



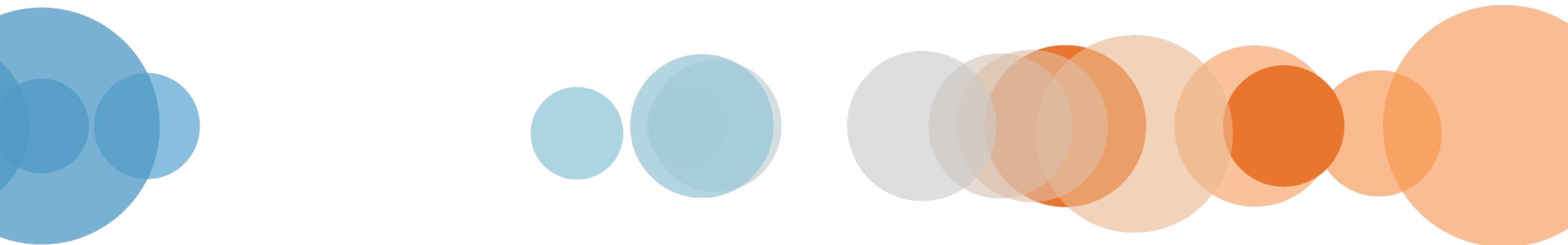
Getting Started with Data

[Presenter Name] | [Presenter Title]

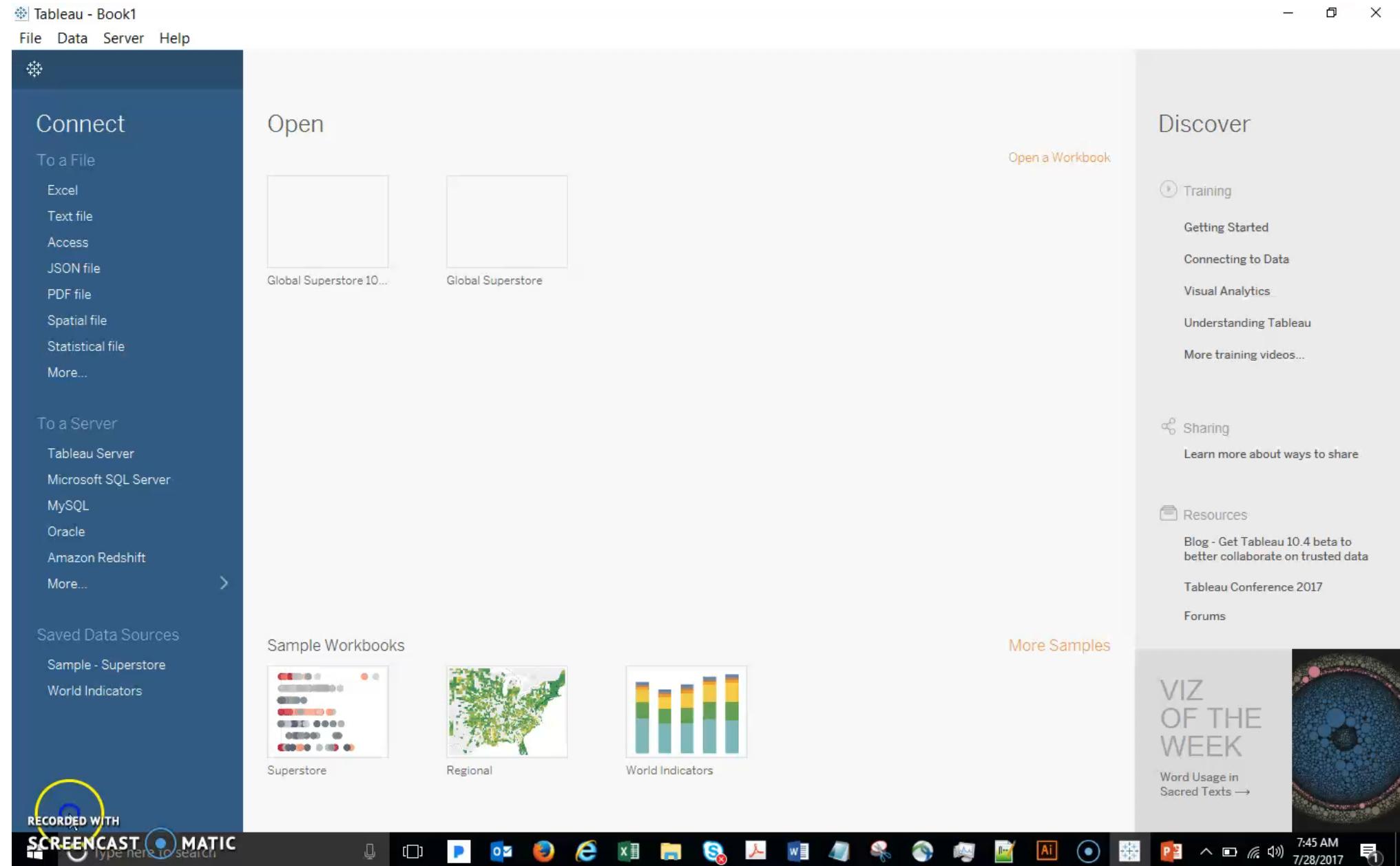
[Date] | [Contact Info (social media, email, etc.)]



Connect to Data Screen

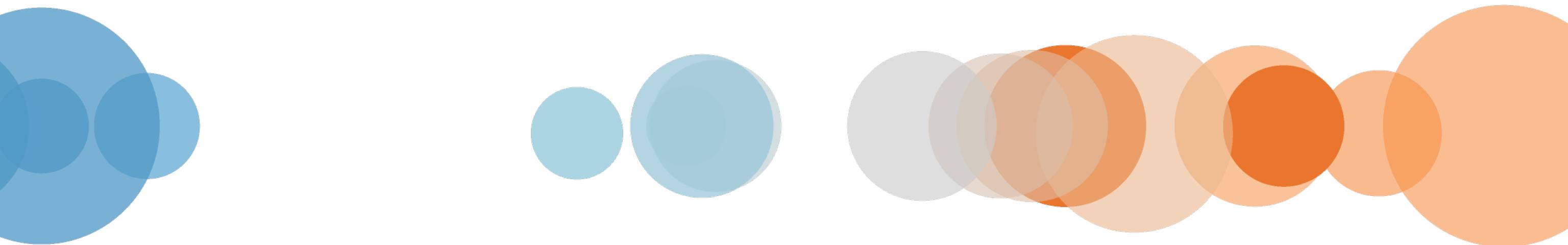


Connect to Data Screen



To connect to the Superstore file, click on Excel, navigate to where you saved the file and click open.

