

7PAM 2000 Applied Data Science 1

Assignment 1: Visualization



Student id: 21077565

Student Name: NAGESWAR PINNOJU

Table of Contents

[Dataset Source 2](#_Toc118899197)

[Line Plot 2](#_Toc118899198)

[Histogram 3](#_Toc118899199)

[Scatter Plot 4](#_Toc118899200)

[References 4](#_Toc118899201)

# Dataset Source

https://www.ethnicity-facts-figures.service.gov.uk/crime-justice-and-the-law/crime-and-reoffending/victims-of-crime/latest/downloads/victims-of-crime-data.csv

# Line Plot

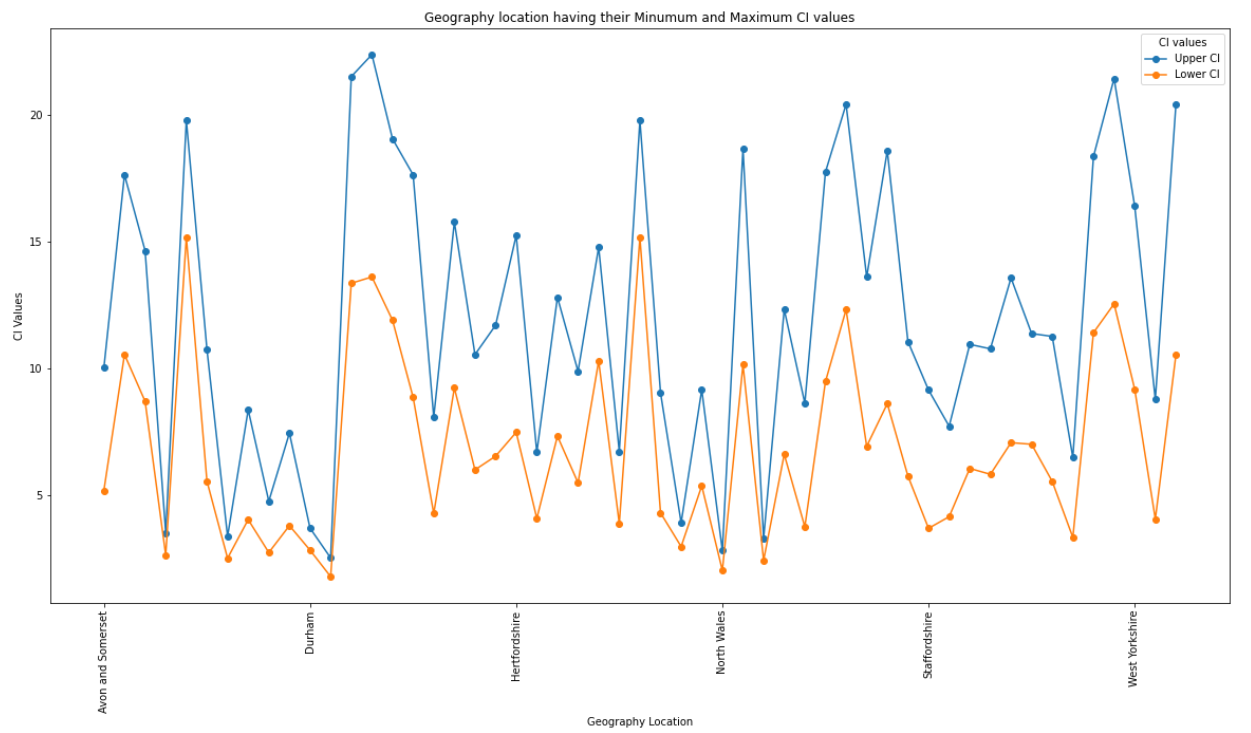


Figure : Line Plot having CI values according to Geography location

Line charts are used to illustrate the relationship that exists between two groups of data, each of which is shown on a separate axis in the chart. Utilizing the pyplot.figure() method in conjunction with this enables the presentation of several charts all inside the confines of a single container. This will offer us with the ability to compare the various charts and provide us control over the appearance and behaviour of the charts (Landup, 2021).

