Appalachia, USA: An Empirical Note and Agenda for Future Research

Russell Weaver

Texas State University, rcweaver@txstate.edu
APPALACHIA, USA: AN EMPIRICAL NOTE AND AGENDA FOR FUTURE RESEARCH

RUSSELL WEAVER
TEXAS STATE UNIVERSITY

ABSTRACT

Drawing regional bright-lines to separate one “kind” of space from another is often complex and nebulous, if not impossible, in social science research. These circumstances generate two countervailing tendencies: they (1) promote multiple conceptualizations of a given ideational region; and (2) increase demand for a standardized operational definition of that region with which to facilitate intertemporal and interdisciplinary empirical research. These two tendencies animate much of the discourse on “Appalachian” geographies in America. While comparatively theoretically-oriented work emphasizes the unbounded, socially constructed nature of Appalachia, empirical research must often represent the region as a bounded spatial unit. The operational definition commonly used in such empirical analyses is the one established by the federal Appalachian Regional Commission in the 1960s. This research note engages in exploratory spatial data analysis to illustrate some pitfalls that are possible when this political/administrative definition functions as a study area in quantitative Appalachian social research.

For over a century, scholars, politicians, and various other stakeholders have attempted to draw the physical bounds of an American Appalachian region (e.g., Fenneman 1916; Shapiro 1978; Gatrell and Fintor 1998; Williams 2002). Like many comparable efforts, the reasons for regionalization in Appalachian studies range from classification and description (e.g., Moore 2005), to planning, governance, and social activism (e.g., Strickland 1999; Wimberley 2010), and to framing or facilitating scientific inquiry (e.g., Cooper et al. 2011). Simply put, the act of partitioning continuous space into discrete components allows users to abstract away from some innumerable complexities and heterogeneities that otherwise stand in the way of describing, explaining, predicting, and controlling parts of the real world. In this sense, well-defined regional boundaries play a large role in furthering the scientific enterprise. Simultaneously, however, the process of drawing regional bright-lines that separate one “kind” of space or phenomenon from another is often complex and nebulous—if not impossible (Smith 2015)—especially when the region of interest is social or cultural in its constitution. These circumstances seem to generate two oppositional tendencies. First, they set the stage for multiple competing theoretical conceptualizations of a given ideational region. Second, they likely increase the demand for a standardized or “official” operational definition of that region with which to facilitate empirical research across time and academic disciplines.

*Address correspondence to: Russell Weaver, Department of Geography, 601 University Drive, San Marcos, TX, USA 78666. Telephone: (512) 245-2170. Email: rcweaver@txstate.edu
These two counter-tendencies have animated much of the discourse on the location(s) and characteristics of “Appalachia” in the United States. While comparatively theoretically-oriented work in cultural geography (e.g., Ulack and Raitz 1982), anthropology (e.g., Batteau 1990), rural sociology (e.g., Billings 2007), political science (e.g., Fisher 2010), history (e.g., Shapiro 1978), and other disciplines has emphasized the socially constructed and unbounded nature of Appalachia, empirical research on the region—despite engaging with these same issues from similar disciplinary perspectives—is often forced [by analytical methods and data availability] to operationalize Appalachia as a bounded spatial unit that either does or does not contain certain phenomena of interest (e.g., Alexander and Berry 2010; Cooper et al. 2011). Within this literature, perhaps the most widely used spatial representation of Appalachia is the definition established by the federal Appalachian Regional Commission (ARC) in the 1960s for targeting its anti-poverty programs (e.g., Williams 2002; Moore 2005; Alexander and Berry 2010; Cooper et al. 2011; Joshi and Gebremedhin 2012). The present research note highlights some pitfalls that are possible when this political/administrative definition functions as a study area in quantitative social science research on Appalachia.

