

FOOD AND COVID-19 RELATIONSHIPS

DATA VISUALIZATION PROJECT



Prepared by

- FURKAN MERT ŞAHİN
16030411042

Submitted to:

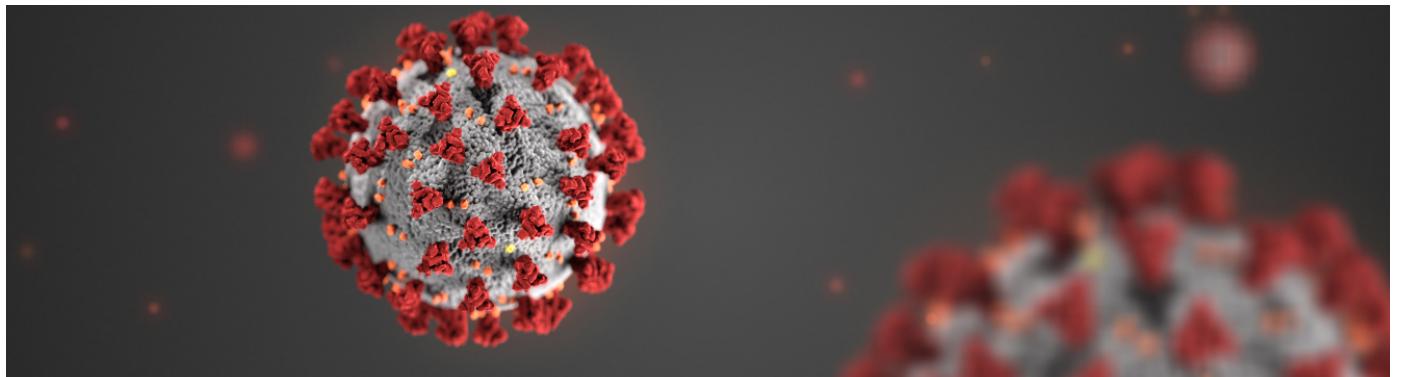
- ASSOC. PROF. HALİL KEMAL
ILTİR

Content of the Data Visualization Project

- Introduction of the Data Visualization Project
- COVID-19 & Food Datasets
- Preparation of COVID-19 & Food Dataset
- Visualization of COVID-19 and Food Datasets
- Conclusion of Data Visualization Project

Introduction of the Data Visualization Project

When the world was about to enter 2020, it encountered an epidemic that would shake people and states, and the virus that caused this epidemic was the COVID-19 virus. The virus, which emerged in China, then spread to Asia and then to Europe and spread all over the world. Millions of people caught this disease and millions of people unfortunately lost their lives. For a long time, the methods used to treat the flu were used to heal people who were sick. However, one of the most important factors that can improve the disease is nutritional habits or nutritional level.



Nutrition level is one of the most important factors that can affect human health positively or negatively. Therefore, in this data visualization project, the nutritional habits of the continents and the effects of these eating habits on recovery were analyzed.

There are a total of 6 continents with dozens of countries in the data sets, and the amount of protein, food and energy consumed by these continents, matplotlib and seaborn, which are data visualization libraries, were used and visualization processes were performed.

COVID-19 & Food Datasets

The data set below shows the percentage of products consumed by countries and continents. The dataset includes food groups such as alcoholic products, animal products, cereals, sugar and vegetables. The total number of data in the data set is 144.

```
# ----- Food Supply Quantity (kg) Intake of Countries Dataset -----
## This dataset includes percentage of food intake (kg) in countries around the world.

country_food_supply = pd.read_csv("Food_Supply_Quantity_kg_Data.csv");
country_food_supply = pd.merge(country_food_supply,worldmeter_coronavirus, on = "Country")
country_food_supply = country_food_supply.drop(["Unit (all except Population)", "Miscellaneous", "Undernourished", "Offals", "Starchy Roots", "Obesity", "Treenuts", "Oilcrops", "Meat", "Stimulants", "Sugar Crops", "Fruits - Excluding Vegetables", "Vegetal Products", "Vegetable Oils", "Eggs", "Confirmed", "Deaths", "Active"], axis=1)
country_food_supply
```

	Country	Alcoholic Beverages	Animal Products	Cereals - Excluding Beer	Sugar & Sweeteners	Vegetables	Recovered	Population	Continent
0	Afghanistan	0.0014	9.4341	24.8097	1.3489	6.7642	0.123374	38928000.0	Asia
1	Albania	1.6719	18.7684	5.7817	1.5367	11.7753	1.792636	2838000.0	Europe
2	Algeria	0.2711	9.6334	13.6816	1.8342	11.6484	0.167572	44357000.0	Africa
3	Angola	5.8087	4.9278	9.1085	1.8495	2.3041	0.056808	32522000.0	Africa
4	Argentina	4.2672	19.3454	8.4102	3.0536	4.3503	3.905192	45377000.0	South America
...
140	Vanuatu	0.2424	6.6244	6.4102	1.3645	3.9175	0.000312	321000.0	Australia/Oceania
141	Vietnam	1.4591	8.5765	16.8740	1.2846	11.9508	0.001526	96209000.0	Asia
142	Yemen	0.0364	5.7874	27.2077	5.0468	3.2135	0.004788	29826000.0	Asia
143	Zambia	5.7360	6.0197	21.1938	1.5632	3.4649	0.290524	18384000.0	Africa
144	Zimbabwe	4.0552	8.1489	22.6240	4.6485	2.3213	0.190964	14863000.0	Africa

Food Supply Quantity (kg) Intake of Continents Dataset

COVID-19 & Food Datasets

Another dataset of the data visualization project shows how much protein countries and continents get from the food groups they consume. The dataset includes food groups such as alcoholic products, animal products, cereals, sugar and vegetables. The total number of data in the data set is 144.

```
# ----- Protein Intake of Countries Dataset -----
## This dataset includes percentage of protein intake from different types of food in countries around the world.

country_protein_supply = pd.read_csv("Protein_Supply_Quantity_Data.csv");
country_protein_supply = pd.merge(country_protein_supply,worldmeter_coronavirus, on = "Country")
country_protein_supply = country_protein_supply.drop(["Unit (all except Population)","Miscellaneous","Undernourished","Offals","Starvation","Obesity","Treenuts","Oilcrops","Meat","Stimulants","Sugar Crops","Fruits - Excluding Vegetables","Vegetable Oils","Eggs","Confirmed","Deaths","Active"], axis=1)
country_protein_supply
```

	Country	Alcoholic Beverages	Animal Products	Cereals - Excluding Beer	Sugar & Sweeteners	Vegetables	Recovered	Population	Continent
0	Afghanistan	0.0000	9.7523	35.9771	0.0000	1.1370	0.123374	38928000.0	Asia
1	Albania	0.1840	27.7469	14.2331	0.0042	3.2456	1.792636	2838000.0	Europe
2	Algeria	0.0323	13.8360	26.5633	0.0000	3.1267	0.167572	44357000.0	Africa
3	Angola	0.6285	15.2311	20.3882	0.0092	0.8133	0.056808	32522000.0	Africa
4	Argentina	0.1704	31.9799	13.6702	0.0049	1.0516	3.905192	45377000.0	South America
...
140	Vanuatu	0.0078	19.8713	14.6387	0.0000	1.5847	0.000312	321000.0	Australia/Oceania
141	Vietnam	0.1555	20.4466	18.5247	0.0056	3.7216	0.001526	96209000.0	Asia
142	Yemen	0.0000	10.0122	35.1179	0.0000	0.5448	0.004788	29826000.0	Asia
143	Zambia	0.4824	9.8925	28.5182	0.0000	0.8039	0.290524	18384000.0	Africa
144	Zimbabwe	0.2929	11.3443	33.1934	0.0000	0.5955	0.190964	14863000.0	Africa

Protein Intake of Continents Dataset

COVID-19 & Food Datasets

The final dataset of the data visualization project shows how much energy (kcal) countries and continents get from the food groups they consume. The dataset includes food groups such as alcoholic products, animal products, cereals, sugar and vegetables. The total number of data in the data set is 144.

```
# ----- Energy Intake of Countries (kcal) Dataset -----
# This dataset includes percentage of energy intake (kcal) from different types of food in countries around the world.

country_kcal_supply = pd.read_csv("Food_Supply_kcal_Data.csv");
country_kcal_supply = pd.merge(country_kcal_supply,worldmeter_coronavirus, on = "Country")
country_kcal_supply = country_kcal_supply.drop(["Unit (all except Population)", "Miscellaneous", "Undernourished", "Offals", "Starchy Roots", "Obesity", "Treenuts", "Oilcrops", "Meat", "Stimulants", "Sugar Crops", "Fruits - Excluding Vegetables", "Vegetal Products", "Vegetable Oils", "Eggs", "Confirmed", "Deaths", "Active"], axis=1)
country_kcal_supply
```

	Country	Alcoholic Beverages	Animal Products	Cereals - Excluding Beer	Sugar & Sweeteners	Vegetables	Recovered	Population	Continent
0	Afghanistan	0.0000	4.7774	37.1186	2.2261	0.7504	0.123374	38928000.0	Asia
1	Albania	0.9120	16.0930	16.2107	3.4422	2.7508	1.792636	2838000.0	Europe
2	Algeria	0.0896	6.0326	25.0112	3.9869	2.0457	0.167572	44357000.0	Africa
3	Angola	1.9388	4.6927	18.3521	2.7539	0.3525	0.056808	32522000.0	Africa
4	Argentina	1.4354	14.9869	16.7927	7.0536	0.8643	3.905192	45377000.0	South America
...
140	Vanuatu	0.2668	7.4519	16.1044	2.4204	0.6099	0.000312	321000.0	Australia/Oceania
141	Vietnam	0.7150	10.9806	26.9833	2.0940	1.9578	0.001526	96209000.0	Asia
142	Yemen	0.0000	3.4667	32.0727	7.0303	0.3636	0.004788	29826000.0	Asia
143	Zambia	1.1925	3.3043	31.5528	2.5590	0.4472	0.290524	18384000.0	Africa
144	Zimbabwe	1.4269	3.9356	29.8044	5.0403	0.2532	0.190964	14863000.0	Africa

Energy (kcal) Intake of Continents Dataset

Preparation of COVID-19 & Food Dataset

COVID-19 & Food datasets need to go through a data preparation process prior to visualization. Missing observation analysis was used in the preliminary preparation of the data. There are 3 missing data in the column with the recoveries in the data set. There is no missing data in other columns.

```
country_food_supply.isnull().sum()      # Columns with Missing Data in the Food Supply Dataset
```

Country	0
Alcoholic Beverages	0
Animal Products	0
Cereals - Excluding Beer	0
Sugar & Sweeteners	0
Vegetables	0
Recovered	3
Population	0
Continent	0
dtype: int64	

Columns with Missing Data in the Food Supply Dataset

```
country_protein_supply.isnull().sum()      # Columns with Missing Data in the Protein Supply Dataset
```

Country	0
Alcoholic Beverages	0
Animal Products	0
Cereals - Excluding Beer	0
Sugar & Sweeteners	0
Vegetables	0
Recovered	3
Population	0
Continent	0
dtype: int64	

Columns with Missing Data in the Protein Supply Dataset

In order to clean up the missing data in the columns and not to spoil the significance of the data set during this process, missing data were filled in by averaging the recovered column. Since the number of missing data is low, filling with the mean approach did not impair the significance of the data sets. The filling process is done with the "fillna" method of the Pandas library, which is one of the Python libraries.

```
# Food Supply Dataset

# Since the data gap in the columns was not too large, the missing data was filled with the value assignment approach.

country_food_supply["Recovered"].fillna(country_food_supply["Recovered"].mean(), inplace=True)

country_food_supply.isnull().sum() # With the value assignment approach, missing data were filled with average data.
```

Country	0
Alcoholic Beverages	0
Animal Products	0
Cereals - Excluding Beer	0
Sugar & Sweeteners	0
Vegetables	0
Recovered	0
Population	0
Continent	0

dtype: int64

Filling in Missing Data in the Food Supply Dataset

```
# Energy (kcal) Supply Dataset

# Note: Since there are the same columns in each dataset, the same operations were performed.

country_kcal_supply["Recovered"].fillna(country_kcal_supply["Recovered"].mean(), inplace=True)

country_kcal_supply.isnull().sum() # With the value assignment approach, missing data were filled with average data.
```

Country	0
Alcoholic Beverages	0
Animal Products	0
Cereals - Excluding Beer	0
Sugar & Sweeteners	0
Vegetables	0
Recovered	0
Population	0
Continent	0

dtype: int64

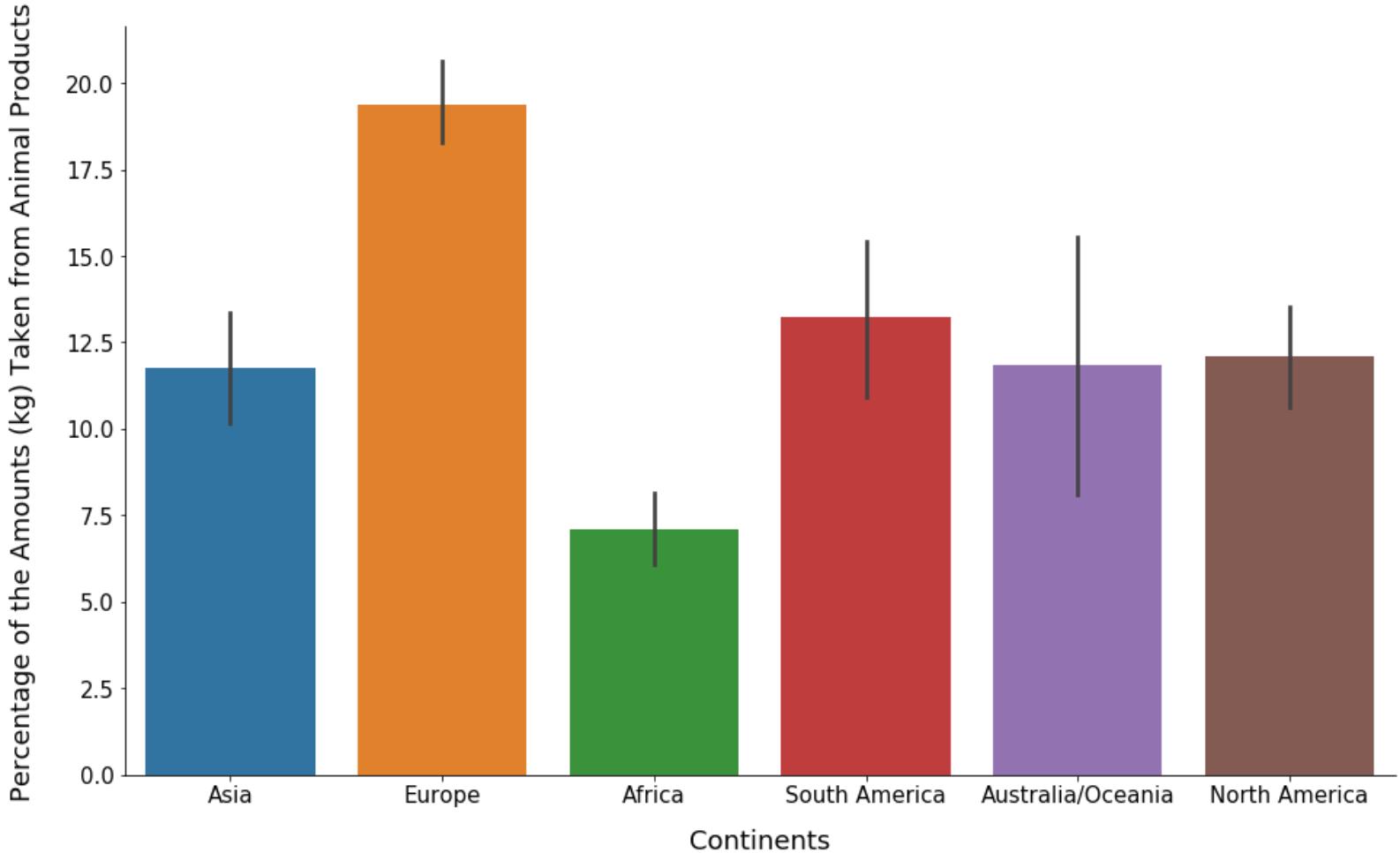
Filling in Missing Data in the Energy (kcal) Supply Dataset

