An Artistic Data Visualization of Pollution at Earlham College

Sara Salloum
Department of Computer Science
Earlham College
Richmond, Indiana
ssallo15@earlham.edu

ABSTRACT

In this paper, I propose ECAir, a system to provide artistic visualization of air pollution data. Air pollution is an ever-growing concern not only for our health but also for the planet as it contributes to global warming [15]. One way to spread awareness about it, is to create informative data visualizations. Data visualization plays a vital role in understanding existing data, as well as in presenting the findings derived from the data. Data visualization is gaining increasing importance in todayâĂŹs world of big data and data science. However, most of the efforts in data visualization are focused on how efficiently we can represent the information and not the aesthetics of the visualization. This study aims at providing a system for the artistic visualization of air pollution data. ECAir employs Arduino devices along with an MQ-135 sensor to collect air pollution data and then uses a novel algorithm to create Mondrian-style visualization. I have tested my system by collecting data on the Earlham College campus.

KEYWORDS

Data Visualization, Air Pollution, Arduino, Informative Art

ACM Reference Format:
Sara Salloum. 2019. An Artistic Data Visualization of Pollution at Earlham College. In *Proceedings of (CS488 Senior Capstone).* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Visualization of information has been around for hundreds, even thousands of years, depending on how we define the term. Among the earliest known examples are the coordinates used by ancient Egyptian surveyors in 200 BC [7]. However, more sophisticated techniques were developed in the 16th century for recording and visualizing data, such as using tables to record physical measurements [8]. Since then, a lot of progress has been made on making it easier and more accessible to create representations of data that can show patterns and trends, and technology has played a big role in that. It is useful in the world of natural science, economics, business, and almost every other field. Governments use it to show the results of domestic surveillance, and academics use it in presentations, papers, and posters [8]. Modern technology and the widespread availability of the internet have provided easy access to data visualization tools, such as Microsoft Excel or Google Developers, where anyone can plug in data and get a chart, a map or another graphic portrayal of quantitative information. However, although very widespread, data visualizations often neglect the aesthetics of the representation and solely focus on the information one can get from the graphs. Typically, informative visualizations live in a file on a computer [19]. That information is not always displayed for the viewer to see; instead, they have to seek it out. Visualizations would be significantly more informative to the public if they were permanently displayed somewhere, the way that artwork typically is. Aesthetics is important here, because putting up pieces of paper with data visualizations that look identical will eventually become a dull feature of the environment, and they will no longer get the attention they were put up to draw. Using art to create unique and interesting data visualizations, that fit into the environment they are placed in and add to it, will draw the viewers to them, drive them to look closer and ask questions, increasing their awareness [4].

Air pollution is an important problem that we have the ability to obtain data on, but lack awareness about. It is poisonous and causes many health problems, and climate change. Yet, there is little awareness of how much damage decisions we make daily cause [10]. ECAir is a tool that collects pollution data on the campus of Earlham College, and displays real-time pollution data in an artistic manner designed to help realize the pollution surrounding us. The data is collected using an Arduino board and an air quality sensor. In order to make the visualization a piece of art, it will follow the style of the artist Mondrian (Figure 1), because of the nature of his work. His artistic style makes use of geometric forms and primary colors, which started out as a way to represent the world without depicting it exactly as it is, but rather an abstract version of it. This kind of system would be especially useful in larger cities, where the visualization can be displayed in public areas such as train stations or bus stops. The display of such information in real time will improve awareness, which would hopefully result in better decision-making that is less harmful to the environment.

Realizing the importance of artistic features in drawing interest and spreading knowledge is important for the world of science. Additionally, an aesthetically-pleasing visualization has the potential to be provocative and engaging [17]. With regards to pollution, if the viewer of the visualization is more provoked, they may, for example, choose not to drive when it is not necessary. The goal of this research is to create a data visualization of air pollution that fits within its environment, while helping raise awareness. This paper will first look at the background in data visualizations and aesthetics, then explain the methodology and provide the results and an evaluation.

This paper contributes a system, ECAir, which uses Arduino and Python's tkinter to collect air pollution data and display it in a Mondrian-style visualizations. This work pushes forward the fields of artistic data visualization and Arduino air quality monitors

CS488 Senior Capstone, May 2019, Earlham College

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM.
https://doi.org/10.1145/nnnnnnn.nnnnnnn

by combining the two. This provides an accessible approach for anyone to build their own monitor, the data collected by which can be displayed as an artistic visualization. The same approach can then be used to display any data that can be collected using an Arduino sensor, and a home could be decorated with informative art.

