INF4000

Zhibo Liu

Word count: 2957

Reg number:210164972

Figure 1

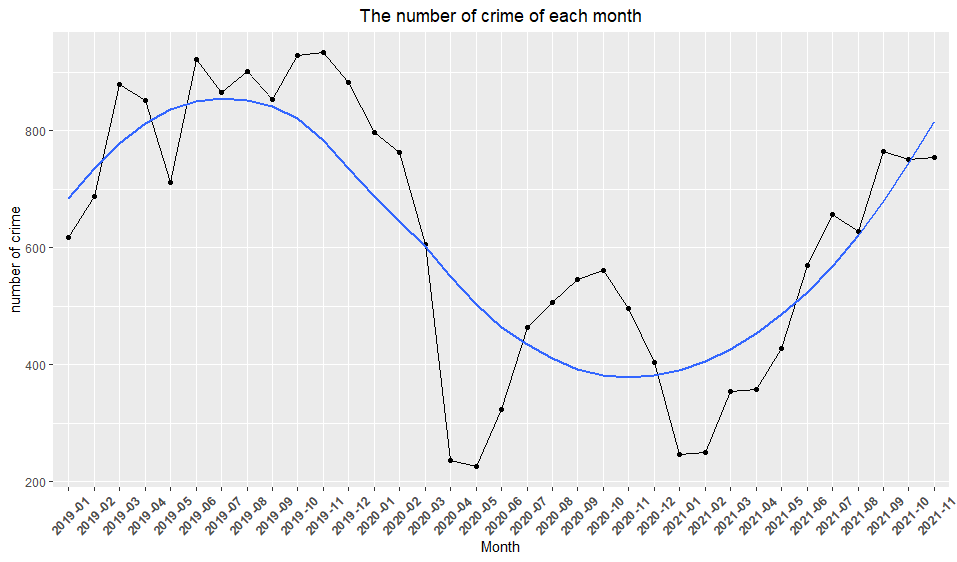


Figure 2

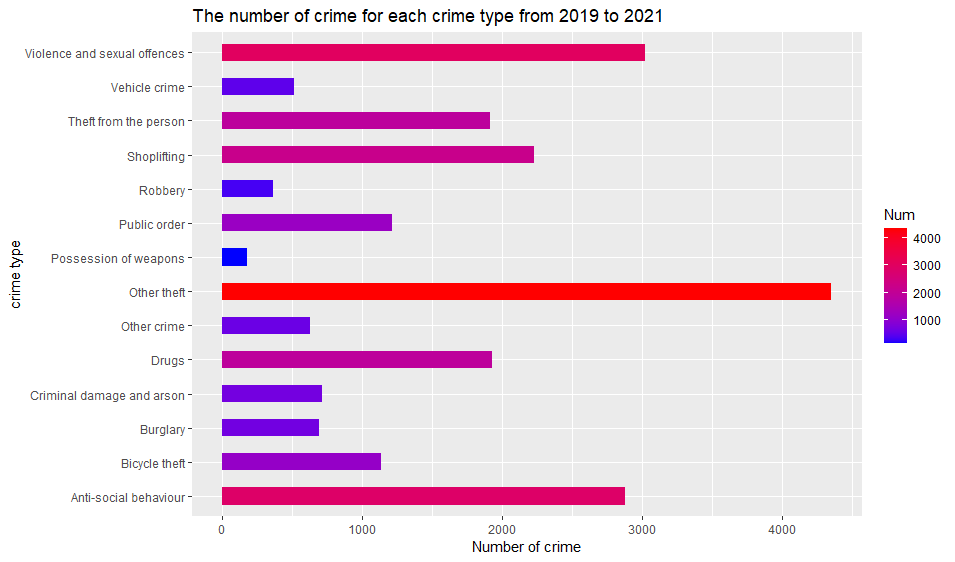


Figure 3

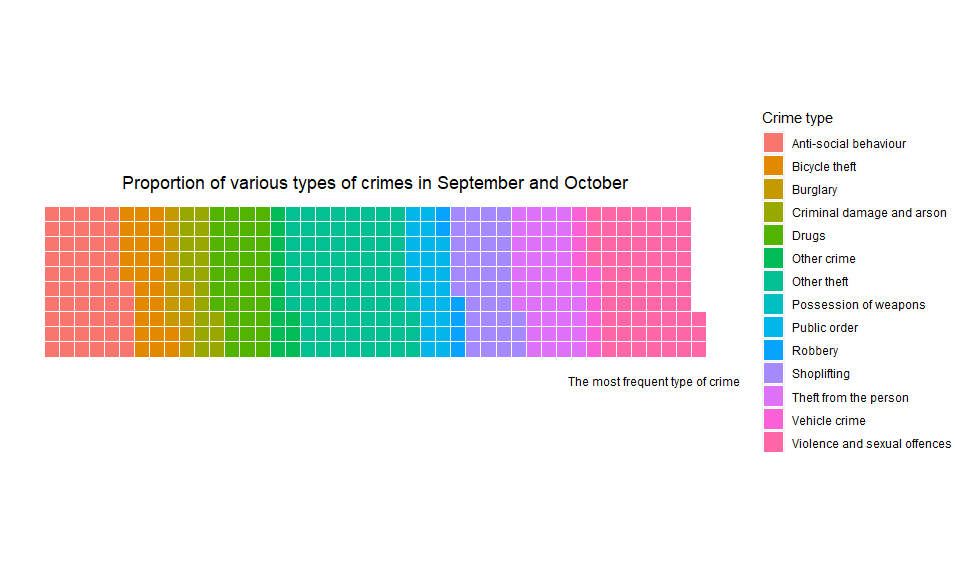
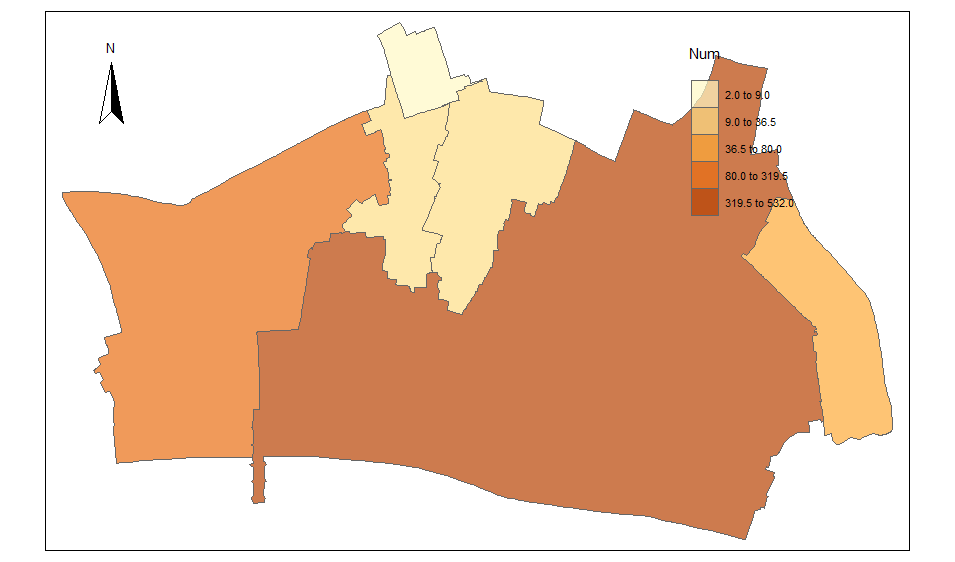


Figure 4 “Other theft” on the city of London map



**Content**

[1. Knowledge building 4](#_Toc19272)

[1.1 Introduction and aims 4](#_Toc6585)

[1.2 City of London dataset 4](#_Toc7225)

[1.3 Visualization 5](#_Toc299)

[2. Theoretical frameworks 5](#_Toc11056)

[2.1 Ask a question 5](#_Toc2049)

[2.2 Search for information 5](#_Toc28829)

[2.3 Structuring the data 6](#_Toc19083)

[2.4 Envision the answer 6](#_Toc3758)

[2.5 Represent the visualization 6](#_Toc25843)

[2.6 Tell a story 6](#_Toc28875)

[3. Accessibility 7](#_Toc32063)

[3.1 Vision 7](#_Toc29723)

[3.2 Content 7](#_Toc30287)

[4. Visualization choice 8](#_Toc18700)

