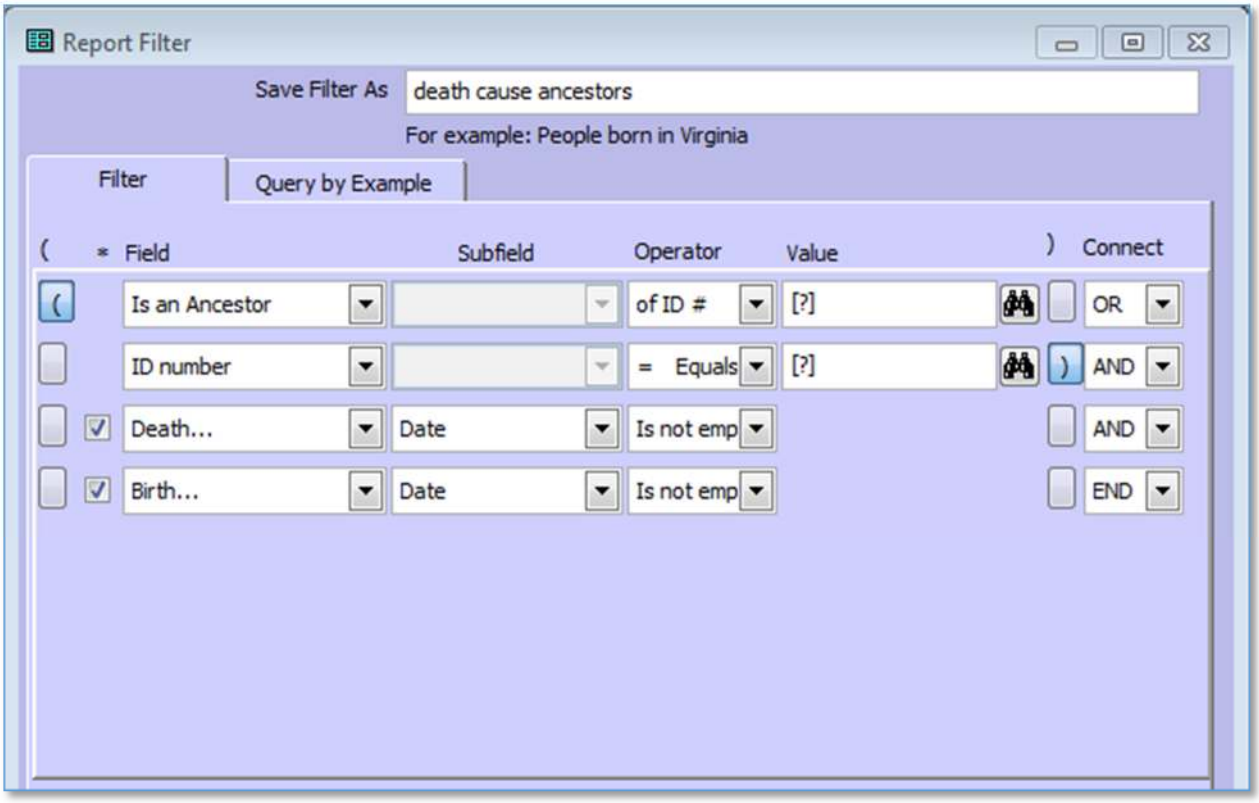
Death Statistics, Charts, Graphs

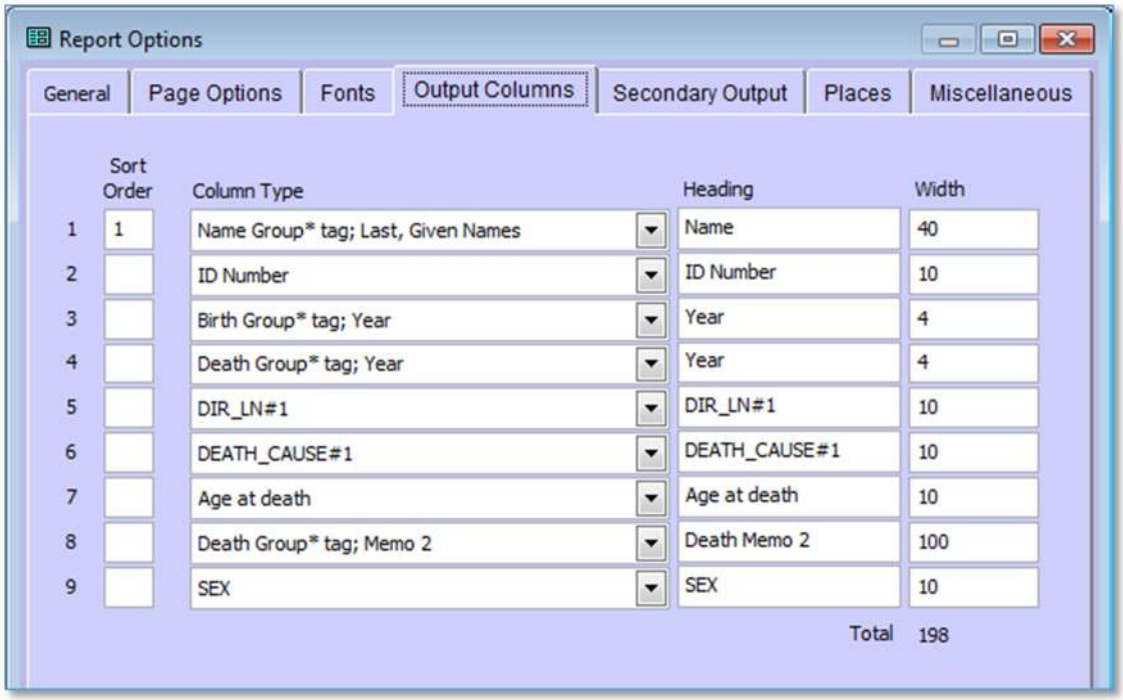
A few years ago, genealogy bloggers posted charts they created of ancestral causes of death. Several bloggers detailed the hours it took to create those charts, but I suspected that the TMG-Excel combination could create equivalent charts in moments. Yes, these charts took less than five minutes to create.

