

The Impact of New Bike Lanes on Urban Transportation Dynamics

Shaya Engelman shaya.engelman82@spsmail.cuny.edu

1 INTRODUCTION

1.1 Background

As urban populations continue to surge, cities around the world are grappling with multifaceted challenges related to traffic congestion, economic productivity, environmental sustainability, and public safety. With more individuals residing in urban areas than ever before, the traditional reliance on motor vehicles has become increasingly untenable and changes are necessary. Many cities opt to attempt to transition more residents towards using bicycles as a mode of transportation. This study seeks to address these challenges by examining the impact of new bike lanes on urban transportation dynamics, specifically focusing on their effects on bike usage, car usage, and pedestrian safety.

The hypothesis posits that adding new bike lanes will lead to increased bike usage, decreased car dependency, and reduced pedestrian accidents, thus contributing to less congestion and enhanced regional productivity. Given the multitude of factors influencing productivity, directly measuring the impact of bike lane changes on it is extremely challenging. Instead, this study accepts the findings by Hartgen and Fields (2009), which concluded that mitigating congestion can significantly bolster economic performance, as a given. Consequently, the focus is on finding increased bike usage as a key indicator. By examining these factors, this study aims to highlight the broader impacts of new bike lanes on urban environments.

The importance of bike usage cannot be overstated. Biking is not only an environmentally friendly mode of transport, but it also promotes healthier lifestyles among urban residents. As cities face rising pollution levels and public health crises related to sedentary lifestyles, increasing bike usage presents a compelling solution. Furthermore, the proliferation of bike-sharing networks, such as Citi Bike, Bay Wheels, and Divvy, has made biking more accessible and

convenient than ever. These systems lower the barriers to entry for potential cyclists, offering a flexible and cost-effective alternative to car travel.

However, the financial implications of constructing bike lanes are considerable. Cities must navigate the complexities of funding such projects, often facing competition for limited municipal budgets. The costs associated with building and maintaining bike lane infrastructure must be weighed against the potential long-term benefits, including reduced traffic congestion and improved public safety. Additionally, the spatial allocation of roadways poses a challenge; dedicating space to bike lanes can result in the reduction of car lanes or parking spots. Urban planners must carefully balance these competing needs, ensuring that the overall transportation network remains efficient and effective.

1.2 Significance

The significance of this research lies in its potential to inform urban planning and policy decisions that can lead to substantial improvements in urban living conditions. As cities evolve, addressing traffic congestion is paramount, as it not only hampers economic productivity but also negatively impacts public safety and environmental quality. Hartgen and Fields (2009) demonstrated that enhancing travel speeds can lead to regional productivity gains of up to 1%, translating into meaningful economic benefits for urban areas. The link between shorter commutes and economic vitality is further reinforced by research from Prud'homme and Lee (1999), which highlights that cities with reduced commute times experience heightened productivity.

Moreover, the need for fewer pedestrian and cyclist accidents is critical in today's urban landscapes. As bike usage rises, so too does the potential for conflicts between cyclists, pedestrians, and motor vehicles. Well-designed bike lanes can help mitigate these risks by creating dedicated spaces for cyclists, thereby reducing accidents, and promoting safer road environments. This aspect of urban safety is not merely a matter of public health; it also affects the economic landscape. Increased safety leads to greater confidence in cycling as a viable transportation option, further encouraging its adoption.

In light of the COVID-19 pandemic, which prompted rapid, diverse responses in over five hundred cities, the need for innovative transport solutions has never

been more pressing. Many urban areas have reallocated street spaces to support walking, cycling, and outdoor commerce. The emergence of these measures underscores the importance of dedicated cycling infrastructure in urban planning, particularly in fostering resilient, adaptive transportation networks that can respond to changing mobility demands (Combs & Pardo, 2021).

Ultimately, this study aims to provide critical insights for city planners and policymakers. By evaluating the trade-offs involved in constructing bike lanes—including initial financial costs, space allocation, and the potential for reduced car dependency—this research will illuminate the long-term benefits of improved safety, reduced commute times, and enhanced economic productivity. As cities continue to navigate the complexities of urbanization, understanding the role of bike lanes in transportation dynamics will be essential in fostering sustainable urban growth.

Understanding the causal relationships between bike lane additions and accident rates is crucial for informing evidence-based policy decisions. By investigating these dynamics, this research aims to provide insights that not only contribute to academic discourse but also offer practical guidance for urban planners and policymakers seeking to enhance cycling safety and infrastructure in their cities.

