



Data Visualization

Dr. Ken Kwong-Kay Wong

Boston, MA
Mar 1, 2023







Data Visualization

Dr. Ken Kwong-Kay Wong

Boston, MA
Mar 1, 2023



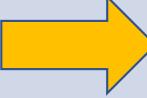
Class 3:

Data Preparation & Maps

May 1, 2023



Course Schedule (Day 3)

Class	Topics
Wed, Mar 1, 2023:	
3 	<p>Data Preparation & Maps</p> <ul style="list-style-type: none">1. Data Interpreter in Tableau Desktop2. Cleaning Messy Survey Data in Excel3. Exploring Tableau Prep Builder<ul style="list-style-type: none">a. Reshaping Data from Wide to Tallb. Joining Data4. Unions in Tableau Desktop5. Different kinds of Joins in Tableau Desktop6. Creating Maps<ul style="list-style-type: none">a. Symbol Mapb. Filled Mapc. Density Mapd. Map Layerse. Pie Chart Mapf. Dual Axis Mapg. Viz in Tooltips on Maph. Mapbox



Course Schedule (Day 4)

Class	Topics
Thurs, Mar 2, 2023:	
4	Different Types of Visual in Business Communications <ul style="list-style-type: none">1. Comparison<ul style="list-style-type: none">a. Column Chartb. Bar Chartc. Bullet Chartd. Lollipop Charte. Line Chartf. Step Chartg. Highlight Tableh. Heatmaps2. Composition<ul style="list-style-type: none">a. Pie Chartb. Donut Chartc. Treemapd. Packed Bubble Charte. Area Chartf. Cumulative Sum with Waterfall Chart3. Relationship<ul style="list-style-type: none">a. Scatterplot4. Distribution<ul style="list-style-type: none">a. Box-and-whisker Plotb. Likert Scale Chart



Course Schedule (Day 5)

Class	Topics
Fri, Mar 3, 2023:	
5	Ethical Considerations in Data Visualization <ol style="list-style-type: none">1. Ethics in Data Visualization<ol style="list-style-type: none">a. Unethical Data Visualizationb. The Misleading Data Dashboardc. 10 Data Visualization Mistakes to Avoidd. A Code of Ethics for Data Visualization Professionals2. Sharing and Publishing Data<ol style="list-style-type: none">a. E-mailing Tableau Workbook fileb. Publishing to Tableau Serverc. Publishing to Tableau Online<ul style="list-style-type: none">- Using the “Ask Data” Functiond. Publishing to Tableau Publice. Publishing to Web Sites and Blogs via HTML Embeddingf. Using Tableau Reader and Tableau Mobile App

Course Schedule (Day 6)

Class	Topics
Mon, Mar 6, 2023:	
6	Calculated Field, Parameters, and Quick Table <ul style="list-style-type: none">1. Create Calculated Field Using Text Operators:<ul style="list-style-type: none">a. SPLITb. LEFT and RIGHTc. LOWER and UPPERd. REPLACEe. DATEADDf. DATEDIFFg. DATEPARTh. DATEPARSEi. CASEj. IF-THEN-ELSE / Group Creationk. IIFl. IFNULL2. Create Parameters<ul style="list-style-type: none">a. What If Analysisb. Text Fields Search3. Quick Table<ul style="list-style-type: none">a. Running Totalb. Cohort Analysis4. The Analytics Pane<ul style="list-style-type: none">a. Constant, Average, and Reference Lineb. Trend Line5. Advanced Visualization Techniques<ul style="list-style-type: none">a. Timelinesb. Gantt Chartc. Bar-in-bar Chartd. Radar Charte. Interactive View: Top 5 Clients



Course Schedule (Day 7)

Class	Topics
Tues, Mar 7, 2023:	
7	Elevating Your Tableau Knowledge <ol style="list-style-type: none">1. What's next after QTM-6032?<ol style="list-style-type: none">a. The Tableau Communityb. The Tableau Conferencec. The Iron Viz competitiond. Blogs about DataVize. Tableau Zen Master2. Fun Stuff – Arts3. Team Presentation



1. Data Interpreter in Tableau Desktop



1. The ideal spreadsheet – required by Tableau

	A	B	C	D	E	F	G
1	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name
2	1	CA-2018-152156	2018-11-08	2018-11-11	Second Class	CG-12520	Claire Gute
3	2	CA-2018-152156	2018-11-08	2018-11-11	Second Class	CG-12520	Claire Gute
4	3	CA-2018-138688	2018-06-12	2018-06-16	Second Class	DV-13045	Darrin Van Huff
5	4	US-2017-108966	2017-10-11	2017-10-18	Standard Class	SO-20335	Sean O'Donnell
6	5	US-2017-108966	2017-10-11	2017-10-18	Standard Class	SO-20335	Sean O'Donnell
7	6	CA-2016-115812	2016-06-09	2016-06-14	Standard Class	BH-11710	Brosina Hoffman
8	7	CA-2016-115812	2016-06-09	2016-06-14	Standard Class	BH-11710	Brosina Hoffman
9	8	CA-2016-115812	2016-06-09	2016-06-14	Standard Class	BH-11710	Brosina Hoffman
10	9	CA-2016-115812	2016-06-09	2016-06-14	Standard Class	BH-11710	Brosina Hoffman
11	10	CA-2016-115812	2016-06-09	2016-06-14	Standard Class	BH-11710	Brosina Hoffman
12	11	CA-2016-115812	2016-06-09	2016-06-14	Standard Class	BH-11710	Brosina Hoffman
13	12	CA-2016-115812	2016-06-09	2016-06-14	Standard Class	BH-11710	Brosina Hoffman
14	13	CA-2019-114412	2019-04-15	2019-04-20	Standard Class	AA-10480	Andrew Allen
15	14	CA-2018-161389	2018-12-05	2018-12-10	Standard Class	IM-15070	Irene Maddox
16	15	US-2017-118983	2017-11-22	2017-11-26	Standard Class	HP-14815	Harold Pawlan
17	16	US-2017-118983	2017-11-22	2017-11-26	Standard Class	HP-14815	Harold Pawlan
18	17	CA-2016-105893	2016-11-11	2016-11-18	Standard Class	PK-19075	Pete Kriz
19	18	CA-2016-167164	2016-05-13	2016-05-15	Second Class	AG-10270	Alejandro Grove
20	19	CA-2016-143336	2016-08-27	2016-09-01	Second Class	ZD-21925	Zuschuss Donatelli



1. In reality, you've got this...in Excel

Open “Hult – Superstore RAW.xlsx” in Excel first.



	A	B	C	D	E	F
2		The Super Store Spreadsheet				
3		Prepared by John Smith (ext. 1234)				
5	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name
6	CA-2018-152156	2018-11-08	2018-11-11	Second Class	CG-12520	Claire Gute
7	CA-2018-152156	2018-11-08	2018-11-11	Second Class	CG-12520	Claire Gute
8	CA-2018-138688	2018-06-12	2018-06-16	Second Class	DV-13045	Darrin Van Huff
9	US-2017-108966	2017-10-11	2017-10-18	Standard Class	SO-20335	Sean O'Donnell
10	US-2017-108966	2017-10-11	2017-10-18	Standard Class	SO-20335	Sean O'Donnell
11	CA-2016-115812	2016-06-09	2016-06-14	Standard Class	BH-11710	Brosina Hoffman
12	CA-2016-115812	2016-06-09	2016-06-14	Standard Class	BH-11710	Brosina Hoffman
13	CA-2016-115812	2016-06-09	2016-06-14	Standard Class	BH-11710	Brosina Hoffman
14	CA-2016-115812	2016-06-09	2016-06-14	Standard Class	BH-11710	Brosina Hoffman
15	CA-2016-115812	2016-06-09	2016-06-14	Standard Class	BH-11710	Brosina Hoffman
16	CA-2016-115812	2016-06-09	2016-06-14	Standard Class	BH-11710	Brosina Hoffman
17	CA-2016-115812	2016-06-09	2016-06-14	Standard Class	BH-11710	Brosina Hoffman
18	CA-2019-114412	2019-04-15	2019-04-20	Standard Class	AA-10480	Andrew Allen
19	CA-2018-161389	2018-12-05	2018-12-10	Standard Class	IM-15070	Irene Maddox



1. It will look problematic in Tableau

Open **Hult – Superstore RAW.xlsx** in Tableau Desktop

Drag “**Orders**” to the Canvas

The screenshot shows the Tableau desktop interface with the following details:

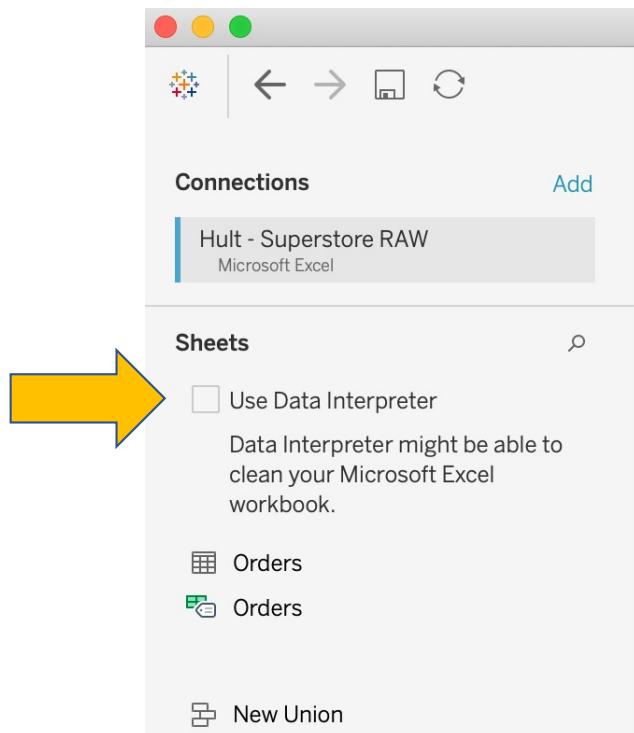
- Connections:** Hult - Superstore RAW (Microsoft Excel)
- Sheets:** Orders (highlighted by a yellow arrow)
- Data Preview:** Shows the first 1,000 rows of the 'Orders' sheet. The columns are labeled F1 through F10. The first few rows of data are as follows:

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
Abc	Orders	Orders	Orders	Orders	Orders	Orders	Orders	Orders	Orders	Orders
Orders	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
The Super Store Spre...										
Prepared by John Sm...										
Order ID	null	null		Ship Mode	Customer ID	Customer Name	Segment	Country/Region	City	State
CA-2018-152156	2018-11-08	2018-11-11	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentuc	
CA-2018-152156	2018-11-08	2018-11-11	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentuc	
CA-2018-138688	2018-06-12	2018-06-16	Second Class	DV-13045	Darrin Van Huff	Corporate	United States	Los Angeles	Califor	
US-2017-108966	2017-10-11	2017-10-18	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	Florida	

13



1. Quick fix: Data Interpreter in Tableau Desktop



- Tableau will **NOT** change your underlying data source
- Click “**Use Data Interpreter**”, wait
- Click “**Review the results**”, it opens up Excel
 - It takes 5-10 seconds
- Done!



1. Quick fix: Data Interpreter in Tableau Desktop

- It removes title
- It removes empty rows and columns
- It removes footnotes
- It removes descriptions



1. Quick fix: Data Interpreter in Tableau Desktop

The screenshot shows the Tableau desktop interface with the following details:

- Connections:** Hult - Superstore RAW (Microsoft Excel)
- Sheets:** Orders (selected), Cleaned with Data Interpreter (checkbox checked)
- Data Preview:** Shows columns: Order ID, Order Date, Ship Date, Ship Mode, Customer ID, Customer Name, Segment, and Country/Region.
- Bottom Panel:** Sort fields, Data source order dropdown, Show aliases, Show hidden fields, 1,000 rows button.

Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country/Region
CA-2018-152156	2018-11-08	2018-11-11	Second Class	CG-12520	Claire Gute	Consumer	United States
CA-2018-152156	2018-11-08	2018-11-11	Second Class	CG-12520	Claire Gute	Consumer	United States
CA-2018-138688	2018-06-12	2018-06-16	Second Class	DV-13045	Darrin Van Huff	Corporate	United States
US-2017-108966	2017-10-11	2017-10-18	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States
US-2017-108966	2017-10-11	2017-10-18	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States
CA-2016-115812	2016-06-09	2016-06-14	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States
CA-2016-115812	2016-06-09	2016-06-14	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States

16



1. Quick fix: Data Interpreter in Tableau Desktop

- But, you must double check to ensure Data Interpreter capture what you want to use (see the pink and green cells)
 - Click “Review the result” link, wait for 10 seconds, check the **last sheet/tab**
 - Look at the tab “**Orders_Orders**”
 - It also tells you that “**A0005 to T9999**” are used.

The screenshot shows the Tableau desktop interface. In the 'Connections' pane, there is one connection named 'Hult - Superstore RAW' from Microsoft Excel. In the 'Sheets' pane, there are two sheets: 'Orders' and 'Orders'. A yellow arrow points from the 'Cleaned with Data Interpreter' checkbox next to the 'Orders' sheet to the 'Review the results.' link below it. This link is described as '(To undo changes, clear the check box.)'. The main area displays a large portion of an Excel spreadsheet with columns labeled A through U. The first few rows show data such as Order ID, Order Date, Ship Date, Ship Mode, Customer Name, Segment, Country/Region, State, Postal Code, Region, Product ID, Category, Sub-Category, Product Name, Sales, Quantity, Discount, Profit, and a Header row. The data spans from row 1 to row 25.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1 Order ID	Order Date	Ship Date	Ship Mode	Customer Name	Customer Segment	Country/Region	State	Postal Code	Region	Product ID	Category	Sub-Catagory	Product Name	Sales	Quantity	Discount	Profit	Header		
2 CA-2018-152156	2018-11-08	2018-11-11	Second Cls	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky	42420	South	FUR-B0-10	Furniture	Bookcases	Bush Some	261.96	2	0	41,9136	Data
3 CA-2018-152156	2018-11-08	2018-11-11	Second Cls	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky	42420	South	FUR-CH-10	Furniture	Chairs	Hon Deluxe	731.94	3	0	219,582	Data
4 CA-2018-138688	2018-06-12	2018-06-16	Second Cls	DV-13045	Darrin Van	Corporate	United States	Los Angeles	California	90036	West	OFF-LA-10C	Office Supp	Labels	Self-Adhesive	14.62	2	0	6,8714	Data
5 US-2017-108966	2017-10-11	2017-10-18	Standard C	SO-20335	Sean O'Don	Consumer	United States	Fort Lauderdale	Florida	33311	South	FUR-TA-10C	Furniture	Tables	Bretford CF	957.5775	5	0.45	-383.031	Data
6 US-2017-108966	2017-10-11	2017-10-18	Standard C	SO-20335	Sean O'Don	Consumer	United States	Fort Lauderdale	Florida	33311	South	OFF-ST-100	Office Supp	Storage	Eldon Fold	22.368	2	0.2	2,5164	Data
7 CA-2016-115812	2016-06-09	2016-06-14	Standard C	BH-11710	Brosina Ho	Consumer	United States	Los Angeles	California	90032	West	FUR-FU-101	Furniture	Furnishing	Eldon Express	48.86	7	0	14,1694	Data
8 CA-2016-115812	2016-06-09	2016-06-14	Standard C	BH-11710	Brosina Ho	Consumer	United States	Los Angeles	California	90032	West	OFF-AR-10C	Office Supp	Art	Newell 322	7.28	4	0	1,9656	Data
9 CA-2016-115812	2016-06-09	2016-06-14	Standard C	BH-11710	Brosina Ho	Consumer	United States	Los Angeles	California	90032	West	TEC-PH-10X	Technology	Phones	Mitel 5320	907.152	6	0.2	90,7152	Data
10 CA-2016-115812	2016-06-09	2016-06-14	Standard C	BH-11710	Brosina Ho	Consumer	United States	Los Angeles	California	90032	West	OFF-BI-100	Office Supp	Binders	DXL Angle-i	18.504	3	0.2	5,7825	Data
11 CA-2016-115812	2016-06-09	2016-06-14	Standard C	BH-11710	Brosina Ho	Consumer	United States	Los Angeles	California	90032	West	OFF-AP-10C	Office Supp	Appliances	Belkin FSG	114.9	5	0	34,47	Data
12 CA-2016-115812	2016-06-09	2016-06-14	Standard C	BH-11710	Brosina Ho	Consumer	United States	Los Angeles	California	90032	West	FUR-TA-10C	Furniture	Tables	Chromcraft	1706.184	9	0.2	85,3092	Data
13 CA-2016-115812	2016-06-09	2016-06-14	Standard C	BH-11710	Brosina Ho	Consumer	United States	Los Angeles	California	90032	West	TEC-PH-10X	Technology	Phones	Konfrel 2SC	911.424	4	0.2	68,3568	Data
14 CA-2019-114412	2019-04-15	2019-04-20	Standard C	AA-10480	Andrew Alli	Consumer	United States	Concord	North Carolina	28027	South	OFF-PA-10C	Office Supp	Paper	Xerox 1967	15.552	3	0.2	5,4432	Data
15 CA-2018-161389	2018-12-05	2018-12-10	Standard C	IM-15070	Irene Madd	Consumer	United States	Seattle	Washington	98103	West	OFF-BI-100	Office Supp	Binders	Fellowes PE	407.976	3	0.2	132,5922	Data
16 US-2017-118983	2017-11-22	2017-11-26	Standard C	HP-14815	Harold Paw	Home Office	United States	Fort Worth	Texas	76109	Central	OFF-AP-10C	Office Supp	Appliances	Holmes Rep	68.81	5	0.8	-123,858	Data
17 US-2017-118983	2017-11-22	2017-11-26	Standard C	HP-14815	Harold Paw	Home Office	United States	Fort Worth	Texas	76106	Central	OFF-BI-100	Office Supp	Binders	Stor-Duri	2,544	3	0.8	-3,816	Data
18 CA-2016-105893	2016-11-11	2016-11-18	Standard C	PK-19075	Pete Kriz	Consumer	United States	Madison	Wisconsin	53711	Central	OFF-ST-100	Office Supp	Storage	Stur-D-Stor	665.88	6	0	13,3176	Data
19 CA-2016-167164	2016-05-13	2016-05-15	Second Cls	AG-10270	Alejandro C	Consumer	United States	West Jordan	Utah	84084	West	OFF-ST-100	Office Supp	Storage	Fellowes SL	55.5	2	0	9.99	Data
20 CA-2016-143336	2016-08-27	2016-09-01	Second Cls	ZD-21925	Zuschuss D	Consumer	United States	San Francisco	California	94109	West	OFF-AR-10C	Office Supp	Art	Newell 341	8.56	2	0	2,4824	Data
21 CA-2016-143336	2016-08-27	2016-09-01	Second Cls	ZD-21925	Zuschuss D	Consumer	United States	San Francisco	California	94109	West	TEC-PH-10X	Technology	Phones	Cisco SPA5	213.48	3	0.2	16,0111	Data
22 CA-2016-143336	2016-08-27	2016-09-01	Second Cls	ZD-21925	Zuschuss D	Consumer	United States	San Francisco	California	94109	West	OFF-BI-100	Office Supp	Binders	Wilson Jon	22.72	4	0.2	7,384	Data
23 CA-2018-137330	2018-12-09	2018-12-13	Standard C	KB-16585	Ken Black	Corporate	United States	Fremont	Nebraska	68025	Central	OFF-AR-10C	Office Supp	Art	Newell 318	19.46	7	0	5,0596	Data
24 CA-2018-137330	2018-12-09	2018-12-13	Standard C	KB-16585	Ken Black	Corporate	United States	Fremont	Nebraska	68025	Central	OFF-AP-10C	Office Supp	Appliances	Acco Six-Ot	60.34	7	0	15,6884	Data
25 US-2019-156909	2019-07-16	2019-07-18	Second Cls	SF-20065	Sandra Flar	Consumer	United States	Philadelphia	Pennsylvania	19140	East	FUR-CH-10	Furniture	Chairs	Global Del	71.372	2	0.3	-1,0196	Data



1. Quick fix: Data Interpreter in Tableau Desktop

Other usages:

No file

- If there are 2 tables sticking in 1 sheet, it can separate them into different sheets

A	B	C	D	E	F	G	H	I	J	K	L	M		
3	Prepared by Dr. Ken Wong.													
4														
5	ID	1	CA-2018-112156	2	CA-2018-112156	3	CA-2018-118688	4	CA-2018-118966	5	CA-2016-118966	6	CA-2016-115812	7
6	Order Date	2018-11-08	2018-11-08	2018-06-12	2017-10-11	2017-10-11	2017-10-11	2017-10-18	2017-10-18	2016-06-09	2016-06-09	2016-06-09	2016-06-09	
7	Ship Date	2018-11-11	2018-11-11	2018-06-16	2017-10-18	2017-10-18	2017-10-18	2017-10-18	2017-10-18	2016-06-14	2016-06-14	2016-06-14	2016-06-14	
8	Customer ID	OS-12320	CG-12320	DU-13945	SO-20335	SO-20335	BIH-11710							
9	Customer Name	Glare Cade	Glare Cade	Darin Van Hufft	Sean O'Donnell	Sean O'Donnell	Brosnia Hoffman							
10	Country	United States	United States	United States	United States	United States	United States	United States	United States	United States	United States	United States	United States	
11	City	Henderson	Henderson	Los Angeles	Fort Lauderdale	Fort Lauderdale	Los Angeles							
12	Postal Code	84240	84240	N/A	33111	N/A	90032	90032	90032	90032	90032	90032	90032	
13	Region	South	South	West	South	South	West							
14	Product ID	FUR-BD-1000175-FUR-HC-00045	FUR-LA-1000240	FUR-TA-10000577	FUR-SI-10000760	FUR-FI-10001487	FUR-AH-10002833	FUR-AM-10002275	FUR-BI-10000390	FUR-PM-10002275	FUR-BI-10000390	FUR-PM-10002275	FUR-BI-10000390	
15														
16	ID	11	12	13	14	15	16	17	18	19	20	21	22	
17	Order ID	CA-2018-112156	CA-2018-112156	CA-2018-118688	CA-2018-118966	CA-2016-118966	CA-2016-118966	CA-2016-115812	CA-2016-115812	CA-2016-115812	CA-2016-115812	CA-2016-115812	CA-2016-115812	
18	Order Date	2018-11-08	2018-11-08	2018-06-12	2017-10-11	2017-10-11	2017-10-11	2017-10-18	2017-10-18	2016-06-09	2016-06-09	2016-06-09	2016-06-09	
19	Ship Date	2018-11-11	2018-11-11	2018-06-16	2017-10-18	2017-10-18	2017-10-18	2017-10-18	2017-10-18	2016-06-14	2016-06-14	2016-06-14	2016-06-14	
20	Customer ID	OS-12320	CG-12320	DU-13945	SO-20335	SO-20335	BIH-11710							
21	Customer Name	Glare Cade	Glare Cade	Darin Van Hufft	Sean O'Donnell	Sean O'Donnell	Brosnia Hoffman							
22	Country	United States	United States	United States	United States	United States	United States	United States	United States	United States	United States	United States	United States	
23	City	Henderson	Henderson	Los Angeles	Fort Lauderdale	Fort Lauderdale	Los Angeles							
24	Postal Code	84240	84240	12480	33111	33111	23405	90032	90032	90032	90032	90032	90032	
25	Region	South	South	West	South	South	West							
26	Product ID	FUR-BD-1000175-FUR-HC-00045	FUR-LA-1000240	FUR-TA-10000577	FUR-SI-10000760	FUR-FI-10001487	FUR-AH-10002833	FUR-AM-10002275	FUR-BI-10000390	FUR-PM-10002275	FUR-BI-10000390	FUR-PM-10002275	FUR-BI-10000390	

Sheet 1

Sheet 2

Sheet 1 Sheet 2



2. Cleaning Messy Survey Data in Excel



2. Cleaning Messy Survey Data in Excel

Common mistake: Cross-tabulation reports with **wide tables**

- Don't try to connect to a fully formatted Excel report that already shows data aggregation (E.g., **with averages** across several data points).
- Instead, try to work with unaggregated raw data if possible
 - **One data point per row**, with each column begin a unique field in the data



2. Cleaning Messy Survey Data in Excel

Temperature Measurement				
	Month	Morning (6:00)	Noon (12:00)	Evening (18:00)
Seattle	April	5	17	11
	May	9	20	15
	June	12	25	18
	Average	8.7	20.7	14.7
New York	April	4	12	9
	May	8	18	13
	June	11	21	16
	Average	7.7	17	12.7
	Average across all	8.2	18.9	13.7

Task: convert this table into a format that Tableau can read.

