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

Morgan Taylor

2021-12-13

# 1 Supplemental Information

This document represents additional information about the data and results.

## 1.1 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=326) |
| --- | --- |
| **Requested FEMA Funds** |  |
| Mean (SD) | 40600000 (255000000) |
| Median [Min, Max] | 420000 [0.000000000000455, 4390000000] |
| **Obligated FEMA Funds** |  |
| Mean (SD) | 40700000 (256000000) |
| Median [Min, Max] | 434000 [0.000000000000455, 4410000000] |
| **Average FEMA Cost Share** |  |
| Mean (SD) | 0.964 (0.0573) |
| Median [Min, Max] | 1.00 [0.675, 1.00] |
| **Federal Agencies Involved** |  |
| Mean (SD) | 6.23 (7.03) |
| Median [Min, Max] | 3.00 [1.00, 55.0] |
| **Percent of Counties Awarded IH** |  |
| Mean (SD) | 0.338 (1.59) |
| Median [Min, Max] | 0.00197 [0, 27.7] |
| **Percent of Counties Awarded PA** |  |
| Mean (SD) | 0.594 (1.56) |
| Median [Min, Max] | 0.357 [0, 27.3] |
| **Percent of Counties Awarded HM** |  |
| Mean (SD) | 0.476 (1.56) |
| Median [Min, Max] | 0.207 [0, 27.3] |
| **Response Duration (days)** |  |
| Mean (SD) | 1400 (898) |
| Median [Min, Max] | 1180 [70.0, 3590] |
| **Incident Duration (days)** |  |
| Mean (SD) | 136 (250) |
| Median [Min, Max] | 14.0 [0, 660] |
| **Declaration Month** |  |
| January | 71 (21.8%) |
| February | 22 (6.7%) |
| March | 22 (6.7%) |
| April | 24 (7.4%) |
| May | 22 (6.7%) |
| June | 25 (7.7%) |
| July | 13 (4.0%) |
| August | 32 (9.8%) |
| September | 31 (9.5%) |
| October | 41 (12.6%) |
| November | 9 (2.8%) |
| December | 14 (4.3%) |
| **Incident Year** |  |
| Mean (SD) | 2020 (2.89) |
| Median [Min, Max] | 2020 [2010, 2020] |
| **Incident Type** |  |
| Biological | 60 (18.4%) |
| Earthquake | 4 (1.2%) |
| Fire | 20 (6.1%) |
| Flood | 61 (18.7%) |
| Hurricane | 73 (22.4%) |
| Other | 15 (4.6%) |
| Severe Ice Storm | 4 (1.2%) |
| Severe Storm(s) | 82 (25.2%) |
| Tornado | 7 (2.1%) |
| **Declaration Type** |  |
| DR | 261 (80.1%) |
| EM | 65 (19.9%) |
| **FEMA Region** |  |
| Region I | 27 (8.3%) |
| Region II | 19 (5.8%) |
| Region III | 25 (7.7%) |
| Region IV | 62 (19.0%) |
| Region IX | 31 (9.5%) |
| Region V | 37 (11.3%) |
| Region VI | 54 (16.6%) |
| Region VII | 20 (6.1%) |
| Region VIII | 24 (7.4%) |
| Region X | 27 (8.3%) |
| **Counties per State** |  |
| Mean (SD) | 75.3 (53.5) |
| Median [Min, Max] | 67.0 [1.00, 254] |
| **State Population** |  |
| Mean (SD) | 9310000 (10300000) |
| Median [Min, Max] | 5710000 [577000, 39500000] |

Figure 1.1 shows the number of disaster declarations per year, and Figure 1.2 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 1.1, which represents FEMA Funding for the 3 Costliest Years (2012 - 2021).

