

Human and Natural Drivers of Wildfires in the Basque Country: Causes and Motivations (1995–2022)

Santiago Gaztelumendi ¹

1 - Basque Meteorology Agency (EUSKALMET), Vitoria-Gasteiz, Basque Country, Spain

Abstract

Understanding the drivers of wildfire ignition is essential for improving predictive models and early-warning systems, particularly in regions with high human activity such as the Basque Country. While natural factors like lightning, wind, temperatures, and drought contribute to fire risk, anthropogenic causes dominate ignition, shaping both the frequency and the area burned. Detailed knowledge of human-driven ignition causes, motivations, spatial, and seasonal patterns is critical to link historical wildfire behavior with environmental and meteorological conditions, and to improve strategies for integral wildfire risk management. This study examines a long-term dataset of 4,054 wildfire incidents, providing a detailed characterization of ignition causes, motivations, seasonal variability, and territorial patterns. These findings represent a first step to transition toward integrated wildfire risk systems that incorporate anthropogenic ignition precursors and fuel status into traditional meteorological wildfire risk frameworks..

2. Data and Methodology

2.1. Data

The study utilizes the official forest fire database of the Basque Country (EGIF matrix, 1995–2022), comprising 4,054 verified wildfire incidents. Extracted variables include Temporal, Geographic/Spatial, and Causal/Motivational attributes, alongside burned area characteristics. Data cleaning organized administrative boundaries into 3 Historical Territories (HT) and 20 operational districts (Fig. 1). The original 82 sub-cause/motivation codes (Table 1) were condensed into 19 consolidated intermediate categories (Table 2).

3. Results and Discussion

3.1 Data preparation

The original dataset classifies all ignitions into five distinct series (see Tab 1) based on their cause and motivation: the 100 Series for lightning strikes, 200 for accidental human escapes, 300 for infrastructure failures, 400 for intentional arson, and 500 for unknown cases. To eliminate the statistical noise of marginal frequencies and establish robust blocks for

Table 1. Original natural and human drivers information

ID	Description
100	Lightning (Direct Natural Cause)
210	Other traditional agricultural burns
211	Agricultural stubble burns
212	Agricultural crop residue / pruning waste burns
213	Vineyard waste / shoot burns
214	Horticultural waste / small plot burns
215	Rural estate and boundary land clearing
220	General pasture management / livestock burns
221	Mountain brushwood and shrubland burns
222	Upland grass and herbaceous burns
230	Other operational forestry waste burns
240	Fireworks (discarded and pyrotechnics)
241	General unmanaged bonfires / recreational fires
242	Recreational bonfires in established campsites
243	Fireworks and pyrotechnics
244	Power lines under construction
245	Other mountain sports and climbing activities
246	Smokers (discarded cigarette butts)
250	General household refuse and trash burns
260	Uncontrolled garbage dumping sites burns
261	Wildland-urban interface trash containers
270	Municipal landfill escapes
280	General interface property maintenance
281	Roadside ditch and embankment land clearing
282	Power transmission line right-of-way clearing
283	Railway corridor structural maintenance
284	Public highway infrastructure repairs
286	Other urban fringe development works
290	Miscellaneous accidental human causes
291	Children playing with fire
292	Hunters and anglers outdoor accidents
293	Backyard household activities
295	Other unclassified non-intentional sources
310	Railway brake friction and line mechanical failures
312	Caterpillar arcing and cargo train failures
320	Electrical distribution lines direct failures
322	Tree branches touching power lines under high wind
323	Defective grid insulators and wire snapping
324	Wildlife electrocution on grid infrastructure
330	General internal combustion engine accidents
332	Commercial vehicle mechanical breakdowns
335	Road traffic accidents and collisions
336	Roadside vehicle accidental fires
331	Agricultural grain harvesting machinery
333	Industrial timber harvesters
334	Mechanical crop balers and agricultural packagers
335	Tractor-mounted mechanical heavy mowers
340	Public military training activities
341	Live-fire exercises in military shooting ranges
342	Handheld angle grinders and metal saw sparks
362	Structural welding in rural properties
363	Portable handheld brush cutters and trimmers
370	Active forest defense and suppression backfires
371	Prescribed research burns / official technical fires
372	Arson: Deliberate agricultural stubble burning
399	Other unclassified mechanical and engine sources
400	Arson: Undetermined final motivation
401	Arson: Deliberate agricultural stubble burning
402	Arson: Deliberate rangeland pasture regeneration
403	Arson: Wildlife control and hunting habitat modifications
405	Arson: Real estate boundary and land disputes
406	Arson: Commercial timber industry disputes
441	Arson: Attempt to alter local timber market prices
443	Arson: Retaliation against forest community consocios
451	Arson: Attempt to create civil unrest and public alarm
452	Arson: Animus against state reforestation programs
453	Arson: Protests against Protected Natural Areas
464	Arson: Personal vendettas and neighbor grievances
482	Arson: Opportunistic vandalism in picnic and park areas
483	Arson: Diagnosed mental illness / pyromania
484	Arson: Occult rituals and occult activities
499	Other unclassified anti-social behaviors
500	Total Undetermined Cause (No physical or human evidence)

