

# Case Study 2: Clustering the PBC Dataset

## Descriptive statistics

### Load Dataset

```
# install.packages("mixAK")
library(mixAK)

## Warning: package 'lme4' was built under R version 4.2.2

data(PBCseq)
PBCseq_uq <- PBCseq[!duplicated(PBCseq$id, fromLast=TRUE),] # Keep last observation per ID
# patients known to be alive and without liver transplantation at 910 days of follow-up
idx <- unique(PBCseq[PBCseq$alive>910,]$id)
dnew910 <- PBCseq[PBCseq$id %in% idx,]
# Keep last observation per ID
dnew910_uq <- dnew910[!duplicated(dnew910$id, fromLast=TRUE),]

dnew910$time <- dnew910$month
dnew910$time <- dnew910$month - mean(dnew910$month, na.rm=TRUE)
dnew910$time2 <- dnew910$time^2

# use only data before 910 days (2.5 years)
dnew910.before <- dnew910[dnew910$day<=910,];
```

### Make Spaghetti Plots to Visualize the Trajectory Patterns

```
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.2.2
library(cowplot)

gp1 <- ggplot(data = dnew910.before, aes(x = month, y = lbili))+
  geom_point(size=2,alpha=0.5) +
  geom_line(aes(x = month, y = lbili,group=id ),size=1.5,alpha=0.5)+
  geom_smooth(method = "loess", size = 3,se = FALSE,span=2)+
  theme_bw() +
  theme(legend.position = "bottom",legend.title=element_blank(),
        plot.title = element_text(size = 12, face = "bold"),
        axis.text=element_text(size=16),
        axis.title=element_text(size=16),
        legend.text=element_text(size=10),
        axis.text.x = element_text(angle = 0 ),
        strip.text.x = element_text(size = 16, angle = 0),
        strip.text.y = element_text(size = 16,face="bold")) +
  xlab("months") + ylab("lbili")
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

```
gp2 <- ggplot(data = dnew910.before, aes(x =month, y =lalbumin))+
  geom_point(size=2,alpha=0.5) +
  geom_line(aes(x = month, y = lalbumin,group=id ),size=1.5,alpha=0.5)+
  geom_smooth(method = "loess", size = 3,se = FALSE,span=2)+
  theme_bw() +
  theme(legend.position = "bottom",legend.title=element_blank(),
        plot.title = element_text(size = 12, face = "bold"),
        axis.text=element_text(size=16),
        axis.title=element_text(size=16),
        legend.text=element_text(size=10),
        axis.text.x = element_text(angle = 0 ),
        strip.text.x = element_text(size = 16, angle = 0),
        strip.text.y = element_text(size = 16,face="bold")) +
  xlab("months") + ylab("lalbumin")
gp3 <- ggplot(data = dnew910.before, aes(x =month, y =lalk.phos))+
  geom_point(size=2,alpha=0.5) +
  geom_line(aes(x = month, y = lalk.phos,group=id ),size=1.5,alpha=0.5)+
  geom_smooth(method = "loess", size = 3,se = FALSE,span=2)+
  theme_bw() +
  theme(legend.position = "bottom",legend.title=element_blank(),
        plot.title = element_text(size = 12, face = "bold"),
        axis.text=element_text(size=16),
        axis.title=element_text(size=16),
        legend.text=element_text(size=10),
        axis.text.x = element_text(angle = 0 ),
        strip.text.x = element_text(size = 16, angle = 0),
        strip.text.y = element_text(size = 16,face="bold")) +
  xlab("months") + ylab("lalk.phos")
gp4 <- ggplot(data = dnew910.before, aes(x =month, y =lsgot))+
  geom_point(size=2,alpha=0.5) +
  geom_line(aes(x = month, y = lsgot,group=id ),size=1.5,alpha=0.5)+
  geom_smooth(method = "loess", size = 3,se = FALSE,span=2)+
  theme_bw() +
  theme(legend.position = "bottom",legend.title=element_blank(),
        plot.title = element_text(size = 12, face = "bold"),
        axis.text=element_text(size=16),
        axis.title=element_text(size=16),
        legend.text=element_text(size=10),
        axis.text.x = element_text(angle = 0 ),
        strip.text.x = element_text(size = 16, angle = 0),
        strip.text.y = element_text(size = 16,face="bold")) +
  xlab("months") + ylab("lsgot")
gp5 <- ggplot(data = dnew910.before, aes(x =month, y = lplatelet))+
  geom_point(size=2,alpha=0.5) +
  geom_line(aes(x = month, y = lplatelet,group=id ),size=1.5,alpha=0.5)+
  geom_smooth(method = "loess", size = 3,se = FALSE,span=2)+
  theme_bw() +
  theme(legend.position = "bottom",legend.title=element_blank(),
```

```

plot.title = element_text(size = 12, face = "bold"),
axis.text=element_text(size=16),
axis.title=element_text(size=16),
legend.text=element_text(size=10),
axis.text.x = element_text(angle = 0 ),
strip.text.x = element_text(size = 16, angle = 0),
strip.text.y = element_text(size = 16,face="bold")) +
xlab("months") + ylab("lplatelet")

plot_grid(gp1,gp2,gp3,gp4,gp5,
          labels=c("(A)", "(B)", "(C)", "(D)", "(E)"), nrow = 1, align = "v" )

```

```

## Warning: Removed 5 rows containing non-finite values (`stat_smooth()`).
## Warning: Removed 5 rows containing missing values (`geom_point()`).
## Warning: Removed 4 rows containing missing values (`geom_line()`).
## Warning: Removed 15 rows containing non-finite values (`stat_smooth()`).
## Warning: Removed 15 rows containing missing values (`geom_point()`).
## Warning: Removed 10 rows containing missing values (`geom_line()`).

```

