Homelessness Response Times in Salt Lake City

Skyler Gray

December 9, 2021

1 Introduction

Homelessness is a perpetual challenge in Salt Lake City. The city government has a protocol for investigating suspicious activity, disbanding homeless communities, and referring individuals to homeless resources in the area. One way city residents and visitors can support these measures is through reporting concerns of homelessness directly to the city government via the web or the SLC Mobile app.

I am interested in investigating the city's recent efforts to resolve concerns of homelessness in the last two years. Specifically, I want to know whether there is a difference in response time among the city's seven city council districts. As a secondary objective, I want to know how the prevalence of COVID-19 affects response time. To investigate this potential difference in performance and the effects of COVID-19 on response time, I propose a Bayesian Poisson regression model to estimate the average number of days needed for the city to respond to and close a homelessness service request.

2 The Data

The data on service requests within Salt Lake City are publicly available through the State of Utah's web page (https://opendata.utah.gov/Government-and-Taxes/Service-Request-SLCMobile/yga5-qpeq/data). Each service request has a creation date, closed date, request type, and GPS coordinates. The data was first filtered down to "Concern of Homelessness" requests between January 1, 2020 and October 31, 2021. Because of both an unrealistic number of cases closed on a given day or the short lag behind closing another case, the requests data was further filtered from about 5,000 to around 1,500 requests (see Figure 1). Those removed included days with five or more requests closed within three minutes of each other as well as days with number of cases closed above the 90th percentile.

COVID-19 new case data for Salt Lake county was obtained from the Center of Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU) (https://github.com/CSSEGISandData/COVID-19/blob/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_confirmed_US.csv). Informa-

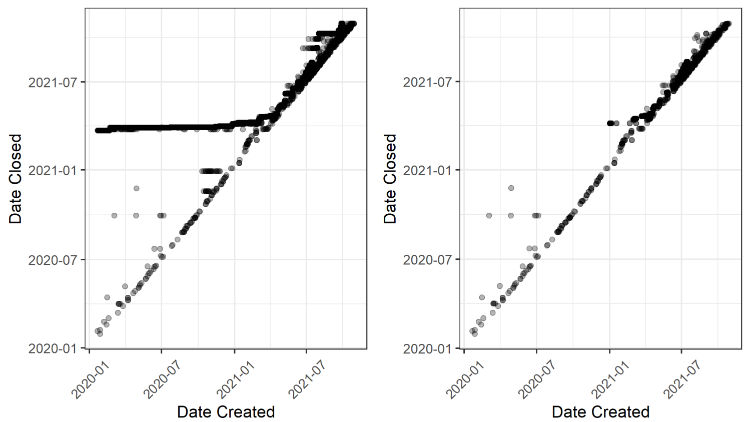

Figure 1: The unfiltered concern of homelessness request data (left) and the filtered data used in the analysis (right)

Variable	**Description**
District	City council district of the request GPS coordinates
New COVID Cases	The 7-day rolling average of new COVID-19 cases in Salt Lake county as of request creation date
Density	Estimated 2019 population density (people/mi^2) of the relevant 2010 Census tract matching the GPS coordinates
Income	Estimated 2019 median household income of the Census tract that the GPS coordinates were located in
Days Open	Number of days until the request was closed; rounded to nearest 24-hour period

Table 1: Brief descriptions of the data to be used in model development

tion on the 2010 Census tract that a request belonged to was evaluated using the GIS data available through the state government's website (https://gis.utah.gov/data/demographic/census/). Lastly, 2019 tract estimates of population density and median income were pulled from the 2020 Salt Lake City Data Book (https://www.slc.gov/hand/wp-content/uploads/sites/12/2020/10/SLC-Data-Book-2020forWeb.pdf).

Figure 3 shows that the response time to any given request is right-skewed and takes values between 0 and 209.

