

Project 2 Description & Report

Arthur Qian, Jade Wang, Xuyuan Liu

A) Data Description

For this project, we used three datasets from Kaggle.com, all from 2022:

1. Instant noodles rating (2022): [Top Ramen Ratings 2022](#)

This dataset provides star ratings for instant noodles from various brands around the world. The main variables include the country, brand, and rating out of five stars. It is named as "noodles_ratings.csv" in our folder.

2. Instant noodle consumption (2022): [World Instant Noodles Consumption 2022](#)

This dataset shows the consumption rate of instant noodles per country, measured in millions of US dollars. It is named "noodles_consumptions.csv" in our folder.

3. World happiness report (2022): [World Happiness Report up to 2022 | Kaggle](#)

This dataset contains the happiness scores of countries globally, along with other indicators that contribute to overall happiness, such as GDP per capita, social support, and life expectancy. We are using the 2022.csv, and we renamed it as "world_happiness.csv" in our folder.

Firstly, to ensure consistency across the datasets, we did several general pre-processings:

- **Standardizing Country Names:** We unified country names across datasets. For example, "United States" was changed to "United States of America," (for matching the naming in our chosen TopoJSON file) and variations like "UK," "United Kingdom," and "Great Britain" were standardized to "United Kingdom." We also ensured consistency for region names like "Taiwan" and "Hong Kong".
- **Selecting Key Variables:** For simplicity, we focused on core variables: happiness scores from the World Happiness Report, noodle consumption data, and the highest-rated noodle brand by country from the ratings dataset. Additional variables in the datasets, such as additional happiness indicators, were not used.
- **Shape Files:** To map the countries on a world map, we used a TopoJSON file from [World Atlas](#) to display country shapes.

For the scatterplot, we used a Python script named “jade_merge.py” to merge “noodles_consumptions.csv” and “world_happiness.csv” into one csv file. The merged version is named “merged_scatterplot_data.csv”. It included country, 2022_consumption, happiness score, and continent columns and removed rows with missing values.

For the bar chart, we used a Python script named “xuyuanSuperclean.py” to merge “noodles_ratings.csv” to filter out the duplicate data inside the csv file.

We found that in the instant noodle rating csv file, there are a lot of country name typos. For example, Ukraine is written as United Kingdomraine. We manually went through each country name listed in the csv file and fixed these typos. Below are some examples:

Find what:	United Kingdomraine	Find what:	Souh Korea
Replace with:	Ukraine	Replace with:	South Korea

B) Visual Design

The general goal of our design is to provide a clear and engaging connection and comparison of noodle consumption, happiness, and top noodle brands by country. Our visual design choices include:

For the main graph—the world map:

- **Map Projection:** We used a Natural Earth projection to provide a balanced global view that keeps country shapes and areas relatively accurate. For a better visualization, we zoom in the default world map to clear out the Antarctica area since there’s no data in there (but zoom out manually it will still be displayed on the map though).
- **Color Mapping for Happiness Scores:** We mapped happiness scores to a gradient color scale, from light yellow (low happiness) to dark purple (high happiness). This color choice ensures contrast and aligns with expectations that darker colors often indicate higher values in a spectrum.
- **Bubble Size for Consumption:** Bubble size represents noodle consumption, with larger bubbles indicating higher consumption. This allows viewers to quickly identify countries with high noodle consumption at a glance.
- **Positioning:** Bubbles are positioned based on each country's geographic centroid. Color and size were mapped directly to happiness scores and noodle consumption, respectively, making it easy to interpret at a glance.

- **Scaling Choices:** Happiness scores were mapped using a quantile color scale to emphasize variations across the happiness spectrum, while noodle consumption data used a square root scale for bubble size to manage large discrepancies in consumption values between countries without overwhelming smaller values.

For the second graph—the scatter plot:

- **Data-to-Visual Mappings:** We picked a scatter plot to show if there is any relationship between happiness score and instant noodle consumption.
 - **Color:** The color of each data point encodes the continent, helping visually separate regional clusters. This use of color facilitates quick identification of geographic trends without overcrowding the chart with labels.
 - **Size and Shape:** Circular shapes are used for data points to provide a consistent mark that balances visual weight and clarity. Each circle corresponds to one country.
- **Scales and Transformations:**
 - **Linear Scale (X-axis):** A linear scale is applied to the happiness score to ensure a proportional representation that is easy to interpret on the horizontal axis.
 - **Logarithmic Scale (Y-axis):** Since instant noodle consumption spans a wide range, a log scale is used to compress the data, making differences at smaller values more distinguishable. This also aligns with the underlying distribution and provides a better visual balance.
 - **Tooltip and Annotations:** The tooltip displays additional data details (country, noodle consumption, happiness score) on hover, providing more information without cluttering the plot. Annotations are also used for axis labels and gridlines to guide the viewer's eye and enhance chart readability.
- **Trend Line:** A regression line captures the trend in the data, and it's mapped based on calculated slope and intercept to provide a visual reference for correlation, with weak correlations noted. This line is essential for users to interpret any general relationship between happiness and consumption.

For the third graph — Bar chart:

- **Color Scale:**

Each bar uses a unique color from a predefined scale to represent differences in categories or values. This makes the visualization intuitive and allows for easy differentiation of data points.

- **Axis Labels and Titles:**

The x-axis and y-axis are clearly labeled with legible fonts, providing essential context for interpreting the data. A descriptive chart title, with x be the rating score and y for the brands that order by the rank of the score, is included to summarize the visualization's purpose.

