

Analysis of Hurricane Helene Impacts and the 2023 “Year of the Trail” Initiative Using AllTrails Usage Data

Zakk Heile

December 24, 2024

Abstract

This report investigates two distinct topics employing monthly AllTrails [1] usage data across North Carolina counties from January 2022 through December 2024: (1) The impact of Hurricane Helene (late September 2024) on Western NC trail visitation, and (2) The 2023 “Year of the Trail” initiative’s effect on trail usage compared to 2022.

Methods include data cleaning, feature engineering (e.g., monthly-to-quarter aggregations, ratio calculations), various normalization options (z-score, min-max), clustering (K-means, Hierarchical, Spectral), and dimensionality reduction (PCA, PaCMAP). We discuss assumptions, present results for both the hurricane impact and the Year of the Trail, and highlight major insights derived from the analysis.

Contents

- 1 Introduction** **3**

- 2 Hurricane Helene Impact Analysis** **4**
 - 2.1 Data Subset and Feature Preparation 4
 - 2.2 Dimension Reduction (PCA, PaCMAP) and Pairwise Visualization 4
 - 2.3 Clustering Methods 7
 - 2.4 Estimating Impact Ratios for Western Counties 11
 - 2.5 Conclusion 15

- 3 Year of the Trail (2023) Analysis** **17**
 - 3.1 Motivation and Data Framework 17
 - 3.2 Statistical and Clustering Methods 17
 - 3.3 Quarter-Based Ratios 19
 - 3.4 Overall Growth 20

- 4 Discussion, Limitations, and Conclusion** **21**

1 Introduction

The AllTrails dataset recorded monthly hiking or trail usage, by county, in North Carolina from January 2022 to December 2024. This information requires an AllTrails user to actively "record" their movement along a trail on the AllTrails app. To investigate Hurricane Helene's impact and the success of the 2023 Year of the Trail, the dataset was *transposed* so that:

- Each row corresponds to one county.
- Each column corresponds to a specific month (e.g., 2024-08).

We note that AllTrails usage can serve only as a proxy or estimate of real visitation; variations in platform popularity or user demographics can create biases in the data. For instance, one may suspect that users are more likely to record hikes at state parks versus local parts on AllTrails. However, by normalizing the rows or comparing *ratios* (e.g., 2024 usage / 2023 usage for the same month and same county), many of these biases likely reduce because we are just focusing on trends for individual counties.

Additionally, certain months or columns can be excluded or aggregated to simplify analysis, including:

- Aggregating months into quarters (e.g., Q1, Q2, Q3, Q4).
- Filtering out columns before August 2024 or focusing only on September–December 2024.
- Computing multiplicative ratios to compare month in one year vs. the same month in the previous year.

To remove the effect of county size or popularity from the clustering, optional *normalizations* can be applied per-row (min-max scaling or z-score scaling).

Overview of Report Organization. First, we focus on the Hurricane Helene analysis, describing the methods, dimensionality reductions, clustering approaches, and results. Afterwards, we present the analyses around the 2023 Year of the Trail initiative, including quarter-based comparisons and overall year-over-year growth.

2 Hurricane Helene Impact Analysis

2.1 Data Subset and Feature Preparation

The 2024 hurricane of interest occurred in late September 2024, with potential spillover effects in October, November, and even December 2024. We therefore focus on:

- **Western NC counties** marked as heavily impacted by the State of North Carolina (e.g., Buncombe, Haywood, Yancey, etc.).
- **Columns from August 2024 to December 2024** to capture pre- and post-hurricane months.

We also create *ratio features*:

$$\text{Ratio}_{(\text{County}, \text{Month})} = \frac{\text{Usage}_{(\text{County}, \text{Month in 2024})}}{\text{Usage}_{(\text{County}, \text{Month in 2023})} + \epsilon} \quad (1)$$

where $\epsilon = 0.01$ is added in the denominator to avoid division by zero. These ratio columns (e.g., 2024-Aug / 2023-Aug, 2024-Sep / 2023-Sep, ..., 2024-Dec / 2023-Dec) form a 5-dimensional feature space if we select August–December 2024. Each county is thus a single point in a 5D space.

2.2 Dimension Reduction (PCA, PaCMAP) and Pairwise Visualization

PCA [2] We apply Principal Component Analysis (PCA) to the 5-dimensional ratio space.

- **Eigenvectors & Eigenvalues:** PCA solves for the eigenvectors of the covariance matrix, projecting data onto orthogonal directions of maximal variance. The first principal component often captures the largest fraction of variance.
- **Example Result:** In our data, the first principal component can nearly capture 80%+ of the total variance.

