UNIVERSITY OF WESTMINSTER

7BUIS009W.2.2020

# Data Visualisation and Dashboarding

# Coursework

## Analysing Pedestrian and Vehicle Traffic Data in London During the COVID-19 Pandemic

Presented by **Jean Boutros** (W1804948)

Last edited on: Thursday, 22 April 2021

## Table of Contents



## Acknowledgment

In this report, I have chosen to remove the appendix and add the images directly into the text. The purpose of doing so is mostly to have a better clarity when reading, and to avoid having to switch back and forth between the text and the visualisations in the appendix.

Kindly note that the report is approximately ten pages in total, and even a little less, if all the paragraph spacing, the formatting, the page breaks and the images are removed.

I hope you enjoy reading it.

## Research Question

What insights can we get on mobility in London, during the three lockdown periods, by analysing data collected on vehicle and pedestrian activity from traffic cameras?

## Rationale

The UK government imposed three major lockdowns during the COVID-19 pandemic between March 2020 and March 2021. As part of the government's effort for faster indicators release in response to the coronavirus pandemic, an experimental dataset has been released by the ONS on the **busyness indices** in the UK. This dataset is generated as a result of analysing traffic cameras images to identify cars, motorbikes, buses, trucks, vans, pedestrians and cyclists (Edwardes, 2020).

By using this dataset, my objective in this project is to identify the periods of lockdown and analyse the patterns of mobility of the different traffic sources. The project will take London as a point of focus.

## Timeline of lockdowns in the UK

The timeline of lockdowns in the UK is as follows (Pa Reporters, 2021):

First Lockdown

- **23 March**: The first UK lockdown is announced. Movement is restraint to limited reasons such as food shopping, exercise, medical need, and essential travel.
- **1 June**: Lockdown measures are eased.
- **4 July:** Lockdown restrictions are eased across England, non-essential shops reopen.

Second Lockdown

- **31 October:** A second national lockdown is announced for four weeks, with the closure of hospitality and non-essential shops.
- **2 December:** England's national lockdown comes to an end and is replaced by a strengthened three-tier system.

Third Lockdown

- **4 January 2021**: A third national lockdown is announced.
- **12 April:** Step 2 of the roadmap out of lockdown begins. (Adams, 2021)

## Data Acquisition

### Who created the data and why?

The data in this dataset was first gathered from traffic cameras providing open access to their footage or images. In the UK, there are thousands of CCTV traffic cameras, spread across the country, but mostly concentrated in big cities and towns. London alone has over 1000 cameras that are publicly accessible through the internet. Among these public cameras are the camera network of the TfL.

The ONS (Office for National Statistics) collected the images from the various openly accessible cameras, including the TfL, and **processed them to compile numerical values that represents 'busyness'.**

## How has the data been compiled?

In order to limit the number of images collected for processing, and as an effort to select only the images that represent the most comprehensive and representative amount of information necessary to compile an indicator of busyness, the ONS created a set of criteria that the camera must match to be selected. The criteria involved assigning scores for each camera based on the number of features represented in their resulting images. The features include among others the following: shops, residential, pavement, cycle lane, bus lane, traffic lights, etc.

Next, to avoid selecting cameras from only few providers such as the TfL, a percentage was assigned to the number of cameras identified as representative of busyness (based on the score mentioned previously) per provider. The selection of cameras was based on the values of the percentage of cameras normalised by provider per total score.

Finally, as a last sampling criteria, geographic areas were kept separate with each a timeseries of its own to avoid having areas with overwhelming contributions compared to others.

Following the selection of cameras, pipelines were put in place to download and process the images on regular basis (every 10 mins). **The pipelines include a deep learning object detection model that identifies categories such as pedestrians, cyclists, cars and buses, then saves these detections to a database.** Subsequently the detections are then aggregated and computed to produce the final indicator of business.

## How reliable can we expect the data to be?

In terms of traffic cameras coverage, they are mostly concentrated in the big towns and cities which could be a problem if we were to create a comparative analysis. However, since this project focuses uniquely on London, we can disregard this issue.

Conversely, an important factor that could affect the reliability of the data, is its accuracy, which depends on a number of external factors. Since the ultimate aim of the images captured through the traffic cameras is to provide a **quantifiable measure of "busyness"**, The positioning of the cameras can affect what objects it can detect. The weather can be another detriment to the quality of image and what can be eventually extracted from it. The object detection algorithm / deep-learning model, is another element to consider when evaluating the accuracy of the data, because each type of algorithms has its strength and weaknesses and in practice, deep learning models can never achieve 100% accuracy.

The aforementioned factors imply that **this data is more suited for estimating trends rather than absolute numbers, and it should not be used as a measure for headcount or number of vehicles or similar deterministic measures.**

## What are the analysis implications?

The main implication on the analysis is that the “busyness” indicator should be used very carefully to **depict trends and not for estimating any precise or even approximate amount of vehicle or pedestrian movements.** It is also not suitable for understanding journeys in terms of departures and destinations (Office for National Statistics, 2020). The “high frequency (hourly) and timely (weekly) data, can help to detect **trends and inflection points in social behaviour**” (*ibid.*).


*Figure 1 Segment of traffic camera video footage from St Giles Circus in London. Cars, pedestrians and buses are clearly visible in the image*

## Data Preparation

The dataset titled '**Traffic camera activity**' is provided by the ONS at the following link: https://www.ons.gov.uk/economy/economicoutputandproductivity/output/datasets/trafficcameraactivity

It is updated on regular basis and access to history is possible in case of any discrepancies.

The dataset is in excel format and contains 4 sheets. The first sheet includes metadata about the other sheets and a link to the ONS's data science blog that describes in detail how the raw data coming from the deep learning model was interpreted and adjusted.

There are 3 different sets of data, 'Non-seasonally adjusted', 'seasonally adjusted' and 'trend'. In this project I will be using the 'Seasonally adjusted' dataset because it allows more accurate interpretations of the numbers. Seasonal adjustment is a statistical adjustment applied to time-series to take into account fluctuations of "sub-yearly observations, e.g., monthly, quarterly, weekly, [that] are often affected by seasonal variations" (Mazzi, Ladiray and Rieser, 2018, p35). The 'Handbook of Seasonal Adjustments' defines seasonality to be "due to the fact that some months or quarters of the year are more important in terms of activity or level." (*ibid.*, p35)

## Preparation the Dataset

The first step before starting the exploratory data analysis is to fix the file format of the dataset to be able to read it properly in Python (pandas) and exporting it eventually to Tableau for visualization.

