Acetaminophen Attenuates Perception of Auditory Stimuli

Lindsay Warrenburg

3/32/20

Previous research has shown that acetaminophen may make you less empathic to others and that acetaminophen can blunt your emotional responses to picture stimuli. Researchers have also found that participants who ingested acetaminophen made attenuated emotional judgments of pleasant and unpleasant pictures from the IAPS.

The goal of the current study was to test whether the emotional blunting effects of acetaminophen extend to the auditory and musical domains.

- **Subgoal 1:** Test whether there is a difference in attenuation among *natural sounds* (like waves or growls), *music*, and *speech*.
- **Subgoal 2:** Test whether acetaminophen blunts *perceived emotion* and *induced emotion* to different extents.

Background

Design

We (Lindsay Warrenburg and Baldwin Way) used a randomized, double-blind, parallel-group, placebo-controlled design to test these hypotheses. The study consisted of two blocks of trial: one for perceived emotion and one for induced emotion. These blocks were counterbalanced across participants. During the drug uptake period, participants completed a number of questionnaires relating to their musical experience, current mood, personality traits, socio-economic status, height and weight, other drugs they may have taken, and so on.

Perceived Emotion Block

The goal of the perceived emotion block was to judge the extent to which a sound conveys a certain affect. For the natural sounds stimuli, we used clips from the International Affective Digital Sounds (IADS). For speech stimuli, we used the Crowd-Sourced Emotional Multimodal Actors Dataset (CREMA-D). And for music stimuli, I used excerpts from film soundtracks, curated by Eerola & Vuoskoski (2011).

- In order to assess the valence of the perceived emotion in the stimuli, we asked *“To what extent does this audio file sound positive?” and “To what extent does this audio file sound negative?’*
- In order to assess the arousal of the perceived emotion in the stimuli, we asked *“To what extent does this audio file sound energetic/arousing?”*
- All three questions utilized an 11-point Likert scale.
- We also asked participants to identify which emotions the audio file represents by checking the appropriate emotions from a list.
  - Participants could select as many or as few emotions as they liked.
  - Once they chose the list of emotion terms, they were asked to choose which emotions (if any) strongly apply to the music.

- By asking participants to choose emotion terms that strongly apply, a three-level response gradient is available for analysis (i.e. does not apply, applies, strongly applies).

Induced Emotion Block

In the section on induced emotion, the stimuli were taken from the same databases as before. The stimuli were different from the ones in the Perceived Emotion section. No speech samples were used because the database instructions explicitly state that they should be used only in studies regarding perceived emotion. Participants were asked

- *"To what extent does this audio file make you feel a positive emotional reaction?"* (on an 11-point Likert scale)
- *"To what extent does this audio file make you feel a negative emotional reaction?"* (on an 11-point Likert scale)
- Once again, participants were asked to identify which emotions the audio file makes them feel by checking the appropriate emotions from a list. They were also asked to choose which emotions strongly apply.

Participants

- 244 participants took part in the experiment.
- A power analysis indicated that 200 participants are needed in order to show a reliable effect, if it is there.

Validation of Stimuli

- We tested the validity of the stimuli -- namely, whether a happy song tended to be perceived as happy, that angry speech was perceived as angry, etc.
- Wefound that the stimuli all behaved as predicted.

Executive Summary

Perceived Emotion

1. The results were consistent with the hypothesis that, **compared to a placebo, acetaminophen blunts people's perception of positive and negative emotions in sound stimuli.** For example, on average, those who took a placebo rated emotional sounds as representing positive emotions at an intensity of 3.99 (out of 10), whereas those who took acetaminophen rated an intensity of positive emotions of 3.58.
2. There was **no difference in arousal ratings between drug and placebo conditions.**

*Chart showing the mean ratings of **perceived emotion** in the drug and placebo conditions*

	Positive Emotion Ratings	Negative Emotion Ratings	Arousal Ratings
Placebo	3.94*	3.86*	4.40
Acetaminophen	3.60*	3.65*	4.33

Induced Emotion

1. Unlike the findings in the study regarding picture stimuli, there was **no difference in the intensity of positive or negative emotions experienced by listeners in the drug and placebo conditions.**

*Chart showing the mean ratings of **induced emotion** in the drug and placebo conditions*

	Positive Emotion Ratings	Negative Emotion Ratings
Placebo	3.74	3.56
Acetaminophen	3.64	3.64

Initialize Workspace

```
/Users/home/Desktop/Research/Tylenol/Analysis
```

```
Click here to toggle on/off the raw code.
```

Goal 1: Long Form Analysis

The initial Qualtrics data was pre-processed in a separate R script. For this part of the analysis, we will build on those dataframes and perform the following tasks:

1. Examination of Missing Data
2. Checking for Duplicates
3. Adding High-Level Summary Features
4. Looking at Mean Differences between Drug and Placebo
    - Overall means
    - Stimulus type (music, natural sounds, and speech)
    - Emotion type (positive, negative, neutral)
    - Arousal and valence levels

We will do this separately for *induced* emotion data and *perceived* emotion data

1a. Induced Emotion Ratings

Read in the data.

```
original inducedLong shape: (4806, 33)
```

Missing Data

We will delete observations with no ID or without a drug/placebo indication. We cannot use these observations for analysis.

```
null shape: (4557, 33)
new inducedLong shape: (4554, 32)
```

Check for Duplicates

Duplicates would indicate an error with the participant ID number fed into the surveys on Qualtrics. These would lead to biased data, so we get rid of them.

```
Number of Induced Stimuli:  18

induced errors in stimuli:  [55, 91, 92, 117, 133, 172]

fixed inducedLong shape: (4338, 32)

induced errors in stimuli:  []
```

Add summary columns

We want to differentiate blunting effects between ***Positive*** and ***Negative*** stimuli, across ***Music***, ***Speech***, and ***Natural Sounds***, and by ***Arousal and Valence*** **(Circumplex model from Russell et al., 1989)** scores.

We will add a new column to summarize each of these features based on emotional theory and *a priori* stimulus analysis.

SoundType	Music	Natural Sounds
Stimulus		
Fear Music 1	241	0
Fear Music 2	241	0
Happy Music 1	241	0
Happy Music 2	241	0
Negative-Valence High-Arousal Human	0	241
Negative-Valence High-Arousal Non-human	0	241
Negative-Valence Low-Arousal Human	0	241
Negative-Valence Low-Arousal Non-human	0	241
Neutral Human	0	241
Neutral Non-human	0	241
Positive-Valence High-Arousal Human	0	241
Positive-Valence High-Arousal Non-human	0	241
Positive-Valence Low-Arousal Human	0	241
Positive-Valence Low-Arousal Non-human	0	241
Sad Music 1	241	0
Sad Music 2	241	0
Tender Music 1	241	0
Tender Music 2	241	0

Comparing Means Between Drug/Placebo

Now we can compare means in ***induced*** emotion ratings in drug and placebo conditions.

First, we prepare the data.

	FixedID	DrugCode	DrugPlacebo	Locus	Stimulus	Positive	Negative	A
2493	94.0	27E	Drug	Induced	Negative-Valence High-Arousal Non-human	0.0	6.0	N M
2093	225.0	27D2	Drug	Induced	Tender Music 2	9.0	0.0	N M

2 rows × 35 columns

Overall Means

We examine the overall mean scores for *positive emotion ratings* and *negative emotion ratings* for participants who took the drug vs. participants who took the placebo.

We then compare the means using t-tests.

```
Induced Emotion Overall Means -- Positive
DrugPlacebo
Drug       3.64
Placebo    3.72
Name: Positive, dtype: float64

t = 0.74
p = 0.46

Induced Emotion Overall Means -- Negative
DrugPlacebo
Drug       3.64
Placebo    3.56
Name: Negative, dtype: float64

t = -0.78
p = 0.43
```

Conclusion: There is no difference in induced emotion ratings for positive emotions or negative emotions between drug and placebo conditions.

Stimulus Type (Music, Natural Sounds)

Although there were no differences in the overall means between drug and placebo conditions, we will look at the mean induced emotion ratings across different stimulus types.

We will *not* conduct t-tests to prevent any potential problems of multiple tests.

```
Induced Emotion Stimulus Type -- Positive
DrugPlacebo  SoundType
Drug         Music             4.72
             Natural Sounds    2.78
Placebo      Music             4.57
             Natural Sounds    3.04
Name: Positive, dtype: float64

Induced Emotion Stimulus Type -- Negative
DrugPlacebo  SoundType
Drug         Music             3.14
             Natural Sounds    4.04
Placebo      Music             3.16
             Natural Sounds    3.89
Name: Negative, dtype: float64
```

Emotion Type (Positive, Negative, Neutral)

Next, we will look at the induced emotion ratings for positive stimuli, negative stimuli, and netural stimuli in both drug and placebo conditions.

```
Induced Emotion Emotion Type -- Positive
DrugPlacebo  PosNeg
Drug         Negative    1.65
             Neutral     1.29
             Positive    6.22
Placebo      Negative    1.76
             Neutral     1.52
             Positive    6.23
Name: Positive, dtype: float64

Induced Emotion Emotion Type -- Negative
DrugPlacebo  PosNeg
Drug         Negative    5.88
             Neutral     3.85
             Positive    1.35
Placebo      Negative    5.78
             Neutral     3.60
             Positive    1.33
Name: Negative, dtype: float64
```

Circumplex Type (Arousal and Valence Info)

Finally, we'll look at induced emotion ratings for different types of emotional stimuli:

- Negative Valence, High Arousal
- Negative Valence, Low Arousal
- Neutral
- Positive Valence, Low Arousal
- Positive Valence, High Arousal

```
Induced Emotion Circumplex -- Positive
DrugPlacebo  Russell
Drug         Neg-Valence High-Arousal    1.43
             Neg-Valence Low-Arousal     1.86
             Neutral                     1.29
             Pos-Valence High-Arousal    6.84
             Pos-Valence Low-Arousal     5.60
Placebo      Neg-Valence High-Arousal    1.56
             Neg-Valence Low-Arousal     1.97
             Neutral                     1.52
             Pos-Valence High-Arousal    6.90
             Pos-Valence Low-Arousal     5.56
Name: Positive, dtype: float64

Induced Emotion Circumplex -- Negative
DrugPlacebo  Russell
Drug         Neg-Valence High-Arousal    6.41
             Neg-Valence Low-Arousal     5.36
             Neutral                     3.85
             Pos-Valence High-Arousal    1.05
             Pos-Valence Low-Arousal     1.66
Placebo      Neg-Valence High-Arousal    6.27
             Neg-Valence Low-Arousal     5.30
             Neutral                     3.60
             Pos-Valence High-Arousal    0.99
             Pos-Valence Low-Arousal     1.67
Name: Negative, dtype: float64
```

Summary

There did not seem to be any significant differences in *induced* emotion ratings in drug vs. placebo conditions.

Before examining *perceived* emotions, we'll clean up the data and look at the pandas Profile Report.

	FixedID	DrugCode	DrugPlacebo	Locus	Stimulus	Positive	Negative	Arc
5	7.0	7J	0	Induced	Fear Music 1	8.0	0.0	No Me
6	9.0	9J	0	Induced	Fear Music 1	4.0	6.0	No Me

2 rows × 35 columns

1b. Perceived Emotion Ratings

Read in the data.

```
original perceivedLong shape: (9842, 33)
```

Missing Data

We will delete observations with no ID or without a drug/placebo indication. We cannot use these observations for analysis.

```
null shape: (9287, 33)
new perceivedLong shape: (9250, 32)
```

Check for Duplicates

Duplicates would indicate an error with the participant ID number fed into the surveys on Qualtrics. These would lead to biased data, so we get rid of them.

```
Number of Perceived Stimuli:  37

perceived errors in stimuli:  [55, 91, 92, 117, 275]

fixed perceivedLong shape: (8880, 32)

perceived errors in stimuli:  []
```

Add summary columns

We want to differentiate blunting effects between ***Positive*** and ***Negative*** stimuli, across ***Music***, ***Speech***, and ***Natural Sounds***, and by ***Arousal and Valence*** scores.

We will add a new column to summarize each of these features based on emotional theory and *a priori* stimulus analysis.

SoundType	Music	Natural Sounds	Speech
Stimulus			
Fear Music 1	240	0	0
Fear Music 2	240	0	0
Fear Music 3	240	0	0
Fear Speech 1	0	0	240
Fear Speech 2	0	0	240
Fear Speech 3	0	0	240
Happy Music 1	240	0	0
Happy Music 2	240	0	0
Happy Music 3	240	0	0
Happy Speech 1	0	0	240
Happy Speech 2	0	0	240
Happy Speech 3	0	0	240
Negative-Valence High-Arousal Human	0	240	0
Negative-Valence High-Arousal Non-human	0	240	0
Negative-Valence Low-Arousal Human	0	240	0
Negative-Valence Low-Arousal Non-human	0	240	0
Neutral Human	0	240	0
Neutral Non-human	0	240	0
Neutral Speech 1	0	0	240
Neutral Speech 2	0	0	240
Neutral Speech 3	0	0	240
Positive-Valence High-Arousal Human	0	240	0
Positive-Valence High-Arousal Non-human	0	240	0
Positive-Valence Low-Arousal Human	0	240	0
Positive-Valence Low-Arousal Non-human	0	240	0
Sad Music 1	240	0	0
Sad Music 2	240	0	0
Sad Music 3	240	0	0
Sad Speech 1	0	0	240
Sad Speech 2	0	0	240
Sad Speech 3	0	0	240
Tender Music 1	240	0	0

SoundType	Music	Natural Sounds	Speech
Stimulus			
Tender Music 2	240	0	0
Tender Music 3	240	0	0
Tender Music 4	240	0	0
Tender Music 5	240	0	0
Tender Music 6	240	0	0

Comparing Means Between Drug/Placebo

Now we can compare means in ***perceived*** emotion ratings in drug and placebo conditions.

We prep the data like we did above.

	FixedID	DrugCode	DrugPlacebo	Locus	Stimulus	Positive	Negative
4058	69.0	2E	Placebo	Perceived	Fear Speech 1	0.0	10.0
4979	192.0	25K2	Drug	Perceived	Happy Speech 1	10.0	0.0

2 rows × 35 columns

Overall Means

We examine the overall mean scores for *positive emotion ratings*, *negative emotion ratings*, and *arousal ratings* for participants who took the drug vs. participants who took the placebo.

We then compare the means using t-tests.

```
Perceived Emotion Overall Means -- Positive
DrugPlacebo
Drug       3.60
Placebo    3.94
Name: Positive, dtype: float64

t = 4.33
p = 0.0

Perceived Emotion Overall Means -- Negative
DrugPlacebo
Drug       3.65
Placebo    3.86
Name: Negative, dtype: float64

t = 2.85
p = 0.0

Perceived Emotion Overall Means -- Arousal
DrugPlacebo
Drug       4.33
Placebo    4.40
Name: Arousal, dtype: float64

t = 1.06
p = 0.29
```

Conclusion:

The results are consistent with the hypothesis that, compared to a placebo, acetaminophen blunts people's perception of positive and negative emotions in sound stimuli (both *p*s < 0.05)

- On average, those who took a placebo rated emotional sounds as representing *positive* emotions at an intensity of 3.94 (on a 0-10 scale), whereas those who took acetaminophen rated an intensity of positive emotions of 3.60.
- Those who took a placebo rated emotional sounds as representing *negative* emotions at an intensity of 3.86 (on a 0-10 scale), whereas those who took acetaminophen rated an intensity of positive emotions of 3.65.

The results are *not* consistent with the hypothesis that, compared to a placebo, acetaminophen blunts people's perception of arousal in sound stimuli.

Stimulus Type (Music, Speech, Natural Sounds)

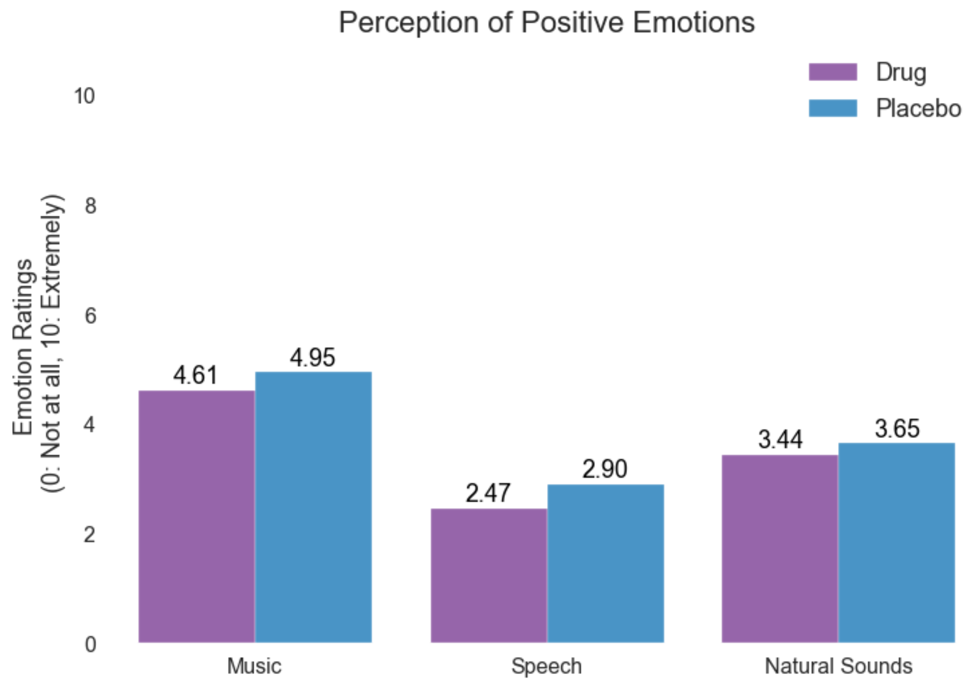
We can examine to see whether the blunting effect of acetaminophen on perceived emotions is similar across different stimulus types.

We will *not* conduct t-tests on this data in order to prevent any potential problems of multiple tests.

```
Perceived Emotion Stimulus Type -- Positive
DrugPlacebo  SoundType
Drug         Music             4.61
             Natural Sounds    3.44
             Speech            2.47
Placebo      Music             4.95
             Natural Sounds    3.65
             Speech            2.90
Name: Positive, dtype: float64

Perceived Emotion Stimulus Type -- Negative
DrugPlacebo  SoundType
Drug         Music             3.18
             Natural Sounds    3.73
             Speech            4.17
Placebo      Music             3.39
             Natural Sounds    3.93
             Speech            4.41
Name: Negative, dtype: float64

Perceived Emotion Stimulus Type -- Arousal
DrugPlacebo  SoundType
Drug         Music             4.79
             Natural Sounds    4.65
             Speech            3.48
Placebo      Music             4.92
             Natural Sounds    4.69
             Speech            3.52
Name: Arousal, dtype: float64
```

Graph Stimulus Type: **Positive**

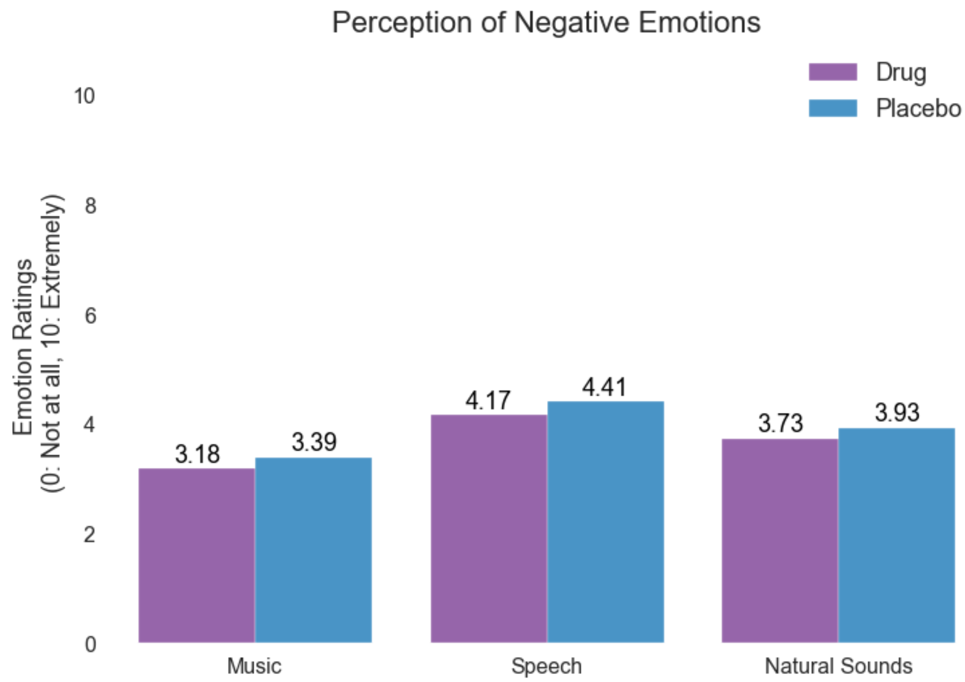
A graph will help showcase the attentuating effect of acetaminophen on perceived positive emotions.



