Is Driving Alone for Long Commutes Associated with Poor Mental Health Outcomes?

Abstract

In this study, we model mental health with relevant predictors to investigate whether driving alone for more than 30 minutes is associated with poor mental health outcomes. We use county-level data on various measures of health outcomes to conduct a best subsets analysis to create models with the explanatory variables that best predict mental health. We then conducted an Extra Sum of Squares test to determine whether driving alone for a long time is associated with poor mental health outcomes. Our analysis shows that driving alone for a long commute has a statistically significant association with poor mental health outcomes when other factors are taken into account; however, we cannot conclude that these results have practical significance.

Introduction

Commuting to work is a common aspect of contemporary American society. According to the US census, in 2019, average one-way commute time is 27.6 minutes. Previous research shows that satisfaction with work commute times, for all types of commute, decreases as the time spent commuting to and from work increases (Olsson et al., 2013). Also, life satisfaction scores are lower for longer commute times (Clark et al, 2013). Clark and his team found that psychological strain is greater for those with longer commute times (2013).

Using commute modes that include interaction with other individuals, such as carpooling and public transportation, was found to have a significant negative association with mental health issues, while driving alone to work was found to have a strong and statistically significant positive association with poor mental health (Ferenchek et al., 2014).

However, there are conflicting results for the effect of long commutes on mental health for different income quintiles, family type, gender, and age. (Clark et al., 2013; Feng Z and Boyle, 2014). Given this rather contradictory and complicated existing literature, we wanted to see if there are any associations between the average number of self-reported poor mental health days in a county and the percentage of the working population that drives alone to work for more than 30 minutes in a US county. Our study is distinguished from previous research as we look at aggregate county-level data in the USA. We conducted a best subsets analysis, considering numerous variables (and their interactions) that might influence mental health before considering the influence of driving alone for long commutes. We hypothesize that the proportion of the population that drives alone for a long commute has a positive association with average poor mental health days in a month, due to factors such as increased boredom and decreased leisure time caused by driving alone for a long commute.

Data

Our first dataset, 2020 County Health Rankings Data, was obtained from County Health Rankings & Roadmaps. This dataset compiles data on various health outcome measures by US county, collected from many different sources such as American Community Surveys, 5-year estimates, Behavioral Risk Factor Surveillance System, and Bureau of Labor Statistics, using telephone surveys and through accessing vital registration systems.[1] We combined data on the area of each county from the Missouri Census Data Center with the population data from the 2020 County Health Rankings dataset to calculate the population density of each county.

We identified 21 explanatory variables that previous research has found to be associated with mental health outcomes. We had two variables in our dataset related to driving alone and long commutes: *% Drive Alone to Work*, which is the percentage of the working population that drives alone to work, and *% Long Commute_Drives Alone*, which is the percentage of the population that drives alone to work for more than 30 minutes one way. To find the percentage of the entire working population that drives alone for more than 30 minutes, we multiplied these two variables to create the variable *LongCommuDriveAlone*, which is our key explanatory variable of interest.

[1] *County Health Rankings & Roadmaps* and MCDC Data Applications (see reference list).

Our response variable, *PoorMentalHealthDays*, shows the average number of mentally unhealthy days reported in the past 30 days. We transformed *Unemployed, MedianIncome,* and *PopulationDensity* using the log function to deal with skewed distributions.(Appendix 1). The *Suicides* and *DrugOverdoseDeaths* variables had many missing values, indicating less than 10 suicides or drug overdose deaths per 10000 people. We encoded these variables to be 1 for more than 12 drug overdose deaths per 10000 people and 0 if less than 12. We omitted observations with missing values for the few that remained in our dataset. Ultimately, we were left with 3,112 counties of all 50 U.S. states, with each county being one observation.

