Computational Statistics manuscript No.
(will be inserted by the editor)

Community engagement and subgroup meta-knowledge: Some factors in the soul of a community

Amelia McNamara

Received: date / Accepted: date

Abstract The Knight Foundation collects data to determine what factors impact community satisfaction, local GDP growth, and interest in Knight news publications. For the 2013 Data Expo at the Joint Statistical Meetings, many participants created graphical explorations of these data. This article focuses on the idea of community meta-knowledge, which is essentially majority group empathy or understanding of how minorities experience their community. For example, the survey asks participants to rate their community "as a place for senior citizens," on a 5-point Likert scale. A city where seniors rated their community in the same way as non-seniors is defined as a community with high meta-knowledge about conditions for seniors. Three minority groups were explored: seniors, families with young children, and racial minorities. In most communities, people outside the minority group tended to under-rate their community, compared to those in the minority group. However, there were some exceptions.

Keywords 2013 Data Exposition, R, ggplot2, Likert scales, meta-knowledge

1 Introduction

Studies have shown that increasing empathy is the best way to improve intergroup relations (Stephan and Finlay, 1999). Therefore, it is of interest to quantify the typical level of empathy in communities across the United States. The Knight Foundation data provides a window to a factor which could be thought of as a proxy for empathy, namely community-meta knowledge. We are defining meta-knowledge as community awareness by those outside a specific subgroup about the conditions for people inside the subgroup. The primary

A. McNamara
8125 Math Sciences Bldg.
Los Angeles, CA 90095-1554
E-mail: amelia.mcnamara@stat.ucla.edu

attempt of this article is to answer the question "Are people outside a specific subgroup aware of the quality of their community for people in that subgroup?" and then, "do communities with high meta-knowledge (those where people outside the subgroup understood conditions for minorities) have higher community satisfaction rates than those with low meta-knowledge?"

This article is one of several related to the Knight Foundation community data from the 2013 Data Expo. For more information on the Expo and the data sets, see (Hofmann and Wickham, 20XX).

2 The Data

The Knight Foundation collects survey data on 26 communities where the Knight brothers own newspapers, including San Jose, CA, State College, PA, Palm Beach, FL, and St. Paul, MN. The foundation has data for three years, starting in 2008. Each data set includes approximately 20 demographic questions and 50-80 survey questions, depending on how distinct questions are defined. The aim of the survey is to gauge what factors are important to community attachment, and it includes questions on a variety of subjects, from "how satisfied are you with this community as a place to live?" to "how many minutes is your daily commute?"

The survey is conducted over the phone by Gallup Poll, and can take place in either English or Spanish. Gallup also performs data analysis for the Knight Foundation, and their yearly reports are available on the Knight Foundation website (Gallup Consulting, 2008; Gallup Poll, 2009, 2010).

The existing data analysis from Gallup is related to a metric they call "Community Attachment." It is difficult to pin down what this variable means, but it's a composite metric composed of Community Loyalty and Community Passion. Both of those metrics, in turn, are composed of several variables. Community Loyalty includes how likely a person says they are to stay in that particular area, how much they would recommend it to friends, and their outlook for the community's future (Gallup Poll, 2010). Community Passion is composed of variables on connectedness and community pride. So, Community Attachment is already a model of what Gallup believes is important to strong communities. The Gallup team has discovered that this composite variable is positively correlated with local Gross Domestic Product (GDP) growth (Knight Foundation, 2010). Because of this relationship, the analysis from Gallup is focused on what other factors correlate with Community Attachment (and therefore, with local GDP growth).

While the Gallup Poll analysis is interesting, it does raise the question of multicolinearity, as factors that are correlated with Community Attachment may simply be correlated with one of the variables that was used to compose it, and may not actually have an impact on local GDP growth.

2.1 Community Survey Rates

As mentioned above, the data were collected by Gallup through telephone surveys in 2008, 2009, and 2010. Participants were a random sample of adults living in 26 "communities" (cities or metro areas of the United States), and at least 400 people were surveyed in each community. The data from 2008 and 2009 had 13822 and 13728 responses, respectively, while the data from 2010 contained 20271 observations. Because the data sets surveyed the same 26 communities, we can calculate the average number of survey participants in each community. In 2008, that average was 531 people, in 2009, 528 people and in 2010, 779 people. The difference in average number of survey participants will be discussed further in Section 2.3.

In most communities, approximately 400 people were interviewed, but certain communities were surveyed much more. It appears that the Knight Foundation was trying to survey places at an approximately similar rate, which is why Philadelphia (for example) was surveyed 1633 times in 2010. To see which places were over- or under-represented in the survey, see Figure 1 for maps showing the percentage of the community that was polled for each polling year.

The population data for the maps in Figure 1 came from the Intercensal population estimates compiled by the Census Bureau. Population estimates are calculated each year between Census years. The estimates do not vary much from the decennial Census count, but yearly estimates were used for the sake of having different population numbers from year to year. For example, the 2010 Census count of the population of Palm Beach, FL was 8,348. In 2009, the estimate was 8,456, and in 2008 it was 8,631.

The maps in Figure 1 show the percentage of the community that was polled, and percentages hover around a mean of 0.07%, with lots of variation. Palm Beach, FL always looks over-represented because the community is small in absolute number of residents and the minimal sample size of 400 was always used, leading to a polling rate around 4%. Large communities like Philadelphia, PA, look under-represented, with a rate around 0.01%. In addition, there is some variation over time, especially on the East side of the US. For example, Akron, OH begins with a polling rate of 0.01%, which rises to 0.07% and then 0.09%, as a result of polling increasing from around 400 residents to more than 1700. It's not clear why Gallup made these polling decisions.

2.2 Scale Lengths

The majority of the Knight data is in the form of responses to survey questions, and most survey questions were answered on a Likert scale (Likert, 1932). However, there was little consistency in the number of levels for the scales. The most common scale was a five-point scale, as in "Not at all satisfied, 2, 3, 4, Extremely satisfied" or "Very bad, 2, 3, 4, Very good." However, many