The image below shows a screen-capture of the dataset viewed in Microsoft Excel. This format presents some challenges: first, the data is not machine friendly but rather human friendly allowing readers to quickly decode it. Secondly, the data is pivoted by day, location and 'busyness' indicator to show one column per day per location and per indicator, which might present a difficulty if we want to dynamically visualise the data, and make comparisons.

| | A | B | C | D | E | F | G | H | |
|---|---|---|---|---|---|---|---|---|---|
| 1 | | London | London | London | London | London | London | North East | Nor |
| 2 | | Cars | Motorbikes | Buses | Trucks | Vans | Pedestrians & cyclists | Cars | Mot |
| 3 | | * London series start 11/03/20 | | | | | | | |
| 4 | | | | | | | | | |
| 5 | 01/03/2020 | * | * | * | * | * | * | 47021 | |
| 6 | 02/03/2020 | * | * | * | * | * | * | 54963 | |
| 7 | 03/03/2020 | * | * | * | * | * | * | 56779 | |
| 8 | 04/03/2020 | * | * | * | * | * | * | 53760 | |
| 9 | 05/03/2020 | * | * | * | * | * | * | 55409 | |
| 10 | 06/03/2020 | * | * | * | * | * | * | 62368 | |
| 11 | 07/03/2020 | * | * | * | * | * | * | 55552 | |
| 12 | 08/03/2020 | * | * | * | * | * | * | 46518 | |
| 13 | 09/03/2020 | * | * | * | * | * | * | 53752 | |
| 14 | 10/03/2020 | * | * | * | * | * | * | 57529 | |
| 15 | 11/03/2020 | 294869 | 506 | 19937 | 22675 | 49062 | 53723 | 54534 | |
| 16 | 12/03/2020 | 298931 | 476 | 19190 | 21816 | 47378 | 46753 | 57639 | |
| 17 | 13/03/2020 | 308105 | 567 | 20138 | 21551 | 46784 | 55337 | 58282 | |
| 18 | 14/03/2020 | 310406 | 487 | 17834 | 13936 | 30738 | 48150 | 52948 | |
| 19 | 15/03/2020 | 282430 | 447 | 14898 | 9937 | 23126 | 38965 | 46074 | |
| 20 | 16/03/2020 | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | [illegible] | |

To make the dataset readable in python there are multiple solutions. In this project I will use the programmatic solution considering that it is the most reliable way of dealing with datasets. This is because when a new version of the dataset is introduced, there is no need for any manual interventions to be made on the excel file, before reading it in Python, to explore the newly added values. This is an essential mindset when building fully automated data science pipelines where no human intervention is required.

The steps performed should be as follows:

- Read the dataset
- Remove all columns unrelated to 'London'
- Fix the multi-level index in columns by eliminating the hierarchical indexing
- Remove the rows that contain notes
- Undo the pivoting by flattening the table (This step is only necessary for certain plotting techniques)

## Data Exploration

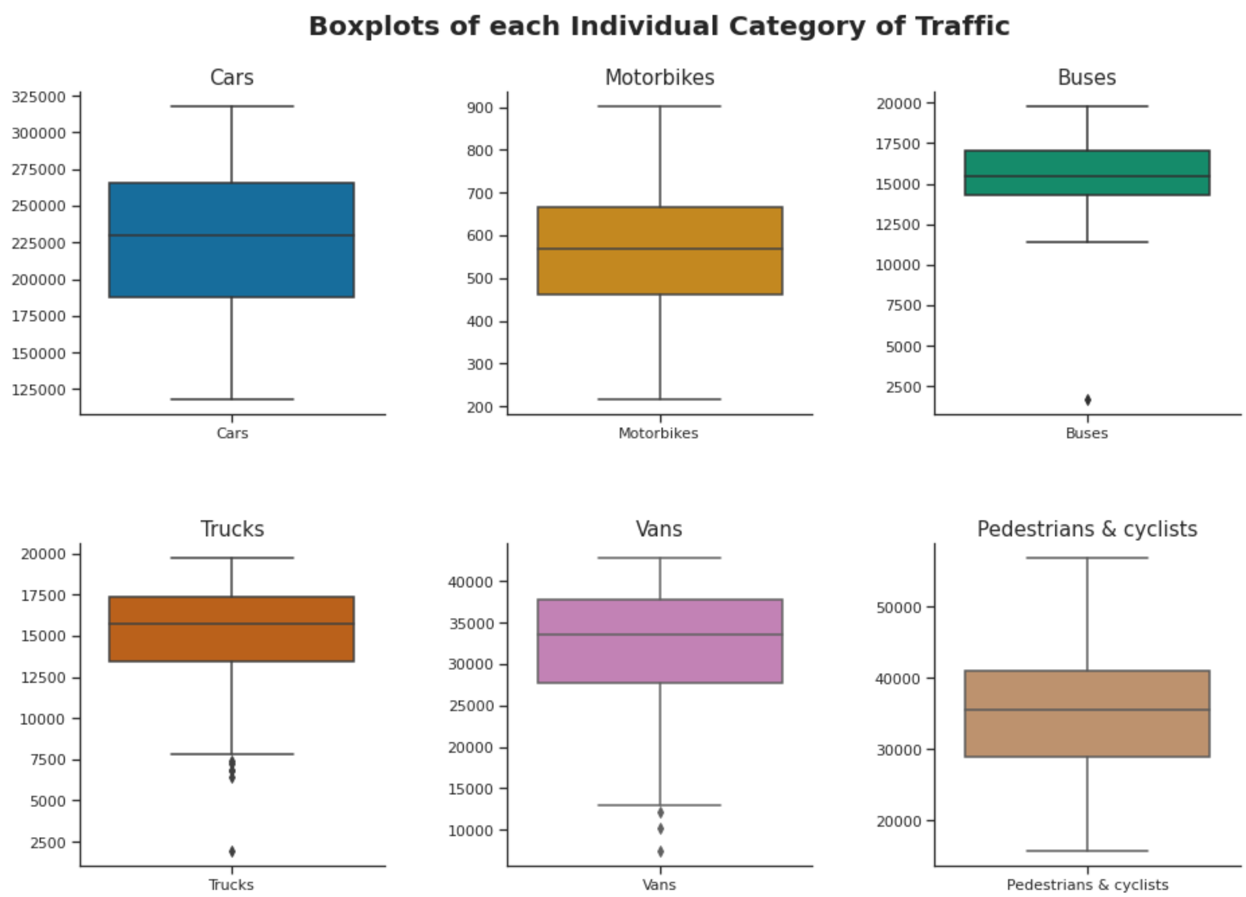
Next, it is essential to explore the data to better understand how it is distributed, what are its characteristics, whether outliers exist, whether imputations should be made, etc. (Myatt and Johnson, 2016)

The first step of exploration is the Initial Data Analysis (IDA) which is an integral part of the broader Exploratory Data Analysis (EDA) that Tukey describes as 'detective work' (Tukey, 1977, p21).

Data exploration relies heavily on visualisation because "*only visualisation can allow an explorer to 'see the whole', 'see in relation', 'look for what is recognisable', and 'attend to particulars'*" (Andrienko and Andrienko, 2006, p21)

### Checking the data distribution

To understand how data is distributed, we graph the data using a boxplot for each category of traffic activity. The boxplot allows us to visualise the smallest and the largest data points, as well as the median, first quartile, third quartile and the inter quartile range (IQR). It provides a summary of "the most frequently occurring characteristics of the pattern of a batch." (Tukey, 1977, p27). Another important function of the boxplot is to identify and visualise the outliers.



The above boxplots reveal important information about the data and about the traffic during COVID.

1. Data for Cars, Trucks and Vans is negatively skewed
2. Outliers exist in the data of Buses, Truck and Vans but Trucks and Buses have a very far outlier that requires addition investigation.
3. Traffic is generated majorly by cars, then Pedestrians & Cyclists and Vans, then Buses, Trucks and finally the Motorbikes generate the least amount of traffic.

