How are you getting home tonight?:

The intersection of nightlife, bus routes, and road safety - a research note

Diana Radovanovici, Filip Szutowicz and Mafalda Candal

Amsterdam University College

Introduction to Geographic Information Systems

Eduardo Dias

December 13th, 2024

Introduction

Transportation planning is a field of urban planning that adapts the current transportation system in a region, focusing on future needs for safety, accessibility, and convenience. An everlasting threat to road safety is alcohol-induced crashes. Consumption of alcohol diminishes reflexes for all pedestrians, bikers and other people on the streets, and it particularly impairs perception, cognition, and action in manual driving (Chen et al., 2018), leading to an increased number of road accidents. Creating appealing alternatives to personal vehicles would minimise the occurrence of “drunk driving" (Farber & Páez, 2009; Peter, 2017). The relation between alcohol-induced crashes and 24h transportation systems, use of UBER service, or proximity to metro stations has been studied.

This case study aims to estimate the relation between the proximity of nightlife establishments to operational night bus stops and the frequency of nearby alcohol-related crashes in Delaware, USA. Public transport networks in US cities are less developed than the ones in Europe (National Research Council, n.d.), making them a more urgent area of analysis. Delaware, the 1st US state, totals just over 5,000 km2. It qualifies for a small scale analysis, adequate for this short case study, but it is preferable over other cities because states have more freedom to implement policy changes, meaning the results presented in this research note can be more easily implemented and have a greater impact. Additionally, Delaware has two areas with higher density of nightlife establishments, and only one of these has access to operational night bus routes, so we believe that might make for clearer results on the influence of access to night buses on the amount of alcohol-induced crashes.

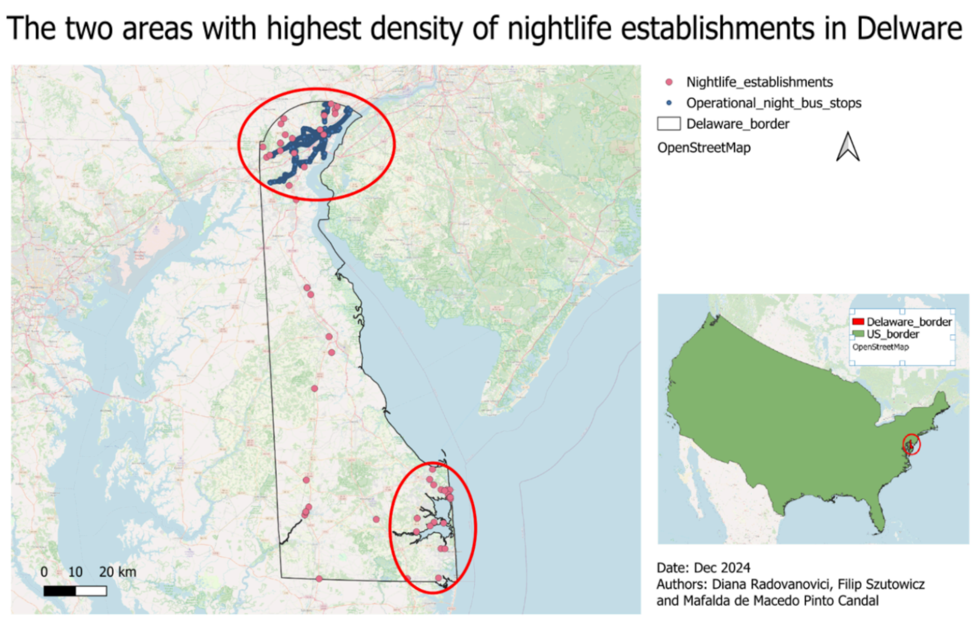

Figure 1: Two areas with highest density of nightlife establishments in Delaware

Data

Layer	Date	Resolution	Source	Format
gadm41_USA	July 2022	N/A	GADM maps	vector
OpenStreetMap	Nov 2024	N/A	OpenStreetMap	vector
geo_export_029a (Crash Dataset)	Dec 2024	N/A	DSHS	vector
DART_Bus_Schedules	October 2019	N/A	DTC	vector

Table 1: data description

The **GADM Database of Global Administrative Areas** is a highly detailed geographic dataset providing precise administrative boundary information for nearly all countries and territories worldwide. The database includes boundaries for multiple administrative levels (e.g., countries, states, districts), offering comprehensive and hierarchical spatial data essential for research, policy-making, and geographic analysis. The GADM Database is used in our analysis to define the administrative boundaries of Delaware for spatial analysis, and of the US for geographic reference. We selected the latest map of the US and its states provided, which was published in July 2022 and ensures accuracy and reliability, as Delaware's borders have suffered so recent changes.

