WORKING WITH DATA IN EXCEL FROM A STUDENT AFFAIRS PERSPECTIVE

ZANE KRATZER

DATA RESEARCH SCIENTIST

LEHIGH UNIVERSITY

OFFICE OF INSTITUTIONAL RESEARCH AND STRATEGIC ANALYTICS

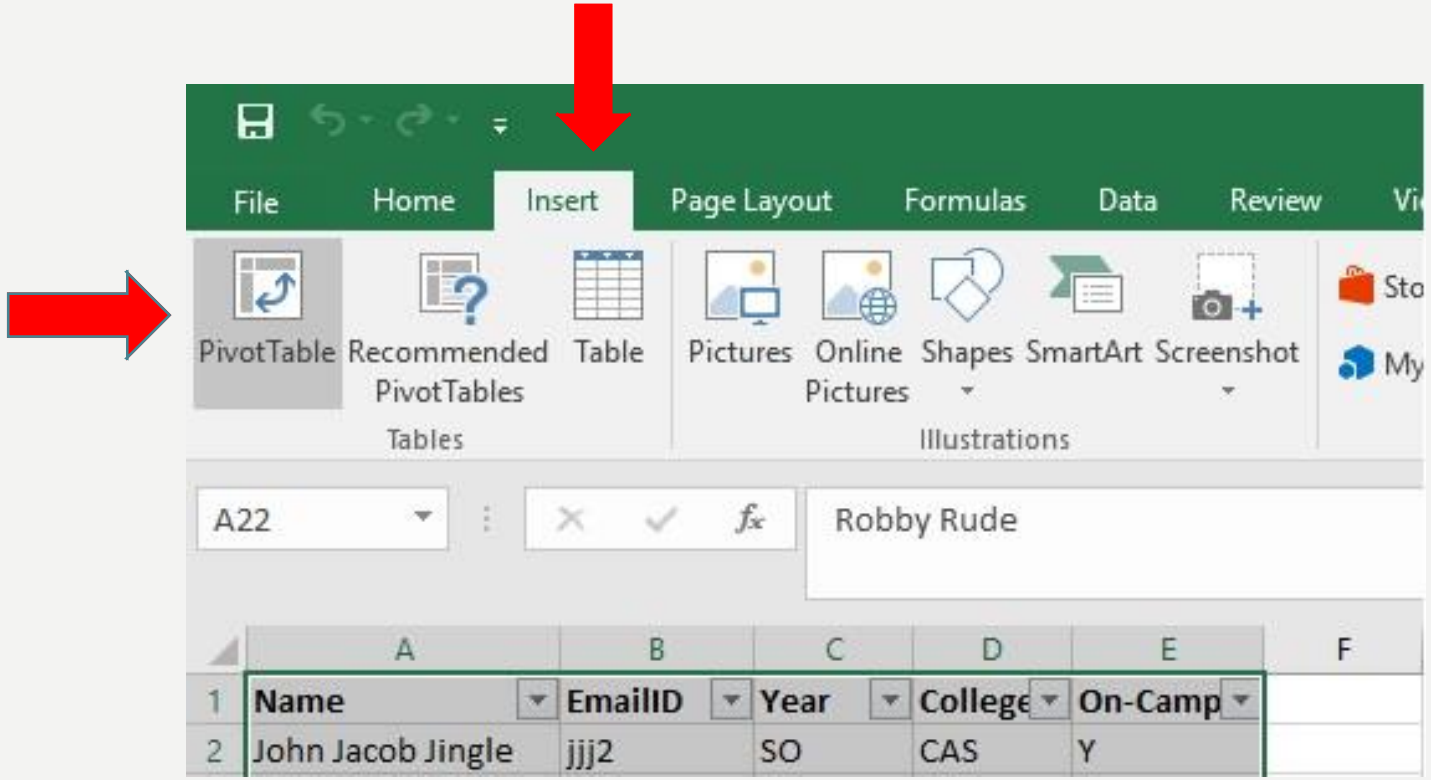
WORKING WITH PIVOT TABLES

- 'Ctrl-A' to select all relevant fields and rows in a spreadsheet
- Be sure to remove empty columns and empty header rows

	A	B	C	D	E	F
1	Name	EmailID	Year	College	On-Camp	
2	John Jacob Jingle	jjj2	SO	CAS	Y	
3	Rammy Ramsbot	rar5	JR	CAS	Y	
4	Jani Lane	jql6	SO	CAS	Y	
5	Roddy Rude	rar6	FR	BUS	Y	
6	Sabintha Vali	slv7	FR	BUS	Y	
7	Lani Lemuel	lvl4	FR	EN	Y	
8	Baba O'Riley	bor5	SO	EN	Y	
9	Harry Henderson	heh205	FR	EN	Y	
10	Rey Skywalker	ros222	SO	CAS	Y	
11	Mega Tron	mat56	SO	CAS	Y	
12	Black Widow	bqw207	JR	BUS	N	
13	Pete Seeger	pas209	SO	BUS	Y	
14	Rory Rockerbock	ryr219	SO	EN	Y	
15	Fred Flinstone	fnf218	SR	IC	N	
16	Dorothy Dale	dtd217	FR	CAS	Y	
17	Suzy Q.	sqq5	SR	IC	N	
18	Betsy Black	blb218	JR	CAS	N	
19	Grimmy Reapa	gyr217	SO	EN	Y	
20	Jasper Jiver	jjj219	SO	IC	Y	
21	Rory Rockerbock	ryr216	SO	EN	Y	
22	Robby Rude	rar6	FR	BUS	Y	
23						

