Variation of Colombian gut microbiome according to the host diet and health

Reyes, Alejandro*. Guerra, John**. Garcia, Angela*. Moreno, Carlos*

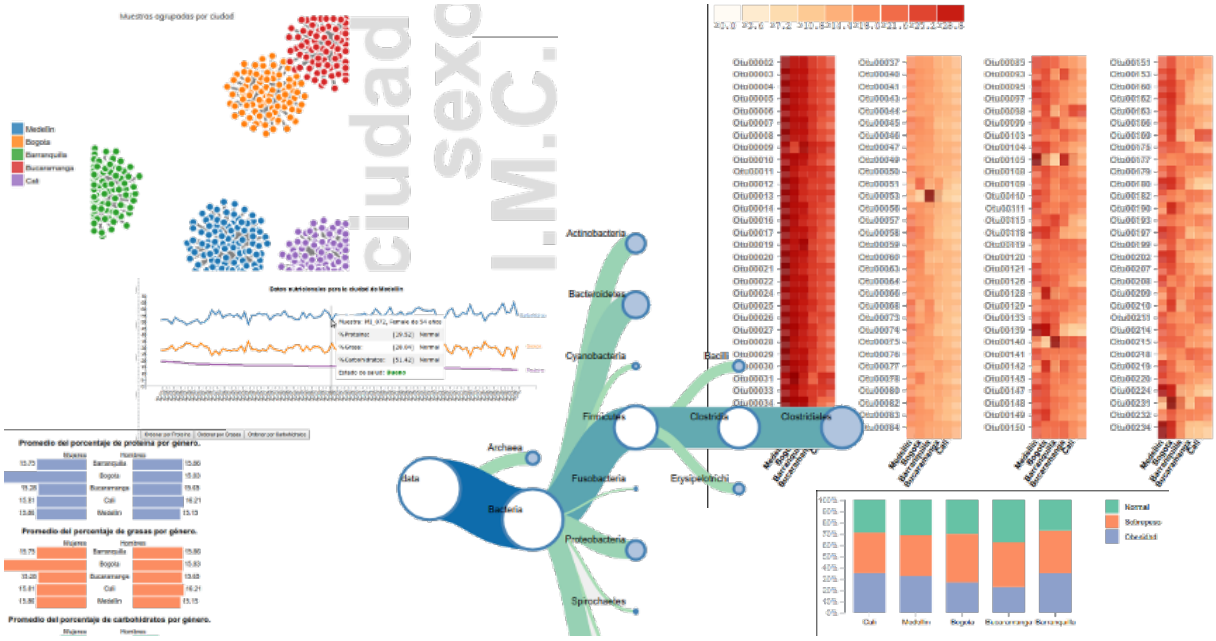

Fig. 1. Mosaic of all the visualizations of the project.

Abstract—The relationship between gut microbiome, diet and health in the colombian population is jet to be determined. From multivariate data available for 441 individuals we designed a visualization that mainly seeks to allow the user first to explore the data alongside the three variables of interest, second to find patterns between the evaluated variables while further use describes details and differences between groups. Of interest there is similarity between all groups of variables. However the most similar data belongs to the cities of Bogo and Medelln. Diet in general is of good quality and does not directly relate to health, but differences in gut microbiota suggest that small changes may be of interest for furhter analysis. This was obtained thanks to the designed tool.

Index Terms— Health, Diet, Nutrition

1 Introduction

Research in medicine and health has described how gut microbiota relates to the health status of their corresponding host. It has also been observed that some diets alter the gut microbiota composition and abundance. Also, diet is of interest in public health to influence health and disease states. However, a pattern between the gut microbiota, diet and health are unknown in the Colombian population. To find the pattern there is multidimensional data available between groups of healthy and unhealthy people. The authors designed the visualization for two main tasks: first to explore the available variables, especially those

- **Research Group in Computational Biology and Microbial Ecology, Universidad de los Andes, Colombia*
- **Max Planck Tandem Group in Computational Biology, Universidad de los Andes, Colombia*
- ***Research group Visual Computing IMAGINE. Universidad de los Andes, Colombia.*
- *Contact: a.reyes@uniandes.edu.co, ja.guerrag@uniandes.edu.com, as.garciav@uniandes.edu.co, c.moreno@uniandes.edu.co*

variables of interest in public health with those from diet and gut microbiota; second, we seek for patterns among variables of health and diet that may alter gut microbiota abundances.

2 THE DATA

A sample of 441 subjects of five different cities in the Colombian population with their corresponding data for diet, health and gut microbiome information were provided from a health institution. The provided data is in three different formats each for every attribute of interest. All data was transformed and treated separately preferably to categorical data with their corresponding description to health and disease states.