Questioning the relevance of the ARC borders is not a novel contribution to Appalachian studies (e.g., Watts 1978; Strickland 1999; Williams 2002), not even to its quantitative branch (e.g., Cooper et al. 2011). That being said, this research note goes beyond simple questioning to illustrate, empirically, through exploratory spatial data analysis, mismatches between the ARC boundaries and common (measurable) perceptions of Appalachia. Importantly, the goal here is not to discover or propose a new set of boundaries that somehow captures Appalachia “better” than the ARC’s or any other existing delineation of the region. Rather, what follows is merely an attempt to understand how adopting the ARC boundaries as a study area in empirical investigations of Appalachia might affect the findings and implications of those analyses. Toward that end, the research note draws on geospatial (1) socioeconomic data from the U.S. Census Bureau, (2) physiographic data from the U.S. Geological Survey, Sayre et al. (2014), and the U.S. Department of Agriculture, and (3) business data from Esri, to perform a series of geovisualization and spatial analytical operations.

First, cluster analysis in data space shows that the demographic, socioeconomic, and topographic attributes that are sometimes used to describe Appalachia map onto territories outside the ARC boundaries, but within physiographic and perceptual definitions of the region. One implication of this finding is that the ARC might define Appalachia too narrowly, if one conflates the region with a collection of characteristics that are regularly attributed to it. Second, following sociologist John Shelton Reed (1982, 1983), and contemporary scholars who build on his work
(Alderman and Beavers 1999; Ambinakudige 2009; Cooper et al. 2011), the
distribution of businesses having the character string ‘Appalachia’ in their name is
examined for clustering in geographic space. Previous research has argued that the
spatial patterning of businesses with symbolic/regional names is a “proxy for
regions” (Liesch et al. 2014: 2). In the present research note, the less sweeping
assertion is made that regional naming patterns convey information about the
geographies of Appalachia perceived by local business owners (e.g., Cooper et al. 2011;
Cooper and Knotts 2013). In other words, place-specific business names bring
particular conceptualizations of the Appalachian region to life. While these
conceptualizations may or may not have salience beyond the business community
(e.g., Liesch et al. 2014), at minimum they present a tangible challenge to the ARC
definition of Appalachia (Cooper et al. 2011). The results from analyzing these
naming patterns for spatial clustering reveal that a core region of perceived
Appalachian identity falls within roughly three-fifths of the area enclosed by the
ARC boundaries, thereby implying that the ARC might define Appalachia too
broadly.

These seemingly inconsistent findings—i.e., that Appalachia may extend far
beyond the ARC borders on the one hand, and that it might occupy considerably
less territory than the ARC region on the other—do well to articulate the central
theme of this note. Namely, representing Appalachia is the “impossible necessity of
Appalachian studies” (Smith 2015). It is at this intersection of the impossible and
the necessary where, as argued above, theoretical and quantitative research on
Appalachia are sometimes separated. Engaging in quantitative research requires
one to abstract away from the multiplicity of regional conceptualizations suggested
by theory. Although such abstractions are necessary if they favor a single
regionalization (viz., the ARC definition) over alternatives, then they might be
insufficient to answer the diverse range of social research questions to which they
are applied. This note concludes by commenting on directions for future research
in this topical area.

PHYSIOGRAPHIC AND ARC DEFINITIONS OF APPALACHIA

Given the objectives outlined above, it is beyond the scope of this note to survey
the literature on Appalachian regional geography comprehensively. Instead, this
section offers only a brief and partial treatment of the [arguably] two most popular
and influential definitions of Appalachia that feature in social science research.
Figure 1 superimposes both sets of boundaries onto their respective spaces in the
eastern U.S. In the first case, “initial attempts to define Appalachia were based
on…physical geography” (Gatrell and Fintor 1998, 885-6). Most notably,
Fenneman’s (1916) classification of the conterminous United States into distinct
physiographic divisions defines the “Appalachian Highlands” as a
region—extending from northern Alabama northeasterly to Maine—that is characterized by “an ancient geomorphic history that formed its rock and surfaces...[and which] is very different from surrounding country in relief, lithology, and natural resources” (Ulack and Raitz 1982, 731). This naturalistic definition served as a primary input to early physical geographic and sociocultural studies of the Appalachian region (see the discussion by Gatrell and Fintor 1998). Later work modified Fenneman’s (1916) physiographic Appalachia to create boundaries that aligned with alternative understandings of the region. For example, Raitz and Ulack used the perimeter of upstate New York’s Mohawk Valley (Figure 1) as a cutoff point—all spaces to the north of this area were excluded from their seminal studies of Appalachian identity (Raitz and Ulack 1981a, 1981b; Ulack and Raitz 1982).

