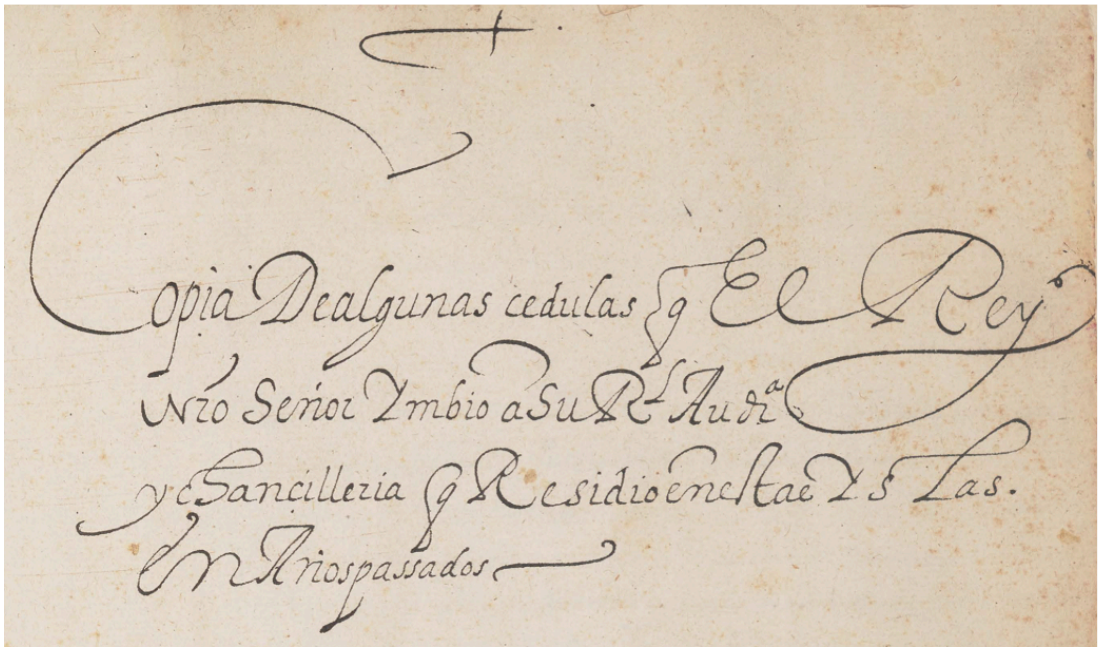
Performing Data Analysis of Spanish Colonial Law in the Philippines
(Platform Tutorial)



Description

This tutorial will show you how to use a Python open-source code to obtain and visualize descriptive statistics from a Spanish Cedulario (collection of royal decrees) from the early colonial Philippines (1565-1600). Any scholar or student interested in colonial Latin America's legal history could use it to analyze data sets created in Microsoft Excel built upon similar primary sources.

Grade Level(s): Undergraduate; Graduate & Professional
Course Subject(s): Digital Scholarship

Learning Objectives

- Understand the basics of performing an statistical analysis using Python
- Learn how to use Jupyter notebook and reutilize this source code to analyze similar data sets
- Explore the fundamentals of data visualization
- Explore the content of the Philippine Cedulario located at the Benson Library

Rights Statement

Creator(s): Abisai Perez, *Digital Scholarship Fellow*, Department of History, The University of Texas at Austin
Date Created: 2021-06-23

Introduction

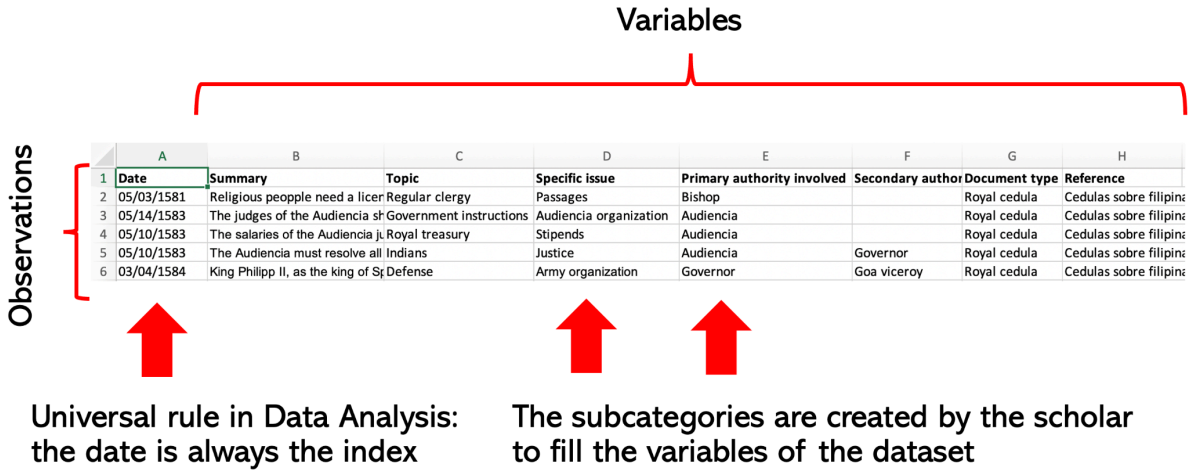
Data Analysis is commonly defined as the process of examining, cleaning, and modeling data so that we can derive some useful information and perform statistical analyses to help answer questions and solve problems in several areas.

