Head/Tail Breaks: A New Classification Scheme for Data with a Heavy-Tailed Distribution

Bin Jiang

Head/Tail Breaks: A New Classification Scheme for Data with a Heavy-Tailed Distribution*

Bin Jiang
University of Gävle, Sweden

This article introduces a new classification scheme—head/tail breaks—to find groupings or hierarchy for data with a heavy-tailed distribution. The heavy-tailed distributions are heavily right skewed, with a minority of large values in the head and a majority of small values in the tail, commonly characterized by a power law, a lognormal, or an exponential function. For example, a country's population is often distributed in such a heavy-tailed manner, with a minority of people (e.g., 20 percent) in the countryside and the vast majority (e.g., 80 percent) in urban areas. This new classification scheme partitions all of the data values around the mean into two parts and continues the process iteratively for the values (above the mean) in the head until the head part values are no longer heavy-tailed distributed. Thus, the number of classes and the class intervals are both naturally determined. I therefore claim that the new classification scheme is more natural than the natural breaks in finding the groupings or hierarchy for data with a heavy-tailed distribution. I demonstrate the advantages of the head/tail breaks method over Jenks's natural breaks in capturing the underlying hierarchy of the data. **Key Words: data classification, head/tail division rule, hierarchy, natural breaks, scaling.**

本文引入了一个新的分类方案—头/尾间隔—来对具有重尾分布的数据进行分组或分层。重尾分布严重右偏，少数的高值集中在头部，大部分的低值集中在尾部，通常由乘幂定律，对数正态分布，或指数函数来表征。例如，一个国家的人口分布经常以这样一个重尾的方式分布，少数人 (例如，20%) 住在农村，绝大多数 (例如80%) 住在城市地区。这种新的分类方案把所有的数据值平均分为两部分，并对在头部（高于平均值）的数值继续重复这一过程，直到头部值不再是重尾分布。因此，类的数目和类的间隔都自然地被确定。因此，我认为，在给具有重尾分布的数据寻找组别或层次时，这个新的分类机制比自然间隔分段法更自然。我演示了该头/尾间隔方法在获取数据的基础层次时，相对詹克斯的自然间隔分段法的优势。关键词：数据分类，头/尾分类规则，层次，自然间隔，缩放。

Este artículo presenta un nuevo esquema de clasificación—rupturas de cabeza/cola—para establecer agrupamientos o jerarquía para datos con una distribución de cola pesada. Las distribuciones de cola pesada están fuertemente sesgadas hacia la derecha, con una minoría de valores grandes a la punta y una mayoría de valores pequeños en la cola, comúnmente caracterizados por una ley de potencia, un log-normal, o una función exponencial. Por ejemplo, una población de un país a menudo se distribuye con mayor concentración en la cola, con una minoría de gente (por ej., el 20 por ciento) en el campo y una vasta mayoría (por ej., el 80 por ciento) en las áreas urbanas. En este nuevo esquema de clasificación la partición de todos los datos se hace alrededor de la media en dos partes y se continúa el proceso iterativamente para los valores (por encima de la media) en la cabeza hasta que los valores de la parte de la punta cesan de ocurrir en distribución de cola pesada. Entonces, el número de clases y los intervalos de clase quedan determinados naturalmente. Sostengo, en consecuencia, que el nuevo esquema de clasificación es más natural que las rupturas naturales para establecer los agrupamientos o jerarquías para datos con una distribución de cola pesada. Demuestro las ventajas del método de las rupturas cabeza/cola frente al de las rupturas naturales de Jenks para capturar la jerarquía subyacente que tienen los datos. **Palabras clave: clasificación de datos, regla de la división cabeza/cola, jerarquía, rupturas naturales, escala.**

Human thinking often involves binary thinking or dualism, which divides things or phenomena into two opposing categories, such as urban–rural, high–low, long–short, and extraordinary–ordinary. Many societal and natural phenomena demonstrate, and can therefore be perceived as, dual parts consisting of an imbalanced contrast; one part is often distinguished as the foreground (the minority) and the other as the background (the majority). For example, in terms of occupied land area, there are far more rural areas than urban areas; 90 percent of the land in Europe is countryside, whereas only 10 percent is urban. On the other hand, far more people live in urban areas than rural areas; 80 percent of people in America and Europe are urban residents and only 20 percent live in the countryside. We also know that there are far more ordinary people (say, 80 percent) than extraordinary people (say, 20 percent); this is often characterized by the 80–20 principle, based on the observation made by the Italian economist Vilfredo Pareto in 1906 that 80 percent of land in Italy was owned by 20 percent of the population. A histogram of the data values for these phenomena would reveal a right-skewed or heavy-tailed distribution. How to map data with the heavy-tailed distribution?

There are a range of data classification methods for statistical mapping, such as equal steps, quantiles, geometric progressions, standard deviation, and natural breaks (Evans 1977; Coulson 1987). The natural breaks method is used to classify data values into different classes according to the breaks or gaps that naturally exist in the data. This is done by plotting the frequency of the data values and then identifying the breaks in the data. This idea of relying on the data frequency to identify class intervals has existed for a long time (Alexander and Zahorchak 1943). To illustrate the method, we examine a statistical data set of rural population densities utilized by Jenks in an early study (Jenks 1963). Figure 1 illustrates the original data with Figure 1A and Figure 1B depicting the spatial distribution and the statistical distribution, respectively. With a few exceptions, most of the data values are centered around 5.5 and lie between 1.6 and 15.0. The exceptions are centered on 18.0, 28.0, 35.0, and 103.4. There are some gaps between the clustered data values, with one of the gaps (between 35.0 and 103.4) being very large. These

gaps constitute the foundation for deriving the class intervals of the natural breaks. Based on the natural breaks, Jenks (1967) developed an optimization method to minimize within-class variance while maximizing between-class variance; this is similar to the concept adopted by Fisher (1958) for data classification. This optimization method is also known as the *goodness-of-variance-fit* (GVF) method and is achieved through iterative computing of the GVF; that is, moving one data value from the class with the largest deviations from the array mean to the class with the lowest deviations from the array mean until the sum of the within-class deviations reaches a minimal value. The resulting classification is called *Jenks's natural breaks* and is widely used in statistical mapping.