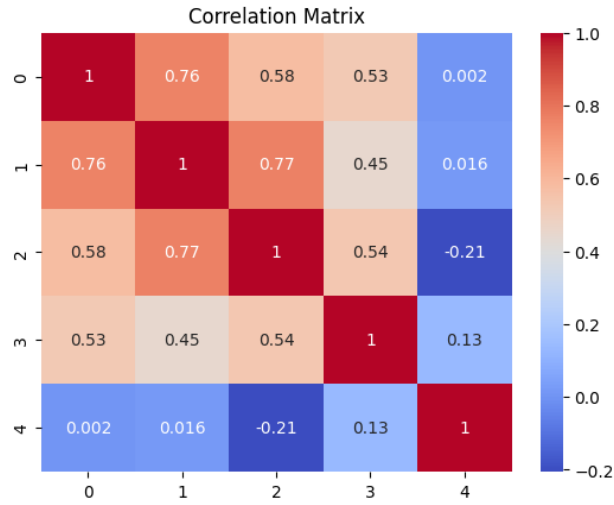


Figure 1: Pairwise Correlation of 5 Ratio Features

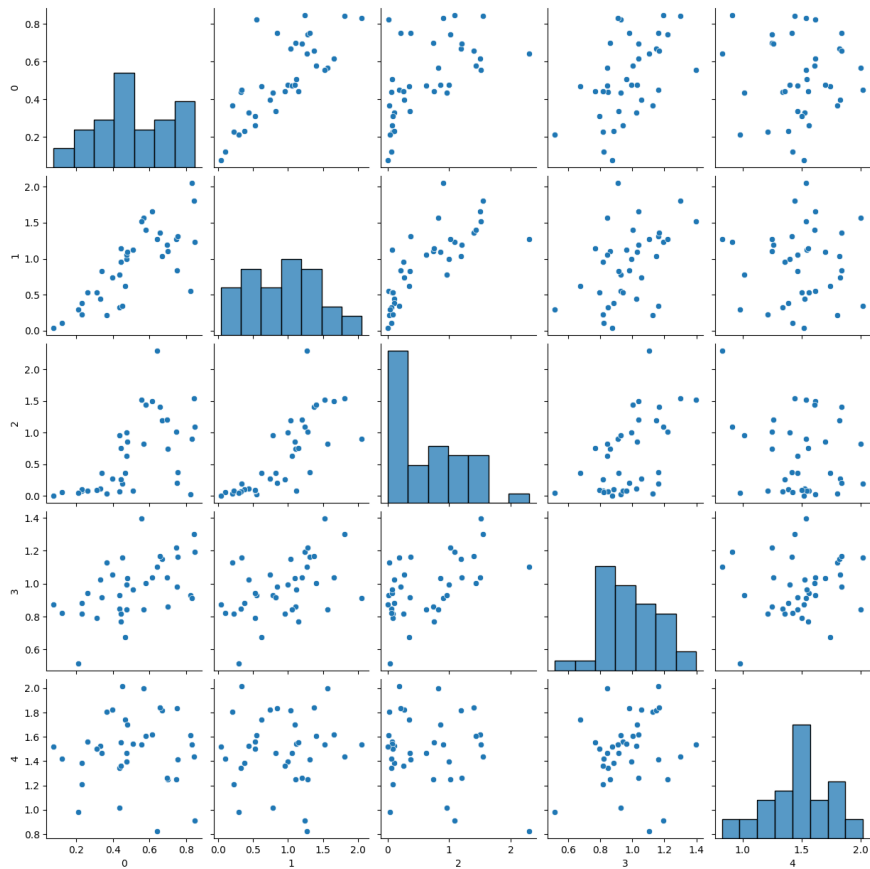


Figure 2: Pairplot of Variables

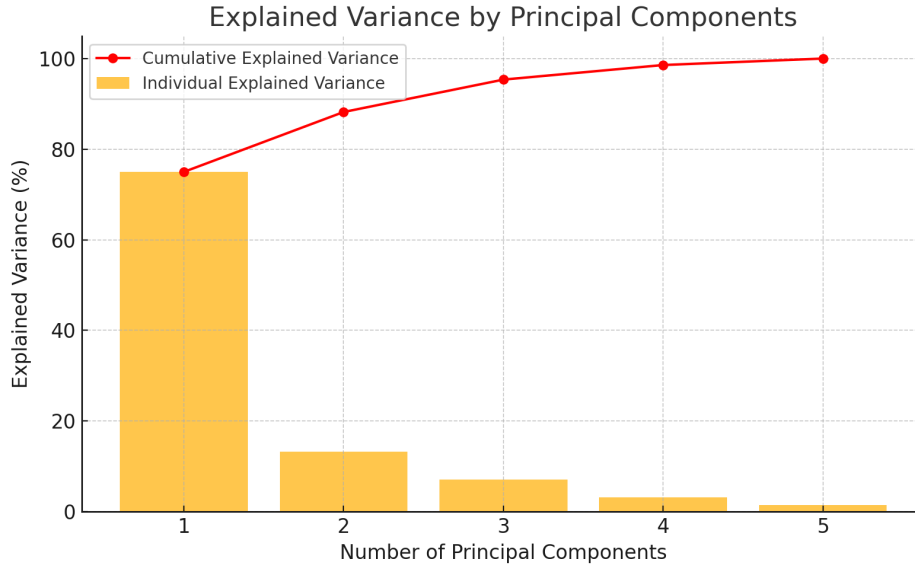


Figure 3: Variance Captured by Projection of Varying Dimensions

Component	Eigenvalue
1	0.568030
2	0.100135
3	0.054172
4	0.024371
5	0.010833

Table 1: Eigenvalues of PCA Components

Reconstruction Type	MSE	Manhattan Distance
1 Component Reconstruction	0.036878	0.142089
2 Component Reconstruction	0.017392	0.102187

Table 2: Reconstruction Errors

Statistic	Value
Average Row Sum	4.528379
Average Row Sum of Squares	5.412948

Table 3: Average Row Statistics for Benchmarking

Based on the above results, we lose minimal information when projecting onto a 2D or even 1D subspace.

North Carolina Counties with PCA Scores

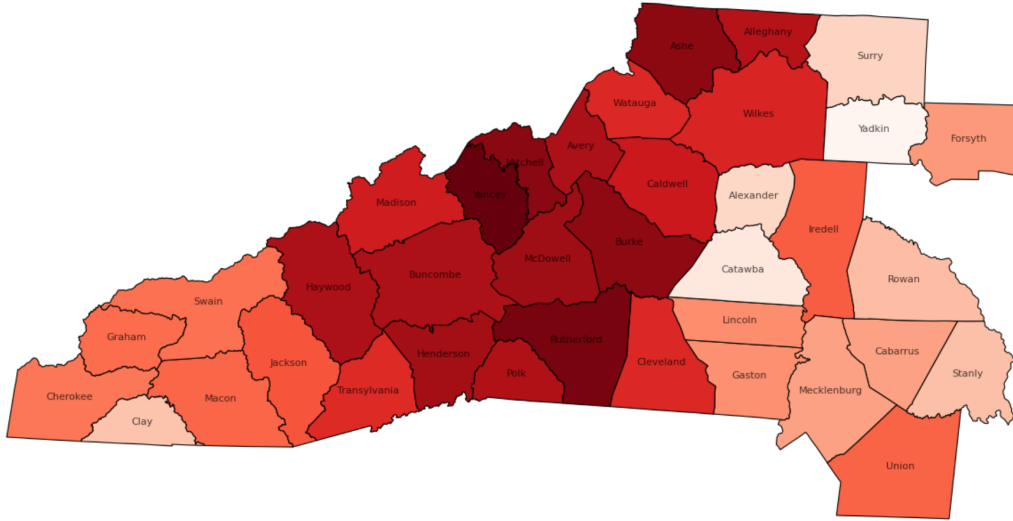


Figure 4: Counties by their projection onto the vector capturing the most variance

PaCMAP. We also use *PaCMAP* [3] to reduce from 5D to 2D for more visually interpretable cluster plots. PaCMAP is a new manifold approach that preserves local and global structures better than older methods.

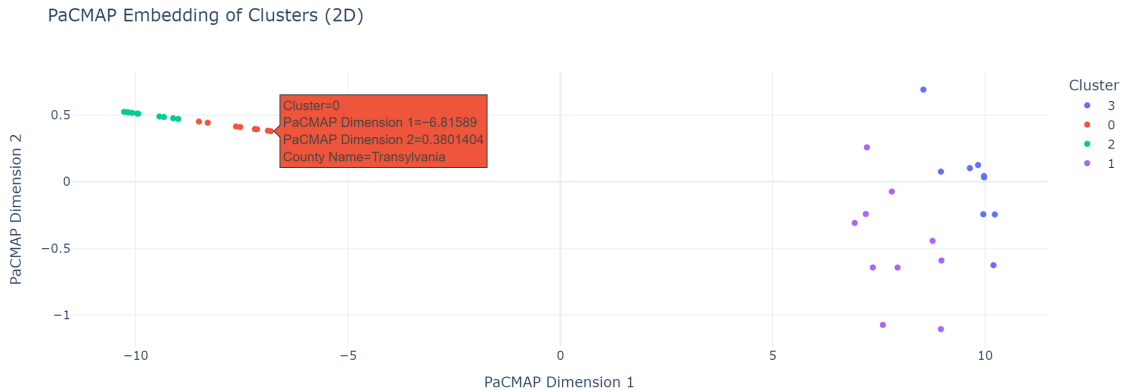


Figure 5: PaCMAP Dimension Reduction

From the dimension reduction, we would expect the data to be logically clustered into 3 or 4 clusters. Our next analysis confirms this.