Visualization of COVID-19 and Food Datasets

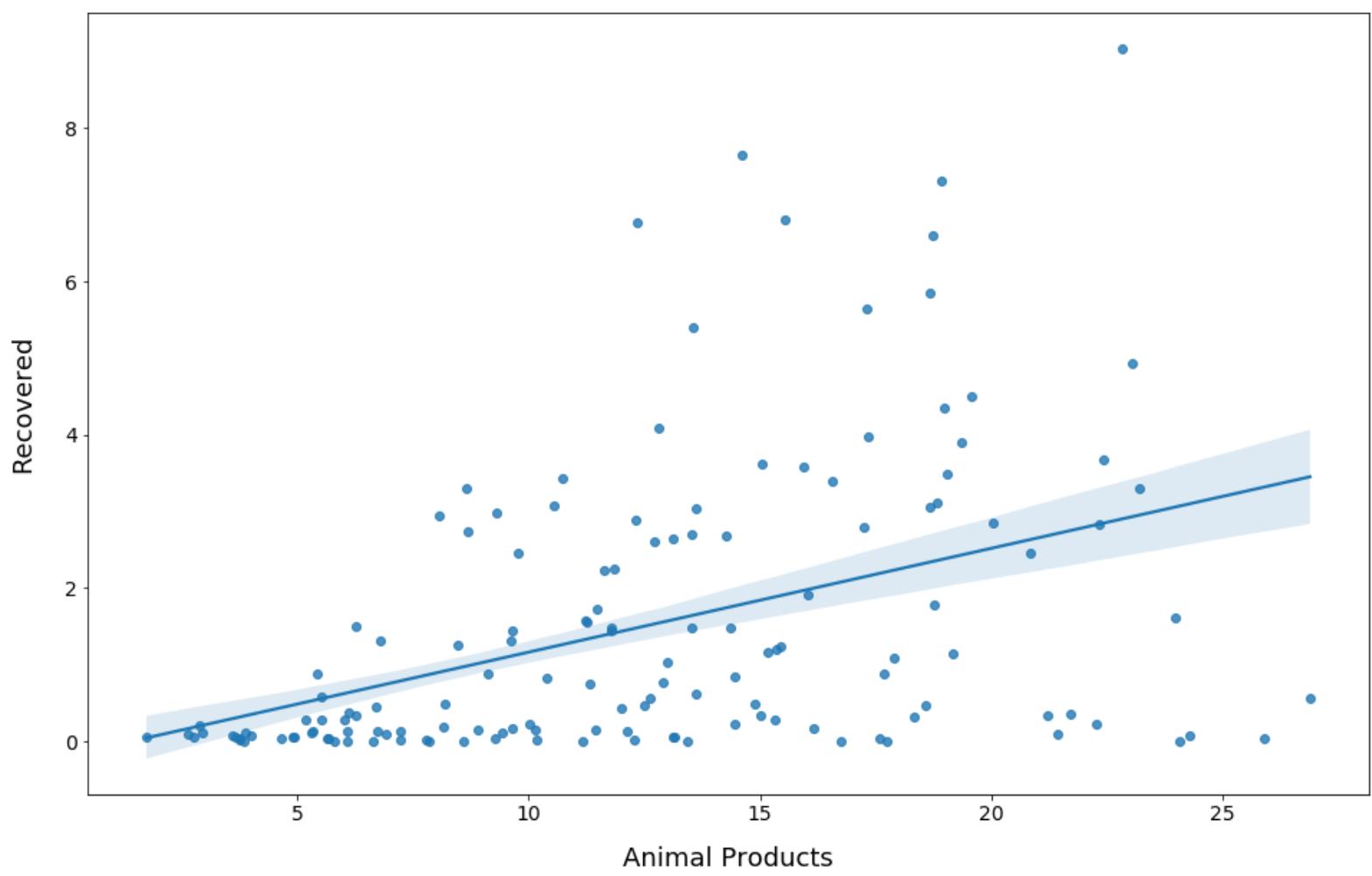
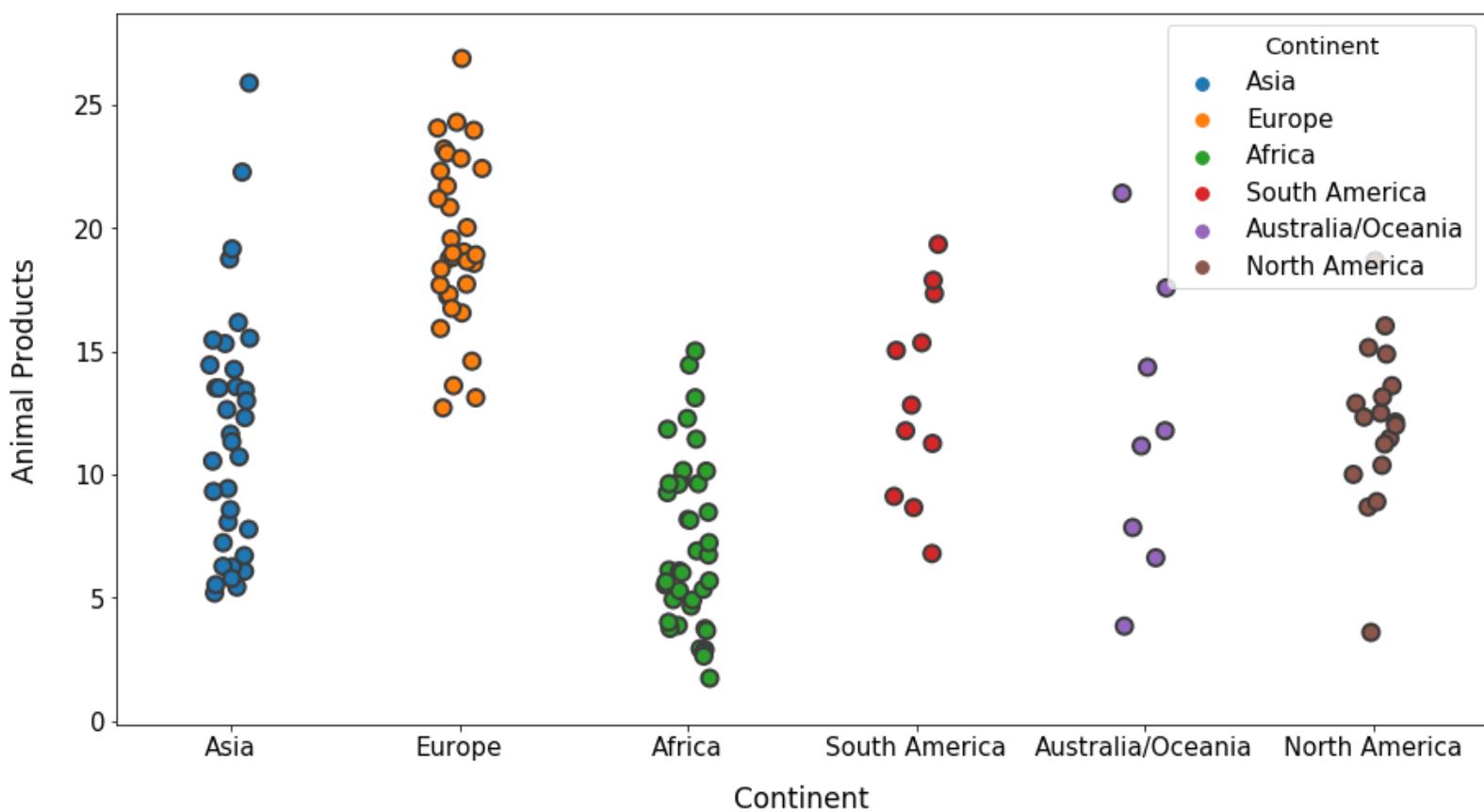
With data visualization, it has been analyzed how animal products, alcohol products, vegetables, sugar products, cereal products contribute to the healing process of people who have coronavirus disease.

1) Animal Products

The bar chart below shows the percentage of animal products consumed by the continents. In the graph, it is understood that the continent that consumes animal products the most is Europe and that the continent that consumes the least is Africa. Factors such as the amount of production and the financial situation of the society affect this picture.



Distribution Chart of Animal Products Intake by Continent (Strip Chart)

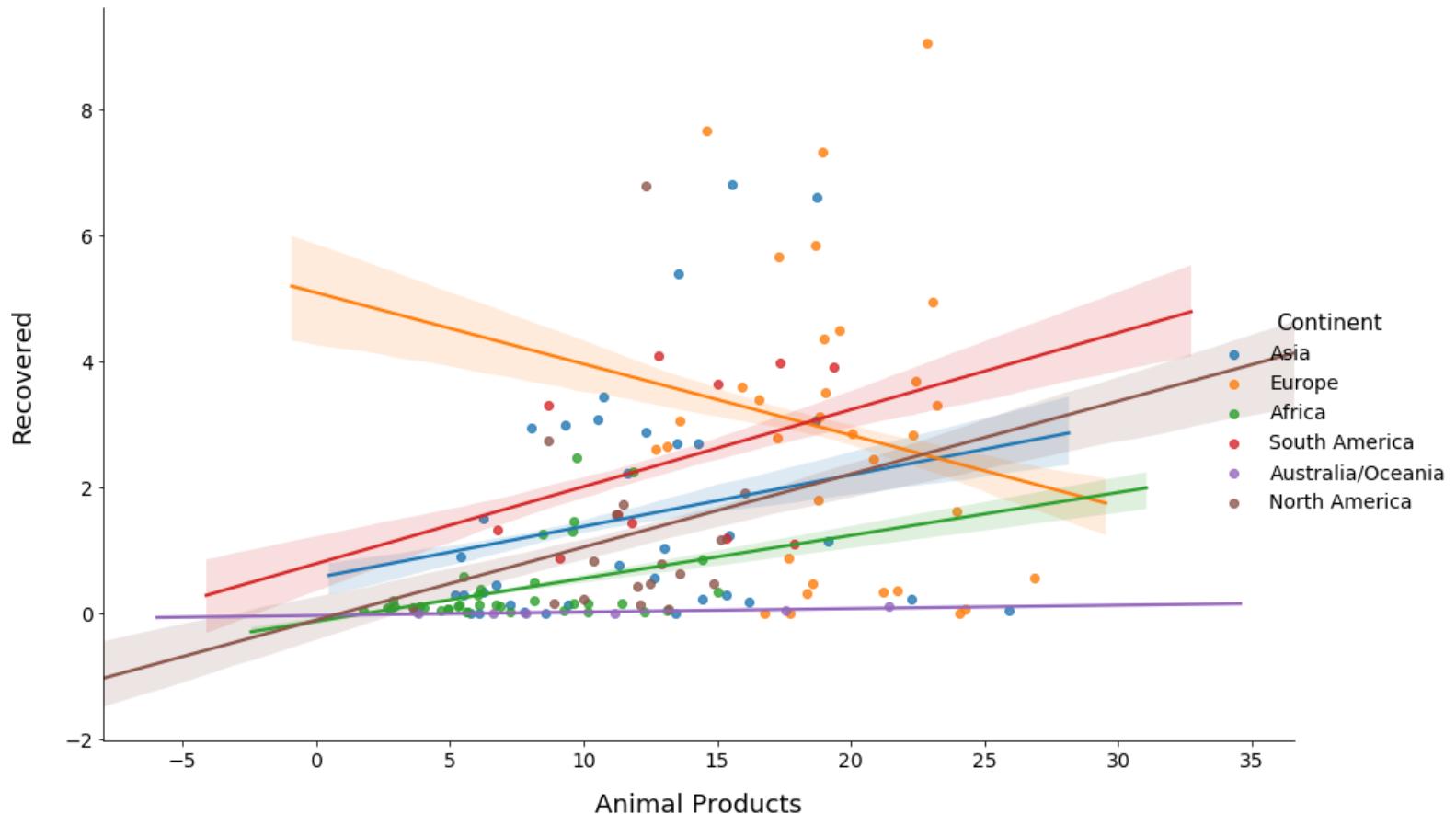


When the regression graph above is examined, it is seen that the increase in animal product consumption has a positive contribution to recovery.

The graph below is a regression graph showing the effect of the amount of animal products consumed by the continents on recovery. The graph says the impact on recovery is negative as consumption of animal products increases in Europe.

In addition, it is observed that the recovery rate of coronavirus patients increases as the consumption of animal products increases in South America, Asia, Africa and North America

Note: There are many factors that can affect coronavirus deaths. COVID-19 & Food datasets visualization studies do not give accurate results, but can provide insights on the subject.

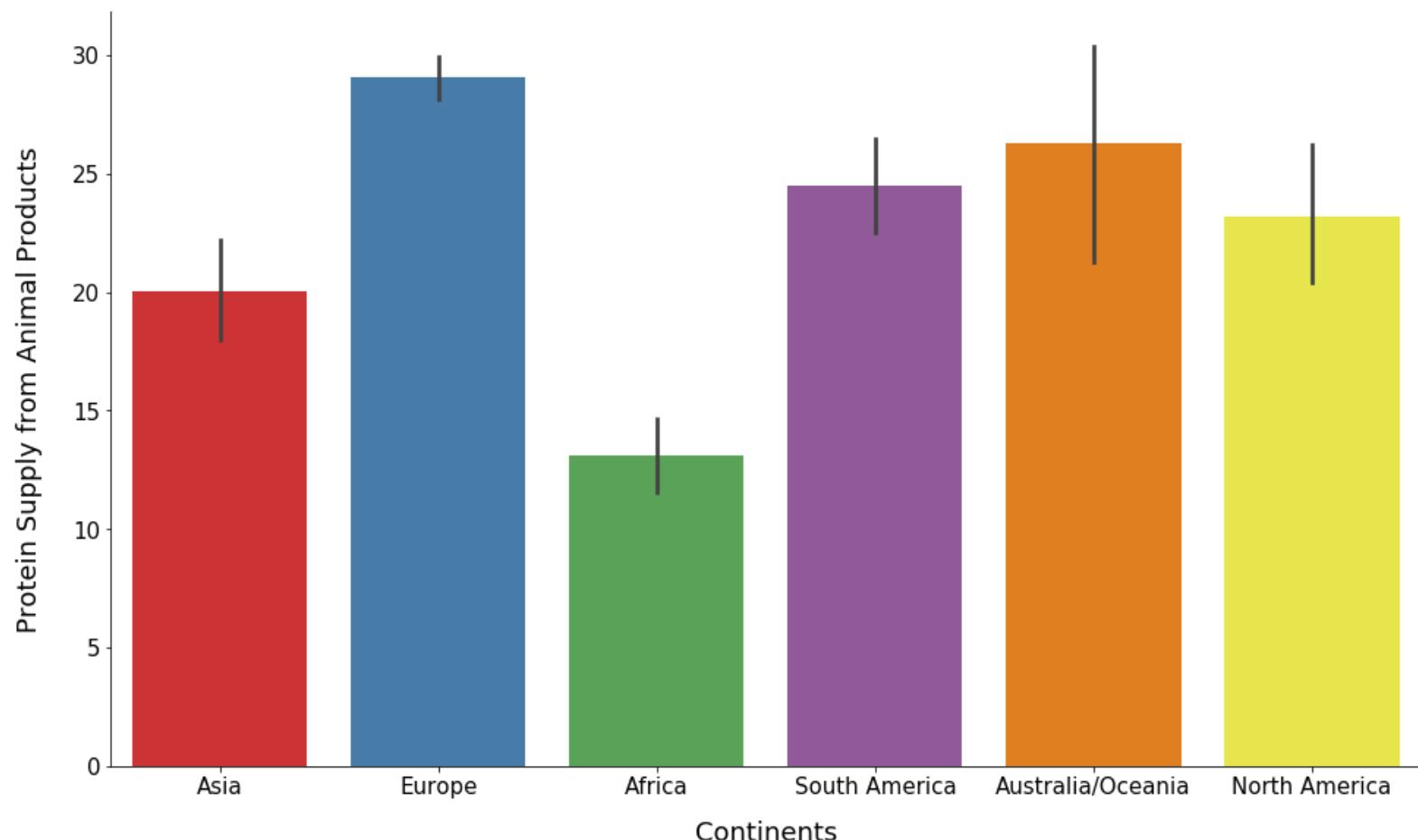


Regression Chart Showing the Effect of Animal Products Consumption on Recovery by Continents

