# Ride Like the Wind Without Getting Winded: *The Growth of E-bike Use*

Team #16432

March 4, 2023

## Executive Summary

With the drastic fall of public transportation during the COVID-19 pandemic, electric bikes have zoomed to the top of the industry, making up 23%[8] of all bike sales in the United Kingdom and even exceeding 70%[9] of all bike-share rides in some parts of America.

When analyzing the rapid rise of the e-bike industry, it is important study not only its past, but its future as well. By doing so, we can make important insights regarding further changes that may come to the transportation industry. To predict future e-bike sales in the United States and the United Kingdom, our mathematical model applies a logistic fit to previous sales data[2], while also considering a saturation point at which the sales rate will level out. To create a model for the United Kingdom, which had no previous sales data, we took the average of our logistic models for the United States and France. By modeling yearly e-bike sales for the US and UK, we projected sales to be 1,690,527.3 and 499,913.9 e-bikes, respectively, in 2025, and 2,466,259.3 and 588,855.7 e-bikes, respectively, in 2028.

This increase in electric bike sales can be attributed to various factors, some of which are much more important than others: environmental awareness, gas costs, electric charging costs/value, and disposable income amounts all may contribute to the observed trends in electric bike sales[2]. Since we were looking strictly at past data, we used exponential regressions to model the sales data, as they had a better fit than the logistic models used in part 1. To narrow our scope, we plotted each factor's data from the provided common data set and distinguished the factors that fit the same exponential model used for the electric bike sales[2]. After finding the 4 variables that best fit the exponential curve for both the US and the UK, we found the linear combination of those variables which most closely approximates the electric bike sales curve using gradient descent optimization. The coefficients for each function in the linear combination represented their weight of importance in modeling the electric bike sales curve. We found that gas prices and the increase in the cost-value benefits of charging and using electric bikes were the main reasons for the increase in electric bike sales from 2008 to 2022 in the UK and the US.

Finally, the advent of e-bikes may result in reduced usage of other modes of transportation. This change in the makeup of the transportation sector may have effects on carbon emissions, traffic congestion, and health and wellness. To determine the impact of e-bikes, we examined the trends in the makeup of the transportation industry, using data from [2]. Unfortunately, we found that there was no statistically significant impact on these factors, since the change from e-bikes was small, and mostly affected mass transportation, an already sustainable and low-impact industry.

# Contents

# 1 The Road Ahead

## 1.1 Defining the Problem

In this problem, we are tasked with creating a model to predict growth in e-bike sales, given historical sales data in the United States, Europe, France, China, India, and Japan[2]. We are then asked to apply this model to predict e-bike sales in the United States and the United Kingdom in 2025 and 2028.

## 1.2 Assumptions and Justification

1. **Electric bikes do not include electric mopeds and sit-down scooters.**
   Justification: While there is some conflict between dataset definitions, the datasets for the United States and the European Union do not include electric mopeds and sit-down scooters[2]. Thus, since the results of our study focus on the United States and the United Kingdom, we will use the definition provided by the datasets that describe them.

2. **There will not be any significant technological advancements that drastically change the affordability or efficiency of e-bikes in the next five years.**
   Justification: While technological advancements can revolutionize an industry and bring it to the mainstream, these advancements take years to develop and are very difficult to predict.

3. **There will not be any significant legal changes that drastically change the ability of the e-bike industry's R&D and sales efforts.**
   Justification: While legal restrictions on industries are not unprecedented, they are usually aimed at harmful sectors that do not align with the government's viewpoint. The United States and the United Kingdom have shown their support[2] of the environmentalist movement, and have let electric cars develop without severe restrictions (seen in the data from [3]), which will likely be reflected in their e-bike policy. Furthermore, these legal restrictions are hard to predict[8].

4. **The number of e-bikes sold before the year 2000 is so low that it can be considered negligible.**
   Justification: e-bikes are a relatively new phenomenon. While they were around before the year 2000, they only hit the mainstream afterward, as seen in the exponential growth of sales in the EU from 2006 in [2]. Furthermore, the provided data only goes back to 2006, so the sales data prior to 2000 is out of the scope of this study.

5. **Only a certain proportion of the population will ever be open to buying e-bikes, and this proportion will not change drastically.**
   Justification: Despite the growing environmentalist movement, western society will still be largely composed of conservatives and moderates, who are unlikely to sacrifice their cars for an e-bike for the sake of the environment[8]. Given our short time frame (5 years), it is acceptable to assume that there will not be a large-scale shift in public opinion that leads to a larger proportion of the population being open to buying e-bikes.

### 1.3 Variables

**Table 1**

| **Symbol** | **Definition** | **Unit / value**[2] |
|---|---|---|
| $b_l(t)$ | The total number of e-bikes (in thousands) sold per 100,000,000 people in a given location $l$ (as a function of year $t$) | Thousands of e-bikes |
| $P_l(t)$ | The total population of a given location $l$ in hundreds of millions (as a function of year $t$) | Hundreds of millions of people |
| S | The average lifespan of an e-bike | 4[6] |

*the subscripts $u, e, f, b$ are used for the United States, the European Union, France, and the United Kingdom.*

### 1.4 The Model

When creating regression models for the sales data[2], the most obvious approach was an exponential regression. However, this model didn't make much sense, as the regressions quickly outpaced the populations of these regions. Looking at this, we realized that there must be a limiting factor for bike sales. This made sense, given assumption 5, since we also know that the average lifespan of an e-bike is 4 years[6]. Therefore, we realized that the sales would, at some point, even out, as the proportion of people open to buying an e-bike would buy them. Thus, at this level point, the entirety of e-bike sales would go to people replacing their old e-bike after four years, making a logistic model work better. This is reinforced by the sales data from France, as we can see a downturn in the rate of sales increase from 2016 to 2019. Thus, we applied logistic regression to the data from France. In this regression, we defined the year 2000 as $t = 0$ and set the y-intercept to 0 (see assumption 4). This gave us the following sales equation (adjusted to reflect values per 100 million people):

$$\begin{aligned} b_f(t) &= \frac{-1016.29}{1 + e^{0.4732(x-18.2814)}} + 1016.29 \\ r^2 &= 0.9759 \end{aligned}$$

From this, we find that the total sales will not exceed 1,016,290 e-bikes (model is in terms of thousands of e-bikes) per 100 million people per year in France. Given the 4-year lifespan of an e-bike[6], this means that the maximum proportion of people who own e-bikes is around 1 in 25 people (4,065,000 per 100 million).

We believe that this capacity is transferable to both the US and the UK. This is because, despite the difference in current sales rates, since these societies share many other characteristics, we believe that the total proportion of people open to ever purchasing e-bikes follows the same logistic model in all three locations.

Using this carrying capacity in conjunction with the sales data from the United States, and again setting the year 2000 as $t = 0$ and the y-intercept to 0 (see assumption 4), we find a logistic regression for e-bike sales (in thousands) per 100 million people in the United States:

$$b_u(t) = \frac{-1016.29}{1 + e^{.3319(x-24.9856)}} + 1016.29$$
$$r^2 = .9095$$

For the United Kingdom, since we do not have past sales data, we must approximate the sales. To do so, we averaged the results from the United States and France. We believe that this is a valid approach since the United Kingdom shares many characteristics with these two countries. Despite being a former part of the European Union and geographically near France, the United Kingdom follows the United States in many of the measured economic trends in datasheet 2[2] (see problem 2 for justification). Thus, the sales of e-bikes in the United Kingdom can be represented by an average of the sales in the United States and France. By averaging the two regression models, we obtain the following equation:

$$b_b(t) = \frac{-508.145}{1 + e^{0.331942(x-24.9856)}} + \frac{-508.145}{1 + e^{0.473195(x-18.2814)}} + 1016.29$$

## 1.5 Results

Using our models, the table below shows the predicted sales of e-bikes (in thousands) in the United States and the United Kingdom in 2025 and 2028, first per 100 million people, and then in total:

**Table 2**

| | 2025 (per 100 m) | 2025 (total) | 2028 (per 100 m) | 2028 (total) |
|---|---|---|---|---|
| United States | 509.358 | 1,690.5273 | 743.087 | 2,466.2593 |
| United Kingdom | 742.521 | 499.9139 | 874.626 | 588.8557 |

*Projected sales for years 2025 and 2028 in the US and UK*

## 1.6 Model Revision

Initially, we used an exponential growth function to model the sales data. As explained above, we revised it to a logistic model to more accurately reflect the data and account for a carrying capacity. However, we still use an exponential growth model in problem 2, since it uses past data for which the exponential model fits better.

## 1.7 Discussion

Our model predicts that the sales rates for e-bikes in the United States and the United Kingdom will increase. For the United States, the rate of increase remained very high, which is to be expected because the United States is currently lower on their sales per capita than France, the European Union, and our prediction for the United Kingdom. In fact, looking at our model for the US, we can see that the rate of sales increase will start to go down around 2025 ($t = 25$). For the UK, however, the rate of sales increase has already started decreasing around 2021, which is reflected in the predicted lower rate of sales increase from 2025 to 2028 (per 100m people).

Strengths:

- This model takes into account an eventual slowing of the rate of sales increase as societies become saturated with e-bikes.
- This model shows a clear difference between the United States' timeline of adopting e-bikes and that of France and the United Kingdom.

Weaknesses:

- This model relies heavily on extrapolations of a small data set.
- This model makes an assumption about how similar the United Kingdom is to France and the United States, which may be inaccurate.
- This model does not account for population growth of the locations used, and instead uses the population from 2021, as found in[1].
- This model fails to account for major technological advancements in the field of e-bikes, instead assuming the pace of advancement will remain constant.

## 1.8 Sensitivity Analysis

Much of this model relies heavily on the five data points of e-bike sales in the United States provided. The model for the US directly extrapolates this data, and the model for the UK is half-based on this data. Furthermore, the carrying capacity of sales was extrapolated from the logistic model of France's data, which makes the US model somewhat dependent on the sales data from France. If we were to change France's sales in 2019 from 388 thousand to 350 thousand, we would get a new limit of 774 thousand e-bikes per 100 million people, which is much lower than the value we used of 1016 thousand e-bikes per 100 million people. This would drastically change our predictions, but we are still confident in our findings given the amount of sales data from France.

## 1.9 Technical Computing Tools

We imported the data into MATLAB and used the 'fit' function to create the exponential model. When making the logistic regression, we instead used Desmos due to its easier interface.

# 2 Shifting Gears

## 2.1 Defining the Problem

In this problem, we are tasked with determining the significance of certain factors in the growth of e-bikes. We are provided historical data on e-bike sales, as well as different economic and social factors, such as gas prices, and public sentiment. We will evaluate the significance of a factor by determining how much impact its trend has on the sales model.