The above graph shows that acetaminophen attenuated the perception of positive emotion similarly in music, natural sounds, and speech stimuli.

Graph Stimulus Type: **Negative**

Now we can look at the attentuating effect of acetaminophen on perceived negative emotions.



The graph shows that acetaminophen also attenuated the perception of negative emotion similarly in music, natural sounds, and speech stimuli.

Comparing this graph to the graph above on positive emotions, we can see that the blunting effect of acetaminophen was similar across ratings of positive and negative emotions.

Emotion Type (Positive, Negative, Neutral)

Now we check whether acetaminophen blunts perceived emotion ratings for positive stimuli, negative stimuli, and netural stimuli.

```
Perceived Emotion Emotion Type -- Positive
DrugPlacebo  PosNeg
Drug         Negative    1.35
             Neutral     1.87
             Positive    6.36
Placebo      Negative    1.65
             Neutral     2.33
             Positive    6.72
Name: Positive, dtype: float64

Perceived Emotion Emotion Type -- Negative
DrugPlacebo  PosNeg
Drug         Negative    6.23
             Neutral     2.68
             Positive    1.40
Placebo      Negative    6.44
             Neutral     3.03
             Positive    1.55
Name: Negative, dtype: float64

Perceived Emotion Emotion Type -- Arousal
DrugPlacebo  PosNeg
Drug         Negative    4.40
             Neutral     1.72
             Positive    5.07
Placebo      Negative    4.39
             Neutral     1.76
             Positive    5.23
Name: Arousal, dtype: float64
```

Circumplex Type (Arousal and Valence Info)

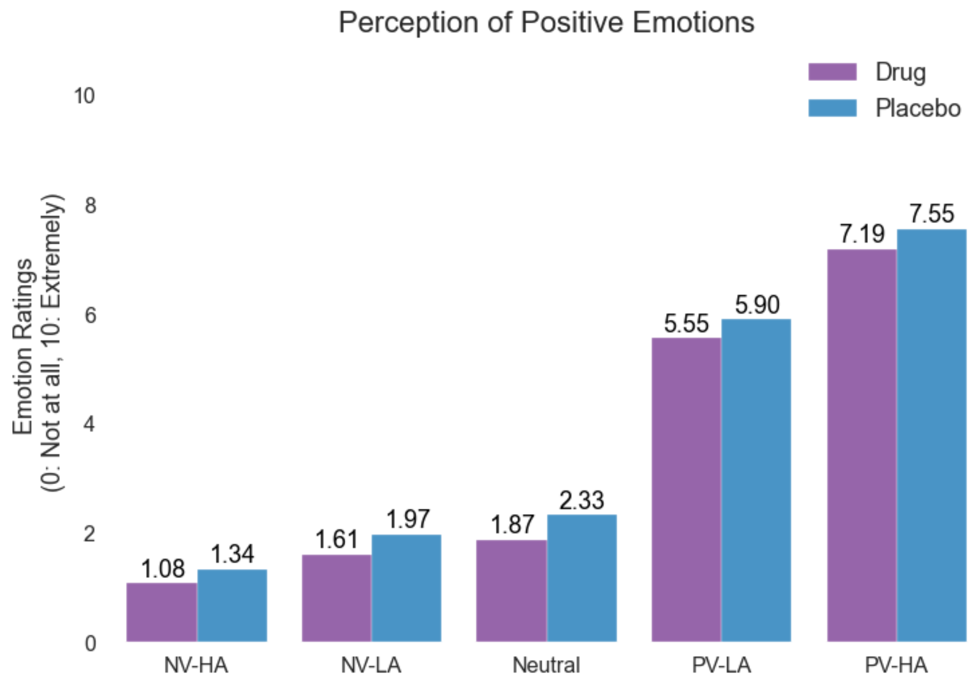
And finally, we can look to see whether the drug attenuation affects perceived emotion ratings for different types of emotional stimuli:

- Negative Valence, High Arousal
- Negative Valence, Low Arousal
- Neutral
- Positive Valence, Low Arousal
- Positive Valence, High Arousal

```
Perceived Emotion Circumplex -- Positive
DrugPlacebo  Russell
Drug         Neg-Valence High-Arousal    1.08
             Neg-Valence Low-Arousal     1.61
             Neutral                     1.87
             Pos-Valence High-Arousal    7.19
             Pos-Valence Low-Arousal     5.55
Placebo      Neg-Valence High-Arousal    1.34
             Neg-Valence Low-Arousal     1.97
             Neutral                     2.33
             Pos-Valence High-Arousal    7.55
             Pos-Valence Low-Arousal     5.90
Name: Positive, dtype: float64

Perceived Emotion Circumplex -- Negative
DrugPlacebo  Russell
Drug         Neg-Valence High-Arousal    6.78
             Neg-Valence Low-Arousal     5.69
             Neutral                     2.68
             Pos-Valence High-Arousal    1.07
             Pos-Valence Low-Arousal     1.73
Placebo      Neg-Valence High-Arousal    6.93
             Neg-Valence Low-Arousal     5.94
             Neutral                     3.03
             Pos-Valence High-Arousal    1.20
             Pos-Valence Low-Arousal     1.89
Name: Negative, dtype: float64

Perceived Emotion Circumplex -- Arousal
DrugPlacebo  Russell
Drug         Neg-Valence High-Arousal    5.54
             Neg-Valence Low-Arousal     3.25
             Neutral                     1.72
             Pos-Valence High-Arousal    6.68
             Pos-Valence Low-Arousal     3.47
Placebo      Neg-Valence High-Arousal    5.46
             Neg-Valence Low-Arousal     3.32
             Neutral                     1.76
             Pos-Valence High-Arousal    6.75
             Pos-Valence Low-Arousal     3.74
Name: Arousal, dtype: float64
```

Graph Circumplex: **Positive**

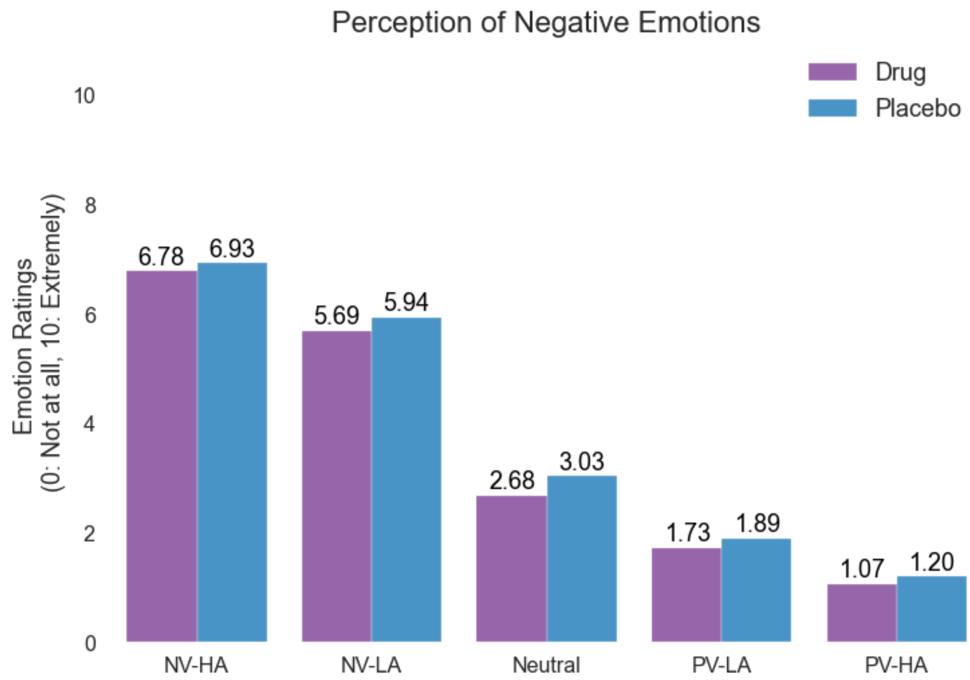
Once again, a graph will help visualize this effect.



Acetaminophen attenuated the perception of positive emotion similarly in stimuli classified as *negative valence-high arousal*, *negative valence-low arousal*, *neutral*, *positive valence-low arousal*, and *positive valence-high arousal*.

Graph Circumplex: **Negative**

And our last graph.



The above graph shows, once again, that acetaminophen blunts the perception of negative emotion similarly in stimuli classified as *negative valence-high arousal*, *negative valence-low arousal*, *neutral*, *positive valence-low arousal*, and *positive valence-high arousal*.

Furthermore, the attenuating effect of acetaminophen was similar across ratings of positive and negative perceived emotions.

Summary

We will make some small changes with the dataframe for future use and summarize the findings with a pandas profile.

	FixedID	DrugCode	DrugPlacebo	Locus	Stimulus	Positive	Negative	A
8	8.0	8J	1	Perceived	Fear Music 1	NaN	NaN	N
9	4.0	4J	1	Perceived	Fear Music 1	3.0	6.0	9

2 rows × 35 columns

1c. Questionnaires

We follow the same data prep steps for the questionnaires, but we won't test any differences in emotion ratings across demographic and psychographic conditions. We'll leave that to the regression analysis.

```
original questionnaires shape: (262, 64)
```

Missing Data

```
new questionnaires shape: (250, 64)

59 columns contain null values

these columns have more than 10% of their values as null/missing:
['WhenLastExerciseHrs', 'MedsInfo', 'MarijuanaInSystem', 'CigarettesInSystem',
'OMSI']

final questionnaires shape: (250, 59)
```

Check for duplicates

```
questionnaire errors in participants:  [91, 92, 117, 144, 193]

fixed questionnaires shape: (240, 59)

questionnaires errors in stimuli:  []
```

2. Imputation of Missing Values

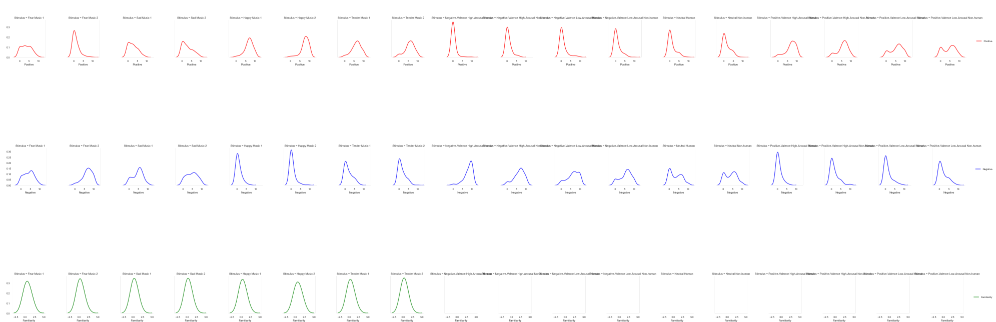
Before we can run a regression analysis, we need to perform some standard feature engineering, such as imputation of missing values, the creation of a single dataframe, and eliminating erroneous data.

In order to tell how we should impute the missing values with the least amount of bias, we need to plot the features that include missing data. The steps are as follows:

1. Find the columns that contain missing data
2. Look at the distributions of these columns
3. Impute using the appropriate method based on the distribution type of the data.

2a. Induced Emotion Ratings

Graph Features



(Note that it was an intentional choice of the researchers to avoid taking familiarity ratings for the natural sounds stimuli)

Because the data are skewed, it's best to use the median for imputation purposes.

We'll impute the values based on their closest neighbors--namely, we'll use the median rating of the feature for *each individual stimulus* to avoid bias.

Impute.

2b. Perceived Emotion Ratings

We repeat the same process for the perceived emotion ratings.

Graph Features

Once again, we'll impute the values based on their closest neighbors--namely, we'll use the median rating of the feature for *each individual stimulus* to avoid bias.

Impute

2c. Combine dataframes

Now we need to create a single dataframe for the regression analysis.

```
perceivedLong shape: (8880, 35)
inducedLong shape: (4338, 35)
questionnaires shape: (240, 59)

[ True  True  True  True  True  True  True  True  True  True  True  True
  True  True  True  True  True  True  True  True  True  True  True  True
  True  True  True  True  True  True  True  True  True  True  True]

perceived/induced shape: (13218, 35)
long shape: (13218, 91)
```

Double check about duplicates or participant ID numbers

All of the data seem to be okay! The only data that are still missing are due to:

- ID numbers in my hypothetical list that didn't correspond to any real participant ID
- Participants who only completed one block of the study (perceived emotion or induced emotion) but not the other one (technical errors, left early, etc.)
- People who completed the pre-test questionnaires but did not complete either stimulus block (perceived or induced emotion). These cases were dropped, as their data won't help the analysis.

3. Feature Engineering

Many of the features--especially in the questionnaire data--need to be transformed before they can be used in any statistical analysis. Some of the categorical columns should be coded on an ordinal, instead of nominal, scale. Others contain messy data that need to be cleaned up using regular expressions.

*In this section, we will use **feature engineering** to perform the necessary tasks to produce a suitable dataframe for analysis.*

```
shape of model dataframe:  (13218, 91)
number of numeric columns:  65
number of categorical columns:  26
number of numeric + categorical columns:  91
```

We can see here that the dataframe currently has 65 numeric columns and 26 categorical columns. We will deal with the numeric columns first, follwed by the categorical columns.

3a. Numeric Column Wrangling

The FixedID column (the participant IDs) are currently numeric. Since the "difference" between persons 1 and 2 should be the same as between persons 1 and 244, the feature should be categorical rather than numeric.

Turn IDs into a Categorical Variable

3b. Categorical Column Wrangling

There are several different ways we need to transform the categorical features. Below, we will perform the following operations:

- Drop theoretically redundant features
- Re-label binary variables as 0 and 1
- Convert ordinal variables from nominal Likert scales to numeric values (e.g., 1-7)
- Transforming uncategorized variables (that require unique transformations)
- One-hot encode certain nominal variables

First, we will examine the unique values in each categorical column.

```
DrugCode unique values:
['8J' '4J' '6J' '7J' '9J' '3J' '12J' '13J' '2J' '10J' '5J' '11J' '14J'
 '1J' '16J' '18J' '19J' '17J' '15J' '20J' '22J' '21J' '23J' '30J' '29J'
 '28J' '27J' '25J' '24J' '26J' '2K' '1K' '5K' '3K' '4K' '9K' '11K' '6K'
 '8K' '7K' '14K' '12K' '18K' '16K' '32J' '13K' '17K' '15K' '24K' '25K'
 '26K' '19K' '22K' '21K' '31K' '32K' '29K' '28K' '2E' '27K' '4E' '1E'
 '23K' '3E' '12E' '10E' '8E' '9E' '11E' '7E' '6E' '14E' '16E' '15E' '17E'
 '13E' '21E' '22E' '19E' '18E' '20E' '26E' '27E' '30K' '30E' '32E' '31E'
 '1D' '28E' '2D' '6D' '4D' '7D' '3D' '5D' '10D' '8D' '11D' '13D' '16D'
 '14D' '20D' '15D' '19D' '17D' '22D' '21D' '26D' '24D' '25D' '23D' '27D'
 '30D' '32D' '28D' '29D' '1J2' '31D' '8J2' '4J2' '5J2' '6J2' '7J2' '3J2'
 '9J2' '14J2' '10J2' '12J2' '11J2' '18J2' '16J2' '17J2' '15J2' '20J2'
 '22J2' '19J2' '13J2' '21J2' '2J2' '24J2' '23J2' '28J2' '26J2' '25J2'
 '29J2' '27J2' '3K2' '4K2' '31J2' '30J2' '2K2' '1K2' '10K2' '6K2' '7K2'
 '9K2' '12K2' '8K2' '11K2' '13K2' '14K2' '16K2' '15K2' '17K2' '18K2'
 '20K2' '19K2' '22K2' '21K2' '24K2' '23K2' '25K2' '26K2' '27K2' '29K2'
 '1D2' '4D2' '3D2' '5D2' '6D2' '7D2' '10D2' '11D2' '9D2' '13D2' '8D2'
 '12D2' '14D2' '21D2' '19D2' '17D2' '15D2' '18D2' '16D2' '23D2' '20D2'
 '22D2' '24D2' '29D2' '31D2' '27D2' '26D2' '1E2' '28D2' '30D2' '25D2'
 '30K2' '31K2' '32K2' '3E2' '5E2' '7E2' '9E2' '4E2' '2E2' '8E2' '6E2'
 '13E2' '16E2' '11E2' '15E2' '14E2' '17E2' '12E2' '19E2' '10E2' '18E2'
```

Drop Redundant Features

Features that are theoretically identical to another feature should be dropped from the analysis, in order to prevent multicollinearity.

DrugCode

```
(13218, 90)
```

Binary Variables

Variables with only two values (for example, ***Yes*** vs. ***No*** responses) should be recoded as 0 and 1.

Locus (Perceived Emotion vs. Induced Emotion)

```
Perceived    67.0
Induced      33.0
Name: Locus, dtype: float64
```

TakeMedsRecentlyYN

```
no      59.0
yes     41.0
Name: TakeMedsRecentlyYN, dtype: float64
```

BirthControlYN

```
no     79.0
yes    21.0
Name: BirthControlYN, dtype: float64
```

ArthritisYN

```
no     94.0
yes     6.0
Name: ArthritisYN, dtype: float64
```

ImmuneDisordersYN

```
no     100.0
yes      0.0
Name: ImmuneDisordersYN, dtype: float64
```

EndocrineDisordersYN

```
no     99.0
yes     1.0
Name: EndocrineDisordersYN, dtype: float64
```

DiabeticYN

```
no             100.0
yes, type 1      0.0
Name: DiabeticYN, dtype: float64
```

Ordinal Variables

Variables on Likert scales should be recoded as ordinal values, as there is an inherent order to these responses.

For example, for the question ***When you take Tylenol, how effective is it at reducing your pain?***, the possible responses were:

- I have never taken Tylenol
- Not effective at all
- Slightly effective
- Moderately effective
- Very effective
- Extremely effective

In this case, we can see that ***Not effective at all*** is less than ***Slightly effective***, and, in turn, ***Slightly effective*** is less than ***Moderately effective*** and so on. By transforming these variables to ordinal, we can capture this difference in any statistical analysis.

Without recoding, it would be difficult to compare ***Extremely effective*** vs. ***Not effective at all***, for instance.

