STATS 744 Final Project:
Data visualization for variables in air quality

Korede Adegboye, Kyuson Lim

December 22, 2021

1 Introduction and Motivation

Various particles in our air remain unseen to the naked eye but are consistently interacted with. Air quality is a significant portion of the environment that surrounds us and could be perceived via continuous data measurements. In fact, during extreme lockdown restrictions of the pandemic before and after photos were taken to demonstrate the smog that is caused by emissions. It is important to note that people were unable to notice such sites because hourly changes in air quality are marginal. Giving an effect that there is minimal changes in the air that surrounds us but photos taking days apart indicate otherwise. Air quality measurements are universally influential to all biological beings in the world. Over time and space, air quality can be accessed. In particular understanding, the air quality of an environment that is highly populated, and is a landmark for factories is of interest. In the year 2020, amongst other changes work (or school) from home was enforced. How can the air quality of 2020 be independently accessed across space and time? How can air quality affect those that commute to different regions of Ontario? What periods give a clearer understanding of the air quality of 2020, and if any the use case of accessing air quality during the respective periods?

Mainly, we propose sub plots and compound plots to be useful and effective for portraying the variables of a time series of air quality concerning including NO_2 and $PM_{2.5}$. What will be displayed are concise attempts to balance memorable data visualizations and simplicity. Through the construction of each graph features of the data were highlighted using various principles of visualization. For relevance, the visuals are curated for an understanding of air quality in 2020. When trying to understand the complexity of time series data, there is much emphasis on its decomposition. Meaning, there exist trends that are basic and insufficiently capture features of the data. To facilitate the observations we display graphics that are absorbable by a general viewer.

1.1 Variables of air quality

The variables of air quality can be categorized by quantitative and qualitative. These variables comprise of hourly measurements of Sulfur Dioxide ($SO2$), Carbon Monoxide (CO), Nitrogen Dioxide ($NO2$), Nitric Oxide (NO), Nitrogen Oxides (NOx), Ground-level Ozone ($O3$), Fine-Particle Matter ($PM2.5$). There also exists station information, where the name and longitude and latitude of each station is given. With limitation of the access and missing data of stations for cities, the main data to be considered as variables in the visualization include Nitrogen Oxides (NOx) and Fine-Particle Matter ($PM2.5$).

To compliment the primary data source, we looked at Ontario geospatial data, population data, and traffic volumes.

2 Visualizations - Story

2.1 Plot 1: Scatter plot on 28 Ontario regions

There are many locations in the Ontario to observe with. The most concerning regions are ones that can be further evaluated. The main interest the plot intend to answer is where regions of Ontario are placed for air quality. The observations of this graphic indicate that Ontario regions tend to be heterogeneous - there exists variation across the regions. The density gives reason to believe that some regions can be analyzed together. The size of Ontario is $1,076,395km^2$ which is larger than usual size of such nation like South Korea ($100,210km^2$). This indicate a qualitative difference to inference for Ontario for 28 regions as well as different perspective of air quality to coexists around 28 different atmosphere.

However, Hamilton appears to be much different from the others. In addition there is variation of air quality variables. Another perspective of what is happening in Ontario, and specifically Hamilton would be feasible. Giving that Ontario is comprised of various environments, it is suggestive of the visual that northern cities are have a lower ratio. In other views, there are some developing rural cities in the mix of low-moderate ratios. Implications of these may lie with those being the standard. This may come from a decrease of any of the 2 variables shown.

For analysis and graph, an optimal features of air quality that surround the atmosphere of cities and shown with scatter plot and for comparison to show how quantitatively location differ for air qualities.

There are 5 major aspects that is shown from the plot:

- **Mean values of $NO_2^{1/3}$ and PM $2.5^{1/3}$ emission** for each city: The quantitative points are shown in the scatter plot with text labelled cities where they are placed as well as qualitative measurement of yearly gas pollution in a 2D topological plot. In this way, the non-parametric rankings of 28 cities are compared for viewers to come up with measured comparisons between cities that they are interested in.

- **Groups within density contour of variation** for each city: 28 cities are plotted with its mean values of yearly $NO_2^{1/3}$ and $PM2.5^{1/3}$ where points fall within 3 different contours, mostly. Each city is regarded to vary between the density contour for their variation of 12 month periods that is plotted by the 2D density graph for scatter points.

- **Categorical clusters of ratio taken between $NO_2^{1/3}$ and $PM2.5^{1/3}$** by NO median as a guideline for colored groups: Found from the median value of monthly NO emission (0.5 as a cut point), there is an obvious two groups clusters of cities that has a different median point when a ratio of $PM2.5/NO$ is distributed in the subplot for comparison.

- **Scatter plot of Hamilton gas emission compared to density plot**: Within 12 months of variation by mean of weekdays and weekend, the Hamilton's individual scatter points of 24 points are shown compared to general density contour of NO_2 and PM 2.5 emission for Ontario regions, which show the variation of highest gas polluted city, Hamilton.

- **Combined legend, box-plot NO graph**: Within 12 months of variation by mean of weekdays and weekend, the box-plot of NO emission in Ontario is taken to be the legend of the color of scatter plot's points for its obvious categorization of two groups. The median is taken to be the cut-off of two colors, which is 0.5 and used as a color variant for 28 cities that are in the scatter plot.

