Visualization of FDA Reports on Adverse Events

Eric Yang and Daniel Tan

May 2nd, 2021

Adverse Events Dataset Description

The adverse events dataset[1] is compiled by the U.S. Food and Drug Administration Center for Food Safety and Applied Nutrition (CSFAN) Adverse Event Reporting System. The agency collects reports from consumers about adverse events involving in-market products regulated by CSFAN . These products include foods, dietary supplements and cosmetics. The reports were created between 2004 to 2017. It is of note that these reports are collected from consumers and do not conclude whether or not the product actually caused adverse events.

Dataset Summary

Variable Name	Description	Data type	No. elements	Other information
RA_Report #	Report ID	String	90.8k	NA
RA_CAERS Created Date	Report creation date	String	90.8k	Range: 01/01/2004 to 06/30/2017
AEC_Event Start Date	Event start date	String	53.7k	NA
PRI_Product Role	Role the product played in event	String	90.8k	2 unique values: suspect or concomitant
PRI_Reported Brand/Product Name	Product name	String	90.8k	45.7k unique values
PRI_FDA Industry Code	FDA industry code[2]	Integer	90.8k	NA
PRI_FDA Industry Name	FDA industry name[2]	String	90.8k	41 unique values
CI_Age at	Victim's age	Integer	52.9k	Mean: 50.8 years, Std. dev:

Adverse Event	at time of incident in years			23.2 years
CI_Gender	Victim's gender	String	90.8k	65% female, 30% males, 5% other
AEC_One Row Outcomes	Outcome experienced by victim	String	90.8k	E.g. 'Non-serious injuries/illness', 'Visited an ER'
SYM_One Row Coded Symptoms	Symptoms experienced by victim encoded by the MEDRA[3] ontology	String	90.8k	33.5k unique values, e.g. 'ovarian cancer', 'choking'

Information to be Derived from Visualization

- Commonly reported product or product types that cause adverse events
- Profile of consumers who report on products that cause adverse events
- Common food products that cause adverse events per age group/gender
- Commonly reported medical outcomes caused by adverse events
- Commonly reported medical symptoms per food or per age group caused by adverse events
- Relationship between symptoms and outcomes caused by adverse events
- Products or product types that commonly cause adverse events at different times of the year

Target Audience

We intend to design and build our visualization primarily for FDA regulatory personnel who are interested in understanding profiles of adverse events caused by in-market products under their jurisdiction. The profiles integrate basic victim, product and medical outcome information to provide regulators a comprehensive view that accounts for all components of an adverse event. Regulators can use this visualization for exploratory analyses and allow these insights to guide regulatory processes and decisions on products under review/in-market. It is important that the target audience and use case is specific in-order to build an effective visualization. However, we also brainstormed some other unintended audience that can potentially benefit from our visualization. They are the following:

- Food, supplement and cosmetic companies and manufacturers can use insights derived from our visualization to guide product development, testing and user-study processes.

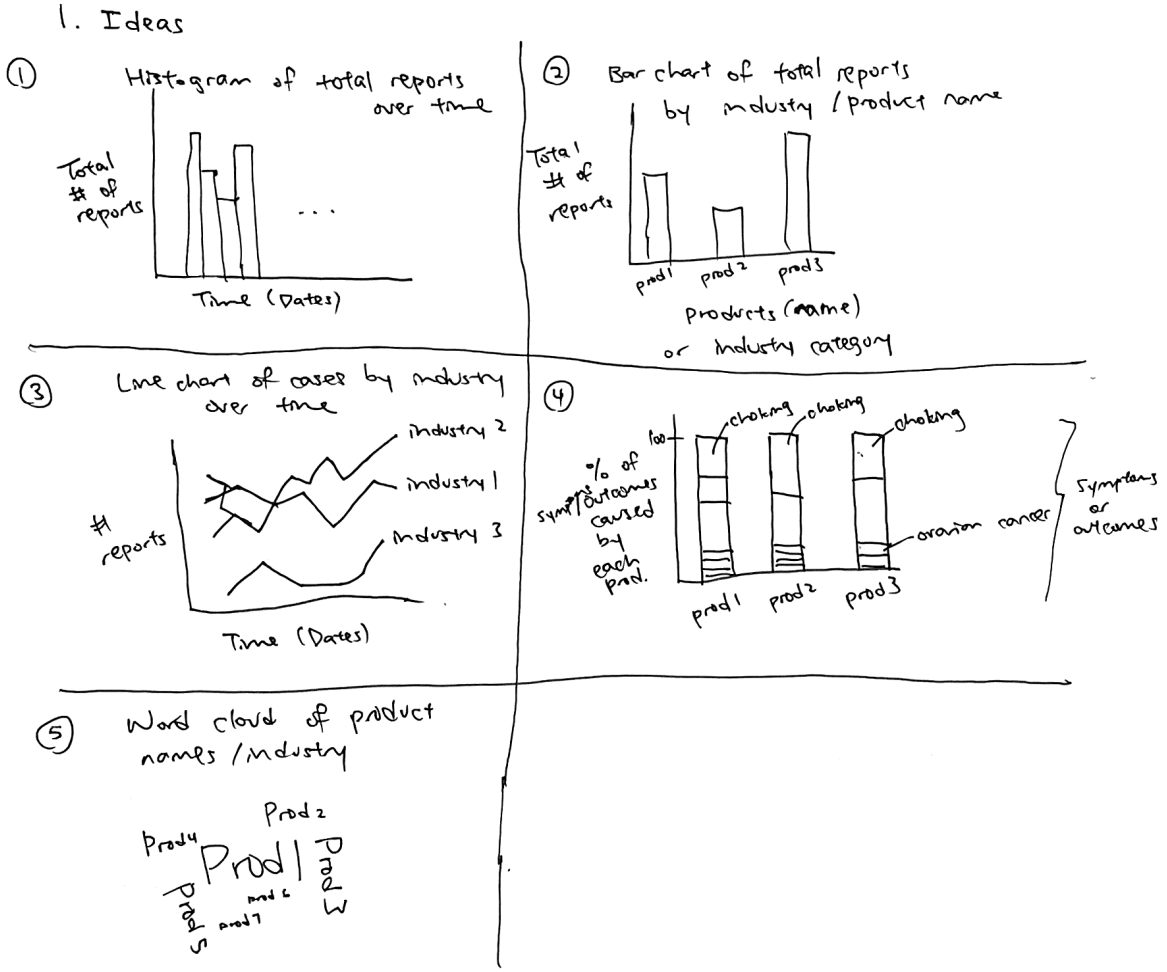
- Health insurance companies who want to understand what types of lifestyle choices (ie diet, cosmetics) that put consumers of specific age groups and gender at risk.
- Hospitals and clinicians who look to find patterns in certain disease causes among patients diagnosed.
- Nutrition scientists and basic science researchers can use the visualization to formulate hypotheses about potential harmful compounds and test whether or not these effects are causative.
- Health conscious consumers of products who want to avoid in-market products that have been widely reported to cause adverse events.

