U.S. Population Distribution by Age in Excel

By Robert Allison

Here's another example in my series showing how to create graphs "on the cheap" – this time I'll be using Microsoft Excel to create a chart showing the population distribution by age in the U.S.

Getting the Data

To get the population data, go to the U.N. data download page, and download the Population by Single Ages Excel table.

Population - Single Ages	**Population by Single Age - Both Sexes (XLSX, 14.02 MB)**	Total population (both sexes combined) by single age. De facto population as of 1 July of the year indicated classified by single age (0, 1, 2, ..., 99, 100+). Data are presented in thousands.

Open the data in Excel, and find the United States year 2022 row (that was row 1148 when I downloaded it). Here's a portion of the table, showing ages 0, 1, 2, and 3.

	Inde	Variant	Type of aggregate or group *	Not	Location code	SDMX code*	Type	Parent code	Year	Total population by single age, both sexes combined (thousands) 0	1	2	3
1142	1125	Estimates/M	United States of America (and dependencies)	8	1111		Special other	1841	2016	4 018	4 049	4 062	4 072
1143	1126	Estimates/M	United States of America (and dependencies)	8	1111		Special other	1841	2017	3 980	4 038	4 079	4 088
1144	1127	Estimates/M	United States of America (and dependencies)	8	1111		Special other	1841	2018	3 915	3 993	4 064	4 107
1145	1128	Estimates/M	United States of America (and dependencies)	8	1111		Special other	1841	2019	3 851	3 921	4 013	4 091
1146	1129	Estimates/M	United States of America (and dependencies)	8	1111		Special other	1841	2020	3 761	3 855	3 933	4 028
1147	1130	Estimates/M	United States of America (and dependencies)	8	1111		Special other	1841	2021	3 721	3 766	3 861	3 938
1148	1131	Estimates/M	United States of America (and dependencies)	8	1111		Special other	1841	2022	3 752	3 728	3 774	3 868
1149	1132	Estimates/M	United States of America (and dependencies)	8	1111		Special other	1841	2023	3 767	3 762	3 739	3 783
1150	1133	Estimates/M	United States of America (and dependencies)	8	1111		Special other	1841	2024	3 784	3 778	3 773	3 748

Copy-n-paste the 2022 row (row 1148), and the header row (row 17) into another sheet in Excel. The resulting table will look something like this:

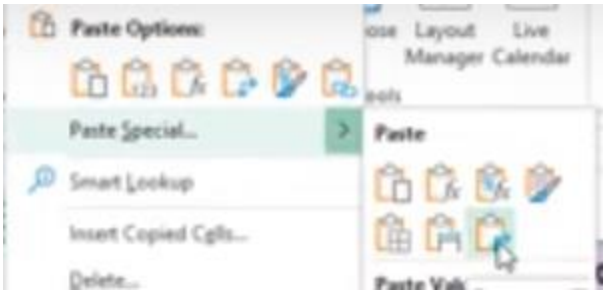
	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Index	Variant	ggregate o	Notes	Location code	SDMX code**	Type	Parent code	Year	0	1	2	3	4
2	1131	Estimates/I	United S	8	1111		Special oth	1841	2022	3 752	3 728	3 774	3 868	3

Massaging the Data

Rather than having the population for each age in a separate column, we want to have an age column, and a population column. Therefore, let's transpose the data. First, select all the age and population cells (just those cells – not the other ones to the left). That's `I1` to `DF2` in my spreadsheet:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Index	Variant	ggregate o	Notes	Location code	SDMX code**	Type	Parent code	Year	0	1	2	3	4
2	1131	Estimates/I	United S	8	1111		Special oth	1841	2022	3 752	3 728	3 774	3 868	3
3														

Right-click the selected values and select "Copy" and then reposition your cursor to another cell, right-click and Paste Special… and select the last Paste icon, which is Transpose:



You'll now get an age column, and a population column:

4		
5	0	3 752
6	1	3 728
7	2	3 774
8	3	3 868
9	4	3 944
10	5	4 038
11	6	4 115
12	7	4 157

Next, insert 3 columns (cumulative pop, pct le age, and pct above age). Here's the equation I used to populate the cumulative pop column (see screen-capture). And the final value at the bottom of the column is 341,907 (remember that's in 1000s … total population is 341,907,000).

