HeadlightsAndAccidents*

Catharina Castillo

December 10, 2025

Abstract goes here.

“‘

1 Introduction

There is much anecdotal evidence supporting the notion that bright headlights are a significant concern and bother for many drivers (Lockert 2025). New headlights are perceived as brighter due to their color and because they are LEDs. While we don't have exact timelines for the deployment of new headlights, 2007 marks the first year cars were available with these headlights, though it was only higher-end cars. As time has passed, LED headlights are now available in most new cars sold (Childress 2025). The National Highway Traffic Safety Administration estimates that about 75% of cars sold in 2023 had LED headlights (Lockert 2025).

Most research now, especially that which draws off findings from the Insurance Institute for Highway Safety(IIHS), states that these new headlights are in fact safer and that glare from these headlights hasn't yet been linked to increases in nighttime car crashes (Delisio 2025).

It is still recognized that bright headlights are a distraction.One amendment that has been approved for deployment in the USA is Adaptive Driving Beam Headlights, which will turn off various LEDs in the headlight array to reduce glare while still keeping the high visibility (Shine 2023). While this was approved in 2022, ADB hasn't been implemented in the US due to compliance challenges car manufacturers face (Lockert 2025).

In the meantime, more and more cars on the road have brighter headlights. Considering the high volume of anecdotal reports that these are distracting and often cause "near misses", as

*Project repository available at: https://github.com/NettleHook/HeadlightsAndAccidents.

well as the scarce research into the potential connection between increasing perceived brightness of headlights and nighttime car accidents, this research paper seeks to use linear regression to explore if there is a connection between time and accidents caused by headlight glare.

Our research will be focused on California, using data obtained from the California Highway Patrol's California Crash Reporting System. Due to the size of that data, we will be limiting our study to accidents that were recorded to have occured in poor lighting that would legally require headlights to be on.

In Section 2, we'll discuss a study done by the IIHS that seeks to explore the same topic. In Section 3, we'll cover the data we used, as well as steps taken to use it for our analysis. Section 4 introduces the final model, while Section 5 introduces an alternate we considered and why it wasn't pursued. Section 6 introduces the results of our analysis, and Section 7 presents the implications of these results, and other limiting factors that may be of consideration.

2 Previous Literature

The Insurance Institute for Highway Safety(IIHS) published published the results of a study in October 2025 where they used data from 11 states. The states included in the study were Connecticut, Florida, Iowa, Illinois, Massachusetts, Michigan, New York, Texas, Washington, and Wisconsin. There was a range of years covered, but generally the data covers from 2013-2016 up to 2023.

#include information on the model/methods they used

The results from the IIHS study was that there was an extremely low number of accidents that listed glare from headlights as a contributing factor, and that there was little change between 2015-2023, which was cited as a period where a lot of changes to headlights occurred. They did find, however, that there was a disproportionate impact of glare on accidents involving 70- and 80- year-old drivers (Brumbelow 2025).

3 Data

Will need to introduce data here, discuss limiting factors:

–Combined crashes and parties datasets so we could get age data and all factors that were noted as part of the accident. Original columns of interest prior to cleaning were: “Collision Id”, “Crash Date Time”, “Day Of Week”, “LightingCode”, “Primary Collision Factor Violation” from the crashes dataset and “CollisionId”, “Other Associate Factor”, “StatedAge” from the parties dataset.

Discuss type of merge used, and what factors will now be left out as a result

Turn this into a table: Final columns in our cleaned dataset were: prop -> continuous variable, if headlights were listed as one of the factors in this crash. This will be our response variable elderlyInvolved -> Boolean, if one of the parties in the crash was 70 or older. Used the DMV's age limit before drivers licenses need to be renewed more frequently. DayofWeek -> Categorical, Day of the Week the accident occured. Included to account for variability in traffic due to the day of the week and commuter patterns. Month -> Categorical, included to account for variability due to monthly changes in traffic patterns Year -> Discrete Numerical,

–some inconsistencies in data noting. Sometimes it was specified if glare/light-based vision obscurement was caused by headlights, but not always. In cases where headlights + head beams weren't specifically noted, we checked the lighting code to confirm if it happened in cases where headlights should be on.