Methods

We first started with a reduced model that included all 21explanatory variables in our dataset that possibly had a meaningful association with poor mental health, and excluded the key explanatory variable of interest, *LongCommuDriveAlone*. We then conducted a best subsets analysis, which essentially compares all possible linear models, and displays the combination of predictors that best explains mental health or gives the largest adjusted R^2 values. We chose to create a model that explained the data well but did not include too many explanatory variables (See Appendix 2 for explanatory variables included in the best subsets model).

We then regressed *PoorMentalHealthDays* against all the variables included in our best subsets model along with all their possible interactions. We chose four interactions to include in our reduced subsets model, by selecting the interaction coefficients that were statistically significant at the 1% significance level. Including the interactions in the final subsets model improved our adjusted R^2 slightly (by approximately 0.03).

Our full model is almost identical to our final subsets model, but it includes our key explanatory variable of interest, *LongCommuDriveAlone* and its interactions with all other explanatory variables. Finally, we conducted an Extra Sum of Squares test to compare the final subsets model and the full model, to see if adding *LongCommuDriveAlone* and its relevant interactions significantly improves the overall fit of the model. (See Appendix 2 for visualizations of how *LongCommuDriveAlone* relates to combinations of other explanatory variables.)

Results

Our final reduced model had an adjusted R^2 value of 0.7621, suggesting that it explains approximately 76.2% of the variation in the data. Note that (from Appendix 3) all the variables included in the model except the variable *FoodInsecurity* were significant in the model. Also, residual plot analyses did not raise any concern for the normality assumption (Appendix 4).

The Extra Sum of Squares (ESS) test of the full model versus the final reduced model resulted in a p-value of less than $2.2e^{-16}$, which suggests that there is a statistically significant difference in goodness-of-fit between the two models (Appendix 5). Additionally, the full model including *LongCommuDriveAlone* and its interactions has a slightly better adjusted R^2 value of 0.7735, indicating that the full model describes the variation in the data more than the final reduced model (Appendix 6). If we plot *LongCommuDriveAlone* against *PoorMentalHealthDays*, we see no significant relationship or pattern (Appendix 7). Therefore, the addition of the *LongCommuDriveAlone* variable seems to slightly

improve the model only when we take into account the effect of other variables (See Appendix 8 for visualizations of these interactions). This supports our hypothesis that there is an association between the proportion of a population that drives alone for a long commute and the number of poor mental health days people experience, when we consider other factors. However, because the difference in the adjusted R^2 value between the reduced model and the full model is only 0.0115, we cannot conclude that this association is practically significant.

Discussion

Our findings agree with previous literature but we do not see a practically significant association between *LongCommuDriveAlone* and *PoorMentalHealthDays*. Previous literature proposed that the negative association could be because commutes like driving alone are active, as opposed to more passive commutes like carpooling where one can be kept occupied through other means like watching a movie (Ferenchak et al., 2013). Our analysis supports previous research that portends that the effect of driving alone on mental health is dependent on other factors like sleep habits, income level, food security and perception of their own physical health.[2] This does not suggest that driving alone for a long commute is not associated with poor mental health outcomes, but rather that the strength of this association could be highly dependent on the interactions between the explanatory variables.

Our study is limited in several ways. Firstly, as we used self-reported poor mental health days as our response variable, our results could be unreliable as poor mental health is subjective (Blackwelder, 2021). In addition, because our data is aggregated at the county level and is not at the individual level as in previous studies, it may be the case that our analysis did not pick up on individual-level patterns that other studies have observed. This also means that our results cannot be extended to the individual level.

Previous research suggests that there is a strong positive association between driving alone and serious mental illness (Ferenchak et al., 2013). However, Ferenchak and team found a negative association between carpooling and serious mental illness which suggests that it isn't the drive itself but the driving alone that could be the distressor. Therefore, our research could be expanded to show how different measures of mental health indicators correlate with commuting alone specifically. Notice that the map of mainland USA shows that some areas have a higher proportion of the population experiencing more poor mental health days than others (Appendix 9). This indicates potential for further regional work in the proportion of people who struggle with poor mental health.