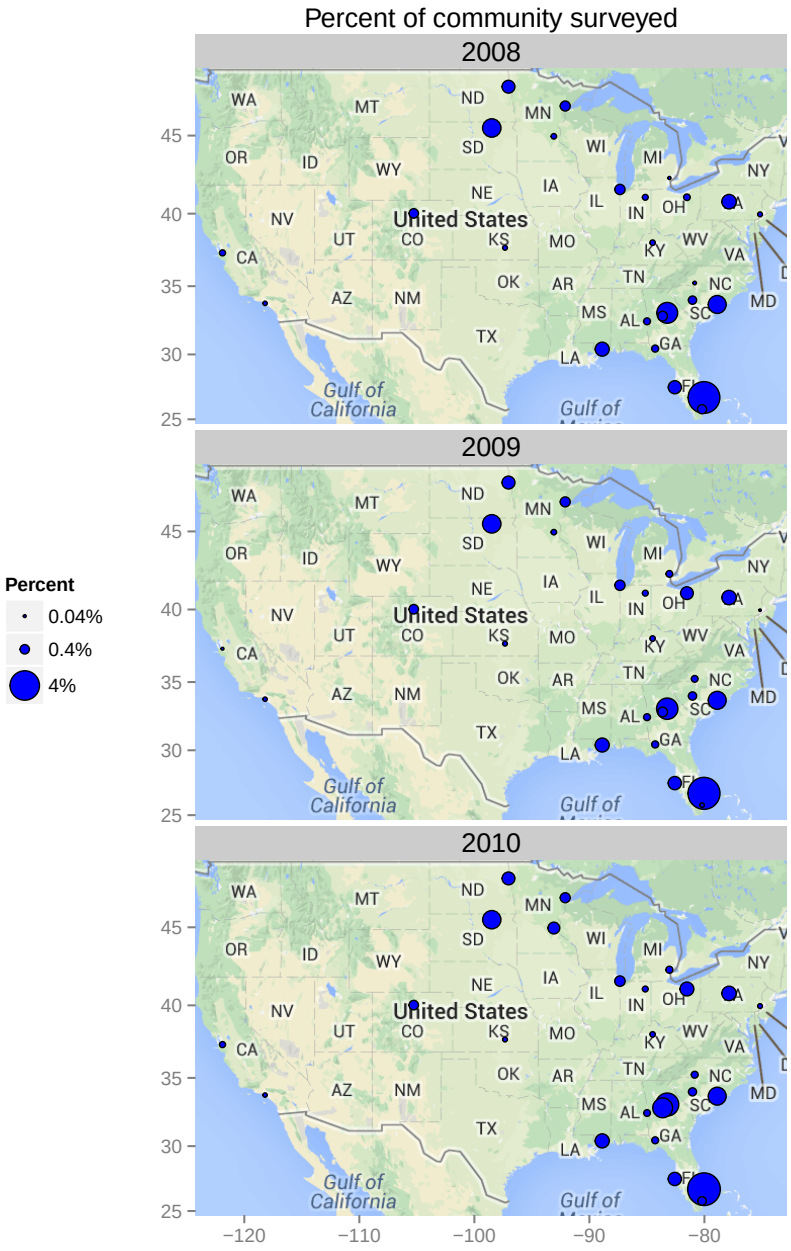

Fig. 1 Maps of yearly survey percentages. The larger the bubble, the larger the percent of community residents surveyed in that year. Notice that some communities are always over-surveyed (for example, Palm Beach, FL) and some always appear under-surveyed (for example, Lexington, KY). The East side of the United States experiences an overall raise in survey rates over time, while the West side stays stable.

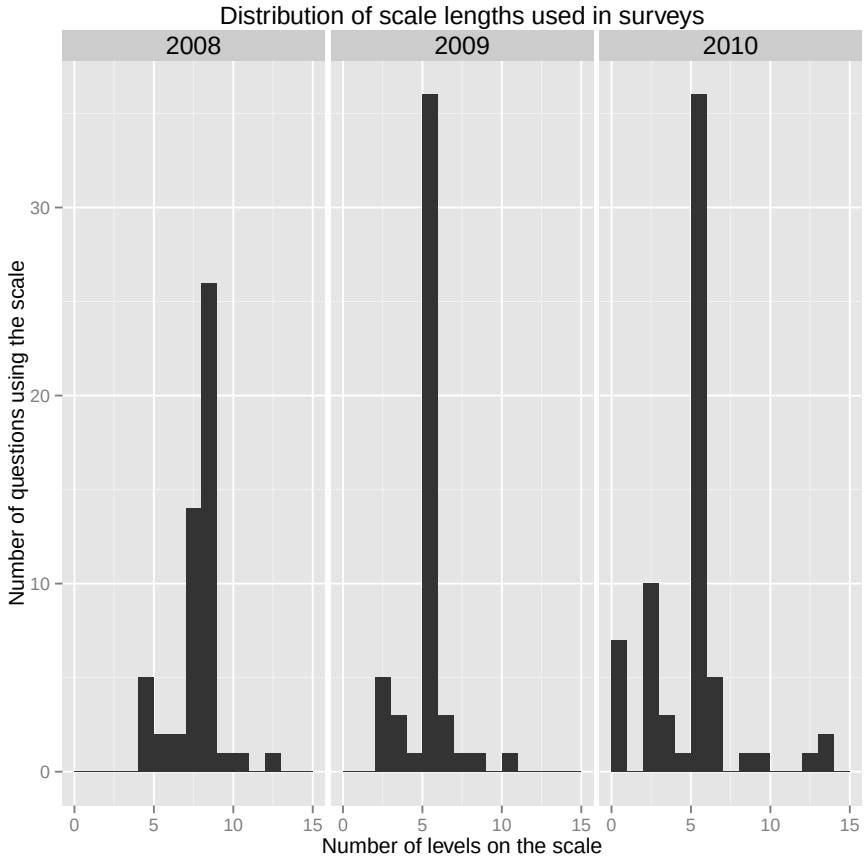

Fig. 2 Scale length distributions for each year of the survey. Notice that the distribution from 2008 is centered around 7, and the 2009 and 2010 distributions are centered around 5. All three distributions have large variation, suggesting that the Knight surveys were quite complex.

other scales wordings (and scale sizes) were used. For the yearly distribution of scale lengths, see Figure 2.

The varied lengths of response scales and the different phrasing of scales even with the same length suggests that this survey was quite long and complex to complete. And though the survey maintains the scale lengths for individual questions over the years, Gallup rescales all the questions down to a 3-point scale to make their analysis simpler. The complete data set provided by the Knight Foundation includes between 156 and 206 variables, depending on the year, but fully half of them are rescaled versions of the original questions. While some researchers have suggested that a 3-point scale is enough (Jacoby and Mattel, 1971), discarding data seems wasteful, especially if participants

have gone to the trouble of rating on a 5- or 7-point scale. So, the remainder of this analysis works on the unscaled variables.

2.3 Missing Data

While the Gallup reports claim the telephone surveys only took 15 minutes, the number of variables collected and the wide range of response scales seems to indicate a much larger time commitment. This raises the question of whether everyone who began the survey completed it. And, as mentioned in Section 2.1, the 2010 data contained many more observations than previous years. The explanation for this difference is missing data, presumably related to surveys that were not fully completed. Table 1 shows the percentage of missing data for each survey year. While the 2008 and 2009 data sets are almost complete, the 2010 data set has about 25% missing data.

Year	Missing	Total	Percent
2008	6	13822	0.04
2009	63	13728	0.46
2010	4947	20271	24.40

Table 1 Percent of missing data from surveys. The 2010 survey has almost 25 percent missing data. The total number of entries in 2010 is also much larger, suggesting that the 2008 and 2009 datasets used a different criteria for inclusion of entries.

The missing data in the 2010 data set is characterized by almost all the demographic information being present, but only one survey question answered (that being, "how satisfied are you with this community as a place to live"). Interestingly, with the incomplete responses removed, the 2010 data set is reduced to 15,000 observations, which is much closer to the 14,000 observations the two prior years. This suggests that incomplete responses were removed in previous years, or that some new survey methodology (i.e. a "short form") was introduced in 2010.

3 Community Satisfaction

Knowing that the question about community satisfaction was the only survey question answered by all respondents, it made sense to see which communities reported the highest levels of community satisfaction.

To visualize this, a set of stacked distribution graphs were created (Robbins and Heiberger, 2011). These stacked distribution graphs are centered around zero and use a diverging color scale to give an overall graphical sense of the amount of positive and negative responses across groups. Figure 3 shows the distribution of responses to the question, "Taking everything into account, how satisfied are you with this community as a place to live?" and is ordered

by the communities with the largest total positive responses in 2008, which highlights the changes in 2009 and 2010.

Looking at Figure 3, we can see that people in State College, PA typically report much greater levels of community satisfaction than people in Detroit, MI or Gary, IN. Brandeton, FL, shows an increase of community satisfaction in 2009 and 2010 compared to 2008, while Milledgeville and Macon, GA show a decrease over the same time period.

4 Behaviors

Another point of interest was the most common behaviors reported by participants. Figure 4 shows the percentage of participants engaging in a variety of behaviors over the three years of the survey. An additional set of questions were introduced in 2010, so those are necessarily blank in the previous years.

Behaviors are arranged by percentage of survey respondents who reported the behavior. Over all three years, the most common behavior was being registered to vote, followed by voting in a local election. The least common responses to the behavior questions (considering all three years) were "worked with other residents to make change in the local community", and "attended a local public meeting in which local issues were discussed." When the additional questions were added in 2010, an even-less-common behavior was added, "gave money or food to an individual in need in your community who is not related to you."

A followup to this question is whether all communities performed these actions at similar rates, or if there were local variations in behavior. Figure 5 shows the difference from the overall rate across all 26 communities and 10 behaviors in 2010. More study is required to determine if the visual differences between communities represent true differences or just random variation, but there are certainly communities that stand out from the rest.

For example, people in St. Paul, MN, San Jose, CA, and State College, PA were much more likely than the average to to provide free shelter to a non-relative, while people in Georgia (both Macon and Milledgeville) were less likely to. This provides a connection to Figure 3, because Macon and Milledgeville were the two communities whose community satisfaction scores decreased the most in 2009 and 2010, while San Jose was one of the most satisfied communities over all three years. One hypothesis is that St. Paul, San Jose, and State College are all university towns, where young people (especially college students) may invite friends and friends-of-friends to stay with them for free. Again, more study is required.

