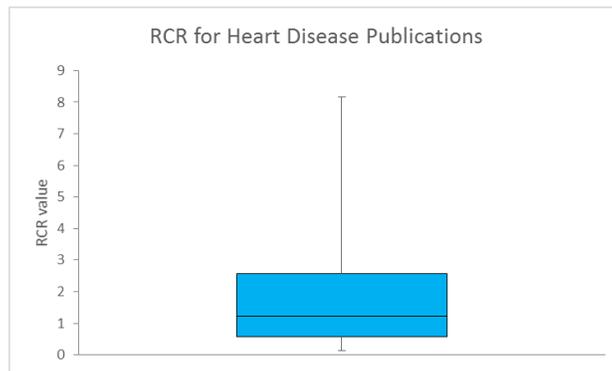


OPA Excel Tips: Creating a box and whisker plot in Excel

What is a box and whisker chart?

A box and whisker plot is a way of showing and comparing distributions. The box represents the middle half of the data (from the 25th to the 75th percentile), with an additional line showing the middle value (the median or 50th percentile). The whiskers show the extreme values and can either show the minimum and maximum values, or as here, the 5th and 95th percentiles, as often these give a better idea of the distribution of the data than the minimum and maximum. Box plots can be drawn vertically or horizontally.



Box and Whisker charts in Excel

Line and bar charts are easy to produce in Excel. In Excel 2016 a new box and whisker plot has been added. In older versions, stock charts exist, and are similar to box and whisker plots but require data in a very specific format, which isn't always possible with data that you wish to convert to a box and whisker plot.

Box and whisker plots can be produced in older versions of Excel though there are a few steps to go through to convert the data to the correct structure and format the chart, which is a stacked bar chart with error bars.

There are other software packages that automatically produce box plots such as Tableau, though most of these require a license. The advantage of Tableau over Excel, as well as there being less formatting required, is that you can add additional information to the chart, such as outliers (values below the 5th percentile, or above the 95th percentile) as additional data points (dots on the chart). It is very complicated to do this in Excel, involving additional data series and secondary axes so this is not show here.

You can produce a single box and whisker plot, or a chart comparing the box and whisker plots for multiple categories. For example the categories could be Fiscal Year, Activity code, RFA, Portfolio. The Y axis will be a numerical range, for example \$s or RCR.

In this tip sheet, we show how to create a single box and whisker plot like the one above, and how to expand the process to show multiple box and whisker plots on the same chart.

Example 1

Step 1. Decide the data you want to present

In this example we will look at Relative Citation Ratio (RCR), recreating the chart that is available in *iCite*. The y-axis will show the RCR value, the x-axis will just have one value – heart disease data.

Step 2. Structure the data

Create a new worksheet in Excel and add the data to it in columns, with one column for each category that you wish to display a box and whisker plot for. In the first example we will produce a chart with a single box and whisker plot.

We have downloaded heart disease publications from *iSearch* for several years. To produce a box plot for RCR we copy the RCR column to a new sheet.

Then create values for:

5th Percentile (=PERCENTILE.INC(x:y,0.05)),

Quartile 1 (=QUARTILE.INC(x:y,1)),

Median (=QUARTILE.INC(x:y,2)),

Quartile 3 (=QUARTILE.INC(x:y,3)),

95th Percentile (=PERCENTILE.INC(x:y,0.95))

Where x is the first cell in your column of data and y is the last cell.

| | A | B | C | D | E |
|----|-------|---|-----------------|----------|---|
| 1 | RCR | | | | |
| 2 | 2.25 | | 5th Percentile | 0.140741 | |
| 3 | 6.13 | | Quartile 1 | 0.566958 | |
| 4 | 0.38 | | Median | 1.227621 | |
| 5 | 1.27 | | Quartile 3 | 2.571595 | |
| 6 | 5.51 | | 95th Percentile | 8.159798 | |
| 7 | 15.69 | | | | |
| 8 | 12.68 | | | | |
| 9 | 5.54 | | | | |
| 10 | 1.03 | | | | |
| 11 | 3.24 | | | | |
| 12 | 0.32 | | | | |

The next step is to calculate the values to display on the chart. To do this we show Quartile 1, Median and Quartile 3 as a stacked bar and the 5th and 95th percentiles become error bars.

