

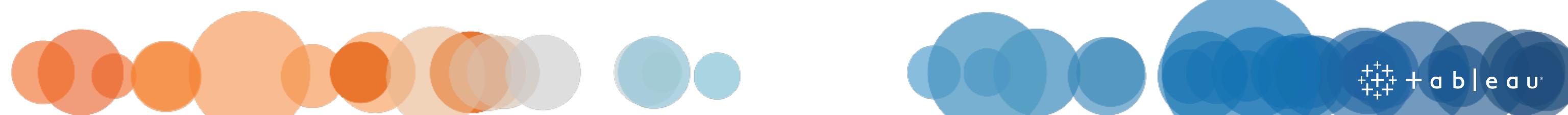


# Connecting to Data

## Data Prep with Text and Excel Files

Miriam McGaugh, PhD

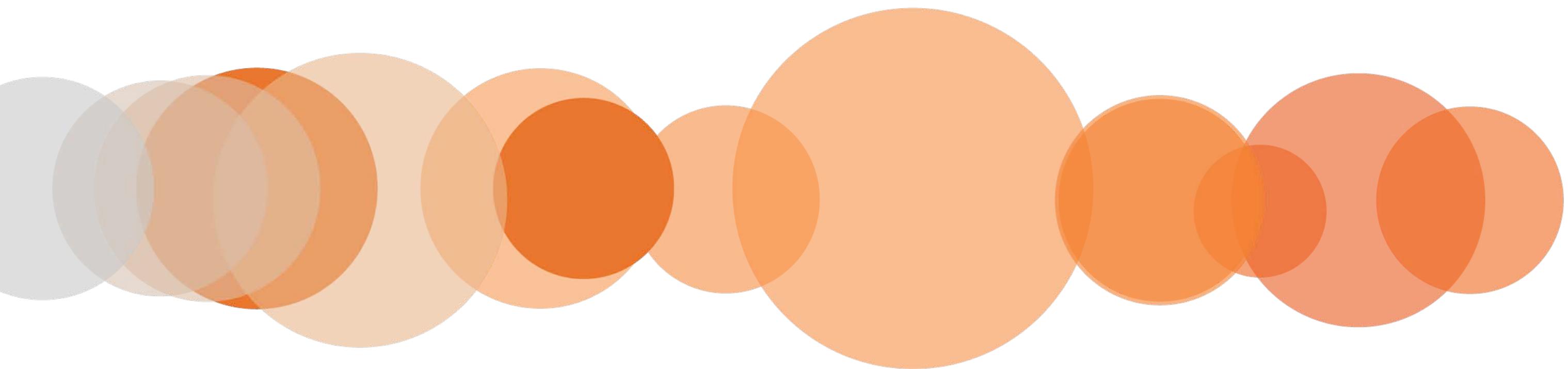
Clinical Assistant Professor, Marketing



# Table of Contents/Agenda

Poorly Formatted Data	Slide 3
Data Interpreter	Slide 9
Pivot	Slide 16
Split	Slide 19
Custom Split	Slide 23

# Poorly Formatted Data

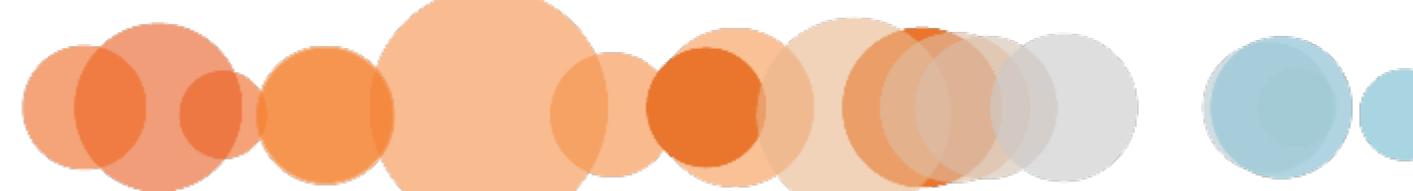


# Poorly Formatted Data

In our first example of connecting to data, we used an Excel file that was already formatted nicely – we just brought it straight into Tableau. In reality, data files are not always so well-formatted. Now we will practice using a file that needs some work before it is ready for analysis.

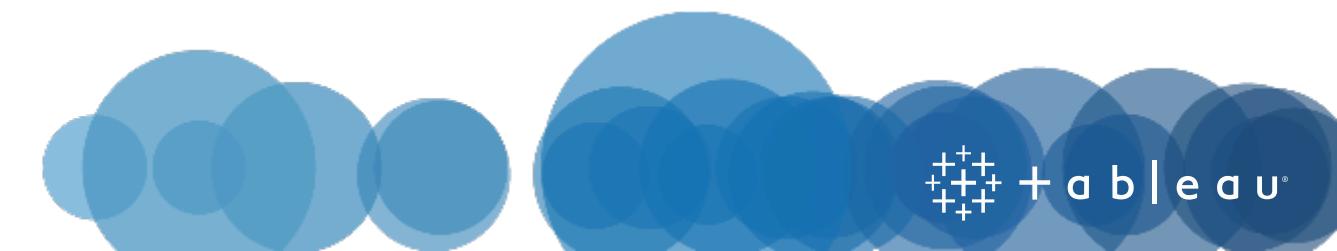
For this example we'll be using the “Data Prep – Flights” Excel file, which you can download from the course website. (Global Superstore, our main dataset, is too well-structured! This gives us a messier example to work with.)

# Poorly Formatted Data



	A	B	C	D	E
1	Date	Employee	Resolved Incidents		
2	1/1/2014	B-002		4	
3	1/1/2014	E-055		1	
4	1/1/2014	E-075		14	
5	1/1/2014	B-066		4	
6	1/1/2014	C-025		17	
7	1/1/2014	E-030		2	
8	1/1/2014	C-001		14	
9	1/1/2014	E-038		4	
10	1/1/2014	C-054		2	
11	1/1/2014	A-081		3	
12	1/1/2014	B-031		14	
13	1/1/2014	D-019		2	
14	1/1/2014	E-096		2	
15	1/1/2014	D-026		0	
16	1/1/2014	E-022		3	

**Open the file in Excel.** Here we have a report in Excel, showing the number of resolved incidents per Employee per month. The “Ideal” tab shows how we wish the data would be formatted – like a database table.