2 PRIOR RESEARCH

The literature on the impact of bike lanes on urban transportation presents a varied landscape, with some studies indicating a positive relationship between bike lane infrastructure and cycling rates, while others report minimal effects. For instance, Hwang et al. (2023) found a significant increase in non-motorized transportation methods in urban areas but did not observe a corresponding rise in biking specifically. Similarly, Buck et al. (2011) identified a strong correlation between the presence of bike lanes and the proximity of bike share stations, suggesting that infrastructure improvements can drive bike-sharing utilization.

Research conducted by Hartgen and Fields (2009) indicated that alleviating traffic congestion through improved travel speeds could enhance regional productivity by up to 1%. This is further supported by Prud'homme and Lee (1999), who demonstrated that shorter commute times contribute significantly to economic efficiency. As such, this study aims to explore whether the installation of bike lanes

can enable cities to achieve these economic benefits by alleviating gridlock and optimizing commute times.

One of the most significant contributions to this field comes from Kraus and Koch (2021), who studied the impact of provisional bike lane infrastructure introduced during the COVID-19 pandemic. Their research revealed that temporary bike lanes led to substantial increases in cycling, with rates rising between 11% and 48% in cities that implemented these lanes. This demonstrates the potential for even short-term infrastructure changes to encourage long-lasting shifts in transportation behavior.

Additionally, Arancibia et al. (2019) studied the economic impact of bike lanes in Toronto and found that replacing on-street parking with bike lanes did not harm local businesses. In fact, businesses saw increased customer spending and foot traffic after the lanes were installed. This finding challenges the perception that bike lanes detract from commercial activity, suggesting instead that they may stimulate economic growth by attracting more foot and bike traffic.

Despite the growing body of literature exploring the relationship between bike lanes and cycling rates, there remains a notable gap in comprehensive research regarding their effects on accidents in general, and, in particular, pedestrian accidents. Buehler and Dill (2015) highlighted that the introduction of bike lanes often leads to positive changes in urban dynamics, indicating that dedicated cycling space can reduce pedestrian injuries by improving overall traffic organization. This study seeks to fill this research gap, providing a holistic view of how bike lanes impact urban transportation dynamics, encompassing bike and car usage, pedestrian safety, and overall traffic patterns.

Expanding this research to multiple cities will yield more reliable and generalizable results. This multi-city approach will aim to identify consistent patterns across different urban environments, enhancing our understanding of bike lanes' impact. It is essential to address the reduction or smaller increase in car ridership. While projecting bike ridership growth is useful, comprehending the shift from car users to bike users is equally critical. This study will specifically investigate this aspect, offering valuable insights into how bike lanes can effectively reduce car usage and promote sustainable urban transportation.

3 METHODOLOGY

3.1 Data Collection

3.1.1 Bike Usage Data

The study utilized bike-sharing data from three cities: Citi Bike in New York City, Bay Wheels in San Francisco, and Divvy in Chicago, all operated by Lyft. Monthly reports on bike-sharing usage were obtained for each city, with each month represented by a separate file. In the case of Chicago, some months were combined in the datasets, necessitating the splitting of these files to ensure monthly granularity. A dedicated dataframe was created for each city to compile the total bike-sharing counts per month.

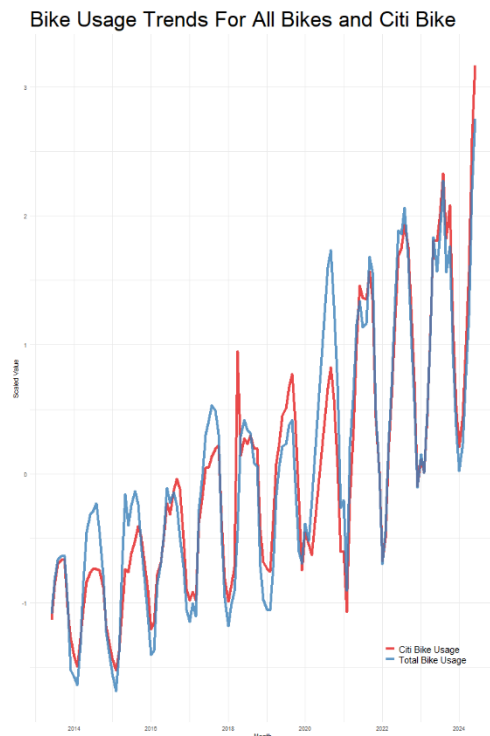


Figure 1; Citi Bike usage data started off with smaller fluctuations than total bicycle usage but as it gains traction it converges to match the range.