–we also included cases when headlights were not on but it was noted that they should've been. This is still relevant to our case of concern, as the brightness of driving lights + the brightness of other's headlights mean some people drive without headlights and aren't aware that they are off. In this way, headlights being off is still related to the brightness of headlights and a potential link to increased accidents due to brighter headlights. Should be noticed we took care to eliminate any cases caused by "dim headlights".

3.1 Missing Data

There was some missing data. It was handled as follows: missing data from Age -> assumed not elderly missing data from LightingCode -> assumed daytime missing factors -> considered cases where headlights weren't involved

3.2 Limitations of the Data

Only available for 2016-2025, which doesn't cover the introduction of LED headlights. Can only extrapolate conclusions based on increasing deployment of brighter headlights in newer cars. All data is human entered, and has slight inconsistencies depending on usage. Primary concern is that "Vision Obscurement" was listed many times without specifying what was causing the obscurement–could have more cases involving headlights that revealed by our search. We assume that the potential inconsistencies[measurement error?] in reporting have at least remained constant over the 9 years we have data for. One glaring limitation is that due to the size of the datasets, we reduced to just accidents that were recorded to have occurred in lighting that would necessitate headlights being on (Dark, Rainy…). Furthermore, due to the high proportion of days without any accidents with headlights listed as a factor in accidents, we excluded days with no recorded accidents from headlights. Table 1 shows the number of days with no reported accidents from headlights by Year.

Plotting the proportion of accidents where headlights were a listed factor against the year we get Figure 1.