[4.1 Line chart 8](#_Toc23483)

[4.2 bar chart 8](#_Toc32301)

[4.3 waffle chart 8](#_Toc21956)

[4.4 Map 9](#_Toc26343)

[5. Implications and Improvements 9](#_Toc11704)

[5.1 Ethical implications 9](#_Toc25803)

[5.2 Improvements 10](#_Toc28271)

[5.2.1 dataset improvement 10](#_Toc4247)

[5.2.2 visualization improvement 10](#_Toc18902)

[Reference : 11](#_Toc26656)

# Knowledge building

## 1.1 Introduction and aims

Safety has always been one of the basic requirements of mankind, not only the safety of the person but also the safety of the living environment. Aksoy (2017) proposed that one of the main causes of human insecurity is a crime in various areas of the city. Therefore, many articles in the past are studies of crime and its location, that is, space criminology (De Melo & Andresen, 2015; Vandeviver & Bernasco, 2017; Aksoy, 2017; Guerry, 1833; Malleson & Andresen, 2016). For example, as early as the 19th century Guerry (1833) drew a crime map in France. The conclusions of these past studies have consistently shown that crime is highly concentrated in a few areas of the city. Steenbeek and Weisburd (2015) stated that a large number of crimes occurred in one street in the city, but there was almost no crime in many other streets. Braga et al. (2011) added that when considering different types of crime, some types such as robbery and violence are more concentrated. For example, Wiles and Costello (2000) pointed out that all crime types in Sheffield are only three kilometres away from the concentration point. This means that all types of crime will occur in a clustered area in the city.

Therefore, the goal of this article is to first analyze the trend of crime data in the city of London, then analyze the most serious crime types, and finally display the distribution of the most serious crime types on the map of the city of London. The research can help urban residents avoid places that are defined as unsafe and reduce the possibility of crime.

## 1.2 City of London dataset

To achieve these goals, this study uses the city of London crime dataset from January 2019 to November 2021. Before selecting a suitable dataset, the author had an understanding of the crime situation in the UK. Moss (2012) indicated that the 2005 European crime survey results named the British capital London as the most crime-stricken city in Europe. And the crime situation in the city of London is the most serious among 33 London towns (CrimeRate, 2021). Therefore, this study sets the city of London as the spatial background of the study. In addition, the author downloads the City of London map from (https://borders.ukdataservice.ac.uk/bds.html). First, choose "England" and "2011 and later". Then select "English Lower Layer Super Output Areas in Boundaries,2011” and City of London and download the shapefile. To link the crime dataset with the LSOA file, the crime dataset also needs to contain the LSOA code. Therefore, the author selects "City of London Police” in “Forces” in (https://data.police.uk/data) obtained a data set including LSOA code and crime types from January 2019 to November 2021.

## 1.3 Visualization

In Figure 1, the crime data from January 2019 to November 2021 are displayed using a combination of line graphs and scatter graphs. It can be seen from the image that the number of crimes is constantly fluctuating, and it can be seen that the number of crimes is highest in September and October each year. From March to April 2020, the number of crimes has dropped significantly to very small. And this period happened to be the time when the British blockade began (Langton & Farrell, 2021). This also verifies the relationship between crime and mobility in the past (Mohler et al., 2020; Halford et al., 2020). Figure 2 shows the number of various types of crime from 2019 to 2021, and Figure 3 shows the ratio of each type of crime in September and October each year. Whether it is all the data or the two-month data, it can be clearly seen that "Other theft" has the largest number of crimes. This is consistent with the conclusion of CrimeRate (2021). Therefore, Figure 4 displays the "Other theft" data on the city of London map. It can be seen that except for the northern central area, the number of crimes in other areas is very high. This phenomenon validates the rational choice perspective and crime pattern theory (Bernasco and Ruiter 2014; Ruiter 2017): the location of the crime is not randomly selected, and the risk and benefit of arrest affect the choice of location. For example, the Museum of London in the area greatly increased the risk.

# Theoretical frameworks

## Ask a question

The ASSERT process starts by asking a question. In the introduction in the previous paragraph, the past research on crime was introduced. Research shows that the use of space criminology finds that crimes are concentrated in some areas. Therefore, this article focuses on the areas and months in which the most serious types of crimes may occur, because know which month is the highest number of crimes in the city of London? (Q1) What is the most serious type of crime? (Q2) And where did it happen? (Q3) It can help the audience avoid these unsafe places and help the city prevent crime in advance.