To enhance the accuracy of the analysis and address concerns regarding the representativeness of bike-sharing data, validation was performed using additional data sources. Automated bike count data, sourced from NYC's open data platform, was compared against the bike-sharing figures. The trends observed in both datasets converged to closely mirror each other, confirming that bike-sharing data serves as a valid proxy for overall bike usage in an urban environment. Consequently, this validation process was not repeated for SF and Chicago. Figure 1.

3.1.2 Traffic Count and Bike Lane Data

Traffic count data was obtained from the open data platforms of NYC and SF.

Accurate traffic count data for Chicago could not be sourced. To quantify the

extent of bike lane infrastructure, bike lane data from NYC was downloaded from which total bike lane footage was calculated using the installation dates. This data was grouped by month and integrated into the NYC dataframe. A similar process was employed for SF, where bike lane data was also aggregated by month. For Chicago, only partial bike lane data was available from two separate websites. This data, provided on a yearly basis, was manually processed, and added to the Chicago dataframe.

3.1.3 Demographic and Accident Data

Population estimates for each city were derived from U.S. Census data and incorporated into all three datasets. Additionally, pedestrian and cyclist accident data were collected from the respective open data platforms for each city. This included total counts of pedestrian accidents, fatalities, cyclist accidents, and fatalities, resulting in six accident-related columns being added to each city's dataframe.

3.2 Data Preparation and Transformation

In preparation for analysis, the dataset underwent several transformations. For instance, the bike-sharing data was aggregated by month to allow for time-series analysis. In Chicago, some filenames in the dataset were renamed for consistency, utilizing a function that converted filenames into a standardized date format. This process ensured that data from all three cities could be accurately compared.

Missing values were identified and addressed through several approaches. An initial assessment quantified the extent of missing data across various variables, which was essential for understanding the dataset's completeness and integrity. Recognizing the potential impact of missing values on the analysis, particular attention was given to key metrics.

Data prior to the introduction of the Citi Bike network in June 2013 was removed. This step was crucial, as the establishment of the bike-sharing program significantly altered cycling dynamics in NYC. By filtering the dataset to include only relevant data, the analysis could more accurately reflect the effects of bike lanes and the bike-sharing system.

To handle missing values for specific variables, a multiple imputation technique was employed. This method helped fill gaps in the data for the total bike lane length variable, enhancing the robustness of the dataset. Additionally, linear regression models were used to impute missing values for total bike counts and traffic volume, ensuring that the imputed values were informed by existing data trends.

These comprehensive data cleaning and preparation steps ensured that the datasets were both complete and ready for rigorous analysis.

3.3 Data Analysis

Next, as a preliminary step to account for the relative size of cities on their transportation patterns, I analyzed bike usage as a proportion of motor vehicle traffic. This metric aimed to provide a standardized measure of bike usage, allowing for more meaningful comparisons across cities. The cumulative sum of bike lane additions was also calculated to assess the impact of infrastructure changes on bike-to-traffic ratios over time.

Since bike usage patterns can be severely impacted by seasonal variations, the time series data was decomposed to isolate the trend and seasonality components. This decomposition process helped identify underlying patterns in bike usage and traffic volume, enabling a more nuanced analysis of the data. After removing the seasonal component, the trend of bike-to-traffic ratios was plotted against the trend of cumulative bike lane additions to explore potential relationships between these variables.

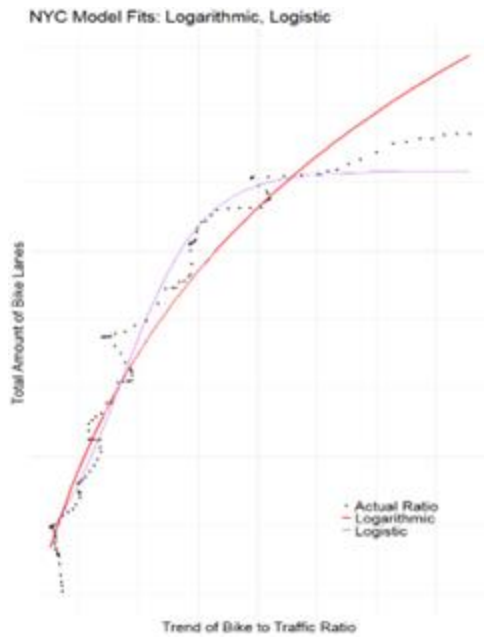


Figure 2; Logistic Regression seems to fit the actual NYC data best, but Logarithmic Regression is likely better for extrapolating future points