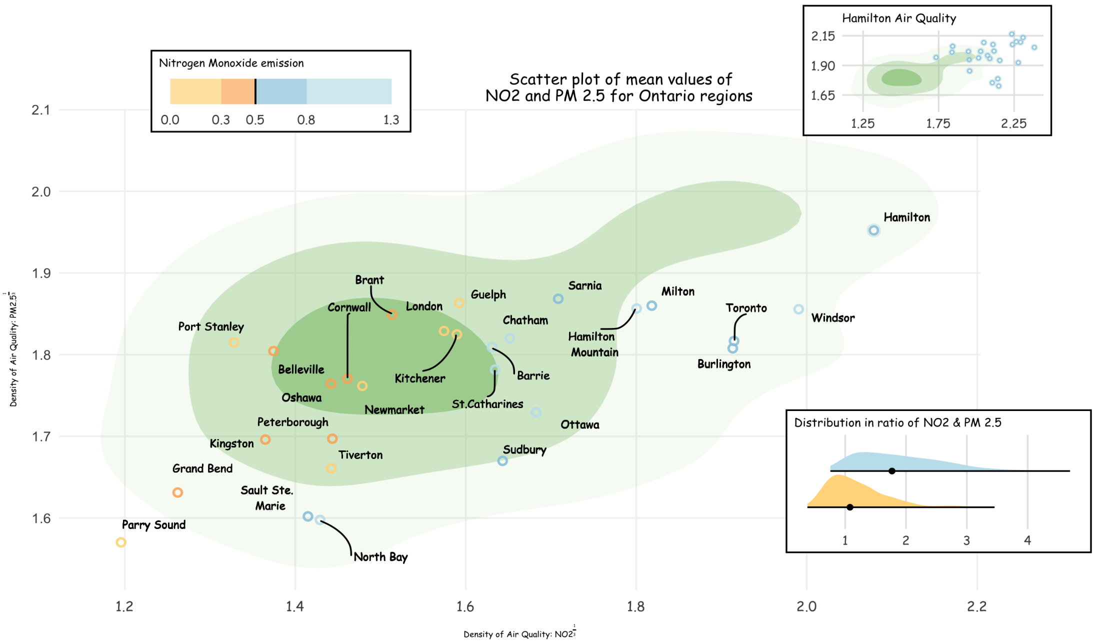

Figure 1: A bivariate Scatter plot of Ontario NO_2 and PM 2.5 with box-plot NO legend plot and scatter plot of Hamilton area, combined with ratio distribution density curve.

The data analysis is as follows. From the box-plot analysis, the data are cleaned by omitting outliers and placed on the scatter plot where points are represented by its location. The side plot shows for the legend how the color is chosen which comes from the median of box-plot as cut-off values. Also, ratio of two variables are distributed on the density curve as well as another scatter plot is described for only Hamilton of 12 month period.

2.1.1 Plot 2: Comparison of weekend and weekdays estimates using GAMM (Generalized Additive Mixed Model)

Correlation between NO_2 and $PM_{2.5}$ is visualized. They both follow an increasing pattern marginally different times. Where it appears there is a slight lag for $PM_{2.5}$. Also suggested by the data is that when the work (school) day begins then emissions are expected to be of most concern on a daily basis. What is interesting is that around 2:00PM–3:00 PM there is a decrease in both variables. At this time of day, many students are released for the day. Before this many would have presumed that parents were picking up their kids as they experienced traffic during this time frame. This graph communicates that to be potentially be bias, and other more natural means of commuting home are being used by students. Seeing that there has been a average decrease of both variables from 2018-2019 to 2020. Which gives relevance to the states of the pandemic.

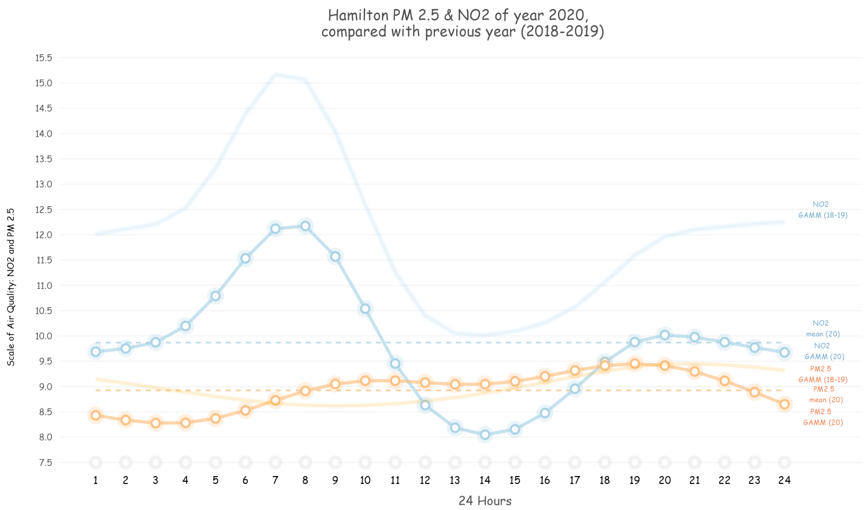

Figure 2: Hourly Time series with estimates using GAMM, shown together with each NO_2 and $PM_{2.5}$ and reference line from 2018-2019. Data Source: Air Quality Ontario

Visually, hourly trends is of significance in various settings. First the data given, the impact of each variable can indicate what a person their air quality to be throughout the day. Where adults and children are found to be at more risk of poor air quality, then advise can be deployed to give caution. This is how air quality is currently being accessed by the general public. Second, another use case of a hourly decomposition can be provided in relation to air cleaning devices used in buildings due to the pandemic. As the devices were developed, they were interested in evaluating how long it may take to clear the hazardous particles in the air. Such periods would presumably be assessed minutely, or hourly.