The preceding discussion shows that data classification involves two issues: the number of classes and the class intervals. This article proposes a classification scheme that naturally determines the number of classes and class intervals based on human binary thinking and applies to data that follow a heavy-tailed distribution. The heavy-tailed distribution is commonly found in many societal and natural phenomena, including geographical systems (Pumain 2006). Small things are far more common in geographic spaces than large things (Jiang 2010), particularly in urban, architectural environments (Salingaros and West 1999; Jiang 2009). For example, there are far more small cities than large ones (Zipf 1949), far more short streets than long ones (Jiang 2009), far more small city blocks than large ones, and far more low-density areas than high-density ones (Jiang and Liu 2012). This is the pattern that underlies geographic spaces and should be reflected in the map.

The remainder of this article is organized as follows. The next section introduces a set of heavy-tailed distributions and elaborates on how they differ from the normal distribution. After that, I propose the head/tail breaks classification scheme and argue that it reflects the underlying hierarchy of heavy-tailed distributions. I then present two case studies mapping U.S. cities and Swedish streets to further illustrate how the new classification scheme reflects the underlying patterns of data. I further discuss the importance of data classification in statistical mapping and several advantages of the head/tail breaks before concluding the article.
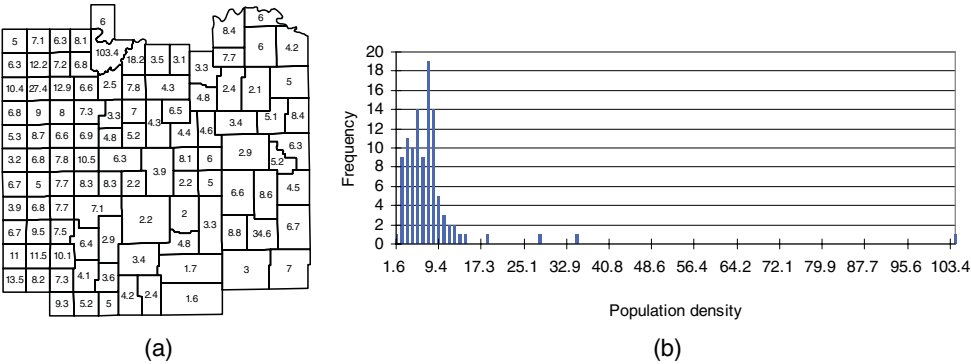
(a)   (b)

Figure 1 *Rural population densities: (A) geographic distribution of the data; (B) statistical distribution of the data. Note: The data concern the population densities for 105 minor civil divisions in eight central Kansas counties. (Color figure available online.)*

Heavy-Tailed Distributions

Existing classification methods are dominated by a Gaussian way of thinking; they focus on high-frequency events but consider low-frequency events to be separated from high-frequency ones. There are large gaps or breaks (shown between the blue bars in Figure 1B) between high-frequency and low-frequency data values, as well as between different levels of low-frequency values. These gaps and breaks constitute the foundation for the natural breaks classification. This article reverses this way of thinking by concentrating on low-frequency values. The lowest frequency value is ranked number 1, the second lowest frequency value is ranked number 2, and so on. The ranking is plotted on the *x*-axis and the corresponding values or sizes on the *y*-axis. This plot, known as the *rank-size distribution* in general and initially used for city sizes and word frequencies

(Zipf 1949), is a typical heavy-tailed distribution (Figure 2)

I focused on low-frequency events because they tend to have high impact, which makes them more important than high-frequency events. Low-frequency events contain far more information than high-frequency events because of their improbable nature. For example, news headlines are low-frequency events that occur rarely and are therefore newsworthy; they convey more information than high-frequency events. Nature, society, and our daily lives are full of rare and extreme events, which are termed *black swan events* (Taleb 2007). This line of thinking provides a good reason to reverse our thinking by focusing on low-frequency events.

Unlike a normal distribution, which is bell-shaped and has two thin tails that reach the *x*-axis rapidly, a heavy-tailed distribution theoretically has a long tail skewed toward the right
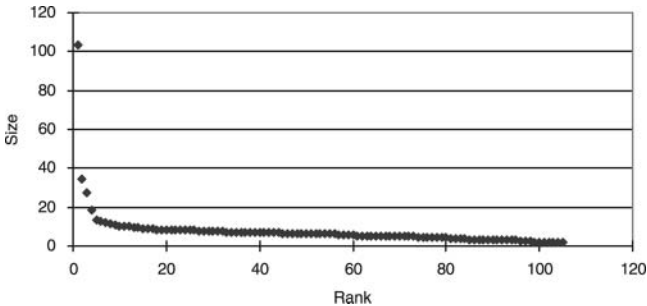

Figure 2 *A typical heavy-tailed distribution based on the rural population density data.*

Size

Head(10%)

m

Tail(90%)

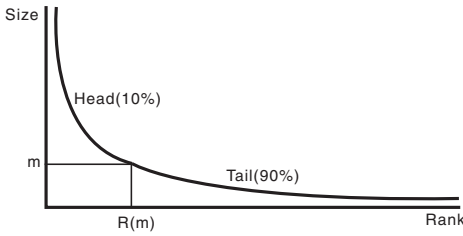R(m)   Rank

Figure 3 *Illustration of the head/tail division rule. Note: A rank-size distribution with the x-axis as the ranking and the y-axis as the corresponding ranked data values. The arithmetic mean is m and the corresponding ranking is R(m). The head and the tail respectively contain imbalanced amounts of data values.*

that approaches but never touches the *x*-axis. The entire distribution line or the long tail consists of the head and the tail, split around an average of the data values. Importantly, the head is composed of a minority of the large data values, whereas the tail is composed of a majority of the small data values. This imbalance between the head and the tail (or respectively the minority and the majority) has been formulated as the head/tail division rule: "Given a variable X, if its values x follow a heavy tailed distribution, then the mean (m) of the values can divide all the values into two parts: a high percentage in the tail, and a low percentage in the head" (Jiang and Liu 2012). Figure 3 illustrates a heavy-tailed distribution in which the head and tail is divided around the mean value *m*, with only 10 percent of data values in the head and 90 percent of data values in the tail. The use of the mean or average to partition the data values into two imbalanced parts is in line with binary thinking of the human mind.

