UNC UNIVERSITY LIBRARIES

Working with Data in Excel

[Download Companion Excel Sheet]

[] indicate sheet names in the companion Excel workbook

[Getting Started]

Shortcuts

To…	Windows	Mac
Find and replace	CTRL + F	+F
Move to the edge of the data region	CTRL+ Arrow key	+Arrow key
Select to the edge of the data region	CTRL+SHIFT+Arrow key	+SHIFT+Arrow key
Select entire column	CTRL+SPACEBAR	CTRL+SPACEBAR
Select entire row	SHIFT+SPACEBAR	SHIFT+SPACEBAR
Enter value into all selected cells	CTRL+ENTER	^+RETURN

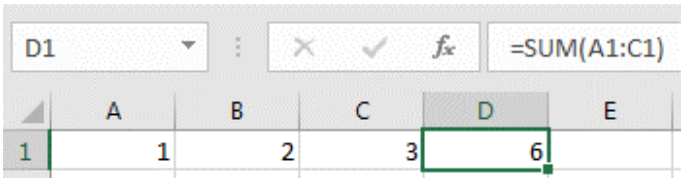
Best Practices and Tidy Data

- Think of each row as an "observation" and each column as a "variable" or "field".
- Avoid color or other formatting alone to encode data.
- Only one table per sheet.
- Each observation should be of the same type (e.g. a student, or a NC county in a given year).
- Try saving your data as a .csv (comma separated values) file. This saves one sheet without formatting.

[Functions]

Using Functions in Excel

- =SUM()
- =AVERAGE()
- Etc.



Working with Text

- **LEFT()** Extracts a specified number of characters from a variable, counting from the left
  - **RIGHT()** same as above, but counting from the right
- **TRIM()** Removes all whitespace aside from since spaces between words
- **CONCATENATE()** combines multiple strings into a single string

Paste Special: Making Functions Permanent

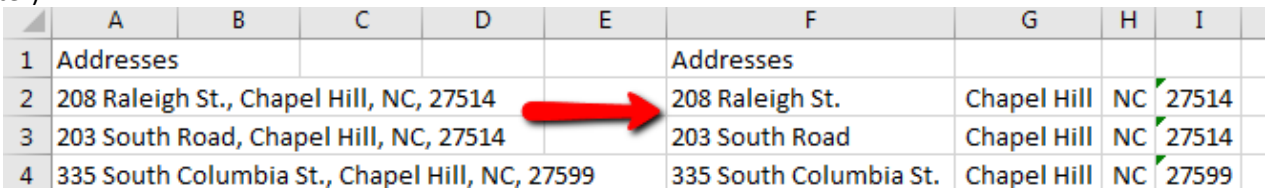
- Right-click: Copy
- Right-click: Paste Special>Values

Paste [Transpose]

- Right-click: Copy
- Right-click: Paste Special>Transpose

Common Problems: [Splitting] on Delimiters

The Text to Columns tool (Data Tab>) lets you split a cell into multiple cells based on width or a special character (delimiter).



Common Problems: Filling [Blanks]

When dealing with human-readable text, we often have categories listed once with the implication that all lines before the next category fall into this group. For example, Bertie Rudolph is a Freshman.

- (PC) Home Tab, Editing > Find & Select> Go to Special> Blanks> OK
- (Mac) Edit Menu>Find>Go To…>Special…>Blanks>OK
- Type =, then hit the up directional arrow. Hit CTRL+Enter (PC) or CMD+Enter (Mac)

	A	B
1	Category	Name
2	Freshman	Janey Hackler
3		Seth Wickham
4		Bertie Rudolph
5		Jonelle Fahnestock
6		Yanira Steffensen
7	Sophomore	Nadia Wiggin
8		Ilana Fortuna

	A	B
1	Category	Name
2	Freshman	Janey Hackler
3	Freshman	Seth Wickham
4	Freshman	Bertie Rudolph
5	Freshman	Jonelle Fahnestock
6	Freshman	Yanira Steffensen
7	Sophomore	Nadia Wiggin
8	Sophomore	Ilana Fortuna

[VLOOKUP]

The VLOOKUP function provides a way to merge or join additional data into a dataset, using a common code or value.

