Twelve Cities: Does lowering speed limits save pedestrian lives?

Jonathan Auerbach and Rob Trangucci

1/14/2017

Summary of Analysis

We investigate whether American cities can expect to achieve a meaningful reduction in pedestrian deaths by lowering the posted speed limit. We present our work in three sections. First we briefly motivate the problem and provide a description of the dataset. Second we fit a log-linear model and compare sources of variation with an analysis of variance. Finally we demonstrate a sample use case. We evaluate the decision to lower many of New York City's posted speed limits from 30 mph to 25 mph. In our evaluation we assume the assignment of speed limits to roads is ignorable given measured covariates, and we calculate the number of lives saved by estimating the causal effect of lowering the speed limit on New York City roads from 30 mph to 25 mph on 25 mph roads. We find some evidence that a lower speed limit does in fact reduce fatality rates, and our estimated causal effect is similar to the traditional before-after analysis espoused by policy analysts. Nevertheless, we conclude that adjusting the posted speed limit in urban environments does not correspond with a reliable reduction in pedestrian fatalities.

We would like to thank both reviewers for their insightful comments and helpful suggestions.

I. Introduction

The Policy Question

Over the last few years, American cities have started to establish aggressive countermeasures that force road users to make safer decisions. These policies are collectively known as Vision Zero and have a stated goal of creating a road system with zero traffic fatalities.

In theory, Vision Zero is a comprehensive road investment strategy that prioritizes safety over vehicle mobility by anticipating human error and then slowing vehicles to the safest possible travel speed. Citywide road redesign and increased levels of enforcement and outreach are integral to a Vision Zero strategy (Tingvall and Haworth 2000). These approaches are effective because they confine drivers, considerably reducing the chance that human error will result in a fatality. Yet such changes are costly to implement and challenging to sustain across an entire city. In practice, policymakers simply mandate that vehicles reduce their travel speed by lowering the posted speed limit below the speed for which the road was originally designed.

Lowering the speed limit is politically expedient since speed limit changes can be implemented immediately and at relatively little cost, and they can be sustained across an entire city at no additional expense. But lowering a speed limit without improvements to road design, enforcement and outreach may do little to reduce fatalities if drivers feel little pressure to comply with the lower limit (Leaf and Preusser 1999). In fact, the National Highway Safety Traffic Administration (NHSTA) rates the countermeasure "reduce and enforce speed limits" three out of five stars for improving safety because research indicates that actual speed is reduced by only a fraction of the change in the posted speed limit (Goodwin et al. 2010).

Twelve major American cities have officially set a Vision Zero strategy in response to vocal advocates: Chicago, San Francisco, New York City, Boston, Los Angeles, Austin, Portland, Seattle, San Jose, San Diego, Washington D.C. and Denver. The majority of these strategies include adjustments to many or all of the posted speed limits. For example, New York City lowered the default citywide speed limit from 30 to 25 mph in 2015. San Francisco uniformly set the speed limit around schools to 15 mph and is considering a default citywide speed limit of 20 mph.

The policy question is twofold. The first is descriptive: which aspects of roads drive pedestrian fatalities? The second is prescriptive: can a speed limit adjustment meaningfully reduce the number of pedestrian deaths? We analyze data from the twelve Vision Zero cities to investigate these questions. To our knowledge this is the first study to systematically examine Vision Zero data across multiple cities.

The Data

Our dataset comprises every pedestrian death in the twelve major American cities with Vision Zero policies between 2010 and 2015 for which there was no missing covariate information. The unit of our analysis, however, is the immediate region or road segment on which each death occurred. We consequently view our dataset as a random sample of regions, and we obtain sampling weights equal to the reciprocal of the probability each region would be included in the sample.

The dataset has the following eight qualitative variables describing each road region: the weather and surface condition (COND), the city (CITY), the year (YEAR), the posted speed limit (SLIM), the presence of various signs or signals (SIGN), the time and lighting (LGHT), the physical road characteristics or built environment (BLTE) and the annual average traffic density (TFFC). These variables can be thought of as batches of categories that describe similar aspects of road regions. The dataset has two quantitative variables: the number of pedestrians exposed to the road region (EXPR) and the estimated number of road regions represented by the observation (WGHT). More information on the dataset can be found in the Appendix.

The first six observations of the dataset are displayed in Table 1 below. Table 2 shows the number of observations within the most common qualitative variable levels. Some of these variables are large, with COND, LGHT and BLTE having 25, 36 and 111 categories each. Throughout this analysis we present the qualitative variable levels in alphabetical order followed by the quantitative variables.

```
kable(round(head(crash_data[c(order(colnames(crash_data)[1:8]),9:10)]),2),
      caption="Example Observations of Dataset")
```

Table 1: Example Observations of Dataset

BLTE	CITY	COND	LGHT	SIGN	SLIM	TFFC	YEAR	EXPR	WGHT
13	3	2	1	1	11	1	1	1300.36	4.69
14	3	2	25	1	6	1	1	196.05	1.90
56	8	2	21	1	8	4	1	398.10	1.77
22	8	2	21	1	8	1	1	497.38	1.72
42	10	2	12	1	9	1	1	458.04	1.68
54	8	24	12	1	8	4	1	317.77	1.60

```
kable(summary(apply(crash_data[,order(colnames(crash_data)[1:8])],2,factor)),
      caption="Summary of Dataset")
```

Table 2: Summary of Dataset

BLTE	CITY	COND	LGHT	SIGN	SLIM	TFFC	YEAR
16 : 358	12 :866	2 :1976	8 :386	1 :1578	7 :898	1:546	1:409
56 : 259	8 :564	23 : 232	21 :357	4 : 476	8 :580	2:830	2:398
14 : 243	6 :255	9 : 198	29 :227	3 : 340	6 :397	3:640	3:443
54 : 192	9 :184	24 : 56	1 :223	6 : 78	10 :170	4:540	4:455
41 : 77	3 :119	3 : 24	34 :221	9 : 24	9 :136	NA	5:406
79 : 63	10 :106	18 : 16	25 :161	2 : 19	14 :135	NA	6:445
(Other):1364	(Other):462	(Other): 54	(Other):981	(Other): 41	(Other):240	NA	NA

Figure 1.1 plots a histogram of the observations by city. Figure 1.2 plots these cities by the most common speed limits, and Figure 1.3 plots them by both the most common speed limit and the last four years of the dataset. Cities are arranged from smallest population at the top to largest population at the bottom (as of the 2010 Census).

```
city_labels <- factor(crash_data$CITY,labels=c(
    "Washington D.C.","Boston","Austin","Denver","Portland","Chicago",
    "Seattle","Los Angeles","San Diego","San Francisco","San Jose",
    "New York City"))
city_labels <- factor(city_labels,
    levels=levels(city_labels)[c(12,8,6,9,11,3,10,7,4,1,2,5)])
crash_viz <- crash_data; crash_viz$city_labels <- city_labels

ggplot(crash_viz) +
  theme(plot.title = element_text(size=9, hjust = .5),
        plot.caption = element_text(hjust = 1)) +
  aes(city_labels) +
  geom_histogram(stat="count") +
  coord_flip() +
  scale_y_continuous(expand = c(0, 0)) +
  labs(x="",y="",
   title="Figure 1.1: Number of Observations per City",
   caption = paste("This figure exhibits the number of deaths in each Vision Zero city",
            "over the 2010 to 2015 sample period. Cities are arranged in increasing",
            "order \nof population. The number of fatalities corresponds loosely",
            "with population size."))
```

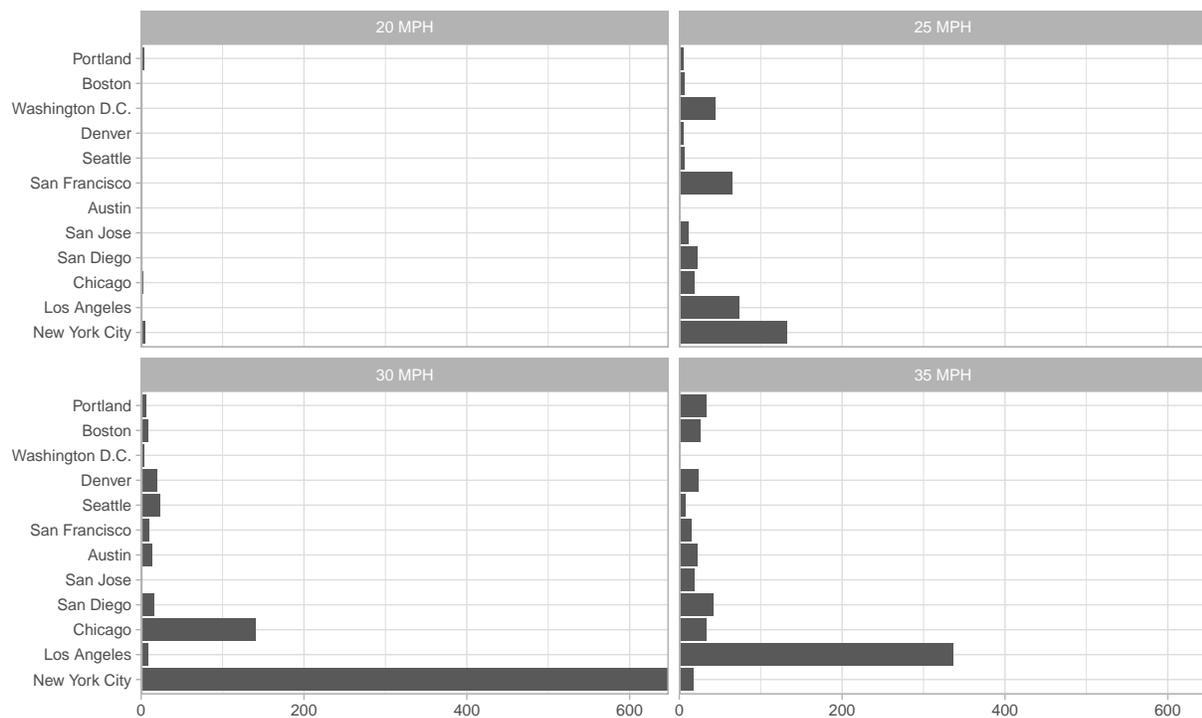Figure 1.1: Number of Observations per City
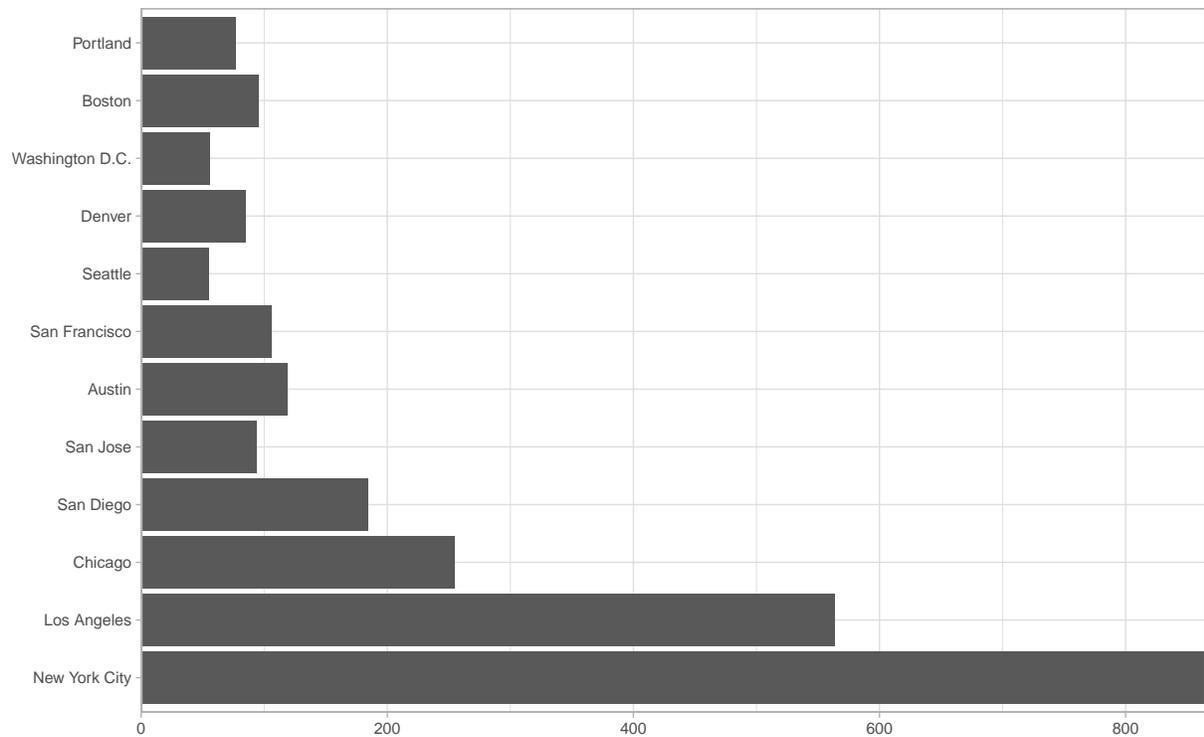
This figure exhibits the number of deaths in each Vision Zero city over the 2010 to 2015 sample period. Cities are arranged in increasing order of population. The number of fatalities corresponds loosely with population size.

3

```
ggplot(crash_viz[crash_viz$SLIM %in% 5:8,]) +
  theme(plot.title = element_text(size=9, hjust = .5),
        plot.caption = element_text(hjust = 1)) +
  aes(city_labels) +
  geom_histogram(stat="count") +
  coord_flip() +
  labs(x="",y="") +
  facet_wrap(~factor(SLIM,labels = c("20 MPH","25 MPH","30 MPH","35 MPH"))) +
  scale_y_continuous(expand = c(0, 0)) +
  labs(x="",y="",
    title="Figure 1.2: Number of Observations per City by Posted Speed Limit",
    caption = paste("This figure exhibits the number of deaths in each Vision Zero city",
              "over the 2010 to 2015 sample period stratified by whether the posted",
              "speed \nlimit on the road was 20, 25, 30 or 35 MPH. Cities are",
              "arranged in increasing order of population. Speed limits of 25, 30 and",
              "35 are observed \nfor every major city."))
```

Figure 1.2: Number of Observations per City by Posted Speed Limit

This figure exhibits the number of deaths in each Vision Zero city over the 2010 to 2015 sample period stratified by whether the posted speed limit on the road was 20, 25, 30 or 35 MPH. Cities are arranged in increasing order of population. Speed limits of 25, 30 and 35 are observed for every major city.

```
ggplot(crash_viz[crash_viz$SLIM %in% 6:8 &
                 crash_viz$YEAR %in% 3:6,]) +
  theme(plot.title = element_text(size=9, hjust = .5),
        plot.caption = element_text(hjust = 1)) +
  aes(city_labels) +
  geom_histogram(stat="count") +
  coord_flip() +
  labs(x="",y="") +
  facet_grid(factor(YEAR,labels = c("2012","2013","2014","2015"))~
```

```
              factor(SLIM,labels = c("25 MPH","30 MPH","35 MPH"))) +
scale_y_continuous(expand = c(0, 0)) +
labs(x="",y="",
  title="Figure 1.3: Number of Observations per City by Posted Speed Limit and Year",
  caption = paste("This figure exhibits the number of deaths in each Vision Zero city",
             "over the 2012 to 2015 sample period stratified by the year each",
             "fatality took \nplace and whether the posted speed limit on the road",
             "was 25, 30 or 35 MPH. Cities are arranged in increasing order of",
             "population. Speed \nlimits of 25, 30 and 35 are not observed for every",
             "major city in every year."))
```

