SHARADA SRINGESWARA, SREEDEVI GATTU, HARITHA SARANGA, AND U DINESH KUMAR

TEACHING NOTE

DATA STORYTELLING: WHAT ARE THE ALTERNATIVES TO CROP RESIDUE BURNING IN INDIA?

Case Summary

Crop residue burning (CRB) is a major factor contributing to air pollution in the northern part of India and Delhi's national capital region (NCR). In November 2021, the NCR consistently recorded an air quality index (AQI) value of more than 450, which is hazardous to human health. Most air quality monitoring stations in the NCR have classified the pollution level as severe. As a result of these bad weather conditions, Delhi residents have experienced multiple health issues such as red eyes, headaches, cough, itchy skin, and itchy throats. This is partly due to farmers in the states surrounding Delhi burning crop residues during the winter months. The Confederation of Indian Industries (CII) has been working with farmers in Punjab and Haryana to encourage them to adopt eco-friendly straw management practices as part of CII's Crop Residue Management (CRM) initiative. CRM's main objective was to eliminate the practice of burning crop residues in the open.

Chandrakant Pradhan, Manager for CRM, wondered how to demonstrate the initiative's results in the upcoming funding agency meeting. As part of his presentation, he wanted to raise the stakeholders' awareness of the alternative methods based on ground realities and the tools that farmers had been using primarily in different districts and villages of Punjab and Haryana. As he weaved through the traffic, several thoughts began to race through his mind. How should the data collected diligently by his team through farmer surveys be presented to potential funders? How best to examine and analyze the data? What valuable insights could the data provide to help raise more funds and support from various stakeholders? Is there enough evidence to show whether this initiative would reduce pollution over time and hence needed to be scaled up?

Teaching Objective

This case can be used in MBA, executive MBA, and Data Science programs to demonstrate data visualization and storytelling techniques. Using this case, students can learn why data visualization is essential for presenting valuable insights from data analysis. In addition, this case helps students understand how to communicate insights through easy-to-understand narratives and visuals. The instructor may use software such as Tableau and Power BI to analyze the case data and build the story.

_____

References and Background Material

The following textbooks and articles may be used as references:

1. Sharada Sringeswara, Purvi Tiwari and U Dinesh Kumar, *Data visualization and Storytelling*, Wiley, New Delhi. 2022.
2. Knaflic, C.N., *Storytelling with Data: A Data Visualization Guide for Business Professionals*. John Wiley & Sons, New York, 2015.

Relevant Industry Papers for Deeper Understanding

1. Schwabish, J.A., 2014. An Economist's Guide to Visualizing Data. *Journal of Economic Perspectives*, 28(1), pp.209-34
2. Segel, E. and Heer, J., 2010. Narrative Visualization: Telling Stories with Data. *IEEE Transactions on Visualization and Computer Graphics*, 16(6), pp.1139-1148.

Learning Objectives

This case illustrates how data visualization and storytelling techniques can be effectively used to provide insights and communicate those insights to decision-makers. The following are the other learning objectives:

1. Learn the important concepts of data visualization.
2. Learn how exploratory data visualization differs from explanatory data visualization.
3. Examine the effectiveness of visual encodings.
4. Learn how to create different types of charts to explore data.
5. Identify the value of business storytelling.

Teaching Plan

The case was written for teaching various concepts in data visualization, storytelling, and for illustrating dashboard development. To ensure that students are familiar with the data prior to the classroom discussion, we request that the instructor send them the accompanying dataset in advance. The instructor may use either Tableau or Power BI during the session. They must inform the students to familiarize themselves with Tableau (or Power BI). The data provided with the case has no missing information and thus can be used directly for exploratory data analysis. Ideally, the instructor should spend 30% time discussing theory and 70% time doing data analysis with Tableau or Power BI as described in the table below.

The TN has been designed for a 90-minute session. The teaching time can be divided as described below:

Discussion Point	Time (Minutes)
Discussion on main dilemma faced by the protagonist of the case to convince the funding agencies	5
Exploratory Data Analysis (descriptive statistics and simple charts)	15
Advanced data visualization using Tableau/Power BI	15
Discussion on data storytelling and relevant storytelling framework for the given dataset	15
Storyboarding	15
Developing 76	10
Discussion on final narration to the funding agencies	15

Suggested Case Questions

Depending on the students' background and skill level, the instructor may select questions from the following list and use any data visualization tool, such as Tableau, Power BI, Dash Python, or R. The outputs presented in this teaching note are based on Tableau.

1. A crucial task for Pradhan is to convince the corporates to fund the CRM initiative. How can he use the data to convince the funding agencies?
2. Perform an exploratory data analysis (EDA) on the data provided with a case and list down all the insights gained.
3. What are the possible data exploration themes from the given data?
4. What are your takeaways from the 2019 and 2020 data?
5. How many different types of narrative data stories are available? Which of them are relevant to the case of CRM data?
6. Discuss what story types are most appropriate for CRM data and what insights can be derived from their application.
7. Based on the 2020 data, what are the key takeaways for funding agencies?
8. As Pradhan presents these insights to various funding organizations and other stakeholders, how can he engage his audience through data visualization?
9. Create a narrative data story for Pradhan to present the findings to his audience.

Recommended Solutions to the Questions

Q1. A crucial task for Pradhan is to convince the corporates to fund the CRM initiative. How can he use the data to convince the funding agencies?

The purpose of this question is to encourage students to think about the strategies that Pradhan should use to convince corporates to fund the CRM project. For Pradhan, the best way to communicate CRM's impact in the short timeframe typically allotted for a CSR funding presentation is to demonstrate its impact through hard data. However, this data-based evidence should not be too complicated to follow and should use easy-to-comprehend visuals, created from the data collected through farmer surveys, that succinctly convey the message to the audience in a few slides. To provide a basic sequence of activities for a data visualization process, the instructor may begin with the following seven stages (Sharada et al., 2022).

Acquire: Acquire the data.
Parse: Create the required structure for the data.
Filter: Remove the noise and filter useful data.
Mine: Examine the hidden patterns in the data.
Represent: Perform simple visual representations to convey complex relationships.
Refine: Improve the representation.
Interact: Provide impactful information to the stakeholders.

Q2. Perform an exploratory data analysis (EDA) on the data provided with a case and list down all the insights gained.

Instructors are recommended to allow students to explore the dataset, for example, generating descriptive statistics on various variables. The distribution of land sizes among adopting farmers during 2019 and 2020 is a good place to begin. The land size distribution for the years 2019 and 2020 is shown in **TN-Figure 1** and **TN-Figure 2** respectively.
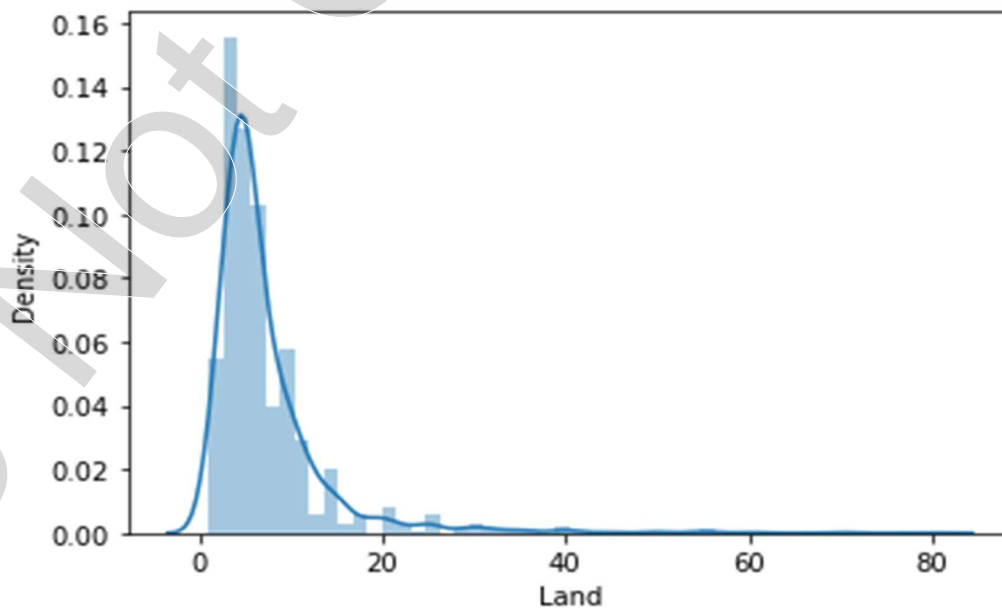
TN-Figure 1