# Poorly Formatted Data

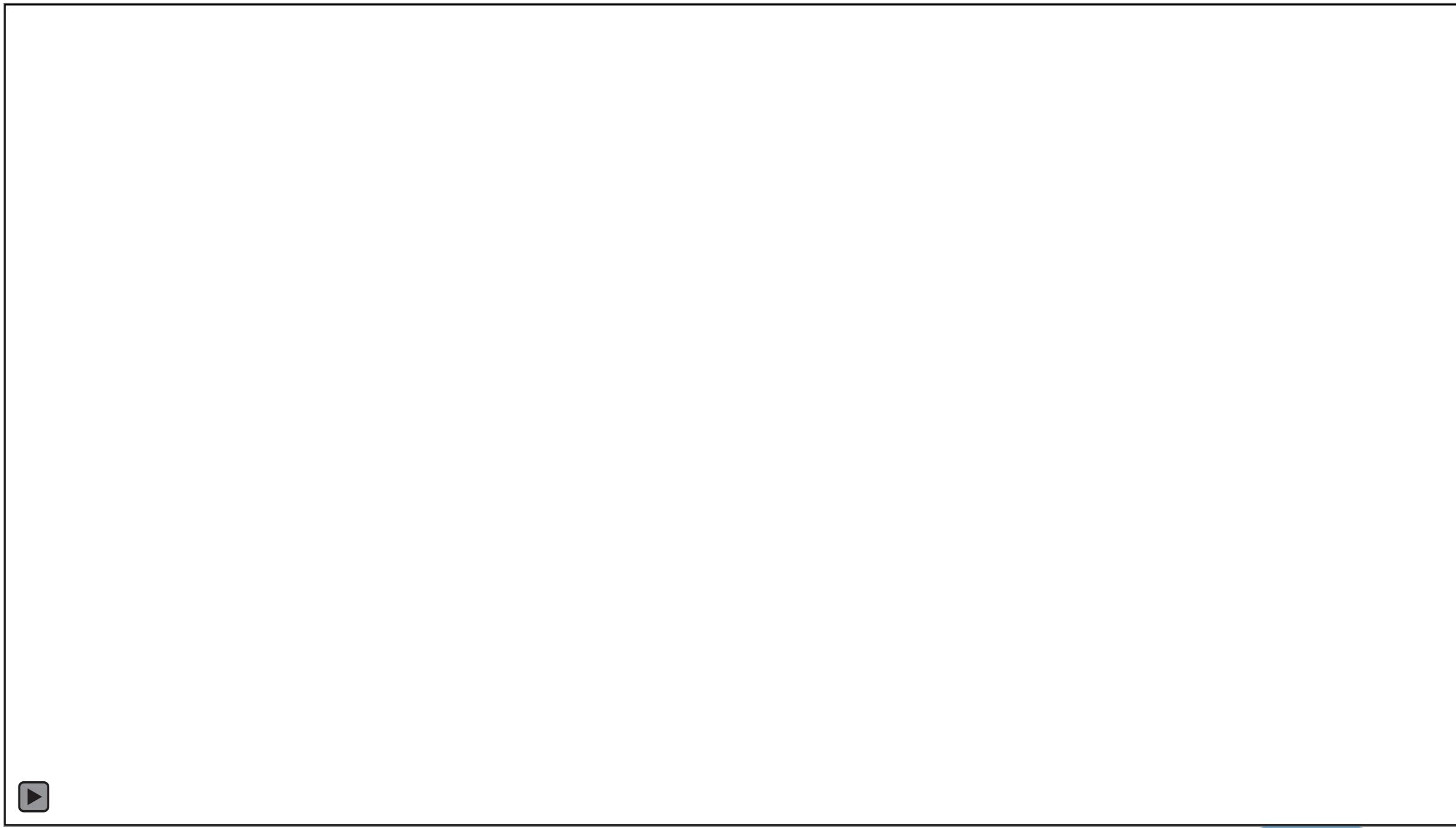
Flights Data Summary <span style="float: right;">B1</span>					
This report was generated on 1-1-15					
Employee	1/1/2014	2/1/2014	3/1/2014	4/1/2014	5/1/2014
B-002	4	1	5	2	3
E-055	1	2	1	3	4
E-075	14	17	16	15	18
B-066	4	4	5	2	5
C-025	17	13	17	18	17
E-030	2	2	1	1	0
C-001	14	14	14	14	13
E-038	4	1	0	4	0
C-054	2	5	4	4	2
A-081	3	2	4	5	2

However, sometimes we receive data that looks more like what we see in the “Resolved Incidents” tab. Luckily, there are several features in Tableau Desktop to help automatically reshape Text and Excel files to get them ready for analysis in Tableau.