21



2. Cleaning Messy Survey Data in Excel

Task

Hult – Temperature_measurement.xlsx

- Can you build a table like this one?
- $3 \times 3 = 9$ rows for Seattle; No need to include “Average”
- Compare your answer at

Hult – Temperature_measurement_w_answer.xlsx

Temperature Measurement				
	Month	Morning (6:00)	Noon (12:00)	Evening (18:00)
Seattle	April	5	17	11
	May	9	20	15
	June	12	25	18
	Average	8.7	20.7	14.7
	Average across all	8.2	18.9	13.7



	A	B	C	D
1	Location	Month	Time of Day	Temperature
2	Seattle	April	6:00	5
3	Seattle	April	12:00	17
4	Seattle	April	18:00	11
5	Seattle	May	6:00	9
6	Seattle	May	12:00	20
7	Seattle	May	18:00	15
8	Seattle	June	6:00	12
9	Seattle	June	12:00	25
10	Seattle	June	18:00	18
11	New York	April	6:00	4
12	New York	April	12:00	12
13	New York	April	18:00	9
14	New York	May	6:00	8
15	New York	May	12:00	18
16	New York	May	18:00	13
17	New York	June	6:00	11
18	New York	June	12:00	21
19	New York	June	18:00	16



2. Discussion (10 minutes)

1. Have you ever received problematic data (from survey/database)?
2. What's the problem? (e.g., missing data)
3. Did you fix it? If Yes, how?





2. Presentria.com

Language
English ▾

 Presentria

Session Number (Required)

Student ID (Optional)

Student Name (Optional)

Remember my student ID and student name

JOIN

Problem with data



Data Preparation - Survey

- Look for...
 - blank data
 - illogical (outlier) data
 - E.g., I work 23 hours a day
 - inconsistent data
 - E.g., I have night classes 5 days a week, and my shift at work is from 6pm till midnight.
 - illegal data
 - E.g., [male = 0, female = 1], but this person has a 5



Data Preparation - Survey

- Handle omission
- If >25% of questions are not answered in that questionnaire, throw it away for sure....it's useless.
- If 25% or less questions are not answered, you have 3 options:
 1. You can also **throw it away**, if the sample size is big enough as you can rely on other questionnaires
 2. Keep the questionnaire but substitute the missing data with **sample average**, if sample size is small
 3. If you have time and resources, try **to deduce the missing answer** from other parts of the questionnaire/record.



2. Cleaning Messy Survey Data in Excel

Let's talk about **Questionnaire / Survey Data**.

Open "**Hult - Reshaped Data.xlsx**" in your **Excel**

There are usually four different elements that need to fit together:

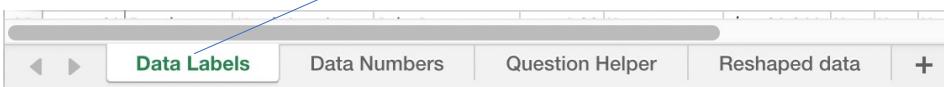
- a. **Demographic** data (e.g., age of respondents, gender, etc.)
- b. Survey responses in **text** format
- c. Survey responses in **numeric** format
- d. **Meta data** that describes the survey data.

With these, we can create a "Reshaped data" sheet



2a. Demographic data

I have 846 rows



	A	B	C	D	E
1	RespID	Gender	Location	Generation	Weight
2	2	Male	South America	Generation X	1
3	4	Female	South America	Baby Boomers	1.44
4	5	Female	South America	Generation X	1
5	6	Male	Antarctica	Baby Boomers	1.44
6	9	Female	Europe	Baby Boomers	1.32
7	12	Female	Europe	Baby Boomers	1.56
8	15	Male	North America	Baby Boomers	1.56
9	16	Male	Antarctica	Baby Boomers	1.44
10	17	Female	Europe	Baby Boomers	1.32
11	18	Male	North America	Traditionalists	0.595
12	22	Male	South America	Generation X	1.32
13	25	Female	South America	Generation X	1.32
14	26	Female	South America	Millenials	0.765
15	27	Male	Europe	Baby Boomers	1.56
16	29	Male	Europe	Generation X	1
17	30	Male	Europe	Baby Boomers	1.32
18	31	Male	Europe	Millenials	0.68
19	33	Male	North America	Generation X	1
20	34	Male	North America	Generation X	1.32
21	36	Female	North America	Millenials	1
22	37	Female	North America	Millenials	0.765
23	38	Female	South America	Baby Boomers	1.32
24	40	Female	North America	Baby Boomers	1.44
25	42	Female	Europe	Baby Boomers	1.44
26	43	Female	Europe	Baby Boomers	1.44
27	45	Male	North America	Baby Boomers	1.44



2b. Survey responses in text format

There are different tabs (sheets), look at the first one

But....we don't like text for answers, we need numbers!

	A	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
1	Respid	Q0	Q100	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q134a	Q134b	Q134c	Q134d	Q134e	Q134f	Q134g	Q134h	Q134i
2	2	No	\$ 98,038	No	No	Yes	No	Yes	No	No	Yes	Small degree	Small degree	Not at all	Small degree	Moderate degree	Small degree	High degree	Small degree	Very high degree	
3	4	No	\$ 138,936	Yes	Yes	Yes	No	No	No	Yes	No	Very high degree	Very high degree	Not at all	Very high degree	Very high degree	Very high degree	Moderate degree	Very high degree		
4	5	Yes	\$ 84,471	No	Yes	Yes	Yes	No	No	No	No	Very high degree	High degree	Moderate degree	Very high degree	High degree	High degree	Very high degree	High degree	Very high degree	
5	6	Don't know	\$ 138,534	No	Yes	Yes	No	No	No	No	No	Very high degree	High degree	Small degree	Moderate degree	High degree	Moderate degree	High degree	High degree	High degree	
6	9	Yes	\$ 68,944	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Very high degree	Very high degree	Moderate degree	Very high degree						
7	12	No	\$ 100,663	No	No	Yes	Yes	No	No	No	No	Moderate degree	Moderate degree	Moderate degree	High degree	High degree	High degree	High degree	High degree	High degree	
8	15		\$ 122,481																		
9	16	Yes	\$ 106,036	Yes	Yes	No	No	No	Yes	No	No	Moderate degree	Very high degree	High degree	Very high degree	High degree	High degree	High degree	High degree	Very high degree	
10	17	Don't know	\$ 81,681	Yes	Yes	Yes	No	No	No	No	Yes	High degree	High degree	High degree	High degree	High degree	High degree	High degree	High degree	Very high degree	
11	18	No	\$ 104,200	No	Yes	No	No	Yes	No	No	Yes	Moderate degree	High degree	High degree	Very high degree	Moderate degree	High degree	High degree	High degree	High degree	
12	22	No	\$ 172,723	No	No	Yes	Yes	No	No	No	No	High degree	High degree	Small degree	High degree	High degree	High degree	High degree	Small degree	High degree	
13	25	Yes	\$ 153,410	Yes	No	Small degree	High degree	Small degree	High degree	Moderate degree	Moderate degree	Very high degree	Moderate degree	Very high degree							
14	26	Yes	\$ 93,194	No	Yes	No	No	No	Yes	No	No	High degree	High degree	Moderate degree	High degree	High degree	High degree	Moderate degree	Moderate degree	High degree	
15	27	Yes	\$ 101,662	Yes	Yes	Yes	No	No	No	No	Yes	Very high degree	High degree	Small degree	Moderate degree	High degree	High degree	High degree	Very high degree	Very high degree	
16	29		\$ 114,216									Very high degree	Very high degree	Small degree	Very high degree	High degree	Moderate degree	Very high degree	Moderate degree	Very high degree	
17	30	No	\$ 97,354	No	No	No	No	Yes	No	No	No	Moderate degree	High degree	High degree	High degree	Moderate degree	Moderate degree	High degree	High degree	High degree	
18	31	Yes	\$ 120,061	No	No	No	No	No	Yes	Yes	No	Moderate degree	High degree	Moderate degree	Moderate degree	Moderate degree	Moderate degree	High degree	Moderate degree	High degree	
19	33		\$ 134,308									High degree	High degree	Not at all	Small degree	High degree	High degree	Moderate degree	High degree	Small degree	
20	34		\$ 146,227																		
21	36	Yes	\$ 110,462	No	Yes	Yes	Yes	No	No	No	No	Yes	Moderate degree	Moderate degree	Not at all	Moderate degree					
22	37		\$ 104,602																		
23	38	Don't know	\$ 119,871	Yes	Yes	Yes	No	No	No	No	No	Yes	High degree	High degree	Small degree	Moderate degree	High degree	Moderate degree	Very high degree	Moderate degree	High degree

Data Labels Data Numbers Question Helper Reshaped data +





2c. Survey responses in numeric format

Look at the 2nd tab (sheet)

So, you need to convert your text answers to numbers like this:

	A	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
1	ResID	Q0	Q100	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q134a	Q134b	Q134c	Q134d	Q134e	Q134f	Q134g	Q134h	Q134i
2	2	0	98,037.68	0	0	1	0	1	0	0	0	1	2	2	1	2	3	2	4	2	5
3	4	0	138,935.50	1	1	1	0	0	0	1	0	1	5	5	1	5	5	5	5	3	5
4	5	1	84,471.00	0	1	1	1	0	0	0	0	1	5	4	3	5	4	4	5	4	5
5	6	2	138,533.55	0	1	1	0	0	0	0	0	1	5	4	2	3	4	3	4	4	4
6	9	1	68,943.60	1	1	1	1	0	1	1	1	0	5	5	3	5	5	5	5	5	5
7	12	0	100,663.20	0	0	1	1	0	0	0	0	1	3	3	3	4	4	4	4	4	4
8	15		122,480.76																		
9	16	1	106,035.60	1	1	0	0	0	1	0	0	1	3	5	4	5	4	4	4	4	5
10	17	2	81,681.30	1	1	1	0	0	0	0	0	1	4	4	4	4	4	4	4	4	5
11	18	0	104,199.70	0	1	0	0	0	1	0	0	1	3	4	4	5	3	4	4	4	4
12	22	0	172,723.10	0	0	1	1	0	0	0	0	0	4	4	2	4	4	4	4	4	2
13	25	1	153,410.40	1	1	1	1	1	1	1	0	1	2	4	2	4	3	3	5	3	5
14	26	1	93,194.00	0	1	0	0	0	1	0	0	1	4	4	3	4	4	4	4	3	4
15	27	1	101,661.78	1	1	1	0	0	0	0	1	0	5	4	2	3	4	4	4	5	5
16	29		114,215.85										5	5	2	5	4	3	5	3	5
17	30	0	97,353.72	0	0	0	0	1	0	0	0	0	3	4	4	4	3	3	4	4	4
18	31	1	120,061.37	0	0	0	0	0	1	1	0	1	3	4	3	3	3	3	4	3	4
19	33		134,308.16										4	4	1	2	4	4	3	4	2
20	34		146,227.20																		
21	36	1	110,462.00	0	1	1	1	0	0	0	0	1	3	3	1	3	3	3	3	3	3
22	37		104,602.30																		
23	38	2	119,871.40	1	1	1	0	0	0	0	0	1	4	4	2	3	4	3	5	3	4



2d. Meta data that describes the survey data

Look at the 3rd tab (sheet)

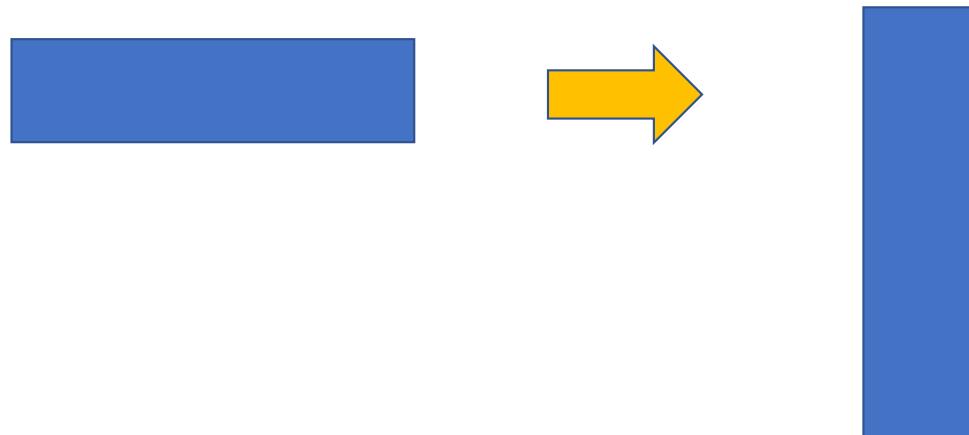
You need a page to show the questions and other details

	A	B	C	D
1	Question ID	Wording	Question Grouping	Question Type
2	Q0	Vote in the upcoming election?	Vote	Yes / No /Maybe
3	Q1	Pulse Rate	What do you measure	Check All
4	Q2	Metabolism	What do you measure	Check All
5	Q3	Blood Pressure	What do you measure	Check All
6	Q4	Temperature	What do you measure	Check All
7	Q5	Galvanic Skin Response	What do you measure	Check All
8	Q6	Breathing	What do you measure	Check All
9	Q7	Perspiration	What do you measure	Check All
10	Q8	Pupil Dilation	What do you measure	Check All
11	Q9	Adrenaline Production	What do you measure	Check All
12	Q134a	Good Job Skills	Indicate degree to which you agree	Likert
13	Q134b	Good Sense of Humor	Indicate degree to which you agree	Likert
14	Q134c	High Intelligence	Indicate degree to which you agree	Likert
15	Q134d	Can Play Jazz	Indicate degree to which you agree	Likert
16	Q134e	Likes the Beatles	Indicate degree to which you agree	Likert
17	Q134f	Good Ability to lift heavy objects	Indicate degree to which you agree	Likert
18	Q134g	Has grace under pressure	Indicate degree to which you agree	Likert
19	Q134h	Is Kind to animals	Indicate degree to which you agree	Likert
20	Q134i	Makes good coffee	Indicate degree to which you agree	Likert



2. Ultimately, we want to put these in the same sheet

- Now, look at the 4th tab called “Reshaped data”
- This is the ultimate kind of spreadsheet that we want to create
- In this example, the data is being “reshaped”, we have transformed it from **Wide** (i.e., many columns), to **Tall** (i.e., many rows)



32



2. Ultimately, we want to put these in the same sheet

i.e., transforming the data from “wide” to “tall”

I have 25004 rows now, not 846!

	Demographic data			Question ID	Numeric responses	Text responses	Meta data				
	A	B	C	D	E	F	G	H	I	J	K
1	RespID	Gender	Location	Generation	Weight	Question ID	Numeric value	Text value	Wording	Question Grouping	Question Type
2	2	Male	South America	Generation X	1	Q0	0	No	Vote in the upcoming election?	Vote	Yes / No / Maybe
3	2	Male	South America	Generation X	1	Q1	0	No	Pulse Rate	What do you measure	Check All
4	2	Male	South America	Generation X	1	Q100	98037.68	98037.68	Salary	Salary	Enter Value
5	2	Male	South America	Generation X	1	Q134a	2	Small degree	Good Job Skills	Indicate degree to which you agree	Likert
6	2	Male	South America	Generation X	1	Q134b	2	Small degree	Good Sense of Humor	Indicate degree to which you agree	Likert
7	2	Male	South America	Generation X	1	Q134c	1	Not at all	High Intelligence	Indicate degree to which you agree	Likert
8	2	Male	South America	Generation X	1	Q134d	2	Small degree	Can Play Jazz	Indicate degree to which you agree	Likert
9	2	Male	South America	Generation X	1	Q134e	3	Moderate degree	Likes the Beatles	Indicate degree to which you agree	Likert
10	2	Male	South America	Generation X	1	Q134f	2	Small degree	Good Ability to lift heavy objects	Indicate degree to which you agree	Likert
11	2	Male	South America	Generation X	1	Q134g	4	High degree	Has grace under pressure	Indicate degree to which you agree	Likert
12	2	Male	South America	Generation X	1	Q134h	2	Small degree	Is Kind to animals	Indicate degree to which you agree	Likert
13	2	Male	South America	Generation X	1	Q134i	5	Very high degree	Makes good coffee	Indicate degree to which you agree	Likert
14	2	Male	South America	Generation X	1	Q2	0	No	Metabolism	What do you measure	Check All
15	2	Male	South America	Generation X	1	Q28-IMP	5	Very Important	Price	Importance	Likert
16	2	Male	South America	Generation X	1	Q28-SAT	1	Not at all satisfied	Price	Satisfaction	Likert

Navigation controls: back, forward, Data Labels, Data Numbers, Question Helper, Reshaped data (highlighted), +, zoom controls, 100%.



2. Reshaping the Data from Wide to Tall

- The goal: to reduce the number of columns from 45 to just 11 (i.e. A – K), and increase the number of rows from 846 to 25,004.
- We do this because Tableau needs the data set in this format.
- But the multimillion-dollar question is....**HOW TO DO IT?**
- I don't have time to do it in Excel one-by-one, cell-by-cell!
- I also can't get my summer intern to do it, as he's busy with another project!

>> **Tableau Prep Builder** comes to rescue!

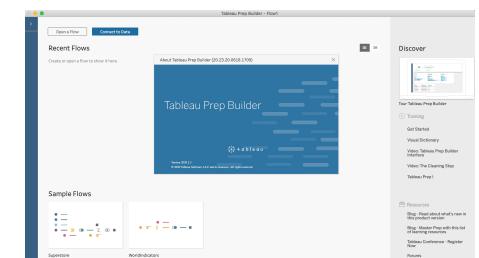
3. Exploring Tableau Prep Builder

a. Reshaping Data from Wide to Tall



3. Exploring Tableau Prep Builder

- Tableau Prep Builder (formerly known as **project maestro**)
- Free with your Tableau subscription; use same license key.
- Tableau Prep also allows you to save your work as a packaged workflow so that you can share the flow and the source data in a single file.
- Download: <https://www.tableau.com/products/prep>
- Tableau Prep projects are called **flows**



<https://www.datarevelations.com/resources/tableau-prep/>

36



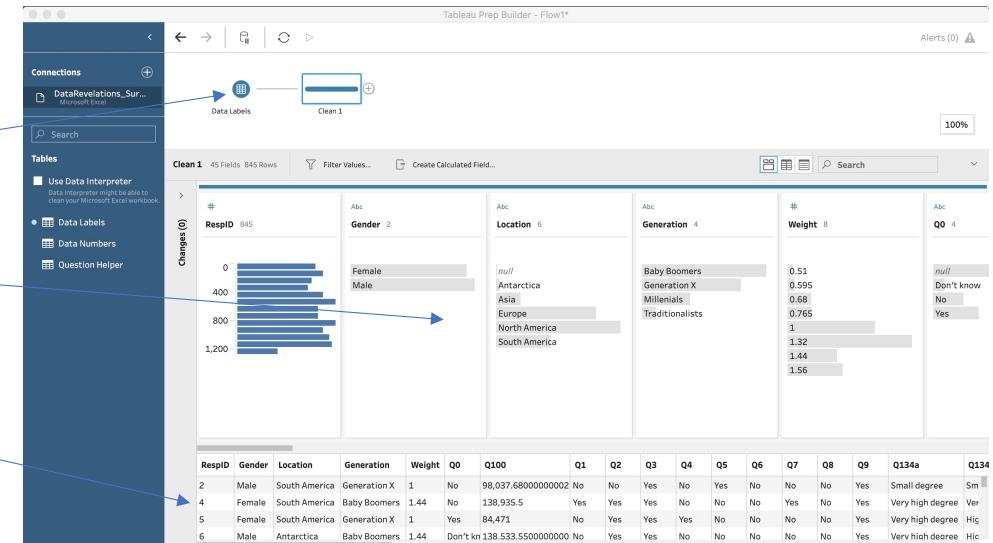
3. Exploring Tableau Prep Builder

- Open Tableau Prep Builder
- >
- Connection +
- To a file, Microsoft Excel
- **Hult - Reshaped Data.xlsx**
- Drag “**Data Labels**” to canvas.
- **Data Labels “+” Add “Clean Step”**

Add a step = let me take a look at the data

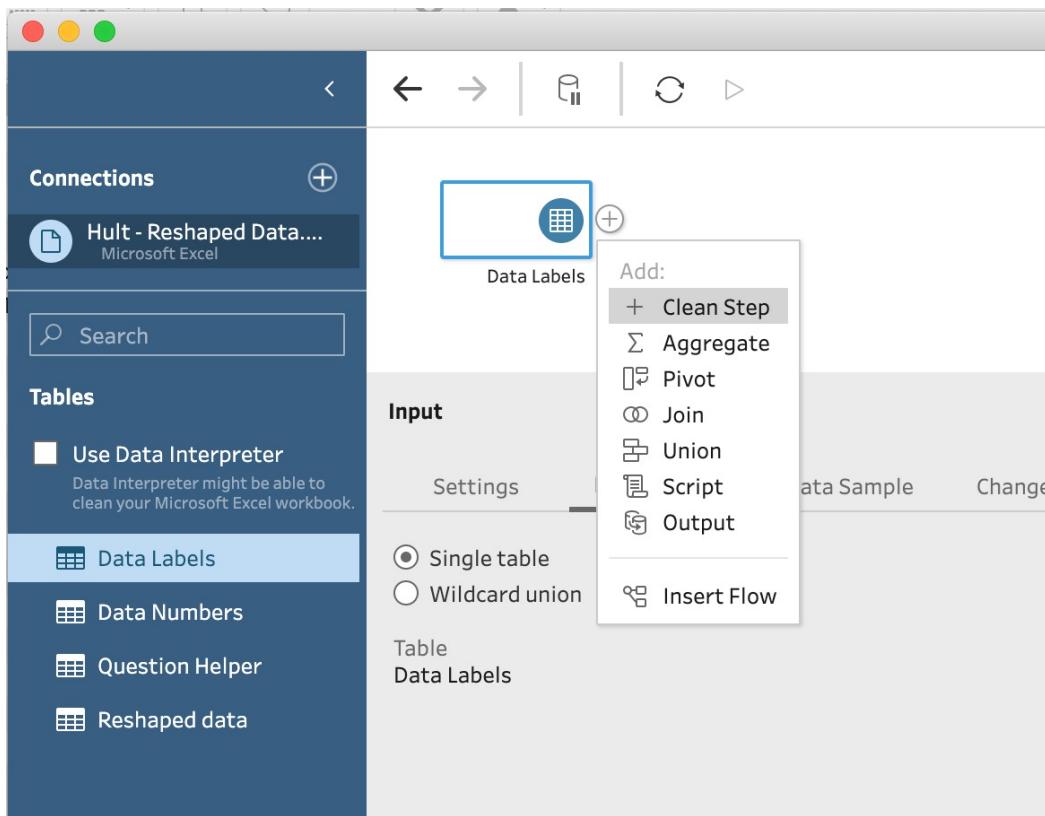
Response profile

Data grid





3. Exploring Tableau Prep Builder

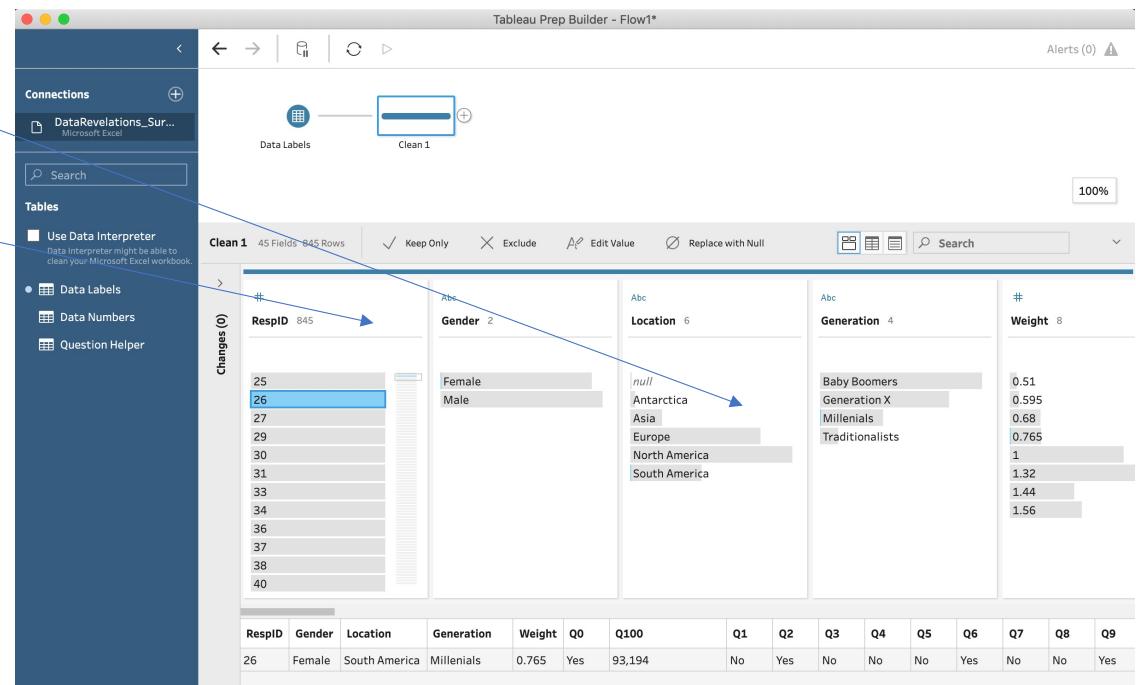


38



3. Exploring Tableau Prep Builder

- It automatically gives you a visual representation of the data!
- Look at the bars
- RespID
- Right side, ..., view state (**Detail**)



39



3. Exploring Tableau Prep Builder

- Let's reshape the data from wide to tall
- Clean 1, +, add **Pivot**.

