Excel Step by Step

1. Used Tabula.technology to convert PDF into Excel.
2. Used text to columns feature with a space, so that the number and percentage were in different columns for each variable. Resulting data sets in first two sheets of “PostDoc\_Analyzed” Excel file.
3. Did a cluster analysis with 3 (and 4 and 5) anchors.
4. Copied and pasted data for Academia, For-Profit, Government, Nonprofit and the associated department name or center.
5. Used VLOOKUP() function in order to look up and match the relevant “teaching” and “non-science related” data point with the correct Department or Center.
6. Numbered the departments/centers for easy use in future analysis.
7. Found the mean and standard deviation for all the data points of each of my 6 variables using the average() and stdev().
8. Used the Standardize() function to find the z scores for all the variables.
9. Made 3 (or 4 or 5 depending) tentative anchors, with associated z scores for each variable. Used VLOOKUP to match the associated z score with the anchor department.
10. Used the Sumxmy2() function to calculate the squared distance between the different anchors and the department/center.
11. Found the minimum squared distance between the anchors and the department/center.
12. Used the Match() function to find between which anchor the most minimum squared distance was.
13. Found the sum of the minimum squared distances using sum().
14. Used Solver to find the 3 (or 4 or 5 depending) anchors that would give the minimum sum of the minimum squared distances.
    1. Set objective cell to the sum of the minimum squared distances.
    2. Checked min.
    3. Changed variable cells to the department/center number.
    4. Added 3 constraints: must be an integer, greater than 1, and less than number of data points (43). Selected Evolutionary as “solving method.”
15. Repeat steps 4-10 on 4 anchors and 5 anchors.
16. Used scatter plot function for the data visualization for four anchors. Added axes and chart titles.