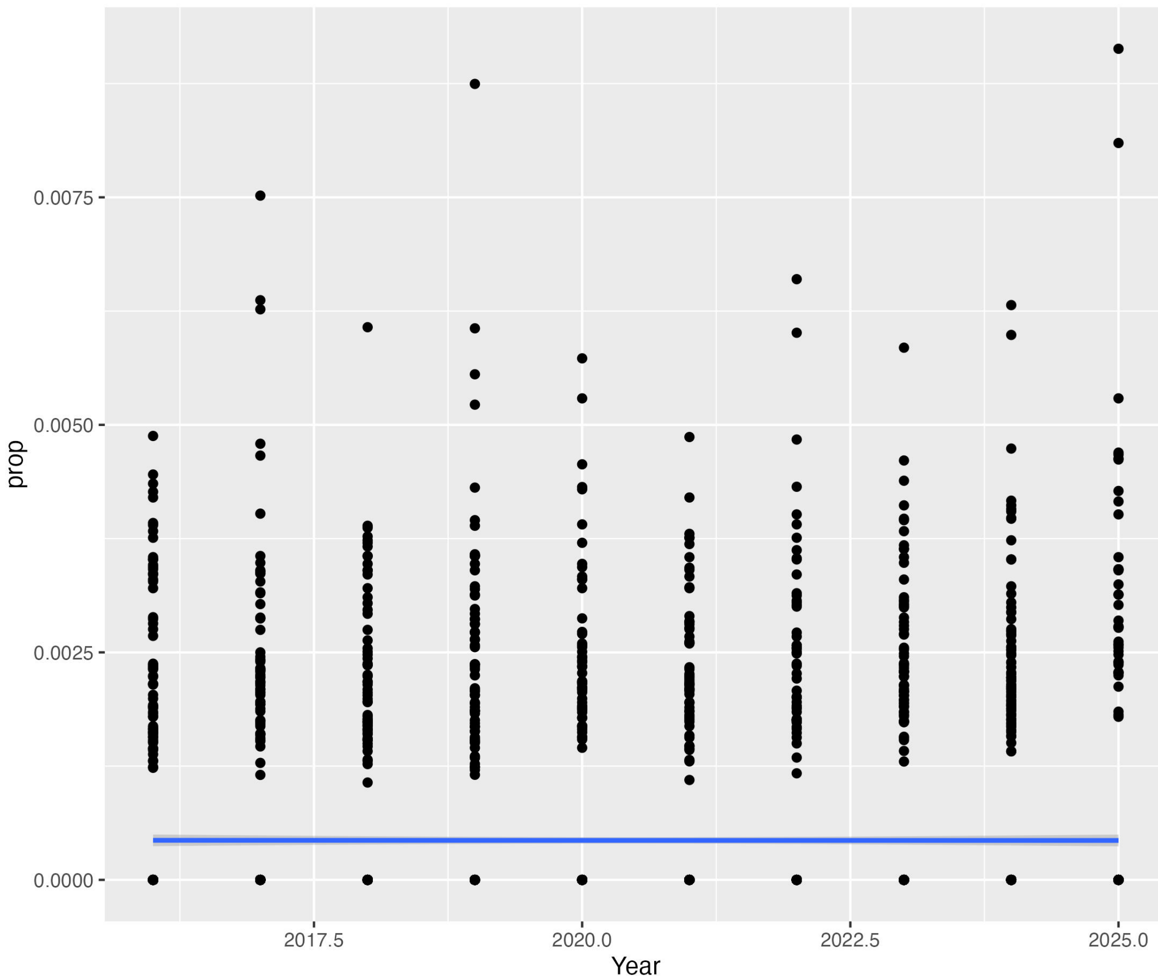

Figure 1: Plot of proportion of headlight accidents against year

The line is very flat, which leaves some concern for interpretability, as we are concerned in the trend of the non-zero points.

Attempting a log transform of the proportion and fitting a LOWESS curve forces the zero points to be dropped, but reveals a positive linear trend, shown in Figure 2.

