DATA 511A: Data Visualization For Data Scientists

C4: Concept Production

Team Viz-a-Viz

By:

Zach Bowyer

Prerit Chaudhary

Sagnik Ghosal

Raman S V

Adithyaa V

MS in DS batch of ’24

12/13/2022

Executive Summary

In this work, we discuss our concept development, research, and implementation of an interactive dashboard that outlines the environmental impacts of food production. We came up with this idea because we all believe that being educated on food affordability, healthy diets, and climate change matters. Although this visualization is helpful for the general public, it will be more relevant for environmental protection agencies, government officials, major event organizers (such as world cups or the Olympics) and agriculturalists. In fact, as part of our out-of-class interviews, we surveyed some potential users falling in the above categories to get detailed feedback on our visualization concept. Our goals in this work were to find various facts about climate emissions when related to food production and report them to our audience. To accomplish this, we had to research, explore, refine, clean, and fuse many different datasets to be able to craft specific visualizations for our dashboard. Additionally, we made two different prototypes of our dashboard, one being a paper prototype and the other being a low fidelity Power BI dashboard. For each prototype, we were able to interview potential users and get feedback which we then used to refine our concept. From there, after changes were made, we presented our finalized dashboard to the class on 12/8, to which we received more feedback, which will be outlined below. Finally, we cap off this paper with a critical evaluation of our success and application of class concepts, as required by the rubric. For ease of access, in addition to putting our dashboard link in the appendix, we include it here - Dashboard Link.

Concept Background

As a team, we believe that telling stories about the environmental impact of food, its' affordability, and the production is important for various reasons. First, food affordability is a key indicator of food choice and access, which in turn drives dietary patterns, food production, and health. Next, climate change has been an ongoing yet slow crisis. Sources show that in the next one-hundred years, given the current global emission trajectory, the earth will have an increased average temperature, the sea level will rise, the oceans will be more acidic, ecosystems will morph, and precipitation patterns will change[1]. Thus, it would be beneficial if we could quantify the role of emissions from food production in climate change. Finally, we believe that exploring the production of various foods is important as we can pinpoint who is responsible for the world's production of foods, why they make specific foods, and what the diets of various geographical locations are. Given these reasons, our stance is that all people should have access to information about their diets, and be able to understand the impact their diets have on both themselves and the rest of the population. While it's straightforward that our audience is the set of all people who consume food, through various personal perspectives, we specifically wanted our stories to be told from a point of view of three main consumer types: Meat eaters, vegetarians and vegans and, those somewhere in-between, as we were of the prior impression that meat-based diets are objectively worse for both the environment and personal health.

Given this background, our goals for this project were the following –

1) Analyze diets from a cost point of view
2) Analyze diets from a nutritional point of view
3) Examine the greenhouse gas emissions of various geographical regions both historically and currently, and draw upon observations to compare/contrast the environmental impact of food production with other less-reported sources of emissions, like private flights that sports teams take.

Further, as outlined in our C1 submission, some of the broader questions we are trying to answer in this data analysis are: “What are good diets from an ecological perspective”, “What regions are responsible for food production and consumption of particular products”, and “How do the emissions from food compare to those from traditionally-direct fossil fuel usage sources”. We also explored if other related works on this topic exists, and we found that various related aspects of this work have been published from sources such as John Hopkins, Nature, Scientific American, etc. [2], [3], [4], [5], [6]

Naturally, for this project to be successful, we needed to ensure that we had access to relevant data sources. Below, for each dataset explored, we will mention its origin, briefly profile it, talk about its cleaning process, and talk about its contribution to the final dashboard.

Worldwide meat consumption - the first dataset explored was a Kaggle dataset on worldwide meat consumption[7]. This dataset was 133KB in size and consisted of 5 columns and 13761 rows. The column headers were ‘Location’ ‘Subject’ ‘Measure’, ‘time’, and ‘Value’. The first column titled 'Location' is a categorical description of which specific country out of 48 is being looked at. The second column, 'subject' is another categorical column that determines what type of meat is being considered (Beef, Pig, Poultry, Sheep). Next, the 'Measure' column is a categorical column that describes the units of the 'Value' column (Kilograms per person annually or thousands of tons per region annually). The fourth column, 'time' is a categorical column that shows the year the data was recorded (1991-2026- which means there are some projections). Finally, the 'Value' column, which is a regressive column, shows the meat consumption for that country, time, meat type, etc (0-131607.31). Fortunately, this dataset did not have to be cleaned, however it was not used in the final dashboard as we found a better data source with a wider scope of details on food items.

Food Products Constituents - the next dataset explored was about the constituents of specific food products [8]. This dataset is 1KB in size, and consists of 5 rows and 20 columns. One important thing to note is that the dataset was handmade from individual sources that are all appropriately referenced in the set as its own column. With that being said, the purpose of this dataset was to show the various nutritional facts (e.g., calories, fat, cholesterol, sodium, vitamins) of beef, goat meat, chicken, wheat, and pig meat. In terms of cleaning, we had to normalize the serving sizes to properly account for calories per kilogram of meat. In terms of final usability, we implemented this dataset in our final dashboard as figure 12, which is the constituent’s view.

Crops and livestock products data - for Figure 14 of the final dashboard, we used a combination of two datasets from the sources listed in references [9] and [10]. The first dataset represented the greenhouse gas emissions per kilogram of food product. It was 4KB in size and consisted of 23 columns and 45 rows. Each column corresponded to various climate metrics such as: “Freshwater withdrawals per 1000kcal (litres per 1000kcal)”. However, we only used the “Total_emissions'' column. Additionally, each row corresponded to a specific food type, such as “Milk”, “Goat”, “Peas”, etc. The second dataset in this pair, which contained details about the production over time of specific foods for all countries, was 20MB in size and had 13 columns and over 100,000 rows. The main four columns that we used here was “Area”, which refers to a country or region, “Item”, which refers to a specific food, “Year”, and “Value”, which corresponds to how many tons of that specific item were produced. For these two datasets, we focused our efforts on fusing the sets together so we could estimate the total emissions of food production for specific food items both by year and by country/region. The fusion process entailed data wrangling and cleaning in the form of dropping rows/columns and filtering, which was all done with python. This cleansed dataset was used in the final dashboard in the food emissions details view.

Emissions Data - for emissions, we came across a website titled, “Our World in Data”, which quantifies the effect of calamities like poverty, hunger, climate change, etc., which is in turn used to help people track what progress is being made to reduce the suffering in the world. From the above source, we leveraged a set of datasets from these references - [11][12][13] - that contained information on the food-wise production in tonnes per year for different countries between the years of 1961-2020. Each food type had a separate excel file and the data was available for food types such as meat, fish, cereals, rice, etc. Each file had 4 columns, respectively titled as “Country Name”, “Country Code”, “Year”, and “Food Produced in tonnes”. Additionally, we came across a dataset [14] which contained data for greenhouse gas emissions across various economic sectors for different countries over the last few years. It was a 1174KB file which consisted of 14 columns and 6151 rows. Each row corresponded to ghg (greenhouse gases) emissions for a country in a given year for various sectors such as agriculture, manufacturing and construction, electricity and heat, buildings, aviation and shipping, etc. The data manipulation process here entailed cleaning in the form of dropping unwanted rows/columns, and filtering, which was again done with python. Eventually the data was output in a format that was directly leveraged to create the visualizations.

Miscellaneous Data - finally, as part of our visualization, we addressed the question of how relevant agricultural emissions are to the climate change crisis compared to other sources of greenhouse gasses. To illustrate these visually, we have used the dataset termed “Interesting Data''. This dataset comprises the different contemporary sources of emissions. The references are either sports events, natural causes, or luxurious activities of the super-rich, which all contribute to greenhouse gas emissions. Specifically, the events we gathered data on are - soccer world cups, Olympics, NBA games, NFL games, IPL team travels, forest wildfires, burning of fossil fuels, and billionaire’s private jet travels. These data sources are from published news articles, blogs, Wikipedia pages, and social media platforms such as Twitter. There are 22 data points collected from 8 countries: Russia, Brazil, India, United Kingdom, United States, Japan, Australia, and South Africa. The number of columns in this dataset is 4 where the first represents the country name, second the event name, third the year, and fourth the details of the emissions in tonnes.

Process Description

In this section, we will describe the various steps we took to prototype, analyze, and modify our visualization concept over the past month. We first started this process with a basic proposal, which was the C1 assignment. In this assignment, we defined our project concept and goals, described our intended users and tasks, linked related works, and explored a handful of datasets.

