Uisge Beatha: Water of Life or Watered Down? Applied Multivariate Methods in Scotch Differentiation

Levi Quintero, Kirill Galiev, Jiaxi Sun, Zion Swinburne

Contents

1 Introduction

jess

Scotch whisky represents one of the United Kingdom's most valuable export commodities, contributing approximately £3.95 billion to the economy in 2015, or about 25% of total food and drink exports. Its global popularity, particularly for blended whiskies—which can consist of 60–80% grain whisky—makes the industry highly vulnerable to counterfeiting(Shand et al. (2017); Storrie (1962); Bower (2016); Scotch

Whisky Association Cereals Working Group (2021)). Counterfeit products not only undermine economic revenue and brand integrity but also pose risks to consumer safety .

Traditional whisky authentication methods rely on techniques such as gas chromatography, which profiles volatile organic congeners, and stable isotope ratio analysis, which can differentiate production origins. While these methods are scientifically robust, they require expensive instrumentation, laboratory-bound environments, and trained specialists. (Shand et al. (2017)) Consequently, they are not practical for rapid, field-based screening in supply chains, retail, or customs inspection.

An alternative approach lies in trace element analysis. Elemental signatures in whisky derive from raw materials (such as water and grain), production equipment, storage vessels, and potential additives. Previous research suggests that these elemental profiles can provide a reliable chemical "fingerprint" to detect fraudulent products and explore provenance (Power et al. (2020)).

Building on this idea, the present study applies Total Reflection X-Ray Fluorescence (TXRF) spectroscopy in combination with multivariate statistical analysis to evaluate whether elemental concentration data can reliably classify whisky samples. Specifically, we investigate whether TXRF measurements across 11 trace elements (P, S, Cl, K, Ca, Mn, Fe, Cu, Zn, Br, and Rb) can: - Differentiate authentic Scotch whiskies from counterfeit products, - Distinguish between blended/grain whiskies, and - Explore whether regional provenance (Highland, Island, Lowland, Speyside) leaves a measurable elemental signature.

The statistical workflow includes data preprocessing (log transformation and Mahalanobis distance assessment), dimensionality reduction by Principal Component Analysis (PCA), classification by Linear Discriminant Analysis (LDA), and clustering using Partitioning Around Medoids (PAM) and hierarchical methods.

Robustness of classification is further tested across distance measures (Euclidean, Manhattan and correlation).

1.1 Our Aims

Reassesing/Using data already collected is both cost-saving and effective within acadamia when resources are limited. We used Shand et al. (2017) XTRF collected trace chem sampels from 7 whisky types. We did so to assess reproducability as well further applications of multivariaite analytical methods to this method of chemical sampling. If successful and widely applicable, this presents a novel, robust and cost affordable way to differentiate counterfeits, grains and malt whiskies.

1.2 Brief description Sharon et al., study

Table 1: Whisky Origin and Chemical Data

Sample_no	Descriptor	Distillery	P	S	Cl	K	Ca	Mn	Fe	Cu	Zn	Br	Rb
1	Blend	Baile Nicol Jarvie	0.152	1.100	0.173	7.860	1.450	0.032	0.027	0.186	0.015	0.002	0.006
2a	Blend	Bells	0.653	1.580	0.238	4.930	1.400	0.019	0.110	0.242	0.021	0.005	0.003
3a	Blend	Chivas	0.375	0.809	0.193	4.310	1.220	0.019	0.044	0.196	0.007	0.003	0.002
4a	Blend	Dewars	0.121	1.160	0.157	3.200	1.140	0.011	0.050	0.189	0.018	0.003	0.003
5a	Blend	Johnnie Walker	0.326	1.090	0.180	5.480	0.526	0.018	0.103	0.286	0.020	0.002	0.002
6a	Blend	The Famous Grouse	0.145	0.615	0.097	2.740	0.416	0.009	0.050	0.208	0.007	0.002	0.001
7a	Blend	Whyte and Mackay	0.067	0.576	0.151	2.360	0.745	0.012	0.047	0.159	0.019	0.003	0.002
8a	Blend	William Grant	0.239	0.748	0.147	2.840	0.976	0.010	0.021	0.137	0.020	0.003	0.002
9a	Counterfeit	Unknown 1	0.089	4.060	0.066	0.336	1.240	0.007	0.154	0.085	0.038	0.005	0.001
10a	Counterfeit	Unknown 2	0.088	14.700	0.072	1.230	1.400	0.006	0.025	0.052	0.018	0.004	0.001
11a	Counterfeit	Unknown 3	0.279	15.900	0.083	0.811	1.360	0.006	0.057	0.038	0.016	0.002	0.002
12a	Counterfeit	Unknown 4	0.320	22.100	0.596	2.320	1.780	0.008	0.019	0.038	0.015	0.068	0.001
13a	Counterfeit	Unknown 5	0.120	26.100	0.071	2.370	1.630	0.010	0.082	0.187	0.194	0.012	0.005
14a	Grain	Grain matured	0.034	2.230	0.252	6.440	1.040	0.013	0.115	0.174	0.019	0.004	0.006
15a	Grain	Grain unmatured	0.084	5.530	0.113	3.250	1.350	0.012	0.076	0.164	0.046	0.010	0.003
16	Highland	Glengoyne	1.040	5.570	0.343	24.200	0.857	0.023	0.197	1.251	0.041	0.004	0.016
17	Highland	Glenmorangie	0.126	0.796	0.245	6.950	0.859	0.035	0.025	0.523	0.011	0.003	0.006
18a	Island	Bowmore	0.914	6.670	0.316	21.100	0.868	0.037	0.148	0.548	0.032	0.007	0.018
19	Island	Bruichladdie	1.630	5.480	0.697	36.500	4.130	0.038	0.288	0.587	0.066	0.034	0.039
20a	Island	Bunnahabhain	2.240	7.540	1.350	36.200	2.120	0.051	0.184	0.580	0.057	0.014	0.037
21	Island	Talisker	0.034	4.850	0.362	5.670	0.607	0.018	0.070	0.277	0.033	0.003	0.006
22a	Lowland	Auchentoshan	0.169	1.460	0.417	11.700	0.681	0.042	0.128	1.320	0.037	0.006	0.012
23a	Lowland	Glenkinchie	0.108	2.450	0.176	7.760	0.738	0.031	0.106	0.434	0.022	0.002	0.007
24	Speyside	Balvenie	0.695	3.850	0.120	20.300	0.765	0.031	0.121	0.380	0.035	0.005	0.024
25	Speyside	Craigellachie	0.096	0.819	0.177	6.110	0.633	0.024	0.094	0.239	0.025	0.005	0.006
26	Speyside	Dufftown	0.883	4.640	0.130	14.000	1.050	0.030	0.078	0.533	0.024	0.002	0.014
27	Speyside	Glen Elgin	0.115	1.350	0.404	9.270	1.400	0.031	0.046	0.195	0.029	0.006	0.009
28	Speyside	Glenburgie	2.000	7.910	0.185	37.700	1.650	0.053	0.134	0.198	0.043	0.008	0.026
29	Speyside	Glennfiddich	0.317	2.720	0.344	12.400	0.660	0.029	0.132	0.519	0.193	0.004	0.013
30	Speyside	Glenrothes	0.953	4.110	0.399	16.700	1.830	0.041	0.137	1.030	0.029	0.007	0.014
31	Speyside	Knockando	0.051	1.030	0.191	5.140	0.605	0.017	0.094	0.432	0.020	0.008	0.005
32	Speyside	Linkwood	0.276	1.050	0.207	6.220	1.010	0.020	0.064	0.769	0.019	0.004	0.006

Note: Chemical concentrations reported in mg per L. All samples analyzed using total reflection X-ray fluorescence (XTRF). Derived from Shand et al. 2017.

1.3 Our hypothesis

As we wish to use all available data to retain maximum information, we wish to employ multiple common cluster analysis methods to see if can produce useful and agreeing groups for whisky differentiation. As Shand et al. (2017) was able to differentiate counterfeits successfully from all other whiskies within a limited principal component (PC) space (PC1-PC3), we wished to attempt to replicate this using k-means, partitioning about mediods (PAM), and agglomerative hierarchical clustering to assess if XTRF is a robust method for counterfeit detection.

Further, we wish to see if the data presents us with any compelling groups other than our pre-applied ones, we will aim at assessing data structure and seeing if groups emerging from cluster analysis agree.

