Evaluating the Rate of Accidents Involving Headlights in California*

Are Brighter Headlights Causing More Accidents?

Catharina Castillo

December 10, 2025

LED headlights in vehicles are recognized as being safer, due to the increased visibility. However, many drivers also report that these brighter headlights are distracting due to glare and vision obscurement. This increased distraction has many concerned that brighter headlights might in fact be causing more accidents. This paper explores if there is a statistically significant relationship between the proportion of accidents involving headlights to all accidents occurring in poor lighting and the year using a multiple linear regression model. Considering the small value of the coefficient found and the multitude of compounding factors, this paper did not reach a concrete conclusion with regards to whether brighter headlights are causing more accidents.

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 LED 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).

Research into the link between LED headlights and car accidents, especially that which draws off findings from the Insurance Institute for Highway Safety(IIHS), states that these new

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

headlights are 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 (ADB), which will turn off various LEDs in the headlight array to reduce glare for other drivers 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 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 analysis to accidents that were recorded to have occurred 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 model 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 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.

The IIHS used nearest-neighbor matching to match drivers in glare-related crashes to drivers in non-glare related crashes. They used two sets to examine driver demographics and environmental conditions to see if there was an increase in glare-related crashes.

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

For this analysis we used data obtained from the California Highway Patrol (CHP)'s California Crash Reporting System, posted publicly at data.ca.gov/dataset/ccrs. The data is available for years 2016-2025. It is also updated daily. We used the newest datasets available on December 2, 2025 for the years 2016-2023 and the newest datasets available on December 7, 2025 for the years 2024-2025.

Both the Crashes and the Parties datasets for 2016-2025 were used. The Crashes dataset is a record of each collision recorded by the CHP. Other information it contains that is relevant to our analysis is the crash date, day of the week, lighting information, and the primary collision factor. The Parties dataset is a record of every party involved in the collision that CHP responded to. The information we took from that dataset was the other factors associated with the crash, as well as the recorded age information for a second model we had considered. More information about the second model can be found in Section 5.

We performed a one-to-many merge using the Collision ID recorded in both datasets.

Due to the size of the data after everything was merged, we decided to reduce our dataset to only consider the crashes that had been recorded as occurring in lighting that required headlights. This corresponds to the following lighting codes: B-DUSK-DAWN, C-DARK-STREET LIGHTS, D- DARK-NO STREET LIGHTS, E-DARK-STREET LIGHTS NOT FUNCTIONING.

Table 1 displays the variables we used in our model, as well as a brief description of what they represent. Day of the Week and Month were included in the model to account for variability in accidents that occur with the changes in traffic patterns on different days of the week and the month.

Table 1: Variables used in the model(s) and a brief description of what they represent. *The elderlyInvolved variable was used for a second model that had been under consideration, discussed further in Section 5.

Name	Type	Description
prop	Continuous Numerical. Bound between 0 and 1	The proportion of all accidents that occurred on the same day that involved headlights as a factor
prop_complement	Continuous Numerical. Bound between 0 and 1	The proportion of all accidents that occurred on the same day that did not involve headlights as a factor
DayofWeek	Categorical	Day of the week the accident occurred on

Name	Type	Description
Month	Categorical	Month the accident occurred on
Year	Discrete Numerical	Year the accident occurred on
elderlyInvolved*	Categorical	1 if someone 70 years old or older was involved in the accident, 0 otherwise.

3.1 Encoding Headlight Involvement

Sometimes headlights were listed as a factor in plain language. Even then, there were some inconsistencies in the record, such as “HEAD LIGHT” and “HDLIGHT”. We made sure to include these instances.

We also included factors listing high beams and headlights being off. The reasoning behind including high beams is that the driver may not be aware that they have their high beams on instead of low beams due to the brightness from the driver’s seat. Additionally, we will be including some instances of glare, but there isn’t a way to distinguish if the glare comes from low beams or high beams from the recorded information. We included headlights being off, as both the brightness of driving lights and the light from other car’s headlights can make it harder for drivers to recognize that their headlights are not on.

Sometimes the accident factors in the Parties database listed glare as a factor, without specifying where the glare came from. We followed the IIHS method of assuming glare in poor lighting was due to headlights (Brumbelow 2025). Our data is already filtered to only include accidents in poor lighting, so the encoding process for accidents with glare only involved confirming that “SUN” wasn’t included as a reason.