5 Meta Knowledge

The primary aim of this article was to address whether communities held meta-knowledge about their city being a good place for subgroups or minorities.

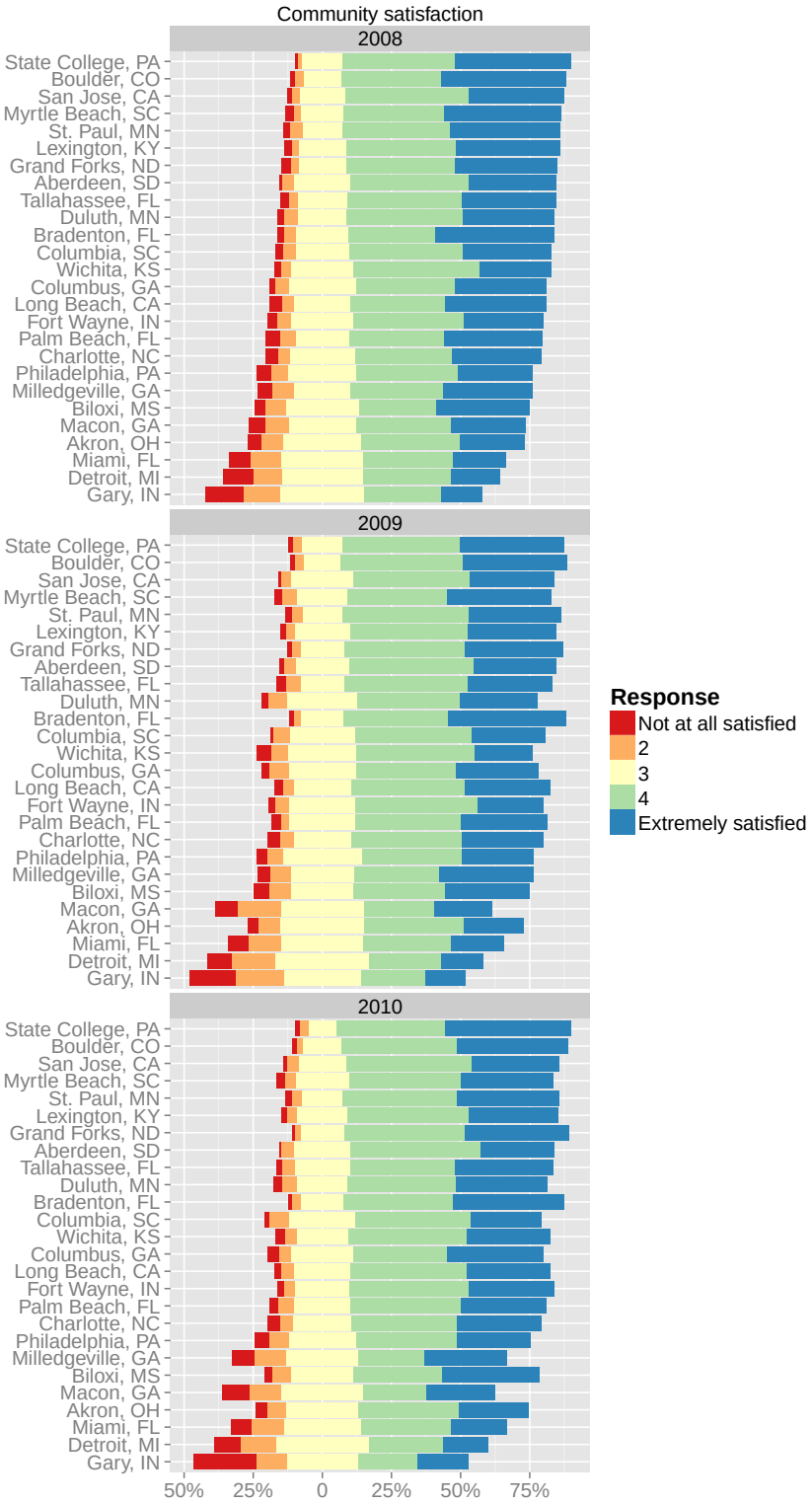

Fig. 3 Responses to the question, "Taking everything into account, how satisifed are you with this community as a place to live?" Communities are ordered by percentage of positive responses in 2008, making it clear the differences in distribution in 2009 and 2010.

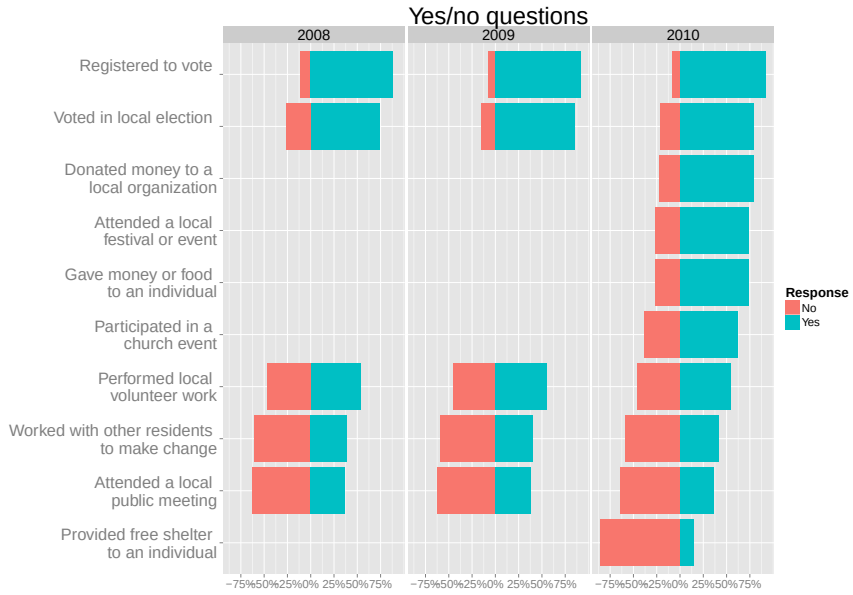

Fig. 4 Responses to yes/no questions about participants' behaviors, comparing all three survey years.

The survey asks a number of questions related to rating the community as a place for subgroups, including: "young, talented college graduates," "immigrants from other countries," "racial and ethnic minorities," "families with young children," "gay and lesbian people," "senior citizens," and "young adults without children." Not all these subgroups were asked to identify themselves in the demographic questions (particularly "gay and lesbian people") so it was not possible to address them all. Instead, we focus on senior citizens, families with young children, and racial and ethnic minorities.

Community meta-knowledge is essentially majority group empathy or understanding of how minorities experience their community. For example, the survey asks participants to rate their community as a place for families with young children on a 5-point Likert scale. A city where participants with children rated their community in the same way as participants without children is defined as a community with high meta-knowledge about conditions for families with young children.

5.1 Meta knowledge about the community as a place for seniors

In order to determine if non-seniors understood how good their community was for seniors, the data was split into two pieces, one of participants aged 55 and older, and the other of participants under 55. For sample sizes of the