MedsEffectiveness

```
['slightly' 'never taken' 'moderately' 'very' nan 'not at all' 'extremely']
```

Politics

```
['conservative' 'somewhat liberal' 'somewhat conservative' 'liberal'
 'moderate' 'very liberal' 'very conservative' nan]
```

HowMuchLastEat

```
['snack' 'full meal' 'light meal' nan]
```

CigarettesPerDay

```
['0' nan '1-10' '11-20']
```

FrequencyTakeMeds

```
['several times a month' 'never taken' 'at least once a month'
 'less than once a year' 'at least once a year' 'several times a year'
 'at least once a week' 'almost everyday' 'several times a week' nan]
```

WhenLastSick

```
['a few months ago' 'a year or more ago' 'a month ago'
 'a couple of weeks ago' 'a week ago' 'a couple of days ago' nan 'today']
```

MarijuanaFrequency

```
['never' 'several times a week' 'almost every day' 'at least once a month'
 nan 'at least once a year' 'several times a year' 'several times a month'
 'at least once a week']
```

YearUniversity

```
['1' '2' '3' 'None' '5' '4' nan]
```

Transforming Uncategorized Variables

There are additional variables that do not fit any of these categories that we need to transform.

Russell --> HighLow

```
Russell unique values:
 Pos-Valence High-Arousal    22.0
Neg-Valence Low-Arousal     22.0
Neg-Valence High-Arousal    22.0
Pos-Valence Low-Arousal     22.0
Neutral                     13.0
Name: Russell, dtype: float64

HighLow unique values:
 High       44.0
Low        44.0
Neutral    13.0
Name: HighLow, dtype: float64
```

PreferredMeds

```
Ibuprofen                          60.0
Acetaminophen                      15.0
Unknown                             9.0
More Than One Type                  6.0
Naproxen Sodium                     4.0
Do Not Take Meds                    3.0
Acetylsalicylic Acid                1.0
Pseudoephedrine Hydrochloride       0.0
Marijuana                           0.0
Calcium Carbonate                   0.0
Baclofen                            0.0
Name: PreferredMeds, dtype: float64
```

LastAlcoholDays

```
['5' '4' '7' '182.5' 'Never' '6' nan '2' '14' '3' '0.416666666666667' '1'
 '30' '34' '10' '9' '365' '28' '24' '21' '270' '11' '60' '1.5'
 '0.583333333333333' '0.5' '90' '23' '15' '0.625' '0.833333333333333'
 '2.5' '300' '8' '0.666666666666667' '240' '0.75' '20' '17'
 '0.708333333333333' '0.333333333333333']
```

One-Hot Encoding of Nominal Variables

```
Stimulus:  ['Fear Music 1' 'Fear Music 2' 'Fear Music 3' 'Happy Music 1'
 'Happy Music 2' 'Happy Music 3' 'Sad Music 1' 'Sad Music 2' 'Sad Music 3'
 'Tender Music 1' 'Tender Music 2' 'Tender Music 3' 'Tender Music 4'
 'Tender Music 5' 'Tender Music 6' 'Fear Speech 1' 'Fear Speech 2'
 'Fear Speech 3' 'Happy Speech 1' 'Happy Speech 2' 'Happy Speech 3'
 'Sad Speech 1' 'Sad Speech 2' 'Sad Speech 3' 'Neutral Speech 1'
 'Neutral Speech 2' 'Neutral Speech 3'
 'Negative-Valence High-Arousal Human'
 'Negative-Valence High-Arousal Non-human'
 'Negative-Valence Low-Arousal Human'
 'Negative-Valence Low-Arousal Non-human' 'Neutral Human'
 'Neutral Non-human' 'Positive-Valence High-Arousal Human'
 'Positive-Valence High-Arousal Non-human'
 'Positive-Valence Low-Arousal Human'
 'Positive-Valence Low-Arousal Non-human']

PosNeg:  ['Negative' 'Positive' 'Neutral']

HighLow:  ['High' 'Low' 'Neutral']

SoundType:  ['Music' 'Speech' 'Natural Sounds']

Gender:  ['female' 'male' 'prefer not to answer' nan]

PoliticalParty:  ['republican' 'democrat' 'libertarian' 'other' nan]

PreferredMeds:  ['Ibuprofen' 'Do Not Take Meds' 'Acetaminophen' 'Other' nan]

Race:  ['White' 'Asian' 'Black' 'Hispanic' 'White and Hispanic' nan
 'White and Asian' 'Other' 'White and Black' 'Black and Hispanic'
 'Hispanic and Asian' 'Black and Other']
```

Race

Instead of having separate categories for one-hot encoded multiple racial identities, like ***White and Hispanic***, we will transform these columns so that if a person identifies with a single race, the value will be ***1*** and if they do not, the value will be ***0***.

So, each person will have the following one-hot encoded variables:

- Race_Asian
- Race_Black
- Race_Hispanic
- Race_Other
- Race_White

If a person identifies as ***Asian***, they will have the following values:

- ***Race_Asian = 1***
- Race_Black = 0
- Race_Hispanic = 0
- Race_Other = 0
- Race_White = 0

If a person identifies as ***White and Asian***, they will have the following values:

- ***Race_Asian = 1***
- Race_Black = 0
- Race_Hispanic = 0
- Race_Other = 0
- ***Race_White = 1***

```
---- Race_Asian ---
0    79.0
1    21.0
Name: Race_Asian, dtype: float64

---- Race_Black ---
0    92.0
1     8.0
Name: Race_Black, dtype: float64

---- Race_Hispanic ---
0    95.0
1     5.0
Name: Race_Hispanic, dtype: float64

---- Race_Other ---
0    99.0
1     1.0
Name: Race_Other, dtype: float64

---- Race_White ---
1    67.0
0    33.0
Name: Race_White, dtype: float64
```

LastTimeTookMeds

This column is very messy.

- There are nonsensical values, like *15 NA* and *_years*.
- There are also values on all different scales, like *>1 years*, *215 days*, and *5 minutes*.

Here, we will transform the values so that the LastTimeTookMeds values are all measured in *days*.

```
before transformation:  ['4 years' '10 years' '1 years' ' weeks' ' never' '1 mon
ths' '3 weeks'
 '6 months' '3 days' '3 years' '2 days' '>1 weeks' ' months' '>1 years'
 '3 months' '>1 months' '48 hours' '2 months' nan '1 weeks' '5 months'
 '2 weeks' '24 hours' '2.5 weeks' '36 hours' ' unknown' '5 minutes'
 '215 days' '2 years' '4 months' '40. hours' '15 NA' '0 NA' ' years'
 '3.5 months' '6 years' '1 hours' '5 days' '20 hours' '4 days' '8 months'
 '8 NA' '4 weeks' '1 days' '5 years']

after transformation:  [1.46000000e+03 3.65000000e+03 3.65000000e+02 7.00000000e
+00
 1.00000000e+00 3.00000000e+01 2.10000000e+01 1.80000000e+02
 3.00000000e+00 1.09500000e+03 2.00000000e+00 9.00000000e+01
 6.00000000e+01            nan 1.50000000e+02 1.40000000e+01
 1.75000000e+01 1.50000000e+00 3.47222222e-03 2.15000000e+02
 7.30000000e+02 1.20000000e+02 1.66666667e+00 1.50000000e+01
 0.00000000e+00 1.05000000e+02 2.19000000e+03 4.16666667e-02
 5.00000000e+00 8.33333333e-01 4.00000000e+00 2.40000000e+02
 8.00000000e+00 2.80000000e+01 1.82500000e+03]
```

	FixedID	DrugPlacebo	Locus	Positive	Negative	Arousal	Familiarity	A
0	8.0	1	0	3.0	5.0	7.0	0.0	(
1	4.0	1	0	3.0	6.0	9.0	0.0	(
2	6.0	1	0	2.0	8.0	10.0	0.0	(
3	7.0	0	0	3.0	3.0	8.0	1.0	(
4	9.0	0	0	4.0	6.0	6.0	0.0	(

5 rows × 144 columns

Check Transformations

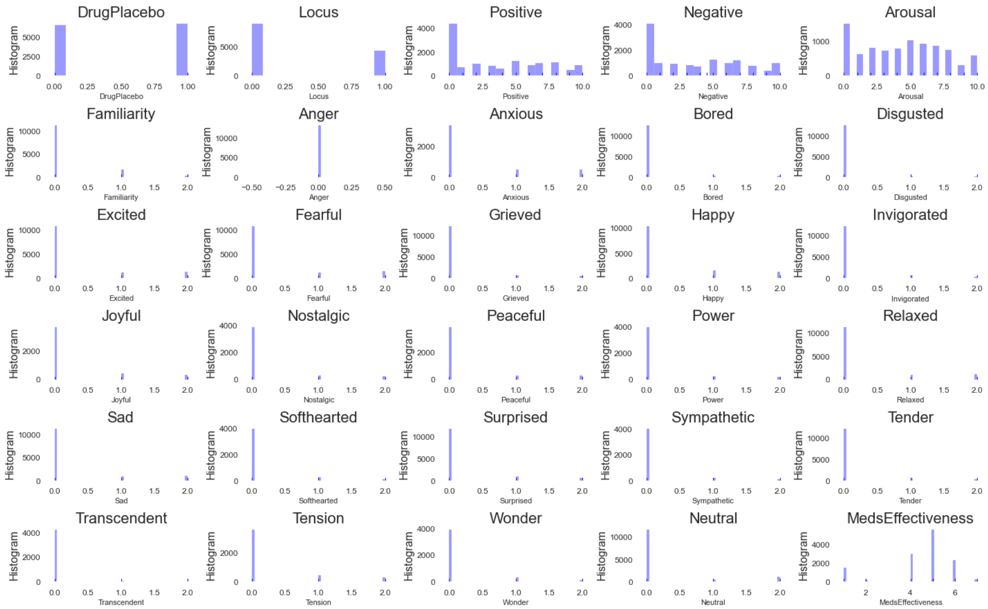
```
shape of model dataframe:  (13218, 144)
number of numeric columns:  143
number of categorical columns:  1
list of cat columns =  ['FixedID']
```

Now, 143 of the 144 features are numeric! The only one that is still categorical is FixedID. We can now proceed with the statistical analysis and exploratory data analysis.

3c. Plot

The next step is to make graphs of the distributions of the variables to see if we need to transform them in any other way.

```
There are 143 features to plot.
```

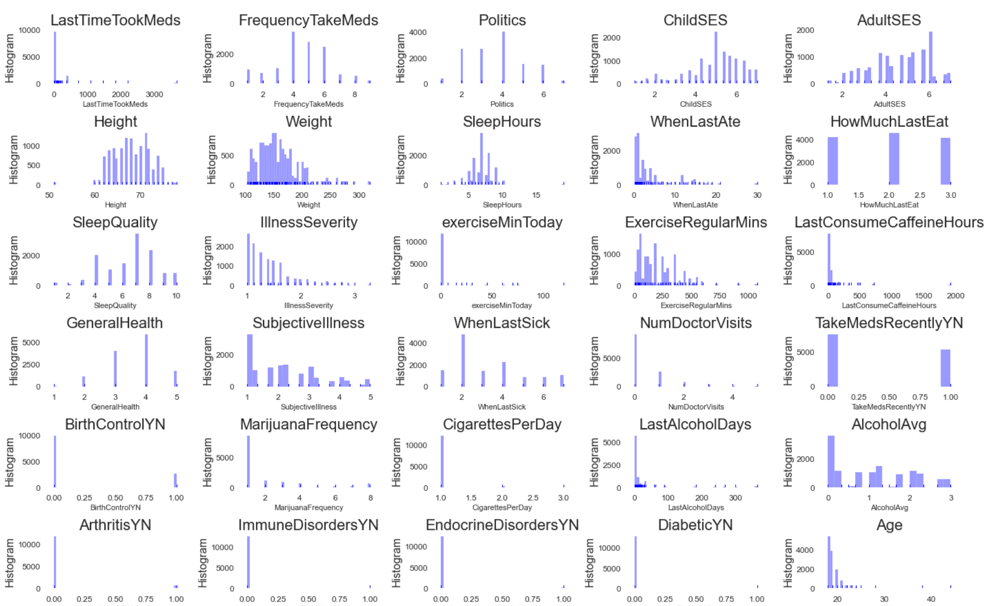


The three emotion ratings (positive, negative, arousal) are right skewed.

- We'll try transforming these later on.

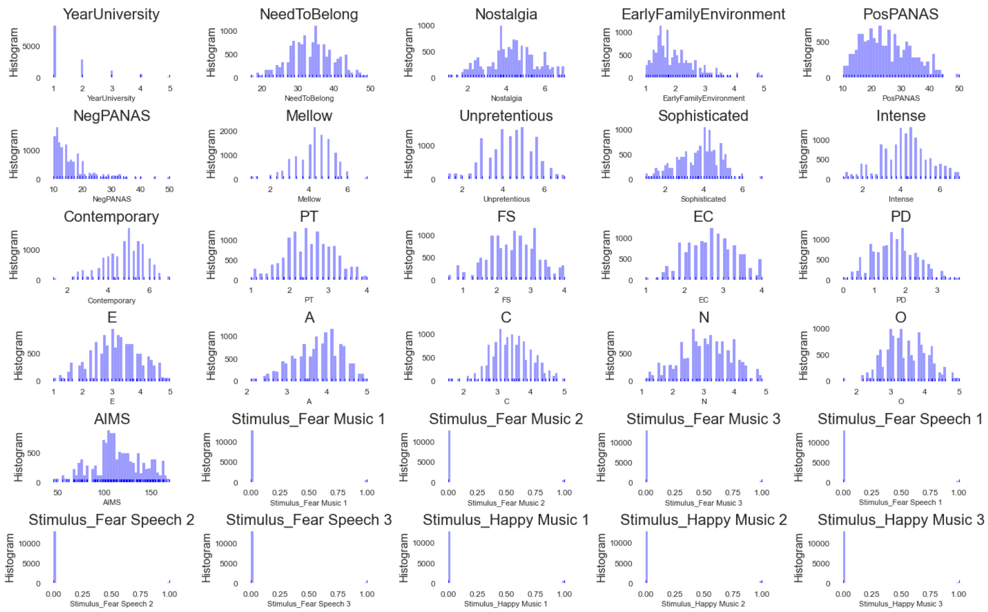
People don't seem to rate any of the specific emotions highly.

- We'll come back to this in the last part of the analysis.



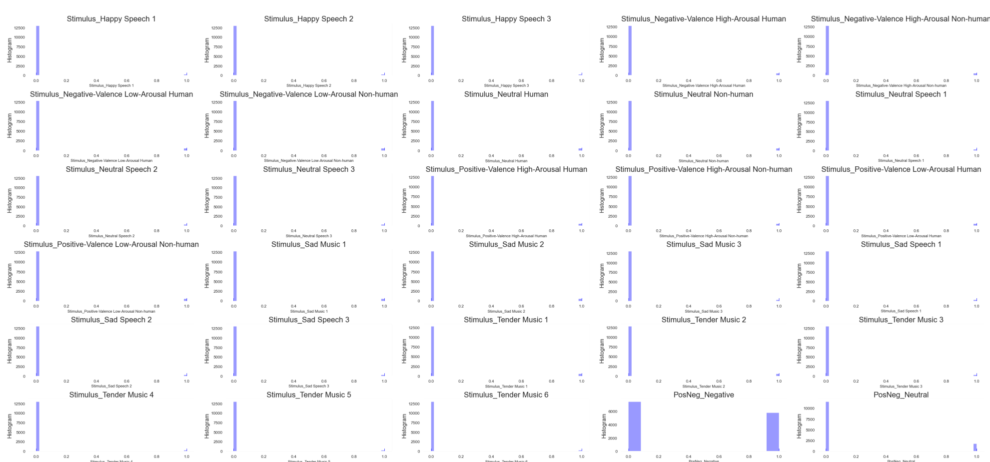
Some variables seem to have little variance, like *DiabeticYN* and *CigarettesPerDay*.

- We'll look for outliers and variability in the next step.

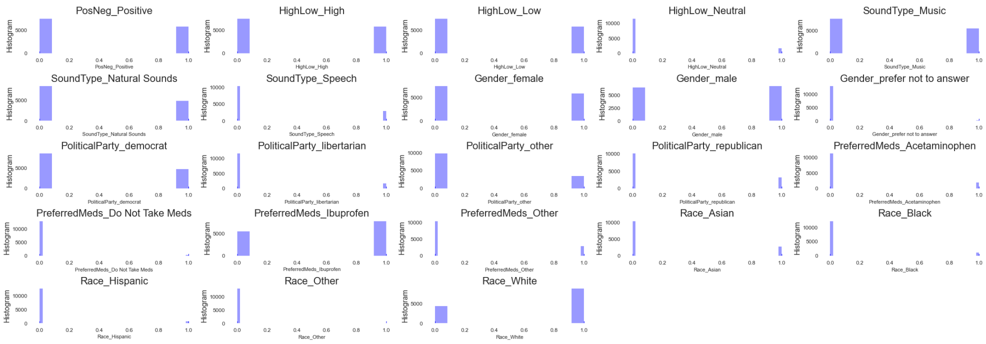


The psychographic variables, like *nostalgia*, *empathy* (PT, FS, EC, and PD), *personality* (E, A, C, N, and O), and *absorption in music* (AIMS) look reasonably normally distributed!

The Stimulus distributions are not useful, as the number of each stimulus was predetermined.



As stated above, the Stimulus distributions are not useful, as the number of each stimulus was predetermined.



Some of the one-hot encoded distributions look skewed

- We will investigate this more in the next section

3d. Feature Selection

We do not want to include features with a high level of missingness in the model, as imputing these values can lead to bias.

Similarly, features where almost every observation is the same will not inform us about how emotion ratings differ between those who took the drug and those who took the placebo.

Therefore, we will delete features with high level of missingness and very low variance

Look for variables that are missing > 10% of the data

Other than the specific emotions (which were not included in both perceived and induced), only 2 features are missing more tan 10% of the data:

- *Arousal* -- 32.8% missing.
- *LastAlcoholDays* -- 30.9% missing.

The arousal values are missing because participants were not asked to rate arousal for the induced emotion stimuli -- we will therefore keep it in the model as is.

We will delete ***LastAlcoholDays***, though, because of its missingness.

Look for variables where one response accounts for 95-100% of the variance

```
Anger
Bored
Disgusted
Transcendent
CigarettesPerDay
ImmuneDisordersYN
EndocrineDisordersYN
DiabeticYN
Stimulus_Fear Music 1
Stimulus_Fear Music 2
Stimulus_Fear Music 3
Stimulus_Fear Speech 1
Stimulus_Fear Speech 2
Stimulus_Fear Speech 3
Stimulus_Happy Music 1
Stimulus_Happy Music 2
Stimulus_Happy Music 3
Stimulus_Happy Speech 1
Stimulus_Happy Speech 2
Stimulus_Happy Speech 3
Stimulus_Negative-Valence High-Arousal Human
Stimulus_Negative-Valence High-Arousal Non-human
Stimulus_Negative-Valence Low-Arousal Human
Stimulus_Negative-Valence Low-Arousal Non-human
Stimulus_Neutral Human
Stimulus_Neutral Non-human
Stimulus_Neutral Speech 1
Stimulus_Neutral Speech 2
Stimulus_Neutral Speech 3
Stimulus_Positive-Valence High-Arousal Human
Stimulus_Positive-Valence High-Arousal Non-human
Stimulus_Positive-Valence Low-Arousal Human
Stimulus_Positive-Valence Low-Arousal Non-human
Stimulus_Sad Music 1
Stimulus_Sad Music 2
Stimulus_Sad Music 3
Stimulus_Sad Speech 1
Stimulus_Sad Speech 2
Stimulus_Sad Speech 3
Stimulus_Tender Music 1
Stimulus_Tender Music 2
Stimulus_Tender Music 3
Stimulus_Tender Music 4
Stimulus_Tender Music 5
Stimulus_Tender Music 6
Gender_prefer not to answer
PreferredMeds_Do Not Take Meds
Race_Hispanic
Race_Other
```

We want to keep all the one-hot encoded variables and specific emotions for later analysis.