We then participated in the FDS activity, which was a brainstorming process on visualization concepts based on our previous research and datasets. To summarize the process, we started from individual brainstorming sheets, where each person would give examples of what visuals they found appropriate. From there, we took those brainstorming ideas together and designed 5 separate sheets. Sheet 1 was a ‘final brainstorming’ sheet where we brought the best of all ideas together and discussed various ideas for how we could filter, refine, combine, and categorize our data. Next, sheets 2, 3, and 4 were various visualization examples. Finally, sheet 5 was a final draft of each section of our dashboard, which would consist of one static view and three dynamic views. Overall, the FDS activity led to productive conversations about how we were going to tackle the project. As stated in our C2 assignment, “one key takeaway from this activity was that we realized that we were better off asking interesting questions about a topic we were passionate about (climate change and world health) before looking for datasets since we knew data was abundant”.

Later on, we were able to put our brainstorming session to test through the C3 assignment, which required us to design a paper prototype of our visualization concept to present to three different users in class. In terms of the prototype itself, we had three main pages. First would be the static page of our dashboard, which consisted of two bar graphs titled “Worldwide Emissions from Agriculture” and “Average Emissions for all Countries.” From the static page, we had two buttons labelled “Figure 1” and “Figure 2”, which, when interacted with, would take the user to more detailed pages. The first detailed page consisted of four different interactive figures. Each figure represented a specific food category: cow, wheat, pig, and chicken. The user could interact with each figure, and when done so, the figure would display a bar chart/slice chart describing the CO2 impact of the consumption of said food category. At a higher level, the four figures represented data for a specific country. Thus, this detailed page also included a filter box, allowing the user to view the figures for any country. Next, the second detailed page was used as a comparison tool to see how the emissions in specific economic sectors differed between countries. For each country, an interactive figure was available that would show a bar chart for the respective economic sector in that country. Naturally, the user could change sectors in that figure as they pleased. Additionally, users could use two dropdown filters on this page to select any pair of countries to compare. Finally, both detailed pages had a back button that would take them to the static page.

For the user interviews, we had Arjun, Tejal, and Rhea try out our workflow and give feedback.

Arjun had quite a few questions on the data sources we used, and he suggested that we would be able to sort emissions on GDP and GDP per capita, and he was unsure on how to proceed in our workflow at a certain point, and mentioned that we could combine bar charts in our country comparisons. Fortunately, he found that comparing agricultural emissions with other contemporary sources was exciting and could navigate through our prototype relatively well. Further, he found some of our filtering options to be intuitive.
Tejal suggested that we should add a sorting filter to side-by-side bar charts and needed some help navigating through the prototype, and thought that comparing individual animals as pictures was a bit confusing. What she did like in our prototype was the fact that our stories were relatively straightforward and easy to follow.

Finally, in our interview with Rhea, she found our titles and labels confusing, noted that some of our buttons weren’t very big, mentioned that some of our plot axes were not clear, and suggested that comparisons between food products should be made in parallel, instead of having just a select few. In terms of positives, Rhea was able to navigate through our concept with ease and mentioned that some of our ideas were unique.

Based on these conversations, we created a few basic dashboard views, which can be seen as Figure 8, Figure 9, and Figure 10. We then demonstrated these views to three other potential users. These users were personal connections of our team members on this project and we decided to keep their personalities anonymous from everyone but the person interviewing them. The first person we interviewed is a professional athlete who follows a nutritionist-approved diet. They were able to understand our concept and mentioned that they look forward to trying to better understand the climate impact of their specific diet. Unfortunately, they needed help at certain times navigating and understanding specific charts.

The 2nd user we interviewed was a person who had transitioned to veganism from a meat-based diet, five years prior. They liked how our dashboard used a static page to show major stories as well as having the ability to drill down into specific information as needed. They had some specific topics they thought would be interesting to add like comparing the water requirements for events to agriculture. They were able to navigate through our prototype relatively easily due to their technological background.

And finally, for our 3rd user interview, we talked to a person who has always adhered to a meat heavy diet. They thought that our prototype was novel and worth exploring in various ways. However, they had trouble understanding the purpose of the static page, and thought we should be able to view more specific foods instead of a select few, which was similar to what Rhea mentioned earlier.

Finally, we capped off our process by presenting our final dashboard to the class on 12/8 via a zoom presentation. This presentation was eight minutes long and consisted of a two-minute segment explaining the various aspects of our work that are covered in this report and a six-minute demonstration of our interactive dashboard. The rest of the class posted constructive feedback about our dashboard on our Slack thread. A few key takeaways from our presentation was that our graphs looked cramped, using animal images as bar charts was somewhat confusing, using time-series whenever applicable is useful, our stories were organized well, and comparing special events to agriculture was interesting to users.

In terms of the contents of our dashboard, we first start with our "Emissions" view, which is our main view which uses Bertin's attributes such as position, color and size to display the CO2 emissions for a few regions across various sectors. We have bifurcated the view into decades for ease of comparison and readability. In our next view, titled "Constituents", we show the nutritional value versus emissions for different food types which allows the viewer to gain a better understanding of which kind of food is healthier as well as sustainable for the environment. Next, we have a view titled "Interesting Data", which allows the user to compare agricultural emissions for a specific country and year to a special event to highlight whether emissions through food production is actually that significant. Next, we have "Emission Details - Food", which allows the user to deep dive on the CO2 emissions. It allows the user to compare and contrast the CO2 emissions across various Food Products, Countries and for different years. Finally, for "Emission Details - GHG", we have created another view which allows for a deeper understanding of the GHG emissions across various sectors for different countries over the last few years.

Based on all the feedback that we received from the above-mentioned sources, we would say that some common issues that users had with our visualization was how the shape of the animals was distorted and whether the area or the height of the animals had any significance. On the other hand, the users appreciated the novelty of the idea where we used images of animals to portray the nutritional content. The view to drill down was also lauded by several viewers as it follows the principle of overview, zoom in and filter on demand.

Further, keeping in view that a majority of feedback clashed with what we believed were aesthetics vs. clarity of thought, we had gone ahead and updated our visuals to better reflect the data while also incorporating the animal/plant figures. This was a minor change that was purely visual and did not entail making changes to the data or the storyboard we had developed so far.

Introspection and Evaluation

We tried to replicate the "The Visualization Mantra", which is a widely known visual information seeking methodology, conceived by Ben Shneiderman [17], which says overview first, zoom and filter, then details-on-demand. This mantra succinctly summarizes the essential elements of interacting with graphically presented information. We have created a static view to reveal the gist of the topic in our first view and then build subsequent tabs for drill down which provides vital details to the context that was set. Further, every data point in this visual can be "drilled-through" to the underlying data. Our schema enables a viewer to understand the bigger picture through efficient and effective exploration of a large data set and also allows them to filter on a subset of a data set using the interface. Stan Pugsley's Data Story Visualisation: A Decision Tree helped us in selecting the right graphic [15]. Recollecting Mackinlay's ranking [16] for encoding data, size is a less evident indicator and requires more time to comprehend and compare than color, which is a very effective indicator for continuous data. We have also tried to incorporate the "Principle of Congruency" that says that the structure and content of the external representation should correspond to the desired structure and content of the internal representation. Graphics are aesthetically appealing and we were able to convey insights that would otherwise need many words to describe. There is a natural correspondence between the graphs as we have used Bertin's attributes such as position, size, shape, and hue to convey the essential information, with one major exception being the animal charts which we felt added some comic relief to an otherwise morbid discussion of pollution. Further, we have ensured that we integrate the principles of graphical integrity and excellence by not displaying data out of context and displaying concise visuals by maintaining consistent and balanced data to ink ratio.

One key aspect where our dashboard zigged while the others zagged, was the use of Microsoft Power BI as our primary visualization software. We chose Power BI as it made it considerably easier for us to share the visual file with our team members. Versioning is an often-faced issue in developing multiple dashboards across team members. And to help ameliorate this to the best possible extent, we first created our entire data model in Power BI. Not just the underlying data cleansing for the multiple datasets, but even the data model generation that helped create a relationship between multiple datasets was designed in Power BI Model view. This helped us create a naive normal form for the datasets and eased interlinking and highlighting of disparate data points in the visuals. Yet, this was not all smooth sailing. We all had access to a Power BI Pro license on our UW email ID's. But, halfway through the development is when we realized that UW doesn't allow map visuals to be hosted on the Power BI server due to the vendor (TomTom) not being included in UW's data security and privacy agreements with Microsoft. Thus, we had to create a temporary trial license for 60 days to help host our dashboard externally and then provide access to specific users via their UW mail ID.