In many respects, a physiographic definition seems appropriate for a region that shares its name with a mountain range. Nevertheless, physical geography yielded
to political geography in Appalachian representations soon after a presidential campaign stop by John F. Kennedy exposed the U.S. to the plight of families in rural West Virginia (Strickland 1999). The “poverty and deplorable living conditions [that Kennedy encountered] moved him to sponsor…the Appalachian Regional Commission, to coordinate…programs that would improve the economy of the…upland region” (Raitz and Ulack 1981b, 106). The Appalachian cause was then taken up by President Johnson, who, as part of his War on Poverty, recommended in 1964 that Congress formally establish the ARC, so that development funds could be distributed to its localities. By 1967, the boundaries of the present-day ARC geographic unit were cemented into place (Strickland 1999). The precise spatial footprint of the region is often described as an outcome of “gerrymandering” (Gatrell and Fintor 1998). Several parts of it, especially Mississippi and central Alabama, are altogether divorced from the physiographic Appalachian Highlands (Fenneman 1916). In addition, the particular combination of states intersected by the ARC boundaries is said to reflect strategic political bargaining that shored up legislative support for the commission by creating opportunities for amenable Congressional representatives to funnel its resources to their home districts (Gatrell and Fintor 1998).

In spite of these political critiques, evidence suggests that the ARC boundaries are relatively successful at enclosing spaces that share similar demographic, social, and economic characteristics (Watts 1978). Moreover, the “Appalachian” parts of ARC states are ostensibly different in their socioeconomic compositions from their non-Appalachian counterparts (e.g., Strickland 1999). Taken together, these statistical regularities, and the aforementioned cross-jurisdictional political accession on the relevant boundaries, lend credibility to the ARC administrative unit as representing a distinctive “region”. Perhaps therefore, combined with its unambiguity and (governmental) legitimacy, the ARC circumscription is sometimes adopted as a bright-line in empirical studies of Appalachia. That is, it operationalizes the concepts of “inside” and “outside” of Appalachia for many quantitative investigations (e.g., Alexander and Berry 2010; Cooper et al. 2011; Joshi and Gebremedhin 2012). To illustrate some potential downsides of this practice, the following sections demonstrate that: (1) defining traits of the ARC region also apply to non-ARC territories in the Appalachian physiographic division; and (2) measurable local perceptions of Appalachia exhibit geographic distributions that are at once more compact and not fully enclosed by the ARC borders. Furthermore, these perceptions reveal different “Appalachias” depending on whether the ARC or physiographic definition of the region is adopted as the starting point/study area.
DATA, METHODS, AND RESULTS

The data for this study come from the sources named earlier: the U.S. Census Bureau, the U.S. Geological Survey (USGS), Sayre et al. (2014), the U.S. Department of Agriculture (USDA), and Esri Business Analyst 2014. For facility, all census data were extracted from databases precompiled in the Esri Business Analyst extension for ArcGIS, as opposed to querying the U.S. Census Bureau data tables directly. Before discussing the variables gathered from these sources, two points must be made. First, nearly all empirical exercises summarized below are executed on an inclusive Appalachian “candidate region,” which is the total spatial union of the areas intersected by either the ARC or physiographic boundaries (Figure 1). Second, the geographic unit of analysis is the county subdivision. Whereas the ARC region, and many empirical studies thereof (e.g., Moore 2005), adopt the county as the analytical or organizational spatial unit, counties are coarse-grained geographies that mask finer resolution heterogeneity. While this statement is true of any aggregated spatial unit (county subdivisions included), county subdivisions are assumed to be more realistic arenas (compared with counties) in which collective socialization processes operate to produce intersubjective place-based identities.

Specifically, in the vast majority of states intersected by the Appalachian candidate boundary (16 of 21), county subdivisions are minor civil divisions that correspond to various legal entities, such as boroughs, towns, and townships. For the five states where this definition does not hold, county subdivisions are delineated “in cooperation with state, tribal, and local officials…[in ways that] usually follow visible features…[and each is given a] local name that identifies its location” (U.S. Census Bureau, n.d.[a]). By Tobler’s (1970) famous First Law of Geography, it follows that social phenomena are likely to be more similar within these micro-level and meaningful spatial units than they are throughout comparatively coarse-grained counties.

