```
# analysis of pretest results

setwd('/home/allgoodguys/Documents/Studying/Lund_PhD/epistles/
005_with-Lima/analysis')
library(lme4)
library(ggplot2)

df = read.csv('data/data_pretest_aggreg.csv') [, -1]
# aggregate(correct ~ emotion, df, mean)
# aggregate(pretestVal ~ emotion, df, mean)

df$pos = (as.character(df$emotion) %in% c('Achievement',
'Amusement', 'Pleasure', 'Relief'))
# df$pos = (as.character(df$emotion) %in% c('Amusement', 'Pleasure',
'Relief'))  # excluding achievement
table(df$pos)

# dur ~ pos
hist(df$duration)
mod0 = lm(duration ~ pos, df)
plot(mod0)
summary(mod0)

# dur ~ em
hist(df$duration)
mod1 = lm(duration ~ emotion, df)
plot(mod1)
summary(mod1)

# intens ~ em
hist(df$pretestInt)
mod2 = lm(pretestInt ~ emotion, df)
plot(mod2)
summary(mod2)

# correct ~ em
hist(df$correct)
mod3 = lm(correct ~ emotion, df)
plot(mod3)  # bad!!!
summary(mod3)  # not really valid, try disaggregated data instead

# ar ~ pos
hist(df$pretestAr) # bad
mod4 = lm(pretestAr ~ pos, df)
plot(mod4) # bad
summary(mod4)

# intens ~ pos
mod5 = lm(pretestInt ~ pos, df)
plot(mod5)
summary(mod5)

## correlations
colnames(df)
```

```
temp = as.matrix(round(cor(df[, c(3:7)]), 3))
# write.csv(temp, '~/Downloads/temp.csv')
ggplot(df, aes(x = pretestVal, y = correct)) + geom_point() +
geom_smooth()
ggplot(df, aes(x = pretestVal, y = pretestInt)) + geom_point() +
geom_smooth()
ggplot(df, aes(x = pretestVal, y = duration)) + geom_point() +
geom_smooth()
ggplot(df, aes(x = pretestInt, y = correct)) + geom_point() +
geom_smooth() + xlab('Intensity, 0 to 6') + ylab('Proportion
correct')

### Disaggregated pretest data
df_disag = read.csv('data/data_pretest_preprocessed.csv')
df_disag$pos = (as.character(df_disag$emotion) %in% c('Achievement',
'Amusement', 'Pleasure', 'Relief'))
table(df_disag$pos)
# aggregate(correct ~ emotion, df_disag, mean)
# aggregate(pretestVal ~ emotion, df_disag, mean)

df1 = df_disag[!is.na(df_disag$pretestVal), ]  # first 20 subjects -
val, ar, int
df2 = df_disag[!is.na(df_disag$correct), ]     # second 20 subjects
- accuracy

## correct ~ em
mod = glmer(correct ~ emotion + (1|subject) + (1|soundName_short),
df2, family = 'binomial')
summary(mod)
drop1(mod, test = 'Chisq')

## int ~ em
hist(df1$pretestInt)
mod = glmer(pretestInt ~ emotion + (1|subject) + (1|
soundName_short), df1, family = 'gaussian')
summary(mod)
plot(mod)  # hm...
drop1(mod, test = 'Chisq')

## Are there differences between +/- emotions in terms of val / ar /
int?
levels(df1$emotion)

boxplot(pretestVal ~ pos, df1)
boxplot(pretestAr ~ pos, df1)
boxplot(pretestInt ~ pos, df1)

table(df1$pretestAr)
hist(df1$pretestAr)

mod1 = glmer(pretestAr ~ pos + (1|subject) + (1|soundName_short),
df1, family = 'gaussian')
plot(mod1)  # not good
# drop1(mod1, test = 'Chisq')
# safer to turn it around:
mod2 = glmer(pos ~ pretestAr + (1|subject) + (1|soundName_short),
df1, family = 'binomial')
drop1(mod2, test = 'Chisq')

# correct ~ pos
mod2 = glmer(correct ~ pos + (1|subject) + (1|soundName_short), df2,
family = 'binomial')
drop1(mod2, test = 'Chisq')

# intens ~ pos
mod2 = glmer(pos ~ pretestInt + (1|subject) + (1|soundName_short),
df1, family = 'binomial')
drop1(mod2, test = 'Chisq')
```