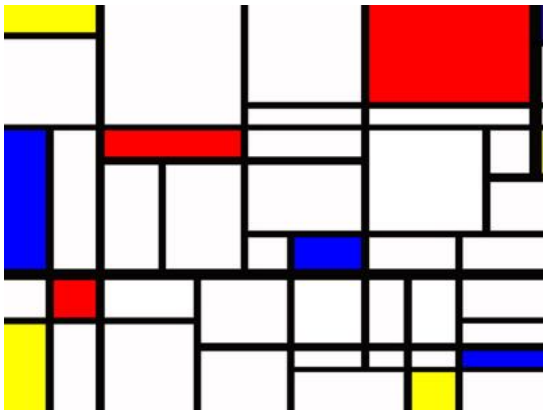

Figure 1: An example of Mondrian's style [1]

2 BACKGROUND

The effects of pollution on our daily lives are often neglected, as well as the importance of aesthetics in data visualizations. Data visualizations are rarely made to be a part of an environment for an extended period of time. When they are, they ought to fit into that environment and add something to it to draw our curiosity, rather than be something unpleasant to look at. This work is informed by three existing lines of research: aesthetics in data visualizations, pollution awareness, and the use of Arduino board to measure air quality.

2.1 Data Visualization and Aesthetics

The visualization of data in an aesthetically pleasing manner is not uncommon in the world of art, where collections of data can be taken and interpreted by artists in abstract ways, for the purpose of making art, rather than displaying quantitative information [21]. However, utilizing art to draw viewers' attention to data requires a different kind of creativity. In creating artistic visualizations of data, one must consider more details, because the main purpose of the piece remains the conveying of information. As far as I have been able to determine, there are not many studies that specifically look at the correlation between aesthetics and user attention, however, Cawthon et al.'s work does explore this. It finds that the attractive visualizations draw the users to look more closely and make them show a higher level of patience. This was measured through presenting 11 different visualizations of the same data to viewers. The visualizations varied from traditional TreeMaps and network diagrams, to more creative, floral-inspired representations. They also had constants in all techniques, such as color, size and text [4]. This is very significant for this project, because a goal of it is to draw users' attention to pollution levels.

Lau et al. identify a visualization field called *information aesthetics*, which forms a link between information visualization and visualization art. It combines different aspects of aesthetics, data and interaction to create something that allows for more profound, long-term impressions on the viewers. The model provides an opportunity for researchers and artists to develop design guidelines for information aesthetics. The goal is to create visualizations that utilize art, without endangering the trustworthiness of the information presented [14]. As an artist and a researcher, this is what I am trying accomplish with this project, creating a visualization which abstracts details but provides sufficient information to spread awareness.

2.2 Pollution and Awareness

Bickerstaff and Walker offer a current picture of the ways in which residents think about the problem of urban air pollution. They state that if a durable improvement is to be achieved, a shift in personal behaviour, in particular the decisions that people make about transport choice, toward a more sustainable way of life is fundamental. This would be remarkably relevant on Earlham's campus, where both students and faculty choose to drive from one building to another, when the longest possible walk would not exceed ten minutes. Moreover, they found that awareness of poor air quality is far from universal. For most people it was a diverse array of localized, physical and social encounters with air pollution that were important in the development of perception. If it was not witnessed first-hand, it did not matter nearly as much. It also showed that the influence of external sources of information on the public is minimal. The authors suggested that an argument can be developed for more visible approaches to public education and information such as air quality messages in public areas. [10]. This goes to show how important a visualization is for spreading awareness about different causes, and it can especially solve a problem like the one described by Bickerstaff and Walker. It may be common knowledge that pollution is bad, but it is far from being something that we are concerned with as much as we should be, considering the damage it causes to our health and the environment.

