MathWorks Math Modeling Challenge 2020

Huron High School

Team # 13403 Ann Arbor, Michigan

Coach: Philip Eliason

Students: Anirudh Cowlagi, Julie Heng, James Xiu, Benjamin Zhang, Derek Zhu

M3 Challenge Technical Computing Award
THIRD PLACE—$1,000 Team Prize

JUDGE COMMENTS

Specifically for Team # 13403:
This paper used technical computing for advanced curve fitting in Part 1 of the problem. They also used numerical integration to find a demand curve for charging along a route, and a greedy algorithm for selecting charging locations based on that curve. One of the strengths of this paper was great communication: at all points the algorithms used were clearly explained in the main text of the paper. Flow-charts were used to make the approach even more clear. Finally, the team leveraged technical computing to produced plots and visualizations that effectively communicated both their approach and final solution to the reader. While we enjoyed the paper's straightforward approach (simple models are often the best models!) the teams which ranked above it introduced more creativity and complexity into the modeling process.

Overall Judging Perspective for Technical Computing Submissions:
The use of technical computing in the final papers was judged on its effectiveness in advancing a papers' modeling, its creativity, and how it was communicated. We rewarded papers where technical computing was used in an essential way, and did not simply replace functionality which could have been implemented in a spreadsheet or on a calculator. We also rewarded clear explanations, even if the underlying algorithm was relatively simple. This year we noticed that technical computing was used in many papers to enhance presentation: some beautiful plots were used to effectively communicate student ideas and final solutions. Such uses of technical computing were also rewarded. Finally, one of the benefits of implementing a model in code is that it is very easy to change modeling choices or parameters to see how these choices effect the final problem solution. Teams which took advantage of this (through sensitivity analysis, testing multiple models, etc.) were rewarded.

Keep on Trucking: *U.S. Big Rigs Turnover from Diesel to Electric*

Team # 13403

Executive Summary

Trucking has grown significantly as a freight distribution mechanism since the construction of the interstate highway system in the late 1950s and early 1960s [1], and the industry is not expected to slow down anytime soon. In the U.S., diesel-fueled semi trucks drive over 150 billion miles annually, enough to reach the moon over 600,000 times [2] [4]. Now, given increasing awareness around the social-environmental costs of diesel burning and reasonable technological innovations to increase fuel efficiency, the trucking industry is poised for electrification. As such, we developed relevant mathematical models to forecast the expansion of electric trucking and its associated costs and benefits in the near future.

We began by anticipating diesel and electric semi truck demographics in 5, 10, and 20 years. We performed a linear regression to determine the demand for trucks and evaluated the point in a new diesel truck's lifespan when it should be replaced to maximize cost efficiency. A gamma/Gompertz distribution was used to estimate the change in age distribution of diesel trucks over time. This mathematical model concluded that all diesel trucks from the current population would be phased out 9 years from now. In other words, 60.17 percent of the 3,154,725 semi trucks on the road in 2025 would be electric, and 100 percent would be electric in 2030 and beyond.

The next step in sustainable electric truck growth is maximizing efficiency of its underlying infrastructure. For this reason, we optimized the number of charging stations and individual chargers per station to sustain truck freight while minimizing traffic along major trucking corridors. We did this by determining the spacing between charging stations and effectively determining the number of Level 3 DC fast chargers present at each charging location. We accomplished the former by relying upon data relating to mean annual traffic along the corridor, a trapezoidal numerical integration scheme, and an iterative algorithm to construct the set of stations based on a computed cutoff (dependent on electric truck battery information, total distance traveled, total charges required, etc.).

Finally, we aimed to create a model that prioritizes national corridors for development, based on community economic, environmental, and social motivations. This method determined the monetary value of four major factors (difference in operating and manufacturing costs, environmental impact based on emissions, cost to build infrastructure, and total communal profit) and was able to rank how to prioritize development along each of the five corridors. We find this order to be, in decreasing development priority: Jacksonville-Washington, D.C., San Antonio-New Orleans, Minneapolis-Chicago, Los Angeles-San Francisco, Boston-Harrisburg.

The trucking industry is ready for a dramatic transformation, and electrification is on the horizon.

Contents

1 Global Assumptions

1. *Class 8 vehicles, including long-haul, short-haul, and regional haul semi trucks, will henceforth be referred to as "trucks."* Light trucks such as pickup trucks will not be included.

2. *Truck type is binary, either diesel or electric.* For simplicity and due to lack of striated data, hybrid trucks will be ignored. For comparison, all trucks, regardless of fuel production, can travel the same distance in the same time frame.

3. *An outlet at a charging station that can serve one vehicle will be called a "charger."* Using a typical gas station as an analogy, a gas station is to a charging station as a fuel pump is to a charger.

2 Problem 1: Shape Up Or Ship Out

2.1 Problem Restatement

Create a model that predicts semi truck demographics in the years 2025, 2030, and 2040, assuming the existence of all essential electric truck infrastructure.

2.2 Local Assumptions

1. *The lifespan of a truck, whether diesel or electric, will be held constant at 12 years* [2]. We assume that all trucks have the same workload, i.e., any electric truck travels the same distance in the same time interval as a diesel truck. This was verified by dividing the lifetime mileage of an electric truck by its annual mileage.

2. *After a 12-year lifespan, trucks will be considered "spent," and any spent truck will be replaced by an electric truck.* This is an approximation, knowing that lifespans can be longer or shorter based on premature accidents, mechanistic failures, etc. Knowing that electric trucks are more cost-efficient and that many major companies are making the switch to electric fleets, any spent truck will be replaced by an electric truck.

3. *The purchase and operating costs of diesel trucks is constant over time.* Marginal vehicle-based (maintenance, tolls, permits, etc.) and driver-based (wages, benefits, etc) costs [5] were aggregated. Inflation over time was not considered.

4. *Electric truck statistics are based on Tesla Semi projections* [6]. Given that current electric trucks are limited and no single such model has become standard, we are basing our estimates on a model at the industry forefront. The Tesla Semi anticipates a million mile battery and 500 mile range on a single charge [7] [8], technologies we can assume are already established within the context of this problem.

5. *The demand for motor vehicles is linear.* Since the United States population growth rate has been relatively constant through the past few decades [9], we can assume that the demand growth rate for motor vehicles, which is proportional to the population, is also constant.

6. *The distribution of the proportion of household vehicles by age is analogous to the distribution of the proportion of trucks by age.* We assume that the average lifespans of household vehicles and trucks are similar.

