* *Initial goals:* What is your goal, what the questions you are trying to answer about your dataset? (Every team should have at least 5 questions)

We were able to answer a number of questions with this dataset. We were able to see the correlation of security breaches over time. This can help us recognize the effectiveness of security in the battle against hackers. As breeches continues to increase over time (as we can see in the Breeches Over Time Graph) companies can see that advance and investing in security will always be an important priority. We also want to see if there is a correlation between the type of breaches and the number of individuals effected by the breach. This information would be useful so companies can recognize what kinds of security they should invest the most in. When we saw we had records of the type of device that was breached in our data set we wanted to know if certain devices are statistically more susceptible to data breaches than others. This information would be similarly useful to our last question. The information can help companies recognize what pieces of hardware they need to be more careful with, if any. A fourth piece of information we wanted to visualize is to see if there is any correlation between the number of breaches and location(state). Lastly, we want to use a mosaic chart to show the relationship between the different types of breaches and the total number of breaches in the dataset. This can also show which breaches occur the most often. This can also be important for decision making on security investments.

* Dataset*:* the source of your dataset, description of your dataset structure, how did you clean up your dataset, and any things related to the used dataset.

Our data is various information regarding data breaches in the US. Some of the information stored is in regards to the data of the data breach, number of individuals effected, the company effected and more. We did have to clean a few things in our dataset. We removed a few columns that we felt were irrelevant to the information we were trying to gather. We also had to fix some of the names of the companies because some of the were misspelled or had and extra punctuation mark. We also had to recode all of the dates to change them from character strings to the actual date datatype. This made it easier for us to sort, gather, and group data by date, which was important to find the correlation between number of breaches over time.