# Poorly Formatted Data

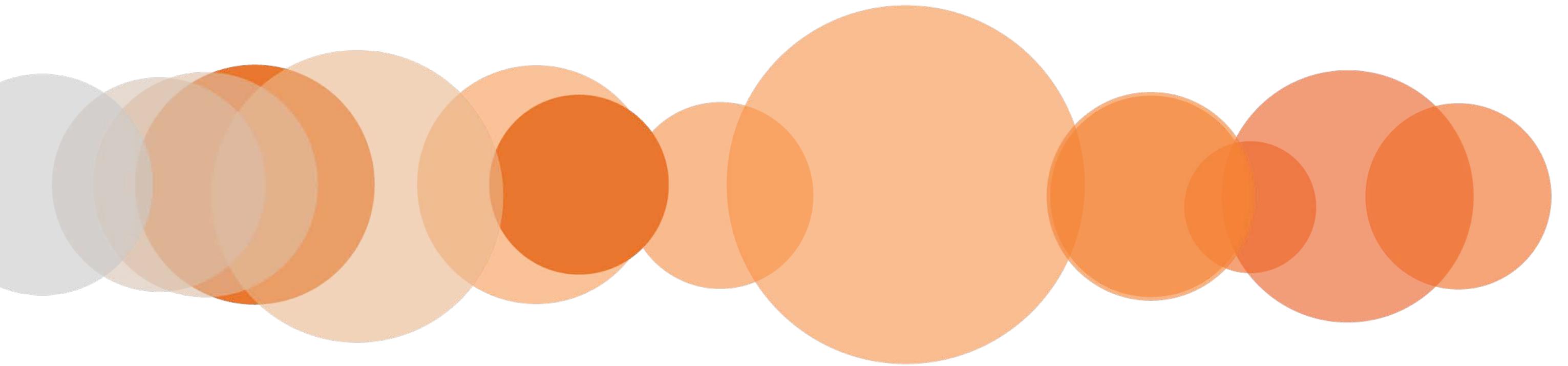
Let's connect to this Excel file and see if we can work with that poorly formatted sheet.

- In Tableau Desktop, **click on Excel, navigate to where you saved the file and click open.**
- **Drag out the “Resolved Incidents” sheet.**



tableau

# Data Interpreter

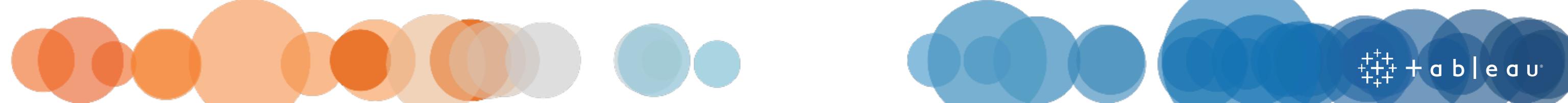


# Data Interpreter

Although Tableau can connect to this sheet, we can see here in the preview that there are some issues. There are no column names, the headers from Excel have a lot of nulls, and so on. Tableau has also recognized this and suggests the [Data Interpreter](#) (Tableau's built-in tool for preparing your data for analysis).

- Click turn on

# Data Interpreter



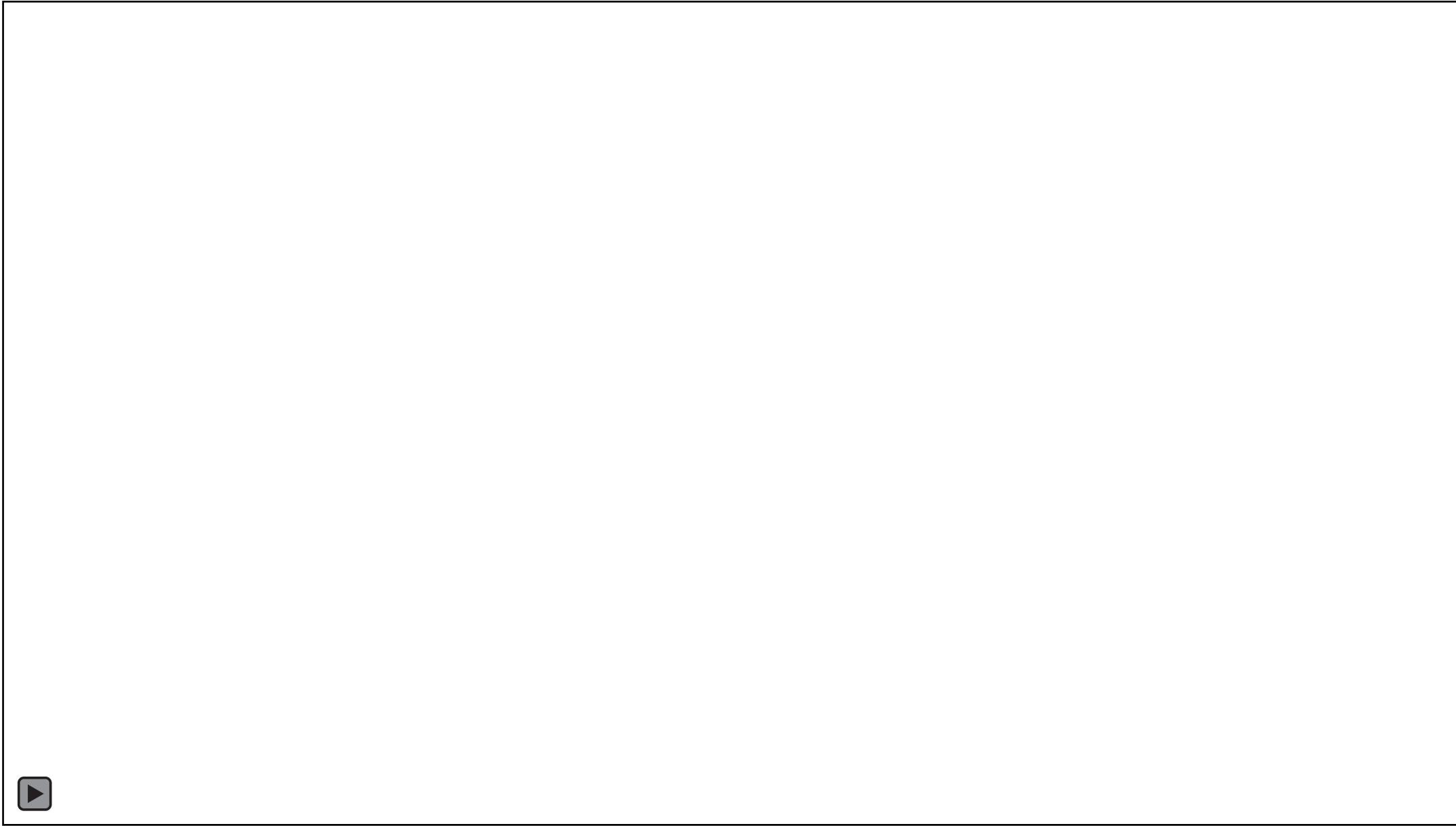
# Data Interpreter

- Now we see that those headers and nulls have been stripped out, and our columns are properly identified!