Finally, we checked the listed vehicle codes if any were provided. From examining the dataset obtained from oag.ca.gov, the vehicle codes we filtered on were 24250, 24604, 24400-24410, 24800, 25250, 25650, 25651, and 26100, which are all related to violations relating to headlight brightness. We did not include violations related to headlight dimness or color. If any of the previously described conditions were met for a collision, that accident was marked as involving headlights.

3.2 Limitations of the Data

There were a few limitations of the dataset. First, is the time period covered. 2016 is the first year available, while LED headlights were first introduced in 2007. This means we might be limited in examining accident patterns before LED headlights became commonplace. Additionally, our dataset covers 2020-2022, during the COVID-19 pandemic. At this time, there were less people on the road. After restrictions lessened, crashes did increase, which may influence the patterns we see ("Traffic Safety Impact of the COVID-19 Pandemic: Fatal Crashes in 2020-2022" 2024). We have attempted to mitigate the effect of the increasing or decreasing number of crashes with time by representing our response variable as a proportion.

All data in the CCRS datasets is human entered, and has slight inconsistencies depending on usage as well as ambiguities. We assume that the potential inconsistencies in reporting have at least remained constant over the 10 years we have data for. One example of ambiguity is that "Vision Obscurement" was listed many times without specifying what was causing the obscurement. As discussed in Section 3.1, we filtered out sun glare, and are working with data on accidents that occurred in poor lighting. Still, due to the ambiguity and our filtering, it's possible we could have more cases involving headlights than were revealed by our search.

There is also a limitation with the roads that may be covered by the CCRS dataset. The California Highway Patrol is responsible for California freeways and other state roads. It is not clear if the CCRS dataset is only collisions that the CHP responds to, and so it isn't certain if this data is limited to freeways. The IIHS study did find that glare-related accidents were more likely to be caused on two-lane roads, which are generally not covered by the CHP.

3.3 The Zero Issue

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

The line is very flat, which leaves some concern for interpretability. Additionally, all our proportion data consists of tiny values. We will want to consider a transformation to increase interpretability.

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

Upon examination of the data, we found that the number of days with no accidents involving headlights was not consistent over the entire period of our consideration. There is some consistency in segments, such as the pre-pandemic years (2016-2019) and the pandemic years (2020-2022). However, looking at all 10 years together, there isn't a clear pattern.

Table 2 shows the total of days with and without accidents caused by headlights

Table 2: Table displaying the number of days with and without accidents caused by headlights, as well as the total. 2025 has less days due to when the dataset was downloaded.

Year	Days without Headlight Accidents	Days with Headlight Accidents	Total
2016	299	67	366
2017	300	65	365

Year	Days without Headlight Accidents	Days with Headlight Accidents	Total
2018	301	64	365
2019	298	67	365
2020	312	54	366
2021	309	56	365
2022	308	57	365
2023	298	67	365
2024	289	77	366
2025	301	39	340

Figure 3 shows a 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.

Since we will be using a log transformation on the data, the zeroes pose a significant problem. As there is never a case where all recorded accidents in poor lighting for a day were caused by headlights, we proceed with running the regression on the complement. We added the prop_complement variable, which measures the proportion of accidents that did not involve headlights. This eliminates the zero problem that occurs when attempting the log transformation.

