

Collect, Report, Analyze, and Use High-Quality Part B Data



Outlier Analyses Step-by-Step Guide

Authors: Danielle Crain, Chris Lysy



January 2020



IDEA DATA CENTER

The IDEA Data Center (IDC) created this publication under U.S. Department of Education, Office of Special Education Programs grant number H₃₇₃Y190001. Richelle Davis and Rebecca Smith serve as the project officers.

The views expressed herein do not necessarily represent the positions or policies of the U.S. Department of Education. No official endorsement by the U.S. Department of Education of any product, commodity, service, or enterprise mentioned in this publication is intended or should be inferred. This product is in the public domain. Authorization to reproduce it in whole or in part is granted.

For more information about IDC's work and its partners, see www.ideadata.org.

Suggested Citation:

Crain, D., and Lysy, C. (2020, January). *Outlier Analyses: Step-by-Step Guide*. IDEA Data Center. Rockville, MD: Westat.



U.S. Office of Special Education Programs



Outlier Analyses: Step-by-Step Guide

Purpose and Intended Audience

Outlier analysis provides an important tool for examining data to identify observations [among local education agencies (LEAs), schools, students] with data that deviate from an established norm so that states can investigate the observations as possible data errors. This guide introduces the principles of outlier analysis and includes six tutorials on completing an outlier analysis. It is a companion for the IDEA Data Quality: Outlier Analyses Tool, an Excel-based tool states can use to identify outliers using the interquartile range approach described in the step-by-step tutorials. IDC designed these two products—the outlier analyses tool and this step-by-step guide-for state personnel responsible for the IDEA 618 and/or 616 data to use. IDEA Part B state staff working with LEAs also can use the products to analyze their local data. Any state staff with the ability to examine and analyze IDEA 618 and/or 616 data also can benefit from these technical assistance (TA) products. Such staff include data managers, Information Technology (IT) personnel, coordinators, and directors.

This guide provides an overview of outlier analysis. It is organized around four questions:

- 1. What is an outlier?
- 2. Why is outlier analysis important for data validity and reliability?
- 3. What action should states take after conducting an outlier analysis?
- 4. How can states conduct and display an outlier analysis?

Question 1: What Is an Outlier?

"An outlier is an observation which deviates so much from the other observations as to arouse suspicions that it was generated by a different mechanism" (Hawkins, 1980). Outlier analysis also includes investigating whether the data are valid or invalid. In the fields of statistics and data mining, outliers may be referred to as anomalies, abnormalities, deviants, and discordant data. When conducting an outlier analysis, states define what value or combination of values are outside the expected norm. These parameters can help specify the LEAs that have data outside of the "normal" parameters the state sets. There may be times when states see differences in LEAs IDC's principles for high-quality data include addressing data quality at all stages—collecting, submitting, analyzing, reporting, and using.

At the most fundamental level, highquality data are timely, accurate, and complete.

- *Timely* data are current per a specific period of time.
- Accurate data are
 - Reliable, that is, consistent across time, methods, and locations; and
 - Valid, that is, representative of what they are designed to measure.
- **Complete** data represent the intended population (e.g., national, state, or local level) and relevant subgroups (e.g., race/ethnicity, grade level, socioeconomic level, gender).

Beyond these fundamental components, high-quality data are also *accessible, usable*, and *secure*.

- Accessible data are readily available in formats that are understandable, user friendly, and practical.
- Usable data promote sound management, strong governance, and dedication to improving results for children and youth with disabilities and their families.
- Secure data are collected and stored with due consideration to maintaining confidentiality and with electronic and physical protections commensurate with the sensitivity of the data.

that the state considers normal. It is up to the states to determine what constitutes a "sufficient" anomaly.

Outlier analysis may identify valid as well as invalid data. Invalid outliers are the target of outlier analysis, as they represent errors in the data. On the other hand, valid outliers may appear to be outside the norm, but investigation demonstrates that the data are not in error. Valid outliers may occur due to random variation, which occurs due to chance and is inherent in a system.

Question 2: Why Is Outlier Analysis Important for Data Validity and Reliability?

Outlier analysis is primarily important because it helps to identify errors in the data, which, when investigated, may reveal systematic errors in data collection, coding, or entry. **Invalid outliers should be corrected, and the processes that resulted in such errors should be fixed**.

Outlier analysis also can be important because it may identify LEAs that are performing better or worse than the norm. Identifying these high or low performers provides opportunities for understanding the factors behind high performance or providing targeted TA where it is needed.

Question 3: What Action Should States Take After Conducting an Outlier Analysis?

After conducting an outlier analysis, states should investigate any identified outliers to understand why the data are so different from the norm. If the data are outside the parameters states set for valid outliers, then states should follow up with the LEAs to determine the root cause of the outlying data. Questions to focus outlier investigations¹ follow.

1. Are the outliers found in just one LEA?

Knowing that all outliers are found in just one LEA can help state personnel focus their investigation into the cause of outliers.

2. Are the same LEAs identified with outliers in more than one data submission

State personnel may want to review LEAs that have outlier data in more than one data submission. The outliers may indicate a need for the LEA to review the data entry or coding policies. They also may indicate that the LEA lacks understanding of the data that the data submissions require.

3. Are multiple outliers commonly identified in the same LEAs?

If an outlier analysis reveals multiple outliers commonly identified in the same LEAs, state personnel may want to review the similarities in demographics or data collection practices in these LEAs or both.

¹ The Office of Elementary and Secondary Education (OESE) compiled a list of possible causes of data quality problems related to the *Elementary and Secondary Education Act* (ESEA) and other data reporting. States can review the list to help determine areas of data quality that the outlier analysis affects (U.S. Department of Education 2006).

4. Are the LEAs with outliers using non-standard data collection definitions?

State personnel may want to review the definitions LEAs use to ensure that, within the state context, the LEAs understand and use the definitions the Office of Special Education Programs (OSEP) provides for the IDEA 618 data collections. For example, outliers in the discipline data could be due to an LEA's interpretation of the terms "suspension" and "expulsion."

5. Are the LEAs with outliers using non-standard methods for aggregating the data?

States that collect aggregated data from LEAs may want to review the methods LEAs use to aggregate studentlevel data to create totals. Inconsistencies in how LEAs aggregate data could lead to outliers. For example, the state education agency (SEA) may want to review the methods LEAs use to aggregate the number of students with disabilities by race/ethnicity to ensure that the LEAs appropriately count each child in the seven categories OSEP requires.

6. Are the LEAs with outliers using non-standard methods to collect the data?

State personnel may want to review whether LEAs are using similar, standard policies and procedures for collecting the data. For example, this can include ensuring that all LEAs use the U.S. Department of Education's race/ethnicity guidelines. It also could include ensuring all locals define suspension and expulsion consistently.

7. Did the small *n*-size affect the analysis?

States can analyze *n* size in two different ways. The *n* size can skew the results of the analysis. If an LEA has a small population, then it may lead to outliers because of the proportion of the population to the rest of the state's LEAs. To investigate if this scenario has occurred, users would disaggregate the data and determine if they are sensible based on the local small populations. The second scenario can lead to outliers when reporting IDEA data. States can review the *n* size used to calculate the Annual Performance Report (APR) for Part B Indicator B4 around discipline and significant discrepancy to determine if outliers have skewed the results of the analyses.

