Multilevel Modeling:Final

Katy Koenig

3/17/2020

Code

Question 2

```
# Read in data
lib_visits <- read.csv("lib_crime_station.csv")
vis_yr <- lib_visits %>% group_by(year) %>% summarise(sum_visits = sum(YTD),
    sum_crime = sum(crime_count))
q2_p1 <- gather(vis_yr, measure, value, -year) %>% ggplot(.,
    aes(year, value)) + geom_col(aes(fill = measure)) + expand_limits(y = 0) +
    theme_minimal() + facet_wrap(~measure, scales = "free") +
    ggtitle("Crime Rates Stagnant While Library Visits Fall")

q2_p2 <- ggplot(lib_visits, aes(YTD)) + geom_density() + theme_minimal() +
    ggtitle("KDE of Library Visits") + xlab("Number of Library Visits") +
    ylab("Count")

q2_p3_title <- paste0("More Library Visits in Lower Crime Areas: ",
    "Harold Washington Library Excluded")
wo_hw <- subset(lib_visits, YTD < 1e+06)
q2_p3 <- ggplot(wo_hw, aes(crime_count, YTD)) + geom_point(aes(color = zone,
    size = num_stations, alpha = 0.6)) + facet_wrap(~year) +
    theme_minimal() + labs(y = "library visits") + ggtitle(q2_p3_title)
```

Question 3

Part A

```
stan_summary <- function(fit) {
    as.data.frame(fit) %>% gather("parameter") %>% group_by(parameter) %>%
        summarize(median = median(value), MAD_SD = mad(value))
}
# pooled model
pooled <- stan_glm(YTD ~ crime_count + num_stations, data = lib_visits)
samples_pooled <- as.data.frame(pooled)
```

Part B

```
# unpooled model intercept, pooled regressor, num_stations
unpooled <- stan_glm(YTD ~ num_stations + period - 1, data = lib_visits)
samples_unpooled <- as.data.frame(unpooled)
```

Question 4

Part A

```
# varying intercept model
multi_one <- stan_glmer(YTD ~ (1 | period) + crime_count + num_stations,
    data = lib_visits)
```

```
## Warning: There were 2 divergent transitions after warmup. Increasing adapt_delta above 0.95 may help
## http://mc-stan.org/misc/warnings.html#divergent-transitions-after-warmup
```

```
## Warning: Examine the pairs() plot to diagnose sampling problems
```

```
samples_multi_one <- as.data.frame(multi_one)
```

Part B

```
# let's add another level w/ community area
multi_two_b <- stan_glmer(YTD ~ (1 | community_area) + (1 | period) +
    crime_count + num_stations, data = lib_visits)
```

```
## Warning: Bulk Effective Samples Size (ESS) is too low, indicating posterior means and medians may be
## Running the chains for more iterations may help. See
## http://mc-stan.org/misc/warnings.html#bulk-ess
```

```
samples_multi_two <- as.data.frame(multi_two_b)
```

Question 5

```
# get medians
med_crime <- median(lib_visits$crime_count)
med_stops <- median(lib_visits$num_stations)

multi_one_inter <- apply(samples_multi_one[, 4:8], 2, function(x) median(x +
    samples_multi_one[, 1]))
multi_one_crime <- med_crime * median(samples_multi_one$crime_count)

multi_two_inter <- apply(samples_multi_two[, 65:69], 2, function(x) median(x +
    samples_multi_two[, 1]))
multi_two_crime <- med_crime * median(samples_multi_two$crime_count)

q5_p <- ggplot(lib_visits) + geom_point(aes(num_stations, YTD,
    alpha = 0.5, color = zone)) + geom_abline(aes(intercept = median(samples_pooled$`(Intercept)`),
    slope = median(samples_pooled$num_stations), linetype = "line1")) +
    geom_abline(aes(intercept = median(samples_unpooled$period) *
        period, slope = median(samples_unpooled$num_stations),
        linetype = "line2")) + geom_abline(aes(intercept = multi_one_inter[period] +
    multi_one_crime, slope = median(samples_multi_one$num_stations),
    linetype = "line3")) + geom_abline(aes(intercept = multi_two_crime +
    multi_two_inter[period], slope = median(samples_multi_two$num_stations),
    linetype = "line4")) + facet_wrap(~period) + theme_minimal() +
    scale_linetype_manual(name = "model", values = c(line1 = "solid",
        line2 = "dotted", line3 = "dotdash", line4 = "longdash"),
        labels = c("pooled", "unpooled", "multilevel_one", "multilevel_two")) +
    ggtitle("Comparison of Models by Time Period") + xlab("Number of L Stations") +
    ylab("Number of Library Visits")
```

