

## MATLAB Semester Project

### Project Title:

Creating a MATLAB program to predict (with graphical visualization) the Total Number of COVID-19 Deaths for a User-chosen US State using Linear Regression.

**Deadline: 12/04/2020, 11:59 PM**

### Instructions:

Write a MATLAB script that implements the following:

1. Reading and displaying the following dataset in a table:  
<https://raw.githubusercontent.com/nytimes/covid-19-data/master/us-states.csv>
2. Provide the user the option of choosing a State (by its name or FIPS-ID whichever you find convenient by your logic).
3. You should then filter your data table for the chosen US State (For example, Texas with FIPS ID- 48). Keep in mind your app should be generically modeled for filtering any state on demand, so your script must not be hard coded for Texas or any single state. The filtered table can be obtained by applying conditional statements or for loop or logical indexing or a combination of them in your MATLAB code.
4. The dates in the table must be converted into day counts. For Texas, the day count starts on 02/07/2020. So 02/07/2020 should be replaced by day #1 and the day number extends up to the latest when the program is executed.
5. You then plot a graph based on Day# and Number of Deaths. For Texas, the graph would look like below with red circle markers:

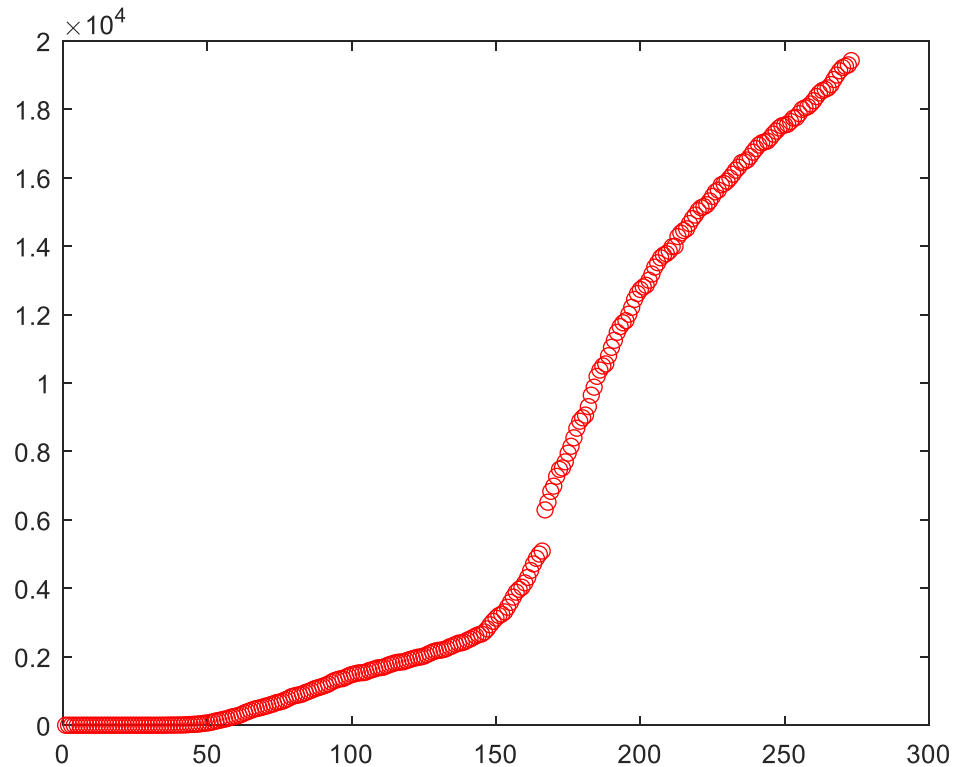


Figure 1: Days (in x-axis) and Number of Deaths (in y axis) plotted

6. From the graph, it is evident that the 'days vs deaths' phenomenon takes an approximately linear relationship from some inflection point of days# (between 100 and 120). At this stage, you let the user to choose (through an input prompt) an inflection point (in terms of day#) from where you shall calculate your linear regression predictive model.
7. Both your data arrays of days and deaths should be modified with their size (number of elements or datapoints) upon that user-chosen inflection point. (For example, 'days' being user\_input\_inflection:end instead of previously being 1:end; similar modifications will be necessary for 'deaths' because you must have same number of data points in the independent and dependent variable arrays to implement the linear regression).
8. At this stage if you implement linear regression for that range of user input inflection point to the last available data point, and if you plot the regression line on top of your observed data of figure 1, you will come across a plot like below for Texas:

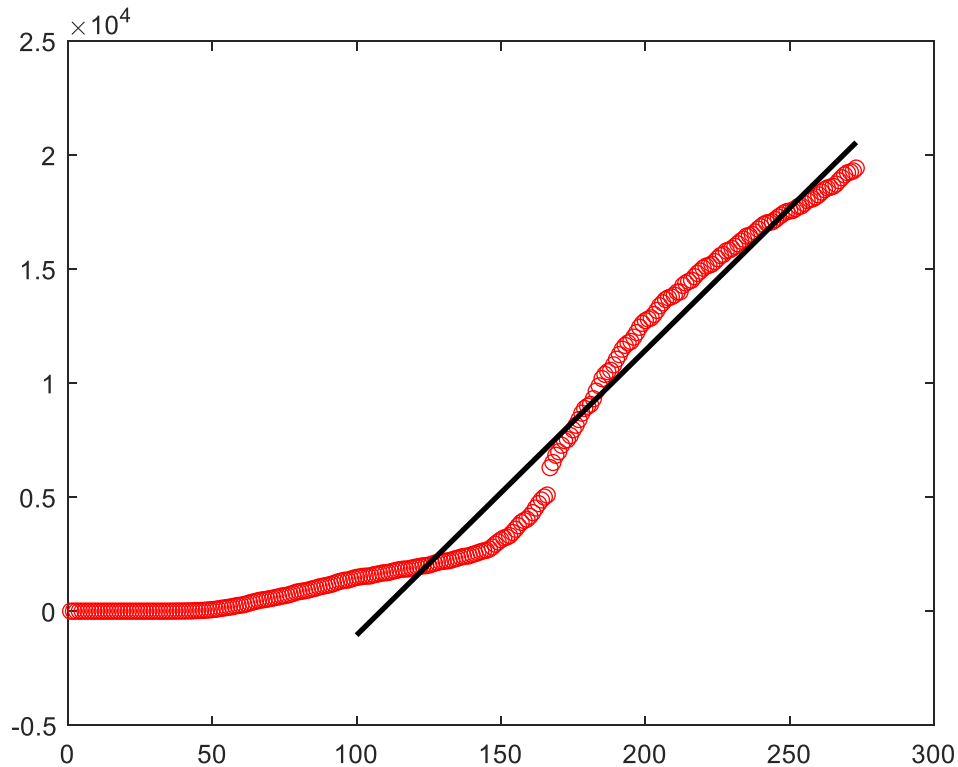


Figure 2: Linear Regression implemented from the user chosen inflection point

9. Your program must determine and display the slope and y-intercept for the regression line.
10. With the obtained linear regression parameters, you will be predicting (and displaying) the probable number of deaths in the state on December 31, 2020 (For Texas, the approximate day number being 325).

**Additional Hints:**

- You will have to save your csv data in an excel file and for reading the excel file the file must be in your MATLAB current folder. Name your excel file as 'CoronaData.xlsx' so that I can run your program in my MATLAB by having my excel file in that name (without having to ask for your excel file).
- You will be using functions like readtable() to read your excel data as table. Table is a special data type in MATLAB; and in different stages of the program, you shall need to use data conversion functions like table2array() to convert your tabular data in column and row vectors or 1-D arrays (Also you may need to use transpose a column vector to get a row vector or vice versa). You will have to explore how you read a single column of a table through appropriate tabular indexing syntax.

- While taking user inputs, using the `inputdlg()` function is encouraged to make your program look like interacting with the user.

**Caution:**

Although way simpler in difficulty level compared to the usual semester projects this course offers, it is still your capstone project for the course and unlike your other home-works and assignments. It demands a holistic problem-solving approach from you, and you shall have to apply a research mindset to explore apply things that you are not necessarily explicitly taught in the class. For evaluation, your submitted .m file (MATLAB script) will be run and in case it does not run with the desired functionality, you shall lose a good part of the available points. You, however, will be evaluated for your logical approaches and efforts; so, in the unlikely case of you don't have a properly functioning program giving yielding the desired output, please still go onto submit your .m file by the due date.

**Encouragement:**

Given what potential and merit you have demonstrated so far, almost the entire class is going to deliver this project in a better way than I (your instructor for the course) did it. The cautionary words are made to make you get started with it right away; the more time you spend with it, the more you learn and the more perfect you get it.