Field Statistics (FS3) Plugin

Colorado School of Mines Field Session
Newmont Exploration Team

Orden Aitchedji, McKenna Duzac, Andreas Foulk and Tanner Lee
Client: Brian Krzys
18 June 2018

THIS PAGE LEFT BLANK INTENTIONALLY

Contents

Introduction

Client Description

Newmont Mining Corporation is primarily a resource extraction company. The first step they take for any project is to locate the sites which they wish to work. These can be anywhere in the world and as such, careful consideration of many variables is taken when picking a site. Climate, road access, resource quality and quantity, and land ownership are just a few of these. In order to efficiently look at these variables worldwide, they leverage Geospatial Information System (GIS) software and heaps of relevant geospatial data. A popular method for observing and interacting with GIS data is through the use of an open-source application QGIS.

Product Vision

Our goal was to produce a plugin for QGIS, an open source GIS program, that can quickly view and analyze any kind of data. By looking at what was missing from other plugins and the previous version of this plugin created by our client Brian, we were able to determine what functionality our plugin should have. We came up with four primary functionalities: attribute visualization and editing, statistics calculations, determining uniquenesses, and producing graphs. These were decided to be the best way to get a quick and comprehensive overview of any kind of data.

The plugin, while being developed for Brian and Newmont, was intended to reach a far greater audience. Its design was intended to be user friendly, fast, efficient, and applicable to a variety of situations. All QGIS users with an interest in visualization and statistical analysis should be able to immediately access and leverage the plugin for their purposes.

Requirements

Functional Specifications

- Run statistical analysis on user-selected data set
    - Analytics dependent on field type (numeric or character)
    - Examples: median, max, min, standard deviation, sum, range, field length
- Allow for multiple graph types (pie chart, scatter plot, histogram, mosaic, etc.)
- Allow for quick navigation between data set's fields
- Indicate what type of data is in a field (character, numeric, etc)
- Ability to export generated graphs
- Support multi-column analysis and comparison
- Data should be able to be log transformed

Non-Functional Specifications

- Target the QGIS 3.X versions
- Utilize QTDesigner/QTCreator for GUI
- Coded in Python 3.6 using Spyder IDE
- Support at least two languages (English and Spanish)
- Follow QGIS coding standards
- Public Github for version control
- Fully functional on all operating systems (Linux, Windows, MacOS)

System Architecture

The following section will detail components of Figure 3.1.

FS3 Interface

Represents the user interactable interface of the plugin. This portion of the architecture was specifically designed and implemented to be an intuitive and efficient methods for the user to access the powerful features of FS3's internals.

FS3 Internals (Analysis)

Represents the first portion of FS3's internal structure. This corresponds to the statistical and analytical calculations that are generated by the plugin. Contains the necessary logic for handling all calculations displayed in the "Statistics" and "Unique" tabs of the FS3 interface.

FS3 Internals (Visuals)

Represents the second portion of FS3's internal structure. This corresponds to the visualization tools that are built in to the plugin. Contains the necessary logic for generating and linking HTML graphs for display within a web-view on the FS3 interface.

Data

Represents the data dependency from the user. This portion of the architecture refers to the necessity from the plugin's user to present data to QGIS. This data may come from various different sources (company database, external storage, online repositories, ect.), but will ultimately be brought onto the user's machine and ported into QGIS.

QGIS

Represents the parent program dependency. This portion of the architecture refers to the parent program of the FS3 plugin. The plugin links to the interface and internals of QGIS to generate statistics and visualizations for any loaded data.

Archive

Represents an optional storage dependency. This portion of the architecture refers to the ability for a user to export visualizations (graphs and charts) to their machine's storage. Visualizations can be stored directly as a PNG image format, or view/modified through a web browser as an HTML file.
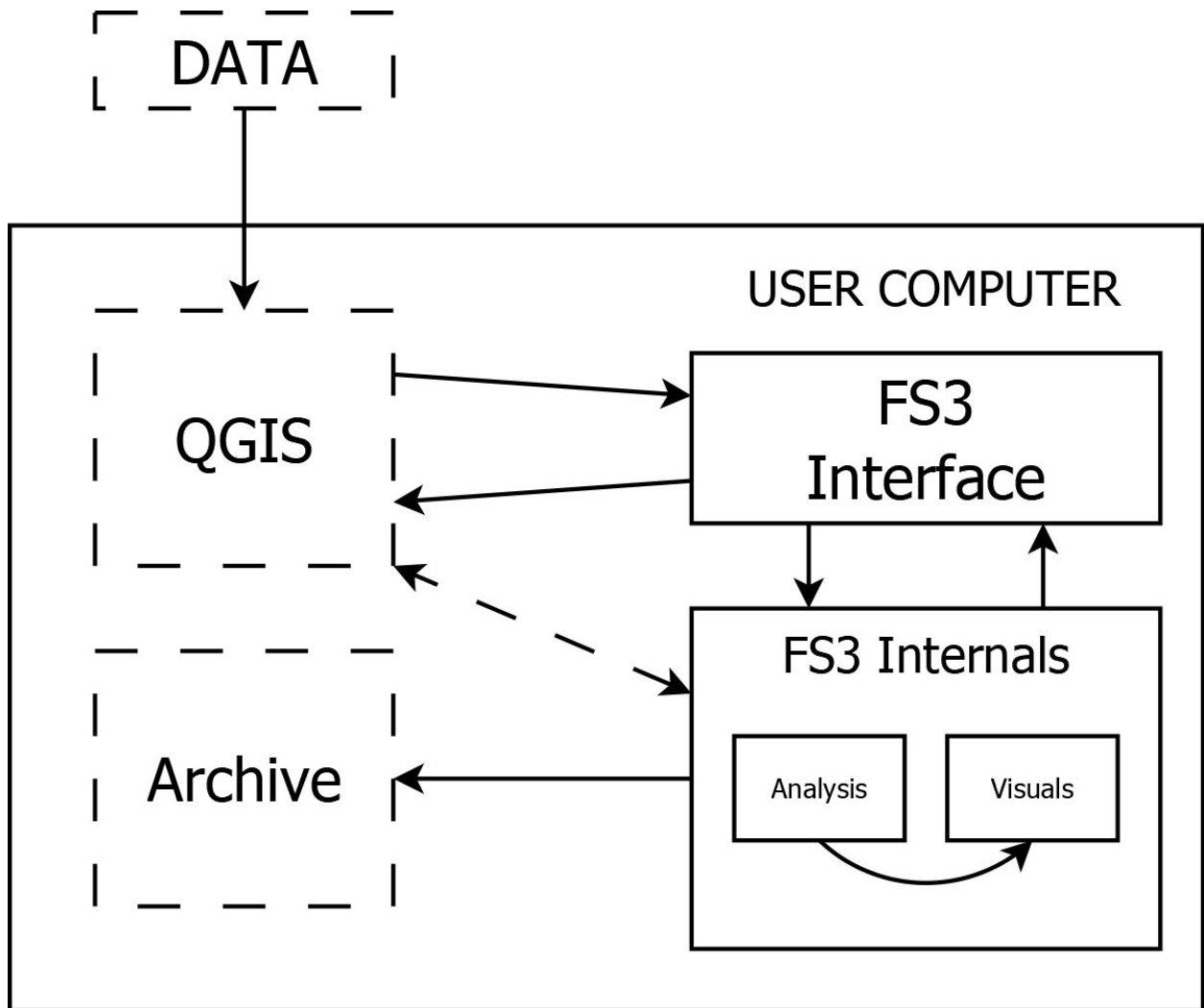
Figure 3.1: System Architecture Overview