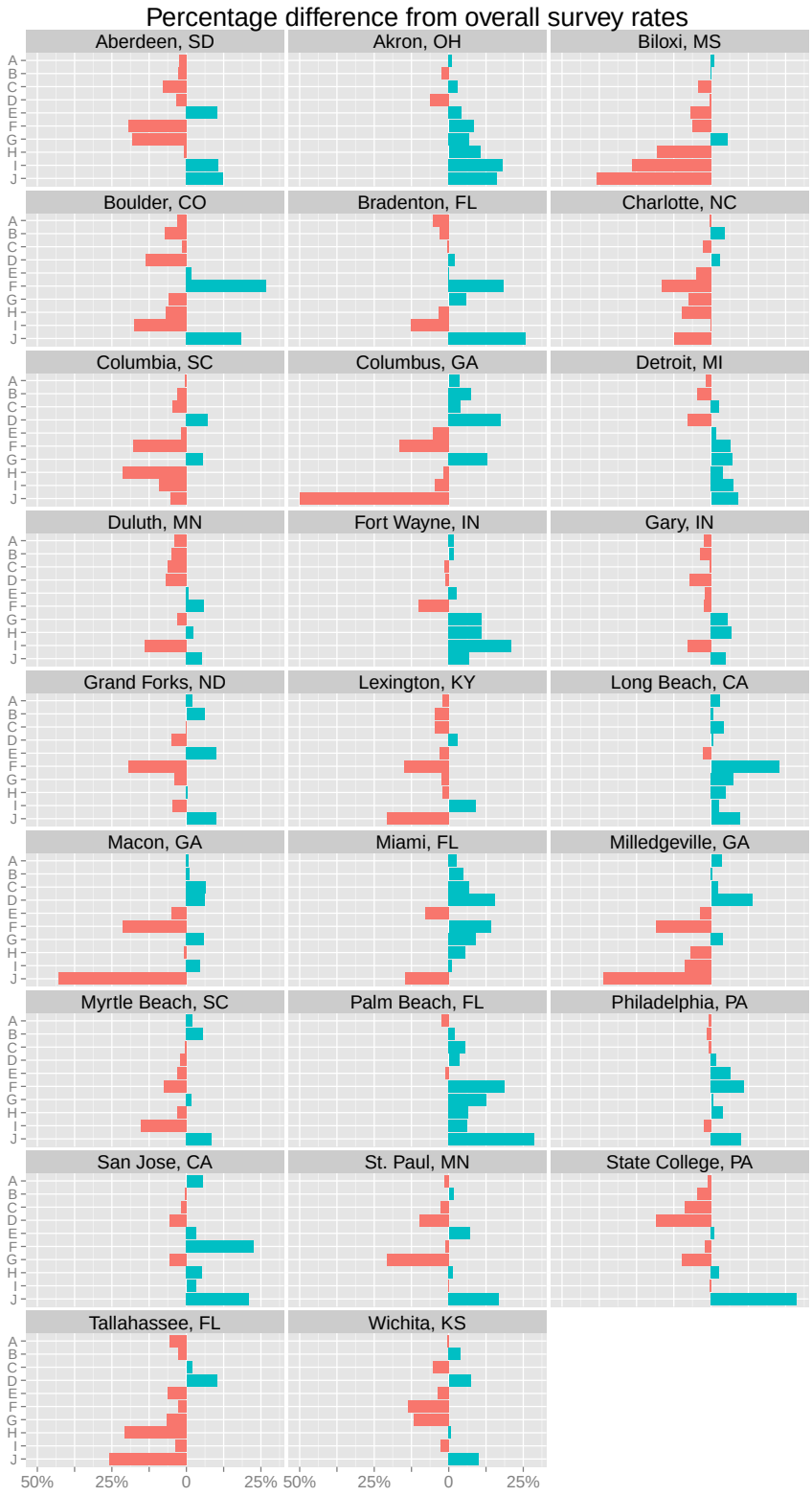

Fig. 5 Difference from overall survey rates (2010 data). The letters A-J represent the activities listed in Figure 4. A: Registered to vote, B: Voted in a local election, C: Donated money to a local organization, D: Attended a local event, E: Gave money or food to an individual, F: Participated in a church event, G: Performed local volunteer work, H: Worked with other residents to make change, I: Attended a local public meeting, J: Provided free shelter to an individual.

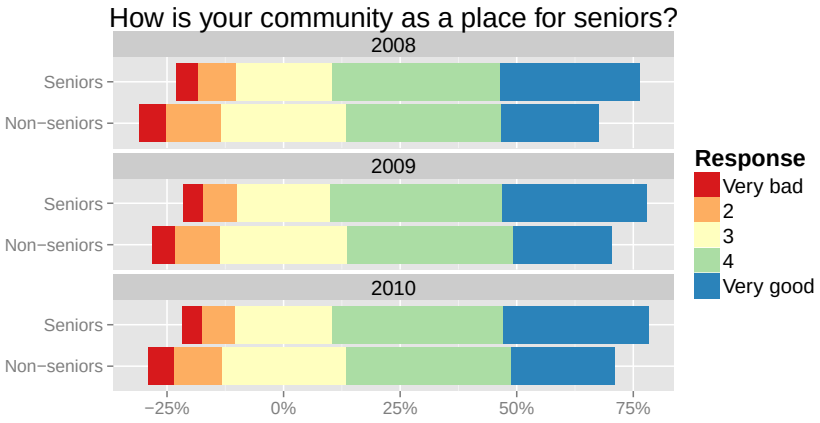

Fig. 6 Responses to the question, "How is your community as a place for seniors?" Seniors are defined as survey participants aged 55 and older, non-seniors are those under 55. Notice that seniors consistently rated their community more highly than did non-seniors over all three survey years.

groups, see Table 2. Interestingly, the split between seniors and non-seniors was roughly 50/50 every year.

Year		Number	Percentage
2008	Seniors	7241	52.39
2008	Non-seniors	6581	47.61
2009	Seniors	6327	47.32
2009	Non-seniors	7044	52.68
2010	Seniors	6974	45.61
2010	Non-seniors	8318	54.39

Table 2 Sample sizes for plots about meta-knowledge about seniors.

The yearly ratings distributions are shown in Figure 6. It appears that non-seniors typically underestimate how good a place is for seniors. People 55 and older rated their community as a better place for seniors than did people under 55, over all three years.

Again, to investigate the local variations in responses, a faceted plot of responses between the two groups was created. This plot can be seen in Figure 7, and it uses 2010 data. Interestingly, almost every community followed the trend of non-seniors underestimating how good their community was for seniors (or seniors boosting their responses). There was only one exception to this rule– Milledgeville, GA. However, this could be due to random variation.

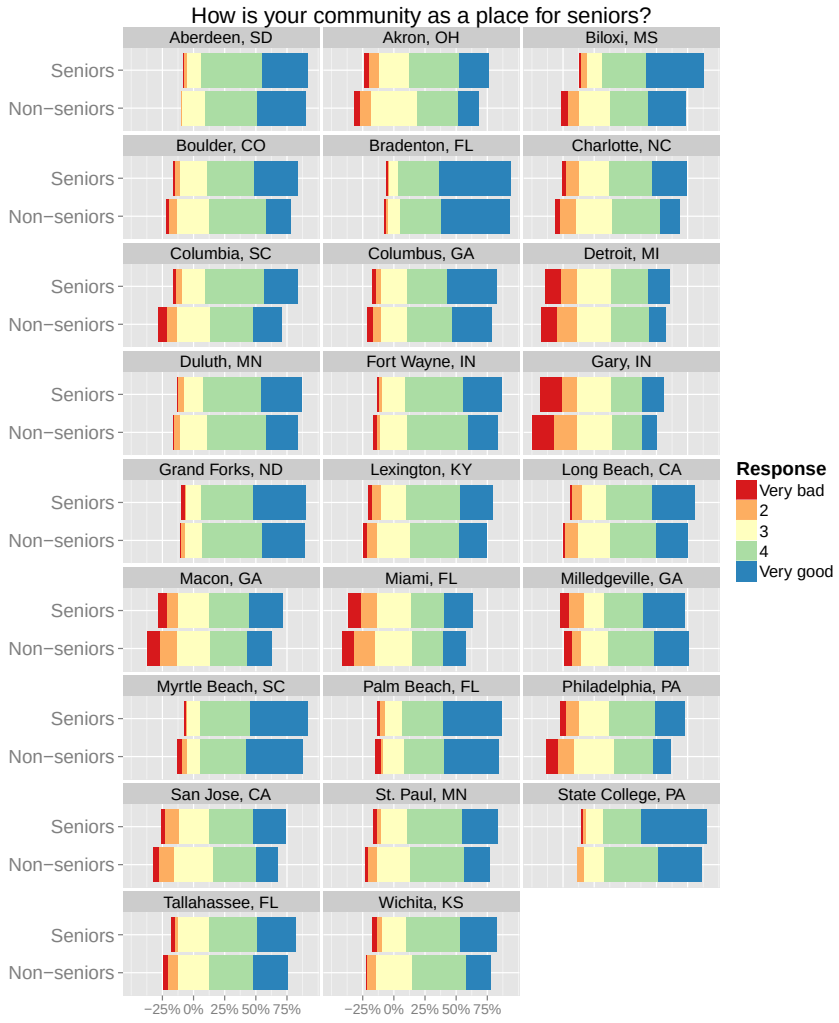

Fig. 7 Responses to the question, "How is your community as a place for seniors?" faceted by community (2010 data). Every community follows the pattern of over-rating by seniors, but some communities have a smaller discrepency between ratings of seniors and non-seniors.

