Comunidad
MEXICANA
DE GESTIÓN PÚBLICA
PARA RESULTADOS

Building global policy inferences for alcoholic beverages harmful consumption: A cluster analysis & machine learning approach.

A Report by The Research Team

-- August 2024 --

Building global policy inferences for alcoholic beverages harmful consumption: A cluster analysis & machine learning approach.

Farfán Mares, G.[1], Esteban Bruera[2] y Pedro Torres[3], Building global policy inferences for alcoholic beverages: A cluster analysis & machine learning approach.

The author(s) would like to thank all the received support for the research, authorship, and/or publication of this article. During the design phase of this work, the Comunidad Mexicana de Gestión Pública para Resultados benefited from research resources provided by a Non-Resident Scholar from the Boston University Global Development Policy Center.

1 Corresponding author (g.farfan.mares@comunidadmexicana.org.mx). Executive Director of the Mexican Community on Public Management for Results & Adjunct Professor of Public Finance, Georgetown University SFS-CLAS, Washington, D.C.
2. Head of research, Mexican Community on Public Management for Results, Mexico City.
3. PhD Candidate in Social Policy, London School of Economics, London, U.K.

First Edition, 2024.
Copyright © Comunidad Mexicana de Gestión Pública para Resultados, A.C., 2024.
Printed and made in Mexico.

Content

Introduction

Since 2014, the Mexican Community of Public Management for Results, A.C. (hereinafter *Mexican Community*), has produced pioneering research on a variety of fiscal and economic dimensions of alcohol beverage consumption, both at national and subnational level. It has also produced a variety of international, global comparisons on a variety of types of beverages.[1] Thus, our research experience and accumulation of knowledge is crucial in the design and development of the present report which uses artificial intelligence extensively.

A sizeable number of individuals consume alcohol for leisure, yet in some countries and regions this phenomenon leads to a harmful pattern which has a direct impact on people's health, their surroundings, health systems, and, ultimately, obligate governments to tackle its negative and generalized impact. In this report, we want to understand the *ecosystem* in which such consumption takes place by differentiating types of beverages (fermented and distilled). We do this because we consider that harmful alcohol consumption is associated with the preferred type of beverages rather than alcoholic beverages or pure alcohol consumption in general. Since we believe that the policy dimension is key and governments, along with other organizations, have the mandate and obligation to tackle harmful alcohol consumption, we use countries as the basis of our unit of analysis.

This research furthers a novel way to understand the environment in which harmful consumption of alcohol around the world is embedded in. Differentiation refers to beverage strength (in terms of pure alcohol content) and their gender impact (female and male consumption). We use a variety of health data and indicators to incorporate the health dimension and social impact in our model as well (disorders and road injuries). We cluster or group countries by their preferred type of beverage, gender, age, health and social indicators and thus place countries by their probability, or likelihood, to belong to a cluster.

Mexican Community proposes an *ecosystem* analysis based on *differentiation* (beverage strength and gender) which uses a variety of machine learning methods as practical and useful tools to build global policy inferences. This analysis aims to find affinities, generate identities, assemble associations, and open new spaces to rethink global policies. We aim to provide harmful alcohol consumption phenomenon' stakeholders with global policy guidelines.

Mexico City, August 2024.

[1] At the end of this report our published research in alcoholic beverages is listed chronologically. Full documents can be accessed on the following link: https://www.comunidadmexicana.org.mx/impuestos-bebidas-alcoholicas

How to rethink global policy

In recent years, information technologies, data science, and publicly available and reliable statistical resources have significantly improved. Progress has taken place both in the health sector and in other areas, such as data on non-communicable diseases (NCDs). This is certainly the case of individual's harmful consumption of alcohol beverages where addictive patterns entail a personal, contextual, and social burden. Not only exists a set of statistics and indicators on the harmful consumption of alcohol at global, national, and subnational levels, but their accessibility has improved enormously. Accordingly, this report takes advantage of public and non-commercial databases produced by a variety of regional and global institutions. It uses artificial intelligence based on a Machine Learning Model (MLM) to rethink global policy in order to prevent harmful consumption of alcoholic beverages. We use this section to describe our claims and their rationale.

From an international and global perspective, what models alcohol beverage harmful consumption and its negative impacts? To answer this question, abundant research has been produced departing from an individual and/or contextual/social dimensions as principal units of analysis. The advancement of research has been able to explain to what extent an individual is more likely to develop addictive behavior when it comes to consuming alcoholic beverages. It has reached the robustness needed to propose a variety of probable scenarios and therefore has come close to predicting future events. Science has reached not only the medical discipline as such, but also those who outside the health community contribute, using data science and other statistical tools, to the public health agenda and ultimately, health as a global public good. In conclusion, there is abundant research which focuses on identifying, analyzing and forecasting the impacts, both individual and social, of harmful consumption of alcoholic beverages.

Nevertheless, we consider that there are macro determinants on how to approach these phenomena that had been relatively unexplored and which offer useful insights and possess future potential. We believe these should substantiate a macro, global perspective which transcends national borders and regional and continental divides as well. In this sense, we set forth that the type and alcohol content of beverages consumption, along with health, life, and risk behavior related with excessive consumption of alcohol are all key components of a model to build global policy inferences. Inferences are crucial to begin a scientific, evidence-based inquiry. "Out of the box" inferences are ambitious but might have the potential to open new areas of thought and research. At the least, we deem that their potential relies on their explanatory power to confirm or refine existing policy approaches (both conceptual and practical) to the prevention and treatment, as well as reduction, of the harmful consumption of alcoholic beverages at global scale. We consider this research to be a way to rethink, or at least refine some of the comparative and global policy approaches a variety of actors had been, correctly, pursuing.

Countries formulate, along with authorities and a collection of public and private national and international institutions, a variety of policies aimed at reducing the harmful consumption of alcohol. Often, these countries converge on the basis that they belong to a group of peers which are grouped in regions based in a variety of criteria (language, culture, economic structure, etc). Yet, in terms of alcohol consumption and its corresponding ills, we might argue that these criteria can guide or misguide our policy assumptions. We might be prone to think that a policy in a country can be replicated partially or entirely to another one or a group of countries. Yet, at first sight or by using descriptive statistics countries might look similar or relatively comparable but they can also have substantial differences in terms of individual and social exposure to harmful consumption of alcoholic beverages. We try to rethink this issue by using hierarchical clustering (a method to group categories), which is a feature of artificial intelligence and standard machine learning models (MLM). We believe that MLM is far more accurate for country profiling and therefore, for a global, all-encompassing policy perspective than a regional or traditional approach.

In short, an underlying leitmotiv of this report tries to avoid the dichotomy between cultural, or idiosyncratic and strictly behavioral components. It proposes to understand alcohol harmful consumption by introducing a dimension of analysis based on the concept of *ecosystem*. We consider that this approach is more suitable to understand and rethink international, or global comparative policy which aims to prevent and treat harmful consumption of this disease. We aspire to reconcile the individual dimensions of such phenomena (magnitude and preference of alcohol consumption) with the health and social impacts, i.e. negative externalities to build a country-based ecosystem to map and identify, by either reinforcing or contesting, country, national-level policies to prevent and treat this disease. This is made by using 31,800 observations in 190 countries and 31 years (from 1990 to 2021).

To reach a global perspective we treat countries as individuals and therefore we consider them our key unit of analysis. We identify countries by their alcohol consumption preferences, i.e. which type of alcoholic beverage like and in what quantity in terms of pure alcohol. Pure alcohol is consumed in different concentrations (% of pure alcohol) and thus is properly associated to types of beverages (fermented, distilled, etc). At the same time, countries present different types and magnitude of negative impacts on individual health. We proceed to understand how these individuals (countries) can be considered by themselves an ecosystem in which specific outputs are registered and therefore prone to be observed and analyzed as an evidence-based knowledge. These outputs are basically three: related disorders, death rates and road injuries all associated with alcohol harmful consumption.

In addition to the above considerations, the negative impact in terms of health disorders or fatalities, particularly upon a gender gap perspective, can be quite different. Even if countries have similar if not identical consumption preferences and magnitudes, they might denote a different impact in terms of alcohol related disorders, fatalities or road injuries with an unequivocal link to alcohol consumption. We incorporate the gender perspective into all our analysis and MLM procedures to isolate the gender dimension and fully identify a *genderized ecosystem*.

Undoubtedly, when it comes to alcohol consumption, geography, identity or other cultural aspects do matter. Indeed, there exist countries where fermented beverages like wine or beer are by far the favorite type of alcoholic beverages. From a health perspective, a country's alcoholic type of beverage preference can be treated as the main source of pure alcohol. This can be true for normal, standard consumers or heavy drinkers. There are several studies that indicate drinking patterns are not related to a specific type of beverage but a type of beverage which is more popular, accessible or affordable. In other words, people that consume alcohol beyond permissible limits and therefore can cause harm to their health and a burden on others can have a variety of availability (type of beverage) for their needs. Of course, beverage preference is a proxy of geography, identity or culture so these rather unquantifiable aspects are internalized in a way.

Without the use of MLM it will be obvious that, for example, the Baltic Sea countries which are characterized by a high level of drinking per capita in a day/week/month/year cannot be compared to other countries which belong to other regions. Policies to prevent and treat the harmful or excessive consumption of alcohol, as often happens, will always be contested on the basis that that they are a group of countries which undeniably have a lot of factors in common and therefore are exceptional. Likewise, and at the other extreme, regions and countries which record no or negligible alcohol consumption because of cultural or religious aspects (or even prohibits consumption) like the Middle East and North African, MENA, countries should be treated separately in terms of analysis and, perhaps more important, for policy purposes. We believe that all countries should be compared, classified and clustered regardless of their proximity or affinities. Finally, it is important to note that “Harmful” is a key word and must be considered in relative terms. This means that “harmful” varies considerably when pondering the elements mentioned above (proximity to a country, countries or belonging to a region or subregion with a path dependent observable trajectory of “x” intensity of consumption either high (Baltic Sea countries), or low (MENA countries). Lastly, the former analysis should consider a differentiated approach among men and women to internalize a gender ecosystem too.

Further in our research, we will try to claim that differences arise between countries and regions precisely because of the nature of the ecosystem in which consumption is embedded. Within such an ecosystem, and despite the lack of available and reliable data which comprises all countries, we suggest some core variables. Once the countries are grouped by their consumption preferences and “environment”, it is contextualized with the probability that this translates into “harmful” consumption or undesirable results derived from alcohol consumption, attributing a probability given its ecosystem.

Literature Review

Research on alcoholic beverages harmful consumption is mostly produced by people and organizations directly or indirectly associated with the health sector. Researchers are often medical doctors or people specialized in the health sector which build research cases from their studies and experience. They use an ample variety of perspectives and disciplines which are familiar to the public health community. Within the health sector, alcohol beverages harmful consumption is often labeled as alcohol use disorder, or AUD. AUD is a health condition resulting from unrestrained and harmful drinking.

There is another branch or community of scholars who are interested in knowledge agendas which are mostly interested in the policy dimension, and more specifically the economic and fiscal substratum of AUD. In the present document, our aim is to bring both the core tradition of public health perspective and the policy and fiscal school of research together, since we consider both as complementary. Although pertinent to the purpose of this research, the literature review included in this section cannot be considered exhaustive.[2] In our view is an updated and a proper sample of some of the most important research pieces on AUD and the harmful consumption of alcoholic beverages that helps to bridge the gap between both the health and policy-fiscal domains.

We will first address some of the key aspects we consider on cross-country comparisons, perspectives on alcohol beverages consumption preferences by type (pure alcohol content), and finally research that uses statistical and machine learning to either better understand AUD or predict potential outputs and outcomes of it based in such models. At the end we include a special revision on gender with a special emphasis on socioeconomic status and taxation.

Cross-country comparisons

It is a fact that in recent years a vast amount of data and numbers of sources directly or indirectly related to AUD have grown considerably. As expected, global parties as the World Health Organization and their regional counterparts such as the Pan American Health Organization had been updating and, in some cases, expanding their databases. Academic institutions from literally all continents had also produced a vast amount of data as researchers as well. Global, regional, and local

[2] This section only mentions or cite the research pieces which are most related to our research. An extensive list of references is included at the corresponding section at the end of this paper.

firms which produce a variety of alcohol-content beverages had also made efforts to produce data, although with restrictions, and "intermediate" organizations such as think tanks, civil society organizations, and the like. We select a couple of examples we believe are closer, at least from inception to our endeavor.

Cross-country comparisons typically address policy comparisons and, in some cases, use health data & indicators in their proposals. For example, Caswell and others (2022) propose an index on alcohol policy (the International Alcohol Control policy index, or IAC), with a special emphasis on stringency and impact. The IAC is comprised of 4 policy domains for the purpose of benchmarking and assessing change in a variety of countries. The IAC is produced by a group of researchers in 12 countries (New Zealand; Australia; England; Scotland; Netherlands; Vietnam; Thailand; South Africa; Turkey; Chile; Saint Kitts and Nevis and Mongolia). The four policy domains are availability, pricing policy, alcohol marketing, drink driving and are associated with alcohol per capita consumption (APC) of commercial (recorded) alcohol (Casswell, et. al., 2022:1). In the opinion of the authors, these 4 policy domains are the most effective and relevant ones.[3] The IAC "not only include data on the legislation pertaining to these policies (stringency), but also measures of the way in which these policies had actually affected key aspects of the alcohol environment (policy impact) using measures of the alcohol environment" (Casswell, et. al., 2022:1-4).

More recently, Díaz and colleagues (2024) develop an alcohol preparedness index (API), to "to assess the existence of alcohol-related public policies for each country … the long-term association of the country's alcohol preparedness index in 2010 with the burden of chronic liver disease, hepatocellular carcinoma, other neoplasms, and cardiovascular disease (is evaluated). The strengthening of alcohol-related public health policies could impact long-term mortality rates from cardiovascular disease, neoplasms, and liver disease. These conditions are the main contributors to the global burden of disease related to alcohol use" (Díaz, Luis Antonio, et. al.: 409). As a tool to assess the establishment of alcohol related public health policies worldwide API finds lower mortality with a higher position in the index and therefore a more stringent policy environment translates into a decline in alcohol morbidity and mortality.

[3] The policy domains are: 1. Trading hours/days of sale (on-premises/off-premises); 2. Outlet density (on-premise/off-premise); 3. Pricing (tax rate calculated as % of price?); 4. Marketing (restrictions, advertising, sponsorship, promotions, product placement); 5. Dring Driving (Blood Alcohol Content, BAC level) and enforcement.

Pure alcohol consumption preferences (beverage type)

At the *Mexican Community*, we have underscored the importance of considering not only drinking patterns/consumption behavior but the type of alcoholic beverage preference, which is a node where identities and affinities converge.[4] Taylor, et. al. goes beyond, by asserting that "alcohol consumption patterns within a community are a reflection of that location's political structure, laws and regulations, societal norms and traditions, dominance of anti-alcohol religious convictions, average income levels and economic standings, as well as motives, behaviors and beliefs" (2019: 28). Indeed, the "ecosystem" in which alcoholic beverage consumption is embedded in matters, but it is difficult to understand its complexities particularly from a comparative, even global, perspective. Always, and for good reason, countries are considered unique.

In our view, first, it is important to disentangle both "normal" or standard consumption to risky or harmful consumption. Of course, the first category might differ a lot between countries since it is a fact that some countries in a region or subregion consume more alcohol *on average* than others. For example, most of the comparative, descriptive statistics indicate that some countries in Eastern Europe, and particularly those around or adjacent to the Baltic Sea consume in average far more alcoholic beverages than their counterparts in Europe or elsewhere. These are not necessarily those which also have the higher prevalence of heavy episodic drinking, alcohol use disorders or related death rates from those.[5] Further in our research, we will try to claim that differences arise precisely because of the nature of the ecosystem in which consumption is embedded. Within such an ecosystem, and despite the lack of available and reliable data which comprises all countries, we suggest some core variables.

The notion of the importance of the environment or setting in which alcohol consumption takes place has been acknowledged by leading reports which perspective is certainly global. For example, it is worth noting that in a recent study the OECD concludes that, "Change in alcohol consumption is also influenced by determinants of alcohol consumption that are beyond policy actions [such as] genetics, demographics, personality traits, expectancies, family and peers and socioeconomic status. Environmental factors refer to social norms that shape drinking behaviors; the economic development of a country; and the availability and affordability of alcohol, which influence drinking patterns and outcomes over the life course" (OECD, 2021:118).

[4] At the end of this report a complete list of published research by *Mexican Community* is included.
[5] In Annex 1 we include some charts and maps for a rapid view of the universe of countries given a sample of some alcoholic consumption on alcoholic beverages and health impacts as well.

As some of the studies the Mexican Community has pointed out, affordability, in terms of the capacity of individuals and households to purchase alcoholic beverages is an important component for analysis.[6] One of the inputs that affects affordability is the presence of taxes. In this line of discussion, Torney, A. and colleagues find a link between harmful drinking and differences in the taxation of each beverage type and therefore recommend any taxation reform effort "should consider the impact of taxes on preferential beverage choice and associated harms". In their comparative study they find that "wine is consumed in greater proportions by risky drinkers in Australia when compared to New Zealand, and the same is true for ready-to-drink alcoholic beverages in New Zealand ... These findings highlight the importance of tailoring policies and prevention efforts targeting beverage specific consumption to the country or region where they are to be implemented." (2023: 6). This is of course a very specific analysis on two cases compared by their beverage preferences, but it underscores the importance of considering beverage preferences in terms of public health, that is, internalizing those preferences in any model, analytic of predictive. This is the purpose of this report.

Machine Learning Methods

Applied research which uses Machine Learning Methods (MLM) to understand alcohol use disorders (AUD) started in the decade of the nineties but it has become, until recently (2014), an emerging trend. Some studies use MLM because of the availability of databases which are worth exploring under these methods. In one of the most comprehensive reviews on the subject, Cresta Morgado and others, find less than 200 articles since 1995 which use MLM for addiction (Substance Use Disorders, or SUD) which is equivalent to only 0.25% of the articles on SUD. The evolution of articles of addiction research applying machine learning according to a PubMed search identify basic research, images, prediction, and text mining groups (2022: 261).[7] Most of the studies cannot be considered "massive data" in terms of comparability on an international-global scale, but they nevertheless are built on abundant data, which explains why artificial intelligence-type methods & tools are used. Data typically comes from hospital and other clinical records which belong to a specific community. MLM are built on small-scale but rather strong robustness and are normally focused on very specific research interests.
As an example of artificial intelligence and machine learning research applications, studies try to understand local, observable phenomena, with a more sophisticated approach. For example, Jonathan Jay uses

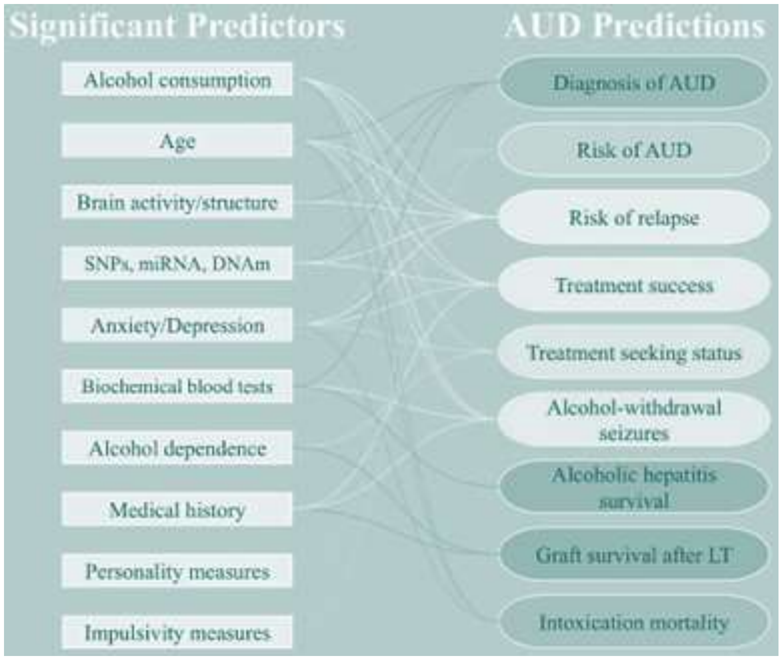
[6] See the Mexican Community paper published in 2024 (full citation at the reference section).
[7] SUD include Alcohol, caffeine, cannabis, phencyclidine, hallucinogens, inhalant opioid, sedative, hypnotic, or Anxiolytic Use Disorders.

data on alcohol outlets and firearms using satellite imagery and machine learning. Based on visual similarity, using high-resolution satellite imagery and machine learning the author concludes that alcohol outlets might systematically be located in certain kinds of environments, providing stronger evidence of a causal effect on nearby firearm violence. By matching places based on visual similarity, and advancement of imagery availability and machine learning the author provides a tool for epidemiologists to understand risks that vary across space (Jay, J.: 2019).

As Cresta Morgado, P., et. al., specify that, by 2020, approximately a third of all articles which use MLM to SUDs focus on prediction, another third to imagery and the last one to text mining. A small proportion focuses on basic research. In the following section specific alcohol use disorders, or AUD prediction is addressed according to the most recent literature (Cresta Morgado, P., et. al., 2022).

Prediction of AUD (Alcohol Use Disorders)

In a recent study, Hurtado, M. and others offer a systematic review on MLM and AUDs. Their results indicate that "after full screening, 70 articles were included in our review. Algorithms developed for AUD predictions utilize a wide variety of different data sources including electronic health records, genetic information, neuroimaging, social media, and psychometric data. Sixty-six of the included studies displayed a high or moderate risk of bias, largely due to a lack of external validation in algorithm development and missing data" (2022: 2). The type of data includes electronic health record (patient chart, vital signs, laboratory data, diagnosis and prescriptions); Genetic data (genome wide association study & polygenic risk score); neuroimaging data (electroencephalography, functional magnetic resonance imaging, and magnetic resonance imaging); polygenic risk score), and psychometric data (surveys, questionnaires, online factors). Data processing uses logistic regression, support vector machine, neural networks, decision trees, random forests, and natural language processing (2022: 7). Specific focus on AUD appears on diagnosis, prediction of severity, risk factors, future alcohol use, treatment outcomes and AUD in adolescents. The MLM studies identify a variety of relevant variables for AUD and AUD-related predictions.



Source: Hurtado, M., et. al., 2022:16.

Ebrahimi, A., et. al., provide a systematic literature review on research which aims to predict the risk of AUD using MLMs for the period between 2010 to 2021. The authors underscore the factors scientists had used to predict AUD. Among the most important are, "history of alcoholism in biological family members, psychological factors, such as level of stress, and personality disorder, behavioral factors, such as gambling problems, and social influences. Health records in hospitals also contain a substantial amount of information [which might be] even more precise and thereby helpful for staff medical decision-making" (2021: p.2). Authors find that the most used features are of demographic nature (age and gender), drinking patterns, school-related variables, psychological profile, and health-related aspects (2021: 10). After an initial screening of 2,355 journal and conference articles, 283 articles were included in the full-text screening phase. Only 12 articles were retained for dataset, machine learning techniques, performance metric and other related characteristics and 9 of them have as a source of data survey questionnaires (the other 4 use: electroencephalogram and genomic data, electronic health records, and magnetic resonance imaging.

Finally, Walters, S., et. al., use MLM to identify predictors of imminent drinking and create tailored messages for at-risk drinkers experiencing homelessness. They use a prediction algorithm (7 variable model) which predicts 80% of all drinking episodes where the 3 strongest predictors of imminent drinking are urged to drink, easy availability of alcohol or social pressure to drink. Messages are tailored to each participant’s current drinking goal (i.e., reduce drinking, stay sober, no drinking goal) and currently relevant drinking triggers. (2021: 7).

Gender Gap

Alcoholic beverages consumption is mostly a male activity. With a few exceptions at subnational level in some countries (Farfán, G. and Bruera, E., 2024), by far, the negative externalities of men alcoholic beverage consumption are higher than those in women. While it is a reality that a gender gap exists, the magnitude of such gap varies considerably among countries, regions, urban/rural areas, socioeconomic status, and other factors, such as background education. In some countries such as the United States, gaps in AUD between men and women are narrowing down. In others, this is in part a result that men are consuming less alcohol although it depends highly on age basis. For example, “among adolescents and emerging adults, narrowing gaps are being driven primarily by faster declines in alcohol use by males than females. Among adults, gaps are narrowing primarily because women are drinking more while men are either drinking less or maintaining their levels” (White, A., 2020: 11). In addition, as M. E. McCaul and others point out, “Women are generally smaller than men and have relatively less total body water and more total body fat. As a result, alcohol is more concentrated in a woman’s body; blood alcohol concentration rises faster and stays elevated longer in women than men. We also know that there are sex differences in brain anatomy, neurochemistry and function” (2019: 3). Ebrahimi, A. and colleagues “highly suggest that researchers consider gender disparities when building predictive models for the prediction of AUD” (2021: 12).

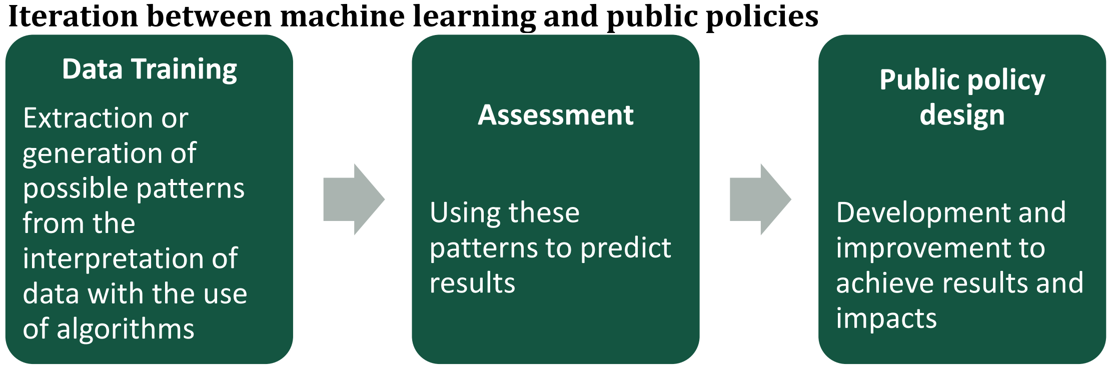
In a recent study for the 32 subnational governments (states) of Mexico, Farfán and Bruera explore data on economic, health and socioeconomic variables related to alcoholic beverages. They find substantial differences among men and women in terms of impact on health, household expenses, and violence. In general, both men have higher figures than women in terms of consumption, expenses and health impact. Nevertheless, disaggregation matters. In some states women might spend more on alcohol than men, considering both fermented (beer) and distilled (liquors) beverages even if the household head is a woman (Farfán, G. and Bruera, E., 2024). This is related to a variety of studies that try to understand how indirect taxation and gender equity interact in the context of alcoholic beverages consumption. Casale, M. D., for example finds that “female-type households across the distribution bear a lower burden of the excise taxes on alcohol and tobacco than male-type households. This finding is consistent with the idea that men have a greater preference for these goods, or that they have greater social permission to consume them, rather than the idea that these are luxury goods that female-type households cannot afford. These results confirm that even after controlling for expenditure class, there are gendered differences in spending patterns, underscoring the importance of an analysis of taxes from a gender perspective that goes beyond standard income/expenditure class comparisons" (Casale, 2012: 39-43).