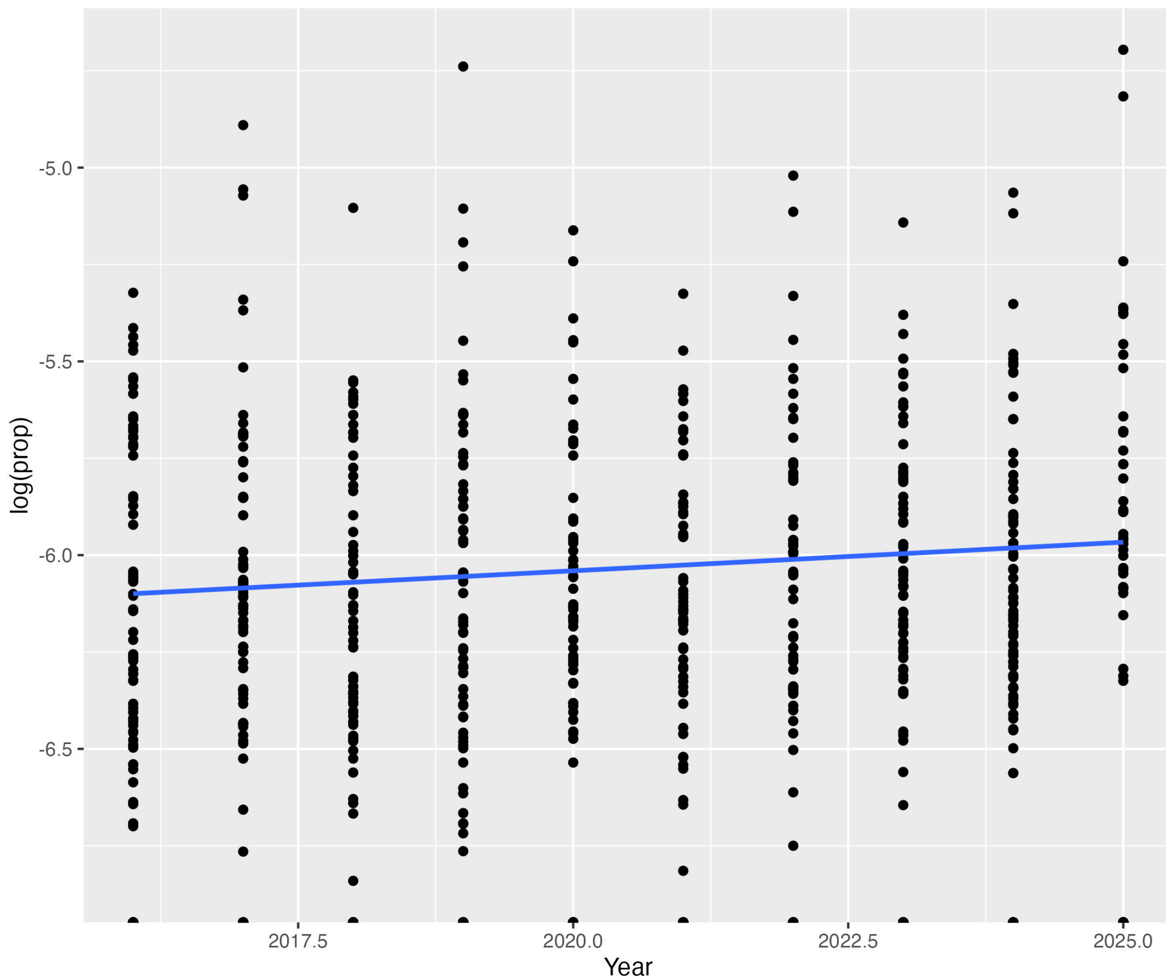

Figure 2: Plot of log(proportion of headlight accidents) against year with a loess curve

Upon examination of the data, we found that the number of days where no accidents involving headlights was not consistent from year to year, with the exception of 2020-2022, though this is likely due to the Covid-19 pandemic. (**table?**) shows the total of days with and without accidents caused by headlights

2016 - 299 2017 - 300 2018 - 301 2019 - 298 2020 - 312 2021 - 309 2022 - 308 2023 - 298 2024 - 289 #double check last date for 2024 data too 2025 - 282 #data gathered only until–find date **Need to emphasize that days with no crashes increased over full dataset

Figure 3 shows to plot of days without accidents caused by headlights against the year. We can see there is no clear linear trend, so removing the days without accidents will certainly invalidate any results we obtain.