2.2 Plot 3: Interval plot on Top 7 regions of Ontario (2020)

Now, the distribution of two variables, NO_2 and $PM2.5$ is shown with interval plots for top 7 cities by quantity of NO_2 emission in Ontario that is found from the previous scatter plot; Hamilton, Windsor, Toronto, Burlington, Milton, and Hamilton Mountain. Even though the median represent the air quality of cities, the variation is clarified with the distribution of interval plot as to be described with. In this plot, the main interest to answer with is the difference of population for rural and urban areas actually matters for air quality as to combine for impact measure which is a numerical values that combines both NO_2, $PM2.5$ and population of cities.

These cities also have very high PM2.5 to be considered as a dangerous places for citizens to live with. Hence, the goal is to come up with global impact on population of each cities with two variables that have same range by the formula

$$\text{Impact } = \left(\frac{\text{population}}{500,000}\right) * (\overline{NO_2}_i + \overline{PM2.5}_i),$$

since the air quality is naturally influential for every individuals human beings in the city. Note that the NO_2 and $PM2.5$ has a same measurement to be compared in the previous line plot to take into account for the consideration. Also, the numerator of 500,000 is taken to be the cut-off of metropolitan city for large population cut-off to be placed for impact values to calculate with. By this cut-off in population of numerator, we are able to better classifies the difference of impact for air pollution between the urban and rural areas to be found from the geospatial point plot.

There are 2 major aspects that is shown from the plot:

- **Distribution of top 7 highest air polluted regions** are compared within interval chart by two variables, NO_2 and PM 2.5: A median contain interval, 45%-55% is highlighted at the center with it variation shown in a separate colors to highlight the different distribution of each city, which are 42.5%-57.5%, 37.5%-62.5%, 25%-75% and 15-85% as intervals.

- **Impact measure of top 7 highest air polluted regions on population** for categorization: By calculation of adding two gas pollution for each city multiplied by the population of each city, an influence of two pollutants is categorized as a measure where the bottom geospatial point map shows the impact of gas pollutants for each city.

Those regions that are identified to have a higher average of $PM2.5$ and NO_2 than other cities in the scatter plot tend to have non-unimodal distribution in their distribution plots.

