Assignment 4: Practicalities; Designing your Final Projects!

Group member names (please self-assign to a group in Canvas!!): Brittney Juliet Allen, Karl Hiner, Sile Yin

Item #1

Within your final project group, consult/review both the (a) Corpus spreadsheet and (b) the Hypotheses.ppt file (especially the definition of a hypothesis) and the "5-point guide to generating a hypothesis". List up to 3 research questions that you think could potentially be answered using symbolic musical data, and for which either: (a) already has an existing corpus, or (b) you can easily see a concrete plan for curating a corpus. **(If "b" then you must fill out Item #4)**

1. Is the "I-V-IV-vi" chord progression in popular music declining in popularity?
2. How original are jazz improvisations?
3. How does the chord complexity evolve in classical music (from Baroque era to Romantic)?

Item #2

Given the available data, and the 5 points to consider in the "5-point guide", come up with one *testable hypothesis* for each of the above research questions (in the same order) that you think is tangible as a final project for this class. (Note: You don't have to worry if it's *really* tangible or not for this class especially since you may not be familiar with statistical techniques—I will help you, but do try anyway! However, you should be able to formulate hypotheses.)

1. The "I-V-IV-vi" chord progression in popular music *is* declining in popularity.
2. Jazz improvisors utilize patterns more than they realize.
3. Classical era works have simplified chords comparing to Baroque era and Romantic era works.

Item #3

Given your hypotheses, for each one, overview the following:
  (a) What data would you need? What is your control(s)?
  (b) How would you create a sample that is representative? Or will you simply use a 'convenience sample'? (This is fine, so long as you are aware of the limitations)
  (c) What concepts or statements will need to be operationalized? (You do not have to write out your operationalizations, but thinking through them may illuminate challenges.)

1.
a) Data: Popular music chord progressions over time, going back some years into recent (last year/two) years. Controls: There is no control group, since there is no intervention.
b) Ideally, there would be a popular music corpus with a statistically significant number of songs relative to all released popular music. In this case, we would have to settle for a small representative sample. The data would ideally be in roman numeral format already, or I could

derive roman numeral chords from MIDI files. The "RS200 Pop / Rock corpus" provided in the corpus spreadsheet, for example, would be a good convenience sample.

c) What is a "chord progression"? When does a progression begin/end, are they all the same length, etc. What is "popular music"?

2.

a) Data: a corpus of jazz solos. As progressions aren't of interest, only symbolic melodic data is required. There is no control for this set of data as this study does not contain intervention.

b) For the purposes of this class, a convenience sample will be used. Ideally, the stratified sampling method, whereby a random selection of improvisations from each era of jazz (big band, bebop, early jazz, etc.) would be compiled compiled, however, due to time restrictions, the Weimar Jazz Database seems to be the most viable option.

c) Operationalized concepts:
- Original - what does that mean, and what is the measure?
- Improvisation - what constitutes an improvisation?

3.

a) Data: Chord annotation for classical music works overtime. Controls: There is no control group, since there is no intervention.

b) We'll use all the possible data from archives of classical music e.g., Yale Classical Archives if we can, otherwise we want to keep the distribution of genres when we sample – quota sampling.

c) Chords, chord complexity, Baroque/Classical/Romantic era