



Supporting a Balanced, Modern and Ready Nation

Welcome!

Critical Skills for Analytics Professionals

MORS Certificate

Giving the Data a Voice



Agenda

Who I am

Data Preparation

Design Principles Overview

Data Visualization

Practice



Supporting a Balanced, Modern and Ready Nation

Disclaimer

The view and opinions expressed or implied in this Section are those of the Instructor and should not be construed to carry the official sanction of or positions of the Sponsors or MORS.



Supporting a Balanced, Modern and Ready Nation

Walt DeGrange

- CANA
 - Senior Director of Analytics
- Education
 - Vanderbilt University (BE-EE)
 - Naval Postgraduate School (MS-OR)
- 21 years Navy Supply Corps Officer
- Teaching
 - Naval Postgraduate School Faculty
 - University of Arkansas MSOM-MSEM
 - NC State MBA Advisor





Supporting a Balanced, Modern and Ready Nation

Administration

- Breaks
- Rules of Engagement
 - MURAL
 - Questions



Supporting a Balanced, Modern and Ready Nation

MURAL

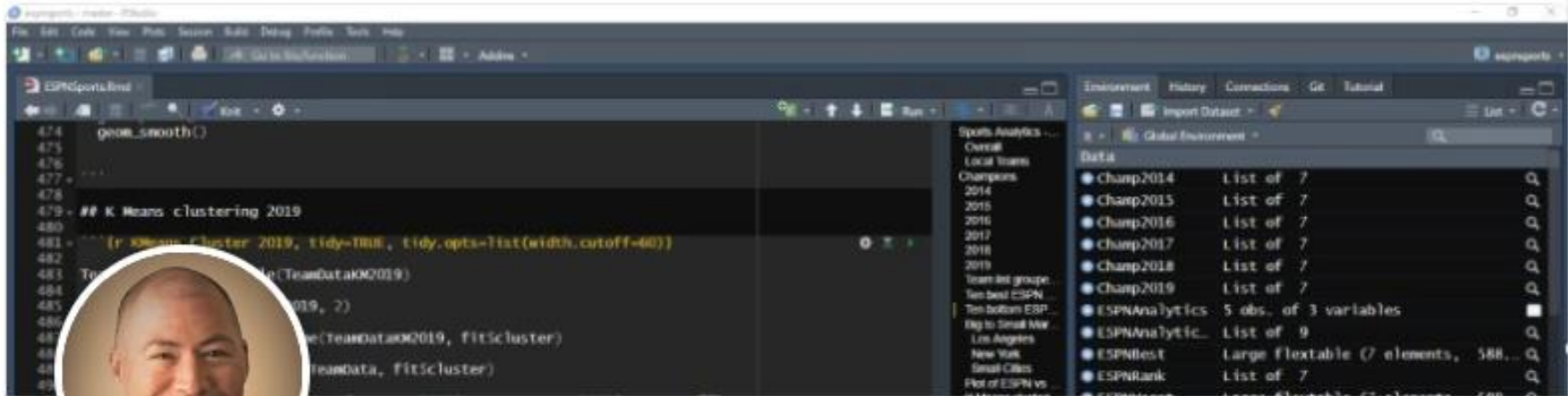
M U R A L


The screenshot displays the MURAL collaborative workspace interface for a session titled "Weekly Team Sync". The interface includes a sidebar on the left with various tool icons. The main workspace contains several interactive elements:

- Agenda:** A pink-bordered box on the left lists the session agenda:
 - 1 Icebreaker & team reflection
 - 2 Brainstorm
 - 3 Action items
- Brainstorm:** A central area with a lightbulb icon and the title "Brainstorm". It features a grid of blue sticky notes. One note in the bottom-left of the grid is circled in pink and labeled "Olivia". To the right, a circular diagram shows four blue sticky notes connected by lines, with a purple sticky note in the center.
- Helpful Links:** A section at the bottom left with a key icon, containing three document icons. A red arrow points from the second icon to a blue sticky note labeled "Alex".
- Action Items:** A section at the bottom center with the title "Action Items". It contains a grid of yellow sticky notes. A thumbs-up icon is placed over one of the notes. A pink heart icon is also visible.
- Collaboration Tools:** At the bottom right, there is a grey box containing a grid of yellow sticky notes, a purple sticky note, and a magnifying glass icon. A green star icon and the name "Grace" are visible below this box.
- Participant Avatars:** A row of avatars at the bottom of the screen shows the participants in the session: a phone icon, a smiley face icon, and five individual avatars, followed by a plus sign icon.




Supporting a Balanced, Modern and Ready Nation






Walt DeGrange


Senior Director of Analytics at CANA - Analytics & Logistics
Operations experts
Chapel Hill, North Carolina, United States
2K followers · 500+ connections



CANA LLC - Analytics & Logistics
Operations Experts



The United States Naval
Postgraduate School



Personal Website [↗](#)



Supporting a Balanced, Modern and Ready Nation

Data Preparation

MORS CSAP

Tuesday, March 11, 2025

Walt DeGrange



Supporting a Balanced, Modern and Ready Nation

Agenda

The Data

- ▶ Instruction

Tidy Data

- ▶ Instruction
- ▶ Guided practice

Exploring Data with Traditional Excel

- ▶ Instruction
- ▶ Guided practice / check for understanding
- ▶ Independent Practice & Assessment



Supporting a Balanced, Modern and Ready Nation

Case Study

National Parks

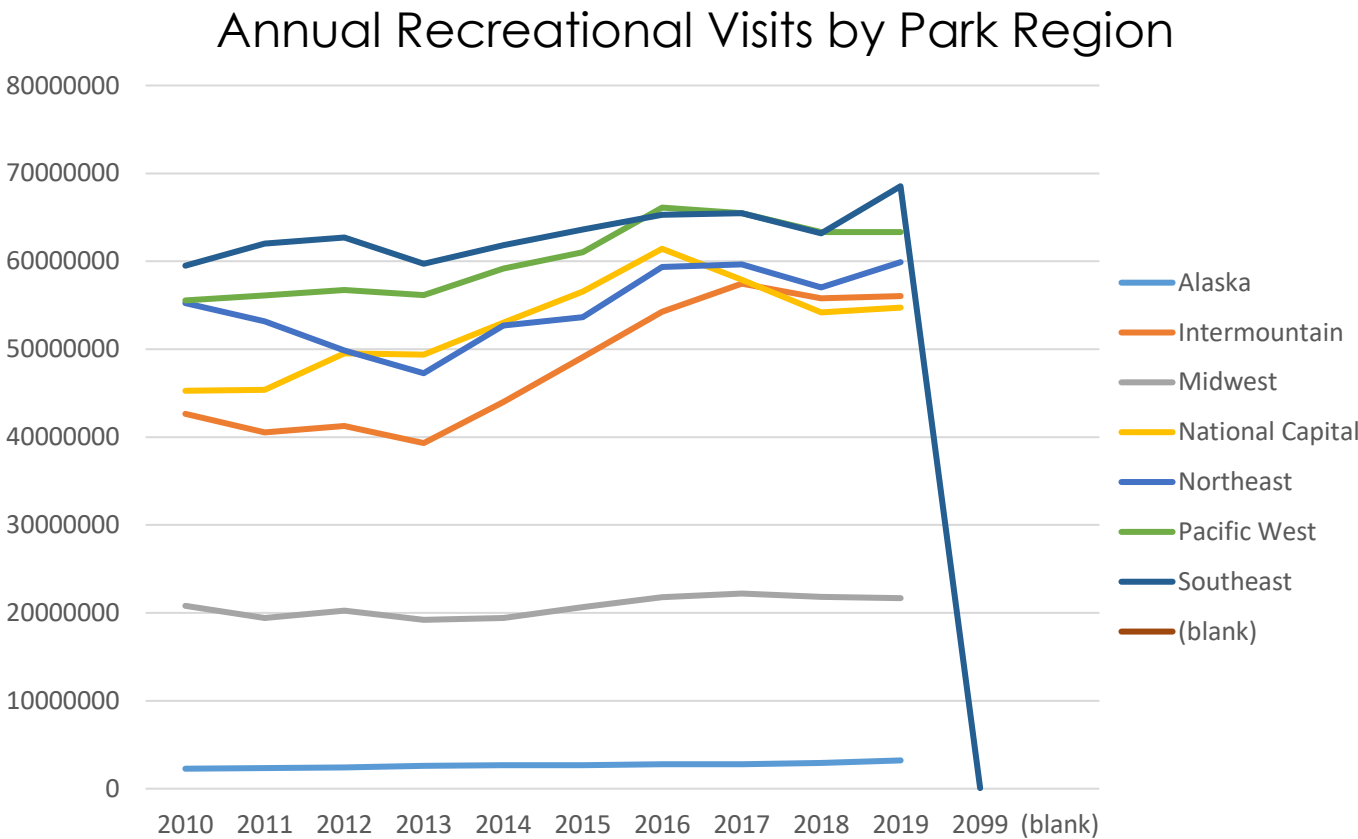
- ▶ National park visitors by year
- ▶ Data fields:
 - ▶ Year
 - ▶ National park name
 - ▶ Region
 - ▶ State
 - ▶ Visitors
 - ▶ Visit Type





