

Tarefa 3 - Classificação e Agrupamento

Tais Bellini

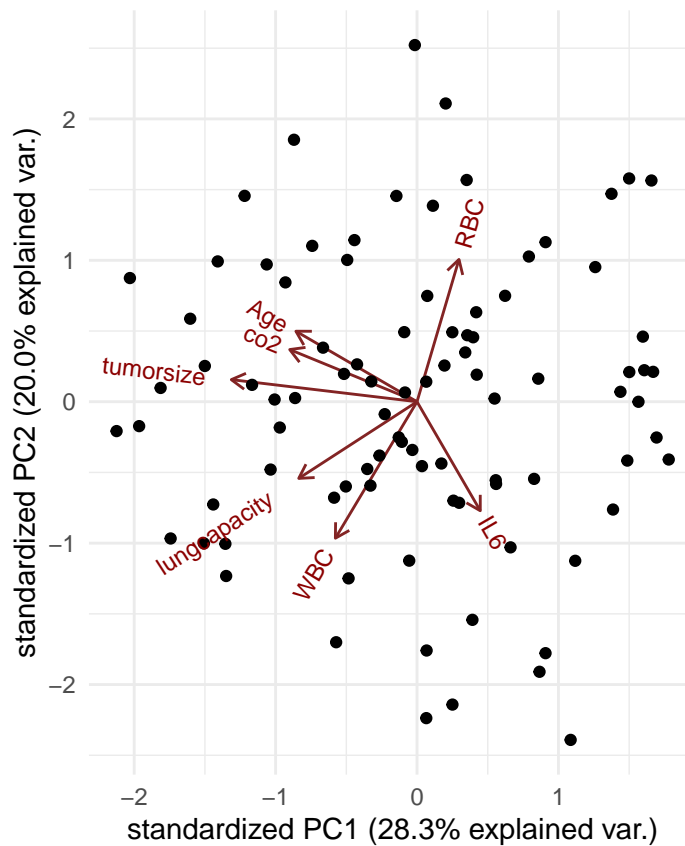
Fevereiro 2020

Questão 1

PCA para as variáveis contínuas **tumorsize**, **co2**, **lungcapacity**, **Age**, **WBC**, **RBC**, **IL6**:

```
## Importance of components:
##              PC1      PC2      PC3      PC4      PC5      PC6      PC7
## Standard deviation  1.4076 1.1828 1.0579 0.8776 0.83657 0.82176 0.59587
## Proportion of Variance 0.2831 0.1998 0.1599 0.1100 0.09998 0.09647 0.05072
## Cumulative Proportion 0.2831 0.4829 0.6428 0.7528 0.85281 0.94928 1.00000
```

Correlação das variáveis com PC1 e PC2:



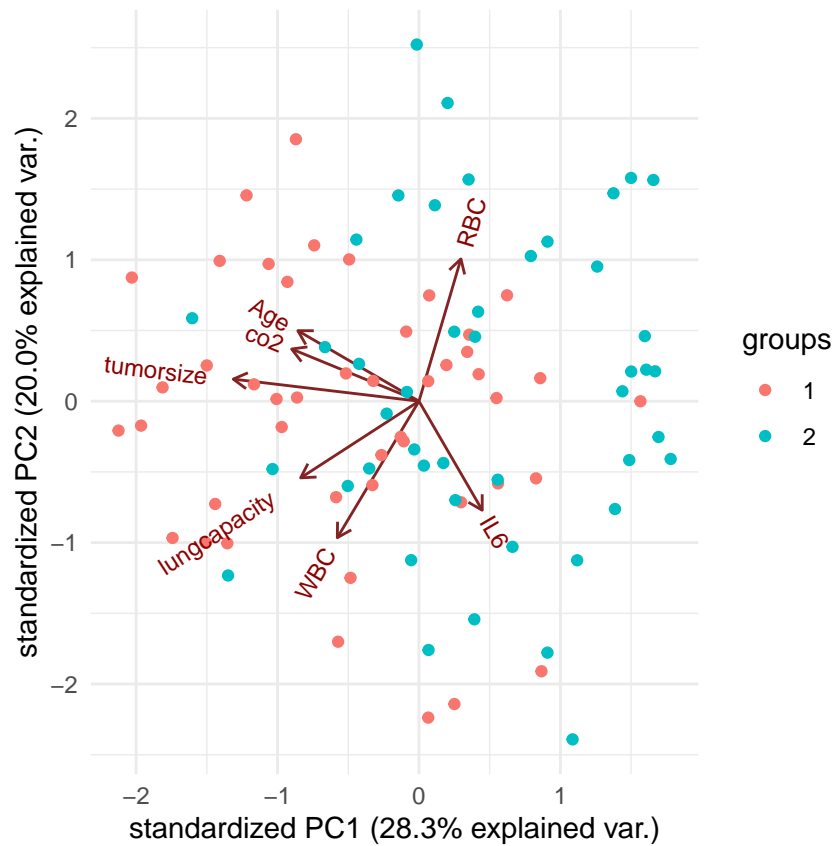
Podemos ver pela tabela e gráficos acima que **RBC** está mais correlacionada com a componente principal 1 (PC1) e que **tumorsize** está mais (inversamente) correlacionada com a componente principal 2 (PC2).