Technical Design

Javascript, HTML, and Plotly

The FS3 plugin contains a graph tab that handles the graphing of selected data fields. Initially, the python library 'matplotlib' was explored for the implementation of graphing but while its functionality was sufficient, it lacked many features present in other visualization libraries. As such, an alternative library 'Plotly', an online collaborative data analysis and graphing tool, was used instead.

The Python API allows you to access all of Plotly's functionality from Python. The Plotly figures can then be shared, tracked, and edited all online while the data provided can also be accessible from the graph. Plotly provides an exceptional toolbar, specific to each types of graphs. Figure 4.1.1 shows useful toolbar options for a line graph. Other graph type like a 'Pie Graph' will have less options in the tool bar, simply due to them being unnecessary in said context.

Figure 4.1.1: Plotly's Toolbar

Aside from the provided toolbar feature, the graph tab for FS3 makes use of an additional graph option button, designed to help users manipulate and personalize the data utilized by plotly. Figure 4.1.2 below shows the graph option dialog box. The user has the ability to input a path within the graph settings, allowing them to export a screenshot of the current graph with the use of the 'Export to PNG' button. Additionally, users can utilize the 'Open as HTML' button to open the same graph with their default web browser. Internet access is not required as the HTML file is temporarily generated locally.

Figure 4.1.2: FS3's Graph Options

Users who have invested in the expanded version of Plotly can utilize the extensive additional features packaged with it when the graph is exported in HTML.

Field Selection List

The initial design and implementation of the FS3 plugin included a single combo box used to select a single field for analysis. This was used to signify which column of the data (provided by the user) should be passed to the plugin and operated on (clearly illustrated in Figure 4.2.1 at the start of the next page). While functional, the feature was limiting in its capabilities. As such, a more robust design was created for the final plugin.

Rather than using a single selection box, the current version of FS3 leverages a QListWidget from PyQt5. The current implementation of the QListWidget can be seen in Figure 4.2.2. This allows the user to select multiple fields from the automatically supplied list of all available fields. Programmatically, this allows the internals to fetch a list of all currently selected fields. This list can then be iterated through to generate and display unique data for each field simultaneously.

Additionally, this meant adding a new capability to the Unique tab, and the calculations it was running for the selected field(s). Initially, when selecting multiple fields, the plugin would simply generate occurrences (the number and percentage of times a unique value appears in the attributes table) for each individual field. This would then populate multiple separate tables in the Unique tab. Again, while functional, the displayed information wasn't particularly insightful. As such, the Unique tab and its logic received an overhaul with the implementation of the multiple field selection capability.

The current implementation of the Unique tab within FS3 now generates unique occurrences for combined sets of data. This means if multiple fields are selected, their values are viewed as a single list. This single list is then stored with all other lists generated from the data and run through the unique tab's processing. The result is a list of unique lists and their respective unique occurance rates. The display of this process can be seen in Figure 4.2.3.

Figure 4.2.1: Original Field Selection (Single Selection Box)

