Case Study 1: Clustering the epileptic.qol Dataset

Group-based trajectory modeling using the flexmix package

```
# install.packages("joineRML")
library(joineRML)
data(epileptic.qol)
# convert days to months
epileptic.qol$time_month <- epileptic.qol$time/30.25
# sort by id and time
epileptic.qol <- epileptic.qol[order(epileptic.qol$id,epileptic.qol$time_month),]

# scaling the clustering variables prior to analysis
epileptic.qol$anxiety_scale <- as.numeric(scale(epileptic.qol$anxiety))
epileptic.qol$depress_scale <- as.numeric(scale(epileptic.qol$depress))
epileptic.qol$aep_scale <- as.numeric(scale(epileptic.qol$aep))
```

group-based trajectory modeling (flexmix package)

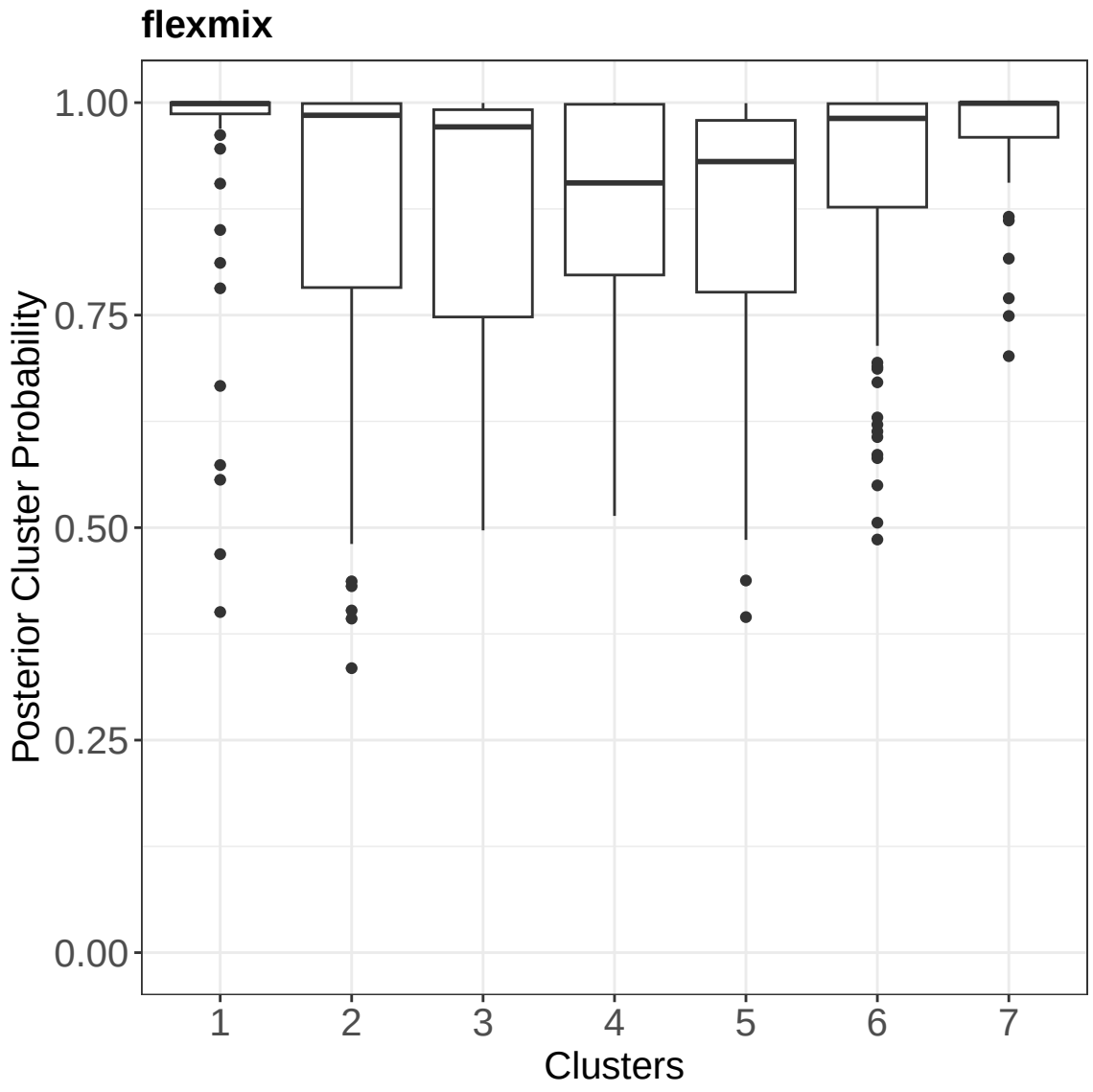
```
# install.packages("flexmix")
library(flexmix)
set.seed(202209)
epileptic.qol$time_month <- epileptic.qol$time/30.25
N <- length(unique(epileptic.qol$id))   # number of subjects
n.obs <- table(epileptic.qol$id)        # number of observations
visit <- NULL
for (i in 1:N){visit <- c(visit,1:n.obs[i])}
epileptic.qol$visit <- visit

epileptic.qol$with.time.month <- epileptic.qol$with.time/30.25
epileptic.qol$anxiety_scale <- as.numeric(scale(epileptic.qol$anxiety) )
epileptic.qol$depress_scale <- as.numeric(scale(epileptic.qol$depress) )
epileptic.qol$aep_scale <- as.numeric(scale(epileptic.qol$aep) )
epileptic.qol$time_month <- epileptic.qol$time/30.25
epileptic.qol.comp <- na.omit(epileptic.qol)

# not run to reduce compiling time
#set.seed(201)
#bic <- NULL
#for (kk in 1:8){
#   fit.flexmix <- flexmix( ~ time_month|id, data = epileptic.qol.comp, k = kk,
#                     model = list(   FLXMRglm(anxiety_scale ~ time_month),
#                                     FLXMRglm(depress_scale ~ time_month),
#                                     FLXMRglm(aep_scale ~ time_month)))
#   bic <- c(bic,summary(fit.flexmix)@BIC)
#}
```

```r
# print the number of clusters with the smallest BIC
#num.clust.flexmix <- which.min(bic); num.clust.flexmix
```

```r
num.clust.flexmix <- 7 # optimal number of clusters based on bic
set.seed(202001)
fit_flexmix <- flexmix( ~ time_month|id, data = epileptic.qol.comp, k = num.clust.flexmix,
                model = list(FLXMRglm(anxiety_scale ~ time_month),
                             FLXMRglm(depress_scale ~ time_month),
                             FLXMRglm(aep_scale ~ time_month)))

# compute and plot the posterior cluster probability
postprob <- apply(posterior(fit_flexmix),1,max)
cluster.tmp <-  apply(posterior(fit_flexmix),1,which.max);

# relabel the clusters
# this step is for making the results compariable to other methods
cluster.flexmix <- (cluster.tmp==5)*1 +
            (cluster.tmp==7)*2 +
            (cluster.tmp==3)*3 +
