Predicting Federal Emergency Management Agency Funding for U.S. Disasters

Morgan Taylor

2021-11-13

# 1 Summary/Abstract

Rapidly increasing inflation rates and worsening effects of climate change are dramatically increasing the financial impacts of disasters. Within the United States, the Federal Emergency Management Agency (FEMA) is responsible for providing federal resources to support states and territories during the response phase as well as tracking the financial costs. All tiers of government would benefit from knowing which disaster declaration characteristics would change the amount of financial support from FEMA. This analysis utilizes machine learning technology to predict the funds requested by states or territories from FEMA as well as the funds eventually obligated to states or territories from FEMA for a disaster declaration. In particular, three models are considered: decision tree, bagged tree, and random forest. Cross-validation and bootstrap resampling structures are utilized, depending on the model, and the fits of the models are evaluated using the root mean square error (RMSE) statistic. The bagged tree model employing bootstrap resampling best fit both the requested and obligated fund models; however, future studies should consider other machine learning models as well as further refinement of the tuning paramters for the ones considered in this analysis. The prediction capabilities created by such models have the potential to dramatically improve comprehension of the key factors affecting financial disaster relief support.

# 2 Introduction

On November 8, 1988, United States (US) President Ronald Reagan signed the Robert T. Stafford Disaster Relief and Emergency Act into law, which formalizes the disaster declaration process and specifies the response responsibilities for each tier of government. Under this law, when a disaster occurs, the affected area (local or state representatives) conducts a damage assessment to estimate the amount of support required from the federal government. That assessment goes through the state government to the Federal Emergency Management Agency (FEMA) to the Department of Homeland Security (DHS) to the US President, who ultimately decides whether the affected region should receive an emergency declaration, a major disaster declaration, or neither. FEMA then deploys resources from across the federal government to support the disaster region, the volume and type of which depend upon the declaration type. However, the federal government does not necessarily incur all of the cost of these resources. Under most circumstances, this pay structure more accurately reflects a reimbursement model; that is, the states or territories submit requests to the FEMA for reimbursement after the expenses have occurred. However, there is then a reconciliation process by which the total funds a state or territory receives is often different than the original request. The funding process was further refined as part of the Sandy Recovery Improvement Act of 2013 as well as The Disaster Recovery Reform Act of 2019. These laws require FEMA to publicly provide the information regarding federal government resources paid for by state and federal government funds for each disaster declaration. As a result, the OpenFEMA Data Sets website now exists and provides information summarizing disaster declarations and mission assignments. While this has improved the accessible information, little has been done to draw conclusions from these data for predicting the financial impacts of disasters.

It would greatly benefit the federal government to be able to predict the amount of funding required for a disaster based on parameters known at the onset of the incident, as it would allow for improved funds flow and budgeting practices. Moreover, it would benefit states and territories to know which parameters increase the amount of federal government support in the aftermath of a disaster. As such, this analysis will examine the use of machine learning models in predicting FEMA expenses for a federally-declared disaster. In particular, a decision tree, bagged tree, and random forest model will be used to predict requested and obligated FEMA funds.

# 3 Methods and Results

*In most research papers, results and methods are separate. You can combine them here if you find it easier. You are also welcome to structure things such that those are separate sections.*

## 3.1 Data import and cleaning

The data were obtained from the OpenFEMA Data Sets website as well as the US Department of Agriculture Economic Research Service website. The mission assignments data was cleaned to represent the requested and obligated FEMA funds for each disaster declaration for each state or territory. Feature engineering was performed to create summary variables for total number of federal agencies involved in resource assistance, the amount of funding requested by the state or territory from FEMA, the amount of funding allocated to the state or territory from FEMA, and the average cost share for FEMA of the resources provided. Similarly, the disaster declaration summaries were also summarized to represent the unique disaster declarations for each state, bounded to the same time frame as the mission assignments data. Feature engineering was performed to create variables for incident duration, response duration, month of declaration, and year of the incident. The population data from the US Department of Agriculture Economic Research Service was incorporated into the disaster summary data, after creating a variable for state. The mission assignment data and declaration data were then joined to create a comprehensive dataset that was used for analysis. The supplementary files contain the RStudio 4.1 code used to process the data. The following R packages were utilized for importing and cleaning the data: here, tidyverse, skimr, and lubridate.

## 3.2 Exploratory analysis

The processed data were then explored to characterize and summarize the relevant features in RStudio 4.1 with the following packages: here, tidyverse, skimr, summarytools, ggplot2, car, gtsummary, scales, cowplot, and table1. The supplementary files contain the code that generates the following results.

Table ?? reflects the results of the exploratory data analysis. Between February 2012 and November 2021, there were 446 disaster declarations for states and territories. Funds requested by states or territories for FEMA ranged from -$855,200.3 to $4,388,540,486, with an average of $29,673,687, and FEMA funds obligated to the states and territories after reconciliation ranged from -$855,200.3 to $440,842,4297 with an average of $29,744,109.

|  | Overall (N=446) |
| --- | --- |
| **Requested FEMA Funds** |  |
| Mean (SD) | 29700000 (219000000) |
| Median [Min, Max] | 109000 [-855000, 4390000000] |
| **Obligated FEMA Funds** |  |
| Mean (SD) | 29700000 (220000000) |
| Median [Min, Max] | 109000 [-855000, 4410000000] |
| **State** |  |
| AK | 8 (1.8%) |
| AL | 10 (2.2%) |
| AR | 10 (2.2%) |
| AZ | 3 (0.7%) |
| CA | 20 (4.5%) |
| CO | 6 (1.3%) |
| CT | 9 (2.0%) |
| DC | 3 (0.7%) |
| DE | 3 (0.7%) |
| FL | 18 (4.0%) |
| GA | 13 (2.9%) |
| HI | 7 (1.6%) |
| IA | 10 (2.2%) |
| ID | 6 (1.3%) |
| IL | 5 (1.1%) |
| IN | 3 (0.7%) |
| KS | 7 (1.6%) |
| KY | 7 (1.6%) |
| LA | 22 (4.9%) |
| MA | 6 (1.3%) |
| ME | 3 (0.7%) |
| MI | 7 (1.6%) |
| MN | 10 (2.2%) |
| MO | 8 (1.8%) |
| MS | 15 (3.4%) |
| MT | 2 (0.4%) |
| NC | 14 (3.1%) |
| ND | 6 (1.3%) |
| NE | 9 (2.0%) |
| NH | 7 (1.6%) |
| NJ | 5 (1.1%) |
| NM | 6 (1.3%) |
| NV | 3 (0.7%) |
| NY | 11 (2.5%) |
| OH | 7 (1.6%) |
| OK | 12 (2.7%) |
| OR | 10 (2.2%) |
| PA | 8 (1.8%) |
| PR | 11 (2.5%) |
| RI | 6 (1.3%) |
| SC | 13 (2.9%) |
| SD | 11 (2.5%) |
| TN | 11 (2.5%) |
| TX | 21 (4.7%) |
| UT | 1 (0.2%) |
| VA | 7 (1.6%) |
| VT | 6 (1.3%) |
| WA | 12 (2.7%) |
| WI | 9 (2.0%) |
| WV | 16 (3.6%) |
| WY | 3 (0.7%) |
| **Incident Year** |  |
| Mean (SD) | 2020 (2.92) |
| Median [Min, Max] | 2020 [2010, 2020] |
| **Declaration Month** |  |
| Mean (SD) | 5.62 (3.56) |
| Median [Min, Max] | 6.00 [1.00, 12.0] |
| **Declaration Type** |  |
| DR | 330 (74.0%) |
| EM | 116 (26.0%) |
| **Incident Type** |  |
| Biological | 78 (17.5%) |
| Chemical | 1 (0.2%) |
| Coastal Storm | 4 (0.9%) |
| Dam/Levee Break | 2 (0.4%) |
| Earthquake | 5 (1.1%) |
| Fire | 22 (4.9%) |
| Flood | 75 (16.8%) |
| Hurricane | 105 (23.5%) |
| Mud/Landslide | 3 (0.7%) |
| Other | 4 (0.9%) |
| Severe Ice Storm | 12 (2.7%) |
| Severe Storm(s) | 117 (26.2%) |
| Snow | 5 (1.1%) |
| Terrorist | 1 (0.2%) |
| Tornado | 11 (2.5%) |
| Volcano | 1 (0.2%) |
| **Incident Duration (days)** |  |
| Mean (SD) | 128 (246) |
| Median [Min, Max] | 12.0 [0, 660] |
| **Response Duration (days)** |  |
| Mean (SD) | 1330 (914) |
| Median [Min, Max] | 1100 [70.0, 3590] |
| **Individuals/Households Program** |  |
| Not Awarded | 267 (59.9%) |
| Awarded | 179 (40.1%) |
| **Public Assistance Program** |  |
| Not Awarded | 12 (2.7%) |
| Awarded | 434 (97.3%) |
| **Hazard Mitigation Program** |  |
| Not Awarded | 117 (26.2%) |
| Awarded | 329 (73.8%) |
| **State Population** |  |
| Mean (SD) | 8640000 (9700000) |
| Median [Min, Max] | 5020000 [577000, 39500000] |
| **Counties per State** |  |
| Mean (SD) | 74.8 (52.3) |
| Median [Min, Max] | 67.0 [1.00, 254] |
| **Federal Agencies Involved** |  |
| Mean (SD) | 4.91 (6.56) |
| Median [Min, Max] | 2.00 [0, 55.0] |
| **Average FEMA Cost Share** |  |
| Mean (SD) | 0.806 (0.354) |
| Median [Min, Max] | 0.978 [0, 1.00] |

