ISTM 6214 Foundations of AI: American Recovery and Reinvestment Act of 2009 (ARRA) Data Analysis

Rachel Aska

2025-02-05

During the financial crisis, the **American Recovery and Reinvestment Act of 2009 (ARRA)** was put into effect to stimulate the economy. Numerous initiatives were sponsored as part of this endeavor to aid in economic rehabilitation and growth. These expenses are documented on the **Federal Stimulus Data** from NYC Open Data, which includes details on the funding amount, project progress, and payments made.

To begin with our analysis, we will be following the next steps:

1. Importing all the necessary libraries.
2. Verifying the content of the dataset
3. Calculating the Total Sum of Payment.Value column.
4. Calculating the Mean of the Payment.Value column.
5. Creating a subset of the data to output the projects whose Project.Status is equal to the completed 50% or more, while excluding fully completed ones.

Importing Necessary Libraries

The following libraries are used in this analysis:

```
library(glue)      #Used to format strings with variables
library(dplyr)     #To manipulate and filter data
library(knitr)     #To create neat tables
library(DT)        #To create interactive Data Tables
library(ggplot2)   #Used to create visualization plots
library(knitr)     #Used to create a report with tables and plots
library(webshot2) #Used to capture screenshots of interactive visuals
```

The next step is to import the fed_stimulus.csv file into RStudio:

```
fed_stimulus <- read.csv("/Users/DELL/Desktop/mydata/fed_stimulus.csv")
```

Next, we will also be corroborating the column names as well as the structure of the Payment.Value column.

```
colnames(fed_stimulus)
```

```
##  [1] "Project.Name"
##  [2] "Project.Description"
##  [3] "Stimulus.Tracker.ID"
##  [4] "Funding.Category"
##  [5] "Funding.Source"
##  [6] "Stimulus.Funding"
##  [7] "Displaced.City.Funding"
##  [8] "All.Other.Funding"
##  [9] "Award.Lead.City.Agency"
## [10] "Project.Lead.City.Agency"
## [11] "Project.Status"
## [12] "X..of.Funds.Spent"
## [13] "Date.Funds.Awarded.by.Fed.State"
## [14] "Date.Funds.Announced.by.NYC"
## [15] "Estimated.Start.Date"
## [16] "Actual.Start.Date"
## [17] "Actual.Completion.Date"
## [18] "Interim.Spending.Deadline"
## [19] "X..of.Funds.to.be.Spent.by.Interim.Spending.Deadine"
## [20] "Final.Spending.Deadline"
## [21] "Contract.Name"
## [22] "Contract.Method"
## [23] "Contract.Status"
## [24] "Contract.ID.."
## [25] "Vendor.Name"
## [26] "Contract.Start.Date"
## [27] "Contract.End.Date"
## [28] "New.or.Existing.Contract"
## [29] "Revised.Contract.Start.Date"
## [30] "Revised.Contract.End.Date"
## [31] "Contract.Value"
## [32] "Payment.Recipient"
## [33] "Payment.Type"
## [34] "Payment.Date"
## [35] "Payment.Description"
## [36] "Payment.Id"
## [37] "Payment.Value"
```

```r
str(fed_stimulus$Payment.Value)
```

```
##  num [1:13553] NA 123241 NA 1306549 NA ...
```

Calculating the summary statistics:

1. Calculating the Total Sum for Payment Value:

If we observe our fed_stimulus.csv dataset, it has a lot of Null values in the Payment.Value column. If a dataset contains missing values, these functions will return NA instead of a result unless you explicitly handle them. In order to deal with them we use na.rm = TRUE" to ignore any of the missing values.

```r
Total_Sum_of_Payments <- sum(fed_stimulus$Payment.Value, na.rm = TRUE)
glue("The Total Sum of the Payments Value column is: {Total_Sum_of_Payments} (in Billions)")
```

```
## The Total Sum of the Payments Value column is: 3180464993.77 (in Billions)
```

2. Calculating the Mean of the Payment.Value column:

```
Mean_of_Payments <- mean(fed_stimulus$Payment.Value, na.rm = TRUE)
glue("The Average Payment Value is: {Mean_of_Payments} (in Thousands)")
```

```
## The Average Payment Value is: 242044.520073818 (in Thousands)
```