- **Bar Spacing and Alignment:**

Bars are evenly spaced and proportionally sized to ensure clarity and prevent overlap. The spacing strikes a balance between minimizing clutter and maintaining distinguishable elements. Mainly focused to provide a more explicit presentation on nations that have ample brands.

- **Gridlines:**

Faint horizontal gridlines are included to help users estimate bar heights relative to the y-axis. The gridlines are subtle to avoid distracting from the main chart elements.

- **Consistent Typography:**

A clean, sans-serif font is used consistently for all text elements, including labels, tooltips, and titles. This ensures readability and aligns with the overall modern design theme with cohesive UX design standard.

C) Interactive Elements

For the world map:

- **Tooltips:** Hovering over a country displays detailed information, including the happiness score, noodle consumption, and top noodle brand for that country. This interaction is designed to be intuitive and unobtrusive, providing additional data on demand without cluttering the visual. Note that if a hovered country has no data for the top noodle brand, this category will not be displayed.
- **Pan and Zoom:** We added zoom controls and a pan feature, allowing users to explore specific regions in more detail. This makes the map accessible on smaller screens and enhances the viewing experience for countries with smaller geographic areas.
- **Reset Button:** The reset button allows users to quickly return to the original map view, enhancing usability by providing an easy way to revert changes after exploring.
- **General Style:** The tooltip and zoom buttons have identical CSS styles as the following scatterplot graph.

The choice of these interactions was aimed at making data exploration simple and engaging. We used standard web conventions (clicking buttons, hovering for details) to ensure discoverability without the need for instructions.

For the scatterplot:

- **Hover Tooltip:** Shows extra information on hover (country, happiness score, noodle consumption) without cluttering the view.
- **Continent Filter Buttons:** Filters data by continent with color-coded buttons, helping users focus on regions. The “Show All” button resets the chart, enhancing control and usability.
- **Zoom and Pan:** Allows users to explore data at different scales. Includes a reset button for easy return to the original view, ensuring flexible yet controlled exploration. We also set a limit on the extent of zooming in and out to make sure users won’t accidentally lose sight of the data points.
- **Dynamic Point Scaling:** Adjusts point size with zoom to maintain readability, making detailed exploration easier.
- **Trend Line Update:** The trend line dynamically adjusts with zoom for accurate scaling.

For the Bar chart:

- **Tooltips:** Hovering over a bar displays a tooltip containing detailed information, such as the category name and exact value. The tooltips appear dynamically near the hovered bar, providing on-demand insights without cluttering the chart.
- **Category Filter Buttons:** Interactive buttons allow users to filter the chart by specific categories. These buttons include a "Show All" option for resetting the view, enabling users to focus on subsets of data based on their interests.
- **Sort Controls:** Users can sort the bars dynamically in ascending order based on their ranks. This feature helps highlight trends, such as identifying the highest or lowest values quickly.
- **Interactive Legend:** As we implemented, clicking items in the legend filters or highlights corresponding bars in the chart. This allows users to explore the data interactively based on legend categories.

Between world map and scatterplot:

- If a user clicks on a country bubble in the world map, the corresponding circle in the scatterplot will be highlighted.

Between world map and bar chart:

- If a user clicks on a country bubble in the world map, the corresponding country will be selected in the bar chart dropdown.

D) The Story

Our visualization provides a unique perspective on the global appeal of instant noodles and its surprising correlation with happiness. Through this project, we hoped to provide answers to topics like: "Do countries with higher noodle consumption have higher happiness scores?", "Which regions have the most beloved noodle brands?", and "which regions consume the most instant noodles".

One key finding is the weak negative correlation between happiness scores and noodle consumption, with a correlation coefficient of -0.36 . This suggests that as happiness scores rise, noodle consumption tends to decrease slightly, although the trend is not strong.

From both the world map and the scatterplot, we saw that Asian countries generally show higher noodle consumption, while regions like Europe and Oceania display lower levels of noodle consumption regardless of happiness scores. The variability across regions suggests that cultural and regional factors likely play a significant role in instant noodle consumption, independent of happiness levels. For instance, higher noodle consumption in Asia could reflect cultural preferences, affordability, and accessibility of instant noodles as a staple or convenient food choice. In contrast, lower consumption in regions like Europe and Oceania might indicate different dietary habits or less reliance on instant noodles, possibly due to alternative food preferences or socioeconomic factors. The bar chart provides extra information on countries' top instant noodle brands and their ratings. With both scatterplot and bar chart being connected to the world map, viewers can dive deep into one specific country to learn more about its instant noodle industry and consumption.

Our visualization intends to inspire curiosity about the connections between cultural habits, like noodle consumption, and broader measures of well-being. We hope viewers find it both

informative and engaging, encouraging them to think about the cultural impact of food on daily life and happiness.

Team Contribution:

Arthur:

Choose the project topic about instant ramen noodles and find datasets. Then, working on the world map, pre examine the data in .csv files to make sure they are usable and fit our project requirement. Also, did a general data unifying across three datasets. Final deliverables include: html, CSS, JavaScript coding for the world map (first graph), and related report writing.

Jade:

Found the dataset about world happiness and suggested analyzing the correlation between happiness and instant noodles. Created the python script to merge datasets for the scatterplot. Designed and implemented the scatterplot. Merged the world map with the scatterplot and edited css files to ensure style consistency.

Xuyuan:

Contributed by brainstorming visualization ideas, designing and developing the bar chart to show the noodle brand rankings that differed by nations, and using Python to merge datasets and remove duplicates, ensuring clean and reliable data for the project.