Supporting a Balanced, Modern and Ready Nation

► What underlying issues impact this graph?



Common data issues

- ▶ Improperly stored data
- ▶ Duplicate records
- ▶ Outliers
- ▶ Missing data
- ▶ Invalid entries





Supporting a Balanced, Modern and Ready Nation

Tidy data

- ▶ Each variable must have its own column
- ▶ Each observation must have its own row
- ▶ Each value must have its own cell

For Excel, this means the first row is the header / list of field names and each row underneath is a record / observation

- ▶ No blank rows
- ▶ No merged cells



Supporting a Balanced, Modern and Ready Nation

This data is tidy

- ▶ Each variable has its own column
- ▶ Each observation has its own row
- ▶ Each value has its own cell

The screenshot shows the Microsoft Excel interface with the 'Home' tab selected. The ribbon includes 'Clipboard', 'Font', and 'Paragraph' groups. The active cell is A1, which contains the text 'ParkName'. The formula bar shows 'ParkName'. The worksheet contains a table with 13 rows and 4 columns. The columns are labeled 'ParkName', 'Rank', 'Value', and 'PercentOfTotal'. The rows list various national parks and their corresponding statistics.

	A	B	C	D
1	ParkName	Rank	Value	PercentOfTotal
2	Grand Canyon NP	1	320,032	16.42%
3	Lake Mead NRA	2	219,510	11.26%
4	Yosemite NP	3	155,578	7.98%
5	Olympic NP	4	96,864	4.97%
6	Great Smoky Mour	5	94,569	4.85%
7	Glen Canyon NRA	6	89,782	4.61%
8	Canyonlands NP	7	77,695	3.99%
9	Saint Croix NSR	8	68,164	3.50%
10	Mount Rainier NP	9	47,703	2.45%
11	Voyageurs NP	10	45,262	2.32%
12	Rocky Mountain N	11	41,213	2.11%
13	Yellowstone NP	12	38,236	1.96%



Supporting a Balanced, Modern and Ready Nation

Tidy or untidy?

