DSIC 210: HW #1 Name: SOLUTION
Spring 2016
Points: 25

Consider the following variable description from the RateMyProfessor dataset. This data is provided on our course website.

Variable / Field	Description
Table_ID	Used to uniquely identify each row of the dataset, 7448 unique values: 1,2, …, 7448
College	College for which the faculty member teaches. 11 unique values: Chicago, CMU, Tufts, Duke, UTexas, Harvard, Yale, MIT, Stanford, Princeton, Rice
Department	Department in which the faculty member teaches. 140 unique values: Accounting, Advertising, …, Writing
Name	Name of Faculty Member, 7413 unique values: Abbott, Andrew, Achar, Pramod, …, Zodrow, George
Total Ratings	Number of Rate My Professor reviews. Min=0, Max = 116
Overall Quality	Average overall quality score. Min=1, Max=5
Easiness	Average easiness score. Min=1, Max=5
Hot	An indicator of whether or not the faculty member is considered hot by at least one reviewer

Table name: Professors

	A	B	C	D	E	F	G	H
1	Table_ID	College	Department	Name	Total Ratings	Overall Quality	Easiness	Hot
2	1	UChicago	Sociology	Abbott, A	6	2	1.3	Not Hot
3	2	UChicago	Mathematics	Achar, Pra	6	4.4	2.8	Hot
4	3	UChicago	English	Alexande	1	1	5	Not Hot
5	4	UChicago	Social Work	Allard, Sc	5	3.2	2.4	Not Hot
6	5	UChicago	Classics	Ando, Clif	1	5	5	Not Hot
7	6	UChicago	Science	Archer, Da	12	4.1	2.7	Hot
8	7	UChicago	Religion	Arnold, Da	0			Not Hot
9	8	UChicago	English	Atkinson,	1	5	4	Hot
10	9	UChicago	History	Austen, R	20	1.9	1.9	Not Hot
11	10	UChicago	Mathematics	Babai, Las	3	4.5	1.3	Not Hot

1. Consider the following applications of the =COUNT() function in Excel. (4 pts)

   Version A: =COUNT(Professors[College])
   Version B: =COUNT(Professors[Total Ratings])
   Version C: =COUNT(Professors[Overall Quality])

Version A:	0
Version B:	7448
Version C:	6441

   a. Which version produces the correct number of rows in this dataset? Discuss.

      Version B appears to be correct. The number of rows in this dataset is 7448.

   b. Figure out what Version A is doing. Why does this not produce the correct value?

      The =COUNT() function counts the number of cells that contain *numbers*.

   c. Find an alternative to the COUNT() function to obtain the count using the Professors[College] column. Verify that this function produces the desired value.

      Two versions provided here – other versions may exist.
      - =COUNTA(Professors[College])
      - =COUNTIF(Professors[College],"<>")

   d. Why does Version C produce a value that is different than Version B?

      Simply because not all professors in this list have an Overall quality score. In fact, 7448 – 6441 = 1007 professors do not have an overall quality score.

2. What is the average overall quality rating for all professors? (1 pt)

   =AVERAGE(Professors[Overall Quality]) = 3.815

3. Using the =AVERAGEIF() function, obtain the average overall quality rating by college. Rank/sort the colleges from highest average overall score to lowest. (3 pts)

College	
Harvard	4.036
Princeton	3.964
MIT	3.926
Stanford	3.905
UChicago	3.893
Rice	3.864
Tufts	3.797
Duke	3.793
Yale	3.786
CMU	3.770
UTexas	3.746

4. Using the =AVERAGEIF() function, obtain the average overall quality rating by Hot. Can we say that a faculty member that is considered hot by at least one reviewer tends to have a higher average overall quality score? How much does Hotness appear to affect the average overall quality score? Briefly discuss. (3 pts)

Hot	4.367
Not Hot	3.561

The average overall quality score for a Hot professor is 4.4 and those not considered hot is 3.6. Thus, it appears that when a professor is considered Hot, we can expect their overall quality score to increase by 0.8 or so.

5. Using the =COUNTIF() function in Excel to obtain the total number of reviews by Department. (2 pts)

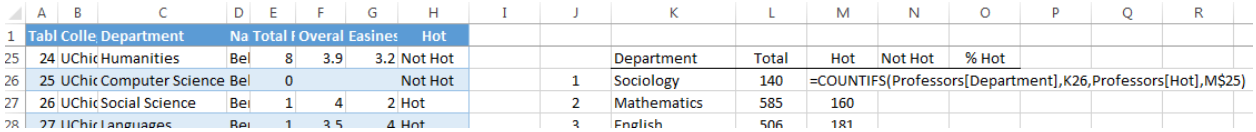
Original Order	Sorted alphabetically by Department

Department	Total
Sociology	140
Mathematics	585
English	506
Social Work	32
Classics	99
Science	146
Religion	85
History	340
Urban Studies	2
Chemical Enginee	2
African Studies	1
Russian Studies	1
Slavic Languages &	1
Media Studies	1
Media Arts	1
Drama	7
Mgmt Science & E	1
East Asian Langua	1

Department	Total
Accounting	24
Advertising	15
Advisor	1
African Studies	1
African-American	4
American Studies	7
Anthropology	143
Arabic	3
Sports	1
Statistics	25
Studio Art	7
Theater	103
Theology	26
Urban Planning	3
Urban Studies	2
Women	1
Women's Studies	28
Writing	118

