Assignment 6 – Text Tables & Data Joins

Professor John Sokol | Due 11/05

Background Information:

Data Joins:

In the business world, most of the time the data you're importing is not just on one spreadsheet. Data is usually divided into multiple spreadsheets in which all the data is connected via a primary key, or a common data field across spreadsheets. This can be explained visually:

Consider an IT analyst at AT&T analyzing consumer data. The analyst wants to import the following data into Tableau that is divided into two separate spreadsheets, user_useage.csv and user_device.csv:

User_device.csv

A	B	C	D	E
use_id	user_id	platform	platform_version	device
22782	26980	ios	10.2	iPhone7,2
22783	29628	android	6	Nexus 5
22784	28473	android	5.1	SM-G903F
22785	15200	ios	10.2	iPhone7,2
22786	28239	android	6	ONE E1003
22787	12921	android	4.3	GT-I9505
22788	28714	android	6	SM-G930F
22789	28714	android	6	SM-G930F
22790	29592	android	5.1	D2303
22791	28775	ios	10.2	iPhone6,2

User_useage.csv

use_id	outgoing_mins_per_month	outgoing_sms_per_month	monthly_mb
22787	21.97	4.82	1557.33
22788	1710.08	136.88	7267.55
22789	1710.08	136.88	7267.55
22790	94.46	35.17	519.12
22792	71.59	79.26	1557.33
22793	71.59	79.26	1557.33
22794	71.59	79.26	519.12
22795	71.59	79.26	519.12
22799	30.92	22.77	3114.67
22801	69.8	14.7	25955.55

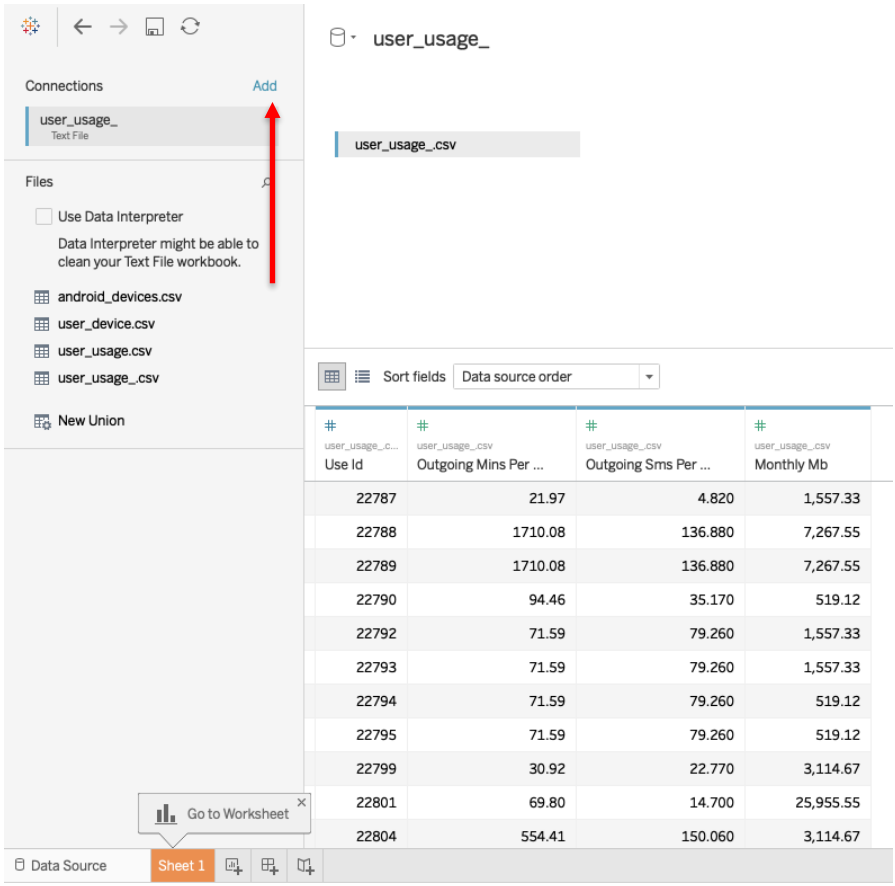
To address the practicality of dividing the data into multiple spreadsheets, there may be several reasons for doing so:

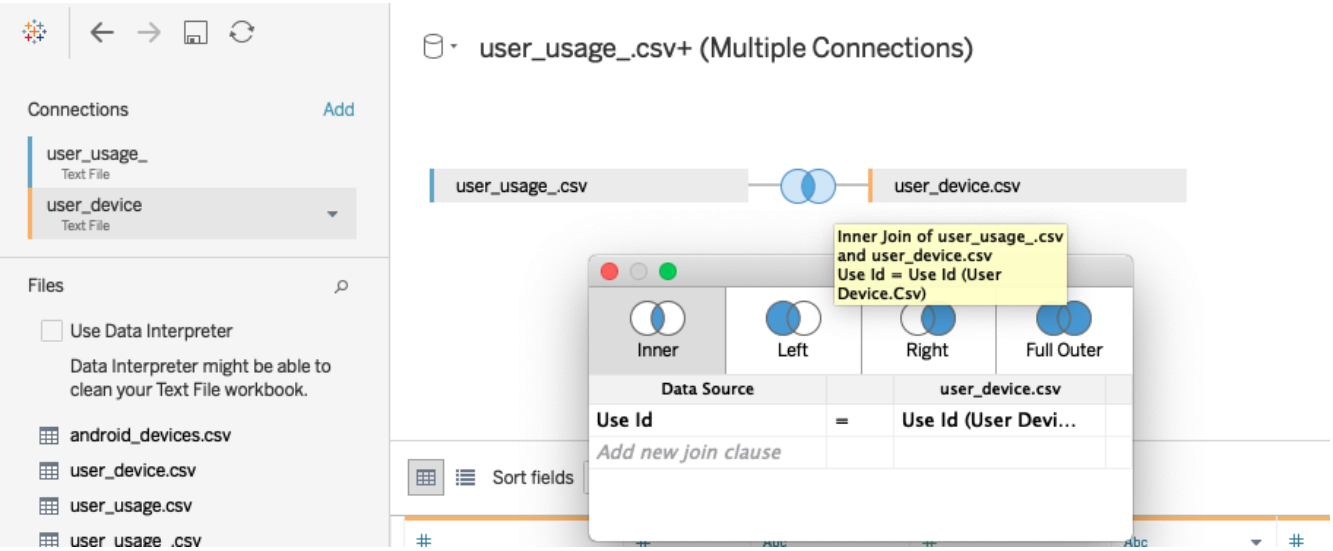
- Abundance of data: The organization has so much raw data, so many data fields, that it is impractical to place all of this data into one table.
- Organizational purposes: In the AT&T example above, if you're only attempting to visualize the distribution of the platform of phone owners (iOS or Android phones predominantly), then data fields such as *outgoing_mins_per_month* would be considered unnecessary for the visualization. But if you needed this data field for some reason, you can join the tables/spreadsheets together via the primary key, *use_id*.

Primary key:

- Common data field that is unique to each row in the table; should be numerical. Primary keys are data fields that are used to join tables/spreadsheets together.
- For example, Stockton University uses your Z number as a common numerical identifier for all the University databases. Stockton could technically use your last and first name for this, but what if they're multiple people with the same name? Plus, a unique numerical value is less prone to errors compared to string values.
- Primary key data fields do not have null values

Import into Tableau either dataset first. Then click 'Add' in the left window pane.
Import the other dataset.





Notice the blue Venn diagram visually shows the datasets joined together. Tableau automatically detects that the Use_id data field is present in both tables and determines the data field as an ideal candidate for joining. However, the data field that you join on can be edited if need be. Once the datasets are joined by an ideal common data field, you need to determine how to join the data together

There are four types of joins to choose from:

- Inner join
  - Selects records that have matching joining conditions in both tables.
  - Removes all rows that do not have matching joining conditions
- Left join
  - Returns all records from the left table, and the matched records from the right table (table2). The result is NULL from the right side, if there is no match.
  - Use left join over the right join.
- Right join
  - Returns all records from the right table (table2), and the matched records from the left table (table1). The result is NULL from the left side, when there is no match.
  - Not used often, if at all. Use left join.
- Full Outer join
  - Returns all the rows from the left table, and all the rows from the right table, regardless if the joining data fields match or not.