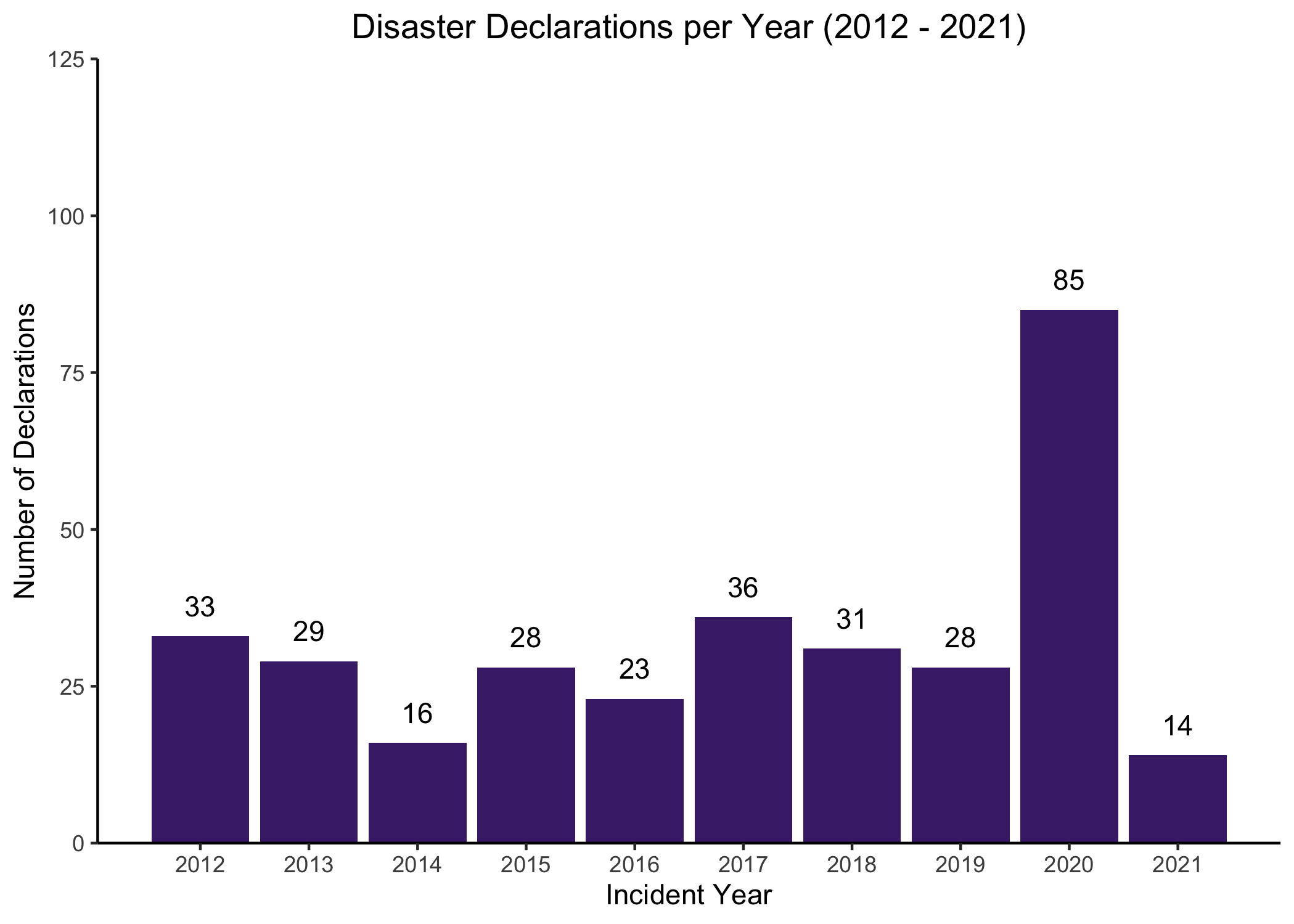


Figure 1.1: Number of disaster declarations per year 2012 - 2021

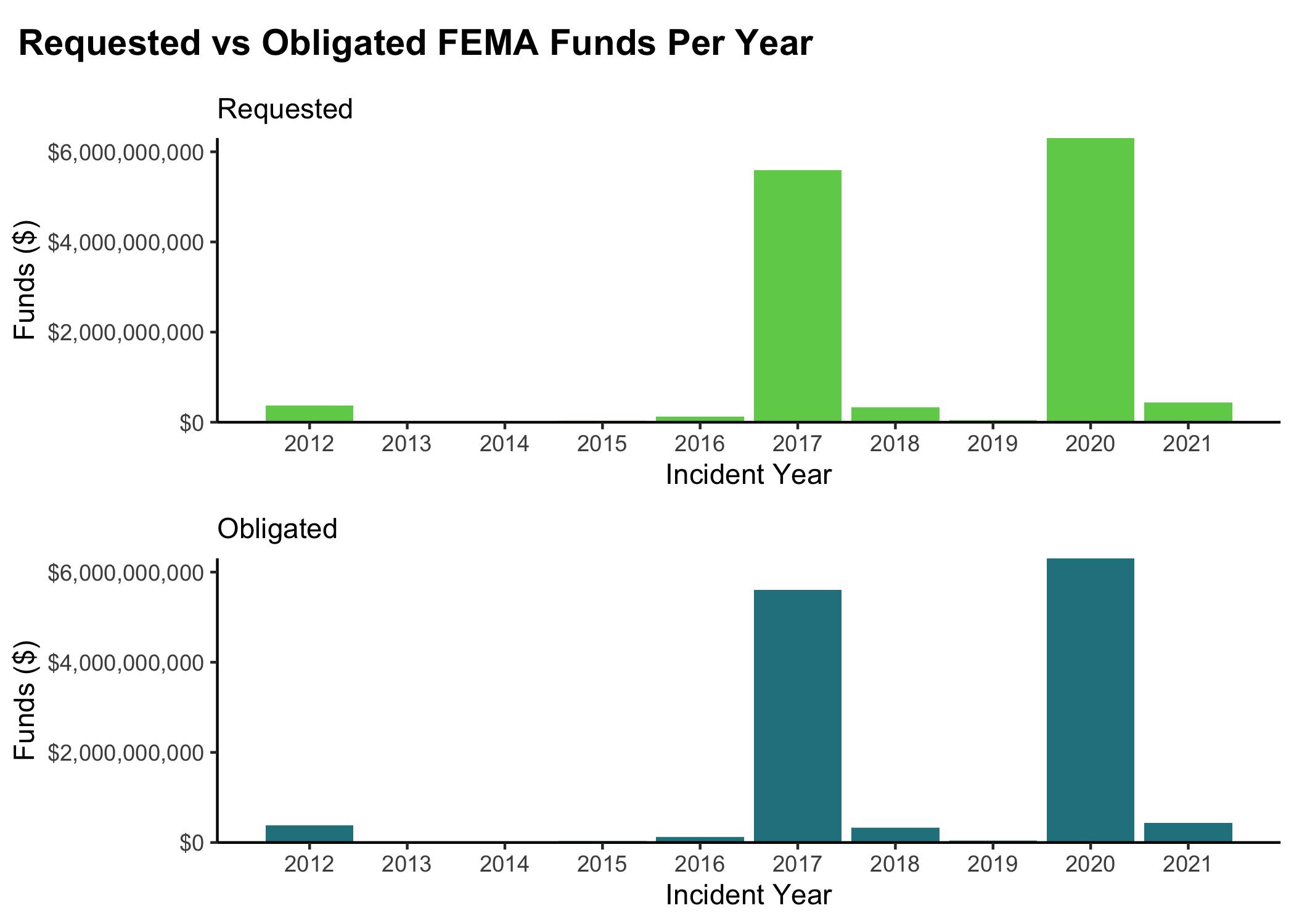


Figure 1.2: State or territory requested FEMA funds and obligated FEMA funds by year 2012 - 2021

Table 1.1: Table 2. FEMA Funding for the 3 Costliest Years (2012 - 2021)

|  | Year | | |
| --- | --- | --- | --- |
| Variable | 2012, N = 331 | 2017, N = 361 | 2020, N = 851 |
| **Requested Funds** | 11,106,321 (55,959,108) | 155,310,630 (735,733,161) | 74,188,913 (115,267,136) |
| **Obligated Funds** | 11,388,973 (56,050,414) | 155,887,178 (739,000,529) | 74,194,235 (115,269,357) |
| 1Mean (SD) | | | |

Figure 1.3 shows the number of disaster declarations per month, and Figure 1.4 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 1.2, which represents FEMA Funding for the 3 Costliest Months.

