

## PUMFs and Pivot Tables: Using Excel to Create Cross-Tabulations from Public User Microdata Files

Excel pivot tables are an excellent tool for summarizing categorical and ordinal data. A pivot table is an interactive worksheet table where you can easily switch row and column headings to view the data in different ways. The resulting tables are very similar to the cross-tabulations that can be created using Beyond 20/20 files.

Excel 2003 and earlier versions have a limit of 256 columns and 65,536 rows which is a constraint when working with most Statistics Canada public user microdata files (PUMFs). However, by subsetting (records) or extracting specific variables most PUMFs can be handled by Excel 2003 and earlier versions. For this exercise we will be working with the public user microdata file produced from the **1996 Sun Exposure Survey**. This survey contains 4023 records and therefore falls well under the limit of Excel 2003 which will be the version used for this exercise.

**NOTE:** Excel 2007 can handle 16,384 columns and 1,048,576 rows.

In this exercise we will try to determine if there is any relationship between leisure sunscreen use and variables such as a person's gender, age, education level, and income. The definition of leisure sunscreen use covers those individuals that spend 30 minutes or more a day in the sun during June-August for leisure activities.

1. Open **sun.sav** with SPSS.
2. We will now do an extract of the following 6 variables from this file:  
**Q405F** – Frequency use sunscreen on face leisure  
**Q405G** – Frequency use sunscreen on body leisure  
**Q601** – Highest level of education completed  
**Q606** – Household income during last 12 months  
**DVAGEP8** – Age group of selected respondent  
**SEX** – Sex of selected respondent

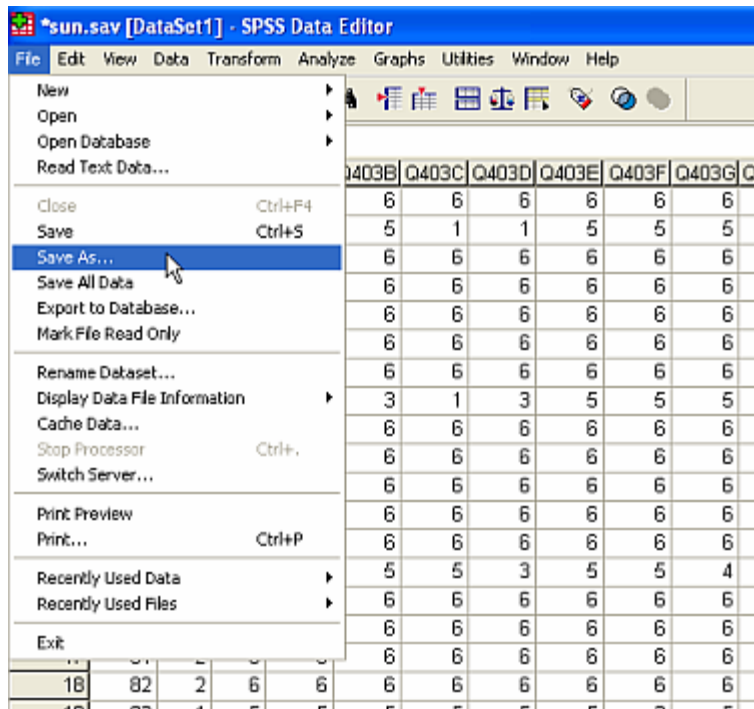
This can be accomplished either by using the following SPSS syntax and then saving the resulting extract as an .xls file:

```
Get file='path:\filename'  
/keep=Q405f, Q405g, Q601, Q606, DVAGEGP8, SEX.  
Save outfile='path:\filename'.  
Execute.
```