	A	B	C
1	Camper	Paid?	Dietary Restrictions
2	Elmo	Yes - \$5.00	
3	Big Bird	No	
4	Oscar	Yes - \$3.00	Trash only



Supporting a Balanced, Modern and Ready Nation

Tidy or untidy?

religion	<\$10k	\$10–20k	\$20–30k	\$30–40k	\$40–50k	\$50–75k
Agnostic	27	34	60	81	76	137
Atheist	12	27	37	52	35	70
Buddhist	27	21	30	34	33	58
Catholic	418	617	732	670	638	1116
Don't know/refused	15	14	15	11	10	35
Evangelical Prot	575	869	1064	982	881	1486
Hindu	1	9	7	9	11	34
Historically Black Prot	228	244	236	238	197	223
Jehovah's Witness	20	27	24	24	21	30
Jewish	19	19	25	25	30	95

Table 4: The first ten rows of data on income and religion from the Pew Forum. Three columns, \$75–100k, \$100–150k and >150k, have been omitted.

Wickham, H. (2014). Tidy Data. Journal of Statistical Software, 59(10), 1 - 23. doi:<http://dx.doi.org/10.18637/jss.v059.i10>



Supporting a Balanced, Modern and Ready Nation

Wide to long

row	a	b	c
A	1	4	7
B	2	5	8
C	3	6	9

(a) Raw data

row	column	value
A	a	1
B	a	2
C	a	3
A	b	4
B	b	5
C	b	6
A	c	7
B	c	8
C	c	9

(b) Molten data

Wickham, H. (2014). Tidy Data. Journal of Statistical Software, 59(10), 1 - 23. doi:<http://dx.doi.org/10.18637/jss.v059.i10>



Supporting a Balanced, Modern and Ready Nation

Check how your data is stored

	M	N	O	P	Q	R	S	T
1	TMR	RFIDSerialNo	TMRDate	Usecode	DamagedCode	AwaitingPickup	FreeTimeLeft	DaysPastFree
2				S	D	FALSE		
3		488459		S	D	FALSE		
4				S	D	FALSE		
5				S	D	FALSE		
6		385960		S	D	FALSE		
7		0980281		S	D	FALSE		
8				S	D	FALSE		
9		52491858		S	D	FALSE		
10				S	D	FALSE		
11				S	D	FALSE		
12				S	D	FALSE		
13				S	D	FALSE		
14				s	u	FALSE		
15				s	u	FALSE		
16				s	u	FALSE		
17				s	u	FALSE		
18				s	u	FALSE		
19				s	u	FALSE		
20				s	u	FALSE		
21				s	u	FALSE		
22				S	D	FALSE		
23	MI0123			S	D	FALSE		
24				S	D	FALSE	0	614
25		492863		S	D	FALSE		
26				S	D	FALSE	0	664
27	MI0183			S	D	FALSE		

► Make sure numbers are properly stored as numbers, and that ID values & serial numbers are stored as text where appropriate

- ▶ Get a sense of the range of values in a column with the Excel Sort & Filter tool

18



Supporting a Balanced, Modern and Ready Nation

Data exploration

- ▶ Sort & Filter also quickly identifies presence of missing values and invalid levels in categorical data

Type	UnitID	OwnerCode
40FT-DRY		G
40FT-DRY		G
40FT-REF		N
20FT-DRY		G
40FT-DRY		G
40FT-DRY		G
20FT-DRY		G
20FT-DRY		G
40FT-DRY		G
20FT-DRY		G
20FT-DRY		G
40FT-DRY		G
40FT-DRY		G
20FT-DRY		G
40FT-DRY		G
20FT-DRY		G
40FT-DRY		G
20FT-DRY		G
40FT-DRY		G
20FT-DRY		G
20FT-DRY		G
20FT-DRY		G
40FT-DRY	LC3IMaterial	C
20FT-DRY	LC3IMaterial	G
20FT-DRY	LC3IBilleting	L
20FT-DRY	LC3D4MATERIALS	G
40FT-DRY	LC3IMotor Pool	U



Supporting a Balanced, Modern and Ready Nation

Conditional formatting

- ▶ Quickly scan for duplicated primary keys
- ▶ Primary keys should uniquely identify a record

