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

Amelia McNamara

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 partipants 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 some factors which could be thought of as proxies 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

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

primary 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 growth 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 itself, 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 collected 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's a little difficult to pin down what this variable is, 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 somewhat of a model of what Gallup believes is important to strong communities. The team has discovered that this composite variable, Community Attachment, is positively correlated with local 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 this 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), 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. This difference 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 these maps came from the Intercensal population estimates compiled by the Census Bureau. Population estimates are calculated each year in between Census years. The estimates don't vary too much from the Census count, but yearly estimates were used for the sake of having slightly different populations 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 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%. And 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 decisions like these were made.
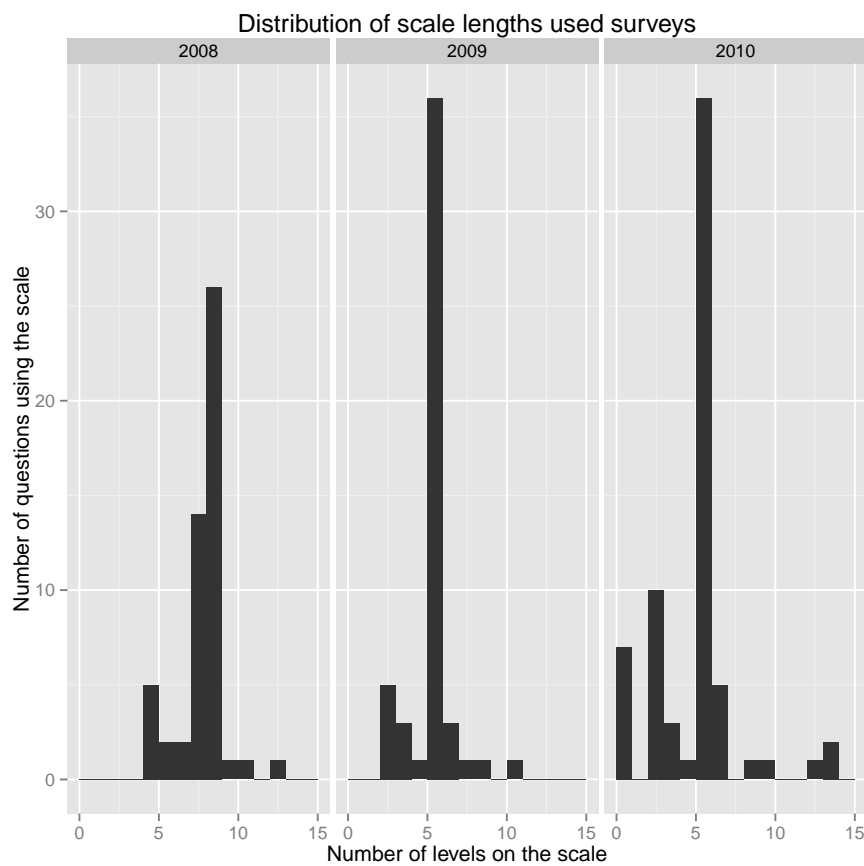
2.2 Scale Lengths

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. However, there was little consistency in the number of levels for the scale. 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 other scales (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
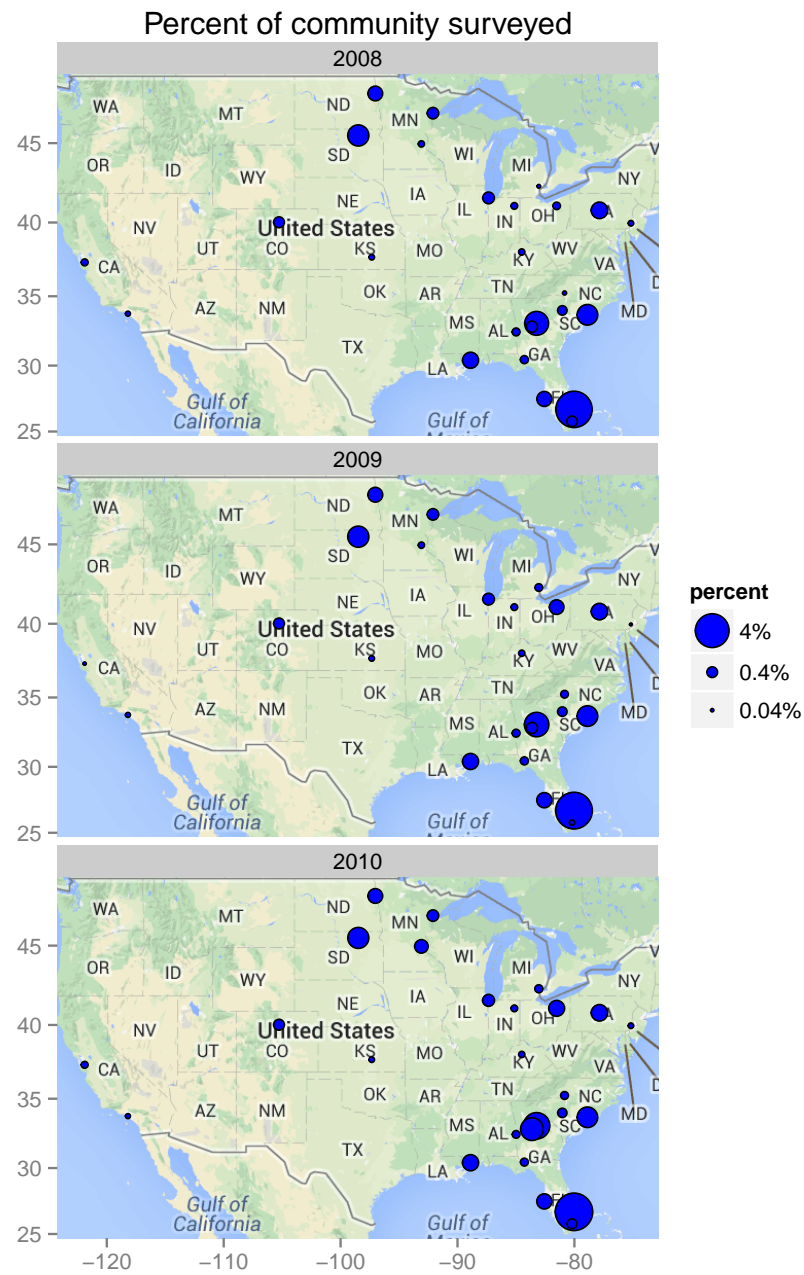
Fig. 1 Maps of yearly survey percentages