	A	B	C	D	E	F	G	H
1	**=VLOOKUP(A3,F3:G9,2,FALSE)**							
2	Code	Test	Outcome			Code	Name	
3	21325	1	A			84571	Sanora Gean	
4	44591	1	B			21111	Lynetta Dostal	
5	21111	2	B			21470	Chin Bisson	
6	57257	2	C			21325	Phung Goffney	
7						57257	Lou Iacovelli	
8						11174	Wilmer Falkner	
9						44591	Loura Goodnight	

=VLOOKUP(A2,E1:F8,2,FALSE)

Parameter	Value	Description
lookup_value	**A3**	***value in our main table*** that we're looking to match in the other table
table_array	**F3:G9**	the ***other table*** we need information from **(note: lock references with $: $F$3:$G$9)**
col_index_num	**2**	the ***column from the other table*** we're looking for
[range_lookup]	**FALSE**	whether you want approximate matches (TRUE) or exact matches (FALSE)

Exercise: [Ex_Main], [Ex_Lookup]

Intro to PivotTables

Pivot_Tables_IPEDS.xlsx

Windows: Insert Tab>PivotTable

Mac: Go to Data Tab>PivotTable>Create Manual Pivot Table…

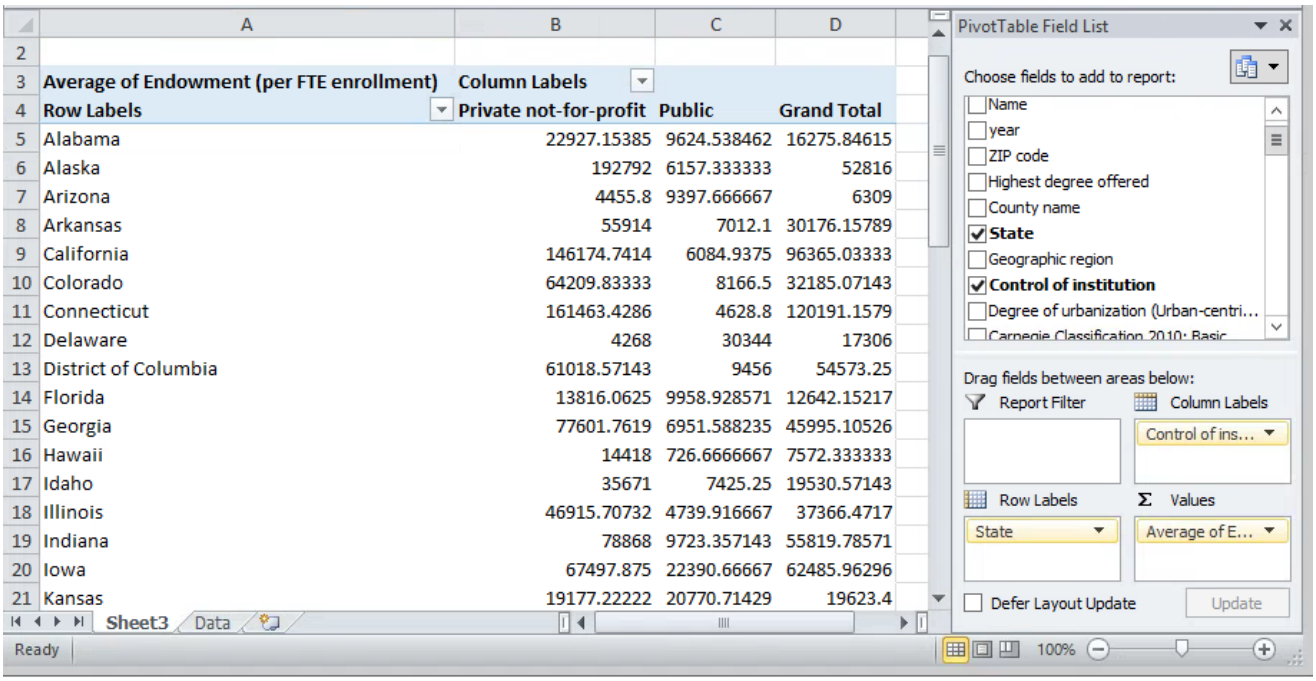
Pivot Tables create cross-tabulations displaying values split out across categories displayed as row and/or column headings.