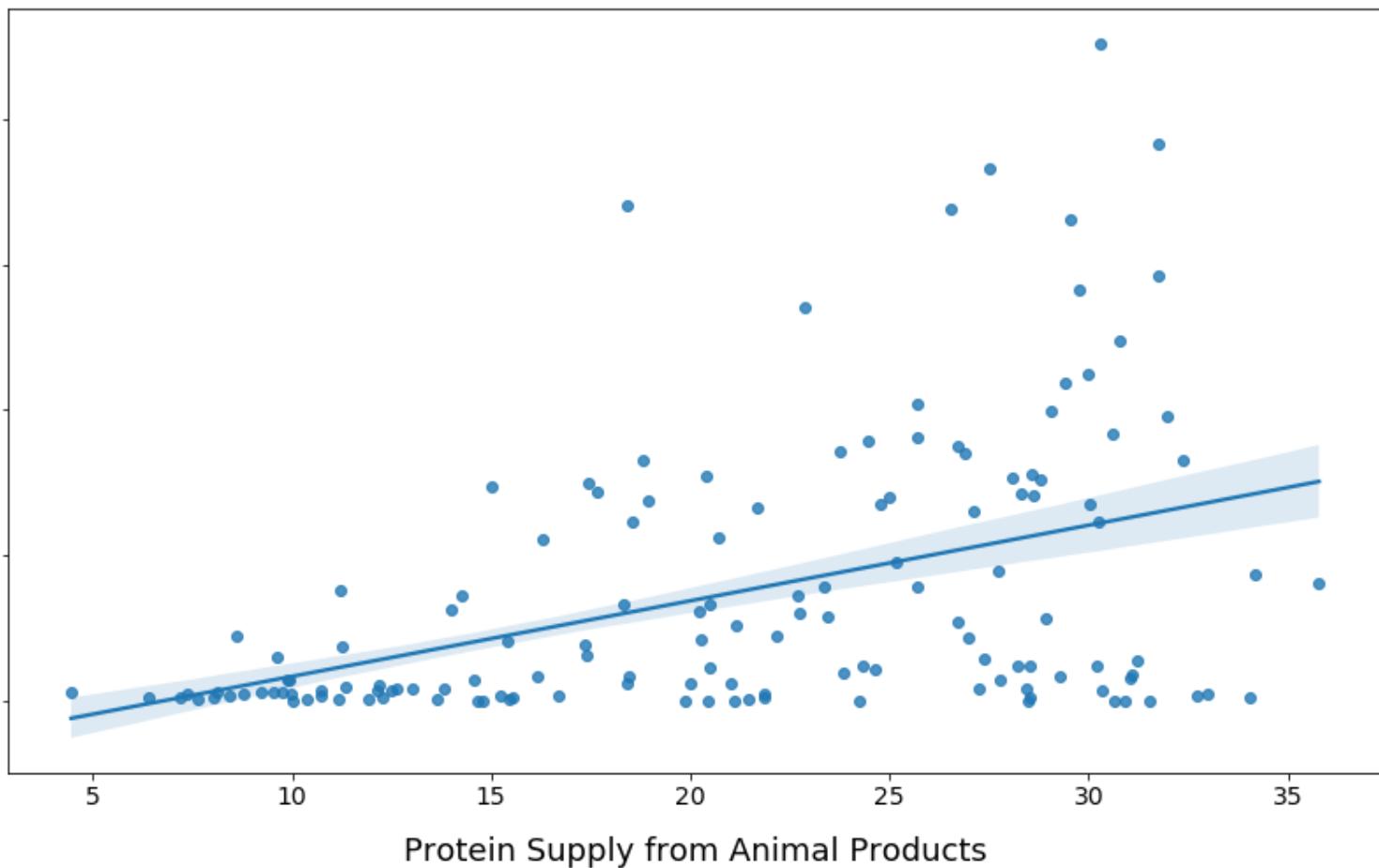
a) Percentage of Protein Supply from Animal Products

Europe, which is the continent that consumes the most animal products, is also seen as the continent that consumes the most protein from animal products.

Although South America receives more animal products than Australia/Oceania, it supplies less protein than Australia/Ocean in terms of protein intake from animal products. Africa occupies the last place in protein supply from animal products as well as in animal product consumption.

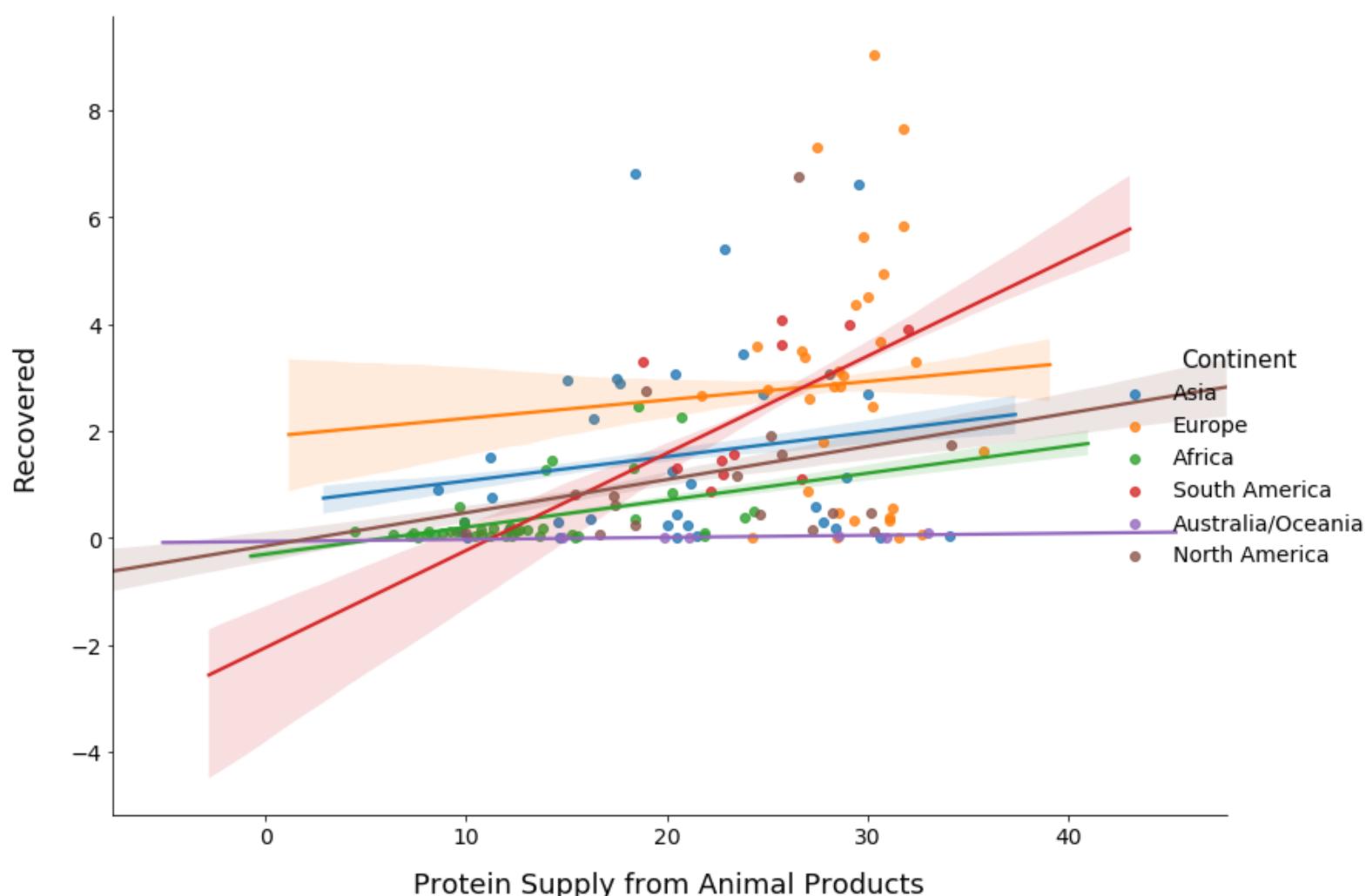


Bar Chart Showing Protein Supply from Animal Products by Continent



The Relationship Between Recovered & Protein Supply from Animal Products

As the amount of protein supplied from animal products increases, the effect of healing increases in a positive sense. However, it cannot be said that the protein supply from animal products is the only factor in recovery. It can be observed that there is only a positive relationship.



The Relationship Between Recovered & Protein Supply from Animal Products by Continents

When the effect of the protein supply from animal products on healing is examined on the basis of continents, the protein supply from animal products has increased the rate of people recovering in Europe, Africa, North America and Asia.

However, it is observed that the protein supply from animal products makes a greater contribution in the South America continent because the slope of the regression line belonging to South America is more positively inclined compared to other continents.

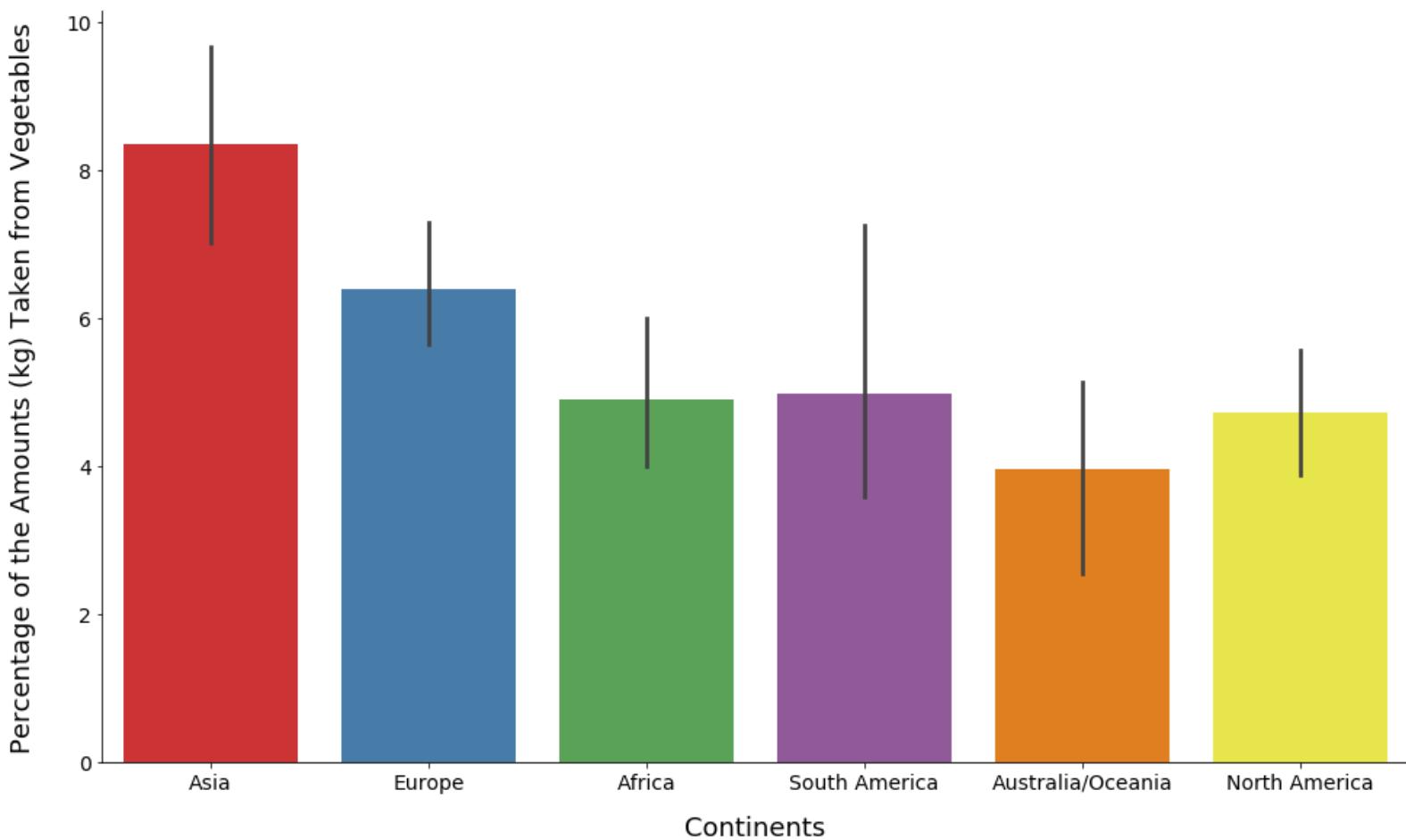
The regression line for Australia/Oceania appears to be neutral. As a result, it is understood that the protein supply from animal products does not have any effect on recovery in this continent.

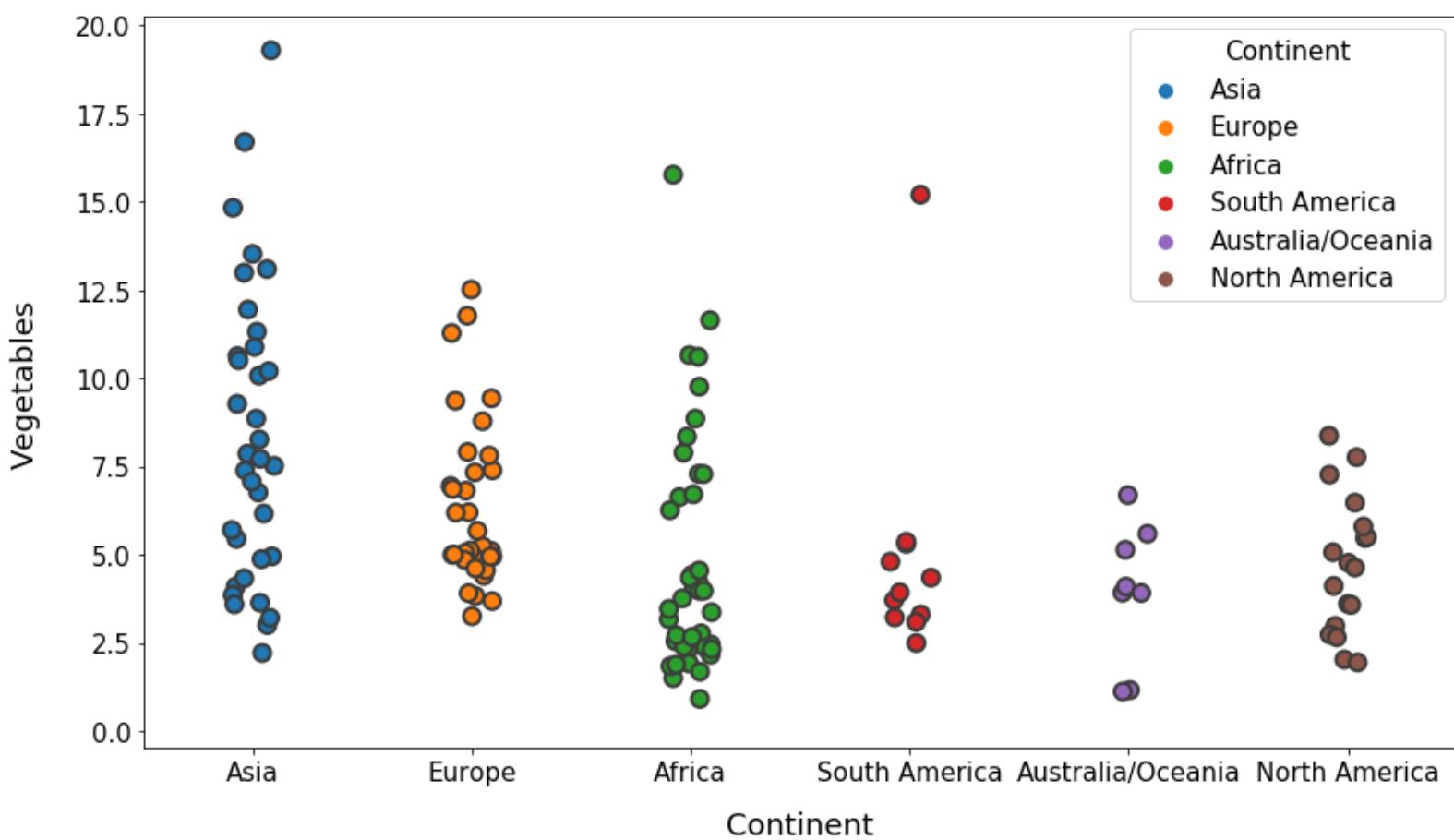
2) Vegetables

When the bar graph of the vegetable category, which is one of the most important food groups in terms of human health, is examined, the continent that consumes the most vegetables is Asia.