Distribution of scale lengths used surveys

Fig. 2 Scale length distributions for each year of the survey.

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 data provided 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 that 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 committment. 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 rows than previous years. The explanation for this difference is missing data, presumately related to surveys that were not fully completed.

While the 2008 and 2009 data sets are almost complete, the 2010 data set has about 25% missing data.

```
##   year missing percent
## 1 2008      NA      NA
## 2 2009      63  0.4589
## 3 2010    4947 24.4043
```

This missing data 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 dataset 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.
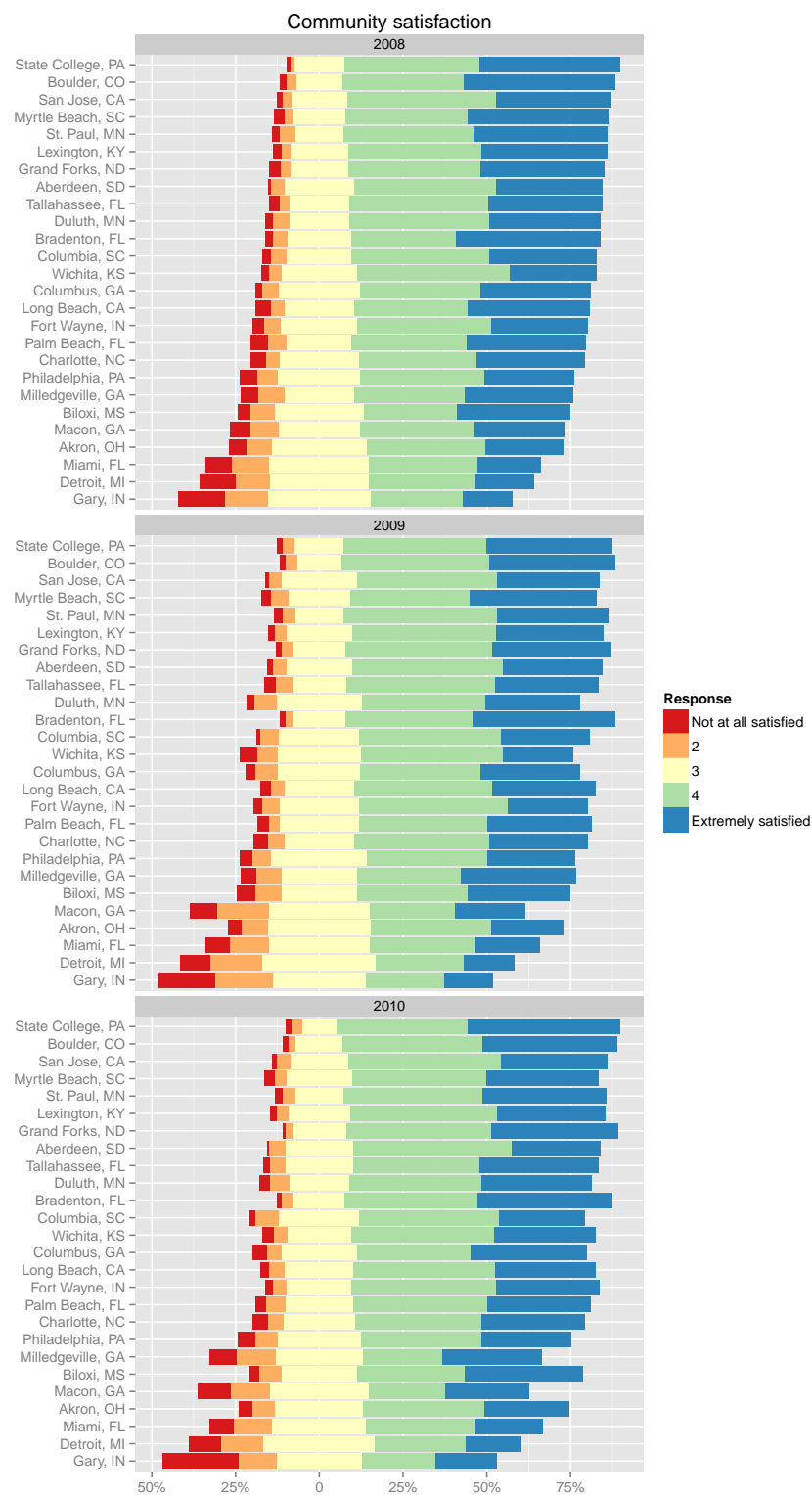
3 Community Satisfaction

