Progress Evaluation - Milestone 6

Title: Customizable Analysis and Visualization Tool for COVID Cases

Team Members:

- Calvin Burns, cburns2017@my.fit.edu (Team Lead)
- Sam Hartle, shartle2017@my.fit.edu
- Stian Olsen, shagboeolsen2017@my.fit.edu
- Nicole Wright, nwright2017@my.fit.edu

Advisor: Dr. Philip Chan, pkc@cs.fit.edu

Client: Dr. Philip Chan, pkc@cs.fit.edu

Progress Matrix for Milestone 6:

Task	Stian	Sam	Nicole	CJ
1) Selecting a variable (column in a dataset) used in the different plots below	-	-	-	100%
2) Plot template for trends over time (line)	50%	-	-	50%
3) Plot template for proportion among categories (pie)	-	-	-	100%
4) Plot template for relationship between possible factors and situations (scatter)	45%	10%	-	45%
5) Plot template for distribution over FL counties (map)	-	-	-	-
Showcase Deliverables and User Manual	10%	30%	30%	30%

Discussion of each accomplished task and obstacles for Milestone 6:

- Task 1: The scope of our project was reduced on March 18th by our client and advisor, Dr. Chan. He recommended that we focus on producing 4 main plot types and reduce the customizability of our product. For all 4 of the plots, the user needs to select what variables they would like to work with. Our initial design of this “card” allowed the user to choose multiple datasets in a multi-select dropdown interface. Below the dataset selection was a variable selection multi-select dropdown. This was dynamically updated based on what datasets were selected in the previous dropdown.

  After meeting with our client on March 29th, Dr. Chan suggested several things. First, we adjust the UI to have the user select dataset/variable pairs. In addition, we now limit the number of variables a user can select based on what plot type is selected. This UI was difficult to build because the data that goes in the dropdowns is dynamic and conditional. We were able to successfully implement the dataset/variables pairs user experience.

  The final suggestion that Dr. Chan made was for the user to be able to search all available variables. We were unable to implement this feature for several reasons. Several attempts were made at the *quick solutions* but we quickly found that some type of caching would be necessary as the computational expense of opening every dataset and saving the variable names was large.

- Task 2: The first plot that Dr. Chan suggested we focus on was the “trends over time” plot, a.k.a. line chart. To implement this feature, we built a new UI/card specific to line charts. The card has start and end date fields which are dynamically filled with the max/min values of the joined datasets the user has selected. Next, the user can select the sampling frequency. Lastly, text input fields dynamically appear for each variable the user has selected and takes input for the names the user wants to appear on the rendered line chart. All parts of this task were successfully implemented.

- Task 3: The second plot we worked on was “proportions among categories”, a.k.a. pie charts. Similar to the line chart, we created a unique UI/card for inputting the necessary fields for a pie chart. The start/end date are dynamically based on the joined datasets. Then the user selects what variable to use as categories followed by what variable to use to aggregate. All elements of this task were completed.

- Task 4: The third chart we worked on was “relationship between possible factors”, a.k.a. scatter plots. Similar to the previous two plots, start and end date fields which are dynamically filled with the max/min values of the joined datasets the user has selected. The user then selects the possible factor and situation from the selected variables. All elements of this task were completed.

- Task 5: The final plot we worked on was a “distribution over counties” plot. Specifically, this would plot the distribution of a variable over Florida counties. We were able to generate the tables for this plot but ran into an issue with our datasets. A custom dataset would be needed or a future interface for associating location codes to county names.

Discussion of contribution of each team member to Milestone 6:

- Stian: Modified the plot utility so it would work with the new scope of the project. Coordinated with CJ to make sure the plot utility would work on the possible inputs.
- CJ: Created wireframes for select dataset and variables card, line card, pie card, scatter card, and map card from Dr. Chan’s suggestions in his “*reduce scope”* email. Created an initial version of the UI that outputted a table for the March 29th meeting.

