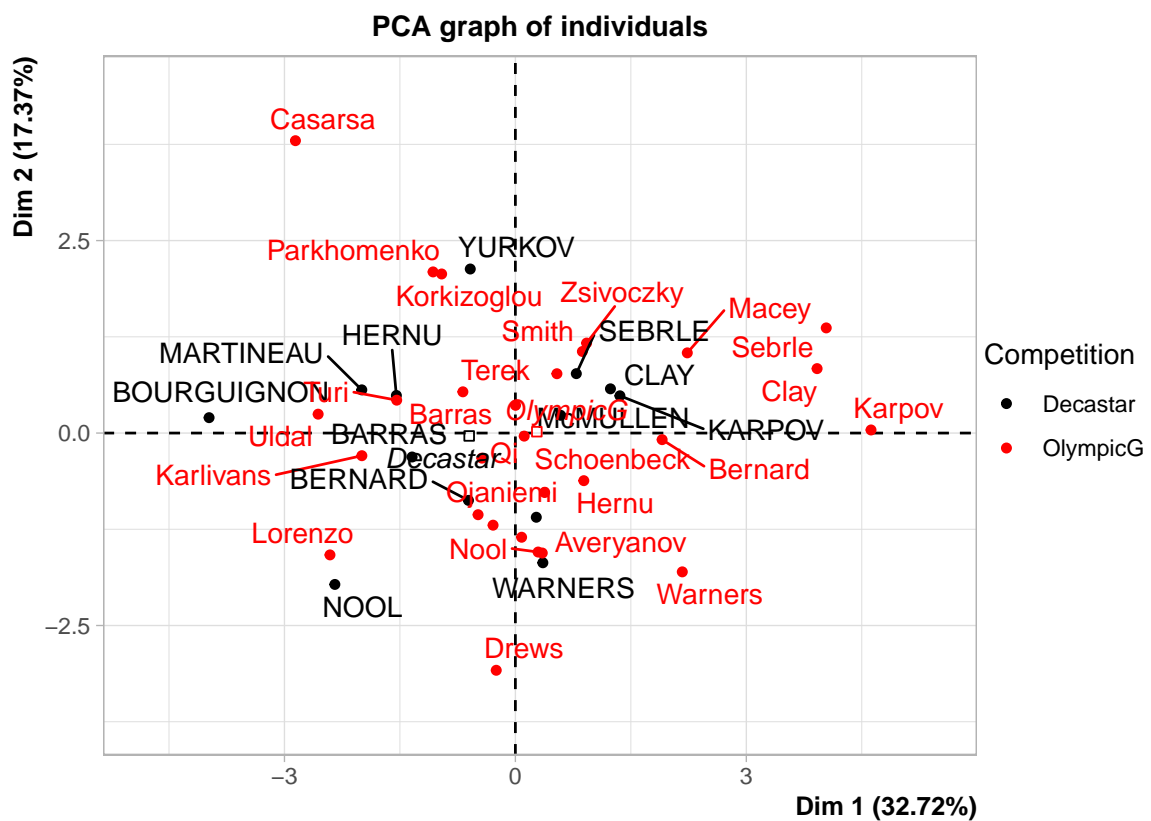
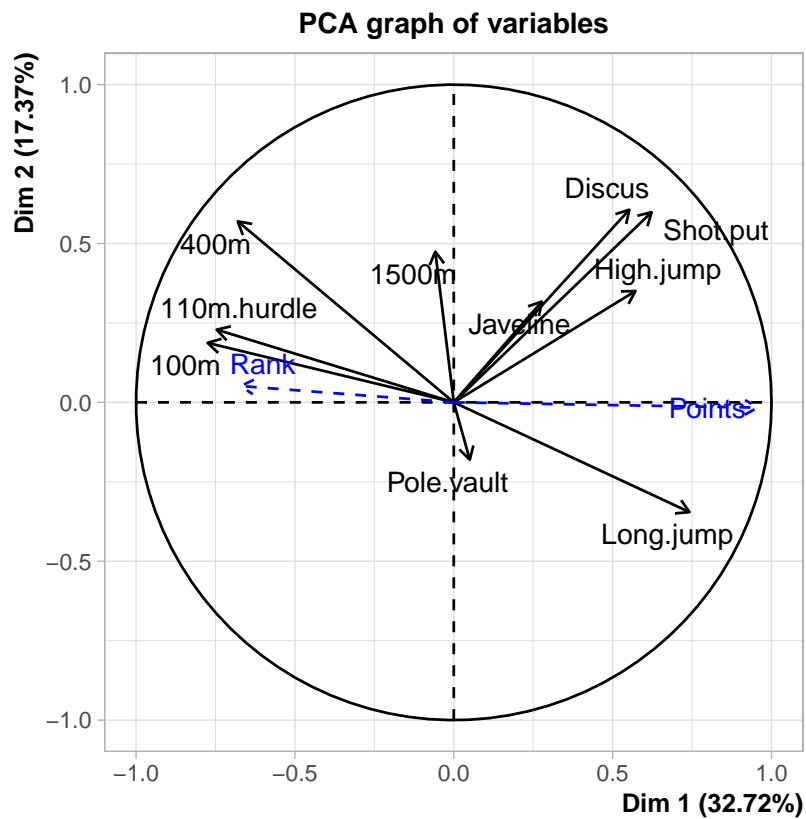
Artigo