For more information on examining root cause, states can review <u>Equity, Inclusion, and Opportunity: Addressing</u> <u>Success Gaps, White Paper.</u>

Question 4: How Can States Conduct and Display an Outlier Analysis?

States can use several possible approaches to conduct an outlier analysis. The <u>IDEA Data Ouality: Outlier Analyses</u> <u>Tool</u> provides assistance with automatic calculation of an outlier analysis. State staff who want to calculate their outlier analysis using their databases and/or programs can use this step-by-step guide instead.

There is no single right way to do an outlier analysis. Staff need to choose an approach and be systematic.

This guide provides six different tutorials covering different methods states can use to identify and visualize outliers. Five of the six tutorials explain approaches using Microsoft Excel. The sixth approach uses Tableau to visualize outliers. Staff should pick the method, or methods, that make the most sense to them.

Identifying an Outlier

Identifying an outlier starts with a simple question, "What Is Normal?" The six tutorials in this guide can help state staff define what is normal and use data visualization to make identified outliers more noticeable.

Outliers are the numbers outside of a range of data considered normal. The following three tutorials present different ways of identifying what is normal:

- Tutorial 1: Systematically Determining What Is Normal Using the Interquartile Range
- Tutorial 2: Qualitatively Defining a Normal Range
- Tutorial 3: Simply Sorting

Data visualization can help support outlier analyses by making identified outliers more noticeable. The following three tutorials present visual approaches that pair well with any of the first three approaches to calculating outlier analysis.

- Tutorial 4: Heat Maps in Excel
- Tutorial 5: Dot Plots in Excel
- Tutorial 6: Dot Plots in Tableau

A Range of Normal

As stated previously, outliers are the numbers outside of a range of data that state staff identified as normal. Here are a few ways to define a normal range.

(Note: Within this section, the guide refers to the edges of the normal range as Upper and Lower Fences.)

Tutorial 1: Systematically Determining What Is Normal Using the Interquartile Range

Tutorial 1 is the typical statistical approach to defining what is normal. This approach uses an easy-to- calculate "interquartile range" to identify a normal range for a provided series of data. This series could be data from across all LEAs in a state for a single measure. Most likely, state staff will analyze a distribution within a single column in Excel.

Step 1.

To start, staff will need an Excel workbook with at least two columns of data. The following example uses districtlevel Indicator B5 data.

X 🚽 🤊	• (> • -	ALC: NO.	Outlier-Analysis-up	lated.xlsx - Microsoft Excel	
File	Home Ins	ert Page Layout	Formulas Data Review View Tableau eDocPrint	er JMP Acrobat	
	iut iopy +	Arial	· 10 · A [^] A [×] ≡ ≡ ≈ ≫·· ≣ ¹ Wrap Text	rcentage	Σ AutoSum * Fill *
v 💞 F	ormat Painter	B I U - H	🗄 * 🤷 * 📥 🛎 🚔 🚝 🎏 🔛 Merge & Center * 🖇	Formatting * as Table * Styles *	Clear * Filter * Select *
Clipbo	ard 🖓	Font	t 🕼 Alignment 🖓	Number 🖙 Styles Cells	Editing
D9	41	• 🤄 🕺 🕇	100%		~
A	В	С	D	E	F
1 District	Enrollment	SWD Enrollment	5A - SWD Served in the Regular Class 80% or More of the Day	5B - SWD Served in the Regular Class Less Than 40% of the Day	5C - SWD Served in Separate Facilities
2 000001	329	35	87.50%	0.00%	0.00%
3 000002	90	13	92.86%	0.00%	0.00%
4 000003	214	30	93.75%	0.00%	0.00%
5 000004	77	0	NR	NR	NR
6 000005	156	32	100.00%	0.00%	0.00%
7 000006	117	55	85.45%	14.55%	0.00%
8 000007	795	113	82.35%	7.06%	3.53%
9 000008	3184	700	87.50%	6.41%	3.97%
10 000009	1073	135	68.38%	10.26%	5.13%
11 000010	17254	4011	43.74%	13.82%	5.65%
12 000011	305	99	95.92%	0.00%	0.00%
13 000012	101	21	93.33%	0.00%	0.00%
14 000013	1290	296	58.76%	5.47%	2.19%
15 000014	991	113	77.32%	11.34%	8.25%
16 000015	317	38	90.91%	0.00%	0.00%
17 000016	2237	473	68.14%	14.77%	4.38%
18 000017	1754	204	73.10%	19.29%	0.00%
19 000018	1460	200	63.33%	10.67%	5.33%
20 000019	3134	545	52.97%	10.27%	5.86%
21 000020	1085	146	53.49%	10.08%	1.55%
22 000021	592	108	50.67%	17 33%	n nn%
	Outlier Analy	sis Step 1 Outl	iier Analysis Step 2 🖉 Outlier Analysis Step 3 🖉 Outlier Analysis S	tep 4 / Outlier Analysid	
Ready 🔛					

Step 2.

The first calculations are for the first and third quartiles, using the following formulas in Excel: "=PERCENTILE(<Cell Range>,0.25)" AND "=PERCENTILE(<Cell Range>,0.75)"; staff should just replace <Cell Range> with the range of values they are checking.

For this example, the formula for the Quartile 1 calculation would be "=PERCENTILE(D2:D949,0.25)," and the Quartile 3 calculation would be "=PERCENTILE(D2:D949,0.75)."

X 🚽 🤊 -	(" + =	-	Outlier-Analysis-upo	lated.xlsx - Microsoft Excel	
File	lome Inse	ert Page Layout	Formulas Data Review View Tableau eDocPrint	er JMP Acrobat	X 🖬 🗕 🕥 A
Paste Clipboa	ut opy * ormat Painter urd ਯ	Arial ••••••••••••••••••••••••••••••••••••	· <u>」</u> · <u>A</u> · A [·] = = ● ≫· 雷Wap Text Gr · <u>③</u> · <u>A</u> · E = 译译 团Merge & Center · S Alignment 5	neral - %,) * 80 + 93 Number Number Conditional Format Cell Formatting as Table ~ Styles Styles Cells Cells Cells Cells	Σ AutoSum * Fill * Soft & Find & Clear * Filter * Select * Editing
D95	2 .	• (* f _x =	PERCENTILE(D2:D949,0.75)		*
A 1 District 934 (000933 935 (000934 936 (000936 938 (000936 938 (000936 938 (000936 940 (000938 941 (000940 942 (000941 943 (000942 944 (000943 945 (000946 946 (000945 947 (000946 948 (000947)	B Enrollment 3092 8059 949 1752 3664 611 55 154 4089 302 192 1331 2566 88 353	C SWD Enrollment 621 1168 170 151 638 111 8 8 37 1022 29 49 197 746 30 10	D 5A - SWD Served in the Regular Class 80% or More of the Day 64 34% 67 12% 54 55% 86 23% 60 95% 72 73% NR 100.00% 39 65% 60 94% 44 87% 100.00% NR	E 5B - SWD Served in the Regular Class Less Than 40% of the Day 10.04% 7.01% 5.30% 0.72% 12.26% 1.14% NR 0.00% 28.43% 0.00% 0.00% 11.17% 7.19% 0.00% NR	F ▲ 5C - SWD Served in Separate Facilities 4.06% 2.76% 2.27% 4.14% 9.14% 9.14% NR 0.00% 1.08% 0.00% 0.00% 1.08% 0.00% 0.00% 3.72% 4.65% 0.00% NR NR
949 000948 950 951	250	6 Ouartile 1	NR 0.61953	NR	NR
952 953		Quartile 3	0.862		
Ready	Outlier Analy	sis Step 1 🔍 Outlie	er Analysis Step 2 🖉 Outlier Analysis Step 3 🏑 Outlier Analysis S	tep 4 / Outlier Analyski 4 III Average: 0.7408125 Count: 4 Sum: 1.481625	

Step 3.