Creating a Subset of the fed_stimulus.csv dataset:

In the next steps we will create a subset of the dataset where the Project.Status column is Completed 50% while excluding Fully Completed projects. We will also only display Project.Name, Project.Value and Project.Status columns to help the viewer understand better and avoid confusion.

```
subset_dataset <- fed_stimulus %>%
  filter(Project.Status == "Completed 50% or more" & Project.Status != "Fully Completed") %>%
  select(Project.Name, Payment.Value, Project.Status)
```

To display the subset_dataset, I chose to display it using kable().

a. Displaying the subset_dataset in a table format using kable()

```
kable(subset_dataset[1:10, ], caption = "Projects Completed 50% or More (Page 1)")
```

Table 1: Projects Completed 50% or More (Page 1)

Project.Name	Payment.Value	Project.Status
830 Amsterdam Houses - Rooftop Water Tank	NA	Completed 50% or more
CUNY pilot program for higher priced trainings in demand occupations expansion	1917.81	Completed 50% or more
CUNY pilot program for higher priced trainings in demand occupations expansion	1917.81	Completed 50% or more
Whitman-Ingersoll Houses - Apartment Renovations 2	NA	Completed 50% or more
Highbridge Gardens Houses - Brickwork Repair and Roofing Replacement	NA	Completed 50% or more
CUNY pilot program for higher priced trainings in demand occupations expansion	16.50	Completed 50% or more
Livonia Terrace	NA	Completed 50% or more
Interior Compactors at Various Manhattan Developments	148400.00	Completed 50% or more
FDNY Emergency Medical Services training program	NA	Completed 50% or more
Adoption Assistance Title IV-E	NA	Completed 50% or more

```
kable(subset_dataset[11:20, ], caption = "Projects Completed 50% or More (Page 2)")
```

Table 2: Projects Completed 50% or More (Page 2)

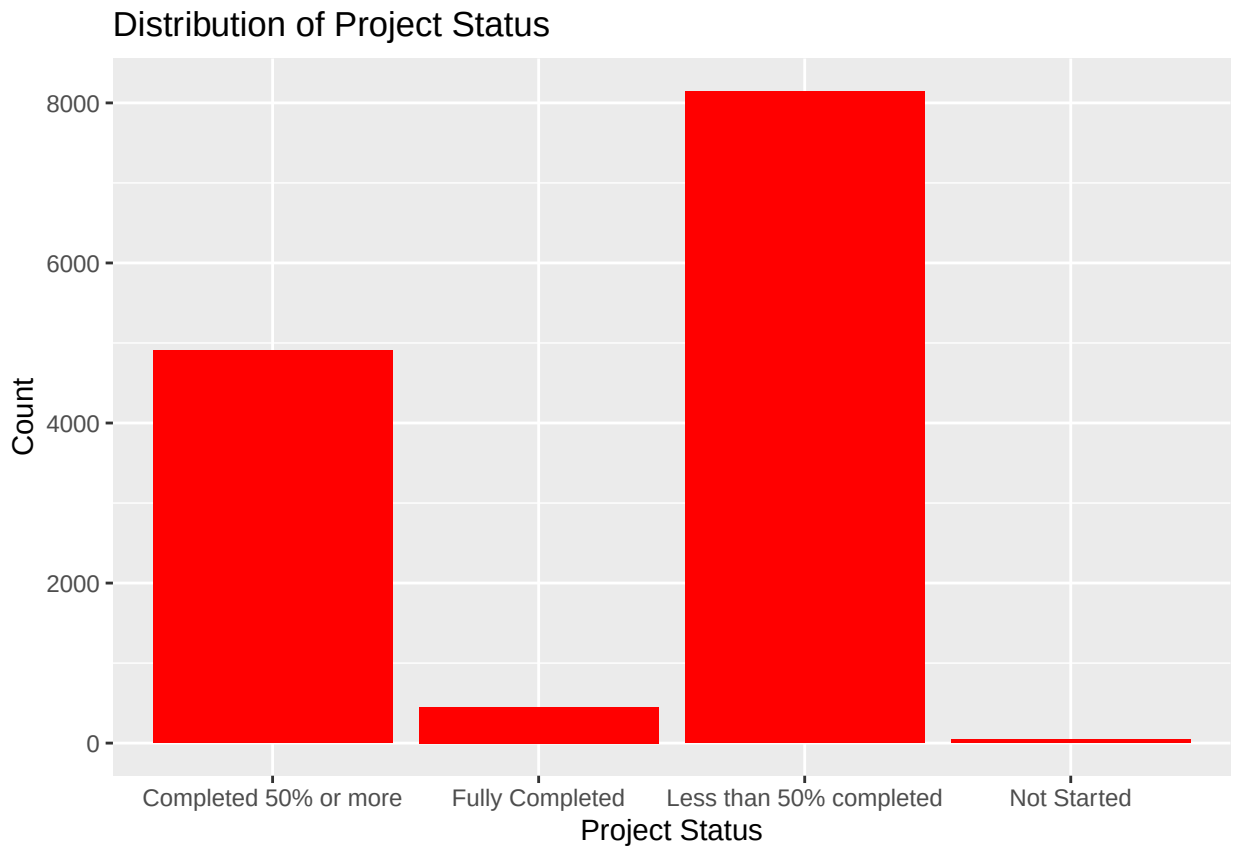
	Project.Name	Payment.Value	Project.Status
11	Older Americans Community Service Employment Program (NCOA Grant)	19680.43	Completed 50% or more
12	Code Violation Removal in Schools	601439.42	Completed 50% or more
13	ARRA Administration	118365.94	Completed 50% or more
14	Parks Opportunity Program - “Spark” electrician training	NA	Completed 50% or more
15	CUNY pilot program for higher priced trainings in demand occupations expansion	1917.81	Completed 50% or more
16	Mott Haven Houses - Structural Repairs to Rooftop Water Tank Enclosure	NA	Completed 50% or more
17	NYC Civic Corps	41602.50	Completed 50% or more
18	Go to School, Get a Job	NA	Completed 50% or more
19	Conlon Lihfe Towers Houses - Roofing Replacement	NA	Completed 50% or more
20	Individual Training Grants - Increase availability	6958.00	Completed 50% or more

