# Voice-Leading and Harmonic Progressions in the Beethoven String Quartets: a Preliminary Report using the Annotated Beethoven Corpus (ABC)

Tim Ruszala

I am greatly indebted to the following people: Hannah Semmelhack, who bore my company and my stress; Luke Soucy, my wonderful partner whose support was as important to this project as my data; and Dmitri Tymoczko, my tireless, quickly responding, and frankly inspirational advisor who, despite the fact that I did not know python or music21 and hadn't touched numbers in years, inspired me to tackle this project.

# Introduction

This report is an account of tools I developed to digitally analyze the Beethoven string quarrtets, as well as a collection of preliminary findings on Beethoven's harmonic and contrapuntal practices in the quartets. In light of Beethoven's extraordinary impact on the musical, harmonic, and contrapuntal styles of later composers, the quantitative knowledge of Beethoven's own procedures is clearly of importance. This paper asks the following questions: With what relative requencies does each roman numeral appear? How often does Beethoven resolve tritones non-conventionally? How often does Beethoven include the leading tone as a passing-tone when moving from IV-I, mixing dominant and predominant functionality? My results suggest that voice-leading rules derived from Bach chorales, as is normally the case in introductory music theory classes, may not be the best model for voice-leadings in Beethoven. My paper concludes by proposing further paths of inquiry using the tools developed here.

This project is enabled by two sources: first, the existence of the *music21* Python library, and secondly the recent (2018) online publication of the Annotated Beethoven Corpus (ABC) by Neuwirth et al.. The *music21* library is a set of tools for computer-aided musicology, including ready-made digital representations of basic musical elements and functions with which to parse them. Developed in 2008 and headed by Michael Cuthbert of MIT, *music21* enables the relatively easy collection of statistics on large corpuses of music, when otherwise such data-collection would be unreasonably time consuming.[1] The ABC dataset contains human-inputted roman numeral analyses of all of Beethoven's string quartets, done by Markus Neuwirth and Daniel Harasim at first individually, and then cross-checked.[2] With these analyses in the romantext format developed by Dmitri Tymoczko, it becomes possible to perform many sorts of quantitative analyses, from determining the average frequency with which Beethoven uses a given chord, to writing a script which can identify non-chord tones. In its current state, however, the ABC leaves something to be desired, with frequent inexpert analyses notated. With a grant from the Center for Digital Humanities at Princeton, this summer under the guidance of Dmitri Tymoczko I plan to clean the data further, and hope to collect new statistics from the improved data by next September.

---

[1] See the music21 "About" page for more: https://web.mit.edu/music21/doc/about/about.html#about

[2] Neuwirth, Markus, Daniel Harasim, Fabian C. Moss, and Martin Rohrmeier. "The Annotated Beethoven Corpus (ABC): A Dataset of Harmonic Analyses of All Beethoven String Quartets." Frontiers in Digital Humanities 5 (2018). doi:10.3389/fdigh.2018.00016. https://www.frontiersin.org/articles/10.3389/fdigh.2018.00016/full

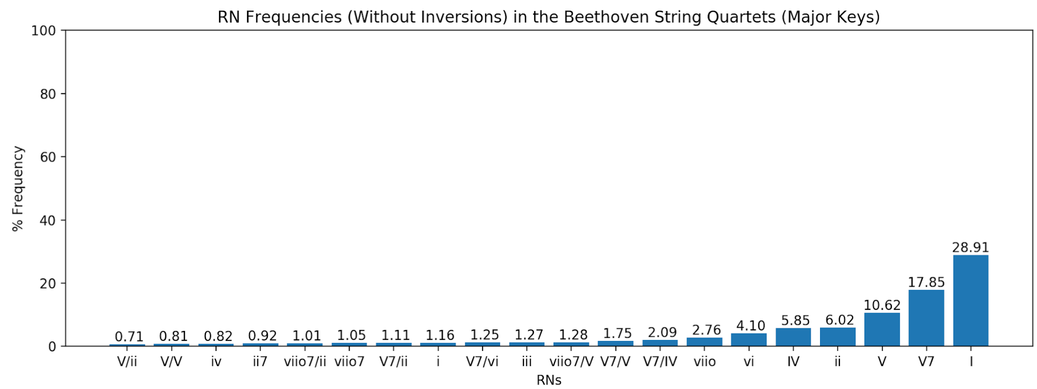
# Results

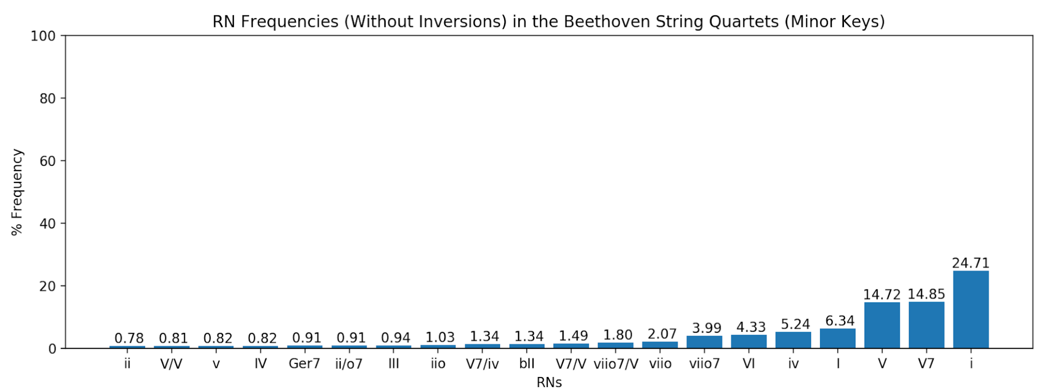
Here I discuss the tentative answers I have found to a series of questions about Beethoven's harmonic and voice-leading procedures.

**Roman Numeral Statistics**

The data presented in this section comes entirely from the analyses of the ABC. This data will likely change significantly after the ABC undergoes significant cleaning over summer 2019, but it nonetheless gives us broad insight into Beethoven's harmonic toolkit.