OpenStreetMap (OSM) is a trusted and versatile geographic data platform, defined by its collaborative, open-source approach. While occasional inconsistencies may arise due to its open-editing model, its rigorous community-driven updates secure its role as a reliable and comprehensive tool for mapping and geospatial analysis. It is a powerful and versatile geographic data platform that allows users to access and extract specific map features such as roads, buildings, waterways, or points of interest. Using the plugin QuickOSM we derived the location of the current nightlife establishments and the roads in Delaware.

The **Delaware Department of Safety and Homeland Security (DSHS)**, serves as the official custodian of statewide crash data in Delaware. This dataset is made available through platforms such as FirstMap and the Open Data Portal, and includes information derived from crash reports submitted by Delaware law enforcement officers. It provides a comprehensive summary of collisions for public exploration and analysis, including the time/date of the crash and “suspected alcohol use by ‘any person’ involved in the crash, excluding passengers” (Delaware, 2022). This data provided the location of alcohol-related crashes during 11pm and 6am in the state of Delaware between the years of 2017-2019. This will be our temporal window of analysis, because our data on bus routes aligns with this time period.

The **Delaware Transit Corporation (DTC)** is the government agency that manages, funds, and plans the public transit operations in Delaware. This dataset is also made available through Open Data Portal and provides comprehensive information about the DART (the service brand operated by the DTC) bus routes. It includes detailed records on routes, stop locations, stop times, and trip-specific details for all bus services. The data is structured to facilitate in-depth analysis of transit operations and planning. From this dataset, we could select the bus stops that offered operational bus routes between 11pm and 6am between the years of 2017-2019.

Methodology

This study case follows a 3-step methodology. Starting with the preprocessing of the data, followed by the creation of heatmaps for each dataset and sampling, and, lastly, the linear regression. These steps are illustrated with a flowchart. The red boxes depict the input layers, while the green boxes contain the output layer of each flowchart.