Figure 2 depicts the candidate region with its county subdivisions symbolized according to the land resource regions (LRR) in which they fall. LRRs are established by the USDA, and they represent groups of “geographically associated major land resource areas” that share physiographic, geologic, climatic, hydrologic, and biologic characteristics (USDA 2006). LRRs do not feature directly in the analyses below; rather, they provide a simple visual means for classifying and describing the lands inside the candidate boundary. Also included in Figure 2 is an indicator of whether a county subdivision is mostly “hilly” or “mountainous.” This information comes from the ecological facet data developed by Sayre et al. (2014, 26), which represents the “finest spatial resolution, globally comprehensive biophysical stratification yet attempted,” and provides “a detailed geospatial delineation of unique physical environments and their associated land cover.” As part of their project, Sayre and colleagues (2014) released a 250m resolution raster dataset that
describes ecological land units (ELUs) across the globe. The ELUs contain information on the defining landforms found in a given space, where landforms are generalized into the categories “Plains,” “Water Bodies,” “Hills,” and “Mountains.” For the current project, these data were joined to the county subdivision boundaries in ArcGIS 10.3 using the Zonal Statistics function, which allowed each county subdivision to be classified according to the landform that makes up most of its land area. In Figure 2, “Hills” and “Mountains” are combined into a single category.

In all, 9,269 county subdivisions are intersected by the Appalachian candidate boundary. However, several of these subdivisions represent areas that are either unpopulated or very sparsely populated, and thus the desired socioeconomic and demographic data (see below) were not available. Moreover, a handful of “edge” county subdivisions that lie mostly outside the candidate boundary (e.g., Manhattan in New York City) were not successfully joined to the ecophysiological datasets described above. The result was that less than 1.5 percent of the county subdivisions were successful in the join process.
subdivisions were dropped from the analytical sample for reasons of missing or incomplete data. The final sample therefore contained the 9,123 county subdivisions pictured in Figure 2. Table 1 presents descriptive statistics for this sample. The U.S. Census demographic and socioeconomic variables listed in the table were chosen for their popularity in the Appalachian literature. Namely, previous studies of the region have asserted that Appalachia is, on balance: (1) rural, which can be operationalized with population density (Reiter et al. 2010); (2) impoverished or working poor (Moore 2005), which can be proxied by the percentage of the population with income below 200 percent of the poverty level (U.S. Census n.d.[b]); (3) racially homogenous and white (Cooper et al. 2011); (4) reliant on extractive and production economies, which is measured here by the percentage of civilian workers in agriculture, mining, and manufacturing occupations (Strickland 1999); and (5) hilly or mountainous (Ulack and Raitz 1982; Smith and Fisher 2012), where this variable was described above. If one adopts an assumption that the co-location of these characteristics distinguish Appalachia from the spaces around it (see Watts 1978; Strickland 1999), then their joint spatial distributions ought to reveal information about the boundaries of an Appalachian region.1 Beyond these variables, Table 1 includes information on the number of businesses in each county subdivision with the character string ‘Appalachia’ in their name. As other researchers have argued, this quantifiable attribute is useful for measuring business owner perceptions of the geographies of an Appalachian region (e.g., Zelinsky 1980; Reed 1982, 1983; Cooper et al. 2011; Lietsch et al. 2014).

K-Means Cluster Analysis

A k-means cluster analysis was performed on the five (non-business) variables enumerated in the preceding section. The k-means algorithm detects clusters in data space, and does not consider geographic connectivity. Explicitly, it identifies an initial seed location based on the data, and it then selects subsequent seeds to be as far as possible in data space from the existing seeds. The number of seeds corresponds to the number of desired clusters or groups (k). Non-seed units are assigned to the seeds to which they are closest in data space. As units are assigned to seeds, the means (centers) of the clusters are iteratively recomputed, and reassignment occurs if a given unit is closer to the center of a cluster other than the one to which it currently belongs. This computation proceeds until an outcome is reached whereby intragroup differences are minimized, and intergroup differences are maximized (Kaufman and Rousseeuw 2005). In other words, the algorithm creates data “regions” characterized by similar values of the selected variables.