Question 6

Part A

```
# helper fn from hw4
parameter_summary <- function(samples) {
    samples %>% as_tibble %>% gather("parameter") %>% group_by(parameter) %>%
        summarize(median = median(value), MAD_SD = mad(value))
}

# grab our regression coefs for the 2nd level line
gamma_0 <- median(multi_one$coefficients["(Intercept)"]) + median(multi_one$coefficients["num_stations"
    med_stops
gamma_1 <- median(multi_one$coefficients["crime_count"])

# multilevel model one
intercepts_multi_one <- parameter_summary(samples_multi_one)
pt_ests <- c()
for (i in 2:6) {
    pt_est <- intercepts_multi_one[1, 2] + intercepts_multi_one[i,
        2] + intercepts_multi_one[8, 2] * med_stops + intercepts_multi_one[7,
        2] * med_crime
    pt_ests <- c(pt_ests, pt_est)
}

crime_medians <- lib_visits %>% group_by(period) %>% summarise_at(vars(crime_count),
    funs(median(., na.rm = TRUE)))
```

```
## Warning: funs() is soft deprecated as of dplyr 0.8.0
## please use list() instead
##
##   # Before:
##   funs(name = f(.))
##
##   # After:
##   list(name = ~ f(.))
## This warning is displayed once per session.
```

```
estimates <- data.frame(estimate = unlist(pt_ests), intercepts_multi_one[2:6,
    3], crime_count = crime_medians$crime_count)
q6_p1 <- ggplot(estimates, aes(x = crime_count, y = estimate,
    ymin = estimate - MAD_SD, ymax = estimate + MAD_SD)) + geom_point() +
    geom_errorbar(linetype = "dotted") + theme_minimal() + geom_abline(intercept = gamma_0,
    slope = gamma_1) + ggtitle("Point Level Estimates & Second Level Regression",
    subtitle = "First Multilevel Model") + theme(plot.title = element_text(size = 12))

# multilevel model two
gamma_0_two <- median(multi_two_b$coefficients["(Intercept)"]) +
    median(multi_one$coefficients["num_stations"]) * med_stops
gamma_1_two <- median(multi_two_b$coefficients["crime_count"])
intercepts_multi_two <- parameter_summary(samples_multi_two)


# predictor: period
pt_ests <- c()
for (i in 63:67) {
    pt_est <- intercepts_multi_two[1, 2] + intercepts_multi_two[i,
        2] + intercepts_multi_two[69, 2] * med_stops + intercepts_multi_two[68,
        2] * med_crime
    pt_ests <- c(pt_ests, pt_est)
}
estimates <- data.frame(estimate = unlist(pt_ests), intercepts_multi_two[63:67,
    3], crime_count = crime_medians$crime_count)
title_1 <- "Point Level Estimates & Second Level Regression"
subtitle_1 <- "for Period Group: Second Multilevel Model"
q6_p2 <- ggplot(estimates, aes(x = crime_count, y = estimate,
    ymin = estimate - MAD_SD, ymax = estimate + MAD_SD)) + geom_point() +
    geom_errorbar(linetype = "dotted") + theme_minimal() + geom_abline(intercept = gamma_0_two,
    slope = gamma_1_two) + ggtitle(title_1, subtitle = subtitle_1) +
    theme(plot.title = element_text(size = 12))


# predictor: zone coefs for reg line
gamma_0_two_b <- median(multi_two_b$coefficients["(Intercept)"]) +
    median(multi_one$coefficients["crime_count"]) * med_crime
gamma_1_two_b <- median(multi_two_b$coefficients["num_stations"])

pt_ests <- c()
for (i in 2:62) {
    pt_est <- intercepts_multi_two[1, 2] + intercepts_multi_two[i,
        2] + intercepts_multi_two[69, 2] * med_stops + intercepts_multi_two[68,
        2] * med_crime
    pt_ests <- c(pt_ests, pt_est)
}

station_medians <- lib_visits %>% group_by(community_area) %>%
    summarise_at(vars(num_stations), funs(median(., na.rm = TRUE)))

estimates <- data.frame(estimate = unlist(pt_ests), intercepts_multi_two[2:62,
    3], num_stations = station_medians$num_stations)
title_2 <- "Point Level Estimates & Second Level Regression"
subtitle_2 <- "for Community Area Group: Second Multilevel Model"

q6_p3 <- ggplot(estimates, aes(x = num_stations, y = estimate,
    ymin = estimate - MAD_SD, ymax = estimate + MAD_SD)) + geom_point(aes(alpha = 0.5)) +
    geom_errorbar(linetype = "dotted") + theme_minimal() + geom_abline(intercept = gamma_0_two_b,
    slope = gamma_1_two_b) + ggtitle(title_2, subtitle = subtitle_2) +
    theme(plot.title = element_text(size = 12))
```

Part B

```
# helper fn from prev hw
lambda <- function(fit) {
    eta <- as.data.frame(fit) %>% dplyr::select(starts_with("b[(Intercept)"))
    numerator <- var(colMeans(eta))
    denominator <- mean(apply(eta, 1, var))
    1 - numerator/denominator
}
multi_models <- list("multilevel_one", "multilevel_two")
```

```
# Bayesian R Squared
model_name <- list("pooled", "unpooled", "multilevel_one", "multilevel_two",
    "multilevel_two_b")
# elpd loo
loo_pooled <- loo(pooled, k_threshold = 0.7)
```

```
## 3 problematic observation(s) found.
## Model will be refit 3 times.
```

```
##
## Fitting model 1 out of 3 (leaving out observation 358)
```

```
##
## Fitting model 2 out of 3 (leaving out observation 360)
```

```
##
## Fitting model 3 out of 3 (leaving out observation 361)
```

```
loo_unpooled <- loo(unpooled, k_threshold = 0.7)
```

```
## All pareto_k estimates below user-specified threshold of 0.7.
## Returning loo object.
```

```
loo_multi_one <- loo(multi_one, k_threshold = 0.7)
```

```
## 5 problematic observation(s) found.
## Model will be refit 5 times.
```

```
##
## Fitting model 1 out of 5 (leaving out observation 357)
```

```
##
## Fitting model 2 out of 5 (leaving out observation 358)
```

```
##
## Fitting model 3 out of 5 (leaving out observation 359)
```

```
##
## Fitting model 4 out of 5 (leaving out observation 360)
```

```
##
## Fitting model 5 out of 5 (leaving out observation 361)
```

```
loo_multi_two_b <- loo(multi_two_b, k_threshold = 0.7)
```

```
## 4 problematic observation(s) found.
## Model will be refit 4 times.
```

```
##
## Fitting model 1 out of 4 (leaving out observation 104)
```

```
##
## Fitting model 2 out of 4 (leaving out observation 106)
##
## Fitting model 3 out of 4 (leaving out observation 357)
##
## Fitting model 4 out of 4 (leaving out observation 361)
```