Land Size Distribution (Year 2019)
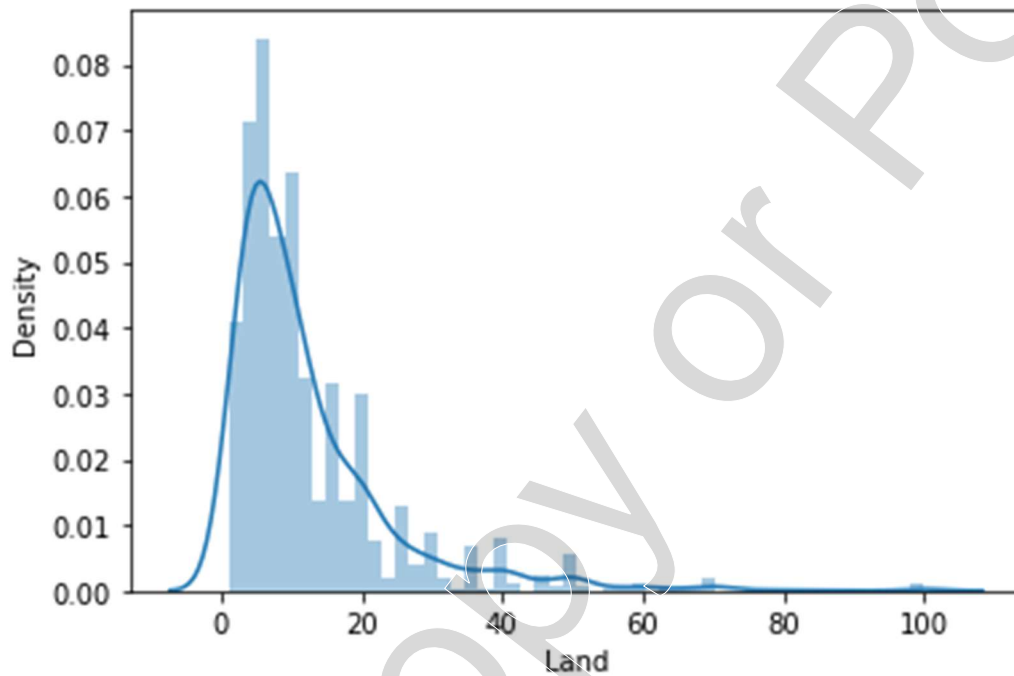
Source: Created by authors

TN-Figure 2

Land Size Distribution (Year 2020)

Source: Created by authors

We can infer from the figures **TN-Figures 1** and **2** that most of the farmers have land sizes of less than 20 acres and some have more than 80 acres. Note that the data sample collected in 2019 and 2020 is from different farmers and regions, therefore, one cannot conclude that the land size per farmer has reduced from 2019 to 2020 based on these two figures. **TN-Figures 3** and **4** show the tree map for number of farmers adopting CRM practices during 2019 and 2020 respectively. Since the CRM intervention has been extended to more villages/regions in 2020, the data sampled also reflects this fact based on **TN-Figures 3** and **4**. For example, the intervention was extended to Sangrur district and 500 sampled farmers in Sangrur have adopted CRM practice and farmers in other districts, such as Patiala, Ludhiana, Sirsa, Fatehabad, Barnala, and Rohtak continue the CRM practices into 2020.

TN-Figure 3
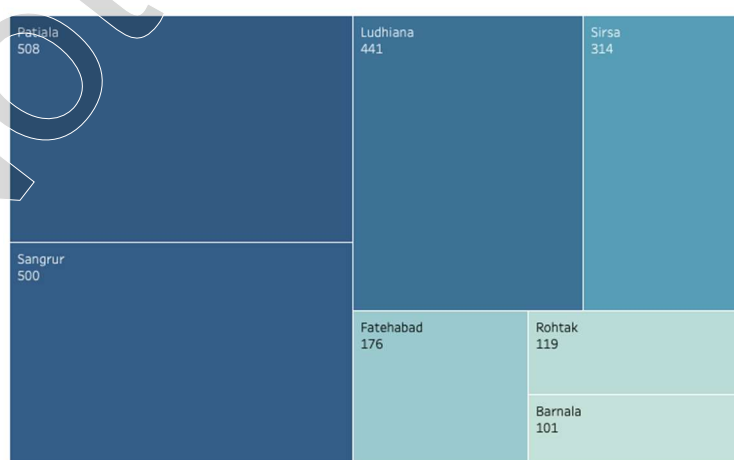
Farmers adopting CRM practices (2019)

