Final Project

Ellie Musson

2024-12-01

Project Topic

Throughout this project, I would like to investigate how the number of fires that occur that are larger than 100 acres across the United States has changed from 1900 until 2018. Is there a correlation between the year and the number of fire occurrences that are above 100 acres? Do more fires of this size occur as time goes on?

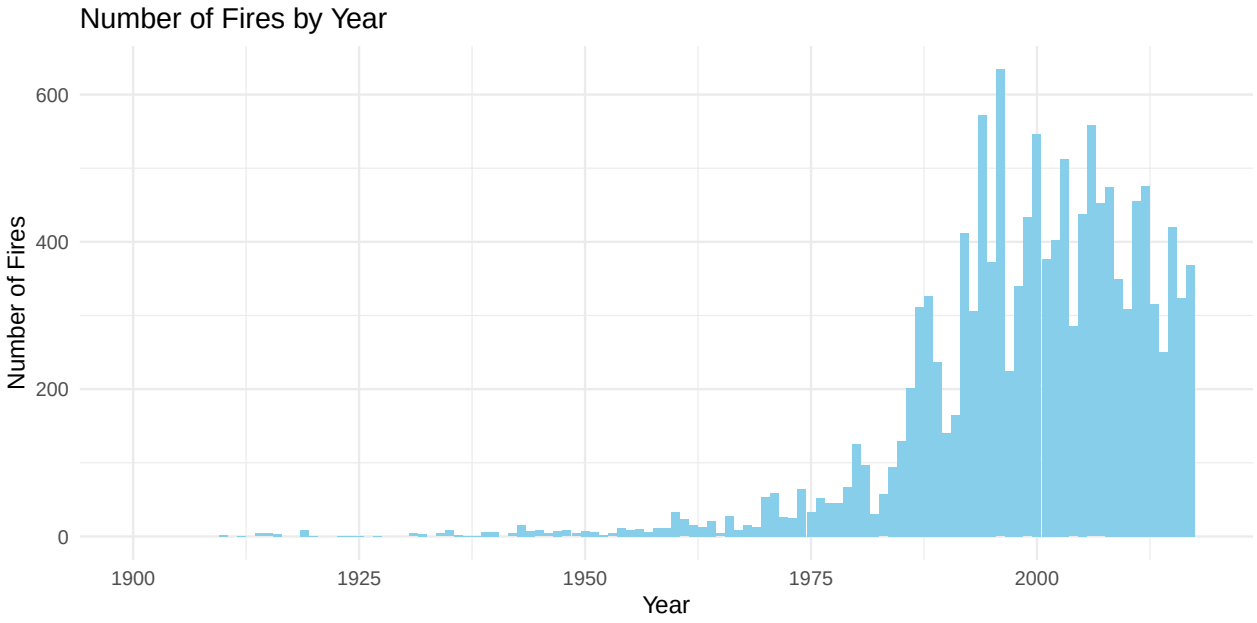
To investigate this question, I have downloaded the data file "National_USFS_Fire_Occurrence_Point_(Feature_Layer).csv" from *data.gov* that has occurrences of fires across the country from the early 1900's until 2018. This data source includes the location of the fire, the year it occurred, the size class (A-G), as well as many extra data points. For the particular question that I am looking to have solved, I will be focusing on the amount of fires that occurred that are class D or above and the year at which they occurred.

Data Description

To start this investigation, I wanted to look at how many fires total occurred each year that are a D class and above. The graph below shows the number of fires that happened each year. At first glance, it appears that fire occurrences really spiked between the years 1990 and 2000. There also seemed to be a steady decline afterward with the fire occurrences above a C class. However, I want to look further into this data to determine if fires are actually changing in frequency as time goes on.

```
## Warning: Removed 6 rows containing missing values (`position_stack()`).
```