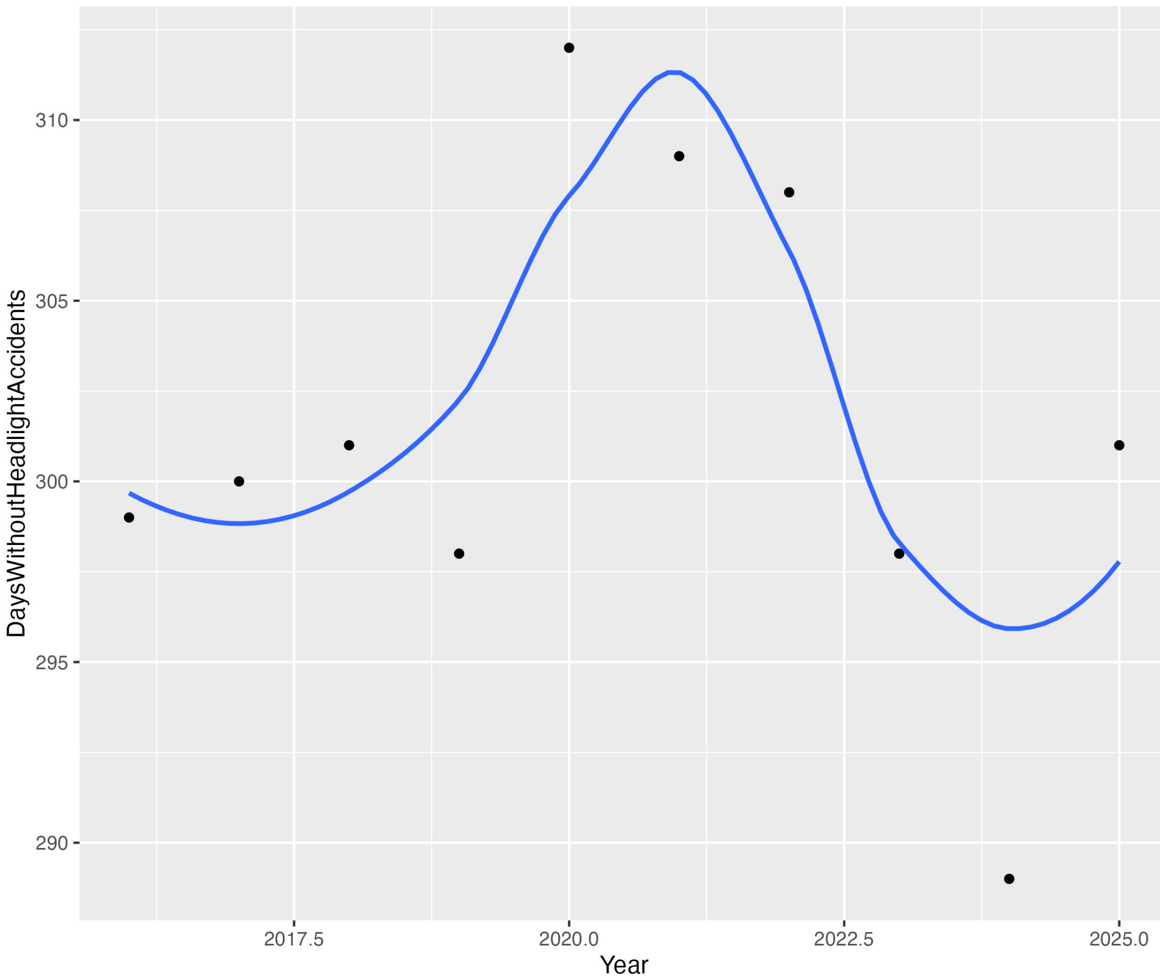

Figure 3: Plot of number of days without accidents involving headlights against the year

Since we will be using a log transform on the data, and our response variable is a proportion, we will proceed by measuring the converse.

4 Methods

We will be using the model:

$$log(Y_i) = \beta_0 + \hat{\beta_D} * \hat{X_D} + \hat{\beta_M} * \hat{X_M} + \beta_T * X_T + \epsilon_i$$

For concision, the categorical variables and their coefficients have been represented with vector notation. Y_i represents the proportion of accidents for the day that were not caused by headlights.

$\hat{X_D}$ represents the categorical variables for the day of the week. The default level represents Sunday. $\hat{\beta_D}$ is vector of coefficients for each weekday except Sunday. $\hat{X_M}$ is the vector of categorical variables representing the month of the year. The default level is January. $\hat{\beta_M}$ is the vector of coefficients for each month except January. Both $\hat{\beta_D}$ and $\hat{\beta_M}$ alter the intercept of the model based on the day of the week and the month, respectively. #describe how

X_T is the coefficient representing the year. With all other variables held constant, we expect a unit change in X_T will result in Y_i increasing by a factor of e^{β_T}. ϵ_i represents the errors.

We will also need to conduct a hypothesis test to affirm that there is a statistically significant relationship between year and the proportion of car accidents not involving headlights. We will use the null hypothesis $\beta_T = 0$ and alternate hypothesis $\beta_T \neq 0$. We will the test statistic $t^* = \frac{b_T}{s\{b_T\}}$. Using a 95% confidence level, our test statistic will need to be compared with $t(1 - \frac{\alpha}{2}; n - 2) = t(0.975; 1549797 - 19) = 1.96$.

The analysis has been implemented using the R-language (R Core Team 2025).

5 Other Models

The IIHS study found that for older drivers, brighter headlights did increase the rate of accidents. As such, we also explored a model that incorporated a categorical variable representing if a participant in the accident was elderly as well as an interaction effect between this variable and the year variable. We defined the cutoff for the elderly category at 70 years old, following when the DMV requires liscences to be renewed more frequently. The model that encorporates the age categorical variable is the following:

$$log(Y_i) = \beta_0 + \hat{\beta_D} * \hat{X_D} + \hat{\beta_M} * \hat{X_M} + \beta_A * X_A + \beta_T * X_T + \beta_{TA} * X_T * X_A + \epsilon_i$$

X_A is the boolean categorical variable representing if someone elderly was involved in the recorded accident. β_A represents the change to the intercept if X_A is true. Due to the interaction effect, if someone elderly is involved in the accident the change to Y_i is a factor of $e^{\beta_T + \beta_{TA}}$.