One of the most popular and effective languages for Data Analysis is Python. This programming language not only allows you to get, clean, and manipulate data, but it also allows you to perform a wide range of analyses. For instance, Python could automate the reading of handwritten text, develop geospatial analysis, or build financial analysis in real-time.

Statistics is a key component of Data Analysis because it comprises a series of methods to collect, interpret, present, and summarize the information extracted from primary sources to turn observations into useful information. Statistics spans three broad areas—descriptive and inferential statistics and probability. While the last two deal with prediction and randomness about future events, descriptive statistics organize and summarize data from past events in visual and numeric ways. Like History, descriptive statistics serve to find patterns throughout time and to frame the plausibility of how something can cause something else.

Creating appropriate datasets in Excel

Before getting into Python, the first steps are to define the data that you want to study, collect it from primary sources, and store it in an appropriate dataset. This project focused on studying a Spanish Cedulario or collection of royal decrees. I entered the data in an Excel spreadsheet by following the well-accepted criteria of Hadley Wickham, who suggests managing data in a tidy format. Wickham explained that “Tidy datasets are easy to manipulate, model and visualize, and have a specific structure: each variable is a column, each observation is a row, and each type of observational unit is a table.”



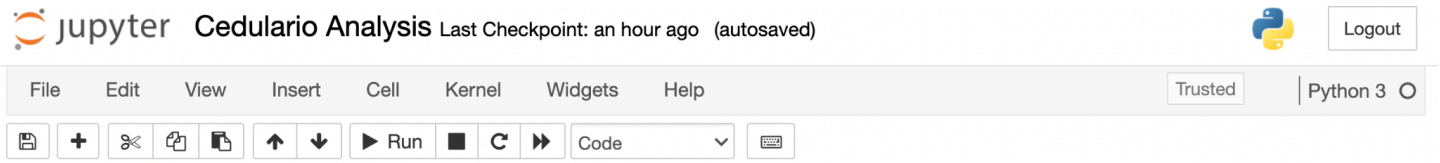
To perform a good statistical analysis, the dataset should contain all the possible variables that reflect or record the content of the information collected. In Statistics, there are two types of variables. They are either numerical variables -represent a counted or measured quantity-, or categorical variables -represent properties or characteristics. The process whereby Spanish authorities produced and implemented a law in the colonial dominions can be approached using categorical variables.

For this project, I developed eight variables. They are: Date, Summary, Topic, Specific issue, Primary authority involved, Secondary authority involved, Document type, Reference.

Any scholar interested in reutilizing this source code WITHOUT MAKING ANY CHANGE should construct a similar dataframe using THE SAME VARIABLES (columns) and writing them EXACTLY as I did (including capitals and spaces).

The subcategories to fill out the observations and variables can be the same as here. But the scholar can create her/his own subcategories for particular research purposes. That will not obstruct the reutilization of the Python code.