Next, staff should calculate the interquartile range, which is simply Quartile 3-Quartile 1. In this example, the formula would be "=D952-D951."

X 🔒 🤊 -	(≅ - ∓	-	Outlier-Analysis-	updated.>	dsx - Microsoft E	xcel		
File	Home Ins	ert Page Layout F	Formulas Data Review View Tableau eDocF	Printer	JMP Acroba	t		X 🖬 🗕 🕥 A
	ut opy -	Arial - 10	0 · A A = = → Wrap Text	General	•			Σ AutoSum - A
Paste	ormat Painter	B <i>I</i> <u>U</u> - ⊞ -	🍐 • 🛕 • 📑 喜 盲 律 律 🖼 Merge & Center •	\$ - %	° • • 00. 00 • • 00. • 0	Conditional Format Cell Formatting * as Table * Styles *	Insert Delete Format	Sort & Find &
Clipbo	ard 🗔	Font	🕞 Alignment 🕞	Nu	umber 🗔	Styles	Cells	Editing
D95	53		52-D951					~
A	В	С	D			E		F 🔺
1 District	Enrollment	SWD Enrollment	5A - SWD Served in the Regular Class 80% or More of the	e Day	5B - SWD Serv	ed in the Regular Class Less	Than 40% of the Day	5C - SWD Served in Separate Facilities
937 000936	1752	151	86.23%			0.72%		4.14%
938 000937	3664	638	60.95%			12.26%		9.14%
939 000938	611	111	72.73%			1.14%	3.41%	
940 000939	55	8	NR			NR	NR	
941 000940	154	37	100.00%			0.00%		0.00%
942 000941	4089	1022	39.65%		28.43%			11.08%
943 000942	302	29	100.00%			0.00%	0.00%	
944 000943	192	49	100.00%		0.00%			0.00%
945 000944	1331	197	60.64%			11.70%	3.72%	
946 000945	2566	746	44.87%			7.19%	4.65%	
947 000946	68	30	100.00%			0.00%		0.00%
948 000947	353	10	NR			NR		NR
949 000948	250	6	NR			NR		NR
950								
951		Quartile 1		0.619525				
952		Quartile 3		0.8621				
953		Interguartile Range		0 242575				
954		interguardie i talige		0.2-2010				
955								
956								
	Outling Apple	icis Step 1 Outline A	Applysis Chap 2 Outline Applysis Chap 2 Outline Applys	in Chan A	Outline Ann	headled at		Y
	Outlier Analy	Sis Step 1 / Outlier A	Millingsis Step 2 Outlier Analysis Step 3 Outlier Analys	is step 4		iysqi] •		
кеаду 🔛								

Step 4.

Using the interquartile range (IQR), staff should come up with a "normal range" by setting up fences. The Lower Fence is the bottom of the range, and the Upper Fence is the top.

Calculate the Lower Fence by subtracting 1.5 times the interquartile range from Quartile 1. [Lower Fence = Quartile 1 - (1.5 * IQR)]

Calculate the Upper Fence by adding 1.5 times the interquartile range to Quartile 3. [Upper Fence = Quartile 3 + (1.5 * IQR)]

In this example, the formula for the Lower Fence would be "=D951-(1.5*D953)," and the formula for the Upper Fence would be "=D952+(1.5*D953)."

[Note: This is subjective. To create a larger range (fewer outliers), staff should multiply by a larger number than 1.5 (maybe 2 or 3). To create a smaller range, staff should not use a multiplier at all. If staff do not use a multiplier, the definition of the percentile indicates that staff will identify about half of their data points.]

X 🚽 🤊 - (*	* -	-		-	Outlier-Analysis	-updated.)	klsx - Microsoft E	xcel				x
File Home	e Insert F	Page Layout F	Formulas Data Re	iew View	Tableau eDocl	Printer	JMP Acroba	t			ه 🕜 ه	- 🗗 🔀
Paste Clipboard	Arial B 2	- 10 ℤ Ψ - 💽 - Font		€ ≫ - i≢ i≢ Alignmen	Wrap Text Merge & Center + t 5	General \$ - %		Conditional Format (Formatting + as Table + Sty Styles	Cell Insert Delete Form Cells	Σ AutoSum at 2 Fill ▼ 2 Clear ▼ Ec	Sort & Find & Filter * Select *	
D954	• (*	<i>f</i> _x =D95	1-(1.5*D953)									~
A 1 District Entro 937 (000936 938 (000937 939 (000938 940 (000939 941 (000940 942 (000941 943 (000942 945 (000944 945 (000944 946 (000947 948 (000947)	B SWD 1 1752 3664 611 55 154 4089 302 192 1331 22666 68 353 553 553 553 553 553 553 553 553 55	C Enrollment 151 638 111 8 37 1022 29 49 197 746 30 10	5A - SWD Served in th	D Regular Clas 86.23% 60.95% 72.73% NR 100.00% 100.00% 60.64% 44.87% 100.00% NR	s 80% or More of the	e Day	5B - SWD Sen	E ed in the Regular Class I 0.72% 12.26% 1.14% NR 0.00% 28.43% 0.00% 0.00% 11.70% 7.19% 0.00% NR	∟ess Than 40% of the Da	/ 5C - SWD Se	F vved in Separate F 4.14% 9.14% 3.41% NR 0.00% 11.08% 0.00% 0.00% 3.72% 4.65% 0.00% NR	acilities
949 000948 950 951	250 Quart	6 tile 1		NR		0.619525		NR			NR	
952	Quart	tile 3				0.8621						
953	Interg	uartile Range			0	0.242575						
955	Lower Fence 0.2550625				•							
956 I	tlier Analysis Step	1 / Outlier A	nalysis Step 2 🖉 Outlie	r Analysis Step	3 Outlier Analys	sis Step	4 🖉 Outlier Ana	ysi[] 4	Ш			•
Ready 🔚										100%		+ .::

Step 5.

Once the Lower and Upper Fences are known, state staff can identify outliers. An outlier would be any number falling below the Lower Fence or above the Upper Fence.