However, we will delete the following columns that had only one value for 95-100% of the observations:

- CigarettesPerDay
- ImmuneDisordersYN
- EndocrineDisordersYN
- DiabeticYN

Delete

```
model shape: (13218, 139)
```

3e. Outlier Detection

For each feature, we should investigate the percent of outliers present.

We will define ***outlier*** as **an observation that is more than 3 standard deviations outside of the mean (in either direction).**

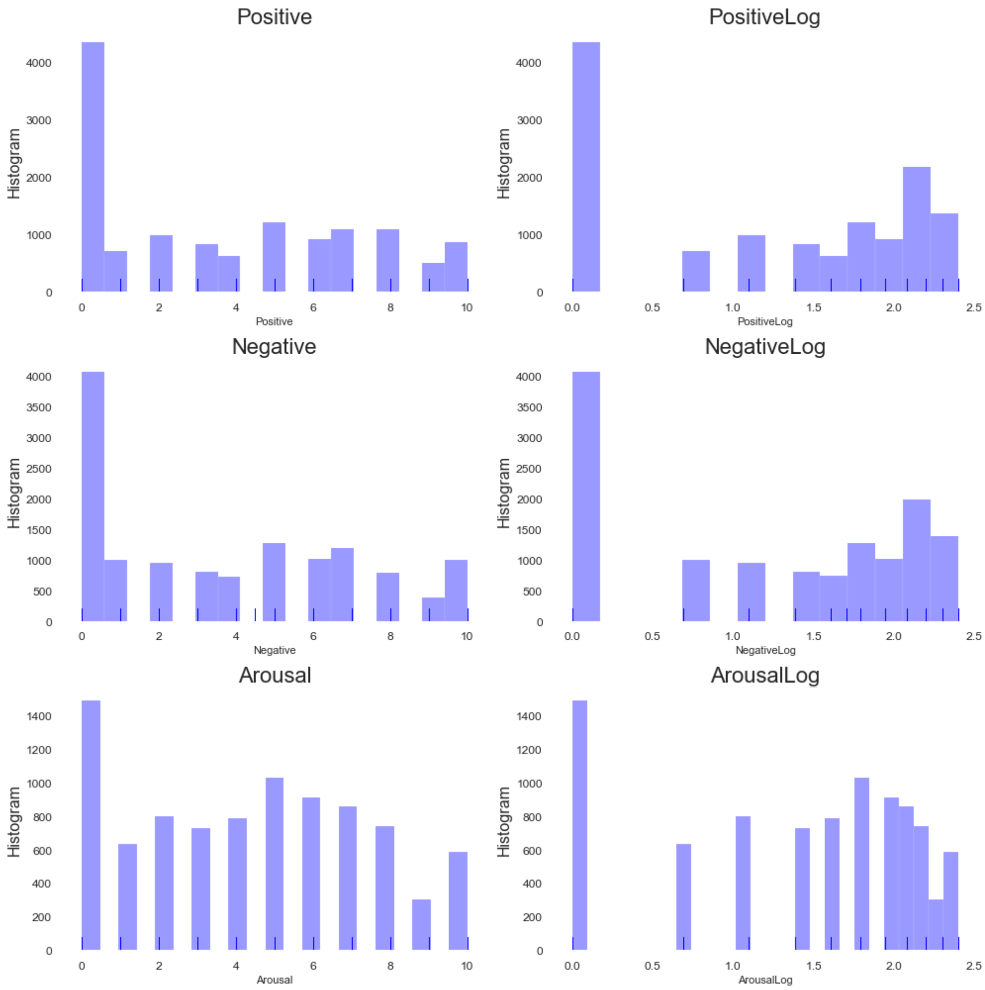
```
Sad
 7 % outliers

Neutral
 8 % outliers

ArthritisYN
 5 % outliers

Race_Black
 7 % outliers

Race_Hispanic
 5 % outliers
```

The only non 'one-hot encoded' value or 'specific emotion value' is ***ArthritisYN*** (with 5% of the cases being outliers). We will still keep it in the analysis for now.

3f. Transform Skewed Variables

Recall that the three emotion ratings (the dependent variables) were right skewed. We will try transforming them to see if that helps the distribution before more normal.

We will use a **log(1+x) transformation**.



The transformation didn't help the skew so we return to the original values to aid in interpretability.

3g. Impute missing values

There are some new missing values (for example, when people completed the stimuli blocks but not the questionnaires, etc.) We will, once again, impute these missing values with the column median.

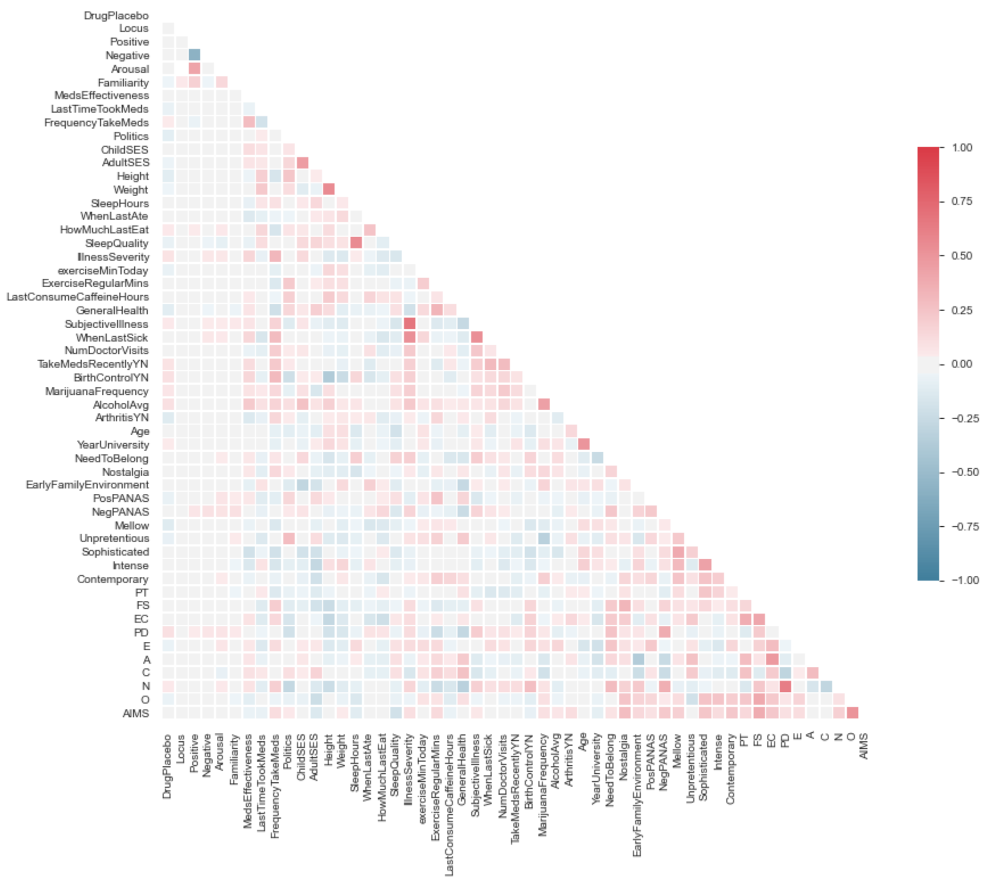
As discussed above, we will not impute the ***arousal*** values or the ***specific emotion values*****(like Wonder and Sympathetic)**, as these values are missing because of the *a priori* study design.

3h. Correlations

We will look for high correlations that may lead to multicollinearity problems.

Plot

```
<matplotlib.axes._subplots.AxesSubplot at 0x155066910>
```



There are a few high correlations, but the variables seem to be mostly reasonably uncorrelated.

We will look for more details, next.

List high correlations

```
maximum absolute value correlation:  0.68
```

The highest correlation (absolute value) is 0.68, so we don't have to worry too much about multicollinearity.

Regularization (L1) will take care of that if needed.

3i. View final dataframe and reset index

Summarize

```
final model shape: (13218, 139)
number of participants:  244
```

The final dataframe consists of 139 features for 244 participants.

Profile

4. Regression

Now that we have completed feature engineering, we can use linear regression to investigate which features contribute to emotion ratings!

Of course, we are the most interested in seeing whether those who took the drug give reduced (blunted) ratings, compared to those who took the placebo.

We will run a few models:

1. Predicting *positive* emotion ratings
2. Predicting *negative* emotion ratings
3. Predicting *arousal* ratings
4. Predicting *all* emotion ratings

For each of these models, we will run a few multiple linear regressions:

1. Simple OLS model
2. OLS model with lasso regularization to penalize complexity
3. Hierarchical/mixed model (with participant ID as the random effect).

   Using ID as a random effect allows us to account for differences among participants and hence perform a within-subjects analysis.

4a. Predicting Positive Emotion Ratings

Data prep

We do not want every single column in these regression models.

The variables we do not need are the following:

- Negative emotion atings and arousal ratings -- we are only focused on positive emotion ratings here.
- Specific stimulus names -- the features of each stimulus are encoded as Locus (perceived/induced), SoundType (music, speech, natural sounds), HighLow (arousal) and PosNeg (valence).
- Specific emotions -- these will be used in a separate analysis later on.

We also want to rename the Positive column to Ratings. This will allow us to combine the positive, negative, and arousal dataframes later on.

```
entire dataframe shape:  (13218, 139)
positive dataframe shape:  (13218, 77)
```

X and Y

- Y variable -- *Ratings*
- X variables -- *All other columns except FixedID (the random effect measured later)*

OLS regression

Note that interactions are already taken care of through the one-hot encoding

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                Ratings   R-squared:                       0.547
Model:                            OLS   Adj. R-squared:                  0.544
Method:                 Least Squares   F-statistic:                     226.5
Date:                Thu, 26 Mar 2020   Prob (F-statistic):               0.00
Time:                        07:56:08   Log-Likelihood:                -29848.
No. Observations:               13218   AIC:                         5.984e+04
Df Residuals:                   13147   BIC:                         6.037e+04
Df Model:                          70
Covariance Type:            nonrobust
==============================================================================
====================
                                 coef    std err          t      P>|t|
[0.025      0.975]
------------------------------------------------------------------------------
--------------------
const                          0.6756      0.383      1.763      0.078
-0.075       1.427
DrugPlacebo                   -0.2338      0.045     -5.230      0.000
-0.321      -0.146
Locus                         -0.1575      0.047     -3.352      0.001
-0.250      -0.065
Familiarity                    0.4672      0.053      8.797      0.000
0.363       0.571
MedsEffectiveness             -0.0260      0.018     -1.420      0.156
-0.062       0.010
LastTimeTookMeds           -2.761e-05   4.86e-05     -0.569      0.570
-0.000    6.76e-05
FrequencyTakeMeds             -0.0753      0.017     -4.381      0.000
-0.109      -0.042
Politics                       0.0839      0.025      3.406      0.001
0.036       0.132
ChildSES                       0.0483      0.022      2.180      0.029
0.005       0.092
AdultSES                      -0.0722      0.022     -3.265      0.001
-0.116      -0.029
Height                         0.0106      0.009      1.240      0.215
-0.006       0.027
Weight                         0.0036      0.001      4.226      0.000
0.002       0.005
SleepHours                    -0.0782      0.017     -4.467      0.000
-0.112      -0.044
WhenLastAte                    0.0009      0.005      0.172      0.864
-0.010       0.012
HowMuchLastEat                 0.1088      0.031      3.466      0.001
0.047       0.170
SleepQuality                  -0.0203      0.016     -1.294      0.196
-0.051       0.010
IllnessSeverity                0.1446      0.077      1.869      0.062
-0.007       0.296
exerciseMinToday              -0.0089      0.002     -5.376      0.000
-0.012      -0.006
ExerciseRegularMins           -0.0003      0.000     -1.847      0.065
-0.001    1.82e-05
LastConsumeCaffeineHours      -0.0003      0.000     -2.566      0.010
```

```
-0.001      -7e-05
GeneralHealth                    0.0377      0.036      1.051      0.293
-0.033       0.108
SubjectiveIllness               -0.0545      0.030     -1.797      0.072
-0.114       0.005
WhenLastSick                     0.0635      0.018      3.596      0.000
0.029       0.098
NumDoctorVisits                  0.0922      0.032      2.854      0.004
0.029       0.156
TakeMedsRecentlyYN              -0.1649      0.051     -3.207      0.001
-0.266      -0.064
BirthControlYN                   0.0696      0.071      0.984      0.325
-0.069       0.208
MarijuanaFrequency              -0.0012      0.015     -0.076      0.940
-0.031       0.029
AlcoholAvg                       0.0404      0.031      1.320      0.187
-0.020       0.100
ArthritisYN                      0.1525      0.112      1.361      0.174
-0.067       0.372
Age                              0.0304      0.011      2.744      0.006
0.009       0.052
YearUniversity                  -0.0099      0.031     -0.314      0.754
-0.071       0.052
NeedToBelong                     0.0062      0.005      1.328      0.184
-0.003       0.015
Nostalgia                       -0.0659      0.020     -3.299      0.001
-0.105      -0.027
EarlyFamilyEnvironment           0.0150      0.038      0.393      0.695
-0.060       0.090
PosPANAS                         0.0083      0.003      2.515      0.012
0.002       0.015
NegPANAS                         0.0219      0.004      5.086      0.000
0.013       0.030
Mellow                          -0.0393      0.027     -1.434      0.152
-0.093       0.014
Unpretentious                   -0.0378      0.028     -1.373      0.170
-0.092       0.016
Sophisticated                    0.1144      0.027      4.189      0.000
0.061       0.168
Intense                         -0.1285      0.024     -5.416      0.000
-0.175      -0.082
Contemporary                     0.0308      0.028      1.113      0.266
-0.023       0.085
PT                               0.0161      0.043      0.375      0.708
-0.068       0.101
FS                              -0.1584      0.039     -4.078      0.000
-0.235      -0.082
EC                              -0.0841      0.052     -1.619      0.105
-0.186       0.018
PD                               0.2485      0.042      5.904      0.000
0.166       0.331
E                                0.1282      0.033      3.877      0.000
0.063       0.193
A                                0.2752      0.051      5.378      0.000
0.175       0.376
C                               -0.0386      0.045     -0.853      0.394
-0.127       0.050
```

```
N                                  -0.1498      0.040     -3.740      0.000
-0.228      -0.071
O                                   0.0853      0.054      1.589      0.112
-0.020       0.191
AIMS                               -0.0008      0.001     -0.697      0.486
-0.003       0.001
PosNeg_Negative                    -1.8991      0.130    -14.631      0.000
-2.154      -1.645
PosNeg_Neutral                     -0.3592      0.130     -2.766      0.006
-0.614      -0.105
PosNeg_Positive                     2.9338      0.130     22.580      0.000
2.679       3.189
HighLow_High                        0.8216      0.130      6.331      0.000
0.567       1.076
HighLow_Low                         0.2131      0.130      1.640      0.101
-0.042       0.468
HighLow_Neutral                    -0.3592      0.130     -2.766      0.006
-0.614      -0.105
SoundType_Music                     0.7015      0.132      5.325      0.000
0.443       0.960
SoundType_Natural Sounds            0.0147      0.132      0.112      0.911
-0.244       0.273
SoundType_Speech                   -0.0406      0.133     -0.305      0.760
-0.301       0.220
Gender_female                       1.0504      0.354      2.970      0.003
0.357       1.744
Gender_male                         0.8519      0.351      2.427      0.015
0.164       1.540
Gender_prefer not to answer         0.5181      0.376      1.376      0.169
-0.220       1.256
PoliticalParty_democrat            -0.7095      0.175     -4.063      0.000
-1.052      -0.367
PoliticalParty_libertarian         -0.3243      0.184     -1.763      0.078
-0.685       0.036
PoliticalParty_other               -1.0292      0.179     -5.745      0.000
-1.380      -0.678
PoliticalParty_republican          -0.8987      0.186     -4.835      0.000
-1.263      -0.534
PreferredMeds_Acetaminophen        -0.5405      0.175     -3.090      0.002
-0.883      -0.198
PreferredMeds_Do Not Take Meds     -0.7738      0.222     -3.492      0.000
-1.208      -0.339
PreferredMeds_Ibuprofen            -0.6816      0.179     -3.800      0.000
-1.033      -0.330
PreferredMeds_Other                -0.9656      0.182     -5.316      0.000
-1.322      -0.610
Race_Asian                          0.3401      0.124      2.733      0.006
0.096       0.584
Race_Black                          0.0082      0.125      0.066      0.948
-0.237       0.253
Race_Hispanic                      -0.2477      0.121     -2.053      0.040
-0.484      -0.011
Race_Other                         -0.7092      0.276     -2.565      0.010
-1.251      -0.167
Race_White                         -0.1107      0.123     -0.902      0.367
-0.351       0.130
==============================================================================
```

```
Omnibus:                      126.402   Durbin-Watson:                   1.815
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              186.558
Skew:                           0.102   Prob(JB):                     3.09e-41
Kurtosis:                       3.545   Cond. No.                     1.68e+16
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correct
ly specified.
[2] The smallest eigenvalue is 1.39e-23. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
```

Conclusions:

- **The measured variables explain 54% of the variance in positive emotion ratings (adjusted R^2 = 0.54)**
- **The ingestion of acetaminophen blunted ratings of positive emotions compared to the ingestion of a placebo.**

Name	Coefficient	p	Interpretation
DrugPlacebo	***-0.2338***	***0***	*Those who **took acetaminophen** rated positive emotions as **less intense** than those who took the placebo. **Namely, acetaminophen blunted the emotional responses.***
PosNeg_Positive	2.9338	0	**Positively-valenced stimuli** resulted in comparatively **high** positive emotion ratings.
PosNeg_Negative	-1.8991	0	**Negatively-valenced stimuli** resulted in comparatively **low** positive emotion ratings.
HighLow_High	0.8216	0	Stimuli that exhibit **high arousal** resulted in comparatively **more intense** positive emotion ratings.
SoundType_Music	0.7015	0	**Musical stimuli** resulted in comparatively **higher ratings of** positive emotion.
Familiarity	0.4672	0	Those who are more **familiar** with the (musical) stimuli rated positive emotions as **more intense.**
A	0.2752	0	Those who are more **agreeable** rated emotions as relatively **more positive** than those who score less high on this personality dimension.
PD	0.2485	0	Those who score higher on the **Personal Distress component of empathy** rated emotions as relatively **more positive** than those who score less high on this trait.
FS	-0.1584	0	Those who score higher on the **Fantasy component of empathy** rated emotions as relatively **less positive** than those who score less high on this trait.
Locus	-0.1575	0.001	Compared to perceived emotion ratings, **induced emotion ratings** were **more positive**.
N	-0.1498	0	Those who are more **neurotic** rated emotions as relatively **less positive** than those who score less high on this personality dimension.

Name	Coefficient	p	Interpretation
Intense	-0.1285	0	Those who **prefer "Intense" music** (Rock, Punk, Alternative, Heavy Metal) rated emotions as relatively **less positive** (broadly, in response to all stimuli) than those who do not like this genre of music.
E	0.1282	0	Those who are more **extraverted** rated emotions as relatively **more positive** than those who score less high on this personality dimension.
Sophisticated	0.1144	0	Those who **prefer "Sophisticated" music** (Blues, Jazz, Bluegrass, Folk, Classical, Gospel, Opera) rated emotions as relatively **more positive** (broadly, in response to all stimuli) than those who do not like this genre of music.
Nostalgia	-0.0659	0.001	Those who scored high on **nostalgia** rated relatively **less intense** positive ratings (although this effect is small).

OLS regression with lasso regularization

There's no way to look at standard errors, p-values, R^2, etc. because the theory for these values is still being developed among statisticians.

Conclusions:

- **This didn't change very much from the original mode.**
- **This chart shows the "non-zero" variables from the original OLS --> these are the only (significant) features kept in the regularized model**
- **Note that I used an arbitrary cutoff of 0.1 coefficient value to count as significant.**

Features Remaining In the Model	OLS Coefficient	OLS **with lasso regularization** Coefficient
DrugPlacebo	***-0.2338***	***-0.213441***
PosNeg_Positive	2.9338	2.679953
PosNeg_Negative	-1.8991	-2.152853
PosNeg_Neutral	-0.3592	-1.489466
Race_Other	-0.7092	-0.566435
Familiarity	0.4672	0.4724
PoliticalParty_other	-1.0292	-0.439776
SoundType_Music	0.7015	0.410356
Race_Asian	0.3401	0.401408
PoliticalParty_republican	-0.8987	-0.343314
HighLow_High	0.8216	0.304188

Features Remaining In the Model	OLS Coefficient	OLS **with lasso regularization** Coefficient
PreferredMeds_Other	-0.9656	-0.270893
Gender_male	0.8519	-0.262979
PD	0.2485	0.238791
TakeMedsRecentlyYN	-0.1649	-0.189407
Gender_female	1.0504	-0.179473
PreferredMeds_Acetaminophen	-0.5405	0.159463
Locus	-0.1575	-0.154902
PreferredMeds_Do Not Take Meds	-0.7738	-0.139533
N	-0.1498	-0.137826
FS	-0.1584	-0.134634
PoliticalParty_democrat	-0.7095	-0.124515
Intense	-0.1285	-0.11814

Mixed Model

FixedID (the participant ID numbers) will be used as the grouping variable (the random effect)

