Information Visualization Final Report

| Catarina Mendonça  75381 | Rúben Vines  76991 |
| --- | --- |

# ABSTRACT

UPDATED—23 December 2016. This sample paper describes the formatting requirements for SIGCHI conference proceedings, and offers recommendations on writing for the worldwide SIGCHI readership. Please review this document even if you have submitted to SIGCHI conferences before, as some format details have changed relative to previous years. Abstracts should be about 150 words and are required.

## Author Keywords

Authors’ choice; of terms; separated; by semicolons; commas, within terms only; this section is required.

## ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous; See<http://acm.org/about/class/1998> for the full list of ACM classifiers. This section is required.

# INTRODUCTION

Our problem’s domain is food and beverage consumption worldwide, which we decided to perceive above all ideas we had, because of the high amount of data found and the various ways of applying such information.

The motivation that lead us to start working on this theme was the fact that we would be able to study each countries’ behavior towards what they consume – their preferences. Not only that, but also, how we could either prove or demystify what we (and maybe wrongly) think is common knowledge (e.g. American’s are the fattest people on Earth).

This types of questions and myths quickly started surging between us, as a group, and we understood the great potential that our theme had in terms of applying it in our *InfoVis* course. The questions we wanted to get an answer from our visualization were:

* What is the most consumed food in the world?
  + Most popular aliment;
  + Highest alcohol consuming regions.
* Do cultural beliefs influence a whole countries’ consumption values?
  + Less alcoholic consumption in Middle East;
  + More cereal (rice) consumption in Asia;
  + Lower amount of bovines eaten in India.

Questions that were related with diet evolution throughout the ages were also relevant and raised, however, due to our dataset and other constrictions they were dismissed, and shall be discussed in other sections of this report.

Actually there was a big factor that wasn’t taken into account, in the initial stages, and wasn’t a huge deal for us in order to perceive this theme. That factor was the lack of tools already existent in the web. However, in retrospective, the most we can find to visualize this type of data is boring static tables and extensive non-interactive text.

After all this introspective and discussion between the group, we decided what tasks we wanted our tool to perform, would be to:

* Explore drink consumption worldwide;
* Explore the FIX ME

We dismissed some of our firstly thought tasks that included comparing information from past eras with the

present, or evolution throughout the time, for the same reason stated before.

# RELATED WORK

Our idea was one hundred percent decided before taking into account external visualizations or influences. However, such visualizations were not useless to us. They might have not served us as a source of inspiration, but they were used to show the possibilities of what we could achieve.

As stated earlier, the amount of interactive tools isn’t abundant covering this topic but there are still a few infographics we can find online with some ease. One of them, can be found in the hyperlink above:

# http://foodtechconnect.com/2012/12/28/top-5-food-health-infographics-of-2012/

# The one and only visualization we found that can we can consider a competitor to our project is a set of interactive pie charts produced and published by National Geographic, that let’s us have access to the diet evolution in the course of the 50 years between the era of 1961 to 2011. Even though, this webpage has some key interactability features, such as selecting pieces of the pie chart, or highlighting some other sections of another, it is highly restrictive due to the low amount of countries they provide for us to interact with (http://www.nationalgeographic.com/what-the-world-eats/).

# THE DATA

The first dataset we came across seemed like very complete and perfect for the task we proposed. It was a static spreadsheet, with over 70 attributes and 80 countries. The attributes of this first dataset were mainly types of consumed food (types of meat, grain, fruit, etc.), nutrients (such as carbohydrates, fats, proteins, etc.) and demographics (obesity, life expectancy, etc.).

After acknowledging the existing of such demographics, we added to our future tasks, comparisons between food consumption habits and life expectancy or obesity of a given region.

Of the over 70 possible attributes, we decided to keep 37 of them either because of the lack of information in a few of the attributes, the non-importance of a certain attribute to our project or simply because we didn’t want or knew how to represent a certain attribute in the visualizations we had in mind.

After analyzing, deciding over and discarding the information we wanted, there was a general agreement that this dataset wasn’t enough to completely fulfill the tasks we had previously thought about. The reason behind that was because there wasn’t nearly enough information about beverages or alcoholic consumptions.

One other factor that influenced the tasks we had previously proposed was the fact that there wasn’t information regarding several years which would deny us the possibility of to take conclusions out of a course of time.

Our second dataset was obviously about beverage consumption and had all countries, information regarding the years 2008 to 2015, and the following types of alcoholic beverages: beer, wine, spirits and others (as the name states it’s a category of beverages that are not included in the others).

This dataset had a bunch of incomplete rows that we considered as outliers and scrapped from the countries’ list. We ended up with 67 countries that fully had information between both datasets, and could represent for the most part every region of the globe. In addition to that, we selected only the year of 2014 to match the information we had from the first dataset.

We didn’t derive any measures because our first dataset had such detail in its attributes that we could simply take the values directly from it.

After we presented the third Checkpoint, where we suggested what our visualizations could be, we had as feedback that our project could easily be more interactive, as in, showing changes in our visualizations after pressing, selecting or hovering other data (for example).

Taking that into consideration, we decided to add the option of giving the possibility of accessing our data either from per capita or absolute values, with a simple button press. And for that, we included two new custom made datasets that had information the same information as the two previous datasets, but with different values. These two datasets, were later normalized to a scale of 1 to 10000kcal so values would appear in the same scale as per capita ones.

A problem that later appeared after having problems with the implementation of our idioms was that every value in our .*json* (JavaScript Object Notation) files was converted into a *string* , causing troubles when we wanted to make numerical operations.

# THE VISUALIZATION

## Overall Description

## Rationale

## Demonstrate the Potential