Preprocessing Data

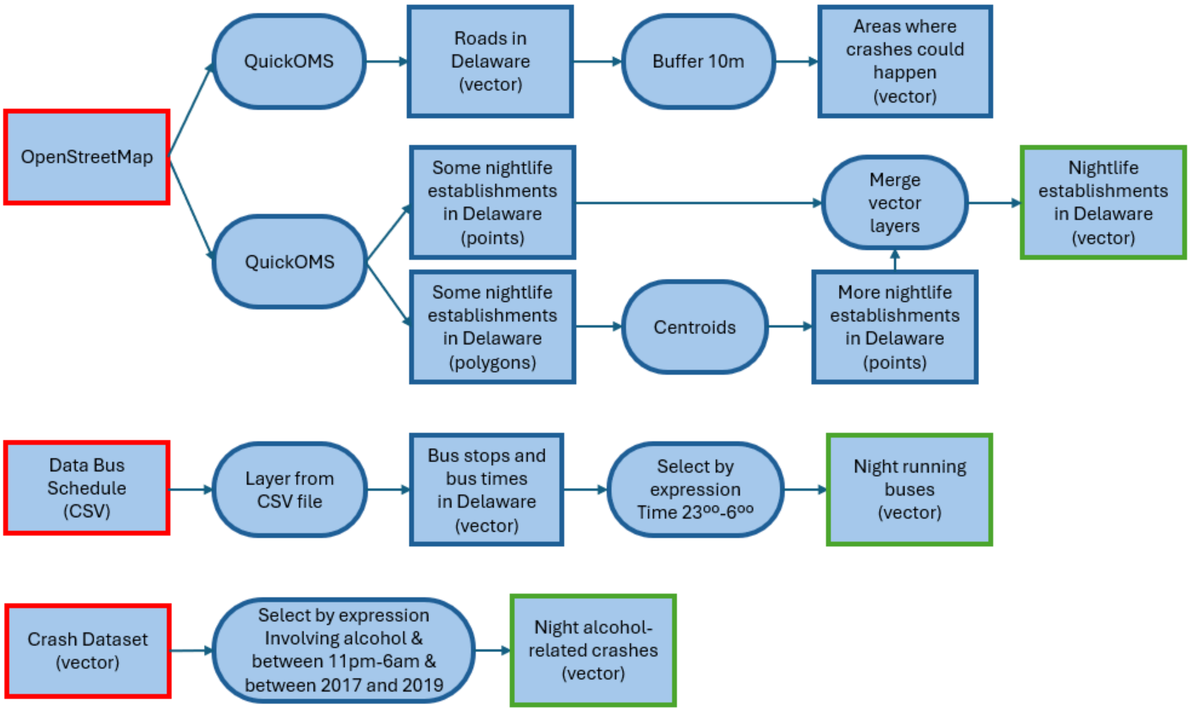

Figure 2: flowchart depicting the method for creating the vector layers for the nightlife establishments, the operational night bus stations 2017-2019, and the alcohol-related crashes 2017-2019

The first step of this analysis involved extracting the relevant data for this analysis from the original datasets. OpenStreetMaps allows users to access and extract specific map features. We used the *QuickOSM* plugin to select the roads in Delaware by creating a layer with the data under highway->motorway, motorway_link, trunk, trunk_link, primary, primary_link, secondary, secondary_link, tertiary, tertiary_link, residential in the area of Delaware. We then used the *buffer* tool to add a 10m buffer to the line representations of the streets to create a layer of the area where crashes could have happened. Following the same method, we selected the night establishments in Delaware by creating a layer with the data under amenity->nightclub, bar, pub, stripclub, swingerclub in the area of Delaware. This process created point and polygon data. The polygon data was converted into point data by use of *centroids* and then merged with the original point layer, creating a final layer with all the night establishments in Delaware. Next, we used the *layer from CSV file* tool to convert the bus CSV dataset into a vector layer. This data involved all the running buses and routes, so we *selected by expression* the ones that ran between 11pm and 6am, creating a layer with only the bus stations where at least one bus route was operational between that time period. Similarly, we selected by expression the alcohol-related crashes that happened between 11pm and 6am in the 2017-2019 timeframe.