2.3 Clustering Methods

We tried several clustering algorithms on this 5-dimensional ratio data:

1. K-Means [4]

2. Agglomerative Hierarchical Clustering [5]

3. Spectral Clustering [6]

All three clustering algorithms created nearly identical clusters for the same number of clusters.

We choose to perform clustering with $k = 4$ clusters because the K-Mean's loss as a function of the number of clustering starts to have diminishing improvements (the second derivative magnitude is decreasing).



Figure 6: Loss as a Function of the Number of Clusters

North Carolina Counties Hierarchical Clustering (k=4)

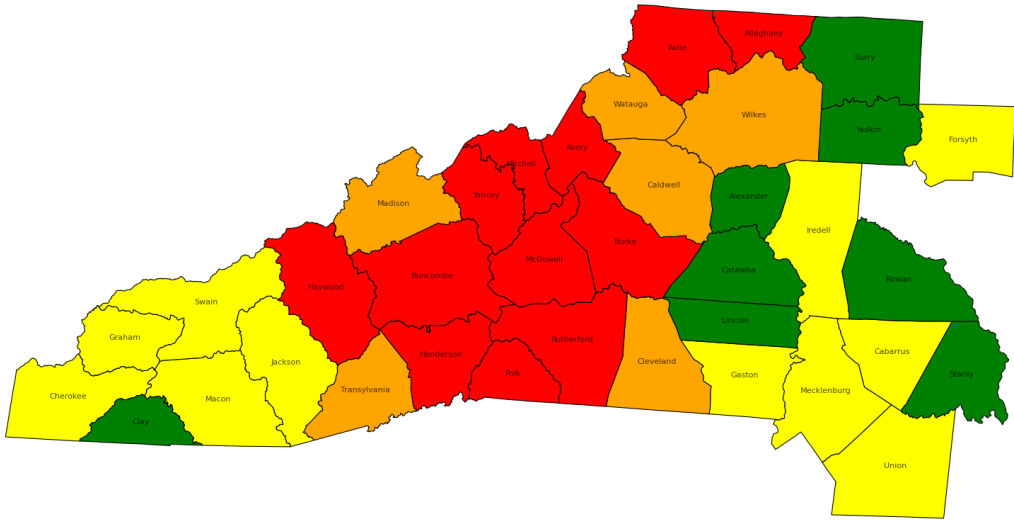


Figure 7: Clustering of 5D Space with 4 Clusters

As Hierarchical Clustering starts with each county in its own cluster and merges them together, we can produce a Dendrogram that shows this iterative process.

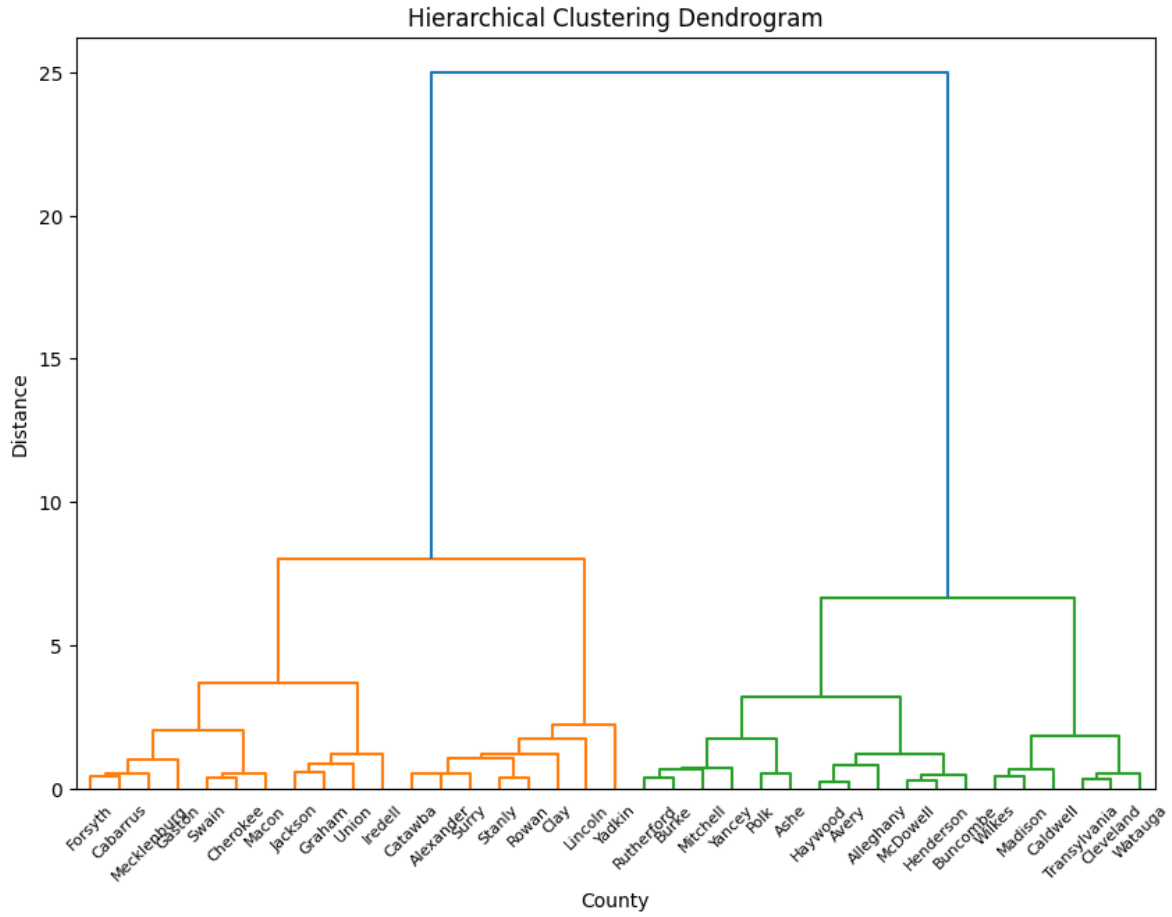


Figure 8: Dendrogram

For interest, we include clusters for a surplus of numbers of clusters, computed with K-Means.