            (cluster.tmp==2)*4 +
            (cluster.tmp==4)*5 +
            (cluster.tmp==6)*6 +
            (cluster.tmp==1)*7

df.new <- data.frame(id=na.omit(epileptic.qol)$id,
              cluster.flexmix=cluster.flexmix,
              postprob=postprob)
df.new.uq <- df.new[!duplicated(df.new$id, fromLast=TRUE),]
cluster.flexmix <- df.new.uq$cluster.flexmix
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.2.2
```

```r
# Posterior cluster probability
bp.flexmix <- ggplot(df.new.uq, aes(x=factor(cluster.flexmix), y=postprob)) +
            geom_boxplot() + ggtitle("flexmix") +
            xlab("Clusters") + ylab("Posterior Cluster Probability") +
        ylim(c(0,1)) +
        theme_bw() +
        theme(legend.position =  "none",
            plot.title = element_text(size = 15, face = "bold"),
            axis.text=element_text(size=15),
            axis.title=element_text(size=15),
            axis.text.x = element_text(angle = 0 ),
            strip.text.x = element_text(size = 15, angle = 0),
            strip.text.y = element_text(size = 15,face="bold"))
bp.flexmix
```




```
per <- paste(round(100*table(cluster.flexmix)/length(cluster.flexmix)),"%",sep="")
df.new.uq$cluster.flexmix <-   factor(cluster.flexmix ,
                    labels=paste("Cluster ",1:num.clust.flexmix," (",per,")",sep=""))
dat.cluster <- data.frame(df.new.uq$id,df.new.uq$cluster.flexmix)
colnames(dat.cluster) <- c("id","cluster.flexmix")
dnew <- merge(epileptic.qol,dat.cluster,by="id")
```

```
library(cowplot)
# plotting the first feature (anxiety) by clusters
p1.flexmix <- ggplot(data =dnew, aes(x =time_month, y = anxiety,
                color=cluster.flexmix,
                linetype=cluster.flexmix,
                fill=cluster.flexmix))+
            ggtitle("flexmix") +
        geom_smooth(aes(x =time_month, y = anxiety,
                    color=cluster.flexmix,
                            linetype=cluster.flexmix,
```

```
                    fill=cluster.flexmix),
                           method = "loess", linewidth = 3,se = FALSE,span=2)+
        theme_bw() +
        theme(legend.position =  "none",
            plot.title = element_text(size = 15, face = "bold"),
            axis.text=element_text(size=15),
            axis.title=element_text(size=15),
            axis.text.x = element_text(angle = 0 ),
            strip.text.x = element_text(size = 15, angle = 0),
            strip.text.y = element_text(size = 15,face="bold")) +
        guides(fill=guide_legend(title=NULL,nrow = 2,byrow=TRUE),
             color=guide_legend(title=NULL,nrow = 2,byrow=TRUE),
               linetype=guide_legend(title=NULL,nrow = 2,byrow=TRUE)) +
