

Consonant & Dissonant Musical Sonorities: A Statistical Approach

**STATS/CSE 780
Project Report**

Konrad Swierczek - 001423065

4/17/23

Introduction

Considerable debate has been devoted to musical consonance and dissonance, the pleasantness of a collection of tones or sounds. The study of this phenomenon goes as far back as Pythagoras, but in the 20th and 21st centuries our understanding of acoustics, auditory perception, and mathematics led to sophisticated solutions using modelling and systematic experimental testing. Theories ranging from “roughness” (the interaction of tones on the basilar membrane of the inner ear) (Plomp and Levelt 1965; Kameoka and Kuriyagawa 1969a, 1969b; Hutchinson and Knopoff 1978, 1979; Mashinter 2006), “harmonicity” or “periodicity” (how similar is a collection of tones to a harmonic series) (Parncutt 1988; Milne, Laney, and Sharp 2016; Harrison and Pearce 2018), and familiarity as facilitated by inculturation Lahdelma, Eerola, and Armitage (2022) have been proposed as solutions to understanding this complex percept. Harrison and Pearce (2020) and Eerola and Lahdelma (2021) set out to compare these various frameworks of consonance and dissonance to human behavioural data; specifically, ratings of consonance or dissonance of western musical sounds. Both studies conclude that a combination of these models or frameworks may be necessary to explain the underlying mechanisms of consonance and dissonance. Here, a re-examination of the datasets explored in Harrison and Pearce (2020) and Eerola and Lahdelma (2021) will attempt to develop an improved composite model for predicting behavioural ratings using these base models. Further, the individual contributions of these base models will be examined in an effort to better understand what features are particularly relevant to the perception of consonance and dissonance. To achieve this goal, both supervised and unsupervised statistical learning methods will be used to train two models to predict human ratings: one using a subset of the features available, and one using all of the features available.¹

Anatomy of Consonance Dataset

Summary

The Anatomy of Consonance dataset (Eerola and Lahdelma 2021) is a collection of 9 datasets with 617 observations from human behaviour experiments on consonance and dissonance perception. It includes 33 feature variables, 7 metadata properties, and 3 normalized rating variables which

¹All materials and reproducible code used are available at <https://github.com/konradswierczek/STATS780>

correspond to the mean, standard deviation and standard error of multiple human participant ratings for a particular sound (observation). The features were selected by Harrison and Pearce (2020) and Eerola and Lahdelma (2021) from the literature based on relevance. Many of these are roughness, harmonicity, or familiarity models, in addition to spectral and numerosity features. The ratings have been normalized between 1 and 10 (see Figure 1), while the feature variables have predominantly continuous values on arbitrary scales based on each individual model. 28 of the feature variables are continuous, while one is categorical (`chord_size`) and was recoded as dummy variables for regression analysis. However, due to the repetitious nature of music in the contemporary western tuning system (Twelve Tone Equal Temperment), some of the features representing harmonicity are transposition invariant and have the same values regardless of their ordering or pitch height. As a results, features such as “`jl_12_tonal`” appear to be pseudo-categorical. They are not treated as categorical variables here as these models have been intended to be continuous in the larger context of sounds more generally.

