Clustering Massachusetts Communities into Regions

Craig E. Wills
Computer Science Department
Worcester Polytechnic Institute

July 2026

Abstract

Massachusetts has 351 cities and towns spread across 14 counties. Counties and other work to divide the state into regions typically results in broad regions with non-specific names. In this work, we cluster Massachusetts communities into cohesive and locally distinctive regions each with a meaningful name to represent it. The result is 35 regions, each with a name, description as well as the list of cities and towns in the region. Regions and communities are also mapped with each region shown in a color representative of the region and each community shown in its primary region if it belongs to more than one.

Roughly 75% of cities and towns belong to a single primary region with the remaining communities in overlapping regions. Two additional maps are created where: 1) communities in overlapping regions are shown in lighter color intensity; and 2) communities in overlapping regions are shown in a blending of region colors.

Interactive versions of all maps generated and shown in this report can be viewed with a Web browser at `https://web.cs.wpi.edu/%7ecew/massregions/`.

1 Introduction

This work was motivated by the desire to develop a better clustering of Massachusetts' 351 cities and towns into regions. Counties are not particularly meaningful in Massachusetts and other work (official and unofficial) tends to result in broad regions with non-specific names. In Computer Science clustering is in an important technique as it helps to understand similarities between members within a cluster and differences between members across clusters. Applying clustering to create regions on a map is enlightening to allow them to be seen and our world to be better understood.

Having lived and worked in Massachusetts for many years, we are familiar with regions such as Cape Cod, the North Shore, Metro West, Central Massachusetts and the Berkshires. However an interest in a more serious and systematic investigation of Massachusetts' regions than what was available led to this work. We also have the means to interactively map these regions.

In developing our map of regions, we did not set out to define a specific number of regions (nor a specific number of communities in each region), but rather sought to cluster communities based on evidence of shared historical, geographic, economic and educational characteristics. We primarily relied on existing regional organizations such as chambers of commerce, governmental planning groups, multi-town school districts and athletic leagues groups both to identify clustered communities in a region as well as to name the region itself. The result is 35 regions ranging from 3 communities in the Mid Cape region to 22 communities in the North Shore region.

In the remainder of this report, Section 2 describes the methodology we used to identify regions and determine the member communities within each region. Section 3 shows a map of the resulting regions with each community shown in its primary region and each region shown in a representative color. Section 4 goes on to discuss questions that arose during our work regarding the region classification for each community with roughly 25% identified in overlapping regions. Section 5 discusses and two approaches (along with maps) for the representation of communities in overlapping regions. We conclude the report in Section 6 with a summary of our results and directions for future work.

2 Methodology

In this section we describe the methodology used to cluster cities and towns into regions. We employ the algorithm shown in Figure 1 with line numbers added for reference.

The algorithm starts with a set of communities C and seeks to both identify a set of regions R and associate the subset of communities belonging to each region. Lines 1-3 initialize the algorithm by starting with a set of "known" regions. For example, in Massachusetts regions such as the North and South Shore, Cape Cod, Metro West and the Berkshires are well-known and such regions could be used as a starting point by identifying the set of communities belonging to each.

Whether or not the algorithm begins with a set of known regions, the main loop of the algorithm begins on line 4 where a community $c_m \in C$ not currently assigned to a region is selected. Line 5 then determines the best fitting regions(s) for the community using heuristics such as shown in lines 6-12. These heuristics give (not always consistent) guidance on the self-defining region in which this community is a member. In many cases the region for a community is obvious, but in some cases the community could legitimately be a member of overlapping regions. In these cases we subjectively assign a primary region to the community, but also record its overlapping

1. Start with the set of communities C and the the set of known regions R (could initially be empty).

2. Foreach region $r_i \in R$ {

    3. Identify the set of communities $c_m \in C$ that are members of region r_i and assign them to the set for r_i.

}

4. Foreach community $c_m \in C$ not currently assigned to a region {

    5. Determine best fitting region(s) for community c_m using heuristics such as {

        6. Find the regional chamber of commerce for the community.

        7. Find regional planning groups in which the community is a member.

        8. Determine if the community is part of a regional school district.

        9. In Massachusetts vocational technical high schools are often regionalized. Determine the technical high school of the community.

