Power of Data Science

Benjamin Dragenice

BOWIE STATE UNIVERSITY

Symbolic Computation using Big Data

CTEC 298

Spring 2021

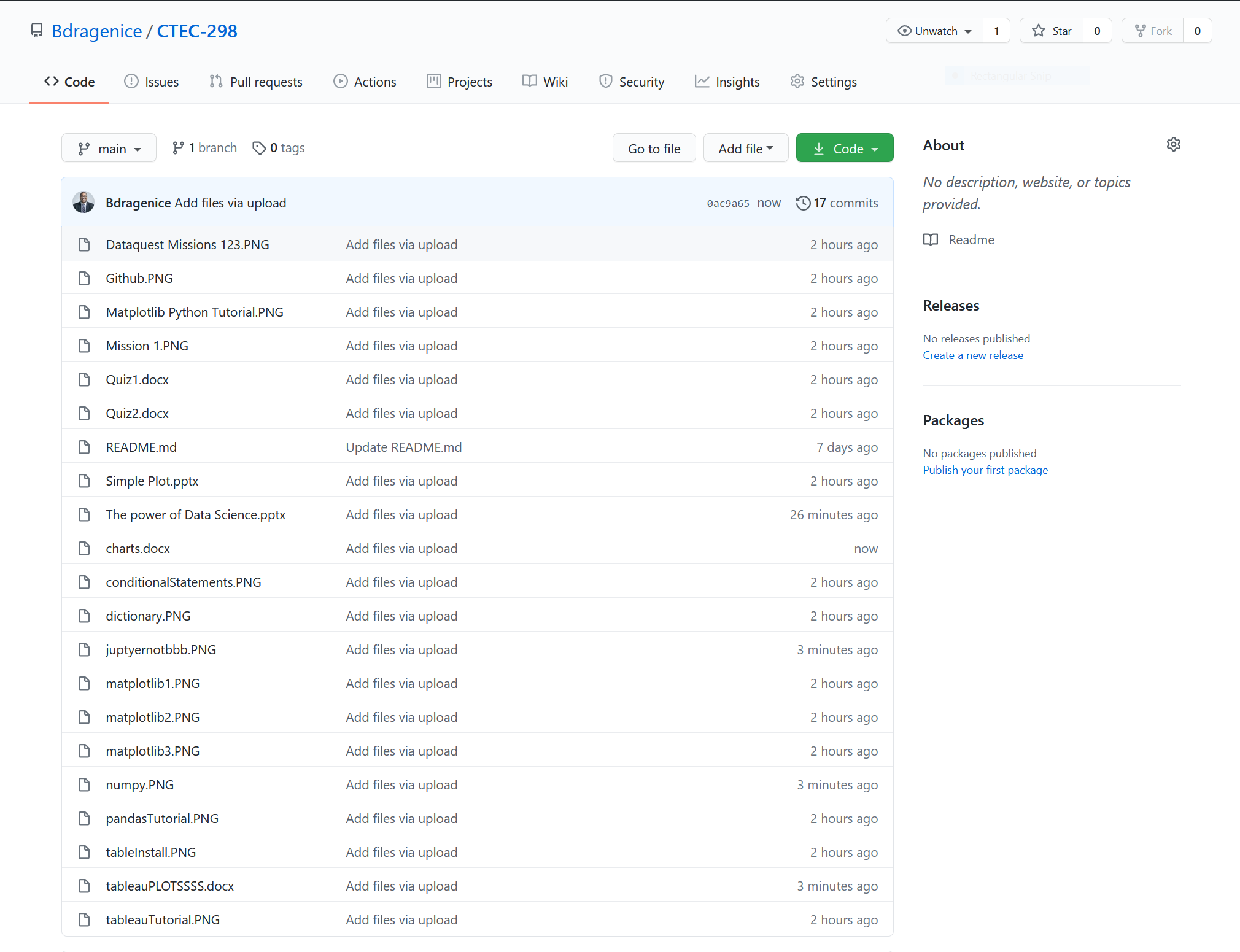
**Introduction**

Data science is an inter-disciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from structured and unstructured data, and apply knowledge and actionable insights from data across a broad range of application domains. Through CTEC 128 and 298 we explore data science and the tools used for it. In this paper we talk about the tools used in CTEC 128 and the new tools we learned through CTEC 298 which includes using python and its libraries to create plots to showcase data.

