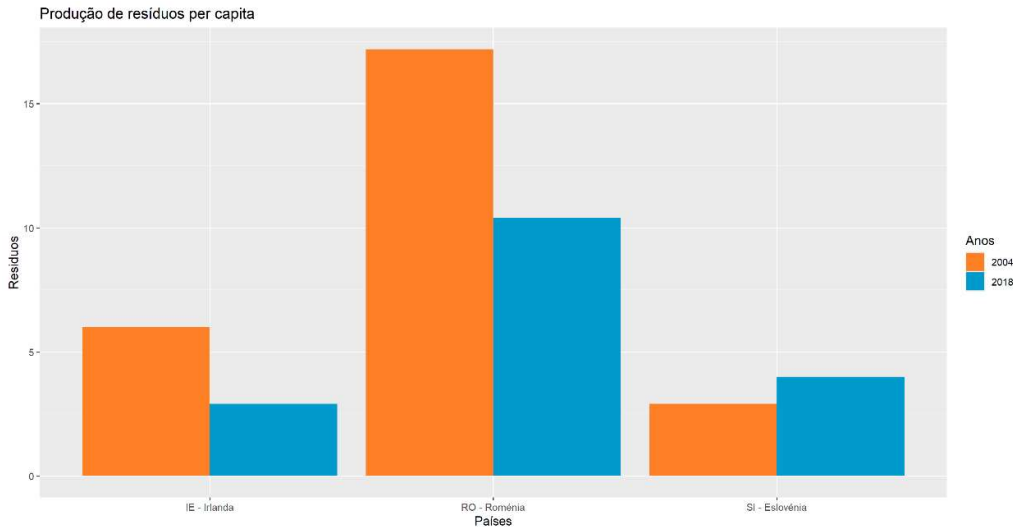
Exercício 1

```r
library(ggplot2)
library(xlsx)
library(reshape2)
residuos <- read.xlsx("ResiduosPerCapita.xlsx", sheetName = "Quadro")
nome <- c(residuos[18,1], residuos[25,1], residuos[35,1])
Ano_2004 <- as.double(c(residuos[18,2], residuos[25,2], residuos[35,2]))
Ano_2018 <- as.double(c(residuos[18,3], residuos[25,3], residuos[35,3]))
df1 <- data.frame(nome,Ano_2004,Ano_2018)
df2 <- melt(df1, id.vars='nome')
Anos <- c('2004','2018')
ggplot(data = df2, aes(x=nome, y=value, fill=variable)) +
geom_bar(stat='identity',position = 'dodge') +
labs(title = "Produção de resíduos per capita", x = "Países",y = "Residuos", fill = "Anos")
scale_fill_manual(labels=Anos, values=c('chocolate1','deepskyblue3'))
```



Através da análise dos gráficos observa-se que a Roménia é o país que mais resíduos produziu per capita em 2004, seguido da Irlanda e depois da Eslovénia.

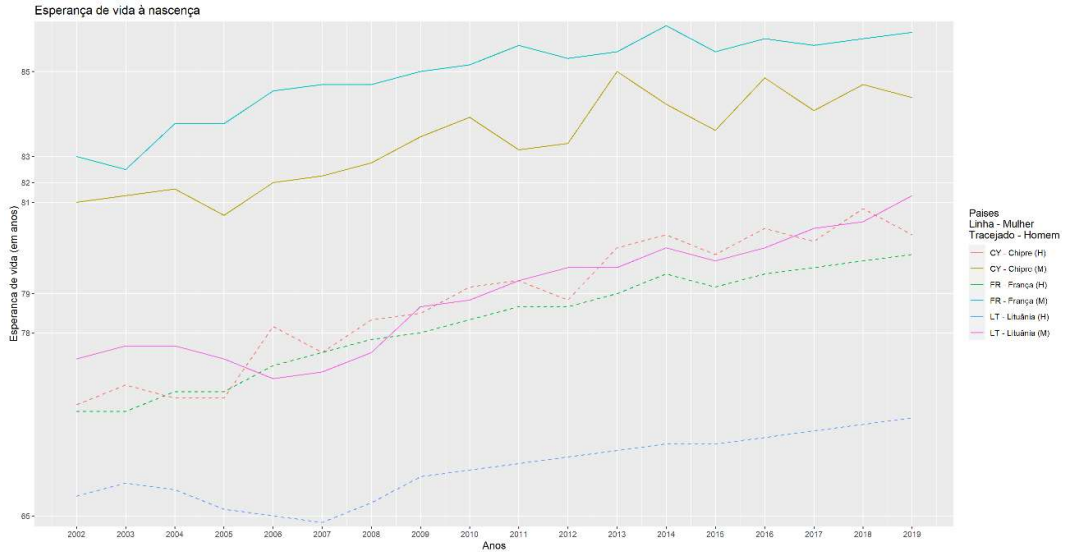
Já em 2018 a Roménia continua como o país com a maior produção de resíduos, apesar de ter sido o país que apresentou a maior queda na produção entre 2004 e 2018.

