Data Visualization of Plant Resistant Towards Plant Disease At PT. East-West Seed Indonesia

D Dharmayanti, A M Bachtiar, M A D Santoso
Departemen Sistem Informasi, Universitas Komputer Indonesia, Indonesia
*dian.dharmayanti@email.unikom.ac.id

Abstract. The purpose of this study is to produce appropriate data visualization in the domain of plant resistance cases used in understanding information for plant pathology researchers. The method used in this study was data visualization. The results in this study shows slow handling of the dangers of attacks by the pest so that production decreases will occur again. An effective way to present information is to turn it into a form of visual information. This research produces visualization designs to be implemented in the form of software prototypes. Thus, plant pathology researchers can understand the information they need with the results of this visualization.

1. Introduction

The agricultural sector can help the government in obtaining foreign exchange, providing raw materials and reducing poverty [1]. But the agricultural sector continues to decline, due to a decline in the problems of production, distribution and affordability. One factor that causes a decrease in production in the agricultural sector is the attack of Plant Pest Organisms (OPT) [2]. PT East West Seed Indonesia is one of the agribusiness companies that conducts research on pests. Based on the results of interviews conducted with one of the plant pathology researchers at PT East West Seed Indonesia, information was obtained that the researchers were still having difficulty understanding the information from the results of research on the resistance of plant varieties to pests. Information about the resistance of plant varieties to pests is presented in the form of an excel file. Annual recapitulation is used to see the development of plant resilience carried out in plant pathology laboratories and also for breeding processes to produce good quality durability. Errors and long periods of time when reading information make a delay in handling pests and slowing down the breeding process. The impact of all this results in a decrease in crop production. Based on current problems, data visualization can help plant pathology researchers to gather information quickly, understand large amounts of information, and can also interact directly with data. Because the effective way of presenting detailed data into information that is easily accepted is by abstraction into visual information [3].

The purpose of this study was to produce appropriate data visualization in the domain of plant resistance cases used in understanding information for plant pathology researchers. Information from the results of visualization can help plant pathology researchers in controlling pest attacks, find out the various types of pests that are owned, know the dominant pests of each plant, and know the spread of pests. The stages carried out in this study consist of 8 steps which complement the book "Data Visualization".
2. Methods
The stages carried out in this study consist of 8 steps which complement the book "Data Visualization" [4]. The method can be seen at Figure 1:

![Figure 1. Methods](image)

All stages in Figure 1 are done in stages. At the stage of designing visualization consists of 6 stages in more detail, namely understanding the context, determining the appropriate form of presentation, eliminating clutter, giving attention to focus on the user, then improving visualization by thinking as a designer and finally telling the visualization results [4]. At the last stage of testing using the usability testing method. This method is evaluative in that the user will be given a task to find out the level of success in completing a given task.

3. Results and discussion
In this section, we will explain in detail each stage and research of the research.

3.1. Information Needs Analysis
Job description of plant pathology researchers at PT Eest West Seed Indonesia is to identify plant pests or diseases, test disease resistance, conduct new identification methods, and make reports. The four jobs were then analyzed to obtain information needs. The results of the analysis found that not all jobs require visualization to present information [5]. Jobs that require visualization are:

a. The work of identifying pests is obtained that the information needed is to determine the distribution of pests that most often arise from each plant by province, to know the distribution of pests per plant in each city/province, and to know the pests of each plant. The information needs were obtained from identification data which consisted of identification of isolates and insects.

b. The disease resistance testing work obtained that the information needed was to determine the number of plant endurance tests over a period of months and to compare the percentage of plants that were resistant to life and vulnerable among tests that had good resistance by month. Both the information needs of the data obtained from the data trial.

3.2. Data Sources Analysis
In the previous step, interviews conducted with one plant pathology researcher produces data used is isolate identification data, insect identification and trial. Data used in the study is a private data derived from plant pathology lab at PT East West Seed Indonesia. The data format used is excel (.xlsx).

Isolate identification data consists of several bacterial categorized data, that is data of bacterial wilt, the data Xanthomonas vesicatoria, Xanthomonas campestris the data, the data and the data bacterial soft rot fruit boltch. Insect identification data consists of data categorized insect, namely the data whitefly and thrips of data. Data trial consists of only one of the data contained test plant resistance to a pest that is data trial.