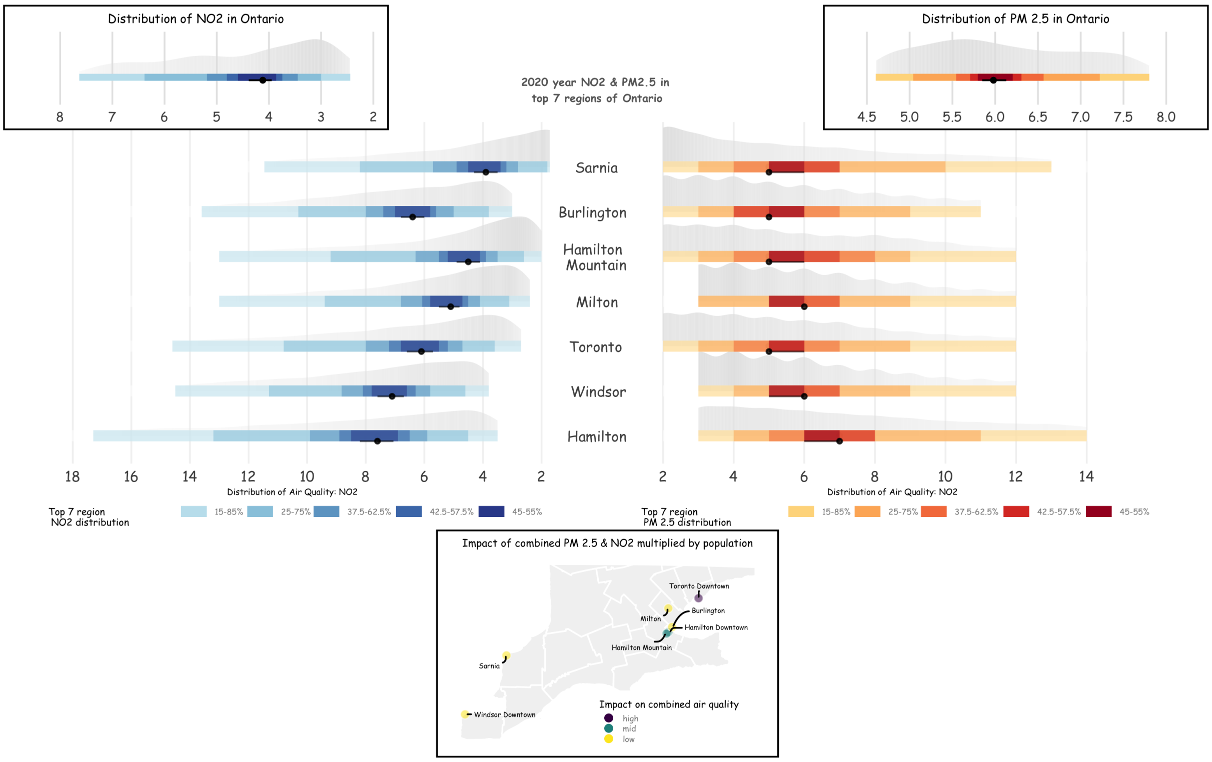

Figure 3: An interval distribution plot of top 7 NO_2 regions along with Ontario's average NO_2 and $PM2.5$ interval plot with an impact geospatial point plot.

While $PM2.5$ is a countable measurement to have peak at each point, the NO_2 is a continuous measurement to have negatively skewed distribution models in all 7 cities. From the categorical classification of top 7 cities for the impact, there is an obvious discrimination to be found from the impact measurement by the multiplier, population of citizens in the city.

While Hamilton is observed to have the highest air pollution of NO_2 and PM 2.5 from the scatter plot for its yearly mean and points of the monthly average, Toronto is found to have the highest impact on its population for the city by the impact numeric value computed. This clarifies the impact to be different for rural and urban areas by the identification of impact values. Not only the air quality that is around us to change along the season and month as it flows around, but also the impact by population clarifies the discrimination of urban and rural areas by its population.

Here it is environmentally reasonable to insert the classifications of regions - urban, suburban, and rural. Therefore, to some extent, the insights proposed by the data of Hamilton can be extended to other rural regions. Also, the geospatial point plot shows the impact of hazardous air quality for it increment based on the population to live in the city, as 3 variables work to create impact numerical values.

There exists evidence that air quality variables differ across regions. Some with marginal differences, others much significant.

2.3 Plot 4: Monthly Univariate geospatial graph on Ontario regions

There are 2 major aspects that is shown from the plot:

- **Monthly variation of NO_2 emission for Ontario regions** in comparison for geospatial map: The Hamilton area is specified in the plot separately from other regions for monthly comparisons. Other regions are also observed to vary by the percentile division of NO_2 emission which is $0-25\%, 25-50\%, 50-75\%$ and $75-100\%$ for comparison.

- **Monthly variation of NO_2 emission for Hamilton region** in magnified square for comparison: Previously, the Hamilton area is found to have highest gas polluted area. A monthly comparison of mean values clarify for variation of NO_2 air pollutants.

In Ontario, it is clear that both variables tend to decrease in the summer. This is expected since flowers can absorb emissions. The geographical representation demonstrates that rural regions adjust to that seasonality much more effectively than urban regions. Again, expected since rural regions are in favour of having more trees and less population.

As we are concerned with Hamilton it is clear from both NO_2 and $PM_{2.5}$ that not much insight can be retained over the months. Rather it indicates that the two variables are relatively constant. With that, there may be other dimensions of time that will be useful.

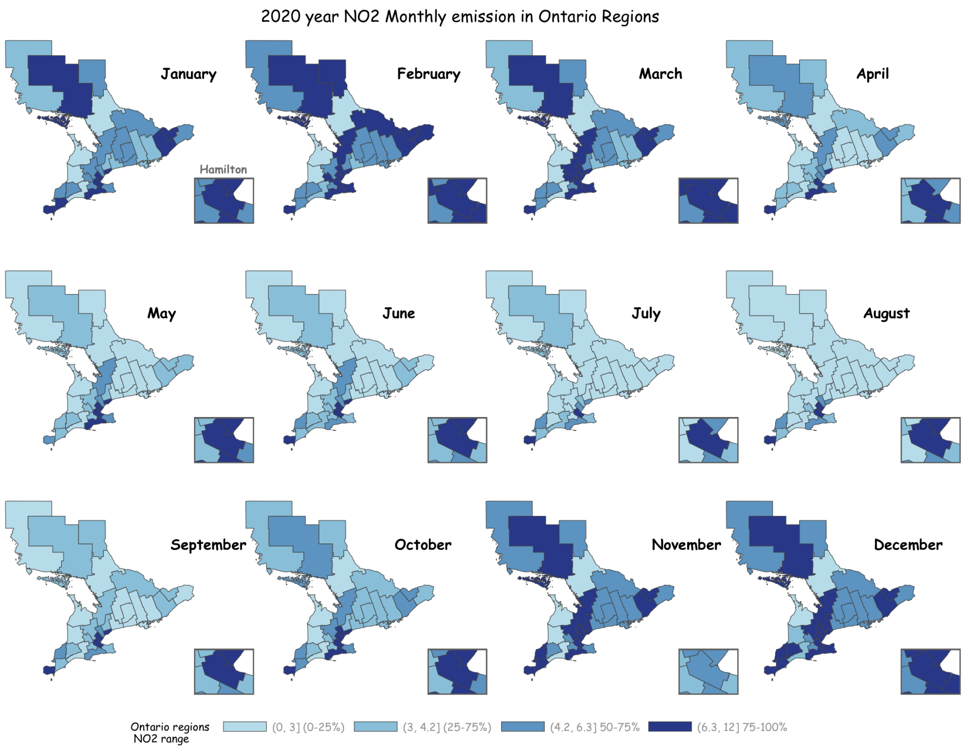

Figure 4: A Univariate geospatial graph of Ontario NO_2

The central part on Ontario including Toronto and Hamilton areas are found to have highest

NO_2 air pollutant over 12 months of period. Although other areas are found to have low emission, constant pattern of high polluted areas are observed to be identified in January, February, March, November, and December for increments. Some of these area includes Burlington, Barrie, Hamilton, and Toronto on particular seasons, Winter and Spring. This clarifies the seasonal changes of air quality to change along particular season by a law of nature.

