Supersize World: Visualization of Global Obesity Trends

Yo-whan Kim*
yowhan

Hyeyoung Shin†
hyshin

1 Introduction

Obesity used to be a problem only in the high-income countries, but it has recently become an emerging problem in the low-income countries as well. In this project, we first highlight an ongoing problem of obesity around the world by presenting the growing trends of obesity rate using multiple interactive and animated charts. Obesity should not be ignored as something trivial because obesity is becoming one of the leading risk factors for early death and is responsible for 4.7 million deaths in 2017 alone [3].

In the second part of the project, we look at the correlation between obesity and daily caloric intake. Excess energy sources is one of few reasons for an increase in obesity, which can happen either from excess energy intake or lack of energy usage, and calories is the standard unit we use to refer to the amount of energy we get from food or drink sources. We visualize the increase in obesity, along with the trend of growing caloric supplies, suggesting that they have a significant correlation. We hope that this project reminds the users of the critical problem of obesity and also encourages the users to protect themselves by engaging in physical activities or being aware of and managing nutrient intake.

2 Related Work

There have been abundant studies on the danger of obesity, as well as on the distribution of obesity by income classes or nations. In 2009, [9] presented data and case studies highlighting the steady rise of obesity among children from 1998 to 2008 in the low-income classes of the United States (they argue that childhood obesity naturally leads to a higher possibility of obesity in adulthood). The paper states that obesity in the low-income classes had been continuously rising from 1998 to 2003, but hit a plateau until 2008. Yet, in 2019, [10] held a world-wide survey of household data from 103 different countries, and concluded that the obesity problem has been transitioning from the wealthy to the middle- and low-income classes even after 2008. Based on this paper, we aim to visualize obesity as a global class-insensitive problem, rather than as a problem just for high-income classes or countries as it used to be.

Many efforts were made to investigate the variables that cause and correlate with obesity. For instance, [4] studies the relationship between sugar intake and obesity in the United States. Their analysis revealed that reduced sugar consumption resulted in slowing down the increase of obesity. As expected, although the sugar consumption trend has been declining, the US obesity rate is still on the rise, suggesting sugar consumption isn't the only reason responsible for the high obesity rate in the US. There are still active researches studying such relationships; for instance, [2] reveals a positive correlation between alcohol intake and obesity.

[6, 7] take a more generalized approach, studying the correlation between obesity and overall daily caloric intake of individuals. They both conclude that daily caloric intake directly has an impact on weight loss and gain. [7], which is a more recent study published in 2018, developed a phone application to accurately track the food consumption and caloric intake of participants, as well as their daily BMI information, for accurate data collection. In our project, we seek to visualize this correlation between obesity and caloric intake, but at a worldwide country-level.

3 Methods

We first present the datasets used for this project, as well as the necessary cleaning, wrangling and augmenting process. Python and Pandas library were mainly used for the data preprocessing.

3.1 Adult Obesity Rate Dataset

The World Health Organization (WHO, [8]) provides obesity rates among adults for each sex by country from 1975 to 2016. For consistency, if a country is missing at least one obesity value for either sex, we drop that country from the datset. The dataset provides mean obesity rates as well as the standard deviation information in string format, so string parsing methods were used to wrangle the data such that we only keep the floating point mean obesity rates.

Because we wanted to group the countries by their continents, we augment the dataset by using Python's CountryInfo library, [1]. The same library is also used to augment the dataset with country population.

3.2 Count of Deaths by Risk Factor Dataset

The Global Health Data Exchange (GHDx) [5] provides a dataset of all causes of death per country per year. Since we want to look at the overall death per cause per year over the world, we aggregated the country data and computed the overall sum of a cause of a death per year. Death caused by cholesterol had missing data for some years, so the missing values were linearly interpolated using neighboring values. We removed age and sex information from the dataset as they are not used in our target visualization.

3.3 Share of Deaths Dataset

GHDx [5] also provides the share of deaths due to obesity by country dataset. Since the dataset was organized in a hierarchical manner (by variables we did not consider for the final project), we reorganize the data similar to how we did in section 3.1.

Although the original data contains continent information for each country, the continent information was only present once for each country (not at every row). The dataset was wrangled such that missing continent values were copied over from the filled rows. Finally, the dataset also contained countries with missing continent information, and these were filled in by the CountryInfo library, [1].