Figure 3.1 shows the number of disaster declarations per year, and Figure 3.2 shows the total FEMA funds (both requested and obligated) by states per year. The substantial peak in 2020 corresponds to the COVID-19 pandemic, and the one in 2017 represents the hurricane season with Hurricanes Irma, Harvey, and Maria. These characteristics can also be seen in Table 3.1, which represents FEMA Funding for the 3 Costliest Years (2012 - 2021).

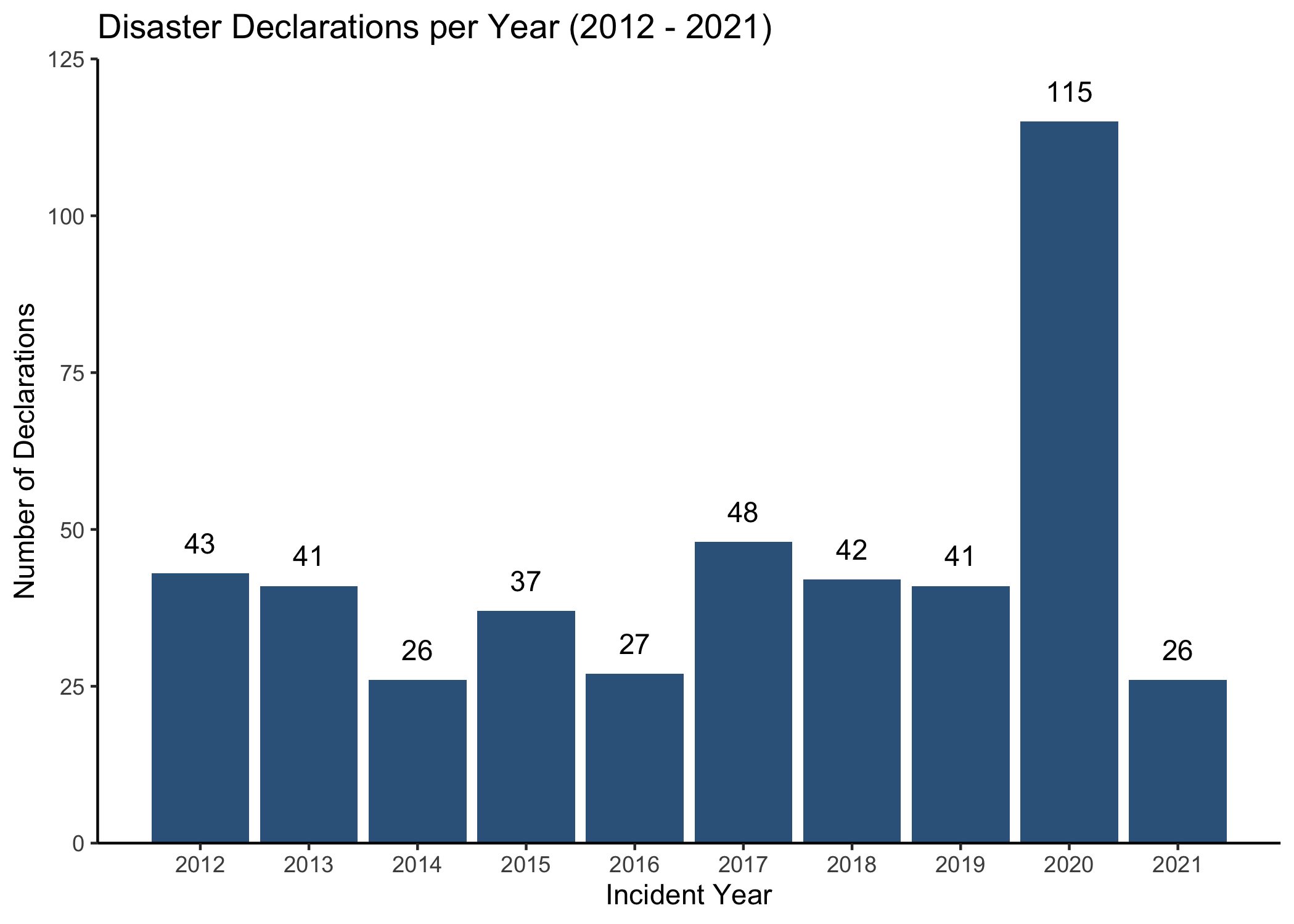


Figure 3.1: Number of disaster declarations per year 2012 - 2021

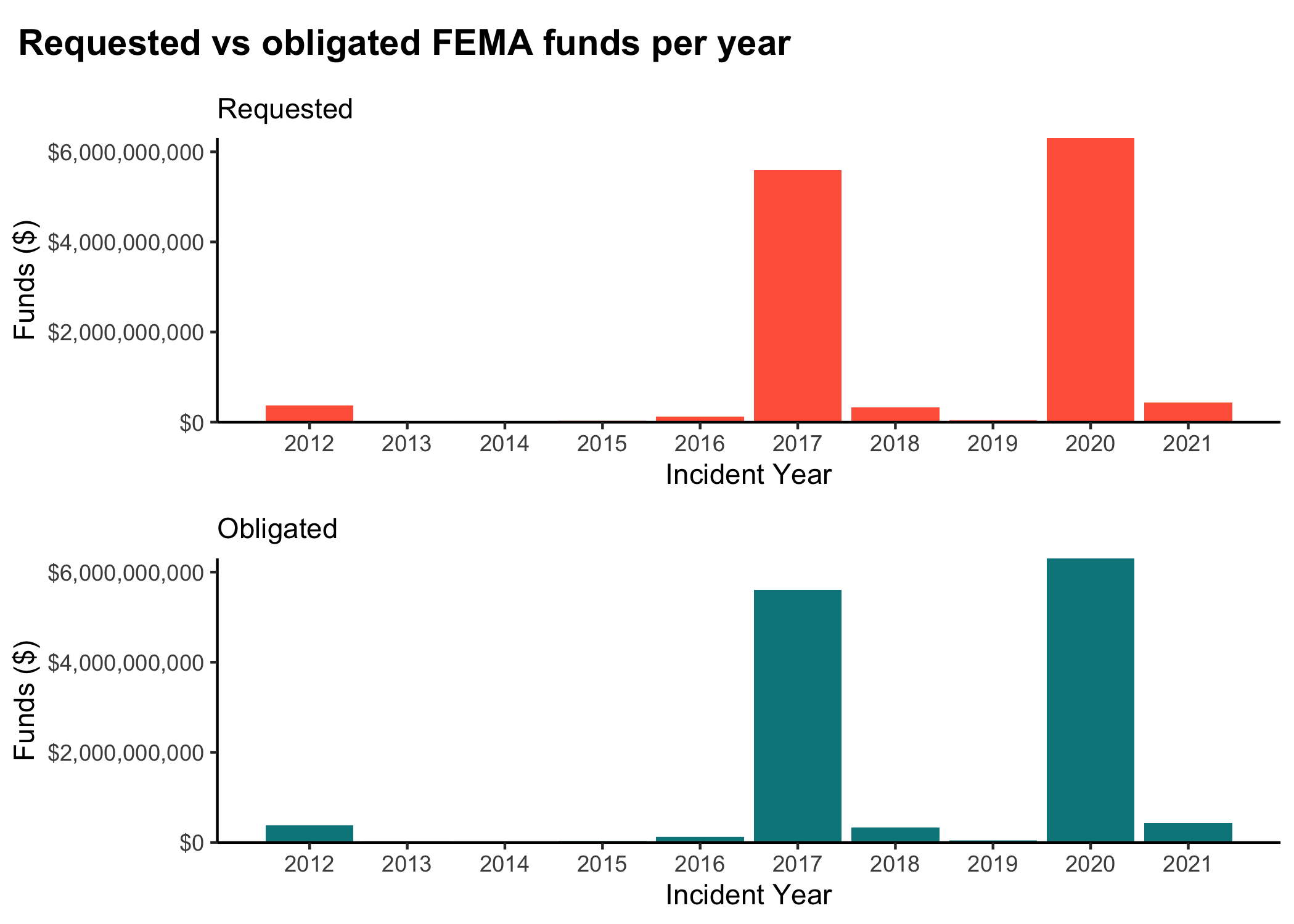


Figure 3.2: State or territory requested FEMA funds and obligated FEMA funds by year 2012 - 2021

Table 3.1: Table 2. FEMA Funding for the 3 Costliest Years (2012 - 2021)

|  | Year | | |
| --- | --- | --- | --- |
| Variable | 2012, N = 431 | 2017, N = 481 | 2020, N = 1151 |
| **Requested Funds** | 8,523,455 (49,075,292) | 116,482,973 (638,527,353) | 54,834,301 (104,214,883) |
| **Obligated Funds** | 8,740,375 (49,166,434) | 116,915,384 (641,357,763) | 54,838,235 (104,217,430) |
| 1Mean (SD) | | | |

Figure 3.3 shows the number of disaster declarations per month, and Figure 3.4 shows the total FEMA funds (both requested and obligated) by states per month. The substantial peak in January likely corresponds to the COVID-19 pandemic, and the ones in August, September, and October represent the peak hurricane season. These characteristics can also be seen in Table 3.2, which represents FEMA Funding for the 3 Costliest Months.