2 EDA

2.1 Exploratory Data Analysis (EDA)

All data analysis were conducted in R version 4.4.3 (R Core Team 2025). Summary statistics for whisky sample chemical trace TXRF data (Table 1) show large differences in range and variability between trace chemical variables magnitude and variability (Table 2). Variables show differences in the range of the magnitude 10^4, with Rb displaying a range of 0.030 and K displaying a range of 36.964. When density plots of by-chemical observations were plotted, all displayed a strong right skew. Further, whisky class observations

(n=32) were highly unbalanced (Grain = 2, Highland = 2, Lowland = 2, Island = 4, Counterfeit = 5, Blend = 8, Speyside = 9) across the observations of these 11 chemical variables (P, S, Cl, K, Ca, Mn, Fe, Cu, Zn, Br, Rb).

Table 2: Summary Statistics for Whisky Chemical Variables

Variable	Mean	Median	SD	Min	Max
P	0.461	0.204	0.575	0.034	2.240
S	5.019	2.585	6.261	0.576	26.100
Cl	0.270	0.188	0.247	0.066	1.350
K	10.262	6.165	10.569	0.336	37.700
Ca	1.192	1.045	0.685	0.416	4.130
Mn	0.023	0.020	0.013	0.006	0.053
Fe	0.095	0.088	0.060	0.019	0.288
Cu	0.380	0.240	0.327	0.038	1.320
Zn	0.037	0.023	0.043	0.007	0.194
Br	0.008	0.004	0.012	0.002	0.068
Rb	0.009	0.006	0.010	0.001	0.039

Note: Values represent mean, median, standard deviation (S.D.), and range (min–max) for each chemical element measured across all whisky samples.

We initially assessed if the TXRF data derived from Shand et al. (2017) (Table 1) followed a multivariate normal distribution ($X \sim N_{11}(\mu, \Sigma)$) via visual comparisons of observation Mahalanobis distances ($d_M^2(X, \mu)$) to their expected quantiles and a QQ plot line (1), observation density plots by factor, as well as by conducting a Henze-Zirkler Test of Multivariate normality ($HZ = 1.325, p < 0.001$; Table 3). As the data clearly did not conform to $X \sim N_{11}(\mu, \Sigma)$ and exhibited starkly different magnitudes in observations, we log transformed the observations and re-aplied the same analysis (Fig. 1; Table 3). Results now conformed to $X \sim N_{11}(\mu, \Sigma)$ with no outliers ($d_M > 0.99$), and displayed distributions generally centered about 0. Anderson-Darling tests of univariate normality confirmed that 9 of the 11 chemical variables displayed normality while Zn ($A^2 = 0.872, p = 0.022$) and Br ($A^2 = 1.075, p = 0.007$) displayed significant departures. This log-transformed data was henceforth used in this analysis.

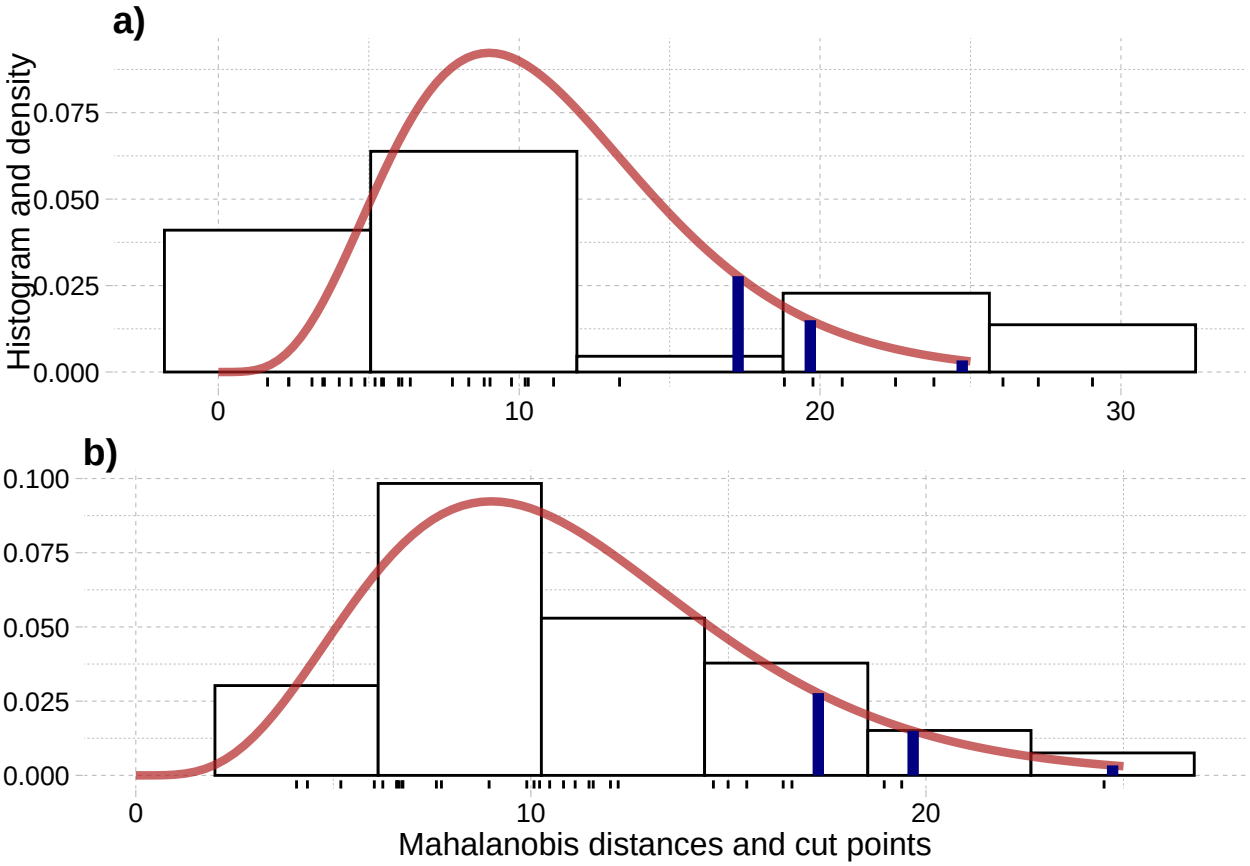

Figure 1: Mahalanobis histogram of multivariate density distribution of whisky observations with cutoff points marked at 0.90,0.95 and 0.99 density quantiles, overlaid with a chi-squared distribution kernel with 11 degrees of freedom.

Table 3: Summary of Surprising Observations by Data Transformation

Distance Category	Count/HZ	%/P-val
Original Data		
Bottom_50%	22.000	68.8
50-75%	2.000	6.2
75-90%	0.000	0
90-95%	1.000	3.1
95-99%	4.000	12.5
Top_1%	3.000	9.4
Henze-Zirkler Test	1.325	<0.001
Log-Transformed Data		
Bottom_50%	17.000	53.1
50-75%	7.000	21.9
75-90%	5.000	15.6
90-95%	2.000	6.2
95-99%	1.000	3.1
Top_1%	0.000	0
Henze-Zirkler Test	0.984	0.137

Note: Distance categories based on Mahalanobis distance quantiles. HZ denotes Henze-Zirkler test statistic for multivariate normality.

In Shand et al. (2017), an LDA analysis was conducted using the first 3 principal components. Due to highly imbalanced design of whisky class sampling, overall homogenity of whisky class covariance ($\Sigma_1 = \Sigma_2 = ... = \Sigma_7$) could not be established as the smallest classes (n = 2) are not $> p = 3$ and thus produced non-invertible covariance matrices. Thus, combined with the small class sizes likely making supervised LDA training unstable, we believe that this rules out the suitability of this method of discriminant analysis.

Due to Shand et al. (2017)'s inability to draw cohesive classification results at the regional level, we sought to allow the data as well as literature evidence of whisky type composition to intuitively guide our selection of the number of assigned clusters (k^*) in this analysis. We first viewed box-plots by variable class and the overall correlation structure of the chemical variables. Drawing upon these observations, differential median trends emerged (particularly for variables Mn, Cu and Rb) between single-origin and blends, grains and counterfeits (Fig. 2). We then isolated the chemical variable correlation structures of the largest classes available to visually assess and compared respective mean vector differences between these using Hotelling T^2 tests.