Research Strategy

Policy inferences

Inference, as conclusions and reasoning reached based on evidence, and *policy*, generally defined as a group of measures to achieve specific goals, are both the main components of this section. By policy we use the public nature of it, which is the implementation by the public and/or private sectors to improve the living standards of people. *Policy inferences* are therefore generally defined here as directives obtained from processing data and machine learning iterations. As the diagram below depicts, data training and assessments were carried out by this study. These two are components for the design of *out of the box* or potential public policies and/or reform for improving the existing ones.

Iteration between machine learning and public policies



Source: Own elaboration.

Data training and assessments are key inputs to build policy inferences to trigger critical revisions and analyses of current policies. Policy inferences are listed in the following pages in terms of cluster analyses and machine learning probabilities.

Policy inference I. Cluster analyses

The first section of this report, *cluster analyses*, brings together global data to build a new characterization of the variables which determine, to a large extent, the harmful consumption of alcoholic beverages. In all, the cluster section offers a new mapping, literally, of the environment of potential harmful consumption. This exercise already offers new insights and probably questions others, in the way that regions that were often thought of as purely geographical or political, are now rethought and rebuilt based on evidence.

After this process countries do not belong anymore to their geographical region or subregion, but to a cluster where in terms of the chosen variable better fits in. This is already a new way to think and evaluate global policies which aim is to reduce the harmful consumption of alcohol beverages and therefore reduce their negative externalities.

In addition to the above elements, country regrouping considerably strengthens policy accuracy, in the way that country grouping also orders and categorizes variables within each cluster. We avoid policy diffusion, or policies which, because of the geographical initial bias, apply common policies which should not be the same. Therefore, we can identify, rank, and prioritize which policies should be considered a priority and, consequently, which policies have better chances or opportunities for improvement. This greatly helps resource allocation, whether human, financial, or institutional resources are in play. In sum, cluster analysis helps to allocate more efficiently resources to avoid waste but also, guarantee or at least increase the chances for a policy to deliver results, in terms of impact or incidence over negative externalities and shifting the status quo. Policy inferences for each cluster follow.

Beverage strength consumption. Is possible to identify a group of countries by beverage preferences and intensity. This by itself indicates the source of pure alcohol consumption and therefore the potential origin of the negative externalities of the harmful consumption of alcoholic beverages by their type. Clusters indicate extreme cases in which beer is in practical terms the predominant source of alcohol while in others are distilled beverages. These general policy inferences can help redirect and improve the policy dialog at global, regional, national or the domestic environment in terms of stakeholder's input.

Alcohol disorders in females. Clusters denote a high heterogeneity, since magnitudes show an important contrast among each other. This is a useful method to characterize clusters and prioritize policies in terms of sex.

Alcohol disorders in males. All clusters denote different number of cases, so this implies policy replication is not feasible among clusters. Sex differentiation helps policy specification since females and males are clustered differently in terms of the type of variable and magnitude.

Death rate in females. Death rate in females caused by alcohol use disorders vary significantly among clusters and do not necessarily correspond to alcohol disorders as such. This means that disorders and death do not necessarily go in hand and different policy responses should be put in place.

Death rate in males. As expected, death rate in males is higher than females. Based on the dispersion type chart, cluster is probably what shows more homogeneity and relatively simple policy prioritization.

Road injury death rate. Clusters show very similar patterns therefore global and relatively homogeneous applied policies should be efficient.

Policy inference II. Variable Importance and Probabilities

Random forest aggregates the results of the individual trees in a manner where the observation is classified into the class for which most of the trees agree. In this same manner, it computes the probability of each class by estimating the proportion of times an observation is classified into each of the classes. The probability of belonging to one of the clusters analyzed according to the independent variables is then analyzed.[8]

Policy Analysis is built on the following criteria:

1. *Policy concentration vs policy diffusion*: to what extent clusters indicate similar magnitudes of probabilities or there are few or just one probability that outpaces all. Similar probabilities indicate opportunities for similar policies (likewise, different probabilities indicate need for different policies).

2. *Policy indifference vs policy monopoly.* Clusters indicate very low or negligible probabilities which turn them *indifferent* to any other policy intervention. The opposite case is policy monopoly or predominance. Clusters with minimal probabilities indicate there is no need to "waste time" focusing on them (likewise, very high probabilities indicate urgent need to address).

[8] Annex 6 contains the charts of all countries displayed in order by probabilities.

3. *Policy irrelevance and strategy.* Policy irrelevance is determined in cases where there are varied and similar probabilities but others which are almost inexistent. This is a key aspect of policy strategy in terms of process. If two or more clusters are very similar in terms of probabilities then policy is not needed but to pursuit/improve the same policies that are being in place. Strategy then becomes key in terms of policy coherence and consistency.

The following table synthesizes a practical way to read clustering and probability charts in terms of policies.

Criteria	Description	Output	Outcome
Policy concentration vs policy diffusion	Similar horizontal color bars indicate similar probabilities and therefore policy problems	Policy replication possible	Best practices, policy transfer.
	Different horizontal color bars indicate different probabilities and therefore policy problems	Policy replication not possible	Tailored, specific policies
Policy indifference vs policy monopoly	Minimal horizontal color bars indicate negligible probabilities and therefore no case for a policy	Policy resources not needed	No action
	Long horizontal color bars indicate high probability and therefore policy priority/urgency	Policy resources needed	Immediate action
Policy irrelevance and strategy	All horizontal color bars have very similar sizes among clusters, therefore no need for action	Policy status quo reinforced	Policy impact evaluation
	Some horizontal color bars have very similar sizes among clusters, need to focus on different with high probabilities	Policy prioritization	Tailored, specific policies

Policy departure (where to start) and *landing policies* (what to leave for the future). This indicates a policy choice that only a more in-depth analysis by the expert or country case researcher can determine. The entry point can be x clusters or y clusters. Policy departure is only identified when a deeper analysis can be made. It is related to the magnitude of available resources for implementing policies. *Policy Prioritization.* Prioritization occurs when probabilities are compared and ranked but, as all policies have limits and even political economic considerations, risk analysis should be performed. Prioritization might entail political economy as well as risk-type analyses.

Concept, theory, and methods

Statistical learning has emerged as a new subfield in statistics, focused on supervised and unsupervised modeling and prediction. Rattan, P., and others define machine learning as “a specific set of techniques within AI that are predicated on “learning” to model patterns in data using mathematical functions … ML (machine learning) diverges from classical statistics with its ability to use higher-dimensional mathematical operations on much larger data sets to decipher complex, nonlinear relationships" (Rattan, P., et. al., 2022: 70). In the opinion of Cunillera, T. and Guilera, G., ML "lies on the intersection of computer science and statistics. It considers the study of human and animal learning in psychology, neuroscience and related fields with the defining goal of creating computer systems that automatically improve with experience … Within artificial intelligence and over the last decades, the study of ML has grown into a broad discipline that has produced fundamental statistical-computational theories of learning processes, motivating the creation of algorithms that are extensively used in commercial software for computer vision, speech recognition, natural language processing, robot control, and other applications as data mining used to discover hidden regularities in the growing volumes of data—big data—(Cunillera, T. and Guilera, G., 2018: 2).

Unsupervised Learning

Statistical learning can derive from supervised or unsupervised models. Both are tools to learn relationships and structure from data. In a supervised approach there is the need to fit a model that relates the response to the predictors, with the aim of accurately predicting the response for future observations (prediction) or better understanding the relationship between the response and the predictors (inference).

By contrast, unsupervised learning describes the somewhat more challenging situation in which for every observation n, we observe i = 1, . . . a vector of measurements but no associated response yi. It is not possible to fit a linear regression model, since there is no response variable to predict. In this setting, we are in some sense working blind (James, G., et. al., 2023: 26). Supervised methods use regression to understand how some features of a x number of observations can predict and outcome. By contrast, unsupervised learning is a set of statistical tools intended for the setting in which we have only a set of features measured on observations. The goal is to discover interesting things about the measurements There are two types of unsupervised learning: principal components analysis for the purpose of data visualization or data pre-processing before supervised techniques are applied, and clustering, a broad class of methods for discovering unknown subgroups in data (James, G., et. al., 2023: 495).

Hierarchical Clustering

Clustering or cluster analysis is a set of statistical learning methods to make groups with observations that share similar characteristics.

These techniques find feature patterns, which are given by a set of related values of the variables analyzed. As Cresta Morgado, P., et. al., state, "Clustering techniques aim to group elements in a dataset, so objects within a group are more similar than those in other groups. For this purpose, the method uses distances (e.g., Euclidean, cosine) between vectors representing objects. In this representation, each dimension in the vectors corresponds to one of the variables or features describing objects. After groups have been found, a domain expert must analyze the groups obtained to interpret the latent causes governing the clustering solution" (Cresta Morgado, et. al., 2022: 267).

On the other hand, in hierarchical clustering, we do not know in advance how many clusters we want; in fact, we end up with a tree-like visual representation of the observations, a tree-based representation of the observations called a dendrogram, that allows us to view at once the clustering obtained for dendrogram each possible number of clusters (James, G., et. al., 2023: 515).

Variable Importance

Variable importance is computed over an out of bag (OOB) sample. The first estimates, that is, to select the optimal number of lags and perform clustering, are made with independent databases, which have the following information:

- *Consumption*: In this base, consumption is separated by volume of alcoholic content. This database has information for 160 countries on an annual basis from 2000 to 2021, resulting in a total of 3,520 observations for 4 variables (Beer, Distillates, Ferments and NoLo's).

- *Alcohol Use Disorders in Females*: There is a base of 206 with a time span from 2000 to 2019, resulting in a total of 6,120 observations.

- *Alcohol Use Disorders in Males*: There is a base of 206 with a time span from 2000 to 2019, resulting in a total of 6,120 observations.

- *Death Rate (alcohol related) in Males*: There is a base of 206 territories/countries/regions, with annual data from 1990 to 2019, resulting in a total of 6,120 observations.[9]

[9] Regarding countries, territories not classified as a country as such are also included, such as the "North Mariana Islands" or "Niue" for example. That's why it exceeds the "official" number of countries.

- *Death Rate (alcohol related) in Females*: It has a base of 206 countries, with annual data from 1990 to 2019, resulting in a total of 6,120 observations.

- *Road Injuries*: database includes 190 countries. The base goes from 1990 to 2019 with an annual periodicity resulting in 3800 observations.

The first database used was the consumption of alcoholic beverages in terms of pure Alcohol Beverage Volume) in 000's 9L Cases published by the IWSR.[10] This database has information for 160 countries with a time span ranging from 2000 to 2021. In order to carry out the exercise, the original measurement was transformed into liters and the percentage share of each alcohol content per year for each country was subsequently calculated.

The second databases consisted of alcohol disorders as such and the rate of deaths attributed to alcohol abuse by gender (Death Rates from Alcohol use Disorders by Gender) obtained from the Institute for Health Metrics and Evaluation, IHME. It has an annual periodicity that goes from 1990 to 2019 and covers 204 countries.

Once the clustering has been carried out, we are left with one observation per country for each base that corresponds to a categorical variable of the cluster in which said country has remained. Subsequently, the names of the countries are unified and the bases are joined to obtain a single base with 129 countries and 14 variables: The categorical clusters and the control variables (% of the male and female population and life expectancy at birth by sex). With this basis the probabilities are estimated.

[10] Compiled by the International Alliance for Responsible Drinking (IARD) and available with no charge at: https://iard.org/science-resources/detail/Alcohol-Consumption-by-Country-and-Beverage-Streng

The Machine Learning Model

Phase 1. Country Grouping based on types of Alcohol Consumption strength: We aim to identify distinct country clusters associated with alcohol consumption and related elements. Employing unsupervised learning techniques, specifically hierarchical clustering, we group countries based on a defined set of variables.

Phase 2. Association between Country Groups and Alcohol-Related Disorders: We analyse the relationship between these identified country clusters and alcohol-related disorders and alcohol-related deaths. To achieve this, we employ supervised learning techniques, specifically regression algorithms. This allows us to assess the impact of different variables on our targeted prediction and understand how these predictions vary concerning various factors of interest.

In broad terms, ML encompasses two primary domains. Supervised learning involves discerning patterns from existing data and extrapolating these to *predict* outcomes in new, unseen data. In a dataset with n observations, consisting of both input (X_i) and output (y_i) variables, our focus lies in uncovering the associations between these inputs and outputs

$$y_i = f(X_i) + \varepsilon_i$$

Where $f(.)$ can take any functional form. The functional form is learned by the model by minimizing a loss function

$$L(y, \hat{y})$$

That indicates how well or bad the model is predicting. For classification analysis a commonly used loss function is the accuracy of the prediction

$$L(y, \hat{y}) = \frac{1}{n}\sum_{i=1}^{n} 1(\hat{y}_i = yi)$$

Through iterative adjustments to the algorithm parameters, the process identifies rules that minimize prediction errors. Typically, this procedure follows a k-fold cross-validation approach, continuously assessing the model's performance within a holdout (or out-of-sample) fold of the data.

For the estimation of $f(.)$, we use decision trees and random forests. Decision trees offer a visual representation advantage. They partition observations into distinct, non-overlapping groups using a recursive binary splitting process. Predictions are derived by taking the mean within each group.

In each iteration of the algorithm, we choose an input j and a cut point s to bifurcate observations into two new regions. These steps continue iteratively until a specified condition, determined by the model's parameters, is satisfied.

However, decision trees often exhibit high variance. This variance implies that if trained on a different dataset, the tree's structure might vary significantly. To mitigate this, we turn to random forests. In this approach, m trees are built similarly to decision trees, but at each splitting point, a subset of both observations and inputs is selected. Finally, the average of all trees is computed for each observation, resulting in more accurate predictions.

Unsupervised learning tackles datasets lacking explicit outputs but containing multiple inputs. For each observation i, we possess a set of inputs p without any corresponding output. The objective revolves around identifying similarities among data points and aiming to diminish the dimensionality of the p inputs into k subgroups of the data for example.

During cluster analysis or data segmentation, the primary aim is to categorize observations into clusters wherein the items within each cluster exhibit greater proximity to one another compared to those classified in separate clusters. This classification hinges upon gauging (dis)similarities or distances between instances of the j_{th}input.

$$D(x_i, x_{i'}) = \sum_{j=1}^{p} d_j(x_{ij}, x_{i'j})$$

The most used distance metric is the *Euclidean distance*

$$d_j = \sqrt{\sum_{j=1}^{p} (x_{ij-} x_{i'j})^2}$$

We opt for hierarchical clustering, an agglomerative technique that merges observations based on a minimal distance metric. This method organizes observations into a dendrogram, presenting a tree-like structure.

Initially, when no clusters exist, the algorithm clusters the two closest observations based on the Euclidean distance. Subsequently, as observations accumulate into clusters, the distance between a new observation and existing clusters is calculated using the Ward linkage method. This method assesses the distance between clusters A and B by considering their centroids and sizes to minimize the increase in total within-cluster variance upon merging

$$IV = \frac{|A| * |B|}{|A| + |B|} * ||c_A - c_B||^2$$

Merging clusters involves considering their relative sizes and the Euclidean distance between their centroids. Ultimately, determining the optimal number of clusters is achieved by minimizing the within-cluster sum of squared variances. This approach aims to identify the ideal number of clusters that effectively minimize the variance within each cluster while maximizing the differentiation between clusters.

$$WCSS(C_i) = \sum_{j=1}^{n_i} ||x_j - c_i||^2$$
$$\arg\min_k \sum_{i=1}^{k} WCSS(C_i)$$

K is chosen using the elbow method, looking for the value where the decrease in WCSS is marginal.

In the following section, 12 Dendrograms will be displayed along with a table with descriptive statistics of each one and their clusters followed by maps and dispersion charts which identify clusters with letters and distinctive colors. Dendrograms correspond to the databases and indicators described earlier in this report ("Knowledge databases rationale"). These dendrograms were built with databases which all include 160 countries. It should be noted that the clusterization process led to the construction of another database, a replica in terms of variables but reduced to 137 countries (discarding countries with a variety of missing values expressed by blank areas, N/A, etc). The resulting databases are then used to use ML for prediction/probabilities whose results are included in the corresponding section (Second Section: Machine Learning

Although in many Machine Learning (ML) models we cannot claim direct causality among explanatory variables on the dependent variable, these models are very useful and often better approaches commonly used to find a relation or association between economic variables and behavioral patterns. of economic agents. The present ML analysis is structured in two phases to address specific inquiries:
).

First section: Cluster Analyses

Cluster dendrograms & indicators

Cluster results and visualizations

A number of clusters are proposed as part of the research strategy. The clusters are the following:

1. Beverage strength (in content of pure alcohol).[11] *Categories are grouped in 4 categories as follows:*

- < 4,5% (NoLos, No and Low Alcohol content beverages).
- = 4,5% (Beer).
- 4,5 and < 30% Medium Alcohol Content (fermented and other alcohol content fortified beverages).
- = 30 - 40% High Alcohol Content (Distilled, liquor, spirit type beverages).

2. Alcohol Use Disorders (per 100 people). Current number of cases of alcohol use disorders per 100 people, in males and females aged age-standardized - Alcohol use disorders - Sex: Male and Females - Age: All Ages (Rate).[12]

3. Death Rate associated with alcohol (Death rate per 100,000 people). Deaths – Cause: Alcohol use disorders - Sex: Male and Females - Age: All Ages (Rate).

4. Road Injury (Death rate per 100,000 people,). Death rate per 100,000 people - Cause: Road injury - Sex: Both - Age: ALL Ages.

For clusters 2 and 3, men and women are clustered separately which results in a total of 6 clusters. We provide 4 visualizations which derive from each cluster: dendrograms, synthesis tables, maps, and dispersion graphs (for policy prioritization).[13]

Beverage by strength (in terms of pure alcohol)

Consumption preferences according to the type of alcoholic beverage vary enormously between countries (Chart 1). Mexico is by far the country that has the greatest preference for beer consumption, while in Saudi Arabia it is zero. The cases of high per capita wine consumption in Europe such as Italy, Portugal, Austria, or Argentina and Chile in the context of Latin America stand out when it comes to the consumption of pure alcohol while countries surrounding or close to the Baltic Sea prefer distilled alcoholic beverages.

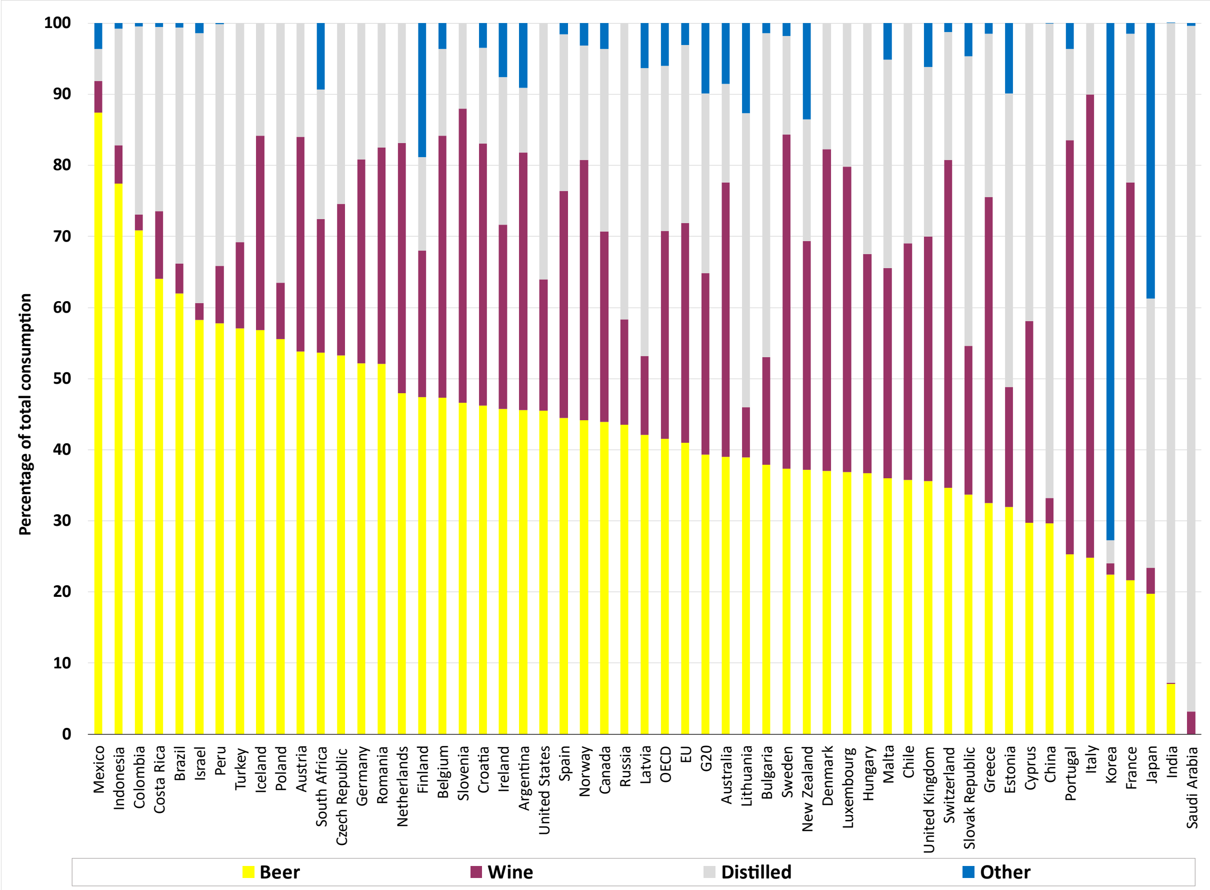
---

[11] For example, Rehm, Jurgen, et. al., (2019) states that, "More research is necessary on the differential effects of level of use and potency on health and other harm, but current regulatory policies for substance use and gambling are often not in line with currently available evidence of harm."

[12] Annex 7 includes a clusterization using Disability Adjusted Life Years DALYs as other variable related with negative externalities. It was discarded and not included in our model since we chose disorders and deaths as more robust.

[13] Dendrograms and the corresponding maps are included in Annex 2 at the end of this document.

Chart 1. Pure alcohol consumption by type of beverage (n=55, %).

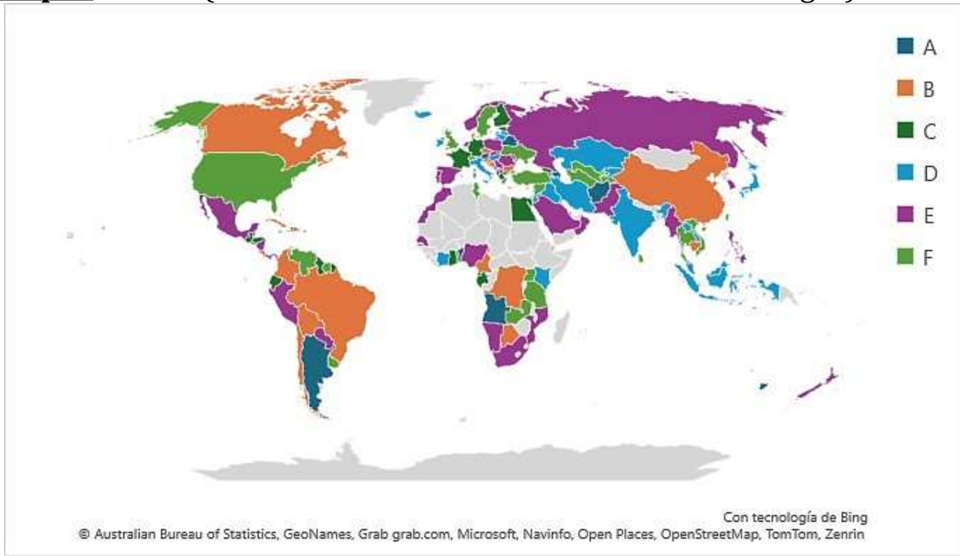


Source: *Preventing Harmful Alcohol Use*, OCDE, 2021.

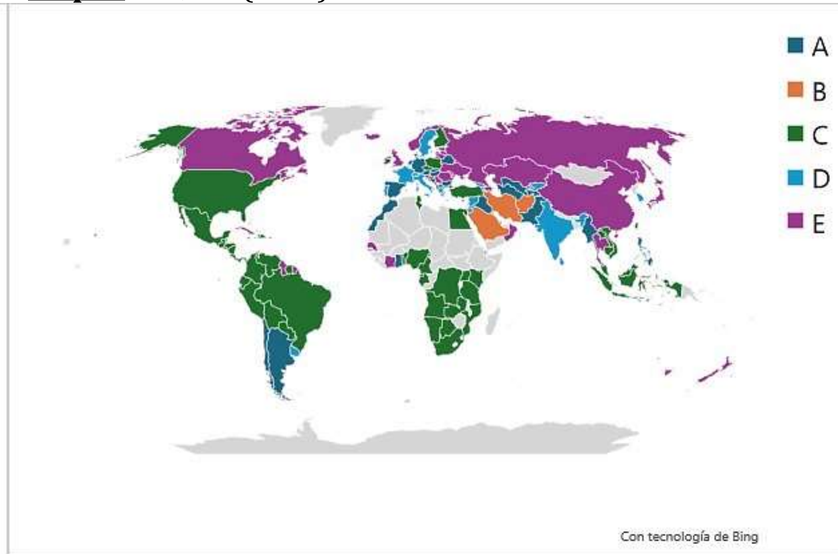
As Taylor, A. W. correctly points out, "… alcohol consumption patterns within a community are a reflection of that location's political structure, laws and regulations, societal norms and traditions, dominance of anti-alcohol religious convictions, average income levels and economic standings, as well as motives, behaviors and beliefs" (Taylor, A. W., et. al., 2019: 28).

In the following section, we build country-based clusters using beverage preference (type of beverage by strength) criterion in terms of % of pure alcohol content. We first portray four maps which depict the resulting number of clusters by type (strength or % of pure alcohol content). Notice that clusters do not necessarily belong to traditional country groupings (by continent, region, subregion, etc). In addition, these are followed by three tables which synthesize and show the composition of each cluster by their types of beverage preferences (similar to Chart 1 but now organized by clusters and including a "no or low alcohol content beverages, or NoLos instead of Table 1 category "Other")

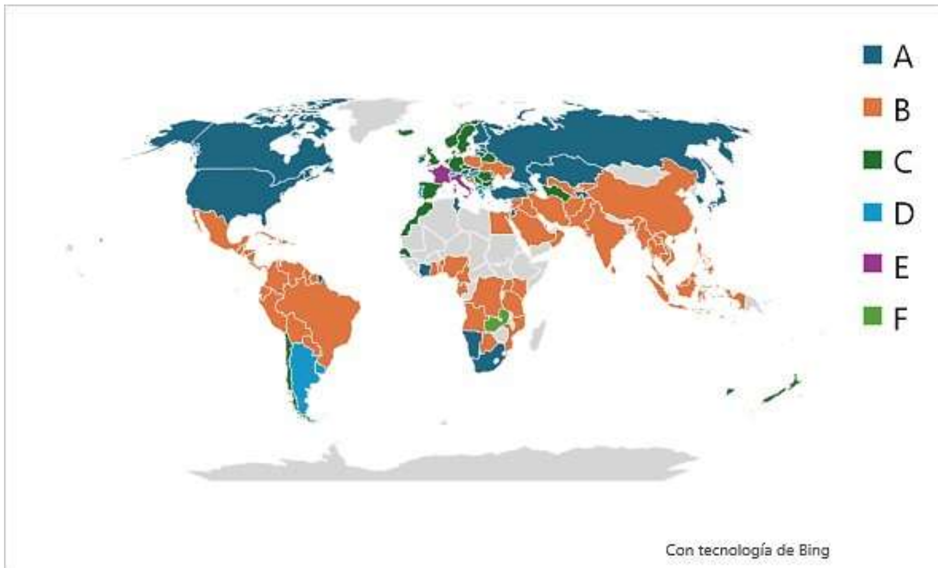
Map 1. < 4,5% (NoLos, No and Low Alcohol content beverages).



