Final Project Part 3: An Analysis of Personal vs Macroeconomic Factors in Car Accident Deaths

Isaac Wetmore (1005756674)

2022-12-21

Introduction:

When two or more individuals are involved in a car accident, blame is typically placed squarely on said individuals. Damages are usually paid to the victim by the defendant deemed to be at "fault." If a victim dies in the accident, a defendant may face serious criminal charges in many legal systems. Placing all the accountability on a single individual involved in a fatal accident, however, may fail the capture the full extent of blame. There can be many factors outside of an individual's control that effects not only the chance of an accident occurring, but also whether the accident is fatal. If a similar type of accident kills someone in country A, but not in country B due to stronger healthcare or road infrastructure, is it truly fair to hold individuals to the same standard of blame in country A and B?

Research into the personal lifestyle factors that affect traffic accidents is extremely extensive. A PubMed article search on the issue will net over 50 000 results alone. Several lifestyle factors have been shown to influence the chance of an accident, such as alcohol intake, personality traits, or even personal emotional factors like whether the individual was going through a divorce or was suffering from depression (Bon de Sousa et al., 2016). Research into the macroeconomic factors influencing traffic accidents, however, is comparatively more limited. Some influential studies have suggested factors such as GDP play a major role in rate of fatal accidents in a country (van Beeck et al., 2000). It was found that the relationship between GDP and car accident death followed a U-shaped curve. An increase in GDP was associated with an increase in car accidents for poorer countries, implying accidents went up as more people could afford traveling by automobile. After a certain point of GDP increase however, car accident deaths would begin to drop again. Another study further broke down this phenomenon, finding this U-shaped relationship disappeared when looking at non-fatal car accidents (Bishai et al., 2005). Together these papers hypothesize that the reason for less car accident deaths as the GDP increases beyond a certain point is due to the increase in health and road infrastructure that starts to slowly develop with greater wealth.

This report will attempt to build a bridge between these two different ways of understanding car accident deaths. Specific care will be spent comparing both lifestyle and macroeconomic factors to assess which subgroup better explains the variance of car accident deaths in a country. The analysis done here will attempt to answer the following research question: do lifestyle factors play a bigger role in the rate of fatal accidents, or are macroeconomic trends a better predictor?

Methods

We will utilize multivariate regression to decode which variables serve as better predictors of fatal car accidents in a country. This analysis will work with 5 variables split into two categories. The first category includes factors related to personal lifestyle choices that may increase the risk of a fatal traffic accident. The second category of variables attempt to address macro trends outside of an individual's control. All indicators come from OCED (fully referenced in bibliography). A summary of the variables, along with their units, is provided in table 1:

Table 1: Summarizing Predictor and Predicting Variables

Variables (Sorted into Categories)	Median	Standard Deviation
Macroeconomic Factors (Outside of Individual)		
GDP (US $ per captia)	40909	12401
Healthcare (% of GDP)	8.80%	2.03%
Road Maintenance (Euro per captia)	65.94	78.19
Personal Factors Factors (Individual Dependent)		
Mobile Broadband Subscriptions (per 100 inhabitants)	81.52	34.24
Alcohol Consumption (L per captia)	9.70	2.01
Dependent Variable		
Car Accident Deaths (per 1 million inhabitants)	53.96	24.02

Before analyzing the model, diagnostics will first be completed to screen for potential leverage points, as well as address issues of variance stabilization and linearity. For identifying influential observations, each point will have its cook's distance, DFFITS, and DFBETAS calculated. For addressing non-linearity, a box-cox power transformation will be applied to each variable to provide the best fit. While this will impact the interpretability of the model, this report is more interested in assessing the extent and significance of each variable, rather than the specific least square estimates between each variable and car accident deaths. Therefore, the transformations are applied to give each independent variable the best chance in showcasing its predictive power and impact on the dependent variable. As a final diagnostic, an analysis of collinearity will be conducted and VIF statistic calculated.