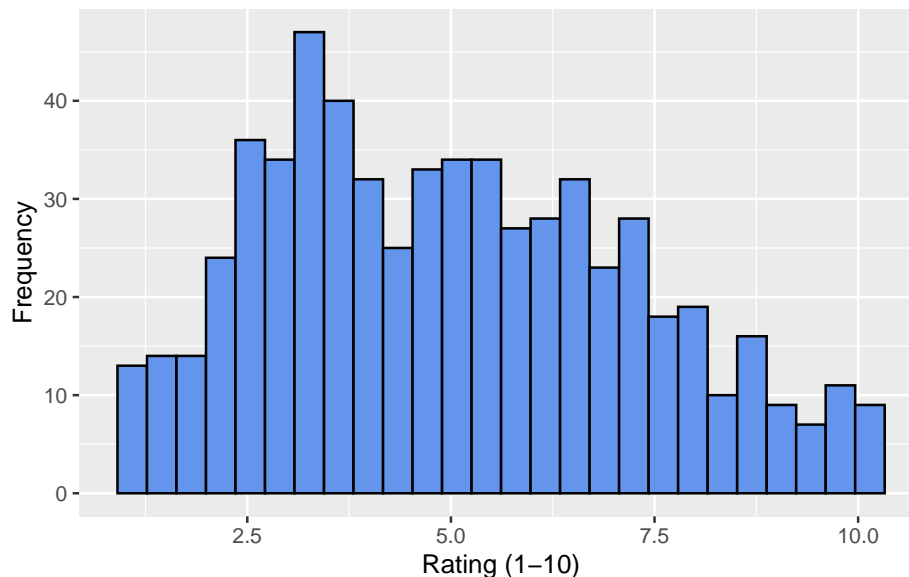


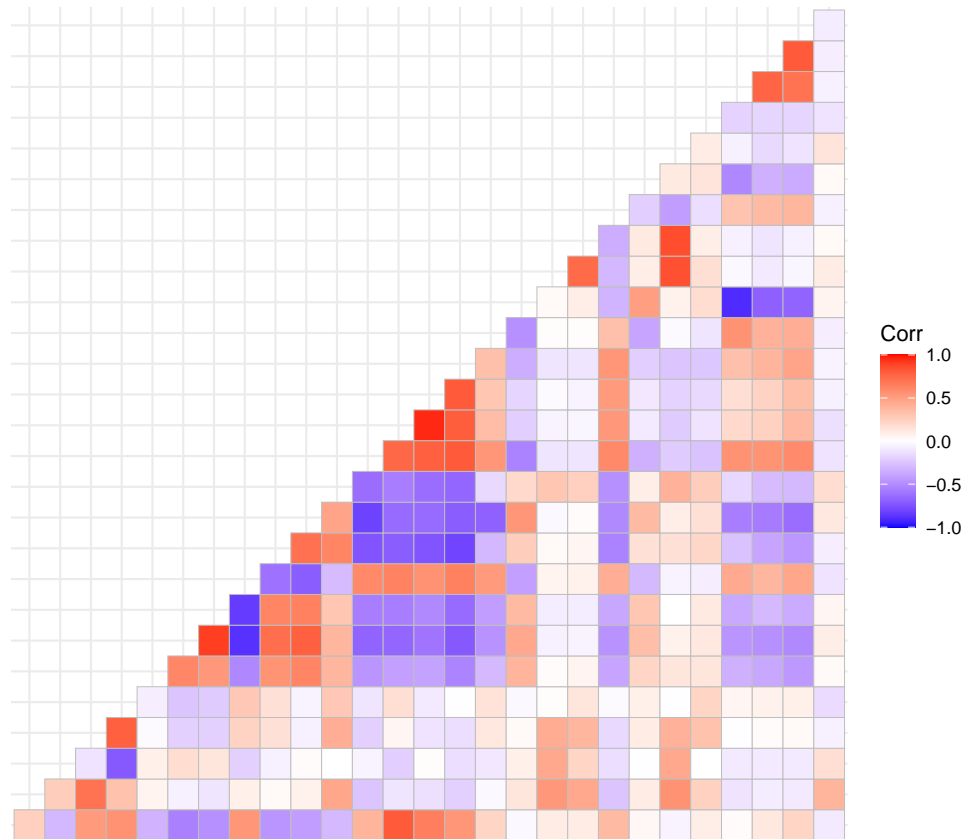
Figure 1: Distribution of ratings across 617 observations. Each observation is the average of all participants in the experiment.

Three feature variables were excluded due to lack of documentation on what they represent, and lack of use in previous analyses (“`TDL`”, “`TDL1`”, “`neg_log_prob`”). All feature variables were scaled before the analysis (see Figure 3) as their scales are generally arbitrary and vary across the models. No outliers were removed from the data since the rating values are already normalized and aggregated from a distribution, and the feature variables come from a predefined geometric

space. No missing values were found for either the rating values or any of the feature variables (see Figure 11 in Supplementary Materials). However, rating variability measures (Standard Deviation and Standard Error) were not included for a large portion of the observations. While this is not discussed in detail by Eerola and Lahdelma (2021), some of the observations may not be summaries of distributions. While is not relevant for the present study since it relies only on the ratings, one of the weaknesses of this dataset is the aggregated nature of the observations: individual participant ratings would provide a more robust sample.

Exploratory Analysis

Figure 3 shows each of the features plotted against ratings. Some of the features such as roughness and harmonicity/periodicity appear to have a linear relation with ratings, while many of the features describing the spectral profile appear to have a weaker association. Correlation analysis between the variables and rating (Figure 2) show that many of the variables have strong positive or negative correlations with the ratings (harmonicity is thought to predict consonance while roughness is thought to predict dissonance). Correlations between variables are highly variable, however high association between many of the variables make this dataset a good candidate for principle component analysis. Figure 1 shows the distribution of rating values, which appears to be slightly skewed or bimodal distribution consistent with the idea that chords are grouped into consonant or dissonant (the peak at a lower value is also consistent with the western music general preference for consonant sounds).



	Correlation Coefficient	Feature Variable
1	-0.727	CorpPop
2	-0.682	hutch_78_roughness
3	-0.626	jl_12_tonal
4	-0.607	CorpClas
5	-0.557	CorpJazz
6	-0.541	parn_88_root_ambig
7	-0.509	wang_13_roughness
8	-0.426	spflux
9	-0.379	vass_01_roughness
10	-0.363	seth_93_roughness
11	-0.130	chord_size
12	-0.046	sharpness
13	-0.025	pitchlow
14	-0.020	sproll

	Correlation Coefficient	Feature Variable
1	-0.010	spcentr
2	0.028	pitchhigh
3	0.035	pitchrange
4	0.040	register_N
5	0.088	register
6	0.218	spirreg
7	0.329	bowl_18_min_freq_dist
8	0.447	keyclar
9	0.464	parn_94_complex
10	0.471	gill_09_harmonicity
11	0.507	milne_13_harmonicity
12	0.570	har_18_harmonicity
13	0.701	har_19_corpus
14	0.743	stolz_15_periodicity

Figure 2: Correlation matrix of all feature variables (top) and correlation coefficients between feature variables and ratings (bottom).