Connecting to Tableau



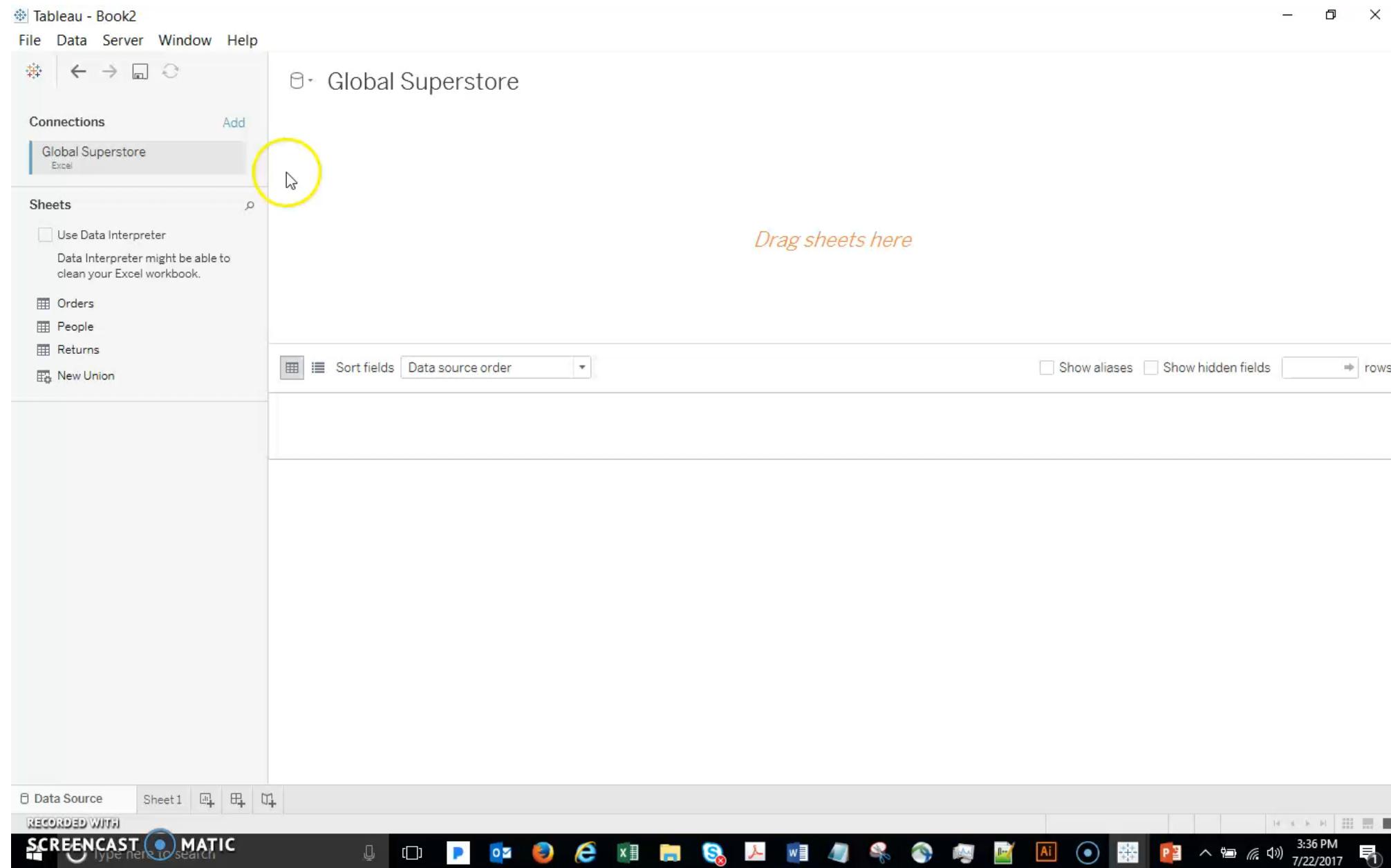
Connecting to Tableau

- Now Tableau brings us to the data connection window.
- Here we can see the name of the file – and here we can click to rename the connection if desired
- Down below, we can see all the sheets in the Excel file.
 - Sheets in Excel are treated the same as tables in databases, and we can choose to connect to a single table or join multiple tables.
- Simply **drag a sheet into the data connection canvas.**
 - Tables can be renamed simply by clicking on the name. The gear icon brings up options related to the data source.
- We can see the data down in the preview pane.

Connecting to Tableau

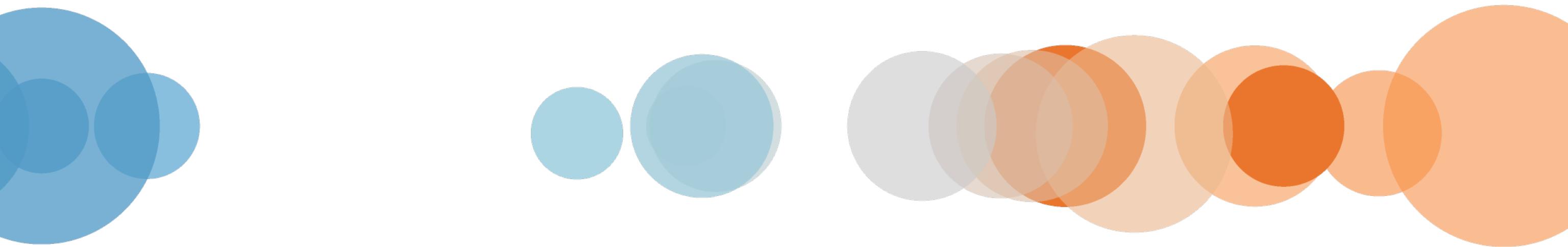
- There's a lot we can do on this screen (we will go over each of these in detail later in the class):
- If our column names aren't ideal, we can **click on the drop down arrow** to the right of the name and select rename.
- **Clicking on the data type icon** allows us to change the default data type for that column.
- If a field contains data that is concatenated, like Order ID and we'd like to split it, we can do that with the **Split option** – either custom or automatic.