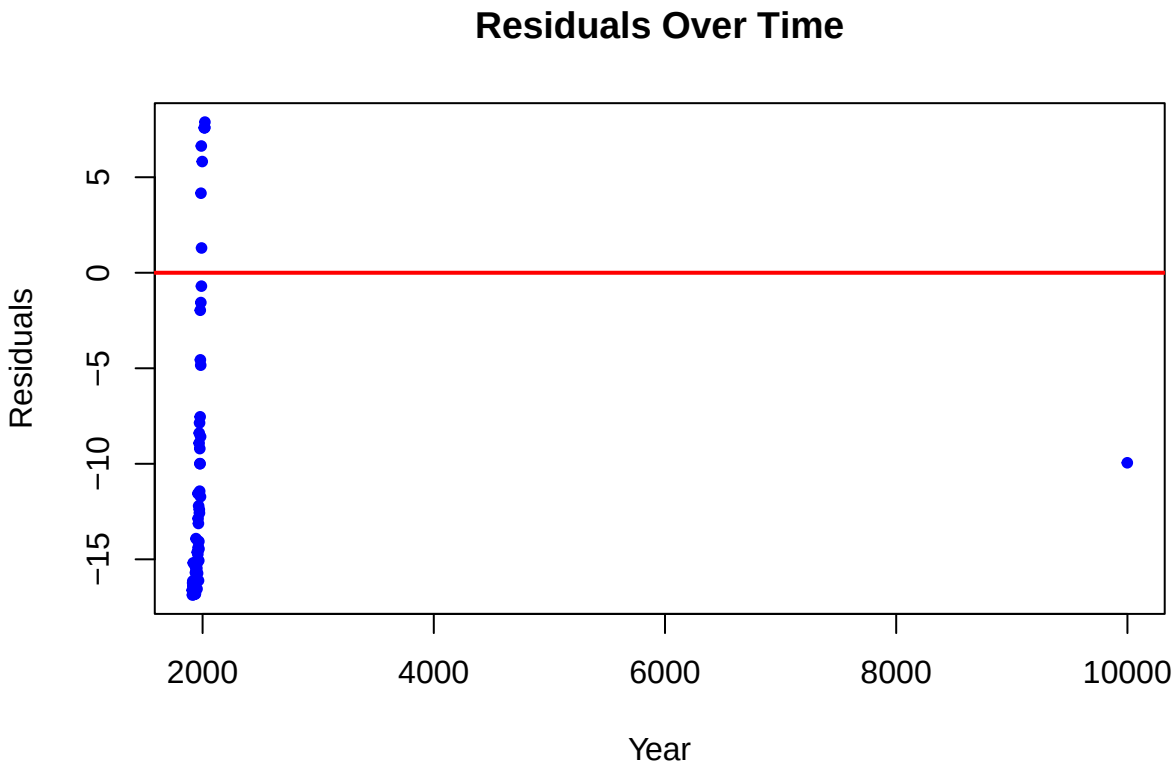
```
## Warning: Removed 1 rows containing missing values (`geom_bar()`).
```



Statistical Methods

I decided to use a Poisson regression model. I chose this model to investigate my original question of whether or not fire occurrences are related to the year. This model also operates under the assumption that all of the fires are independent of each other.