2.3 Variables

Table 1: Variables

Variable	Description	Units
t	Time in years after 2019	years
$C(t)$	Total number of trucks in year t	trucks
$D(t)$	Number of diesel trucks in year t	trucks
$E(t)$	Number of electric trucks in year t	trucks
$R(t)$	Number of diesel trucks replaced by electric trucks in year t	trucks
P_D	Purchase cost of 1 diesel truck	dollars
P_E	Purchase cost of 1 electric truck	dollars
O_D	Operational cost of 1 diesel truck	dollars/mile
O_E	Operational cost of 1 electric truck	dollars/mile
t^*	Age of a truck	years

2.4 Solution

To model the total number of electric trucks, $E(t)$ for each year, we needed to model the number of total trucks ($C(t)$) and diesel trucks ($D(t)$) for each year. Since trucks are either electric or diesel, we used the equation

$$E(t) = C(t) - D(t) \tag{1}$$

2.4.1 Determination of C(t)

First, we modeled $C(t)$. We began by finding the general regression for growth of truck population from 1999 to 2017; data was sourced from the Bureau of Transportation Statistics [11].

We performed a linear regression for this data, as we assumed that the demand for motor vehicles in the United States is approximately linear. Our resulting equation had $R^2 = 0.738$.

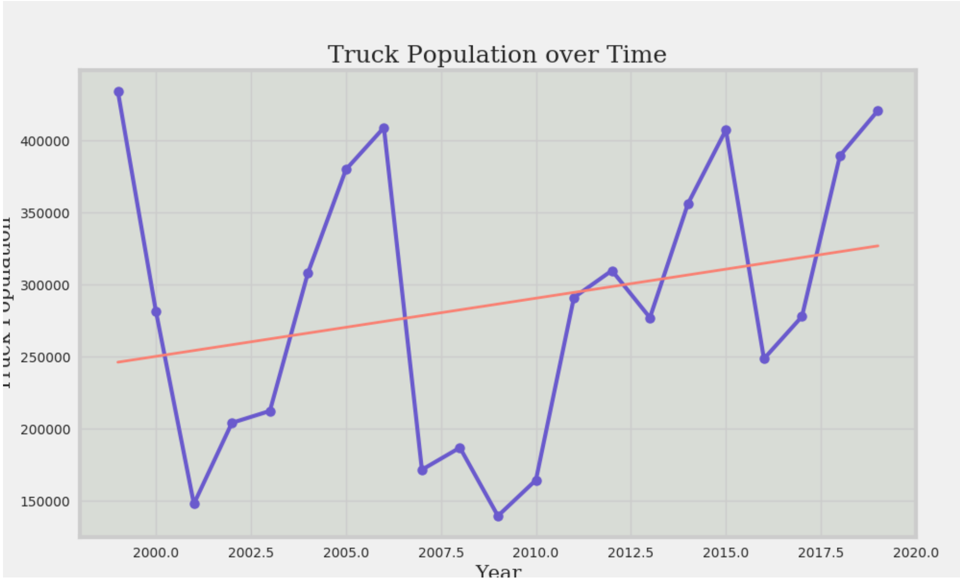

Figure 1: Flowchart Describing Numerical Implementation

$$C(t) = 44669t + 2886711 \tag{2}$$

2.4.2 Determination of $D(t)$

Next, we found the point in time at which a relatively new diesel truck should be replaced by an electric one. We calculated this age t^* by solving the piecewise inequality involving the respective purchase and operational costs of both diesel and electric trucks when the time to the end of lifespan is greater than seven years:

$$P_E + O_E(7 * 80,000 + 110,000(5 + t^*)) < P_D + O_D(7 * 80,000 + 110,000(5 + t^*)) \tag{3}$$

Solving the inequality yields $t^* < 3.6$.

Next, we needed to determine $D(t)$. We sought to determine the distribution of diesel trucks as a function of age. This was determined by utilizing existing data about the age proportion distribution of household vehicles [10] and generalizing these results to diesel trucks using **Assumption 5**. This yields the data table shown below:

Table 2: Truck Age Distributions in 2019

Age	Proportion	Age	Proportion
$0-1$	0.067	$11-12$	0.056
$1-2$	0.069	$12-13$	0.054
$2-3$	0.067	$13-14$	0.046
$3-4$	0.063	$14-15$	0.042
$4-5$	0.060	$15-16$	0.039
$5-6$	0.051	$16-17$	0.035
$6-7$	0.052	$17-18$	0.029
$7-8$	0.044	$18-19$	0.024
$8-9$	0.057	$19-20$	0.021
$9-10$	0.056	$20-21$	0.016
$10-11$	0.054		

Now, in order to determine the number of diesel trucks present at any time t, we simply determined the number of diesel trucks that are replaced at the end of each year by electric trucks, following **Assumption 4**. Following the work in the previous section, we simply "replaced" the diesel trucks whose ages fall in the following range: $(0, t^*) \cup (12, \infty) = (0, 3.6) \cup (12, \infty)$. Furthermore, we were able to approximate our age distribution using a scaled gamma/Gompertz distribution [12] due to the following reasons:

1. Given that it is traditionally used as an aggregate-level model of customer lifetime, we may apply it to an analogous situation modeling age distribution of diesel trucks
2. The gamma/Gompertz distribution has nonzero probability density for $t = 0$
3. The gamma/Gompertz distribution has decaying behavior as $t \to \infty$, which is in line with the assumptions we have presented above

We used Python, Scipy's curve fitting functionality, and the general functional form of the gamma/Gompertz distribution to fit the data in **Table 3**. This implementation is shown in the Appendix. The functional form of the gamma/Gompertz pdf, as well as the fitted curve, is shown below:

$$PDF(x) = \frac{bse^{bx}\beta^s}{(\beta - 1 + e^{bx})^{s+1}} \tag{4}$$

In the above, $b, s, \beta > 0$ are constants determined by the curve-fitting process.

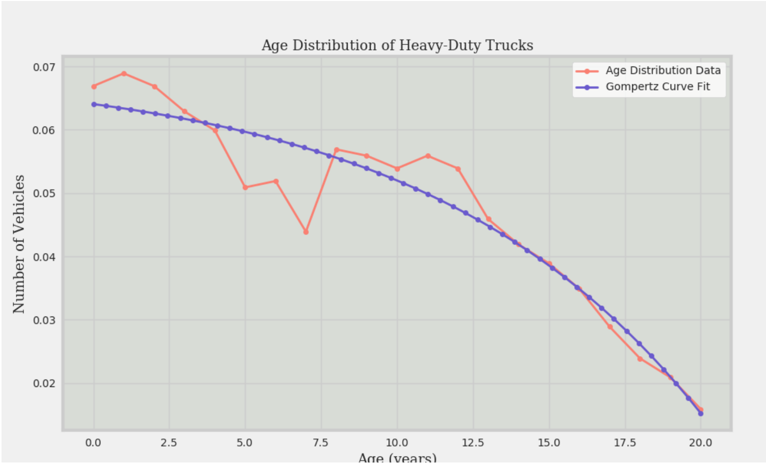

Figure 2: Age Distribution of Heavy-Duty Diesel Trucks with Gamma/Gompertz Fitting