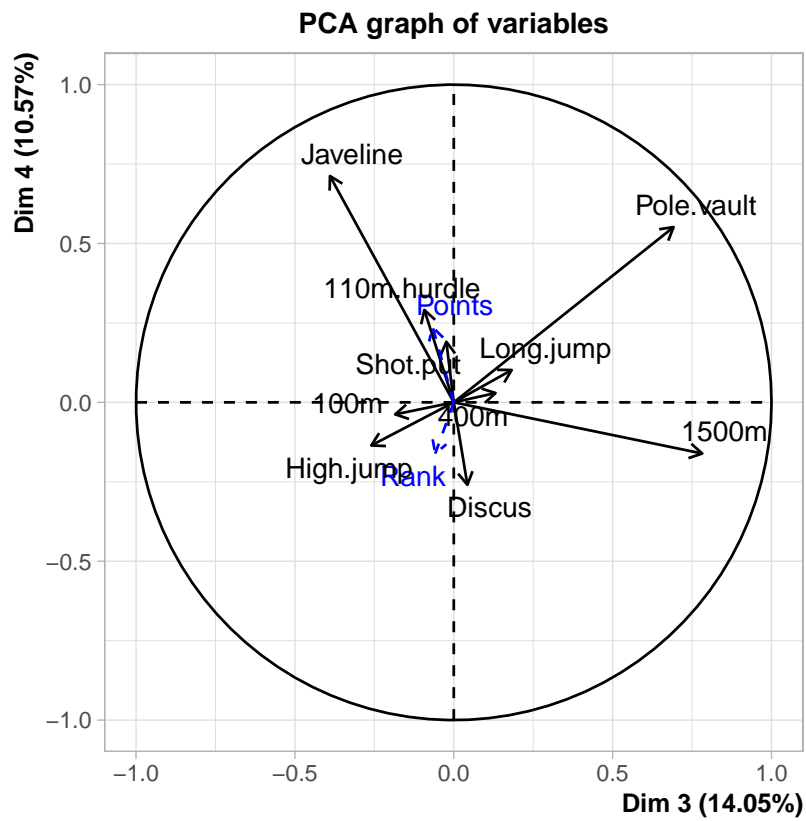
O artigo selecionado foi o “FactoMineR: An R Package for Multivariate Analysis” (LÊ; JOSSE; HUSSON, 2008)[1]. Este é o artigo que apresenta o pacote *FactoMineR*. O banco de dados utilizado nos gráficos nos exemplos é o Decathlon data (HUSSON; PAGÈS, 2005)[2]

Gráficos contidos no artigo

Primeiramente, os gráficos propostos para o banco de dados Decathlon, para uma análise de componentes principais.