6. Use the =COUNTIFS() function in Excel, obtain the total number of Hot faculty members by Department. Repeat to obtain the total number of Not Hot faculty members by Department. Use these value to obtain the % of faculty members that are consider Hot by Department. Finally, sort this list to obtain the following. (6 pts)

My =COUNTIFS() function is provided below. Need to check two conditions (Department, and Hot). The M$25 will force Excel to use row 25 when copy down is used to obtain counts for remaining Departments.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Tabl	Colle	Department	Na	Total F	Overal	Easines	Hot										
25	24	UChic	Humanities	Bel	8	3.9	3.2	Not Hot			Department	Total	Hot	Not Hot	% Hot			
26	25	UChic	Computer Science	Bel	0			Not Hot		1	Sociology	140	=COUNTIFS(Professors[Department],K26,Professors[Hot],M$25)					
27	26	UChic	Social Science	Bel	1	4	2	Hot		2	Mathematics	585	160					
28	27	UChic	Languages	Bel	1	3.5	4	Hot		3	English	506	181					

Top and Bottom 5 after sorting by % Hot.

Top 5 Departments for Hotness					
Department	Total	Hot	Not Hot	% Hot	% Not Hot
Women	1	1	0	100.0	0.0
Dramatic Arts	1	1	0	100.0	0.0
Slavic Languages &	1	1	0	100.0	0.0
Media Arts	1	1	0	100.0	0.0
Archaeology	4	3	1	75.0	25.0
African-American	4	3	1	75.0	25.0
Latin American St	3	2	1	66.7	33.3

Bottom 5 Departments for Hotness					
Department	Total	Hot	Not Hot	% Hot	% Not Hot
Biomedical	1	0	1	0.0	100.0
Forestry	2	0	2	0.0	100.0
African Studies	1	0	1	0.0	100.0
Russian Studies	1	0	1	0.0	100.0
Media Studies	1	0	1	0.0	100.0
Mgmt Science & E	1	0	1	0.0	100.0
East Asian Langua	1	0	1	0.0	100.0

Note: Several Departments have 0% Hot; thus, any of these 5 will suffice for bottom 5.

Top 5 Departments for Hotness			Bottom 5 Departments for Hotness	
1	Women		1	African Studies
2	Dramatic Arts		2	Russian Studies
3	Slavic Languages & Literatures		3	Media Studies
4	Media Arts		4	Mgmt Science & Engineering
5	Archaeology		5	East Asian Languages

7. Obtain a table akin to the one provided above, but here only include Departments that have a minimum of 10 reviews. This will prevent low counts from adversely influencing the rankings. (3 pts)

Top 5 Departments for Hotness					
Department	Total	Hot	Not Hot	% Hot	% Not Hot
Physical Education	14	8	6	57.1	42.9
Social Science	55	26	29	47.3	52.7
German	13	6	7	46.2	53.8
Experimental College	13	6	7	46.2	53.8
Physical Ed	25	11	14	44.0	56.0
French	23	10	13	43.5	56.5
Ethnic Studies	18	7	11	38.9	61.1

Bottom 5 Departments for Hotness					
Department	Total	Hot	Not Hot	% Hot	% Not Hot
Geography	26	4	22	15.4	84.6
Chinese	14	2	12	14.3	85.7
Public Policy	50	6	44	12.0	88.0
International Studies	17	1	16	5.9	94.1
Finance	20	1	19	5.0	95.0
Journalism	22	1	21	4.5	95.5

*Departments with 10 or more reviews

Top 5 Departments for Hotness*			Bottom 5 Departments for Hotness*	
1	Physical Education		1	Chinese
2	Social Science		2	Public Policy
3	German		3	International Studies
4	Experimental College		4	Finance
5	Physical Ed		5	Journalism
*Departments with 10 or more reviews				

8. Use a Stacked Google Bar Chart (code provided below) to create a visualization of the table provided in the problem above. In order to successfully make this chart, you will need to calculate the % Hot and % Not Hot for each Department that is included on this graph.

- Top 5 and Bottom 5 as sorted by % Hot.

Department	Total	Hot	Not Hot	% Hot	% Not Hot
Physical Education	14	8	6	57.1	42.9
Social Science	55	26	29	47.3	52.7
German	13	6	7	46.2	53.8
Experimental College	13	6	7	46.2	53.8
Physical Ed	25	11	14	44.0	56.0
French	23	10	13	43.5	56.5
Chinese	14	2	12	14.3	85.7
Public Policy	50	6	44	12.0	88.0
International Studies	17	1	16	5.9	94.1
Finance	20	1	19	5.0	95.0
Journalism	22	1	21	4.5	95.5

- The Javascript vector elements used to create the Google Stacked Bar Chart

```
['Physical Education',57.1,42.9],
['Social Science',47.3,52.7],
['German',46.2,53.8],
['Experimental College',46.2,53.8],
['Physical Ed',44,56],
['Chinese',14.3,85.7],
['Public Policy',12,88],
['International Studies',5.9,94.1],
['Finance',5,95],
['Journalism',4.5,95.5]
```

- The Google Stacked Bar Chart of % Hot and % Not Hot for the Top 5 and Bottom 5 departments.