4. Conclusions

- ✓ **Anthropogenic Dominance:** Human activity is the driver of wildfire regimes in the Basque Country, accounting for over 80% of total ignitions and 83% of total burned area, while natural lightning strikes remain at lees than 2%.
- ✓ **Negligence vs. Arson Severity:** Negligent ignitions (uncontrolled agricultural/yard burns) represent the most prevalent source, dominating small-to-medium events, whereas deliberate arson episodes disproportionately dictate large-scale, high-severity fires.

2.2. Exploratory and Descriptive Data Analysis

Absolute and relative proportions of ignitions and total burned area were computed and cross-analyzed for each HT across the 19 consolidated categories (Fig. 2). Absolute ignition frequencies were cross-tabulated across four independent temporal axes: annual trends, monthly seasonal distributions, weekly labor cycles, and 24-hour diurnal cycles. Every temporal panel was stacked and colored by HT to account for geographic asymmetry, linking human activity schedules and regional reporting variations with localized landscape configurations (Fig. 3, 4, 5, and 6)

1.Introduction

Understanding the complex interactions between human and natural drivers of wildfire ignitions is essential for improving early-warning systems and ignition forecasting models, particularly in regions characterized by intense anthropization and complex bioclimatic gradients. The Basque Autonomous Community (BAC) provides a critical and highly representative case study to address this challenge due to its acute socio-ecological and environmental heterogeneity, straddling the transitional boundary between the humid Atlantic Cantabrian front and the drier, warmer Mediterranean Ebro basin (Fig. 1). This bioclimatic pivot is physically driven by a rugged

2.3. Characterization of ignition drivers and predictive modeling

To decode the behavioral mechanisms governing fire origins in each district, a supervised classification model was trained using a Random Forest algorithm based exclusively on spatial and temporal features. The dataset was partitioned into an 80% training set for model construction and a 20% independent testing set for validation. Predictor variables included raw temporal indices and engineered features: Hour, Month, Weekday, HT_Label, District_Label, Hour_Period (Morning, Afternoon, Evening, Night), Season_Label (Winter, Spring, Summer, Autumn), and a socio-labor binary indicator (Is_Weekend; YES/NO). The forest was built with 500 decision trees; the variables randomly sampled at each split (mtry) were optimized programmatically to maximize information gain and minimize out-of-bag error. Variable predictive power was evaluated via the Mean Decrease in Gini Index.

topography, where mountain ranges segregate the landscape into distinct fire environments defined by strong spatial and phenological gradients in weather, land-use, and vegetation fuel connectivity, within which different human practices condition ignitions (see Fig. 1), in a territory where less than 2% of wildfires are due to natural causes.

Previous research in the region has established a solid foundation on these environmental drivers, focusing on the characterization of critical fire-conductive synoptic patterns (Gaztelumendi et al. 2019), comparative macro-regional assessments within the Cantabrian arc (Gaztelumendi et al. 2025a), spatiotemporal pattern assessments (Gaztelumendi 2026b), and the evaluation of

weather drivers (Gaztelumendi et al. 2026c, 2026d).

To transcend purely meteorological indicators, current operational frameworks (Gaztelumendi et al. 2020) require a migration toward comprehensive models that integrate human behavior as a central component of wildfire risk. This study addresses this need by analyzing the motivational and causal patterns of ignitions across the region's operational districts (Fig. 1). Through a combination of traditional data analysis and artificial intelligence/machine learning techniques (Gaztelumendi et al. 2025c), this research aims to optimize the Basque operational wildfire risk framework (Gaztelumendi et al. 2020, 2025b; GV 2025).

2.4. Actionable Knowledge Extraction via Decision Tree Paths

To translate multi-variable interactions into operational intelligence, the final RF model (Global Accuracy = 28.47%) was directly interrogated. Programmatic extraction of highly coherent terminal leaf-nodes through native conditional paths mapped out exactly 15 "Critical Ignition Scenarios" where spatial and temporal dimensions converge. Each scenario was filtered through quality control thresholds: a minimum empirical volume (Vol_N ≥ 10) independent fires per cell) and a minimum conditional predictive probability (Confidence ≥ 35%), discarding mathematical noise to provide direct deterministic rules.