D6 =SUM($E$5:E6)

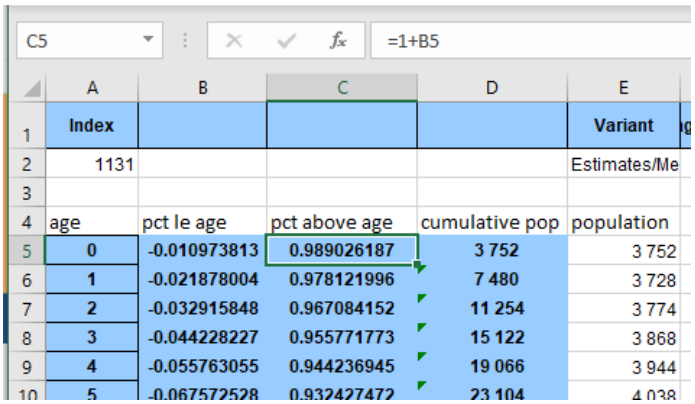
	A	B	C	D	E
1	Index				Variant
2	1131				Estimates/Me
3					
4	age	pct le age	pct above age	cumulative pop	population
5	0	-0.010973813	0.989026187	3 752	3 752
6	1	-0.021878004	0.978121	7 480	3 728
7	2	-0.032915848	0.967084152	11 254	3 774
8	3	-0.044228227	0.955771773	15 122	3 868
9	4	-0.055763055	0.944236945	19 066	3 944
10	5	-0.067572528	0.932427472	23 104	4 038

Here's the equation I used to calculate the percent of the population less than or equal to the age. I multiply it by -1, because I want it to plot these bar segments below the zero line in my chart.

B5 =-1*(D5/$D$105)

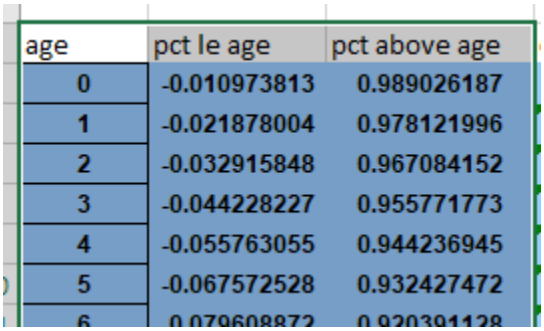
	A	B	C	D	E
1	Index				Variant
2	1131				Estimates/Me
3					
4	age	pct le age	pct above age	cumulative pop	population
5	0	-0.010973813	0.989026187	3 752	3 752
6	1	-0.021878004	0.978121996	7 480	3 728
7	2	-0.032915848	0.967084152	11 254	3 774
8	3	-0.044228227	0.955771773	15 122	3 868
9	4	-0.055763055	0.944236945	19 066	3 944
10	5	-0.067572528	0.932427472	23 104	4 038
11	6	-0.079608872	0.920391128	27 219	4 115

And I populate the percent above the age as =1+B5.