## 2.2 Assumptions and Justification

1. **Electric bike sales increase is independent of population growth**
   Justification: The populations of the United Kingdom and the United States are not increasing at a great enough rate to be a significant factor in the increase in sales of electric bikes compared to other factors such as those used in our model[1].
2. **Multiple factors can affect e-bike sales with different magnitudes**
   Justification: Given the complexity of our society, it is unlikely that only one factor will impact e-bike sales, and much more likely that an amalgamation will do so.
3. **For a factor to impact e-bike sales, it must have a high correlation with the e-bike sales on a year by year basis**
   Justification: While it is theoretically possible for the randomness of a factor to influence customer decisions (for example, if the price of pork is constantly fluctuating, shoppers may buy steak instead, because they want a price they can expect), it is unlikely for this impact to be that large, and thus it will likely not be a significant factor.

## 2.3 Variables

**Table 3**

| Symbol | Definition | Unit |
|---|---|---|
| $b_l(t)$ | The total number of e-bikes (in thousands) sold per 100,000,000 people in a given location $l$ (as a function of year $t$) | Thousands of e-bikes |

*the subscripts $g, b, e$ are used for gas, batteries, and environment*

## 2.4 The Model

We examined the given data and created regression models for the number of bike sales in the United States, France, and the United Kingdom. We used exponential regression as it fit the data best. We did not need to use a logistic regression, as we are not using the model to extrapolate predicted sales, and rather just looking at past data. We used the year 2000 as $t = 0$ to further simplify calculations (see problem 1, assumption 4). We found the functions in MATLAB for the raw data and then divided them by the current population in hundreds of millions of each location[1] to find the number of sales per 100,000,000 people. For the United Kingdom, since there was no sales data provided, we averaged the United States and France models, for the same reasons as problem 1. Below are the functions:

**Table 4**

| **Location** | **function** | **r-squared value** |
|---|---|---|
| Model | $b_l(t) = A \cdot e^{Bt}$ | - |
| United States | $b_u(t) = .686 \cdot e^{.2729t}$ | .9182 |
| France | $b_f(t) = 1.8538 \cdot e^{.3055t}$ | .9663 |
| United Kingdom | $b_b(t) = .343 \cdot e^{.2729t} + .9269 \cdot e^{.3055t}$ | - |

*Exponential regression models for US, France, and UK*

Because of the exponential nature of the model we developed for electric bike sales growth, we determined the relevance of factors contributing to these sales based on how well they fit the exponential model. Given the data sets provided for various factors, we involved those with a value of $r^2 > 0.8$ in our model.

**Table 5**

| **Variable name** | **Function** | $r^2$ **value** |
|---|---|---|
| Gas Price in UK Over Time | $g_b(t) = 73.33 \cdot e^{0.0323t}$ | .8330 |
| Gas Price in US Over Time | $g_u(t) = 1.675 \cdot e^{0.03318t}$ | .5889 |
| Cost (US$/kW-hr) | $c(t) = 1516 \cdot e^{-0.1526t}$ | .9899 |
| Gravimetric Energy Densities (W-hr/kg) | $d(t) = 145.9 \cdot e^{-0.04278t}$ | .9499 |
| United Kingdom's population's value of environmental issues in top 3 | $p(t) = 0.8614 \cdot e^{0.1631t}$ | .8675 |
| United States citizens' average disposable income over time | $i(t) = 28780 \cdot e^{0.01518t}$ | .9419 |

*Exponential regressions for various relations with time*

We involved 4 different variables/functions in the UK and the US models. Other variables we tested include disposable income in the UK and Interest levels in environmental issues in the US, both of which had $r^2 < 0.8$ when fit to the exponential model used in part 1. However, we did involve the US gas prices in our US model due to gas prices' importance in the UK model. Additionally, the US gas prices data did follow an exponential trend until 2013 (0.913 $r^2$ value), until it became increasingly sporadic. For these reasons, we made an exception for this specific variable. Another important note is that the use of different units for each function makes for completely different scales for each function. However, all of these functions were standardized to match one scale.

In order to assess the degree to which a factor affected e-bike sales, we decomposed the total sales functions into a linear combination of the above equations à la Fourier Analysis[5]. In other words, we found a vector such that

$$b_l(t) \approx \begin{bmatrix} f_1(t) & f_2(t) & \cdots & f_n(t) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

where each $f_n(t)$ parameterizes a factor affecting e-bike sales.

The first problem we ran into was magnitude: some equations had a disproportionate numerical value compared to others. To solve this, we normalized each function by defining

$$f^1(t) = \frac{f(t)}{f(0)}.$$

As all of our functions are exponential, this ensures $f^1(0) = 1$.

Now that we have a set of normalized functions, we can decompose our sales function and measure how much of an impact each has. For the UK, we used MATLAB to generate a set of 100 data points $(t, b_b(t))$ on $[8, 22]$ (because the function for bike sales in the UK is based on data for the US and France, for which we have data corresponding only to 2008-2022). We then fit

$$u \cdot g_b^1(t) + v \cdot c^1(t) + s \cdot d^1(t) + t \cdot p^1(t)$$

to the data points[4].

For the US, our $r^2$ value for the population's value of environmental issues was sufficiently low ($< 0.5$) that we decided to omit it as a factor. Thus, we factored in average disposable income (which had a poor fit for the UK). We maintained the set of 100 data points on $[8, 22]$ from the UK:

$$u \cdot g_u^1(t) + v \cdot c^1(t) + s \cdot d^1(t) + t \cdot i^1(t).$$

## 2.5 Results

Using MATLAB to fit the data for the UK, we got

$$\begin{bmatrix} u \\ v \\ s \\ t \end{bmatrix} = \begin{bmatrix} 4372 \\ -815.3 \\ -3998 \\ 57.51 \end{bmatrix}$$

meaning

$$B_b(t) := 4372 \cdot g_b^1(t) - 815.3 \cdot c^1(t) - 3998 \cdot d^1(t) + 57.51 p^1(t) \approx b_b(t).$$

And indeed we compute

$$I(B_b, b_b) = \int_8^{22} \left[B_b(t) - b_b(t)\right]^2 \, dx = 169.0710$$

which measures the difference between the two functions over the interval[7]. Changing any of these values increases the value of $I(B_b, b_b)$.

Doing the same for the US, we got

$$\begin{bmatrix} u \\ v \\ s \\ t \end{bmatrix} = \begin{bmatrix} -35231 \\ -1509 \\ 19695 \\ 16522 \end{bmatrix}$$

meaning

$$B_u(t) := -35231 \cdot g_u^1(t) - 1509 \cdot c^1(t) + 19695 \cdot d^1(t) + 16522 \cdot i^1(t) \approx b_u(t).$$

And once again,

$$I(B_u, b_u) = \int_8^{22} \left[B_u(t) - b_u(t)\right]^2 \, dx = 149.7849$$

which, similarly to that for the UK, is a minimum.

Now we can look at the relative weights of each function. We do this by examining the coefficients. To make it easier, we divide each coefficient by the one with the lowest magnitude and take the absolute value. Looking at the UK first, we get

$$\begin{bmatrix} u \\ v \\ s \\ t \end{bmatrix} = \begin{bmatrix} 4372 \\ -815.3 \\ -3998 \\ 57.51 \end{bmatrix} \to \begin{bmatrix} 75.933 \\ 14.193 \\ 69.437 \\ 1 \end{bmatrix}.$$

This tells us gas prices and battery density have a much higher impact on e-bike sales than energy cost, which in turn has a much higher impact than the population's value of environmental issues.
Similarly, examining the values we got for the US, we get

$$\begin{bmatrix} u \\ v \\ s \\ t \end{bmatrix} = \begin{bmatrix} -35231 \\ -1509 \\ 19695 \\ 16522 \end{bmatrix} \to \begin{bmatrix} 23.3472 \\ 1 \\ 13.0516 \\ 10.9489 \end{bmatrix}$$

meaning gas prices carry about twice the weight of battery density and disposable income, both of which carry a much higher weight than energy cost.

## 2.6 Model Revision

We initially considered using a Fourier transform to compute the decomposition; however, it quickly became apparent that using a nonlinear regression worked better and more efficiently. We also decided to normalize the functions because we realized that the decomposition model disproportionately weighted the more numerically significant functions.

## 2.7 Discussion

While it is interesting that we can approximate our function as a linear combination of others (and indeed, there are some cool applications of Linear Algebra to the more general concept of bases for $C(\mathbb{R})$), we don't spend much time on this consideration because we normalized the functions in order to 'prepare' for our approximation.

While we lose scale, normalizing the functions allows us to see the impact they each have on the normalized sales function. This revealed some interesting underlying patterns, notably (and unfortunately somewhat predictably) that people's polled value of environmental issues doesn't have a significant correlation with actions taken to actually mitigate these issues.

Strengths:

- This model accounts for a variety of potential factors rather than being limited to just one.
- This model effectively highlights which factors have higher or lower impacts on e-bike sales.

Weaknesses:

- We were unable to factor the population's value of environmental issues into the analysis for the US or average disposable income into the analysis for the UK.
- Machine learning methods could be more useful for accounting for more variables and generating extremely accurate weights for each variable since they can have multiple of layers to optimize the given function

## 2.8 Sensitivity Analysis

Though our model effectively classifies the importance of 4 different variables to the growth of electric bikes in the UK and US, our model does only account for four variables, when in reality, there are hundreds of less important variables that could have collectively made a sizeable impact on the growth of electric bike sales.

## 2.9 Technical Computing Tools

We used MATLAB to compute the regression models for each of the functions and then to decompose the sales function into a linear combination of the factors.

# 3 Off the Chain

## 3.1 Defining the Problem

In this problem, we are tasked with quantitatively defining the impact of e-bikes on emissions, traffic congestion, health and wellness, and other factors. We are given usage statistics in the form of passenger miles, for different forms of transportation between 1990 and 2020. To complete this task, we will first show the extent to which e-bikes have had an effect on these usage statistics, and then that these statistics are representative of the impacts provided.

## 3.2 Assumptions and Justification

1. **Changes in public transportation during 2020 cannot be attributed to e-bikes.**
   Justification: The impact of the COVID-19 outbreak on transportation rates is much more significant than the impact of e-bikes, demonstrated by the dramatic decline[2] of public transportation.

2. **All recorded e-bike sales are first-hand from manufacturer to customer.**
   Justification: While second-hand sales are not entirely negligible, they are hard to track, so the provided datasets most likely reflect first-hand sales from manufacturer to customer. Furthermore, given the short lifespan of e-bikes[6], it is unlikely that people will be willing to sell their bikes second-hand in the brief time before they break.

3. **e-bike sales will eventually level out, as they reach a carrying capacity in a location.**
   Justification: As shown in problem 1, eventually the sales will level out.