Já a Eslovénia sobe a sua produção, quando comparada com 2004 (sendo o único dos três em que tal se verificou), passando para o segundo maior produtor de resíduos deste grupo de países.

Por fim a Irlanda é o país que menos produz, tendo o seu valor em 2018 baixado para o mesmo valor que a Eslovénia apresentava em 2004.

Exercício 2

```r
library(ggplot2)
library(xlsx)
ficheiro <- read.xlsx("EsperancaVida.xlsx", sheetIndex = 1)
Homens <- ficheiro[51:68,c(43,51,57)]
Mulheres <- ficheiro[51:68,c(77,85,91)]
Anos <- c(2002:2019)
df1 <- data.frame(Homens, Anos)
df2 <- data.frame(Mulheres, Anos)
ggplot() + geom_line(data=df1, aes(x=Anos, stat(y=Homens[,1]), col="CY - Chipre (H)"), lty = 2) +
geom_line(data=df1, aes(x=Anos,stat(y=Homens[,2]), col="FR - França (H)"), lty = 2) +
geom_line(data=df1, aes(x=Anos, stat(y=Homens[,3]), col="LT - Lituânia (H)"), lty = 2) +
geom_line(data=df2, aes(x=Anos, stat(y=Mulheres[,1]), col="CY - Chipre (M)")) +
geom_line(data=df2, aes(x=Anos, stat(y=Mulheres[,2]), col="FR - França (M)")) +
geom_line(data=df2, aes(x=Anos, stat(y=Mulheres[,3]), col="LT - Lituânia (M)")) +
scale_y_discrete(breaks = seq(60, 90)) + scale_x_continuous(breaks = seq(2002, 2019)) +
labs(title = "Esperança de vida à nascença", y = "Esperanca de vida (em anos)", col = "Paises\nLinha - Mulher\nTracejado - Homem")
```



Através da análise dos gráficos observa-se que entre 2002 e 2019 as esperanças médias de vida aumentaram, independentemente do género ou do país escolhido.

Para além disso ainda se observa que todos os países têm uma esperança média de vida mais alta nas mulheres do que nos homens, quando comparados apenas dados do mesmo país.