Since we selected our variables prior to building any models, the [what are the variable selection methods by metric called?]. However, the model involving the age categorical variable had a standard error greater than the estimated coefficient for both X_A and $X_T * X_A$. The results are shown in Table 1.

Table 1: Coefficient estimates of b_A and b_{TA} and the standard errors

b_A	$s(b_A)$	b_{TA}	$s(b_{TA})$
-0.001519	0.003701	0.0000007488	0.000001831

As the standard error is so large, we cannot reliably make a determination about what effect the age categorical variable has on the proportion of accidents that don't involve headlights.

Further, by comparing the residuals of the two models, we can see that the X_A variable from this dataset does not explain much of the variability. Figure 4 shows the residuals for this model, which can be compared with Figure 5, which shows the residuals for the model we used. The two plots are fairly similar, aside from the duplicate points in the model with the age categorical variable.

6 Results

The results for the model give that b_T is -3.002e-06, with a standard error of 2.916e-07.

The t-value for this is -10.296. Comparing against our student AAAA. $abs(-10.296) > 1.96$, so we conclude that this may be statistically significant.

However, it is worth noting that the assumptions about errors that are required for the hypothesis test to apply have been broken. The residual graph can be viewed in Figure 5. Our response variable is bound between 0 and 1, meaning the errors can't be normally distributed (?). Additionally, the data points are heavily skewed towards Y_i = 1, which breaks the assumption of constant variance (?).

7 Discussion

While our hypothesis test may not be reliable, there are other limitations even if we assume the relationship between Year and the proportion of accidents not involving headlights is accurate. While the negative slope does imply that headlights are causing more accidents in poor lighting, the coefficient is so small that the effect may be negligible. #might want to calculate how many accidents this ends up being for one of the years.

Furthermore, it has been separately noted that there has been an increase in car crashes post-pandemic ("Traffic Safety Impact of the COVID-19 Pandemic: FatalCrashes in 2020-2022" 2024), though we've attempted to limit this by making our response variable the proportion of all accidents that occur in conditions requiring headlights. It is still possible that an increase in car accidents involving headlights may be related to the increase in accidents overall.