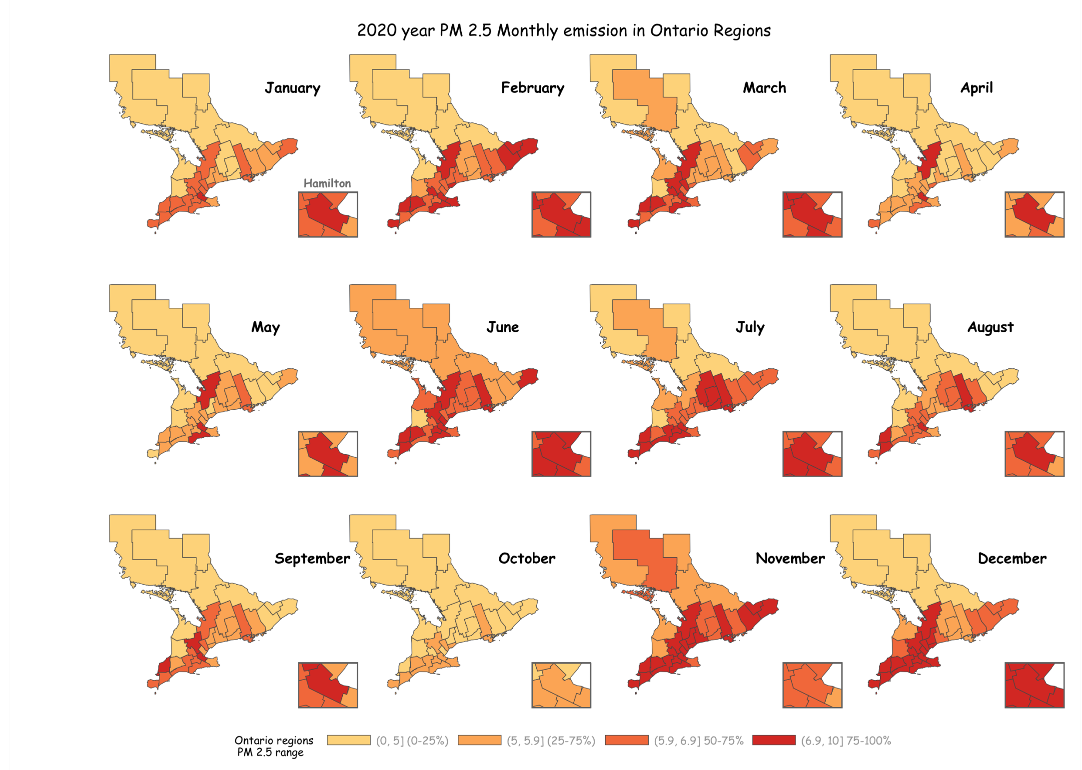

Figure 5: A Univariate geospatial graph of Ontario PM 2.5

Although NO_2 and $PM2.5$ shows a similar pattern of increment and seasonality from previous plots, most areas that have high emission of $PM2.5$ maintain its quantitative measurement for the year. This shows the constant pollution that human being could make for populated areas that have constant high $PM2.5$ range for the year. However, most areas that are identified to have high percentile range (75-100%) are also observed to have high percentile (75-100%) during 12 months period. Even though there is a variation and difference compare to the trend of NO_2, the seasonal changes of air quality to change along particular season by a law of nature is inevitable for most regions with its color changes in pattern to detect with.