Knowing that the question about community satisfaction was 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 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

behaviorsec) 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

Fig. 3 Community satisfaction with cities ordered by positive responses in 2008
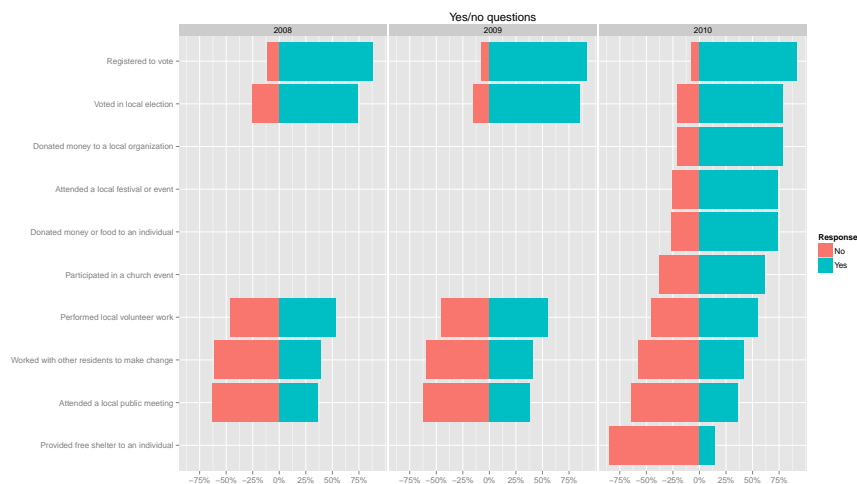
Fig. 4 Percent participation in local activities over all three years

previous years. Behaviors were 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 behaviors (considering all three years) were performing working to make change, and attending public meetings. When the additional questions were added in 2010, an even-less-common behavior was added, donating free shelter to a non-relative.

Then, the question arose whether all communities performed these actions at similar rates, or if there were local variations in behvaior. 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 communitites 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 satisifed 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.
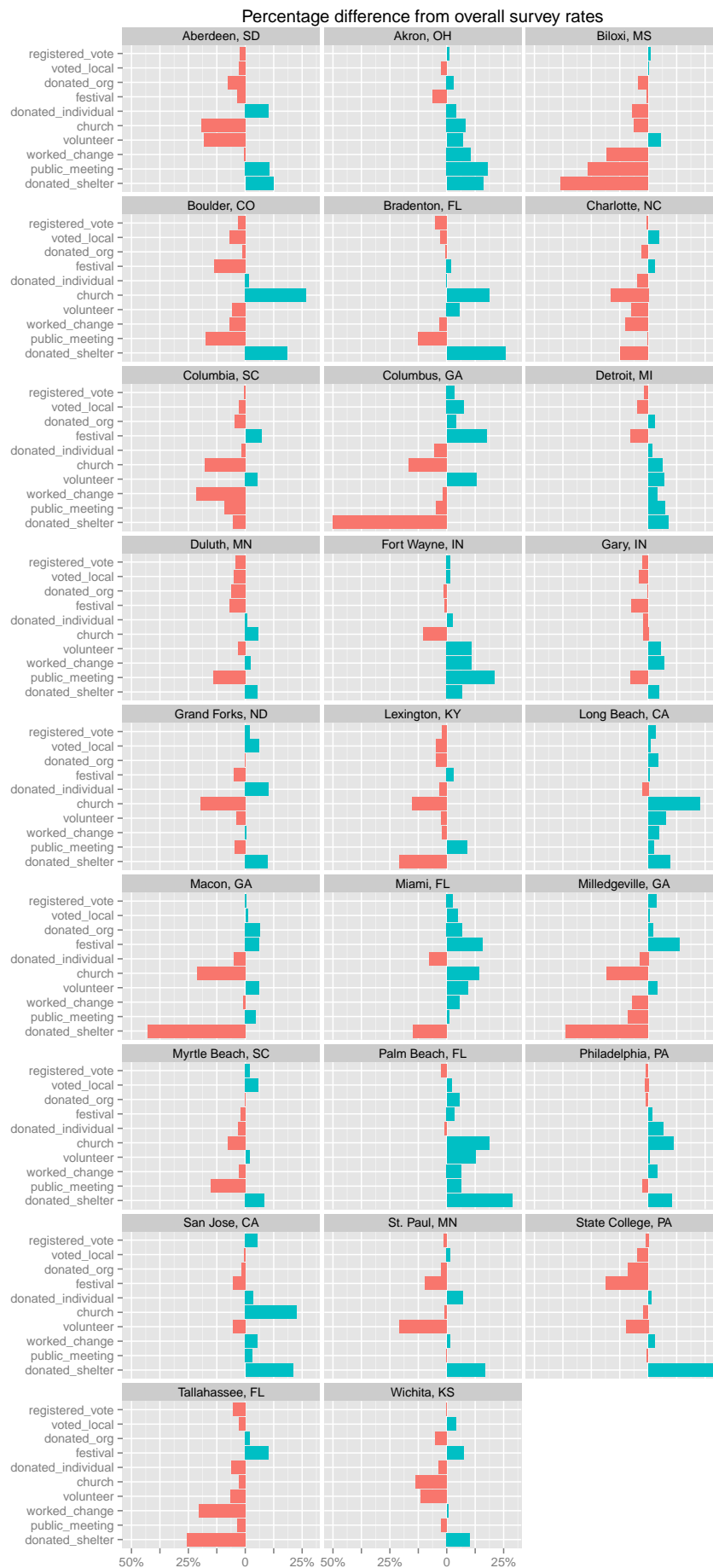
Fig. 5 Difference from overall survey rates (2010 data)