To make sure that there are no inaccuracies in the data, we take a closer look at the outliers that were identified by the boxplot to try to understand what happened and whether they are faulty and require deletion or they are correct and require no additional intervention.

To explore these values, we use the IQR method (Sharma, 2018) which is the same method used in the boxplot.

| Date | Buses | Trucks | Vans |
|---|---|---|---|
| **2020-04-10** | - | 6863.0 | 12150.0 |
| **2020-04-13** | - | 6460.0 | 10169.0 |
| **2020-05-08** | - | 7391.0 | - |
| **2020-05-25** | - | 7241.0 | - |
| **2020-12-25** | 1654.0 | 1907.0 | 7386.0 |
| **2021-01-01** | - | 6843.0 | - |

The table above shows the values of the outliers and their dates. When the dates are cross-checked with the holidays in the UK (i team, 2020), we notice that the outliers fall exactly on these days.

The following list shows the holidays that correspond the dates where outliers were identified:

- 2020-04-10: Good Friday
- 2020-04-13: Easter Monday
- 2020-05-08: Early May bank holiday
- 2020-05-25: Spring bank holiday
- 2020-12-25: Christmas Day
- 2021-01-01: New Year’s Day

The conclusion is that it is completely normal for traffic activity to be low on days where businesses are closed, and people are not working. This is especially true for Vans, and Trucks. It's also normal for the traffic of the Buses to be very low on Christmas Day because the bus network does not operate except two special routes, the m1 and m2 (Christmas & New Year bus times, 2020). **This means that the outliers are not inaccurate and consequently, we do not need to impute or modify them.**

## Labelling the Data

In order to get better insights, and because the dataset is a time-series, labelling the data can be very helpful in understanding the traffic activity of each category at each phase of lockdown. The labels that were added to the dataset are: Pre-Lockdown 1, Lockdown 1, Pre-Lockdown 2, Lockdown 2, Pre-Lockdown 3 and Lockdown 3. Each record of the dataset containing a measure of traffic activity on a certain date, was labelled based on the phase of lockdown in which its date falls (the dates were specified in an earlier section).

Labelling can also provide a visual reference to enable an easier comparison between different time periods.

Finally, by labelling the data, we can start identifying correlations between different traffic categories at different phases of lockdowns, that are otherwise impossible if we treat the whole time-series as one continuous period.

## Smoothing the Data

In time-series, usually the values fluctuate considerably between each date and this can be difficult to interpret when plotting the data. The fluctuations are also not systematically important to visualise especially in the context of this project where we are looking at behaviour over long periods that is naturally expected to fluctuate rather than looking at small fluctuations that occur unexpectedly and that could have a significant effect.

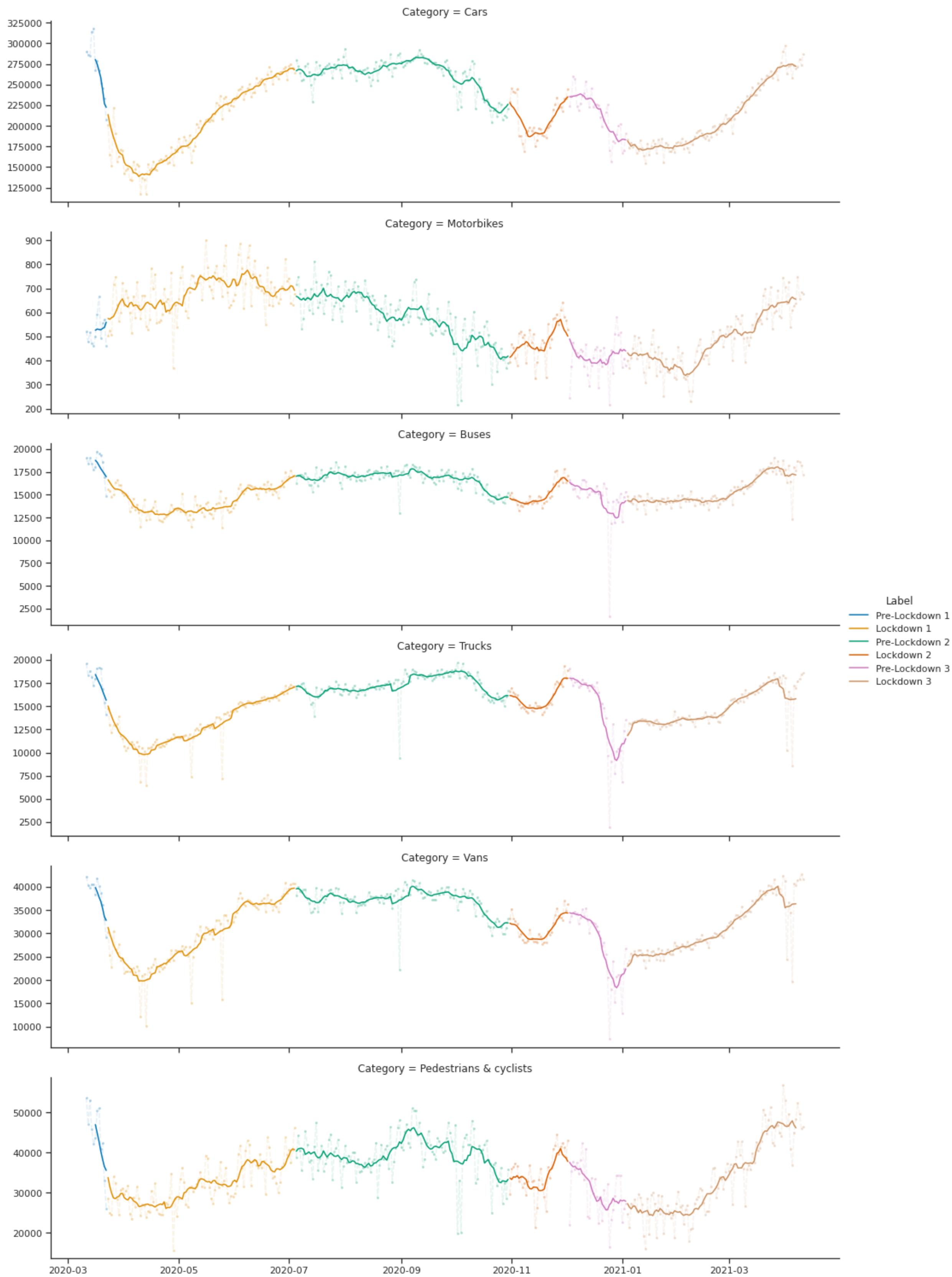
The solution that was used for the project is by applying a transformation method that evaluates data over a sliding window and then calculates a rolling average (Cleveland, 1985; Cleary, 2016; McKinney, 2017). The visualisation below was created for illustrative purposes only and shows the car traffic activity in blue before smoothing and in orange after smoothing.


## Rough Visualisation

As a last step in the exploratory data analysis, and before exporting the modified dataset to be used in Tableau for visualisation, two quick visualisations are created.

- The first visualisation is a line plot showing the time-series values for each category of traffic activity. This plot is intended to show if any final inconstancies or discontinuities exists in the data. In general, once the dataset is exported to be used in Tableau, the best practice is to have a clean version to avoid any misrepresentations and to solely focus on the design process of the data visualisation and the story-telling that will be the key to communicating the findings.