Figure 1.3: Number of Observations per City by Posted Speed Limit and Year
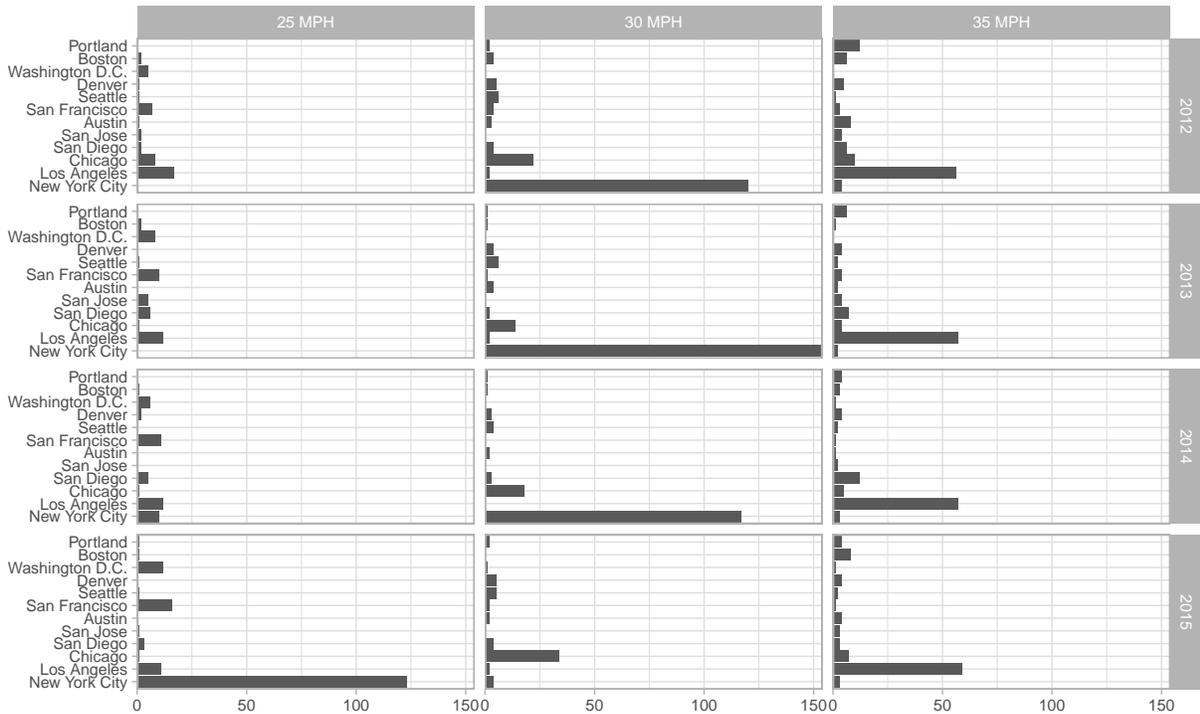
This figure exhibits the number of deaths in each Vision Zero city over the 2012 to 2015 sample period stratified by the year each fatality took place and whether the posted speed limit on the road was 25, 30 or 35 MPH. Cities are arranged in increasing order of population. Speed limits of 25, 30 and 35 are not observed for every major city in every year.

The 2015 reduction in New York City's default speed limit from 30 to 25 mph is apparent in Figure 1.3. In this analysis, we construct a model for the years 2010-2014 and estimate the causal effect of this 2015 policy. Reserving the observations made during the year 2015 for testing our model serves two purposes. First, an independent testing set establishes stronger evidence that the model is predictive of the potential number of fatalities that would occur as a result of either speed limit policy. Second, the training set reflects available information before the policy was adopted, and consequently an analysis similar to ours could have been performed before the policy was implemented. Coincidentally, our results are comparable to a naïve version of the "before-after" analysis favored by policymakers.

II. Model and Inference

Policymakers consider a wide variety of road characteristics when setting the posted speed limit. As a result, the relationship between pedestrian deaths and the speed limit is potentially confounded by any of the large number of qualitative variable categories in our dataset. We summarize the pairwise association of the eight qualitative variables using Cramer's V statistic in Figure 2. We find that no two covariates are completely

associated although nearly every covariate exhibits moderate association. This suggests we should account for all eight qualitative variables in our initial model, and possibly some interactions.

```r
cramer.v <- function(X1,X2)
  sqrt(chisq.test(X1, X2)$statistic / ((length(X1) *
      (min(length(unique(X1)),length(unique(X2))) - 1))))
cor_mat <- matrix(ncol = 8, nrow = 8,
                  dimnames = list(colnames(crash_data)[1:8],
                                  colnames(crash_data)[1:8]))
for(i in 1:8) for(j in 1:8) cor_mat[i,j] <- cramer.v(crash_data[,i],crash_data[,j])

ggcorr(data=NULL,cor_matrix = cor_mat,
       low = "steelblue", mid = "white", high = "darkred",
       label = TRUE, label_size = 3, size = 3) + theme_void() +
  theme(
        plot.title = element_text(size=9, hjust = 0),
        plot.caption = element_text(size = 7, hjust = 1),
        legend.position = "none") +
  labs(
   title="Figure 2: Correlation of Qualitative Variables using Cramer's V",
   caption = paste("This figure exhibits the marginal correlation",
              "of the eight qualitative variables using Cramer's V statistic.",
              "\nA value of 0 indicates that the two variables have little",
              "association, and a value of 1 indicates that the \ntwo variables",
              "are equal to each other. From this figure we observe that all of the",
              "variables are \nmoderately associated. The largest association exists",
              "between the posted speed limit, \ntraffic density and the built",
              "environment."))
```

Figure 2: Correlation of Qualitative Variables using Cramer's V

	COND	CITY	YEAR	SLIM	SIGN	LGHT	BLTE	TFFC
TFFC								
BLTE								0.4
LGHT							0.2	0.2
SIGN						0.2	0.3	0.2
SLIM					0.1	0.2	0.4	0.3
YEAR				0.2	0.1	0.1	0.2	0.1
CITY			0.1	0.3	0.2	0.2	0.3	0.3
COND		0.1	0.1	0.2	0.1	0.1	0.3	0.1

This figure exhibits the marginal correlation of the eight qualitative variables using Cramer's V statistic. A value of 0 indicates that the two variables have little association, and a value of 1 indicates that the two variables are equal to each other. From this figure we observe that all of the variables are moderately associated. The largest association exists between the posted speed limit, traffic density and the built environment.

With all eight qualitative variables, the data constitute a high dimensional contingency table where each cell is a low, non-zero count. It is conventional to employ a log-linear model for such data. Causal estimates can then be calculated by imputing the values of missing cells representing unobserved counterfactual outcomes and then comparing those outcomes that address the question of interest. However, the validity of these estimates depends on a variety of assumptions, and we briefly highlight three major assumptions. See Gelman and Hill (2006) or Imbens and Rubin (2015) for a general discussion.

Assumptions

We assume the assignment of posted speed limits to road regions is strongly ignorable given the observed covariates. This implies that road regions with similar covariate values (in a particular city, near a school, up a hill, etc.) are equally likely to be designated a certain speed limit. In particular, among two such road regions, the one with fewer pedestrian deaths cannot be more or less likely to post any speed limit.

This assumption would be violated if speed limit policy were based on an omitted variable that also influenced the fatality rate. For example, under Oregon state law, Portland can reduce speed limits to 20 mph provided additional signage indicating the presence of pedestrians is displayed (Essex and Teigen 2016). Failing to consider the presence of signs and signals would distort the estimated relationship of interest between the posted speed limit and pedestrian fatalities by inadvertently incorporating the relationship between signs and fatalities, which is not of explicit interest. Consultation with experts of speed limit policy informed our choice of covariates, and we believe we have accounted for all important information used by policymakers to set speed limits.

We assume no interference between the units of analysis, the road regions. This assumption would be violated if, for example, the fatality rate in any road region depended on the posted speed limit of a neighboring road region. While such interferences are possible, we anticipate meaningful interactions between urban roads to

be rare since the factors leading to a collision are generally contained within a single road region.

We rely on the following zero-truncated, log-linear model to predict the counterfactual outcomes. Let y_{ij} denote the ith death in the jth covariate strata. The joint probability distribution can be factorized:

$$
\begin{aligned}
\epsilon &\sim \text{Normal}(0, \sigma_\epsilon) \\
\alpha_{i} &\sim \text{Normal}(0, \sigma_i) \\
\bar{y}_{.j} &\sim \text{Poisson}^+(\exp(\mu + \alpha_1^{\text{SLIM}} + \alpha_2^{\text{CITY}} + \alpha_3^{\text{YEAR}} + \alpha_4^{\text{COND}} + \alpha_5^{\text{SIGN}} \\
&\qquad\qquad + \alpha_6^{\text{LGHT}} + \alpha_7^{\text{BLTE}} + \alpha_8^{\text{TFFC}} + \epsilon_j + \beta \cdot \log(\text{EXPR}_{.j})))
\end{aligned}
$$

To complete the specification of our first model (Model 1), we also put diffuse normal priors on μ, β and the σ. In the Comparison section we consider a second model (Model 2) where we add interactions between some of the explanatory variables.

Computation

We sample from the posterior distribution of the parameters in the preceding model, excluding year 2015 from the dataset. Specifically, we use RStan (Stan Development Team 2016) to run four chains for 2,000 iterations each. This yields 4,000 posterior samples after discarding the first 1,000 iterations of each chain. In the Stan generated quantities block, we calculate the finite sample standard deviations of each qualitative variable. We also estimate the expected potential outcome for each road region in the dataset in the year 2015, first supposing the region had a 25 mph posted speed limit and then supposing it had a 30 mph posted speed limit. We use these estimates in the Criticism section to derive the posterior predictive distribution for model evaluation and in the Policy Analysis section to estimate the causal effect of New York City's default speed limit reduction. The corresponding Stan code can be found in the Appendix.

```r
G <- 8
J <- apply(crash_data[,1:8],2,function(x) length(unique(x)))
stan_dat <- aggregate(cbind(1,EXPR)~.,crash_data,sum)
stan_dat <- stan_dat[order(stan_dat$YEAR),]
stan_dat_list <- with(stan_dat,
                    list(N_train = which.max(stan_dat$YEAR)-1,
                        N = nrow(stan_dat),
                        J = J,
                        G = G,
                        COND = COND,
                        CITY = CITY,
                        YEAR = YEAR,
                        SLIM = SLIM,
                        SIGN = SIGN,
                        LGHT = LGHT,
                        BLTE = BLTE,
                        TFFC = TFFC,
                        count = V1,
                        EXPR = EXPR))
model <- stan_model(file = 'model.stan')
fit <- sampling(model, data = stan_dat_list, iter = 2000,
                chains = 4, refresh = 10, cores = 4)
```

Analysis of Variance

Figures 3 and 4 display the Analysis of Variance using the posterior draws from the first model (Model 1) outlined above. Figure 3 shows the inner 50 and 95 percent quantiles of the finite sample standard deviation. The fat box represents the inner 50 percent range of the parameter, and the thin line represents the inner 95 percent range. We loosely interpret these regions as corresponding to the likely and possible values of the parameter respectively. Figure 3 demonstrates that time of day and lighting (LGHT) explain a significant portion of the pedestrian death rate. This makes sense as this variable captures the varying use of roads each day (i.e. commuting, tourism, dinner, the bar scene, etc.). City (CITY) and the built environment (BLTE) also explain much variation in pedestrian deaths. Surprisingly, there is little variation of deaths within cells (cell), years and levels of traffic (TFFC) suggesting that, through time of day and lighting, city and the built environment, we have adjusted for major sources of confounding.

```r
btwn <- c("COND_sd","CITY_sd","YEAR_sd",
          "SLIM_sd","SIGN_sd","LGHT_sd",
          "BLTE_sd","TFFC_sd","cell_sd")
coefs <- data.frame(extract(fit,pars=btwn))
coef_ggplot <- data.frame(coef_mean=apply(coefs,2,mean),btwn = btwn)
coef_ggplot$upper50 <- apply(coefs,2,quantile,probs=.75)
coef_ggplot$lower50 <- apply(coefs,2,quantile,probs=.25)
coef_ggplot$upper95 <- apply(coefs,2,quantile,probs=.975)
coef_ggplot$lower95 <- apply(coefs,2,quantile,probs=.025)


ggplot(coef_ggplot, aes(substr(btwn,1,4), coef_mean)) +
  theme(plot.title = element_text(size=9, hjust = .5),
        plot.caption = element_text(hjust = 1)) +
  geom_linerange(aes(ymin = lower50, ymax = upper50),size=2) +
  geom_linerange(aes(ymin = lower95, ymax = upper95)) +
  coord_flip() +
  labs(y=expression(hat(sigma)(.)),x="",
    title="Figure 3: Model 1, Analysis of Variance, Between",
    caption = paste("This figure exhibits inner 50 and 95 percent intervals for the",
                    "finite-population standard deviations of each batch of",
                    "explanatory variables. Time of day \nand lighting, city and the",
                    "built environment explain a substantial amount of variation in",
                    "the fatality rate. There is relatively more uncertainty around",
                    "\nthe importance of road surface condition and signs/signals.",
                    "Year, traffic density and speed limit account for relatively",
                    "little variation")) +
  scale_y_continuous(limits=c(0,1.1*max(coef_ggplot$upper95)),
                      expand = c(0, 0))
```

Figure 3: Model 1, Analysis of Variance, Between
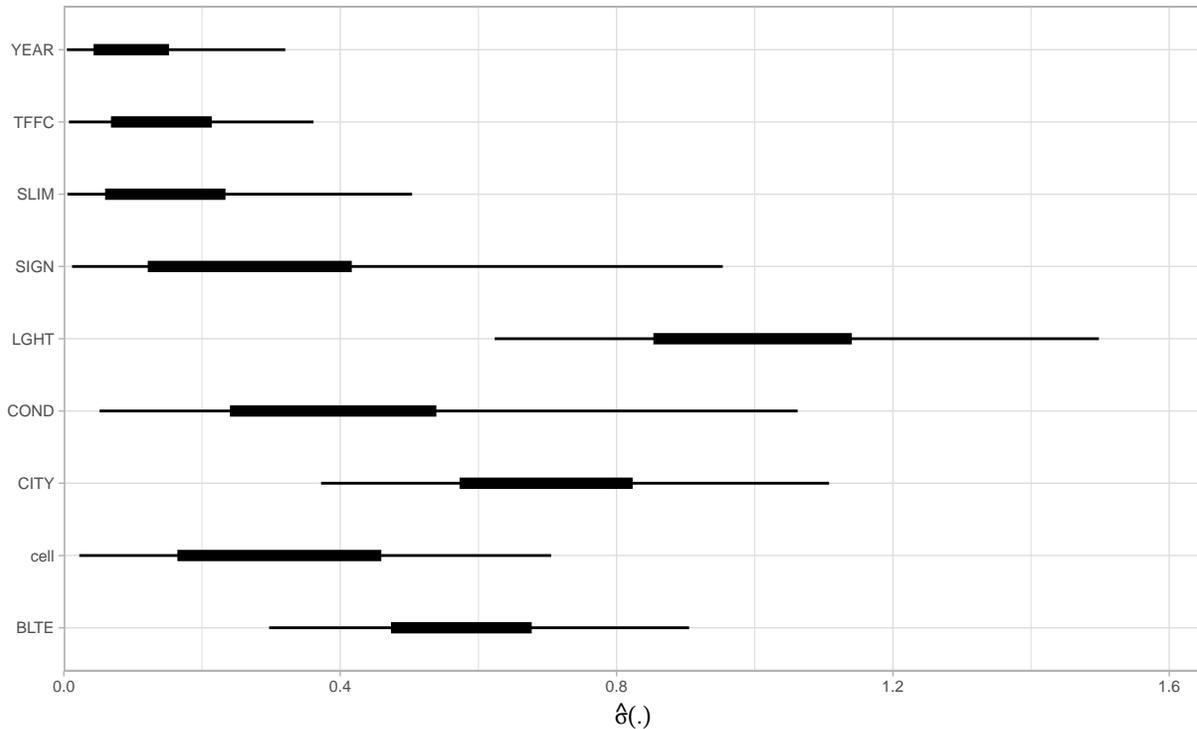
This figure exhibits inner 50 and 95 percent intervals for the finite–population standard deviations of each batch of explanatory variables. Time of day and lighting, city and the built environment explain a substantial amount of variation in the fatality rate. There is relatively more uncertainty around the importance of road surface condition and signs/signals. Year, traffic density and speed limit account for relatively little variation