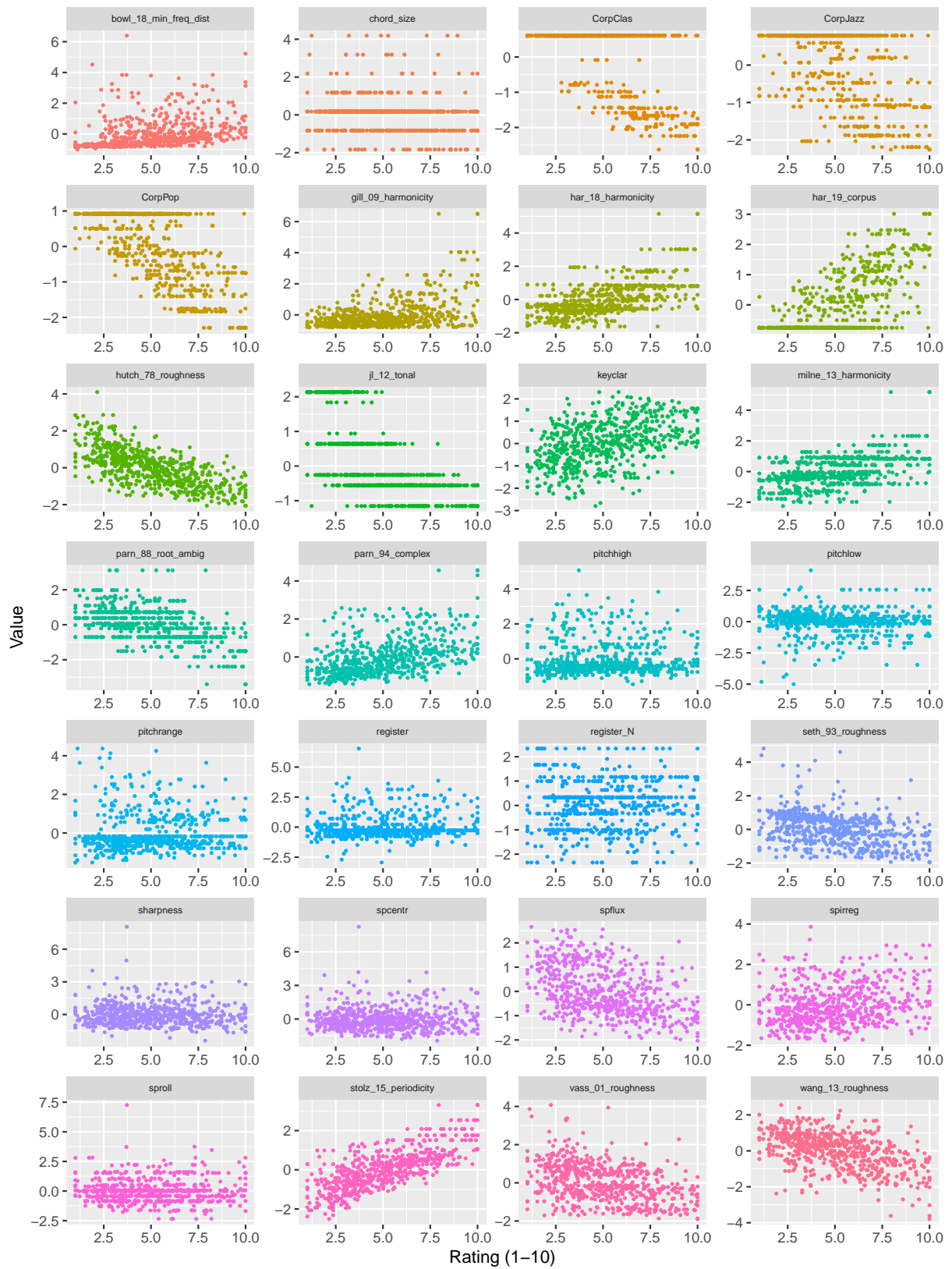


Figure 3: Scatterplots for each feature variable against rating values. Continuous variables have been scaled. Note that features such as `jl_12_tonal` appear to be categorical but are not intended to be when considering the broader context of sounds.

Methods

First, linear regression using forward stepwise subset selection was performed on the dataset to create a model that predicts rating while also selecting the best features out of the larger pool. Bayesian Information Criterion was used to select the size of the model since the state-of-the-art model only uses 3 components and thus a conservative selection criterion is appropriate to avoid overfitting. However, the above approach does not guarantee the subset of features selected will represent the most suitable group of features, much less an approximation of human perception since all of these features have in their own way been implicated in perceptual processes. A second linear regression will be performed on components generated using principal component analysis. The performance of these two different approaches was compared directly on a hold-out dataset, and also to the state-of-the-art model. These two approaches use different underlying assumptions to develop a model for consonance and dissonance: one uses as much information as possible, relying on the redundancy in the dataset, while the other focuses exclusively on the best subset of features and is therefore a simpler model. All analyses were performed in R (R Core Team 2020)

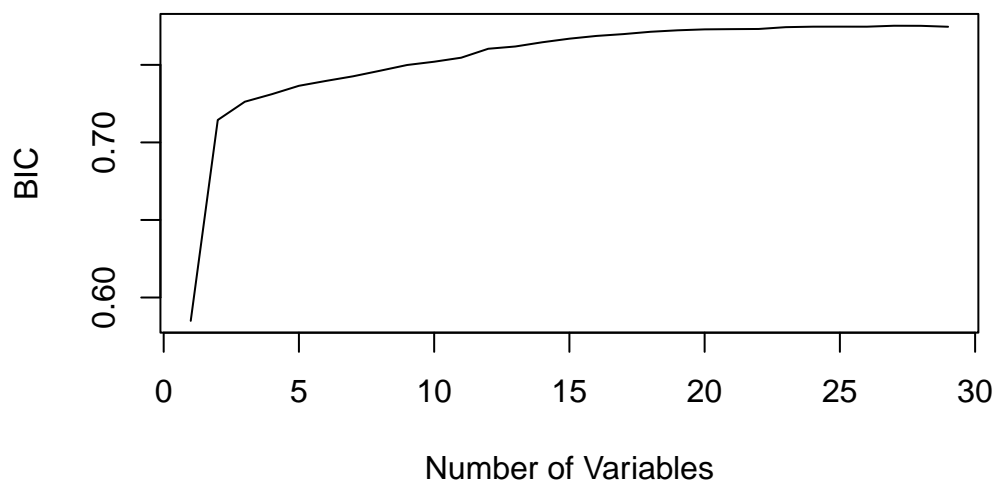


Figure 4: Bayesian Information Criterion for given amount of variables. 7 was selected as an ideal minimum based on this plot.

Results

Bayesian Information Criterion minimum value determined 5 features as the optimal model size. These 10 features and their regression coefficients are identified in Figure 12. Most notably, familiarity and harmonicity features were prominent in the selection process, with spectral features also being included. Figure 6 shows each feature in the model individually.

Principle component analysis show that 7 components explained over 80% of the variance in the dataset, with the first component only accounting for ~35% (Figure 5). Further analysis of the component loadings show that component 1 is predominant represented by roughness and harmonicity features, while subsequent components are represented by spectral features, familiarity, and harmonicity (Figure 7). A multiple linear regression model using the first 7 components was used to predict ratings: the first two dimensions are visualized in Figure 8.