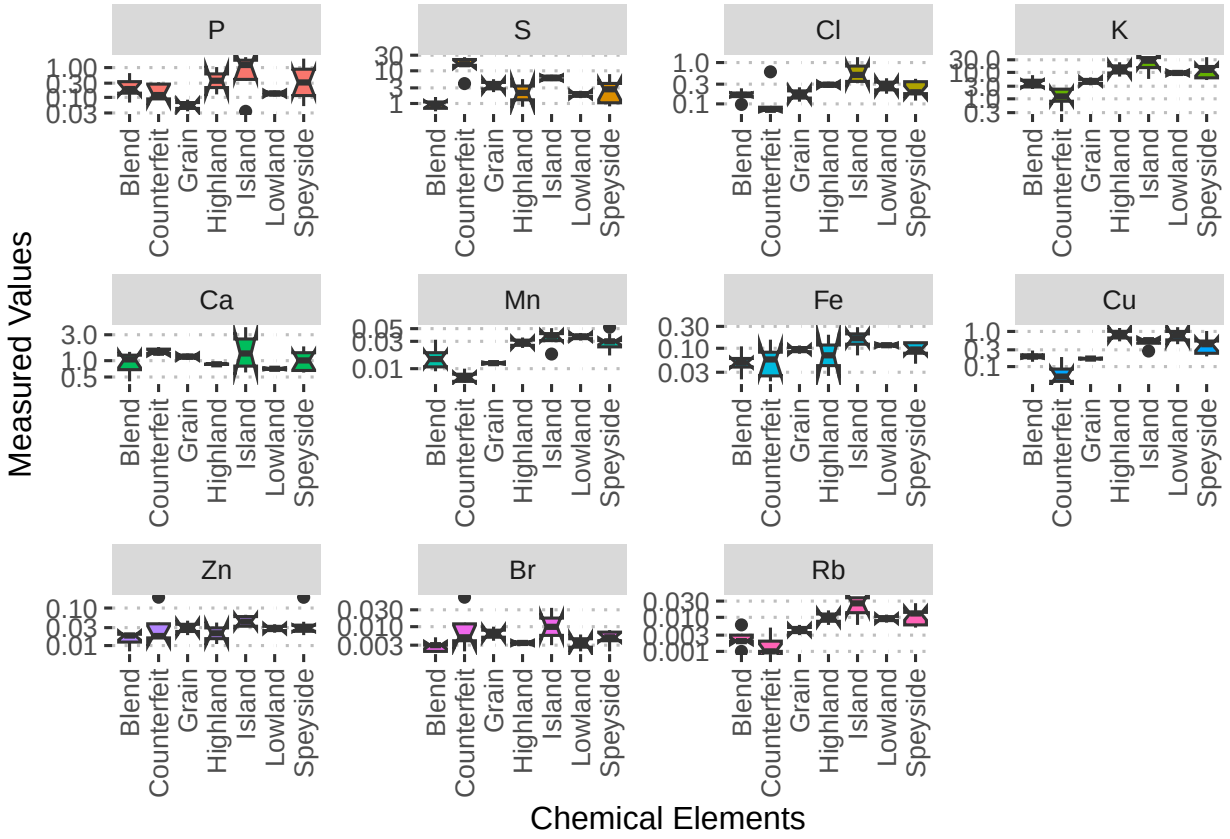

Figure 2: Panel boxplots of log-transformed measurements of observations (mg/L) faceted by chemical (P, S, Cl, K, Ca, Mn, Fe, Cu, Zn, Br, Rb), with observations grouped by whisky type (Blend, Counterfeit, Grain, Highland, Lowland, Speyside, Island).

Counterfeits (Counterfeit, Speyside: $T^2 = 7083$, p = 0.009), Blends (Counterfeit, Blend: $T^2 = 928{,}150$, p = 0.009) and Speyside (Speyside, Blend: $T^2 = 213.48$, p = 0.026) whisky classes all displayed significant mean vector differences and visually differing correlation structures while Island and Speyside (Both single-region origin malted-barley whiskies) did not (Island, Speyside: $T^2 = 474$, p = 0.042). Further, blended whiskies are composed primarily of grains (wheat or maize, 60-80%) with little malted-barley, while single origin whiskies are of only of malt character (Storrie 1962; Bower 2016; Kew et al. 2016; Scotch Whisky Association Cereals Working Group 2021).

BoxM tests for the homogeneity of variances was unable to be performed within our feature space as all class $n < p$, but correlation plots indicated likely differences in covariance matrices. These differences were to be visually supported when observations were projected within three-dimensional principal component (PC) space, and counterfeits appeared completely linearly separable from all other observations.

Thus we re-aggregated the sampled whiskies into three logical overarching predictive whisky classes: "Counterfeits" (n=5), "Grains & Blends" (a combination of grain and blended classes; n=10), and "Provenance" (all whiskies of a single-region origin; n=17). This reclassification is further supported as both grain type (barley vs. other) and region of production being shown to influence chemical composition of whiskies under other analytical methods, thus we should expect the regional fingerprint to be diluted within grain & blended whiskies (Kew et al. 2016; Roullier-Gall et al. 2020).

We then reassessed correlation plots of chemical observations for each of these three classes, as well as mean vector differences between them. We found striking correlation structure differences between the classes (particularly between counterfeit and provenance whiskies) as well as significant mean vector differences (3;

4).

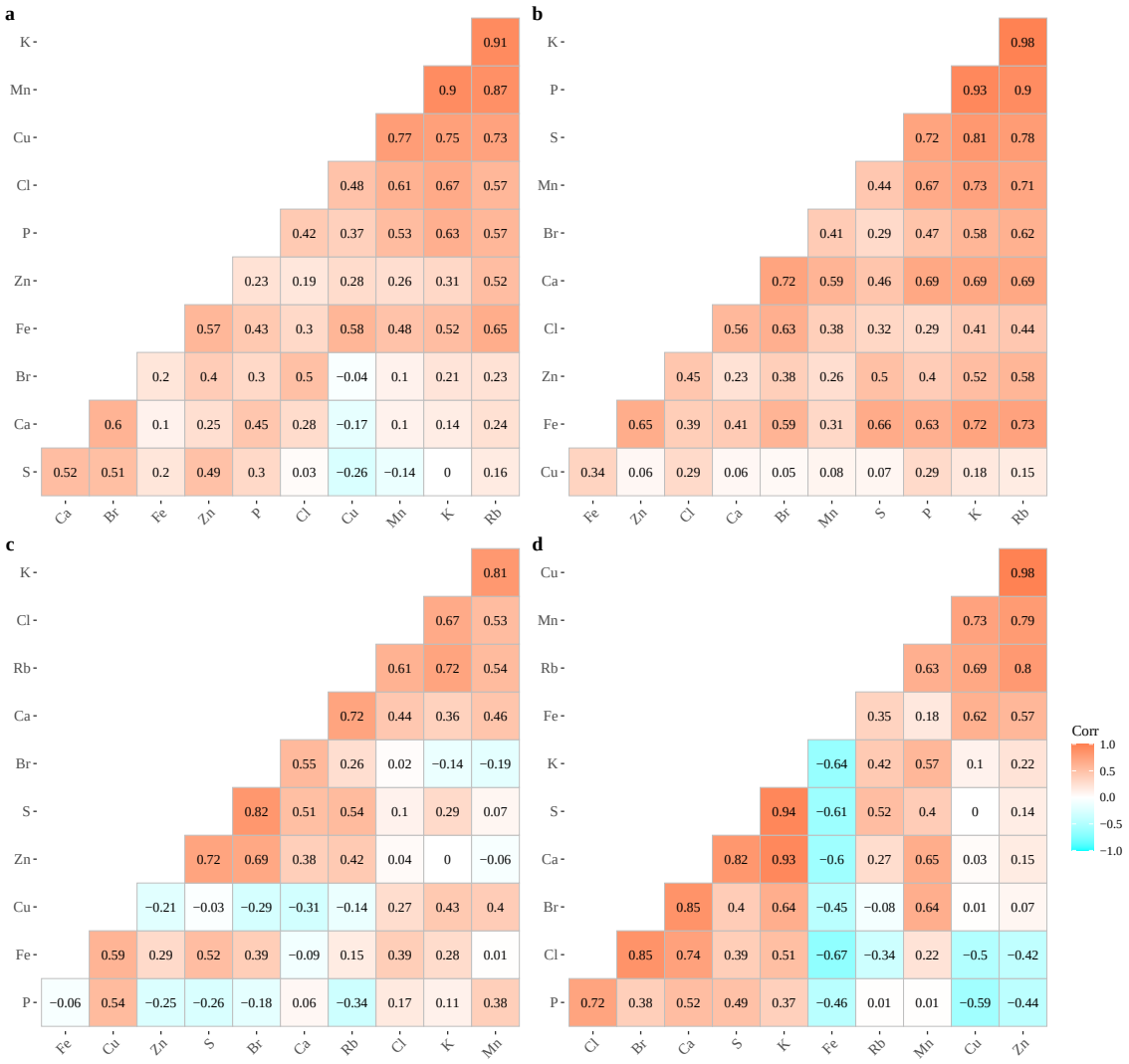

Figure 3: Correlation plots of whisky trace elements:(a) Correlation structure between all observations, (b) Correlation structure between whiskies of provenance, (c) corrrelation structure between blended whiskies, (d)correlation structure between counterfeit whiskies