In general, a heavy-tailed distribution can also be observed when assessing the probability density function (PDF) of the data values. It should be noted that low-frequency events are in the head in rank-size distributions but in the tail in PDFs. In this article, to avoid confusion, the terms *head* and *tail* refer to the rank-size distribution rather than the PDF. Given a variable of interest *x* and the probability of occurrence of the variable value *y*, the most typical heavy-tailed distribution is a power law, expressed as

$$y = x^{-\alpha}, \qquad (1)$$

where α is termed the power law exponent.

Taking the logarithm of both sides in Equation 1, we obtain

$$\ln y = -\alpha \ln x. \qquad (2)$$

The distribution line is straight when graphed as a double-logarithm plot. This straight distribution line, which is the hallmark of a power law, is often used to detect the power law through double-logarithm plots, so called log-log plots. This method, which is essentially based on the least squares method, has been criticized, however, for creating errors or bias when drawing conclusions regarding the existence of a power law distribution. Other methods based on the maximum likelihood have been proposed as being more appropriate for detecting a power law distribution. In this article, I adopt the latter method in detecting a heavy-tailed distribution. It should be noted that mathematics are kept at the simplest level in this article; interested readers should refer to Clauset, Shalizi, and Newman (2009) and the references therein for more details.

A heavy-tailed distribution can also be characterized by the lognormal function, which has the format

$$y = \frac{1}{\sqrt{2\pi\sigma^2}x} \exp\left(\frac{-(\ln(x) - \mu)^2}{2\sigma^2}\right), \qquad (3)$$

where the parameters μ and σ are the mean and standard deviation of $\ln(x)$, respectively. This is a normal distribution for $\ln(x)$, but the variable *x* itself is heavy-tailed distributed; that is, lognormal.

In the literature, particularly in the statistical physics literature, the tail of a heavy-tailed distribution is considered to be heavier or more long-tailed than that of an exponential distribution. The "heavier" or "more long-tailed" is what was referred to earlier for the tail approaching but never touching the *x*-axis. Therefore, the exponential distribution is excluded from the class of heavy-tailed distributions with a good reason. Physicists believe that the mechanisms underlying events that follow the "strictly defined" heavy-tailed distribution are fundamentally different from those that follow an exponential distribution.

In the simplest style, an exponential distribution is expressed as

$$y = e^x. \tag{4}$$

Taking the logarithm of both sides in Equation 4, we obtain

$$\ln y = x. \tag{5}$$

Equation 5 indicates that a linear relationship exists between x and the logarithm of y. This is different from the power law distribution, for which there is a linear relationship between the logarithm of x and logarithm of y. This article, however, loosely considers the exponential distribution to be a part of the heavy-tailed distribution family, for two reasons: (1) each of the three functions illustrated earlier are right-skewed, and (2) it is sometimes extremely difficult to differentiate among the three distributions with real-world data sets. Instead of following any of the three distributions exactly, real-world data sets often follow degenerate heavy-tailed distributions, such as a power law with an exponential cutoff or a stretched exponential distribution (Clauset, Shalizi, and Newman [2009], and references therein). In this study, all heavy-tailed distributions were considered to be a good approximation of a power law function.

A Novel Data Classification Scheme: Head/Tail Breaks

This article proposes a data classification scheme in which the number of classes and the class intervals are both naturally determined by the very property of the heavy-tailed distributions of the data. As discussed earlier, a heavy-tailed distribution consists of a head and a tail, each of which contains an unbalanced proportion of data values. This imbalance between the head and tail, or between a few large values and many small values, can be expressed as "far more small things than large things." Perceptually, the tail constitutes a background on which the head is distinguished. Interestingly, this imbalance recurs in the head; the values in the head can be divided again around an arithmetic mean into many small values and a few large values. In other words, the values demon-strate a hierarchical structure. This inherent hierarchy constitutes the foundation of the proposed classification scheme, which is termed here *head/tail breaks.*

The head/tail breaks method partitions the data values into two parts around the arithmetic mean and continues the partitioning for values above the mean iteratively until the head part values are no longer heavy-tailed distributed. Therefore, the number of classes and the class intervals are both naturally determined by the data or the data's very property of heavy-tailed distribution. In this way, the head/tail breaks classification method differs fundamentally from the nested means classification method, which iteratively partitions the values in both the head and the tail (Scripter 1970). The nested means start by partitioning the data values into two classes around the mean: those above the mean and those below the mean. This process continues for each of the two classes and using the new mean for each class will result in four classes. This process can continue to get eight classes, sixteen classes, and so on. In fact, there is no need to further classify or partition the values in the tail or below the mean because the variation in the tail is small enough.

This classification scheme is illustrated by applying it to the previously mentioned rural population density data. As noted, the data are heavy-tailed distributed. The first arithmetic mean is 7.6, which partitions all of the data into two groups: those above the mean or in the head and those below the mean or in the tail. The values in the head are also heavy-tailed distributed. The second mean, for the values in the head, is 14.0. Four values exceed 14.0; namely, 18.2, 27.4, 34.6, and 103.4. The third mean for these four values is 45.9, which can be used to partition the four values into two groups, in which the value 103.4 constitutes one group. Eventually, the classification scheme ends up with four classes, as shown in Figure 4A, which is put in comparison with the natural breaks classification (Figure 4B).