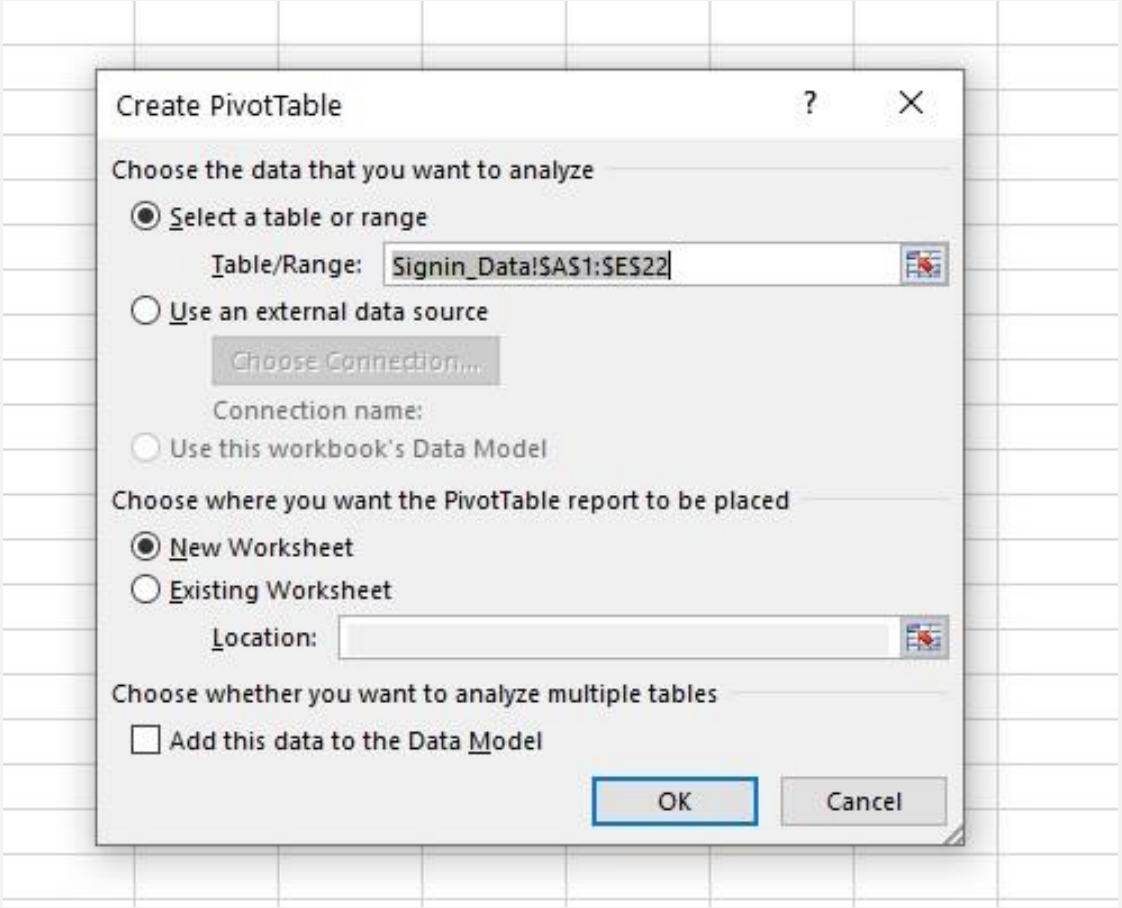
WORKING WITH PIVOT TABLES

- Insert > Pivot Table



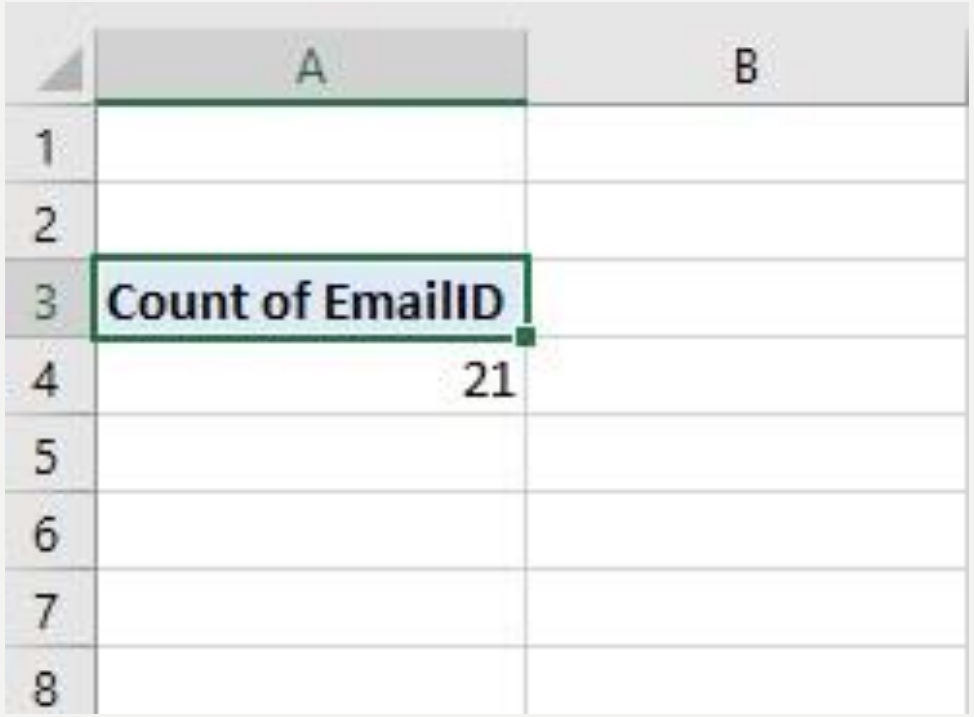
WORKING WITH PIVOT TABLES

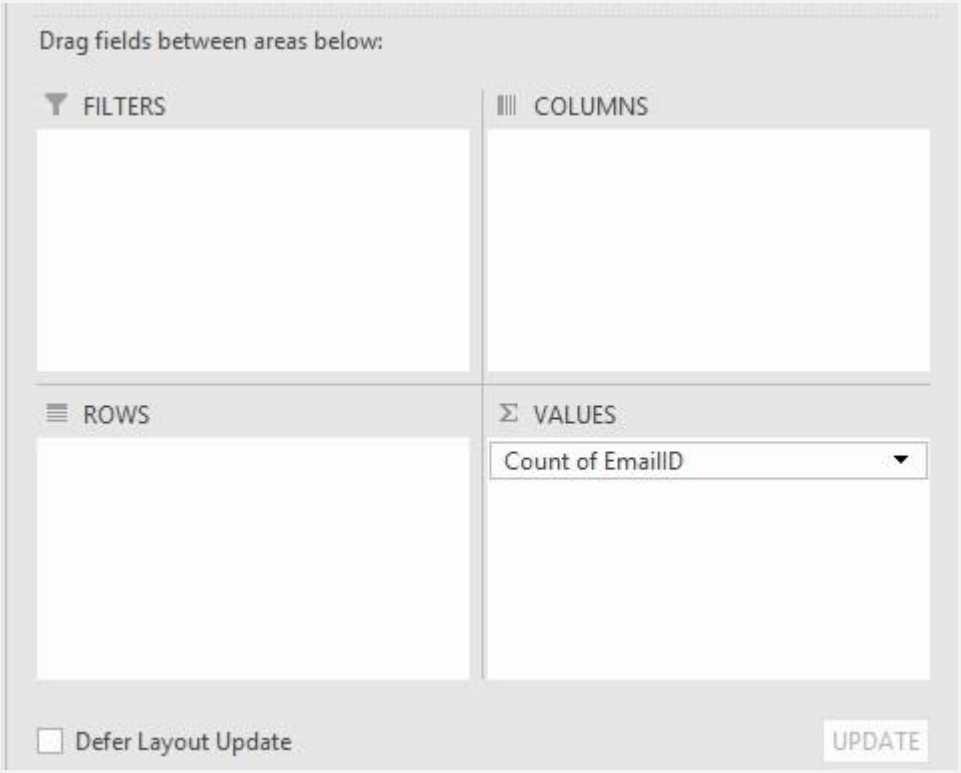
- Choose between ‘New Worksheet’ and ‘Existing Worksheet’
- This tells Excel where to paste your new Pivot Table, on the current