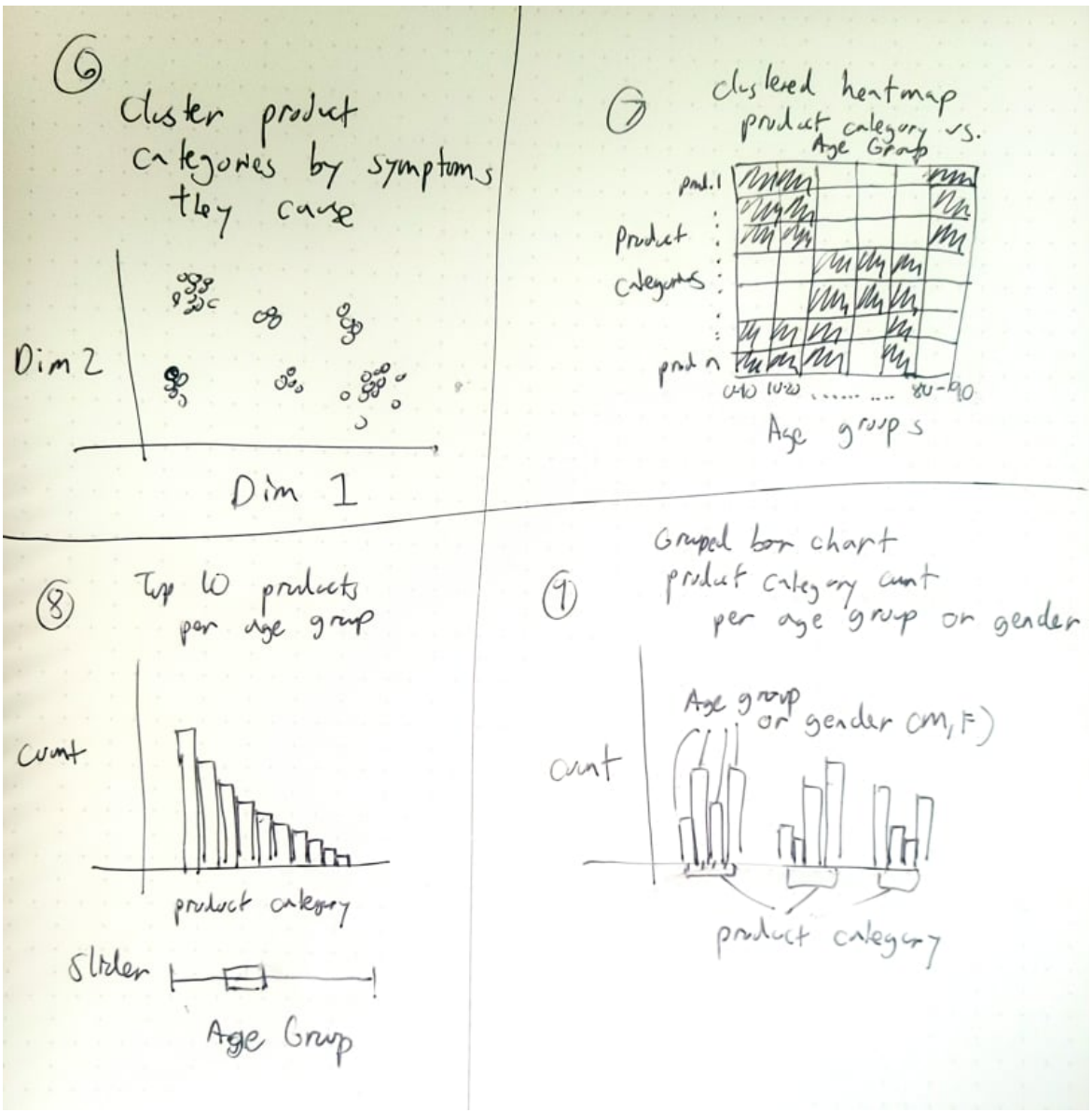
Potential Visualization Tasks

- Identify any trends or correlation of a product's/product type's incidence over time
- Show age distributions of a product/product type and identify what products disproportionally affect certain age groups
- For a specific product, find differences in incidence by gender
- Identify order of most commonly to least commonly reported product/product type overall or by time of the year
- Cluster products or product type by age group/gender
- Find or order most commonly reported symptoms or medical outcomes per product/product type
- Identify association between age group and symptoms