## 2.2 Search for information

To answer the question accurately, it is necessary to use meaningful data. First of all, various types of crimes and their numbers must be included in the data. Secondly, to visualize the data in the form of a map, a variable with a geographic location is also required. Finally, in addition to space, the months with the most serious crimes must be explored, so the data is also required to be time series. Finally, according to the requirement of meaningful data proposed by Few (2009), the author chose the City of London Police dataset (https://data.police.uk/data) after comparing several datasets. The data includes the number, location(LSOA code) of each type of crime from January 2019 to November 2021. Clear and consistent enough, high-volume time series data can help discover new information.

## 2.3 Structuring the data

After obtaining the data, the author cleans, constructs and organizes the data. Mainly use the Month variable and Crime type variable which is categorical in the data set. The Month variable is the time and continuous variable from January 2019 to November 2021. The author grouped the data by Month after clearing the missing values and summed up the total number of crimes per month. Then, filter out the data for September and October each year, group them by crime types, and summarize the numbers of each type. Finally, the data of other theft will be filtered and left joined with the shape data of the City of London to realize spatial visualization.

## 2.4 Envision the answer

Exploratory data analysis based on the processed data and the three questions. First of all, for question Q1, we need to analyze the changes in the number of crimes over time. So use a line chart and scatter chart to analyze the relationship between time and the number of crimes, and add a smoothed line to observe the trend. Then for Q2, what needs to be analyzed is the various types of crimes and the number and ratio of categorical, so the bar chart and the waffle chart are used respectively. Finally, Q3 needs to find the location with the most crimes, so the method of spatial data visualization is used.

## 2.5 Represent the visualization

In Figure 1, the x coordinate is the continuous data of the month, and the y coordinate is the number of crimes, so Cartesian coordinates are used. Select 45 degrees for the x coordinate and adjust the title to the middle. In Figure 2, the author adjusts the width of the histogram to 0.5 and fills in the number of crimes, and adjusts the colour range to be red for high and blue for low and add a title. Figure 3 uses the waffle chart to show the ratio. It removes the content and space of all extension axis lines, tick marks, and tick labels for neatness. In Figure 4, the transparency is adjusted to 0.7 and the map compass is added in the upper left corner.

## 2.6 Tell a story

The four visualizations in this article coherently tell the audience about crime in the city of London. Figure 1 shows the most serious crimes in September and October to the percentages of each crime type shown in Figures 2 and 3. Then Figure 4 shows areas with serious crimes. Three questions were answered in succession to let the audience understand how to avoid unsafe areas.

# Accessibility

When designing a visualization, it is necessary to consider accessibility. This is because the abilities of people are diverse, and there are differences in the ability to acquire knowledge or the ability to perceive (Ferster, 2012; Matzen et al., 2018; Kirk, 2016). So when making data visualization, It is necessary to consider whether the audience will not be able to properly understand the meaning due to lack of ability or knowledge gap.

## 3.1 Vision

Vision is the main perception that humans accept knowledge, and colour plays a central role in data visualization (Lin et al., 2013; Atkinson, 2002). Appropriate colour tone helps us to identify and distinguish categories. So when discussing the accessibility of visualization. However, Ferster (2012) pointed out that human vision ranges from full marks to blindness, which also includes some form of colour blindness. Inappropriate colour usage may misunderstand people with colour blindness and affect the accessibility of visualization. So when discussing the accessibility of visualization, whether the audience can see the conclusion is very important. For example, in this study, Figure 1 shows the trend of the number of crimes over time. Simply observing the line chart may prevent some viewers from observing the overall trend. So the author adds a smoothed curve to help understand the general trend. The wafer diagram in figure 3 uses multiple colours to indicate different types of crimes. However, people with colour blindness or colour weakness may not be able to distinguish them, so they cannot understand the proportion of each type of crime, which greatly reduces accessibility. However, Figure 2, which also uses multiple colours, performs well. This is because the grid in the background of the histogram allows colour blindness to compare which type has more numbers without using colours. The colours used in Figure 4 may not be obvious, and some viewers may not be able to recognize them.

