Chapter 5

Natural Break Maps

Before an in-depth analysis of the acoustic data is conducted, the F1 and F2 mean values for each speaker will be presented geographically in this section in a pre-theoretic manner, without the construction of isoglosses or manually-defined ranges of values. This will enable the actual distributions for each vowel to be observed in an unobstructed way. Furthermore, it facilitates direct and objective comparison with similar maps provided in Chapter 10 in the ANAE.

Each map in this section presents each speaker's mean value for the automatic F1 and F2 measurements for 17 vowel classes. The points for each speaker are color coded to correspond to four ranges of values that span the entire range of possible values for each vowel. For ease of comparison, the color scheme of the ANAE (Labov et al. 2006:77) was adopted. Under this representation, F1 values are arranged from lowest to highest in the order red, yellow, green, blue; i.e., a red symbol represents the highest vowel and a blue symbol represents the lowest vowel. For F2, values are arranged from lowest to highest in the order blue, green, yellow, red (the opposite order from F1); i.e., a red symbol represents the most front vowel, and a blue symbol represents the most back vowel.

The values for the four ranges in each map were determined by using the Natural Break

algorithm in MapInfo (MapInfo Professional, version 7.8). This procedure attempts to find natural clusters of individual data points by iteratively minimizing the distance between the individual data values and the average value for each range until the clusters are as homogenous as possible. The MapInfo documentation does not provide any details about the implementation of their algorithm (such as how the clusters are initialized or whether multiple rounds of clustering are performed to avoid local minima), but they cite Jenks and Caspall (1971) as the source for their algorithm. In essence, the algorithm developed in that paper is a one-dimensional *k*-means clustering procedure[1] where the stopping criteria take into account a set of four accuracy indices defined by Jenks and Caspall (1971). For an initial examination of cartographic data, the use of such clustering techniques for determining the ranges on the map is preferable to defining the ranges by hand. This is because this procedure minimizes any subjective bias that could arise due to the researcher's preconceived notions of which groupings are important.

The natural break maps in this chapter were created from a subset of 45,756 tokens from the total of 112,848 automatic vowel measurements that were extracted from the corpus (see Section 3.8 for information about how these tokens were selected). The mean F1 and F2 values were computed from these tokens for the following 17 vowel classes: /i/, /e/, /ae/, /æN/, /o/, /ʌ/, /u/, /iy/, /ey/, /ayV/, /ay0/, /oy/, /Tuw/, /Kuw/, /ow/, /aw/, and /oh/. Three pairs from this set of vowel classes represent allophones of the same phonemic class: 1) /æN/ signifies tokens of /æ/ occuring before nasals, 2) /ayV/ signifies tokens of /ay/ that occur before voiced consonants or word-finally, whereas /ay0/ signifies /ay/ before voiceless consonants, and 3) /Tuw/ signifies /uw/ after coronal onsets, whereas /Kuw/ signifies /uw/ after non-coronal onsets.

The maps below contain F1 and F2 means for these 17 vowel classes for the 88 speakers listed in Tables 3.1 through 3.11. They contain tokens from all stylistic contexts that were

[1]For a description of the *k*-means clustering algorithm see, for example, Bishop (2006:424–428).

availble for each speaker. Thus, tokens from word lists, reading passages, and interviews are combined. While not ideal, this combination of data from different stylistic contexts was deemed necessary in order to provide a more complete geographic representation, since interview data is missing for a subset of the speakers. For example, for most of the DARE speakers, only the tokens from the "Arthur the Rat" reading passage were analyzed. Additionally, a few speakers in my corpus were only recorded reading the word list—time constraints prevented the completion of a full interview. Thus, data from all stylistic contexts are combined in the maps, so that these speakers who lack interview datacan also be included. Furthermore, studies have shown that the NCS and the merger of /o/ and /oh/, two of the most important changes in the dialect boundary region around Erie, are occurring below the level of consciousness. Thus, the amount of style-shifting between different contexts is minimal, and the combination of data from these three sources is justified.

A few of the speakers who were only able to provide word list data and who were recorded in the early stages of my fieldwork did not provide any examples of the vowels /iy/, /ey/, /u/, /oy/, /ay/, and /uw/. (An early version of the word list did not contain any tokens of these vowels.) In these cases, the speaker's symbol is not shown on the map, but the symbols for the other speakers from the same town are left in place. For example, the maps for /u/ in Figures 5.13 and 5.14 show that the speaker from Ford City and four speakers from Erie are missing.