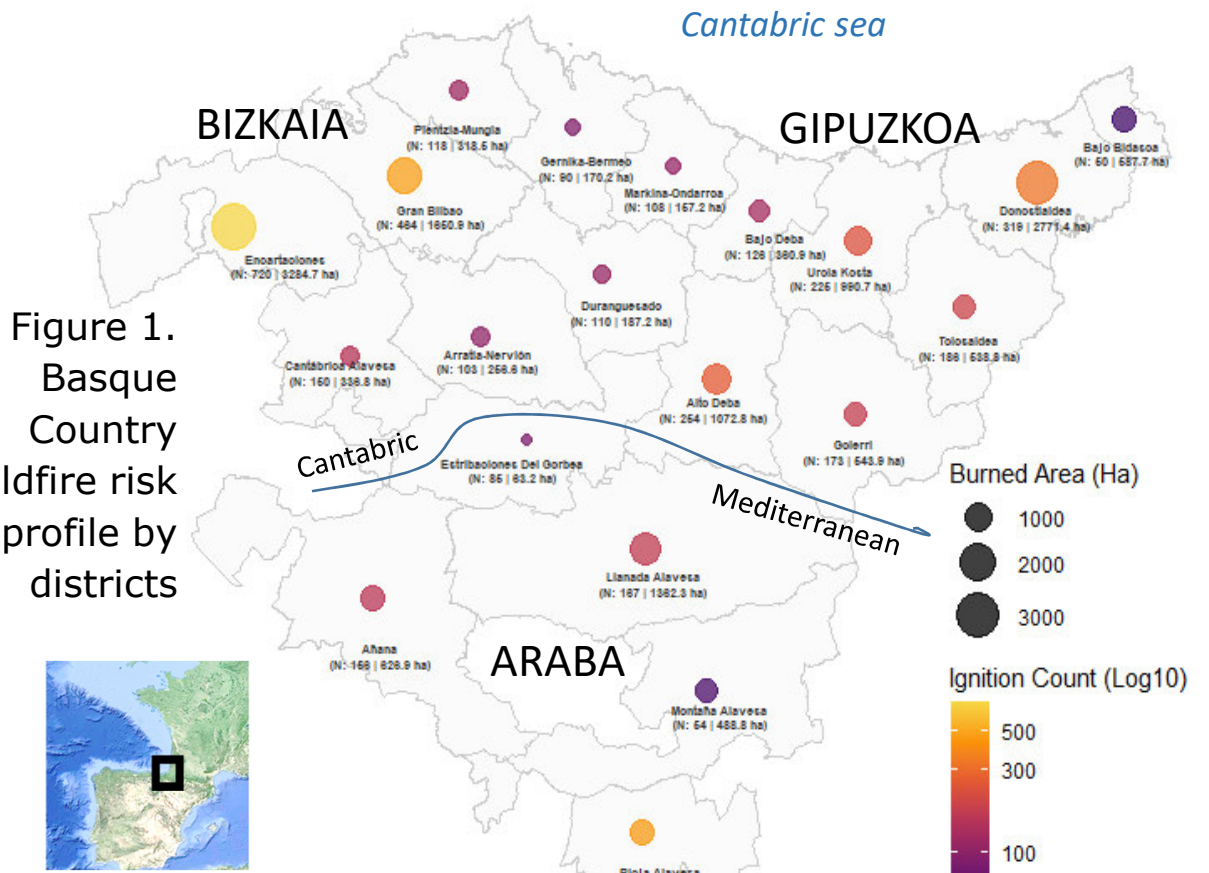


Figure 1. Basque Country wildfire risk profile by districts

2.5. Tools

Data curation, calculations, and spatial representations utilized R (R Core Team, 2026), leveraging the tidyverse ecosystem including dplyr (Wickham et al., 2023a), tidyr (Wickham et al., 2023b), ggplot2 (Wickham, 2016), purrr (Wickham & Henry, 2023), and stringr (Wickham, 2019) for data manipulation and visualization. Predictive modeling was implemented via the randomForest package (Breiman et al., 2024), while vector data processing and geospatial workflows were executed using the sf package (Pebesma et al., 2026).

3.2. Exploratory and Descriptive Analysis

Different aggregations, counts, and statistics are calculated for the 19 consolidated Causal-Motivational Fire Categories to characterize the dataset. The exploratory and descriptive analysis of the dataset is structured into two main components: first, an evaluation of territorial distributions and structural severity (Figure 2), followed by a multi-axis assessment of temporal and socio-labor cycles (Figures 3, 4, 5, and 6)