2.4 Plot 5: Daily Estimates using GAMM (Generalized Additive Mixed Model)

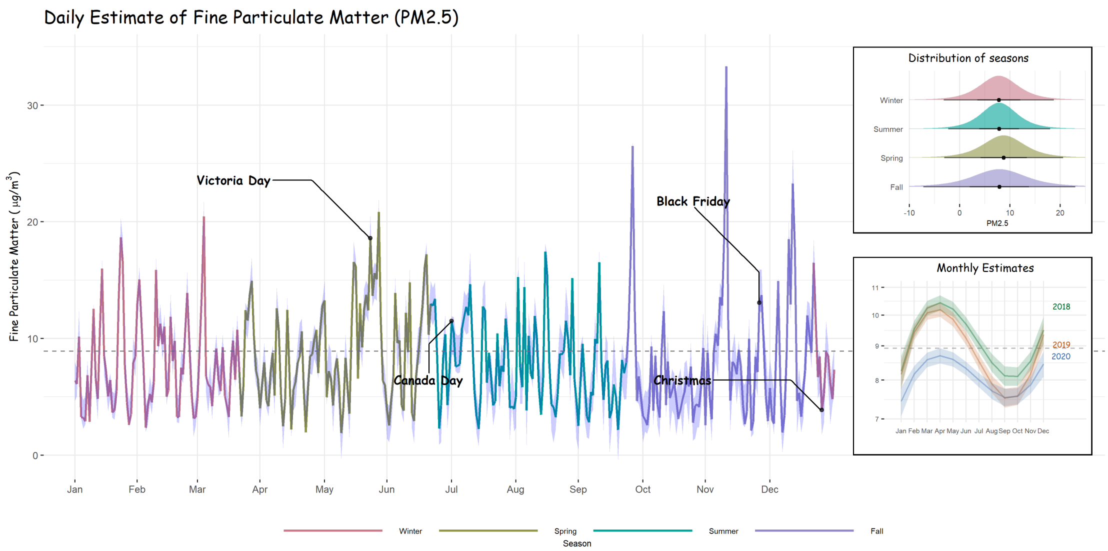

Figure 6: Daily Time series with estimates using GAMM, shown together with each seasons respective distribution, and smoothed trend lines of previous years. Data Source: Air Quality Ontario

This visual provides the viewer with an understanding of how fine particulate matter (pm2.5) varies within the year 2020. The graphic was constructed where the audience is interested in specific days of the year. Here the daily estimates fluctuate throughout the year. As depicted on Black Friday there was an increase in $PM_{2.5}$. Amongst other days it would be interesting to see if there exists a trend of eventful days, or increases are due to unknown sources at this time.

The next layer of the graph includes the seasonal trend. It pivots from the plot displayed previously, where months were of interest. That vague insight of there not being many variations across months can now be reassessed. We know concerning other regions, seasons appear to have a moderately larger impact on the effect on $PM_{2.5}$. Here it remains that $PM_{2.5}$ does not appear to vary much for the year 2020. This view of the plot is given by the distribution subplot in the top left. In other words, fine particulate matter appears to have homogeneous distributions, with spring having a slightly higher central point.

The trend in the top right subplot indicates otherwise. Concerning seasons, 4 turning points appear. It indicates that fine particulate matter is at its highest during warmer climates and tends to decrease during colder. This suggests that for 2020, an increase of fine particulate matter for warmer climates could be described as natural occurrences. Ones we would expect to see in previous years. But for colder climates, it appears to be more event-oriented. Additionally, the reference plot shows that Hamilton's environment has been decreasing. Thus the various fluctuations suggest a healthy trend of air quality. Leaving the user to advocate understanding the days that show a decrease. Wondering more about a given day. Note that it may very well be that fine particulate matter has its classification of seasons, which is inconsistent with what is displayed. A visual reminder that December only contains a small portion of winter is given to learning about $PM_{2.5}$ concerning the winter break. $PM_{2.5}$ fluctuates below a reference line. Suggesting that the states before and after winter break could be a potential route of research.

Referring back to plot 3 (or Figure 3 and Figure 4), we discussed that $PM_{2.5}$ did not present itself as different across the months of 2020. This was solely due to the colors being chosen for regions of Ontario. Now information is retained that is (for the most part) independent of other regions. Note, Hamilton can still be marginally dependent on surrounding regions.

Hence $PM_{2.5}$ in Hamilton is not entirely homogeneous within. At different levels of decomposition of the time series, it may appear so. About the motivation of the study, it was mentioned that air quality is difficult to understand daily. But here the data suggests that it is more insightful to grasp. After understanding the underlying data, then assumptions can be portrayed onto months and years. If daily data indicates such a finding then it would be interesting to see the outlook of hours within a day.

3 Methodology

3.1 Plot 1: Scatter plot on 28 Ontario regions

The intention of the visual was to translate aspects of the motivation, so that the audience can sense the same difficulties that we were introduced to. Leaving the viewer general thoughts and assisting them with keeping an open-mind when learning about air quality. It mainly attends to the appearance of information. Although it still has some suggestive features. The positioning, colors scale, and labels create effective discrimination, ranking and ratioing.

These elements are framed within the idea of compound plots. A viewer first interacts with the density and the labels, then recognizes the color scheme of the points are related to the subplots that visibly simple to understand. Notice that majority of the graphic is light, but the labels are dark. The innovative part of the plot is on legend of color difference by the cut of NO median taken from the box plot of NO emission of all Ontario regions. Divided by its median of the NO emission, the color of two categorical groups of cities are colored on the scatter plot for apparent difference in the scatter plot.

Since interest is in Hamilton, a subplot that is slightly away from the view of the others is displayed. The visual was given as a means to begin developing the story and understanding the variations of air quality and variations across regions. A view that it best facilitated as being general. The innovative legend (top left subplot) is a distribution interval of the values, grouped by the two variables.

In an attempt leverage the flexibility of *ggplot* before considering other packages, it was redundant to annotate 28 regions where there labels will have to be chosen effectively. Hence, *ggrepel* added position labels and arrows, and restricted the overlapping of points. Across Ontario, three air quality variables were available. To begin to give the audience an understanding that air quality differs 2 variables would suffice. For more precise distribution difference in the scatter plot, scales are taken to the power of 1/3 to show more obvious difference between cities for its air qualities. It is possible to see both ratio of distribution difference between two variables, NO_2 and $PM2.5$ and regions themselves from the subplot of density plot.