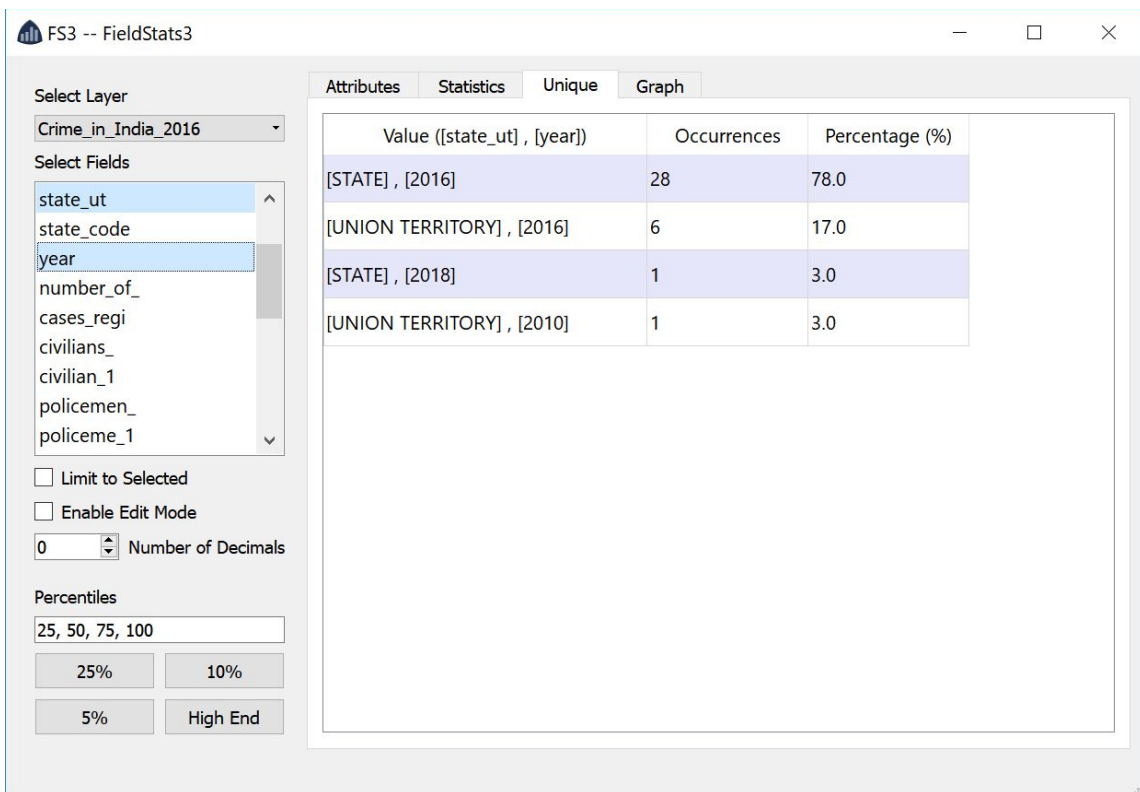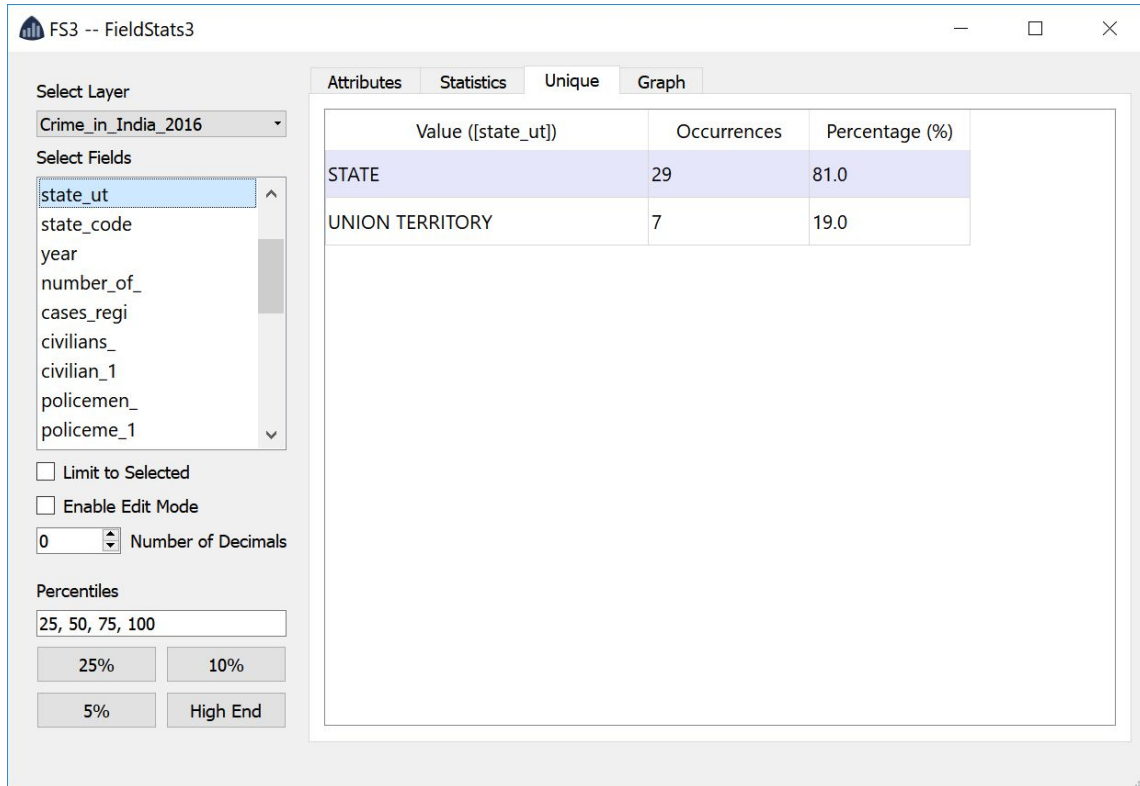*Figure 4.2.2: Updated Field Selection (Multiple Selection List)*

Figure 4.2.3: Unique Values with Multiple Fields

Design Decisions

Language: Python

Python was chosen for our plugin because our client advised us to utilize it. Brian believed python would be the best choice for creating the application because developing a project of this scope in python is much simpler than in C++. Furthermore, QGIS also suggests writing plugins in python as it can easily utilize both the Qt5 library and the QGIS API.

Statistics

The statistics generated by our plugin include: minimum value, maximum value, mean value, median value, number/percent filled, number/percent empty, sum value, standard deviation, coefficient of variation, and user-selected percentiles. These statistics were chosen because our client's original plugin calculated them.

Unique with Multiple Fields Selected

When the user has selected two or more fields for comparison, the fields column will display each data type in brackets with commas in between (Example: [Rome] , [1600]). This allows the data in QGIS to be observed similarly to data in a SQL database.

Graph Types

Users will be given the following graph options: pie chart, histogram, scatterplot, and line graphs. These were the options given in the original plugin designed by our client, so we are moving forwards with these.

Pylint

Pylint was chosen for our code style checker because it was highly customizable and has a large amount of documentation.

QT Design

Instead of hand-writing and debugging our user interfaces, we used QT Design to speed up the process and allow for easier manipulation.

QT Linguistic

QT Linguistic was used to detect the test in our user interfaces and assist in creating the translation files. We used this tool as the translation files created by hand did not properly connect to our application.

Results

Unfinished Features

Our team successfully completed all client required features of our project and used the remaining time to add additional features. These additional features were requested by the client as "It would be cool if you have time", but were not strictly required. Among these features were multi-field support, empty cell handling, QGIS data editing capabilities, HTML export options, and automatic interface refreshing.

