

Mini-Project 4

Stat 212: Interim 2021

Marco Ruiz, Ivana Kramarevsky, Karen Jensen

Introduction 12/15

The data we will be using comes from the MPD Use Of Force data set, a Minneapolis government data set on police records involving use of police force from 2008 to 2020, along with accompanying information. It is meant to represent all use of police force in metropolitan areas. Instead of focusing on the use of force itself, we would like to know the following:

- 1. Do all Minneapolis districts have the same average calls to 911?
- 2. Is there a linear relationship between the total number of 911 calls neighborhoods make and the year?

A 2019 Star Tribune article titled “Big gaps in 911 response in Minneapolis” explored the relationship between police response time to high-priority 911 calls and the amount of calls from different districts in Minneapolis and St. Paul area. They found that there is a very short response time to the areas that have a higher number of high-priority 911 calls. They provide a map of the Twin-Cities and show that the center of Minneapolis (downtown area) experience low response rates, while the very south of Minneapolis experience high response rates. From this information, we expect to see the downtown area with significantly higher total 911 call rates and the south of Minneapolis with significantly lower total 911 call rates. The article only reported call rates for 2017, so we have no reference for the total call rates from 2008 to 2020.

Methods 23/25

Each time a Minneapolis police officer uses force against a civilian, it is recorded into the database with accompanying information such as the time (the year), location (neighborhood), and general 911 call information (the number of calls made to 911 in the same location). Since there were too many neighborhoods to explore, we grouped them into the eleven official Minneapolis districts and removed the instances of police force used in neighborhoods that fell outside of those districts (less than 600 instances). Finally, we added up the total number of 911 calls for each neighborhood (per year) based on the district they were in to give the total number of 911 calls that the district makes (per year).

For our first research question, our null hypothesis is: all districts have the same average 911 call rate per year. Our alternative hypothesis is: there is at least one district that has an average call rate that is different from the other districts per year. The explanatory variable is the districts, while the response variable is the total number of calls a district makes. We will compare means using an ANOVA test (and Tukey’s Honest Significant Difference test if we find at least one mean that is significantly different from the rest).

For our second research question, our null hypothesis is: there is no linear relationship between the number of 911 calls made by neighborhoods and the year (or $\beta_1 = 0$ in $\hat{Calls} = \beta_0 + \beta_1 * Year$). Our alternative relationship is: there is a linear relationship between the total number of calls made and the year ($\beta_1 \neq 0$). The explanatory variable is the year, while the response variable is the total number of calls a neighborhood makes. We will explore the relationship using a linear regression model.

