

THE SOUL OF WINE

The Map and the Soul

On the Statistical Independence of Terroir and Cultural Identity in Wine Regions

The Soul of Wine Research Project · April 2026

59 wine regions · 6 identity dimensions · 2 independent clustering systems

Abstract

Wine regions are classified in two fundamentally different ways: by what the land gives them — geology, climate, topography — and by what they have made of themselves — narrative, ambition, historical depth, cultural character. This study asks whether these two modes of classification agree. Across 59 wine regions spanning 16 countries, we construct independent clustering solutions for terroir (Layer 2: six factual geographic and climatic features, TF-IDF vectorised, $k=7$) and cultural identity (Layer 1: six expert-scored identity dimensions — D1 through D6 — derived from anthropological descriptions, $k=6$). Testing their agreement using the Adjusted Rand Index (ARI), the result is unambiguous: $ARI = 0.023$, $\chi^2 p = 0.131$. The two classifications are statistically independent. A region's physical terroir tells us almost nothing about its cultural identity cluster. This finding places wine region analysis within a broader family of dual-classification studies spanning cross-cultural psychology, historical linguistics, biogeography, and neuroscience, where physical substrate and emergent identity repeatedly diverge. For the wine world, the implication is precise: terroir shapes what grapes can become; history, ambition, and culture determine what a region chooses to be.

1. Introduction

The dominant framework for understanding wine regions is geographic. Appellation systems — from Burgundy's grands crus to Napa's sub-AVAs — are organised by land. The working assumption, rarely questioned, is that physical terroir is the primary organising principle of wine regional identity. Regions in similar climatic and geological settings, the logic runs, should develop similar characters. The map, in this view, is the soul.

This study tests that assumption directly. We apply two independent classification systems to the same set of 59 wine regions: one grounded in physical terroir features, one grounded in cultural and anthropological identity dimensions. We then ask a simple question: do regions that cluster together by terroir also cluster together by identity? The answer, measured with statistical precision, is no.

Terroir shapes what grapes can become. History, ambition, and culture determine what a region chooses to be.

This finding is not an argument against terroir. It is an argument for recognising that identity is a second system — one that operates in parallel to geography, governed by its own logic. A region with harsh volcanic soils may develop an identity of survival and struggle, or it may develop one of ancient civility and philosophical depth. A young New World region with ideal growing conditions may channel those conditions into reinvention and ambition, or into quiet confidence. The physical substrate creates a range of possible identities; which one emerges is determined by history, culture, and human choice.

The structure of this paper is as follows. Section 2 surveys related literature across four academic fields where the same independence between physical substrate and emergent identity has been documented. Section 3 describes the data and methodology. Section 4 presents results. Section 5 discusses implications. A closing section, written for the wine-focused reader, translates the key findings into the language of regions, metaphors, and wine character.

2. Related Literature

The question posed by this study — whether two independent classification systems applied to the same objects yield the same groupings — is not new. It is, in fact, a recurring discovery across strikingly different disciplines. In each case, researchers have found that physical substrate and emergent cultural or functional organisation diverge in ways that are both statistically significant and intellectually revealing.

2.1 Cross-Cultural Psychology: Ronen, Shenkar, and the GLOBE Study

The most direct precedent comes from cross-cultural psychology. Ronen and Shenkar (1985, 2013) constructed cultural clusters of countries based on similarity in work-related attitudes — a classification grounded in values, behaviours, and social norms rather than geography. Their landmark 2013 paper in the *Journal of International Business Studies* produced a triple-nested cultural map whose boundaries frequently violated geographic expectation: the religion and language brought by the Spanish to Latin America failed to produce a cohesive cultural cluster analogous to the tight Anglo cluster; the Netherlands appeared in the Nordic group; Israel in the Germanic group. Cultural identity, they found, is shaped by geography but not determined by it.

