CSC110Y1-F Fall 2020 - Fundamentals of Computer Science 1
Course Project Proposal

Ching Chang Letian Cheng Arkaprava Choudhury Hanrui Fan

November 5, 2020

1. # Part 1

There are 2 main types of citations: text Butler (2020) and Parenthetical (Government., 2020).

command citet Marshall (2020) is the same as Melillo et al. (1993).

References

Butler, R. (2020). Calculating deforestation figures for the amazon. https://rainforests.mongabay.com/amazon/deforestation_calculations.html.

Government., B. (2020). Queimadas. instituto nacional de pesquisas espaciais. http://queimadas.dgi.inpe.br/queimadas/portal-static/situacao-atual/.

Marshall, M. (2020). Planting trees doesn't always help with climate change. https://www.bbc.com/future/article/20200521-planting-trees-doesnt-always-help-with-climate-change.

Melillo, J., McGuire, A., Kicklighter, D., Moore III, B., and Vörösmarty, C. (1993). Global climate change and terrestrial net primary production. 363:234–240. https://www.bbc.com/future/article/20200521-planting-trees-doesnt-always-help-with-climate-change.

Solution. ■

2. Part 2

- Give an overview of any background knowledge necessary for the reader to understand the problem you are studying.
- Provide context for the problem and motivate why you have chosen your research question.
- Your research question should be in *bold*; it should be fairly concise, but can be more than one sentence.

Solution. ■

3. Part 3

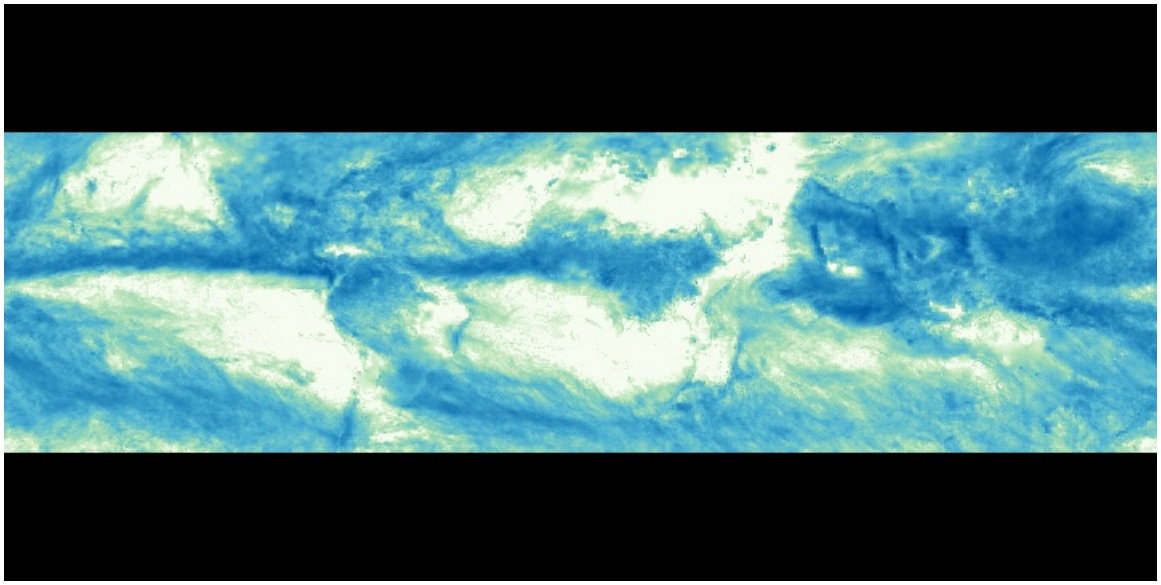
- State the source (e.g., government/organization website) and format (e.g., text, csv, json, image) of the dataset, and give some sample data contained inside that dataset.
- Don't be afraid to cobble together your own dataset, such as creating a collection of images that are related. Or to combine two datasets from different sources.
- You will also submit a small sample of your dataset to MarkUs along with your project proposal document. (See more below)

The data we show here is based on joint estimates by Brazilian National Institute of Space Research and the United Nations Food and Agriculture Organization with map data provided by MapBiomas.
The data is presented as a text-based table, but we will convert it into a csv file. The data shows the area of the Amazon rainforest in Brazil, which accounts for about 60For these data we did some pre-processing, we first removed items in the data that had NA (unrecorded or incomplete data), these included data up to 1985 and 2019, but we needed to record data from pre-1970 as this is the baseline data for the subsequent percentage change.
We then selected data from the start of 1988 because our team considered the absence of the change in deforestation loss data between 1978-1987 to be unrepresentative, so we filtered the 1978-1987 data as well.
At the same time, these data contain four columns including estimated remaining forest cover, average annual forest area loss, percentage of forest cover compared to 1970 area, and total forest area lost since 1970.