worksheet where your data resides or in a new bland worksheet.
- Also, you can check the table range here to make sure that Excel is grabbing the correct rows and columns.



WORKING WITH PIVOT TABLES

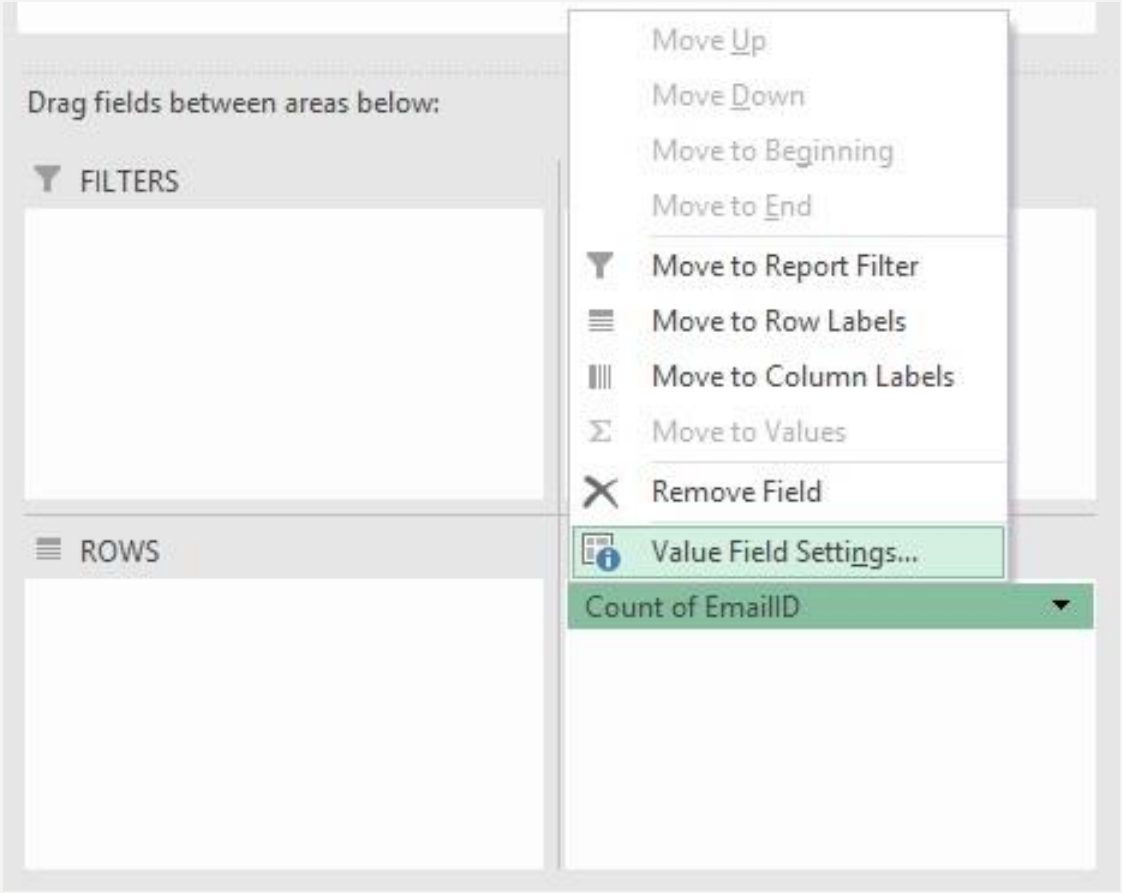
- Values: The field to be summarized (i.e. EmailID provides a Count of all EmailID values)