Connecting to Tableau



tableau

Live versus Extract



Live versus Extract

Something to consider before we begin analyzing our data is if we want to connect live or extract.

- Connecting live leaves the data in the database or source file.
- This is best when we want to leverage a high performance database's capabilities, or to get up-to-the-second changes in data visualized in Tableau.
- A live data connection means that Tableau sends queries to your database and retrieves up to date data whenever the Tableau dashboard is updated.
- A live data connection is best and should be used when the live connection between the Tableau dashboard and your data is possible.
- That being said, sometimes connecting live can result in a slow experience, depending on the database.

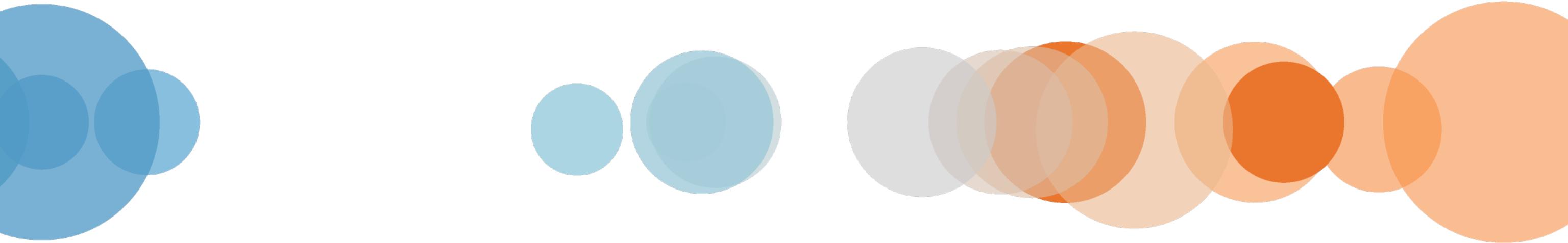
Live versus Extract

- The other option is to extract the data into Tableau's high performance in-memory data engine.
- Which means that Tableau sends queries to your extracted or static database and not the underlying data.
- Any updates to the underlying database will not update the Tableau dashboard. You will have to refresh the Tableau extract to get the up-to-date data.
- TDE should be used when a live connection to your data is not possible, when the live data connection is too slow, or in such cases where the size of the data is extremely large.
- In general, TDE should be used when your data is located on your own computer, a real data feed to your dashboard is not needed and when the data size is sufficiently large

When to use Live or Extract?

- A Tableau data extract is recommended when your database is too slow for interactive analytics
- or when you need to be offline and will not always have an internet or network connection to your data.
- A live connection is recommended when you have a fast database or when you need up to the minute
 - Bus schedules
 - Currency exchange

Connecting to Multiple Tables



Connecting to Multiple Tables

Next, what if we realize we need to bring in additional data?

- To add columns from *other tables* in the *same data source* (i.e. other worksheets in our Excel file), we need to edit our data connection.
- To do so, **click on the Data Source tab**.
- Let's join our *returns table* to the *orders table*.
 - Double click or **drag out Returns**.

Types of Tables Joining

Table A

id	First name	Last name
1	Aaron	Cab
2	John	Smith
3	Clark	Kent
6	Barry	Allen

Table B

id	Age	Place
1	30	CA
2	22	OH
3	33	DC
5	11	CE

Inner Join

id	First name	Last name	Age	Place
1	Aaron	Cab	30	CA
2	John	Smith	22	OH
3	Clark	Kent	33	DC

Types of Tables Joining

Table A

id	First name	Last name
1	Aaron	Cab
2	John	Smith
3	Clark	Kent
6	Barry	Allen

Table B

id	Age	Place
1	30	CA
2	22	OH
3	33	DC
5	11	CE

Left Join

id	First name	Last name	Age	Place
1	Aaron	Cab	30	CA
2	John	Smith	22	OH
3	Clark	Kent	33	DC
6	Barry	Allen	Null	Null

Types of Tables Joining

Table A

id	First name	Last name
1	Aaron	Cab
2	John	Smith
3	Clark	Kent
6	Barry	Allen

Table B

id	Age	Place
1	30	CA
2	22	OH
3	33	DC
5	11	CE

Right Join

id	First name	Last name	Age	Place
1	Aaron	Cab	30	CA
2	John	Smith	22	OH
3	Clark	Kent	33	DC
5	Null	Null	11	CE

Types of Tables Joining

Table A

id	First name	Last name
1	Aaron	Cab
2	John	Smith
3	Clark	Kent
6	Barry	Allen

Table B

id	Age	Place
1	30	CA
2	22	OH
3	33	DC
5	11	CE

Full Join

id	First name	Last name	Age	Place
1	Aaron	Cab	30	CA
2	John	Smith	22	OH
3	Clark	Kent	33	DC
5	Null	Null	11	CE
6	Barry	Allen	Null	Null

Connecting to Multiple Tables

- Click on the join icon to show the details.
- The default join clause is shown here.
- Tableau has figured out that Market is a common field between these two tables.
- If desired, we could edit the join clause, or even create a new one
- For these data we could also select a left join (explain the difference).
- For now we'll change it to Left and close the dialog.

Again, we can view the data down here and verify the data we'll be connecting to.

It looks good, so we'll click back onto our sheet.

We now have columns from both Orders and Returns in our data window and are ready to start our analysis!

Connecting to Multiple Tables

Tableau - Book1

File Data Worksheet Dashboard Story Analysis Map Format Server Window Help

Data Analytics

Pages

Columns Category

Rows AVG(Sales)

Search

Dimensions

- Category
- City
- Country
- Customer ID
- Customer Name
- Market
- Order Date
- Order ID
- Order ID - Split 1
- Order ID - Split 2
- Order ID - Split 3
- Order Priority
- Postal Code
- Product ID
- Product Name
- Region
- Segment

Marks

- Automatic
- Color
- Size
- Label
- Detail
- Tooltip

Sum(Sales)

Sum(Number ...)