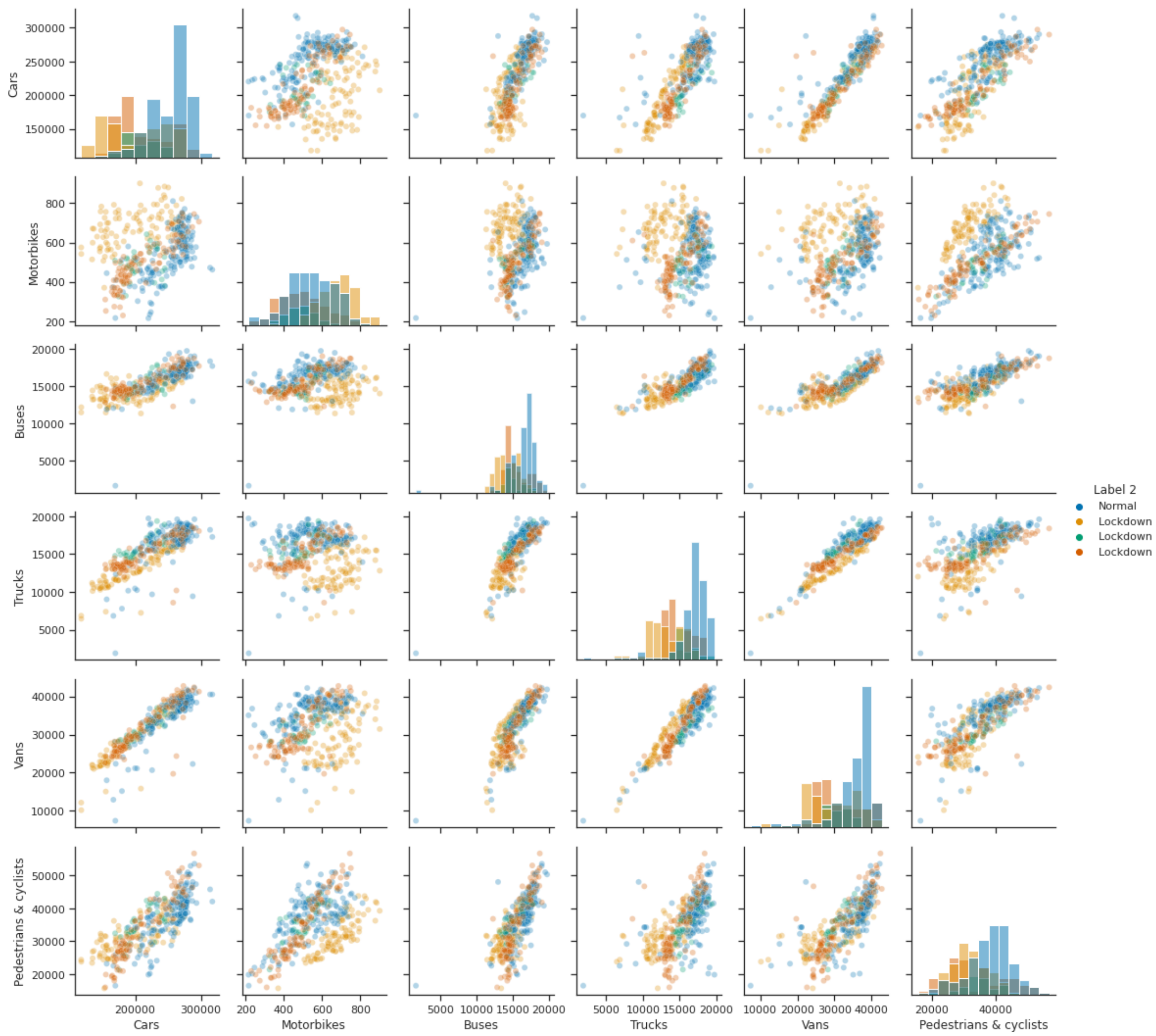
  The described plot is shown below:


Category = Cars
Category = Motorbikes
Category = Buses
Category = Trucks
Category = Vans
Category = Pedestrians & cyclists
Label
Pre-Lockdown 1
Lockdown 1
Pre-Lockdown 2
Lockdown 2
Pre-Lockdown 3
Lockdown 3
2020-03
2020-05
2020-07
2020-09
2020-11
2021-01
2021-03

- The second visualisation is a pair grid with scatterplots displaying the relationships between the different variables (the different categories of traffic activities) and on the diagonal the grid displays a histogram that shows the distribution of the data for each variable individually. The importance of this plot is to identify strong correlations between the variables such as in the case of cars and vans or cars and pedestrians, and the absence of any significant correlations such as in the case of motorbikes with all other variables.

  Awareness of these relationships is essential to understanding the data to be able to shape the story that will be told late, through the final visualisations in Tableau.

  The described plot is shown below:



## Data Visualisation

Departing from Tufte's sixth principle for the analysis and display of data, which claims that 'analytical presentations ultimately stand or fall depending on the quality, relevance, and integrity of their content' (Tufte, 2006, p136), and given the ubiquity of data related to COVID-19, an innovative dataset was chosen for this project which relies on unconventional data collection mechanisms. However, and in order to drive meaning from the data, it is equally important to consider the other 5 principles of visual integrity that Tufte (2006) proposed. Additionally, considering that such data is a critical decision-making tool for the UK government when fighting a global pandemic, it is important to visualise the data beyond the bare 'representational' aim, in the sense of simply communicating numeric values, and instead make the visualisations 'significatory', making them "constitutive of a reality rather than as a mere reflection of it" (Pryke, 2010, p435). A last element to consider when visualising decision-making graphs is what Pollock and Campagnolo (2015) call the 'subitizing effect' which lies in capacitating the viewers to "make rapid, accurate, and confident judgements" (Pollock and Campagnolo, 2015, p4) when looking at simpler graphs with smaller number of elements, allowing them to arrive easily at certain conclusions or decisions.