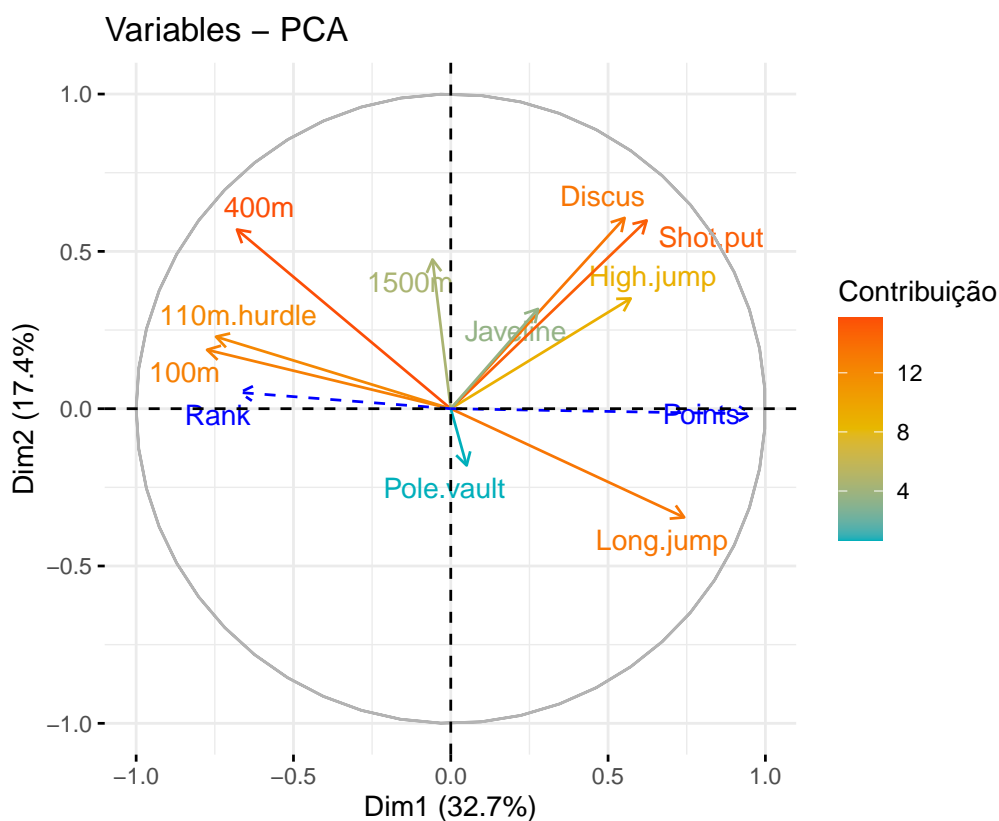
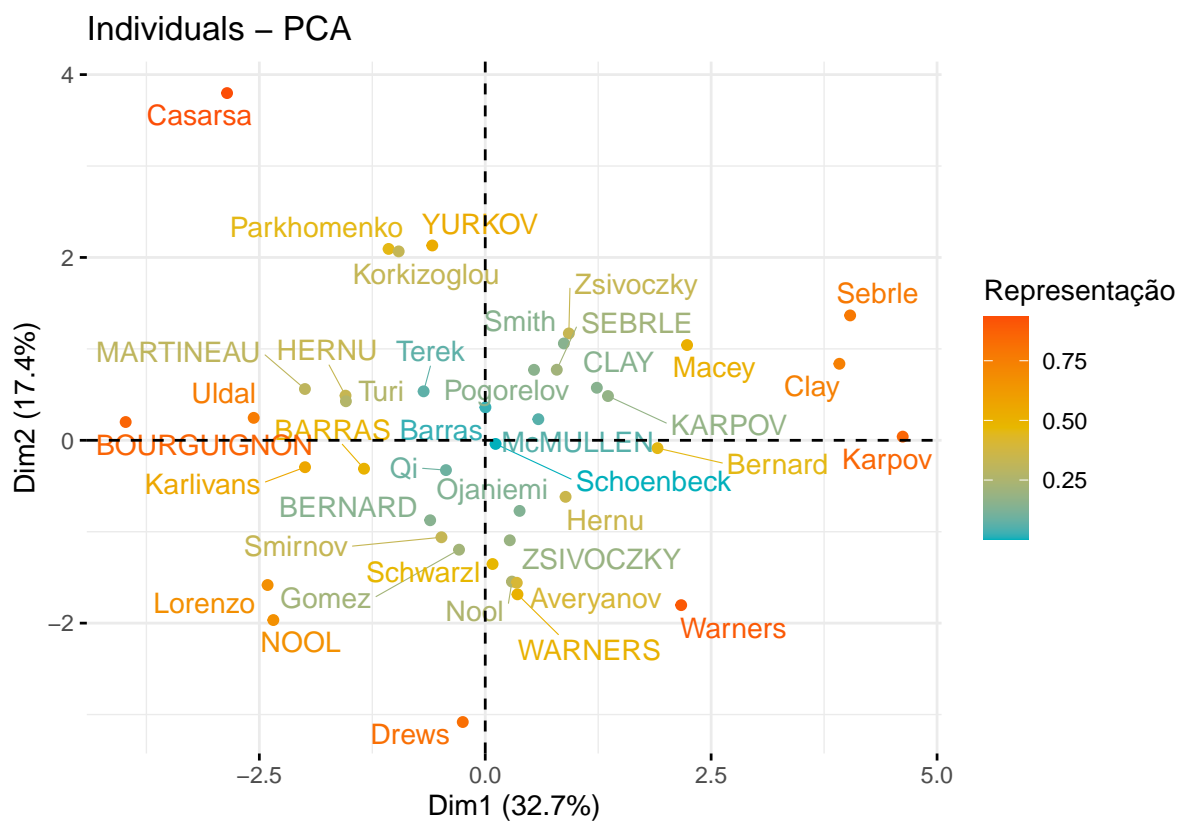