**Summary of CTEC 128 Paper**

Marriage, according to Britannica.com, is a legally and socially sanctioned union that is regulated by laws, rules, customs, beliefs, and attitudes that prescribe the rights and duties of the partners and accords status to their possible offspring. For this project our group used data provided by the American Community Survey.  Within the data the ACS had an expectation of five groups married, separated, divorced, widowed, and never married.  As a group we decided to use data specifically from 2019 although the ACS had data predating back to 1996. With this data we focused on what factors lead to a successful marriage, and if these factors make a difference.  As a team our brainstorming process was the most important part of making this project. Personality traits were going to be hard to determine why people get married with data, so we focused on external things that could be tracked. We started with four major groups: economics, social, housing, and demographic.  Economics covered subtopics such as one's employment status, income, or occupation.  For social reasons we thought it would be best to talk about education, status, or previous relationships. As for housing touched points such as if someone had a house, and transportation. Demographics had to be the most important topic out of them all because we knew that the statistics could change for each race.  At the end we decide to narrow down our factors down to three out of many for our presentation.  Our focus was directed to education level, employment status, and race. After in depth research we narrowed the factors down to three. These three are employment status, educational attainment and race. The results of our project led us to multiple conclusions. When looking at the effect of employment status on a marriage we have concluded that people who are employed have the highest chances of attaining a successful marriage. When deciding what level of education has the biggest benefit on a marriage we have concluded that people with a master’s degree have the highest chance of attaining a successful marriage. Lastly, when looking at what effect a person’s race has on a marriage we have concluded that white people have the highest chance of attaining a successful marriage.