The continent that consumes the least vegetables is Australia/Oceania. Factors such as nutritional culture, fertile soil, agricultural production policy and purchasing power of the society may cause this graph to emerge.

The African continent, which is in the last place in the consumption of animal products, tries to meet its food needs from vegetables.

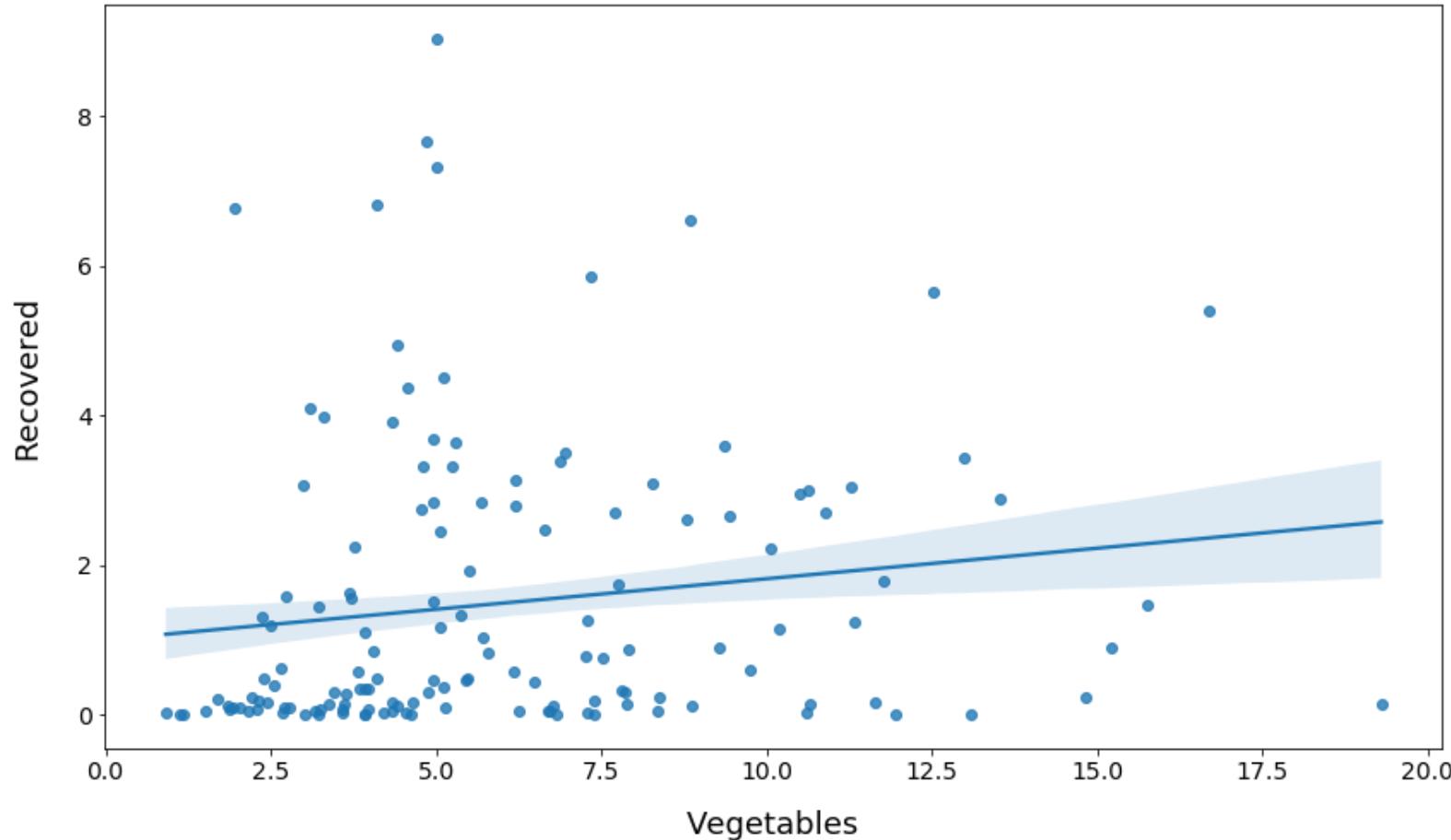




Distribution Chart of Amounts (kg) Taken from Vegetables by Continent (Strip Chart)

Consumption values in the Asian continent vary greatly in the distribution of vegetables within the continents. The same is true for the African continent. Issues such as the fact that many different cultural groups live on these continents and whether there is agricultural land to be cultivated are effective in the distribution of these data.

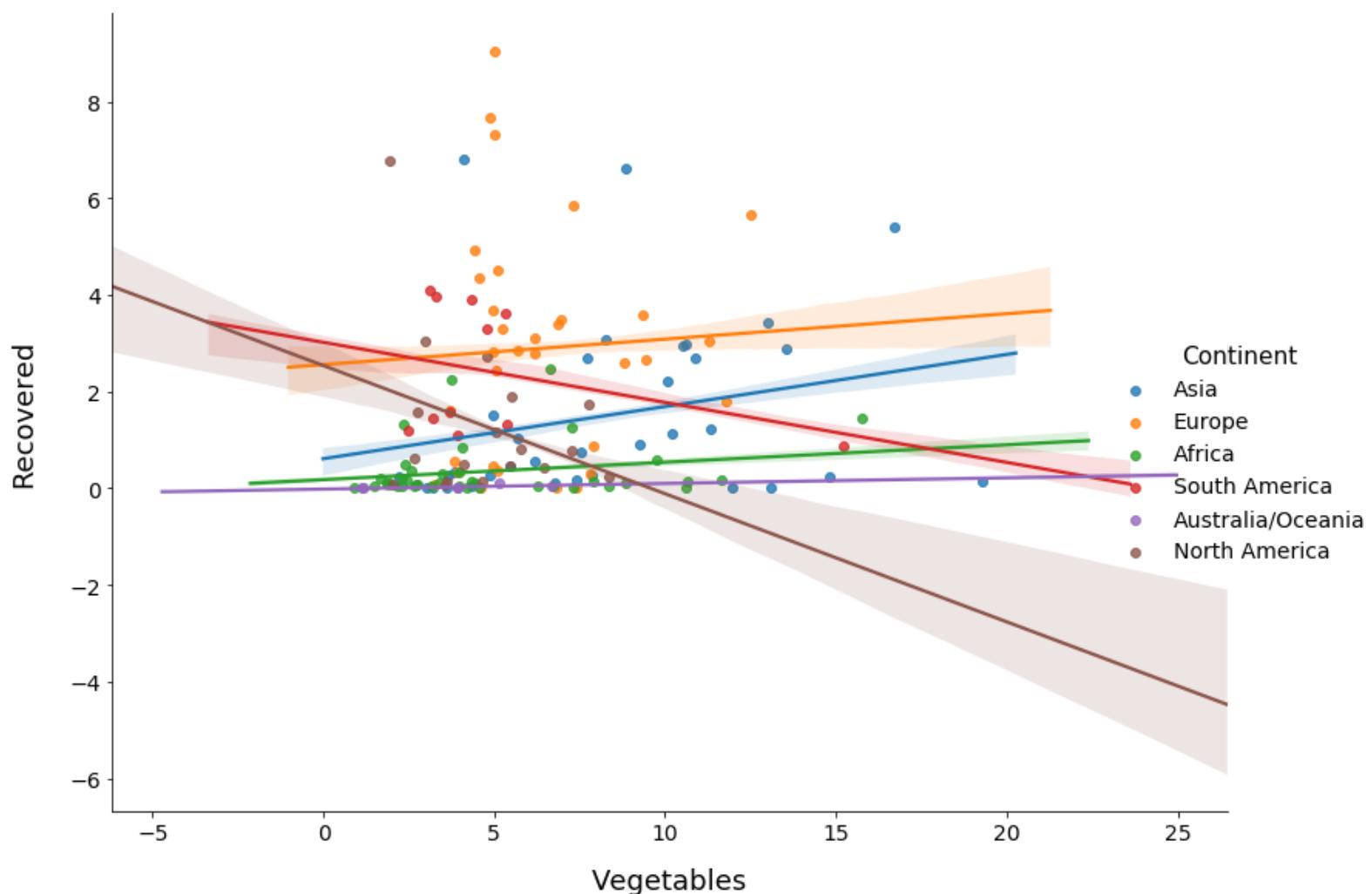
Vegetable consumption rates are similar in other continents such as Europe. Vegetable consumption rates do not vary much in these continents.



The Relationship Between Recovered & Amounts (kg) Taken from Vegetables

In the regression graph, it is clear that there is a positive effect between vegetable consumption and recovery, but this relationship is not very strong because the slope of the regression line is low.

In addition, since there is no scattered image and linear relationship in the distribution of the data, the relationship between vegetable consumption and recovery is low.



The Relationship Between Recovered & Amounts (kg) Taken from Vegetables by Continents

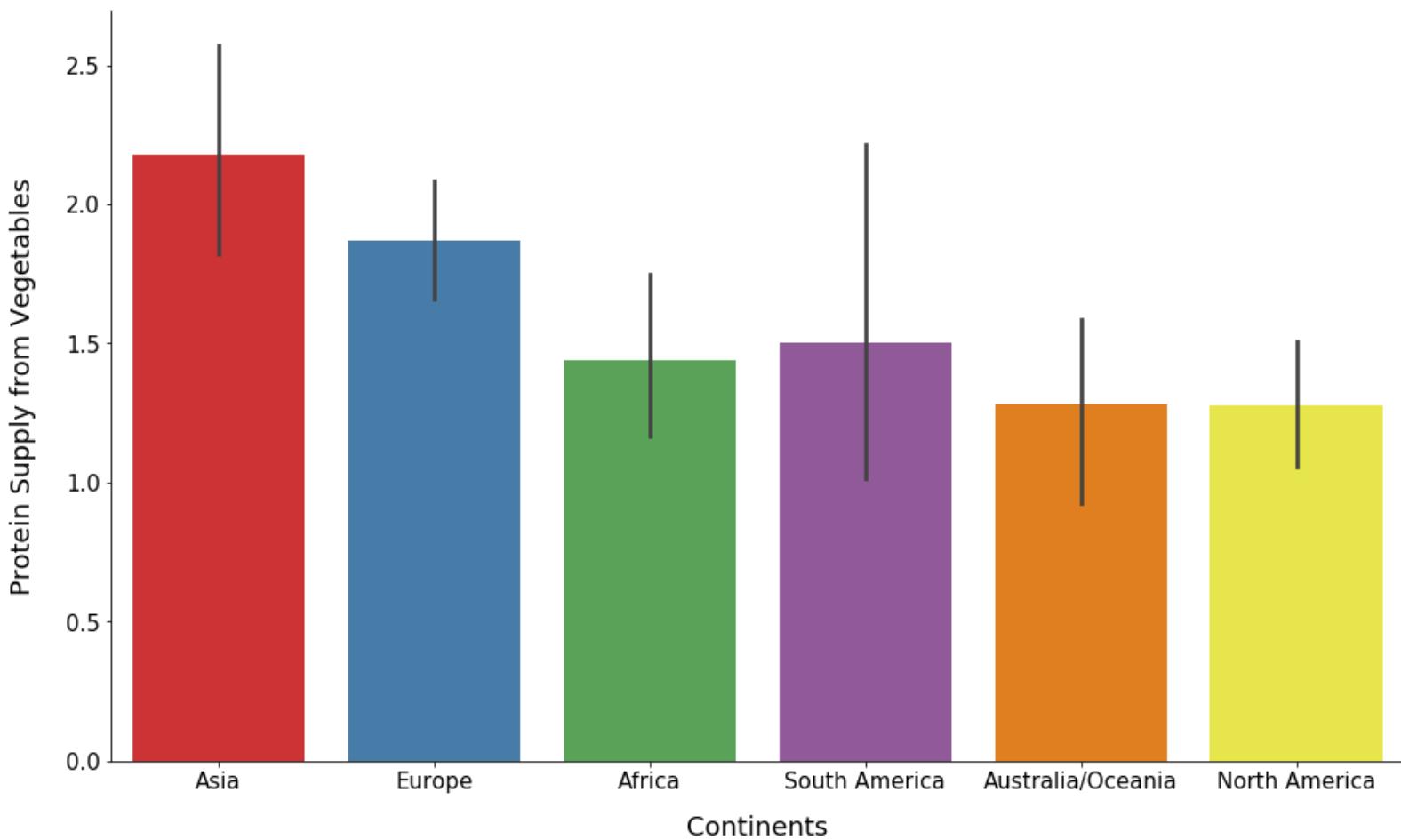
When examining the relationship between vegetable consumption and recovered on a continent basis, it is understood that recovery rates decrease when the amount of vegetable intake increases in North America and South America. However, this does not necessarily mean that vegetables are harmful. This result may occur because there are not enough data and there are many conditions that affect deaths rates.

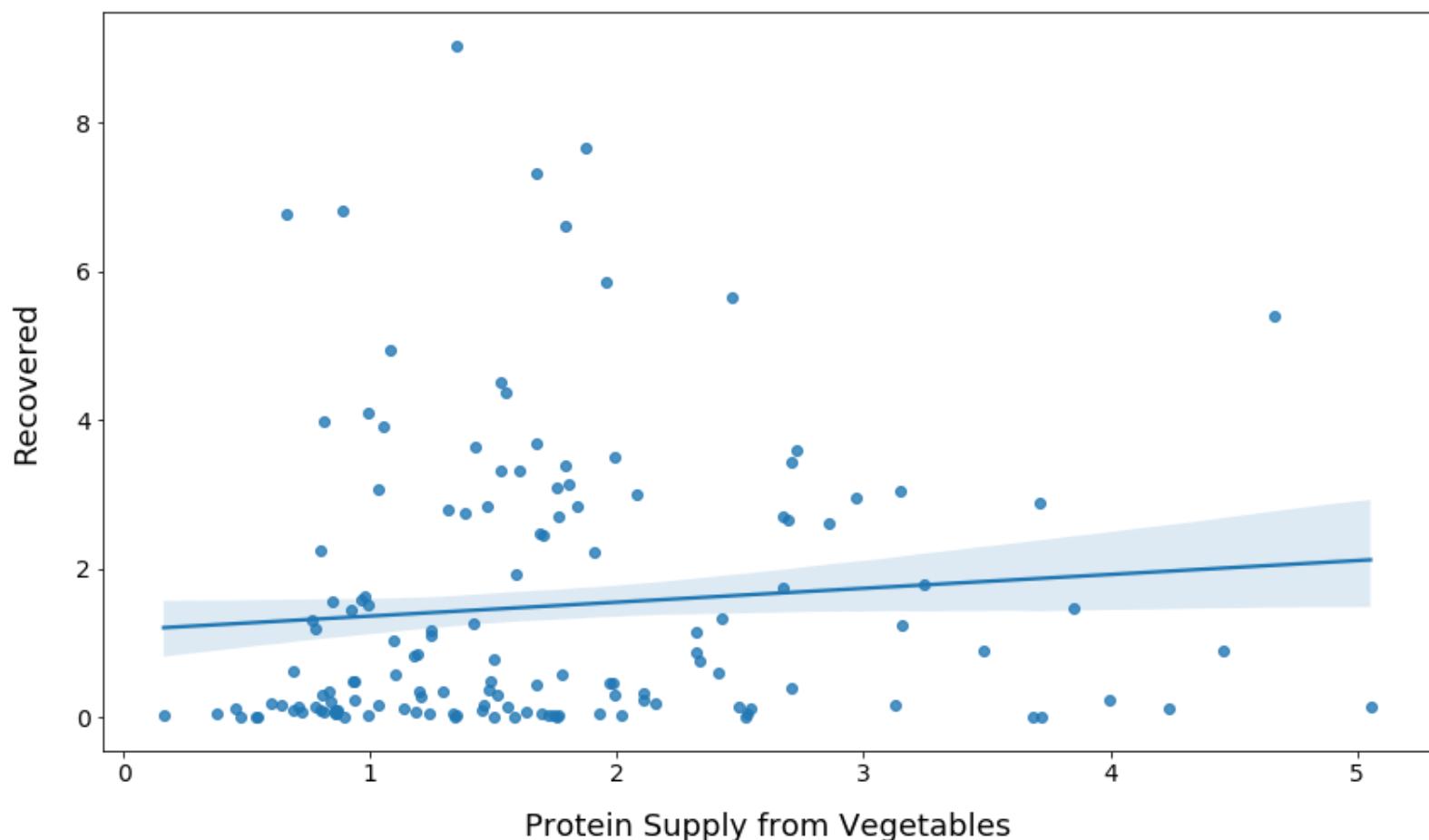
The increase in vegetable consumption in Asia and Europe causes an increase in recovery rates. In Africa and Australia/Oceania, the effect is less than in other continents.

a) Percentage of Protein Supply from Vegetables

There is a great similarity between the results in the bar graph of protein supply from vegetables and the bar graph of consumption of vegetables. As in consumption, the Asian continent is the continent with the highest protein intake compared to other continents in protein supply from vegetables.