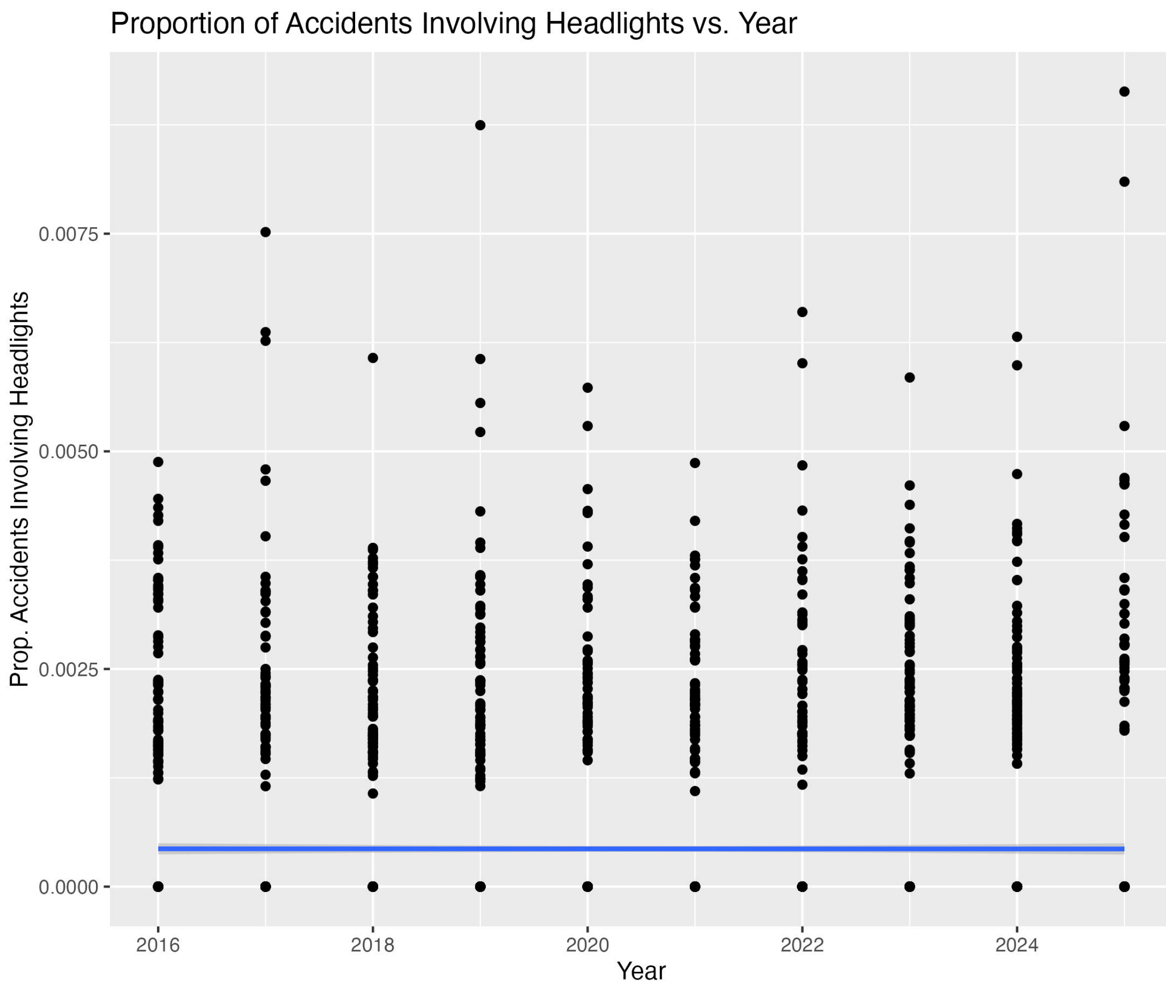

Figure 1: Plot of proportion of accidents involving headlights against year

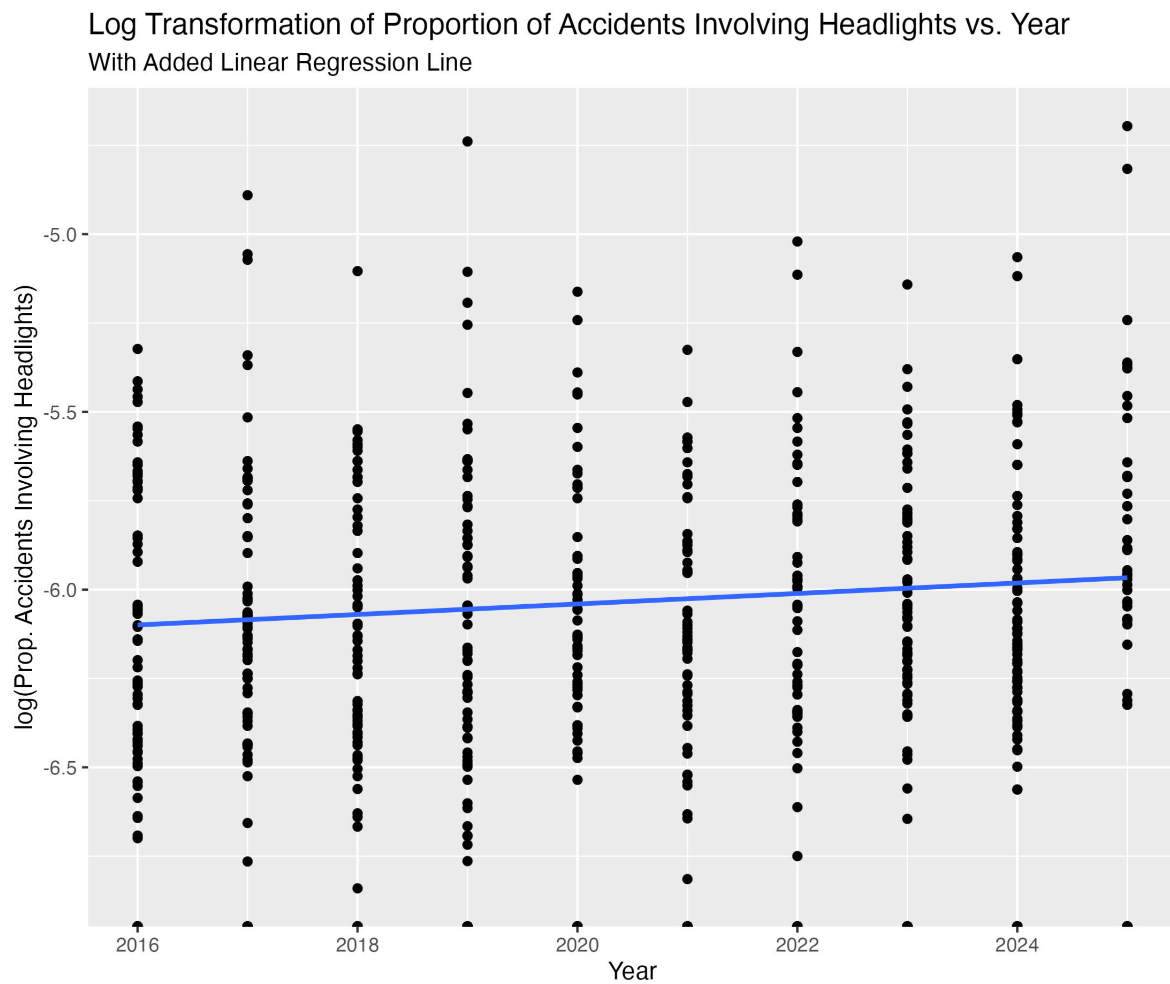

Figure 2: Plot of log(proportion of accidents involving headlights) against year with a linear regression line. Zeroes have been dropped

