# London Fire Brigade: Introducing the London Fire Brigade Dataset

Tom Eisenbeiß, Christopher Guth, Nico Heinrich, Therese Riedemann

## Introduction

The London Fire Brigade (LFB) is the busiest fire and rescue service in the country and one of the largest fire and rescue organisations in the world. It was established in 1866 as a direct consequence of a devastating fire at London’s Cotton Wharf in 1861. This fire raged for 2 weeks before being extinguished, resulting in a huge financial burden on insurance companies that traditionally employed firefighters at the time.

Insurance companies then requested a public London firefighting service to relieve them of their responsibilities and control their escalating compensation costs leading to the Metropolitan Fire Brigade Act being passed in 1865. Accordingly, London firefighters are now public servants. The LFB is headed by the London Fire Commissioner. However, certain decisions (e.g. the LFB’s annual budget and the Community Risk Management Plan) need additional approval by the Mayor of London. The work of the LFB is monitored by the Fire Committee of the London Assembly and by His Majesty’s Inspectorate of Constabulary and Fire & Rescue Services. In addition, a number of specialised boards within the LFB manage special aspects of the LFB work, such as performance or finances.

The LFB operates under a tight budget, with a forecasted annual budget of around £500 m in 2025/2026. Roughly 75% of this budget is spent on current and retired personnel. Around 13% of the budget is spent on the maintenance of premises and supplies and services. Most of the remaining budget is spent on transportation. The budget plan of the LFB faces 2 specific challenges: 1) It is assessed by the LFB’s routine work and, 2) the LFB’s capacity to quickly respond in major and resource-intensive incidents, despite being rare, such as the Grenfell Tower incident or multi-sited events in the midst and aftermath of the St. Jude storm should not be compromised by the budget plan.

Therefore, a particular focus of the LFB’s work is a 1) reduction of incident numbers by e.g. charging mechanisms of false alarms or by educational work of firefighters during Home Fire Safety Visits and 2) a reduction of attendance times, i.e. the time interval between the emergency call and the arrival at the incident site. An acceleration of attendance times calls for the maintenance of the highest possible number of fire stations on the one hand and faster operational procedures on the other. Noteworthy, operational procedures are partially beyond control (e.g. acts of nature and/or traffic).

The present report analyses all documented incidents and mobilisations of the LFB between the years of 2009 and 2024. We identify time, site-, and service-dependent fluctuations of incident numbers and time-, weather-, traffic- and equipment-dependent changes of LFB attendance times that lays the ground for a future model-based optimisation of incident numbers and mobilisation times

## Objectives

The main goal of the present project is a prediction of mobilisation times and incident numbers of the LFB on the basis of variables such as time, traffic, location, equipment and service type.

The LFB was contacted to gain a better insight into certain categories of the dataset (e.g. specification of national costs), a reply is still pending.

## Framework

The data for the current study were obtained from the LFB website and are publicly available ([London Fire Brigade Incident Records](https://data.london.gov.uk/dataset/london-fire-brigade-incident-records), [London Fire Brigade Mobilisation Records - London Datastore](https://data.london.gov.uk/dataset/london-fire-brigade-mobilisation-records)).

Specifically, the following datasets were the basis of the present report:

1. [LFB Incident data from 2009 - 2017 (CSV, 313 MB)](https://data.london.gov.uk/download/london-fire-brigade-incident-records/73728cf4-b70e-48e2-9b97-4e4341a2110d/LFB%20Incident%20data%20from%202009%20-%202017.csv)
2. [LFB Incident data from 2018 onwards (CSV, 256 MB)](https://data.london.gov.uk/download/london-fire-brigade-incident-records/f5066d66-c7a3-415f-9629-026fbda61822/LFB%20Incident%20data%20from%202018%20onwards.csv)
3. [LFB Mobilisation data from January 2009 - 2014 (XLSX, 118 MB)](https://data.london.gov.uk/download/london-fire-brigade-mobilisation-records/fcbd2e97-b5bf-4117-a50f-d596181bc8d3/LFB%20Mobilisation%20data%20from%20January%202009%20-%202014.xlsx)
4. [LFB Mobilisation data from 2015 - 2020 (XLSX, 113 MB)](https://data.london.gov.uk/download/london-fire-brigade-mobilisation-records/329ef4d3-7ebb-4198-8bcf-92ad6711e50d/LFB%20Mobilisation%20data%20from%202015%20-%202020.xlsx)
5. [LFB Mobilisation data 2021 - 2024 (CSV, 152 MB)](https://data.london.gov.uk/download/london-fire-brigade-mobilisation-records/3ff29fb5-3935-41b2-89f1-38571059237e/LFB%20Mobilisation%20data%20from%202021%20-%202024.csv)

## Description of data

The incident data and the mobilisation data were each merged into one respective datasheet and are described below.

### Incidents dataset

The LFB incident datasheet consists of 39 columns and 1,793,315 rows. Only 4 columns contain quantitative variables. These 4 columns represent the target variables (1. FirstPumpArriving\_AttendanceTime, 2. SecondPumpArriving\_AttendanceTime, 3. PumpMinutesRounded and 4. National Cost (£)). The majority of columns (30 out of 39 columns) consists of categorical variables, most of them containing more than 10 categories. The remaining columns are either date variables (4 columns) or unique variables (1 column). On average, the dataset contains 11.4% of missing cells, however, the number of missing cells is distributed unevenly across the dataset.

The variables containing no or less than 1% of missing cells are:

1. Date of call
2. Call year
3. Time of call
4. Hour of call
5. Incident Group
6. Stop code description
7. Property category
8. Property type
9. Address qualifier
10. Postcode district
11. UPRN
12. USRN
13. IncGeo borough code
14. IncGeo borough name
15. Proper Case
16. IncGeo ward code
17. IncGeo ward name
18. Easting rounded
19. Northing\_rounded
20. FRS
21. Incident station ground
22. Num station with pumps attending
23. Num pumps attending
24. Pump count
25. Pump minutes rounded
26. National cost (£)
27. Num calls

The variables containing 50% or more missing cells are:

1. Special service type
2. Postcode full
3. Easting\_m
4. Northing\_m
5. Latitude
6. Longitude
7. Second pump arriving attendance time
8. Second pump arriving deployed from station

### Mobilisation dataset

The LFB mobilisation datasheet consists of 24 columns and 2,496,800 rows. Quantitative variables are present in 3 columns (1. TurnoutTimeSeconds, 2. TravelTimeSeconds, 3. FirstPumpArriving\_AttendanceTime, 2. SecondPumpArriving\_AttendanceTime, 3. AttendanceTimSeconds). The majority of columns consists of categorical variables, most of them containing more than 10 categories. On average, the dataset contains 15.0% of missing cells, the number of missing cells being distributed unevenly across the dataset.