The Cause of Death Report

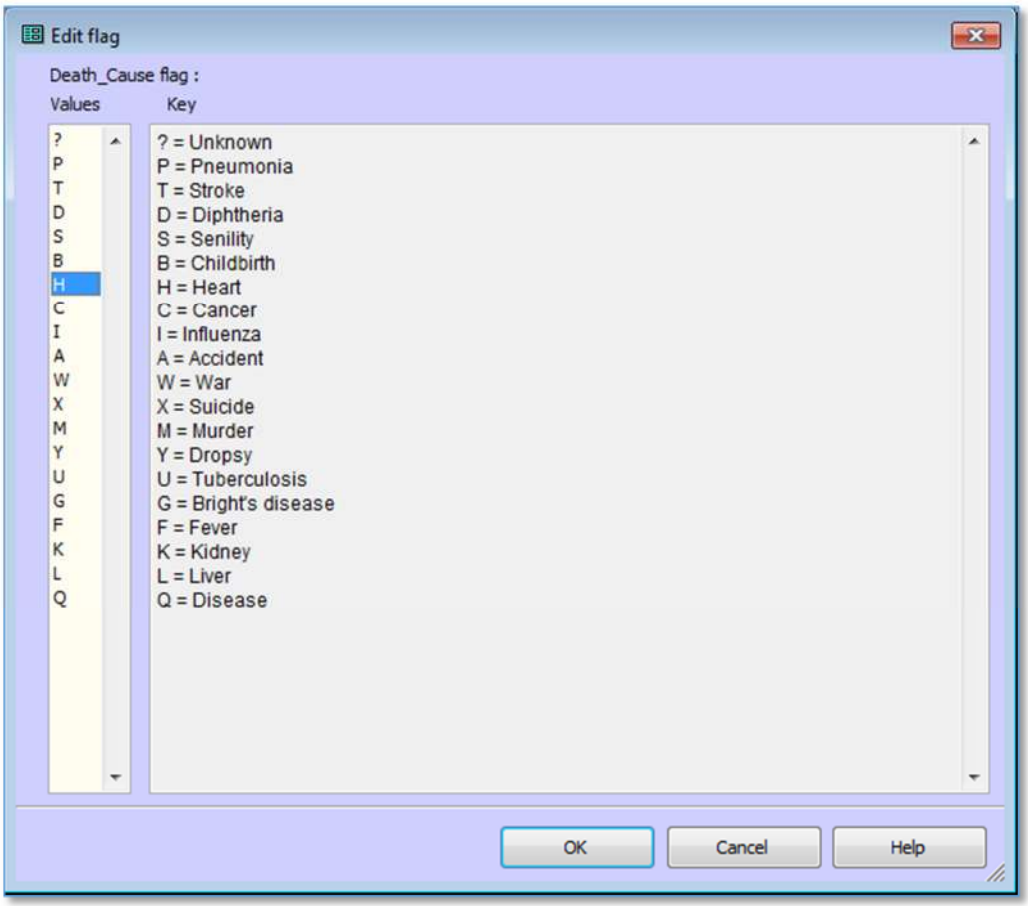
The illustrated report filters for ancestral lines only. An informative variation could compare descendants of defined progenitors. Because I wanted to view statistics on age at death, the filter includes two lines mandating non-empty birth and death dates.



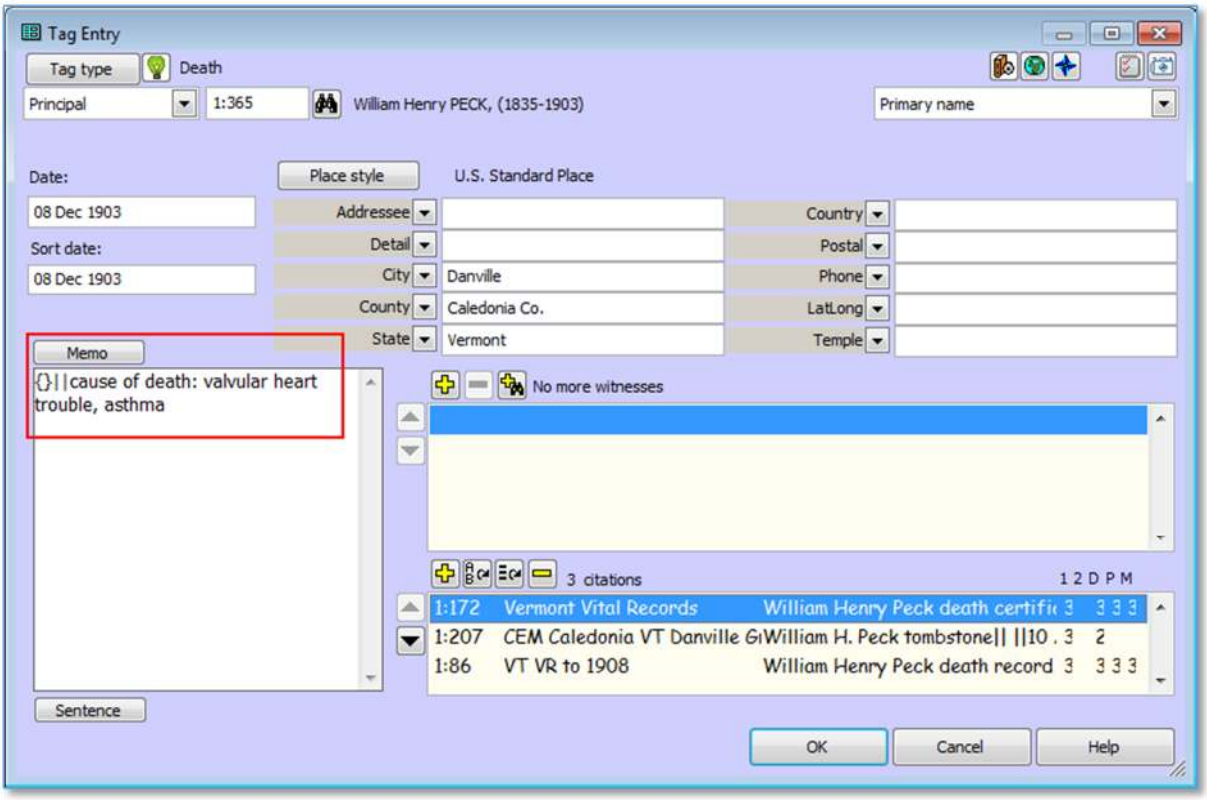
The output columns are defined to include data on the direct lines (flag), the user-defined cause of death (flag), the gender (flag), and the age at death. My data entry protocol places the official cause of death in the Death tag's Memo2, and that datum is included in this report. The person's name and ID number, as well as the years of birth and death, are also included. Technically, none of these are required for a statistical report, but they do add information to the Excel list. (Side note: TMG currently allows only single-value flags. If multi-value flags are allowed in HRE, one could use one of the *International Classifications of Diseases* for the flag values, creating more detailed and informative charts.)