2.3 Measuring Air Quality

Kim and Paulos created a tool called InAir, used for measuring, visualizing, and learning about indoor air quality. InAir creates a visualization of the data it collects, but the visualization is purely informative, and has nothing to offer in terms of aesthetics and fitting into its environment [12]. Devarakonda et al. and Ali et al. both also created Arduino devices to measure air quality [5] [2]. Arduino is an open-source hardware and software company, which produces different types of Arduino boards for affordable prices [3]. These boards are essentially mini computers. They have a microcontroller and a certain amount of memory, as well as pins to connect different sensors. They are typically used to create a cost-effective, accessible and easy-to-build devices. Arduino provides an integrated development environment (IDE) that is capable of running on all major operating systems and has support for a simplified C/C++ programming language. ECAir uses an air quality sensor called the MQ-135, which measures levels of NH3, C6H6,

NOx, smoke and CO2 in the air. Various studies have found a significant variation of outdoor air pollution at a small scale, but in the cases of those studies, the small scale was close to a highway or an interstate, where traffic is much more condensed, with large vehicles being a major part of it [22] [6]. In this research, there is a much less significant source of vehicle emissions, that being a parking lot.

This project incorporates advancements in aesthetics in data visualizations, air pollution and the use of Arduino boards. By creating a tool that displays an artistic real-time pollution data visualization on campus using Arduino, it will hopefully, over time, spread awareness about pollution, and be used on a larger scale.

3 METHODOLOGY

The development of ECAir can be divided into three main modules: the tools required for collecting the data, data collection, and the visualization. This section describes each one of these steps in more detail.

3.1 Tools

The tool made to collect the data is built with an Arduino Uno board, using an SD card shield, which provides the ability to collect data without being connected to a computer. The sensor used for measuring air quality is the MQ-135, which provides one number combining the concentrations of NH_3 C_6H_6, NO_x, smoke and CO_2 in the air [9]. If one of them increases, the value provided increases, but it does not give specific information about each one. It is a simple module that gives a general indication of the air quality. It is adequate for the purposes of this project, because the project is focused on the visualization, rather than the study of exact air particles. Another reason for choosing a sensor that provides only one number is to abstract details but provide sufficient information, and that works perfectly with the style of Mondrian. In the preliminary data collection, ECAir does not use a GPS sensor because the points at which the data is collected are fixed. However, one can be added if the points were dynamic (e.g. if a drone flew the device around campus). Another situation in which it would be useful is to use the coordinates to create the mapped visualization. This was the original idea for this project, which was to use Basemap for the visualization, but that works best for larger maps.

3.2 Data Collection

The data is collected around entry points of campus buildings, for example, the front and back doors of Barrett Hall. This is because most buildings have one entry point close to a parking lot, and one close to the heart of campus. The heart of campus is a green area (G in figure 2), in which motor vehicles are not permitted. We would expect that proximity to a parking lot would result in worse air quality, because motor vehicle emissions are a major contributor to air pollution [20].

All the data collection points are shown in figure 2. A is facing a parking lot, B is the entryway opposite of that. C, D, E and F are the same. The buildings were chosen to represent each of the three types of buildings on campus. One academic, at points E and F, one dorm, at points C and D, and a building with food/health facilities at points A and B. Next, the data files are taken from the SD card and used to fill the colors of the rectangles in the Mondrian visualization, which is the most refined version of the data collected. The use of a database, such as PostgreSQL, would facilitate making this a real-time tool, but it is not in the scope of this project.

The way the color is chosen is determined by the range of the data collected. The color for every rectangle or location is therefore relative to the six points at which air quality was measured. This decision was made because ECAir cares about the points it is concerned with. Are there areas on the campus of Earlham College that have worse air quality than others? The question is within the campus, rather than checking for air quality as compared to a general standard of what is good or bad. If the value at point A is in the highest third of the data, it will be filled red, if it is in the middle third, it will be filled yellow, and if it is in the lowest third it will be filled with blue. The color choice was based on traffic lights, where yellow means slow down, or in this case, beware. Red is stop, which in this case is this area has poor air quality. This leaves blue to represent the cleanest air out of the 6 areas.

Figure 2: Data Collection Points

As part of preliminary data gathering, a single device will be used to collect the data in one day, to minimize variation in temperature, humidity and other weather-related factors that can influence air quality. Additionally, the representation is based on a value that is the average of multiple data entries collected every 3 seconds over two minutes in each location. 3 seconds was found to be an appropriate time after some experimentation, because it gives enough time for the data printing to go smoothly. However, in the long term and as part of the future work, the data collection will be a continuous process, as this will be a real time air monitoring device.