5.2 Meta knowledge about the community as a place for families with young children

For the exploration of meta knowledge about communities as a place for families with young children, we split participants between those who reported having dependent children under the age of 18 living in their household and those who did not. Intuitively, it makes sense that a parent of an older child (say, a teenager) would have higher meta-knowledge about the community as

a place for families with young children than a participant who never had children or whose children have grown up and moved away. While the data included more granular demographic details about the ages of the children in the households, splitting the data into participants with children and those without made the groups closer in size. To see the sample sizes and percentages, see Table 3.

Year		Number	Percentage
2008	Childless households	9372	69.16
2008	Families with children	4179	30.84
2009	Childless households	9610	71.43
2009	Families with children	3844	28.57
2010	Childless households	11179	73.80
2010	Families with children	3969	26.20

Table 3 Sample sizes for plots about meta-knowledge about families with young children.

Figure 8 shows the difference between the ratings of the groups over the three years of the survey. As in Section 5.1, we can see that in-group ratings were slightly higher than out-group ratings, but the difference is not nearly as significant as in Section 5.1 on seniors. In fact, both groups seem to give about the same ratings overall.

Again, we broke this down by community for the 2010 data, as seen in Figure 9. Communities generally reflect the larger trend of good meta-knowledge outside the subgroup. Interestingly, though, the variation apparent in Figure 9 is more about the overall ratings of the communities as places for raising kids. There was agreement both inside and outside the subgroup, but people seem to agree that State College, PA is a good place for families with young children, and Gary, IN and Macon, GA, are not.

5.3 Meta knowledge about the community as a place for racial and ethnic minorities

Investigating meta knowledge about racial and ethnic minorities was more difficult than the previous subgroups, because each year of data collection used a slightly different set of possible answer choices to the question "Which of these groups best describes your racial background?" and because there were so many participants who refused to answer (particularly in 2010).

To see the overall distribution of responses to the demographic race question, see Figure 10.Notice that in 2010, the largest category was "Refused," so Figure 11 shows the distributions without "Refused" responses. For absolute numbers and percentages for each of the race responses, see Table 4. As the table shows, the sample sizes for individual minority race responses were somewhat small each year, so for the plots about subgroup meta knowledge, we combined all the minority responses into one group that we refer to as Nonwhite. This category contains all participants who reported a race that was not

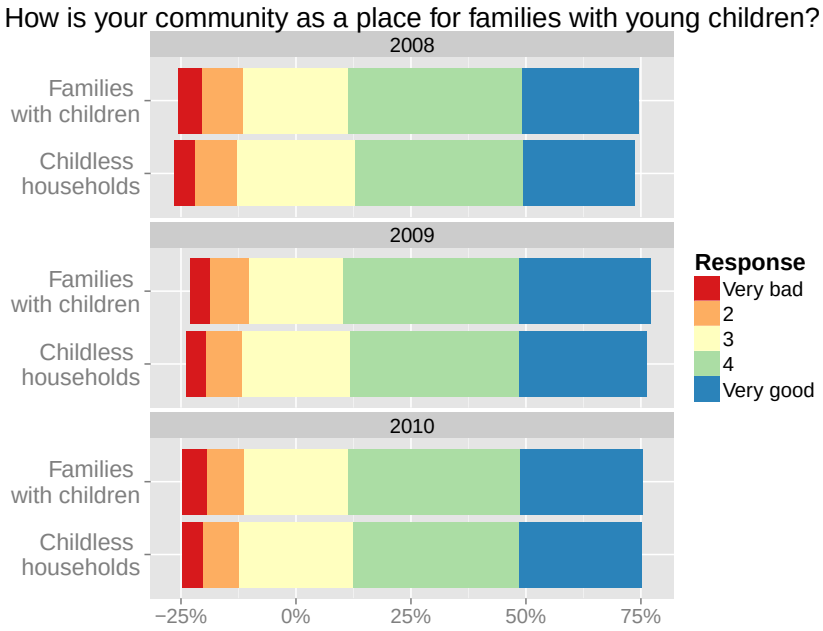

Fig. 8 Responses to the question, "How is your community as a place for families with young children?" Families with children are defined as any household with children under the age of 18, Childless households are any households without children, whether or not the residents have adult children living elsewhere. Ratings were consistently similiar between groups and across survey years.

White, but does not include participants who declined to give a response to the question. For a condensed summary of what the White/Non-white criteria means for the overall percentages, see Table 5.

As with the previous explorations of meta-knowledge, we look at the overall difference between in-group and out-group responses to the question, "how is your community as a place for racial and ethnic minorities?" Again, the data is split into two groups, one called White and one called Non-white. For comparison of the distribution of responses between groups, see Figure 12. Interestingly, Whites were rating their communities as better places for minorities than Non-whites in 2008 and 2009, but in 2010 Whites began underrating their communities.

Again, we want to see the individual variation between communities. In the past few sections (on the quality of communities for seniors and families with children) we have been studying the 2010 data at the community level. However, as can be seen in Table 4, the 2010 data has extremely low response rates to the demographic question about race and therefore, we chose to look at the 2009 data for community-level variation. The community-level responses are shown in Figure 13. There is a lot of variation between the communities

Fig. 9 Responses to the question, "How is your community as a place for families with young children?" faceted by community (2010 data). Interestingly, most communities follow the pattern of similar ratings by both groups, but there is a lot of variation between communities in terms of the overall rating.

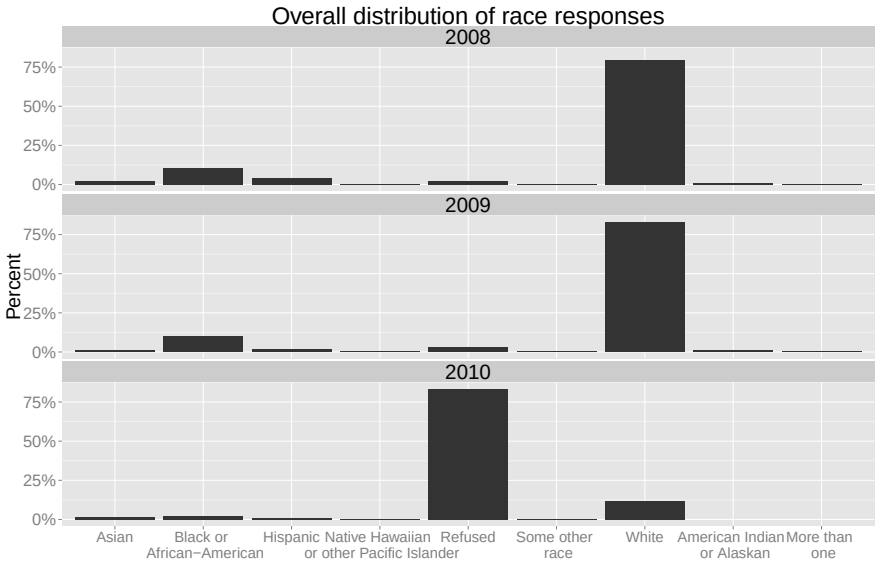

Fig. 10 Responses to the question, "Which of these groups best describes your racial background?" Notice that 2010 shows an overrepresentation of Refused responses, compared to the other years.

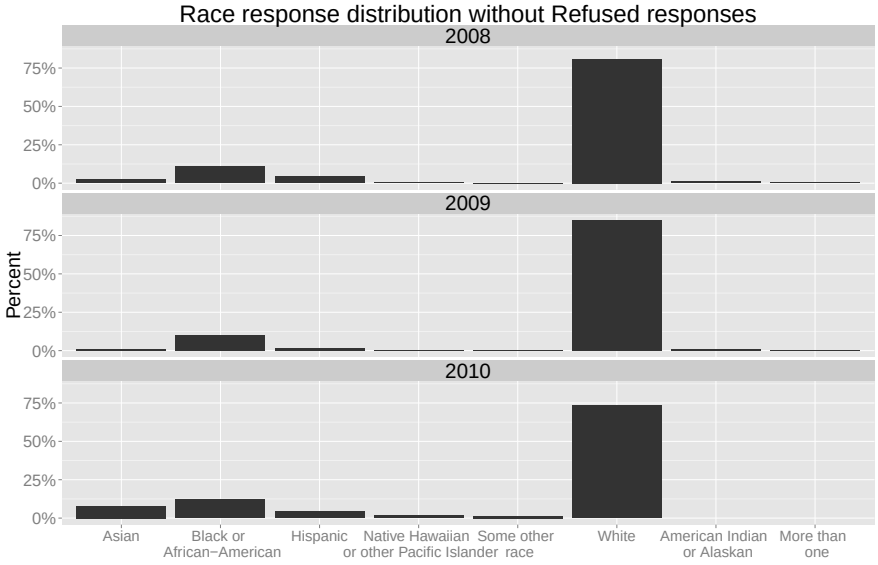

Fig. 11 The distribution of responses to the question, "Which of these groups best describes your racial background?" with Refused responses removed.