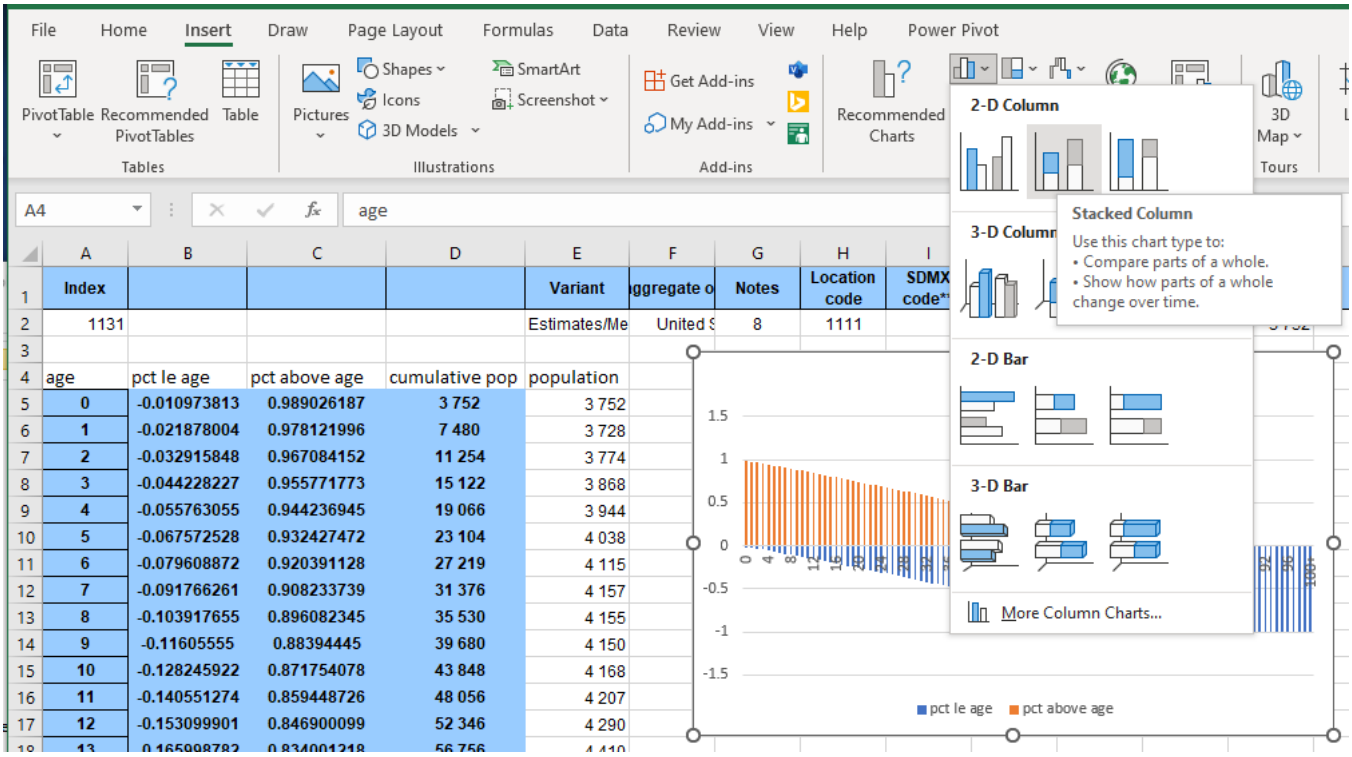


Creating the Chart

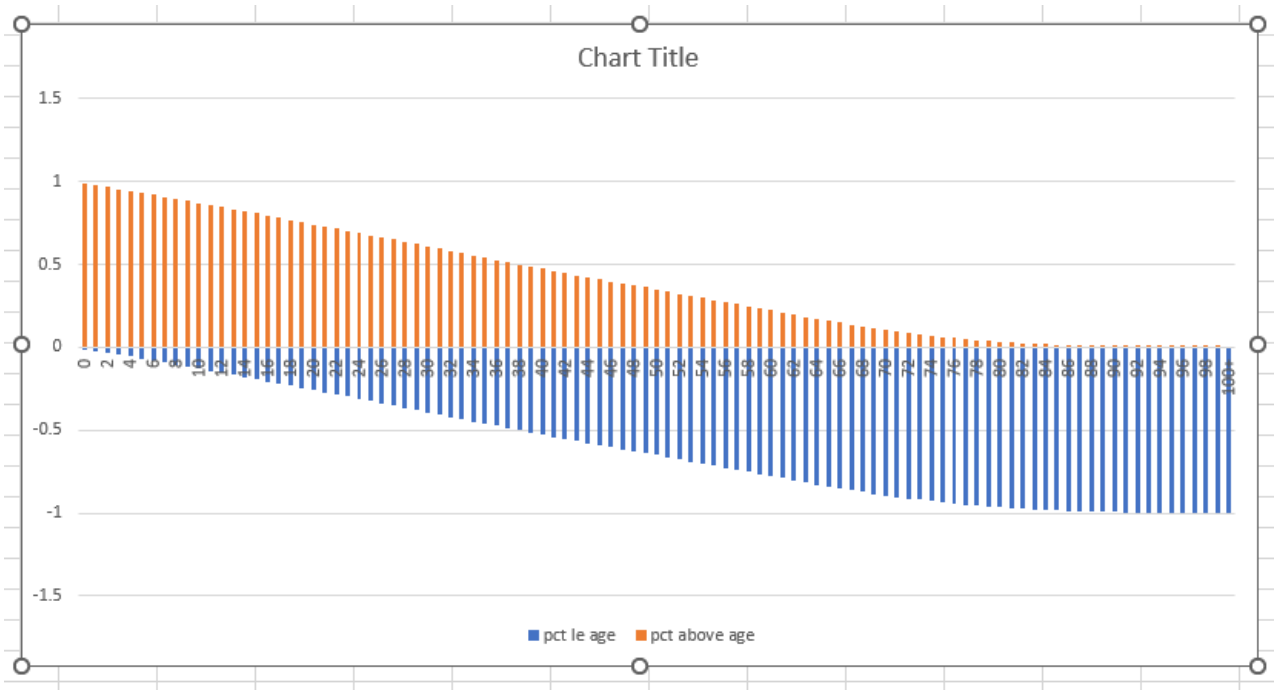
To create the chart, select the age, pct le age, and pct above age values:



Then go to Excel's "Insert" tab, and select 2-D Column, Stacked Column chart:

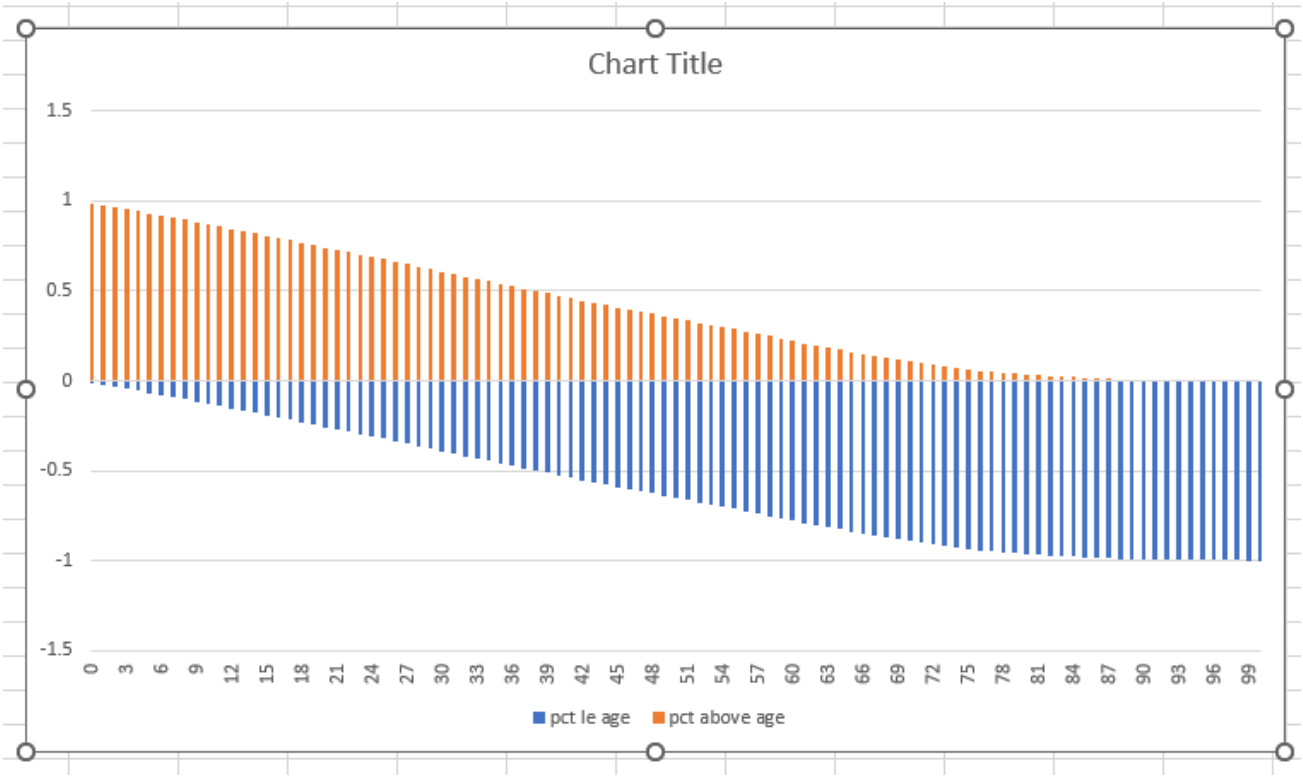


Here’s the default chart, with no customizations. It needs a few enhancements, but it’s basically what we want:



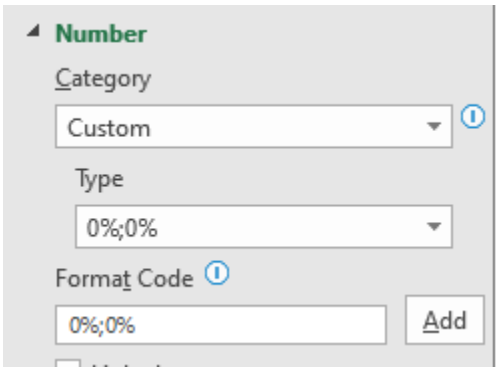
Customizing the Chart

First, move the x-axis labels to the bottom, rather than the middle of the chart. Select the x-axis values, right-click and select “Format axis…”. Change Labels->Label Position from “Next to axis” to “Low”. Now the values are along the bottom side of the chart:

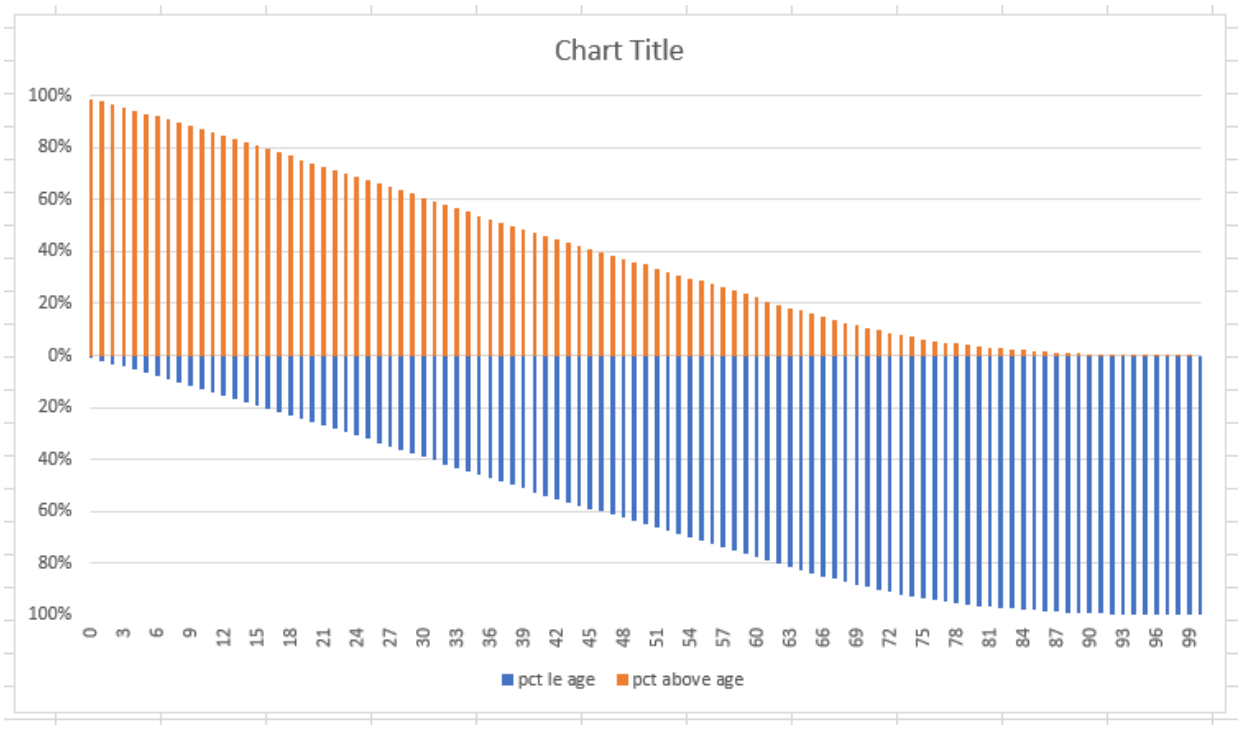


Select the axis on the left side of the chart, right-click and select “Format Axis…” and change the Minimum and Maximum values:

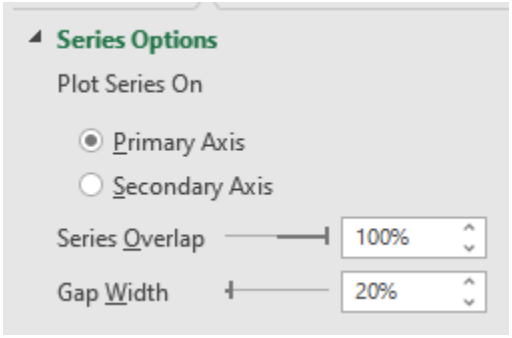
Then change the Number->Category to Percentage, with 0 decimal places, and change the Format Code from 0% to 0%;0% and click “Add” (to make the negative numbers show up without the negative sign). Note that this also changed the Category from Percentage to Custom.

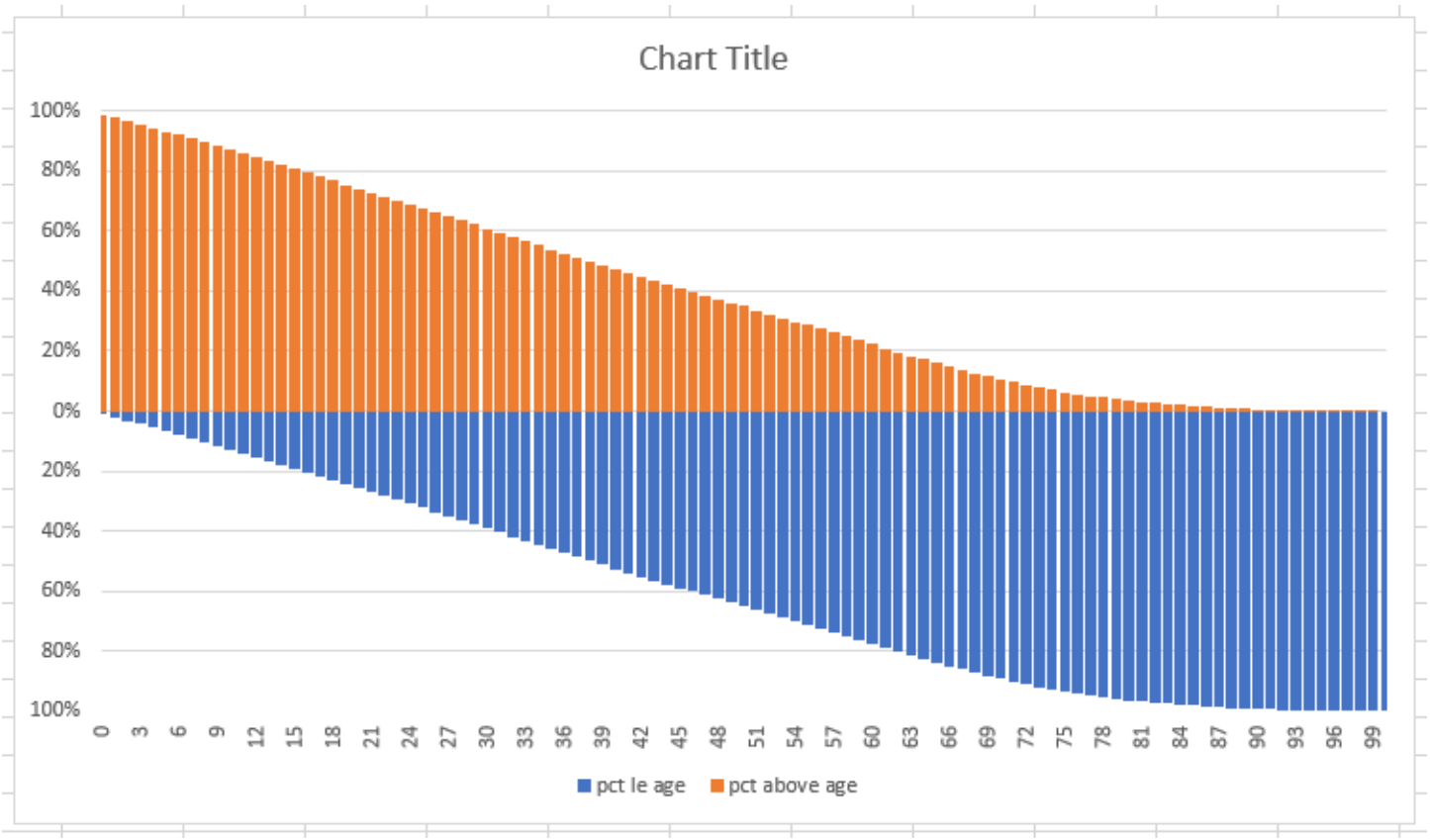


The chart now looks like this:



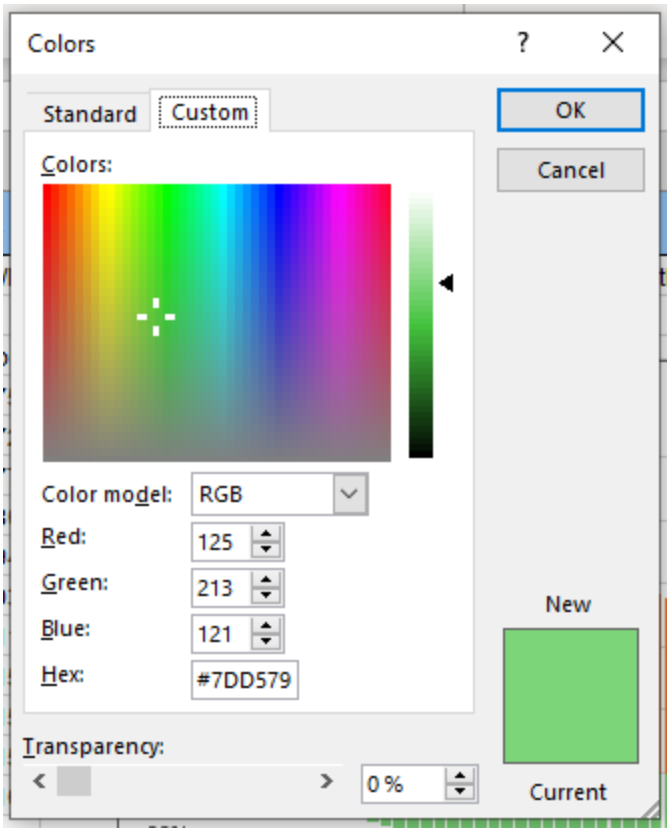
Next, let's make the bars wider. Select the bars on the left, right-click and "Format data series…", and modify the gap width and overlap until the bars and spaces look good to you. Here are the values I used:





Now let's change the colors of the bars. Click the legend until only one item is selected:

Right-click and select "Format Legend entry…". Click the paint bucket beside Fill, Color, and use whichever color editing interface you prefer. I like using the More Colors… color editor. Do this for both of the legend colors.


Colors
Standard
Custom
OK
Cancel
Colors:
Color model:
RGB
Red:
125
Green:
213
Blue:
121
Hex:
#7DD579
New
Transparency:
0 %
Current