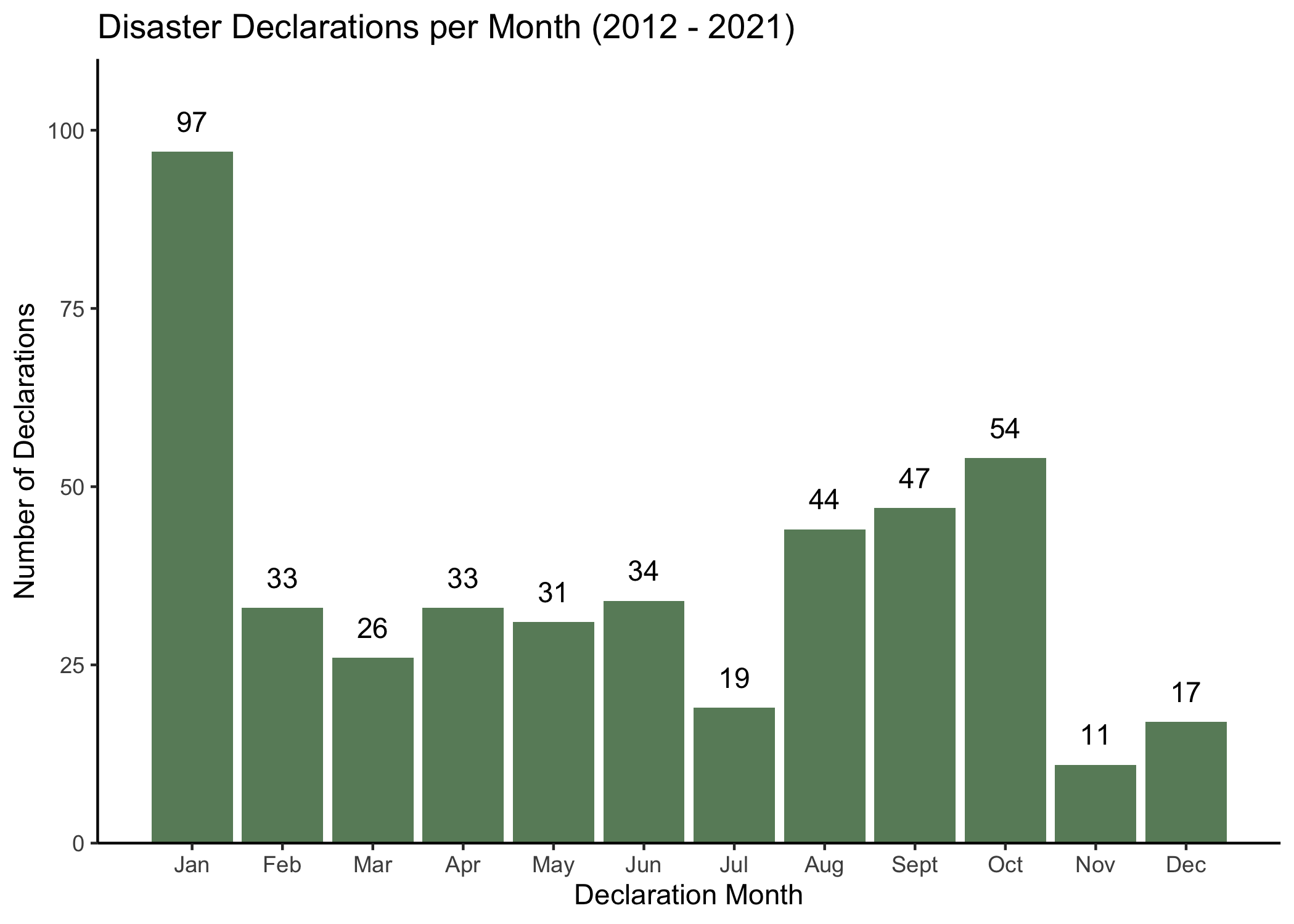


Figure 3.3: Number of disaster declarations per month for years 2012 - 2021

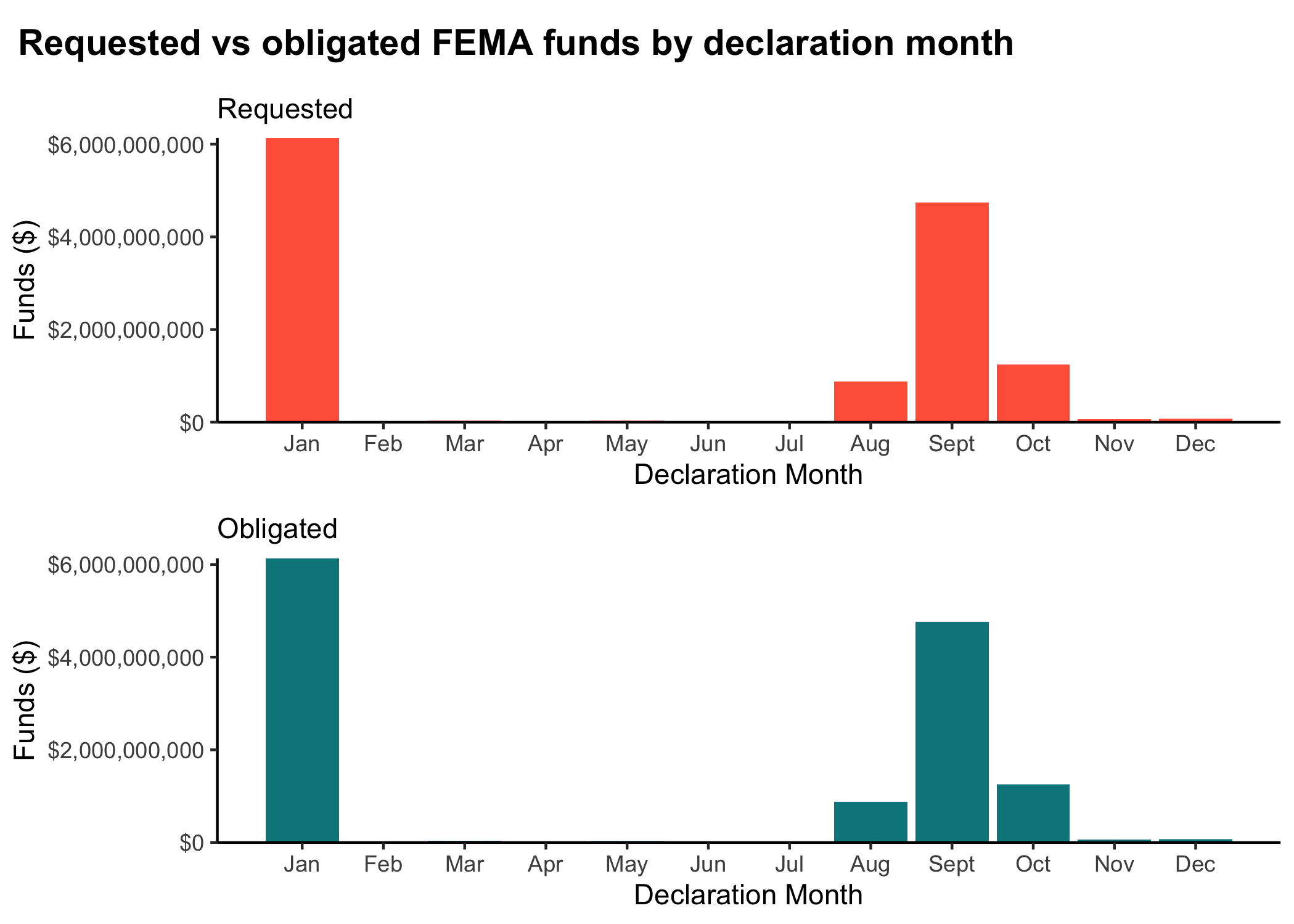


Figure 3.4: State or territory requested FEMA funds and obligated FEMA funds by month for 2012 - 2021

Table 3.2: Table 3. FEMA Funding for the 3 Costliest Months (2012 - 2021)

|  | Month | | |
| --- | --- | --- | --- |
| Variable | Jan1 | Sept1 | Oct1 |
| **Requested Funds** | 63,229,783 (111,136,115) | 100,831,890 (639,779,995) | 22,992,573 (101,460,411) |
| **Obligated Funds** | 63,239,449 (111,135,744) | 101,252,822 (642,677,256) | 23,165,582 (101,470,263) |
| 1Mean (SD) | | | |

All 50 states as well as Washington, DC were included in the final, combined dataset. Table 3.3 provides the five most disaster-prone states and summarizes their requested FEMA funding and received FEMA funding over the study period.

Table 3.3: Table 4. FEMA Funding for 5 Most Disaster-Prone States 2012 - 2021

|  | State | | | | |
| --- | --- | --- | --- | --- | --- |
| Variable | CA, N = 201 | FL, N = 181 | LA, N = 221 | TX, N = 211 | WV, N = 161 |
| **Requested Funds** | 62,605,465 (169,900,730) | 27,337,218 (68,075,813) | 34,318,344 (90,594,708) | 36,586,287 (119,143,196) | 6,659,588 (25,471,101) |
| **Obligated Funds** | 62,605,465 (169,900,730) | 27,337,218 (68,075,813) | 34,318,344 (90,594,708) | 36,637,886 (119,167,108) | 6,660,260 (25,470,955) |
| 1Mean (SD) | | | | | |

Incident duration ranged from zero days to 660 days (or almost two years), as seen by Figure 3.5. The incident duration of 660 days likely corresponds to the COVID-19 pandemic, which is an ongoing incident of almost two years. Further analysis of all variables can be found in the supplementary information.

