

# Análise de Dados

Marcelo Augusto Germano Artur

27 de março de 2025

## Limpeza e Pré-processamento

Para realizar uma análise precisa, é necessário possuir uma base de dados adequada para evitar problemas durante o processo. Então, o primeiro procedimento que deve ser feito é a limpeza e pré-processamento da base de dados. Para essa etapa, realizei alguns processos para melhorar os dados e facilitar a análise.

### 1) Acessando a base de dados e instalando os pacotes:

Primeiro, é necessário acessar a base de dados. Como a base de dados se encontrava na mesma pasta que o código, usei a função `read.delim()` para acessar o arquivo. Em seguida instalei os pacotes que serão utilizados durante todas as etapas da análise.

```
Exam_1 <- read.delim("D:/Exam_1.txt") dados
<- Exam_1

options(repos = c(CRAN = "https://cran.r-project.org"))
install.packages("outliers")

library(outliers)

install.packages("dplyr")
library(dplyr)
install.packages("ggplot2")
library(ggplot2)
```

### 2) Renomear colunas:

Após acessar o arquivo e instalar os pacotes, utilizei a função `names()` para renomear as colunas, visando torná-las mais intuitivas para quem as veja. Além da função `summary()` para coletar algumas informações sobre as colunas.

```
# Renomeando as colunas da base de dados
```

```
names(dados)[1:10] <- c("Escola", "Nota_Exame",
"Genero_Escola", "Media_Escola", "Faixa_VR", "Faixa_Admissoao", "Nota_LR", "Sexo_Es
tudante", "Tipo_Escola", "ID_Estudante")
```

```
# Coletando informações das colunas
```

```
## Coluna "Nota_Exame"
```

```
summary(dados$Nota_Exame)
```

```
##      Min.   1st Qu.   Median     Mean   3rd Qu.     Max.
## -3.666072 -0.699505  0.004322 -0.000114  0.678759  3.666091
```

```
## Coluna "Media_Escola"
```

```
summary(dados$Media_Escola)
```

```
##      Min.   1st Qu.   Median     Mean   3rd Qu.     Max.
## -0.75596 -0.14934 -0.02020  0.00181  0.21053  0.63766
```

```
## Coluna "Nota_LR"
```

```
summary(dados$Nota_LR)
```

```
##      Min.   1st Qu.   Median     Mean   3rd Qu.     Max. ## -2.93495 -
0.62071  0.04050  0.00181  0.61906  3.01595
```

### 3) Valores em branco:

Em sequência, iniciei a limpeza dos dados, verificando a existência de linhas com valores em branco. Usei a função `is.na()` para procurar valores em branco, e a partir da função `mutate()`, criei uma coluna na base de dados. Após isso, como a função `is.na()` retorna TRUE ou FALSE, com TRUE = 1 e FALSE = 0, utilizei a função `sum()` para somar os valores da coluna "Valores\_Branco", e caso o resultado fosse diferente de 0, existiriam linhas com valores em branco. Como a função `sum()` retornou o valor 0, conclui que não existiam linhas com valores em branco, então exclui a coluna "Valores\_Branco" e continuei com a limpeza dos dados.

```
VB <- is.na(dados) print(VB)
```

```
dados <- dados %>%
  mutate(Valores_Branco = VB)
```

```
sum(dados$Valores_Branco)
```

```
dados$Valores_Branco <- NULL
```

### 4) Formato dos dados:

Nesse momento, realizei a checagem do formato dos dados na minha base de dados usando a função `str()`, para identificar se existia alguma coluna com o formato errado. Identifiquei que as colunas "Escola" e "ID\_Estudante" estavam no formato int, mas como essas colunas

correspondem a identificação da escola e do estudante, respectivamente, deveriam estar no formato chr. Usei a função `as.character()` para alterar o formato de int para chr, e em seguida, novamente com a função `str()`, confirmei o sucesso dessa alteração.

```
str(dados)

## 'data.frame':    4059 obs. of  10 variables:
## $ Escola      : int  1 1 1 1 1 1 1 1 1 1 ...
## $ Nota_Exame  : num  0.261 0.134 -1.724 0.968 0.544 ...
## $ Genero_Escola : chr  "mixed" "mixed" "mixed" "mixed" ...
## $ Media_Escola : num  0.166 0.166 0.166 0.166 0.166 ...
## $ Faixa_VR    : chr  "mid 50%" "mid 50%" "mid 50%" "mid 50%" ...
## $ Faixa_Admissao: chr  "bottom 25%" "mid 50%" "top 25%" "mid 50%" ...
## $ Nota_LR     : num  0.619 0.206 -1.365 0.206 0.371 ...
## $ Sexo_Estudante: chr  "F" "F" "M" "F" ...
## $ Tipo_Escola  : chr  "Mxd" "Mxd" "Mxd" "Mxd" ...
## $ ID_Estudante : int  143 145 142 141 138 155 158 115 117 113 ...

dados$Escola <- as.character(dados$Escola)
dados$ID_Estudante <- as.character(dados$ID_Estudante)

str(dados)

## 'data.frame':    4059 obs. of  10 variables:
## $ Escola      : chr  "1" "1" "1" "1" ...
## $ Nota_Exame  : num  0.261 0.134 -1.724 0.968 0.544 ...
## $ Genero_Escola : chr  "mixed" "mixed" "mixed" "mixed" ...
## $ Media_Escola : num  0.166 0.166 0.166 0.166 0.166 ...
## $ Faixa_VR    : chr  "mid 50%" "mid 50%" "mid 50%" "mid 50%" ...
## $ Faixa_Admissao: chr  "bottom 25%" "mid 50%" "top 25%" "mid 50%" ...
## $ Nota_LR     : num  0.619 0.206 -1.365 0.206 0.371 ...
## $ Sexo_Estudante: chr  "F" "F" "M" "F" ...
## $ Tipo_Escola  : chr  "Mxd" "Mxd" "Mxd" "Mxd" ...
## $ ID_Estudante : chr  "143" "145" "142" "141" ...
```

