Exploratory Data Analysis, Regression, Prediction and Optimization Models using National Bridge Inventory Database

Haritha Appojala

January, 2018

A thesis submitted to the faculty of the Graduate School of the
University at Buffalo, State University of New York
in partial fulfillment of the requirements for
the degree of Master of Science
Department of Industrial and Systems Engineering

Acknowledgements

I would first like to thank my thesis advisor, Dr. Jun Zhuang for constant support, guidance and patience. Prof. Zhuang always guided whenever I ran into a trouble spot or had a question about my research or writing. He consistently allowed this paper to be my own work, but guided me in the right direction whenever he thought I needed it. I am thankful for my advisor's direction, devotion and encouragement which kept me focused.

I would like to thank Jessica Brooke, Thomas A. of Argonne National Laboratory for sharing the US National Bridge Inventory data set and letting me work on it.

I would also like to thank Dr. Zhuang's team who were involved in weekly group meetings. Without their passionate participation and inputs for my presentations, my thesis progress would not have been smooth.

Finally, I must express my profound gratitude to my parents and to my friends for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them. Thank you.

Contents

List of Figures

List of Tables

Abstract

The research work on United States National Bridge Inventory (NBI) data is motivated by the state and federal funds allocated by government to roads and bridges (structures). The bridges are widely classified into structurally deficient and functionally obsolete structures based on their performances. The present research takes into account, the historical funding data to assess the correlation between funds being allocated and the performance of bridges. This research includes intensive data analysis and presents regression models and time series models to portray the improvement of bridge performance over years. An attempt is made to recognize the states with higher number of low performance bridges by proposing regression models with funds as input. The results of the regression model showed that the funds are not in proportion of the functional obsolesce of the bridges. So, a mathematical problem is proposed which optimizes the way in which funds are allocated among the states. The optimization model is applied to national bridge inventory data for the year 2014 and the results show that the funds are allocated to the states with higher number of structurally deficient or functionally obsolete structures which in turn increases the performance of bridges. Time Series analysis is performed with the obtained results (optimized funds) by fitting Auto Regressive Integrated Moving Average (ARIMA) and Seasonal Auto Regressive Integrated Moving Average (SARIMA) models. The prediction results shows that the percentage of functional obsolete structures tend to decrease in the next 10 years. A description of the proposed model and its implementation are presented along with a case study. While this research focuses on bridge fund allocation based on sufficiency rating, details on future improvements to incorporate other significant bridge parameters are outlined.

Section 1

Introduction and Literature Review

Bridges play an important role in roadway networks as they connect places decreasing the difficulty to commute. The failure of bridges causes excessive public and private losses. So, it is important to establish an effective maintenance policy to keep bridges sufficiently safe and serviceable throughout their service lives. However, decisions related to the prioritization of bridges for repair purposes and allocation of necessary funds became complex due to the fact that the most existing bridges are old and the funds available for repair are distributed based on state policies. One of the problems related to highway bridges is that the cost of maintaining a network of bridges with an acceptable level of service is more than the available budgeted funds. This results in the increased number of structurally deficient bridges and functionally obsolete bridges.

About 42% of the 574,000 highway bridges in the United States were reported by FHWA to be structurally deficient or functionally obsolete. The performance of Highway Bridges is majorly influenced by the Federal and State funds allocated per each state. These funds are used for the construction of new bridges, and the maintenance and rehabilitation of existing bridges. The condition or performance of the bridge can be evaluated by using the bridge performance indicators explained in FHWA (Federal Highway Administration) Recording and Coding Guide.

The National Bridge Inspection Standards (NBIS), implemented by the Federal Highway Administration, maintains specifications for the inspection and inventory of bridges on public roads. Inspection information is collected annually through this program and is reported to the Federal government. They maintain this information within the National Bridge Inventory (NBI) database. NBI Information, including condition and appraisal ratings, is the basis for the prioritization and allocation of federal funding. The Federal Highway Administration utilizes the NBI information to isolate states eligible for federal funds and to allocate funds to the States. In the initial stages of the program, the focus was to target limited available funds for replacement activity. In course of time, the process was modified to incorporate consideration of bridge rehabilitation activity. Funds are disbursed through the Highway Bridge Repair and Replacement Program (HBRRP) and a Special Bridge Program. HBRRP utilizes the sufficiency rating to allocate funds and the Special Bridge Program utilizes bridge rating to allocate funds. The sufficiency rating formula is calculated utilizing a point deduction system. The formula considers structural adequacy and safety, serviceability and functional obsolescence, and essentiality for public use. A bridge in perfect condition would receive a rating of 100. Deficiencies reduce the score received within each of the categories. The resulting rating is determined through the summation of the structural (55%), functional (30%), and essentiality (15%) scores. Special conditions, such as an exceptionally large detour length, may reduce the sufficiency rating further. Bridges with a rating of 50 or less are eligible for replacement funding. A sufficiency rating between 50 and 80 designates eligibility for rehabilitation funds. Each State has flexibility in selecting bridge projects subject to these eligibility constraints. Approximately 20 percent of all bridges (excluding culverts which are treated equivalently for funding purposes) are eligible for replacement funds. In addition, approximately 35 percent of bridges have a sufficiency rating between 50 and 80 and are eligible for rehabilitation funds. The Special Bridge Program, considers the sufficiency rating, the total project cost, Average Daily Traffic (ADT) and Average Daily Truck Traffic (ADTT), and the funding disbursed through the HBRRP program.

Bridge maintenance costs have increased in several developed countries. Now it is more expensive to maintain damaged bridges than to build new ones. Several bridges have deteriorated considerably since years due to factors such as the increase in traffic volume, weather conditions, the increase in the weights of vehicles and the structural aging. However, due to budget limitations, funds for such rehabilitation must be taken from sources that were originally procured for the construction of new bridges.

The efficient use of funds for the well being of bridge networks requires an effective bridge asset management technology. It is particularly important to optimize future bridge maintenance, repair and rehabilitation activities with the given funding and to request suitable future funding based on reliable Bridge Management System (BMS) outcomes. A BMS is a computer-based decision support system (DSS) which is used to determine the best possible strategy that ensures an adequate level of safety for bridges at the lowest possible life cycle cost. Based on the results of a deterioration model, BMS provides various important future estimations for the planning of maintenance, repair and rehabilitation activities.

This research is mainly focused on characterizing the influence of funds allocated on Bridge performance. Also, the relationship between various bridge performance indicators and funds allocated is extracted. This report walks through the process of data filtering and data analysis with supporting visualizations (plots and graphs). Taking into account, the results of data analysis, regression models and time series models are proposed to assess the bridge performance. The scope of the research can be extended by introducing regression models with funds as an input and creating optimization/Mathematical model to optimize the funds allocated for respective States.

Section 2

Data Analysis

2.1 Initial data

The National Bridge Inventory (NBI) data provides various bridge performance indicators. The explanation of these bridge performance variables is explained in FHWA Recording and Coding Guide. The bridge data for each state is obtained in ASCII format. The expenditure data comprises the federal funds, state funds allocated and the money spent on physical maintenance, construction, reconstruction, rehabilitation etc. This expenditure data is acquired from Census data and USDOT FHWA Highway Statistics.

2.2 Interpretation of NBI Data

Each Bridge performance indicator is clearly understood from FHWA Recording and Coding Guide. After careful consideration of the description various variables, we selected the following items as key variables for initial analysis.

- Item 1 – State code
- Item 27 – Year Built
- Item 29 – Average Daily Traffic (ADT)
- Item 30 – Year of Average Daily Traffic

- Item 42 – Type of Service
- Item 67 – Structural Evaluation
- Item 94-97 – Improvement costs
- Item 106 – Year Reconstructed
- Item 113 – Scour Critical Bridges
- Item SR - Sufficiency Rating (SR represents sufficiency rating)

2.2.1 State Code, Year Built, ADT and Year of ADT (Average Daily Traffic)

State code is assigned as per FIPS (Federal Information Processing Standards) codes for States and FHWA (Federal Highway Administration) Region codes. Year built represents the year in which construction of the structure was completed. Average daily traffic (ADT) is the daily traffic volume for the inventory route. ADT is calculated based on the Average Daily Truck Traffic corresponding to a bridge. If the bridge is closed, ADT is collected from before the closure occurred. The year in which Average Daily Traffic is collected is termed as Year of ADT.

2.2.2 Type of Service

Type of Service is of two kinds:

- Type of Service on Bridge - The type of service on bridge is explained using the codes shown in the Table 2.1. Here the codes 1, 4, 5 are considered for Data Analysis as the research is focused on Highway Bridges (see Table 2.1). The codes includes the highways with rail road and pedestrian service.
- Type of Service Under Bridge -The type of service under the bridge includes the codes shown in the Table 2.1. Here the codes 1, 4, 6, 8 are considered based on Highway bridges. These codes include the services under the bridge like highways with or without pedestrians, waterways and railroads under the bridges.

2.2.3 Structural Evaluation

The structural evaluation of a Bridge iis defined as the classification of bridges into either Structurally Deficient or Functionally Obsolete based on condition rating and appraisal rating of items

Code	Type of service on Bridge-Description
1	Highway
2	Railroad
3	Pedestrian - bicycle
4	Highway - railroad
5	Highway -pedestrian
6	Overpass structure at an interchange or second level of a multilevel interchange
7	Third level (Interchange)
8	Fourth level (Interchange)
9	Building or Plaza
Code	Type of service under Bridge-Description
1	Highway , with or without pedestrian
2	Railroad
3	Pedestrian - bicycle
4	Highway - railroad
5	Waterway
6	Highway - Waterway
7	Railroad - Waterway
8	Highway - Waterway - railroad
9	Relief for waterway

Table 2.1: Codes and Description of Type of Service of the Bridges

specified in Reporting and Coding guide by FHWA.

- Structurally Deficient

    1. A condition rating of 4 or less for

        - Item 58 - Deck; or
        - Item 59 - Superstructures; or
        - Item 60 - Substructures; or
        - Item 62 - Culvert or Retaining Walls

    2. An appraisal rating of 2 or less for

  - Item 67 - Structural Condition; or
  - Item 71 - Water Adequacy

- Functionally Obsolete

  1. An appraisal rating of 3 or less for
     - Item 68 - Deck Geometry; or
     - Item 69 - Under Clearances; or
     - Item 72 - Approach Roadway alignment; or
  2. An appraisal rating of 3 for
     - Item 67 - Structural Condition; or
     - Item 71 - Water Adequacy

2.2.4 Improvement Costs

Improvement costs include Bridge Improvement Cost, Roadway Improvement Cost and Total Project Cost.

Bridge Improvement cost is the estimated cost of the proposed bridge or major structure improvements.This cost includes only bridge construction costs, excluding roadway, right of way, detour, demolition, preliminary engineering, etc.

Roadway improvement costs represent the cost of the proposed roadway improvement. This includes only roadway construction costs, excluding bridge, right-of-way, detour, extensive roadway realignment costs, preliminary engineering, etc. Roadway improvement costs are calculated for bridges eligible for the Highway Bridge Replacement and Rehabilitation Program. In the absence of a procedure for estimating roadway improvement costs, a guide of 10 percent of the bridge costs is suggested.

2.2.5 Year Reconstructed

Year reconstructed is the year of most recent reconstruction of the structure. A bridge is defined as reconstructed if the type of work performed is eligible for funding under any of the Federal-aid funding categories. The eligibility criteria would apply to the work performed regardless of whether all State or local funds or Federal-aid funds were used.

2.2.6 Scour Critical Bridges

Scour critical bridges describes a code for each bridge based on the location of waterway. The codes ranges from 0 to 9 and N,U,T (see Table 2.2).

Code	description
N	Bridge not over waterway.
U	Bridge with "unknown" foundation that has not been evaluated for scour.
T	Bridge over "tidal" waters that has not been evaluated for scour, but considered low risk.
9	Bridge foundations on dry land well above flood water elevations.
8	Bridge foundations determined to be stable for assessed or calculated scour conditions.
7	Countermeasures have been installed. Bridge is no longer scour critical.
6	Scour calculation/evaluation has not been made.
5	Bridge foundations determined to be stable for calculated scour conditions.
4	Bridge foundations determined to be stable for calculated scour conditions.
3	Bridge is scour critical;bridge foundations determined to be unstable.
2	Bridge is scour critical;Immediate action is required to provide scour countermeasures.
1	Bridge is scour critical; field review indicates that failure is imminent. Bridge is closed to traffic.
0	Bridge is scour critical. Bridge has failed and is closed to traffic.

Table 2.2: Code Description for Scour Critical Bridges

2.2.7 Sufficiency Rating

This variable is used as one of the primary performance indicators of resilience. Bridge Sufficiency Rating is an aggregate measure of a bridge's sufficiency to remain in service. Highway bridge data is evaluated by calculating four separate factors.

1. S_1 - Structural Adequacy and Safety
2. S_2 - Serviceability and Functional Obsolescence
3. S_3 - Essentiality for Public Use
4. S_4 - There are also special reductions (up to 13%) related to detour length,traffic safety features and structure types.

The maximum value of S_1 is 55%, S_2 is 30% and S_3 is 15%. S_4 can take values upto 13%. Now, sufficiency rating is calculated by $(S_1 + S_2 + S_3 - S_4)$. A detailed description of the calculation of sufficiency rating is given in Recording and Coding guide for Nation's bridges.The result is a percentage in which 100 percent represents an entirely sufficient bridge and zero percent represents an entirely insufficient or deficient bridge.

Highway Bridges considered structurally deficient or functionally obsolete with a sufficiency rating less than 80 were considered for the selection list. The bridges with sufficiency rating less than 80 are considered structurally deficient and those bridges with sufficiency rating less than 50 are considered as functionally obsolete. Based on this classification, those bridges appearing on the list with Sufficiency rating less than 50 are eligible for replacement or rehabilitation and those bridges with sufficiency rating less than 80 are considered for rehabilitation.

2.3 Data Cleansing and Visualization

The initial NBI data was in ASCII format. This raw ASCII data is converted to .csv files by using python script. The bridge data for all the states in Unites States for the years between 1994 and 2015 are generated. For the preliminary analysis, the data of the year 2014 is considered. The data for the year 2014 is filtered by considering the key variables and appropriate visualizations are made. Also, the historical data for the years 1994-2014 is aggregated from individual years and the trends over years are visualized.

2.3.1 Bridges in Each State

As an initial step, the number of bridges in each state is analyzed and a plot is made. From the Figure(2.1), it is seen that, maximum number of bridges are present in California, Tennessee, New York, New Mexico and Pennsylvania.

2.3.2 Relationship between Number of Structures and Funds Allocated

A scatter plot for number of bridges and federal funds is plotted to see the relation between the funds allocated and the bridges present in each state (see Figure 2.2). A spatial plot is made to represent the number of bridges and the federal funds in each state (see Figure 2.3). The color gradient from red to green represents the increase of funds from lower value to higher value. Moreover, the larger the size of the blob, larger is the number of bridges in each state.

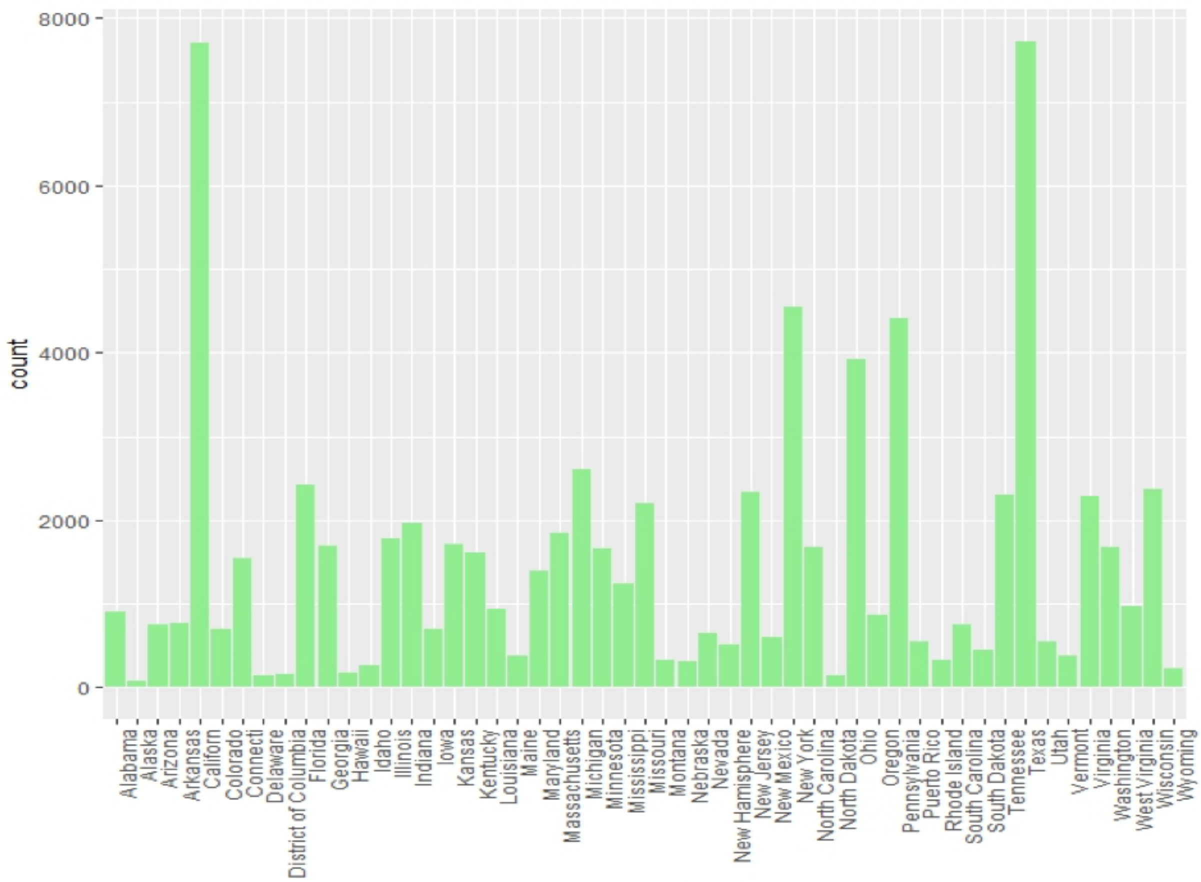


Figure 2.1: Number of Bridges in Each State (2014)