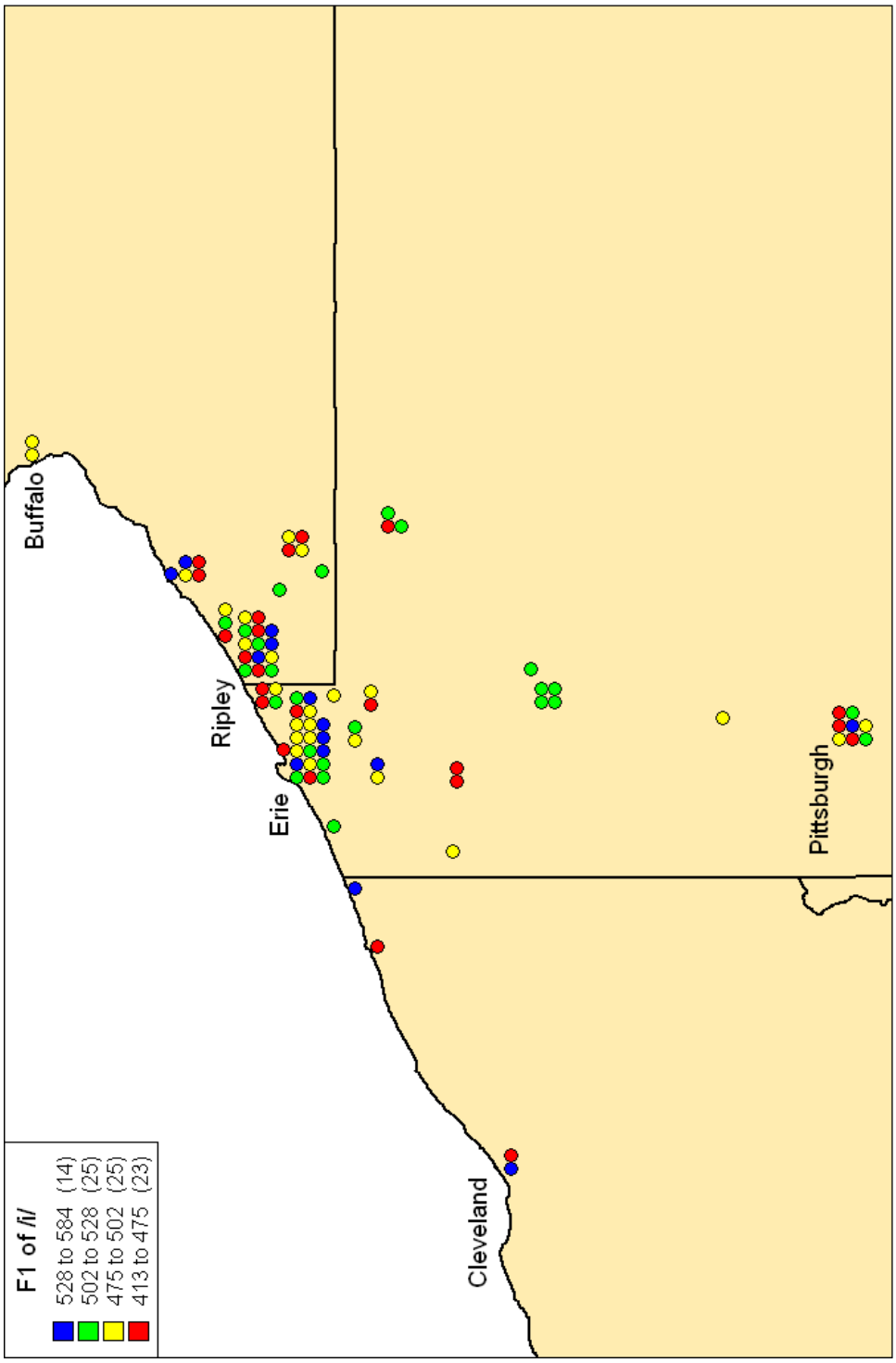

Figure 5.1: Natural break map for F1 of /i/

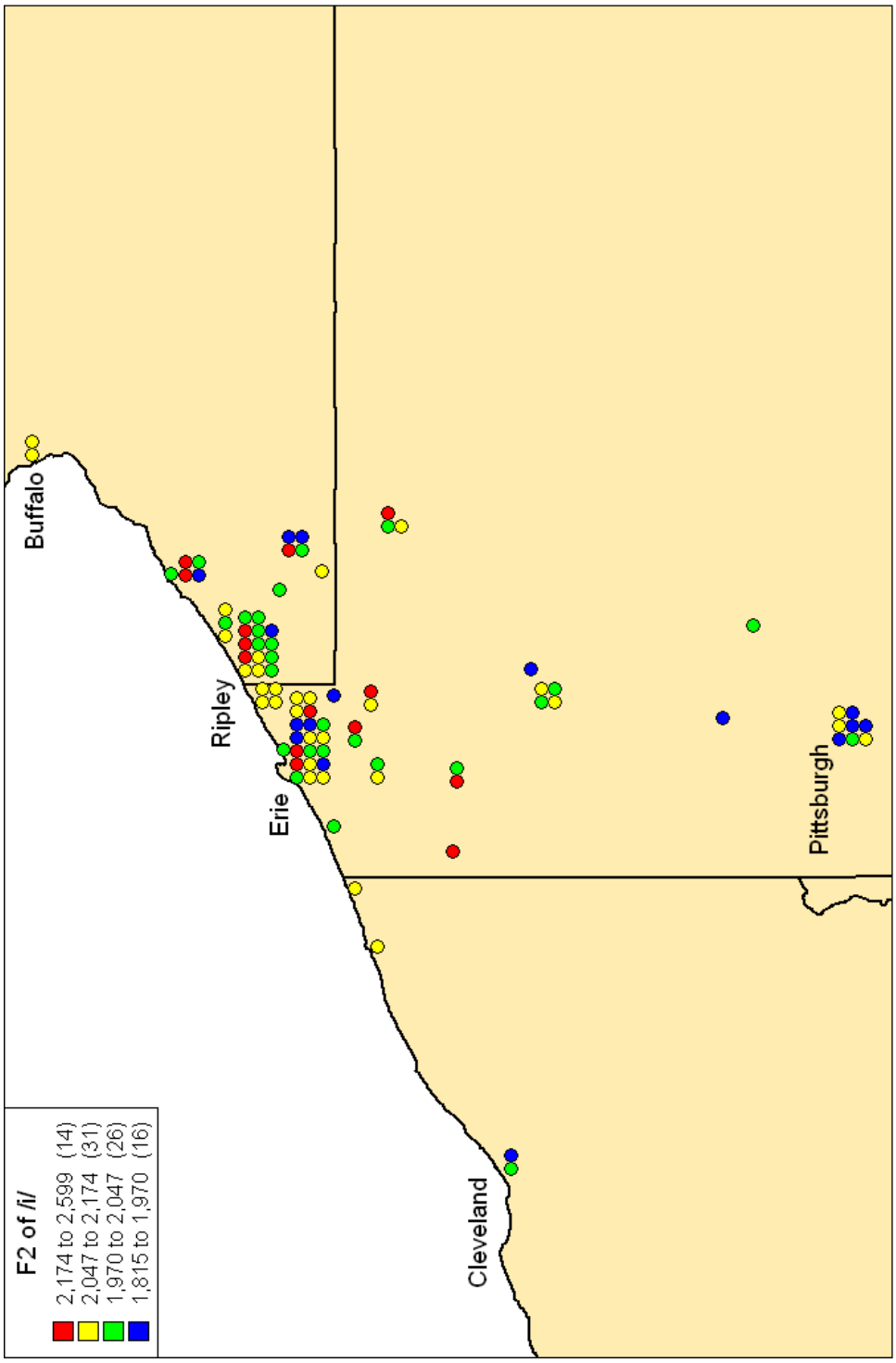

Figure 5.2: Natural break map for F2 of /i/

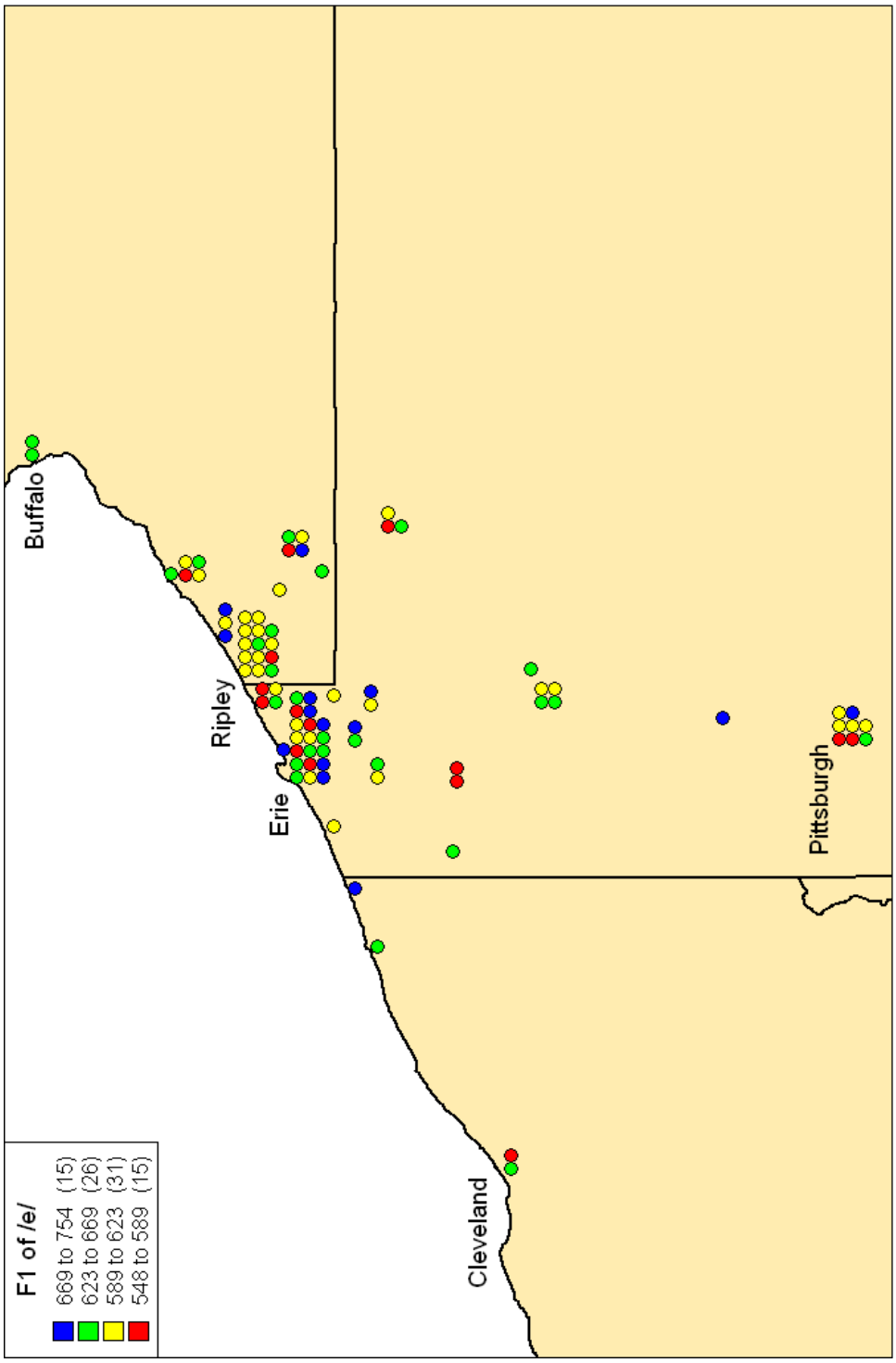

Figure 5.3: Natural break map for F1 of /e/

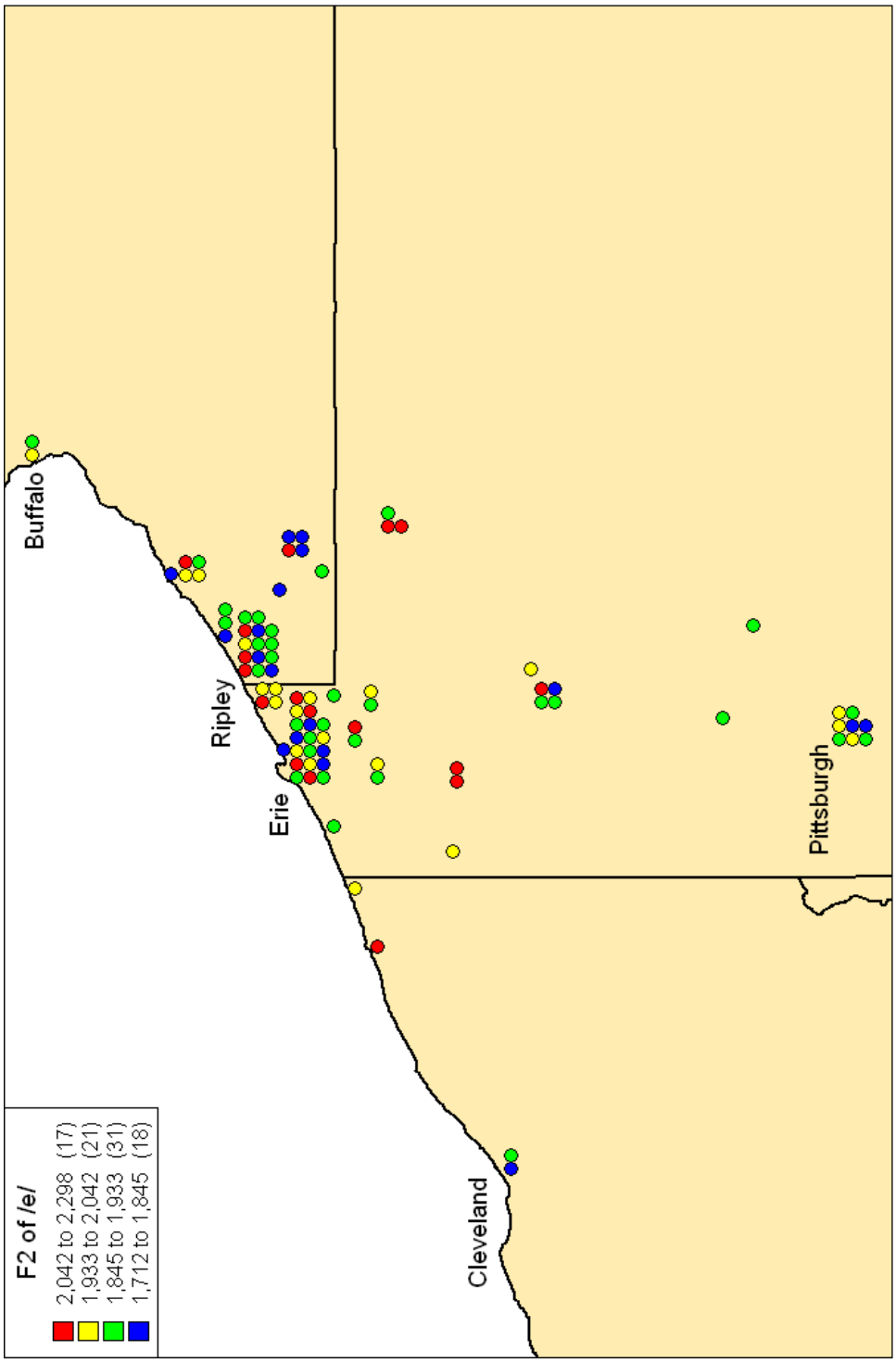

Figure 5.4: Natural break map for F2 of /e/

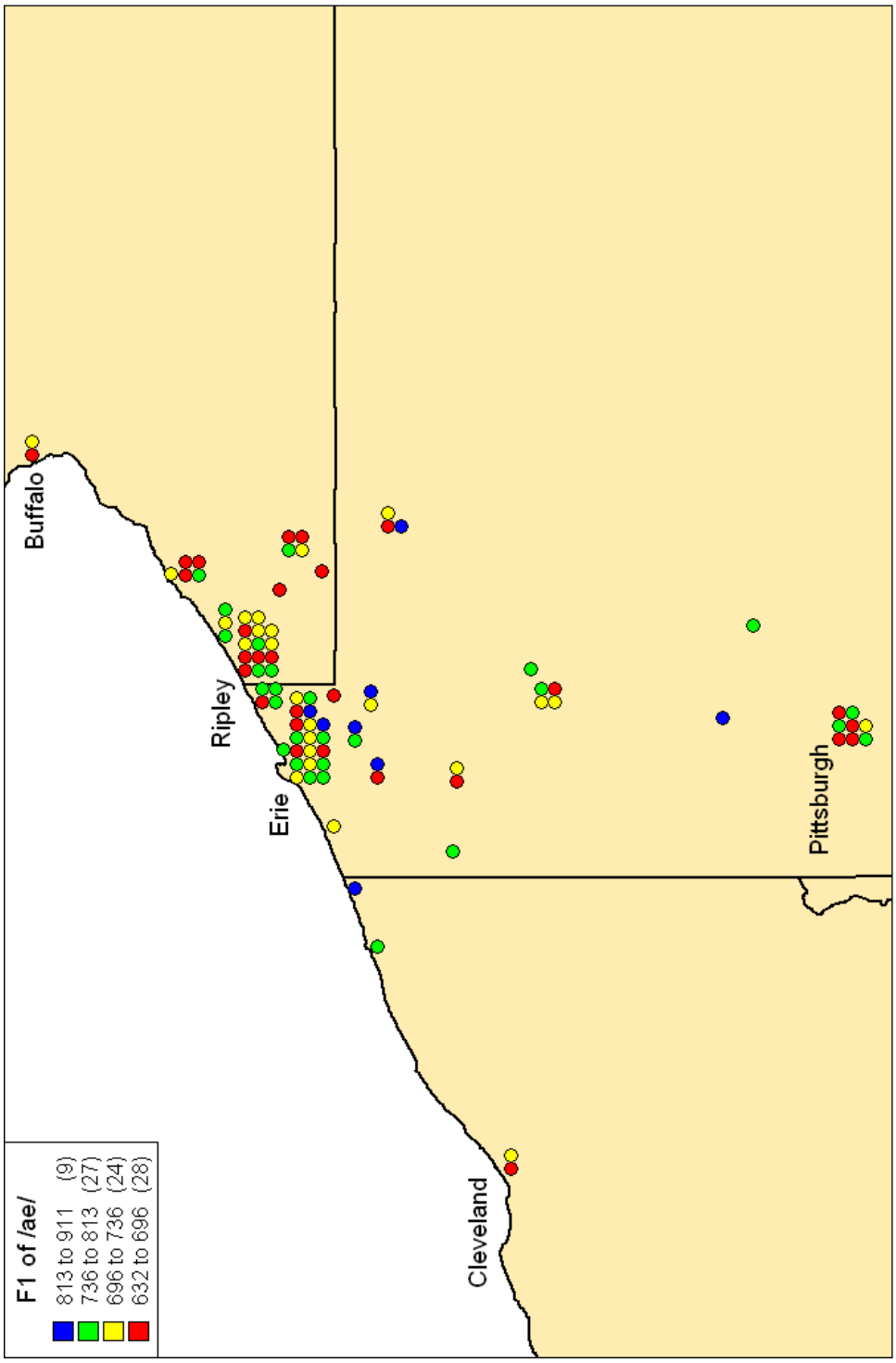

Figure 5.5: Natural break map for F1 of /æ/

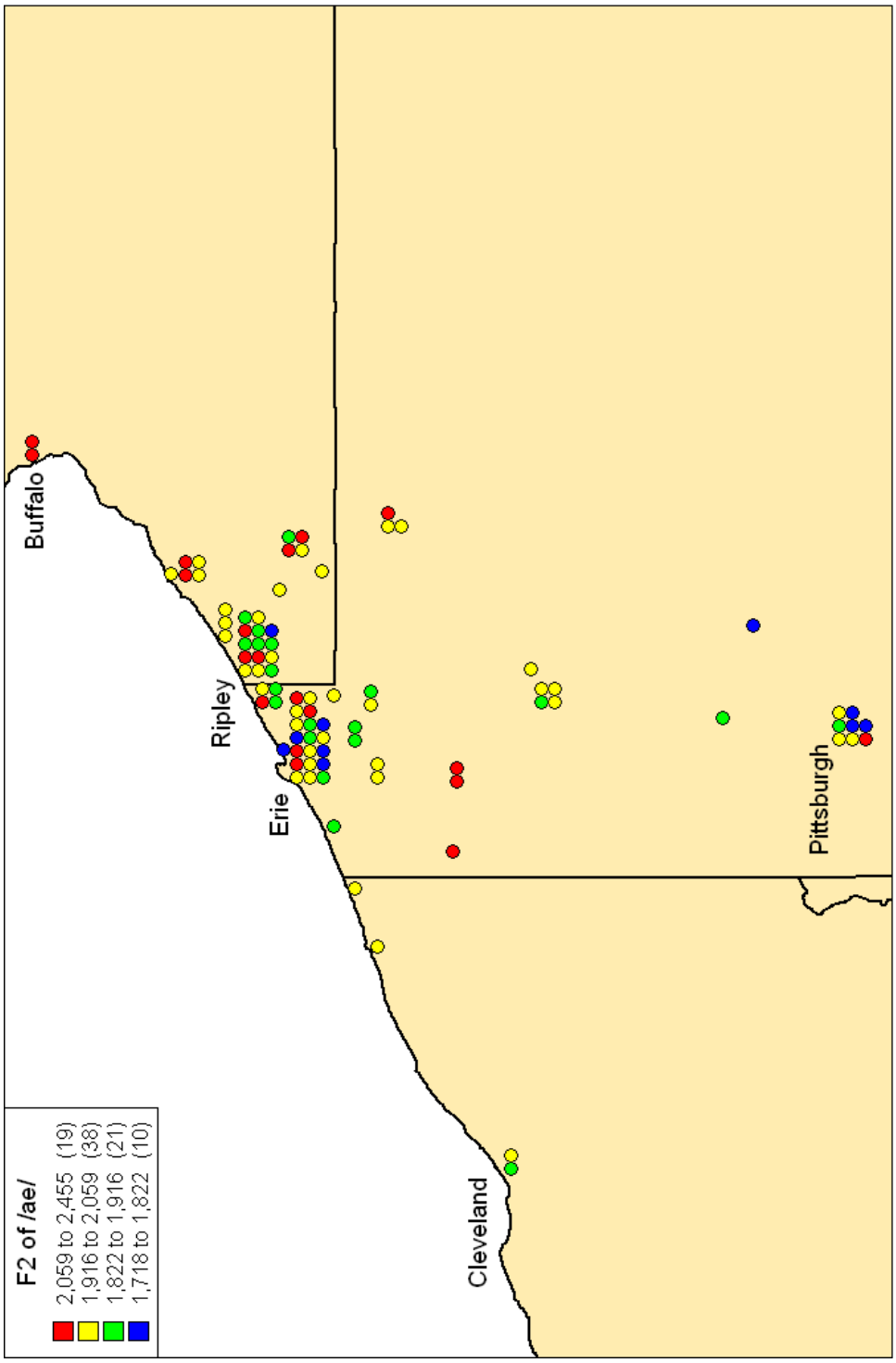

Figure 5.6: Natural break map for F2 of /æ/

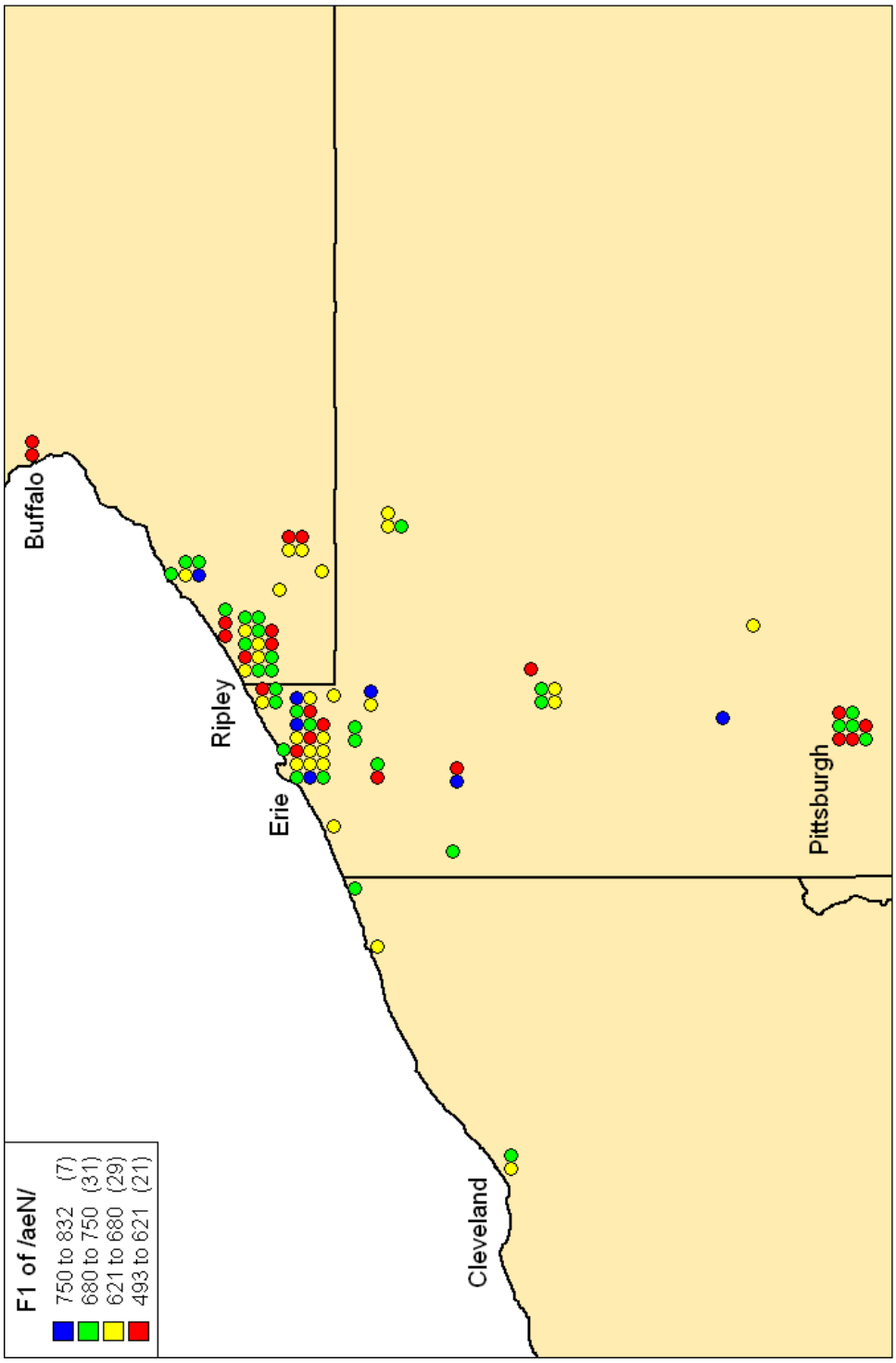

Figure 5.7: Natural break map for F1 of /æN/ (/æ/ before nasals)

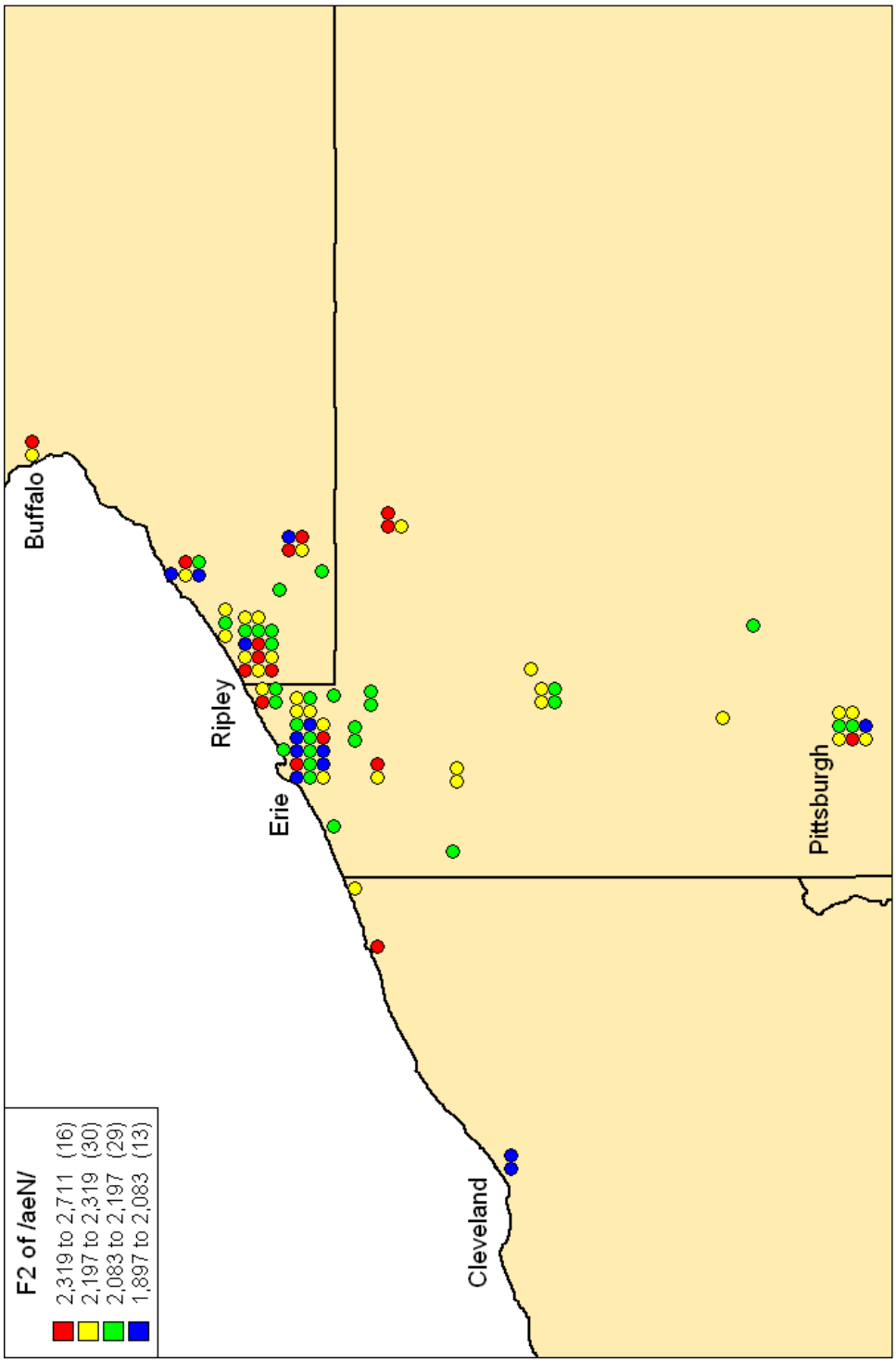

Figure 5.8: Natural break map for F2 of /æN/ (/æ/ before nasals)