Conclusion

Overall, we believe that we have tried to do justice to the initial premise and tried to explain the impact of food production on the environment. The main reason for selecting this topic was that it resonates with the challenges the world is facing every day and we have tried to convey the message through quantified data. We have also tried to relate this to emissions in other sectors and special events happening across the world to highlight the dissonance between popular discourse and ground reality. Additionally, we have also been able to provide a quick view on the nutritional and emission related data of 5 most common food items. We believe we have achieved all of the above while also adhering to the key principles and frameworks that we had learnt as a part of this course.

References

Hosted Online Dashboard: Power BI Dashboard. Please reach out to Adithyaa V (adi279@uw.edu) if you are unable to access the dashboard via this link.

[1] "Future of Climate Change." climatechange.chicago.gov

https://climatechange.chicago.gov/climate-change-science/future-climate-change#:~:text=Key%20global%20projections,mitigation%20of%20greenhouse%20gas%20emissions. (Accessed Dec. 09, 2022).

[2] Johns Hopkins Center for a Livable Future. "Food and Climate Change." foodsystemprimer.org https://www.foodsystemprimer.org/food-production/food-and-climate-change/ (Accessed Dec. 09, 2022).

[3] Johns Hopkins Center for a Livable Future. "Country-specific dietary shifts to mitigate climate and water crises." foodsystemprimer.org https://clf.jhsph.edu/publications/country-specific-dietary-shifts-mitigate-climate-and-water-crises (Accessed Dec. 09, 2022).

[4] Hedenus, F., Wirsenius, S. & Johansson, D.J.A. The importance of reduced meat and dairy consumption for meeting stringent climate change targets. *Climatic Change* 124, 79–91 (2014). https://doi.org/10.1007/s10584-014-1104-5

[5] Tilman, D., Clark, M. Global diets link environmental sustainability and human health. *Nature* 515, 518–522 (2014). https://doi.org/10.1038/nature13959

[6] Seth Wynes. "Professional Sports Leagues Need to Reduce Their Carbon Footprint." scientificamerican.com https://www.scientificamerican.com/article/professional-sports-leagues-need-to-reduce-their-carbon-footprint/ (accessed Dec. 09, 2022).

[7] Vagif Aliyev, "Worldwide Meat Consumption." (2020). Distributed by kaggle. https://www.kaggle.com/datasets/vagifa/meatconsumption

[8] Syndigo, "Food Database." (2020). Distributed by nutritionix. https://www.nutritionix.com/

[9] Vivek, "Environment Impact of Food Production." (2020). Distributed by kaggle. https://www.kaggle.com/datasets/selfvivek/environment-impact-of-food-production

[10] The Food and Agriculture Organization (FAO), "Crops and livestock products." (2022). Distributed by FAO. https://www.fao.org/faostat/en/#data/QCL

[11] Hannah Ritchie, Pablo Rosado and Max Roser, "Agricultural Production." (2022). Distributed by Our World in Data. https://ourworldindata.org/agricultural-production

[12] Hannah Ritchie, Pablo Rosado and Max Roser, "Meat and Dairy Production." (2017). Distributed by Our World in Data. https://ourworldindata.org/meat-production

[13] Hannah Ritchie and Max Roser, “Fish and Overfishing.” (2021). Distributed by Our World in Data. https://ourworldindata.org/fish-and-overfishing

[14] Hannah Ritchie, Pablo Rosado and Max Roser, “Emissions by sector.” (2020). Distributed by Our World in Data. https://ourworldindata.org/emissions-by-sector

[15] Stan Pugsley. “Data Story Visualization: A Decision Tree.” tdwi.org https://tdwi.org/articles/2020/10/06/bi-all-data-story-visualization-decision-tree.aspx (Accessed Dec. 09, 2022).

[16] Mickael Causse. “The Mackinlay ranking of perceptual task.” researchgate.net https://www.researchgate.net/figure/The-Mackinlay-ranking-of-perceptual-task_fig2_221098028

[17] J. Heer and B. Shneiderman, "Interactive dynamics for visual analysis", Communications of the ACM, vol. 55. (2012)

[18] Lecture Slides for Data511: Data Visualization for Data Scientists

[19] Tufte, Edward R., 1942-. “The visual display of quantitative information.” Cheshire, Conn.: Graphics Press. (2001)

[20] Few, Stephen. “Information dashboard design: the effective visual communication of data.” Beijing; Cambridge [MA]: O'Reilly. (2006)

[21] Few, Stephen author. “Now you see it: simple visualization techniques for quantitative analysis.” Burlingame, California: Analytics Press. (2009)