3 EDA

Table 2 and Figure 4 provide insight into the response time across districts. The mean response times are in the range of 14 to 16 days except for district 6. However, we do not have nearly as large of a sample in

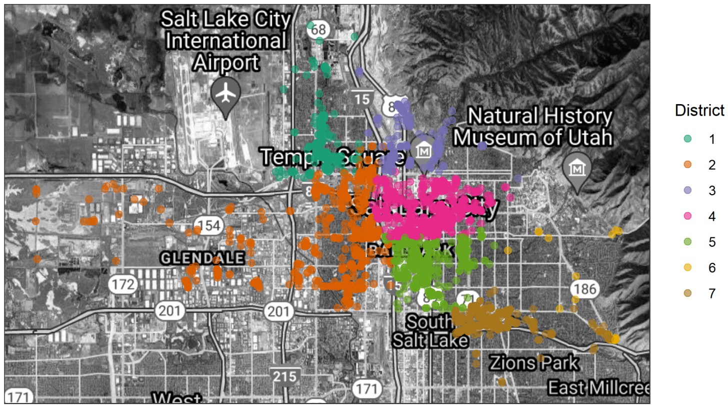

Figure 2: Reports of homelessness concerns since January 1, 2021. Each observation is colored by its respective city council district.

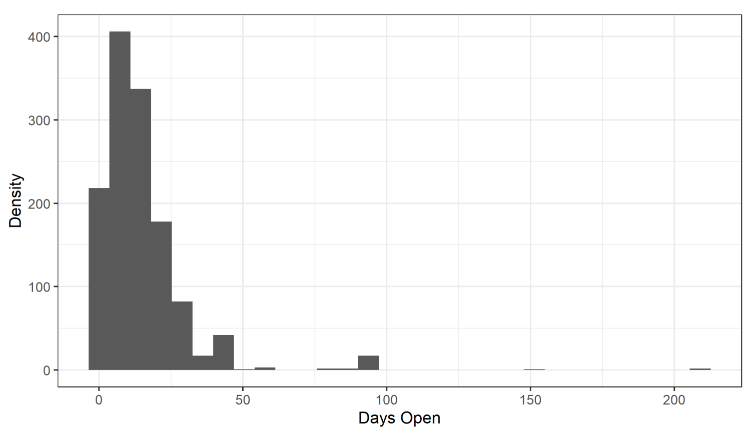

Figure 3: The distribution of days until a report of concern of homelessness is closed.

district 6 as we do in all other districts. The boxplots of Figure 4 suggest that, on the log scale, all districts except district 6 appear to be similarly distributed.

District	n	Days Open
1	159	14.66
2	327	14.13
3	56	16.46
4	360	15.47
5	250	12.56
6	6	30.50
7	150	14.07

Table 2: Sample sizes and mean response time across districts

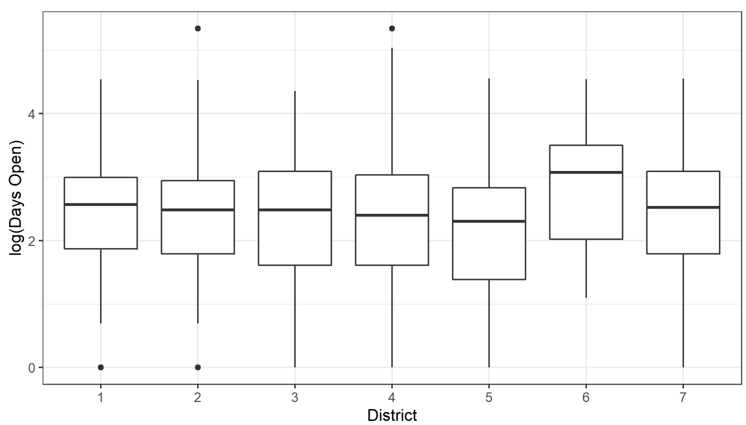

Figure 4: Visualization of the distribution of the log days open across districts. Only district 6 appears to have a different distribution than the other districts, but district 6 has significantly less observations behind its distribution than all the others.

Multiple variables will be included in the model as potential predictors of average response time. These three variables are the 7-day rolling average of new covid cases, the log-scaled median household income, and population density. The independent effects of these variables on log(Days Open) are explored in Figure 5.

4 Methods

4.1 Model: Poisson Regression

To understand the district effect on mean response time after accounting for other potential confounding variables, I propose a Bayesian Poisson regression model with the logit link. Let Y_i be the number of days until the ith request concerning homelessness in Salt Lake City is closed. Additionally, let μ_i be the average number of days until a request is closed for the ith and let X_i be a 1×10 matrix of the covariates for request i and let β be a 10×1 matrix with coefficients for the intercept, districts 2 through 7, the 7-day rolling