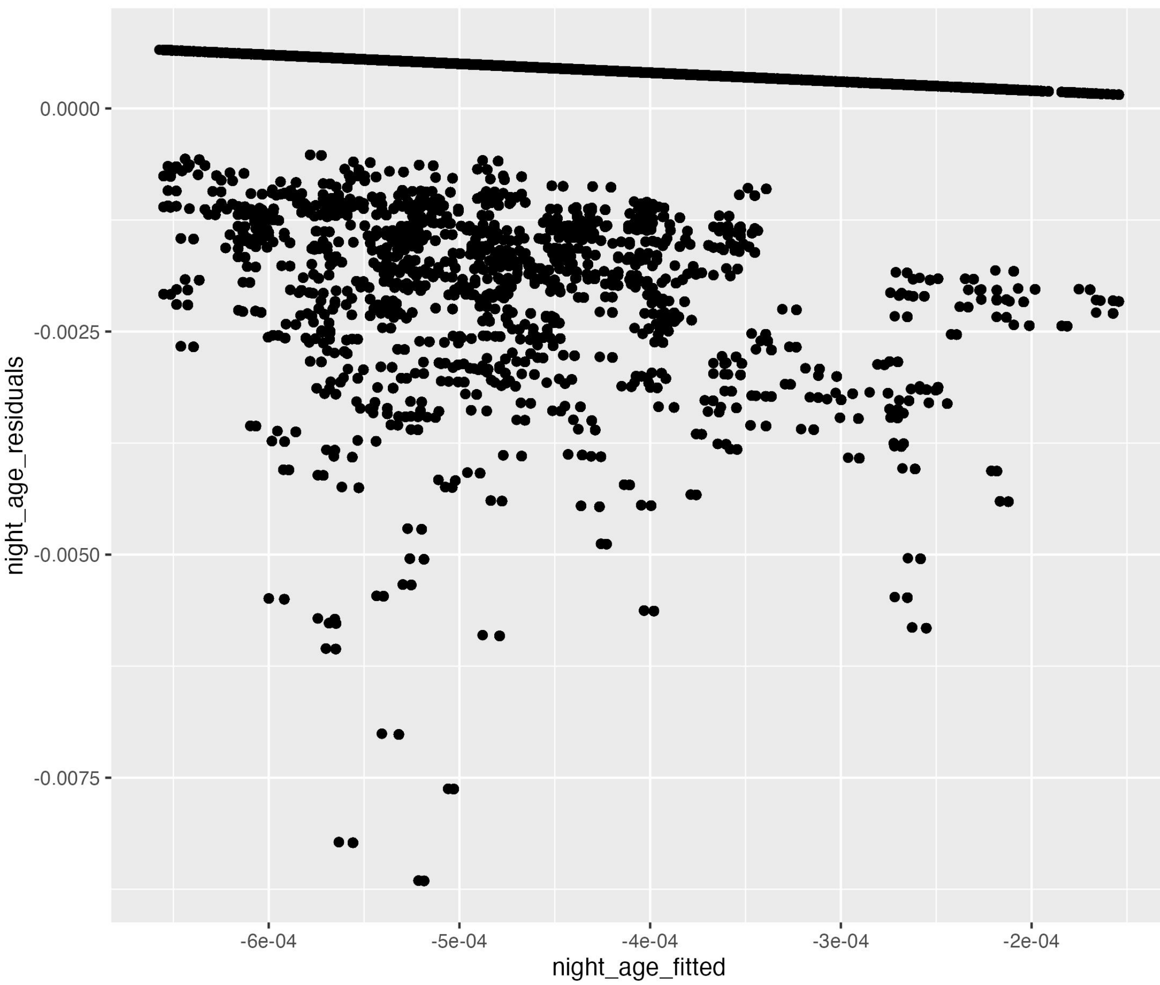

Figure 4: A plot of the residual against the fitted values for the model that incorporates a variable for age