Creating Visualizations:

Visualizations are crucial in R, particularly when dealing with data analysis and statistical modelling. It helps us identify the patterns and trends which might not be that easy to catch in raw data. It also helps us present our findings to to non-technical

a. Creating a visualizations with Bar Plot:

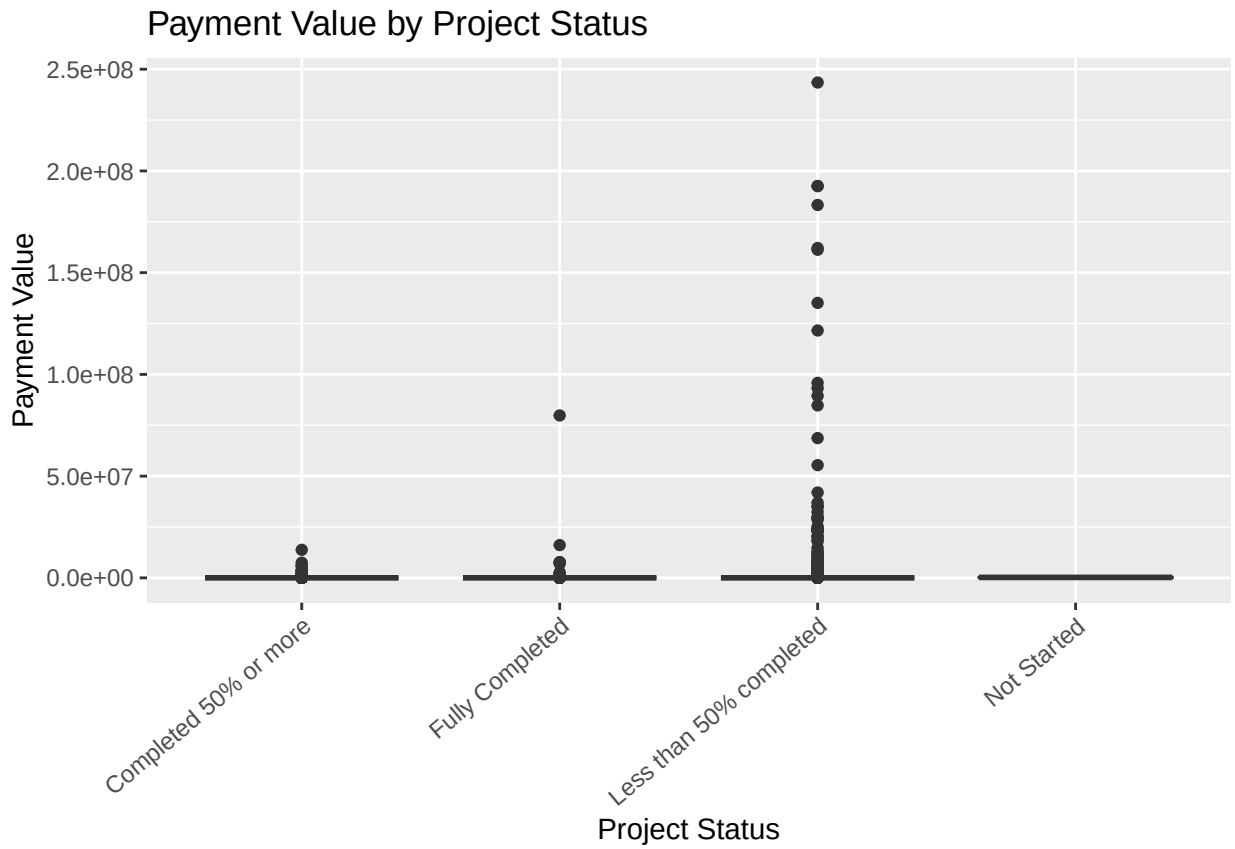
```
ggplot(fed_stimulus, aes(x = Project.Status)) +
  geom_bar(fill = "red") +
  labs(title = "Distribution of Project Status",
       x = "Project Status",
       y = "Count")
```