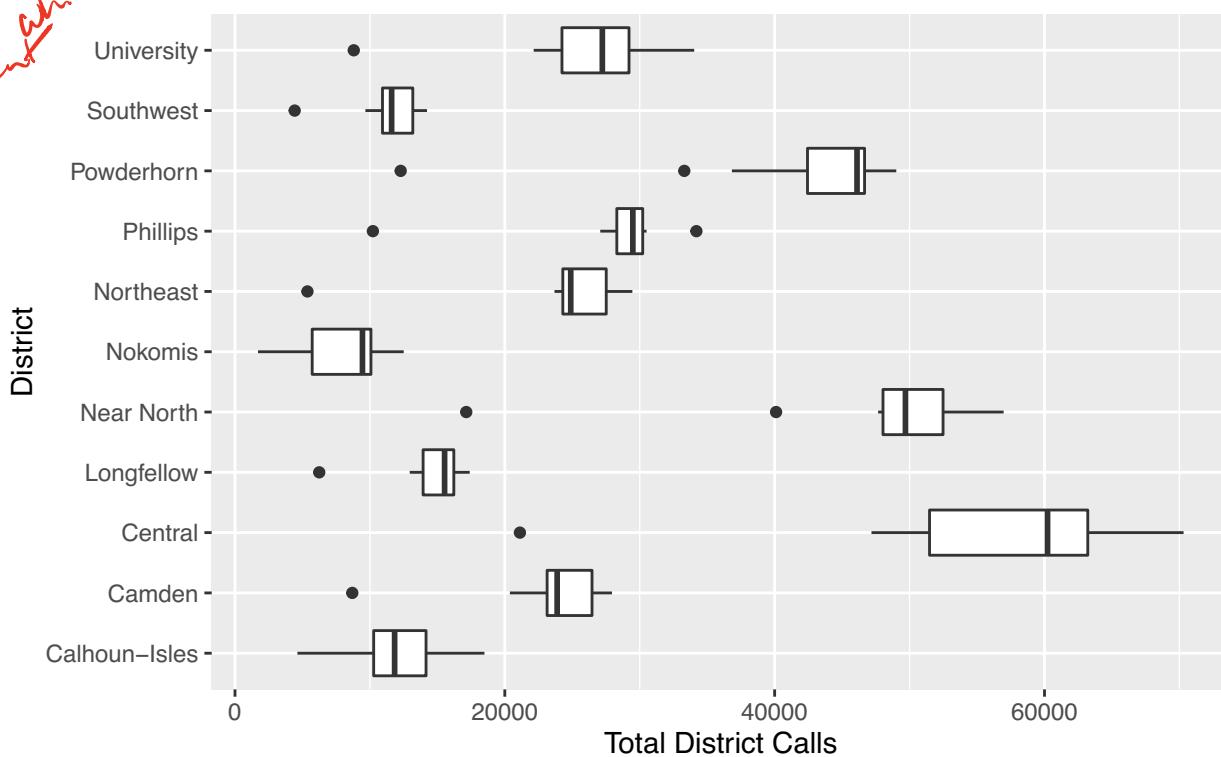
Results 22/25

District 911 calls

Figure 1 shows visibly different median calls made to 911 for most districts. Figure 1 also shows odd distributions. We can see approximately normal distributions in Nokomis and Calhoun-Isles but the rest have significant outliers or strong skews (which is verified with histograms representing the same information). This may be because there are only 12 or less years that are being plotted. Variability is roughly equal throughout the districts. Assuming that one district can't raise another district's call count, the conditions are good enough to perform an ANOVA and Tukey Honest Significant Difference test, though we should be wary. We see an F value of 67.96 and a p-value of about zero in our test results. We have strong evidence to reject the null hypothesis. The following pairings of districts have no significant differences in the mean 911 calls made per year (p-value is less than 5%), but are significantly different across pairings:

- I get the overall ranking for listing the pair, but it still seems counterintuitive which are different which
- 1. Longfellow, Nokomis, Southwest, and Calhoun-Isles (seen clustered in Figure 1 for less than 20,000 calls)
- 2. Northeast, Phillips, University, and Camden (seen clustered in Figure 1 for more than 20,000 calls and less than 40,000 calls)
- 3. Near North and Powderhorn (seen clustered in Figure 1 for more than 40,000 calls)
- 4. Near North and Central (seen clustered in Figure 1 for more than 40,000 calls)