Source: Created by authors

TN-Figure 4

Farmers adopting CRM practices (2020)
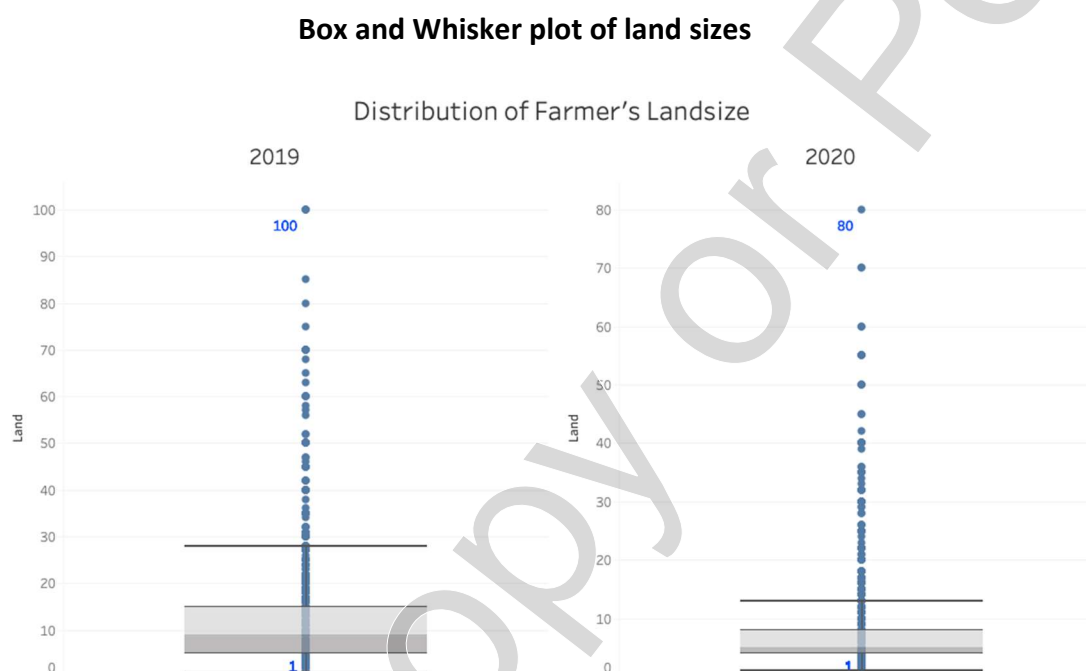

Source: Created by authors

To gain insight into the distribution of land size and outliers, students can use box plots (box and whisker plots) on the land size. **TN-Figure 5** shows that farmers with land sizes as small as 1 acre and as large as 100 acres have adopted the initiative. It shows several outliers as we can see many records beyond the whisker.
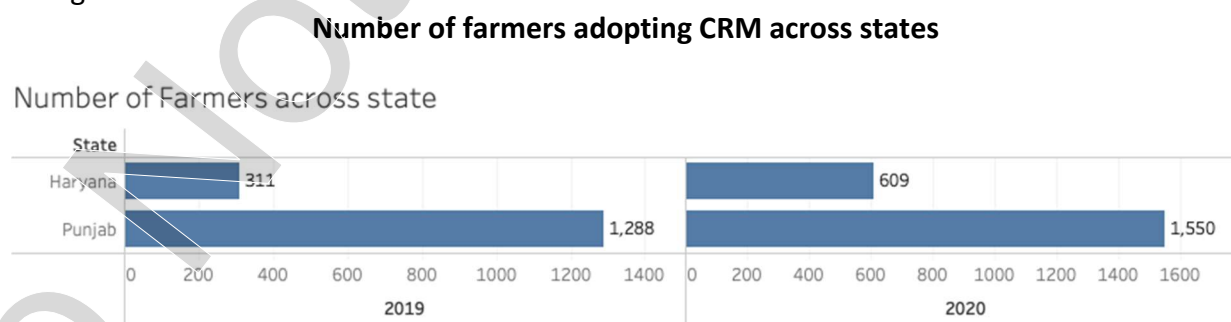
TN-Figure 5

Box and Whisker plot of land sizes

Source: Created by authors

Students may be asked to explore the number of farmers adopting CRM across states using the sample data. As the size of intervention increases, the initiative seems to have influenced more farmers in Punjab than in Haryana to adopt CRM practices during 2020 compared to that in 2019, as can be seen from **TN-Figure 6**.
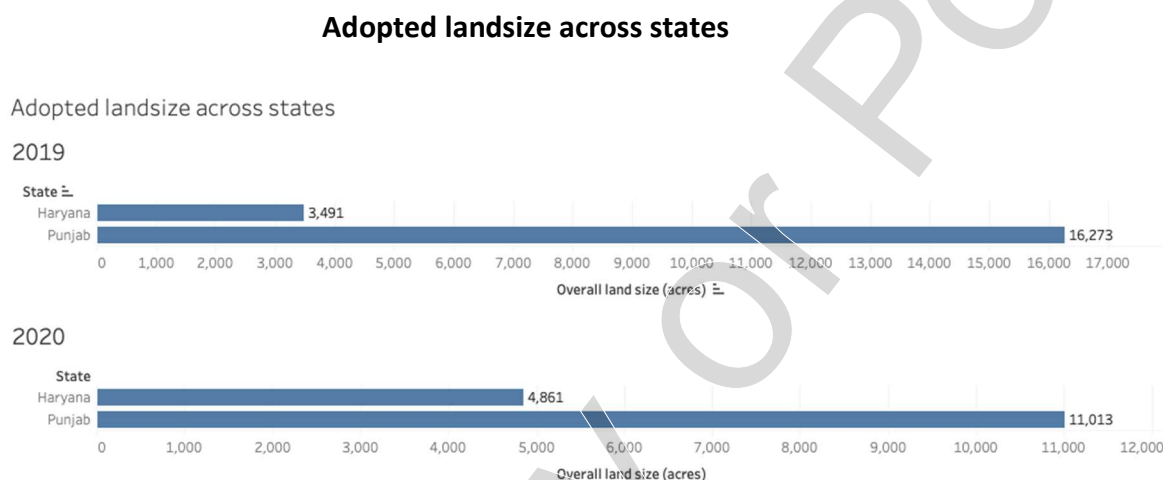
TN-Figure 6

Number of farmers adopting CRM across states

Source: Created by authors

Students can be asked to estimate the total land size adopted for CRM initiatives across states using the sample data. From the chart below (**TN-Figure 7**), we can infer that the total land size has increased significantly in Haryana in 2020, while it has reduced in Punjab. This could be because the regions where the intervention was extended to in 2020 had more small-scale farmers (more in Punjab than in Haryana). However, this also depends on how the sampling was carried out in each of the states. Therefore, the instructor should caution the students against drawing

such conclusions based on these figures alone but take a look at the overall population data for confirmatory evidence.

TN-Figure 7

Adopted landsize across states

Source: Created by authors

One can explore the percentage of farmers adopting the CRM practice. From the following chart (**TN-Figure 8**), we can infer, based on sample data, that while 50% of farmers were still doing crop residue burning in 2019, by 2020, majority (89.3%) of the farmers began to adopt sustainable practices.

TN-Figure 8

% of farmers adopting Sustainable practices
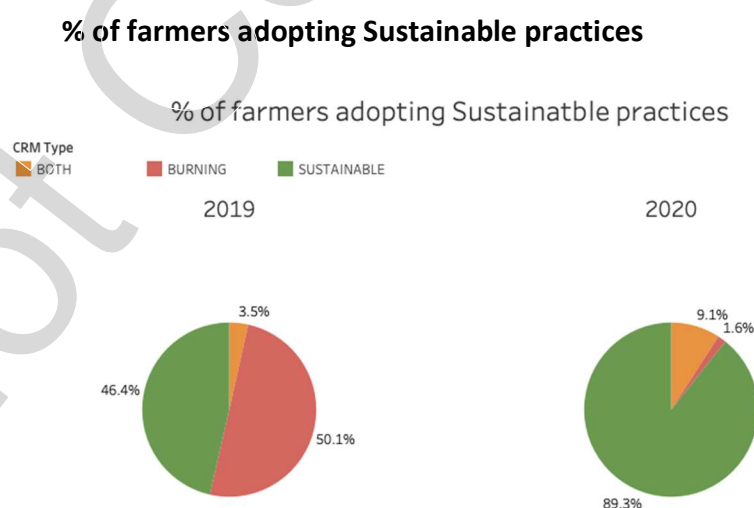
Source: Created by authors

Q3. What are the possible data exploration themes from the given data?

Following are a few data exploration themes to explore based on the data provided:

1. How many farmers have participated in the CRM initiative in 2019 and 2020?
2. How many farmers have adopted CRM in the states and corresponding districts?

3. A comparison of the size of the land intervened through CRM initiative across the state and its corresponding districts.
4. Compare how many farmers have adopted CRM practices across the states and their corresponding districts.
5. Has water and fertilizer consumption and weed and pest infestation increased or decreased between 2019 and 2020?
6. Compare and contrast farmers' concerns with CRM practices.
7. Analyze the tools used by farmers. Find out how farmers source these tools. What is the performance of the tools in terms of the time taken per acre to clear a field?

Q4. What are your takeaways from the 2019 and 2020 data?

Discuss the types of charts that can be used to work with the current data. Explore different types of charts and select those that provide stakeholders with important insights hidden in the data. Explore how the available data can be used to create seven basic story types. With only 2-year data, a change over time and intersection data story types that discuss trends over time cannot be used. However, a contrast data story type can be used for analysis based on the available data. To compare changes in the two years, we can use maps and butterfly charts. From 2019 to 2020, the number of framers reached through the CRM initiative increased. Maps and butterfly charts created using 2019 and 2020 data are shown in **TN-Figure 9**. Use principles of graphical integrity by maximizing data-ink ratio. In **TN-Figure 9**, the values of the bar charts are directly labeled by which we can avoid displaying axis tick marks leading to removal of unnecessary data-ink.