Appendix

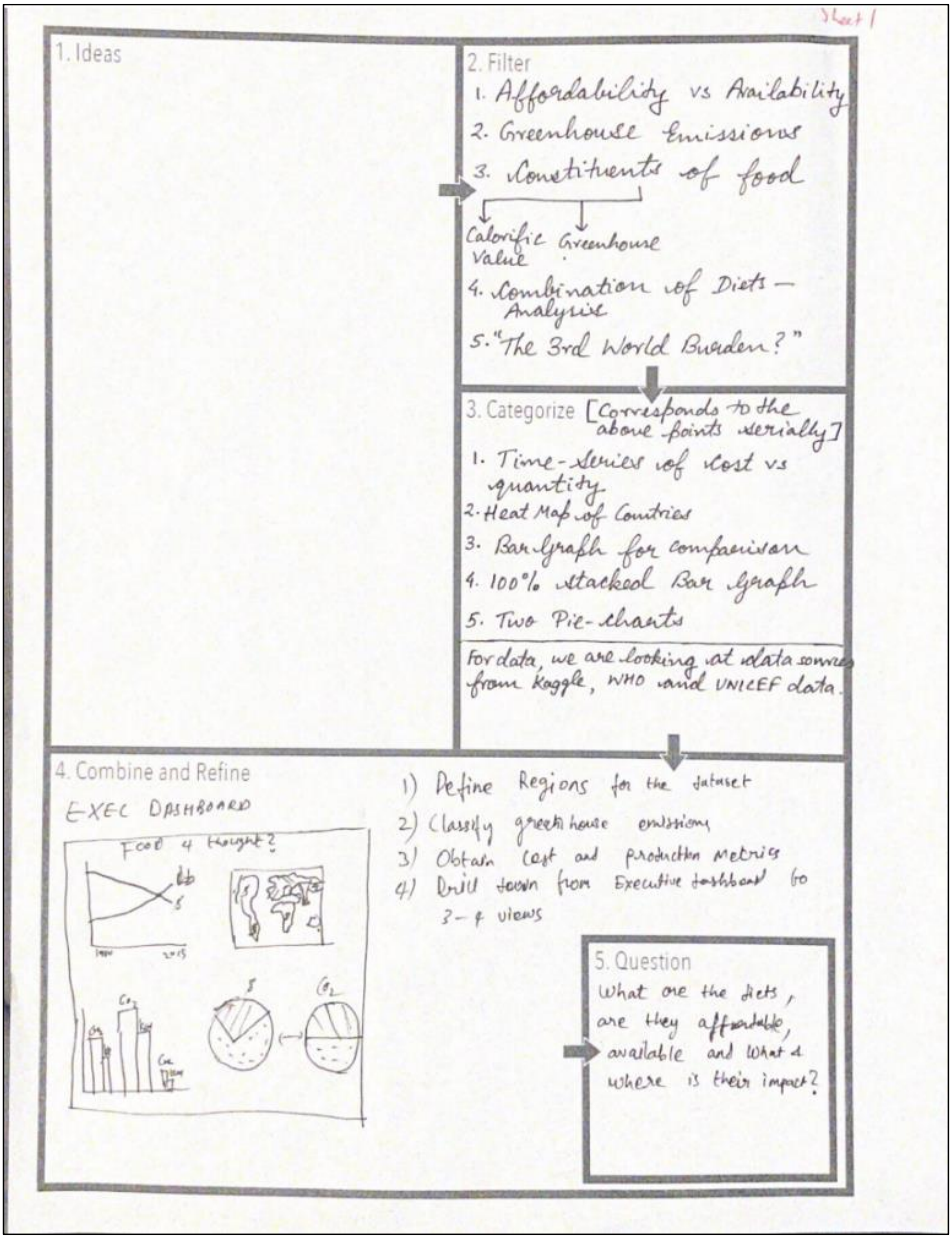

Figure 1: FDS - Sheet 1

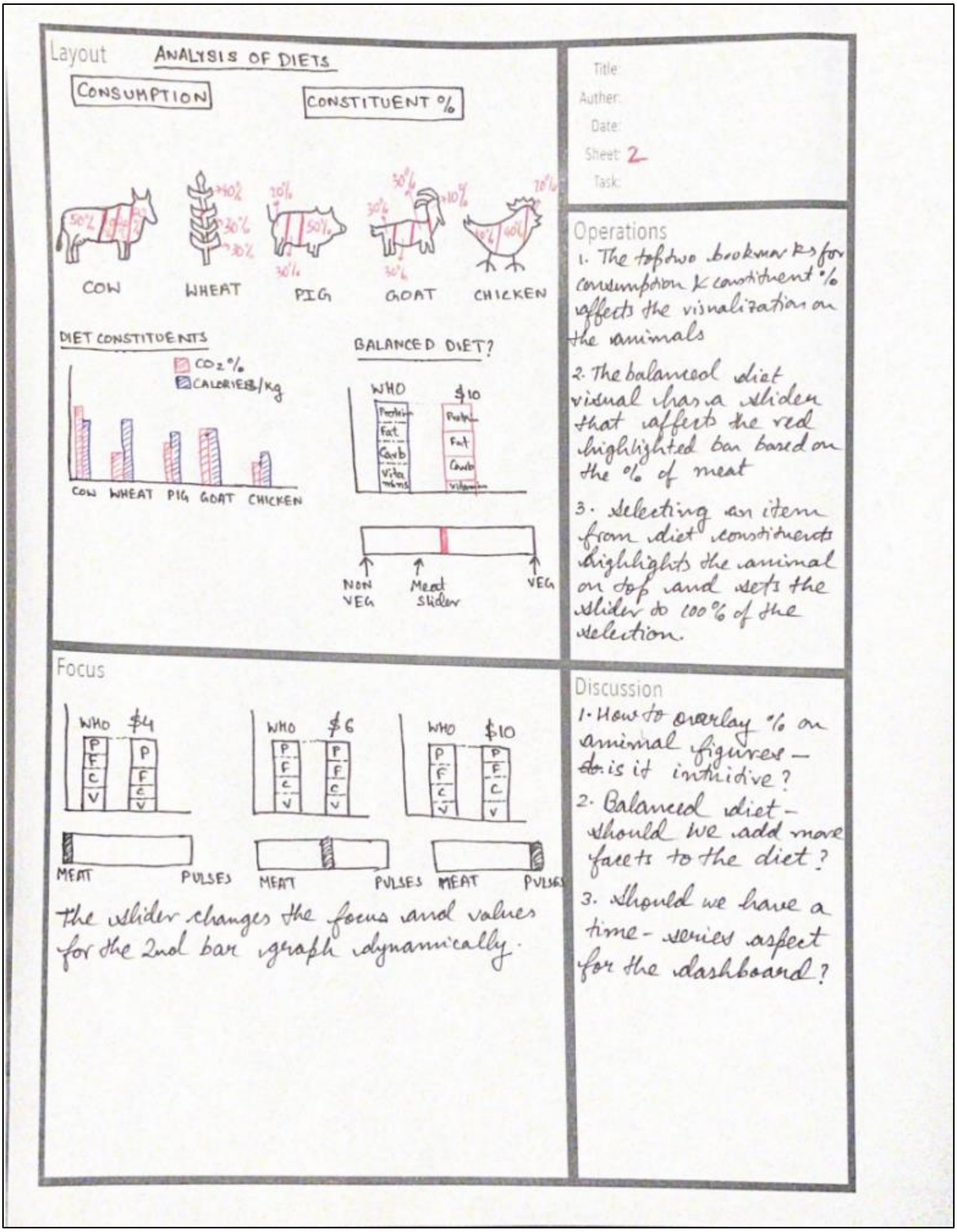

Figure 2: FDS - Sheet 2

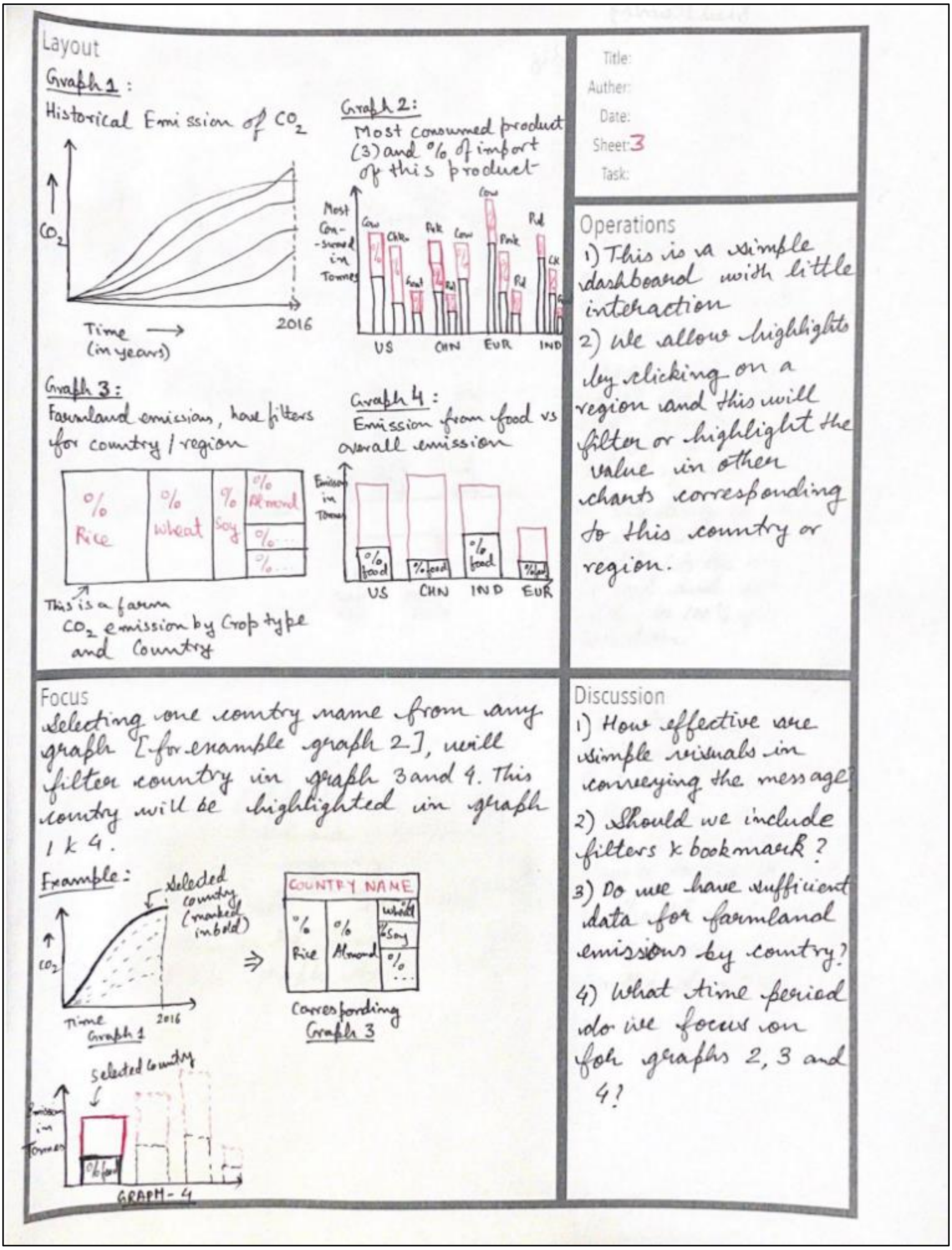

Figure 3: FDS - Sheet 3

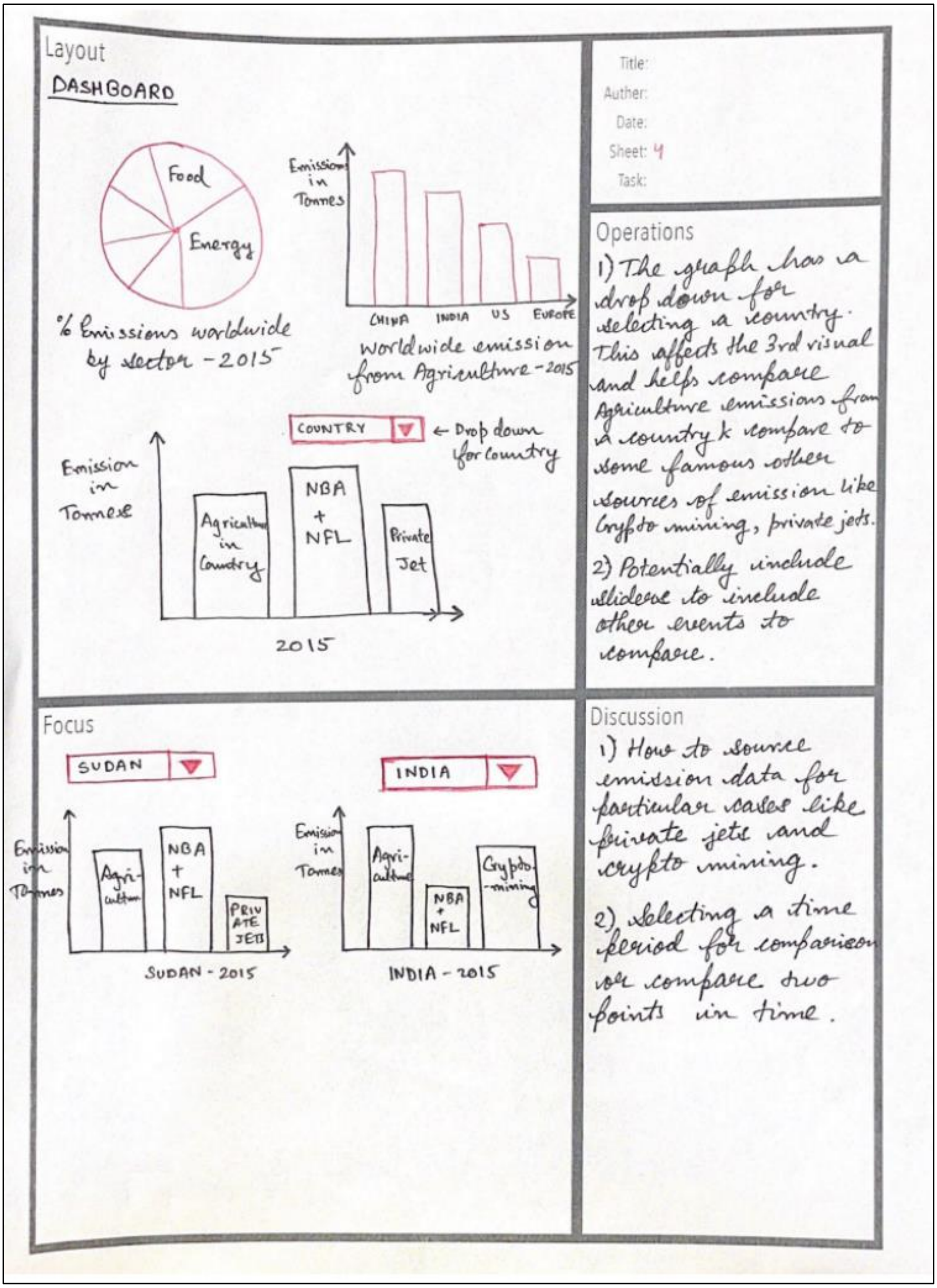

Figure 4: FDS - Sheet 4

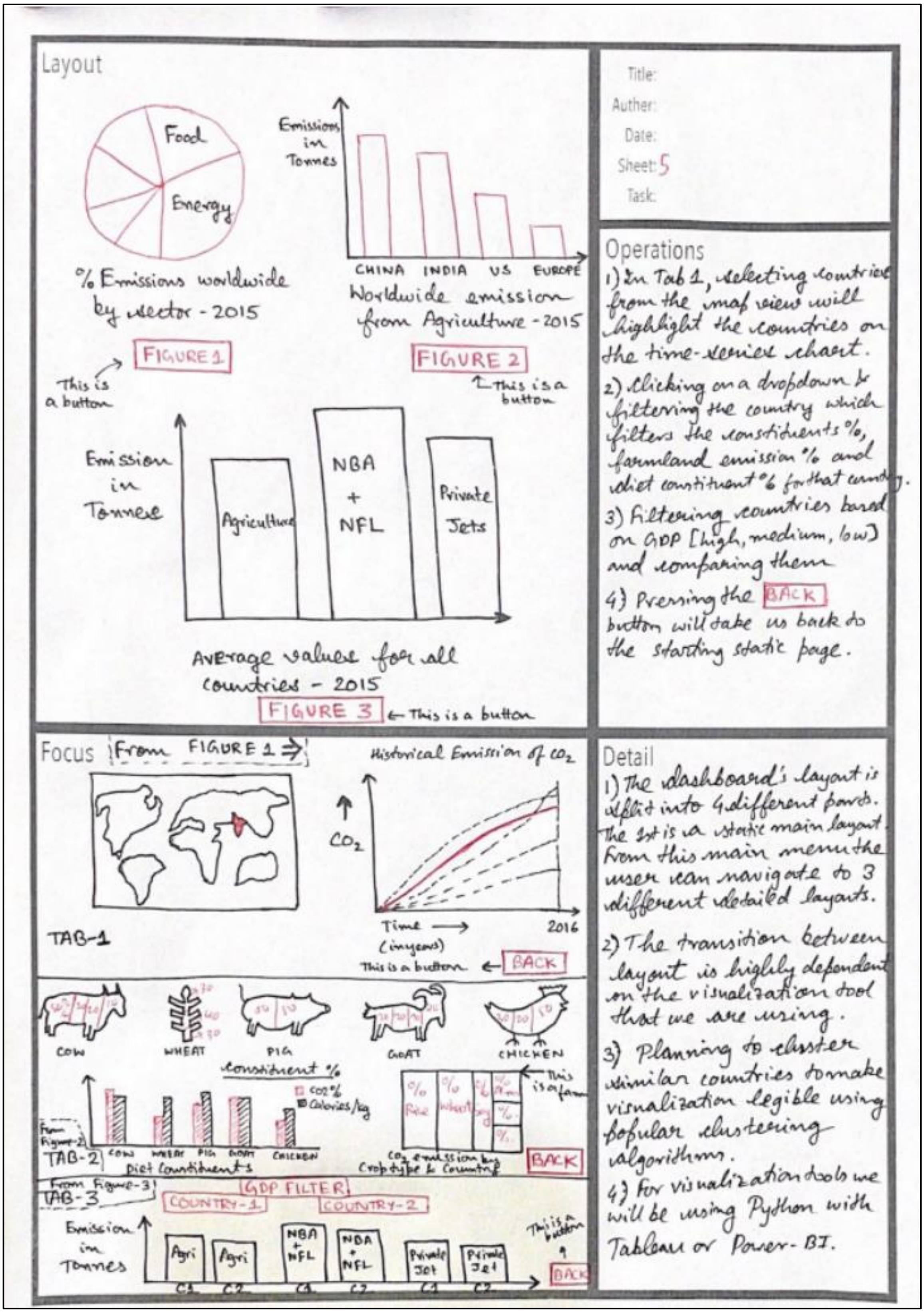