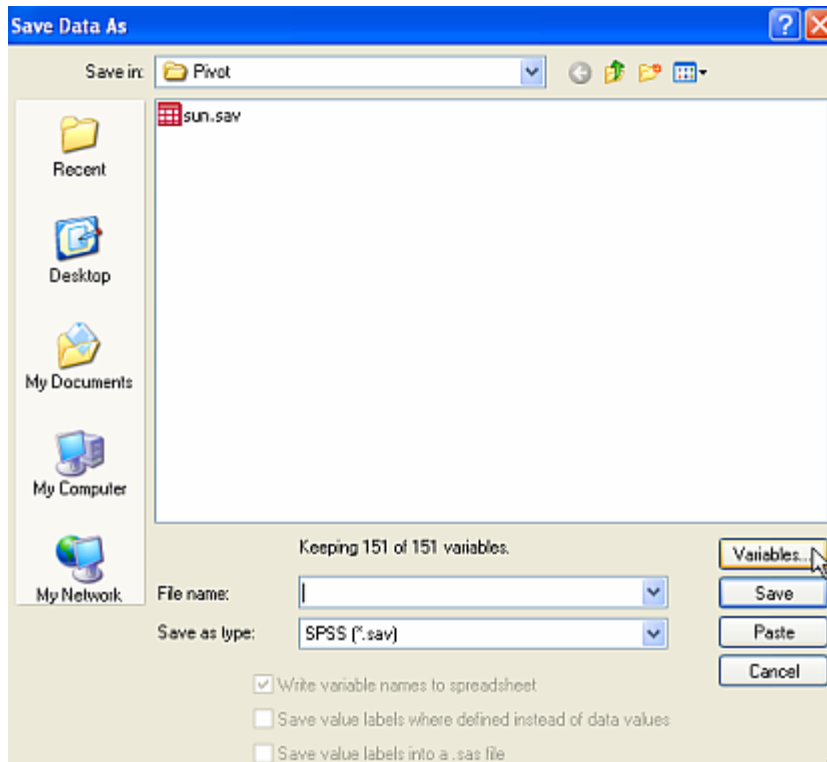
OR

By selecting the variables when you save your output as an .xls file. This is the option that we will use.

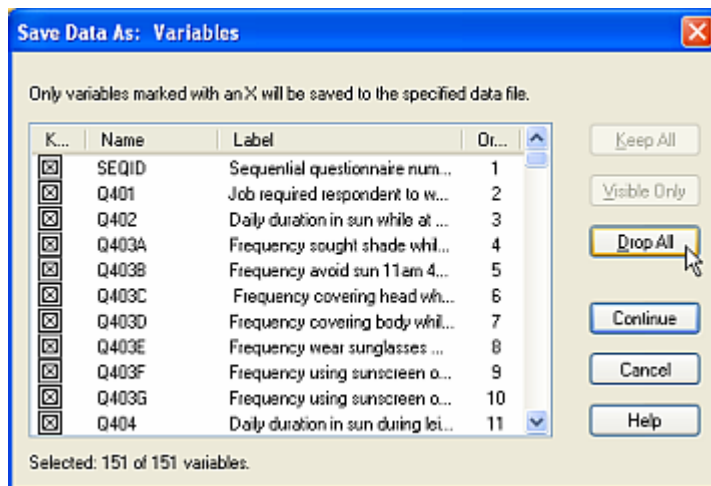
3. Select **File** and then **Save As...**



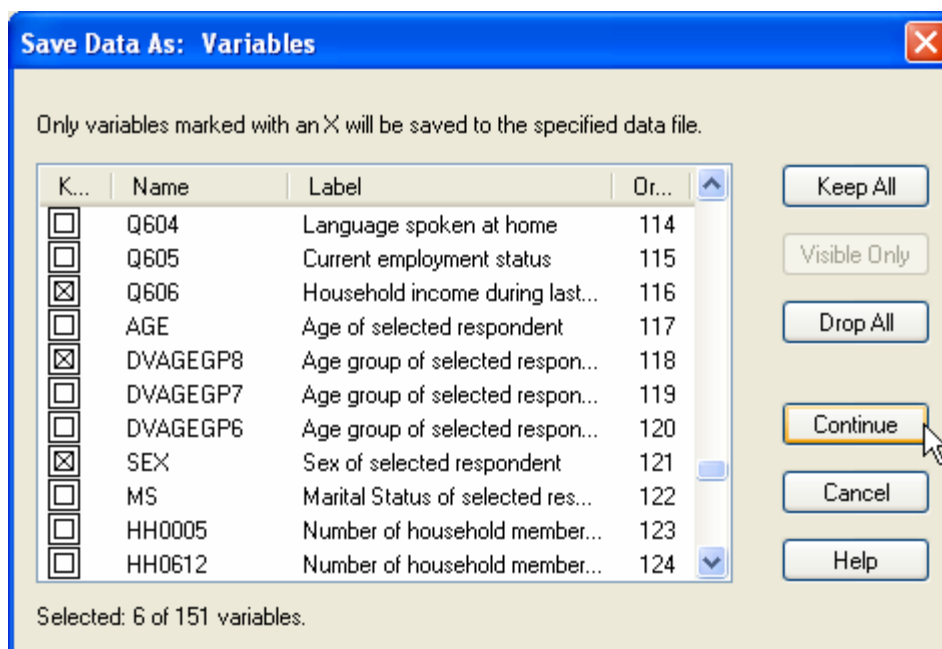
4. In the **Save Data As** box click on the **Variables...** button.



5. First drop all the variables with the **Drop All** button and then select the specific ones that you want by clicking in the box to the left of the variable.