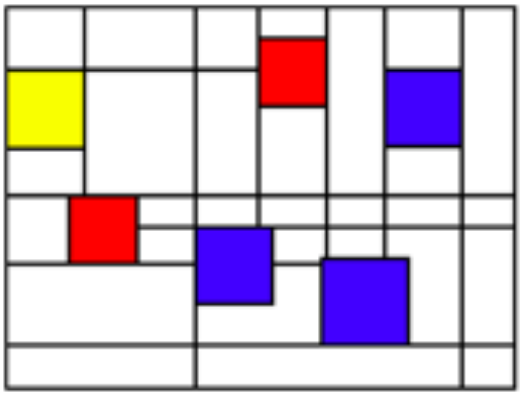

Figure 3: A Mondrian-inspired visualization of the current weather in six cities around the world [14]

3.3 Data Visualization

The visualization is a mapping of the data onto the campus of Earlham College, where the data collection points are the colorful rectangles in a Mondrian painting. Red will indicate a high value returned by the sensor, yellow will be medium and blue will be low. In this case, low is good and high is bad. The lengths of the lines represents the distance between campus buildings. The third element in a Mondrian painting, the size of the rectangles, is determined by the size of area where the data is being collected. For example, a rectangle is be drawn around the Barrett Hall parking lot (Figure 4). The idea to use an artist's style is inspired by Skog et al.'s use of Mondrian's style to create informative art. A visualization of the current (real-time) weather in six cities around the world (Figure 3) represented each city by a colored square, the temperature by the size of the square and weather conditions by the color.

Tkinter, a Python GUI library, is used to create the visualization on top of a map of Earlham's campus. Lines were drawn over a faded map of the campus, and the positions of the lines were determined by the rectangles they needed to create with their intersection around the points of data collection. The outcome, for the purposes of this project, is an image, but again, in future work, it would be make into a real-time map that is constantly changing based on the data collected.

4 RESUTLS/EVALUATION

The resulting visualization is the locations in Figure 2 filled in on Figure 4 to create Figure 5, or a variation of it.

The results show that all points but A fall into the same category, which is the best air quality in the range. However, it is likely that the data collected at point A was affected by the initialization of the MQ-135 sensor, which starts with a high value and takes some time to plateau. If this is the case, the entirety of the data is at least partially affected. However, overall, there seems to be little variation across the areas where air quality was measured between 16:00 and 17:00. It was anticipated that at that time, there would be many moving vehicles because it is the time frame in which many people leave the campus. However, there were few moving cars, which may be why there was no significant difference, especially

Figure 4: The empty grid for the visualization

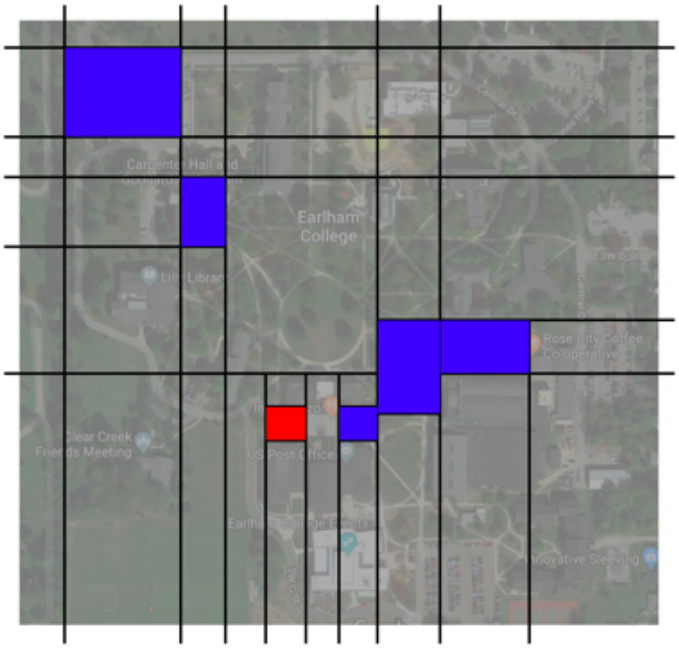

Figure 5: The Mondrian-inspired visualization of air quality on the campus of Earlham College

considering the studies mentioned earlier that found a significant variation over small areas that are close to a highway.

Although this might make ECAir less useful at Earlham College, it does not hinder it from working in larger scale, where the difference in air quality is significant enough to show up. Nothing in

the setup of the device prevents it from being used on a large scale, but it would, of course, need to be tested for that scale, which is beyond the scope of this project.

I intend to collect more data between 8-9 am to compare to the current data, with a more purposeful attempt to avoid the sensor initialization issue. I attempted to collect the data earlier to it can be in this version of the paper, but I had a technical issue that I later fixed and should be able to have it very soon.

