* **Statistical/ Hypothetical Questions**: Is there an air pollutant that was generated the most?
* **Outcome of your EDA.** Since my dataset was large with over 16 years of data, the EDA process helped me go through each variable and understand their significance better. For example, NO2 and SO2 were measured in billions and O3 and CO were in millions. Once I noticed the difference in measurements, I knew which variables I could compare for certain questions.
* **What do you feel was missed during the analysis?** In the beginning of the project I was focused more on getting my program to plot correctly that I wasn’t always making sure my questions were being answered. Once my code worked properly and my graphs were how I wanted them to be, I went back and made sure I had incorporated my questions into it.
* **Were there any variables you felt could have helped in the analysis?** My dataset was large, it started on 1/1/2000 and went to 12/31/2016 and would have trouble with Excel opening it up.
* **Were there any assumptions made you felt were incorrect?** In the beginning I assumed all 4 pollutants had the same measurements and started comparing them equally, then I noticed the difference in measurements (billion and million) and I had to change my methods.
* **What challenges did you face, what did you not fully understand?** Some of the challenges I faced were when it came to the analysis when I created the graph I wanted and then I needed to understand the significance of what it meant. I needed to go back and review the terms to make sure I was using the proper graph.