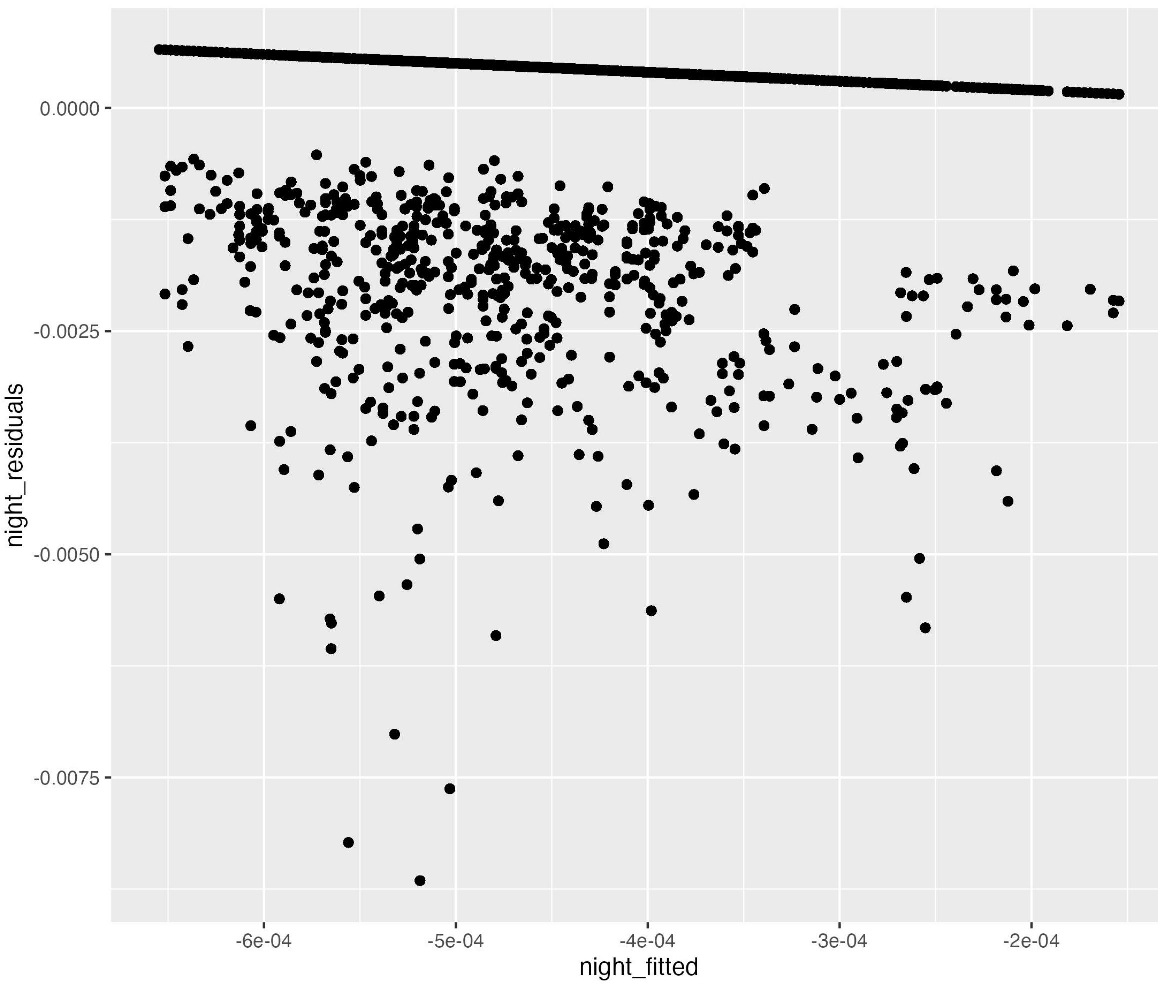

Figure 5: A plot of the residual against the fitted values for the model that doesn't incorporate a variable for age

Another factor that has already been shown to be resulting in more dangerous car accidents is larger truck size ("Vehicles with Higher, More Vertical Front Ends Pose Greater Risk to Pedestrians2" 2023). This has also been cited as a cause behind increased headlight glare as well. Larger SUVs and LTVs are more likely to have their headlights at a level where they reflect off the rearview and sideview mirrors of passenger cars (Lockert 2025). If there is an increase in the number of car accidents involving headlights, it's possible it could be due to the fact that SUVs and LTVs are getting larger rather than more cars having bright headlights.

Finally, there are other factors that could be counteracting the effects of brighter headlights, such as new safety features in cars. These would include features such as Lane Keep Assist, but also automatic headlights.

–

Brumbelow, Matthew L. 2025. "Headlight Glare in Police-Reported Crash Data: Prevalence, Contributing Factors, and Potential Effects." IIHS. https://www.iihs.org/api/datastoredocument/bibliography/2347.

Childress, Ally. 2025. "Are Car Headlights Much Brighter All of a Sudden—or Is It Your Imagination?" Reader's Digest. https://www.rd.com/article/car-headlight-brightness/.

Delisio, Ellen R. 2025. "What's the Story with Headlight Glare? Newer LED Headlights Seem Brighter and More Blinding Than Their Halogen Predecessors. But Change Is on the Way." YourAAAToday. https://magazine.northeast.aaa.com/daily/life/cars-trucks/whats-the-story-with-headlight-glare/.

Lockert, Melanie. 2025. "Why Do Headlights Seem so Bright? Headlights Are in Fact Brighter These Days. Here's What You Can Do about It." AAA. https://mwg.aaa.com/via/car/why-do-headlights-seem-so-bright.

R Core Team. 2025. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Shine. 2023. "Matrix LED Headlights: Redefine Adaptive Front-Lighting with Smart High Beam Technology." Global Lighting Forum. https://www.shine.lighting/threads/matrix-led-headlights-redefine-adaptive-front-lighting-with-smart-high-beam-technology.79/.

"Traffic Safety Impact of the COVID-19 Pandemic: FatalCrashes in 2020-2022." 2024. AAA Foundation for Traffic Safety. https://aaafoundation.org/wp-content/uploads/2024/07/202407-AAAFTS-Impact-of-COVID.pdf.

"Vehicles with Higher, More Vertical Front Ends Pose Greater Risk to Pedestrians2." 2023. IIHS. https://www.iihs.org/news/detail/vehicles-with-higher-more-vertical-front-ends-pose-greater-risk-to-pedestrians.