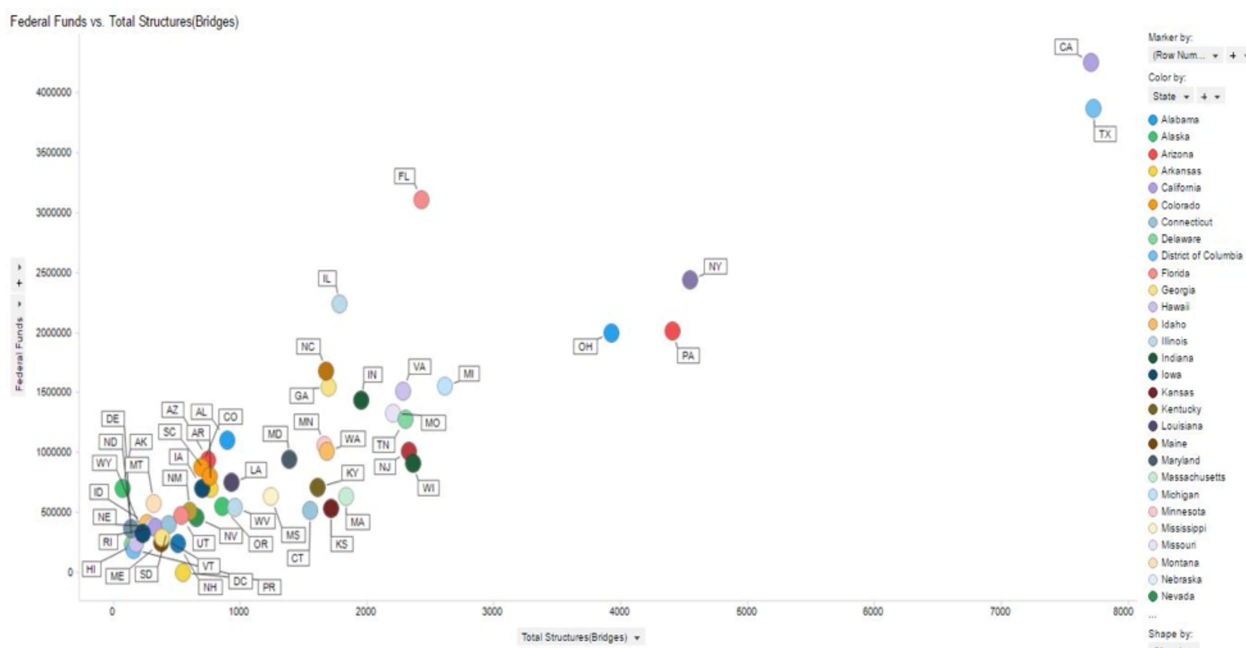


Figure 2.2: Federal Funds vs Number of bridges

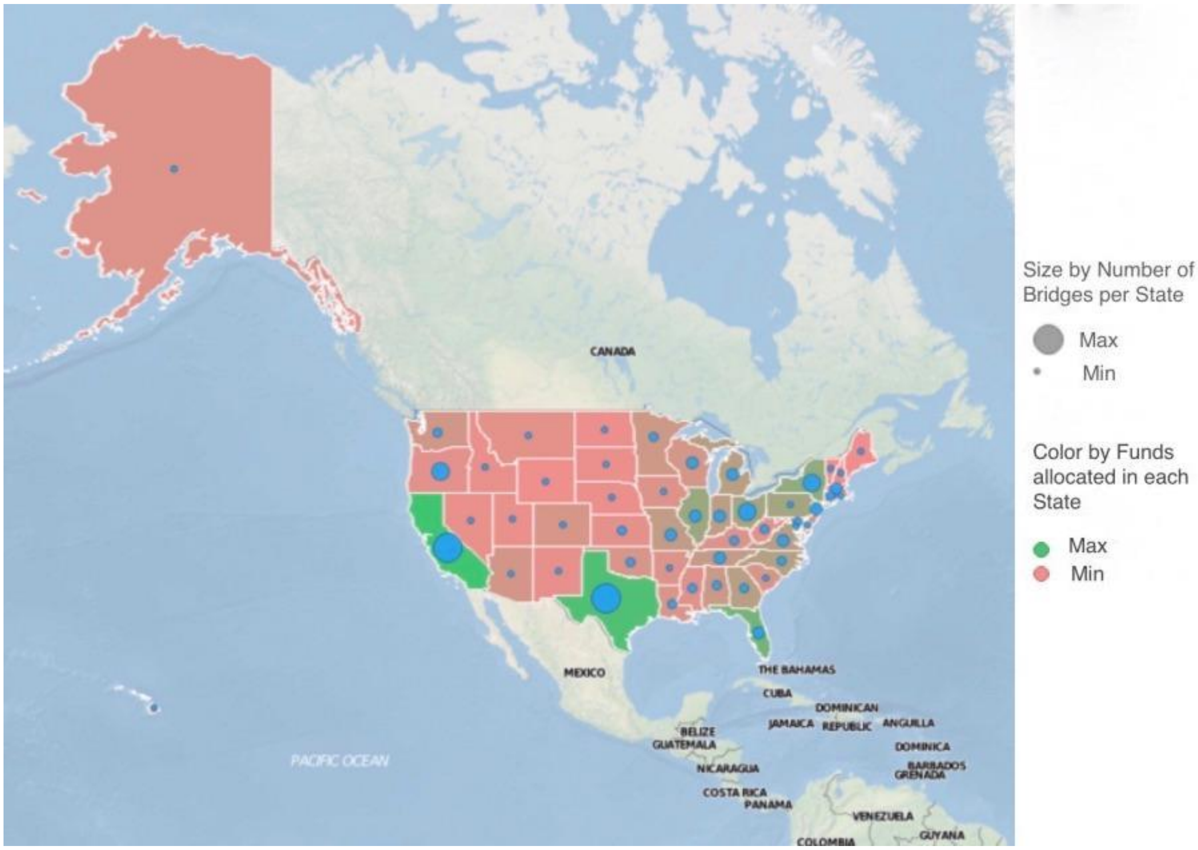


Figure 2.3: Distribution of Bridges and Funds

From the two plots it is seen that, the federal funds and the number of structures are linearly related.

2.3.3 Bridges based on Sufficiency Rating

Sufficiency rating plays a key role in the performance of bridges. We classified the bridges into two categories based on sufficiency rating- Structurally Deficient and Functionally Obsolete. Besides analyzing the total number of bridges in each state, we calculated the percentage of number of structurally deficient and functionally obsolete structures and plotted the results on the graph. The plot (see Figures 2.4, 2.5) is made in such a way that the decreasing order of percentage of structurally deficient structures as per the state code is visualized.

2.3.4 Bridge Improvement Costs and Structurally Deficient Structures

The relationship between bridge improvement costs and structurally deficient structures is obtained through a scatter plot (see Figure 2.6). While plotting, the state NY is ignored because the bridge

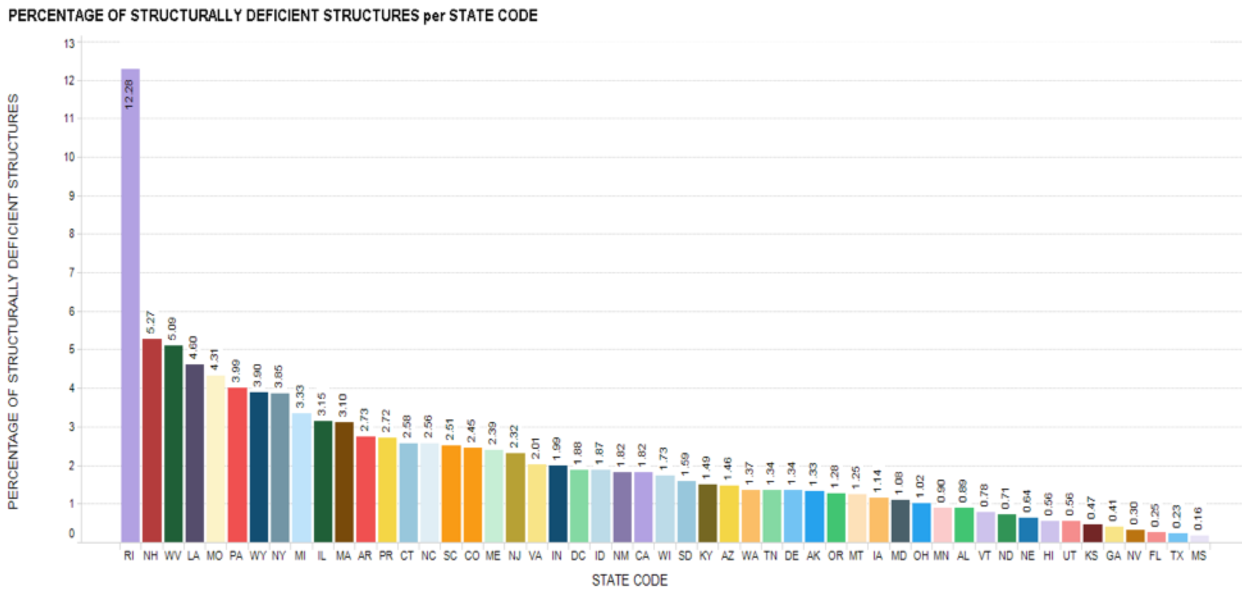


Figure 2.4: Percentage of Structurally Deficient Structures

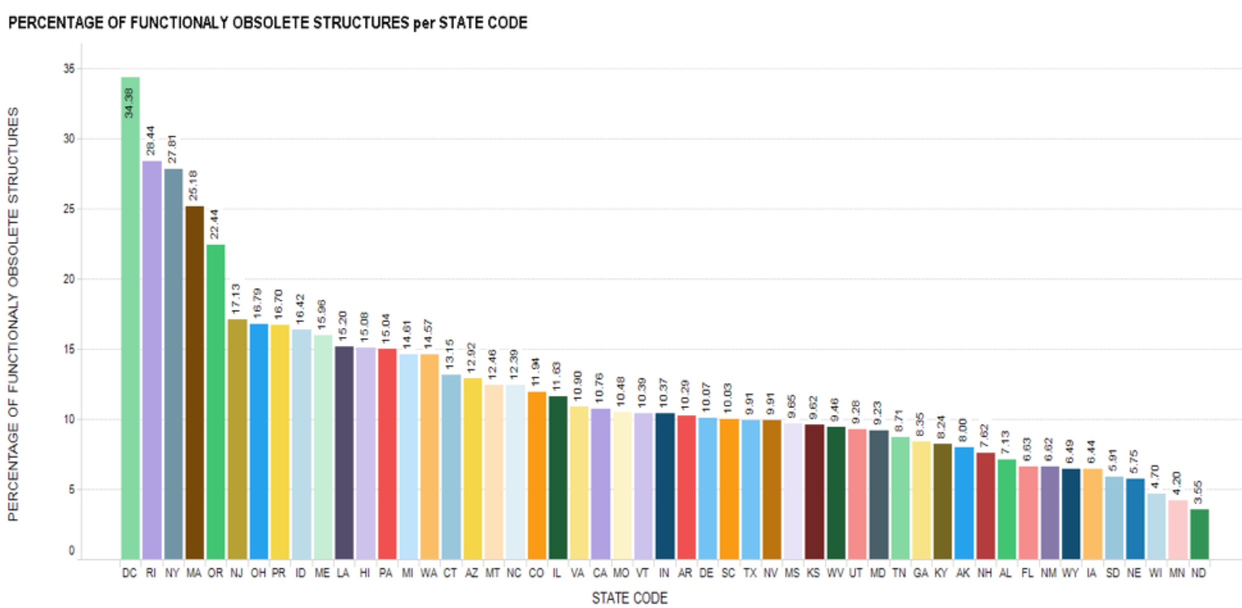


Figure 2.5: Percentage of Functionally Obsolete Structures

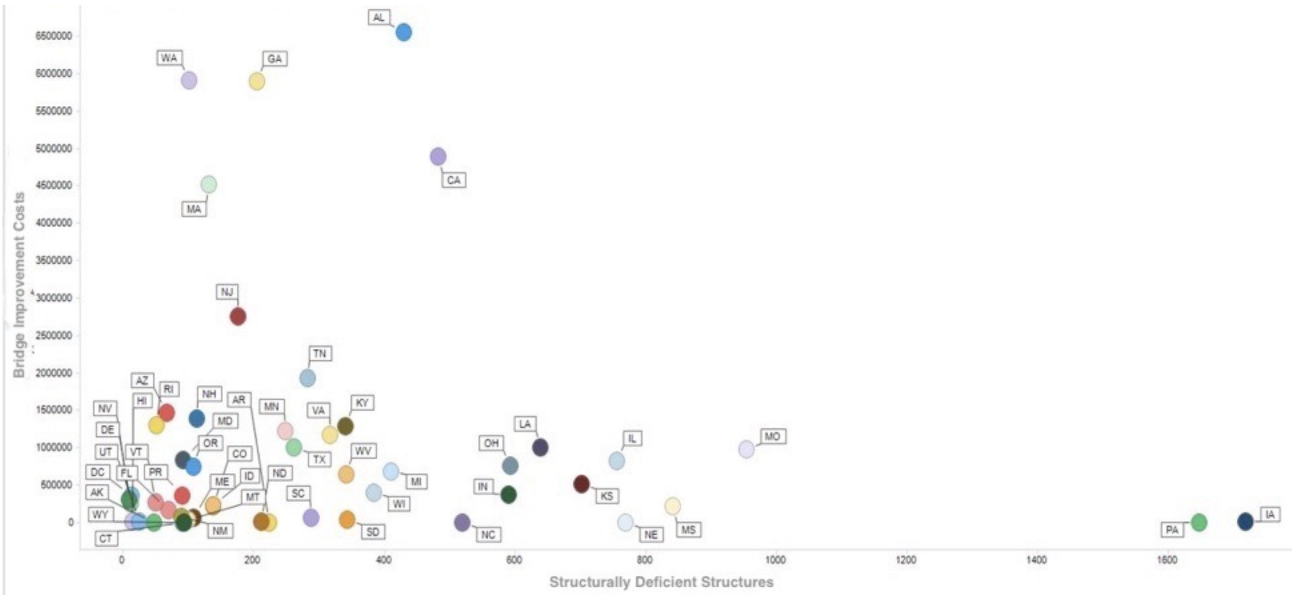


Figure 2.6: Bridge Improvement Costs vs Structurally Deficient Structures

improvement costs were quite high in NY. From the scatter plot, it is observed that the bridge improvement costs are not proportionate. For example, consider Pennsylvania(PA) from the plot. It is seen that the bridge improvement costs are very less even though the structurally deficient structures are high. On the contrary, bridge improvement costs are quite high in Washington(WA) for minimal structurally deficient structures.

2.3.5 Roadway Improvement Costs and Structurally Deficient Structures

Roadway improvement costs represent the cost of the proposed roadway improvement in thousands of dollars. This shall include only roadway construction costs, excluding bridge, right-of-way, detour, extensive roadway realignment costs, preliminary engineering, etc. The relationship between roadway improvement costs and structurally deficient structures is obtained through a scatter plot (see Figure 2.7). From the scatter plot, it is observed that the Roadway improvement costs are not proportionate. For example, consider the state Iowa(IA) from the plot. It is seen that the roadway improvement costs are very less even though the structurally deficient structures are high. On the contrary, roadway improvement costs are quite high in Georgia(GA) for less number of structurally deficient structures.

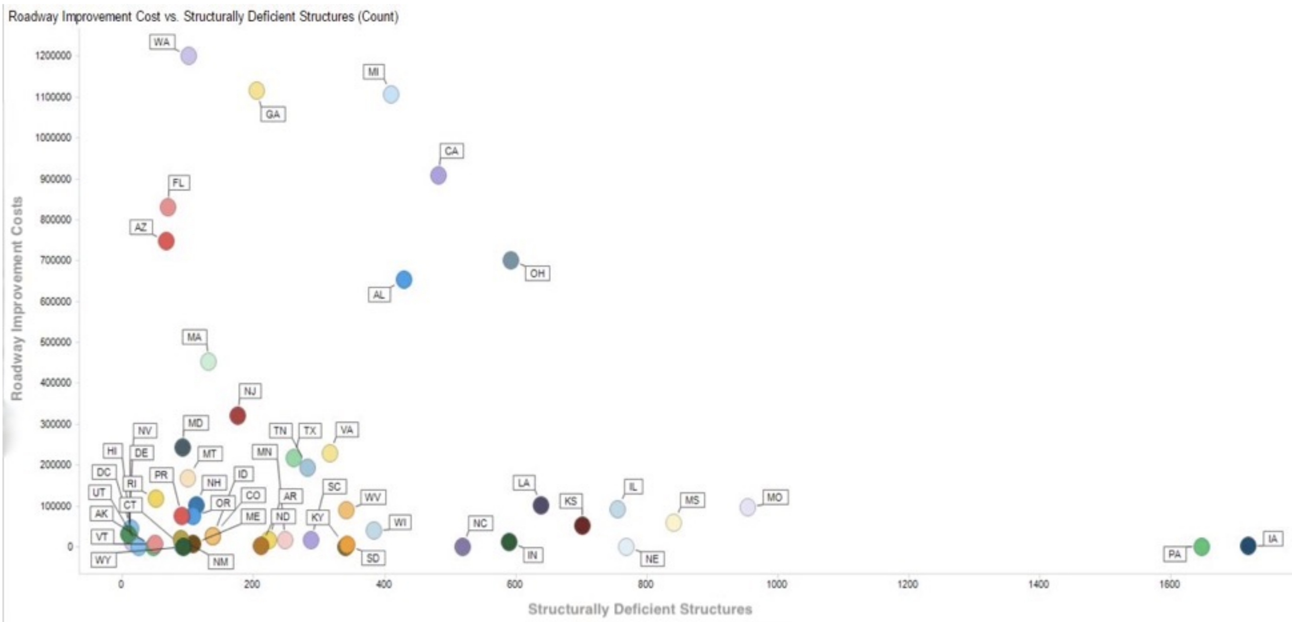


Figure 2.7: Roadway Improvement Costs vs Structurally Deficient Structures

2.3.6 Trend of Funds

The historical data for the funds in all states from 1994 to 2014 is collected from Census data of State and local expenditure of all states. The following Figure (2.8) show the trend of total funds, federal funds and state funds from 1994-2014. It is seen that federal funds are allotted in the same way in all the years and the state funds followed a linear trend.

2.3.7 Trend of Percentage of Improvement Costs

The percentage of total improvement costs is also plotted. Here the percentage of total improvement costs is calculated by

$$\%TotalImprovementCosts = \frac{TotalImprovementCosts}{Totalfunds} X100$$

From the Figure (2.9), It is observed that there is a spike in the year 2007.

