Kelsey Evans
LING 211
Professor Zanuttini
16 April 2020

Shopping at Meijer's? An Examination of Possessive Business Names in Speech

Contents

Introduction

Many people take extreme pride in their local linguistic features. Sometimes, this is justified, such as the Southeast Michigan phenomenon of calling the night before Halloween "Devil's Night." Sometimes, it is not justified – I have seen social media posts boasting about the "local" feature of saying "yeah no" to mean "no" and "no yeah" to mean "yes," only for the poster to find out that in fact, the majority of young people across the country have phrases like this in their dialect. My question is about another lexical item that is supposedly also unique to Michigan (or perhaps the Midwest more generally). This feature is the use of the <'s> possessive form for business names even in a context where it was not possessive initially. Michiganders like to joke about this being a quirky feature of Michigan speech, which is a fun thing to consider (I have this feature myself). But once in a while I will hear someone from a completely different part of the country who also has it, and I wonder if it is in fact local after all.

Hypothesis

I have two hypotheses. The first is that this feature is indeed more concentrated in Michigan, or at least the Midwest, and the second is that this construction is only possible with a proper noun that could conceivably sound like a personal name; in other words, something like "Costco's" is ungrammatical, because it is short for the Cost Company, and therefore couldn't be someone's name, while something like "Meijer's" is grammatical to people who have this feature.

In symbols:

$\mu_{MI} > \mu_{Other}$ / $\mu_{Midwest} > \mu_{Other}$

$\mu_{Name} > \mu_{Not\ a\ Name}$

where μ is the average score given to a question on a 1-5 scale.

History

The most coherent history and prediction of this phenomenon that I could locate is an interview with Ann Curzan, a linguist at the University of Michigan, with Michigan Public Radio in 2015. Curzan suggests that there could be several reasons for this phenomenon. One likely one is that the iconic Michigan company, Ford Motor Company, used to be called "Ford's Factory" when Henry Ford started it initially. For other companies, this is true as well – there are records of calling the JC Penney store "JC Penney's" since as early as 1940. She also admits that it could be for a reason as simple as all the stores that DO have a possessive, like Busch's or McDonald's. But this final reason doesn't explain why this is a Michigan (or, as she notes, perhaps a wider Midwest/Rust Belt) phenomenon. It could be because of the small-town culture that is perhaps more pervasive in these regions, where stores would be named after a person – Curzan says it lends a "friendly, homey feeling." Many restaurants, for instance, are titled after someone's name to create a familiar feeling (Though this is outside the scope of my research as these restaurants already have the possessive in the name – it is not added by speakers).

Curzan also speaks a bit on my second hypothesis, agreeing that this phenomenon is likely limited to stores that could theoretically be someone's name. This is why we can get "Ford's" from people who have this feature, but not "GM's," even though they are both car companies

concentrated in the Midwest. With the history backing up my predictions, I decided to continue researching this phenomenon on my own.

Methodology

There is surprisingly little existing information about this phenomenon. Beyond the Michigan Public Radio interview, almost all of what I found was speculation on social media by people who had lived in Michigan for a long time trying to figure out the phenomenon, as well as a few similar articles from other regions, such as from Pittsburg and other cities. Many online quizzes about Michigan dialect also ask about the phenomenon, but unfortunately, none of them provided an aggregate of the results at the end that I could look into. Therefore, I decided to design my own survey.

Around 550 people were surveyed, cut down to 501 people after removing incorrectly-answered surveys and surveys where the participant had not lived in any particular location more than a few years. The participants were solicited over several forms of social media. I reached out through Facebook and Instagram. One issue that I ran into while collecting data this way is that a very Midwest-heavy Facebook community happened to answer the survey far more than any other participants, leaving me with the majority of the responses being from Midwest participants. This would be something for me to look out for in future studies.

The survey contains three demographic questions and 30 sentence questions. 15 out of the 30 questions will be analyzed; the other 15 were there so as not to make the point of the survey extremely obvious. I will explain the questions further:

Demographic questions

1. How old are you?
2. Where have you lived the longest in your life? (2-letter state code or name of country if you are not from the United States)
3. How many years have you lived there?

Though there is no evidence of this being an age-based phenomenon, I figured it wouldn't hurt to check. Question 2 is asked to divide the participants by region, and Question 3 is asked so

that I would instantly remove any results where they had answered a number lower than 8 (to only capture people who had lived somewhere a long time).

The answers to Question 1 ranged from age 15 to age 74. The answers to Question 2 included 30 of the 50 states, as well as three other countries. The answers to Question 3 ranged from participants having lived in a place from 8 years to 74 years (the 74-year old participant lived in Michigan their entire life).

Survey Questions

The survey questions are random in the survey, but I will display them in the sections that I selected for easier explanation. Each question, the participants rate on a scale from 1-5. The following is the instructions they are given:

This question and the following questions will each contain a sentence. Rate the sentence between 1 and 5 where 1 is "not acceptable" and 5 is "totally acceptable."