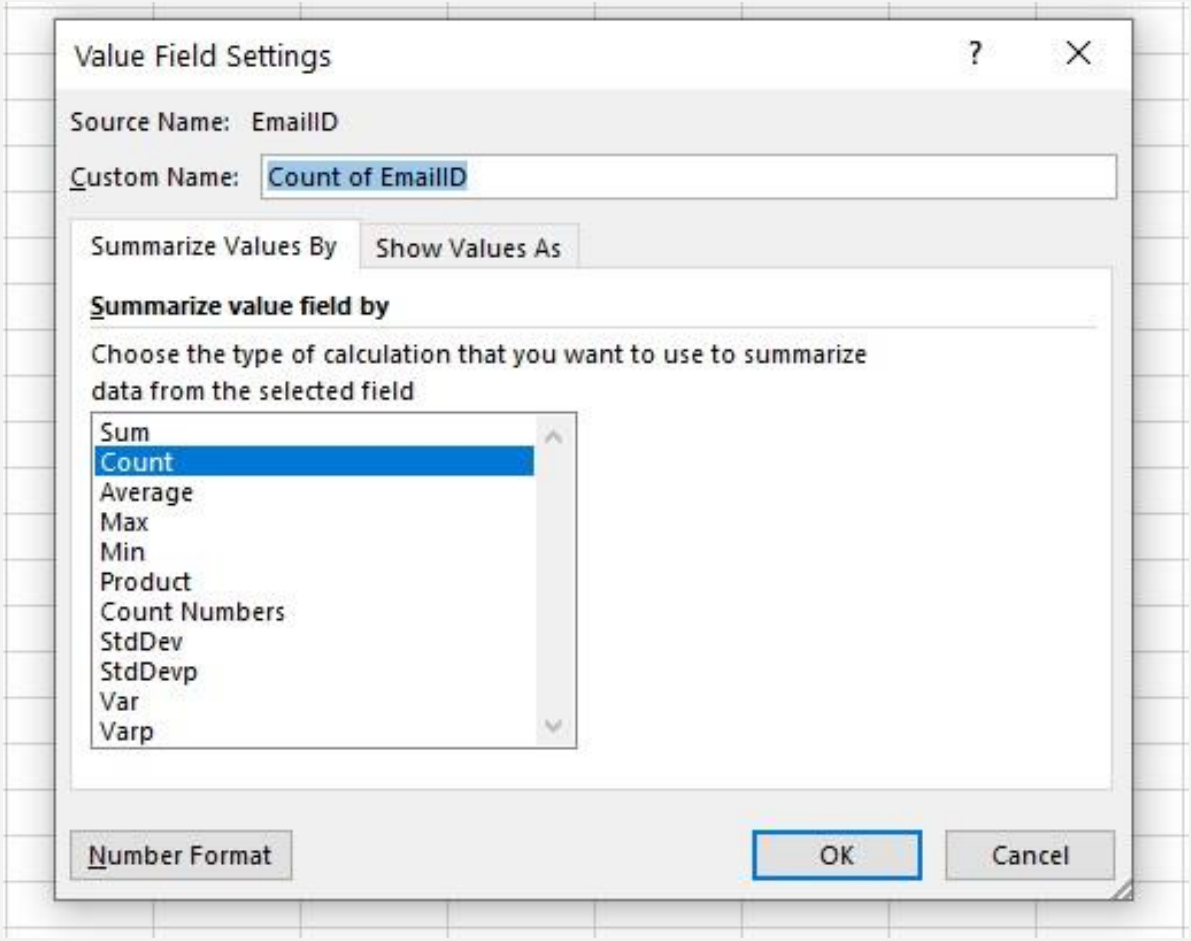
WORKING WITH PIVOT TABLES

- To change how the sum works, click on the dropdown arrow and select "Value Field Settings"



WORKING WITH PIVOT TABLES

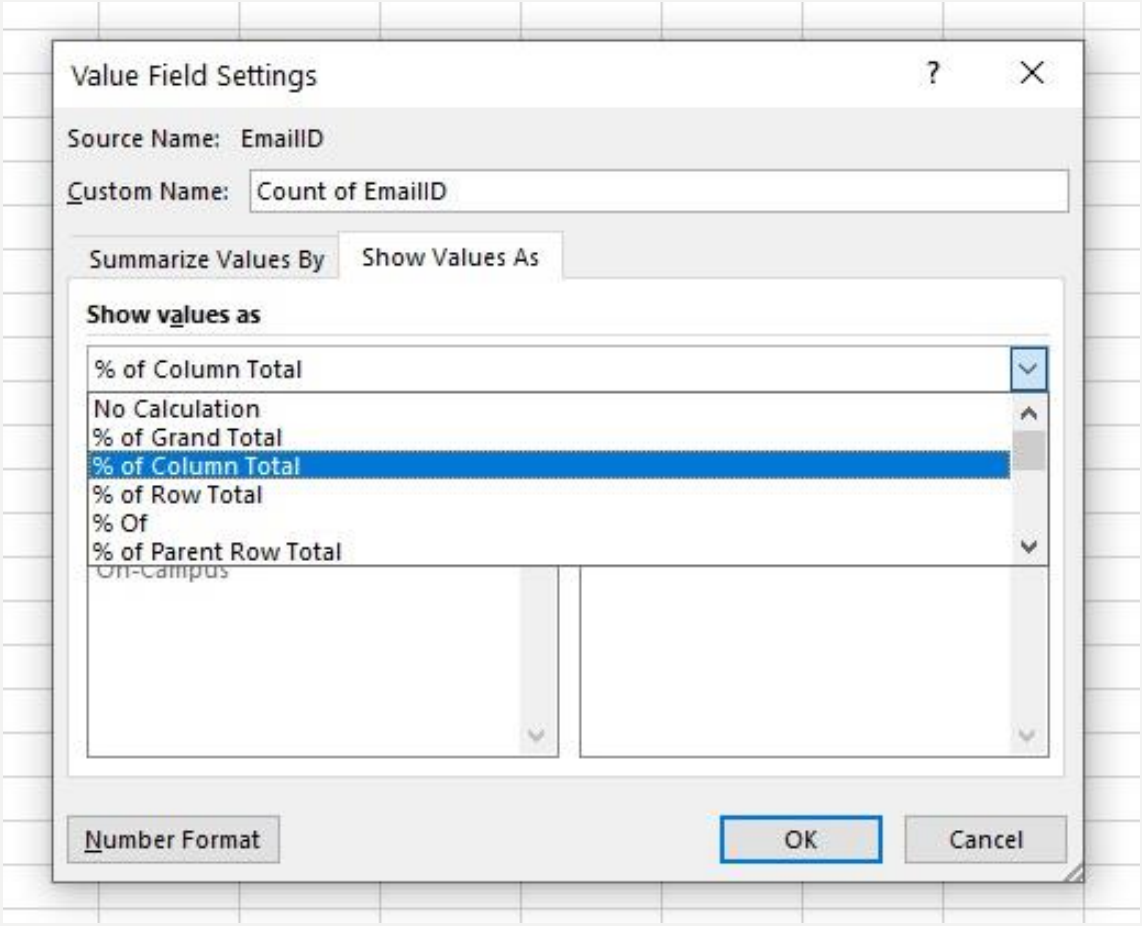
- Change the "Summarized value field by" to Sum, Count, Average, Max, Min, etc.
- Note that if your selected field is a text column then numeric options such as Sum and Average will not work



WORKING WITH PIVOT TABLES

- To show the summed values as a percentage or other form, click on the "Show Value As" tab and select an option such as "% of Column Total"