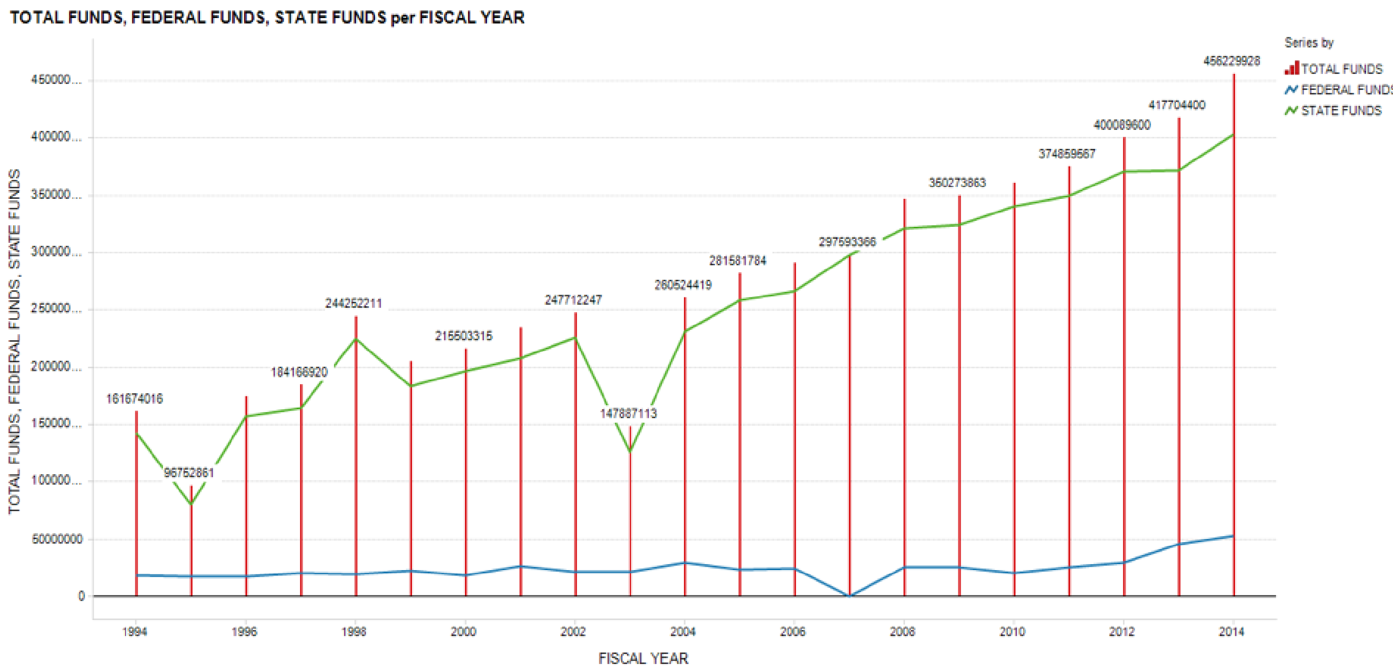


Figure 2.8: Trend of Funds (1994-2014)

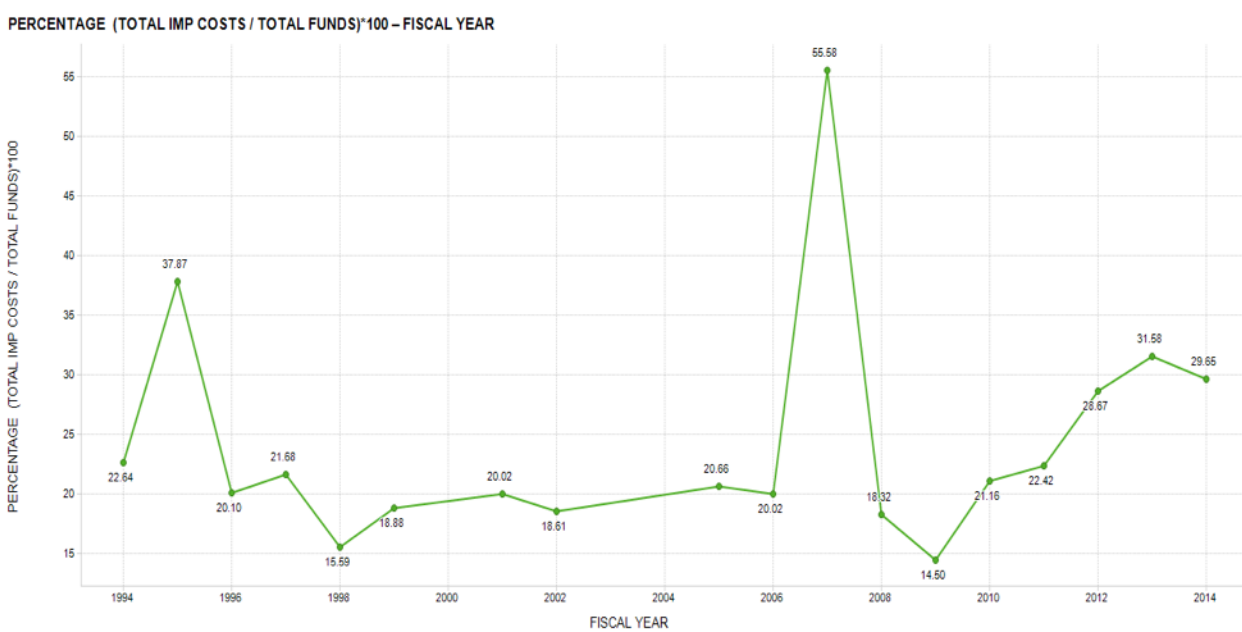


Figure 2.9: Percentage of Improvement Costs (1994-2014)

Reason for Spike in Improvement Costs

The reason for spike in 2007 is that, there is no Federal Funds Data available for the fiscal year 2007. Percentage of improvement costs are calculated using

$$\frac{Total\ Improvement\ Costs}{Total\ funds} X100$$

And Total funds is the sum of Federal and State funds. Due to unavailability of federal fund data for the year 2007, the denominator (Total funds) decreased. This resulted in increase of percentage of improvement costs in 2007.

2.3.8 Minimum and Maximum Age of Bridges in each State

The age of bridges in each state is calculated by considering the year of construction.

$$Age\ of\ Bridge = Current\ Year - Year\ of\ Construction$$

Then the minimum age and maximum age of bridges in each state is identified and plotted (see Figure 2.10).

2.3.9 Reconstructed Bridges

A plot is made to see the trend of number of bridges reconstructed between 1920 and 2015 (see Figure 2.11). It is seen that the number of bridges reconstructed is high during 1985-1995. Similarly, a bar plot is made for structure reconstructed in each state to extract the decreasing order of states with reconstructed bridges (see Figure 2.12). By using these factors, bridges that are not renovated or reconstructed since the time they are constructed were extracted. Structurally Deficient and Functionally Obsolete structures are identified among those bridges, so that necessary measures can be adopted (federal and state funds) for the reconstruction of those bridges to ensure safety.

- Highway Bridges considered Structurally Deficient or Functionally Obsolete with a sufficiency rating of 80 or less will be eligible for rehabilitation.
- And those with sufficiency rating less than 50 will be eligible for replacement.

The Table 2.3 shows the number of bridges that are not reconstructed since their construction and are eligible for replacement.

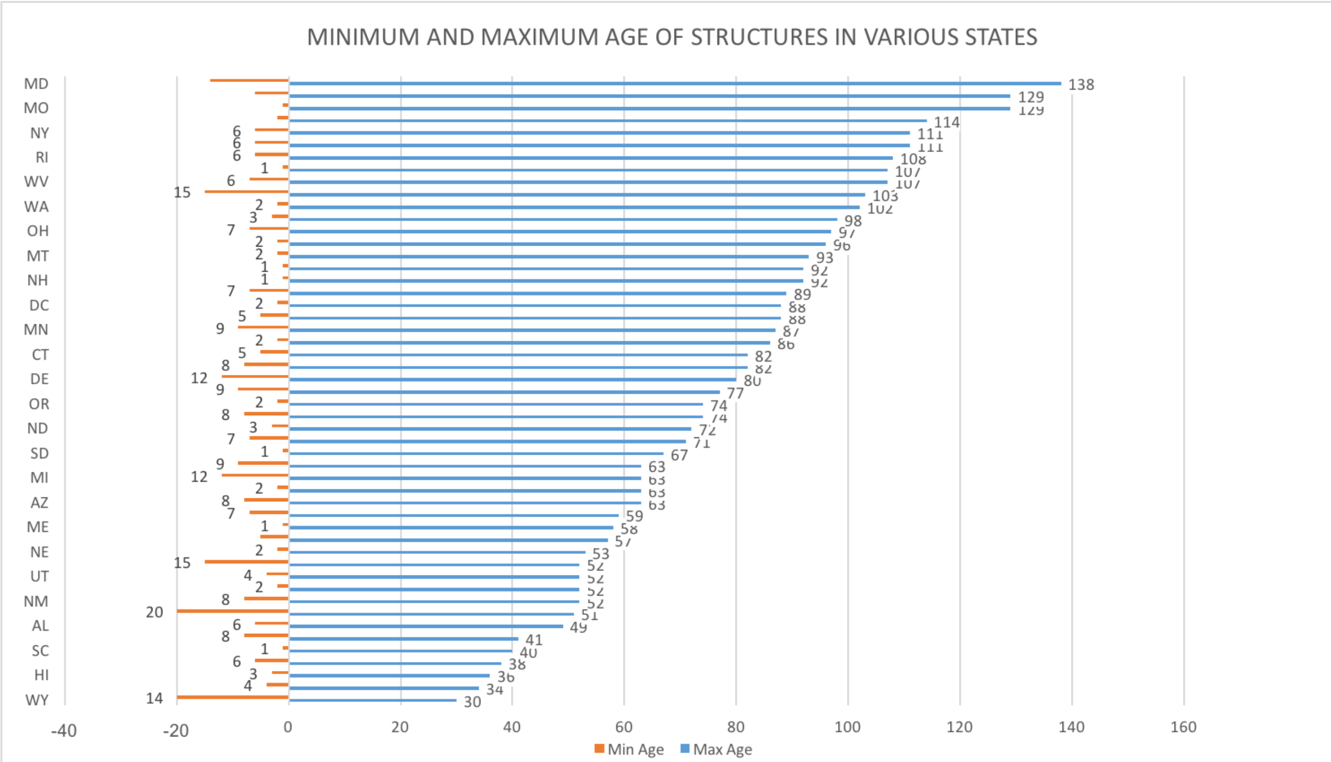


Figure 2.10: Minimum and Maximum age of Bridges

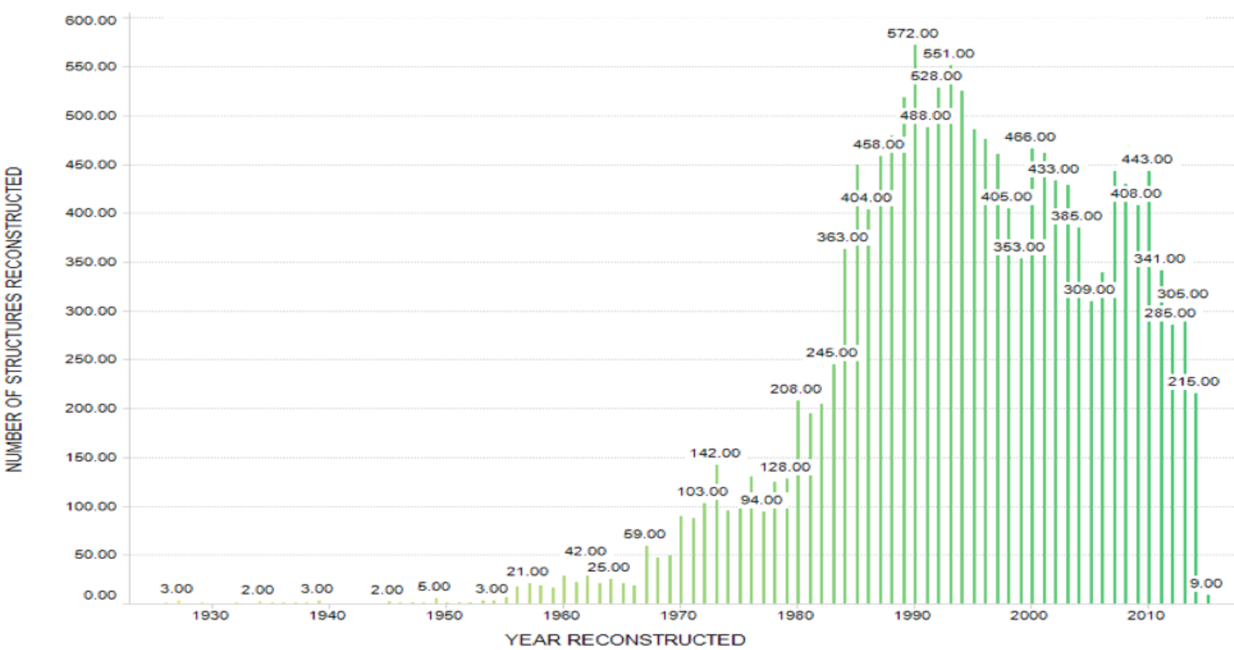


Figure 2.11: Bridges Reconstructed (1920-2015)

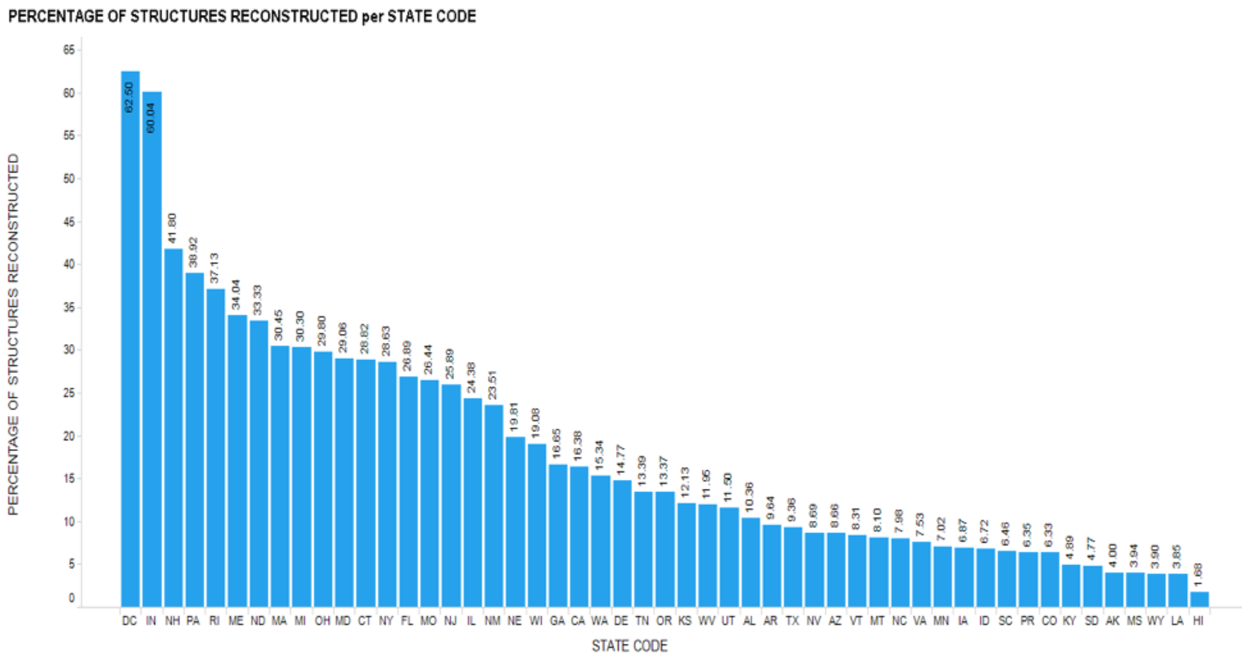


Figure 2.12: Percentage of Bridges Reconstructed

StateCode	Number of Bridges Eligible for Replacement
AR	2
GA	1
IA	1
IL	1
LA	4
MI	1
MO	5
NC	3
NH	2
NJ	2
NY	4
PA	6
RI	1
TN	2
VA	2
WA	1

Table 2.3: Number of Bridges Eligible for Replacement in each State (Note : States not included in the table have 0 in second column)

Map Chart

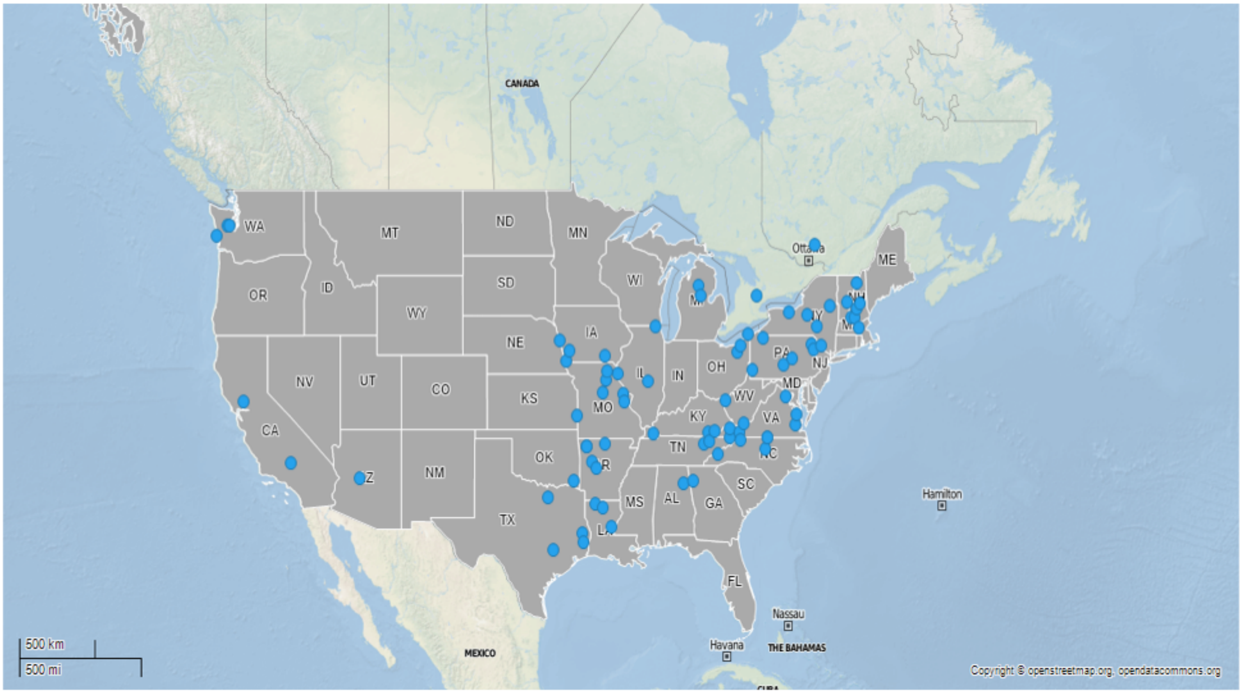


Figure 2.13: Spatial Analysis showing Location of Bridges to be Replaced/Rehabilitated

An attempt has been made to represent the number of bridges that are to be rehabilitated/ reconstructed on the map of United States (see Figure 2.13).

Section 3

Regression Analysis - California

3.1 Calculating Funds with Inflation Values

The Funds for the state CA are calculated by including the inflation rates (by considering 2014 as reference). The funds of all the years are converted by taking the year 2014 as reference. A graph is plotted to observe the trend of funds with inflation rate included (see Figure 3.1).

3.2 Bridge Variables included in the Analysis

As first step in regression analysis, we filtered out some bridge variables (by using the description given in "Recording and Coding Guide" by U.S. DOT) which have trivial effect on Funds Allotted. Example: StateCode, Structure number, Inventory route, Record type, Route signing Pref, Route Number etc.

The bridge variables that were filtered from analysis explained codes, numbers, material of structure, direction of routes, location, latitudes, longitudes. Also we filtered variables that explained performance/Resilience of the bridges because inclusion of sufficiency rating compensates those variables.

Bridge variables included in the Analysis:

- Sufficiency Rating
- Year Built

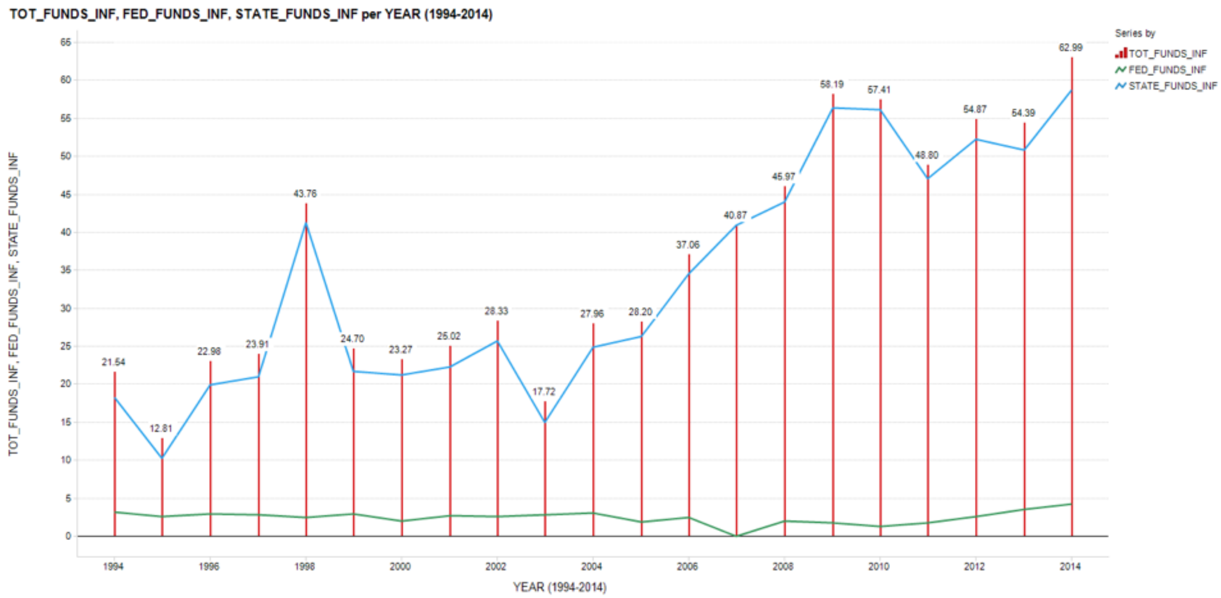


Figure 3.1: Trend of Funds with Inflation Rate included

- Year Reconstructed
- Base Highway Network
- Maintenance Responsibility
- Functional Class of Inventory Route
- Average Daily Traffic
- Historical significance
- Structure Open/Posted/Closed
- STRAHNET Highway Designation
- Highway System of Inventory Route

3.2.1 Sufficiency Rating

Sufficiency Rating is explained in the previous Section (2.2.7) of the document.