3.4 Daily Caloric Supply dataset

The Food and Agriculture Organization of the United Nations [11] provides the daily caloric supply dataset by country. Since we wanted to study the correlation between obesity and caloric supply, the dataset from section 3.1 was merged with this new dataset. If a country was not present in one of these two datasets, that country was dropped in our merged version. Although the daily caloric supply dataset had values dating back to 1961, the older years were not considered as we were not able to fetch the obesity data for those years. As continent information was already augmented in section 3.1, it was not necessary to further augment the merged dataset.

*e-mail: yowhan@mit.edu
†e-mail: hyshin@mit.edu

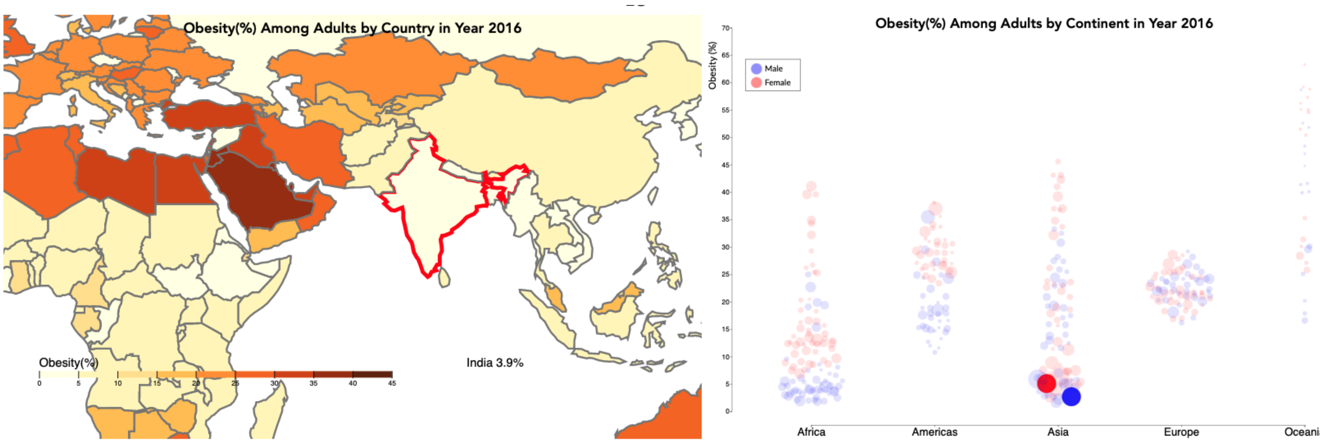

Figure 1: World Obesity Map and Beeswarm Chart with *India* as the user selected country.

4 Results

4.1 World Map and Beeswarm Chart of Adult Obesity

Using the adult obesity rate dataset from section 3.1, we create a side-by-side world map and animated beeswarm chart (using d3 force) to visualize the increasing trend of obesity rates around the world. We implement an automated play feature that iterates through the years to update the two charts accordingly. The user can also choose to use a slider to manually select a year to view. We also give the users the freedom to choose which continents to show on the beeswarm chart in case the user only wants to focus on a few specific continents.

We chose to use a choropleth map to help users easily catch that the obesity data is given for each country (not for major cities or regions). Color encoding for the map was chosen such that the color scale has a high contrast for low and high values for dramatic change effects. We also wanted to make sure that the color scheme goes along with the major theme color (yellow-orange scale) of the overall project. We implemented zoom by mouse wheel and pan by click-and-drag for intuitive exploration of the world map. Finally, users can hover over countries to see the numerical obesity value along with the country name, and further click a country of interest on the map to pin the obesity rate on the bottom of the map. The border of the selected country is highlighted, and the respective country is accented on the beeswarm map as well.

The main goal of the beeswarm chart was to help visualize the shift in distributions of obesity rates per continent. Countries were grouped by their continents in the x-axis, and the y-axis was used to show the obesity rates. Users can easily notice the circle data points generally shifting upwards as the years pass. Color encodings were used to represent the obesity rates of each sex, and radius of the circles were decided based on the population of the respective country. Again, we show the quantitative obesity rate upon hovering. Finally, we allow the user to set the 'spatial density' variable using a slider. Lower spatial density value groups the countries by continents more tightly and vice versa, and as a result, users can study the continent-level trends with lower spatial density or observe the worldwide-level trends more easily with a higher spatial density value.

Figure 1 shows an example usage of the two charts. For instance, if we are interested in the change of obesity in India, we can select the country on the map by clicking, and the respective circles in the beeswarm charts are highlighted. Using the automated play button or the slider, we can then study the change in trends of obesity in India.

4.2 Number of Deaths by Obesity and Top Risk Factors

In this section, we wanted to highlight how obesity has become a serious problem over the years by stressing that it is causing a lot of death and is actually becoming one of the leading factors of death.