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 a number of subgroups or minorities.

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 racial and ethnic minorities, families with young children, and senior citizens.

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. The yearly ratings distributions are shown in Figure 6. It appears that non-seniors 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 in the two groups was created. This plot can be seen in Figure 7, although it is specifically on the 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.

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

For this facet of meta knowledge, we consider the difference between community ratings as a place for families with young children, split between participants who reported having dependent children under the age of 18 living in their household and those who did not. While the data included more granular demographic details about the ages of the children in the households,
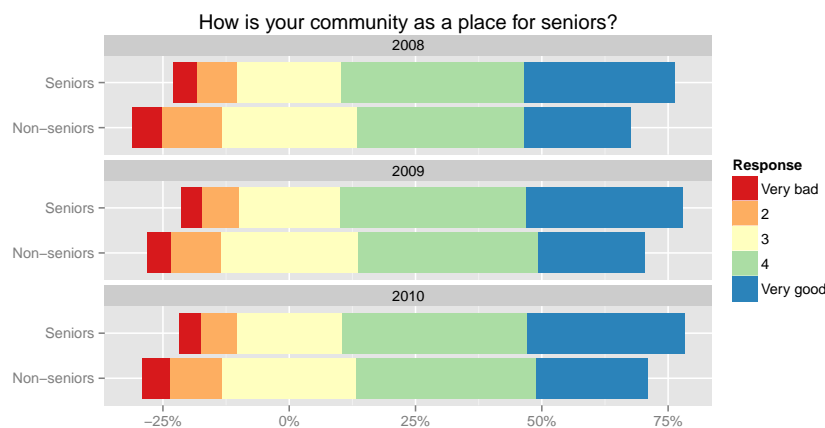
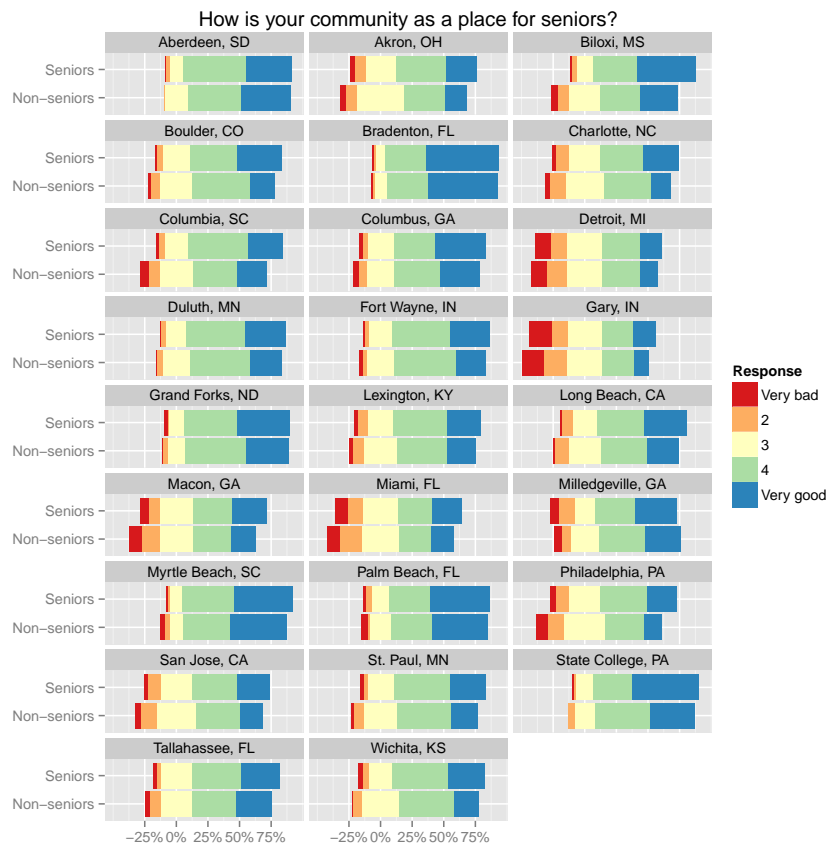**Fig. 6** How is your community as a place for seniors?

splitting the data into partipants with children and those without made the groups closer in size. And, 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.

