Customizable Analysis and Visualization Tool for COVID Cases

Milestone 3

Team Members

► Calvin Burns, cburns2017@my.fit.edu (Team Lead)

► Sam Hartle, shartle2017@my.fit.edu

► Nicole Wright, nwright2017@my.fit.edu

► Stian Olsen, shagboeolsen2017@my.fit.edu

Faculty Advisor/Client

- Dr. Philip Chan, pkc@cs.fit.edu

Progress Matrix

Task	Completion %	Stian	Sam	Nicole	CJ	To do
1. Continue Feature 4.1 (Customizable Operations on Variables)	60%	30%	30%			Fixing UI details and interacting with more backend operations
2. Small GUI demo that integrates lab testing and mask mandate data	80%		40%	10%	30%	Add the mask mandate data to the plot
3. Consider different options for saving plots	100%	30%			70%	None

Task 1 - Continue Feature 4.1 (Customizable Operations on Variables)

- Model complete for an operation

- Utilizing SymPyCharField which allows the user to input any operation

- Can create a plot based on public or private datasets

  - Filter data based on columns, variables and operations

  - Visualize the data filtered data by selecting plot type

To do: Fixing UI details and interacting with more backend operations

# Task 2 Summary - Integrating lab testing and mask mandate data

- Decided on charting software (Highcharts)

- Created a csv importer which is able to import any csv file

- Backend functions needed to format the data in the data frames are completed

  To do: Add the mask mandate data to the plot

Task 3 - Consider different options for saving plots

- Dealing with large amounts of data

- Currently storing the data in a pickle file

    - Store the data as binary

    - Python way to serialize data

    - We don't need to worry about running out of memory

- New idea: Storing the data frames based on user sessions. Store the data in a cache and keep it there while the user is active. Data not used will be taken away from the cache after a certain time. The idea is to lower the time it takes to load data