What is ignored in this graphic are the axis labels. Although it is given for reference and display of authenticity of the information. To further engage with the ratio, it would be compelling to view what exists when that information is separated. Distribution intervals help to visualization uncertainty of point estimates. According to (Wilke, Fundamentals of Data Visualization) it has been noted that the plot could be difficult to interpret when the strips fade to white. As such, intervals were chosen to alleviate that challenge. While this is the manner that it can used, we also displayed it in relation the transparency of points. Half-eye plots provide a different perspective to the distribution of the respective air qualities. The viewer can freely choose which they are most comfortable with.

The graph communicates a story that we are able to pick up on. But we have had much more experience with the data to quickly grasp cues, and thus bias towards an effective visual is plausible. Arguments can be made against this graph where it can be visually busy (filled with noise). To better organize the plot, the information could have been split by the traditional and possibly cleaner approach given by compound plots. The plot aims to capture a relationship between 2 variables in respect to all regions available.

An alternative view would be display a ranked bar graph with flipped coordinates of the proportional mean values of each city and both variables. This would be quick for the viewers to follow but may not be retained as it wont be memorable. With such a distinction may be best suited for the last graph. Since it usually the most complex, the idea of memorable is naturally introduced to capture the information.

3.1.1 Plot 2: Comparison of weekend and weekdays estimates using GAMM (Generalized Additive Mixed Model)

The plot acts as a supplementary visual for Figure 1. The key components are the appearance of points with the line and a blurred line. The glow effect on the points bring the viewers attention to the main trend lines of 2020 to be evaluated. While the blurred lines in the background aims to act as a reference in a non-distracting form. Such a display is suitable when various lines are presented on the same graph. Comparison of 4 lines are best understood by the audience.

Although it is not shown here, more lines, say 6 can be made visible. With 4 of the most recent years at the forefront (given points and shadow) and the remaining 2 as references. The reference (blurred) lines are not to be interacted with as much as the others. There importance is minimal but can be closely looked at to strengthen ones understand.

3.2 Plot 3: Interval plot on Top 7 regions of Ontario (2020)

Extracting the minimal but suggestive properties of the previous graph for a adhesive story was considered. Here, it was still an aim to provide the audience with a general understanding but more interesting and direct implications than that of the previous graphic. The two colors from the previously plot are again used for consistency. At first it was of interest to find a way display the distribution of all regions. Then realising that this would become a visual that is difficult for the audience to read and retain information from. So a ranking of top 7 cities in respect to the ratios are displayed. The idea of ranks is one that is traditionally displayed via bar graphs.

Taking the average for each region would oversimplify the data in a way that implies concrete information can be perceived by the user. Thus, focus was on highlighting variation of data in respect to different levels (variable, across region, within region). With each rank there is a combination of box plot and density. Information first given by the gradient distribution interval.

3.3 Plot 4: Monthly Univariate geospatial graph on Ontario regions

This graph gives an image that combines geographical regions (shapes) with that of a sequential color scale for increasing and decreasing trends of air quality. The selection of intervals were chosen to be quarterly as to reflect a balanced amount of visual differences between regions. Information absorbed in this way leads to direct assumptions of the data. Thus, it is more so suggestive than it is informative. To emphasize that Hamilton is also of interest a facet zoom was used to showcase Hamilton trends. Viewers may then assess the surrounding regions of Hamilton. Note a label of Hamilton is given once as repeating the information would be repetitive and does not add information.

Monthly trends can be useful for understanding the impact of changes related to an environment. For instance, when companies release climate action plans it is reasonable to review if such goals are being met. Clearly conditions (variables) will slightly vary. Where regions may be now by smaller communities or buildings.

3.4 Plot 5: Daily Estimates using GAMM (Generalized Additive Mixed Model)

The central purpose of the visual was to support associations amongst two qualitative variables (Month and Season) in respect to fine particulate matter ($PM_{2.5}$). It focuses on demonstrating to the viewer that there is valuable information underlying the variations of $PM_{2.5}$ in which they can grasp the information quickly. Generally, it highlights notable deviations in respect to the days of the year. A generalized additive mixed model was used for various reasons. When accessing the data quality, it was discovered that approximately 14% of the data was missing. Interpolation with GAMM was used to estimate these missing values and possibly account for other noise attained from the initial measurements of the data. The other significant use of GAMM was the ability to tune the smooth terms. In the case of the daily time series, 366 terms were chosen to detrend the hourly data. Whereas for the subplot of smooth trends, 4 terms were used giving the line much more of a functional form. All the plots contain within this graphic make use of the GAMM. To be transparent with the results, the confidence intervals assist with visually the uncertainty of estimates. It is displayed such that it does not visually interrupt the many various fluctuations of the estimated trend.

The subplots played a significant role in this visual. It allows the audience to easily retain cues that may be not as evident in the main plot. Additionally, it can be leveraged re-iterate a viewer's understanding of the main plot. Note, there is a difference in understanding between the main and subplot. The main plot leads one to various perceptions, whereas the subplots give basic results that are solely intended for the understanding of the main plot.