        10. High school athletic leagues often are formed based on communities in a region. What is the league of the town high school and who are the other communities in the league?

        11. What do wiki pages for the community and potential regions suggest? While not authoritative sources, these pages are typically written with local knowledge of a community or region and provide supporting evidence.

        12. AI-based results can also be used with caution as inaccuracies were found, but the results can also point to other authoritative sources.

    }

    13. If the best fit for community c_m is in existing region r_i {

        14. Assign c_m to r_i.

    }

    15. Else {

        16. Create new region r_i and add to set R.

        17. Add c_m and other previously unassigned communities to r_i.

        18. Potentially adjust overlapping regions for previously assigned communities with the availability of the new region r_i.

    }

}

Figure 1: Clustering Algorithm to Identify Regions and Their Associated Communities

region(s).

Lines 11 and 12 show that we do make use of wiki and AI-based results in our work. Wiki results do not have a clear source and can be subjective, but also represent "local" information, which is what we seek to make use of. Wiki pages are also pointers to more authoritative sources. AI-based results can also point to authoritative sources for their results, although the results must be used with caution as we found inaccuracies in what was returned.

Lines 13 and 14 of the algorithm show that the best region fit for a community may be an existing region. Lines 16 and 17 show that if a new region is identified then this community and other unassigned communities may be added to it. New regions may also be identified as subsets of previously identified regions. For example a "Cape Cod" region was eventually broken into four separate regions. Finally, line 18 shows that the availability of a new region may also lead to adjustments for previously assigned communities who are now in overlapping regions.

The main loop of the algorithm beginning on line 4 continues to iterate until all communities in set C have been assigned to at least one of the regions in R. We note that our algorithm does not have a pre-defined number of regions nor does it have a pre-determined size (min or max) on the number of communities within a region. While we prefer to identify more specific regions when we can, our algorithm is guided to find clusters of communities with evidence for treating them as a named and recognized region.

Our algorithm finds self-revealing regions that have meaning to people based local organizations. Thus our work is not about creating "new" regions, but rather assembling the specific regions that are locally, but maybe not widely, known and stitching them together into a mosaic that encompasses the entire state.

3 Resulting Regions

Making use of this methodology we clustered the 351 towns and cities of Massachusetts into 35 distinct regions. The resulting map is shown in Figure 2 with an interactive version of the map available on the project Web site.

This colorful map shows each Massachusetts community in one of 35 regions. Colors for each region were selected to allow visual distinction of each region as well as be representative of the region itself. For example, the Greater Boston area is shown in Continental Blue, an official color of the City of Boston. The Islands region is shown in bright pink for the Oak Bluffs' ginger bread cottages and the Cranberry region is shown in cranberry. Rockport's Barn Red Motif #1 represents Cape Ann while Smiley Face yellow is shown for the Greater Worcester region. Finally, Great Barrington Green is shown for the Southern Berkshires region.

Hovering over a community on the interactive version of the map shows its name, region and county. The list below the map shows each region's name and color as well as the number of cities and towns within the region. Clicking on a region shows its description including a list of towns and the representative color for the region.

Roughly 75% of communities are associated with a single region while the remainder belong to overlapping regions. For these latter communities, we identify a primary region for each with the map showing the color of the primary region for each community. We also show the overlapping towns both in our description of each region as well as in additional maps discussed later in the report.