Summary of Testing

The parent program that our plugin is associated with works on all operating systems. Our team successfully developed and tested the FS3 plugin on four operating systems (Mac OS, Windows, Arch Linux, Ubuntu). We also had an extensive library of unit tests that were used to verify all calculations made by the Statistics and Unique tabs. These tests ensured that all calculations for things like Standard Deviation, Median, and Uniqueness were functioning correctly.

Results of Usability Tests

Our plugin has been released to the public QGIS plugin repository. Our project is now available to all QGIS users, and as of the completion of this document has been downloaded 88 times and has seven five-star reviews. In addition to a feature request template, we have an issue template on our Github for users of the plugin to report bugs they encounter while using the software.

Future Work

Even though our project is fully functioning, we would like to add these additional features:
- Additional language support
- Additional graph types and options
- An export feature that exports a full report of a given layer

Lessons Learned

- Python is a powerful programming language, but can be hard to maintain for larger projects.
- The importance of maintaining a clean code style and repository. Pylint is a useful tool to help ensure that the project's codebase periodically receives this necessary refactoring/maintenance.
- Reading resource/library documentation is a crucial part of any interesting project.
- Assistant programs and scripts are not crutches, but helpful tools.

Appendices

Appendix A: Instructions for Downloading

Downloading the plugin is extremely easy, as it is now hosted on QGIS's official plugin repository. To download the plugin on any version of QGIS 3, simply access the repository within the application and navigate to FS3. Then the plugin can be added to QGIS by simply clicking the install button.
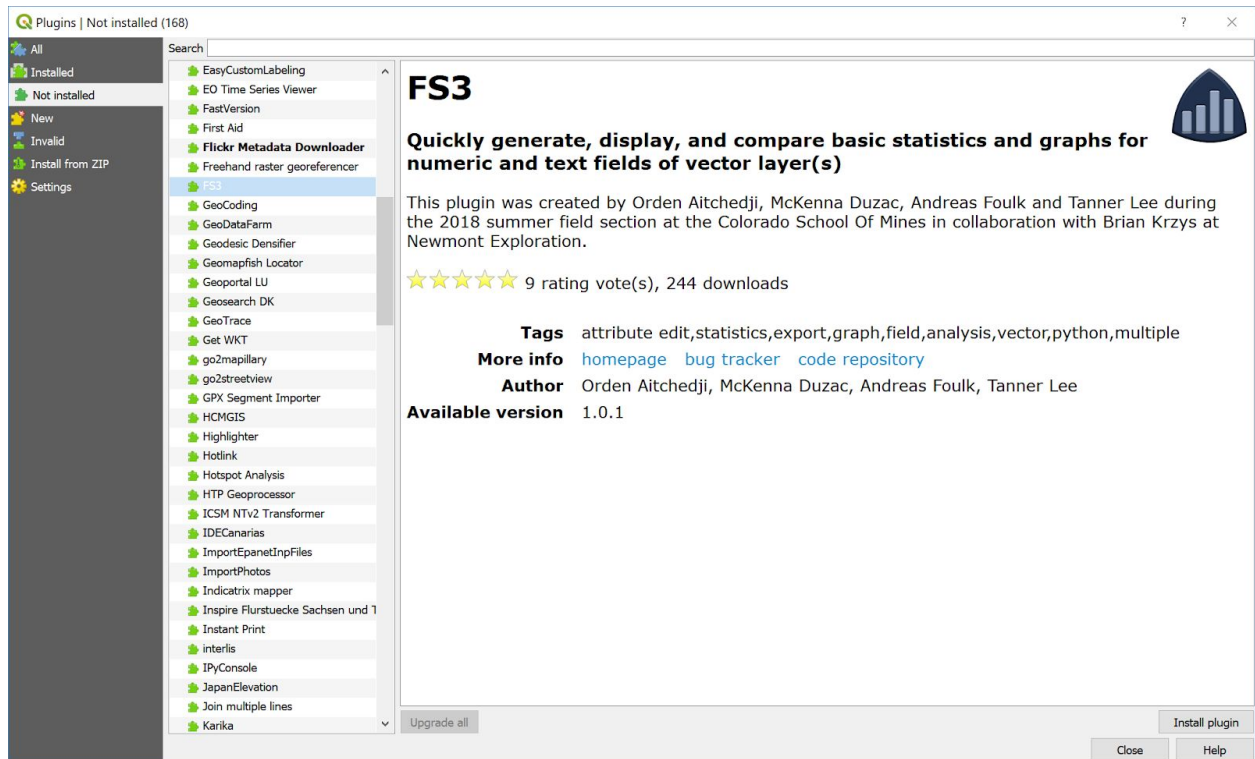