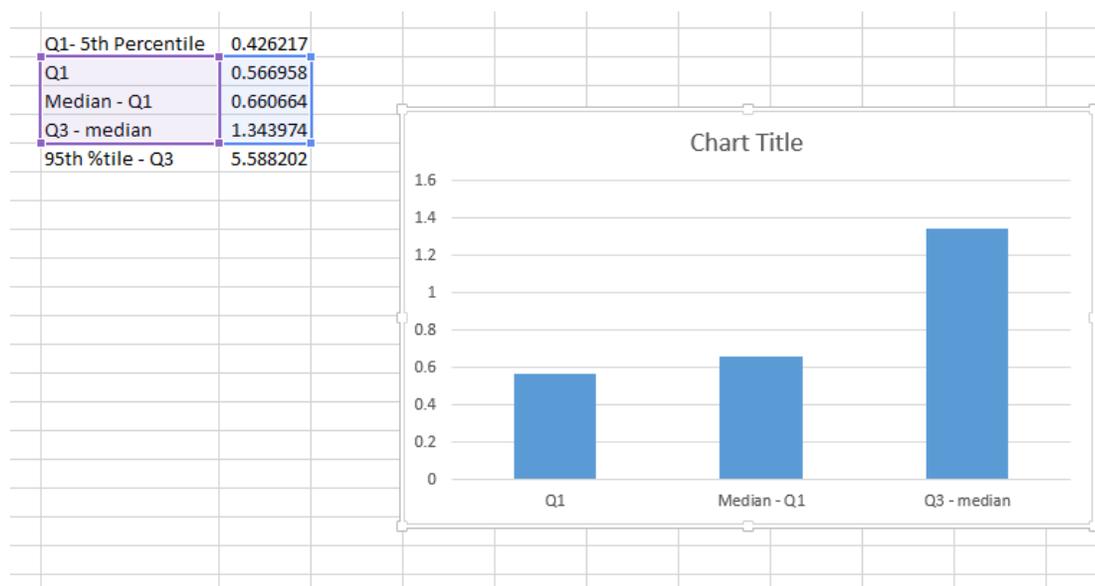
If we use the values that we have just calculated then we would be overestimating the median and quartile 3 as we would be adding the actual value to the stacked bar chart. We need to calculate the difference between Quartile 1 and the Median and the Median and Quartile 3 to correctly display the data as a stacked bar chart.

The error bars will show the 5th and 95th percentiles but to present these as error bars we need to calculate the distance from Quartile 1 to the 5th percentile and from Quartile 3 to the 95th percentile. These are calculated in an additional column.

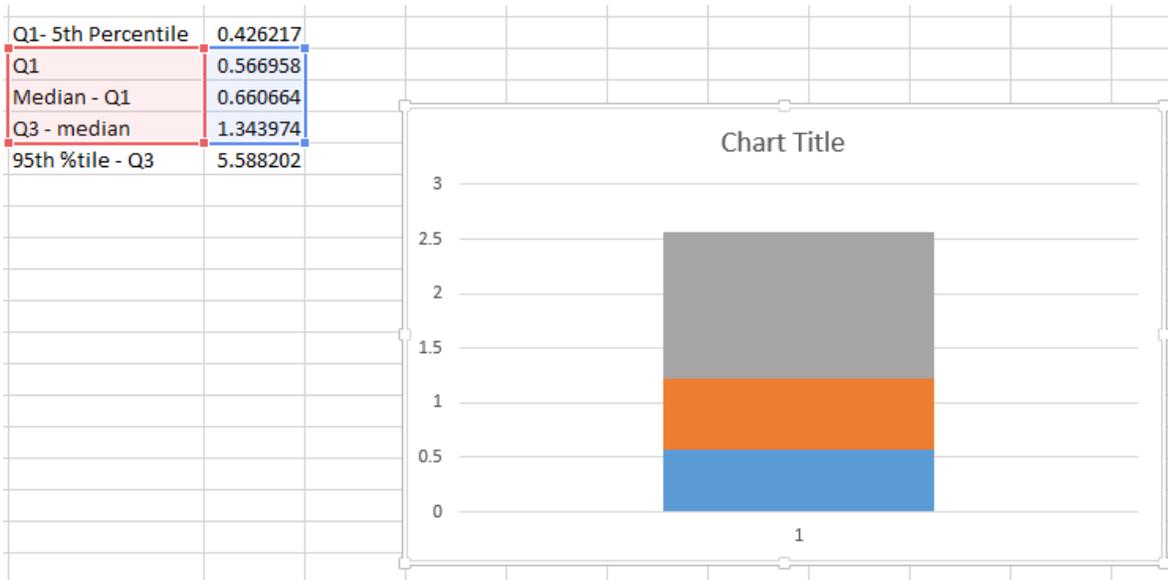
| | A | B | C | D | E | F | G | H |
|----|-------|-----------------|----------|---|---|--------------------|----------|---|
| 1 | RCR | | | | | | | |
| 2 | 2.25 | 5th Percentile | 0.140741 | | | Q1- 5th Percentile | 0.426217 | |
| 3 | 6.13 | Quartile 1 | 0.566958 | | | Q1 | 0.566958 | |
| 4 | 0.38 | Median | 1.227621 | | | Median - Q1 | 0.660664 | |
| 5 | 1.27 | Quartile 3 | 2.571595 | | | Q3 - median | 1.343974 | |
| 6 | 5.51 | 95th Percentile | 8.159798 | | | 95th %tile - Q3 | 5.588202 | |
| 7 | 15.69 | | | | | | | |
| 8 | 12.68 | | | | | | | |
| 9 | 5.54 | | | | | | | |
| 10 | 1.03 | | | | | | | |
| 11 | 3.24 | | | | | | | |