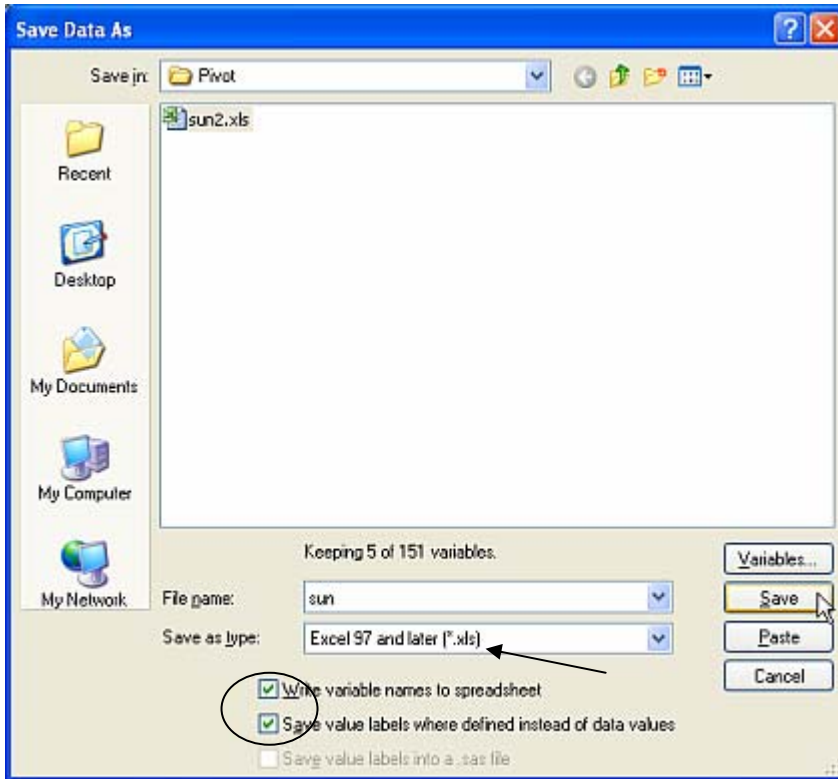


6. Select the following 6 variables: **Q405F**, **Q405G**, **Q601**, **Q606**, **DVAGEP8**, **SEX**. Click **Continue** to proceed.



7. On the next page do the following:

- **Save as type:** *Excel 97 and later (\*.xls)*
- Check off the box **Save value labels where defined instead of data values**
- Make sure that the box **Write variable names to spreadsheet** is checked
- Name the file **Sun.xls**
- Click on **Save**



8. Close **sun.sav**.

9. Open **sun.xls**.

10. Let's change the variable labels:

**Q405F** to **Sunscreen Face**

**Q405G** to **Sunscreen Body**

**Q601** to **Education**

**Q606** to **Household Income**

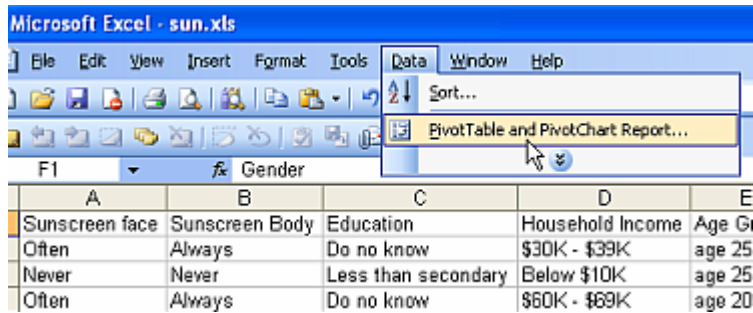
**DVAGEP8** to **Age Group**

**SEX** to **Gender**

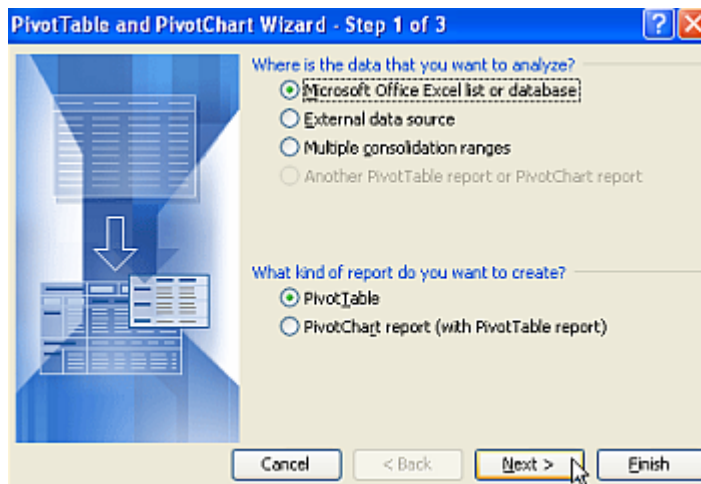
	A	B	C	D	E	F
1	Sunscreen face	Sunscreen Body	Education	Household Income	Age Group	Gender
2	Often	Always	Do no know	\$30K - \$39K	age 25-34	Female
3	Never	Never	Less than secondary	Below \$10K	age 25-34	Male
4	Often	Always	Do no know	\$60K - \$69K	age 20-24	Female
5	Valid skip	Valid skip	Do no know	\$50K - \$59K	age 55-64	Male
6	Often	Often	Complete secondary	Do not know	age 45-54	Female
7	Never	Never	Less than secondary	\$20K - \$29K	age 65-74	Male
8	Valid skip	Valid skip	Complete secondary	\$10K - \$19K	age 65-74	Female
9	Never	Never	Less than secondary	\$20K - \$29K	age 65-74	Male