The GLOBE study (House et al., 2004) extended this framework to 62 societies across ten cultural clusters, surveying 17,300 middle managers on nine cultural dimensions and six leadership styles. Like Ronen and Shenkar, GLOBE found that cultural clustering and geographic clustering regularly diverged — and that the divergences were as informative as the alignments. The study also introduced a methodologically important distinction between cultural practices (what is) and cultural values (what should be), a duality that maps cleanly onto our separation of *terroir* (what the land gives) from identity (what the region has chosen to make of it).

Our study extends the Ronen-Shenkar tradition from nations to wine regions, substituting attitudinal survey instruments with NLP-derived identity dimensions, and formalising the independence test using the Adjusted Rand Index rather than noting mismatches qualitatively.

2.2 Historical Linguistics: Genetic vs. Areal Classification

Linguistics has long distinguished between two modes of grouping languages that regularly conflict. Genetic classification groups languages by common ancestry — the Indo-European family, Romance languages, Germanic languages. Areal classification, the concept of the *Sprachbund*, groups languages by geographic contact: features that spread across language boundaries regardless of genealogical origin. The Balkan *Sprachbund* is the canonical example, where Albanian, Greek, Romanian, Bulgarian, and Macedonian — from distinct genealogical families — converged structurally through prolonged proximity and contact.

The parallel to this study is almost exact. Our *terroir* clustering is the areal system — grouping regions by shared physical environment and proximity. Our identity clustering is the genealogical system — grouping by inherited cultural character, transmitted through history, winemaking tradition, and narrative. And just as in linguistics, the two systems regularly disagree. Santorini, Etna, and the Hunter Valley form a kind of vinicultural *Sprachbund* — sharing a harsh-condition identity without sharing *terroir* type or geographic proximity.

Recent computational work in linguistics (Longobardi et al., 2015; EPJ Data Science, 2024) has formalised this distinction using syntactic distance matrices and phylogenetic clustering — the same class of methods employed in this study.

2.3 Biogeography: Environment vs. Evolutionary History

Biogeography provides perhaps the longest-standing empirical record of divergence between physical substrate and emergent biological organisation. Regions with similar contemporary climates can differ enormously in species assemblages because evolutionary history — isolation, extinction events, plate tectonics — is an independent force. Australia and the Afrotropics share broad climatic zones but harbour radically different biological communities. East Asia and eastern North America share temperate forest conditions but diverge in angiosperm diversity due to differential glacial history.

The methodological contribution of biogeography to this study is direct: the Adjusted Rand Index, used here to measure independence between *terroir* and identity clustering, was introduced in the biogeographic clustering literature (Hubert & Arabie, 1985) and remains the standard metric for comparing alternative clustering solutions across fields. A biogeography paper (PMC, 2016) explicitly describes its use in palaeoecological cluster comparison — providing the methodological citation chain for our application.

The core insight transfers directly: similar physical environments produce different communities because history intervenes. For wine regions, similar *terroir* types produce different identities because culture intervenes.

2.4 Neuroscience: Structural vs. Functional Brain Connectivity

The most conceptually elegant parallel comes from network neuroscience. Researchers routinely build two maps of the brain: structural connectivity (SC), describing physical white matter tracts between regions measured by diffusion MRI; and functional connectivity (FC), describing which regions activate together during rest or cognitive tasks, measured by fMRI. The central finding across decades of research is that SC and FC are related but diverge systematically in their network organisation.

Crofts et al. (Scientific Reports, 2022) formalised this divergence through a measure of structure-function clustering in weighted multiplex brain networks, finding that functional connectivity exhibits stable departures from anatomically-constrained patterns — that the brain operates in regimes where dynamical switching between functional states is actively promoted by the decoupling of structure from function. The implication: physical wiring creates the conditions for function; the emergent functional organisation follows its own logic.

Neuroscience Concept	Soul of Wine Equivalent
Structural connectivity (white matter tracts)	Terroir (geology, climate, topography)
Functional connectivity (emergent co-activation)	Cultural identity (emergent character)
SC ≠ FC in network organisation	Terroir clusters ≠ Identity clusters (ARI ≈ 0)
Structure constrains but does not determine function	Terroir shapes but does not determine identity
Functional states reflect history, task, and dynamics	Identity reflects history, culture, and human ambition

Across all four fields, the same intellectual structure recurs: two independent classification systems, one grounded in physical substrate, one in emergent cultural or functional organisation, applied to the same objects, yield different groupings. This study is the first to demonstrate this pattern in wine, with a formal independence test.