Step 3. Create the chart

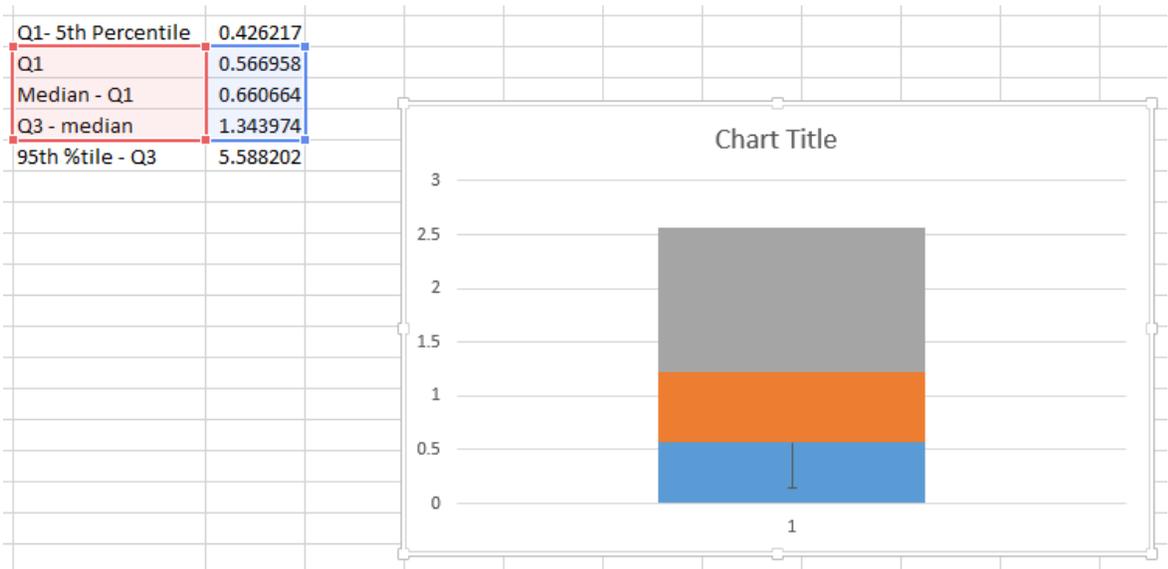
We're now ready to create the box plot.



But we don't want separate bars. Right click on the chart > Select data... > Switch Row/Column > OK