The variables containing no or less than 1% of missing cells are:

1. Call year
2. Hour of call
3. Resource mobilisation Id
4. Resource code
5. Performance reporting
6. Date and time mobilised
7. Date and time mobile
8. Date and time arrived
9. Attendance time seconds
10. Deployed from station code
11. Deployed from station name
12. Deployed from location
13. Pump order
14. Plus code code
15. Plus code description
16. Easting rounded

The variables containing 50% or more missing cells are:

1. Delay code Id
2. Delay code description
3. Borough name
4. Ward name

## Relevance

The following groups of variables were considered most relevant to this project: temporal variables (e.g. Month of Call), incident characteristics (e.g. Incident Group), property characteristics (e.g. Property category), operational constraints (e.g. Delay code Id).

## Visualisations and Statistics

All data were prepared, processed and visualised using Jupyter Notebook. Data visualisation was performed with the help of the Matplotlib and Seaborn package, statistical analyses were performed using the scipy.stats module.

## Pre-Processing and feature engineering

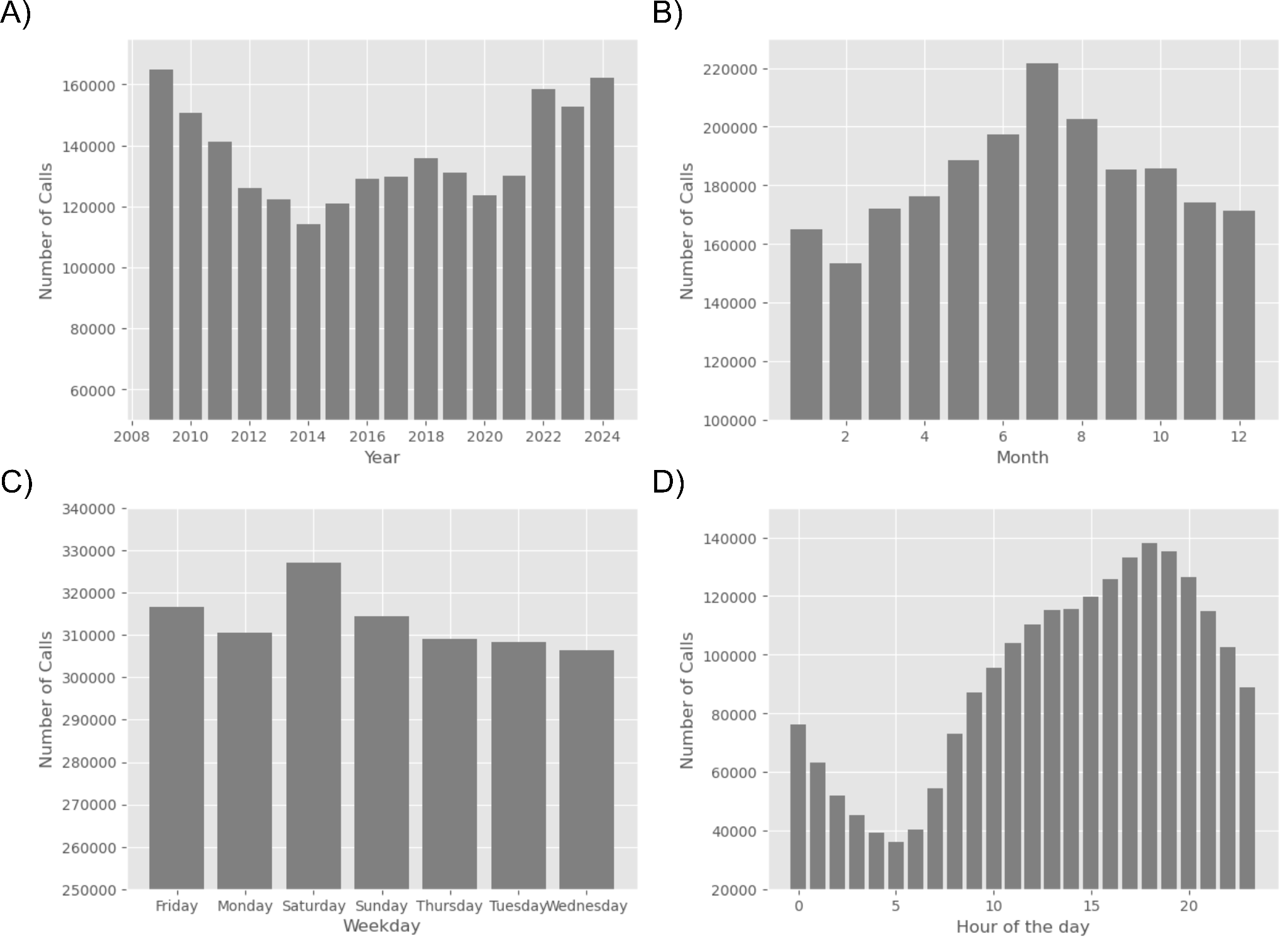
Data processing was performed in a stepwise fashion. The first step consisted of an exploration of the dataset using common functions from the pandas library. Months and weekdays were derived from the category “Date of Call”, time units of all mobilisation time variables were changed from seconds to minutes. Next, we performed an in-depth data quality profiling using YData Profiling, however its usage was discontinued due to RAM limitations. Following data exploration, the dataset was cleaned from missing values and from redundant categories. Identification of errors in the data (such as outliers or possible typographical errors), encoding and feature selection is still ongoing and the datasets are updated continuously.

## Results

### 1) Time-dependent incident fluctuations

We initially investigated time-dependent fluctuations in incident numbers. To this end, we analysed the number of emergency calls as a function of the year, month, weekday and hour of the day (Fig. 1) and conducted ANOVA tests on the data. To ensure ANOVA is applicable, we transformed the data to represent the number of incidents per hour for each hour. We found that the number of incidents is significantly influenced by all time units investigated (year, month, weekday and hour), the *p*-values being extremely small (largest of the 4 *p*-values: ).

Therefore, we conclude that time plays a crucial role in the number of incidents.

Figure 1. Histogram showing year-dependent (A), month-dependent (B), weekday-dependent (C) and hour of the day-dependent (D) fluctuations in emergency call (call) numbers.

### 2) Site-dependent incident fluctuations

Next, we analysed site-dependent influences on the number of emergency calls. In this analysis, the term ‘site’ encompasses London boroughs, property category and property type. Therefore, we plotted the number of emergency calls as a function of London boroughs, of property type and property category (Fig. 2). The highest number of incidents is observed in Westminster, followed by Camden, Southwark and Tower Hamlets. The lowest incident numbers are found in City of London (Fig. 2 A). In order to test a correlation between population density and incident numbers, we took the population density numbers of individual London boroughs according to the Office for National Statistics ("Mid-Year Population Estimates, UK, June 2022") and performed a Pearson correlation test with the numbers of incidents per London borough. The data show a clear correlation between population density and emergency call counts (correlation coefficient: 0.57, *p*-value = 0.0006, Fig. 2 B).