5 RELATED WORK

The related work can be split into two categories: aesthetics in data visualizations, and the use of Arduino board to measure air quality. What distinguishes this project is that it combines those lines of research to create an Arduino data visualization tool for air pollution.

5.1 Data Visualization

Cawthon and Moere study the effect of aesthetics on the usability of data visualizations. Through their study of creating different visualizations, presenting them to an audience and getting feedback, they find that viewers show a higher level of patience when looking at attractive visualizations. This find is crucial to this project, because that is exactly what is needed to spread awareness about air pollution.

Kosara proposes a classification of several types of information visualization based on aesthetic criteria, introducing the notions of artistic and pragmatic visualizations. The goal of pragmatic visualization is for the user to be able to thoroughly understand and analyze the data. On the other hand, the goal of artistic visualization is to communicate a concern. The data need not be deeply understood, but it should still provide information based on real data. As artistic visualization is what this project creates, in order to communicate a concern regarding air quality.

Skog et al create a visualization of bus departure time data based on Mondrian's style, and argue that information visualization can be employed to make dynamic data more useful and appealing. This project is similar to mine in that it uses Mondrian's style to convey information, but different in that its purpose is not to spread awareness.

5.2 Arduino and Air Quality

Ali et al. built an Arduino air quality system, which has more specifications, such humidity and temperature monitoring, which would be a great future work for this project. In this case, simplicity is key, and the more elements we add, the harder it becomes to abstract away details. The techniques used in this system are informative, but much of it would not be necessary in building a tool which has the purpose of spreading awareness. The human brain is only capable of holding so much information at a time, and and giving viewers the least amount of informative data they need is the best way to make them remember it.

Kim and Paulos create a simple air quality measurement system, called inAir, which provides a visualization of the data. The goal of that work is indeed to spread awareness and provide a simple visualization for the home, but that is (1) meant for use in a home (2) provides a pragmatic visualization, which does not catch the eye of the viewer as well as an artistic visualization would, especially in a public area.

6 FUTURE WORK

With an addition of Wi-Fi enabled Arduino boards, ECAir can be set up to provide an aesthetically pleasing, real-time visualization for pollution data in large cities, which it is mostly set up to be able to do. To make the visualization even more inviting, connecting it to a website could allow access to the visualizations, and possible interactive features. An example of those features is the ability to check what the visualization looks like on a certain day or at a certain time. It can also be changed to make the color choice absolute instead of relative, which can also be an interactive website option.

7 CONCLUSION

ECAir is a preliminary device for one that can be more elaborate, but use the same basic principles. Using affordable and accessible tools to measure air pollution, or any variable, and use an algorithm to convert that data into an aesthetically pleasing data visualization. Using Wi-Fi, a database, and well-packaged devices, this could be made into a real-time tool which can be used in large cities and areas where air pollution is a true problem. This will invite the viewers to look closer, and will add an artistic element to the environment, in the form of informative art.

8 ACKNOWLEDGEMENTS

I would like to thank Dr. Ajit Chavan for his detailed feedback and help giving this project a better scope. Additionally, I want to thank Dr. Charlie Peck for taking the time to answer my questions and helping me get a better understanding of the hardware and software that will be used. I also want to thank Dr. David Barbella for constantly providing feedback and guidance throughout this research. Finally, I would like to express my special thanks to my fellow computer science students, who helped me in numerous ways.

REFERENCES

[1] [n. d.]. Mondrian's 100 year legacy. https://musartboutique.com/mondrian-art-hundred-year-legacy/. Accessed: 2019-03-29.

[2] Akram Syed Ali, Zachary Zanzinger, Deion Debose, and Brent Stephens. 2016. Open Source Building Science Sensors (OSBSS): A low-cost Arduino-based platform for long-term indoor environmental data collection. *11th International Conference Information Visualization* 100 (2016), 12.

[3] Yusuf Abdullahi Badamasi. 2014. The working principle of an Arduino. In *2014 11th International Conference on Electronics, Computer and Computation (ICECCO)*. IEEE, 1–4.