To determine how the age distribution of these diesel trucks changes over time, we simply iterated the process of removing diesel trucks from the population of trucks each year. This way, every point that wasn't removed by the filtering process would be mapped from the point $(t, y) \to (t + 1, y)$, where y is the number of vehicles at age t. Then, we obtained the graph shown below:

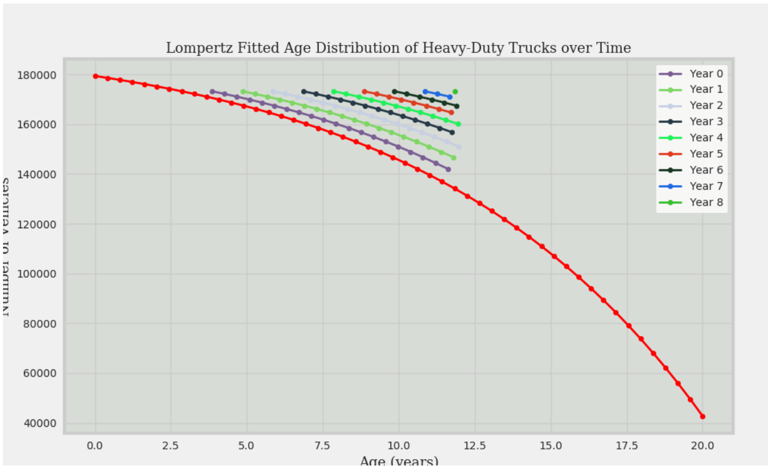

Figure 3: Gamma/Gompertz Fitting Iterated over 20 Years (t = years from 2019)

For each year, the total number of diesel cars, $D(t)$, will be the sum of the diesel cars for each age, given by our gamma/Gompertz distribution as the value of the integral below the curve. After 9 years, all diesel trucks older than 3.6 years in 2019 will now be at least 12.6 years old, so they will be have been "spent." Thus, after 9 years, there will be no diesel trucks left from the initial population, so $D(t) = 0$, and

$$E(t) = C(t) - D(t) = C(t), t \geq 9 \tag{5}$$

We now computed the percentage of total trucks that will be electric in the years 2025, 2030, and 2040 with our model. For year 2025, or $t = 6$, the linear model $C(t)$ gives the total number of trucks as $C(6) = 44669 * 6 + 2886711 = 3,154,725$ and our gamma/Gompertz distribution yields the total number of trucks replaced as electric trucks from 2019 to 2025, $R(6)$, as $1,630,047$. So the total number of diesel trucks in year 2025 is this value subtracted from the number of total trucks in 2019, or $D(6) = 2,886,711 - R(6) = 1,256,664$. Then $E(6) = C(6) - D(6) = 3,154,725 - 1,256,664 = 1,898,061$. The percentage of total trucks that are electric in 2025 will be $\frac{1{,}898{,}061}{3{,}154{,}725} \times 100 = \boxed{60.17\%}$.

For years 2030 and 2040, or $t = 11$ and $t = 21$, since $t \geq 9$, from (refer to equation above) we said that $E(t) = C(t)$ for $t \geq 9$, so the percentage of total trucks that will be electric in these years will be $\boxed{100\%}$. For $t = 11$, $E(t) = C(t) = 44669(11) + 2886711 = 3378070$ trucks. For $t = 21$, $E(t) = C(t) = 44669(21) + 2886711 = 3824760$ trucks.

2.5 Validation

Table 3: Summary of Results

Year	C(t)	E(t)	$\frac{E(t)}{C(t)}$ Percentage
2025	3,154,725	1,898,601	60.17 percent
2030	3,378,070	3,378,070	100 percent
2040	3,824,760	3,824,760	100 percent

The results of our data make sense in context. Given the analysis presented earlier, diesel truck use should eventually phase out, due to the apparent differences in operating/manufacturing costs. We expect that if all truck producers were to only make electric trucks from today, then all the trucks will be electric within 10 years, and our model gives us this outcome in 9 years. In 6 years, we expect a majority of trucks on the road to be electric. Furthermore, comparing our results to existing projections of the number of electric trucks in circulation, we found fairly close agreement. According to a study by P&S Market Research, the electric truck population is expected to reach a size of 1,508,100 trucks in 2025, which is mostly in alignment with our own predictions for 2025. The discrepancy can likely be explained by the assumption of a "seamless transition" to electric trucking by existing firms. Relaxing this condition would likely decrease the predicted value of $E(t)$.

2.6 Sensitivity Analysis

Change in **Assumption 1** is accounted for directly in $D(t)$ determination. Increasing or decreasing the lifespan will result in a change in D(t) dependent on the magnitude of the change and the exact parameters of the fitted gamma/Gompertz D(t) function, Changes in **Assumption 2** and **Assumption 3** are also accounted for in the determination of t^*, whereby a change in proportion of trucks replaced and/or time-dependent changes in purchase or operating costs would result in proportional changes in the value of t^* (and thus $D(t)$).

2.7 Strengths and Weaknesses

The strengths of this model arise from the ability to provide a closed-form solution for the number/proportion of electric trucks present at any instance in time over the next 20 years. The model predicting $D(t)$ is quite robust in its ability to produce the time-dependent distribution of diesel trucks regardless of small changes in the raw data of distributions, as the gamma/Gompertz distribution may still be fitted (conditions 2 and 3 in the justification of this model will always hold). Furthermore, the model effectively separates the raw demand for trucks in $C(t)$ and the cost-sensitivity of demand in determination of the cutoff time, t^*. Despite the strengths of the model, there are still some changes that may be applied to the model in the future. The approximation of $C(t)$ using a straight line regression may be improved by determining the exact nature of the relationship between time and the total supply/demand for trucks, such as by using a composition of various functions, numerically solving a system of differential equations.

3 Problem 2: In It For The Long Haul

3.1 Problem Restatement

Develop a model that optimizes the number of electric charging stations and chargers per station along highways to sustain current truck freight.

3.2 Local Assumptions

1. *Drivers will always charge their batteries up to exactly 80% and will never let their battery charge fall below 20%.* This range optimizes battery life, which is better within 20% and 80% charge [2].

2. *Drivers will only charge at a charging station if they cannot reach the following charging station with 20%.*

3. *The potential distance traveled from any one percent of charge is constant.*

4. *Distance traveled from 80% to 20% battery charge is 300 miles.* This is based on the fact that a full Tesla Semi charge is 500 miles [8], so the distance traveled in miles between 80% to 20% is $\frac{80-20}{100} * 500 = 300$.