- CO2 data. https://ourworldindata.org/grapher/annual-co-emissions-by-region

Entity	Year	Annual CO2 emissions
South America	2010	1059.075401726349
South America	2011	1062.6963727613836
South America	2012	1125.3927407850085
South America	2013	1166.250041134838
South America	2014	1198.3039835470736
South America	2015	1170.890808102911
South America	2016	1134.3849989543753
South America	2017	1125.5205007738507
South America	2018	1112.9340357582964

- precipitation data. https://neo.sci.gsfc.nasa.gov/view.php?datasetId=TRMM_3B43M
  This is the Global rainfall map in August, 2016. We can extract the exact numbers using colour table provided.



here is the colour table: 

- Deforestation data. https://rainforests.mongabay.com/amazon/deforestation_calculations.html

Period	Estimated Natural Forest Cover	Deforestation (INPE)
2015	3,413,662	6,207
2016	3,406,796	7,893
2017	3,399,308	6,947
2018	3,390,835	7,900

4. A computational plan for your project. (300–500 words)

Solution. ■

4. Part 4

- Describe the kinds of computations you plan to perform, such as: data transformation/filtering/aggregation, computational models, and/or algorithms.
- Explain how your program will report the results of your computation in a visual and/or interactive way. You don't need to go into a lot of details here, but it should be clear what you plan to do.

Technical requirement: for your project, you must use at least one Python library/module that we have not covered in this course, *or* use plotly or pygame to a much larger extent than what what have given you so far in this course. (See examples and note in the next section).

- In this part of your proposal, you should also describe one new library you intend to use, how you will use it, and why it is appropriate. Refer to specific functions, data types, and/or capabilities of the library that make it relevant for solving the problem you wish to solve.

Solution.

We first create a function that parses the `html` element of the stats on the website as a string, and converts it to a `nested array` so that it's easier to work with. This will involve using a `for loop`, `if statements`, and an `accumulator` keeping track of the data parsed so far. Using this function, we will collect our data for deforestation in the Amazon rain-forest over the past few decades. With the data now converted into a form that we can easily manipulate, we shall focus on analysing the data using our own functions.

For this project, we use smooth polynomial fitting to relate two of the variables in our `nested list`. Now, although there exist readily available functions that would do the same in the module scipy, we try implementing our own functions for the same, to test our learning from the course.

We split the mathematical algorithm for this problem using top-down design. Firstly, the main function would have two lists, `l_x, l_y`, of same length as input (for the two variables), along with an integer n (where $n \leq$ `len(l_x)`; representing degree of intended polynomial). The function body would have calls to helper functions. Note: this is only a rough outline and the exact technical details may be changed based on the results after testing the functions.

Firstly, we have a function to calculate the perpendicular distance of one point from a given polynomial. As opposed to the naive approach to the problem, we use Newton-Raphson method *repeatedly* to estimate a solution for the derivative of the expression for the difference between the point and the polynomial, hence finding the coordinates of the foot of perpendicular, and consequently, the length of perpendicular.

The first estimate for the polynomial will be trivial, and we will then run the simplex algorithm to minimize the sum of the squares of the perpendiculars using two more

helper functions, yielding the polynomial regression model. We now move towards plotting the resulting graph using the matplotlib library.

We also plan to write a function to calculate the coefficient of determination to check whether the graph shows an appropriate relationship between the independent variable (forest cover) and the dependent variable.

In addition to the graph, we plan to create an interactive text-based report of our data, where the user inputs a value for the independent or dependent variable, and the program will provide the corresponding dependent or independent value, coefficient of determination, or the slope of the tangent at the point, depending on which one the user asks for. The output will be text-based, and will require string concatenation, and if statements to check whether to add trivial information to the report.

The input/output model will use while loops and input prompts to keep the program interactive. We also extrapolate the data to yield the predictions about future data using the interactive i/o model. Finally, we use the extrapolated data to summarize the upcoming significant years where the dependent variable will reach a certain milestone.

■