We first plotted a line chart, shown in Figure 2, that represents the total number of deaths per cause by year. The x-axis represents years from 1990 to 2017 and the y-axis represents the number of deaths. Each line in the graph represents a death-cause in the dataset from section 3.2. We first used different colors for each death cause, but too many colored lines were present and labels were hard to read. Most importantly the obesity trend line was hard to point out in this type of representation, and it was challenging to notice its increasing trend and rank as well. We tried implementing a color encoding in which the obesity line gets an unique color, but we were not able to efficiently solve the problem of complex lines and text layout.

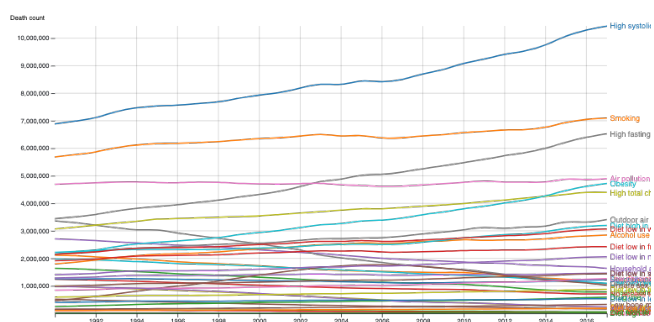

Figure 2: Initial approach, but the chart was not used due to its complexity and confusing color encoding.

We then switched our visualization to a bar race graph. This graph also represents the total number of deaths per cause by year. The x-axis is the number of deaths and the y-axis is the causes of death in hierarchical order. To make the graph less clustered, we only show the top 12 ranks per year instead of all 35 causes. We added a play button to automatically iterate over the years from 1990 to 2017. To make the animation smoother between years for better readability, we interpolated our data to generate the frames for year transitions and rank change animations. Lastly, to stress obesity over other causes, we color-coded obesity in orange and all other causes in grey. Again, we also wanted to make sure the color encoding matched the overall theme color.

By presenting our data as a bar race graph, we hoped that the

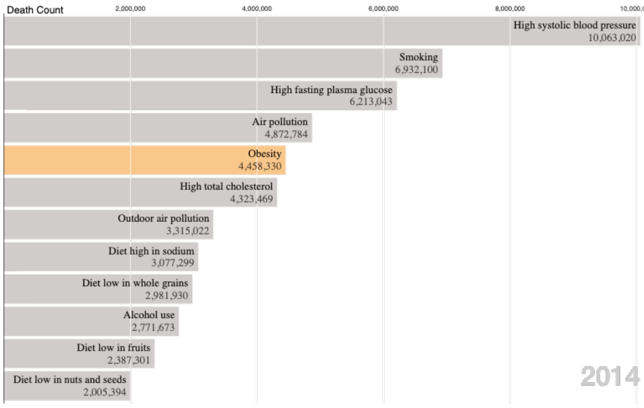

Figure 3: Snapshot of the bar race graph. Highlighted bar represents the death count due to obesity.

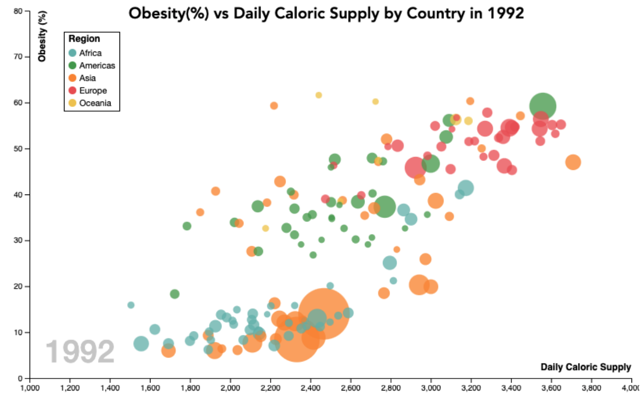

Figure 4: Gapminder-inspired chart presenting general increase in both obesity rate and daily caloric supply.

rank of obesity steadily climbing up over the years was shown more effectively compared to our initial line graph.

4.3 Share of Death by Obesity Map

In the previous section, we provided a visualization that shows the growing count of deaths by obesity. Since some may believe that the increasing number of deaths by obesity might be due to the general increase in world population rather than because of the obesity problem worsening, we wanted to further provide a visualization showing the share of death by obesity in percentages rather than in absolute counts. We use a choropleth map to show the share of death by obesity, and the same encodings from section 4.1 were used. Dataset described in section 3.3 is used to create this visualization.

