

EFFECTIVENESS OF CLASSIFICATION METHOD AND COLOR SYMBOL SCHEME ON CHOROPLETH MAP OF POPULATION DENSITY IN SPECIAL REGION OF YOGYAKARTA

Zulfa Nur'aini 'Afifah^{1*} Sudaryatno²

¹ Dept. of Geographic Information Science, Universitas Gadjah Mada, Bulaksumur, Yogyakarta 55281,
zulfa.nuraini@mail.ugm.ac.id

² Dept. of Geographic Information Science, Universitas Gadjah Mada, Bulaksumur, Yogyakarta 55281, sudaryatno@ugm.ac.id

KEY WORDS: classification method, color symbol scheme, choropleth map, proportion assesment, eye-tracking

ABSTRACT:

The need for presenting information in maps is increasingly high in various scientific fields. All scientific fields need to present effective data for decision making. Good decision making based on maps requires good understanding but not all scientific fields are familiar with using maps. Supporting factors for easy maps to understand are classification method and color symbol scheme. The purpose of this study was to select and test the classification method and the most effective color symbol scheme for mapping population density in the Special Region of Yogyakarta. The classification methods used in this study are constant interval, arithmetic progression, geometric progression, quantile, standard deviation and dispersal graph. The effectiveness test method for the most effective classification method is the proportion assessment. The color symbol scheme used in this study is a sequential color scheme, diverging color schemes, Corel Draw color schemes and color symbol schemes provided in ArcMap 10.3 software. The effectiveness test method for the most effective color symbol scheme is conventional eye tracking. The results showed that according to the proportion test the most effective classification method was the arithmetic interval classification method with results of 0.26. The most effective color symbol scheme in accordance with the effectiveness test using the conventional eye tracking method shows that the most effective color symbol scheme is a diverging color scheme. The important aspects to consider are average answering duration of 8.15 seconds, the accuracy of the answer is 98.9%, and easiness level of symbolization readings is 341. This research can be one of the references on the most effective classification method and reference regarding the selection of the most effective color symbol scheme on Choropleth Map of Population Density in Special Region of Yogyakarta, so that further research can continue the analysis of appropriate classification methods for demographic data. The method discussed in this study is also expected to be applicable to other data.

1. INTRODUCTION

Ways to understand a map are by the class of data classification result which is given and the symbolization which are used. Analysis of the classification method on the classes of data classification results given and visualization of data concerning with the symbolization used are important to analyze. According to Bandrova and others (2014), the data that are analyzed or represented usually are classified with the aim to simplify the choropleth map.

Moreover, they are aimed to minimize the observation error of an individual user of the map on a class of classification results which is symbolized by a certain color. Wei and others (2017) claimed that the important purpose of data classification are to maximize the homogeneity on each of the classes of classification results where similar values can be presented using similar pattern/color and map user will be able to explore special pattern of geographic phenomenon which is represented.

Kraak and Ormelling (2007) explained that choropleth map is a thematic map which is suitable to visualize economy social data. One of the economy social data which can be represented using choropleth map is demographic data. Demographic data is a data that consists of population of an administrative region. Administrative region which has a good demographic data and required to be represented is Special Region of Yogyakarta.

The choice of study region in Special Province of Yogyakarta is due to the region being one of the province in

Indonesia which has a high population density for an area of 3.185,80 km². In 2016, the population density of Special Province of Yogyakarta was in the 4th position in Indonesia Statistic Year 2017 by Statistics Indonesia with population of 1.188 per km². Therefore, this research attempts to investigate the effectiveness on the choice of classification method and color symbol scheme on choropleth map of population density in Special Region of Yogyakarta, especially for population density data.

The aims of this research are as follows:

1. To choose the most effective classification method for choropleth mapping of population density in Special Region of Yogyakarta
2. To present the data using the most effective classification method for choropleth mapping of population density in Special Region of Yogyakarta
3. To test and find out the most effective color symbol scheme to present choropleth map based on the case study of Demographic Mapping in Special Region of Yogyakarta
4. To present the data using color symbol scheme as the most effective visual variable for choropleth mapping of population density in Special Region of Yogyakarta.

2. STUDY AREA AND DATASET

The selection of the study area in the Special Province of Yogyakarta is because the Special Region of Yogyakarta is one of the provinces in Indonesia which has a fairly high population density in the area of 3,185.80 km². In 2016, the population density of the Special Province of Yogyakarta ranked 4th in the Indonesian Statistics 2017 by the Central Bureau of Statistics, which amounted to 1,188 residents per km².