Figure A.1: The FS3 page in the QGIS Plugin Repository

Appendix B: Examples

The images that follow are examples of data analysis and visualization that can be created within the FS3 plugin.
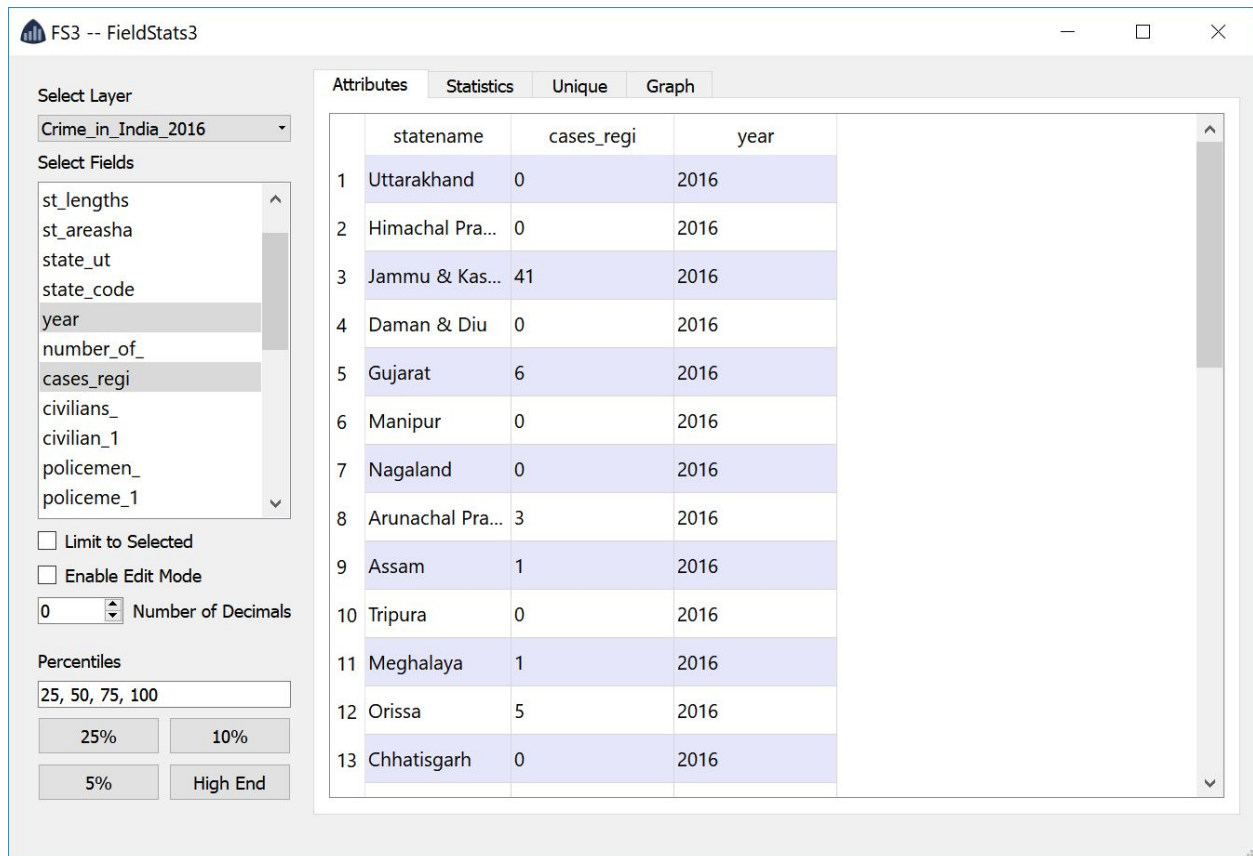
Figure B.1: Indian States with their number of registered crimes
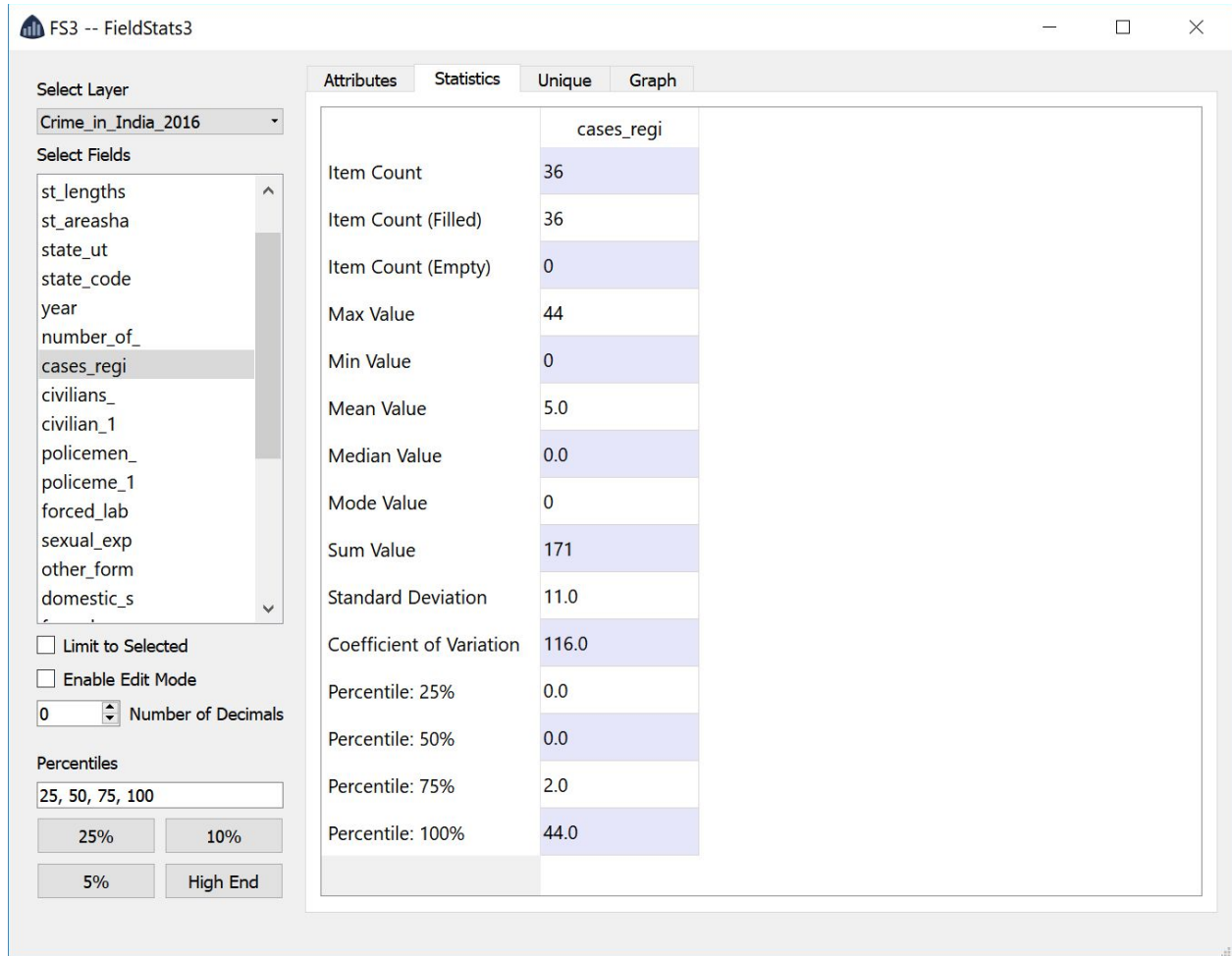
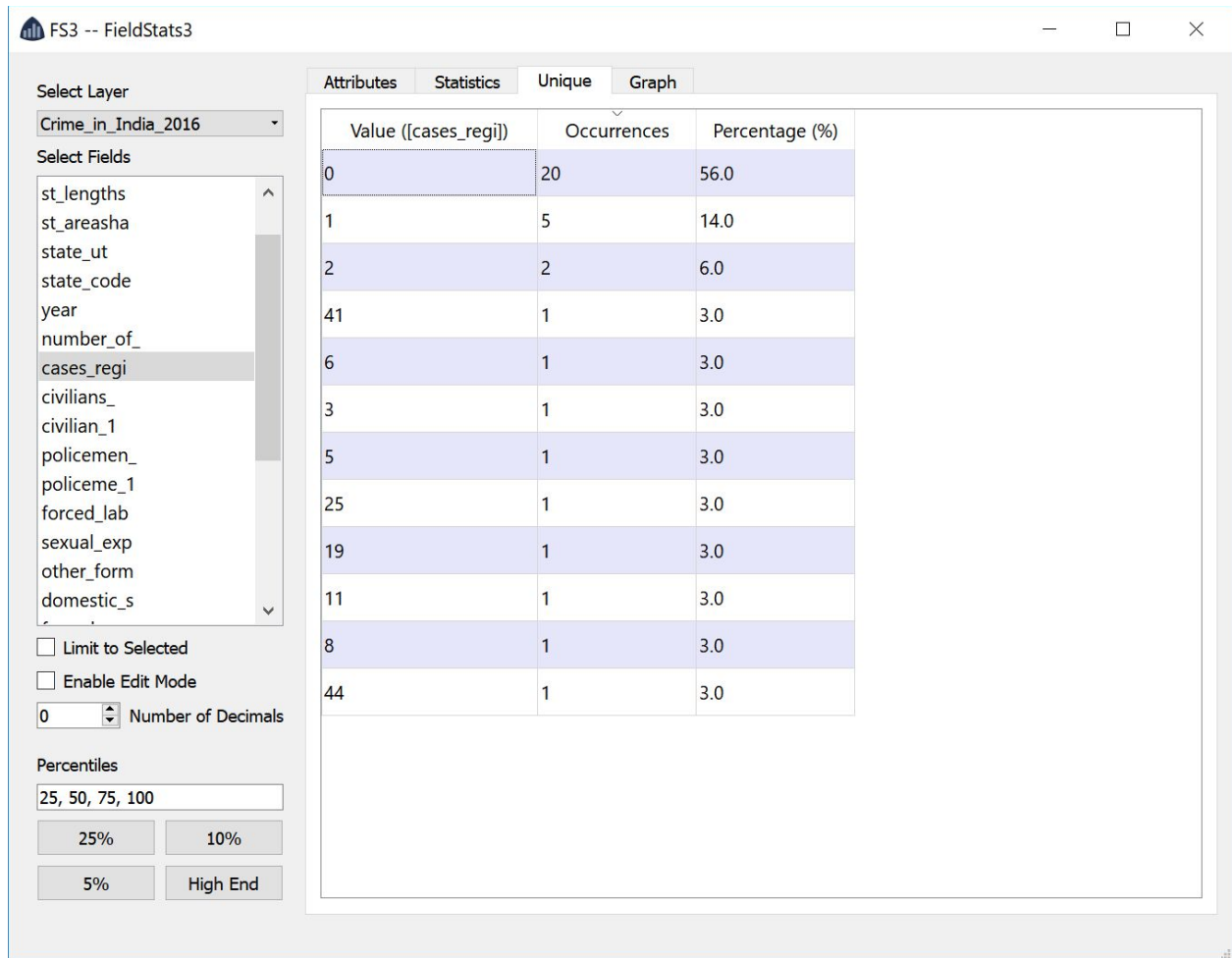*Figure B.2: Statistics on the number of registered crimes in India*