Despite potential limitations in this study, our research provides insight into the variables that influence poor mental health, as well as the association between driving alone for a long commute and mental health outcomes at the county level. This is helpful in enhancing our understanding of factors correlated with mental health and the health risks that could arise due to driving alone for more than 30 minutes.

References

[2] Our coefficient of *LongCommuDriveAlone* was significant when we did not include its interaction terms with the other explanatory variables, but became insignificant when we did include them.

Ku, Benson S., Jianheng Li, Cathy Lally, Michael T. Compton, Benjamin G.Druss. 2021. "Associations between mental health shortage areas and county-level suicide ratesamong adults aged 25 and older in the USA, 2010 to 2018." *General Hospital Psychiatry* 70: 44-50. .https://doi.org/10.1016/j.genhosppsych.2021.02.001

Blackwelder A, Hoskins M, Huber L. 2021. "Effect of Inadequate Sleep on Frequent Mental Distress." *Prev Chronic Dis* 18. https://www.cdc.gov/pcd/issues/2021/20_0573.htm

Clark, B., Chatterjee, K., Martin, A. et al. 2020. "How commuting affects subjective wellbeing." *Transportation* 47: 2777–2805. https://doi.org/10.1007/s11116-019-09983-9

"Explore health rankings: Measures & data sources." *County Health Rankings & Roadmaps.* (n.d.). https://www.countyhealthrankings.org/explore-health-rankings/measures-data-sources

Feng Z, Boyle P. 2014. "Do Long Journeys to Work Have Adverse Effects on Mental Health?" *Environment and Behavior*. 46 (5):609-625. doi:10.1177/0013916512472053

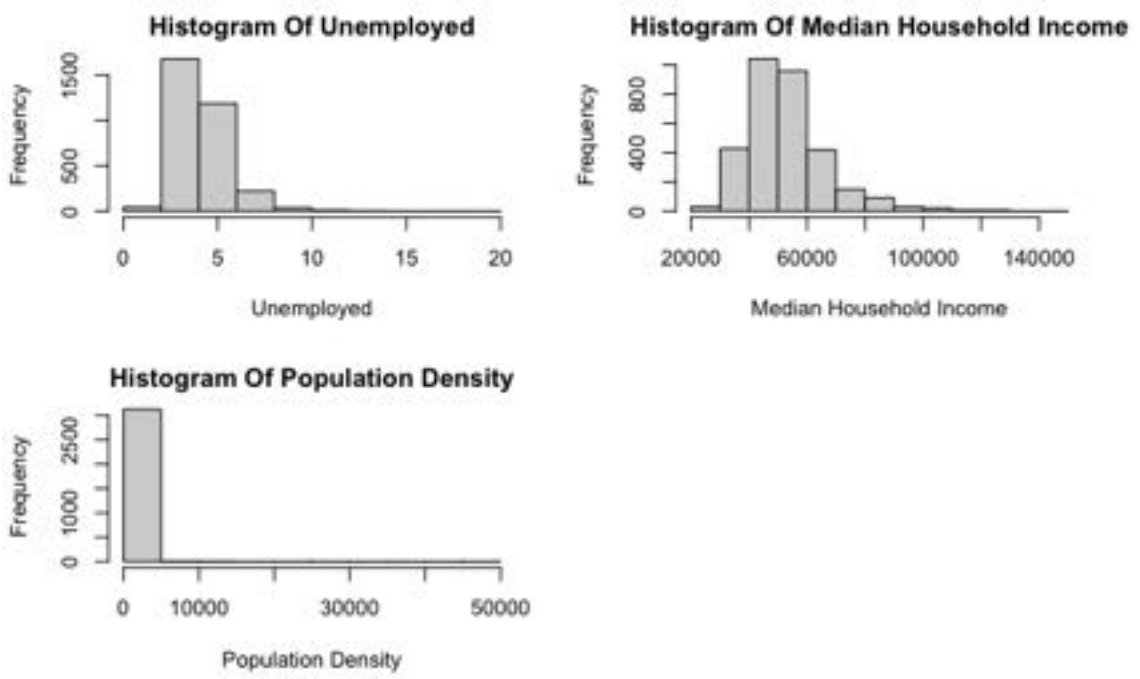
Ferenchak, N. N. and F. N Katirai, M. 2015. "Commute mode and mental health in major metropolitan areas." *The International Journal of Transportation Research* 7: 92-103 https://www.tandfonline.com/doi/full/10.1179/1942787514Y.0000000040?scroll=top&needAccess=true