The screenshot shows the Tableau Prep Builder interface. On the left, there's a sidebar with 'Connections' (Hult - Reshaped Data....), 'Tables' (Data Labels, Data Numbers, Question Helper, Reshaped data selected), and 'Changes (0)'. The main workspace shows a flow starting with 'Data Labels', followed by a 'Clean 1' step. A context menu is open over the 'Clean 1' step, with 'Pivot' highlighted. Other options in the menu include 'Clean Step', 'Aggregate', 'Join', 'Union', 'Script', 'Output', and 'Insert Flow'. Below the 'Clean 1' step, there are sections for '#', 'RespID 845', 'Gender', and 'Location'.

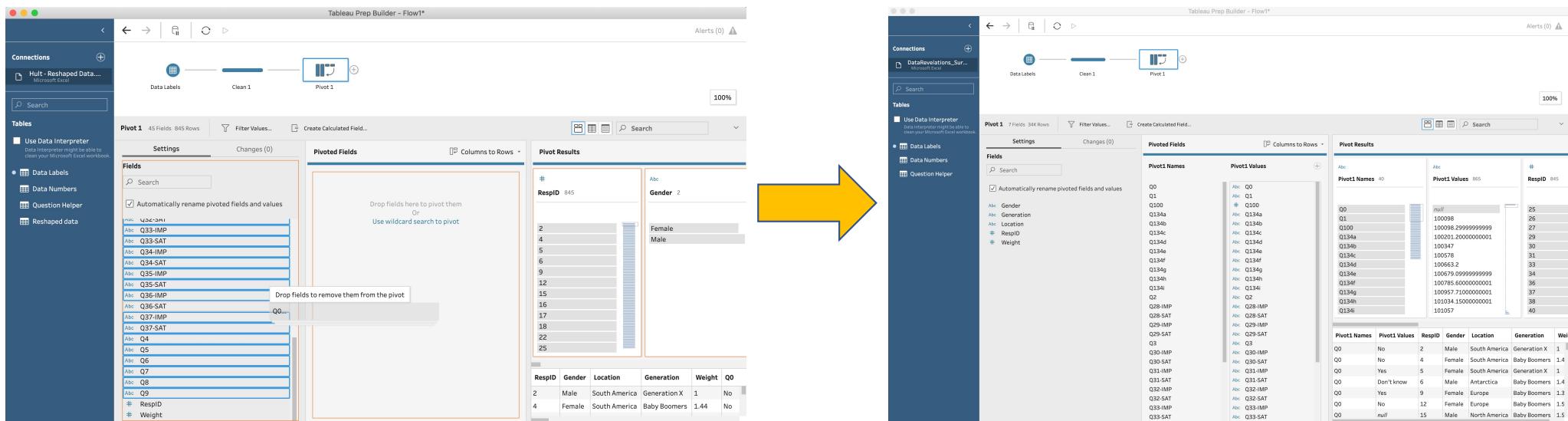
Recap: too wide!

	A	F	G	H	I	J	K	L	M	N	O	P
1	RespID	Q0	Q100	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9
2	2	No	\$ 98,038	No	No	Yes	No	Yes	No	No	No	Yes
3	4	No	\$ 138,936	Yes	Yes	Yes	No	No	No	Yes	No	Yes
4	5	Yes	\$ 84,471	No	Yes	Yes	Yes	No	No	No	No	Yes
5	6	Don't know	\$ 138,534	No	Yes	Yes	No	No	No	No	No	Yes
6	9	Yes	\$ 68,944	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No
7	12	No	\$ 100,663	No	No	Yes	Yes	No	No	No	No	Yes



3. Exploring Tableau Prep Builder

- Drag all fields except the demographics fields and RespID, into the ‘**Pivoted Fields**’ area.
 - That is, we need **Q0 to Q9**
 - [click Q0, shift, click Q9, drag]





3. Exploring Tableau Prep Builder

- Look at the 2nd table “**Pivot Results**”
- In “Pivot1 Name”, ..., **Rename field**
- Change “Pivot1 Name” to “Question ID”, the next one to “Text Responses”

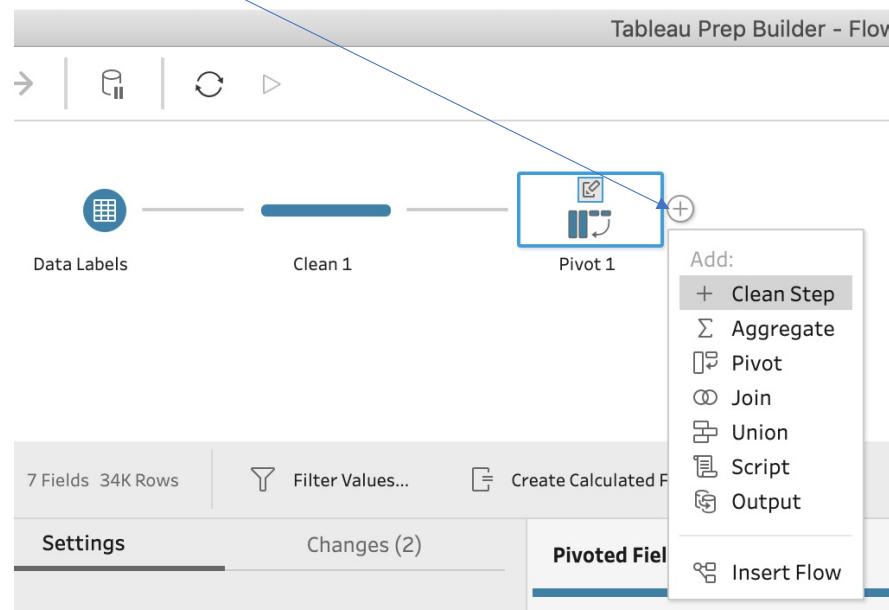
Remember, we're using info from the “Data Label” sheet, so it contains Text Responses

The screenshot shows the Tableau Prep Builder interface with the 'Pivot Results' step selected. On the left, the 'Pivoted Fields' pane lists various fields like Q0, Q1, Q100, etc. The main area shows a table with two columns: 'Pivot1 Names' and 'Pivot1 Values'. A context menu is open over the first column, with 'Rename Field' highlighted. The resulting table on the right has columns 'Question ID' and 'Text Responses'. The 'Question ID' column contains values like 40, 865, and 845, while the 'Text Responses' column contains text values such as 'null', '100098', and '100098.299999999999'. An arrow points from the 'Pivot1 Name' column to the 'Question ID' column, and another arrow points from the 'Pivot1 Values' column to the 'Text Responses' column.

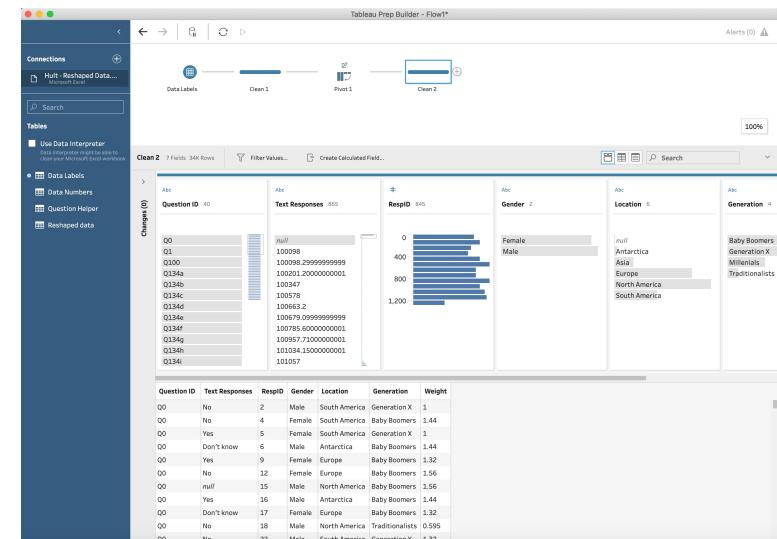


3. Exploring Tableau Prep Builder

- Pivot 1 +, add a **clean step**



Adding a clean step is like seeing the result





3. Exploring Tableau Prep Builder

- Next, we import and reshape the **numeric responses**
- Drag “**Data Numbers**” into the canvas on the upper right-hand corner.
- **Deselect the demographic fields** (Gender, Location, Generation, and Weight)

The screenshot shows the Tableau Prep Builder interface. In the top right, there's a flow pane with steps: Data Labels → Clean 1 → Pivot 1 → Clean 2. A blue box highlights the 'Data Numbers' step. Below the flow pane is the 'Input' tab, which includes tabs for Settings, Multiple Files (selected), Data Sample, and Changes (4). The 'Multiple Files' tab shows a list of selected fields: Data Numbers (Fields selected: 46 of 50). In the bottom half of the screenshot, there's a detailed view of the 'Data Numbers' editor. It has a header 'Data Numbers Fields selected: 46 of 50' and a 'Filter Values...' button. Below is a table titled 'Select the fields to include in your flow, apply a filter, or change data types. To see and clean your data, add a cleaning step in the flow pane.' The table columns are: Type, Field Name, Original Field Name, Changes, and Sample Values. The rows show the following data:

Type	Field Name	Original Field Name	Changes	Sample Values
#	ResID	ResID		2, 4, 5
Abc	Gender	Gender		Male, Female
Abc	Location	Location		South America
Abc	Generation	Generation		Generation X, Baby Boomers
#	Weight	Weight		1, 1.44
#	Q0	Q0		0, 1
#	Q100	Q100		98,037.6800000002, 138,935.5, 84,471
#	Q1	Q1		0, 1
#	Q2	Q2		n/a



3. Exploring Tableau Prep Builder

- In Canvas, Data Numbers +, add **Pivot**.
- Drag all the fields except the RespID (last one at the bottom) to the **Pivoted Fields**.

I.e., Drag Q0 to Q9

Pivot1 Names	Pivot1 Values	RespID	F46	F47	F48	F49	F50
Q0	0	2	null	null	null	1	
Q0	0	4	null	null	null	null	
Q0	1	5	null	null	null	null	
Q0	2	6	null	null	null	null	
Q0	1	9	null	null	null	null	
Q0	0	12	null	null	null	null	

45



3. Exploring Tableau Prep Builder

- In the 2nd “Pivot Results” table, rename “Pivot1 Names” to “Question ID”, and the next one to “Numeric Responses”

The screenshot shows the Tableau Prep Builder interface with a data flow. It starts with two tables: "Data Labels" and "Data Numbers". "Data Labels" flows through a "Clean 1" step, then "Pivot 1", and finally "Clean 2". "Data Numbers" flows through "Pivot 2". The "Pivot 1" step is highlighted with a blue border. The "Pivot Results" table is displayed on the right, showing two columns: "Pivot1 Names" and "Pivot1 Values". The "Pivot1 Names" column lists various question IDs (Q0, Q1, Q100, etc.) and other fields like F46, F47, F48, F49, F50, and RespID. The "Pivot1 Values" column contains numerical values corresponding to these names. Below the table is a bar chart titled "Question ID" with 40 categories, showing values ranging from 0 to over 100,000. To the right of the chart is another bar chart titled "Numeric Responses" with 852 categories, showing values ranging from 0 to 1,200. At the bottom, a preview of the data is shown in a grid format with columns for Question ID, Numeric Responses, RespID, and several F fields.

Question ID	Numeric Responses	RespID	F46	F47	F48	F49	F50
Q0	0	2	null	null	null	null	1
Q0	0	4	null	null	null	null	null
Q0	1	5	null	null	null	null	null
Q0	2	6	null	null	null	null	null
Q0	1	9	null	null	null	null	null
Q0	0	12	null	null	null	null	null
..

Remember, we're using info from the “Data Numbers” sheet, so it contains Numeric Responses

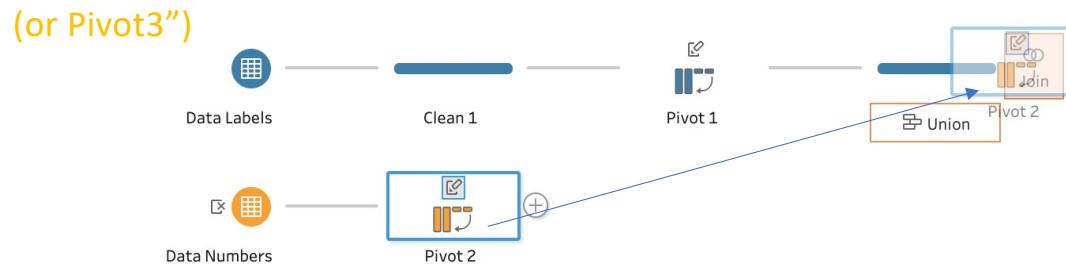
3. Exploring Tableau Prep Builder

- a. Reshaping Data from Wide to Tall
- b. Joining Data



3. Exploring Tableau Prep Builder

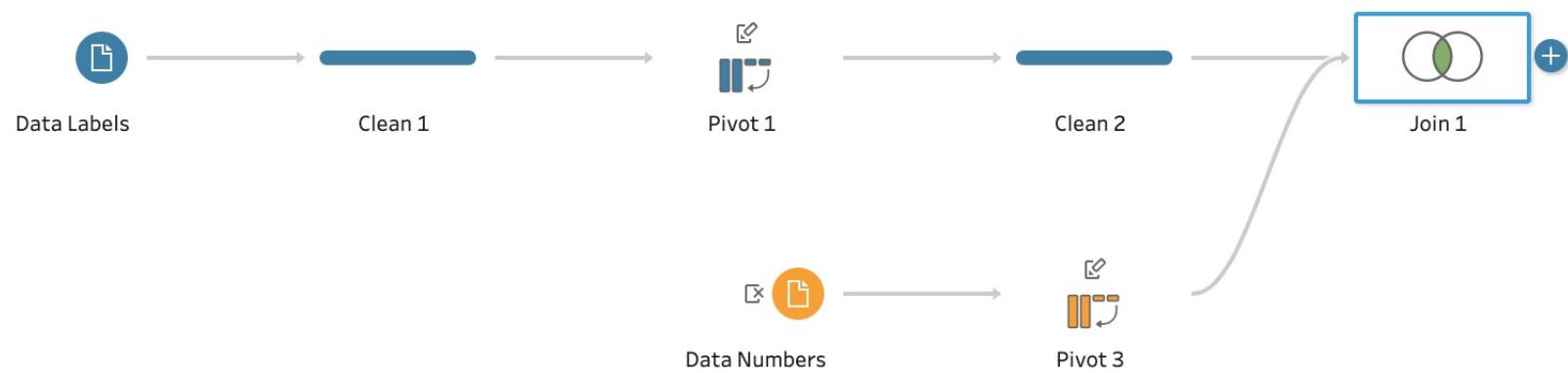
- Let's **join** the Text and Numeric Responses
- Now that we have both the text and numeric results reshaped from wide to tall (i.e., **pivoted**), we need to merge the data so that all the numeric responses line up with all the text responses.
 - That is, for every Resp ID we want to make sure the text and numeric responses for each Question ID line up properly.
- Drag “**Pivot2**” box to “**Clean2**”: Join (orange color); wait for 10 seconds



48



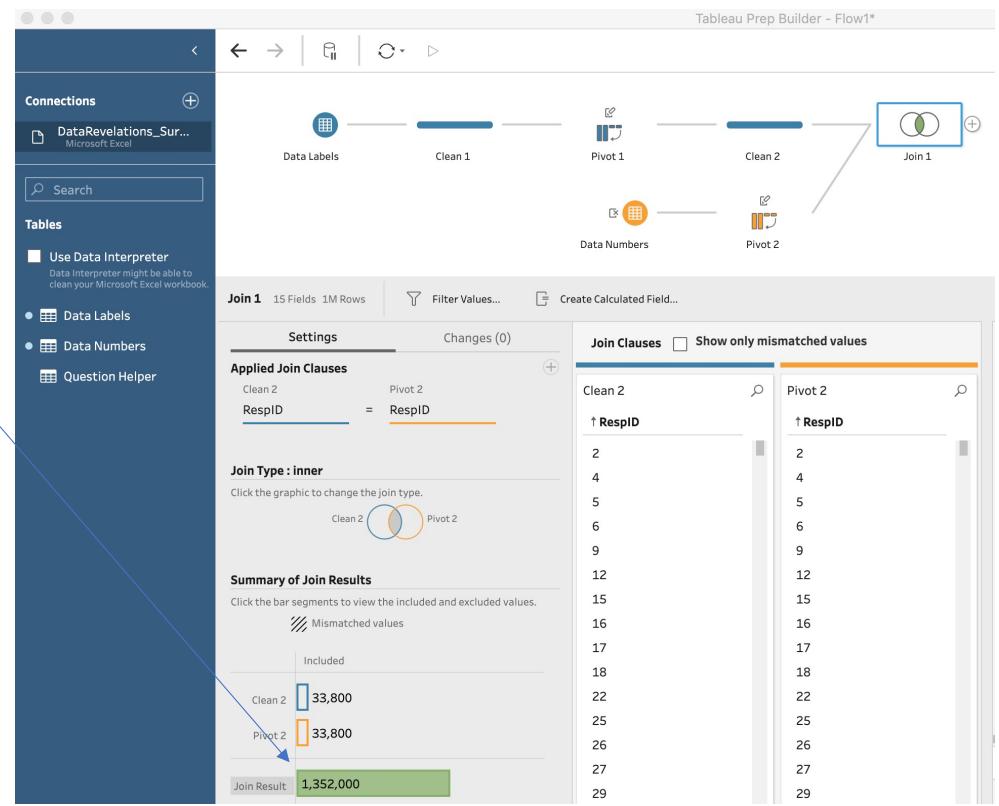
3. Exploring Tableau Prep Builder





3. Exploring Tableau Prep Builder

- *Tableau Prep will see that there's a match on RespID and create a join using just that field.*
 - *This will create over one million rows of data because it generates every possible combination of text and numeric result for each RespID.*

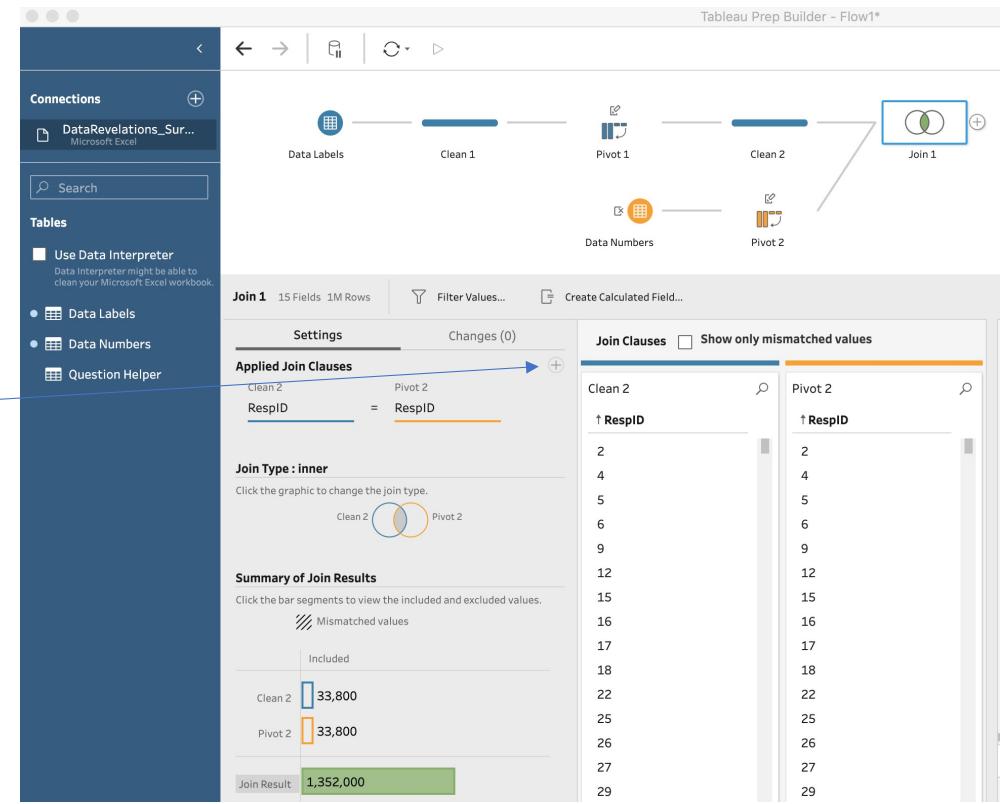


50



3. Exploring Tableau Prep Builder

- We need to indicate that we also want to join using the Question ID field.
 - If your screen is small, put your cursor in the grey area, then scroll up
 - Click the circular “+” icon next to “Applied Join Clauses” to specify the additional field.



51



3. Exploring Tableau Prep Builder

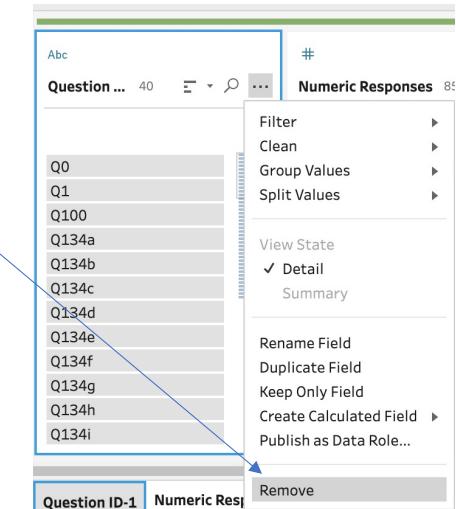
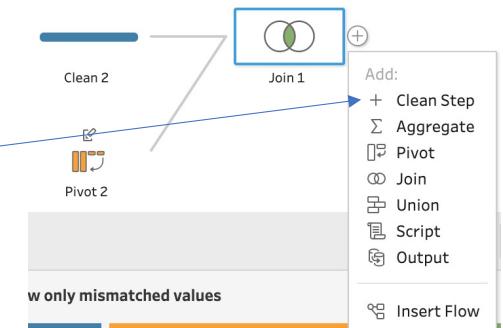
- Select “Question ID” under Clean2
- Select “Question ID” under Pivot2
- *This shows that in addition to RespID we also want to join using Question ID.*

In some rare situations, the joining will fail and you'll see some exclamation marks. In that case, press the ← button a few times and just re-do the steps.