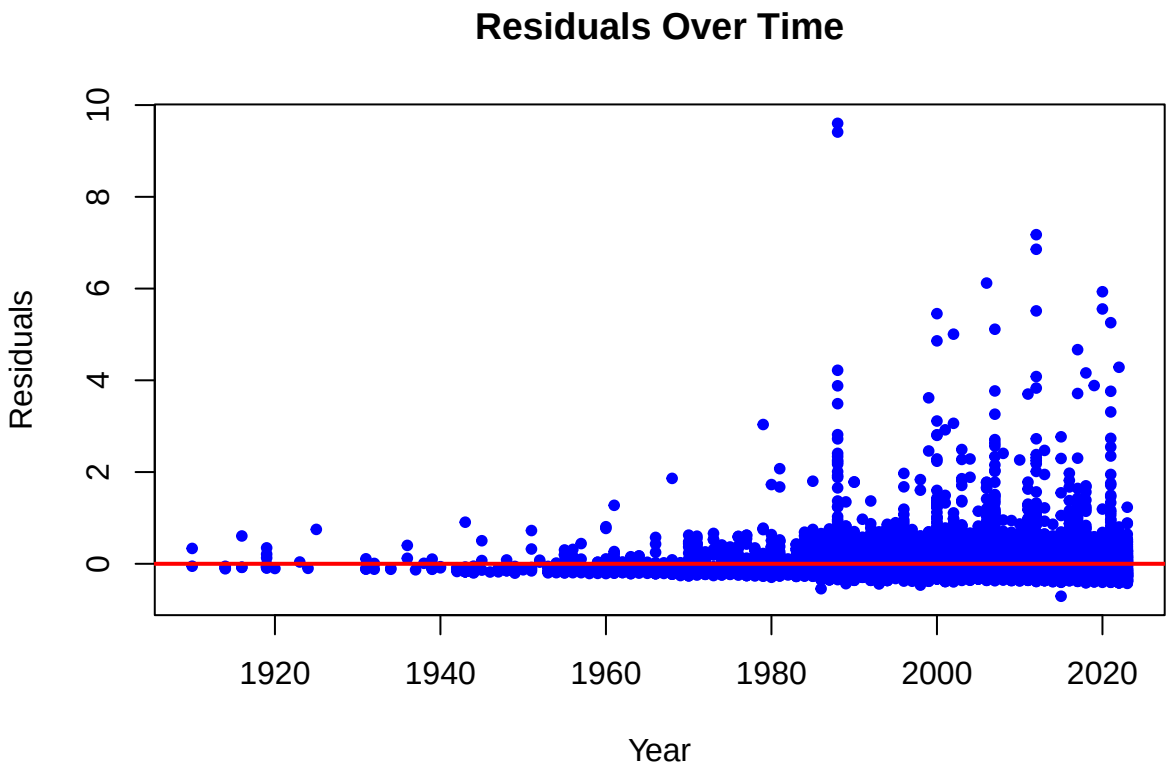
The summary statistics above highlight that the year of which fires take place is directly coordinated with how many fires above class C occur with a p-val of 1.69e-12. Naively, one might suggest that it appears that the year of when fires occur has a direct impact on the number of them that are above a C class. However, looking at the residual plot below, it is clearly seen that there is no random dispersion of the data points.



Furthermore, the correlation value between the year and the count of the fires is -0.04813243. Since the value is very close to zero, it suggests that there is no linear relationship between these two variables.

Since it appears that the year of which the fires occur is not correlated to how many fires happen, another model that I used uses the year the fire occurred and the location (in latitude and longitude) to predict the size of the fires. This model is a Quasi-Poisson which allows for non-integer response variables since the locations are in latitude and longitude coordinates. This makes the model more accurate since it makes the locations of where the fires occurred more exact.

This model estimates the size of the fires that happen by using the year and coordinates of where the fire happens. Each of these estimators is statistically significant, since the p-values are far below 0.05. The residual graph also looks far better than the previous model that only uses the year to predict the number of fires that happen. There is more random dispersion in the graph below compared to the other residual graph.



Results

Thus, the original question of whether or not there is a correlation between the year and the number of fires that are above a C class has a more complicated answer than a simple yes or no. The first model demonstrates that the year alone does not directly influence the number of fires over 100 acres that occur.

However, the other model that predicts the size of the fire from the year, as well as the coordinates of where the fires occurred is a better model. Therefore, you can say that the year fires occur is one of the variables that predicts the size of the fire. Having said that, the year by itself is not directly related to the size, so other predictor variables are necessary to sufficiently state that year is related to the size of fires that occur. By adding variables such as latitude and longitude, the model provides a much better fit, highlighting the importance of these factors alongside the year.

Future Work

Although this model has established a baseline for understanding fire size, significant opportunities remain to improve its predictive power and accuracy. Future work should focus on expanding the scope of predictors, improving model design, and leveraging more sophisticated statistical and machine learning techniques to capture the complex dynamics of fire behavior. This data set does not include descriptions of the environment, so perhaps a new a data set that includes all environment descriptions as well could be used.

Some more future analysis that could occur for this model is outlier analysis. Investigating and potentially removing or explaining influential data points could reduce model bias. This could potentially increase the randomness in the distribution in the residuals.