3.2.1. Exploratory and Descriptive Analysis

Ignition and Severity Asymmetry

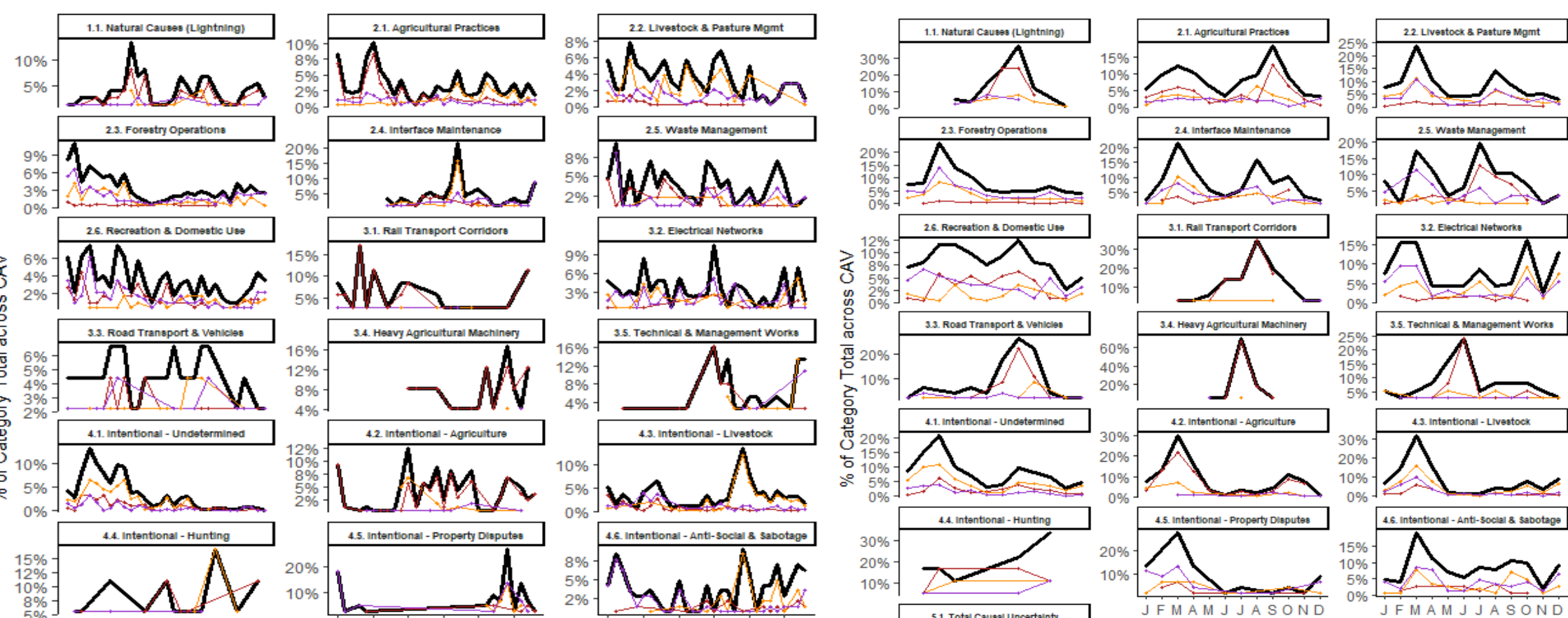


Figure 3. Empirical interannual trends f BAC and HT, percentage relative to ea category's total in the BAC.

3.2.2. Multi-Axis Temporal Patterns and Socio-Labor Cycles

Annual Trends (Fig 3)

Long-term interannual series show a general stabilization and reduction in both 4.1. Intentional - Undetermined Motivation and 5.1. Total Causal Uncertainty over the recent decades,

indicating an improvement in reporting and forensic investigation accuracy.

Traditional livestock and pasture management practices (2.2) display noticeable structural variations and reductions in frequency in the post-2012 period, probably coinciding with the implementation of burning authorization frameworks.

Monthly Patterns and Seasonality (Fig 4)

Atlantic Spring Regime: Incidents related to 2.2. Livestock & Pasture Management and 4.3. Intentional - Livestock Practices concentrate heavily during March, driven almost exclusively by the records in Bizkaia (BI) and Gipuzkoa (GI) during pre-vegetative clearance periods.

Mediterranean Summer Regime: Conversely, 1.1. Natural Causes (Lightning) and 3.4. Heavy Agricultural Machinery failures peak strictly within July and August, presenting an operational footprint restricted to the southern agricultural plains of Araba (AR).

Weekly Behavior (Fig 5)