Heatmaps and sampling

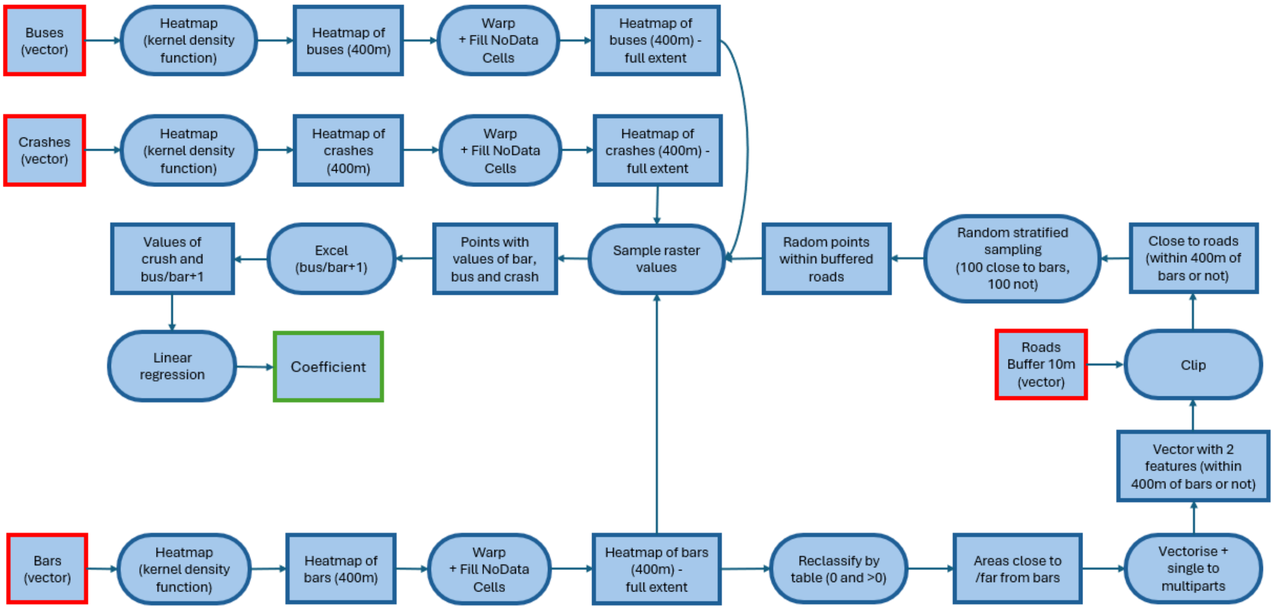

Figure 3: flowchart depicting the method for creating the heatmaps for the three main points datasets, the coefficient (bar/bus+1) and random sampling within buffered roads

The second step involved starting by creating heatmaps for each of the datasets which we will address as buses, bars and crashes for simplicity. We established a baseline value of 400m for the radius of the heatmap for buses, as 0.25 miles is put forward as the acceptable walking distance in U.S. research studies (Yang & Diez-Roux, 2012). For consistency, we chose 400m as the radius for the three baseline heatmaps. These values are revisited in the sensitivity analysis stage. Using the heatmap (kernel density function) tool, we create heatmaps for each dataset, followed by using the *warp* tool to reproject the layers onto the same UTM zone 18N, EPSG:32618 projection and then using the *Fill NoData Cells* tool, altering the extent of the layer so that it covers the whole of Delaware and not only the area covered by the points in the heatmap.

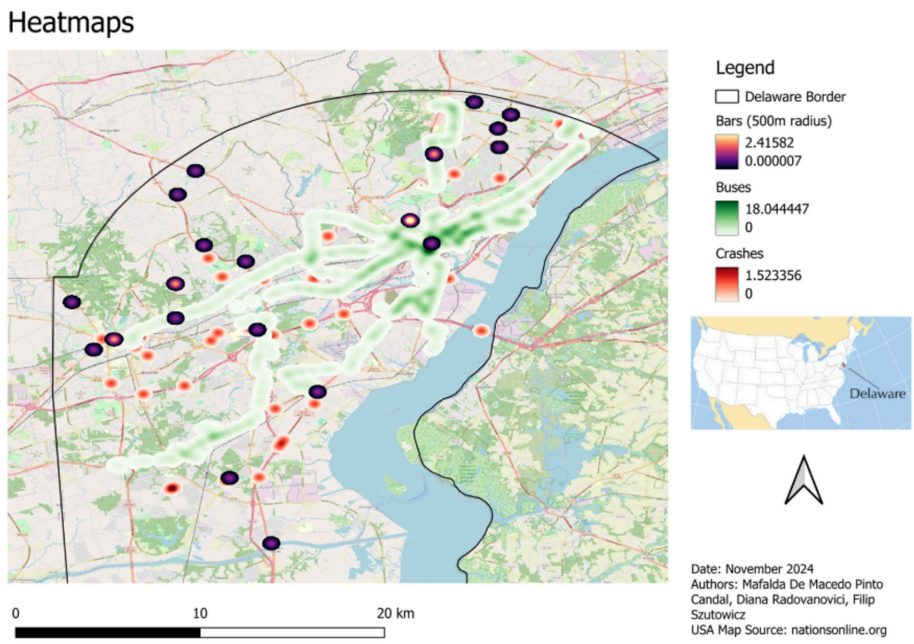

Figure 4: Visualisations of the base heatmaps for buses, bars and crashes

The final step of our methodology is running a linear regression, using the crashes as the dependent variable. We decided to perform stratified sampling in order to reduce the number of datapoints which have zero values for the parameters. The decision to stratify by the density of bars followed from preliminary attempts at non-stratified sampling and initial stratification by crashes. In both cases the number of non-negative values for the bar density was the lowest among the three parameters, which proved harmful for the analysis. Thus, stratification by bar density was the alternative chosen aiming to decrease the number of data points with parameter values equal to zero.