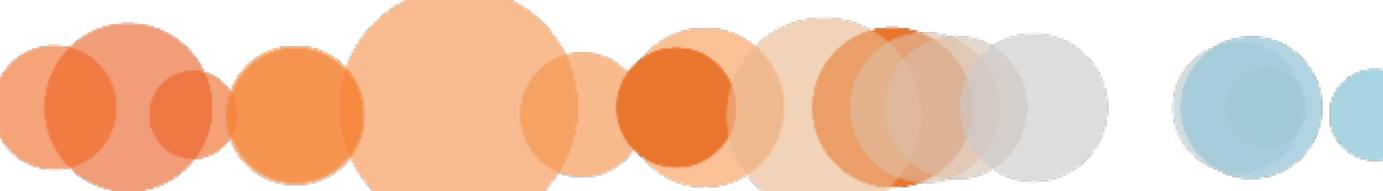
# Data Interpreter

- If we want more specifics on what the Data Interpreter did, we can **click “Review Results”** on the right. This will open an Excel file describing the changes.
- If we **click to the tab we used, Resolved Incidents**, we see which fields are being used as headers, in red, and which are considered data, in green

# Data Interpreter

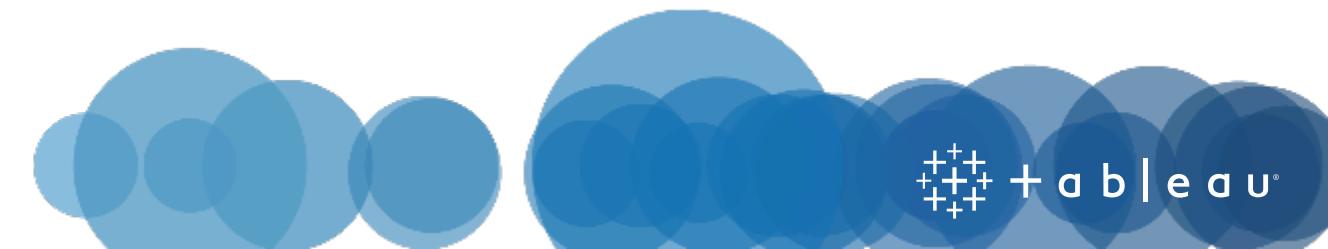


# Data Interpreter

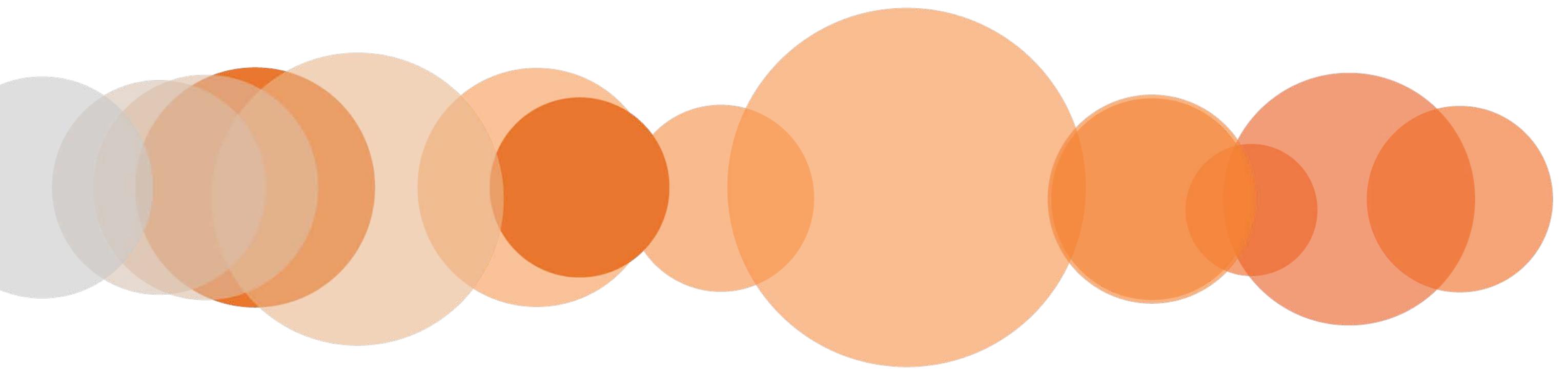


	A	B	C	D	E
1	Date	Employee	Resolved Incidents		
2	1/1/2014	B-002		4	
3	1/1/2014	E-055		1	
4	1/1/2014	E-075		14	
5	1/1/2014	B-066		4	
6	1/1/2014	C-025		17	
7	1/1/2014	E-030		2	
8	1/1/2014	C-001		14	
9	1/1/2014	E-038		4	
10	1/1/2014	C-054		2	
11	1/1/2014	A-081		3	
12	1/1/2014	B-031		14	
13	1/1/2014	D-019		2	
14	1/1/2014	E-096		2	
15	1/1/2014	D-026		0	
16	1/1/2014	E-022		3	

Before we go back to Tableau and our data connection, let's take one more look at that "Ideal" tab. Note that instead of having a column for each month with data underneath, in this format, there is a "Date" column and each row contains the number of resolved incidents for each unique combination of date and employee. This data is in the preferred format for analysis: taller, with more rows, rather than wider, with more columns. Let's see if we can do that in Tableau.