## 5) Outliers

Na sequência, iniciei a procura por outliers na base de dados. Realizei uma varredura nas três colunas que correspondem a notas, “Nota\_Exame”, “Media\_Escola” e “Nota\_LR”. Iniciei a procura usando a função `boxplot()` para plotar os gráficos e identificar se existiam outliers, em seguida usei a função `which()` para identificar em quais linhas da base de dados estavam os outliers, e finalmente, extrai os outliers e criei um data.frame único para eles. Repeti esse processo para todas as colunas que correspondem a notas.

```

# Coluna "Nota_Exame"

bp_NE<- boxplot(dados$Nota_Exame)

outliers_NE <- bp_NE$out
print(outliers_NE)

linhas_outliers_NE<- which(dados$Nota_Exame %in% outliers_NE)
print(linhas_outliers_NE)

outliers_identificados_NE <- dados[linhas_outliers_NE, ]
print(outliers_identificados_NE)

df_NE <- data.frame(outliers_identificados_NE)
print(df_NE)

# Coluna "Media_Escola"

bp_ME <- boxplot(dados$Media_Escola)

outliers_ME <- bp_ME$out
print(outliers_ME)

linhas_outliers_ME <- which(dados$Media_Escola %in% outliers_ME)
print(linhas_outliers_ME)

outliers_identificados_ME <- dados[linhas_outliers_ME, ]
print(outliers_identificados_ME)

df_ME <- data.frame(outliers_identificados_ME)
print(df_ME)

# Coluna "Nota_LR"

bp_LRT <- boxplot(dados$Nota_LR)

outliers_LRT <- bp_LRT$out
print(outliers_LRT)

linhas_outliers_LRT <- which(dados$Nota_LR %in% outliers_LRT)
print(linhas_outliers_LRT)

outliers_identificados_LRT <- dados[linhas_outliers_LRT, ]
print(outliers_identificados_LRT)

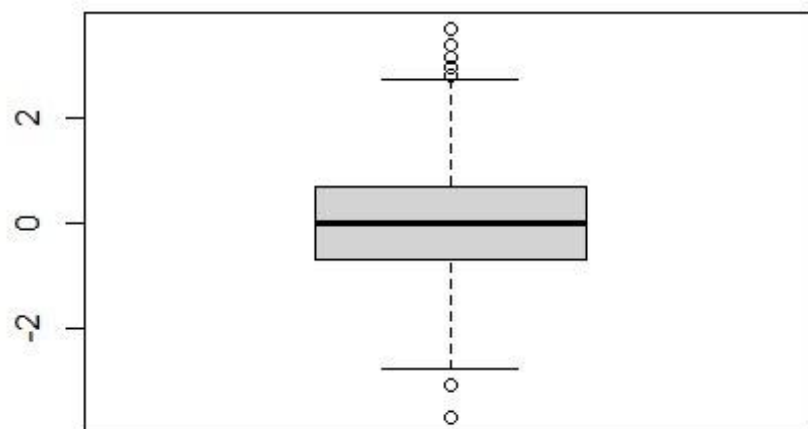
df_LRT <- data.frame(outliers_identificados_LRT)
print(df_LRT)

```

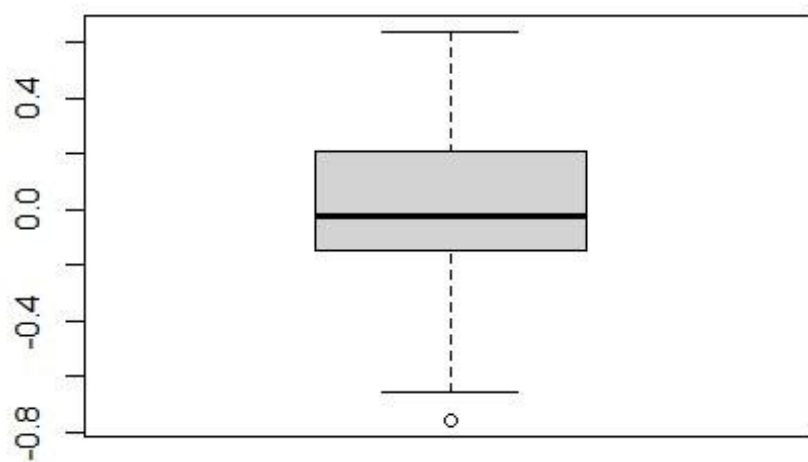
### 5.1) Gráficos

Gráficos gerados com o uso da função boxplot() durante a busca por outliers

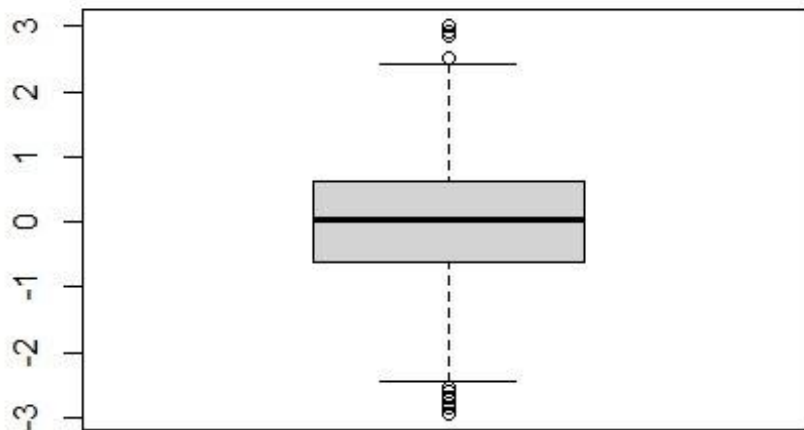
```
bp_NE<- boxplot(dados$Nota_Exame) # Coluna "Nota_Exame"
```



```
bp_ME <- boxplot(dados$Media_Escola) # Coluna "Media_Escola"
```



```
bp_LRT <- boxplot(dados$Nota_LR) # Coluna "Nota_LR"
```



## Análise

Após todo o trabalho de limpeza e pré-processamento dos dados, é hora de iniciar a análise dos dados. Para iniciar uma análise, é necessário saber o que deseja descobrir com a análise, ter em mente quais perguntas você quer responder, para assim poder encontrar as respostas que deseja. Nesse caso, baseei minha análise em quatro perguntas chaves: 1) Como o desempenho dos alunos de escolas mistas se diferenciam dos alunos de escolas únicas? 2) Como o desempenho dos alunos de escolas masculinas se diferenciam do desempenho das alunas de escolas femininas? 3) Como essa divisão afeta a pontuação no teste de raciocínio verbal (VR)? Com esses questionamentos em mente, iniciei minha análise. 4) O que os outliers mostram sobre o desempenho dos alunos?

### 1) Como o desempenho dos alunos de escolas mistas se diferenciam dos alunos de escolas únicas?

Para responder essa primeira pergunta, busquei comparar o desempenho dos estudantes de escolas mistas (Mxd) e únicas (Sngl) usando a pontuação normalizada no exame e a pontuação padronizada no teste LR.

```
# Separando os dados referente a cada tipo de escola
```

```
notas_mxd <- dados %>%  
  filter(Tipo_Escola == "Mxd") notas_sngl  
<- dados %>% filter(Tipo_Escola ==  
  "Sngl")
```