In this example, that would be any number less than 25.6 percent and any number over 122.6 percent. Staff would consider the value of 15.38 percent an outlier because it falls below 25.6 percent.

(Note: With data that are widely dispersed, it is likely that the Lower Fence could be a negative number or the Upper Fence could fall outside of the possible range of values. Staff could interpret a negative Lower Fence as having no Lower Fence with no low outliers.)

X	🚽 🔊 •	(≅ - ∓	-	Outlier-Analysis-	updated.xlsx - Microsoft	Excel	
F	ile H	ome Inse	rt Page Layout F	ormulas Data Review View Tableau eDocP	rinter JMP Acrob	at	a 🕜 🗆 🗟
		it ipy *	Arial • 10	· A A = = = the the two of two	Percentage	Conditional Format Call Invest Dalate Format	∑ AutoSum * Ar
1.	Fo 🗸 Fo	rmat Painter	B 1 0 * .	· · · · · · · · · · · · · · · · · · ·	\$ ° % • .00 →.0	Formatting * as Table * Styles * * *	
	Clipboa	rd G	Font	🕞 Alignment 🕞	Number 5	Styles Cells	Editing
	D68	7 -	fx 15.38	1%			~
- 4	A	В	С	D		E	F E
678	000677	1070	241	43.22%		12.56%	4.02%
679	000678	609	60	90.38%		0.00%	3.85%
680	000679	2870	531	81.26%		5.13%	2.97%
681	000680	5111	810	50.57%		7.06%	7.49%
682	000681	92	16	NR		NR	NR
683	000682	36	5	NR		NR	NR
684	000683	1205	109	64.86%		2.70%	0.00%
685	000684	713	106	79.55%		15.91%	4.55%
686	000685	1909	196	69.06%		11.51%	2.94%
687	000686	76	17	15.38%		0.00%	0.00%
688	000687	1367	338	67.91%		1.25%	10.77%
689	000688	1304	472	63.85%		18.18%	7.58%
690	000689	1128	188	70.83%		11.90%	1,19%
691	000690	4334	777	74.53%		9.70%	4.50%
692	000691	298	87	84.72%		9.72%	0.00%
693	000692	13	0	NR		NR	NR
694	000693	55	8	NR		NR	NR
695	000694	1084	217	70.41%		14.29%	8.16%
696	000695	618	49	100.00%		0.00%	0.00%
697	000696	2	0	NR		NR	NR
698	000697	2099	495	70.33%		12.66%	6.63%
699	892000	1371	286	69 16%	3 20%		
14 4		Outlier Anal	lysis Step 2 Outlier	Analysis Step 3 Outlier Analysis Step 4 Outlier Analysis	ysis Step 5 🖉		
Rea	ady 🛅						

Tutorial 2: Qualitatively Defining a Normal Range

Another way to identify outliers would be to use qualitative information. State staff can identify the numbers that seem out of place given what they know about their state's data, or they can ask colleagues about their own expectations based on experience. Staff should determine a range that seems to make sense, then look at the data.

Turning to the last example, let's say that after speaking with colleagues, staff determined that they should look further at districts if the districts had any less than 50 percent for Indicator B5A. Staff should just go ahead and identify those cases.

🔣 🛃 🤊 • (🍽 • =	-	Outlier-Analysis-upd	lated.xlsx - Microsoft Excel	X
File Home Inser	rt Page Layout	Formulas Data Review View Tableau eDocPrint	er JMP Acrobat	a 🕜 🗖 🖉 🗙
Paste Stormat Painter	Arial •	· 10 · A · A · = = → · □· Wrap Text Per · ③ · ▲ · = = □ · 章 章 望 图 Merge & Center · S	rcentage	► AutoSum * Fill * Sort & Find & Clear * Filter * Select *
D11 -	(<i>f_x</i> 4	3.74%	Number is Styles cens	V
A B 1 District Enrollment 2 000001 329 3 000002 90 4 000003 214 5 000004 77 6 000005 156 7 000006 117 8 000007 795 9 000008 3184 10 000009 1073 11 000010 1254 12 000011 317 15 000014 991 16 000015 317 17 000016 2237 18 000017 1754	C SWD Errollment 35 13 0 0 32 55 113 700 135 135 135 135 135 135 135 135 135 135	D 5A - SWD Served in the Regular Class 80% or More of the Day 87.50% 92.86% 93.75% NR 100.00% 85.45% 82.35% 87.50% 68.38% 43.74% 95.92% 93.33% 58.76% 77.32% 90.91% 68.14% 77.10%	E 5B - SWD Served in the Regular Class Less Than 40% of the Day 0.00% 0.00% 0.00% NR 0.00% 14.55% 7.06% 6.41% 10.26% 13.82% 0.00% 5.47% 11.34% 0.00% 14.7% 11.34%	F ⊂ 5C - SWD Served in Separate Facilities 0.00% 0.00% 0.00% NR 0.00% 0.00% 0.00% 5.13% 3.53% 3.97% 5.13% 5.65% 0.00% 0.00% 2.19% 8.25% 0.00% 4.38% 0.00%
20 000019 3134 21 000020 1085	545	52.97% 52.90%	10.67%	5.86%
21 000020 1085 22 000021 592 II ↓ → H Outlier Analys Ready 🔚	108 sis Step 1 Outli	5.49% 50.67% er Analysis Step 2 / Outlier Analysis Step 3 / Outlier Analysis St	17 33%	

Tutorial 3: Simply Sorting

Sometimes the easiest way to do an analysis would be simply to sort the data. The goal of an outlier analysis is not to support statistical analyses but, rather, to identify potential data issues. There is no harm in singling out data that staff would usually consider part of a normal range.

Step 1.

Staff should highlight their range of data.

🗶 🛃 🍤 -	(* - =			Outlier-Analysis-update	ed.xlsx - Microsoft Excel	training the second second	1	×
File	lome In	isert Page Layout	t Formulas Data Review View	Tableau eDocPrinter	JMP Acrobat			a 🕜 🗖 🛱 🗙
From From Access Web	From Fr Text S Get Exter	rom Other Sources * Connecti nal Data	g g Refresh → Edit Links Connections 2↓ 2↓ X ↓ 2↓ X ↓ 2↓ X ↓ 5 ort Fi Connections Sort	ter Advanced Stiller	to Remove Data Consolidate V Data Tools	What-If nalysis • Group Ungroup Subt	●별 Show Detail =별 Hide Detail otal	
A1		• (* fx	District					*
A	В	С	D		E		F	
1 District	Enrollment	SWD Enrollment	t 5A - SWD Served in the Regular Class 80%	or More of the Day	5B - SWD Served in the Regular Class	Less Than 40% of the Day	5C - SWD Served in Separa	te Facilities
2 000001	329	35	87.50%		0.00%		0.00%	
4 000002	214	30	92.00%		0.00%		0.00%	
5 000004	77	0	NR		NR		NR	
6 000005	156	32	100.00%		0.00%		0.00%	
7 000006	117	55	85.45%		14.55%		0.00%	
8 000007	795	113	82.35%		7.06%		3.53%	
9 000008	3184	700	87.50%		6.41%		3.97%	
10 000009	1073	135	68.38%		10.26%		5.13%	
11 000010	17254	4011	43.74%		13.82%		5.65%	
12 000011	305	99	95.92%		0.00%		0.00%	
13 000012	101	21	93.33%		0.00%		0.00%	
14 000013	1290	296	58.76%		5.47%		2.19%	
15 000014	991	113	17.32%		11.34%		8.25%	
17 000015	317	38	90.91%		0.00%		0.00%	
18 000017	1754	475	00.14% 73.10%		14.7770		4.30%	
19 000018	1460	204	63 33%		10.67%		5 33%	
20 000019	3134	545	52 97%		10.07%		5.86%	
21 000020	1085	146	53 49%		10.08%		1.55%	
22 000021	592	108	50 67%		17 33%		0.00%	
	utlier Ana	lysis Step 1 / Ou	Itlier Analysis Step 2 Outlier Analysis Step 3	Outlier Analysis Step	4 Outlier Analysia 4			► <u> </u>
Ready 🔚					Average: 370.6913398	Count: 5694 Sum: 1685904.213	■□ □ 100% -	· · · · ;;