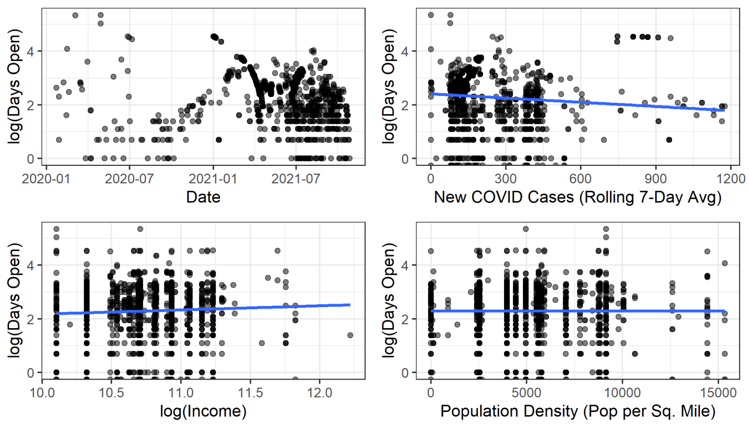

Figure 5: Exploratory plots of potential predictors against the log number of days a report is open.

average of new COVID cases in Salt Lake county, median household income, and population density. Then I propose the following model for response time:

$$Y_i|\mu_i \overset{ind}{\sim} Pois(\mu_i),$$

$$log(\mu_i) = X\beta,$$

$$\beta_1, ..., \beta_{10} \sim N(\mu_0, \tau_0^2),$$

where the β coefficients have diffuse priors centered on 0. My initial choice of a prior is $\mu_0 = 0$ and $\tau_0^2 = 1/100$ because I am not sure whether an effect is positive or negative for most β estimates and want the data to do the estimating.

I sampled from the posterior distribution using two different methods. First, I sampled from the posterior distribution via a univariate slice sampler for each β. Second, I used the probabilistic programming language (PPL) Stan to sample from the posterior distribution. In order to stabilize sampling from the posterior, I centered and scaled the three continuous variables before sampling from the posterior. After, I rescaled the coefficients again. So β_8 will be interpreted as the effect of an increase in Census tract population by 1,000 people/mi^2 on the log mean response time, holding all other variables constant. β_9 is interpreted as the effect of an increase of log median household income on log mean response time, and β_{10} is the effect of an increase in the 7-day rolling average of new COVID-19 cases on log mean response time.

4.2 Diagnostics

Table 3 shows the convergence diagnostics of both the slice sampler and Stan approaches to sampling from the posterior distribution. I could sample more draws using the slice sampler to potentially reduce the $\hat{R}$ calculations, but for the sake of saving myself time, I am going to just use the much larger sample from Stan.

	Estimate		**ESS**		$\hat{R}$	
Coefficient	**Slice**	**Stan**	**Slice**	**Stan**	**Slice**	**Stan**
β_1	2.671	2.671	2252	21441	1.52	1
β_2	-0.013	-0.013	2757	25318	1.33	1
β_3	0.108	0.107	4714	35126	1.48	1
β_4	0.062	0.062	2695	25432	1.47	1
β_5	-0.149	-0.149	3133	28564	1.52	1
β_6	0.731	0.731	7986	45810	1.52	1
β_7	-0.037	-0.036	3532	30704	1.13	1
β_8	0.0085	0.0085	8294	50664	1.5	1
β_9	0.0035	0.0032	7701	43888	1.52	1
β_{10}	0.0005	0.0005	9000	65405	1.52	1

Table 3: Both samplers produced the same coefficient estimates up to Monte Carlo error, but Stan yields a much larger effective sample size and lower $\hat{R}$ than my slice sampler and in a fraction of the time. The high $\hat{R}$ values for the slice sampler suggest that the chains have not converged yet, which may be due to the relatively small sample size.

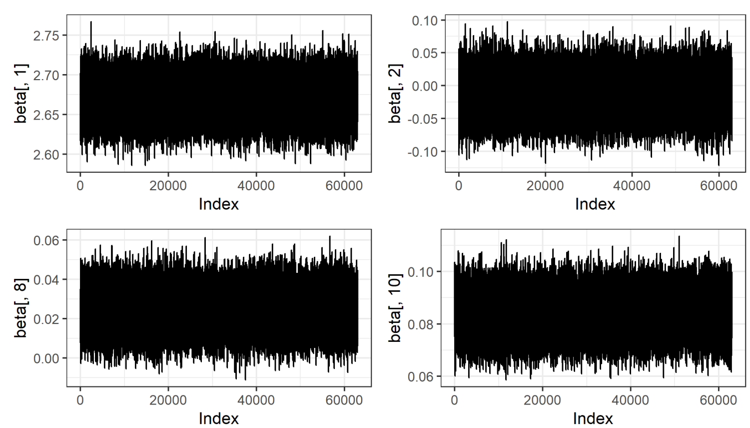

Figure 6: Trace plots for posterior draws via Stan.