```
                        Mixed Linear Model Regression Results
================================================================================
=========
Model:                        MixedLM          Dependent Variable:           Ra
tings
No. Observations:             13218            Method:                       RE
ML
No. Groups:                   244              Scale:                        4.
9020
Min. group size:              18               Likelihood:                   -2
9542.5375
Max. group size:              55               Converged:                    No
Mean group size:              54.2
--------------------------------------------------------------------------------
---------
                                 Coef.    Std.Err.    z    P>|z|    [0.025
0.975]
--------------------------------------------------------------------------------
---------
const                             24.794 743061.093  0.000 1.000 -1456348.186 14
56397.774
DrugPlacebo                       -0.233      0.123 -1.895 0.058       -0.475
0.008
Locus                             -0.150      0.045 -3.340 0.001       -0.239
-0.062
Familiarity                        0.390      0.054  7.284 0.000        0.285
0.495
MedsEffectiveness                 -0.026      0.050 -0.507 0.612       -0.125
0.073
LastTimeTookMeds                  -0.000      0.000 -0.212 0.832       -0.000
0.000
FrequencyTakeMeds                 -0.075      0.048 -1.571 0.116       -0.168
0.018
Politics                           0.084      0.068  1.234 0.217       -0.049
0.217
ChildSES                           0.047      0.061  0.770 0.441       -0.073
0.167
AdultSES                          -0.070      0.061 -1.155 0.248       -0.190
0.049
Height                             0.011      0.024  0.471 0.637       -0.035
0.058
Weight                             0.004      0.002  1.558 0.119       -0.001
0.008
SleepHours                        -0.075      0.048 -1.552 0.121       -0.169
0.020
WhenLastAte                        0.001      0.015  0.045 0.964       -0.029
0.030
HowMuchLastEat                     0.110      0.087  1.266 0.205       -0.060
0.280
SleepQuality                      -0.024      0.043 -0.557 0.577       -0.108
0.060
IllnessSeverity                    0.147      0.214  0.689 0.491       -0.271
0.566
exerciseMinToday                  -0.009      0.005 -1.946 0.052       -0.018
0.000
ExerciseRegularMins               -0.000      0.000 -0.659 0.510       -0.001
```

```
0.001
LastConsumeCaffeineHours           -0.000      0.000 -0.934 0.351        -0.001
0.000
GeneralHealth                       0.035      0.099  0.354 0.724        -0.159
0.229
SubjectiveIllness                  -0.052      0.084 -0.626 0.531        -0.217
0.112
WhenLastSick                        0.062      0.048  1.274 0.203        -0.033
0.157
NumDoctorVisits                     0.091      0.089  1.018 0.308        -0.084
0.266
TakeMedsRecentlyYN                 -0.166      0.142 -1.166 0.243        -0.444
0.113
BirthControlYN                      0.065      0.195  0.332 0.740        -0.318
0.448
MarijuanaFrequency                 -0.000      0.042 -0.003 0.997        -0.083
0.082
AlcoholAvg                          0.041      0.084  0.485 0.628        -0.124
0.206
ArthritisYN                         0.170      0.302  0.562 0.574        -0.423
0.763
Age                                 0.029      0.031  0.956 0.339        -0.031
0.089
YearUniversity                     -0.008      0.087 -0.088 0.930        -0.177
0.162
NeedToBelong                        0.006      0.013  0.477 0.633        -0.019
0.031
Nostalgia                          -0.067      0.055 -1.210 0.226        -0.175
0.041
EarlyFamilyEnvironment              0.013      0.105  0.124 0.902        -0.193
0.219
PosPANAS                            0.009      0.009  0.946 0.344        -0.009
0.026
NegPANAS                            0.022      0.012  1.853 0.064        -0.001
0.045
Mellow                             -0.038      0.076 -0.497 0.619        -0.186
0.110
Unpretentious                      -0.033      0.076 -0.432 0.666        -0.181
0.116
Sophisticated                       0.115      0.075  1.522 0.128        -0.033
0.262
Intense                            -0.127      0.066 -1.945 0.052        -0.256
0.001
Contemporary                        0.028      0.076  0.366 0.714        -0.122
0.177
PT                                  0.016      0.118  0.135 0.893        -0.215
0.247
FS                                 -0.155      0.107 -1.450 0.147        -0.365
0.055
EC                                 -0.086      0.143 -0.596 0.551        -0.367
0.196
PD                                  0.248      0.115  2.154 0.031         0.022
0.474
E                                   0.128      0.091  1.407 0.160        -0.050
0.305
A                                   0.272      0.141  1.924 0.054        -0.005
0.549
```

```
C                                 -0.043      0.124 -0.342 0.733       -0.287
0.201
N                                 -0.150      0.111 -1.353 0.176       -0.368
0.067
O                                  0.088      0.148  0.591 0.554       -0.203
0.378
AIMS                              -0.001      0.003 -0.271 0.786       -0.007
0.006
PosNeg_Negative                  -13.396 451632.948 -0.000 1.000  -885197.708  8
85170.916
PosNeg_Neutral                  3176.709
PosNeg_Positive                   -8.562 451684.161 -0.000 1.000  -885293.249  8
85276.126
HighLow_High                     -11.506
HighLow_Low                      -12.116
HighLow_Neutral                -3201.251
SoundType_Music                    0.407 120061.108  0.000 1.000  -235315.040  2
35315.855
SoundType_Natural Sounds          -0.312 113747.479 -0.000 1.000  -222941.273  2
22940.650
SoundType_Speech                  -0.365 119121.505 -0.000 1.000  -233474.225  2
33473.495
Gender_female                      1.059      0.980  1.080 0.280       -0.862
2.979
Gender_male                        0.852      0.972  0.877 0.381       -1.053
2.757
Gender_prefer not to answer        0.530      1.042  0.509 0.611       -1.513
2.573
PoliticalParty_democrat          -10.223
PoliticalParty_libertarian        -9.839
PoliticalParty_other             -10.549
PoliticalParty_republican        -10.412
PreferredMeds_Acetaminophen        8.967
PreferredMeds_Do Not Take Meds     8.736
PreferredMeds_Ibuprofen            8.829
PreferredMeds_Other                8.546
Race_Asian                         0.342      0.345  0.992 0.321       -0.334
1.019
Race_Black                        -0.004      0.346 -0.011 0.991       -0.681
0.674
Race_Hispanic                     -0.240      0.334 -0.719 0.472       -0.894
0.414
Race_Other                        -0.700      0.766 -0.914 0.361       -2.201
0.801
Race_White                        -0.112      0.340 -0.329 0.742       -0.777
0.554
Group Var                          0.664
================================================================================
=========
```

Conclusions:

- **The model did not converge**

- **Basically no features are significant in the mixed model (because they are so redundant with the questionnaire data)**

4b. Predicting Negative Emotion Ratings

Data prep

```
entire dataframe shape:  (13218, 139)
negative dataframe shape:  (13218, 77)
```

X and Y

OLS regression

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                Ratings   R-squared:                       0.488
Model:                            OLS   Adj. R-squared:                  0.485
Method:                 Least Squares   F-statistic:                     179.0
Date:                Thu, 26 Mar 2020   Prob (F-statistic):               0.00
Time:                        07:56:34   Log-Likelihood:                -30464.
No. Observations:               13218   AIC:                         6.107e+04
Df Residuals:                   13147   BIC:                         6.160e+04
Df Model:                          70
Covariance Type:            nonrobust
==============================================================================
====================
                                 coef    std err          t      P>|t|
[0.025      0.975]
------------------------------------------------------------------------------
--------------------
const                          1.9981      0.401      4.978      0.000
1.211       2.785
DrugPlacebo                   -0.1973      0.047     -4.212      0.000
-0.289      -0.105
Locus                         -0.2381      0.049     -4.837      0.000
-0.335      -0.142
Familiarity                    0.0048      0.056      0.086      0.932
-0.104       0.114
MedsEffectiveness             -0.0431      0.019     -2.244      0.025
-0.081      -0.005
LastTimeTookMeds            1.593e-05   5.09e-05      0.313      0.754    -
8.38e-05       0.000
FrequencyTakeMeds              0.0042      0.018      0.232      0.817
-0.031       0.039
Politics                       0.0773      0.026      2.997      0.003
0.027       0.128
ChildSES                      -0.0681      0.023     -2.938      0.003
-0.114      -0.023
AdultSES                       0.0104      0.023      0.451      0.652
-0.035       0.056
Height                         0.0043      0.009      0.476      0.634
-0.013       0.022
Weight                         0.0014      0.001      1.583      0.113
-0.000       0.003
SleepHours                     0.0341      0.018      1.862      0.063
-0.002       0.070
WhenLastAte                    0.0040      0.006      0.700      0.484
-0.007       0.015
HowMuchLastEat                 0.1865      0.033      5.669      0.000
0.122       0.251
SleepQuality                  -0.1064      0.016     -6.466      0.000
-0.139      -0.074
IllnessSeverity                0.1792      0.081      2.211      0.027
0.020       0.338
exerciseMinToday               0.0023      0.002      1.316      0.188
-0.001       0.006
ExerciseRegularMins           -0.0004      0.000     -2.395      0.017
-0.001   -7.36e-05
LastConsumeCaffeineHours      -0.0005      0.000     -3.963      0.000
```

```
-0.001      -0.000
GeneralHealth                     -0.0959      0.038     -2.549      0.011
-0.170      -0.022
SubjectiveIllness                 -0.0597      0.032     -1.881      0.060
-0.122       0.003
WhenLastSick                       0.0704      0.018      3.808      0.000
0.034        0.107
NumDoctorVisits                    0.0450      0.034      1.331      0.183
-0.021       0.111
TakeMedsRecentlyYN                -0.1845      0.054     -3.425      0.001
-0.290      -0.079
BirthControlYN                    -0.0029      0.074     -0.039      0.969
-0.148       0.142
MarijuanaFrequency                 0.0279      0.016      1.750      0.080
-0.003       0.059
AlcoholAvg                         0.1604      0.032      5.004      0.000
0.098        0.223
ArthritisYN                        0.3687      0.117      3.141      0.002
0.139        0.599
Age                               -0.0354      0.012     -3.054      0.002
-0.058      -0.013
YearUniversity                    -0.0283      0.033     -0.860      0.390
-0.093       0.036
NeedToBelong                      -0.0006      0.005     -0.132      0.895
-0.010       0.009
Nostalgia                         -0.0428      0.021     -2.044      0.041
-0.084      -0.002
EarlyFamilyEnvironment            -0.0409      0.040     -1.024      0.306
-0.119       0.037
PosPANAS                           0.0084      0.003      2.434      0.015
0.002        0.015
NegPANAS                           0.0342      0.005      7.582      0.000
0.025        0.043
Mellow                             0.0769      0.029      2.680      0.007
0.021        0.133
Unpretentious                     -0.0967      0.029     -3.353      0.001
-0.153      -0.040
Sophisticated                      0.0369      0.029      1.290      0.197
-0.019       0.093
Intense                           -0.1078      0.025     -4.336      0.000
-0.156      -0.059
Contemporary                       0.0440      0.029      1.520      0.129
-0.013       0.101
PT                                -0.0047      0.045     -0.104      0.917
-0.093       0.084
FS                                -0.0467      0.041     -1.148      0.251
-0.127       0.033
EC                                 0.2329      0.054      4.279      0.000
0.126        0.340
PD                                 0.2260      0.044      5.125      0.000
0.140        0.312
E                                  0.0246      0.035      0.709      0.478
-0.043       0.092
A                                  0.0073      0.054      0.136      0.892
-0.098       0.112
C                                  0.0350      0.047      0.738      0.460
-0.058       0.128
```

```
N                                    -0.0835      0.042     -1.990      0.047
-0.166      -0.001
O                                    -0.0439      0.056     -0.781      0.435
-0.154       0.066
AIMS                                 -0.0026      0.001     -2.088      0.037
-0.005      -0.000
PosNeg_Negative                       3.2290      0.136     23.745      0.000
2.962       3.496
PosNeg_Neutral                        0.3566      0.136      2.622      0.009
0.090       0.623
PosNeg_Positive                      -1.5875      0.136    -11.662      0.000
-1.854      -1.321
HighLow_High                          0.8920      0.136      6.561      0.000
0.626       1.159
HighLow_Low                           0.7495      0.136      5.505      0.000
0.483       1.016
HighLow_Neutral                       0.3566      0.136      2.622      0.009
0.090       0.623
SoundType_Music                       0.3946      0.138      2.859      0.004
0.124       0.665
SoundType_Natural Sounds              0.9590      0.138      6.948      0.000
0.688       1.230
SoundType_Speech                      0.6445      0.139      4.622      0.000
0.371       0.918
Gender_female                         2.0843      0.371      5.625      0.000
1.358       2.811
Gender_male                           1.7295      0.368      4.704      0.000
1.009       2.450
Gender_prefer not to answer           1.3418      0.394      3.403      0.001
0.569       2.115
PoliticalParty_democrat              -1.2575      0.183     -6.874      0.000
-1.616      -0.899
PoliticalParty_libertarian           -0.8885      0.193     -4.612      0.000
-1.266      -0.511
PoliticalParty_other                 -1.4953      0.188     -7.967      0.000
-1.863      -1.127
PoliticalParty_republican            -1.2431      0.195     -6.384      0.000
-1.625      -0.861
PreferredMeds_Acetaminophen          -0.8895      0.183     -4.853      0.000
-1.249      -0.530
PreferredMeds_Do Not Take Meds       -1.7970      0.232     -7.740      0.000
-2.252      -1.342
PreferredMeds_Ibuprofen              -0.9961      0.188     -5.301      0.000
-1.364      -0.628
PreferredMeds_Other                  -1.2019      0.190     -6.316      0.000
-1.575      -0.829
Race_Asian                            0.1082      0.130      0.830      0.407
-0.147       0.364
Race_Black                           -0.1553      0.131     -1.186      0.236
-0.412       0.101
Race_Hispanic                         0.0401      0.126      0.317      0.751
-0.208       0.288
Race_Other                           -0.8999      0.290     -3.107      0.002
-1.468      -0.332
Race_White                           -0.2653      0.129     -2.062      0.039
-0.517      -0.013
==============================================================================
```

```
Omnibus:                       29.193   Durbin-Watson:                   1.818
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               34.597
Skew:                           0.047   Prob(JB):                     3.07e-08
Kurtosis:                       3.232   Cond. No.                     1.68e+16
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correct
ly specified.
[2] The smallest eigenvalue is 1.39e-23. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
```

Conclusions:

- **The measured variables explain 49% of the variance in negative emotion ratings (adjusted R^2 = 0.49)**
- **The ingestion of acetaminophen blunted ratings of negative emotions compared to the ingestion of a placebo.**

Name	Coefficient	p	Interpretation
DrugPlacebo	***-0.1973***	***0***	*Those who **took acetaminophen** rated negative emotions as **less intense** than those who took the placebo. **Namely, acetaminophen blunted the emotional responses.***
PosNeg_Negative	3.229	0	**Negatively-valenced stimuli** resulted in comparatively **high** negative emotion ratings.
PosNeg_Positive	-1.5875	0	**Positively-valenced stimuli** resulted in comparatively **low** negative emotion ratings.
SoundType_Natural Sounds	0.959	0	**Natural Sounds** resulted in comparatively **higher ratings** of negative emotion *(and its effect is comparatively large next to speech and especially compared to music)*
HighLow_High	0.892	0	Stimuli that exhibit **high arousal** resulted in comparatively **more intense** negative emotion ratings.
HighLow_Low	0.7495	0	Stimuli that exhibit **low arousal** resulted in comparatively **higher** negative emotion ratings *(but the effect is smaller than the high arousal music)*.
SoundType_Speech	0.6445	0	**Speech** resulted in comparatively **higher ratings** of negative emotion *(and its effect is comparatively large with musical stimuli)*
SoundType_Music	0.3946	0.004	**Musical stimuli** resulted in comparatively **higher ratings** of negative emotion *(but the effect is small compared to speech and natural sounds)*
Locus	-0.2381	0	Compared to perceived emotion ratings, **induced emotion ratings** were **less negative**.
EC	0.2329	0	Those who score higher on the **Empathic Concern component of empathy** rated emotions as relatively **more negative** than those who score less high on this trait.

Name	Coefficient	p	Interpretation
PD	0.226	0	Those who score higher on the **Personal Distress component of empathy** rated emotions as relatively **more negative** than those who score less high on this trait.
Intense	-0.1078	0	Those who **prefer "Intense" music** (Rock, Punk, Alternative, Heavy Metal) rated emotions as relatively **less negative** (broadly, in response to all stimuli) than those who do not like this genre of music.
Unpretentious	-0.0967	0.001	Those who **prefer "Unpretentious" music** (Pop, Country, Religious) rated emotions as relatively **less negative** (broadly, in response to all stimuli) than those who do not like this genre of music (although this effect is small).
N	-0.0835	0.047	Those who are more **neurotic** rated emotions as relatively **more negative** than those who score less high on this personality dimension (although this effect is small).
Mellow	0.0769	0.007	Those who **prefer "Mellow" music** (Dance/Electronica, New Age, World) rated emotions as relatively **more negative** (broadly, in response to all stimuli) than those who do not like this genre of music (although this effect is small).
Nostalgia	-0.0428	0.041	Those who scored high on **nostalgia** rated relatively **less intense** negative ratings (although this effect is small).