Step 2.

Staff should sort by the measure.

🔀 🛃 🍯 = (🖻 = =	Outlier-Analysis-updated.xlsx - Microsoft Excel	
File Home Insert Pag	age Layout Formulas Data Review View Tableau eDocPrinter JMP Acrobat	a 🕜 🗖 🛪
From From From Other Access Web Text Sources* Get External Data	Existing Connections Connections 1 Image: Connections Image: Connections	e Show Detail =∃ Hide Detail e G
A2 🔻 🖱	<i>f</i> _x 000001	~
A B I 1 District Enrollment SWD En 2 000001 329 3 3 000002 90 1 4 000003 214 3 5 000004 77	C D E inrollment 5A - SWD Served in the Regular Class 80% or More of the Day 5B - SWD Served in the Regular Class Less Than 40% of the Day 35 87.50% 0.00% 13 92.86% 0.00% 30 93.75% 0.00% 0 ND ND	F 5C - SWD Served in Separate Facilities 0.00% 0.00% 0.00%
6 000005 156 3 7 000006 117 5	55 Sort 10%	0.00%
9 000007 795 1 9 000008 3184 7 10 000009 1073 1:	113 700 Aladd Level ★ Delete Level ★ Options ♥ My data has headers 11% 11% 26%	3.53% 3.97% 5.13%
11 000010 17254 40 12 000011 305 9 13 000012 101 305	1011 Sort by SA - SWD Served in the Values Smallest to Largest 10%	5.65% 0.00%
13 000012 101 2 14 000013 1290 2' 15 000014 991 1	296 17% 113 34%	2.19% 8.25%
16 000015 317 3 17 000016 2237 4 18 000017 1754 2	38 00% 473 77% 204 02%	0.00% 4.38% 0.00%
19 000018 1460 20 20 000019 3134 5 21 000020 1025 1	200 545 ОК Салсе 27%	5.33% 5.86%
21 000020 1085 1 22 000021 592 1 14 4 → → Outlier Analysis Step 1	146 102 112 112 112 112 112 112 112	1.55%
Ready 🔛	Average: 370.6913398 Count: 5694 Sum: 1685904.213	

Step 3.

Staff should look more deeply into the values at the top and bottom of the sort.

X 🚽 🤊	• (° + -				No. of Concession, Name	Outlier-Analysis-	updated.	xlsx - Microsoft Exce	el .			100		
File	Home	Insert Page	Layout	Formulas Data	Review View	Tableau eDocF	rinter	JMP Acrobat						a 🕜 🗕 🖶 🗙
From Fro Access W	om From eb Text Get Ext	From Other Sources * Co ernal Data	Existing	Refresh All v Connections	s 2 J 2 Z Z J Sort Fill Sort 8	Reapply Advanced	Text to Column	Remove Data s Duplicates Validati Data	Consolidat	e What-If Analysis •	Group Ungroup	Subtot	부클 Show Detail 프릴 Hide Detail tal	
	41	- (n	f _x Dis	trict										*
A	В	C			D				E			-	F	
1 Distric 2 00003	t Enrollme 3 0	nt SWD Enro 111	ollment 5/ 1	A - SWD Served in the	Regular Class 80% 0.00%	or More of the Da	y 5B	- SWD Served in t	he Regular Cla 0.009	iss Less Tha %	an 40% of the D	ay 50	C - SWD Served in Sepa 100.00%	arate Facilities
3 00008	7 1	145	5		0.00%				0.009	6			16.36%	
5 00064	1 0	36			0.00%				0.009	% %			100.00%	
6 00082	3 0	51			0.00%				100.00)%			0.00%	
7 00083	7 62	56			0.00%				5.269	%			0.00%	
8 00068	6 76	17			15.38%				0.009	6			0.00%	
9 00081	5 137	14			16.67%				0.009	6			0.00%	
10 000549	578	50			18.18%				11.36	%			0.00%	
11 00023	1 1920	544	4		21.24%				24.82	%			15.99%	
12 00010	2 996	189	9		24.76%				33.50	%			12.14%	
13 000447	7 342	52			25.00%				16.67	%			6.25%	
14 00031	7 187	32			25.81%				16.13	%			0.00%	
15 00042	399	40			25.81%				3.239	%			0.00%	
16 000834	16868	446	8		26.32%				29.55	%			5.64%	
17 000153	3 28893	907:	3		28.02%				35.17	%			2.47%	
18 00064	7 3329	511	1		30.57%				21.76	%			4.92%	
19 000573	3 962	239	9		30.87%				4.359	16			7.83%	
20 000470) 135	44			33.33%				19.44	%			2.78%	
21 000884	1081	351	1		33.65%				26.98	%			6.67%	
22 00007	/85	9.1			33 71%	1.0.11		/ a 18 . 1	11 24	0/_			0.00%	
	Outlier Ar	alysis Step 1	Outlier	Analysis Step 2 Ou	clier Analysis Step 3	<u>Outlier</u> Analys	s Step 4	Outlier Analys	{I] 4					
Ready								Avera	ige: 370.6913398	Count: 569	94 Sum: 168590	4.213		• •

Data Visualization Support for Outlier Analyses

Data visualization can help support outlier analyses by making identified outliers more noticeable. Here are a few ways state staff can use data visualization to support their outlier analyses.

Tutorial 4: Heat Maps in Excel

By using conditional formatting in conjunction with the approaches shown previously, staff can make outliers stand out. This can be particularly important when there are multiple columns of data or many rows.

Step 1.

Staff should highlight the range of data and click on the *Conditional Formatting* button in the *Home* tab.

