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
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”

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Row Labels</td>
<td>Count of EmailID</td>
</tr>
<tr>
<td>4</td>
<td>FR</td>
<td>28.57%</td>
</tr>
<tr>
<td>5</td>
<td>JR</td>
<td>14.29%</td>
</tr>
<tr>
<td>6</td>
<td>SO</td>
<td>47.62%</td>
</tr>
<tr>
<td>7</td>
<td>SR</td>
<td>9.52%</td>
</tr>
<tr>
<td>8</td>
<td>Grand Total</td>
<td>100.00%</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

![Screenshot of Excel Pivot Table settings](image)
WORKING WITH PIVOT TABLES

• Rows and Columns: Fields to be inserted into either the Row or Column of the pivot table
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($B$2: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
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
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).
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