Responses

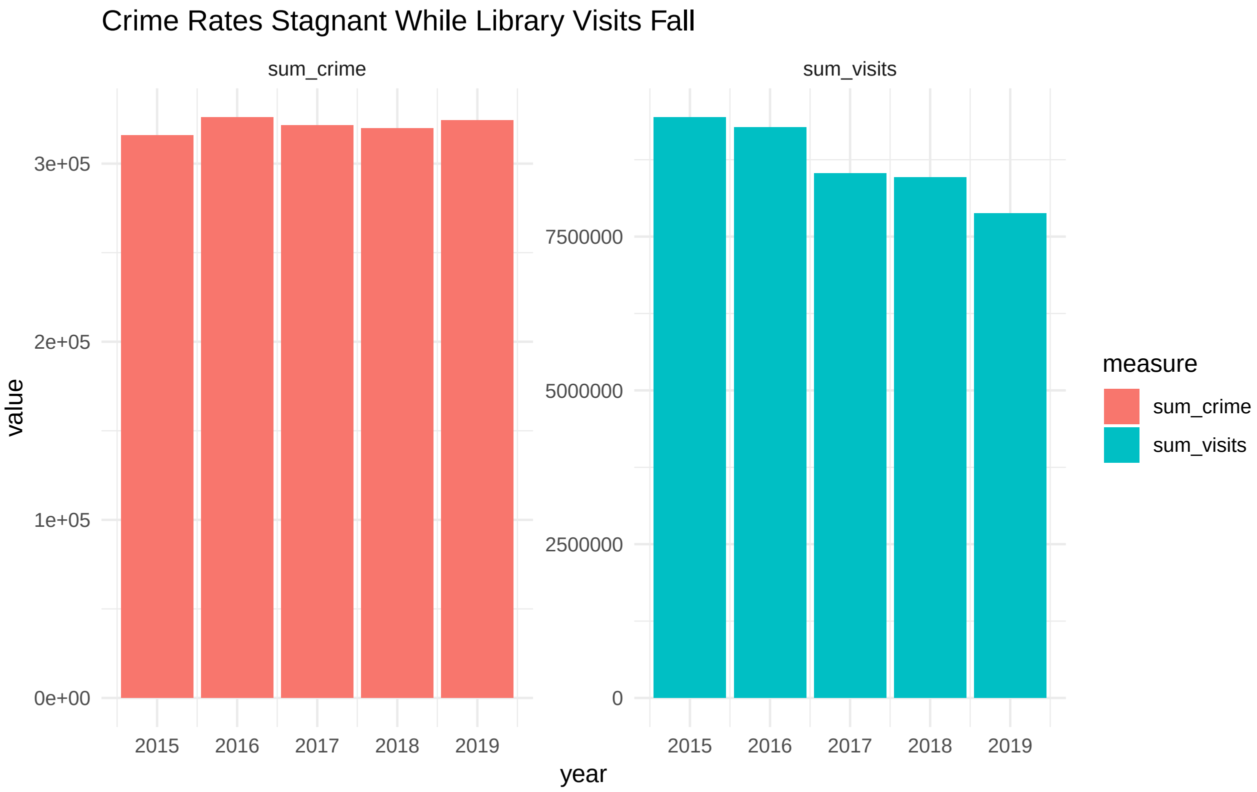
Question 1

In my analysis below, I wanted to analyze the impact of access and safety on use of a public utility. Therefore, I collected data from the City of Chicago data portal regarding library visits (a public utility), reported crime (a proxy for safety) and L stop locations (a proxy for access) for 2015-2019. I aggregated the data by year and community area and merged this information with library visit information, so that each row consists of a library branch, year (group: period), number of library visits that year (outcome variable: YTD), number of crimes in the community area in which the library is located for a given period (group-level predictor: crime_count), and number of L stations in the community area in which the library is located (observation-level predictor: num_stations). For more information regarding the data cleaning process, see this notebook. As I used data from 2015 to 2019, there are five groups in this dataset, with each group having 79-80 observations. (Note: I used reported crime data. Any crimes that occur which are not reported are therefore missing from this dataset. Conversely, if someone falsely reports a crime, it would be included in this dataset.)

I believe a multilevel model will be useful here because we can capture both time-variant aspects, e.g. trends in popularity of buying Kindle books instead of renting books from library, as well as time-invariant aspects, e.g. when something is more easily accessible, more people use it. Additionally, while crime rates and library visits vary year-to-year, each year's rate is not independent of others, therefore, I believe a multilevel model will better capture the effects of perceived safety on resource use. When I enrich my initial multilevel model (in 4b), I add another second layer of community area. I believe a multilevel model will be better than pooled regression due to collinearity issues regarding community area and library locations. I believe a multilevel model will be more informative than an unpooled model as there are not many libraries in each community area.