        xlab("Time (months)") + ylab("anxiety") +
        ylim(c(min(dnew$anxiety,na.rm=TRUE),max(dnew$anxiety,na.rm=TRUE)))+
        scale_color_manual(values=c("green", "black","blue","red",
                        "purple","goldenrod3","dimgray" ))+
        scale_fill_manual(values=c("green", "black","blue","red",
                        "purple","goldenrod3","dimgray"))

# plotting the second feature (depress) by clusters
p2.flexmix <- ggplot(data =dnew, aes(x =time_month, y = depress,
                color=cluster.flexmix,
                linetype=cluster.flexmix,
                fill=cluster.flexmix))+
        ggtitle("flexmix") +
         geom_smooth(aes(x =time_month, y = depress,
                    color=cluster.flexmix,
                           linetype=cluster.flexmix,
                    fill=cluster.flexmix), method = "loess",
                           linewidth = 3,se = FALSE,span=2)+
        theme_bw()  +
        theme(legend.position = "none",
            plot.title = element_text(size = 15, face = "bold"),
            axis.text=element_text(size=15),
            axis.title=element_text(size=15),
            axis.text.x = element_text(angle = 0 ),
            strip.text.x = element_text(size = 15, angle = 0),
            strip.text.y = element_text(size = 15,face="bold")) +
        guides( fill=guide_legend(title=NULL,nrow = 2,byrow=TRUE),
                color=guide_legend(title=NULL,nrow = 2,byrow=TRUE),
                linetype=guide_legend(title=NULL,nrow = 2,byrow=TRUE)) +
        xlab("Time (months)") + ylab("depress") +
            ylim(c(min(dnew$depress,na.rm=TRUE),max(dnew$depress,na.rm=TRUE)))+
        scale_color_manual(values=c("green", "black","blue","red",
                        "purple","goldenrod3","dimgray" ))+
        scale_fill_manual(values=c("green", "black","blue","red",
                        "purple","goldenrod3","dimgray"))
# plotting the third feature (aep) by clusters
p3.flexmix <- ggplot(data =dnew, aes(x =time_month, y = aep,
            color=cluster.flexmix,
            linetype=cluster.flexmix,
            fill=cluster.flexmix))+
```

```
        ggtitle("flexmix") +
        geom_smooth(aes(x =time_month, y = aep,
                    color=cluster.flexmix,
                           linetype=cluster.flexmix,
                    fill=cluster.flexmix), method = "loess",
                           linewidth = 3,se = FALSE,span=2)+
        theme_bw() +
        theme(legend.position = "none",
            plot.title = element_text(size = 15, face = "bold"),
            axis.text=element_text(size=15),
            axis.title=element_text(size=15),
            axis.text.x = element_text(angle = 0 ),
            strip.text.x = element_text(size = 15, angle = 0),