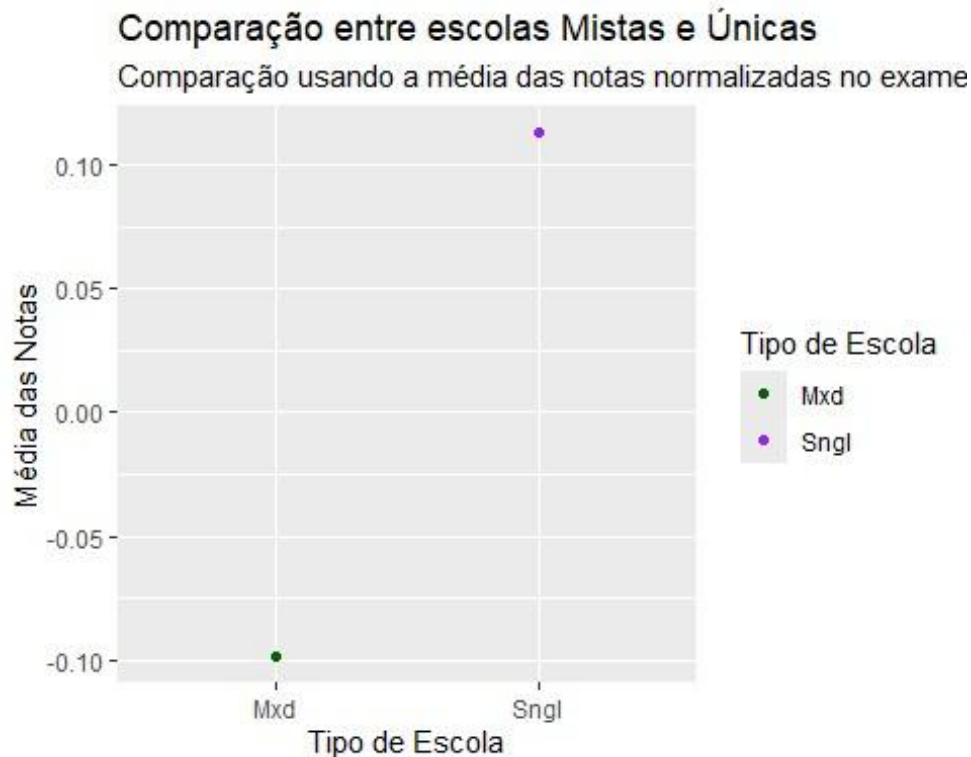
```
# Análise usando a pontuação normalizada no exame
```

```
library(ggplot2)
```

```
media_NE_mxd <- mean(notas_mxd[,2])  
media_NE_sngl <- mean(notas_sngl[,2])
```

```
dados_NE <- data.frame(Tipo_Escola = c("Mxd","Sngl"), medias_NE <-  
  c(media_NE_mxd, media_NE_sngl))
```

```
ggplot(dados_NE, aes(x = Tipo_Escola, y = medias_NE, color = Tipo_Escola)) +  
  geom_point(stat = "identity") +  
    scale_color_manual(values = c("darkgreen", "purple")) + labs(title =  
  "Comparação entre escolas Mistas e Únicas", subtitle = "Comparação usando  
  a média das notas normalizadas no exame", x = "Tipo de  
  Escola", y = "Média das Notas", color = "Tipo de Escola")
```



```
# Análise usando a pontuação padronizada no teste LR
```

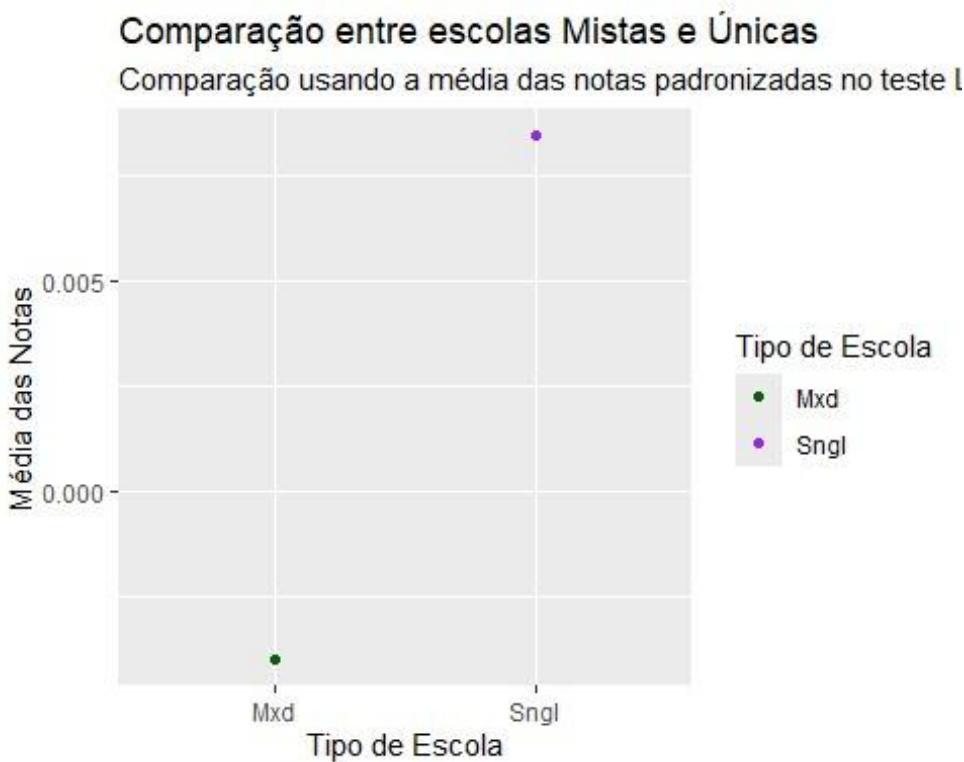
```
library(ggplot2)
```

```
media_LR_mxd <- mean(notas_mxd[,7])
```

```
media_LR_sngl <- mean(notas_sngl[,7])
```

```
dados_LR <- data.frame(Tipo_Escola = c("Mxd", "Sngl"), medias_LRT <-  
c(media_LR_mxd, media_LR_sngl))
```

```
ggplot(dados_LR, aes(x = Tipo_Escola, y = medias_LRT, color = Tipo_Escola))  
+  
  geom_point(stat = "identity") +  
  scale_color_manual(values = c("darkgreen", "purple")) +  
  labs(title = "Comparação entre escolas Mistas e Únicas", subtitle =  
"Comparação usando a média das notas padronizadas no teste LR", x = "Tipo de  
Escola", y = "Média das Notas", color = "Tipo de Escola")
```



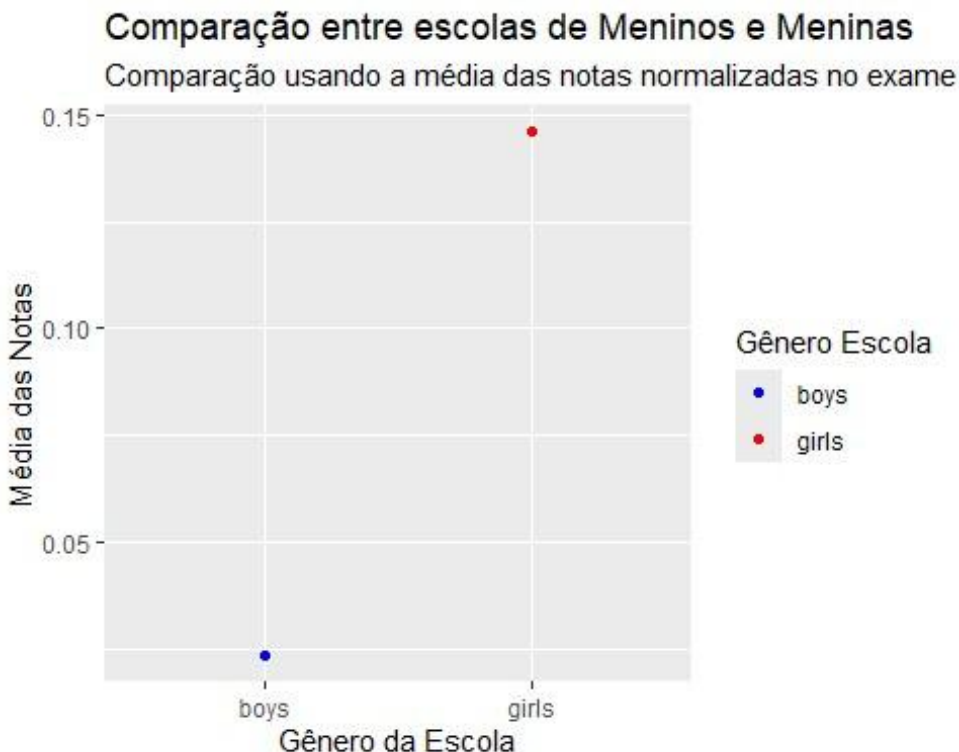
## 2) Como o desempenho dos alunos de escolas masculinas se diferenciam do desempenho das alunas de escolas femininas?