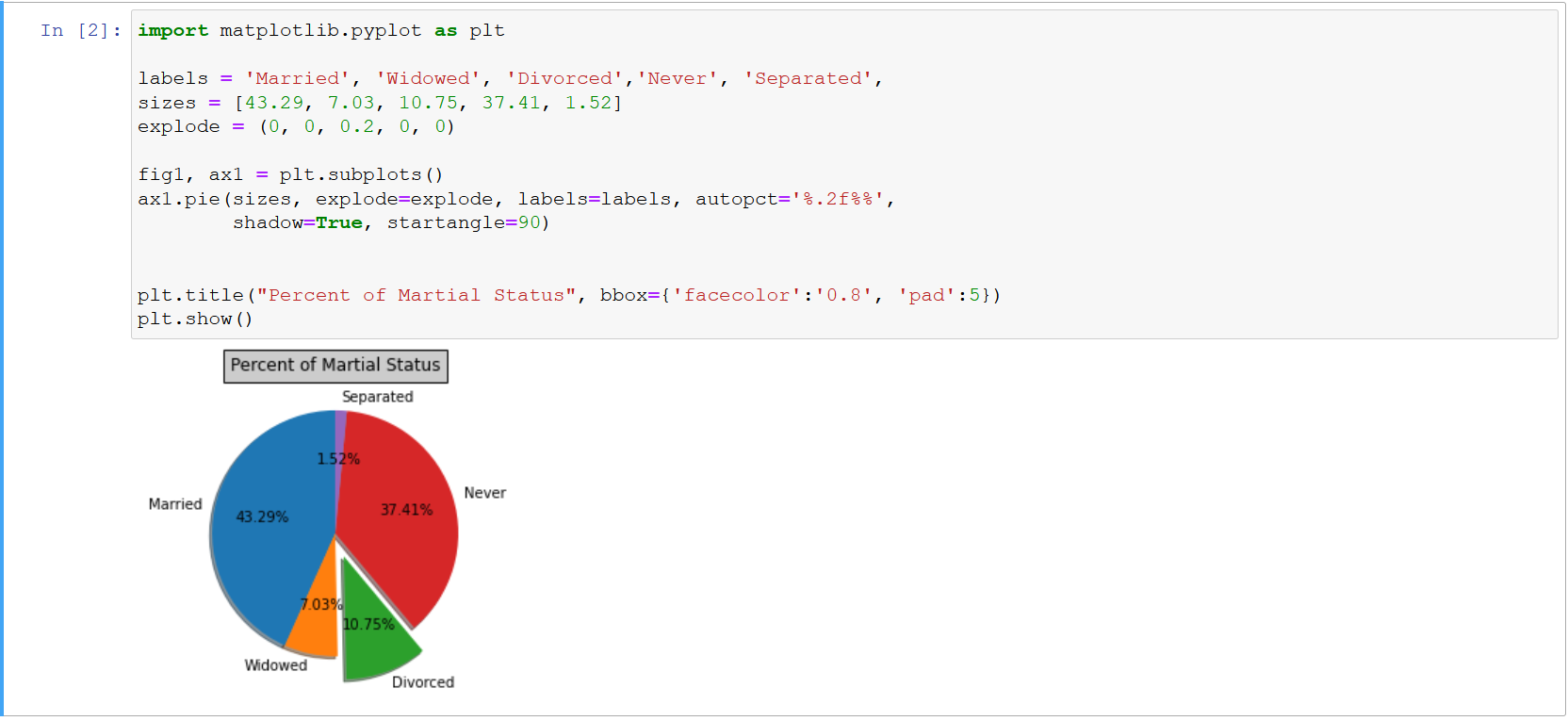
**CTEC Material Submitted**



This image shows all the CTEC 298 material that was submitted. This image also show the GitHub repository that was used to house all of these files to be viewed publicly. In CTEC 298 we covered Python and its various libraries to plot charts/graphs. We first started the semester completing Dataquest Tutorials so we can become more familiar with Python. After completing those we worked on completing tutorials on Matplotlib, Numpy, and Pandas. These are the 3 tools within Python that we used to complete our assignments. Matplotlib, Numpy, and pandas were all used to create the six plots that will be shown later in this paper. We also learned about Tableau, which is used to create visuals with Excel data. The charts made with Tableau is also shown later in this paper.