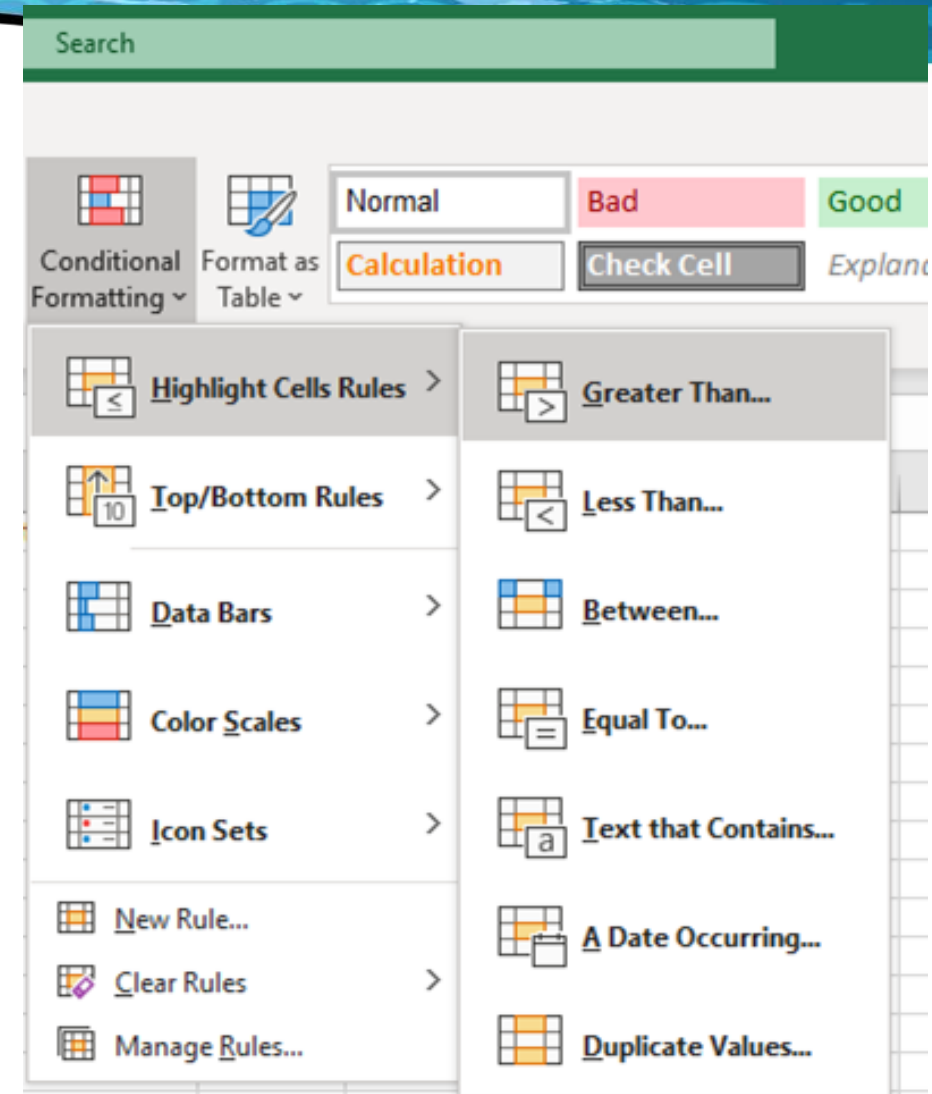
The screenshot shows the Microsoft Excel interface with the 'Conditional Formatting' menu open. The spreadsheet data is as follows:

	A	B	C	D	E	F
	CountryName	LocationName	ContainerNo	PoolIn	PendingBuyoutInd	Type
1	IRAQ	ABU GHURAYB AAFES	UACU4957873	FALSE	FALSE	40FT-DRY
2	IRAQ	ABU GHURAYB AAFES	TEXU7019579	FALSE	FALSE	40FT-DRY
3	IRAQ	ABU GHURAYB AAFES	ALRU400058	FALSE	FALSE	40FT-REF
4	IRAQ	ABU GHURAYB AAFES	TOLU2396084	FALSE	FALSE	20FT-DRY
5	IRAQ	ABU GHURAYB AAFES	SEAU7903244	FALSE	FALSE	40FT-DRY
6	IRAQ	ABU GHURAYB AAFES	CLHU8591088	FALSE	FALSE	40FT-DRY
7	IRAQ	ABU GHURAYB AAFES	TRLU4589344	FALSE	FALSE	40FT-DRY
8	IRAQ	ABU GHURAYB AAFES	GLDU2031125	FALSE	FALSE	20FT-DRY
9	IRAQ	ABU GHURAYB AAFES	POCU0273732	FALSE	FALSE	20FT-DRY
10	IRAQ	ABU GHURAYB AAFES	SCZU3234072	FALSE	FALSE	40FT-DRY
11	IRAQ	ABU GHURAYB AAFES	CAXU6452953	FALSE	FALSE	20FT-DRY
12	IRAQ	ABU GHURAYB AAFES	CAXU9765150	FALSE	FALSE	40FT-DRY
13	IRAQ	ABU GHURAYB AAFES	TRLU4353488	FALSE	FALSE	40FT-DRY
14	IRAQ	ABU GHURAYB AAFES	SVWU6556526	FALSE	FALSE	20FT-DRY
15	IRAQ	ABU GHURAYB AAFES	MSKU6268260	FALSE	FALSE	40FT-DRY
16	IRAQ	ABU GHURAYB AAFES	NOSU4437388	FALSE	FALSE	40FT-DRY
17	IRAQ	ABU GHURAYB AAFES	GATU4017715	FALSE	FALSE	40FT-DRY
18	IRAQ	ABU GHURAYB AAFES				

The 'Conditional Formatting' menu is open, showing options like 'Highlight Cells Rules', 'Top/Bottom Rules', 'Data Bars', 'Color Scales', 'Icon Sets', 'New Rule...', 'Clear Rules', and 'Manage Rules...'. The 'Duplicate Values...' option is highlighted under 'Highlight Cells Rules'.