TN-Figure 9
Maps and butterfly chart for comparing the number of farmers who adopted CRM during 2019 and 2020
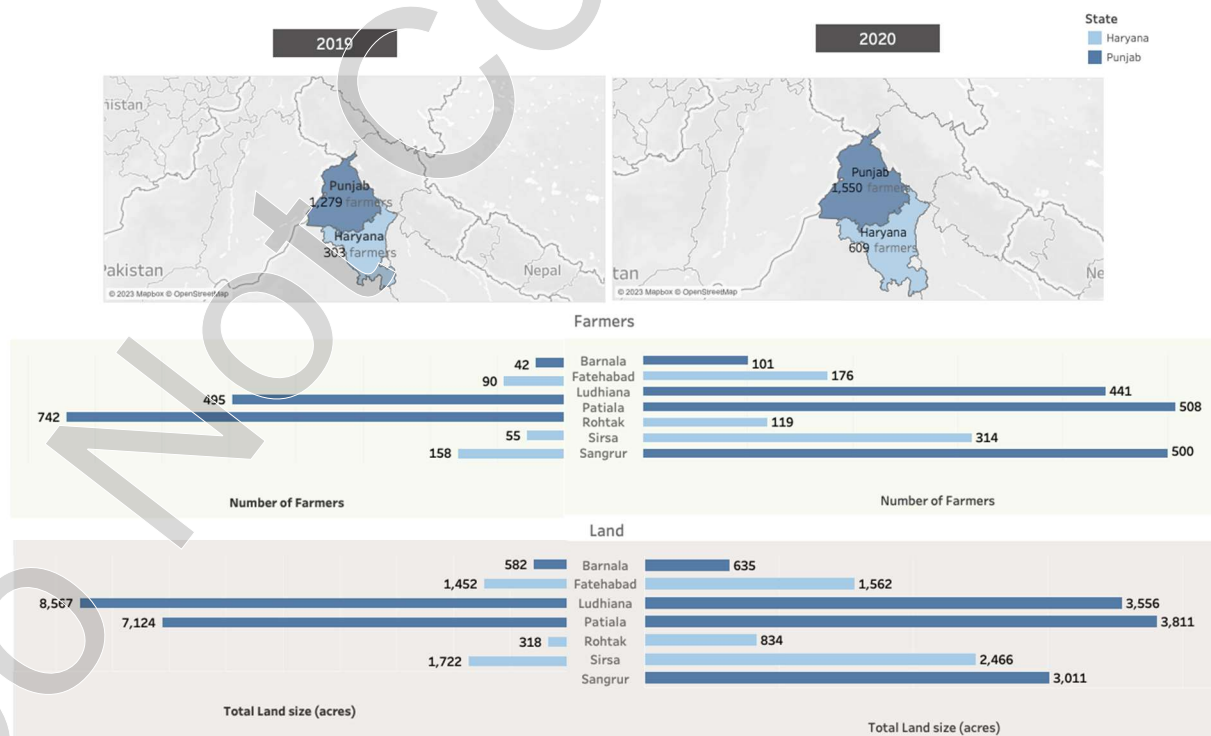
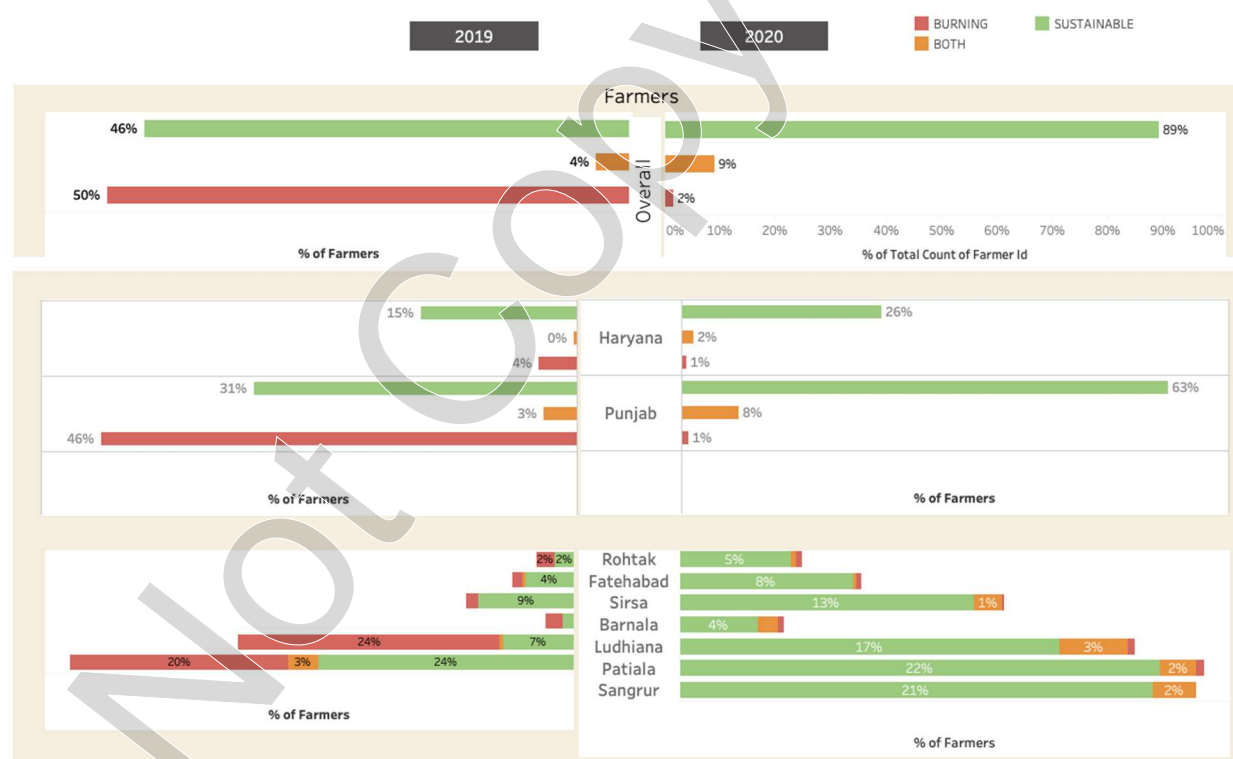**Source:** Created by authors

It can be seen from **TN-Figure 9** that Punjab's Sangrur district was added in 2020. A comparison of 2019 and 2020 shows that the sample size from Patiala contained the highest number of farmers in both years. However, in terms of land size, Ludhiana was the leading district in the sample in 2019, whereas Patiala was the leading district in 2020. The number of farmers and the intervened land size in the collected data sample reduced from 2019 to 2020 in both Patiala and Ludhiana. This could be because of the addition of new districts into the CRM initiative, which reduced the amount of data collected from the older districts.

The contrast data story type can be used along with the drill down story type, where we can plot the comparison in terms of the number of farmers adopting and by drilling down to geographical details of sustainable practices adopted. **TN-Figure 10** presents the visuals showing the overall, state-wise, and district-wise breakup of adoption rates across 2019 and 2020. Based on the visual in **TN-Figure 10**'s upper panel, 89% of the farmers have adopted sustainable practices in 2020 compared to 46% in 2019.