3.2.2 Year Built and Reconstructed

Year Built is the year in which the structure is constructed and the Year Reconstructed is the year in which the structure is reconstructed(refer to Section 2.2.5).

3.2.3 Base Highway Network

This variable is used to evaluate whether funding priorities are given to roadways that are part of the National Highway System (NHS), which may also help to reconcile any trends or influences of state vs. federal funding sources.

- Inventory Route is not on the Base Network
- Inventory Route is on the Base Network

3.2.4 Maintenance Responsibility

This variable is used to evaluate whether funding priorities are given to roadways based upon the owner of the roadway. This may also be used to reconcile any trends or influences of state vs. federal funding sources, or funding provided by other agencies. For example, tollways may receive different prioritization as they are facilities that generate their own funding streams, which may only require supplemental funding from other sources.

- 01 = State Highway Agency
- 02 = County Highway Agency
- 03=Town or Township Highway Agency
- 04 = City or Municipal Highway Agency
- 11 = State Park, Forest, or Reservation Agency
- 12 = Local Park, Forest, or Reservation Agency
- 21 = Other State Agencies
- 25 = Other Local Agencies
- 26 = Private (other than railroad)
- 27 = Railroad
- 31 = State Toll Authority
- 32 = Local Toll Authority

- 60 = Other Federal Agencies
- 61 = Indian Tribal Government
- 62 = Bureau of Indian Affairs
- 63 = Bureau of Fish and Wildlife
- 64 = U.S. Forest Service
- 66 = National Park Service.
- 67 = Tennessee Valley Authority
- 68 = Bureau of Land Management
- 69 = Bureau of Reclamation
- 70 = Corps of Engineers (Civil)
- 71 = Corps of Engineers (Military)
- 72 = Air Force
- 73 = Navy/Marines
- 74 = Army
- 75 = NASA
- 76 = Metropolitan Washington Airports Service
- 80 = Unknown

Here Maintenance responsibility is divided into three levels.

- **Level 1** 01 = State Highway Agency
- **Level 2**

  02 = County Highway Agency

  03 = Town or Township Highway Agency

  04 = City or Municipal Highway Agency
- **Level 3** Rest everything (one level)

3.2.5 Functional Class of Inventory Route

This variable is used to evaluate whether or not funding priorities are given to more major roadways (e.g., interstates vs state roadways). This will be used in conjunction with other similar variables in anticipation of issues with data completeness or adequacy.

Rural:

- Principal Arterial - Interstate
- Principal Arterial - Other
- Minor Arterial
- Major Collector
- Minor Collector
- Local

Urban:

- Principal Arterial - Interstate
- Principal Arterial - Other Freeways or Expressways
- Other Principal Arterial
- Minor Arterial
- Collector
- Local

Functional class of inventory route is divided into different levels.

- **Level 1**

  01 = Rural

  02 = Principal Arterial – Interstate

- **Level 2**

  11 = Urban

12 = Principal Arterial – Interstate

Principal Arterial - Other Freeways or Expressways

- **Level 3** Rest everything (one level)

3.2.6 Average Daily Traffic

This variable is used to evaluate whether or not funding priorities are given to bridges that carry higher traffic volumes. ADT is a measure of traffic (veh/day), that is collected for a specified period of time that is ranges from a month to less than a year. ADT may reflect some seasonal variability in volumes.

3.2.7 Historical Significance

This variable is used to evaluate whether or not funding priorities are given to bridges with historical significance.

- Bridge is on the National Register of Historic Places.
- Bridge is eligible for the National Register of Historic Places.
- Bridge is possibly eligible for the National Register of Historic Places (requires further investigation before determination can be made) or bridge is on a State or local historic register.
- Historical significance is not determinable at this time. Bridge is not eligible for the National Register of Historic Places.

3.2.8 Structure Open/Posted/Closed

This variable is primarily evaluated for the "P" value which would indicated some extent of functional obsolescence (i.e., it is not able to carry current-day design loads). This variable then becomes somewhat of a proxy for functional obsolescence in the absence of data in related fields (i.e., data completeness issues).
P = Posted for load (may include other restrictions such as temporary bridges which are load posted)

3.2.9 STRAHNET Highway Designation

This variable is used to evaluate whether funding priorities are given to roadways of national importance (STRAHNET is the Strategic Highway Network, a designation given to roadways by the US Department of Defense indicating that roadways are essential for providing access to critical national defense facilities, or facilitating the movement of goods for national security and defense purposes).

- The inventory route is not a STRAHNET route.
- The inventory route is on a Interstate STRAHNET route.
- The inventory route is on a Non-Interstate STRAHNET route.
- The inventory route is on a STRAHNET connector route.

3.2.10 Highway System of Inventory Route

This variable is used to evaluate whether funding priorities are given to roadways that are part of the National Highway System. This will be used in conjunction with other similar variables in anticipation of issues with data completeness or adequacy.

- Inventory Route is not on the NHS
- Inventory Route is on the NHS

3.3 Data Extraction and Cleansing

Initially we extracted the data for the state of California (CA) from 1994-2014 using excel. The average age of the bridges is assumed to be 45 years. The data had missing values causing hindrance for the analysis. Hence cleansed the data by removing missing value rows. The data thus obtained is processed using Microsoft excel and R programming to get the values for the selected bridge variables (refer to section 3.2). There is no data pertaining to the predictor variables for the years 2004,2003,2000 in the base dataset and hence the values of predictor variables for the years 2000, 2003 and 2004 were ignored. There are few zeros (0) in the data in the columns Base Highway Network and ADT. The presence of these zeros in the data affects the fit of the regression model.So we replaced 0 in Base Highway Network with NAs and extracted the data for the missing data of

ADT from the later years. We also observed that for the year 2007, there is spike in data (refer Section 2.3.7) which resulted in outliers. To remove the outliers, the year 2007 is ignored for the further analysis. After preparing the data, we created linear regression models which are explained in the next section.

3.4 Regression Models

1. A linear regression model (3.2) with Total Funds as the input or scalar dependent variable is created to model the relationship between total funds and the other variables used in regression analysis. The results of the following regression model are explained in Section 3.6.

```
> linearCAedit<-lm(TotalFunds~.-Federal-State,data=regressionCA,na.action=na.exclude)
> summary(linearCAedit)

Call:
lm(formula = TotalFunds ~ . - Federal - State, data = regressionCA,
    na.action = na.exclude)

Residuals:
         1          2          3          4          5          6          7          8
   -5460.9     6048.4     -924.5     -927.2     4657.3   -20952.9    22271.8  4150987.5
         9         10         11         12         13         14
-4151200.5     2352.7    19925.4   -30798.9     6827.5    -2805.5

Coefficients:
                                    Estimate Std. Error t value Pr(>|t|)
(Intercept)                       -1.033e+08  8.358e+07  -1.236    0.342
SuffRating.80                      1.572e+06  1.132e+07   0.139    0.902
SuffRating.50                      1.914e+06  1.454e+07   0.132    0.907
AvgAge45                           5.580e+02  2.080e+04   0.027    0.981
Base.Highway.Network               5.574e+03  2.094e+03   2.662    0.117
Maintenance1                      -3.136e+03  1.225e+04  -0.256    0.822
Functional.Class.Of.Inventory.Rte.1.11 -8.607e+04  8.491e+04  -1.014    0.417
ADT                               -1.026e+02  1.233e+02  -0.832    0.493
HistoricalSignificance.123        -9.523e+04  1.452e+05  -0.656    0.579
StructPosted.Closed.P              1.299e+06  2.419e+06   0.537    0.645
STRAHNET.Highway.Designation.123  -1.832e+03  9.899e+03  -0.185    0.870
Highway.System.Of.Inventory.Route.1  6.387e+04  4.044e+04   1.579    0.255

Residual standard error: 4151000 on 2 degrees of freedom
  (3 observations deleted due to missingness)
Multiple R-squared:  0.9901,    Adjusted R-squared:  0.9356
F-statistic: 18.16 on 11 and 2 DF,  p-value: 0.05332
```

Figure 3.2: Linear Regression Model with Total Funds as input

2. A second linear regression model (3.3) with Federal Funds as the input or scalar dependent variable is created to model the relationship between Federal funds and the other variables(excluding total funds and state funds) used in regression analysis. The results of this

regression model are explained in Section 3.6.

```
> linearCAfederal<-lm(Federal~.-TotalFunds-State,data=regressionCA,na.action=na.exclude)
> summary(linearCAfederal)

Call:
lm(formula = Federal ~ . - TotalFunds - State, data = regressionCA,
    na.action = na.exclude)

Residuals:
      1       2       3       4       5       6       7       8       9      10      11      12
  15882  -17591    2689    2697  -13545   60939  -64775  271535 -270916   -6843  -57951   89575
     13      14
 -19857    8159

Coefficients:
                                   Estimate Std. Error t value Pr(>|t|)
(Intercept)                       1.625e+07  5.830e+06   2.787    0.108
SuffRating.80                    -5.238e+05  7.899e+05  -0.663    0.575
SuffRating.50                     4.263e+05  1.014e+06   0.420    0.715
AvgAge45                         -4.340e+02  1.451e+03  -0.299    0.793
Base.Highway.Network             -1.044e+02  1.461e+02  -0.715    0.549
Maintenance1                     -1.168e+03  8.543e+02  -1.367    0.305
Functional.Class.Of.Inventory.Rte.1.11  9.387e+02  5.922e+03   0.158    0.889
ADT                               6.490e+00  8.597e+00   0.755    0.529
HistoricalSignificance.123        6.103e+03  1.013e+04   0.602    0.608
StructPosted.Closed.P             3.162e+05  1.688e+05   1.874    0.202
STRAHNET.Highway.Designation.123  4.117e+01  6.905e+02   0.060    0.958
Highway.System.Of.Inventory.Route.1  -1.078e+03  2.821e+03  -0.382    0.739

Residual standard error: 289600 on 2 degrees of freedom
  (3 observations deleted due to missingness)
Multiple R-squared:  0.9805,    Adjusted R-squared:  0.8736
F-statistic: 9.166 on 11 and 2 DF,  p-value: 0.1024
```

Figure 3.3: Linear Regression Model with Federal Funds as input

3. Other linear regression model (3.4) with State Funds as the input or scalar dependent variable is created to model the relationship between State funds and the other variables(excluding total funds and federal funds) used in regression analysis. The results of the following regression model are explained in Section 3.6.

3.5 Explanation of Variables used in Regression

Sufficiency.80 represents the proportion of bridges with sufficiency rating less than 80. Sufficiency.50 represents the proportion of bridges with sufficiency rating less than 50. AvgAge45 represents the proportion of bridges with age less than or equal to 45. Base.Highway.Network represents the number of bridges with inventory route on base highway network. Maintenance1 represents the bridges whose responsibility is with State Highway Agency. Functional.Class.of,Inventory.Rte.1.11

```
> linearCAstate<-lm(State~.-TotalFunds-Federal,data=regressionCA,na.action=na.exclude)
> summary(linearCAstate)

Call:
lm(formula = State ~ . - TotalFunds - Federal, data = regressionCA,
    na.action = na.exclude)

Residuals:
       1        2        3        4        5        6        7        8        9       10
  -21343    23640    -3613    -3624    18202   -81892    87047  3879452 -3880285     9195
      11       12       13       14
   77876  -120374    26684   -10965

Coefficients:
                                     Estimate Std. Error t value Pr(>|t|)
(Intercept)                        -1.196e+08  7.816e+07  -1.530    0.266
SuffRating.80                       2.095e+06  1.059e+07   0.198    0.861
SuffRating.50                       1.487e+06  1.360e+07   0.109    0.923
AvgAge45                            9.920e+02  1.945e+04   0.051    0.964
Base.Highway.Network                5.678e+03  1.958e+03   2.900    0.101
Maintenance1                       -1.968e+03  1.145e+04  -0.172    0.879
Functional.Class.Of.Inventory.Rte.1.11 -8.701e+04  7.940e+04  -1.096    0.388
ADT                                -1.091e+02  1.153e+02  -0.946    0.444
HistoricalSignificance.123         -1.013e+05  1.358e+05  -0.746    0.533
StructPosted.Closed.P               9.824e+05  2.263e+06   0.434    0.706
STRAHNET.Highway.Designation.123   -1.873e+03  9.258e+03  -0.202    0.858
Highway.System.Of.Inventory.Route.1  6.495e+04  3.782e+04   1.717    0.228

Residual standard error: 3882000 on 2 degrees of freedom
  (3 observations deleted due to missingness)
Multiple R-squared:  0.991,     Adjusted R-squared:  0.9413
F-statistic: 19.94 on 11 and 2 DF,  p-value: 0.0487
```

Figure 3.4: Linear Regression Model with State Funds as input

represents the functional class of inventory with rural and urban. ADT is the average daily traffic. HIstoricalSignificance123 represents the proportion of bridges on National Register of Historic places, bridges eligible for the National Register of Historic Places and bridges possibly eligible for National Register of Historic places. Strucuture.Posted.Closed represents the bridges which are posted for load, which includes temporary bridges that are load posted.

3.6 Regression Results and Fixes

From the regression summary results, it is observed that inspite of having a good R^2 value (98% - 99%), few observations are deleted due to missingness. Due to the presence of NA values, the linear regression function (lm) deletes the rows with NA values and hence resulting in loss of data. Leaving out available data points deprives the data of some amount of information. One of the fixes to prevent loss of data is mean-substitution. Applying mean substitution leaves the mean unchanged (which is desirable) but decreases variance, which may be undesirable. To overcome these, "mice" library is used to impute the data. The MICE (Multivariate Imputation via Chained Equations) package in R, helps in imputing missing values with plausible data values. These plausible values are drawn from a distribution specifically designed for each missing data point.

After fixing NAs using mice library, three regression models (refer Figures 3.5,3.7,3.9) are created with the imputed data and residual plots (refer Figure 3.6,3.8,3.10) are also generated for better understanding of the regression models. Residuals are leftover of the outcome variable after fitting a model (predictors) to data and they reveal unexplained patterns in the data by the fitted model. Using this information, not only could we check if linear regression assumptions are met, but we could improve your model in an exploratory way.

First Regression Model - Total Funds (Without NAs) : (refer Figure 3.5)

The first regression model (refer Figure 3.5) is created with Total Funds as predictor variable to establish the relationship between the predictor and response variables. To improve the regression model, residual plots are made and interpreted. From the Figure (refer Figure 3.6),in the Residuals vs fitted Values plot, it is seen that the residuals are equally spread around the red line.This shows that the residuals follow a linear pattern. The Q-Q plot shows that the residuals are not deviating from the straight line. This shows that the residuals are normally distributed expect the points 11,14,15. In the Scale-Location plot, it is seen that residuals spread wider and wider along the range of the fitted values. Hence the red smooth line is not horizontal and shows a steep angle. In the plot Residuals vs Leverage, the dotted line shows the cook's distance. We see that the points 11, 15 are outside the cook's distance. From the interpretation of residual plots, we consider the points 11,14,15 as outliers. To improve the linear regression model, we removed outliers and fitted a new regression model.

Second Regression Model - Federal Funds (Without NAs) : (refer Figure 3.7)

The second regression model (refer Figure 3.7) is fitted with Federal Funds as continuous predictor variable and residual plots are made to analyze and make the regression model better. From the Figure (refer Figure 3.8),in the Residuals vs fitted Values plot, it is seen that the residuals are not equally distributed around the red line.This shows that the residuals follow a non-linear pattern. In Q-Q plot it is seen that most of the residuals are deviating from the straight line. This shows that the residuals are not normally distributed. In the Scale-Location plot, it is seen that residuals are concentrated in the left corner. But the residuals are expected spread along the range of the fitted values which is not seen. Considering the above results, the linear regression model with federal funds as input is not used in further analysis even though the R^2 value is 90%.

```
> tempData2lm<-lm(TotalFunds~.-Federal-State,data=tempData2)
> summary(tempData2lm)

Call:
lm(formula = TotalFunds ~ . - Federal - State, data = tempData2)

Residuals:
       1        2        3        4        5        6        7        8        9       10 
  513433  -442971  -401070  -422576 -1969826  2011681  4451988  4846958 -3455230  -725438 
      11       12       13       14       15       16       17 
-6524005 -3017821  1636007  4614306  -619454   705397 -1201380 

Coefficients:
                                     Estimate Std. Error t value Pr(>|t|)   
(Intercept)                        -1.516e+08  6.178e+07  -2.454  0.05765 . 
SuffRating.80                      -2.130e+07  8.111e+06  -2.625  0.04679 * 
SuffRating.50                       3.026e+07  9.784e+06   3.092  0.02709 * 
AvgAge45                           -4.223e+04  6.660e+03  -6.340  0.00144 **
Base.Highway.Network                1.317e+03  1.511e+03   0.872  0.42330   
Maintenance1                        1.973e+04  6.394e+03   3.086  0.02728 * 
Functional.Class.Of.Inventory.Rte.1.11  8.749e+04  1.417e+04   6.176  0.00162 **
ADT                                -2.822e+02  9.715e+01  -2.905  0.03359 * 
HistoricalSignificance.123          8.180e+04  1.401e+05   0.584  0.58462   
StructPosted.Closed.P              -1.449e+05  1.571e+06  -0.092  0.93011   
STRAHNET.Highway.Designation.123    1.559e+04  7.732e+03   2.017  0.09979 . 
Highway.System.Of.Inventory.Route.1 -1.357e+04  4.349e+03  -3.121  0.02624 * 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5338000 on 5 degrees of freedom
Multiple R-squared:  0.9718,    Adjusted R-squared:  0.9097 
F-statistic: 15.64 on 11 and 5 DF,  p-value: 0.003542
```

Figure 3.5: Linear Regression Model with Total Funds as Input(Without NAs)