Note that "acceptable" has NOTHING to do with what is considered "standard" or "correct" English but has only to do with whether the sentence is acceptable to you personally.

In other words, read the sentence as it's written. If it "sounds right," it's acceptable. If any part of the sentence does not sound right, then it's not acceptable.

Next I will explain the questions. I will briefly share the not-analyzed questions, then move onto the ones I used in my analysis.

Extra Questions (not analyzed)

Of the questions not to be analyzed, there are three groups.

GROUP 1:

(1) Trader Joe's has the best cheese crackers.
(2) My mom used to wear Levi's jeans.
(3) My brother works in the produce section at Wegmans.
(4) My prom dress was from Macy's at the mall.
(5) The Shamrock Shake at McDonald's is very good.

GROUP 2:

(1) My friend likes the free samples they give at ULTA.
(2) My alarm clock is from Bed Bath and Beyond.
(3) Our washing machine is from General Electric.
(4) Fresh Thyme has more expensive produce than other grocery stores.
(5) My computer from Apple has more storage than my old computer.

GROUP 3:

(1) There's so much traffic on the way to the mall anymore.
(2) My brother was done dinner, so our mom let him get a slushie from 7/11.
(3) Mom decided the car needed washed, so we went to the Splash Car Wash.
(4) I'm going to Kohl's later; do you want to come with?
(5) We might could go to the Mobil station to get gas.

GROUP 1 and GROUP 2 questions are all controls – business names that already have possessives, for GROUP 1, and business names with no possessive at all for GROUP 2. These were all rated a 4.25/5 or higher on the acceptability scale, and most were higher than 4.5, so I would consider them good controls. GROUP 3 questions ask about grammatical properties that are not present in Standard American English. They are included to steer participants away from the true purpose of the survey. These questions were rated 2.54, 1.61, 3.16, 4.67, and 2.47 respectively.

The only one that is rated highly is (4); to me, it makes sense that this one is the highest rated, as "come with" is widely used in the Midwest, and that is where most of the survey participants were from.

Questions for analysis

Of the questions for analysis, there are two groups.

GROUP 1:

(1) My dad works at Ford's.
(2) At Nordstrom's they have a nice shoe selection.
(3) I love getting gift cards to Barnes and Noble's.
(4) My dad works for Ford's.

(5) I like shopping at Aldi's because the shelves are very organized.
(6) Have you been to the Starbucks inside Kroger's?
(7) The parking lot at Meijer's was so full yesterday.
(8) These high-top shoes are from Zara's.
(9) JC Penney's is there when you first enter the mall from the east side.
(10) I buy all my sewing supplies at Joann's.

GROUP 2:

(1) My younger sister spends so much time in Target's.
(2) I need to get my phone fixed at T-Mobile's.
(3) The grocery store where I go to college is Stop and Shop's.
(4) You can only go into Costco's if you have a membership.
(5) They were out of batteries at Walmart's.

GROUP 1 questions are businesses that I believe can be made possessive, based on my hypothesis that business names can be made possessive if they sound like they could be a person's name. GROUP 2 questions are businesses I believe cannot be made possessive. It may look like I mistakenly put Ford's for both (1) and (4), but it was actually a small curiosity I had about whether there was a difference in working "for" someone or working "at" a place. In the following results section, the variables in R code are labelled as follows:

NA.3.2 means "not analyzed, group 3, question 2"
A.1.9 means "analyzed, group 1, question 9" etc.

All the questions I'll be working with from here on out will be of the form A.1.x or A.2.x, and I will include the questions as much as possible for clarity.

Results & Commentary

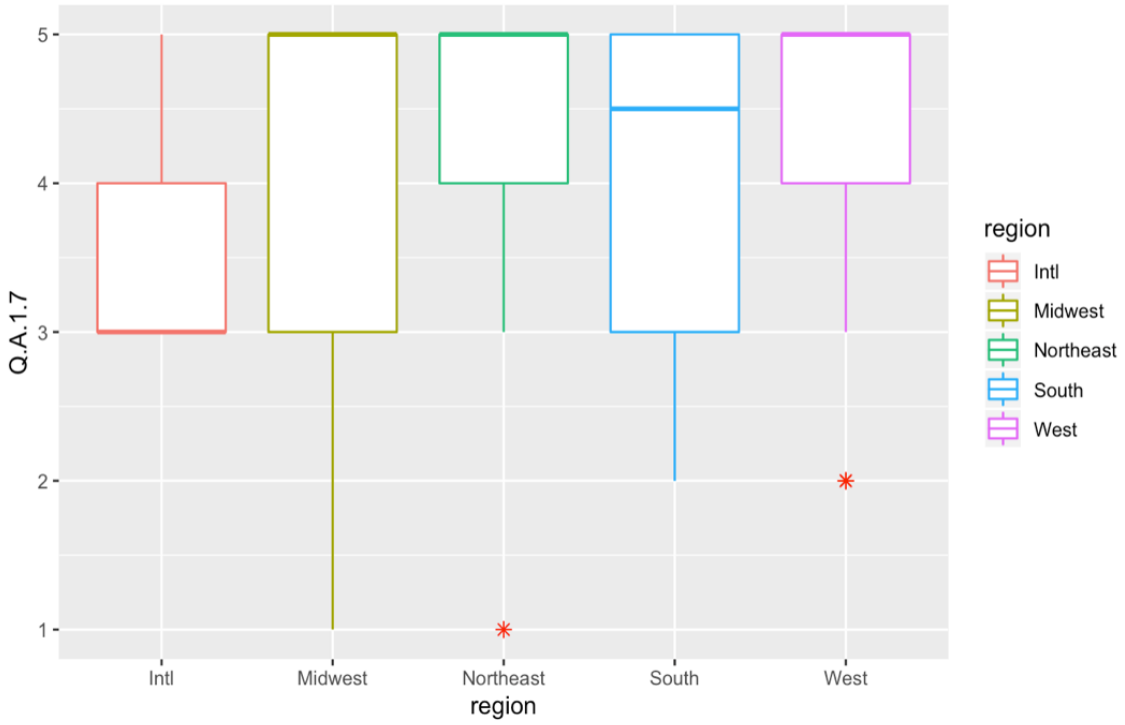
I have tried to include everything possible directly in this paper, but I will also provide in parenthesis the section in the R code document to look to if necessary.