Therefore, it is necessary to map the population density of the Special Region of Yogyakarta so that the data can be easily understood using the choropleth map with the classification method and the most effective color symbol scheme.

Administratively the Special Region of Yogyakarta Province consists of 4 districts, which are divided into 78 sub-districts, and 438 villages.

Geographically, the Special Province of Yogyakarta is located in the south-central part of Java Island, more precisely located at 8° 30' - 7° 20' South Latitude, and 109° 40' - 111° 0' East Longitude according to figure 1.

Administratively, the Special Province of Yogyakarta borders on several regions:

- North side: Boyolali Regency, Central Java Province
- West side: Purworejo Regency, Central Java Province
- East: Wonogiri Regency, Central Java Province
- South: Indian Ocean



Figure 1 Map of Research's Area

Tools and Materials

Research Tools

- a. Laptop to process the data which is equipped with:
 - ArcGIS 10.3 : used to visualize seconder data and create choropleth data
 - SPSS/Microsoft Excel : used to process the data in statistics
 - Corel Draw : used to manually arrange color schemes
 - Google Chrome : used to access website of Color Brewer (arranger of color schemes)
- b. Stationary

Research Materials dan Datasets

- a. Data of Total Population in Special Region of Yogyakarta
- b. Data of Vector (Shapefile) in Special Region of Yogyakarta
- c. Questionnaire
- d. Printed Map/Album of Conventional Choropleth Map

3. METHODOLOGY

a. Data Classification

The stages of making data classifications (according to Jenks and Coulson 1963):

1. Determine the data type to adjust the data classification method
2. Limiting the number of classes
Sturges = $1 + 3.3 \log (n)$
Description: n = amount of data
3. Limiting the number of classes, determining the lowest class boundaries and following other boundaries which are continuations of the lowest class limit.

Classification Methods:

- Constant Interval Method

$$Interval = \frac{Highest\ Value - Lowest\ Value}{Total\ Class}$$

- Arithmetic Progression Method

Arithmetic Progression Method uses interval which is obtained from the highest and lowest value from the available data which then be divided with the total classes counted beforehand, similar to interval on constant interval classification. This interval is used as a variable which will be multiplied by specific coefficient on each additional class in every class of classification result.

- Geometric Progression Method

Geometric Progression Method uses different interval from the previous. Interval value is obtained from this formula:

$$Log\ Interval = \frac{Highest\ Log\ Value - Lowest\ Log\ Value}{Total\ Class}$$

- Quantile Classification Method

Quantile classification method uses a different approach with the previous 3 methods. This method uses the data sequence which has been arranged from the lowest to highest value. After the data is arranged, they will be classified according to the total class where each class has the same total members with upper limit and lower limit which cannot be controlled.

- Standard Deviation

Standard deviation method uses standard deviation as an interval for data classification where average calculation of all the data that needs to be classified is required. Standard deviation is used as a standard for division of value from data.

- Dispersal Graph

Dispersal Graph uses a different approach from all of the classification methods used in this research, which is using graphic approach. All of the data value is entered in a graphic that shows the relationship between X (data) and Y (data value). Classification is carried out by observing all data values which tends to have similar value and those similar values can be entered into the same class.

b. Most Effective Classification Method Test (Classification Method Effectiveness Test)

In the article by Kurniati dan Rahardjo (2012), it was explained that proportion test is a wrong statistic method test where the test is carried out by measuring the errors resulted from some of the classification methods which involved absolute distance compared to the length of class. The smaller the value being obtained, the higher the chances classification method able to represent good spatial distribution. Below are mathematic formulas that are used in proportion test:

1. Median value that are obtained are median value of all the population density data which has been classified.

$$Median = Lowest Value + \frac{Highest Value - Lowest Value}{Total Class}$$

2. Absolute distance is a value which is obtained after the calculation of median value, which then be operated on all of the population density data in each sub-district so the value of absolute distance for all of the sub-districts can be obtained. Preceding that, the absolute distance will then be used as a comparison value for the calculation of proportion value.

$$Absolute Distance = | Density Value - Median |$$

3. Class length is the value which is obtained from each classes of classification method. Class length then be used as a comparison value for the calculation of proportion value.