TN-Figure 10
Butterfly chart for comparison – percentage of farmers adopting CRM during 2019 and 2020
(Note: % values are rounded off to nearest percentage)

Source: Created by authors

One can create a contrast data story type along with drill down story type using the farmer's feedback data, which is Likert scale data providing the farmers' subjective judgment of the performance impact of adopting CRM on various aspects of farming (**TN-Figure 11**). **TN-Figure 11** shows that fertilizer consumption, weed infestation, and pest infestation reduce when farmers adopt sustainable practices. Besides, 69-71% of farmers felt that water consumption for the next crop decreases with the adoption of sustainable practices, which is good, given the water scarcity and reducing water tables in these two states.

TN-Figure 11

Comparison chart using farmers' feedback data during 2019 and 2020.
(Note that in this figure, we are only comparing the positive and negative feedbacks from the farmers during 2019 and 2020)
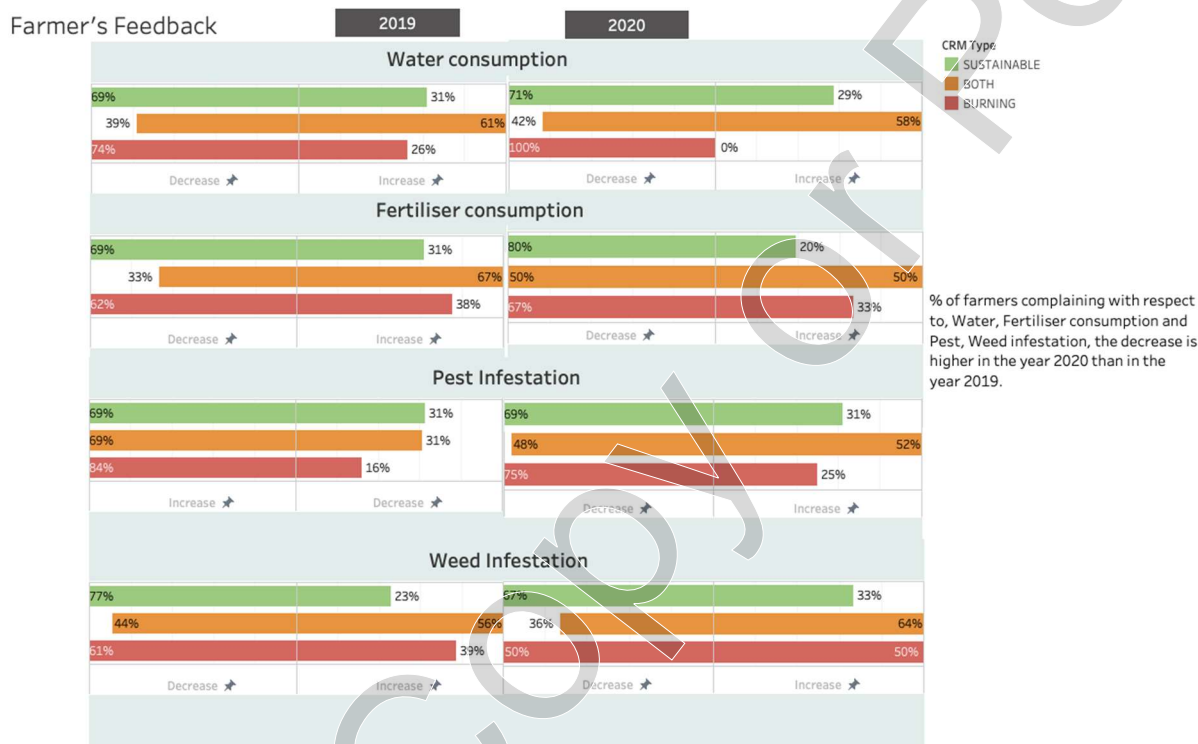
Source: Created by authors

Q4. How many different types of narrative data stories are available? Which of them are relevant to the case of CRM data?

Stories can be narrated using data from different perspectives. The instructor can discuss the following seven basic data story types constructed using Tableau.[1] These visual narratives can be used to tease out stories from the data and present facts linearly. These different narrative data stories can be used to trigger the stakeholder thought process. One or more of these types can be utilized to explain insights from the data to the audience.

1. Change over time: Narrative through visualization showing how a metric of choice has gradually changed.
2. Drill down: Lead your audience toward an insight from a big-picture view to a focused view.
3. Zoom out: Reveal insights by taking the audience from a hyper-local view to the big picture.
4. Contrasts: Outline the differences between data points.
5. Intersection: As a metric changes gradually, highlight interesting details about what happened when the data intersected.
6. Factors: Explain the factors that lead to a metric.
7. Outliers: Explore and present any interesting insights from the outlier data.

Discuss relevant story types for the current data. We can use drill down and factors story types to show how various CRM practices have been adopted. As the current data consists of only two years of data, changes over time and

---

[1] http://mediashift.org/2015/06/exploring-the-7-different-types-of-data-stories/, last accessed on May 3, 2023.

intersection story points would not be helpful in showcasing how the various metrics change over a long-time horizon.

Q5. Discuss what story types are the most appropriate for CRM data and what insights can be derived from their application.

Factor data stories can be applied to these data to highlight the detailed analysis and the resulting insights. Considering the data from 2020, as highlighted in **TN-Figure 12**, we can consider metrics such as the various CRM practices adopted and determine the factors such as tools, time utilized by tools, and tool sources to show how to conduct detailed analysis of the CRM data and tell the story about how the farmers in Punjab and Haryana adopted the CRM practices.

It is clear from **TN-Figure 12** that most farmers had adopted soil incorporation, mainly using rotavator, followed by super seeder and mulcher. However, the reason for using these tools like rotavator, super seeder, and mulcher was clearly not to save time taken for managing each acre of land. For example, a rotavator typically takes 1.7 hours per acre, whereas the Super Straw Management System (Super-SMS) takes only 1.1 hours. A mulcher, however, takes 1.9 hours per acre. The main reason for using these tools could be that many farmers had access to rotavators, followed by super seeders and mulchers, either through cooperative societies or through ownership. Moreover, Super-SMS was an expensive tool, hence owned by very few farmers and not easily available through cooperative societies. Therefore, although it was less time-consuming to prepare the land, only few farmers were able to use it.

TN-Figure 12

Tools used by farmers for different CRM practices and their sourcing details (2020)
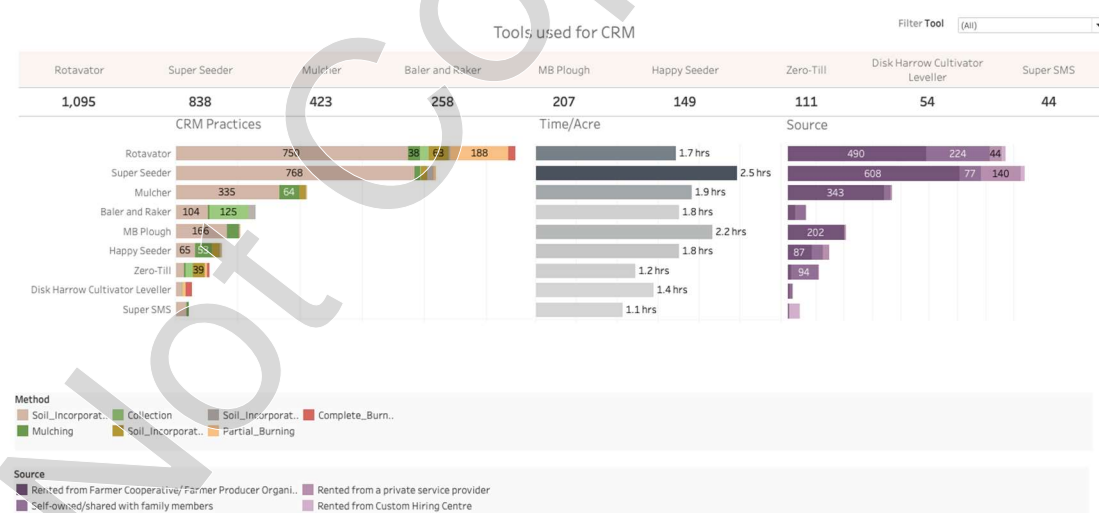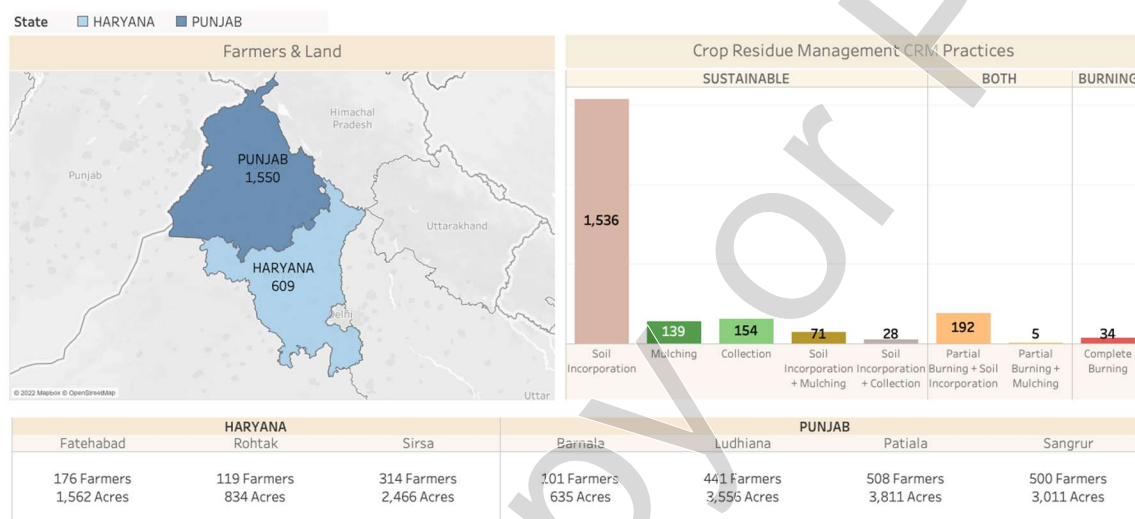

Source: Created by authors

Q6. Based on the 2020 data, what are the key takeaways for funding agencies?

As shown in **TN-Figure 13**, a high-level summary of the 2020 survey shows that CRM intervention was undertaken in four districts of Punjab and three districts of Haryana. The first map in **TN-Figure 13** shows the number of farmers who participated in the survey sample collection in each state. The adjacent bar chart shows the adoption of various sustainable practices, burning or both (some farmers resorted to partial burning of the straw and then used sustainable practices such as soil incorporation or mulching to manage the remaining straw). As seen in this figure, soil incorporation has been widely adopted by farmers, followed by collection and mulching. The table in **TN-Figure**

13 shows the number of farmers who participated in the sample survey from the seven districts of Punjab and Haryana and their landholding.

TN-Figure 13

Farmland size adopted for sustainable practices across the two states
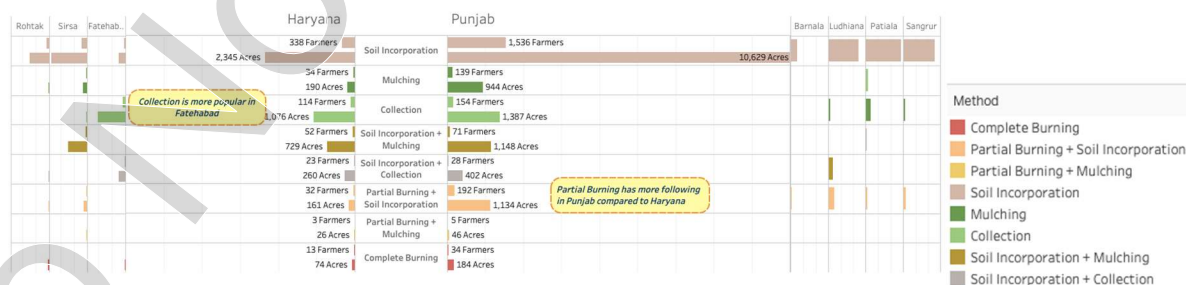

Source: Created by authors

A comparison of CRM practices adopted by farmers in Haryana and Punjab is shown in **TN-Figure 14**. In Haryana's Fatehabad district, collection (where straw gets collected in the form of bales and sold as animal feed etc.) appears to be the most popular practice, whereas farmers in Punjab are adopting soil incorporation. Partial burning along with soil incorporation is more closely followed in Punjab than in Haryana. Pradhan can use his own field knowledge to interpret these results to understand why certain practices are more popular in some geographies than in others.

TN-Figure 14

Comparison of methods adopted by farmers across different districts of Haryana and Punjab
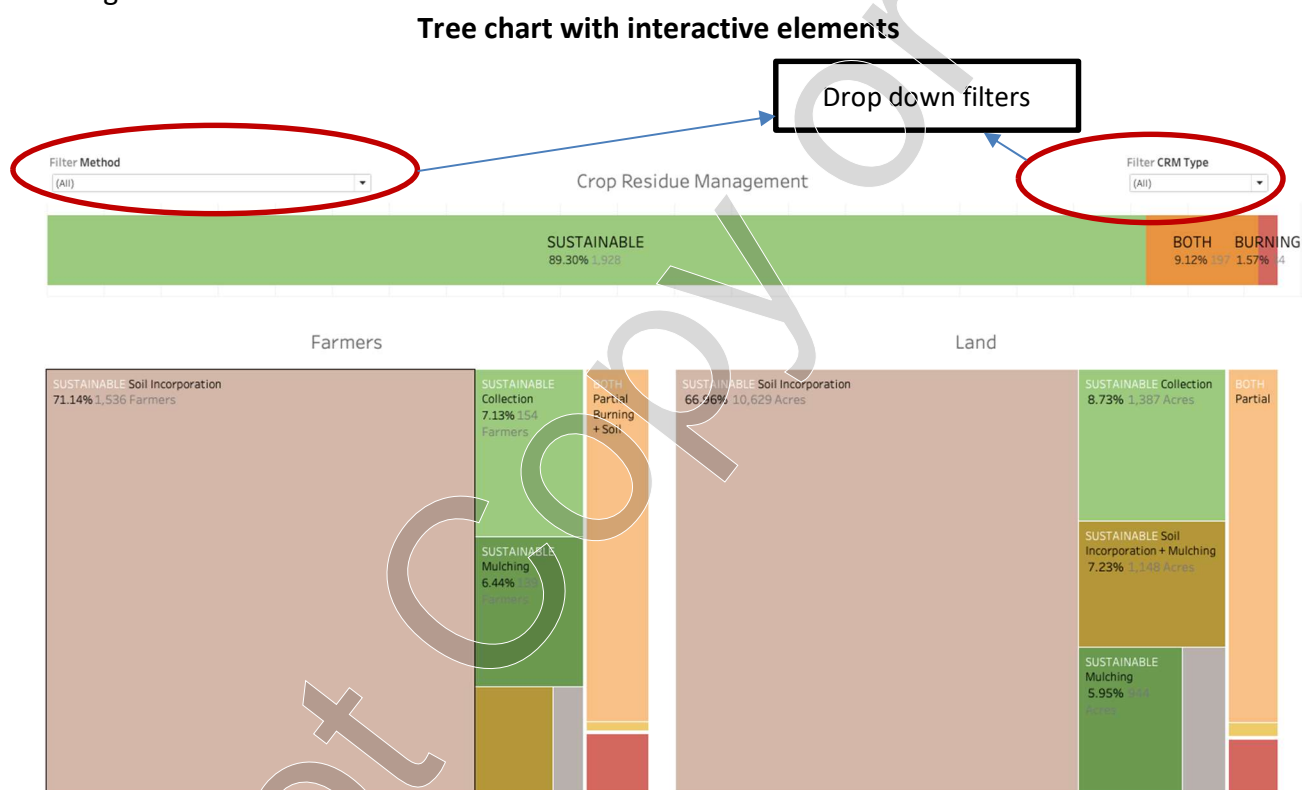

Source: Created by authors

Q7. As Pradhan presents these insights to various funding organizations and other stakeholders, how can he engage his audience through data visualization?