Now let's create a pivot table:

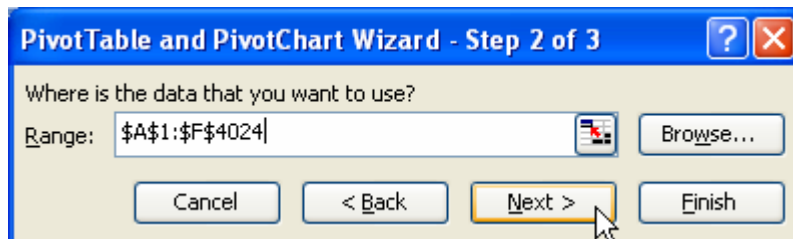
11. Select **Data** from the menu bar and then **PivotTable and PivotChart Report**.



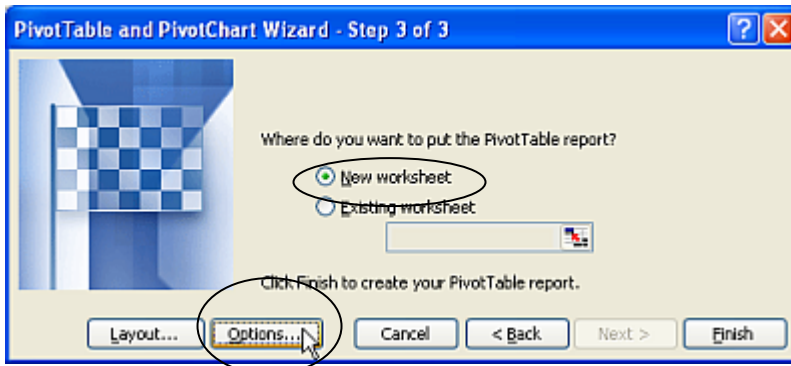
12. On the next screen (*Step 1 of the PivotTable Wizard*) accept the defaults and click **Next**.



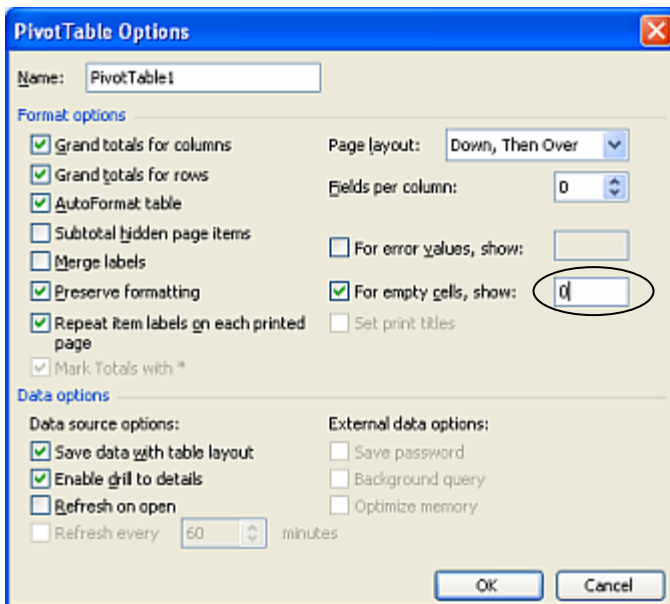
13. Step 2 prompts you for where the data is. Enter **\$A\$1:\$F\$4024** into the box and click **NEXT**. This is the entire data set including the variable labels. CAUTION: Excel may default the range to \$A\$2:\$F\$4024 which would not include the necessary variable labels.