It is worth elaborating on the pattern illustrated by the head/tail breaks. The first class ranges from the minimum value (1.6) to the first mean (7.6) and constitutes the background for the second class. The second class ranges from the first mean (7.6) to the second mean (14.0) and constitutes the background for the third
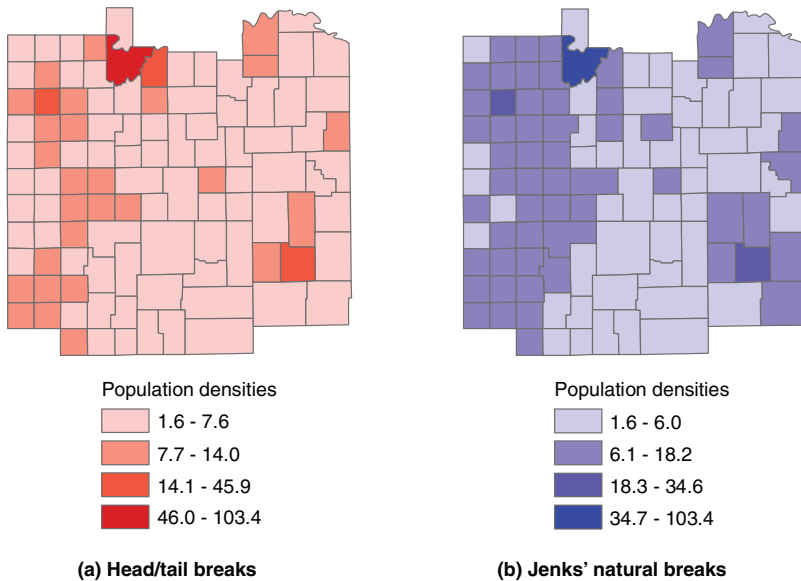
Figure 4 *Head/tail breaks (A) compared with Jenks's natural breaks and (B) using the rural population density data. (Color figure available online.)*

class. The third class ranges from the second mean (14.0) to the third mean (45.0) and constitutes the background for the fourth class. The fourth class ranges from the third mean (45.0) to the maximum value (103.4) and constitutes the foreground. The lowest (or first) class is the background and the highest (or fourth) class is the foreground. All of the other middle classes serve as both a background for the immediate upper class and a foreground for the immediate lower class. Importantly, both the backgrounds and the foregrounds, which correspond to the tail and the head, create an imbalanced contrast; that is, a small foreground versus a vast large background. This is exactly the underlying pattern for data with a heavy-tailed distribution. This pattern can also be called scaling or hierarchy, which should be retained in the map.

A good classification method should largely reflect the pattern that underlies the data. In other words, any classification method that distorts the underlying pattern of data is considered a poor classification. The pattern underlying heavy-tailed distributed data is scaling or hierarchy, which can be expressed as far more smaller things than larger ones. The notion of

scaling of geographic space refers to the hierarchy that is characterized by the heavy-tailed distributions (Jiang 2010). The pattern illustrated by Jenks's natural breaks (Figure 4B) clearly lacks this scaling property; for example, the two lowest classes occupy approximately same percentage of areas or data values, so no hierarchy is shown. It should be noted that the expression far more smaller things than larger ones can sometimes be stated the other way around. For example, in a city center where almost all buildings are skyscrapers, it could be stated as "far more larger things than smaller ones." In this case, the minority of smaller values constitute the head and the majority of the larger values constitute the long tail.

Case Studies: Head/Tail Breaks in Comparison with Jenks's Natural Breaks

I conducted two case studies to further illustrate the advantages of the head/tail breaks scheme in capturing the underlying hierarchy of the data. The two case studies contain considerably more data values than the small example

shown earlier because the heavy-tailed distributions usually appear in systems that are large in terms of time or space, which implies that large systems have evolved over a long time period or developed in a large spatial setting. It is often the case that the larger the systems, the more obvious or striking the heavy-tailed distributions. Another reason is the perspective from which we observe the world. For example, a street network is not heavy-tailed distributed in terms of the degree of junction connectivity, but it is heavy-tailed in terms of the degree of the connectivity of individual streets (Jiang, Zhao, and Yin 2008). The correct perspective and scope must be chosen to observe the scaling of geographic space.

The first case study concerns the statistical mapping of U.S. cities. According to the 2000 U.S. Census (e.g., National Weather Service 2011), there are 42,934 cities in the United States. For the purposes of this study, 5,242 cities with populations greater than 8,000 were selected for mapping. The selected city sizes are ranked by population; the smallest city has 8,000 people and the largest city has 7,322,564. The total population in the 5,242 cities is 284,916,464. The city sizes do not appear to be normally distributed; there are two reasons for this. First, both the smallest size and the largest size are far from the average size of the city, which is 54,353 (284,916,464/5,242 = 54,353). Second, the ratio of the largest size to the smallest size is very large: 7,322,564/8,000 = 915. Further examination of the city sizes demonstrates that they are indeed power-law distributed. Figure 5 illustrates the log-log
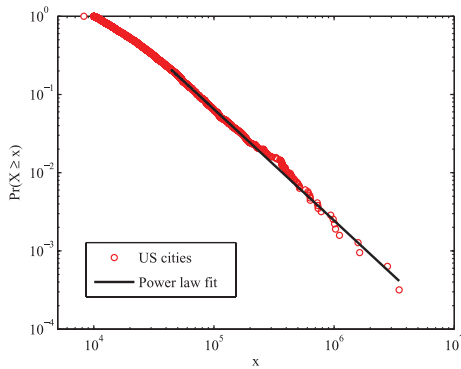
Table 1 *Three partitions leading to four hierarchical levels of U.S. cities*

No. of cities	No. in head	% in head	No. in tail	Mean
5,242	744	14	4,498	54,353
744	121	16	623	268,370
121	18	15	103	1,111,291

plot, which indicates that the power law fits across two decades of the data values.