# Pivot

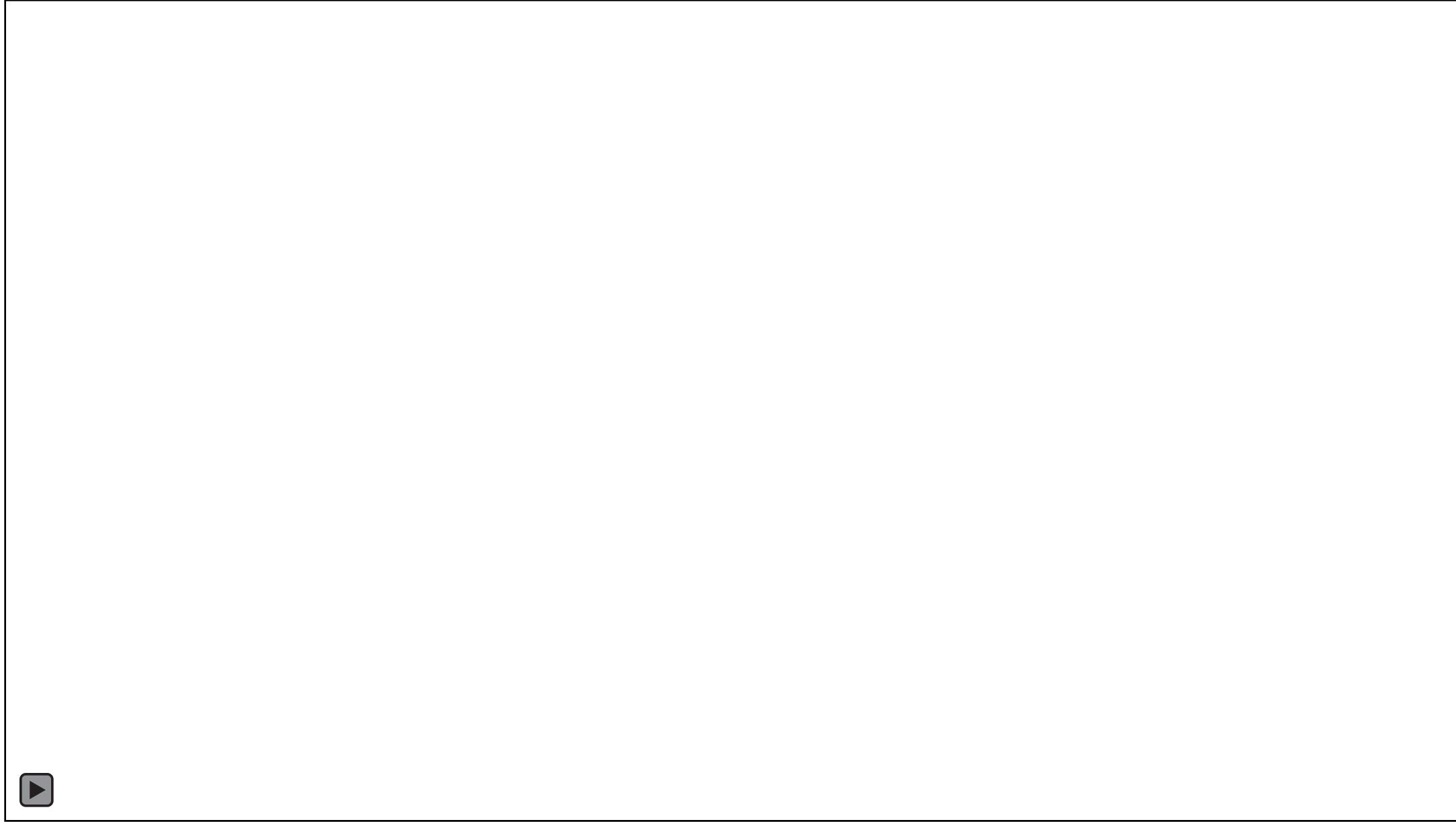


# Pivot

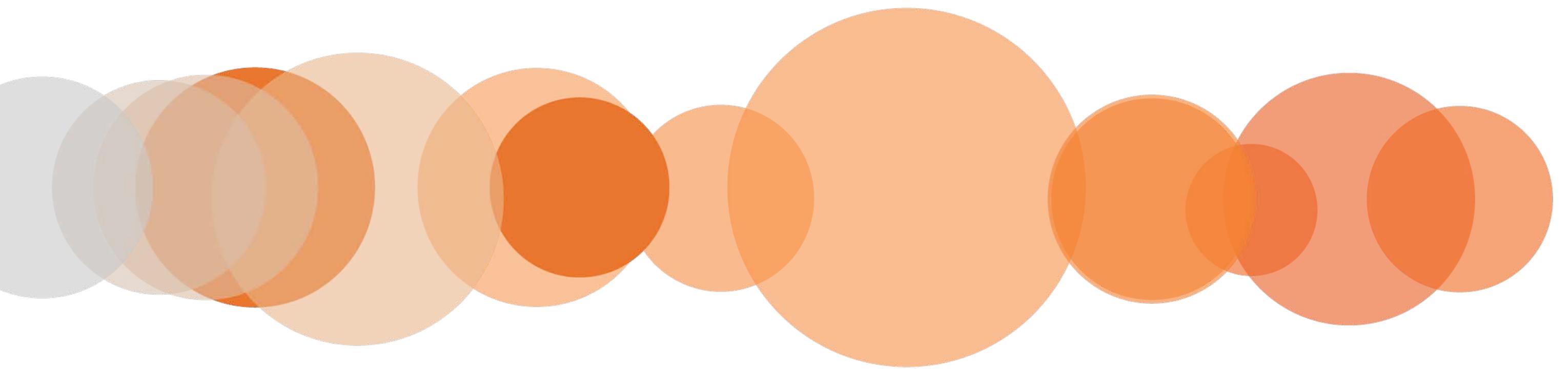
Back in Tableau, we want to change the format from that column-per-month layout into a single date column and a single column for Resolved Incidents.