MCDC Data Applications. Missouri Census Data Center. (n.d.). https://mcdc.missouri.edu/cgi-bin/broker?_PROGRAM=utils.uex2dex.sas&path=%2Fdata%2Fpl942020&dset=uscounties&view=0

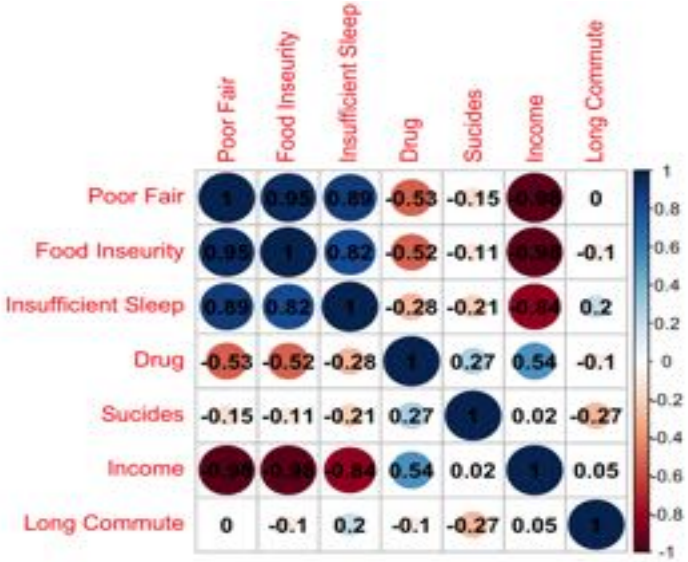
Olsson, L.E., Gärling, T., Ettema, D. et al. 2013 "Happiness and Satisfaction with Work Commute." *Soc Indic Res* 111:255–263. https://doi.org/10.1007/s11205-012-0003-2

Appendix

Appendix 1: Histogram of the distribution of explanatory variables which indicate a skewed distribution (*Unemployed*, *MedianHouseholdIncome, PopulationDensity*).



Appendix 2: Corrplot for Explanatory Variables in Full Model. [3]



[3] Note: ** Since *PoorPhysicalHealthDays* was highly correlated with *PoorMentalHealthDays*, we chose *PoorFairHealth* as an indicator of physical health to avoid overfitting. *PoorFairHealth* shows the percentage of adults who report poor or fair health, when asked to rate their health as Excellent, Very Good, Good, Fair, or Poor.