Para responder essa questão, comecei separando os dados apenas das escolas únicas. Em seguida comparei o desempenho usando as mesmas métricas da primeira pergunta.



### ### Separação dos dados correspondentes a cada tipo de escola única

```
dados_escolas_sngl_boys <- dados %>%  
  filter(Genero_Escola == "boys") dados_escolas_sngl_girls  
<- dados %>% filter(Genero_Escola == "girls")  
  
# Análise usando a pontuação normalizada no exame  
  
library(ggplot2)  
  
media_NE_sngl_boys <- mean(dados_escolas_sngl_boys[,2]) media_NE_sngl_girls  
<- mean(dados_escolas_sngl_girls[,2])  
  
dados_NE_sngl <- data.frame(Genero_Escola = c("boys", "girls"),  
  medias_NE_sngl = c(media_NE_sngl_boys, media_NE_sngl_girls))  
  
ggplot(dados_NE_sngl, aes(x = Genero_Escola, y = medias_NE_sngl, color =  
  Genero_Escola)) +  
  geom_point(stat = "identity") +  
  scale_color_manual(values = c("blue", "red")) + labs(title = "Comparação  
entre escolas de Meninos e Meninas", subtitle = "Comparação usando a média  
das notas normalizadas no exame", x = "Gênero da  
Escola", y = "Média das Notas", color = "Gênero Escola")
```



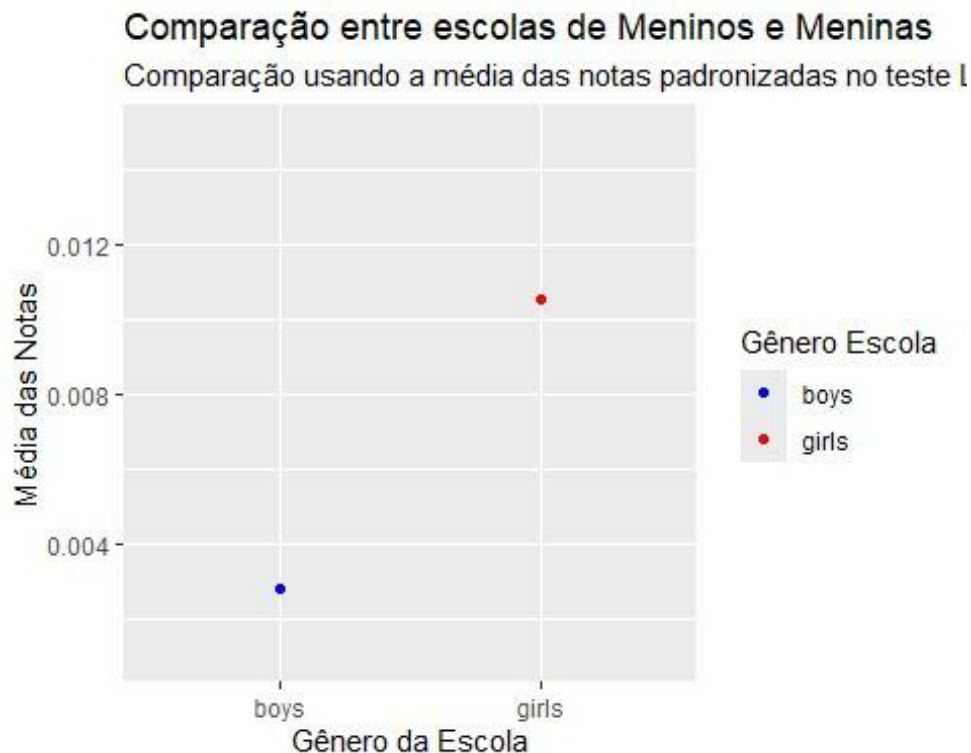
# Análise usando a pontuação padronizada no teste LR

```
library(ggplot2)

media_LR_sngl_boys <- mean(dados_escolas_sngl_boys[,7]) media_LR_sngl_girls
<- mean(dados_escolas_sngl_girls[,7])

dados_LR_sngl <- data.frame(Genero_Escola = c("boys", "girls"),
medias_LR_sngl = c(media_LR_sngl_boys, media_LR_sngl_girls))

ggplot(dados_LR_sngl, aes(x = Genero_Escola, y = medias_LR_sngl, color =
Genero_Escola)) +
  geom_point(stat = "identity") + scale_color_manual(values = c("blue",
"red")) + scale_y_continuous(limits = c(0.001,0.015)) + labs(title =
"Comparação entre escolas de Meninos e Meninas", subtitle = "Comparação
usando a média das notas padronizadas no teste LR", x = "Gênero da Escola",
y = "Média das Notas", color = "Gênero Escola")
```



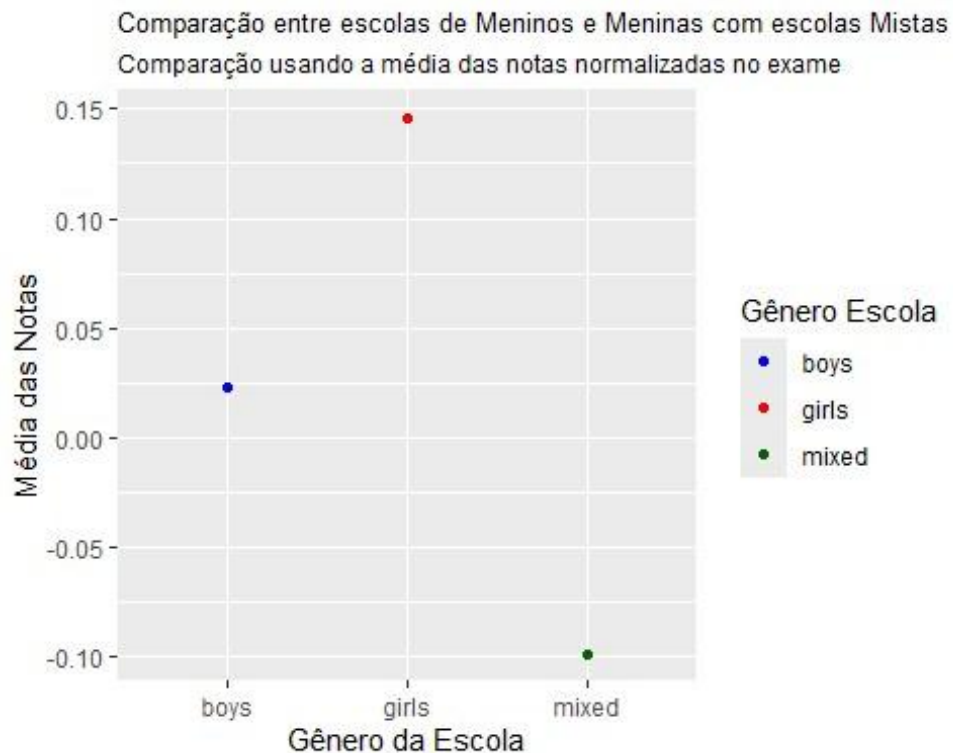
*Observação:*

Decidi realizar uma comparação entre as escolas mistas e as únicas novamente, mas desta vez separei as únicas em masculinas e femininas. Esse ponto mescla a pergunta 1) com a 2).