OLS regression with lasso regularization

Conclusions:

- **This didn't change very much from the original mode.**
- **This chart shows the "non-zero" variables from the original OLS --> these are the only (significant) features kept in the regularized model**
- **Note that I used an arbitrary cutoff of 0.1 coefficient value to count as significant.**

Features Remaining In the Model	OLS Coefficient	OLS **with lasso regularization** Coefficient
DrugPlacebo	***-0.2338***	***-0.170217***
PosNeg_Negative	-1.8991	2.784016
PosNeg_Positive	2.9338	-2.03266
PreferredMeds_Do Not Take Meds	-0.7738	-0.752959
PoliticalParty_other	-1.0292	-0.588811
Race_Other	-0.7092	-0.56928

Features Remaining In the Model	OLS Coefficient	OLS **with lasso regularization** Coefficient
PosNeg_Neutral	-0.3592	-0.552684
PoliticalParty_democrat	-0.7095	-0.357943
PoliticalParty_republican	-0.8987	-0.340561
SoundType_Music	0.7015	-0.262282
Race_Asian	0.3401	0.257602
Locus	-0.1575	-0.236797
TakeMedsRecentlyYN	-0.1649	-0.192641
Gender_female	1.0504	0.183484
Gender_male	0.8519	-0.183096
Race_Hispanic	-0.2477	0.172875
PreferredMeds_Other	-0.9656	-0.16225
HowMuchLastEat	0.1088	0.150585
PD	0.2485	0.146022
Intense	-0.1285	-0.119339
PreferredMeds_Acetaminophen	-0.5405	0.117664
WhenLastSick	0.0635	0.094308
N	-0.1498	-0.092503
Politics	0.0839	0.080004
A	0.2752	-0.074981
HighLow_High	0.8216	0.07116
PreferredMeds_Ibuprofen	-0.6816	0.063047
E	0.1282	-0.055616
ChildSES	0.0483	-0.049005
FS	-0.1584	-0.047957
NumDoctorVisits	0.0922	0.047311
Nostalgia	-0.0659	-0.046546
NegPANAS	0.0219	0.033516
AdultSES	-0.0722	0.030117
SleepHours	-0.0782	0.022598
Familiarity	0.4672	0.01509
PosPANAS	0.0083	0.011698

Mixed Model

```
                  Mixed Linear Model Regression Results
================================================================================
==========
Model:                   MixedLM            Dependent Variable:            R
atings
No. Observations:        13218              Method:                        R
EML
No. Groups:              244                Scale:
5.3771
Min. group size:         18                 Likelihood:                    -
30153.4846
Max. group size:         55                 Converged:                     N
o
Mean group size:         54.2
--------------------------------------------------------------------------------
----------
                            Coef.    Std.Err.    z    P>|z|     [0.025
0.975]
--------------------------------------------------------------------------------
----------
const                        6.683 2059769.340  0.000 1.000  -4037067.041  4
037080.406
DrugPlacebo                 -0.207       0.130 -1.586 0.113        -0.463
0.049
Locus                       -0.231       0.047 -4.900 0.000        -0.324
-0.139
Familiarity                 -0.071       0.056 -1.261 0.207        -0.181
0.039
MedsEffectiveness           -0.043       0.053 -0.807 0.420        -0.148
0.062
LastTimeTookMeds             0.000       0.000  0.158 0.875        -0.000
0.000
FrequencyTakeMeds            0.004       0.050  0.087 0.931        -0.094
0.103
Politics                     0.082       0.072  1.134 0.257        -0.060
0.223
ChildSES                    -0.068       0.065 -1.054 0.292        -0.195
0.059
AdultSES                     0.009       0.065  0.132 0.895        -0.118
0.135
Height                       0.005       0.025  0.215 0.830        -0.044
0.054
Weight                       0.001       0.003  0.523 0.601        -0.004
0.006
SleepHours                   0.038       0.051  0.745 0.456        -0.062
0.138
WhenLastAte                  0.005       0.016  0.315 0.753        -0.026
0.036
HowMuchLastEat               0.185       0.092  2.020 0.043         0.006
0.365
SleepQuality                -0.111       0.045 -2.439 0.015        -0.200
-0.022
IllnessSeverity              0.187       0.226  0.828 0.408        -0.256
0.631
exerciseMinToday             0.002       0.005  0.364 0.716        -0.008
0.011
```

```
ExerciseRegularMins             -0.000       0.000 -0.771 0.441        -0.001
0.001
LastConsumeCaffeineHours        -0.000       0.000 -1.390 0.164        -0.001
0.000
GeneralHealth                   -0.092       0.105 -0.880 0.379        -0.298
0.113
SubjectiveIllness               -0.061       0.089 -0.687 0.492        -0.235
0.113
WhenLastSick                     0.073       0.051  1.429 0.153        -0.027
0.174
NumDoctorVisits                  0.043       0.095  0.451 0.652        -0.143
0.228
TakeMedsRecentlyYN              -0.189       0.150 -1.254 0.210        -0.483
0.106
BirthControlYN                   0.012       0.207  0.059 0.953        -0.393
0.418
MarijuanaFrequency               0.029       0.045  0.641 0.521        -0.059
0.116
AlcoholAvg                       0.155       0.089  1.738 0.082        -0.020
0.330
ArthritisYN                      0.304       0.320  0.950 0.342        -0.323
0.932
Age                             -0.033       0.032 -1.033 0.302        -0.097
0.030
YearUniversity                  -0.023       0.092 -0.255 0.799        -0.203
0.156
NeedToBelong                     0.001       0.014  0.049 0.961        -0.026
0.027
Nostalgia                       -0.047       0.058 -0.807 0.419        -0.162
0.067
EarlyFamilyEnvironment          -0.042       0.111 -0.381 0.703        -0.260
0.176
PosPANAS                         0.008       0.010  0.805 0.421        -0.011
0.026
NegPANAS                         0.035       0.013  2.762 0.006         0.010
0.060
Mellow                           0.073       0.080  0.909 0.363        -0.084
0.230
Unpretentious                   -0.091       0.080 -1.130 0.258        -0.248
0.067
Sophisticated                    0.028       0.080  0.354 0.724        -0.128
0.184
Intense                         -0.111       0.069 -1.606 0.108        -0.247
0.025
Contemporary                     0.048       0.081  0.595 0.552        -0.110
0.206
PT                               0.003       0.125  0.023 0.982        -0.242
0.248
FS                              -0.048       0.113 -0.425 0.671        -0.270
0.174
EC                               0.236       0.152  1.555 0.120        -0.061
0.534
PD                               0.208       0.122  1.707 0.088        -0.031
0.447
E                                0.026       0.096  0.273 0.785        -0.162
0.214
A                                0.016       0.150  0.107 0.915        -0.277
```

```
0.309
C                               0.028       0.132  0.212 0.832        -0.230
0.286
N                              -0.081       0.117 -0.692 0.489        -0.312
0.149
O                              -0.035       0.157 -0.224 0.823        -0.342
0.272
AIMS                           -0.002       0.003 -0.700 0.484        -0.009
0.004
PosNeg_Negative                 0.905 1130753.423  0.000 1.000  -2216235.080  2
216236.890
PosNeg_Neutral                 38.721 6263343.490  0.000 1.000 -12275888.942 12
275966.385
PosNeg_Positive                -3.910 1130434.240 -0.000 1.000  -2215614.307  2
215606.487
HighLow_High                   -0.818 1227962.135 -0.000 1.000  -2406762.378  2
406760.741
HighLow_Low                    -0.963 1227713.185 -0.000 1.000  -2406274.589  2
406272.664
HighLow_Neutral               -42.043 6720455.194 -0.000 1.000 -13171892.184 13
171808.098
SoundType_Music                -0.424
SoundType_Natural Sounds        0.109
SoundType_Speech               -0.204
Gender_female                   2.110       1.038  2.033 0.042         0.076
4.143
Gender_male                     1.767       1.029  1.717 0.086        -0.250
3.784
Gender_prefer not to answer     1.384       1.104  1.253 0.210        -0.780
3.547
PoliticalParty_democrat         1.936 2229494.508  0.000 1.000  -4369727.003  4
369730.874
PoliticalParty_libertarian      2.281 2229494.508  0.000 1.000  -4369726.657  4
369731.220
PoliticalParty_other            1.679 2229494.508  0.000 1.000  -4369727.260  4
369730.617
PoliticalParty_republican       1.920 2229494.508  0.000 1.000  -4369727.018  4
369730.859
PreferredMeds_Acetaminophen    -4.094 2229494.508 -0.000 1.000  -4369733.033  4
369724.844
PreferredMeds_Do Not Take Meds -4.984 2229494.508 -0.000 1.000  -4369733.922  4
369723.955
PreferredMeds_Ibuprofen        -4.192 2229494.508 -0.000 1.000  -4369733.131  4
369724.746
PreferredMeds_Other            -4.394 2229494.508 -0.000 1.000  -4369733.332  4
369724.545
Race_Asian                      0.109       0.365  0.297 0.766        -0.608
0.825
Race_Black                     -0.141       0.366 -0.385 0.700        -0.859
0.576
Race_Hispanic                   0.021       0.353  0.060 0.952        -0.671
0.714
Race_Other                     -0.892       0.811 -1.100 0.271        -2.482
0.698
Race_White                     -0.288       0.360 -0.802 0.423        -0.993
0.417
Group Var                       0.747       0.041
```

```
================================================================================
==========
```

Conclusions:

- **The model did not converge**
- **Basically no features are significant in the mixed model (because they are so redundant with the questionnaire data)**

4c. Predicting Arousal Ratings

Data prep

```
arousal dataframe shape:  (13218, 77)
entire dataframe shape:  (13218, 139)
arousal dataframe shape:  (8880, 77)
```

X and Y

Note that we also delete the ***Locus*** (Perceived or Induced) column because all arousal ratings are related to perceived emotion.

OLS regression

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                Ratings   R-squared:                       0.390
Model:                            OLS   Adj. R-squared:                  0.386
Method:                 Least Squares   F-statistic:                     81.78
Date:                Thu, 26 Mar 2020   Prob (F-statistic):               0.00
Time:                        07:57:01   Log-Likelihood:                -20454.
No. Observations:                8880   AIC:                         4.105e+04
Df Residuals:                    8810   BIC:                         4.154e+04
Df Model:                          69
Covariance Type:            nonrobust
================================================================================
==================
                                   coef    std err          t      P>|t|
[0.025      0.975]
--------------------------------------------------------------------------------
------------------
const                            1.6504      0.489      3.378      0.001
0.693       2.608
DrugPlacebo                     -0.1015      0.057     -1.774      0.076
-0.214       0.011
Familiarity                      0.6654      0.071      9.352      0.000
0.526       0.805
MedsEffectiveness               -0.0378      0.024     -1.606      0.108
-0.084       0.008
LastTimeTookMeds                 0.0001   6.21e-05      2.020      0.043    3.
71e-06       0.000
FrequencyTakeMeds               -0.0260      0.022     -1.183      0.237
-0.069       0.017
Politics                         0.0530      0.032      1.684      0.092
-0.009       0.115
ChildSES                         0.0436      0.028      1.539      0.124
-0.012       0.099
AdultSES                         0.0017      0.028      0.059      0.953
-0.054       0.057
Height                           0.0037      0.011      0.342      0.732
-0.018       0.025
Weight                           0.0004      0.001      0.331      0.741
-0.002       0.003
SleepHours                      -0.0896      0.022     -4.017      0.000
-0.133      -0.046
WhenLastAte                     -0.0173      0.007     -2.482      0.013
-0.031      -0.004
HowMuchLastEat                   0.1787      0.040      4.457      0.000
0.100       0.257
SleepQuality                    -0.1054      0.020     -5.291      0.000
-0.144      -0.066
IllnessSeverity                  0.0199      0.099      0.202      0.840
-0.173       0.213
exerciseMinToday                -0.0069      0.002     -3.245      0.001
-0.011      -0.003
ExerciseRegularMins           2.524e-05      0.000      0.122      0.903
-0.000       0.000
LastConsumeCaffeineHours         0.0003      0.000      2.331      0.020    5.
48e-05       0.001
GeneralHealth                   -0.0618      0.046     -1.352      0.176
```

```
-0.151      0.028
SubjectiveIllness                    0.0284      0.039      0.733      0.463
-0.047      0.104
WhenLastSick                         0.0573      0.023      2.527      0.012
0.013       0.102
NumDoctorVisits                      0.0735      0.041      1.780      0.075
-0.007      0.155
TakeMedsRecentlyYN                  -0.0647      0.066     -0.984      0.325
-0.194      0.064
BirthControlYN                       0.1500      0.091      1.657      0.098
-0.027      0.327
MarijuanaFrequency                  -0.0257      0.019     -1.319      0.187
-0.064      0.012
AlcoholAvg                           0.1291      0.039      3.303      0.001
0.052       0.206
ArthritisYN                          0.3384      0.145      2.326      0.020
0.053       0.624
Age                                 -0.0083      0.014     -0.588      0.556
-0.036      0.019
YearUniversity                       0.0706      0.040      1.768      0.077
-0.008      0.149
NeedToBelong                         0.0110      0.006      1.832      0.067
-0.001      0.023
Nostalgia                           -0.0533      0.026     -2.087      0.037
-0.103     -0.003
EarlyFamilyEnvironment              -0.0971      0.049     -1.995      0.046
-0.193     -0.002
PosPANAS                             0.0240      0.004      5.688      0.000
0.016       0.032
NegPANAS                             0.0164      0.005      2.976      0.003
0.006       0.027
Mellow                              -0.0324      0.035     -0.928      0.353
-0.101      0.036
Unpretentious                       -0.1594      0.035     -4.544      0.000
-0.228     -0.091
Sophisticated                        0.1047      0.035      2.996      0.003
0.036       0.173
Intense                              0.0205      0.030      0.675      0.500
-0.039      0.080
Contemporary                         0.1279      0.035      3.607      0.000
0.058       0.197
PT                                   0.0613      0.055      1.122      0.262
-0.046      0.168
FS                                  -0.1671      0.050     -3.357      0.001
-0.265     -0.070
EC                                   0.1969      0.066      2.965      0.003
0.067       0.327
PD                                   0.2692      0.054      4.997      0.000
0.164       0.375
E                                    0.2895      0.043      6.797      0.000
0.206       0.373
A                                    0.1713      0.065      2.628      0.009
0.044       0.299
C                                   -0.2885      0.058     -4.998      0.000
-0.402     -0.175
N                                   -0.0338      0.051     -0.661      0.508
-0.134      0.066
```

```
O                                0.0136      0.069      0.198      0.843
-0.121       0.148
AIMS                            -0.0049      0.002     -3.274      0.001
-0.008      -0.002
PosNeg_Negative                  0.7421      0.166      4.483      0.000
0.418       1.067
PosNeg_Neutral                  -0.3712      0.165     -2.244      0.025
-0.696      -0.047
PosNeg_Positive                  1.2796      0.166      7.718      0.000
0.955       1.605
HighLow_High                     2.4821      0.166     14.994      0.000
2.158       2.807
HighLow_Low                     -0.4604      0.166     -2.777      0.006
-0.785      -0.135
HighLow_Neutral                 -0.3712      0.165     -2.244      0.025
-0.696      -0.047
SoundType_Music                  0.7480      0.168      4.455      0.000
0.419       1.077
SoundType_Natural Sounds         1.1220      0.168      6.679      0.000
0.793       1.451
SoundType_Speech                -0.2196      0.168     -1.308      0.191
-0.549       0.109
Gender_female                    1.1362      0.452      2.514      0.012
0.250       2.022
Gender_male                      1.0100      0.449      2.251      0.024
0.131       1.889
Gender_prefer not to answer      0.7875      0.481      1.638      0.102
-0.155       1.730
PoliticalParty_democrat         -1.2030      0.223     -5.394      0.000
-1.640      -0.766
PoliticalParty_libertarian      -0.4850      0.235     -2.065      0.039
-0.945      -0.025
PoliticalParty_other            -1.5415      0.229     -6.738      0.000
-1.990      -1.093
PoliticalParty_republican       -1.2120      0.238     -5.102      0.000
-1.678      -0.746
PreferredMeds_Acetaminophen     -0.7674      0.224     -3.433      0.001
-1.206      -0.329
PreferredMeds_Do Not Take Meds  -1.4631      0.283     -5.168      0.000
-2.018      -0.908
PreferredMeds_Ibuprofen         -1.0057      0.229     -4.389      0.000
-1.455      -0.557
PreferredMeds_Other             -1.2053      0.232     -5.194      0.000
-1.660      -0.750
Race_Asian                       0.3715      0.159      2.336      0.020
0.060       0.683
Race_Black                       0.5499      0.160      3.439      0.001
0.236       0.863
Race_Hispanic                    0.1529      0.154      0.991      0.322
-0.149       0.455
Race_Other                       0.4364      0.353      1.235      0.217
-0.256       1.129
Race_White                       0.2457      0.157      1.566      0.117
-0.062       0.553
==============================================================================
Omnibus:                       22.511   Durbin-Watson:                   1.768
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               18.176
```

```
Skew:                           0.020   Prob(JB):                     0.000113
Kurtosis:                       2.782   Cond. No.                     1.49e+16
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The smallest eigenvalue is 1.19e-23. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
```

Conclusions:

- **The measured variables explain 39% of the variance in the arousal ratings (adjusted R^2 = 0.39)**
- **There was *no difference* in arousal ratings between those who took acetaminophen and those who took a placebo.**

Name	Coefficient	p	Interpretation
HighLow_High	2.4821	0	Stimuli that exhibit **high arousal** resulted in comparatively **more energetic** ratings.
PosNeg_Positive	1.2796	0	**Positively-valenced stimuli** resulted in comparatively **high energy** ratings.
SoundType_Natural Sounds	1.122	0	**Natural Sounds** resulted in comparatively **higher energy** ratings *(and its effect is comparatively large next to music and especially compared to speech, which is not significant)*
SoundType_Music	0.748	0	**Musical stimuli** resulted in comparatively **higher energy** ratings *(compared to speech, which was non-significant)*
PosNeg_Negative	0.7421	0	**Negatively-valenced stimuli** resulted in comparatively **higher energy** ratings.
Familiarity	0.6654	0	Those who are more **familiar** with the (musical) stimuli rated the stimuli as being **higher in energy.**
HighLow_Low	-0.4604	0.006	Stimuli that exhibit **low arousal** resulted in comparatively **lower energy** ratings.
E	0.2895	0	Those who are more **extraverted** rated emotions as relatively **more energetic** than those who score less high on this personality dimension.
C	-0.2885	0	Those who are more **conscientious** rated emotions as relatively **less energetic** than those who score less high on this personality dimension.
PD	0.2692	0	Those who score higher on the **Personal Distress component of empathy** rated emotions as relatively **more energetic** than those who score less high on this trait.
EC	0.1969	0.003	Those who score higher on the **Empathic Concern component of empathy** rated emotions as relatively **more energetic** than those who score less high on this trait.

Name	Coefficient	p	Interpretation
A	0.1713	0.009	Those who are more **agreeable** rated emotions as relatively **more energetic** than those who score less high on this personality dimension.
FS	-0.1671	0.001	Those who score higher on the **Fantasy component of empathy** rated emotions as relatively **less energetic** than those who score less high on this trait.
Unpretentious	-0.1594	0	Those who **prefer "Unpretentious" music** (Pop, Country, Religious) rated emotions as relatively **less energetic** (broadly, in response to all stimuli) than those who do not like this genre of music (although this effect is small).
Contemporary	0.1279	0	Those who **prefer "Contemporary" music** (Rap/Hip Hop, Soul/R&B, Funk, Reggae) rated emotions as relatively **more energetic** (broadly, in response to all stimuli) than those who do not like this genre of music (although this effect is small).
Sophisticated	0.1047	0.003	Those who **prefer "Sophisticated" music** (Blues, Jazz, Bluegrass, Folk, Classical, Gospel, Opera) rated emotions as relatively **more energetic** (broadly, in response to all stimuli) than those who do not like this genre of music.
Nostalgia	-0.0533	0.037	Those who scored high on **nostalgia** rated the stimuli aas being relatively **less energetic** than those who score low on this trait (although this effect is small).

OLS regression with lasso regularization

Conclusions:

- **This didn't change very much from the original mode.**
- **This chart shows the "non-zero" variables from the original OLS --> these are the only (significant) features kept in the regularized model**
- **Note that I used an arbitrary cutoff of 0.1 coefficient value to count as significant.**

Features Remaining In the Model	OLS Coefficient	OLS **with lasso regularization** Coefficient
PosNeg_Neutral	-0.3712	-2.166824
HighLow_High	2.4821	1.47148
HighLow_Low	-0.4604	-1.47148
PosNeg_Positive	1.2796	0.865751
Familiarity	0.6654	0.686584
Race_Black	0.5499	0.609665
SoundType_Natural Sounds	1.122	0.589999

Features Remaining In the Model	OLS Coefficient	OLS **with lasso regularization** Coefficient
PoliticalParty_other	-1.5415	-0.542389
Gender_male	1.01	-0.517634
PreferredMeds_Do Not Take Meds	-1.4631	-0.470649
ArthritisYN	0.3384	0.393543
Gender_female	1.1362	-0.362125
PoliticalParty_libertarian	-0.485	0.351171
Race_Asian	0.3715	0.343275
PosNeg_Negative	0.7421	0.328635
PoliticalParty_republican	-1.212	-0.32835
PoliticalParty_democrat	-1.203	-0.315477
C	-0.2885	-0.312441
EC	0.1969	0.251219
PreferredMeds_Acetaminophen	-0.7674	0.217798
PreferredMeds_Other	-1.2053	-0.213326
SoundType_Music	0.748	0.208025
HowMuchLastEat	0.1787	0.201207
A	0.1713	-0.193694
FS	-0.1671	-0.180844
EarlyFamilyEnvironment	-0.0971	-0.172594
PD	0.2692	0.154598
AlcoholAvg	0.1291	0.144398
Unpretentious	-0.1594	-0.127286
E	0.2895	0.124883
Sophisticated	0.1047	0.105058

Mixed Model