There are multiple valid approaches to answering this problem. The first is working backwards: create a full multivariate model including all variables, then analyze each variable's relevance. A model selection algorithm can be used here to screen out variables not relevant in predicting car accident deaths. The second approach is more bottom up: create two separate multivariate models, one accounting for personal factors, and another accounting for macro contributors, and compare the models through multiple metrics to assess which one better explains the variance in car accident deaths. This analysis will take both approaches to provide the fullest picture.

For the full model approach, alongside a standard least square estimate for the model's coefficients, we will use both stepwise and LASSO shrinkage regression to assess which variables are more relevant towards predicting car accident deaths. We will then cross-validate these selected models to assess their predictive accuracy. These two approaches should paint a strong picture of which variables from either category are relevant in predicting traffic deaths.For the second approach, an ANOVA table will be created for both models. Weight will be placed on the adjusted Rˆ2 values for each model, alongside BIC and AICc calculations to fully assess the significance of each model. We can conclude that the model with the higher adjusted Rˆ2 can better account for the variance of car accident deaths.

Results

Of the influential points noted, there were two countries worth addressing. The first is Luxembourg, which was found to be an aggressive outlier due to its extremely high GDP per capita. Luxembourg possesses an extremely low population coupled with a high proportion of absurdly wealthy citizens, heavily skewing this statistic. Being right in the center of Europe and thus a high travel route for people passing through, it seems plausible that its relatively high traffic death rate is due from accidents involving citizens from different countries. Due to its odd characteristics, it was decided its inclusion only hurts our model with little predictive benefit and was thus removed. Doing so removes multiple very poor leverage points (see figure 1). Poland was also removed due to odd entries in road infrastructure, suggesting an extraordinary circumstance that should not be generalized.