Socio-Labor Synchronization: Forest

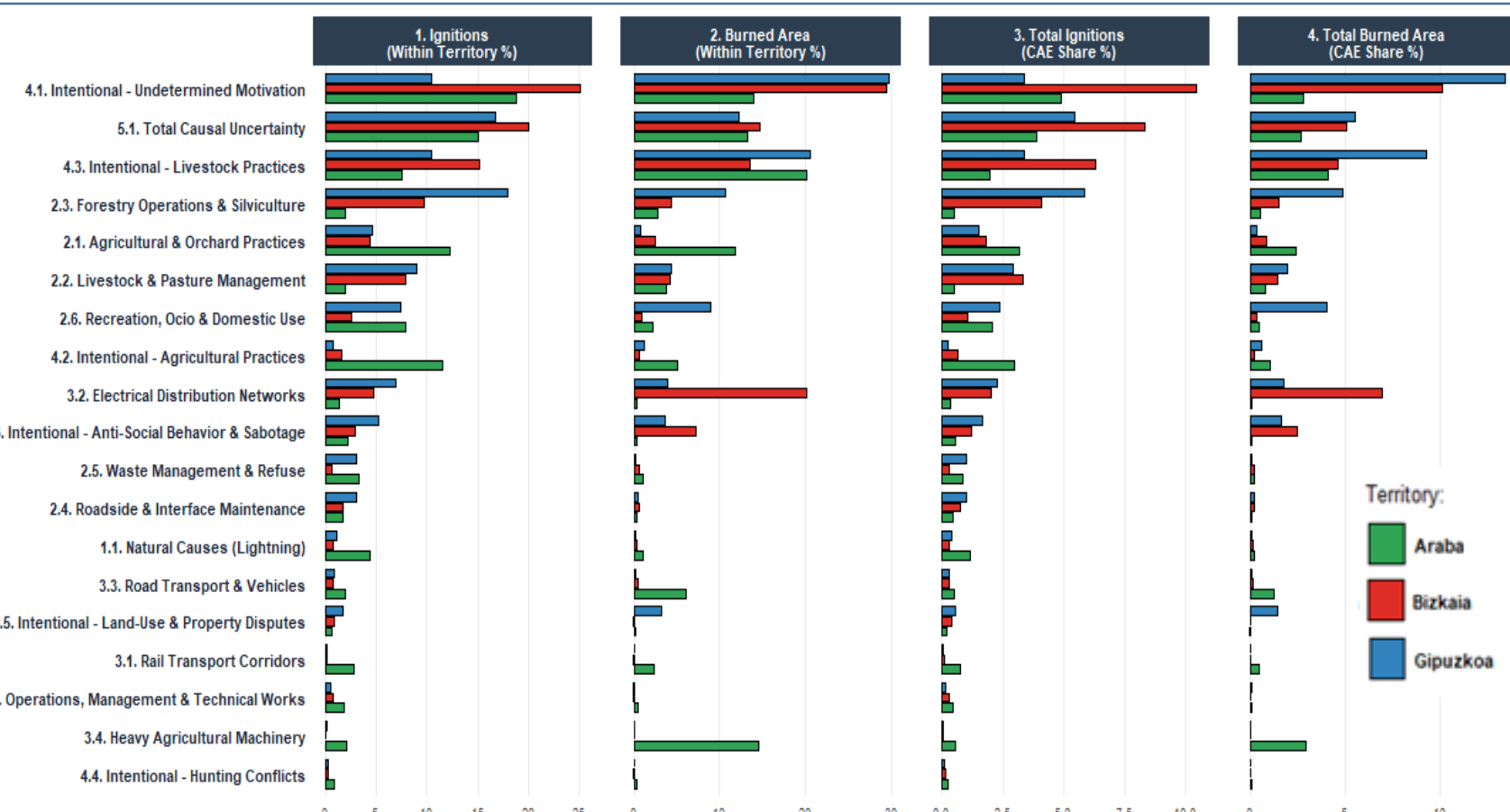


Figure 2. Multi-scale analytical matrix of causal motivational fire risk. scales.

