CGT 270 Data Visualization
Module 1 ● Week 2
Lab 2: Parsing Data

Name: **Chia-Hua (Joy) Lin**

The goal of this lab is to understand the structure of data. In this lab you will change data into a format that tags each part of the data with its intended use. After completing this lab every element of the data, you selected (Tableau dataset) and the two (2) additional datasets you acquired in lab last week will be broken into its individual parts.

What you should be able to do (at the end of this lab):

Remember	*Describe* what happens in the **parse** stage.
Understand	*Describe* the data in detail according to the parsing specifications.
Apply	*Demonstrate* the ability to change data into a useful format for future processing.
Evaluate	*Categorize* the data according to parsing specs.
Analysis	*Identify* specific features about the data.
Create	*Generate* a parsed listing of the data.

Part I: Tableau Data Set
In this document:
Answer the following questions about your Tableau data:

1. List the name of the Tableau Dataset you selected in the Acquire Lab:
   a. Top Baby Names by State
2. How many rows (records) are in the data set?
   a. There are 10506 records in the data set.
3. How many columns (variables) are in the data set?
   a. There are 5 columns in the data set.
4. What assumptions are you making about the data?
   a. I am assuming that the data is complete with everyone born being registered. I'm also going to assume that the Name Occurrences column only counts people with this name that was born in this year, instead of a tally of all the living people with this name.

Use the following link to provide a parsing of the Tableau Data.
https://tinyurl.com/Parse-Worksheet

For Visualization Assignment select: **Training Data**
Then provide the parsed data and list your assumptions.

Part II: Additional Data Set #1
In this document:
Answer the following questions about your Tableau data:

1. List the name of the Tableau Dataset you selected in the Acquire Lab:
   a. Unisex names table
2. How many rows (records) are in the data set?
   a. There are 919 records in the data set.
3. How many columns (variables) are in the data set?
   a. There are 5 variables in the data set.

4. What assumptions are you making about the data?
    a. The process of filtering for unisex names was not described, so I am pretty much assuming that these are all the possible unisex names in the US that are used.

Use the following link to provide a parsing of the first additional data set from Lab 1.
https://tinyurl.com/Parse-Worksheet

For Visualization Assignment select: **Lab assignment**
Then provide the parsed data and list your assumptions.

Part II: Additional Data Set #2

In this document:
Answer the following questions about your Tableau data:

1. List the name of the Tableau Dataset you selected in the Acquire Lab:
    a. Adjusted name combinations list
2. How many rows (records) are in the data set?
    a. There are 400 records in the data set.
3. How many columns (variables) are in the data set?
    a. There are 6 variables in the data set.
4. What assumptions are you making about the data?
    a. I am assuming that this data is not actually the real numbers of these specific first name and last name combinations. I am also assuming that the Adjustment column is there to compensate for the possible amount of unregistered people with that name, and that the adjustment will make the estimates more accurate instead of less accurate.

Use the following link to provide a parsing of the second additional data set from Lab 1.
https://tinyurl.com/Parse-Worksheet

For Visualization Assignment select: **Lab assignment**
Then provide the parsed data and list your assumptions.

What to submit:
1. This document saved as a PDF file: **LastnameFirstInitial-CGT270Fall2021-Lab2Parsing.pdf**
2. Parse Activity Worksheet (.pdf) for the Tableau data set.
3. Parse Activity Worksheet (.pdf) for 1st additional data set.
4. Parse Activity Worksheet (.pdf) for the 2nd additional data set.