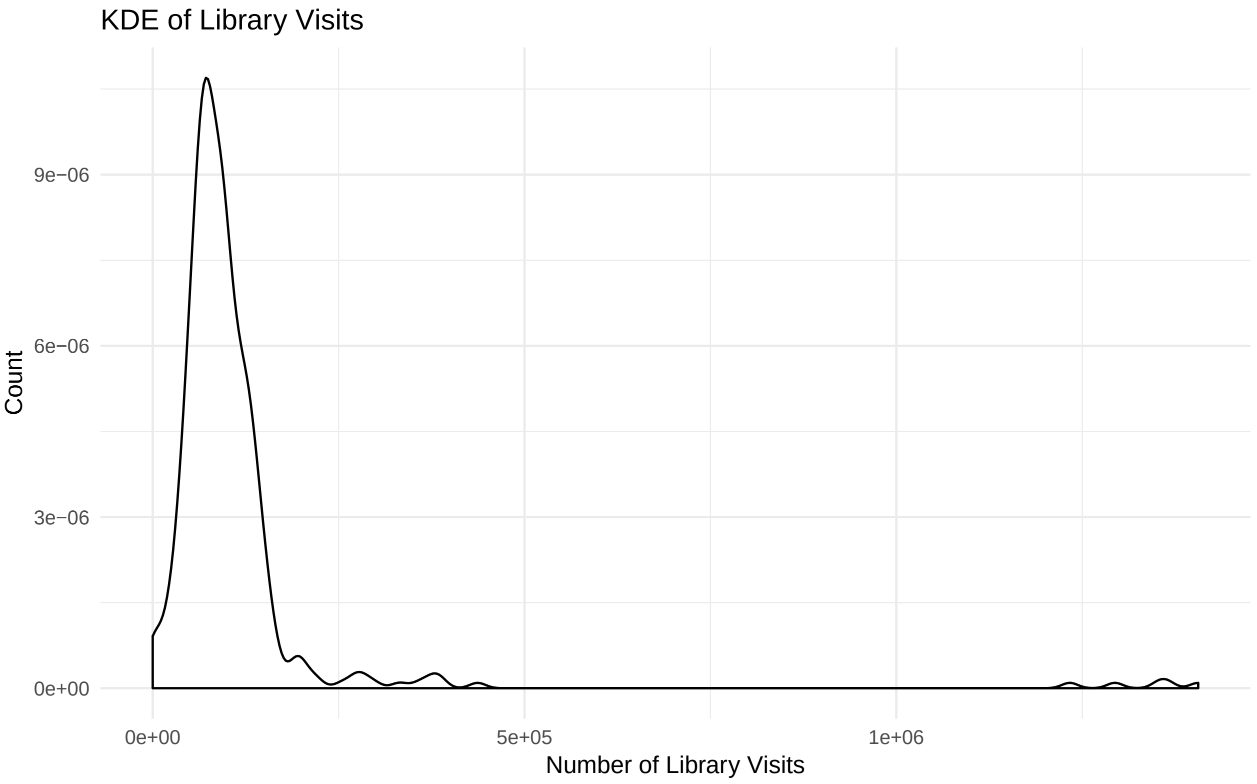
Question 2

```
q2_p1
```



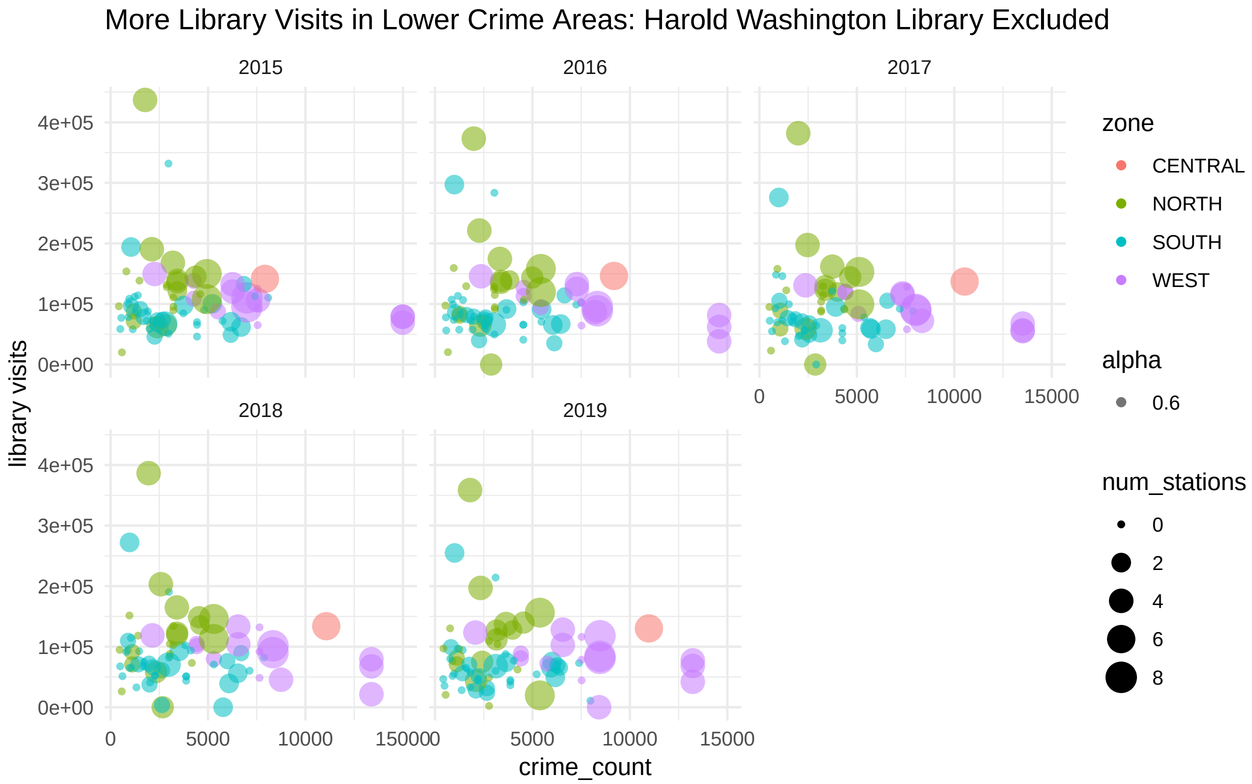
Crime Rates Stagnant While Library Visits Fall

```
q2_p2
```



