

```

# Eduardo Freire Mangabeira 116137191
# Trabalho 1 de AED
# 1)
# item a)
nomes = read.table("eduardobrasil.txt", col.names=c("UF","Pop","Freq","Taxa"))
UF <- nomes$UF
Freq <- nomes$Freq
barplot(Freq, main = "Frequência por estado",
ylim=c(0,200000),xlab="Estados",ylab="Eduardos",font.lab=2,names.arg=UF)
# item b)
Pop <- nomes$Pop
barplot(Freq/Pop, main = "Eduardos por habitante por estado",
ylim=c(0,0.006),xlab="Estados",ylab="Eduardos/habitante",font.lab=2,names.arg=UF)
# item c)
# Para fazer o item c) criei um novo dataset com os dados das regiões
nomes_regiao = read.table("regioes.txt", col.names=c("UF","Pop","Freq","Taxa"))
Freq <- nomes_regiao$Freq
Pop <- nomes_regiao$Pop
# Sul
Freq_sul = sum(Freq[1:3])
Pop_sul = sum(Pop[1:3])
Taxa_sul = Freq_sul/Pop_sul*100000
# Sudeste
Freq_sudeste = sum(Freq[4:7])
Pop_sudeste = sum(Pop[4:7])
Taxa_sudeste = Freq_sudeste/Pop_sudeste*100000
# Nordeste
Freq_nordeste = sum(Freq[8:16])
Pop_nordeste = sum(Pop[8:16])
Taxa_nordeste = Freq_nordeste/Pop_nordeste*100000
# Centro-Oeste
Freq_centro_oeste = sum(Freq[17:20])
Pop_centro_oeste = sum(Pop[17:20])
Taxa_centro_oeste = Freq_centro_oeste/Pop_centro_oeste*100000
# Norte
Freq_norte = sum(Freq[21:27])
Pop_norte = sum(Pop[21:27])
Taxa_norte = Freq_norte/Pop_norte*100000

Nome_Regioes <- c("Sul", "Sudeste", "Nordeste", "Centro-Oeste", "Norte")
Freq_Regioes <- c(Freq_sul, Freq_sudeste, Freq_nordeste, Freq_centro_oeste, Freq_norte)
Pop_Regioes <- c(Pop_sul, Pop_sudeste, Pop_nordeste, Pop_centro_oeste, Pop_norte)
Taxa_Regioes <- c(Taxa_sul, Taxa_sudeste, Taxa_nordeste, Taxa_centro_oeste, Taxa_norte)
Regioes <-
data.frame("Regioes"=Nome_Regioes,"Freq_reg"=Freq_Regioes,"Pop_Regioes"=Pop_Regioes,"Taxa_Regioes"=Taxa_Regioes)
pie(Regioes$Freq, labels=Regioes$Regioes, main="Frequência de Eduardos por Região")

# 2)
alunos = read.table("alunos.txt", col.names=c("Matéria","Nota","Opinião","Tempo"))
summary(alunos)
Nota = alunos$Nota
Tempo = alunos$Tempo

# media
media_notas = mean(Nota)
media_tempo = mean(Tempo)

# mediana
mediana_notas = median(Notas)
mediana_tempo = median(Tempo)

# quartis
quartil_notas = quantile(Nota)
quartil_tempo = quantile(Tempo)

# desvio padrão
desvio_padrao_notas = sd(Nota)
desvio_padrao_tempo = sd(Tempo)

# variância
variancia_notas = var(Nota)
variancia_tempo = var(Tempo)

# coeficiente de variação
coef_var_notas = desvio_padrao_notas/media_notas
coef_var_tempo = desvio_padrao_tempo/media_tempo

# histogramas
hist(Nota, main="Frequência de notas", ylab="Quantidade de alunos", xlab="Notas", col="lightblue", breaks = 9)
hist(Tempo, main="Tempo de estudo", ylab="Quantidade de alunos", xlab="Tempo de estudo",col="purple", breaks = 13)

# boxplots
boxplot(Tempo, main="Tempo de estudo")
boxplot(Nota, main="Notas")

```

```

# gráficos de barras
Materia = table(alunos$Matéria)
barplot(Materia, main = "Frequência de alunos por matéria")
Opinioao = table(alunos$Opinião)
barplot(Opinioao, main = "Opinião dos alunos sobre a escola")

# gráficos de setores
pie(Materia)
pie(Opinioao)

# Notas por materia(split)
Notas_materias = split(alunos$Nota, alunos$Matéria)
mean(Notas_materias$Química)
hist(Notas_materias$Química, main="Notas dos alunos em Química", breaks=11, col="grey",
xlab="Notas",ylab="Frequência")

# Comparativo entre avaliação e desempenho(subset)
# Ruim
Ruim = subset(alunos, Opinião == "Ruim", select = Nota)
mean(Ruim$Nota)
hist(Ruim$Nota, main="Notas dos alunos que acharam a escola ruim", col="magenta", breaks = 5,
xlab="Notas",ylab="Frequência")
# Regular
Regular = subset(alunos, Opinião == "Regular", select = Nota)
mean(Regular$Nota)
hist(Regular$Nota, main="Notas dos alunos que acharam a escola regular", col="yellow", breaks = 7,
xlab="Notas",ylab="Frequência")
# Boa
Boa = subset(alunos, Opinião == "Boa", select = Nota)
mean(Boa$Nota)
hist(Boa$Nota, main="Notas dos alunos que acharam a escola boa", col="lightblue", breaks = 7,
xlab="Notas",ylab="Frequência")

# Ótima
Otima = subset(alunos, Opinião=="Ótima", select = Nota)
mean(Otima$Nota)
hist(Otima$Nota, main="Notas dos alunos que acharam a escola ótima", col="green", breaks = 5,
xlab="Notas",ylab="Frequência")

# Comparativo entre boa/ótima e Regular/ruim(merge)
# avaliação positiva
positivo = merge(Otima, Boa)
median(positivo$Nota)
boxplot(positivo, main = "Notas dos alunos que avaliaram como boa ou ótima")
# avaliação negativa
negativo = merge(Regular, Ruim)
median(negativo$Nota)
boxplot(negativo, main = "Notas dos alunos que avaliaram como regular ou ruim")

# não entendi muito bem como funciona o aggregate (aggregate)
df_alunos = data.frame(alunos)
agg_materias = aggregate(df_alunos$Nota, by= list(tempo=alunos$Tempo), FUN = mean)
boxplot(agg_materias)

```