4. **If there is no causal relationship between e-bike sales and another factor, it is reasonable to assume that there will not be a significant causal relationship in the future.**
   Justification: All the math modeling is based on the assumption that trends will continue. We are only able to draw conclusions based on the data we have, so we must assume that the conclusions we draw about the past will continue to be true in the future.

### 3.3 Variables

**Table 6**

| **Symbol** | **Definition** | **Unit / value** |
|---|---|---|
| $b_l(t)$ | The total number of e-bikes (in thousands) sold in a given location $l$ (as a function of year $t$) | Thousands of e-bikes |
| $P_l(t)$ | The total population of a given location in hundreds of millions $l$ (as a function of year $t$) | Hundreds of millions of people |
| S | The average lifespan of an e-bike | 4[6] |
| $B_l(t)$ | The cumulative number of e-bikes (in thousands) present in a given location $l$ (as a function of year $t$) | Thousands of e-bikes |
| $D_{lm}(t)$ | The total passenger distance traveled (in billions of kilometers) in a given location $l$ of a given mode $m$ (as a function of year $t$) | Billions of passenger kilometers |
| $A_l(t)$ | The average passenger distance traveled (in kilometers) in a given location $l$ (as a function of year $t$) | Kilometers |

*The subscripts $u, b$ are used for the United States, and the United Kingdom respectively*
*The subscripts $p, m, e$ are used for passenger vehicles (cars, trucks, taxis, motorbikes), mass transportation (trains, buses, ferries), and e-bikes respectively*

### 3.4 The Model

We started by taking the per capita sales models from problem 2 (see Table 4). We multiplied these expressions by the populations of these locations (found in [1]) to find the total sales per year in each location. These models are shown below:

**Table 7**

| **Location** | **function** | **r-squared value** |
|---|---|---|
| Model | $b_l(t) = A \cdot e^{Bt}$ | - |
| United States | $b_u(t) = 2.2768 \cdot e^{.2729t}$ | .9182 |
| United Kingdom | $b_b(t) = .2309 \cdot e^{.2729t} + .6240 \cdot e^{.3055t}$ | - |

*Total sales per year in US and UK*

We then created a loss function based on the lifespan of an e-bike. Since e-bikes last for 4 years on average, we simply shifted the sales function over 4 units and made it negative. This makes sense because four years after sale, the e-bike will break. Below is an example loss function for the United States:

$$-2.2768 \cdot e^{.2729(t-4)}$$

We then integrated the sum of the sales and loss functions to create a function for the net number of e-bikes in a location $l$. Below is an example of this function for the United States:

$$B_u(t) = 8.3430\left(e^{.2729t} - e^{.2729(t-4)}\right)$$

Our next step was to find the average number of kilometers traveled per person in a location $l$ as a function of year $t$. To do so, we wrote out the following equation:

$$P_l(t) \cdot A_l(t) = D_{lm}(t) + D_{lp}(t) + B_l(t) \cdot A_l(t)$$
$$A_l(t) = \frac{D_{lm}(t) + D_{lp}(t)}{P_l(t) - B_l(t)}$$

We then applied this equation to the United States and the United Kingdom to find the average kilometers traveled per person each year from 2000-2019. Below are some of the results from the US:

**Table 8**

| **Year** | **Population** | **AVG kilometers per person** |
|---|---|---|
| 2000 | 282162411 | 25352.04695 |
| 2005 | 295516599 | 26944.17828 |
| 2010 | 309327143 | 26520.11992 |
| 2015 | 320738994 | 26991.3837 |

*Modeled kilometers traveled per person in years 2000, 2005, 2010, and 2015*

We then used these values in conjunction with the models for total e-bikes ($B_l(t)$) to predict the total passenger miles traveled by e-bike each year from 2000-2019 (we started at 2000 because as stated in problem 1, we assume e-bike sales to be negligible before 2000, and we disregarded 2020 due to COVID confounding the data)

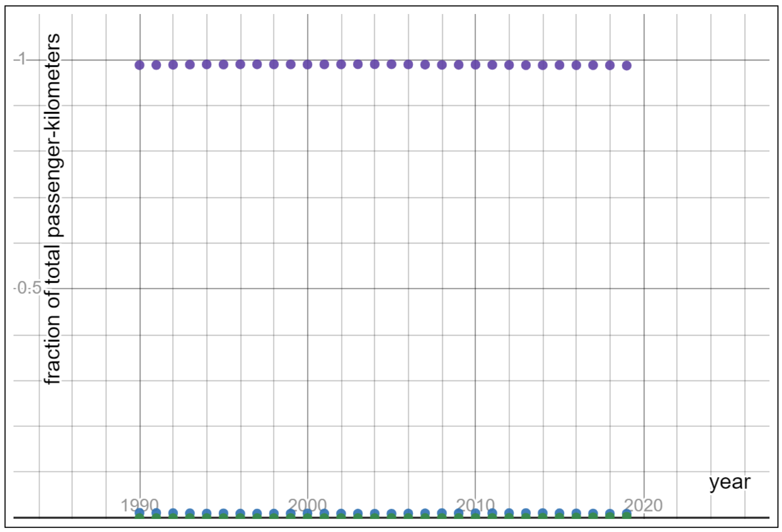
Now that we had $D_{lm}(t)$, $D_{lp}(t)$, and $D_{le}(t)$, we could start to show relationships between the three. Our first step was to define each of these in terms of the total passenger miles traveled $D_l(t)$ (which is simply the population * the average kilometers traveled per person). To do so, we created the following ratios for the United States and the United Kingdom defined from 2000-2019:

$$\frac{D_{lm}(t)}{D_l(t)};\ \frac{D_{lp}(t)}{D_l(t)};\ \frac{D_{le}(t)}{D_l(t)}$$

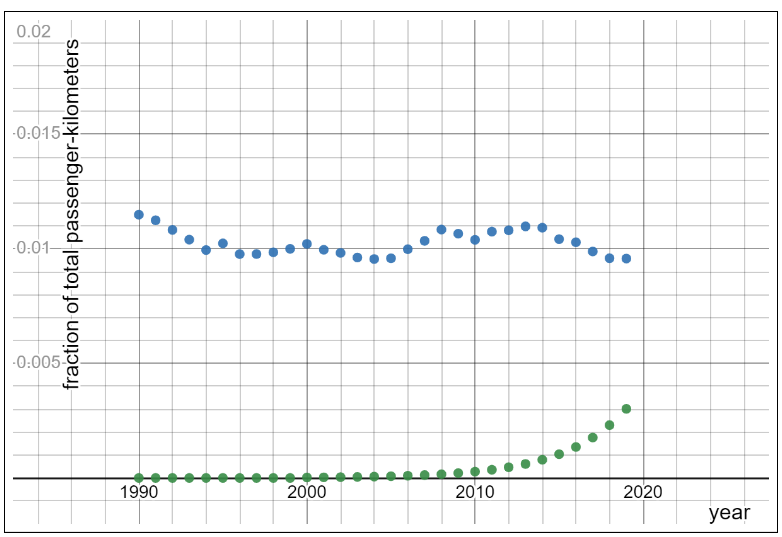
The sum of these three ratios will always be 1, since $D_l(t)$ is, by definition, $D_{lm}(t) + D_{lp}(t) + D_{le}(t)$

When we graph these ratios over the time period 2000-2019 for the United States and the United Kingdom, we see the following:

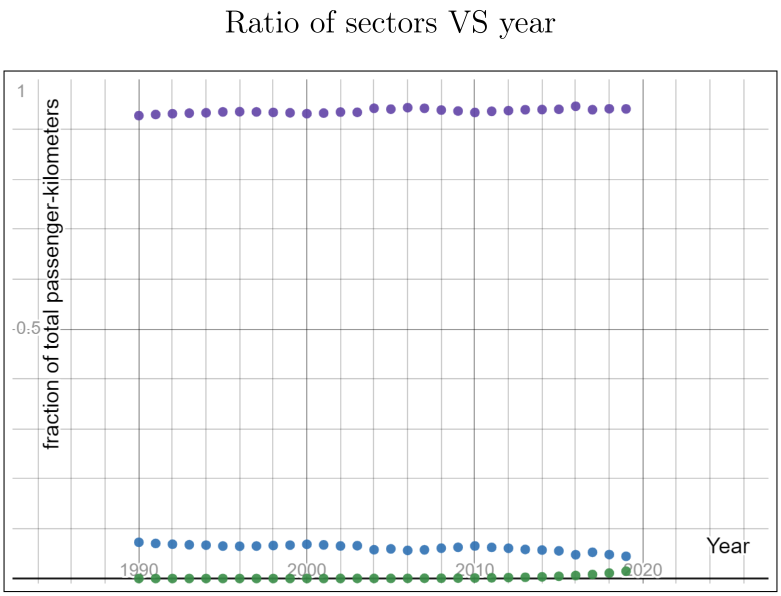
Ratio of sectors VS year

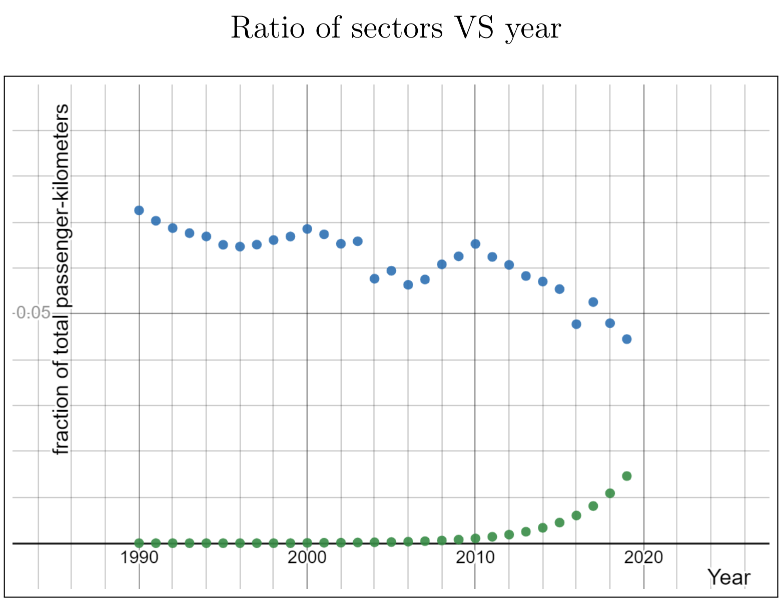



Ratio of sectors VS year




*$\frac{D_{ue}(t)}{D_u(t)}$ (green), $\frac{D_{up}(t)}{D_u(t)}$ (purple), and $\frac{D_{um}(t)}{D_u(t)}$ (blue) VS time for the United States*




$\frac{D_{be}(t)}{D_b(t)}$ *(green)*, $\frac{D_{bp}(t)}{D_b(t)}$ *(purple), and* $\frac{D_{bm}(t)}{D_b(t)}$ *(blue) VS time for the United Kingdom*

Here, it is obvious that the e-bike has had less of an effect on the transportation industry in the United States and more of an effect in the United Kingdom. This can be attributed to the greater proportion of vehicles that e-bikes make up in the United Kingdom[2].