- To do this easily, we'll simply **select all the date columns**. Click on the first, scroll if necessary, then shift click on the last. We'll **open the menu and select “Pivot”**
- This pivot feature essentially merges the information from the original columns and rows into two new columns – Pivot field names, and Pivot field values.
- We can see that “Pivot field names” is actually our Date, so we can **click to open the menu and select rename**.
- Similarly, “Pivot field values” can be **renamed “Resolved Incidents”**

# Pivot



Split



# Split

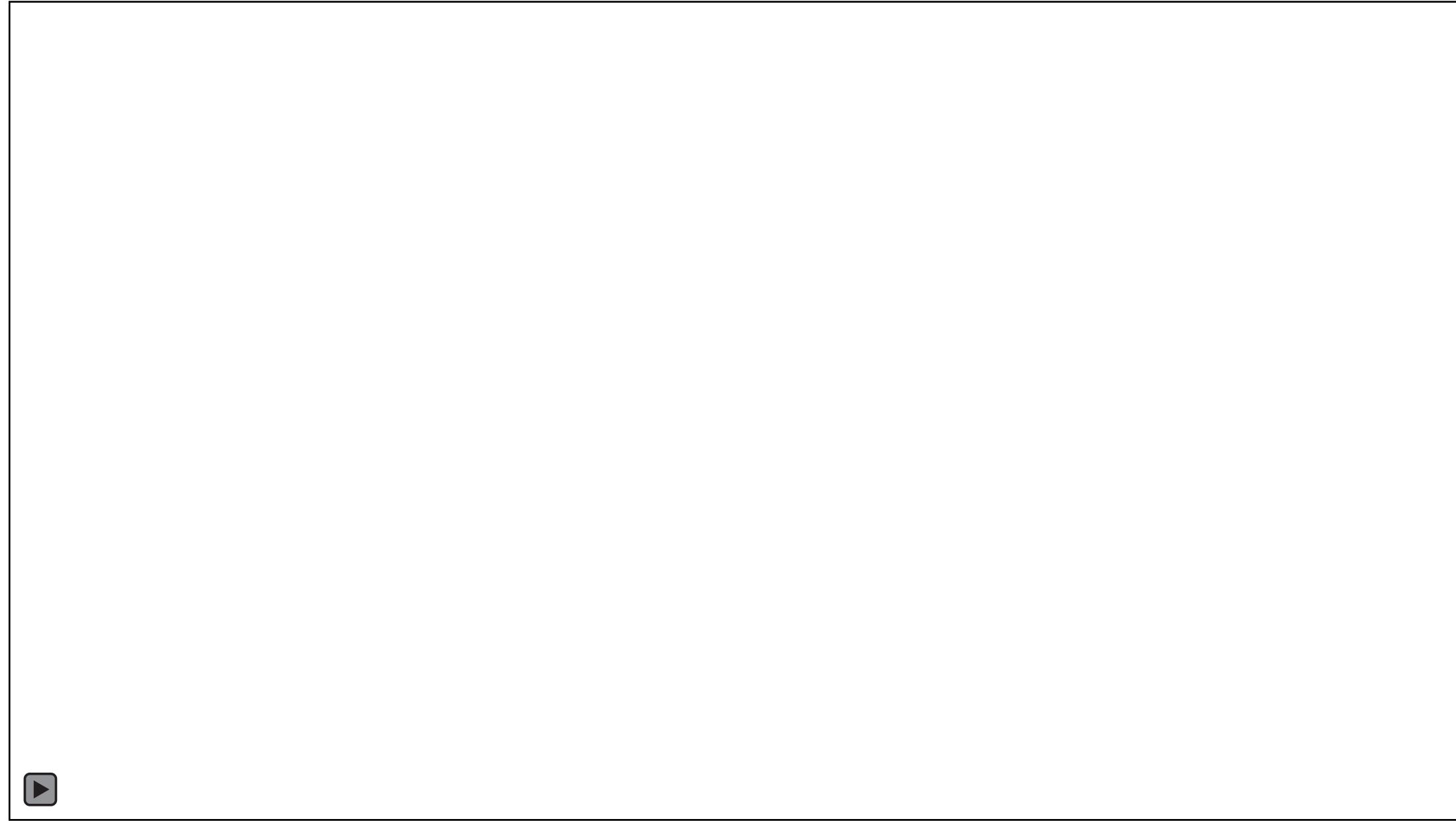
There's one more thing we can do to prepare this data. Note that the "Employee" field is actually two pieces of information – a location code, A, B, C, D, or E, followed by an Employee ID number. We can split the column based on the hyphen delimiter:

- Click to **open the menu and select Split**
- There are now two new fields – Employee - Split 1 and Split 2
- We'll use the **Metadata Grid view (click the icon to the left of Sort Fields)** to rename our split fields

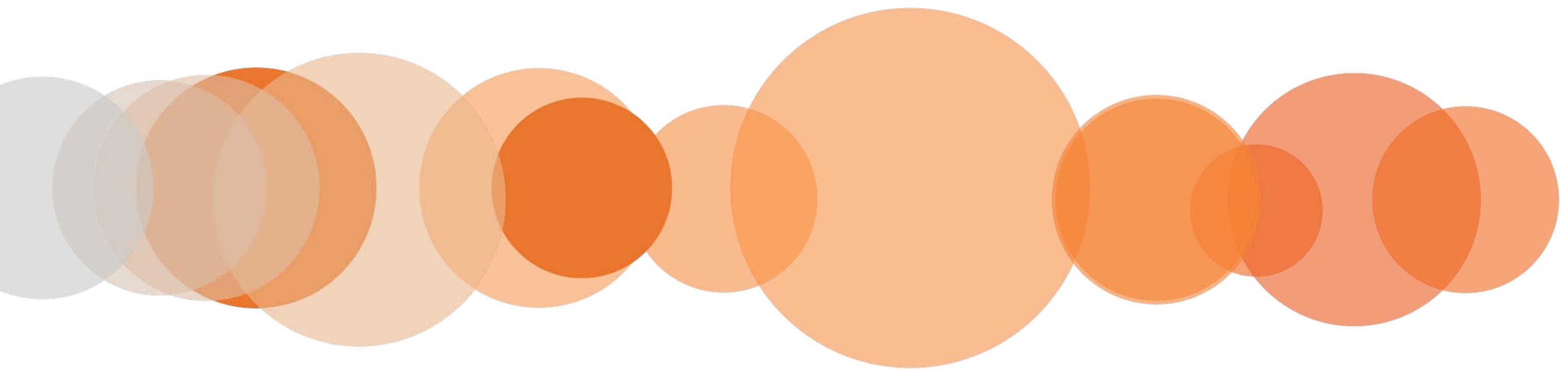
# Split

- **Click on the name** to edit in-line, **Split 1 should be Location**, and we'll hit tab, **Split 2 should be Employee ID**
- There's an Abc next to the Date field indicating this column is considered a String. We know it's actually a Date, though, so we can **click on the Abc and select Date** to update the data type.
- Now if we click on Sheet 1, we'll see nice tidy data ready for analysis!