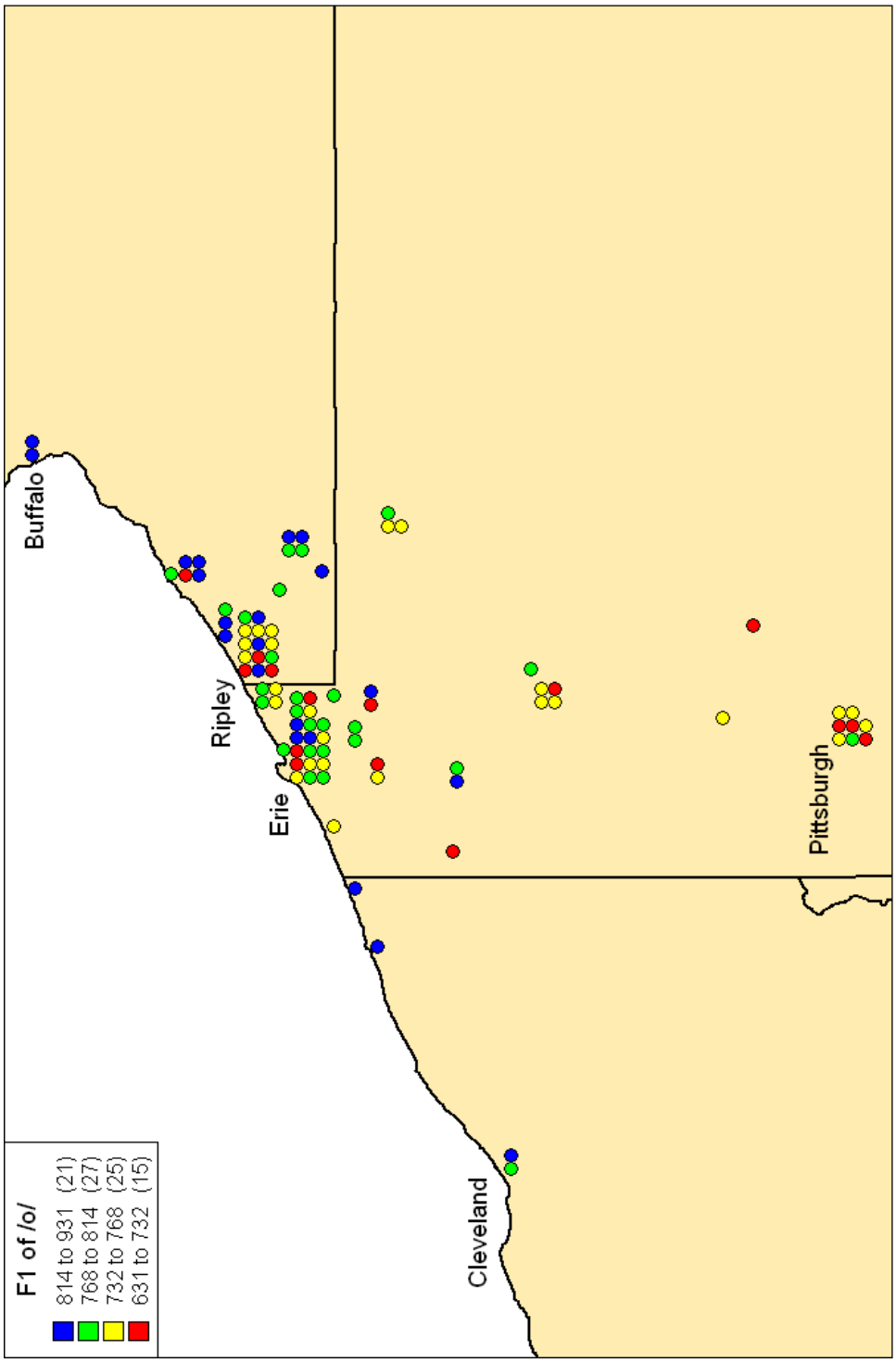

Figure 5.9: Natural break map for F1 of /o/

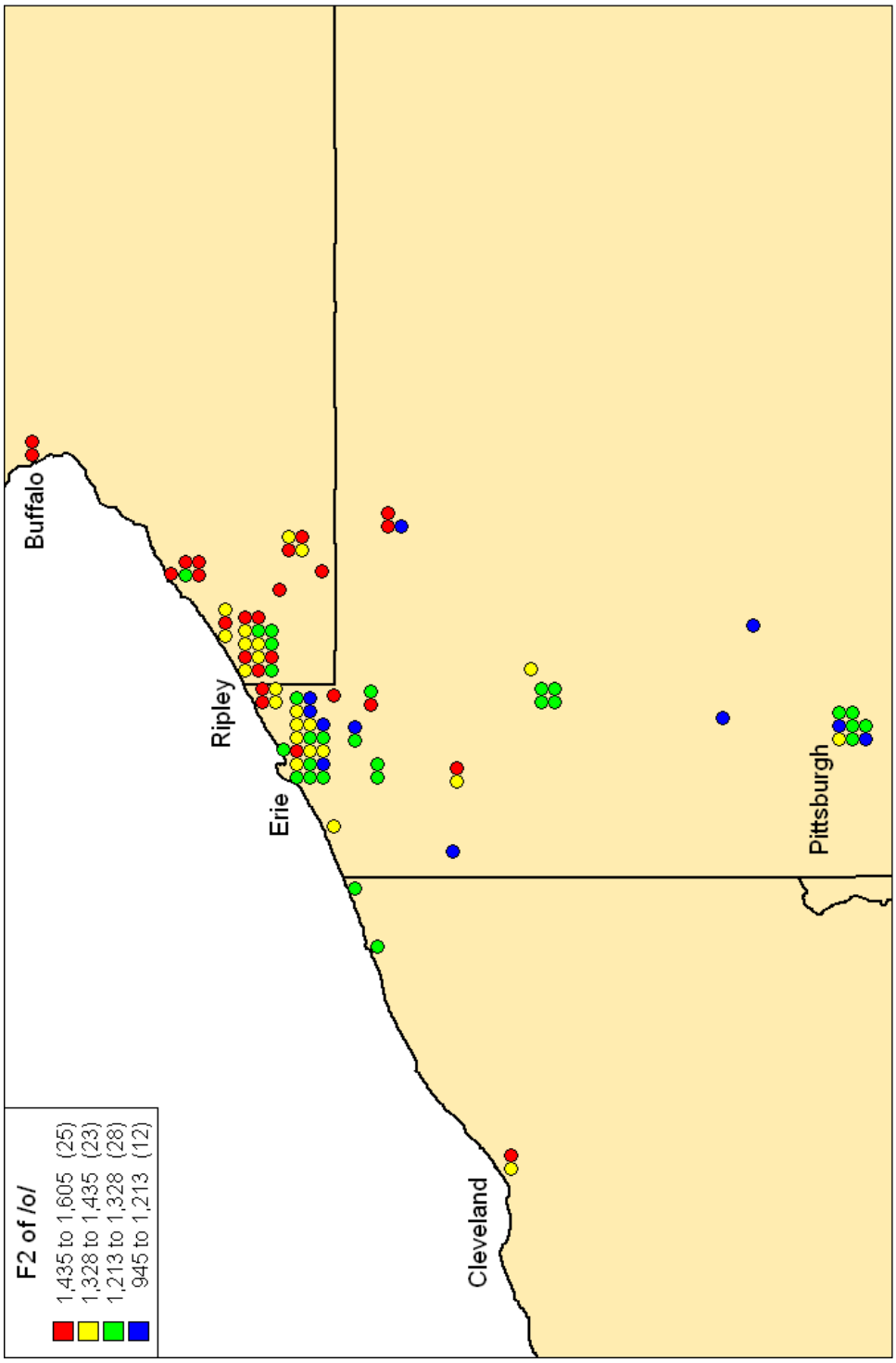

Figure 5.10: Natural break map for F2 of /o/

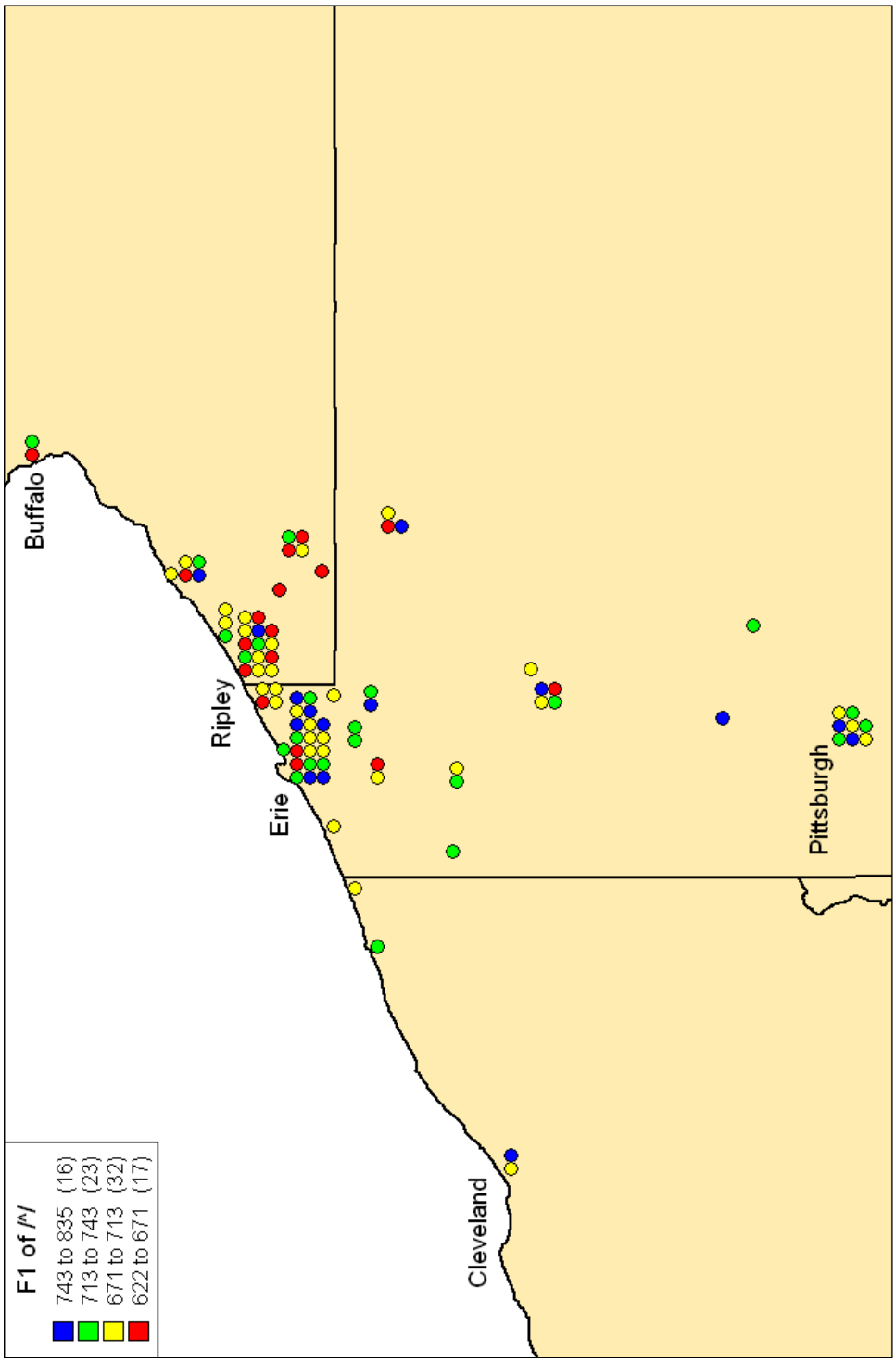

Figure 5.11: Natural break map for F1 of /ʌ/

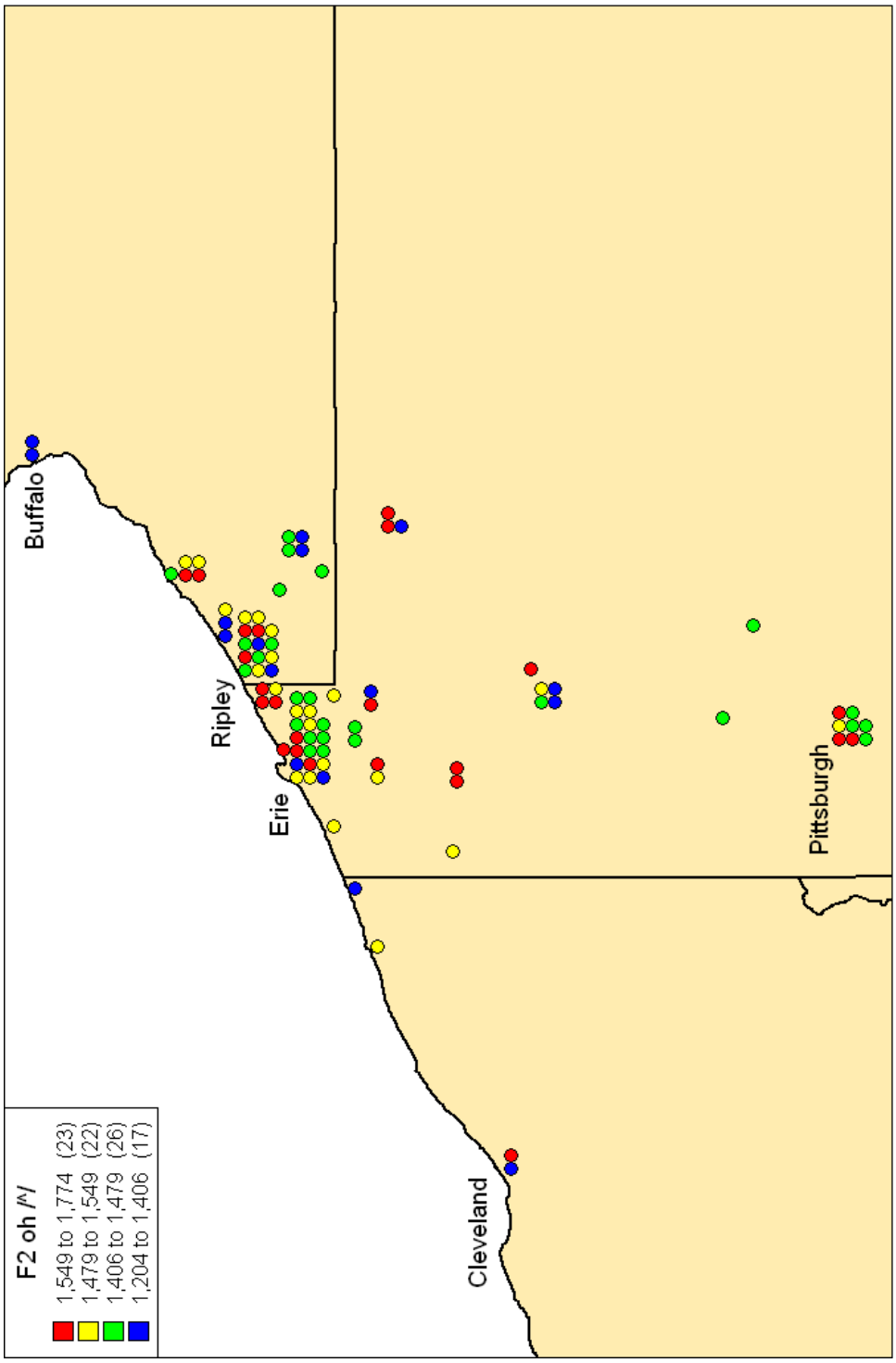

Figure 5.12: Natural break map for F2 of /ʌ/

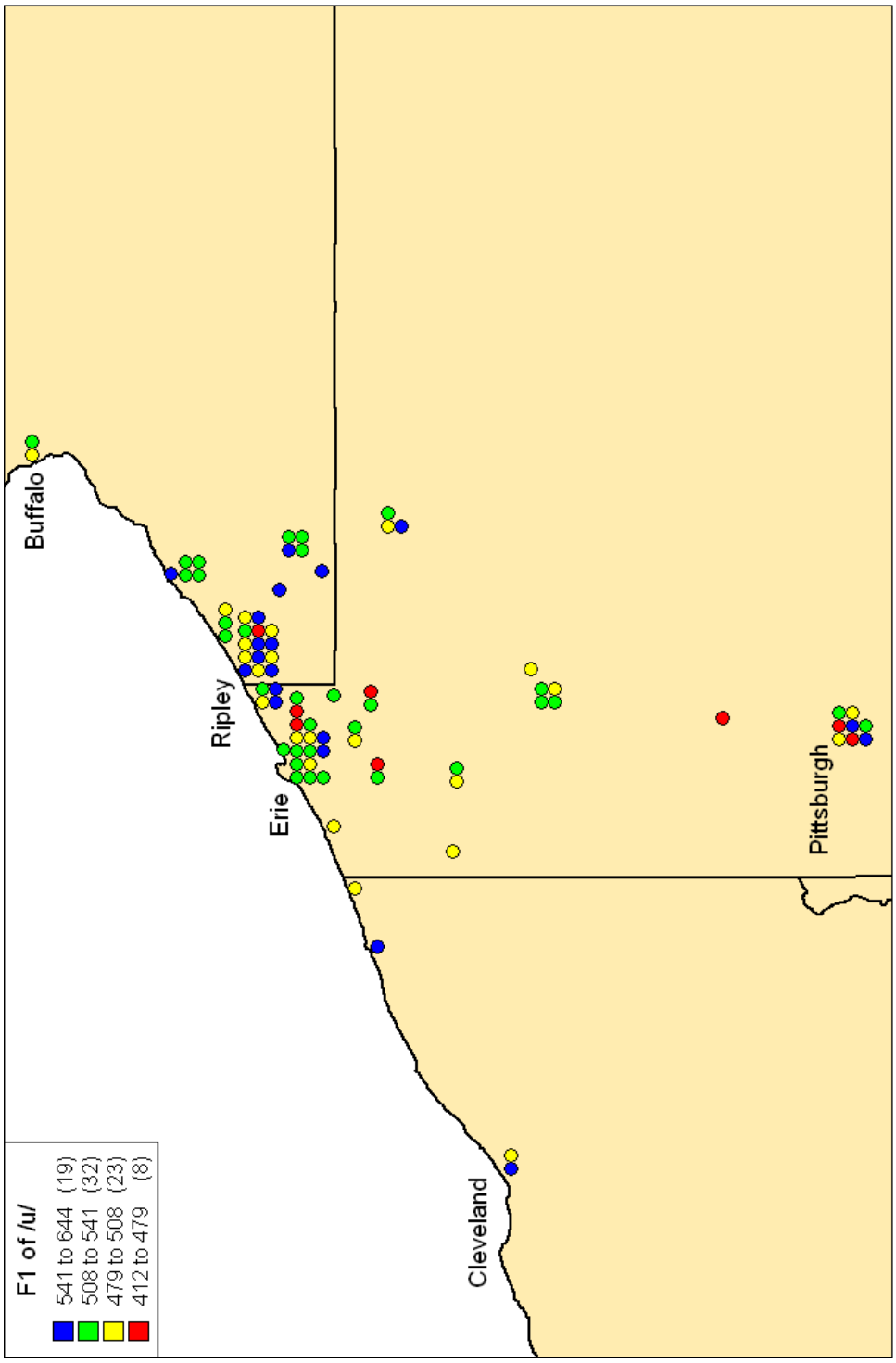

Figure 5.13: Natural break map for F1 of /u/