Below are two graphs presenting the relative percent frequencies of the twenty most common roman numerals in the string quartets, abstracting away from inversions, and separated between major and minor tonal areas:



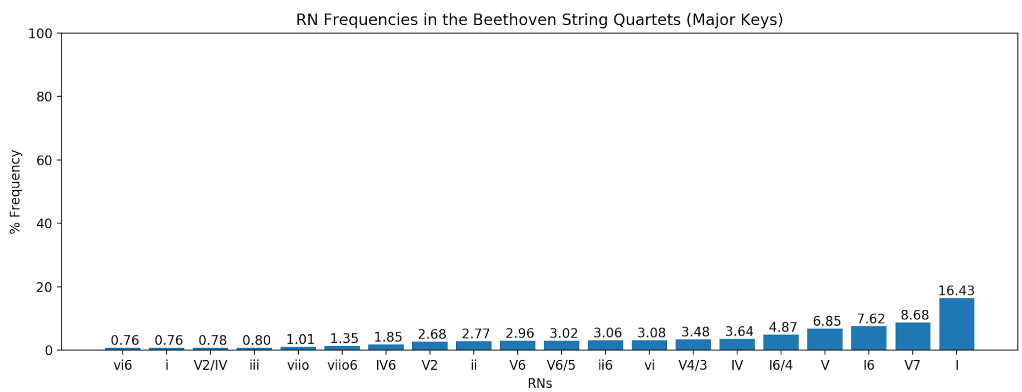

A few items of note:

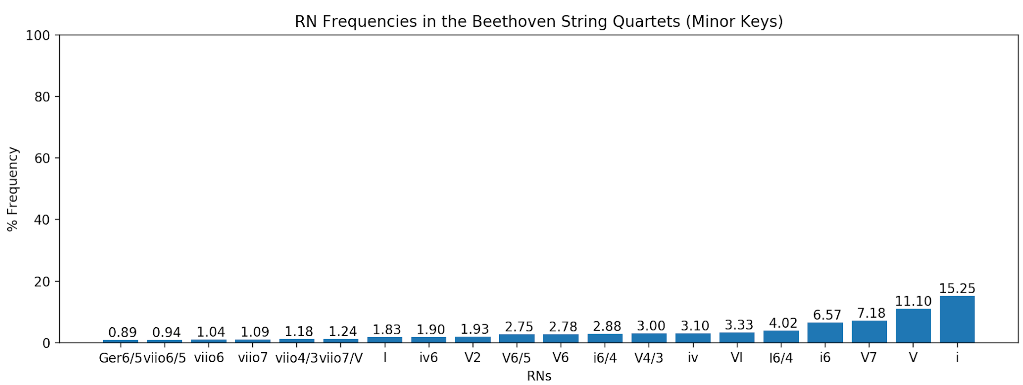
- As one would expect, the tonic and dominant chords take the lead by far in both modes. In major modes, tonic and dominant chords alone result in approximately 58% of Beethoven's chords, and 60% in minor modes.
- In the major mode, the ii chord is the fourth most common chord at 6%, but in minor, iio is thirteenth at 1.03%, and iio7 is fifteenth at 0.91%. This large disparity shows that Beethoven systematically used the supertonic chords differently between

major and minor modes. No obvious predominant replacement for ii in minor can be discerned from the data.

- The percent frequency of 1.27% for iii chords in the major suggests to me questionable analyses on the part of the ABC.
- The prominence of certain unexpected chords in both charts, such as the minor tonic and minor subdominant chords in major, and the major tonic and major subdominant chords in minor, if not simply mode mixture, are likely due either to temporary tonicizations, or also to the analysts choice not to mark small-scale temporary modulations with a key change. I also predict that, in the case of the I chord in the minor mode chart, that its enormous presence is partially due to mislabelings of V/iv chords as I chords.

Below are the 20 most common roman numerals again, this time with inversions included:




More items of note:

- In the major mode, it seems that the ii chord is approximately as likely to occur in root position as it is in first inversion. This goes against the notion held by some that ii6 chords are much more likely than ii chords. Similarly, viio seems to appear with a very similar frequency as its supposedly more common first inversion cousin. More data on the inversional frequencies of individual chords below.

- In minor, I6/4 has a frequency of 4%. This is a very unexpected result, especially considering that it appears even more often than the minor mode's native i6/4, at 2.88%. This is likely due to non-notated modulations to the parallel major in the ABC.

I provide here also the relative frequencies with which a given chord appears in each of its inversions. This is particularly interesting data with reference to the question of the ii vs. ii6 chords and viio vs. viio6 chords as noted above, as well as for better understanding Beethoven's use of inversions:

Major modes:

V

| RN | % Frequency |
|---|---|
| V6/4 | 6.62764675503529 |
| V6 | 26.82045102427268 |
| V | 64.57221552762954 |

V7

| RN | % Frequency |
|---|---|
| V2 | 15.908453993461 |
| V4/3 | 18.26249416160673 |
| V6/5 | 19.23400280242877 |
| V7 | 46.595049042503504 |

I

| RN | % Frequency |
|---|---|
| I6/4 | 18.010870244268133 |
| I6 | 25.92615730617855 |
| I | 56.062972449553314 |

ii

| RN | % Frequency |
|---|---|
| ii6/4 | 3.696622052262587 |
| ii | 44.582536647546206 |
| ii6 | 51.7208413001912 |

viio

| RN | % Frequency |
|---|---|
| viio6/4 | 14.053254437869821 |
| viio | 38.83136094674556 |
| viio6 | 47.11538461538461 |

IV

| RN | % Frequency |
|---|---|
| IV6/4 | 4.783258594917788 |
| IV6 | 32.40657698056801 |
| IV | 62.810164424514205 |

Items of Note:

- While ii6 and viio6 are indeed more common than their respective root position chords, the difference is not by so much as to rule out the normative use of ii and viio, provided that the ABC analyses analyze these chords correctly.
- V7 appears by far most often in root position, and its inversions are each approximately equally frequent with each other. The effect of the "dissonant" fourth above the bass in the case of the V4/3 seems not to matter much.

Minor Modes:

V

| RN | % Frequency |
|---|---|
| V6/4 | 5.535643170051772 |
| V6 | 19.315013938669853 |
| V | 73.59617682198328 |

V7

| RN | % Frequency |
|---|---|
| V2 | 12.16895742541066 |
| V4/3 | 15.186054307743882 |
| V6/5 | 19.711699631243714 |
| V7 | 52.93328863560174 |

I

| RN | % Frequency |
|---|---|
| i6/4 | 9.599681338378808 |
| i6 | 28.579964150567616 |
| i | 61.7805218084047 |

iio

| RN | % Frequency |
|---|---|
| **iio6/4** | 4.854368932038835 |
| **iio** | 30.097087378640776 |
| **iio6** | 65.0485436893204 |

viio

| RN | % Frequency |
|---|---|
| **viio6/4** | 16.856492027334852 |
| **viio** | 27.33485193621868 |
| **viio6** | 55.80865603644647 |