Sheet 1

Category

Avg.Sales

Technolo... Furniture Office Supplies

467.9 416.2 121.1

For horizontal bars try

0 or more Dimensions

1 or more Measures

Measure Values

Data Source Sheet 1

RECORDED WITH 10 columns SUM of AVG(Sales): 1,005.2

SCREENCASTOMATIC

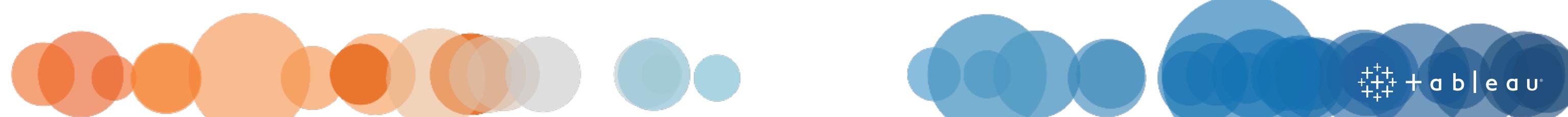
3:27 PM 7/22/2017

Category	Avg.Sales
Technology	467.9
Furniture	416.2
Office Supplies	121.1

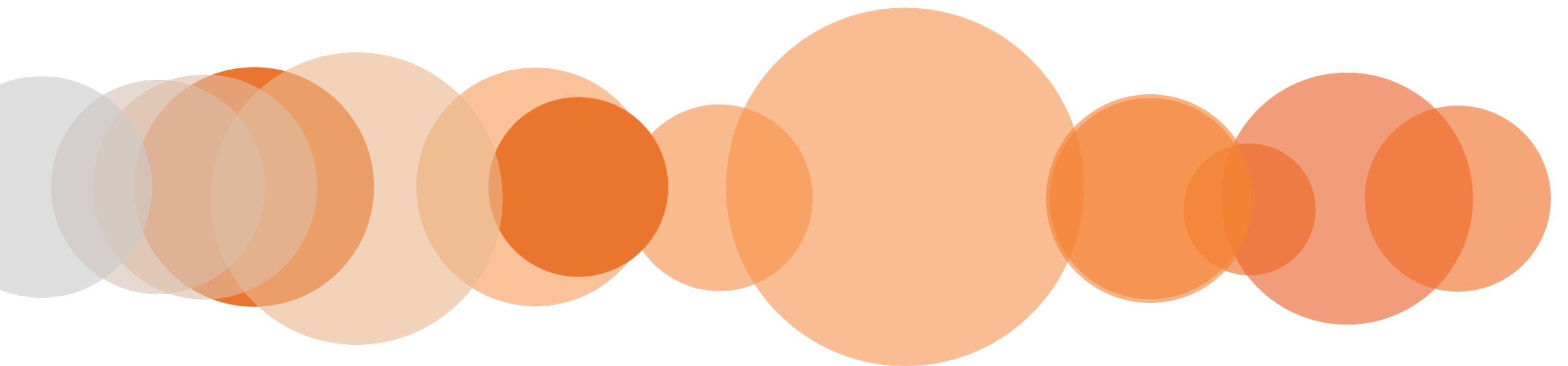


Connecting to Data

Data Prep with Text and Excel Files



Poorly Formatted Data

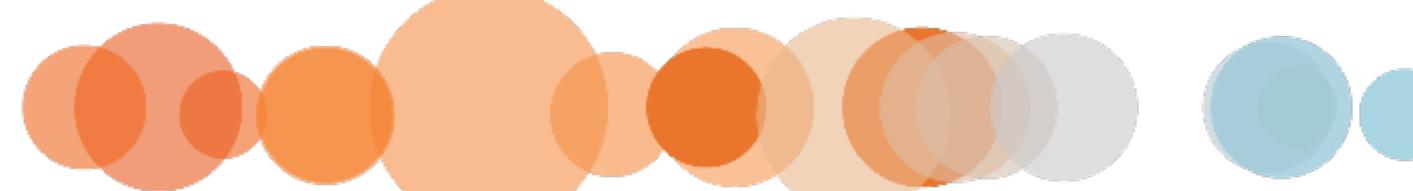


Poorly Formatted Data

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	F-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 B1					
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.**

File Data Server Help

Connect

To a File

- Excel
- Text file
- Access
- JSON file
- PDF file
- Spatial file
- Statistical file
- More...

To a Server

- Tableau Server
- MySQL
- Oracle
- Amazon Redshift
- Microsoft SQL Server
- More...

Sample Workbooks

- Superstore
- Regional
- World Indicators

Open

- Global Superstore
- SUMMER TEST ...
- Enrollment Com...
- Academic Depar...
- Course Assessm...
- Course Assessm...
- Analysis of Degr...
- ECDB Sankey
- 12 Term Major C...
- Program Array ...
- Program Array ...
- Enrollment Com...
- Enrollment Clas...
- HG Grads - Post ...

Discover

Open a Workbook

- Training
- Getting Started
- Connecting to Data
- Visual Analytics
- Understanding Tableau
- More training videos...

Viz of the Week

- Data Role T Models

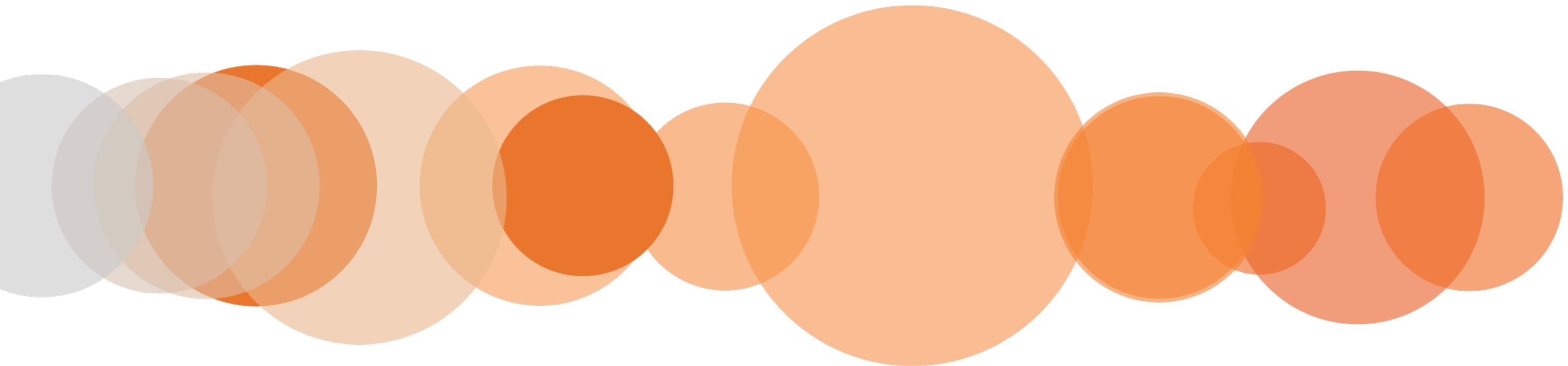
Resources

- Blog - Get Tableau 10.4 beta to better collaborate on trusted data
- Tableau Conference 2017
- Forums

RECORDED WITH
2017 SCREENCASTOMATIC

++
++
++
+ a b l e a u

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

File Data Server Window Help

Connections Add

Copy of Data Prep - Flights Excel

Sheets

Use Data Interpreter
Data Interpreter might be able to clean your Excel workbook.