3.3. Data Analysis based on Information Needs
Analysis of the data mapping is done by listing the attributes of the data needed in the analysis of information needs. List mapping attributes based on their information needs can be found at Table 1:
Table 1. Mapping attributes on the information

| No. | Information needs                                                                 | Attributes of the data needed                             |
|-----|-----------------------------------------------------------------------------------|-----------------------------------------------------------|
| 1.  | Knowing the distribution of the pest or disease occurs most often per plant by province | Location, crop, disease                                   |
| 2.  | Knowing what cities are attacked by the pest of each plant per province             | Location, crop, disease                                   |
| 3.  | Knowing collection pest or disease of each plant and Knowing the amount of endurance testing plants that have good resistance per month | Crop, disease, No trial, crop, population, The percentage of resistant life, The percentage of susceptible life, order date |
| 4.  | Knowing the percentage of resistant and susceptible vibrant per test that has good endurance by month | No trial, % susceptible live, % resistant live           |

3.4. Methods Analysis based on Information Needs

After mapping the attributes of each need for information, the mapping method for the extraction of knowledge. Not all the necessary information needs to mengestraksi knowledge method. The second and third information needs can be directly visualized, so it does not need a method for the extraction. Here mapping methods to the needs of the information, can be found at Table 2:

Table 2. Mapping methods to the needs of the information

| No. | Information needs                                                                 | Method                          |
|-----|-----------------------------------------------------------------------------------|--------------------------------|
| 1.  | Knowing the distribution of the pest or disease occurs most often per plant by province | Statistics (Mode)              |
| 2.  | Knowing the amount of endurance testing plants that have good resistance per month | Clustering (K-Means)           |
| 3.  | Knowing the percentage of resistant and susceptible vibrant per test that has good endurance by month | Clustering (K-Means)           |

a. Statistical

The need of information using statistical techniques is knowing the pest or disease occurs most often per plant in each province. The method used to perform the extraction of knowledge by finding mode. Frequency most of a pest from each plant in a province will determine these pests into pest is most often arise. The data used is data isolate and insect. To facilitate the search mode, do the grouping data based on crop and location. After that sought the mode of each group by calculating the frequency of occurrence.

b. Clustering

Information needs to use clustering techniques is knowing the amount of endurance testing plants that have good resistance per month and determine the percentage of resistant and susceptible vibrant per test that has good resistance per month[6]. The algorithm used for extracting knowledge is a k-means algorithm [7,8]. The data used is the trial from January through March 2018. The variables used were persentase_rc_hidup (x) and persentase_sc_hidup (y). The first thing to do is to determine the value of k. The value of k for this case is 2. Then, initialization of centroid 1 and 2 (x; y) with the centroid 1 (0.21; 0.31) and centroid 2 (0.32; 0.22) which is drawn from a range of variables x and y, then arithmetic seldom each member with the nearest centroid. Keep going until each cluster member unchanged [9]. The results of this calculation is done until iteration 10. The number of members of the cluster 1 there were 29 (51.79%) and the number of members of the cluster 2 there were 27 (48.21%).

3.5. Design Visualization

Design visualization there are four stages of understanding the context, the selection of visualization, minimizing clutter and giving attention [10]. The fourth step is done for every need of information[12].
a. The first information needs: to know the distribution of the pest or disease occurs most often per plant by province

There are two layers to the first information needs. Here's an explanation of each layer.

1) layer 1
   Visualization can be viewed at Figure 2:

   ![Figure 2. Results visualization needs first information (layer 1)](image)

   At layer 1 information to be conveyed is the risk of pest attack on a plant, whereas the information to be conveyed is in the form of geographical visualization of the presence or absence of a potential attack of the pest by province in Indonesia. Preferred forms of visualization for presenting this information is choropleth map. The reason for using the visualization of this is because it can present information from categorical data divided by geography. Lastly is giving attention to visualization. Preattentive attribute used is color. Color is used to inform the meaning of the area. The red color means the area potentially affected pest and light gray is an area that does not exist.