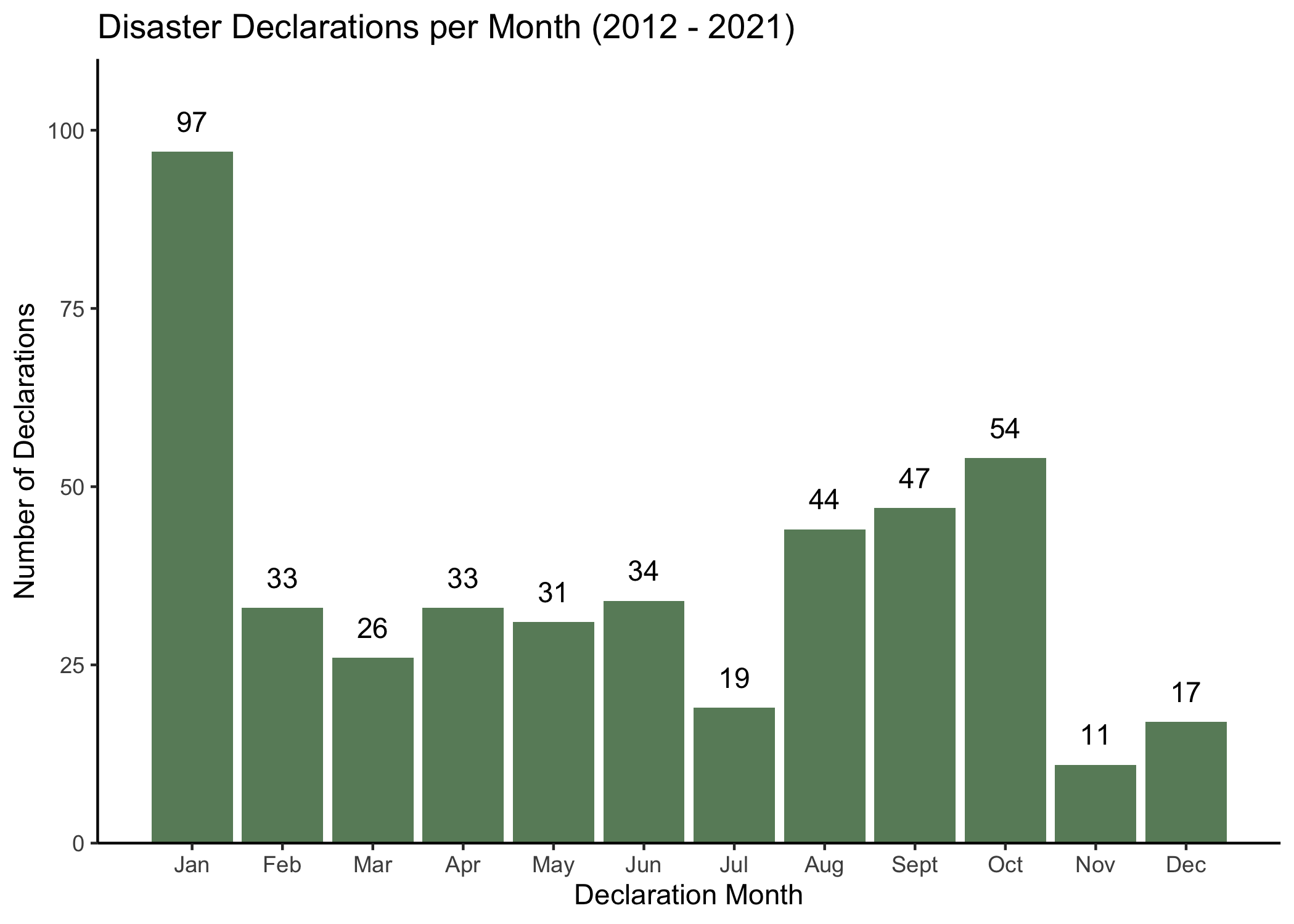


Figure 1.3: Number of disaster declarations per month for years 2012 - 2021

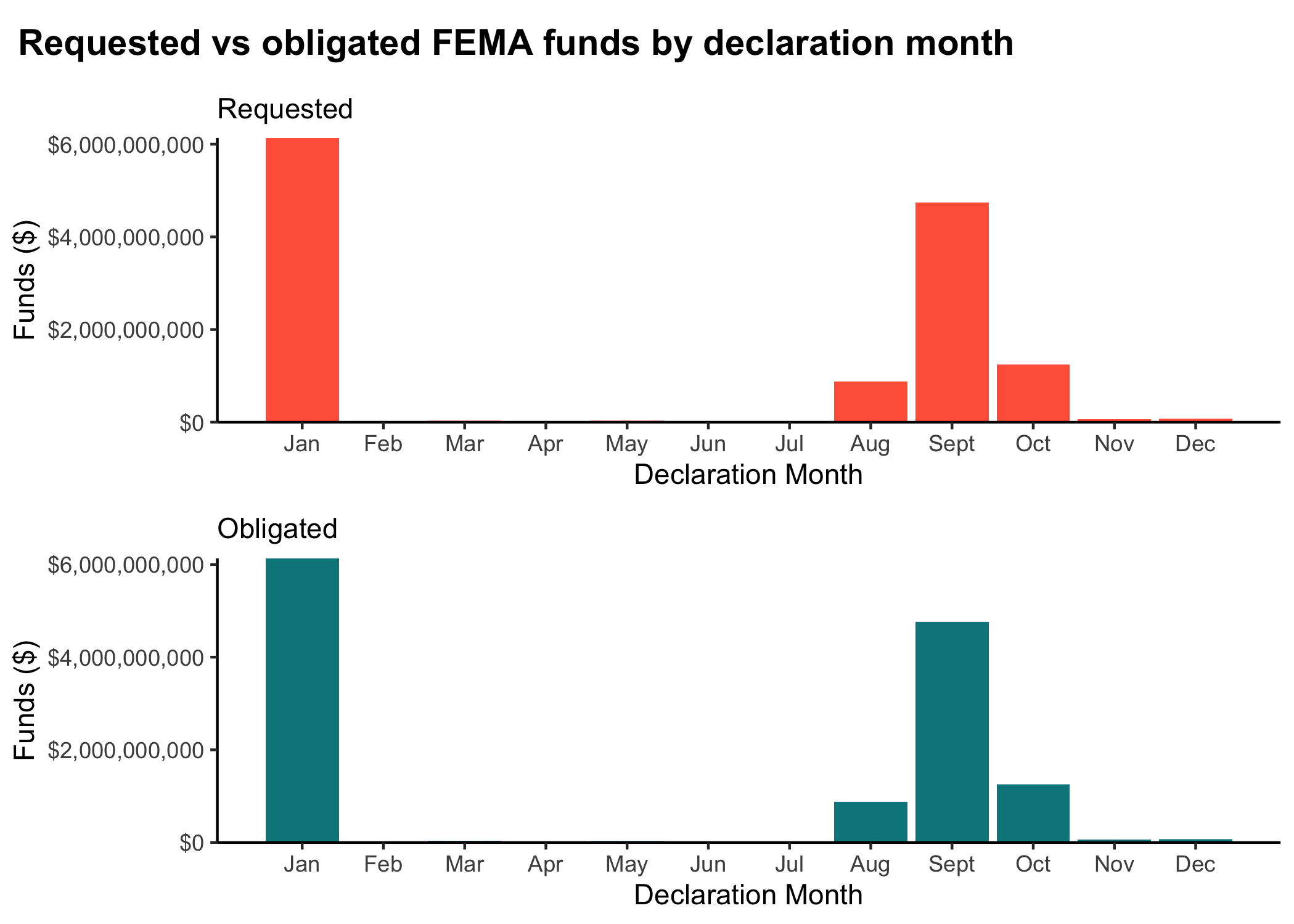


Figure 1.4: State or territory requested FEMA funds and obligated FEMA funds by month for 2012 - 2021

Table 1.2: 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 1.3 provides the five most disaster-prone states and summarizes their requested FEMA funding and received FEMA funding over the study period.

Table 1.3: Table 4. FEMA Funding for 5 Most Disaster-Prone States 2012 - 2021

|  | State | | | | |  | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Variable | CA, N = 191 | FL, N = 131 | LA, N = 161 | MN, N = 101 | OK, N = 101 | SD, N = 101 | TX, N = 161 | WA, N = 111 | WV, N = 101 |
| **Requested Funds** | 65,900,489 (173,898,625) | 37,851,532 (78,320,708) | 47,187,723 (104,127,805) | 4,758,425 (13,184,247) | 6,413,888 (14,086,311) | 2,727,818 (7,026,623) | 48,019,502 (135,429,624) | 14,599,599 (47,096,442) | 10,655,340 (32,155,687) |
| **Obligated Funds** | 65,900,489 (173,898,625) | 37,851,532 (78,320,708) | 47,187,723 (104,127,805) | 4,758,425 (13,184,247) | 6,413,888 (14,086,311) | 2,727,818 (7,026,623) | 48,087,226 (135,451,571) | 14,599,599 (47,096,442) | 10,656,415 (32,155,346) |
| 1Mean (SD) | | | | | | | | | |

Incident duration ranged from zero days to 660 days (or almost two years), as seen by Figure 1.5. 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.