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 before. 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 groups, 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).
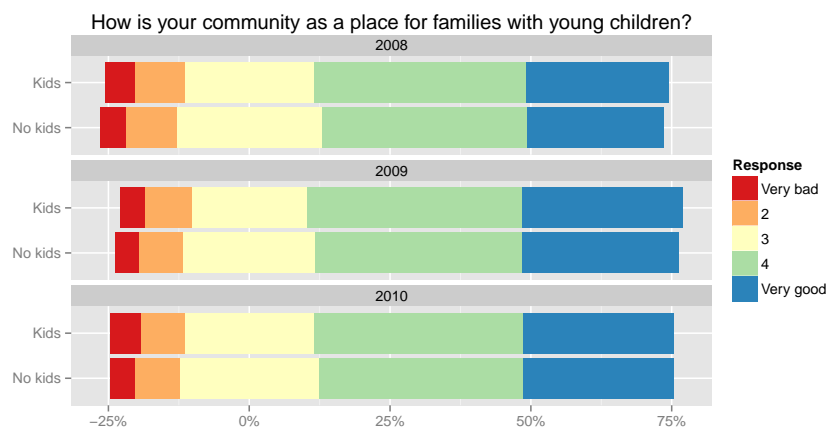
Fig. 7 How is your community as a place for seniors, faceted by city (2010 data).

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.

The sample sizes were somewhat small each year,

```
## Source: local data frame [22 x 4]
## Groups: year
##
##                                          race    n       frac year
## 1                                        Asian  214  0.0784170 2010
## 2                   Black or African-American  326  0.1194577 2010
## 3                                     Hispanic  113  0.0414071 2010
## 4  Native Hawaiian or other Pacific Islander   45  0.0164896 2010
## 5                              Some other race   36  0.0131916 2010
## 6                                        White 1995  0.7310370 2010
```
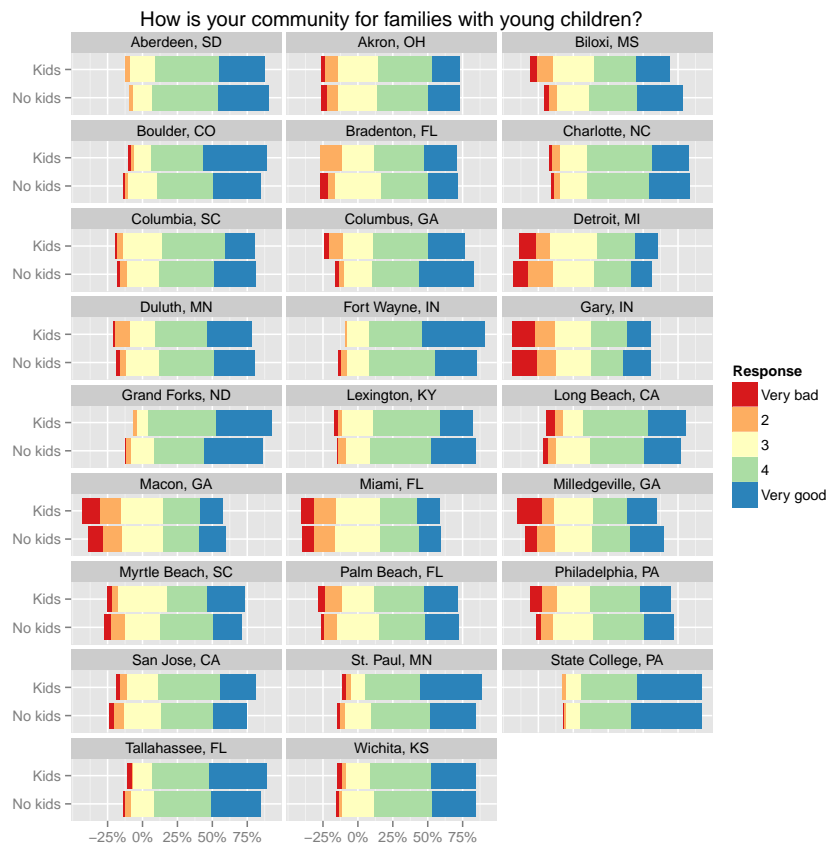
Fig. 8 How is your community as a place for families with young children?

```
## 7                     American Indian or Alaskan   150 0.0112444 2009
## 8                                          Asian   173 0.0129685 2009
## 9             Black or African-American  1363 0.1021739 2009
## 10                                    Hispanic   208 0.0155922 2009
## 11                              More than one    53 0.0039730 2009
## 12 Native Hawaiian or other Pacific Islander    26 0.0019490 2009
## 13                             Some other race     7 0.0005247 2009
## 14                                       White 11360 0.8515742 2009
## 15                                    Hispanic   563 0.0417099 2008
## 16                              More than one    78 0.0057786 2008
## 17             American Indian or Alaskan   142 0.0105201 2008
## 18                                       Asian   314 0.0232627 2008
## 19             Black or African-American  1438 0.1065343 2008
## 20 Native Hawaiian or other Pacific Islander    36 0.0026671 2008
## 21                             Some other race    12 0.0008890 2008
## 22                                       White 10915 0.8086383 2008
```

So for the plots about subgroup meta knowledge, we combined all the non-white responses into one group. These sample sizes are shown below.

