

BUSINESS AND SPORTS ANALYTICS

Collaborated with Agent Rouge

For the assignment, we picked the 'Federal Stimulus Dataset' from New York City which is an open dataset which includes attributes ranging from the projects in the city, funding categories, funding sources, the funding amount and the amount of funds spent on these projects. The funding for the projects was the field that we found most relevant from the dataset that falls in the business category. The average amount of funding across the various funding categories is shown in the visualization below in Fig a. The average funding is high for Medicaid Relief and Education compared to the other categories and hence we have plotted another visualization excluding these categories shown in Fig b.

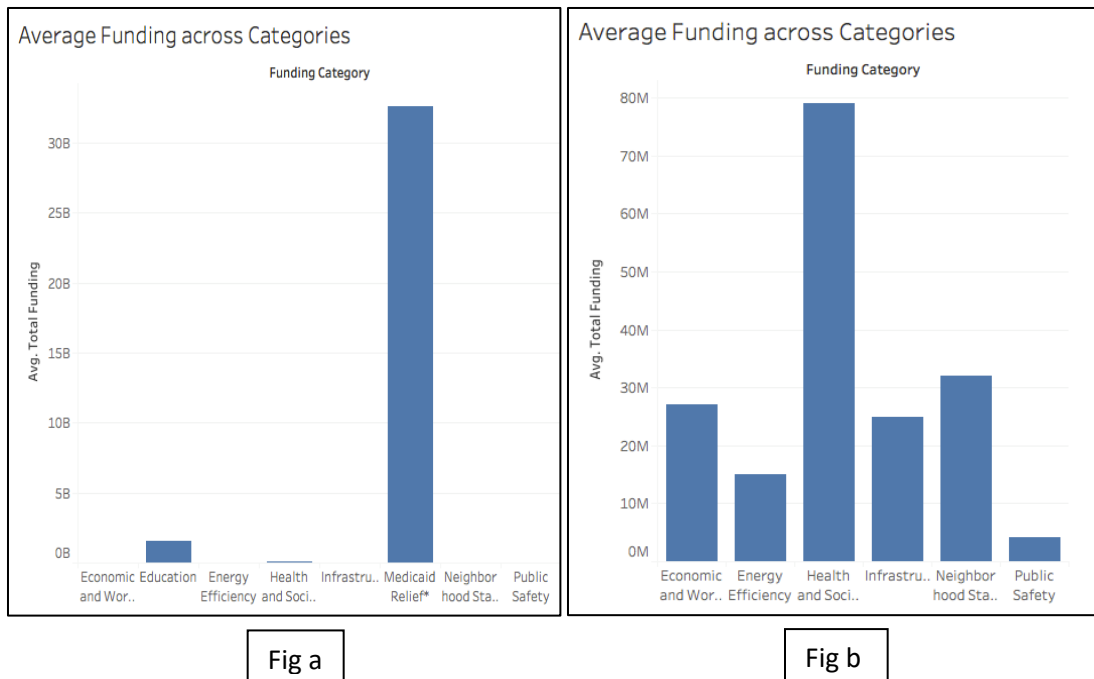


Fig a

Fig b

Main Insights:

The important insights based on the funding for categories in the dataset that we came across are as follows:

1. Comparing the amount of funds spent on projects for a category versus the total funding for the respective category.
 - On comparison we found that there were categories for which the average amount of funds spent were greater than the average funding for the project. The visualization for the same is shared in the evidence section as shown in Fig c. This can help the Mayor's office of Operations in the city of New York to list down flaws in the budget for Medicaid Relief, Infrastructure and Public Safety and either provide more additional funding or look into utilization of these funds more efficiently.
2. The relation between sources of funding for categories to the timeliness of the projects in the respective categories.
 - From the visualization shared in Fig e and Fig f we observe that, a majority of delayed projects for a category belong to a particular funding source. The office of Operations can look into these funding sources to check for reasons like: whether funds from these sources

are received on time, is the delay in funding due to non-compliance of norms by the source or disagreement between the parties involved.

Evidence:

We have shared detailed visualizations created in Tableau as a part of the evidence for the insights mentioned above.