É possível diminuir a dimensionalidade. Sugeriria utilizar 3, que explicaria 64.28% da variabilidade e ainda é possível de visualizar.

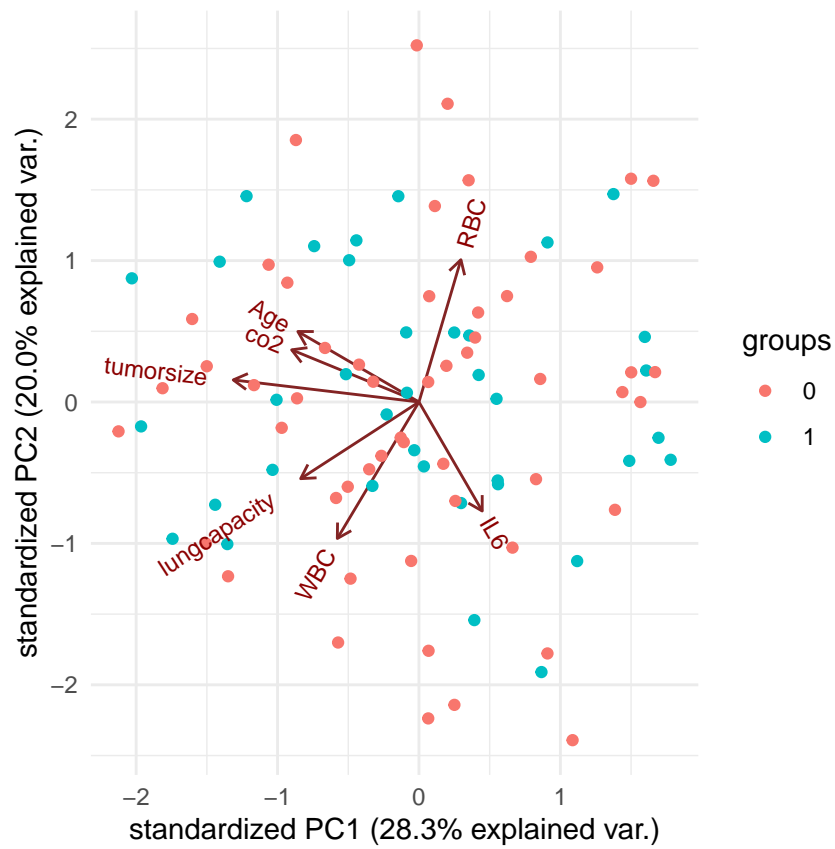
Questão 2

Gráficos usando os níveis das variáveis categóricas como grupos (**remission**, **Married**, **FamilyHx**, **SmokingHx**, **Sex**, **DID**)