In addition, we analysed the number of incidents according to property category and property type. We found that dwellings account for nearly 50% of all incidents, followed by non-residential buildings, indicating that most LFB operations are indoor operations. Altogether, outdoor operations (incl. road vehicles) account for roughly 20% of all incidents (Fig. 2 C).

In agreement with above data, the highest number of incidents occurs in private housing, either in blocks of flats or single houses (Fig. 2 D).

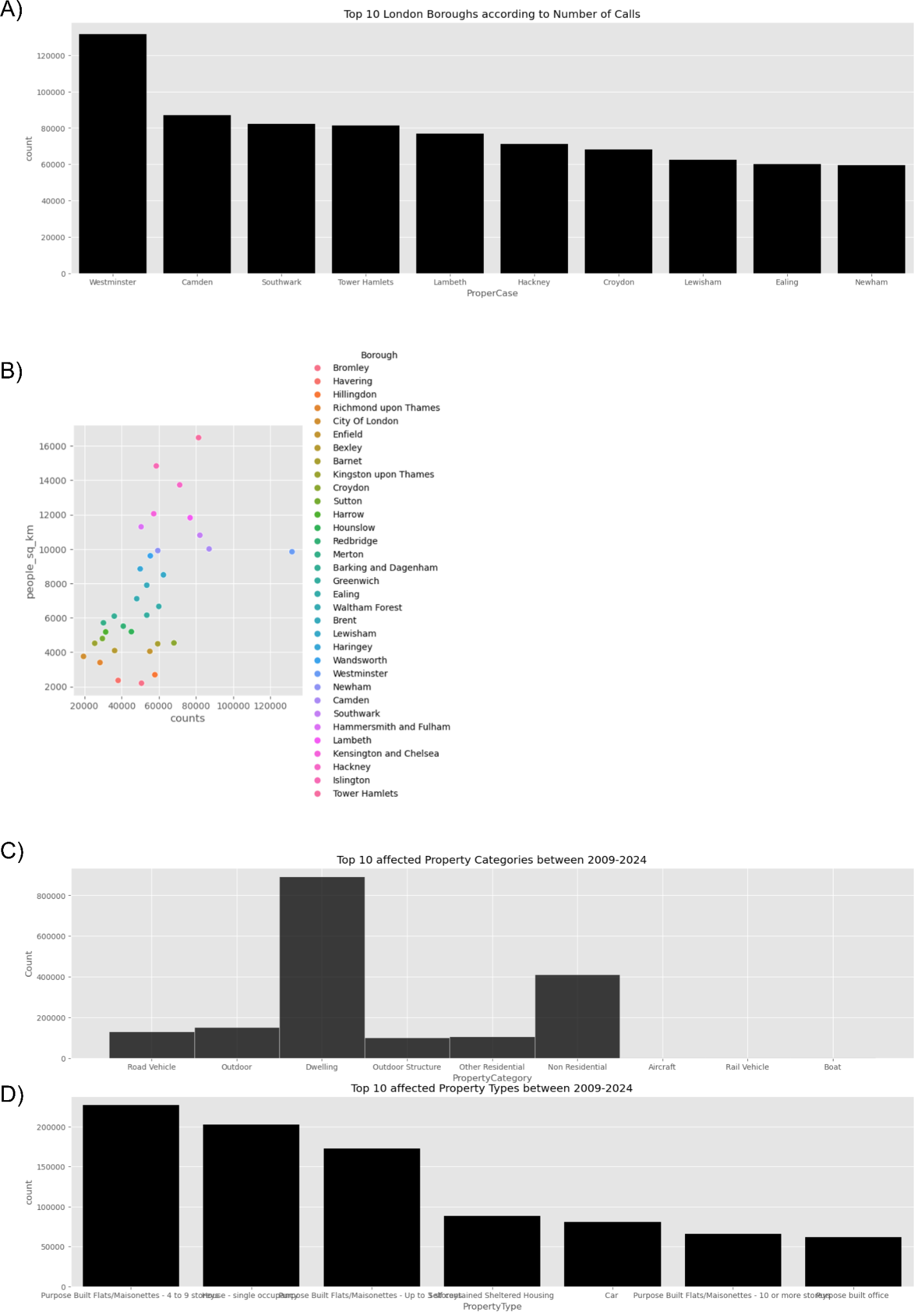


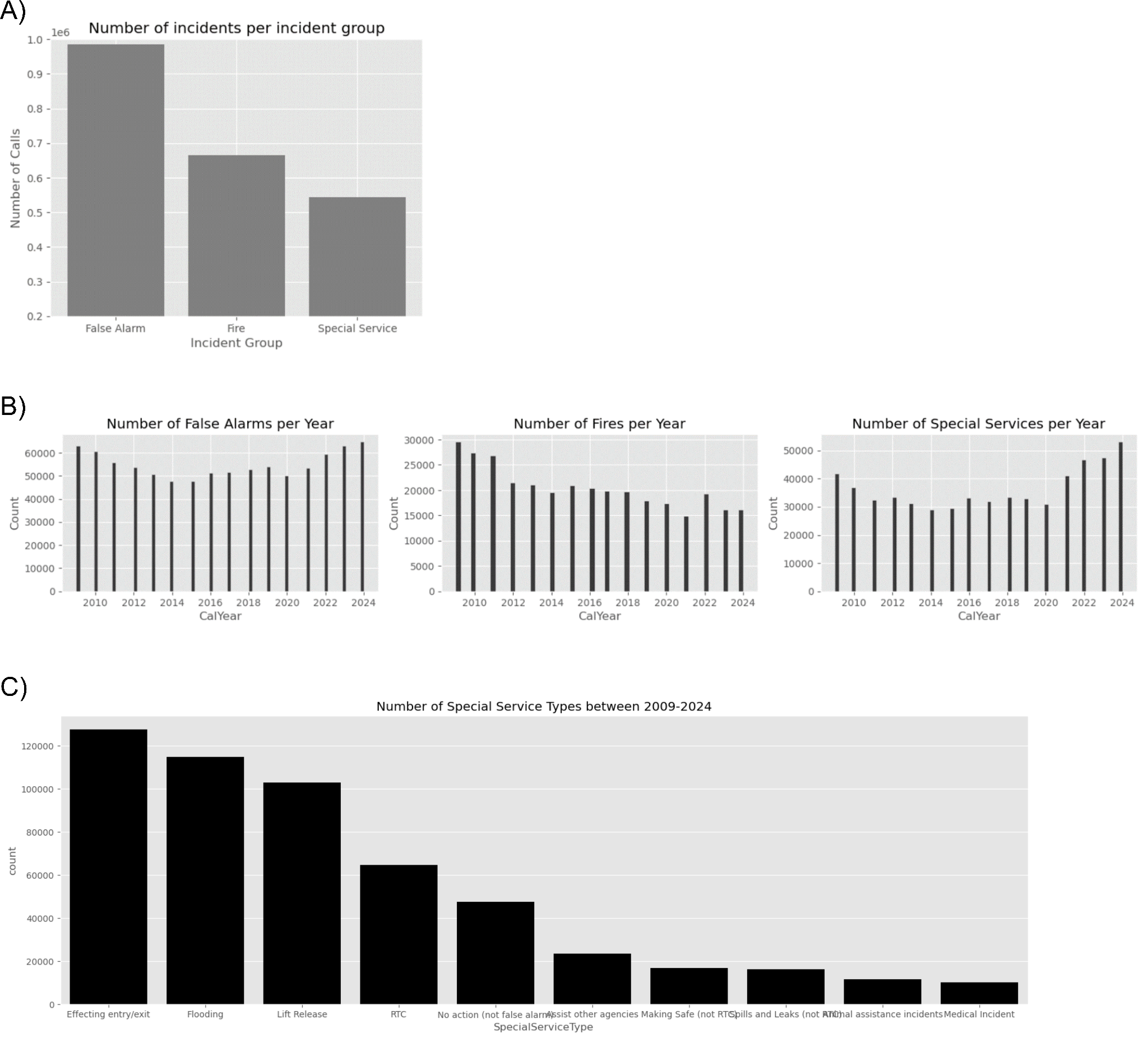
Figure 2. Histogram showing the top 10 number of incidents in individual London boroughs (A). The scatter plot depicts the population density in individual London boroughs (people\_sq\_km) as a function of the number of incidents (B). Histogram showing the number of incidents according to property category (C) and property type (D).