The only difference in this chart is that although North America consumes more vegetable products than Australia/Oceania, both continents receive equal amounts of protein in protein supply from vegetables.

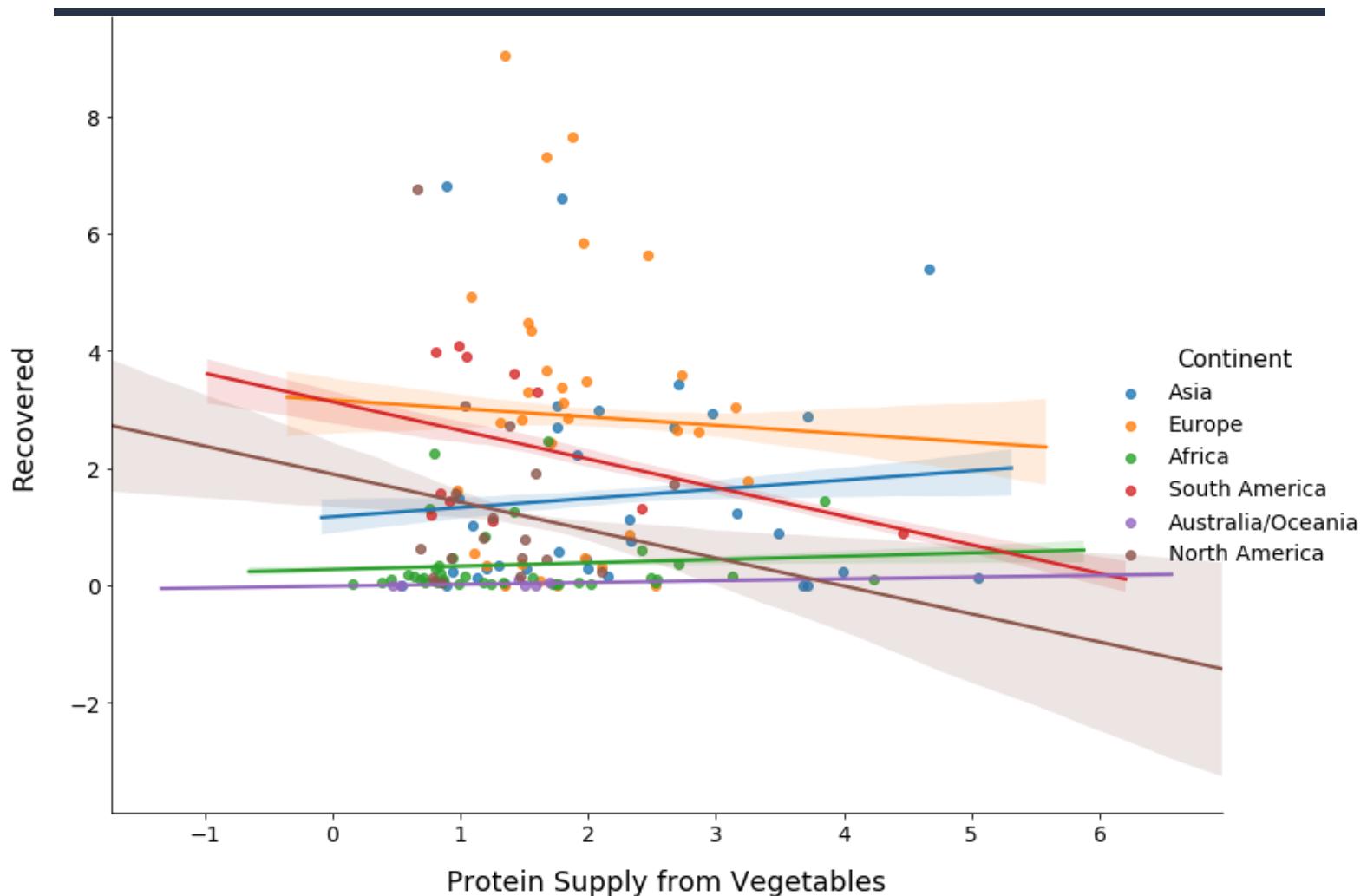




The Relationship Between Recovered & Protein Supply from Vegetables

In the above regression graph, the increase in protein supply from vegetables makes little positive contribution to recovery rates. As in the vegetable consumption graph, since the slope of the line in this regression graph is very low, the strength of the relationship between the two variables remains weak. At the same time, the data are not distributed linearly, which confirms this result.





The Relationship Between Recovered & Protein Supply from Vegetables by Continents

In the graph above, there is a small positive effect on recovery rates as the protein supply from vegetables increases in Asia. One of the reasons for this may be that the Asian continent consumes a lot of vegetables, but although Europe is the continent that consumes the most vegetables after Asia, the increase in protein supply from vegetables has a slightly negative effect on the recovery rates in this continent.

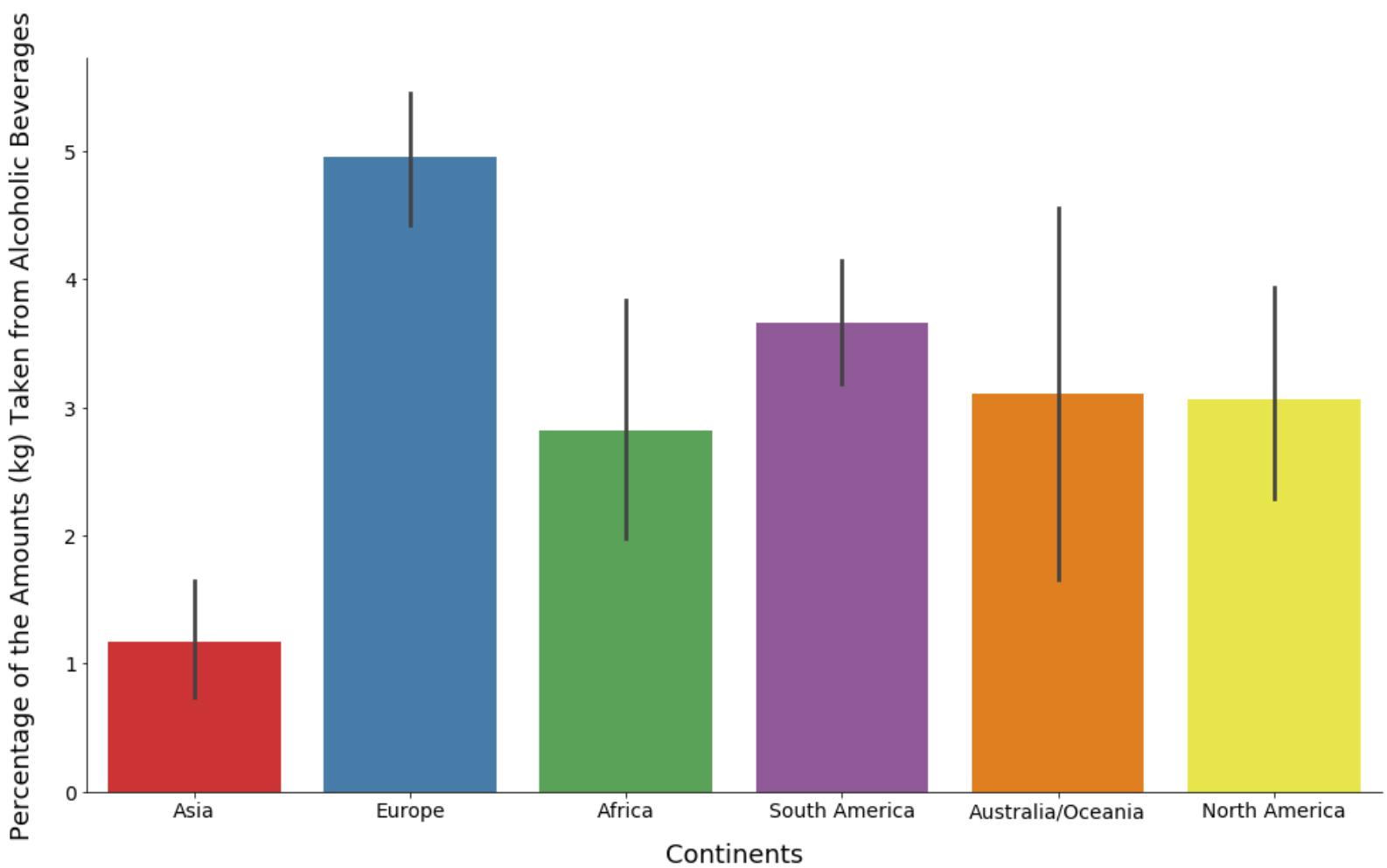
As a result, high vegetable consumption or increased protein supply from vegetables does not have a positive effect. It is clear that there are other powerful factors that influence the rate of recovery.

The situation in other continents is very different, because the increase in protein supply from vegetables in the Americas does not contribute to recovery rates.

3) Alcoholic Beverages

Europe is one of the continents that consumes the most alcoholic beverages in the category of alcoholic beverages. It is understood that the continent that consumes the most after Europe is South America.

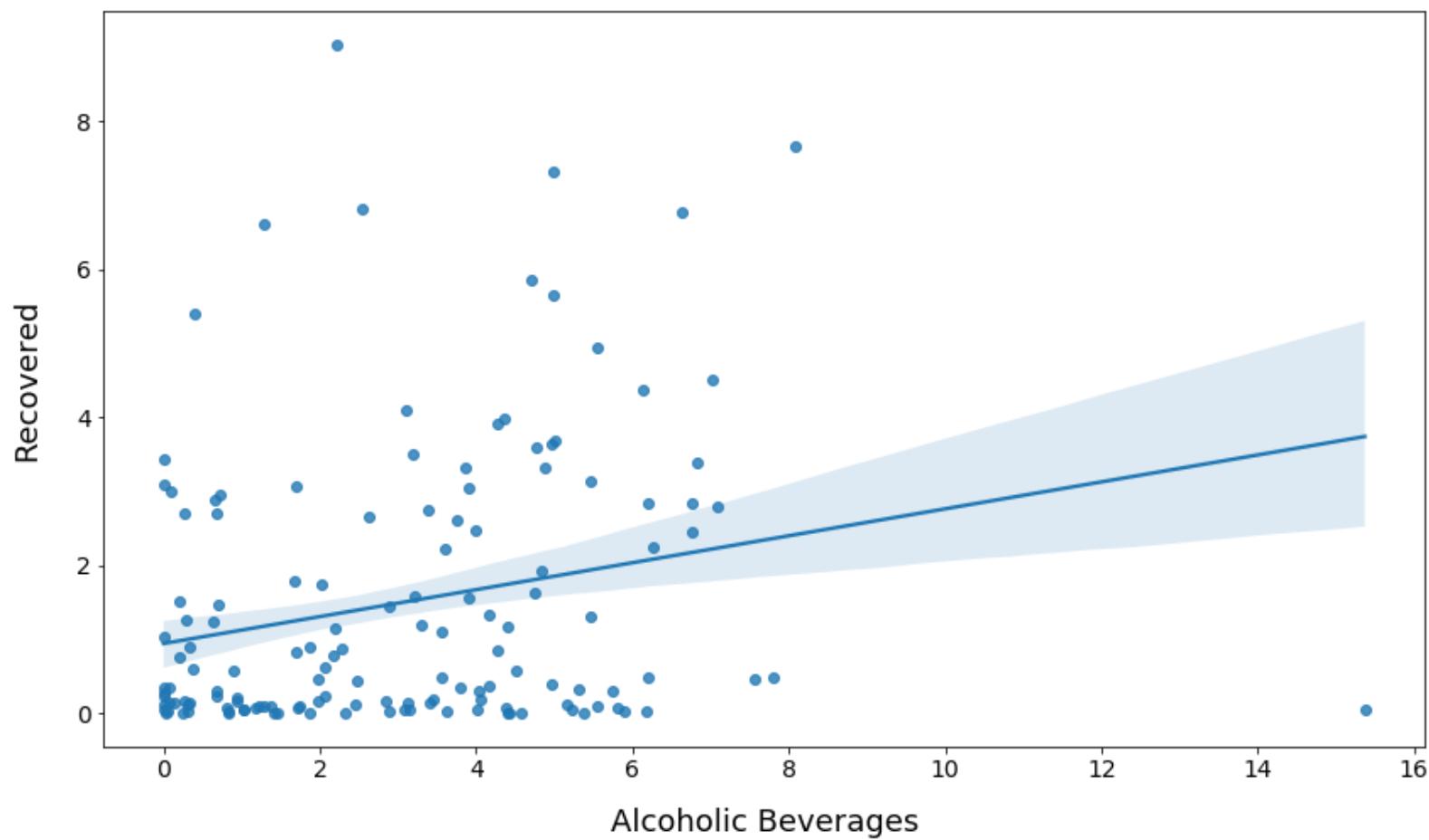
The continent that consumes the least amount of alcoholic beverages is Asia. The reasons for consuming so little alcohol in the Asian continent may be their eating and drinking habits and the dense Muslim population in this continent.



The graph below is a regression graph showing the relationship between consumption of alcohol beverages and the recovery rate of coronavirus patients. In this graph, increased consumption of alcoholic beverages leads to increased recovery rates. Although the slope of the regression line in the graph is not high, this result emerges.

However, this graph certainly does not give us the conclusion that alcoholic beverages are beneficial to health. As previously stated, there are many factors that affect the recovery and death rates of coronavirus patients.

In this regression graph, it is seen that the consumption of alcoholic beverages, even if small, has a positive contribution to recovery rates.

