Assignment #1
Names: Karl Hiner

Part One: From Question to Hypothesis

With your partner, share your "brainstorming" ideas for interesting research questions that you may want to work towards as a part of your final projects (don't worry, you don't have to commit to something just yet!)

Ideally, research should focus on questions of the highest importance. The conscientious researcher always tries to answer the most important questions first.

Unfortunately, the most important questions cannot always be answered given our limited resources and limited understanding. Instead, we might focus on answering questions for which we have the skills and resources to answer. Sometimes, these questions might seem trivial or uninteresting.

The best research endeavors to connect good "top-down" questions with good "bottom-up" research resources. Empirical research methods can provide helpful resources for answering questions about music, but these methods should not dictate our research agendas. In learning how to use do empirical research, it is equally important to learn to ask creative questions about music. The purpose of this task is to get you thinking about larger issues in computational analysis for music, and to become excited about carrying out this research. As a researcher, what do you *really* want to know?

Between the two of you, come up with at least THREE and up to five questions that you think are really interesting and list them below.

1. Can bird songs be found in popular music? (In some known cases, composers or musicians have done this intentionally. But what about unintentionally? To what extent do we need to stretch our similarity measure to get a qualified "yes"?)
2. What can instances of ML-generated music tell us about the corpus it's trained on? (By analogy, ML-generated images can tell us useful things about the dataset it's trained on, e.g. its gender or racial skewness. Note that this is an explorative question.)
3. Are the melodic rhythms of breath-controlled instruments (including singing) different from non-breath-controlled instruments? (Explorative interest here is to characterize the impact of natural breathing patterns on how we compose music more broadly.)
4. Is the I–V–vi–IV progression in pop music becoming less popular? (I heard a podcast recently with Rick Beato, who made the popular YouTube video "The Four Chords That Killed POP Music!", claiming that popular music may be slowly moving on from this ubiquitous chord progression.)
5.

Of your questions, pick *one* that you both agree is the most interesting of the questions that seem *feasible* (acknowledging that you may not know what skills you will acquire yet! Use your best judgement.) This will be the question that you use to complete the remainder of this assignment.

Is the I–V–vi–IV progression becoming less popular?

From your selection, work towards brainstorming (theoretically) *testable hypotheses* that stem from the question. A good method for this is to work from the question towards a theory or conjecture that stems from the question. Recall that a conjecture is a speculative claim (i.e., "probably…", while a theory is an explanatory framework (i.e., "because…"). None of this has to be based in any kind of fact, you are free to draw on your own intuition at this stage.

Hypothesis: Yes, the I–V–vi–IV progression is becoming less popular.

Part Two: Sampling

1. Given your question of interest, what is the *population* of interest?

   This question asks about the population of *chord progressions in pop music* over time. To answer the question, we would need to compare chord progressions across a representative sample of pop music, at different periods of time.

2. Without worrying about what is possible or not, how would you obtain a sample that is *representative* of this population? In other words, what would you sample, and which sampling method would you use? *N.B.: before you think about sampling you may have to operationalize a bit. Make notes.*

   First, I would choose a set of time periods to compare across. The chosen time periods should be equal in duration, and if more than two periods are selected, the time between each period should be roughly the same. It would probably be most illuminating to choose a period of one year, with *no gap* between each period.

   Then, for each time period, I would obtain a set of pop songs that were produced (realistically, *released*), within the time period. The notion of "pop songs" needs operationalizing, but I tentatively think Top-N lists would be effective sources.) The set of pop songs should be statistically significant with respect to the total number of pop songs produced over the period. Finally, I would find every chord progression (needs operationalizing) across all songs in each time period, to produce the final population for each period.

   At this stage, I would inspect the distribution of unique chord progressions *across all time periods*, and likely end up discarding some number of very rare progressions, considering them as noise. (In practice I would want to verify what may be due to the operational difficulty of categorizing into "chord progressions" when making this cutoff/cleanup determination, as opposed to actually being rare progressions.)

3. Explain your rationale for the sample selection and method. (Note you can do this "inline" above if it is easier)

   My main rationale is to be careful to use the same procedure for each period of comparison, trying to make the only characteristic varying between the samples in each period be as representative as possible of the characteristics that *actually vary* between the true (real-world) population of chord progressions.

   One more thing I would note is that all sampling done from each time period should be done randomly. That is, as much as possible, the results of any experiment should be independent of the incidental order of the derived chord-progression dataset.

Part Three: From Hypothesis to Operationalization

1. Given your hypothesis that your group came up with, what terms will need to be defined operationally? Write those below. Think through how your decisions might affect the outcomes to your question. Take notes to recall why you may have "thrown out" some option or idea (you may forget later and have to do it all over again!)

   * Pop music
   * Chord progression
   * I–V–vi–IV progression
   * Popular (Could reword the question to "… less frequent"? Still, this notion needs explanation even if it's a straightforward frequency comparison.)

2. Write down all necessary operational definitions below (you should have at least 1-2 and possibly many more.)

    * **Pop music:** Any song in the Billboard Top-100 for the year in question
    * **Chord progression:** Ooph.. help? I'll throw out a bad definition to beat: Determine a dominant key for each song (really want each *section*, which is equally hard to define. Also, often there's not a single dominant key). Derive the sequence of chords from the song. Translate each chord sequence to the number relative to the key determined in the first step. Find every "I" (major root) chord with at least 3 non-"I" chords following it, and consider that group of 4 chords a "chord progression."
    * **I–V–vi–IV progression:** Using the definition of "chord progression" above, this is the sequence of chords: Major-1, Major-5, Minor-6, Major-4, within the dominant key in which the sequence resides.
    * ***Popular* chord progression:** Relative frequency of the progression with respect to all chord progressions within the time period in question.