General Assembly

STUDY GUIDE

BEYOND BASIC CHARTS: BUBBLE PLOTS AND HEAT MAPS

Key Terms

Bubble Plot: An effective visualization for three quantitative continuous variables.

Heat Map: A visualization of unordered, comparable data using conditional formatting.

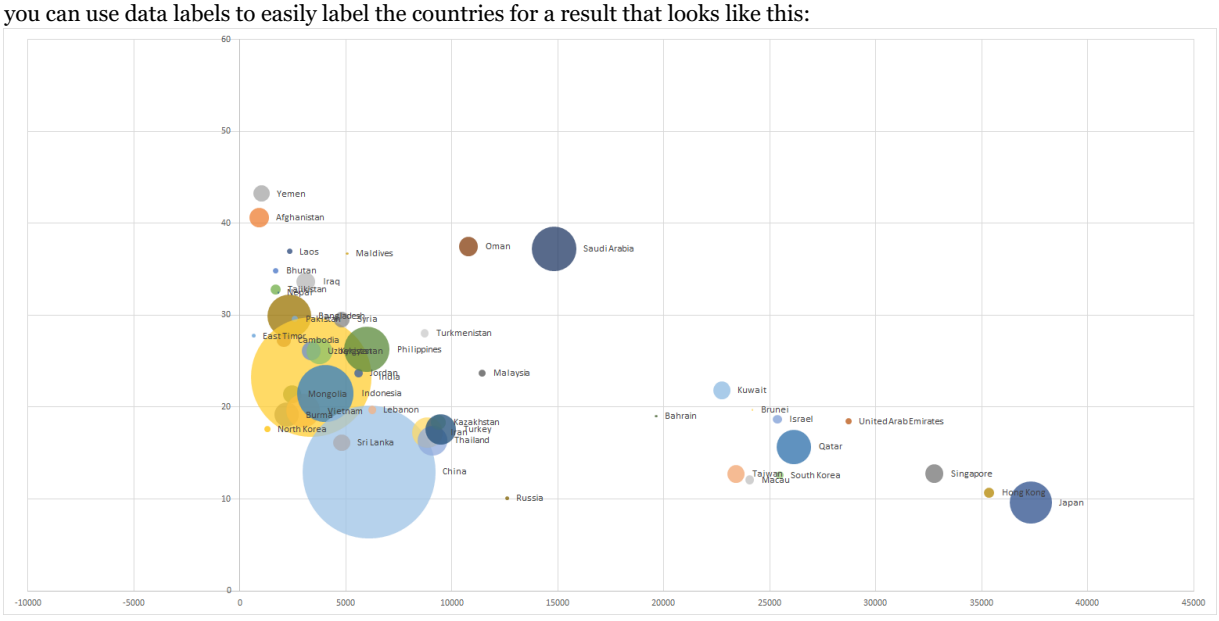
Cheat Sheet

1. Making and Interpreting Bubble Plots

- The easiest way to make a bubble plot is to first put the data in the order Excel wants it: x-axis, then y-axis, then size of bubble. So if we want per-capita income to be our x-axis, birth rate to be our y-axis, and population to represent the size of our bubble, then we'd have a data set that looks like this:

	A	B	C	D
1	regionName	perCapitaIncome	birthRate	population
2	East Timor	650	27.75	997853
3	Afghanistan	910	40.63	28717213
4	Yemen	1040	43.23	19349881
5	North Korea	1300	17.61	2712315
6	Bhutan	1690	34.82	2139549
7	Tajikistan	1690	32.78	6863752
8	Nepal	1820	32.46	329684
9	Cambodia	2080	27.28	13124764

- Highlighting all of the numbers and navigating to Insert >> Bubble Plot produces a one-color bubble plot. Each country only shows up once, and we want each country to be its own color, so we'd have one color per bubble. By right clicking on a bubble, choosing "Format Data Series," then "Vary colors by point" under the spill-paint icon, we can automatically make the bubbles different colors. If you're on a PC, you can use data labels to easily label the countries for a result that looks like this:



- If you want multiple bubbles per color or want a legend, right click on the chart and use the Select Data menu.
- What insights can we gain from this? The bubbles aren't in any noticeable pattern, so the x- and y-axes probably aren't related — that is,

per-capita income doesn't affect birth rate in a predictable way. Some of the smaller bubbles are to the right.

2. Making and Interpreting Heat Maps

- Can be made from PivotTables or regular data.
- Are typically for unsortable data.
- Standard heat maps have a categorical variable plotted vertically on the left side, a categorical variable plotted horizontally across the top, and a single quantitative variable that is tabulated for each cross section:

Average of wage	Column Labels					
Row Labels	NA	north_east	nothern_central	rural_area	south	Grand Total
Agricultural	1.29	1.62	1.40	1.23	1.25	1.31
Business_and_Repair_Service	1.64	1.71	1.75	1.33	1.59	1.66
Construction	1.60	1.83	1.66	1.85	1.56	1.62
Entertainment	1.26	1.72	0.87	1.81	1.22	1.19
Finance	1.71	2.11	1.78		2.02	1.88
Manufacturing	1.75	1.86	1.78	1.62	1.76	1.78
Mining	2.00		1.93	1.85	1.85	1.92
Personal_Service	1.71	1.88	1.43	1.48	1.38	1.55
Professional_and_Related Service	1.60	1.65	1.32		1.62	1.53
Public_Administration	1.73	1.79	1.91		1.73	1.78
Trade	1.55	1.60	1.48	1.63	1.41	1.50
Transportation	1.83	1.73	1.93	1.90	1.93	1.89
Grand Total	**1.65**	**1.75**	**1.63**	**1.58**	**1.62**	**1.65**

- In the example above, the average wage is the single quantitative variable, the sector is the categorical vertical variable, and the region is the categorical horizontal variable.
- Using different quantitative variables in a single heat map can become misleading or confusing, especially if they are on different scales or have different units.