Health and diet data was categorized according to Colombian health policy standards, including the technical document for the Recommendations for Energy and Nutrient Intake for the Colombian population (RIEN in Spanish), approved under the Resolution 3803 of the Ministry of Social Protection in 2016. Gut microbiome data was summarized to their taxonomic information using the bioiformatic tool Qiime, designed for ecological studies, given that gut microbiome data was found in terms of absolute abundances for different OTUs (Operational Taxonomic Units), artifitial representations of different species of microorganisms present in faecal samples of each individual. Further data preprocessing was done to transform the data to the corresponding tsv, json and csv formats required for the visualization. All data can be found in `https://github.com/Cuntaquinte/va_proyecto/tree/master/data`.

All data visualizations were designed in D3.js version 5.

3 TARGET AUDIENCES

For the general user, the visualization describes sociodemographic variables of the Colombian population sample. Also, the visualization describes the Colombian diet, showing how good or wrong is the Colombian diet. Additionally, it illustrates the main taxonomic groups of bacteria found in The gut microbiome of individuals. For users with previous knowledge in the areas of biological sciences and health, the visualization describes trends in macronutrient consumption, the nutritional stifferences in gut microbiome taxons through different variables of interest in public health. This differences reflect some previous knowledge known in studies of populations different to Colombian.

4 THE VISUALIZATION

4.1 Multiple graphs

The first main task seeks to explore the information described by the different variables or attributes available in a sample of the Colombian population. In terms of Tamara Musnzner [1] we seek to explore features. Other tasks that can be solved through the visualization include summarizing data by city, sex and nutritional state as described by body mass index.

From the multiple diagrams designed to describe the data we observed that more than three quarters of the population reported a normal diet, being Cali the city with highest percentage of people with a good diet. However, two thirds of the population in the sample are overweight or obese, being Bucaramanga the healthiest city. Overweight and obesity are considered dependent of a low quality data consumption, a data which is inconsistent with the reported diet quality, suggesting that people know how to eat well but the population is still dealing with health problems as consequence of not eating as recommended.

In terms of the available nutritional information by city, the line-chart shows apparently no pattern by macronutrient content. After ordering for each of the macronutrients is possible to see that carbohydrate and lipids have an inverse proportion, where a higher consumption of carbohydrates is related to lower consumption of lipids. On the other hand, there seems to be a direct relationship between lipid and protein consumption.

Looking for more information from the gut microbiome, in the heatmap we observe clusters of OTUs with differential presence, on one hand we observe that the first section holds 31 OTUs that are present trough out all cities in at least 30 samples, the second block shows OTUs with the lowest presence alongside all cities, while the other three sections are mixed and show that the closes gut microbiome similarity is between the cities of Bogot and Medelln.

4.2 Dendrogram

The second main task was to identify patterns between gut microbiome abundances, nutritional status and some sociodemographic variables of interest in public health, specifically the city and sex. The designed based on the availability of taxonomic classification of each OTU wich is a hierarchical classification of organisms used by biologist.

The dendrogram shows that across all variables the abundances of taxonomic groups are similar and the most abundant kingdom of microorganisms present are bacteria, from them the most abundant phylum are Firmicutes, the most abundant Class are Clostridia, the order Clostridiales and the most abundant family within is Ruminococcaceae. This group is the most abundant across all evaluated attributes.

Searching for differences between factors of the body mass index classification, we observed that Obese population has a highest proportion of Firmicutes than Bacteroidetes phylum than overweight individuals. Also, Between normal and overweight individuals there is an inverse proportion of Verrucomicrobia and Proteobacteria being Verrucomicrobia of highest abundance in the data. While comparing all three factors most microbiota groups change very little, the most marked difference shown was that Obese individuals had lower abundances of Bacterodales (family Bacteroidaceae). From the above microbial taxonomic groups Bacterodaceae and Verrucomicrobia have been related to healthy status while Proteobacteria groups pathogenic bacteria, an interesting data, as Proteobacteria are mostly found in overweight individuals.

By analyzing the dendrogram across all five cities, we observed again the closest similarity between the gut microbiome composition between Bogot and Medelln, while the apparently healthiest Bucaramanga has marked differences in the Bacteroidales and Verrucomicrobia group, both reported in healthy status. Finally, the analysis by sex showed that men have more Bacteroidetes than women.