Adding data: Click and drag to areas at the bottom of "PivotTable Fields"

Columns and Rows: The *categories* on the edges of the table

- Multiple categories on a single axis will be nested

Values: The *numbers* shown in the cells of the table (each cell matches **all** categories it lines up with)

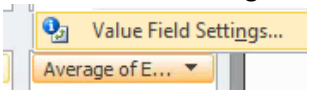


Average of Endowment (per FTE enrollment)	Column Labels		
Row Labels	Private not-for-profit	Public	Grand Total
Alabama	22927.15385	9624.538462	16275.84615
Alaska	192792	6157.333333	52816
Arizona	4455.8	9397.666667	6309
Arkansas	55914	7012.1	30176.15789
California	146174.7414	6084.9375	96365.03333
Colorado	64209.83333	8166.5	32185.07143
Connecticut	161463.4286	4628.8	120191.1579
Delaware	4268	30344	17306
District of Columbia	61018.57143	9456	54573.25
Florida	13816.0625	9958.928571	12642.15217
Georgia	77601.7619	6951.588235	45995.10526
Hawaii	14418	726.6666667	7572.333333
Idaho	35671	7425.25	19530.57143
Illinois	46915.70732	4739.916667	37366.4717
Indiana	78868	9723.357143	55819.78571
Iowa	67497.875	22390.66667	62485.96296
Kansas	19177.22222	20770.71429	19623.4

In most cases, there will be **many** rows in your dataset that fall into both categories, so we need to summarize or aggregate the data. In the example above, there are likely many public and private universities in any given state.

Aggregation for Values

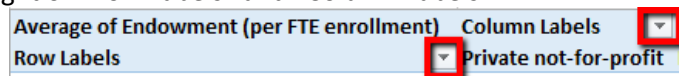
- Click on a field in the Values area and choose "Value Field Settings" to change the default aggregation



- "Summarize Values By" determines the mathematical function used to summarize the cells
  - Basic frequencies are available via the Count function for **any** field.
- "Show Values As" allows more complex calculations based on other cells (e.g. percent of total)
- *Mac: Click the "i" icon on the Value. "Show Values As" is located under the "Options>>" tab.*

Sorting and Filtering Columns and Rows

- Use the arrows at the right of "Row Labels" and "Column Labels"



- "More Sort Options" provides advanced sorting by Values (*Mac: Not Available)*
- Filter with the check boxes next to unique values (*Mac: Not Available)*
  - Advanced filtering available in "Label Filters" and "Value Filters"

Filters

- **Filters:** Dragging a field to the Filters area will create a filter box similar to those in a regular spreadsheet

PivotTable Exercises

1. What are the Enrolled Totals in public and private schools (see Control of institution)?

2. Which Geographic Region has the highest average ACT Composite 75th percentile score? How many regions have average scores below the national (Grand Total) average?

3. Which category of Degree of Urbanization contains the most public (Control of Institution) universities?
   a. What percent of the Applications total go to each of the top two Degrees of Urbanization? (Hint: Use the Show Values As tab in Value Field Settings)

Next Steps:

- Power Query (Windows-only): Loading and filtering large datasets
    - Data Tab> Get & Transform
- Data Validation: Control data entry to prevent errors
    - Data Tab> Data Validation

Getting Help:

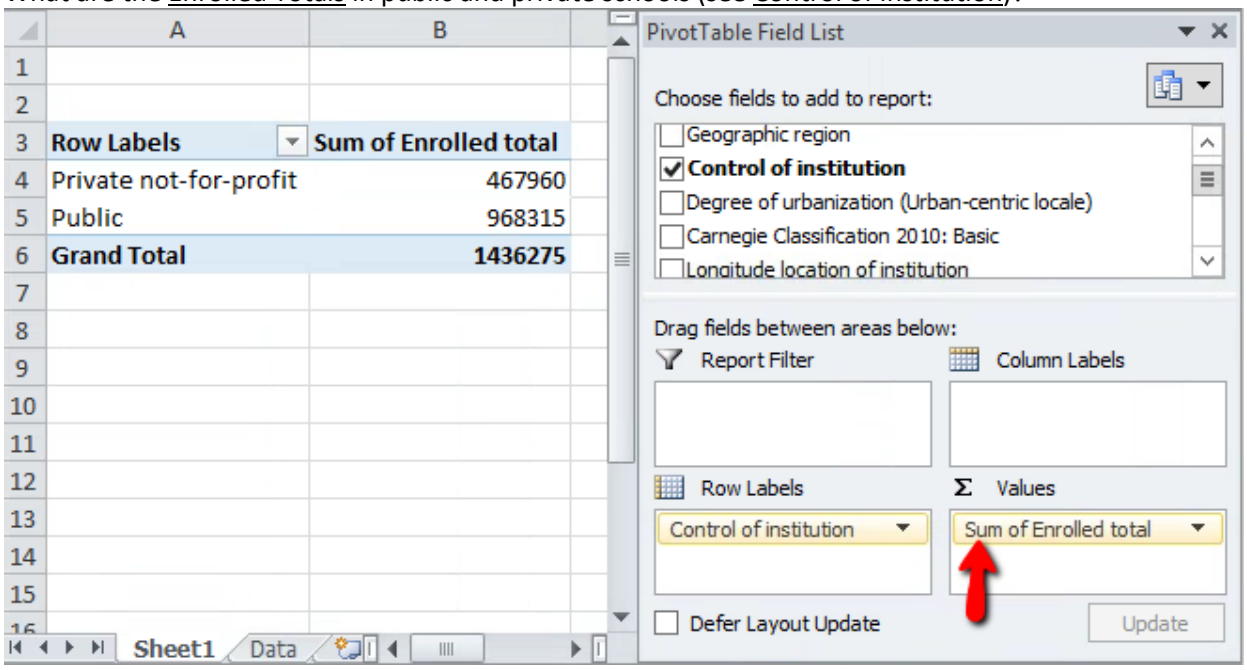
- Lynda.com provides training videos (free to UNC affiliates) on a wide variety of Excel functions
- Matt Jansen (guides.unc.edu/mattjansen) of the Davis Library Research Hub is available for one-on-one consultations.

PivotTable Exercise
SOLUTIONS BELOW

SOLUTIONS

DISCLAIMER: For many of these questions, there are multiple ways to get to the correct values. These are merely one way you could get the desired results.