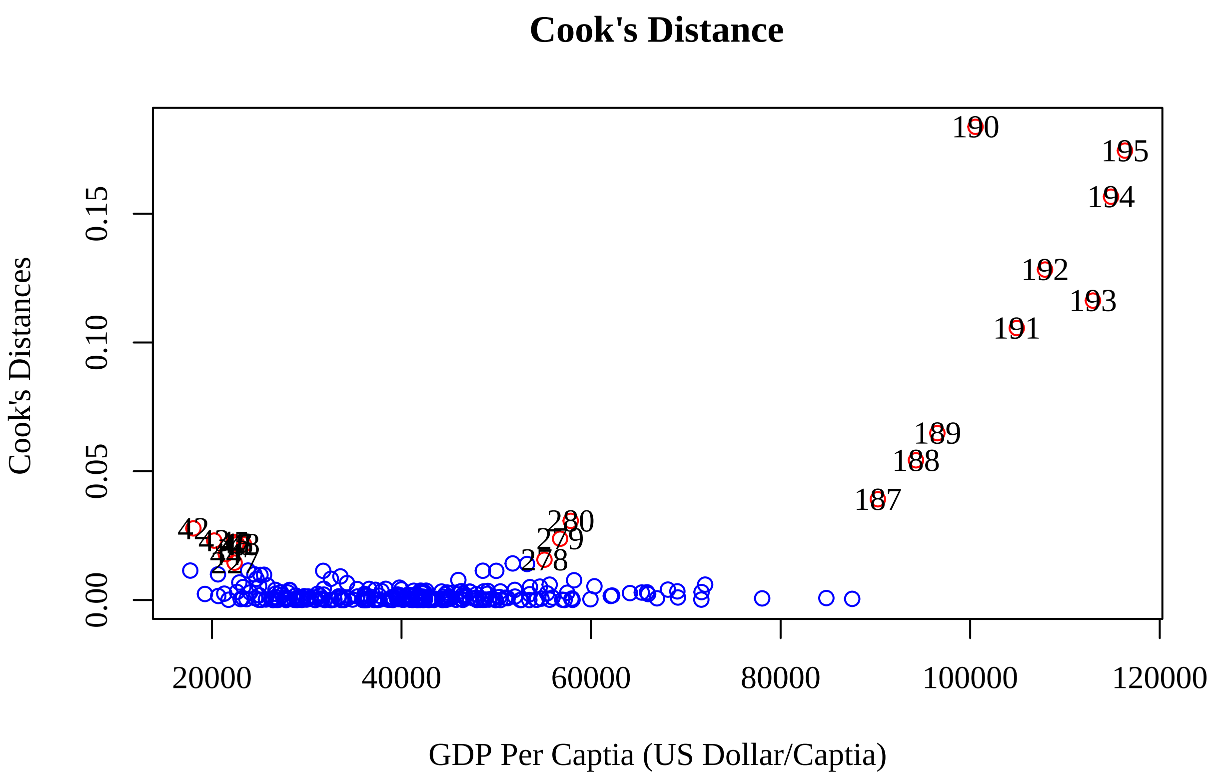

Figure 1: Cooks Distance Example: Points 187-195 all come from Luxembourg and were removed.

The results of the full model approach are shown in table 2. After completing a power transformation, both variance stabilization and normality of data was achieved (see appendix A1). We can note that the only variables that can reject the null hypothesis with significant p values are GDP per capita and healthcare as a percentage of GDP. The MLE show that for every 1% increase in traffic accident rate, there is a 0.87% decrease in GDP per capita. No factor related to the personal lifestyle trends of the country were found to have a significant relationship with car deaths. The LASSO shrinkage singled out GDP per capita as the primary variable driving prediction. The stepwise model using AIC selected healthcare and road maintenance alongside GDP per capita as relevant variables, making macro variables the only relevant predictors chosen by any approach.

Table 2: T-Tests for Significance of Least Square Estimates

Variables	P-statistics	Can We Reject H_0?
GDP (US $ per captia)	0.00	Yes
Healthcare (% of GDP)	0.02	Yes
Road Maintenance (Euro € per captia)	0.10	No
Mobile Broadband Subscriptions (per 100 inhabitants)	0.35	No
Alcohol Consumption (L per captia)	0.26	No

The results of the ANOVA tests from both the personal and macro model are shown in table 3. The only variable with a significant adjusted R^2 value is GDP per capita. Since both methods of model selection only chose macro variables, only the macro models were cross validated. The AIC model performed slightly better compared to the LASSO model that only assessed GDP per capita (see appendix A2 & A3). These results suggest that macro factors serve as much better predictors for car accident deaths in a country, and moreover, GDP per capita is the primary variable with a significant correlated relationship with car accident deaths. Healthcare and Road infrastructure have only a marginal predictive impact on car accident deaths.

Table 3: Comparison of Models

Models	R Squared Value Adjusted	AICc	BIC	Mean Absolute Error from Cross-Validation
Model With Macroeconomic Variables	0.41	-579	-557	0.014
Model with Personal Lifestyle Variables	0.13	-481	-462	NA
Model with GDP per captia Only	0.39	-574	-560	0.020

Discussion

The following analysis sought to decode relevant variables in understanding and possibly predicting the chance of an individual getting into a fatal car accident. Personal lifestyle factors, such as a country's relationship with alcohol or reliance on phones, did not seem to have any relevant impact on a country's rate of car deaths. Furthermore, GDP was consistently shown to be the most relevant predictor of car accidents, as previous literature has suggested. However, deviating from literature, this analysis found a surprising lack of evidence that healthcare funding and quality of road care meaningfully relate to car accident deaths. It should be noted that while there was a correlation between healthcare per capita and car accident death, this statistic has a high collinearity with GDP per capita. Controlling for this by making healthcare a percentage of GDP eliminates its relationship with traffic deaths. These results imply that there may be more going on that contributes to GDP's relationship with traffic deaths than what meets the eye. Healthcare and infrastructure alone do not seem to explain car accident deaths sufficiently.