In the NYC dataset, this visual analysis revealed a strong resemblance to a logistic curve, suggesting a saturation point for bike lane additions. However, the data also exhibited a slight upward trend after the inflection point, instead of the typical flattening seen in logistic curves. Additionally, a logistic curve is an illogical fit for bike lane additions, as it is unlikely for these to be a strict saturation point where no more bike lanes can be added, while the effect slowing down as more bike lanes are added is more likely. After further exploration, a logarithmic model was considered as a potentially better fit for

the data as it too fit the data reasonably well while also allowing for the continuing upward trend in the data. Figure 2.

Building on these visual observations, a series of regression models were developed to explore potential relationships between bike lane additions, bike-to-traffic ratios, and safety outcomes. The preliminary results of these exploratory models are summarized in Table 1, highlighting potential trends rather than definitive conclusions. The Log Lane-Bike Ratio Model, which plots cumulative bike lane additions against the logarithm of the bike-to-traffic ratio trend, showed a very strong fit ($R\text{-squared} = 0.9426$). This suggests that 94.26% of the variance in cumulative bike lane additions can be explained by the model, with highly significant coefficients ($p\text{-value} < 2e\text{-}16$). This statistical significance confirms a strong relationship between bike lane additions and the log-transformed bike-to-traffic ratio, reinforcing the infrastructure's impact on bike usage trends.

In contrast, the Linear Bike Ratio-Injuries Model, which analyzed the relationship between the bike-to-traffic ratio and total injuries and deaths, had a low $R\text{-squared}$ value (0.01389), meaning the bike-to-traffic ratio explained little of the variance in

injuries and deaths. The non-significant p-value (0.183) further suggests that the relationship is not statistically meaningful. Given the small effect size of the bike-to-traffic ratio relative to overall traffic, it's likely that the ratio alone is insufficient to explain variations in injury rates.

The Linear Traffic-Injuries Model demonstrated a more substantial relationship between motor vehicle traffic volume and injuries and deaths, with an R-squared of 0.1403. Although this explains only 14.03% of the variance, the highly significant p-value (1.22e-05) indicates a strong correlation between higher traffic volumes and increased accident rates.

While a direct statistically significant relationship between the bike-to-traffic ratio and total injuries and deaths was not identified, a clear positive relationship exists between motor vehicle traffic volume and total injuries and deaths. Consequently, there must be a connection between fewer cars on the road (indicated by a higher bike-to-traffic ratio) and fewer injuries and deaths. This relationship was not captured in the linear model likely due to the small effect size relative to total traffic volume. Establishing this connection, along with a positive logarithmic relationship between bike lane additions and the bike-to-traffic ratio (p-value < 2e-16, R-squared = 0.9426), infers that adding bike lanes correlates with fewer injuries and deaths.

Table 1; Summary outputs from the various regression models created. Note the high and statistically insignificant value in the second model.

| Model | Term | Estimate | Standard Error | t-Value | P-Value | R-Squared |
|---------------------------|---------------------------------------|---------------|----------------|-----------|----------------------|-----------|
| Log Lane-Bike Ratio Model | Log-Transformed Bike-to-Traffic Ratio | 2.069258e+06 | 4.529407e+04 | 45.684966 | 1.07923649553373e-80 | 0.9426408 |
| Bike Ratio-Injuries Model | Bike-to-Traffic Ratio | -4.641118e+01 | 3.470183e+01 | -1.337428 | 0.183473386112087 | 0.0138887 |
| Traffic-Injuries Model | Motor Vehicle Traffic Volume | 5.170000e-04 | 1.136000e-04 | 4.552947 | 1.22288388016561e-05 | 0.1403196 |

A similar approach was utilized for the San Francisco data. First a variety of patterns were visualized to determine if the data easily fits any of the common regression lines. As with the NYC data, both a logarithmic and logistic regression line were relative fits. However, neither of them fit as closely as the NYC data did. For the San Francisco data, the logarithmic model seems to be a better fit than the logistic one, further reinforcing that the logarithmic model is a better starting point for analysis. Figure 3.