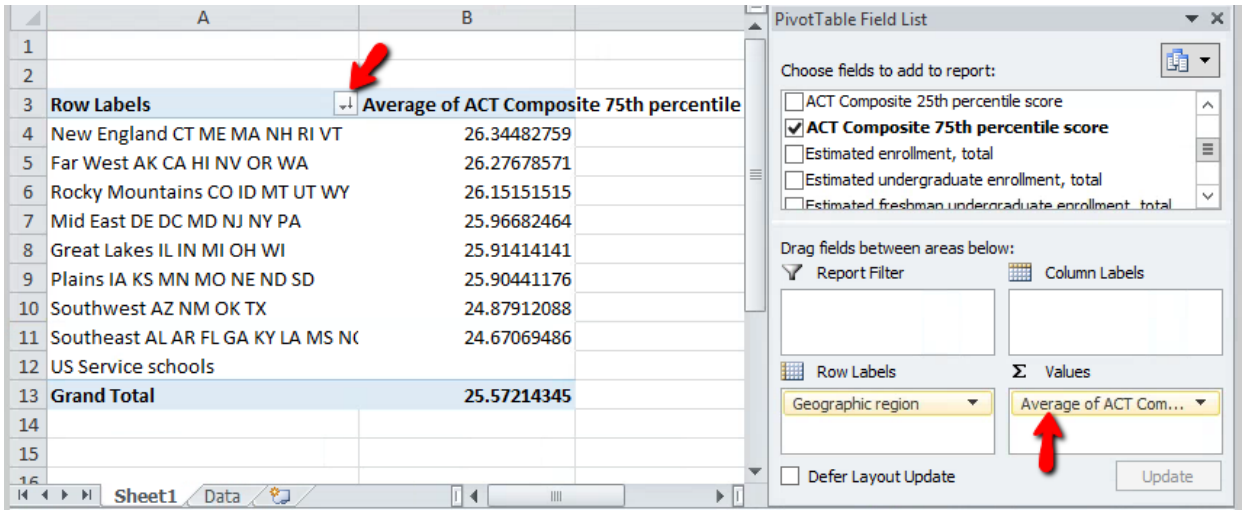
4. What are the Enrolled Totals in public and private schools (see Control of institution)?



Instructions:

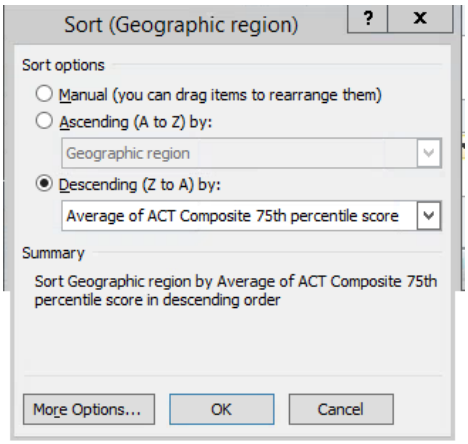
- Drag Control of institution to Row Labels to layout the table.
- Drag Enrolled total to Values.
- Click Enrolled total in Values and select Value Field Settings…
    - Change Summarize Values by to **Sum** instead of the default Count.

5. Which Geographic Region has the highest average ACT Composite 75th percentile score? How many regions have average scores below the national (Grand Total) average?

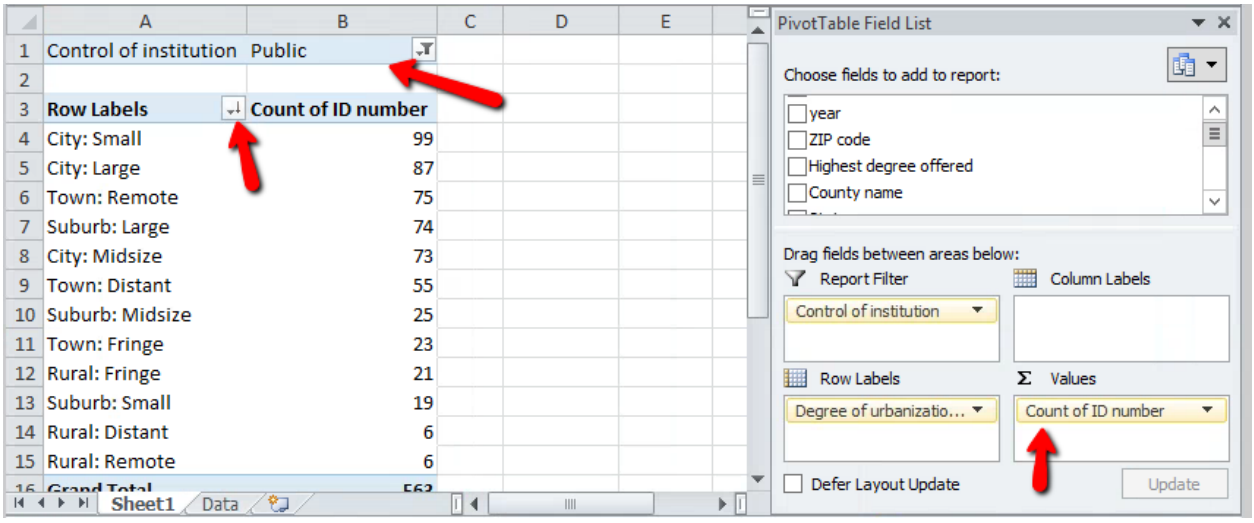


Instructions:

- Drag Geographic region to Row Labels.
- Drag ACT Composite 75th percentile score to Values.
- Click ACT Composite 75th percentile score in Values and select Value Field Settings…
    - Change Summarize Values by to **Average**.
- Sort the table by clicking the arrow icon in the corner of **Row Labels** on the table itself (see arrow)
    - Click More Sort Options, then select Descending (Z to A) and set to Average of ACT Composite 75th percentile score



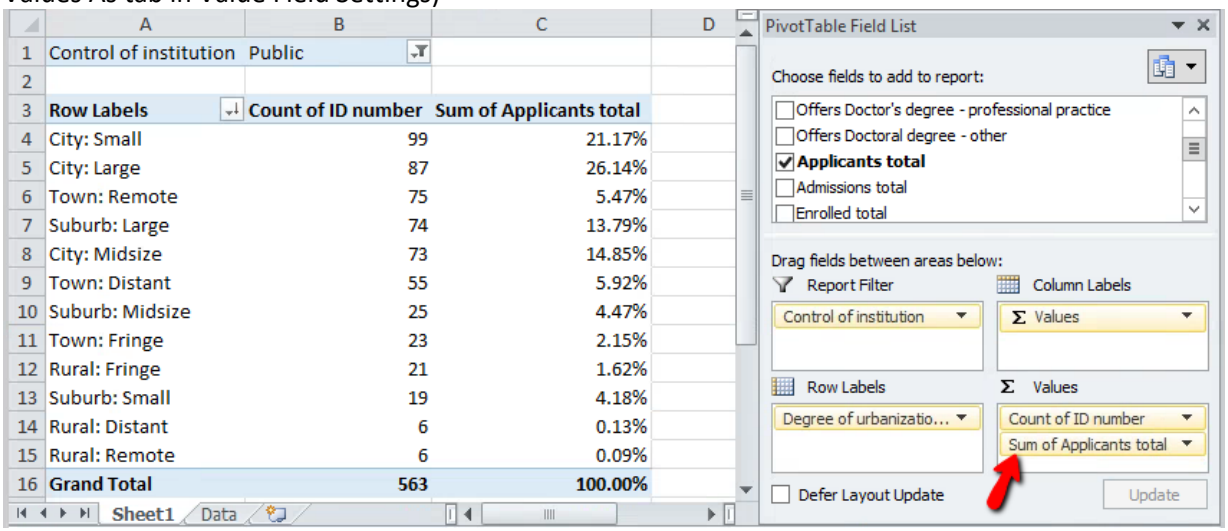
6. Which category of Degree of Urbanization contains the most public (Control of Institution) universities?



Instructions:

- Drag Degree of Urbanization to Row Labels. Drag ID number to Values.
- Click ID number in Values, then Value Field Settings>Summarize Values by to set aggregation to **Count.**
  - **Note:** ANY variable that doesn't have missing values will work here. The **Count** aggregation is essentially counting the number of filled cells in the given categories. So some values will have slightly different counts due to empty cells.
- Drag Control of Institution to Report Filter. Use the Filter appearing at the top of the screen to select **Public**.
- Use the Row Labels arrow, then More Sort Options to sort by Count of ID number.

a. What percent of the Applications total go to each of the top two Degrees of Urbanization? (Hint: Use the Show Values As tab in Value Field Settings)



Instructions

- Assume that we keep the Public filter active. Leave the earlier work in place.
- Drag Applicants total to Values and place it under Count of ID number.
- Click Applicants total and select Value Field Settings.
  - Under Summarize Values by, choose **Sum**
  - Under Show Values As, choose **% of Column Total**