5. *All stations will use Level 3 chargers.* We assume that the charging option with the most long-term economic viability, Level 3, will be chosen.

6. *It takes 30 minutes to charge a truck up to 80%.* Level 3 chargers, which have 1,600 in kW power, take 30 minutes to charge a truck to 80%.[13]

7. *Each charging station will have 9 individual chargers.* This is derived from calculating the ratio of Tesla's existing Superchargers and Supercharger stations (8.86) [14], which is approximately equal to 9. We are standardizing the number of chargers per station as a constant rather than the number of stations (with variable chargers) with the expectations that it would be difficult to arbitrarily build exits containing charging stations on existing highways so stations would have to be built at existing exits. In that case, specific information about the exit locations of highways would have to be known before developing a model, which is unrealistic.

8. *At any given point on the highway, there is the constant distribution of trucks that have a certain percentage of their battery charged.*

9. *The Annual Average Daily Truck Traffic (AADTT) at any point on the highway can be modeled as a piecewise linear function connecting every pair of consecutive highway mile marker with data.*

10. *On highway markers without AADTT data, we multiplied the AADTT by the average of known AADTT% values along the corridor.*

11. *Every corridor starts with a charging station.* Since we have no information about the route before the specific corridor interval, we built the algorithm around a charging station at the first point.

3.3 Variables

Table 4: Variables

Variable	Description	Units
R	Total number of charges per day	charges/day
I_i	Interval length, distance between consecutive mile markers with data	miles
A_i	AADTT for an interval, trucks per interval per day	trucks/day
C	Total chargers along corridor	chargers
T	Total distance traveled by all trucks in a day	miles
S	Total number of stations	stations
$f(x)$	The AADTT x miles away from the origin of the corridor	trucks/day
M	$M = \{m_i \mid i = 1, 2, \cdots, S\}$ where m_i is the number of miles from the origin of the corridor for station i, and $m_1 = 0$	

3.4 Solution

3.4.1 Determining the Total Number of Chargers

We first need to find the total number of chargers, C, for each route. By **Assumption 6**, each truck will charge for 30 minutes (0.5 hours) at a time, so each charger will be used at max $\frac{24}{0.5} = 48$ times in a 24-hour day. This value, R, is the total number of charges a day. We also have that $R = \frac{T}{300}$, as the total number of truck miles traveled per day equals the number of charges times the miles that the truck can travel per charge, which we calculated in **Assumption 4**. T is also equal to $\sum$(Interval Length) $\times$ (Trucks In Interval Per Day) $= \sum I_i \times A_i$, as we are summing over the miles the trucks traveled across all intervals that make up a corridor. Thus, we arrive at this equation:

$$C = \frac{T}{14400} = \frac{\sum I_i \times A_i}{14400} \tag{6}$$

We can obtain the Interval Length and Trucks in Interval Per Day from the M3 Corridor Data, where an Interval is a part of the highway between two consecutive highway mile markers and the Trucks in Interval Per Day refers to the AADTT values in our data. We now define a function $f(x)$ which takes x, the highway mile marker, to $f(x)$, which is the number of trucks that passes highway mile marker x. Setting $0 = x_0 < x_1 < \cdots < x_n = b$ (b is total mileage of the route) where $x_i \in M$, where M is the set of highway mile markers, the total number of truck miles traveled can be computed as the integral $\int_0^b f(x)dx$. This integral can be approximated by a midpoint Riemann sum $\sum_{i=1}^{n}(x_i - x_{i-1})f(\frac{x_i+x_{i-1}}{2})$. Then by **Assumption 3**, we assume that $f(x)$ is linear when x is in between two consecutive elements in M, so $f(\frac{x_i+x_{i-1}}{2}) = \frac{f(x_i)+f(x_{i-1})}{2}$ by linearity. This gives us the equation

$$T = \sum_{i=1}^{n}(x_i - x_{i-1})\frac{f(x_i) + f(x_{i-1})}{2} \tag{7}$$

Thus, we can compute T while only being provided values from the Corridor Data, which also allows us to compute C, or the total number of charges per route, as well. Also, note that since f is linear between two consecutive highway mile markers by **Assumption 3**, our midpoint Riemann sum form is also equivalent to the Trapezoidal Rule Sum, which is the sum we used with our interval determination Python program.

3.4.2 Spacing Out the Stations

Now we must determine how to space out the stations. Based on **Assumption 7**, there are nine chargers at every station. Therefore, by the pigeonhole principle, there are $S = \lceil C/9 \rceil$ stations needed to account for every charger. We want to evenly distribute the stations along the route by total miles traveled all trucks. Therefore, we place a station every $\frac{T}{S}$ miles, with one station at the start of the route by **Assumption 11**. In other words, we are finding m_i such that for every $(m_i, m_{i+1}), i \in [0, S-1]$,

$$\frac{T}{S} = \int_{m_i}^{m_{i+1}} f(x)dx \tag{8}$$

Combining the m_i will give us M, which is the set of locations of the charging stations.

3.4.3 Number of Stations and Chargers Along Each Route

We now give our calculated results of the total number of stations (S) and chargers (C) along each of the five routes [15]:

Table 5: Summary of Results

Route	C	S
San Antonio-New Orleans	521	58
Minneapolis-Chicago	224	25
Boston-Harrisburg	282	32
Jacksonville-Washington, D.C.	394	44
Los Angeles-San Francisco	261	29

3.5 Validation

We are able to test the success of our model by numerically implementing the algorithm described above. We start by inputting the parameters C, T as determined for each data set (given by start and end points of journey). Then, we must linearly interpolate between subsequent (distance, AADTT) values in order to guarantee the accuracy of our numerical integration scheme (we want a resolution of $dx = 0.5$ miles). Next, we integrate from (m_i, x) using a trapezoidal integral. If the value of this integral equals or exceeds $\frac{T}{S}$, the cutoff value above which a new station is needed, then the value x is appended to the set M. A flowchart of this process is demonstrated below:

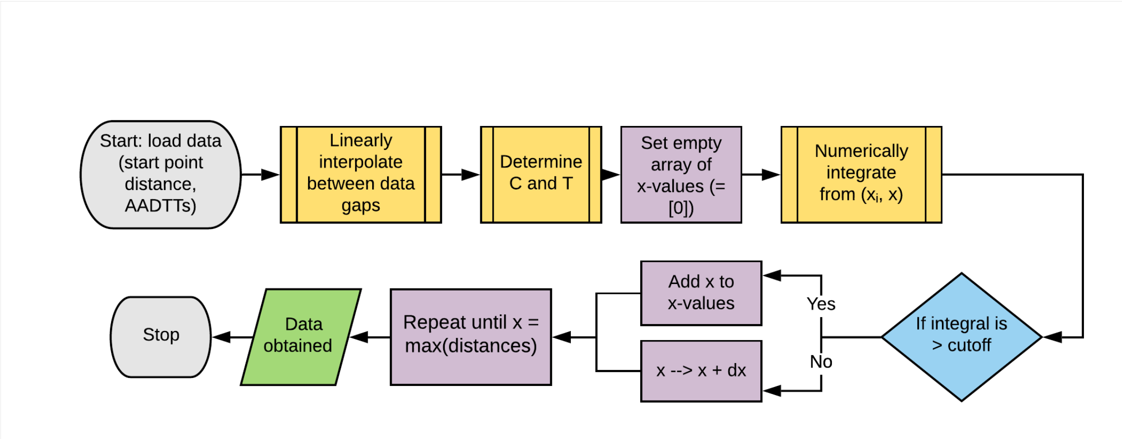

Figure 4: Flowchart Describing Numerical Implementation

Next, we examine the validity of our approach on the data sets provided, considering the journeys: 1) San Antonio-New Orleans 2) Minneapolis-Chicago 3) Boston-Harrisburg 4) Jacksonville-Washington D.C. 5) Los Angeles-San Francisco [15]. We provide both the number of stations and the locations of the stations for all the journeys in the table below. In addition, we present a graph that plots AADTT as a function of distance, with the array M plotted along the x-axis, for the Minneapolis-Chicago route.

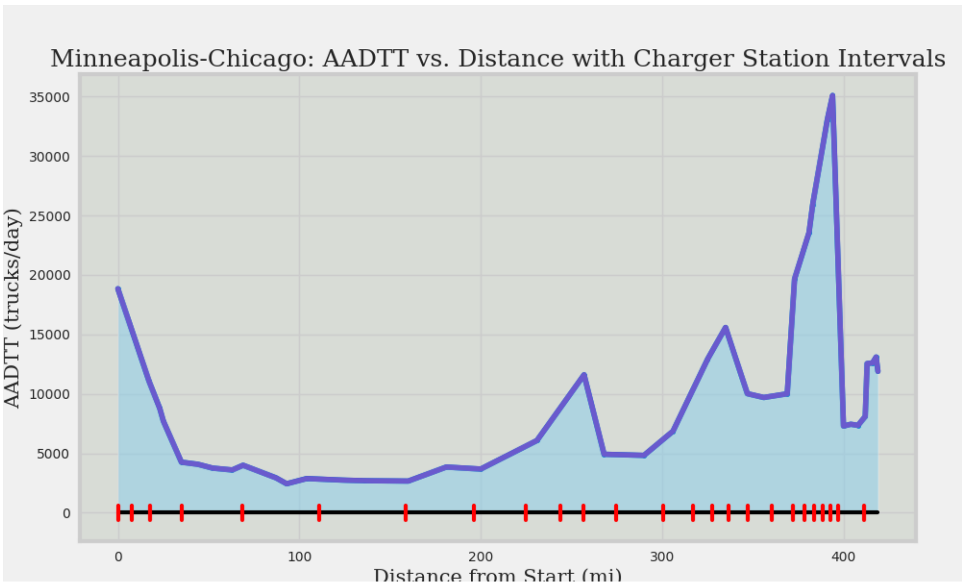

Figure 5: Interval Generation (Minneapolis-Chicago)

From the graph, we see that our model produces the desired result, as distances with higher AADTT values have charging stations placed at closer intervals. Furthermore, when we multiply the total number of stations by 9 (chargers/station), we recover approximately the value of C (the small discrepancy is because we can only have an integer number of charging stations): (25 stations)$\times$(9 chargers/station) = 225 chargers $>$ 224 chargers ✓.

Table 6: Stations Placed Along the Routes (By Mile Number) [1]

Station Number	SanAn-NOLA	Minn-Chi	Bos-Hburg	Jax-DC	SF-LA
1	0	0	0	0	0
2	18.04	7.64	17.24	16.76	8.80
3	27.48	17.56	28.36	27.96	16.72
4	38.04	35.00	37.12	39.12	25.40
...	...	...	...	...	...
24	220.64	396.88	291.00	411.32	230.72
25	228.24	411.44	301.16	430.36	243.00
26	236.28	N/A	312.32	451.24	254.68
...	...	...	...	...	...
57	519.76	N/A	N/A	N/A	N/A
58	524.04	N/A	N/A	N/A	N/A

3.6 Sensitivity Analysis

We may perform sensitivity analysis by investigating the parameters affecting the cutoff point, $\frac{T}{C}$. The first of these is the number of chargers per station. We allow this number to vary from 6 to 12. Performing this analysis on the Minneapolis data set, we find that the length of set M (the number of stations) varies as follows:

$$6 \rightarrow 38, 7 \rightarrow 32, 8 \rightarrow 28, 9 \rightarrow 25, 10 \rightarrow 23, 11 \rightarrow 21, 12 \rightarrow 19$$

We may similarly vary the parameter T, which is directly a function of the aggregate total of AADTT over all intervals are described in **Equation 6**. However, we observe that the parameter T is inversely proportional to the total number of chargers required. Thus, the product of T and $\frac{1}{C}$ is constant! This means the density of the list M is unchanged as a result of any variations in T, indicating it is completely independent of this parameter! Thus, we may easily extend our solutions between paths that have a roughly 1:1 correspondence between each other in terms of AADTT as a function of distance. That is, given the corridors $f(x)$ and $g(x)$, if there exists an isomorphism between f and g, knowing the intervals in one corridor, as well as the upper/lower bounds of the other corridor, we may easily determine the intervals for the other corridor. A very useful result indeed!

3.7 Strengths and Weaknesses

The model developed is strong due to its ability to directly account for mean traffic at any point along a path, allowing for strong correlation to be seen between traffic density and charging station requirements. Furthermore, the developed model is

[1] In this data table, N/A means that the data values are not applicable, or station M does not exist along this route because the route has less than M stations.

quite resistant to changes in parameters concerning battery health/status or changes in **Assumption 7**. That is, it can be altered trivially to account for developments in battery health/incorporating time dependence (nonconstant upper-bound of "ideal" battery capacity).

A deficiency in our model arises from our interpolation between consecutive highway mile markers using a straight line, since we do not have AADTT data between consecutive mile markers. This is particularly relevant for data points that have large distances between mile markers. However, this weakness can be simply remedied by deliberately retrieving AADTT data for the largest intervals.

4 Problem 3: I Like To Move It, Move It

4.1 Problem Restatement

Design a model that evaluates and ranks national corridors for transition to electric trucks, based on community economic, environmental, and social motivations.