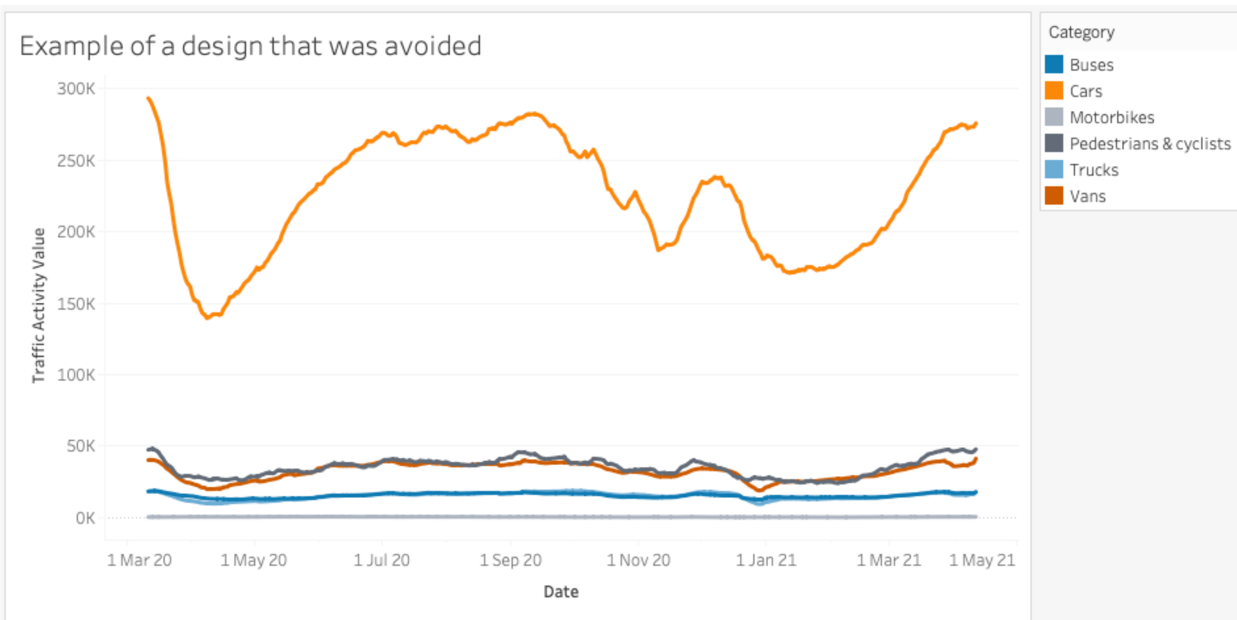
### Visual Encoding and Data Representation

Kirk (2016) describes encoding as "the right blend of marks and attributes that most effectively will portray the angle of analysis you wish to show your viewers". In his description he emphasises the need to choose the right visualisations for the right audience. On that basis, and adjunctly with the earlier ideas of significatory and subitizing function of the graphs, as well as the criticality of the analysed data for timely decision-making, the choice of visual encoding can be narrowed down to choosing a functional visualisation encoding that is able to convey an explanatory portrayal of the available data to a knowledgeable audience. It is therefore necessary to have a clean, and pragmatic tone for the visuals, avoiding any embellishment and chart-junk.

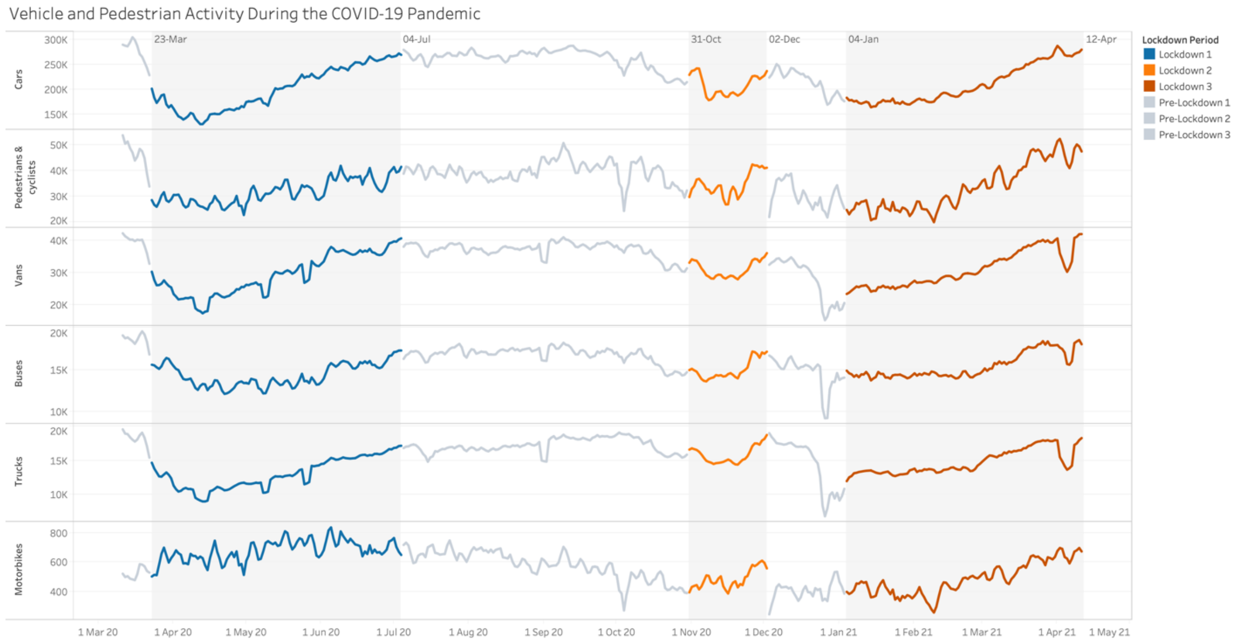
For this project, two sets of visualisations have been produced:

- **The first is for the exploratory data analysis:** In this part, tables were mostly used to ensure the technical validity of the data and to verify the correctness of the operations applied on the data such as transformations, indexing, dropping columns, etc.
  Boxplots were also used to quantitively visualise the data and identify any issues or outliers.
  Lastly, line plots were used to visualise the time-series and to explore the significance of changes over time for the different category of traffic such as pedestrians, cars, vans, etc. Data visualisation in this section were generated for quick and rough observations, so not much emphasis has been spent on the design process. Nevertheless, as much as practically possible, axis were properly labelled, all colours were colour-blind friendly, and data-ink ratio remaining reasonably high.
- **The second set of visualisations is the one intended for the final findings:** For these visualisations, Tableau Desktop Professional Edition version 2020.4.2 was used, and careful consideration was taken with the choice of graphs and the layout.

- Firstly, all graphs are created with a colour palette specific for colour blindness. This is particularly import for people who have colour deficiency and who constitute 10% of the world population. (Kirk, 2012, p100)
- Secondly, because the main data is a time-series, and because the aim of the project is to analyse “how quantitative values for different categories have changed over time” (Kirk, 2016, p225), a line chart was used to visualise the value of the data points using joined-up lines.
  In order to compare the different traffic categories over time, It would have been possible to plot the values for traffic activity over one chart. However, this would have resulted in cluttering the data region, making it difficult to read the values for categories that have similar values such as pedestrians and vans, and obscuring the values that are closer to zero such as motorbikes (Cleveland, 1985). The image below illustrates the design decision that was avoided.




- Alternatively, to eliminate the clutter in the data region, the data was graphed on “juxtaposed panels” (Cleveland, 1985, p40) as shown in the image below:




The trends of Moving Average of Cars, Moving Average of Pedestrians & cyclists, Moving Average of Vans, Moving Average of Avg. Buses, Moving Average of Trucks and Moving Average of Motorbikes for Date Day. Colour shows details about Lockdown Period.

- In line with Cleveland's recommendation for "not overdo[ing] the number of tick marks" (Cleveland, 1985, p39), Tableau was configured to show the dates in a 'day month year' format with minimal amount of Major ticks representing the first of each month and without additional minor ticks, leaving us with only "the most useful and meaningful interval for [...] time axis labelling" (Kirk, 2016, p226)
- Because "line charts do not always need the y-axis to start at zero" (Kirk, 2016, p227) and because the zero baseline does not have any significance in the interpretation of our data, it was removed from all the graphs, allowing an superior visual representation of the smaller variations of data which would have been otherwise compressed if the y-axis started at zero.
- To help the readers in decoding faster the visualisation, a mapping technique was used on the graph, which according to Tufte, allows the visualisation to "outperform purely pictorial representations for presenting, explaining and documenting evidence" (Tufte, 2006, p13). In the logic of mapping, and in order to create a sense of guiding lines, a grey background was added under the lockdown periods to add contrast between them and the non-lockdown periods. Also, the non-lockdown data points were coloured in light grey to deemphasise them.