Exploration of Visualization Design Space

Broad idea generation





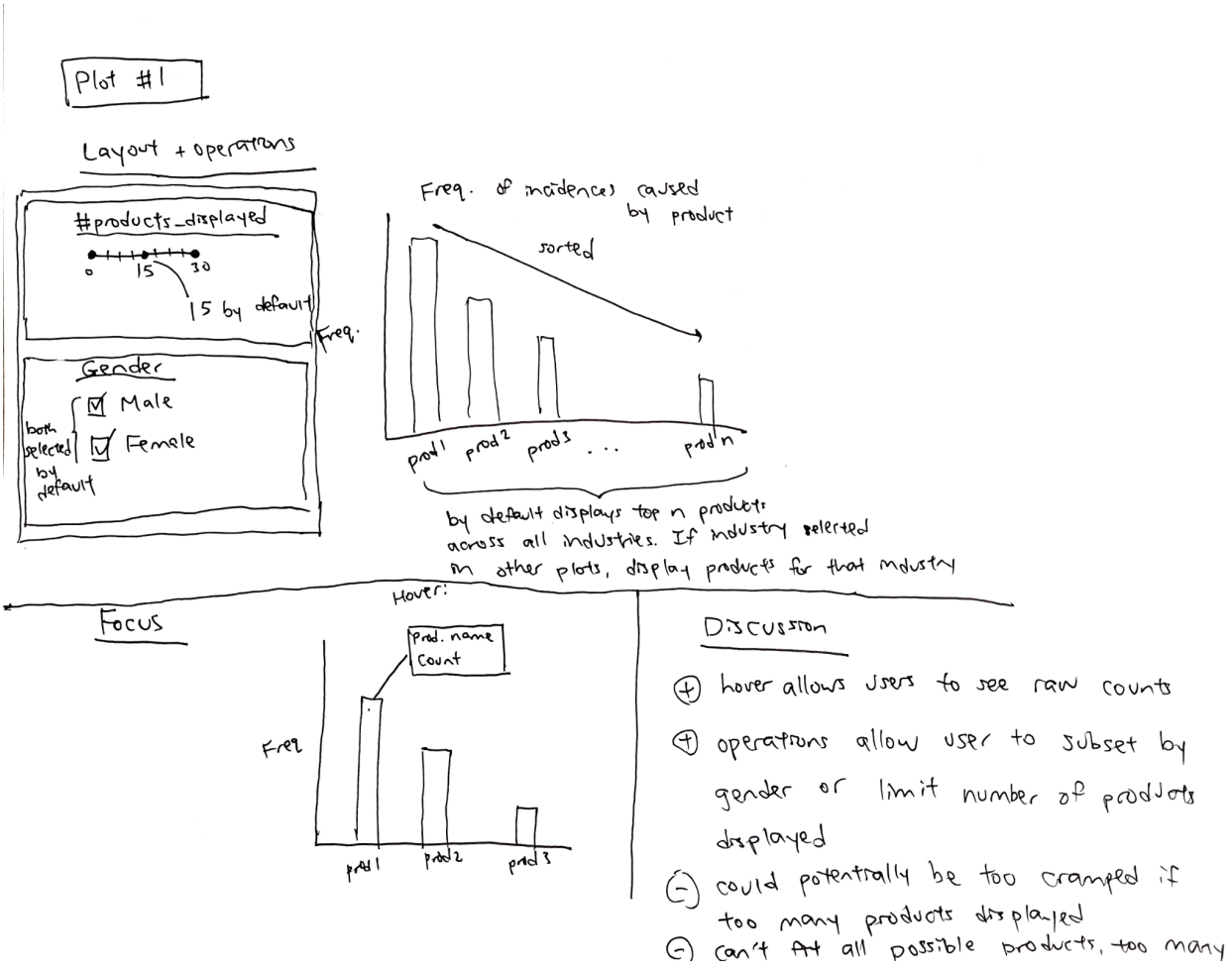
Filter, categorization, strengths and weaknesses of initial ideas

- Ideas 1 and 3 have time components, we can potentially combine those. Idea 1 is very simplistic, 3 allows for more filtering operations (ie group by product or industry)
- Idea 5 shows product make up in all cases, but word clouds might not be great if we want a quantitative visualization.
- Idea 9 is similar to idea 2 except grouped by a specific variable. We can choose either one based on the question we want to answer or what granularity of detail we wish to capture. Idea 4 is also similar, it may not be ideal to implement a stacked bar chart as proportions may be difficult to see.
- Idea 6 is unique, although it requires computation to reduce dimensions.
- Ideas 7 and 8 allow us to see how age groups relate with products or industry categories, they can be combined.

Combine and refine

1. Top products vs count histogram: by default, display top n products, can display industry specific products with industry selection in other plots.
2. Time (year) vs count line plot, each line is a broad industry, might need to create fewer broad categories since there are currently 41 industries.
3. We came up with a new idea not sketched above. Treemap where broad categories are age groups, and within age groups, proportions indicate industry makeup of that age group.

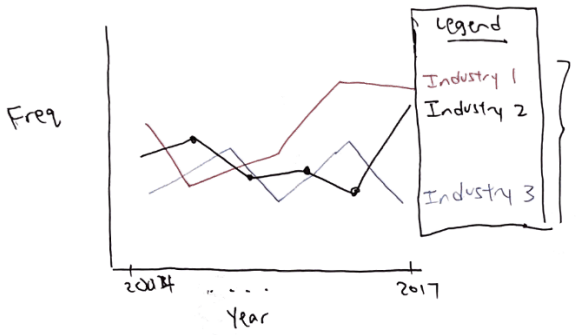
Layouts, focus, operations, and discussion of plot #1



Layouts, focus, operations, and discussion of plot #2

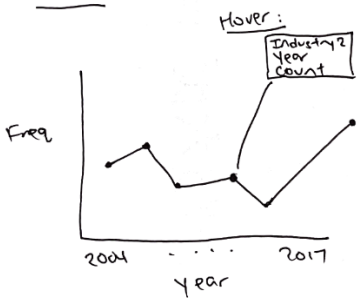
Plot #2

Layout + Operations



since there are currently 41 industries in the table, we will need to create broader industry categories ~10 so the line plot is not too cluttered and colors can be unique + not overwhelming and easily discernable.