Table 4: Hotelling's T^2 Test Results

Comparison	T2_Statistic	P_Value
Provenance vs Counterfeit	1181.90	0.000
Provenance vs Grain/Blend	137.44	0.000
Grain/Blend vs Counterfeit	474.57	0.042

Note:
Hotelling's T^2 tests were conducted on log-transformed whisky descriptors by the reaggregated classes providence, blends & grains, and counterfeits.

This led us to conduct Principle Component Analysis (PCA) to ascertain the magnitude and directionality of each chemical driver within a reduced space, as to infer and categorize chemical differences between groups later drawn from cluster analysis. Principal components analysis was conducted via using the prcomp() function on our scaled, log-transformed data. We returned to Shand et al. (2017)'s LDA procedure with our newly aggregated classes, and performed a global BoxM test to assess the equality of variance matrices for by group for PC1-PC3, which we found to be heterogeneous ($X^2_{12}, p = 0.041$) and thus discarded LDA as an analysis option.

We then implemented k-means, partitioning around medoids (PAM), and agglomerative hierarchical clustering to draw data-supported groups to test our predictive ones. In doing so, we scaled our log-transformed observations, except in the case of correlation distance hierarchical clustering which was conducted on the log-transformed data. Further, we wished to see if we could draw general consensus between these methods in support of using XTRF chemical composition sampling as a widely applicableand robust method to produce suitable data for general multivariate analysis to categorize whiskies by counterfeit, blend, or single origin status.

2.2 K-means Clustering

K-means analysis was conducted on the scaled-log transformed data using the kmeans() function with the default recommended Hartigan-Wong algorithm, iterations set to 100 (iter. max = 100), and 50 random starts (nstart = 50) for all values of k. Setting random initializations to 50 reduces the chance of poor initial centroid allocation, while allowing 100 maximum iterations ensures convergence during the process of iterative group allocation of data points.

This process was generated for $k = 1 - 10$, allowing k^* to be chosen via visual assessment of total within sum of squares (WSS) reductions with concurrent silhouette plot analysis.

2.3 Partitioning Around Medoids (PAM) Clustering

PAM clustering was conducted using the pam() function from the cluster package with Euclidean distance (Kaufman and Rousseeuw, n.d.). k^* assessment was conducted via assessing silhouette plots and widths as well as via comparing clustering results to k-means clustering for respective k^*.

2.4 Agglomerative Hierarchical Clustering

All agglomerative hierarchical clustering was performed using hclust(), with the original clustering results of Shand et al. (2017) reproduced via using euclidian distance with complete linkage. We attempted to find consistent hierarchical clustering results to these using other distances emphasizing absolute differences (Minkowski [a=3] and Chebyshev) and associated linkages (complete and Ward) as well as develop our own agglomerative hierarchical clustering models with better performance. Of those trialed, Euclidian (Ward linkage), Manhattan (complete and Ward linkage), and Correlation ($1 - r$, Ward linkage) distances were retained for comparison and analysis.

2.5 Quality Metrics

After k^* was established confusion matrices were produced for all clustering results and compared to our proposed groups (Provenance, blend/grain, and counterfeit whiskies) with global quality statistics (Acc, $F1\text{-score}_M$, TNR_M, $F1\text{-score}_\mu$, PPV_μ, TPR_μ) calculated. The best scoring clustering model was then assessed via class-wise quality metrics (Acc_i, MR_i, PPV_i, TPR_i, TNR_i, $F1\text{-score}_i$).

3 Results

3.1 PCA Analysis:

The first three principal components contained 79.16% of our datas variance within the PC feature space (S.D. - PC1 = 2.265, PC2 = 1.549, PC = 1.085; Variance explained - PC1 = 46.64%, PC2 = 21.82%, PC3 = 10.71%). Further principal components are associated with rapidly decreasing proportions of variance explained (4; 5).

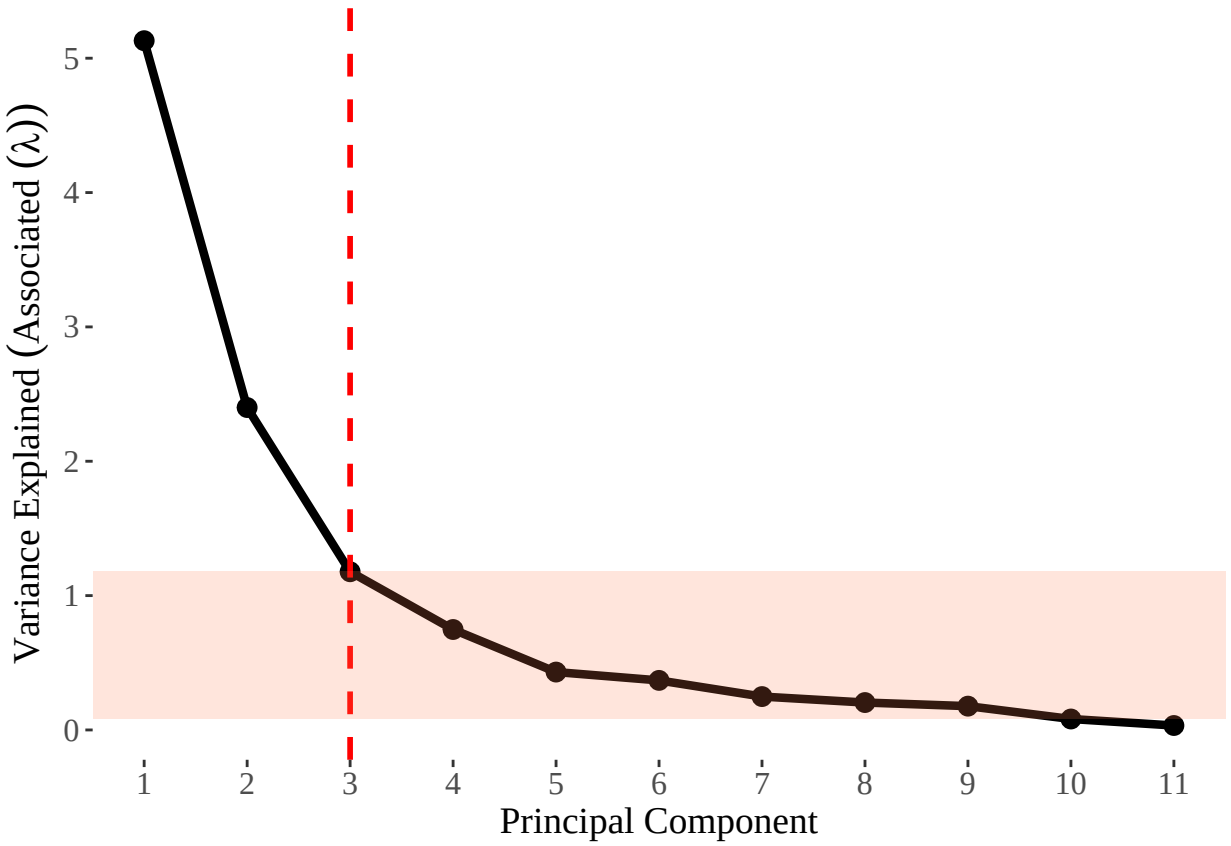

Figure 4: Elbow plot of principal components (PC) and associated eigen values (variance). The red line denotes where approximately 80% of the data's variance is contained (0.7916), with the remaining 20% associated with the red shading over PC 4-11

Table 5: Standardized Log-data PCA Summary

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11
Standard deviation	2.2650	1.5491	1.0852	0.8647	0.6564	0.6074	0.4985	0.4517	0.4210	0.2857	0.1818
Proportion of Variance	0.4664	0.2182	0.1071	0.0680	0.0392	0.0335	0.0226	0.0186	0.0161	0.0074	0.0030
Cumulative Proportion	0.4664	0.6846	0.7916	0.8596	0.8988	0.9323	0.9549	0.9735	0.9896	0.9970	1.0000

Note: Standardized Principal Components of the scaled log-transformed whisky trace chemical observations and their associated standard deviation, with proportional variance explained.