$$Class Length = Upper Limit - Lower Limit$$

4. Proportion value is result of comparison operation of absolute distance value and class length which is obtained for each sub-district that has been classified according to each classification method. The proportion value is obtained from each classes of the result from classification methods, which is continued by calculation of average of proportion value for all population density data in each sub-district. Classification method which has the lowest proportion value is the most effective method.

$$Proportion Value = \frac{Absolute Distance}{Class Length}$$

c. Symbolization and Visualization

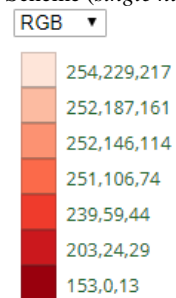
Cartographic elements that explain the best colors in the presentation of population density maps are not specifically found. however, there are several sources that explain red in mapping can be used for major highways, roads, urban areas, airports, special interest sites, military sites, place names, buildings, and borders (Rosenberg, 2019). The study area mapped is included in the urban areas so that it can use red. in addition, in the Cartography book by Geoff Hatchard (2018) there is a population density map of choropleth with the main color red. the research conducted by Kurniati and Raharjo (2012) also uses red to present a population density choropleth map. In Indonesia there is no specific standard that sets colors to represent maps of population density, so it is hoped that this research can be a reference for making population density maps.

Color symbol scheme is a color symbol generator system which is formed in the form of area dimension, which is used for choropleth mapping. Color symbol scheme is generated from <http://colorbrewer2.org> which consists of sequential color symbol scheme (*single hue and multi hue*), Diverging and qualitative.

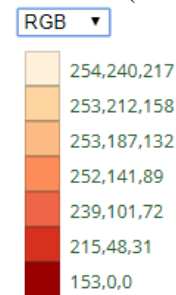
However, in this research the color symbol scheme that will be analyzed is Sequential Color Schemes, both single hue and multi hue, Manual Color Schemes and Diverging Color Schemes with the addition of color symbol schemes available in the software ArcGIS 10.3 which is *ArcMap Color Ramp*. The calculation of total classes for the data is 8 classes of population density.

Color Symbol Scheme:

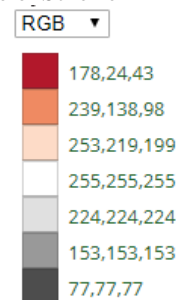
- Sequential Color Scheme (*single hue*)



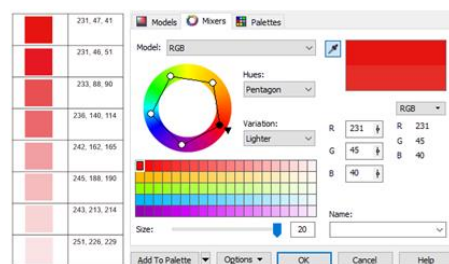
- Sequential Color Scheme (*multi hue*)



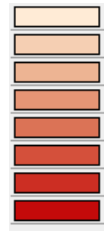
- Diverging Color Scheme



- Color Scheme created by *Corel Draw*



- Original Color Symbolization Scheme ArcGIS 10.3



d. Most Effective Color Symbol Scheme Test

According to the research done by Brychtova and Coltekin (2015), *Eye Tracking* method is the method to find out the quickness of map user in differentiating colors where in this research, aspects used are the duration in reading the symbols, duration in answering the questions on the questionnaire and the accuracy of the reader’s answers.

This method is applied in the test conventionally by using printed map and questionnaire. Duration calculated by stopwatch in unit of seconds. This method is used in the test of the most effective color symbol scheme with these stages as follows:

- **Creating A Map User Sample Scheme**

The expected sample map user involved in this research is those estimated as a target for creating choropleth map. The sample user can be students or university students in the age range of 20-25 years old and staff who are from the government or private agencies with the age above 25 years, which will be taken by using sampling method of purposive sampling due to it being compatible with the scheme created by the map creator with a specific aim or means.

- **Creating Questionnaire**

Questionnaire which is created to test the most effective color symbol scheme consists of:

1. Choropleth Map of Population Density in Special Region of Yogyakarta using different color symbol scheme which is printed in A3 size
2. Questions regarding the map

- **Questionnaire Distribution**

Distribution of questionnaire is held in Special Region of Yogyakarta. The distribution prioritize map user from field of study who needs representation of data statistics specifically the population density in the form of choropleth map. The expected map users are expected to be competent according to the information stated in the map. Sampling scheme above can represent targeted user of choropleth map.