As before, 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?" This is shown in Figure 12. Interestingly, Whites were rating their communities as better places for minorities than minorities themselves in 2008 and 2009, but in 2010 they began under-rating.

Again, we want to see the individual variation between communitites. Although we've been showing responses over the 26 communities for 2010 data for the past few section, in this case the responses to the demographic question

How is your community for families with young children?

Fig. 9 How is your community as a place for families with young children? by community (2010 data)

about race had such low response rates that we chose to look at the 2009 data. The community-level responses are shown in Figure 13.

```
## Source: local data frame [3 x 3]
## Groups:
##
##    Year Number Percent
## 2 2008  10915    79.17
## 4 2009  11360    85.16
## 6 2010   1761    64.36
```

There is a lot of variation between the communities for this particular facet of meta knowledge. Some communities saw the White respondents over-scoring their community as a place for minorities, some were under-scoring. For the most part, however, Whites over-rated their communities as a place for racial
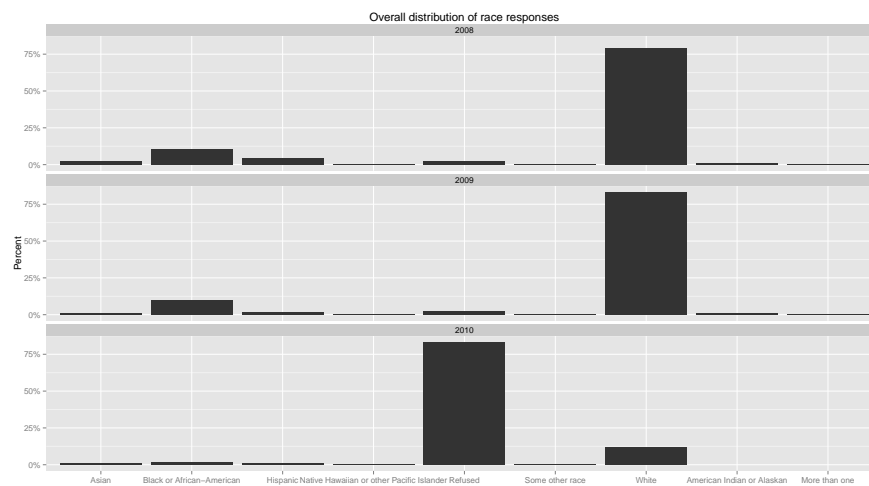
Fig. 10 Which of these groups best describes your racial background?
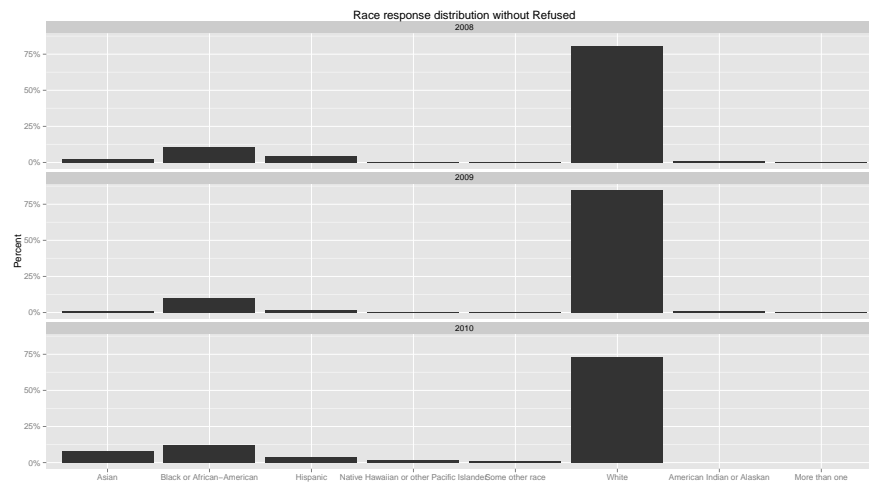
Fig. 11 Distribution of races, with Refused responses removed.

and ethnic minorities, compared to minorities themselves. 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.
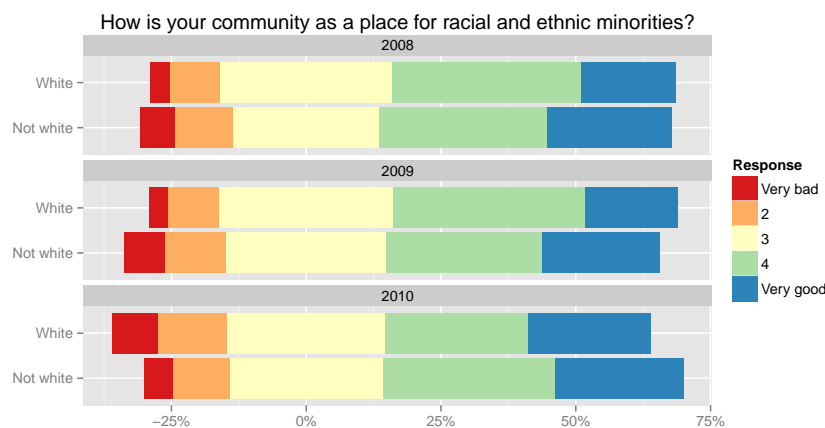
Fig. 12 How is your community as a place for racial and ethnic minorities?