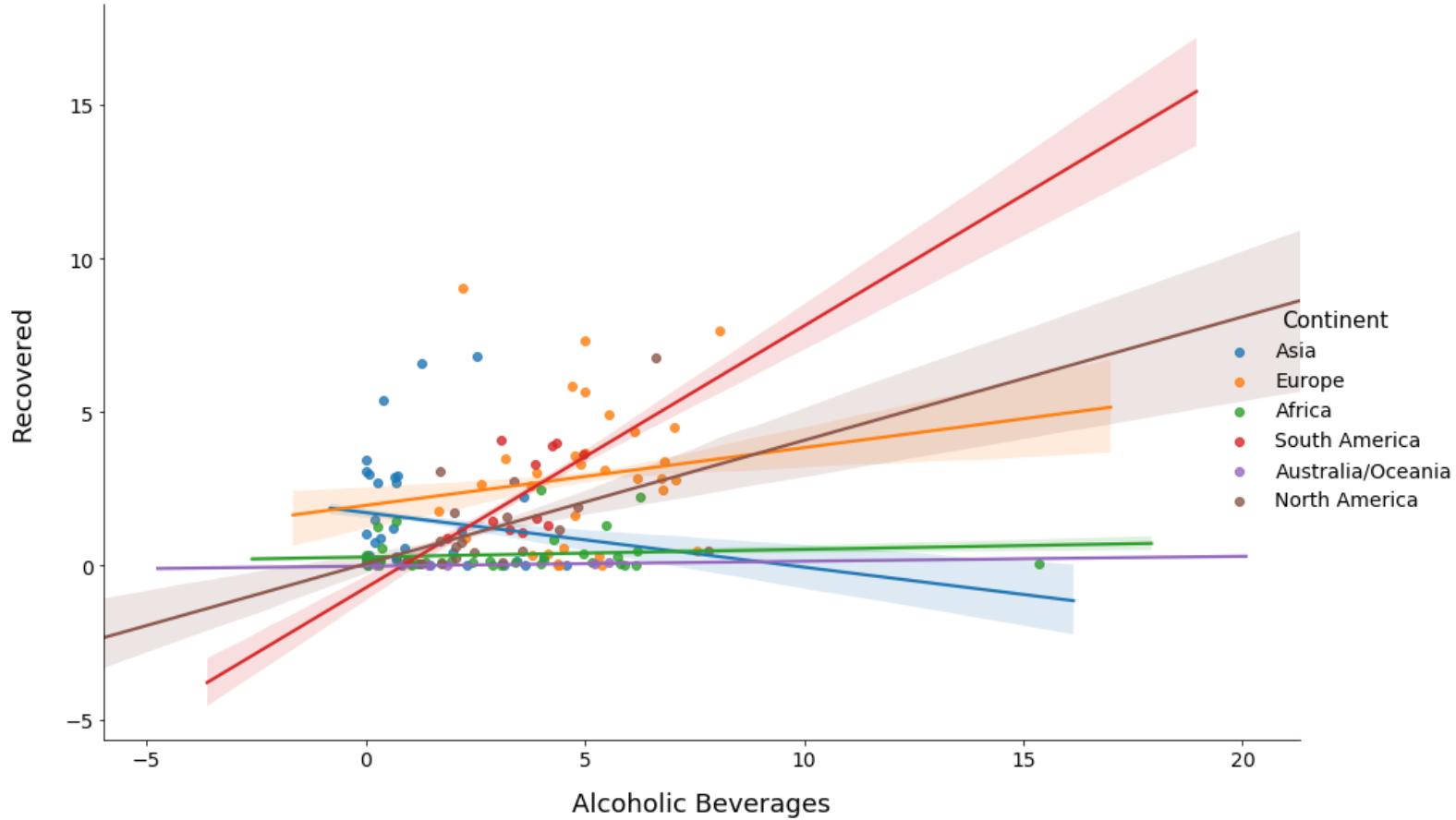


The Relationship Between Recovered & Amounts (kg) Taken from Alcoholic Beverages

When the relationship between recovery rates and alcoholic beverage consumption is analyzed on the basis of continents, the slope of the regression line, especially for South America, is very high. In this case, the increase in the consumption of alcoholic beverages in this continent greatly contributes to the recovery rates.

Likewise, the slope of the brown-colored regression line belonging to North America is not as high as that of South America, but it seems to have a higher slope. Consumption of alcoholic beverages also positively affects recovery rates in this continent.

The only continent where the consumption of alcoholic beverages does not affect the recovery rates positively is the Asian continent because the slope of the regression line belonging to Asia in the graph indicates that the relationship between the consumption of alcoholic beverages and the recovery rates is negative. In other words, consumption of alcoholic beverages decreases recovery rates in Asia.



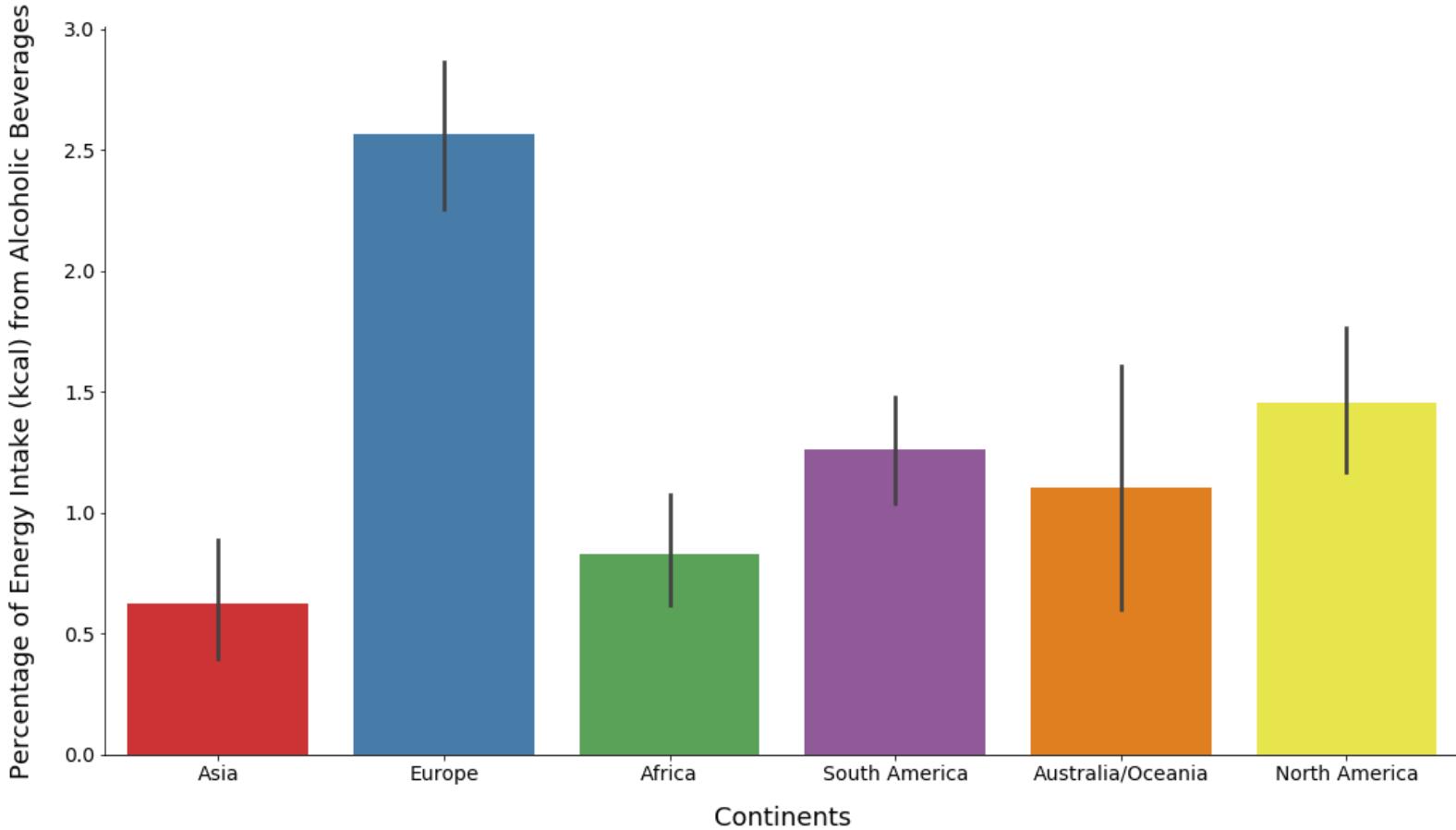
The Relationship Between Recovered & Amounts (kg) Taken from Alcoholic Beverages by Continents

Finally, the continents where the consumption of alcoholic beverages does not contribute to recovery rates are Africa and Australia/Oceania. This result arises because the slope of the regression curves for these continents is almost zero.

3a) Percentage of Energy Intake (kcal) from Alcoholic Beverages

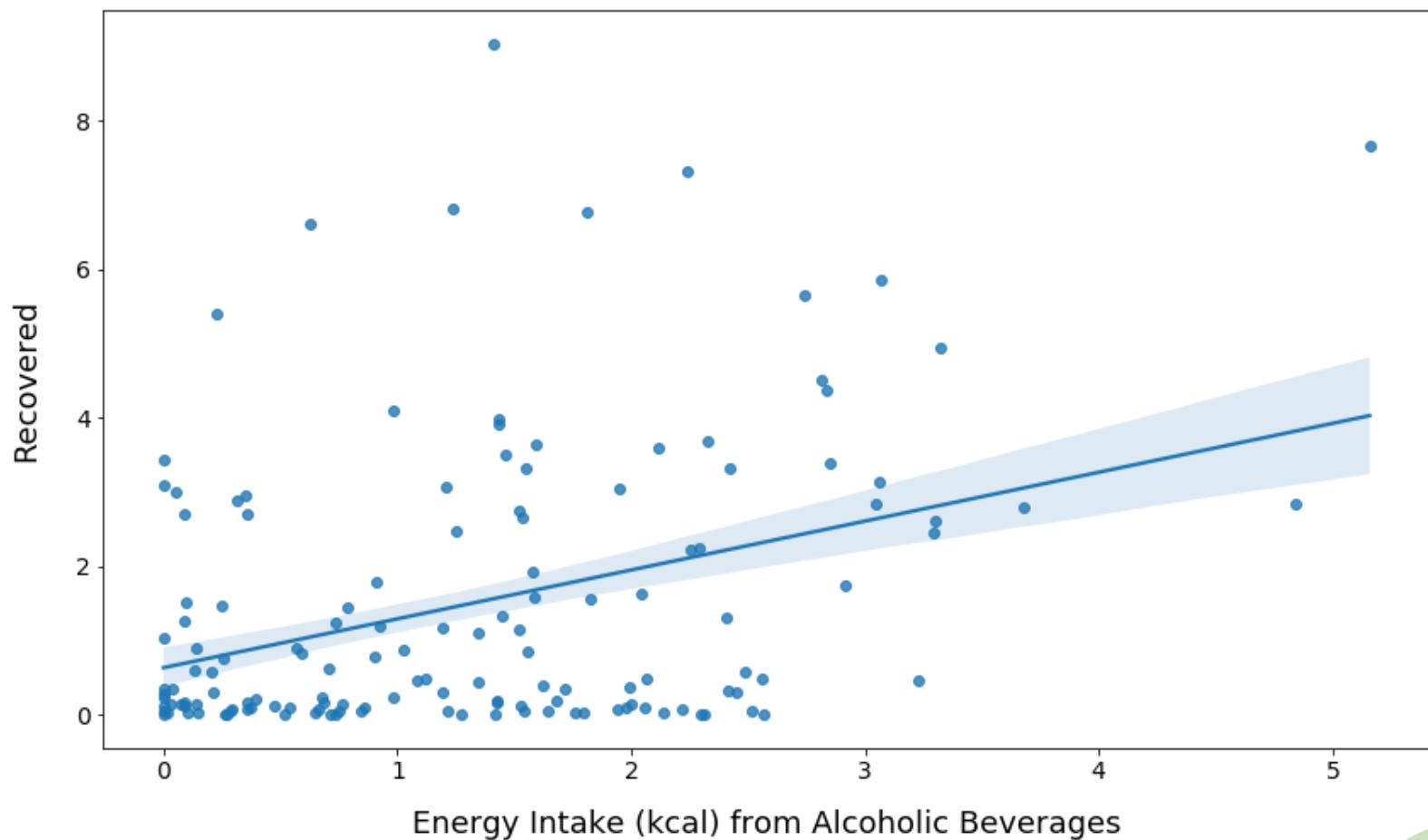
Energy rates from alcoholic beverages produce similar results as in alcoholic beverage consumption. The European continent, which is the leader in alcoholic beverage consumption, is also the continent that gets the most energy from alcoholic beverages. The Asian continent, which is in the last place in the consumption of alcoholic beverages, is also in the last place in the energy intake from alcoholic beverages.

The only difference in this chart is that Australia/Ocean, the continent that consumes slightly more alcoholic beverages than North America, lags behind North America in energy intake from alcoholic beverages.



The regression graph showing the relationship between the two variables below shows that the increase in energy intake from alcoholic beverages has a positive effect on recovery rates.

When the slope of the regression line is examined, it is clear that there is a directly proportional relationship between the two variables, even if there is not a very strong relationship.

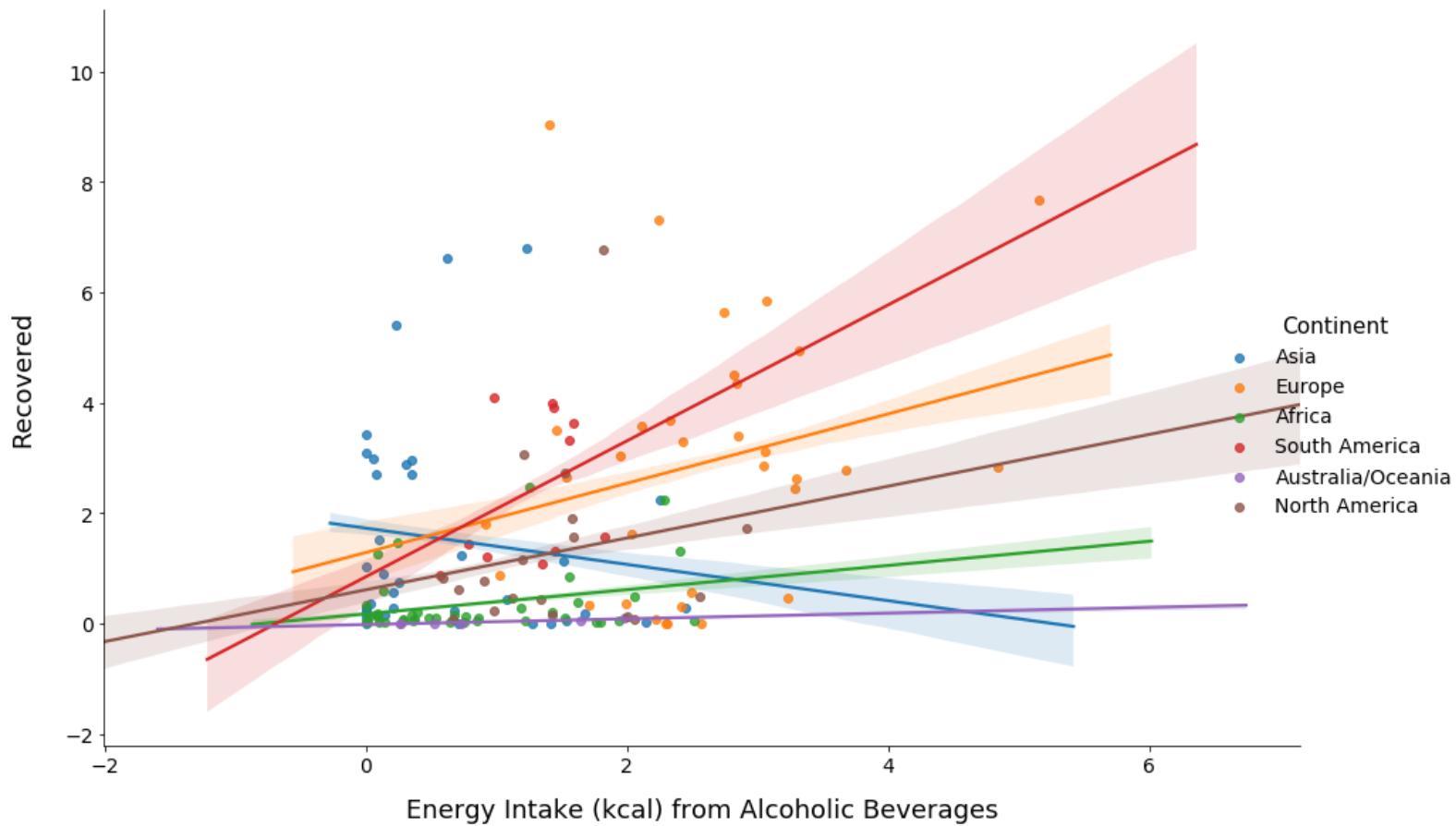


The Relationship Between Recovered & Energy Intake (kcal) from Alcoholic Beverages

In the regression graph of the continents, when the amount of energy taken from alcoholic beverages increases, recovery rates also increase in South America, North America and Europe. The increase in Africa is less than in other continents because the slope of the regression graph belonging to the African continent is closer to the zero point.

As in the consumption of alcoholic beverages, the amount of energy taken from alcoholic beverages in this graph has a negative effect on the recovery rates in the Asian continent.

In the Australia/Oceania continent, where alcoholic beverage consumption almost does not affect the recovery rate, the amount of energy taken from alcoholic beverages has almost no effect on the recovery rates.