```
                   Mixed Linear Model Regression Results
================================================================================
========
Model:                     MixedLM         Dependent Variable:            Rat
ings
No. Observations:          8880            Method:                        REM
L
No. Groups:                240             Scale:                         5.1
277
Min. group size:           37              Likelihood:                    -20
122.0205
Max. group size:           37              Converged:                     No
Mean group size:           37.0
--------------------------------------------------------------------------------
--------
                              Coef.    Std.Err.    z    P>|z|    [0.025
0.975]
--------------------------------------------------------------------------------
--------
const                         -1.845
DrugPlacebo                   -0.105       0.157 -0.668 0.504       -0.412
0.202
Familiarity                    0.573       0.070  8.192 0.000        0.436
0.710
MedsEffectiveness             -0.037       0.065 -0.577 0.564       -0.164
0.089
LastTimeTookMeds               0.000       0.000  0.735 0.462       -0.000
0.000
FrequencyTakeMeds             -0.026       0.060 -0.435 0.663       -0.144
0.092
Politics                       0.054       0.086  0.629 0.529       -0.115
0.224
ChildSES                       0.043       0.078  0.554 0.580       -0.109
0.196
AdultSES                       0.003       0.078  0.038 0.970       -0.149
0.155
Height                         0.004       0.030  0.135 0.893       -0.055
0.063
Weight                         0.000       0.003  0.127 0.899       -0.006
0.006
SleepHours                    -0.089       0.061 -1.458 0.145       -0.209
0.031
WhenLastAte                   -0.017       0.019 -0.903 0.366       -0.055
0.020
HowMuchLastEat                 0.180       0.110  1.635 0.102       -0.036
0.395
SleepQuality                  -0.106       0.055 -1.940 0.052       -0.213
0.001
IllnessSeverity                0.018       0.271  0.066 0.948       -0.513
0.548
exerciseMinToday              -0.007       0.006 -1.184 0.237       -0.018
0.005
ExerciseRegularMins            0.000       0.001  0.043 0.966       -0.001
0.001
LastConsumeCaffeineHours       0.000       0.000  0.846 0.397       -0.000
0.001
```

```
GeneralHealth                  -0.063      0.125 -0.505 0.613      -0.309
0.182
SubjectiveIllness               0.030      0.106  0.287 0.774      -0.178
0.238
WhenLastSick                    0.057      0.062  0.915 0.360      -0.065
0.179
NumDoctorVisits                 0.073      0.113  0.642 0.521      -0.149
0.295
TakeMedsRecentlyYN             -0.063      0.180 -0.349 0.727      -0.416
0.291
BirthControlYN                  0.151      0.248  0.608 0.543      -0.336
0.638
MarijuanaFrequency             -0.025      0.053 -0.474 0.635      -0.130
0.079
AlcoholAvg                      0.130      0.107  1.208 0.227      -0.081
0.340
ArthritisYN                     0.342      0.399  0.856 0.392      -0.440
1.124
Age                            -0.008      0.039 -0.214 0.831      -0.084
0.068
YearUniversity                  0.069      0.110  0.632 0.527      -0.146
0.284
NeedToBelong                    0.011      0.016  0.685 0.494      -0.021
0.043
Nostalgia                      -0.054      0.070 -0.773 0.440      -0.191
0.083
EarlyFamilyEnvironment         -0.095      0.134 -0.715 0.475      -0.357
0.166
PosPANAS                        0.024      0.012  2.084 0.037       0.001
0.047
NegPANAS                        0.017      0.015  1.107 0.268      -0.013
0.046
Mellow                         -0.031      0.096 -0.324 0.746      -0.219
0.156
Unpretentious                  -0.158      0.096 -1.646 0.100      -0.347
0.030
Sophisticated                   0.104      0.096  1.084 0.278      -0.084
0.292
Intense                         0.020      0.083  0.245 0.807      -0.143
0.183
Contemporary                    0.127      0.097  1.308 0.191      -0.063
0.318
PT                              0.060      0.150  0.402 0.688      -0.233
0.354
FS                             -0.168      0.137 -1.227 0.220      -0.435
0.100
EC                              0.198      0.182  1.085 0.278      -0.159
0.555
PD                              0.269      0.148  1.823 0.068      -0.020
0.559
E                               0.290      0.117  2.478 0.013       0.061
0.519
A                               0.172      0.179  0.962 0.336      -0.179
0.523
C                              -0.289      0.158 -1.822 0.068      -0.599
0.022
N                              -0.035      0.140 -0.248 0.804      -0.310
```

```
0.240
O                              0.012       0.188  0.064 0.949       -0.357      
0.381
AIMS                          -0.005       0.004 -1.192 0.233       -0.013      
0.003
PosNeg_Negative                2.346                                            
PosNeg_Neutral               -19.298                                            
PosNeg_Positive                2.885                                            
HighLow_High                   4.758                                            
HighLow_Low                    1.817                                            
HighLow_Neutral               22.437                                            
SoundType_Music                0.355   56791.604  0.000 1.000  -111309.144  11
1309.854
SoundType_Natural Sounds       0.694   26497.446  0.000 1.000   -51933.346   5
1934.734
SoundType_Speech              -0.647   55920.062 -0.000 1.000  -109601.955  10
9600.661
Gender_female                  1.154       1.239  0.931 0.352       -1.275      
3.583
Gender_male                    1.023       1.231  0.831 0.406       -1.389      
3.435
Gender_prefer not to answer    0.807       1.319  0.612 0.541       -1.778      
3.392
PoliticalParty_democrat      -19.722 3779355.264 -0.000 1.000 -7407419.924 740
7380.479
PoliticalParty_libertarian   -19.008 3779355.264 -0.000 1.000 -7407419.209 740
7381.194
PoliticalParty_other         -20.067 3779355.264 -0.000 1.000 -7407420.269 740
7380.135
PoliticalParty_republican    -19.734 3779355.264 -0.000 1.000 -7407419.936 740
7380.467
PreferredMeds_Acetaminophen   17.747 3779355.264  0.000 1.000 -7407382.454 740
7417.949
PreferredMeds_Do Not Take Meds  17.056 3779355.264  0.000 1.000 -7407383.146 740
7417.258
PreferredMeds_Ibuprofen       17.509 3779355.264  0.000 1.000 -7407382.692 740
7417.711
PreferredMeds_Other           17.310 3779355.264  0.000 1.000 -7407382.892 740
7417.512
Race_Asian                     0.368       0.436  0.844 0.398       -0.487      
1.224
Race_Black                     0.547       0.439  1.247 0.212       -0.313      
1.407
Race_Hispanic                  0.156       0.423  0.370 0.712       -0.673      
0.986
Race_Other                     0.440       0.969  0.454 0.650       -1.459      
2.340
Race_White                     0.236       0.430  0.549 0.583       -0.607      
1.079
Group Var                      1.064                                            
================================================================================
========
```

Conclusions:

- **The model did not converge**
- **Basically no features are significant in the mixed model (because they are so redundant with the questionnaire data)**

4d. Predicting All Ratings

Data prep

Here, we combine the *positive*, *negative*, and *arousal* dataframes into a single dataframe.

```
positive shape: (13218, 78)
negative shape: (13218, 78)
arousal shape: (8880, 78)

pos/neg colnames:  [ True  True  True  True  True  True  True  True  True  True
True  True
  True  True  True  True  True  True  True  True  True  True  True  True
  True  True  True  True  True  True  True  True  True  True  True  True
  True  True  True  True  True  True  True  True  True  True  True  True
  True  True  True  True  True  True  True  True  True  True  True  True
  True  True  True  True  True  True  True  True  True  True  True  True
  True  True  True  True  True  True]

pos/ar colnames:  [ True  True  True  True  True  True  True  True  True  True
True  True
  True  True  True  True  True  True  True  True  True  True  True  True
  True  True  True  True  True  True  True  True  True  True  True  True
  True  True  True  True  True  True  True  True  True  True  True  True
  True  True  True  True  True  True  True  True  True  True  True  True
  True  True  True  True  True  True  True  True  True  True  True  True
  True  True  True  True  True  True]

ar/neg colnames:  [ True  True  True  True  True  True  True  True  True  True
True  True
  True  True  True  True  True  True  True  True  True  True  True  True
  True  True  True  True  True  True  True  True  True  True  True  True
  True  True  True  True  True  True  True  True  True  True  True  True
  True  True  True  True  True  True  True  True  True  True  True  True
  True  True  True  True  True  True  True  True  True  True  True  True
  True  True  True  True  True  True]

alldata shape: (35316, 78)
```

X and Y

OLS regression

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                Ratings   R-squared:                       0.093
Model:                            OLS   Adj. R-squared:                  0.092
Method:                 Least Squares   F-statistic:                     50.46
Date:                Thu, 26 Mar 2020   Prob (F-statistic):               0.00
Time:                        07:57:19   Log-Likelihood:                -91076.
No. Observations:               35316   AIC:                         1.823e+05
Df Residuals:                   35243   BIC:                         1.829e+05
Df Model:                          72
Covariance Type:            nonrobust
================================================================================
==================
                                   coef    std err          t      P>|t|
[0.025      0.975]
--------------------------------------------------------------------------------
------------------
const                            1.2371      0.276      4.478      0.000
0.696       1.779
DrugPlacebo                     -0.1878      0.038     -4.991      0.000
-0.262      -0.114
Locus                           -0.3237      0.045     -7.269      0.000
-0.411      -0.236
Familiarity                      0.3066      0.045      6.784      0.000
0.218       0.395
MedsEffectiveness               -0.0343      0.015     -2.222      0.026
-0.065      -0.004
LastTimeTookMeds              2.722e-05   4.09e-05      0.666      0.506   -5.
29e-05       0.000
FrequencyTakeMeds               -0.0339      0.014     -2.342      0.019
-0.062      -0.006
Politics                         0.0747      0.021      3.605      0.000
0.034       0.115
ChildSES                         0.0029      0.019      0.153      0.878
-0.034       0.039
AdultSES                        -0.0219      0.019     -1.175      0.240
-0.058       0.015
Height                           0.0064      0.007      0.886      0.376
-0.008       0.020
Weight                           0.0020      0.001      2.751      0.006
0.001       0.003
SleepHours                      -0.0382      0.015     -2.597      0.009
-0.067      -0.009
WhenLastAte                     -0.0025      0.005     -0.538      0.591
-0.011       0.007
HowMuchLastEat                   0.1570      0.026      5.942      0.000
0.105       0.209
SleepQuality                    -0.0745      0.013     -5.646      0.000
-0.100      -0.049
IllnessSeverity                  0.1269      0.065      1.951      0.051
-0.001       0.254
exerciseMinToday                -0.0041      0.001     -2.979      0.003
-0.007      -0.001
ExerciseRegularMins             -0.0003      0.000     -1.908      0.056
-0.001    7.01e-06
LastConsumeCaffeineHours        -0.0002   9.73e-05     -2.098      0.036
```

```
-0.000   -1.34e-05
GeneralHealth                      -0.0390      0.030     -1.290      0.197
-0.098       0.020
SubjectiveIllness                  -0.0349      0.026     -1.368      0.171
-0.085       0.015
WhenLastSick                        0.0642      0.015      4.314      0.000
0.035       0.093
NumDoctorVisits                     0.0697      0.027      2.562      0.010
0.016       0.123
TakeMedsRecentlyYN                 -0.1478      0.043     -3.416      0.001
-0.233      -0.063
BirthControlYN                      0.0640      0.060      1.074      0.283
-0.053       0.181
MarijuanaFrequency                  0.0036      0.013      0.282      0.778
-0.021       0.029
AlcoholAvg                          0.1075      0.026      4.173      0.000
0.057       0.158
ArthritisYN                         0.2880      0.095      3.043      0.002
0.102       0.474
Age                                -0.0041      0.009     -0.442      0.659
-0.022       0.014
YearUniversity                      0.0031      0.026      0.118      0.906
-0.049       0.055
NeedToBelong                        0.0050      0.004      1.266      0.205
-0.003       0.013
Nostalgia                          -0.0545      0.017     -3.238      0.001
-0.087      -0.021
EarlyFamilyEnvironment             -0.0342      0.032     -1.067      0.286
-0.097       0.029
PosPANAS                            0.0123      0.003      4.455      0.000
0.007       0.018
NegPANAS                            0.0252      0.004      6.949      0.000
0.018       0.032
Mellow                              0.0065      0.023      0.281      0.779
-0.039       0.052
Unpretentious                      -0.0901      0.023     -3.892      0.000
-0.135      -0.045
Sophisticated                       0.0832      0.023      3.618      0.000
0.038       0.128
Intense                            -0.0838      0.020     -4.197      0.000
-0.123      -0.045
Contemporary                        0.0599      0.023      2.573      0.010
0.014       0.106
PT                                  0.0180      0.036      0.498      0.619
-0.053       0.089
FS                                 -0.1174      0.033     -3.589      0.000
-0.182      -0.053
EC                                  0.1045      0.044      2.390      0.017
0.019       0.190
PD                                  0.2441      0.035      6.888      0.000
0.175       0.314
E                                   0.1284      0.028      4.604      0.000
0.074       0.183
A                                   0.1491      0.043      3.465      0.001
0.065       0.233
C                                  -0.0731      0.038     -1.920      0.055
-0.148       0.002
```

```
N                                -0.0967      0.034     -2.869      0.004
-0.163      -0.031
O                                 0.0198      0.045      0.438      0.661
-0.069       0.108
AIMS                             -0.0025      0.001     -2.527      0.012
-0.004      -0.001
PosNeg_Negative                   0.6284      0.094      6.675      0.000
0.444       0.813
PosNeg_Neutral                   -0.1628      0.094     -1.730      0.084
-0.347       0.022
PosNeg_Positive                   0.7715      0.094      8.182      0.000
0.587       0.956
HighLow_High                      1.2057      0.094     12.809      0.000
1.021       1.390
HighLow_Low                       0.1942      0.094      2.059      0.039
0.009       0.379
HighLow_Neutral                  -0.1628      0.094     -1.730      0.084
-0.347       0.022
SoundType_Music                   0.5603      0.096      5.838      0.000
0.372       0.748
SoundType_Natural Sounds          0.6139      0.096      6.399      0.000
0.426       0.802
SoundType_Speech                  0.0628      0.097      0.649      0.517
-0.127       0.253
Gender_female                     1.4605      0.298      4.906      0.000
0.877       2.044
Gender_male                       1.2206      0.295      4.132      0.000
0.642       1.800
Gender_prefer not to answer       0.8962      0.317      2.829      0.005
0.275       1.517
PoliticalParty_democrat          -1.0349      0.147     -7.043      0.000
-1.323      -0.747
PoliticalParty_libertarian       -0.5747      0.155     -3.714      0.000
-0.878      -0.271
PoliticalParty_other             -1.3326      0.151     -8.839      0.000
-1.628      -1.037
PoliticalParty_republican        -1.1059      0.156     -7.069      0.000
-1.413      -0.799
PreferredMeds_Acetaminophen      -0.7306      0.147     -4.962      0.000
-1.019      -0.442
PreferredMeds_Do Not Take Meds   -1.3271      0.186     -7.116      0.000
-1.693      -0.962
PreferredMeds_Ibuprofen          -0.8806      0.151     -5.834      0.000
-1.176      -0.585
PreferredMeds_Other              -1.1098      0.153     -7.260      0.000
-1.409      -0.810
Race_Asian                        0.2579      0.105      2.462      0.014
0.053       0.463
Race_Black                        0.0814      0.105      0.774      0.439
-0.125       0.288
Race_Hispanic                    -0.0396      0.102     -0.390      0.697
-0.239       0.159
Race_Other                       -0.4925      0.233     -2.116      0.034
-0.949      -0.036
Race_White                       -0.0824      0.103     -0.797      0.425
-0.285       0.120
RatingType_Arousal                0.8362      0.096      8.697      0.000
```

```
0.648       1.025
RatingType_Negative                0.1939      0.095      2.038      0.042
0.007       0.380
RatingType_Positive                0.2070      0.095      2.175      0.030
0.020       0.394
==============================================================================
Omnibus:                    13596.035   Durbin-Watson:                   1.916
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             1985.370
Skew:                           0.206   Prob(JB):                         0.00
Kurtosis:                       1.914   Cond. No.                     2.33e+16
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The smallest eigenvalue is 1.93e-23. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
```

Conclusions:

- **The measured variables explain 9% of the variance in overall emotion ratings (adjusted R^2 = 0.09)**
- **The ingestion of acetaminophen--overall--blunted ratings of positive emotions compared to the ingestion of a placebo.**
- **The low variance explained means that people used there is little similarities in how people used the three types of scales (positive emotions, negative emotions, and arousal).**
- **It's likely that there's effects going in opposite directions--for example, people use the opposite criteria to judge positive and negative emotion ratings--so it makes sense the R^2 is much closer to 0.**
- **The order of the features (in terms of absolute values of the coefficients) is very similar to the order that they are entered into the equation during the regression. This hints that the feature importances can't be determined reliably.**

Name	Coefficient	p	Interpretation
DrugPlacebo -0.1878 0	*Those who* ***took acetaminophen*** *rated emotions as* ***lower overall*** *compared to those who took the placebo.* ***Namely, acetaminophen blunted the emotional responses.***		
HighLow_High	1.2057	0	Stimuli that exhibit **high arousal** resulted in comparatively **higher overall ratings**.
RatingType_Arousal	0.8362	0	**Arousal ratings** were overall comparatively high in number (compared to positive and negative emotion ratings).
PosNeg_Positive	0.7715	0	**Positively-valenced stimuli** resulted in comparatively **higher overall ratings**.