4.4 Investigating the Correlation Between Obesity and Caloric Supply

In the second part of the project, we present visualizations that show the correlation between obesity and daily caloric supply, and the dataset from section 3.4 is used throughout this section.

First, we use a gapminder-inspired chart to show the obesity and daily caloric supply change for years from 1975 to 2013, as shown in Figure 4. We plot the daily caloric supply on the x-axis, and the obesity rate on the y-axis. Years can be iterated through automatically using the play button, or manually by using the slider. Color encoding reveals the continent of the country that the respective circle datapoint represents, and circle radius represents the population. Similar to the other visualizations, we allow the users to filter the continents to be shown on the chart using a checkbox, and if the user is interested in a specific country, we provide a textbox to search and pin a country to track. The pinned country gets highlighted on the chart accordingly. Note that the users can also find out the country name of each circle datapoint upon hovering. Finally, transition animations for each circle are added such that even dramatic position changes of the circles look smooth.

We hope to show the users that both obesity rates and daily caloric supply had generally increased from 1975 to 2013, and that these two variables have positive correlation with each other. One can easily notice that the overall distribution of data points has shifted up and right, especially when the manual slider is used to compare the years 1975 and 2013 back to back.

4.5 Side by Side Obesity and Caloric Supply Multi-Line Charts

Finally, we present two interactive line charts to further shed light on the correlation between obesity and caloric supply. This visualization provides a better view of the two variables at a lower level. That is, we can better observe the yearly trends of a specific continent or country of interest. While the first line chart plots the change in obesity rates over the past three decades, the second line chart is used to plot the respective change in calories once a country is selected by the user. Refer to Figure 5 for an example usage of this side by side multi-line graph.

We provide the checkbox and textbox filtering system described in the previous section. Color encoding is used to represent the continent information again, and for consistency, we use the same color scheme as the previous visualization. Upon hovering over a line on the obesity chart, the name of the country representing that line is revealed in the title and as an annotation. Then, the selected country's change in daily caloric supply is plotted on the chart on the right. As mentioned above, the user can also select a country using the textbox, rather than hovering. If that is the case, we disable the hovering functionality until the user deletes the selected country in the textbox.

When one or multiple continents are selected to be shown, users can quickly notice from the obesity chart that the overall trend of obesity rate is increasing. Upon hovering over the lines, we wanted the users to notice that while most of the countries' daily caloric intake trends also have an upward trend, there are a discernible number of countries with no significant change or even decline in caloric intake. (We encourage the users to search for such countries, and we also give examples of both categories in the visualization description.) On the published web-based visualization, we inform the users that while calories is the standard we use, higher numerical caloric value doesn't definitely indicate it will lead to more weight gain, and that nutrient composition is a more crucial determining factor of weight gain than numerical caloric values. By providing example countries that had growth in obesity rate but decline in daily caloric intake, we wanted to further solidify this argument.

5 Discussion

In section 4, we described the visualization products as well as their intentions and what we hope the users to learn from each chart. In this section, we share the results of our small user study. We gathered data from two graduate students in MIT Korean Graduate Students Association (KGSA). As a pre-survey, both participants were asked to rank what they believe to be the top seven causes of premature death out of 35 choices. They were then asked to select all continents that they think had an increase in obesity rate and daily caloric intake over the past two decades. Both participants did not include obesity in their choices of top seven causes of premature death, and one participant did not include Africa in their choices of continents with increased obesity and caloric intake. After both