To this end, we used *reclassify by table* tool to obtain a raster with value 0 for the cells with 0 density, and with value 1 for cells that had bar density larger than zero. We then used the *vectorise* and *single to multiparts* tools to create a vector layer with two features - bars within 400m or not. The next step involves restricting this area to the parts where crashes could have happened at all, meaning using the *clip* tool to restrict the two-feature layer to the buffered roads. This new layer is now ready for the stratified random sampling, which yielded 100 points from within the 400m heatmap for bars and 100 points beyond the 400m.

In alignment with our research question, we wanted to understand the interconnection between proximity to bars and operational bus stops and the occurrence of crashes, leading to the decision to not include the bars and buses as two independent variables, but as a coefficient created with the expression (bus/bar+1) where bus alludes to the operational night bus stations heatmap and bar to the nightlife establishments heatmap. The (+1) added on the denominator avoids the creation of very large values when the density of bars is close to zero. This computation was done in Excel only on the sampled points as a less computationally heavy alternative to the *raster calculator* operation on the entire raster. This coefficient can be interpreted as the relative density of bus stops compared to nightlife establishments in a given area, offering a normalized measure to assess how the proximity of these features collectively influences the occurrence of crashes.

Finally, the last step consists of running a linear regression with our bus/bar+1 coefficient as the independent variable and the car crashes as the dependent variable, producing the final coefficient and statistical parameters presented in the following results section.

Results

As detailed in the Methodology section, we will use a simple linear regression $Y = \beta_0 + \beta_1 X$, Y representing the Crashes variable, and X the ratio term of Buses/(Bars+1). The initial case for our regression considers that radius of 400m around the bus stations, nightlife establishments and car crashes, values also used in similar papers such as Jackson and Owens' (2011). Unfortunately, the results for these values are not significant, the model having an R^2 of 0 (Table 2), so 0% of the crashes could be explained. As we will see in the Discussion section, an attempt to improve the accuracy of the results was made by choosing a sample of 2000 points, but the R^2 remained 0. This led us to perform a thorough sensitivity analysis, looking at various radius values for our independent variables, as it is highly likely that car crashes which are further than 400m from bars and bus stations are actually related to their presence, or lack thereof.

Sensitivity Analysis

Bus Radius (m)	100			400			800		
	R^2	β	p-value	R^2	β	p-value	R^2	β	p-value
Bar Radius (m)									
400	0	-0.06	0.491	0	-0.01	0.541	0	0	0.497
2000	0.02	0.1	0.073	0.01	0.01	0.277	0.01	0	0.128
3600	0	0	0.905	0.01	0.01	0.113	0.02	0	0.056
5200	0	-0.1	0.894	**0.05**	**0.05**	**0.002**	0.03	0.01	0.018

Table 2: results of sensitivity analysis, where flow β represents the coefficient of the buses/(bars+1) independent variable, bolded results belong to the best regression for our data, and a sample size of 200 points was used.

For the sensitivity analysis, the radiuses for the nightlife establishments were calculated using the same methodology as Levine (2017), who looked into how many miles away from bars do car crashes happen. Their conclusion was that 45% of the car crashes happen within 0.25 miles, and for every extra mile the likelihood of a car crash happening increased by 42%. So, the radius values chosen are 0.25, 1.25, 2.25 and 3.25 miles, which mean 400, 2000, 3600 and 5200 meters. The bus station radii were chosen to be 100, 400 and 800 meters, similarly to Jackson and Owens' (2011) paper that explored a related research question, but with a focus on metro stations. The radius used to determine the relevant area around car crashes remained 400m for all of the regressions and Table 2 presents the results.