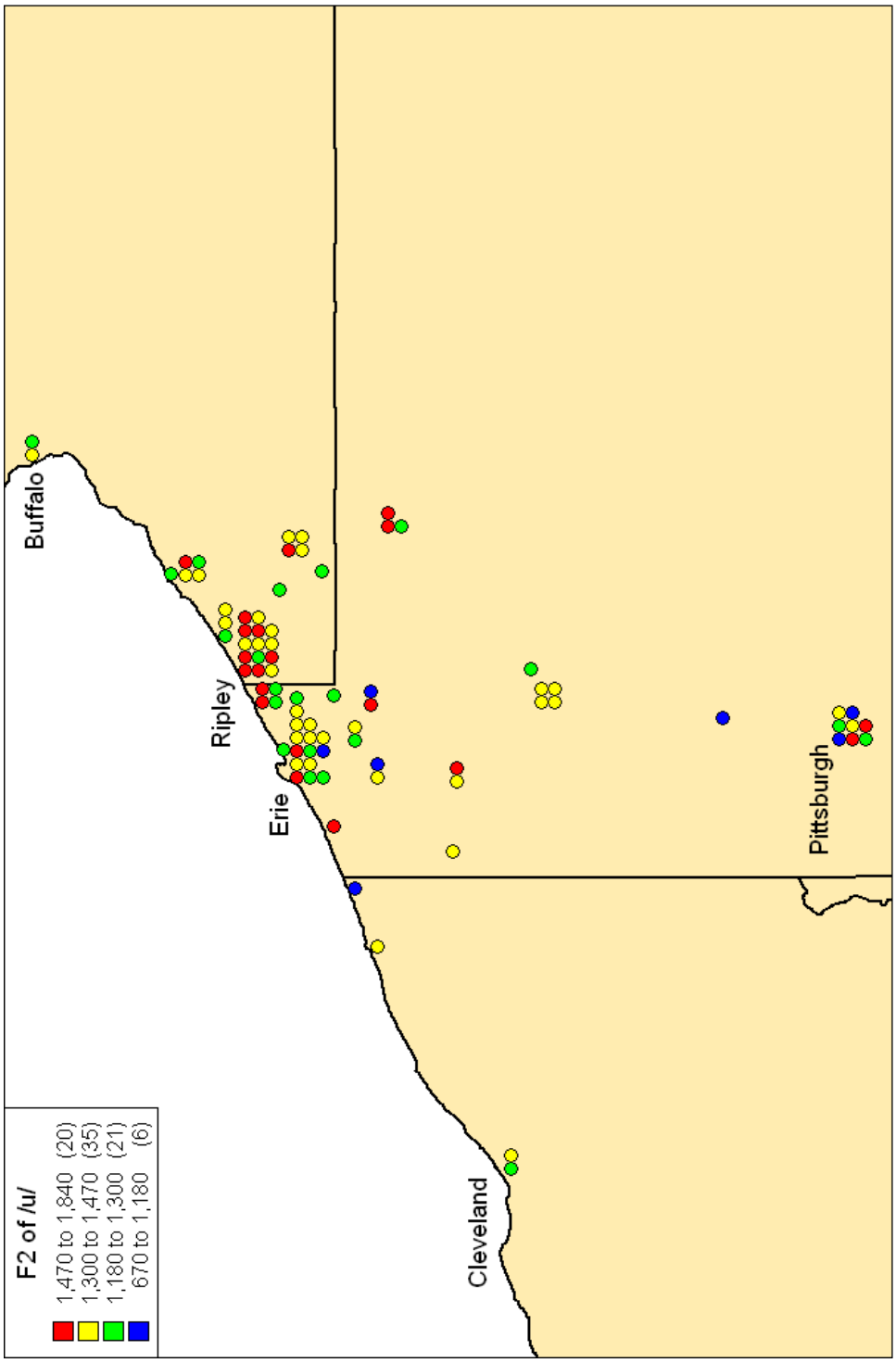

Figure 5.14: Natural break map for F2 of /u/

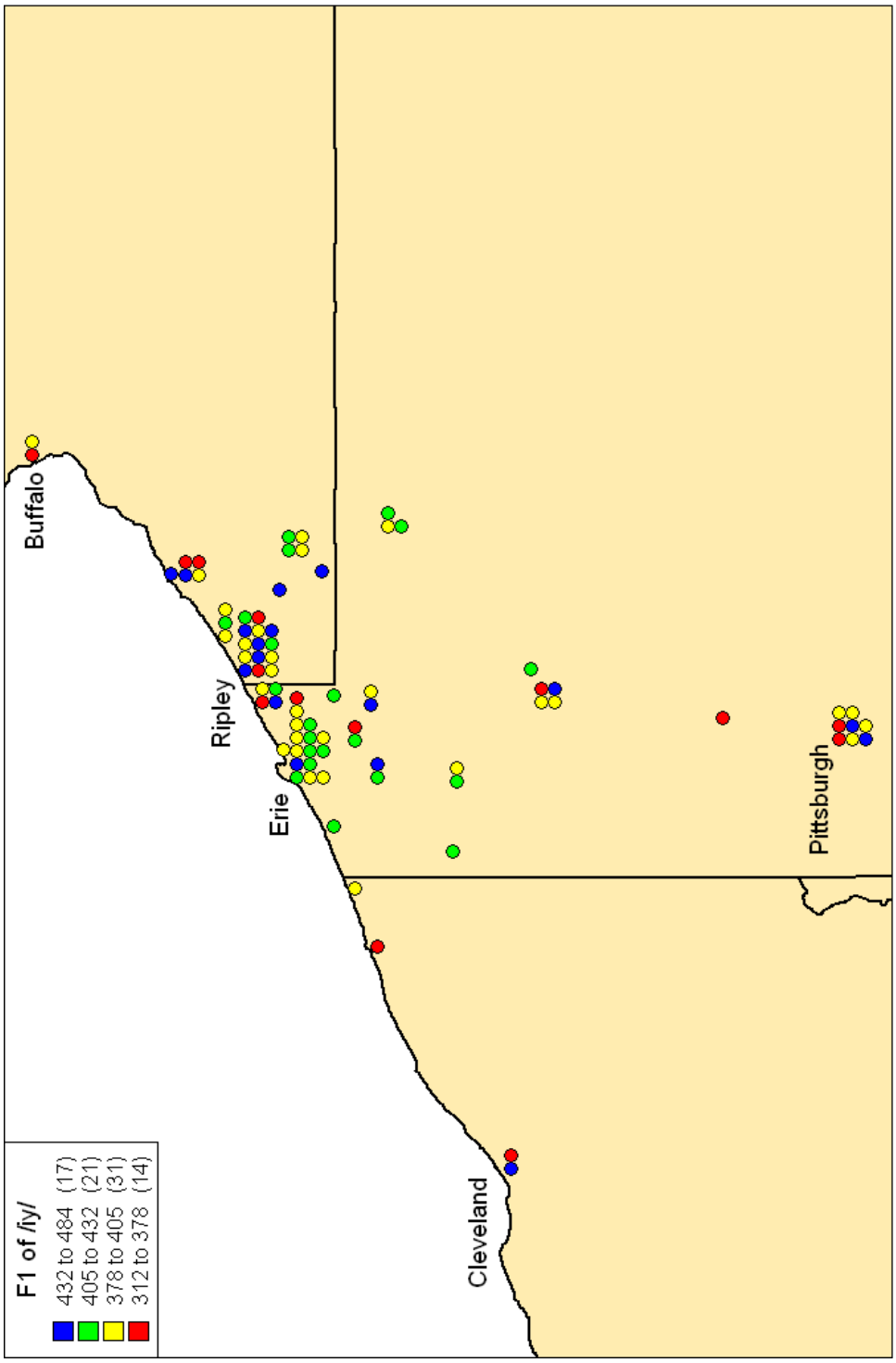

Figure 5.15: Natural break map for F1 of /iy/

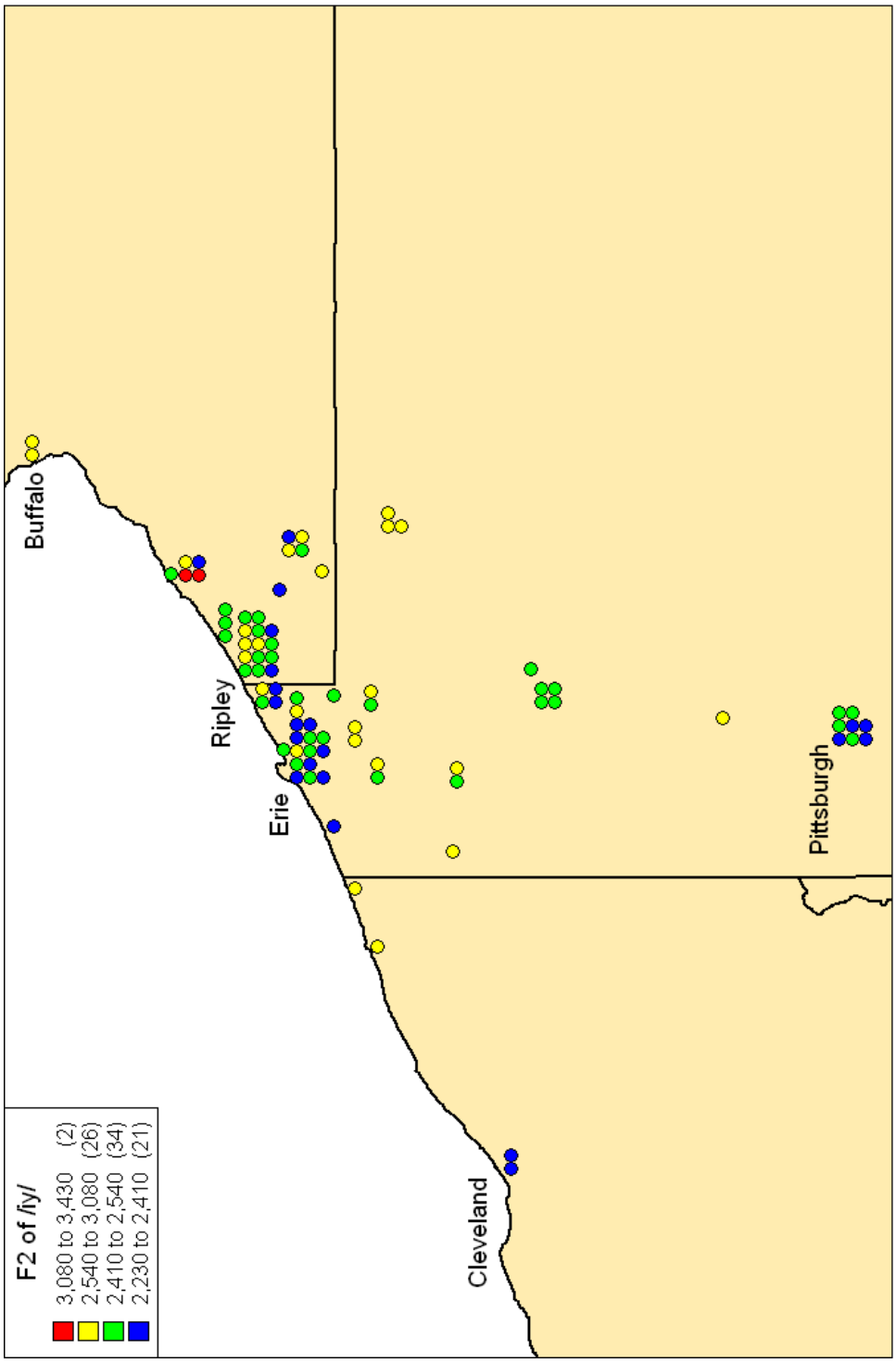

Figure 5.16: Natural break map for F2 of /iy/

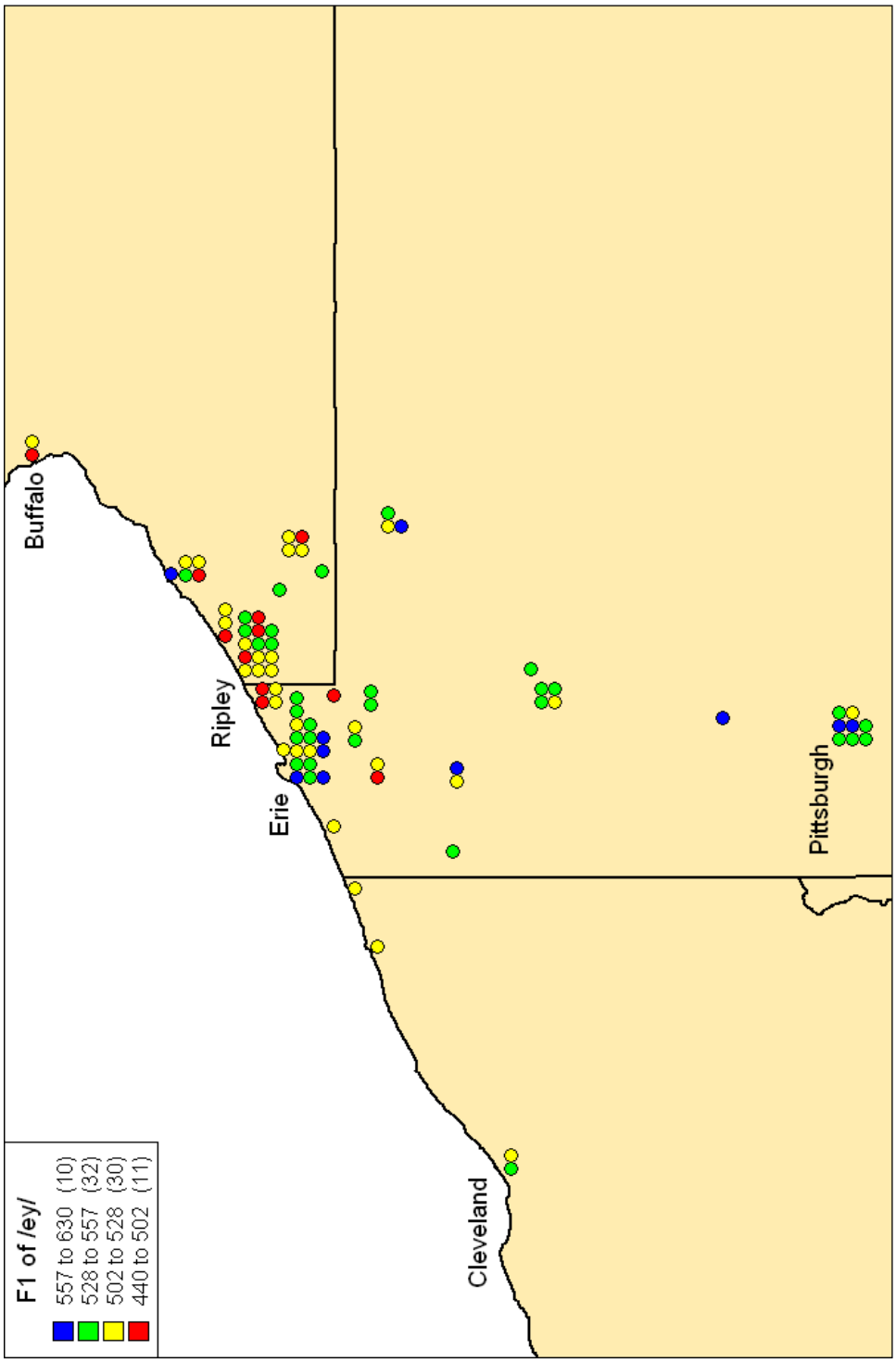

Figure 5.17: Natural break map for F1 of /ey/

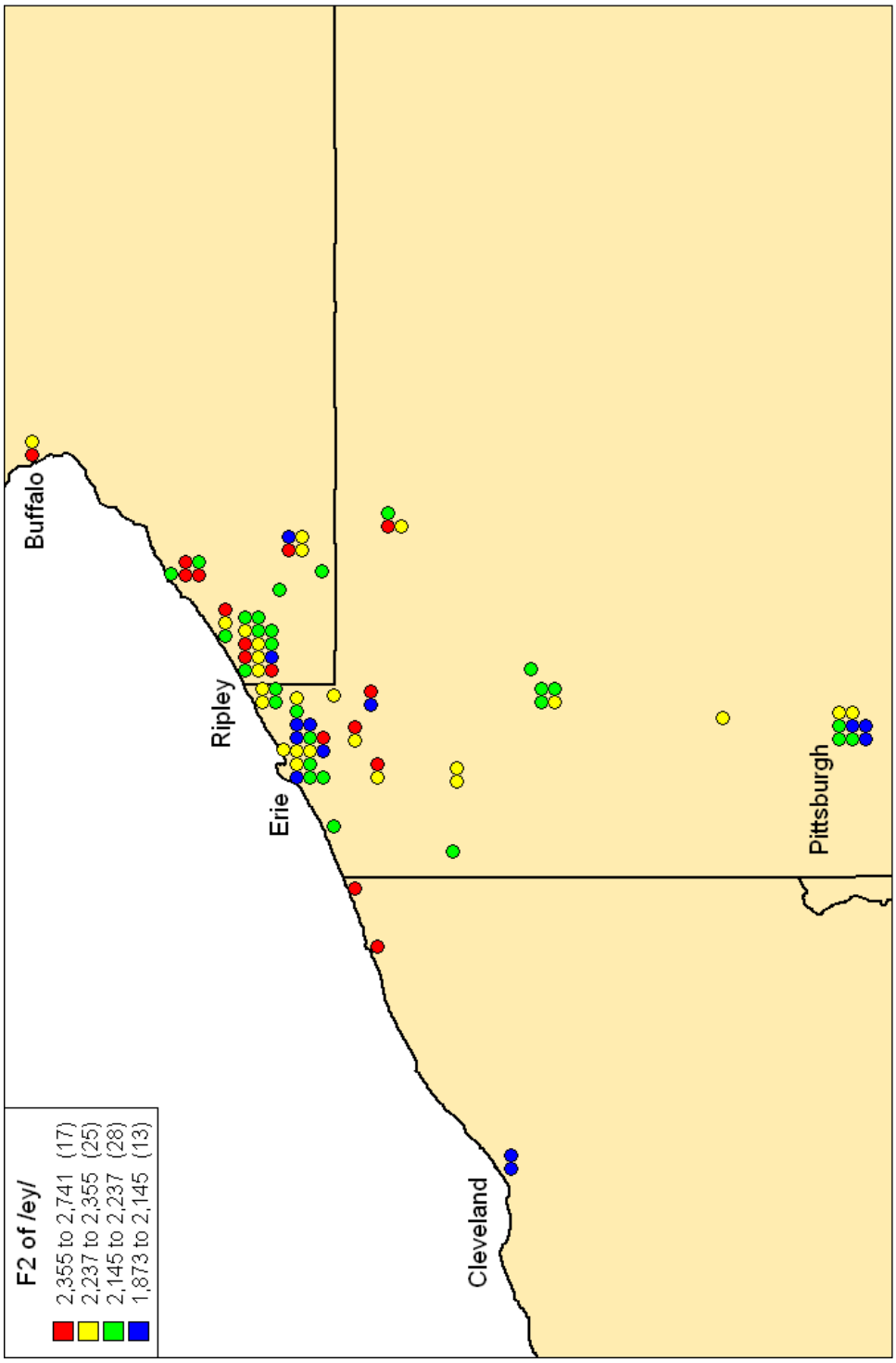

