Response Summary:

1. Student Information *

First Name	David
Last Name	Luo
Major	Data Science
Course *(e.g. CGT 270-001)*	CGT-003
Term *(e.g. F2019)*	S2022

2. Email Address *
(University Email Address is required.)
luo354@purdue.edu

3. Visualization Assignment *
- Lab Assignment

Analyze

4. Basic Descriptors: for each data component from the Parse Worksheet, identify basic descriptors (basic statistics). Explain *
For identifiers like "Code" and "Name" basic descriptors are number of entries. For all columns, They include continuous data and contain basic statistics such as Min, Max, Range, Mean, Median, Variance/Standard Deviation. If the data were pivoted to make "Year" into a field, Year would be a discrete variable with basic descriptors, Min, Max, Range.
*Note: the data is not tidy. i.e. The table is wider than what would be optimal for analysis
*The data needs to be pivoted in to a more tidy format to get more basic statistics
Year:Range = 2005 to 2014
Name:Count = 32
Industry and Commercial:Min =182.7
Industry and Commercial:Max = 2712.5
Industry and Commercial:Mean = 585.9
Domestic:Min = 16.5
Domestic:Max = 870.7
Domestic:Mean = 474.9
Total:min = 643.8
Total:max = 3567.6
Total:mean = 1319.0

5. Categorize: consider what is similar and what is different? Categorize the data. Are the variables categorical (normal, ordinal, or rank). Are they quantitative (discrete or continuous)? Show categories. Explain. *
Categorical variables are nominal, includes things like "Name", "Code", and Type of Emission. Year is a discrete variable. All observations (CO2 emissions) are continuous.

6. Temporal: is the data streaming data? How is it stored (all at one time, over several years in years, days, minutes, seconds)? Explain. *
No, the data is not streaming data. The values are from 2005 to 2014 and can't be updated by users.

7. Range and Distribution: what is the distribution of the data? Few values, small size, evenly spread, sparse or dense? Explain. *

The data has a wide range of values over a large time. The data is also dense. The spread of the values is relatively even with some outliers including the borough of Westminster.

Evaluate

8. Questions and Assumptions: list at least 3 questions you plan to answer with the data or list the questions if they were provided. Must be complete sentences and end in a question mark. What assumptions are you making? *

Question 1	Is population corelated with emissions?
Question 2	Combined with geospatial data, Is there a correlation between latitude and pre capita emissions?
Question 3	Is emissions going up per year for each borough?
Assumptions	Latitude is defined from the "center" of each borough.