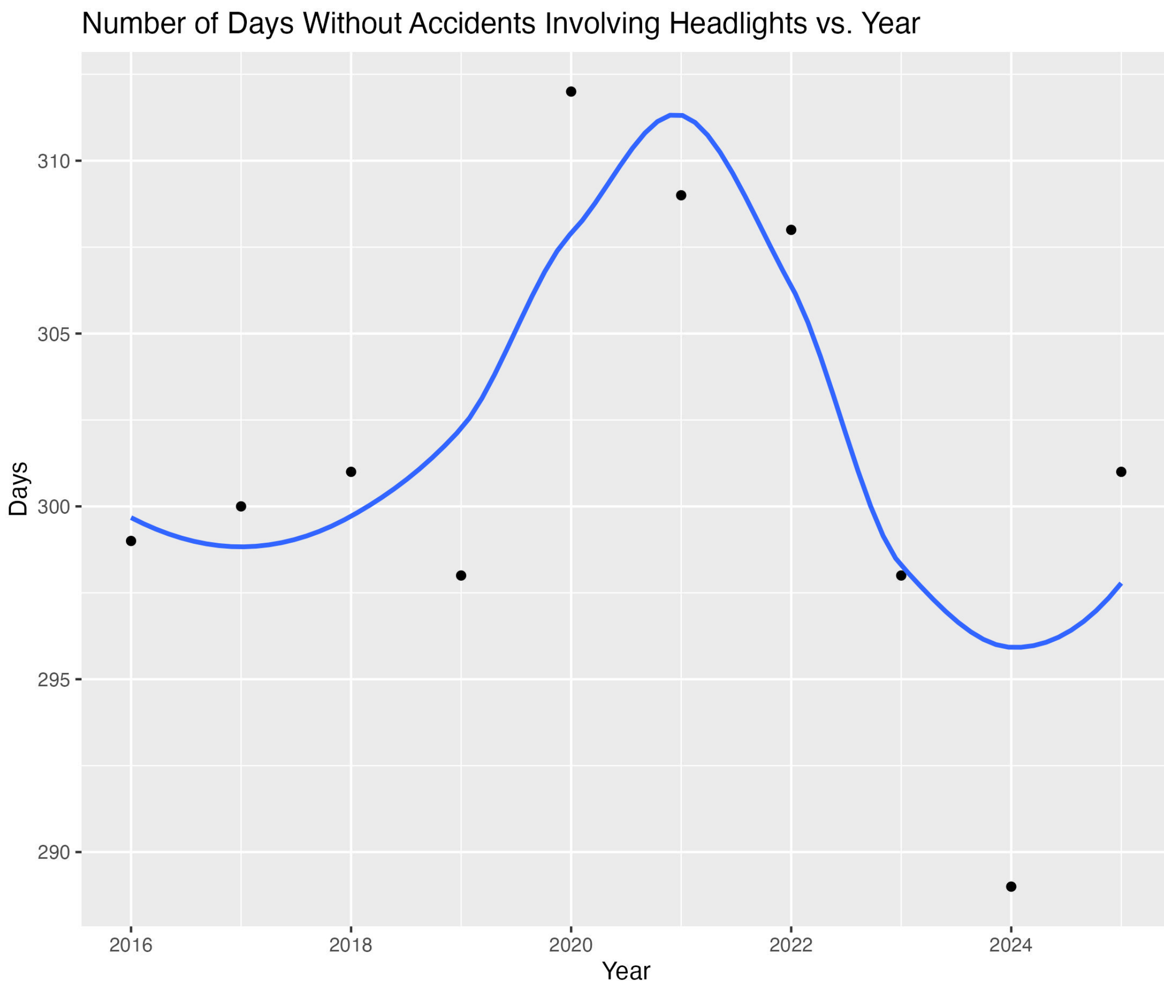

Figure 3: Plot of number of days without accidents involving headlights against the year

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 a 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 model based on the day of the week and the month, respectively. Holding the year constant, Y would change by a factor of $e^{\hat{\beta_{D,j}}+\hat{\beta_{M,k}}}$, where j represents the index of $\hat{\beta_D}$ that is the coefficient corresponding to the day of the week and k represents the index of $\hat{\beta_M}$ that is the coefficient corresponding to the month.

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 use 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 ability to apply this type of hypothesis test has key assumptions that should be met to assure its validity. One of the assumptions is that errors are independent between observations. Another is that the variance of the errors remains constant. Finally, errors are assumed to be distributed in a way that approximates the normal distribution. We know we may already have some trouble with this last assumption, as the response variable is bound between 0 and 1.

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. We also explored a model that incorporated a categorical variable representing if a participant in the accident was elderly. This model also included an interaction effect between the new variable and the year variable. We defined the cutoff for the elderly category at 70 years old, following when the DMV requires licenses to be renewed more frequently. The model that incorporates 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 $e^{\beta_T + \beta_{TA}}$, given every other variable is held constant.

Since we selected our variables prior to building any models, we didn't use model criterion metrics such as Mallow's C_p or AIC to compare the two models. 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 3.

Table 3: 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 the estimate for β_T, b_T is -3.002e-06, with a standard error of 2.916e-07. The t-value for this is -10.296. Comparing this t-value against the value calculated in Section 4, we get $|-10.296| > 1.96$, so we would conclude that we reject the null hypothesis $\beta_T = 0$.

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 distributed normally. Additionally, the data points are heavily skewed towards $Y_i = 1$, which breaks the assumption of homogeneous variance. This makes our hypothesis test unreliable, and we are not able to rely on it to reject the null hypothesis which states that $\beta_T = 0$.