# Split



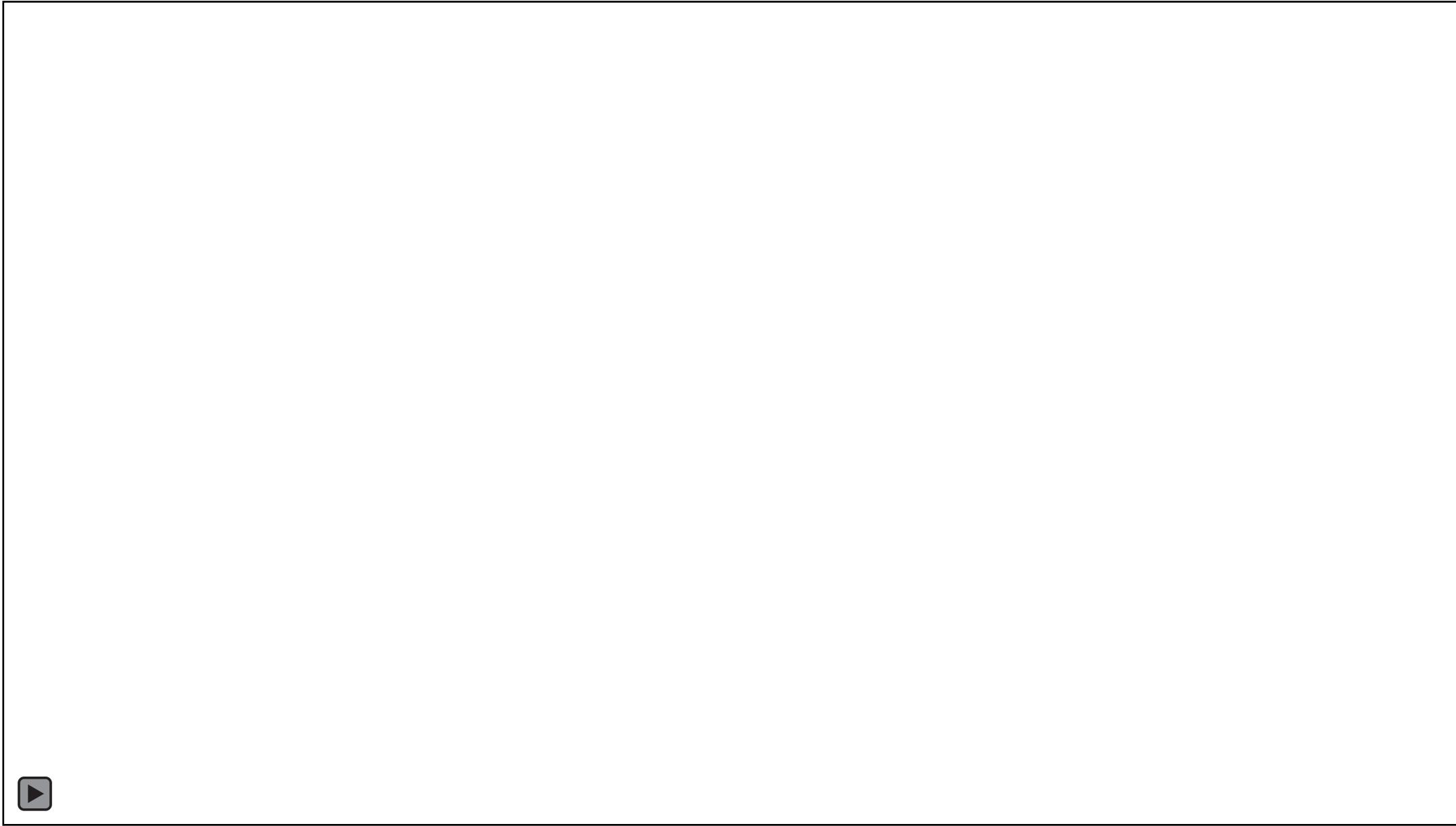
# Custom Split



# Custom Split

- Let's create a viz now: **bring Employee ID to the view, Resolved Incidents to Columns, and sort it.**
- It's clear from this view that there are really two groups of employees – some who resolve a much higher number of incidents than others. Looks like some employees are often able to get through more cases, and they have a Tier II designation.

# Custom Split



tableau

# Custom Split

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
	Employee	1/1/2014	2/1/2014	3/1/2014	4/1/2014	5/1/2014	6/1/2014	7/1/2014	8/1/2014	9/1/2014	#####	#####	#####	Header
1	B-002	4	1	5	2	3	0	3	1	2	0	2	5	Data
2	E-055	1	2	1	3	4	1	4	0	2	1	4	0	Data
3	E-075-II	14	17	16	15	18	16	14	17	12	13	14	12	Data
4	B-066	4	4	5	2	5	0	0	2	0	1	0	3	Data
5	C-025-II	17	13	17	18	17	17	12	15	17	17	14	15	Data
6	E-030	2	2	1	1	0	3	5	5	0	2	4	1	Data
7	C-001-II	14	14	14	14	13	18	17	14	13	18	15	14	Data
8	E-038	4	1	0	4	0	2	5	0	2	2	2	2	Data
9	C-054	2	5	4	4	2	3	0	5	5	5	3	5	Data
10	A-081	3	2	4	5	2	2	2	4	1	4	2	0	Data
11	B-031-II	14	14	14	14	15	13	15	14	12	16	12	18	Data
12	D-019	2	3	0	0	4	4	1	2	5	0	5	5	Data
13	E-096	2	0	4	4	5	3	3	0	5	4	2	0	Data
14	D-026	0	2	0	2	5	3	1	0	0	2	5	4	Data
15	E-022	3	3	4	3	4	2	0	3	2	3	3	1	Data
16	C-015	1	5	3	5	2	1	3	3	1	1	5	2	Data