Figure 5: FDS - Sheet 5

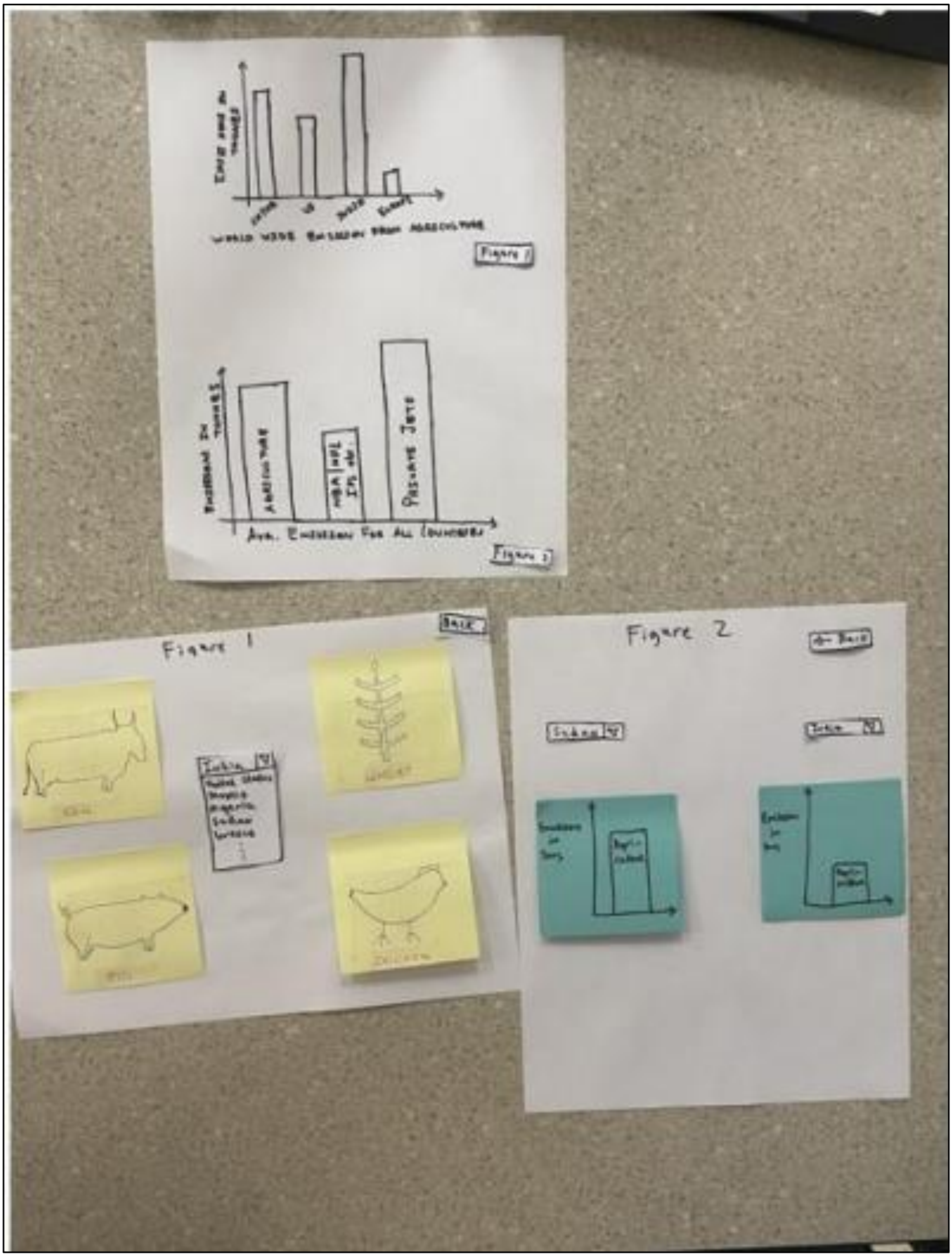

Figure 6: FDS - Low fidelity prototype - In class 1

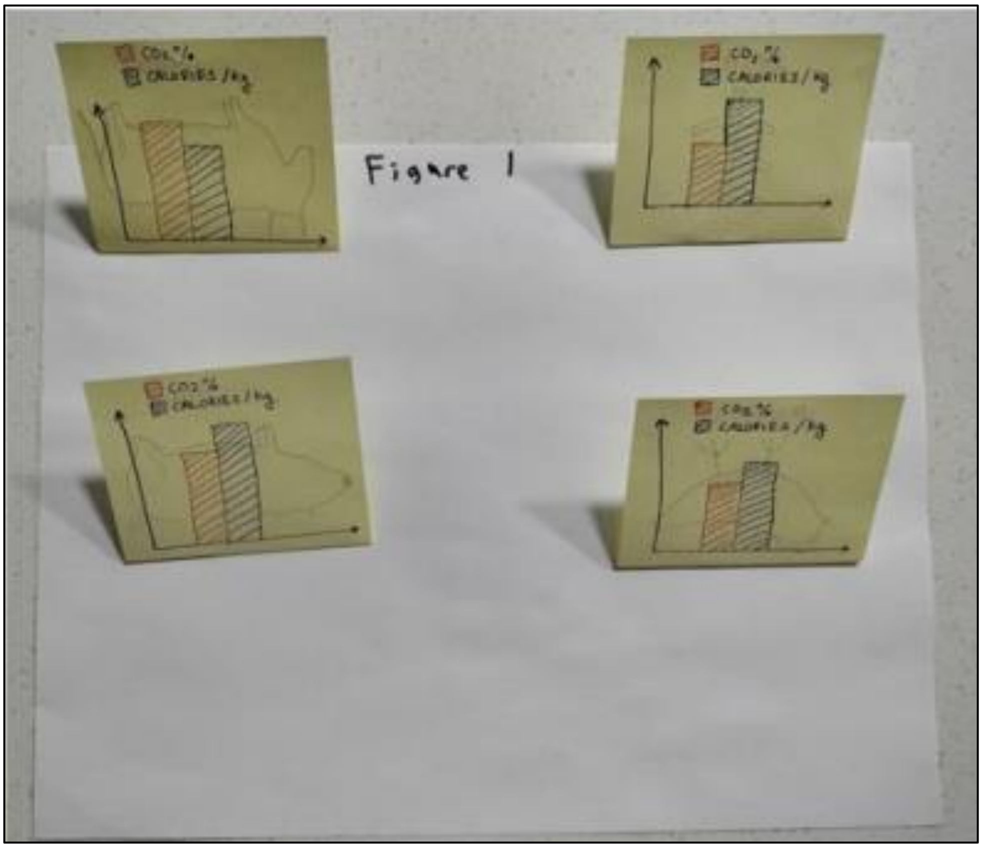

Figure 7: FDS - Low fidelity prototype - In class 2

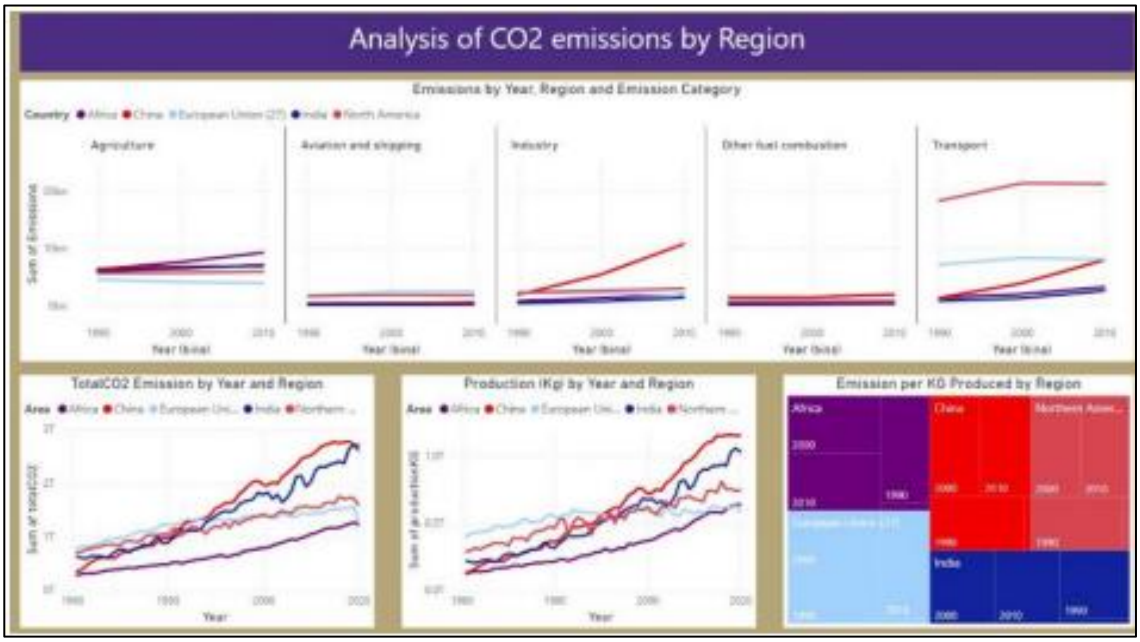

Figure 8: Power BI View: First chart of the static page from