3. Exploring Tableau Prep Builder

- To inspect our results, we **add a new “Clean Step”**.
- We see that everything lines up perfectly, but we also have **redundant** RespID and Question ID fields.
 - Question ID and **Question ID-1**
 - RespID and **RespID-1**
- Go to Question ID-1, ..., select “**Remove**”
- Scroll to the right, remove “**RespID-1**”



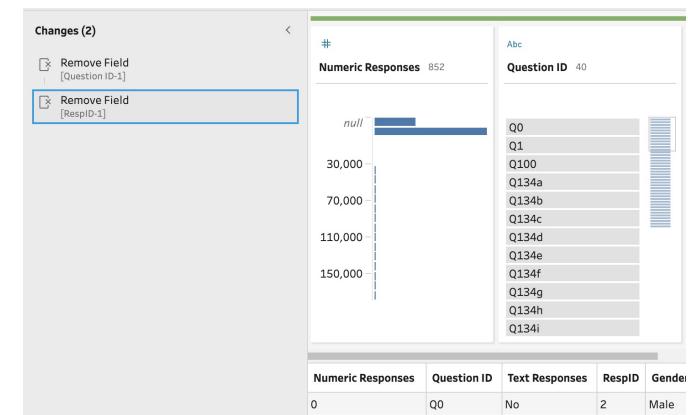
53



3. Exploring Tableau Prep Builder

Let's rename this step

- Go to “**Clean 3**” box, right click, rename it as “**Remove fields**”
- Click the “>” on the strip to view the change details



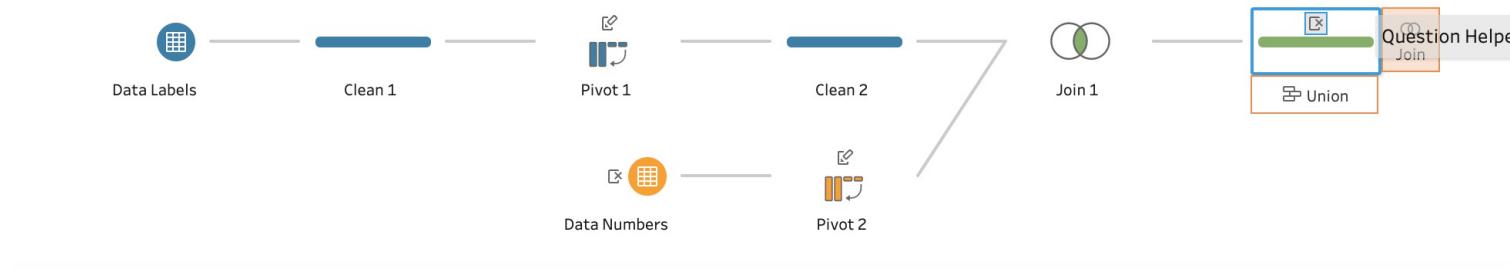
54



3. Exploring Tableau Prep Builder

After linking Data Labels and Data Numbers, let's add the 3rd sheet “Question Helper” Metadata

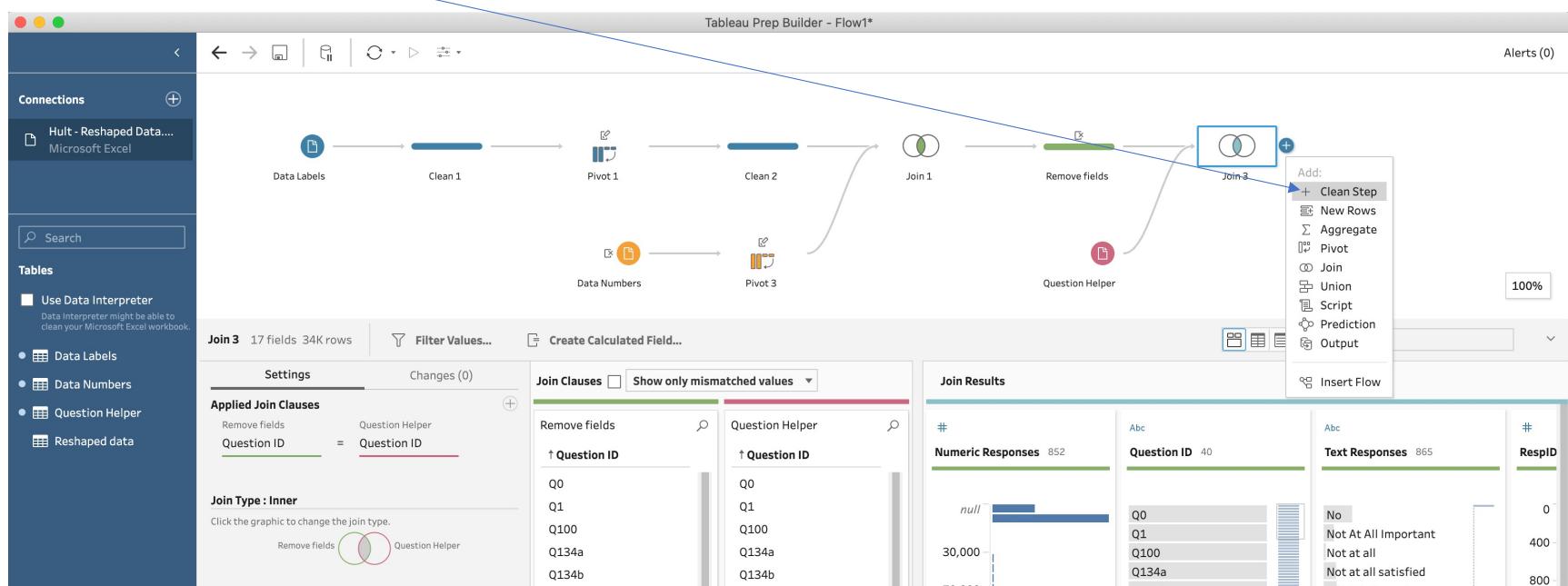
- Drag “**Question Helper**” from the left-hand side panel and put it on top of the “Join” box located on the right-hand side of “**Remove fields**”





3. Exploring Tableau Prep Builder

- Add “Clean Step” to refresh

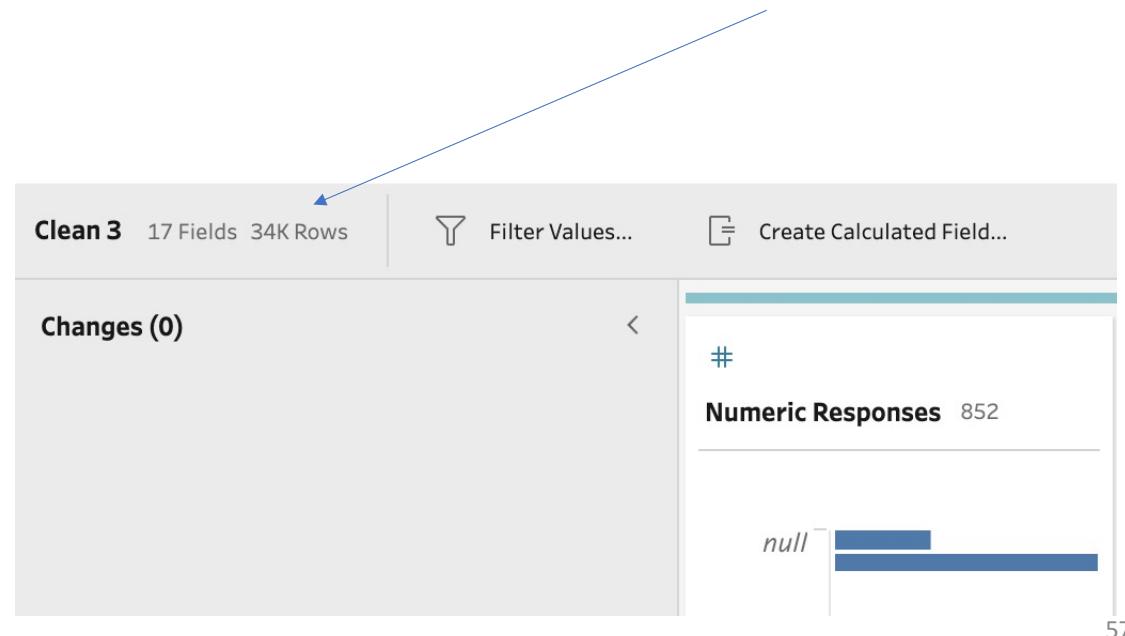


56



3. Exploring Tableau Prep Builder

- Note that at this point the data has morphed from containing 45 columns and 845 rows into 12 columns and over **34,000 rows**.

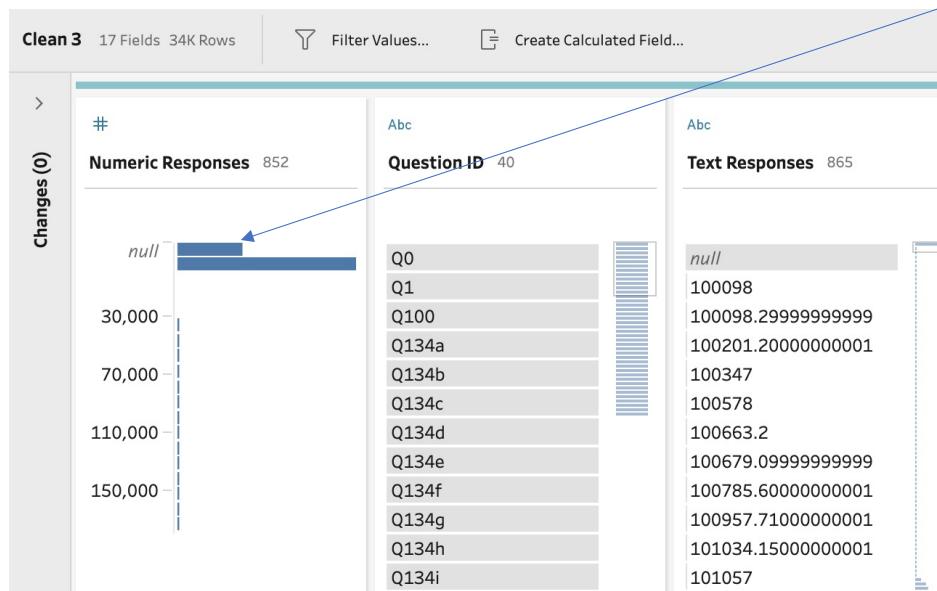


57



3. Exploring Tableau Prep Builder

- To remove Null, in the Profile pane, right click the null bar and select “Exclude”.

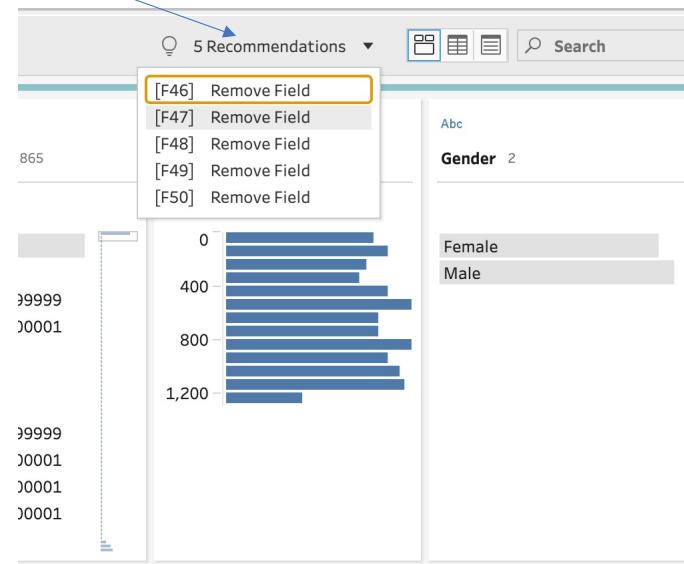


58



3. Exploring Tableau Prep Builder

- Look at the “**5 Recommendations**”
- Remove these 5 irrelevant columns one-by-one, Apply



59



3. Exploring Tableau Prep Builder

- Now, your screen should look something like this:

The screenshot shows the Tableau Prep Builder interface with a data flow diagram and a detailed data preview.

Data Flow Diagram:

```
graph LR; A[Data Labels] --> B[Clean 1]; B --> C[Pivot 1]; C --> D[Clean 2]; D --> E[Join 1]; E --> F[Remove fields]; F --> G[Join 3]; G --> H[Clean 3]; I[Data Numbers] --> J[Pivot 3]; J --> K[Question Helper];
```

Data Preview:

Changes (6):

- Numeric Responses:** 851 rows. Bar chart showing values from 0 to 180,000. Rows include Q0, Q1, Q100, Q134a, Q134b, Q134c, Q134d, Q134e, Q134f, Q134g, Q134h, Q134i.
- Question ID:** 40 rows. Bar chart showing responses from No to Yes. Rows include Q0, Q1, Q100, Q134a, Q134b, Q134c, Q134d, Q134e, Q134f, Q134g, Q134h, Q134i.
- Text Responses:** 864 rows. Bar chart showing responses from No to Yes. Rows include No, Not All Important, Not at all, Not at all satisfied, Of Little Importance, Satisfied, Small degree, Somewhat dissatisfied, Very Important, Very Satisfied, Very high degree, Yes.
- RespiD:** 845 rows. Bar chart showing values from 0 to 1,200. Rows include 0, 400, 800, 1,200.
- Gender:** 2 rows. Bar chart showing Female and Male.
- Location:** 6 rows. Bar chart showing null, Antarctica, Asia, Europe, North America, South America.
- Generation:** 4 rows. Bar chart showing Baby Boomers, Generation X, Millennials, Traditionalists.

Table:

Numeric Responses	Question ID	Text Responses	RespiD	Gender	Location	Generation	Weight	Question ID-1	Wording	Question Grouping	Question Type
0	Q0	No	2	Male	South America	Generation X	1	Q0	Vote in the upcoming election?	Vote	Yes / No / Maybe
0	Q1	No	2	Male	South America	Generation X	1	Q1	Pulse Rate	What do you measure	Check All
98,037,68000000002	Q100	98037,6800000000008	2	Male	South America	Generation X	1	Q100	Salary	Salary	Enter Value
2	Q134a	Small degree	2	Male	South America	Generation X	1	Q134a	Good Job Skills	Indicate degree to which you agree	Likert
2	Q134b	Small degree	2	Male	South America	Generation X	1	Q134b	Good Sense of Humor	Indicate degree to which you agree	Likert
1	Q134c	Not at all	2	Male	South America	Generation X	1	Q134c	High Intelligence	Indicate degree to which you agree	Likert
2	Q134d	Small degree	2	Male	South America	Generation X	1	Q134d	Can Play Jazz	Indicate degree to which you agree	Likert
2	Q134e	Medium degree	2	Male	South America	Generation X	1	Q134e	Indicate degree to which you agree	Indicate degree to which you agree	Likert



3. Exploring Tableau Prep Builder



Finally, we can export the results [i.e., to create a Tableau Data Extract (.hyper) or .csv for Excel]. Here, let's create a file called “**Output.hyper**”

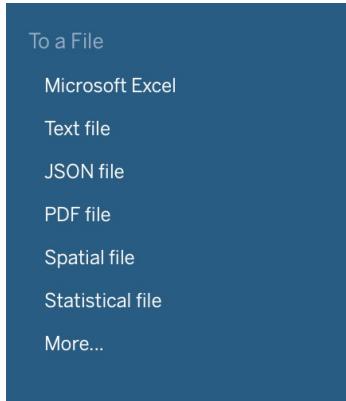
- In the Canvas, click the “+” sign next to the last “Clean” icon to add “**Output**”.
- Click “**Browse**” to choose the output location, and press “**Run Flow**” at the bottom.

The screenshot shows the Tableau Prep Builder interface. On the left, a vertical configuration pane for an 'Output' step is displayed. It includes fields for 'Save output to' (set to 'File'), 'Name' ('Output'), 'Location' ('/Users/kenwong/Desktop'), 'Output type' ('Tableau Data Extract (.hyper)'), and 'Write Options' (with 'Create table' selected). Two blue arrows point from the text above to this pane: one from the 'Browse' button and another from the 'Run Flow' button at the bottom. On the right, a large white box displays the 'Finished Running Flow' summary. It features a green checkmark icon, the text 'Output.hyper', and 'Total time 00:02'. A blue arrow points from the text 'Done!' to the 'Done' button at the bottom right of the summary box.



3. Let's check

- Open up Tableau Desktop
- File > New (wait for 10 seconds)
- Data Source tab, To a File:
More...
 - Output.hyper
- The Extract box should automatically appear in the canvas. If not, drag “Extract (Extract.Extract)” to the Canvas.
- Done!



Extract ID	Question ID	Text Responses	RepID	Gender	Location	Generation	Weight	Question ID-1
0.000000 Q0	No	2	Male	South America	Generation X	1.000000 Q0	Vote in the upcoming...	Vote
0.000000 Q0	No	4	Female	South America	Baby Boomers	1.440000 Q0	Vote in the upcoming...	Vote
1.000000 Q0	Yes	5	Female	South America	Generation X	1.000000 Q0	Vote in the upcoming...	Vote
2.000000 Q0	Don't know	6	Male	Antarctica	Baby Boomers	1.440000 Q0	Vote in the upcoming...	Vote
1.000000 Q0	Yes	9	Female	Europe	Baby Boomers	1.320000 Q0	Vote in the upcoming...	Vote
0.000000 Q0	No	12	Female	Europe	Baby Boomers	1.560000 Q0	Vote in the upcoming...	Vote
1.000000 Q0	Yes	16	Male	Antarctica	Baby Boomers	1.440000 Q0	Vote in the upcoming...	Vote
2.000000 Q0	Don't know	17	Female	Europe	Baby Boomers	1.320000 Q0	Vote in the upcoming...	Vote
0.000000 Q0	No	18	Male	North America	Traditionalist	0.959000 Q0	Vote in the upcoming...	Vote
0.000000 Q0	No	22	Male	South America	Generation X	1.320000 Q0	Vote in the upcoming...	Vote
1.000000 Q0	Yes	25	Female	South America	Generation X	1.320000 Q0	Vote in the upcoming...	Vote
1.000000 Q0	Yes	26	Female	South America	Millennials	0.765000 Q0	Vote in the upcoming...	Vote
1.000000 Q0	Yes	27	Male	Europe	Baby Boomers	1.560000 Q0	Vote in the upcoming...	Vote
0.000000 Q0	No	30	Male	Europe	Baby Boomers	1.320000 Q0	Vote in the upcoming...	Vote



4. Unions in Tableau Desktop



4. New Union in Tableau Desktop

- We've just learned how to use Tableau Prep Builder to "join" different sheets or tables together, so that Tableau Desktop can read it properly.
- However, if your data is spread over several files, or different sheets/tables, you can merge them by creating a **new union**.
 - You do this usually in Tableau Desktop, not Tableau Prep Builder even though it's possible.
 - The "New Union" option will only appear after you've opened up an Excel file, not a .hyper file.

Sheets

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

Sheet1

New Union



4. Join vs. Union

- **Join**

- Sheets/table probably in **different structure**
- You can do it in either Tableau Prep Builder, or Tableau Desktop

- **Union**

- Sheets/table in the **same structure** data in different files or sheets
- You do it in Tableau Desktop



4. Unions

- For best results, the tables that you combine using a union **must have the same structure**.
- That is, each table must have the same number of fields, and related fields must have matching field names and data types.

May2016				June2016				July2016				No file
DAY	CUSTO MER	PURCH ASES	TYPE	DAY	CUSTO MER	PURCH ASES	TYPE	DAY	CUSTO MER	PURCH ASES	TYPE	
4	Lane	5	Credit	1	Lisa	3	Credit	2	Mario	2	Credit	
10	Chris	6	Credit	28	Isaac	4	Cash	15	Wei	1	Cash	
28	Juan	1	Credit	28	Sam	2	Credit	21	Jim	7	Cash	

A union of these tables creates the following single table that contains all rows from all tables.

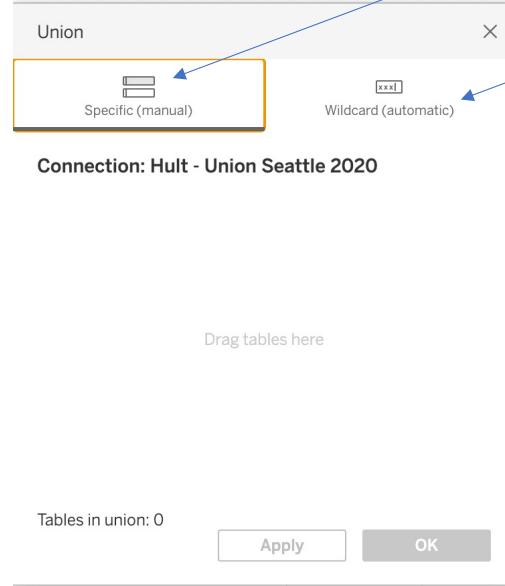
Union

DAY	CUSTOMER	PURCHASES	TYPE
4	Lane	5	Credit
10	Chris	6	Credit
28	Juan	1	Credit
1	Lisa	3	Credit
28	Isaac	4	Cash
28	Sam	2	Credit
2	Mario	2	Credit
15	Wei	1	Cash
21	Jim	7	Cash



4. Unions

- To union different sheets within the same Excel file, do **Specific (manual)** union
- To union different Excel files within a folder, do **Wildcard (automatic)** search:



67



4. Unions

- To union different sheets within the same Excel file, do **Specific (manual)** union:
- In **Tableau Desktop**, **File > New** (wait for 10 secs)
- Open “**Hult – Union Seattle and New York.xlsx**”
 - It has 2 different sheets, titled Seattle and New York

	A	B	C	D
1	Location	Month	Time of Day	Temperature
2	Seattle	April	6:00	5
3	Seattle	April	12:00	17
4	Seattle	April	18:00	11
5	Seattle	May	6:00	9
6	Seattle	May	12:00	20
7	Seattle	May	18:00	15
8	Seattle	June	6:00	12
9	Seattle	June	12:00	25
10	Seattle	June	18:00	18
11				

G10

Seattle

	A	B	C	D
1	Location	Month	Time of Day	Temperature
2	New York	April	6:00	4
3	New York	April	12:00	12
4	New York	April	18:00	9
5	New York	May	6:00	8
6	New York	May	12:00	18
7	New York	May	18:00	13
8	New York	June	6:00	11
9	New York	June	12:00	21
10	New York	June	18:00	16
11				

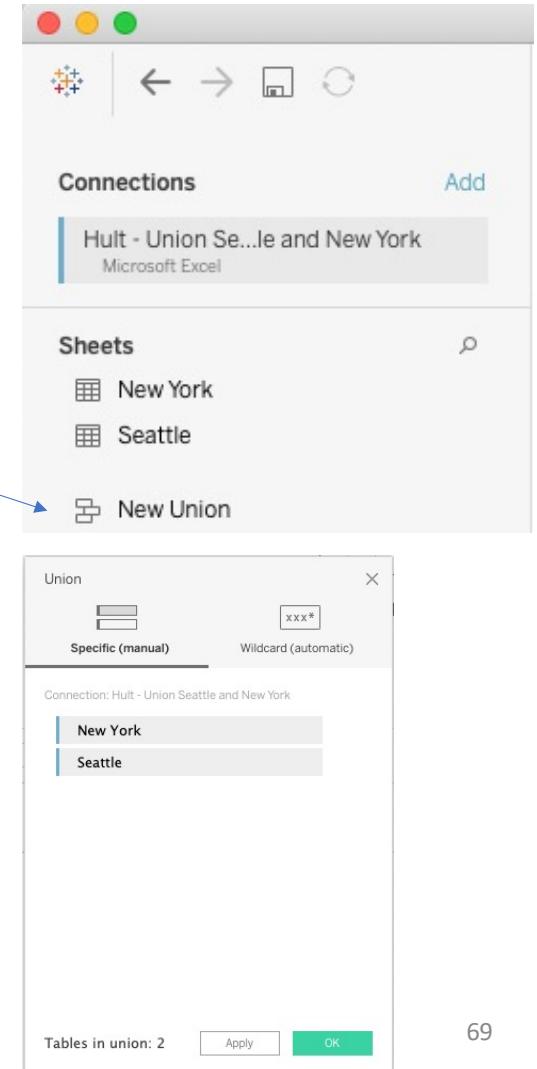
E15

New York



4. Unions

- Drag “New Union” to Canvas
- Under “Specific (manual), drag “New York” and “Seattle” to the Union window, Apply, OK.



69



4. Unions

A union is created.

The screenshot shows the Tableau Data Source interface. On the left, the 'Connections' pane displays a single connection named 'Hult - Union Seattle and New York (Microsoft Excel)'. Below it, the 'Sheets' pane lists three sheets: 'New York', 'Seattle', and 'New Union'. A blue arrow points from the text 'A union is created.' to the 'New Union' sheet icon. The main workspace is titled 'Union (Hult - Union Seattle and New York)' and contains a single data source named 'Union'. A message at the bottom says 'Need more data? Drag tables here to relate them. Learn more'. The data preview table has the following columns and data:

Abc Union Location	Abc Union Month	Union Time of Day	# Union Temperature	Abc Union Sheet	Abc Union Table Name
New York	April	1899-12-30 6:00:00 ...	4	New York	New York
New York	April	1899-12-30 12:00:00...	12	New York	New York
New York	April	1899-12-30 6:00:00 ...	9	New York	New York
New York	May	1899-12-30 6:00:00 ...	8	New York	New York
New York	May	1899-12-30 12:00:00...	18	New York	New York
New York	May	1899-12-30 6:00:00 ...	13	New York	New York
New York	June	1899-12-30 6:00:00 ...	11	New York	New York
New York	June	1899-12-30 12:00:00...	21	New York	New York
New York	June	1899-12-30 6:00:00 ...	16	New York	New York
Seattle	April	1899-12-30 6:00:00 ...	5	Seattle	Seattle
Seattle	April	1899-12-30 12:00:00...	17	Seattle	Seattle
Seattle	April	1899-12-30 6:00:00 ...	11	Seattle	Seattle
Seattle	May	1899-12-30 6:00:00 ...	9	Seattle	Seattle
Seattle	May	1899-12-30 12:00:00...	20	Seattle	Seattle

At the bottom, there are tabs for 'Data Source' (selected), 'Sheet 1' (highlighted in orange), and other options. A 'Go to Worksheet' button is also visible.

Once it's done, you can now close the window,
don't save.