IV

| RN | % Frequency |
|---|---|
| **iv6/4** | 3.321364452423698 |
| **iv6** | 37.61220825852783 |
| **iv** | 59.06642728904848 |

viio7

| RN | % Frequency |
|---|---|
| **viio6/5** | 17.806451612903228 |
| **viio2** | 22.064516129032256 |
| **viio7** | 26.838709677419352 |
| **viio4/3** | 33.29032258064516 |

Items of Note:

- Because of its marked frequency in minor but not in major, here I have also included inversion statistics for the viio7 chord. Of interest here is that viio4/3 is actually significantly more frequent than the root position viio7. That all inversions of the viio7 chord are fairly interchangeable is unsurprising, given the fully diminished seventh chord's inversional intervallic symmetry.
- Compared with major, V, V7, and I each increasingly favor root position sonorities, the most dramatic case seen in V. Almost 75% of Beethoven's V chords are in root position.
- Although in major, iio6 and ii are approximately equally common, as are viio and viio6, it is clear that in minor this is not the case. Indeed, here the first inversion triads are favored to their root positions. However, it should be remembered that iio chords in minor are much less common than ii chords in major, so the dataset for iio is much smaller, with only 206 chords for iio in minor vs. 3138 for ii in major.

**Doubling Statistics**

In this section I provide charts detailing the percentages with which a given chord member is doubled when exactly four voices are present. Under “Doubling distribution,” the the first digit of the given number is to be understood as the number of pitches on the corresponding chord degree. For example, ‘211’ means that there are two voices on the root, one on the third, and one on the fifth. ‘1111’ means that the root, third, fifth, and seventh of a seventh chord each have one chord member.

I

| **Doubling Distribution** | **Percent Frequency** |
|---|---|
| **Other (less than 1% each)** | 2.252252252 |
| **103** | 1.401401401401 |
| **130** | 2.00200200200 |
| **400** | 4.304304304304 |
| **202** | 5.405405405405 |
| **301** | 5.555555555555 |
| **112** | 8.808808808808 |
| **220** | 11.56156156156 |
| **121** | 13.36336336336 |
| **310** | 14.26426426426 |
| **211** | 31.08108108108 |

I6

| **Doubling Distribution** | **Percent Frequency** |
|---|---|
| **Other (less than 1% each)** | 2.383531961 |
| **031** | 1.84182015167930 |
| **040** | 2.81690140845070 |
| **130** | 3.14192849404117 |
| **013** | 3.68364030335861 |
| **310** | 4.44203683640303 |
| **022** | 5.30877573131094 |
| **220** | 8.66738894907909 |
| **211** | 21.3434452871072 |
| **121** | 21.5601300108342 |
| **112** | 24.8104008667388 |

I6/4

| **Doubling Distribution** | **Percent Frequency** |
|---|---|
| **Other (less than 1% each)** | 2.232854864 |
| **031** | 1.27591706539074 |
| **004** | 1.75438596491228 |
| **301** | 2.23285486443381 |
| **022** | 3.82775119617224 |
| **103** | 6.69856459330143 |
| **013** | 7.49601275917065 |
| **121** | 10.8452950558213 |
| **202** | 13.2376395534290 |
| **211** | 14.0350877192982 |
| **112** | 36.3636363636363 |

IV

| **Doubling Distribution** | **Percent Frequency** |
|---|---|
| **Other (less than 1% each)** | 1.982378855 |
| **301** | 1.98237885462555 |
| **103** | 2.20264317180616 |
| **202** | 2.42290748898678 |
| **400** | 3.30396475770925 |
| **130** | 4.40528634361233 |
| **310** | 6.60792951541850 |
| **112** | 10.3524229074889 |
| **220** | 15.6387665198237 |
| **121** | 25.5506607929515 |
| **211** | 25.5506607929515 |

ii

| **Doubling Distribution** | **Percent Frequency** |
|---|---|
| **Other (less than 1% each)** | 2.247191011 |
| **301** | 1.12359550561797 |
| **202** | 2.62172284644194 |
| **400** | 3.37078651685393 |
| **130** | 7.11610486891385 |
| **310** | 8.98876404494382 |
| **112** | 9.73782771535580 |
| **220** | 13.1086142322097 |
| **121** | 22.0973782771535 |
| **211** | 29.5880149812734 |

ii6

| **Doubling Distribution** | **Percent Frequency** |
|---|---|
| **Other (less than 1% each)** | 1.935483871 |
| **013** | 2.25806451612903 |
| **310** | 2.58064516129032 |
| **031** | 2.90322580645161 |
| **130** | 5.16129032258064 |
| **220** | 5.80645161290322 |
| **022** | 7.09677419354838 |
| **211** | 16.4516129032258 |
| **112** | 20.6451612903225 |
| **121** | 35.1612903225806 |

V

**Doubling Distribution**
**Percent Frequency**
**Other (less than 1% each)**
1.647785788

| | | |
|---|---|---|
| **130** | 1.3388259526261 | |
| **103** | 2.0597322348094 | |
| | **202** | 4.1194644696189 |
| | **301** | 5.2523171987641 |
| **220** | 6.2821833161688 | |
| | **121** | 6.9001029866117 |
| **112** | 7.2090628218331 | |
| **400** | 11.637487126673 | |
| **310** | 12.358393408856 | |
| | **211** | 41.194644696189 |

V7

**Doubling Distribution**
**Percent Frequency**
**Other (less than 1% each)**
3.458425313

| | |
|---|---|
| **4000** | 1.0301692420897 |
| **0121** | 1.0301692420897 |
| **2200** | 1.2509197939661 |
| **1210** | 1.3245033112582 |
| **3010** | 1.6188373804267 |
| **1102** | 1.9867549668874 |
| **1201** | 1.9867549668874 |
| **1012** | 2.0603384841795 |
| **3100** | 2.0603384841795 |
| **3001** | 2.1339220014716 |
| **2002** | 2.2075055187637 |
| **2020** | 2.4282560706401 |
| **1021** | 3.3848417954378 |
| **1120** | 3.8999264164827 |
| **2011** | 10.301692420897 |
| **2110** | 10.301692420897 |
| **2101** | 10.375275938189 |
| **1111** | 37.159676232523 |

viio6

**Doubling Distribution**
**Percent Frequency**

| | |
|---|---|
| **301** | 0.9090909090909 |
| **202** | 0.9090909090909 |
| **103** | 0.9090909090909 |
| **130** | 1.8181818181818 |
| **310** | 2.7272727272727 |
| **220** | 2.7272727272727 |
| **013** | 2.7272727272727 |
| **031** | 3.6363636363636 |
| **211** | 10 |
| **022** | 15.454545454545 |
| **112** | 18.181818181818 |

Observations:

- Beethoven's doubling tendencies are clearly dependent on the inversion he uses. In root position I chords, for example, Beethoven is more than twice as likely to use the most common doubling configuration (all notes present with root doubled) as he is the next configuration (root tripled and fifth absent), but in first inversion I chords, all the configurations which contain each chord tone are approximately equally likely.
- The preponderance of doubling configurations where the root is missing (i.e. of the form 0xx) suggest imaginative analyses in the ABC. So do configurations where only one chord tone is present.
- In the V7 chord, Beethoven is almost twice as likely to use a complete voicing (1111) rather than leave out either the third or the fifth (2011 and 2101). This is significant

because in a V7-I progression, either the V7 or the I chord must be incomplete to be in compliance with standard voice-leading procedures. However, to judge this more clearly it is necessary to isolate the V7 chords that are specifically in V7-I progressions from the V7 chords at large.