Other than targeted sample user, sample who filled the questionnaire are expected to be able to represent the ability of the user in reading the map with color scheme used in choropleth map symbolization, both experts and novices. Questionnaire is filled along with album of maps which consists of 5 maps with different color scheme of choropleth map symbolization.

- **Results Tabulation by Statistic**

The results in the questionnaire is tabulated in the form of statistic according to sample classification which resulted the information as follows:

1. Right answers on all five maps according to the color scheme used
2. Wrong answers on all five maps according to the color scheme used
3. Time needed to read the symbolization information on five maps according to the color scheme used
4. Map that is the easiest to understand

The result of this tabulation can be presented by using statistic graphic that is easier to understand, which from the information above can be concluded the most effective color symbol scheme which continued on by Choropleth Map of Population Density in Special Region of Yogyakarta Using Classification Method and the Most Effective Color Symbol Scheme.

4. RESULT AND DISCUSSIONS

a. The Most Effective Classification Method

Table 1 Calculation Summary of Proportion Test

Proportion Test Results	Constant Interval Method	Arithmetic Progression Method	Geometric Progression Method	Quantile Classification Method	Standard Deviation	Dispersal Graph	The Most Effective Classification Method According to the Lowest Proportion Value
Total	22.60	20.35	20.71	23.74	21.85	22.85	0.26 (Constant Interval Method)
Average	0.29	0.26	0.27	0.30	0.28	0.29	0.26 (Constant Interval Method)

The calculation result of the proportion test for each population density data of each sub-district in Special Region of Yogyakarta from all of the classification methods shows that interval arithmetic is the most effective method due to the result produced is the smallest proportion value which is 0.26, according to table 1 presented above.

This result represents that all of the population density data occupy or are the members of the correct classification class. Arithmetic interval shows a unique results because according to the population density data, there are only 7 classes of classification. Additionally, classification method of arithmetic interval is expected to give good result of special distribution.

Visualization of the arithmetic interval results on choropleth map (figure 2) is symbolized with color symbol scheme from the software ArcGIS. Other than the results of arithmetic interval, visualization of other classification method results are carried out as well. The visualization result shows that from all of the classification method, the special distribution of the population density data using arithmetic interval is good and ideal.

Alongside that, there are 7 classes of population density. Classification classes which are in odd numbers are expected to make it easier for the map user. Map user obtained the average value or average class as a comparison for the remaining 3 classes above and below it.

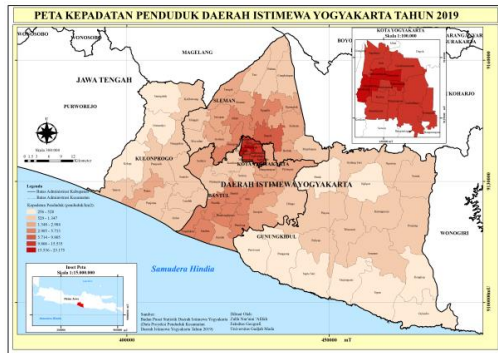


Figure 2 Choropleth Map Using the Most Effective Classification Method

b. The Most Effective Color Symbol Scheme

Effectiveness test of color symbol scheme is carried out using conventional *eye-tracking*. This is caused by the limited technology facilities such as laboratory and *eye-tracker* equipment which at least requires 10 computer with inviting the respondents or map user to the laboratory. Other than that, limited technical issues such as lighting, screen resolution and other technical condition which can disrupt the color interpretation of the user. Color is a sensitive visual variable which requires conventional method that makes the handling of the map user be uniform. It is also to minimize biasness towards the results of the color symbol reading.

Effectiveness test requires visualization of population density data using 5 different color schemes in the form of choropleth map. The map is printed on an ivory paper in the size A3 with the scale of 1:300,000 which allows the user to observe clearly the result of choropleth map, similar to the previous two results which has been discussed.

Choropleth map with different color symbol scheme is marked without differentiating the title of the map. The first map uses color symbol scheme which is available in the software ArcGIS 10.3 called *ArcMap Color Ramp*. The second map uses Manual Color Schemes which is created using the software *Corel Draw*. The third map uses Diverging Color Schemes. The fourth map uses Sequential Color Schemes in Multi Hue and the last map uses Sequential Color Schemes in Single Hue.