70



4. Unions

- To union different Excel files within a folder, do **Wildcard (automatic)** search:

- You have 2 files in a folder:

- “Hult – Union Seattle **2020.xlsx**”, with a sheet titled “seattle”
 - “Hult – Union New York **2020.xlsx**”, with a sheet titled “newyork”

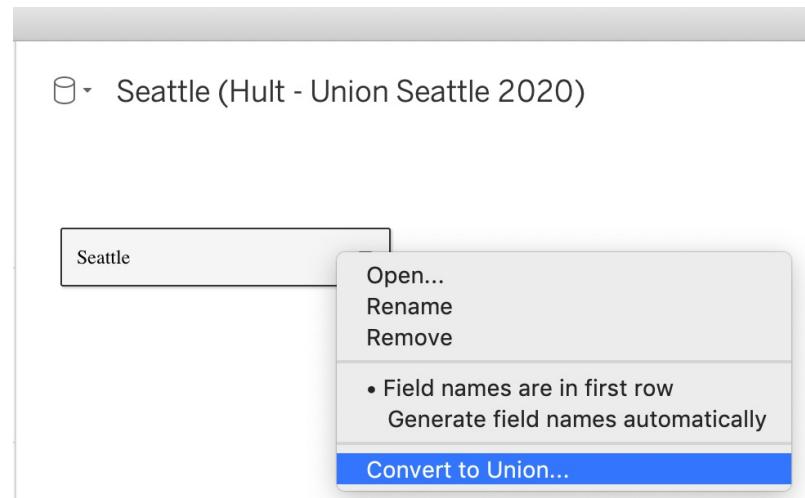


Try to find out which part of the file name is in common! ***2020.xlsx**



4. Unions

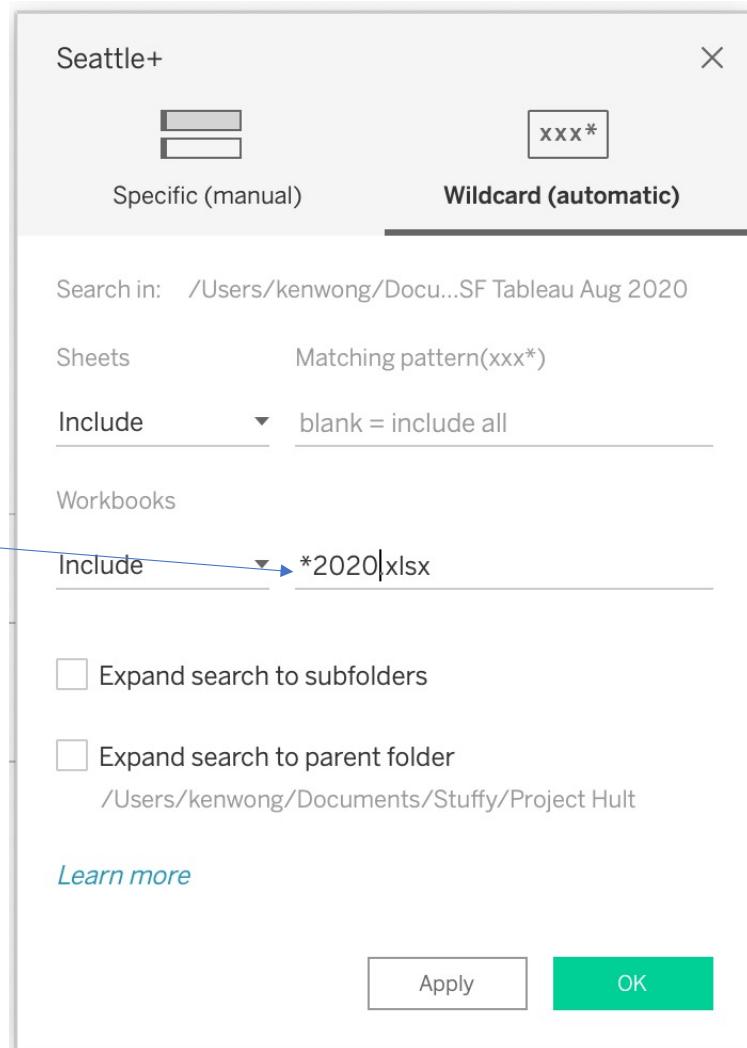
- In Tableau Desktop, **File > New** (wait for 10 secs)
- Connect to “**Hult – Union Seattle 2020.xlsx**”
- In the Canvas, right click on “**Seattle**”, “Convert to Union...”
 - *Alternatively, drag “New Union” to Canvas and then drag “Seattle” to the Union window*





4. Unions

- Go to the “Wildcard (automatic)” tab
- Sheets: Include: *blank*
- Workbooks: Include: ***2020.xlsx**
- Apply, OK
- It will now find the New York data and combine it with the Seattle one
- In the Canvas, right click “Seattle”, rename it as “**Unioned Temperature**”.

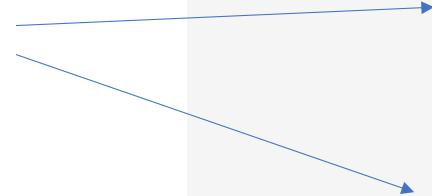




4. Unions

- A union is created.

Both New York and Seattle data are now shown in the same workbook!



The screenshot shows the Tableau interface with a single connection named "Hult - Union Seattle 2020". Under the "Sheets" section, there are two sheets: "Seattle" and "New Union". A tooltip "Unioned Temperature" is visible over a small icon. The main workspace displays a table with the following columns: Abc Seattle+ Location, Abc Seattle+ Month, Time of Day, # Seattle+ Temperature, Abc Seattle+ Path, and Abc Seattle+ Sheet. The data includes rows for New York and Seattle in April, May, and June, with various temperature values and path numbers. The bottom of the screen shows the Tableau ribbon with tabs like Data Source, Sheet 1, etc., and a status bar indicating "Ken Wong".

Abc Seattle+ Location	Abc Seattle+ Month	Time of Day	# Seattle+ Temperature	Abc Seattle+ Path	Abc Seattle+ Sheet
New York	April	1899-12-30 6:00:00 ...	4	SF Tableau Aug 2020...	NewYork
New York	April	1899-12-30 12:00:00...	12	SF Tableau Aug 2020...	NewYork
New York	April	1899-12-30 6:00:00 ...	9	SF Tableau Aug 2020...	NewYork
New York	May	1899-12-30 6:00:00 ...	8	SF Tableau Aug 2020...	NewYork
New York	May	1899-12-30 12:00:00...	18	SF Tableau Aug 2020...	NewYork
New York	May	1899-12-30 6:00:00 ...	13	SF Tableau Aug 2020...	NewYork
New York	June	1899-12-30 6:00:00 ...	11	SF Tableau Aug 2020...	NewYork
New York	June	1899-12-30 12:00:00...	21	SF Tableau Aug 2020...	NewYork
New York	June	1899-12-30 6:00:00 ...	16	SF Tableau Aug 2020...	NewYork
Seattle	April	1899-12-30 6:00:00 ...	5	SF Tableau Aug 2020...	Seattle
Seattle	April	1899-12-30 12:00:00...	17	SF Tableau Aug 2020...	Seattle
Seattle	April	1899-12-30 6:00:00 ...	11	SF Tableau Aug 2020...	Seattle
Seattle	May	1899-12-30 6:00:00 ...	9	SF Tableau Aug 2020...	Seattle
Seattle	May	1899-12-30 12:00:00...	20	SF Tableau Aug 2020...	Seattle



5. Different kinds of Joins in Tableau Desktop



5. Joining – in Tableau Desktop

- File → New; open up **Hult – Superstore.xls**
- We've got different sheets, such as Orders, People, and Returns.
- Let's say we want to combine these 3 sheets together...to add the **account managers name** and **return info** to the order sheet.
 - Are these sheets having the same structure? If Yes, use Union. If No, use Join.
 - Since these 3 sheets have different structures, we'll use **Join** in Tableau Desktop.

Country/Region										
A	B	C	D	E	F	G	H	I	J	
Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer Customer Segment	Customer Customer Segment	Customer Customer Segment	Country/Region City		
1	CA-2018-152156	11/8/18	11/11/18	Second Class	CG-12520 Claire Gut Consumer	United Sts Henderson	United Sts Henderson	United States		
2	CA-2018-152156	11/8/18	11/8/18	Second Class	CG-12520 Claire Gut Consumer	United Sts Henderson	United Sts Henderson	United States		
3	CA-2018-138688	6/12/18	6/16/18	Second Class	DV-13045 Darrin Var Corpone	United Sts Los Angeles	United Sts Los Angeles	United States		
4	US-2017-108966	10/11/17	10/18/17	Standard Class	SQ-02035 Sean O'De Consumer	United Sts Fort Laud.	United Sts Fort Laud.	United States		
5	US-2017-108966	10/11/17	10/18/17	Standard Class	SQ-02035 Sean O'De Consumer	United Sts Fort Laud.	United Sts Fort Laud.	United States		
6	CA-2016-115812	6/9/16	6/14/16	Standard Class	BH-1710 Brossina H Consumer	United Sts Los Angel.	United Sts Los Angel.	United States		
7	CA-2016-115812	6/9/16	6/14/16	Standard Class	BH-1710 Brossina H Consumer	United Sts Los Angel.	United Sts Los Angel.	United States		
8	CA-2016-115812	6/9/16	6/14/16	Standard Class	BH-1710 Brossina H Consumer	United Sts Los Angel.	United Sts Los Angel.	United States		
9	CA-2016-115812	6/9/16	6/14/16	Standard Class	BH-1710 Brossina H Consumer	United Sts Los Angel.	United Sts Los Angel.	United States		
10	CA-2016-115812	6/9/16	6/14/16	Standard Class	BH-1710 Brossina H Consumer	United Sts Los Angel.	United Sts Los Angel.	United States		
11	CA-2016-115812	6/9/16	6/14/16	Standard Class	BH-1710 Brossina H Consumer	United Sts Los Angel.	United Sts Los Angel.	United States		
12	CA-2016-115812	6/9/16	6/14/16	Standard Class	BH-1710 Brossina H Consumer	United Sts Los Angel.	United Sts Los Angel.	United States		
13	CA-2016-115812	6/9/16	6/14/16	Standard Class	BH-1710 Brossina H Consumer	United Sts Los Angel.	United Sts Los Angel.	United States		
14	CA-2016-114412	4/15/19	4/20/19	Standard Class	AA-10480 Andrew A Consumer	United Sts Concord	United Sts Concord	United States		
15	CA-2018-161389	12/5/18	12/10/18	Standard Class	IM-15070 Irene Madi Consumer	United Sts Seattle	United Sts Seattle	United States		
16	US-2017-118983	11/22/17	11/26/17	Standard Class	HP-4815 Harold Pv Home Off	United Sts Fort Worth	United Sts Fort Worth	United States		
17	US-2017-118983	11/22/17	11/26/17	Standard Class	HP-4815 Harold Pv Home Off	United Sts Fort Worth	United Sts Fort Worth	United States		
18	CA-2016-105893	11/11/16	11/18/16	Standard Class	PK-19075 Pete Kriz Consumer	United Sts Madison	United Sts Madison	United States		
19	CA-2016-167164	5/13/16	5/15/16	Second Class	AG-0270 Alejandro Consumer	United Sts West Jord	United Sts West Jord	United States		
20	CA-2016-143336	8/27/16	9/1/16	Second Class	ZD-21925 Zuschuss I Consumer	United Sts San Franci	United Sts San Franci	United States		
21	CA-2016-143336	8/27/16	9/1/16	Second Class	ZD-21925 Zuschuss I Consumer	United Sts San Franci	United Sts San Franci	United States		
22	CA-2016-143336	8/27/16	9/1/16	Second Class	ZD-21925 Zuschuss I Consumer	United Sts San Franci	United Sts San Franci	United States		
23	CA-2018-137330	12/9/18	12/13/18	Standard Class	KB-16585 Ken Black Corpone	United Sts Fremont	United Sts Fremont	United States		

Ryan Ch3

People						
A	B	C	D	E	F	G
1 Person	Region					
2 Anna Andreadi	West					
3 Chuck Magee	East					
4 Kelly Williams	Central					
5 Cassandra Brandow	South					
6						
7						
8						
9						
10						
11						
12						
13						
14						
15						
16						
17						
18						
19						
20						
21						
22						
23						

Ready

Returns	
A	B
1 Returned	Order ID
2 Yes	CA-2016-100762
3 Yes	CA-2016-100762
4 Yes	CA-2016-100762
5 Yes	CA-2016-100762
6 Yes	CA-2016-100867
7 Yes	CA-2016-100867
8 Yes	CA-2016-102652
9 Yes	CA-2016-102652
10 Yes	CA-2016-102652
11 Yes	CA-2016-103373
12 Yes	CA-2016-103744
13 Yes	CA-2016-103744
14 Yes	CA-2016-103940
15 Yes	CA-2016-103940
16 Yes	CA-2016-103940
17 Yes	CA-2016-103940
18 Yes	CA-2016-105229
19 Yes	CA-2016-105229
20 Yes	CA-2016-105270
21 Yes	CA-2016-108609
22 Yes	CA-2016-108861
23 Yes	CA-2016-109918
24 Yes	CA-2016-109918

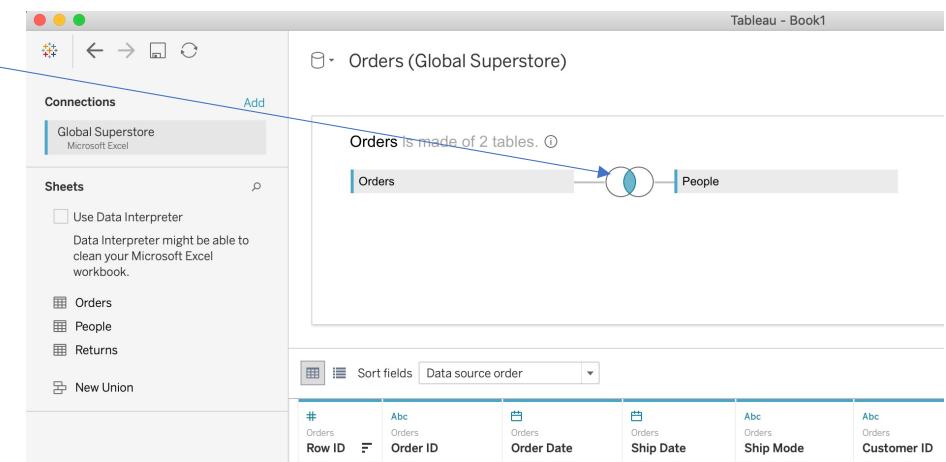
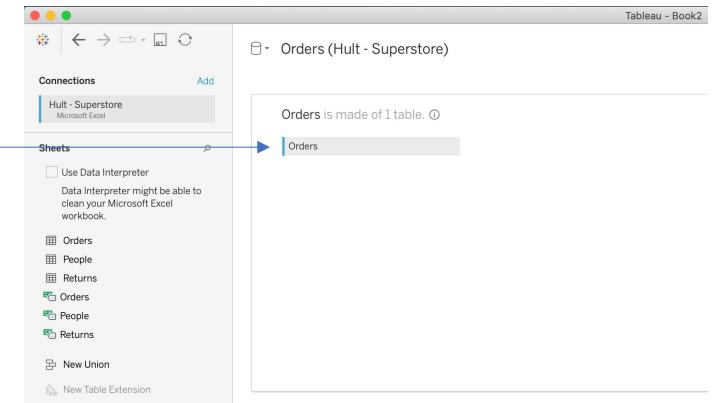
Ready

76



5. Joining – in Tableau Desktop

- Drag “Orders” to Canvas
- Double click “Orders” in the canvas to enter the physical layer
- Drag “People” to the right-hand side of Orders to join
- Click the join icon to change join type



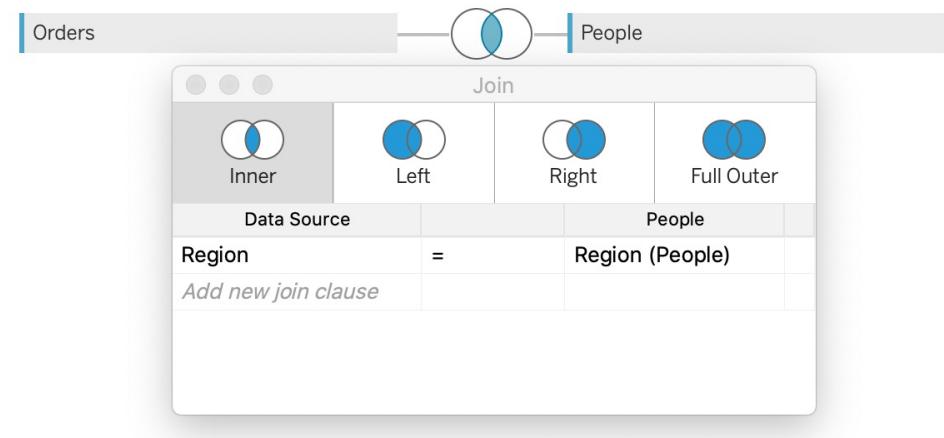
Ryan Ch3



5. Joining – in Tableau Desktop

- Data source is the “Order” sheet
- Since the “Region” column appears in both the “Orders” sheet and the “People” sheet, Tableau automatically select “Region” on both sides.
 - 23 fields, 9994 rows
 - No change after inner join

Orders is made of 2 tables. ⓘ



78



5. Joining – in Tableau Desktop

The screenshot shows a Tableau desktop view with a joined dataset. At the top, there's a tooltip for 'rows' showing '100'. Below the header, there are two newly added columns: 'Region (People)' and 'Person'. The data consists of nine rows with the following values:

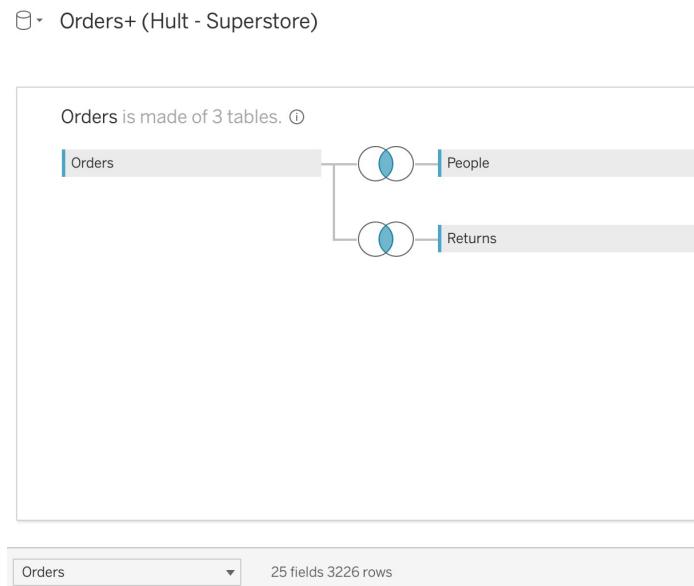
# Orders	# Orders	# Orders	Abc People	Abc People
Quantity	Discount	Profit	Person	Region (People)
2	0.000000	41.91	Cassandra Brandow	South
3	0.000000	219.58	Cassandra Brandow	South
2	0.000000	6.87	Anna Andreadi	West
5	0.450000	-383.03	Cassandra Brandow	South
2	0.200000	2.52	Cassandra Brandow	South
7	0.000000	14.17	Anna Andreadi	West
4	0.000000	1.97	Anna Andreadi	West
6	0.200000	90.72	Anna Andreadi	West
3	0.200000	5.78	Anna Andreadi	West

Newly added columns
You can later remove
the duplicate column
“Region (People)”



5. Joining – in Tableau Desktop

- Drag “Returns” to the right-hand side of Orders to join
- “23 fields, 9994 rows” becomes “25 fields, 3226 rows”



Newly added columns

You can later remove the duplicate column "Order ID (Returns)"

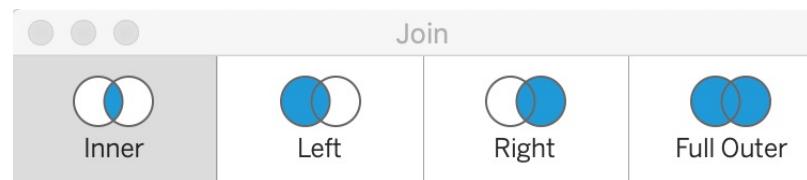
The screenshot shows the Tableau Data View pane. At the top, there is a filter bar with "100" and "rows" buttons. The main area displays a data grid with the following columns:

#	Abc Profit	Abc Person	Abc Region (People)	Abc Returns	Abc Order ID (Returns)
2.4824	Anna Andreadi	West	Yes	CA-2016-143336	
2.4824	Anna Andreadi	West	Yes	CA-2016-143336	
2.4824	Anna Andreadi	West	Yes	CA-2016-143336	
16.0110	Anna Andreadi	West	Yes	CA-2016-143336	
16.0110	Anna Andreadi	West	Yes	CA-2016-143336	
16.0110	Anna Andreadi	West	Yes	CA-2016-143336	
7.3840	Anna Andreadi	West	Yes	CA-2016-143336	
7.3840	Anna Andreadi	West	Yes	CA-2016-143336	
7.3840	Anna Andreadi	West	Yes	CA-2016-143336	



5. Joining – in Tableau Desktop

- **Inner join:** Joins records where there is a **matching field in both datasets**. Using an inner join to combine tables produces a new virtual table that contains values that have matches in both tables.





5. Joining – in Tableau Desktop

- **Left join:** Joins records from the left and right sides of your equation when there is a match.
 - Using a left join to combine tables produces a new virtual table that contains **all values from the left table** and corresponding matches from the right table.
 - When there is no corresponding match from left to right, you will see **a null value**.





5. Joining – in Tableau Desktop

- **Right join:** Joins all the records from the data on the right side of your equation and any matching records from the left side.
 - Using a right join to combine tables produces a table that contains all values from the right and corresponding matches from the left.
 - When a value in the right table doesn't have a corresponding match in the left table, you see a null value.

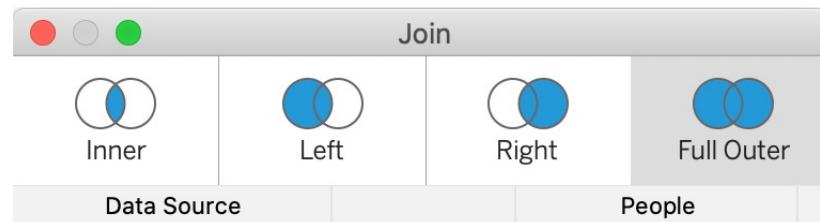


83



5. Joining – in Tableau Desktop

- **Full Outer join:** Joins all the records from each dataset together, even when there is no join—and rarely used. Using a full outer join to combine tables produces a table that contains all values from both tables.
- If a value from either table doesn't have a match with the other table, you see a **null** value.
- *Note: Null = empty cells. To exclude it, right click on it and select exclude.*



84



6. Creating Maps



Presentria.com

Language
English ▾

 Presentria

Session Number (Required) _____

Student ID (Optional) _____

Student Name (Optional) _____

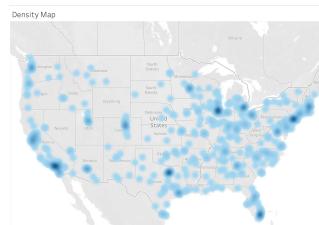
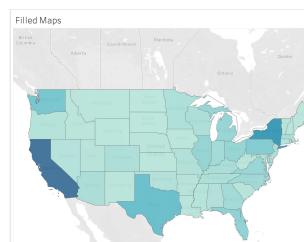
Remember my student ID and student name

JOIN

Maps – US Election



6. Maps – The Anatomy of a Tableau Map



Loth, Ch6

a. Symbol maps

- Locations are marked with circles, squares, or custom shapes
- Form, size, color of these marks can vary

b. Filled maps

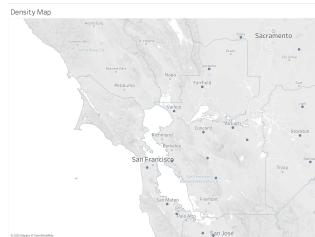
- A.k.a. choropleth maps
- Geographic areas are shaded

c. Density maps

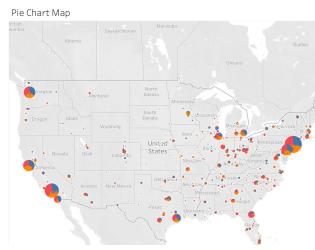
- A.k.a. heatmaps
- Areas of relative concentration are colored intensely



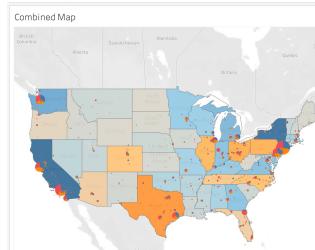
6. Maps – The Anatomy of a Tableau Map



d. Map Layers



e. Pie Chart Map

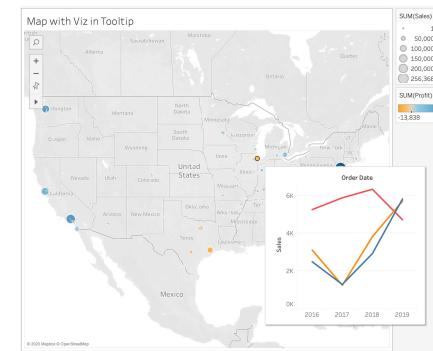


Loth, Ch6

f. Dual Axis Map

g. Viz in Tooltips on Map

You can add secondary information via a pop-up chart in the tooltips





6. Creating Maps

a. Symbol Map



6. Maps – Symbol Maps

- File > New, Connect to **Hult - Superstore.xls**
- Drag “**Orders**” to the Canvas
- Name “Sheet 1” as “**Symbol Map**”
- In the Data pane, if there’s a globe symbol next to the field name, you can just double-click it to generate a map!
- Double-click
 - City
 - Country/Region
 - State
 - And then the measure **Sales**
 - Now, the size of the circles reflects the sum of sales revenue there
- Double-click
 - **Profit**
 - Now, the color of the circles reflects the profits (dark blue is high profit, orange is losing money!)