We further examine the hierarchy of the data values. The first arithmetic mean (54,353) groups all of the cities into two categories: those above the mean and those below the mean. There are 744 cities above the mean, which accounts for 14 percent (a minority) of all cities. The second mean on these 744 cities is 268,370. This value can be used to further partition the 744 cities into two categories. Of the 744 cities above the mean, only 121 (16 percent) have populations larger than the average of 268,370 (again, a minority). The third mean, 1,111,291, derived from the 121 cities, partitions the 121 cities into two groups, the result of which is shown in Table 1. The eighteen cities in the last head are no longer heavy-tailed distributed, which means it is necessary to stop the partitioning. Eventually, the three partitions lead to the creation of four classes or hierarchical levels. The four class intervals are determined by the smallest size (8,000), the largest size (7,322,564), and the three means: [8,000, 54,353], [54,354, 268,370], [268,371, 1,111,291], [1,111,292, 7,322,564].

The four classes are mapped using graduated circles and compared with classes generated using Jenks's natural breaks scheme (Figure 6). The map illustrated by the head/tail breaks reflects the underlying pattern better than the map generated using Jenks's natural breaks. The pattern based on Jenks's natural breaks looks very flat (not much variance), whereas the map based on the head/tail breaks is full of changes or surprising changes, both locally and globally. This scaling pattern is more obvious when it is assessed with Table 1. In view of this, the map based on the head/tail breaks scheme reflects the hierarchy of the data values much better than the one based on Jenks's natural breaks.
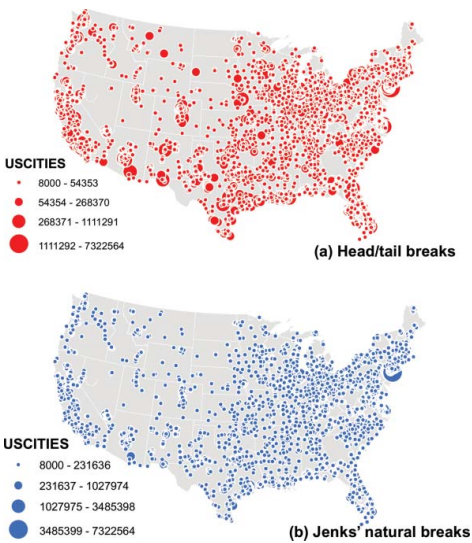

Figure 5 *Power law distribution of U.S. cities. (Color figure available online.)*

Figure 6 *Classification of U.S. cities based on (A) the head/tail breaks and (B) Jenks's natural breaks. (Color figure available online.)*

Figure 7 *Heavy-tailed distribution for the connectivity of Swedish streets. (Color figure available online.)*

The second case study maps the underlying hierarchy of 160,000 Swedish streets. The streets are either named streets with unique names or natural streets that constitute good continuity in perception (Jiang and Claramunt 2004; Jiang, Zhao, and Yin 2008). The degree of connectivity of individual streets was mapped; that is, the number of other streets that connect or intersect each individual street. The degree of connectivity ranges from 1 to 1,040. Therefore, the ratio of the most connected to the least connected streets is $1,040/1 = 1,040$, which is a clear indicator of a heavy-tailed distribution. A less pronounced fit to a power law distribution was uncovered as illustrated in Figure 7; it is definitely a heavy-tailed distribution. Due to the distribution, the data values can be partitioned nine times around different means iteratively. Table 2 shows the result of the classification, where the head always accounts for a minority of streets. The nine partitions lead to ten classes or hierarchical levels, which are mapped using a spectral color legend as shown in Figure 8.

It is difficult to determine which of the two maps in Figure 8 is better solely on the basis of the visual effect because the high density of streets causes overlap. Most blue lines
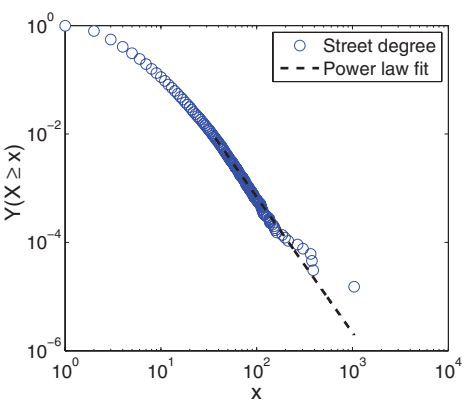
in the two maps are covered by lines in warm colors, such as red, orange, and yellow, which makes it difficult to visually assess the difference between the two maps. In terms of capturing the underlying hierarchy of the data values, however, the left pattern is indeed superior to the right one. This point can only be seen or be assessed from Table 2, which is associated with Figure 8. The right pattern does not adopt the inherent hierarchical levels; instead, it addresses the gaps that naturally exist in the data. The head/tail breaks scheme captures and reflects the underlying hierarchy more precisely than the natural breaks scheme, however.

The second case study has ten classes, which is slightly more than the number that the human mind can hold in the memory simultaneously (Miller 1956). In this case, the number