My cause of death flag values were defined gradually over time. A more coherently defined list might be beneficial, but I haven't taken the time to develop it. Flag values are consistent, so it's easier to create a chart from these, rather than from the more detailed descriptions found on the death certificate.

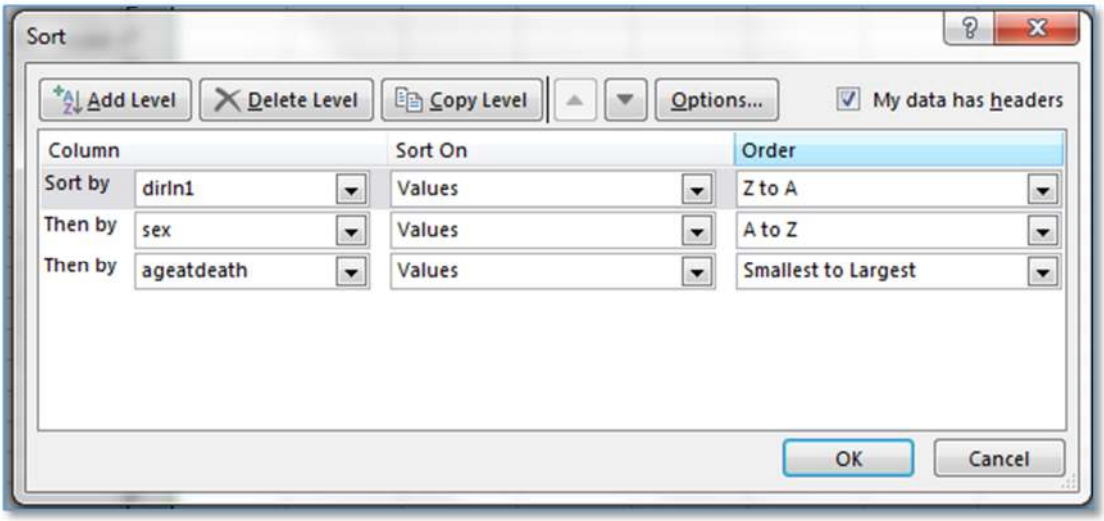


The death certificate's cause of death is entered in the death tag's Memo2, shown below. William Henry Peck's death had two equal causes, valvular heart trouble and asthma. Only the heart condition is indicated by the cause of death flag value (H).