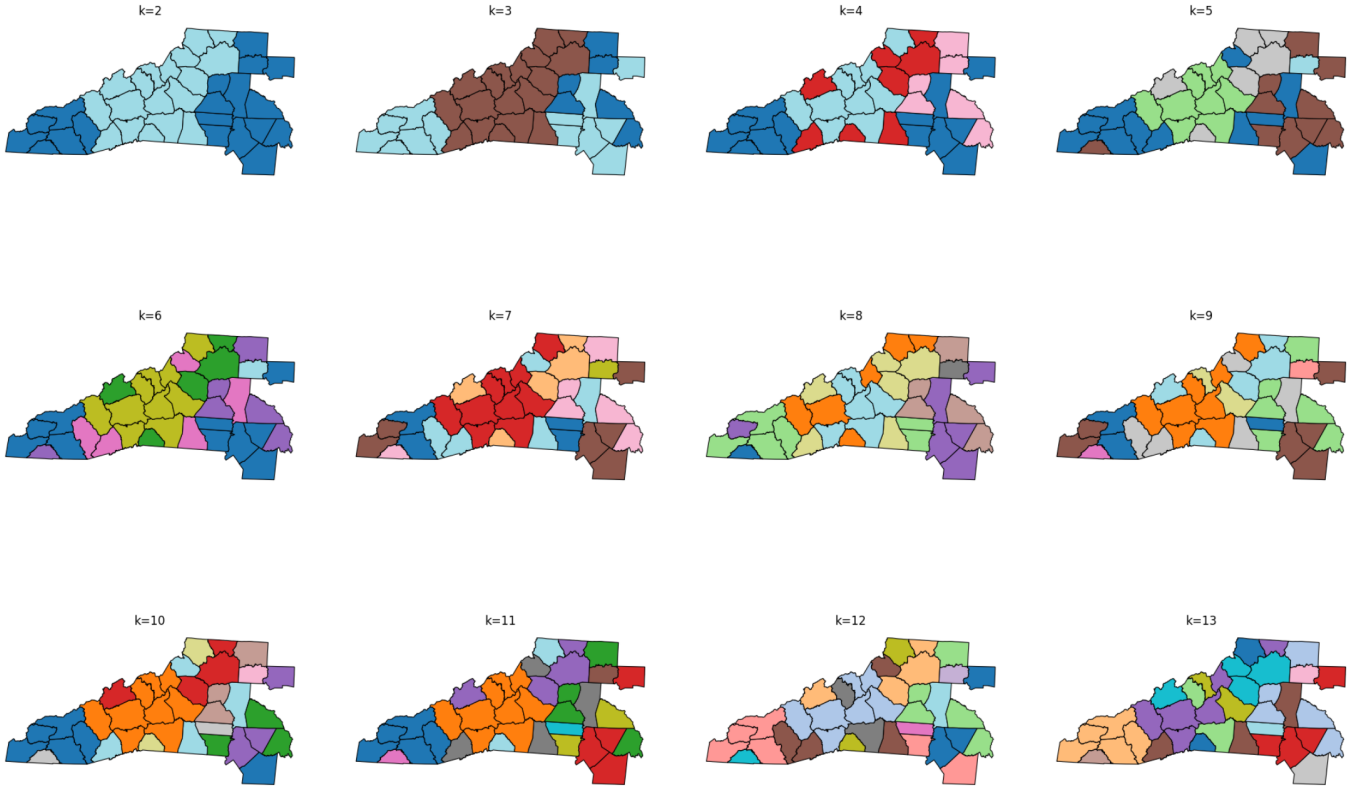


Figure 9: Varying Number of Clusters

2.4 Estimating Impact Ratios for Western Counties

Beyond clustering, we quantify the *impact* of Hurricane Helene by:

1. **Compute an average growth factor** from 2023 to 2024 for all *non-impacted* counties (i.e., a baseline ratio).
2. **Estimate expected 2024 usage for impacted counties** by multiplying each impacted county's 2023 usage by that baseline ratio.
3. **Divide actual 2024 usage by the estimated 2024 usage**, so:

$$\text{Impact Ratio} = \frac{\text{Actual 2024 Usage (Impacted County)}}{\text{Expected 2024 Usage (Impacted County)}}. \quad (2)$$

A value > 1 indicates the actual usage was higher than expected (i.e., possibly not as severely impacted), and a value < 1 indicates the county underperformed relative to the baseline (likely a hurricane-related drop).