the paper prototype that describes emissions by country and category (Was a work in progress at that time)

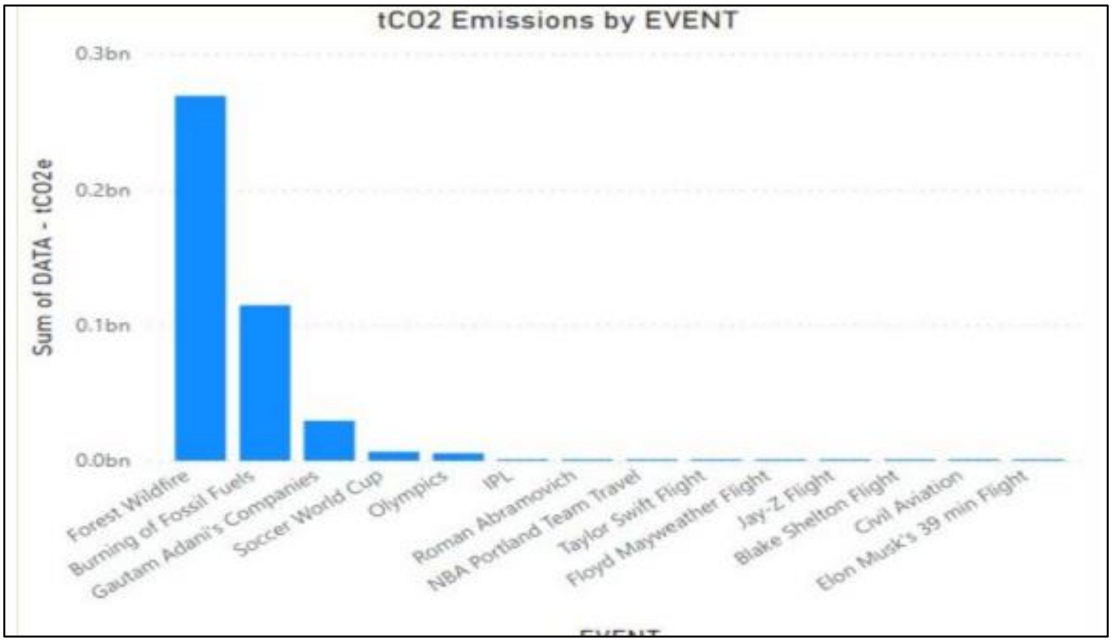

Figure 9: Power BI View: Second chart of the static page from the paper prototype that describes emissions from different sectors (Was a work in progress at that time)