### 3) Service-dependent incident fluctuations

Next, we analysed the number of incident groups of all recorded incidents. False alarms account for nearly 45%, fires for roughly 30% of all recorded incidents. Special services (non-fire operations) account for the remaining 25% of incidents (Fig. 3 A). A closer look at the evolution of false alarm, fire and special service numbers per year reveals that the number of false alarms could successfully be reduced with the introduction of charging repeated false alarm offenders in the early 2010s, and increased again with the lift of charge management in recent years. The numbers of fires show a steady decline from 2009 to 2024. In contrast, the numbers of special services exhibit a steady increase since the year 2020 (Fig. 3 B).

Looking into special service types, we found that the most frequent service provided by the LFB is that of effecting entry or exit, followed by flooding, lift releases and road traffic collisions (RTC, Fig. 3 C).

The data clearly show that false alarms account for a high number of emergency calls and we hypothesize that reductions in false alarms will help to increase the efficiency and financial scope of LFB operations.

Figure 3. Histogram showing the number of all recorded incidents per incident group (A). Histograms showing the numbers of false alarms (left panel), fires (middle panel) and special services (right panel) per year (B). Histogram showing the number of special service type operations of all recorded incidents (C).

### 4) Time-dependent mobilisation time variations

The second part of the report is dedicated to analysis of mobilisation times. Generally, mobilisation times are defined as the time interval between an emergency call and the attendance to the site. Accordingly, mobilisation times can be divided into turnout times (i.e. the dispatch time from the operation room to the rescue vehicle) and the travel time (i.e. the time spent inside the rescue vehicle). The sum of both times is called attendance time.

We initially tested whether mobilisation times differed throughout the recorded time span, across a given month of a year, across weekdays and across a single day. Therefore, we compared the number of incidents per year (Fig. 4 A), per month (Fig. 4 B), per weekday (Fig. 4 C) and per hour of the day (Figs. 4 D-F). Analysis of year-dependent mobilisation time variations discloses a steady increase in mobilisation times in recent, post-COVID, years.

Furthermore, we find month-dependent mobilisation time differences: Mobilisation times are shortest in spring and tend to increase towards the summer and then tend to decline again.

With regard to weekdays, the shortest, respectively highest, mobilisation times are detected on Sundays, respectively Fridays.

Lastly, mobilisation times were analysed across a single day. To this end, a 24 hour day was split into 8 bins of 3 hours each and we plotted the number of incidents within each bin as a function of pump arrival. In addition, we plotted the mean pump arrival time of each hour of the day as a bar chart. We find a sine-wave-like pattern in mobilisation times: Mobilisation times are shorter towards the evening and late night and tend to increase by the early dawn. From around 7 AM onwards, mobilisation times decrease until around 10 AM where mobilisation times begin to prolong again.

In summary, we find that time has a significant impact on mobilisation times. The p-values of the 4 corresponding ANOVA tests (does year/month/weekday/hour of day affect the mobilisation time) are all smaller than .