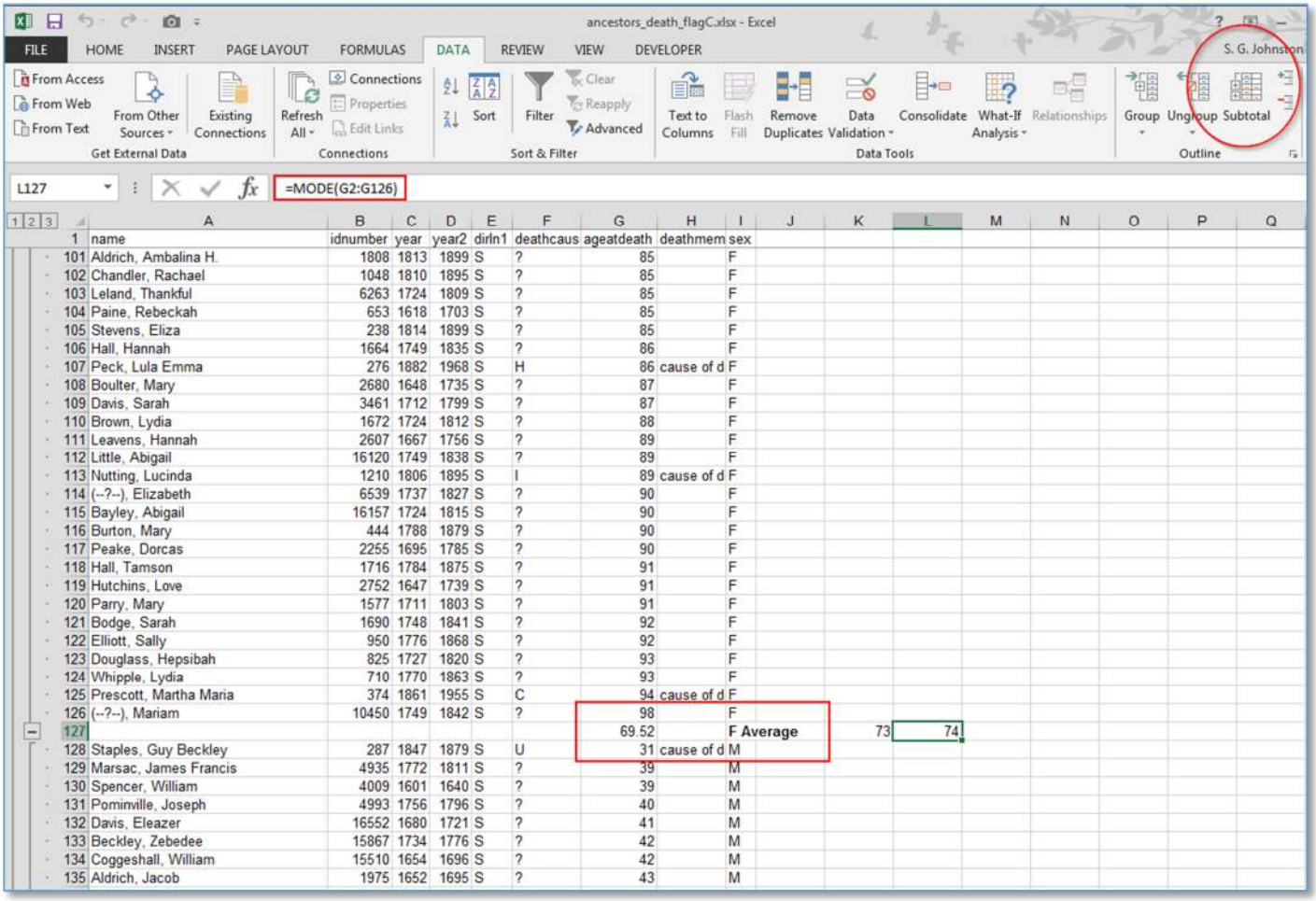


Determining the Average Age at Death and Other Statistics

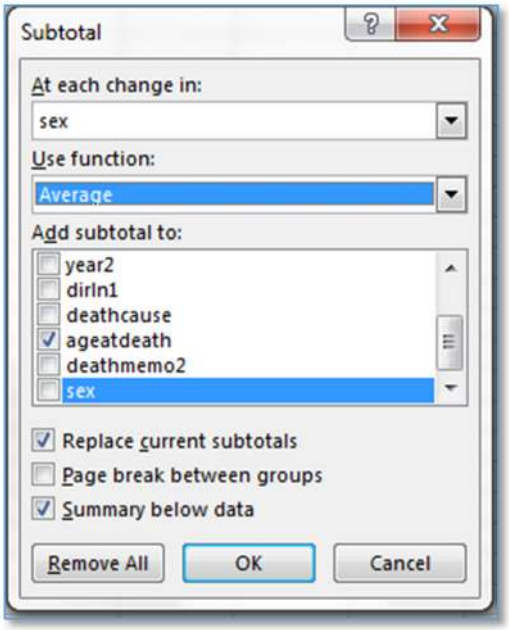
One of the first steps in analyzing spreadsheet data is sorting it into logical groups. I wanted to compare mortality statistics between my children's two direct lines and between genders within those two lines. I wanted to know (1) youngest and oldest ages at death, (2) average age at death, (3) median age at death, and (4) the mode age at death. The spreadsheet was sorted as shown.



One can insert calculation rows to calculate these statistics, but the **Subtotal** function, if available in your spreadsheet program, is easy and fun to use. In Excel 2013, this function is found under the **Data** tab. Row 127 shows the calculations for the females in the S direct line. The average age at death is calculated by the Subtotal function, and its settings are shown below the report screenshot. The median and mode ages at death are calculated by formula. The minimum and maximum ages at death can be included via formulas, but for a relatively short list, eyeballing the first and last entries is simple.