Ideal

Irregular Delimiter

Resolved Incidents

Tiers

New Union

Resolved Incidents (Copy of Data Prep - Flights)

Connection Live Extract

Filters 0 | Add

Resolved Incidents

Sort fields Data source order ▾

Show aliases Show hidden fields 32 ➔

Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc	Abc
Resolved Incidents	Resolved Inci...											
F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	Resolve
Flights Data Summary	null	n										
This report was gener...	null	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	10/1/2014	11/1/2014	
B-002	4	1	5	2	3	0	3	1	2	0	2	2
E-055	1	2	1	3	4	1	4	0	2	1	4	
E-075	14	17	16	15	18	16	14	17	12	13	14	
B-066	4	4	5	2	5	0	0	2	0	1	0	
C-025	17	13	17	18	17	17	12	15	17	17	14	
E-030	2	2	1	1	0	3	5	5	0	2	4	

RECORDED WITH SCREENCASTOMATIC

Go to Worksheet

Data Source

Sheet1

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

File Data Server Window Help

← → ⌂ ⌂

Connections Add

Copy of Data Prep - Flights Excel

Cleaned with Data Interpreter
[Review the results](#). (To undo changes, clear the check box.)

Ideal Irregular Delimiter Resolved Incidents Tiers New Union

Resolved Incidents

Connection Live Extract Filters 0 | Add

Sort fields Data source order ▾ Show aliases Show hidden fields 26 ➔

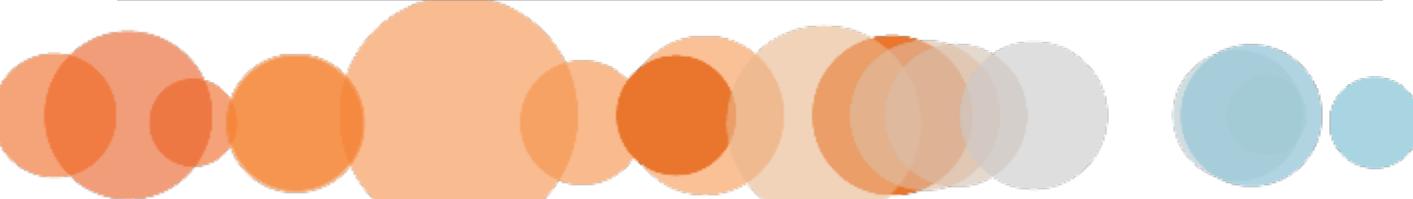
Employee	1/1/2014	2/1/2014	3/1/2014	4/1/2014	5/1/2014	6/1/2014	7/1/2014	8/1/2014	9/1/2014
Resolved Incidents									
B-002	4	1	5	2	3	0	3	1	
E-055	1	2	1	3	4	1	4	0	
E-075	14	17	16	15	18	16	14	17	
B-066	4	4	5	2	5	0	0	2	
C-025	17	13	17	18	17	17	12	15	
E-030	2	2	1	1	0	3	5	5	
C-001	14	14	14	14	13	18	17	14	
E-038	4	1	0	4	0	2	5	0	
C-054	2	5	4	4	2	3	0	5	

RECORDED WITH SCREENCASTOMATIC

Go to Worksheet Data Source Sheet1

tableau

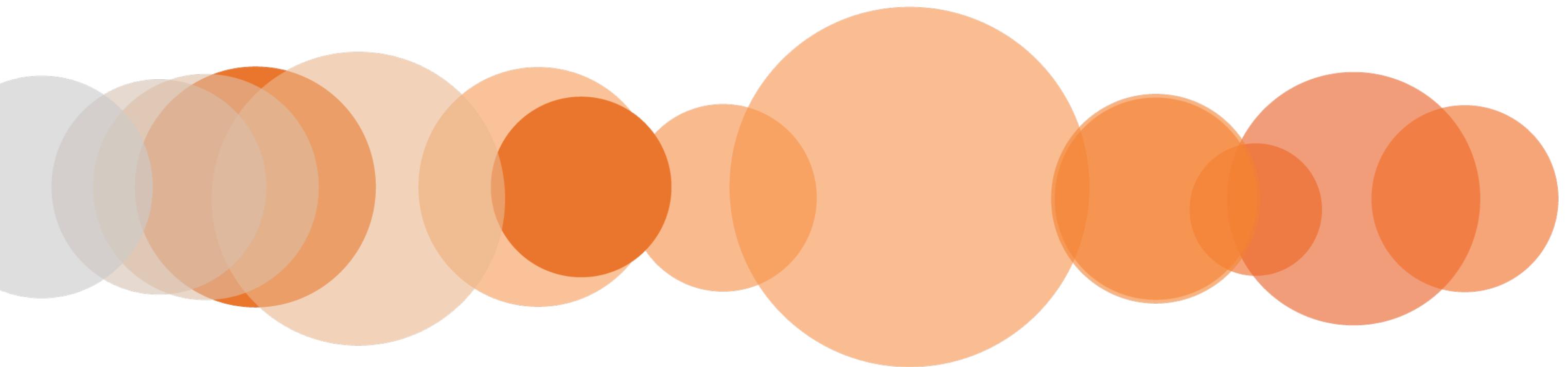
Data Interpreter



	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	F-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

File Data Server Window Help

Connections Add

Copy of Data Prep - Flights Excel

Cleaned with Data Interpreter Review the results. (To undo changes, clear the check box.)