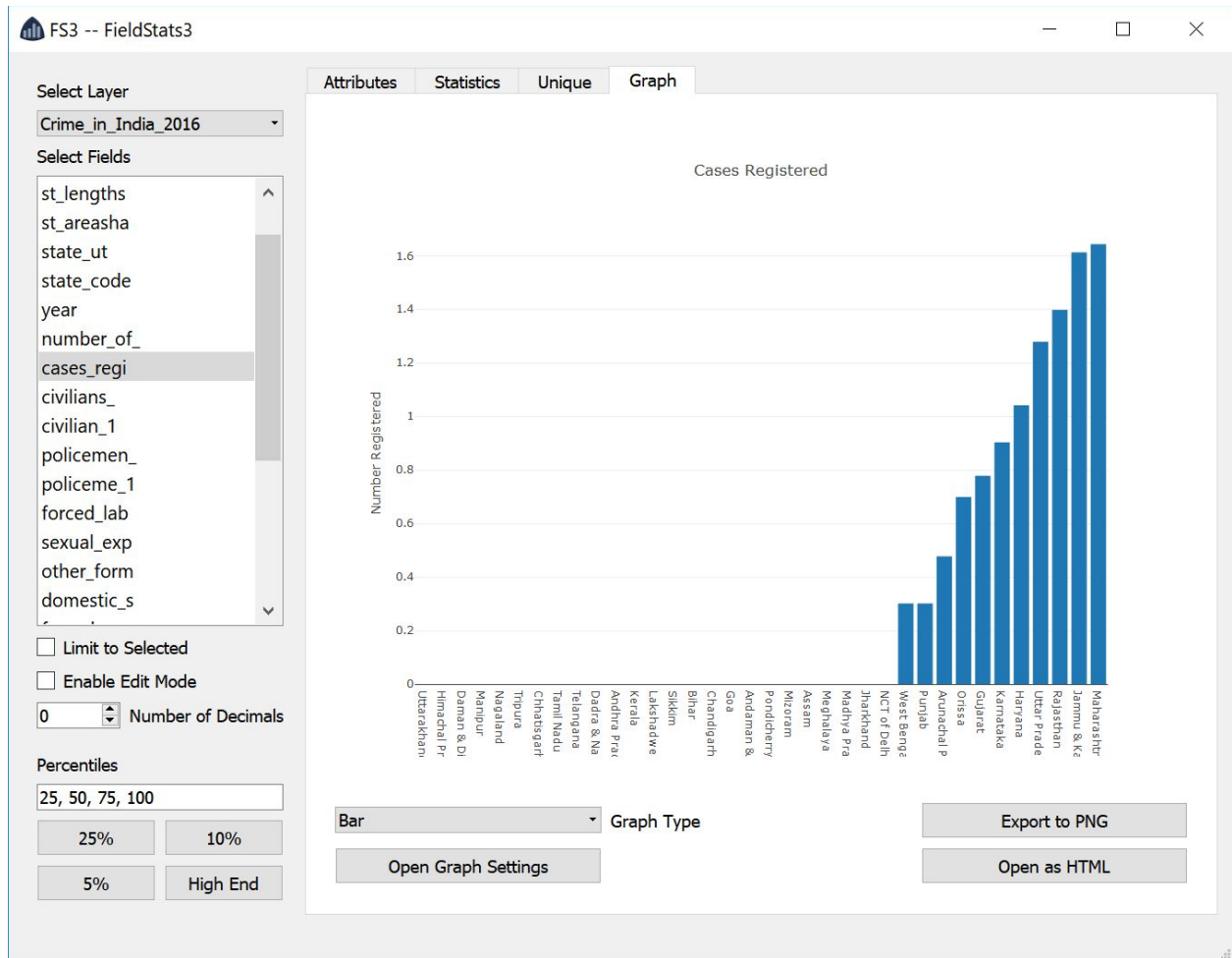*Figure B.3: Unique occurrences for number of registered crimes in India*