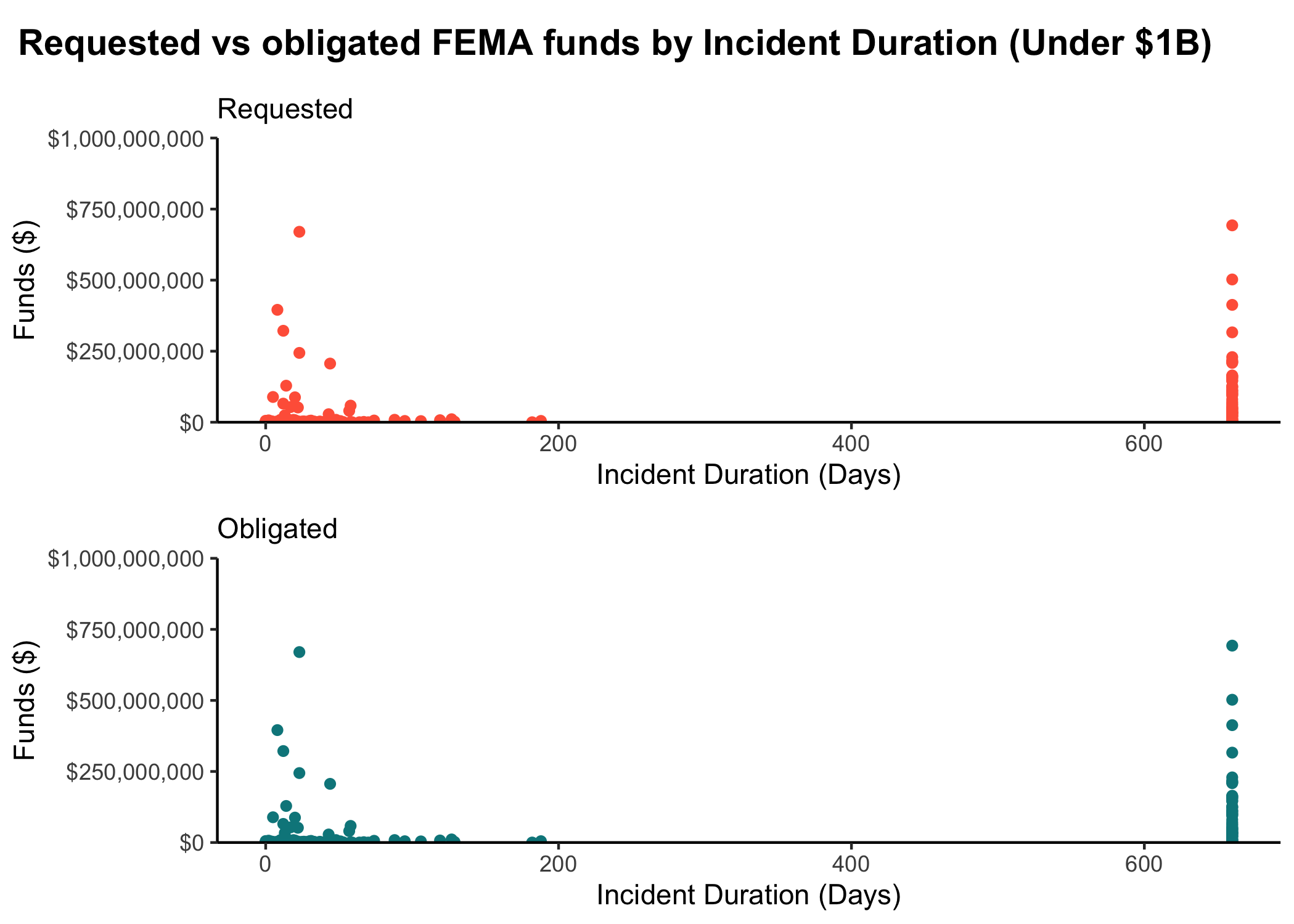
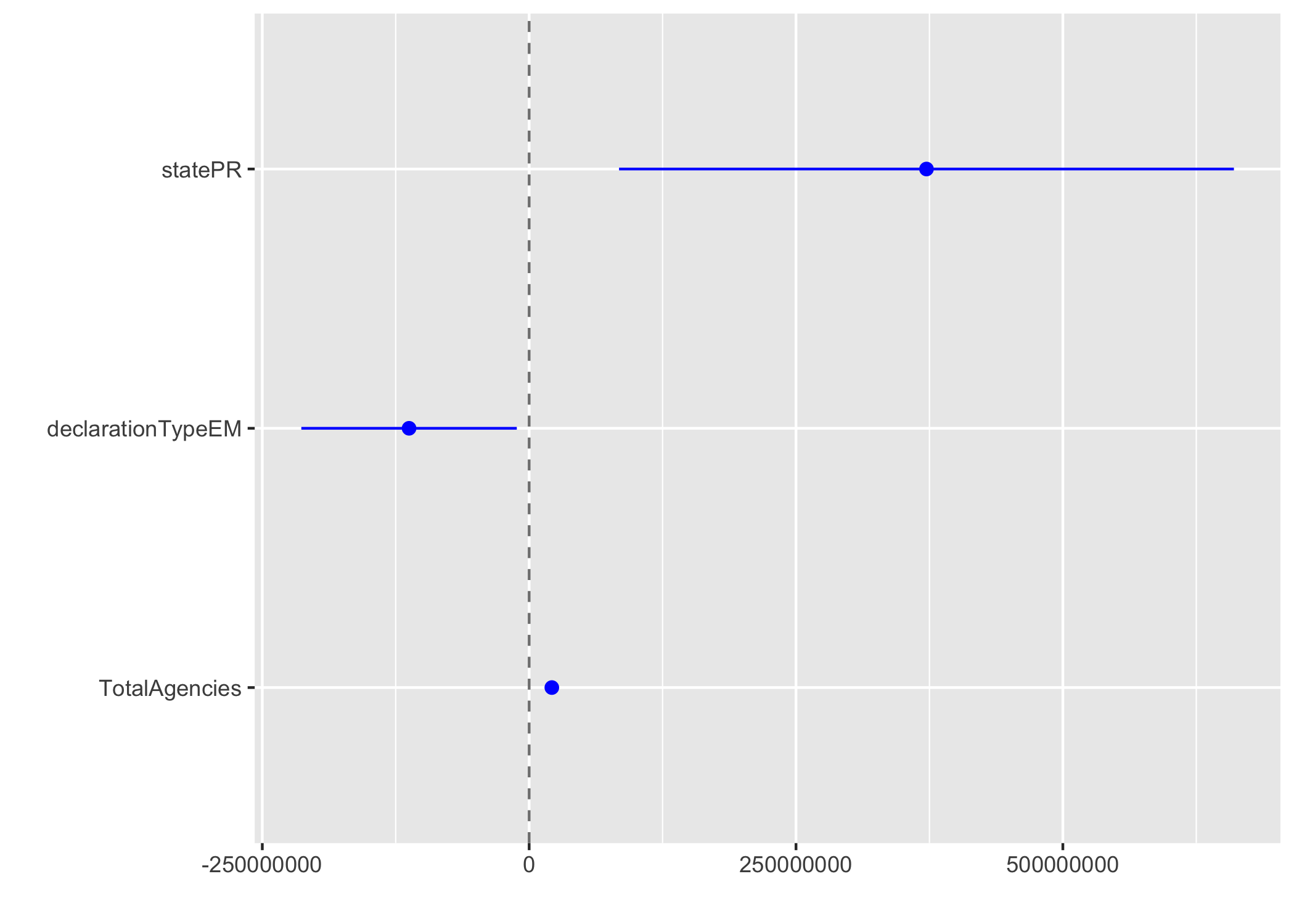