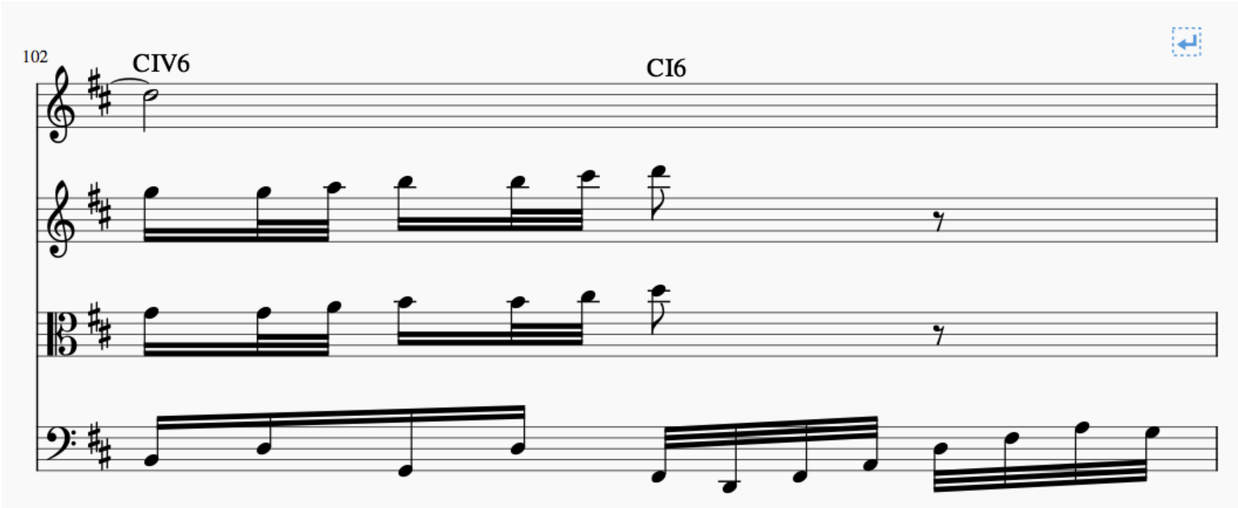
- It is much more likely that the ii chord double its third rather than its root. This is in compliance with the idea that ii6 acts like an altered IV chord.
- Approximately 65% of the time, the V chord doubles, triples, or (in these analyses) quadruples its root, rather than another chord member.

**Leading-tone Passing Tones in IV-I progressions**

One thing that my program is able to do well is to point out sections of the score which may be of interest to the human researcher. I wrote a program to help me find a very specific yet extremely interesting voice-leading: I-IV progressions where the scale degree 4 moves to 3, and scale degree 6 moves to 8 (1) via a passing tone on the leading tone. Following the upcoming work of Dmitri Tymoczko, this voicing is interesting because it gives the predominant IV chord a sort of dominant function, as the tritone formed by 4-3 and 7-8 is resolved *à la V7*, though over top a tonic pedal. Bach used this specific voice-leading for the IV-I progression approximately 40% of the time in his chorales (as known from the upcoming work by Dmitri Tymoczko), and so if Beethoven uses the voice-leading at a significantly different frequency, that suggests that Bach chorales are not the best model for modeling voice-leadings into the classical era.

I checked each and every one of the 431 IV-I progressions (in any inversions) by hand to see whether this voice-leading featured in any of them. I found a total of only five individual excerpts (i.e., excluding measures where the figure is repeated nearly note for note), which seems to confirm the inequivalence of voice-leadings in Bach chorales and Beethoven string quartets. I list the relevant excerpts here, accompanied by brief comments:

Op. 18 no. 5, third movement, var. 5 mm. 102




Here, the resolving tritone appears in an implied voice, not an actually sounding one. The G on the upbeat of 1 is clearly heard as moving down to F#, while the C# in the violin 2 just before beat 2 moves up to D. However, the C# is not sounded simultaneously against the G, although if played on an open string the tritone could be indeed actualized. The destabilizing dominant effect is also greatly mitigated by the very quick passing of the leading tone, a 32nd note in *Andante cantabile.*

This progression is repeated two bars later, at mm. 104 as well.
Op. 18 no. 5, third movement, closing section, mm. 131-132


This excerpt comes shortly after the one discussed above. Here, the resolving tritone does indeed sound. That two out of five of these voice-leadings occur within a single movement of one quartet is partially explained by Beethoven's motivic consistency within a given movement, and within entire quartets more generally. (Note that the final two examples come from different movements in the same quartet; five instances in three quartets.) In this and the previous example, the resolving tritone was generated from simple scalar motion, obviously derived from the simple scale which forms the theme for this set of variations.

Op. 130, first movement, mm 146


Another simple ascending scale. This leading-tone passing-tone again is disguised by a quick tempo: *Allegro*.

Op. 132, second movement, mm. 197-198


To put this voice-leading into the same category as the rest is not so obvious, since the penultimate G# in the violin I part before the I chord is perhaps best read as a neighbor note, rather than a passing tone. However, because the G# does indeed appear as a passing tone from F# to A the beat before, and since a tritone is indeed present and resolves as prescribed, it seems largely to work in the same way as the other listed voice-leadings.

Op. 132, third movement, “Heiliger Dankgesang,” mm. 71-72


Here, note that while the ABC analysts have marked the I6 as occurring on beat 2 of mm. 72, I analyse it as occuring on beat 1, reading the bass G of mm. 72 as a suspension in light of the confidence with which the violin I ascends to the D. Again, this voice-leading has many complications which disguise its dominant character. A brisk 32nd note C# leading tone (in *Andante* tempo) anticipates its resolution, and the G delays its resolution via a double neighbor tone (perhaps better described as a suspension + lower chromatic neighbor) until the next beat.