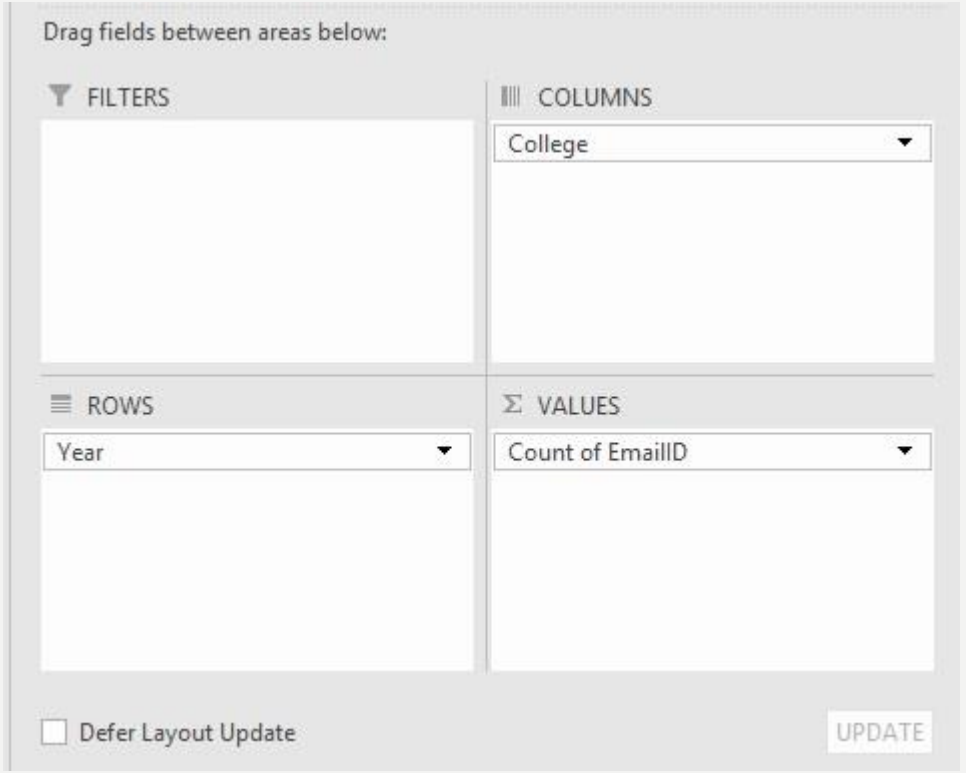
	A	B	C
1			
2			
3	Row Labels	Count of EmailID	
4	FR	28.57%	
5	JR	14.29%	
6	SO	47.62%	
7	SR	9.52%	
8	Grand Total	100.00%	
9			



WORKING WITH PIVOT TABLES

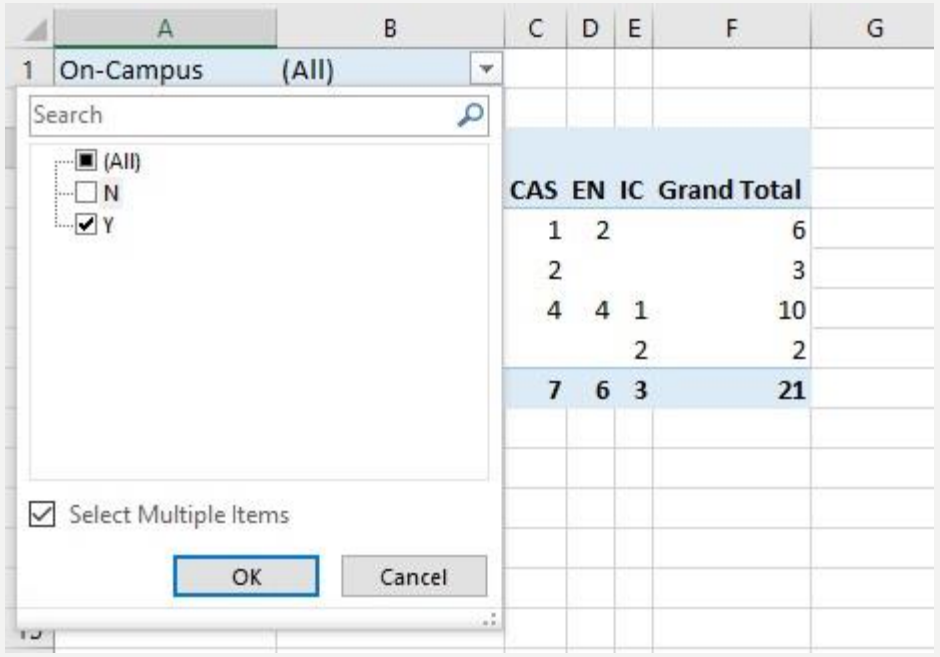
- Rows and Columns: Fields to be inserted into either the Row or Column of the pivot table

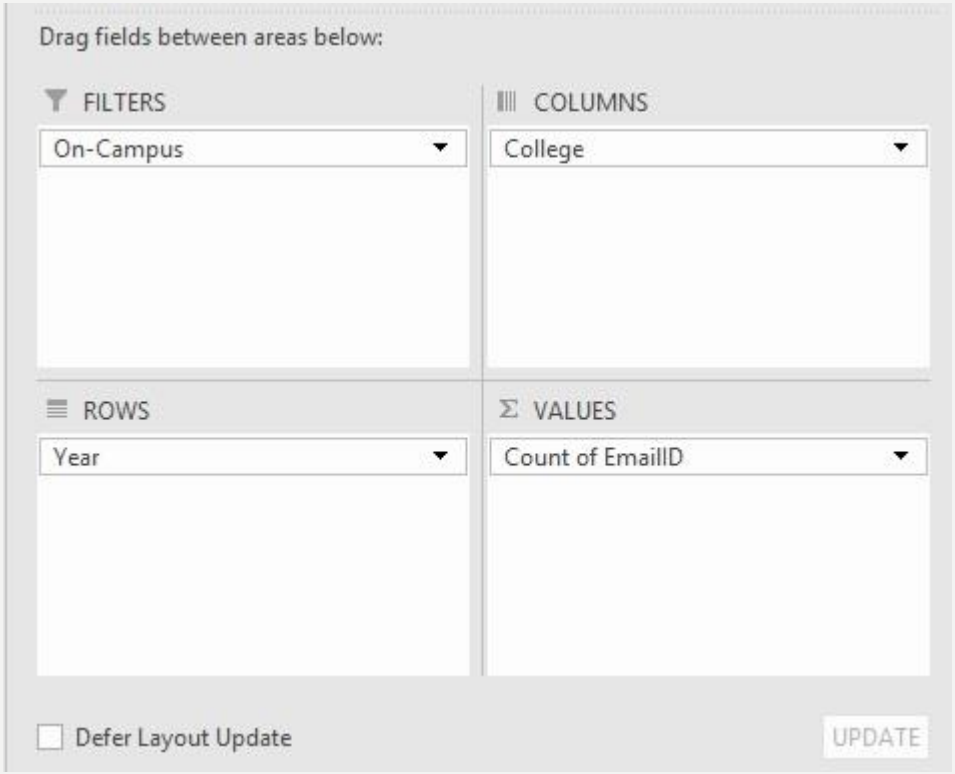
Count of EmailID	Column Labels				
Row Labels	BUS	CAS	EN	IC	Grand Total
FR	3	1	2		6
JR	1	2			3
SO	1	4	4	1	10
SR				2	2
Grand Total	5	7	6	3	21