The sensitivity analysis allowed us to find a better performing model for our project, and if we take the relevant areas around bars to have a radius of 5200 m, and around bus stations and car crashes a radius of 400 m, then our resulting regression has R^2 0.05, and a coefficient β_1 of 0.05. This coefficient has a p-value of 0.002, which means that the significance level is under 5% and the confidence level over 95%. The results show that 5% of alcohol-related car crashes that occur during the night can be explained by our model and an increase of one unit of the Bus/(Bar+1) coefficient leads to an increase of the car crashes by 0.05 units. So, the more bus stations are around nightlife establishments, the more car crashes are occuring within their proximity. This is contrary to our expectation, so further research should be done to verify and explain this result.

Finally, for reasons further detailed in the Discussion section, we will look at the flipped ratio of Bar/(Bus+1). The regression using this new term as our X has R^2 0.02 and a β_1 of 0.07 with p-value less than 0.01. While the coefficient has a significance level of 1%, only 2% of the crashes are explained by this model. An increase of one unit of the Bar/(Bus+1) coefficient would lead to an increase of the car crashes by 0.07 units. The positive coefficient confirms the positive correlation found by our best regression using the Bus/(Bar+1) coefficient. Further impacts of the flipped ratio can be explored in future papers. Another attempt at exploring the relationship between bus stations, bars and car crashes was recoding the sampled data such that a bus station density bigger than 0 reclassifies the bus station variable to 1 and a density equal 0 reclassifies it to 0. This turns our independent variable X into a binary variable that is 1 if there are bus stations around a bar and 0 if not, showing us if just the presence of bus stations around bars has an impact, regardless of their number. Unfortunately, this regression model had a R^2 of 0, so no additional insights were provided.

Discussion

Within the area of this research, there were relatively low numbers and densities of the nightlife establishments. The same holds for the night buses. They are only present in the area of one city – Wilmington. The remaining vast area of the state is devoid of night buses. Lastly, the numbers of car accidents involving alcohol were naturally moderately low. Those facts resulted in a very low number of points with non-zero densities of all three parameters among the sampled points. Moreover, the number of points sampled for the purposes of the sensitivity analysis was too low considering this scarcity of non-zero data. Larger sample size, like 2000 points used for the base case scenario was a more optimal number. However, this sampling resulted in the value of R^2 equal to 0 and a high p value for the slope. The reason behind those results is explained further. The above explains low statistical significance of the sensitivity analysis.

The treatment of the two parameters as one variable, the ratio of bus density over bar density incremented by one, further exacerbated the amount of the values of the independent variable equal to zero. Whenever the bus density is zero, any potential non-zero value of the bar density at this point is disregarded in further statistical analysis. However that decreases the amount of information, it results from the fundamental idea behind this parameter. Namely, this paper focuses on the density of the buses near nightlife establishments. This means that the loss of data about bus stops farther away from bars is meaningful and important for the study.

This issue could be alleviated by performing stratified random sampling with multiple strata. In this case, the desired number of points with non-zero values of all three parameters should be set. That would provide better possibilities to draw meaningful conclusions from the results.

However, there is one more limitation of the bus/bar ratio we used as an independent variable. The hypothesis expects that the number of crashes should decrease with the increase in this ratio. This is because an increase in ratio means that there are more buses per bar at a given point, so less people would drive under the influence. So the expectation is a negative correlation between the dependent and independent variable. However, it was very unlikely to obtain a negative coefficient in the regression. This is attributed to the prevalence of data points with zero values of both variables, mentioned above. This leads to strong attraction of the linear regression line to pass through the origin. All the data points have non-negative values of both variables, so the most likely result of the regression is a slightly increasing line, or a line very close to $y=0$.

A possible improvement is to inverse the ratio, using the variable bar density over bus density incremented by one. Treating this inverted ratio as an independent variable is possibly a more reasonable strategy. The expectation is that it has a positive correlation with crashes. That can improve the results of the analysis significantly considering high concentrations of points at the origin (so with values of both variables equal to zero). However, the prevalence of those points is lower than with the original ratio considering the scenario with stratified sampling by the density of bars. This is because the multiple non-zero values of bar density resulting from the stratified sampling are now in the numerator, while the denominator (bus density) cannot make the ratio equal zero. The results obtained from the regression with this variable showed that it explains some variability of crashes (R^2 equal to 0.02) with a very low p-value.