Ideal

Irregular Delimiter

Resolved Incidents

Tiers

New Union

Resolved Incidents (Copy of Data Prep - Flights)

Connection Live Extract Filters 0 | Add

Resolved Incidents

Sort fields Data source order ▾ Show aliases Show hidden fields 26 ➔

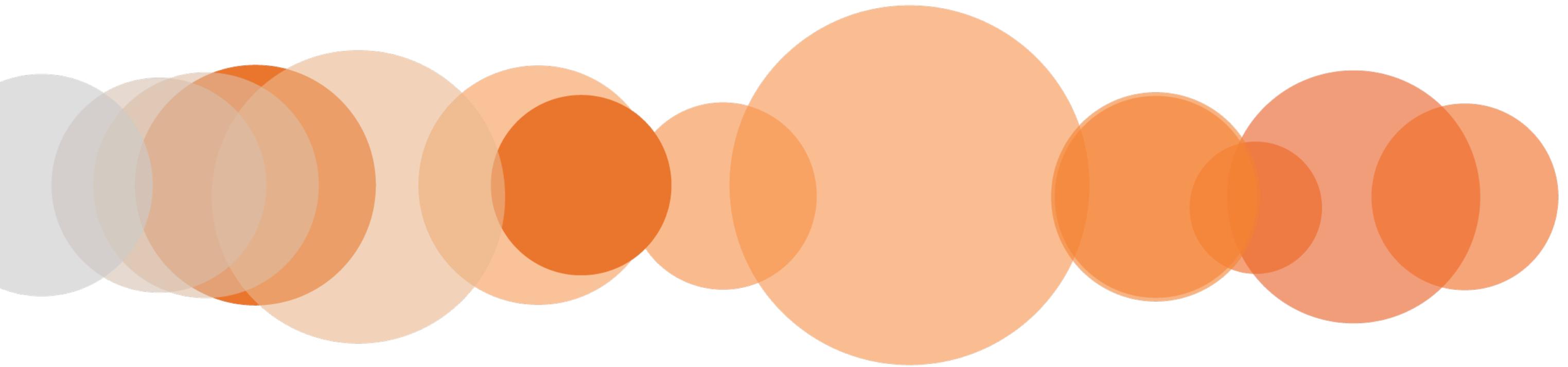
Abc	#	#	#	#	#	#	#	#	#	#
Resolved Incidents	Resolved In									
Employee	1/1/2014	2/1/2014	3/1/2014	4/1/2014	5/1/2014	6/1/2014	7/1/2014	8/1/2014	9/1/2014	9/1/2014
B-002	4	1	5	2	3	0	3	1		
E-055	1	2	1	3	4	1	4	0		
E-075	14	17	16	15	18	16	14	17		
B-066	4	4	5	2	5	0	0	2		
C-025	17	13	17	18	17	17	12	15		
E-030	2	2	1	1	0	3	5	5		
C-001	14	14	14	14	13	18	17	14		
E-038	4	1	0	4	0	2	5	0		
C-054	2	5	4	4	2	3	0	5		

RECORDED WITH SCREENCASTOMATIC

Go to Worksheet Data Source Sheet1

++
++
+ a b | e a u

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

File Data Server Window Help

Connections Add

Copy of Data Prep - Flights Excel

Cleaned with Data Interpreter
Review the results. (To undo changes, clear the check box.)

Ideal
Irregular Delimiter
Resolved Incidents
Tiers
New Union

Resolved Incidents (Copy of Data Prep - Flights)

Connection Live Extract Filters 0 | Add

Resolved Incidents

Sort fields Data source order ▾

Show aliases Show hidden fields 312 ➔

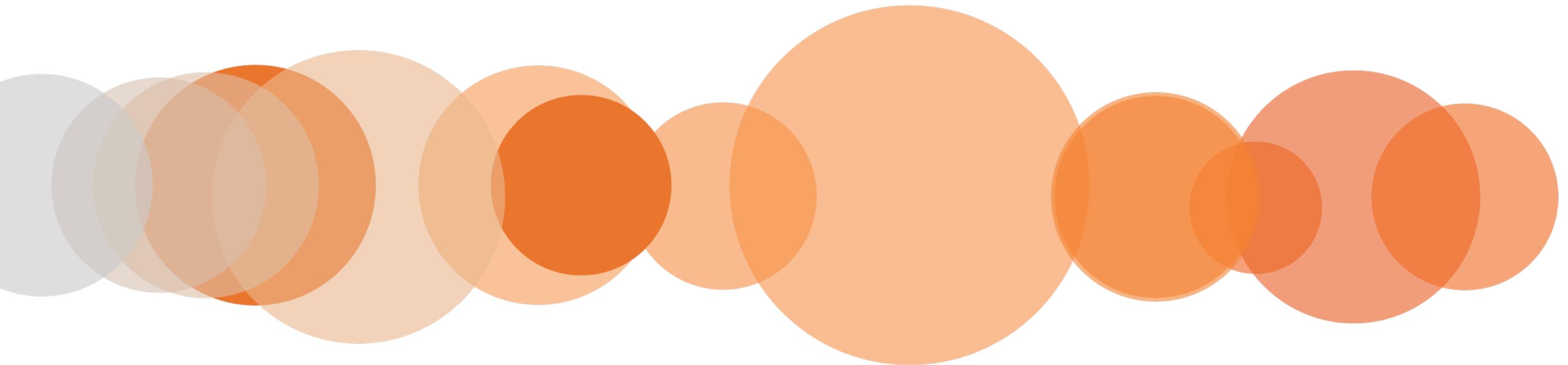
Date	Resolved Incidents	Employee
1/1/2014	4	B-002
1/1/2014	1	E-055
1/1/2014	14	E-075
1/1/2014	4	B-066
1/1/2014	17	C-025
1/1/2014	2	E-030
1/1/2014	14	C-001
1/1/2014	4	E-038
1/1/2014	2	C-054

RECORDED WITH SCREENCASTOMATIC

Go to Worksheet Data Source Sheet1

tableau

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

The screenshot shows the Tableau desktop interface with a custom split sheet configuration. The top navigation bar includes File, Data, Worksheet, Dashboard, Story, Analysis, Map, Format, Server, Window, and Help. The left sidebar displays the Data pane with a single item: Resolved Incidents (Cop...). The Dimensions pane lists Date, Employee, Employee ID (highlighted with a yellow circle), Location, and Measure Names. The Measures pane lists Resolved Incidents, Number of Records, and Measure Values. The central workspace is titled "Sheet 1" and contains three blank drop zones labeled "Drop field here". The bottom status bar indicates the analysis was recorded with SCREENCASTOMATIC, shows the current sheet as "Sheet 1", and includes icons for Data Source, Refresh, and Undo/Redo.