WORKING WITH PIVOT TABLES

- Filters: Fields to be used to filter out specific values from the data



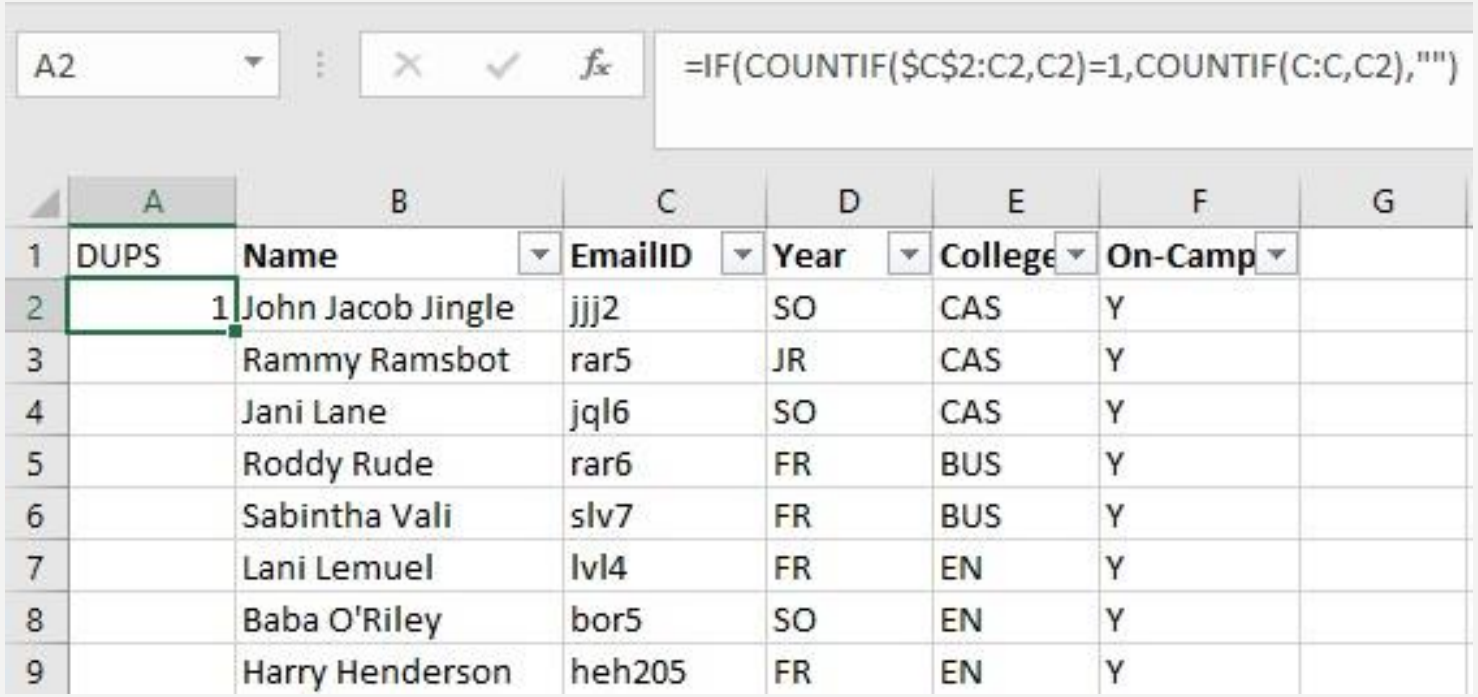


IDENTIFYING DUPLICATE RECORDS

- Counting Duplicate Records by Formula

  =IF(COUNTIF(B2:B2,B2)=1,COUNTIF(B:B,B2),"")

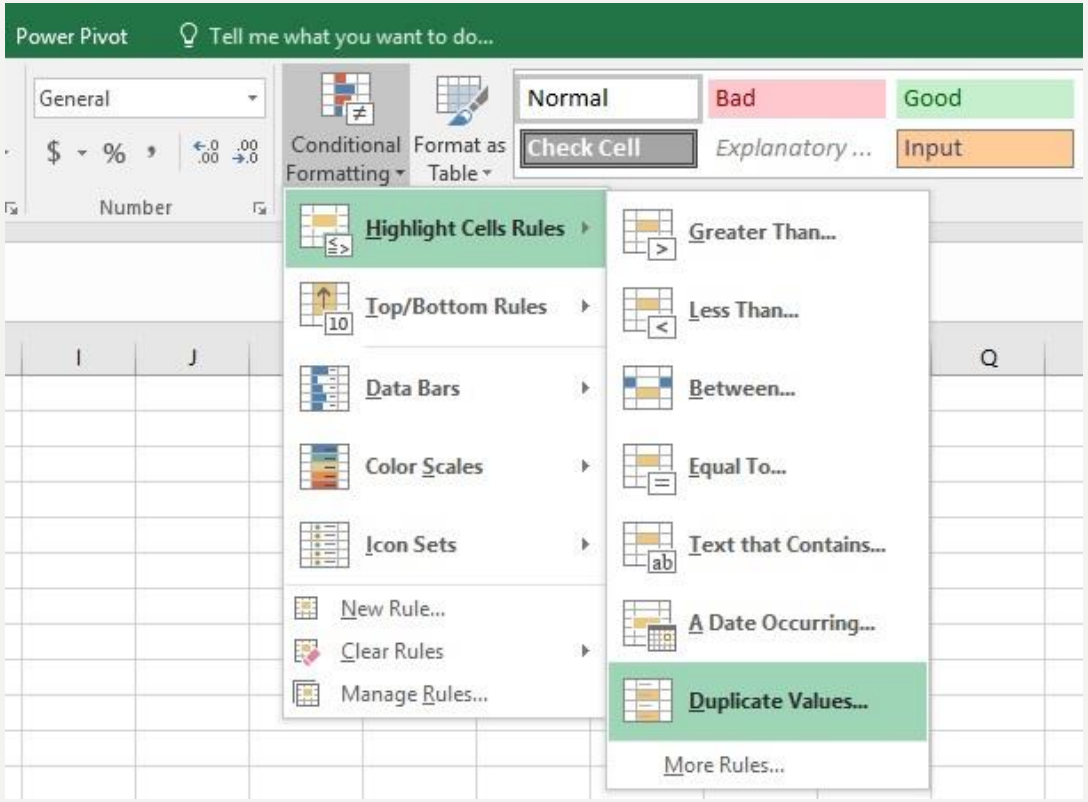
- Replace 'B' with whatever column your target field is in
- If your data does not begin on row 2, replace '2' with whatever row your data begins on