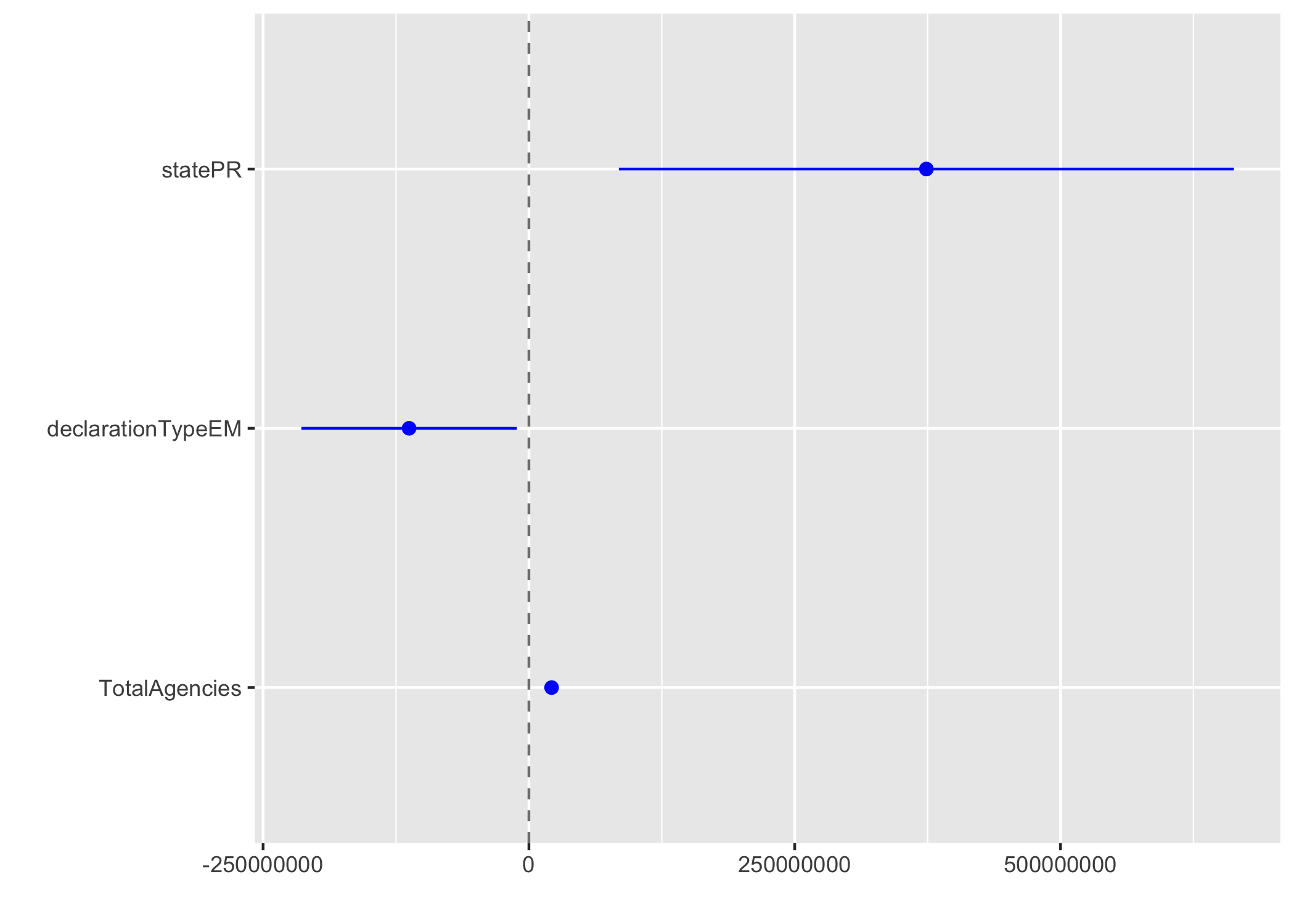