Demo - Generic Dataset CSV Uploader, Public/Private Data Pages, View Dataset
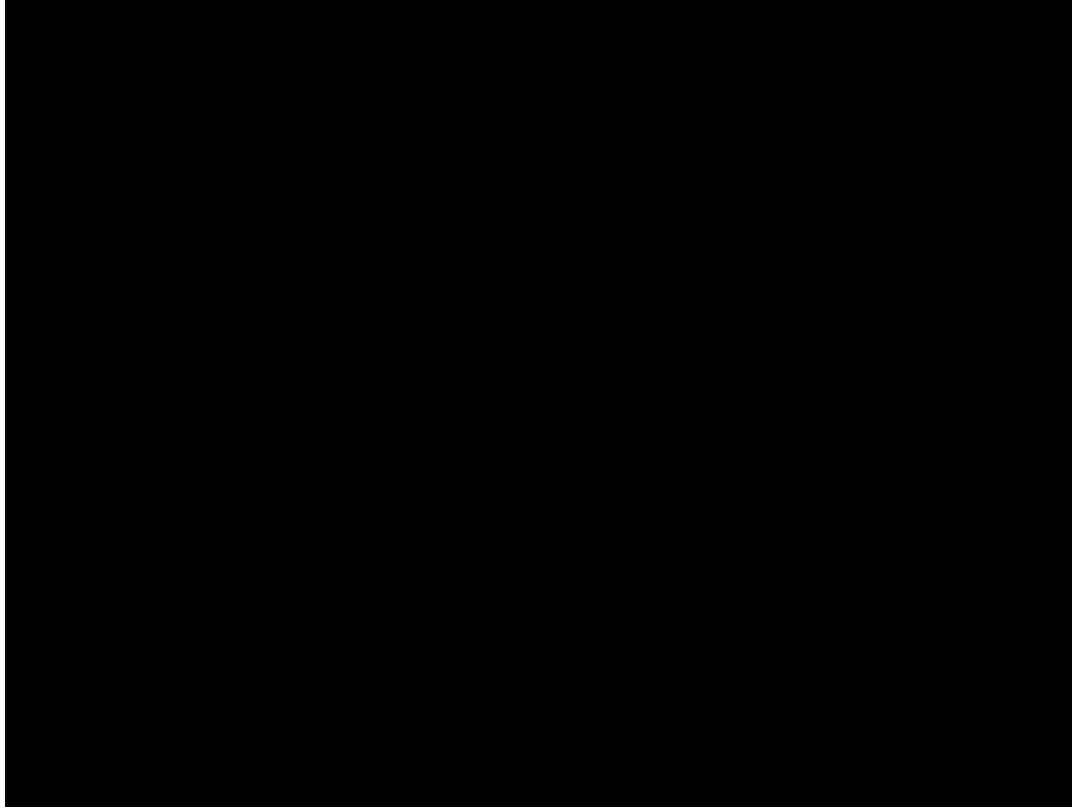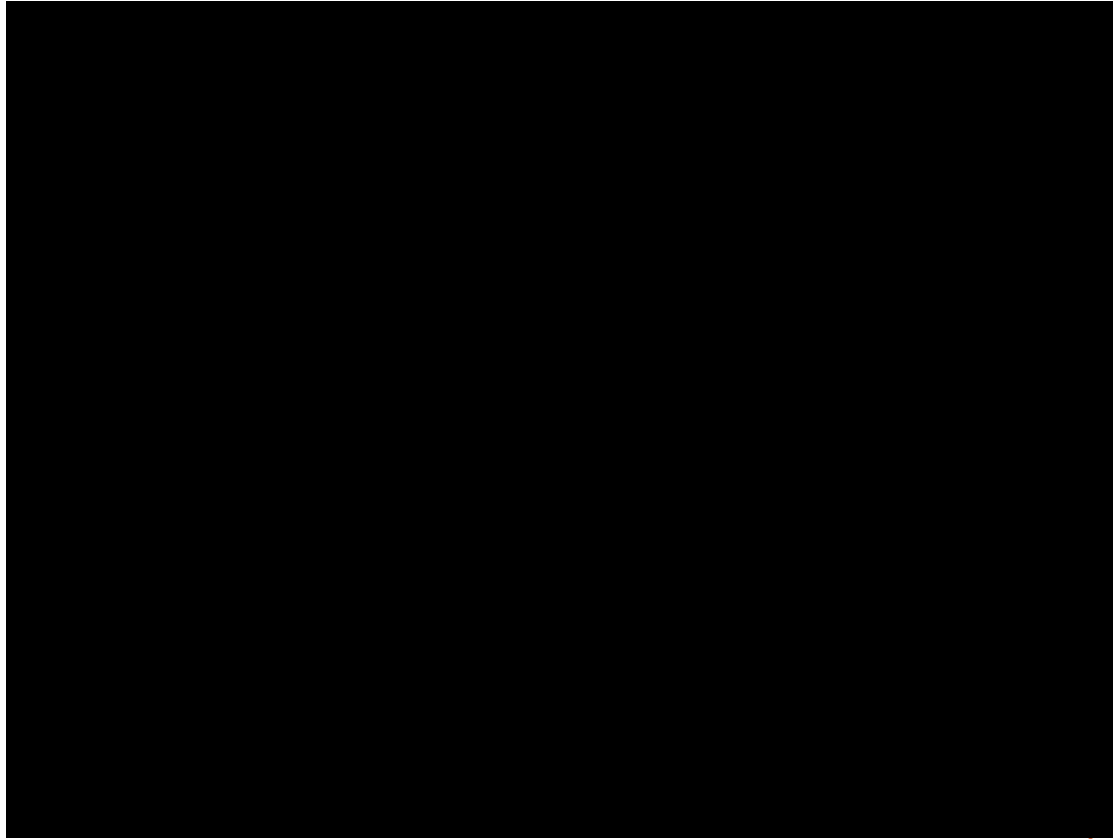
Conversion and Compression of Datasets