3. Data and Methodology

3.1 The 59 Regions

The study covers 59 wine regions: 39 Old World regions across 10 countries (Austria, Croatia, France, Germany, Greece, Hungary, Italy, Portugal, Slovenia, Spain) and 20 New World regions across 6 countries (Argentina, Australia, Chile, New Zealand, South Africa, USA). Regions were selected to provide broad geographic and stylistic coverage while maintaining sufficient identity distinctiveness for clustering to be meaningful. Each region carries a canonical soul metaphor — a single word or phrase distilling its essential identity character, assigned through anthropological analysis prior to and independent of the clustering.

3.2 Layer 1: Identity Classification (D-Score System)

Each region was described in a standardised 280–320 word identity narrative (Layer 1), written to capture cultural, historical, and anthropological character while remaining blind to any geographic or terroir information. From these narratives, six identity dimensions (D1–D6) were scored by the primary researcher acting as subject matter expert (SME), on an integer scale from –2 to +2:

Dimension	+2 Pole	−2 Pole
D1 — Interiority ↔ Exteriority	Interior, place-defined	Outward-facing, market-oriented
D2 — Struggle ↔ Ease	Extreme difficulty, endurance	Pleasure, comfort, ease
D3 — Tradition ↔ Reinvention	Deep custodial tradition	Radical reinvention, disruption
D4 — Individual ↔ Collective	Strongly individual, solitary	Strongly collective, communal
D5 — Urgency ↔ Timelessness	Urgent, present-focused	Timeless, eternal
D6 — Earthly ↔ Transcendent	Deeply grounded, earthly	Transcendent, spiritual

The six D-scores were Z-score standardised before clustering. PCA was applied to reduce dimensionality (capturing 100% of variance from the six input dimensions). K-means clustering (k=6, n_init=20, random_state=42) was applied to the standardised D-score matrix, yielding the identity clustering solution.

3.3 Layer 2: Terroir Classification

Each region was also described in a standardised six-field factual profile (Layer 2) covering: climate type, predominant soil composition, topographic character, principal grape varieties, winemaking techniques, and historical position. These fields were TF-IDF vectorised independently of the Layer 1 descriptions, using a separate vectoriser (ngram range 1–2, max 800 features, sublinear TF scaling). PCA reduced the feature space to 10 components (capturing 32.2% of variance). K-means clustering with k=7 was then applied to produce the terroir clustering solution. The terroir clustering is fully model-derived — cluster count, membership, and naming are determined entirely by the algorithm with no manual override.

3.4 Independence Test

To test whether the two clustering solutions agree, we compute the Adjusted Rand Index (ARI; Hubert & Arabie, 1985). ARI measures the fraction of region pairs whose cluster co-membership agrees across the two solutions, corrected for chance agreement. ARI = 1.0 indicates perfect agreement; ARI = 0.0 indicates chance-level agreement; ARI < 0 indicates less agreement than expected by chance. We supplement ARI with a chi-square test of independence on the 6×7 cluster membership contingency table.

4. Results

4.1 Model Performance

The identity clustering (D-scores, k=6) achieves a silhouette score of 0.3029, indicating meaningful cluster structure. The terroir clustering (Layer 2 TF-IDF, k=7, fully model-derived) achieves a silhouette score of 0.2457. The near-zero ARI between the two solutions (ARI = 0.023), combined with a chi-square p-value of 0.131, confirms that knowing a region’s terroir cluster provides essentially no information about its identity cluster. The two systems are orthogonal.