## Testing and Improving the Visualisation

When the plot with juxtaposed graphs was tested with two different persons from different backgrounds, it turned out to be a little bit more complicated to be decoded.

**The first respondent**, a female in her 30s and works in the media industry, when she was presented with the visualisation, without any additional explanation and was asked to explain what she sees, she was able to clearly explain the function of the graph however, few points arose:

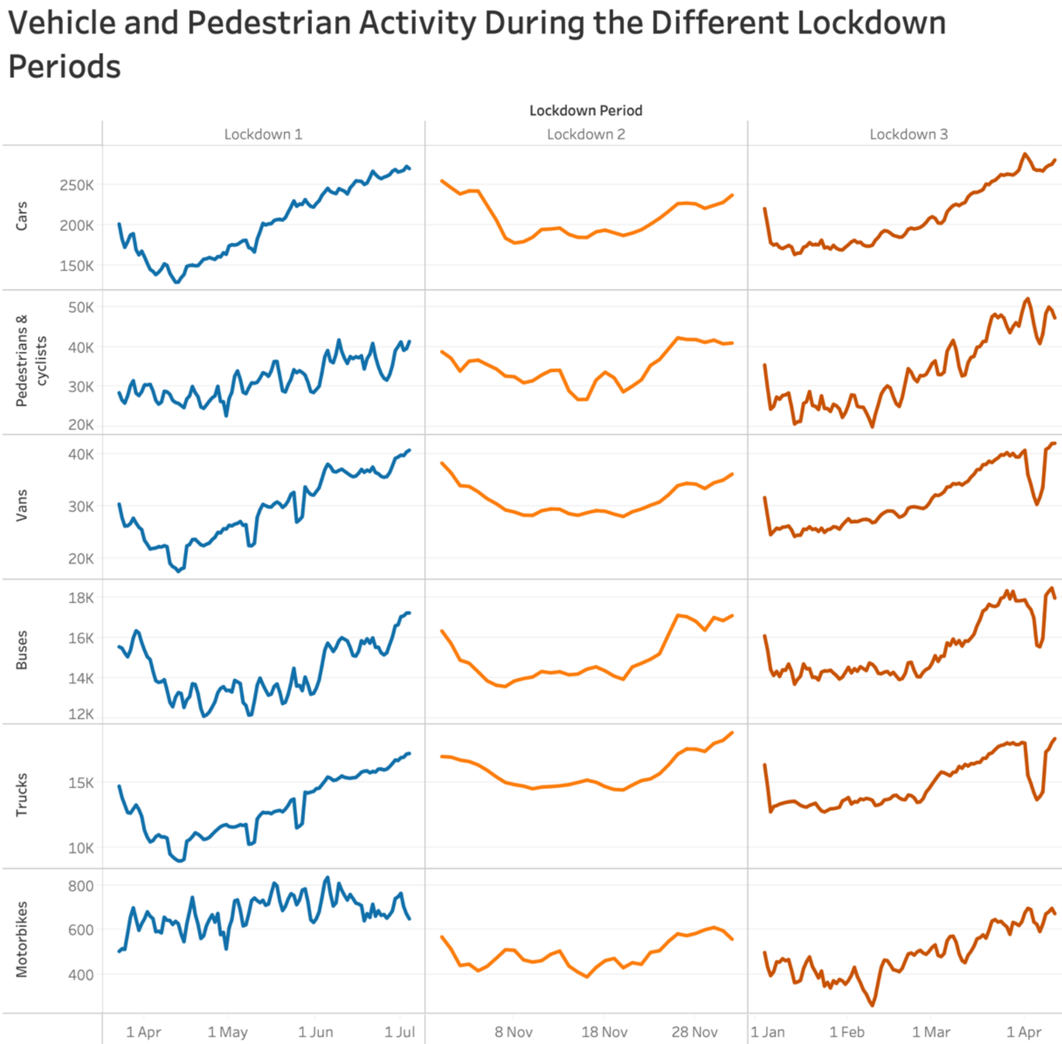
- She didn’t notice the difference between the coloured (lockdown periods) and the greyed-out areas (non-lockdown periods).
- She inquired about the large dips, mostly occurring in the non-lockdown periods.
- She mentioned that the graph requires a substantial amount of concentration to be properly read.
- She interpreted the values of traffic activities as literal number of cars, pedestrians etc. However, the numbers are only quantitative measure of busyness.

**The second respondent**, a male in his 30s and works as computer engineer, was presented with the visualisation in the same way as the first respondent:

- He found it easy to read the graph but found it a bit cluttered.
- He complained about the curves being vertically squeezed and not showing the full extent of variation in data.
- He argued that it’s easy to interpret the graph in terms of traffic behaviour, however, he found It difficult to give any justification about what triggers such behaviour because the data analysed is not placed within the whole COVID-19 context. Instead, it is being used in isolation of other important data that could give us a broader perspective about the situation.

Based on the above feedback, I decided to simplify the graph as much as possible by removing more parts that turned out to be less important for the readers and for the research question. Consequently, the non-lockdown phases were removed, the 3 lockdown phases were emphasised even further, and a caption was added as a disclaimer about the busyness indicator.

The resulting visualisation became as follows:



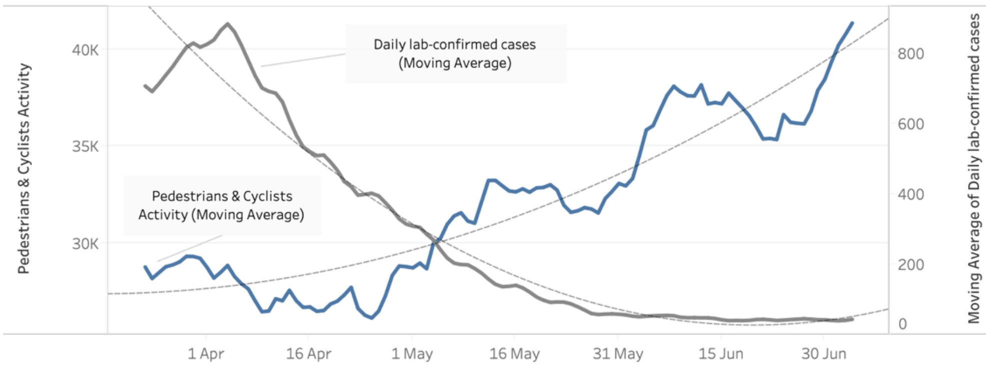

The numbers representing the activity of Cars, Pedestrians & Cyclists, Vans, Buses, Trucks and Motorbikes are quantitative measures of **busyness** and depict trends but should not be used for estimating any precise or even approximate amount of vehicle or pedestrian movements.

In a **second round of feedback**, when the same respondent was asked about the new chart, it was clearly easier for her to read it quicker and understand it without much effort. Unfortunately, by fixing the chart, a new issue arose in the proportions of the x-axis sections of lockdown periods. The proportions are not faithful to the actual length of each lockdown period, and while the second lockdown is supposed to be the shortest, it occupies as much space as the first and second lockdown. It was a technical challenge to solve this issue with Tableau and sadly I had to drop it due to time constraints.