There are multiple limitations in this study. The validity of the lifestyle predictors is questionable, which may explain the weak relationship with car accident deaths found. Mobile subscriptions were intended to provide a measurement of how prevalent phone usage (and the resulting distracted driving) was in a country, but OCED data only looks at specific mobile plans and not SMS. Since data was limited to OCED, there were not many indicators related to country wide lifestyle decisions. In terms of the models itself, there was some non-linearity that could not be explained by the model (see figure 2) that may have hurt the model's predictive ability. This problem fits in line with previous literature suggesting GDP is non-linear. Box-cox transformations can still create strong predictions, but only to a certain extent. Future analysis should better account for lifestyle variables, as well as attempt to fit variables such as GDP with non-linear regression models.

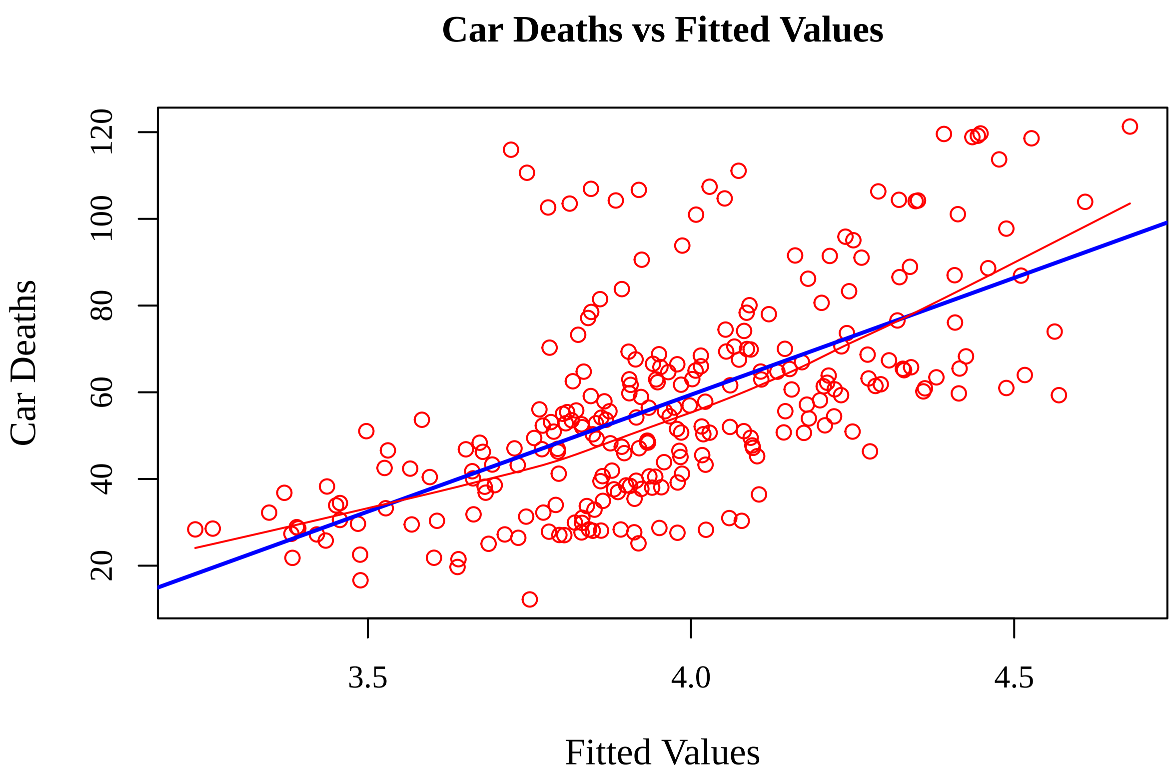

Figure 2: Fitted values of Full Model. Note the slight unexplained non-linearity.