1This statement should not be interpreted as the author attaching a positive truth-value to the referenced assumption.
**TABLE 1. DESCRIPTIVE STATISTICS FOR THE SAMPLE (N = 9,123 COUNTY SUBDIVISIONS).**

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>MEAN</th>
<th>STD. DEV.</th>
<th>MIN</th>
<th>MAX</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population Density (persons per square km)</td>
<td>269.43</td>
<td>647.48</td>
<td>0.01</td>
<td>20,004.00</td>
</tr>
<tr>
<td>% of the Population with Income Below 200% of the Poverty Level (% of population classified as “in poverty” or “working poor”)</td>
<td>0.33</td>
<td>0.14</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>% of the Total Population that is White (Not Hispanic or Latino)</td>
<td>0.88</td>
<td>0.15</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>% of Civilian Employees whose Occupation is “Mining,” “Agriculture,” or “Manufacturing”</td>
<td>0.16</td>
<td>0.07</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Majority of County Subdivision is Hilly or Mountainous (1=Yes, 0=No)</td>
<td>0.40</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Count of “Appalachian” Businesses</td>
<td>0.18</td>
<td>0.85</td>
<td>0.00</td>
<td>25.00</td>
</tr>
<tr>
<td>“Appalachian” Businesses as a % of All Businesses (x 100)</td>
<td>0.05</td>
<td>0.28</td>
<td>0.00</td>
<td>11.11</td>
</tr>
</tbody>
</table>

*Dichotomous variable (mean is the fraction of n classified as hilly or mountainous)*

An initial run of the k-means cluster algorithm was executed in ArcGIS 10.3 to determine the optimal number of groups (k) for the analysis. This number was four, the k value for which the mean Calinski-Harabasz pseudo F-statistic was most statistically significant (NB: this test statistic is a ratio comparing intragroup homogeneity with intergroup heterogeneity) (Calinski and Harabasz 1974). Figures 3a and 3b map the results of the k-means cluster analysis for the candidate region relative to the ARC boundaries. Table 2 presents the descriptive statistics for the four groups (data clusters) produced by the method. The maps show that the ARC boundaries preponderantly enclose what are classified in Table 2 as “Rural Uplands” and “Rural-Suburban Uplands.” Consistent with some extant descriptions of Appalachia, the former group is associated (on average) with hills and mountains, low population density, high poverty/working poor populations, significant extractive and production economies, and almost no racial diversity. The latter group is (on average) hilly and mountainous, rural, and racially homogeneous, but has higher average income and is less reliant on extractive and production occupations than the former group.

In the southernmost geographic portions of the ARC region (refer to the red outline), and the Great Valley of eastern Tennessee, most of the county subdivisions...
TABLE 2. GROUP CLASSIFICATIONS AND MEAN INTRA-GROUP VARIABLE VALUES

<table>
<thead>
<tr>
<th>CLUSTER</th>
<th>MEAN % POVERTY</th>
<th>MEAN % WHITE, MINING, &amp; PROD.</th>
<th>MEAN % AG., OCCS.</th>
<th>% HILLY OR MOUNTAINOUS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rural Uplands</td>
<td>96.18</td>
<td>0.45</td>
<td>0.95</td>
<td>0.22</td>
</tr>
<tr>
<td>Rural-Suburban Uplands</td>
<td>148.53</td>
<td>0.27</td>
<td>0.94</td>
<td>0.13</td>
</tr>
<tr>
<td>Rural-Suburban Lowlands</td>
<td>109.17</td>
<td>0.40</td>
<td>0.86</td>
<td>0.20</td>
</tr>
<tr>
<td>Urban-Suburban Lowlands</td>
<td>693.21</td>
<td>0.22</td>
<td>0.81</td>
<td>0.10</td>
</tr>
</tbody>
</table>

are classified as “Rural-Suburban Lowlands.” As Table 2 describes, these areas feature mostly flat lands and valleys, low population density, high poverty, and significant extractive/production economies. Relative to their uplands counterpart, however, they have higher racial diversity. Note well that this statement is not
meant to label these areas as racially diverse—on average, the population is still 86% white and non-Hispanic (Table 2). Thus, the “Rural-Suburban Lowlands” exhibit relative, not absolute, racial diversity in the candidate region. Finally, most of the areas beyond the eastern edge of the ARC boundaries, especially from northern Virginia through the New England coast, are classified as “Urban-Suburban Lowlands.” These territories feature high population densities, flat lands, proportionally low impoverished/working poor populations, relatively insignificant extractive/production economies, and the most racial diversity of all the groups (Table 2; NB: as above, this is not to call these areas racially diverse in an absolute sense—only in a relative sense).