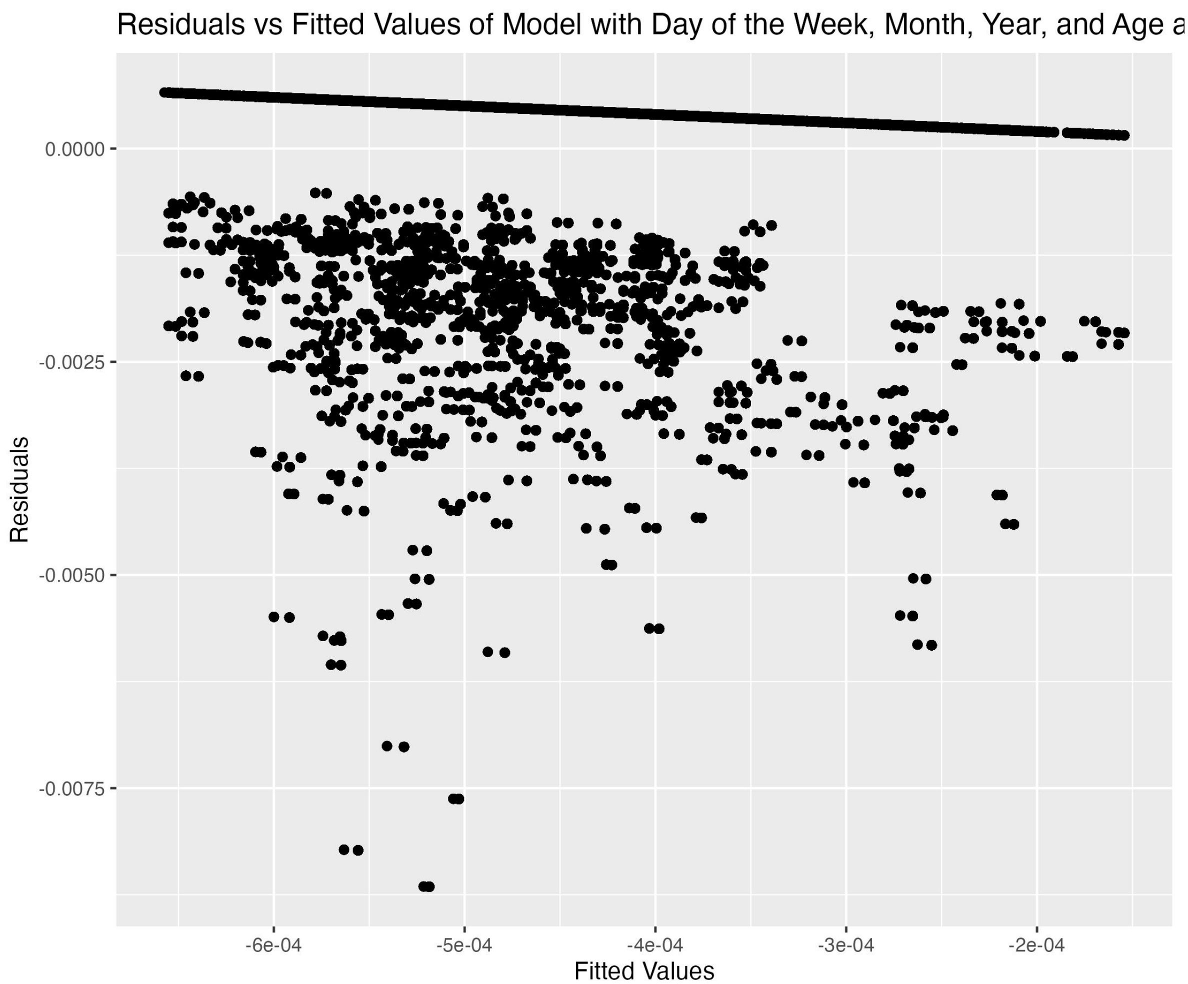

Figure 4: A plot of the residual against the fitted values for the model that incorporates a variable for age