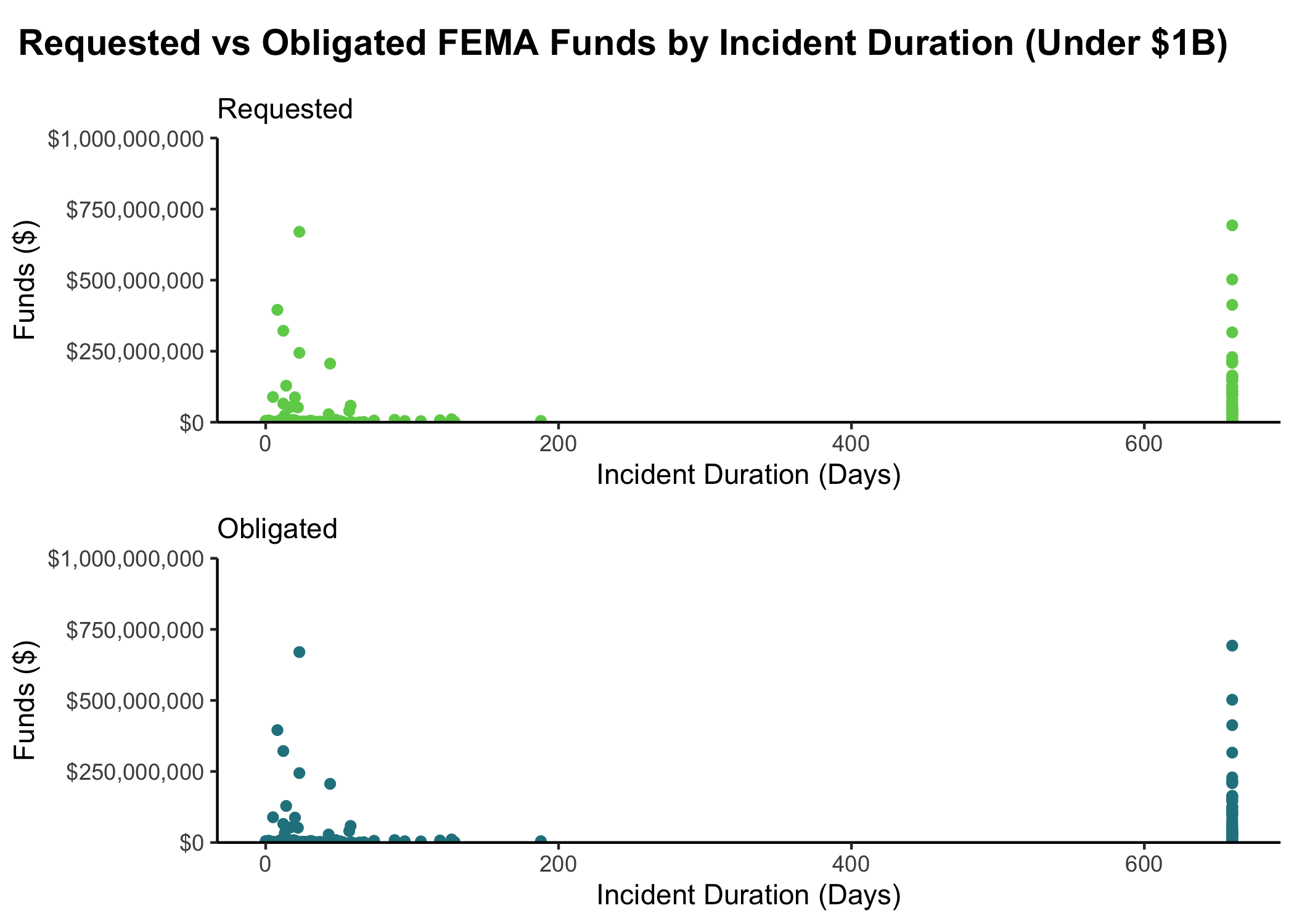
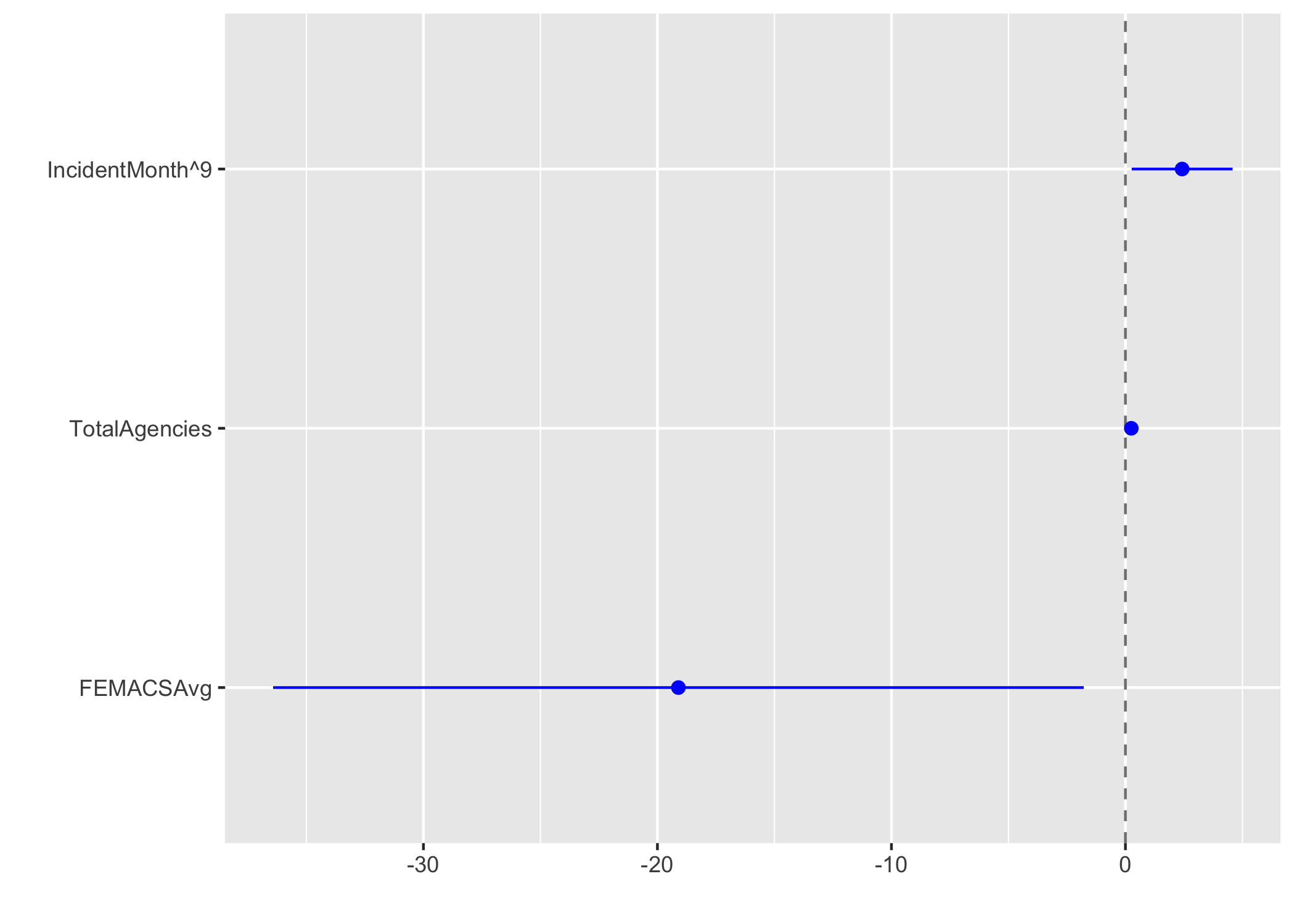