Race	Number of Responses	Percent	Year
Hispanic	563	4.17	2008
More than one	78	0.58	2008
American Indian or Alaskan	142	1.05	2008
Asian	314	2.33	2008
Black or African-American	1438	10.65	2008
Native Hawaiian or other Pacific Islander	36	0.27	2008
Some other race	12	0.09	2008
White	10915	80.86	2008
American Indian or Alaskan	150	1.12	2009
Asian	173	1.30	2009
Black or African-American	1363	10.22	2009
Hispanic	208	1.56	2009
More than one	53	0.40	2009
Native Hawaiian or other Pacific Islander	26	0.19	2009
Some other race	7	0.05	2009
White	11360	85.16	2009
Asian	214	7.84	2010
Black or African-American	326	11.95	2010
Hispanic	113	4.14	2010
Native Hawaiian or other Pacific Islander	45	1.65	2010
Some other race	36	1.32	2010
White	1995	73.10	2010

Table 4 Absolute response numbers and percentages for race responses, with Refused responses removed.

for this particular type of meta knowledge. Some communities saw the White respondents over-scoring their community as a place for minorities, while some were under-scoring. For the most part, however, Whites over-rated their communities as a place for racial and ethnic minorities, compared to Non-whites. Grand Forks, ND, was a particularly bad offender– Whites over-rated it as a place for minorities, and minorities themselves rated it as one of the worst communities in 2009.

There was a lot of additional variation in response distribution. For example, San Jose, CA is highly rated as a place for minorities both by people in- and -outside the subgroup. And it was slightly under-rated by Whites, which is probably a good sign of meta knowledge and empathy. Going the opposite direction are Grand Forks, ND, Myrtle Beach, SC, and Macon, GA.

Year	Percent White	Percent Non-white
2008	80.86	19.14
2009	85.16	14.84
2010	73.10	26.90

Table 5 Percent of dataset that is White versus Non-white. 2010 saw an increase in the percentage of survey respondants who were identifying as Non-white, but it also saw a marked decrease in any respondants reporting race, as can be seen in Table 4.

For another view of the relationship between ratings (as seen in Figure 13) see Figure 14, which shows the relationship between total positive responses by

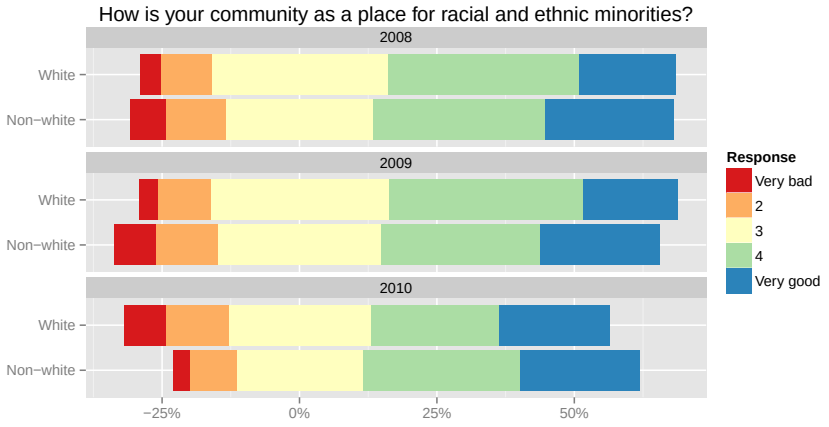

Fig. 12 Responses to the question, "How is your community as a place for racial and ethnic minorities?" White denotes survey respondants who listed their race as White, and Non-white is all other race responses (not including survey participants who refused to report a race). Notice the difference between the 2008/2009 responses and the 2010 responses, but also refer to Table 4 for the absolute sample sizes for each year– 2010 has a much smaller sample of responses to the question overall.

Whites versus total positive responses by Non-whites to the question, "How is your community as a place for racial and ethnic minorities?" Figure 14 makes it clear that there are communities that are under- and over-rated by Whites, and those that are closer to the 1-1 (or y=x) line. However, even if a community is on the 1-1 line, it may fall below the mean rating for communities overall. Gary, IN is a good example. While it is rated almost exactly the same by Whites and Non-whites, those ratings fall substantially below the mean ratings of communities overall. So, there is agreement that Gary is not a good place for minorities.

5.4 Generalizing with meta knowledge

Ideally, we could use the information gathered about community meta knowledge on subgroup experiences to determine more about the community itself. It's possible that meta knowledge would correlate with other measures we are interested in (for example, overall community satisfaction or Gallup's favorite Community Attachment variable).

For the purpose of this analysis, we chose to define meta-knowledge about the community's quality for minorities as the difference between the total percentage of positive responses to the question, "How is your community as a place for racial and ethnic minorities?" by Non-whites and Whites, as seen in equation (1).

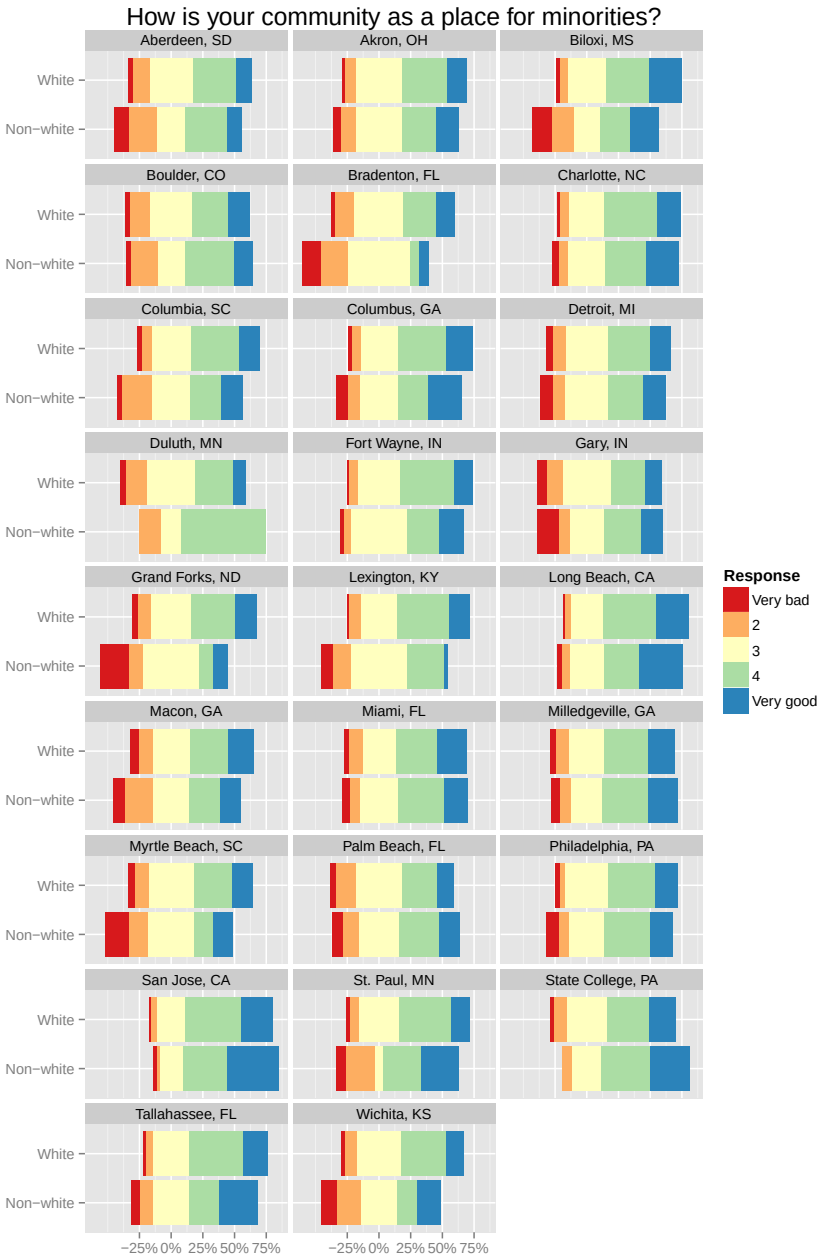

Fig. 13 Responses to the question, "How is your community as a place for racial and ethnic minorities?" faceted by community (2009 data).

