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

Name: **Cristina Pascua**

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. 10506
3. How many columns (variables) are in the data set?
   a. 5
4. What assumptions are you making about the data?
   a. I assume there are records for each of the 50 states and each state contains the most popular names for male and female babies born for each year between 1910 to 2012.

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. NamesByState_CA
2. How many rows (records) are in the data set?
   a. 394179
3. How many columns (variables) are in the data set?

        a.   4
4. What assumptions are you making about the data?
        a.   I assume the data records all names and the number of babies with that name for every year between 1910 and 2020 in the state of California

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.   Top Baby Names
2. How many rows (records) are in the data set?
        a.   3003
3. How many columns (variables) are in the data set?
        a.   5
4. What assumptions are you making about the data?
        a.   I assume the data summarizes the top 25 baby names for males and females for each year between 1960 and 2019 in California.

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.