Statistical Analysis using Python in Jupyter notebook



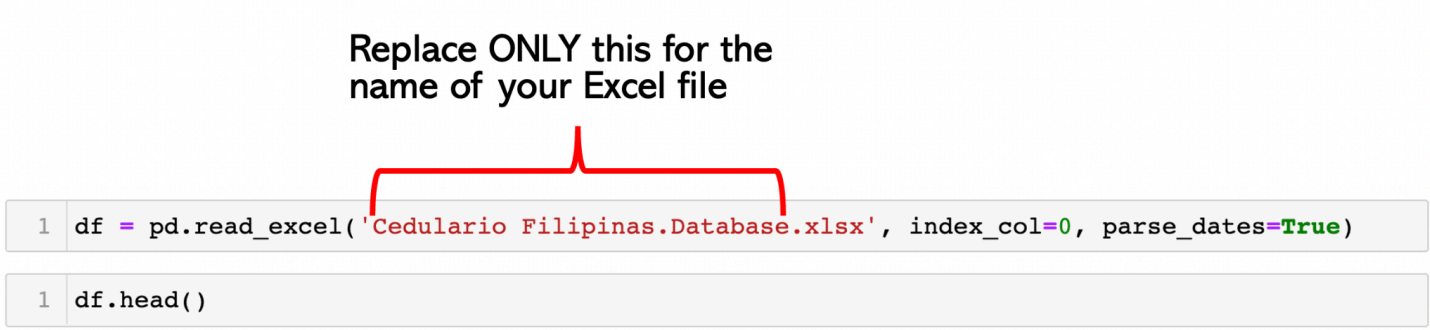
Jupyter Notebook is an open-source web application that allows Python programmers to create and share their source codes. In this way, people can reutilize, modify or improve the work of others. As a Digital Scholarship Fellow, I developed my research project in a Jupyter notebook.

Thus, the scholar interested in this code should first install Python and Jupyter notebook in her/his computer. A complete and detailed guide on how to install and use Python in Jupyter Notebook for Liberal Arts scholars can be found here: https://programminghistorian.org/en/lessons/jupyter-notebooks.

After making the proper installation, most of the job has been done! The following steps are only about loading the data, obtaining the statistical results and visualizing the data with only typing enter! You only need to copy and paste EACH LINE OF CODE IN THE SAME ORDER FOLLOWED in my Jupyter notebook. Refer to my original Jupyter notebook saved as a HTML file [INSERT LINK] to see the entire notebook, follow and copy the lines of code, and check the final results of my analysis.

All the Python libraries necessary for the analysis are already loaded in the first line of code. Please, do not modify anything, unless indicated.

1. **Reading data**: Load your Excel file by introducing the name of the Excel spreadsheet you want to analyze. In this case, the ONLY INFORMATION YOU NEED TO REPLACE IS “Cedulario Filipinas.Database” After that, you only have to type enter to see the data frame loaded in the Jupyter notebook with Python.



Date	Summary	Topic	Specific issue	Primary authority involved	Secondary authority involved	Document type	Reference
05/03/1581	Religious peopple need a license if they want ...	Regular clergy	Passages	Bishop	NaN	Royal cedula	Cedulas sobre filipinas, 1581 a 1594," BLAC, G...
05/14/1583	The judges of the Audiencia shall have pre-emi...	Government instructions	Audiencia organization	Audiencia	NaN	Royal cedula	Cedulas sobre filipinas, 1581 a 1594," BLAC, G...

2. **Getting tables of frequency**: A frequency table is the first step to identify patterns of the topics and issues. It also allows us to transform categorical data into numerical data, so we can build visualizations to analyze them. If you uploaded the dataset following the previous instructions and incorporating EXACTLY THE SAME NAME OF VARIABLES OR COLUMNS, the only thing you need to do is to type enter and that is everything! You will immediately get the tables of frequency!

Table of frequency by topic

```
freq_table_topic = df['Topic'].value_counts().to_frame()
freq_table_topic.rename(columns = {'Topic': 'Frequency'}, inplace=True)
freq_table_topic
#ONLY TYPE ENTER!
```

	Frequency
Government instructions	56
Royal treasury	28
Indians	25
Secular clergy	15
Trade	14
Defense	12
Report request	10
Chinese	8
Regular clergy	7

3. **Visualizing data**: After preparing the frequency tables, we can proceed then to visualize our data. This project has implemented the basic charts (bar charts, pie charts, and time series) used in descriptive statistics to perform the visualization of the data. Again, if your dataset uses THE SAME NAME OF VARIABLES OR COLUMNS, the only thing you need to do is to type enter!

```
ax1 = freq_table_topic.plot.barh(color='lightcoral', figsize=(10, 6))
ax1.set_title('Frequency by topic', fontsize=18)
ax1.grid(axis='x')
ax1.set(facecolor = "whitesmoke")
ax1.get_legend().remove()
#ONLY TYPE ENTER!
```