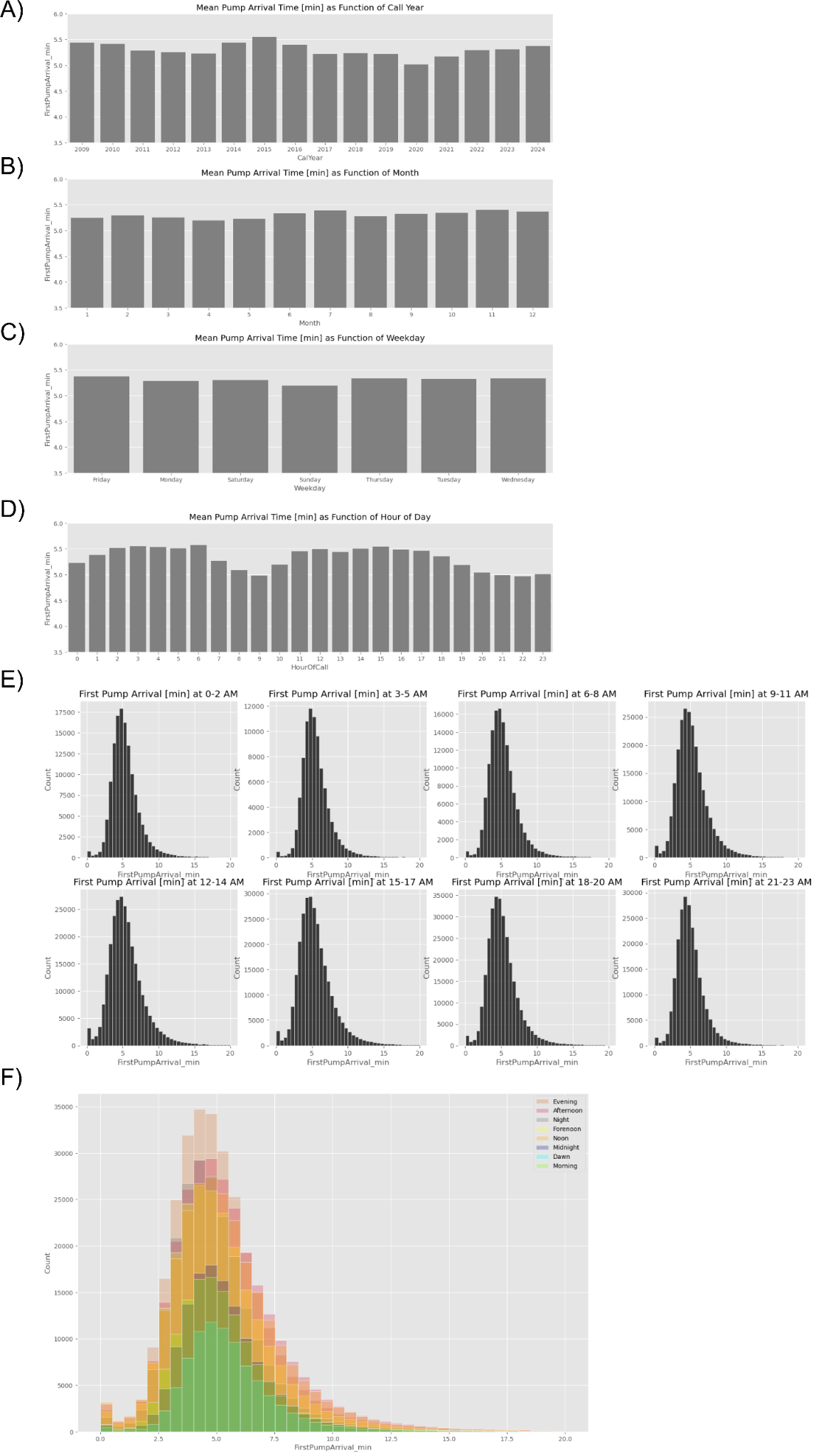
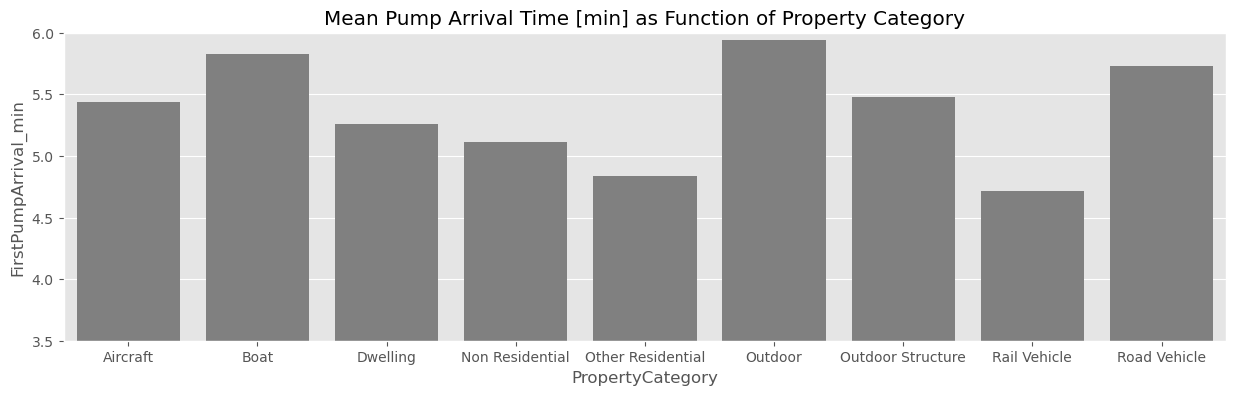


Figure 4. Histogram showing the number of all recorded incidents per year (A), per month (B), per weekday (C) and per hour of the day (D). Histogram showing the number of pump arrivals (binwidth = 30 s) per 3-hour bin of a single day (E). Visualisation of all histograms depicted in (E) merged into a single graph to display differences in incident numbers (F).

### 5) Site-dependent mobilisation time variations

Next, we analysed site-dependent mobilisation time differences. To this end, we plotted the mean mobilisation time as a function of the property category (Fig. 5). With the exception of rail vehicles, mobilisation times tend to be shorter for indoor operations compared to outdoor operations.

Figure 5. Histogram showing the mean first pump arrival time as a function of the property category.

### 6) Traffic, weather- and equipment-dependent mobilisation time variations

Next, we analysed differences in all recorded mobilisation times according to delay code description. Altogether, the category delay code consists of 9 features: 1) not held up (i.e. normal conditions), 2) roadworks/traffic, 3) traffic calming conditions, 4) wrong address, 5) radio problems, 6) outside duty, 7) non-normal weather conditions, 8) defective equipment, 9) drills. As mentioned before, mobilisation or attendance time is split into a turnout and into a travel time. Here, we analysed the impact of the different delay codes on either sub-time and the overall attendance time. We find that turnout times are delayed by radio problems and/or defect equipment and/or drills (Fig. 6). In contrast, we find that travel times are mostly influenced by reporting the correct/incorrect address and during drills (Fig. 7). In summary, the attendance time is influenced by a combination of equipment, radio and address and to a milder degree by traffic or weather (Fig. 8, Fig. 9).