4.2 Local Assumptions

1. *Corridors with higher monetary benefit per capita should be targeted for development first. For nonfinancial factors, a reasonable monetary cost can be determined that results from the factor.* By unifying every factor into a common unit, we make them easy to compare. Since this monetary metric can be converted from/into other impacts such as community excitement per person, comparing the monetary impact per person is reasonable.

2. *All necessary infrastructure will be completed within one year.* With necessary funding and labor, it takes less than one year to build an EV charging station, including the building of the circuitry and chargers and gaining approval. Thus, it is reasonable to assume that all stations will be built in one year.

3. *Eight years after all infrastructure is completed, 100% of diesel trucks will be replaced by electric trucks.* Based on our model in Problem 1, all diesel trucks will be converted to electric trucks in eight years.

4. *Factors that are dependent on a truck's lifetime will be assessed for the first 10 years of a truck's lifespan.*

5. *Trucks that go through the corridor will take one single full trip from one end to the other.* We assume this because most electric trucks are long-haul and travel journeys longer than the length of the corridor.

6. *All existing semi trucks will eventually be replaced by electric trucks.* From Section 2, we calculated that this is true after 10 years.

4.3 Variables

Table 7: Variables

Variable	Description
O	Difference in operating and manufacturing cost between diesel and electric
I	Cost of building infrastructure for electric trucks
E	Difference in cost of emissions of diesel and electric
R	Difference in cost of toxic runoff of diesel and electric
P	The number of people living within 5 miles of a corridor
T	Total monetary difference per capita between diesel and electric
U	Total monetary profit the corridor makes from trucks using pumping stations
D	Monetary profit the corridor makes from trucks using pumping per day
C	Number of chargers needed in a corridor
N	The number of unique trucks using a corridor in 2017
L	Corridor length
S	Number of stations per corridor

In this table, the variables C, E, R, F are the diesel costs minus the electric costs, so for the benefits of using electric trucks to exceed that of using diesel trucks, we want these variables to be positive. J is also positive as having more jobs in the electric infrastructure benefits the corridor. U is also positive as it represents a profit towards the corridor. I is negative as it represents an initial cost.

4.4 Solution

Our overall metric is T, which is calculated by summing together the monetary impact of various factors before dividing by the number of people living near the corridor:

$$T = \frac{O + I + E + U}{P} \tag{9}$$

We will now describe how the monetary impact for each factor was calculated.

1. **O** - The savings per electric truck over 10 years is the the manufacturing cost Man_d and operating cost Ope_d of diesel trucks minus the manufacturing cost Man_e and operating cost Ope_e of electric trucks. By **Assumption 8**, we can assume that all existing trucks will be replaced with electric trucks. Therefore,

$$[19][20] O = ((117430 + 1434500) - (180000 + 1197000))N \tag{10}$$

$$O = 174930 * N \tag{11}$$

N here is the number of unique cars that pass through the corridor over a 10 year lifetime. We assume simply that N is 3 times the average AADTT for a region (this is reasonable because all cars passing through corridor will not be unique everyday, but the number will certainly be a few factors greater than AADTT).

2. **I**. This is the fixed cost associated with installation of Electric Vehicle Storage Equipment (EVSE) at a location. Of course, based on the work in §2, we know that there are C chargers along a corridor. We may express the cost, I, by the following equation:

$$I = -(Lifetime\_Cost + Installation\_Cost) \times C \tag{12}$$

Using the information in the ASDA Energy Summary [22], we find $Lifetime\_Cost = -4,000$ and $Installation\_Cost = -12,000$, so that

$$I = -16,000 \times C \tag{13}$$

3. **E** - This is the difference in monetary impact resulting from the emissions avoided through the use of electric trucks. Emissions from diesel trucks account for 29% of all transportation sector emissions, which account for 20% of all US emissions [16]. Calculating from 5.4 billion metric tons of carbon dioxide (CO2) equivalent greenhouse gas (GHG) in the US annually, we get 5.4 billion $\times 0.29 \times 0.2 = 313$ million metric tons of CO2 equivalent resulting from diesel trucks [17]. According to a Stanford study, the cost of a single ton of CO2 equivalent emissions is around \$220 [16] [18]. This gives a approximate social cost of 68 billion from GHG emissions resulting from diesel trucks [21]. Since the number of trucks in the US is approximately 3.68 million, this yields a cost of \$187120 per truck per year. Finally, we multiply by $N = 3$ times AATTD (as before).

4. **U** - This is the total communal profit that the corridor makes from trucks using its charging stations through 10 years. We can model this variable as a unit of time t, where t is the number of days since the beginning of the first day of truck usage. It suffices to write $U = D \times 3650$, where D is the profit a corridor makes from trucks using its charging stations every day. D is equal to the product of the average number of charges a truck makes per day, the average number of trucks going through the corridor, and the cost per charge. The average number of charges a truck makes per day is $\frac{L}{300}$, as a truck will go through the entire length of the corridor once by **Assumption 7**. The average number of trucks going through the corridor is $\frac{\text{avg}(AADDT)}{S}$, where S is the number of stations in the corridor. The cost per charge for a 1600 kW truck, which is the most efficient truck, is $1600\text{kW} \times 0.5\text{hr} \times \frac{\$0.05}{\text{kWh}} = \frac{\$4.00}{kWh}$. Thus, multiplying these three things together, we get the equation

$$D = \frac{L \times \text{avg}(AADDT)}{75S} \tag{14}$$

and from this,

$$U = \frac{146L \times \text{avg}(AADDT)}{3S} \tag{15}$$

5. **P** - This is the total population around the interstate in a five mile radius, summed at 2 mile increments. Data was taken from the US census. However, not all data [25] was available from the census; the remaining routes' populations were found by adding populations of the communities the corridor passes through.

4.5 Ranking of Corridors

We plugged in values for each corridor to obtain the following values.

Corridor	O	I	E	U	P	T
SanAn-NOLA	6296400000	-8336000	6732000000	5276138	9639686	1351.22
Minn-Chi	5249807145	-3584000	5613001350	8005080	8487254	1280.42
Bos-HBurg	5711897731	-4512000	6107060467	5165419	12747688	927.20
Jax-DC	4749065162	-6304000	5077616840	4514946	5244684	1873.31
LA-SF	7332613461	-4176000	7839901184	5980280	13245275	1145.64

Therefore, the rank of the corridors in decreasing development priority is

1. Jacksonville-Washington, D.C.
2. San Antonio-New Orleans
3. Minneapolis-Chicago
4. Los Angeles-San Francisco
5. Boston-Harrisburg

4.6 Validation

The results of our data make sense in context. The T metric, which is around a thousand dollars per person over 10 years, seems right. The Jacksonville-Washington, D.C. corridor has a much larger T value than the other four corridors, because the corridor has many trucks traveling along the highways, due to its very long length, but not a lot of people living near this route. The Boston-Harrisburg corridor produces a much lower T value because there are fewer trucks on the corridor compared to the high population around the highway.