KDE of Library Visits

```
q2_p3
```



As we can see in our first plot above, library visits overall decrease slightly in our five year period whereas crime rates seem to hold pretty steady. In the second plot, we see that our data is not exactly normal: most libraries have a similiar number of visits, we have a few extreme outliers.

Note: in the third data visualization, Harold Washington Library was not included as it drown out almost all other data points. We also include a 'zone' variable corresponding to the "sides" of Chicago (north/south/west/central), e.g. Hyde Park would be in the south zone whereas the Loop comprised almost all of the central zone. This is only for illustrative purposes and is not used in the analysis above. Additionally, in this plot we see that zones are somewhat related to number of L stations: the south zone has considerably smaller dots on average, meaning there are fewer L stops in that community area. We also see here that crime is somewhat related to library visits: areas with a high amount of library visits also have lower than average crime rates.

Question 3

Part A

Below we present the coefficients for the following pooled model:

$y \sim N(\alpha + \beta_1 x_i + \beta_2 u_i, \sigma_y^2)$

using our variables, we have:

$YTD \sim N(\alpha + \beta_1 num\_stops_i + \beta_2 crime\_count_i, \sigma_y^2)$

```
stan_summary(pooled)
```

```
## # A tibble: 4 x 3
##   parameter      median  MAD_SD
##   <chr>           <dbl>   <dbl>
## 1 (Intercept)   82013.  9585.
## 2 crime_count     -10.8    2.22
## 3 num_stations  41516.  2660.
## 4 sigma        117180.  4049.
```

Our intercept of ~82000 means that we would predict any given library in any given period to have 82000 visits when there are zero L stations and 0 crimes in the area. The coefficient for ~11 for our crime_count variable implies that for each crime committed, we expect to see a decrease in number of library visits of 11, ceteris paribus. A coefficient of ~42,000 for our num_stations variable means that for ever L stations that in a library's community area, we would expect to see an increase in annual library visits of 42,000, ceteris paribus.

Part B

Below we present the coefficients for the following unpooled model: $y \sim N(\alpha_{j[i]} + \beta_1 x_i, \sigma_y^2)$

using our variables, we have:

$YTD \sim N(\alpha_{j[i]} + \beta_1 num\_stops_i, \sigma_y^2)$ where J is our time periods and $j \in J$.

```
stan_summary(unpooled)
```

```
## # A tibble: 3 x 3
##   parameter     median MAD_SD
##   <chr>          <dbl>  <dbl>
## 1 num_stations  38407.  2460.
## 2 period        10857.  2189.
## 3 sigma        123416.  4442.
```

Our coefficient of ~10,500 for our period variable reflects that for each time period, we see an increase of 105,00 library visits, ceteris paribus, while our coefficient of ~38,500 for num_stations means that for each additional L station in a library's community area, we expect to see an increase in library visits of 38,500, ceteris paribus.

Question 4

Part A

Below we present a summary for the following model: $y_i \sim N(\alpha_{j[i]} + \beta x_i, \sigma_y^2)$

Second level: $\alpha_j \sim N(\mu_\alpha, \sigma_{alpha}^2)$

Using our variables our model becomes:

$YTD_i \sim N(\alpha_{j[i]} + \beta num\_stop_i, \sigma_y^2)$

Second level: $\alpha_j \sim N(\mu_\alpha, \sigma_\alpha)$

where J represents our groups, or periods in our analysis, and $j \in J$.

```
# varying intercept model
stan_summary(multi_one)
```

```
## # A tibble: 10 x 3
##    parameter                                        median      MAD_SD
##    <chr>                                             <dbl>       <dbl>
##  1 (Intercept)                                      82261.     11397.
##  2 b[(Intercept) period:1]                   1125.      5347.
##  3 b[(Intercept) period:2]                    866.      5526.
##  4 b[(Intercept) period:3]                   -107.      5390.
##  5 b[(Intercept) period:4]                   -166.      5332.
##  6 b[(Intercept) period:5]                  -2053.      6060.
##  7 crime_count                                -10.8        2.26
##  8 num_stations                             41635.      2667.
##  9 sigma                                   117026.      4149.
## 10 Sigma[period:(Intercept),(Intercept)] 52722741.  73717956.
```

In our summary above, we see the following:

1.) crime_count has a coefficient of approximately -10, meaning that for every crime committed during a given time period, we expect to see a decrease of 10 library visits, ceteris paribus. We also see a coefficient of approximately 41,500 for num_stations, meaning that for every additional L stations in a library's community area, we expect to see an increase in library visits of roughly 41,500 per year, ceteris paribus.