Map 2. = 4,5% (Beer).



Map 3. 4,5 and < 30% Medium Alcohol Content (fermented and other alcohol content fortified beverages).



Map 4. = 30 - 40% High Alcohol Content (Distilled, liquor, spirit type beverages).

Table 1.

Clusters descriptive analytics (# of countries)

Category Content (% of pure alcohol)	NoLos <4,5%	Beer 4,5%	Fermented 4,5 - 30%	Distilled >30%	Category Content (% of pure alcohol)
1. Cluster A (# of countries)	19	38	44	17	1. Cluster A (# of countries)
2. Cluster B (# of countries)	26	5	67	3	2. Cluster B (# of countries)
3. Cluster C (# of countries)	19	54	31	54	3. Cluster C (# of countries)
4. Cluster D (# of countries)	27	18	12	68	4. Cluster D (# of countries)
5. Cluster E (# of countries)	38	42	2	5	5. Cluster E (# of countries)
6. Cluster F (# of countries)	28		1	2	6. Cluster F (# of countries)
7. Cluster G (# of countries)				7	7. Cluster G (# of countries)
8. Cluster H (# of countries)				1	8. Cluster H (# of countries)
Total of countries:			157		

This table shows a relatively homogeneous participation of NoLos in all clusters. Beer is concentrated in Cluster 1A, 3C, and 5E. Cluster 2B in fermented is by far the one which contains the highest number of countries (followed by 1A & 3C), a similar figure for 4D for distilled beverages (followed by 3C clusters.

Table 2.

Clusters descriptive analytics (percent of participation)

Category Content (% of pure alcohol)	NoLos <4,5%	Beer 4,5%	Fermented 4,5 - 30%	Distilled >30%	Category Content (% of pure alcohol)
1. Cluster A (% of total)	12.1	24.2	28.0	10.8	1. Cluster A (% of total)
2. Cluster B (% of total)	16.6	3.2	42.7	1.9	2. Cluster B (% of total)
3. Cluster C (% of total)	12.1	34.4	19.7	34.4	3. Cluster C (% of total)
4. Cluster D (% of total)	17.2	11.5	7.6	43.3	4. Cluster D (% of total)
5. Cluster E (% of total)	24.2	26.8	1.3	3.2	5. Cluster E (% of total)
6. Cluster F (% of total)	17.8		0.6	1.3	6. Cluster F (% of total)
7. Cluster G (% of total)				4.5	7. Cluster G (% of total)
8. Cluster H (% of total)				0.6	8. Cluster H (% of total)
Total of percent:			100		

In addition to table 1 which takes the number of countries as the basis of analysis, table 2 presents such figures in terms of percentage of participation for policy prioritization policy on a global scale. For example, a higher number of countries indicate a larger impact on the policies revised or improved, largely because of proposing similar policy measures to similar policy challenges.

Table 3.

Clusters descriptive analytics (average % of pure alcohol consumption)

Category Content (% of pure alcohol)	NoLos <4,5%	Beer 4,5%	Fermented 4,5 - 30%	Distilled >30%	Category Content (% of pure alcohol)
1. Cluster A (% of total)	0.4%	67.1%	11.8%	19.8%	1. Cluster A (% of total)
2. Cluster B (% of total)	1.2%	8.1%	3.5%	69.0%	2. Cluster B (% of total)
3. Cluster C (% of total)	0.4%	89.5%	20.9%	10.8%	3. Cluster C (% of total)
4. Cluster D (% of total)	0.7%	49.2%	35.2%	3.3%	4. Cluster D (% of total)
5. Cluster E (% of total)	1.5%	77.0%	55.2%	47.7%	5. Cluster E (% of total)
6. Cluster F (% of total)	0.8%		13.3%	99.4%	6. Cluster F (% of total)
7. Cluster G (% of total)				31.8%	7. Cluster G (% of total)
8. Cluster H (% of total)				54.6%	8. Cluster H (% of total)
Total of percent:			100		

In addition to table 1 & 2, this table presents such figures in terms of percentage of pure alcohol consumption for policy prioritization policy on a global scale in terms of beverage preference and type as a source of pure alcohol. From a global, comparative perspective these figures are strong indicators of pure alcohol consumption, regardless or their source in terms of type of beverage.

As we can appreciate in tables 1, 2, and 3, participation by type of beverage, regardless of the indicator, is quite low in NoLos and more important in distilled beverage, a result relatively expected because of alcohol strength. Nevertheless, variations among clusters indicate that the source of pure alcohol comes from very different types of beverages. The extreme case is Cluster 6F which indicates that for those countries distilled beverages is the only source of pure alcohol (though with small participation in terms of number of countries). There are clusters with high participation and number of countries (3C in Tables 1 and 2) have in terms of pure alcohol source, beer as predominant. Cluster 5E has beer, fermented and distilled beverages as almost equal sources of pure alcohol (see table 3).

In general, policy priorities can diverge or converge in terms of the scope of countries or the magnitude of pure alcohol consumption. There are other clusters which deserve a unique approach and might have difficulty in sharing or borrowing best practices. Clusters with common percentages of participation by type of beverage should be more able to share "best practices" or build upon a common agenda (prevention, treatment, regulation, taxation, etc). Participation of the type of beverage in pure alcohol consumption signals the origin of the source and therefore specific targeted policies should be implemented.

Clustering results

In the following pages, clusters are displayed in terms of dendrograms, maps and tables.[14] As for number of female cases of alcohol use disorders per 100 people, there is a very marked territorial division. Cluster 1 shows North Africa, Arab and Asian countries with the lowest number of cases per 100 inhabitants. Cluster 2 (the second highest with an average of 1.79 cases per 100 inhabitants) includes the USA, Russia and Eastern Europe. The third cluster is composed of Mexico and Central America as well as Central and South Africa. These have the second lowest average of all clusters. The 4th Cluster contains England, France and Germany as well as Australia among others. Finally, the fifth cluster contains only Greenland with the highest average (2.97 cases per 100 inhabitants).

[14] , A complete list of countries identified by cluster can be found in Annex 4 and dendrograms, along with their corresponding maps in Annex 3 of this document.

Map5. Female alcohol disorders per 100 people.

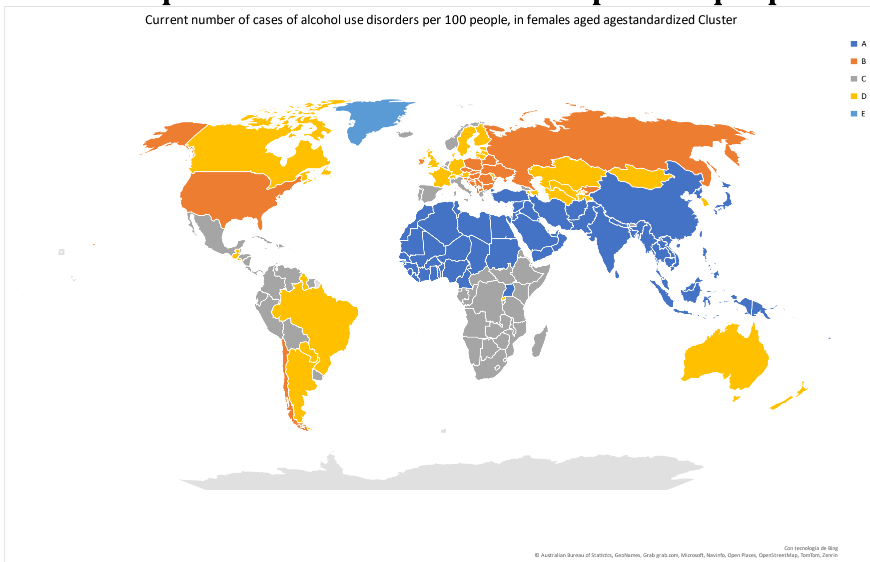


Table 4: Female alcohol disorders per 100 people.

Cluster	# of cases
1 – A	0.34
2 – B	1.79
3 – C	0.77
4 – D	1.24
5 - E	2.97

The following map contains # of male cases where the 1st cluster has a mean of 0.63 cases (North African countries and some Muslim countries of the Arabian Peninsula and some Southeast Asian countries) and the 2nd cluster is composed of middle African countries and three Latin American countries (Peru, Ecuador and Bolivia, with an average of 1.39 cases per 100 people.

Map 6. Male alcohol disorders per 100 people.

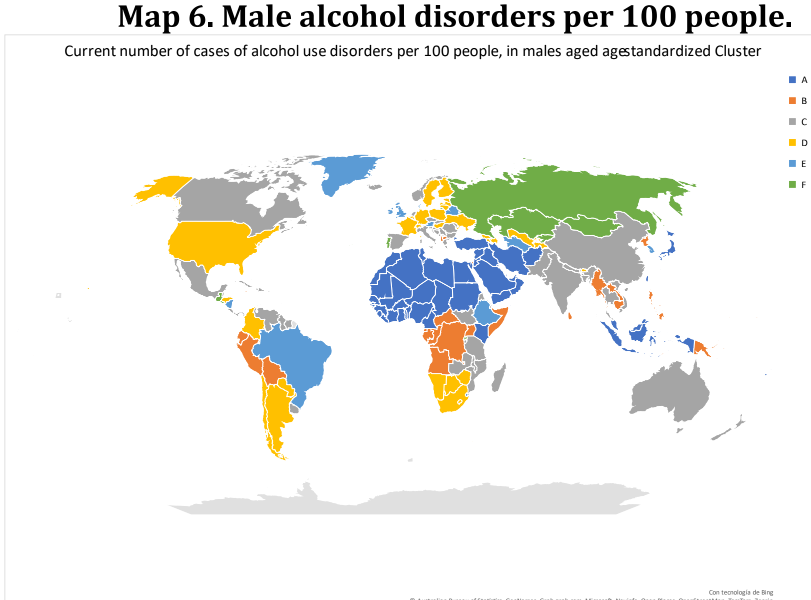


Table 5: Male alcohol disorders per 100 people.

Cluster	# of cases
1 – A	0.632
2 – B	1.396
3 – C	2.362
4 – D	3.487
5 - E	4.347
6 – F	5.756

Cluster 3 is composed of Mexico, Canada, Uruguay and the Guyanas in America, European countries and the large Asian countries such as China and India among others. Cluster 4 comprises the USA as well as Northern Europe and Southern Africa. Finally, cluster 5, with the highest number of cases per 100 inhabitants, is composed of Russia, Mongolia, Kyrgyzstan and Portugal in continental Europe, together with Scotland. On the American side are El Salvador and Guatemala.

Finally, regarding the death rate associated with alcohol consumption disorders for women, Belarus stands out, the only member of this group. In second place (cluster 6) are Eastern European countries (Estonia, Ukraine, Latvia and Lithuania).

Map 7. Females Death Rate (per 100,000 people, all ages).

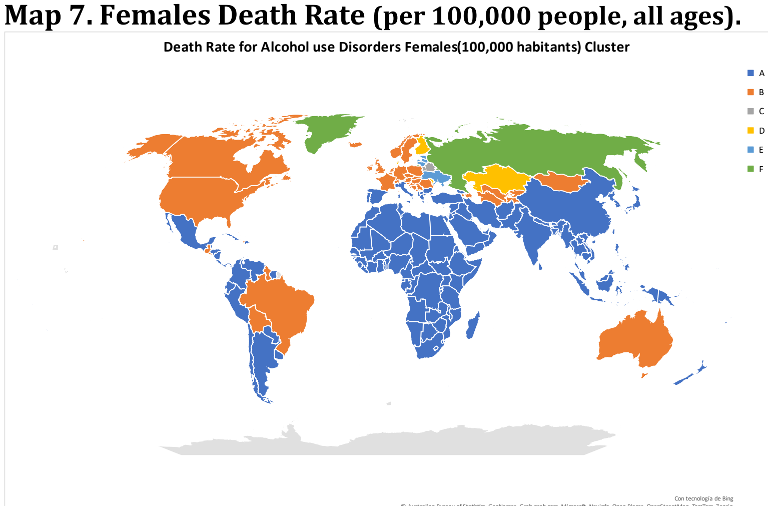


Table 6: Female death rate (per 100,000 people, all ages).

Cluster	Death rate Females
1 – A	0.22
2 – B	1.62
3 – C	13.32
4 – D	5.15
5 - E	7.79
6 - F	11.86

The same exercise is repeated, but now for men, in contrast to the death rate for women, the values for men are much higher. Once again, the Eastern European countries (Belarus, Estonia, Ukraine and Russia) are grouped with the highest average, followed by cluster 5 where there are 2 Central American countries (El Salvador and Guatemala) and two Eastern European countries (Lithuania and Latvia), Greenland and in the Caribbean Saint Kitts and Nevis.

Map 8. Male Death Rate (per 100,000 people, all ages).

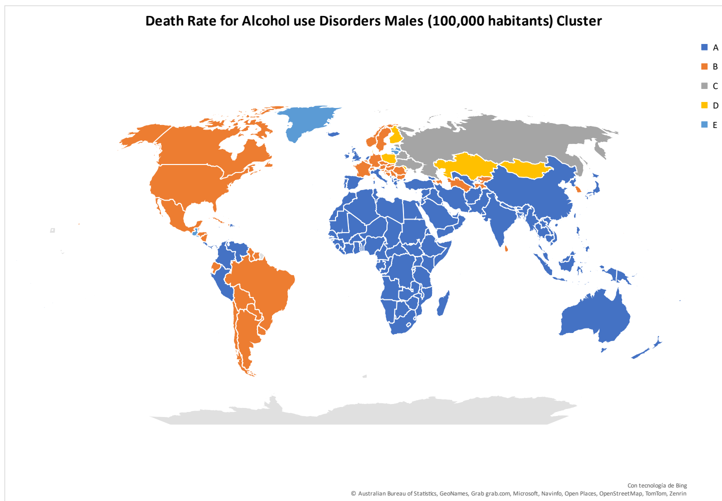


Table 7: Male Death Rate (per 100,000 people, all ages).

Cluster	Death rate Males
1 – A	1.38
2 – B	6.79
3 – C	38.66
4 – D	18.07
5 - E	26.14

In the rate of deaths due to traffic accidents, Canada, Europe and Australia have the lowest average (6.84 per 100,000 inhabitants). The next group of countries with more accidents is North America except for Canada, Central America and Peru, Chile, Argentina and Uruguay for South America with an average of 13.51 accidents per 100,000 inhabitants. Russia, China and former members of the USSR have an average of 19.60 in this same cluster for the Americas are Brazil, Bolivia, Paraguay, Colombia and Ecuador.

Map 9. Road injury (Death rate per 100,000 people, both sexes, all Ages).

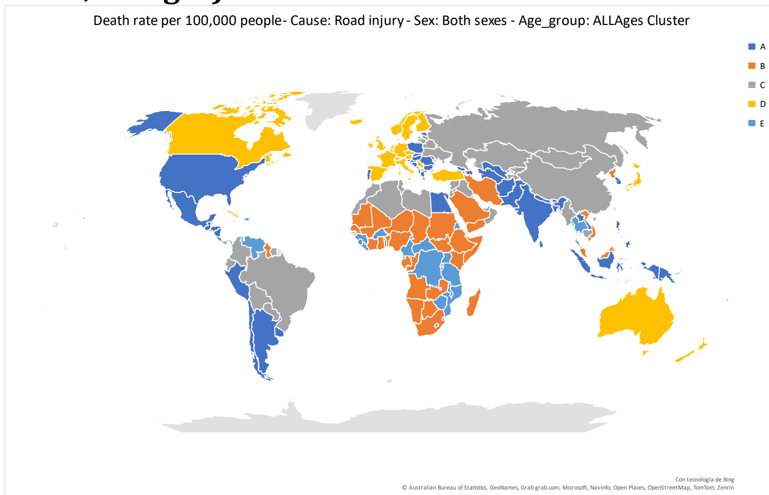


Table 8: Road injury (Death rate per 100,000 people, both sexes, all Ages).

Cluster	Road injury death rate
1 – A	13.51
2 – B	25.75
3 – C	19.60
4 – D	6.84
5 - E	32.37

Cluster 2 (the second highest in terms of accidents) includes African countries and the Arabian Peninsula. Finally, cluster 5, with the highest number of accidents, includes countries in Middle Africa and Venezuela in the Americas.

Second section: variable importance and probabilities

Variable importance

Variable importance is measured by permutating each variable one at a time and estimating the effect that including it has over the accuracy of the model. Since the values may range indefinitely, they are normalized so that the variable with the higher importance takes the value of 1. The rest are then interpreted as the relative contribution relative to the most important variable. Variable importance is computed over an out of bag (OOB) sample.[15]

In this section we assess the importance of independent variables in the determination of each dependent variable. The data set is composed of the previous clustering results for each country. As a result, we have a database composed of categorical variables for each variable according to the cluster in which they were grouped. We were able to build a complete database for **127 countries**. Those that did not have observations in one or more variables were discarded, so the number of clusters per variable may change from the original ones. In the case of Alcohol Use Disorders for Females we also lose the last cluster in which only Greenland was located, since it included missing values for another variable and had to be eliminated. The same happens for the Death Rate due to alcohol use disorders in Females, where we lose two clusters (5 and 6) due to the missing values for Greenland, Russian Federation, Estonia, Latvia and Lithuania that comprise these clusters.

Table 9. Cluster transition (variable importance, VI).

Cluster	Number of Clusters (Individual Dataset)	Number of Clusters (Merged Dataset)	Missing Cluster	Country With NA
Beer Consumption	5	5		
Spirits Consumption	8	8		
Fermented Beverages Consumption	6	6		
NoLo's Consumption	6	6		
Alcohol use disorders per 100 people, in males (aged age-standardized)	6	6		
Alcohol use disorders per 100 people, in females(aged age-standardized)	5	4	#5	Greenland *
Death Rate Alcohol use disorders females (per 100,000 people, all ages)	6	6		
Death Rate Alcohol use disorders males (per 100,000 people, all ages)	5	5		
Road injury (Death rate per 100,000 people, both sexes, all Ages)	5	5		

* NA in Consumption by ABV

The resulting variables are (# of clusters):

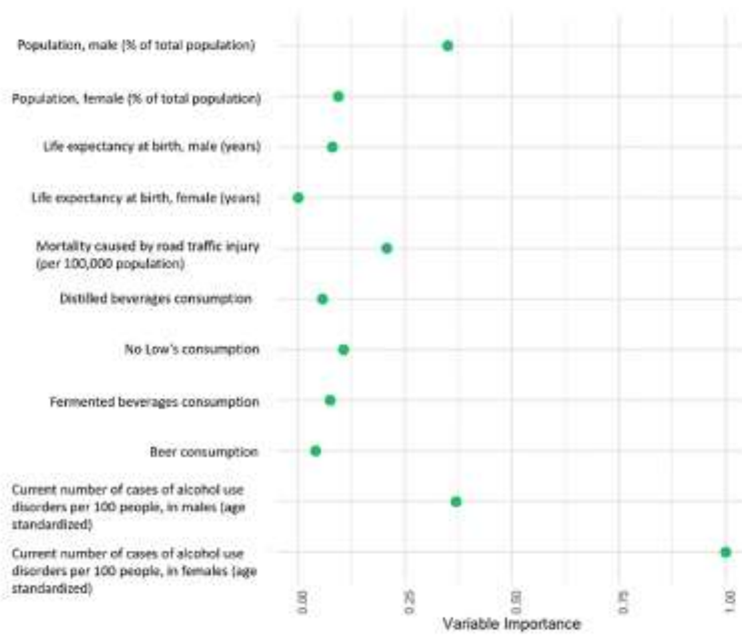
[15] Annex 5 contains the optimal number of Clusters Elbow Graphs.

1. Beer Consumption (1 – 5)
2. Spirits Consumption (1 – 8)
3. Fermented Beverages Consumption (1 – 6)
4. NoLo's Consumption (1 – 6)
2. Males Alcohol use disorder (1- 6)
3. Female Alcohol use disorder (1 - 4)
4. Female Death Rate in Alcohol Use Disorders (1 - 6)
5. Male Death Rate in Alcohol Use Disorders (1 - 5)
6. Cluster of Road Injury (1 - 5)

Additionally, control variables such as the percentage of males and females by country for the most recent year and Life expectancy at birth (years) for males and females for the most recent year were added.

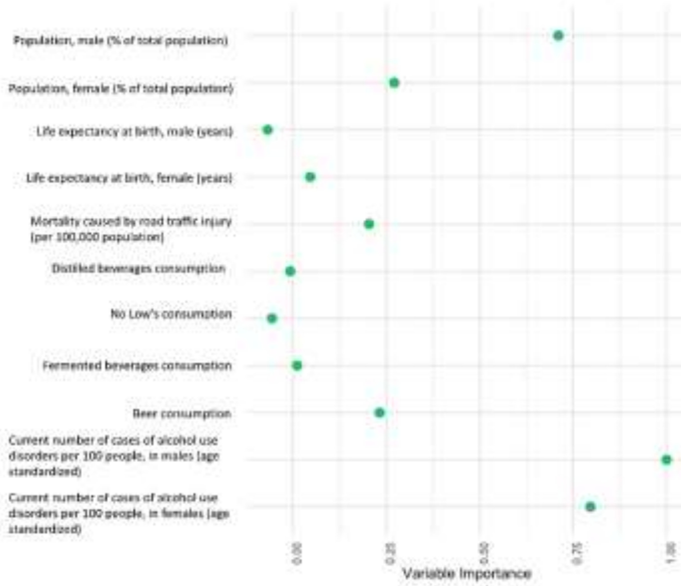
Estimation for the following models was made: 1. Death Rate due to alcohol use disorders in Females; 2. Death Rate due to alcohol use disorders in Males; and finally, 5. Road Injuries. We compute the importance of each independent variable in the determination of our dependent variable. In the charts below we can observe the importance of each variable. The values are normalized so the higher one is always 1.

Chart 2. V.I.: Death Rate Females



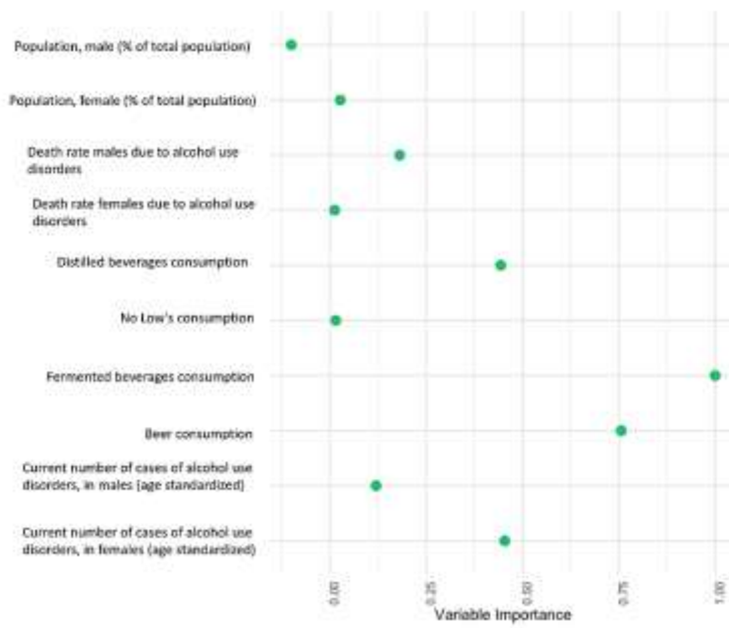
This is the importance of each variable in describing Death Rate Females. The variable that best explains the model is Alcohol Use Disorders Females. Second, Alcohol Use Disorders Males. The percentage of the male population is also a good predictor, although this is only a control and not a cluster. *In terms of* *consumption, it seems that the consumption of NoLos is the one that helps the most in predicting female deaths, followed by ferments, then distillates and finally beer.*

Chart 3. V.I.: Death Rate Males



In terms of the death rate for men (Death Rate Males), the best descriptor is Alcohol Use Disorders Males, followed by that of women (Alcohol Use Disorders Females). Of our control variables, the percentage of the male population is also important for the prediction of the model. *For consumption, the consumption of beer stands out by far, followed by ferments*. Distillates and No Low's seem not to be relevant in the prediction.

Chart 4. V.I.: Road Injuries



To describe car accidents, the most important variable is the clusters for consumption of fermented beverages, followed by the cluster for beer consumption. Thirdly, there is a technical tie between the alcohol use disorders of women (Alcohol use Disorders Females) and the Distillates consumption cluster. In general, we can say that alcohol consumption clusters are important in the prediction of the Automobile Accident model.

The explanation for why fermented drinks is strongly associated with accidents is that surely in the clusters of countries where the main source of alcohol comes from wines it explains a lot of the accidents but in other clusters where the sources of alcohol are varied or are not linked to fermented beverages, it is not.

Probabilities

Random forest aggregates the results of the individual trees in a manner where the observation is classified into the class for which most of the trees agree. In this same manner, it computes the probability of each class by estimating the proportion of times an observation is classified into each of the classes. The probability of belonging to one of the clusters analyzed according to the independent variables is then analyzed.[16]

Policy Analysis is built on the following criteria:

1. *Policy concentration vs policy diffusion*: to what extent clusters indicate similar magnitudes of probabilities or there are few or just one probability that outpaces all. Similar probabilities indicate opportunities for similar policies (likewise, different probabilities indicate need for different policies).

2. *Policy indifference vs policy monopoly.* Clusters indicate very low or negligible probabilities which turn them *indifferent* to any other policy intervention. The opposite case is policy monopoly or predominance. Clusters with minimal probabilities indicate there is no need to "waste time" focusing on them (likewise, very

[16] Annex 6 contains the charts of all countries displayed in order by probabilities.

high probabilities indicate urgent need to address).

3. *Policy irrelevance and strategy.* Policy irrelevance is determined in cases where there are varied and similar probabilities but others which are almost inexistent. This is a key aspect of policy strategy in terms of process. If two or more clusters are very similar in terms of probabilities then policy is not needed but to pursuit/improve the same policies that are being in place. Strategy then becomes key in terms of policy coherence and consistency.

