PROJECT\_A

EDGAR

5/17/2022

This project shows the impact of hierarchical clustering on decision making.

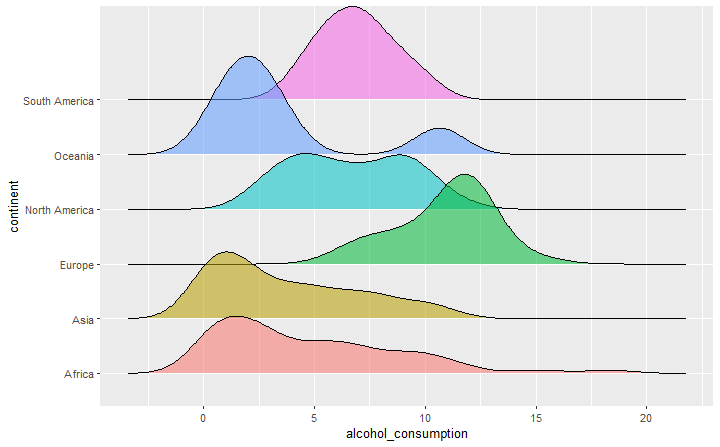
Problem

The goal of this project to identify which region of the world is more preferable for alcohol selling. Or, on other words, on which region should pay attention the company which is going to export alcohol drinks.

Approach

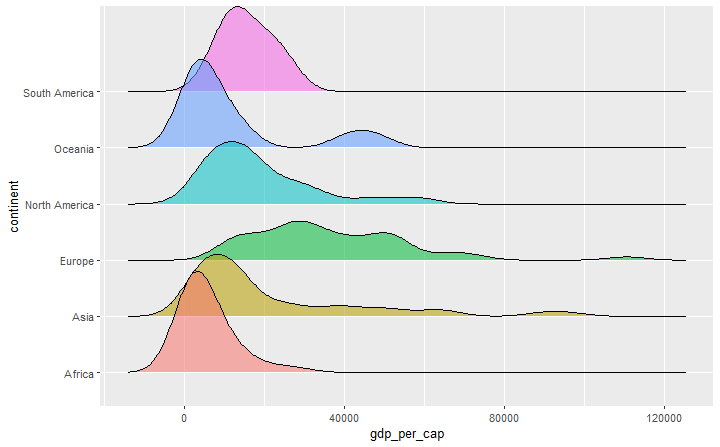
I will use three variables: GDP per capita, population and alcohol consumption for one person (data is given). I shall divide countries on four groups to identify the best cluster.

# Alcohol consumption plot



This plot shows that from six continents the best for our goal is Europe if we take as a variable only alcohol consumption. South and North Americas follow the Europe by alcohol consumption.

# GDP plot



If we take GDP per capita to compare continents we shall have this graph. And again Europe is in the first place, North America and Asia are follow Europe.