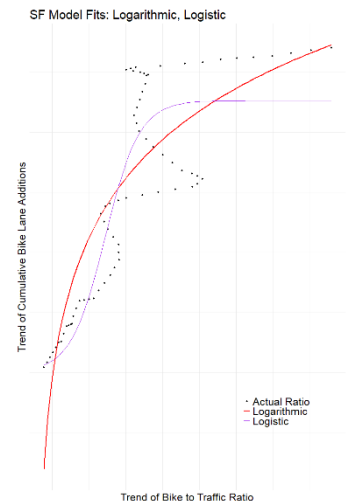


Figure 3

A similar approach was utilized for the San Francisco data. First, a variety of patterns were visualized to determine if the data easily fit any of the common regression lines. As with the NYC data, both a logarithmic and logistic regression line were reasonable fits. However, neither of them fit as closely as the NYC data did. For the San Francisco data, the logarithmic model emerged as the better fit compared to the logistic model, further reinforcing that the logarithmic model is a more suitable starting point for analysis.

The Log Lane-Bike Ratio Model, which analyzed cumulative bike lane additions against the logarithm of the bike-to-traffic ratio, yielded a strong fit ($R^2 = 0.714$). This indicates that 71.4% of the variance in cumulative bike lane additions can be explained by the logarithmic transformation of the bike-to-traffic ratio. The statistically significant coefficient ($p\text{-value} < 2e-16$) underscores the robustness of this relationship, highlighting the importance of infrastructure changes in shaping bike usage patterns.

In contrast, the Linear Bike Ratio-Injuries Model examined the relationship between the bike-to-traffic ratio and total injuries and deaths. This model revealed a moderate fit, with an R^2 value of 0.1837, suggesting that the bike-to-traffic ratio explains only 18.37% of the variance in injuries and deaths. The significant p -value (0.000192) indicates a meaningful relationship, pointing to the potential influence of increased bike usage on safety outcomes.

Finally, the Linear Traffic-Injuries Model analyzed the impact of motor vehicle traffic volume on total injuries and deaths. Although the model showed a very low

R-squared value (0.02044), suggesting limited explanatory power, it indicated that motor vehicle traffic has a strong correlation with injuries, as evidenced by the significant intercept ($p\text{-value} < 2e-16$) and a coefficient indicating a small effect size for traffic volume.

Overall, while these preliminary models highlight different aspects of the relationship between bike usage and safety, they collectively suggest that increasing bike infrastructure may contribute positively to reducing injuries and deaths in San Francisco; however, further analysis is necessary to draw definitive conclusions. The findings emphasize the complexity of urban cycling dynamics and the need for targeted interventions to enhance safety while promoting bike usage.

4 NEXT STEPS

The preliminary regression models provided valuable insights but were not conclusive, indicating that further, more robust analyses are necessary. They reveal a statistically significant relationship but not exactly what it is, it isn't linear or strictly logarithmic either. Additionally, while being statistically significant, the effect size of bike lane additions on accidents is relatively small, making it harder to model as well. However, the models serve as a great exploratory analysis into what the next steps of the research should be.

4.1 Decision Trees: Uncovering Nonlinear Interactions and Thresholds

Incorporating Decision Trees emerged as a logical next step. This method excels in capturing non-linear patterns and identifying thresholds where the relationship between bike lanes and injury rates shifts. Decision Trees do not assume proportionality in outcomes with each added mile of bike lane but instead reveal critical junctures, such as when lane density effectively separates cyclists from vehicle traffic, leading to significant changes in safety outcomes.

Decision Trees also automatically detect interactions between variables without needing explicit specification. For example, they could reveal how the effectiveness of bike lanes varies in high-traffic versus low-traffic areas or across different seasons—both relevant factors given the complex data interactions. Given that the decision-makers involved in bike lane installation are often

government workers who require a clear and valid basis for their decisions, the interpretability of Decision Trees is especially valuable. Their branching structure provides intuitive insights into the conditions where bike lanes most effectively reduce injury rates, aiding stakeholders in setting context-specific thresholds for bike lane expansions. Ensemble methods like Random Forests may also be explored to enhance these findings, improving reliability by combining multiple trees to capture subtle trends across cities and time periods.