Appendix

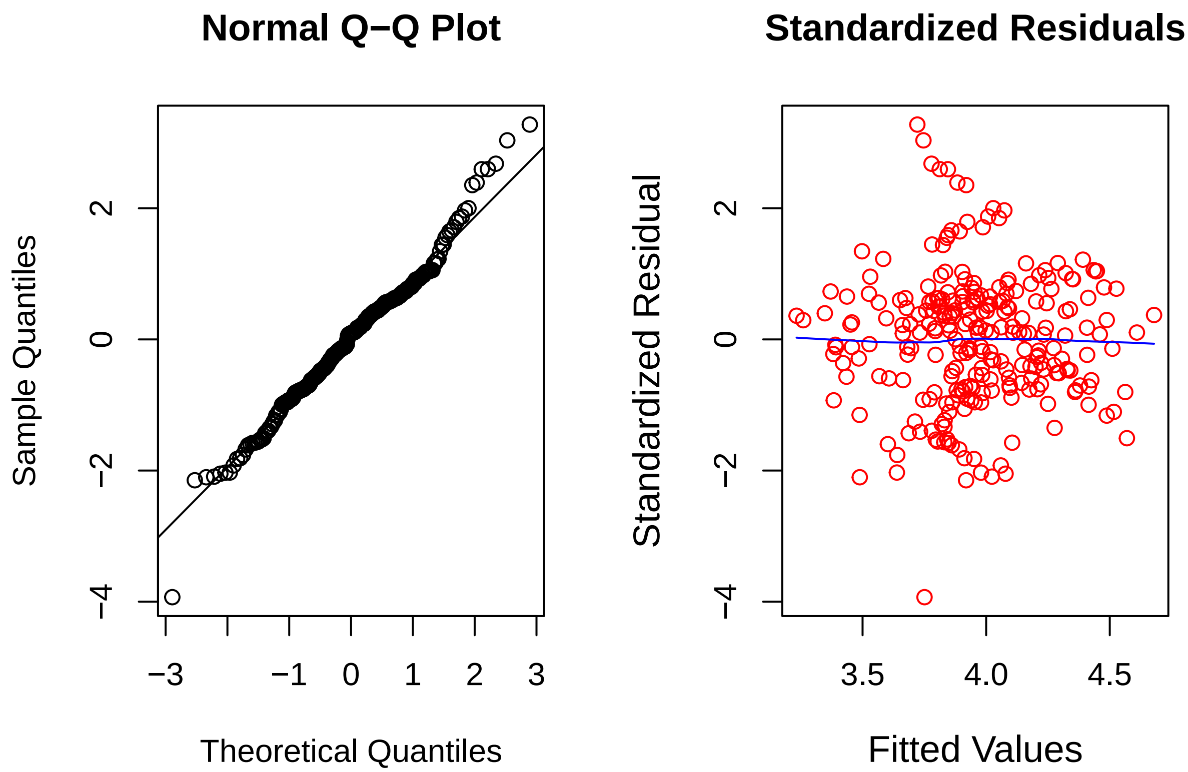

Figure A1: QQ-Plot and Standerdized Variance Graphs Showcasing Constant Variance and Normality of Errors

