Response Summary:

1. Student Information *

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

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

3. Visualization Assignment *
- Lab Assignment

Understand

4. Parse Data: List each field and its data type. Refer to Fry (page 8-9, 2007) for examples of description of different data types (string, float, character, integer), you can also create user defined types (some combination that uniquely identifies data like the Index type in the Fry 2007 page 9 example) *

Carbon-Emissions-Borough (Training Data)
Code - String
Name - String
Category - String
Year - Integer (2005 - 2014)
CO2 Emissions - Float
Workless-Households-Borough (Extra Dataset 1)
Code - String (Does not match previous)
Name - String (Matches previous)
Working Households/thousands - Integer
Working Households/percent - Float
Mixed Households/thousands - Integer
Mixed Households/percent - float
Workless Households/thousands - integer
workless Households/percent - float
year - integer (2004 - 2020)
Earnings-Residence-Borough (Extra Dataset 2)
Code - String (same as Extra Dataset 1)
Area - String (similar to name)
Year - Integer (2002-2022)
Pay (£) - Float
conf % - Float
*note, datasets 2 and 3 are different from the Acquire worksheet due to a misunderstanding of directions in the Acquire step. Datasets come from the London government website https://data.london.gov.uk/

5. Assumptions: List any assumptions you are making about the data and/or the visualization challenge (aka the project) *

The three datasets all refer to the same geographical areas with different measures. A map may be a useful visualization to see how the three datasets are related. The datasets have the same or very similar names/codes so joins should be possible. The datasets are formated differently and have some missing values so cleaning may be necessary. In datasets 2 and 3, I assume the symbol "!" represents missing data.