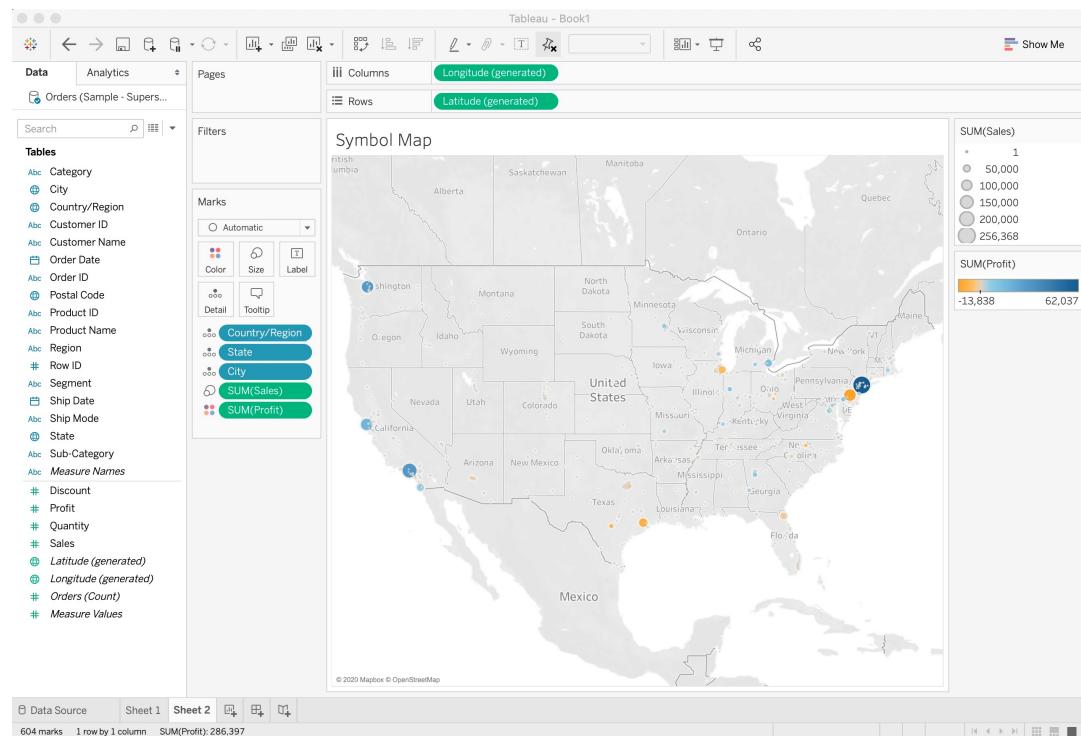
Country/Region



6. Maps – Symbol Maps

Tips:

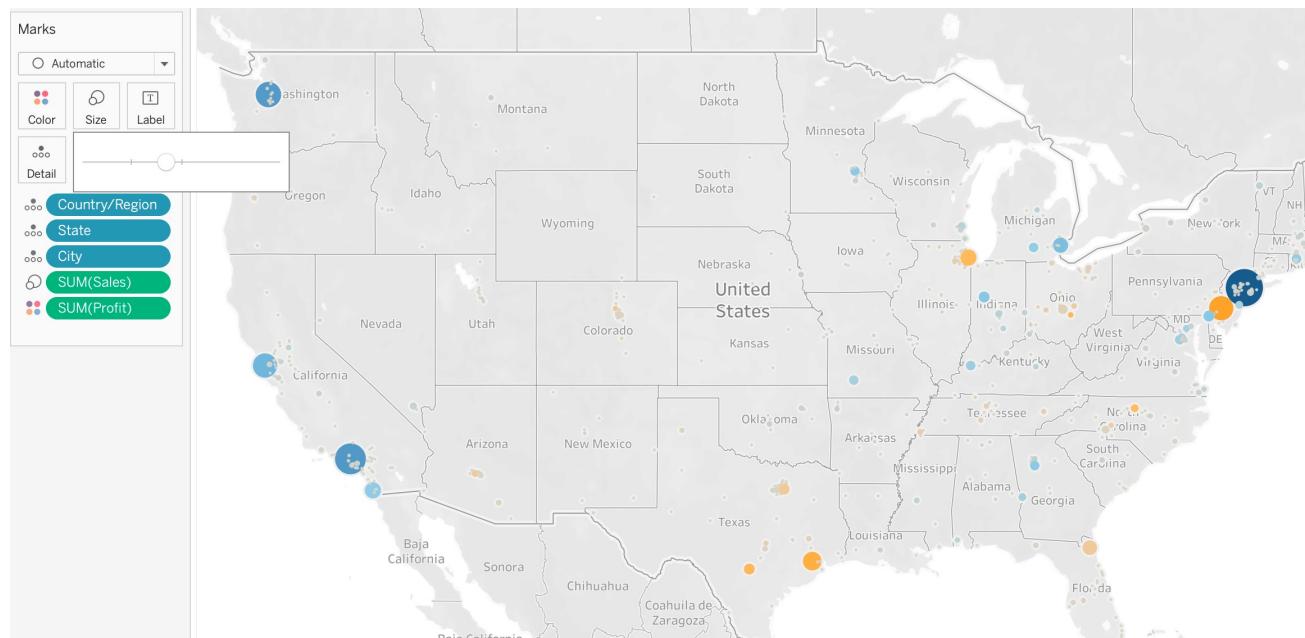
- Use the mouse wheel to quickly zoom in/out of a map
- Holding down the **Shift** key while moving the mouse can pan (drag) the map around.





6. Maps – Symbol Maps

- You can go to the Marks card, click the “Size” button, to adjust the size of the symbol by moving the slider.



92



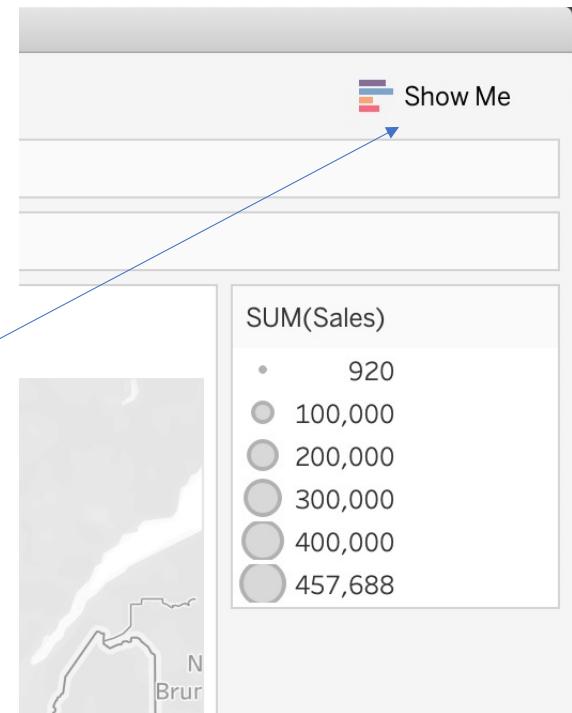
6. Creating Maps

- a. Symbol Map
- b. Filled Map



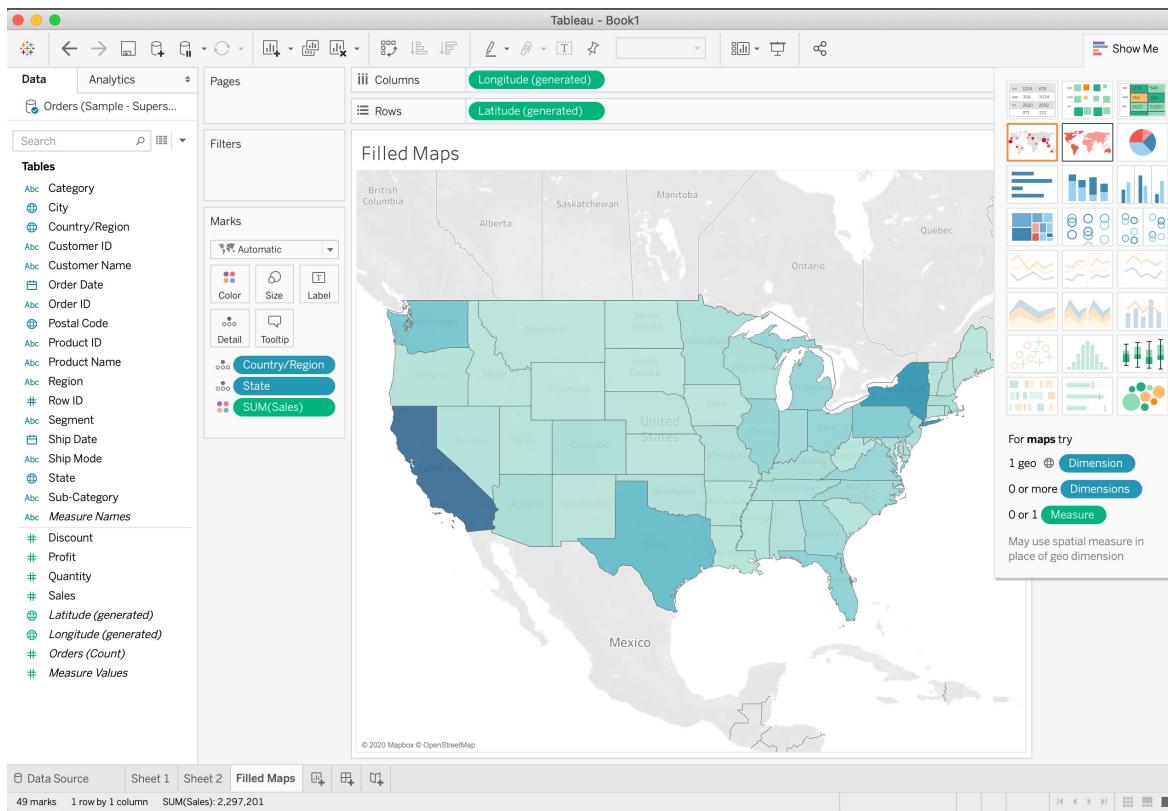
6. Maps – Filled Maps

- Let's say you want to focus on a larger geographical area (e.g., States) instead of a city, we'll use filled maps
- **Hult - Superstore.xls**
- Create a new sheet, call it "**Filled Maps**"
- In the Data pane, double click
 - State
 - Country/Region
 - [City can only be shown as points on the map so we'll not use it here]
 - And then the measure Sales
- Open the **Show Me** menu
 - Click the button for "Filled maps" (2nd row, 2nd one)





6. Maps – Filled Maps



Done!

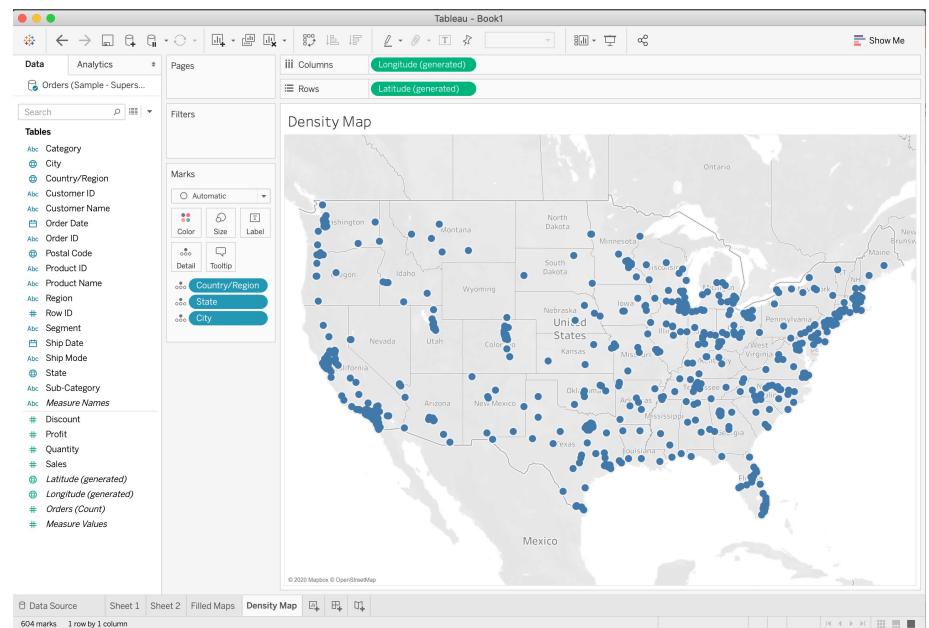
6. Creating Maps

- a. **Symbol Map**
- b. **Filled Map**
- c. **Density Map**



6. Maps – Density Maps (a.k.a. Heatmaps)

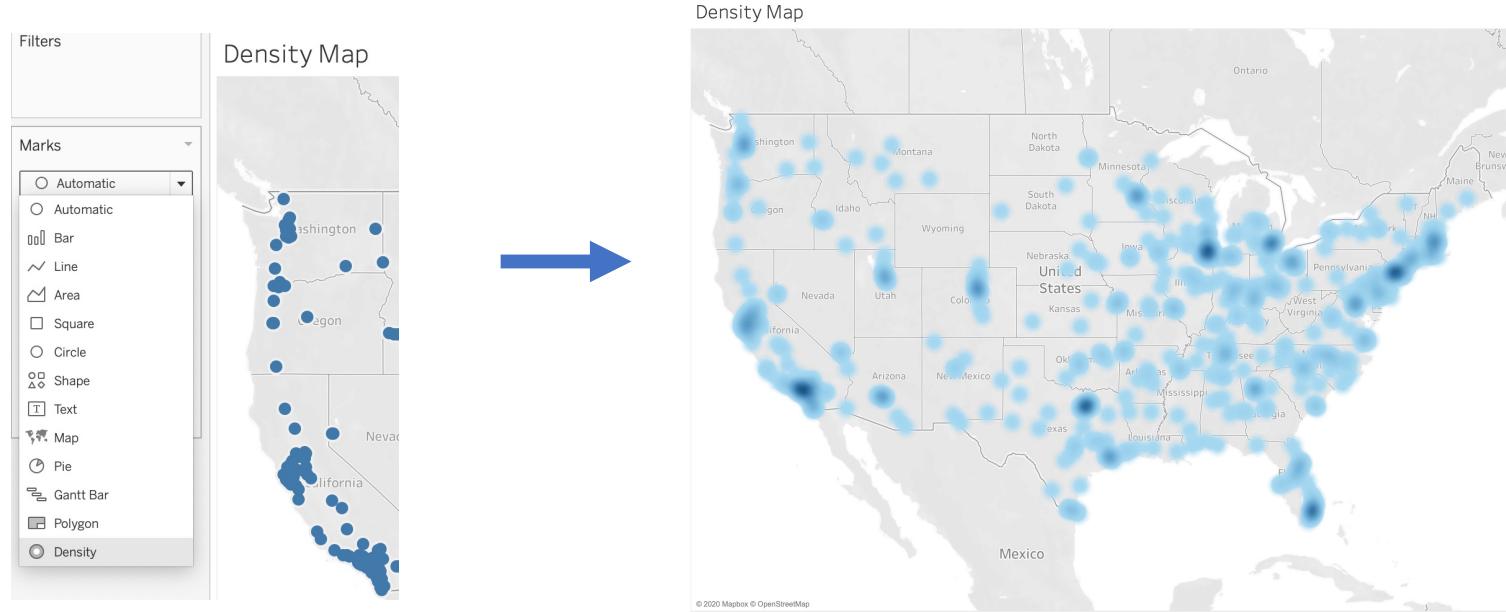
- A good alternative to is to use density maps (getting popular)
- **Hult - Superstore.xls**
- Create a new sheet and call it “**Density Map**”
- Double-click
 - City
 - Country/Region
 - State
- However, it looks too crowded as the dots are overlapping each other.





6. Maps – Density Maps

- So, in the Marks card, change “Automatic” to “**Density**” to create your density map. Done!



98



6. Maps – Density Maps

- Let's think about the kind of data that can utilize this map type.

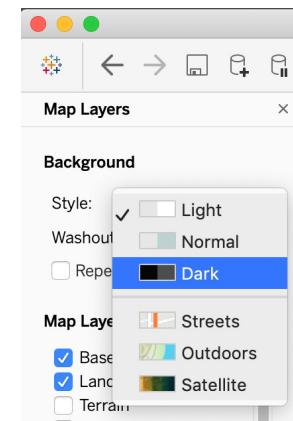
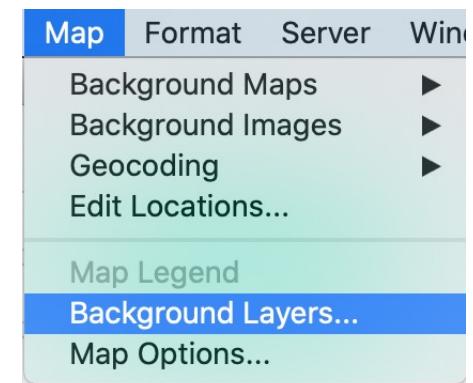
6. Creating Maps

- a. **Symbol Map**
- b. **Filled Map**
- c. **Density Map**
- d. **Map Layers**



6. Maps – The Map Layers

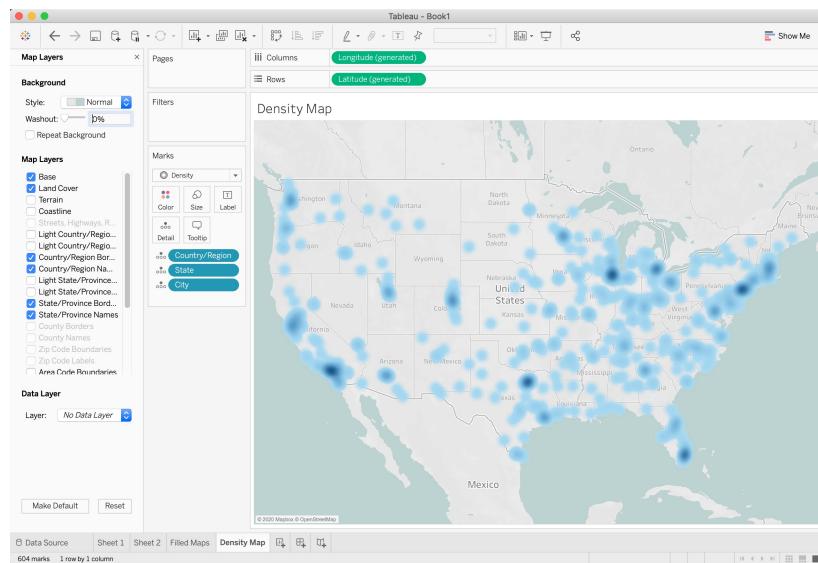
- For this demo, change the density map back to **automatic** in the Marks card.
- Default background map: Light, Normal, and Dark.
- To change it, go to the “map” menu at the top, select **“Background Layers...”**
- The “Map Layers” pane appears on the left and you can select the Background style, such as Normal or **Dark**.



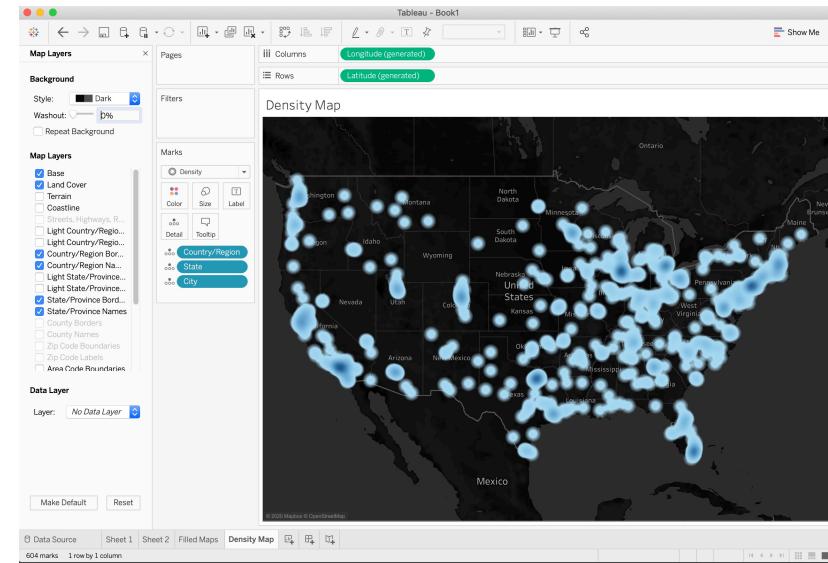
101



6. Maps – The Map Layers



Normal

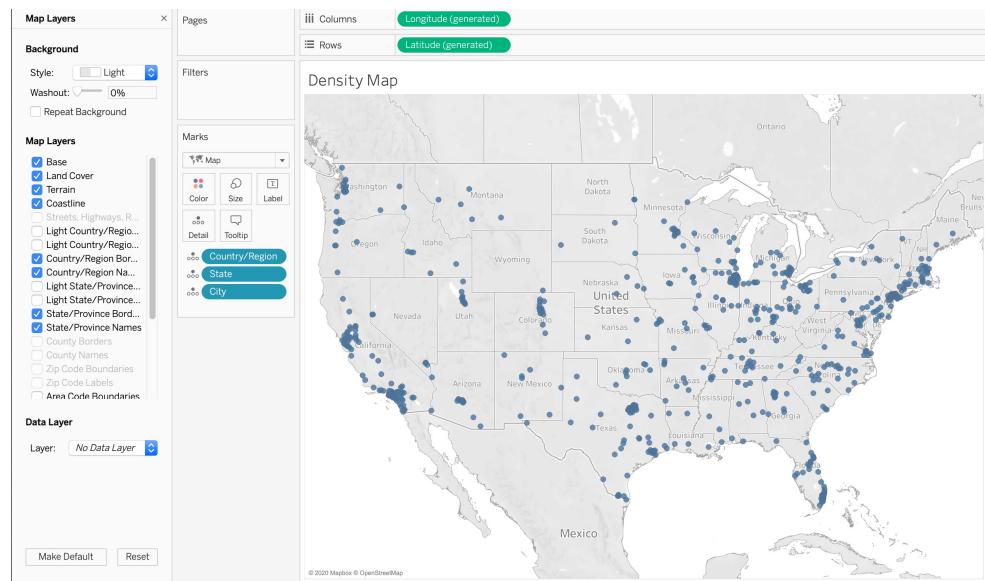


Dark



6. Maps – The Map Layers

- Go back to **Light**
- You can try to check “**Terrain**” and “**Coastline**” to see how it changes on the map!



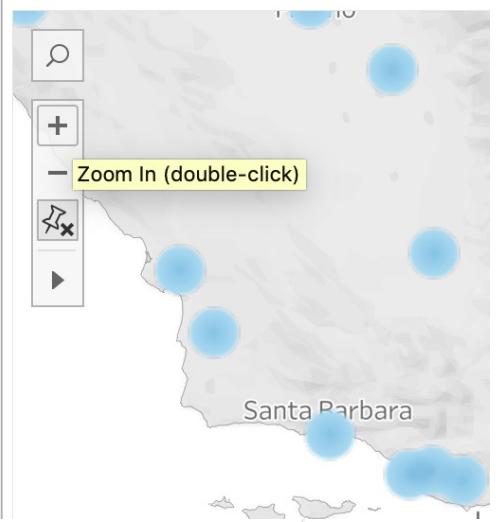
103



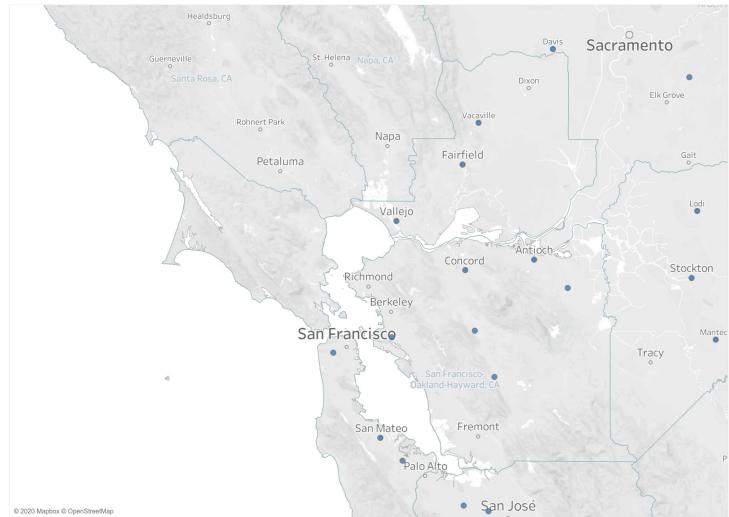
6. Maps – The Map Layers

- If you press “+” to zoom in a bit, you can also check “City”
 - If you don’t see the “+” icon, just click on the map once. Alternatively, you can use your mouse’s scrolling wheel
 - E.g., zoom into San Francisco, and then check “US Metro Boundaries”

Density Maps



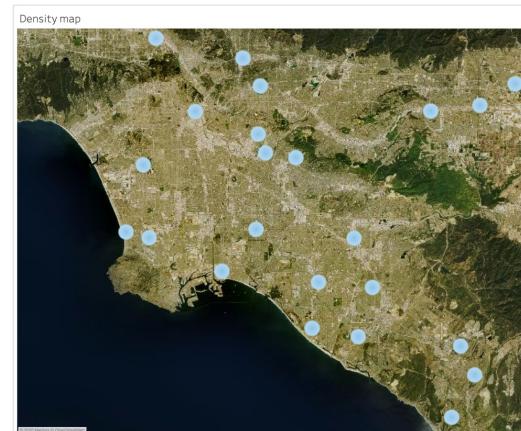
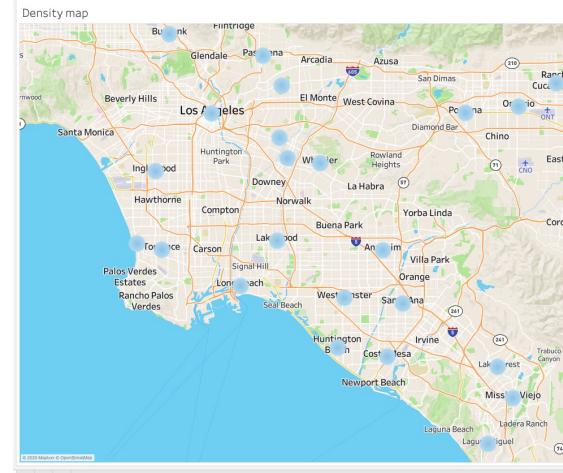
Density Map





6. Maps – The Map Layers

- TIPS: Holding down the **Shift** key while moving the mouse can pan (drag) the map around.
- Zoom In, Background: Streets, Outdoor, Satellite
- Washout



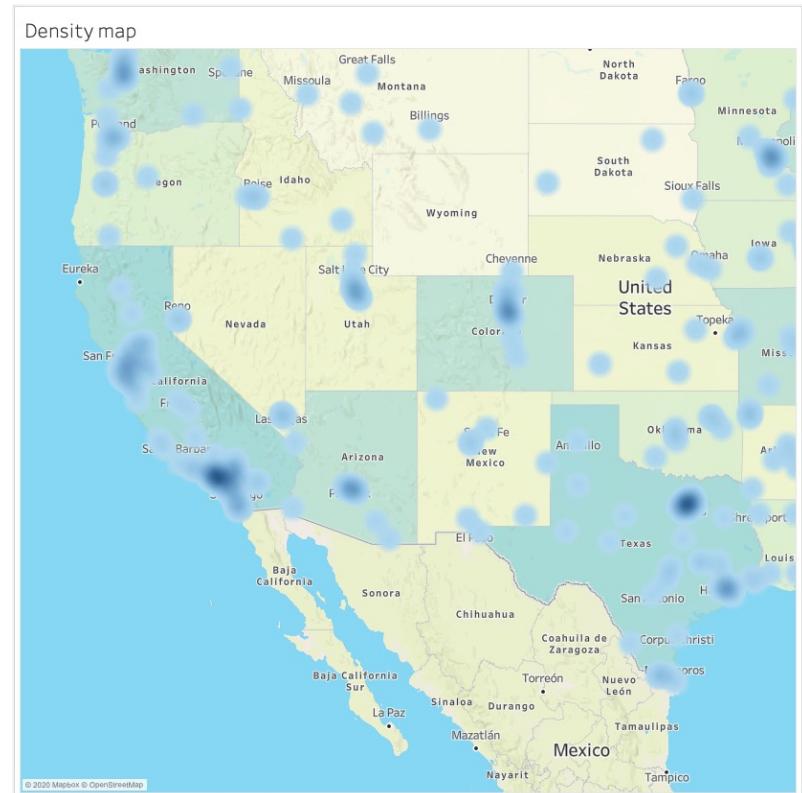
105



6. Maps – The Map Layers

- Data Layer
 - Population
 - By State, County, ZIP Code, Census Tract

What if you're the marketing director for Ann Taylor / Olsen?