## 3.2 Content

To improve accessibility, past research indicated that other senses such as subtitles, titles or coordinates can be used to help the audience better understand visualization (Mueller, 2003; Kirk, 2016). Mueller (2003) pointed out that high accessibility in data visualization requires clear text, easy to distinguish background and content. For example, in this article, the author in Figure 1 and Figure 2 adds labels for the horizontal and vertical coordinates and chooses the year and month of the x-axis at 45 degrees to avoid covering each other. However, the size of the ordinate in Figures 1 and 2 is not appropriate, which may cause the audience to be unsuitable for reading when using a smartphone. The author in Figure 3 deleted all the content and space of the extension axis, tick marks, and tick labels to allow the audience to avoid unnecessary interference. Figure 4 shows the area where the crime occurred, but the map does not show the location of some construction facilities. This has become a knowledge gap for the audience who do not understand the city of London. They cannot understand the characteristics of areas with low crime rates.

# Visualization choice

Choosing the right visualization method needs to start with the problem that the visualization wants to solve. Select the appropriate method to visualize in the taxonomy of data visualization through the goal of visualization (Kirk, 2012; Borkin et al., 2013).

## Line chart

The first visualization is to express the trend of the number of crimes over time, so the "Showing changes over time" method in taxonomy is selected. The chart that the author chooses to use is a line chart or a sparklines chart. The advantage of the line chart is that the range displayed by time is very complete and can clearly show the trend (Kirk, 2012). However, Trautner and Bruckner (2021) stated that the weakness is that it may be difficult to distinguish and explain when there are more lines. For the sparklines diagram, its advantage can show complex visualization in a small space (Kirk, 2012). But the disadvantage is that segmented visualization cannot let the audience understand the overall trend at a glance. Therefore, to clearly show the trend over time, the author uses a line chart.

## bar chart

The purpose of the second visualization is to compare the number of various types of crimes, so choose the "Comparing categories" method. Possible visualization methods are bar graphs and word clouds. The biggest advantage of the bar chart is that it is easier for the audience to understand (Cleveland & McGill, 1984). Although the word cloud can also allow the audience to clearly understand the largest number of variables, the font size is the number of variables (Kirk, 2012). But its disadvantage is that the volume of a small number of variables is too small, which may be inconvenient to read. The goal of Figure 2 is to not only let the audience know which crime types are the most, but also let them know what types of crimes are. Therefore, the bar chart is used to clearly display all variables and quantities.

## 4.3 waffle chart

The third purpose is to compare the ratio of crime types. The author can choose either a pie chart or a waffle chart in the "part-to-whole relationships" method. The pie chart visually displays the ratio of each part, but it will be extra complicated when there are too many categories. The waffle chart is similar to the square pie chart, but the study of Borkin et al. (2013) found that the unique and uncommon figures of arguing are more memorable than ordinary figures (circles, lines). Thus, the author uses a more impressive waffle chart.

## 4.4 Map

The purpose of the last visualization is to find the most severe areas, and there is no doubt that spatial visualization is used here. To indicate the most severe areas, the data is combined with the census LSOA data for visualization. The map can show areas with different crime levels in different colours. But the disadvantage is obvious. There are no more details about the building on the map, which confuses some audiences and reduces accessibility. An alternative method is to use a dot plot map combined with spatial data representation (TopoJSON, such as Google Maps). This can show more details on the map and may help understand why crimes happen frequently.

# Implications and Improvements

## 5.1 Ethical implications

Visualization is an influential technology that can magnify the views of its designers, and can also magnify the ethical problems caused by prejudice (Hepworth, 2020; O'Brien, 2017). Many scholars say that designers should consider that visualization will cause The ethical influences of the people and need to be responsible for them (Monmonier, 1991; Harley, 1990). Monmonier (1991) pointed out that ethical considerations should start from the collection of data until the end of visualization. Unintentional omissions in the process may cause ethical problems. The goals of visualization in this research include exploring trends, comparing crime types and identifying areas. All three visualization tasks can produce bias (Brehmer & Munzner 2013). The final map visualization of this study contains less information and may cause viewers to misunderstand the crime situation in the region.