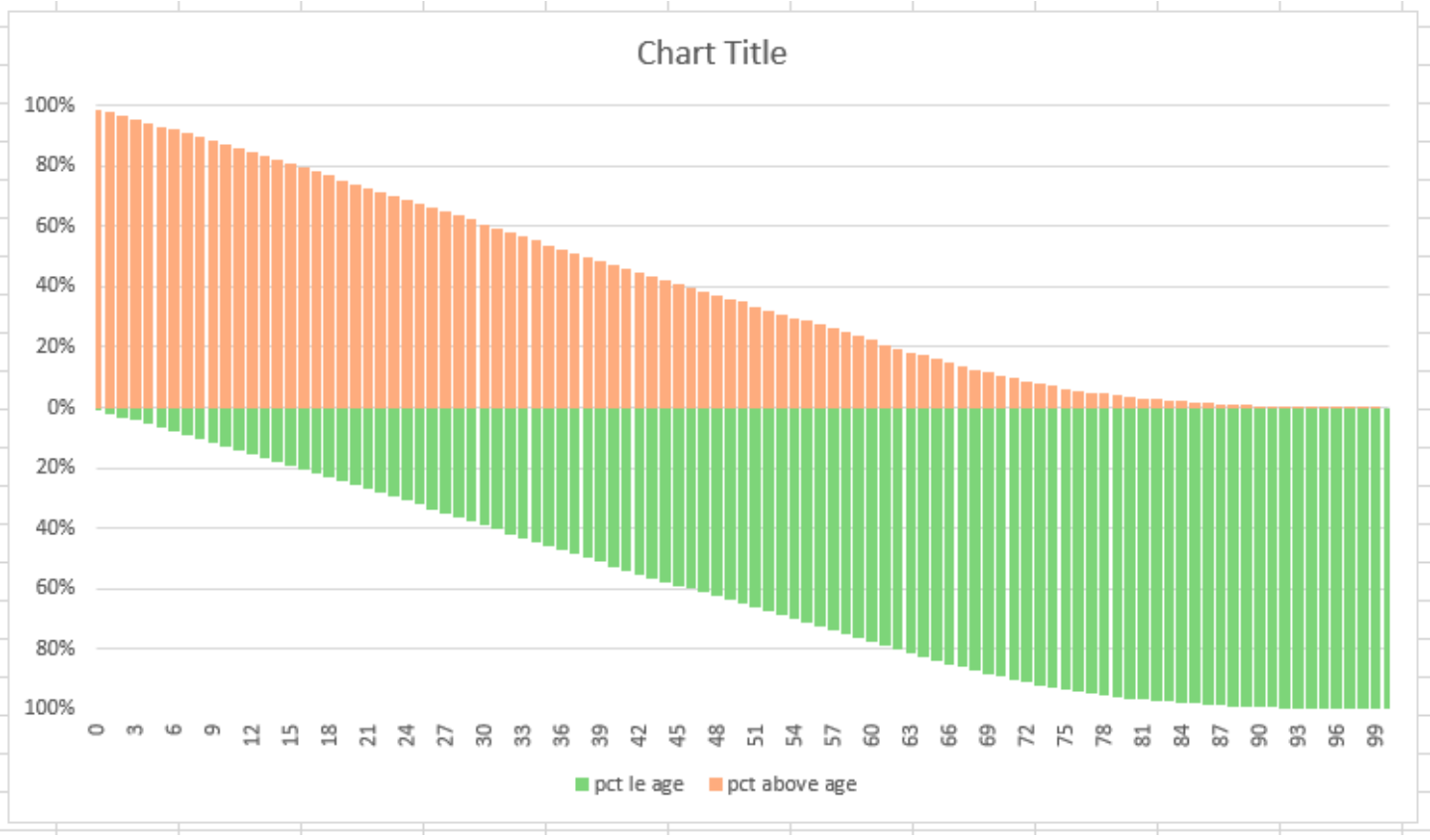
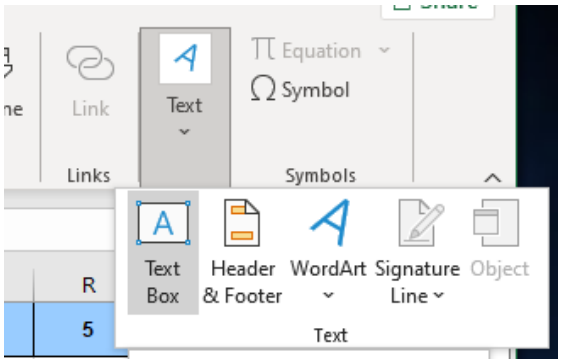


Chart Title
100%
80%
60%
40%
20%
0%
20%
40%
60%
80%
100%
0
3
6
9
12
15
18
21
24
27
30
33
36
39
42
45
48
51
54
57
60
63
66
69
72
75
78
81
84
87
90
93
96
99
pct le age
pct above age

And for the final enhancements, let's work on the text. Change the title to something descriptive such as "**U.S. Age Distribution in Year 2022**".

Click on the color legend, then right-click and "Delete" it.

And now add three pieces of text in arbitrary positions ("**Your Age**," "**% of population older than you**," and "**% of population younger than you**") by going to Excel's "Insert" tab, and selecting Text->Text Box.



And for a final bit of fine-tuning, resize the graph to adjust the bar spacing and horizontal axis auto-labeling until it suits you (note that you might not be able to get it *exactly* like you want).

Final Chart

Here's the final chart – I think it looks pretty good for something created with an Excel license that cost me less than $50. 😊 It even looks pretty good when compared to my old SAS Chart where the licenses cost thousands of dollars!