*# Análise usando a pontuação normalizada no exame*

```
dados_NE_obs <- data.frame(Genero_Escola = c("boys", "girls", "mixed"),
media_NE = c(media_NE_sngl_boys,media_NE_sngl_girls,media_NE_mxd))

ggplot(dados_NE_obs, aes(x = Genero_Escola, y = media_NE, color =
Genero_Escola)) +
  geom_point(stat = "identity") + scale_color_manual(values = c("blue",
"red", "darkgreen")) + theme(plot.title = element_text(size = 10),
plot.subtitle = element_text(size = 9)) + labs(title = "Comparação
entre escolas de Meninos e Meninas com escolas Mistas", subtitle =
"Comparação usando a média das notas normalizadas no exame", x = "Gênero
da Escola", y = "Média das Notas", color = "Gênero
Escola")
```

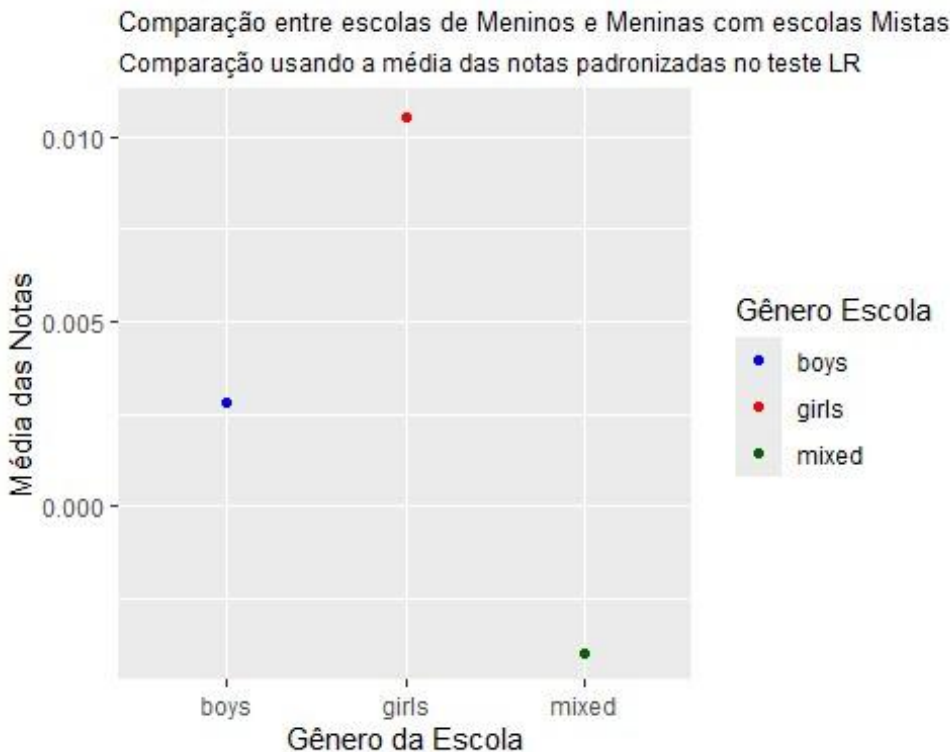


```
# Análise usando a pontuação padronizada no teste LR
```

```
library(ggplot2)
```

```
dados_LR_obs <- data.frame(Genero_Escola = c("boys", "girls", "mixed"),  
media_LR = c(media_LR_sngl_boys, media_LR_sngl_girls, media_LR_mxd))
```

```
ggplot(dados_LR_obs, aes(x = Genero_Escola, y = media_LR, color =  
Genero_Escola)) +  
  geom_point(stat = "identity") + scale_color_manual(values = c("blue",  
"red", "darkgreen")) + theme(plot.title = element_text(size = 10),  
plot.subtitle = element_text(size = 9)) + labs(title = "Comparação entre  
escolas de Meninos e Meninas com escolas Mistas", subtitle = "Comparação  
usando a média das notas padronizadas no teste LR", x = "Gênero da  
Escola", y = "Média das Notas", color = "Gênero  
Escola")
```



### 3) Como essa divisão afeta a pontuação no teste de raciocínio verbal (VR)?

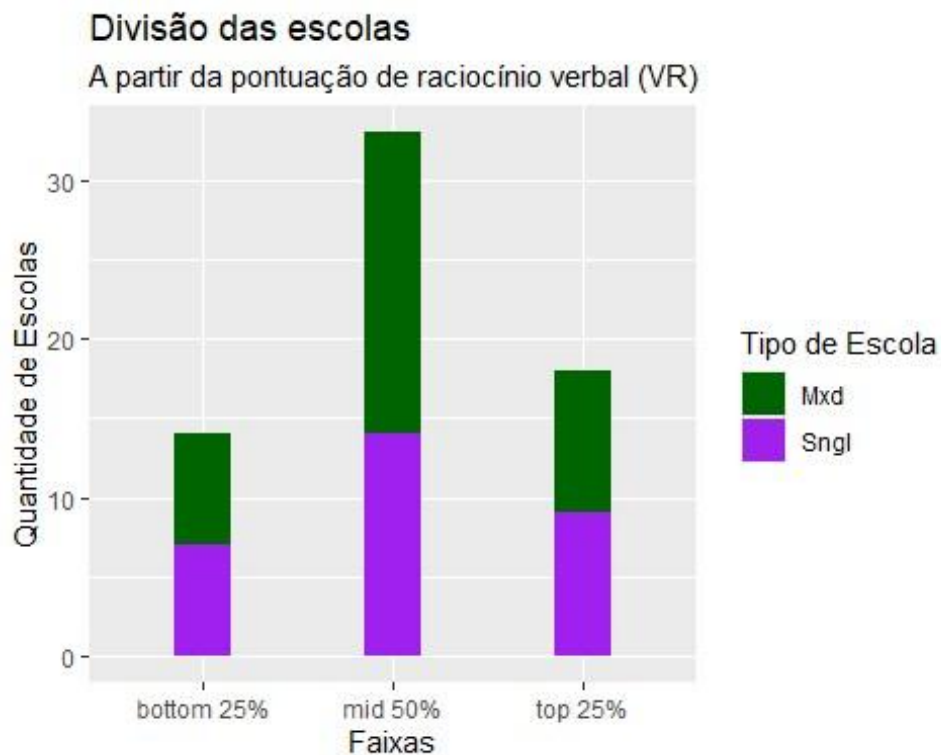
Para responder o terceiro questionamento, utilizei a pontuação no teste de raciocínio verbal (VR). Utilizei o conceito de 1º e 3º quartil para analisar, que nesse caso aparecem no formato “bottom 25%” e “top 25%”, respectivamente. Inicialmente comparei escolas mistas e únicas, e em sequência analisei cada caso separadamente.

```
# Definindo a quantidade de escolas
```

```
dados$Qtd_Escolas <- ifelse(duplicated(dados$Escola), 0, 1)
```

```
# Comparando escolas mistas com únicas
```

```
ggplot(dados, aes(x = Faixa_VR, y = Qtd_Escolas, fill = Tipo_Escola)) +  
geom_bar(stat = "identity", width = 0.3) + scale_fill_manual(values =  
c("darkgreen", "purple")) + labs(title = "Divisão das escolas", subtitle =  
"A partir da pontuação de raciocínio verbal (VR)", x = "Faixas", y =  
"Quantidade de Escolas", fill = "Tipo de Escola")
```



```
# Separando os dados de cada tipo de escola
```

```
# Sngl
```