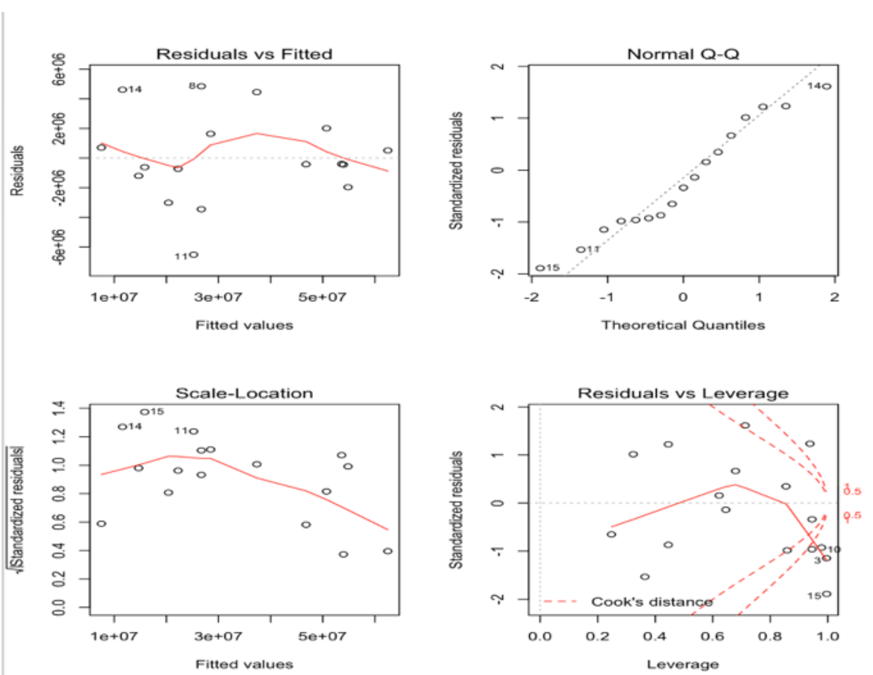


Figure 3.6: Residual Plots (Total Funds -Without NAs)

```
> tempData2lmFederal<-lm(Federal~.-TotalFunds-State,data=tempData2)
> summary(tempData2lmFederal)

Call:
lm(formula = Federal ~ . - TotalFunds - State, data = tempData2)

Residuals:
      1       2       3       4       5       6       7       8       9      10      11      12 
 391348 -354262   44469   63484 -186890 -204936  114114  176325 -366126  -95168  185156  395540 
     13      14      15      16      17 
  55995  121143    5320 -297986  -47527 

Coefficients:
                                    Estimate Std. Error t value Pr(>|t|)
(Intercept)                        5.443e+06  4.767e+06   1.142    0.305
SuffRating.80                     -7.118e+05  6.260e+05  -1.137    0.307
SuffRating.50                      8.535e+05  7.551e+05   1.130    0.310
AvgAge45                          -7.544e+02  5.139e+02  -1.468    0.202
Base.Highway.Network              -2.297e+01  1.166e+02  -0.197    0.852
Maintenance1                      -1.927e+02  4.934e+02  -0.391    0.712
Functional.Class.Of.Inventory.Rte.1.11  7.911e+02  1.093e+03   0.724    0.502
ADT                                1.499e+00  7.497e+00   0.200    0.849
HistoricalSignificance.123        -4.164e+03  1.081e+04  -0.385    0.716
StructPosted.Closed.P             -8.237e+03  1.212e+05  -0.068    0.948
STRAHNET.Highway.Designation.123   1.199e+02  5.967e+02   0.201    0.849
Highway.System.Of.Inventory.Route.1 -7.961e+01  3.356e+02  -0.237    0.822

Residual standard error: 411900 on 5 degrees of freedom
Multiple R-squared:  0.9043,	Adjusted R-squared:  0.6938 
F-statistic: 4.295 on 11 and 5 DF,  p-value: 0.05991
```

Figure 3.7: Linear Regression Model with Federal Funds as Input (without NAs)

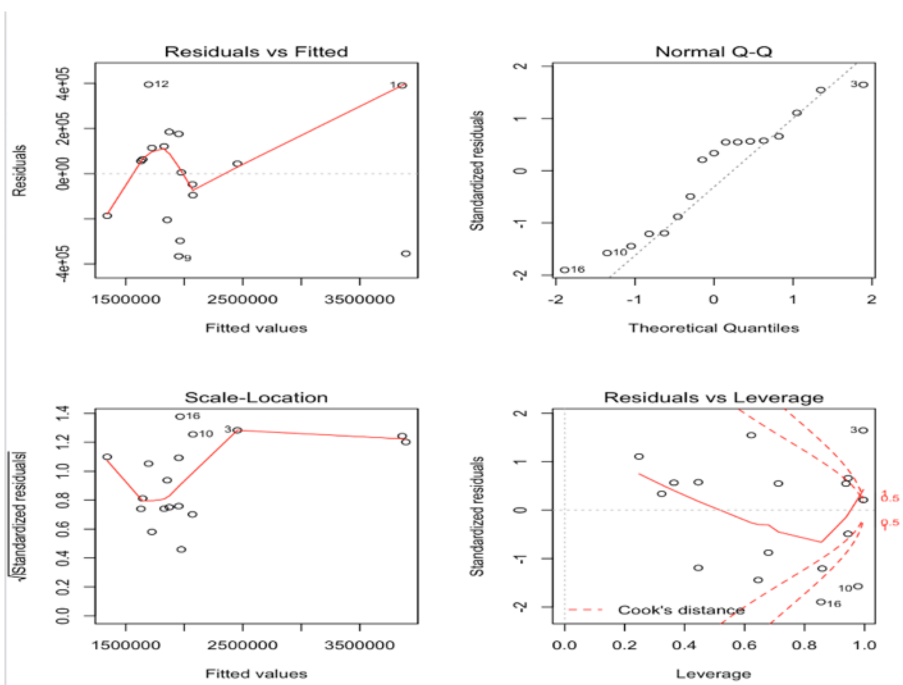


Figure 3.8: Residual Plots (Federal Funds -without NAs)

```
> tempData2lmState<-lm(State~.-Federal-TotalFunds,data=tempData2)
> summary(tempData2lmState)

Call:
lm(formula = State ~ . - Federal - TotalFunds, data = tempData2)

Residuals:
      1        2        3        4        5        6        7        8        9       10 
 122086   -88710  -445539  -486060 -1782937  2216617  4337874  4670633 -3089104  -630270 
     11       12       13       14       15       16       17 
-6709161 -3413360  1580011  4493163  -624775  1003383 -1153853 

Coefficients:
                                  Estimate Std. Error t value Pr(>|t|)   
(Intercept)                     -1.571e+08  6.150e+07  -2.554  0.05104 . 
SuffRating.80                   -2.058e+07  8.075e+06  -2.549  0.05133 . 
SuffRating.50                    2.940e+07  9.741e+06   3.019  0.02946 * 
AvgAge45                        -4.147e+04  6.630e+03  -6.255  0.00153 **
Base.Highway.Network             1.340e+03  1.504e+03   0.891  0.41386   
Maintenance1                     1.992e+04  6.365e+03   3.130  0.02595 * 
Functional.Class.Of.Inventory.Rte.1.11  8.669e+04  1.410e+04   6.148  0.00166 **
ADT                             -2.837e+02  9.672e+01  -2.934  0.03250 * 
HistoricalSignificance.123       8.596e+04  1.395e+05   0.616  0.56461   
StructPosted.Closed.P           -1.366e+05  1.564e+06  -0.087  0.93377   
STRAHNET.Highway.Designation.123  1.547e+04  7.697e+03   2.010  0.10063   
Highway.System.Of.Inventory.Route.1  -1.349e+04  4.329e+03  -3.116  0.02637 * 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5314000 on 5 degrees of freedom
Multiple R-squared:  0.971,	Adjusted R-squared:  0.9072 
F-statistic: 15.22 on 11 and 5 DF,  p-value: 0.003775
```

Figure 3.9: Linear Regression Model with State Funds as Input (without NAs)

Third Regression Model - State Funds (Without NAs) : (refer Figure 3.9)

The third regression model (refer Figure 3.9) is created with State Funds as predictor variable. Residual plots are also made to validate the linear regression model. From the Figure (refer Figure 3.10),in the Residuals vs fitted Values plot, it is seen that the residuals are equally spread around the red line.This shows that the residuals follow a linear pattern. The Q-Q plot shows that the residuals are concentrated around the straight line. This shows that the residuals are normally distributed expect the points 11,14,15. In the Scale-Location plot, it is seen that residuals spread much wider along the range of the fitted values. Hence the red smooth line is not horizontal and shows a steep angle. In the plot Residuals vs Leverage, the dotted line shows the cook's distance. We see that the points 11, 15 are outside the cook's distance. From the interpretation of residual plots, we consider the points 11,14,15 as outliers.

All the above results show that the residual plots of State Funds regression model are similar to that of Total Funds. Hence a linear regression model is fitted by removing the outliers, with Total Funds as continuous predictor variable (refer Figure 3.11). Also residual plots are made to understand the regression model (refer Figure 3.12). From Residual vs Fitted plot, we can say that the residuals are following a linear trend as all the points are spread around the mean (red line). In

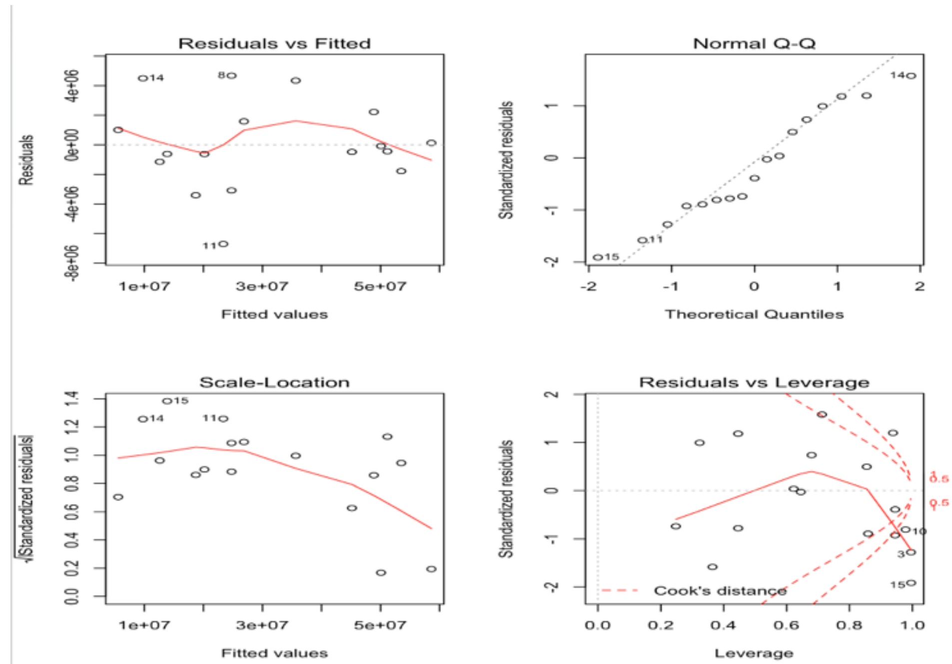


Figure 3.10: Residual Plots (State Funds - without NAs)

Scale-Location plot, all the residual points are exactly distributed along the range of fitted points. Hence the red line is horizontal without an inclination. Also, in Residuals vs Leverage plot, it is seen that there are no extra points outside the cook's distance. Wrapping up all these results, we can say that the regression model thus created is good.

3.7 Updating the Definition of Variables used in Regression

During the modeling of regression models, we considered two variables Maintenance Responsibility and Functional Class of Inventory route. In the previous regression model (refer Figure 3.11), we saw that Maintenance responsibility has the highest positive coefficient and Functional Class of Inventory route has a negative coefficient. In previous sections (Refer Sections 3.2.4 and 3.2.5), we understood that there are different sub levels for Maintenance Responsibility and Functional Class of Inventory route. The coefficients we obtained in the previous regression model are due to the combined effect of all the levels in the variables. So we clustered few levels based on the highway importance, rural and urban functional classes.

The variables added to the regression analysis are

```
> tempData2lmnoOutilers<-lm(TotalFunds~.-Federal-State,data=tempData2noOutilers)
> summary(tempData2lmnoOutilers)

Call:
lm(formula = TotalFunds ~ . - Federal - State, data = tempData2noOutilers)

Residuals:
      1        2        3        4        5        6        7        8        9       10 
-304473   239361   -88417  -270841   446422   218209  -308135  4441775 -3860413   167192 
     11       12       13       14 
-1156916    41186   522216   -87164 

Coefficients:
                                         Estimate Std. Error t value Pr(>|t|)
(Intercept)                            -1.133e+08  7.896e+07  -1.435    0.288
SuffRating.80                          -5.716e+06  2.544e+07  -0.225    0.843
SuffRating.50                           8.988e+06  3.276e+07   0.274    0.810
AvgAge45                               -7.080e+03  4.378e+04  -0.162    0.886
Base.Highway.Network                    4.777e+03  3.766e+03   1.268    0.332
Maintenance1                            6.371e+01  2.143e+04   0.003    0.998
Functional.Class.Of.Inventory.Rte.1.11 -5.699e+04  1.768e+05  -0.322    0.778
ADT                                    -1.218e+02  1.901e+02  -0.641    0.587
HistoricalSignificance.123             -7.491e+04  2.433e+05  -0.308    0.787
StructPosted.Closed.P                   1.312e+06  2.154e+06   0.609    0.604
STRAHNET.Highway.Designation.123        6.297e+01  2.014e+04   0.003    0.998
Highway.System.Of.Inventory.Route.1     5.334e+04  8.002e+04   0.667    0.574

Residual standard error: 4293000 on 2 degrees of freedom
Multiple R-squared:  0.991,     Adjusted R-squared:  0.9414 
F-statistic: 19.98 on 11 and 2 DF,  p-value: 0.04861
```

Figure 3.11: Linear regression model with Total Funds as Input (no Outliers)

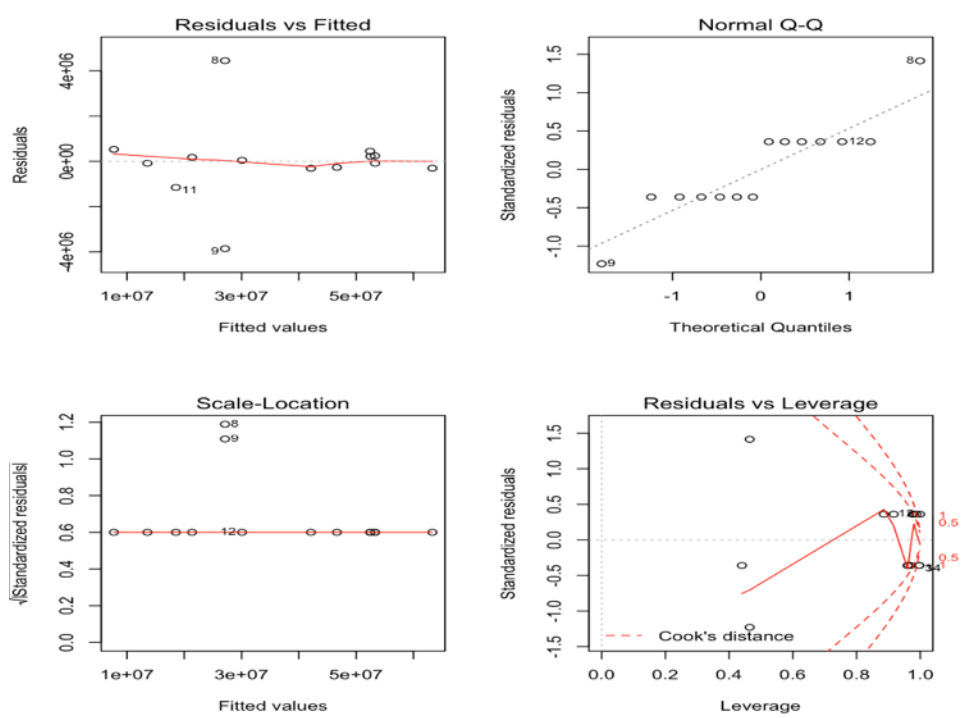


Figure 3.12: Residual Plots (Total Funds - no Outliers)

```
> tempData2lmState<-lm(St_inf_1000~.-Year-Fed_inf_1000-TF_inf_1000-Functional.Class.Of.Inventory
.Rte.1.11,data=tempData2infl)
> summary(tempData2lmState)

Call:
lm(formula = St_inf_1000 ~ . - Year - Fed_inf_1000 - TF_inf_1000 - 
    Functional.Class.Of.Inventory.Rte.1.11, data = tempData2infl)

Residuals:
       1        2        3        4        5        6        7        8        9       10 
-1.98721  0.74728 -0.46522  4.36489  0.38519 -3.11685 -2.57785 -0.27385  0.31178  2.75593 
      11       12       13       14       15       16       17 
-0.07701  1.88727 -0.13806 -1.53205 -0.24235 -3.23727  3.19538 

Coefficients:
                                      Estimate Std. Error t value Pr(>|t|)  
(Intercept)                         -2.796e+01  1.345e+02  -0.208   0.8487  
SufficiencyRating.80                -2.185e-01  7.353e+00  -0.030   0.9782  
SufficiencyRating.50                 4.682e+00  8.297e+00   0.564   0.6120  
AvgAge45                            -3.761e-03  3.951e-03  -0.952   0.4114  
Base.Highway.Network                 5.532e-03  1.681e-03   3.291   0.0460 *
MaintenanceResposibility1            1.077e-02  5.736e-03   1.877   0.1572  
MaintenanceResposibility234         -4.755e-02  2.637e-01  -0.180   0.8684  
FunctionalClass_rural               -7.059e-02  1.593e-02  -4.430   0.0214 *
FunctionalClass_urban                1.677e-02  2.931e-03   5.723   0.0106 *
ADT                                 -1.300e-04  9.415e-05  -1.381   0.2611  
HistoricalSignificance.123          -2.920e-02  1.461e-01  -0.200   0.8544  
StructPosted.Closed.P               -4.722e-01  1.488e+00  -0.317   0.7718  
STRAHNET.Highway.Designation.123     8.179e-03  8.146e-03   1.004   0.3894  
Highway.System.Of.Inventory.Route.1 -6.287e-03  3.458e-03  -1.818   0.1667  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.991 on 3 degrees of freedom
Multiple R-squared:  0.9817,	Adjusted R-squared:  0.9024 
F-statistic: 12.38 on 13 and 3 DF,  p-value: 0.03068
```

Figure 3.13: Updated Linear Regression Model with State Funds as Input

- MaintenanaceResponsiblity1 - Represents the bridges whose Maintenance is with State Highway Agency.
- MaintenanceResponsibility234 - Represents the bridges whose maintenance is either with County Highway Agency or Town/Township Highway Agency or City/Municipal Highway Agency.
- FunctionalClass_rural - Represents the bridges with rural functional class of inventory.
- FunctionalClass_urban - Represents the bridges with urban functional class of inventory.

With these changes, a linear regression model with State funds as predictor variable is fitted (3.13) to observe the effect of State funds on updated bridge performance factors. Here the Federal funds are not considered as they are approximately constant for all the states throughout all the years. So the federal funds have nominal effect on the bridge performance indicators. Also the total funds and state funds follow similar trend. Hence we created a regression model with State funds as input. To evaluate the goodness of the model, residual plots are plotted (refer Figure3.14).

