HOMEWORK #01 (9 pts)

Summary

Create stunning data visualizations as well as descriptive statistics and share it with the class.

Instructions:

There are 2 questions on this assignment. You may use any dataset(s) you wish for the assignment. You may select your dataset from kaggle, UCI Machine Learning Repository or another publicly available datasets. You can find many others on the Web. Pick something that interests you.

Upload your jupyter solution notebook(s) and dataset(s). Don't forget to put code descriptions (markdown or comments), mention about your references/sources in your notebook.

Each student will share their final visualization in class via jupyter notebook. Please plan for a demo of up to 10 to 15 minutes. Assignments not presented in the class will not be graded.

1. Data Visualization

You can use the Matplotlib, Seaborn etc. python libraries to create your visualization(s)

a. Describe your dataset

b. Select 2-5 columns (attributes / features) from your dataset, use different and rich statistics functions to give info about important attributes of your dataset.

b. Use different charts for visualization.

2. Geospatial Data Visualization

GeoPandas (geopandas.org) is an open source project to make working with geospatial data in python easier. GeoPandas extends the datatypes used by pandas to allow spatial operations on geometric types.

a. Find a geospatial dataset and describe your dataset

b. Plot geospatial data (population density, covid cases, geometry etc.) using different functions for visualization.

Grading *(1st question: 6 points, 2nd question: 3 points)*

Output:

1. Jupyter Notebook File : Share your codes and comments with a jupyter notebook (s)

2. Complementary File : In a pdf file share your data source, dataset description, libraries yo have used.

Each group will share their final visualization in class via jupyter notebook. Please plan for a demo of up to 8 to 10 minutes. Assignments not presented in the class will not be graded.