14. On Step 3 bullet **New worksheet** and click on **Options....**



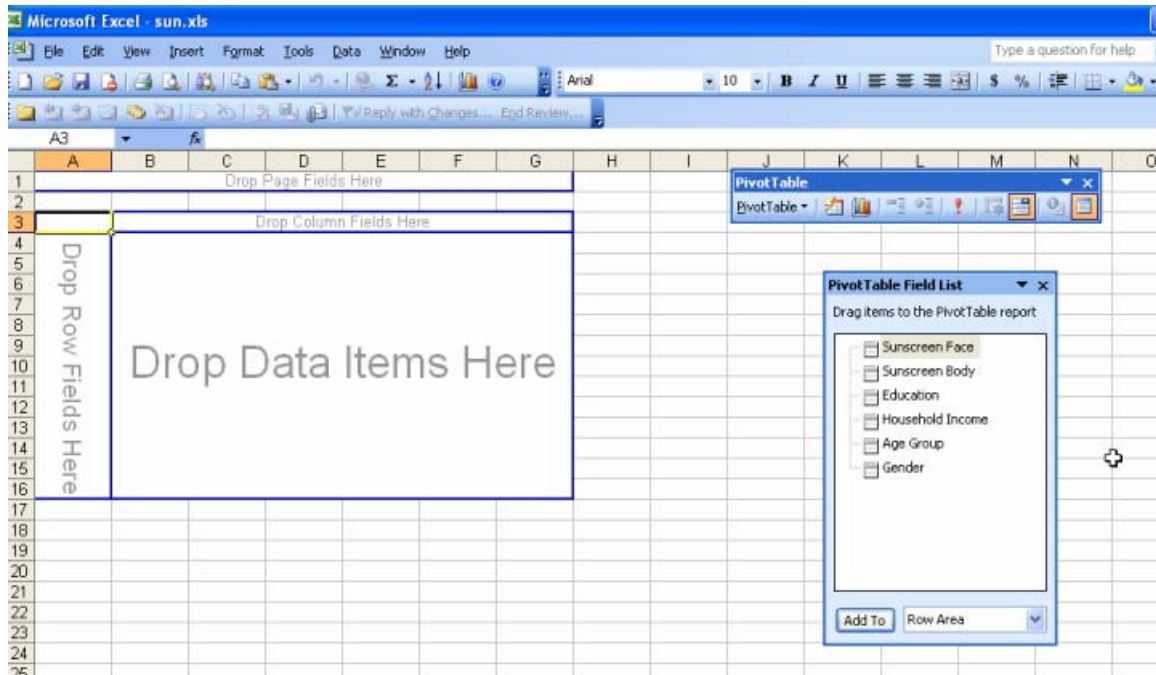
15. Enter **0** in the box beside **For empty cells, show:** and then click **OK**.



16. On the next screen (*Step 3 of 3*) click on **Finish**.



We are now ready to populate our pivot table. This is very much like building a B20/20 table.



Let's do a simple frequency first.

17. Click and drag the **Sunscreen Body** variable tile to where it says **Drop Row Fields Here**. Then click and drag the **Sunscreen Body** variable tile (from the *PivotTable Field List*) to where it says **Drop Data Items Here**. Your resulting table should look like this:

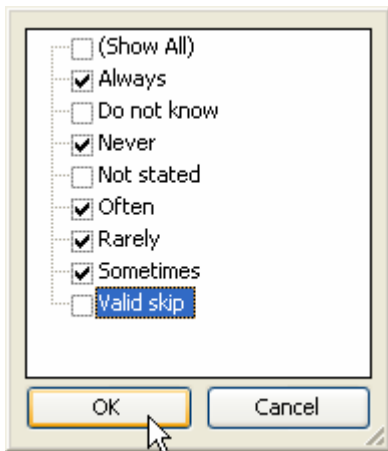
	A	B
1	Drop Page Fields Here	
2		
3	Count of Sunscreen Body	
4	Sunscreen Body	Total
5	Always	524
6	Do not know	1
7	Never	1119
8	Not stated	14
9	Often	490
10	Rarely	404
11	Sometimes	623
12	Valid skip	848
13	Grand Total	4023

Let's get rid of the missing values now for this variable.

18. Click on the **Sunscreen Body** drop-down box on the table.

1		
2		
3	Count of Sunscreen Body	
4	Sunscreen Body	Total
5	Always	524
6	Do not know	1
7	Never	1119
8	Not stated	14
9	Often	490
10	Rarely	404
11	Sometimes	623
12	Valid skip	848
13	Grand Total	4023

19. Deselect the **Do not know**, **Not stated**, **Valid skip** checkboxes in the list of value labels and click **OK**. Those value labels will disappear from the table.



Let's now put the values into a logical order from **Always** to **Never** – by default the system sorts them in alphabetical order.