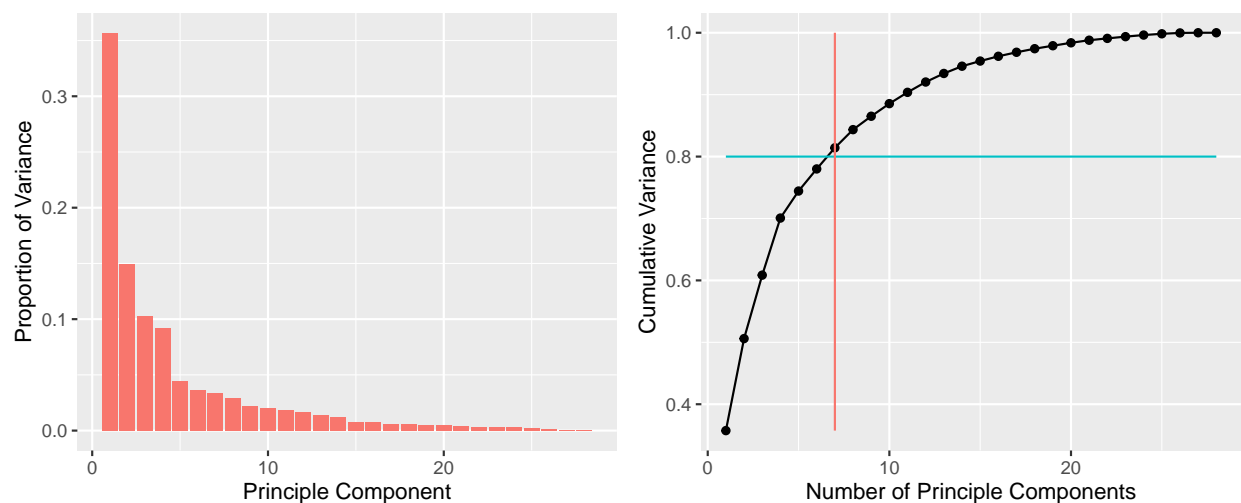


Figure 5: Two plots showing the proportion of variance explained by each principal component. The first plot shows the proportion of variance for each component while the second shows the cumulative proportion. 7 components were selected as they account for over 80% of the variance (see lines).

Accuracy of the models was evaluated by taking the absolute difference between the prediction and actual rating value across all observations in the hold-out set. Figure 10 shows the results of this test procedure, including predictions by the state-of-the-art model developed in Harrison and Pearce (2020). The feature selection model did not outperform the state-of-the-art model, but did have lower variability. However, the PCA model marginally outperformed the state-of-the-art model on both average rating error and variability.

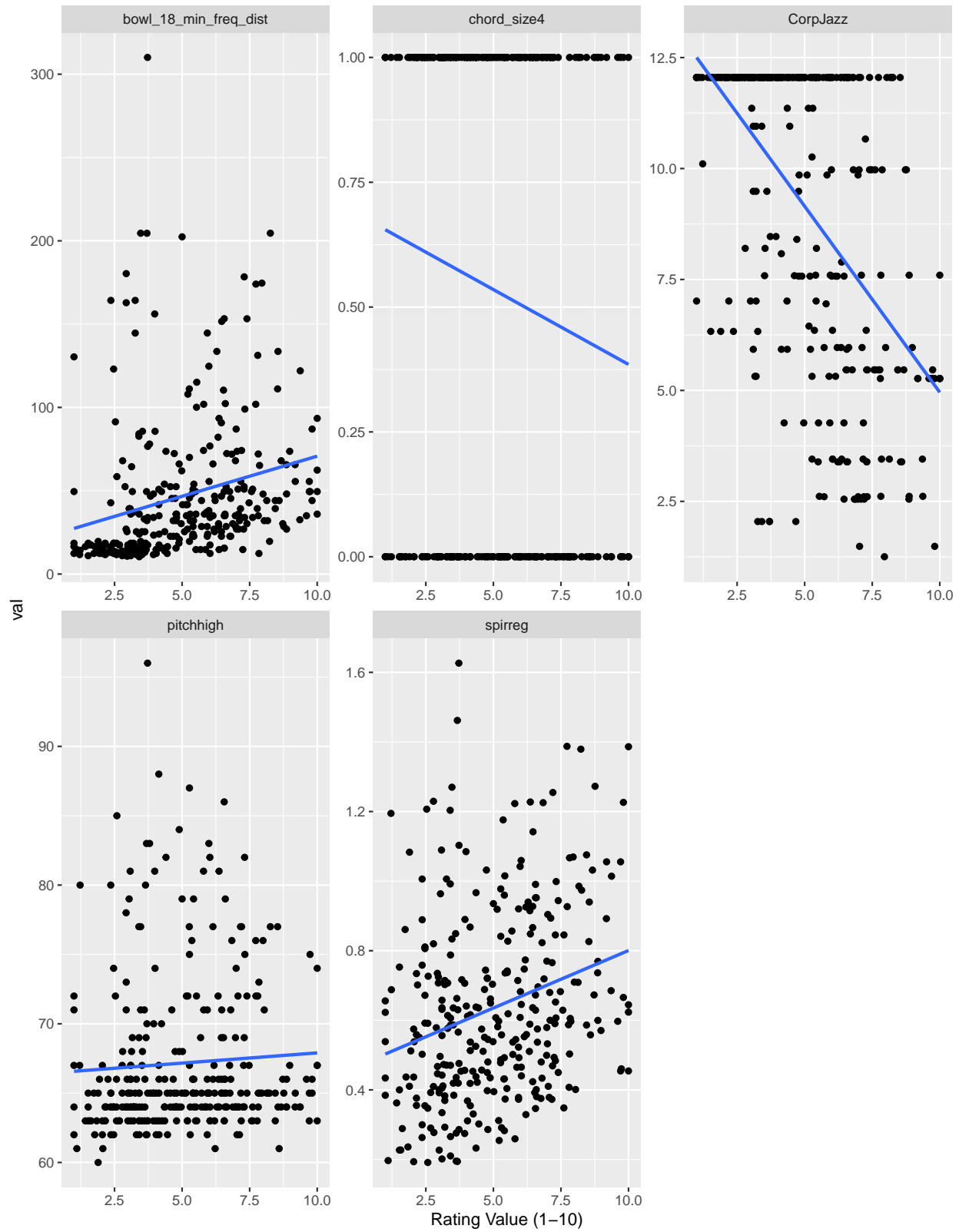


Figure 6: Regression lines for each variable selected against rating values.