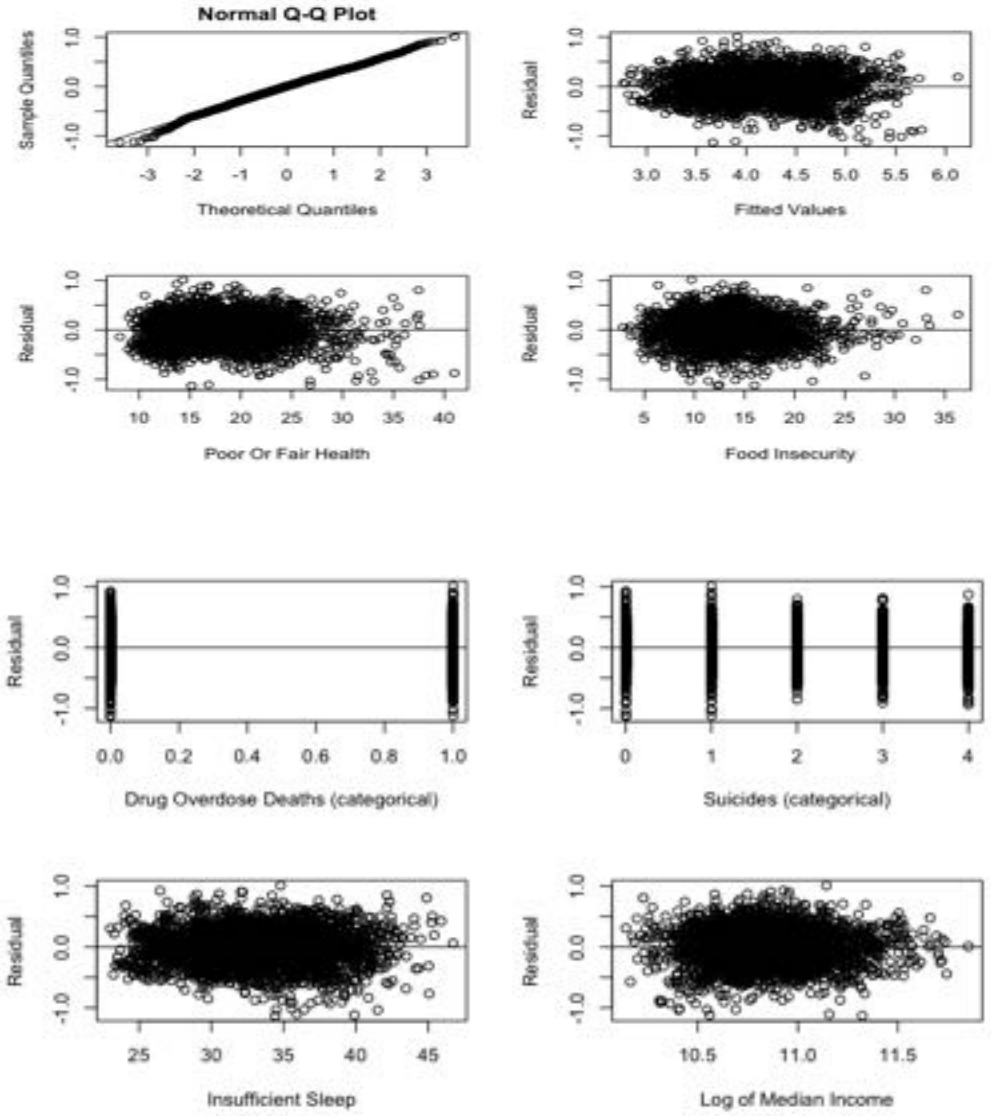
Appendix 3: Final Reduced Model Summary. The model is our chosen best subset model including significant interactions between the explanatory variables.

```
Call:
lm(formula = finalData2$PoorMentalHealthDays ~ finalData2$PoorFairHealth +
    finalData2$FoodInsecurity + finalData2$InsuffSleep + finalData2$DrugOverdoseDeathsCat +
    finalData2$SuicideCat + finalData2$LogMedianIncome + finalData2$PoorFairHealth *
    finalData2$FoodInsecurity + finalData2$FoodInsecurity * finalData2$InsuffSleep +
    finalData2$FoodInsecurity * finalData2$LogMedianIncome +
    finalData2$InsuffSleep * finalData2$LogMedianIncome, data = finalData2)

Residuals:
     Min       1Q   Median       3Q      Max
-1.13116 -0.19067  0.01058  0.20133  1.00847

Coefficients:
                                                  Estimate Std. Error t value Pr(>|t|)
(Intercept)                                     -27.205854   2.840316  -9.578  < 2e-16 ***
finalData2$PoorFairHealth                         0.016162   0.006650   2.430  0.01514 *
finalData2$FoodInsecurity                        -0.106067   0.101962  -1.040  0.29830