Figure 3.5: Incident duration for requested FEMA funds and obligated FEMA funds under $1 billion for 2012 - 2021

Prior to fitting the data to machine learning models, standard simple and multivariable regression models were fitted using the tidymodels framework in RStudio 4.1. Both outcomes of interest were fitted, using incident duration as the primary predictor of interest. Figure @ref(fig:req\_MVR\_bp) and @ref(fig:obl\_MVR\_bp) show the significant variables fitted in the multivariable regression models for requested and obligated FEMA funds, respectively.



(#fig:req\_MVR\_bp)Significant predictors in multivariate regression model for requested FEMA funds



(#fig:obl\_MVR\_bp)Significant predictors in multivariate regression model for obligated FEMA funds

Table 3.4 shows the comparisons of the models, using the root mean square error (RMSE) and r-squared statistics. The requested and obligated funding models performed almost identically. None of the models fit the data well, which suggests more sophisticated modeling practices are necessary to adequately characterize the data.

Table 3.4: Linear model comparisons

|  |  |  |
| --- | --- | --- |
|  | Obligated | Requested |
| MVR: rmse | 1.998684e+08 | 1.990141e+08 |
| MVR: rsq | 3.641982e-01 | 3.643611e-01 |
| SLR: rmse | 2.498472e+08 | 2.488021e+08 |
| SLR: rsq | 6.466200e-03 | 6.538900e-03 |

## 3.3 Full analysis

*Use one or several suitable statistical/machine learning methods to analyze your data and to produce meaningful figures, tables, etc. This might again be code that is best placed in one or several separate R scripts that need to be well documented. You want the code to produce figures and data ready for display as tables, and save those. Then you load them here.*

Three models were considered for the prediction analysis: decision tree, bagged tree, and random forest, and each were fit for both outcomes of interest (e.g. requested and obligated funds). Prior to fitting the models, the data were subset based on outcome of interest, and some feature engineering was performed. Due to the limited values in certain categories, the state variable was summarized into the ten FEMA regions and the least frequent disaster types were added to the “Other” category. The data were split such that 66.7% were used for training the models and the remaining 33.3% were used for testing the final chosen model, stratified by the incident type. Cross-validation was used for the decision tree and random forest models, using a five-fold, five-repeat resampling structure, also stratified by the incident type. The bagged tree model instead implemented a bootstrap resampling structure, repeated ten times, again stratified by the incident type. The models were then compared based on the root mean square error (RMSE), the standard metric for regression models. The analysis was completed in RStudio 4.1 with the following packages: here, tidyverse, tidymodels, skmir, broom.mixed, rpart.plot, vip, glmnet, doParallel, ranger, and baguette.

The identified decision tree model for each outcome is illustrated by Figures 3.6 and 3.7. Both outcomes identified the same decision point features: total number of federal agencies involved in the response, duration of the incident, state population, state counties, and whether the individual / housing assistance program was awarded.

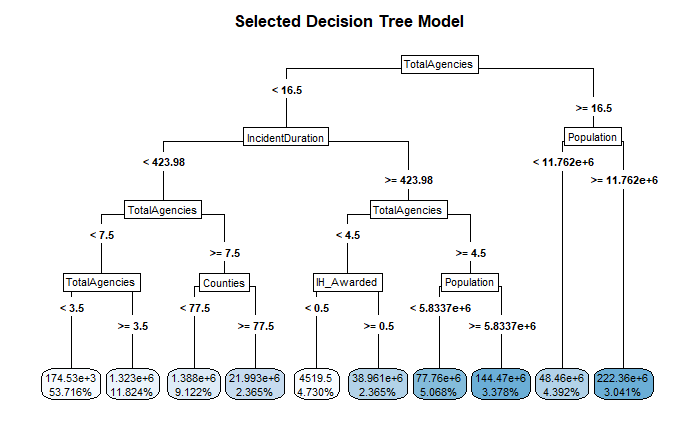


Figure 3.6: Selected decision tree for predicting requested FEMA funds

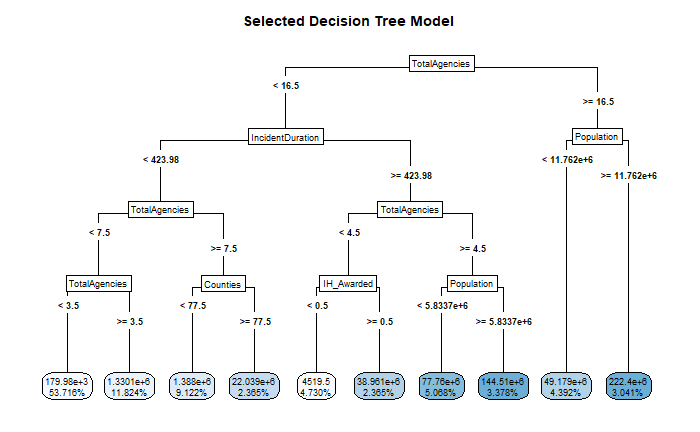


Figure 3.7: Selected decision tree for predicting obligated FEMA funds

The most important variables in the identified random forest model for each outcome is illustrated by Figures 3.8 and 3.9. Both outcomes identified the same important features: incident duration, total number of agencies involved, individual/housing assistance program awardance, and incident month. Interestingly, however, the obligated funds model includes days of the entire response as the fifth most important feature, whereas the requested funds model includes it as the eleventh most important feature. Unfortunately, the tidymodels framework does not support visualization of the bagged tree model.

