- **Due 11:59pm on March 23, 2023. Late penalty: 20 points off per week. Total 110 points**
- This is an **individual homework**
- Write programs to solve the following problems. Your results should be accessible through the web page at http://obi.kean.edu/~xxxx/CPS5721
- You project main page index.html should have 2 sub-links - HW2.1 and HW2.2 on the main page.
- Please set permission mode to **705** for all your files (HTML, Python, etc).
- You need to submit the homework through the class website. http://vader.kean.edu/students
- **You can use Google Charts, D3.js, plotly.js to draw graph, or JpGraph library to generate images.**
- You need to write programs in **Python** to get the data from the database, process the data, call API or function to calculate, and print the results and explanation for each question on the browser.
- Each Python the program should be under your **cgi-bin** folder on **obi.kean.edu**.
  The cgi-bin folder should be created at ~/public_html/cgi-bin/
- Questions are based on the tables in **datamining** DB. You must clearly list the question # for each answer.
- All the link and programs should call programs under same folder. No link to eve or other servers.
- **You must read the data from the database and calculate in the programs. Do NOT hardcode any results. Otherwise, the corresponding function will have 0 point.**
- **Note: Your project will be 0 if the system detects plagiarism in your programs. It is ILLEGAL to share your codes or copy/modify/submit other people's codes under your name.**

Please use table uszips table in the simplemaps database for the following 2 questions. Do NOT hardcode your results/values in the program. You must directly get data from the table in your database instance on imc.kean.edu server.

HW2.1 (55 points) Scatter plot and historgram. The goal is to analyze and visualize the data and show the results in scatter and histogram plots on the browsr.

- Please write a Python program named "**scatter_histogram.py**" for this question.

1. _____ (10 points) Create a table **County**(state, county, income, population, home_value, education_college_or_above) in your database that includes 50 states and DC. **Note** : education_college_or_above is the ratio indicating the county has % of population with college or higher degree. The home_value indicates the average of the home value in the county.

2. _____ (10 points) Please create a scatter plot with income as the X-axis and education_college_or_above as the Y-axis.

   For questions 3 and 4, you will need to divide the income into six ranges based on 20K intervals: < 20K, 20-40K, 40-60K, 60-80K, 80-100K, and > 100K.

3. _____ (10 points) Please create a histogram using income as the X-axis and the number of counties in each income range with education_college_or_above as the Y-axis.

4. _____ (10 points) Please create a histogram using income as the X-axis and the total population of education_college_or_above in each income range as the Y-axis? You may create two series for questions 3 and 4 on the same histogram chart, or you may create two different charts.

5. _____ (15 points) Please analyze these 3 plots and discuss the advantages of each chart in representing the data. Additionally, please share your findings regarding these 2 attributes.

Please proceed to the second page.

HW2.2 (55 points) Boxplot. The goal is identify if any counties are outliers in both attributes.

- Please write a Python program named "**boxplot.py**" for this question based the two attributes home_value and education_college_or_above in your **County** table.
- You can put both plots on the same chart or two different charts. You need to show the numbers, not just display the boxplot.

6. _____ (20 points) Please create a boxplot and indicate the first quartile (Q1), median (Q2), third quartile (Q3), upper limit, lower limit, and any outliers for home_value.

7. _____ (20 points) Please create a boxplot and indicate the first quartile (Q1), median (Q2), third quartile (Q3), upper limit, lower limit, and any outliers for education_college_or_above.

8. _____ (15 points) Please provide your findings and conclusions regarding the outliers of the home_value and education_college_or_above attributes? Are they from the same counties, and which state has more outliers? Additionally, what factors may contribute to them being outliers??