finalData2$InsuffSleep                            1.268755   0.089485  14.178  < 2e-16 ***
finalData2$DrugOverdoseDeathsCat1                 0.134296   0.013609   9.868  < 2e-16 ***
finalData2$SuicideCat1                            0.008401   0.019796   0.424  0.67134
finalData2$SuicideCat2                            0.055777   0.019020   2.933  0.00339 **
finalData2$SuicideCat3                            0.102195   0.018822   5.430 6.08e-08 ***
finalData2$SuicideCat4                            0.140510   0.017875   7.860 5.24e-15 ***
finalData2$LogMedianIncome                        2.317512   0.249303   9.296  < 2e-16 ***
finalData2$PoorFairHealth:finalData2$FoodInsecurity   0.002614   0.000484   5.400 7.16e-08 ***
finalData2$FoodInsecurity:finalData2$InsuffSleep     -0.009743   0.000569 -17.123  < 2e-16 ***
finalData2$FoodInsecurity:finalData2$LogMedianIncome  0.036373   0.009088   4.002 6.42e-05 ***
finalData2$InsuffSleep:finalData2$LogMedianIncome    -0.101124   0.007790 -12.981  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2937 on 3098 degrees of freedom
Multiple R-squared:  0.7631,    Adjusted R-squared:  0.7621
F-statistic: 767.7 on 13 and 3098 DF,  p-value: < 2.2e-16
```