PC1 variance was composed by all positive loadings primarily from P = 0.31, Cl = 0.315, K = 0.404, Mn = 0.379, Fe = 0.310, Cu = 0.327, Zn = 0.241, and Rb = 0.415 with minor contributions from Br,Ca and S

(<0.20). The largest loadings contributing to PC2 are those contributing little to PC1 such as Br (0.454), Ca (0.475) and S (0.534), with others generally having negative or small contributions. PC3 further differentiates the PC space with varying large positive and negative inputs (Table 6).

Table 6: Principal Component Loadings: First Three Components

	P	S	Cl	K	Ca	Mn	Fe	Cu	Zn	Br	Rb
PC1	0.311	0.094	0.315	0.404	0.149	0.379	0.310	0.327	0.241	0.183	0.415
PC2	0.118	0.534	0.006	-0.161	0.475	-0.239	-0.004	-0.347	0.244	0.454	-0.074
PC3	-0.196	0.223	-0.422	-0.131	-0.301	-0.132	0.469	0.118	0.570	-0.205	0.088

Note: Loading contributions of chemical variables to the first three standardized principal components of the scaled log-transformed whisky trace element observations.

Within the PC space counterfeit samples are generally composed of positive PC2 scores (elevated levels of Br, Ca and S) and negative PC1 scores (reduced levels of the rest of the chemical trace elements). Inversely, most single-origin whiskies are characterized by positive PC1 scores (elevated levels of P, Cl, K, Mn, Fe, Cu, Zn) and negative PC2 scores (reduced levels of S, Br and S) with some variation. Blends and grains also are generally sit negatively below both the PC2 and PC1 abscissus and thus are primarily composed of low to average levels of most trace elements (Fig. 5).

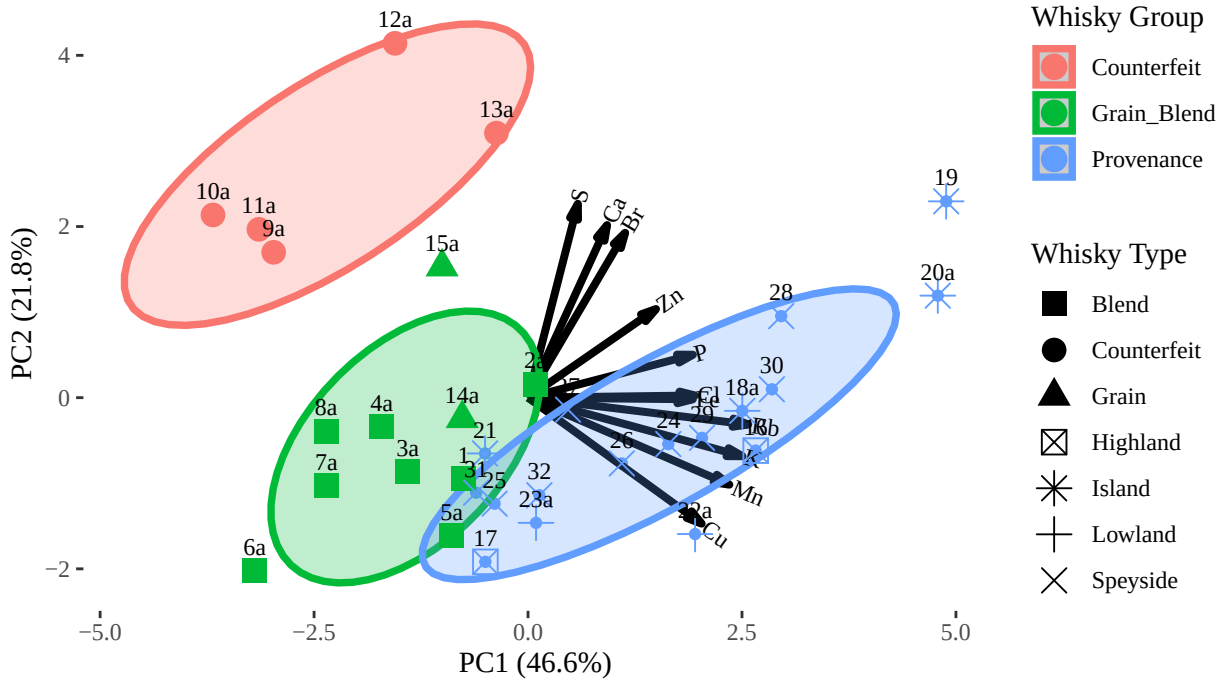

Figure 5: Biplot within the first two principal components feature space with chemical variable loadings, with data points colored by group (provenance, blend/grain and counterfeit) and shapes associated by class (counterfeit, grain, blend, highland, lowland, speyside, island). Confidence ellipses represent group distribution estimated within one standard deviation based on a multivariate normal distribution.

3.2 K-means

With the stated starting parameter metrics, we produced k-means clustering algorithms using $k = 1, 2, ..., 10$. We then extracted total within sum of squares (WSS) for each model, and via an elbow plot of WSS as a function of k, visually assessed a parsimonious k^* candidate visually. The sharpest shift in proportional reduction of WSS occured at $k = 3$, explaining 37.37% of all WSS composing the remaining variance reductions in $k = 2 - 10$ (Fig. 6). For $k = 3$, 51.6% of variance is explained (Between SS/Total SS), average class silhouette width is 0.30, WSS = 164.9 and Between SS = 176.1. This shows moderate fit of clusters overall and in comparison to $k = 4$ displays a fairly equitable fit, though cluster size ($k = 3$: 10,6,16; $k = 4$: 8,2,6,16) indicates more balanced clustering from $k = 3$ modelling. $k = 4$ isolates 2 points which appear highly differentiated in our feature space from all observations (Island 18, Island 19).