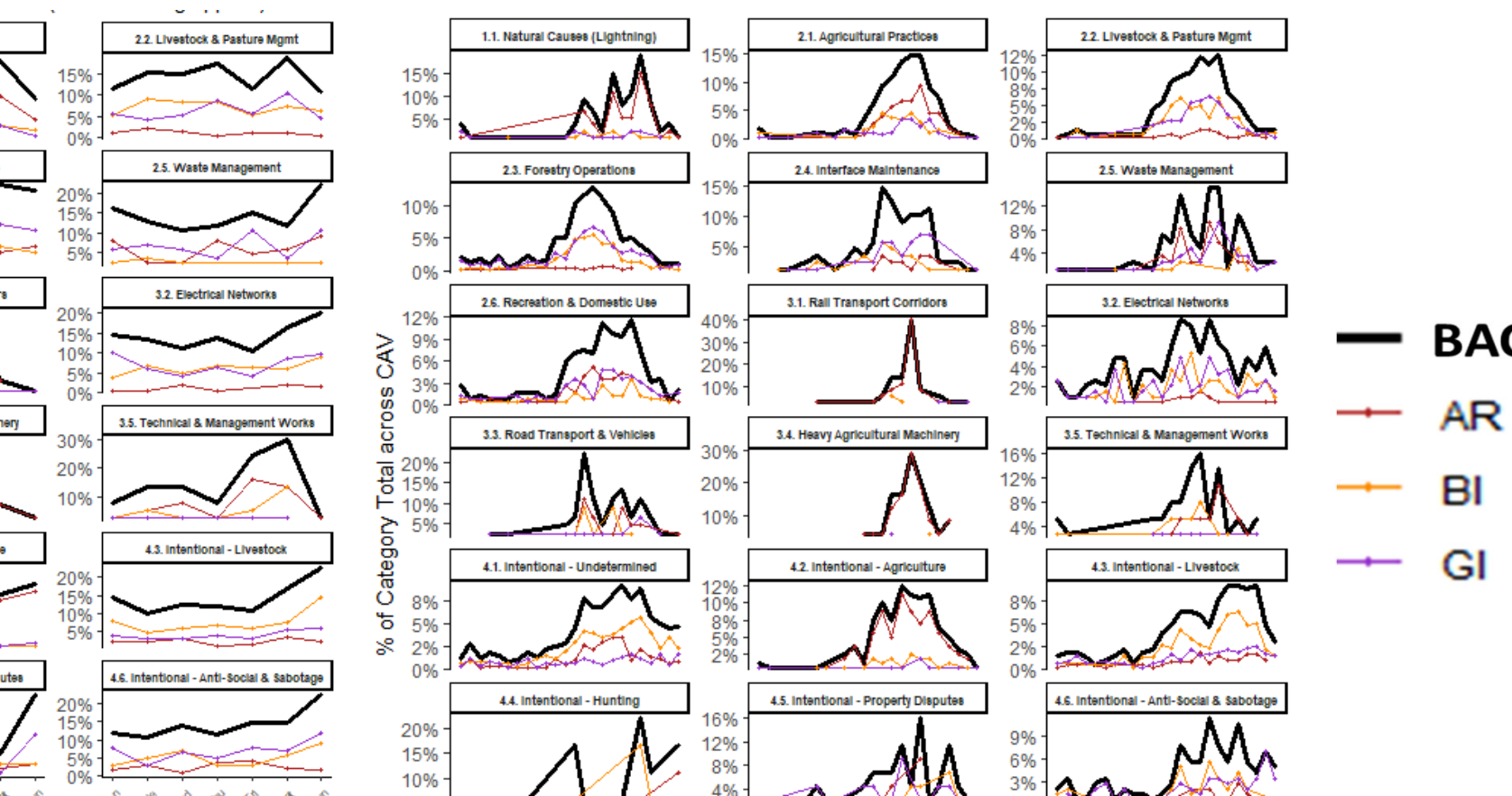


Figure 4. Empirical monthly seasonality category-relative distribution

Figure 5. Empirical weekly day category-relative distribution

Weekend and Recreational Increments: Activities linked to 2.6. Recreation & Domestic Use and deliberate rangeland practices (4.3) display higher relative proportions on Saturdays and Sundays.

Hourly Behavior (Fig 6)

Circadian Distribution: Intentional livestock categories (4.3) and specific vandalism scenarios (4.6) exhibit a marked concentration in the late afternoon and evening hours, specifically between 16:00 and 21:00, establishing a distinct nocturnal pattern compared to daytime operational accidents.