4.2 Support Vector Machines (SVM): Capturing Complex Patterns with Limited Data

Following Decision Trees, SVM stands as the next viable approach, given its strength in managing complex, non-linear relationships even with smaller datasets. Utilizing the “kernel trick,” SVM can map data into higher dimensions, capturing nuanced relationships between variables. For instance, SVM could reveal where safety improvements plateau at specific bike lane densities, distinguishing high- and low-injury areas shaped by factors such as traffic volume and seasonality. With its margin-based optimization, SVM offers a robust model resistant to overfitting, enabling generalization across cities like New York, San Francisco, and Chicago, where dataset size and data diversity pose challenges.

4.3 Expanding the Analysis through Cross-City Comparison and Complex Relationships

Comparing New York City, San Francisco, and Chicago will help validate the findings and test their generalizability across different urban settings, revealing any significant cross-city differences in how bike infrastructure influences safety. However, due to the lack of traffic volume data in Chicago, these comparisons will focus on models that exclude traffic variables, ensuring consistency across datasets.

Additionally, the analysis will examine more complex relationships within the data. Interactions between bike lane length, traffic volume, and seasonality may reveal nonlinear dynamics that basic linear models cannot capture. Advanced models could potentially uncover latent variables or mediators that better explain variations in accident rates and bike usage.

4.4 Exploring Causal Inference Techniques

Finally, applying causality-focused methods will clarify the direction and strength of the relationships between bike lane additions, bike usage, traffic patterns, and accident rates. Techniques such as difference-in-differences (DID) and instrumental variable (IV) methods will help determine if bike lane expansions directly impact bike-to-traffic ratios and reduce injuries and fatalities. These methods address potential confounding factors, strengthening the research's analytical framework for a deeper understanding of how urban cycling infrastructure affects transportation and safety outcomes.

4.5 Consideration of Neural Networks

Neural Networks were initially considered as part of the analysis but were ultimately discarded due to the dataset's relatively small size, which limits their effectiveness. Additionally, Neural Networks lack interpretability, making them less suitable in a setting where transparent and justifiable insights are crucial for government decision-makers involved in bike lane planning and safety improvements.

Appendix

The data for this project was sourced from various locations and compiled into distinct datasets for each city, containing all relevant information. These datasets, along with the code used to create and analyze them, can be found in the following GitHub repository. Please note that the original data files are too large to upload there. I will work on finding an alternative method to share these files. The code for the analyses done are found there is well. As of this writing this code can be found in a file called Draft.rmd. Additionally, all figures and tables in this document were generated using R code from this repository.

[Shayaeng/Data698 \(github.com\)](https://github.com/Shayaeng/Data698)

References

1. **Combs, T. S., & Pardo, C. F. (2021).** Shifting streets COVID-19 mobility data: Findings from a global dataset and a research agenda for transport planning and policy. *Transportation Research Interdisciplinary Perspectives*, 9, 100322. <https://doi.org/10.1016/j.trip.2021.100322>
2. **Rérat, P., Haldimann, L., & Widmer, H. (2022).** Cycling in the era of Covid-19: The effects of the pandemic and pop-up cycle lanes on cycling practices. *Transportation Research Interdisciplinary Perspectives*, 15, 100677. <https://doi.org/10.1016/j.trip.2022.100677>
3. **Karpinski, E. (2021).** Estimating the effect of protected bike lanes on bike-share ridership in Boston: A case study on Commonwealth Avenue. *Case Studies on Transport Policy*, 9(3), 1313-1323. <https://doi.org/10.1016/j.cstp.2021.06.015>
4. **Buck, D., & Buehler, R. (2011).** Bike Lanes and Other Determinants of Capital Bikeshare Trips. TRB 2012 Annual Meeting Paper. Virginia Tech Alexandria Center.
5. **Prud'homme, R., & Lee, C.-W. (1999).** Size, Sprawl, Speed and the Efficiency of Cities. *Urban Studies*, 36(11), 1849-1858. <http://www.jstor.org/stable/43198143>
6. **Hartgen, D. T., & Fields, M. G. (2009).** Gridlock and Growth: The Effect of Traffic Congestion on Regional Economic Performance. Reason Foundation.
7. **Kraus, S., & Koch, N. (2021).** Provisional COVID-19 infrastructure induces large, rapid increases in cycling. *Proceedings of the National Academy of Sciences*, 118(15), e2024399118. <https://doi.org/10.1073/pnas.2024399118>