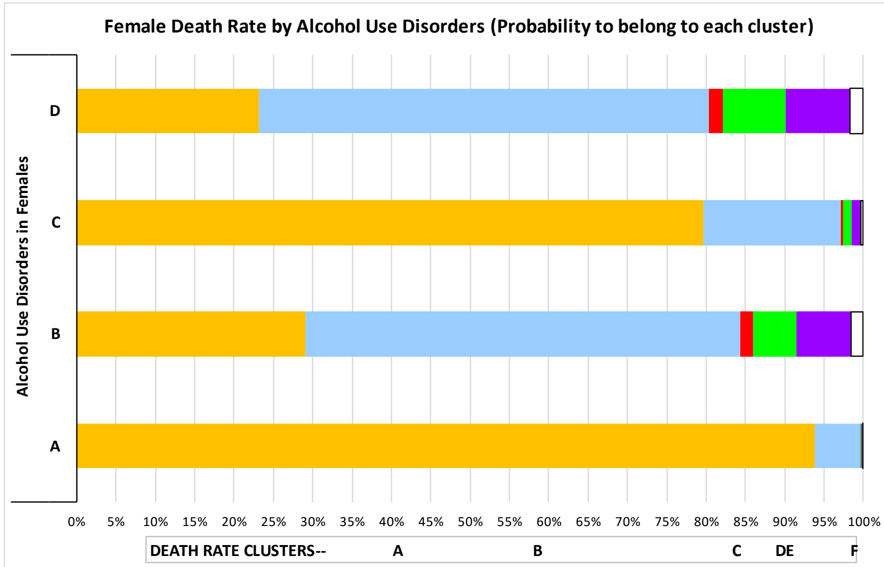
The following table synthesizes a practical way to read clustering and probability charts in terms of policies.

Criteria	**Description**	**Output**	**Outcome**
Policy concentration vs policy diffusion	Similar horizontal color bars indicate similar probabilities and therefore policy problems	Policy replication possible	Best practices, policy transfer.
	Different horizontal color bars indicate different probabilities and therefore policy problems	Policy replication not possible	Tailored, specific policies
Policy indifference vs policy monopoly	Minimal horizontal color bars indicate negligible probabilities and therefore no case for a policy	Policy resources not needed	No action
	Long horizontal color bars indicate high probability and therefore policy priority/urgency	Policy resources needed	Immediate action
Policy irrelevance and strategy	All horizontal color bars have very similar sizes among clusters, therefore no need for action	Policy status quo reinforced	Policy impact evaluation
	Some horizontal color bars have very similar sizes among clusters, need to focus on different with high probabilities	Policy prioritization	Tailored, specific policies

Policy departure (where to start) and *landing policies* (what to leave for the future). This indicates a policy choice that only a more in-depth analysis by the expert or country case researcher can determine. The entry point can be x clusters or y clusters. Policy departure is only identified when a deeper analysis can be made. It is related to the magnitude of available resources for implementing policies. *Policy Prioritization.* Prioritization occurs when probabilities are compared and ranked but, as all policies have limits and even political economic considerations, risk analysis should be performed. Prioritization might entail political economy as well as risk-type analyses.

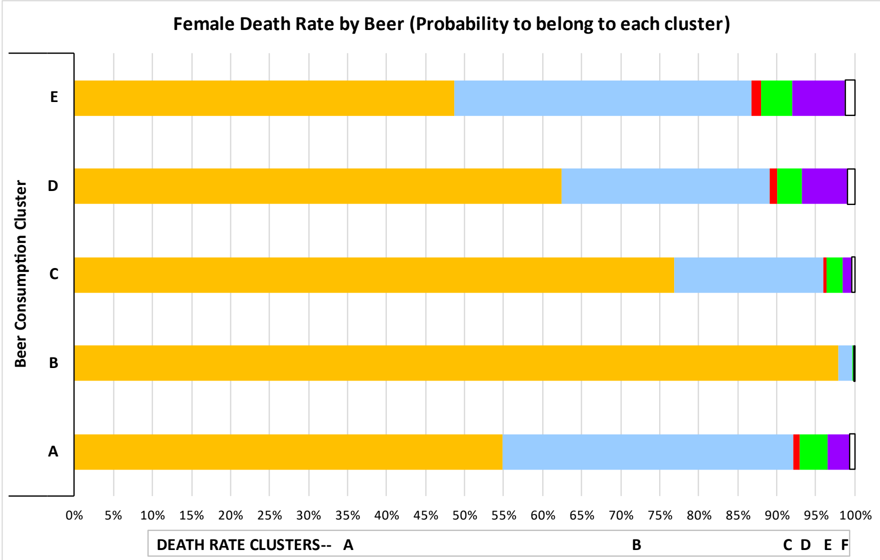
Female Death Rates Clusters

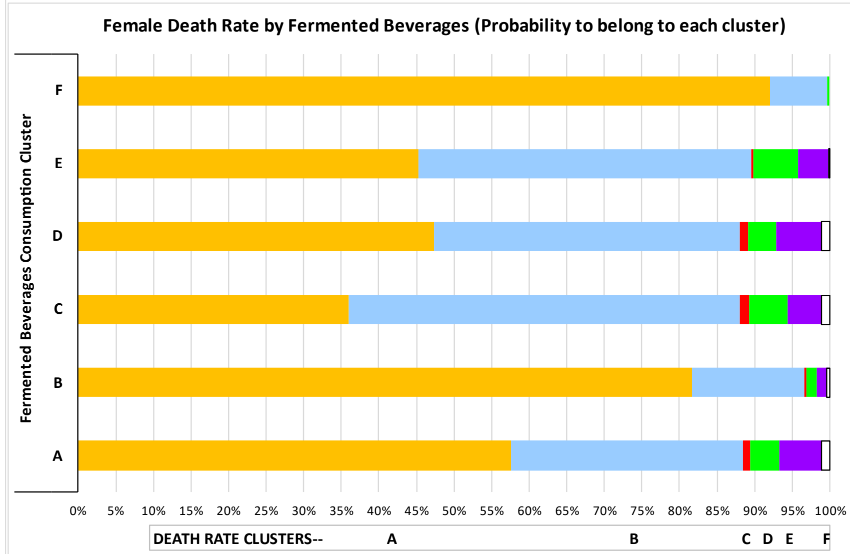
This section presents the probability of belonging to one of the 6 clusters of Death Rate in Females (see horizontal bar below for identifying purposes) due to alcohol use disorders by country according to all the independent variables.

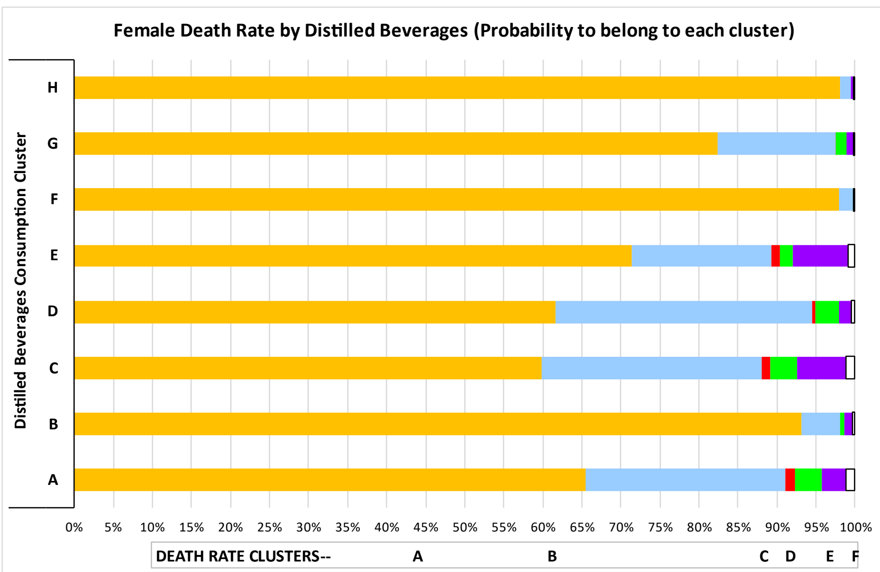


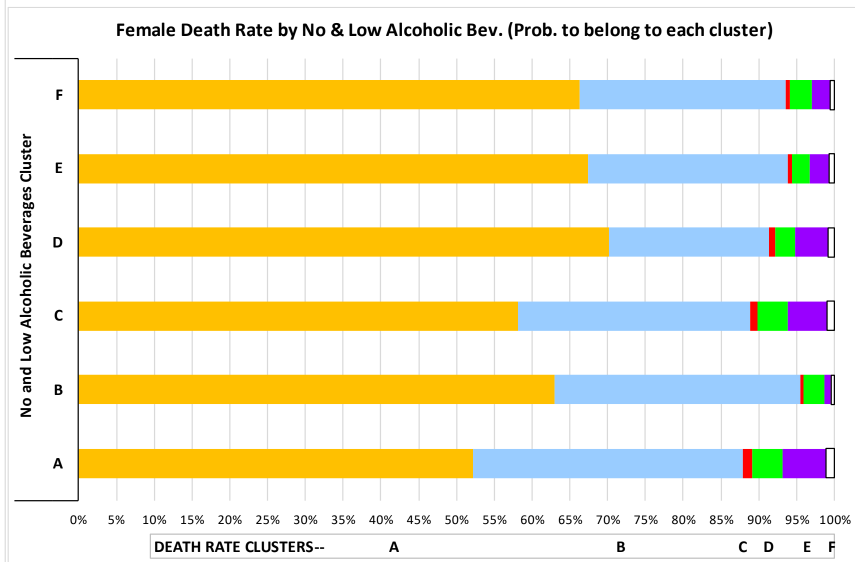
Analysis: Death rate clusters A and B (vertical reading, yellow & blue colors) define policy priorities for prevalence of alcohol use disorders in females. In terms of *policy concentration vs policy diffusion,* policy opportunities show a strong concentration that denotes symmetry. In terms of *policy indifference vs policy monopoly*, the first can be exemplified by Deaths clusters C (red color indicating very low probabilities), and to a lesser extent D (green on the left upper side of the chart). Finally, in terms of *policy irrelevance and strategy*, those irrelevant (to be ruled out) vs those strategic (top priority) are easily identified by looking at the longer length of rows or lines, which indicate high probability and therefore a stronger opportunity of improvement for any policy implementation (for example death rate clusters C, D, E, and F). This points out policy departure (where to start) and landing policies (what to leave for the future). Prioritization should also consider in this case to begin policy departure from Death rates or Prevalence of alcohol disorders clusters.

In these 4 charts, Female Death Rate by the Four types of alcoholic beverages according to their strength are portrayed. At first glance, all clusters indicate important variations among each type of beverage type and strength in terms of their potential to impact death rate in females. Beer has one cluster that presents and evident higher probability (C), fermented beverages indicate two salient clusters (B & F), distilled beverages present 3 clusters with high probabilities (B, F and H). NoLos has significant probability variation among all clusters which can be explained by the very low presence in the market and source of pure alcohol consumption. From a general perspective, beer and wine have strong variations among clusters while distilled beverages clusters are more homogeneous despite the 3 clusters already mentioned.



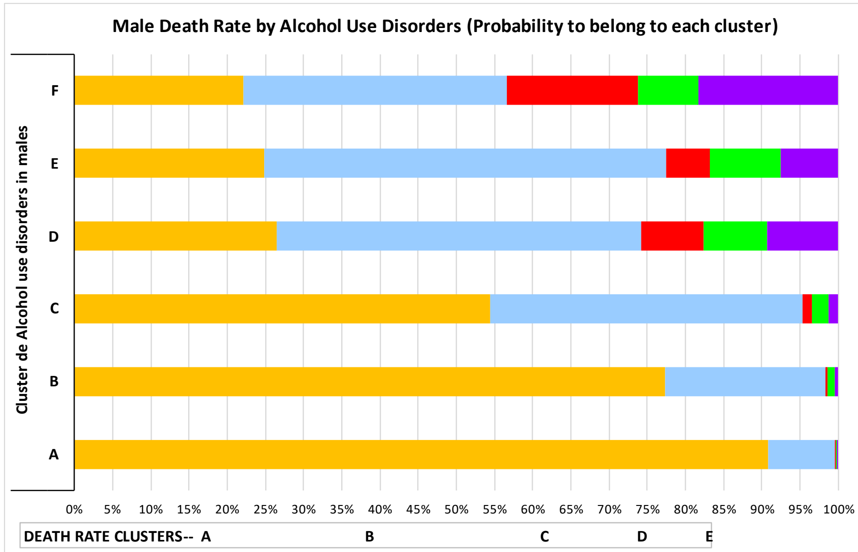






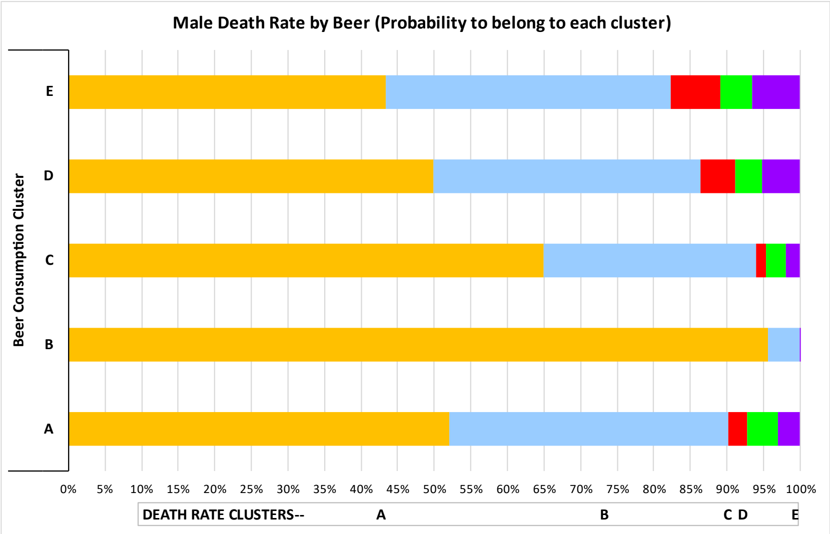
Male Death Rates Clusters

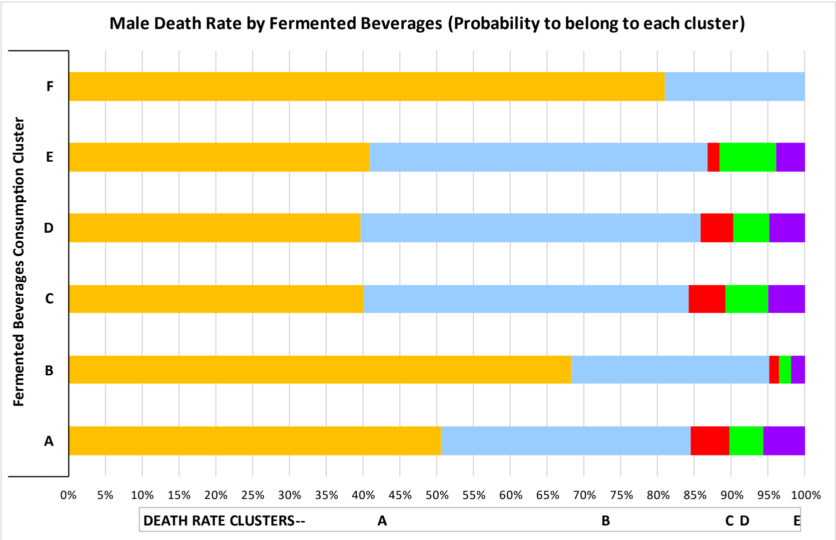
Each country's individual probability of being part of one of the 5 Prevalence cluster of Alcohol use Disorders Males according to its classification in the other clusters is presented below.

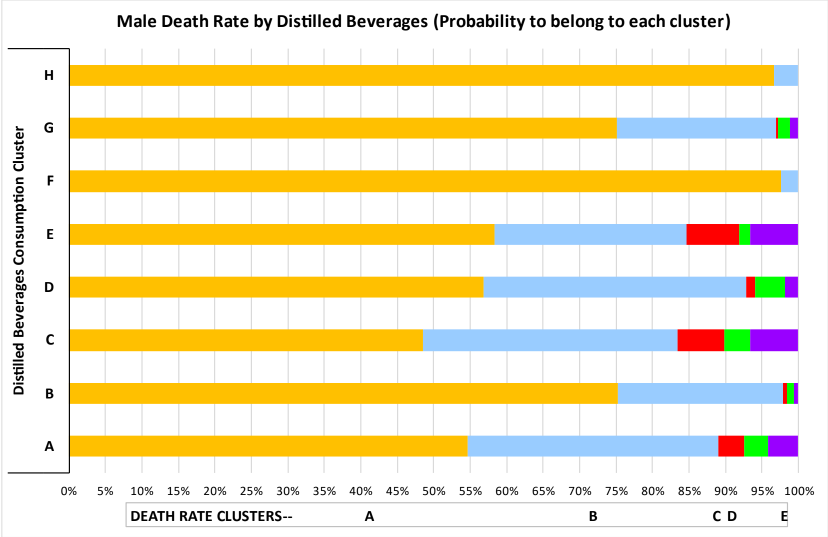


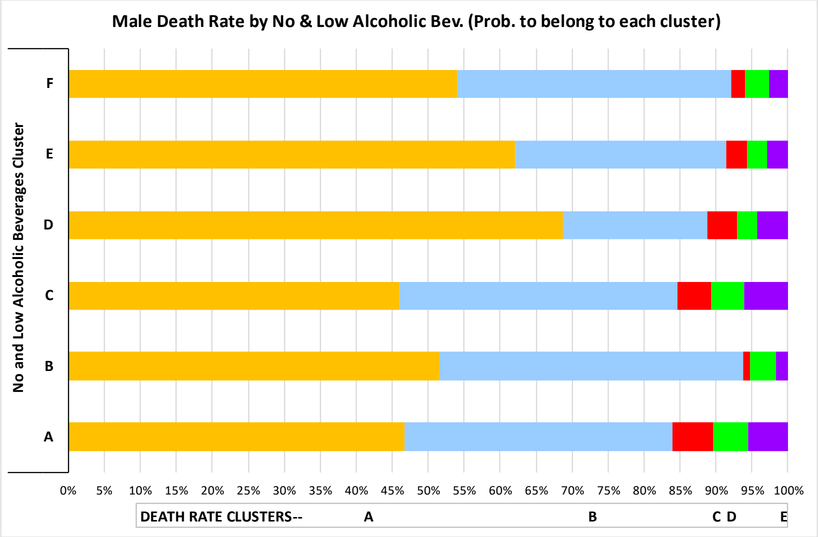
Analysis: Clusters A, B and C indicate an apparent policy priority (in descending order) in terms of a global policy inference. Cluster A and to a lesser extent, B, denote similarities and therefore potential policy transfer (replication of best practices in one cluster borrowed from other), but other clusters seem to have a variety of probabilities (more colors in horizontal lines) which indicate global policies are not easily replicated and probably should not be borrowed. For example, clusters D, E and F require a variety of policies for each of the probabilities. In general, higher probabilities indicate that death is strongly associated with alcohol use disorders, i.e. there is more probability that disorders lead to loss of life.

In these 4 charts, Male Death Rate by the Four types of alcoholic beverages according to their strength are portrayed. At first glance, all clusters indicate important variations among each type of beverage type and strength in terms of their potential to impact death rate in females. Beer has one cluster that presents and evident higher probability (B) which varies from females (C), fermented beverages indicate 2 salient clusters (B & F) which are the same as in females, distilled beverages present 3 clusters with high probabilities (B, F and H), where B indicates a relatively lower probability than in males. NoLos has significant probability variation among all clusters which can be explained by the very low presence in the market and source of pure alcohol consumption. As it happens in females, from a general perspective, beer and wine have strong variations among clusters while distilled beverages clusters are more homogeneous despite the 3 clusters already mentioned.

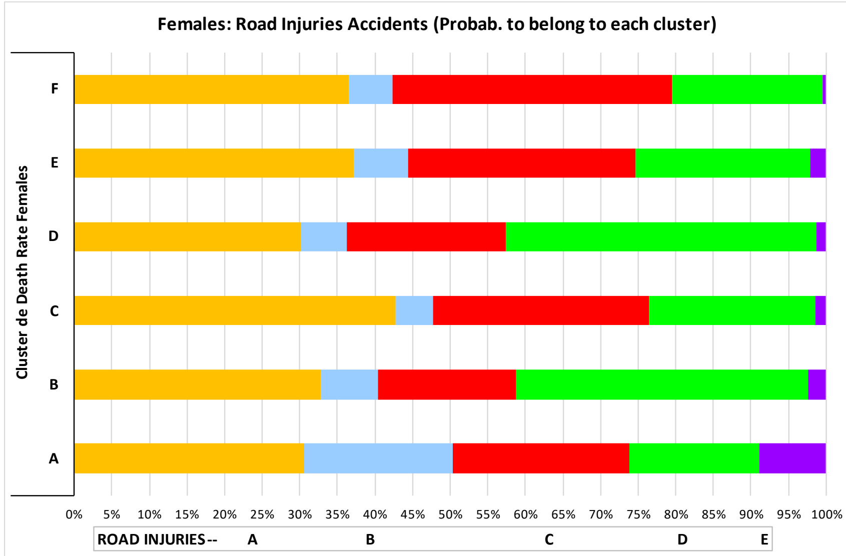


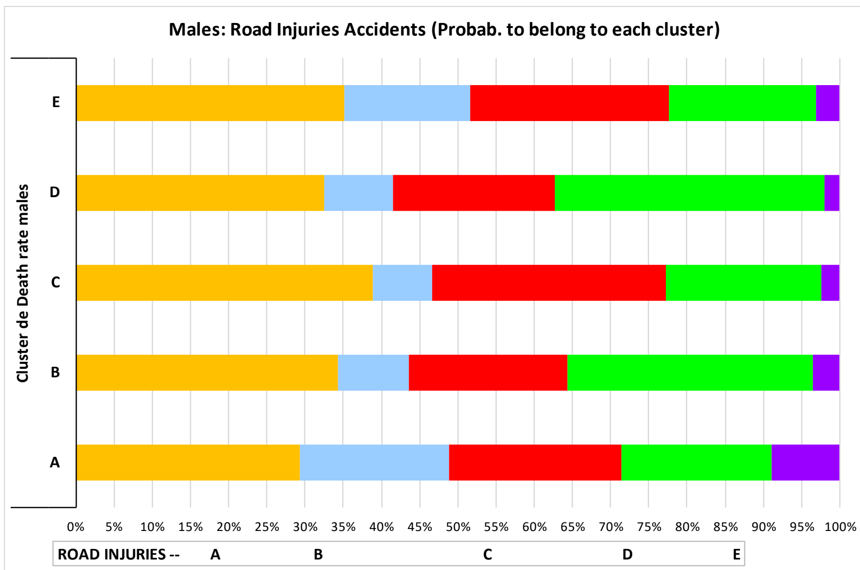






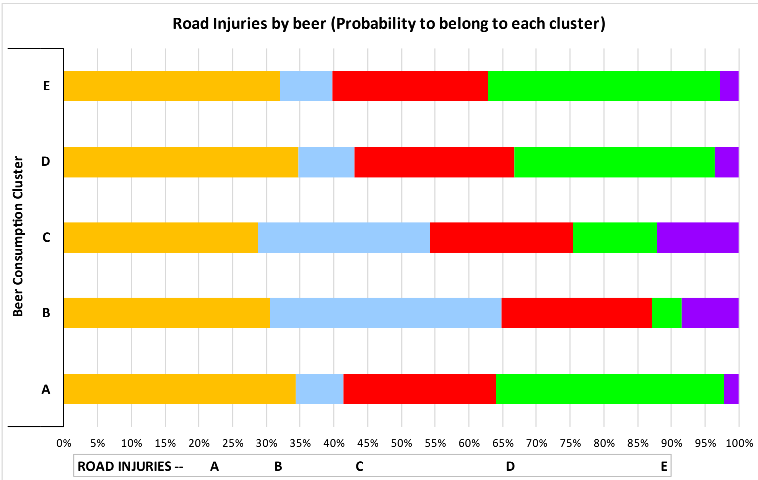
Cluster of Road Injuries

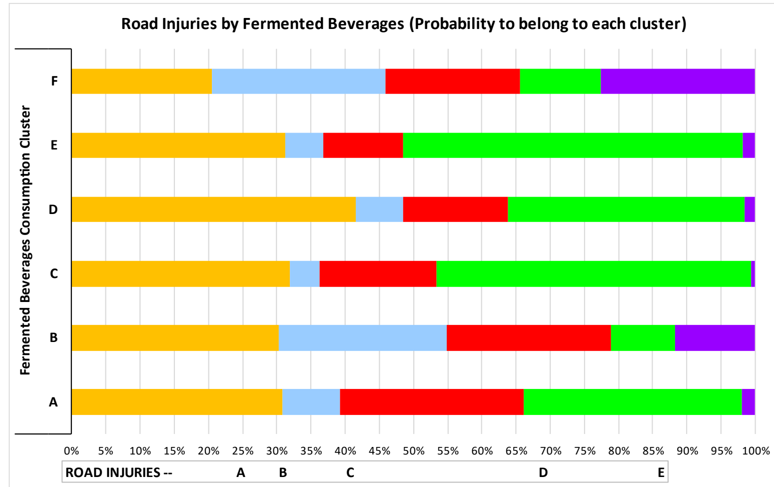


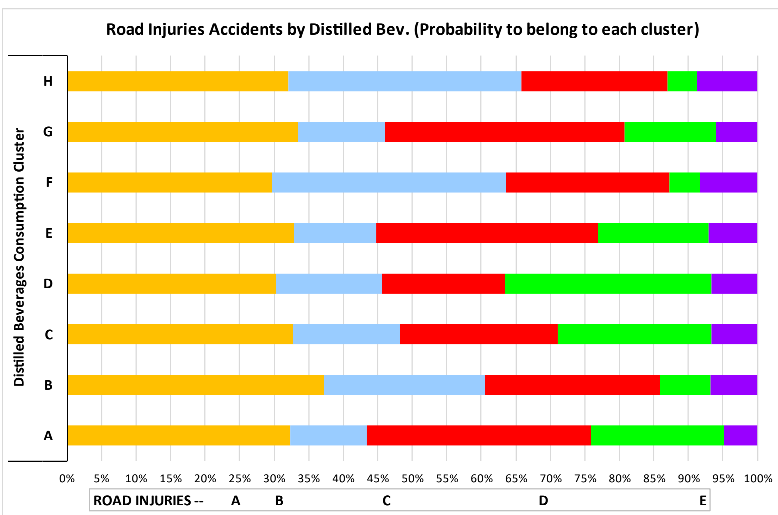


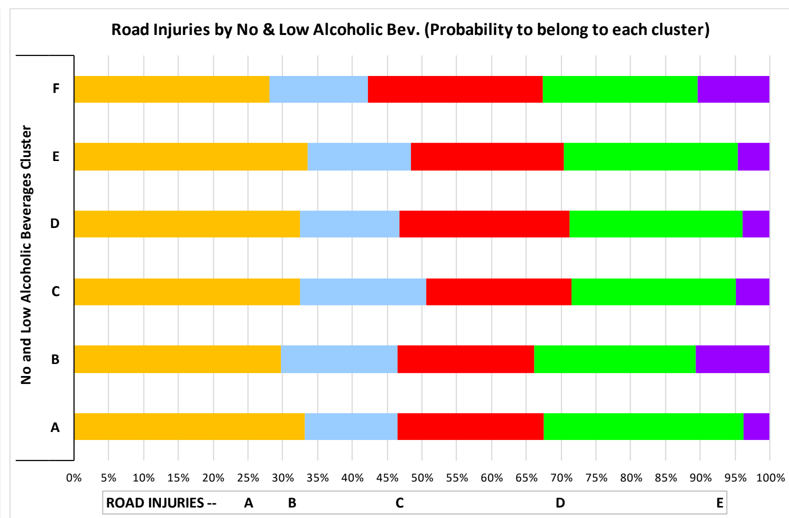
Analysis: In both cases all clusters and probabilities show very similar patterns and magnitudes. While females have more clusters than men the first 3 clusters (A, B, and C) are very similar. In conclusion, road injures, death rate and gender does not appear to indicate a gender gap.