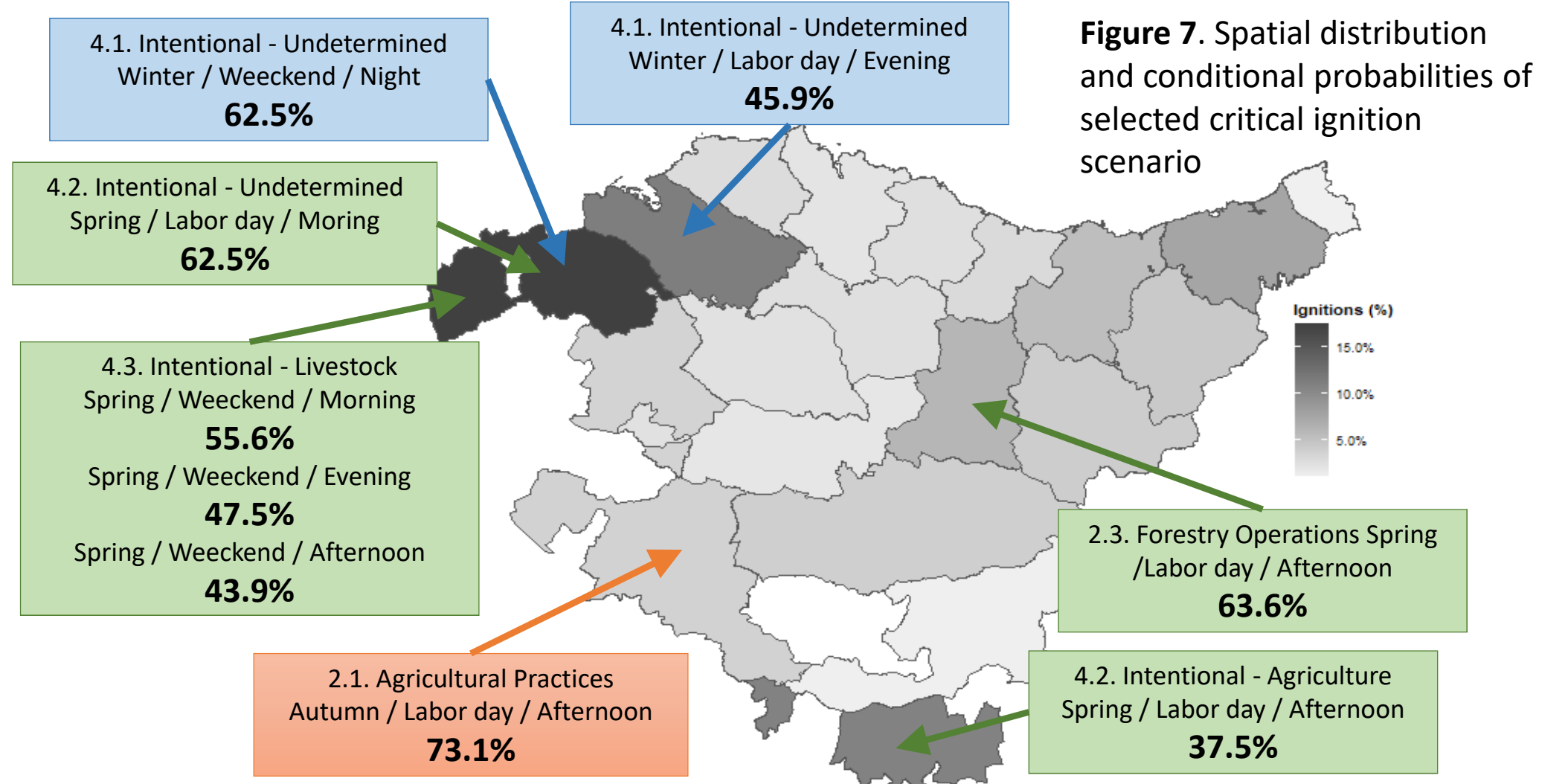


Figure 7. Spatial distribution and conditional probabilities of selected critical ignition scenario

3.3. Random Forest

3.3. 1. Feature Importance and Global Performance

Global Accuracy:

The optimized ensemble model achieved a 28.47% global accuracy over the independent test set.

Predictive Power:

This performance is five times higher than random chance (1/19 = 5.26%) for a 19-class classification problem.

Behavioral Signatures:

High predictive accuracy confirms that human-driven ignitions are not random but follow stable spatiotemporal patterns.

Gini Key Drivers:

District_Label emerged as the primary predictor in the Mean Decrease Gini ranking, followed by Hour and Month.

Macro-scale Stabilizers:

Engineered features (Is_Weekend and Hour_Period) successfully reduced variance in the decision nodes.

3.3.2. Critical Ignition Scenarios via Decision Tree Paths

Direct interrogation of the best model forest's conditional paths programmatically isolated 15 critical ignition scenarios (Vol_N ≥ 10; Confidence ≥ 35%), uncovering high-resolution regional patterns (see tab 3). An graphical representation of some of then are shown in Fig 7.

Table 3. Critical Ignition Scenarios

District Baseline	Season	Weekend?	Time Period	Target Category (Prediction)	Conf. (%)
AR, RIBERAS	Autumn	NO	Afternoon	2.3. Forestry Operations	73.1%
GI, GOIERRI	Spring	NO	Afternoon	2.3. Forestry Operations	63.6%
BI, ENCARTACIONES	Winter	NO	Morning	4.1. Intentional - Undetermined	62.5%
BI, ENCARTACIONES	Spring	YES	Night	4.1. Intentional - Undetermined	62.5%
BI, ENCARTACIONES	Spring	YES	Morning	4.3. Intentional - Livestock	55.6%
BI, ENCARTACIONES	Spring	YES	Evening	4.3. Intentional - Livestock	47.5%
BI, GRANA BLIBAO	Winter	NO	Evening	4.1. Intentional - Undetermined	45.9%
BI, ENCARTACIONES	Winter	NO	Evening	4.3. Intentional - Livestock	44.2%
BI, ENCARTACIONES	Spring	NO	Evening	4.1. Intentional - Undetermined	43.9%
GI, DONOSTIALDEA	Winter	NO	Evening	4.3. Intentional - Livestock	43.9%
BI, ENCARTACIONES	Spring	NO	Night	4.1. Intentional - Undetermined	41.4%
BI, ENCARTACIONES	Winter	NO	Afternoon	4.1. Intentional - Undetermined	39.0%
AR, RIOJA ALAVESA	Spring	NO	Afternoon	4.2. Intentional - Agriculture	37.5%
GI, DEBA GARAIA	Winter	NO	Afternoon	2.3. Forestry Operations	36.1%

Consequently, the operational district scale (District_Label) consolidates as a preferable methodological alternative for structuring predictive risk frameworks

- ✓ **Predictive Model Calibration:** The optimized Random Forest framework reached a 28.47% global accuracy over 19 distinct categories, multiplying random chance by five and proving that human-driven ignition profiles possess precise, quantifiable signatures.
- ✓ **Operational Leaf-Node Extraction:** Programmatic path analysis isolated 15 critical ignition scenarios Day-to-day agricultural clearing marks maximum

predictability in Riberas (73.1% Conf), and spring forestry operations dominate in Goierri (63.6% Conf).