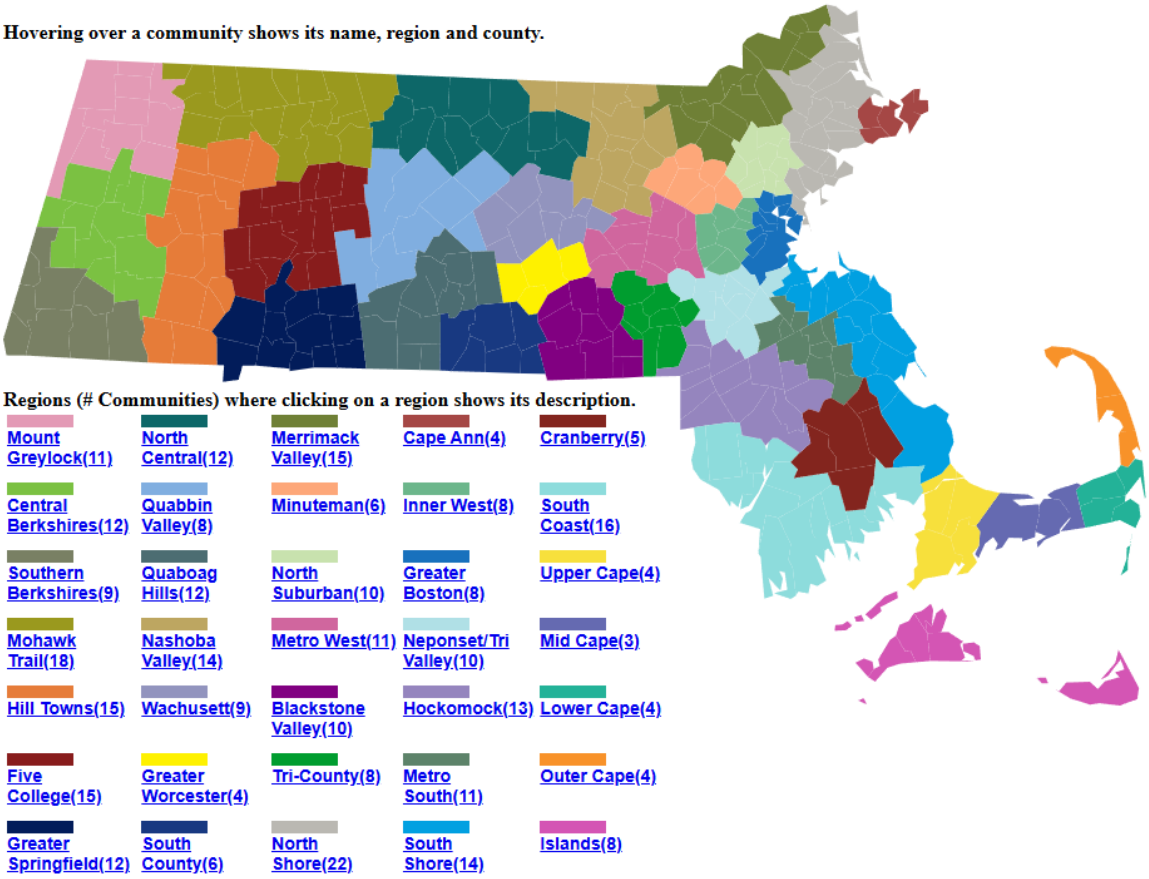

Figure 2: Regions of Massachusetts

4 Discussion

We try to take an evidence-based approach in the identification of regions and the assignment of communities to these regions. However, questions can legitimately be raised on how cluster cities and towns into a region. In the following we provide further discussion on questions about the assignment of communities to regions for some of the resulting regions.

4.1 Greater Boston

The Metropolitan Area Planning Council includes 101 cities and towns in its Greater Boston region [1]. However, its own description recognizes the diversity of these communities in that "the 101 cities and towns include a diverse mix of coastal communities, older industrial centers, rural towns, and urban neighborhoods."

In our work we employ a much tighter definition of the Greater Boston region focused on Boston and its inner core communities. Our practical definition is to include cities and towns with Massachusetts Bay Transportation Authority (MBTA) subway "T" (not commuter rail) stations. This region definition results in a region with Boston and its adjoining communities, although also results in communities in overlapping regions, which we discuss further.

In particular, are the coastal communities Quincy and Braintree part of the Greater Boston region as they have T stops or are they part of the South Shore? Similarly, is Revere part of Greater Boston or the North Shore? In each case we primarily identify them with their coastal region, but identify Greater Boston as an overlapping region for these communities.

Similarly, what about Milton and Newton, which are communities immediately south and west of Boston. In each case we primarily include them in an overlapping region. The end result is that our Greater Boston region includes Boston, Brookline and inner core communities immediately north of Boston.