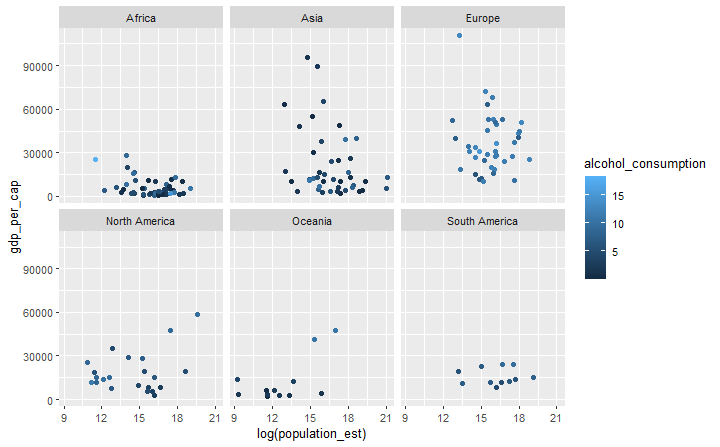
# Population plot

# 

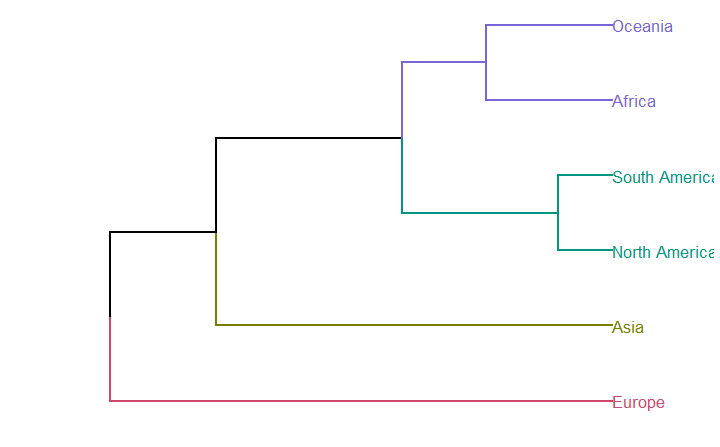
If we pay attention on this graph we shall see that Asia has the most population, then North and South America.

# Combination of three variables

Each of variables separately is important, however we need overall picture. In this graph I tried to combine all variables to show which continent is the best for our project. The most amount of light blue points (countries with high consumption) has Europe. Americas also look good, but position of Europe is much preferable.



# Hierarchical clustering



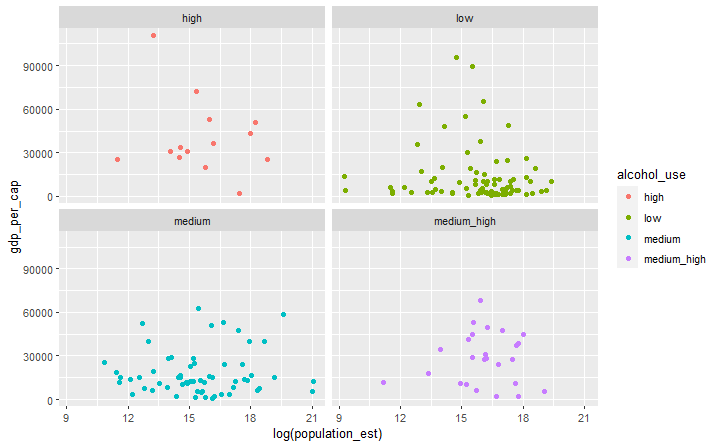
At least if we try to use hierarchical clustering we will have overall picture of continents. Europe still the first, but we can notice a strong cluster of Americas. Hierarchical clustering shows that position of Europe maybe not as strong as it seems after previous analysis.

After having some information about continents let’s look on countries to identify exact clusters which are the best for alcohol selling.

# General overview

First of all let’s see overall picture. By alcohol consumption I divided all countries on four groups high, medium\_high, medium and low.

|  |  |
| --- | --- |
| Groups | Alcohol consumption |
| high | more than 12 liters per person |
| medium\_high | from 10 to 12 liters per person |
| medium | from 5 to 10 liters per person |
| low | less than 5 liters per person |



# Which countries in 2015 had high alcohol use?

Here is the list of countries which have high alcohol consumption (higher than 12 liters per person). As we can notice 12 from 14 countries are from Europe and most of them are from central and east Europe.

1 Austria Europe

2 Bulgaria Europe

3 Czechia Europe

4 Estonia Europe

5 France Europe

6 Germany Europe

7 Ireland Europe

8 Latvia Europe

9 Lithuania Europe

10 Luxembourg Europe

11 Russia Europe

12 Seychelles Africa

13 Slovenia Europe

14 Uganda Africa

# Which countries in 2015 had medium\_high alcohol use?

If we take the group with alcohol consumption less than 12 but more than 10 liters we will have 25 countries and 16 of them from Europe.

|  |  |  |
| --- | --- | --- |
|  |  |  |
| 1 | Australia | Oceania |
| 2 | Belgium | Europe |
| 3 | Burkina Faso | Africa |
| 4 | Cyprus | Europe |
| 5 | Denmark | Europe |
| 6 | Dominica | North America |
| 7 | Finland | Europe |
| 8 | Greece | Europe |
| 9 | Hungary | Europe |
| 10 | Laos | Asia |
| 11 | Moldova | Europe |
| 12 | Mongolia | Asia |
| 13 | Montenegro | Europe |
| 14 | New Zealand | Oceania |
| 15 | Nigeria | Africa |
| 16 | Poland | Europe |
| 17 | Portugal | Europe |
| 18 | Romania | Europe |
| 19 | Slovakia | Europe |
| 20 | South Korea | Asia |
| 21 | Spain | Europe |
| 22 | Switzerland | Europe |
| 23 | Tanzania | Africa |
| 24 | Ukraine | Europe |
| 25 | United Kingdom | Europe |