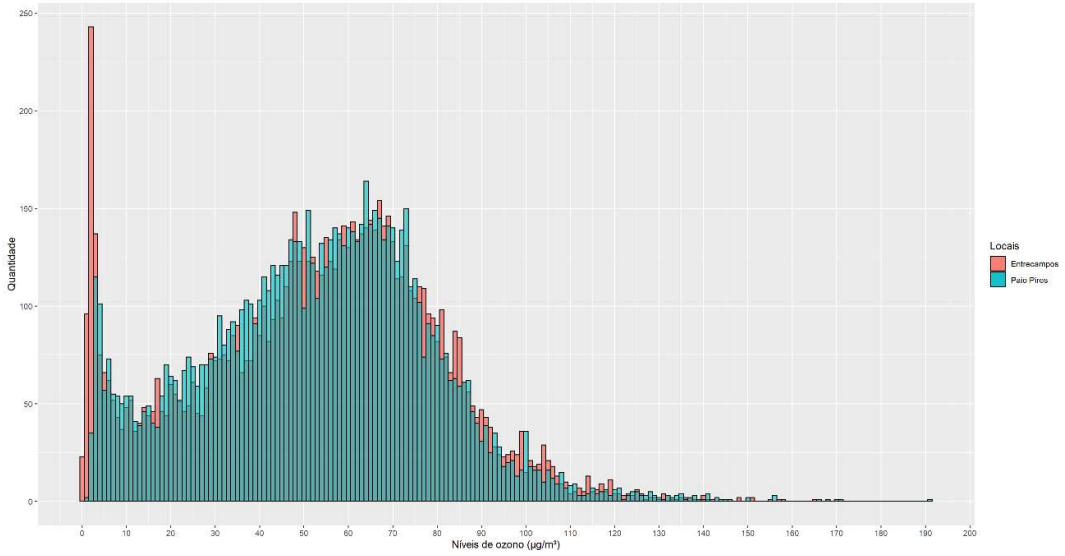
Comparando todos, podemos concluir que a esperança de vida média nas Mulheres é mais alta na França, depois no Chipre e por fim na Lituânia.

Já nos Homens o Chipre encontra-se em primeiro (apenas sendo ultrapassado em 2004 e 2005 pela França), seguido da França e por fim da Lituânia.

Para além disso podemos ver que de entre todas as esperanças de vida das mulheres, a da Lituânia é a única que, dependente do ano, é ultrapassada por Homens do Chipre e de França.

Exercício 3

```r
library(ggplot2)
library(xlsx)
Ar <- read.xlsx("QualidadeARO3.xlsx", sheetName = "Sheet1")
Entrecampos <- as.double(Ar$Entrecampos)
Paio.Pires <- as.double(Ar$Paio.Pires)
df1 <- data.frame(Entrecampos,Paio.Pires)
ggplot() +
geom_histogram(data = df1, aes(x=Entrecampos, fill= "Entrecampos"), binwidth = 1, colour="black", alpha = 0.8) +
geom_histogram(data = df1, aes(x=Paio.Pires, fill= "Paio Pires"), binwidth = 1,  colour ="black", alpha = 0.6) +
scale_x_continuous(breaks=seq(0,200, by = 10)) +
labs(x = "Níveis de ozono (µg/m³)", y = "Quantidade", fill = "Locais")
```



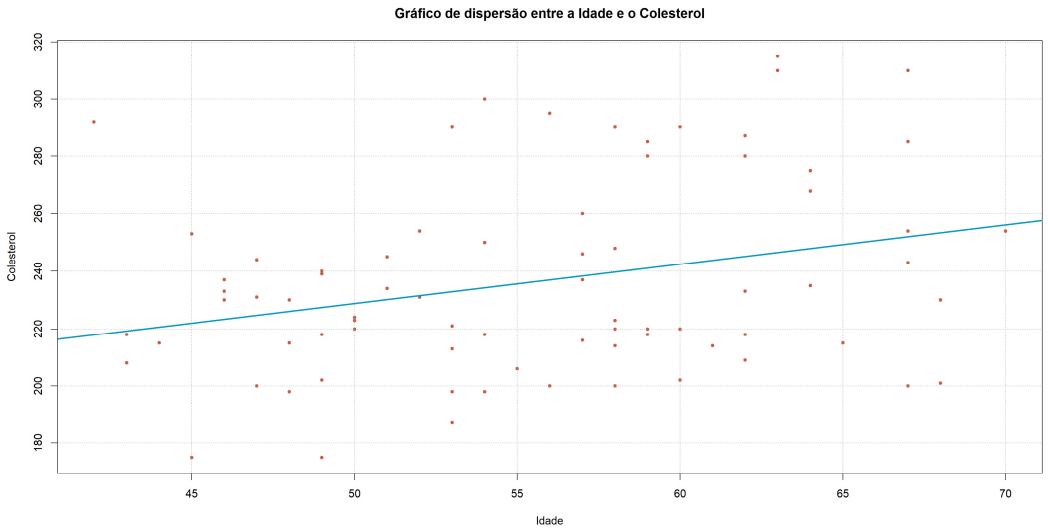
Pela observação da sobreposição dos dois histogramas, podemos observar que Entrecampos apresenta várias alturas em que o nível de ozono é consideravelmente baixo (abaixo de 10 µg/m^3), sendo muito menos visto este comportamento em Paio-Pires.