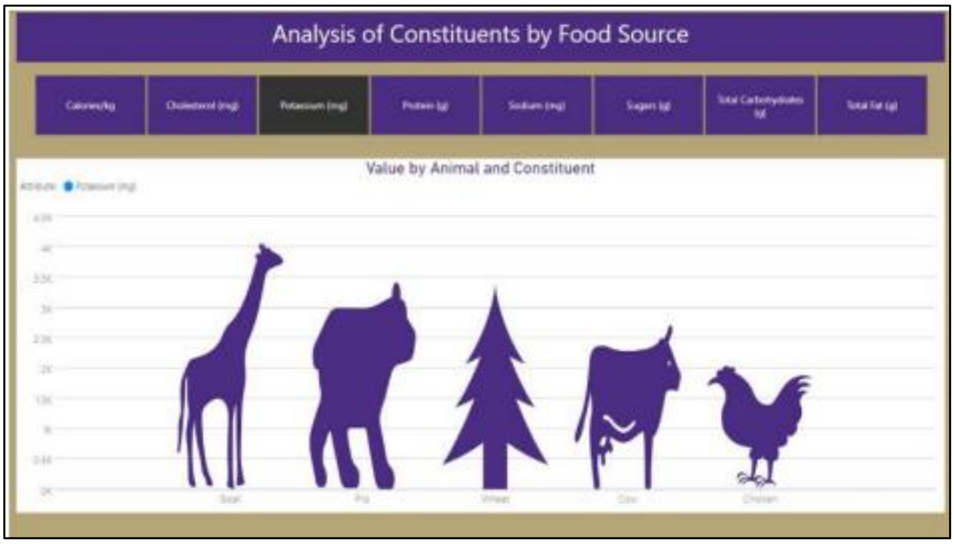

Figure 10: Power BI View: “Figure 1” from the paper prototype that describes emissions from different food types along with constituents (Was a work in progress at that time)