The following four figures display the individual drivers of pedestrian deaths. Figures 4.1 and 4.2 show the 50 and 95% quantiles of the mean effects. Figures 4.3 and 4.4 show just the city effects and speed limit effects respectively. Figure 4.4 shows that pedestrians are, in general, at similar risk of death at 30 as compared with 25 MPH roads. Of course, this is not a comparison of the speed limit effects within the same stratum. We make this comparison in the Policy Analysis section.

```r
wthn <- c("COND_e","CITY_e","YEAR_e","SLIM_e",
          "SIGN_e","LGHT_e","BLTE_e","TFFC_e")
coefs <- extract(fit,pars=wthn)
coef_ggplot <- data.frame(wthn = character(),
                          numb = character(),
                          coef_mean=numeric(),
                          upper50 = numeric(),
                          lower50 = numeric(),
                          upper95 = numeric(),
                          lower95 = numeric())
for(var in seq_along(coefs)){
  coefs_temp <- coefs[[var]]
  coef_ggplot_temp <- data.frame(wthn = wthn[var],
                                 numb = paste0(wthn[var],1:ncol(coefs_temp)),
                                 coef_mean=apply(coefs_temp,2,mean))
  coef_ggplot_temp$upper50 <- apply(coefs_temp,2,quantile,probs=.75)
  coef_ggplot_temp$lower50 <- apply(coefs_temp,2,quantile,probs=.25)
  coef_ggplot_temp$upper95 <- apply(coefs_temp,2,quantile,probs=.975)
  coef_ggplot_temp$lower95 <- apply(coefs_temp,2,quantile,probs=.025)
  coef_ggplot <- rbind(coef_ggplot,coef_ggplot_temp)
}
```

```
ggplot(coef_ggplot[coef_ggplot$wthn != "BLTE_e",]) +
  theme(plot.title = element_text(size=9, hjust = .5),
        plot.caption = element_text(hjust = 1)) +
  aes(numb, coef_mean,color=substr(wthn,1,4)) +
  geom_linerange(aes(ymin = lower50, ymax = upper50),size=2) +
  geom_linerange(aes(ymin = lower95, ymax = upper95)) +
  coord_flip() +
  scale_x_discrete(breaks = NULL) +
  theme(legend.position = c(.85, 0.3)) +
  scale_color_discrete(guide = guide_legend(reverse = TRUE)) +
  labs(title="Figure 4.1: Model 1, Analysis of Variance, Within",
       y=expression(hat(mu)(.)),x="",
       color = "",
       caption = paste("This figure exhibits inner 50 and 95 percent intervals for the",
                       "effects from seven batches of explanatory variables. Effects",
                       "are colored by batch and \nare interpreted as the log of the",
                       "expected multiplicative increase in the fatality rate compared",
                       "to the average road region. Time of day and \nlighting and",
                       "city explain a substantial amount of variation in the fatality",
                       "rate."))
```

Figure 4.1: Model 1, Analysis of Variance, Within
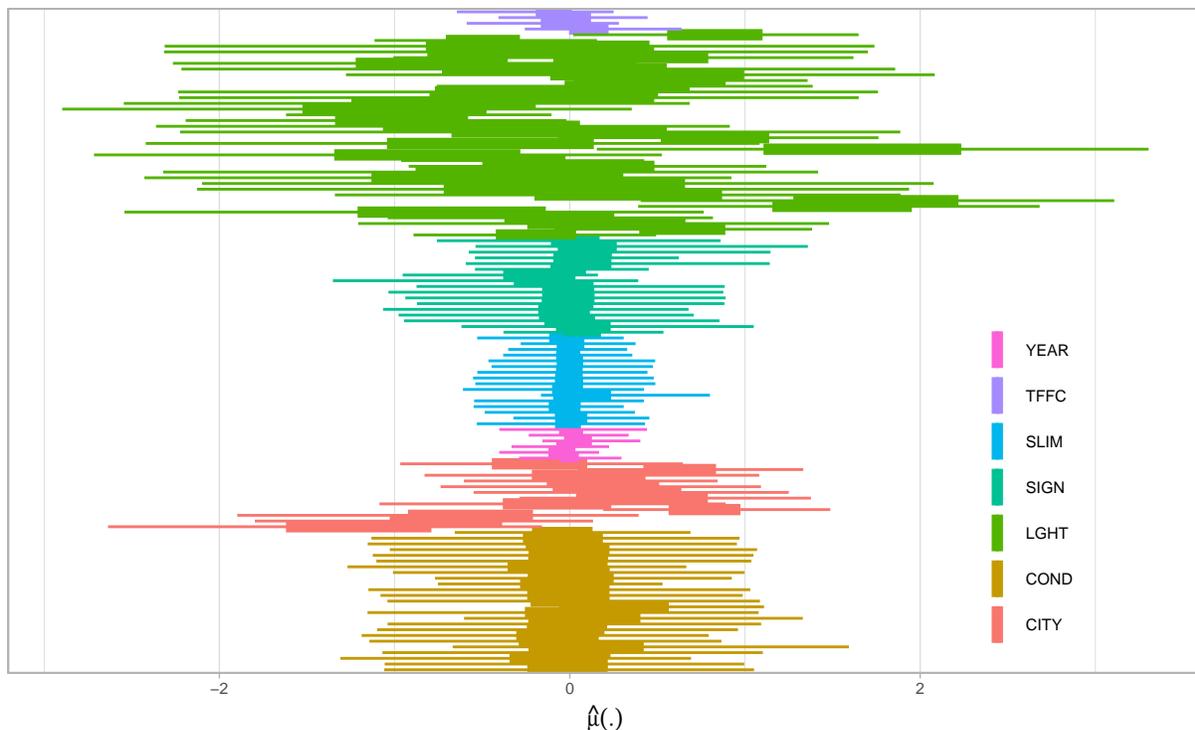
This figure exhibits inner 50 and 95 percent intervals for the effects from seven batches of explanatory variables. Effects are colored by batch and are interpreted as the log of the expected multiplicative increase in the fatality rate compared to the average road region. Time of day and lighting and city explain a substantial amount of variation in the fatality rate.

```
ggplot(coef_ggplot[coef_ggplot$wthn =="BLTE_e",]) +
  theme(plot.title = element_text(size=9, hjust = .5),
        plot.caption = element_text(hjust = 1)) +
  aes(numb, coef_mean) +
  geom_linerange(aes(ymin = lower50, ymax = upper50),size=2, alpha = .5) +
```

```
geom_linerange(aes(ymin = lower95, ymax = upper95), alpha = .5) +
coord_flip() +
scale_x_discrete(breaks = NULL) +
theme(legend.position = "none") +
labs(title="Figure 4.2: Model 1, Analysis of Variance, Within",
     y=expression(hat(mu)("BLTE")),x="",
     color = "",
     caption = paste("This figure exhibits inner 50 and 95 percent intervals for the",
                     "batch of built environment effects. Coefficients are",
                     "interpreted as the log of the expected \nmultiplicative",
                     "increase in the fatality rate compared to the average road",
                     "region."))
```

Figure 4.2: Model 1, Analysis of Variance, Within
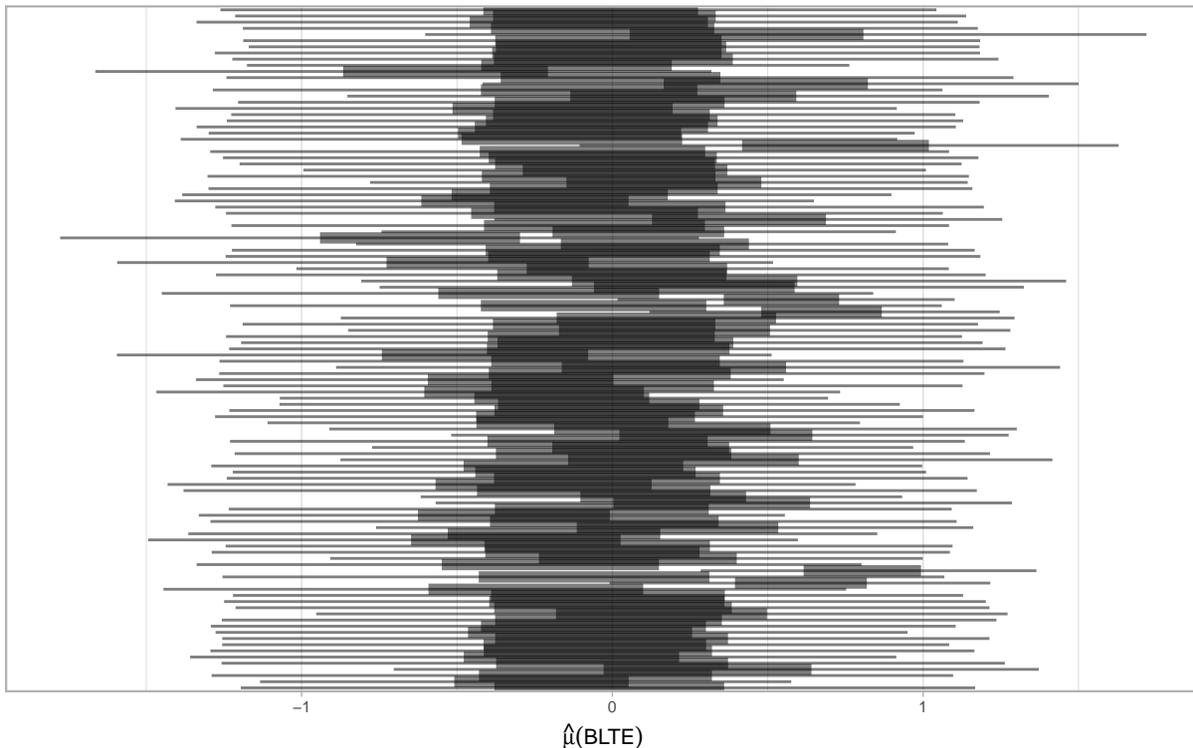
$\hat{\mu}$(BLTE)

This figure exhibits inner 50 and 95 percent intervals for the batch of built environment effects. Coefficients are interpreted as the log of the expected multiplicative increase in the fatality rate compared to the average road region.

```
city_anova <- data.frame(coef_ggplot[coef_ggplot$wthn == "CITY_e",],
                 city_labels = c("Washington D.C.","Boston","Austin",
                                 "Denver","Portland","Chicago",
                                 "Seattle","Los Angeles","San Diego",
                                 "San Francisco","San Jose",
                                 "New York City"))
city_anova$city_labels <- factor(city_anova$city_labels,levels=levels(city_labels))

ggplot(city_anova) +
  theme(plot.title = element_text(size=9, hjust = .5),
        plot.caption = element_text(hjust = 1)) +
  aes(city_labels, lower50 + (upper50 - lower50)/2) +
  geom_linerange(aes(ymin = lower50, ymax = upper50),size=7) +
```

```r
geom_linerange(aes(ymin = lower95, ymax = upper95)) +
geom_text(aes(label = city_labels),color="white",size=2) +
coord_flip() +
scale_x_discrete(breaks = NULL) +
theme(legend.position = "none") +
labs(title="Figure 4.3: Model 1, Analysis of Variance, Within",
     y=expression(hat(mu)("CITY")),x="",
     color = "",
     caption = paste("This figure exhibits inner 50 and 95 percent intervals for the",
                     "batch of city effects. Coefficients are interpreted as the log",
                     "of the expected multiplicative \nincrease in the fatality rate",
                     "compared to the average road region. New York City and Los",
                     "Angeles have roughly twice the average fatality rate while",
                     "\nWashington D.C. has roughly one third, holding all else",
                     "equal."))
```

Figure 4.3: Model 1, Analysis of Variance, Within
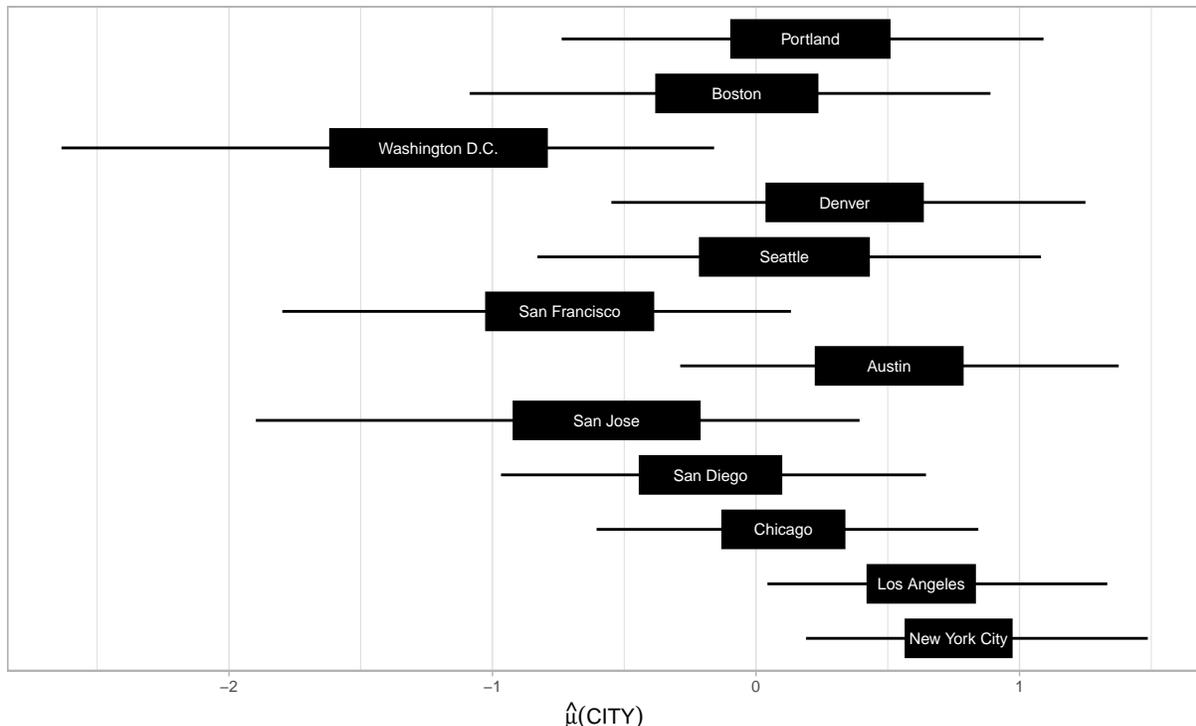
This figure exhibits inner 50 and 95 percent intervals for the batch of city effects. Coefficients are interpreted as the log of the expected multiplicative increase in the fatality rate compared to the average road region. New York City and Los Angeles have roughly twice the average fatality rate while Washington D.C. has roughly one third, holding all else equal.

```r
ggplot(data.frame(coef_ggplot[coef_ggplot$wthn == "SLIM_e",][5:9,],
                  slim_names = c("20 MPH","25 MPH","30 MPH",
                                 "35 MPH","40 MPH"))) +
  theme(plot.title = element_text(size=9, hjust = .5),
        plot.caption = element_text(hjust = 1)) +
  aes(numb, lower50 + (upper50 - lower50)/2) +
  geom_linerange(aes(ymin = lower50, ymax = upper50),size=7) +
  geom_linerange(aes(ymin = lower95, ymax = upper95)) +
  geom_text(aes(label = slim_names),color="white",size=3) +
  coord_flip() +
```

```
scale_x_discrete(breaks = NULL) +
theme(legend.position = "none") +
labs(title="Figure 4.4: Model 1, Analysis of Variance, Within",
     y=expression(hat(mu)("SLIM")),x="",
     color = "",
     caption = paste("This figure exhibits inner 50 and 95 percent intervals for",
                     "some speed limit effects. Effects are interpreted as the log",
                     "of the expected multiplicative \nincrease in the fatality rate",
                     "compared to the average road region. With this model,",
                     "individual speed limit effects are not discernable."))
```