### 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, Gallup's favorite community attachment variable). However, it is hard to decide how to rate communities on their meta knowledge.

Clearly, communities whose out-of-subgroup participants rated their city the same way that in-subgroup participants did should be rated more highly. But, having a city that is good for minorities is also a positive factor, and should be taken into account. A city that is bad for children where everyone (parents and non-parents alike) knows it is not necessarily a great community.

One first step toward this sort of analysis is presented in Figure 14, which shows the relationship between community attachment (Gallup's composite measure from all their analysis material) and meta knowledge. In this case, we are defining meta knowledge as the difference between the total positive ratings given by whites versus minorities to the question, "how is your community as a place for racial and ethnical minorities?" If you look at Figure 13, you can see that whites highly over-rate their community as a place for minorities in Grand Forks (and therefore have a negative difference score), while whites in State College under-rate their community (and therefore have a positive difference score). Gary, IN manages a score of almost exactly zero– that is, whites and minorities are rating it in the same way. Plotted against the mean community attchment for communities, there is very little perceptable trend, and there is a lot of scatter around high mean community attachment values.

So, instead of using the calcuated difference between scores by whites and minorities, we can plot the absolute difference. This is shown in Figure 15.
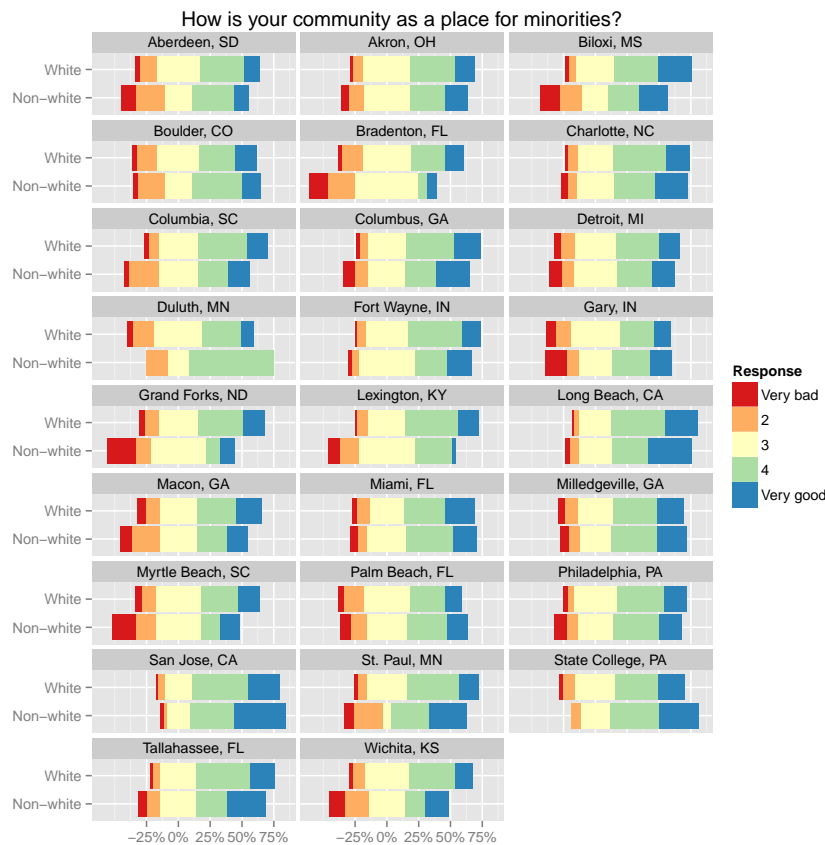
Fig. 13 How is your community as a place for racial and ethnic minorities? (2009 data)

Now we can see a positive linear relationship between the variables, but again, 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 minorities had the highest community attachment scores. Communities that had good meta knowledge (like Gary, IN) have low community attachment scores.