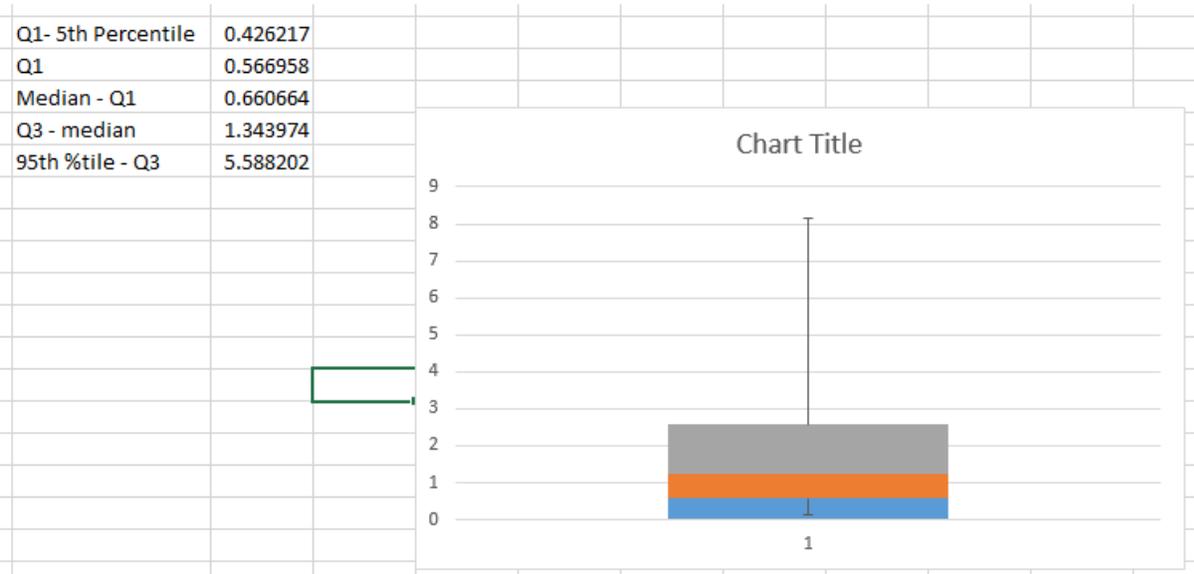


Now to add the error bars. Left click on the bottom bar. In the top menu tabs, select Design > Add Chart Element > Error Bars > More Error Bars Options. Under Direction, select Minus. Then scroll down and select Custom > Specify Values. In the Negative box, select the Q1-5th Percentile value and click OK. An error bar will appear on the chart.

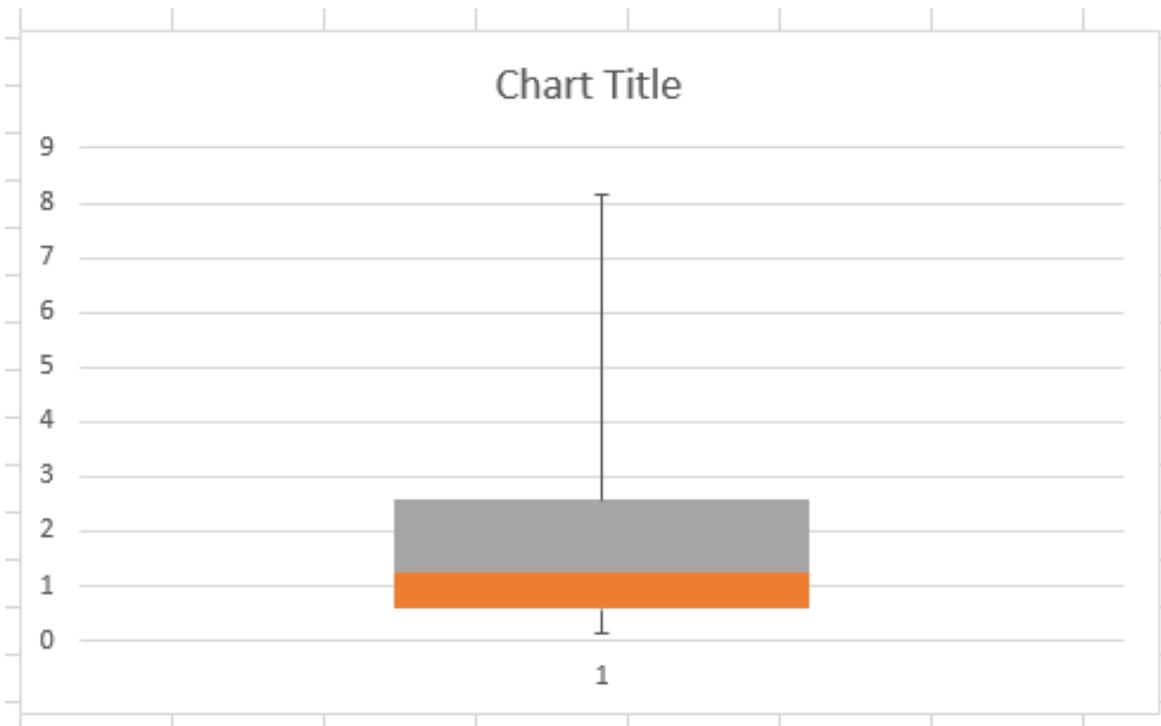


Do the same thing for the 95th Percentile. Left Click on the top bar, then go through the same process but when you get to More Error Bars Options, under Direction, select Plus. Then scroll