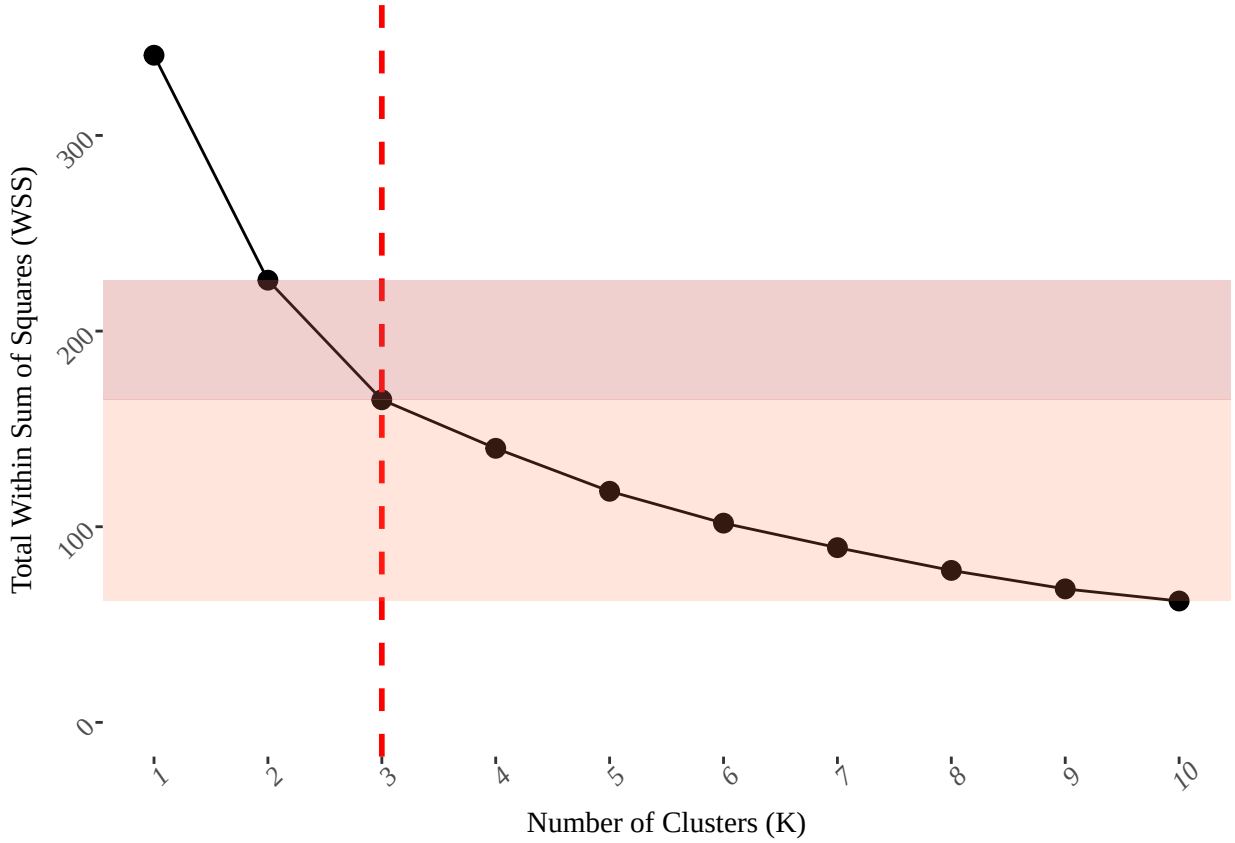

Figure 6: Elbow plot of k-means derived clusters for k = 1,2,..,10 using euclidian distance and the Hartigan-Wong algorithm. The dashed red line indicates likely optimal clustering at k=3, from visual assesment of proportion of variance reduced. Dark red shading indicates 37.37% of the variance explained from k=3 in comparison to every further group addition beyond k=3

Silhoutte plots (Fig. 7) indicate that for $k = 3$ clusters 2 ($S_2 = 0.36$) and 3 ($S_3 = 0.28$) have moderate fit, while cluster 1 ($S_1 = 0.15$) displays weak to insubstantial structure; only one observation (15 = young grain) displays a truly poor fit to a group ($S_i < 0$). Clustering suitability remains similar for groups 1 and 3 when $k = 4$, but the 2 points forming a new group indicate a very tight, suitable cluster ($S_4 = 0.49$; WSS = 140.08) while S_2 is reduced to 0.30. This trade-off only reduces average silhouette width by 0.02, indicating that there is minimal reduction in group fit and that this is a feasible cluster model as well.

Table 7: K-means Clustering Comparison

Metric	K = 3	K = 4
Cluster Sizes	10, 6, 16	8, 2, 6, 16
Variance Explained	51.6%	58.9%
Avg Silhouette	0.30	0.28
Total Within SS	164.9	140.08
Between SS	176.1	200.92
Total SS	341	341

Note:
Summary statistics for k-mean cluster analysis with k=3 and k=4.

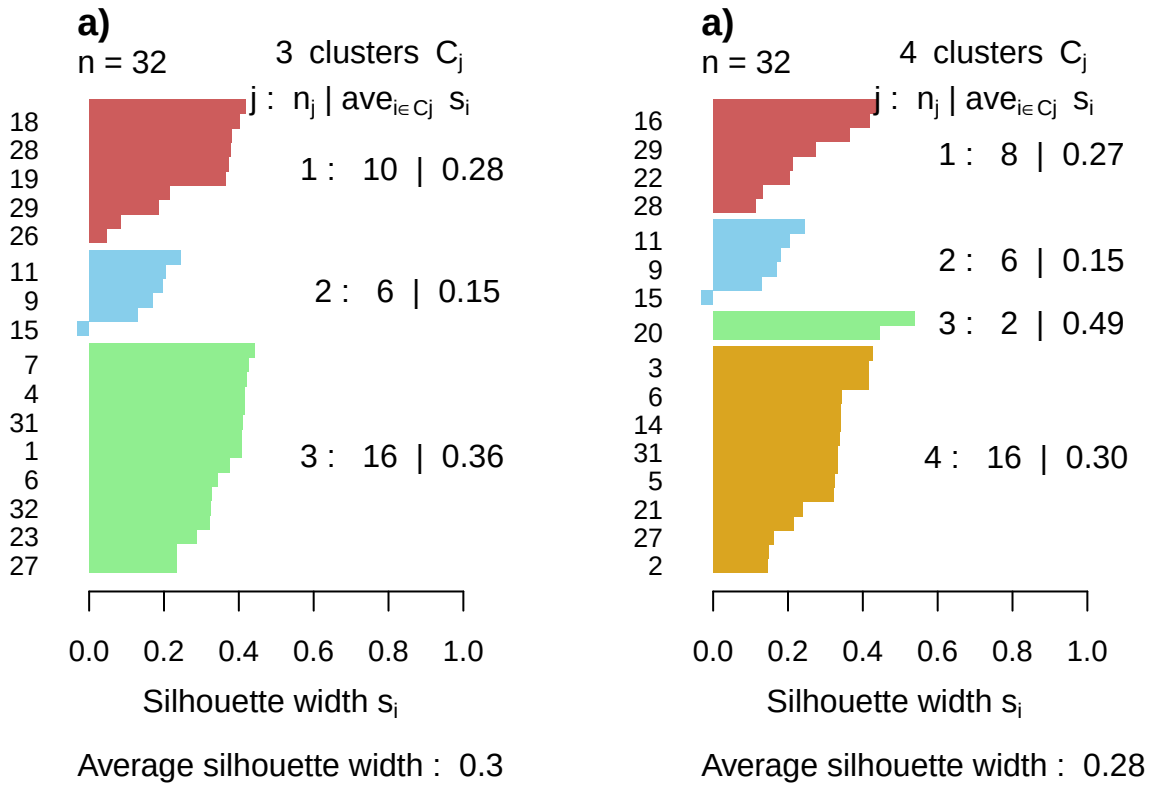

Figure 7: Silhouette plots for cluster candidates k=3 and k=4 displaying average silhoute scores. Silhouette scores range from -1 to 1, with suitable cluster structure indicated by higher positive scores.

The fourth group composed of 2 island observations ($k = 4$) sits in the feature space characterized by both positive PC1 and PC2 scores, quite different from other group trends. This aligns with Shand et al. (2017)'s linear discriminant analysis (LDA) results utilizing the first three principal components, in which they were able to classify 2 out of 4 island whiskies correctly. For $k = 3$ these observations are clustered along with 8 other single-origin whiskies, and all other cluster classifications remain the same. All five counterfeits have been correctly clustered, with one false positive which aligns with our poor fitting sample 15 (young grain). Further, all other grains and blends have been correctly classified into one group, but with 7 false positives

of provenance whiskies also classified into that group.

```
## Too few points to calculate an ellipse
```