Table 2 *Nine partitions leading to ten classes of Swedish streets*

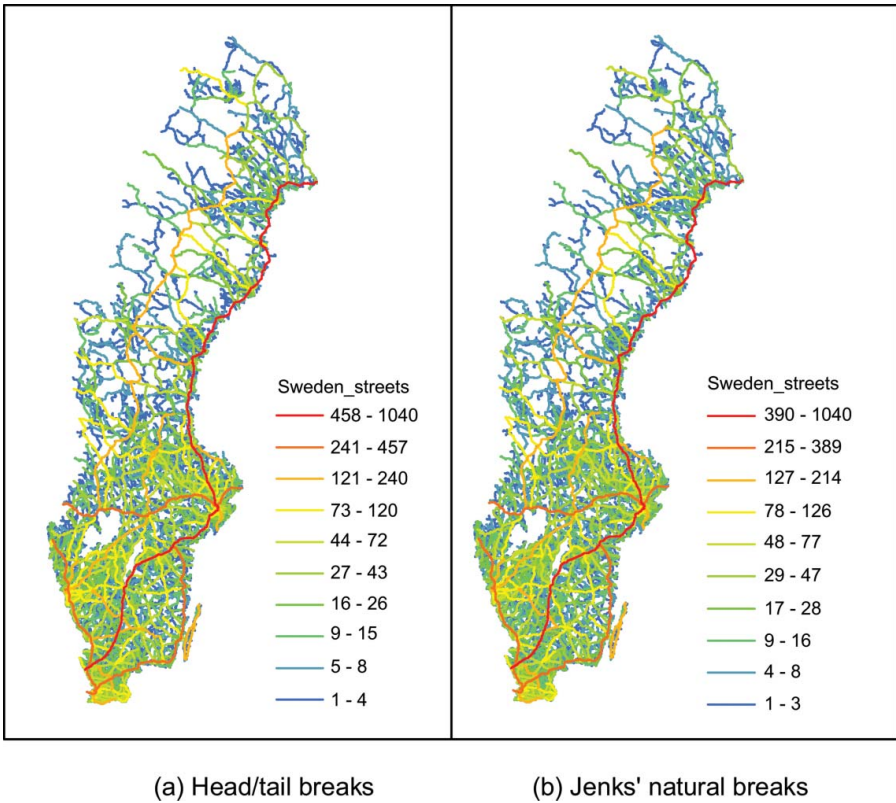| No. of streets	No. in head	% in head	No. in tail	Mean
166,479	47,021	28	119,458	3.6
47,021	14,382	31	32,639	8.0
14,382	4,235	29	10,147	14.8
4,235	1,294	31	2,941	26.0
1,294	363	28	931	43.2
363	101	28	262	71.7
101	21	21	80	119.8
21	6	29	15	239.6
6	1	17	5	457.4

Figure 8   *Classification of Swedish streets in terms of street connectivity based on (A) the head/tail breaks and (B) Jenks's natural breaks. (Color figure available online.)*

of classes can be reduced to seven (for example) by merging the top four classes into a single class. It is important to note that the merging does not affect the fundamental hierarchy. When this is performed, the two maps (based on the head/tail breaks and Jenks's natural breaks schemes) look quite different visually or in terms of the coloring (Figure 9). This is probably an advantage of the head/tail breaks compared to the natural breaks. Every time the number of classes changes, the natural breaks must recompute all of the class intervals by minimizing within-class variance and maximizing between-class variance. There is no need, however, for this to happen with the head/tail breaks to maintain the derived hierarchy or classes. The upper classes are simply merged into one and all of the lower class intervals remain unchanged.

In perceiving a map such as Figure 9A or Figure 8A, one should not be influenced by the visual effect, which incorrectly suggests that there are "far more long streets than short ones." In fact, there are far more short streets (which are, unfortunately, overlapped by long ones on the top) than long ones. In view of this and with respect to Table 2, the pattern on the left represents the hierarchy of the data, whereas the pattern on the right indicates the breaks in the data.

The two case studies have shown how the head/tail breaks classification scheme captures the inherent hierarchy of the data. Interestingly, a lower class constitutes a background in which the upper classes are distinguished. In other words, an upper class constitutes a foreground for the immediate lower class. Importantly, the background and the foreground
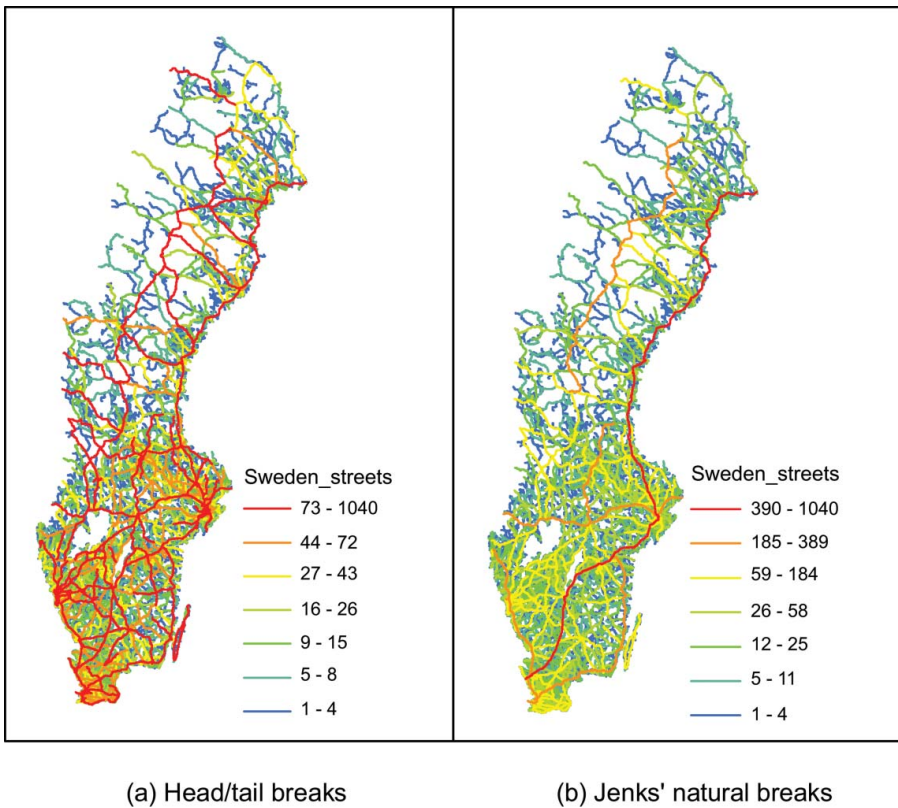
Figure 9 *Reduction of the ten classes into seven classes for (A) head/tail breaks and (B) Jenks's natural breaks. (Color figure available online.)*

are in an imbalanced contrast. This is in line with the binary thinking of the human mind. Therefore, this article argues that the binary thinking–based classification is more natural than the natural breaks for data with a heavy-tailed distribution.