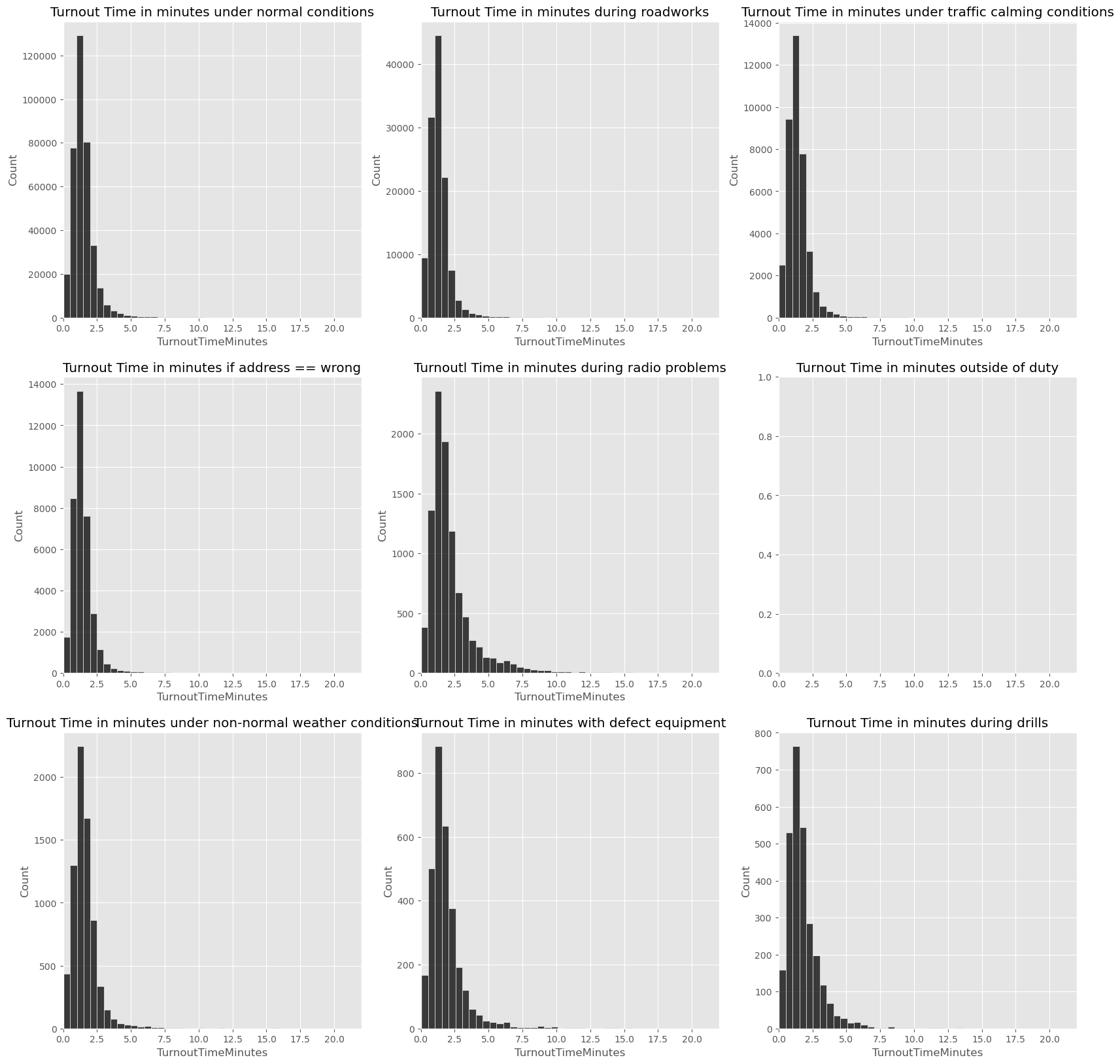


Figure 6. Histograms showing the turnout time (binwidth = 30 s) at different delay code conditions (as indicated by plot title).

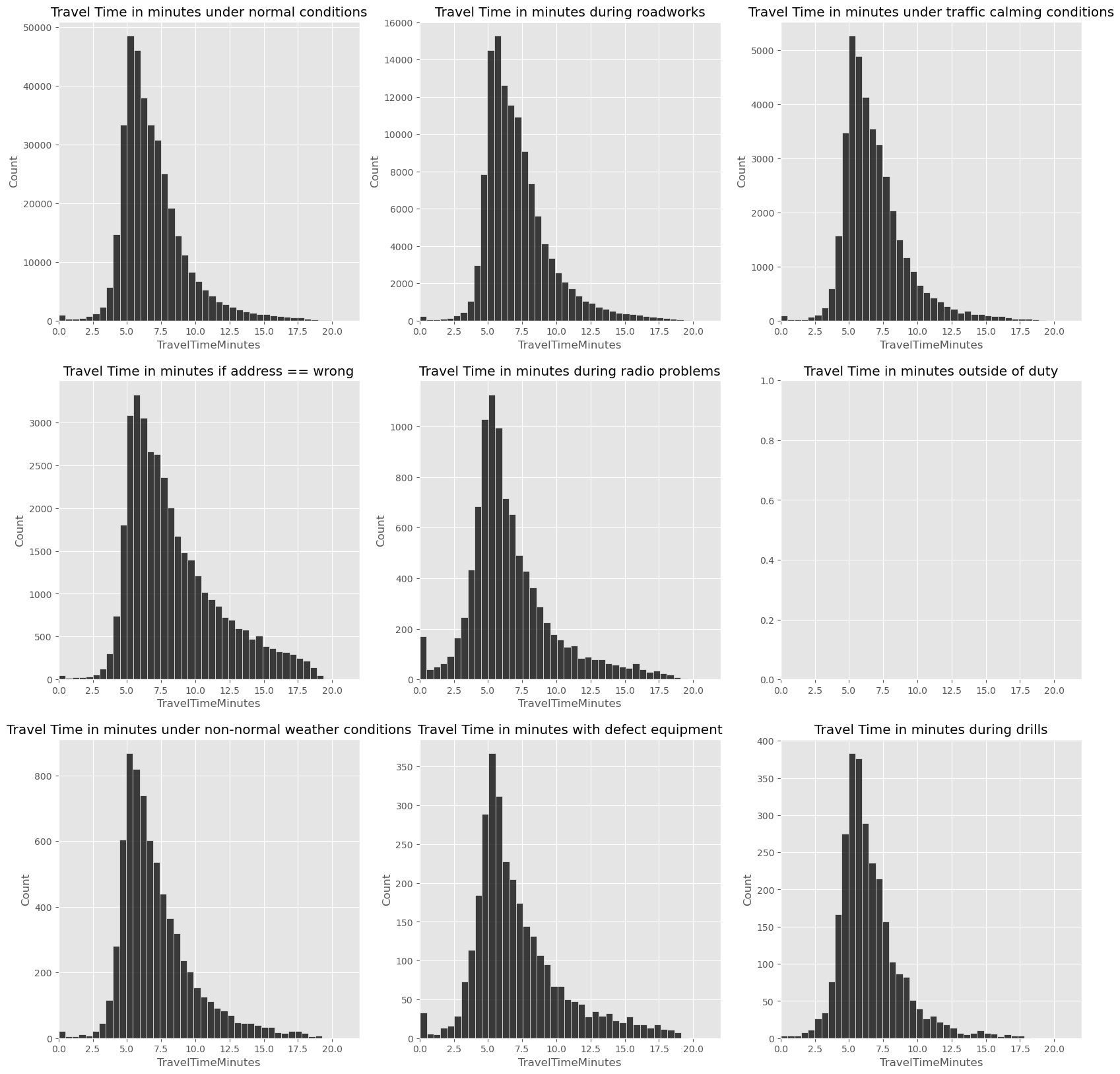


Figure 7. Histograms showing the travel time (binwidth = 30 s) at different delay code conditions (as indicated by plot title).