Focus



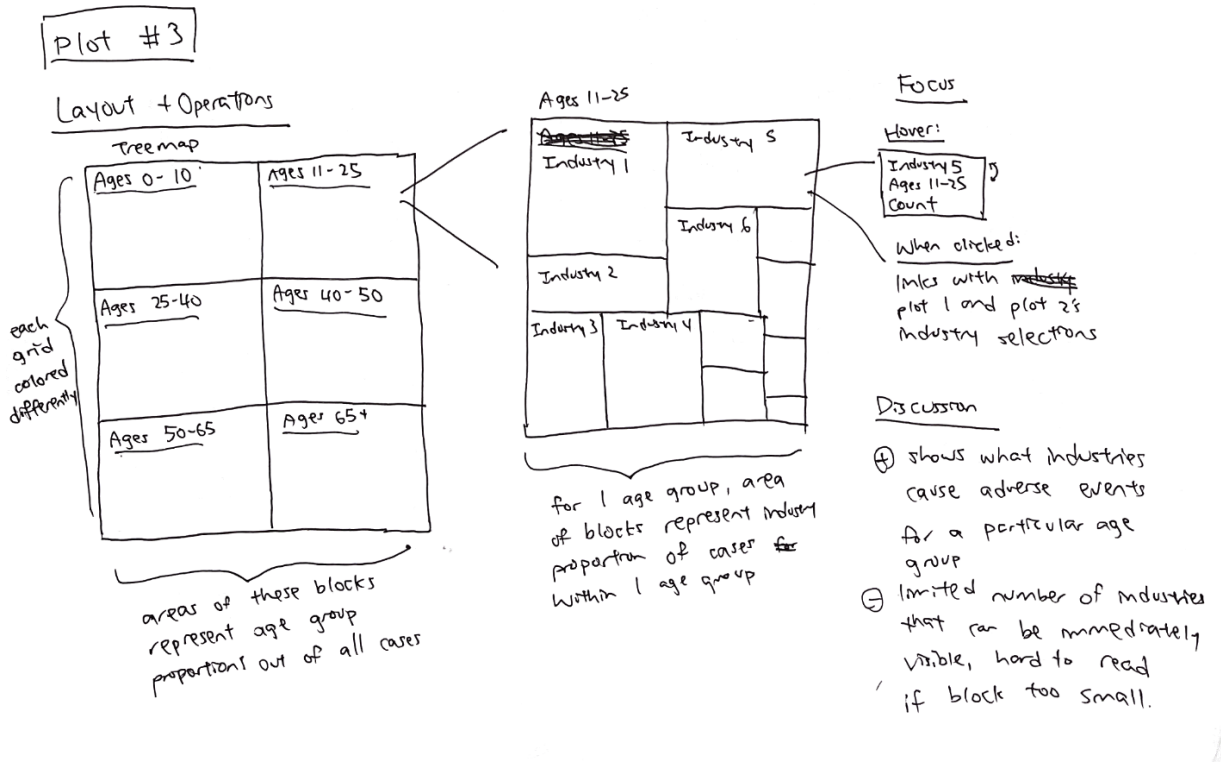
When clicked:
- other lines fade into background
~~(one selected)~~
(only 1 industry can be selected at once)
- also links with plot 3's industry selection and plot 1 displays products for that industry

Discussion

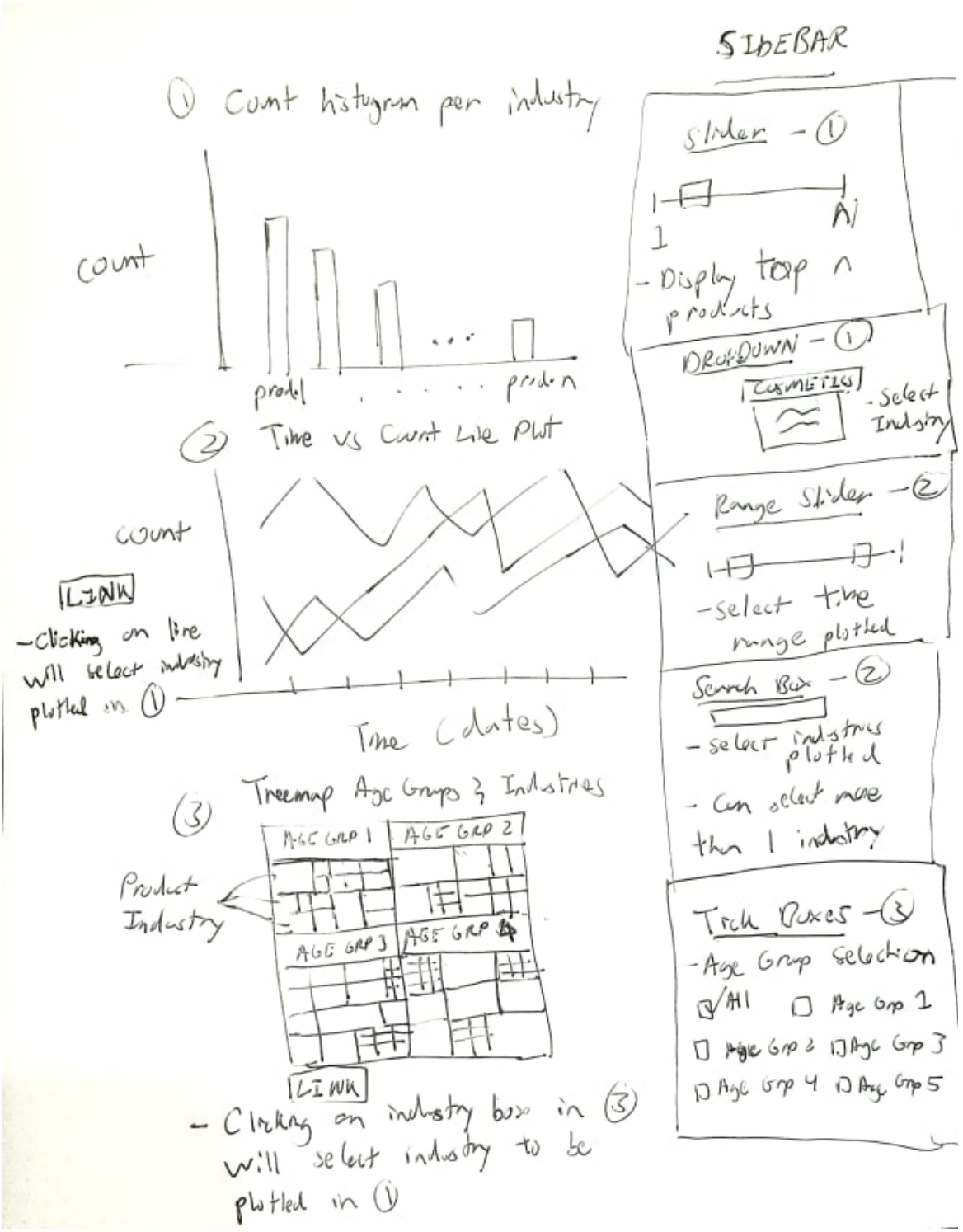
⊕ show changes over time, has some industries improved/gotten worse