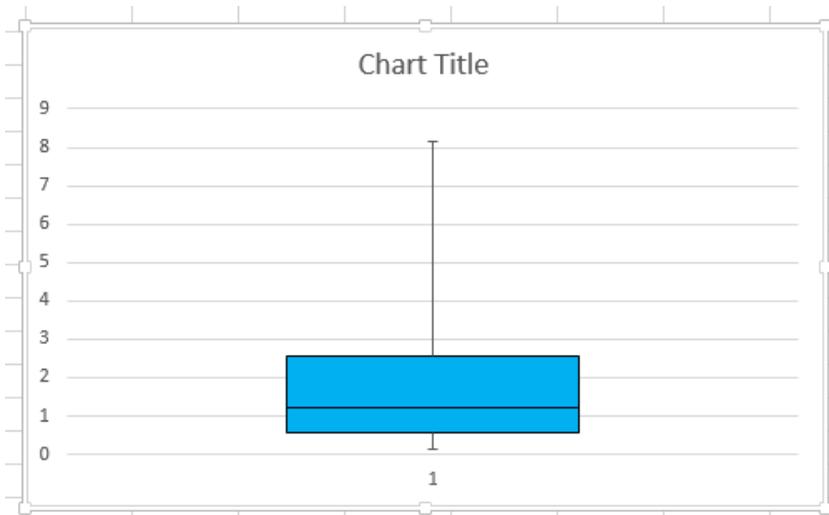
down and select Custom > Specify Values. In the Positive box, select the 95th Percentile – Q3 value and click OK. An error bar will appear on the chart.



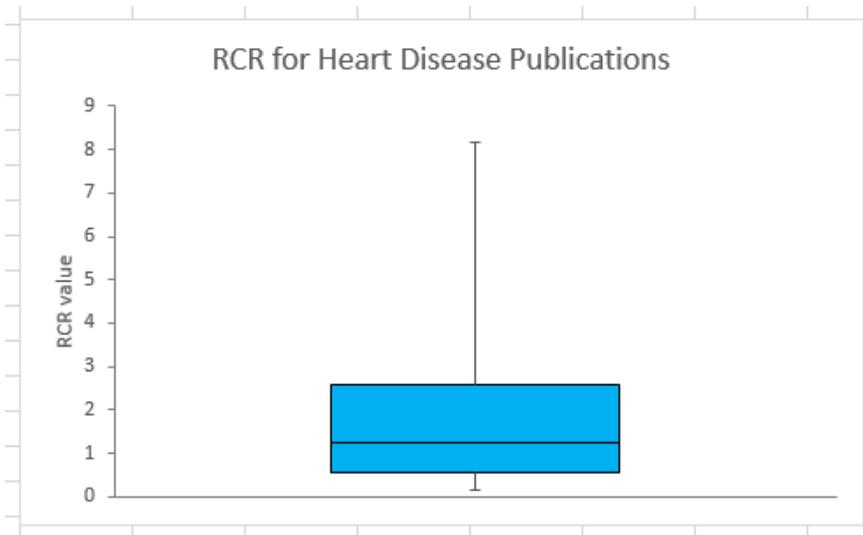
Now we just need to tidy up the formatting. Right click on the bottom bar, select Fill > No Fill.



Then we can change the color of the other two blocks to match, and add borders. Right click on each block in turn, select Fill and pick a color, then right click and select Outline and the color e.g. black.



You can also change or remove the label on the x axis, add a title for the Y axis and remove the grid lines.



You can also add labels if you wish, or add them as text boxes if you are presenting the chart in a Powerpoint slide.

Example 2: Multiple Box and Whisker plots on the same chart

In this example we will use the same dataset but show Box and Whisker plots for individual years.