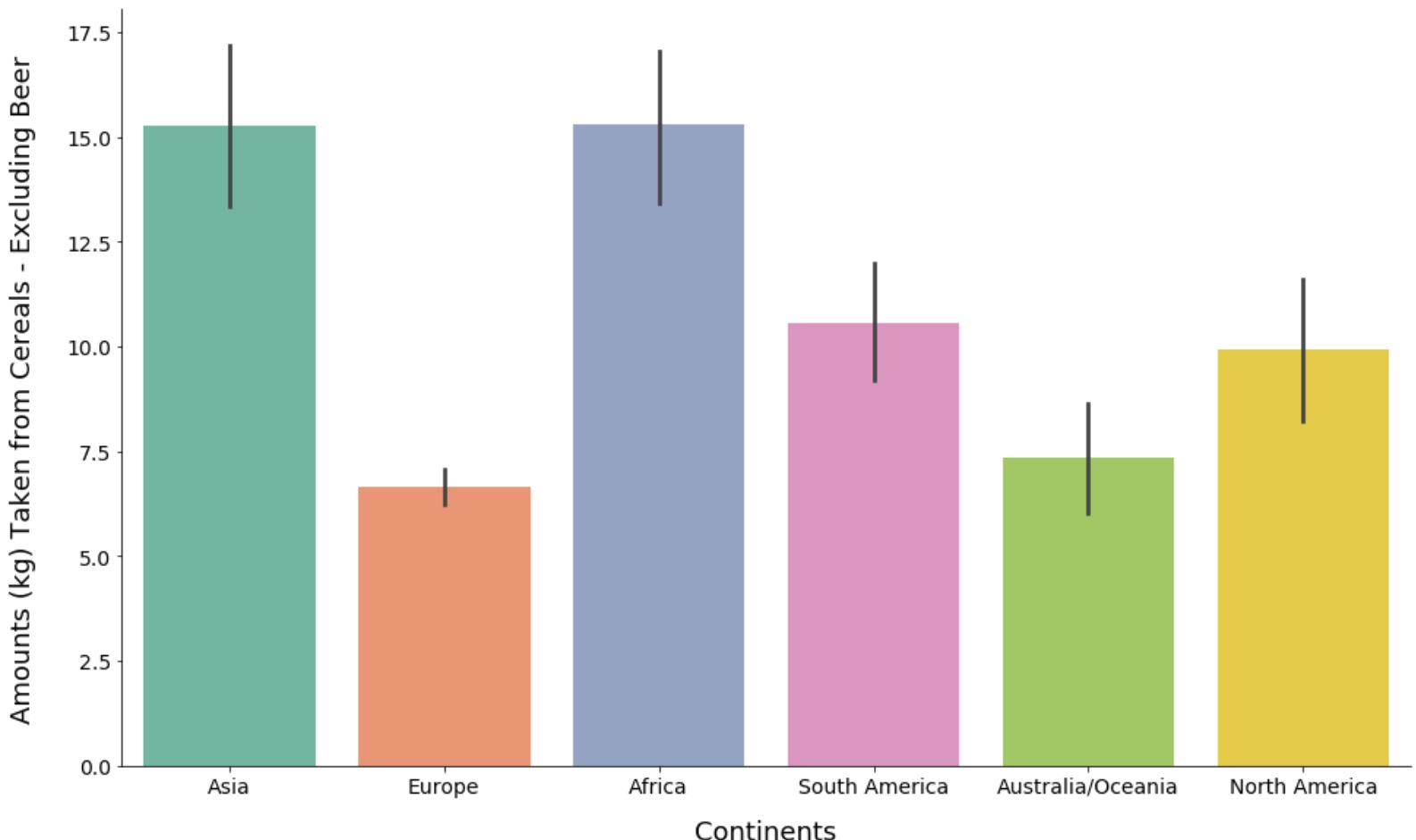
The Relationship Between Recovered & Energy Intake (kcal) from Alcoholic Beverages by Continents

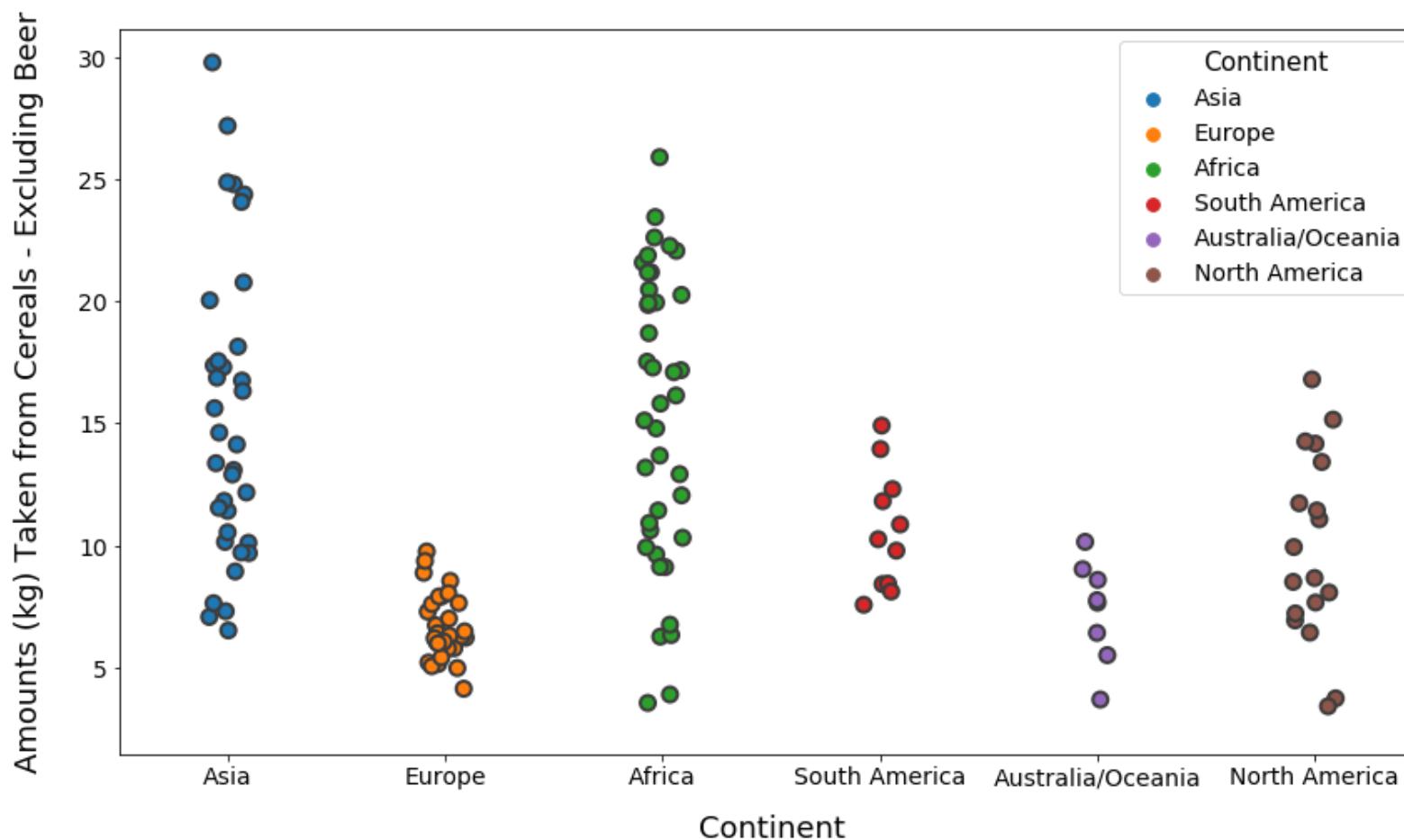
4) Cereals - Excluding Beer

The continents in the first place in cereals consumption are Asia and Africa. Asian and African continents have almost the same consumption rates. The continent with the lowest cereals consumption is Europe.

The main reasons for the emergence of these results are the financial situation of the countries in Asia and Africa and their cereals-based agriculture and difficult conditions in agriculture force countries to agricultural production.

In continents such as Europe and North America, where developed countries are located, cereals consumption is less than in other continents, since protein-heavy foods are consumed and the purchasing power of the society is better than in other countries.

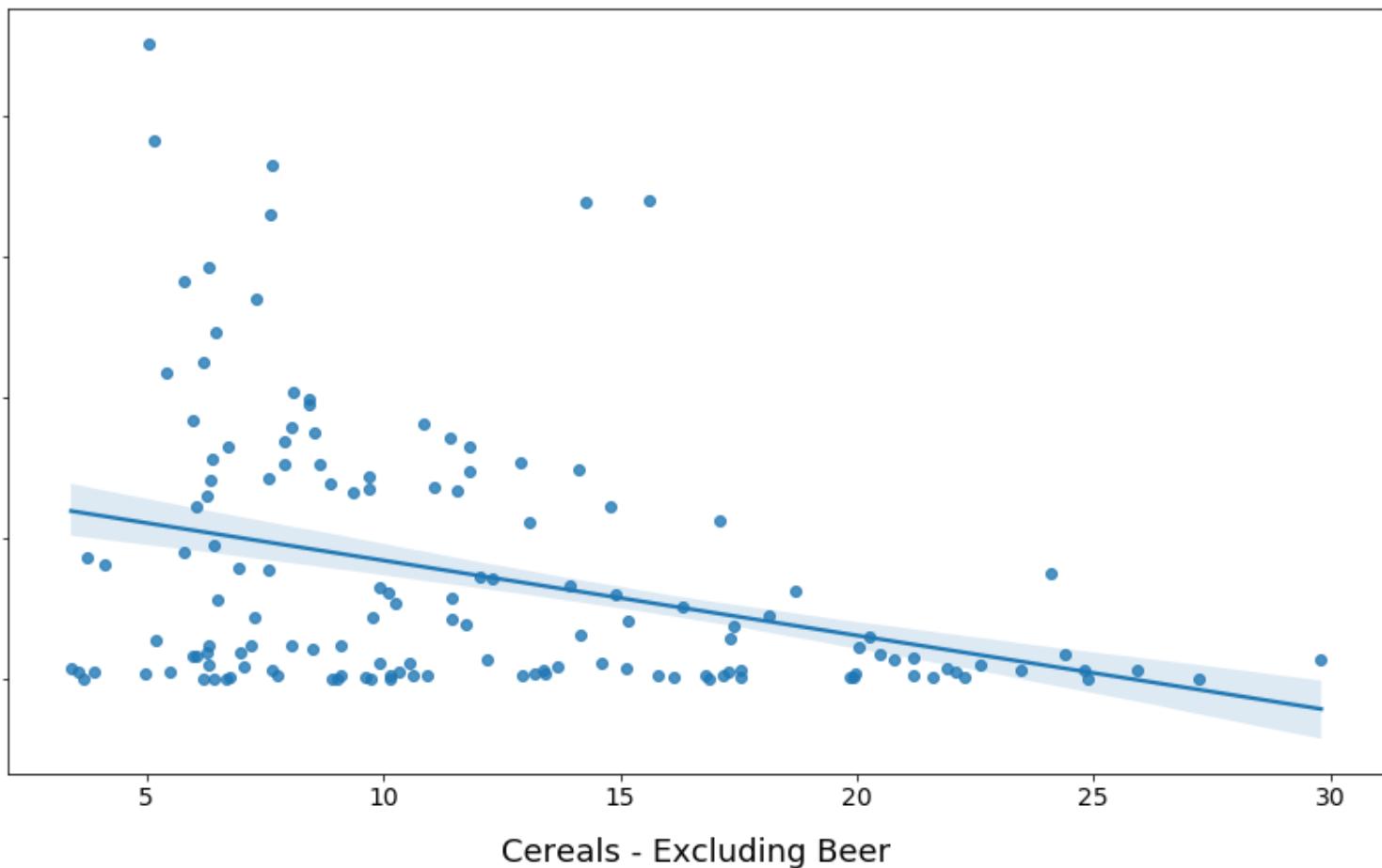




Distribution Chart of Amounts (kg) Taken from Cereals by Continent (Strip Chart)

The chart above shows the cereals rates consumed by the countries in the continents. For example, countries in Asia generally consumed different proportions of cereals. Issues such as agricultural policies, population of the country, fertile agricultural lands and irrigation facilities can cause cereals consumption rates to be so different. The same situation in Asia is also found in the African continent. In the African continent, the cereals consumption rates give very different rates according to the countries.

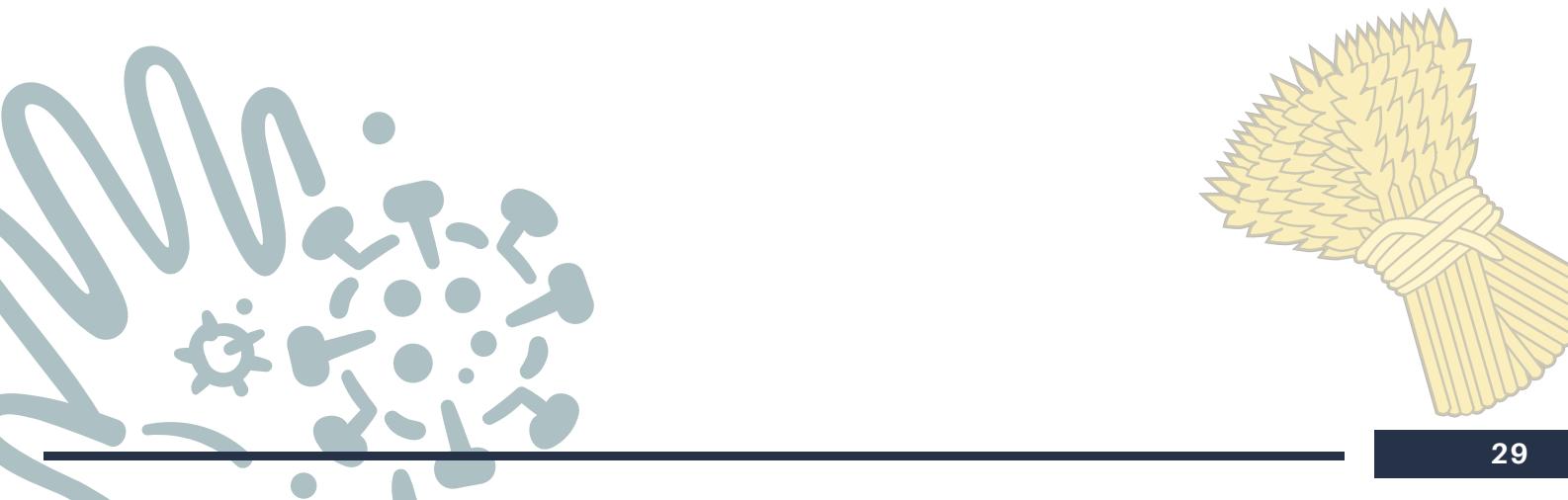
The continent where consumption rates are very close is the European continent. Countries in the European continent consume cereals at almost similar rates. Issues such as well-adjusted supply-demand balance, well-system agricultural policies and the fact that the population of Europe is not as much as in Asia and Africa may have caused these rates to occur.