++
++
++
+ a b l e a u °

Custom Split

Flights Data Summary													
This report was generated on 1-1-15													
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
B-002	4	1	5	2	3	0	3	1	2	0	2	5	Data
E-055	1	2	1	3	4	1	4	0	2	1	4	0	Data
E-075-II	14	17	16	15	18	16	14	17	12	13	14	12	Data
B-066	4	4	5	2	5	0	0	2	0	1	0	3	Data
C-025-II	17	13	17	18	17	17	12	15	17	17	14	15	Data
E-030	2	2	1	1	0	3	5	5	0	2	4	1	Data
C-001-II	14	14	14	14	13	18	17	14	13	18	15	14	Data
E-038	4	1	0	4	0	2	5	0	2	2	2	2	Data
C-054	2	5	4	4	2	3	0	5	5	5	3	5	Data
A-081	3	2	4	5	2	2	2	4	1	4	2	0	Data
B-031-II	14	14	14	14	15	13	15	14	12	16	12	18	Data
D-019	2	3	0	0	4	4	1	2	5	0	5	5	Data
E-096	2	0	4	4	5	3	3	0	5	4	2	0	Data
D-026	0	2	0	2	5	3	1	0	0	2	5	4	Data
E-022	3	3	4	3	4	2	0	3	2	3	3	1	Data
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 3rd 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

File Data Server Help

Connect

To a File

- Excel
- Text file
- Access
- JSON file
- PDF file
- Spatial file
- Statistical file
- More...

To a Server

- Tableau Server
- MySQL
- Oracle
- Amazon Redshift
- Microsoft SQL Server
- More...

Saved Data Sources

Sample - Superstore

RECORDED WITH
SCREENCASTOMATIC

Open

Step 1 Viz

Global Superstore

SUMMER TEST ...

Enrollment Com...

Academic Depar...

Academic Depar...

Course Assessm...

Course Assessm...

Analysis of Degr...

ECDB Sankey

12 Term Major C...

Program Array ...

Program Array ...

Enrollment Com...

Enrollment Clas...

Sample Workbooks

Superstore

Regional

World Indicators

Discover

Open a Workbook

Training

Getting Started

Connecting to Data

Visual Analytics

Understanding Tableau

More training videos...

Viz of the Week

Data Role T Models

Resources

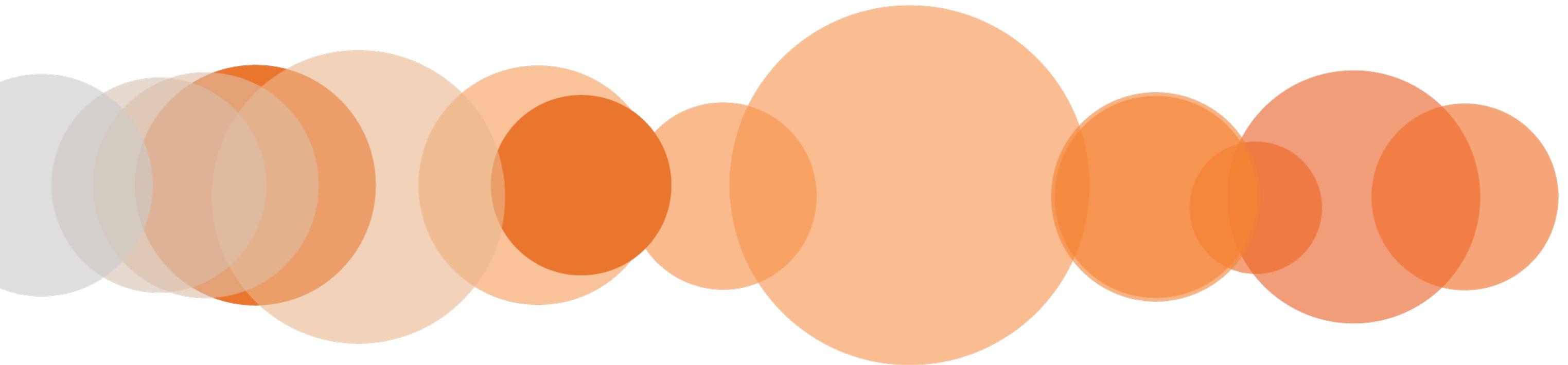
Blog - Get Tableau 10.4 beta to better collaborate on trusted data

Tableau Conference 2017

Forums

++
++
++
+ a b l e a u

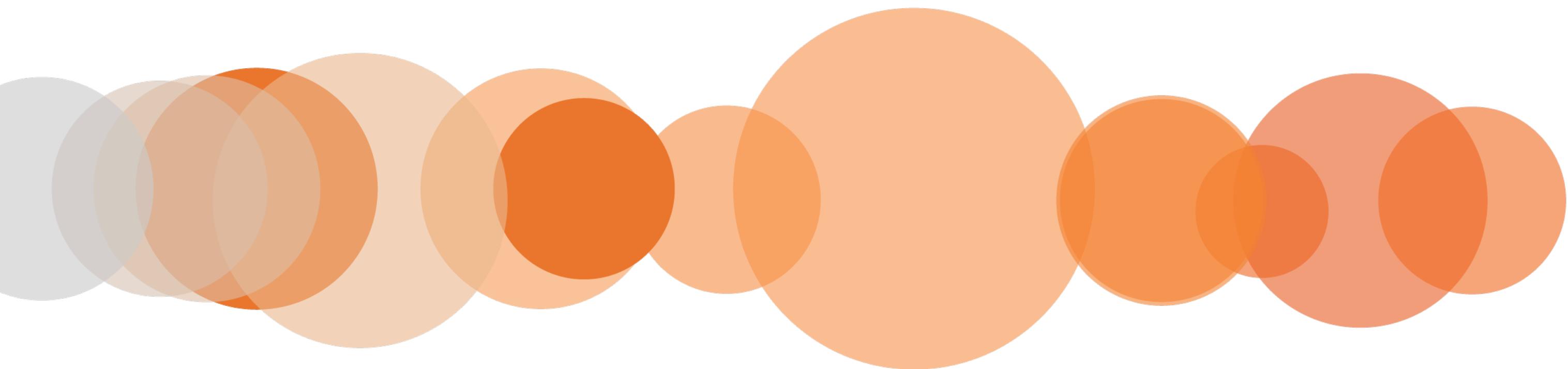
Connect to PDF



Connection to PDF

- Tableau can connect to tables in PDF as well and clean those up
- You can select which range of pages you want
- Each page is labeled as a table and a new data source is added
- Drag first page and use data interpreter
- Drag other pages to union