To summarize my findings in this section, while IV-I progressions with resolving diatonic tritones do certainly exist in Beethoven’s string quartets, they are few and far between, and often disguised by their quick passing or with embellishments. Dmitri Tymoczko’s hypothesis that this unique voice-leading is not ubiquitous in the music of the Classical era is supported by this evidence. One might also note that, since four out of five of these progressions being IV-I6 progressions, with the bass voice resolving down by step, that these progressions in Beethoven are somewhat similar to V4/2 chords.

**Resolution of the Tritone in V7-I Progressions**

Introductory music theory classes never fail to teach that, in a V7-I progression, the leading tone resolves up by step and the chordal seventh resolves down by step. It is therefore an interesting question to see how often Beethoven diverts from this schema.

While checking each possible instance of a voice-leading phenomenon is a very robust method, with certain problems such as this one, the size of the data is prohibitive. I wrote a script to return all V7-I progressions in any inversion which feature non-traditional resolutions of the V7’s tendency tones. For the present time, I could only check resolutions using the Voicing object, i.e. only with actually sounding voices rather than implied voices.

This results in certain inaccuracies when implied voices are present, as a traditional resolution in an implied voice can easily be obscured by a sounding note in the other implied voice. Because of this, I first searched the collection of “unambiguous” VLs, those VLs without any implied voices, so that the problem of the implied voice could be bracketed off. Furthermore, a yet-to-be-solved bug resulted in only approximately half of all V7-I progressions being examined by the program, exactly 2166 out of 2420 (major) + 1171 (minor) = 3591 total. (Note the disparity in magnitude between major and minor.) As for how this bug affects the accuracy of my result, I cannot at the moment say, although the data is still somewhat instructive nonetheless.

Here are my results, first with the unambiguous VLs:

| Tritone Resolution | Major Count | Minor Count | Total | % of total |
|---|---|---|---|---|
| 4->3, 7->1 | 168 | 67 | 235 | 77.30263158 |
| 4->3, 7->5 | 7 | 3 | 10 | 3.289473684 |
| 4->5, 7->1 | 11 | 11 | 22 | 7.236842105 |
| 4->5, 7->5 | 1 | 0 | 1 | 0.3289473684 |
| Other | 18 | 18 | 36 | 11.84210526 |

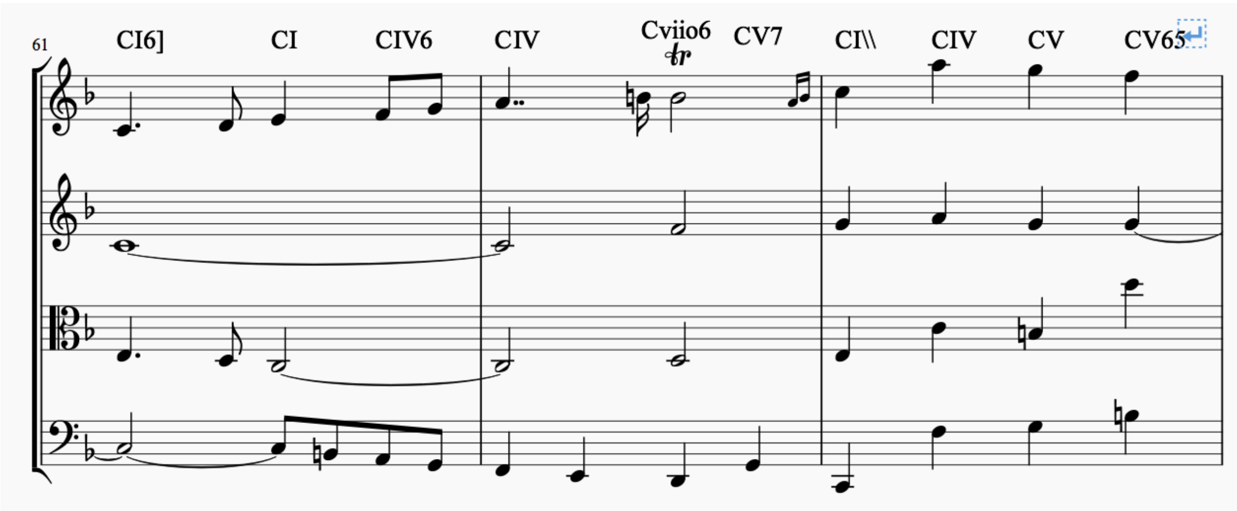
Because the number of 4->5 resolutions is rather shocking, I checked each one by hand to verify the result. Out of the 23 such instances:

- Sixteen were passing V4/3 chords, where the chordal seventh did indeed move up by step. Does this suggest that the chordal seventh to a certain extent loses its tendency when the fifth is in the bass? That would seem arbitrary. I think rather that the result suggests that the passing V4/3 chord is used by Beethoven less as a V7 chord in its own right, and more as a non-functional voice-leading chord. The chordal seventh in such a chord would have as much of a tendency to step downward as does the fourth scale degree in any other context, i.e. not very much at all. That the leading-tone in this situation always seems to resolve upward I think may be attributed to the fact that, unlike the subdominant scale degree, the leading tone exerts an upward pull in any context.
- Four instances were false positives due to incorrect RN analyses.

- One instance was a false positive due to the chordal seventh's resolving downward in a different instrument. This result is due to a flaw in the current implementation of the digital representation of voicings, which for now cannot record voices in one part resolving in another.

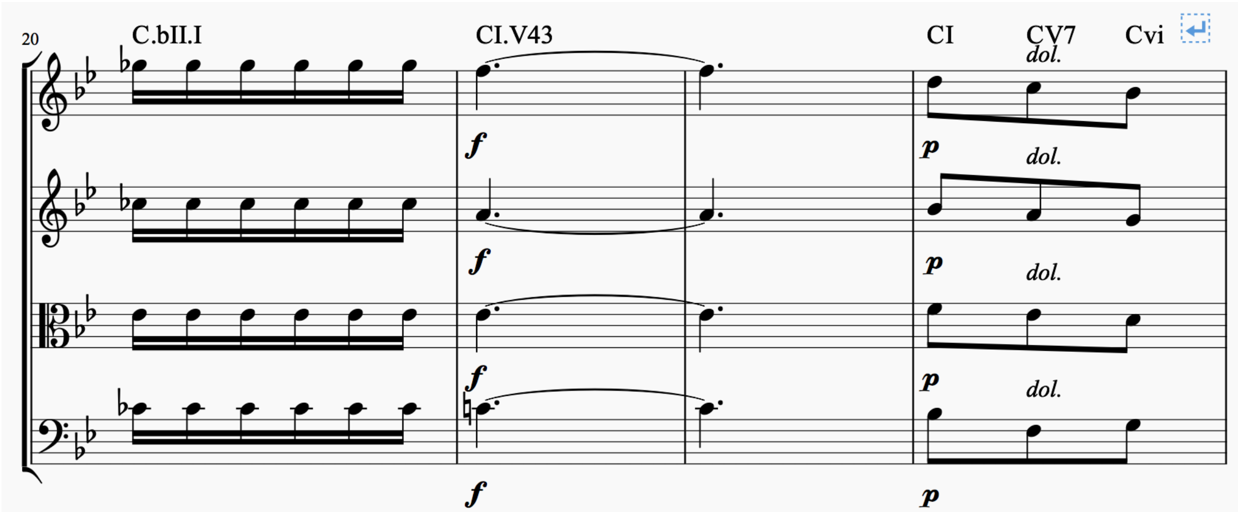
The two remaining examples may be worth showing here:

Op. 59 no. 1, first movement, mm. 61-63




The chordal seventh in mm. 62 in the violin II part certainly does move up by step. However, V7 here is not very strong as a harmony in its own right: rather, this appears to be a viio6-I progression where the bass touches the dominant on its way down from the supertonic to the tonic. The tendency of the fourth scale degree not being nearly as strong in the viio6 chord as it is in V7, this would explain the unusual behavior.

Op. 59 no. 1, second movement, mm. 20-23




Here we again have a V4/3 chord (in Bb major), although there is not any passing motion. The chordal seventh in the Violin II part indeed moves upward, but the unusual resolution is quickly undone when the Violin II immediately back to the chordal seventh, and then resolves correctly to scale degree 3.

For comparison, here is the same table documenting resolutions of tendency tones in V7-I progressions, this time drawing from *all* VLs, not just the unambiguous ones:

| Tritone Resolution | Major Count | Minor Count | Total | % of total |
|---|---|---|---|---|
| 4->3, 7->1 | 1144 | 359 | 1503 | 69.39058172 |
| 4->3, 7->5 | 92 | 14 | 106 | 4.893813481 |
| 4->5, 7->1 | 52 | 26 | 78 | 3.601108033 |
| 4->5, 7->5 | 9 | 0 | 9 | 0.4155124654 |
| Other | 359 | 111 | 470 | 21.6989843 |

When the accuracy of this program is improved, it may become possible to gather information about the relative tendency of the various tendency tones. Which resolves unconventionally more often? Chordal sevenths or leading tones? One could extend this line of thought beyond just V7-I progressions, to gather information about the "tendency" of chord tones of less obvious tendency: for example chordal sevenths in seventh chords built on tones other than the dominant, or the chord tones of the Neapolitan chord.

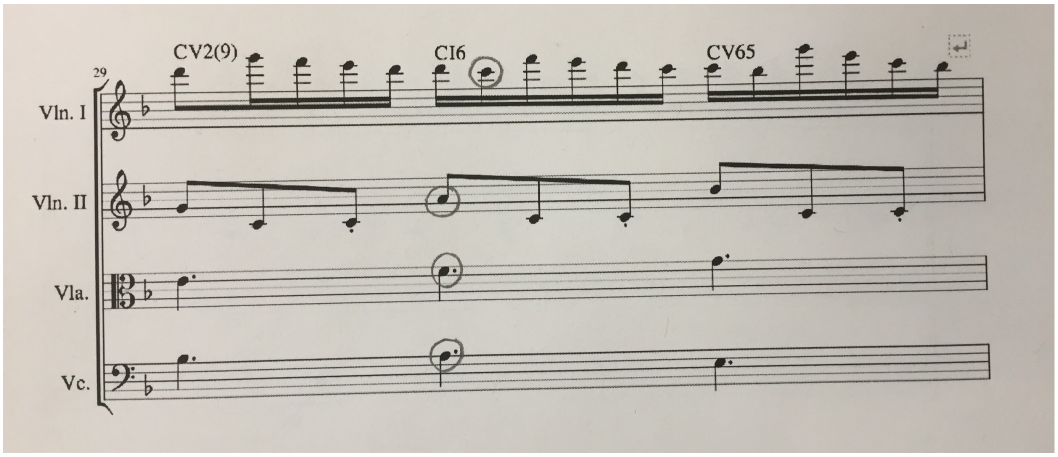
# Methods

In this section I discuss the key files, data structures, and functions with which I gathered these statistics: how they work, their inaccuracies, and future plans for improvement.

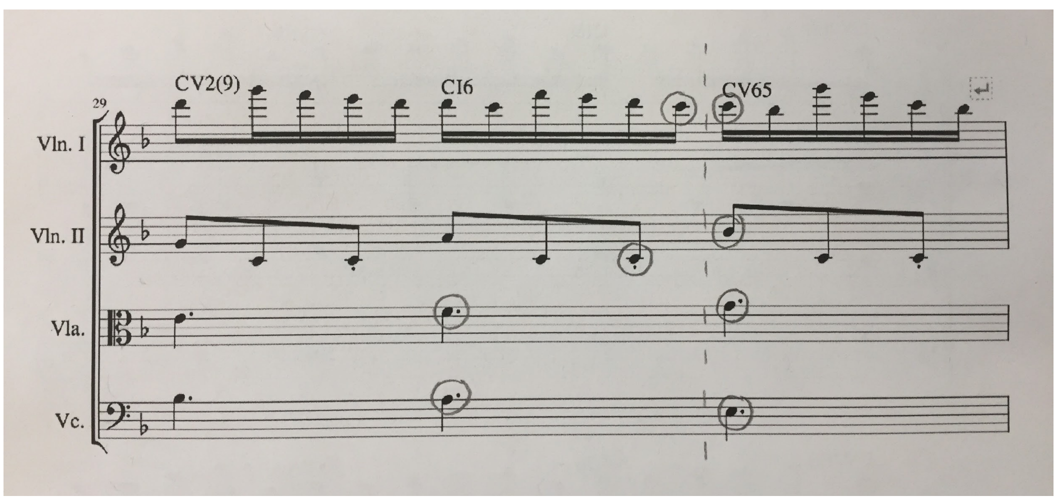
**voicing.py**
This program actually contains two different major classes which in a future update will be separate files: the Voicing object and the VL (voice-leading) object. Voicing and VL together form one of two representations with which I handle voice-leadings in the string quartet, the other of which are the BigVoicing and BigVL objects.