While the personal vehicle proportion seems undisturbed, the mass transportation proportion seems to be affected by e-bikes. To quantify this effect, we applied linear regression for $\frac{D_{bm}(t)}{D_b(t)}$ from 1990-2010, and linear regression for $\frac{D_{bm}(t)}{D_b(t)}$ from 2010-2019, which is when the e-bike started to take a non-negligible share of the transportation industry[2]. Here are the regression equations:

$$\begin{aligned}
\frac{D_{bm}(t)}{D_b(t)} &= -0.000513528t + 1.09182;\ (1990 \leq t \leq 2010) \\
r^2 &= .5457 \\
\frac{D_{bm}(t)}{D_b(t)} &= -0.00219267t + 4.47232;\ (2010 \leq t \leq 2019) \\
r^2 &= .9319
\end{aligned}$$

## 3.5 Results

From this, we can see that the arrival of e-bikes is correlated with the change in the rate at which $\frac{D_{bm}(t)}{D_b(t)}$ was changing (i.e. it has changed the speed at which mass transportation was losing its proportion of total person-kilometers per year in the United Kingdom)

Realistically, however, the proportion of mass transportation was already on the decline[2], and the downturn from 2010-2019 is similar to the downturn from 1990 to 1996, before e-bikes were prevalent. Thus, we cannot attribute the downturn from 2010-2019 to e-bikes without knowing much more about other factors. Correlation is not causation, and it would be irresponsible to assume that this is a causal relationship.

Furthermore, the mass transportation industry has a small (less than 10%) share of the total transportation in the UK. Also, mass transportation is already more sustainable than personal transportation, and contributes much less to carbon emissions, traffic congestion, and health and wellness. Therefore, even if this was a causal relationship, the effects would be negligible.

Given all of this, we came to the conclusion that e-bikes have not significantly impacted carbon emissions, traffic congestion, or health and wellness, in both the US and the UK.

## 3.6 Model Revision

Originally, we compared the raw passenger-kilometers of personal and mass transportation before and after e-bikes became prevalent. While this did allow us to reach the same conclusion, we worried that we were not accounting for population growth. Thus, we revised the model to use the ratio of mass transportation and personal transportation to the net passenger-kilometers in each location. In this way, we could be sure that our data was not confounded by population growth.

## 3.7 Discussion

Ultimately, our model has led us to conclude that e-bikes have not significantly impacted carbon emissions, traffic congestion, or health and wellness, in both the US and the UK. Interestingly, there seemed to be somewhat of a relationship in the UK, though it was not enough to be statistically significant.

Strengths:

- Accounts for population growth and minimizes the effect of other trends by comparing sectors to the transportation industry as a whole
- Results in a normalized and easy-to-understand graph that can be compared to other locations due to the sum of the three functions always equaling 1

Weaknesses:

- May still be confounded by unknown variables, resulting in a lack of relationship when one may actually be present
- Fails to account for future years where e-bikes may become much more prevalent, instead assuming that the trend of mass transportation to e-bike will continue and that the e-bikes will eventually reach capacity

## 3.8 Sensitivity Analysis

Much of this model relies heavily on the five data points of e-bike sales in the United States provided. The model for the US directly extrapolates this data, and the model for the UK is half-based on this data. Thus, another added data point may have a large impact on our conclusions.

## 3.9 Technical Computing Tools

We used the MATLAB 'fit' function to create the exponential models, like we did in problem 2. We used functions in Excel to compute the values for $A_l(t)$ since it is a discrete set of values and did not require the use of MATLAB.

# 4 Conclusion and Reflections

First, we used past sales data to estimate e-bike sales in the United States and the United Kingdom in 2025 and 2028. We used a logistic regression model to extrapolate the sales data from the United States and France, and then averaged the two models to estimate e-bike sales in the United Kingdom. We used a logistic model to reflect the idea that only a subsection of a population would be open to buying an e-bike, and validated our model by finding that it predicted 1 in 25 people would own an e-bike, which seems like a reasonable proportion.

Next, we investigated the factors that led to the increase in e-bike sales in the United States and the United Kingdom given data sets of important variables/factors. Some of these factors included gas prices, the public's valuation of environmental issues over time, and the cost/value of electric vehicle charging. Given 6 data sets, we were able to establish 4 initial factors contributing to the growth of electric bike sales in the US and the UK. We chose these factors based on their fit of the exponential model we established previously. In order to derive the importance of each factor to the growth of electric bike sales, we found a linear combination of all of the factors' exponential curve fits to most closely approximate the function of the electric bike sales. Through the process of gradient descent, we were able to find the coefficients/weights of each factor in the optimal linear combination. This showed that the most important factors to the growth in electric bike sales in both the US and the UK were gas prices and battery density.

Finally, we worked to predict the effects e-bikes would have on carbon emissions, traffic congestion, and health and wellness. We used transportation and population data from 1990 to 2019 to find the share of the transportation industry private vehicles, mass transportation, and e-bikes had respectively over the years. In the United Kingdom, we saw the percent of mass transportation go down when e-bikes started to become popular. We were not able to establish a causal relationship, however, and given the small market share of mass transportation, concluded that the advent of e-bikes had little to no effect on carbon emissions, traffic congestion, and health and wellness.

# References


[1] *Population, total - united states, data,* https://data.worldbank.org/indicator/SP.POP.TOTL?locations=US.

[2] *Ride like the wind without getting winded: The growth of e-bike use, mathworks math modeling challenge 2023,* https://m3challenge.siam.org/node/596.

[3] *Global ev outlook 2022 - data product*, May 2022.

[4] *Least-squares function approximation*, Sep 2022.

[5] G. Beylkin and L. Monzón, *On approximation of functions by exponential sums*, Applied and Computational Harmonic Analysis, 19 (2005), p. 17–48.

[6] L. Bonk, *How long do electric bikes last: Ebike lifespans explained*, Mar 2023.

[7] C. L. (https://math.stackexchange.com/users/82404/claude leibovici), *Exponential sum approximation.* Mathematics Stack Exchange. URL:https://math.stackexchange.com/q/3717613 (version: 2020-06-13).

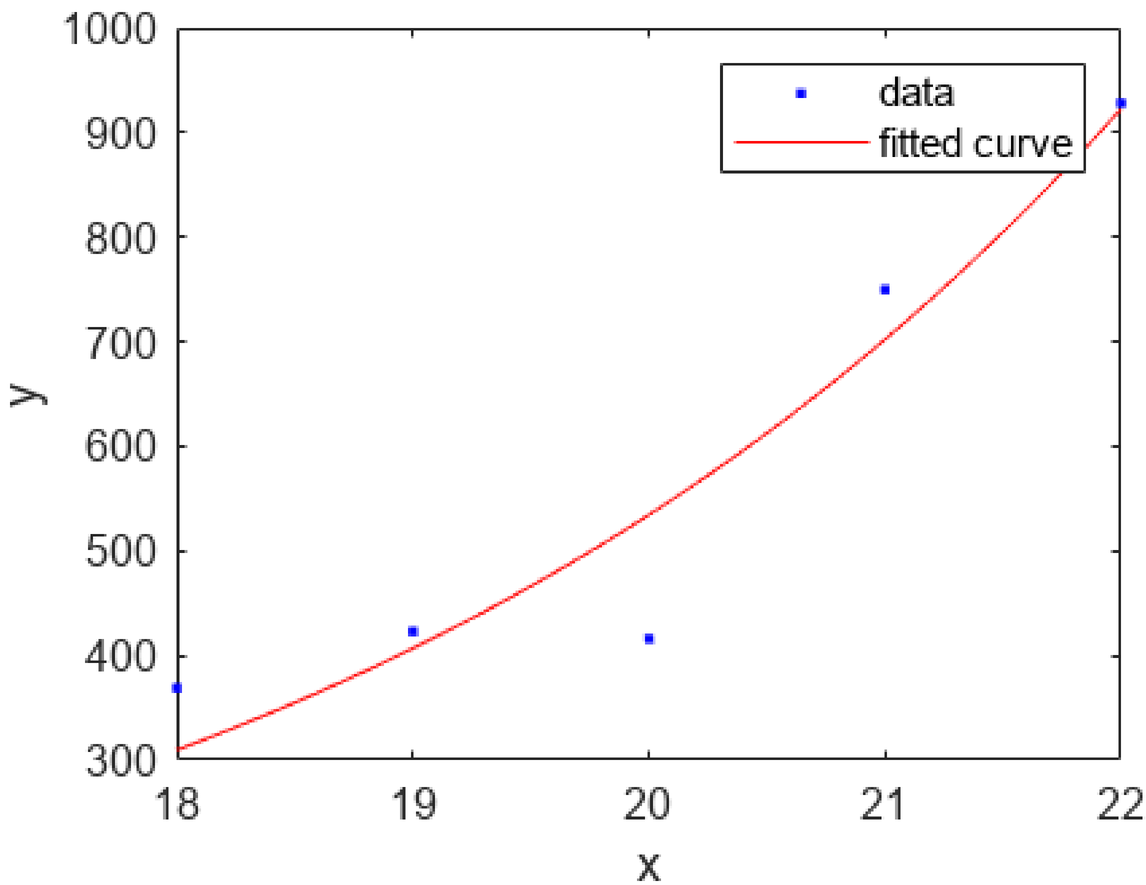
[8] P. Norman, *Electric bikes and the law: Everything you need to know about ebike regulations*, Feb 2023.

[9] E. Ramos, *Covid fueled the rise of the e-bike. see where ridership grew in the u.s.*, Jun 2021.