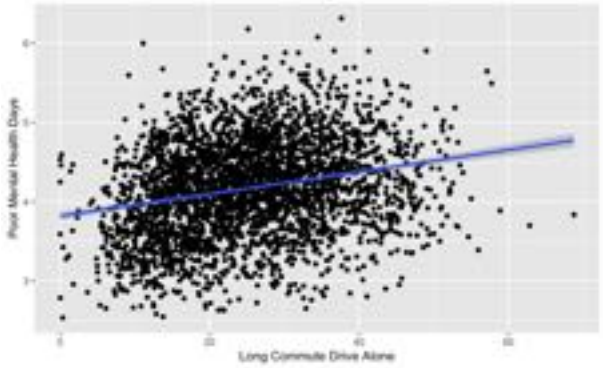
Appendix 4: Results from the Extra Sum Squares test of the Final Reduced Model versus the Full Model. The low p-value suggests a significant difference in the goodness-of-fit between the two models.

```
Analysis of Variance Table

Model 1: finalData2$PoorMentalHealthDays ~ finalData2$PoorFairHealth +
    finalData2$FoodInsecurity + finalData2$InsuffSleep + finalData2$DrugOverdoseDeathsCat +
    finalData2$SuicideCat + finalData2$LogMedianIncome + finalData2$LongCommuDriveAlone +
    finalData2$PoorFairHealth * finalData2$FoodInsecurity + finalData2$FoodInsecurity *
    finalData2$InsuffSleep + finalData2$FoodInsecurity * finalData2$LogMedianIncome +
    finalData2$InsuffSleep * finalData2$LogMedianIncome + finalData2$PoorFairHealth *
    finalData2$LongCommuDriveAlone + finalData2$FoodInsecurity *
    finalData2$LongCommuDriveAlone + finalData2$InsuffSleep *
    finalData2$LongCommuDriveAlone + finalData2$DrugOverdoseDeathsCat *
    finalData2$LongCommuDriveAlone + finalData2$SuicideCat *
    finalData2$LongCommuDriveAlone + finalData2$LogMedianIncome *
    finalData2$LongCommuDriveAlone
Model 2: finalData2$PoorMentalHealthDays ~ finalData2$PoorFairHealth +
    finalData2$FoodInsecurity + finalData2$InsuffSleep + finalData2$DrugOverdoseDeathsCat +
    finalData2$SuicideCat + finalData2$LogMedianIncome + finalData2$PoorFairHealth *
    finalData2$FoodInsecurity + finalData2$FoodInsecurity * finalData2$InsuffSleep +
    finalData2$FoodInsecurity * finalData2$LogMedianIncome +
    finalData2$InsuffSleep * finalData2$LogMedianIncome
  Res.Df    RSS  Df Sum of Sq      F    Pr(>F)
1   3088 253.60
2   3098 267.21 -10   -13.606 16.567 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Appendix 5: Residual plots for the final reduced model. Assumptions of normality and equal variance are met.



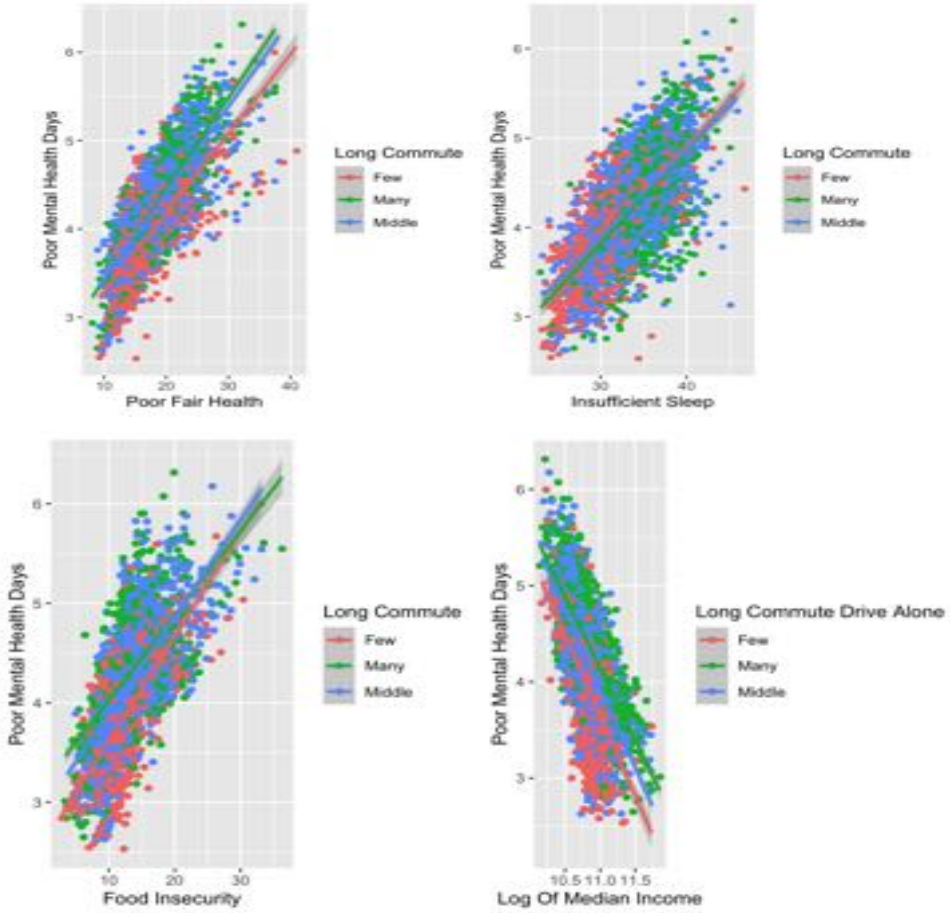
Appendix 6: Plot of Long Commute Drive Alone against Poor Mental Health Days. We can observe no significant relationship or pattern.



Appendix 7: Full Model Summary.