Figure 4.4: Model 1, Analysis of Variance, Within
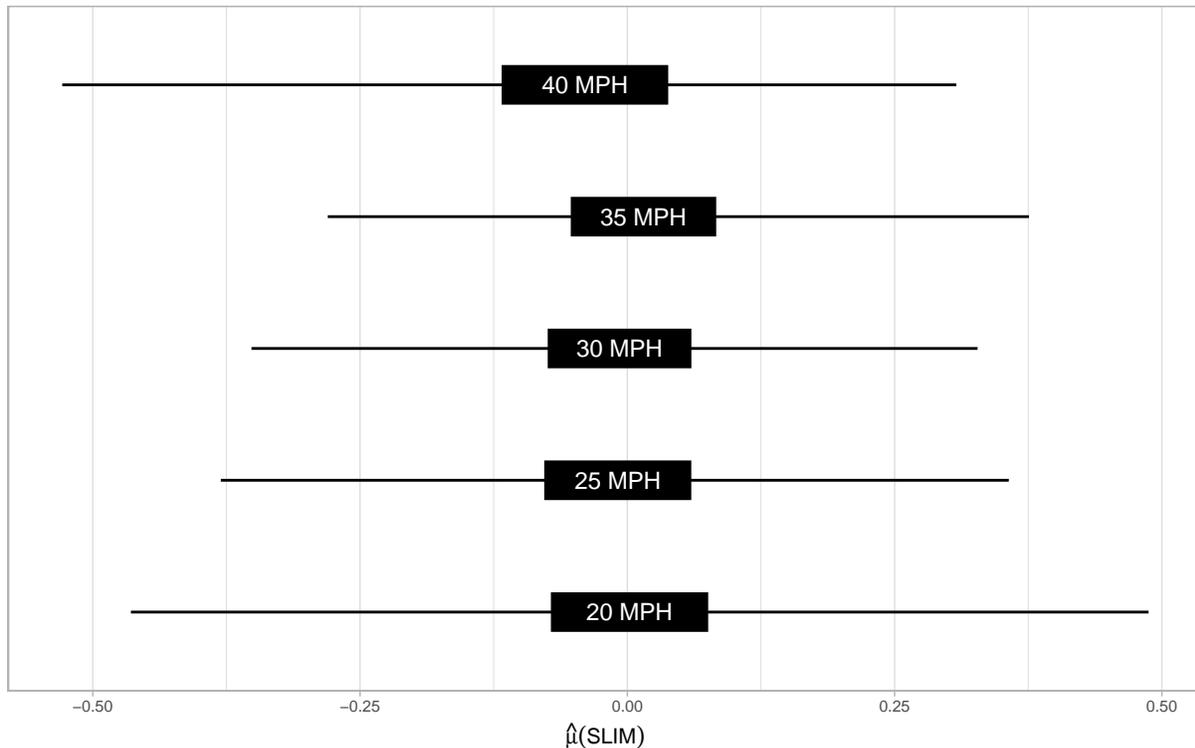
This figure exhibits inner 50 and 95 percent intervals for some speed limit effects. Effects are interpreted as the log of the expected multiplicative increase in the fatality rate compared to the average road region. With this model, individual speed limit effects are not discernable.

Comparison

Figure 3 suggests dropping variables YEAR and TFF and adding the following interactions to the regression: CITYxSLIM, LGHTxSLIM, BLTExSLIM, CONDxSLIM, LGHTxCOND and LGHTxCONDxSLIM. Figure 5 shows the resulting ANOVA plot using the posterior draws from this second model (Model 2) after repeating the same steps outlined in the Computation section. The within cell error term is much smaller than in Figure 3, suggesting that the new covariates jointly explain more variation than the original model. In addition, Figure 6.1, unlike Figure 4.5, shows a slightly higher average fatality rate at 30 than 25 MPH roads. Such a difference is small but potentially meaningful.

```
crash_data_int <- crash_data
crash_data_int <- rbind(crash_data_int,crash_data_int[crash_data_int$YEAR==6,])
crash_data_int$SLIM_actual <- crash_data_int$SLIM
crash_data_int$SLIM[crash_data_int$YEAR==6] <-
  sort(rep(c(6,7),sum(crash_data_int$YEAR == 6)/2))
```

```r
for(int in c("CITYxSLIM","LGHTxSLIM","BLTExSLIM","CONDxSLIM","LGHTxCOND"))
  crash_data_int[int] <- as.numeric(factor(interaction(
    eval(parse(text=paste0("crash_data_int$",strsplit(int,"x")[[1]])[[1]])),
    eval(parse(text=paste0("crash_data_int$",strsplit(int,"x")[[1]])[[2]])),
    labels = 1:length(unique(interaction(
    eval(parse(text=paste0("crash_data_int$",strsplit(int,"x")[[1]])[[1]])),
    eval(parse(text=paste0("crash_data_int$",strsplit(int,"x")[[1]])[[2]]))))))
  ))
crash_data_int$LGHTxCONDxSLIM <-  as.numeric(
  factor(with(crash_data_int,interaction(SLIM,LGHT,COND)),
         labels = 1:length(unique(with(crash_data_int,interaction(SLIM,LGHT,COND))))))
)
crash_data_int$TEST <- crash_data_int$YEAR == 6
crash_data_int <- crash_data_int[,c(1:2,4:7,9:18)]
G <- 12
J <- apply(crash_data_int[, c(1:6,10:15)], 2, function(x) length(unique(x)))
stan_dat_int <- aggregate(cbind(1,EXPR)~., crash_data_int, sum)
stan_dat_list <- with(stan_dat_int,
                      list(N_train = which.max(stan_dat_int$TEST)-1,
                           N = nrow(stan_dat_int),
                           J = J,
                           G = G,
                           COND = COND,
                           CITY = CITY,
                           SLIM = SLIM,
                           SIGN = SIGN,
                           LGHT = LGHT,
                           BLTE = BLTE,
                           CITYxSLIM = CITYxSLIM,
                           LGHTxSLIM = LGHTxSLIM,
                           BLTExSLIM = BLTExSLIM,
                           CONDxSLIM = CONDxSLIM,
                           LGHTxCOND = LGHTxCOND,
                           LGHTxCONDxSLIM = LGHTxCONDxSLIM,
                           count = V1,
                           EXPR = EXPR))
model_int <- stan_model(file = 'model_int.stan')
fit_int <- sampling(model_int, data = stan_dat_list, iter = 2000,
                    chains = 4, refresh = 10, cores = 4,
                    control = list(adapt_delta = 0.99))
```

```r
btwn <- c("COND_sd","CITY_sd","SLIM_sd","SIGN_sd","LGHT_sd","BLTE_sd",
          "cell_sd","CITYxSLIM_sd", "LGHTxSLIM_sd","BLTExSLIM_sd",
          "CONDxSLIM_sd","LGHTxCOND_sd","LGHTxCONDxSLIM_sd")
coefs <- data.frame(extract(fit_int,pars=btwn))
coef_ggplot <- data.frame(coef_mean=apply(coefs,2,mean),btwn = btwn)
coef_ggplot$upper50 <- apply(coefs,2,quantile,probs=.75)
coef_ggplot$lower50 <- apply(coefs,2,quantile,probs=.25)
coef_ggplot$upper95 <- apply(coefs,2,quantile,probs=.975)
coef_ggplot$lower95 <- apply(coefs,2,quantile,probs=.025)
coef_ggplot$btwn <- substr(coef_ggplot$btwn,1,nchar(as.character(coef_ggplot$btwn))-3)

ggplot(coef_ggplot, aes(btwn, coef_mean)) +
  theme(plot.title = element_text(size=9, hjust = .5),
```

```
        plot.caption = element_text(hjust = 1)) +
geom_linerange(aes(ymin = lower50, ymax = upper50),size=2) +
geom_linerange(aes(ymin = lower95, ymax = upper95)) +
coord_flip() +
scale_y_continuous(limits=c(0,1.1*max(coef_ggplot$upper95)),
                   expand = c(0, 0)) +
labs(y=expression(hat(sigma)(.)),x="",
     title="Figure 5: Model 2, Analysis of Variance, Between",
        caption = paste("This figure exhibits inner 50 and 95 percent intervals",
                        "for the finite-population standard deviations of each",
                        "batch of explanatory variables. \nCompared to Model 1,",
                        "Time of day and lighting and city remain strong",
                        "predictors of the fatality rate. Weather and surface",
                        "condition has \nincreased in importance. In addition,",
                        "interactions between the weather and surface condition",
                        "and speed limit and the built environment \nand speed",
                        "limit appear moderately predictive."))
```

Figure 5: Model 2, Analysis of Variance, Between
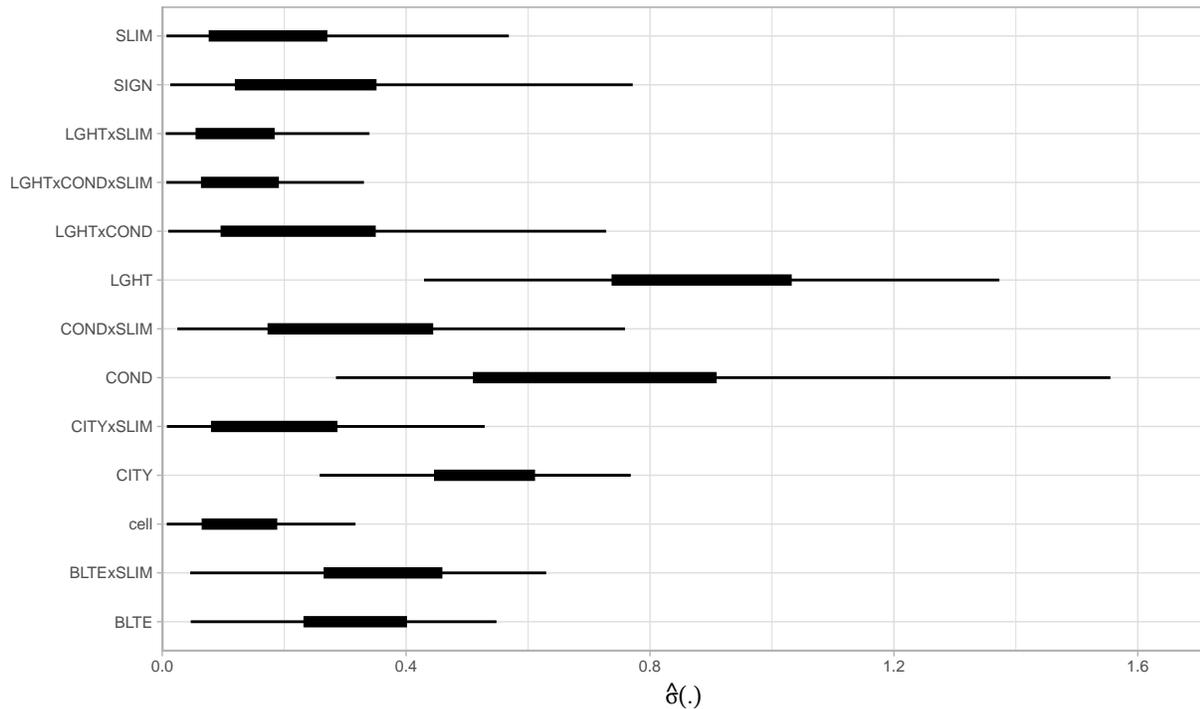
This figure exhibits inner 50 and 95 percent intervals for the finite–population standard deviations of each batch of explanatory variables.
Compared to Model 1, Time of day and lighting and city remain strong predictors of the fatality rate. Weather and surface condition has
increased in importance. In addition, interactions between the weather and surface condition and speed limit and the built environment
and speed limit appear moderately predictive.

```
wthn <- c("SLIM_e","COND_e","BLTExSLIM_e","CONDxSLIM_e")
coefs <- extract(fit_int,pars=wthn)
coef_ggplot <- data.frame(wthn = character(),
                          numb = character(),
                          coef_mean=numeric(),
                          upper50 = numeric(),
                          lower50 = numeric(),
                          upper95 = numeric(),
```

```
                          lower95 = numeric())
for(var in seq_along(coefs)){
  coefs_temp <- coefs[[var]]
  coef_ggplot_temp <- data.frame(wthn = wthn[var],
                                 numb = paste0(wthn[var],1:ncol(coefs_temp)),
                                 coef_mean=apply(coefs_temp,2,mean))
  coef_ggplot_temp$upper50 <- apply(coefs_temp,2,quantile,probs=.75)
  coef_ggplot_temp$lower50 <- apply(coefs_temp,2,quantile,probs=.25)
  coef_ggplot_temp$upper95 <- apply(coefs_temp,2,quantile,probs=.975)
  coef_ggplot_temp$lower95 <- apply(coefs_temp,2,quantile,probs=.025)
  coef_ggplot <- rbind(coef_ggplot,coef_ggplot_temp)
}

ggplot(data.frame(coef_ggplot[coef_ggplot$wthn == "SLIM_e",][5:9,],
                  slim_names = c("20 MPH","25 MPH","30 MPH",
                                 "35 MPH","40 MPH"))) +
  theme(plot.title = element_text(size=9, hjust = .5),
        plot.caption = element_text(hjust = 1)) +
  aes(numb, lower50 + (upper50 - lower50)/2) +
  geom_linerange(aes(ymin = lower50, ymax = upper50),size=7) +
  geom_linerange(aes(ymin = lower95, ymax = upper95)) +
  geom_text(aes(label = slim_names),color="white",size=3) +
  coord_flip() +
  scale_x_discrete(breaks = NULL) +
  theme(legend.position = "none") +
  labs(title="Figure 6.1: Model 2, Analysis of Variance, Within",
       y=expression(hat(mu)("SLIM")),x="",
       color = "",
       caption = paste("This figure exhibits inner 50 and 95 percent intervals for",
                       "some speed limit effects. Effects are interpreted as the log",
                       "of the expected multiplicative \nincrease in the fatality rate",
                       "compared to the average road region. With this model, slight",
                       "average speed limit effects are observable although \nthese",
                       "differences are likely too small to be meaningful."))
```