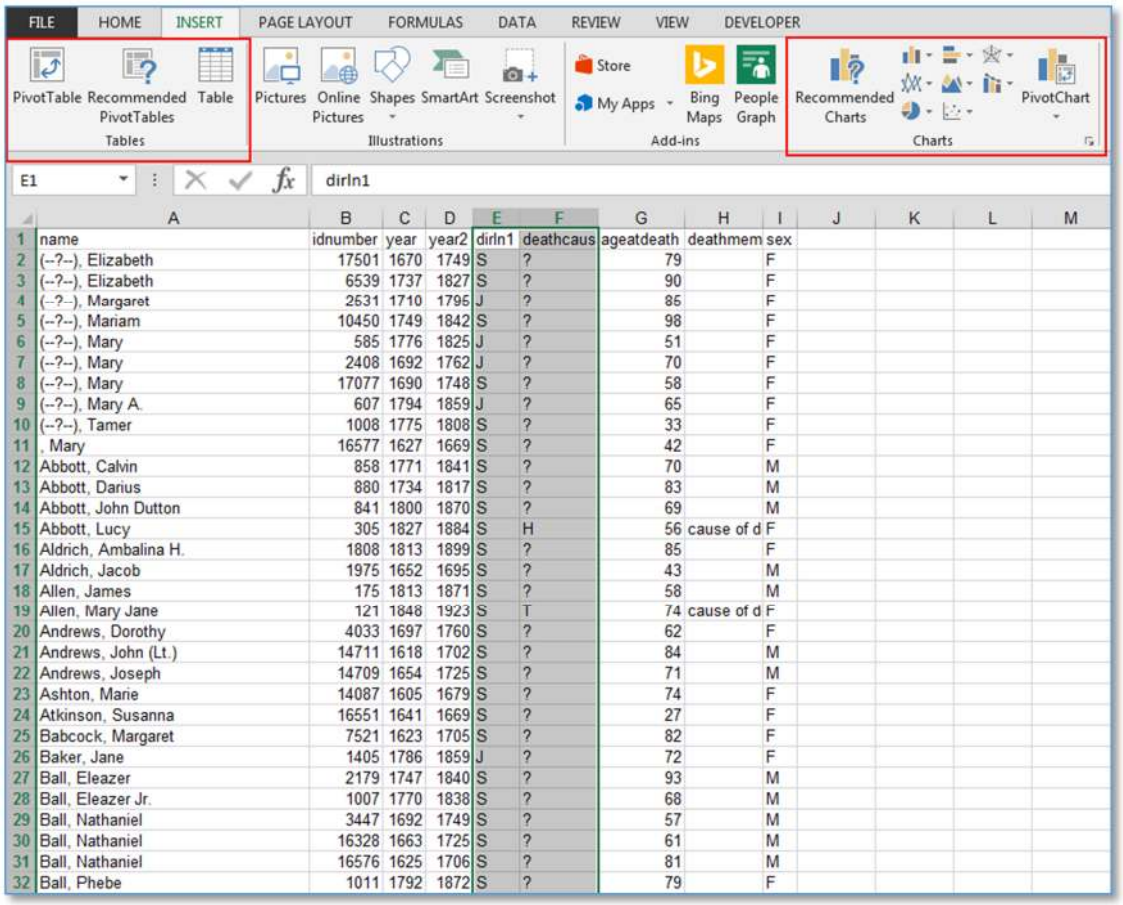


	A	B	C	D	E	F	G	H	I	J	K	L
1	name	idnumber	year	year2	dirln1	deathcaus	ageatdeath	deathmem	sex			
101	Aldrich, Ambalina H.	1808	1813	1899	S	?	85		F			
102	Chandler, Rachael	1048	1810	1895	S	?	85		F			
103	Leland, Thankful	6263	1724	1809	S	?	85		F			
104	Paine, Rebeckah	653	1618	1703	S	?	85		F			
105	Stevens, Eliza	238	1814	1899	S	?	85		F			
106	Hall, Hannah	1664	1749	1835	S	?	86		F			
107	Peck, Lula Emma	276	1882	1968	S	H	86	cause of d	F			
108	Boulter, Mary	2680	1648	1735	S	?	87		F			
109	Davis, Sarah	3461	1712	1799	S	?	87		F			
110	Brown, Lydia	1672	1724	1812	S	?	88		F			
111	Leavens, Hannah	2607	1667	1756	S	?	89		F			
112	Little, Abigail	16120	1749	1838	S	?	89		F			
113	Nutting, Lucinda	1210	1806	1895	S	I	89	cause of d	F			
114	(--?--), Elizabeth	6539	1737	1827	S	?	90		F			
115	Bayley, Abigail	16157	1724	1815	S	?	90		F			
116	Burton, Mary	444	1788	1879	S	?	90		F			
117	Peake, Dorcas	2255	1695	1785	S	?	90		F			
118	Hall, Tamson	1716	1784	1875	S	?	91		F			
119	Hutchins, Love	2752	1647	1739	S	?	91		F			
120	Parry, Mary	1577	1711	1803	S	?	91		F			
121	Bodge, Sarah	1690	1748	1841	S	?	92		F			
122	Elliott, Sally	950	1776	1868	S	?	92		F			
123	Douglass, Hepsibah	825	1727	1820	S	?	93		F			
124	Whipple, Lydia	710	1770	1863	S	?	93		F			
125	Prescott, Martha Maria	374	1861	1955	S	C	94	cause of d	F			
126	(--?--), Mariam	10450	1749	1842	S	?	98		F			
127							69.52		F	Average	73	74
128	Staples, Guy Beckley	287	1847	1879	S	U	31	cause of d	M			
129	Marsac, James Francis	4935	1772	1811	S	?	39		M			
130	Spencer, William	4009	1601	1640	S	?	39		M			
131	Pominville, Joseph	4993	1756	1796	S	?	40		M			
132	Davis, Eleazer	16552	1680	1721	S	?	41		M			
133	Beckley, Zebedee	15867	1734	1776	S	?	42		M			
134	Coggeshall, William	15510	1654	1696	S	?	42		M			
135	Aldrich, Jacob	1975	1652	1695	S	?	43		M			