**Plot Deliverables**

**Pie Chart**



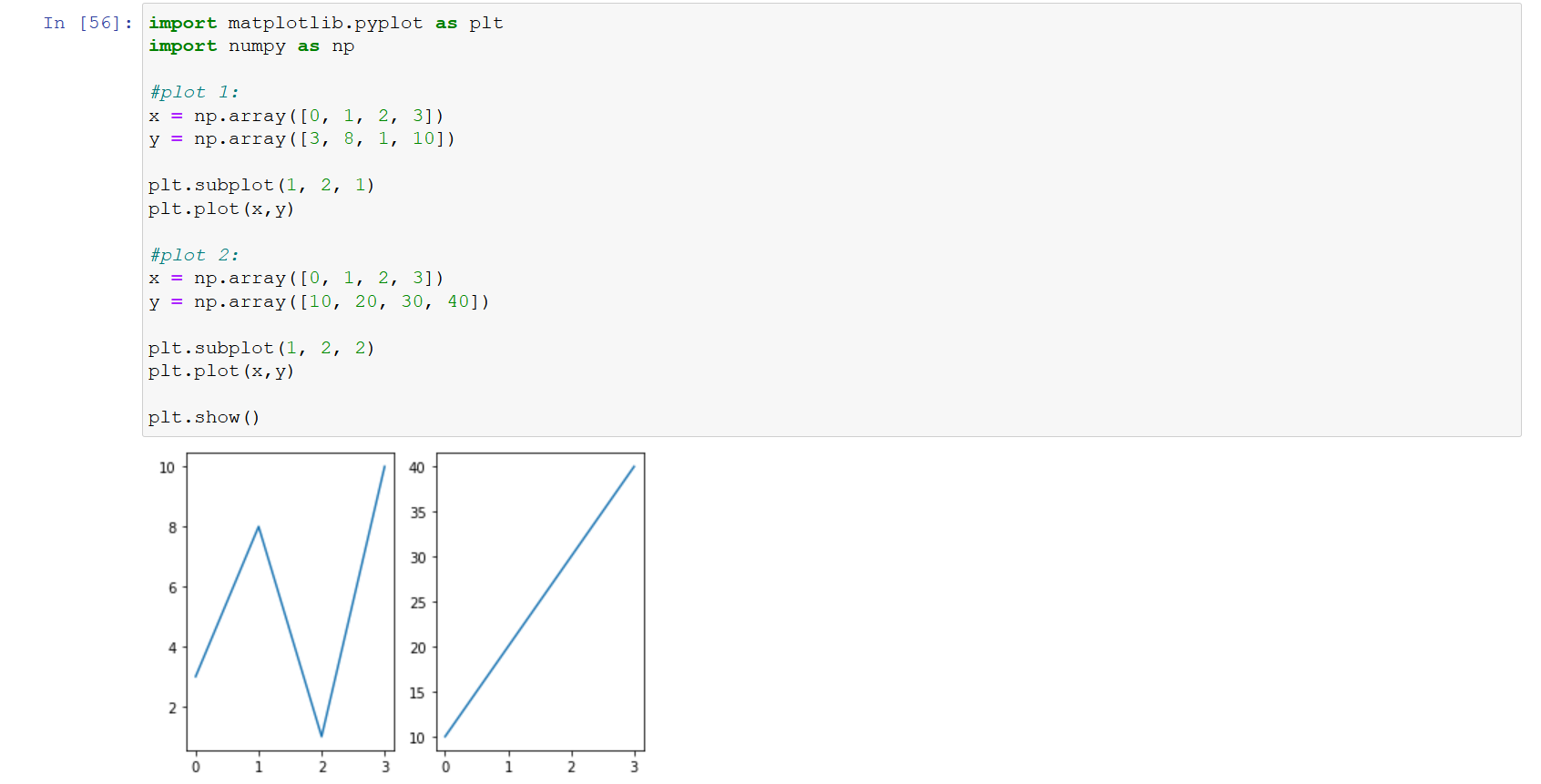
**Bar Graph**



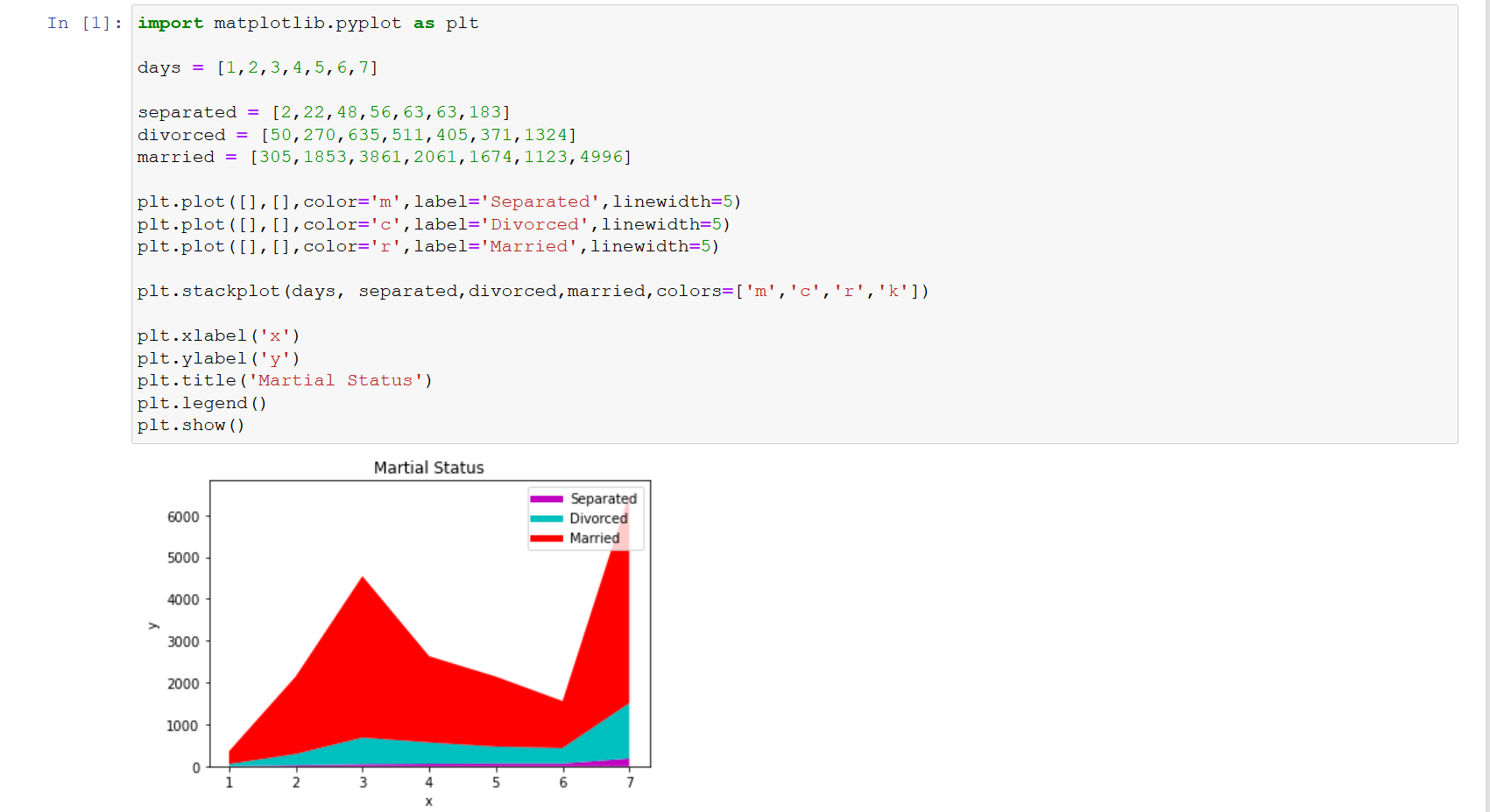
**Scatter Plot**



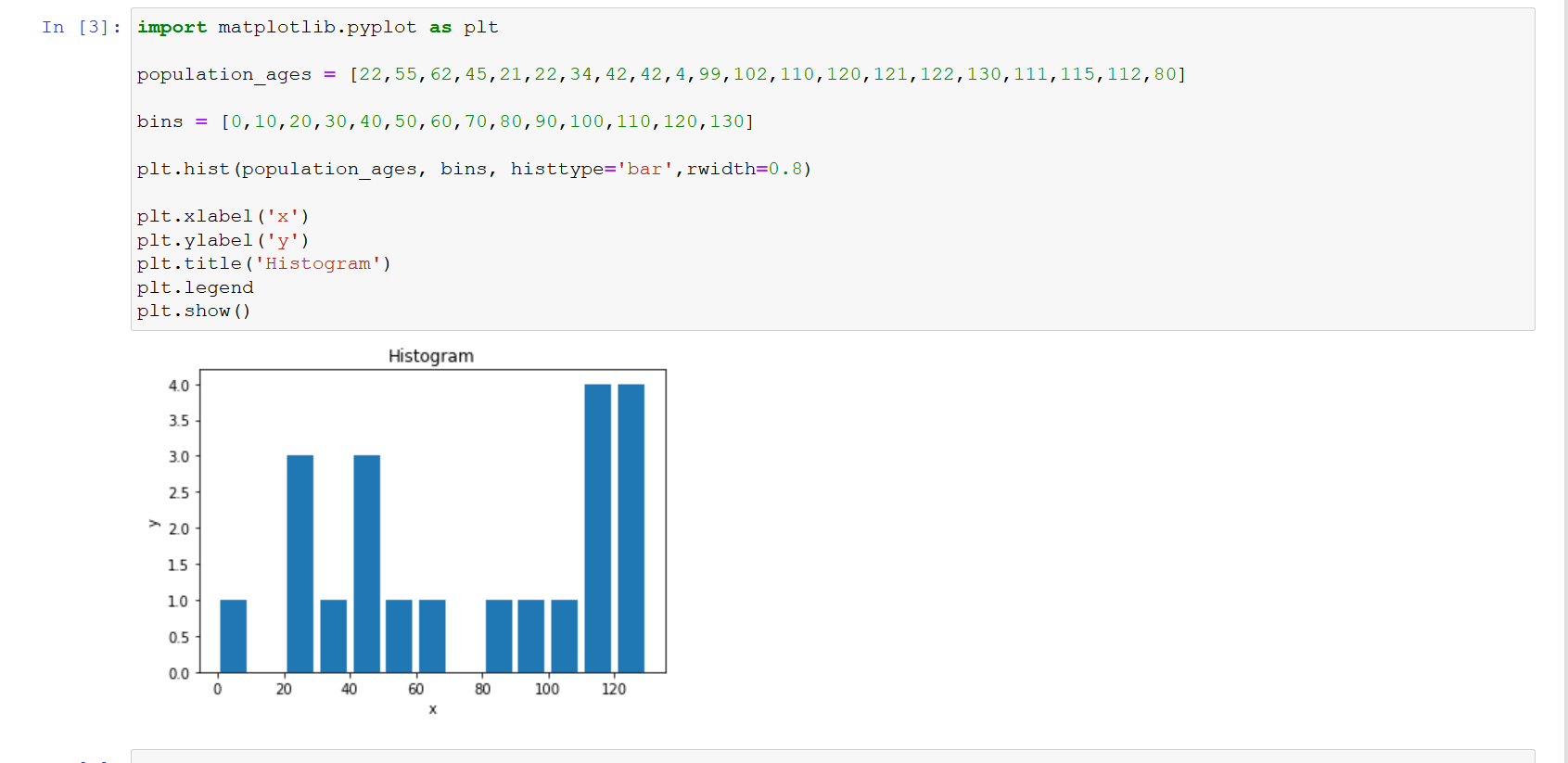
**Multiplot**



**Area Plot**



**Histogram**



**Tableau Plots**

****

**Conclusion**

In conclusion this course expanded on what we learned in CTEC 128. We used just excel and GUI tools to chart data in CTEC 128. In CTEC 298 we programmed using Python and its libraries in order to plot charts and graphs. We learned python through Dataquest tutorials. We learned Matplotlib, Numpy, and Pandas through the various tutorials. Matplotlib, Numpy, and Pandas are powerful