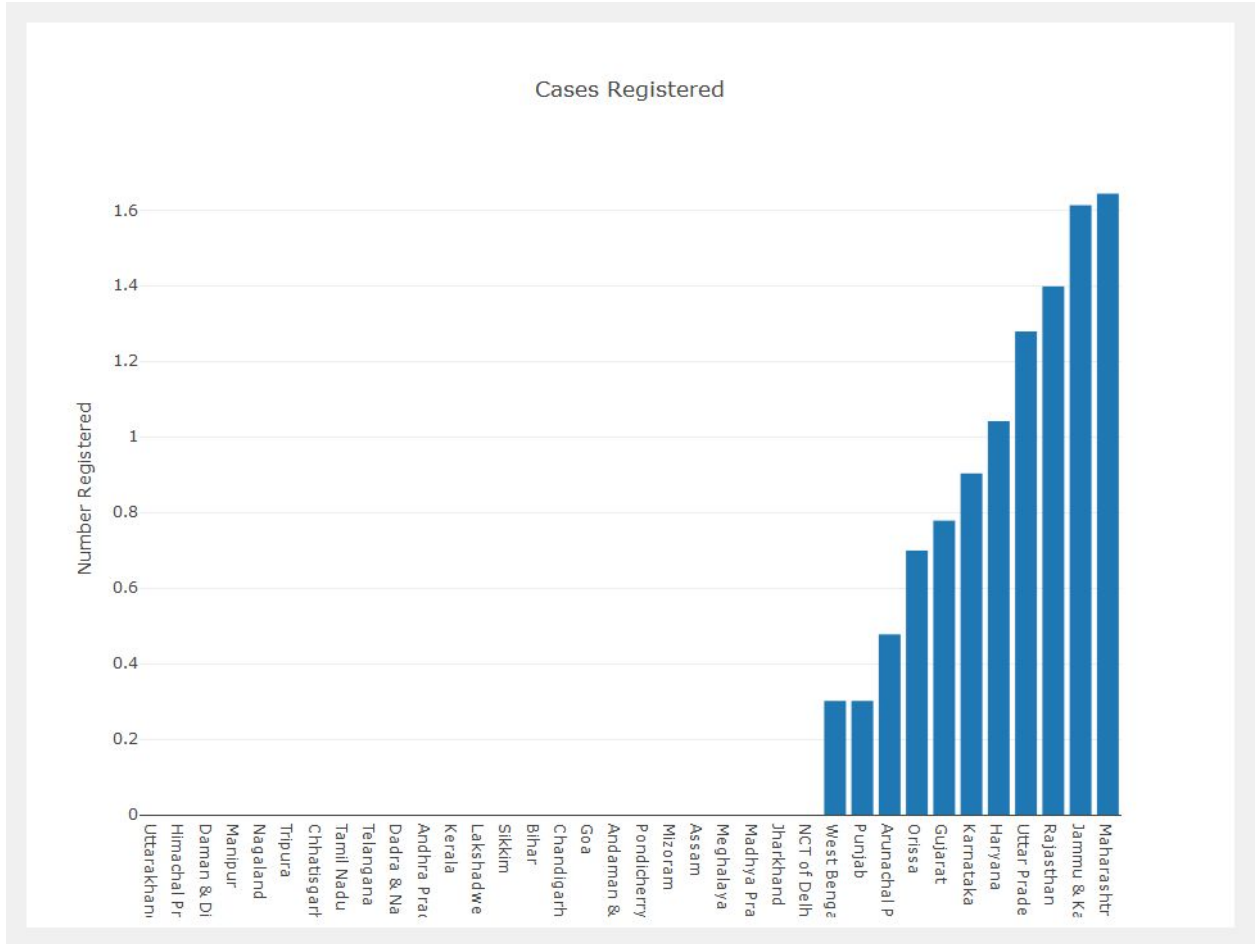*Figure B.4: Graph for cases registered with their State's names*