4.7 Sensitivity Analysis

Our model is quite robust to changes in each parameter because we have effectively uncoupled the effects of each parameter, making their impact entirely independent of the other parameters. Also, we notice that the metric, T is directly related to each of the quantities O, I, E, U and inversely proportion to the quantity P. Thus, the change in parameter ΔT will be given by

$$\frac{\Delta O + \Delta I + \Delta E + \Delta U}{\Delta P} \tag{16}$$

The changes, $\Delta O, \Delta I, \Delta E, \Delta P$ can be determined simply using the explicit forms of the formulas provided above, altering each condition as necessary to produce a corresponding change in $\Delta O, \Delta I, \Delta E, \Delta P$.

4.8 Strengths and Weaknesses

Our model takes into account diverse environmental and economic factors ranging from direct costs of producing diesel and electric trucks to indirect impacts on gas emissions and jobs. In addition, by associating the importance of each factor with a monetary value, we provide a clear way of quantifying and comparing each factor. This is advantageous to using different units of comparison such as utility per capita, since the conversion to other units can be subjective and less accurate.

One weakness in our model is that we were unable to reliably calculate the population P. This is because census data on population density around highways was incomplete, only including a few interstates. Our other method was inaccurate as people who did not live in larger communities were not counted and communities are not always fully located within 5 miles of the corridor.

5 Further Studies

5.1 Model 1

A better approximation of $C(t)$ could be obtained by determining the exact nature of the relationship between time and the total supply/demand such as by determining effects of various parameters and appropriate weights, then numerically integrating the resulting differential equations. Of course, knowing that this problem assumed all electric infrastructure in place, a more realistic approach could be determined incorporating the construction of said infrastructure.

5.2 Model 2

Having AAADT values for the large stretches of highway that currently represent gaps in the data would lead to a more accurate general model. Additionally, plotting out specific existing gas stations and surrounding services (truck stops, convenience stores) would allow a more practical building plan to incorporate existing markets.

5.3 Model 3

More precisely stratified data in each section would be more telling for this section. We were unable to determine, for example, the exact monetary impact of runoff caused by incompletely combusted diesel dripping from exhaust pipes or human error in diesel administration. Specific environmental monetary costs from long-term fuel production costs, from hydraulic fracturing or offshore oil drilling for example, would be valuable.

References

[1] U.S. Department of Transportation. *Interstate System Dwight D. Eisenhower National System of Interstate and Defense Highways*. Federal Highway Administration. URL.

[2] *Keep on Trucking Information Sheet*. MathWorks Math Modeling Challenge 2020. URL.

[3] *Truck Production Data*. MathWorks Math Modeling Challenge 2020. URL.

[4] NASA Science. *How Far Away Is the Moon?*. September 30, 2019. URL.

[5] Alan Hooper and Dan Murray. *An Analysis of the Operational Costs of Trucking: 2017 Update*. American Transportation Research Institute, 2017. URL.

[6] *Tesla Semi*. URL.

[7] Daniel Oberhaus. *Tesla May Soon Have a Battery That Can Last a Million Miles*. Wired, September 23, 2019. URL.

[8] Peter Holley. *Tesla's latest creation: An electric big rig that can travel 500 miles on a single charge*. Washington Post, November 17, 2017. URL.

[9] U.S. Census Bureau. *U.S. Population*. March 1, 2020. URL.

[10] Wolf Richter. *Average Age of Cars and Trucks by Household Income and Vehicle Type over Time*. Wolf Street, August 21, 2018. URL.

[11] Bureau of Transportation Statistics. *Truck Profile*. October 28, 2019. URL.

[12] German Rodriguez. *Parametric survival models*. Revised Spring 2010, Princeton. URL.

[13] *Battery Data*. MathWorks Math Modeling Challenge 2020. URL.

[14] *Superchargers: Charge on the Road*. Tesla 2020. URL.

[15] *Corridor Data*. MathWorks Math Modeling Challenge 2020. URL.

[16] Center for Energy and Climate Solutions. *Federal Vehicle Standards*. URL.

[17] American Trucking Associations. *Reports, Trends and Statistics*. URL.

[18] David Quiros et al. *Greenhouse gas emissions from heavy-duty natural gas, hybrid, and conventional diesel on-road trucks during freight transport*, August 6, 2017. URL.

[19] Liam Tung. *Tesla's all-electric Semi truck: Prices start at $150,000 and you can reserve one today*, ZDNet, November 24, 2017. URL.

[20] I Wagner. *Average price of new Class 8 trucks in the United States from 2010 to 2018*, ZDNet, September 13, 2019. URL.

[21] Ker Than. *Estimated social cost of climate change not accurate, Stanford scientists say*, January 12, 2015. URL.

[22] *Review of Asda Energy Prices for Electricity and Gas.* URL.

[23] Contributors of Encyclopedia Britannica, *Fermat's Principle*. Encyclopedia Britannica, Encyclopedia Britannica Inc., October 28, 2016. URL.

[24] Kendall A Atkinson. *An Introduction to Numerical Analysis*. John Wiley and Sons, New York, 2nd Edition, 1989.

[25] U.S. Census Bureau *Interstate Population Density Profile* URL.