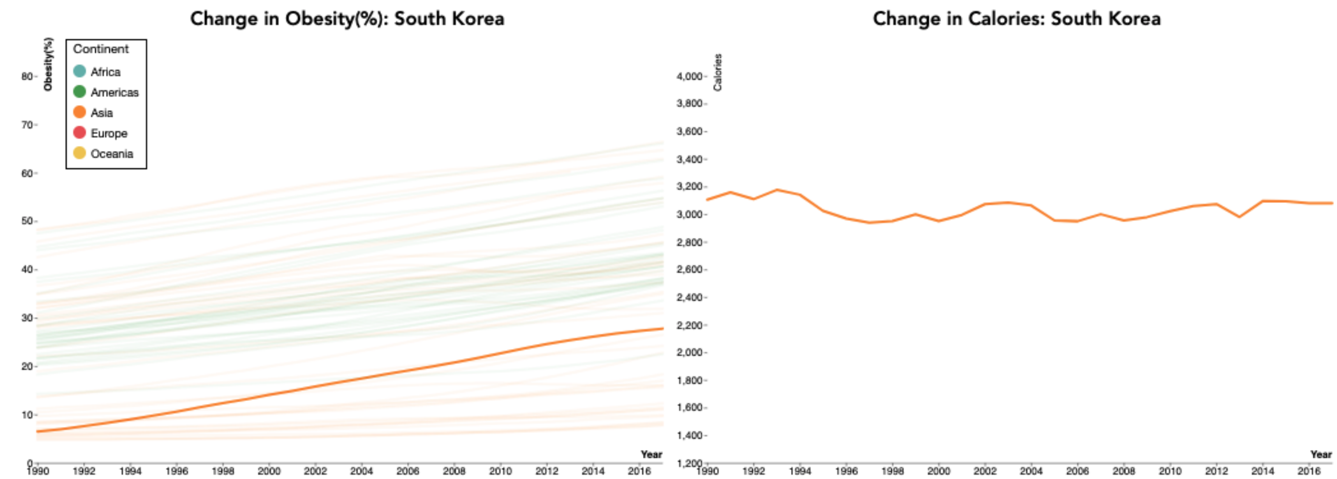

Figure 5: Interactive side by side change in obesity and daily caloric supply line graphs, with *South Korea* chosen as the user selected country via textbox input.

participants interacted with our site, both were surprised that obesity is one of the top five leading death factors.

As feedback for visual encodings, both participants preferred the bar race chart over the line chart mentioned in section 4.2, stating that the bar race chart grasped their attention more quickly. One participant further stated that the first world map was essential for her to fully understand the obesity rate changes across the world. The other participant stated that the chart of "side by side view of obesity and caloric supply" was astonishing since a lot of countries didn't seem to have a significant increase in caloric intake. Both participants agreed that after looking at the visualizations, they now understand the seriousness of the obesity problem and that they should be more aware of their caloric intakes.

6 Future Work

First, we would like to integrate different types of variables that might have correlations with obesity and visualize their correlation. Particularly, we would like to explore the data of sugar consumption rate and alcohol intake, as mentioned in section 2. After looking at these different types of excess energy, we could also visualize which of those variables has the strongest correlation with obesity.

As discussed before, nutrient composition is a more influential factor than the mere caloric value when it comes to weight gain or loss. We believe that changes in the three major nutrients (i.e. protein, fat, carbohydrates) intake rates data should be collected to create informative interactive visualizations as future works.

References

[1] P. Chandro. Country info python open library. 2018.

[2] G. Colditz, E. Giovannucci, E. Rimm, M. Stampfer, B. Rosner, F. Speizer, E. Gordis, and W. Willett. Alcohol intake in relation to diet and obesity in women and men. *The American Journal of Clinical Nutrition*, 1991.

[3] M. Desai, R. Sharma, I. Riaz, S. Sudhanshu, R. Parikh, and V. Bhatia. Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128.9 million children, adolescents and adults. *Yearbook of Paediatric Endocrinology*, 2018. doi: 10.1530/ey.15.13.20

[4] S. Faruque, J. Tong, V. Lacmanovic, C. Agbonghae, D. Minaya, and K. Czaja. The dose makes the poison: Sugar and obesity in the united states – a review. *Pol J Food Nutr Sci.*, 2019.

[5] T. I. for Health Metrics and Evaluation. Global burden of disease collaborative network. *Global Burden of Disease Study*, 2018.

[6] C. Hill, B. Weir, L. Fuentes, A. Garcia-Alvarez, D. Anouti, and L. Cheskin. Relationship between weekly patterns of caloric intake and reported weight loss outcomes: Retrospective cohort study. *JMIR Mhealth Uhealth*, 2018.

[7] M. Nestle. Increasing portion sizes in american diets: More calories, more obesity. 2003.

[8] W. H. Organization. The global health observatory body mass index dataset. 2021.

[9] A. Sharma, L. Grummer-Strawn, K. Daleniu, D. Galuska, M. Anandappa, E. Borland, H. Mackintosh, and R. Smith. Obesity prevalence among low-income, preschool-aged children - united states, 1998-2008. *Epidemiology Program Office, Centers for Disease Control and Prevention (CDC)*, 2009.

[10] T. Templin, T. Hashiguchi, B. Thomson, J. Dieleman, and E. Bendavid. The overweight and obesity transition from the wealthy to the poor in low- and middle-income countries: A survey of household data from 103 countries. *Plos Medicine*, 2019.

[11] the United Nations Food Agriculture Organization. Faostat. 2021.