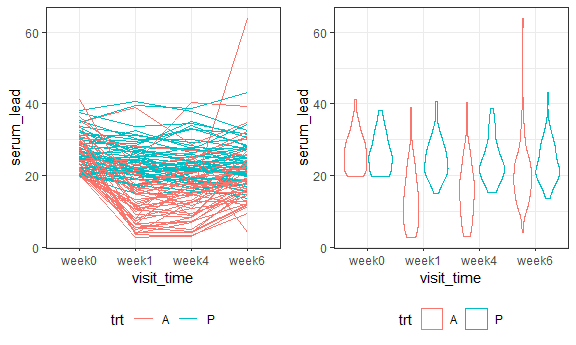
p8157\_LDA\_hw1

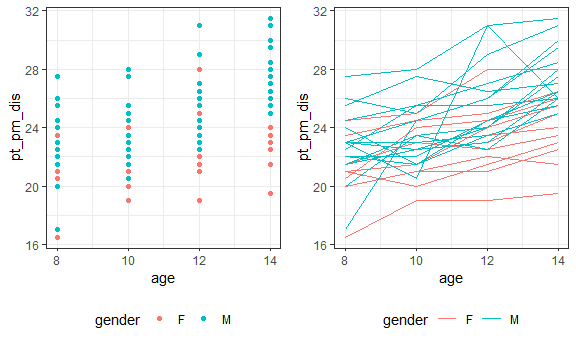
Chunxiao Zhai cz2544

Part B

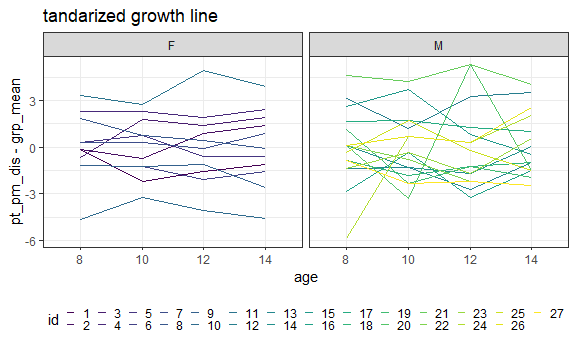
tlc = read.table("TLC.dat")  
colnames(tlc) = c("id", "trt", "week0", "week1", "week4", "week6")  
# matplot(t(tlc), type = "l")  
tlc\_df = as.data.frame(tlc) %>%   
 pivot\_longer(  
 week0:week6,  
 names\_to = "visit\_time",   
 values\_to = "serum\_lead")   
   
tlc\_p1 = tlc\_df %>%   
 ggplot( aes(x=visit\_time, y = serum\_lead, group = id, color = trt))+  
 geom\_line()  
  
tlc\_p2 = tlc\_df %>% group\_by(trt) %>%   
 ggplot( aes(x=visit\_time, y = serum\_lead, color = trt))+  
 geom\_violin()  
  
tlc\_p1+tlc\_p2



dental = read.table("dental.dat")  
colnames(dental) = c("id", "gender", "8", "10", "12", "14")  
dtl\_df = as.data.frame(dental) %>%   
 pivot\_longer(  
 "8":"14",  
 names\_to = "age",   
 values\_to = "pt\_pm\_dis") %>%   
 mutate(age = as.numeric(age))   
  
dtl\_p1 = dtl\_df %>%   
 ggplot( aes(x=age, y = pt\_pm\_dis, color = gender))+  
 geom\_point()  
   
dtl\_p2 = dtl\_df %>%   
 ggplot( aes(x=age, y = pt\_pm\_dis, group = id, color = gender))+  
 geom\_line()  
  
dtl\_mean = dtl\_df %>% mutate(age=as.factor(age)) %>%   
 group\_by(age,gender) %>% mutate(grp\_mean=mean(pt\_pm\_dis)) %>% ungroup()  
dtl\_p3 = dtl\_mean %>% mutate(id=as.factor(id)) %>%   
 ggplot( aes(x=age, y = pt\_pm\_dis-grp\_mean,group = id, color = id))+  
 geom\_line()+ facet\_grid(~gender) +  
 labs( title = "tandarized growth line" ) +   
 viridis::scale\_color\_viridis(  
 name = "id",   
 discrete = TRUE)+  
 theme(legend.key.size = unit(.25, "cm"))+  
 guides(col = guide\_legend(ncol=14))  
  
  
dtl\_p1+dtl\_p2



dtl\_p3



zeb = read.table("ZERBE2.DAT")  
colnames(zeb) = c("grp","id", "0", "0.5", "1", "1.5","2","3")  
  
#EDA  
zeb\_df = as.data.frame(zeb) %>%   
 pivot\_longer(  
 "0":"3",  
 names\_to = "time",   
 values\_to = "serum\_pip") %>%   
 mutate(grp = factor(grp, labels=c("ctl","ob")))  
  
zeb\_p1 = zeb\_df %>%   
 ggplot( aes(x = time, y = serum\_pip, color = grp))+  
 geom\_boxplot()  
   
zeb\_p2 = zeb\_df %>%   
 ggplot( aes(x = time, y = serum\_pip, color = grp, group = id))+  
 geom\_line()+ facet\_grid(~grp)  
  