I understand that a written explanation is not the best way to convey these join concepts. Here is a visual that shows an example of the two most important joins, the inner join and left join:

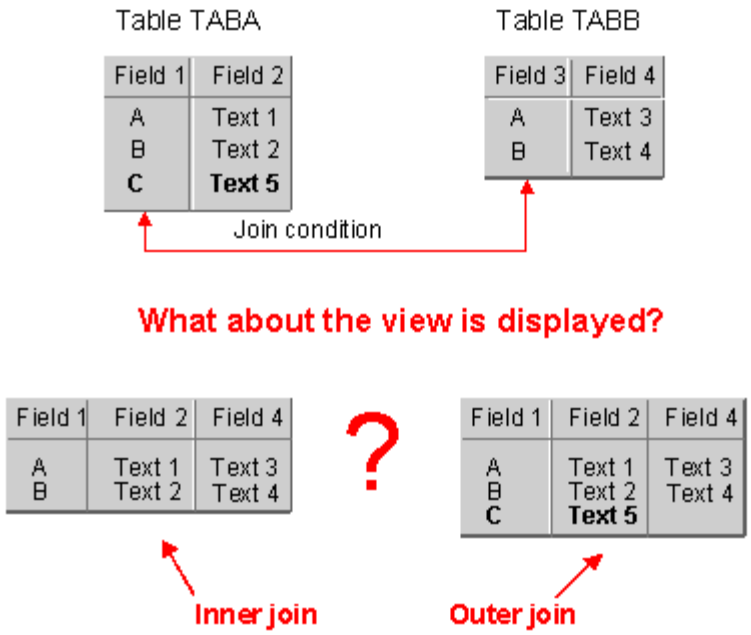


Table TABA and Table TABB are joined together by Field 1 and Field 3, respectively. FYI, the data fields that you are joining on do not have to be named the same.

Inner join

- Since the data value of C is in TABA, but not in TABB, this row is not included in the inner joined table.
- Another inner join example:

Dataset - A

ID	Name	Height
1	A	1
3	B	2
5	C	2
7	D	2
9	E	2

Dataset - B

ID	Name	Weight
2	A	2
4	B	3
5	C	4
7	D	5

Inner Join : Merged Dataset

ID	Name	Height	Weight
5	C	2	4
7	D	2	5

Left join

- With the left join, all of the data in the table that is joined on the left is included. However, if the joined data value is present in the left table but not in the right table, then the left data value is included, but none of the right table data is included, indicated by null values.
- You may see left joins be referred to as LEFT JOIN or LEFT OUTER JOIN; they're both the same, including OUTER is completely optional. This is also applicable to the right join.
- Another left join example:

In Tableau, the default join type is the inner join.
This Explanation of joins on Quora is also very cogent and helpful.

Tableau Text Tables:
Earlier in the semester, our textbook mentioned how text tables are NOT visualizations, and should preferably be avoided. But many people have become so accustomed to Excel spreadsheets that they request the addition of the raw data in the Tableau dashboard. Text tables can add content to your analysis and provide context for your visualizations. Therefore, creating text tables for use within dashboard is a very important skill to add to your Tableau repertoire.

I will be teaching Tableau text tables as an instructional video with the Sample Superstore dataset. This video is available on Blackboard.

Deliverables:

Data Joins

- Submit a Tableau *packaged* workbook that consists of both user_useage.csv and user_devices.csv imported. Inner join the two spreadsheets.

Text Tables

- In the same Tableau packaged workbook, create the same text table that I created with the Sample Superstore dataset.
- In the same Tableau packaged workbook, import the LBI dataset. Create a text table of the LBI dataset that consists of the following five data fields.
    - Owner
    - Home Address
    - Home Town
    - Home State
    - Home Zip
- All formatting practices as seen in the video must be adhered to for the text table.