2.) When we view our intercepts for each time period, we see that library visits decrease over time: For example, if we are in the first time period, we expect a library with zero L stations and no crimes in this time period to have approximately 83,000 library visits (where our intercept coefficient is ~82,000 and our period 1 coefficient is ~1,100). In our fifth time period, we expect a library with zero L stations and no crimes in this time period to have approximately 80,000 library visits (where our intercept coefficifient is again ~82,000 and our period 5 coefficient is approximately -2,000).

3.) Above, we also see that there is more variation within groups than between groups as our σ_y is significantly larger than our σ_α. Specifically:

$\sigma_\alpha \approx \sqrt{52,723,000} \approx 7,260$

$\sigma_y \approx 117,000$

Part B

Expanding on the model in Part A: $y_i \sim N(\alpha^{period}_{j[i]} + \alpha^{community\_area}_{k[i]} + \beta * num\_stop_i, \sigma^2_y)$

Second levels:

$\alpha^{period}_j \sim N(\mu^{period}_\alpha, \sigma^2_{period})$

$\alpha^{community\_area}_k \sim N(\mu^{community\_area}_\alpha, \sigma^2_{community\_area})$

where J are our period groups in our analysis and K are our community area groups where $j \in J$ and $k \in K$.

```
# let's add another level w/ community area
head(stan_summary(multi_two_b))
```

```
## # A tibble: 6 x 3
##   parameter                          median MAD_SD
##   <chr>                               <dbl>  <dbl>
## 1 (Intercept)                        46037. 22739.
## 2 b[(Intercept) community_area:1]   -63500. 24081.
## 3 b[(Intercept) community_area:10]  -11097. 19952.
## 4 b[(Intercept) community_area:12]   50302. 24091.
## 5 b[(Intercept) community_area:13]  105536. 23780.
## 6 b[(Intercept) community_area:14] -123049. 23423.
```

```
tail(stan_summary(multi_two_b))
```

```
## # A tibble: 6 x 3
##   parameter                                       median        MAD_SD
##   <chr>                                            <dbl>         <dbl>
## 1 b[(Intercept) period:5]                       -1.02e 4       5209.
## 2 crime_count                                   -2.17e 0          3.78
## 3 num_stations                                   4.67e 4       6657.
## 4 sigma                                          2.67e 4       1032.
## 5 Sigma[community_area:(Intercept),(Intercept)]  1.41e10 2567091431.
## 6 Sigma[period:(Intercept),(Intercept)]          8.92e 7   68771388.
```

1.) crime_count has a coefficient of approximately -2, meaning that for every crime committed during a given time period, we expect to see a decrease of 2 library visits, ceteris paribus. We also see a coefficient of approximately 46,500 for num_stations, meaning that for every additional L stations in a library's community area, we expect to see an increase in library visits of roughly 46,500 per year, ceteris paribus. As compared to our previous model, the number of stations has more influence in predicting library visits while crime count has less influence in our second model.

2.) When we view our intercepts for each time period, we again see that library visits decrease over time although this rate of change has increased as compared to our previous partial pooling model: For example, if we are in the first time period, we expect a library with zero L stations, no crimes in this time period and located in community area 12 to have approximately 103,900 annual (where our intercept coefficient is ~46,000, our period 1 coefficient is ~7,900 and our community area 12 coefficient is approximately 50,000). For the same library in period 5, we expect to 10,000 fewer visits, ceteris paribus.

3.) Additionally, we see vast differences in the coefficients for community area: For example, community area 28 has a coeffcent of approximately -305,000 while community area 19 has a coefficient of ~53,000, meaning that all else equal, we expect to see a library in community 28 have ~358,000 fewer visits than a library in community area 19.

4.) Above, we also see that there is more variation between our community area groups $\sigma_\alpha^{community\_area}$ than either variation between our period groups σ_α^{period} or variation within our groups σ_y. Specifically,

$\sigma_\alpha^{period} \approx \sqrt{89,248,400} \approx 9,400$

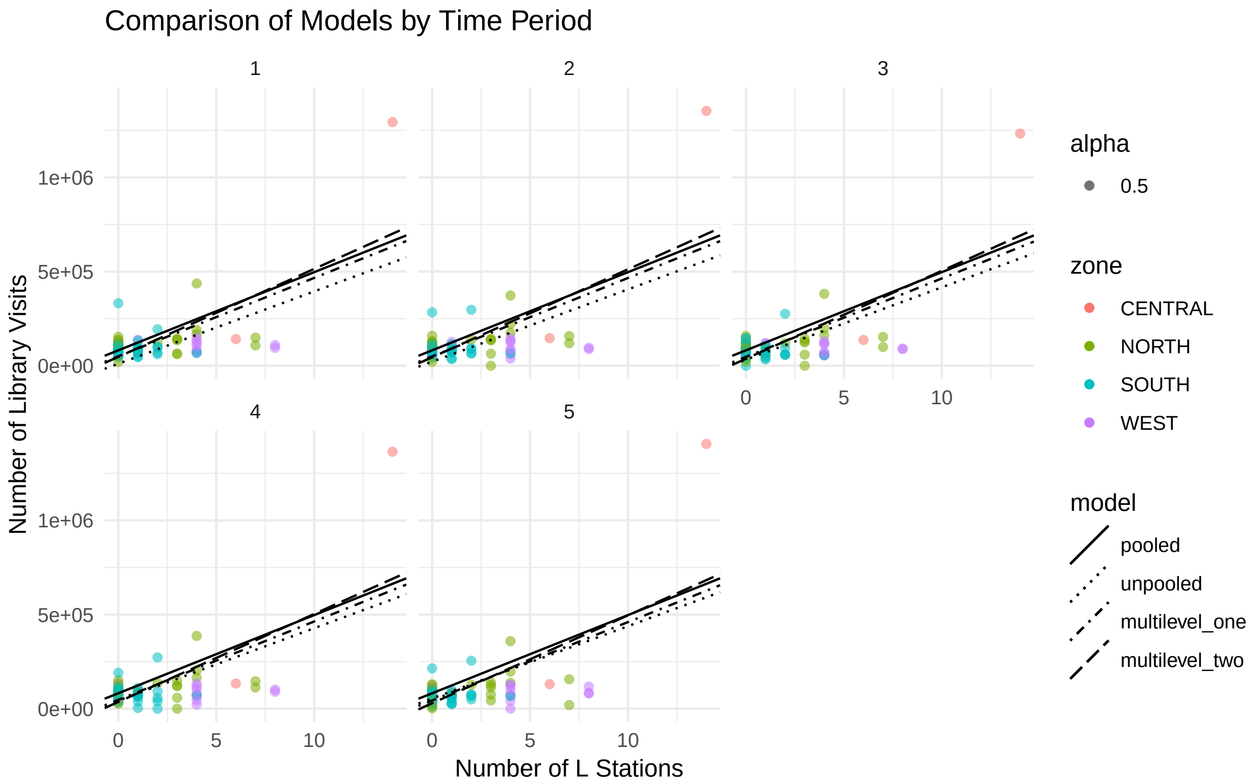
$\sigma_\alpha^{community\_area} \approx \sqrt{14,081,250,000} \approx 118,600$