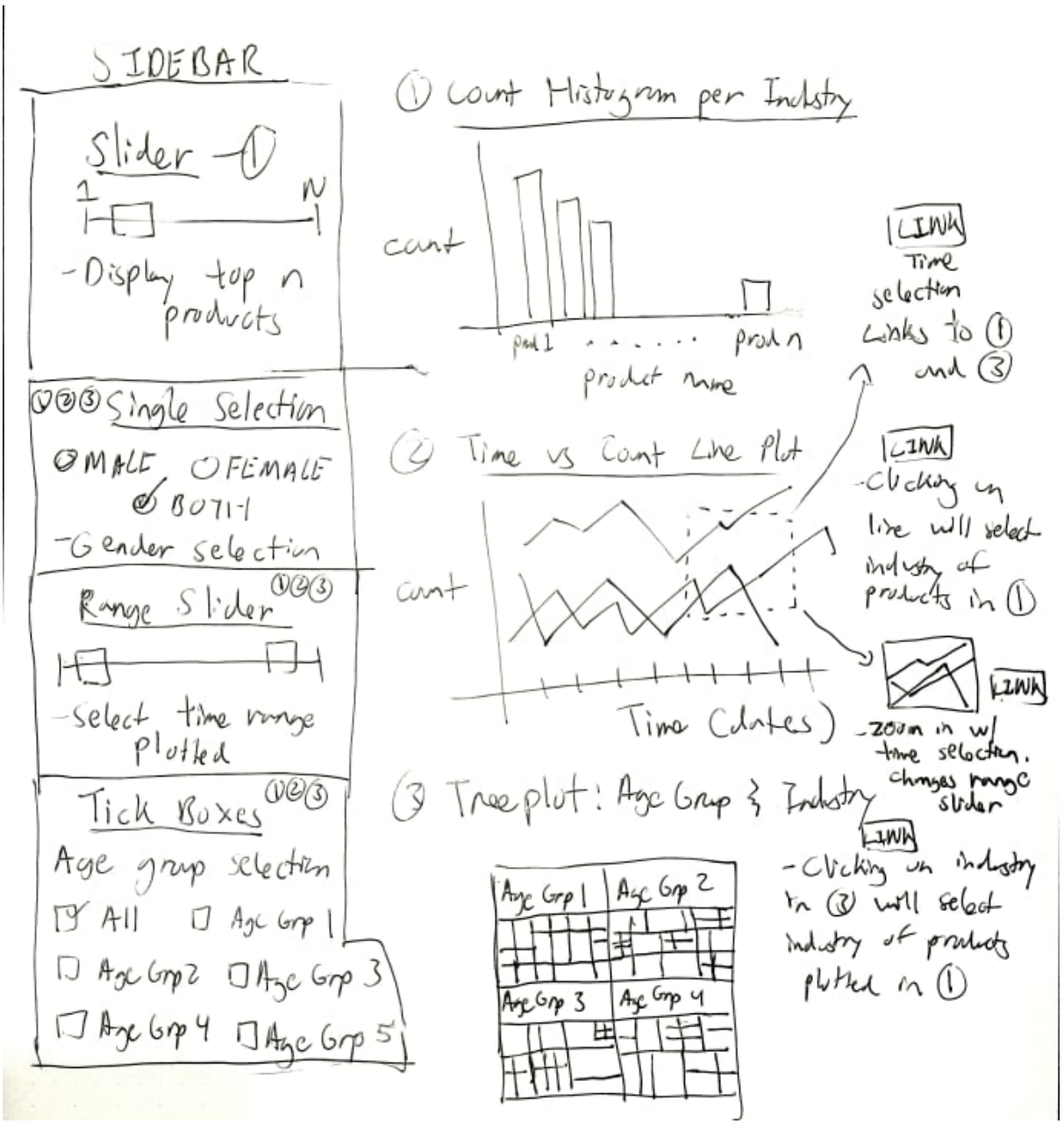
⊖ limited number of industries that can be displayed, need to create broader categories

Layouts, focus, operations, and discussion of plot #3



Overall layout designs, operations and linkage





For the 2nd layout design, we remove the sidebar buttons allowing for selection of product industry/category. In addition we added selection boxes for choosing gender shown in each of the plots. We also create a time highlighting/selection mechanism for our 2nd plot which will link to the two other plots as well as dynamically change the range slider for time.

Potential visualization challenges

Because our dataset is relatively large, with almost one hundred thousand samples and each column containing thousands of unique values, it is difficult to view the whole dataset at once and draw meaningful insights from the visualization. Therefore, we

plan to implement a lot of modularity and interactions for users to subset the dataset based on their questions of interest. In addition, we may need to create broader categories in our final design so as to not clutter our plots if we represent our visualizations with too detailed levels.

Implementation plan

To implement our application, we plan on programming in R, along with packages such as tidyverse and plotly for data wrangling and visualization. The linked brushing and filtering interactions between the widgets will be implemented using the crosstalk package. The visualizations will be deployed as an RShiny application.

Our design will consist of three plots that are linked and will interact with each other. The first plot displays the frequency of reports per product. Because there are more than thirty thousand unique products, we implemented a selector that allows the users to choose the number of products that will be displayed on the bar chart. The remaining side bar operations allow users to subset the dataset, with the selections affecting all three plots. Our second sidebar widget is a single selection checkbox which allows for selection of gender. We will also implement a range slider to select time range as well as tick boxes for filtering for age group, all these sidebar widgets will link to all plots and will be implemented via the crosstalk package.

Additional functionality include zooming in/highlighting time ranges in the second plot. This will dynamically update the time range slider as well as the first and third plot. This can be implemented using event handlers that let us update the UI controls in response to user interactions with the Plotly plot. In the third plot, clicking on specific product categories/industries will also link to the first two plots and filter for their output for the chosen category. These can specifically be implemented using an event_data("plotly_click") function call.