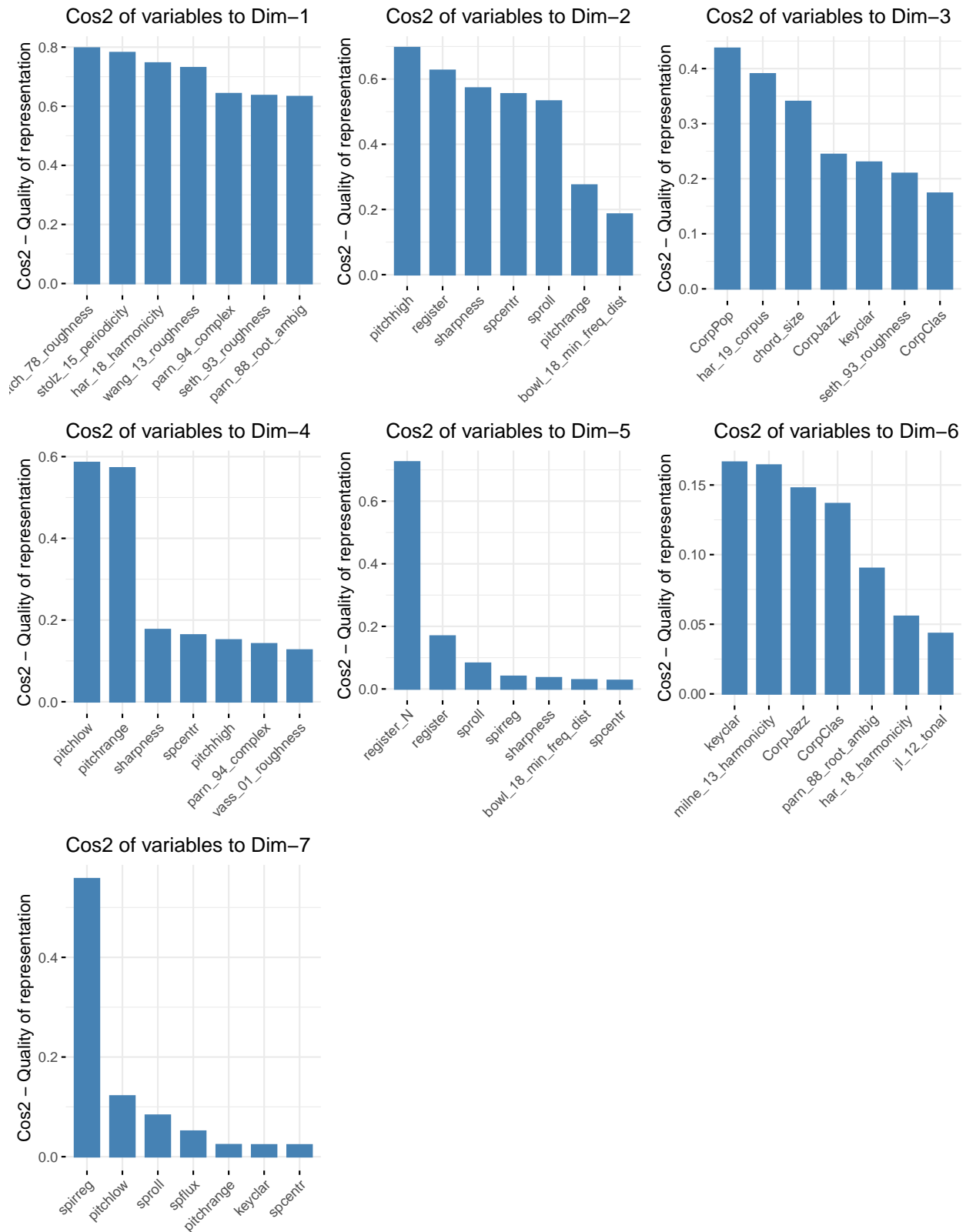


Figure 7: Cos2 of 7 most influential variable loadings for the first 7 principal components (notated as 'dim' here). Component 1 is predominantly driven by roughness and harmonicity, component 2 is driven by the spectral quality of the sound, and component 3 is driven by familiarity.

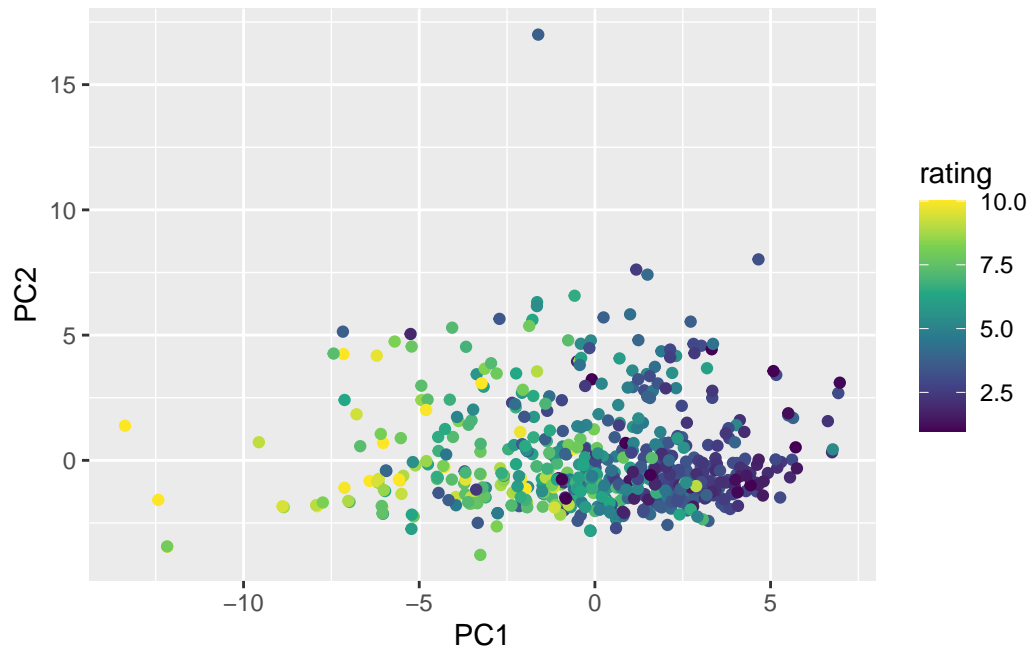


Figure 8: First two principal components plotted against ratings in colour. The first two components account for ~50% of the variance in the data.