	A	B	C	D	E	F	G
1	DUPS	Name	EmailID	Year	College	On-Camp	
2	1	John Jacob Jingle	jjj2	SO	CAS	Y	
3		Rammy Ramsbot	rar5	JR	CAS	Y	
4		Jani Lane	jql6	SO	CAS	Y	
5		Roddy Rude	rar6	FR	BUS	Y	
6		Sabintha Vali	slv7	FR	BUS	Y	
7		Lani Lemuel	lvl4	FR	EN	Y	
8		Baba O'Riley	bor5	SO	EN	Y	
9		Harry Henderson	heh205	FR	EN	Y	

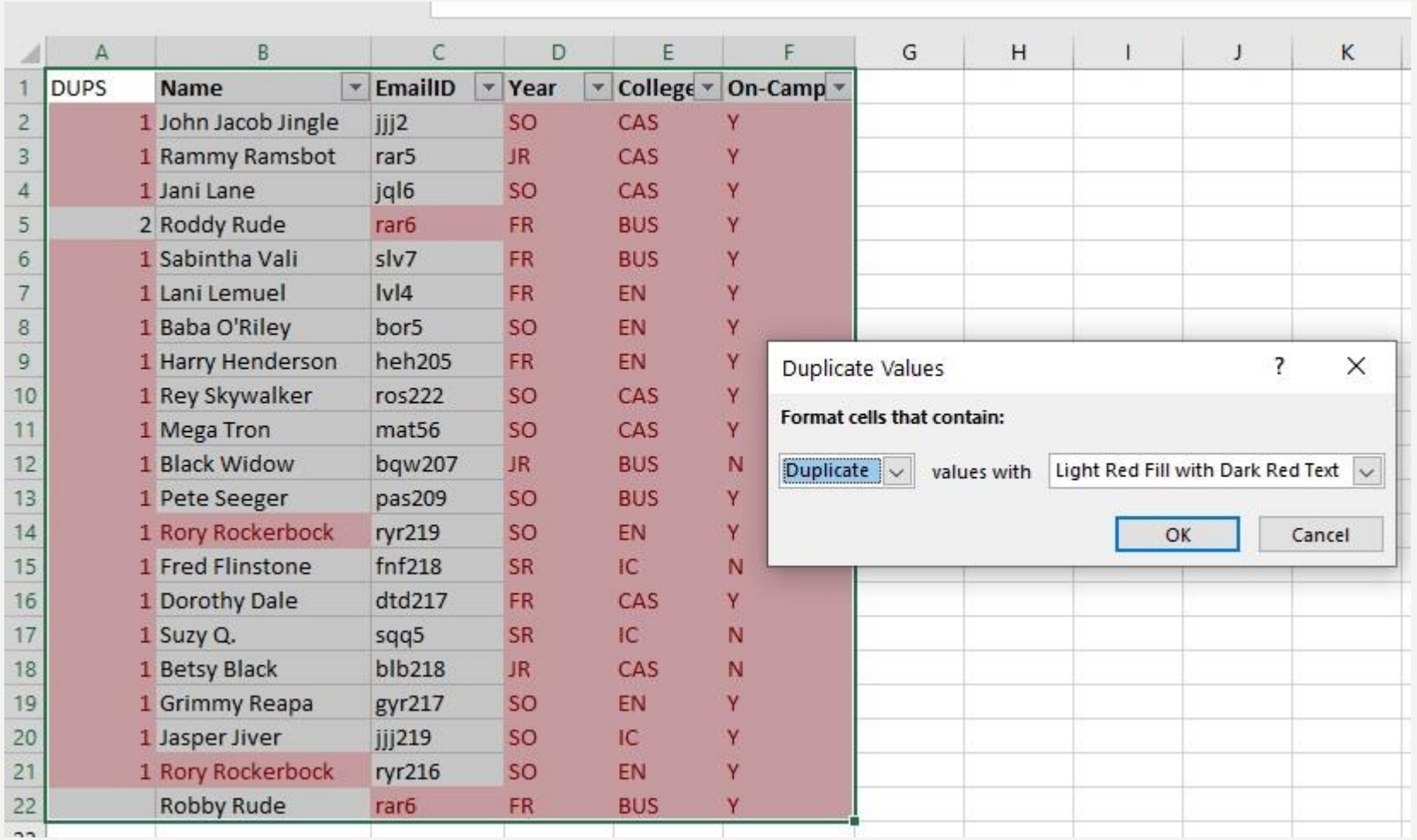
IDENTIFYING DUPLICATE RECORDS

- Finding Duplicate Records with Conditional Formatting
- ‘Ctrl-A’ to select all relevant fields and rows in a spreadsheet
- Conditional Formatting > Highlight Cells Rules > Duplicate Values