# Code Appendix

## Problem 1

```
%The US electric bike sales data fitted to an exponential model
USList = transpose([369,423,416,750,928])
x = transpose([2018,2019,2020,2021,2022])
x = x-2000
f = fit(x, USList, 'exp1')
plot(f, x, USList)
```

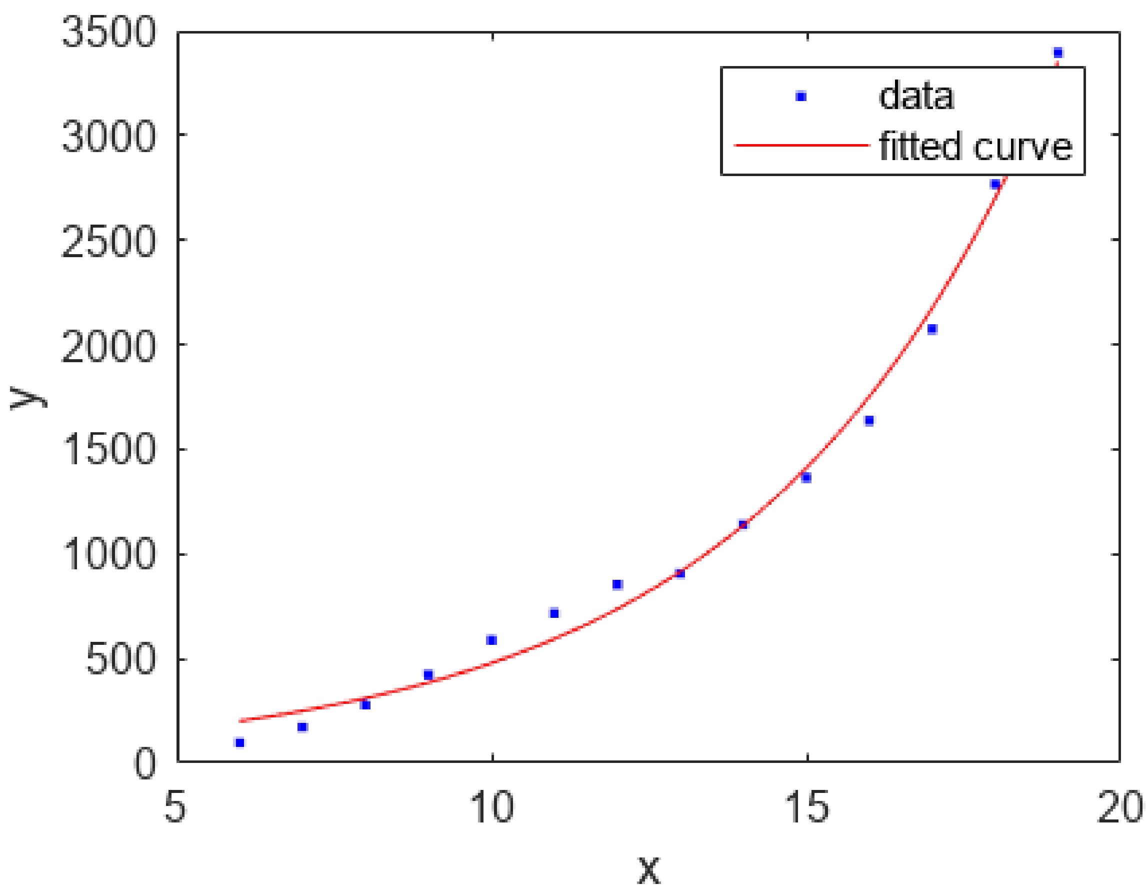



```
[curve2,gof2] = fit(x, USList, 'exp1')
```

```
curve2 =
     General model Exp1:
     curve2(x) = a*exp(b*x)
     Coefficients (with 95% confidence bounds):
       a =       2.276  (-5.366, 9.918)
       b =      0.2729  (0.1131, 0.4328)
```

```
gof2 =
           sse: 2.0136e+04
       rsquare: 0.9182
           dfe: 3
    adjrsquare: 0.8909
          rmse: 81.9270
```

```
%The Europe electric bike sales data fitted to an exponential model
EuropeList = transpose([98,173,279,422,588,716,854,
    907,1139,1364,1637,2074,2767,3397])
x = transpose([2006,2007,2008,2009,2010,2011,2012,2013,2014,
        2015,2016,2017,2018,2019])
x = x-2000
f = fit(x, EuropeList, 'exp1')
plot(f, x, EuropeList)
```

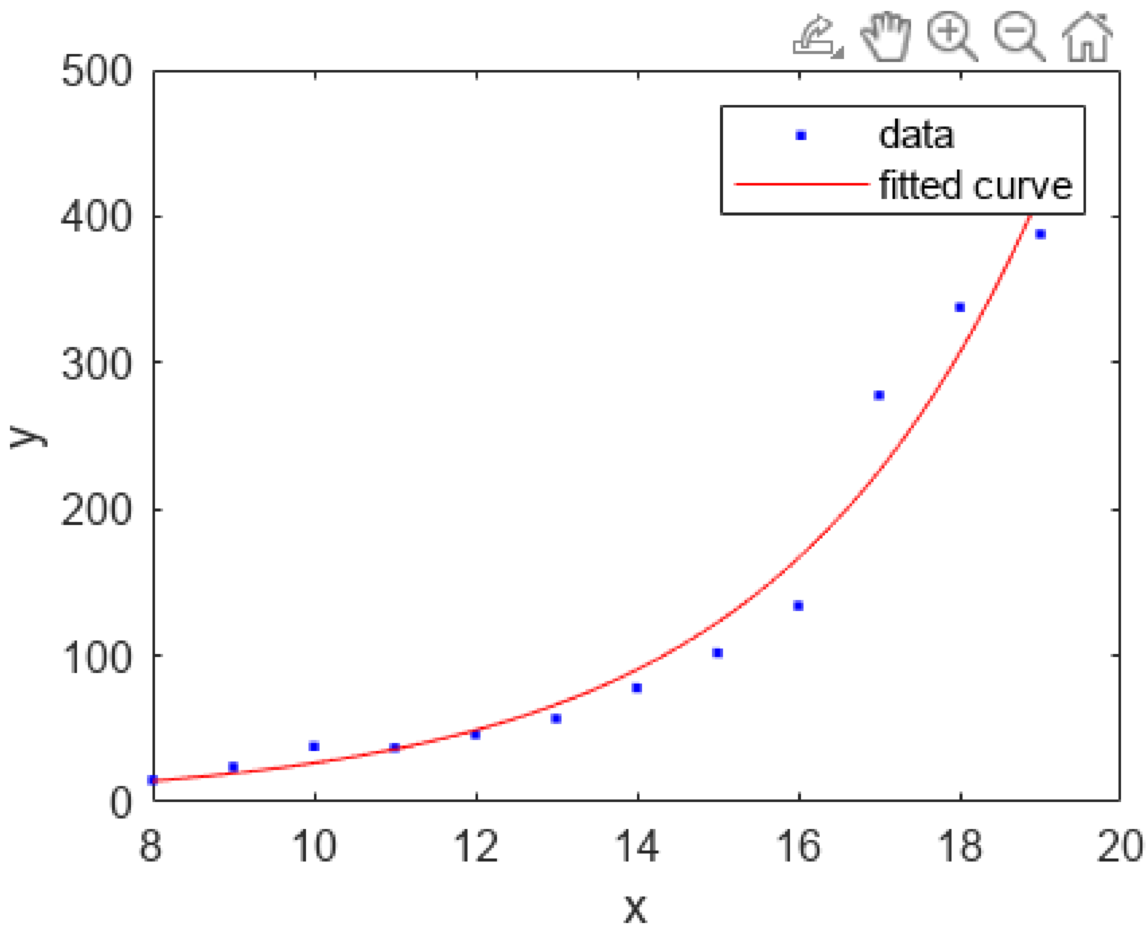



```
[curve4,gof4] = fit(x, EuropeList, 'exp1')
```

```
curve4 =
```

```
        General model Exp1:
        curve4(x) = a*exp(b*x)
        Coefficients (with 95% confidence bounds):
          a =       55.02  (40.37, 69.68)
          b =      0.2163  (0.2009, 0.2317)
   gof4 =
              sse: 9.1573e+04
          rsquare: 0.9928
              dfe: 12
       adjrsquare: 0.9922
             rmse: 87.3560
```

```
%The France electric bike sales data fitted to an exponential model
FranceList = transpose([15,24,38,37,46,57,78,102,134,278,338,388])
x = transpose([2008,2009,2010,2011,2012,2013,2014,
        2015,2016,2017,2018,2019])
x = x-2000
f = fit(x, FranceList, 'exp1')
plot(f, x, FranceList)
```



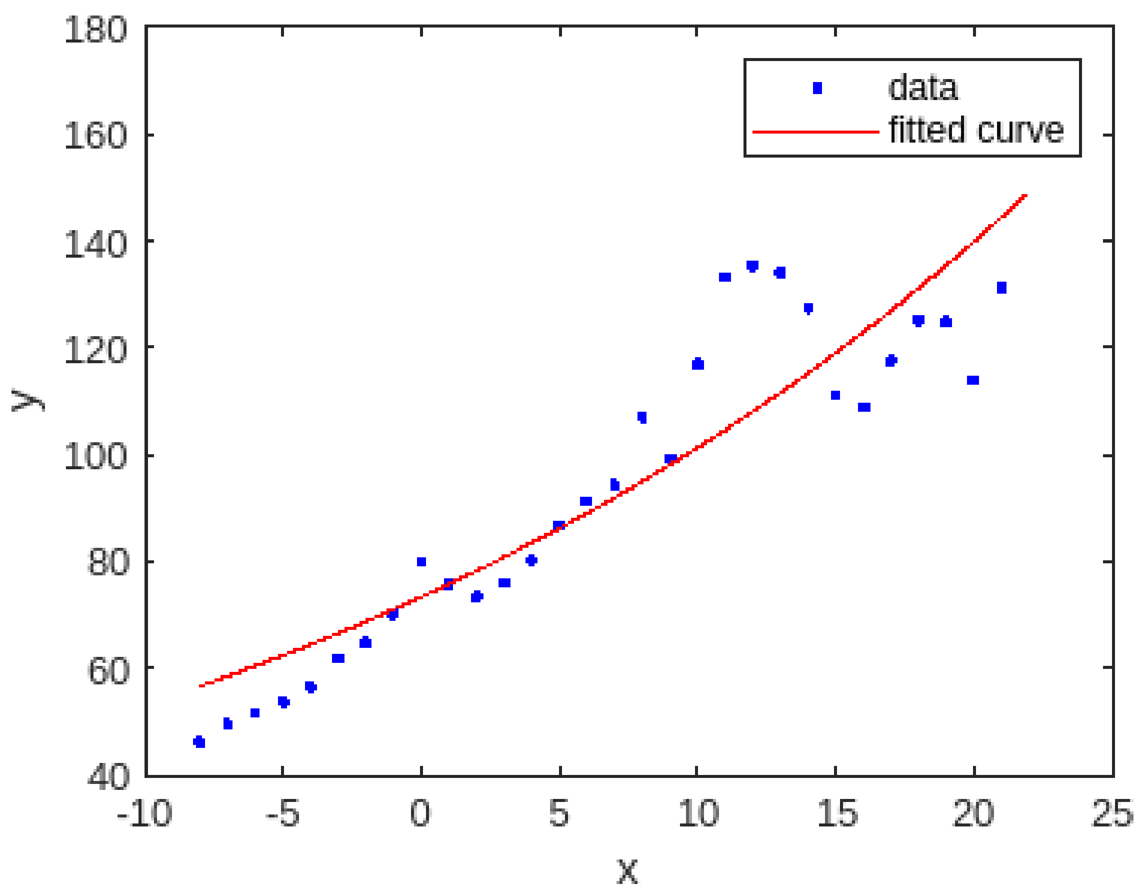
```
[curve3,gof3] = fit(x, FranceList, 'exp1')
```

```
curve3 =
     General model Exp1:
     curve3(x) = a*exp(b*x)
     Coefficients (with 95% confidence bounds):
       a =       1.256  (-0.04148, 2.553)
       b =      0.3055  (0.2477, 0.3632)
gof3 =
           sse: 6.3726e+03
       rsquare: 0.9663
           dfe: 10
    adjrsquare: 0.9629
          rmse: 25.2441
```