Discussion

These results, particularly of the principle component analysis, build on the findings of Harrison and Pearce (2020) and Eerola and Lahdelma (2021), showing that harmonicity and roughness are the most significant contributors to the subjective perception of consonance and dissonance. Further, the spectral qualities of the sound as well as numeroisty appear to be significant contributors despite being fairly unassociated on their own and their exclusion in many of the models in this dataset. Future work may analyze the principle component loadings in more detail to remove features that may be redundant and not actively contributing to interpretability and prediction accuracy of the model. This dataset includes a variety of information, particularly in the case of the selection model, suffers from the curse of dimensionality which may account for the worsened accuracy. Other dimensionality reduction and statistical techniques such as commonality analysis may help with identifying redundancy in the data to overcome this issue. However, given the marginal improvement in accuracy of the PCA model and the increased complexity in comparison to the state-of-the-art model, these may not be practical improvements. The computational cost of the PCA model is significant considering all these models must be calculated for each simultaneous

collection of pitches in a piece of music (generally in the order of hundreds). While the cost is decreased for the feature selected model, these approaches are still impractical for uses in Music Information Retrieval and music classification/recommendation. The pseudo-categorical nature of harmonicity variables discussed above should be explored in more detail. If in a larger collection of sounds (an appropriate dataset is discussed in Eerola and Lahdelma (2021)) shows that western musical sounds are categorical for harmonicity, statistical approaches in modelling consonance and dissonance should be revised to reflect this. An additional challenge with the PCA model relates to reproduction, since prediction is based on principle components, not specific features. However, the loadings of each component can be used to weigh the original features for prediction of new data.

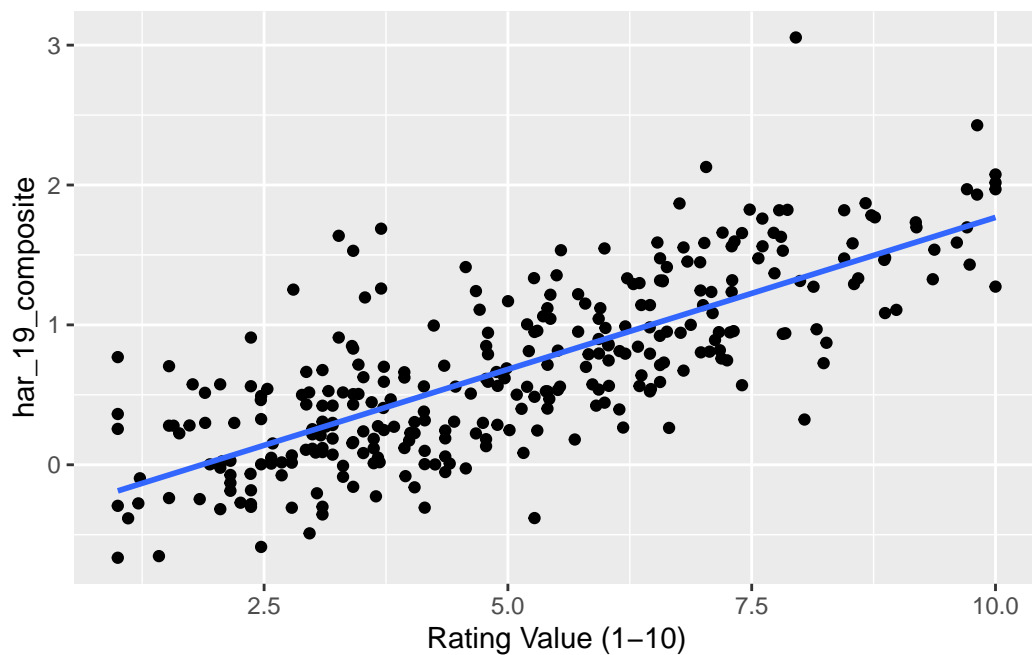


Figure 9: Current state-of-the-art model as developed in Harrison & Pearce 2019 (line), plotted against the test set (points). This model is a composite which uses harmonicity, roughness, and familiarity.

	Model	Mean Error	Standard Deviation
1	Stepwise Linear Reg.	1.46	1.15
2	PCA Linear Reg.	0.93	0.75
3	Harrison 2019	1.12	0.88

Figure 10: Model accuracy for three discussed models. The first two models were developed in this study, while Harrison 2019 is the current state-of-the-art model. Mean and standard deviation error was computed by taking the absolute difference between the prediction and actual value in the test set. Both models developed here slightly outperform Harrison 2019.