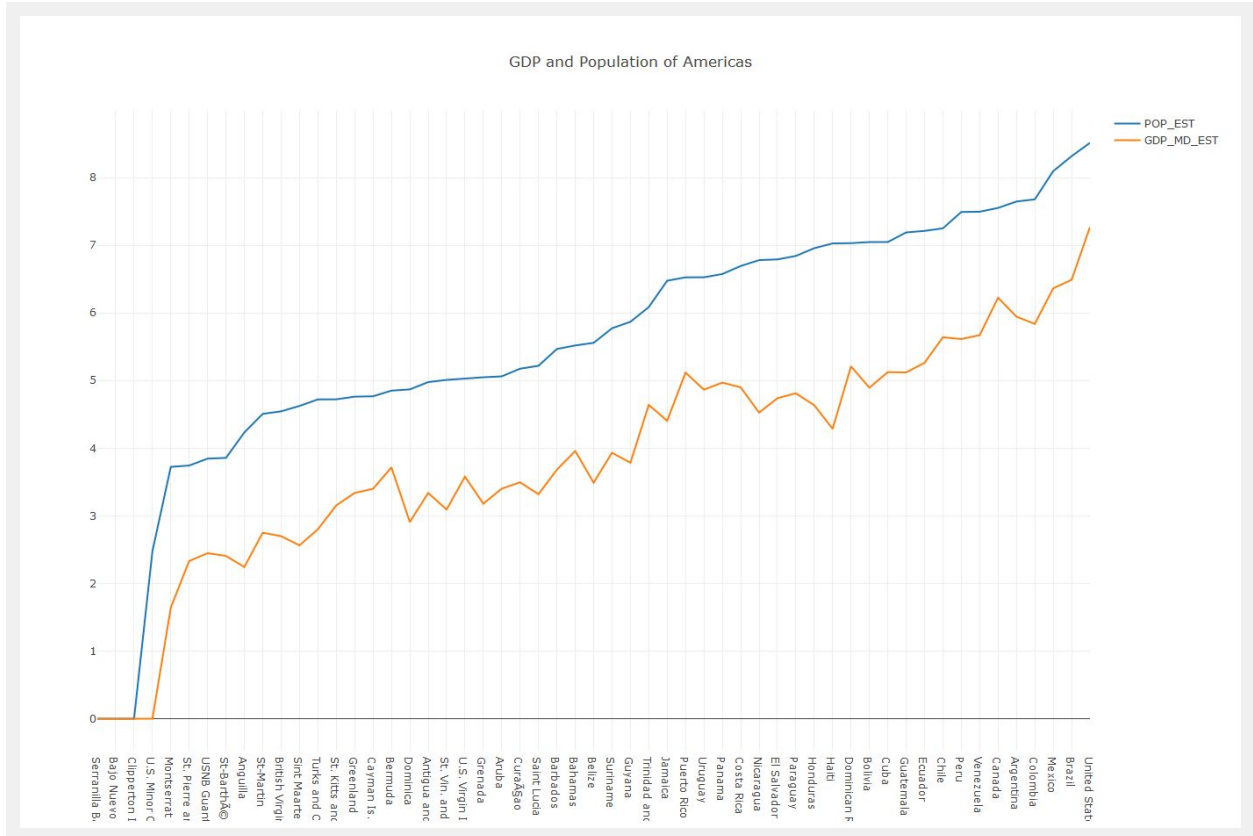*Figure B.5: Exported version of Figure B.4's graph*

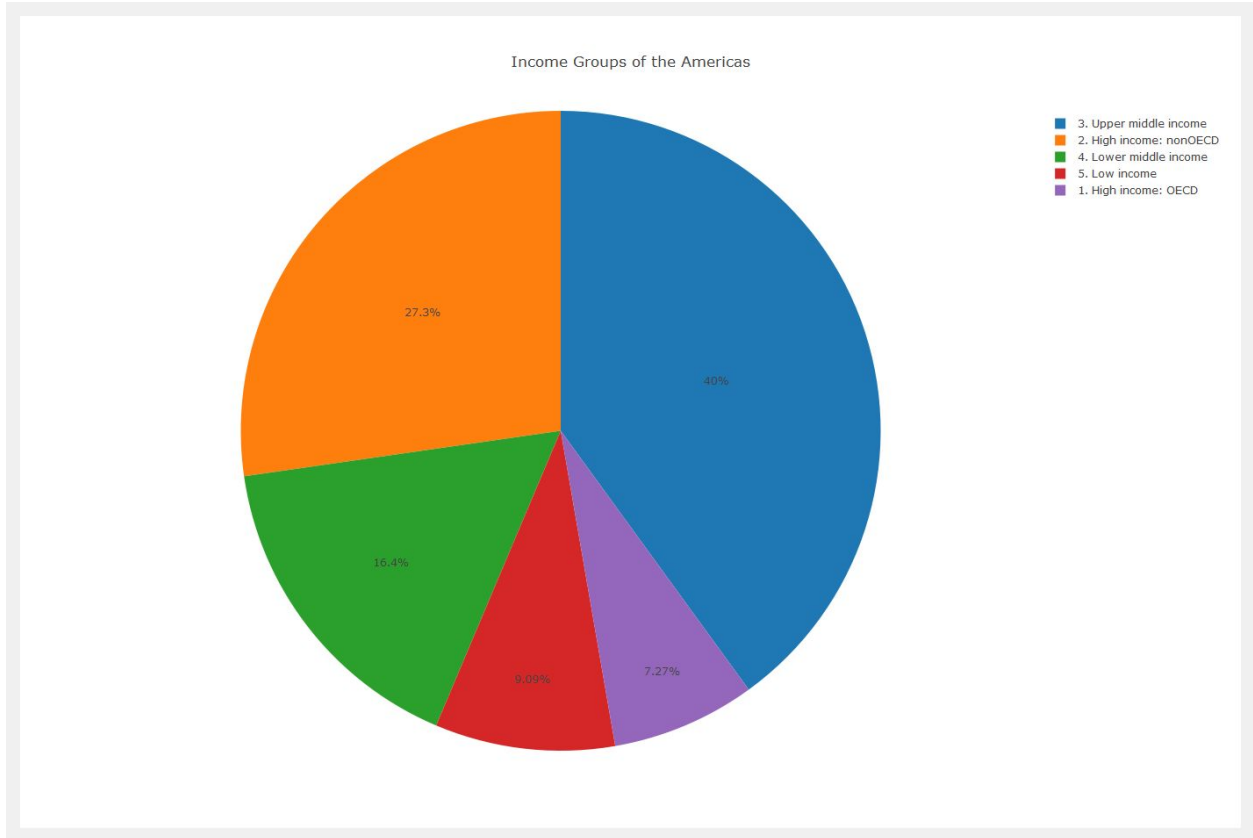*Figure B.6: Graph of the Population and GDP of countries in the Americas*

Figure B.7: Pie Chart of the Income Groups in the Americas