# Problem 2

### Factor Models

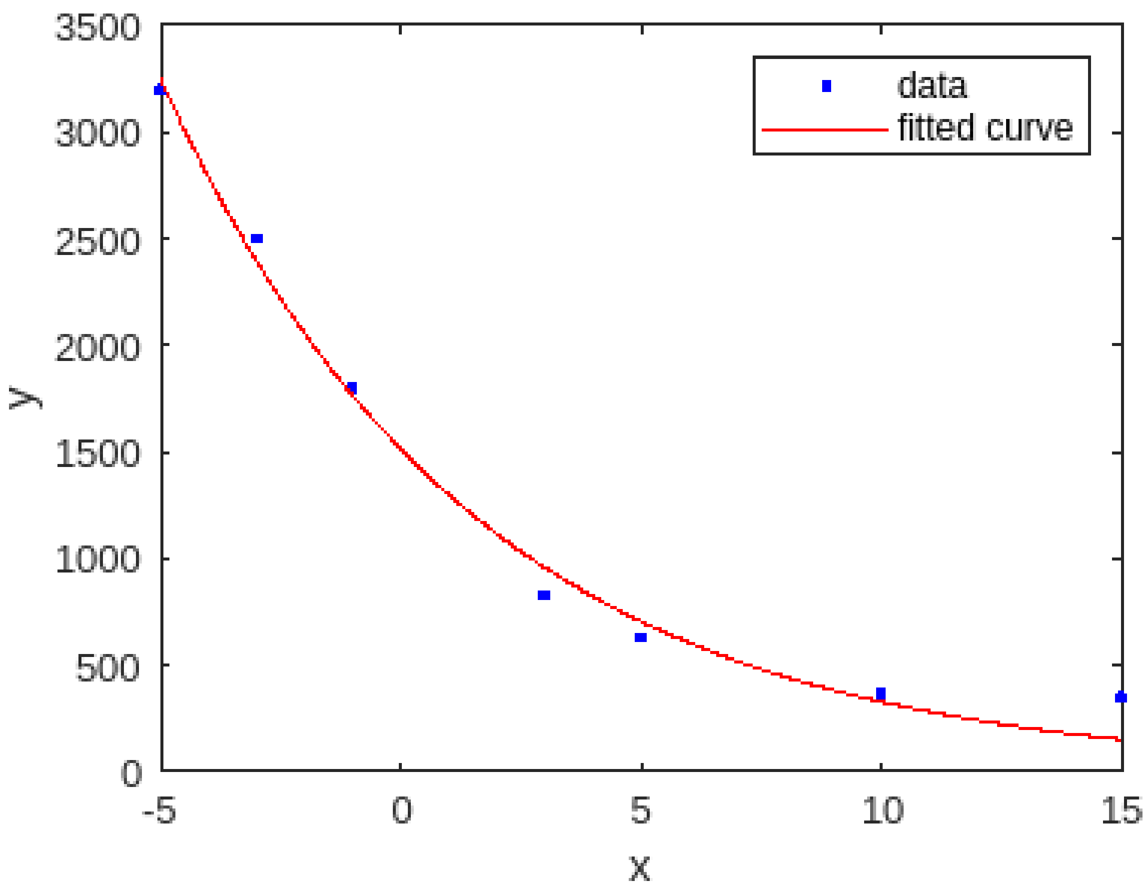
```
%UK gas price data fitted to an exponential model
gasdata = [46.07,49.44,51.58,53.77,56.52,61.82,64.8,70.16,79.93,75.72,
73.24,76.04,80.22,86.75,91.32,94.24,107.08,99.29,116.9,133.27,135.39,
134.15,127.5,111.13,108.85,117.59,125.2,124.88,113.95,131.27,164.73]
gasyears = [1992,1993,1994,1995,1996,1997,1998,1999,2000,2001,2002,2003,
2004,2005,2006,2007,2008,2009,2010,2011,2012,2013,2014,2015,2016,2017,
        2018,2019,2020,2021,2022]

plot_exponential(gasdata,gasyears)
```




```
curve6 =
     General model Exp1:
     curve6(x) = a*exp(b*x)
     Coefficients (with 95% confidence bounds):
       a =       73.33  (67.2, 79.45)
       b =      0.0323  (0.02649, 0.0381)
```

```
gof6 =
           sse: 4.7650e+03
       rsquare: 0.8386
           dfe: 29
    adjrsquare: 0.8330
          rmse: 12.8183
```

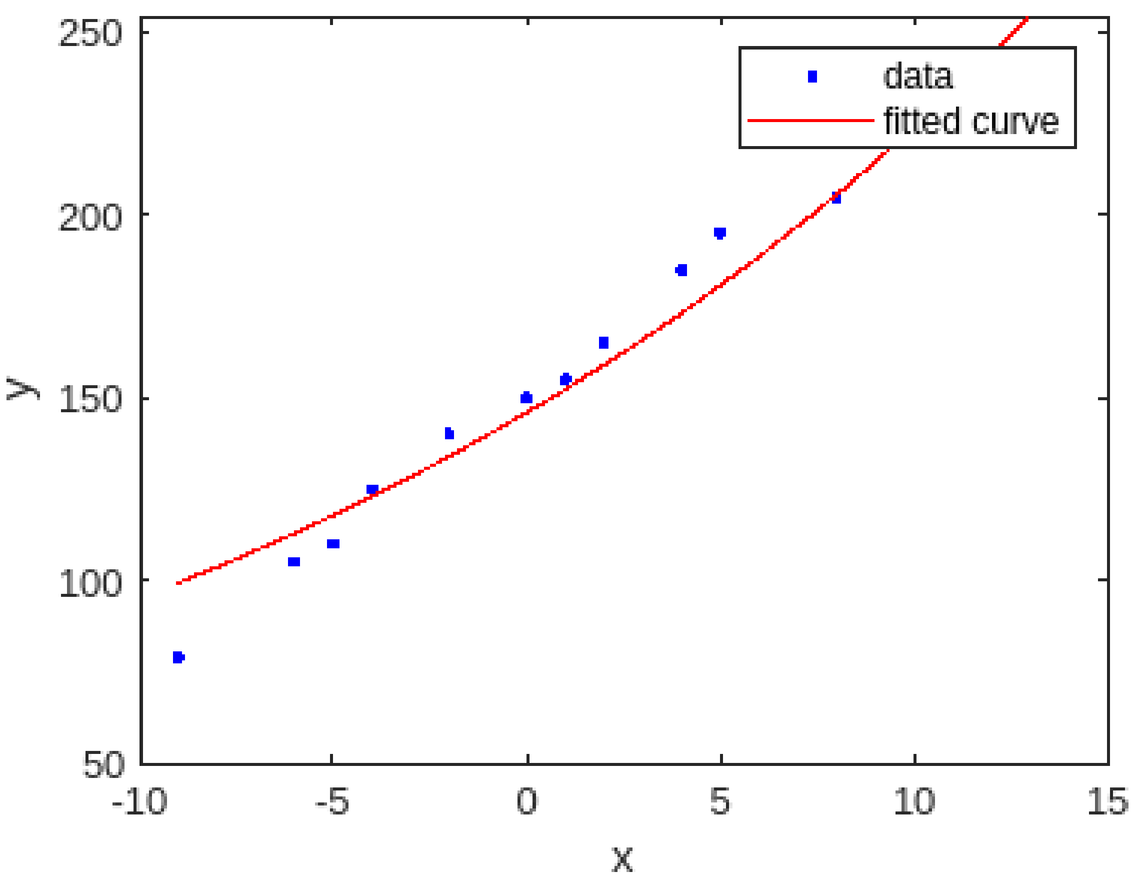
```
%Cost KwH of electricity for electric bikes fitted to an exponential model
costdata = [3200, 2500, 1800, 830, 630, 370, 350]
costyears = [1995, 1997, 1999, 2003, 2005, 2010, 2015]

plot_exponential(costdata,costyears)
```




```
curve6 =
     General model Exp1:
     curve6(x) = a*exp(b*x)
     Coefficients (with 95% confidence bounds):
       a =        1516  (1351, 1682)
       b =     -0.1526  (-0.1792, -0.1261)
```

```
gof6 =
           sse: 7.7336e+04
       rsquare: 0.9899
           dfe: 5
    adjrsquare: 0.9879
          rmse: 124.3671
```