# Hotelling test  
zeb\_h01 = cbind(zeb[1],zeb[3:8])  
colnames(zeb\_h01) = c("grp", "t0", "t0.5", "t1", "t1.5","t2","t3")  
fit\_a = hotelling.test(.~grp, data = zeb\_h01)  
fit\_a

## Test stat: 8.5531   
## Numerator df: 6   
## Denominator df: 26   
## P-value: 3.495e-05

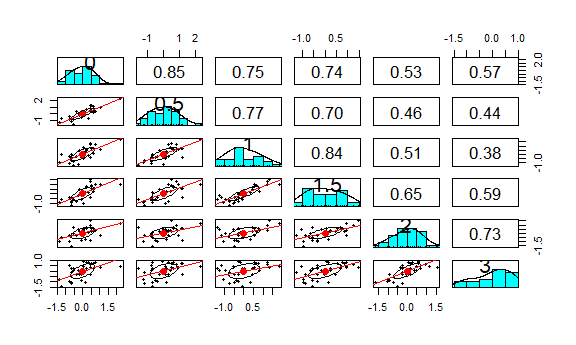
# reject H0 (a) at 0.05 alpha level  
  
zeb\_h02 = zeb\_h01 %>% mutate(it1=t0.5-t0, it2=t1-t0.5,   
 it3=t1.5-t1, it4=t2-t1.5,   
 it5=t3-t2) %>% select(grp,it1:it5)  
fit\_b = hotelling.test(.~grp, data = zeb\_h02)  
fit\_b

## Test stat: 8.1805   
## Numerator df: 5   
## Denominator df: 27   
## P-value: 8.344e-05

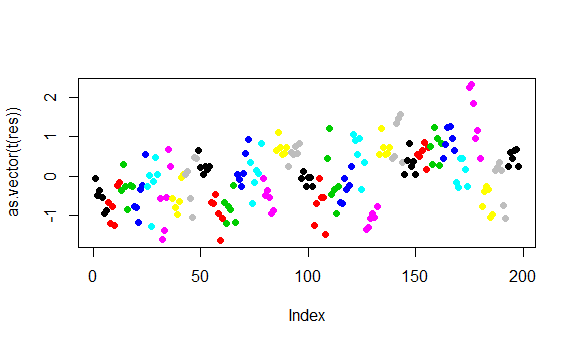
# reject H0 (b) at 0.05 alpha level  
  
fit\_c = hotelling.test(it5~grp, data = zeb\_h02)  
fit\_c

## Test stat: 0.41711   
## Numerator df: 1   
## Denominator df: 31   
## P-value: 0.5231

# can not reject H0 (a) at 0.05 alpha level  
  
# EDA  
mu = apply(zeb[3:8],2,mean)  
res = sweep(zeb[3:8],2,mu)  
cols = rep(1:33,rep(6,33))  
pairs.panels(res,lm=TRUE)



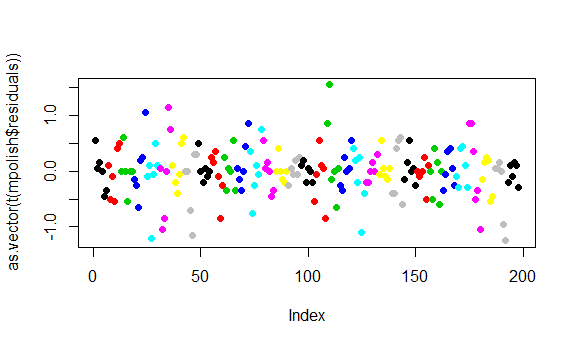
# rev vs mean plot  
plot(as.vector(t(res)), col=cols, pch=19)



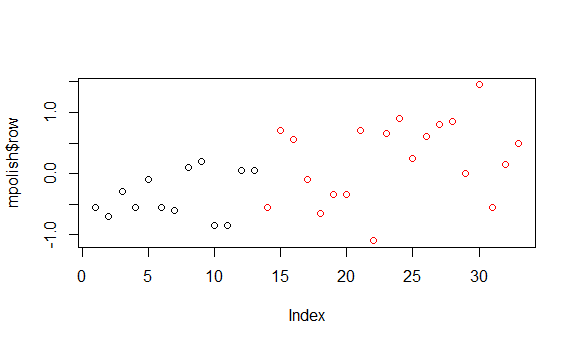
# rev vs median plot  
mpolish = medpolish(zeb[3:8])

## 1: 64  
## 2: 60.5  
## Final: 60.05

plot(as.vector(t(mpolish$residuals)), col=cols,pch=19)



# subject vs median  
plot(mpolish$row, col=zeb$grp)



# ACF  
acf(as.vector((t(mpolish$residuals))))