4.2 The Six Identity Clusters

The D-score model yields six identity clusters. Each is named for its dominant character and annotated with representative regions and their soul metaphors:

Identity Cluster	n	Character	Representative Regions & Metaphors
Old World Interior	10	Inward, terroir-defined, historically dense	Burgundy (Devotion), Northern Rhône (Solitude), Piedmont (Philosophy), Mosel (Poetry)
Outward Ease	5	Approachable, extroverted, pleasure-forward	Beaujolais (Joy), Provence (Pleasure), Veneto (Commerce), Marlborough (Assertion)
New World Reinvention	12	Intentional identity-building, constructed character	Napa Valley (Ambition), Swartland (Rebellion), Sicily (Resurrection), Stellenbosch (Aspiration)
Old World Exterior	7	Old World provenance, outward-facing identity	Bordeaux (Business), Champagne (Society), Rioja (Patience), Châteauneuf-du-Pape (Family)
Against the Odds	10	Marginality and struggle as central character	Santorini (Survival), Jura (Eccentricity), Hunter Valley (Defiance), Barossa Valley (Fortitude)
The Moderates	15	Balanced, pluralistic, no single force dominates	Tuscany (Art), Margaret River (Composure), Willamette Valley (Idealism), Alsace (Duality)

4.3 The Seven Terroir Clusters

The terroir model, fully model-derived with $k=7$, yields seven geographically coherent clusters reflecting climate type, soil composition, and topographic character:

Terroir Cluster	n	Top TF-IDF Terms	Members
Iberian Continental	6	spain, rioja, do	Douro, Catalonia, Galicia, Ribera del Duero, Rioja, Mendoza
Southern Hemisphere & International	12	zealand, new zealand, australia	Alto Adige, Patagonia, Barossa Valley, Hunter Valley, Margaret River, Maipo Valley, Central Otago, Hawke's Bay, Marlborough, Stellenbosch, Swartland, Napa Valley
American West Coast	7	ava, wineries, california	Paso Robles, Santa Barbara, Santa Cruz Mountains, Sonoma, Walla Walla, Willamette Valley, Columbia Valley
French Viticultural	10	france, côte, de	Alsace, Beaujolais, Bordeaux, Burgundy, Champagne, Châteauneuf-du-Pape, Jura, Loire, Northern Rhône, Provence
Germanic Rhine	5	germany, vdp, mosel	Baden, Mosel, Nahe, Pfalz, Rheingau
Austrian Danube	3	dac, veltliner, austria	Kamptal, Wachau, Wagram
Mediterranean & Volcanic	16	italy, di, indigenous	Steiermark, Dalmatian Coast, Macedonia, Santorini, Tokaj, Campania, Etna, Friuli-Venezia Giulia, Liguria, Piedmont, Sardinia, Sicily, Tuscany, Veneto, Goriška Brda, Finger Lakes

4.4 Key Region Movements: Where the Map Diverges from the Soul

The most revealing finding is not the aggregate statistics but the individual region trajectories — cases where a region’s terroir cluster membership would lead us to predict one identity, but the actual identity cluster tells a different story:

Region	Terroir Cluster	Identity Cluster	The Divergence
Santorini	Mediterranean & Volcanic	Against the Odds	Shares terroir type with coastal Mediterranean regions, but volcanic isolation and Assyrτικο’s survival character drive an identity of extremity rather than ease
Sicily	Mediterranean & Volcanic	New World Reinvention	Shares Mediterranean terroir with Tuscany and Sardinia, but its identity is one of radical reinvention — an Old World region behaving like a New World pioneer
Santa Cruz Mtns	American West Coast	Old World Interior	A New World region in California whose obsessive, private, inward-facing identity places it among Burgundy and Piedmont — not among its geographic neighbours
Barossa Valley	Southern Hem. & Intl.	Against the Odds	International terroir conditions suggest commercial ease, but the Barossa’s identity is forged through multi-generational endurance against drought and vine-pull pressures
Catalonia	Iberian Continental	New World Reinvention	Grouped with Iberian neighbours by terroir vocabulary, but its political and cultural assertion of distinctiveness gives it the urgency of a New World identity
Finger Lakes	Mediterranean & Volcanic	New World Reinvention	Terroir vocabulary groups it with Mediterranean regions, but its identity is built on deliberate construction and the urgency of proving itself
Dalmatian Coast	Mediterranean & Volcanic	Old World Exterior	Mediterranean terroir typically correlates with interior or struggle-based identities nearby, but Dalmatia’s deep calm and outward ease place it with Bordeaux and Champagne
Walla Walla	American West Coast	Outward Ease	Shares terroir with Columbia Valley and Willamette Valley, but its communal, outward-facing identity has nothing in common with the determination of its terroir peers