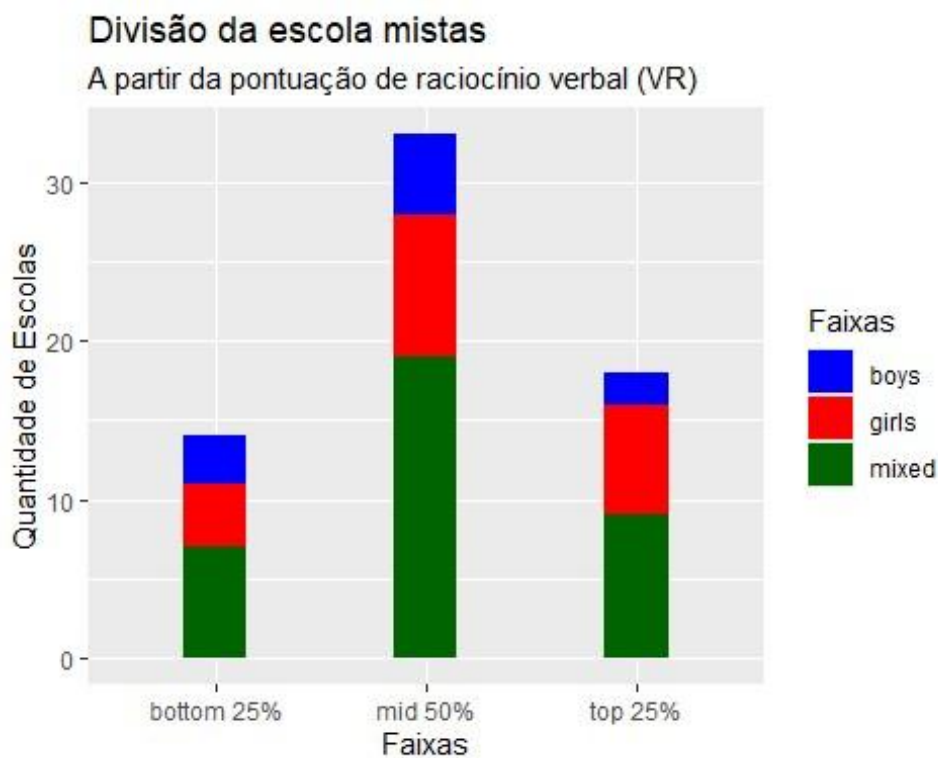
```
dados_escolas_sngl <- dados %>%  
filter(Tipo_Escola == "Sngl")
```

```
# Todas
```

```
dados_escolas_VR <- dados %>%  
select(Faixa_VR, Genero_Escola, Qtd_Escolas)
```

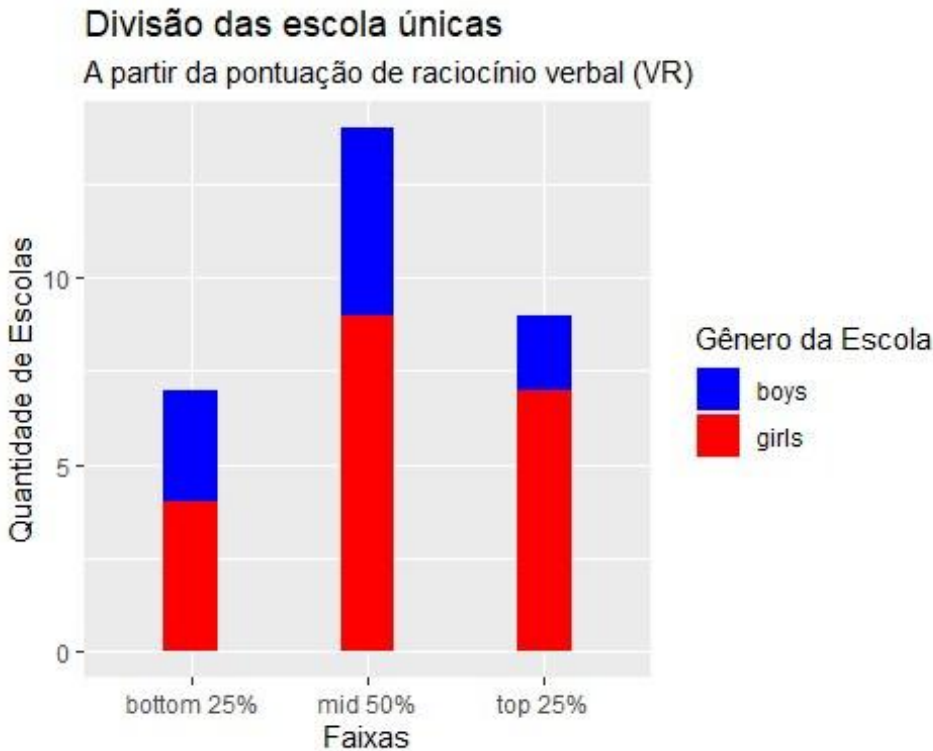
```
# Análise todas as escolas
```

```
ggplot(dados_escolas_VR, aes(x = Faixa_VR, y = Qtd_Escolas, fill =  
Genero_Escola)) +  
  geom_bar(stat = "identity", width = 0.3) +  
  scale_fill_manual(values = c("blue", "red", "darkgreen")) + labs(title =  
"Divisão da escola mistas", subtitle = "A partir da pontuação de raciocínio  
verbal (VR)", x = "Faixas", y = "Quantidade de Escolas", fill = "Faixas")
```



```
# Análise escolas únicas
```

```
ggplot(dados_escolas_sngl, aes(x = Faixa_VR, y = Qtd_Escolas, fill =  
Genero_Escola)) + geom_bar(stat = "identity", width = 0.3) +  
scale_fill_manual(values = c("blue", "red")) + labs(title = "Divisão das  
escolas únicas", subtitle = "A partir da pontuação de raciocínio verbal (VR)",  
x = "Faixas", y = "Quantidade de Escolas", fill = "Gênero da Escola")
```