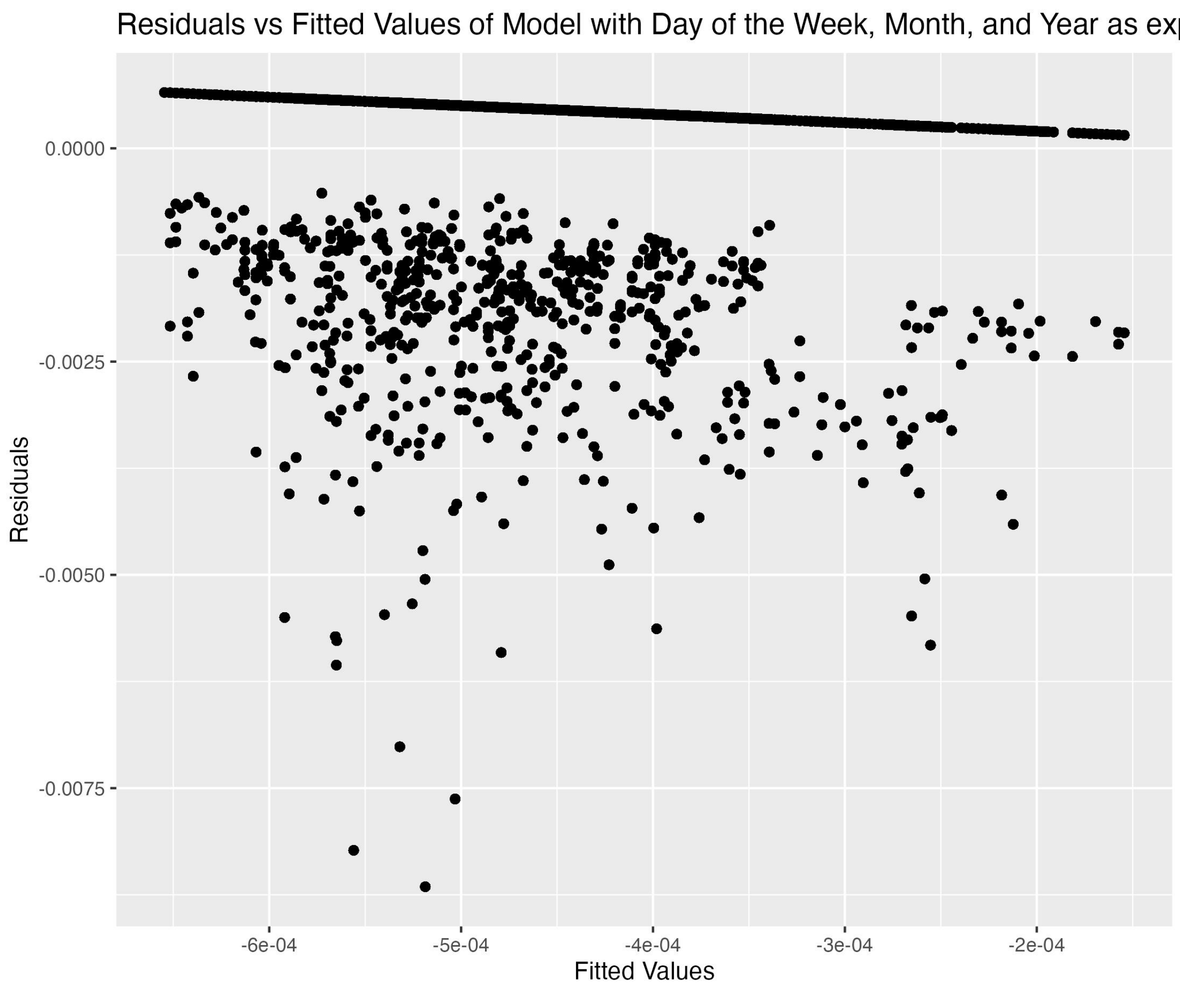

Figure 5: A plot of the residual against the fitted values for the model that doesn't incorporate a variable for age

7 Discussion

There are other limitations even if we assume there is a relationship between Yyar and the proportion of accidents not involving headlights. While the negative slope found by our model does imply that headlights are causing more accidents in poor lighting, the coefficient is so small that the effect may be negligible. This is compounded by the fact that our hypothesis test was unable to reject the null hypothesis. We have no certainty that there is a relationship between year and the proportion of accidents that involve headlights.

Even if we continue with the assumption that there is a relationship between year and the proportion of accidents that involve headlights, it has been separately noted that there has been an increase in car crashes post-pandemic ("Traffic Safety Impact of the COVID-19 Pandemic: Fatal Crashes 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.

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 and blind spot indicators, but also automatic headlights.

To conclude, despite the model revealing a slight negative relationship between time in years and the proportion of accidents not caused by headlights in poor lighting conditions, we cannot conclude that there is an increase in accidents caused by headlight glare with each year. There are too many compounding factors to consider. Additionally, we do have to keep in mind the limitations of the data used, which we described previously in Section 3.2.

–

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: Fatal Crashes 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.