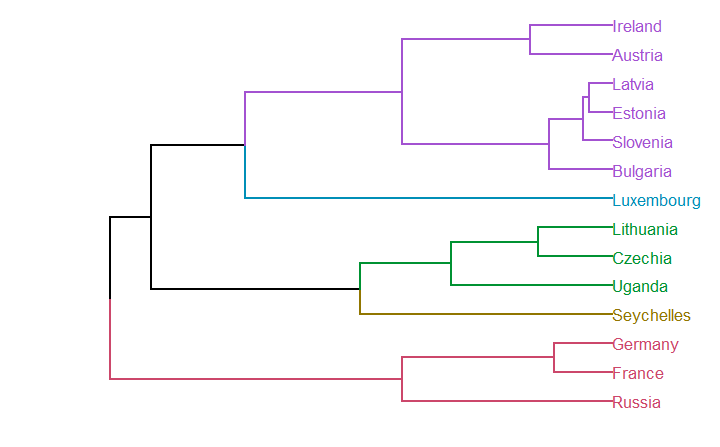
# What countries had medium and low alcohol use?

We can do the same analysis with countries which have medium (less 10 more 5) and low (less 5) consumption but it is too big data and not as important. Just several statistic numbers which may be useful for us.

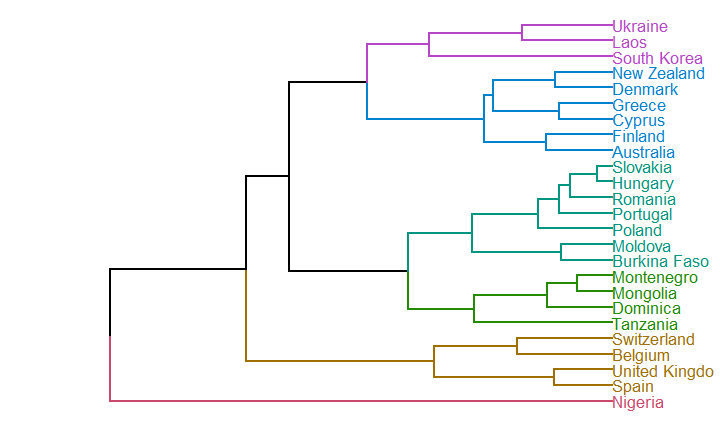
|  |  |  |
| --- | --- | --- |
|  | Number of countries | |
| continent | medium | low |
| Asia | 11 | 29 |
| Europe | 11 | 8 |
| Africa | 17 | 30 |
| Oceania | 0 | 10 |
| North America | 13 | 8 |
| South America | 9 | 2 |

# Plot cluster dendrogram for hierarchical clustering of “high” countries.

From the plot below we can see the cluster which includes Russia, Germany and France very strong and the most preferable for alcohol selling.



# Plot cluster dendrogram for hierarchical clustering of “medium\_high” countries.



Unfortunately “medium” and “low” countries are too much that’s why country names on dendogram will not be visible.

**Conclusion**

Our analysis shows that the best continent for alcohol selling is Europe and the best cluster is that includes Russia, German and France. Even more, if we pay attention we will notice that many countries from “high” and “medium\_ high” list are in the same geographical region (central and north-east Europe). The worst continent for alcohol selling is Asia (0 “high” countries, 3 “medium\_high” countries, 11 “medium” countries, and 29 “low” countries). Clustering gives as the overall picture of the countries which are similar or very close by the variables which we believe are the most important.

For example, we used to think about Lithuania, Latvia and Estonia as very similar countries and if I was alcohol distributor who is interested in that region I would think about this three countries as one market because of their soviet past, European present, geography, policy, etc. However, clustering shows me that even if they are from “high” list, Lithuania is different from its neighbors from point of view variables which I am interested the most. After this analysis, I would enter Latvia and Estonia market and then think about Slovenia and Bulgaria, instead of Lithuania, because they are from the same cluster.

Of course, for final decision companies need much more information about taxation, government policy, etc. However, this analysis shows overall picture and should be basis for further decisions.