Step 1. Format the data

The data needs to be in a similar format to the previous one, but with individual years in separate columns. To do this we'll filter the original dataset on Fiscal Year then copy the RCR values for that year to a new column on the worksheet we are using to create the chart.

| | A | B | C | D | E | F |
|----|------|------|------|------|------|-------|
| 1 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 |
| 2 | 0.34 | 0.52 | 0.62 | 0.29 | 1.40 | 2.25 |
| 3 | 4.54 | 0.94 | 0.92 | 0.63 | 0.92 | 6.13 |
| 4 | 0.48 | 1.64 | 2.31 | 0.72 | 0.00 | 0.38 |
| 5 | 2.75 | 0.16 | 0.24 | 0.38 | 0.73 | 1.27 |
| 6 | 2.21 | 1.54 | 1.28 | 3.13 | 4.52 | 5.51 |
| 7 | 1.23 | 0.54 | 0.40 | 1.31 | 0.78 | 15.69 |
| 8 | 0.65 | 7.90 | 0.11 | 1.00 | 2.23 | 12.68 |
| 9 | 0.05 | 2.70 | 1.17 | 1.05 | 1.88 | 5.54 |
| 10 | 2.20 | 1.41 | 2.06 | 1.03 | 0.81 | 1.03 |
| 11 | 0.16 | 0.51 | 0.40 | 1.52 | 1.77 | 3.24 |
| 12 | 0.00 | 0.58 | 0.67 | 0.31 | 1.13 | 0.32 |
| 13 | 1.32 | 0.40 | 1.44 | 1.50 | 1.89 | 0.25 |
| 14 | 0.48 | 0.60 | 0.44 | 1.69 | 6.87 | 1.44 |
| 15 | 1.94 | 1.44 | 1.41 | 1.46 | 0.44 | 1.40 |
| 16 | 2.71 | 4.70 | 4.70 | 0.68 | 1.24 | 1.84 |
| 17 | 2.39 | 0.83 | 2.00 | 1.43 | 4.56 | 1.15 |
| 18 | 0.56 | 0.14 | 0.45 | 0.29 | 0.52 | 2.24 |
| 19 | 1.42 | 0.25 | 0.99 | 0.47 | 0.38 | 3.86 |
| 20 | 1.47 | 1.00 | 4.50 | 2.46 | 1.96 | 0.00 |
| 21 | 2.05 | 1.64 | 0.91 | 0.45 | 3.42 | 2.52 |
| 22 | 2.49 | 0.96 | 1.76 | 2.22 | 0.00 | 0.55 |

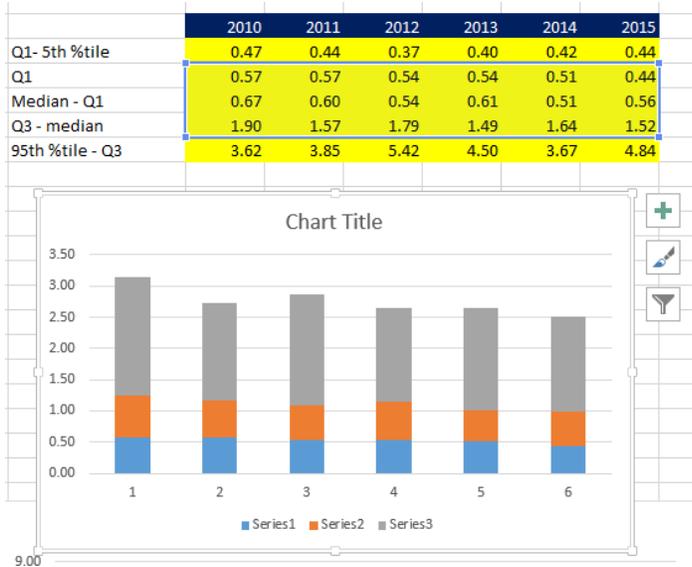
Step 2. Calculate the values required for the chart

We can then calculate the values required for the chart, as we did above, using the same formula for each year of data but making sure the cell values include all the data for an individual year