References

- Eerola, Tuomas, and Imre Lahdelma. 2021. "The Anatomy of Consonance/Dissonance: Evaluating Acoustic and Cultural Predictors Across Multiple Datasets with Chords." *Music & Science* 4: 20592043211030471.
- Harrison, Peter, and Marcus T Pearce. 2018. "An Energy-Based Generative Sequence Model for Testing Sensory Theories of Western Harmony." *arXiv Preprint arXiv:1807.00790*.
- . 2020. "Simultaneous Consonance in Music Perception and Composition." *Psychological Review* 127 (2): 216.
- Hutchinson, William, and Leon Knopoff. 1978. "The Acoustic Component of Western Consonance." *Journal of New Music Research* 7 (1): 1–29.
- . 1979. "The Significance of the Acoustic Component of Consonance in Western Triads." *Journal of Musicological Research* 3 (1-2): 5–22.
- Johnson-Laird, Phil N, Olivia E Kang, and Yuan Chang Leong. 2012. "On Musical Dissonance." *Music Perception: An Interdisciplinary Journal* 30 (1): 19–35.
- Kameoka, Akio, and Mamoru Kuriyagawa. 1969a. "Consonance Theory Part i: Consonance of Dyads." *The Journal of the Acoustical Society of America* 45 (6): 1451–59.
- . 1969b. "Consonance Theory Part II: Consonance of Complex Tones and Its Calculation Method." *The Journal of the Acoustical Society of America* 45 (6): 1460–69.
- Lahdelma, Imre, Tuomas Eerola, and James Armitage. 2022. "Is Harmonicity a Misnomer for Cultural Familiarity in Consonance Preferences?" *Frontiers in Psychology* 13: 124.
- Mashinter, Keith. 2006. "Calculating Sensory Dissonance: Some Discrepancies Arising from the Models of Kameoka & Kuriyagawa, and Hutchinson & Knopoff."
- McDermott, Josh H, Alan F Schultz, Eduardo A Undurraga, and Ricardo A Godoy. 2016. "Indifference to Dissonance in Native Amazonians Reveals Cultural Variation in Music Perception." *Nature* 535 (7613): 547–50.
- Milne, Andrew J, Robin Laney, and David B Sharp. 2016. "Testing a Spectral Model of Tonal Affinity with Microtonal Melodies and Inharmonic Spectra." *Musicae Scientiae* 20 (4): 465–94.
- Parncutt, Richard. 1988. "Revision of Terhardt's Psychoacoustical Model of the Root (s) of a Musical Chord." *Music Perception* 6 (1): 65–93.
- Parncutt, Richard, and Graham Hair. 2011. "Consonance and Dissonance in Music Theory and

- Psychology: Disentangling Dissonant Dichotomies.” *Journal of Interdisciplinary Music Studies* 5 (2).
- Plomp, Reinier, and Willem Johannes Maria Levelt. 1965. “Tonal Consonance and Critical Bandwidth.” *The Journal of the Acoustical Society of America* 38 (4): 548–60.
- R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Zajonc, Robert B. 2001. “Mere Exposure: A Gateway to the Subliminal.” *Current Directions in Psychological Science* 10 (6): 224–28.

Supplementary Materials

	Category	Missing Values
1	Ratings	0
2	Features	0
3	Rating Variability	-254

Figure 11: Missing values for each data type.

	Feature	Coefficient
1	(Intercept)	8.70
2	chord_size4	0.25
3	pitchhigh	-0.03
4	bowl_18_min_freq_dist	0.02
5	spirreg	0.51
6	CorpJazz	-0.30

Figure 12: Regression coefficients for forward-stepwise subset selection.

STATS 780 Project Report

17

```

# knitr code chunk configuration
knitr::opts_chunk$set(echo=FALSE, warning=FALSE, message=FALSE)

knitr::opts_chunk$set(fig.pos = "H", out.extra = "")

#####

# Imports

packages <- c("tidyverse", "dplyr", "readr", "cluster", "factoextra", "ggplot2",
              "knitr", "ggcorrplot", "tidyr", "tree", "ggpubr", "gridExtra",
              "leaps")

lapply(packages, library, character.only = TRUE)

#####

# Set random seed for reproducible samples.

set.seed(75)

# Anatomy of Consonance Dataset: Tuomas Eerola
# https://github.com/tuomaseerola/anatomy-of-consonance
# Retrieve data from GitHub repo.

temp <- paste(tempfile(), ".zip", sep = "")

options(timeout = 60 * 10)

"https://github.com/tuomaseerola/DCD/archive/refs/heads/master.zip" %>%
download.file(temp)

experiment1 <- unz(temp, "DCD-master/data/DCD_predictors.csv") %>%
  read_csv()

"github.com/tuomaseerola/anatomy-of-consonance/archive/refs/heads/main.zip" %>%
download.file(temp)

link <- "anatomy-of-consonance-main/data/experiment2_data.csv"
experiment2 <- unz(temp, link) %>%
  read_csv()

link <- "anatomy-of-consonance-main/data/experiment3_data.csv"
experiment3 <- unz(temp, link) %>%
  read_csv()

# Save data locally.

save(experiment3, file = "data/experiment3.RData")
save(experiment2, file = "data/experiment2.RData")
save(experiment1, file = "data/experiment1.RData")

# Load data from local files.