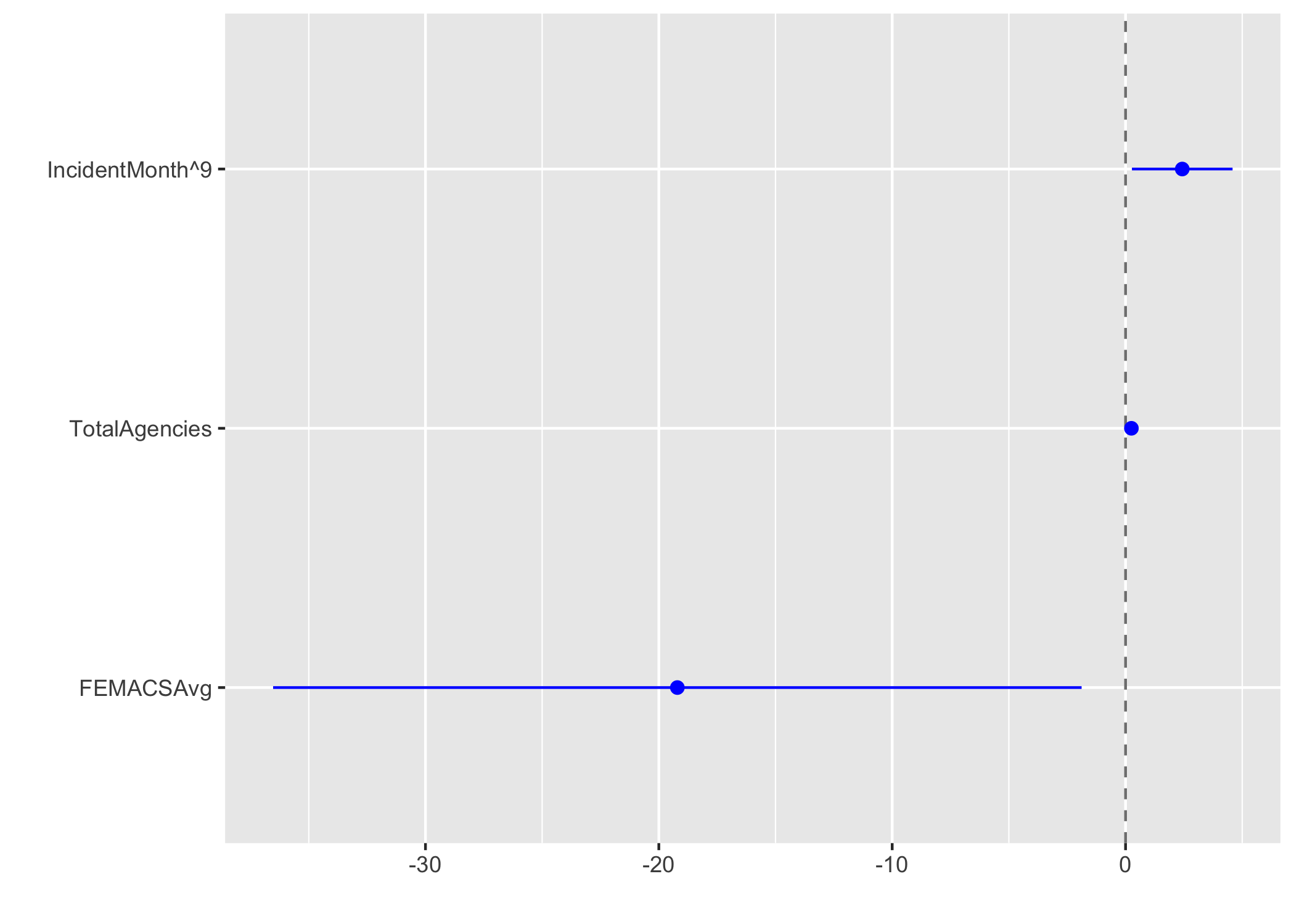