$\sigma_y \approx 26,600$

In comparison with our first multilevel model, we see that for both regressions variation within groups σ_y is larger than variation between our period groups σ_α^{period}.

Question 5

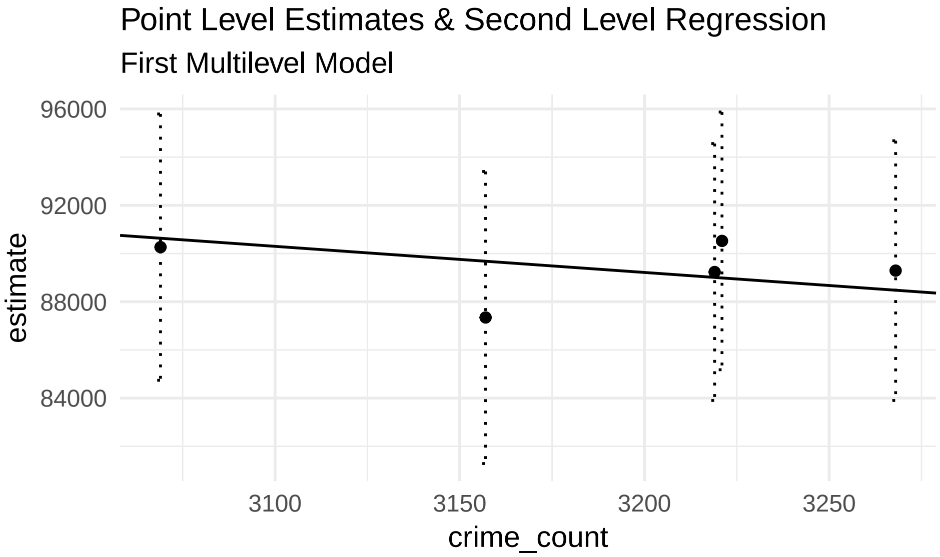
```
q5_p
```



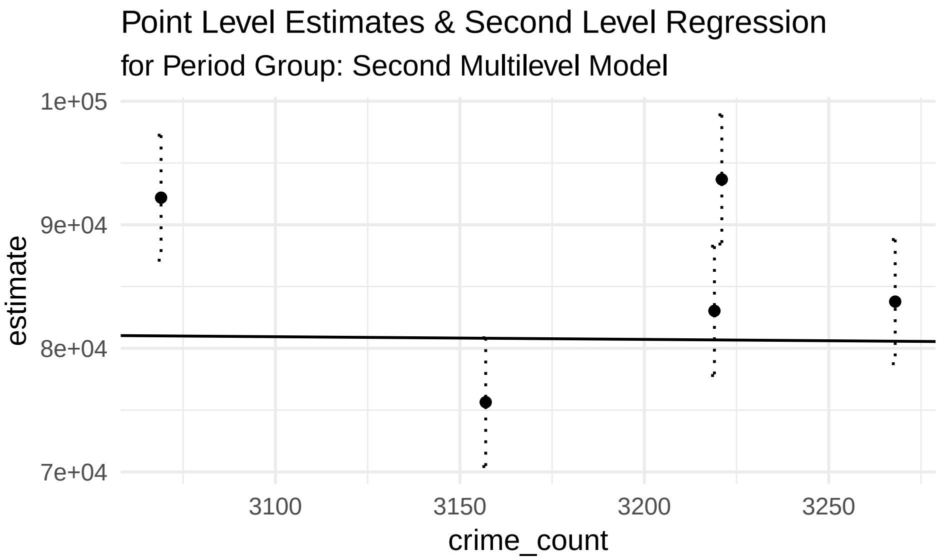
Question 6

Part A

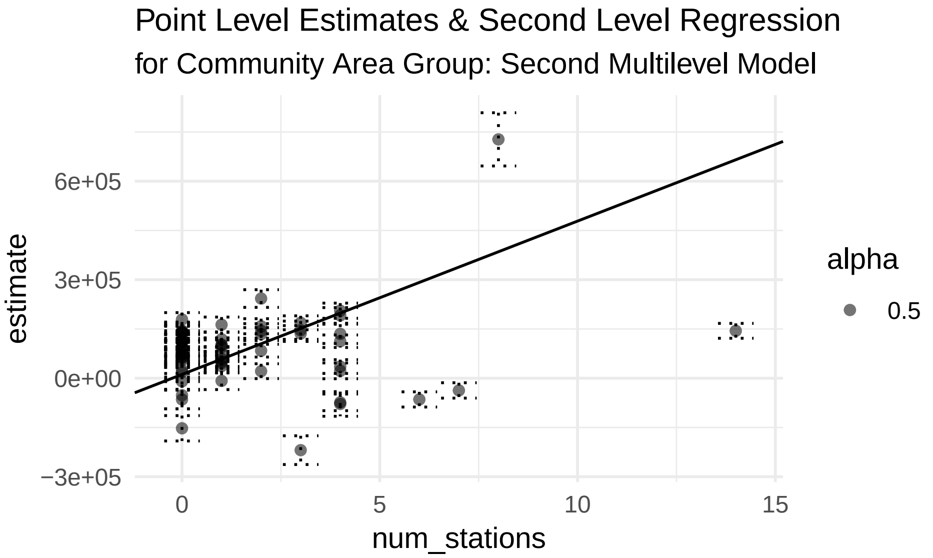
```
q6_p1
```



```
q6_p2
```



```
q6_p3
```



Part B

```
i <- 1
for (model in list(multi_one, multi_two_b)) {
    print(multi_models[i])
    print(lambda(model))
    i = i + 1
}
```

```
## [[1]]
## [1] "multilevel_one"
##
## [1] 0.8752999
## [[1]]
## [1] "multilevel_two"
```

```
##
## [1] 0.0331414
```

As shown above, we see significant pooling in our first multilevel model between time periods, while our second multilevel model has very little pooling between groups.

Question 7

```r
# Bayesian R Squared
i <- 1
for (model in list(pooled, unpooled, multi_one, multi_two_b)) {
    print(model_name[i])
    print(quantile(bayes_R2(model), c(0.25, 0.5, 0.75)))
    i = i + 1
}
```

```
## [[1]]
## [1] "pooled"
##
##       25%       50%       75%
## 0.3636258 0.3869482 0.4091120
## [[1]]
## [1] "unpooled"
##
##       25%       50%       75%
## 0.3610157 0.3835865 0.4052534
## [[1]]
## [1] "multilevel_one"
##
##       25%       50%       75%
## 0.3662819 0.3891600 0.4118540
## [[1]]
## [1] "multilevel_two"
##
##       25%       50%       75%
## 0.9663004 0.9680615 0.9696897
```

```r
# elpd loo
loo_compare(loo_pooled, loo_unpooled, loo_multi_one, loo_multi_two_b)
```

```
##             elpd_diff se_diff
## multi_two_b    0.0       0.0
## pooled      -568.0      61.0
## multi_one   -568.9      61.2
## unpooled    -583.9      56.2
```

Looking at our Bayesian R^2 values for each of our models, our second multilevel model significantly better than the Bayesian R^2 values for our other models, demonstrating that most of the variance in our data is explained by this model and therefore, implying that a fair amount of variance can be explained by community area/location.

When we compare our leave-one-out log scores for our models, we see that our second multilevel model performs significantly better than our pooled, unpooled and first multilevel models. It is also interesting to note that our pooled model performs better than our first multilevel model.

Question 8