Voicing is an intelligent representation of a functional harmonic verticality. Given a roman numeral and its location within the score, Voicing stores the chordal verticality at that location, with non-chord-tones filtered out. The process of filtering out NCTs is as follows: for each part in the score, the program checks whether the note at the specified location belongs to the chord specified by the roman numeral. If the note belongs to the roman numeral, this note is a chord tone and is stored. If not, the program identifies it as an NCT, checks the following note, and continues until it finds a chord tone, or until it reaches the end of the domain of the given roman numeral, at which point it marks the voice as resting. That a NCT identifier may be as simple as this is indebted to the fact that the assignment of a roman numeral already decides which notes are chord tones and which ones are not. When the program comes across a rest, it is treated as an NCT, a (fairly safe, to my ear) assumption that voice-leadings are heard accross rests. Chordal simultaneities within a single instrument are also represented.

Below is an example of how the program represents a Voicing in the score, where circled notes show inclusion. Notice how the Violin I D is correctly identified as a NCT, as it does not belong to I6 in the key of F major:




The **VL object** consists of two Voicing objects, and calculates the intervals between each voice of the voicing. Over the course of its initialization, the VL object actually updates the first Voicing, such that the first Voicing represented is of that just before the chordal switch. The program starts at the boundary between the two chords, and using a procedure very similar to the one outlined above, this time works backwards to find the latest CTs in the roman numeral's domain. The process looks like this, imagining we continue the above, where the dotted line shows the chord boundary:

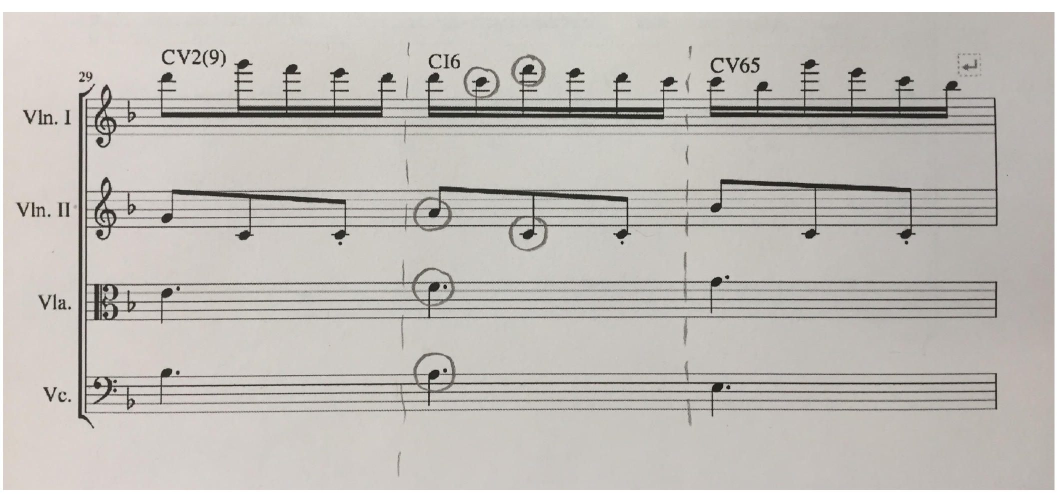



The VL and Voicing objects together have two major setbacks. First, Voicing can only recognize voices which are actually sounding at a given moment, and ignores voices which are only implied. In the above example, the very clear G-A-Bb voice in the Violin II is not recognized as such. (As an aside, the question of whether "implied" voices such as these are treated the same as actually sounding voices in Beethoven's voice-leading procedures is one of my concerns in this project.) Furthermore, each "voice" in the Voicing object is tied to the instrument playing it, and therefore is ignorant of cases in which a line is passed smoothly from one instrument to another, such as in cases where Beethoven revoices a chord between two instruments to make an otherwise unplayable simultaneity playable. Both of these inadequacies of representation are complemented and partially mitigated by the BigVL object.

In order to gauge the current accuracy of Voicing and VL, I checked each of the VLs recorded against the first sixty voice-leadings in Op. 18 no. 3 movement 1. The result was that for two out of the sixty VLs, one instrumental part was incorrect. These two errors were on the more harmless side. The first failed to find the seventh of a V7 in a V->V7 progression. This is a bug that only occurs a small percentage of the time specifically in progressions of the form X->X7, and should not have any bearing on my results here. The other bug left out a note where there should have been one. The effect of this latter bug is a false negative rather than a false positive, which is preferrable. Though I will work to improve the Voicing object in the future, for now its innaccuracy seems reasonable innocuous.

**bigVoicing.py and bigVL.py**
The BigVoicing object is the second representation of a voicing I used in this project. Whereas the Voicing object treats voicings essentially as slices in time with basic adjustments for filtering NCTs, the BigVoicing object treats voicings much more expansively. Within the domain prescribed by a given roman numeral, the BigVoicing includes all sounded notes which belong to the chord specified by the roman numeral. In the case of the G-A-Bb Violin II voice mentioned above, BigVoicing represents each of the three implied voices with equal weight. BigVoicing also does not discriminate between part boundaries, although it does record which pitches are sounded by which instruments.

Here is an example of which notes BigVoicing collects:



The nature of BigVoicing means that its data will always contain some amount of data noise. While it does a good job of representing implied voices, it will also see implied voices where there are actually none. For example, BigVoicing will read scales as containing many voices, the notes of scales usually do not display the behavior of voices (the above example actually being an interesting exception!). Functions to remove noise in the BigVoicing have undergone partial development, although much work remains to be done on this front.

BigVL connects two BigVoicings. At its current state, it does not decide which voices "go" to another voice, and thus does not calculate an interval structure, either, but remains useful for data storage.

For now, where BigVoicing is truly useful is cleaning the data provided by Voicing and VL such that only the most unambiguous VL objects remain. The most obvious voice-leadings in the Beethoven quartets are those with the following qualities, with "degree" understood to mean the number of voices recorded:

1. The degree of the first Voicing in VL is equal to that of the second Voicing, i.e. there is a 1:1 correspondence between each of the actually sounded notes from one Voicing to the next.
2. The degree of part of each Voicing is equal to the degree of each part in the BigVoicing; i.e. there are no implied voices which the Voicing object could have missed.

This scheme is particularly strict, since it requires that BigVoicing not record any implied voices in the second Voicing, either. This disqualifies a large number of VL objects which would otherwise appear quite unambiguous, due to the appearance of another chord tone long after what a human analyst would call the voice-leading proper, although still within the domain of a given roman numeral. While in future updates this situation is to be avoided, I chose to leave this schema as strict as it is for the sake of obtaining as truthful data as is currently possible. Out of 22970 voice-leadings recorded in the quartets, 1157 are designated "unambiguous." I use the term "unambiguous" throughout the paper to describe these voices.