In these 4 charts, Road Injuries by the Four types of alcoholic beverages according to their strength are portrayed. At first glance, all clusters indicate important similarities among each type of beverage type and strength in terms of their potential to impact. It is important to note that only NoLos indicates that policies among clusters do not need to be different. In contrast, there are cases in beer, fermented, and distilled beverages that present very similar probabilities (yellow & red horizontal areas in all clusters for all beverages). In contrast, blue & green horizontal areas indicate very different probabilities, some of them practically negligible (see green B cluster in beer, blue C, D, & E clusters in fermented and green areas in B, F, and H clusters in distilled). Low probabilities indicate policy indifference or irrelevance. In contrast, purple horizontal areas relevance are rare (B & C clusters in beer, B and F clusters in fermented).









Conclusions & Policy Recommendations

This report has set forth an ambitious approach to rethink how to build global policy inferences for alcoholic beverages harmful consumption by using artificial intelligence. It does so by using a Machine Learning Model which utilizes unsupervised learning and hierarchical clustering and obtain predictions from probabilities obtained from variable identification. The model is able to predict the percentage of probability for a country to present 3 phenomena: disorders, losses of lives and road injuries associated with the consumption of alcoholic beverages. Among these, it distinguishes between males and females so it is possible also to predict the gender gap for all phenomena.

We consider that this document has the potential to guide and open new areas of research by rethinking how countries, from a comparative and global perspective, compare between each other. We base our findings on the idea that countries are indeed ecosystems in which crucial variables interact and create the atmosphere or environment that explains the negative externalities of the harmful consumption of alcoholic beverages.

Building upon a regrouping (clusterization) of countries and phenomena associated with alcohol consumption we propose a probability model which characterizes each country, i.e. defines and deepens our knowledge on which specific variables are crucial to build a country's profile for research and policy purposes. We consider that this profiling is a proper way to confirm or question our understanding of the harmful consumption of alcohol, either from knowledge or policy agenda.

We pretend that both our proposal for a new clusterization or characterization of countries has the potential to rethink and perhaps reevaluate existing policies, guide government and non-governmental organizations and contribute to the analytic capabilities of those interested in the topic.

Policy recommendations

Aside from the potential policy recommendations that had been stated, in very broad terms, previously we consider that our research can be useful in terms of policy.

It has been often the case that country-level debates focus on narratives and ideas which are part of common sense, or conventional wisdom. At *Mexican Community*, we had developed a variety of pieces of research which had addressed the issue of the source or component that harms human health, i.e. pure alcohol, or, the product that is mostly or likely associated with such public problem. In this respect, there are country-level debates often focused on the negative impact of a specific type of beverage, in terms of its alcohol content or type (fermented, distilled). We consider that the policy dichotomy that focuses on pure alcohol content consumption in a specific type of beverage, or a type of beverage per se present important obstacles to the understanding and correct profiling of cases (countries).

We claim that the above dichotomy, to build policy upon alcohol content or type of beverage has the potential to be surmounted by fully incorporating the ecosystem in which such harmful consumption is embedded. This ecosystem has the potential to avoid political economy and implementation problems. This avoidance comes from the fact that we are trying to base our claims by more abundant, robust, and systematic data which goes beyond national or country borders.

By accepting the fact that the higher probabilities are placed in a specific sector of alcoholic beverages in terms of their negative externalities, we build a more efficient policy making process. This improved policymaking process is based on better and more efficient policy prioritization in terms of evaluating existing policies or building new ones. In any case, our model is a strong instrument to evaluate whether public and private, always scarce resources, are well placed. In any case, we hope our model triggers a new discussion on global policies and directives oriented towards a better and more prevention and treatment for people harmed by excessive and addictive consumption of alcoholic beverages.

References

References

Aswani, Serina, No-and low-alcohol category value surpasses $11 bn in 2022, the IWSR, December 2022, 2 p.p.

Brunori, Paolo, Paul Hufe and Daniel Mahler, The Roots of Inequality: Estimating Inequality of Opportunity from Regression Trees and Forests, IZA Institute of Labor Economics, 14689, August 2021.

Bryant, Lucy, Nudge theory and alcohol policy: how nudge frames drinkers and industry, Institute of Alcohol Studies, London, United Kingdom, 2020, 33 p.p.

Casale, Maria Daniela, Indirect taxation and gender equity: evidence from South Africa, Feminist Economics, 18, 2012, p. 25-54

Casswell, Sally, et. al., Benchmarking alcohol policy based on stringency and impact: The International Alcohol Control (IAC) policy index, PLOS Global Public Health, April 2022, 13 p.p.

Cresta Morgado, Pablo, Martin Caruso, Laura Alonso Alemany and Laura Acion, Practical foundations of machine learning for addiction research, The American Journal of Drug and Alcohol Abuse, 48, 2022, p. 260

Critchlow, Nathan and Crawford Moodie, Understanding the broader impact of alcohol marketing: time for a research agenda which includes adults, Alcohol and Alcoholism, 56, , Medical Council on Alcohol and Oxford University Press, April, 2021, p. 614- 616.

Cunillera, Toni, and Georgina Guilera, Twenty years of statistical learning: from language, back to mchine learning, Scientometrics, 117, 2018, p. 1-8

Díaz, Luis Antonio, et. al., Association between public health policies on alcohol and worldwide cancer, liver disease and cardiovascular disease outcomes, Journal of Hepatology, 80, March 2024, 409-418

Ebrahimi, Ali, et. al., Predicting the risk of alcohol use disorder using machine learning: a systematic literature review, IEEE Access, Vol. 9, 2021, 16 p.p.

Foster, K. T., B. M. Hicks, W. G. Iacono, and M. McGue, ``Alcohol use disorder in women: Risks and consequences of an adolescent onset and persistent course,'' Psychol. Addictive Behaviors, vol. 28, no. 2, p. 322, 2014.

Farfán, Gabriel, and Bruera, Esteban, Incidencia del consumo nocivo de bebidas alcohólicas en las entidades federativas de México -- Métricas de recursos públicos, desigualdad, salud, seguridad pública, violencia y género, Comunidad Mexicana de Gestión Pública para Resultados, A.C., Mexico City, May, 2024, 74 p.p.

Howlett, Neil, et. al., A systematic review and behaviour change technique analysis of remotely delivered alcohol and/or substance misuse interventions for adults, Drug and Alcohol Dependence, 239, 2022, 14 p.p.

Hurtado, Myrna, et. al., Machine Learning Applications and Advancements in Alcohol Use Disorder: A Systematic Review", MedR, 2022, p. 32

IAS, IAS Response to first draft of the WHO Global alcohol action plan 2022 to strengthen implementation of the Global Strategy to Reduce the Harmful Use of Alcohol, August, 2021, 8 p.p.

Institute of Alcohol Studies, IAS Response to first draft of the WHO Global alcohol action plan 2022 to strengthen implementation of the Global Strategy to Reduce the Harmful Use of Alcohol, IAS, August 2021, p. 8

International Tax and Investment Center, Guidebook for the Taxation of Distilled Spirits, London, U.K., 2000, p. 44

J. Bi, J. Sun, Y. Wu, H. Tennen, and S. Armeli, ``A machine learning approach to college drinking prediction and risk factor identication,'' ACM Trans. Intell. Syst. Technol., vol. 4, no. 4, pp. 124, Sep. 2013.

James, Gareth (Gareth Michael), Daniela Witten, Trevor Hastie, and Robert Tibshirani. An Introduction to Statistical Learning : With Applications in R. Second edition. New York, NY: Springer, 2021.

Jay, Jonathan, Alcohol outlets and firearm violence: a place-based case–control study using satellite imagery and machine learning, Inj Prev, 2019, p. 61-66

Lakshika Maduwanthi Ruberu, Thanthirige, et. al., Joint risk prediction for hazardous use of alcohol, cannabis, and tobacco among adolescents: A preliminary study using statistical and machine learning, Preventive Medicine Reports, 25, 2022, p. 6

Li, Can, Physiological data analysis – Alcohol drinking prediction using statistical and deep learning methods, Thesis, University of Missouri-Columbia, 2017, 70 p.p.

Manthey, Jakob, et. al., The Impact of Raising Alcohol Taxes on Government Tax Revenue: Insights from Five European Countries, Applied Health Economics and Health Policy, February 22, 2024,

McCaul, M. E., D. Roach, D. S. Hasin, C. Weisner, G. Chang, and R. Sinha, ``Alcohol and women: A brief overview,'' Alcoholism, Clin. Exp. Res., vol. 43, no. 5, p. 774, 2019.

Nicholls, Emily, The marketing and consumption of no and low alcohol drinks in the UK, IAS, 2022, 50 p.p.

Rattan, Puru, Daniel D. Penrice, and Douglas A. Simonetto, Artificial Intelligence and Machine Learning: What you always wanted to know but were afraid to ask, Gastro Hep Advances, 1, 2022, p. 70-78.

Rehm, Jurgen, et. al., Regulatory Policies for Alcohol, other Psychoactive Substances and Addictive Behaviours: The Role of Level of Use and Potency. A Systematic Review, International Journal of Environmental Research and Public Health, 2019

Rehm, Jurgen, Pol Rovira, Jakob Manthey, and Peter Anderson, Reduction of Alcoholic Strength: Does it matter for public health?, Nutrients, 15, February 2023. 14 p.p.

Room, Robin and Jurgen Rehm, “Harm per litre” as a concept and a measure in studying determinants of relations between alcohol consumption and harm, International Journal of Drug Policy, 115, 2023, 6 p.p.

Sassi, Franco, et. al., Equity impacts of price policies to promote healthy behaviours, The Lancet Taskforce on NCDs and economics 4, April 2018, 12 p.p.

T. Rosenström, F. A. Torvik, E. Ystrom, N. O. Czajkowski, N. A. Gillespie, S. H. Aggen, R. F. Krueger, K. S. Kendler, and T. Reichborn-Kjennerud, ``Prediction of alcohol use disorder using personality disorder traits: A twin study,'' Addiction, vol. 113, no. 1, pp. 1524, Jan. 2018.

Taylor, Anne W., et. al., Clusters of alcohol abstainers and drinkers incorporating motives against drinking: a random survey of 18 to 34 year olds in four cities in four different continents", AIMS Public Health, 6, January 2019, p. 15-33.

The IWSR Group, The IWSR Database, , 2022, p. 38

Torney, Alexandra, Robin Room, Taisia Huckle, Sally Casswell, and Sarah Callinan, Beverage-specific consumption trends: A cross-country, cross-sectional comparison, International Journal of Drug Policy 117, 2023, 6 p.p.

Walters, Scott T., et. al., Using machine learning to identify predictors of imminent drinking and create tailored messages for at-risk drinkers experiencing homelessness, Journal of Substance Abuse Treatment, 127, 2021, 9 p.p.

White, A., ``Gender differences in the epidemiology of alcohol use and related harms in the United States,'' *Alcohol Res., Current Rev.*, vol. 40, no. 2, p. 1, 2020.

WHO, Global Status Report on Road Safety, Management of Noncommunicable Diseases, Disability, Violence and Injury Prevention (NVI), Geneva, Switzerland, 2018, 424 p.p.

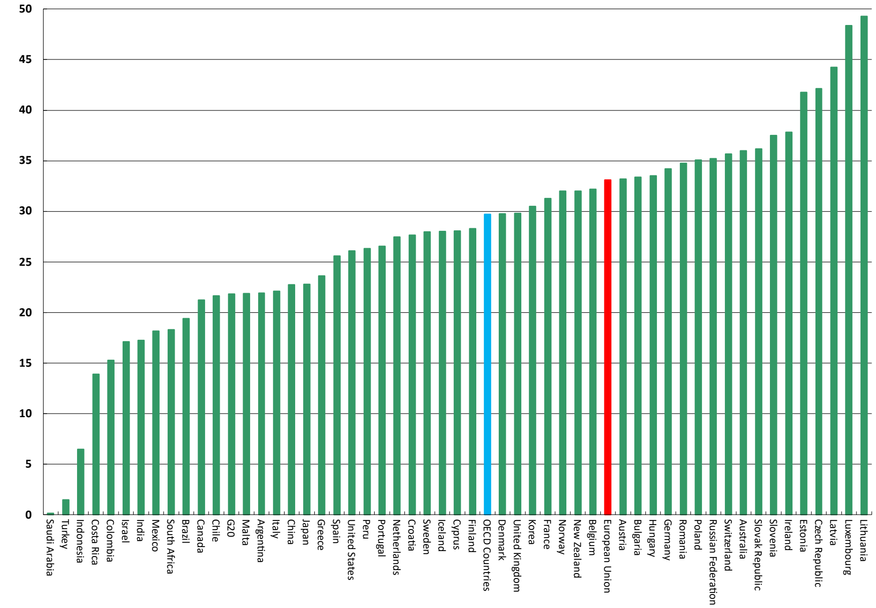
WHO, Global Status Report: Alcohol Policy, WHO-Department of Mental Health and Substance Abuse, Geneva, Switzerland, 2004, 209 p.p.

World Health Organization, Political declaration of the third high-level meeting of the General Assembly on the prevention and control of non-communicable diseases, January 2022, 35 p.p.

Annexes

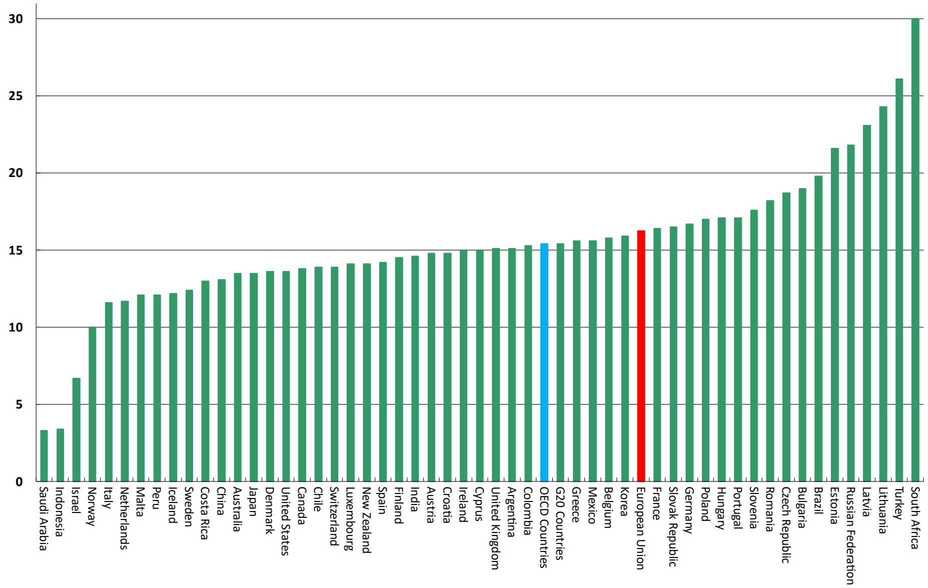
Annex 1. Alcoholic Beverages Consumption and Health Comparative Indicators.

Prevalence of heavy episodic drinking (Percentage of adult population (aged 15+) with at least one occasion of heavy episodic drinking in the past 30 days, 2016)



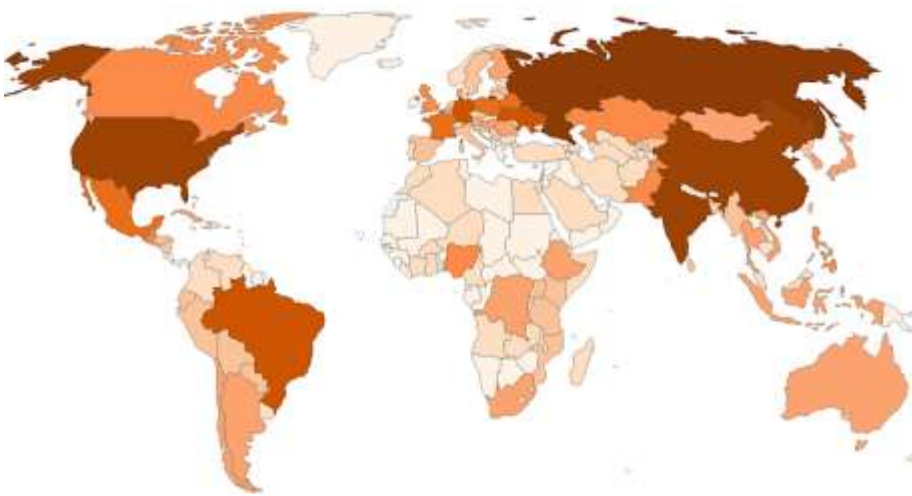
Source: Own elaboration with data from OECD, 2021.

Alcohol consumption, drinkers only (Total per capita (15+) alcohol consumption (in litres of pure alcohol) for drinkers only, 2016.



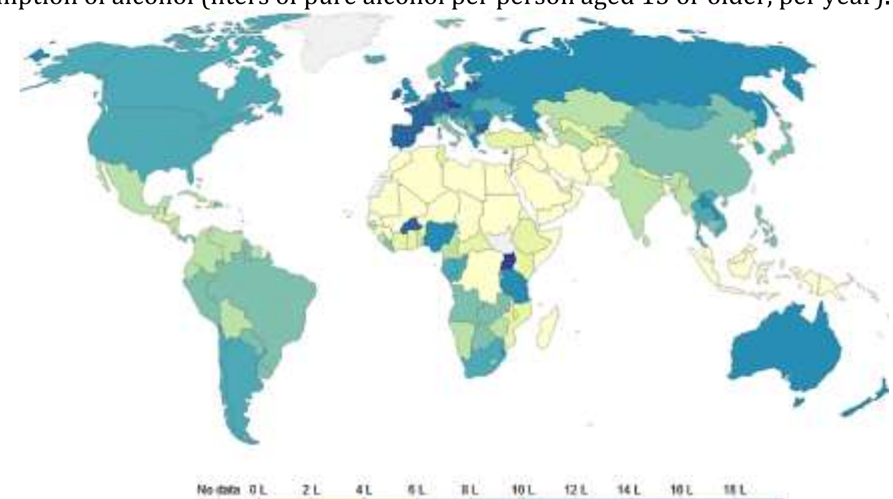
Source: Own elaboration with data from OECD, 2021.

Deaths from alcohol use disorders, 2019.
Annual number of deaths from alcohol use disorders.



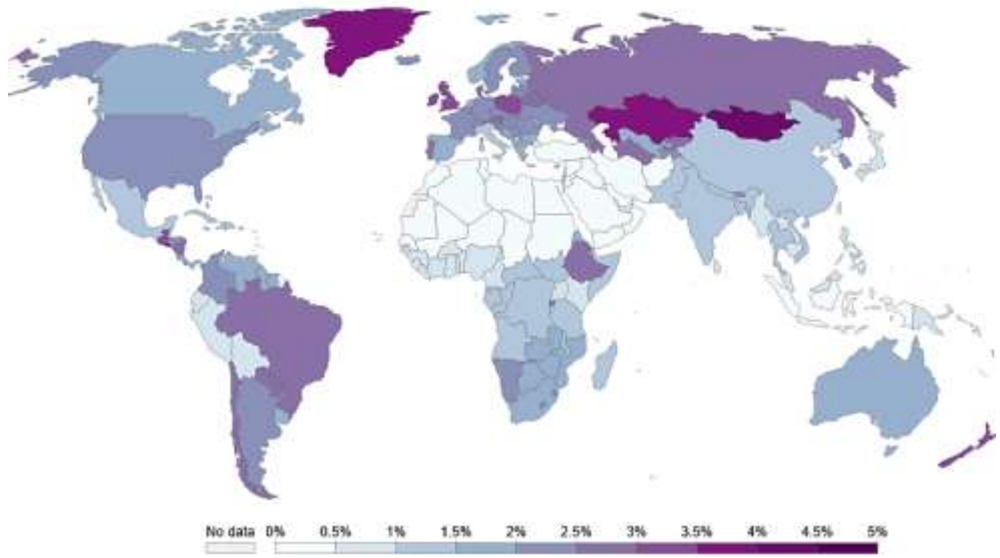
Source: Our World in Data, 2024.

Alcohol consumption per person, 2018.
Consumption of alcohol (liters of pure alcohol per person aged 15 or older, per year).



Source: Our World in Data, 2024.

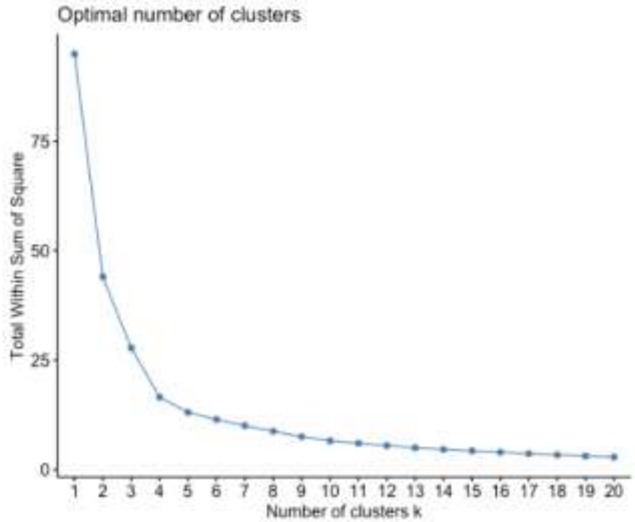
Share of population with an alcohol use disorder, 2019.
Alcohol dependence is defined by the International Classification of Diseases as the presence of three or more indicators of dependence for at least a month within the previous year.

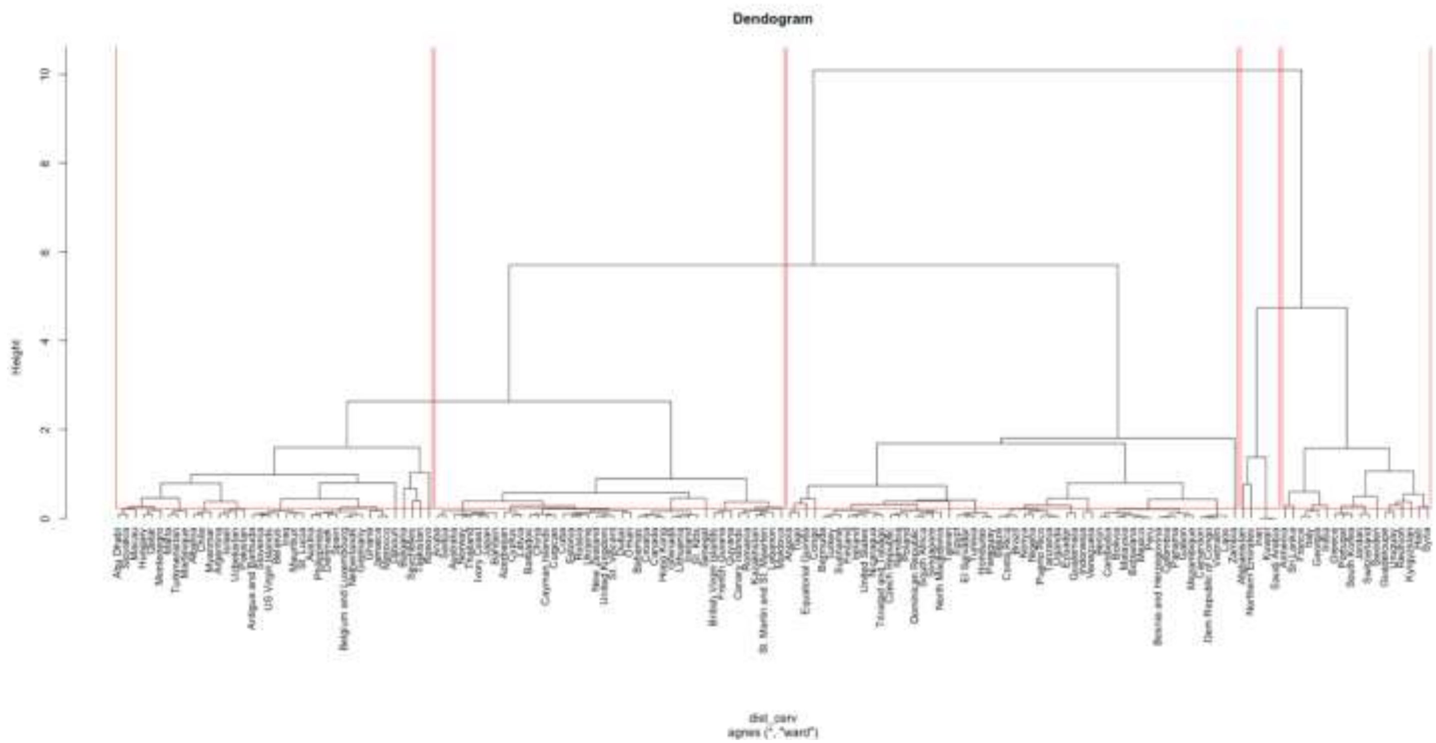


Source: Own elaboration with data from OECD, 2021.

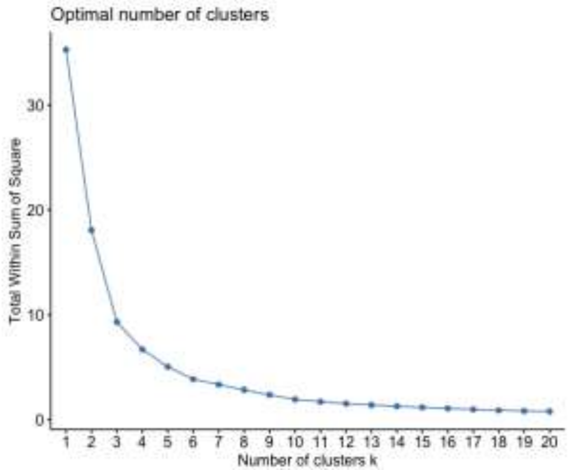
Annex 2. Alcoholic Beverages Consumption by Strength (content of pure alcohol) clusterization.

