

Author's Response to Reviewers' Comments
Submission COST-D-14-00128R1
Amelia McNamara

Author's overall comments to reviewers: Thank you so much for all your thoughtful and detailed comments! I know this was just a minor revision, but I feel like these changes made a big difference in the quality of my paper.

Reviewers' comments:

Reviewer #1:

Page 4, Figure 1 - Is this the Google Maps API? Be sure that there are no copyright issues

Given Reviewer #2's comments and my concern about the API, I chose to replace the maps with a faceted set of line plots for each community. This unfortunately does not show the geographic distribution, but it does make it easier to make exact comparisons between communities.

Page 8, Figure 3 - Very nice, I like the color scale. Is the X axis for this kind of plot standard? I understand each bar is 100%, so the X axis is a little confusing to me.

This is a fairly standard practice. In the reference by Robbins and Heiberger, this type of X-axis scale is shown. My motivation for adding it was to ensure that readers could tell that each bar was 100%, and have reference lines for standard percentages.

Page 11, Table 2, Page 13, Table 3 - I wonder if this table can be presented in a better way - Rather than repeating the grouping variable (i.e., senior or non-senior) maybe this can be split in some way. It's not bad the way it is but I'm imagining how it would be presented if instead there had been, say, 10 years in the dataset.

Thanks for this comment, it really made me rethink these tables. I've now replaced them with grouped bar charts, which I think convey the same information even more clearly, and could be extended to more years if additional data were available (and, since this is a reproducible article, it's good to support the addition of more data!). I've also changed Table 5 to be in line with Tables 2 and 3. Table 4 has been left a table because I intended for that to be additional support for figures 10 and 11. (All table and figure numbers refer to my previously-submitted draft, not the current one.)

Page 13, Lines 28-30. I agree the data shows this, but given that the data suggests negative feelings towards communities like Gary and Macon, and positive feelings towards State College, relative to the overall difference in ratings, are the ratings for a good place to raise a family particularly high for State College and low for Gary and Macon?

I'm not sure I fully understood this comment, but upon re-reading this section I realized my sentences did not really make sense. I've now re-written it to make it clearer.

Page 13, Line 43 - Need space after Figure 10. Added.

Page 16, Figures 10 and 11 - Was the choice to not assign any kind of color to the bars intentional? I understand that it can be a touchy issue because of dealing with race, but I

think it might be nice to spruce up those plots a bit. Also, the y axis scale should go up to 100% just so the relative sizes of the bar are not misleading.

I have extended the y-axis scale to 100%, but I have chosen to leave the black-and-grey scale. Adding color to the bars would not contribute any additional information, and the printed version of this document will not be in color.

Page 17, Line 36 - meta-knowledge - Actually i think this occurs elsewhere in the document, sometimes including the hyphen and sometimes not.

This has been changed to be consistently “meta-knowledge” rather than “meta knowledge” throughout the document.

Page 17, Line 44 and 45 - Respondents Fixed here and a few other places, and ran an additional spell check.

Page 20, Figure 14 - Maybe make the figure extend from (0, 0) to (1, 1). Possibly for figure 15 and 16 as well.

I looked at this for all three graphs, and in each case the data did so little space-filling that it was hard to see the individual points (instead, all the values were clustered in one area and there was a lot of blank space everywhere else). I also considered standardizing the scales in some other way but couldn’t find a scale that made sense in all three cases.

Reviewer #2:

The paper is good but need some revision. The ideas and some of the plots presented in the paper are great. However deeper research is needed to better define those concepts and to relate them with the community attachment. For instance the Meta-Knowledge should be clearly defined, not only thought examples including a reference explicitly in the text.

There are some general suggestions about how the paper is presented. The order of plots has to be improved because difference among where the plot and where is referred in the text makes harder to read the paper.

I’ve done what I can to improve the plot positioning, but most of that is due to L^AT_EX.

Perhaps it is possible to combine the yearly information when there is no different patten on each year. There are some plot legends that mixed numbers and words that should be homogenized and the table captions need to be move up to the of the table.

I am not sure what you mean about “mixed numbers and words.” I have used the ASA style guide, including the style of writing out numbers less than ten and using arabic numerals for numbers greater than ten, while also keeping statistics such as the mean in arabic numerals. Most of the tables have now been removed, so hopefully that addresses your comment about the table captions— there seems to be a typo in your comment so I don’t fully understand it.

Section 2: The data.

This section has 5 pages long but is not clear why is important or interesting, we think should be shorter. About Figure 1, you may want to choose one year (or combine the information)

since the overall structure seem to be stable, if you want to highlight the change over time for some communities then Figure 1 is not the best options, is hard to see the yearly evolution for each community. Probably Gallup has some methodological reasons for use the survey design they are using, if you think that these patterns in survey rates may bias the data in some way you need to be clear in which way.