Data exploration

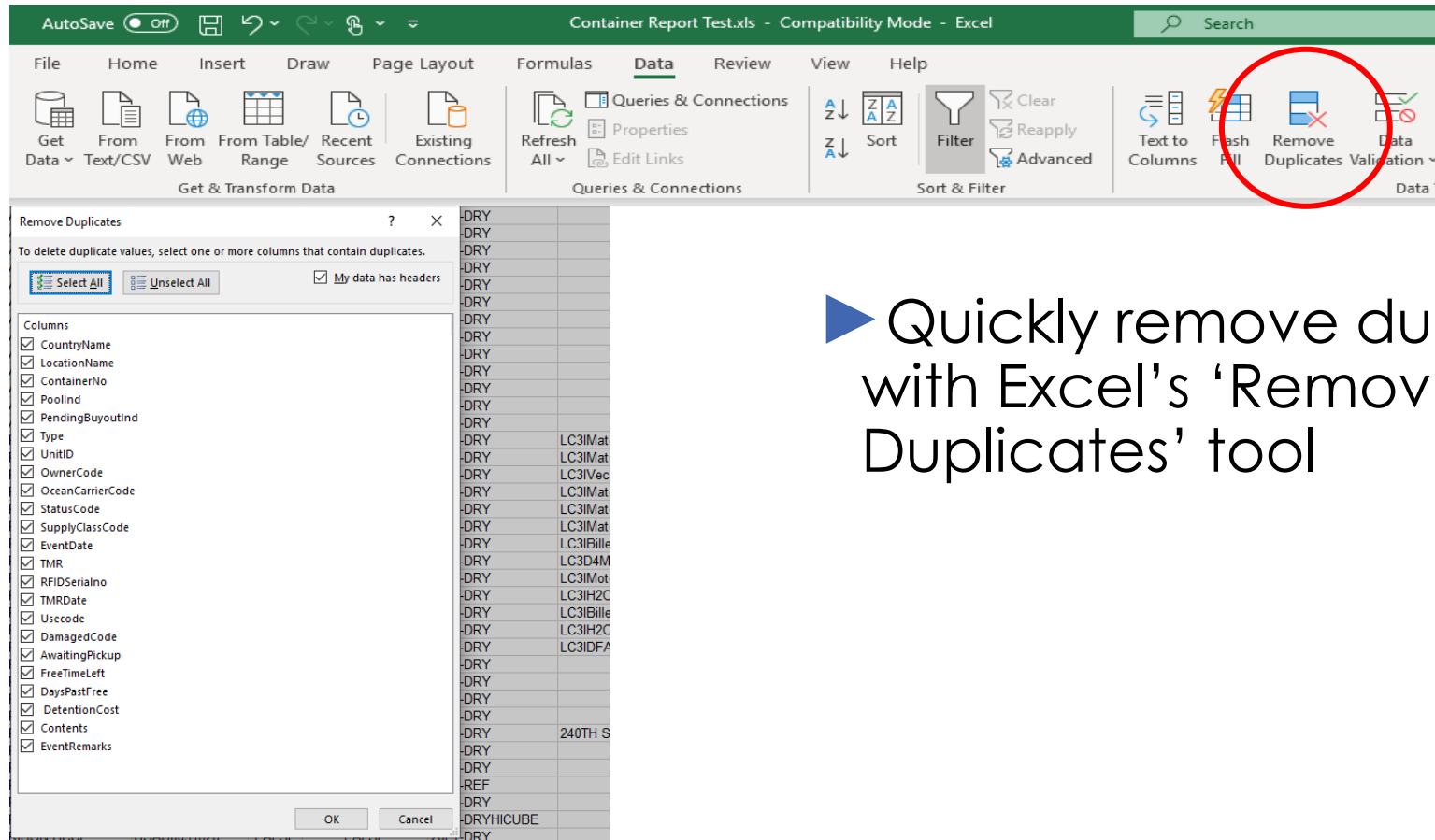
- Highlight outliers and invalid numeric entries with conditional formatting





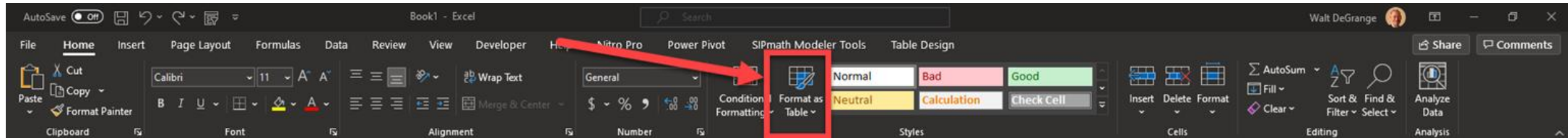
Supporting a Balanced, Modern and Ready Nation

Remove duplicates



► Quickly remove duplicates with Excel's 'Remove Duplicates' tool

Functions to summarize data with Format as Table



- ▶ SUM()
- ▶ COUNTA()
- ▶ MIN()
- ▶ MAX()
- ▶ Range: Max()-Min()
- ▶ AVERAGE()
- ▶ MEDIAN()
- ▶ MODE()
- ▶ VAR()
- ▶ STDEV()



Supporting a Balanced, Modern and Ready Nation

Learn More

- Alexander, M., Kusleika, R. and Walkenbach, J. **Excel 2019 Bible**. Wiley Press. 2018.
- Wickham, Hadley. **Tidy data**. Journal of Statistical Software 59.10 (2014): 1-23.