[4] Nick Cawthon and Andrew Vande Moere. 2007. The effect of aesthetic on the usability of data visualization. *11th International Conference Information Visualization* (July 2007), 11. https://doi.org/10.1109/IV.2007.147

[5] Srinivas Devarakonda, Parveen Sevusu, Hongzhang Liu, Ruilin Liu, Liviu Iftode, and Badri Nath. 2013. Real-time air quality monitoring through mobile sensing in metropolitan areas. In *Proceedings of the 2nd ACM SIGKDD international workshop on urban computing*. ACM, New York, NY, USA, 15. https://doi.org/10.1145/2505821.2505834

[6] Marieke B. Dijkema, Ulrike Gehring, Rob T. van Strien, Saskia C. van der Zee, Paul Fischer, Gerard Hoek, and Bert Brunekreef. 2010. A comparison of different approaches to estimate small-scale spatial variation in outdoor NO2 concentrations. *Environmental health perspectives* 119, 5 (2010), 670–675.

[7] Michael Friendly. 2006. A Brief History of Data Visualization. In *Handbook of Computational Statistics: Data Visualization*, C. Chen, W. Härdle, and A Unwin (Eds.). Vol. III. Springer-Verlag, Heidelberg. (In press).

[8] Michael Friendly. 2008. Milestones in the history of thematic cartography, statistical graphics, and data visualization. *Handbook of data visualization* 100 (Oct. 2008), 41.

[9] Hanwei Electronics CO. [n. d.]. *Technical Data MQ-135 Gas Sensor*. Hanwei Electronics CO.

[10] Bickerstaff Karen and Walker Gordon. 2001. Public understandings of air pollution: the 'localisation' of environmental risk. *Global Environmental Change* 11, 2 (2001), 133–145.

[11] Sunyoung Kim and Eric Paulos. 2009. inAir: measuring and visualizing indoor air quality. In *Proceedings of the 11th international conference on Ubiquitous computing*. ACM, New York, NY, USA, 81–84. https://doi.org/10.1145/1620545.1620557

[12] Sunyoung Kim and Eric Paulos. 2010. InAir: sharing indoor air quality measurements and visualizations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1861–1870. https://doi.org/10.1145/1753326.1753605

[13] Robert Kosara. 2007. Visualization criticism-the missing link between information visualization and art. *11th International Conference Information Visualization* (June 2007), 631–636. https://doi.org/10.1109/IV.2007.130

[14] Andrea Lau and Andrew Vande Moere. 2007. Towards a Model of Information Aesthetics in Information Visualization. *11th International Conference Information Visualization* (July 2007), 87–92. https://doi.org/10.1109/IV.2007.114

[15] Veerabhadran Ramanathan and Yan Feng. 2009. Air pollution, greenhouse gases and climate change: Global and regional perspectives. *Atmospheric environment* 43, 1 (2009), 37–50.

[16] A Susana Ramírez, Steven Ramondt, Karina Van Bogart, and Raquel Perez-Zuniga. 2019. Public Awareness of Air Pollution and Health Threats: Challenges and Opportunities for Communication Strategies To Improve Environmental Health Literacy. *Journal of Health Communication* 24, 1 (2019), 75–83. https://doi.org/10.1080/10810730.2019.1574320 arXiv:https://doi.org/10.1080/10810730.2019.1574320 PMID: 30730281.

[17] Johnny Rodgers and Lyn Bartram. 2011. Exploring ambient and artistic visualization for residential energy use feedback. *IEEE transactions on visualization and computer graphics* 17, 12 (2011), 2489–2497.

[18] Johnny Rodgers and Lyn Bartram. 2011. Exploring ambient and artistic visualization for residential energy use feedback. *Transactions on Visualization and Computer Graphics* 17, 12 (Dec. 2011), 2498–2497. https://doi.org/10.1109/TVCG.2011.196

[19] Tobias Skog, Sara Ljungblad, and Lars Erik Holmquist. 2003. Between Aesthetics and Utility: Designing Ambient Information Visualizations. *IEEE Symposium on Information Visualization* (Oct. 2003), 7. https://doi.org/10.1109/INFVIS.2003.1249031

[20] H.Y. Tong, W.T. Hung, and C.S. Cheung. 2000. On-road motor vehicle emissions and fuel consumption in urban driving conditions. *Journal of the Air & Waste Management Association* 50, 4 (2000), 543–554.

[21] Fernanda B. Viegas and Martin Wattenberg. 2007. Artistic data visualization: Beyond visual analytics. *International Conference on Online Communities and Social Computing* 17, 12 (July 2007), 182–191.

[22] Yifang Zhu, William C. Hinds, Seongheon Kim, Si Shenc, and Constantinos Sioutasc. 2002. Study of ultrafine particles near a major highway with heavy-duty diesel traffic. *Atmospheric environment* 36, 27 (2002), 4323–4335.