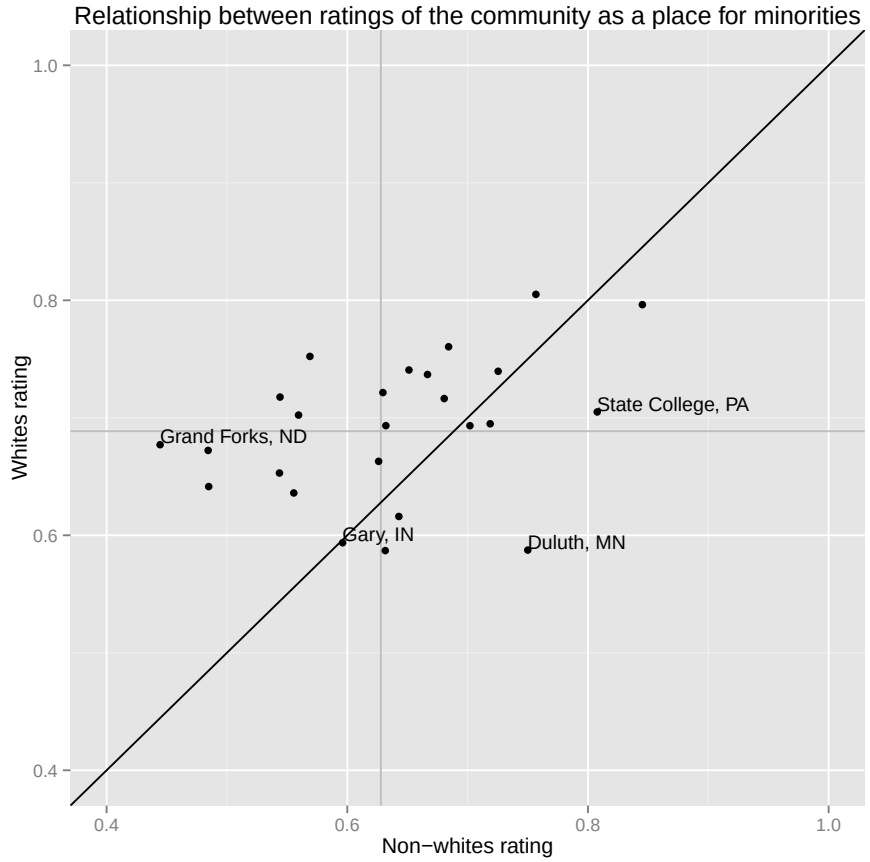

Fig. 14 Relationship between positive responses to the question, "How is your community as a place for minorities?" comparing ratings of Whites and Non-whites. Each community is represented, and the plot uses 2009 data. Some communitites, like State College, PA, are under-rated by Whites, some are over-rated, like Grand Forks, ND, and some are rated the same by both groups, like Gary, IN. The black line shows y=x, for comparison. Grey lines at x=0.63 and y=0.69 show the mean ratings by each group.

$$MK = \sum R^{+}_{\text{Non-whites}} - \sum R^{+}_{\text{Whites}} \tag{1}$$

Using this definition, communities where Whites highly over-rate their community as a place for minorities have a negative MK score, while communities where Whites under-rate their community have a positive MK score. Communities where the ratings by both groups are roughly equal (like Gary, IN) will have a MK score of zero. For a visual explanation of this rating system, see Figure 14. Points above the $y = x$ line will have a negative MK score, those below the line will have a positive score, and those on the line will be zero.

This definition puts emphasis on communities where out-of-subgroup participants rated their city the same way that in-subgroup participants did, based on the assumption that empathy is important to communities (Stephan and Finlay, 1999) However, looking at Figure 14, we can see that there is another important factor, which does not seem to correlate with meta-knowledge– ratings of the community by minorities themselves.

Ideally, a community would be good for minorities, and Whites and Non-whites alike would know it. But, communities with scores near zero are spread over the whole range of the ratings by Non-whites. In other words, there are many communities with meta-knowledge but they really just have awareness of how bad their community is for minorities. The fact that White and Non-white residents of Gary, IN are in agreement that it is not a good place for minorities does not suggest it is a great place to live.

In order to explore this, we plotted raw MK score against community satisfaction (from Section 3) and did not find a trend. Instead, it seemed like high MK scores, both positive and negative, were associated with higher community satisfaction. So, Figure 15 shows the relationship between the absolute value of meta knowledge score, $|MK|$, and community satisfaction.

Figure 15 shows a positive linear relationship between the variables, but it's not what we would expect. Communities that had the highest absolute difference in scoring of the question "how is your community as a place for minorities?" between Whites and Non-whites (and therefore the highest $|MK|$ score) had the highest community satisfaction scores. Communities that had low meta knowledge score (like Gary, IN) have low community satisfaction scores. This seems counter-intuitive, as we would expect communities that had more self awareness to be more satisfied, but that's not what we see in the plot. However, the baseline Non-white rating of the community as a place for minorities is not displayed in this plot.

A followup hypothesis is that meta-knowledge may not actually be good for a community. Perhaps the situation has to get quite bad before a community becomes meta-aware of how it actually treats minorities. In order to explore this hypothesis, we created the plot seen in Figure 16. Figure 16 shows the same data as Figure 15, with the addition of the points being sized by the percent of positive responses to the question, "how is your community as a place for minorities?" by Non-white respondents. From this plot, it seems plausible that the ratings by Non-white respondents are related to the relationship between community satisfaction and $|MK|$ score.

In order to further study the trend, we fit the model seen in Table 6. This model takes into account the ratings of the community as a place for minorities, rated by Non-white respondents, as well as the $|MK|$ score, in predicting community satisfaction.

All the terms in the model are significant at the 10% level, which admittedly isn't very high, but suggests there may be something to the model. The model tells us that without any information about $|MK|$ or Non-white ratings, we would predict a community satisfaction rating of 47%. However, for every one unit increase in $|MK|$, we would see a 9 point increase in community