4.2 Inland Towns of the North and South Shore

The North and South Shore regions stretching north and south from Boston clearly include coastal communities in each direction. What about the next "layer" of communities not directly on the coast, but adjacent to those coastal communities. As shown on the map, these communities are generally included as part of the "Shore" region with overlapping regions (e.g. Merrimack Valley, Cranberry) identified for some of them.

The resulting North and South Shore regions are each relatively larger in terms of number of communities, but other than Cape Ann as a separate region on the north coast there are no clear justifications for smaller regions.

4.3 Western Boston Suburbs

The Metro West region generally includes suburban communities west of Boston between the Route 128/I-95 and I-495 corridors, but what about the communities between the two regions? As shown in Figure 2, these communities around and inside the Route 128/I-95 corridor along the Charles River are clustered as the Inner West region, although historic Lexington is primarily included in the Minute Man region.

4.4 Southeast Massachusetts

The broad area of Massachusetts between Boston, Rhode Island and Cape Cod is generally known as Southeast Massachusetts and includes a large number of cities and towns. Communities to the east and south of this area cluster into the South Shore and South Coast regions, but what about the remaining inland communities? As shown in Figure 2, these communities cluster into four, sometimes overlapping, Metro South, Cranberry, Hockomock and Neponset/Tri Valley regions.

4.5 Central Massachusetts

Worcester County is typically identified as Central Massachusetts, but what are its distinct regions? We define a Greater Worcester region with Worcester and its immediately adjoining communities. The remaining Central Massachusetts communities north, west and south of Greater Worcester then cluster into a number of (sometimes overlapping) regions.

4.6 Western Massachusetts

Western Massachusetts encompasses the Pioneer Valley on either side of the Connecticut River as well the Berkshires. Are there regions within each of these areas? What about the large area in between? Our resulting clustering groups the Pioneer Valley communities into the Mohawk Trail, Five College and Greater Springfield regions. Similarly, the Berkshires area is grouped into the Mount Greylock, Central Berkshires and Southern Berkshires with the Hill Towns region between these larger areas.

4.7 Communities in More Than Two Overlapping Regions

Roughly 25% of cities and towns are in more than one overlapping region, but based on our research almost all of them are identified in no more than two of the regions. However, we identify three communities south of Boston in three overlapping regions. These communities are Canton, Sharon and Stoughton, which are each in the Neponset/Tri Valley, Metro South and Hockomock regions.

4.8 Communities Without a Clear Primary Region

In some cases, communities that did not have strong affinity for an existing region led to the identification of new regions. However in other cases, communities were not necessarily a strong fit for an existing region but identification of an additional region was also not obvious. We cite three examples where in each case the communities are associated with the best fitting, but necessarily a strong fitting region:

1. The town of Whately is on the edge of the Five College region, but not clearly identified with Mohawk Trail region.

2. Boxford, Georgetown and Middleton are inland North Shore region towns, but further removed from the coast yet also not strongly part of the Merrimack Valley region.

3. Towns such as Berlin, Boxborough, Clinton and Maynard are in the I-495 area between I-290 and Route 2, which puts each of them on the edge of regions such as the Metro West, Minute Man, Wachusett, Nashoba Valley and North Central regions without a strong affinity for any of them.

5 Representing Communities in Overlapping Regions

Figure 2 shows each city and town in one and only one region. However as we have discussed, roughly 25% of communities belong to more than one of our 35 regions. In our associated Web site for the project, we include the list of communities in each region and indicate which belong on multiple regions.

We also take two separate approaches for representing these overlapped communities in maps of the regions. In the first approach we use the same colors for communities in each region, but show the color in a lighter intensity for those communities in the region that overlap with another region. The resulting map is shown in Figure 3 and available in interactive form on our Web site.