EDIT: We modified our design plan during implementation. Most noticeably, we changed our third plot to a stacked bar chart to easier compare quantities of product industries for each age group. Because there is no inherent hierarchical structure, we abandoned our initial plan of implementing a treemap.

Screenshots

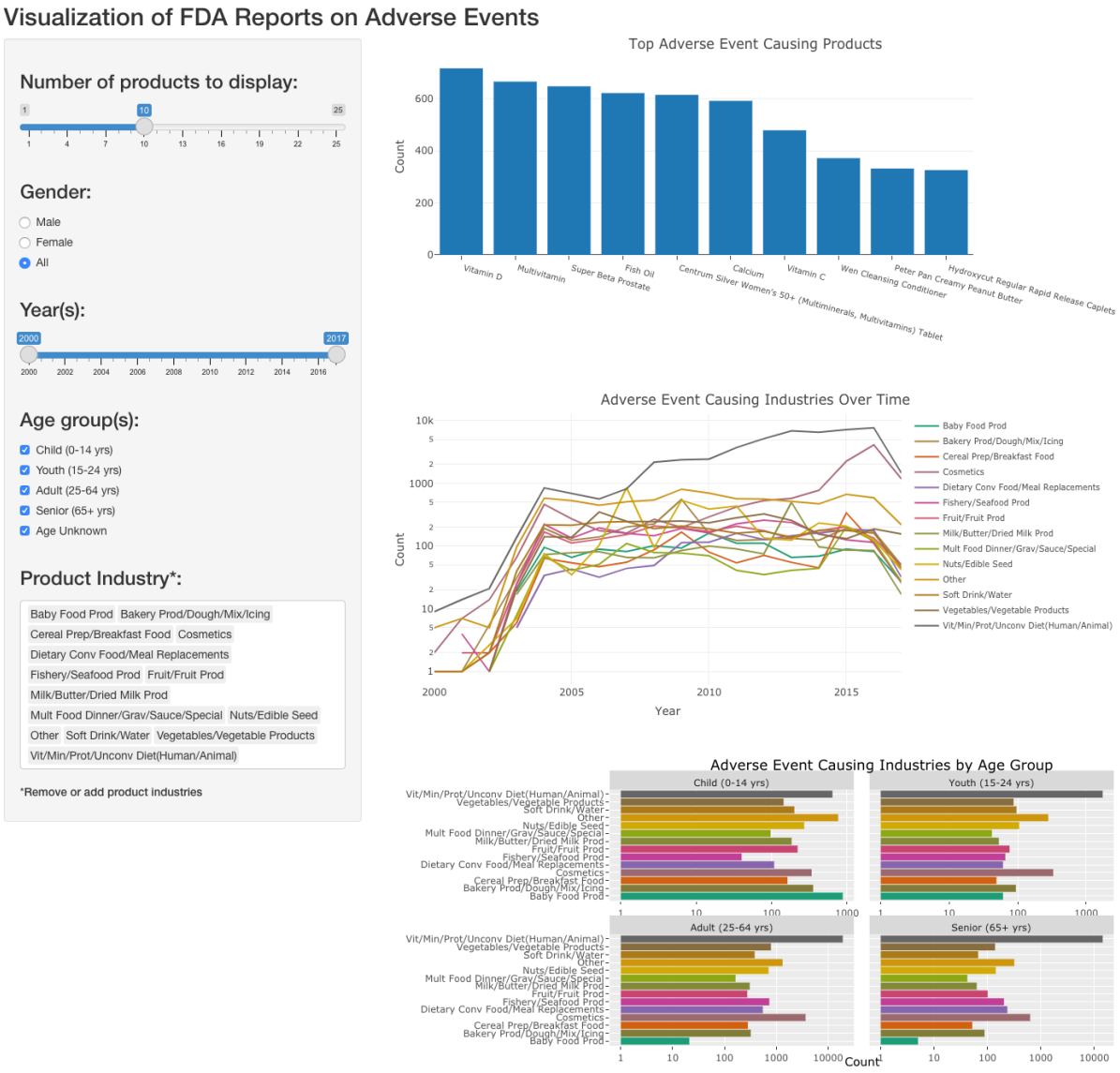


Figure 1. Overview of Shiny app upon first loading. Sidebar on the left allows for data subsetting. Three plots on the right are stacked on top of one another.