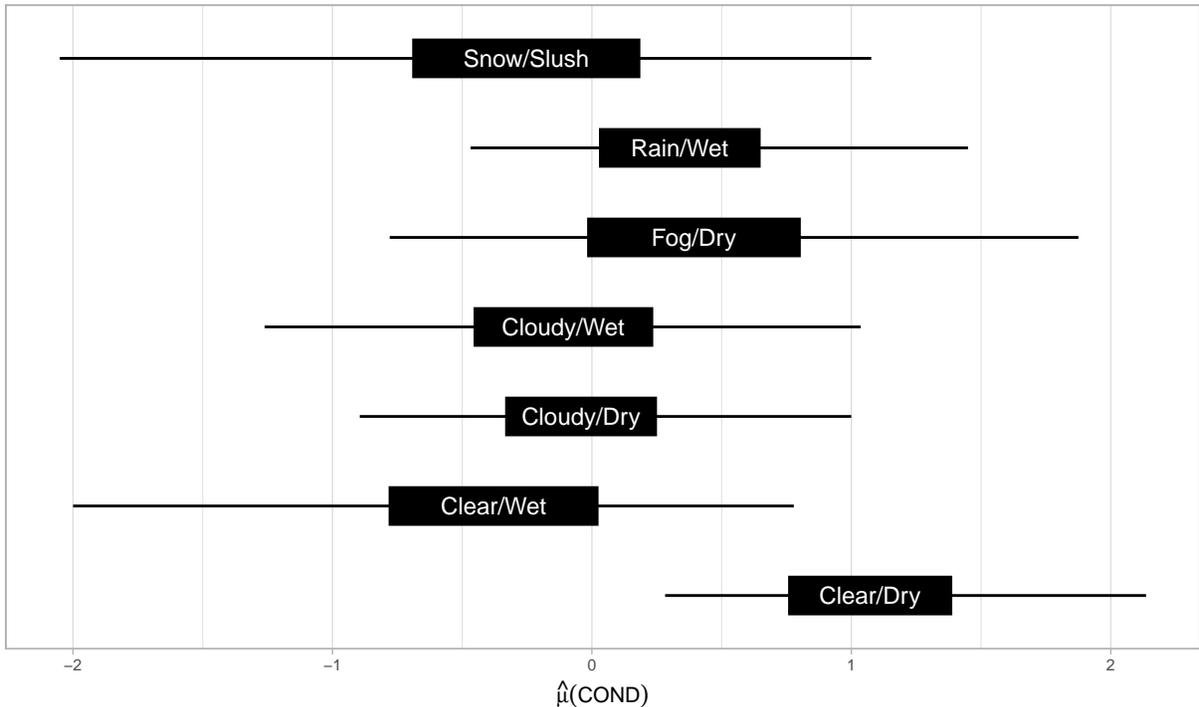Figure 6.1: Model 2, Analysis of Variance, Within
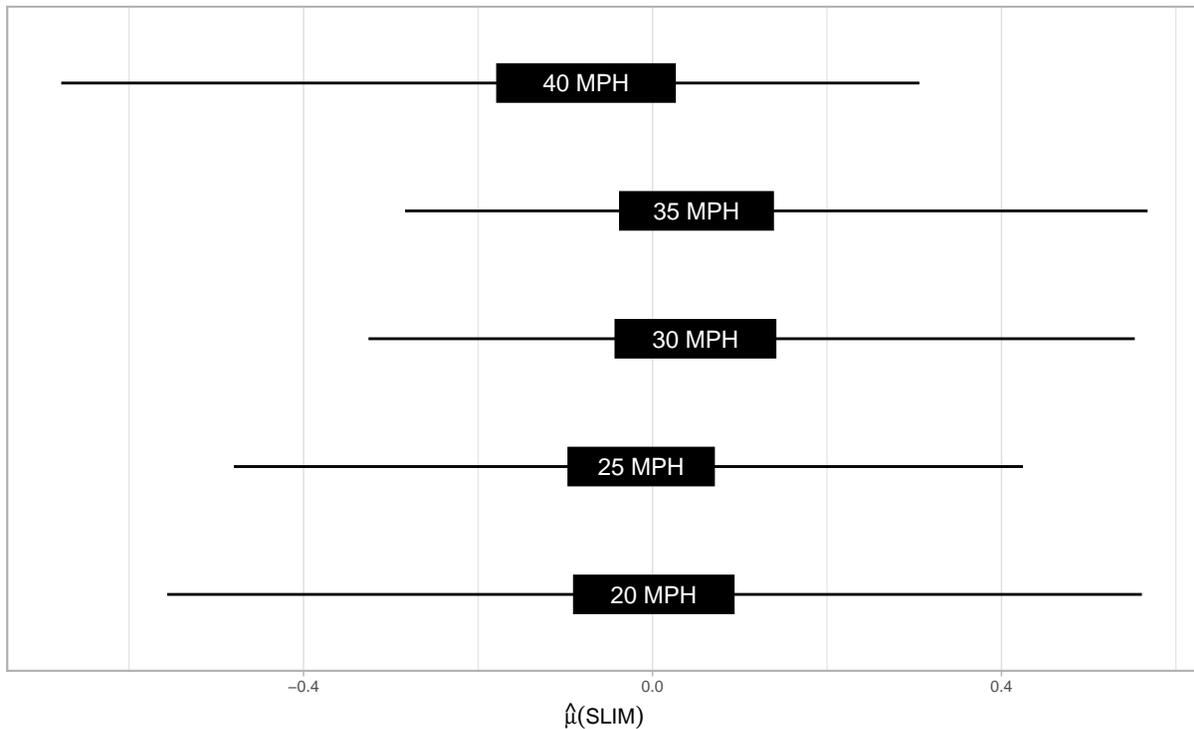
This figure exhibits inner 50 and 95 percent intervals for some speed limit effects. Effects are interpreted as the log of the expected multiplicative increase in the fatality rate compared to the average road region. With this model, slight average speed limit effects are observable although these differences are likely too small to be meaningful.

While the average speed limit effect is small, Figures 6.3 and 6.4 display some meaningful interactions. For example, 25 mph roads regions appear safer than 30 mph road regions during cloudy weather and at large, four lane intersections.

```r
ggplot(data.frame(coef_ggplot[coef_ggplot$wthn == "COND_e",][c(2,3,9,15,18,23,24),],
      cond_names = c("Clear/Dry",  #weather is clear/road is dry
                     "Clear/Wet",  #weather is clear/road is wet
                     "Rain/Wet",   #weather is rain/road is wet
                     "Snow/Slush", #weather is snow/road is slush
                     "Fog/Dry",    #weather is fog/road is dry
                     "Cloudy/Dry", #weather is cloudy/road is dry
                     "Cloudy/Wet"  #weather is cloudy/road is wet
                     ))) +
  theme(plot.title = element_text(size=9, hjust = .5),
        plot.caption = element_text(hjust = 1)) +
  aes(cond_names, lower50 + (upper50 - lower50)/2) +
  geom_linerange(aes(ymin = lower50, ymax = upper50),size=7) +
  geom_linerange(aes(ymin = lower95, ymax = upper95)) +
  geom_text(aes(label = cond_names),color="white",size=3) +
  coord_flip() +
    scale_x_discrete(breaks = NULL) +
  theme(legend.position = "none") +
  labs(title="Figure 6.2: Model 2, Analysis of Variance, Within",
       y=expression(hat(mu)("COND")),x="",
       color = "",
       caption = paste("This figure exhibits inner 50 and 95 percent intervals for",
                       "some weather and surface condition effects. Effects are",
```

```
                 "interpreted as the log of the expected \nmultiplicative",
                 "increase in the fatality rate compared to the average road",
                 "region. Clear weather and dry roads have significantly larger",
                 "fatality rates \npossibly reflecting the fact that more",
                 "pedestrians are on roads and interacting with vehicles.",
                 "Rainstorms also appear to create hazardous \nconditions that",
                 "put pedestrians at increased risk."))
```

Figure 6.2: Model 2, Analysis of Variance, Within

This figure exhibits inner 50 and 95 percent intervals for some weather and surface condition effects. Effects are interpreted as the log of the expected multiplicative increase in the fatality rate compared to the average road region. Clear weather and dry roads have significantly larger fatality rates possibly reflecting the fact that more pedestrians are on roads and interacting with vehicles. Rainstorms also appear to create hazardous conditions that put pedestrians at increased risk.

```
ggplot(data.frame(coef_ggplot[coef_ggplot$wthn == "CONDxSLIM_e",][
    c(12,15,24,25,28,35,45,46),],
    cond_names = c("Clear/Dry",
                   "Rain/Wet",
                   "Cloudy/Dry",
                   "Cloudy/Wet",
                   "Clear/Dry",
                   "Rain/Wet",
                   "Cloudy/Dry",
                   "Cloudy/Wet"
                   ),
    slim = c(rep("25 MPH",4),rep("30 MPH",4)))) +
theme(plot.title = element_text(size=9, hjust = .5),
      plot.caption = element_text(hjust = 1)) +
aes(paste0(cond_names,slim), lower50 + (upper50 - lower50)/2, color = slim) +
geom_linerange(aes(ymin = lower50, ymax = upper50),size=7) +
geom_linerange(aes(ymin = lower95, ymax = upper95)) +
geom_text(aes(label = cond_names),color="white",size=3) +
```

```
coord_flip() +
scale_x_discrete(breaks = NULL) +
theme(legend.position = c(.8, 0.8)) +
scale_color_discrete(guide = guide_legend(reverse = TRUE)) +
  labs(title="Figure 6.3: Model 2, Analysis of Variance, Within",
      y=expression(hat(mu)("CONDxSLIM")),x="",
      color = "",
      caption = paste("This figure exhibits inner 50 and 95 percent intervals for",
                      "some interaction effects between the posted speed limit and",
                      "the weather and surface \ncondition. Effects are interpreted",
                      "as the log of the expected multiplicative increase in the",
                      "fatality rate compared to the average road region. \nLower",
                      "speed limits coincide with lower fatality rates during cloudy",
                      "weather possibly because pedestrians remain on the road but",
                      "driver \nvision is relatively obstructed. There does not",
                      "appear to be much difference between speed limits during",
                      "clear weather."))
```

Figure 6.3: Model 2, Analysis of Variance, Within
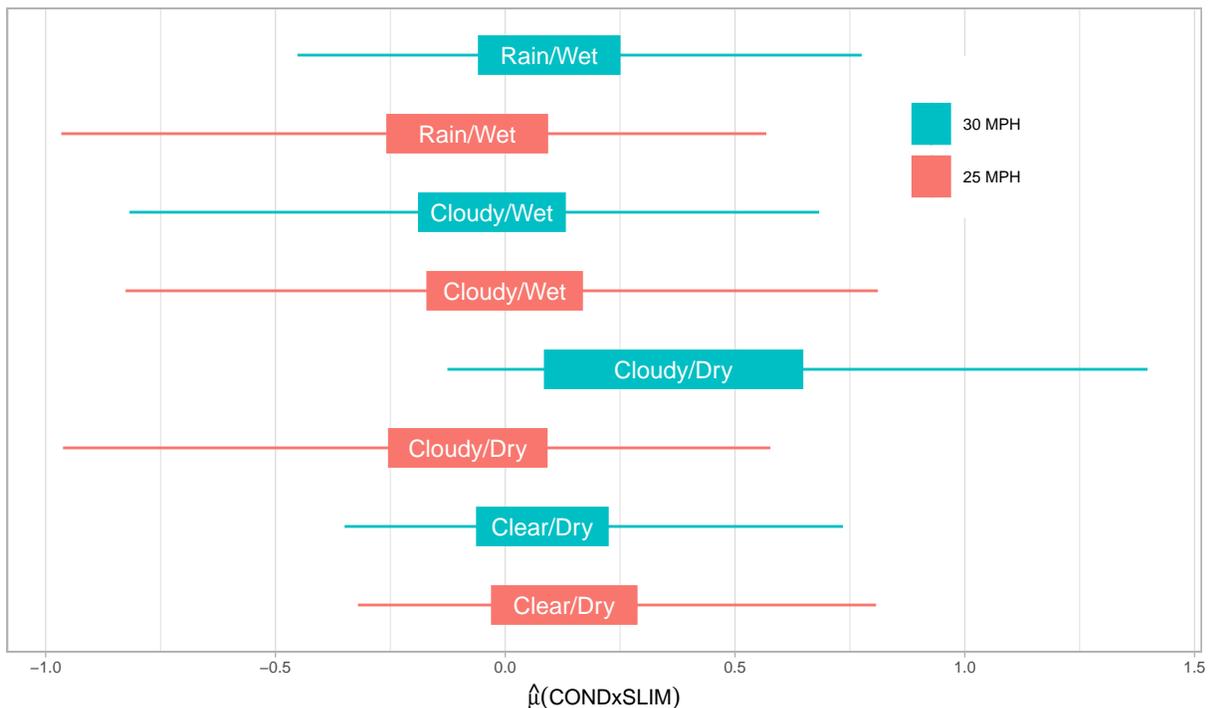
This figure exhibits inner 50 and 95 percent intervals for some interaction effects between the posted speed limit and the weather and surface condition. Effects are interpreted as the log of the expected multiplicative increase in the fatality rate compared to the average road region. Lower speed limits coincide with lower fatality rates during cloudy weather possibly because pedestrians remain on the road but driver vision is relatively obstructed. There does not appear to be much difference between speed limits during clear weather.

```
ggplot(data.frame(coef_ggplot[coef_ggplot$wthn == "BLTExSLIM_e",][
        c(26,28,57,59,105,107,138,140),],
      cond_names = c("Two Lane Road",
                    "Two Lane Intersec",
                    "Four Lane Road",
                    "Four Lane Intersec",
                    "Two Lane Road",
                    "Two Lane Intersec",
                    "Four Lane Road",
```

```
                      "Four Lane Intersec"
                     ),
     slim = c(rep("25 MPH",4),rep("30 MPH",4)))) +
theme(plot.title = element_text(size=9, hjust = .5),
      plot.caption = element_text(hjust = 1)) +
aes(paste0(cond_names,slim), lower50 + (upper50 - lower50)/2, color = slim) +
geom_linerange(aes(ymin = lower50, ymax = upper50),size=7) +
geom_linerange(aes(ymin = lower95, ymax = upper95)) +
geom_text(aes(label = cond_names),color="white",size=3) +
coord_flip() +
scale_x_discrete(breaks = NULL) +
theme(legend.position = c(.1, 0.8)) +
scale_color_discrete(guide = guide_legend(reverse = TRUE)) +
labs(title="Figure 6.4: Model 2, Analysis of Variance, Within",
     y=expression(hat(mu)("BLTExSLIM")),x="",
     color = "",
     caption = paste("This figure exhibits inner 50 and 95 percent intervals for",
                     "some interaction effects between the posted speed limit and",
                     "the built environment. Effects \nare interpreted as the log",
                     "of the expected multiplicative increase in the fatality rate",
                     "compared to the average road region. Lower speed limits",
                     "\nappear to coincide with lower fatality rates across the",
                     "major road types with the largest decreases in intersections."))
```

Figure 6.4: Model 2, Analysis of Variance, Within
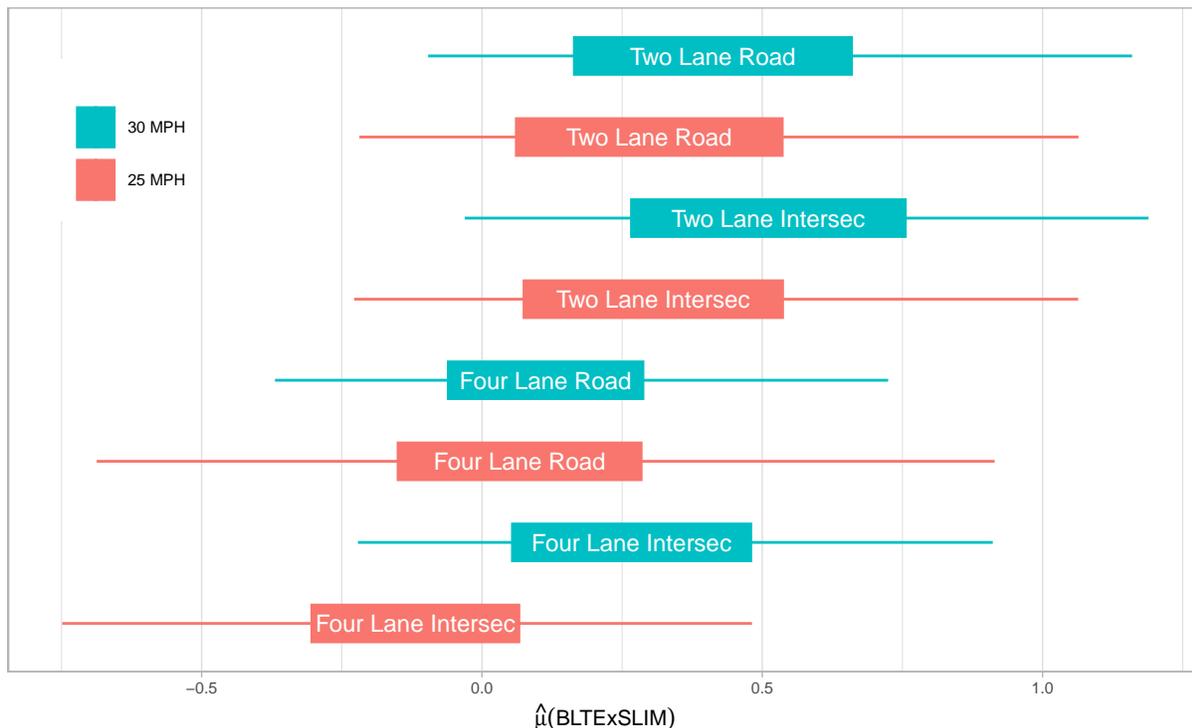
This figure exhibits inner 50 and 95 percent intervals for some interaction effects between the posted speed limit and the built environment. Effects are interpreted as the log of the expected multiplicative increase in the fatality rate compared to the average road region. Lower speed limits appear to coincide with lower fatality rates across the major road types with the largest decreases in intersections.