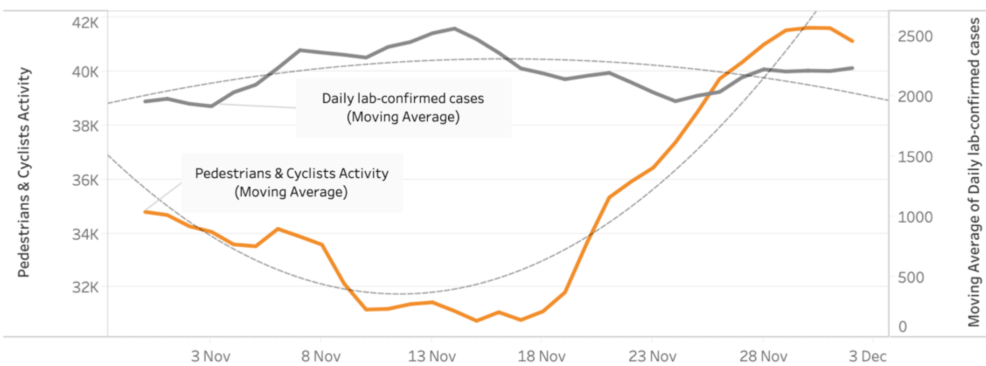
**To address the second feedback**, an additional dataset was added to the project to show the daily number of COVID-19 positive cases. The dataset can be downloaded here: https://coronavirus.data.gov.uk/details/download

The time-series of the positive cases was then plotted with each category of traffic at each phase of lockdown. The colour coding of each lockdown phase was kept for better visual continuity. The image below represents one example of the new visualisation that was added, and that illustrations the pedestrian and cyclists activity in comparison with COVID-19 confirmed cases:

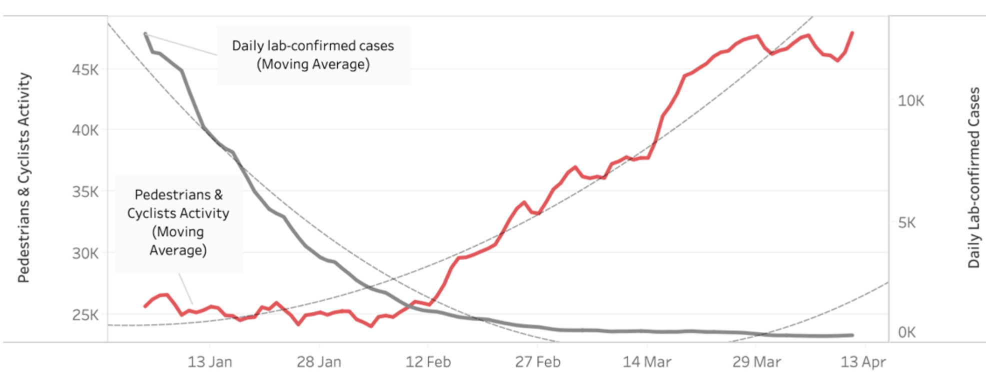
Pedestrian & Cyclists Activity In Lockdown 1 Compared to the Daily Number of COVID-19 Positive Cases




Pedestrian & Cyclists Activity In Lockdown 2 Compared to the Daily Number of COVID-19 Positive Cases




Pedestrian & Cyclists Activity In Lockdown 3 Compared to the Daily Number of COVID-19 Positive Cases



## Key Findings

The departure point of the research was to figure out what insights can be inferred on mobility in London, during the three lockdown periods, by analysing data collected on vehicle and pedestrian activity from traffic cameras.

When the traffic activity is plotted for every category (Cars, Vans, etc..) and for each lockdown period, the most visible and obvious insight is the similarity between all traffic categories in terms of having an ascending trend of increased activity over time.

It seems that at the beginning of each lockdown, when it is first announced, activity of all traffic drops down very quickly to an absolute minimum, then gradually starts picking up even though the lockdown is still in place. What can be deduced from this is that the lockdown is only efficient at reducing the activity of people and vehicles for a truly short time. It is not an efficient long-term solution since as the graphs show, people and vehicles resume gradually their activity regardless of the government's lockdown measures.

It is also noticeable how little effect the lockdowns had on motorbikes whose activity was the least affected by the lockdown. In fact, if we analyse the post-lockdown periods, the motorbikes' activity is lower than during the lockdown. This could be due largely to the fact that motorbikes in London are mostly business owned and their main function is the delivery so of goods so the demand increases on them during the lockdown and when the country reopens, the demand on home deliveries decreases so their activity decreases consequently.

In terms or ranking, cars activity is the highest, and it generates 80% more traffic than any other traffic source. Pedestrians and cyclists come second after cars, then vans, buses and trucks.

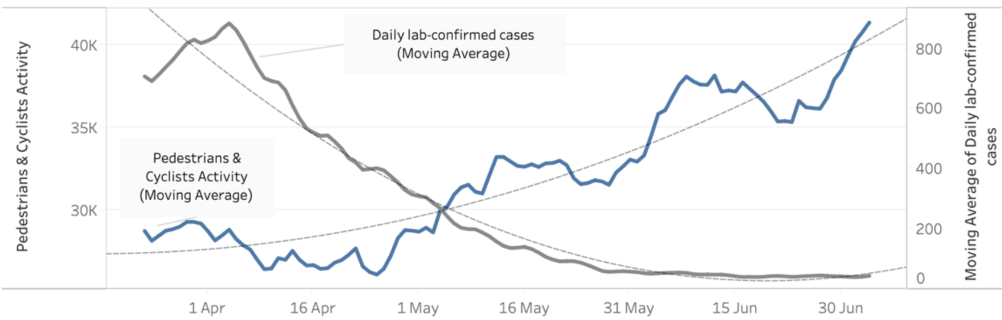
There's a very strong correlation between the activity of car, trucks and vans. This implies that lockdown measures when implemented, they affect the three of them in a similar manner, unlike Pedestrians and motorbikes who are less affected.

The most important finding in this dataset is actually not the conclusions that can be drawn but rather the questions that it raises about the effectiveness of the lockdown measures, the behaviour of people over time, the economic questions of offer, demand, supply chains and fulfilment. For instance, the assumption for trucks and vans is to have a similar activity to the motorbikes since the demand on supply chains increases during lockdown because people start stocking food and groceries. But the reality is different as shown through this graph.

To be able to answer these questions, it's important to look at this data wholistically with data from other areas such as economic data. By doing so, the data can be interpreted in a broader context that reflects better the core reasons behind the observed phenomenon.

