Project.

The story of our favorite professor continues. Now she has been thinking about the cost of her medical procedure and wondering what things might influence the different costs for the same procedure across the country. From listening to the national news broadcasts and reading various news papers, she has heard that sometimes large pharmaceutical companies will encourage physicians to promote procedures that use the pharmaceuticals that have the highest profit margin. Therefore she is concerned that there might be a correlation between the cost of her procedure and how much money (including goods and services) a pharmaceutical company might be paying a physician's office. She wants to quantify the correlation between the cost of the service and the money spent on the office.

From her previous investigation, she knows that Medicare provides the cost of various procedures, and that the US government requires the pharmaceutical companies to disclose how much money they pay to each physician (see https://openpaymentsdata.cms.gov/). She is interested in the correlation between the total cost of all Medicare billings and the total of all monies paid to a physician's office.

Here is what she wants to see:

1. A Venn diagram showing:
   (a) The total number of Medicare payment records by address,
   (b) The total number of pharmaceutical payment records by address, and
   (c) The number of intersections between the two sets of data.
2. A X-Y plot showing pharmaceutical payments vs. medicare billings.
3. The correlation between pharmaceutical payments and medicare billings.

Here are her constraints:

1. She has both undergrad and graduate students.
2. The students can form 2 person teams, where all team members are either undergraduates, or graduates.
3. From the undergrads, she wants all of the above deliverables for the state of VA.
4. From the grad students, she wants all the above deliverables for the entire intersection between the two data sets.
5. Because of the importance of this information, she wants a short presentation (about 5 minutes) from each team.

A simple explanation of correlation (with sample data) is found at:
https://www.mathsisfun.com/data/correlation.html