5. Implications

5.1 For Terroir Theory

This study does not diminish terroir. D1 — Interiority/Exteriority — varies enormously across the 59 regions, confirming that some regions are genuinely more terroir-legible than others. What it challenges is the conflation of terroir with identity. Terroir scholars from Demossier (2001) to Trubek (2008) have argued that terroir is partly a social construction — a story producers and institutions tell about the relationship between land and wine. This study provides quantitative support for that argument: the terroir narrative that a region adopts is statistically independent of the physical terroir it actually possesses.

5.2 For Wine Region Identity Studies

The six identity clusters offer a new vocabulary for describing wine regions that transcends geography. A region is not merely Mediterranean or Germanic or Antipodean; it is, in the language of this study, inward or outward, struggling or at ease, constructed or inherited. These identity dimensions — D1 through D6 — provide a multidimensional scorecard that can be applied to any region, including new appellations seeking to understand and communicate their character.

Three key findings from the D-score analysis are worth highlighting. First, the Struggle ↔ Ease axis (D2) is the dimension that most strongly differentiates Old World from New World regions, more so than Interiority (D1) — confirming that lived difficulty is the primary axis of wine regional identity. Second, the Individual ↔ Collective axis (D4) is the primary driver of New World identity cluster membership, consistent with the observation that where terroir's constraints are weakest, human agency fills the void. Third, Urgency ↔ Timelessness (D5) and Earthly ↔ Transcendent (D6) are moderately correlated, suggesting that regions with a strong sense of temporal permanence also tend toward the grounded rather than the aspirational.

5.3 For Academic Research Across Fields

This study joins a family of findings across disciplines in which physical substrate and emergent identity are demonstrably separable. The Ronen-Shenkar tradition established this for national cultures; biogeography for species communities; historical linguistics for language families; neuroscience for brain networks. The underlying principle is consistent: complex systems develop identities that are shaped by but not determined by their physical conditions. History, isolation, ambition, and cultural transmission are independent forces that leave statistically detectable signatures.

The methodological contribution of this study to that broader literature is the use of expert-scored identity dimensions (the D-score system) as a substitute for survey instruments. Rather than asking human subjects about their values and attitudes — the approach of Hofstede, GLOBE, and Ronen-Shenkar — we derive identity dimensions from expert-curated textual descriptions. This approach may be applicable to other domains where survey collection is impractical but textual documentation is rich.

6. For the Wine Reader

A plain-language guide to the key findings

If you have skipped here from the beginning, here is what this study found, in plain language.

We took 59 wine regions from across the world and classified them in two completely different ways. First, we grouped them by what the land gives them: climate, soil, topography — the physical terroir. Second, we grouped them by who they are: their historical depth, the ambition of their winemakers, the strength of their cultural story, how assertively they express their own character. Then we asked: do the two groupings agree?

Knowing where a region sits on the map tells you almost nothing about where it sits in the landscape of identity.

They do not. Knowing where a region sits on the map — what climate it has, what soils lie beneath its vines — tells you almost nothing about where it sits in the landscape of identity. A region battered by volcanic heat and wind may develop an identity of survival, or one of ancient philosophical contemplation, or one of extroverted joy. The terroir creates the conditions; the identity is chosen, inherited, fought for, and built over generations.

The Six Identity Types

Old World Interior · Devotion, Solitude, Philosophy, Monumentality

These regions look inward. They are defined by the weight of their own history and the specificity of their terroir's expression. They do not need to explain themselves. Burgundy, the Northern Rhône, Piedmont, Wachau — the wines of these regions communicate through precision, tradition, and a kind of magnificent gravity. They are the hardest wines in the world to understand, and the most rewarding when you do.