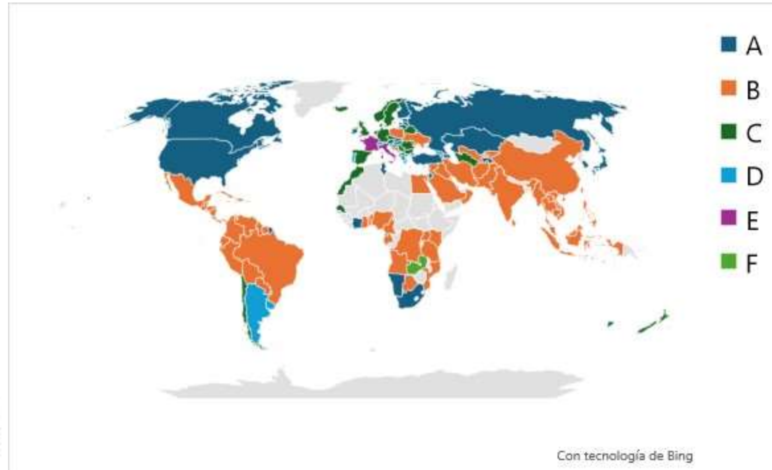
Beer optimal number of clusters, map and dendrogram (5)

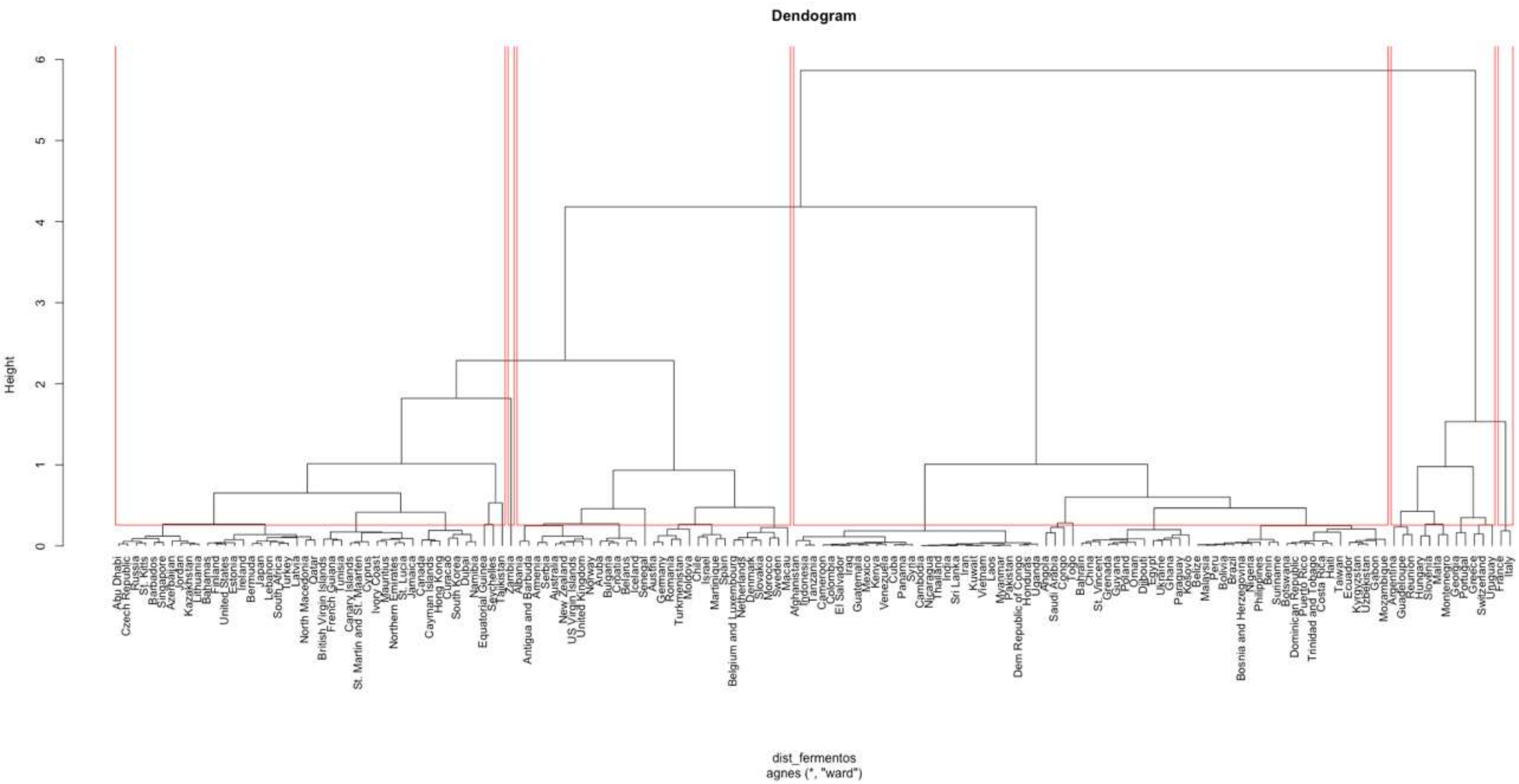




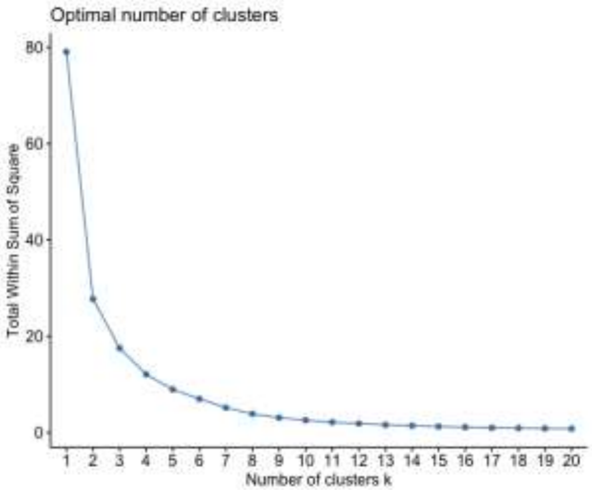
Ferments optimal number of clusters, map and dendrogram (6)



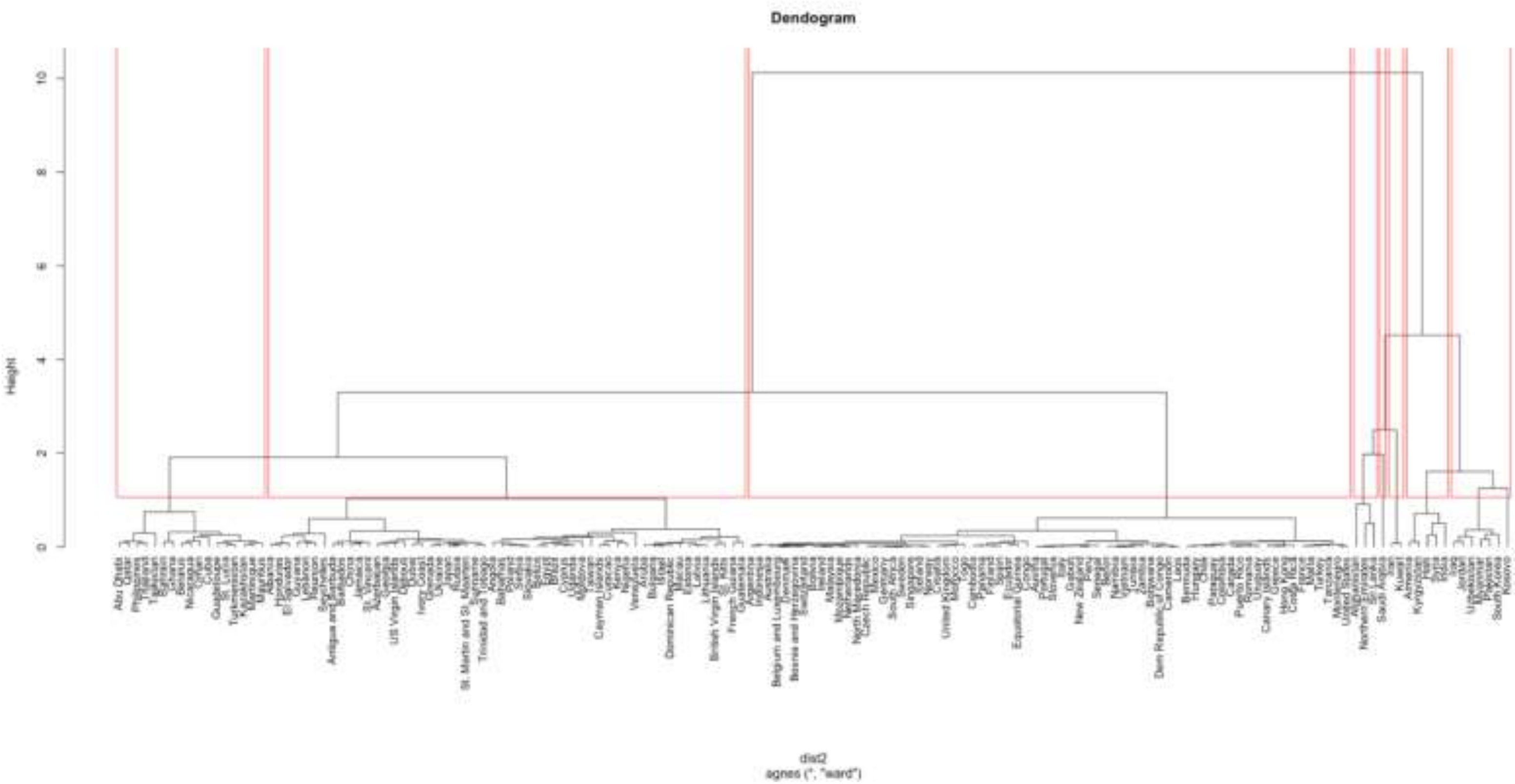




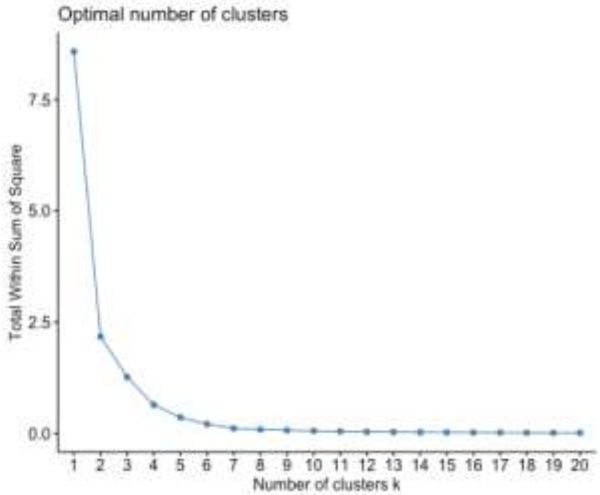
Distilled optimal number of clusters, map and dendrogram (8)

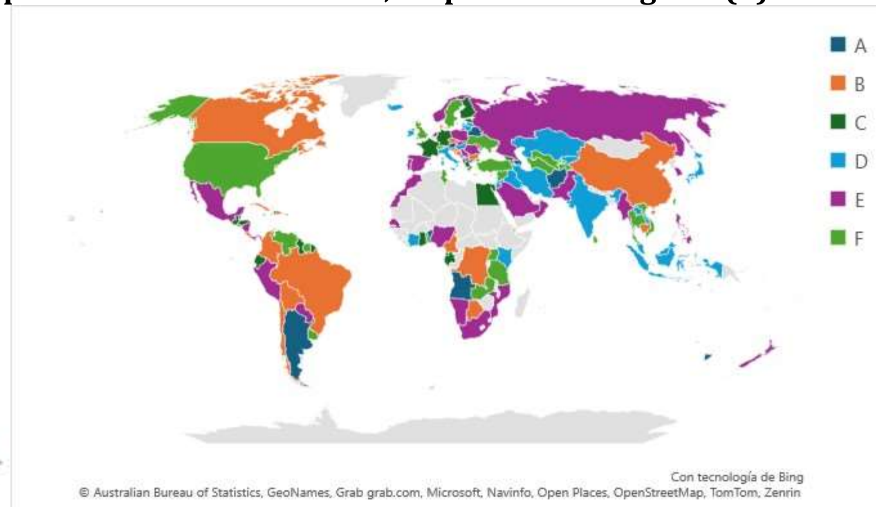


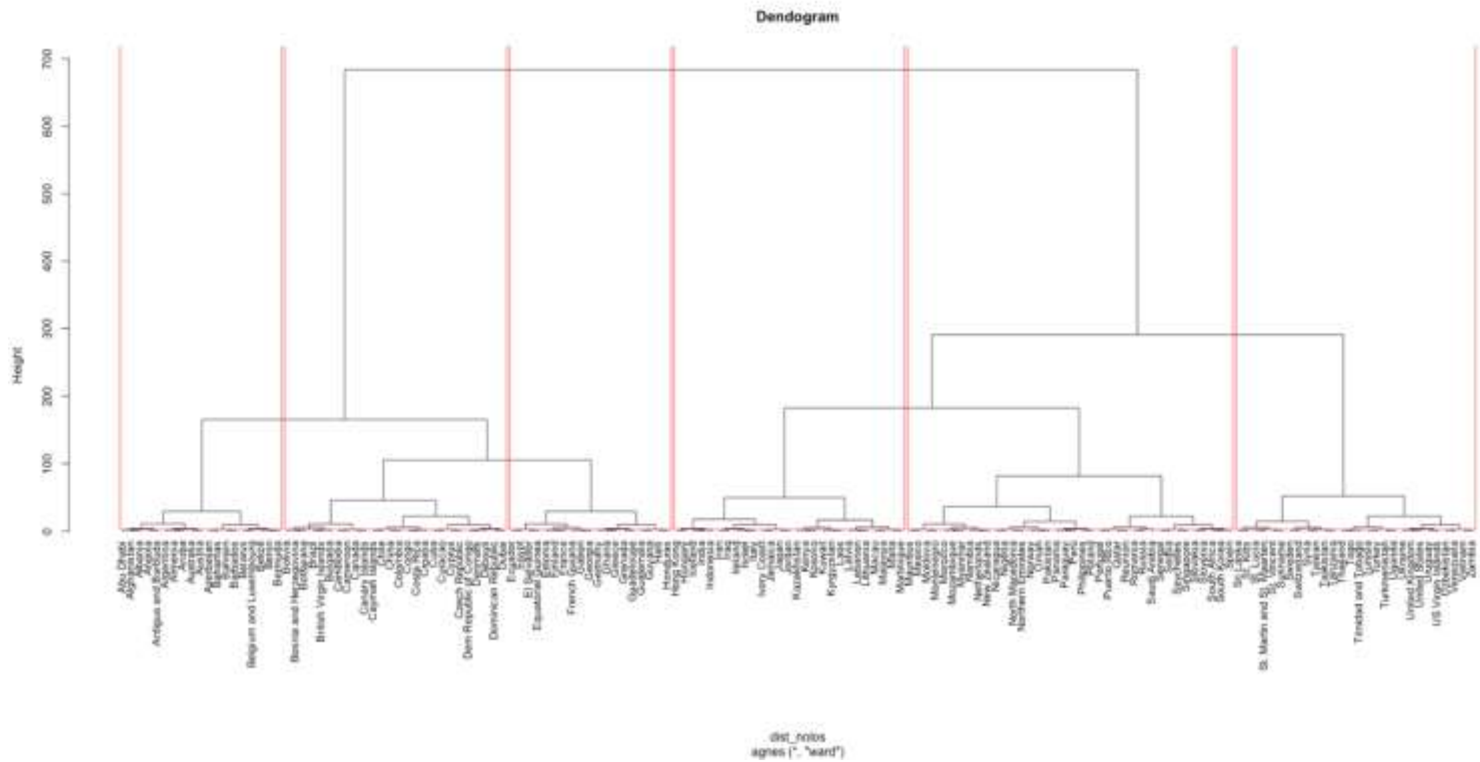
Distilled Dendrogram



No and Low Alcohol Beverages optimal number of clusters, map and dendrogram (6)







Annex 3. List of countries by clusters of alcohol content

No and Low Alcoholic Drinks (NoLo	Beer	Fermented	Distilled
A	**A**	**A**	**A**
Abu Dhabi	Abu Dhabi	Abu Dhabi	Abu Dhabi
Afghanistan	Albania	Azerbaijan	Bahrain
Albania	Antigua and Barbuda	Bahamas	Belarus
Angola	Argentina	Barbados	Cuba
Antigua and Barbuda	Austria	Bermuda	Ghana
Argentina	Belarus	British Virgin Islands	Guadeloupe
Armenia	Belgium and Luxembourg	Canada	Kazakhstan
Aruba	Bulgaria	Canary Islands	Martinique
Australia	Chile	Cayman Islands	Mauritius
Austria	Denmark	Curacao	Nicaragua
Azerbaijan	Djibouti	Cyprus	Oman
Bahamas	Germany	Czech Republic	Philippines
Bahrain	Ghana	Dubai	Qatar
Barbados	Hungary	Equatorial Guinea	St. Lucia
Belarus	Iraq	Estonia	Tajikistan
Belgium and Luxembourg	Israel	Finland	Thailand
Belize	Jamaica	French Guiana	Turkmenistan
Benin	Kosovo	Hong Kong	**B**
Bermuda	Macau	Ireland	Afghanistan
B	Malta	Ivory Coast	Northern Emirates
Bolivia	Martinique	Jamaica	Sri Lanka
Bosnia and Herzegovina	Mauritius	Japan	**C**
Botswana	Montenegro	Jordan	Albania
Brazil	Morocco	Kazakhstan	Angola
British Virgin Islands	Myanmar	Latvia	Antigua and Barbuda
Bulgaria	Netherlands	Lebanon	Aruba
Cambodia	Pakistan	Lithuania	Azerbaijan
Cameroon	Philippines	Mauritius	Bahamas
Canada	Qatar	Namibia	Barbados
Canary Islands	Seychelles	North Macedonia	Belize
Cayman Islands	Slovakia	Northern Emirates	Brazil
Chile	Slovenia	Qatar	British Virgin Islands
China	Spain	Russia	Bulgaria
Colombia	St. Lucia	Seychelles	Cayman Islands
Congo	Tajikistan	Singapore	China
Costa Rica	Turkmenistan	South Africa	Curacao
Croatia	US Virgin Islands	South Korea	Cyprus
Cuba	Uzbekistan	St. Kitts	Djibouti
Curacao	**B**	St. Lucia	Dominican Republic
Cyprus	Afghanistan	St. Martin and St. Maarten	Dubai
Czech Republic	Iran	Tajikistan	Egypt
Dem Republic of Congo	Kuwait	Tunisia	El Salvador
Denmark	Northern Emirates	Turkey	Estonia
Djibouti	Saudi Arabia	United States	French Guiana
Dominican Republic	**C**	**B**	Georgia
Dubai	Angola	Afghanistan	Grenada
C	Belize	Angola	Guatemala
Ecuador	Benin	Bahrain	Guyana
Egypt	Bermuda	Belize	Honduras
El Salvador	Bolivia	Benin	Israel
Equatorial Guinea	Bosnia and Herzegovina	Bolivia	Ivory Coast
Estonia	Botswana	Bosnia and Herzegovina	Jamaica
Finland	Brazil	Botswana	Japan
France	Cambodia	Brazil	Kenya
French Guiana	Cameroon	Cambodia	Latvia
Gabon	Colombia	Cameroon	Lebanon
Georgia	Congo	China	Lithuania
Germany	Costa Rica	Colombia	Macau
Ghana	Czech Republic	Congo	Moldova
Greece	Dem Republic of Congo	Costa Rica	Nigeria
Grenada	Dominican Republic	Cuba	Poland
Guadeloupe	Ecuador	Dem Republic of Congo	Reunion
Guatemala	Egypt	Djibouti	Russia
Guyana	El Salvador	Dominican Republic	Serbia
Haiti	Equatorial Guinea	Ecuador	Seychelles
Honduras	Finland	Egypt	Slovakia
D	Gabon	El Salvador	St. Kitts
Hong Kong	Guatemala	Gabon	St. Martin and St. Maarten
Hungary	Honduras	Ghana	St. Vincent
Iceland	Indonesia	Grenada	Suriname
India	Ireland	Guatemala	Taiwan
Indonesia	Kenya	Guyana	Trinidad and Tobago
Iran	Laos	Haiti	Uganda
Iraq	Malaysia	Honduras	Ukraine
Ireland	Mexico	India	US Virgin Islands
Israel	Mozambique	Indonesia	Venezuela

D

Italy
Ivory Coast
Jamaica
Japan
Jordan
Kazakhstan
Kenya
Kosovo
Kuwait
Kyrgyzstan
Laos
Latvia
Lebanon
Lithuania
Macau
Malaysia
Malta
Martinique

E

Mauritius
Mexico
Moldova
Montenegro
Morocco
Mozambique
Myanmar
Namibia
Netherlands
New Zealand
Nicaragua
Nigeria
North Macedonia
Northern Emirates
Norway
Oman
Pakistan
Panama
Paraguay
Peru
Philippines
Poland
Portugal
Puerto Rico
Qatar
Reunion
Romania
Russia
Saudi Arabia
Senegal
Serbia
Seychelles
Singapore
Slovakia
Slovenia
South Africa
South Korea
Spain

F

Sri Lanka
St. Kitts
St. Lucia
St. Martin and St. Maarten
St. Vincent
Suriname
Sweden
Switzerland
Syria
Taiwan
Tajikistan
Tanzania
Thailand
Togo
Trinidad and Tobago
Tunisia
Turkey
Turkmenistan
Uganda
Ukraine
United Kingdom
United States
Uruguay
US Virgin Islands
Uzbekistan
Venezuela
Vietnam
Zambia

C

Namibia
Nicaragua
Nigeria
North Macedonia
Panama
Paraguay
Peru
Poland
Puerto Rico
Singapore
South Africa
Suriname
Taiwan
Tanzania
Togo
Trinidad and Tobago
Tunisia
Turkey
Uganda
United States
Venezuela
Vietnam
Zambia

D

Armenia
France
Georgia
Greece
Guadeloupe
Haiti
India
Italy
Jordan
Kyrgyzstan
Portugal
Reunion
South Korea
Sri Lanka
Sweden
Switzerland
Syria
Uruguay

E

Aruba
Australia
Azerbaijan
Bahamas
Bahrain
Barbados
British Virgin Islands
Canada
Canary Islands
Cayman Islands
China
Croatia
Cuba
Curacao
Cyprus
Dubai
Estonia
French Guiana
Grenada
Guyana
Hong Kong
Iceland
Ivory Coast
Japan
Kazakhstan
Latvia
Lebanon
Lithuania
Moldova
New Zealand
Norway
Oman
Romania
Russia
Senegal
Serbia
St. Kitts
St. Martin and St. Maarten
St. Vincent
Thailand
Ukraine
United Kingdom

B

Iran
Iraq
Kenya
Kosovo
Kuwait
Kyrgyzstan
Laos
Malaysia
Mexico
Mozambique
Myanmar
Nicaragua
Nigeria
Oman
Pakistan
Panama
Paraguay
Peru
Philippines
Poland
Puerto Rico
Saudi Arabia
Sri Lanka
St. Vincent
Suriname
Syria
Taiwan
Tanzania
Thailand
Togo
Trinidad and Tobago
Uganda
Ukraine
Uzbekistan
Venezuela
Vietnam

C

Albania
Antigua and Barbuda
Armenia
Aruba
Australia
Austria
Belarus
Belgium and Luxembourg
Bulgaria
Chile
Croatia
Denmark
Germany
Iceland
Israel
Macau
Martinique
Moldova
Morocco
Netherlands
New Zealand
Norway
Romania
Senegal
Serbia
Slovakia
Spain
Sweden
Turkmenistan
United Kingdom
US Virgin Islands

D

Argentina
Georgia
Greece
Guadeloupe
Hungary
Malta
Montenegro
Portugal
Reunion
Slovenia
Switzerland
Uruguay

E

France
Italy

F

Zambia

D

D

Argentina
Australia
Austria
Belgium and Luxembourg
Benin
Bermuda
Bolivia
Bosnia and Herzegovina
Botswana
Cambodia
Cameroon
Canada
Canary Islands
Chile
Colombia
Congo
Costa Rica
Croatia
Czech Republic
Dem Republic of Congo
Denmark
Ecuador
Equatorial Guinea
Finland
France
Gabon
Germany
Greece
Hong Kong
Hungary
Iceland
Indonesia
Ireland
Italy
Laos
Malaysia
Malta
Mexico
Montenegro
Morocco
Mozambique
Namibia
Netherlands
New Zealand
North Macedonia
Norway
Panama
Paraguay
Peru
Portugal
Puerto Rico
Romania
Senegal
Singapore
Slovenia
South Africa
Spain
Sweden
Switzerland
Tanzania
Togo
Tunisia
Turkey
United Kingdom
United States
Uruguay
Vietnam
Zambia

E

Armenia
Haiti
India
Kyrgyzstan
Syria

F

Iran
Kuwait

G

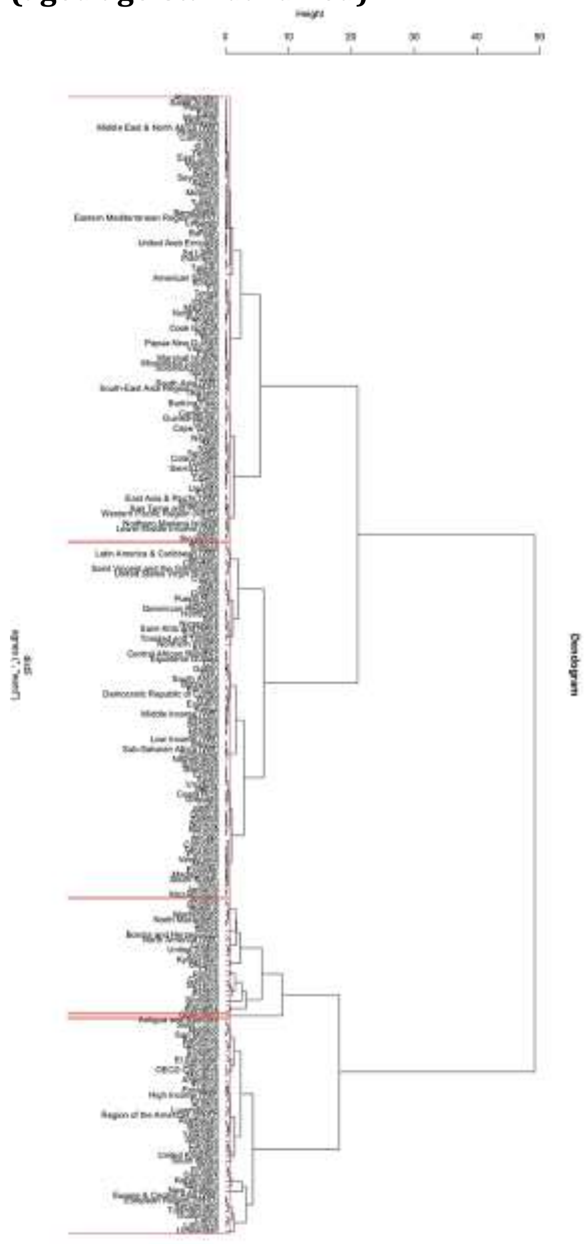
Iraq
Jordan
Kosovo
Myanmar
Pakistan
South Korea
Uzbekistan