X	😰 🔄 🔊 Y 🔍 🖓 🗮 🕐 V 🕄 🚽 Outlier-Analysis-updated.xix - Microsoft Excel								
F	ile H	lome Inse	ert Page Layout I	Formulas Data Review View Tableau eDocPrin	r JMP Acrobat	∧ 😗 🗆 🗗 ×			
ſ	× 0	ut	Arial * 10	A A → = = ≫ · = Wrap Text Pe	centage 🔹 📑 📑 📑	Σ AutoSum - A			
Pa	Paste Format Painter B I U - O -			🕭 * 🗛 * 📰 🚍 津 津 🖾 Merge & Center * 💲	% , 00 00 Conditional Format Cell Insert Delete Format Formatting v as Table v Styles v v v v	Sort & Find &			
	Clipboa	ird 5	Font	Alignment 5	Number is Styles Cells	Editing			
	D94	.9 .	fx NR			*			
1	A	В	С	D	F	F E			
1	District	Enrollment	SWD Enrollment	5A - SWD Served in the Regular Class 80% or More of the D	v 5B - SWD Served in the Regular Class Less Than 40% of the Day	5C - SWD Served in Separate Facilities			
2	000001	329	35	87.50%	0.00%	0.00%			
3	000002	90	13	92.86%	0.00%	0.00%			
4	000003	214	30	93.75%	0.00%	0.00%			
5	000004	77	0	NR	NR	NR			
6	000005	156	32	100.00%	0.00%	0.00%			
7	000006	117	55	85.45%	14.55%	0.00%			
8	000007	795	113	82.35%	7.06%	3.53%			
9	000008	3184	700	87.50%	6.41%	3.97%			
10	000009	1073	135	68.38%	10.26%	5.13%			
11	000010	17254	4011	43.74%	13.82%	5.65%			
12	000011	305	99	95.92%	0.00%	0.00%			
13	000012	101	21	93.33%	0.00%	0.00%			
14	000013	1290	296	58.76%	5.47%	2.19%			
15	000014	991	113	77.32%	11.34%	8.25%			
16	000015	317	38	90.91%	0.00%	0.00%			
17	000016	2237	473	68.14%	14.77%	4.38%			
18	000017	1754	204	73.10%	19.29%	0.00%			
19	000018	1460	200	63.33%	10.67%	5.33%			
20	000019	3134	545	52.97%	10.27%	5.86%			
21	000020	1085	146	53.49%	10.08%	1.55%			
22	000021	592 Outline Area	108	50 67%	17 33%	0.00%			
14 4		Outlier Ana	ilysis Step 2 / Outlier	Analysis Step 3 / Outlier Analysis Step 4 / Outlier Analysis	Step 5 / U 4				
L Rea	idy 🔛				Average: 72.92% Count: 948 Sum: 64461.42%				

Step 2.

If staff determined the normal range by using the interquartile range, they should select *Highlight Cells Rules*, then select *Less Than*. In the box that asks for a number, they should enter the cell location of the Lower Fence value. They also can adjust the format (color/font/additional symbols).

🔣 🔛 🔊 • (🖻 • 🗦	Outlier-Analysis-updated.xlsx - Microsoft Excel
File Home Insert Page Layout	Formulas Data Review View Tableau eDocPrinter JMP Acrobat 🛆 🕜 🖂 🎘
↓ Cut ↓ Copy + Paste ↓ Format Painter Clipboard ↓ Format Painter	$\begin{array}{c c c c c c c c c c c c c c c c c c c $
D954 • 🗇 f N	۰ ۲
A B C 937 000936 1752 151 938 000937 3664 638 939 000938 611 111 940 000939 55 8 941 000940 154 37 942 000942 302 29 943 000943 192 49 945 000944 1331 197 946 000947 1526 746 947 000947 353 10 948 000948 250 6	D E F 86.23% 0.72% 4.14% 60.95% 12.26% 9.14% 72.73% 1.14% 3.41% NR NR NR 100.00% 0.00% 0.00% 100.00% 0.00% 0.00% 100.00% 0.00% 0.00% 100.00% 0.00% 0.00% 100.00% 60.64% 3.72% 44.87% -sto 9354 3.72% 44.87% -sto 9354 0.00% NR OK Cancel NR NR NR NR
950 Quartile 1 951 Quartile 3 953 Interquartile Rang 964 Lower Fence 955 Upper Fence 956 956 957 Uter Analysis Step 2 ∠ Out Point Image: Content Analysis Step 2 ∠ Out	0.619525 0.8621 0.242575 0.2556525 1.2259625 I 2.259625 I 2.25976 I 2.259776 I 2.25976 I 2.25976 I 2.259776 I 2.25976 I 2.259776 I 2.259776 I 2.259776 I 2.25977777777777777777777777777777777777

Staff should do the same for the Upper Fence but use *Greater Than*.

Step 3.

If staff determined the normal range using a qualitative approach, they should follow the same instructions as in Step 2. They should enter the number directly into the *LESS THAN* box or place the Lower Fence value into a cell in the worksheet and refer to that.

	Image: Second							
File Home Insert Page Layout	Formulas Data Review View Tableau eDocPrinter	JMP Acrobat	a 🕜 🗖 🛪					
B I I Paste I I I I I	0 • A [*] A [*] = = = 8 ⊗ -	age Conditional Format Cell Formating as Table Styles Cells	∑ AutoSum * Arr Arr Fill * Sort & Find & C Clear * Filter * Select * Editing					
D949 •			~					
A B C 937 [00396 1752 151 938 [00937 3664 638 939 [00938 611 111 940 [000393 55 8 941 [000940 1564 37 942 [000941 4089 1022 943 [000942 302 29 944 [000943 192 49 945 [000942 302 29 944 [000944 1331 197 945 [000946 68 30 947 [000942 326 2566 746 947 [000948 250 6	D 86 23% 72.73% NR 100.00% 100.00% 60.64% 44.87% 100.00% NR NR NR	E 0.72% 12.26% 1.14% NR 0.00% at are LESS THAN: State with Light Red Fill with Dark Red Text v OK Cancel	F □ 4.14% 9.14% 3.41% NR 0.00% 0.00% 0.00% 0.00% 0.00% 3.72% 4.65% 0.00% 0.00% NR NR NR					
950 250 0 951 Quartile 1 952 Quartile 3 953 Interquartile Range 964 Lower Fence 955 Upper Fence 956 957 It + ▶ ↓ Outler Analysis Step 2	0.619525 0.8621 0.242575 0.2556625 1.2259625 1.2259625 1.2259625	95 (9) / I (

Alternative Quick Heat Maps

A quicker approach to using heat maps in Excel is using the *Color Scales* feature found under the *Conditional Formatting* drop-down menu. This will automatically color a set of selected cells based on the range of values. Staff also can use data bars and icon sets to identify possible outliers quickly.