Visually, the ARC eastern border appears to approximate a socioeconomic border. In the main, the county subdivisions immediately east of the ARC belong to different data clusters than the areas inside the region. Simultaneously, Figure 3b shows that the “Rural Uplands” and “Rural-Suburban Uplands” classifications that predominate within the ARC borders also extend beyond them, particularly into northeastern spaces intersected by or proximate to the Appalachian Trail (AT). The spatial concurrence of (1) socioeconomic, physical, and demographic attributes that underlie many perceptions of Appalachia, and (2) the AT, is noteworthy in that the AT is symbolic of “Appalachia” for numerous individuals (Kyle et al. 2003). It therefore “structures many popular definitions” of the region (Cooper et al. 2011, 457). In this sense, while the northern parts of the physiographic Appalachian Highlands are excluded from the ARC’s version of the region, these spaces have at least three properties that might qualify them as “Appalachian.” First, they are within the physical geographic definition of the region (Figure 1). Second, they share the social and demographic characteristics that are sometimes attributed to the region and its peoples by observers (Figure 3a). Third, they are physically connected to a symbolic representation of the region, namely the AT (Figure 3b). For these reasons, studies of Appalachian society that begin their inquiries from the ARC boundaries risk underestimating the extent of the region. More will be said about this implication below.

Detecting Geographic Clusters of Perceived Appalachian Identity

Recall that place-specific business names are tangible means for representing business owners’ perceptions of a region (e.g., Zelinsky 1980; Reed 1982, 1983; Cooper and Knotts 2013). Cooper et al. (2011) applied this reasoning to study perceptions of Appalachian regional “identity” in cities of 10,000 or more people within the ARC borders. Specifically, the authors tallied the number of businesses with ‘Appalachian’ in their name for each city, to calculate a continuous random variable that they called an “A-score” (Cooper et al. 2011). Given each city’s (1) point location and (2) A-score, which was used a surrogate measure of Appalachian
identity, Cooper et al. (2011) interpolated a geographic surface of Appalachian business naming inside the ARC boundaries. This process enabled the authors of that study to identify a “core” area of perceived Appalachian identity within the ARC region that includes eastern Kentucky, northeastern Tennessee, southern West Virginia, western Virginia, and a small piece of western North Carolina (Cooper et al. 2011, 464). The current section draws on Esri Business Analyst data to perform a similar analysis for the candidate study area pictured in Figure 1.

In total, there are 1,747 businesses with the word ‘Appalachia’ in their name (Table 1; hereafter, “Appalachian businesses”) included in the national Esri Business Analyst dataset. Of these, 1,633 (93.5 percent) are located in the candidate study area (Figure 1). This spatial distribution suggests strongly that, despite living in an “increasingly interconnected and mobile society,” Appalachian business naming is a relatively place bound phenomenon (Cooper and Knotts 2013, 7). That being said, the Esri data are used here to assign each county subdivision in the study area an “A-score,” where an A-score in this analysis is the total count of Appalachian businesses in a county subdivision divided by the total number of businesses reported for that subdivision by Business Analyst. This value is then multiplied by 100 to convert it into a percentage. Proceeding in this manner has a marked advantage over earlier, city-level studies (e.g., Alderman and Beavers 1999; Cooper et al. 2011), in that it does not require interpolation. Rather, the full geographic extent of the candidate region is covered, which means that a “core” area of Appalachian business naming can plausibly be detected with known data. Figure 4 maps the spatial distribution of A-scores in the study area. The map shows that Appalachian business naming is clearly concentrated within the ARC region, but not exclusive to it.