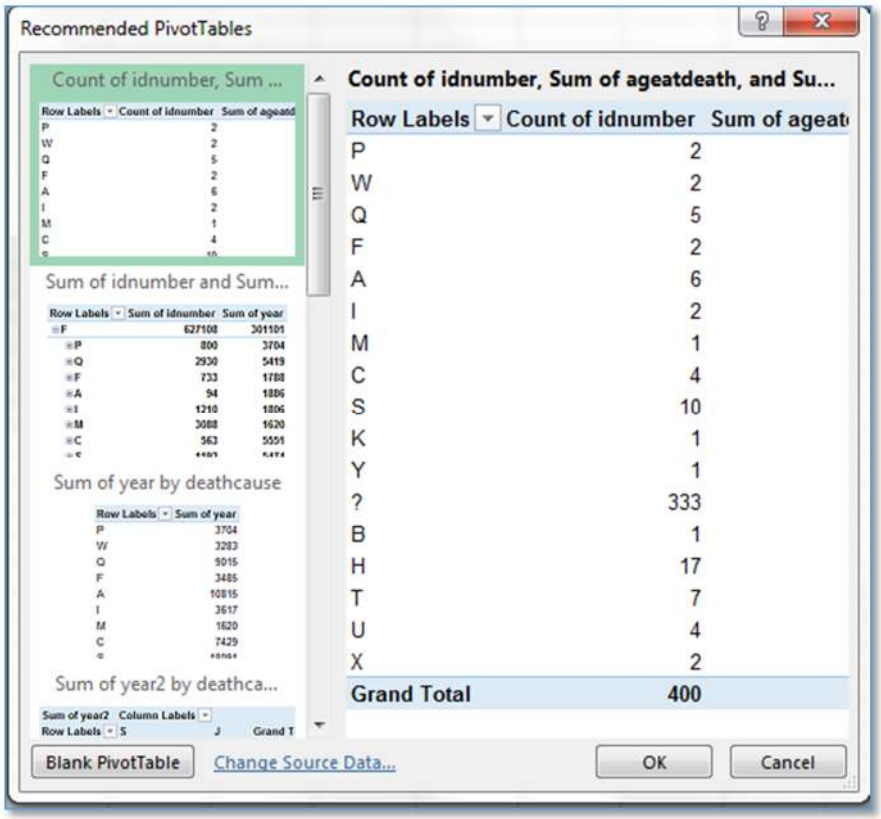
Creating Charts

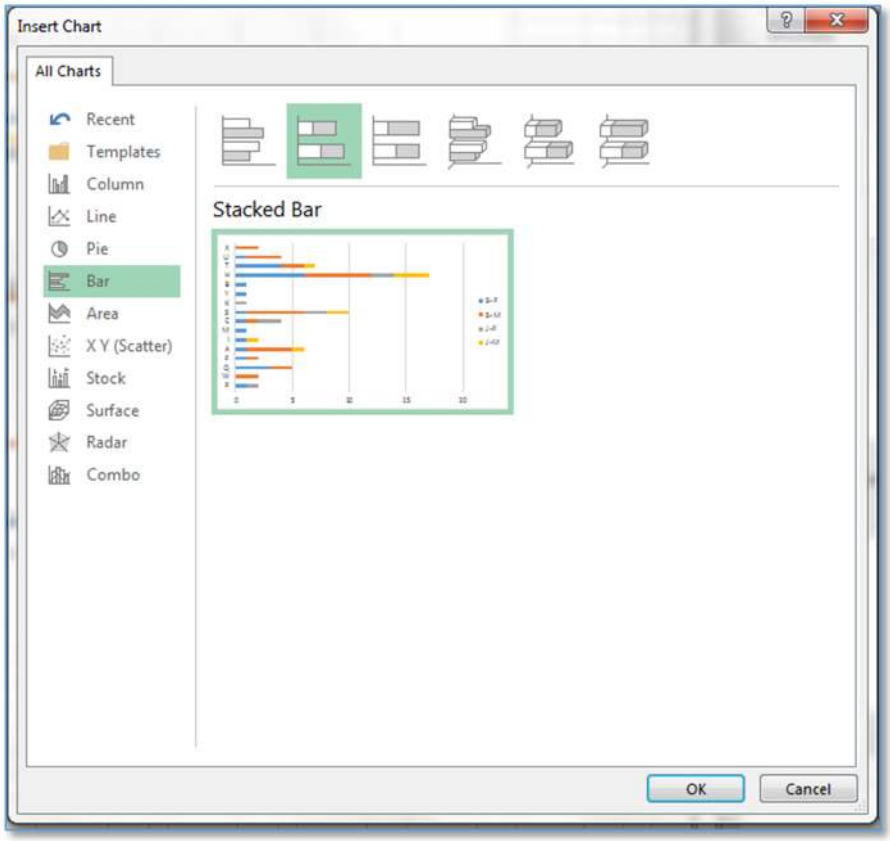
Charts are fun to create and make beautiful and informative illustrations. The **Insert** tab contains two areas relevant to charts: the *Pivot Table* area and the *Charts* area. Both areas include "Recommended" functions. Begin by selecting the data you want in your chart, and play with the available options until you get the chart you want.



	A	B	C	D	E	F	G	H	I
1	name	idnumber	year	year2	dirln1	deathcaus	ageatdeath	deathmem	sex
2	(--?--), Elizabeth	17501	1670	1749	S	?	79		F
3	(--?--), Elizabeth	6539	1737	1827	S	?	90		F
4	(--?--), Margaret	2631	1710	1795	J	?	86		F
5	(--?--), Mariam	10450	1749	1842	S	?	98		F
6	(--?--), Mary	585	1776	1825	J	?	51		F
7	(--?--), Mary	2408	1692	1762	J	?	70		F
8	(--?--), Mary	17077	1690	1748	S	?	58		F
9	(--?--), Mary A.	607	1794	1859	J	?	65		F
10	(--?--), Tamer	1008	1775	1808	S	?	33		F
11	, Mary	16577	1627	1669	S	?	42		F
12	Abbott, Calvin	858	1771	1841	S	?	70		M
13	Abbott, Darius	880	1734	1817	S	?	83		M
14	Abbott, John Dutton	841	1800	1870	S	?	69		M
15	Abbott, Lucy	305	1827	1884	S	H	56	cause of d	F
16	Aldrich, Ambalina H.	1808	1813	1899	S	?	85		F
17	Aldrich, Jacob	1975	1652	1695	S	?	43		M
18	Allen, James	175	1813	1871	S	?	58		M
19	Allen, Mary Jane	121	1848	1923	S	T	74	cause of d	F
20	Andrews, Dorothy	4033	1697	1760	S	?	62		F
21	Andrews, John (Lt.)	14711	1618	1702	S	?	84		M
22	Andrews, Joseph	14709	1654	1725	S	?	71		M
23	Ashton, Marie	14087	1605	1679	S	?	74		F
24	Atkinson, Susanna	16551	1641	1669	S	?	27		F
25	Babcock, Margaret	7521	1623	1705	S	?	82		F
26	Baker, Jane	1405	1786	1859	J	?	72		F
27	Ball, Eleazer	2179	1747	1840	S	?	93		M
28	Ball, Eleazer Jr.	1007	1770	1838	S	?	68		M
29	Ball, Nathaniel	3447	1692	1749	S	?	57		M
30	Ball, Nathaniel	16328	1663	1725	S	?	61		M
31	Ball, Nathaniel	16576	1625	1706	S	?	81		M
32	Ball, Phebe	1011	1792	1872	S	?	79		F