In the past, there were also examples of prejudice caused by map visualization. For example, Hepworth and Church (2018) compared the Racial Terror Lynchings map (Visualization 1) and the Map of White Supremacy Mob Violence (Visualization 2) in their research on two visual maps of California lynchings. . The biggest difference between two visualizations is difference in the use of data. Visualization 2 includes African or Latino Americans, Aboriginals, and other races. However, the data in Visualization 1 only includes data on lynching by African Americans, so the results of the visualization introduce bias. It uses colour to draw attention to the southeastern United States, making viewers think that lynchings are rare in California. This also broke the California prison propaganda and people's understanding of the lynching situation.

## 5.2 Improvements

In summary, the four data visualizations in this article need to be changed and improved.

### 5.2.1 dataset improvement

In terms of data sets, the city of London crime dataset from January 2019 to November 2021 used in this study may not be sufficient. If the author chooses a larger time range, using 5 or more years of the city of London crime data can better reflect the trend of crime over time. More records of various types of crimes can also make the conclusion of unsafe places in cities more convincing.

### 5.2.2 visualization improvement

The visual selection and visual design of the four visualizations in this article can also be improved. These improvements can also increase their accessibility. For the second and third visualizations, the author used multiple colours to distinguish each crime type. The use of multiple colours may be very unfriendly to people with colour blindness or colour weakness. Therefore, there may be two improvements here: the first is to use a colour spectrum that can be recognized by people with colour blindness, such as a colour spectrum designed for red-green colour blindness. The second chapter is to add text labels to the visualization to help understand. The improvement to Figure 4 is to use a dot plot map combined with the spatial data representation of Google Maps. This approach can complement the location of buildings in the city of London and may improve accessibility and improve the understanding and memory of unsafe locations.

# Reference :

Aksoy. (2017). Geography of Crime and Its Relation to Location: The City of Balıkesir (Turkey). IOP Conference Series. Materials Science and Engineering, 245(7), 72012. <https://doi.org/10.1088/1757-899X/245/7/072012>

Atkinson, J. (2002). “The Developing Visual Brain” Oxford, UK.: Oxford University Press.

Bernasco, Wim, Ruiter, Stijn. 2014. “Crime Location Choice.” Pp. 691–99 in Encyclopedia of Criminology and Criminal Justice, edited by Bruinsma, G., Weisburd, D.. New York: Springer.

Borkin, Vo, A. A., Bylinskii, Z., Isola, P., Sunkavalli, S., Oliva, A., & Pfister, H. (2013). What Makes a Visualization Memorable? IEEE Transactions on Visualization and Computer Graphics, 19(12), 2306–2315. https://doi.org/10.1109/TVCG.2013.234

Braga, A. A., Hureau, D. M., & Papachristos, A. V. (2011). The relevance of micro places to citywide robbery trends: A longitudinal analysis of robbery incidents at street corners and block faces in Boston. *Journal of Research in Crime and Delinquency*, *48*(1), 7-32.

Brehmer, M., & Munzner, T. (2013). A multi-level typology of abstract visualization tasks. IEEE transactions on visualization and computer graphics, 19(12), 2376-2385.

Cleveland, & McGill, R. (1984). Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. Journal of the American Statistical Association, 79(387), 531–554. <https://doi.org/10.1080/01621459.1984.10478080>

CrimeRate. (2021). Crime and Safety in City of London. Retrieve from: <https://crimerate.co.uk/london/city-of-london>

De Melo, Matias, L. F., & Andresen, M. A. (2015). Crime concentrations and similarities in spatial crime patterns inaBrazilian context. Applied Geography (Sevenoaks), 62, 314–324. https://doi.org/10.1016/j.apgeog.2015.05.012

Ferster, B. (2012). Interactive visualization : Insight through inquiry. MIT Press.

Few, S. (2009). Now you see it: simple visualization techniques for quantitative analysis (No. Sirsi) i9780970601988).

Halford, E., Dixon, A., Farrell, G., Malleson, N., & Tilley, N. (2020). Crime and coronavirus: Social distancing, lockdown, and the mobility elasticity of crime. *Crime Science,* *9*(1), 1–12.