Figure 1.
Total 911 Calls Minneapolis Districts Make Per Year



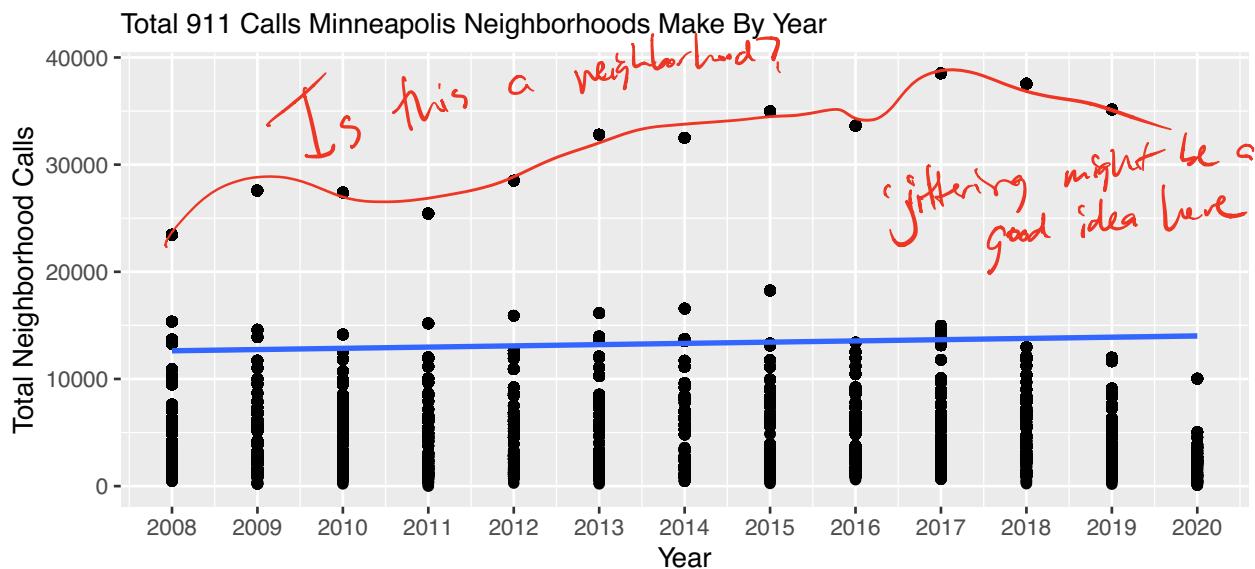
Neighborhood 911 calls

Should, rand

The correlation between the year and the number of neighborhood calls is weak ($r = 0.03758689$). Figure 2 shows that the data points has, very vaguely, a positive direction with neighborhoods making more calls as the year increases (with the exception of 2020). This plot helps explain the low correlation value. The equation of the best fitted line shown in Figure 2 is $\hat{Calls} = -218830.53 + 115.27 * Year$. We are 95% confident that the true slope is between 79.84 and 150.69. The t-value for the linear regression model is 6.377 with a p-value

near 0, meaning the slope is significant such that *if* there is a linear trend between the data points, then this is this linear relationship. However, the R-squared value is 0.001378, which indicates that almost none of the variability in the plot can be explained by the model.

Figure 2.



Next, we look at the conditions for using the linear regression model, especially because of our R-squared value. The Residuals vs. Fitted plot (see appendix) indicates that the residuals may have non-constant variance. Our Normal Q-Q plot shows a strong S-shape, indicating that the residuals are not normally distributed (see appendix). In fact, the residuals have a bimodal distribution. These are deep violations of the conditions necessary for linear regression. While our p-value would suggest that we should reject the null hypothesis because our slope indicates that there is a relationship between neighborhood calls and year, we fail to reject the null hypothesis because it is clear that is inappropriate to perform linear regression on a nonlinear distribution of data.

Discussion 21/25 - Room for more discussions

Not all Minneapolis districts have the same average 911 call rates for years 2008 to 2020. For the 2008-2020 year range, districts along the south, southwest, and southeast borders of Minneapolis (group 1) have the lowest 911 call rates, as opposed to their neighbors near the center of Minneapolis (group 3 and 4, particularly Central), who experience the highest call rates. We cannot make any causation claims at the moment, and it would be difficult to conduct an experiment where causal claims can be made in reference to 911 calls. Future research could be done by continuing to collect data for the next few years and separating the total number of 911 calls into high-priority and low-priority calls.

There is no linear relationship for total number of 911 calls each neighborhood makes per year from 2008 to 2020. The change in 911 calls that are made by Minneapolis neighborhoods since 2008 might be better described by more advanced regression models. If we were to treat the data as linear, then we would be 95% confidence that the rate of 911 calls an individual neighborhood makes increased by 79.84 and 150.69 calls for each year since 2008. However, we think it is best to look for other models using more advanced techniques in the future.

What variables would you collect?
How could your first result be used?

References

Roper, E., & Palazzolo, A. (MARCH 10, 2019). Big gaps in 911 response in Minneapolis; A Star Tribune analysis of more than 70,000 calls shows disparities in response times across the Twin Cities. Star Tribune. <https://www.startribune.com/emergency-callers-wait-longer-for-police-responses-at-minneapolis-border/506438091>

Variable Codebook

Variable name	Original name	Description	Type	Levels/Encoding
Neighborhood (original)	Neighborhood	The Minneapolis neighborhood that the use of force occurred in	categorical	identifier
District	District	The Minneapolis district that the use of force occurred in	categorical	identifier
Neighborhood calls (original)	TotalNeighborhoodCallsForYear	The total number of 911 calls made in the neighborhood for a given year	numeric	none
District calls	TotalDistrict	The total number of 911 calls made in the district for a given year	numeric	none
Year (original)	Year	The year the data was taken	numeric	none

5/5 Supplemental Materials

5/5 overall style + format

Appendix of Plots

