Problem 1 To run this and the next exercise, first load the `R` packages:

```
load(evd,mev,ismev,scales,lubridate,gridExtra,ggplot2,dplyr,tidyr,ggdist,ggpubr,xts)
```

The problem below should familiarise you with some of the R tools used to fit the GEV and GPD models to data. The dataset used contains hourly precipitation (mm) from 1 January 1970 to 31 December 1986 at Eskdalemuir, in southern Scotland. To load it use

```
load("eskrain.RData")
```

This is a time series object, and it may be easier to work with a dataframe, for instance, using the code

```
time.seq <- seq(from=min(date(eskrain)), to=max(date(eskrain)), length=149016)
precip_numeric <- as.numeric(eskrain)
esk.rain <- data.frame(date=as.Date(time.seq), precip=precip_numeric)
```

Try to plot the precipitation time series, and highlight the points corresponding to the exceedances for a threshold u using (e.g.)

```
u <- 5
plot_esk <- plot(esk.rain, type="h", ylab="Hourly rainfall (mm)", xlab="Time")
points(esk.rain[esk.rain$precip > u,], col="red", cex=.25, pch=20)
abline(u, 0, col="red")
```

(a) We first fit the GEV to the daily, weekly and monthly maxima, made using the code

```
### start by taking daily maxima
daily.max <- apply(matrix(esk.rain$precip, ncol=24, byrow=T), max)

### then use daily maxima to compute weekly maxima
weekly.max <- apply(matrix(daily.max, ncol=7, byrow=T), max)

### finally, take monthly maxima, with months of 30 days (!)
monthly.max <- apply(matrix(daily.max, ncol=30, byrow=T), max)
```

Now try to model each of these series and investigate the resulting fit. Do you expect problems with the daily maxima? You could use the code

```
fit.weekly <-  fgev(weekly.max)
par(mfrow=c(2, 2))
plot(fit.weekly)
```

(b) For each fitted model, give the MLEs and their standard errors, and use them to compute 95% confidence intervals for the parameters. Profile log likelihood confidence intervals could instead be computed using

```
par(mfrow=c(1,3))
plot(profile(fit.monthly))
```

What do you conclude? Are these intervals similar to those from the MLEs?

(c) To test whether the shape parameter for the monthly maxima might be zero, use a formal test based on the likelihood ratio statistic between a GEV with three parameters and the Gumbel model, which sets $\xi = 0$:

```
fit.monthly0 <- fgev(monthly.max, shape=0)
ratio <- fit.monthly0$dev-fit.monthly$dev
qchisq(0.95,1)
p <- 1-pchisq(ratio1,1)
```

What do you conclude if you work with a significance level $\alpha = 0.05$?

Problem 2 Consider the Eskdalemuir data again, now considering the exceedances of a threshold u.

(a) The choice of u can be difficult., but a good choice should lead to a stable model for higher thresholds. The mean residual life plot (mean excess plot) is often used:

```
mrlplot(esk.rain$precip)
```

Comment on the stability (or not) for various thresholds u.

One can also look at the estimated parameters of the GPD over a certain range of u:

```
tcplot(esk.rain$precip, tlim=c(0.1,7), model="gpd", nt = 20)
```

Comment on the resulting stability plots. Do the results agree with those for the mean excess plot?

(b) For a chosen threshold u, fit a GPD model to the exceedances and comment on the resulting estimates. To investigate the resulting fit you may use the code

```
# here we take a fixed threshold u=5, but you can choose u based on part (a)
thresh.fit_gpd <- fpot(esk.rain$precip, threshold=5, start=list(scale=1.2,shape=0.1))
# next, we look at the resulting diagnostic plots
par(mfrow=c(1,2))
plot(thresh.fit_gpd)
```

Problem 3 This follows on from the previous `R` exercises using the hourly rainfall from Eskdalemuir, but now estimates return levels. Here's the code to read the data and make what you need:

```
library(evd)
library(lubridate)
load("eskrain.RData")
time.seq <- seq(from=min(date(eskrain)), to=max(date(eskrain)), length=149016)
precip_numeric <- as.numeric(eskrain)
esk.rain <- data.frame(date=as.Date(time.seq), precip=precip_numeric)

plot(esk.rain,type="h") # reality check - the maximum over 17 years is around 15 mm

daily.max <- apply(matrix(esk.rain$precip, ncol=24, byrow=T), 1, max)
weekly.max <- apply(matrix(daily.max, ncol=7, byrow=T), 1, max)
monthly.max <- apply(matrix(daily.max, ncol=30, byrow=T), 1, max)
```

(a) We first fit the GEV to weekly and monthly maxima with the return level for $m = 10$ years (set using `prob`) as one of the parameters:

```
m <- 10
fit.w <- fgev(weekly.max, prob=1/(m*52)) # fit to weekly maxima
fit.m <- fgev(monthly.max, prob=1/(m*12)) # fit to monthly maxima
#  compare profile log likelihoods for the two fits
par(mfrow=c(2,3))
plot(profile(fit.w))
plot(profile(fit.m))
```

You can see the estimates by typing (e.g.) `fit.w`. Do the return level estimates and confidence intervals for the two fits agree? Do you think one is more appropriate than the other, based on your previous analysis of the data?

Try refitting using return levels for higher values of m (25 years, 50 years, ...). What happens to the MLEs and confidence intervals? Why do you think this is? It might help to check the fits in the original parametrization:

```
(fgev(weekly.max))
(fgev(monthly.max))
```

and think about the fitted upper limit (if any) for rainfall.

(b) We now fit the generalized Pareto approximation with the return level for 10 years (set using `mper`) as one of the parameters. We use `npp` to set the number of background observations in each year:

```
u <- 5; m <- 10
(fit <- fpot(esk.rain$precip, threshold=u, mper=m, npp=365.25*24))
par(mfrow=c(1,2))
plot(profile(fit))  # profile log likelihoods
```

Is the profile plot for the return level consistent with the reality check and with the results in (a)? Try setting a larger value for `mper` (e.g., 25 or 100) and see how the profile changes. What is the effect of varying the threshold u?