Harley J. B.: Cartography, ethics and social theory. Cartographica: The International Journal for Geographic Information and Geovisualization 27, 2 (1990), 1–23. 2

Hepworth. (2020). Make Me Care: Ethical Visualization for Impact in the Sciences and Data Sciences. In Design, User Experience, and Usability. Interaction Design (Vol. 12200, pp. 385–404). Springer International Publishing. <https://doi.org/10.1007/978-3-030-49713-2_27>

Katherine Hepworth, & Christopher Church. (2018). Racism in the Machine: Visualization Ethics in Digital Humanities Projects. Digital Humanities Quarterly, 12(4).

Kirk, A. (2016). Data visualisation: A handbook for data driven design. Sage.

Kirk. (2012). Data visualization [electronic resource] : a successful design process ; a structured design approach to equip you with the knowledge of how to successfully accomplish any data visualization challenge efficiently and effectively (1st edition.). Packt Pub.

Langton, Dixon, A., & Farrell, G. (2021). Six months in: pandemic crime trends in England and Wales. Crime Science, 10(1), 6–6. <https://doi.org/10.1186/s40163-021-00142-z>

Lin, Fortuna, J., Kulkarni, C., Stone, M., & Heer, J. (2013). Selecting Semantically-Resonant Colors for Data Visualization. Computer Graphics Forum, 32(3pt4), 401–410. <https://doi.org/10.1111/cgf.12127>

Matzen, Haass, M. J., Divis, K. M., Zhiyuan Wang, & Wilson, A. T. (2018). Data Visualization Saliency Model: A Tool for Evaluating Abstract Data Visualizations. IEEE Transactions on Visualization and Computer Graphics, 24(1), 563–573. <https://doi.org/10.1109/TVCG.2017.2743939>

Malleson, & Andresen, M. A. (2016). Exploring the impact of ambient population measures on London crime hotspots. Journal of Criminal Justice, 46, 52–63. <https://doi.org/10.1016/j.jcrimjus.2016.03.002>

Mohler, G., Bertozzi, A. L., Carter, J., Short, M. B., Sledge, D., Tita, G. E., Uchida, C. D., & Brantingham, P. J. (2020). Impact of social distancing during covid-19 pandemic on crime in Los Angeles and Indianapolis. *Journal of Criminal Justice,* *68,* 101692.

Moss. (2012). Property Crime in London, 1850-Present. Contemporary British History, 26(3), 437–439. <https://doi.org/10.1080/13619462.2012.701909>

Monmonier M.: Ethics and map design: Six strategies for confronting the traditional one-map solution. Cartographic Perspectives, 10 (1991), 3–8. 2

Mueller, J. (2003). Accessibility for everybody: Understanding the section 508 accessibility requirements . New York: Apress.

O'Brien, S. (2017). Interpretations of data in ethical vs. unethical data visualizations (Order No. 10681043). Available from ProQuest Dissertations & Theses A&I. (1976992451). Retrieved from <https://www.proquest.com/dissertations-theses/interpretations-data-ethical-vs-unethical/docview/1976992451/se-2?accountid=13828>

Ruiter, Stijn . 2017. “Crime Location Choice: State-of-the-art and Avenues for Future Research.” Pp. 398–420 in The Oxford Handbook on Offender Decision Making, edited by Bernasco, W., Elffers, H., Van Gelder, J.-L.. Oxford, UK: Oxford University Press.

Steenbeek, W., & Weisburd, D. L. (2015). Where the action is in Crime? An examination of variability of crime across different spatial units in the Hague, 2001-2009. Journal of Quantitative Criminology, 1-21.http://dx.doi.org/10.1007/s10940-015-9276-3

Trautner, & Bruckner, S. (2021). Line Weaver: Importance‐Driven Order Enhanced Rendering of Dense Line Charts. Computer Graphics Forum, 40(3), 399–410. https://doi.org/10.1111/cgf.14316

Vandeviver, & Bernasco, W. (2017). The geography of crime and crime control. Applied Geography (Sevenoaks), 86(September), 220–225. https://doi.org/10.1016/j.apgeog.2017.08.012

Guerry, A. M. (1833). Essai sur la Statistique Morale de la France. Paris: Crochard

Wiles, P., & Costello, A. (2000). The'road to nowhere': the evidence for travelling criminals (Vol. 207). London: Home Office.