Connecting to Google Sheets

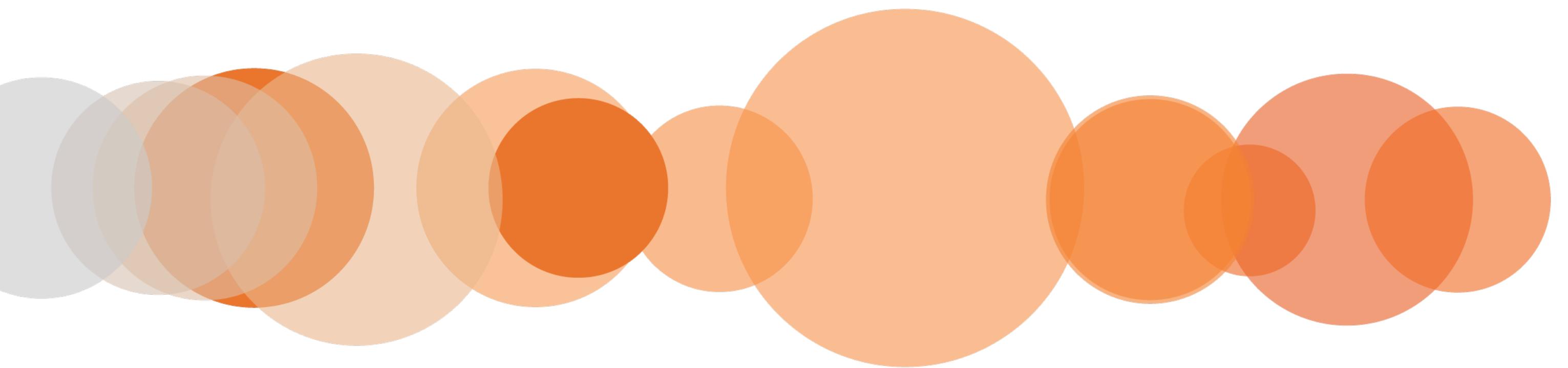


Connect to Google Sheets

- Upload your data into your google drive
- In the connect to data options, choose google sheets
- Give Tableau access to your google drive using Google credentials
- Pick the file
- You are ready

PDFs can be Messy

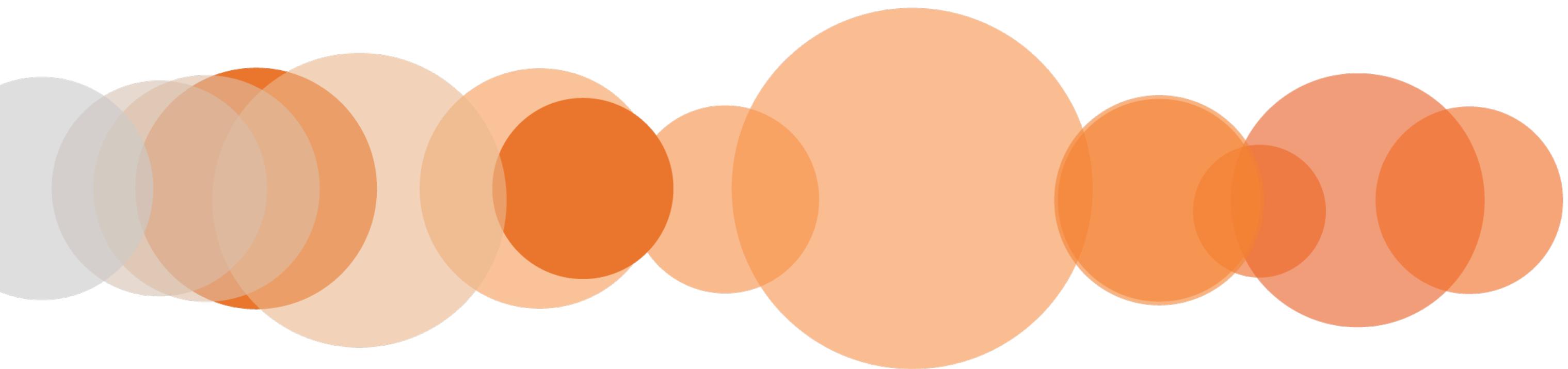
- Choose page 14 from the water resources sheet
- Tableau found 3 possible options to interpret the sheet
 - Table 2 and 3 is the bottom and top of the table
- Let's union table 3 and 2
- Turn on interpreter
- Let's cure the mismatch between inflows and F1 by merging them
- Lots of Nulls
 - Single row divided to 2
 - Header was read as a row
- Also bunch of extra totals exist
- It is filter time (hydroelectricity appears twice with data and nulls so leave for now)
- Rename headers and pivot then remove the extra nulls
- Use Alias for better water sources renaming



Odata Connector

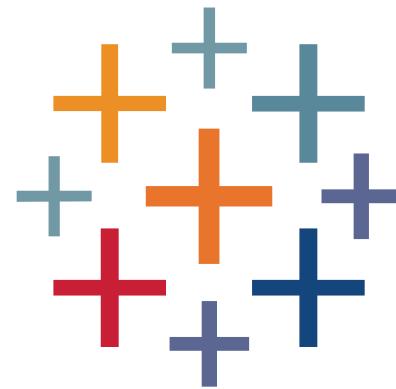
- OData is intended to be used to access information from a variety of sources including, but not limited to, relational databases, file systems, content management systems, and traditional websites.
- To start an OData connection, first we click on OData from the Tableau
- From the OData connection window, enter the URL to where the data is located.
- <https://services.odata.org/V3/Northwind/Northwind.svc/Employees>
- If no authentication information is necessary, click on "None".
- We can click "Okay", and the connection will pull in the available data for you to analyze.

Connecting to Web Data Connector



Web Data Connector

- It can connect to any data accessible over http
- It can even access json and xml data
- To build your own data connector you need some programming



+ a b l e a u®