Impact Ratio - October

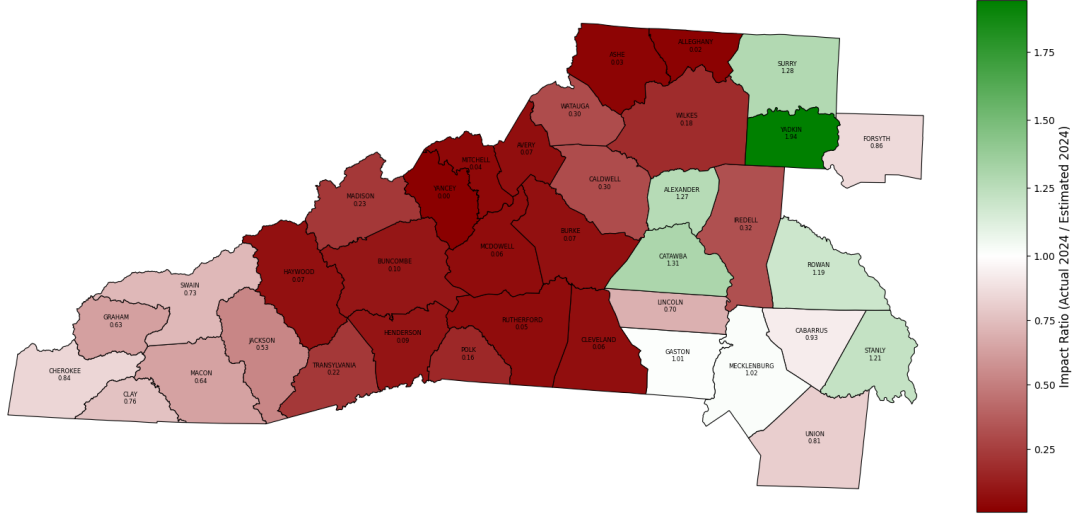


Figure 10: October Impact Ratios

Impact Ratio - November

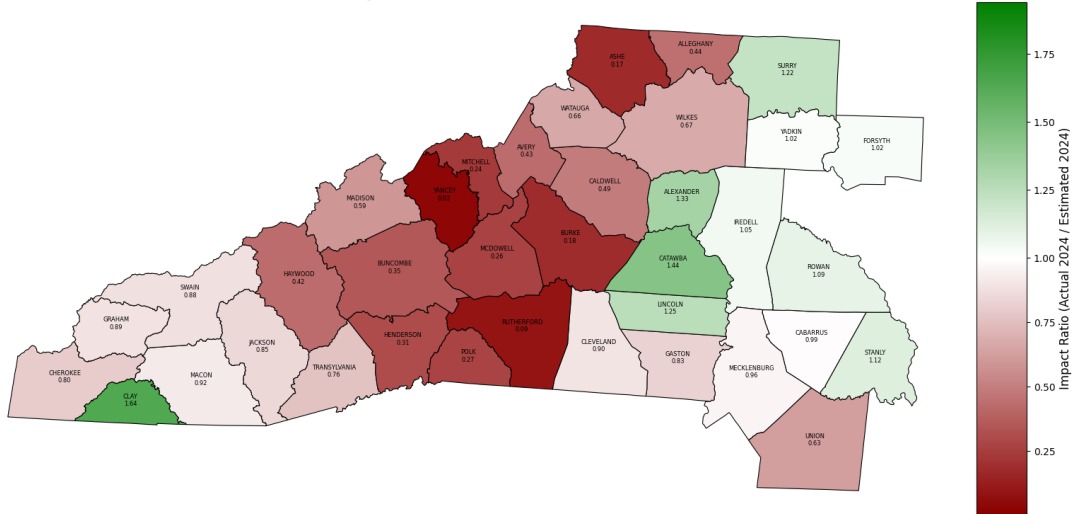


Figure 11: November Impact Ratios

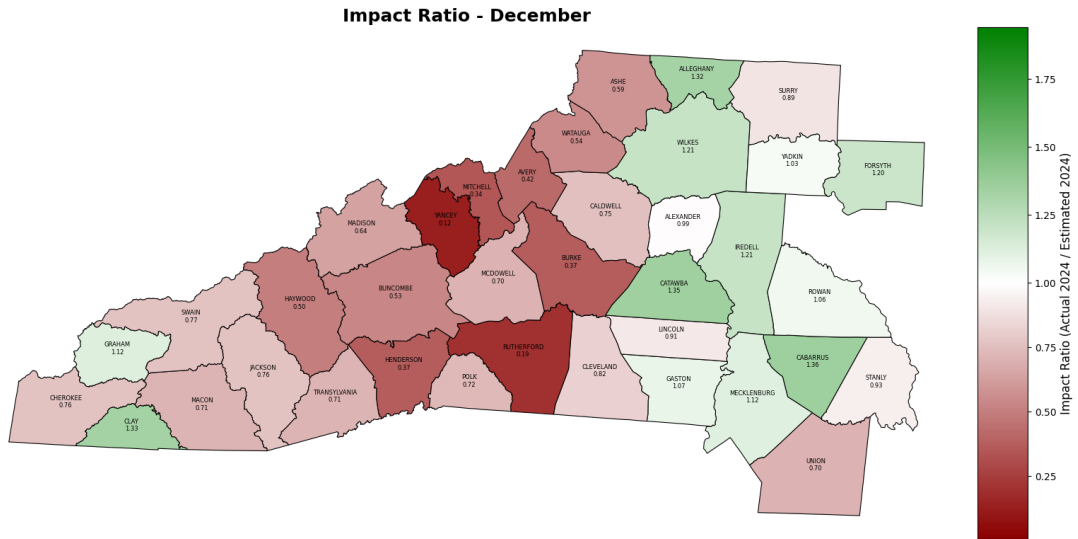


Figure 12: December Impact Ratios

In particular, we bring your attention to Yancey County, going from 1640 recorded hikes/uses in October 2023 to 8 in October 2024.

We also created impact ratios for page views instead of recordings, which were very similar to those above. For instance, Buncombe was 0.11 in October instead of 0.1, Swain was 0.73 instead of 0.69, and so on. An observation that holds for the majority of counties is that page views were less affected than hike recordings. This suggests that people were curious about what trails were open during the aftermath of the hurricane.

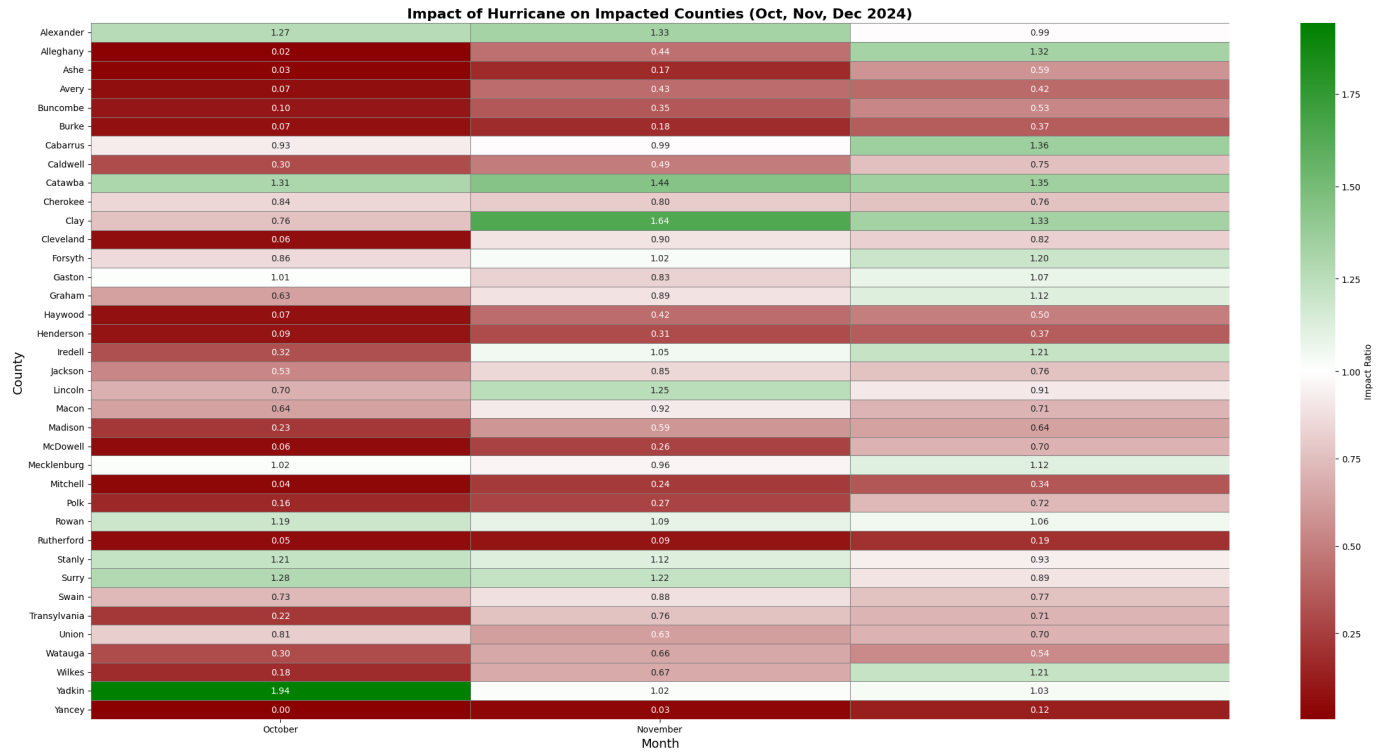


Figure 13: Numerical Comparison

Other items of note:

All state parks west of I-77 were closed through at least October 31, 2024:

- Chimney Rock State Park - Rutherford County
- Crowders Mountain State Park - Gaston County
- Elk Knob State Park - Watauga County
- Gorges State Park - Transylvania County
- Grandfather Mountain State Park - *Avery / Caldwell / Watauga counties
- Lake James State Park - Burke and McDowell Counties
- Lake Norman State Park - Iredell County
- Mount Jefferson State Natural Area - Ashe County
- Mount Mitchell State Park - Yancey County
- New River State Park - Ashe and Alleghany counties
- Rendezvous Mountain - Wilkes County
- South Mountains State Park - Burke County
- Stone Mountain State Park - Alleghany and Wilkes counties

Most events and programs scheduled for October were canceled at ALL state parks across NC.

Opened Nov 1 - Grandfather Mountain, Gorges, Crowders Mountain, Lake Norman and Rendezvous Mountain.

Opened Nov 19 or soon after - Elk Knob, Lake James, New River and Stone Mountain state parks and Mount Jefferson Natural Area - at least partially.

State Parks still closed - Chimney Rock, Mount Mitchell, South Mountains

2.5 Conclusion

In summary, the data suggests that certain western counties show significantly lower 2024 usage than the *expected* usage (i.e., Impact Ratio < 1). Other counties remain stable or even show slight gains. The cluster structures align moderately with geographic groupings and severity of impact, validating the initial hypothesis that Hurricane Helene had a region-specific effect.