Figure 1.5: Incident duration for requested FEMA funds and obligated FEMA funds under $1 billion for 2012 - 2021

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 1.4 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 1.4: Linear model comparisons

|  | Obligated | Requested |
| --- | --- | --- |
| MVR: rmse | 3.2825340 | 3.2821346 |
| MVR: rsq | 0.5767613 | 0.5765888 |
| SLR: rmse | 4.6045374 | 4.6012644 |
| SLR: rsq | 0.1672034 | 0.1678456 |

## 1.2 Full analysis

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 1.6 and 1.7. 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.

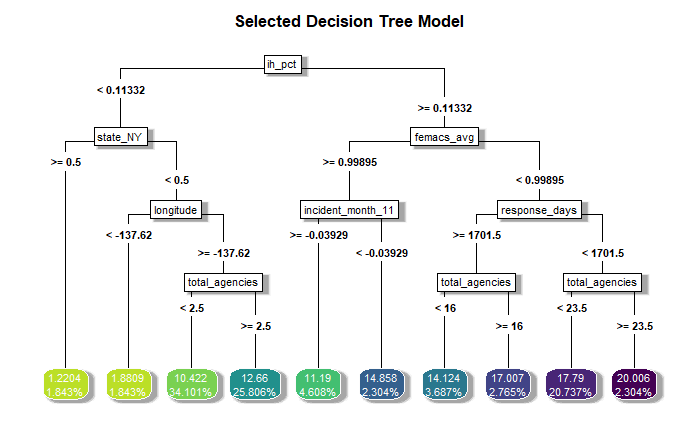


Figure 1.6: Selected decision tree for predicting requested FEMA funds

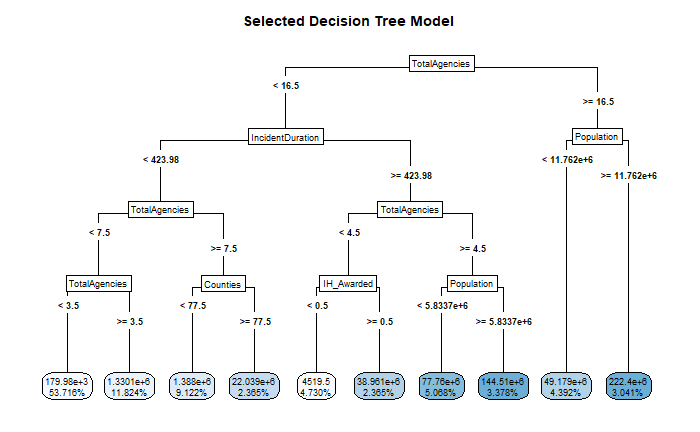


Figure 1.7: Selected decision tree for predicting obligated FEMA funds

The most important variables in the identified random forest model for each outcome is illustrated by Figures 1.8 and 1.9. 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.