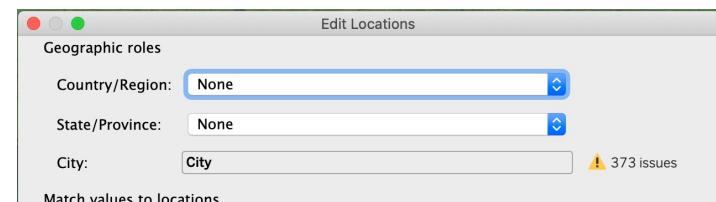
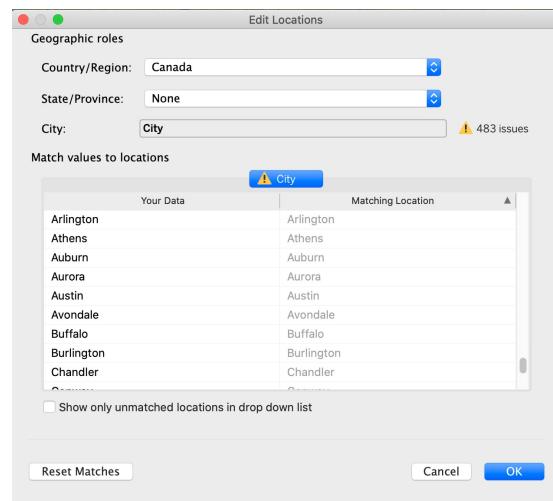
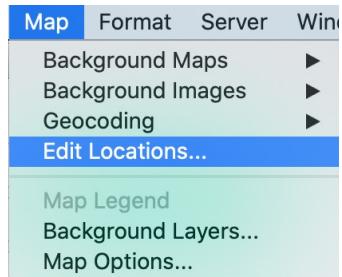


100



6. Maps – Editing the Locations

- If you have non-US geospatial data, remember to set the country correctly under the Map menu.
- If you need to show both US and Canada data, set “Country/Region” as “None”



107

6. Creating Maps

- a. **Symbol Map**
- b. **Filled Map**
- c. **Density Map**
- d. **Map Layers**
- e. **Pie Chart Map**

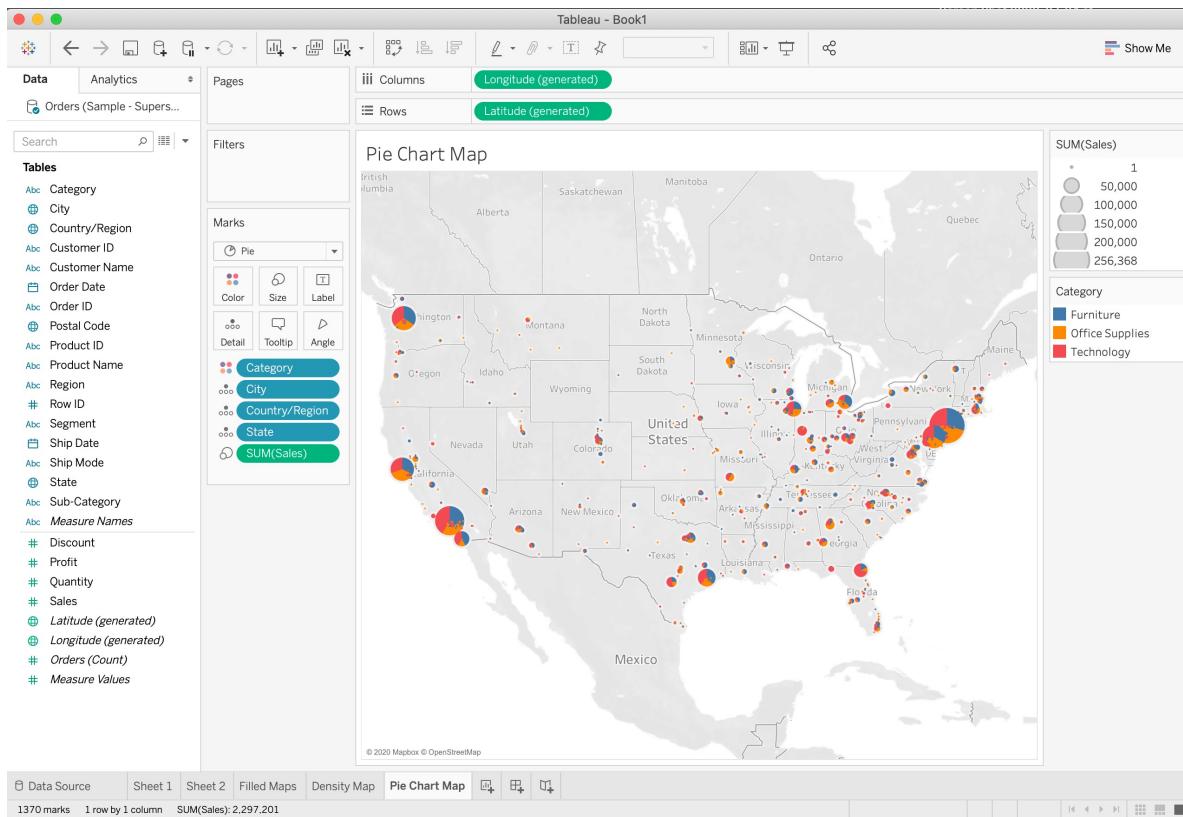


6. Maps – Pie Chart Map

- New sheet, call it “**Pie Chart Map**”
- Double click
 - City
 - Country/Region
 - State
- Drag “**Sales**” to “**Size**” in the Marks card.
- In the Marks card, change “Automatic” to “**Pie**” [Nothing happens]
- Drag “**Category**” to “**Color**” in the Marks card.
- In the Marks card, click “**Size**”, to increase the Pie Chart size.



6. Maps – Pie Chart Maps

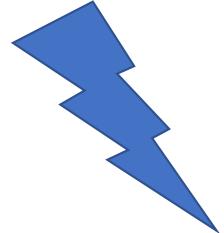


But, it looks a bit too much in the middle.

How about we filter out those smaller cities and just focus on the big ones?



6. Maps – Pie Chart Maps



- Since we only want to exclude some data inside a particular category and not to exclude the whole category, we can't just add Sales to the Filters card.
- Hence, we have to use a **Level of Detail (LOD) Expression** to create a **new calculated field**.
 - We'll learn more about it the next few sessions, but here's an example.

I want to filter the map based on sales revenue!

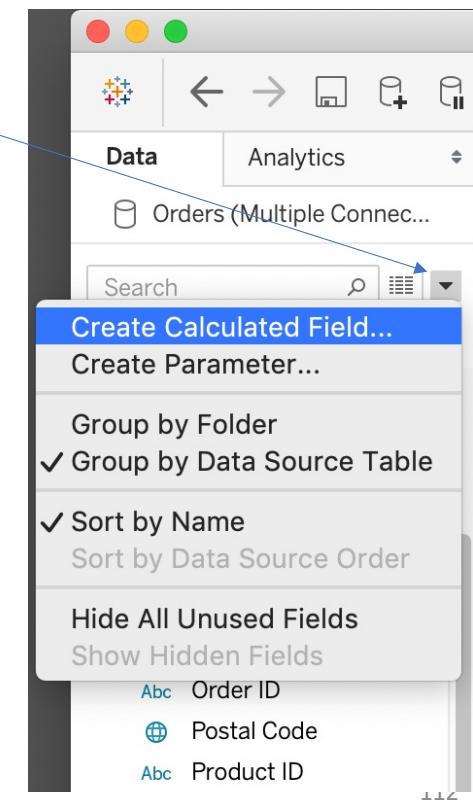
You cannot just add Sales to the Filters card, as you would filter the individual categories. So, use the above LOD Expression to create a new measure for filtering purpose.

111



6. Maps – Pie Chart Maps

- Data pane, click the little arrow at the top
- Click “Create Calculated Field...”





6. Maps – Pie Chart Maps

- Replace “Calculation1” with “**Sales (w/o category)**”
- Below the line, type the following:
{EXCLUDE[Category]:SUM([Sales])}
- *“I want to exclude something within the category and that is based on the sum of sales”*
- Click Apply, OK



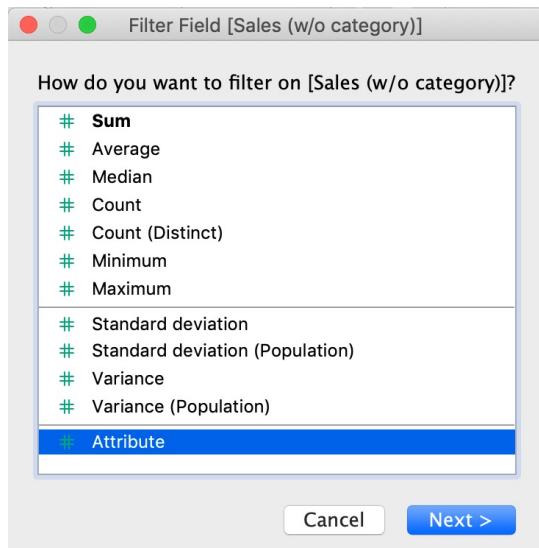
{EXCLUDE[Category]:SUM([Sales])}

113



6. Maps – Pie Chart Maps

- Then, add this newly created field (measure or pill) to the **Filters card**, select “Attribute”, next, Apply, OK.

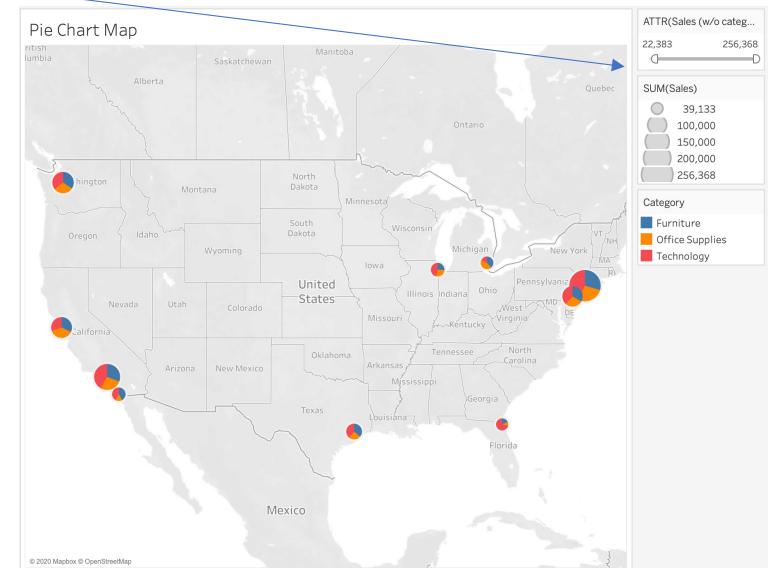
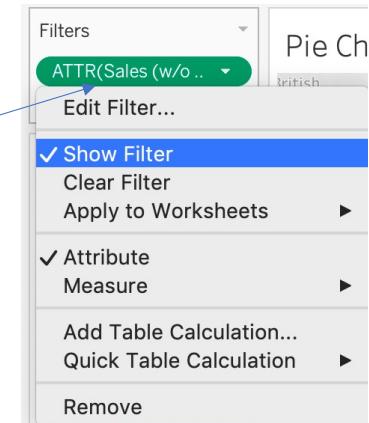


The screenshot shows the Tableau Data pane with the 'Analytics' tab selected. The 'Tables' section lists various dimensions and measures. The 'Filters' card is open, showing a single pill named 'ATTR(Sales (w/o categor...))'. The 'Marks' section shows options for Color, Size, Label, Detail, and Tooltip. The 'Dimensions' section includes Country/Region, State, City, and various generated latitude and longitude fields. The 'Measures' section includes SUM(Sales) and SUM(Profit).



6. Maps – Pie Chart Maps

- In the Filters card, right-click “ATTR(Sales(w/o cat..”, select “Show Filter”
- In the Slider control, you can now filter the map based on sales revenue! It’s super cool, isn’t it!?



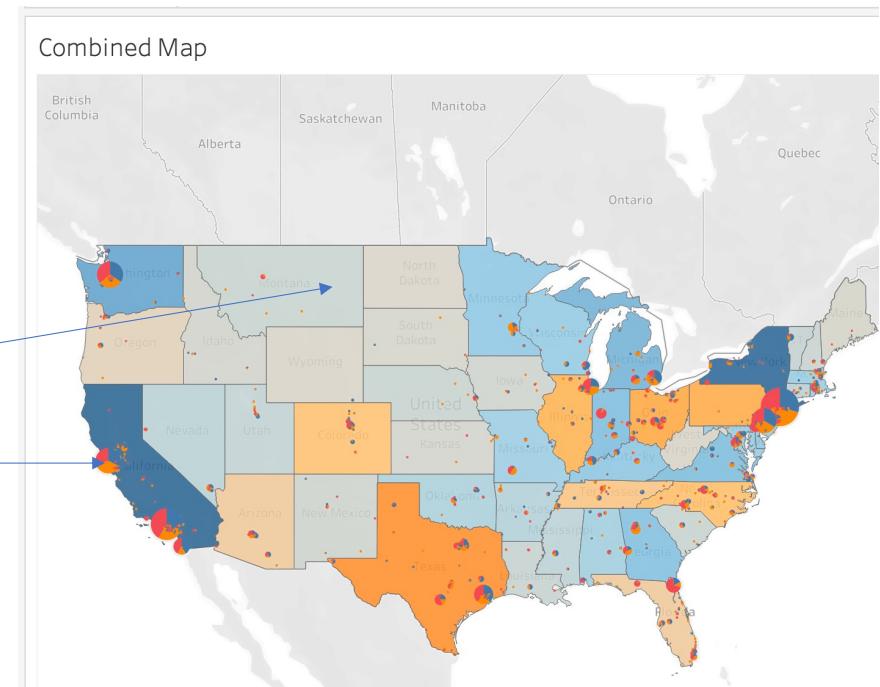
6. Creating Maps

- a. **Symbol Map**
- b. **Filled Map**
- c. **Density Map**
- d. **Map Layers**
- e. **Pie Chart Map**
- f. **Dual Axis Map**



6. Maps – Dual Axis Map

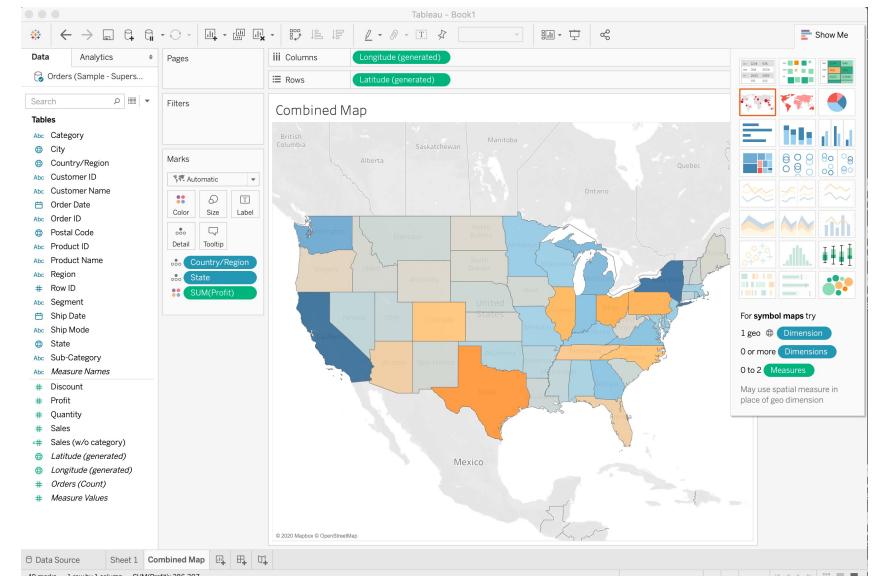
- How about we create a Dual Axis Map in which we combine 2 maps into one?
 - The first map is a **filled map** showing profit
 - The second map is a **pie chart map** showing revenue and the 3 categories





6. Maps – Dual Axis Map

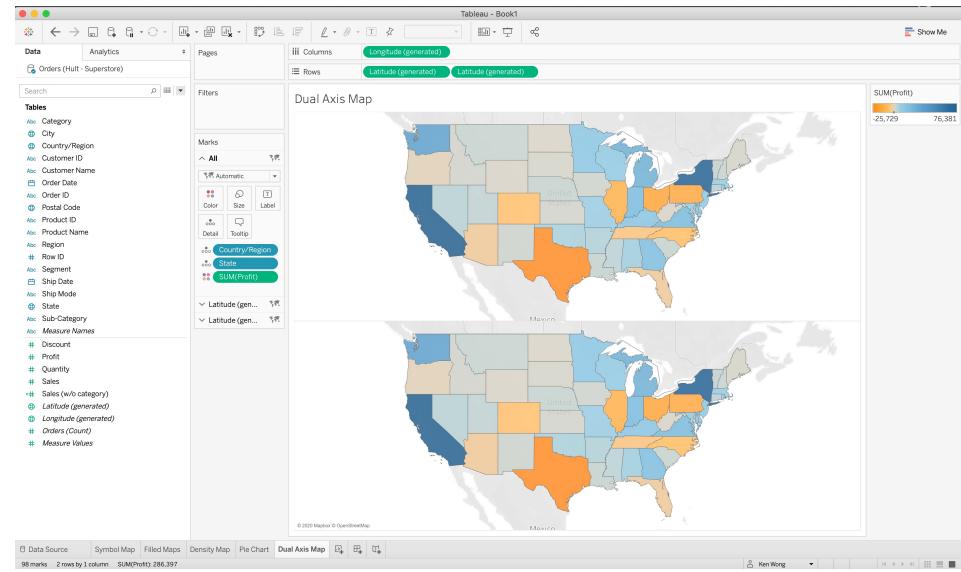
- Create a new sheet, call it “**Dual Axis Map**”
- Now, let’s build the 1st map (filled map):
- In the Data pane, double click
 - State
 - Country/Region
 - [City can only be shown as points on the map so we’ll not use it here]
 - And then the measure **Profits**





6. Maps – Dual Axis Map

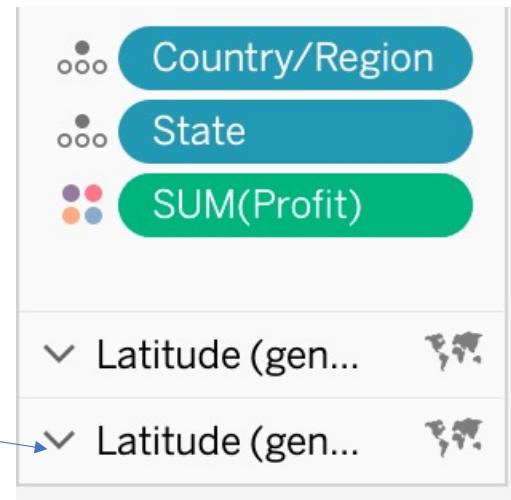
- Open the Show Me menu
 - Click the button for “Filled maps” (2nd row, 2nd one). Now, the filled map showing Profits is created.
- **Now, let's create the 2nd map**
 - *Stay on the same page*
- In the Data pane, drag “Latitude (Generated)” to Rows.
 - Yes, it's appearing twice! That's why it's called dual axis.





6. Maps – Dual Axis Map

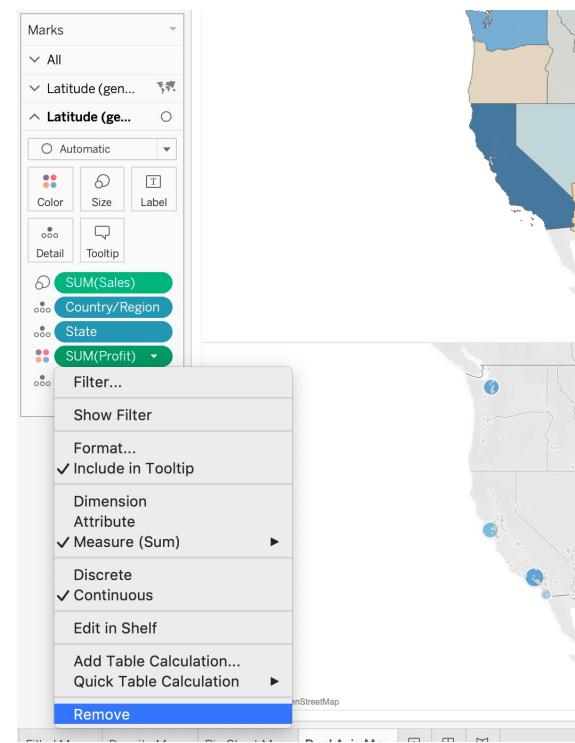
- You'll see that in the Marks card, there are two "Latitude (gen..)" showing the 2 different maps.
- Now, click the 2nd "Latitude (gen..)" in the Marks card





6. Maps – Dual Axis Map

- Now, let's continue to create the 2nd map (Pie Chart Map) on the same page:
- Drag “**City**” to the Marks card and **put it below the pills**.
- Drag “**Sales**” to **Size**.
- Remove “**SUM(Profit)**”
 - Right click on it, select Remove

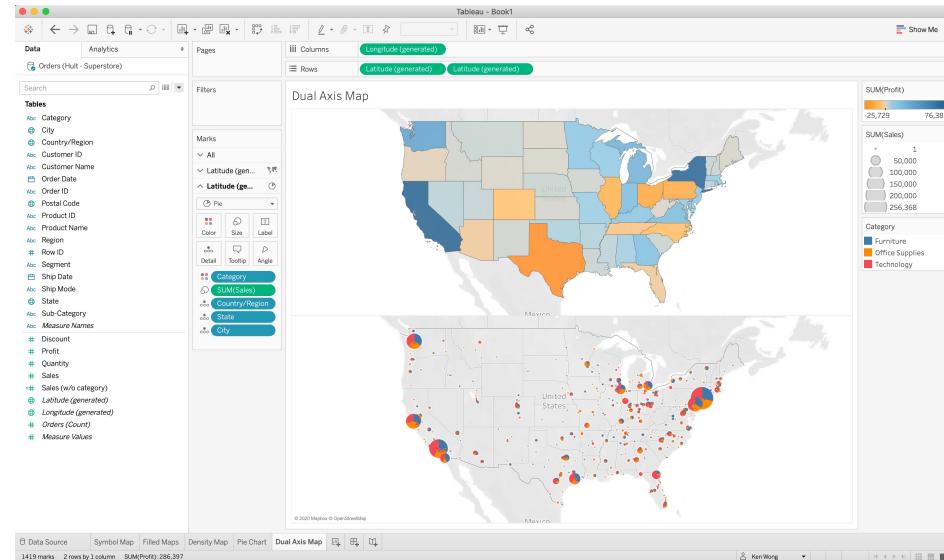


121



6. Maps – Dual Axis Map

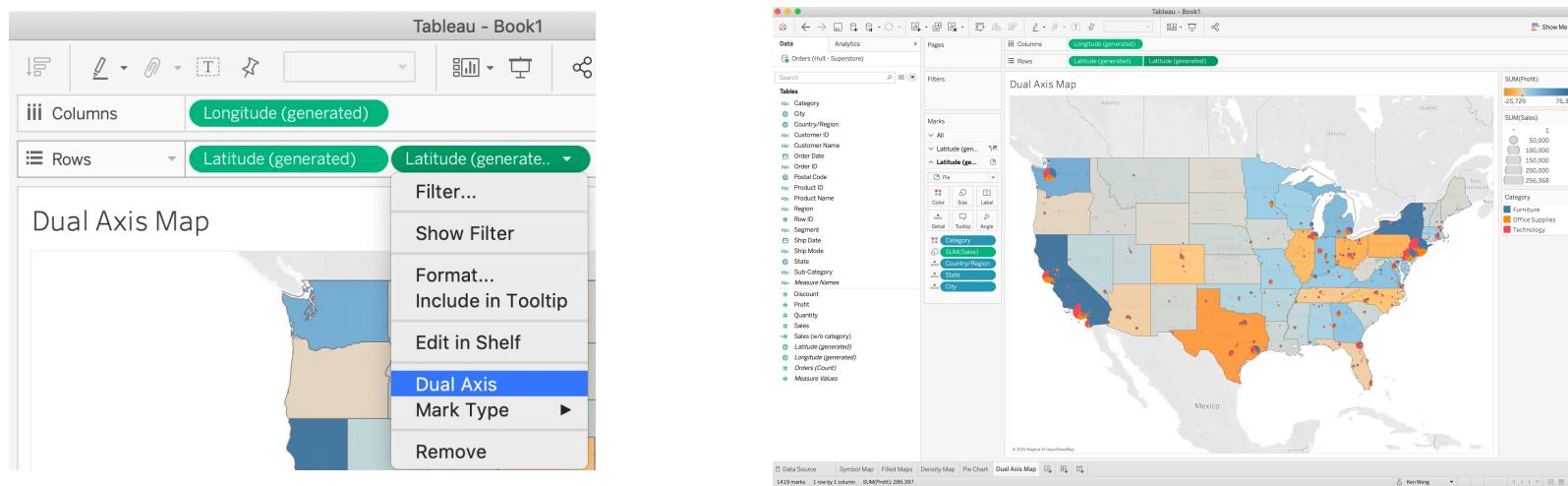
- In the Marks card, change “Automatic” to “Pie” [Nothing happens]
- Drag “Category” to Color in the Marks card.
- In the Marks card, click size, increase the size of the Pie Chart





6. Maps – Dual Axis Map

- Now, it's time to combine the two maps. Go to the 2nd “Latitude (generated)” pill in the Rows, click the context menu on the right, and select “Dual Axis”! Done!