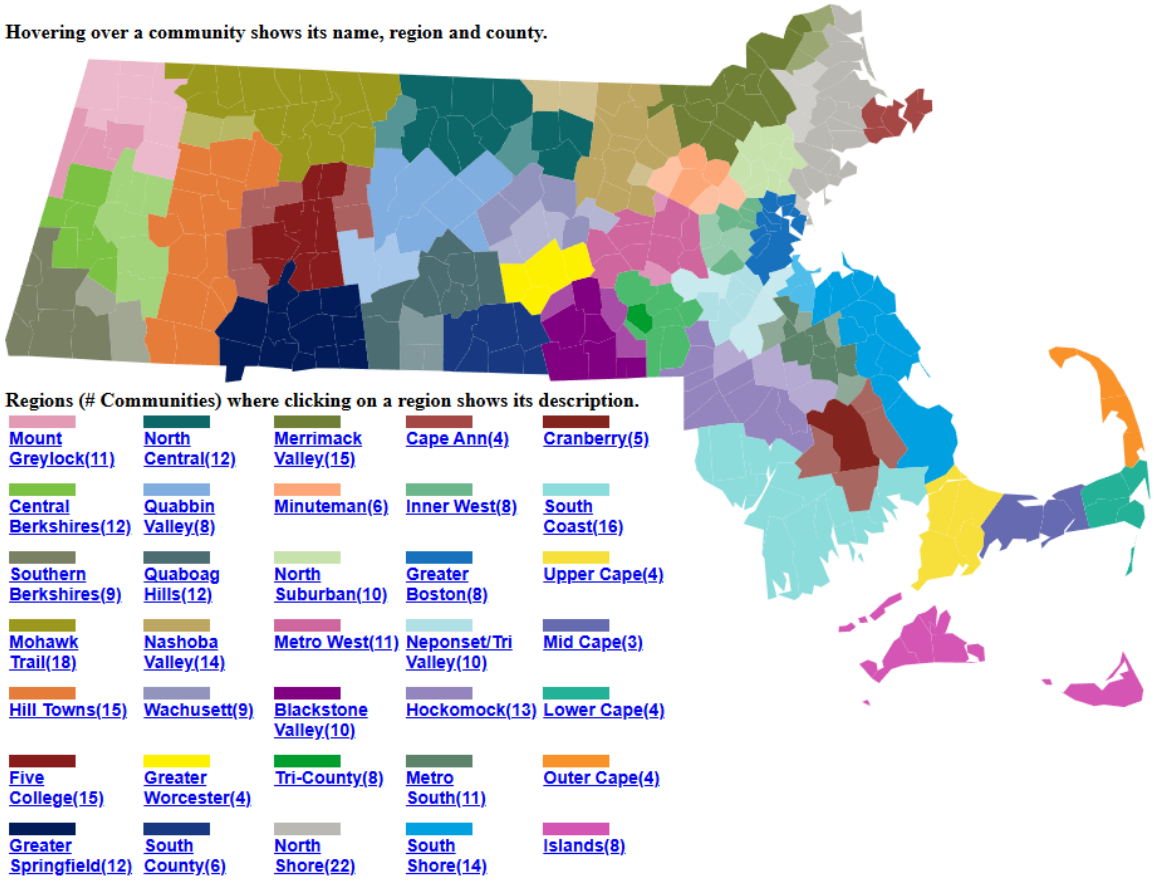

Figure 3: Map of Massachusetts with Lighter Color Intensity for Communities in Overlapping Regions

An example of this change in color intensity is for the Cranberry region in southeast Mas-

sachusetts. The town of Middleborough is shown in full color intensity while Plympton and Carver to the east are shown in lighter intensity as they overlap with the South Shore region. Similarly, the region towns of Lakeville and Rochester are shown in lighter intensity because they overlap with the South Coast region.

In the second approach we blend the overlapping region colors for towns in multiple regions with a stronger weighting for the color of the primary region. The resulting map is shown in Figure 4 and available in interactive form on our Web site.