Name	Coefficient	p	Interpretation
PosNeg_Negative	0.6284	0	**Negatively-valenced stimuli** resulted in comparatively **higher overall ratings** (but not as much as positively-valenced stimuli).
SoundType_Natural Sounds	0.6139	0	**Natural Sounds** resulted in comparatively **higher overall ratings**(compared to music and especially compared to speech, which wasn't significant)*
SoundType_Music	0.5603	0	**Natural Sounds** resulted in comparatively **higher overall ratings**(compared to speech, which wasn't significant)*
Locus	-0.3237	0	Compared to perceived emotion ratings, **induced emotion ratings** were **lower in overall score**.
Familiarity	0.3066	0	Those who are more **familiar** with the (musical) stimuli rated stimuli **comparatively higher overall** compared to people less familiar with the music.
PD	0.2441	0	Those who score higher on the **Personal Distress component of empathy** rated emotions as relatively **higher overall** than those who score less high on this trait.
RatingType_Positive	0.207	0.03	**Positive emotion ratings** were overall comparatively high in number (compared to negative ratings).
HighLow_Low	0.1942	0.039	Stimuli that exhibit **low arousal** resulted in comparatively **high overall ratings** (but the effect is smaller than the high arousal music).
RatingType_Negative	0.1939	0.042	**Negative emotion ratings** were overall comparatively high in number (but low compared to arousal ratings and positive emotion ratings).
A	0.1491	0.001	Those who are more **agreeable** rated emotions as relatively **higher overall** than those who score less high on this personality dimension.
E	0.1284	0	Those who are more **extraverted** rated emotions as relatively **higher overall** than those who score less high on this personality dimension.

Name	Coefficient	p	Interpretation
FS	-0.1174	0	Those who score higher on the **Fantasy component of empathy** rated emotions as relatively **lower overall** than those who score less high on this trait.
EC	0.1045	0.017	Those who score higher on the **Empathic Concern component of empathy** rated emotions as relatively **higher overall** than those who score less high on this trait.
N	-0.0967	0.004	Those who are more **neurotic** rated emotions as relatively **lower overall** than those who score less high on this personality dimension.
Unpretentious	-0.0901	0	Those who **prefer "Unpretentious" music** (Pop, Country, Religious) rated emotions as relatively **lower overall** (broadly, in response to all stimuli) than those who do not like this genre of music (although this effect is small).
Intense	-0.0838	0	Those who **prefer "Intense" music** (Rock, Punk, Alternative, Heavy Metal) rated emotions as relatively **lower overall** (broadly, in response to all stimuli) than those who do not like this genre of music.
Sophisticated	0.0832	0	Those who **prefer "Sophisticated" music** (Blues, Jazz, Bluegrass, Folk, Classical, Gospel, Opera) rated emotions as relatively **higher overall** (broadly, in response to all stimuli) than those who do not like this genre of music.
Contemporary	0.0599	0.01	Those who **prefer "Contemporary" music** (Rap/Hip Hop, Soul/R&B, Funk, Reggae) rated emotions as relatively **higher overall** (broadly, in response to all stimuli) than those who do not like this genre of music (although this effect is small).
Nostalgia	-0.0545	0.001	Those who scored high on **nostalgia** rated relatively **lower overall** (although this effect is small).

OLS regression with lasso regularization

Conclusions:

- **This didn't change very much from the original mode.**
- **This chart shows the "non-zero" variables from the original OLS --> these are the only (significant) features kept in the regularized model**
- **Note that I used an arbitrary cutoff of 0.1 coefficient value to count as significant.**

Features Remaining In the Model	OLS Coefficient	OLS **with lasso regularization** Coefficient
DrugPlacebo	***-0.1878***	***-0.162255***
RatingType_Arousal	0.8362	0.871276
Gender_prefer not to answer	0.8962	-0.617293
PoliticalParty_other	-1.3326	-0.524402
HighLow_High	1.2057	0.505651
HighLow_Low	0.1942	-0.505651
PosNeg_Positive	0.7715	0.463598
PreferredMeds_Do Not Take Meds	-1.3271	-0.451093
PoliticalParty_republican	-1.1059	-0.343582
Race_Asian	0.2579	0.332973
Locus	-0.3237	-0.322413
PosNeg_Negative	0.6284	0.320646
Familiarity	0.3066	0.317096
ArthritisYN	0.288	0.314295
Gender_male	1.2206	-0.293518
PoliticalParty_democrat	-1.0349	-0.256109
RatingType_Positive	0.207	0.242152
Race_Other	-0.4925	-0.231538
RatingType_Negative	0.1939	0.229102
PreferredMeds_Other	-1.1098	-0.212399
PD	0.2441	0.182542
PoliticalParty_libertarian	-0.5747	0.178751
TakeMedsRecentlyYN	-0.1478	-0.167184
EC	0.1045	0.165425
SoundType_Natural Sounds	0.6139	0.158922
PreferredMeds_Acetaminophen	-0.7306	0.155727
HowMuchLastEat	0.157	0.145694
FS	-0.1174	-0.113215

Features Remaining In the Model	OLS Coefficient	OLS **with lasso regularization** Coefficient
AlcoholAvg	0.1075	0.106252
SoundType_Music	0.5603	0.10102

Mixed Model

Conclusions:

- **The model did not converge --> the matrix was singular**
- **Basically no features are significant in the mixed model (because they are so redundant with the questionnaire data)**

4e. Predicting Drug vs. Placebo from a Logisitic Regression

In order to use people's ratings to predict whether they were given the drug or a placebo, we have to change the data back to a wide format.

Namely, there should be one row for each participant.

Turn one-hot encoded stimuli back into a single column

First, we have to turn all the one-hot encoded values back to a single categorical column.

```
(13218, 139)
(13218, 139)
```

Stimulus

PosNeg

HighLow

SoundType

PoliticalParty

PreferredMeds

Gender

Rearrange columns

```
(13218, 89)
```

Separate into perceived/induced/questionnaires

We need to separate the three study components (perceived emotion ratings, induced emotion ratings, and questionnaire responses) into separate dataframes.

Each of these dataframes should have 244 rows--one for each participant.

Questionnaires

First, we need to pull out all of the questions from the questionnaires.

```
wideQ shape:  (244, 57)
```

	FixedID	DrugPlacebo	MedsEffectiveness	LastTimeTookMeds	FrequencyTakeMeds
0	1.0	0	4.0	1095.0	1.0
1	2.0	1	4.0	365.0	4.0
2	3.0	0	1.0	1.0	5.0
3	4.0	1	1.0	3650.0	1.0
4	5.0	0	5.0	180.0	4.0

5 rows × 57 columns

Perceived

Next, we need to extract all of the perceived emotion information.

We ideally want there to be several columns per stimulus. For Fear Music 1, for example:

- Perceived_Positive_Fear Music 1
- Perceived_Negative_Fear Music 1
- Perceived_Arousal_Fear Music 1
- Perceived_Familiarity_Fear Music 1

In order to create these columns, we need to first stack all of the ratings (positive, negative, arousal, familiarity) on top of each other under a single column: ***Ratings***. That's what is done here.

```
wideP: (35520, 6)
```

	FixedID	DrugPlacebo	Stimulus	Rating	Locus	RatingType
0	1.0	0	Neutral Speech 2	1.0	Perceived	Positive
1	1.0	0	Tender Music 5	0.0	Perceived	Familiarity
2	1.0	0	Fear Speech 1	0.0	Perceived	Familiarity
3	1.0	0	Happy Speech 3	0.0	Perceived	Familiarity
4	1.0	0	Neutral Speech 1	0.0	Perceived	Familiarity

Induced

We repeat the above process for induced emotion.

Note that there are no arousal ratings for induced emotion.

```
wideI: (13014, 6)
```

	FixedID	DrugPlacebo	Stimulus	Rating	Locus	RatingType
0	1.0	0	Fear Music 2	2.0	Induced	Positive
1	1.0	0	Tender Music 1	6.0	Induced	Negative
2	1.0	0	Happy Music 2	3.0	Induced	Negative
3	1.0	0	Neutral Human	6.0	Induced	Negative
4	1.0	0	Positive-Valence High-Arousal Non-human	3.0	Induced	Negative

Make Wide P and I dataframe

Next, we stack the induced and perceived dataframes on top of each other. Then, we unstack them in order to make the wide dataframe.

The resulting dataframe will have 244 rows--one for each participant.

For each stimulus, there are 7 ratings. For example:

- Induced_Positive_Fear Music 1
- Induced_Negative_Fear Music 1
- Induced_Familiarity_Fear Music 1
- Perceived_Positive_Fear Music 1
- Perceived_Negative_Fear Music 1
- Perceived_Arousal_Fear Music 1
- Perceived_Familiarity_Fear Music 1

```
null shape: (7, 204)
null values:  Empty DataFrame
Columns: [0]
Index: []
widePI shape: (244, 204)
```

	FixedID	DrugPlacebo	Induced_Familiarity_Fear Music 1	Induced_Familiarity_Fear Music 2	Induc
0	1.0	0	0.0	0.0	0.0
1	2.0	1	0.0	0.0	0.0
2	3.0	0	0.0	0.0	0.0
3	4.0	1	2.0	1.0	1.0
4	5.0	0	0.0	0.0	0.0

5 rows × 204 columns

Combine Wide Questionnaires, Perceived, and Induced into one dataframe

Finally, we create the wide dataframe with info about the questionnaires, perceived, and induced blocks.

```
shapes equal?  True
wide shape:  (244, 259)
```

	FixedID	DrugPlacebo	MedsEffectiveness	LastTimeTookMeds	FrequencyTakeMeds
0	1.0	0	4.0	1095.0	1.0
1	2.0	1	4.0	365.0	4.0
2	3.0	0	1.0	1.0	5.0
3	4.0	1	1.0	3650.0	1.0
4	5.0	0	5.0	180.0	4.0

5 rows × 259 columns

Logistic Regression predicting Drug vs. Placebo

Now we can perform logistic regression, with DrugPlacebo as the dependent variable.

Correlations

Examining correlations for multicollinearity purposes.

Regression with L1 regularization

Because there's so many columns (especially compared to rows), we need to use a regularization parameter. L1 regularization is the best fit, since it can fit feature values to 0.

Conclusions:

- **The model did not converge, even with regularization**
- **The matrix was singular, indicating that some features perfectly predict others**
- **Therefore, the logistic regression cannot be carried out in this forma**

3. Are Emotion Terms Used Differently

The last goal of this analysis is to test whether those who ingested acetaminophen and those who took a placebo believe the same emotions apply to the stimuli.

Recall that participants were asked to answer the following questions about the specific emotions:

- Identify which emotion(s) the audio file [*represents* or *makes you feel*] by checking the appropriate emotion(s) from the following list. You may select as few or as many as you like
- Given this list of emotion terms you chose, which one(s), if any, strongly apply?

These responses were coded on a 3-point scale:

- Does not apply = 0
- Applies = 1
- Strongly applies = 2

We will investigate this question graphically.

3a. Make dataframe

First we need to make a dataframe with all the specific emotion ratings.

```
perceived shape:  (8880, 52)
induced shape:  (4338, 52)
emotions shape:  (13218, 52)
```

3b. Plot distributions of Drug vs Placebo

In order to compare the specific emotions rated by those who took the drug vs. those who took the placebo, an exploratory graph can give us a lot of information. We will first do this with the perceived emotion ratings and then with the induced emotion ratings.

Perceived Emotion

Dataframe creation

There were fewer emotions examined in the perceived emotions section. We will delete the ones not studied first.

Means

Examine the means of emotions from drug conditions vs. placebo conditions.

Note: The possible range is from 0-2, NOT from 0-1!

	Anger	Bored	Disgusted	Excited	Fearful	Grieved	Happy	Invigora
Drug	0.0	0.07	0.06	0.32	0.32	0.13	0.34	0.13
Placebo	0.0	0.08	0.05	0.30	0.31	0.12	0.33	0.12

Percents

Examine the percent of time the emotion term applipercent of time the emotion term applies or strongly applieses or strongly applies to the stimuli.

	Anger	Bored	Disgusted	Excited	Fearful	Grieved	Happy	Invi
Drug	0.0	4.623632	4.645968	20.348448	20.303775	9.157918	23.051150	9.69
Placebo	0.0	5.677947	3.497615	19.486714	19.327731	8.289802	22.711787	8.53

Summary Table

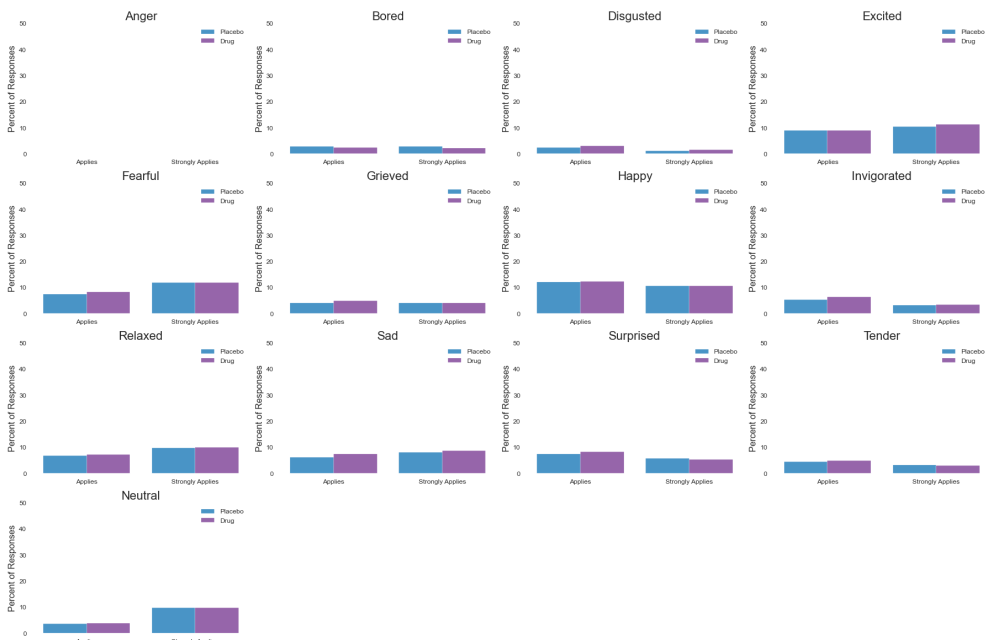
Next, we will make a summary table comparing the following features of the emotions:

- Sum of the emotion terms (for each emotion, summing the 0s, 1s, and 2s)
- Mean of the emotion terms (with a possible range of 0-2)
- The SD of the emotion terms
- The percent of people who said each term "Applies" to the stimuli
- The percent of people who said each term "Strongly applies" to the stimuli
- Whether the person doing the ratings ingested the drug or placebo
- The emotion in question

```
perceivedSummary shape: (26, 8)
26
```

	Emotion	DrugPlacebo	Mean	SD	variable	value
12	Neutral	Placebo	0.229843	0.608772	Applies	3.633886
27	Bored	Placebo	0.084261	0.363529	Strongly Applies	2.748126
23	Surprised	Drug	0.190083	0.511101	Applies	8.286799
50	Tender	Drug	0.111459	0.402012	Strongly Applies	3.127094
4	Fearful	Placebo	0.311379	0.671362	Applies	7.517602

Plot



Conclusions:

- There doesn't seem to be much of a difference between the placebo and drug conditions.
- I will not follow up with statistical tests.

Induced Emotion

Dataframe creation

Means

	Anger	Anxious	Bored	Disgusted	Excited	Fearful	Grieved	Happy
Drug	0.0	0.37	0.06	0.05	0.25	0.27	0.10	0.26
Placebo	0.0	0.30	0.06	0.06	0.24	0.27	0.08	0.25

2 rows × 23 columns

Percents

	Anger	Anxious	Bored	Disgusted	Excited	Fearful	Grieved	Ha
Drug	0.0	24.453552	3.825137	4.326047	17.167577	17.896175	7.331512	19.1
Placebo	0.0	20.168067	4.435107	4.154995	16.246499	18.300654	6.255836	18.3

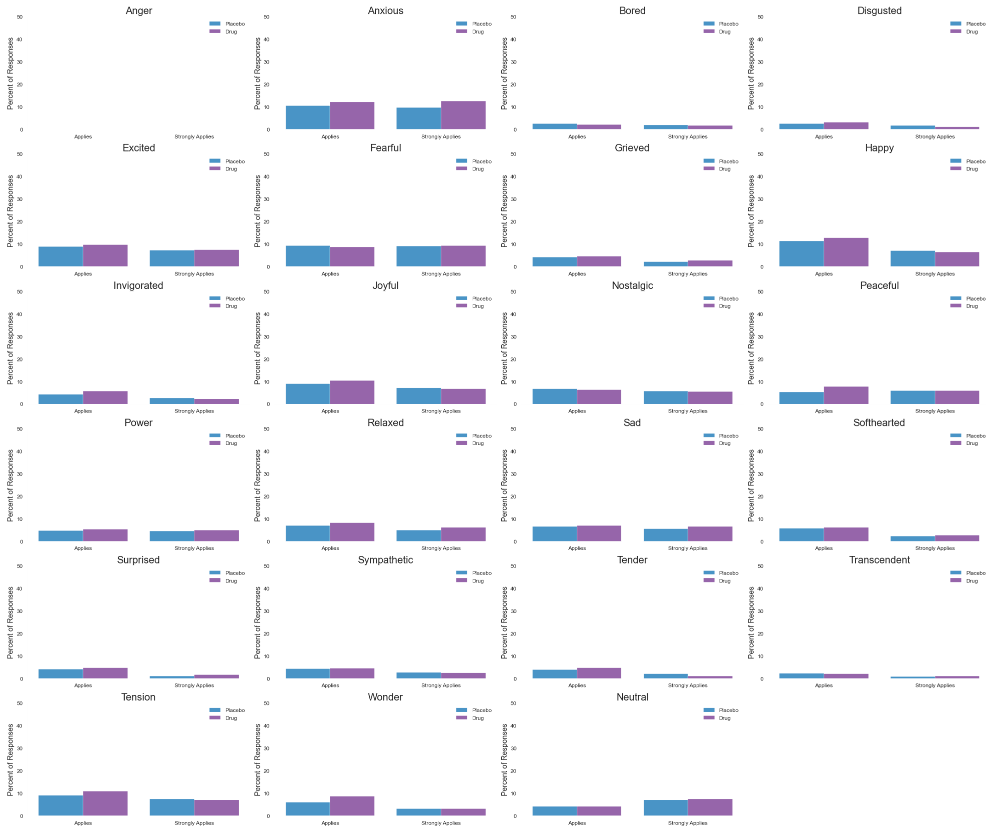
2 rows × 23 columns

Summary Table

```
inducedSummary shape: (46, 8)
46
```

	Emotion	DrugPlacebo	Mean	SD	variable	value
77	Invigorated	Drug	0.099271	0.363706	Strongly Applies	2.140255
40	Sympathetic	Drug	0.094718	0.367407	Applies	4.553734
38	Softhearted	Drug	0.114754	0.394196	Applies	6.102004
27	Excited	Drug	0.246357	0.578948	Applies	9.699454
20	Tension	Placebo	0.236228	0.571987	Applies	8.963585

Plot



Conclusions:

- There doesn't seem to be much of a difference between the placebo and drug conditions.
- I will not follow up with statistical tests.

4. Post-Hoc Tests