1. Average funding versus Average amount spent across categories of projects.

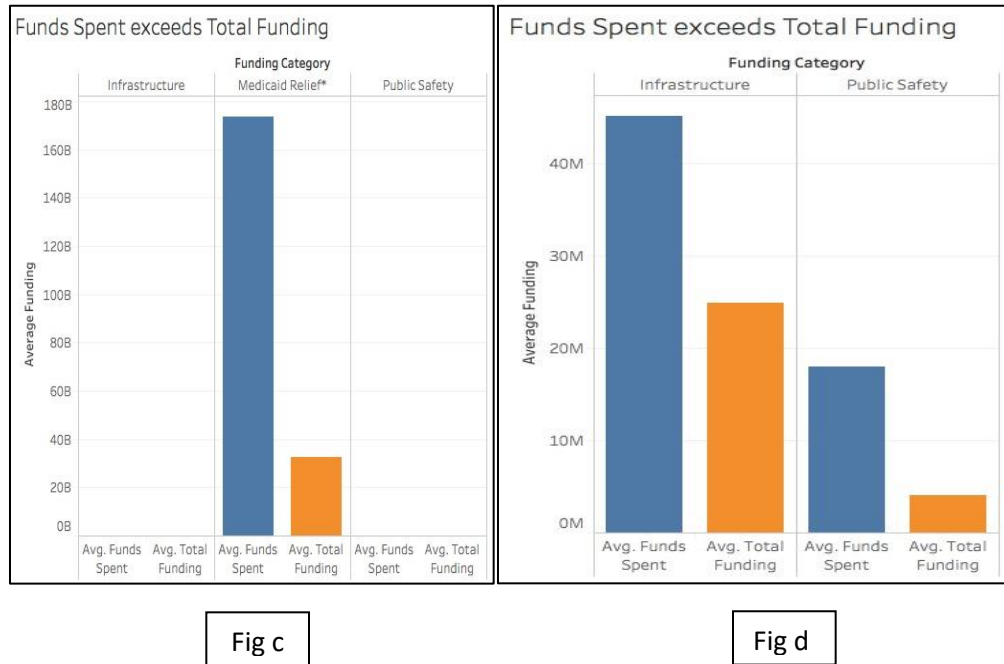


Fig d is the same visualization as in Fig c excluding the bar graph for Medicaid relief to display the value for the other two categories.

2. Sources of funding versus categories to the timeliness of the projects

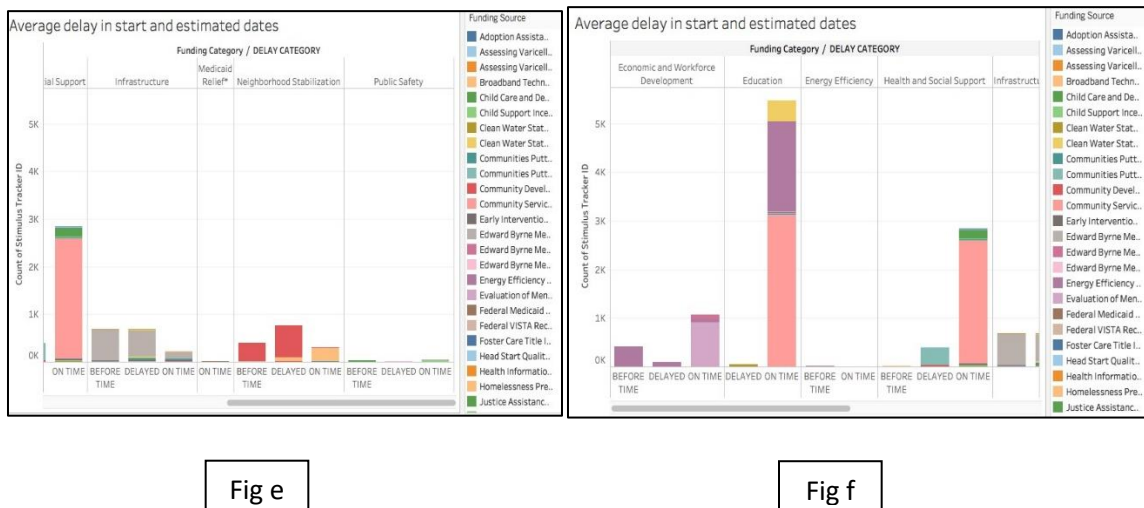


Fig g and Fig h shown below are a numeric representation of the number of projects delayed in each of the categories based on the funding sources.

Average delay in start and estimated dates										
DELAY CATEGORY	Funding Category / Funding Source									
	Infrastructure									Medicaid Relief*
	Broadband Tech..	Clean Water S..	Clean Water S..	Public Housin..	Surface Transp..	Surface Transp..	Transit Capital..	Transit Capital..	Transit Capital..	Federal Medicaid..
BEFORE TI..		4		656		2			27	
DELAYED		12	3	561	41	48	7	15	2	
ON TIME	2	9		153		38	1		23	9

Fig g

Average delay in start and estimated dates													
DELAY CATEGORY	Funding Category / Funding Source												
	Neighborhood Stabilization					Public Safety							
	Community Dev..	Federal VISTA R..	Homeless P..	Neighborhood St..	Tax Credit A..	Edward Byrne ..	Edward Byrne ..	Edward Byrne ..	Justice Assista..	Justice Assista..	Port Securit..	Port Securit..	Transportation
BEFORE TI..	376	5			12				28				
DELAYED	673		75		19			18	2		1		
ON TIME	2		280	1	20	8	2		20	22		3	

Fig h

Method:

The method describes how most of the derived columns were obtained and the data required for each of the visualizations shared as a part of the assignment.

1. We derived the column for average funding by computing the average of a derived column named total funding. The derived column named total funding is obtained by summation of the following columns in the dataset: stimulus funding, displaced city funding, other funding.
2. The Funds spent was obtained by using the factor given as percentage of funds spend as of date.
3. These derived columns along with the categories for funding were used for the first visualization.
4. The next derived column for timeliness of a project was obtained by categorizing based on the difference between actual start date and estimated start date. The categories were as follows:
 - a. Difference is 0 – On time,
 - b. Difference is negative – Before time
 - c. Difference positive – delayed
5. This column along with the funding source and funding category were used for the second visualization in the bar graph format and the tabular format with numerical data.
6. The value for the rows and columns is easy to understand for each of the visualization.