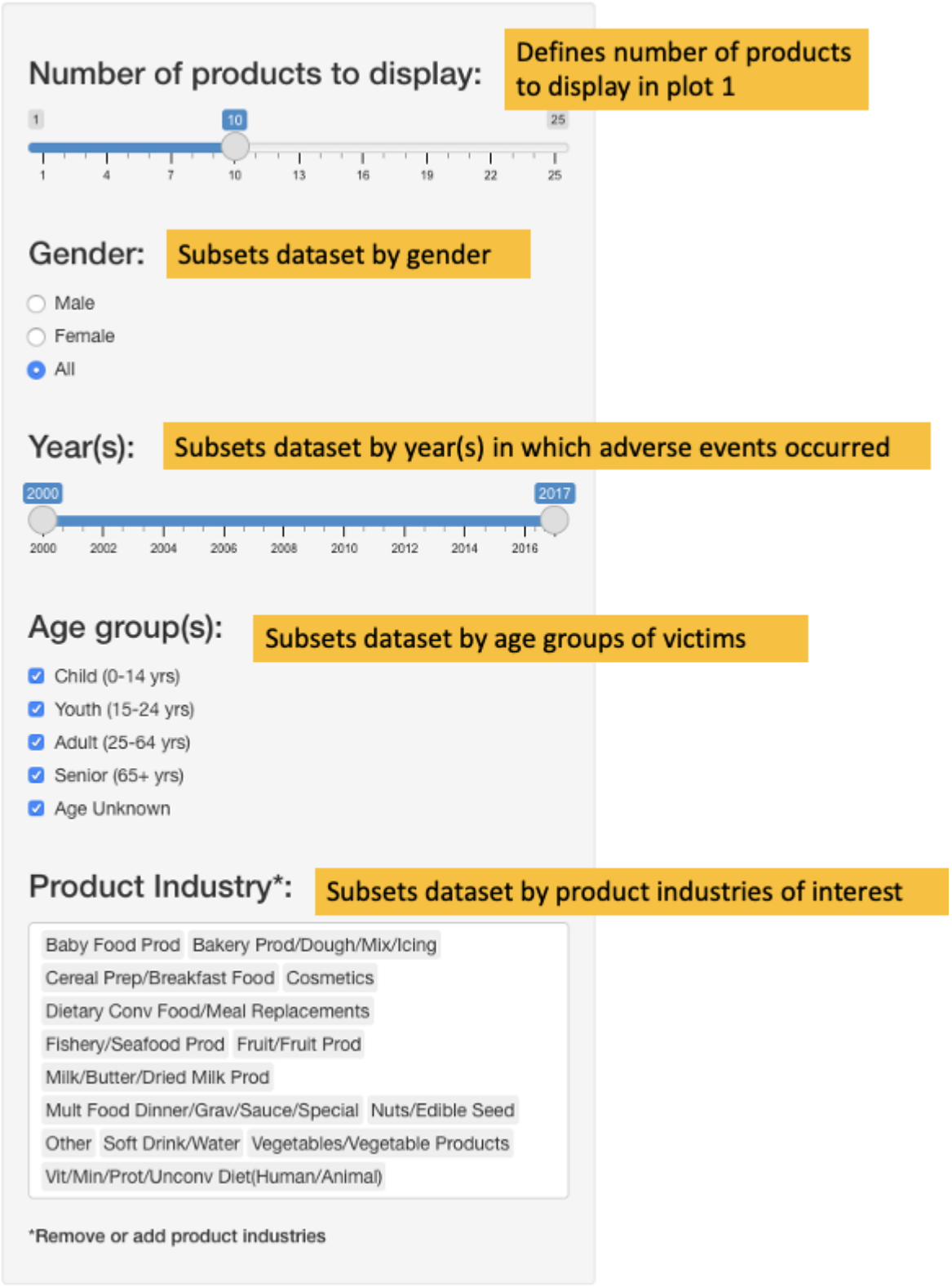


Figure 2. Specifications of selection tools in sidebar.

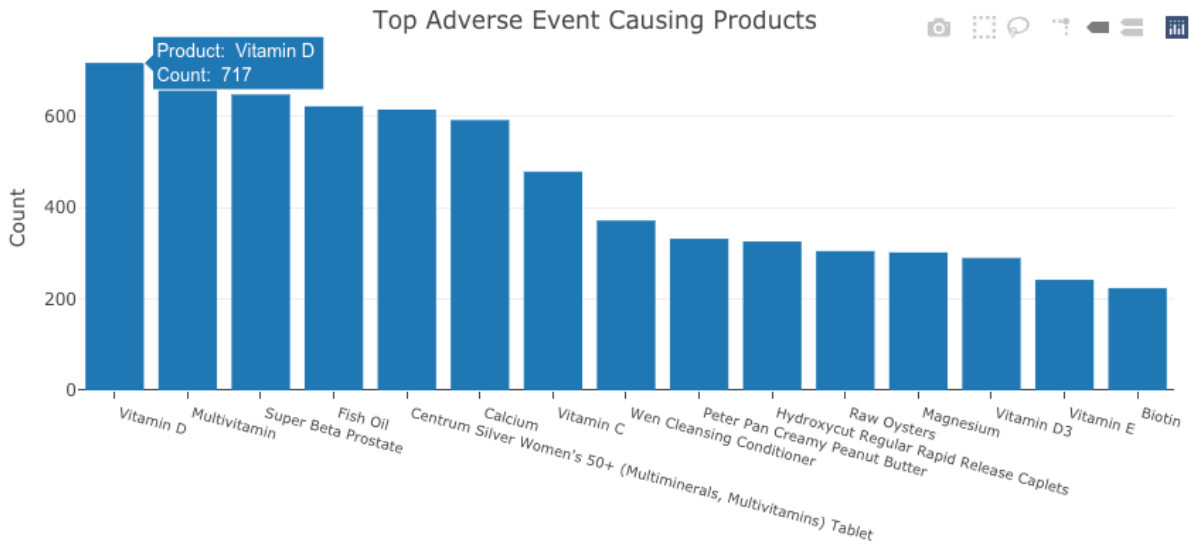


Figure 3. Plot 1 displays the top adverse event causing products based on user subsetting and visualization setting. The 'number of products to display' is set to 15 here. Hovering displays exact quantities.

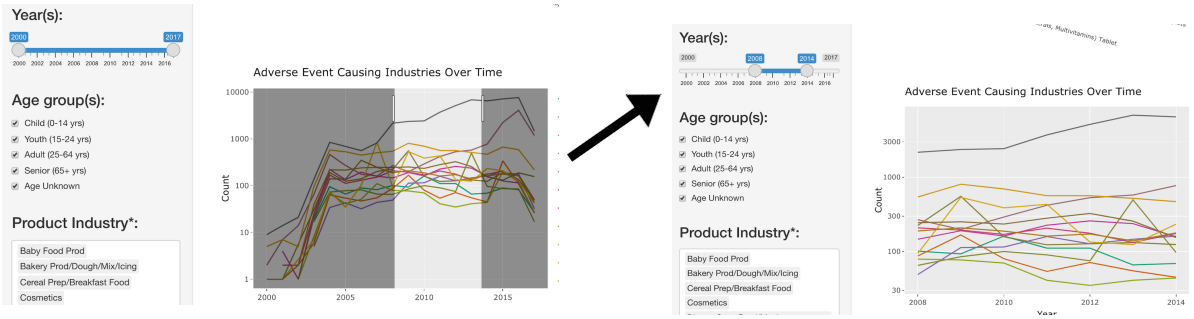


Figure 4. Plot 2 displays the adverse event causing industries over time. Zooming in Plot 2 dynamically updates the time range slider. The interaction between dragging and modifying the range slider is bi-directional. Clicking on industries on the legend also adds or removes lines from the plot. Hovering displays exact quantities.