Discussion

Data classification is a fundamental issue in statistical mapping, and it is well known that different classification methods can lead to very different visual patterns, sometimes in a distorted or biased manner (e.g., MacEachren 1994; Monmonier 1996; Brewer and Pickle 2002). Because of the difficulty of classification, continuous classes or no-class-interval maps have been suggested (Tobler 1973). Classification is an important step toward effective

mapping, as human perceptual limits require effective classification to successfully map underlying patterns. One such visual pattern is the scaling pattern studied in this article, which is pervasive in geographic space; that is, far more small things than large ones. This pattern is reflected in a great deal of geographic data. Mapping of geographic data should reflect the pattern of geographic space and a good classification scheme must capture the underlying pattern of the data distribution.

Data classification is, to a large extent, a process of generalization by differentiating between vital and trivial things. In fact, human thinking also generalizes when receiving information from the real world. For example, news headlines generally deserve more attention than trivial matters. This article introduces a new way of thinking by concentrating on

low-frequency things because these tend to be vital, important, and high impact. The head/tail breaks classification is driven by this new way of thinking; the resulting upper classes capture the important and vital things in the heads, whereas trivial things are relegated to the tail as the lowest class.

The head/tail breaks method demonstrates several advantages compared with the natural breaks and other classification schemes. First, the head/tail breaks captures the image or hierarchy that appears in reality (far more smaller things than larger things). Second, the number of classes and the class intervals are naturally determined based on the hierarchy embedded in the data. The third advantage is that the number of classes matches well the human memory limit, which is approximately seven (Miller 1956). This third advantage contrasts strongly with the nested-means classification scheme, which creates too many classes because of the binary partitioning of the head and the tail, and more critically, generates unnecessary and redundant breaks for the data values in the tail.

Some conventional classification methods pay attention to the scaling pattern or heavy-tailed distribution of data. Fundamentally, for data with a heavy-tailed distribution, class intervals cannot be imposed in a linear manner (e.g., as equal steps) but rather in a nonlinear style. This nonlinear property is behind the use of nonlinear class interval methods, such as arithmetic and geometric progression methods. The widths of the class intervals increase in size at an arithmetic (or additive) rate. If the first class is two units, then the second and third classes would be four and six units, respectively. For the geometric progression, the widths of the class intervals increase in size at a geometric (or multiplicative) rate. If the first class is two units, then the second and third classes would be four and eight units, respectively. Both arithmetic and geometric progression methods are not commonly used, however, because neither the number of the classes nor the class intervals can be simply determined.

Conclusion

This article has proposed a novel classification scheme, the head/tail breaks, for data that are heavy-tailed distributed or right skewed. For the head/tail breaks scheme, the number of classes and class intervals are both naturally determined based on inherent hierarchical levels of the data. The class intervals are iteratively derived from the arithmetic means until the data values in the head are no longer heavy-tailed distributed. The head/tail breaks scheme captures the hierarchy of the data and is based on the binary thinking of the human mind. In this sense, this classification scheme is more natural than the natural breaks scheme for data with a heavy-tailed distribution.

The head/tail breaks method demonstrates a number of advantages over the commonly used methods for classifying data with a heavy-tailed distribution. The most important of these is that the classification scheme captures the essence of the data distribution that is scaling or hierarchy. Another notable advantage is the scheme's simplicity; the class intervals are iteratively derived from the arithmetic means (cf. Appendix). In this way, both the number of classes and the class intervals are naturally determined. If there are too many classes (e.g., more than seven), the upper classes can simply be merged into a single class without affecting the lower classes. Although the advantages of the head/tail breaks method are clear, it remains to be seen how humans will respond to the new classification scheme compared to traditional ones. ∎

Literature Cited

Alexander, J. W., and G. A. Zahorchak. 1943. Population-density maps of the United States: Techniques and patterns. *Geographical Review* 33 (3): 457–66.

Brewer, C. A., and L. Pickle. 2002. Evaluation of methods for classifying epidemiological data on choropleth maps in series. *Annals of the Association of American Geographers* 92 (4): 662–81.

Clauset, A., C. R. Shalizi, and M. E. J. Newman. 2009. Power-law distributions in empirical data. *SIAM Review* 51:661–703.

Coulson, M. R. C. 1987. In the matter of class intervals for choropleth maps: With particular reference to the work of George F Jenks. *Cartographica* 24 (2): 16–39.

Evans, I. S. 1977. The selection of class intervals. *Transactions of the Institute of British Geographers* 2 (1): 98–124.

Fisher, W. D. 1958. On grouping for maximum homogeneity. *Journal of the American Statistical Association* 53:789–98.

Jenks, G. F. 1963. Generalization in statistical mapping. *Annals of the Association of American Geographers* 53 (1): 15–26.

———. 1967. The data model concept in statistical mapping. *International Yearbook of Cartography* 7:186–90.

Jiang, B. 2009. Street hierarchies: A minority of streets account for a majority of traffic flow. *International Journal of Geographical Information Science* 23 (8): 1033–48.

———. 2010. Scaling of geographic space and its implications. Position paper presented at Las Navas 20th Anniversary Meeting on Cognitive and Linguistic Aspects of Geographic Space, Las Navas del Marques, Avila, Spain.

Jiang, B., and C. Claramunt. 2004. Topological analysis of urban street networks. *Environment and Planning B: Planning and Design* 31:151–62.

Jiang, B., and X. Liu. 2012. Scaling of geographic space from the perspective of city and field blocks and using volunteered geographic information. *International Journal of Geographical Information Science* 26 (2): 215–29.

Jiang, B., S. Zhao, and J. Yin. 2008. Self-organized natural roads for predicting traffic flow: A sensitivity study. *Journal of Statistical Mechanics: Theory and Experiment* (July):P07008.

MacEachren, A. M. 1994. *Some truth with maps: A primer on symbolization and design*. Washington, DC: Association of American Geographers.