Figure 5.18: Natural break map for F2 of /ey/

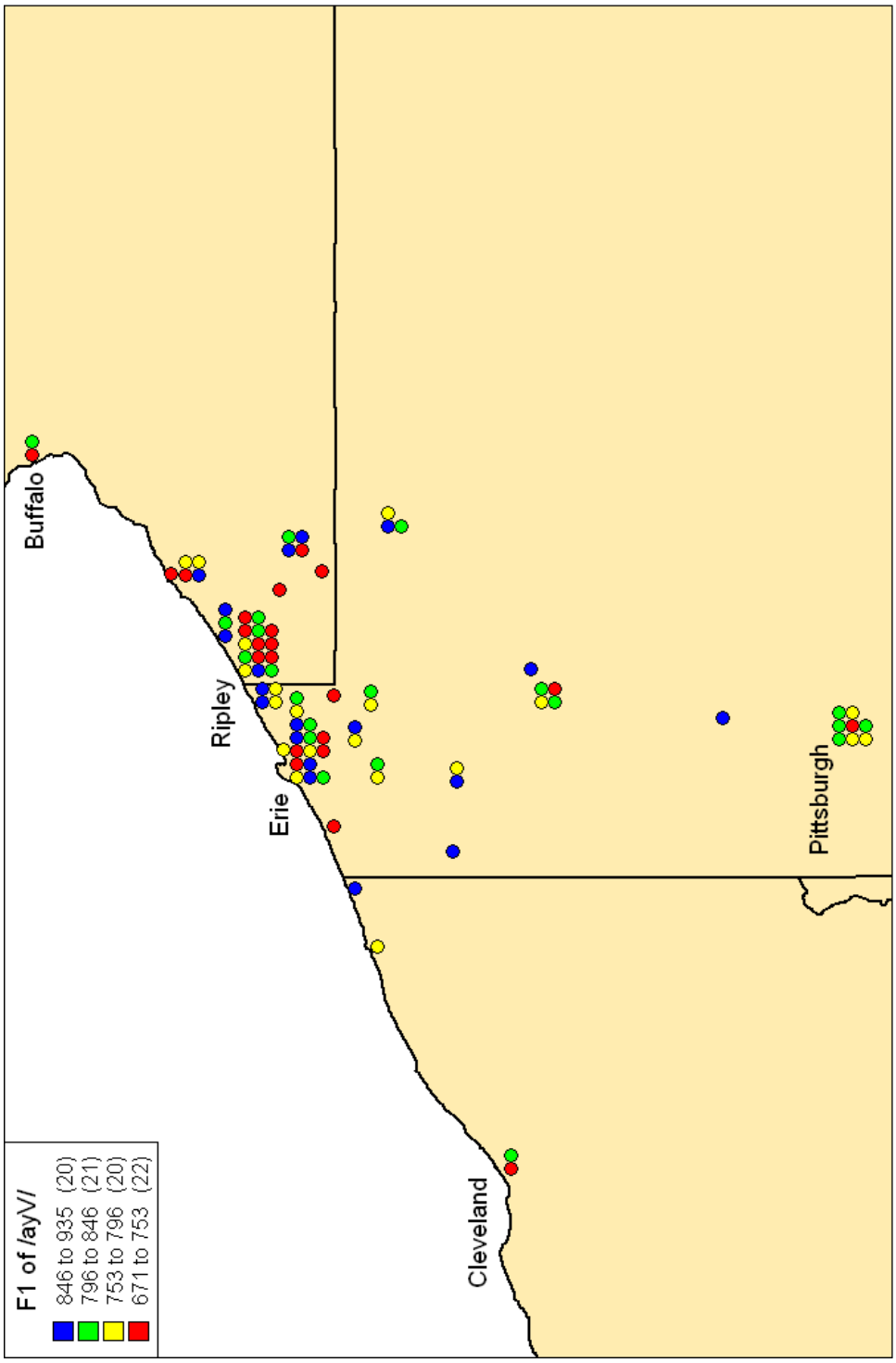

Figure 5.19: Natural break map for F1 of /ayV/ (/ay/ before voiced codas and word-finally)

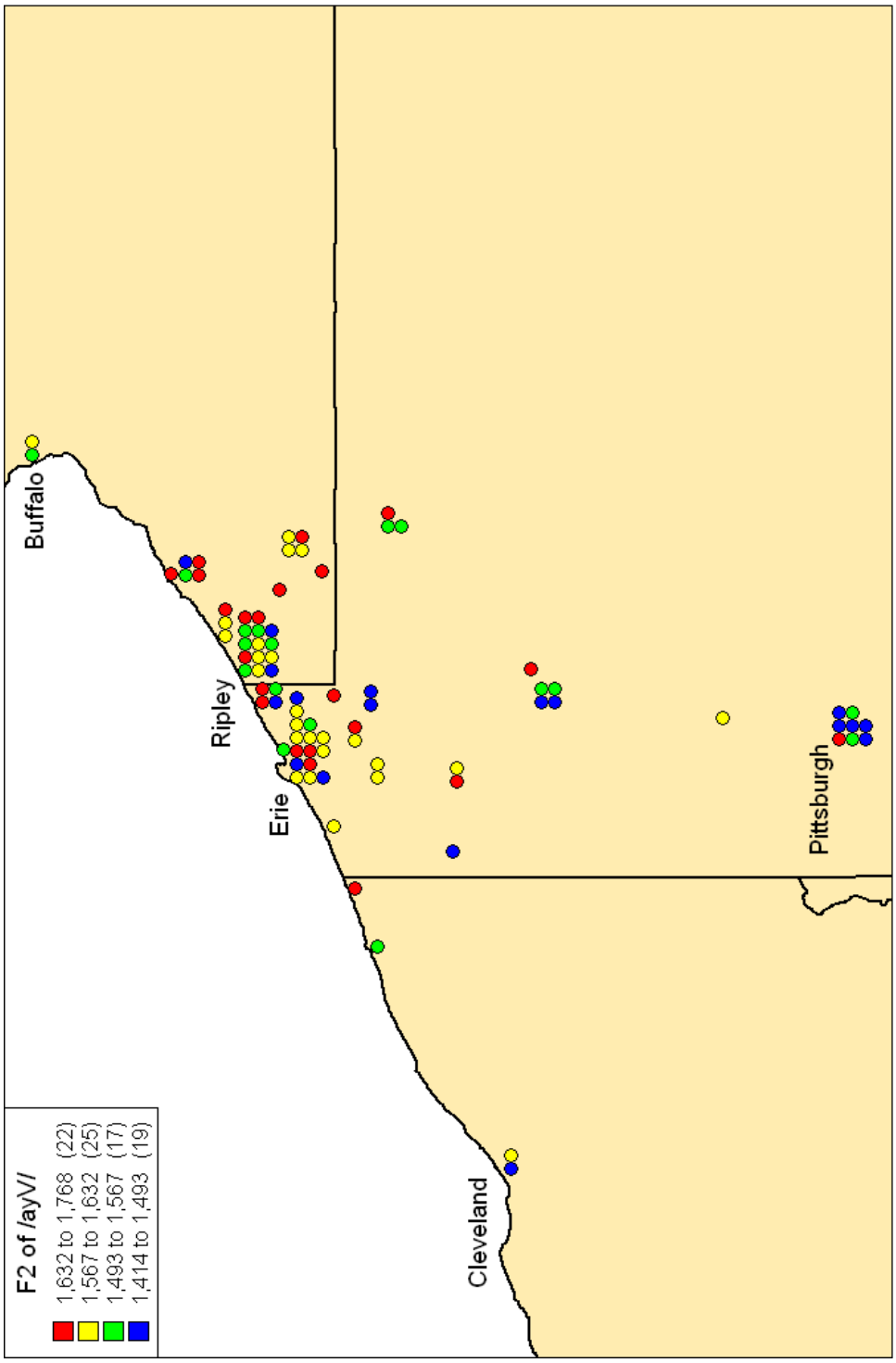

Figure 5.20: Natural break map for F2 of /ayV/ (/ay/ before voiced codas and word-finally)

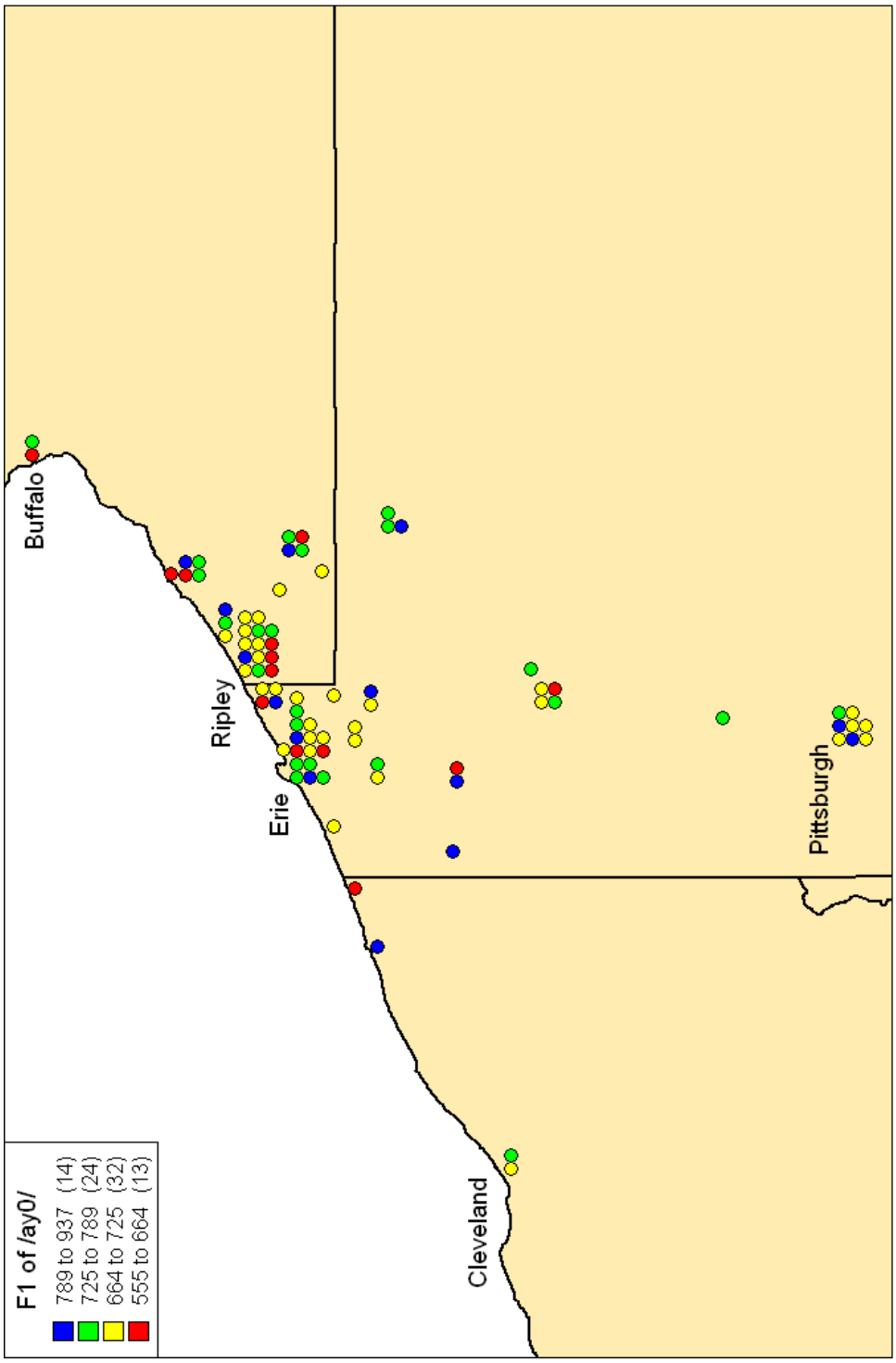

Figure 5.21: Natural break map for F1 of /ay0/ (/ay/ before voiceless codas)