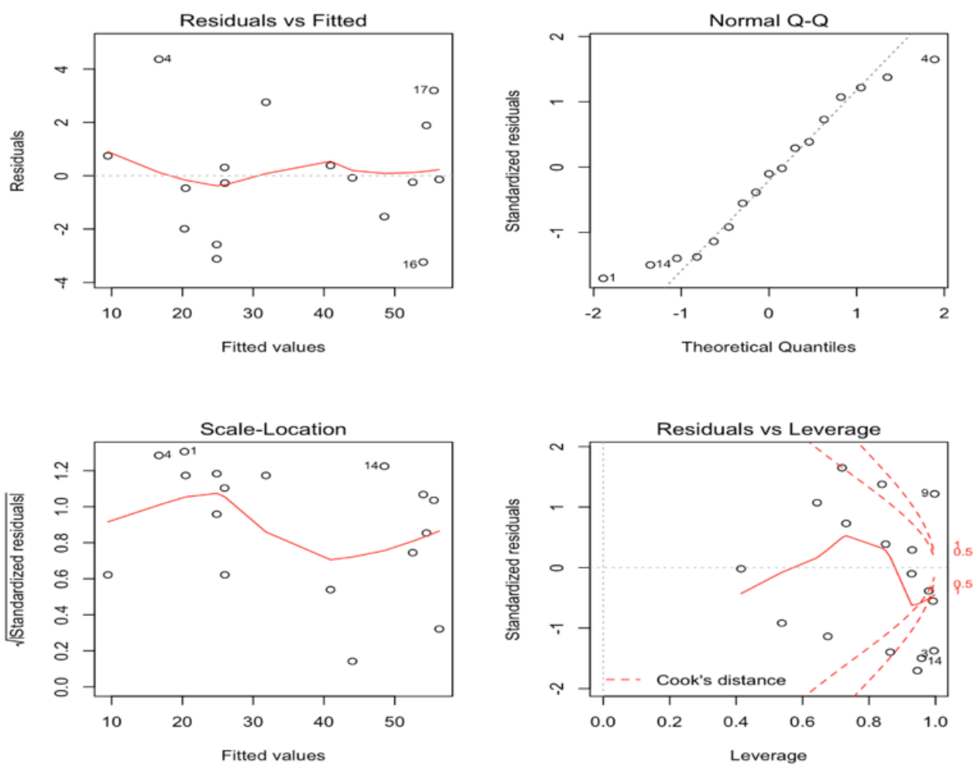


Figure 3.14: Residual Plots Updated (State Funds)

```
> CA_FuncObsol_lm<-lm(Funds~CA_funcObs,data=CA_FuncObsol)
> summary(CA_FuncObsol_lm)

Call:
lm(formula = Funds ~ CA_funcObs, data = CA_FuncObsol)

Residuals:
    Min      1Q  Median      3Q     Max 
-24.954 -14.598   1.252  12.959  23.984 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)  
(Intercept)  65.8520    32.6453   2.017   0.0619 .
CA_funcObs   -0.4649     0.4996  -0.931   0.3668  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 16.05 on 15 degrees of freedom
Multiple R-squared:  0.05458,   Adjusted R-squared:  -0.008448 
F-statistic: 0.866 on 1 and 15 DF,  p-value: 0.3668
```

Figure 3.15: Effect of Funds on Functional Obsolesce

3.8 Regression Results with updated Variable Definition

From the regression model (3.13), it is observed that with one million increase in state funds, the bridges with sufficiency rating less than 80 decreased which is evident from the negative coefficient. The coefficient of average age is observed to be negative. This implies that with the increase in state funds, the bridges with average age less than 45 decreased. This in turn implies that the average age of the bridges increased with one million increase in funds. The variables Base Highway Network and functional class of Inventory route are the significant factors. One of the observations is that, the ADT decreased with the increase in funds. From the residual plots (refer Figure3.14), it is seen that the residuals are normally distributed in Q-Q plot. In residual vs fitted plot, the residuals are distributed around mean. Inspite of the presence of some residual points outside the cook's distance (dotted red line), the regression model showed a good R^2 value of 90%.

In addition to the above regression models, a regression model is fitted to find the effect of funds allotted on Functional Obsolesce (number of functionally obsolete structures)in California [refer Figure 3.15]. From the model, it is seen that, for every one million increase in State funds, the percentage of functionally obsolete structures in California decreased by a factor 0.4649. However, it would be more interesting to see how the funds are being allocated based on the Functional Obsolete structures. Hence a linear regression model is created with Functional Obsolete structures as predictor variable with state funds as response variable[refer Figure 3.16]. Surprisingly, It is seen that for every one percent increase in Functional Obsolete structures in California, the state funds

```
> CA_FuncObsol_lm1<-lm(CA_funcObs~Funds,data=CA_FuncObsol)
> summary(CA_FuncObsol_lm1)

Call:
lm(formula = CA_funcObs ~ Funds, data = CA_FuncObsol)

Residuals:
    Min      1Q  Median      3Q     Max 
-13.109  -5.554  -1.603   4.985  14.948 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  69.0726     4.9093  14.070 4.78e-10 ***
Funds        -0.1174     0.1262  -0.931    0.367    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.064 on 15 degrees of freedom
Multiple R-squared:  0.05458,	Adjusted R-squared:  -0.008448 
F-statistic: 0.866 on 1 and 15 DF,  p-value: 0.3668
```

Figure 3.16: Allocation of Funds based on Functional Obsolesce

allotted decreased by a factor 0.1174. This shows that the state funds are not being allocated based on functional obsolesce or structural deficiency. There is a need to pursue a way to distribute funds proportionate with the number of structurally deficient structures and functionally obsolete structures. This can be done by creating an optimization model which is discussed in next chapter.

Section 4

Funds Allocation - Optimization Model

In this section, a brief description of the problem is provided and an abstract Resource Allocation Mathematical model is proposed for optimized distribution of funds among various states in Unites States. The parameters used in the model are based on the regression results from the previous section.

4.1 Problem Statement and Assumptions

Here, a list all the states are considered among which the Total Funds allocated by U.S. government are to be distributed, where Total Funds is the sum of Federal funds and State Funds.

$$TotalFunds = FederalFunds + StateFunds$$

From the regression analysis, it is found that the federal funds are allotted for each state in the same trend every year. Hence we made the following assumptions.

Assumptions

- The minimum funds allotted in each state should be equal to Federal funds.
- Also, maximum funds allotted per state cannot exceed the total funds.
- Average age of bridges is 45 years.
- Each and every state should be allotted State Funds irrespective of any circumstances.

Notation	Description
S	set of all States in U.S
B	set of all bridges present in different states
r_{ij}	Sufficiency rating of Bridge i in State j
d_{ij}	Binary variable for Structurally deficient Structures i in State j
f_{ij}	Binary variable for Functionally Obsolete Structures i in State j
${n^d}_j$	Number of Structurally deficient structures in state j
${n^f}_j$	Number of Functionally Obsolete structures in state j
a_{ij}	Age of Bridge i in State j
${a^d}_j$	Average age of Structurally Deficient Bridges in State j
${a^f}_j$	Average age of Functionally Obsolete Bridges in State j
${T^{fed}}_j$	Federal Funds allotted per state j
${T^{st}}_j$	State Funds allotted per state j
T_j	Total funds allotted per state j
T	Total Funds allotted for all the states
x_j	Funds to be allotted in each state j

Table 4.1: Table of Notation used in Optimization Model

- There are no natural calamities or disasters occurring in that particular year.

The objective of this Mathematical model is to maximize the funds allocated in the states with more number of Structurally Deficient and functionally Obsolete structures subject to constraints. Firstly, basic notations used in the mathematical model are defined and then an abstract model is developed.

4.2 Parameters used in Modeling

The Table 4.1 shows the variables used throughout the optimization model. The set of all States in U.S. is denoted by S and the set of all Bridges present in different states is denoted by B. Given the sets S and B, the notations used in the formulation can be defined as follows.

- r_{ij} : Sufficiency rating of Bridge i in State j, where $i \in B$ and $j \in S$

- d_{ij} : Binary variable for Structurally deficient Structures i in State j

$$\begin{aligned} d_{ij} &= 1, when \quad r_{ij} < 80 \\ d_{ij} &= 0, otherwise \end{aligned}$$

- f_{ij} : Binary variable for Functionally Obsolete Structures i in State j

$$\begin{aligned} f_{ij} &= 1, when \quad r_{ij} < 50 \\ f_{ij} &= 0, otherwise \end{aligned}$$

- n_j : Total number of Bridges in state j, where $j \in S$.

- ${n^d}_j$: Number of Structurally deficient structures in state j, where $j \in S$, and

$$ {n^d}_j = \sum_{i=1}^{B} d_{ij} \quad , j \in S $$

- ${n^f}_j$: Number of Functionally Obsolete structures in state j, where $j \in S$, and

$$ {n^f}_j = \sum_{i=1}^{B} f_{ij} \quad , j \in S $$

- a_{ij} : Age of Bridge i in State j, where $i \in B$ and $j \in S$

- ${a^d}_j$: Average age of Structurally Deficient Bridges in State j, where $j \in S$ and

$$ {a^d}_j = \frac{\sum_{i=1}^{B} d_{ij} a_{ij}}{\sum_{i=1}^{B} d_{ij}} $$

- ${a^f}_j$: Average age of Functionally Obsolete Bridges in State j, where $j \in S$ and

$$ {a^f}_j = \frac{\sum_{i=1}^{B} f_{ij} a_{ij}}{\sum_{i=1}^{B} f_{ij}} $$

- ${T^{fed}}_j$: Federal Funds allotted per state j where $j \in S$

- ${T^{st}}_j$: State Funds allotted per state j where $j \in S$
- T_j :Total funds allotted per state j where $j \in S$, where

$$T_j = T^{fed} + T^{st}$$

- T : Total Funds allotted for all the states, where

$$T = \sum_{j}^{S} T_j, where \quad j \in S$$

4.3 Objective and Constraints

We are concerned with the problem of optimally allocating the funds to bridges based on the number of structurally deficient and functionally obsolete structures in respective states. We need to find the proportion of funds to be allotted in each state with minimum funds as federal funds and the minimum age of bridges as 45 years. Let's introduce some mathematical notation.

Let $S = \{$ AK, AL, AR, AZ, CA, ... WV, WY $\}$ be the set of states in United States and $B = \{1, 2, 3, 4, \ldots \text{n}\}$ be the bridges in all the states. For convenience, we considered the set of states as $S = \{1, 2, 3, 4, \ldots 52\}$. The structurally deficient structures in each state are extracted by taking sufficiency rating, r_{ij} as reference. This is done by considering a binary variable d_{ij}. Similarly one more binary variable, f_{ij} is considered to identify the functionally obsolete structures by taking sufficiency rating, r_{ij} as reference.

We extracted all the structurally deficient and functionally obsolete structures in each state as follows. Let us consider State j. Now for structurally deficient structures, the binary variables are $d_{1j}, d_{2j}, \ldots d_{nj}$.

d_{1j} represents the binary variable for bridge 1 in state j

d_{2j} represents the binary variable for bridge 2 in state j
$\vdots$

d_{nj} represents the binary variable for bridge n in state j

If the sufficiency rating r_{1j} is less than 80, then $d_{1j} = 1$, else $d_{1j} = 0$
If the sufficiency rating r_{2j} is less than 80, then $d_{2j} = 1$, else $d_{2j} = 0$
If the sufficiency rating r_{nj} is less than 80, then $d_{nj} = 1$, else $d_{nj} = 0$

The total number of structurally deficient structures in each state j, ${n^d}_j$ are calculated with the formula below.

$$
\begin{gathered}
{n^d}_j = d_{1j} + d_{2j} + d_{3j} + \ldots + d_{nj} \\
\text{For state 1, } {n^d}_1 = d_{11} + d_{21} + d_{31} + \ldots + d_{n1} \\
\text{For state 2, } {n^d}_2 = d_{12} + d_{22} + d_{32} + \ldots + d_{n2} \\
. \\
. \\
\text{For state n, } {n^d}_j n = d_{13} + d_{23} + d_{33} + \ldots + d_{n_j n}
\end{gathered}
$$

Now for functionally obsolete structures, the binary variables are $f_{1j}, f_{2j}, \ldots f_{nj}$.

f_{1j} represents the binary variable for bridge 1 in state j

f_{2j} represents the binary variable for bridge 2 in state j

$\vdots$

f_{nj} represents the binary variable for bridge n in state j

If the sufficiency rating r_{1j} is less than 50, then $f_{1j} = 1$, else $f_{1j} = 0$

If the sufficiency rating r_{2j} is less than 50, then $f_{2j} = 1$, else $f_{2j} = 0$

$\vdots$

If the sufficiency rating r_{nj} is less than 50, then $f_{nj} = 1$, else $f_{nj} = 0$

The total number of functionally obsolete structures in each state j, ${n^f}_j$ are calculated with the formula below.

$$
\begin{gathered}
{n^f}_j = f_{1j} + f_{2j} + f_{3j} + \ldots + f_{nj} \\
\text{For state 1, } {n^f}_1 = f_{11} + f_{21} + f_{31} + \ldots + f_{n1} \\
\text{For state 2, } {n^f}_2 = f_{12} + f_{22} + f_{32} + \ldots + f_{n2} \\
. \\
. \\
\text{For state n, } {n^f}_n = f_{1n} + f_{2n} + f_{3n} + \ldots + f_{n_j n}
\end{gathered}
$$

In this way, the total number of structurally deficient structures, ${n^d}_j$ and functionally obsolete structures, ${n^f}_j$ in state j are calculated. Now, the average age of structurally deficient and functionally obsolete structures are calculated by considering the binary variables d_{ij}, f_{ij} and the age of the bridge a_{ij}. The average age of structurally deficient structures is calculated using the following expression

$$a^d{}_j = \frac{d_{1j}a_{1j} + d_{2j}a_{2j} + \ldots + d_{nj}a_{nj}}{d_{1j} + d_{2j} + \ldots + d_{nj}}$$

$$\text{For state 1, } a^d{}_1 = \frac{d_{11}a_{11} + d_{21}a_{21} + \ldots + d_{n1}a_{n1}}{d_{11} + d_{21} + \ldots + d_{n1}}$$

$$\text{For state 2, } a^d{}_2 = \frac{d_{12}a_{12} + d_{22}a_{22} + \ldots + d_{n2}a_{n2}}{d_{12} + d_{22} + \ldots + d_{n2}}$$

.

.

$$\text{For state n, } a^d{}_n = \frac{d_{1n}a_{1n} + d_{2n}a_{2n} + \ldots + d_{nn}a_{nn}}{d_{1n} + d_{2n} + \ldots + d_{nn}}$$

The average age of functionally obsolete structures is calculated using the following expression

$$a^f{}_j = \frac{f_{1j}a_{1j} + f_{2j}a_{2j} + \ldots + f_{nj}a_{nj}}{f_{1j} + f_{2j} + \ldots + f_{nj}}$$

$$\text{For state 1, } a^f{}_1 = \frac{f_{11}a_{11} + f_{21}a_{21} + \ldots + f_{n1}a_{n1}}{f_{11} + f_{21} + \ldots + f_{n1}}$$

$$\text{For state 2, } a^f{}_2 = \frac{f_{12}a_{12} + f_{22}a_{22} + \ldots + f_{n2}a_{n2}}{f_{12} + f_{22} + \ldots + f_{n2}}$$

.

.

$$\text{For state n, } a^f{}_n = \frac{f_{1n}a_{1n} + f_{2n}a_{2n} + \ldots + f_{nn}a_{nn}}{f_{1n} + f_{2n} + \ldots + f_{nn}}$$

Objective

Let x_j be the decision variable which is defined as the funds to be allotted in each state j, where $j \in S$. The objective now is to maximize the funds allotted in each state.

$$Maximize \quad (n^d{}_j + n^f{}_j)x_j$$

Constraints

$$x_j > T^{fed}{}_j \tag{4.1}$$

$$x_j < T_j \tag{4.2}$$

$$\sum_{j=1}^{S} x_j \leq T \tag{4.3}$$

$$\frac{a^d{}_j + a^f{}_j}{2} \geq 45 \tag{4.4}$$

$$x_j > 0 \tag{4.5}$$

- Constraint 1 (refer Equation 4.1) guarantees that each state is allotted a minimum funds equal to Federal funds.
- Constraint 2 (refer Equation 4.2) states that the funds allocated in each state cannot exceed the Total funds to be distributed.
- Constraint 3 (refer Equation 4.3) indicated that sum of the funds allocated in each state should not exceed total funds.
- Constraint 4 (refer Equation 4.4) ensures that the funds are allocated to bridges whose average age is atleast 45 years.
- Constraint 5 (refer Equation 4.5) is a non negativity constraint.

4.4 Heuristic Approach to the Optimization Model

This section presents a heuristic approach to solve the optimization problem of funds allocation. We created a method, which takes number of structurally deficient structures in each state, number of functionally obsolete structures, Total federal funds in each state, Total funds, average age of structurally deficient bridges in each state, average age of functionally obsolete structures in each state as input parameters. We defined the total state funds for all the states as TS and number of states as n. In the first step of the algorithm, we have calculated the total state funds by subtracting the federal funds from total funds. In step 5, the total number of structurally deficient and functionally obsolete bridges with average age less than 45. n step 8, we initialized a list that contains the optimized set of funds for all the states. In step 12, the funds to be allocated for each state are calculate by the proportion of bridges. The Table (4.2) shows the optimized funds for each state based on the structurally deficient and functionally obsolete bridges. It is seen that for the state Connecticut, the state funds increased by 80% in accordance with the total number of deficient and obsolete bridges. In the table, the values marked in bold letters show the increase in state funds after running the heuristic algorithm. The values with negative sign indicates that the funds are decreased in those states due to less number of structurally deficient and functionally obsolete structures.

