UCSD Data Science Bootcamp Final Project Proposal, 5/2/20

COVID-19 Machine Learning Analysis

Team Members:

* David Jaimes
* Grant Thompson
* Arundhati Chakraborty
* Alexis Perumal

Assignment

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

1. Your visualization must include a Python Flask–powered RESTful API, HTML/CSS, JavaScript, and at least one database (SQL, MongoDB, SQLite, etc.).
2. Your project should fall into one of the below four tracks:
   * A custom “creative” D3.js project (i.e., a nonstandard graph or chart)
   * A combination of web scraping and Leaflet or Plotly
   * A dashboard page with multiple charts that update from the same data
   * A “thick” server that performs multiple manipulations on data in a database prior to visualization (**must be approved**)
3. Your project should include at least one JS library that we did not cover.
4. Your project must be powered by a data set with at least 100 records.
5. Your project must include some level of user-driven interaction (e.g., menus, dropdowns, textboxes).
6. Your final visualization should ideally include at least three views.

Project Objective

* Visualize current COVID-19 outbreak timeseries data, at the country, state and county level (if possible) to understand the spread of the pandemic.

Project Description/Outline

* The user will see a web dashboard with multiple views and a selector. (See below.)
  + A selector will be available so the user can pick the country of interest, or WW (all countries).
    - Metadata for the selected country will be shown in a simple table.
    - A time series line chart (or area) for the selected country will show a cases curve and a death curve.
  + We will build an animated horizontal bar chart that can show growth over time.
  + World map showing cases and/or deaths. Time permitting we’ll add an animation illustrating change over time, perhaps a regional heat map.
  + We may enable a way to view and compare a few multi-selected countries.
  + Time permitting, we will show drill down below country level: state, and possibly counties where supported by the dataset.
* Technologies:
  + Backend: Python Flask, Mongo DB
  + New JS library: TBD
  + Data Retrieval: Use the Johns Hopkins dataset from a REST endpoint if possible, but we will likely need to pull it from the [Johns Hopkins github repo](https://github.com/CSSEGISandData/COVID-19) and pull it into our MongoDB. No web scraping needed for this project.
    - Time permitting: We’ll automate the data retrieval with a scheduler.
  + Front End: JavaScript, D3, Leaflet.js

Rough Breakdown of Tasks

* Get the JH Dataset with its new schema (just released this week), study it and determine what we will use, and how we will retrieve it. - Alexis
* Design how we will store the dataset in Flask / MongoDB database. – Arun
* Create a top level index.html, css, and JavaScript files. - Grant
* Build the Visualizations, controls, etc. – All
* Setup automated scheduling. – TBD
* Publish the Dashboard
* Using the dashboard, do data exploration, analysis and build a story of the pandemic, key insights and findings, and predictions.
* Build a presentation for the class, **Wed., 4/8/20**
* Marketing: Distribute the link to our page in Social Media to drive hits.
* Final Submittal in github/BCS, **Fri., 4/10/20**

Figure: Wireframe of UI

* + A screenshot of a cell phone

    Description automatically generated

Figure: Style Guide

A screenshot of a cell phone

Description automatically generated