Goals

- To find and interpret clusterings of Bach chorales => allow musicians to connect musical and contextual features.
- To determine features (independent of relative positions) that distinguish individual SATB voices in Bach chorales.

Methodology

- Scraped scores encoded in the Humdrum file format, using kern representation, from CCARH
- Used MIT’s music21 framework to extract features from 183 Bach chorales and their 183 * 4 individual voices

Example features:
- Duration of melodic arc = Average number of notes separating peaks and troughs
- Average melodic interval = Average # of semitones between consecutive notes
- Melodic tempi = average # of notes per second and average note duration

Categorized features into 8 buckets:
- Dissonance
- Prevalence of major vs. minor chords
- Note range
- Average melodic interval
- Chromaticism vs diatonicism
- Melodic contour
- Melodic tempi
- Phrase structure
- Amount of arpeggiation

Clustering Voices

Algorithm: K-Means, K implicitly defined to be 4 (SATB)

Error rate: 0.321

Results: Bass line most distinct, soprano next

Voice Prediction

Algorithm: Softmax / Multinomial Logistic Regression, CV = 10

10-fold CV Error Rate = 0.255

p-Values < 0.05:

- Average Melodic Interval (largest for Bass)
- Range of Individual Voices (largest for Bass)
- Number of Repeated Notes (most for inner voices)

Interpretation (a basic model of phrasing): To highlight independence of lines, performers can emphasize
- Highest and lowest notes
- Transitions between notes

Future Work

In future, can apply this methodology
- with different musical features (reflecting diff. styles of music theory)
- with different composers and types of compositions
In addition, analyze our data with respect to contextual features (e.g., map clusters to date of composition, intended occasion of performance, etc.)