Our impact ratios and clusters closely align with the "Hurricane Helene Damage and Needs Assessment" by NC OSBM [7]:

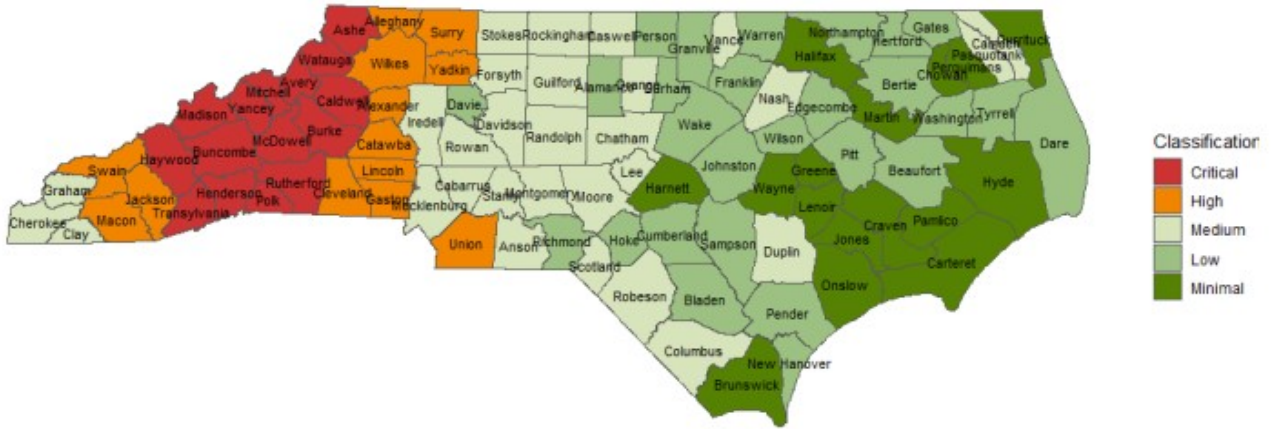


Figure 14: From the "Hurricane Helene Damage and Needs Assessment" by NC OSBM

3 Year of the Trail (2023) Analysis

3.1 Motivation and Data Framework

The state of North Carolina declared 2023 the “Year of the Trail,” sparking media and local promotions. We aim to see if the initiative corresponded to *increased AllTrails usage* relative to 2022. Similar to the hurricane analysis, we compare 2023 vs. 2022 usage through:

- **Ratios:** $Usage_{2023}/Usage_{2022}$ by county and month.
- **Quarterly Aggregations:** Grouping Jan–Mar (Q1), Apr–Jun (Q2), Jul–Sep (Q3), Oct–Dec (Q4).

We can either analyze the 12-dimensional monthly feature space (all 12 months in 2023 vs. 2022) or just compare total usage year-over-year (1-dimension).

3.2 Statistical and Clustering Methods

We again apply:

- **Clustering on Ratios** to see if counties being similar in growth lines up geographically.

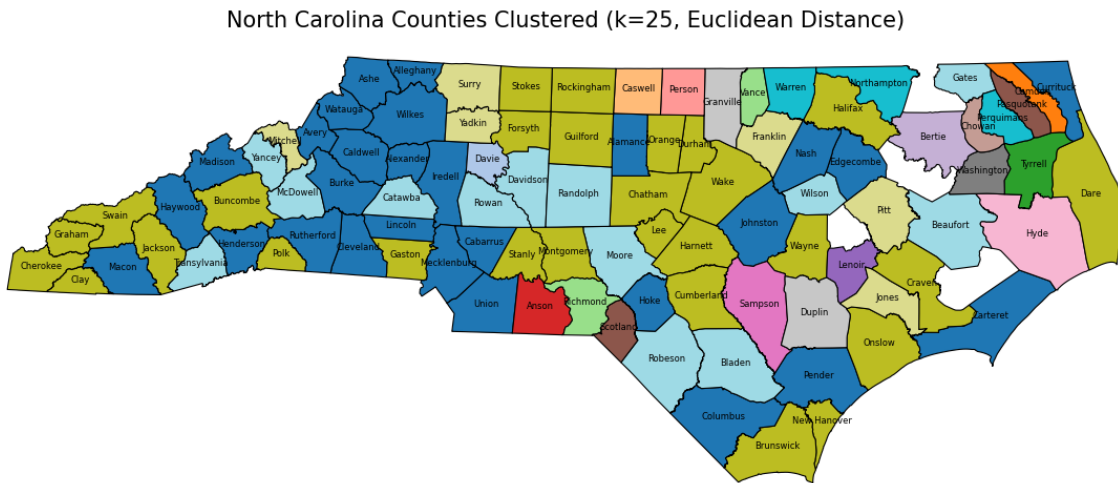


Figure 15: YOTT Clusters

- **PCA, Pairplots, PacMap** to reduce dimensionality from 12 months to 1 or 2 components.