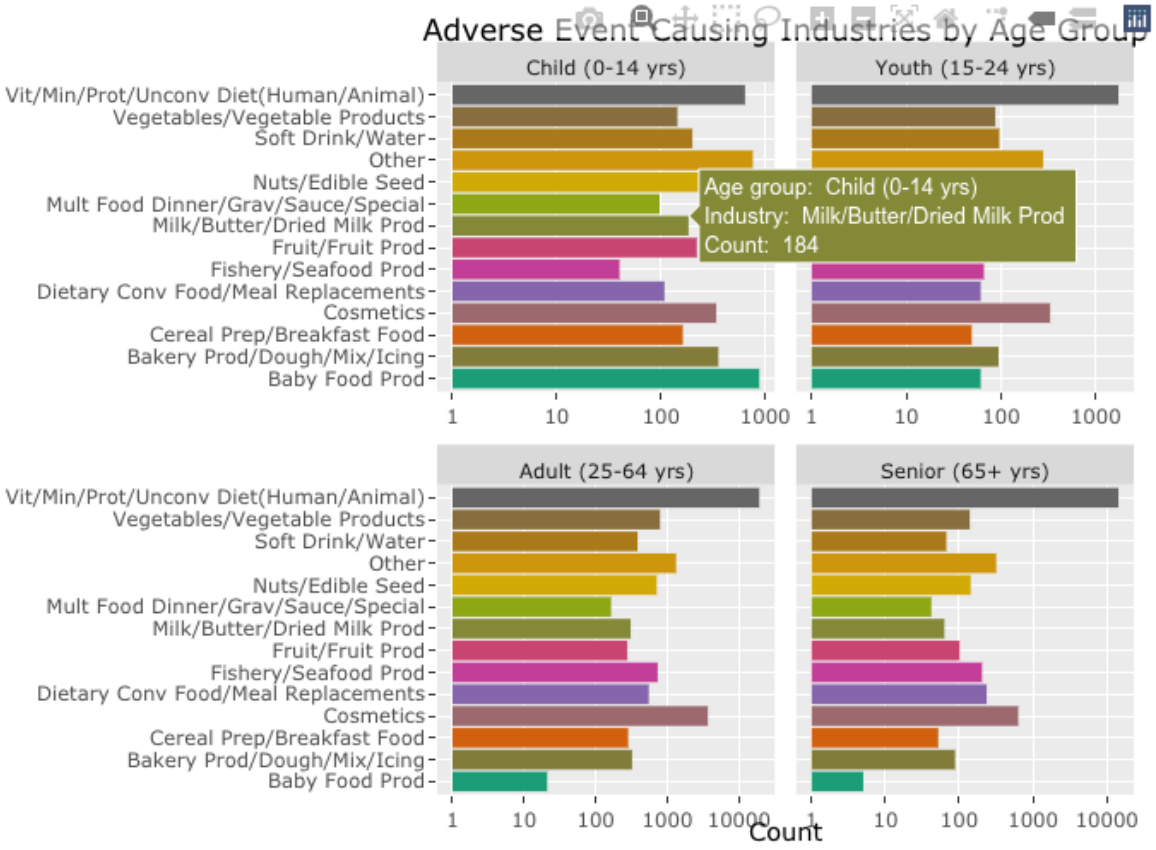


Figure 5. Plot 3 displays the adverse event causing industries faceted by age groups. The industries are colored consistently with plot 2. Users are also able to remove uninterested age groups or product industries from the plot. Hovering displays exact quantities.

Future Work

Because of the heterogeneity of the dataset, we were not able to utilize every feature for the scope of this project. In addition, there were many potential alternatives to plots, displays and interactions that can help augment our current implementation. Below, we listed some concrete steps to take for future work.

1. Add more interactions between plots. For example, clicking a line in plot 2 or a bar in plot 3 will highlight that particular product industry in the other plots, and update the input selection widget in the sidebar.
2. Create a k-means cluster plot of similar products based on different variables i.e. symptoms, outcomes, gender, age group, etc.
3. Visualize most commonly reported symptoms or medical outcomes per product/product type. If we are comparing between products, we can use a stacked bar chart or layered bar chart. Alternatively, if we want to illustrate the contribution of individual outcomes/symptoms counts to the total, we can use a stacked bar chart.
4. Visualize associations between age group and symptoms. This can be implemented via a grouped bar chart or layered bar chart.

5. For a specific product/industry, find differences in overall-incidence or symptom/outcome-specific incidence by gender. This can be done through a simple bar chart, or a line plot with a separate line for each male and female if we are interested in changes in event incidence caused by that product over time.
6. Create heatmap as another way to show change in adverse event incidence for each product/product industry over months of a single year or over several years. Clustering can also be done to identify similarities between products/product industries. Alternatively, instead of depicting industry vs time, we can create a heatmap of absolute values of the pairwise correlation coefficients between industries.

Team Member Contributions

Daniel:

- Explored dataset makeup, key statistics
- Brainstormed visualization tasks, audience, and information to be derived
- Created design sheets of overall layout
- Identified visualization challenges and implementation plan
- Created plot 2 and plot 3, along with their interactions with time range slider, gender radio buttons, age group checkbox, and product industry input selection. Implemented zoom functionally of plot 2 connected to time range slider update
- Created screenshots and brainstormed areas for future work
- Demonstrated Shiny App functionality in video submission

Eric:

- Identified dataset, explored dataset makeup, key statistics
- Brainstormed visualization tasks, audience, and information to be derived
- Created design sheets of specific plots
- Identified visualization challenges and implementation plan
- Cleaned data, created sidebar and plot 1, along with its interactions with sidebar inputs
- Created screenshots and brainstormed areas for future work
- Created powerpoint presentation

References

1. "Adverse Food Events". *Kaggle.Com*, 2021, https://www.kaggle.com/fda/adverse-food-events?select=CAERS_ASCII_2004_2017Q2.csv.
2. "Product Codes And Product Code Builder". *U.S. Food And Drug Administration*, 2021, https://www.fda.gov/industry/import-program-resources/product-codes-and-product-code-builder.

3. "Medical Dictionary For Regulatory Activities Terminology (Meddra) | NCBO Bioportal". *Bioportal.Bioontology.Org*, 2021, https://bioportal.bioontology.org/ontologies/MEDDRA?p=summary.