IDENTIFYING DUPLICATE RECORDS

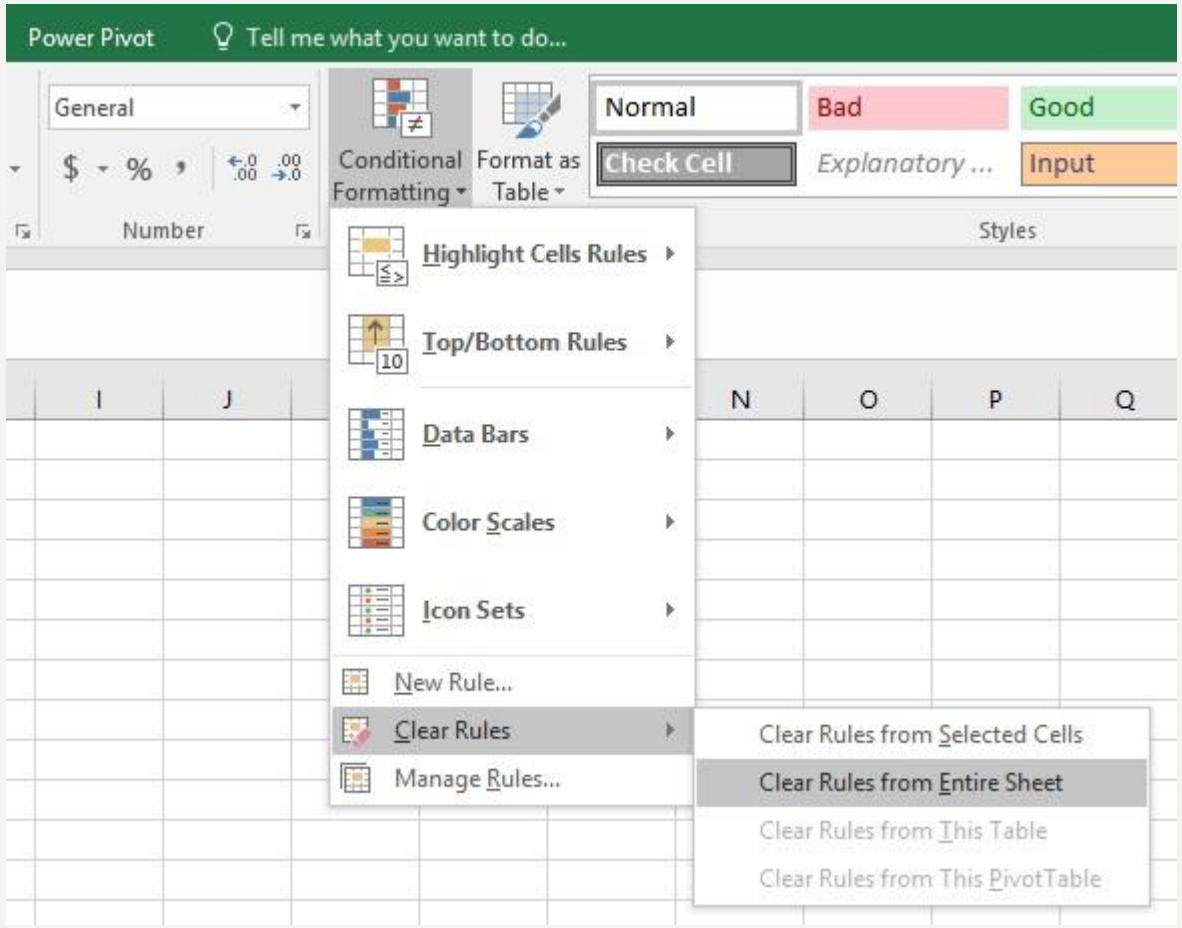
- Check default settings on Duplicate Values dialog box on select OK



	A	B	C	D	E	F
1	DUPS	Name	EmailID	Year	College	On-Camp
2	1	John Jacob Jingle	jjj2	SO	CAS	Y
3	1	Rammy Ramsbot	rar5	JR	CAS	Y
4	1	Jani Lane	jql6	SO	CAS	Y
5	2	Roddy Rude	rar6	FR	BUS	Y
6	1	Sabintha Vali	slv7	FR	BUS	Y
7	1	Lani Lemuel	lvl4	FR	EN	Y
8	1	Baba O'Riley	bor5	SO	EN	Y
9	1	Harry Henderson	heh205	FR	EN	Y
10	1	Rey Skywalker	ros222	SO	CAS	Y
11	1	Mega Tron	mat56	SO	CAS	Y
12	1	Black Widow	bqw207	JR	BUS	N
13	1	Pete Seeger	pas209	SO	BUS	Y
14	1	Rory Rockerbock	ryr219	SO	EN	Y
15	1	Fred Flinstone	fnf218	SR	IC	N
16	1	Dorothy Dale	dtd217	FR	CAS	Y
17	1	Suzy Q.	sqq5	SR	IC	N
18	1	Betsy Black	blb218	JR	CAS	N
19	1	Grimmy Reapa	gyr217	SO	EN	Y
20	1	Jasper Jiver	jjj219	SO	IC	Y
21	1	Rory Rockerbock	ryr216	SO	EN	Y
22		Robby Rude	rar6	FR	BUS	Y

IDENTIFYING DUPLICATE RECORDS

- To remove formatting, go to Conditional Formatting > Clear Rules > Clear Rules from Entire Sheet



MERGING MULTIPLE TABLES

Tools for merging tables, pulling data from one table into another:

- VLOOKUP

  https://exceljet.net/excel-functions/excel-vlookup-function
- INDEX MATCH

  https://exceljet.net/index-and-match
- Power Query

  https://www.howtoexcel.org/power-query/the-complete-guide-to-power-query/

COMBINING YEARS/MULTIPLE EVENTS

- What is the unit of analysis?
- If the student record must be unique, then the data should be in a wide format, with each row representing a unique student and all subsequent fields or measurements for the student added as extra columns.
- If the unit of analysis is the point of measurement for each student (a specific term, date, event), then the data should be in a long format, with each row representing a unique data entry for each student. Students will have duplicate rows if data is recorded for them more than once (i.e. for every event or semester attended).

OIRSA CONTACT INFO

For general OIRSA contact and information:

Website: oirsa.lehigh.edu

Email: oir@lehigh.edu

Phone: 610-758-5890

To request our services: oirsa.lehigh.edu/make-request

For specific questions regarding this presentation:

Zane Kratzer, Data Research Scientist

Email: zlk205@lehigh.edu

Phone number: 610-758-5918