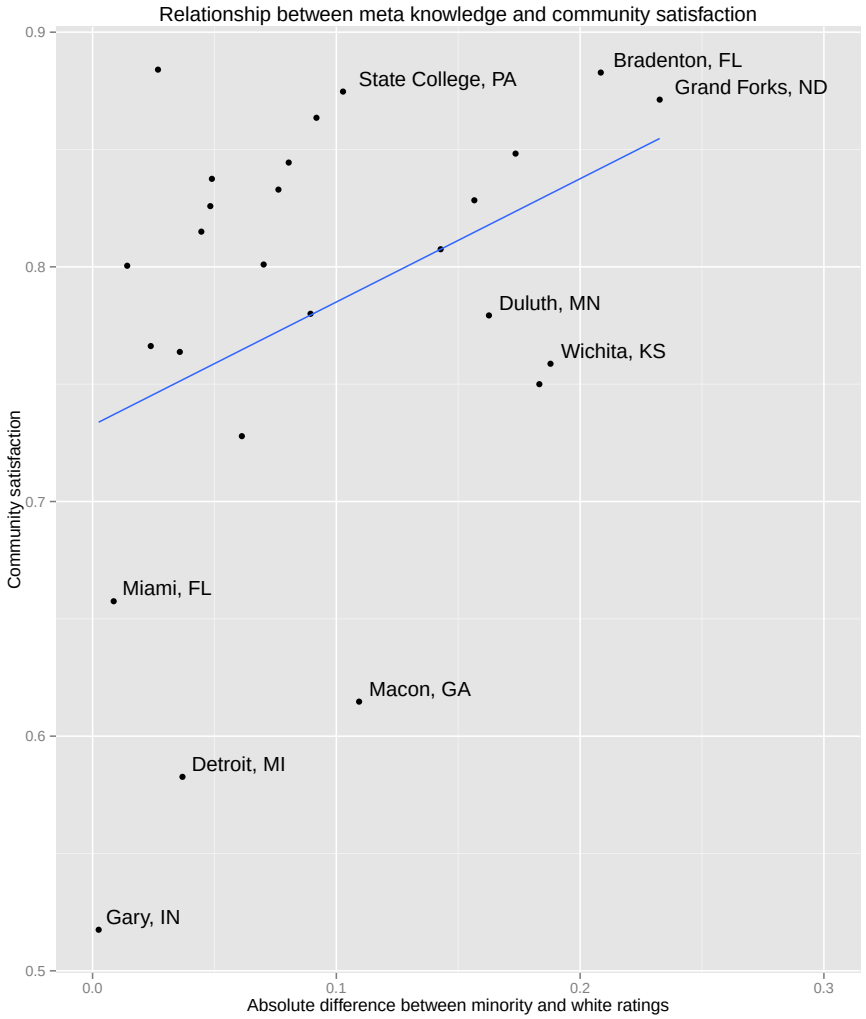

Fig. 15 Relationship between community satisfaction and $|MK|$. Communities in the lower left corner had good agreement between Whites and Non-whites about whether their community was a good place for minorities, but a low community satisfaction. Communities in the upper right had a lot of discrepency between Whites and Non-whites, but high community community satisfaction. (2009 data).

satisfaction rating. Similarly, for every one point increase in Non-white ratings, we would expect to see a 3.5 point increase in community satisfaction rating.

Interestingly, this relationship can also be seen with Gallup Poll's favorite composite variable, Community Attachment (discussed in Section 2). For those model estimates, see Table 7.

Both these models explain only a small portion of the variability in the data, but intuitively it makes sense that there would be other factors influenc-

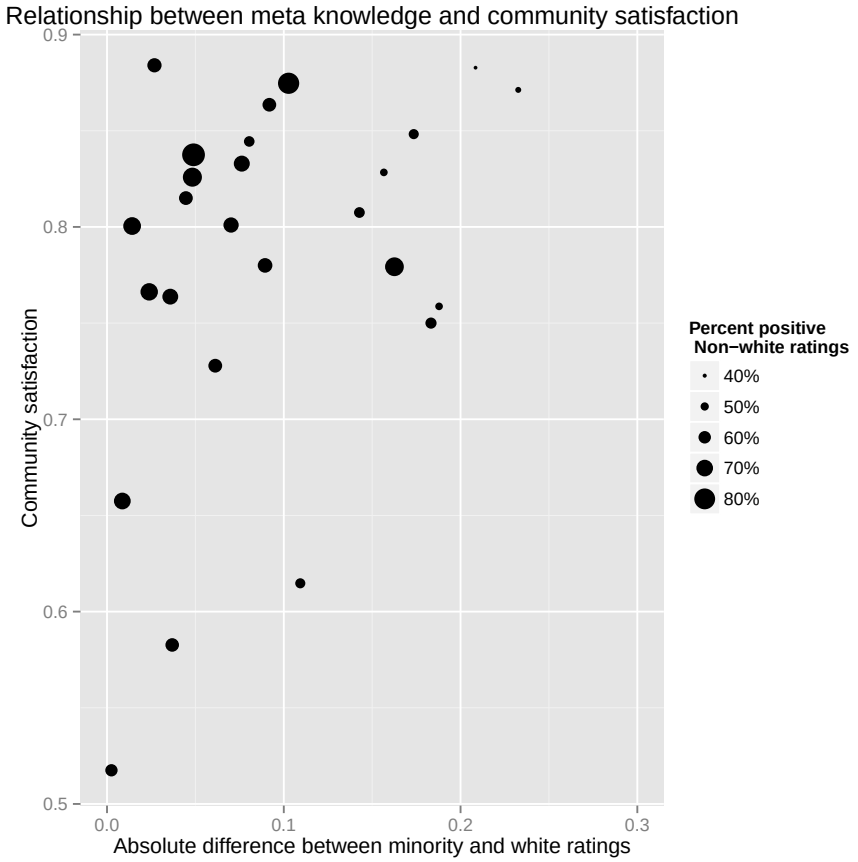

Fig. 16 Relationship between community satisfaction and $|MK|$, with points sized by Non-white ratings of their community as a place for minorities.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.4744	0.1520	3.12	0.0048
'|MK|'	0.9009	0.3345	2.69	0.0130
'Non-white raings'	0.3555	0.2054	1.73	0.0969

Table 6 Linear model predicting 2009 community satisfaction scores using $|MK|$ and ratings by Non-white respondants.

ing overall community satisfaction or attachment. However, it is interesting to think about the relationships shown here. Essentially, communities where Whites were over- or under-rating their community as a place for minorities have higher community satisfaction, as do communities with higher Non-white ratings of the community as a place for minorities. This is interesting, because it suggests that overall community satisfaction is related to the satisfaction of a minority group.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	2.6855	0.4740	5.67	0.0000
‘|MK|‘	3.5247	1.0432	3.38	0.0026
‘Non-white raings‘	1.1830	0.6407	1.85	0.0778

Table 7 Linear model predicting average 2009 Community Attachment scores using $|MK|$ and ratings by Non-white respondants.

Of course, even these minor conclusions are not very conclusive. We're working with a non-random sample of communities, and a small one at that. However, the relationships studied here can point to areas for further study.

6 Conclusions and Further Work

Through this exploration of the Knight Foundation's Soul of the Community data, we have been able to examine variability between communities and theorize about what might affect those variations. In Section 3 we saw which communities reported the highest levels of community satisfaction, namely State College, PA and Boulder, CO. In Section 4, we explored the common behaviors that survey participants engaged in, including registering to vote and voting in local elections. We noticed that college towns tend to have higher than typical rates of providing shelter to a non-relative (the couchsurfing effect?).

Then, in Section 5 we began to explore community meta knowledge, or how aware survey participants are about their community as a place for minority groups. We noticed that non-seniors almost always under-rate how good their community is for seniors, while communities tend to have high meta-knowledge about how good their community is for families with young children (though there are exceptions). Finally, we explored meta knowledge about how good a community is for racial and ethnic minorities. This was the most challenging, and had the highest potential for payoff, but the results were different than we expected. We were able to expose variation between cities, years, and groups, but it is still not clear how useful this meta knowledge could be as a measure of the soul of the community.

Looking at it community satisfaction, we found that the linear trend was in the opposite direction than we had expected. However, when used in conjunction with overall Non-white ratings of a community, meta knowledge could be used to somewhat successfully predict both community satisfaction and Community Attachment. For future work, it would be fascinating to track down why this trend is so different than one's assumptions. Are people happier in communities with low meta knowledge, or where people in- and outside subgroups have very different experiences?

It would also be interesting to explore whether meta knowledge on other subgroups, like seniors and families with children, can be used to model community satisfaction or Community Attachment. Unfortunately, this was outside the scope of this exploration.

References

Gallup Consulting. Soul of the community: Presentation to the program directors, June 2008.

Gallup Poll. Soul of the community knight communities overall, September 2009.

Gallup Poll. Soul of the community knight communities overall, November 2010.

Heike Hofmann and Hadley Wickham. Soul of the community. *Computational Statistics*, 20XX.

Jacob Jacoby and Michael S Mattel. Three-point Likert scales are enough. *Journal of Marketing Research*, 8(4):495–500, 1971.

Knight Foundation. Knight soul of the community 2010: Why people love where they live and why it matters: a national perspective, 2010.

Rensis Likert. A technique for the measurement of attitudes. *Archives of Psychology*, 140:1–55, 1932.

Naomi B. Robbins and Richard M. Heiberger. Plotting Likert and other rating scales. In *Section on Survey Research Methods - JSM 2011*, 2011.

Walter G Stephan and Krystina Finlay. The role of empathy in improving intergroup relations. *Journal of Social Issues*, 55(4):729–743, 1999.