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N |
|----|------|------|------|------|------|-------|---|-----------------|----------|----------|----------|----------|----------|----------|
| 1 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | | | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 |
| 2 | 0.34 | 0.52 | 0.62 | 0.29 | 1.40 | 2.25 | | 5th Percentile | 0.10 | 0.13 | 0.17 | 0.14 | 0.08 | 0.00 |
| 3 | 4.54 | 0.94 | 0.92 | 0.63 | 0.92 | 6.13 | | Quartile 1 | 0.570145 | 0.567892 | 0.539862 | 0.543547 | 0.506544 | 0.442265 |
| 4 | 0.48 | 1.64 | 2.31 | 0.72 | 0.00 | 0.38 | | Median | 1.243318 | 1.170016 | 1.084285 | 1.152562 | 1.019021 | 0.998139 |
| 5 | 2.75 | 0.16 | 0.24 | 0.38 | 0.73 | 1.27 | | Quartile 3 | 2.573357 | 2.170203 | 2.336719 | 2.10108 | 2.15205 | 2.07361 |
| 6 | 2.21 | 1.54 | 1.28 | 3.13 | 4.52 | 5.51 | | 95th Percentile | 6.20 | 6.02 | 7.75 | 6.60 | 5.82 | 6.91 |
| 7 | 1.23 | 0.54 | 0.40 | 1.31 | 0.78 | 15.69 | | | | | | | | |
| 8 | 0.65 | 7.90 | 0.11 | 1.00 | 2.23 | 12.68 | | | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 |
| 9 | 0.05 | 2.70 | 1.17 | 1.05 | 1.88 | 5.54 | | Q1- 5th %tile | 0.47 | 0.44 | 0.37 | 0.40 | 0.42 | 0.44 |
| 10 | 2.20 | 1.41 | 2.06 | 1.03 | 0.81 | 1.03 | | Q1 | 0.57 | 0.57 | 0.54 | 0.54 | 0.51 | 0.44 |
| 11 | 0.16 | 0.51 | 0.40 | 1.52 | 1.77 | 3.24 | | Median - Q1 | 0.67 | 0.60 | 0.54 | 0.61 | 0.51 | 0.56 |
| 12 | 0.00 | 0.58 | 0.67 | 0.31 | 1.13 | 0.32 | | Q3 - median | 1.90 | 1.57 | 1.79 | 1.49 | 1.64 | 1.52 |
| 13 | 1.32 | 0.40 | 1.44 | 1.50 | 1.89 | 0.25 | | 95th %tile - Q3 | 3.62 | 3.85 | 5.42 | 4.50 | 3.67 | 4.84 |
| 14 | 0.48 | 0.60 | 0.44 | 1.69 | 6.87 | 1.44 | | | | | | | | |

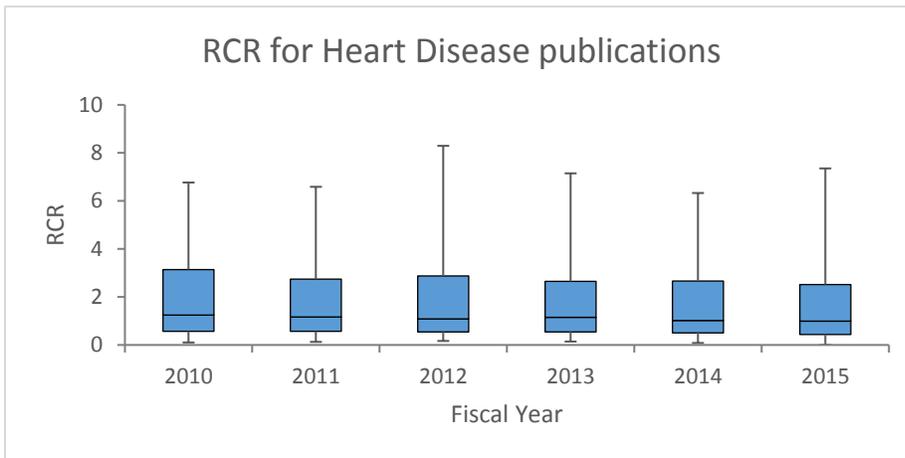
Step 3. Produce the chart and format it

Select the values for the chart in the same way, though this time the chart will default to the stacked bar that we are after without the need to switch rows and columns.



We can then add error bars in the same way as before, selecting a group of cells instead of the single value we selected previously.

With some reformatting we end up with something like the version below which tells us that there is little variation in RCR over the years.



More help

There is more help on formatting charts in the Excel help. There is plenty more detail about box and whisker plots on line which can be found via google or another search engine.

If you have any questions or comments about this tip sheet, contact OPA training:
OPA_Training@mail.nih.gov