In the above bar plot, it is observed, that the majority of projects fall under Less than 50% completed. A significant number of projects have Completed 50% or more and few projects are fully completed. We can infer that many projects struggle to reach completion,

b. Creating Visualizations with Box Plot to compare the distribution of payment accross status:

```
ggplot(fed_stimulus, aes(x = Project.Status, y = Payment.Value)) +
  geom_boxplot(fill = "blue") +
  labs(title = "Payment Value by Project Status",
       x = "Project Status",
       y = "Payment Value") +
  theme(axis.text.x = element_text(angle = 40, hjust = 1))
```



In the above Box Plot, we can observe that many projects struggle to reach completion, and most are stagnant at an incomplete status.

Conclusion:

In this analysis, we delved into the American Recovery and Reinvestment Act of 2009 (ARRA) dataset to compute the summary statistics and determine projects that were more than 50% complete and while excluding the ones already completed. In my analysis, it was found that the overall sum of payments was $3,180,464,993.77 (or 3.18 Billion) while the average value of a payment is $242,044.52. These results indicate large investments within federal stimulus projects; however, the presence of partially completed projects suggests possible delays or inefficiencies.

The above analysis was limited by the availability of project timelines and did not include any data on project costs. Given the low completion rate, further analysis into the causes of delay to improve completion rates is would be suggested. We also observe Fully completed projects are rare, which may indicate inefficiencies in project execution or funding allocation.

To conclude, this report emphasizes the monitoring of federal stimulus projects to see that funds are used both on-time and effectively. By addressing the challenges we identified, policy makers can make the most out of these vital investments for economic recovery.

References:

1. Prof Shubhash Sharma (2025). ISTM 6214: Foundations of AI. Lecture notes, [George Washington University] from (https://blackboard.gwu.edu/ultra/courses/_415418_1/cl/outline)
2. Open Data NYC. (n.d.). Federal Stimulus Data. Data City of New York. Retrieved February 6, 2025, from (https://data.cityofnewyork.us/Business/Federal-Stimulus-Data/ivix-m77e)
3. DeepSeek AI. (2025). AI Drafting Assistance. Retrieved from (https://chat.deepseek.com/a/chat/s/92e2e08d-151e-441f-8793-553ee01cf099)
4. Packages: {ggplot2}, {dplyr}, {tidyverse}, {knitr}, {rmarkdown}
5. Software: R, RStudio, TeX Live, or Overleaf.