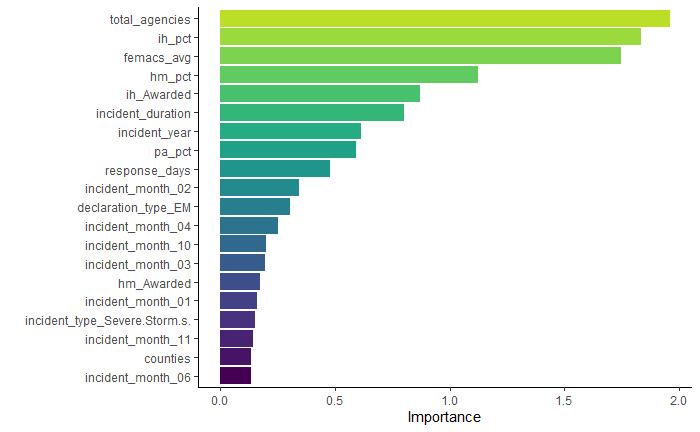


Figure 1.8: 20 most important variables for the random forest model for requested FEMA funds

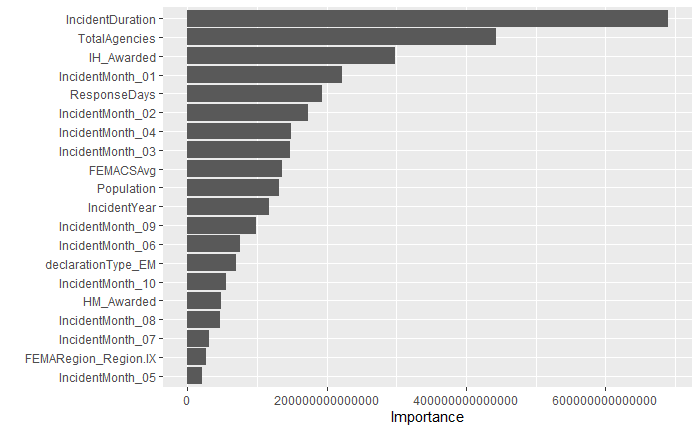


Figure 1.9: 20 most important variables for the random forest model for obligated FEMA funds

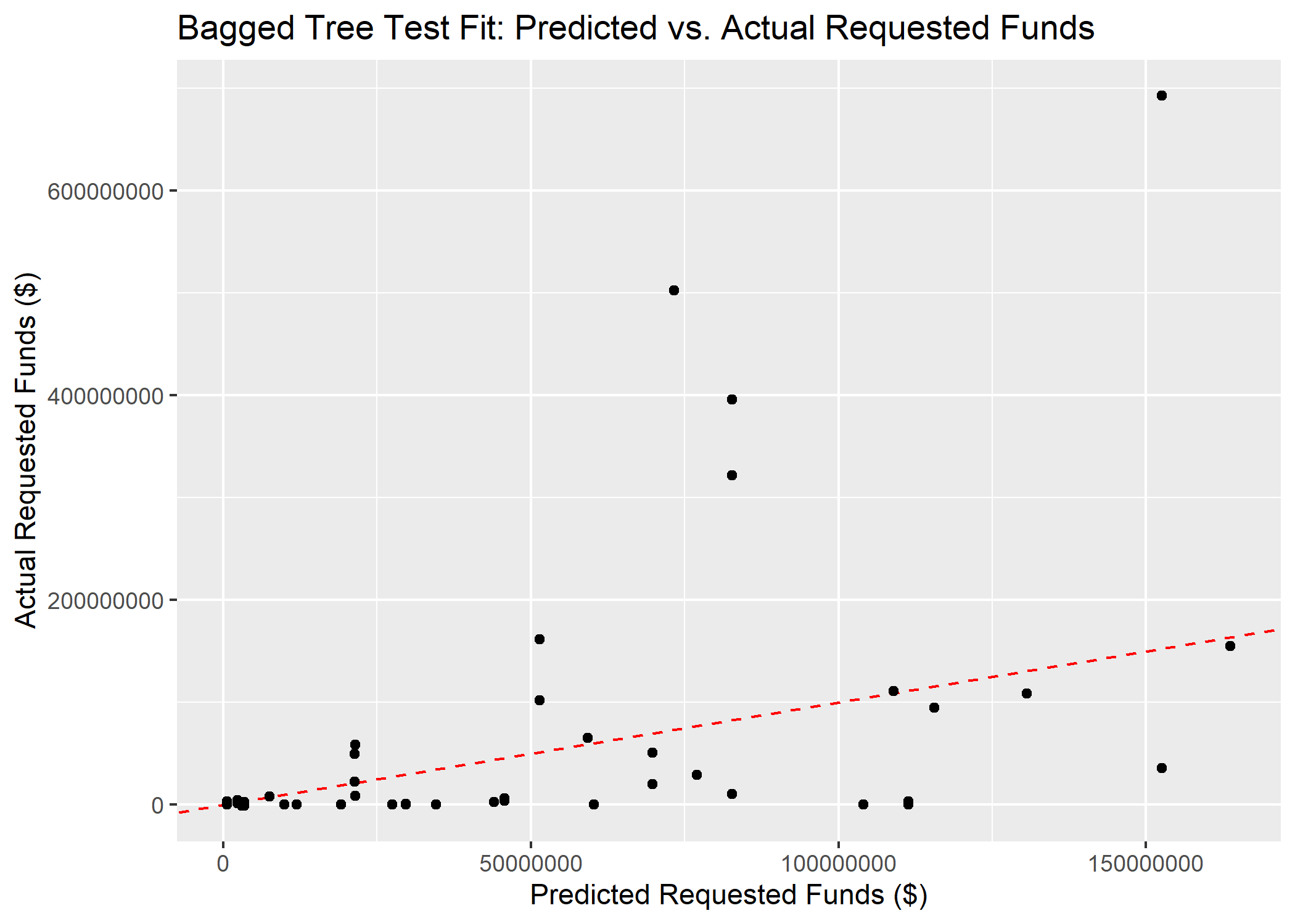
Table 1.5 shows the model comparisons for estimating requested FEMA funding, and Table 1.6 similarly shows the model comparisons for estimating obligated FEMA funding. In both outcomes, the bagged tree had the lowest RMSE; however, the standard error estimate of the RMSE was higher in the bagged tree models than for the other models.