At the first view of the graphic, lines appear and categorical colours. This appearance gives immediate cues to the time series. By the length and overall size of the line, the attention is greater than that of the subplots. Note that the bottom subplot is on a different scale. Rather it can be referred to as a magnified scale. Where a range of fewer values is considered. Although analogous scales are trusted most often. Here it was important to highlight a range of the data that daily estimates, unfortunately, could not capture.

The labels on the x-axis are given by the month rather than days, to lead the user to focus more so on the seasons and months. The legend is used to show the main plot and the top subplot are coloured by season. Such a scale is known to be qualitative to showcase the seasonal categories. As for the bottom subplot, the legend was removed and direct labeling was introduced to enable the viewer to quickly recognize which trend line is associated with which year. Minor grid lines were removed to alleviate visual noise. Partially transparent confidence intervals were used in the bottom subplot to be more visually appealing. The colors were also chosen to be soft on the eyes, and not distract the viewer. Positioning and white space were given attention to ensure the user experience was seamless.

Use of the *mgcv* and *gratia* package were used to generate the GAMM. A drawback of this was long rendering times. To account for this, models were saved in rData files. For larger amounts of data this strategy becomes more important. Use of *gratia* may be limited from a data visualization point of view. To continue with further decomposition, the extraction of the trend was required. This then enabled the more flexible and computationally friendly *ggplot* package. Another side-effect of the *mgcv* was that it gives the estimating curve in terms of weights. Using the predict method or extracting the y-intercept of the model could assist with reverting back to the response units. The trend displayed is from a bootstrap sample. While various bootstrap sample could have been displayed, we chose to show a confidence interval to show the uncertainty.

It is questionable whether or not the daily trend is captured effectively. It very much could have been shown but as a reference plot not the main. An alternative display could be given by a graph related to bar plots where the count of days that fall above the mean line and there is a count for those that fall above. Keeping idea of seasonality, bar graphs for each season can be constructed. It stands as a replacement to much of what is shown. It can be visually understandable in a few seconds and viewers can trust the results. Smooth curves appear to be most effective when a potential trend is evident. Otherwise it can simply leave others with results that they cant be claimed confidently.

4 Conclusion

In search of appealing visualizations that can be communicated effectively we find that the feasible graphics are ones that highlight feature(s) of the data. Air quality in Hamilton, most importantly does suggest distinct assumptions given the period. The most suggestive features of the data is that there was a decrease of air quality for the year 2020, air quality distributions have multiple modes - implying that there exists sub regions or climates that have an effect, emissions in the winter are of concern since flowers are not as active in absorbing emissions. It is meaningful to the audience to be aware of such factors when defining problems and proposing solutions. To process further development of plots, a closer look at uncomplicated plots would be of interest. Appeal can be added by first be formally evaluated to assess whether or not the data is communicating more than what is given. If so, then a natural transition to more complicated plots, will benefit from appeal to engage the viewer.

5 Acknowledgement

The copyright of plot 1: scatter plot on Ontario region belongs to the authorization of Kyuson and Korede as created by own ideas and combination on density contour, scatter points and text labels.

Also, the copyright of plot 2: impact rankings on top 7 Ontario region belongs to the authorization of Kyuson and Korede as created by own ideas and combination on point colors and mathematical formula to be used with.

References

[1] David Wachsmuth *Making beautiful inset maps in R using sf, ggplot2 and cowplot* `https://upgo.lab.mcgill.ca/2019/12/13/making-beautiful-maps/`

[2] Cedric Scherer, *Data Visualization Computational Ecology: Daily Temperatures in Berlin, Germany (Topic: Visualizing Uncertainty).* `https://www.cedricscherer.com/top/dataviz/`

[3] Matthew Kay, *Slab + interval stats and geoms.* `https://cran.r-project.org/web/packages/ggdist/vignettes/slabinterval.html`

[4] World Population Review, *Ontario Population 2021.*
`https://worldpopulationreview.com/canadian-provinces/ontario-population`

[5] Roy Francis, *Introduction To ggplot2.*
`https://nbisweden.github.io/workshop-r-1911/Exercises/ggplot_roy/lab/ggplot_lab.html`

[6] Edzer Pebesma, Roger Bivand, *Spatial Data Science with applications in R, Chapter 12 Spatial Interpolation.*
`https://keen-swartz-3146c4.netlify.app/interpolation.html`

[7] Claus O. Wilke, *Mixing different plotting frameworks.*
`https://wilkelab.org/cowplot/articles/mixing_plot_frameworks.html`

[8] Gavin Simpson, *Introducing gratia.*
`https://fromthebottomoftheheap.net/2018/10/23/introducing-gratia/`

[9] Kamil Slowikowski, *ggrepel examples.*
`https://ggrepel.slowkow.com/articles/examples.html`

[10] Ontario, Ministry of environment, conservation and parks, *Air Pollutant Data.*
`http://www.airqualityontario.com/history/`

[11] The R graph gallery 2018, *Time Series.*
`https://www.r-graph-gallery.com/time-series.html`

[12] Claus O. Wilke, *Fundamentals of Data Visualization.*
`https://clauswilke.com/dataviz/preface.html`

[13] Len Kiefer, *Watch your charts glow up with R and ggfx.*
`http://lenkiefer.com/2021/03/17/watch-your-charts-glow-up-with-r-and-ggfx/`