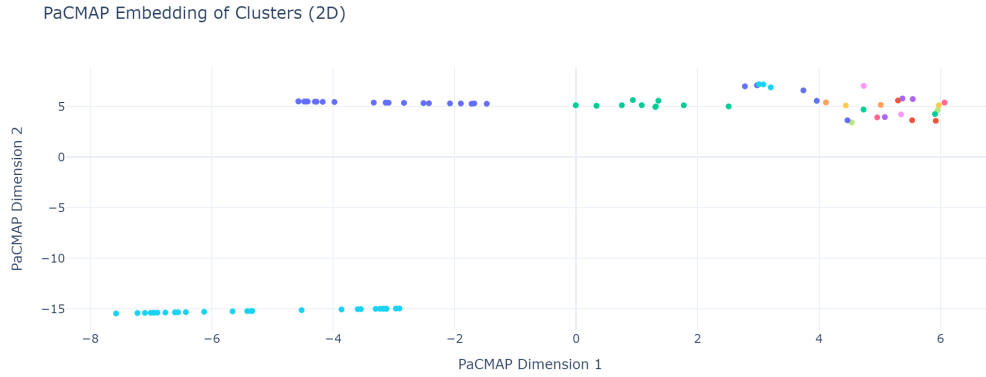


Figure 16: 12D to 2D Reduction

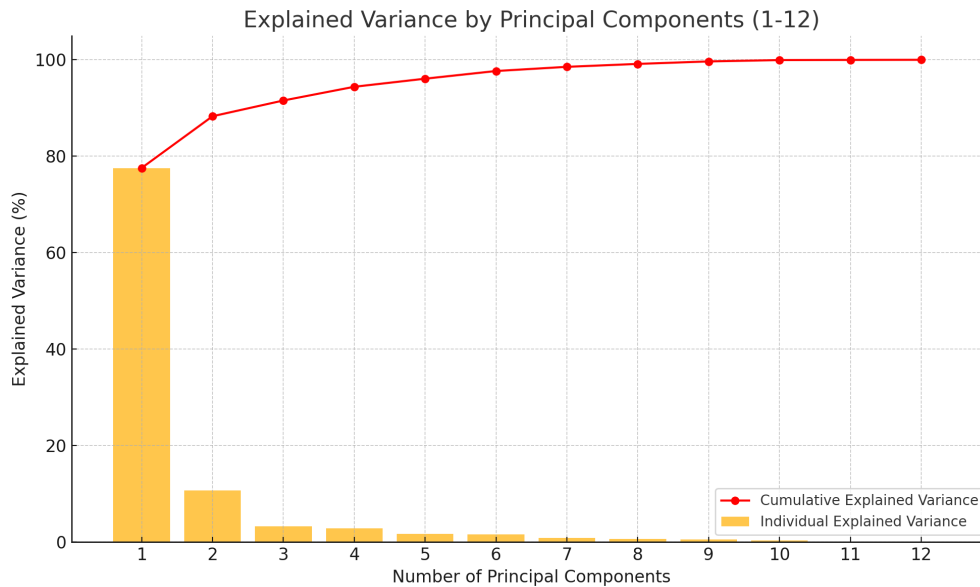


Figure 17: Variance Captured by Principal Components

Cautions. We remind the reader that usage could be influenced by platform popularity growth or new trails listed on AllTrails. That is to say that these ratios could be inflated if AllTrails has grown in popularity in the past year. However, assuming the growth of the platform was consistent across counties, we can still compare the ratios to other counties, considering relative growth and not interpreting an absolute number.

3.3 Quarter-Based Ratios

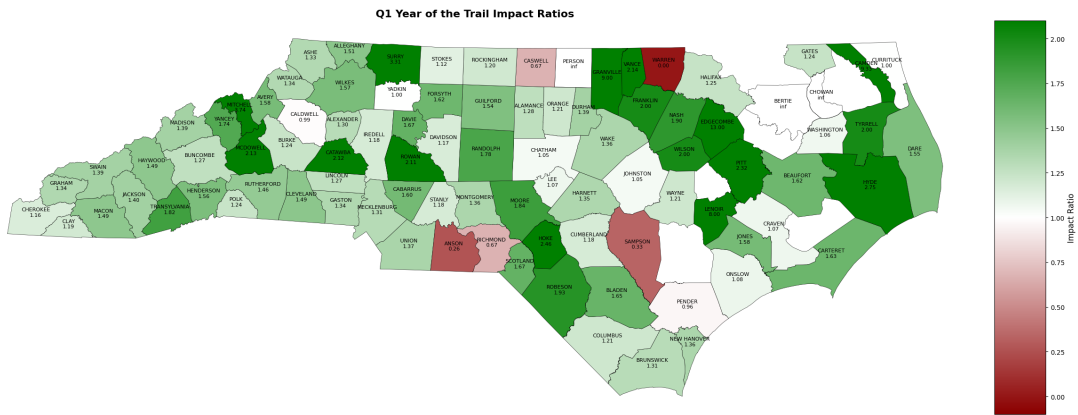


Figure 18: Quarter 1 YOTT Growth Ratios

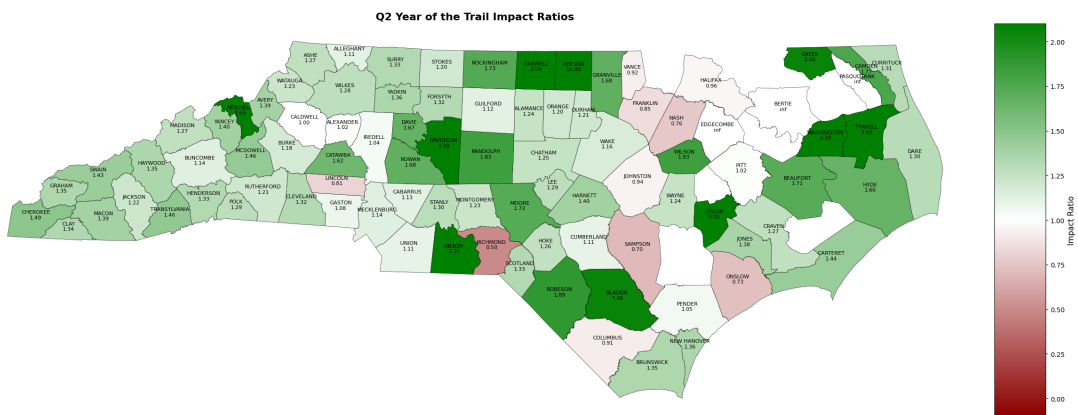


Figure 19: Quarter 2 YOTT Growth Ratios

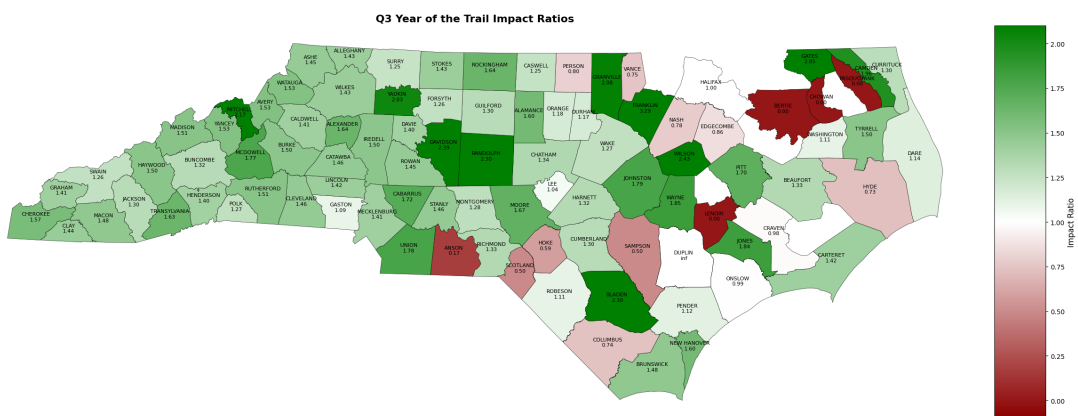


Figure 20: Quarter 3 YOTT Growth Ratios

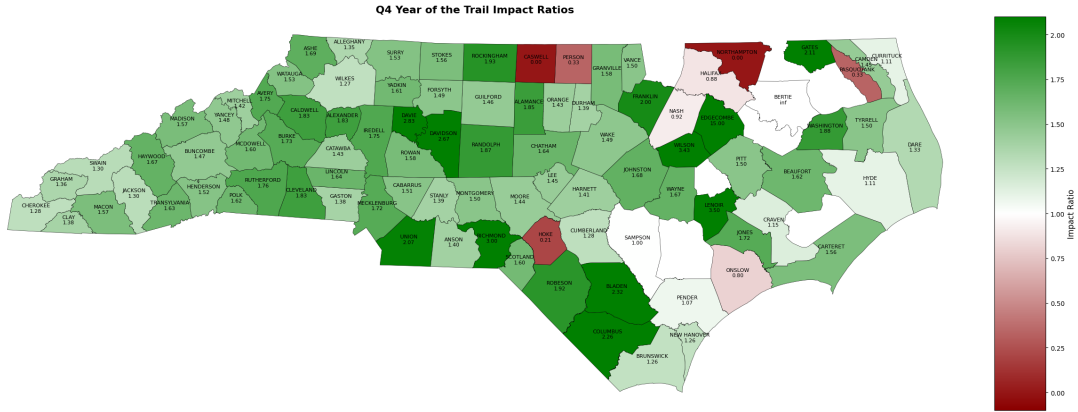


Figure 21: Quarter 4 YOTT Growth Ratios

An interesting finding is that the Q4 and Q3 ratios (2023 vs. 2022) was generally higher than Q1 or Q2, suggesting that the Year of the Trail's effect might have *increased* later in the year. This could indicate that awareness campaigns and events ramped up over time, or a general awareness of trails going into this fall combined with the initiative to yield stronger usage growth in Q3/Q4 that was not seen in the previous year.