The sampling scheme of map user undergoes changes due to condition at the field. The plan of sampling scheme of map user adjusts to the condition at the field. Map user samples which are students represents the user age from 20-25 years old. Samples from institution staff (age over 25 years old) can represent the variety of field of study which uses map as a policy holder and public policy maker. Map user samples should originate from different field of study which can represent various education background of the map user. Population density map can be used by map user of students who are carrying out research or materials for studying.

Moreover, map can be used by institution staff for the necessity of policy maker or advanced level of research. These map user samples originated from various institute. The students are from UGM (Universitas Gadjah Mada). The institute staff are from various institutes such as BPBD (Indonesian National Board for Disaster Management), BAPPEDA (Development Planning Institutional Sub-

National Level), BMKG (Meteorological, Climatological, and Geophysical Agency), BLH (Environmental Agency), Social Services and PSKK UGM (Center for Population and Policy Studies Universitas Gadjah Mada).

Division of study is divided into 50% for map user samples who are familiar with the map and another 50% for map user samples who are unfamiliar with the map. Familiar map user sample are grouped according to background of study in the field of Earth Science. On the other hand, those map user sample who are unfamiliar with the map has various background of study.

The number of respondents who were samples of map users in this study were as many as 90 people. Respondents joined in this study consisted of 30 familiar map students, 30 unfamiliar map students, 15 familiar map institution staffs and 15 unfamiliar map institution staffs. There is 1 respondent who has color impairment, he is a familiar map institution staff. Based on the results obtained, the respondents claimed that the conventional *eye-tracking* method had no effect in reading the population density maps compared in this study. The results obtained were also not much different from other respondents. However, the results will be different if the respondents who have color impairment are unfamiliar map respondents.

The following is a review of the most effective color symbol schemes of each of the important aspects in conventional *eye-tracking*:

- Average Answering Duration

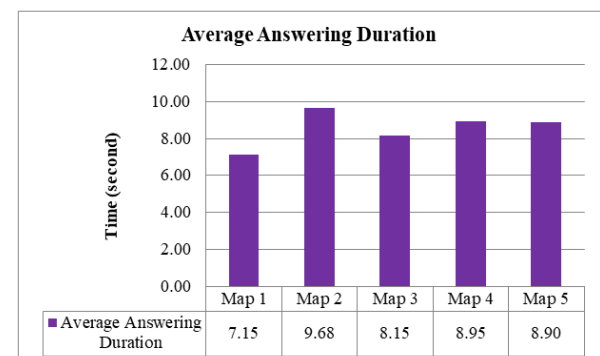


Figure 3 Average Answering Duration Graphic

Average duration of all the map users shows the easiness in reading symbolization of sub-districts as an object of interest. Based on the Figure 3, the average duration of all the fastest map users in answering on map 1 is 7.15 seconds. Factors that affects the duration to answer are selection of sub-districts and color symbol schemes.

The selection of sub-districts can be seen from distribution of spatial symbol as a comparison. If there is a sub-district as an object of interest in the position which is relatively easy to differentiate with other sub-districts, hence it will simplify the map users in identifying and interpreting the correct symbols.

The selection of color symbol scheme which is RGB composition. RGB composition of each color symbol gradation in population density affect greatly the easiness in reading the symbols. The greater the difference contrast between class symbols of population density, the easier the

symbols can be differentiate from one class to another with shorter duration.

The color symbol scheme from ArcGIS on map 1 seemed to be the most effective because the user requires shorter time from all of the maps. Color symbol of Patuk Sub-district in terms of spatial, is easier to compare from the other sub-districts.

The second most effective color symbol scheme is on map 3 which is Diverging color symbol with the average duration to answer of 8.15 seconds. This is caused by color difference which is greatly contrasted between population density symbols and other class population density symbols. On the other hand, the most difficult color symbol scheme and requires more time to read is the scheme from map 2 which is created by Corel Draw.

This is resulted by color differences which is not contrasted and relative position of sub-district as an object of interest which is harder to compare with other classes symbol of population density. The conclusion obtained from all of the results of the duration to answer is that the most effective color symbol scheme is from ArcGIS.