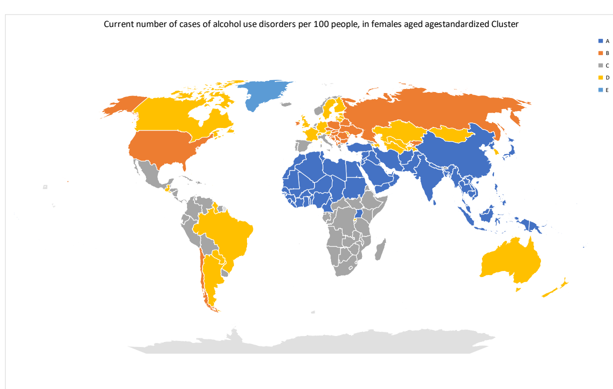
H

Saudi Arabia

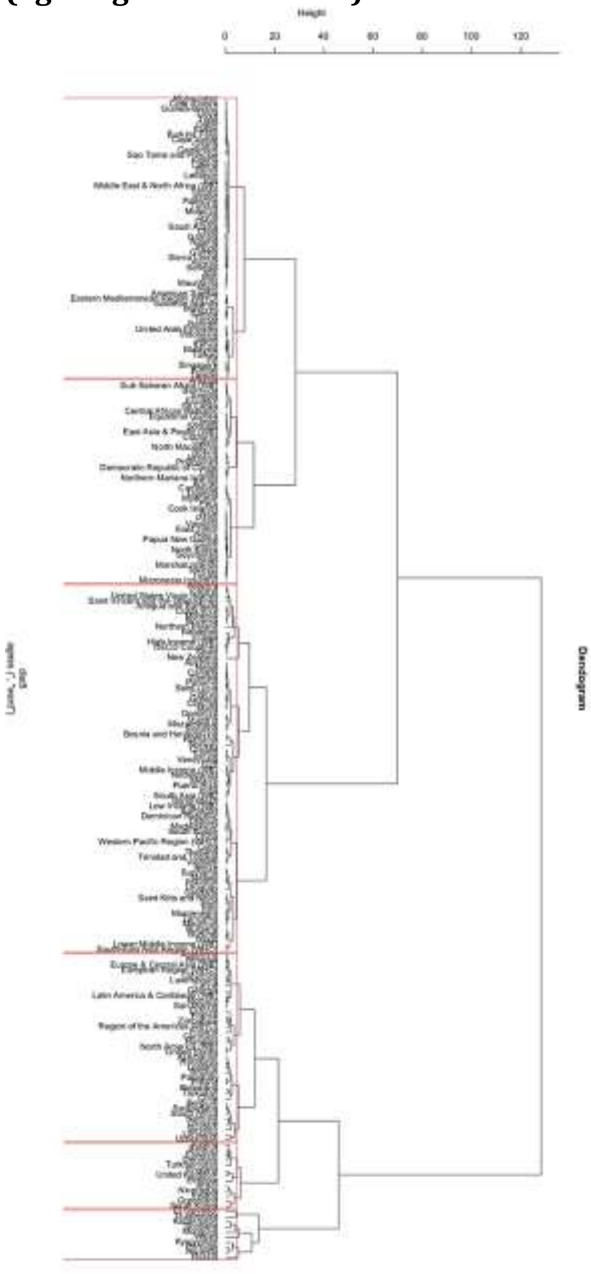
Annex 4. Alcohol Use Disorders, Death Rate, and Road Injuries Dendrograms

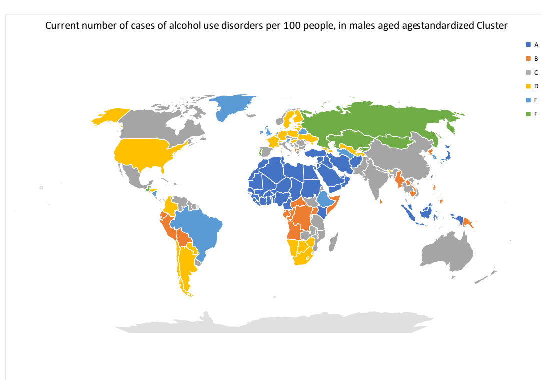
Alcohol use disorders per 100 people, in females (aged age-standardized)



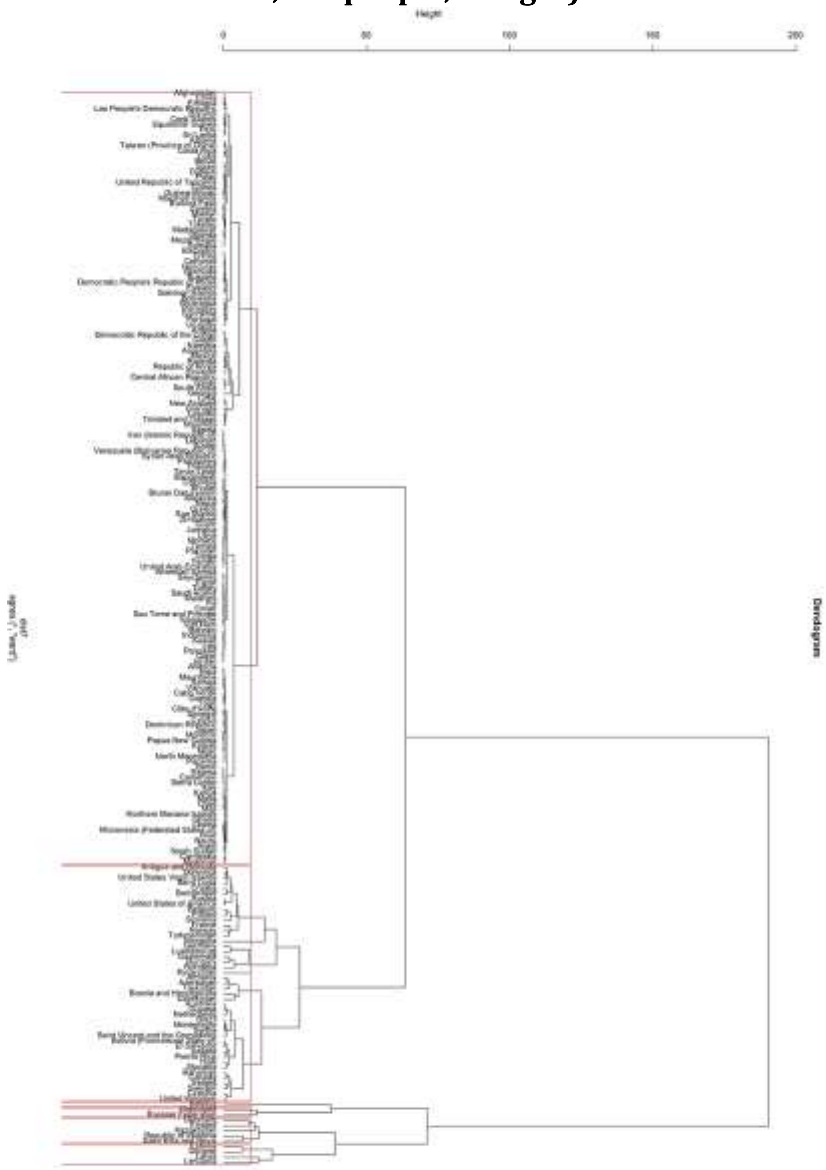


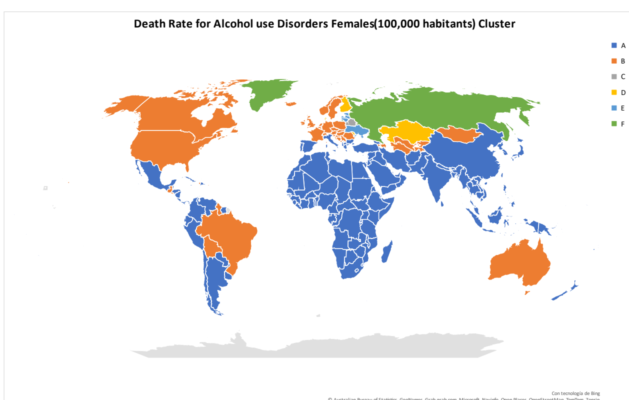
Alcohol use disorders per 100 people, in males (aged age-standardized)



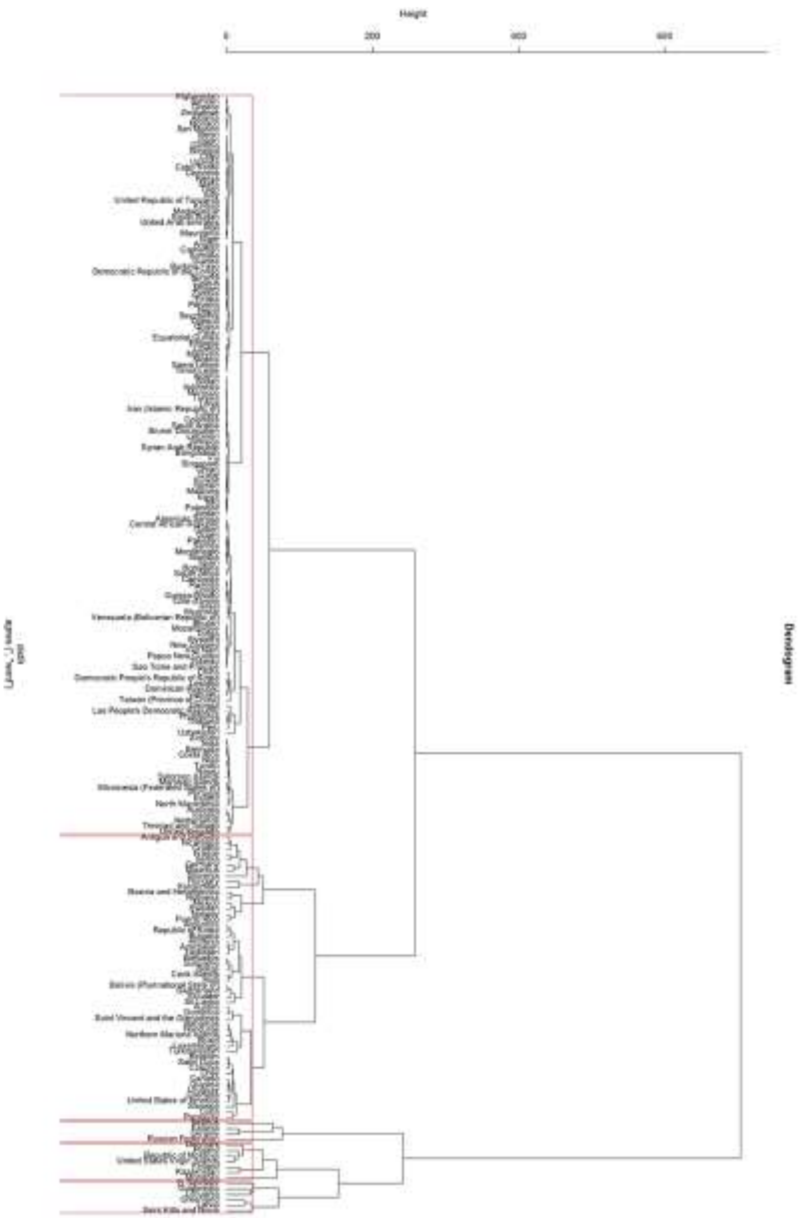


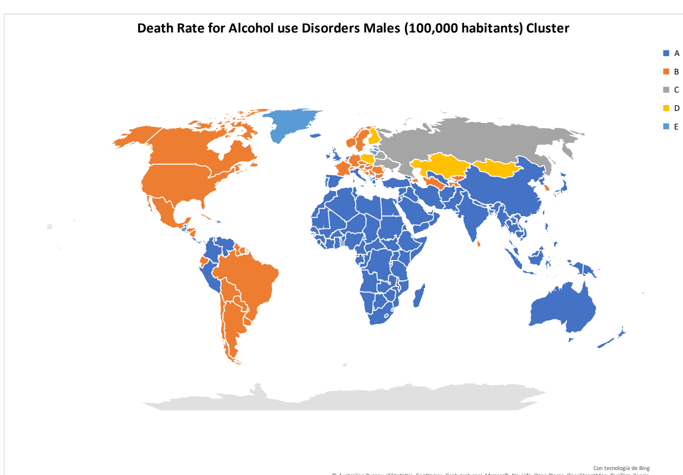
Death Rate Alcohol use disorders females (per 100,000 people, all ages).



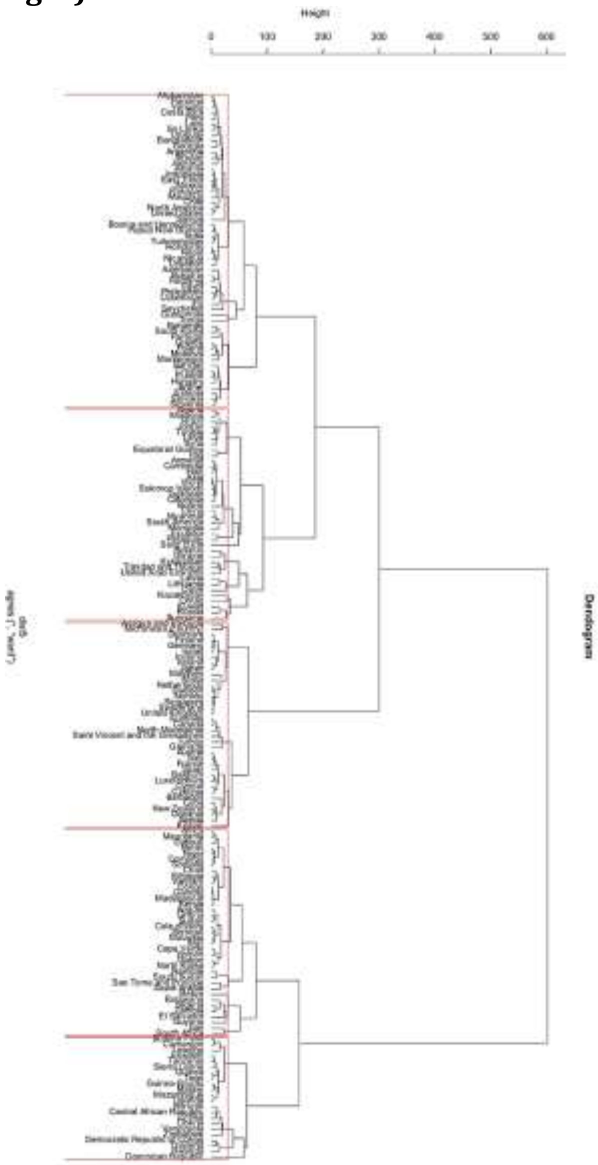


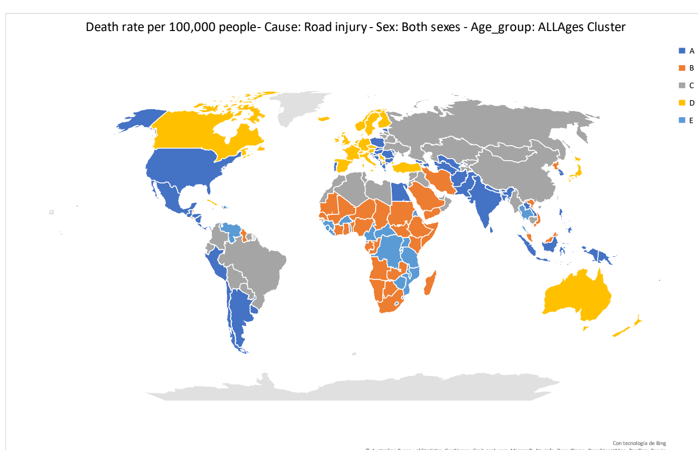
Dendrogram 7: Death Rate Alcohol use disorders males (per 100,000 people, all ages).





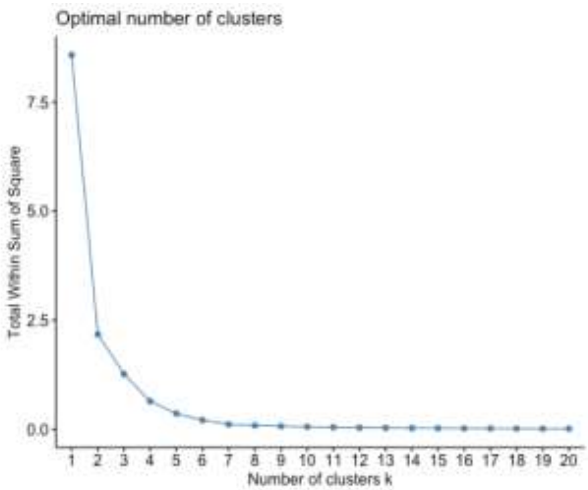
Road injury (Death rate per 100,000 people, both sexes, all Ages).



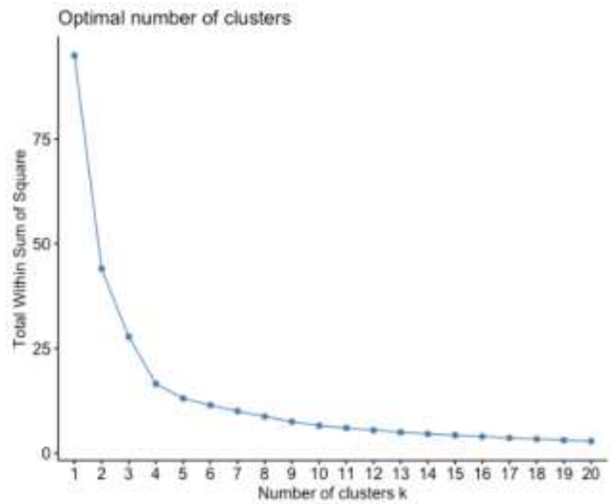


Annex 5. Optimal number of Clusters Elbow Graphs

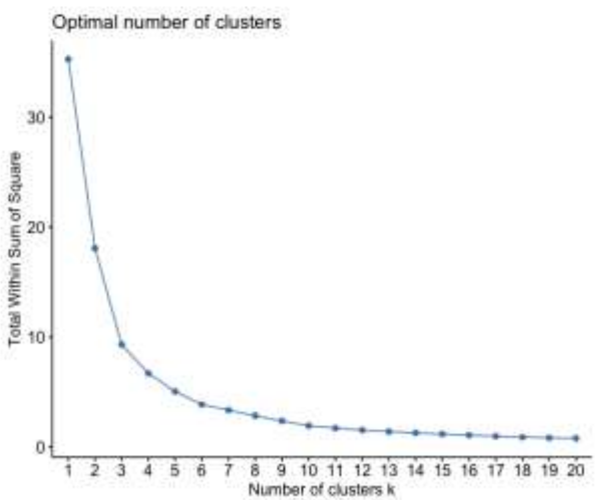
No Low Alcohol Beverages optimal number of clusters (6)



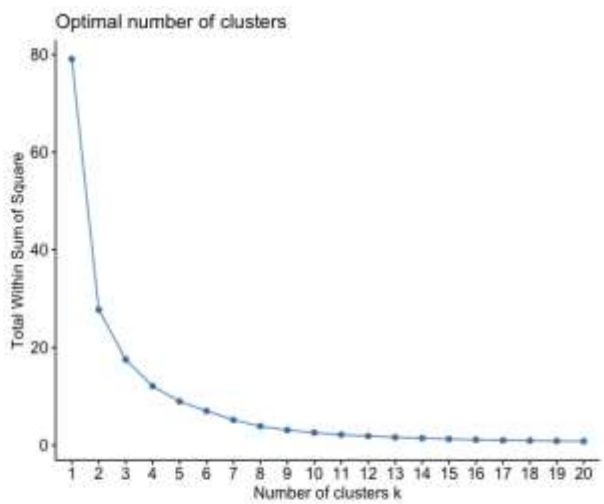
Beer optimal number of clusters (5)



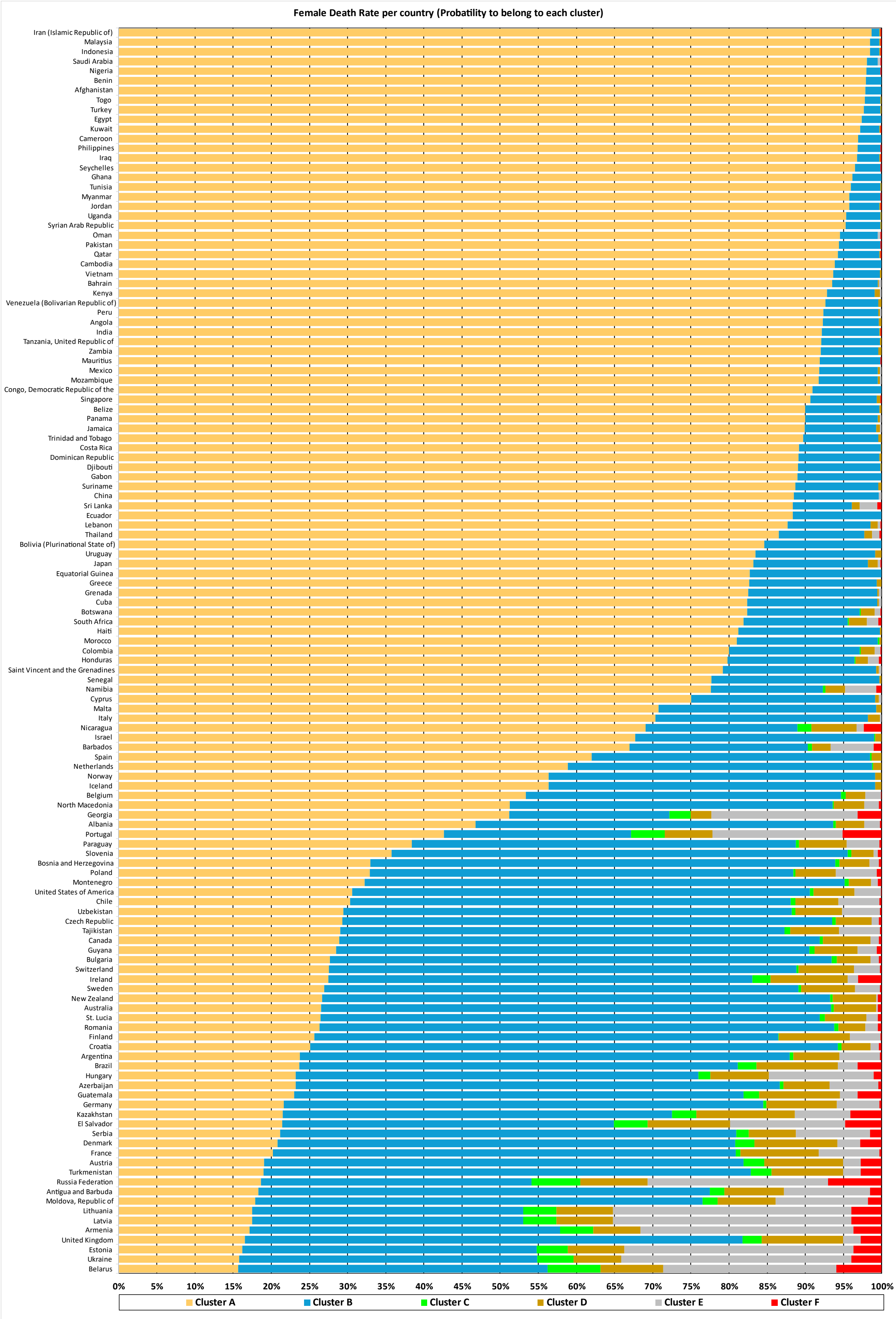
Ferments optimal number of clusters (6)

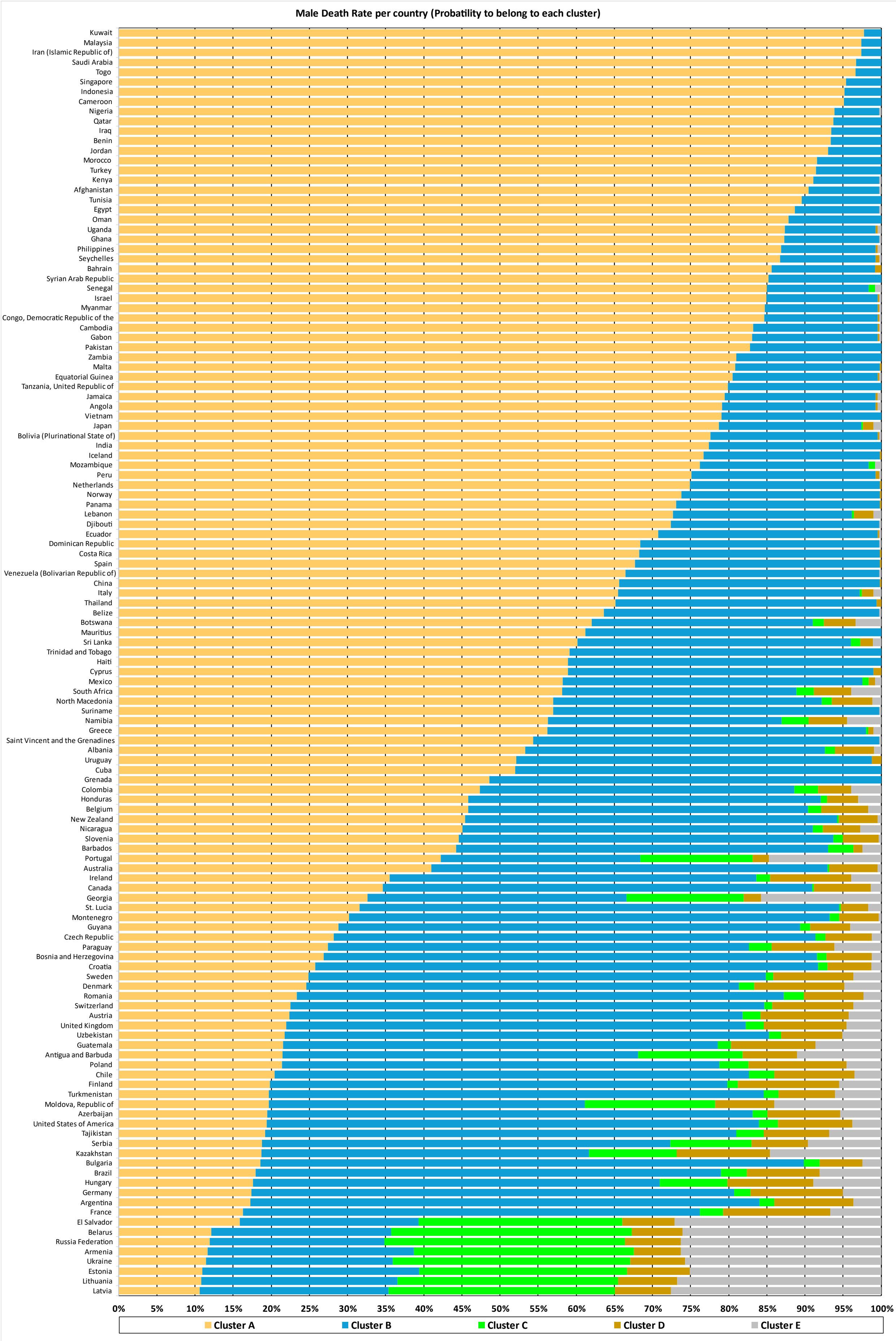


Distilled optimal number of clusters (8)