5 USABILITY TEST RESULTS

Report corresponding to the usability test carried out with real users for the project end of Visual Analytics.Each of the main tasks was evaluated by at least two users independently

- **Navigability:**the site is user friendly and the information is easily found.
- **Convenience:** that the final interaction with the user on the website is done in time and appropriate form.
- **Functionality:**all theVizto run correctly.For example: that there is not broken links, erroneous information messages, or error messages in the browser.
- **Actuality:** that the published information be updated and report the dates of completion.

Based on the previous principles for the design of our visualization, it was granted to each ”evaluation” by selected users to test the application, a score in each variable (navigability, convenience, functionality) according to the analysis made during the time that each test lasted, like this:

Scale 1-5

Table 1. Usability first score

Excellent	Very good	Good	Regular	Bad	N/A
5	4	3	2	1	0

Main task No 1: Explore the information described by the different aspects collected (variables or attributes) available in a sample of the population Colombian

Which for purposes of effectiveness of the test it is presented totheuseras: **What is the degree of correlation between the health, nutrition and intestinal microbiota of the sample interviewed in 5 cities of Colombia.**

Users interviewed 4. You were asked to navigate in a running application (part of what will be the development complete), and the rest of the visualization in a mockup that could be scratched and evaluated. The verification questions at the end of the test were:

a What do you think about the navigability?

b What do you think of the information displayed?

c What do you think about the execution time?

d What do you think about aesthetics?

e What can you suggest to improve visualization?

The first part of the visualization will be the first interaction with the users, then through interaction with the viz you can expand the tools to achieve the proposed insights.

Suggestions:

- More differentiable colors.
- Improve execution time.
- Expand the correlation information, additional to the displayed data.
- Improve usage information.
- Change the node icons or the viz interaction in the cluster, as it guides the user to navigate as if you could discover more information by clicking on the nodes.
- Add titles.
- Arming the correlations on request.
- Use pop-ups to expand the information when interacting.
- Evaluate if a heatmap is the best option to show the desired correlation or it would require a different viz or additional ones.

Task 2: Identify patterns among the topics of intestinal microbiota, diet and health a sample of the Colombian population.

Usability check questions dendrograma:

a Which kingdom has more subdivisions or associated taxa?

b What is the species of Archaea present in the Colombian population?

c How many families of Bacteria has the Order of the Burkholderiales ?. Your taxonomy is: Bacteria / Betaproteobacteria / Burkholderiales

d What do you like about the visualization?

e What do you think can be improved in the visualization?

The dendrogram is based on the selection of previous data resulting from clustering in the network, with this it is possible to show the taxonomy of the UTOs in an orderly manner, where it is possible identify which are the different groups present in the selected subgroup (taxon), with it is expected to identify patterns at different taxonomic levels. Table 4.

Aspects to improve:

1. It is not possible to detect a direct relationship between the microbiota and the state of health, due to the failure in the information filters .
2. The names of the different taxa are not clear due to the superposition of the link with the text of the nodes .
3. After searching for specific taxa, it was evident that visualization was cut when the tree reached a depth greater than 4, which is insufficient since there are 6 levels of information .
4. Highlight the groups of bacteria with the greatest presence in the data
5. Describe by color the groups of bacteria that occur in the different quantitative data categories, eg. To observe differences between the presence of certain taxa between men and women.

Positive aspects:

6. The taxonomy of the OTUs present in the data is clearly and clustered.
7. It is possible to identify specific groups especially for those users who have previous knowledge in the bacterial taxonomy.

Global evaluation of the visualization: Value: moderate. Relevance: Important but requires work.

Table 2. Usability scale score

Insight	Navegability	Convenience	Functionality	Average	importance
Identify correlation between nutrition, health and microbiota	4.5	3.3	3.2	3.6	1
Organization of data	4.6	3.7	3.6	3.9	2
Discover patterns	4.0	4.1	3.8	3.9	1
Esthetic	4.7	4.1	4.0	4.2	3

Table 3. Usability second score

Insight	Navegability	Convenience	Functionality	Present	Average	Position
Identify dependencies between different topics	5	3.3	3.7	10	1	
Classification of data in terms of healthy or not healthy	4.0	3.4	3.3	3.4	7	2
Finding data extremes	5	5	5	3	18	3

6 Conclusion

The present visualization allows the exploration of data from different attributes, but with further use of the tool we found clusters, tendencies and details of changes between gut microbiome, diet and health. Some findings reflect the changes between groups that by other methodologies are hidden. In general after the usability tests the tool shows promises to describe the data.

Acknowledgments

The authors wish to thank to our client Alejandro Reyes, for the support and advice.

References

[1] T. Munzner. *Visual Analysis and Design*. CRC Press. Taylor and Francis Group, first edition, 2015.