```

```
load("data/experiment3.RData")
# Specify feature/model variables.
vars <- c(7, 11:27, 29:34, 37:39, 42)
# Split observations into training and testing sets: 50/50 split.
train <- sample(nrow(experiment3), 0.5 * nrow(experiment3))
# Distribution of ratings
binwidth <- function(col) ((max(col)-min(col))/sqrt(length(col)))
ggplot(experiment3, aes(x = rating)) +
  geom_histogram(binwidth = binwidth(experiment3$rating),
                 colour = "black", fill = "cornflowerblue") +
  xlab("Rating (1-10)") +
  ylab("Frequency")
# Correlations between feature variables and rating values
cors <- tibble("Correlation Coefficient" =
  round(cor(experiment3[, colnames(experiment3[, vars])],
            experiment3$rating), 3)[, 1],
  "Feature Variable" = colnames(experiment3[, vars])) %>%
arrange(`Correlation Coefficient`)
ggarrange(
# Correlations between feature variables
  round(cor(experiment3[, vars]), 3) %>%
  ggcorrplot(insig = "blank", type = "lower") +
  theme(axis.text.x=element_blank(),
        axis.ticks.x=element_blank(),
        axis.text.y=element_blank(),
        axis.ticks.y=element_blank()),
  ggarrange(
    ggtexttable(cors[1:14, ]),
    ggtexttable(cors[15:28, ]),
    ncol = 2),
  ncol = 1
```

```
)  
# Scaled feature variables and ratings.  
experiment3[, c(3, vars)] %>%  
mutate_at(2:29, ~(scale(.) %>% as.vector)) %>%  
gather(feature, val, -rating) %>%  
ggplot(aes(x = rating, y = val, colour = feature)) +  
  geom_point(size = 0.5) +  
  facet_wrap(~feature, scales = "free", ncol = 4) +  
  ylab("Value") +  
  xlab("Rating (1-10)") +  
  theme(strip.text = element_text(size = 6),  
    legend.position = "none",  
    plot.margin = margin(0, 0.5, 0, 0.5, "cm"))  
# Modifying categorical variable before analysis  
lm_data <- experiment3 %>%  
mutate(chord_size = as.factor(chord_size))  
# Forward stepwise subset selection  
regfit.full <- regsubsets(rating ~ ., data = lm_data[train, c(3, vars)],  
  nvmax = 28, method = "forward")  
# BIC for model complexity.  
plot(  
  summary(regfit.full)$adjr2,  
  xlab = "Number of Variables",  
  ylab = "BIC",  
  type = "l")  
bic_min <- which.min(summary(regfit.full)$bic)  
# Find out coefficients in the best model.  
coefi <- coef(regfit.full, id = which.min(summary(regfit.full)$bic))  
# Principle component analysis  
pca <- prcomp(experiment3[, vars], scale = TRUE)  
pca_df <- tibble(`prop` = round(pca$sdev^2/sum(pca$sdev^2),3), `pc` = c(1:28),
```

```

    `cum` = cumsum(pca$sdev^2 / sum(pca$sdev^2)))
ggarrange(
# Variance explained by each principle component.
  ggplot(pca_df, aes(x = pc, y = prop)) +
    geom_bar(stat="identity", aes(fill = "#6495ed")) +
    xlab("Principle Component") +
    ylab("Proportion of Variance") +
    theme(legend.position = "none"),
# Cumulative variance explained for n principle components.
  ggplot(pca_df, aes(x = pc, y = cum, group = 1)) +
    geom_line() +
    geom_point() +
    geom_line(aes(y = 0.8, colour = "red")) +
    geom_line(aes(x = 7, colour = "black")) +
    xlab("Number of Principle Components") +
    ylab("Cumulative Variance") +
    theme(legend.position = "none"),
  ncol = 2
)

# Mutating dataset for dummy variables in regression.
lm_data2 <- lm_data %>%
mutate(chord_size2 = ifelse(experiment3$chord_size == 2, 1, 0)) %>%
mutate(chord_size3 = ifelse(experiment3$chord_size == 3, 1, 0)) %>%
mutate(chord_size4 = ifelse(experiment3$chord_size == 4, 1, 0)) %>%
mutate(chord_size5 = ifelse(experiment3$chord_size == 5, 1, 0)) %>%
mutate(chord_size6 = ifelse(experiment3$chord_size == 6, 1, 0)) %>%
mutate(chord_size7 = ifelse(experiment3$chord_size == 7, 1, 0)) %>%
mutate(chord_size8 = ifelse(experiment3$chord_size == 8, 1, 0))
# Plotting each variable in multiple linear regression analysis.
lm_data2[-train, c("rating", names(coefi)[-1])] %>%
gather(`feature`, `val`, -`rating`) %>%

```

```
ggplot(aes(x = rating, y = val)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  xlab("Rating Value (1-10)") +
  facet_wrap(~feature, scales = "free")

# PCA Variable loadings: top 7 variables for top 7 components.
grid.arrange(
  fviz_cos2(pca, choice = "var", axes = 1, top = 7),
  fviz_cos2(pca, choice = "var", axes = 2, top = 7),
  fviz_cos2(pca, choice = "var", axes = 3, top = 7),
  fviz_cos2(pca, choice = "var", axes = 4, top = 7),
  fviz_cos2(pca, choice = "var", axes = 5, top = 7),
  fviz_cos2(pca, choice = "var", axes = 6, top = 7),
  fviz_cos2(pca, choice = "var", axes = 7, top = 7),
  ncol = 3
)

# PCA linear regression: first two components plotted with ratings in colour.
pca_data <- as_tibble(pca$x) %>%
  add_column(rating = experiment3$rating)
lm.fit_2 <- lm(rating ~ ., data = pca_data[train, ])
as_tibble(pca$x) %>%
  add_column(rating = experiment3$rating) %>%
  ggplot(size = 0.5, aes(colour = rating, x = PC1, y = PC2)) +
    geom_point() +
    scale_colour_viridis_c() +
    guides(size = FALSE)

# Predictions on hold-out set for each model.
test.mat <- model.matrix(rating ~ ., data = lm_data2[-train, c(3, vars, 45:51)])
pred <- test.mat[, names(coefi)] %*% coefi
pred_2 <- predict(lm.fit_2, pca_data[-train, ])
lm.sota <- lm(rating ~ har_19_composite, data = experiment3[, c(3, 28)])
```

```

pred_3 <- predict(lm.sota, experiment3[-train, ])
# Harrison 2019 model against ratings.
experiment3[-train, c(3, 28)] %>%
ggplot(aes(x = rating, y = har_19_composite)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  xlab("Rating Value (1-10)")
# Prediction accuracy of each model
ggtexttable(
  tibble(`Model` = c("Stepwise Linear Reg.", "PCA Linear Reg.", "Harrison 2019"),
    `Mean Error` = c(round(mean(abs(experiment3$rating[-train] - pred)), 2),
                      round(mean(abs(experiment3$rating[-train] - pred_2)), 2),
                      round(mean(abs(experiment3$rating[-train] - pred_3)), 2)
    ),
    `Standard Deviation` = c(round(sd(abs(experiment3$rating[-train] -
                                      pred)), 2),
                              round(sd(abs(experiment3$rating[-train] -
                                      pred_2)), 2),
                              round(sd(abs(experiment3$rating[-train] -
                                      pred_3)), 2)
    ))) +
  theme(plot.margin = margin(-5, 0, 0, 0, "cm"))
# Missing value analysis for ratings and feature variables
ggtexttable(tibble(
  `Category` = c("Ratings", "Features", "Rating Variability"),
  `Missing Values` = c(table(is.na(experiment3[, 7]))["FALSE"] -
                        nrow(experiment3[, 7])
    , table(is.na(experiment3[, vars]))["FALSE"] - prod(dim(experiment3[, vars])),
    table(is.na(experiment3[, 4:5]))["FALSE"] - prod(dim(experiment3[, 4:5]))
  )
))

```

```
# Subset Selection feature coefficients.  
ggtexttable(tibble(Feature = names(coefi), Coefficient = round(coefi, 2))) +  
theme(plot.margin = margin(-2, 0, 0, 0, "cm"))
```