123

6. Creating Maps

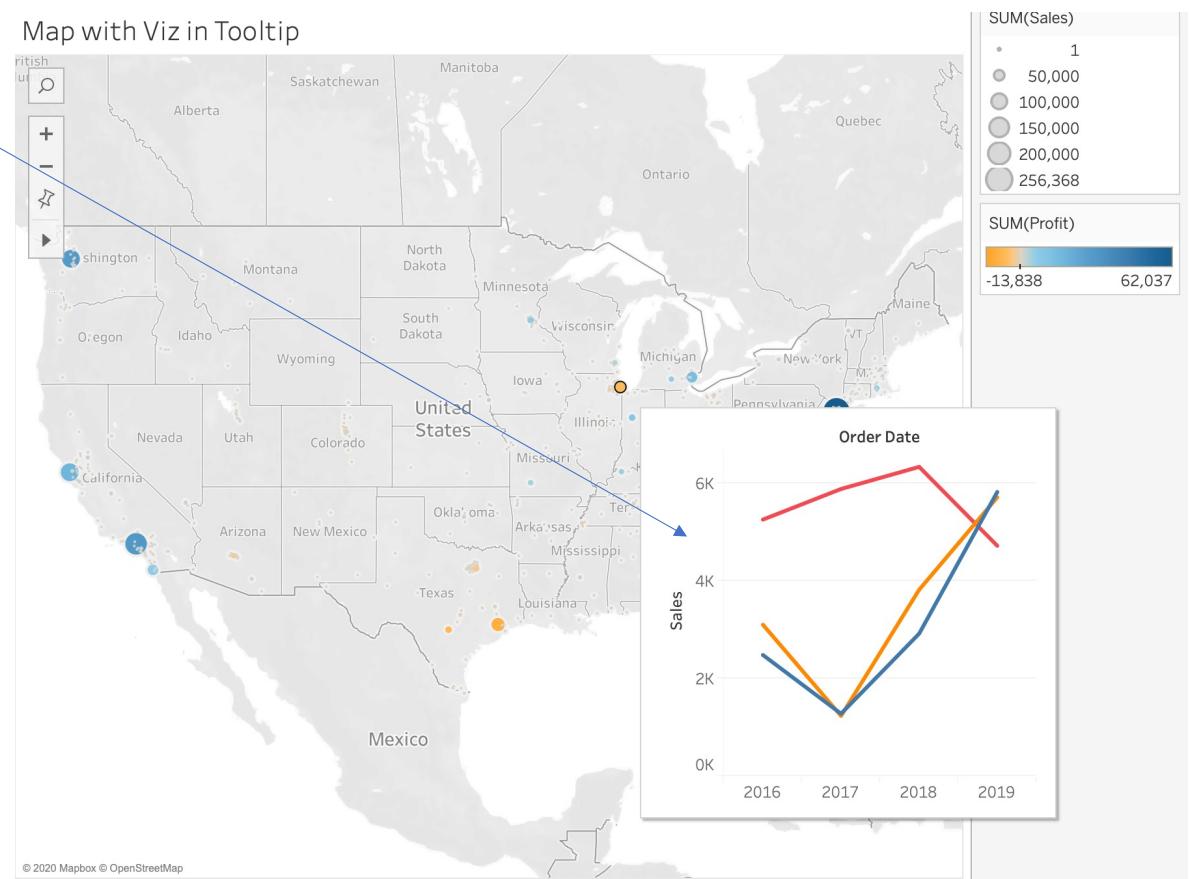
- a. Symbol Map
- b. Filled Map
- c. Density Map
- d. Map Layers
- e. Pie Chart Map
- f. Dual Axis Map
- g. Viz in Tooltips on Map



6. Viz in Tooltip

- It's a great way to add information when you hover over individual marks on a map.
- Not restricted to map, but very useful in maps

See Tableau Public example



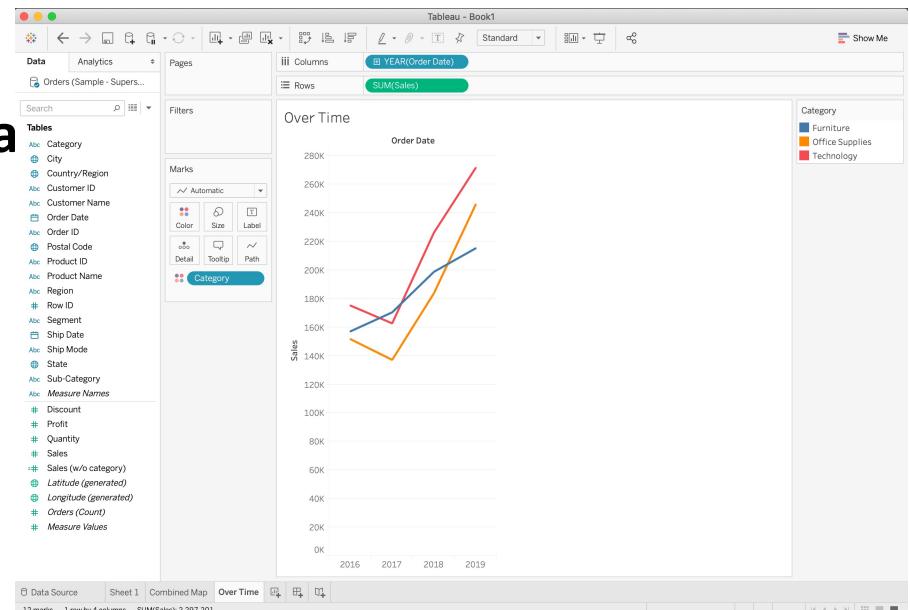


6. Viz in Tooltip

- Hult - Superstore.xls

Step 1: Create the smaller, second chart

- New sheet, call it **Over Time**
 - Sales > Rows
 - Order Date > Columns
 - Category > **Color** in Marks



126



6. Viz in Tooltip

Step 2: Create the larger, first/main chart

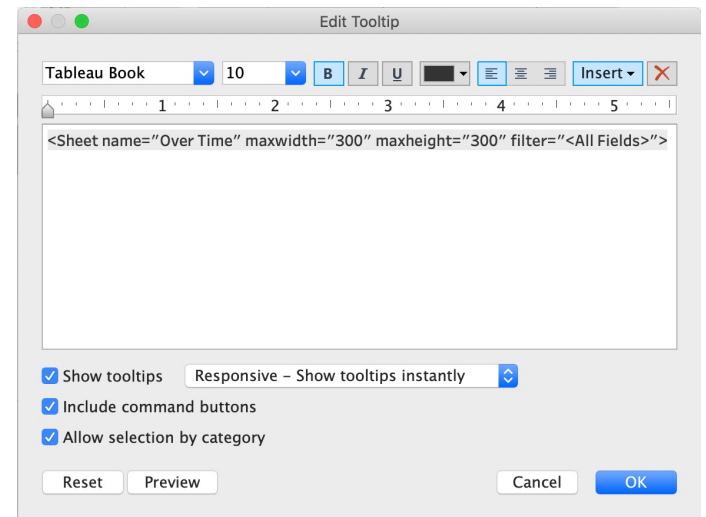
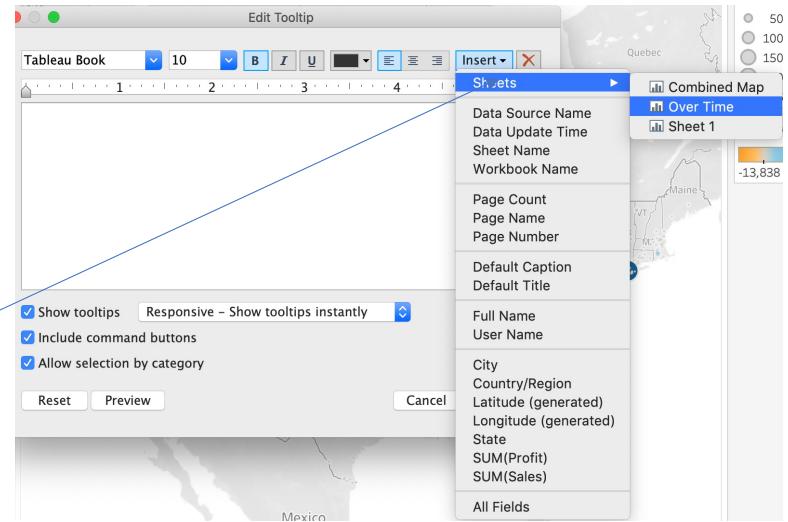
- Create a symbol map as the main chart in a new sheet called “Map with Viz in Tooltip”
- Double-click
 - City
 - Country/Region
 - State
 - And then the measure Sales
 - Now, the size of the circles reflects the sum of sales revenue there
- Double-click
 - Profit
 - Now, the color of the circles reflects the profits (dark blue is high profit, orange is losing money!)



6. Viz in Tooltip

Step 3: Embedding the Chart in Tooltips

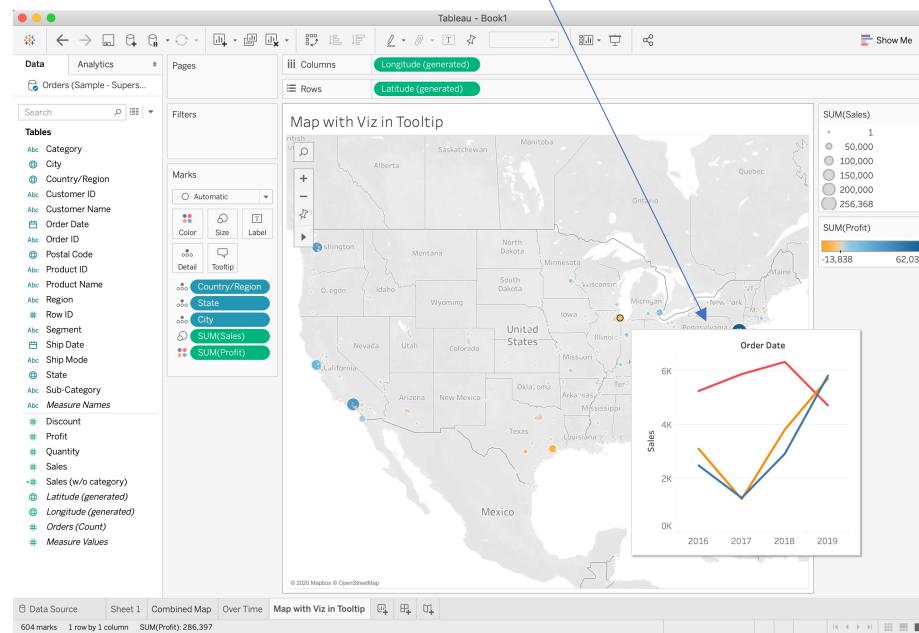
- Click “**Tooltips**” in the Marks card.
- Delete the settings in the middle of the “Edit Tooltip” window
 - Highlight everything, delete
- Click the “**Insert**” button in the upper right-hand corner, and select Sheets > “**Over Time**”, OK.





6. Viz in Tooltip

- Now, the Over Time data is shown when I hover over different cities on the symbol map. Done!



129

6. Creating Maps

- a. **Symbol Map**
- b. **Filled Map**
- c. **Density Map**
- d. **Map Layers**
- e. **Pie Chart Map**
- f. **Dual Axis Map**
- g. **Viz in Tooltips on Map**
- h. **Mapbox**



6h. Mapbox.com

- Log into Mapbox.com
- Account (upper-right hand corner) → **Studio**
- “New Style” button
- Choose a template: **Monochrome** (pink)
 - Customize Monochrome
- “Styles” (upper-left hand corner)

The screenshot shows the Mapbox Studio interface. At the top, a dark header bar displays "Styles > Monochrome" and "Edited 1 minute ago". Below the header are tabs for "Components", "Layers", and "3D". A toolbar contains icons for adding layers, search, and other editing tools. The main area shows a pink-toned map of a residential area labeled "Bloomfield". At the bottom left, there's a note about "Administrative boundaries".

The screenshot shows the Mapbox Studio account settings and template selection interface. At the top right, the account name "designerhouse" and handle "@M_1" are visible, along with a "Paused" status indicator. Below this is a sidebar with links for "Settings", "Account", "Studio", and "Documentation". Under "Account", the "Sign out" option is highlighted. The main area is titled "Choose a template" and shows several template options: "San Francisco Streets" (selected), "San Francisco Monochrome", "Outdoors", "Navigation", "Satellite Streets", and "Blank". A "Customize Streets" button is located at the bottom right of the template section.



6h. Mapbox.com

- Click the “Share Your Style” icon

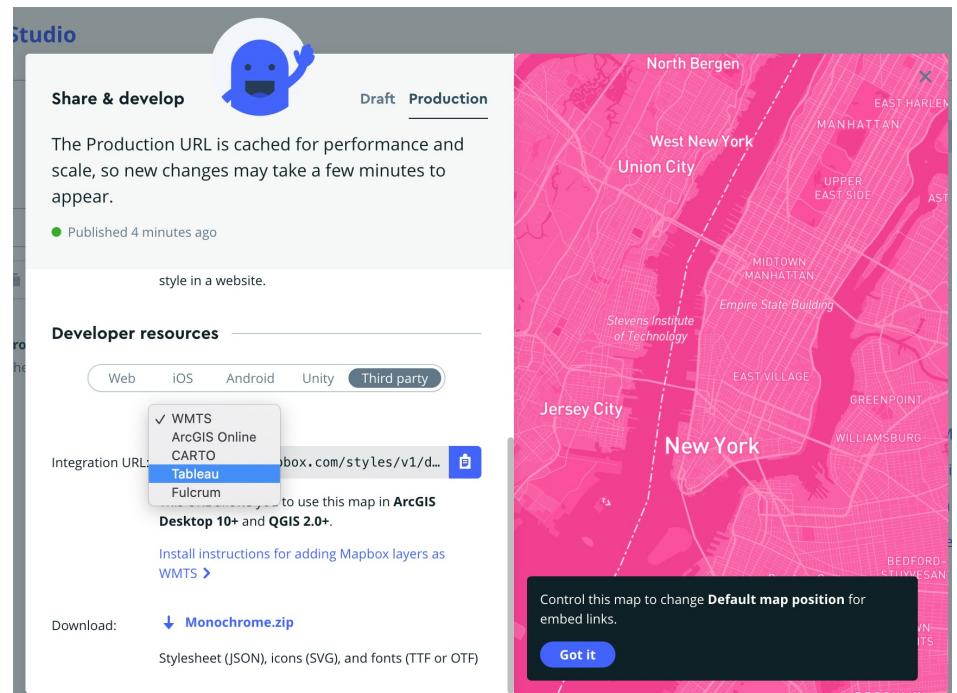
The screenshot shows the Mapbox Styles interface. At the top, there are buttons for "New folder" and "New style". Below that is a search bar with a magnifying glass icon and the word "Search". To the right of the search bar are navigation arrows and icons for creating a new folder or style. Underneath, there are filters for "0 selected" and "All" (with a dropdown arrow), and sorting options for "Date edited" (also with a dropdown arrow). A message indicates "1-1 of 1 file". The main list area shows a single item: "Monochrome" by "ew Yor" (likely a placeholder for "New York"). This item has a small thumbnail image, the name "Monochrome", and a status message "Published 29 minutes ago". To the right of this item are three icons: a share icon, a download icon, and a more options icon (three dots). At the bottom of the list area are two navigation arrows.



6h. Mapbox.com

- Select “Third Party” → Tableau in the pull-down menu
- Then, copy the “Integration URL” by clicking that blue icon

Tableau ▾
Integration URL: <https://api.mapbox.com/styles/v1/d...> 

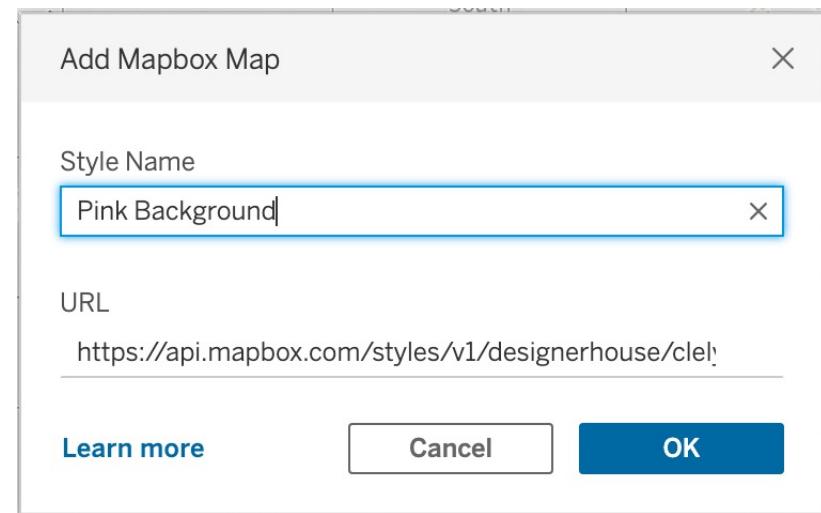
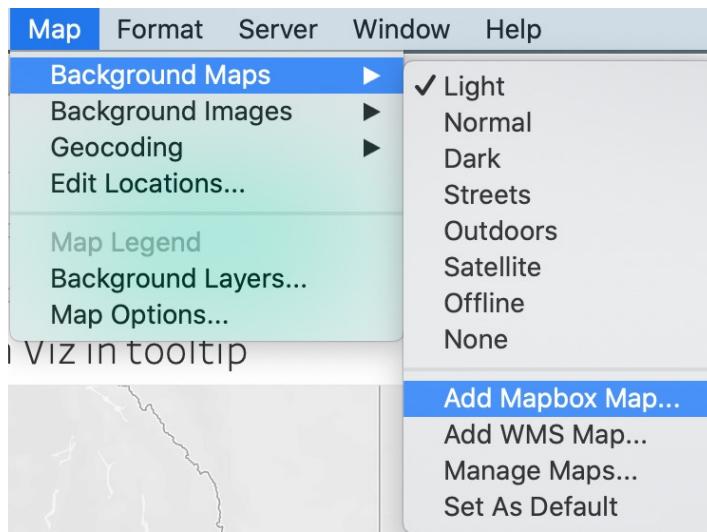


The screenshot shows the Mapbox Studio interface. On the left, there's a sidebar with a blue cartoon character icon. Below it, the "Share & develop" section shows a message about cached production URLs. It includes a green "Published 4 minutes ago" badge. In the "Developer resources" section, a dropdown menu is open over a "Third party" button, showing options like WMTS, ArcGIS Online, CARTO, Tableau (which is highlighted in blue), and Fulcrum. An integration URL is provided for Tableau. To the right, a map of New York City is displayed with various neighborhoods labeled. A black overlay box at the bottom right says "Control this map to change Default map position for embed links." and has a "Got it" button.



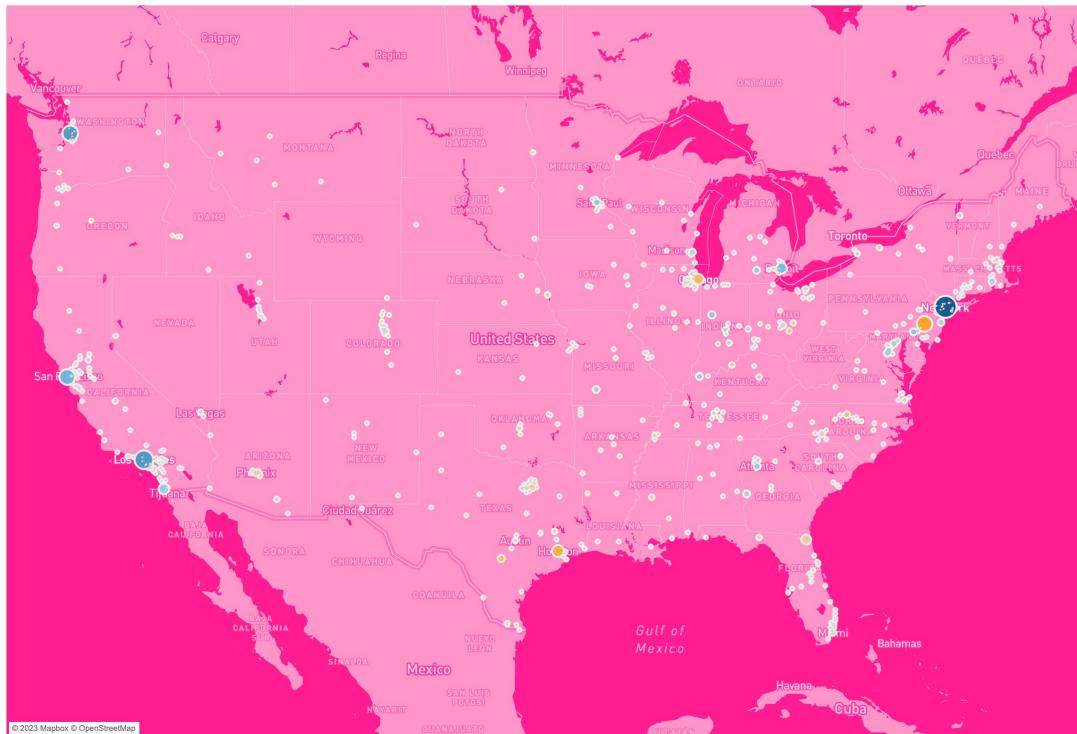
6h. Mapbox.com

- In Tableau, click **Map > Background Maps > Add Mapbox Map...**.
 - Then enter a style name and paste in the Integration URL, OK.





6h. Mapbox.com



Done!

135



Mapbox.com

The screenshot shows the Mapbox website homepage. At the top, there's a green header bar with the Mapbox logo and a link to "Learn how Boston Children's Hospital uses Mapbox to create Vaccines.gov". Below the header, the main navigation menu includes "Products", "Solutions", "Developers", "Company", "Pricing", and "Blog", along with a "Go to account" button. The main title "Maps and location for developers" is prominently displayed in large white font. A subtitle below it reads "Precise location data and powerful developer tools to change the way we navigate the world." A blue "Start mapping for free" button is located below the subtitle. The bottom section features several images of Mapbox products: a car dashboard with a navigation screen, a smartphone displaying a map, and another smartphone showing a 3D terrain view.

136



Video: MapBox



Presentria.com

Language
English ▾

 Presentria

Session Number (Required) _____

Student ID (Optional) _____

Student Name (Optional) _____

Remember my student ID and student name

JOIN

Tableau software & Class attendance

Thank You