One way to explore the distribution from Figure 4 for a “core” area of Appalachian business naming is through local statistics for geographic cluster detection (Rogerson and Yamada 2009). For present purposes, the Getis-Ord Gi* statistic is used to compare the local A-score values for a given county subdivision and its neighbors to the local average that would be expected for that subdivision under a null hypothesis of no spatial clustering or dispersion (Rogerson and Yamada 2009). This operation was carried out in ArcGIS 10.3 for the candidate study area using a neighborhood radius of 100km, which was chosen for its relevance in other regional influence models (e.g., Sinton and Huber 2007). A false discovery rate correction was applied to account for multiple testing (Castro and Singer 2006). The results allow one to identify so-called “hot” and “cold” spots of Appalachian business naming, at prescribed levels of confidence, by comparing the false-discovery-rate-corrected positive and negative Gi* statistics to relevant critical values. The results from this exercise are summarized in Figure 5a, where
a statistically significant cluster, or contiguous “core” of Appalachian business naming, extends from north Georgia in the south to western Maryland in the north. A less significant, small, and disconnected cluster is also found in central Pennsylvania. This latter area lies close to the AT (Figure 5a), and it marks a zone of transition from plains to hills (Figure 2). For these reasons, it is plausible that the demographic, social, and physical attributes of this area (Figure 3), as well as its nearness to the AT, combine to give local business owners the perception that it is part of “Appalachia.”

There are at least three more notable takeaways from Figure 5a. First, the “core” area of Appalachian business naming—inclusive of the cluster in Pennsylvania—aligns almost perfectly with the representation of Appalachia made by roughly half the regional “insiders” who participated in Ulack and Raitz’s cognitive mapping studies (1982, 742 [at Figure 4]). This impressive correspondence between Appalachian resident perceptions and the clustering of
APPALACHIA, USA

Figure 5. Spatial Cluster Analysis Results for (A) The Full Candidate Region and (B) The ARC Spatial Definition.

Appalachian business names lends credibility to the latter as a measure of perceptual regions (e.g., Zelinsky 1980; Reed 1982, 1983; Cooper et al. 2011). Second, the “core” area depicted in Figure 5a is substantially larger than its counterpart from comparable city-level studies that rely on spatial interpolation (Cooper et al. 2011, 464). As suggested above, the known, county subdivision-level data leveraged here cover the entire study area, which means that no interpolation is needed to detect clustering in the A-score distribution. In that sense, the hotspot shown in Figure 5a is conceivably a more representative “core” of Appalachian business naming relative to earlier attempts. Finally, observe that despite their (1) proximity to the AT (Figure 3b), (2) demographic and socioeconomic similarities to county subdivisions inside the ARC (Figure 3a), and (3) isolated patches of relatively high A-scores (Figure 4), no areas to the northeast of the ARC boundaries were found to contain high clusters of Appalachian business naming. In fact, virtually no areas north of the Mason-Dixon Line, which separates Pennsylvania and Maryland, are within the business name hot spots (save for the small Pennsylvania cluster already discussed). In contrast to the previous set of results, this finding implies that the ARC boundaries might overestimate the extent of the Appalachian region.

Apart from the above, one of the most consequential points associated with Figure 5 is that local cluster detection methods like Gi* compare local (neighborhood) data with global (study area) data (Rogerson and Yamada 2009).
Thus, the “core” shown in Figure 5a is dependent upon the bounds of the study. Performing an identical analysis on only those county subdivisions that fall within the ARC boundaries—i.e., adopting the ARC definition as the study area—produces a much smaller “core” that essentially excludes Ohio, Maryland, Pennsylvania, and large portions of Kentucky and West Virginia (Figure 5b). This result illuminates the dangers in making inferences about Appalachian social phenomena when starting from the ARC boundaries. Hence, if there is no a priori reason to assume that the ARC borders act as functional constraints on the social or behavioral phenomena under investigation, then quantitative researchers may wish to revisit their conceptual models and identify more appropriate geographic frames for their analyses (for similar arguments from urban studies, see: Kwan 2012, and Weaver 2014).