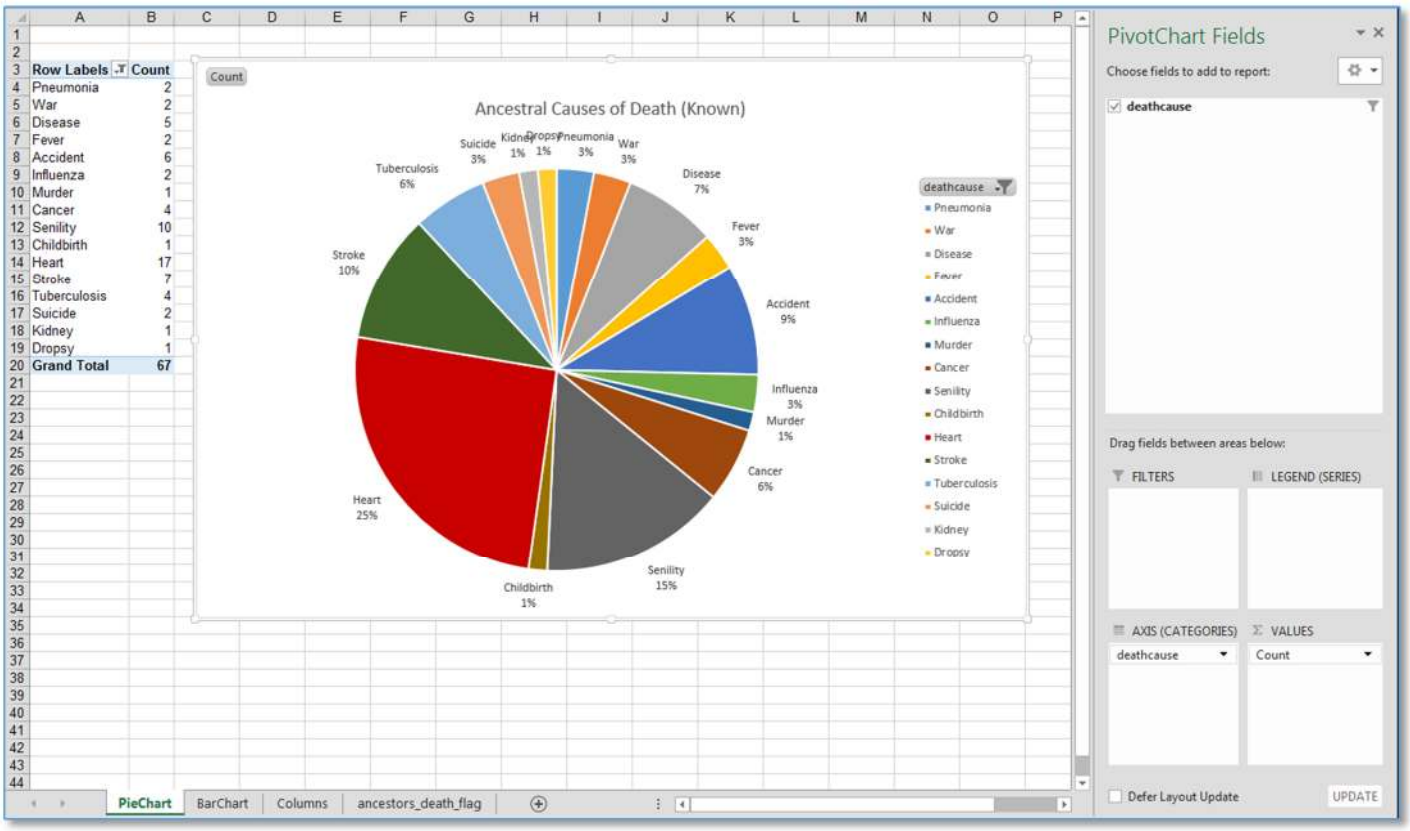
For most of these charts, I began with a blank pivot table, played with its options, and inserted the blank chart based on the pivot table data.