Já na zona onde o nível de ozono encontra-se entre 20 e 85 µg/m^3 (zona mais comum dos níveis de ozono), podemos observar um maior número de registos em Paio-Pires, sendo várias vezes maior que Entrecampos nessa zona.

Por fim, na zona superior 85 µg/m^3 já existem poucos registos, notando-se assim que não é tão comum um nível de ozono superior a 85 µg/m^3. Mesmo assim, dos registos existentes pode-se observar que a maior parte são de Entrecampos.

Exercício 4

```r
library(ggplot2)
library(xlsx)
Utentes <- read.xlsx("Utentes.xlsx", sheetIndex = 1)
plot(Utentes$Idade, Utentes$Colesterol, pch=20,  col = "coral3",
xlab = "Idade", ylab = "Colesterol", main = "Gráfico de dispersão entre a Idade e o Colesterol")+
grid(col = "darkgrey") +
abline(lm(Utentes$Colesterol~Utentes$Idade), col="deepskyblue3", lwd = 2)
```



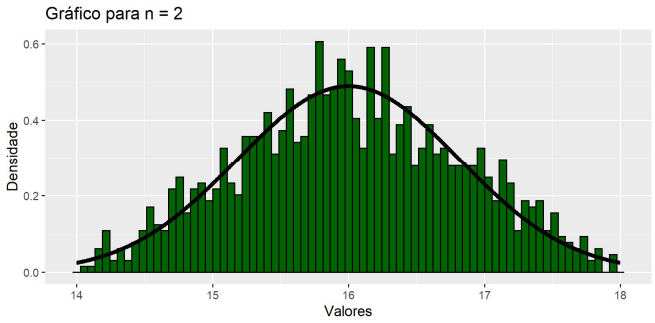
Através do gráfico podemos observar que a tendência é que quanto mais velho o utente maior será o seu colesterol, apesar de que podem existir utentes que tem um colesterol alto mesmo tendo uma idade baixa e vice-versa.

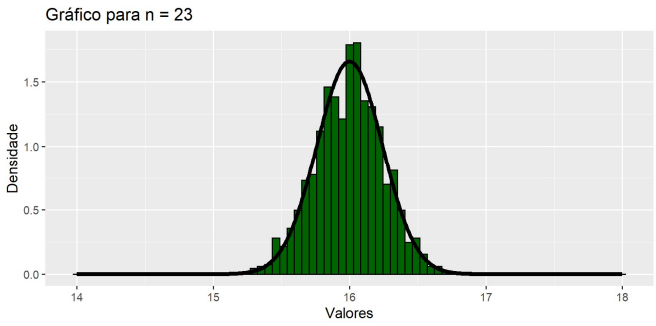
Mesmo assim observando o gráfico de dispersão podemos ver que os pontos estão consideravelmente afastados uns dos outros. A explicação para tal pode dever-se a termos uma amostra muito baixa de utentes (76 utentes), levando a uma média pouco precisa ou também pode ser devido a não existir uma relação tão grande entre o colesterol e a idade.