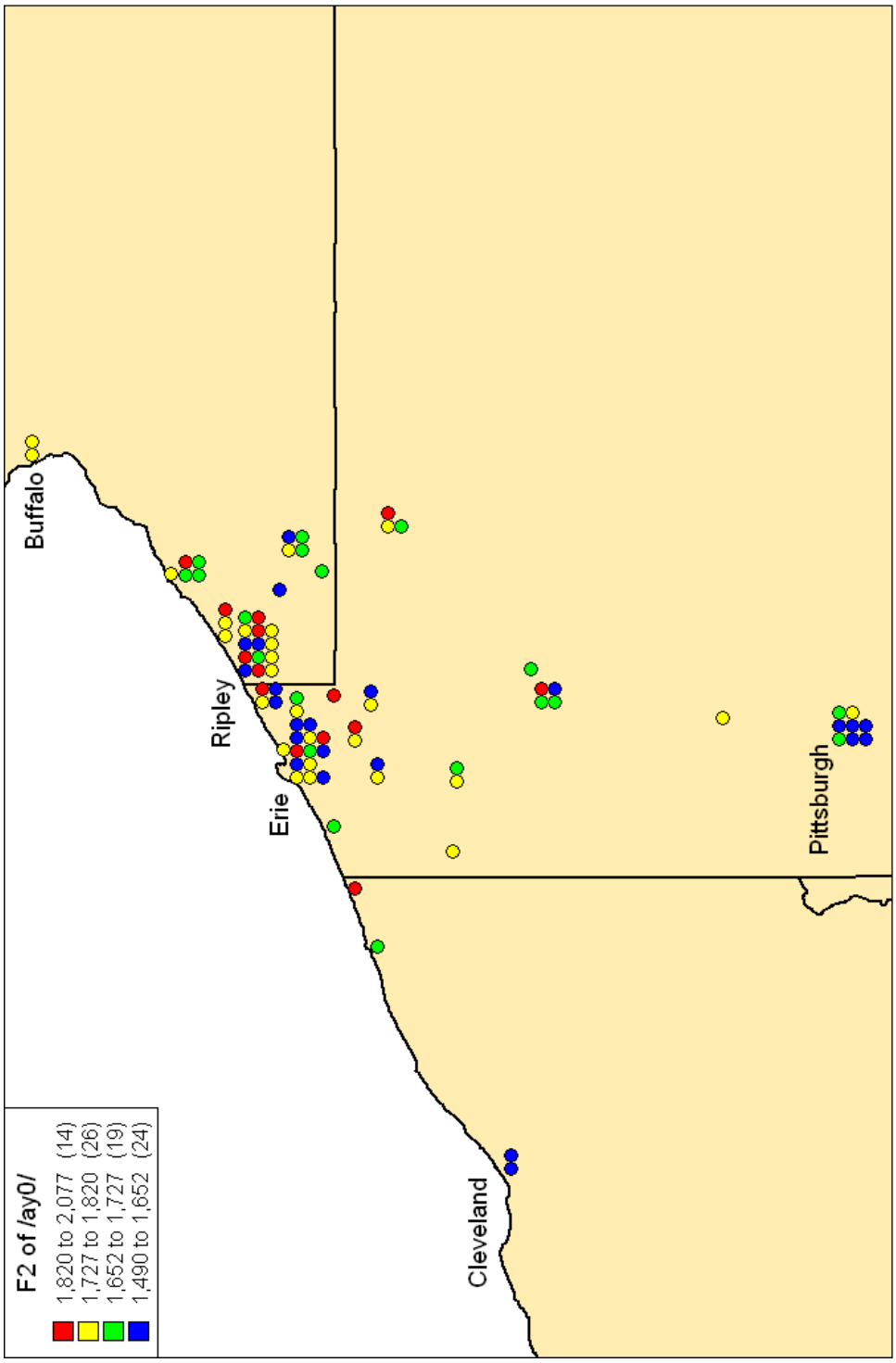

Figure 5.22: Natural break map for F2 of /ay0/ (/ay/ before voiceless codas)

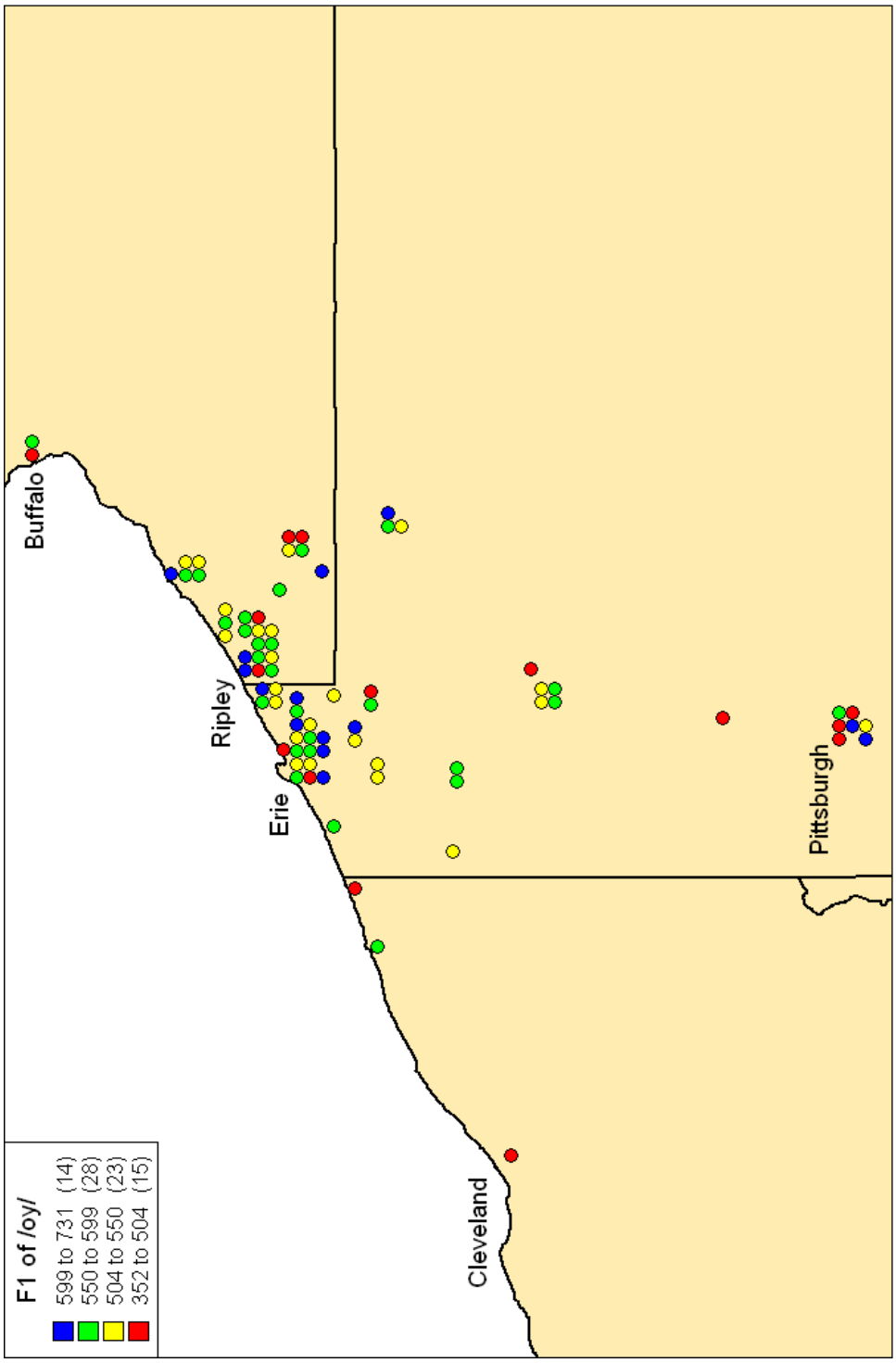

Figure 5.23: Natural break map for F1 of /oy/

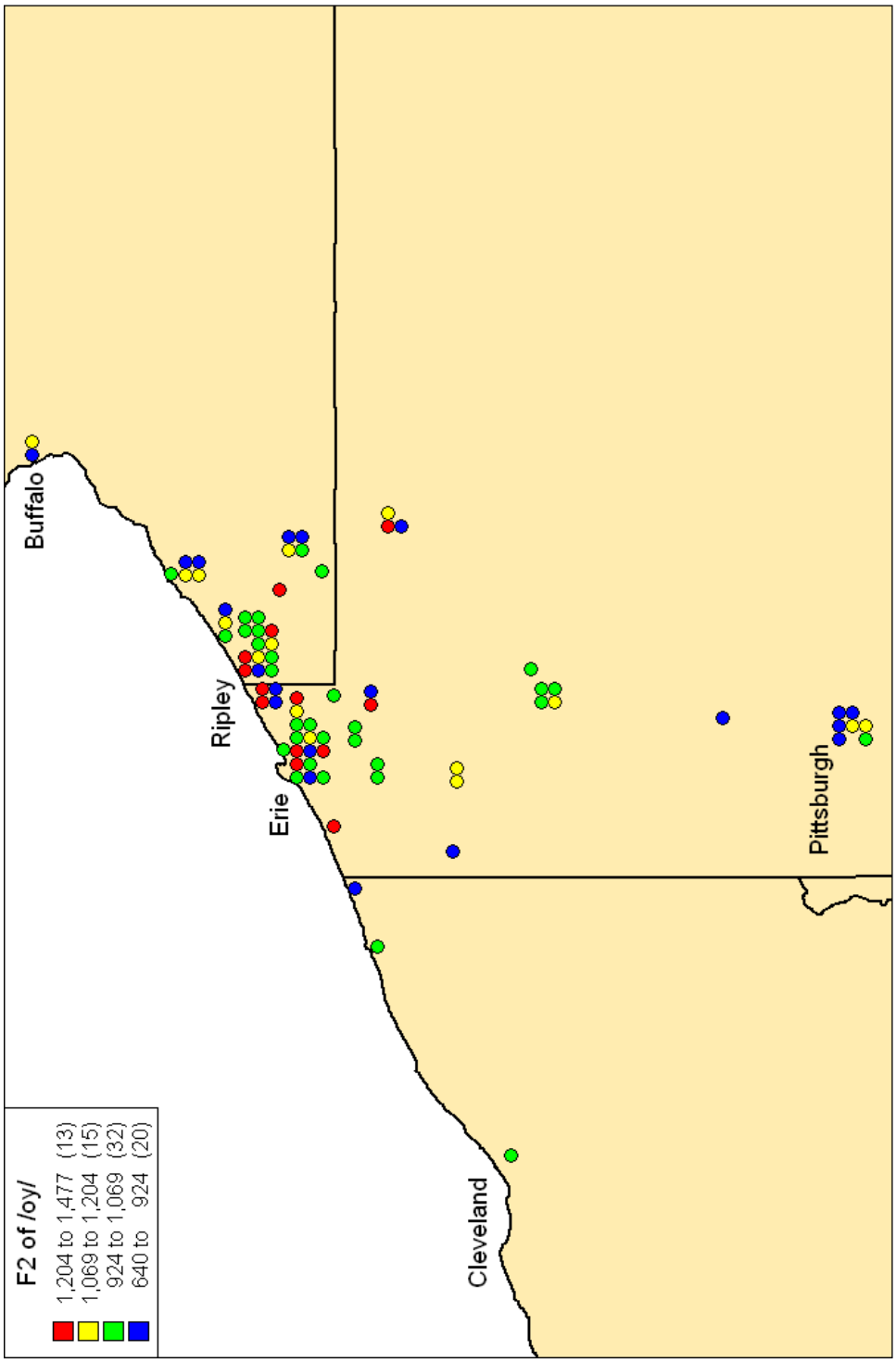

Figure 5.24: Natural break map for F2 of /oy/

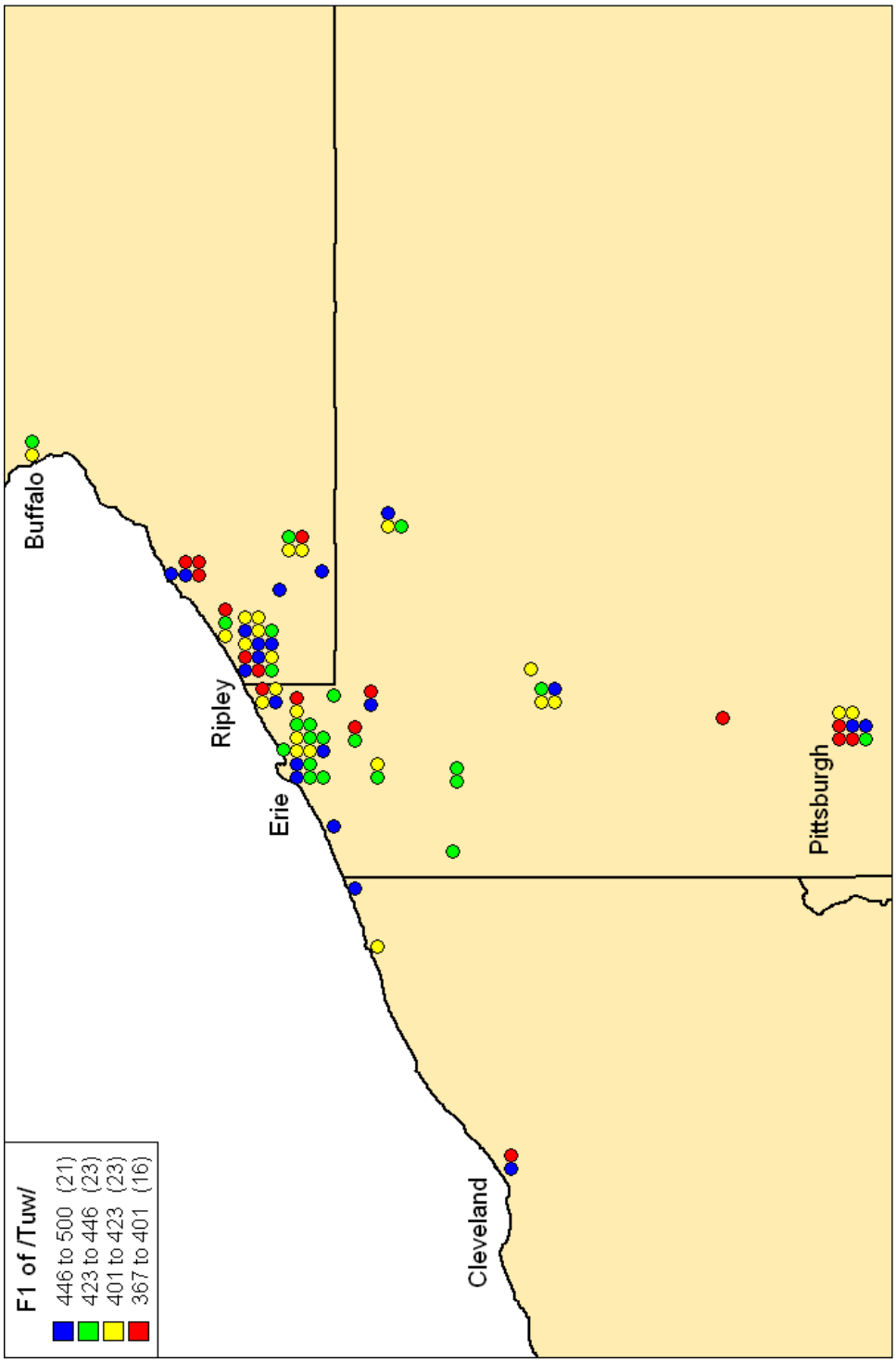

Figure 5.25: Natural break map for F1 of /Tuw/ (/uw/ after coronal onsets)