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 communitities 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 **??**, we explored the common
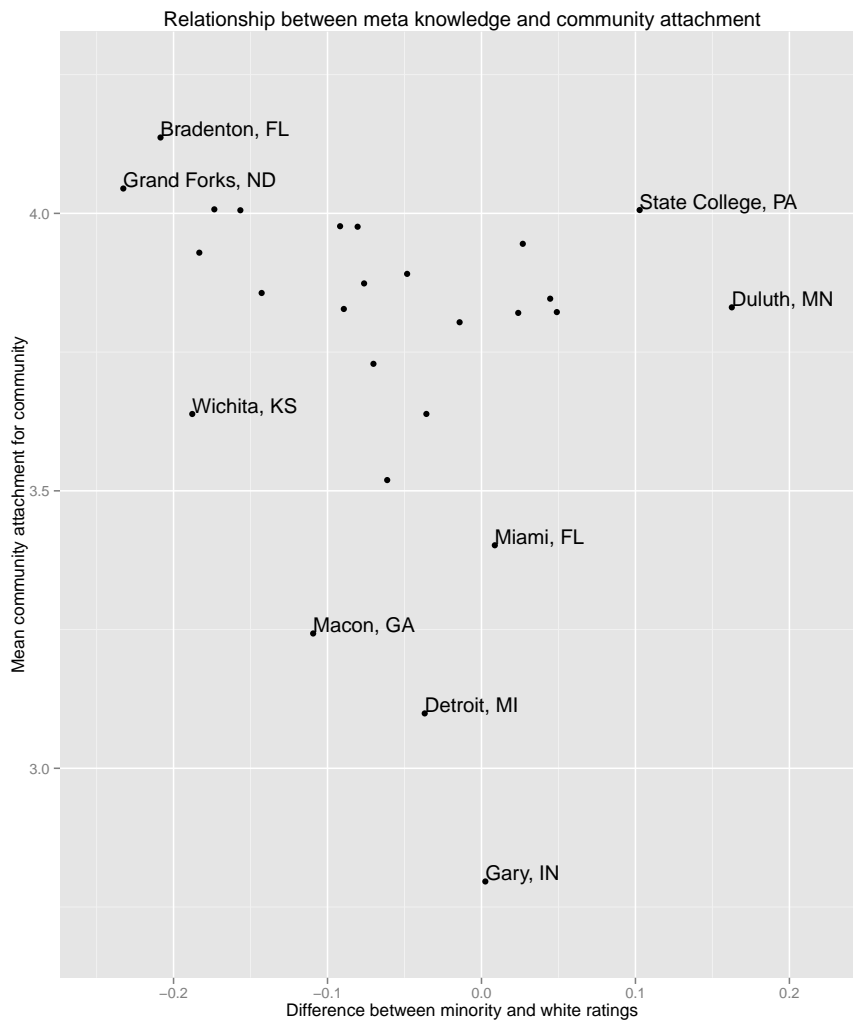
Fig. 14 Relationship between mean community attachment and meta knowledge (2009 data)

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
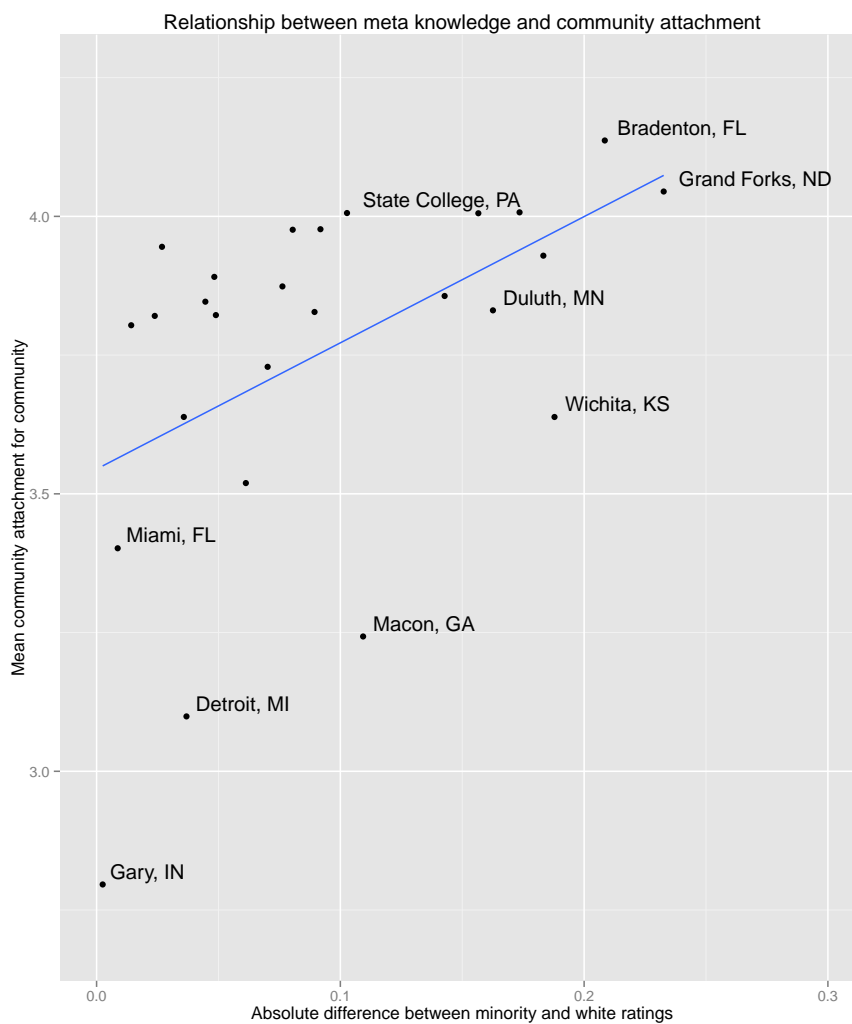
Fig. 15 Relationship between mean community attachment and absolute value of difference (meta knowledge) (2009 data)

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 in relation to Gallup Poll's favorite measure, community attachment, we found that the linear trend was in the opposite direction than we had expected.

For future work, it would be fascinating to track down why this trend is so different than one's assumptions. Is it because community attachment includes some facet of meta knowledge already? Are people happier in communities with low meta knowledge, or where people in- and outside subgroups have very different experiences? More exploration is necessary.

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.

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.