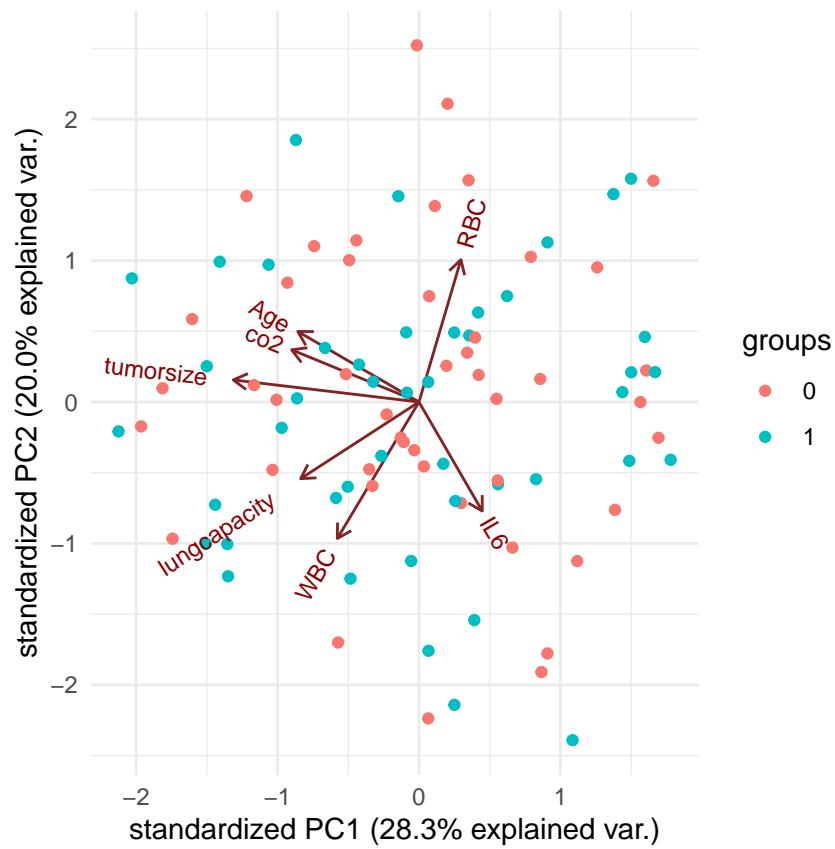
Agrupado por **remission**:



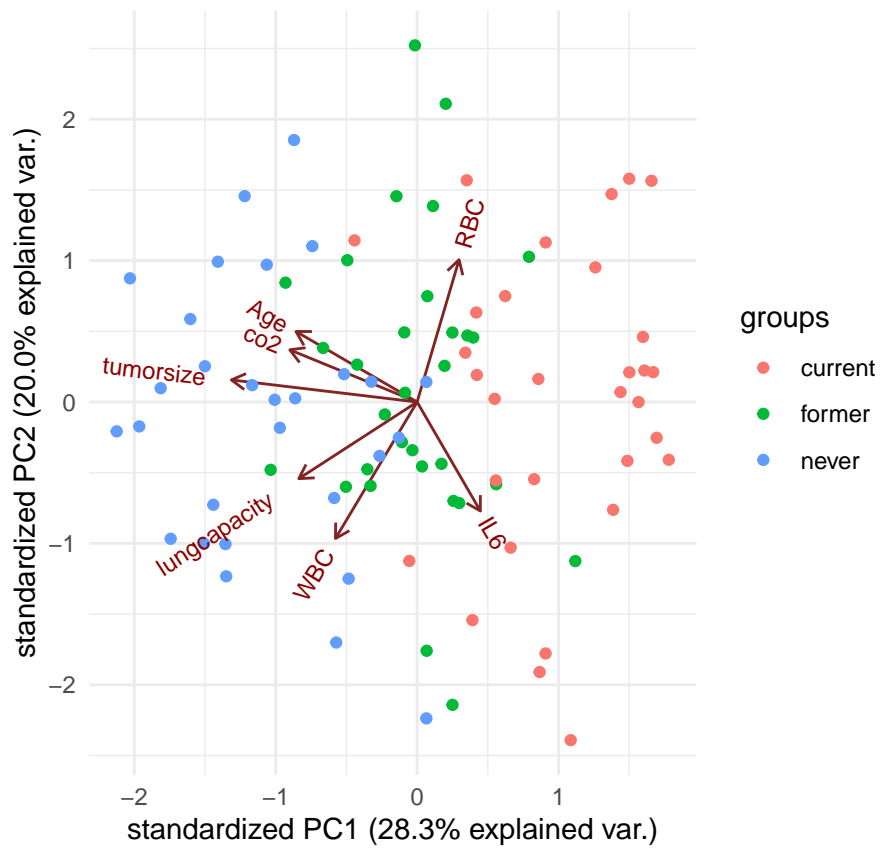
Agrupado por **Married**:



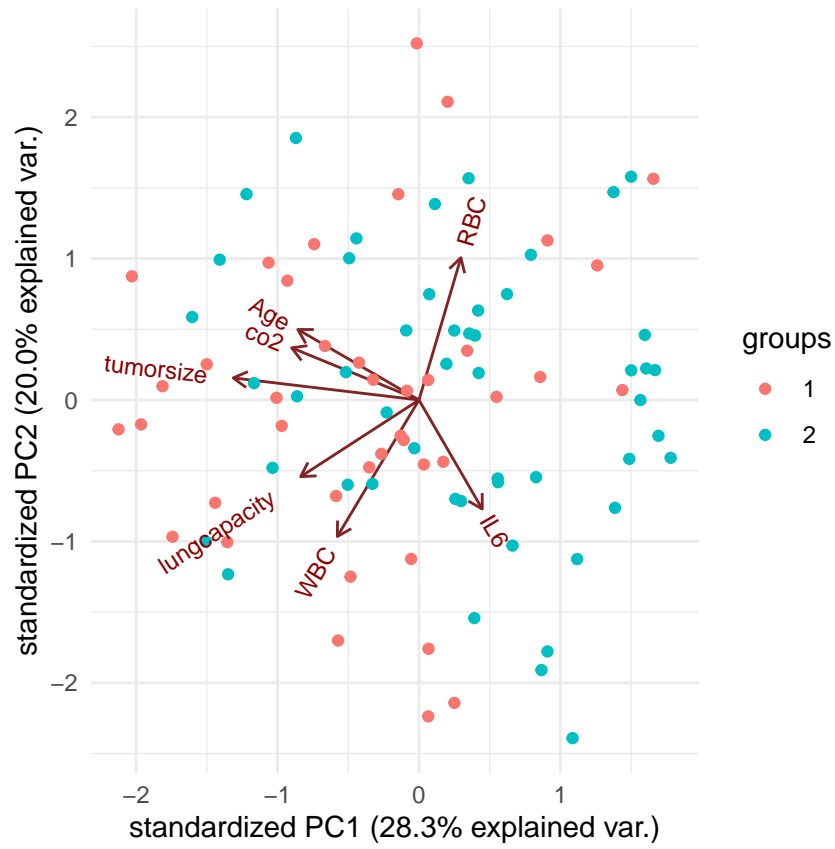
Agrupado por **FamilyHx**:



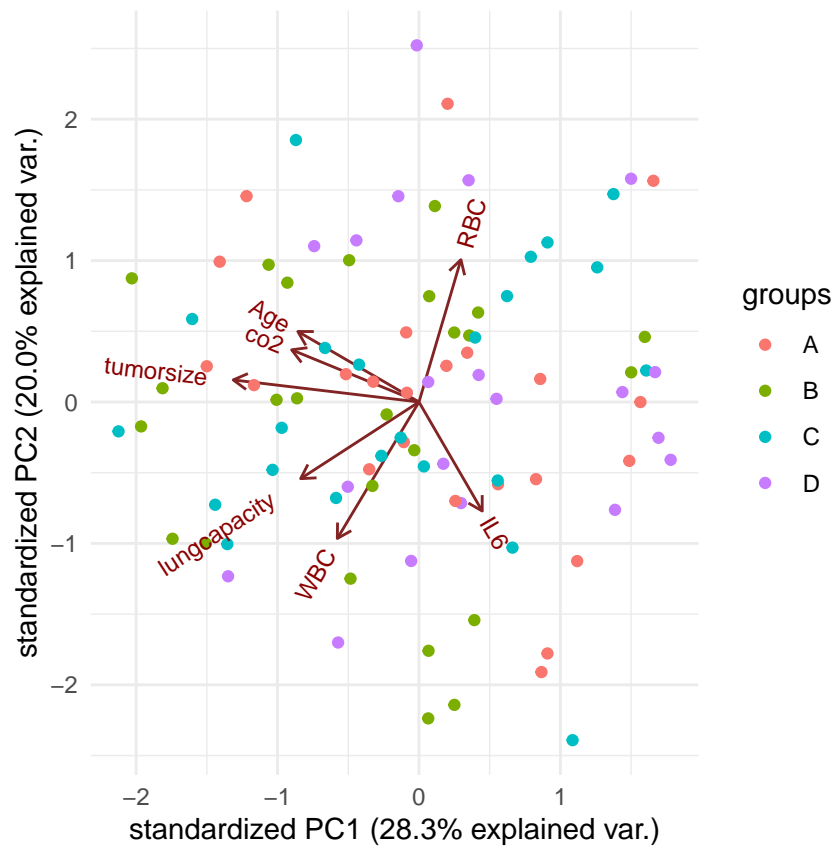
Agrupado por **SmokingHx**:



Agrupado por **Sex**:



Agrupado por **DID**:



A variável que mais foi possível diferenciar foi **SmokingHx**.

Agrupamento por **SmokingHx** com elipses:

