CSISA

Landscape Diagnostic Survey of Crop Production Practices in India

Guidelines for

Krishi Vigyan Kendra (KVK)

BILL & MELINDA GATES foundation

CIMMYT
International Maize and Wheat Improvement Center

Indian Agricultural Statistics Research Institute

IFPRI

IRRI

INTRODUCTION, OBJECTIVES & CONTENTS

Cereal Systems Initiative for South Asia (CSISA) in collaboration with Indian Council of Agriculture Research (ICAR) initiated landscape diagnostic survey (LDS) to bridge the existing data-gap around current production practices of major cereal crops. The survey has to be conducted by respective *Krishi Vigyan Kendra* (KVK) in its geographical domain and it is planned to be implemented by almost 200 KVKs across Indian states with technical and financial support from CSISA. Agriculture Technology Application Research Institute (ATARI) and State Agriculture University (SAU) hold the responsibility of coordinating and facilitating the activities required to initiate and undertake the survey.

LDS is designed in a way that data is collected from randomly selected farmers spread uniformly within a KVK domain/district. Survey questionnaire covers all production practices from land preparation to harvesting including detailed sections on fertilizer use, weed control and irrigation application. Data is captured through electronically enabled Open Data Kit (ODK) tool on mobile phone or tablet. Hence, all the data points would be geo-tagged. Use of ODK for data collection reduces the errors in data collection and automate the process of data compilation. All data collected by all the KVKs gets aggregated centrally at a server owned by Indian Agricultural Statistics Research Institute (IASRI). Datasets can be downloaded by designated persons of IASRI and CSISA for analysis and sharing with policy makers.

There are two objective of LDS:

- **Generate data-based EVIDENCES about current crop production practices**
- **Derive better INSIGHTS for informed decision making**

This document provides detailed guidelines to survey implementer about LDS. It covers following three sections:

1. **Survey Methodology**
2. **Open Data Kit (ODK)**
3. **Survey Questionnaire & Data Analytics**

1. Survey Methodology

Survey is the research method of gathering data by asking questions to people who are thought to have desired information. Survey methodology defines the sampling of individual units from a population and associated techniques.

CONTEXT

What – Information related to current crop production practices being applied by farmers at landscape level. *Do we know how many farmers apply herbicide for weed control in kharif 2018 in your district?*

Why – We should generate data based evidence for proper targeting and sharpening extension strategy. *Do we know which production variable(s) affect crop yield the most in your district?*

Where – We need to know the farming situation of the whole district and shouldn't be restricted to few villages. *Do we have mechanism to know the district scenario?*

When – We need to find-out a suitable timeframe for survey based on cropping season, farmers' availability and our planned schedule. The estimated labor requirement to finish the survey in one KVK is 30 person-days.

Who – There is provision for hiring suitable staff for implementing the survey. But, it is advisable to continuously monitor the survey by concerned KVK staff.

How – There has to be fairly strong sampling procedure so that results can be generalized. We need to adopt new survey tools to minimize data collection errors and reduce data accumulation time.

Data Collection

There are several methods of data collection that can be applied in field surveys. These methods fall into two broad categories:

Quantitative methods – Data are collected in a random sample of 'observation units', typically farm households. Random sampling is required to ensure that the sample is representative of a larger underlying population, e.g. farm households in the district. From each observation unit, the same set of information is elicited using a structured questionnaire. If the observation unit is the farm household, this means that each respondent farmer is asked exactly the same questions.

Appropriate statistical methods are used for data analysis. For example, we collect data on wheat yields from a random sample of farm households in district X. If our sample is sufficiently large, the average yield we find in our sample will be an adequate estimate of the average yield that farmers in district X attain overall.

Qualitative method – This approach is commonly used when the research topic is complex and requires deeper understanding. Focus group discussion (FGD) and key informant interviews (KII) are some of the very popular methods to collect qualitative data. Other than in the quantitative approach, we will prepare an interview guideline for data collection, rather than a questionnaire with fully formulated questions and potential response options. The guideline helps us ensure that we cover all relevant aspects during the interviews/discussions, but each such event will differ from the other. For instance, when we discuss a given topic with two groups of farmers separately, the two groups will almost certainly give different kinds of reactions and inputs, leading the discussion in somewhat different directions. It is the task of the researcher to react flexibly and follow up on such diverse inputs, rather than sticking to a list of pre-defined questions, as is done in a quantitative survey.

Both quantitative and qualitative methods have their own merits and limitations. None is 'better' than the other; rather, the two approaches complement each other, and which of the two is more appropriate depends on the research question to be addressed. If we want to get a 'representative' picture of what practices farmers are using, how these technologies are performing, and what farmers' perceptions are regarding the benefits of these practices and the constraints to their adoption, we need to use the quantitative approach. If we want to delve into great depth or get farmers' views on sensitive or highly complex issues, or we want to investigate particularly contrasting cases/settings, we should pursue the qualitative approach. When we follow a qualitative approach, we often select the villages where we conduct FGDs or KIIs according to certain criteria, e.g. villages with good market access versus very remote villages; this means that, in contrast to the quantitative approach, we often use purposive sampling rather than random sampling to select our research villages. Findings from qualitative research cannot be generalized to the population, but they can be used to highlight (contrasting) cases or conditions that require further investigation. Consequently, the sample size (e.g. number of selected villages) in qualitative research is usually very small.

Often, a 'mixed-methods' approach is recommended, combining the strengths of both quantitative and qualitative approaches. Following is an overview of the advantages and disadvantages of the two approaches.

	Quantitative	**Qualitative**
Advantages	• Results can be extrapolated to a larger, underlying population • Efficient and easy digital data collection using structured questionnaire • Relatively quick basic statistical data analysis	• Information can be obtained relatively quickly and inexpensively • More suitable for sensitive or complex issues than quantitative approach • Flexibility to follow up on unexpected aspects as they arise during data collection
Disadvantages	• Relatively costly and time-consuming, depending on sample size • Less suitable for sensitive or highly complex issues (e.g. power relations etc.)	• Results cannot be extrapolated to a larger population (e.g., each FGD represents a case study) • Data collection and analysis require greater skill than applying a structured questionnaire

SAMPLING

In the context of field surveys, sampling is a process in which a predetermined number of respondents are selected from a larger population. The methodology used for selecting respondents from a larger population depends on the type of analysis being performed. All sampling methods can broadly be categorized into two:

- Probability sampling
- Non-probability sampling

The difference lies between the above two is whether the sample selection is based on randomization or not. In case of randomization, every element gets equal chance/probability to be selected and to be part of survey. Before start, it is important to understand basic terminologies used in sampling.

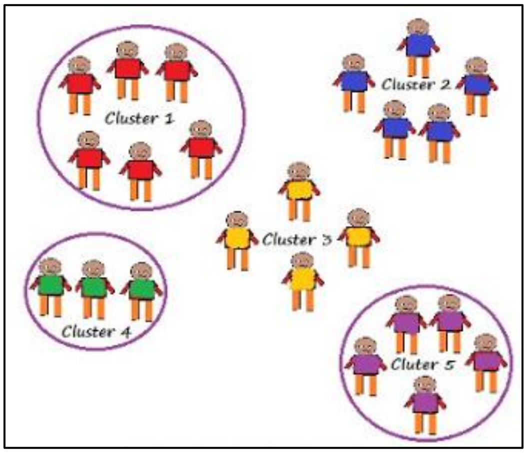
1. Element (or observation unit): This is the unit about which information is sought.
2. Sampling unit: This is the element or elements available for selection at a given stage in the sampling process.

3. Sampling frame: This is the list of sampling units available for selection.
4. Population (or universe): This is the aggregate of all the elements defined prior to selection of the sample.

The ongoing production practice survey (PPS) of cereal crops used single stage cluster sampling, a type of probability/random sampling method.

Probability sampling → Cluster sampling → Single stage cluster sampling

In cluster sampling, entire population is divided into clusters and then the clusters are randomly selected. To apply single stage cluster sampling, samples are drawn randomly from the selected clusters. All the elements of the cluster are used for sampling.



Accordingly, PPS selected villages within a district considering village as one cluster then selected farm households within each cluster/village. So, in our perspective, the above terminologies refer to:

Population: all villages of a district then all households of the village

Sampling frame: rural villages with >30 & <5000 households then all households of the village

Sampling units: villages first then farm households

Elements: farm households

In broad terms, the sampling process comprises of the following five steps:

Step 1: Define the population
Step 2: Select a sampling procedure
Step 3: Construct the sampling frame
Step 4: Determine the sample size
Step 5: Select the sample

SAMPLE SIZE

The larger the sample size, the more precise our estimates will be, such as average yields or the percentage of farmers using a given technology. In other words, with a larger sample, we can be more confident that our results will be relatively close to what we would find in the population as a whole. Related to this, the larger the sample size, the more likely we are to detect statistically significant differences between groups (e.g., differences in wheat yields between farmers who sowed before November 15 and those who sowed thereafter). However, the gains in precision decrease quickly at the margin with increasing sample size.

We suggest that KVKs aim at a sample size of 210 randomly selected farm households in their district to assess farmers' current practices; for most purposes, this sample size achieves a good balance between data precision on the one hand and cost of data collection on the other. We further suggest that sample households be spread across 30 randomly selected villages to capture an adequate degree of across-village variation, e.g. in terms soil conditions, infrastructure, and market access (factors which may influence the outcomes that we're interested in).

VILLAGE SELECTION

Probability proportional to size method of random sampling was used to select villages. It refers to a sampling technique where the probability that a particular sampling unit will be chosen in the sample is proportional to some known variable such as number of households. It can also be called "unequal probability sampling" because you are actually increasing the odds that a subject will be chosen in the sample based on its size. It is used when the populations of sampling units vary in size. If the sampling units are selected with equal probability, then the likelihood of a sampling unit with a large population being selected for the survey is actually less than the likelihood of elements from a sampling unit with a small population. This reduces standard error and bias by increasing the likelihood that a sampling unit from a larger population will be chosen over a sampling unit from a smaller population. To illustrate this method, consider the example of four villages of varying sizes given in the table below:

Village name	Number of households (HHs)	Cumulative number of HHs	HH ID range	Probability of selection
A	200	200	1 -200	200/1000 = 20%
B	300	500	201 – 500	300/1000 = 30%
C	100	600	501 – 600	100/1000 = 10%
D	400	1000	601 – 1000	400/1000 = 40%

To select villages using probability proportional to size method, we generate random numbers within the range 1 – max. HH ID. In the example above, we would type the formula =randbetween(1, 1000) into Excel; a village is selected if the random number falls within its HH ID range, thus making the probability of its selection proportionate to its size. For example, the random number 461 would fall into village B (HH ID range 201 – 500); hence, village B would be selected. We would continue generating random numbers (pressing the F9 key) until the desired number of villages is selected. If a random number falls within an already selected village, we simply continue pressing F9 until we get a random number that falls within a new village. This method for selection of villages was done based on the 2011 census data which contain the number of resident households in each village of a given district.

HOUSEHOLD SELECTION

Once the 30 villages are selected using probability proportional to size method, 7 households in each villages need to be selected through simple random sampling. In the simple random sample there is only one type of sampling unit, for instance all households residing in one village. Simple random sampling is a sampling technique where every item in the population has an equal chance of being selected in the sample.

This means that we need a complete list of households in that one village. This is our sampling frame for household selection. PPS used voter list of respective village to construct sampling frame. These voter lists of villages were downloaded from election commission websites of the respective states. These list are generally available in pdf version. Unique house numbers were treated as single household. Using 'R' software, these pdf type voter lists were processed in batch to generate random house numbers. The output was available as single excel file with 30 worksheets (one sheet per village) having desired random numbers for survey. This is an efficient way of doing household level randomization.

The process can also be done alternatively using MS Excel. But, you need to enlist all unique house numbers of the village. Once, it is compiled, number the households consecutively from 1 to max, where max stands for the total number of households in the village. For example, if there are 150 households in the village, your numbers would run from 1 through 150. Open an MS Excel spreadsheet and select cell A1. Use the function 'randbetween' to create a random number that lies between a specified minimum and maximum. The minimum is usually '1', i.e. the first element in our sampling frame. The maximum depends on the number of elements in our list. In our example it is 150; we therefore type:

=randbetween (1,150) and press Enter

Assume you want to select 7 households randomly: select cell A1, click on the lower right corner of cell A1 and drag it down until you reach cell A7. You now have a list of 7 random numbers available, which all lie between 1 and 150. Now simply copy the random numbers and paste them in column B as values. Now, tick off all the households that have been selected according to the list of random numbers.

PLOT SELECTION FOR CROP-CUT

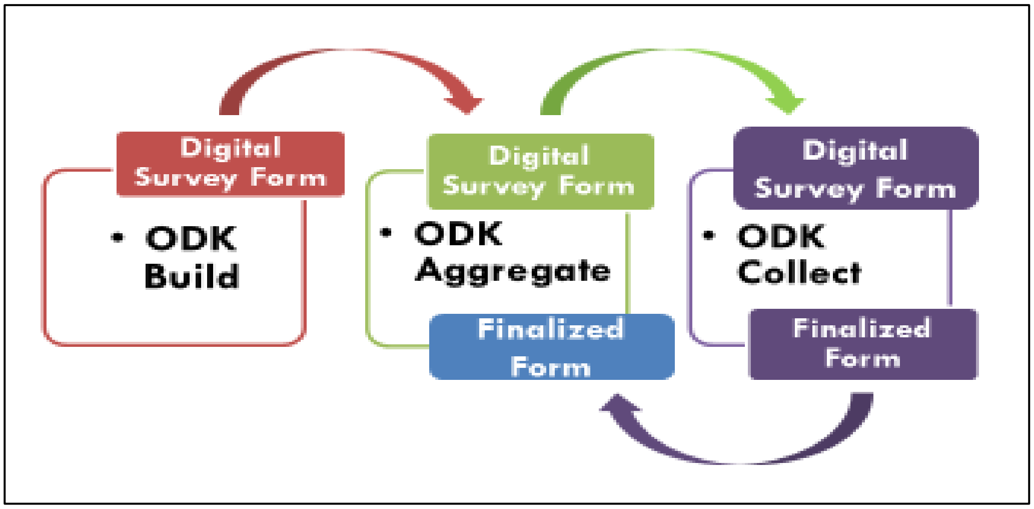
Follow these 10 steps:

I. Refer to the selected 7 households in this village
II. Select the farmer whom you meet first out of these selected 7
III. Ask him for his largest wheat/rice plot – consider this largest plot for crop-cut
IV. Take farmer's consent for crop-cut
V. Crop-cut has to be taken from 2 spots in the selected largest plot
VI. Size of each of these 2 spots (quadrants) are 2m X 2m
VII. Get on the corner of the plot, move diagonally for almost 5 meters and select your first spot here for taking samples
VIII. Similarly, repeat the procedure from the another corner of this plot and mark second spot
IX. Finish crop-cut from these two spots and record
   - Total above ground biomass
   - Grain weight and
   - moisture percent
X. Use Open Data Kit (ODK) Form – 'Crop Cut Form' to enter these readings along with other basic information asked in this form

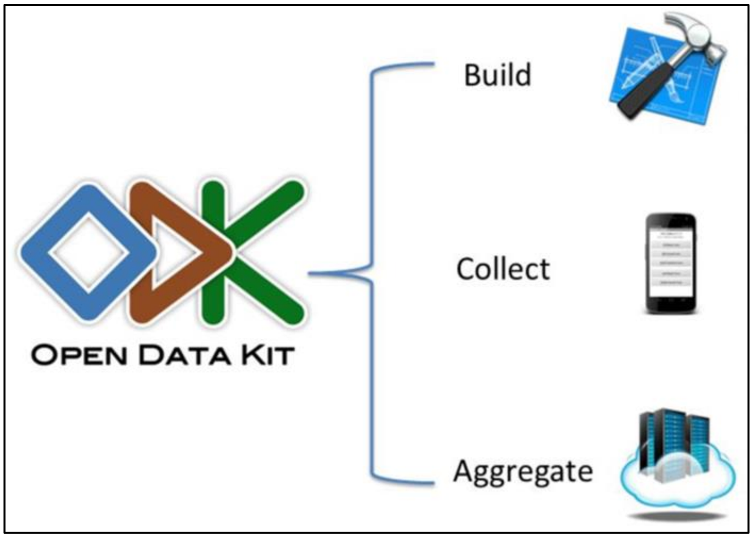
2. Open Data Kit (ODK)

The ODK is free and open-source software for collecting, managing, and using data in resource-constrained environments. It allows the collection of data offline and submit the data, when internet connectivity is available. It is a digital data collection tool that has many advantages over conventional paper based form. It reduces collection time, improves data quality and saves data generation cost. Developers and researchers at Department of Computer Science and Engineering, University of Washington had founded ODK. ODK began as a Google sponsored sabbatical project in April of 2008. The first two deployments of the tool happened in Uganda and Brazil (https://docs.opendatakit.org/). ODK is an open-source tool meaning, the source code is available for free and is licensed to permit customization by users. These are generally developed as a public collaboration and made freely available. Compared to conventional paper based data collection, ODK provides great ease by automating data compilation. In large scale survey, data compilation itself require huge resources and task is very much error-prone. Whereas, ODK was found easy to use and easy to scale even in resource-constrained environments.

There are three major components (Build, Collect & Aggregate) that jointly form the data ecosystem in ODK.



ODK Build: This is used for designing a questionnaire for ODK. ODK Build is a form designer with a drag-and-drop user interface. Build is an HTML web application and works best for designing simple forms. Alternatively, XLSForm is a form standard created to help simplify the authoring of forms in Excel. XLSForms are simple to get started with but allow for the authoring of complex forms. Forms designed with Excel can be converted to XForms that can be used with ODK tools.



ODK Collect: It is an op Android app that is used in survey-based data gathering. It supports a wide range of question and answer types, and is designed to work well without network connectivity. ODK Collect renders forms into a sequence of input prompts. Users work through the prompts and can save the submission at any point. Finalized submissions can be sent to a server. Collect supports location, audio, images, video, barcodes, signatures, multiple-choice, free text, and numeric answers.

ODK Aggregate: It is a Java application that stores, analyzes, and presents XForm survey data collected using ODK Collect. It supports a wide range of data types, and is designed to work well in any hosting environment. With Aggregate, data collection teams can:

- Host blank XForms used by ODK Collect
- Store and manage XForm submission data
- Visualize collected data using maps and simple graphs
- Export and publish data in a variety of formats

Accordingly, the workflow for data collection through ODK system is as follows:

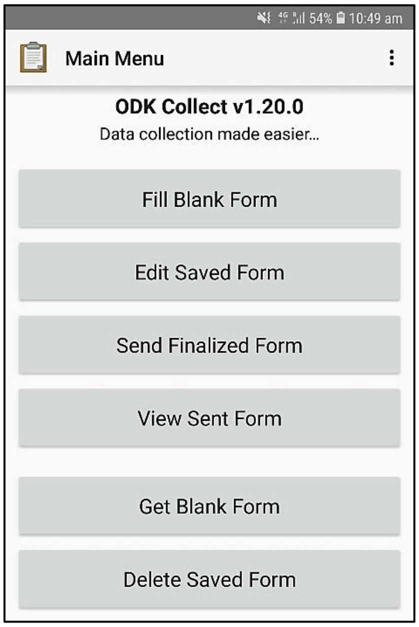
i. Design the form (questionnaire)
ii. Download a questionnaire for data collection
iii. Collect the data, **even if device is offline**

iv. Submit collected data to ODK Aggregate

v. Access aggregated data for use

The mobile app i.e. ODK Collect, to be used by enumerators can be downloaded from Google Play Store. The updated version (v1.20.0) of the app contains six buttons and their functions are self-explanatory. Once the mobile app gets linked with the hosting server, these buttons rightly perform following functions.

Get Blank Form – It is used to download desired survey forms in the data collection device from server. Internet connectivity is required.



Fill Blank Form – It is used to fill-in the information in the form while conducting the survey. It works offline.

Edit Saved Form – If enumerator wishes to add/change some information in the surveyed form before sending to the server, it can be saved in the device. This button can be used for doing edits.

Send Finalized Form – It is used to send single or multiple surveyed forms from collection device to the server. Internet connectivity is required.

View Sent Form – If you wish to see how many forms you have sent through a particular device, it generates the list of sent forms.

Delete Saved Form – It is used to delete blank form, if the current form is obsolete or an updated version of blank form has to be used. This button can also be used to delete filled-in forms if users don't want to submit it on server. It mostly happens in case of form testing.

BENEFITS OF ODK

There are several reasons for preferring ODK in the current production practice survey. As the diagnostic survey is quite large in terms of sample size, spread and length of questionnaire, manual data compilation would have been extremely difficult to handle. The respondents of this survey are farmers and they are mostly located in hinterlands. So, we wanted a tool that can work uninterrupted in such setting. Another factor of choosing ODK was the confidence of CSISA's technical team in handling the tool. ODK had been used by CSISA for almost five years for collecting monitoring data. Considering these factors, it was decided to go with ODK for the current landscape survey. In general, ODK provides another benefits over conventional paper based survey system. The key benefits are as follows:

Cost: There are many elements of cost. Electronic devices of course cost more than paper. But, when we factor in the requirement of hiring, training and employing data entry staff for the paper processes, in addition to buying and setting up the data entry machines, it ends up being costlier.

Speed and Efficiency: This is the most obvious advantage of digital data collection over paper-based system. Digital data collection reduces both data collection time and also the time required to analyze and distribute results. One of the main issues with paper version is its in-field administration if changes arise. While digital forms can be updated and pushed to enumerators quickly and automatically.

Data quality: Digital data collection not only reduces the possibility of error at the point of in-field collection, but it can also automate data correction. Paper can be lost, destroyed, or mishandled in a number of ways, which can create problems later if the data needs to be re-accessed. Digital data, on the other hand, can be easily and inexpensively stored, copied, backed up.

Visibility and Tracking: Another important advantage of digital data collection is tracking. Paper process doesn't tell us anything about what's going on in real time, but with a digital platform, as soon as an enumerator completes and submits a form, the data is accessible to all stakeholders. We can check who has sent this, from where it has come and is there any discrepancy. Data managers can contact back the data collector in case of need.

FUNCTIONALITIES

ODK provides wide range of functionalities right at the time of questionnaire designing that improve data quality and restrict users to enter incorrect data. Some of these features are:

Skip patterns: A questions with skip patterns is very common in any form of survey. For example, we may only want to ask a respondent about irrigation frequency, if their response to a previous question on whether they have irrigation facility is "yes". These types of skip patterns can only be enforced on digital surveys, with a conditional question only appearing based on the response to a previous question. An example of a skip pattern question is as below:

Do you have irrigation facility? [ ] Yes / [ ] No

If Yes to question above, how many times you irrigated your crop [Number Entry_____]

For paper based questionnaires, proper recording of such skip pattern kinds of questions are entirely reliant on the enumerator skills, knowledge of the questionnaire and keenness, leaving plenty of room for error.

Entry limits: This kind of restriction is usually vital especially for numeric types of questions. For digital surveys, it is possible to restrict entries, by having minimum and maximum values. For example, when taking the second split of urea applied in days after seeding, it cannot be less than the value of days (10-30) entered for first split. We can restrict conditional entry to higher value of first split in days. Any entry below that is therefore rejected.

Type questions: Survey questions happen to be of different types. These can be numeric, alpha-numeric, and dates, among other types. ODK ensures that entries are limited to their type, so we don't have a text response for a numeric question. Form developer is also able to control date format through pop-up calendar, furnishing options as single select or multiple select, pre-populating basic information such as area details, etc.

Optional vs mandatory questions: In digital data collection, we have control over whether a question is mandatory or optional. In this case, enumerator doesn't miss responses for questions that are considered essential for the survey. For example, you cannot move forward with the interview unless you fill the response about variety type. This means that the data available for analysis is usually pretty clean and ready for analysis.

GEO-TAGGING

One of the best features of ODK-based survey is geo-referencing. Current available mobile hand-sets can capture geo-location even without having internet and mobile connectivity. It adds great credibility in data we collect through ODK. All the locations (largest plot of respondents) of production practice survey henceforth are geo-tagged. It further allows us to layer this data with other parameters such as, soil profile, weather condition, etc.

3. Survey Questionnaire & Data Analytics

LDS questionnaire covers detailed questions around cereal crop production practices. Some of the major areas of information to be collected are as follows:

Landholding characteristics – Total cultivated land, area under major cereal crop, area of the largest crop land, soil type, drainage class, GPS coordinate of the largest plot, perception of soil quality, etc.

Crop establishment – Method of land preparation, sowing/transplanting method, sowing/transplanting date, type of variety, name of variety, source of seed, etc.

Fertilizer use – Types of fertilizer applied, doses of fertilizers at different growth stages, source of fertilizer, source of information about fertilizer usage, fertilizer availability, etc.

Weed information – Top infesting weeds identified with weed poster, method of control, time of control, etc.

Irrigation application – Number of irrigation, irrigation availability, source of irrigation, irrigation decisions, etc.

Pest control – Degree of pest/disease infestation, method of control, chemical use, dose of chemical used, etc.

Production constraints – Occurrence of draught/flood, severity of stress, stage at which crop was affected, etc.

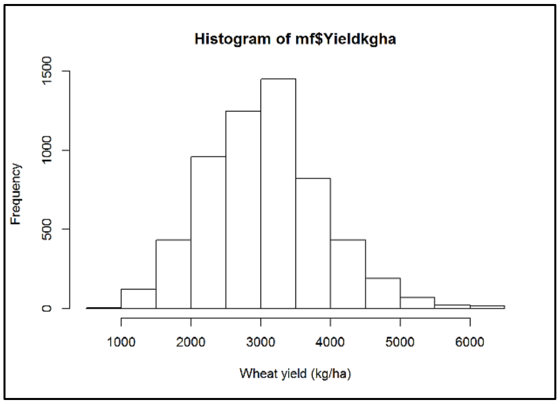
Harvesting – Harvesting method, harvesting time, threshing method, yield of largest plot, total production, production trend over last few years, etc.

Besides these, there are few more questions on household characteristics. This questionnaire is built on ODK Collect. To complete the survey of one farmer, it takes around 30 minutes in ideal condition.

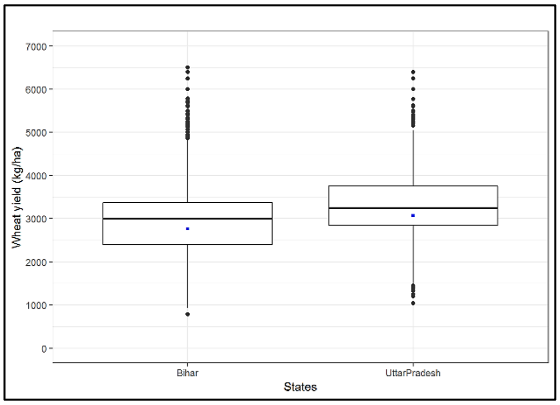
Data accumulated on the server can be downloaded any time for further cleaning (if required) and analysis. Simple analysis can be applied for individual dataset of single district/KVK and advance analytics can be done for cluster of districts.

Data analysis is the process of evaluating data using analytical tools to derive useful information for informed decision making. Major types of analysis planned for LDS datasets are as follows:

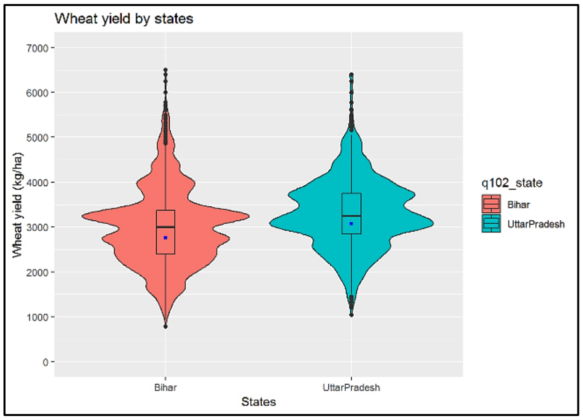
Histogram: A histogram is an accurate representation of the distribution of numerical data. It is an estimate of the probability distribution of a continuous variable. It differs from a bar graph, in the sense that a bar graph relates two variables, but a histogram relates only one. To construct a histogram, the first step is to "bin" (or "bucket") the range of values i.e. divide the entire range of values into a series of intervals and then count how many values fall into each interval. Histograms give a rough sense of the density of the underlying distribution of the data, and often for density estimation. The given histogram is of wheat yield of surveyed farmers.



Boxplot: A boxplot is a method for graphically depicting groups of numerical data through four quartiles. Box plots may also have lines extending vertically from the boxes (whiskers) indicating variability outside the upper and lower quartiles, hence specifically termed as box-and-whisker plot. Outliers may be plotted as individual points. Box plots are non-parametric: they display variation in samples of a statistical population without making any assumptions of the underlying statistical distribution. Box plots can be drawn either horizontally or vertically. Box plots received their name from the box in the middle. The given boxplot represents wheat yields of two surveyed states.

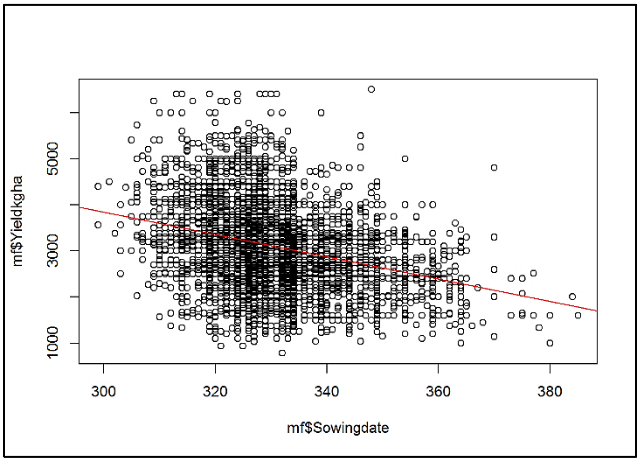


Violin plot: Violin plots are similar to boxplots, except that they also show the probability density of the data at different values, usually smoothed by a kernel density estimator. Typically a violin plot will include all the data that is in a box plot. So, a violin plot is more informative than a plain boxplot. While a boxplot only shows summary statistics such as mean/median and interquartile ranges, the violin plot shows the full distribution of the data. The difference is particularly useful when the data distribution is multimodal (more than one peak). In this case a violin plot shows the presence of

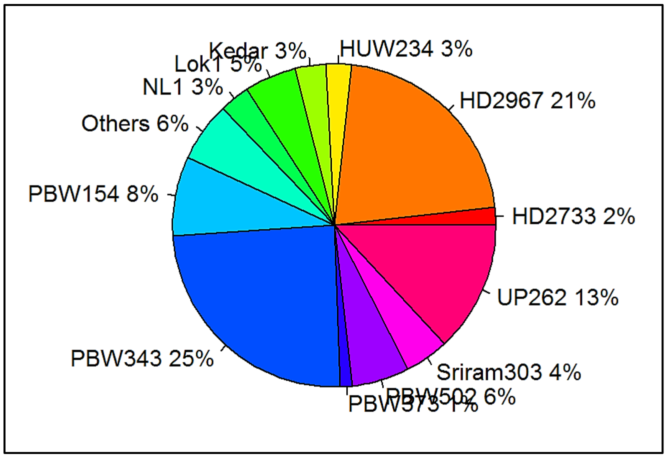


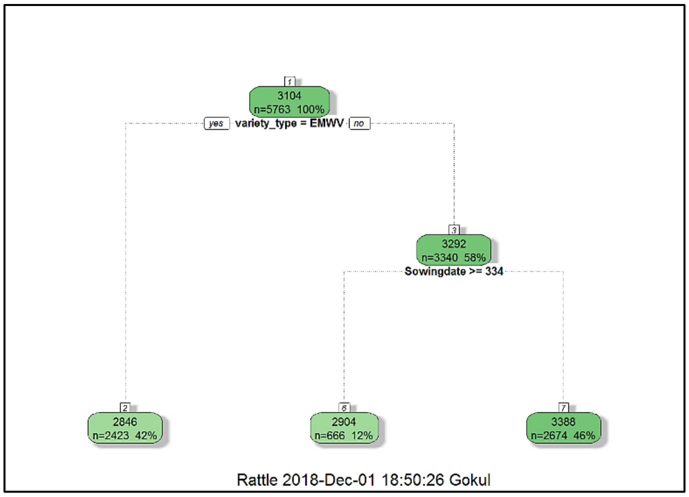
different peaks, their position and relative amplitude. The given violin plot compares wheat yields of two surveyed states.