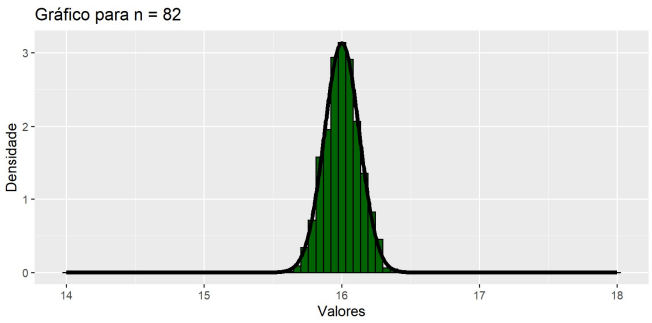
Exercício 6

- Semente = 168
- Dimensões das amostras = 2, 23, 82
- Parâmetros da distribuição uniforme: X~Unif(14,18)

```r
library(ggplot2)
seed = 168
amostras = 1190
n = 2
set.seed(seed)
Valores_rand <- 1:amostras
Valor= seq(14,18, by= 4/(amostras-1))
for (i in 1:amostras){
Valores_rand[i] = mean(runif(n, min = 14, max = 18))
}
Var = ((18-14)^2)/12
desvio <- sqrt((Var/n))
distri <- dnorm(Valor, mean = 16, sd = desvio)
data <-data.frame(Valores_rand)
ggplot() +
geom_histogram(data = data, aes(x = Valores_rand , y = after_stat(density)), col="Black", fill = "DarkGreen", bins = 75)+
geom_line(data = data, aes(x=Valor ,y=distri), size = 1.5) +
labs(x = "Valores", y = "Densidade", title = sprintf("Gráfico para n = %d",n))
```







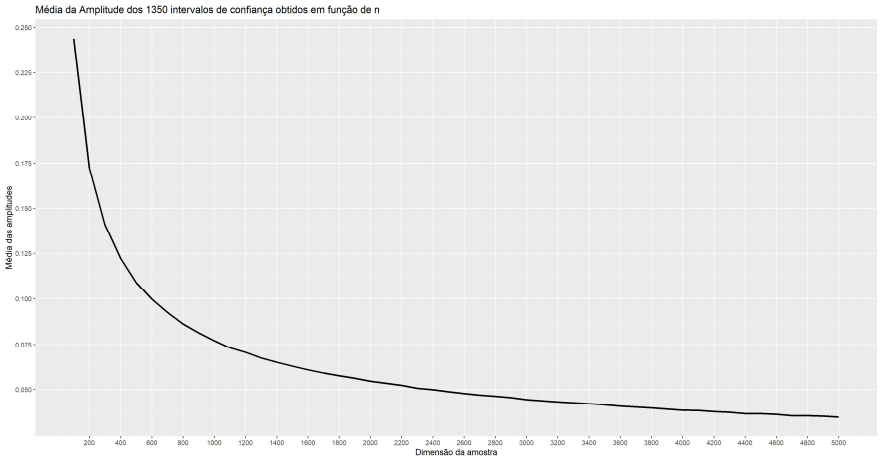
Analisando os histogramas podemos observar que quanto maior o n, mais perto será a distribuição média do valor esperado da distribuição uniforme. Isto acontece porque na distribuição uniforme todos os valores dentro do intervalo têm a mesma probabilidade de sair e, portanto, numa amostra de grande dimensão a média de todos os valores será mais perto do meio dos dois valores, ou seja, neste caso (18+14) /2 = 16. Já em uma amostra de pequena dimensão, como por exemplo n=2, pode acontecer o caso de sair os dois valores perto dos extremos do intervalo o que leva a que as médias fiquem bastante deslocadas do centro.

Quanto à distribuição normal, quando comparada com os histogramas, observa-se que esta fica uma aproximação cada vez melhor quanto maior a dimensão da amostra, estando no caso de n=82 bastante próxima.

Exercício 9

- Semente = 744
- m = 1350
- Lambda = 0.61
- Gama ou (1-alfa) = 0.95

```r
library(ggplot2)
set.seed(744)
m = 1350
n = seq(100,5000, by = 100)
lambda = 0.61
gama = 0.95
Media = 0
X <- matrix(0, nrow = 50, ncol = m)
for (i in 1:50){
for (j in 1:m){
  X[i,j] <- mean(rexp(n,lambda))
}}
b =  qnorm(1-(1-gama)/2)
amp = (2*b)/(X*sqrt(n))
for (i in 1:50){
Media[i] = mean(amp[i, 1:m])
}
df <- data.frame(Media)
ggplot() + geom_line(data = df, aes(x = n, y = Media), size = 1) +
labs(x = "Dimensão da amostra", y = "Média das amplitudes",
title = "Média da Amplitude dos 1350 intervalos de confiança obtidos em função de n") +
scale_x_continuous(breaks=seq(200,5000, by = 200)) +
scale_y_continuous(breaks=seq(0.050,0.25, by = 0.025))
```