Miller, G. A. 1956. The magic number seven, plus or minus two: Some limits on our capacity for processing information. *The Psychological Review* 63 (2): 81–97.

Monmonier, M. 1996. *How to lie with maps.* 2nd ed. Chicago: University of Chicago Press.

National Weather Service. 2011. National Oceanic and Atmospheric Administration's national weather service. http://www.weather.gov/geodata/catalog/national/html/cities.htm (last accessed 15 February 2012).

Pumain, D. 2006. *Hierarchy in natural and social sciences*. Dordrecht, The Netherlands: Springer.

Salingaros, N. A., and B. J. West. 1999. A universal rule for the distribution of sizes. *Environment and Planning B: Planning and Design* 26:909–23.

Scripter, M. W. 1970. Nested-means map classes for statistical maps. *Annals of the Association of American Geographers* 60 (2): 385–93.

Taleb, N. N. 2007. *The black swan: The impact of the highly improbable*. London: Allen Lane.

Tobler, W. 1973. Choropleth maps without class intervals? *Geographical Analysis* 5:262–65.

Zipf, G. K. 1949. *Human behaviour and the principles of least effort*. Cambridge, MA: Addison Wesley.

BIN JIANG is Professor of Computational Geography and GeoInformatics in the Department of Technology and Built Environment of the University of Gävle, Gävle 80176, Sweden. E-mail: bin.jiang@hig.se. His research interests include geospatial analysis, geovisualization, and complexity modeling for urban structure and dynamics.

Appendix A: Short Tutorial on How to Perform the Head/Tail Breaks Classification in ArcMap

Assuming that you have the USCITIES shapefile added into ArcMap as a map layer, open the attribute table associated with the shapefile. Find the field named POP_1990, right-click the column, and choose Statistics. . . . This will open a pop-up window similar to Figure A1.

On the left side of Figure A1, you can see that the total number of cities is 5,242, the lowest size is 8,000, and the largest size is 7,322,564. The frequency distribution to the right side looks L-shaped, which means it is a heavy-tailed distribution. Copy the first mean (54,353) and go to menu Selection > Select By Attributes. A window will appear that looks like Figure A2.

Choose the correct layer (e.g., USCITIES) and the correct field (e.g., POP_1990) for the selection "POP_1990" $>=$ 54352.625715, as shown in Figure A2. This is how we selected larger cities above the first mean (m_1) or in the first head. Continuing this process will obtain the second and third means (m_2, and m_3). Next, we can conduct a manual classification by imposing the class intervals, [8,000, 54,353], [54,354, 268,370], [268,371, 1,111,291], [1,111,292, 7,322,564] to perform the classification. In general, the class intervals are determined by the lowest size (min), largest size (max), and the individual means (e.g., m_1, m_2, and m_3), i.e., [min, m_1], [m_1+1, m_2], [m_2+1, m_3], [m_3+1, max].
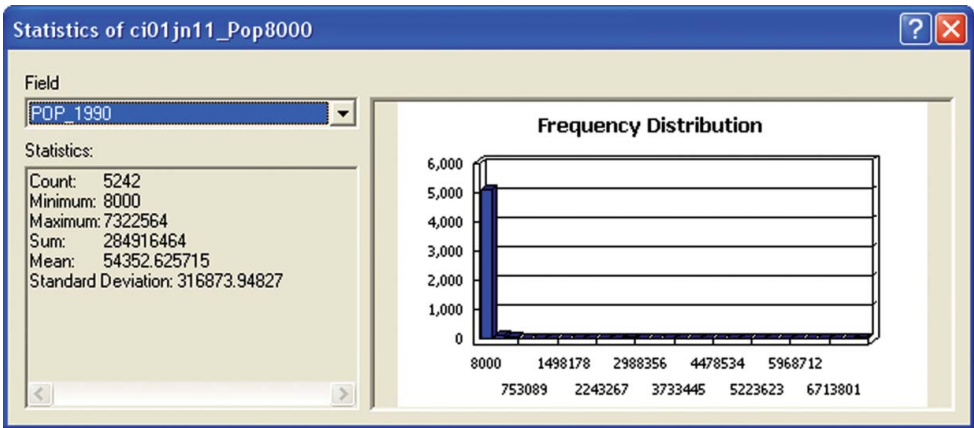
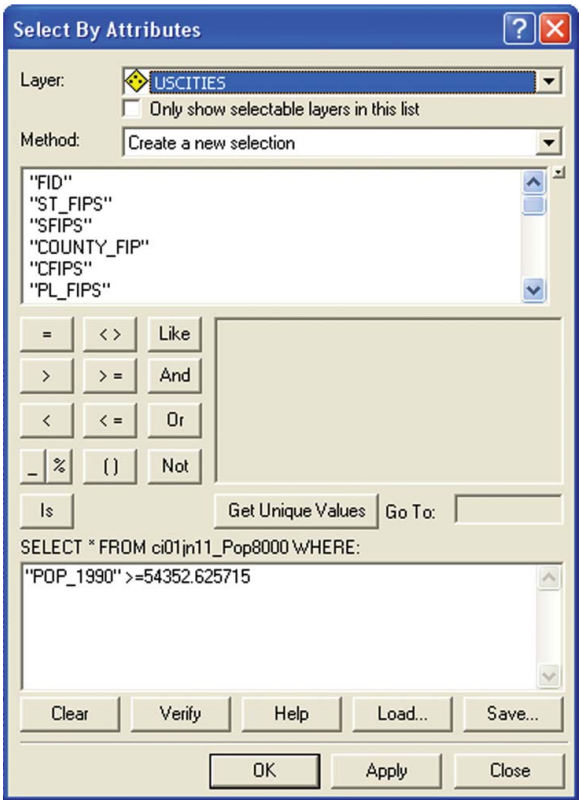**Figure A1**   *Pop-up window for statistics. (Color figure available online.)*

Figure A2   *Selection by attributes window. (Color figure available online.)*