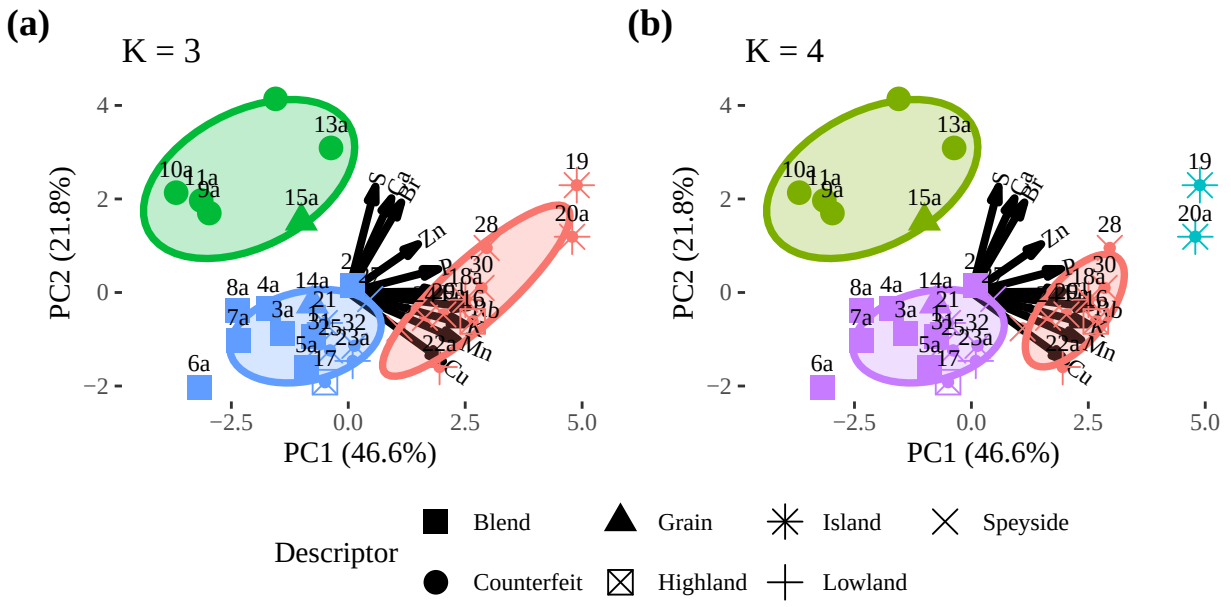

Figure 8: Biplots of data-points within principal components feature space by k-means clusters with ellipses indicating a confidence interval within one standard error. Legend denotes original whisky class. (a) Displays k=3 model with cluster 1 indicated by red, cluster 2 by green, and cluster 3 by blue colouration. (b) Displays k=4 model with cluster 1 indicated by red, cluster 2 green, by cluster 3 by aqua, and cluster 4 by purple coloration.

3.3 PAM

Overall, PAM clustering at $k = 3$ produced fairly equitable results to k-means clustering at the same cluster size, but at $k = 4$ produced much different results (Table 8). At $k = 3$ the average silhouette width was 0.294 with relatively similar by-group s_i, and group sizes of similar observations with medoids 4 (Blend), 10 (counterfeit) and 18 (island). However, for $k = 4$ the average silhouette width decreased to 0.149, with group 1 (Fig. 9) being separated into two which drastically reduced that clusters separation metric from 2.304 to 1.839 per cluster, heavily reducing cluster s_i for that group as well as for the cluster primarily composed of counterfeits (Table 8). Due to both the reduction in silhouette fit and inconsistancies with k-means modelling at $k = 4$, we will distinguish $k^* = 3$ as our robust and optimal group number.

For k^* clustering results were almost identical to k-means, except counterfeits were isolated completely with no false positives and the young grain observation (15) was included in cluster 1 along with the other 9 previously grouped grains and blends, and the 7 provenance whiskies.

Table 8: PAM Clustering Results

K	Cluster	Size	Medoid	Avg Diss.	Separation	Avg Silhouette
1	17	4	2.234	2.304	0.344	3
2	5	10	2.962	2.653	0.138	3
3	10	18	2.277	2.304	0.287	3
1	8	4	1.875	1.839	0.080	4
2	9	25	1.848	1.839	0.192	4
3	5	10	2.962	2.653	0.074	4
4	10	18	2.277	2.304	0.204	4

Note: Summary statistics for PAM cluster analysis with k=3 and k=4. K=3 Overall Avg Silhouette: 0.294; K=4 Overall Avg Silhouette: 0.149.

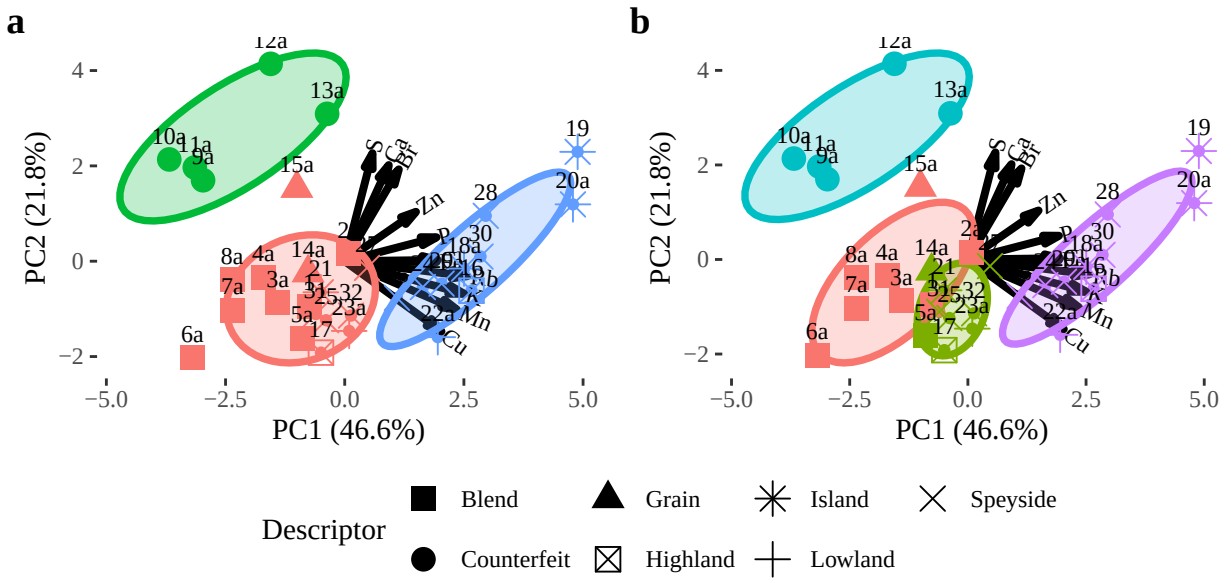

Figure 9: Biplots of data-points within principal components feature space by PAM clustering with ellipses indicating a confidence interval within one standard error. Legend denotes original whisky class. (a) Displays k=3 model with cluster 1 indicated by red, cluster 2 by green, and cluster 3 by blue colouration. (b) Displays k=4 model with cluster 1 indicated by red, cluster 2 green, by cluster 3 by aqua, and cluster 4 by purple coloration.

3.4 Hierarchical clustering

Using k^*, we reproduced dendrogram results from Shand et al. (2017) using agglomerative hierarchical clustering with euclidean point-to-point distance and complete linkage, but could find no other distance

and linkage combination which were consistent with these results (Fig. 10). These included implementing various cases of power distances emphasizing large differences in our feature space such as the Chebyshev (linkage: complete and Ward) and the Minkowski (a=3; linkage: complete and Ward), as the model from Shand et al. (2017) separated those observations belonging to more distinctly different regions of our feature space (observations 18:island, 19:island, and counterfeit whiskies) while retaining a third cluster of fairly homogeneous observations across blends and provenance whiskies.

As we found these results relatively non-useful, we proposed finding distance and linkage combinations which would reinforce our previous k-means and PAM clustering results at k^*. We found the using the euclidian (Ward linkage) and the Manhattan (both complete and Ward) produced equal cluster results to k-means clustering at k^*. Further we applied a Pearson correlation coefficient distance ($d = 1 - r$) between whisky observations combined with a Ward linkage, such that whiskies should be aggregated by similar relative patterns of chemical variables, even if absolute concentrations differ. We thought this method of profiling whiskies this way may yield interesting results, and clustered counterfeits as k-means clustering did but only classified 7 rather than 10 provenance whiskies into a single group, adding the remaining three to a larger pool of blend, grain and provenance whiskies (Fig. 10).