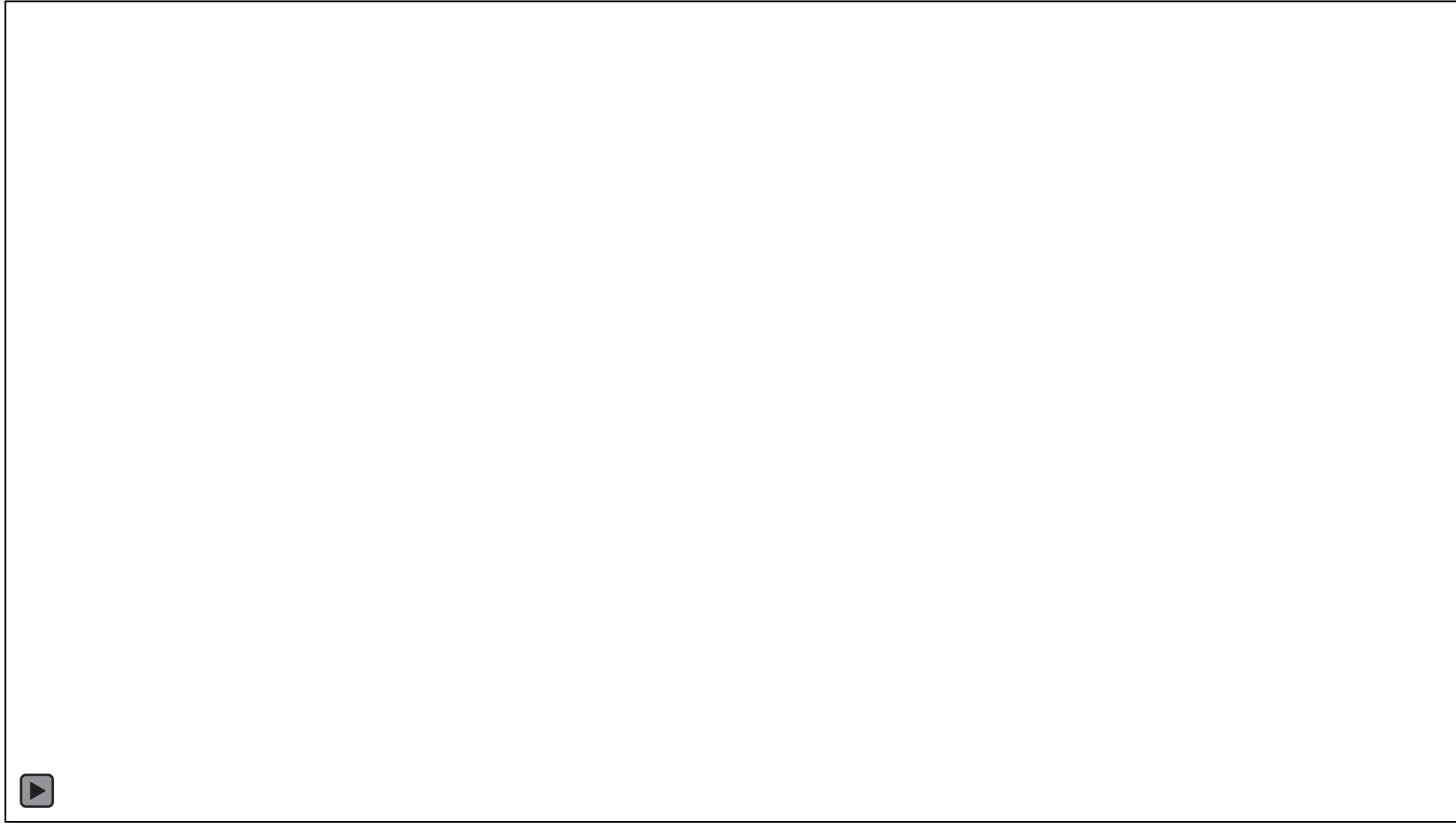
If we look at our original data set in Excel, we see there's a tab called Tiers. This report adds a -II to the end of an employee ID if they're tier II. Because not all rows have this -II, a standard split won't work. Let's see if we can create a viz that incorporates this Tier designation.

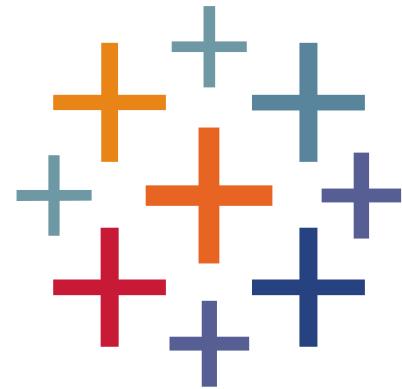
[[side note: both Split and Custom Split require consistent delimiters. If our data has irregular delimiters, Tableau won't be able to split out the data using these options.]]

# Custom Split

- Open a new Tableau file and recreate the viz, this time using the Tiers sheet from Excel.
- Remember to use the Data Interpreter, and Pivot the dates again
- Click on the Employee column to open the menu and select Custom Split
  - We can choose our delimiter, we'll use a hyphen
  - And now we can say we want to have 3 columns
  - This forces Tableau to break off that 3<sup>rd</sup> column with the tier II indicator
  - Now students should all try to finish the viz by themselves: rename all columns, and do your bar chart as above, and this time color the bars by Tier.

# Custom Split





+ a b l e a u®