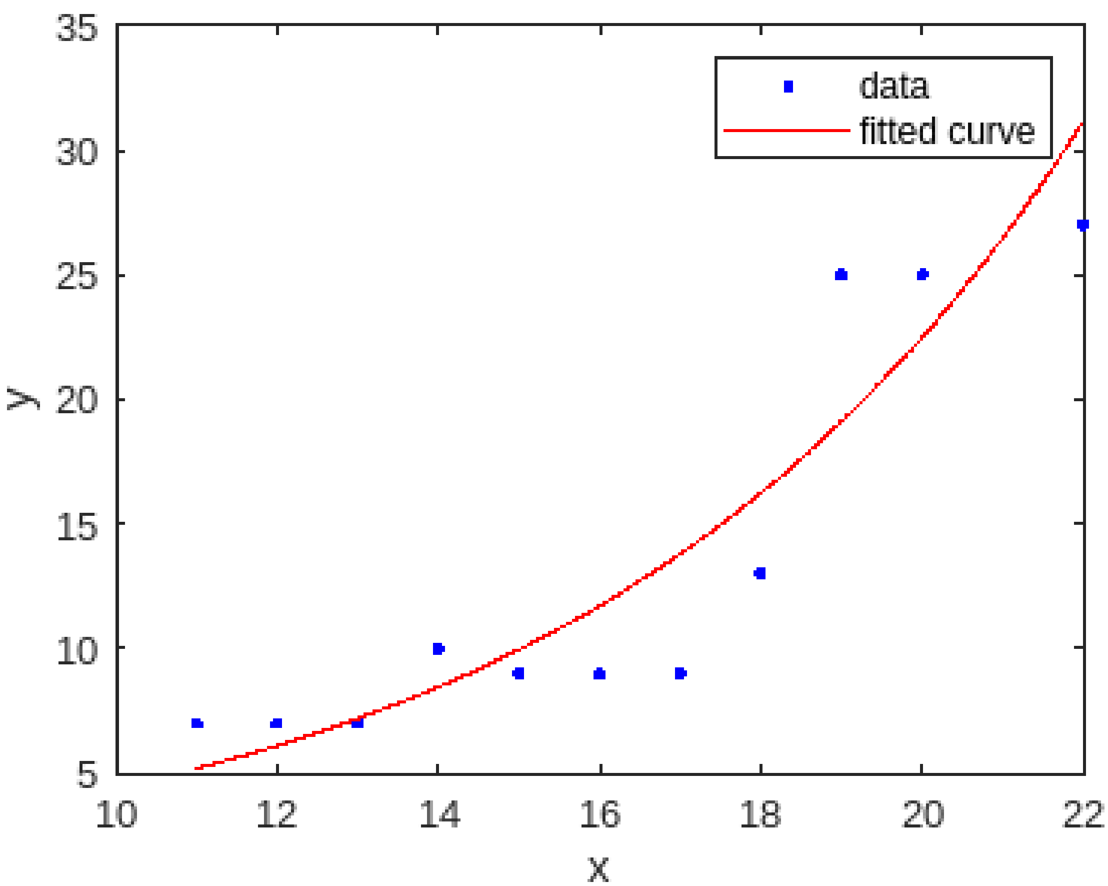
```
%Gravimetric Energy Densities (W-hr/kg) of electric bike batteries fitted
%to an exponential model
densitydata = [79,105,110,125,140,150,155,165,185,195,205,240]
densityyears = [1991,1994,1995,
1996,1998,2000,2001,2002,2004,2005,2008,2013]

plot_exponential(densitydata,densityyears)
```




```
curve6 =
     General model Exp1:
     curve6(x) = a*exp(b*x)
     Coefficients (with 95% confidence bounds):
```

```
       a =       145.9  (138.4, 153.4)
       b =     0.04278  (0.03585, 0.04972)
gof6 =
           sse: 1.1901e+03
       rsquare: 0.9499
           dfe: 10
    adjrsquare: 0.9449
          rmse: 10.9091
```

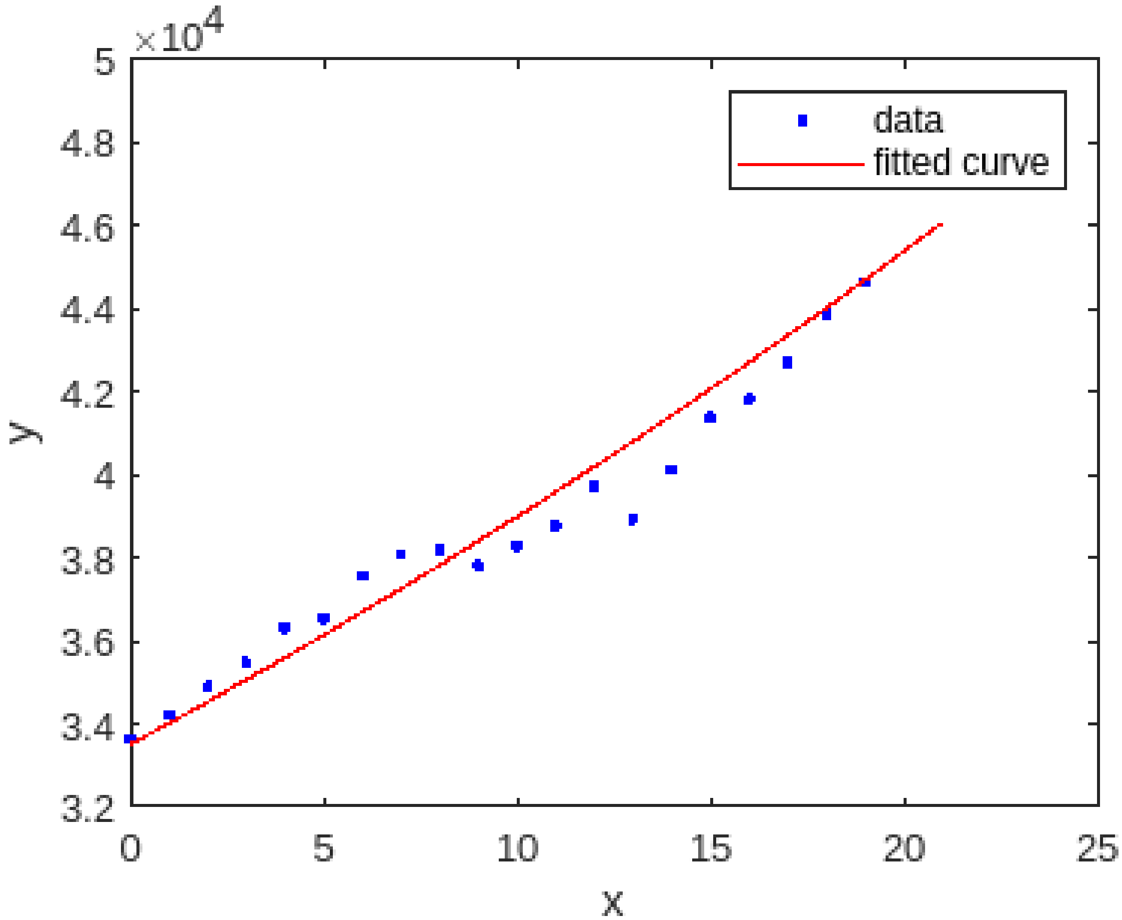
```
%Percentage of survey respondants who included "Climate Change and the
%Environment" in a list of the top three most important issues facing the
%country fitted over an exponential model
voteyears = [2011,2012,2013,2014,2015,2016,2017,2018,2019,2020,2021,2022]
votedata = [7,7,7,10,9,9,9,13,25,25,30,27]

plot_exponential(votedata,voteyears)
```




```
curve6 =
     General model Exp1:
```

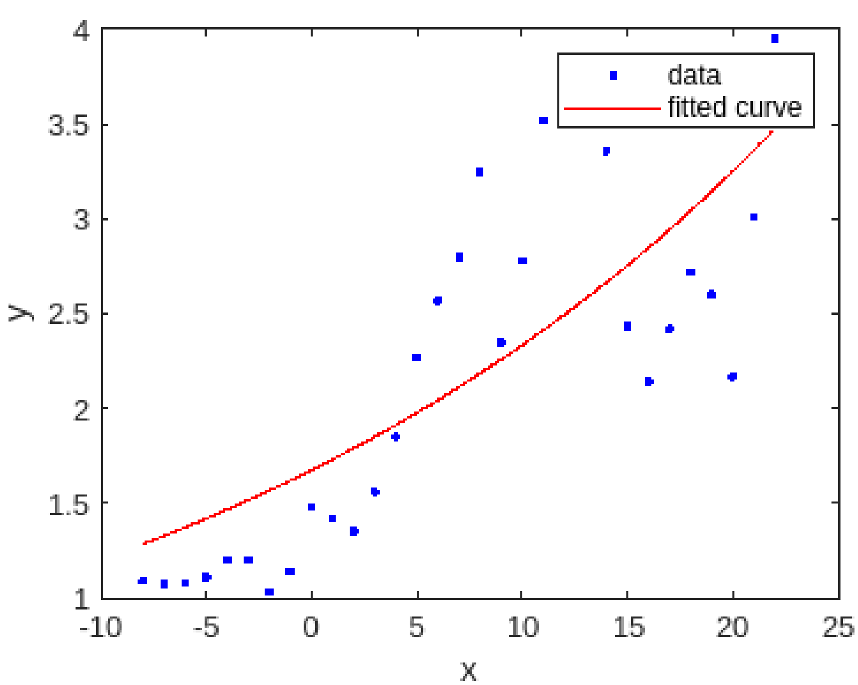
```
      curve6(x) = a*exp(b*x)
      Coefficients (with 95% confidence bounds):
        a =       0.8614  (-0.03063, 1.753)
        b =       0.1631  (0.1108, 0.2154)
gof6 =
            sse: 118.9774
        rsquare: 0.8675
            dfe: 10
     adjrsquare: 0.8542
           rmse: 3.4493
```

```
%Average disposable income for US citizens fitted to an exponential model
incomedata = [33645,34216,34894,35474,36325,36526,37570,38093,38188,37814,
38282,38769,39732,38947,40118,41383,41821,42699,43886,44644,47241,48219]
incomeyears = [2000,2001,2002,2003,2004,2005,2006,2007,2008,2009,2010,
2011,2012,2013,2014,2015,2016,2017,2018,2019,2020,2021]

plot_exponential(incomedata,incomeyears)
```



```
curve6 =
     General model Exp1:
     curve6(x) = a*exp(b*x)
     Coefficients (with 95% confidence bounds):
       a =    3.35e+04  (3.272e+04, 3.428e+04)
       b =     0.01518  (0.01342, 0.01694)
gof6 =
           sse: 1.9325e+07
       rsquare: 0.9419
           dfe: 20
    adjrsquare: 0.9390
          rmse: 982.9721
```

```
%US gas price data fitted to an exponential model
gasdata = [1.09,1.07,1.08,1.11,1.20,1.20,1.03,1.14,1.48,1.42,1.35,1.56,
1.85,2.27,2.57,2.80,3.25,2.35,2.78,3.52,3.62,3.51,3.36,2.43,2.14,2.42,
2.72,2.60,2.17,3.01,3.95]
gasyears = [1992,1993,1994,1995,1996,1997,1998,1999,2000,2001,2002,2003,
2004,2005,2006,2007,2008,2009,2010,2011,2012,2013,2014,2015,2016,2017,
2018,2019,2020,2021,2022]

plot_exponential(gasdata,gasyears)
```




```
curve6 =
     General model Exp1:
```

```
      curve6(x) = a*exp(b*x)
      Coefficients (with 95% confidence bounds):
        a =       1.675  (1.391, 1.958)
        b =     0.03318  (0.02147, 0.04489)
gof6 =
           sse: 10.2208
       rsquare: 0.5889
           dfe: 29
    adjrsquare: 0.5747
          rmse: 0.5937
```

```
function p = plot_exponential(data,years)
    years = years - 2000
    data = transpose(data)
    years = transpose(years)
    f = fit(years, data, 'exp1')
    plot(f, years, data)
    [curve6,gof6] = fit(years, data, 'exp1')
end
```