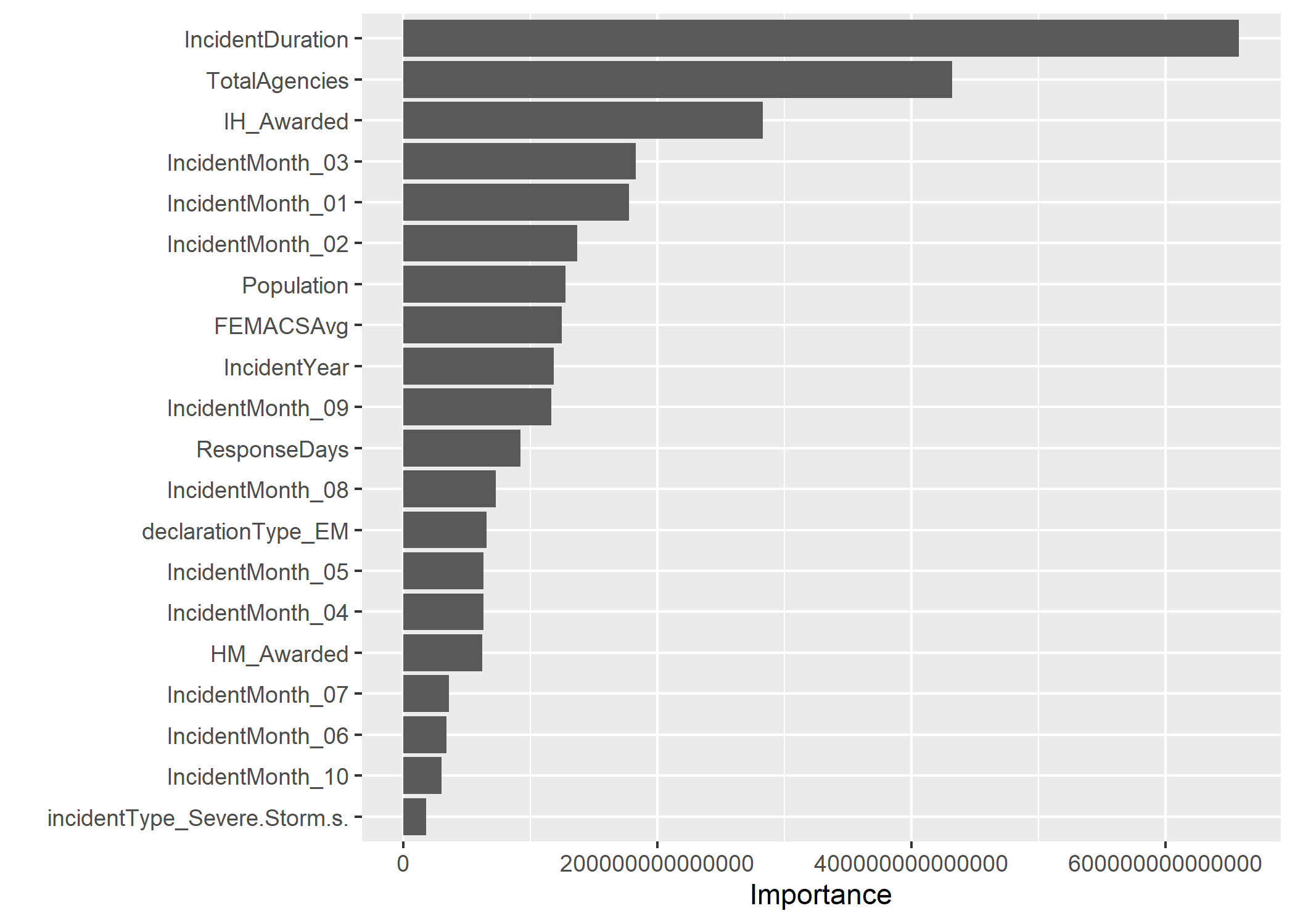


Figure 3.8: 20 most important variables for the random forest model for requested FEMA funds

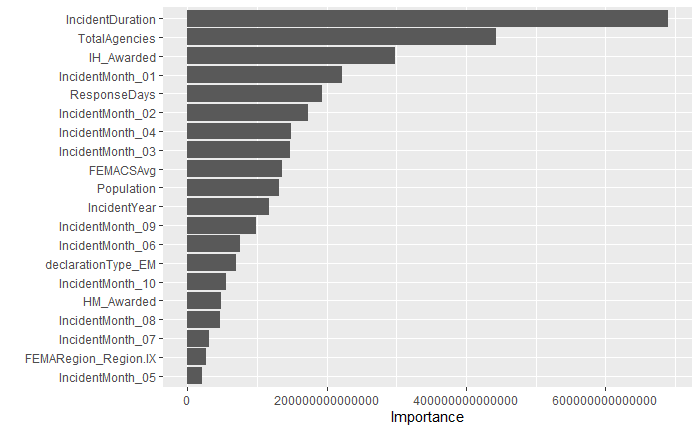


Figure 3.9: 20 most important variables for the random forest model for obligated FEMA funds

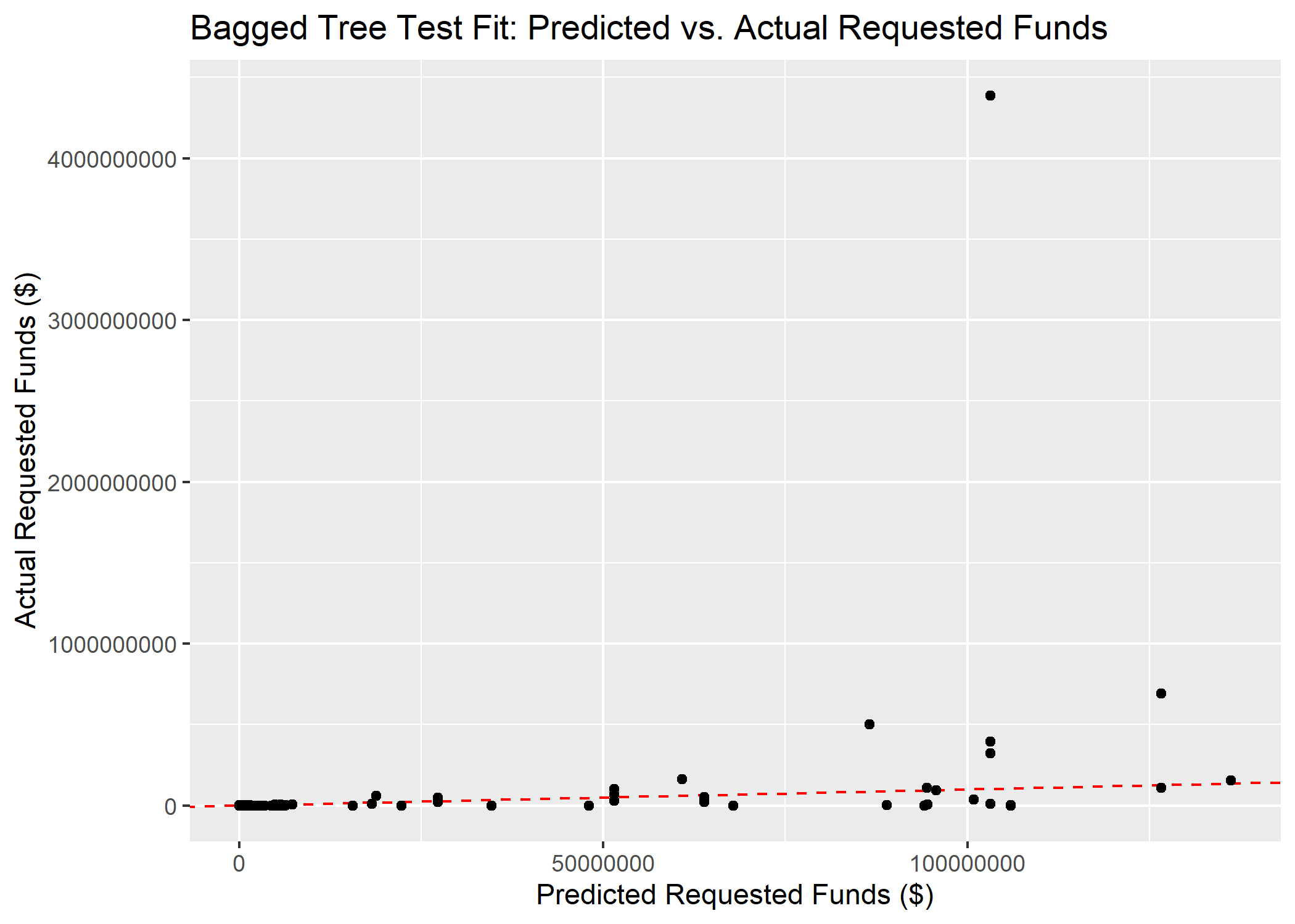
Table 3.5 shows the model comparisons for estimating requested FEMA funding, and Table 3.6 similarly shows the model comparisons for estimating obligated FEMA funding. In both outcomes, the bagged tree had the lowest RMSE; however, the standard error estimate of the RMSE was higher in the bagged tree models than for the other models.