tools for data science. In conclusion this class was a success in expanding our knowledge about data science and its tools.

**References**

Herston, K. (2016, September 30). Biggest Factor in Divorce Is Husband's Employment Status, Says Harvard Study. Retrieved December 20, 2020, from <https://herstontennesseefamilylaw.com/2016/10/05/biggest-factor-in-divorce-is-husbands-employment-status-says-harvard-study/>

Crowley, J. C. (2020, November 3). What Are Common Reasons for Divorce? Survive Divorce. <https://www.survivedivorce.com/common-reasons-for-divorce#8>

US Census Bureau. (2020a, October 16). About the Survey. The United States Census Bureau. <https://www.census.gov/programs-surveys/acs/about.html>

Marriage. Retrieved December 20, 2020, from <https://www.britannica.com/topic/marriage>

Matplotlib.pyplot.plot¶. (n.d.). Retrieved February 25, 2021, from https://matplotlib.org/devdocs/api/\_as\_gen/matplotlib.pyplot.plot.html#matplotlib.pyplot.plot

Simple plot¶. (n.d.). Retrieved February 25, 2021, from https://matplotlib.org/devdocs/gallery/lines\_bars\_and\_markers/simple\_plot.html

Line graph - when to use it? (n.d.). Retrieved February 25, 2021, from https://www.edrawsoft.com/chart/when-to-use-line-graph.html

https://www.youtube.com/watch?v=SiCyTcudoSE - Python Tutorial : Basic plots with python matplotlib