   1) layer 2
   At layer 2 the information you want to convey is the details of pests that attack a province. The form of visualization chosen to present this information is simple text. The reason is simple text matching to present one or two information. Last is to pay attention to visualization using bold letters, to emphasize the name of the pest.

b. The second information needs: know what cities are attacked by the pest of each crop in each province.
   There are two layers to the information needs of both. Here's an explanation of each layer.
   Visualization can be viewed at Figure 3:
Figure 3. Results visualization of the information needs of both (layer 1)

1) layer 1
   At layer 1 information to be conveyed is a city affected by pests, whereas the information to be conveyed in the form of geographical visualization of the presence or absence of a potential attack of the pest by province in Indonesia. Preferred forms of visualization for presenting this information is choropleth map. The reason for using the visualization of this is because it can present information from categorical data divided by geography. The latter is the awarding of attention. Preattentive attribute used is color. The red color means that the affected areas and ash means no data.

   The end result visualization can be viewed at Figure 4:

   Figure 4. Results visualization of the information needs of both (layer 2)

2) layer 2
   At layer 2 information to be conveyed is detail pest that attacks a town of the city, whereas the information to be conveyed by displaying a list of names of pests that attack a city. Preferred forms of visualization for presenting this information is simple text. The reason being simple text matching to present one or two pieces of information. Lastly is giving attention to the visualiasai. Preattentive attribute used are bold font-style, to give emphasis to the name of the pest.

c. The third information needs: know the collection of pests or disease of each plant.
   Visualization can be viewed at Figure 5:
d. The information to be conveyed is a collection of pests from each plant, while the information to be conveyed to display a list of pests per crop. Preferred forms of visualization for presenting this information is simple text. The reason being simple text matching to present one or two pieces of information. The latter is the awarding of attention. Preattentive attribute used is bold font-style in the name of the plant for emphasis.

e. The fourth information needs: know the amount of endurance testing plants that have good resistance per month.
Visualization can be viewed at Figure 6:

![Figure 5. Results visualization of the information needs of the third](image)

The Number Of Trials that have Good Endurance in 2018

![Figure 6. Results visualization fourth information needs](image)

The information to be conveyed is the conclusion of testing on a monthly basis, while the information to be conveyed by showing the amount of testing that have good plant resistance of each month. Preferred forms of visualization for presenting this information is simple text. The reason being simple text matching to present one or two pieces of information. The latter is the awarding of attention. Preattentive attribute used is the font-size and font-style bold green color also on the amount of testing as well as font-style bold and green colors on the label testing. The green color defines a good emotion to the audience.

f. The fifth information needs: know the percentage of resistant and susceptible vibrant per test that has good endurance by month.
Visualization can be viewed at Figure 7.
The information to be conveyed is plant resistance testing results every month while the information to be conveyed by showing the percentage of resistant and susceptible plants that have good endurance of each month. Preferred forms of visualization for presenting this information is a horizontal bar chart. The reason is because it is suitable for presenting the data categories and variables that has a long name. The latter is the awarding of attention. Preattentive attribute used is color. The green color that has a good sense are used to represent the testing that has most excellent durability.

3.6. Acceptance Testing of Visualization
Improved visualization of results can be seen in Figure 8:

Acceptance testing was conducted to determine the design visualizations created previously can easily be understood by the audience or not [11]. Testing is done to one influent members in plant pathology lab by asking three questions for each visualization.

a. What information you can from this visualization?
b. Do you feel any difficulty while reading this visualization?
c. Is there any feedback that you want to associated visualization?

Based on the results of the acceptance tests, only the visualization of the five information requirements needs to be improved. The improvement made is to remove the percentage scale and replace it with a label percentage in each bar. This is done to clarify the percentage of each bar.

3.7. Prototype testing with Usability Testing
After visualization of any information requirements are designed and tested finished, the next step is to implement it in prototype form. Tests carried out to the same person when performing acceptance testing. The purpose of this test is the value of the efficiency and effectiveness of the completion of each task is greater equal to 70%. Following the task given to the participants at the time of testing:
a. Task 1: Go to the page containing the spread of pests from plant visualization cayenne most widely attacked by province in Indonesia.
b. Task 2: Go to the page containing the spread of the pest visualization of cucumber plants in West Java
c. Task 3: Go to the page that contains the visualization of each plant pest collections.
d. Task 4: Go to the page that contains the visualization of the information on the number of trial which has good durability in January 2018,
e. Task 5: Go to the page that contains the visualization on the percentage of resistant and susceptible plants that live on each trial that has good durability in February.