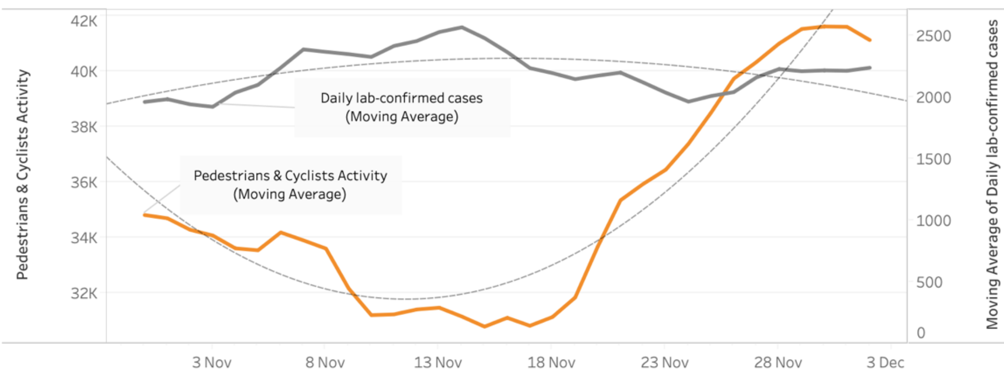
As an example, I included an additional dataset showing the number of daily lab-confirmed positive cases. When I placed this data with the traffic activity data, new findings surfaced. A sample of the generated visualisations is the one for pedestrians and cyclists and it is shown below:

Pedestrian & Cyclists Activity In Lockdown 1 Compared to the Daily Number of COVID-19 Positive Cases

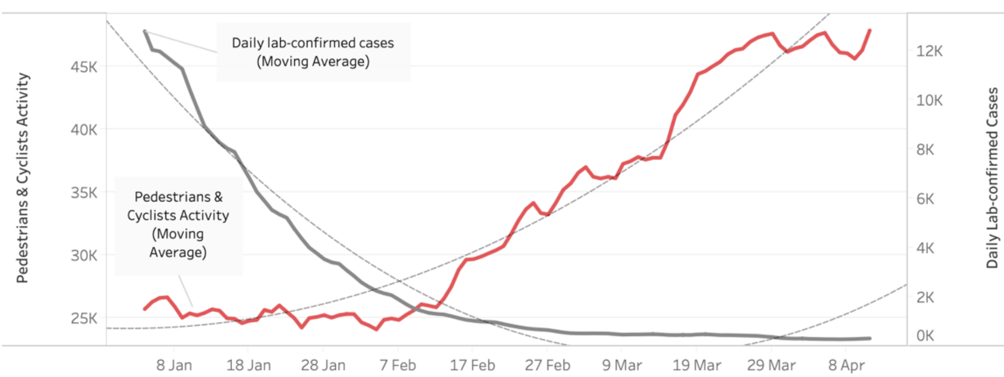



The numbers representing the activity of Cars, Pedestrians & Cyclists, Vans, Buses, Trucks and Motorbikes are quantitative measures of **busyness** and depict trends but should not be used for estimating any precise or even approximate amount of vehicle or pedestrian movements.

Pedestrian & Cyclists Activity In Lockdown 2 Compared to the Daily Number of COVID-19 Positive Cases




Pedestrian & Cyclists Activity In Lockdown 3 Compared to the Daily Number of COVID-19 Positive Cases




What the above visualisation shows is that the activity of pedestrians and cyclists is highly inversely correlated with the daily lab-confirmed cases, therefore, when the number of cases start to drop, the activity picks up again. Another very important finding is that the absolute value of the number of positive cases is not important by itself, and we can see that in lockdown 1 when the max number of daily cases was at its peak 800 and then dropped, the pattern of pedestrian and cyclists activity was the same as lockdown 2 when the number of positive cases peaked at 12,000 daily cases. However, in contrary to what was assumed in the previous section about the inefficiency of the lockdown on traffic, it seems to have a positive effect in reducing the number of cases. The catch from this graph is the assumption that the activity of pedestrians and cyclists is more affected by the hype around the increase and decrease in the number of cases regardless of the actual number itself. This hype has so much to do with the traditional media, social media and the other informal communication tools that are used to spread such information. This assumption, however, requires additional data to be better understood.

## Conclusion

As much as it can be tempting to make inferences and come out with conclusions by looking at a dataset and some data visualisations, it is equally important to be very careful about what can be deduced and how the information should be presented. That's because there's a big responsibility when making statements about existing data and it can easily be misinterpreted and used to spread misinformation. Also, when analysing data, the context should be very clear and data should be looked at in a multi-dimensional manner, and not as an isolated set of measures. The previous section clearly demonstrates this conclusion and proves the need to look at the data and around it if one was to better understand it.

## References

Adams, Z. (2021). Roadmap out of lockdown: Key dates for lockdown lifting in the UK.

Andrienko, N. and Andrienko, G. (2006). *Exploratory Analysis of Spatial and Temporal Data*. Springer.

Christmas & New Year bus times. (2020). Available from https://www.morebus.co.uk/christmas# [Accessed 19 April 2021].

Cleary, M. (2016). *Python Data Science Handbook*, First. O'Reilly Media, Inc.

Cleveland, W.S. (1985). *The Elements of Graphing Data*. Wadsworth Advanced Books and Software.

Edwardes, A. (2020). Estimating vehicle and pedestrian activity from town and city traffic cameras. *Data science for public good*. Available from https://www.turing.ac.uk/research/research-projects/project-odysseus-understanding-london-busyness-and-exiting-lockdown#main-content [Accessed 10 April 2021].

i team. (2020). When is the next bank holiday? August long weekend date – and the full list of 2020 UK holidays. Available from https://inews.co.uk/news/uk/bank-holiday-2020-when-next-august-monday-uk-holidays-378790 [Accessed 15 April 2021].

Kirk, A. (2012). *Data Visualization: A Successful Design Process*. Packt Publishing Ltd.

Kirk, A. (2016). *Data Visualisation: A Handbook for Data Driven Design*. SAGE Publications Ltd.

Mazzi, G.L., Ladiray, D. and Rieser, D.A. (2018). *Handbook on Seasonal Adjustment 2018 edition*. Publications Office of the European Union. Available from https://doi.org/10.2785/279605.

McKinney, W. (2017). *Python for Data Analysis*, Second. O'Reilly Media, Inc.

Myatt, G.J. and Johnson, W.P. (2016). *Making sense of data I - A Practical Guide to Exploratory Data Analysis and Data Mining*, Second. John Wiley & Sons, Inc.

Office for National Statistics. (2020). Using traffic camera images to derive an indicator of busyness: experimental research. Available from https://www.ons.gov.uk/economy/economicoutputandproductivity/output/methodologies/usingtrafficcameraimagestoderiveanindicatorofbusynessexperimentalresearch [Accessed 11 April 2021].

Pa Reporters. (2021). Covid timeline: Key moments in the past year since the UK's first lockdown. Available from https://www.independent.co.uk/news/uk/home-news/covid-timeline-uk-boris-johnson-b1820412.html [Accessed 15 April 2021].

Pollock, N. and Campagnolo, G.M. (2015). Subitizing Practices and Market Decisions. *Making Things Valuable*, 15 (1), 89–108. Available from https://doi.org/10.1093/acprof:oso/9780198712282.003.0005.

Pryke, M. (2010). Money's eyes: The visual preparation of financial markets. *Economy and Society*, 39 (4), 427–459. Available from https://doi.org/10.1080/03085147.2010.510679.

Sharma, N. (2018). Ways to Detect and Remove the Outliers.

Tufte, E.R. (2006). Beautiful evidence. 213.

Tukey, J.W. (1977). *Exploring Data Analysis*, Second. Addison-Wesley Publishing Company.