```
Call:
lm(formula = finalData2$PoorMentalHealthDays ~ finalData2$PoorFairHealth +
    finalData2$FoodInsecurity + finalData2$InsuffSleep + finalData2$DrugOverdoseDeathsCat +
    finalData2$SuicideCat + finalData2$LogMedianIncome + finalData2$LongCommuDriveAlone +
    finalData2$PoorFairHealth * finalData2$FoodInsecurity + finalData2$FoodInsecurity *
    finalData2$InsuffSleep + finalData2$FoodInsecurity * finalData2$LogMedianIncome +
    finalData2$InsuffSleep * finalData2$LogMedianIncome + finalData2$PoorFairHealth *
    finalData2$LongCommuDriveAlone + finalData2$FoodInsecurity *
    finalData2$LongCommuDriveAlone + finalData2$InsuffSleep *
    finalData2$LongCommuDriveAlone + finalData2$DrugOverdoseDeathsCat *
    finalData2$LongCommuDriveAlone + finalData2$SuicideCat *
    finalData2$LongCommuDriveAlone + finalData2$LogMedianIncome *
    finalData2$LongCommuDriveAlone, data = finalData2)

Residuals:
     Min       1Q   Median       3Q      Max
-1.19635 -0.18628  0.01113  0.19357  1.03290

Coefficients:
                                                                  Estimate Std. Error t value Pr(>|t|)
(Intercept)                                                     -1.908e+01  2.887e+00  -6.609 4.54e-11 ***
finalData2$PoorFairHealth                                       -1.094e-02  7.978e-03  -1.371 0.170445
finalData2$FoodInsecurity                                       -1.384e-01  1.042e-01  -1.328 0.184268
finalData2$InsuffSleep                                           1.054e+00  9.426e-02  11.186  < 2e-16 ***
finalData2$DrugOverdoseDeathsCat1                                2.033e-01  3.496e-02   5.814 6.73e-09 ***
finalData2$SuicideCat1                                           5.551e-02  4.935e-02   1.125 0.260714
finalData2$SuicideCat2                                           2.396e-02  4.723e-02   0.507 0.612044
finalData2$SuicideCat3                                           1.248e-01  4.671e-02   2.672 0.007587 **
finalData2$SuicideCat4                                           1.351e-01  4.188e-02   3.226 0.001267 **
finalData2$LogMedianIncome                                       1.599e+00  2.540e-01   6.295 3.51e-10 ***
finalData2$LongCommuDriveAlone                                  -1.881e-02  4.165e-02  -0.452 0.651595
finalData2$PoorFairHealth:finalData2$FoodInsecurity              2.050e-03  4.793e-04   4.276 1.96e-05 ***
finalData2$FoodInsecurity:finalData2$InsuffSleep                -8.046e-03  5.849e-04 -13.755  < 2e-16 ***
finalData2$FoodInsecurity:finalData2$LogMedianIncome             3.808e-02  9.313e-03   4.089 4.45e-05 ***
finalData2$InsuffSleep:finalData2$LogMedianIncome               -8.251e-02  8.227e-03 -10.029  < 2e-16 ***
finalData2$PoorFairHealth:finalData2$LongCommuDriveAlone         1.608e-03  2.101e-04   7.650 2.67e-14 ***
finalData2$FoodInsecurity:finalData2$LongCommuDriveAlone        -1.374e-03  2.003e-04  -6.863 8.10e-12 ***
finalData2$InsuffSleep:finalData2$LongCommuDriveAlone           -6.846e-04  1.823e-04  -3.755 0.000176 ***
finalData2$DrugOverdoseDeathsCat1:finalData2$LongCommuDriveAlone -2.792e-03  1.288e-03  -2.168 0.030265 *
finalData2$SuicideCat1:finalData2$LongCommuDriveAlone           -1.847e-03  1.824e-03  -1.012 0.311405
finalData2$SuicideCat2:finalData2$LongCommuDriveAlone            1.080e-03  1.735e-03   0.623 0.533603
finalData2$SuicideCat3:finalData2$LongCommuDriveAlone           -1.052e-03  1.704e-03  -0.617 0.536985
finalData2$SuicideCat4:finalData2$LongCommuDriveAlone           -1.096e-05  1.595e-03  -0.007 0.994519
finalData2$LogMedianIncome:finalData2$LongCommuDriveAlone        3.393e-03  3.629e-03   0.935 0.349898
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2866 on 3088 degrees of freedom
Multiple R-squared:  0.7752,    Adjusted R-squared:  0.7735
F-statistic: 462.9 on 23 and 3088 DF,  p-value: < 2.2e-16
```

Appendix 8: Interaction between Long Commute Drive Alone and other Explanatory Variables



Appendix 9: Distribution of *PoorMentalHealthDays* across Mainland United States.