Outward Ease · Joy, Sentimentality, Duality, Assertion

These regions reach outward. Their wines communicate pleasure before complexity, invitation before challenge. Beaujolais does not ask you to study it; it asks you to drink it. Loire invites you into something personal. Alsace navigates between two cultural worlds with extraordinary grace. These are not simple regions — they are regions that have chosen to meet you more than halfway.

New World Reinvention · Ambition, Rebellion, Idealism, Resurrection

These regions build their identity deliberately. Where Old World regions inherit, New World Reinvention regions construct. Napa Valley assembled the vocabulary of greatness from borrowed parts and made it distinctly its own. Swartland rejected the established order entirely. Willamette Valley committed to Pinot Noir before the evidence warranted it. These are regions driven by human ambition — and it shows in the wines.

Old World Exterior · Business, Society, Patience, Family

These regions carry Old World credentials with outward-facing confidence. Bordeaux and Champagne are the supreme examples: ancient in origin, global in orientation, commercially articulate in a way that Burgundy, for all its greatness, is not. These regions have translated their historical depth into institutional power. Their wines are among the world's most recognisable precisely because the regions have worked hard to make them so. Of the six clusters, Old World Exterior is the most internally diverse. Bordeaux and Champagne define the extreme end —

maximally outward, maximally collective, maximally traditional — while Baden, Pfalz, and Rioja express the same institutional character more moderately. Dalmatian Coast is the cluster's most distinctive member: its strongly interior orientation runs counter to the group's outward tendency, but its deep timelessness and tradition place it here rather than elsewhere. The cluster holds together not because its members are identical, but because they share a mode of engaging with the world — institutional, tradition-anchored, outward-facing — even as they arrive at that mode through very different paths.

Against the Odds · Survival, Poetry, Awakening, Eccentricity, Defiance

These regions are defined by difficulty. Santorini's Assyrtiko grows in volcanic baskets on an island where Phylloxera could not travel and the wind never stops. Mosel's Riesling clings to near-vertical slate slopes where mechanisation is impossible. Etna's winemakers farm around the rim of an active volcano. The Hunter Valley produces Semillon in brutal heat and humidity and ages it into something miraculous. Struggle is not incidental to these regions — it is the source of their character.

The Moderates · Clarity, Dialogue, Composure, Serendipity

These regions are balanced and pluralistic. No single force — terroir, history, ambition, struggle — dominates. This is not a criticism. Steiermark's clarity, Campania's memory, Margaret River's composure — these are regions that have found equilibrium rather than extremity. Their wines reflect that balance: consistently good, rarely transcendent, always honest.

The deepest implication for the wine world is this: when a new region seeks to establish its identity, it should not look primarily at its terroir for answers. Terroir tells it what it can grow. History, ambition, cultural narrative, and the forcefulness with which it asserts its own character — these determine what it becomes. The map is not the soul. The soul has to be built.

References

- Bowen, S. (2015). *Divided Spirits: Tequila, Mezcal, and the Politics of Production*. University of California Press.
- Crofts, J.J. et al. (2022). Structure-function clustering in weighted brain networks. *Scientific Reports*.
- Demossier, M. (2001). The quest for identities: Consumption of wine in France. *Anthropology of Food*, S1.
- Demossier, M. (2018). *Burgundy: The Global Story of Terroir*. Berghahn Books.
- EPJ Data Science (2024). Exploring language relations through syntactic distances and geographic proximity.
- House, R.J. et al. (2004). *Culture, Leadership, and Organizations: The GLOBE Study of 62 Societies*. Sage.
- Hubert, L. & Arabie, P. (1985). Comparing partitions. *Journal of Classification*, 2(1), 193–218.
- Longobardi, G. et al. (2015). Across language families: Genome diversity mirrors linguistic variation within Europe. *American Journal of Physical Anthropology*.
- Ronen, S. & Shenkar, O. (2013). Mapping world cultures: Cluster formation, sources and implications. *Journal of International Business Studies*, 44(9), 867–897.
- Trubek, A. (2008). *The Taste of Place: A Cultural Journey into Terroir*. University of California Press.