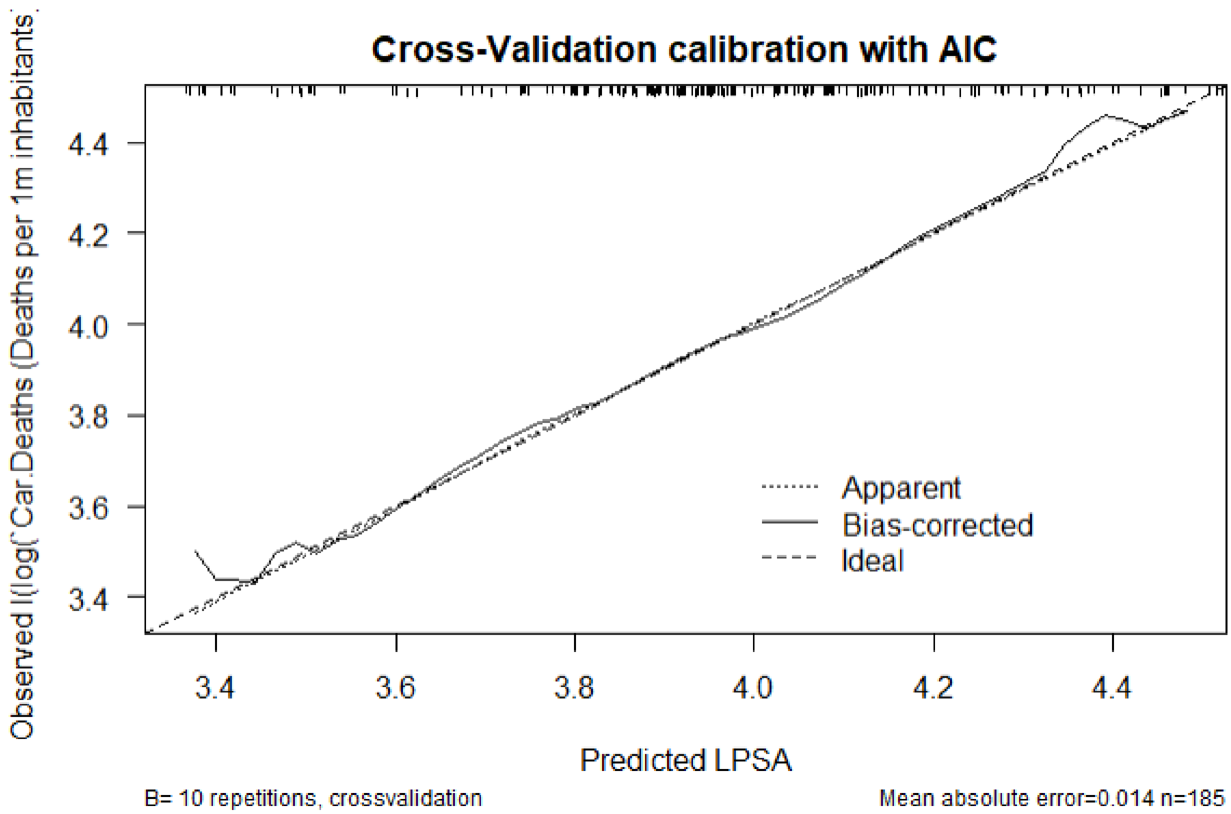

Figure A2: Cross-Validation Calibration using AIC Varible Selection

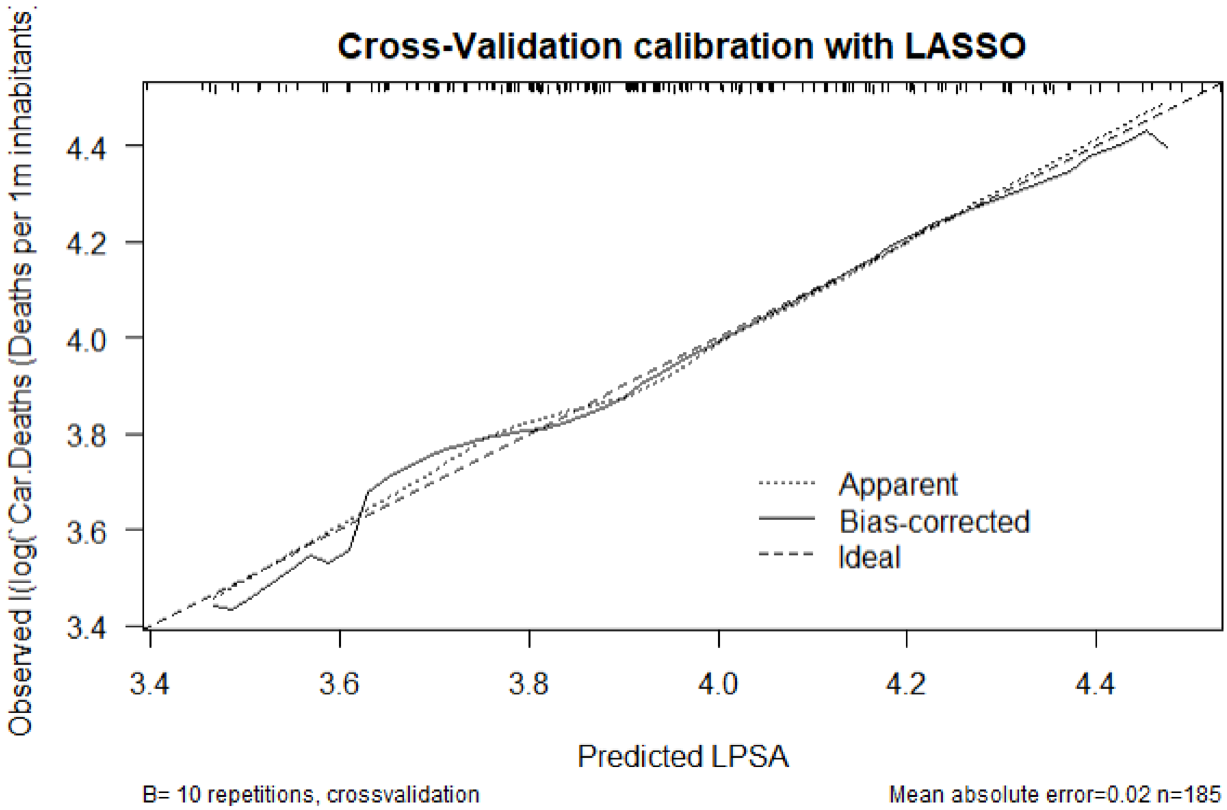

Figure A3: Cross-Validation Calibration using LASSO Varible Selection

References:

Bishai, D., Quresh, A., James, P., & Ghaffar, A. (2005). National Road Casualties and Economic Development. Health Economics, 15(1), 65–81. https://doi.org/10.1002/hec.1020

Bon de Sousa, T., Santos, C., Mateus, C., Areal, A., Trigoso, J., & Nunes, C. (2016). Road traffic accidents and self-reported Portuguese car driver's attitudes, behaviors, and opinions: Are they related? Traffic Injury Prevention, 17(7), 705–711. https://doi.org/10.1080/15389588.2016.1150591

van Beeck, E. F., Borsboom, G. J. J., & Mackenbach, J. P. (2000). Economic development and traffic accident mortality in the industrialized world, 1962–1990. International Journal of Epidemiology, 29(3), 503–509. https://doi.org/10.1093/intjepid/29.3.503

OECD (2022), Road accidents (indicator). doi: 10.1787/2fe1b899-en (Accessed on 20th December 2022)

OECD (2022), Alcohol consumption (indicator). doi: 10.1787/e6895909-en (Accessed on 20th December 2022)

OECD (2022), Health spending (indicator). doi: 10.1787/8643de7e-en (Accessed on 20th December 2022)

OECD (2022), Gross domestic product (GDP) (indicator). doi: 10.1787/dc2f7aec-en (Accessed on 20th December 2022)

OECD (2022), Mobile broadband subscriptions (indicator). doi: 10.1787/1277ddc6-en (Accessed on 20th December 2022)

OECD (2022), Population (indicator). doi: 10.1787/d434f82b-en (Accessed on 20th December 2022)

OECD (2022), Infrastructure maintenance (indicator). doi: 10.1787/c73dc965-en (Accessed on 20th December 2022)