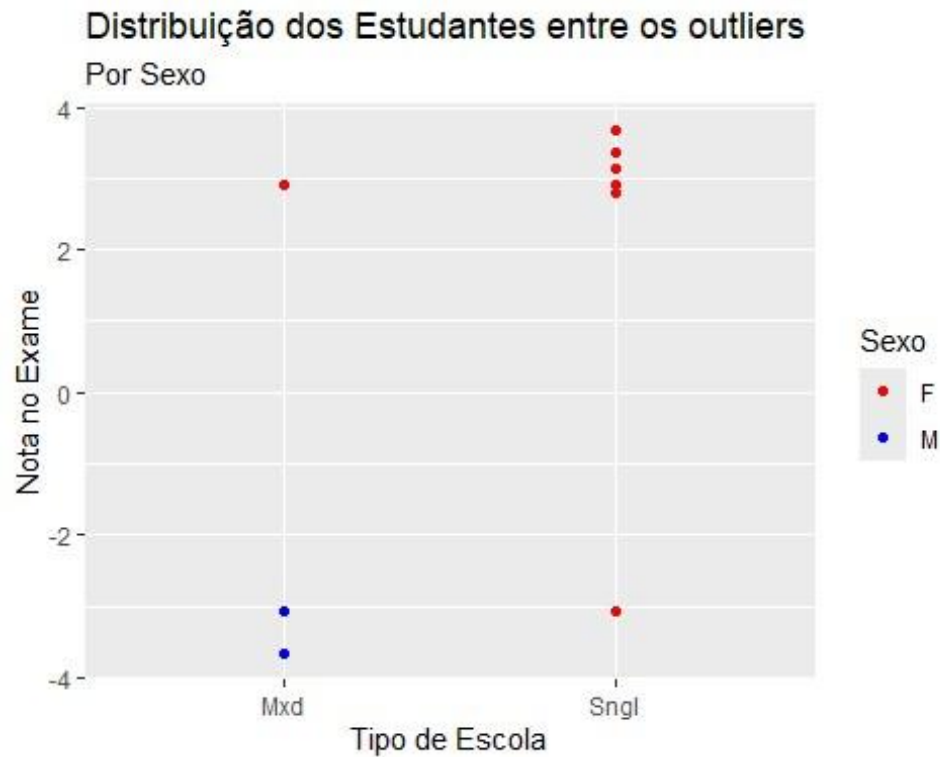
#### 4) O que os outliers mostram sobre o desempenho dos alunos?

Para a última pergunta, usei os dados sobre outliers que havia coletado durante o préprocessamento. Analisei os dados das colunas “Nota\_Exame”, “Media\_Escola” e “Nota\_LR”, fazendo comparações por tipo de escola e por sexo, quando possível.

*# Análise dos outliers da coluna "Nota\_Exame"*

*## Análise por sexo dos estudantes e por tipo de escola*

```
ggplot(df_NE, aes(x = Tipo_Escola, y = Nota_Exame, color = Sexo_Estudante)) +  
  geom_point(stat = "identity") +  
  scale_color_manual(values = c("red", "blue")) +  
  labs(title = "Distribuição dos Estudantes entre os outliers", subtitle =  
"Por Sexo", x = "Tipo de Escola", y = "Nota no Exame", color = "Sexo")
```

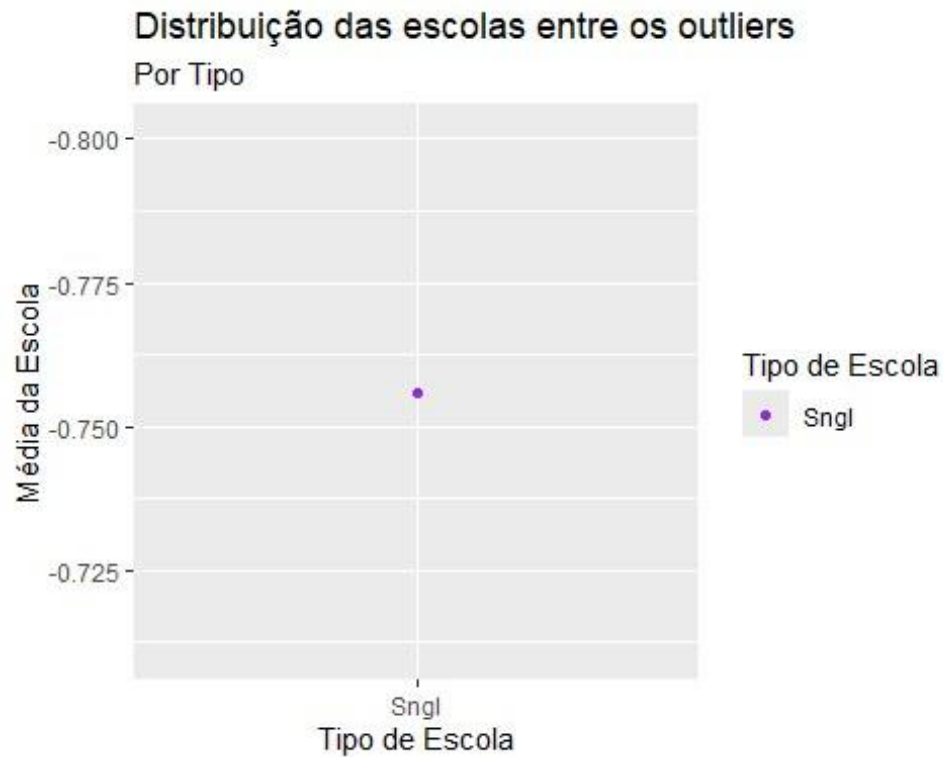


# Análise dos outliers na coluna "Media\_Escola"

## Análise por tipo de escola

```
ggplot(df_ME, aes(x = Tipo_Escola, y = Media_Escola, color = Tipo_Escola)) +  
geom_point(stat = "identity") + scale_y_reverse() +  
scale_color_manual(values = c("purple", "darkgreen")) +  
labs(title = "Distribuição das escolas entre os outliers", subtitle =  
"Por Tipo", x = "Tipo de Escola", y = "Média da Escola", color = "Tipo  
de  
Escola")
```