The Relationship Between Recovered & Amounts (kg) Taken from Cereals

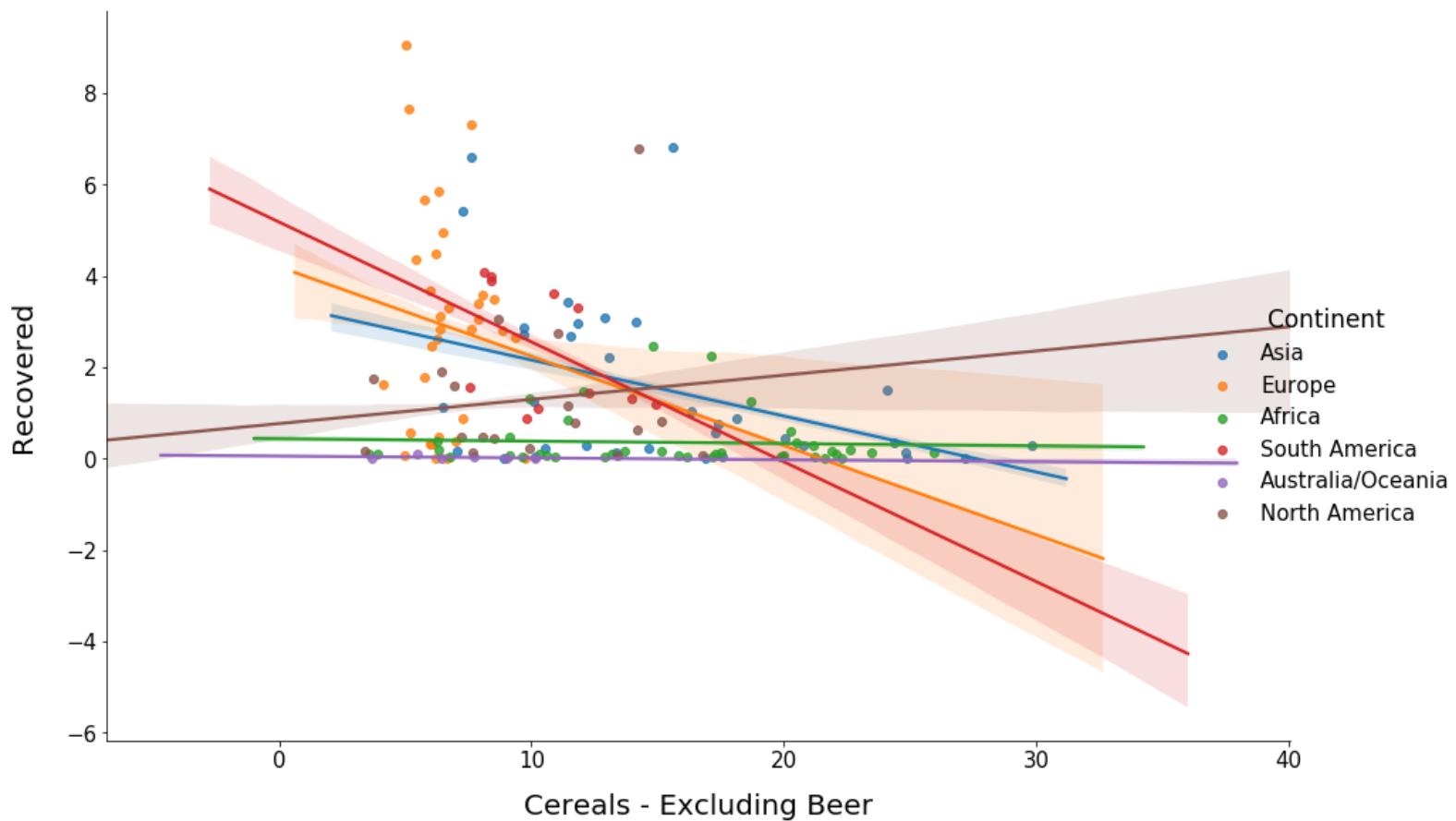
When cereals consumption in the continents increases, the recovery rates of patients with coronavirus are negatively affected. In short, recovery rates are declining.

It is seen that the slope of the regression line is negative. At the same time, it is understood how strong the relationship between the two variables is.



The increase in cereals consumption in South America reduces the recovery rates to lower levels. The slope of the regression line for South America shows how strong the relationship between these two variables is. The situation in South America also exists in continents such as Asia and Europe.

There is no continent that has a positive effect on the recovery rates of the increase in cereals consumption. Cereals consumption has almost no effect on recovery rates in Africa and Australia/Oceania continents.

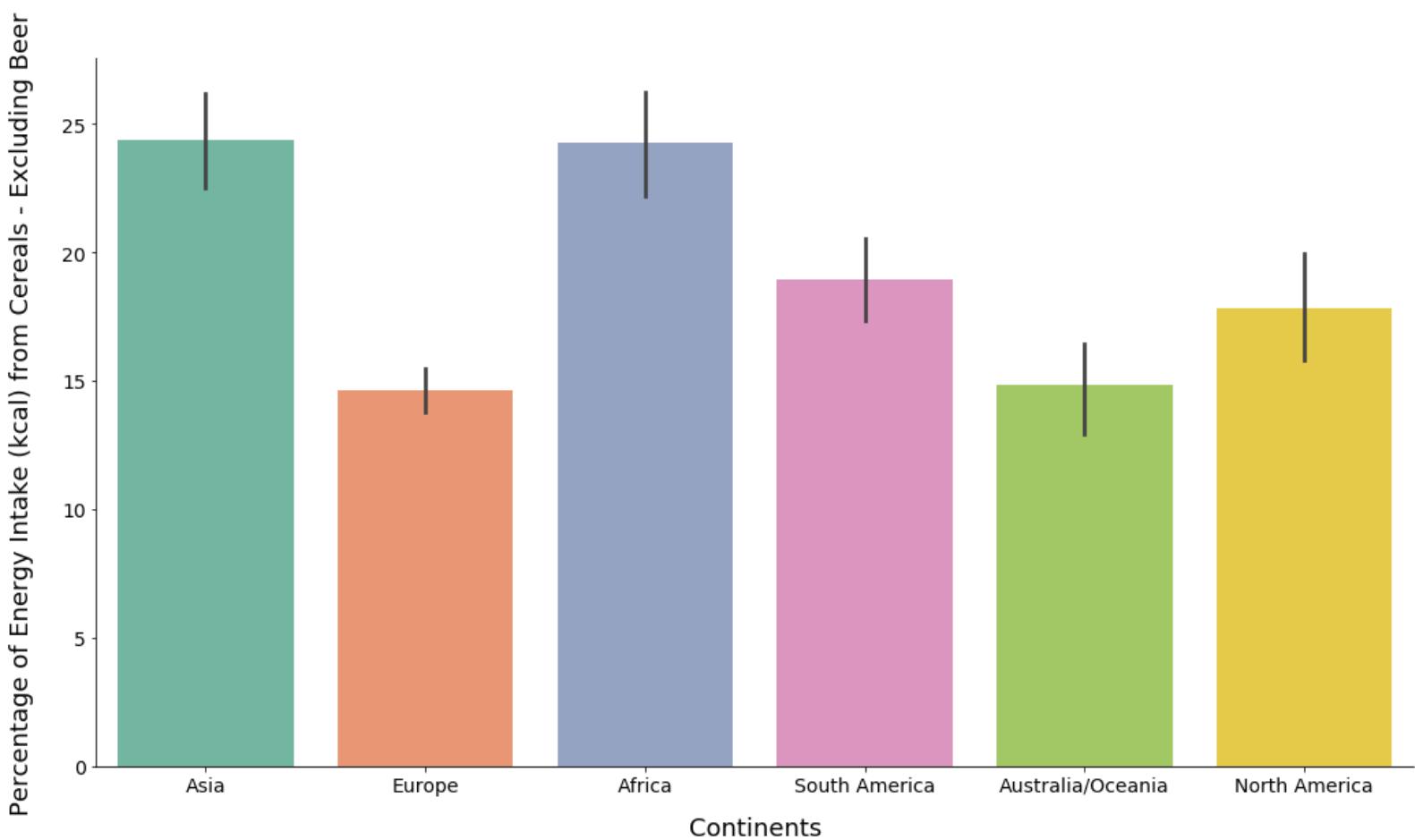


The Relationship Between Recovered & Amounts (kg) Taken from Cereals by Continents

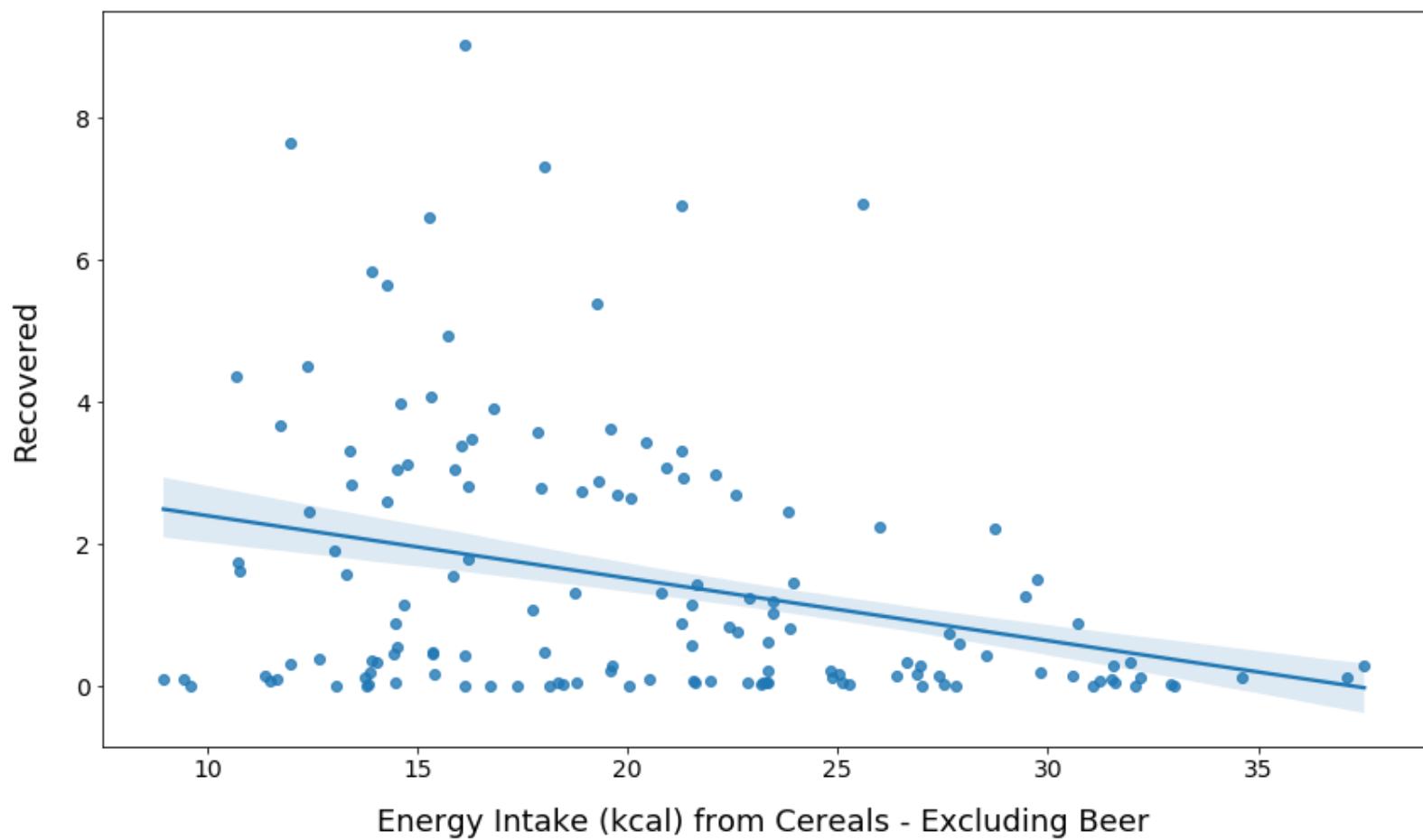
4a) Percentage of Energy Intake (kcal) from Cereals - Excluding Beer

Asian and African continents have the highest energy intake from cereals. After Asia and Africa continents, South America is the continent with the highest energy intake from cereals.

Finally, as seen in the chart, Europe, the continent that consumes the least cereals, is the continent with the least energy intake from cereals.



As the amount of energy taken from cereals increases, the recovery rate of coronavirus patients decreases. The slope of the line in the regression graph indicates that the relationship between the two variables is negative. In addition, the relationship between these two variables cannot be said to be very strong.



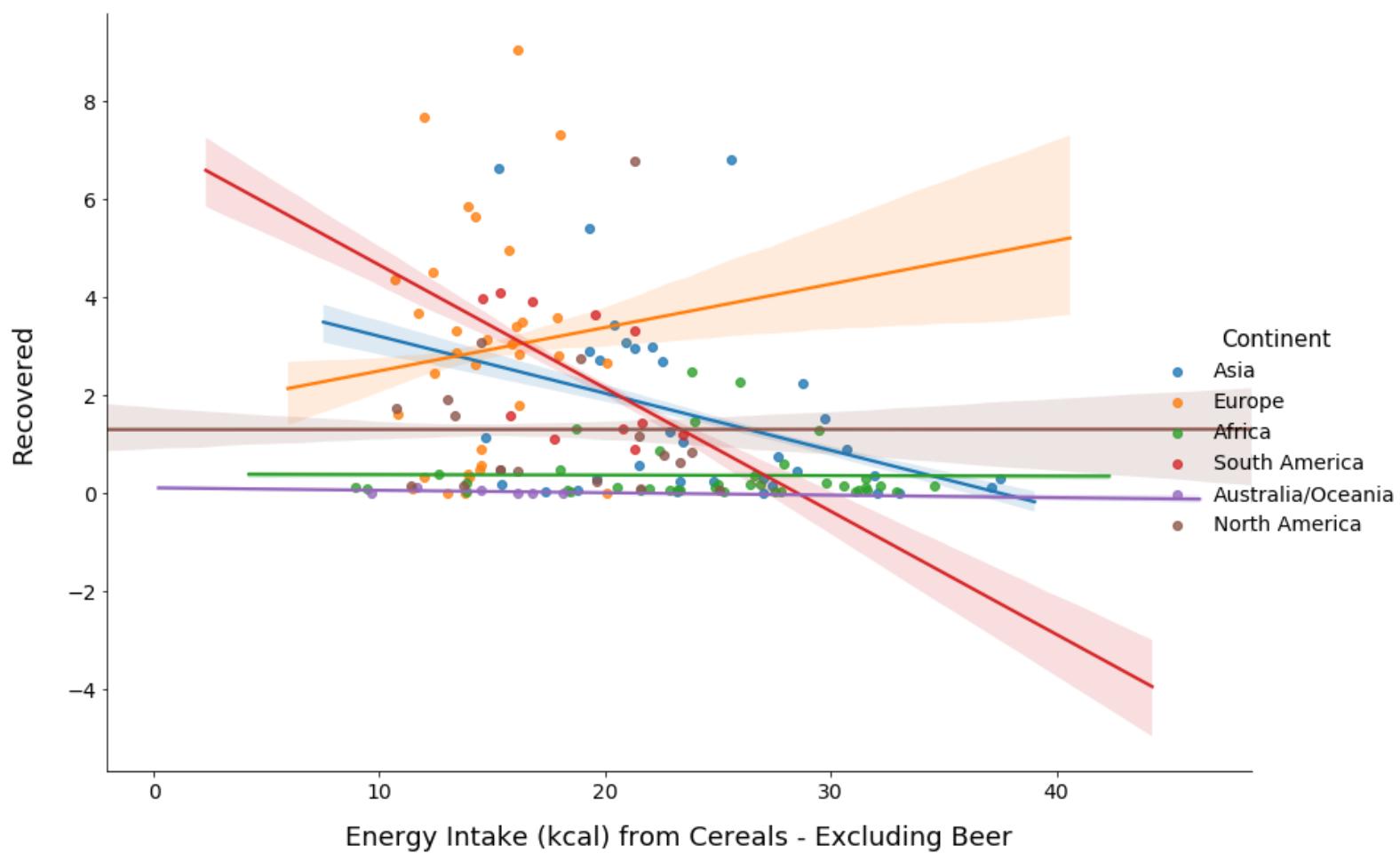
The Relationship Between Recovered & Energy Intake (kcal) from Cereals



Due to the increase in the amount of energy taken from cereals, the continent that experienced the sharpest decline in recovery rates is South America. The slope of the regression line for South America gives this result. After South America, another continent where the amount of energy taken from cereals reflects negatively on recovery is Asia with a blue regression line.

The only continent that has a positive effect on the recovery rates of the energy taken from cereals is the European continent. As the energy intake from cereals increased in the European continent, the recovery rates also increased.

In continents such as North America, Australia/Oceania and Africa, the amount of energy taken from cereals seems to have almost no effect on recovery.



The Relationship Between Recovered & Energy Intake (kcal) from Cereals by Continents

Conclusion of Data Visualization Project

One of the most important factors affecting human health is nutrition habits, and with this data visualization project, it was observed how some food groups and dietary habits contribute to the recovery process of coronavirus patients. At the same time, while analyzing within the project, evaluations were made according to the continents.

As a result of the analyzes made with data visualization tools, animal products, vegetables, alcoholic beverages contributed positively to the recovery rates. These food and beverage groups contributed differently to the healing process. Only the cereals products group could not contribute to the recovery. At the same time, consuming too much of a product group or getting too much protein and energy from that product group does not have a large positive relationship with recovery rates.

Another result is that although a continent consumes a product group more, the amount of protein or energy it receives is less than the other continent. The reason for this may be the development of agriculture, fertile lands and the population of the continent. As a result, it doesn't matter how much something is consumed. What matters is how much valuable nutritional value is taken from the food consumed.

In conclusion, many food groups in the data set and the nutritional values taken from these food groups have contributed positively to the recovery process, but there are many factors that affect coronavirus deaths and recovery. The data visualized with this data set does not give direct results, but it provides insight into the subject.