20. To do the sort:

- Click on the value that you want to move
- Move the mouse pointer to the top or bottom edge of the cell until you get a doubleheaded arrow
- Click and drag the item to the position that you want. In the example below **Rarely** is being dragged down to **Sometimes**

	A	B
1	Drop Page Fields Here	
2		
3	Count of Sunscreen Body	
4	Sunscreen Body	Total
5	Always	524
6	Often	490
7	Rarely	404
8	Sometimes	623
9	Never	1119
10	Grand Total	3160

	A	B
1	Drop Page Fields Here	
2		
3	Count of Sunscreen Body	
4	Sunscreen Body	Total
5	Always	524
6	Often	490
7	Sometimes	623
8	Rarely	404
9	Never	1119
10	Grand Total	3160

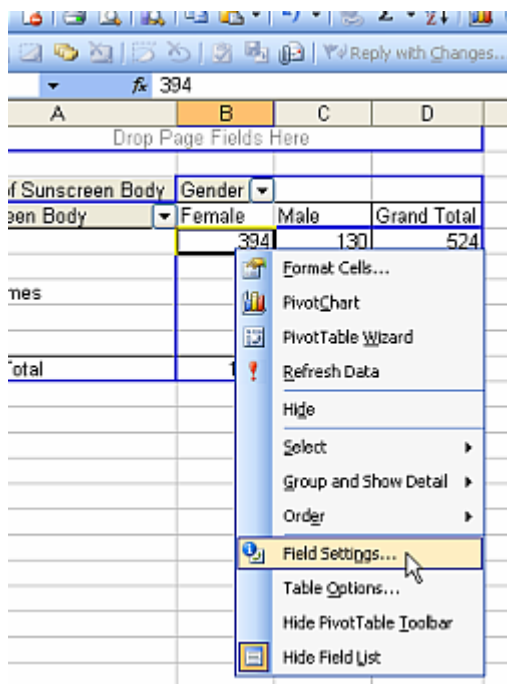
Now let's create a two-way table showing **Gender** cross-tabulated with **Sunscreen Body**.

21. To do this, drag the **Gender** variable tile to the column heading to the right of **Sunscreen Body**. You should now have the following two-way table:

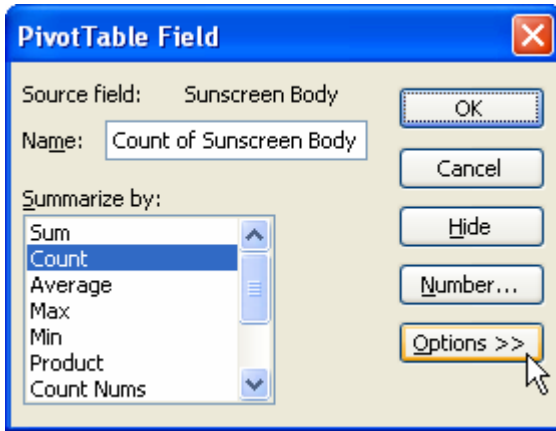
	A	B	C	D
1	Drop Page Fields Here			
2				
3	Count of Sunscreen Body	Gender		
4	Sunscreen Body	Female	Male	Grand Total
5	Always	394	130	524
6	Often	300	190	490
7	Sometimes	327	296	623
8	Rarely	156	248	404
9	Never	375	744	1119
10	Grand Total	1552	1608	3160
11				

Now let's calculate a percentage within each gender – this would require the data to be expressed as a column %.

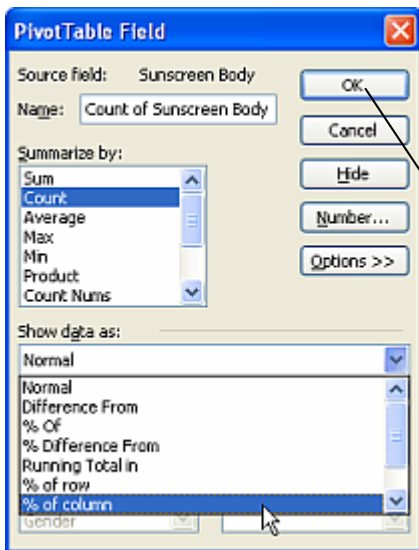
22. Right click on a data value in the table and from the drop down select **Field Settings...**



23. From the **PivotTable Field** box select **Options >>**.



24. Click on the drop-down button **Show data as:** and change **Normal** to **% of column**. Click on **OK**. Your table values will change from counts to column percentages.

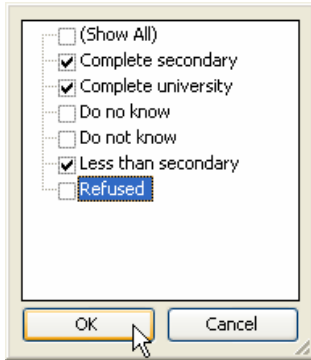


	A	B	C	D
1	Drop Page Fields Here			
2				
3	Count of Sunscreen Body	Gender		
4	Sunscreen Body	Female	Male	Grand Total
5	Always	25.39%	8.08%	16.58%
6	Often	19.33%	11.82%	15.51%
7	Sometimes	21.07%	18.41%	19.72%
8	Rarely	10.05%	15.42%	12.78%
9	Never	24.16%	46.27%	35.41%
10	Grand Total	100.00%	100.00%	100.00%
11				

**Question:** From this table can you tell if there is any difference in sunscreen usage between males and females?

Now let's see if there is any relationship between sunscreen use and educational attainment.

