

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

Name: <Enguang Liu>

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. Answer the following questions and complete the table for each dataset.

1. List the name of the Tableau Dataset you selected in the Acquire Lab: TopBabyNamesbyState
2. How many rows (records) are in the data set? 10507
3. How many columns (variables) are in the data set? 5
4. What assumptions are you making about the data? People might don't like a public name and they do not want their baby's name appear too many times. The information tells me every different states could has their own preferred name.

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

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

Tableau Data Set

In the table below list each variable and its data type (add more rows as needed):

	Variable	Data type
1	State	String
2	Gender	Character
3	Year	Integer
4	Top name	String
5	Occurrence	Integer

Lab 2: Parsing Data

You may add more rows and attach additional pages if needed.

Lab 2: Parsing Data**Additional Data Set #1**

1. List the name of the first (1st) additional data set you acquired in the Acquire Lab:
name_gender_dataset
2. How many rows (records) are in the data set? 65535
3. How many columns (variables) are in the data set? 4
4. What assumptions are you making about the data? James are the most popular name in this data set, and Aarshi is the less popular name.

In the table below list each variable and its data type (add more rows as needed):

	Variable	Data type
1	Name	String
2	Gender	Character
3	Count	Integer
4	Probability	Float

You may add more rows and attach additional pages if needed.

Additional Data Set #2

1. List the name of the second (2nd) additional data set you acquired in the Acquire Lab: StateNames
2. How many rows (records) are in the data set? 65534
3. How many columns (variables) are in the data set? 6
4. What assumptions are you making about the data? I assume that all popular name in USA is around 65530. The most important variable should be count, since it shows the occurrence of each names.

In the table below list each variable and its data type (add more rows as needed):

Lab 2: Parsing Data

	Variable	Data type
1	ID	Integer
2	Name	String
3	Year	Integer
4	Gender	Character
5	State	String
6	Count	Integer

You may add more rows and attach additional pages if needed.