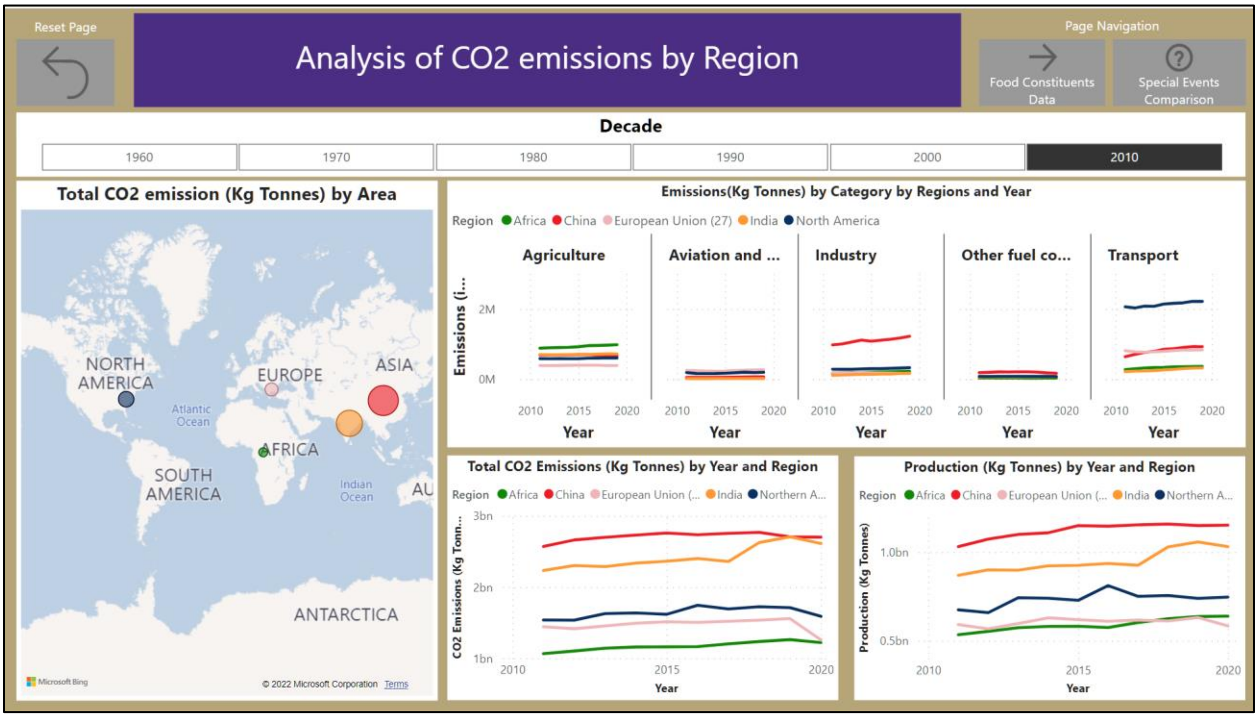

Figure 11 - CO2 Emissions page on final dashboard

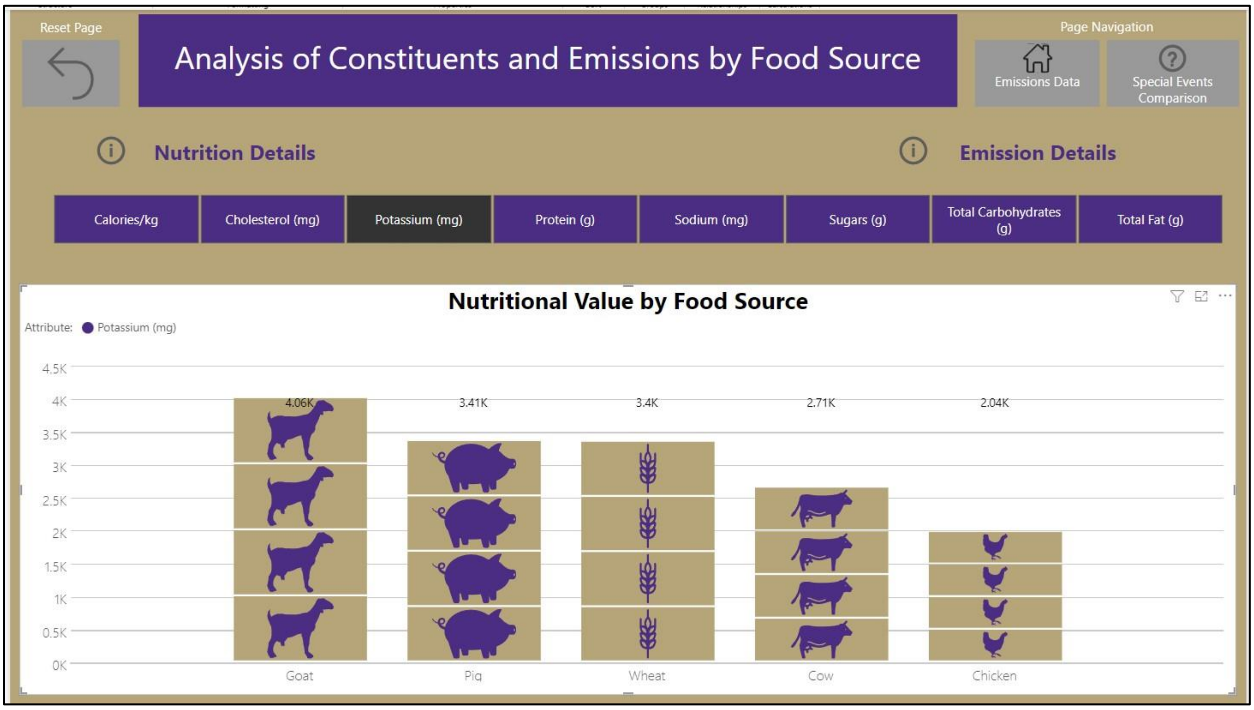

Figure 12 - Food constituents/emissions page on final dashboard

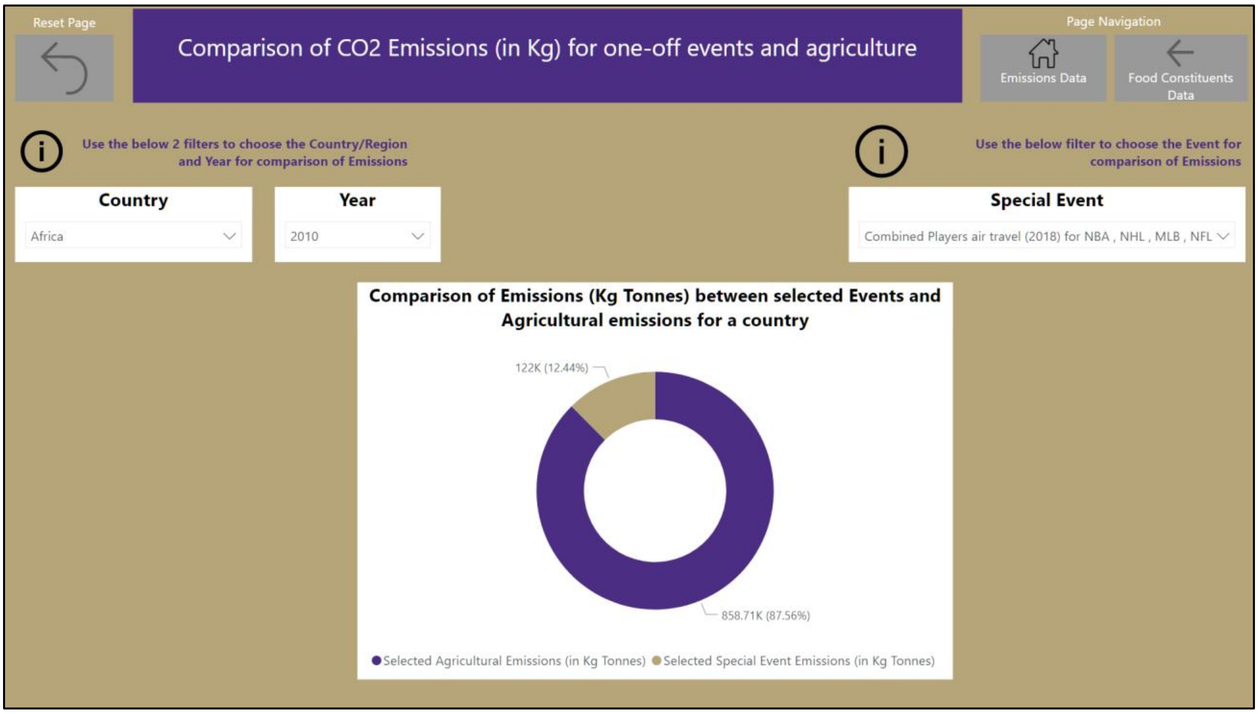

Figure 13 - Special events comparison to agriculture page on final dashboard

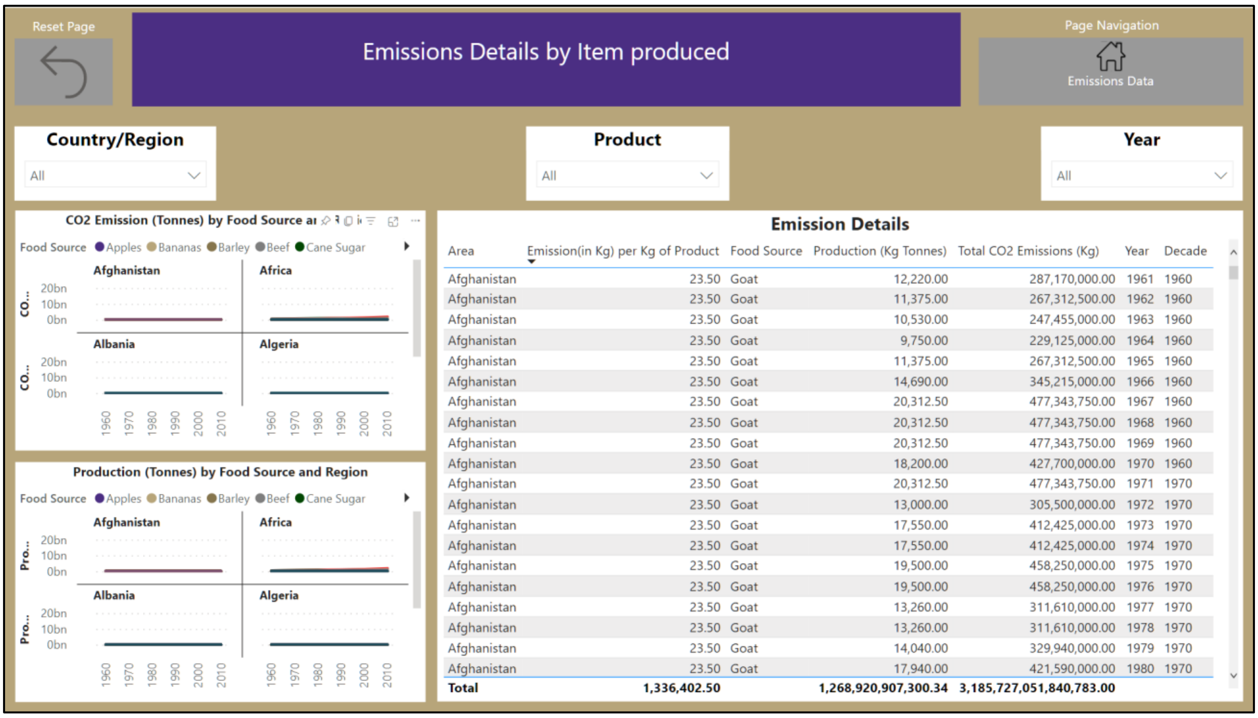

Figure 14 - Emission details by item produced page on final dashboard

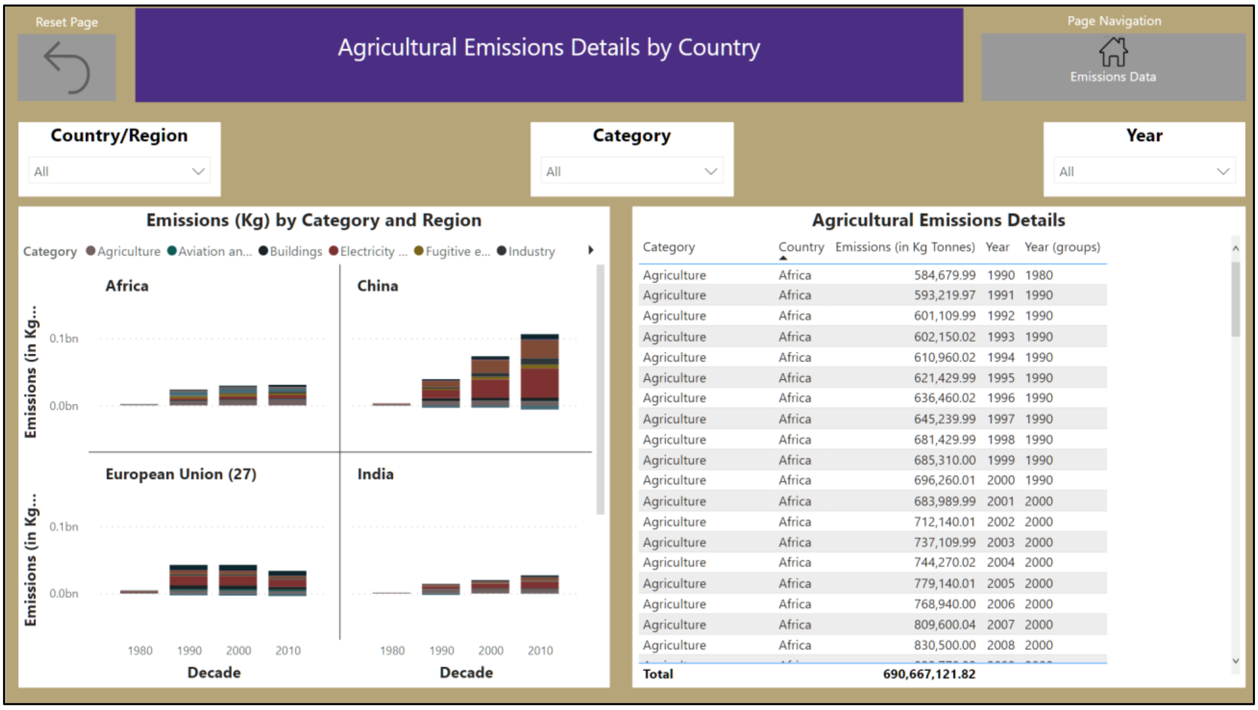

Figure 15 - Agricultural emission details by country