A different limitation stems from the nature of the subject of this study – cars. People leaving bars, potentially under the influence of alcohol, can drive a long way and cause an accident far away from the bar. With the available data it is impossible to possibly attribute those crashes to any nightlife establishment.

Last technical assumption is that we used the current (2024) version of the Open Street Map to get data about nightlife establishments. Meanwhile, the data we used for crashes and buses is for the years 2017-2019. The nightlife establishment distribution might have changed over this period, especially considering the COVID-19 pandemic that might have caused closure of many businesses (Chandler et al., 2021).

Conclusion

The methods of this study underwent many changes. However, considerations on them allowed us to arrive at the ideas for improvements. The modified ratio used as a variable brought much more significant results. Using it, with high statistical significance we managed to explain 0.02 of the alcohol related crashes in the state of Delaware. This result is highly relevant if it is interpreted as 2% of alcohol-related crashes that can now be prevented, having the potential to save many lives. Future research in Delaware and further locations can build upon this paper's progressive adjustments, setbacks and successes.

References

Chandler, M., Cole, G., Kunkle, G., & Wial, H. (2021). *How the Coronavirus Recession and Recovery Have Affected Businesses and Jobs in the 100 Largest Metropolitan Areas Second Quarter of 2020 through Second Quarter of 2021*. https://home.treasury.gov/system/files/271/Q2.21-ICIC-Recession-Recovery-Tracker-Report-Draft-Formatted-with-Cover-11.3.21.pdf

Chen, H., Yuan, X., Ye, H., Chen, L., & Zhang, G. (2018). The effect of alcohol on the physiological performance of the driver. *International Journal of Crashworthiness*, *24*(6), 656–663. https://doi.org/10.1080/13588265.2018.1511226

Connor, J., Norton, R., Ameratunga, S., & Jackson, R. (2004). The Contribution of Alcohol to Serious Car Crash Injuries. *Epidemiology*, *15*(3), 337–344. https://doi.org/10.1097/01.ede.0000120045.58295.86

Delaware. (2022, July 25). *Public Crash Data*. Delaware.gov. https://data.delaware.gov/Transportation/Public-Crash-Data/827n-m6xc/about_data

Farber, S., & Páez, A. (2009). My car, my friends, and me: a preliminary analysis of automobility and social activity participation. *Journal of Transport Geography*, *17*(3), 216–225. https://doi.org/10.1016/j.jtrangeo.2008.07.008

Jackson, C. K., & Owens, E. G. (2011). One for the road: Public transportation, alcohol consumption, and intoxicated driving. *Journal of Public Economics*, *95*(1-2), 106–121. https://doi.org/10.1016/j.jpubeco.2010.09.010

Laan, C. M., & Piersma, N. (2021). Accessibility of green areas for local residents. *Environmental and Sustainability Indicators*, *10*, 100114. https://doi.org/10.1016/j.indic.2021.100114

Levine, N. (2017). The location of late night bars and alcohol-related crashes in Houston, Texas. *Accident Analysis & Prevention*, *107*, 152–163.

https://doi.org/10.1016/j.aap.2017.05.010

Map Features - OpenStreetMap Wiki. (n.d.). Wiki.openstreetmap.org. https://wiki.openstreetmap.org/wiki/Map_Features

National Research Council. (n.d.). Making Transit Work: Insight from Western Europe, Canada, and the United States -- Special Report 257. In *nap.nationalacademies.org*. National Academic Press. https://nap.nationalacademies.org/read/10110/chapter/4#23

Peter, F. (2017, May). *Responsible options : empirical analyses on the effects of alternative transportation on drunk driving*. Utexas.edu. https://repositories.lib.utexas.edu/items/f71b4f33-4579-4c6e-81f0-42c39977722b

Yang, Y., & Diez-Roux, A. V. (2012). Walking Distance by Trip Purpose and Population Subgroups. *American Journal of Preventive Medicine*, *43*(1), 11–19. https://doi.org/10.1016/j.amepre.2012.03.015

Declarations

All the authors equally participated in the steps.