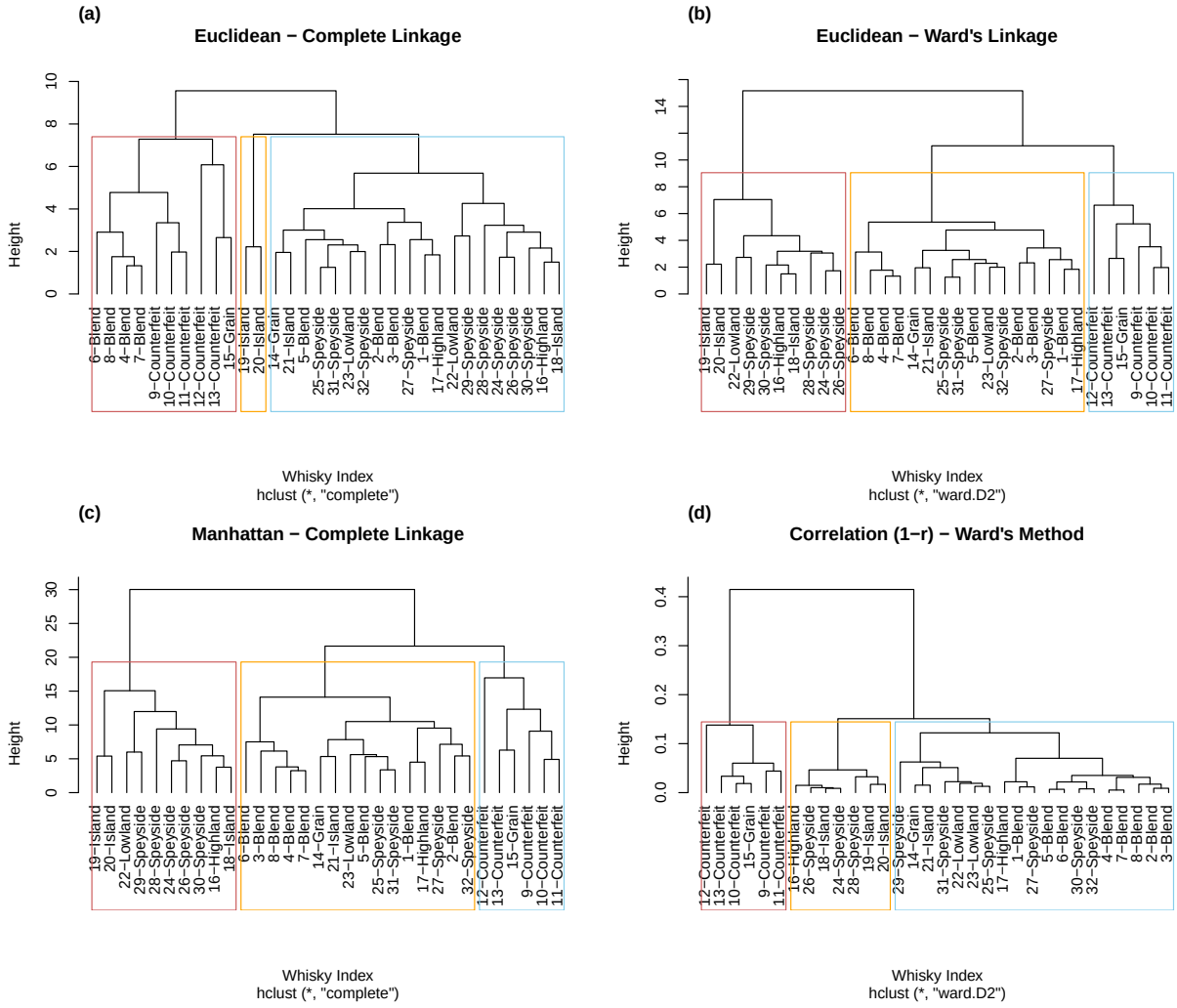

Figure 10: Agglomerative hierarchical clustering plots for whisky data. Panel (a) denotes the repordiced figure from Shand et al., 2017 using euclidean distance with a complete linkage; (b) denotes the euclidean distance using Ward linkage; (c) denotes the manhattan distance using complete linkage; (d) denotes the correlation distance (1-r) using Ward Ward linkage. Coloured boxes segregate selected k=3 groups given the distance and linkage.

3.5 Quality metrics

Confusion matrices were calculated, and global and classwise quality metrics derived. K-means (at k^*), Manhattan agglomerative (complete and Ward), and Euclidian agglomerative (Ward) all displayed equal clustering results. Global quality statistics displayed that PAM clustering consistantly performed the best with the highest metrics: $OAcc$ (78.1%), $AAcc$ (0.854), $F1\text{-score}_M$ (82.7%), TNR_M (89.4%), $F1\text{-score}_\mu$ (79.3%), PPV_μ (83.5%), and TPR_μ (79.3%). K-means, euclidian (Ward) agglomerative and Manhattan (complete and Ward) agglomeration modelling displayed similar performance ($OAcc = 0.75$; Table 9, followed by poorer performance by correlation distance agglomeritive clustering. Euclidean (complete) clustering performed the poorest overall across all metrics (Table 9.

Pam class-wise performance shows that this model classified all counterfeits correctly, aligning with Shand et al. (2017)'s LDA results (Table 10. However, our provenance class displayed a specificity of 68.2% (TNR_i) and a 58.8% recall (TPR_i) showing that approximately 41% of this class were misclassified as blended whiskies (false negatives). This, along with perfect precision and our weaker recall shows that a

Table 9: Clustering Method Performance Comparison

Method	Overall Acc.	Average Acc.	F1 (Macro)	TNR (Macro)	F1 (Micro)	TNR (Micro)	TPR (Micro)
Consensus	0.750	0.833	0.781	0.882	0.758	0.801	0.758
PAM	**0.781**	**0.854**	**0.827**	**0.894**	**0.793**	**0.835**	**0.793**
Correlation HC	0.656	0.771	0.704	0.836	0.677	0.735	0.677
Euclidean HC	0.375	0.583	0.404	0.711	0.452	0.479	0.452

Note: Global confusion quality metrics for evaluating clustering method performance at k =3; the best performing model is highlighted with grey background. Consensus results refer to aggreeing clustering models (k-means, hierarchical (euclidean [Ward], manhattan [Ward & complete]))

very strong true positive signal, but moderately strong chance of incurring false negatives. Inversely, our presision (PPV_i) for blended and grains was 58.8%, showing that approximately 41% of predictions within this class were wrong (false positives), with our specificity (TNR_i) of 68.2% indicating a moderate lack of class differentiation ability between blended/grain and provenance whiskies .Both classes sharing a balanced accuracy (Acc_i) of 78.1%, though their deficits and strengths differ.

Table 10: PAM Class-wise Performance

Class	TP	TN	FP	FN	ACC_i	MR_i	PPV_i	TPR_i	TNR_i	F_class
Counterfeit	**5**	**27**	**0**	**0**	**1.000**	**0.000**	**1.000**	**1.000**	**1.000**	**1.000**
Blended	10	15	7	0	0.781	0.219	0.588	1.000	0.682	0.741
Provenance	10	15	0	7	0.781	0.219	1.000	0.588	1.000	0.741

Note: Counts and derived quality metrics for PAM clustering performance by class.

4 Discussion

Reference

Bower, Julie. 2016. "Scotch Whisky: History, Heritage and the Stock Cycle." *Beverages* 2 (2): 11.

Kaufman, L., and Peter J. Rousseeuw. n.d. "PAM Clustering Algorithm."

Kew, Will, Ian Goodall, David Clarke, and Dušan Uhrín. 2016. "Chemical Diversity and Complexity of Scotch Whisky as Revealed by High-Resolution Mass Spectrometry." *Journal of the American Society for Mass Spectrometry* 28 (1): 200–213.

Power, Aoife C, Caoimhe Ní Néill, Sive Geoghegan, Sinéad Currivan, Mary Deasy, and Daniel Cozzolino. 2020. "A Brief History of Whiskey Adulteration and the Role of Spectroscopy Combined with Chemometrics in the Detection of Modern Whiskey Fraud." *Beverages* 6 (3): 49.

R Core Team. 2025. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Roullier-Gall, Chloé, Julie Signoret, Christian Coelho, Daniel Hemmler, Mathieu Kajdan, Marianna Lucio, Bernhard Schäfer, Régis D Gougeon, and Philippe Schmitt-Kopplin. 2020. "Influence of Regionality and Maturation Time on the Chemical Fingerprint of Whisky." *Food Chemistry* 323: 126748.

Scotch Whisky Association Cereals Working Group. 2021. "Scotch Whisky Cereals Technical Note." The Scotch Whisky Association. https://www.scotch-whisky.org.uk/media/1900/cereals-technical-note-6th-edition-240821.pdf.

Shand, Charles A, Renate Wendler, Lorna Dawson, Kyari Yates, and Hayleigh Stephenson. 2017. "Multivariate Analysis of Scotch Whisky by Total Reflection x-Ray Fluorescence and Chemometric Methods: A Potential Tool in the Identification of Counterfeits." *Analytica Chimica Acta* 976: 14–24.

Storrie, Margaret C. 1962. "The Scotch Whisky Industry." *Transactions and Papers (Institute of British Geographers)*, no. 31: 97–114.