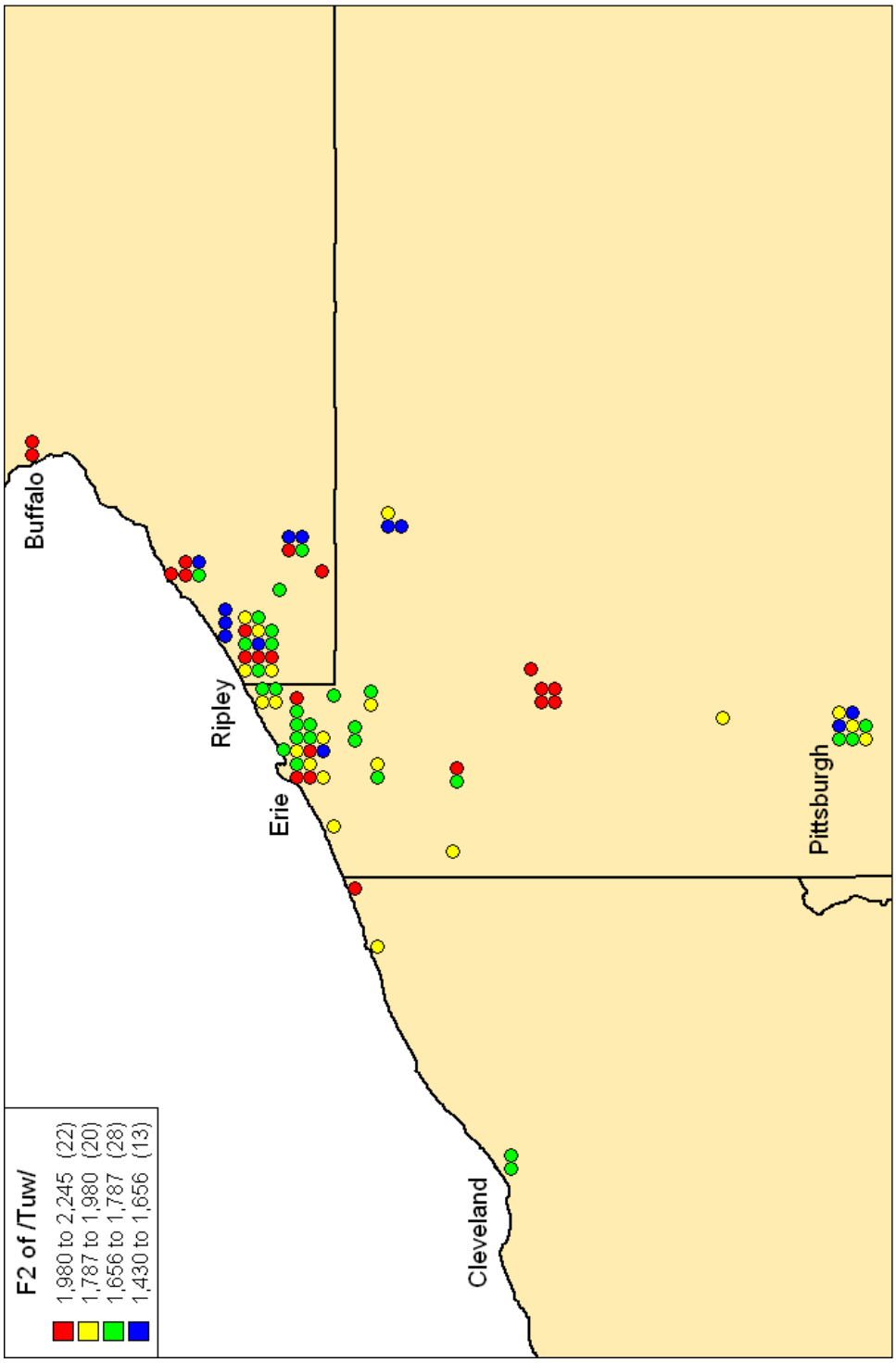

Figure 5.26: Natural break map for F2 of /Tuw/ (/uw/ after coronal onsets)

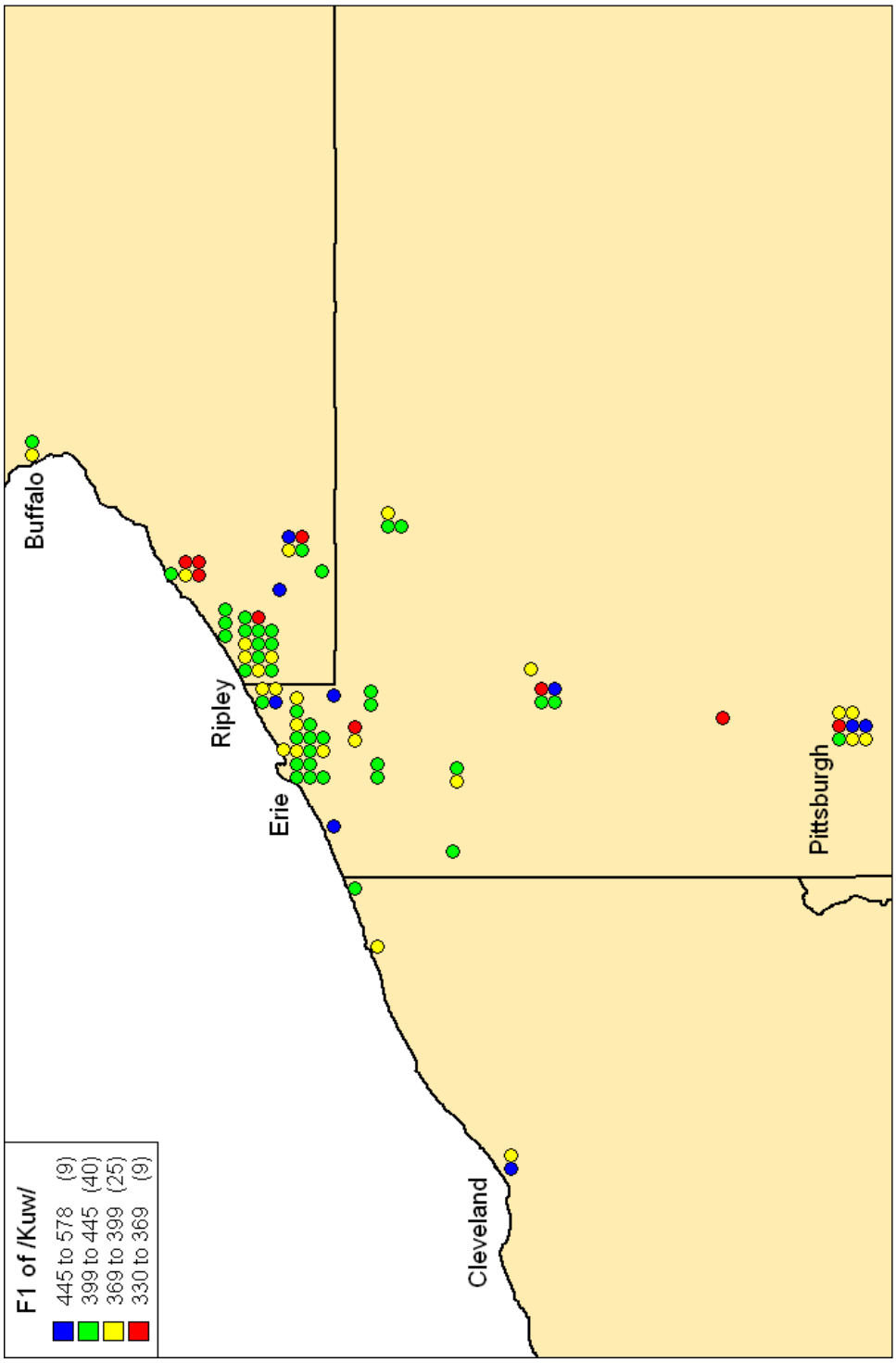

Figure 5.27: Natural break map for F1 of /Kuw/ (/uw/ after non-coronal onsets)

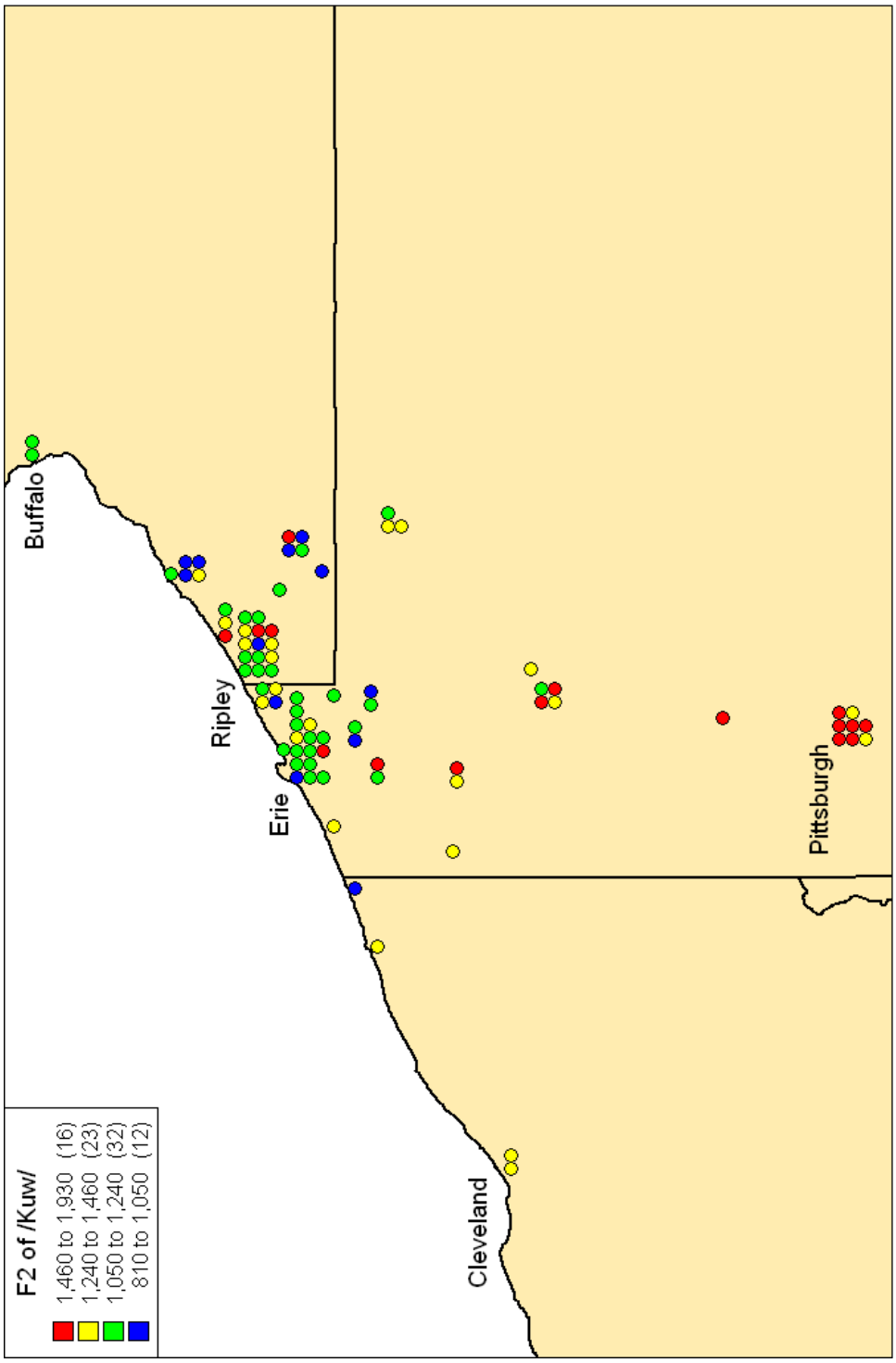

Figure 5.28: Natural break map for F2 of /Kuw/ (/uw/ after non-coronal onsets)