We also note that these growth ratios were essentially the same when looking at page views instead of hike recordings.

3.4 Overall Growth

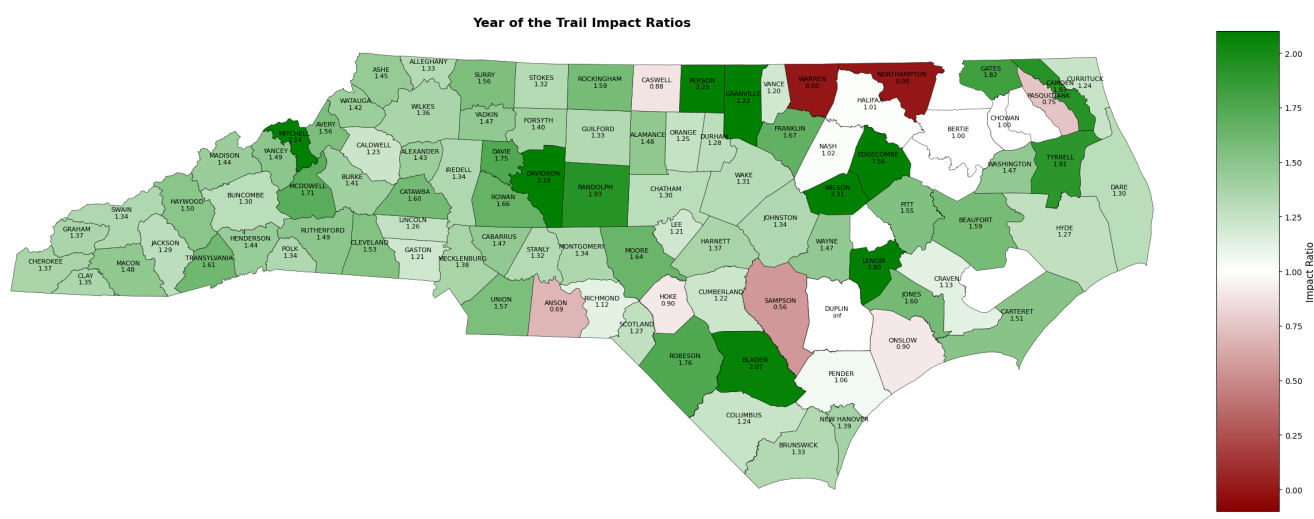


Figure 22: Overall YOTT Growth

4 Discussion, Limitations, and Conclusion

Key Insights.

- **Hurricane Helene Impact.** Western NC counties showed a measurable dip (90%) in 2024 usage relative to expected baseline growth, suggesting the hurricane substantially disrupted typical trail visitation. Clustering in the 5D feature space was consistent with known impacted areas.
- **Year of the Trail (2023).** 90 counties had positive growth in trail use from 2022 to 2023, with the majority of counties' usage growing by 30-40%, indicating substantial growth in usage. Data hints that Q3/Q4 were particularly robust, possibly reflecting momentum from the initiative.

Data and Method Limitations.

- **AllTrails as a Proxy.** Not all hikers use AllTrails, and platform usage may change over time for reasons unrelated to actual trail visitation.
- **Bias Towards State Parks.** It is plausible that AllTrails users are more likely to record hikes to state parks compared to local parks so hikes at state parks could be over-represented in this sample compared to the overall data distribution. This could mean that counties with state parks could appear more impacted by our metrics if state parks were closed for longer than local parks.
- **Estimating Visitation if no Hurricane.** Even though our attempt to multiply 2023 visitation per county by the growth the rest of the state saw from the same month one year ago is a good start, it is possible that counties have different growth trajectories and this is an overgeneralization. However, compared to the alternative of calculating a growth multiple based on say August 2024 to August 2023, or another month right before the hurricane, this has the advantage of preserving the season and general weather and climate trends that cannot be replicated at another timestamp.

Conclusion. We presented a comprehensive pipeline for cleaning and transforming monthly AllTrails usage data, applying various clustering, dimensionality reduction, and data analysis techniques to address:

1. The local *disruption* from Hurricane Helene in late September 2024 in Western NC.
2. The *spatially broader* usage growth from the 2023 Year of the Trail.

Results affirm that certain western counties faced sharp visitation drops in the fall of 2024, while 2023 usage displayed a noticeable positive jump over 2022, especially in later quarters. Future work could incorporate more external validation (e.g., official park statistics, AllTrails user growth) to confirm the magnitude of these patterns.

References

- [1] “AllTrails: Trail Guides and Maps for Hiking, Camping, and Running,” Available: <https://www.alltrails.com/>.
- [2] I. Jolliffe, “Principal Component Analysis,” *Springer Series in Statistics*, 2002. DOI: <https://doi.org/10.1007/b98835>.
- [3] Y. Wang, H. Huang, C. Rudin, and Y. Shaposhnik, “Understanding How Dimension Reduction Tools Work: An Empirical Approach to Deciphering t-SNE, UMAP, TriMap, and PaCMAP for Data Visualization,” *Journal of Machine Learning Research*, vol. 22, no. 201, pp. 1–73, 2021. Available: <http://jmlr.org/papers/v22/20-1061.html>.
- [4] J. MacQueen, “Some Methods for Classification and Analysis of Multivariate Observations,” in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1967, pp. 281–297.
- [5] G. N. Lance and W. T. Williams, “A General Theory of Classificatory Sorting Strategies: 1. Hierarchical Systems,” *The Computer Journal*, vol. 9, no. 4, pp. 373–380, 1967. DOI: <https://doi.org/10.1093/comjnl/9.4.373>.
- [6] A. Ng, M. Jordan, and Y. Weiss, “On Spectral Clustering: Analysis and an Algorithm,” in *Advances in Neural Information Processing Systems*, 2002, pp. 849–856.
- [7] “Hurricane Helene Damage and Needs Assessment,” Office of State Budget and Management, North Carolina, 2024. Available: <https://www.osbm.nc.gov/hurricane-helene-dna/open#:~:text=Updated%20estimates%20indicate%20damage%20and,tornadoes%20generated%20by%20the%20storm>.
- [8] W. McKinney, “Data Structures for Statistical Computing in Python,” in *Proceedings of the 9th Python in Science Conference*, 2010, pp. 56–61. DOI: <https://doi.org/10.25080/Majora-92bf1922-00a>.
- [9] J. Jordahl, “GeoPandas: Python Tools for Geographic Data,” 2014. Available: <https://geopandas.org/>.