Criticism

We can check both of the preceding models by comparing the posterior predictions for the number of deaths on the 2015 sample roads with the actual number of deaths observed on those roads (Gelman et al. 2014). This ensures the models do not over predict or under predict the number of deaths in each stratum. We use a p-value, the percentage of samples from the predictive distribution exceeding the observed number of deaths, to indicate the suitability of the models for prediction. Figures 7.1 and 7.2 demonstrate that the original model without interactions coincides with the observed data for 30 mph road regions but under predicts the observed data for 25 mph road regions. Figure 7.3 shows that the second model with interactions better fits the observed data. We will use this model in the Policy Analysis section.

```r
rpois_trunc <- function(log_rate){
  rate <- exp(log_rate)
  cum_cdf <- exp(-rate) / (1 - exp(-rate)) * rate
  x <- 1; u <- runif(1)
  while (cum_cdf < u) {
    x <- x + 1;
    cum_cdf <- cum_cdf * (1 + rate / x);
    }
  x
}

pred30 <- apply(extract(fit,c("mu_indiv_pred30"))[[1]], c(1,2), rpois_trunc)
city_30 <- apply(pred30[,stan_dat$SLIM[stan_dat$YEAR == 6] == 7],1,sum)
pvalue_30 <- paste("p-value:",round(sum(city_30 > sum(stan_dat$V1[
    stan_dat$SLIM == 7 & stan_dat$YEAR == 6]))/length(city_30),2))

qplot(city_30) +
  theme(plot.title = element_text(size=9, hjust = .5),
        plot.caption = element_text(hjust = 1)) +
  geom_vline(xintercept = sum(stan_dat$V1[stan_dat$SLIM == 7 &
                                          stan_dat$YEAR == 6]),
             linetype=2) +
  geom_text(aes(label=pvalue_30),x=80,y=750) +
  scale_x_continuous(expand = c(0,0)) +
  scale_y_continuous(expand = c(0,50), limits = c(0,NA)) +
  labs(title=
"Figure 7.1: Model 1, Posterior Draws of Deaths in 2015 on 30 mph roads (unweighted)",
    x = "", y = "",
    caption = paste("This figure exhibits a histogram of draws from the posterior",
                    "predictive distribution of the total number of fatalities on the",
                    "30 mph road regions in \n2015 set aside as a test set. The dotted",
                    "line represents the observed number of fatalities on those road",
                    "regions. The p-value is the proportion \nof simulations exceeding",
                    "the number observed in 2015. The purpose of this figure is to",
                    "ensure that the model is predictive of the fatality rate \nof the",
                    "roads that did not have their posted speed limits reduced to 25",
                    "mph. Model 1 appears consistent with the total number of",
                    "fatalities \non these road regions."))
```

Figure 7.1: Model 1, Posterior Draws of Deaths in 2015 on 30 mph roads (unweighted)
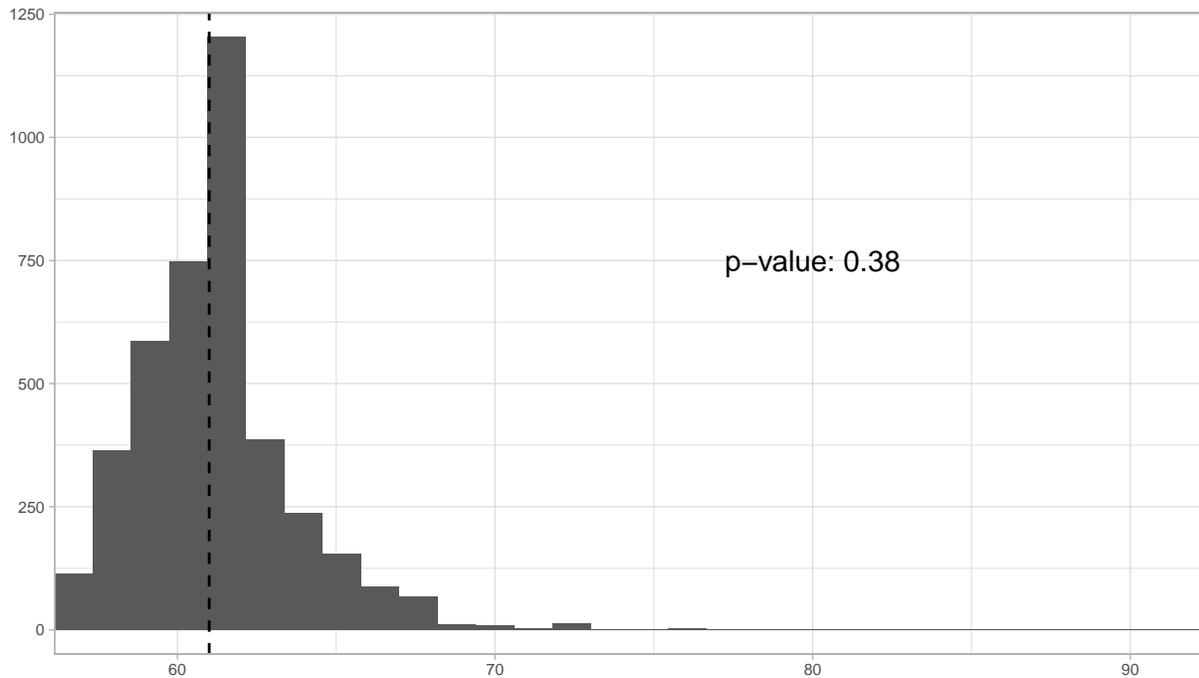
p–value: 0.38

This figure exhibits a histogram of draws from the posterior predictive distribution of the total number of fatalities on the 30 mph road regions in 2015 set aside as a test set. The dotted line represents the observed number of fatalities on those road regions. The p–value is the proportion of simulations exceeding the number observed in 2015. The purpose of this figure is to ensure that the model is predictive of the fatality rate of the roads that did not have their posted speed limits reduced to 25 mph. Model 1 appears consistent with the total number of fatalities on these road regions.

```
pred25 <- apply(extract(fit,c("mu_indiv_pred25"))[[1]], c(1,2), rpois_trunc)
city_25 <- apply(pred25[,stan_dat$SLIM[stan_dat$YEAR == 6] == 6],1,sum)
pvalue_25 <- paste("p-value:",round(sum(city_25 > sum(stan_dat$V1[
    stan_dat$SLIM == 6 & stan_dat$YEAR == 6]))/length(city_25),2))

qplot(city_25) +
  theme(plot.title = element_text(size=9, hjust = .5),
        plot.caption = element_text(hjust = 1)) +
  geom_vline(xintercept = sum(stan_dat$V1[stan_dat$SLIM == 6 &
                                          stan_dat$YEAR == 6]),
             linetype=2) +
  geom_text(aes(label=pvalue_25,x=200,y=500)) +
  scale_x_continuous(expand = c(0,0)) +
  scale_y_continuous(expand = c(0,50), limits = c(0,NA)) +
  labs(title=
"Figure 7.2: Model 1, Posterior Draws of Deaths in 2015 on 25 mph roads (unweighted)",
    x = "", y = "",
    caption = paste("This figure exhibits a histogram of draws from the posterior",
                    "predictive distribution of the total number of fatalities on the",
                    "25 mph road regions in \n2015 set aside as a test set. The dotted",
                    "line represents the observed number of fatalities on those road",
                    "regions. The p-value is the proportion \nof simulations exceeding",
                    "the number observed in 2015. The purpose of this figure is to",
                    "ensure that the model is predictive of the fatality rate \nof the",
                    "roads that had their posted speed limits reduced to 25 mph.",
                    "Model 1 appears to systematically under predict the total number",
                    "of\n fatalities on these road regions."))
```

Figure 7.2: Model 1, Posterior Draws of Deaths in 2015 on 25 mph roads (unweighted)
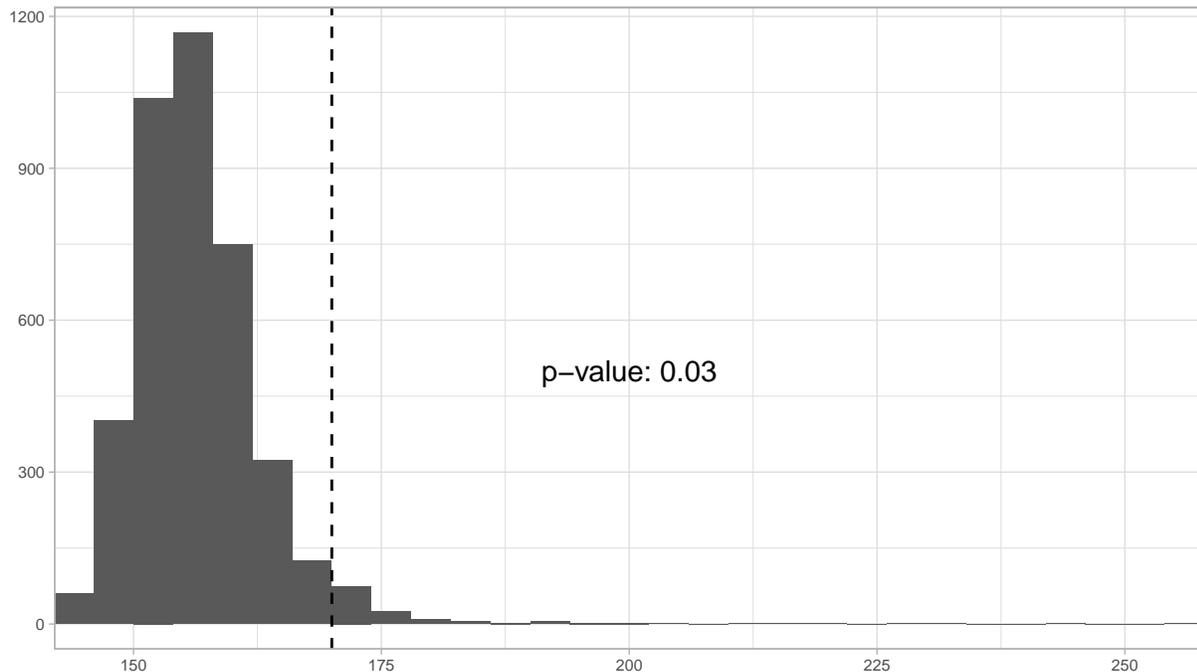
p–value: 0.03

This figure exhibits a histogram of draws from the posterior predictive distribution of the total number of fatalities on the 25 mph road regions in 2015 set aside as a test set. The dotted line represents the observed number of fatalities on those road regions. The p–value is the proportion of simulations exceeding the number observed in 2015. The purpose of this figure is to ensure that the model is predictive of the fatality rate of the roads that had their posted speed limits reduced to 25 mph. Model 1 appears to systematically under predict the total number of fatalities on these road regions.

```r
pred_int <- apply(extract(fit_int,c("mu_indiv_pred"))[[1]], c(1,2), rpois_trunc)
city_25_int <- apply(pred_int[,stan_dat_int$SLIM_actual[stan_dat_int$TEST == TRUE] == 6 &
                      stan_dat_int$SLIM[stan_dat_int$TEST == TRUE] == 6],1,sum)
pvalue_25_int <- paste("p-value:",round(sum(city_25_int > sum(stan_dat_int$V1[
   stan_dat_int$SLIM_actual == 6 & stan_dat_int$TEST == TRUE &
   stan_dat_int$SLIM == 6]))/length(city_25_int),2))

qplot(city_25_int) +
  theme(plot.title = element_text(size=9, hjust = .5),
        plot.caption = element_text(hjust = 1)) +
  geom_vline(xintercept = sum(stan_dat_int$V1[stan_dat_int$SLIM_actual == 6 &
                                              stan_dat_int$TEST == TRUE &
                                              stan_dat_int$SLIM == 6]),
             linetype=2) +
  geom_text(aes(label=pvalue_25_int,x=200,y=500)) +
  scale_x_continuous(expand = c(0,0)) +
  scale_y_continuous(expand = c(0,50), limits = c(0,NA)) +
  labs(title=
"Figure 7.3: Model 2, Posterior Draws of Deaths in 2015 on 25 mph roads (unweighted)",
   x = "", y = "",
   caption = paste("This figure exhibits a histogram of draws from the posterior",
                   "predictive distribution of the total number of fatalities on the",
                   "25 mph road regions in \n2015 set aside as a test set. The dotted",
                   "line represents the observed number of fatalities on those road",
                   "regions. The p-value is the proportion \nof simulations exceeding",
```

```
                          "the number observed in 2015. The purpose of this figure is to",
                          "ensure that the model is predictive of the fatality rate \nof the",
                          "roads that had their posted speed limits reduced to 25 mph. Model",
                          "2 appears consistent with the total number of fatalities on",
                          "these \nroad regions."))
```

Figure 7.3: Model 2, Posterior Draws of Deaths in 2015 on 25 mph roads (unweighted)
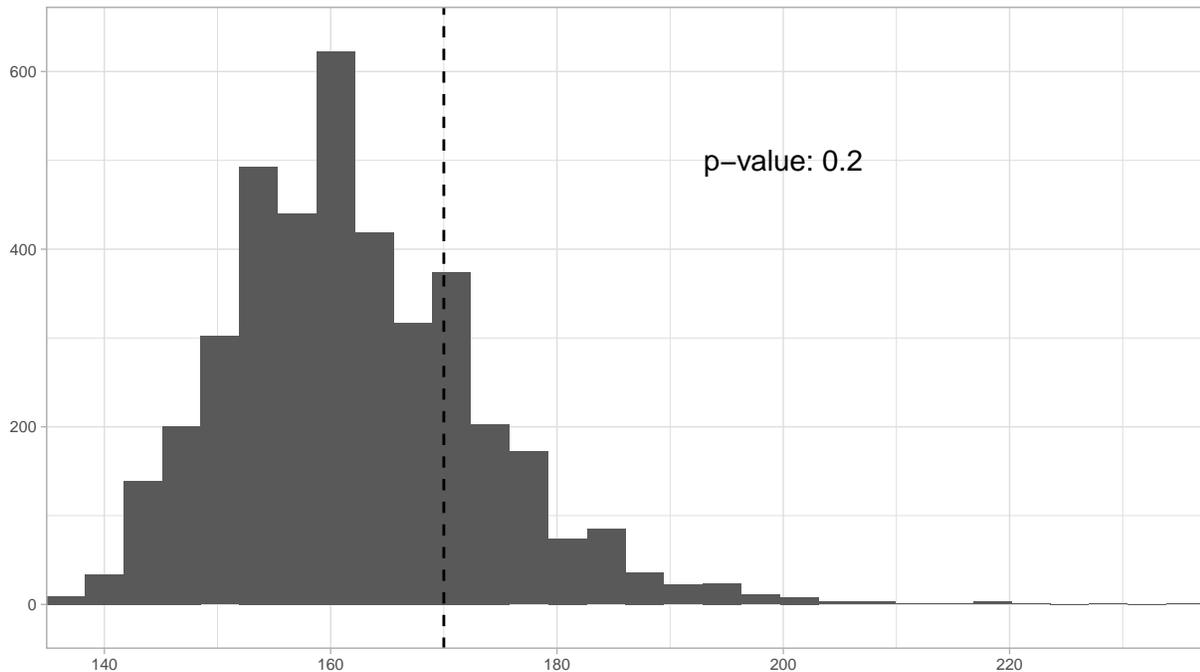
This figure exhibits a histogram of draws from the posterior predictive distribution of the total number of fatalities on the 25 mph road regions in 2015 set aside as a test set. The dotted line represents the observed number of fatalities on those road regions. The p–value is the proportion of simulations exceeding the number observed in 2015. The purpose of this figure is to ensure that the model is predictive of the fatality rate of the roads that had their posted speed limits reduced to 25 mph. Model 2 appears consistent with the total number of fatalities on these road regions.