            strip.text.y = element_text(size = 15,face="bold")) +
        guides( fill=guide_legend(title=NULL,nrow = 2,byrow=TRUE),
                color=guide_legend(title=NULL,nrow = 2,byrow=TRUE),
                linetype=guide_legend(title=NULL,nrow = 2,byrow=TRUE)) +
        xlab("Time (months)") + ylab("aep") +
        ylim(c(min(dnew$aep,na.rm=TRUE),max(dnew$aep,na.rm=TRUE)))+
        scale_color_manual(values=c("green", "black","blue","red",
                    "purple","goldenrod3","dimgray" ))+
        scale_fill_manual(values=c("green", "black","blue","red",
                    "purple","goldenrod3","dimgray"))
#-------------------------------------------------------------------------------------#
# extract a legend
legend.flexmix <- get_legend(ggplot(data =dnew, aes(x =time_month, y = depress,
                color=cluster.flexmix,
                linetype=cluster.flexmix,
                fill=cluster.flexmix))+
                        ggtitle("flexmix") +
                        geom_smooth(aes(x =time_month, y = depress,
                          color=cluster.flexmix,
                                        linetype=cluster.flexmix,
                          fill=cluster.flexmix), method = "loess",
                                    linewidth = 3,se = FALSE,span=2)+
                        theme_bw()  +
                        theme(legend.position = c(0.5,0.5),
                              legend.text =element_text(size = 12),
                              plot.title = element_text(size = 15, face = "bold"),
                              axis.text=element_text(size=15),
                              axis.title=element_text(size=15),
                              axis.text.x = element_text(angle = 0 ),
                              strip.text.x = element_text(size = 15, angle = 0),
                              strip.text.y = element_text(size = 15,face="bold")) +
                        guides(fill=guide_legend(title=NULL,nrow = 7,byrow=TRUE),
                               color=guide_legend(title=NULL,nrow = 7,byrow=TRUE),
                               linetype=guide_legend(title=NULL,nrow = 7,byrow=TRUE)) +
                        xlab("Time (months)") + ylab("depress") +
                        ylim(c(min(dnew$depress,na.rm=TRUE),max(dnew$depress,na.rm=TRUE)))+
                        scale_color_manual(values=c("green", "black","blue","red",
                           "purple","goldenrod3","dimgray"))+
                        scale_fill_manual(values=c("green", "black","blue","red",
                           "purple","goldenrod3","dimgray"))
```

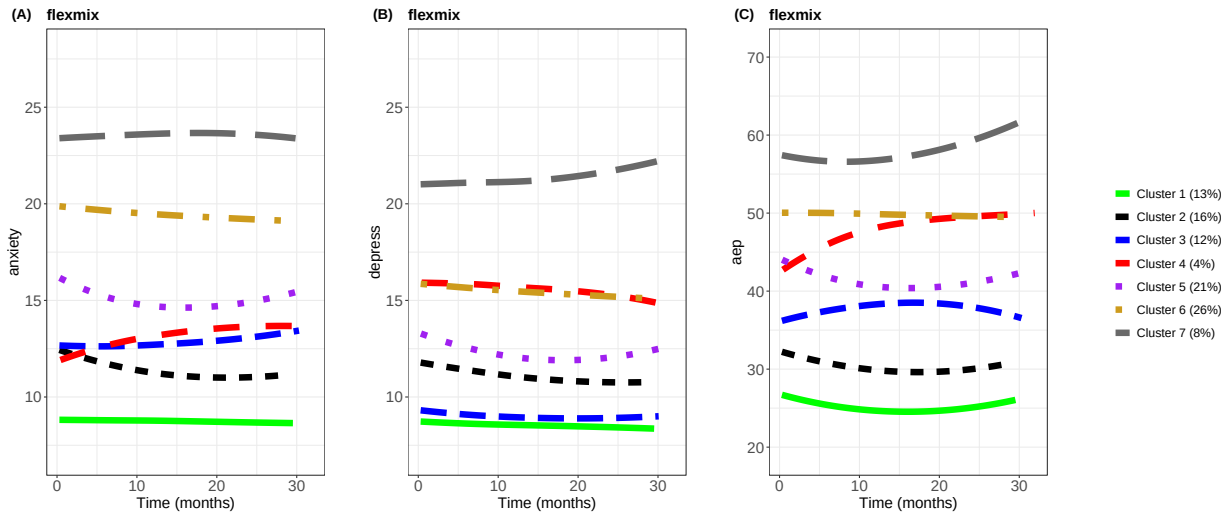
```
)
```

```