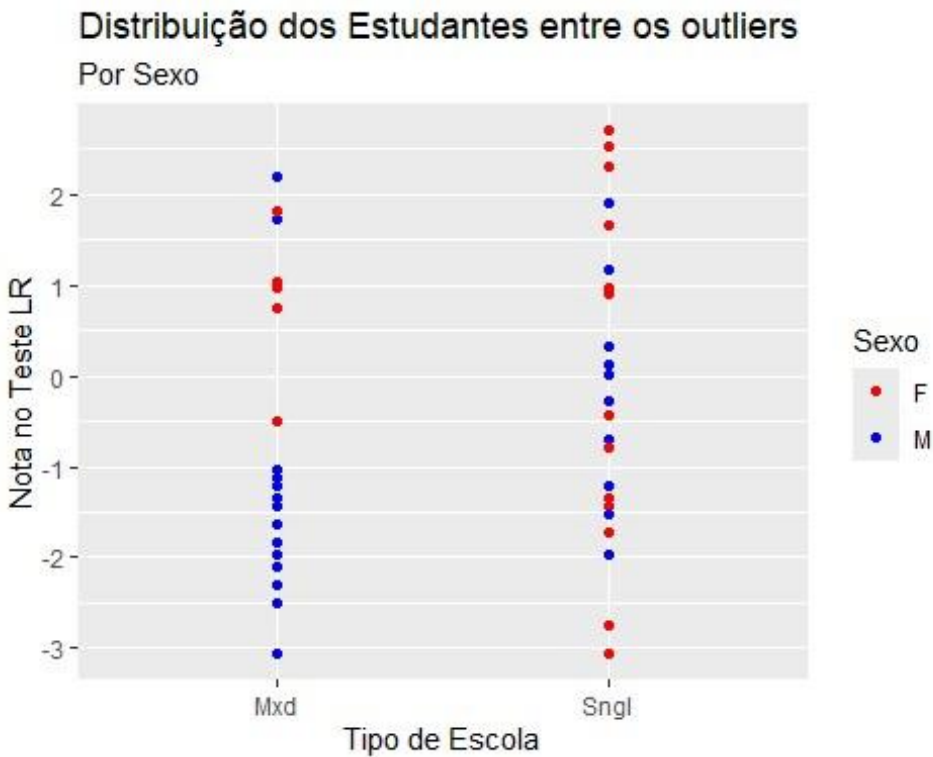




```
# Análise dos outliers da coluna "Nota_LR"
```

```
## Análise por sexo dos estudantes e por tipo de escola
```

```
ggplot(df_LRT, aes(x = Tipo_Escola, y = Nota_Exame, color = Sexo_Estudante))  
+  
  geom_point(stat = "identity") +  
  scale_color_manual(values = c("red", "blue")) + labs(title =  
"Distribuição dos Estudantes entre os outliers", subtitle = "Por Sexo", x  
= "Tipo de Escola", y = "Nota no Teste LR", color = "Sexo")
```



## Apresentação de resultados, Levantamento de Hipóteses e Conclusão

### Resultados

Após concluir a análise dos dados, com base nos resultados, é possível responder as perguntas feitas no início do processo, além de levantar algumas hipóteses.

De início, foi feita a comparação do desempenho das escolas mistas e únicas, tendo como base as notas normalizadas no exame e as notas padronizadas no teste LR de todos os estudantes. Para deixar claro o critério de avaliação, em ambos os processos, tanto de normalização, como de padronização, a curva das notas se comporta como a normal, tendo a maioria das ocorrências mais próximas do centro, e menos ocorrências em ambas as extremidades. Além disso, como na grande maioria das avaliações, quanto maior a nota, melhor ela é, e isso também vale para esse caso. Com base nisso, é possível perceber uma dominância no desempenho das escolas únicas contra as mistas, tendo em vista que na média de ambas as notas, as escolas únicas estão acima das mistas. Concluindo, nesse cenário analisado, é possível afirmar que as escolas únicas levam vantagem sobre as mistas no quesito desempenho, não sendo possível afirmar que a separação seja o principal fator causador desse fato, mas certamente é um dos mais principais.

Em seguida, direcionei a análise apenas para as escolas únicas, usando como base os mesmos critérios. Assim como foi observado na comparação entre escolas mistas e únicas, existe uma

larga diferença no desempenho de escolas exclusivamente femininas e escolas exclusivamente masculinas, com as escolas femininas muito acima. Na comparação das médias no exame e no teste LR, a diferença observada entre escolas femininas e masculinas é maior do que a observada entre únicas e mistas, levando a hipótese de que as escolas femininas são muito superiores comparada com as demais.

Na sequência, refiz a comparação entre escolas mistas e únicas, mas dessa vez separando as únicas em femininas e masculinas. Após essa análise, a hipótese levantada na análise anterior foi comprovada, as escolas femininas são de fato superiores a todas as outras, com uma vantagem ainda maior sobre as escolas mistas.

Partindo para a próxima análise, comparei o desempenho das escolas no teste de raciocínio verbal (VR). Nessa análise, realizei três comparações: A distribuição nas faixas de pontuação entre escolas mistas e únicas, a distribuição nas faixas de pontuação entre escolas mistas, femininas e masculinas e a distribuição nas faixas de pontuação entre escolas femininas e masculinas. Na comparação entre mistas e únicas, é perceptível um equilíbrio entre elas na faixa “top 25%” e “bottom 25%”, e uma vantagem para as mistas na faixa “mid 50%”. Com base nisso, fica claro que mesmo tendo menos escolas únicas na base de dados, elas praticamente ocupam metade das vagas entre as melhores escolas, evidenciando uma superioridade na qualidade de ensino em relação as escolas mistas. Partindo para a comparação entre todos os tipos de escolas, tudo que foi descoberto na primeira comparação é reiterado, acrescentando apenas a disparidade entre escolas femininas e masculinas. Na última comparação, a disparidade entre escolas femininas e masculinas é confirmada. De maneira oposta ao que ocorre entre escolas mistas e únicas, no cenário de comparação entre femininas e masculinas, as escolas para garotas fazem valer a superioridade numérica e dominam, com certa tranquilidade, a faixa das melhores escolas.

Na última parte da análise, utilizei os outliers que encontrei na fase de pré-processamento dos dados para basear meu estudo. Comparei a distribuição dos outliers nas notas do exame, médias de entrada nas escolas e notas no teste LR. Começando nas notas normalizadas no exame, é perceptível uma diferença no desempenho no exame entre meninos e meninas, sendo todos os outliers positivos (maiores notas) vindo de meninas, e dois terços dos negativos vindo de meninos. Essa diferença se mantém entre as escolas, com as mistas tendo apenas um outlier positivo, e entre as únicas, todos os outliers positivos vieram de meninas, corroborando com a superioridade feminina constatada nas análises anteriores. Nas médias de entrada nas escolas, há apenas um outlier, sendo ele negativo e vindo de uma escola única. Nas notas padronizadas no teste LR ocorreu a maior presença de outliers, porém, mantendo o padrão apresentado nas notas no exame, os outliers positivos dominados por escolas únicas e meninas, enquanto os outliers negativos são, em sua maioria, de escolas mistas e meninos.

## Conclusão

Após concluir toda a análise da base de dados, avaliar os resultados obtidos e estudar as hipóteses levantadas, chego à conclusão de que existe uma disparidade no desempenho dos estudantes de escolas mistas e únicas. Os dados analisados são bem claros em relação a essa diferença, de todas as formas que eles foram trabalhados e apresentados, sempre evidenciaram a superioridade das escolas únicas, principalmente femininas, comparadas com as escolas mistas. Até mesmo os dados extremos (outliers) mostraram esse cenário, um claro domínio feminino entre as maiores notas, e uma predominância masculina e de escolas mistas nas notas mais baixas. Por fim, afirmo que é necessário uma análise mais profunda dessa situação. Os dados usados são muito versáteis, mas têm suas limitações, tornando necessária a busca por mais informações contundentes para análises futuras. Uma sugestão seria acrescentar dados sobre a localização das escolas, situação econômica da região onde se encontram e idade dos estudantes.