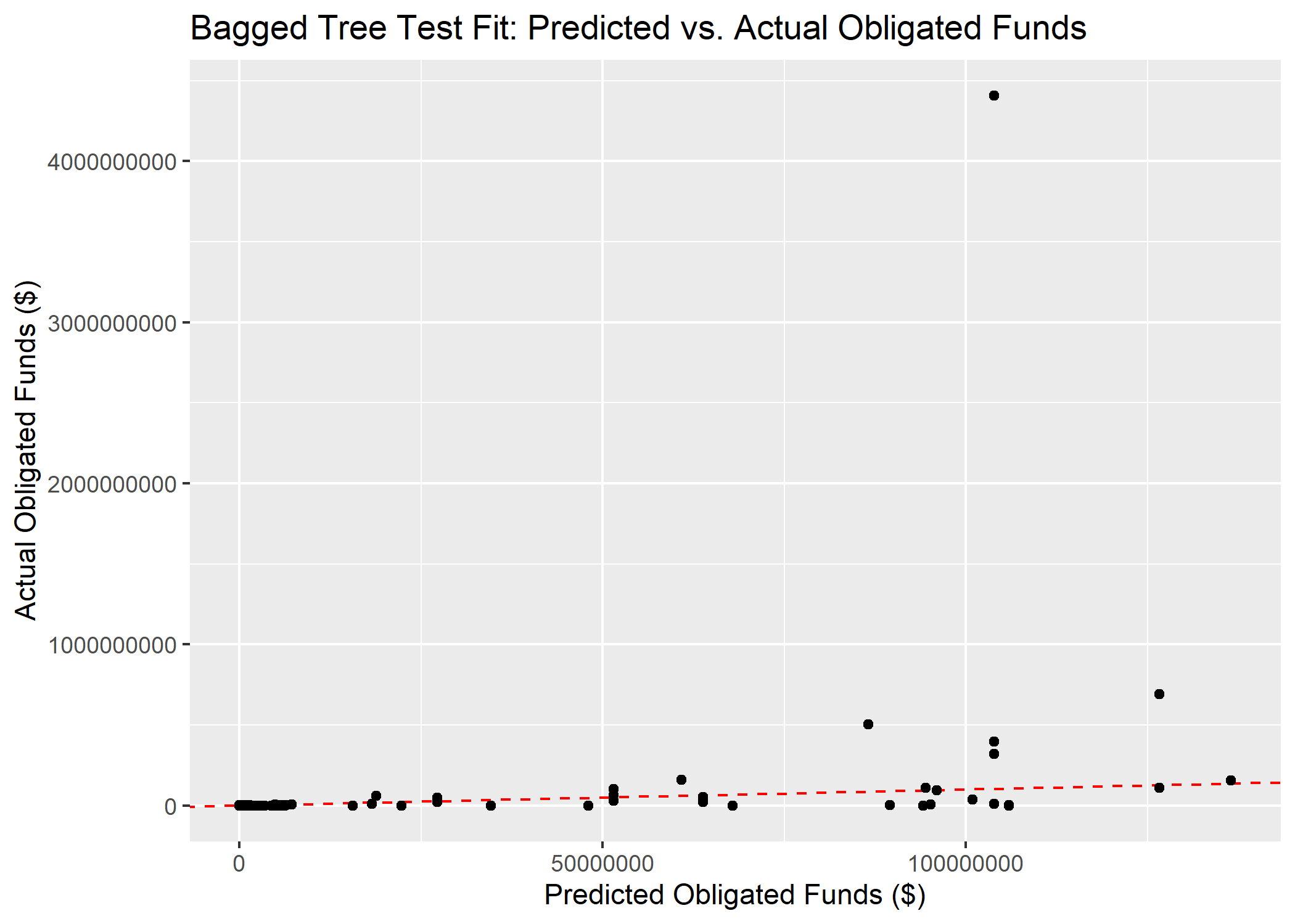
Table 3.5: ML model comparison for requested FEMA funding

|  |  |  |
| --- | --- | --- |
| rmse | SE | model |
| 45805848 | 5238182 | Bagged Tree |
| 49238144 | 4788041 | RF |
| 50834310 | 4273545 | Tree |
| 64177983 | 0 | Null - Train |

Table 3.6: ML model comparison for obligated FEMA funding

|  |  |  |
| --- | --- | --- |
| rmse | SE | model |
| 45732084 | 5245947 | Bagged Tree |
| 49203267 | 4760076 | RF |
| 50798220 | 4275946 | Tree |
| 64188313 | 0 | Null - Train |

The bagged tree model was still chosen as the best model of the three, and it was fitted to the testing data sample. The RMSEs for the testing and training datasets were substantially different. Figure ?? shows the predicted and actual requested funds values, and Figure ?? does the same for obligated FEMA funds. The supplementary information contains the code for the models as well as all diagnostic plots. 



(#fig:BT\_pred\_act\_test-obl)Predicted vs actual obligated FEMA funds for the bagged tree fit

# 4 Discussion

This analysis applies machine learning modeling techniques to predict the requested and obligated funds from FEMA in the aftermath of a disaster.

## 4.1 Summary and Interpretation

*Summarize what you did, what you found and what it means.*

## 4.2 Strengths and Limitations

This analysis shows promise but requires further tuning and validation. THe usefulness of the model is limited by the broad variance present in the relatively small dataset, which potentially leads to biases. Moreover, the COVID-19 pandemic created extensive outliers in the dataset, which further complicate the prediction algorithms. While it is tempting to remove the outlier from a statistics perspective, it makes less sense scientifically, as the outlier disasters are often the ones where funding prediction would be the most useful.

## 4.3 Conclusions

*What are the main take-home messages?*

# 5 References