Scatter plot: A scatter plot also called a scatter diagram, is a type of plot using Cartesian coordinates to display values for typically two variables for a set of data. The data are displayed as a collection of points, each having the value of one variable determining the position on the horizontal axis and the value of the other variable determining the position on the vertical axis. A scatter plot can suggest various kinds of correlations between variables with a certain confidence interval. If the pattern of dots slopes from upper left to lower right, it indicates a negative correlation. A line of best fit (alternatively called 'trendline') can be drawn in order to study the relationship between the variables. The given scatter plot is indicates correlation between wheat yield and wheat sowing dates.



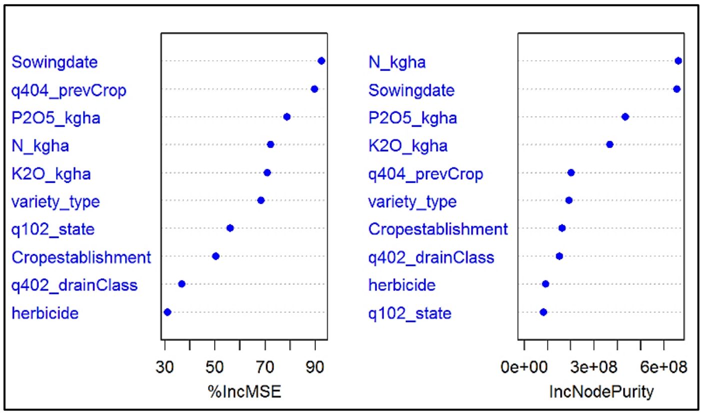
Pie chart: A pie chart (or a circle chart) is a circular statistical graphic, which is divided into slices to illustrate numerical proportion. In a pie chart, the arc length of each slice (and consequently its central angle and area), is proportional to the quantity it represents. Pie charts are very widely used tool in the business world. Research has shown that it is difficult to compare different sections of a given pie chart, or to compare data across different pie charts. Pie charts can be replaced in most cases by other plots such as the bar chart, box plot or dot plots. The given pie chart shows proportion of farmers using different wheat varieties.





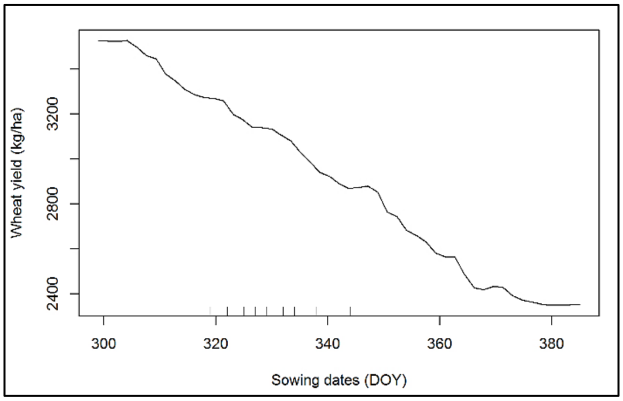
Rattle 2018-Dec-01 18:50:26 Gokul

Classification and regression tree (CART): CART is a term used to describe decision tree algorithms that are used for classification and regression learning tasks. A decision tree is a supervised machine learning algorithm. It has a tree-like structure with its root node at the top. Decision tree algorithms are nothing but if-else statements that can be used to predict a result based on data. A classification tree is an algorithm where the target variable is fixed or categorical. A regression tree refers to an algorithm where the algorithm is used to predict its value. The interpretation of results summarized in CART is fairly simple. The given CART predicts wheat yield based on variety type and sowing date.



Variable importance plot: Random forests can be used to rank the importance of variables in a regression or classification problem. Variable importance plot provides a list of the most significant variables in descending order. The top variables contribute more to the model than the bottom ones and also have high predictive power. Variable importance is calculated by the sum of the decrease in error when split by a variable. Then, the relative importance is the variable importance divided by the highest variable importance value. The given plot ranks variables of importance for wheat yield.

Partial dependence plot (PDP): These plots are graphical visualizations of the marginal effect of a given variable on an outcome. Typically, these are restricted to only one or two variables. In linear regression with a single independent variable, a scatter plot of the response variable against the independent variable provides a good indication of the nature of the relationship. If there is more than one independent variable, things become more complicated as this does not take into account the effect of the other independent variables in the model.


Wheat yield (kg/ha)
Sowing dates (DOY)