Pradhan can use interactive visualization to engage the audience in the insights presented. This allows the audience to select their area of interest for deeper understanding. Unlike static graphs, interactive data visualization allows users to engage with data. As shown in **TN-Figures 15** and **16**, one can include interactive elements in visualizations, such as dropdown filters and highlight the action. For example, by selecting a method, the chart is filtered and highlights the CRM methods adopted by farmers.
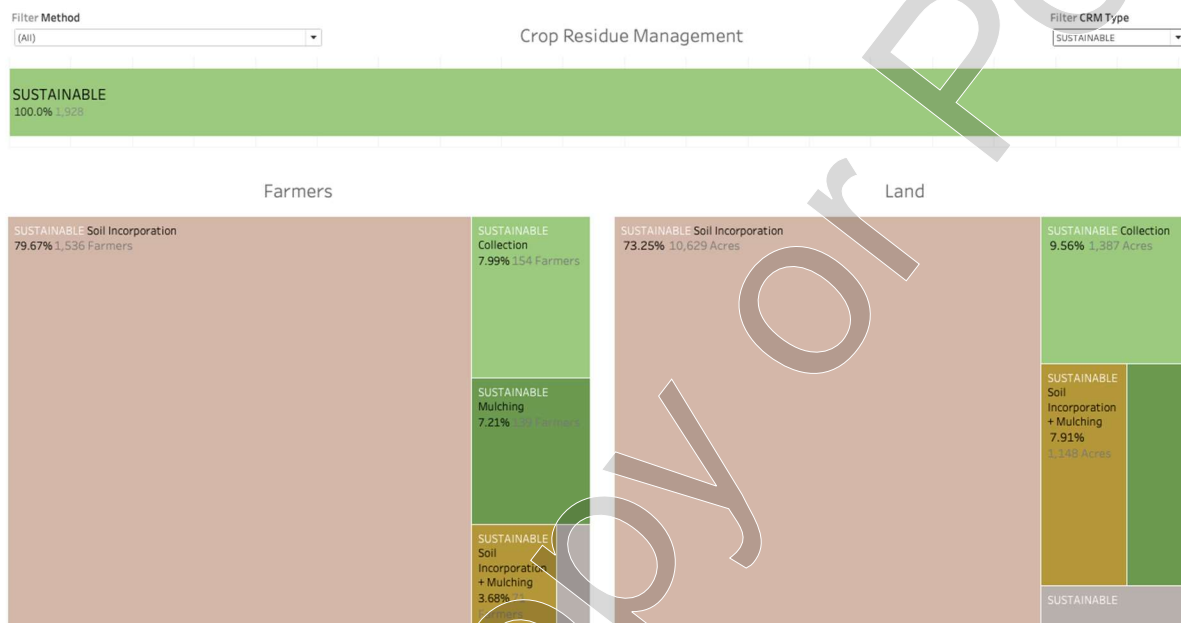
TN-Figure 15

Tree chart with interactive elements

Source: Created by authors

TN-Figure 16

CRM methods and farmland tree chart showing percentages of farmers and land under various sustainable practices

Source: Created by authors

Q8. Create a narrative data story for Pradhan to present findings to his audience.

To answer this question, the instructor can discuss various story structures available. Using a specific story structure, the instructor can build the narrative for these data. For instance, using **TN-Figure 17**, Pradhan can start with an introduction to the issue of CRB, its implications for people living not just in nearby villages but also in the Delhi NCR region, which is affected because of the pollution spillover due to wind direction. He can then present alternative methods of managing this crop residue through sustainable and eco-friendly mechanisms, such as soil incorporation, mulching, and collection. Using **TN-Figure 18**, he can then talk about the CRM initiative and how the CII Foundation has been creating awareness among farmers in various villages of Punjab and Haryana and encouraging them to adopt these sustainable practices. He can use **TN-Figure 19** to discuss the sustainable methods that are more popular among farmers and then compare data from 2019 and 2020 to discuss the steady growth in adoption. By showing the visuals created using farmer's feedback data (e.g., **TN-Figure 22**), Pradhan can demonstrate how these sustainable methods are reducing their costs and the need for irrigation water. Finally, by showing the tools and sources available for farmers and how their availability or lack of it influences the adoption rate of various sustainability practices (**TN-Figure 21**), he can pitch for more funding and list the areas of opportunities for stakeholder involvement.

Sample data story board in Tableau (Refer to **TN-Figures 17** to **22**) with details for each slide has been presented below. The instructor should engage students in a discussion based on the content in each slide.
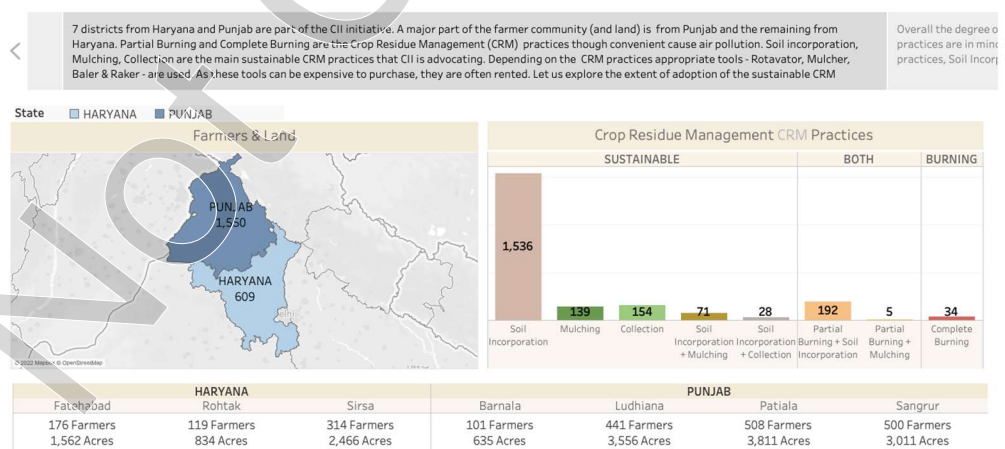
TN-Figure 17
Storyboard (Slide 1) – Introduction to problem statement (crop residue burning and its impact on Delhi NCR)
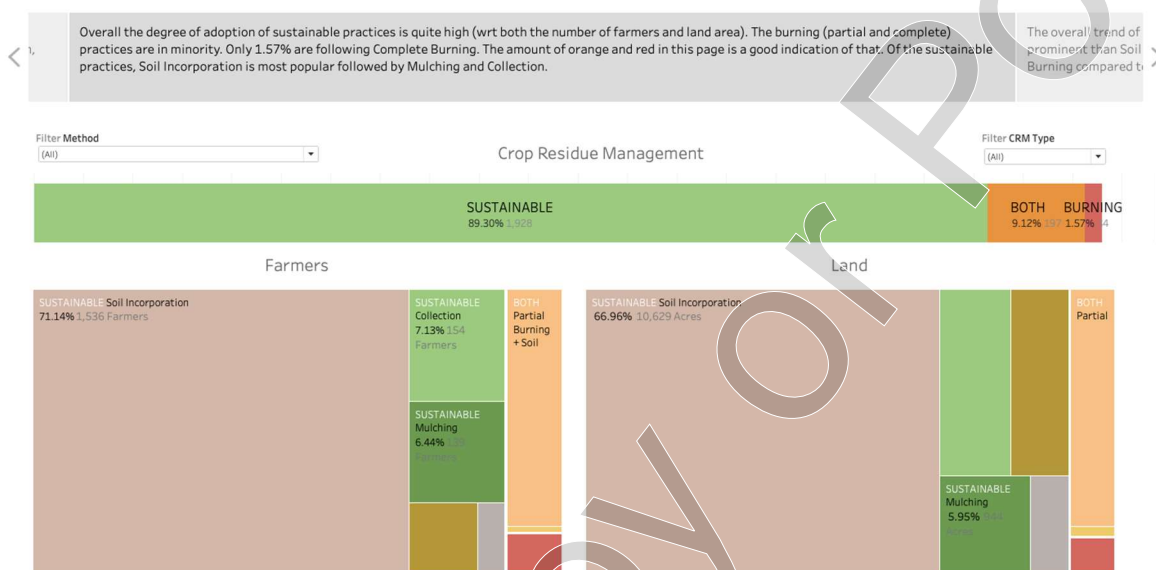