The line plot depicts the relationship between the confidence interval (CI) values and the geographic location. There are two distinct categories of CI value: upper CI values and lower CI values. The graph that includes the group by function, which is now operating on the geography column. In this, the values are grouped according to the geographical location, and if any geography location has more than one CI value of upper and lower, it will compute the mean of those values. Additionally, the values are grouped according to the geography location. There are 53 distinct values that are included inside the geography property, and it is easily discernible from the graph that there are 53 marking points along the line.

The legend is used as CI values in the graph, and the legend is presented in accordance with the colour of the column; for example, blue is used for higher CI, and orange is used for lower CI. Both axes of the line plot have appropriate titles and labels attached to them. Matplotlib, a software package, is used to generate a line plot.

# Histogram

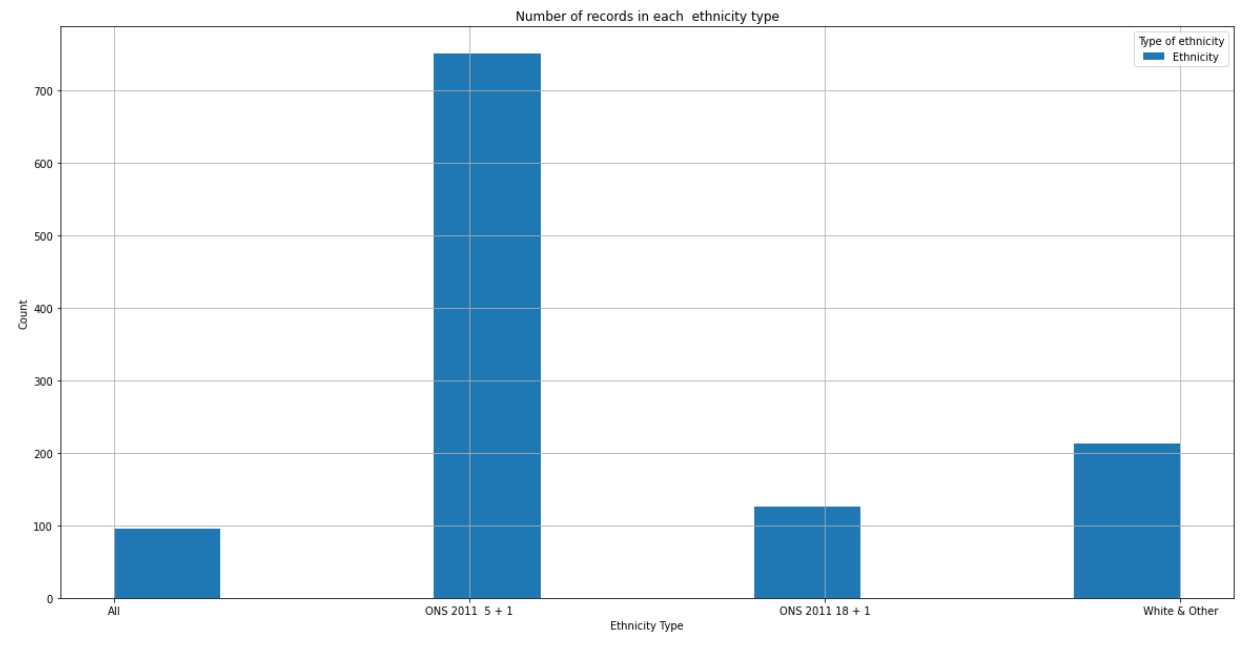


Figure : Histogram for Ethnicity Type

The frequency distribution may be shown graphically in a histogram. On a graph, the number of observations that take place within each set period is represented by a point. In order to generate a histogram, there is use of the hist() function that is made available by Matplotlib. A histogram is the output that is generated by the hist() function, which takes an array of integers as its input. After that, the histogram is constructed with the help of this array (Nik, 2020).

Here, the histogram is displaying the number of observations of the attribute Ethnicity. In essence, there is a count of the attribute. Within the attribute, there are four types of ethnicity: "ONS 2011 5 + 1", "ONS 2011 2011 18 + 1", and in the very last position, there is the type White & other. It is possible to get the following conclusion after examining the graph: "ONS 2011 5 + 1" contains the greatest number of observations. The legend that associates the blue hue with a certain ethnic group is employed here to convey this information.