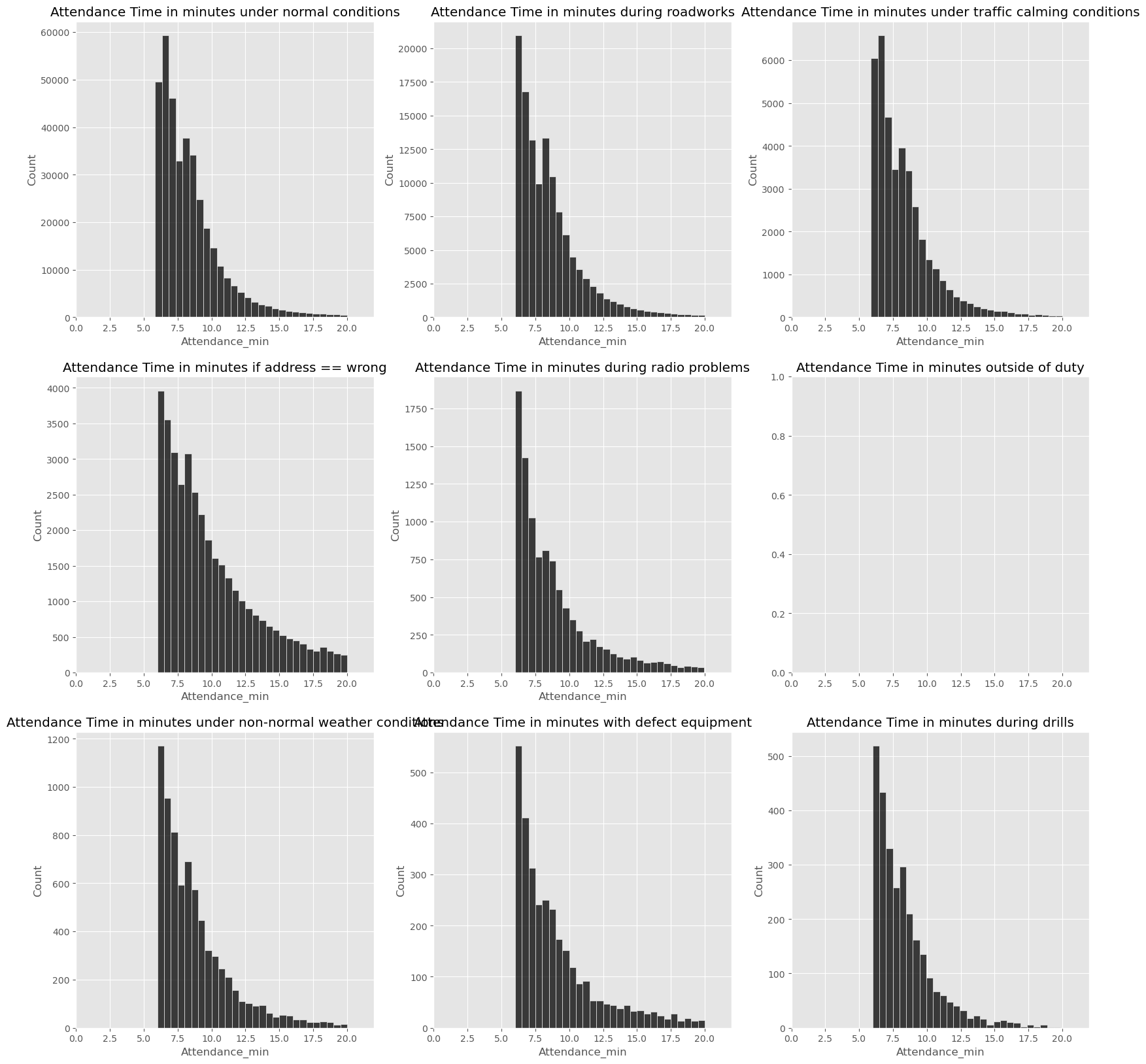


Figure 8. Histograms showing the attendance time (binwidth = 30 s) at different delay code conditions (as indicated by plot title).

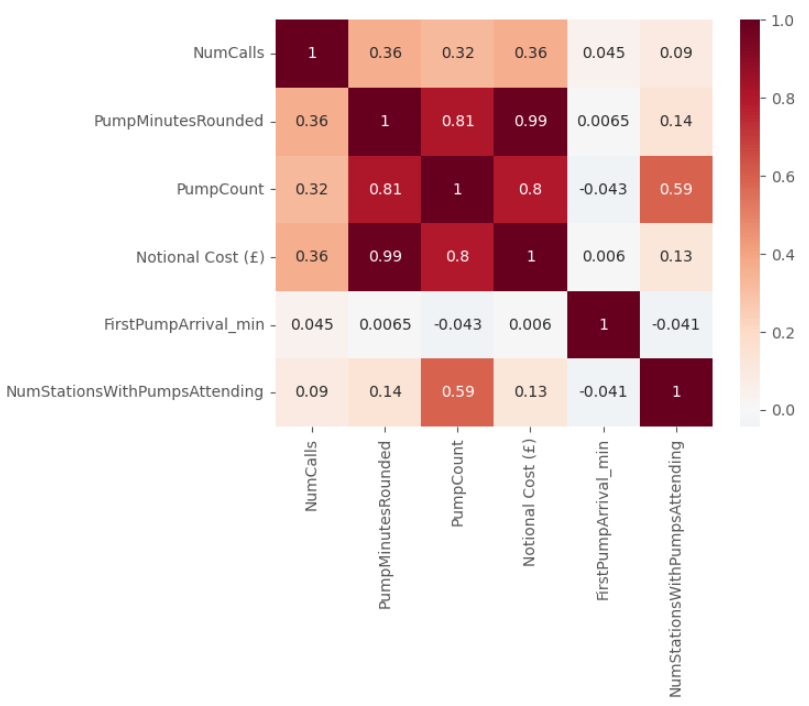


Figure 9. Heatmap showing correlations between selected categories.

### 7) Service-type dependent pump operations and derived national costs

The last part of the present report investigated the factors that predominantly impact on national costs. We hypothesised that the number of pumps and the pump duration directly reflect the extent of a given incident. Therefore, we first analysed the number of pumps in the three incident groups and, unsurprisingly, found this number to be highest in fires and lowest in false alarms (Fig. 10 A). Similarly, the highest pump durations were found in fires, the lowest in false alarms (Fig. 10 B). Given that pump numbers and pump durations reflect the extent of an incident, national costs were significantly higher in fires compared to special services and/or false alarms (Fig. 10 C). This finding is further supported by a high correlation between national costs and pump duration and pump counts whereas pump arrival times do not correlate with national costs (Fig. 10 D).