III. Policy Analysis

A naïve before and after study would compare the number of pedestrian deaths in New York City in 2015 with the number of pedestrian deaths in New York City in 2014 or the 2010-2014 average. These comparisons are extremely popular among policy analysts because they are simple to calculate and simple to interpret: the same units are compared before and after treatment. The use of before-after analysis in transportation research is discussed at length in Hauer (2005) and Hauer (2015).

The number of pedestrian deaths in New York City each year is displayed in Figure 8.1, and a before-after comparison would estimate four or six lives saved respectively. This estimate, however, fails to account for possible factors that influence the number of pedestrian deaths and that changed in tandem with speed limits. As a result, we cannot rule out the influence of additional policy changes or other external factors.

```
ggplot(aggregate(V1 ~ YEAR,
                 stan_dat[stan_dat$CITY==12,],sum)) +
  theme(plot.title = element_text(size=9, hjust = .5),
        plot.caption = element_text(hjust = 1)) +
  aes(YEAR+2009,V1) +
  geom_line() +
  scale_x_continuous(limits=c(2009.5,2015.5),
```

```
                    expand = c(0, 0)) +
  scale_y_continuous(limits=c(0,200),expand = c(0,0)) +
  labs(x = "", y = "",
       title="Figure 8.1: Number of Pedestrian Deaths in NYC",
       caption = paste("This figure exhibits the annual number of pedestrian deaths",
                       "in New York City over the study period. New York City had",
                       "roughly 140 pedestrian \ndeaths each year except for 2013 when",
                       "there were 168. This reflects a 20 percent increase. City",
                       "officials have attributed the fatality decline \nfrom 2013 to",
                       "2015 to Vision Zero policy."))
```

Figure 8.1: Number of Pedestrian Deaths in NYC
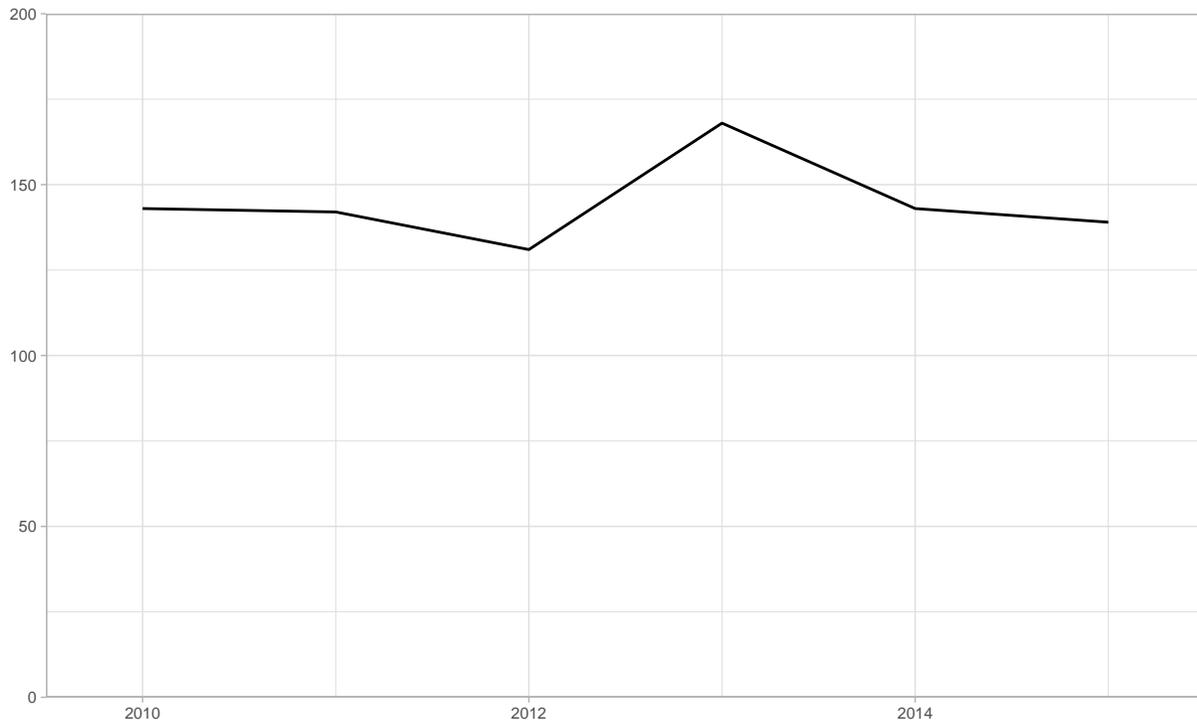
This figure exhibits the annual number of pedestrian deaths in New York City over the study period. New York City had roughly 140 pedestrian deaths each year except for 2013 when there were 168. This reflects a 20 percent increase. City officials have attributed the fatality decline from 2013 to 2015 to Vision Zero policy.

For example, the predictive importance of LGHT and COND that we identified in the ANOVA plots suggest that fatalities are largely driven by changes in weather, lighting and road surface conditions. Figure 8.2 shows the number of rainstorms in New York City relative to 2010 follows a similar trend to the fatality rate relative to 2010. If we interpret Figure 6.2 to suggest that rain storms increase the number of pedestrian deaths than Figure 8.2 implies that the number of pedestrian fatalities in New York City have actually increased after adjusting for the number of rain storms. Of course this is one of several plausible explanations for the trend of pedestrian fatalities. The purpose of Figure 8.2 is to simply motivate the necessity of accounting for major sources of confounding.

```
yearly_rainstorm <- function(x) {
 url <- paste0("https://www.wunderground.com/history/airport/jfk/",
        x,"/1/1/CustomHistory.html?dayend=31&monthend=12&yearend=",
        x,"&req_city=NA&req_state=NA&req_statename=NA&format=1")
 weather <- read.csv(url)
 sum(grepl("Rain",weather$Events) & weather$Mean.Wind.SpeedMPH > 11.20402)
}
```

```
deaths <- aggregate(V1 ~ YEAR,stan_dat[stan_dat$CITY==12,],sum)
deaths$deaths_norm <- deaths$V1/deaths$V1[1]
deaths$rain_storm <- sapply(2010:2015,yearly_rainstorm)
deaths$rain_storm_norm <- deaths$rain_storm/deaths$rain_storm[1]

ggplot(deaths) +
  theme(plot.title = element_text(size=9, hjust = .5),
        plot.caption = element_text(hjust = 1)) +
  geom_line(aes(YEAR+2009,deaths_norm)) +
  geom_line(aes(YEAR+2009,rain_storm_norm), linetype=2) +
  geom_text(x = 2014, y = 1.25, label = "Rain Storms") +
  geom_text(x = 2014, y = .9, label = "Fatalities") +
  scale_x_continuous(limits=c(2009.5,2015.5),
                     expand = c(0, 0)) +
  scale_y_continuous(limits=c(0,1.5),expand = c(0,0)) +
  labs(x = "", y = "",
       title="Figure 8.2: A Plausible Explanation of Pedestrian Deaths in NYC",
       caption = paste("This figure exhibits the annual number of pedestrian deaths",
                       "in New York City over the study period divided by the number",
                       "of deaths in 2010 \n(solid line) and the annual number of",
                       "rainstorms in New York City over the study period divided by",
                       "the number of rainstorms in 2010 \n(dashed line). While Figure",
                       "8.1 suggested that fatalities have fallen relative to 2013,",
                       "Figure 8.2 suggests that, relative to the \nnumber of rain",
                       "storms, fatalities have actually increased in 2015."))
```

Figure 8.2: A Plausible Explanation of Pedestrian Deaths in NYC
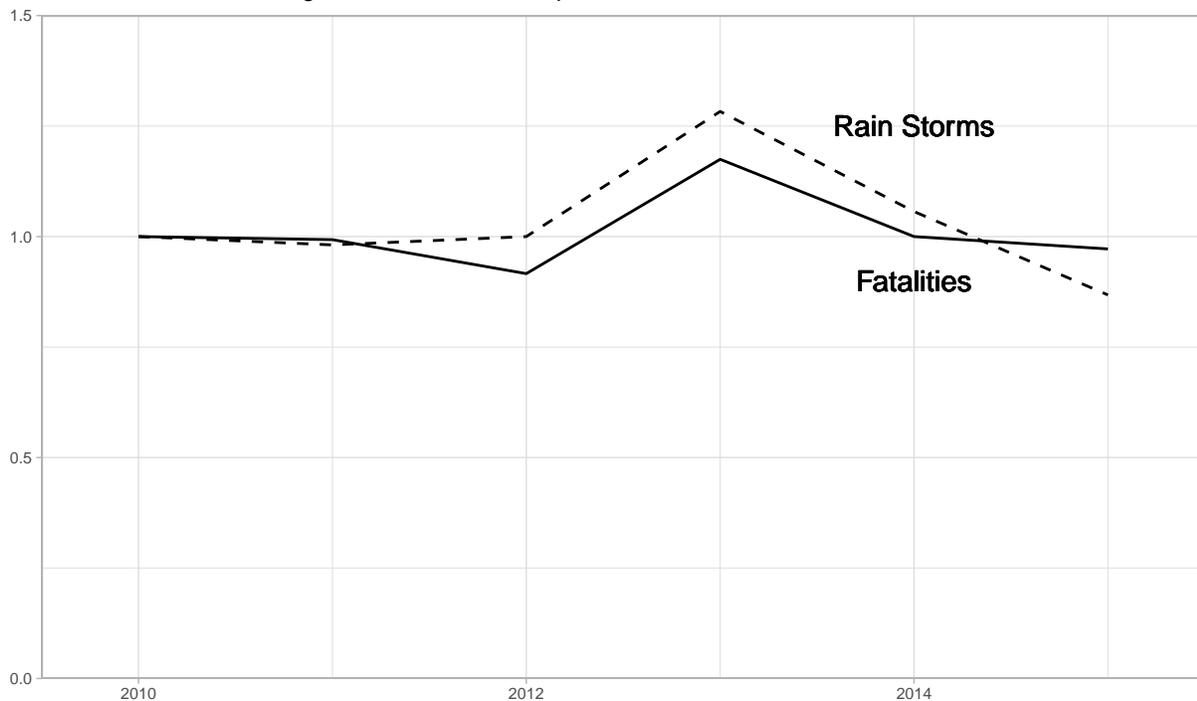
This figure exhibits the annual number of pedestrian deaths in New York City over the study period divided by the number of deaths in 2010 (solid line) and the annual number of rainstorms in New York City over the study period divided by the number of rainstorms in 2010 (dashed line). While Figure 8.1 suggested that fatalities have fallen relative to 2013, Figure 8.2 suggests that, relative to the number of rain storms, fatalities have actually increased in 2015.

We conclude by calculating the effect of New York City's speed limit reduction using the posterior predictions from both models in the previous section. Ignoring the sampling weights, the estimate corresponds to the set of road regions in New York City for which there was a death in 2015. Using the sampling weights, the estimate corresponds to all road regions in New York City. Table 3 shows that under Model 2, the most frequent outcome after lowering the speed limit is coincidentally similar to the before-after estimate of four or six.

```
nyc_pred25 <- pred25[,stan_dat$SLIM[stan_dat$YEAR==6] == 6 &
                       stan_dat$CITY[stan_dat$YEAR==6] == 12]
nyc_pred30 <- pred30[,stan_dat$SLIM[stan_dat$YEAR==6] == 6 &
                       stan_dat$CITY[stan_dat$YEAR==6] == 12]
nyc_pred25_int <- pred_int[,stan_dat_int$SLIM_actual[stan_dat_int$TEST==TRUE] == 6 &
                            stan_dat_int$SLIM[stan_dat_int$TEST==TRUE] == 6 &
                            stan_dat_int$CITY[stan_dat_int$TEST==TRUE] == 12]
nyc_pred30_int <- pred_int[,stan_dat_int$SLIM_actual[stan_dat_int$TEST==TRUE] == 6 &
                            stan_dat_int$SLIM[stan_dat_int$TEST==TRUE] == 7 &
                            stan_dat_int$CITY[stan_dat_int$TEST==TRUE] == 12]
WGHT <- stan_dat$WGHT[stan_dat$SLIM == 6 &
                       stan_dat$CITY == 12 &
                       stan_dat$YEAR == 6]
WGHT_int <- stan_dat_int$WGHT[stan_dat_int$SLIM == 6 &
                               stan_dat_int$SLIM_actual == 6 &
                               stan_dat_int$CITY == 12 &
                               stan_dat_int$TEST == TRUE]

results <- rbind(
  round(quantile(apply( nyc_pred25 - nyc_pred30, 1, sum),
                 c(.025,.25,.5,.75,.975)),2),
  round(quantile(apply( nyc_pred25 - nyc_pred30, 1,
                 function(x,w) sum(x*w), w = WGHT),
                 c(.025,.25,.5,.75,.975)),2),
  round(quantile(apply(nyc_pred25_int - nyc_pred30_int, 1, sum),
                 c(.025,.25,.5,.75,.975)),2),
  round(quantile(apply(nyc_pred25_int - nyc_pred30_int, 1,
                  function(x,w) sum(x*w), w = WGHT_int),
                 c(.025,.25,.5,.75,.975)),2)
  )

rownames(results) <- c("Model 1", "Model 1 (weighted)","Model 2","Model 2 (weighted)")
kable(results, caption="Estimated Number of Deaths due to Lowering Speed Limit
      from 30 to 25 MPH on 25 MPH roads in NYC")
```

Table 3: Estimated Number of Deaths due to Lowering Speed Limit from 30 to 25 MPH on 25 MPH roads in NYC

	2.5%	25%	50%	75%	97.5%
Model 1	-8.00	-3.00	0.00	3.00	9.00
Model 1 (weighted)	-20.50	-6.64	0.00	6.98	21.39
Model 2	-23.00	-11.00	-3.00	4.00	23.00
Model 2 (weighted)	-55.49	-26.04	-8.35	10.07	56.54

The estimated probability of witnessing an increase in deaths after the policy change is 0.37 percent. We conclude that while the policy appears beneficial, there is too much uncertainty in the parameter estimates

and thus too much variation in the predictive distribution, for changes in the number of pedestrian deaths each year to be meaningfully attributed to policy.