1. Open CSV as a Pandas DataFrame.
2. Convert given "Date Fields" to datetime objects.
3. Convert obvious data types. (i.e. strings with only numbers get converted to integers)
4. Anything left as a string gets converted to Categories
   a. Only store string value once, replace string values in table with keys to respective string value. (i.e. Gender categories = {1: "Male", 2: "Female", 3: "Unknown"})
5. Export compressed Dataframe as a Pickle file and save on disk.
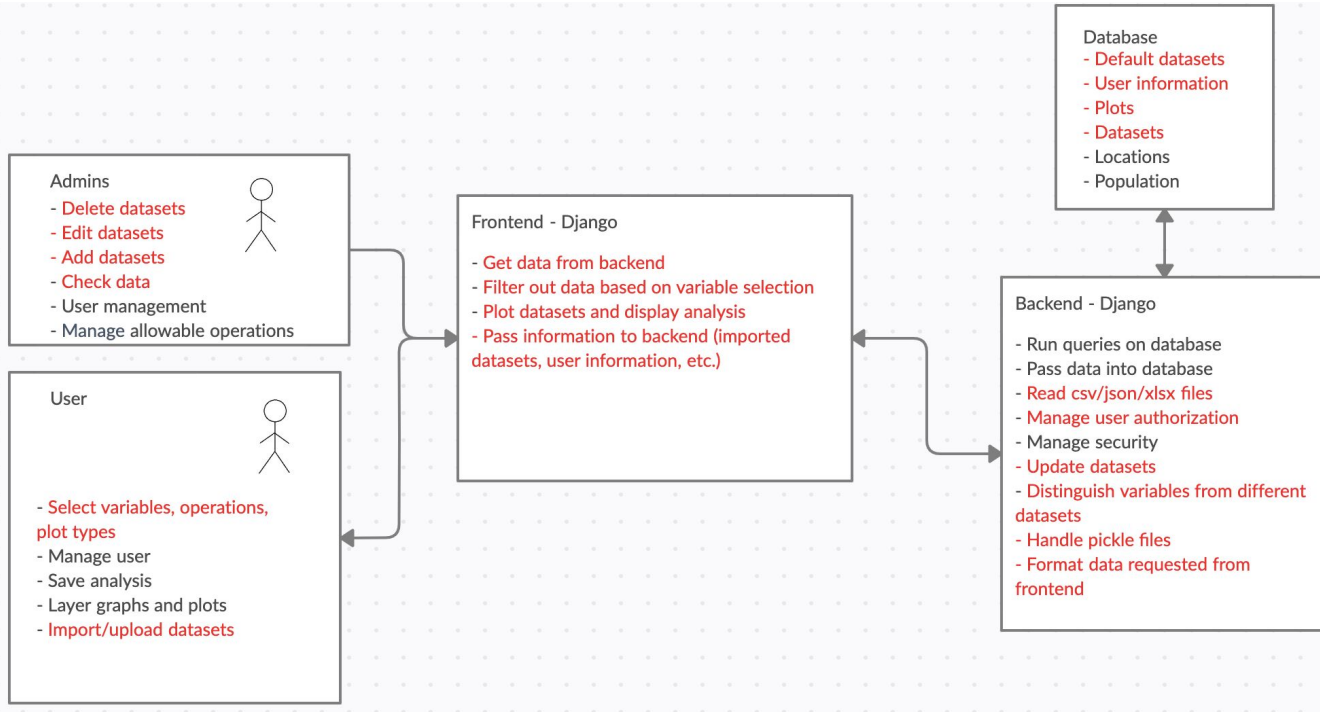
Results:

- Original CSV: ~110 Mb
- Raw Dataframe(before compression): ~600 Mb
- Final: ~40 Mb

Demo - Create and View Plot

Design - System Architecture

Database
- Default datasets
- User information
- Plots
- Datasets
- Locations
- Population

Admins
- Delete datasets
- Edit datasets
- Add datasets
- Check data
- User management
- Manage allowable operations

Frontend - Django
- Get data from backend
- Filter out data based on variable selection
- Plot datasets and display analysis
- Pass information to backend (imported datasets, user information, etc.)

Backend - Django
- Run queries on database
- Pass data into database
- Read csv/json/xlsx files
- Manage user authorization
- Manage security
- Update datasets
- Distinguish variables from different datasets
- Handle pickle files
- Format data requested from frontend

User
- Select variables, operations, plot types
- Manage user
- Save analysis
- Layer graphs and plots
- Import/upload datasets

Task Matrix for Milestone 4

Task	Stian	Sam	Nicole	CJ
1. Continue work on scatter plot	Example for Dr. Chan on why we would use a scatter plot for our application Finish development on plot type (85%)	Assist as needed (5%)	Assist as needed (5%)	Assist as needed (5%)
2. Continue work on male/female pie chart	Assist with different chart types (10%)	Assist with various operations as needed (10%)	Add additional "cards" to Create Plot interface which shows general plot details (20%)	Work on Create Plot GUI and using specified filters/operations instead having to "program" (60%)
3. Continue work on creating lab testing plot	Assist with different chart types and operations research (10%)	Plot positivity rate vs. date Focus on operations: count, sum, division (60%)	Assist as needed (5%)	Ideally this will be able to done through GUI as well Continue to build Create Plot GUI abstractly (25%)

Questions?