Thank you for this comment. I did want to highlight the change in survey rates, but I realized that a set of maps was probably not the best way to display this. Instead, I've created a set of faceted line plots. This loses the geographic structure of the data (i.e. you can't see the West-East relationship) but it makes it easier to make numeric comparisons.

As for the length of the section, because this was a data-driven project, and the data does not represent a random sample, I felt it was important to explain some of the exploratory data analysis that went into developing the questions I went on to answer.

Figure 2 has not clear message, why the number of items in a question is important? We think there is no need for this figure.

Again, this is a representation of the exploratory data analysis that was done to start this project. It is also interesting to note because the varying length of response items suggests that the survey would have been hard to complete and that there may be data that is even less standardized than usual Likert scale data.

Missing data analysis is nice. You may want to explain how you count the missing values more explicitly. For example in 2008 you have only 6 missing observations. This means 6 totally incomplete observations? Or a row is consider as missing when have at least one missing response?

I've made this explicit in the text. One particular survey question was used as an indicator for missing data in 2010, and the same question was used in 2008 and 2009 for consistency, although those data sets were not sparse in the same way.

Section 3: Community Satisfaction

Explain clearly why you choose community satisfaction instead of attachment.

This is explained in Section 2— community attachment is an arbitrary composite measure created by Gallup, while community satisfaction is a question answered by respondents directly.

Figure 3 you should include a general comment about community satisfaction, for example: What is the range of community satisfaction percent across communities? I have added this.

Is difficult to see the community change across years with this plot and actually the overall structure seem to be stable over time. You may want to combine the 3-years information or select a year. It is not clear, the comment about Bradenton in this plot (satisfaction rates go from 77% to 80%)

Although I take your point that the overall structure is rather stable over time (and variation over the years may be random) I like the ability to compare across years, so I am keeping all

three plots. I have added some explicit references to values in communities that increased (Bradenton) and decreased (Macon) to make it clearer what I am indicating.

Section 4: Behaviors

What is the relation of this section with the rest of the paper? Are these behaviors related with the attachment? or the community satisfaction? or the meta-knowledge?

With the time to revise the paper, I have integrated the results from this section into the section of the paper. I thought that behaviors might be another factor that improved community satisfaction (again, I do not like the attachment variable, I am only referencing it in this paper because of Gallup's use of it).

Figure 4 do not need the 3 years, only 2010 is enough. Figure 5 looks interesting, you should highlight some interesting facts suggested this plot. Currently the plot show the rates in 10 different behaviors categories but only the shelter issue is commented in the text. For example Palm Beach, Miami, Fort Wayne and Akron present positive rate bigger than the average in most of the dimensions.

I've chosen to keep all three years to make it clear that not all the questions were present in the survey in 2008 and 2009. Otherwise, I agree with you that the pattern is the same. I have added additional exposition about the difference from averages.

Section 5: Meta knowledge

The MK index from equation 1 not depends of the specific dimension (senior, race, family) you should change the order and include a general definition at the beginning of the section. As a way to measure the meta knowledge concept.

Thanks, this makes sense. I've moved the MK index up in the paper, and introduced subscripts to indicate the dimension.

Why only race is include in the MK index? Our suggestion is to compute the index for the tree dimensions, and use all of them to characterize the meta knowledge of a particular community. In particular we suggest changing figures 7, 9 and 13 by figures type 14, using the MK index on each dimension.

I did change the paper to use all the subgroups for MK calculations. In response to the comment about the plots, I still want to maintain the ability to see the overall response distribution for each community, but I did add additional scatterplots to make the overall trends more clear.

Respect to the distribution about race question you can comment the response rate but you included two plots and two tables to show this, is too much (Figure 10, 11 and table 4 and 5), you may want to reduce this part with a comment and a table instead.

I agree that it was overkill before. I've taken out one of the plots, one of the tables, and converted one table to a plot (following a suggestion from Reviewer # 1). I also modified the remaining plot to include information from the table (absolute numbers). Again, because this is a submission to the Data Expo special issue, I have chosen to keep plots rather than

tables in most instances.

The linear model for the community satisfaction needs to be improved. You can use all the years information (why you select only information form 2009?), the 3 meta knowledge indices and include other variables (Gallup metrics variables, demographic ones etc). The model response is a rate, and some of the values are close to 1, may you can try a transformation of the response (a logit transform for instances) or using a generalized linear model instead with a beta distribution. In any case after select a model and fit you should comment on the model goodness-of-fit.

I've actually chosen to remove the models, as I realized you were right about the necessity of taking a logit transform or using glm with a beta distribution. The coefficients in those models became complex enough that I didn't think it was worth the effort for the interpretation, particularly because this model is not something that is useful for prediction.

My only goal for producing a model was to provide more explanatory suggestions about relationships between variables. In particular, because it is clear that the association is not causal and predicting community satisfaction isn't a goal in and of itself, it doesn't necessarily make sense to make predictions.

Instead, I've included more plots and information about correlation.