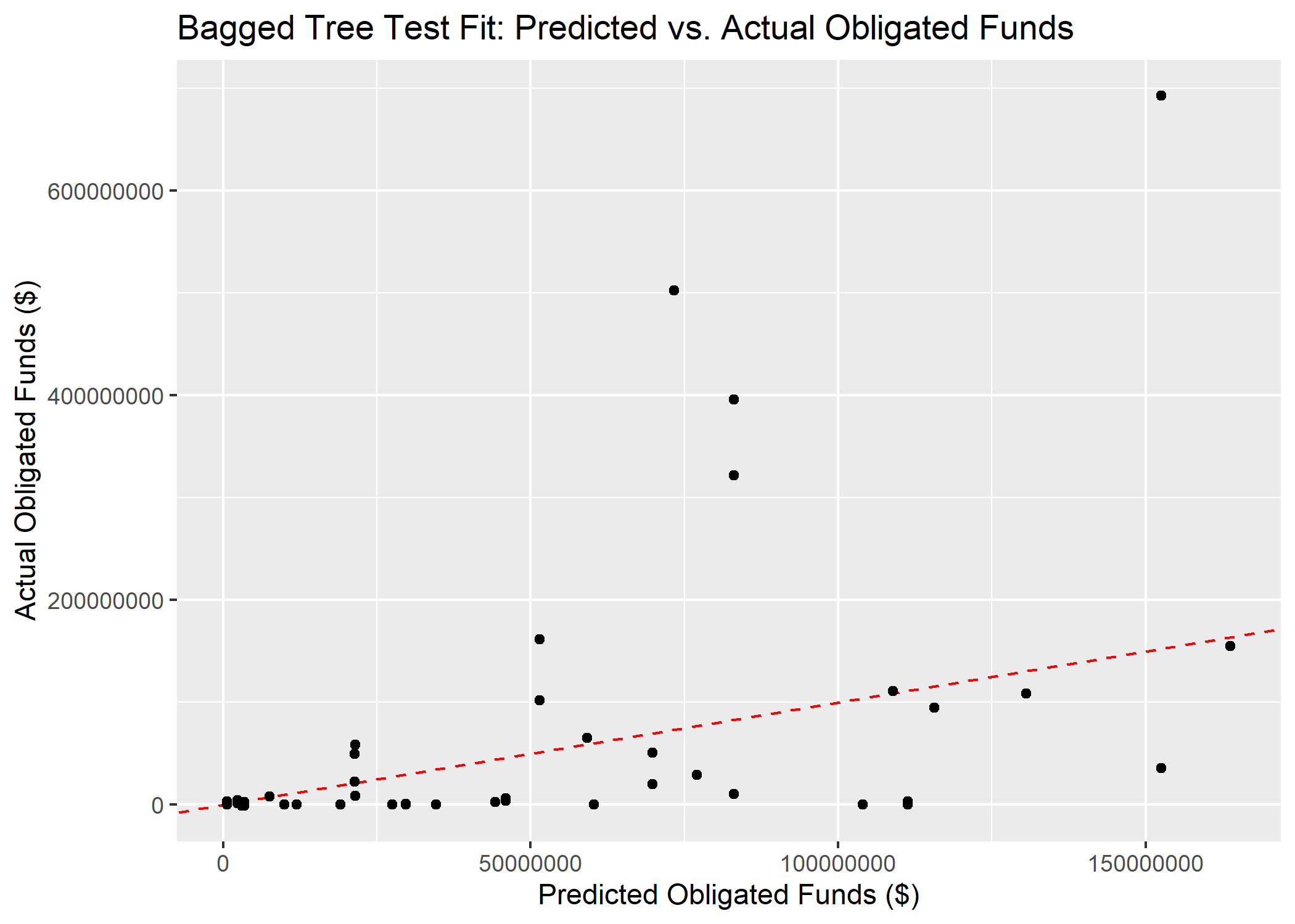
Table 1.5: ML model comparison for requested FEMA funding

| rmse | SE | model |
| --- | --- | --- |
| 5.388000 | 2.3019 | LASSO |
| 5.394000 | 2.3023 | EN |
| 5.552000 | 2.2316 | RF |
| 5.679000 | 2.1802 | Tree |
| 5.276854 | 0.0000 | Null - Train |

Table 1.6: ML model comparison for obligated FEMA funding

| rmse | SE | model |
| --- | --- | --- |
| 44963346 | 5423209 | Bagged Tree |
| 49244359 | 4706033 | RF |
| 53092104 | 4338514 | Tree |
| 64289213 | 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