- Answers Accuracy

Menurut anda, berapakah kepadatan penduduk Kecamatan Wirobrajan, Kota Yogyakarta (dalam Inset KANAN ATAS) di dalam peta ini? (Peta 3: Diverging)

89 / 90 correct responses

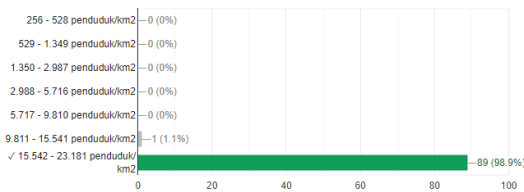


Figure 4 Answers Distribution

Based on Figure 4, all of the students map user samples are able to correctly answer the questions on map 3. The percentage of correct answers are 15,542 – 23,181 population/km² with 98.9%. Whereas the percentage of wrong answers are on option 9,811 – 15,541 population/km² with 1.1%. Wrong answers can be caused by the age of the map user, flight hours of map user in reading maps and the similarity of RGB composition on color symbol classes of population density as an object of interest.

According to the results presented on the graphic, Diverging color symbol scheme which is used on map 3 is the most ideal and effective to visualize the population density with 2 primary colors divided into 7 gradation color or 7 RGB composition. This is caused by RGB composition which is contrast from one another on each classes symbol of population density. The map users claims that map is the easiest to read compared to the other maps.

- Easiness Level of Symbolization Reading

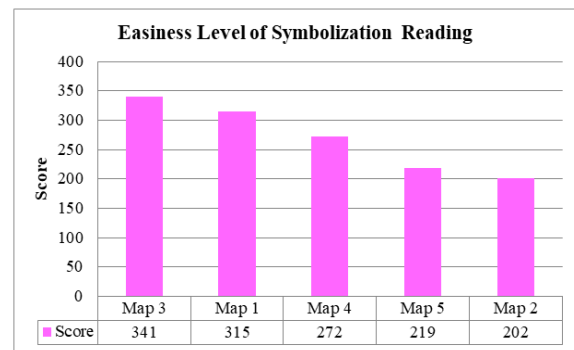


Figure 4.4 Easiness Level of Symbolization Reading

Level of easiness in reading symbolization on five maps are very easy, easy, average, difficult and very difficult. The highest score is 5 for maps that are easy to read and lowest score is 1 for maps that are difficult to read. Graphic summary in Figure 4.4 above shows that the score of level of easiness in reading symbolization are based on the amount of user who ranked the maps according to the level of easiness in reading symbolization.

The amount of map user who chooses map 3 as the most easiest is the highest among the other maps which obtained a score of 341. The conclusion that can be obtained from the level of easiness in reading symbolization is diverging color scheme on map 3 is the most effective color symbol scheme for both map users aging 20 – 25 years old and above 25 years old.

Important aspects which shows the first color symbol scheme are duration to answer, answers accuracy and easiness in reading symbolization. The average fastest duration to answer are on map 1 with 7.15 seconds and on map 3 with 8.15 seconds. The most accurate answers is on map 3 with 98.9%.

The highest score for easiness in interpreting reading of symbolization is on map 3 with 341. Therefore, it can be conclude that the most effective color symbol scheme for map users who are students aging 20 – 25 years old is diverging color scheme.

This is based on the sum of maximum class which can be differentiated by humans in the theory of visual variable isolation by Bertin (1983). Map 3 (Figure 4.5) is one of the color schemes from the website *Color Brewer*. Other color symbol schemes such as sequential color schemes, both single hue or multi hue is considered as less effective based on the results which has discussed previously.

Effectiveness result of color symbol scheme can be used to symbolize other maps which uses choropleth mapping method or quantitative mapping. Diverging color schemes are ensured to be used in mapping that requires more than 8 classification classes. This is based on cartographic theorem and feedbacks from all of the map user samples. Other than diverging color schemes, original color symbol schemes by *ArcGIS* (on Figure 4.6) can be used as an alternative.