🗶 🛃 🍠 🗸	(™ - -		Outlier-Analysi	s-updated.xlsx - I	Microsoft Excel			
File	Home Inse	rt Page Layout I	Formulas Data Review View eDocPrinter	JMP Acr	obat			a 🕜 🗖 🐼
۵ الم الم	ut	Arial × 10	0 · A A ■ = = ≫· H Wrap Text	Percentag	ge v		🏣 🏋 🏢	Σ AutoSum - Z
Paste	ormat Painter	B <i>I</i> <u>U</u> - <u>□</u> -	💁 • A • 📰 🚍 🚎 🛱 🖼 Merge & Cer	nter - \$ - %	• • • • • • • • • •	Conditional Format Cell	Insert Delete Format	Sort & Find &
Clipbo	ard G	Font	Alianment	S Nu	imber 5	Formatting * as Table * Styles *	Cells	Editing
D94	19	fx (=				Highlight Cells Rules >		Y
A	В	С	D					F 🖻
1 District	Enrollment	SWD Enrollment	5A - SWD Served in the Regular Class 80% or Mor-	e of the Day	5B - SWD Serve	Top/Bottom Rules →	an 40% of the Day	5C - SWD Served in Separat
2 000001	329	35	87.50%					0.00%
3 000002	90	13	92.86%			Data Bars		0.00%
4 000003	214	30	93.75%					0.00%
5 000004	77	0				Color Scales		NR
6 000005	156	32	100.00%					0.00%
7 000006	117	55	85.45%					0.00%
8 000007	795	113	82.35%			icon Sets		3,53%
9 000008	3184	700	87.50%			Rew Bule	Blue - White - Rec	I Color Scale
10 000009	1073	135	68.38%			Class Bular	Displays a two o	r three color 13%
11 000010	17254	4011	43.74%				gradient in a rar	nge of cells. The .65%
12 000011	305	99	95.92%			Manage <u>R</u> ules	value in the cell.	.00%
13 000012	101	21	93.33%			0.00%		0.00%
14 000013	1290	296	58.76%			5.47%		2.19%
15 000014	991	113	77.32%			11.34%		8.25%
16 000015	317	38	90.91%			0.00%		0.00%
17 000016	2237	473	68.14%			14.77%		4.38%
18 000017	1754	204	73.10%			19.29%		0.00%
19 000018	1460	200	63.33%			10.67%		5.33%
20 000019	3134	545	52.97%			10.27%		5.86%
21 000020	1085	146	53.49%			10.08%		1.55%
22 000021	592	108	50.67%			17.33%	0.00%	
23 000022	3862	399	64.38%	1.92%				
24 000023	564	107	86.67%			2.22%		4.44%
нари	Outlier Ana	lysis Step 2 / Outlier	Analysis Step 3 🟒 Outlier Analysis Step 4 📜 Outlie	r Analysis Step	5 2 14			▶ []
Ready Scr	oll Lock 🛅				Averag	ge: 72.92% Count: 885 Sum: 64	4461.42% 🔳 🗆 🗉	100% 🗩 🗸 🕂

Tutorial 5: Dot Plots in Excel

Another approach to identifying outliers involves visualizing the data. An easy approach is using an in-cell formula that will create a simple dot plot next to the data. Once the formula creates the plot, staff simply look for data points that do not seem to fit with the others.

By using an Excel formula to create an in-cell chart, dot plots always will remain in line with the data. Excel's standard chart functions also can assist in identifying outliers but, ultimately, the standard chart functions will be disconnected from the original data.

Step 1.

This example uses Excel's REPT function to create this visual. Staff should start by selecting a cell next to the first data point. Basically, staff are instructing Excel to put in a series of repeating blanks based on the data and then add some kind of character at the end.

In this example, the formula looks like this:

=REPT ("",B2*50)&"|"

In order to get the dot at the end, staff should use the Wingdings font. The *50 multiplier creates the spacing. If staff desire less spacing, they should use a smaller multiplier.



Step 2.

Staff then just need to copy the formula down once they are happy with the look of the first cell. That is all that it takes!

X 🖬 🧉	- (* -	Ŧ	-			-		Outlier-	Analysis-u	pdated.xls	x - Microsoft I	Excel	-						-		• ×
File	Home	Insert	Page Layout	Formulas D	Data Revie	w Viev	/ Tableau eDo	ocPrinte	er JMP	Acrob	at									۵ 🕜	- # ×
B×	Cut	1	Wingdings * 1	0 * A* A*	= = =	89×-	Wrap Text	Ger	neral	Ŧ	5		Normal	Bad	*	+	*		Σ AutoSum	· 27	A
Paste	Format Pa	inter	B I <u>U</u>	ð - <u>A</u> -		信用 行用	Merge & Center	- \$	- %,	€.0 .00 •.€ 00.	Conditional	Format	Good	Neutral	Ŧ	Insert	Delete	Format	Q Clear *	Sort &	Find &
Clip	board	5	Font	G		Alignm	ent	5	Number	r G	ronnaturg	0,10010	Styles				Cells		E	diting	Juice
	F2	-	fx =RE	PT(" ".D2*50)8	8"I"																~
			0			D								-							-
1 Distric	t Enrolla	ont 9	SWD Enrollment	5A - SWD S	erved in the F	Conular Cl	ass 80% or More of t	he Da	4					_						56	A SWD
2 00000	1 32	9	35	574-0110-0	cived in the i	87 50	%	ne Da	· -									•			
3 00000	2 90		13			92.86	%												•		
4 00000	3 21	1	30			93.75	%												•		
5 00000	4 77		0						•												
6 00000	15 15	5	32			100.00	%												•		
7 00000	6 11	7	55			85.45	%			•											
8 00000	7 79	5	113			82.35	%											•			
9 00000	8 318	4	700			87.50	%											•			
10 00000	9 107	3	135			68.38	%									•					
11 00001	0 172	54	4011			43.74	%						•								
12 00001	1 30	5	99			95.92	%												•		
13 00001	2 10	1	21			93.33	%												•		
14 00001	3 129	0	296			58.76	%							•							
15 00001	4 99	1	113			77.32	%										•				
16 00001	5 31	7	38			90.91	%												•		
17 00001	6 223	7	473			68.14	%									•					
18 00001	7 175	4	204			73.10	%									•					
19 00001	8 146	0	200			63.33	%								•						
20 00001	9 313	4	545			52.97	%							•							
21 00002	108	5	146			53.49	%							•							
22 00002	Outle	2 Anahu	108 cir Stop 2 Outlin	r Analysis Ston	2 Outlin	50 67	top 4 Outlior An	alucie	Stop 5	0 1			_	•							× 11
Ready		n Aildly:		a Anayab Step	J _ Outle		Jucep - J Outlier An	ary 515	Step 5	00									100% 🗩		÷

Step 3.

It is easy to change the font or change the character and get a different look. Here is an alternative using the same exact formula with the Arial Black font in 10 point instead of Wingdings. The font change also changes the spacing, so staff should set up cells within the same column using the same font.