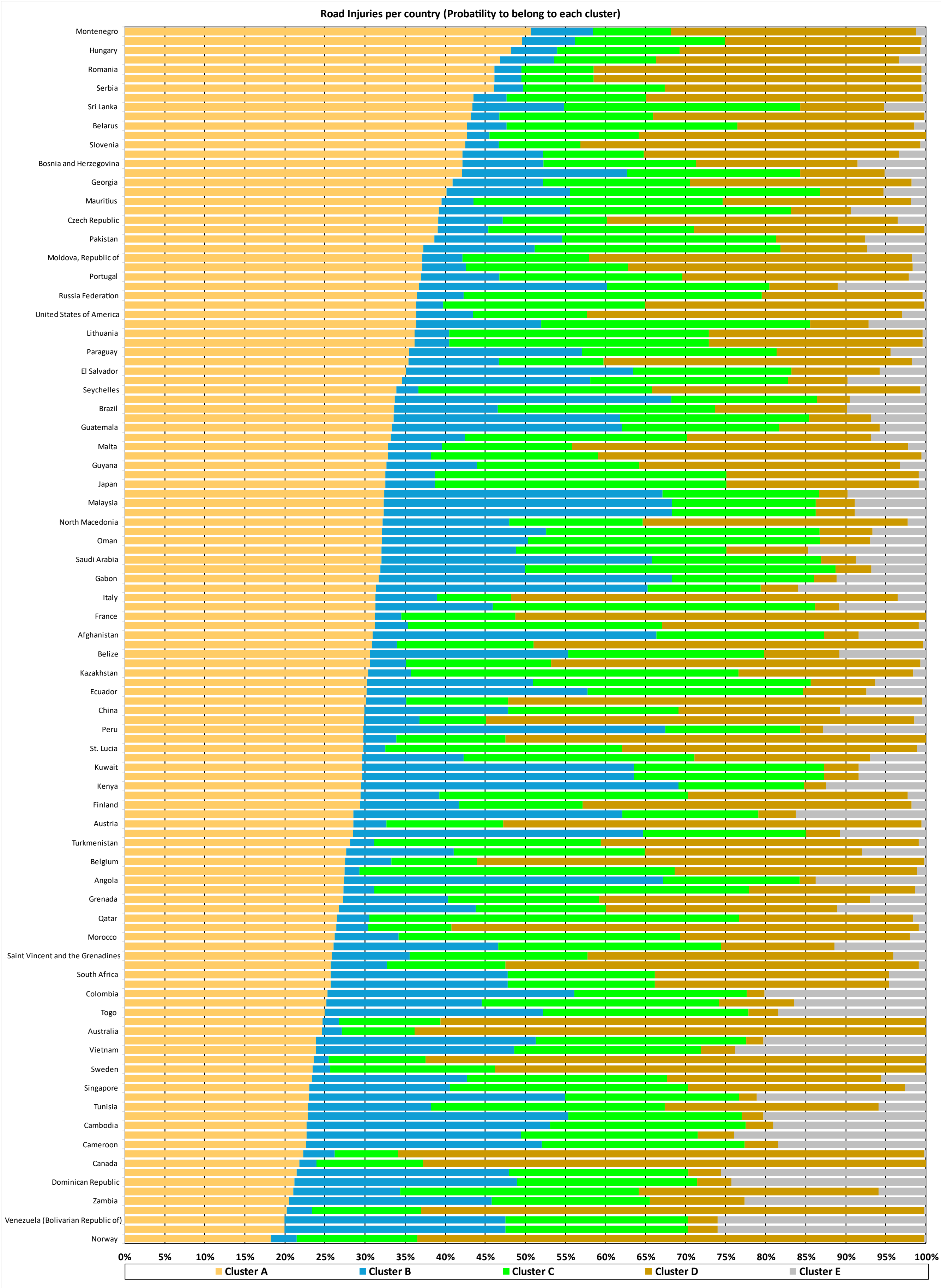


Annex 6. List of countries by probabilities.





Male Death Rate per country (Probatility to belong to each cluster)
Kuwait
Malaysia
Iran (Islamic Republic of)
Saudi Arabia
Togo
Singapore
Indonesia
Cameroon
Nigeria
Qatar
Iraq
Benin
Jordan
Morocco
Turkey
Kenya
Afghanistan
Tunisia
Egypt
Oman
Uganda
Ghana
Philippines
Seychelles
Bahrain
Syrian Arab Republic
Senegal
Israel
Myanmar
Congo, Democratic Republic of the
Cambodia
Gabon
Pakistan
Zambia
Malta
Equatorial Guinea
Tanzania, United Republic of
Jamaica
Angola
Vietnam
Japan
Bolivia (Plurinational State of)
India
Iceland
Mozambique
Peru
Netherlands
Norway
Panama
Lebanon
Djibouti
Ecuador
Dominican Republic
Costa Rica
Spain
Venezuela (Bolivarian Republic of)
China
Italy
Thailand
Belize
Botswana
Mauritius
Sri Lanka
Trinidad and Tobago
Haiti
Cyprus
Mexico
South Africa
North Macedonia
Suriname
Namibia
Greece
Saint Vincent and the Grenadines
Albania
Uruguay
Cuba
Grenada
Colombia
Honduras
Belgium
New Zealand
Nicaragua
Slovenia
Barbados
Portugal
Australia
Ireland
Canada
Georgia
St. Lucia
Montenegro
Guyana
Czech Republic
Paraguay
Bosnia and Herzegovina
Croatia
Sweden
Denmark
Romania
Switzerland
Austria
United Kingdom
Uzbekistan
Guatemala
Antigua and Barbuda
Poland
Chile
Finland
Turkmenistan
Moldova, Republic of
Azerbaijan
United States of America
Tajikistan
Serbia
Kazakhstan
Bulgaria
Brazil
Hungary
Germany
Argentina
France
El Salvador
Belarus
Russia Federation
Armenia
Ukraine
Estonia
Lithuania
Latvia
0%
5%
10%
15%
20%
25%
30%
35%
40%
45%
50%
55%
60%
65%
70%
75%
80%
85%
90%
95%
100%
Cluster A
Cluster B
Cluster C
Cluster D
Cluster E


Road Injuries per country (Probatility to belong to each cluster)
Montenegro
Hungary
Romania
Serbia
Sri Lanka
Belarus
Slovenia
Bosnia and Herzegovina
Georgia
Mauritius
Czech Republic
Pakistan
Moldova, Republic of
Portugal
Russia Federation
United States of America
Lithuania
Paraguay
El Salvador
Seychelles
Brazil
Guatemala
Malta
Guyana
Japan
Malaysia
North Macedonia
Oman
Saudi Arabia
Gabon
Italy
France
Afghanistan
Belize
Kazakhstan
Ecuador
China
Peru
St. Lucia
Kuwait
Kenya
Finland
Austria
Turkmenistan
Belgium
Angola
Grenada
Qatar
Morocco
Saint Vincent and the Grenadines
South Africa
Colombia
Togo
Australia
Vietnam
Sweden
Singapore
Tunisia
Cambodia
Cameroon
Canada
Dominican Republic
Zambia
Venezuela (Bolivarian Republic of)
Norway
0% 5% 10% 15% 20% 25% 30% 35% 40% 45% 50% 55% 60% 65% 70% 75% 80% 85% 90% 95% 100%
Cluster A
Cluster B
Cluster C
Cluster D
Cluster E

Annex 7. Disability Adjusted Life Years DALYs

Since mortality does not offer the complete picture of the morbidity endured by individuals from different populations, DALYs present a temporal measure of the years of life lost due to premature mortality produced by disorders attributed to alcohol consumption since these generate premature deaths. These bases present a time length that goes from 1990 to 2019 with a sample of 204 countries separated into two bases, one for females and the other for males.

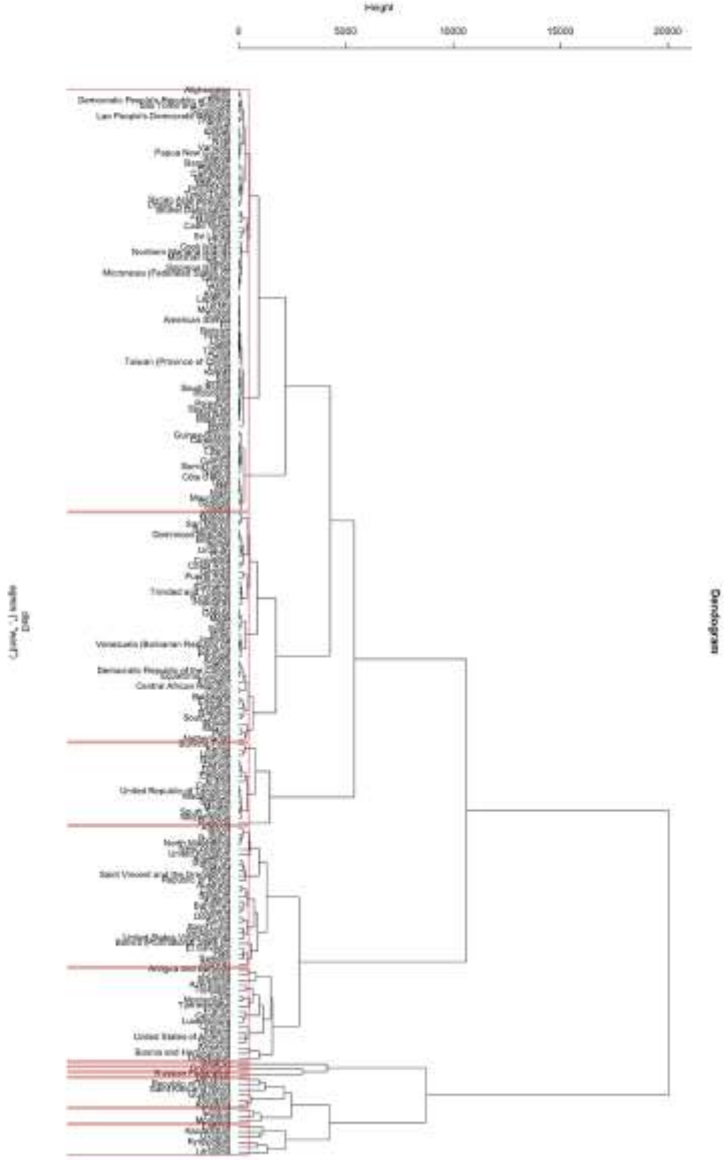
Disability-adjusted life years (commonly known as DALYs) attributed to alcohol consumption divided by gender (DALY rates from alcohol use disorders by gender) were use along with disorders and deaths associated with alcohol consumption. The database was obtained from the Institute for Health Metrics and Evaluation (IHME).[17]

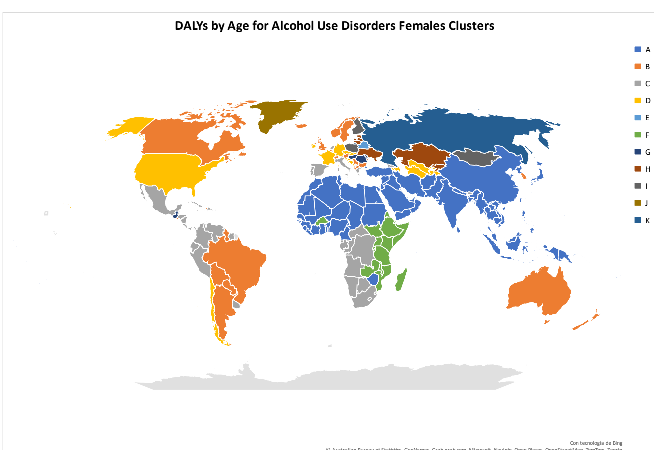
DALYs females. Age is a major determinant for the magnitude of the alcohol beverage consumption impact in females. Distance between age groups in clusters is significant with common patterns for two tier of age ranges (two groups of a range of age move locate and "move" in similar patterns). Policy focal point gains from clusterization since it indicates gender prioritization in terms of age per cluster with the potential for prioritization ordered in rankings.

DALYs males. Age is a major determinant for the magnitude of the alcohol beverage consumption impact in males in half of the clusters only. Distance between age groups in clusters is significant in 3 clusters with common patterns for two tier of age ranges. Policy focal point gains from clusterization since it indicates gender prioritization in terms of age per cluster with the potential for prioritization ordered in rankings.

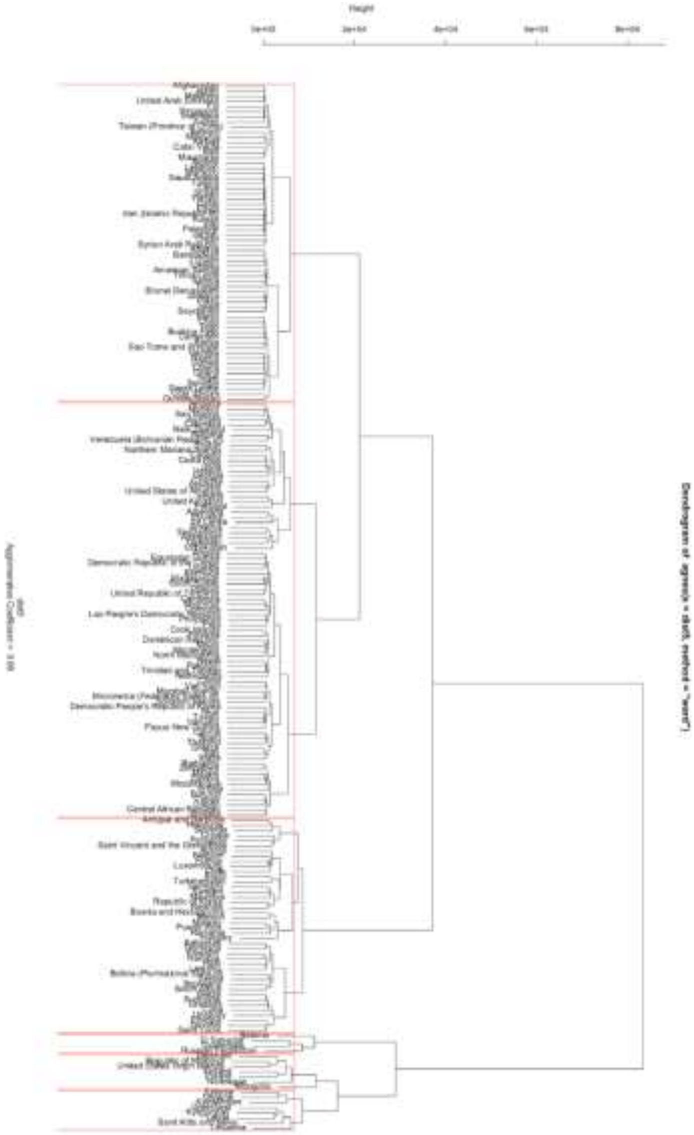
[17] https://www.healthdata.org/

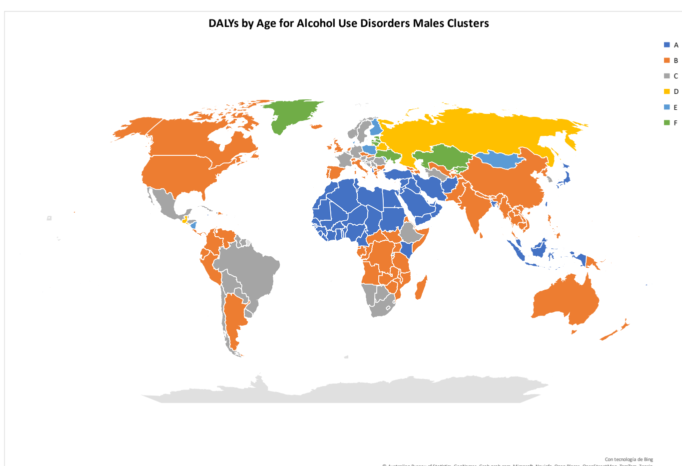
DALYs - Alcohol use disorders Females.





DALYs - Alcohol use disorders Males.





Age divide

Dendrogram 2 Age width: It is possible to build country-based clusters using DALYs per age?

1. DALYs (Disability-Adjusted Life Years) - Alcohol use disorders - Sex: Male and Females- Age: 5-14 years (Rate).
2. DALYs (Disability-Adjusted Life Years) - Alcohol use disorders - Sex: Male and Females - Age: 15-49 years (Rate).
3. DALYs (Disability-Adjusted Life Years) - Alcohol use disorders - Sex: Male and Females - Age: 50-69 years (Rate)..
4. DALYs (Disability-Adjusted Life Years) - Alcohol use disorders - Sex: Male and Females - Age: 70+ years (Rate)

DALYs - Alcohol use disorders Females.

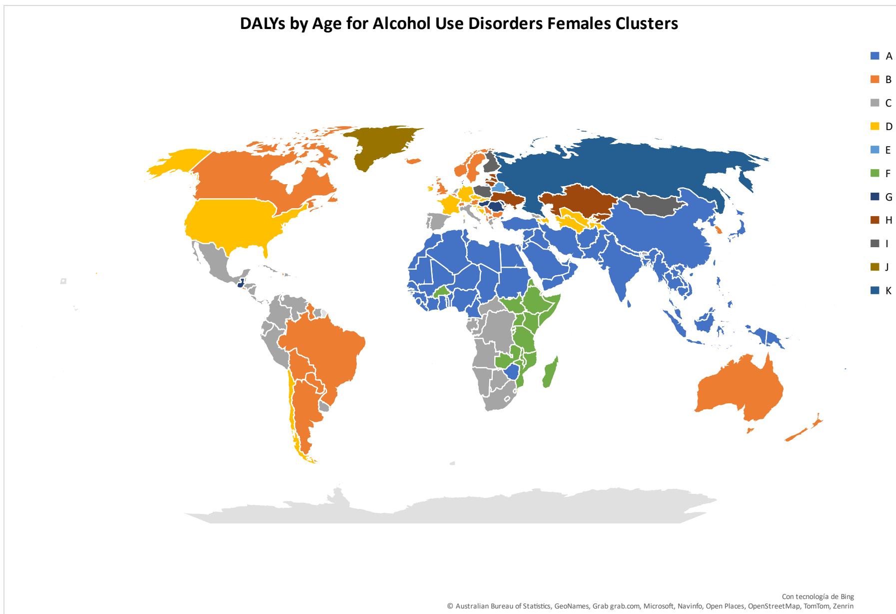
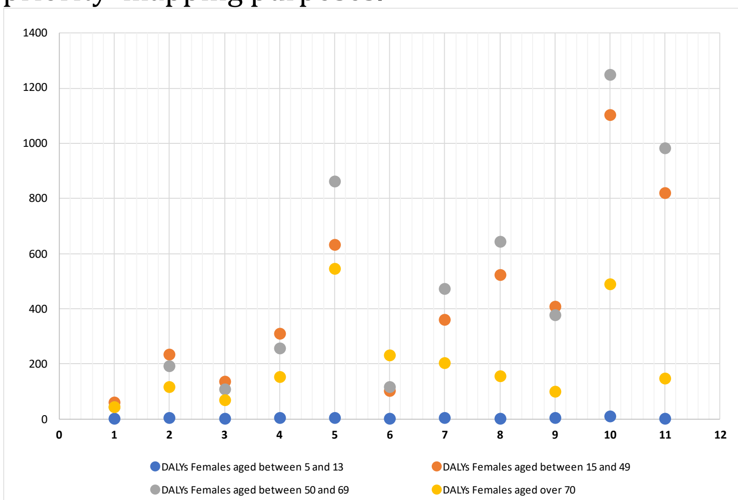


Table 2: DALYs – Female Alcohol use disorders

Cluster	DALYs Females aged between 5 and 14	DALYs Females aged between 15 and 49	DALYs Females aged between 50 and 69	DALYs Females aged over 70
1 – A	1.67	59.74	45.39	45.04
2 – B	4.45	233.93	191.53	116.25
3 – C	3.45	138.16	107.83	69.92
4 – D	5.36	309.87	258.22	152.91
5 - E	3.94	632.66	863.20	545.14
6 – F	2.43	102.50	118.51	230.83
7 - G	3.93	361.68	473.47	203.64
8 - H	3.54	522.53	642.97	156.01
9 – I	4.84	410.0	377.47	99.55
10 – J	11.98	1103.0	1250.28	488.71
11 - K	3.35	821.75	983.59	147.78

In this section, clusters are made for DALYs by age groups for women. The second column shows the averages per cluster for DALYs in women aged 5 to 14 years. Given that this group concentrates ages that are not allowed to drink, we observe low levels, cluster 10 stands out for women between 5 and 14 years old with the highest value (11.98), said cluster is comprised of Finland, Poland and Mongolia. In contrast to the consumption clusters, both Finland and Poland are located in the same cluster (5) where the consumption of fermented beverages other than beer is the highest of the consumption groups.

Cross-sectionally, cluster 10 has higher DALY values regardless of age for women. This cluster far exceeds the average values for ages 15 to 49, 50 to 69, and 70 or older. On the contrary, cluster 1 has the lowest values by age range. This cluster mainly includes Muslim countries, as well as small island countries. A better policy interphase is to visualize the product of the clusterization process as a policy framework. The following chart re organizes clusters, and within them, countries for priority' mapping purposes.



DALYs - Alcohol use disorders Males.

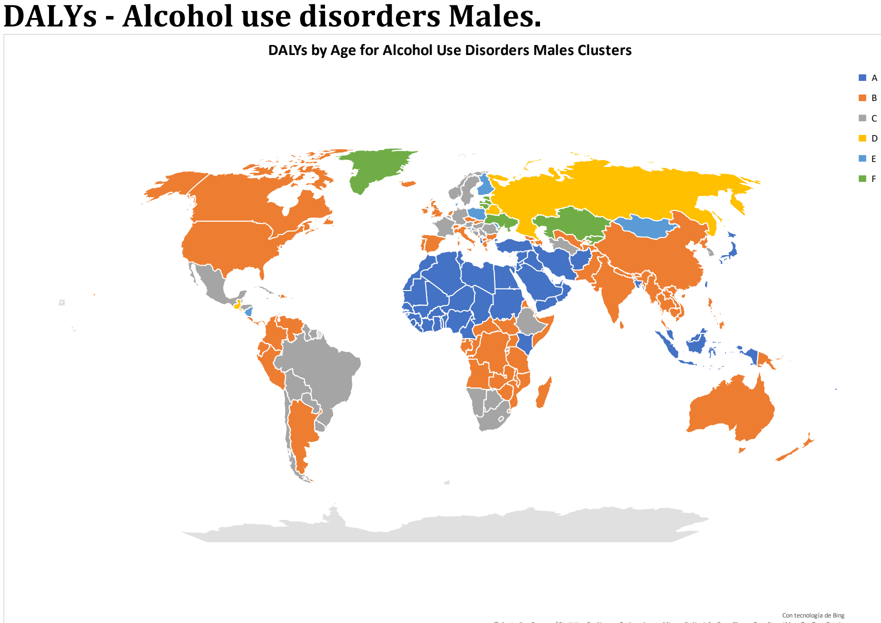
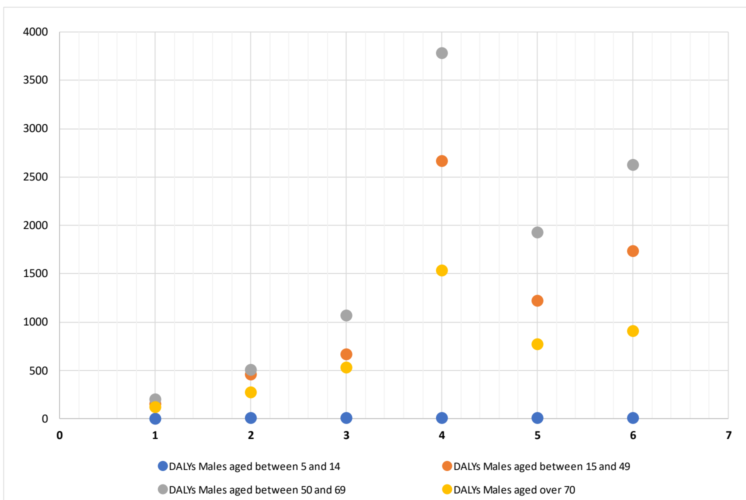


Table 3: DALYs – Male Alcohol use disorders

Cluster	DALYs Males aged between 5 and 14	DALYs Males aged between 15 and 49	DALYs Males aged between 50 and 69	DALYs Males aged over 70
1 – A	2.44	153.56	198.85	116.90
2 – B	4.63	455.89	504.03	276.89
3 – C	4.97	666.40	1067.80	527.75
4 – D	6.08	2665.32	3786.05	1536.07
5 - E	5.59	1219.34	1929.48	770.73
6 – F	6.10	1732.61	2630.62	905.13

Table three shows the averages from the previous year but now for men. 6 main clusters are identified. Cross-sectionally, it is cluster 6 that has higher values for men regardless of age. This cluster is mainly made up of Eastern European nations along with Greenland and Saint Kitts and Nevis. As for the minimum values, these are observed in cluster 1, again composed of Arab countries, small islands and some African countries. A better policy interphase is to visualize the product of the clusterization process as a policy framework. The following chart re organizes clusters, and within them, countries for priority' mapping purposes.



comunidadmexicana.org.mx
hola@comunidadmexicana.org.mx

@MEXGPR

Comunidad Mexicana de Gestión Pública para Resultados, A.C.
Calle Presidente Carranza 137
Col. Villa Coyoacán
C.P. 04000
Alcaldía Coyoacán
Ciudad de México.

Mexico City August 2024.

Comunidad Mexicana research on alcoholic beverages *

2018.

1. Análisis de la política de ingresos tributarios y no tributarios para bebidas alcohólicas en México en los gobiernos federal, estatal y municipal con un enfoque de salud pública.

2019.

2. Analysis of Tax and Non-Tax Revenue Policy for Alcoholic Beverages in Mexico at Federal, State and Municipal Levels, with a special focus on Public Health.

2021.

3. Modernización del IEPS a bebidas alcohólicas. Salud y Progresividad (en coautoría con Luis Foncerrada, Anel Rodríguez y Joaquín Sánchez).

4. Análisis de la política fiscal para bebidas alcohólicas en México. Un reporte para los 3 niveles de gobierno en el marco de la pandemia COVID 19.

2022.

5. Marco fiscal aplicable a las bebidas alcohólicas: un análisis comparativo Determinantes estructurales y comparativos fiscales y tributarios.

2023.

6. Incidencia del consumo nocivo de bebidas alcohólicas en las entidades federativas de México -- Métricas de recursos públicos, desigualdad, salud, seguridad pública, violencia y género.

2024.

7. Building global policy inferences for alcoholic beverages harmful consumption: A cluster analysis & machine learning approach.

2025 (forthcoming).

8. Behavioral strategies to reduce harmful alcohol in Mexico: a comparative approach with Germany and the United Kingdom.

* Documents are listed chronologically and originally published language.