# Scatter Plot

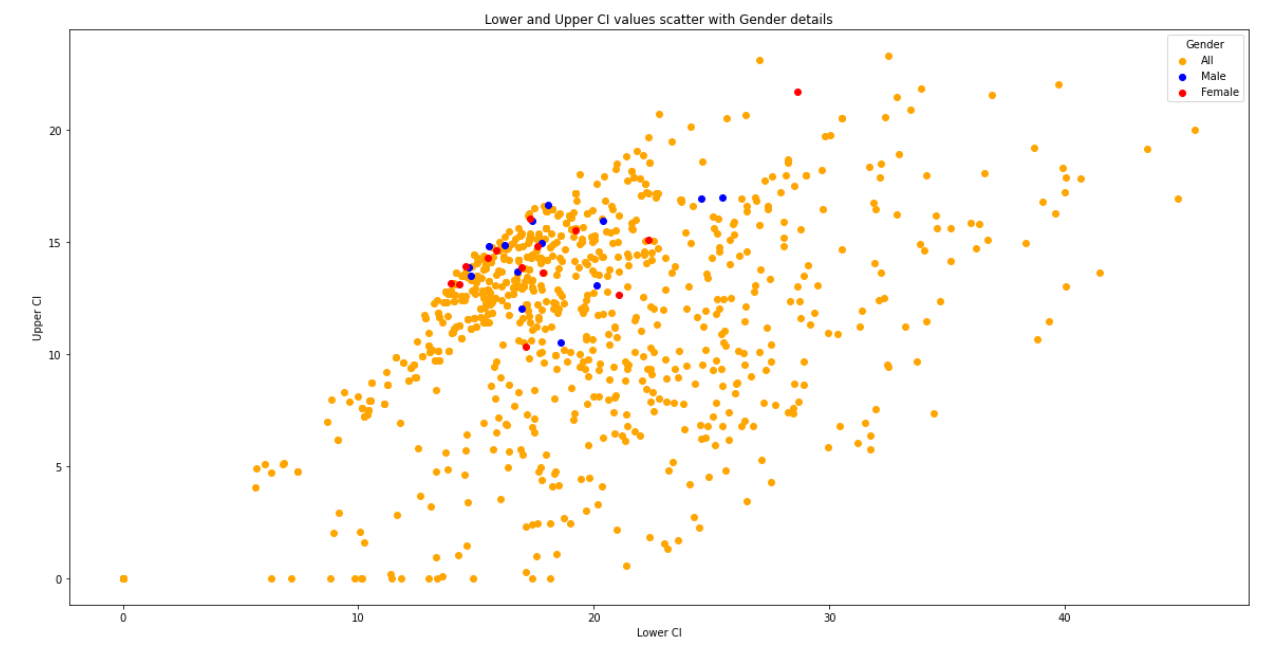


Figure : Scatter plot having info about CI values according to the Gender

The link between the two may be analysed with the use of scatter plots, which make use of dots to illustrate the relationship between several different variables. For the purpose of generating a scatter plot, the scatter() function that is part of the matplotlib package may be used. The relationship between many variables, as well as the influence that altering one variable has on another, may often be graphically shown by using scatter plots (Verma, 2022).

The scatter plot displays the values of CI so that the Lower CI values are located on the x axis, while the Upper CI values are located on the y axis. The scatter plot displays a variety of colours because it is segmented based on the gender characteristic, as shown in the legend. This allows for the inclusion of a wide range of colours. There are three different sorts of values that make up the gender, and they are all male or female, with some dispersal across the three colours that make up the base of the gender. In a nutshell, it demonstrates the values of the CI in relation to the gender.

# References

Landup, D. (2021) *Matplotlib line plot - tutorial and examples*, *Stack Abuse*. Stack Abuse. Available at: https://stackabuse.com/matplotlib-line-plot-tutorial-and-examples/ (Accessed: November 9, 2022).

Nik (2020) *Creating a histogram with Python (matplotlib, pandas) • datagy*, *datagy*. Available at: https://datagy.io/histogram-python/ (Accessed: November 9, 2022).

Verma, A.S. (2022) *Scatter plot in Matplotlib - Scaler Topics*, *Scaler Topics*. Available at: https://www.scaler.com/topics/matplotlib/scatter-plot-matplotlib/ (Accessed: November 9, 2022).