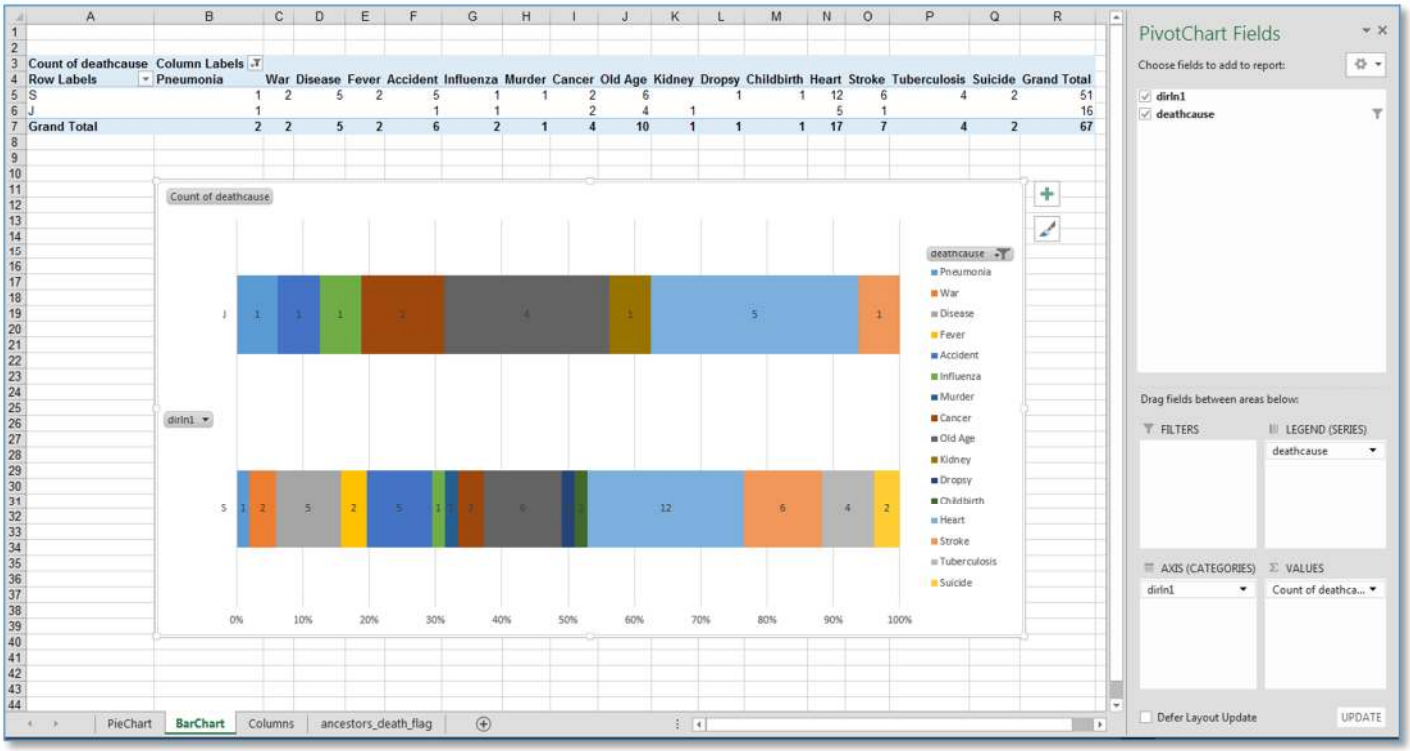
The Pie Chart

This pie chart is equivalent to the charts discussed by various genealogy bloggers. It shows the various causes of death found among the direct ancestors and the number of deaths in each group. These charts are interactive. The "unknown" cause of death has been filtered out, so the percentages shown are relative to all the known causes of death in this group. Actual flag names have been substituted for TMG's single-character flag values. This chart took less than five minutes to create.



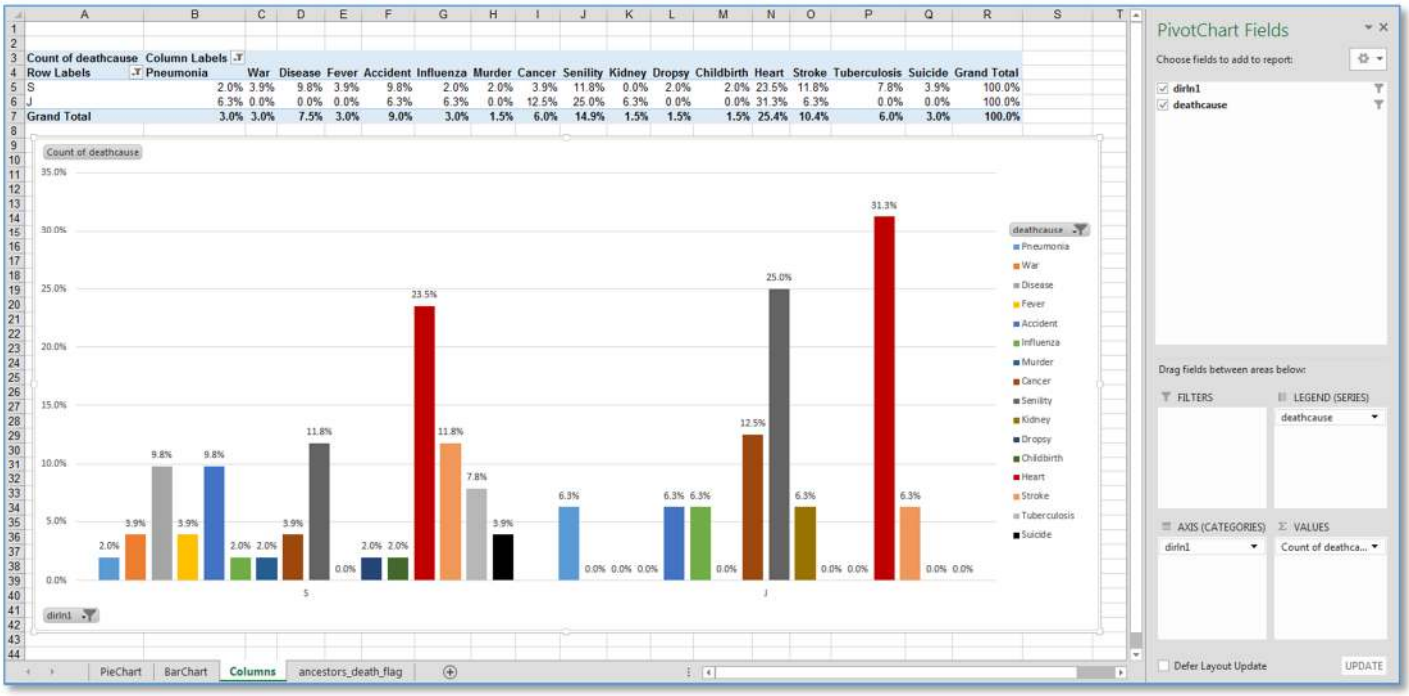
The Bar Chart

This bar chart compares causes of death in two different direct lines. As above, the "unknown" cause of death has been filtered out. Because both bars are set up as 100% bars, and there are far fewer known causes of death in one line, this chart is not as informative as the pie chart.



The Columns Chart

This columns chart is more informative than the preceding bar chart. The two lines of descent are compared side-by-side, and cause of death percentages are relative to the parent series, the specific line of descent. Note that "Heart" related deaths are the largest group in both lines, but they constitute 31.3% of the known causes of death in the "J" line and 23.5% of those deaths in the "S" line. Currently, there are no known kidney-related causes of death in the "S" line, and there are no known murders, suicides, or childbirth deaths in the "J" line.



Count of deathcause	Column Labels																	
Row Labels	Pneumonia	War	Disease	Fever	Accident	Influenza	Murder	Cancer	Senility	Kidney	Dropsy	Childbirth	Heart	Stroke	Tuberculosis	Suicide	Grand Total
S	2.0%	3.9%	9.8%	3.9%	9.8%	2.0%	2.0%	3.9%	11.8%	0.0%	2.0%	2.0%	23.5%	11.8%	7.8%	3.9%	100.0%
J	6.3%	0.0%	0.0%	0.0%	6.3%	6.3%	0.0%	12.5%	25.0%	6.3%	0.0%	0.0%	31.3%	6.3%	0.0%	0.0%	100.0%
Grand Total	3.0%	3.0%	7.5%	3.0%	9.0%	3.0%	1.5%	6.0%	14.9%	1.5%	1.5%	1.5%	25.4%	10.4%	6.0%	3.0%	100.0%

Experiment with your data using your spreadsheet program's chart features. You might be surprised at what you can create quickly and easily.