The success rate of the test was 3, which:

a. 0%: declare failure, where tasks are not successfully done.
b. 50%: declare success, but in completing the assigned task without the need to be given directions.
c. 100%: declare success, in completing the task run smoothly in terms of time and also the objectives achieved

Evaluation of testing usability testing can be seen in Table 3.

| Task | Anomaly | Level of success | Time (sec) |
|------|---------|-----------------|------------|
| 1    | There is no | 100%            | 6 sec      |
| 2    | Clicking on the map section West Java Indonesia | 100% | 10 sec |
| 3    | There is no | 100%            | 3 sec      |
| 4    | There is no | 100%            | 15 secs    |
| 5    | Clicking card month | 100%    | 10 sec |

Results of the test will show success with views of efficiency and effectiveness. Efficient each task obtained from the task completion time of each participant, while effectiveness is seen from the success of the task. Efficient value and effectiveness of each task is 100%. It states that the purpose of usability testing is achieved because the value of greater efficiency and effectiveness equal to 70% (See Figures 9-15).

![Distribution](distribution.png)

Figure 9. Interface Prototype First Information Needs (Layer 1)
Figure 10. Interface Prototype First Information Needs (Layer 2)

Figure 11. Prototype Interface Information Needs Both (Layer 1)
**Figure 12.** Prototype Interface Information Needs Both (Layer 2)

**Figure 13.** Interface Prototype Third Information Needs
4. Conclusion
Based on the results of testing usability testing as seen from the efficiency and effectiveness, the resulting visualization is right for the visualization of plant pathology researcher at PT East West Seed Indonesia. Facilitate visualization of plant pathology researcher to obtain information for doing their job. In this study, the role is taken of plant pathology researcher. Therefore, in the next study is expected to take on other roles related to plant resistance.

Acknowledgment
The author would like to thank PT East-West Seed Indonesia for providing data and information relating to the object of research and UNIKOM as providing facilities in completing this research.
References

[1] M. Sodiq, *Strategic Plan 2015-2019 Body Awareness And Human Resources Development Of Agriculture*. 2015. Surabaya: Ministry of Agriculture.

[2] Foster, A. D., & Rosenzweig, M. R. (2004). Agricultural productivity growth, rural economic diversity, and economic reforms: India, 1970–2000. Economic Development and Cultural Change, 52(3), 509-542.

[3] NAS and AM Bachtiar. 2016. Open Data Interactive Data Visualization Provincial Government Jakarta: Subject Economic And Financial Area.

[4] A. Kirk. 2012. *Data Visualization: a successful design process*, 1st ed. Birmingham: Packt Publishing Ltd.

[5] AM Bachtiar and D. Dharmayanti. 2018. Data Visualization at Trading Sector Using Government's Open Data. 118(20), pp. 523-530.

[6] L. Morissette and S. Chartier. 2013. The k -means clustering technique: General considerations and implementation in Mathematica. 9(1), pp. 15-24.

[7] E. Kandogan and U. Interfaces. 2012. Just-in-Time Annotation of Clusters, Outliers, and Trends in Point-based Data Visualizations. *IEEE Symp. Vis. Anal. Sci. Technol. 2012*, Pp. 73-82.

[8] C. Cîmpanu and L. Ferariu. 2012. Survey of Data Clustering Algorithms. *Bul. Institutului Politeh. din Iasi, Tome LVIII (LXII)*, 3(LXII) , pp. 23-42.

[9] M. Shamim Reza. 2015. "Study of Multivariate Data Clustering Based on K-Means and Independent Component Analysis," *Am. J. theor. Appl. Stat.,* 4(5), p. 317.

[10] B. Martin and B. Hanington, "Universal Methods of Design," pp. 2-480, 2012.

[11] Trenner, L., & Bawa, J. (2012). The Politics of Usability: a practical guide to designing usable systems in industry. Springer Science & Business Media.

[12] CN Knaflic. 2015. *Storytelling With Data*. 8. June. John Wiley & Sons, Inc.