- ✓ **Agricultural Sector (2.1 & 4.2):** Wildfire risk peaks in Araba/Álava during March, driven by stubble and boundary burning before spring sowing; this practice becomes highly hazardous when aligned with dry south wind episodes. A more pronounced, purely accidental risk window occurs in August due to machinery sparks during the cereal harvest.
- ✓ **Livestock Sector (2.2 & 4.3).** Concentrated in the Atlantic basin (Encartaciones), this sector shows a bimodal pattern driven by scrub clearance for

pasture regeneration. The main peak occurs from February to April (highest in March), and a secondary one in October-November. Both windows exploit dry, intense south winds to ensure high-velocity biomass combustion

- ✓ **Reporting Quality Trends:** Reduction in unknown causes over recent decades.
- ✓ **Future work:** Upgrading current weather-based operational networks into integrated wildfire risk platforms by incorporating fuel status and anthropogenic ignition precursor models.

Acknowledgements

This study has been partially funded by the European Union through the Interreg Sudoe Project USE4FOREST (Project Code: SO2.4/1.1/E0024). Additionally, this research was supported by the Directorate of Natural Heritage and Climate Change Adaptation (Vice-Ministry of Environment) and the Directorate of Emergency Attention and Meteorology (Vice-Ministry of Civil Protection) of the Basque Government.



The author expresses his gratitude to the global open-source and open-data communities for providing accessible, high-utility datasets and shared platforms of public utility. Special recognition is extended to the R

package maintainers, whose shared frameworks significantly advance data science and foster the democratization of code, enabling the reproducibility of complex computational pipelines and spatial workflows in environmental research.

Brennan, L., Cutler, A., Llew, A., & Wiener, M. (2023). randomForest: Breiman and Cutler's Random Forests for Classification and Regression. R package version 4.7-12. Available at: CRAN (randomforest). [1]

Gaztelumendi, S., Gomez de Segura, I., D., Ochoa de Alde, K., & E. E. (2025). A 21st century synoptic climatology of fire-weather conditions in Basque Country. EMS2025-524. https://doi.org/10.5194/emms2025-524

Gaztelumendi, S. (2026a). Spatial and Temporal Patterns of Wildfires in the Basque Country (1995-2022). EMS Annual Meeting 2026, Vitoria, Netherlands, 6-11 Sep 2026. EMS2026-586. https://doi.org/10.5194/emms2026-586

Gaztelumendi, S., Ochoa de Alde, K. (2026b). Empirical Analysis of Meteorological Drivers on Wildfire Dynamics in the Basque Country. Advances in Forest Fire Research - 2026. X. Viegas & L.M. Ribeiro (Eds.) ICFR 26, Coimbra, Portugal.

GV. Plan especial de emergencias por riesgo de incendios forestales en la comunidad autonoma vasca (2025). Servicio central de publicaciones de Gobierno Vasco. Administracion de la Comunidad Autonoma del Pais Vasco Departamento de Seguridad. ISBN: 978-84-403-8812-1.

Pebesma, E., Barend, R., Riecke, E., Sumner, M., Cook, I., Kallit, T., Lovelace, R., Wickham, H., Dorn, J., Miller, A., Pedersen, T.J., Baston, D., Dunnington, D., & Couriel, A. (2026). sf: Simple Features for R. R

Core Team (2026). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: R Project Website.

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. ISBN: 978-1-329-46027-4.

Wickham, H. (2019). stringr: Simple, Consistent Wrappers for Common String Operations. R package version 1.5.1. Available at: CRAN stringr.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, J., Hester, J., Kuhn, M., Pedersen, T. J., Miller, E., Riecke, S. M., Miller, K., Dorn, J., Robinson, D., Seidel, D. P., Sporn, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., & Yutani, H. (2023). Welcome to the tidyverse. Journal of Open Source Software. 4(3), 1-6.

Wickham, H., François, R., Bryan, J., & Barret, M. (2023a). dplyr: A Grammar of Data Manipulation. R package version 1.1.4. Available at: CRAN dplyr.

Wickham, H., Vaughan, D., & Gihm, M. (2023b). tidy: Tidy Messy Data. R package version 1.0.0. Available at: CRAN tidy.

Wickham, H., & Henry, L. (2023). purrr: Functional Programming Tools. R package version 1.0.2. Available at: CRAN purrr.