  After suggestions from Dr. Chan, reworked the UI with wording changes. Then reworked the “select datasets and variables” card to use dataset/variable pairs. Then built the line, pie, scatter, and map cards for user input. Developed and created custom database models for saving the unique information for each subclass of a Plot. Worked on the backend to save the data from the create plot form. Wrote get_dataframe methods for each subclass of plot to reproduce the necessary dataframe from the selected datasets/variables. Coordinated with Stian to fix bugs and update the plot render utility.
- Sam: Did majority of the work on finalizing showcase deliverables (poster, ebook, demo video, and banner image). In addition, outlined the format for and worked with Nicole to complete the user manual.
- Nicole: Assisted Sam, Stian, and CJ with project finalizations. Helped work on and edit showcase deliverables and wrote the user manual.

Lessons Learned:

- "*Project Scope"* - During the fall semester, we proposed an ambitious project plan. After a few milestone iterations, it became clear that the task for creating plots was much more complicated and involved than we had anticipated. A large number of our features depended on the completion of the Create Plot page. Numerous design iterations were made on the Create Plot page and we still don't have a UI that is exactly what we originally wanted. We should have descoped earlier on when we realized the Create Plot page wasn't a one or two milestone task.
- *"Target Audience"* - A closer (and earlier) identification of what kind of user would be interested in using our application would have aided in the development process. This would have allowed us to tailor development to be more or less intuitive depending on the type of user.

Date(s) of meeting(s) with Client/Advisor (same) during Milestone 6:

- March 29th
- April 12th

Client feedback on the current milestone: See Faculty Advisor Feedback below

Faculty Advisor feedback for Milestone 6:

- Notes from 3/29/2021
    1) Create Plot Page
        a) X/Y can say horizontal/vertical in layman's terms
        b) Use similar wording to email notes from 3/17 for descriptions
        c) For scatter plot, two variables should be selected
            i) Make help note explaining this in the UI
        d) Wording/Vocabulary
            i) Select Dataset *and Variable* (additional wording)
            ii) *Select Plot* instead of Plot Information
        e) Select Dataset and then one variable
            i) Repeat for each dataset selected by user
            ii) Picking one pair at a time
            iii) Would like to limit selection of number of variables to 2 for certain plots
                (1) Pie chart
                (2) Scatter plot
            iv) Search for a variable without knowing which dataset it belongs to

(1) Lower priority

2) Move table to end of selection before the end of project
    a) For DEBUG purposes
3) Between the UI and plotting, please discuss and decide the table (or dataframe, ...) format to be passed from the UI to each of the 4 plotting modules
    a) For example, for "trends over time" (line plot):
        i) column 1 is date (or week or month)
        ii) column 2 is the first trend variable (a value could be NA),
        iii) column 3 (if it exists) is the second trend variable
4) Once this is decided/standardized, integration of the UI and the plotting modules would be easy
5) If the x-axis is too crowded with labels, you can use fewer labels, every 7 (14 ...) days. Maybe at most 20? labels on the x-axis, and you can calculate the interval between each label.
6) The video is due on Apr 7 for Showcase, I suggest you to aim for a scenario/demo that can be achieved by a user from beginning to end (uploading datasets, creating a plot, saving a plot). The "trends over time" plot template probably is more important than the other 3, so I suggest at least making "trends over time" work by Apr 7.

- Notes for 4/12/2021
    1) X axis should be less crowded
    2) If range is wide apart, use 2nd y-axis to utilize entire vertical axis
    3) Try to increase diagram size to reduce crowding on x-axis and utilize entire screen space for the plot
    4) Automatic redirect from successful plot creation page to the plot display itself
    5) Add ability to rename axes on scatter plots
    6) Do not need multiple colors on scatter plots unless they have specific meanings
    7) Wording for Laymans terms for scatter plots
        a) After possible factor (horizontal axis)
        b) After situations (vertical axis)
    8) Axis labels for all plots
        a) Specifically scatter plot as of right now
    9) Translate county names to county codes used by the mapping library, preferably in the plotting module right before calling the mapping library
    10) Add line/pie/scatter/map to each technical term
    11) Limit number of variables to 2 for trends over time