Source: Created by authors

TN-Figure 18
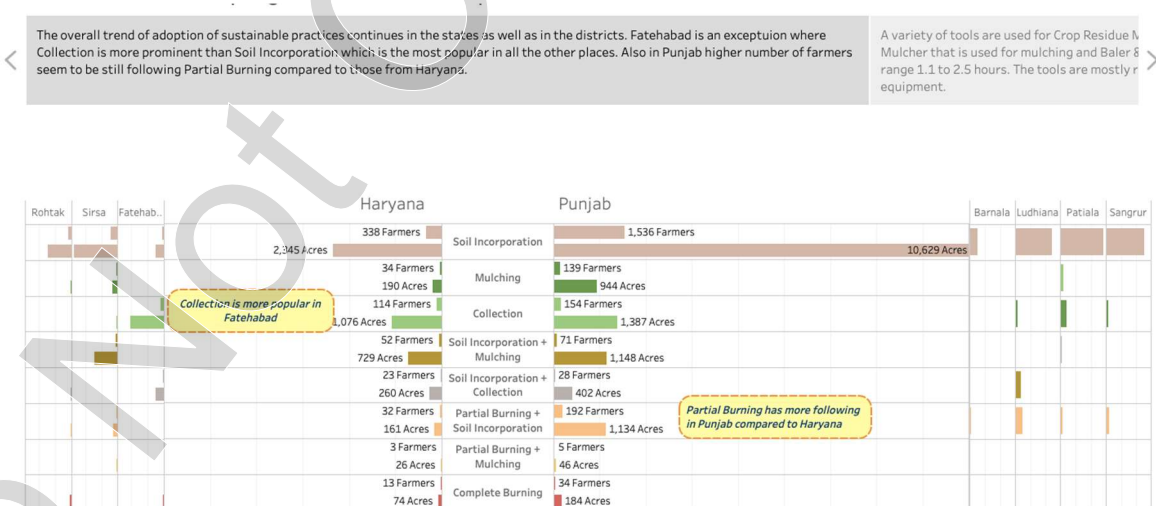## Storyboard (Slide 2) – Details on the Crop Residue Management initiative

Source: Created by authors

TN-Figure 19

**Storyboard (Slide 3) – Popular sustainable practices adopted by farmers**

Source: Created by authors

TN-Figure 20

Storyboard (Slide 4) – Comparison of popular sustainable practices adopted by farmers across different districts of Haryana and Punjab
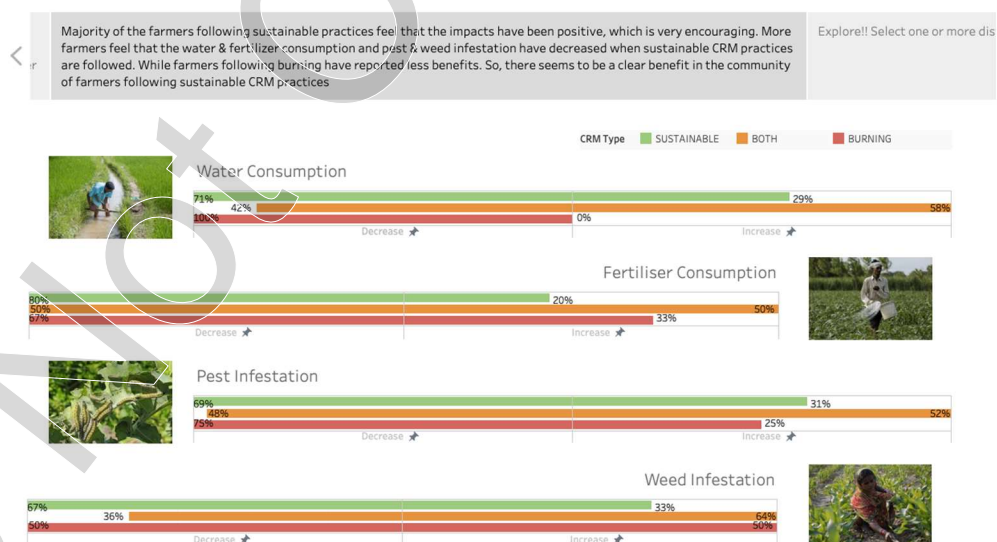

Source: Created by authors

TN-Figure 21
Storyboard (Slide 5) – Tools used by farmers (that depend not only on usage time but also on availability) for different CRM practices and their sourcing details
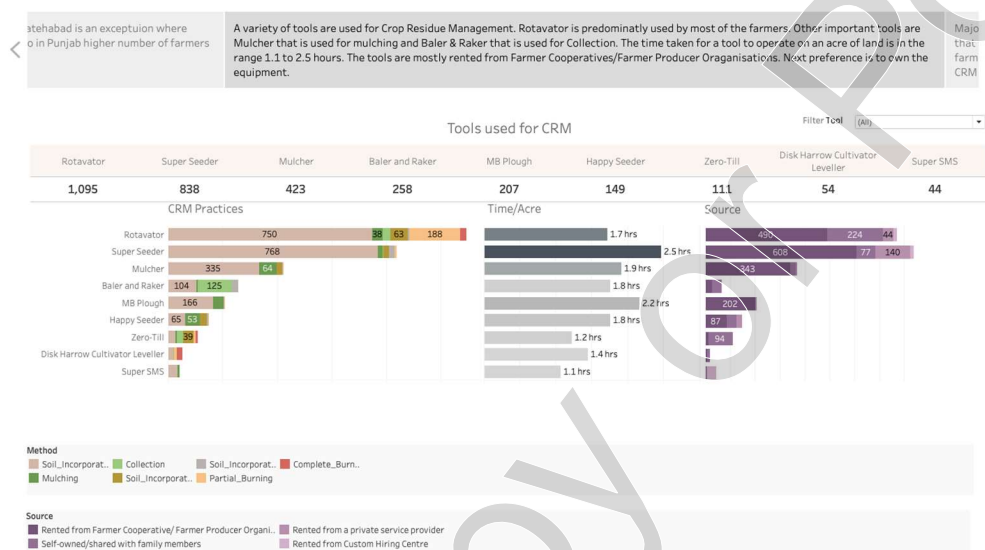
Source: Created by authors

TN-Figure 22

Storyboard (Slide 6) – Farmers' feedback data

Source: Created by authors