4.3 Sensitivity Analysis

Due to the model structure's simplicity, the sensitivity analysis I could investigate only pertains to giving the β coefficients either stronger or weaker priors centered in different locations. Given the intercept β_1 is highly correlated with all other β_j, creating an unrealistically strong prior will dramatically influence our β estimates. Given my lack of necessary time to explore it, I will let this commentary on a sensitivity analysis suffice. Because of the large sample size we have, we can expect the β coefficients to be fairly robust against prior center as well as varying precision levels.

5 Results

The primary question behind this analysis is whether, after accounting for several potential confounding variables, there is an apparent difference in city government response time to homelessness concerns among Salt Lake City's seven city council districts. Figure 7 shows us that, even after accounting for uncertainty, a request from every district except district 6 should expect to be resolved within about 13 to 17 days. I would be interested to see the city government's input on these observed differences, as an imbalance in request investigation could mean neglecting larger needs in districts that may be overlooked, which could be districts 3 and 4.

Again, because district 6 only has 6 observations behind its estimated effects, I believe we do not have enough information to really observe any useful pattern as compared to all other districts.

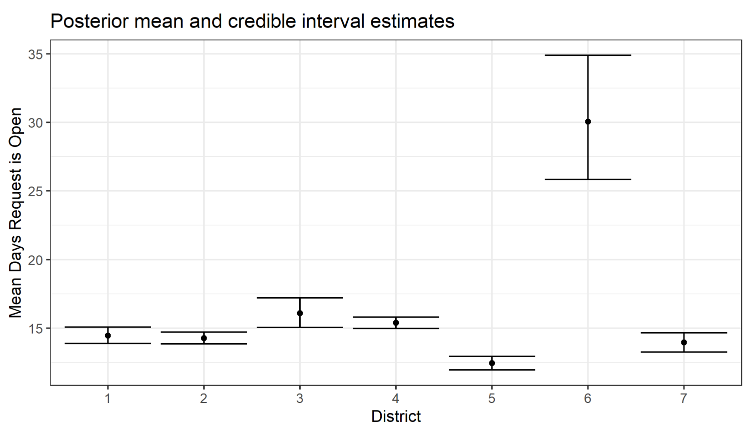

Figure 7: Posterior estimates and 95% credible intervals of mean days a homelessness concern request is open.

The secondary goal of this analysis is to understand the relationship between COVID-19 prevalence and average response time to concerns of homelessness. The 95% credible interval for the 7-day rolling average of new COVID cases is (0.0004, 0.0005). Because the data for this analysis is not a random sample, we cannot conclude causation, but there is a positive correlation between the rolling average of new COVID cases and the average response time for a request.

Coefficient	Estimate	Lower	Upper
Density	0.0085	0.0027	0.0142
Log(Income)	0.0032	-0.0598	0.0657
New Cases	0.0005	0.0004	0.0005

Table 4: Posterior mean and 95% credible intervals.

6 Conclusion

I have taken an initial dive into Salt Lake City's publicly available data on requests submitted to the city's Department of Health. It is interesting to note that there may be some discrepancies in equitable resource allocation to resolve homelessness concerns among Salt Lake City's seven city council districts, but it is hard to call anything a true concern without speaking to a domain expert. Additionally, the model follows one's intuition that a dramatic rise in new COVID-19 cases should slow down the city government's efforts to resolve problems that involve increased contact with others.

There are so many ways this project could develop down the road. For one thing, I have not made contact with the right person within the Salt Lake City Department of Health. Making contact with the department and learning more about the investigation process of these reports will be crucial to performing some helpful analysis.

Additionally, there are several routes of improvement open to this model. First, I believe this first model does not reflect the uncertainty behind the mean effects of district 6 on response time. The next step to approaching this may be to develop a hierarchical model with an unknown precision parameter for each β_j. In addition to more accurately reflecting our uncertainty in the model, we may consider checking for overdispersion by modeling the random error behind response time as a negative binomial distribution rather than as Poisson error.