Algorithm 1 - Heuristic Algorithm to optimize the funds allocated

Require: ${n^d}_j, {n^f}_j, {T^{fed}}_j, T, {a^d}_j, {a^f}_j$

$TS \leftarrow T - (n) * {T^{fed}}_j$

$Initialize\ number\ of\ bridges : totalBridges = 0$

for each state j in S **do**

  if $\frac{{a^d}_j + {a^f}_j}{2} \geq 45$ **then**

    $totalBridges += {n^d}_j + {n^f}_j$

  end if

end for

$Initialize\ a\ list\ for\ state\ funds\ in\ each\ state : list = [\,]$

for each state j in S **do**

  $T_j = {T^{fed}}_j$

  if $\frac{{a^d}_j + {a^f}_j}{2} \geq 45$ **then**

    ${T^{St}}_j += \frac{{T^{fed}}_j * ({n^d}_j + {n^f}_j)}{totalBridges}$

  end if

  $list.append({T^{St}}_j)$

end for

State	StateCode	StructurallyDefcient	FunctionallyObsolete	Total Bridges	Actual State Funds	Optimized State Funds	Percentage Change in Funds
Alabama	AL	8	64	72	2,284,776	1038890.172	-54.53
Alaska	AK	1	6	7	1,172,528	101003.2112	-91.39
Arizona	AZ	11	97	108	2,388,235	1558335.258	-34.75
Arkansas	AR	21	79	100	2,022,292	1442903.017	-28.65
California	CA	140	829	969	14,350,351	13981730.23	-2.57
Colorado	CO	17	83	100	2,570,992	1442903.017	-43.88
Connecticut	CT	40	204	244	1,942,764	3520683.361	**81.22**
Delaware	DE	2	15	17	1,290,769	245293.5129	-81.00
District of Columbia	DC	3	55	58	386,000	836883.7498	**116.81**
Florida	FL	6	161	167	8,538,417	2409648.038	-71.78
Georgia	GA	7	142	149	2,742,926	2149925.495	-21.62
Hawaii	HI	1	27	28	396,972	404012.8447	**1.77**
Idaho	ID	5	44	49	808,028	707022.4783	-12.50
Illinois	IL	56	207	263	6,331,169	3794834.934	-40.06
Indiana	IN	39	203	242	2,613,215	3491825.301	**33.62**
Iowa	IA	8	45	53	2,024,650	764738.5989	-62.23
Kansas	KS	8	165	173	1,726,954	2496222.219	**44.54**
Kentucky	KY	24	133	157	2,945,695	2265357.736	-23.10
Louisiana	LA	43	142	185	1,773,962	2669370.581	**50.48**
Maine	ME	9	60	69	705,428	995603.0816	**41.13**
Maryland	MD	15	128	143	2,785,005	2063351.314	-25.91
Massachusetts	MA	57	463	520	2,923,022	7503095.688	**156.69**
Michigan	MI	87	382	469	3,418,532	6767215.149	**97.96**
Minnesota	MN	15	70	85	2,997,473	1226467.564	-59.08
Mississippi	MS	2	120	122	1,412,711	1760341.681	**24.61**
Missouri	MO	95	231	326	3,420,873	4703863.835	**37.50**
Montana	MT	4	40	44	764,000	634877.3274	-16.90
Nebraska	NE	2	18	20	1,555,074	288580.6034	-81.44
Nevada	NV	2	65	67	778,519	966745.0213	**24.18**
New Hampshire	NH	27	39	66	722,118	952315.9911	**31.88**
New Jersey	NJ	54	399	453	7,492,167	6536350.666	-12.76
New Mexico	NM	11	40	51	1,017,555	735880.5386	-27.68
New York	NY	175	1265	1440	11,391,121	20777803.44	**82.40**
North Carolina	NC	43	208	251	4,414,304	3621686.572	-17.96
North Dakota	ND	1	5	6	880,309	86574.18101	-90.17
Ohio	OH	40	659	699	5,150,663	10085892.09	**95.82**
Oregon	OR	11	193	204	2,222,566	2943522.154	**32.44**
Pennsylvania	PA	176	663	839	8,200,466	12105956.31	**47.63**
Rhode Island	RI	41	95	136	511,265	1962348.103	**283.82**
South Carolina	SC	19	76	95	1,864,286	1370757.866	-26.47
South Dakota	SD	7	26	33	650,630	476157.9956	-26.82
Tennessee	TN	31	201	232	1,987,379	3347534.999	**68.44**
Texas	TX	18	766	784	19,681,911	11312359.65	-42.52
Utah	UT	3	50	53	1,481,738	764738.5989	-48.39
Vermont	VT	3	40	43	531,802	620448.2972	**16.67**
Virginia	VA	46	249	295	5,316,330	4256563.9	-19.93
Washington	WA	23	245	268	4,802,004	3866980.085	-19.47
West Virginia	WV	49	91	140	1,314,915	2020064.224	**53.63**
Wisconsin	WI	41	111	152	3,316,270	2193212.586	-33.87
Wyoming	WY	9	15	24	594,039	346296.724	-41.70

Table 4.2: Results of Optimization with Heuristic Approach

Section 5

Time Series Analysis

In this section, we create Time Series Models to obtain an understanding of the underlying factors that produced the observed data. Also, we attempt to fit models and proceed to forecasting.

Time series Objects for Total Funds, Federal Funds and State funds are created. (Refer Figure 5.1). A time series object is a vector (univariate) or matrix (multivariate) with additional attributes, including time indices for each observation, the sampling frequency and time increment between observations, and the cycle length for periodic data. Such objects are of the ts (time series) class, and represent data that has been observed at (approximately) equally spaced time points. The ts() function takes several arguments, the first one is the data itself. It can be a vector object or a matrix. The other arguments are start, end and frequency where the start and end arguments allow us to provide a start date and end date for the series. Finally the frequency argument lets us specify the number of observations per unit of time. In Figure 5.1, three time series objects are created. The data is Total funds (`TF_inf_1000`) for the first Time Series Object (TotFunds), Federal funds (`Fed_inf_1000`) for the second Time Series Object (FedFunds) and State Funds (`St_inf_1000`) for the third Time Series Object (StFunds). For the three time series objects, the start and end dates are given as 1994 and 2014 respectively and we considered frequency to be 1 as we are dealing with yearly data. We used the command class() to cross validate that the objects created belong to ts (time series) class.

The advantage of creating and working with time series objects of the ts (time series) class is that many methods are available for utilizing time series attributes, such as time index information and calling plot() on a ts (time series) object will automatically generate a plot over time. We

```
> TotFunds<-ts(TS_CAinflation$TF_inf_1000,frequency = 1, start = 1994, end = 2014)
> class(TotFunds)
[1] "ts"
> FedFunds<-ts(TS_CAinflation$Fed_inf_1000,frequency = 1, start = 1994, end = 2014)
> class(FedFunds)
[1] "ts"
> StFunds<-ts(TS_CAinflation$St_inf_1000,frequency = 1, start = 1994, end = 2014)
> class(StFunds)
[1] "ts"
```

Figure 5.1: Time Series Objects

```
> plot(as.xts(TotFunds), major.format = "%Y",main="Time Series of Total Funds",ylab="Total Funds
(ts)")
> plot(as.xts(FedFunds), major.format = "%Y",main="Time Series of Federal Funds",ylab="Federal F
unds(ts)")
> plot(as.xts(StFunds), major.format = "%Y",main="Time Series of State Funds",ylab="State Funds(
ts)")
> plot.ts(StFunds,col = "blue", lwd = 2,main="Time series of StateFunds and TotalFunds",ylab="St
ateFunds(Ts-Blue),TotalFunds(Ts-Red) ")
> lines(TotFunds, type='b', col = "red", lwd = 2)
```

Figure 5.2: Commands for Time Series Plots

created plots for the time series objects thus created.[refer Figure 5.3] using the commands shown in the Figure 5.2. From the Time Series plot of Total funds, it is seen that funds increased from the year 1994 to 1998 and then again dropped in the year 2003. Later on the funds allocated increased progressively with very few drops. From the time series plot of federal funds, we see that the federal funds are distributed between 2 and 3 millions of dollars except for the year 2007. There is huge drop in the federal funds in the year 2007 due to unavailability if data for the year 2007 (see Section 2.3.7). Except for 2007, the federal funds allotted are approximately constant for all the years. In the time series plot of state funds, it is seen that the state funds followed the same trend as the total funds. To justify the statement, a fourth plot is made to compare the total funds (red) and state funds (blue). We can say that the total funds and state funds followed the same trend from the plot and also because federal funds are constantly allotted.

5.1 ARIMA model

The historical data obtained from Federal Highway Administration is obtained. The data lists State Disbursements for Highways, total disbursements for. Highways and disbursements by states for State administered Highways. With the historical data obtained from Federal Highway Administration, in this research, ARIMA (Auto Regressive Integrated moving Average) models are used to

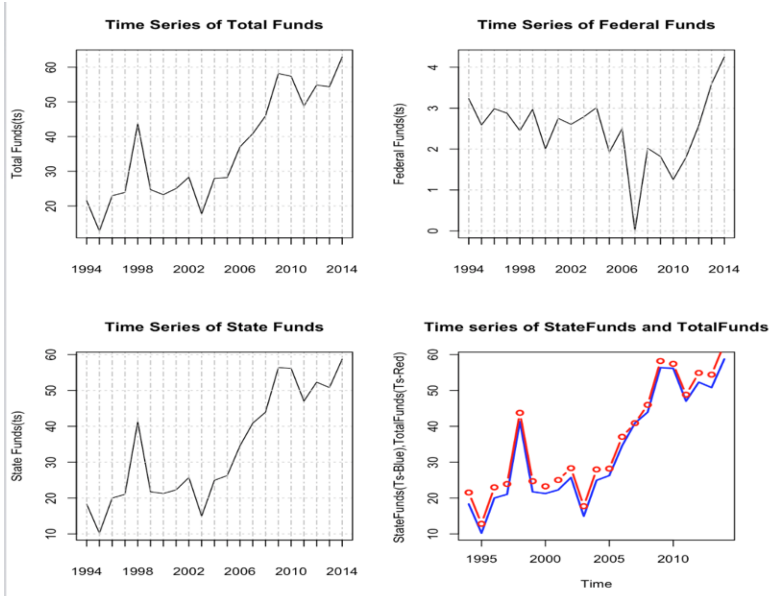


Figure 5.3: Pots for Time Series Objects

forecast and make predictions. An Auto Regressive Integrated Moving Average (ARIMA) model is fitted to a time series data to better understand the data and predict future points in the series. The AR part of ARIMA indicates that the evolving variable of interest is regressed on its own lagged or prior values. The MA part means that the regression error is actually a linear combination of error terms whose values occurred at various times in the past. The I (for "integrated") indicates that the data values have been replaced with the difference between their values and the previous values (and this differencing process may have been performed more than once). The purpose of AR, MA and I is to make the model fit the data as well as possible. The reason behind choosing ARIMA model is that it is a flexible forecasting model which utilizes historical data and also, it can be used as a foundation for more complex models.

Forecasting with ARIMA comprises the following stages - Model Identification, Parameter estimation, Diagnostic checking. These three stages are repeated until a suitable model is obtained. This process is represented by a flowchart [refer Figure 5.4]. Model Identification is done by analyzing the ACF and PACF plots. In R, the functions acf() and pacf() are used to calculate and plotting of ACF and PACF. Then the order of the model is obtained. Once the order of ARIMA model is obtained, in the next step Parameter estimation, the function arima() is used to estimate the parameters. The coefficients, residuals and the Akaike Information Criterion (AIC) obtained after fitting the model are the main factors used in estimation. Diagnostic checking is done by using the function tsdiag() which is used to test the presence of non randomness.

Detailed description of steps in ARIMA model

Autocorrelations and Choosing Model Order:

AutoCorrelation Function or ACF plots are used to choose the parameters for ARIMA models.ACF plots display correlation between a series and its lags. In addition to suggesting the order of differencing, ACF plots help in determining the order of the MA(q) model. Partial autocorrelation plots (PACF) display correlation between a variable and its lags that is not explained by previous lags. PACF plots are useful when determining the order of the AR(p) model.

ACF and PACF plots are made for Total Funds(see Figure 5.5) and State Funds (refer Figure 5.6). The blue lines shown in the plots represent the 95% significant boundaries.

Findings :

In Figures 5.5,5.6, as there are more significant auto correlations, differencing is opted for the series. In Figures 5.5, 5.6 , it is seen that there are significant auto correlations at lag 1 and 2 (for differenced series). In Figures 5.5,5.6 , it is seen that partial correlation plots show significant spike at lag 2. So the ARIMA parameters for Total funds (p, d, q) are chosen as $(2, 1, 2)$. And, the ARIMA parameters for Total funds (p, d, q) are also chosen as $(2, 1, 2)$.

ARIMA model for Total Funds:

After obtaining (p, d, q) values from ACF and PACF plots, an ARIMA model is fit for Total Funds [refer Figure 5.7]. It is seen that the data used for creating the model is differenced by using log(TotFunds) and hence the value of d is taken as 1. As the data is collected on an yearly basis, the period of ARIMA model is taken as 1. The coefficients are used in the calculation of the funds after a period t.The AIC is computed from the conditional sum-of-squared errors and not from the true maximum likelihood function. The best model is usually considered to be the one with lowest variance and lowest AIC value.

ARIMA model for State Funds:

After obtaining (p, d, q) values from ACF and PACF plots, an ARIMA model is fit for State Funds [refer Figure 5.8]. We have used log(StFunds) to make the series stationary by differencing and hence the value of d is taken as 1. As the data is collected on an yearly basis, the period of ARIMA

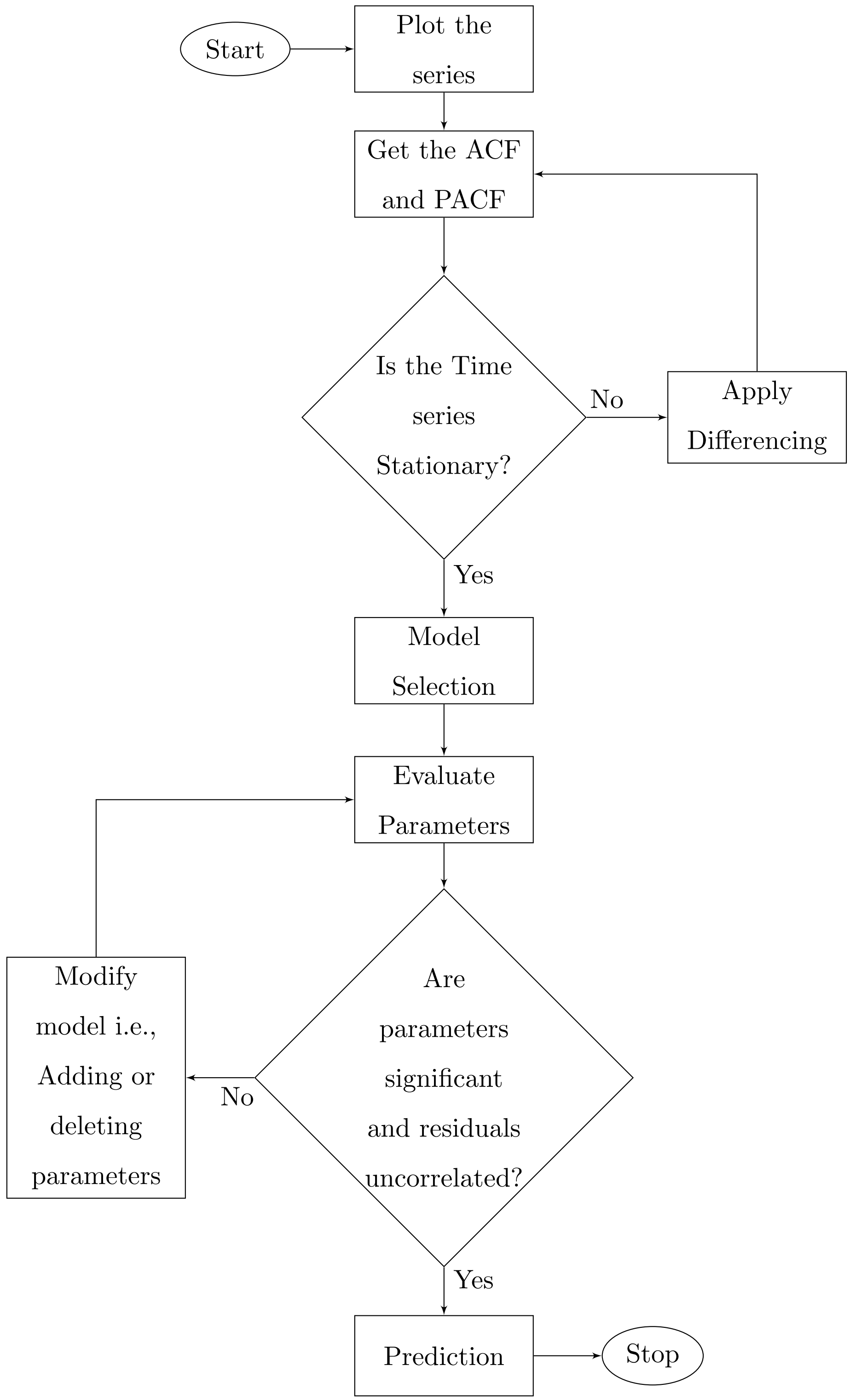


Figure 5.4: Flow Chart for building ARIMA Model and Forecasting

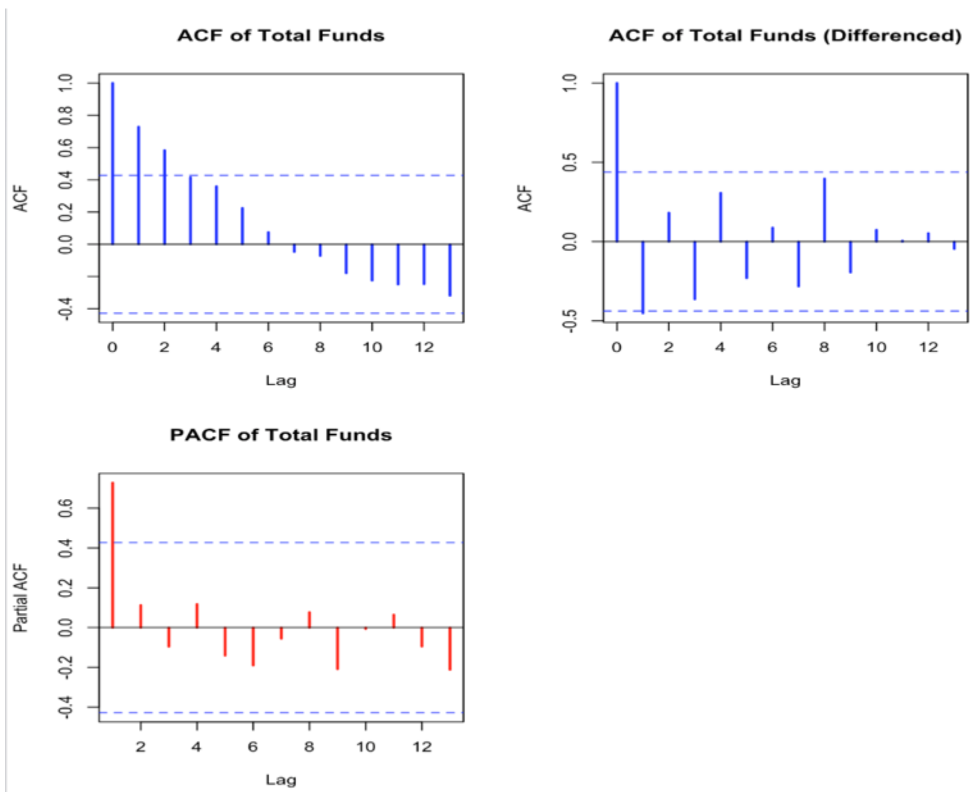


Figure 5.5: ACF and PACF Plots to Calculate (p,d,q) of ARIMA Model (Total Funds)

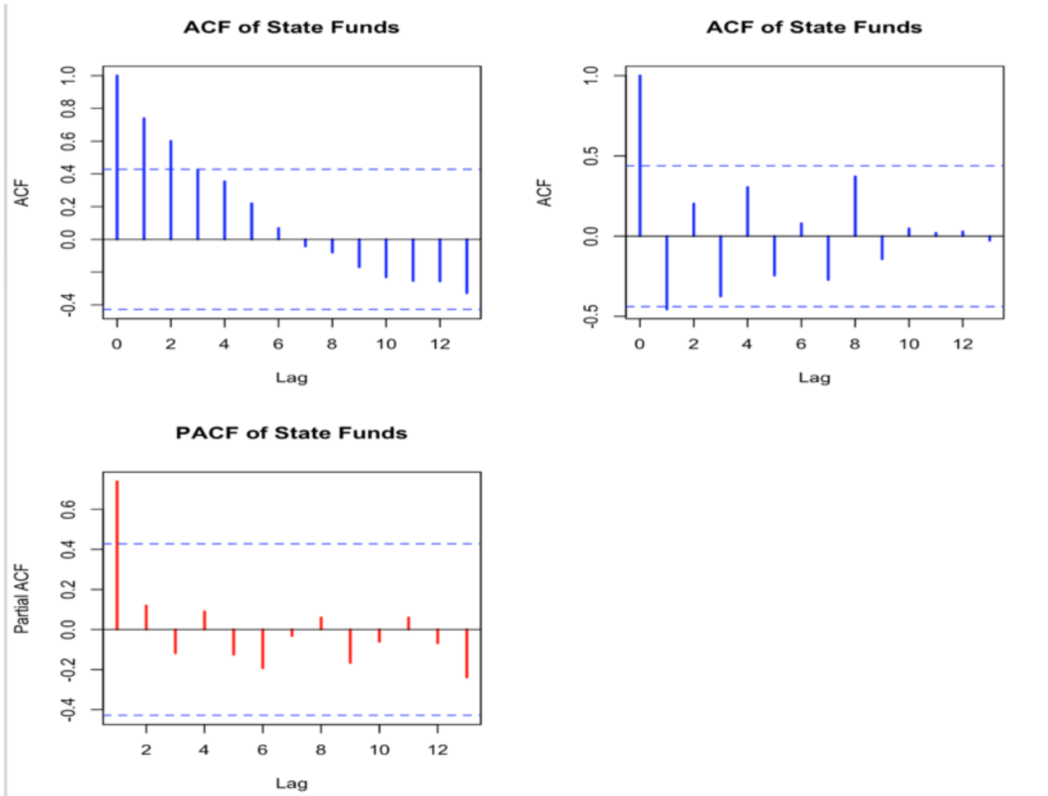


Figure 5.6: ACF and PACF Plots to Calculate (p,d,q) of ARIMA Model (State Funds)

```
> fitTotFunds <- arima(log(TotFunds), c(2, 1, 2),seasonal = list(order = c(2, 1, 2), period = 1))
> fitTotFunds

Call:
arima(x = log(TotFunds), order = c(2, 1, 2), seasonal = list(order = c(2, 1,
    2), period = 1))

Coefficients:
          ar1     ar2      ma1      ma2     sar1    sar2     sma1     sma2
      -0.6301  0.2898  -0.1610  -0.8390  -0.6301  0.2898  -0.1610  -0.8390
s.e.   0.7652  0.6941   0.4961   0.4531   0.7652  0.6941   0.4961   0.4531

sigma^2 estimated as 0.04778:  log likelihood = -2.4,  aic = 22.81
```

Figure 5.7: ARIMA model for Total Funds

```
> fitStFunds <- arima(log(StFunds), c(2, 1, 2),seasonal = list(order = c(2, 1,2), period = 1))
> fitStFunds

Call:
arima(x = log(StFunds), order = c(2, 1, 2), seasonal = list(order = c(2, 1,
    2), period = 1))

Coefficients:
          ar1     ar2      ma1      ma2     sar1    sar2     sma1     sma2
      -0.6131  0.2964  -0.1788  -0.8211  -0.6131  0.2964  -0.1788  -0.8211
s.e.   0.6795  0.6025   0.4715   0.4346   0.6795  0.6025   0.4715   0.4346

sigma^2 estimated as 0.05994:  log likelihood = -4.53,  aic = 27.06
```

Figure 5.8: ARIMA model for State Funds

model is taken as 1. The coefficients are used in the calculation of the funds after a period t. The best model is usually considered to be the one with lowest variance and lowest AIC value. We observed that the AIC value of ARIMA model with state funds is higher than the AIC value of ARIMA model with total funds. So we can say that the ARIMA model with Total funds is the best fit.