25. Move the **Gender** variable tile back to the **PivotTable Field List** and replace it with the **Education** variable tile. Let's get rid of the **Do not know**, **Do no know**, and **Refused** value labels. To accomplish this click on the **Education** drop-down and uncheck these values.



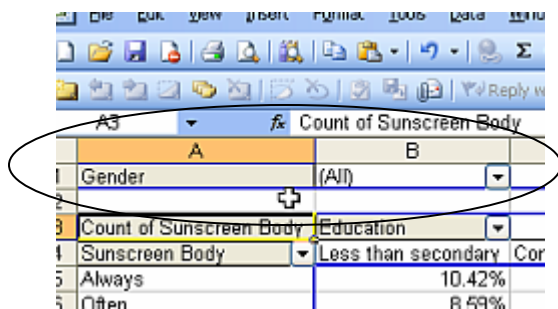
26. Resort the remaining value labels in order from **Less than secondary** to **Complete university**. Your table should now look like this:

Sunscreen Body	Less than secondary	Complete secondary	Complete university	Grand Total
Always	10.42%	17.67%	19.71%	16.03%
Often	8.59%	14.62%	19.35%	13.94%
Sometimes	14.85%	20.12%	21.70%	18.89%
Rarely	11.97%	13.02%	13.38%	12.80%
Never	54.37%	34.57%	25.86%	38.35%
Grand Total	100.00%	100.00%	100.00%	100.00%

**Question:** Is there a difference in sunscreen usage when educational attainment is looked at?

Let's leave this table as it is and consider whether changing **Gender** from **All** to **Female** would have an impact on this table.

27. To do this click and drag the **Gender** variable tile to the top line of the table where it says **Drop Page Fields Here**.



28. Click on the drop-down beside **All**, select **Female** and then click **OK**.

Gender	Female		
Count of Sunscreen Body	Education		
Sunscreen Body	Less than secondary	Complete secondary	Complete university
Always	17.70%	28.04%	26.64%
Often	12.11%	17.91%	21.45%
Sometimes	16.15%	22.47%	21.11%
Rarely	13.98%	7.60%	12.11%
Never	40.06%	23.99%	18.69%
Grand Total	100.00%	100.00%	100.00%

Let's change the display of the **Sunscreen Body** data to row percentages to see the percentage difference within each sunscreen usage category.

29. Right click on any of the data cells and select **Field Settings...**

Gender	Female		
Count of Sunscreen Body	Education		
Sunscreen Body	Less than secondary	Complete second	
Always	17.70%	28.04%	
Often			
Sometimes			
Rarely			
Never			
Grand Total			

30. In the **PivotTable Field** box change **Show data as:** to **% of row** and click **OK**.

PivotTable Field

Source field: Sunscreen Body

Name: Count of Sunscreen Body

Summarize by: Count

Show data as: % of row

Our new table looks like this now:

A	B	C	D	E
Gender	Female			
Count of Sunscreen Body	Education			
Sunscreen Body	Less than secondary	Complete secondary	Complete university	Grand Total
Always	19.00%	55.33%	25.67%	100.00%
Often	18.84%	51.21%	29.95%	100.00%
Sometimes	21.14%	54.07%	24.80%	100.00%
Rarely	36.00%	36.00%	28.00%	100.00%
Never	39.69%	43.69%	16.62%	100.00%
Grand Total	26.77%	49.21%	24.02%	100.00%

Now let's get really fancy and nest the variable **Age Group** with the **Sunscreen Body** variable.

31. Click on the **Age Group** variable tile and then drag it over to the **Sunscreen Body** variable tile and drop it there. Your table should now look like this:

B4		fx Age Group				
	A	B	C	D	E	F
1	Gender	Female				
2						
3	Count of Sunscreen Body	Education				
4	Sunscreen Body	Age Group	Less than secondary	Complete secondary	Complete university	Grand Total
5	Always	age 15-19	64.52%	35.48%	0.00%	100.00%
6		age 20-24	10.00%	55.00%	35.00%	100.00%
7		age 25-34	8.64%	54.32%	37.04%	100.00%
8		age 35-44	12.82%	57.89%	29.49%	100.00%
9		age 45-54	10.00%	57.50%	32.50%	100.00%
10		age 55-64	25.93%	59.26%	14.81%	100.00%
11		age 65-74	27.78%	72.22%	0.00%	100.00%
12		age 75 plus	40.00%	60.00%	0.00%	100.00%
13	Always Total		19.00%	55.33%	25.67%	100.00%
14	Often	age 15-19	56.00%	40.00%	4.00%	100.00%
15		age 20-24	13.33%	66.67%	20.00%	100.00%
16		age 25-34	12.28%	50.88%	36.84%	100.00%
17		age 35-44	2.27%	50.00%	47.73%	100.00%
18		age 45-54	22.22%	58.33%	19.44%	100.00%
19		age 55-64	33.33%	63.33%	13.33%	100.00%

Now try this one on your own. You will need to remove all the current variable tiles by dragging them back to the **PivotTable Field List**.