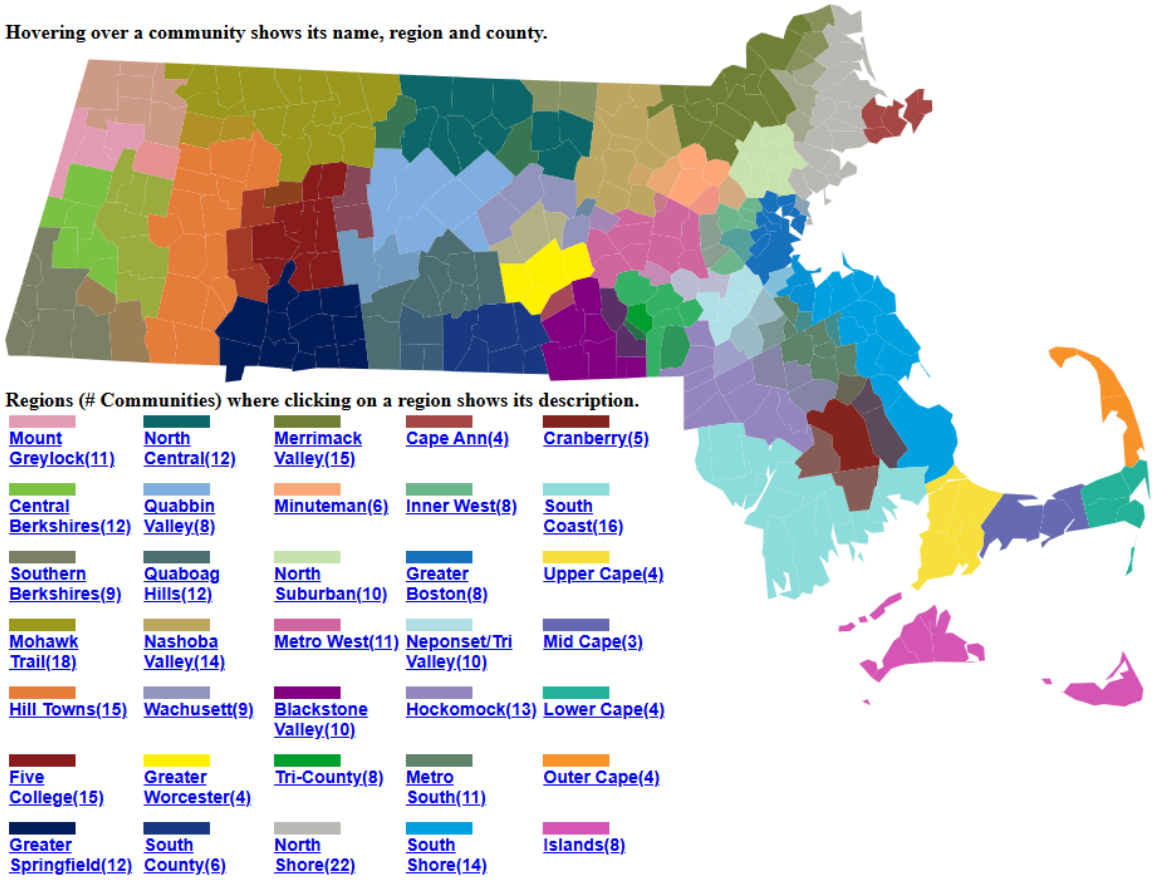

Figure 4: Map of Massachusetts with Blended Colors for Communities in Overlapping Regions

An example of this color blending is again evident for the Cranberry region. The town of Middleborough is shown in the cranberry color of the region while Plympton and Carver to the east are shown in a blending of the Cranberry and South Shore region colors. Similarly, the region towns of Lakeville and Rochester are shown in a blending of colors from the Cranberry and South Coast regions.

6 Summary and Future Work

In summary, this work applies the technique of clustering to identify regions for the 351 cities and towns in Massachusetts. We cluster Massachusetts communities into cohesive and locally

meaningful regions each with a descriptive name to represent it. The result is 35 regions, each with name, description as well as the list of cities and towns in the region.

We have not created "new" regions, but rather assembled the specific regions that are locally, but maybe not widely, known and stitched them together into a mosaic that encompasses the entire state. This mosaic is available as an interactive map with each region shown in a color representative of the region and each community shown in its primary region if it belongs to more than one.

Roughly 75% of cities and towns belong to a single primary region with the remaining communities in overlapping regions, which are represented in two additional interactive maps.

There are at least two directions for future work. First, there will always be the fine-tuning of regions and region membership as more evidence is uncovered on the primary and overlapping regions for communities. Second, there is the opportunity to apply the same self-identifying region methodology of this work to discover and map regions in other states.

References

[1] Metropolitan Area Planning Council. Greater Boston's 101 Cities and Towns. https://www.mapc.org/resource-library/101-cities-towns/.