# Project Part 1

Sara Altman & Hershel Mehta
October 25, 2017

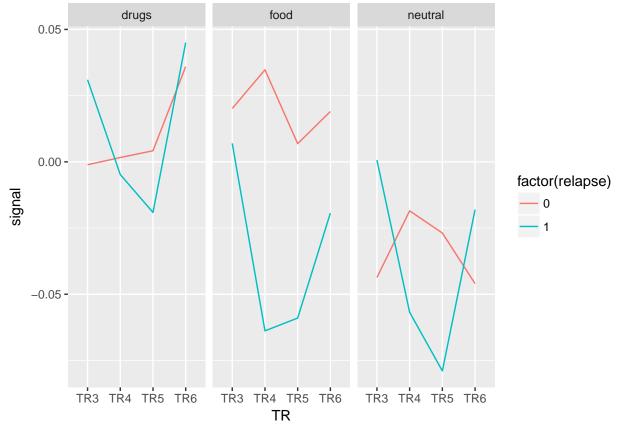
#### Load Data

```
df_trials <- read_csv("relapse_trials.csv")</pre>
## Parsed with column specification:
## cols(
##
     subjects = col_character(),
##
    roi = col_character(),
    condition = col character(),
##
    trial = col_character(),
##
##
    TR3 = col_double(),
##
    TR4 = col_double(),
    TR5 = col_double(),
##
##
     TR6 = col_double()
## )
df_subjects <- read_csv("relapse_subjects.csv")</pre>
## Parsed with column specification:
## cols(
##
     .default = col_double(),
##
     subjid = col_character(),
##
     obstime = col_integer(),
##
     censored = col_integer(),
    relIn6Mos = col_integer(),
##
##
    poly_drug_dep = col_integer(),
##
    depression_diag = col_integer(),
##
     ptsd_diag = col_integer(),
    age = col_integer(),
##
     clinical_diag = col_integer(),
     bis = col_integer()
## )
## See spec(...) for full column specifications.
df_subjid <- df_subjects %>%
  select(subjid, relapse = relIn6Mos)
```

## **Exploratory Data Analysis**

## Time Series (Relapse x Condition)

```
df_trials %>%
  left_join(df_subjid, by = c("subjects" = "subjid")) %>%
  group_by(relapse, condition) %>%
  summarise_at(vars(starts_with("TR"), -trial), mean) %>%
  ungroup() %>%
  slice(1:6) %>%
  gather(key = "TR", value = "signal", TR3:TR6) %>%
  ggplot(aes(x = TR, y = signal)) +
  geom_line(aes(group = relapse, color = factor(relapse))) +
  facet_wrap(~condition)
```

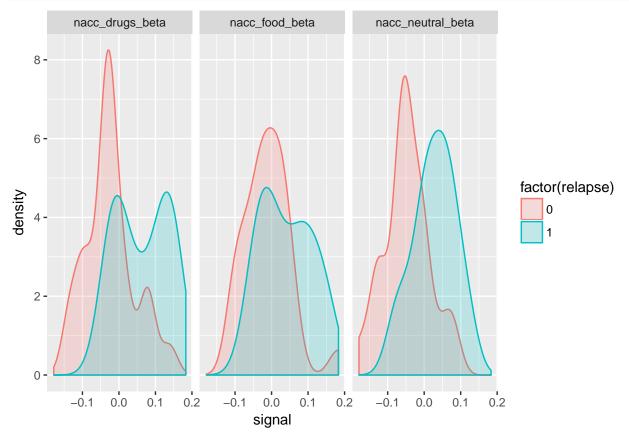


#### Some thoughts:

- Seems like relapsers have blunted brain activation across the board when compared within condition
- Notice that TR4 for drugs and food are decently stimulating for non-relapsers, but it looks like food is very blunted for relapsers. Perhaps a salient feature would be TR4 (drugs) TR4 (food)!
- These appear to be different from Kelly's results from her poster, is it because our sample size has increased? Check with Kelly...

#### Density Plot of Betas (Relapse x Condition)

```
df_subjects %>%
  select(subjid, relapse = relIn6Mos, starts_with("nacc")) %>%
  select(relapse, ends_with("beta")) %>%
  gather(key = "condition", value = "signal", nacc_drugs_beta, nacc_food_beta, nacc_neutral_beta) %>%
  ggplot(aes(x = signal, y = ..density.., color = factor(relapse), fill = factor(relapse))) +
  geom_density(alpha = 0.2) +
  facet_wrap(~ condition)
```



This is a plot of the extracted coefficient for nacc\_betas, which shows that relapsers should have a higher mean beta than non-relapsers...

- Looks like nacc\_drugs\_beta captures some differentiating factor between relapsers and non-relapsers! Perhaps we should try a classification using the nacc\_betas!
- Do we need any kind of transformation here?
- We should find out what regression these coefficients are extracted from to help interpret...

## TODOS

- Try k-fold subject-level classification of relapse in 6 months, using: 1) nacc\_drugs\_beta (extracted betas approach) and 2) TR4 (drugs) TR4 (food) (raw activation approach)
- Try to find some behavioral metrics that also discrimnate relapsers from non-relapsers and establish, then do a k-fold cross-validated model to see how that compares with the predictive accuracy of our model using just brain activity
- Try to find a baseline model of predicting relapsers vs non-relapsers... Perhaps we could try something like a lasso model