**Optional Pivot Table Exercise:**

Create a table with **Sunscreen Face** as the column variable, **Household Income** as the row variable, **Sunscreen Face** as the data and **Gender** as the *page field*. Drop the missing value labels (*do not know, refused, not stated, etc.*) for both the **Household Income** and **Sunscreen Face** variables; resort the **Household Income** value labels from **No income** to **\$80K+** in increasing \$ value; resort the **Sunscreen Face** value labels from **Always** to **Never** in decreasing frequency; change the **Gender** from **All** to **Female**; and change the data from counts to row percentages. Your final product should like something like this table:

Household Income	Always	Often	Sometimes	Rarely	Never	Grand Total
No income	14.29%	14.29%	28.57%	28.57%	14.29%	100.00%
Below \$10K	16.47%	12.94%	18.82%	8.24%	43.53%	100.00%
\$10K - \$19K	34.72%	12.50%	13.89%	6.48%	32.41%	100.00%
\$20K - \$29K	36.77%	19.73%	11.66%	12.11%	19.73%	100.00%
\$30K - \$39K	37.27%	15.00%	20.45%	6.36%	20.91%	100.00%
\$40K - \$49K	43.48%	13.66%	10.56%	10.56%	21.74%	100.00%
\$50K - \$59K	43.80%	17.36%	14.05%	9.09%	15.70%	100.00%
\$60K - \$69K	37.76%	24.49%	13.27%	7.14%	17.35%	100.00%
\$70K - \$79K	48.33%	18.33%	10.00%	10.00%	13.33%	100.00%
\$80K+	49.67%	14.57%	17.88%	9.27%	8.61%	100.00%
Grand Total	38.60%	16.10%	14.83%	8.87%	21.61%	100.00%

**Question:** Does income influence sunscreen usage?

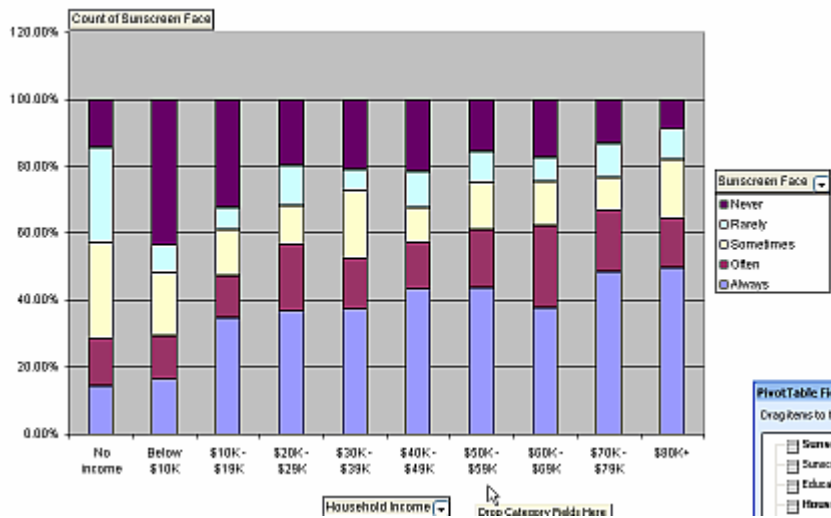
This concludes our tour of the Excel pivot table.

## Optional Pivot Chart Exercise:

Create a bar chart with the data from the table just created in the optional pivot table exercise

Right click on any data cell and select **PivotChart**.

	A	B	C
1			
2	Gender	Female	
3			
4	Count of Sunscreen Face	Sunscreen Face	
5	Household Income	Always	Often
6	No income	14.29%	14.29%
7	Below \$10K	16.47%	12.94%
8	\$10K - \$19K	34.1	
9	\$20K - \$29K	36.1	
10	\$30K - \$39K	37.1	
11	\$40K - \$49K	43.1	
12	\$50K - \$59K	43.1	
13	\$60K - \$69K	37.1	
14	\$70K - \$79K	48.1	
15	\$80K+	49.1	
16	Grand Total	38.1	
17			



As demonstrated above, you can quickly display your pivot table data in a pivot chart. You can also easily modify your pivot chart in much the same fashion as we did with pivot tables. The bar graphs and pie charts produced with pivot charts are an effective tool for displaying categorical and ordinal data.