Prediction of Total and State Funds:

The forecasting or prediction of a fitted ARIMA model is done by using *predict()* command with the training data (70%) for the next 10 years.The same process is implemented for test data set (30%). We calculated error rate for both the training and test data set and achieved a lesser error rate for the training data set. So,Prediction plot (refer Figure 5.9) is drawn for training data set of total funds. From the prediction plot (5.9), it is seen that the funds allotted tend to increase for the next 10 years.

5.2 Time Series Analysis for Functionally Obsolete Structures

Time series analysis is performed to find the variation of percentage of functionally obsolete structures over the time period of 1994-2014. Initial data cleansing is performed before performing time series analysis. Imputed the data using mice package (ref Figure 5.10). Imputation is the process of replacing missing data with substituted values. Missing data creates problems for analyzing data, so imputation one of the ways to avoid the long process of deleting cases that have missing values. In R, imputation is performed with mice package. mics package helps in imputing missing values with plausible data values. These plausible values are drawn from a distribution specifically

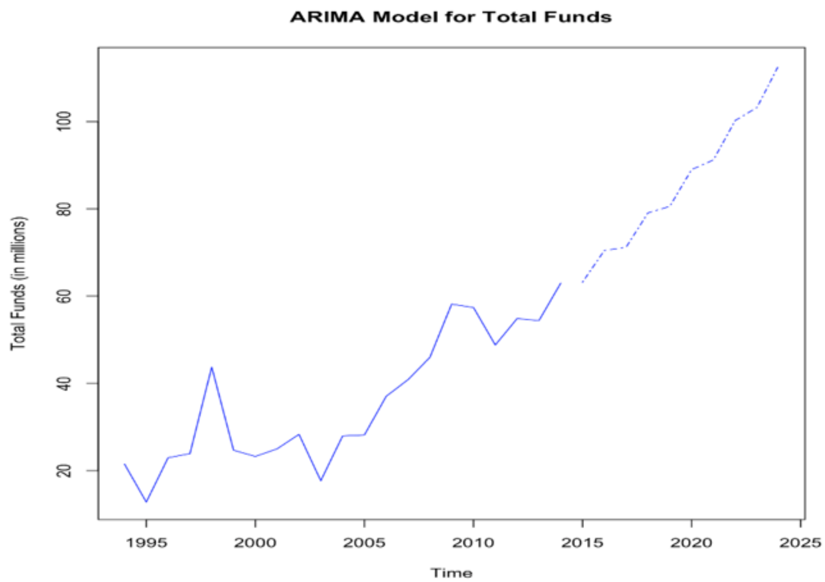


Figure 5.9: Prediction of Fitted Model for Total Funds

```
CA_FuncObsol_NA<-read.csv("/Users/haritha/Documents/Thesis/2017.1.Bridge/OCIA State Policy
methods(mice)
CA_FuncObsol_NA_impute <- mice(CA_FuncObsol_NA,m=5,maxit=50,meth='cart',seed=1)
CA_FuncObsol_NA_impute$imp$Funds
CA_FuncObsol_NA_impute <- complete(CA_FuncObsol_NA_impute,3)
str(CA_FuncObsol_NA_impute)
summary(CA_FuncObsol_NA_impute)
```

Figure 5.10: Data Cleansing using MICE package

designed for each missing data point. A time series object is created (refer Figure **??**) as mentioned in previous sections in this chapter. A time series plot is drawn at Figure 5.11. It is seen that the percentage of functionally obsolete structures decreased from 1994 to 2000 and the again increased and followed a decreasing pattern. From this, it is seen that the data follows a seasonal pattern. Hence a SARIMA model is created for the time series data.

SARIMA model for Functional Obsolesce

SARIMA is the Seasonal Auto Regressive Integrated Moving Average model.ACF and PACF plots are made to get (p,d,q) values for the Time series model. In Figure 5.12a , it is seen that there are significant auto correlations at lag 1. In Figures 5.12b, it is seen that partial correlation plots show significant spike at lag 2. We can also see that there is a sudden drop in the ACF value, which shows that the time series is stationary. Hence there is no necessity of differencing for this time

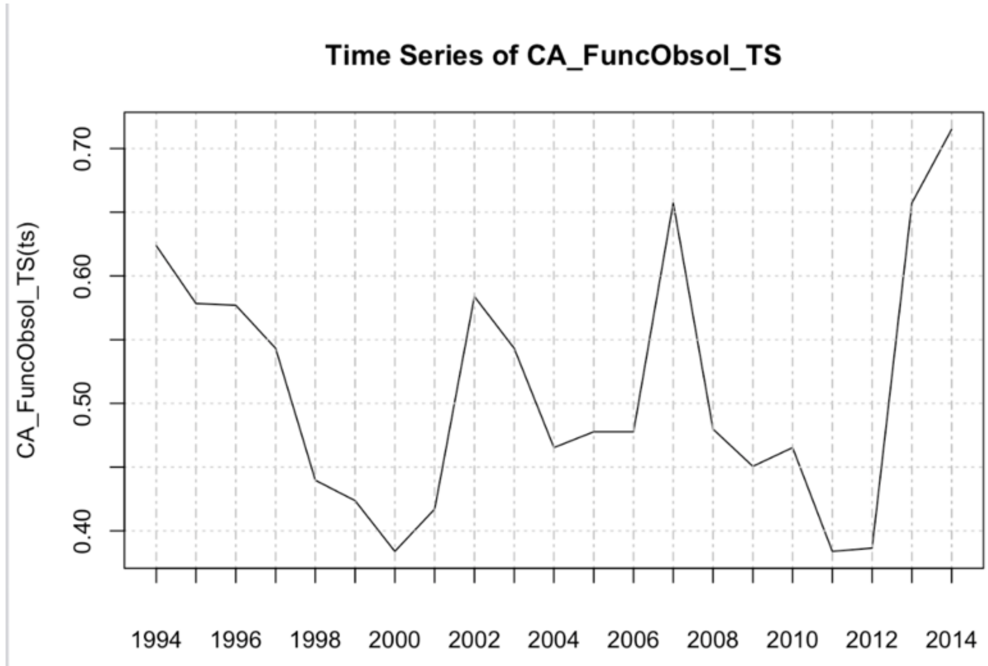
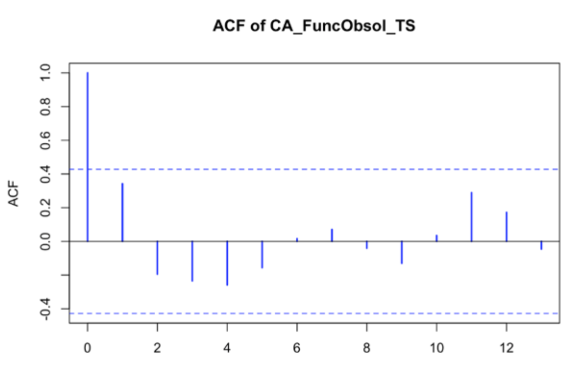
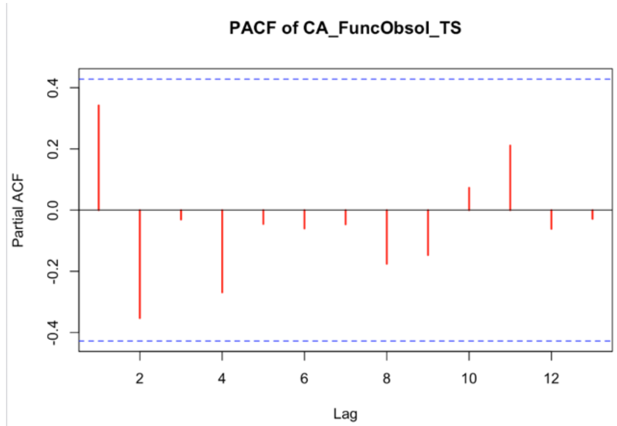


Figure 5.11: Time Series Analysis for Percentage of Functionally Obsolete Structs



(a) ACF plot to calculate (p,d,q)



(b) PACF plot to calculate (p,d,q)

```
> fitCA_FuncObsol_TS

Call:
arima(x = log(CA_FuncObsol_TS), order = c(1, 0, 2), seasonal = list(order = c(1,
    0, 2), period = 1))

Coefficients:
         ar1      ma1      ma2     sar1     sma1     sma2  intercept
      0.4721  -0.5000  -0.5000  0.4721  -0.5000  -0.5000    -0.7155
s.e.  0.3579   0.4919   0.4532  0.3579   0.4919   0.4532     0.0077

sigma^2 estimated as 0.01401:  log likelihood = 11.85,  aic = -7.71
```

Figure 5.13: SARIMA Model for Functional Obsolesce

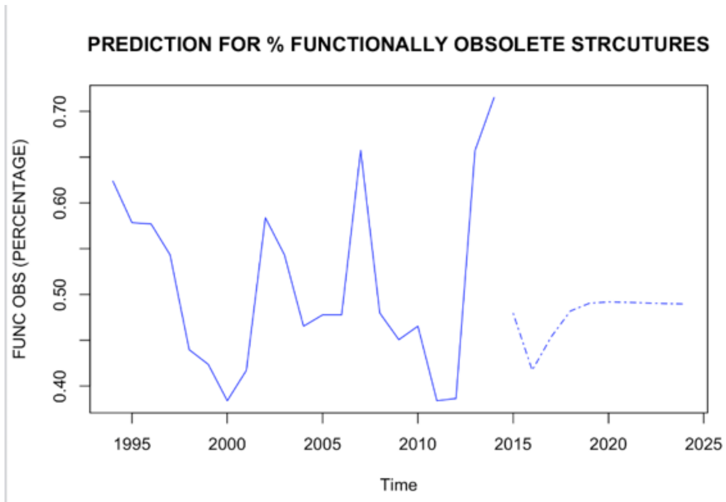


Figure 5.14: Prediction for Functional Obsolesce

series. So the SARIMA parameters (p, d, q) are chosen as $(1, 0, 2)$ as there is no differencing. With the obtained (p, d, q) values, the SARIMA model is fit (refer Figure 5.13). The results of fitted model show that the variance is very less which makes the model better. It is also seen that AIC value is negative. Negative AIC indicates less information is lost than a positive AIC and therefore a better model. As a next step in the time series analysis, predictions are made for the next 10 years for the percentage of functionally obsolete structures (refer Figure 5.14). From the Figure 5.14, it is seen that the percentage of functional obsolete structures tend to decrease in the next 10 years, which can be considered as a good sign.

Section 6

Conclusions and Future Research Directions

6.1 Conclusions

In this thesis, we presented extensive exploratory data analysis of National Bridge Inventory data. Various visualization were made to depict the trends followed in the data. Regression analysis is also carried out with the data and various regression models were fitted with Total funds, federal funds and state funds as predictor variables. Most significant factors like sufficiency rating and average age of bridges are extracted from the variables used in the regression. After analyzing the regression results thoroughly, we concluded that there is a huge potential for the optimized distribution of funds among the states to decrease the structural deficiency and functional obsolesce of the bridges. An optimization model for allocating funds was proposed. This model ensures that the funds are allocated in proportion with the number of structurally deficient structures and functionally obsolete structures. A heuristic approach was proposed for this model and was applied to the National Bridge Inventory for the year 2014. The case study showed that the funds are distributed in proportion to the number of deficient bridges. Time series models were also fitted to understand the underlying factors that produces patterns and trends in the data. ARIMA (Auto regressive Integrated Moving Average)and SARIMA (Seasonal Auto regressive Integrated Moving Average) models were fitted to predict the funds to be allotted in the future and we concluded from the models that the funds allotted tend to increase for the next 10 years. To find the variation of percentage of functionally obsolete structures, time series analysis was done and predictions showed that the functionally obsolesce decreases over the next 10 years with the optimized fund distribution.

6.2 Future Research Directions

In this section of thesis, we discuss the various future research scope for this problem and how it can be improvised. In the optimization model, we assumed that the re were no natural disasters occurring the particular year. But in general the model can be improved by including the weather conditions as a dependent variable and also by including other significant factors obtained in the regression. In this way, the funds allocated will be much more accurate.

The material of the bridge also plays an important role in the strength of the bridge. Sufficiency rating of each bridge can be recalculated by including the type of material with which the bridge is constructed. This would lead to change in the number of structurally deficient and functionally obsolete bridges, meaning that even though a bridge is considered structurally deficient (considering the current sufficiency rating), it might be still efficient if it is built with strong metals.

In the current research, we proposed a way for the optimized allocation of funds in each state. This research can be further extended by narrowing down to the optimized allocation of funds to each bridge. This can be done by considering the sufficiency rating of each bridge instead of the number of deficient bridges. In this way, we can find the respective bridges which are to be funded to reduce the structural deficiency or functional obsolesce.

Bibliography

[1] RECORDING AND CODING GUIDE FOR THE STRUCTURE INVENTORY AND APPRAISAL OF THE NATION'S BRIDGES. Report No. FHWA-PD-96-001. Office of Engineering, FHWA, U.S. Department of Transportation, Bridge Division, Bridge Management Branch, Washington, D.C, 1995.

[2] Hegazy, T., Elbeltagi, E., and Elbehairy, H. (2004). Bridge Deck Management System with Integrated Life Cycle Cost Optimization. Transportation Research Record: Journal of the Transportation Research Board, No. 1866, TRB, National Research Council, Washington, D.C., pp. 44-50.

[3] Lee, Jaeho., Kamalarasa Kamalarasa, Michael Blumenstein & Yew-Chaye Loo(2008). Improving the reliability of a Bridge Management System (BMS) using an ANN-based Backward Prediction Model (BPM). ***Automation in Construction***.

[4] Gattulli, Vincenzo & Leonardo Chiaramonte (2005). Condition Assessment by Visual Inspection for a Bridge Management System. ***Computer-Aided Civil and Infrastructure Engineering***. Vol.20(1). p95-107

[5] EDGAR P. SMALL, TERRY PHILBIN, MICHAEL FRAHER, GEORGE P. ROMACK. Current Status of Bridge Management System Implementation in the United States. ***TRB Transportation Research Circular***

[6] Kumares C. Sinha & Tien F. Fwa. On the Concept of Total Highway Management. ***TRANSPORTATION RESEARCH RECORD 1229***

[7] FOAD FARID, DA VID W. JOHNSTON, BASHAR S. RIHANI, AND CHWEN-]INQ CHEN. Feasibility of Incremental Benefit-Cost Analysis for Optimal Budget Allocation in Bridge Management Systems. ***TRANSPORTATION RESEARCH RECORD 1442***

[8] Nasir G. Gharaibeh, Michael I. Darter and Donald R. Uzarski (1999). DEVELOPMENT OF PROTOTYPE HIGHWAY ASSET MANAGEMENT SYSTEM. ***J. Infrastruct. Syst.*** . Vol.5(1). p61-68.

[9] Mohamed, H. A., A. O. Abd-El-Halim, and A. G. Razaqpur. Use of Neural Networks in Bridge Management Systems. ***In Transportation Research Record 1490***. TRB, National Research Council, Washington, D.C., 1995, pp.1-8.

[10] National Research Council (U.S.). Transportation Research Board. (1995). Management and maintenance of bridge structures. Washington, D.C.: National Academy Press.

[11] Highway Bridge Replacement and Rehabilitation Program. 10th Annual Report of the Secretary of Transportation to the Congress of the United States. Bridge Division, Office of Engineering, FHWA, U.S. Department of Transportation, 1991.

[12] Ayaho Miyamoto, Kei Kawamura and Hideaki Nakamura (2001). Development of a bridge management system for existing bridges. ***Advances in Engineering Software***. Vol.32(1). p821 - 833.

[13] Zucchini W & Nenadic O (2011). Time series analysis with R-Part I. ***Document de cours***.

[14] Highway Statistics (2007). ***Book***. http://www.fhwa.dot.gov/policyinformation/statistics/2007/

[15] Gharaibeh, Nasir G., Yi-Chang Chiu & Patrick L. Gurian (2006). Decision Methodology for Allocating Funds across Transportation Infrastructure Assets ***Journal of Infrastructure Systems***. Vol. 12(1). p1-9.

[16] Raiffa, H., Pratt J. W. & R. Schlaifer (1995). Introduction to statistical decision theory. across Transportation Infrastructure Assets ***Book. MIT Press, Cambridge, Mass.***.

[17] Highway Capacity Manual (2000). Transportation Research Board. ***Transportation Research Board***. http://www.gsweventcenter.com/Draft_ SEIR_ References%5C2000_TRB.pdf

[18] Chan, Weng Tat., T. F. Fwa & J. Y. Tan (2003). Optimal Fund-Allocation Analysis for Multidistrict Highway Agencies. ***Journal of Infrastructure Systems***. Vol. 9(4).

[19] http://www.usinflationcalculator.com