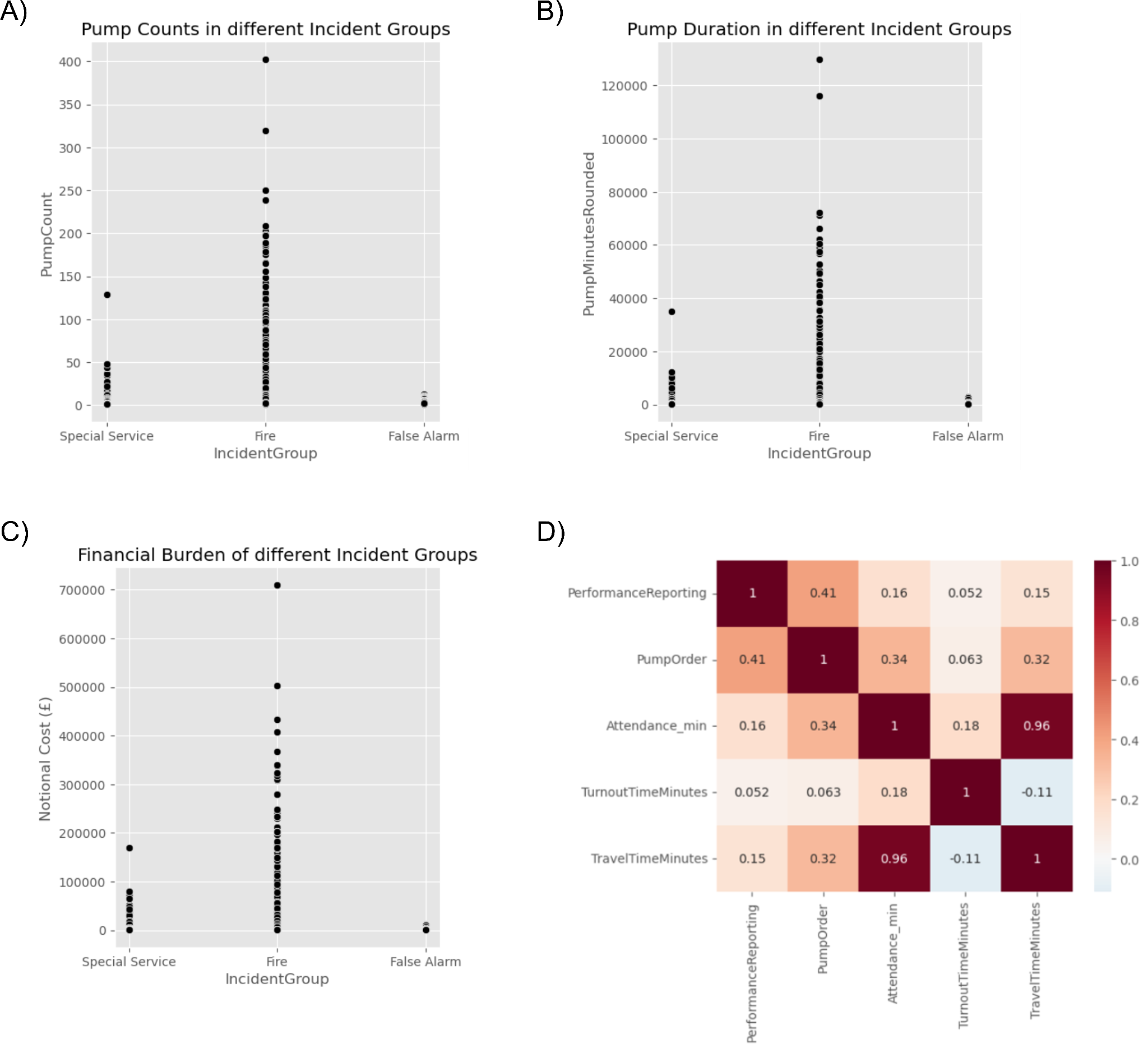
****

Figure 10. Scatter plot showing individual pump count of all recorded incidents between 2009 and 2024 (A). Scatter plot showing individual pump durations of all recorded incidents between 2009 and 2024 (B). Scatter plot showing national costs of all recorded incidents between 2009 and 2024 (C). Heatmap showing correlations between national costs of selected categories (D).

## Discussion

The current report described time- and site-dependent changes in incident numbers. In addition, it identified important variables that have a significant impact on mobilisation times. The work of the LFB is complex and multi-dimensional as it involves the fight against fires, floodings and forces of nature but it also provides help in road collisions, medical incidents and chemical spills. Given the broad operational activities of the LFB, changes in incident numbers are the result of complex interactions between many variables. However, identification of cardinal features impacting on incident numbers will help to reduce incident numbers in the future. False fire alarms represent such a cardinal feature: The data provided here clearly show that false alarms are the predominant contributor to the number of recorded incidents. Costs for responses to false fire alarms add up to around £4 m per year tying up important resources. The introduction of a policy to charge repeated offenders of false fire alarms in the early 2010s resulted in a robust decline of false fire alarms and, consequently, of total incident numbers and suspension of this policy in recent years caused incident numbers to rise again. Another factor contributing to high incident numbers are road traffic collisions. The introduction of congestion fees for vehicles driving in and around London resulted in a reduction of traffic density. It is speculated here that this also contributed to a reduction of total incident numbers. Similarly, the data shown here indicate that the COVID pandemic caused another massive reduction in total incident numbers. A lift of all COVID restrictions by 2022 together with the end of the false alarm charging policy caused a high increase in total incident numbers, however, other contributing factors remain unknown and need to be identified. The data shown here depict a clear correlation between population density of individual London boroughs and incident numbers. In addition, most incident operations are reported in private and business buildings. In an effort to reduce incident numbers, the LFB introduced Home Fire Safety Visits and offers educational work on fire alarm systems.

Analysis of mobilisation times revealed that defect equipment or radio problems result in a delayed turnout time, whereas the travel time to the incident site is further influenced by traffic, non-normal weather conditions and to a high degree by the correct designation of the incident address. In addition, we find that mobilisation durations are influenced by time. Given that the traffic density also shows a time-dependent pattern, reductions in traffic density will help to decrease mobilisation times. Interestingly though, the attendance time is shortest at 9 AM and during the night despite contrasting traffic densities indicating that other factors such as number of staff on night/day duty contribute to mobilisation times. Another factor playing a role in attendance time is the distance from the fire station to the incident site. Under the Johnson mayoralty, the LFB budget was cut leading to the closure of fire stations and an increase in mobilisation times in 2014-2016. A reopening of fire stations as a consequence of the Grenfell Tower incident resulted in an acceleration of mobilisation times.

The LFB is bound to offer the highest possible service in the shortest possible time for the worst possible incident on a budget that is assessed by a combination of routine work and hypothetical major incidents. Therefore, any measure to reduce incident numbers and mobilisation times is important to operate within the allocated budget.