## Warning: Removed 53 rows containing non-finite values (`stat_smooth()`).
```

```
plot_grid(p1.flexmix,NULL,p2.flexmix,NULL,p3.flexmix,NULL,legend.flexmix,
        labels=c("(A)","", "(B)","","(C)","",""),
            ncol = 7,
            rel_widths = c(1,0.1,1,0.1,1,0.1,0.5))
```

```
## Warning: Removed 55 rows containing non-finite values (`stat_smooth()`).
## Removed 53 rows containing non-finite values (`stat_smooth()`).
```

```
## Warning: Removed 86 rows containing non-finite values (`stat_smooth()`).
```




```
# weighted cox model by the posterior cluster probability
# Keep last observation per id
dnew_uq <- merge(dnew[!duplicated(dnew$id, fromLast=TRUE),],df.new[,c(1,3)], by="id")
fit <- survfit(Surv(with.time.month, with.status2) ~ cluster.flexmix,
               data = dnew_uq)
res.cox <- coxph(Surv(with.time.month, with.status2) ~ cluster.flexmix,
                 weights=postprob, data =  dnew_uq)
pvalue <-  ifelse(summary(res.cox)$sctest[3] >= 0.0001,
                  summary(res.cox)$sctest[3],'<0.0001')

# Visualize with survminer
library(survminer)
```

```
## Warning: package 'ggpubr' was built under R version 4.2.2
```

```
library(survival)
names(fit$strata) <- paste("Cluster ",1:num.clust.flexmix," (",per,")",sep="")
gp_survival.flexmix <- ggsurvplot(fit, data = dnew_uq,title="flexmix",
                          risk.table = TRUE,
                 risk.table.y.text.col = TRUE,
                 pval = pvalue,
                          legend = "bottom", # conf.int = TRUE,
                          xlab = "Time (months)",
                 legend.title="Clusters",
                          ggtheme = theme_bw() +
```

```
                                theme(legend.position ="none",legend.title=element_blank(),
                                      plot.title = element_text(size = 15, face = "bold"),
                                      axis.text=element_text(size=15),
                                      axis.title=element_text(size=15),
                                      strip.text.x = element_text(size=15),
                                      strip.text.y = element_text(size=15)))

gp_survival.flexmix$plot <- gp_survival.flexmix$plot +
        guides(fill=guide_legend(title=NULL,nrow = 3,byrow=TRUE),
             color=guide_legend(title=NULL,nrow = 3,byrow=TRUE),
             linetype=guide_legend(title=NULL,nrow = 3,byrow=TRUE))+
        scale_color_manual(values=c("green", "black","blue",
                                    "red","purple","goldenrod3","dimgray" ))+
        scale_fill_manual(values=c("green", "black","blue",
                                   "red","purple","goldenrod3","dimgray" ))
gp_survival.flexmix$plot
```