Appendix

Problem 1: Iterative Algorithm (Code)

```
import numpy as np                    # collection of math functions
import pandas as pd                   # to process csvs

import matplotlib.pyplot as plt          #plotting library
from scipy.optimize.minpack import curve_fit  #curve fitting
method
from Bio.phenotype.pm_fitting import gompertz, fit    #provides
general functional form for Gamma/Gompertz


# styles for plotting
import seaborn as sns
import matplotlib
import matplotlib.style as style

matplotlib.rcParams.update(matplotlib.rcParamsDefault)
matplotlib.rcParams['font.family'] = "serif"
style.use('seaborn-notebook')
style.use('fivethirtyeight')
sns.set_style("whitegrid")


#store data in csv as DataFrame
data = pd.read_csv('Age_Distribution.csv')
total = 2800000      #total number of diesel trucks
cutoff1 = 3.6        #lower bound on trucks to be replaced
cutoff2 = 12         #lifespand of diesel truck


#separate into x (times), y (relative frequencies)
x = data['Year'].values
y = (data['Truck Frequency'].values)/100

params, cov = curve_fit(gompertz, x, y, maxfev = 5000000)   #
store parameters of function and covariance of model fit using
curve_fit

fig, ax = plt.subplots(1,1, figsize = (10,6))        #initialize
plots
```

```
ax.plot(x, y, '.-', color = 'salmon', label='Age Distribution Data', linewidth = 2, markersize = 10)  #plot raw data
x1 = np.linspace(x.min(), x.max())
ax.plot(x1, gompertz(x1, *params), '.-b', color = 'slateblue', label='Gompertz Curve Fit', linewidth = 2, markersize = 10)  #plot fitted Gamma/Gompertz
ax.set_xlabel('Age (years)', fontname = 'serif', fontsize = 13)
ax.set_ylabel('Number of Vehicles', fontname = 'serif',fontsize = 13)
ax.set_title('Age Distribution of Heavy-Duty Trucks', fontname = 'serif', fontsize = 13)
ax.set_facecolor('xkcd:light grey')
ax.legend(loc='best', frameon=True)
fig.savefig('Age Distribution.png')    #save figure


time = 20  #number of years to iterate over
i = 0
x_0 = np.linspace(x.min(), x.max())  #initial conditions on x, y
y_0 = gompertz(x1, *params)
x_i = x_0
y_i = y_0
plt.clf()
fig, ax = plt.subplots(1,1, figsize = (10,6))
ax.plot(x_0, y_0*total, '.r-', linewidth = 2, markersize = 10)
ax.set_xlabel('Age (years)', fontname = 'serif', fontsize = 13)
ax.set_ylabel('Number of Vehicles', fontname = 'serif',fontsize = 13)

ax.set_title('Lompertz Fitted Age Distribution of Heavy-Duty Trucks over Time', fontname = 'serif', fontsize = 13)
#ax.fill_between(x_0,y_0, facecolor='red', alpha=0.5)
integral = []

while i < 20:
    x_i = x_i + 1   #step i-th year by 1
    bool_keep = np.all((x_i > cutoff1, x_i < cutoff2), axis=0)  #bool ages not removed
    bool_removed1= x_i < cutoff1                                  #bool ages removed (less than cutoff1)
    bool_removed2= x_i > cutoff2                      #bool ages removed (greater than cutoff2)

    removed_x_i1 = x_i[bool_removed1]            #ages removed (less than cutoff 1)
    removed_y_i1 = y_i[bool_removed1]            #population removed (less than cutoff 1)
    integral1 = np.trapz(removed_y_i1, removed_x_i1)              #integrate --> total number of trucks replaced (less than cutoff 1)

    removed_x_i2 = x_i[bool_removed2]            #ages removed (greater than cutoff 1)
    removed_y_i2 = y_i[bool_removed2]            #population removed (greater than cutoff 1)
    integral2 = np.trapz(removed_y_i2, removed_x_i2)      #integrate --> total number of trucks replaced (greater than cutoff 1)
```

```
        integral.append(integral1 + integral2)   #number of trucks
replaced/year


        x_i = x_i[bool_keep]         #times kept
        y_i = y_i[bool_keep]         #cars removed
        cc = np.random.rand(3)         #color randomly
        if len(x_i) > 0:
            ax.plot(x_i, y_i*total, '.-', color = cc, linewidth = 2,
markersize = 10, label = 'Year ' + str(i)) #plot distribution
        i += 1
    ax.legend(loc='best', frameon=True)
    ax.set_facecolor('xkcd:light grey')
    fig.savefig('Iterate over Time.png')  #save figure

    n = 5                      # number of years
    first_n = np.sum(integral[0:n])*total       # add elts. of, times
 total = value of E(n) (solely from replacement)
```

Problem 2: Interval-Determining Algorithm (Code)

```
    import numpy as np
    import pandas as pd
    import matplotlib.pyplot as plt

    import seaborn as sns
    import matplotlib
    import matplotlib.style as style

    matplotlib.rcParams.update(matplotlib.rcParamsDefault)
    matplotlib.rcParams['font.family'] = "serif"
    style.use('seaborn-notebook')
    style.use('fivethirtyeight')
    sns.set_style("whitegrid")


    total_chargers = 224              #parameter 'C'
    total_miles = 3220760             #parameter 'T'
    cutoff = (6/total_chargers)*total_miles    #cut off value of
integral

    #read data of (distance, AADTT) for Minneapolis

    data = pd.read_csv('MINNEAPOLIS.csv')
    distances = data['approximate number of miles from Minneapolis'].
values
    AADTTS = data['AADTT'].values

    #populate distances between consecutive values, interpolate
linearly to predict AADTT
    distances_populated = np.linspace(np.min(distances), np.max(
distances), 10000)
    AADTT_populated = np.interp(distances_populated, distances,
AADTTS)

    #initialize plots
```

```python

    fig, ax = plt.subplots(1,1, figsize = (10, 6))
    ax.plot(distances, AADTTS,'.-')
    ax.plot(distances_populated, AADTT_populated, color = 'slateblue')
    ax.fill_between(distances_populated,AADTT_populated, facecolor='skyblue', alpha=0.5)

    #initialize list of stations, establish step size, initialize running parameter 'x'
    x_values = [0]
    dx = 0.5
    x = 0

    #before reaching end point (max_distance)
    while x <= np.max(distances_populated):
        x_values_numpy = np.asarray(x_values)      #convert to numpy array
        bool_x = x_values_numpy < x                #boolean array to determine lower bound, x_i
        x_values_numpy = x_values_numpy[bool_x]    #all values of M less than x
        left_bound = x_values[-1]                  #largest value = left_bound
        #print(left_bound)

        #determine region to perform integral over
        bool_keep = np.all((distances_populated > left_bound, distances_populated < x), axis=0)

        AADTTS_keep = AADTT_populated[bool_keep]
        distances_keep = distances_populated[bool_keep]

        #trapezoidal integration with distances, AADTTS
        integral = np.trapz(AADTTS_keep, distances_keep)

        #if value of integral > cutoff, add x to M
        if integral>cutoff:
            x_values.append(x)

        #increment x
        x +=dx

    #slightly unwieldy method of plotting the intervals!

    power = 600*np.ones(len(x_values))

    points = np.column_stack((x_values, power))
    for pt in points:
        ax.plot( [pt[0],pt[0]], [-600,pt[1]], linewidth = 3, zorder=5, color ='red')

    #make plots look nice!
    ax.set_facecolor('xkcd:light grey')
    ax.plot([0, np.max(distances)], [0,0], linewidth = 3, color = 'k')
    ax.set_xlabel('Distance from Start (mi)', fontname = 'serif', fontsize = 15)
```

```
    ax.set_ylabel('AADTT (trucks/day)', fontname = 'serif', fontsize
    = 15)
    ax.set_title('Minneapolis-Chicago: AADTT vs. Distance with
    Charger Station Intervals', fontname = 'serif', fontsize = 18)

    #save figure
    fig.savefig('Minneapolis.png')
```