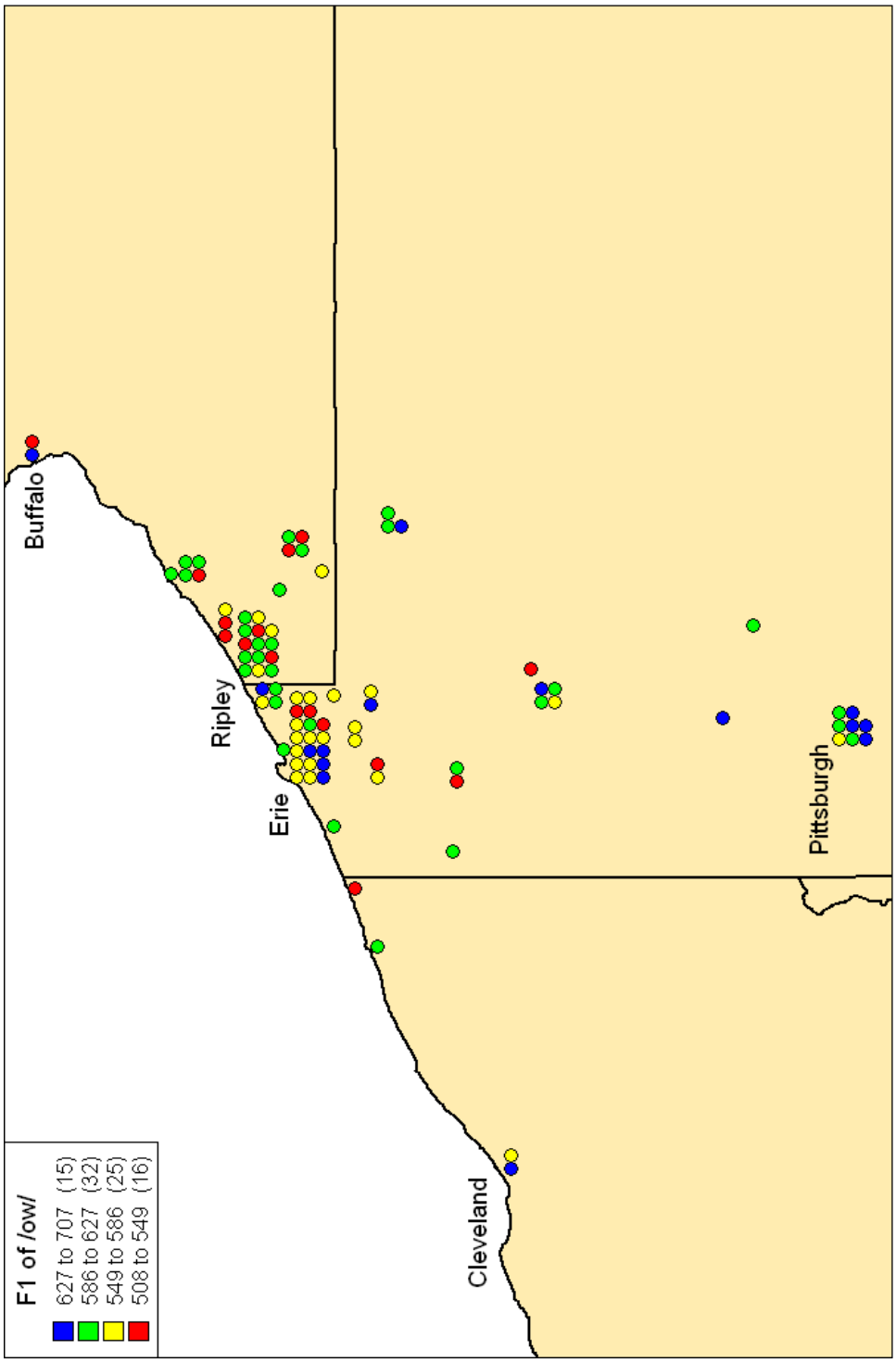

Figure 5.29: Natural break map for F1 of /ow/

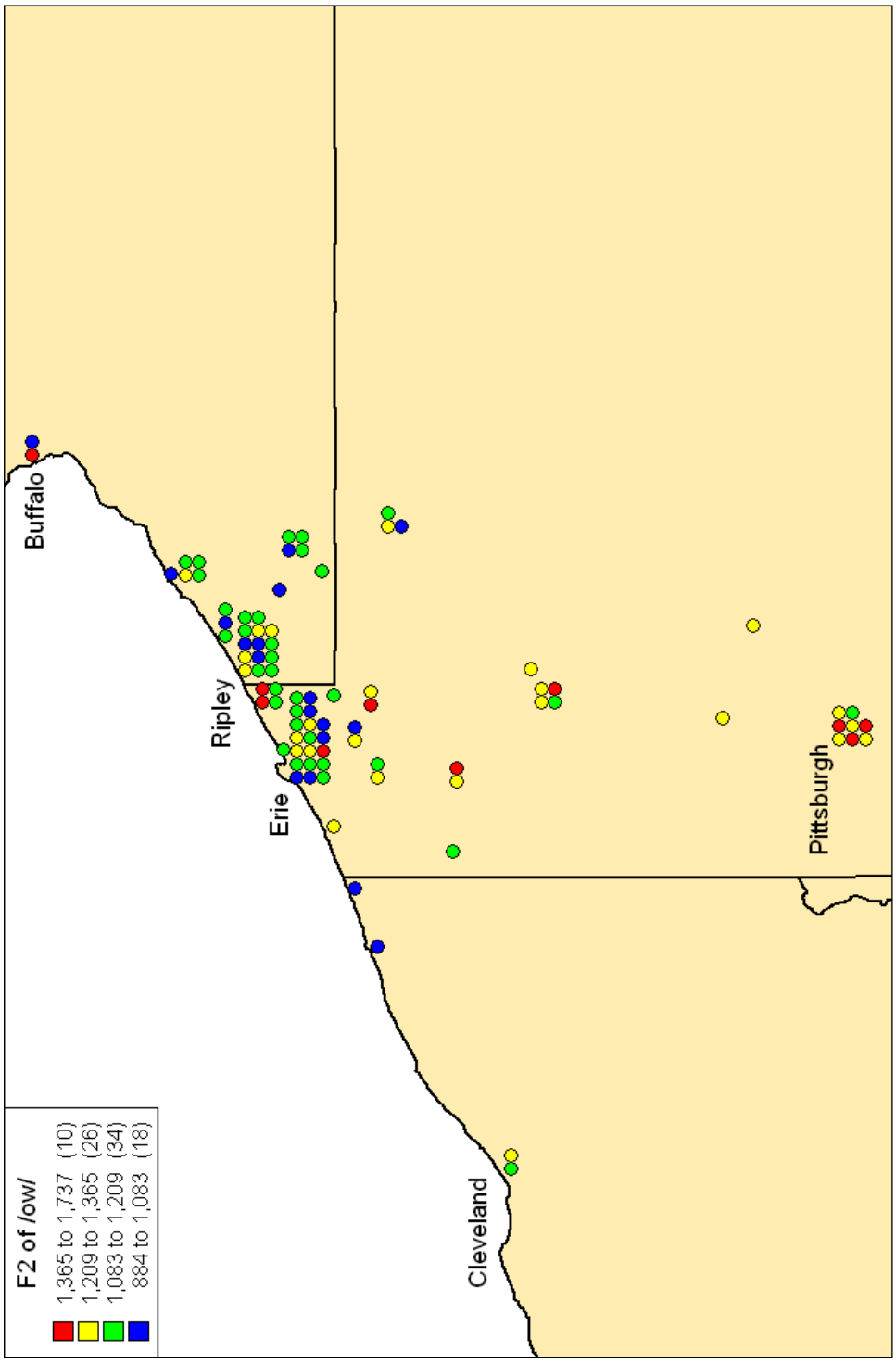

Figure 5.30: Natural break map for F2 of /ow/

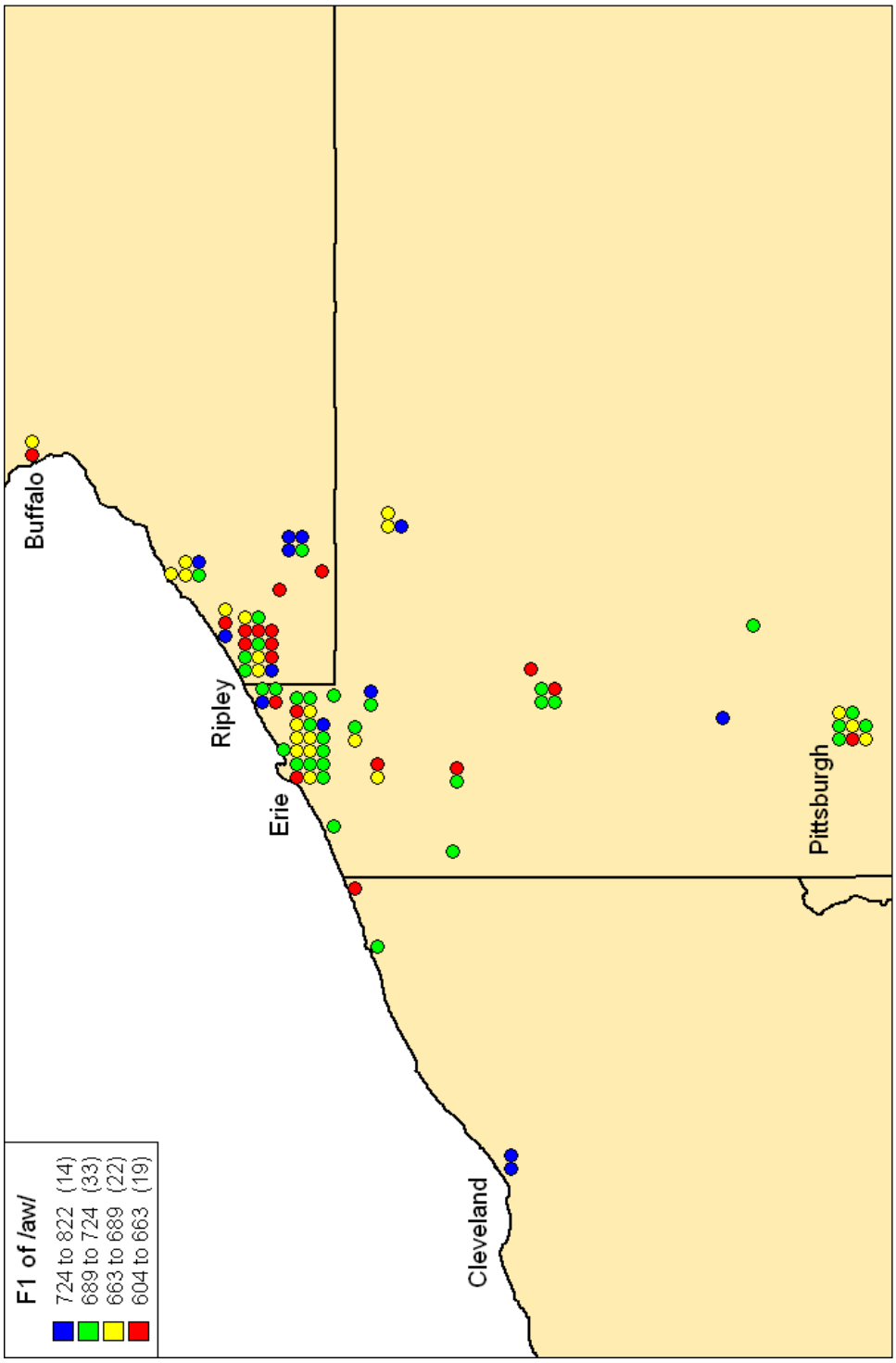

Figure 5.31: Natural break map for F1 of /aw/

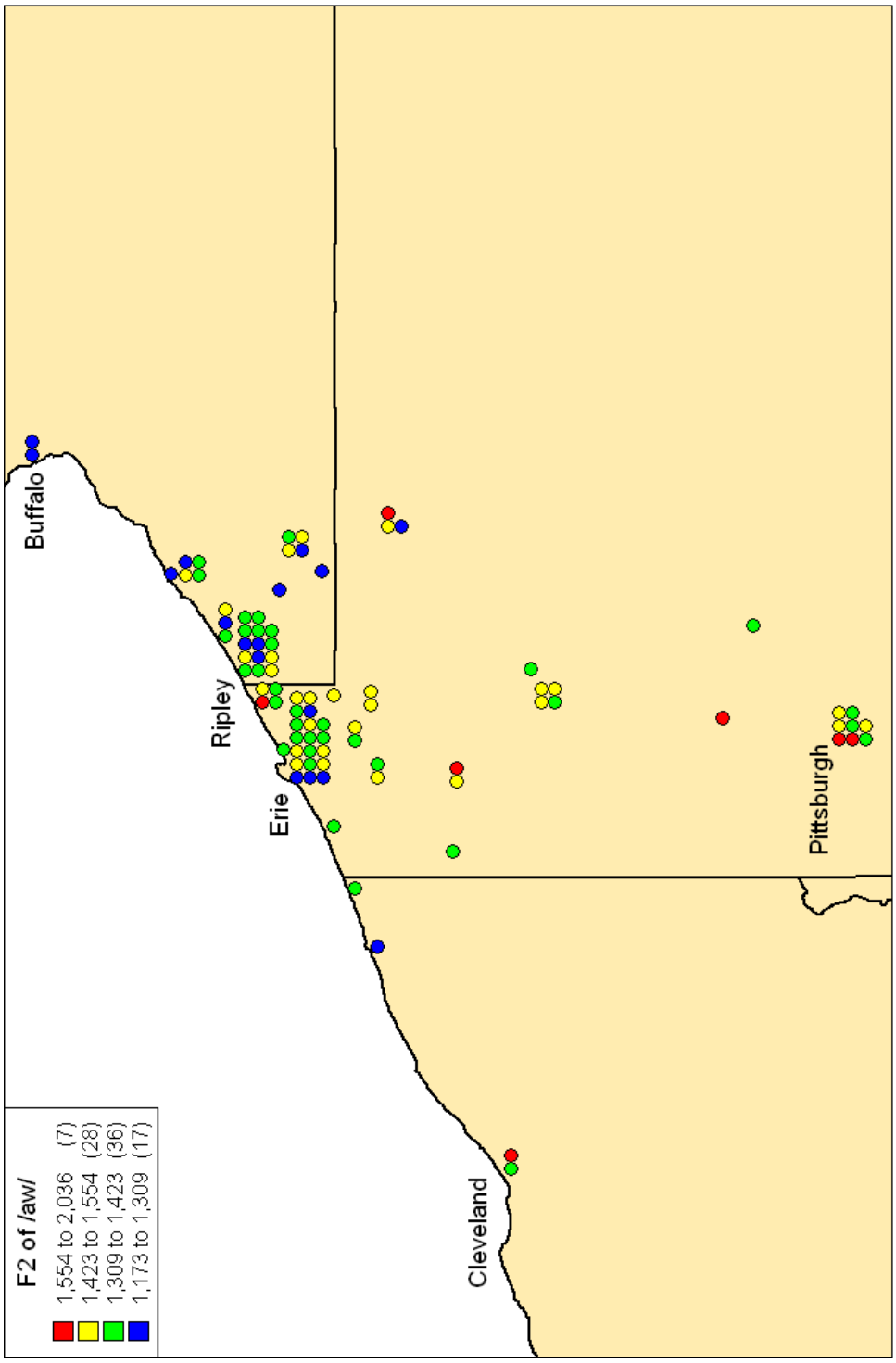

Figure 5.32: Natural break map for F2 of /aw/

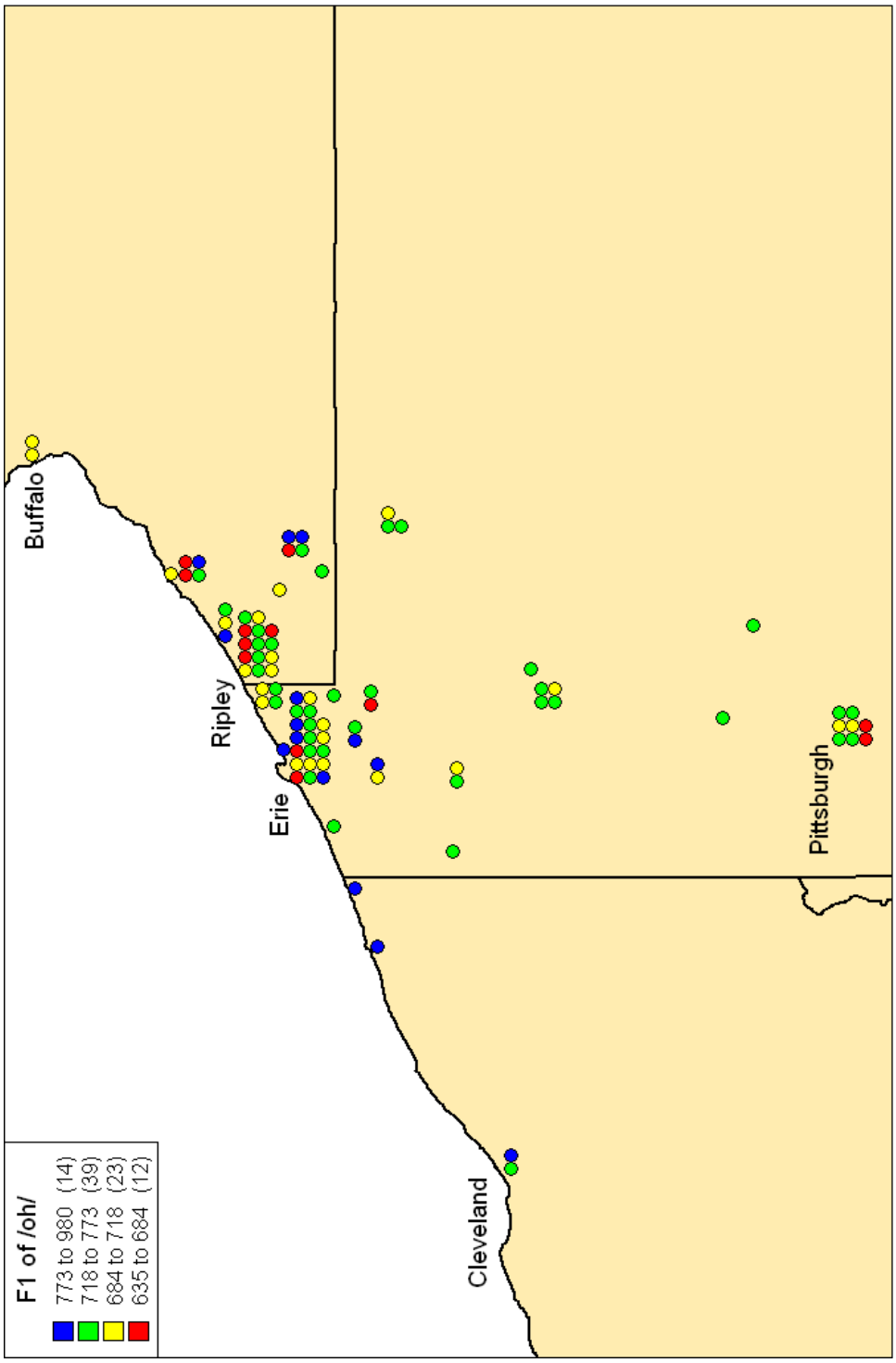

Figure 5.33: Natural break map for F1 of /oh/

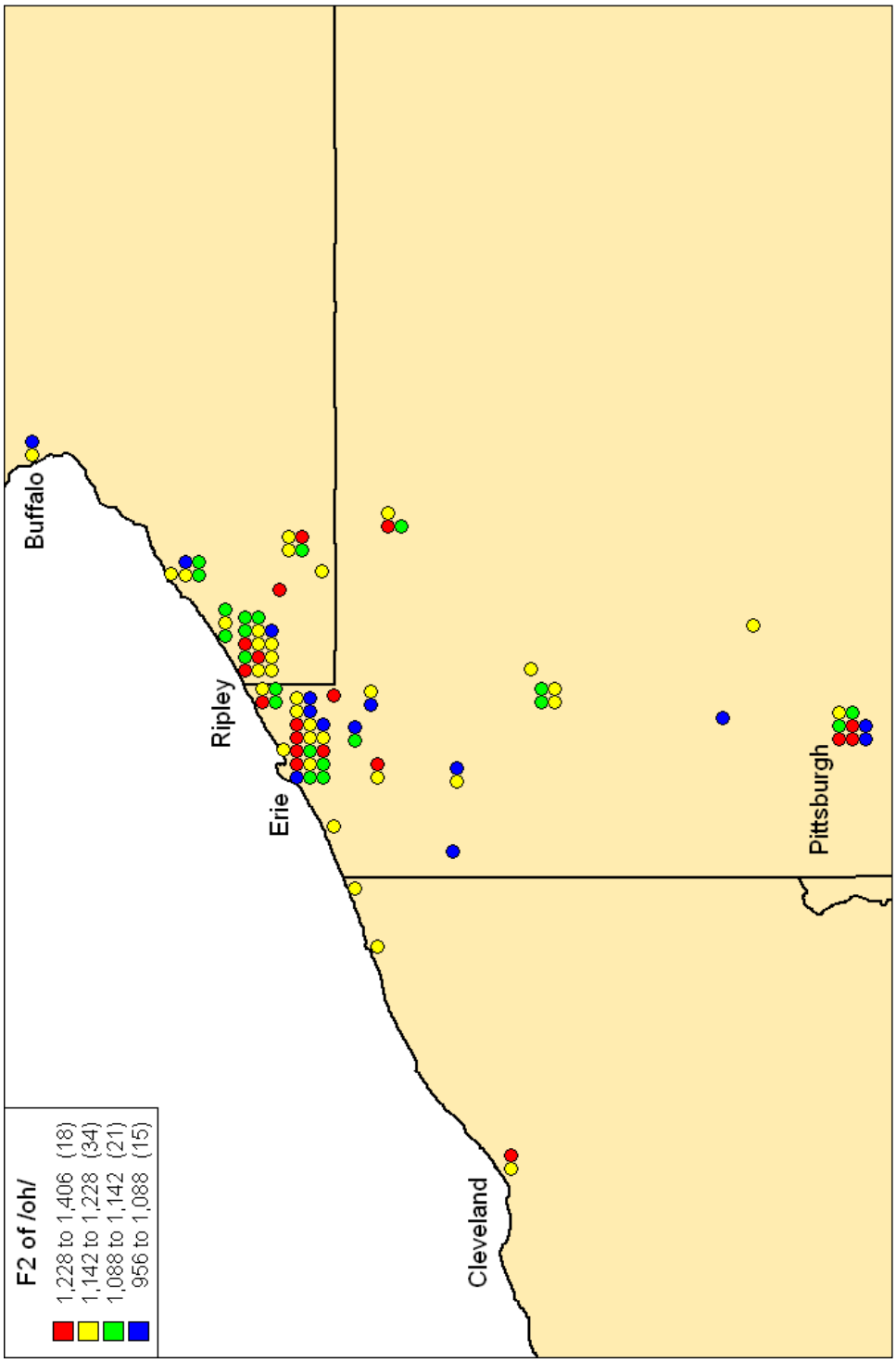

Figure 5.34: Natural break map for F2 of /oh/