Above, we can see both benefits and issues using multilevel regression: our first multilevel model performed more poorly than our pooled model as evident by the leave-one-out analysis. Although all of our first three models (pooled, unpooled and multilevel model with only period at the second level) did not perform well as evident by the Bayesian R^2 values. I believe the autoregressive period second level might have performed better had there been many more time periods: because there were only five time periods, the model pooled its estimates dramatically.

Our second multilevel model outperformed our previous models significantly as we can see in our leave-one-out analysis and our Bayesian R^2 comparison. We also see much smaller error bars when plotting our second level regression line of community area than we do for our second level regression line for periods in our second multilevel model as well as much smaller error bars than when we plot our second level regression line of our first multilevel model, representing that there are significantly smaller standard errors (and therefore less uncertainty) in the coefficients for community area in our second multilevel model. It is important to note that we do see a decrease in standard errors for time period from our first multilevel model to our second as well, and in the same plot, we can visually see the less pooled results in our second model than our first, as confirmed when we calculate the pooling factor, λ in question 6, part b. These results reflect that library visits is very much a function of community area, which is both unsurprising given Chicago's nickname, "the City of Neighborhoods", and the City's historical segregation/redlining policies and subsequent distribution of resources.

Generally, we do see that crime has a negative impact on library visits, although the standard error in our multilevel model is quite large, so the impact may be less strong than the relationship between very positive relationship between number of L stations and library visits. It is important to note the limitations and desired improvements of this analysis. There are obvious improvements for this analysis as evident by the unrealistic predictions of negative library visits for some community areas. Therefore, it may be better to use a log-linear model instead of the entirely linear models we used in this analysis. Additionally, it may be beneficial to add more information regarding the library itself, e.g. the resources available, as regressors but unfortunately, that data was not readily available. Finally, while our second multilevel model with varying intercepts for community area and time period performed significantly better than our pooled and unpooled models, it could be useful to compare models with varying intercepts and varying slopes or only varying slopes to appropriately model library visits.