**VL and BigVL Degree distribution**

In this section I present two charts which show the degree distribution my program collected using the VL and BigVL objects, respectively, which show something of the number of voices involved in the string quartet voice leadings. Degree here is of the form, (#voices in first VL, #voices in second VL). For example, a chorale texture where each instrument plays only one note in between two adjacent chords has the degree (4, 4). If each instrument played a double stop on the first chord, where two notes were doubled between the instruments, and a single note on the second chord, VL would return a degree of (8, 4), but BigVL would return a degree of (6, 4). The rows represent the degree of the first voice leading, and the columns that of the second.

VL degree distribution:

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 13 | 24 | 6 | 14 | 37 | 8 | 5 | 0 | 0 | 0 | 0 |
| 1 | 4 | 580 | 265 | 224 | 362 | 55 | 7 | 0 | 0 | 0 | 0 |
| 2 | 6 | 109 | 1981 | 731 | 673 | 97 | 16 | 0 | 2 | 0 | 0 |
| 3 | 14 | 117 | 451 | 4083 | 1936 | 156 | 23 | 10 | 4 | 0 | 0 |
| 4 | 32 | 529 | 769 | 1523 | 18452 | 886 | 124 | 47 | 12 | 3 | 4 |
| 5 | 1 | 92 | 108 | 167 | 802 | 1489 | 118 | 20 | 7 | 3 | 0 |
| 6 | 0 | 15 | 33 | 34 | 102 | 110 | 272 | 23 | 3 | 2 | 0 |
| 7 | 2 | 14 | 4 | 10 | 38 | 13 | 23 | 16 | 3 | 0 | 0 |
| 8 | 2 | 3 | 1 | 2 | 6 | 4 | 7 | 4 | 13 | 0 | 0 |
| 9 | 0 | 0 | 0 | 1 | 2 | 0 | 2 | 4 | 3 | 4 | 0 |
| 10 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 2 | 0 |

As would be expected, by far the most VLs have a degree of (4, 4). (But note that the degree (4, 4) does not simply indicate four instruments playing four notes, but also includes the group of two instruments playing two notes, two playing one note and one playing two notes, etc.) Also notice that the VL degrees where both chords are of the same degree (the line $y = -x$) are more common than the surrounding degrees. This reflects a tendency of Beethoven's to stay with a certain texture (e.g. double stops to double stops) rather than change sporadically.

BigVL Degree Distribution (excerpt):

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 349 | 112 | 89 | 105 | 166 | 74 | 66 | 25 | 13 | 9 | 3 | 4 | 7 |
| 1 | 113 | 168 | 153 | 171 | 213 | 122 | 72 | 37 | 32 | 7 | 3 | 4 | 5 |
| 2 | 103 | 128 | 499 | 497 | 465 | 213 | 121 | 107 | 29 | 16 | 9 | 23 | 3 |
| 3 | 109 | 165 | 381 | 1609 | 1693 | 605 | 336 | 146 | 77 | 36 | 30 | 10 | 17 |
| 4 | 162 | 243 | 505 | 1467 | 5794 | 1669 | 989 | 518 | 291 | 92 | 113 | 46 | 28 |
| 5 | 70 | 138 | 236 | 602 | 1702 | 1425 | 669 | 433 | 209 | 118 | 73 | 46 | 31 |
| 6 | 46 | 49 | 161 | 364 | 919 | 725 | 759 | 406 | 198 | 107 | 69 | 33 | 21 |
| 7 | 34 | 40 | 95 | 211 | 497 | 396 | 408 | 396 | 144 | 72 | 49 | 29 | 22 |
| 8 | 16 | 13 | 40 | 83 | 256 | 259 | 213 | 158 | 163 | 59 | 41 | 24 | 12 |
| 9 | 8 | 3 | 12 | 52 | 90 | 124 | 92 | 84 | 74 | 74 | 51 | 22 | 7 |
| 10 | 3 | 6 | 16 | 34 | 80 | 85 | 53 | 53 | 53 | 47 | 42 | 32 | 22 |
| 11 | 5 | 4 | 7 | 10 | 40 | 44 | 56 | 39 | 20 | 15 | 32 | 22 | 25 |
| 12 | 4 | 5 | 16 | 10 | 30 | 14 | 24 | 26 | 25 | 23 | 14 | 24 | 15 |

Immediately one notices that the distribution in the BigVL degree graph is far more even than the distribution in the VL degree graph. This is partially due to the "noise" in the data described earlier, and partially due to the greater number of voices that any given BigVL may record. VL capped off at 10; theoretically it could have hit 16 if Beethoven wrote quadruple stops for each instrument. On the other hand BigVL, taking place over a span of time, may hit a far greater range.

The careful eye will also notice that this graph betrays that something is going wrong here, indicated by the 349 (0,0) degree instances, as well as the 13 (0,0) instances in V. Both VL and BigVL are in need of a little more work.

**Other programs in the library:**
**JPa.py** is the main file of the library, extracting the data from the .txt romantext analysis files and making calls to all other programs in the library. **progs.py** stores information about the VL objects and BVL objects in a searchable dictionary. **getStats.py** contains a motley of methods with which I collected all the statistics about the collected data. **bigScore.py** houses the .mxl file for the given quartet, from which the Voicing and BigVoicing objects are actually extracted.

# Conclusion

The potential of the tools developed for this project extend beyond just the problems that I have chosen to explore here. With the basic infrastructure already in place, it should not be difficult, for example, to find information about Beethoven's doubling procedures (is he more likely to double the fifth than the third, and to what degree?), to extract all actually sounding parallel fifths for human perusal, to compile a collection of non-traditional harmonic progressions (provided the correctness of the ABC corpus), to quantify the pull of tendency tones (and chords not usually thought of as tendency tones) and much more.

This being said, the Voicing and BigVoicing objects are still wanting. Before many of my results may become definitive, the accuracy of the Voicing object must be much more heavily tested and improved. After both the Voicing and BigVoicing objects behave as expected, the next step will be to integrate them more fully in order to allow for the exploration of the behavior of implied voices. This line of inquiry holds great potential: we will be able to determine to what extent implied voices act similarly to actually sounding voices, and if not, determine to what extent rules such as the exclusion of parallel fifths and octaves, or the behavior of tendency tones, are relaxed.

It is also my hope that this paper served as a demonstration of many of the capabilities of music21, as well as computer-aided musicology in general. Certainly, even at this preliminary stage, interesting results were found, and more interesting paths were suggested.