**UK Data**

```
% first we will define our functions that we've fit already
% Gas price - corresponds with u
g = @(x) 73.33*exp(0.0323*x)
% normalize
g1 = @(x) g(x)/g(0)

% Cost KwH - corresponds with v
c = @(x) 1516*exp(-0.1526*x)
% normalize
c1 = @(x) c(x)/c(0)

% Battery Density - corresponds with s
b = @(x) 145.9*exp(0.04278*x)
% normalize
b1 = @(x) b(x)/b(0)

% Environmental issues voting priority - corresponds with t
p = @(x) 0.8614*exp(0.1631*x)
% normalize
p1 = @(x) p(x)/p(0)
```

```
% Sales (function to approximate)
S = @(x) 0.5*(0.686*exp(0.2729*x)+1.8538*exp(0.3055*x))
% normalize
S1 = @(x) S(x)/S(0)


% now we can optimize stuff
% we will optimize between 2008 and 2022
% generate a bunch of data points
x1 = transpose(linspace(8,22))%,10000)
y1 = arrayfun(S1,x1)

% https://en.wikipedia.org/wiki/Least-squares_function_approximation
% regress
x0 = [1 1 1 1]

fitfun = fittype( @(u,v,s,t,x) double(u*g1(x)+v*c1(x)+s*b1(x)+t*p1(x)) )
```

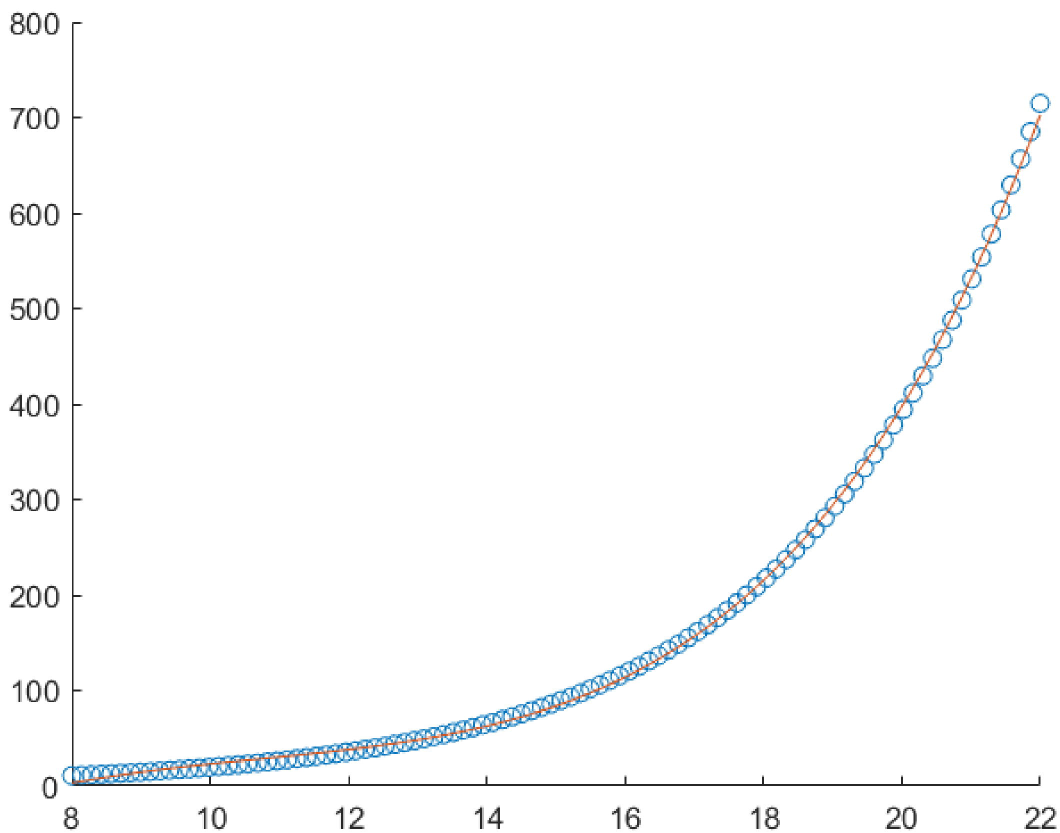
```
fitfun =
     General model:
     fitfun(u,v,s,t,x) = double(u*g1(x)+v*c1(x)+s*b1(x)+t*p1(x))
```

```
[fittedfunction,gof] = fit(x1,y1,fitfun,'StartPoint',x0)
```

```
fittedfunction =
     General model:
     fittedfunction(x) = double(u*g1(x)+v*c1(x)+s*b1(x)+t*p1(x))
     Coefficients (with 95% confidence bounds):
       u =        4372  (4161, 4583)
       v =      -815.3  (-883.5, -747.1)
       s =       -3998  (-4183, -3813)
       t =       57.51  (56.15, 58.86)
gof =
           sse: 1.3054e+03
       rsquare: 0.9996
           dfe: 96
    adjrsquare: 0.9996
          rmse: 3.6876
```

```
c = coeffvalues(fittedfunction)

% Plot results
scatter(x1,y1)
hold on
```

```
plot(x1,fittedfunction(x1))
hold off
```




```
% Test accuracy
I = integral(
    @(x) (c(1).*g1(x)+c(2).*c1(x)+c(3).*b1(x)+c(4).*p1(x)-S1(x)).^2,
    8,
    22
)
```

```
I = 169.0710
```

## US Data

```
% first we will define our functions that we've fit already
% Gas price - corresponds with u
g = @(x) 1.675*exp(0.03318*x)
% normalize
```

```
g1 = @(x) g(x)/g(0)

% Cost KwH - corresponds with v
c = @(x) 1516*exp(-0.1526*x)
% normalize
c1 = @(x) c(x)/c(0)

% Battery Density - corresponds with s
b = @(x) 145.9*exp(0.04278*x)
% normalize
b1 = @(x) b(x)/b(0)

% Average Disposable income - corresponds with t
p = @(x) 28780*exp(0.01518*x)
% normalize
p1 = @(x) p(x)/p(0)

% Sales (function to approximate)
S = @(x) 0.686*exp(0.2729*x)
% normalize
S1 = @(x) S(x)/S(0)


% now we can optimize stuff
% we will optimize between 2008 and 2022
% generate a bunch of data points
x1 = transpose(linspace(8,22))%,10000)
y1 = arrayfun(S1,x1)

% https://en.wikipedia.org/wiki/Least-squares_function_approximation
% regress
x0 = [1 1 1 1]

fitfun = fittype( @(u,v,s,t,x) double(u*g1(x)+v*c1(x)+s*b1(x)+t*p1(x)) )
```

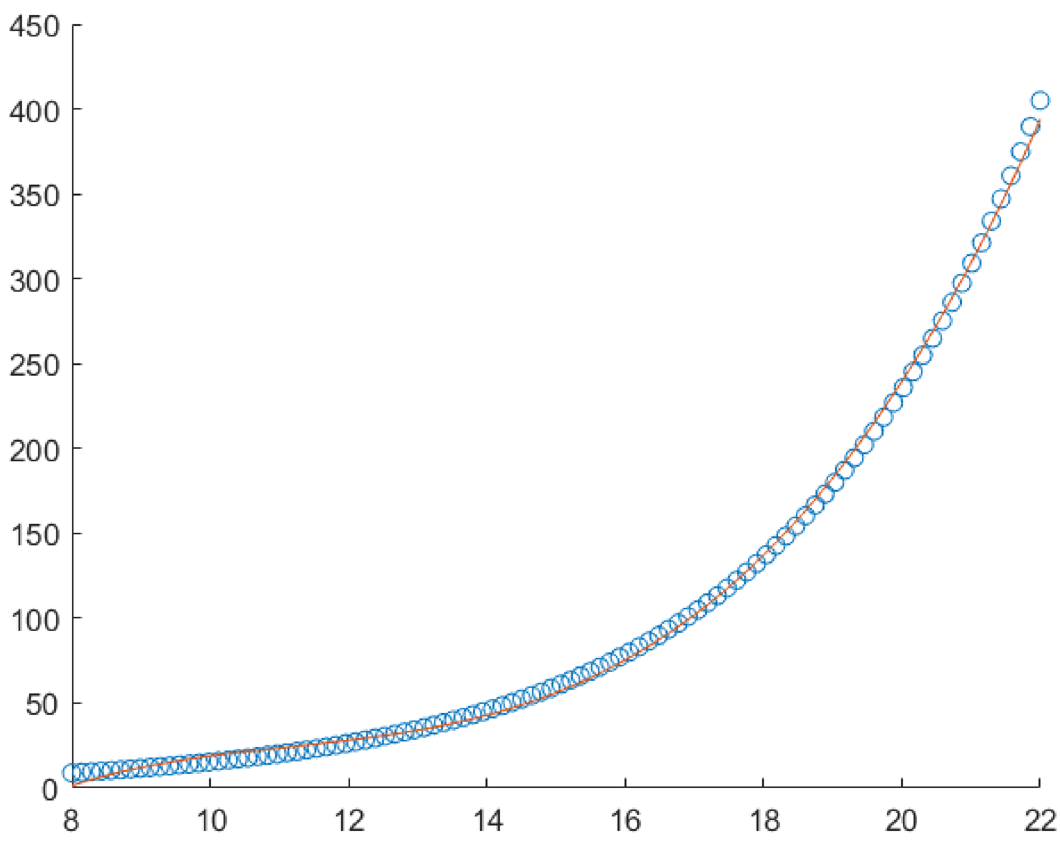
```
fitfun =
     General model:
     fitfun(u,v,s,t,x) = double(u*g1(x)+v*c1(x)+s*b1(x)+t*p1(x))
```

```
[fittedfunction,gof] = fit(x1,y1,fitfun,'StartPoint',x0)
```

```
fittedfunction =
     General model:
     fittedfunction(x) = double(u*g1(x)+v*c1(x)+s*b1(x)+t*p1(x))
     Coefficients (with 95% confidence bounds):
```

```
        u =  -3.523e+04  (-3.673e+04, -3.373e+04)
        v =       -1509  (-1626, -1392)
        s =    1.97e+04  (1.889e+04, 2.05e+04)
        t =   1.652e+04  (1.576e+04, 1.728e+04)
gof =
            sse: 1.1564e+03
        rsquare: 0.9990
            dfe: 96
     adjrsquare: 0.9989
           rmse: 3.4708
```

```
c = coeffvalues(fittedfunction)

% Plot results
scatter(x1,y1)
hold on
plot(x1,fittedfunction(x1))
hold off
```



```
% Test accuracy
I = integral(
    @(x) (c(1).*g1(x)+c(2).*c1(x)+c(3).*b1(x)+c(4).*p1(x)-S1(x)).^2,
    8,
    22
)
```

```
I = 149.7849
```