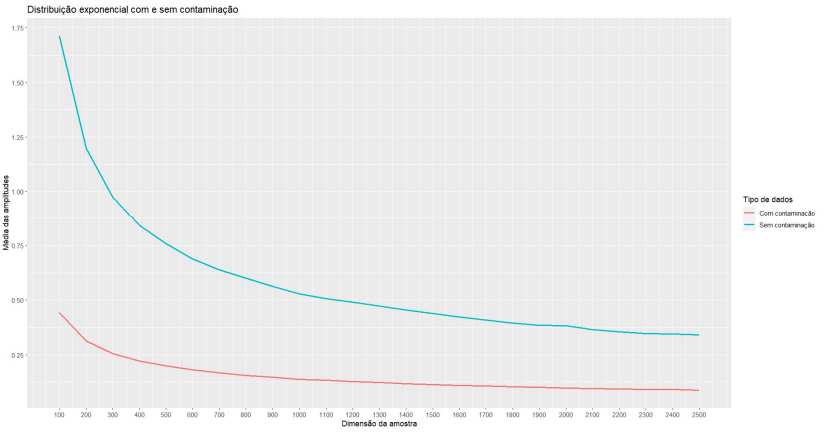


Através da análise do gráfico podemos ver que à medida que a dimensão das amostras aumenta a média de amplitudes dos intervalos de confiança diminuem. Isto ocorre, porque à medida que temos mais amostras, obtemos um resultado cada vez mais preciso e, portanto, a média das amplitudes dos intervalos de confiança vai diminuir.

Exercício 10

- Semente = 291
- m = 1050
- Lambda = 3.15
- Lambda.C = 0.11
- ϵ = 0.1
- Gama ou (1-alfa) = 0.99

```r
library(ggplot2)
set.seed(291)
m = 1050
n = seq(100,2500, by = 100)
lambda1 = 3.15
lambda2 = 0.11
gama = 0.99
Media1 = 0
Media2 = 0
Cont = 0.1
X <- matrix(0, nrow = 25, ncol = m)
Xc <- matrix(0, nrow = 25, ncol = m)
for (i in 1:25){
  for (j in 1:m){
    X[i,j] <- mean(rexp(n,lambda1))
    Xc[i,j]<- mean(rexp(n,lambda2))
    Xc[i,j] = X[i,j]*(1-Cont) + Xc[i,j]*Cont
}}
b =  qnorm(1-(1-gama)/2)
amp1 = (2*b)/(X*sqrt(n))
amp2 = (2*b)/(Xc*sqrt(n))
for (i in 1:25){
  Media1[i] = mean(amp1[i, 1:m])
  Media2[i] = mean(amp2[i, 1:m])
}
nomes <- c(Med = "Sem contaminação", Medc = "Com contaminação")
df <- data.frame(Media1)
ggplot() + geom_line(data = df, aes(x = n, y = Media1, col=nomes[1]), size = 1) +
geom_line(data = df, aes(x = n, y = Media2, col=nomes[2]), size = 1) +
labs(x= "Dimensão da amostra", y = "Média das amplitudes", col = "Tipo de dados",
title = "Distribuição exponencial com e sem contaminação") +
scale_x_continuous(breaks=seq(100,2500, by = 100)) +
scale_y_continuous(breaks=seq(0.25,1.75, by = 0.25))
```



Através dos gráficos podemos observar que a curva sem contaminações tem uma média de amplitudes mais alta do que a contaminada, pois a contaminada tem a influência da distribuição exponencial original (a não contaminada) e de uma outra distribuição exponencial de valor mais pequeno. Como o peso da distribuição exponencial contaminada é a média ponderada das duas, em que o peso da distribuição exponencial de valor mais pequeno é o ϵ, então esta tem de ser menor que a original.