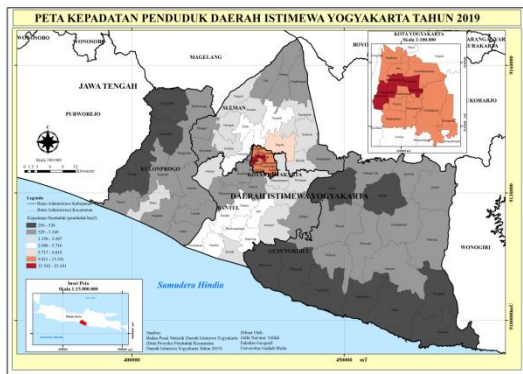
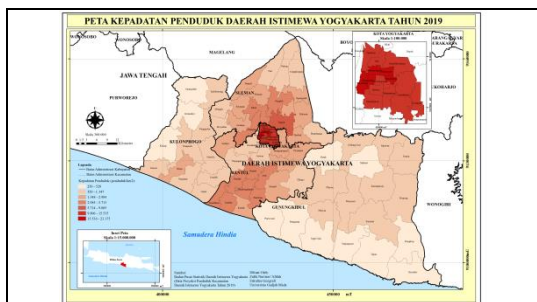
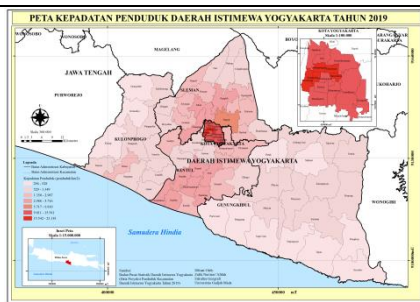


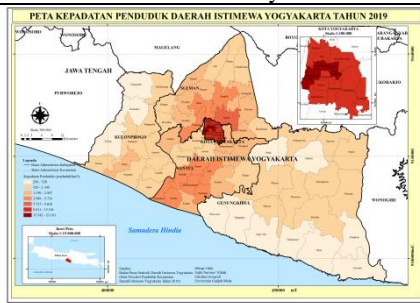
Figure 4.5 Choropleth Map Using the Most Effective Color Symbol Scheme



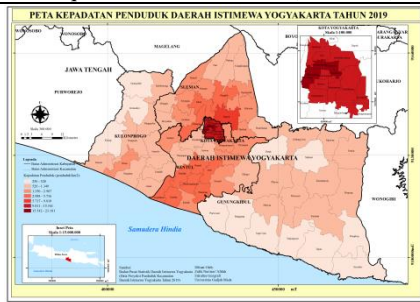
Original Color Symbolization Scheme ArcGIS 10.3



Color Scheme created by Corel Draw



Sequential Color Scheme (multi hue)



Sequential Color Scheme (single hue)

Figure 4.6 Choropleth Map Using Other Color Symbol Scheme

5. CONCLUSION

Based on the research that has been done, the conclusion is obtained to answer the research objectives as follows:

Visualization which has used the most effective color symbol scheme and classification method which are arithmetic interval and diverging color scheme is the most effective map compared to the others hence it is easier to read and understand.

ACKNOWLEDGEMENTS

All studies were supported by Gadjah Mada University. The author would like to thank the organization that has supported this research. The author is also grateful to all team members who have collaborated in completing this research well.

REFERENCES

- Bandrova, Temenoujka. 2014. *Lecture Notes in Geoinformation and Cartography: Thematic Cartography for the Society*. Switzerland: Springer International Publishing
- Central Jakarta Statistic Center, 2017. *Indonesian Statistics 2017*. Central Jakarta: Central Statistics Agency
- Brewer, Cynthia A, others. 2013. ColorBrewer in Print: A Catalog of Color Schemes for Maps. *Cartography and Geographic Information Science Journal*, 30:1
- Brychtova, Alzbeta dan Coltekin, Arzu. 2015. Discriminating classes of sequential and qualitative colour schemes. *International Journal of Cartography*, 1:1
- Hatchard, Geoff. 2018. *Cartography on* https://www.e-education.psu.edu/natureofgeoinfo/c3_p17.html. Pennsylvania: The Pennsylvania State University
- Kraak, Menno-Jan dan Ormeling, Ferjan. 2007. *Kartografi dan Visualisasi Data Geospasial*. Translator: Sukendra Martha, etc. Yogyakarta: Gadjah Mada University Press
- Kurniati, Erna dan Rahardjo, Noorhadi. 2012. Evaluasi Metode Klasifikasi Dalam Pembuatan Peta Kepadatan Penduduk DIY Dengan Permukaan Statistik Dan Uji Proporsi. *Jurnal Bumi Indonesia*
- Rosenberg, Matt. 2019. *The Role of Colors on Maps on* <https://www.thoughtco.com/colors-on-maps-1435690>. Thought.co
- White, Travis M. ,et al. 2017. Trends and Issues in the Use of Quantitative Color Schemes in Refereed Journals. *Annals of the American Association of Geographers*, 107:4

Revised August 2019