X 🖬 🤊	• (°" • =	and a second			0	utlier-Anal	ysis-updated.xl	ax - Microsoft	Excel	towney .					-		
File	Home Ins	ert Page Layout	Formulas Data Review	v View	r Tableau eDoo	Printer	JMP Acro	pat								۵ ()	- # ×
n 10 10 10 10 10 10 10 10 10 10 10 10 10	Cut Copy ~	Arial Black - 1		89	Wrap Text	Text		1		Normal	Bad	*		*	Σ AutoSum	Ż	A
Paste 💞	Format Painter	B <i>I</i> <u>U</u> - <u>⊞</u> -	· · <u>A</u> · ≡ ≡ ≡	ie ie	Merge & Center *	\$ - %	· · .0 .0	Formatting	as Table *	Good	Neutral	-	Insert	Delete Format	2 Clear *	Filter * S	elect *
Clipb	oard 🗔	Font	G	Alignme	ent G	Nu	umber	a la		Styles				Cells	Ec	Jiting	
6	E1 ·	• (* fx															*
A	В	С		D							E						=
1 District	t Enrollment	SWD Enrollment	5A - SWD Served in the R	egular Cla	ass 80% or More of th	e Day										5B	- SWD
2 000001	1 329	35		87.50%	%				1								
3 000002	2 90	13		92.869	%				1								
4 000003	3 214	30		93.759	%				1								
5 000004	1 77	0					1										
6 000005	5 156	32		100.00	%				1								
7 000006	5 117	55		85.45%	%				1								
8 000007	7 795	113		82.35%	%				1								
9 00000	3 3184	700		87.50%	%				1								
10 000009	9 1073	135		68.389	%			1									
11 000010	17254	4011		43.749	%			1									
12 000011	1 305	99		95.929	%				1								
13 000012	2 101	21		93.339	%				1								
14 000013	3 1290	296		58.769	%			- I									
15 000014	1 991	113		77.329	%												
16 000015	5 317	38		90.919	%				- I								
17 000016	5 2237	473		68.149	V6			1									
18 000017	7 1754	204		73.109	%			1									
HAPH	Outler Anz	alvsis Step 2 Outlie	er Analysis Step 3 Outlier	Analysis S	ten 4 Outlier Ana	lysis Ster	5 91		I 4								► ÎÎ
Ready P		anyono occupie / Oddine	Cucler	111111/06 0	June Ana	iyos otep						-			100% (-)		()
	-																

Tutorial 6: Dot Plots in Tableau

Using an interactive visualization program like Tableau can be useful if there is a large amount of data. Such a program allows the user to visualize hundreds of rows and multiple measures quickly and easily. If visualizing public data, state staff can use Tableau Public for free. For private data, staff would need at least a personal version of the Tableau desktop license.

Step 1.

Staff should open the Excel data in Tableau. The measures staff use should be under the *Measures* space in Tableau, and descriptive variables, such as district, should be under *Dimensions*. If this is not the case, staff should right-click on the variable and convert.



Step 2.

Staff should place the measure variable in the *Columns* box and the district variable in the *Rows* box. The default should be a bar graph. If satisfied with the bar chart, staff should stop here.



Step 3.

Staff should click on the Show Me tab and select the bar chart.



Step 4.

To change to a dot plot, staff should just change the *Marks* from Automatic to Circle.



Step 5.

To create a line to identify the Lower and Upper Fences, staff should just right-click on the x-axis and select *Add Reference Line, Band, or Box.* Then, in the *Value* box, select *Create New Parameter*.

Tableau - Book2			
File Data Worksheet Dashboa	rd Story Analysis Maj	p Format Server Window Help	
← → ■ ● ‱ • 교 □		🛓 🖅 🖉 - Abc 📊 - Normal 🚽 📈 - 🛱	Ilit Show Me
Data • Data • Coutlier Analysis Step 5 (Outlie • Dimensions • Acc District Abc Abc Measure Names Weasures • # 5A - SWD Served in the Regu # Enrollment # SVD Enrollment ** Number of Records # Measure Values	Pages Filters Marks Color Size Label Detail Toottip	Add Reference Line, Band, or Box Image: Band Image: Band Band Distribution Box Plot Scope Entire Table Per Pane Per Cell Line Value: SLM(SA - SWD Served in the Regular Class 80% or More of the Day) Average Label: Computation Fill Above: None Fill Above: None CK Cancel Apply 20 22 24 SA - SWD Served in the Regular Class 80% or More of the Day	64 nulls -
040		h - D	14 4 5 N A
1949 marks 949 rows by I column	SUIVICIA - SWD Served in t	ne regular Class 60 % or More of the Day): 647.620	[4 4 P P] •

Step 6.

Staff should change the name of the parameter to "Fence" and set the current value as the Lower Fence Value and click *OK*. Then, staff should click *Apply* and *OK*.

Tableau - Book2			
File Data Worksheet Dashboa	rd Story Analysis Map	Format Server Window Help	
← → 🖪 🔒 🌆 • 🖓 🖬	<mark>, • ¶, •</mark> ⊗ • 13 4	Create Parameter	III Show Me
Data ≎ Outlier Analysis Step 5 (Outlie Dimensions Abc	Pages	Name: Lower Fence Comment >> Properties	
Abc Measure Names	Filters Marks • Circle	Current value: 1255/6625 Display format: Automatic Allowable values: All Clist Range	•
Measures # 5A - SWD Served in the Regu # Enrollment # SWD Enrollment -# Number of Records # Measure Values	Color Size Asia	OK Cancel Hine: Fill Above: None • Fill Below: None • OK Cancel Apply 20 22 SA - SWD Served in the Regular Class 80% or More of the De	64 nulls 24 26 28 30 32
	Sheet 1 time to		
949 marks 949 rows by 1 column	SUM(SA - SWD Served in the	e Regular Class 80% or More of the Day): 047.820	

Step 7.

To include both the Upper and Lower Fences, staff should just follow the same steps a second time.

Staff also may opt to use the *Parameters* box that the program created on the screen to shift the Fence based on what they are trying to see.

🔀 Tableau - Book2									
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help									
← → 🖪 🔒 🖬 • 🔐 🕠	- Q - 17 H	+ 🖉 - Abc	📷 • Normal 🔹 🛏 🗶 • 🛱	III Show Me					
Data ÷	Pages	Columns	SUM(5A - SWD Served i						
Dimensions		= Rows	District						
Abc District	-		UTATION .						
Abc Measures Measures # 5A - SWD Served in the Regu # Enrollment # SWD Enrollment -# Number of Records	Marks Marks Color Size Label Detail Tooltip	District 000539 000539 000540 000542 000542 000543 000544 000545 000546 000546 000546 000546 000546 000549 000550		▲ 0.2556625 0					
# Measure Values		000552 000553 000554 000555	0.0 0.5 1.0 1.5 2.0 2.5 3.0 5A - SWD Served in the Regular Class 80% or More of the Day 5.0	-					
# Lower Fence	Sheet 1 to the								
		P 1 CL 000	M (1) D \ (77.02)						

Conclusion

This guide introduces the principles of outlier analysis and includes a handful of approaches state agency staff can use to identify and visualize outliers. There are many others. Staff should find a set of approaches that works well for their state and their data and then apply the approaches systematically.

State staff should visit <u>www.ideadata.org</u> to get in touch with their <u>IDC State Liaison</u> if they have any questions about these approaches.

References

Hawkins, D.M. (1980). Identification of Outliers. Netherlands: Springer.

U.S. Department of Education. Office of Elementary and Secondary Education. (2006, April). *Improving Data Quality for Title I Standards, Assessments, and Accountability Reporting: Guidelines for States, LEAs, and Schools.* Retrieved January 3, 2020, from <u>https://www2.ed.gov y guid > standardsassessment > nclbdataguidance</u>