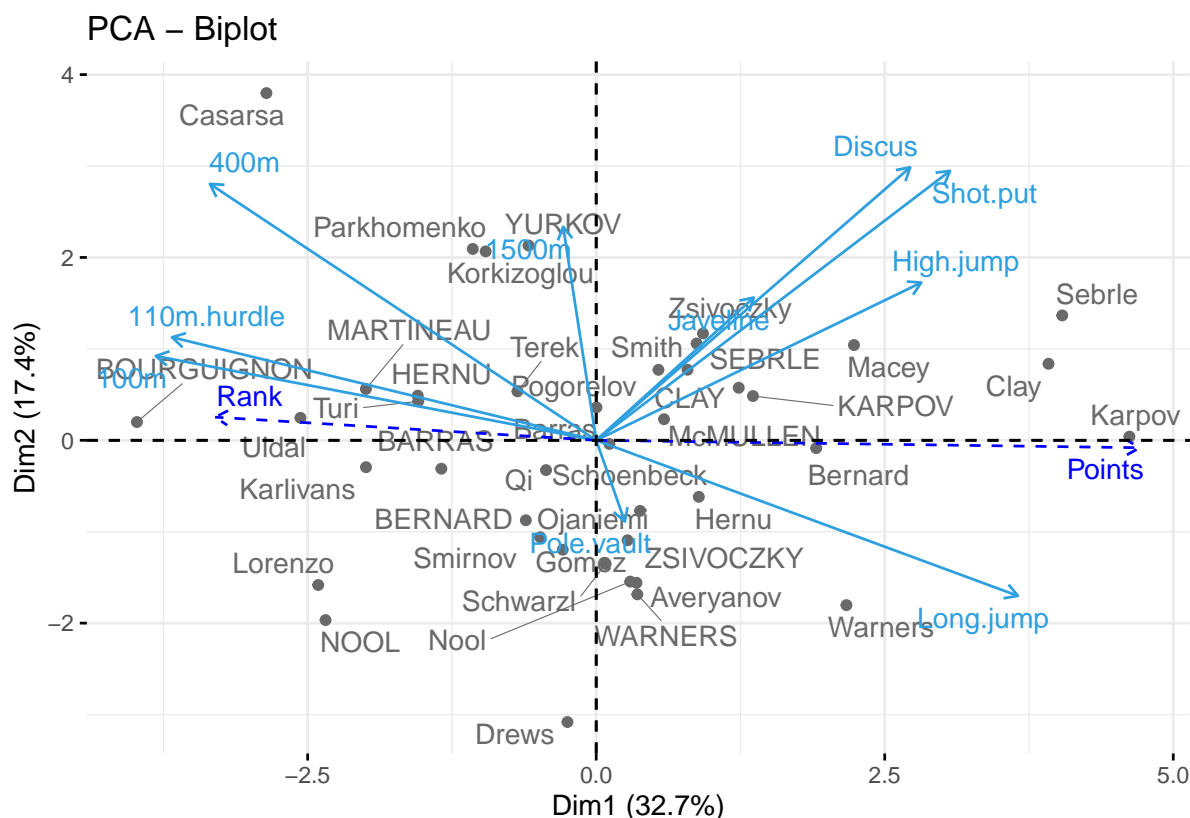
Gráficos - Solução alternativa

Uma possível alternativa seria utilizar os gráficos contidos no pacote *factoextra*, que trazem uma abordagem visual similar porém com algumas peculiaridades específicas do pacote



No caso desses dois gráficos, são praticamente idênticos aos produzidos pelo *FactoMineR*, porém esteticamente mais agradáveis.

É possível ainda fazer um *biplot* de *gabriel*



Em que adicionamos também juntamente às observações, os vetores das componentes que estão influenciando o comportamento nas dimensões apresentadas.

Comentários:

Com estes gráficos, conseguimos perceber semelhanças com um outro banco apresentado didaticamente pelo Prof. von Borries em aula, que versava sobre corrida em olimpíadas. Notamos que em geral é possível notar pela análise de componentes principais as diferenças de rendimento entre atletas de corridas de longa distância para atletas de corrida de curta distância. Neste caso específico, notamos a semelhança nas componentes de corridas de curta distância, e sua disparidade com outras modalidades divergentes como salto e arremesso de vara.

Desta forma, conseguimos visualizar claramente a possibilidade de redução de dimensionalidade neste banco de dados, visto que pelo último gráfico (*biplot* de *gabriel*), percebemos o quanto apenas duas dimensões estão sendo suficientes para explicar a variabilidade do conjunto de dados, além de permitir uma análise clara do quanto do rendimento de cada atleta está sendo influenciado por cada componente.

Referências:

[1] LÊ, Sébastien; JOSSE, Julie; HUSSON, François. FactoMineR: an R package for multivariate analysis. Journal of statistical software, v. 25, p. 1-18, 2008.

[2] Husson, François; PAGÈS, Jérôme. Statistiques Générales pour Utilisateurs. Presses Universitaires de Rennes (2005).