CONCLUSIONS

Representing Appalachia is the “impossible necessity of Appalachian studies” (Smith 2015): impossible, in that there are “many Appalachias” (Fisher and Smith 2012); and necessary, to the extent that analyzing regions typically requires spatial delimitation (e.g., Williams 2002). From this nexus of impossibility and necessity emerge countervailing tendencies toward (1) multiple theoretical conceptualizations of Appalachia, and (2) the practical search for a standardized operational definition with which to facilitate intertemporal and interdisciplinary quantitative research. Concerning the latter, the “official” Appalachian regional definition established by the ARC, an agency of the U.S. federal government, has proven useful. Besides its unambiguity, constancy, and legitimacy as an administrative jurisdiction, statistical evidence suggests that the ARC region exhibits relative internal homogeneity in its demographic and socioeconomic composition when compared with the territories that surround it (Watts 1978; Strickland 1999). Whether for these or other reasons, the ARC definition recurrently enters quantitative Appalachian research—sometimes to operationalize the concepts of “inside” and “outside” of the Appalachian region (e.g., Alexander and Berry 2010; Joshi and Gebremedhin 2012); and other times to set the bounds of a study area in which to ultimately challenge the ARC delineation (e.g., Cooper et al. 2011).

This research note performed exploratory spatial data analysis to communicate some potential downsides to using the ARC boundaries in either of these ways. In the first place, when Appalachia was conceptualized as a set of socioeconomic, demographic, and physical characteristics, the ARC borders were found to underestimate the joint spatial distribution of those characteristics (Figure 3). Second, when patterns of Appalachian business naming practices were analyzed for spatial clustering, the ARC borders overestimated the extent of the resultant region (Figure 5a). Moreover, when this spatial analysis was re-executed exclusively on the
county subdivisions inside the ARC boundaries (i.e., when the ARC region served as the study area), a much smaller “core” of Appalachian business naming was detected relative to the analysis carried out on the combined physiographic/administrative candidate region (Figure 5). Collectively, these results have two important implications for quantitative Appalachian studies. First, they reaffirm well-established claims that there is no clear cut, generalizable solution to the question of “where is Appalachia?” (Raitz and Ulack 1981a, 212). Second, and more significantly, they caution that employing the ARC definition to overcome this “multiple Appalachias” (Fisher and Smith 2012) challenge in empirical research might lead to inconclusive or specious inferences depending on the question at hand. That is, as far as quantitative findings hinge on study area definition (e.g., Figure 5), inferences in empirical Appalachian research might be based on too much or too little information (refer to Figs. 3 and 5).

At first these conclusions might make the future of quantitative Appalachian studies sound bleak. Much to the contrary, however, they open several doors for establishing vital new research programs. Perhaps the most fruitful of these opportunities is for quantitative researchers to undertake comparative empirical studies that seek to answer the same research question(s) for multiple operational definitions of Appalachia (e.g., Griffin and Thompson 2002). In addition, cognitive mapping exercises like those facilitated by Raitz and Ulack (1981a; Ulack and Raitz 1982) can be updated to leverage new Geographic Information Systems (GIS) technologies, and to include a wider range of participants and stakeholders. Such studies would allow researchers to map and combine diverse perceptions of Appalachia, which could feature in a variety of empirical analyses. Relatedly, future work could even incorporate the Canadian sections of the Appalachian uplands into studies of the region, to determine whether the physical phenomena that appear to separate certain populations in America from their non-Appalachian counterparts (e.g., Strickland 1999) perform similar functions across national borders.

Similarly, Appalachia—wherever it is, has been, or will be located—is a promising and plentiful source of empirical rural social research. Yet this statement is only true insofar as we see the region and quantitative research on it as existing in a dialectical relationship. The more the ARC boundaries operationalize Appalachia in empirical research—even in cases where the end goal is to challenge the ARC boundaries (see especially the difference between Figure 5a and Figure 5b)—the greater will be the ARC’s influence on our perceptions, descriptions and explanations of “Appalachian” phenomena. In turn, this influence limits our abilities to produce new insights through “creative marginality,” or working on the margins, outside familiar constructs (Lobao 1996). Thus, whereas this research note may not have answered outstanding questions about the geographies of Appalachia, my
hopes are that it generates a lot more of them—and that these questions in turn inspire scores of new, critical waves of quantitative Appalachian studies.

**AUTHOR BIOGRAPHY**

**Russell Weaver** is an Assistant Professor in the Department of Geography at Texas State University. He studies place-based cooperation and collective action, processes of community change, and relationships between spatial contexts and individual and group behavior. His recent work appears in the *Annals of the Association of American Geographers*, *The Professional Geographer*, the *Journal of Geography*, and *Ecological Economics*, among other journals.

**REFERENCES**