IV. Appendix

Data Sources

Our primary dataset is the Fatality Analysis Reporting System (FARS) collected by the National Highway Traffic Safety Administration. FARS contains a census of vehicle related deaths in the United States with a large number of covariates detailing the road, vehicle and persons involved. Road refers to the general area in which the death took place. This is generally a road segment or an intersection.

We supplement the FARS data with a few other sources. These sources are not included in the References section but their websites have been linked to here. We obtain sampling weights for each FARS observation using the GES. Missing New York City speed limit data in FARS is obtained from NYC DOT using the shapefiles of speed limits. The pedestrian population on each road was calculated following the Census Bureau using the TPP and PDB. Average annual daily traffic was estimated from HPMS. The TFFC variable is discretized by the average number of vehciles per second.

Stan Code

```
//Model 1
data {
  int<lower=1> N_train;// number of training data points
  int<lower=1> N;        // number of total data points
  int<lower=1> G;        // number of groupings
  int<lower=3> J[G];     // group sizes
  int COND[N];           // index for surface condition/inclement weather
  int CITY[N];           // index for city
  int YEAR[N];           // index for year
  int SLIM[N];           // index for posted speed limit
  int SIGN[N];           // index for signs and signals i.e. school zone/work zone
  int LGHT[N];           // index for light and time
  int BLTE[N];           // index for built environment
  int TFFC[N];           // index for traffic volume
  vector[N] EXPR;        // population exposed
  int count[N];          // number of pedestrian deaths
}
transformed data {
  vector[N] offset;
  offset = log(EXPR);
}
parameters {
  real offset_e;
  vector[J[1]] COND_eta;
  vector[J[2]] CITY_eta;
  vector[J[3]] YEAR_eta;
  vector[J[4]] SLIM_eta;
  vector[J[5]] SIGN_eta;
  vector[J[6]] LGHT_eta;
  vector[J[7]] BLTE_eta;
  vector[J[8]] TFFC_eta;
  vector<lower=0>[G + 1] sds;
```

```
    vector[N_train] cell_eta;
    real mu;
}
transformed parameters {
    vector[J[1]] COND_e;
    vector[J[2]] CITY_e;
    vector[J[3]] YEAR_e;
    vector[J[4]] SLIM_e;
    vector[J[5]] SIGN_e;
    vector[J[6]] LGHT_e;
    vector[J[7]] BLTE_e;
    vector[J[8]] TFFC_e;
    vector[N_train] cell_e;
    vector[N_train] mu_indiv;

    COND_e = sds[1]     * COND_eta;
    CITY_e = sds[2]     * CITY_eta;
    YEAR_e = sds[3]     * YEAR_eta;
    SLIM_e = sds[4]     * SLIM_eta;
    SIGN_e = sds[5]     * SIGN_eta;
    LGHT_e = sds[6]     * LGHT_eta;
    BLTE_e = sds[7]     * BLTE_eta;
    TFFC_e = sds[8]     * TFFC_eta;
    cell_e = sds[G + 1] * cell_eta;

    for(n in 1:N_train)
    mu_indiv[n] = mu + offset_e * offset[n]
                     + COND_e[COND[n]]
                     + CITY_e[CITY[n]]
                     + YEAR_e[YEAR[n]]
                     + SLIM_e[SLIM[n]]
                     + SIGN_e[SIGN[n]]
                     + LGHT_e[LGHT[n]]
                     + BLTE_e[BLTE[n]]
                     + TFFC_e[TFFC[n]]
                     + cell_e[n];
}
model {
    COND_eta ~ normal(0,1);
    CITY_eta ~ normal(0,1);
    YEAR_eta ~ normal(0,1);
    SLIM_eta ~ normal(0,1);
    SIGN_eta ~ normal(0,1);
    LGHT_eta ~ normal(0,1);
    BLTE_eta ~ normal(0,1);
    TFFC_eta ~ normal(0,1);
    cell_eta ~ normal(0,1);
    offset_e ~ normal(0,1);
    sds      ~ normal(0, 1);
    mu       ~ normal(0, 10);

 for(n in 1:N_train){
     target += poisson_log_lpmf(count[n] | mu_indiv[n]);
     target += -log1m_exp(-exp(mu_indiv[n]));
```

```
 }
}
generated quantities {
  real COND_sd;
  real CITY_sd;
  real YEAR_sd;
  real SLIM_sd;
  real SIGN_sd;
  real LGHT_sd;
  real BLTE_sd;
  real TFFC_sd;
  real cell_sd;
  vector[N - N_train] mu_indiv_pred25;
  vector[N - N_train] mu_indiv_pred30;
  vector[N - N_train] cell_e_pred;

  COND_sd = sd(COND_e);
  CITY_sd = sd(CITY_e);
  YEAR_sd = sd(YEAR_e);
  SLIM_sd = sd(SLIM_e);
  SIGN_sd = sd(SIGN_e);
  LGHT_sd = sd(LGHT_e);
  BLTE_sd = sd(BLTE_e);
  TFFC_sd = sd(TFFC_e);
  cell_sd = sd(cell_e);

  for (n in 1:(N - N_train)){
    cell_e_pred[n]     = normal_rng(0, sds[G+1]);
    mu_indiv_pred25[n] = mu + offset_e * offset[N_train + n]
                            + COND_e[COND[N_train + n]]
                            + CITY_e[CITY[N_train + n]]
                            + YEAR_e[YEAR[N_train + n]]
                            + SLIM_e[6]
                            + SIGN_e[SIGN[N_train + n]]
                            + LGHT_e[LGHT[N_train + n]]
                            + BLTE_e[BLTE[N_train + n]]
                            + TFFC_e[TFFC[N_train + n]]
                            + cell_e_pred[n];
    mu_indiv_pred30[n] = mu + offset_e * offset[N_train + n]
                            + COND_e[COND[N_train + n]]
                            + CITY_e[CITY[N_train + n]]
                            + YEAR_e[YEAR[N_train + n]]
                            + SLIM_e[7]
                            + SIGN_e[SIGN[N_train + n]]
                            + LGHT_e[LGHT[N_train + n]]
                            + BLTE_e[BLTE[N_train + n]]
                            + TFFC_e[TFFC[N_train + n]]
                            + cell_e_pred[n];

  }
}

//Model 2
data {
```

```
    int<lower=1> N_train;// number of training data points
    int<lower=1> N;      // number of total data points
    int<lower=1> G;      // number of groupings
    int<lower=3> J[G];   // group sizes
    int COND[N];         // index for surface condition/inclement weather
    int CITY[N];         // index for city
    int SLIM[N];         // index for posted speed limit
    int SIGN[N];         // index for signs and signals i.e. school zone/work zone
    int LGHT[N];         // index for light and time
    int BLTE[N];         // index for built environment
    int CITYxSLIM[N];
    int LGHTxSLIM[N];
    int BLTExSLIM[N];
    int CONDxSLIM[N];
    int LGHTxCOND[N];
    int LGHTxCONDxSLIM[N];
    vector[N] EXPR;      // population exposed
    int count[N];        // number of pedestrian deaths
}
transformed data {
  vector[N] offset;
  offset = log(EXPR);
}
parameters {
  real offset_e;
  vector[J[1]] COND_eta;
  vector[J[2]] CITY_eta;
  vector[J[3]] SLIM_eta;
  vector[J[4]] SIGN_eta;
  vector[J[5]] LGHT_eta;
  vector[J[6]] BLTE_eta;
  vector[J[7]] CITYxSLIM_eta;
  vector[J[8]] LGHTxSLIM_eta;
  vector[J[9]] BLTExSLIM_eta;
  vector[J[10]] CONDxSLIM_eta;
  vector[J[11]] LGHTxCOND_eta;
  vector[J[12]] LGHTxCONDxSLIM_eta;
  vector<lower=0>[G + 1] sds;
  vector[N_train] cell_eta;
  real mu;
}
transformed parameters {
  vector[J[1]]  COND_e;
  vector[J[2]]  CITY_e;
  vector[J[3]]  SLIM_e;
  vector[J[4]]  SIGN_e;
  vector[J[5]]  LGHT_e;
  vector[J[6]]  BLTE_e;
  vector[J[7]]  CITYxSLIM_e;
  vector[J[8]]  LGHTxSLIM_e;
  vector[J[9]]  BLTExSLIM_e;
  vector[J[10]] CONDxSLIM_e;
  vector[J[11]] LGHTxCOND_e;
  vector[J[12]] LGHTxCONDxSLIM_e;
```

```
  vector[N_train]      cell_e;
  vector[N_train]       mu_indiv;


  COND_e            = sds[1]   * COND_eta;
  CITY_e            = sds[2]   * CITY_eta;
  SLIM_e            = sds[3]   * SLIM_eta;
  SIGN_e            = sds[4]   * SIGN_eta;
  LGHT_e            = sds[5]   * LGHT_eta;
  BLTE_e            = sds[6]   * BLTE_eta;
  CITYxSLIM_e       = sds[7]   * CITYxSLIM_eta;
  LGHTxSLIM_e       = sds[8]   * LGHTxSLIM_eta;
  BLTExSLIM_e       = sds[9]   * BLTExSLIM_eta;
  CONDxSLIM_e       = sds[10]  * CONDxSLIM_eta;
  LGHTxCOND_e       = sds[11]  * LGHTxCOND_eta;
  LGHTxCONDxSLIM_e  = sds[12]  * LGHTxCONDxSLIM_eta;
  cell_e            = sds[G+1] * cell_eta;


  for(n in 1:N_train)
  mu_indiv[n] = mu + offset_e * offset[n]
                  + COND_e[COND[n]]
                  + CITY_e[CITY[n]]
                  + SLIM_e[SLIM[n]]
                  + SIGN_e[SIGN[n]]
                  + LGHT_e[LGHT[n]]
                  + BLTE_e[BLTE[n]]
                  + CITYxSLIM_e[CITYxSLIM[n]]
                  + LGHTxSLIM_e[LGHTxSLIM[n]]
                  + BLTExSLIM_e[BLTExSLIM[n]]
                  + CONDxSLIM_e[CONDxSLIM[n]]
                  + LGHTxCOND_e[LGHTxCOND[n]]
                  + LGHTxCONDxSLIM_e[LGHTxCONDxSLIM[n]]
                  + cell_e[n];
}
model {
  COND_eta           ~ normal(0,1);
  CITY_eta           ~ normal(0,1);
  SLIM_eta           ~ normal(0,1);
  SIGN_eta           ~ normal(0,1);
  LGHT_eta           ~ normal(0,1);
  BLTE_eta           ~ normal(0,1);
  CITYxSLIM_eta      ~ normal(0,1);
  LGHTxSLIM_eta      ~ normal(0,1);
  BLTExSLIM_eta      ~ normal(0,1);
  CONDxSLIM_eta      ~ normal(0,1);
  LGHTxCOND_eta      ~ normal(0,1);
  LGHTxCONDxSLIM_eta ~ normal(0,1);
  cell_eta           ~ normal(0,1);
  offset_e           ~ normal(0,1);
  sds                ~ normal(0,1);
  mu                 ~ normal(0,10);

 for (n in 1:N_train){
    target += poisson_log_lpmf(count[n] | mu_indiv[n]);
    target += -log1m_exp(-exp(mu_indiv[n]));
```

```
 }
}
generated quantities {
  real COND_sd;
  real CITY_sd;
  real SLIM_sd;
  real SIGN_sd;
  real LGHT_sd;
  real BLTE_sd;
  real CITYxSLIM_sd;
  real LGHTxSLIM_sd;
  real BLTExSLIM_sd;
  real CONDxSLIM_sd;
  real LGHTxCOND_sd;
  real LGHTxCONDxSLIM_sd;
  real cell_sd;
  vector[N - N_train] mu_indiv_pred;
  vector[N - N_train] cell_e_pred;

  COND_sd           = sd(COND_e);
  CITY_sd           = sd(CITY_e);
  SLIM_sd           = sd(SLIM_e);
  SIGN_sd           = sd(SIGN_e);
  LGHT_sd           = sd(LGHT_e);
  BLTE_sd           = sd(BLTE_e);
  CITYxSLIM_sd      = sd(CITYxSLIM_e);
  LGHTxSLIM_sd      = sd(LGHTxSLIM_e);
  BLTExSLIM_sd      = sd(BLTExSLIM_e);
  CONDxSLIM_sd      = sd(CONDxSLIM_e);
  LGHTxCOND_sd      = sd(LGHTxCOND_e);
  LGHTxCONDxSLIM_sd = sd(LGHTxCONDxSLIM_e);
  cell_sd           = sd(cell_e);

  for (n in 1:(N - N_train)){
    cell_e_pred[n]    = normal_rng(0, sds[G+1]);
    mu_indiv_pred[n] = mu + offset_e * offset[N_train + n]
                          + COND_e[COND[N_train + n]]
                          + CITY_e[CITY[N_train + n]]
                          + SLIM_e[SLIM[N_train + n]]
                          + SIGN_e[SIGN[N_train + n]]
                          + LGHT_e[LGHT[N_train + n]]
                          + BLTE_e[BLTE[N_train + n]]
                          + CITYxSLIM_e[CITYxSLIM[N_train + n]]
                          + LGHTxSLIM_e[LGHTxSLIM[N_train + n]]
                          + BLTExSLIM_e[BLTExSLIM[N_train + n]]
                          + CONDxSLIM_e[CONDxSLIM[N_train + n]]
                          + LGHTxCOND_e[LGHTxCOND[N_train + n]]
                          + LGHTxCONDxSLIM_e[LGHTxCONDxSLIM[N_train + n]]
                          + cell_e_pred[n];
  }
}
```

V. References

Essex, Shinkle, Amanda, and Anne Teigen. 2016. "Transportation Review: Trends in State Speed Legislation."

Gelman, Andrew, and Jennifer Hill. 2006. *Data Analysis Using Regression and Multilevel/Hierarchical Models.* Cambridge University Press.

Gelman, Andrew, John B Carlin, Hal S Stern, and Donald B Rubin. 2014. *Bayesian Data Analysis.* Vol. 2. Chapman & Hall/CRC Boca Raton, FL, USA.

Goodwin, Arthur H, Libby J Thomas, William L Hall, and Mary Ellen Tucker. 2010. "Countermeasures That Work: A Highway Safety Countermeasure Guide for State Highway Safety Offices."

Hauer, Ezra. 2005. "Cause and Effect in Observational Cross-Section Studies on Road Safety." *Unpublished Manuscript.*

———. 2015. *The Art of Regression Modeling in Road Safety.* Vol. 38. Springer.

Imbens, Guido W, and Donald B Rubin. 2015. *Causal Inference in Statistics, Social, and Biomedical Sciences.* Cambridge University Press.

Leaf, William A, and David F Preusser. 1999. *Literature Review on Vehicle Travel Speeds and Pedestrian Injuries.* US Department of Transportation, National Highway Traffic Safety Administration.

Stan Development Team. 2016. *RStan: The R Interface to Stan* (version 2.14.1). http://mc-stan.org.

Tingvall, Claes, and Narelle Haworth. 2000. "Vision Zero: An Ethical Approach to Safety and Mobility." In *6th Ite International Conference Road Safety & Traffic Enforcement: Beyond.* Vol. 1999.