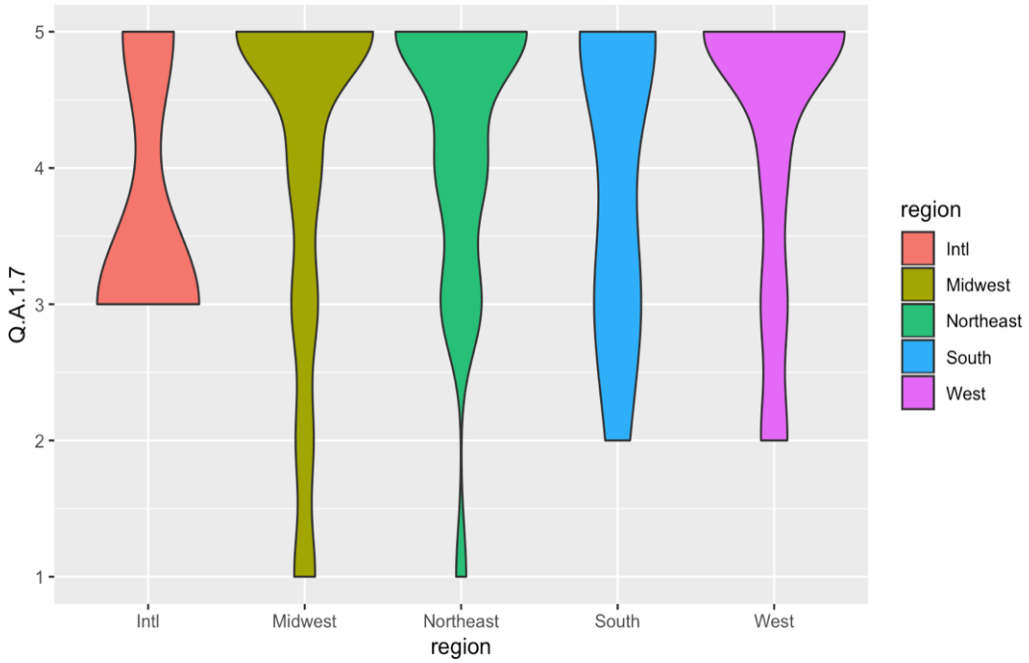
Hypothesis 1

First, I will look at my first hypothesis – the idea that this possessive construction is more acceptable in the Midwest than in other regions. I will look at the following: boxplots, maps, and ANOVA test by region.

Plots

In the R code document (SECTION 2: PLOTS) are boxplots and violin plots for all the A.1.x questions (ones I expect can be possessive) by region. I cannot average all the questions together because most of the questions do not have a strong enough correlation with each other for this to be statistically significant. Below is an example of each plot. The example question is A.1.7 ("Meijer's").





Unfortunately, none of these boxplots, at least by sight, seem convincing that Midwest participants have these constructions at a much higher rate than others.

ANOVA

To test my hypothesis more formally, I ran ANOVA tests (analysis of variance by region) in R for each question. Below are the F-scores and p-values for each question. (In R: SECTION 3: ANOVA)

Question	**F-score**	**p-value**
A.1.1 (Ford's)	3.63	0.006
A.1.2 (Nordstrom's)	2.10	0.079
A.1.3 (Barnes and Noble's)	0.168	0.954
A.1.4 (Ford's)	1.63	0.164
A.1.5 (Aldi's)	1.68	0.154
A.1.6 (Kroger's)	0.148	0.964
A.1.7 (Meijer's)	0.331	0.857
A.1.8 (Zara's)	1.21	0.306
A.1.9 (JC Penney's)	2.98	0.019
A.1.10 (Joann's)	0.770	0.545

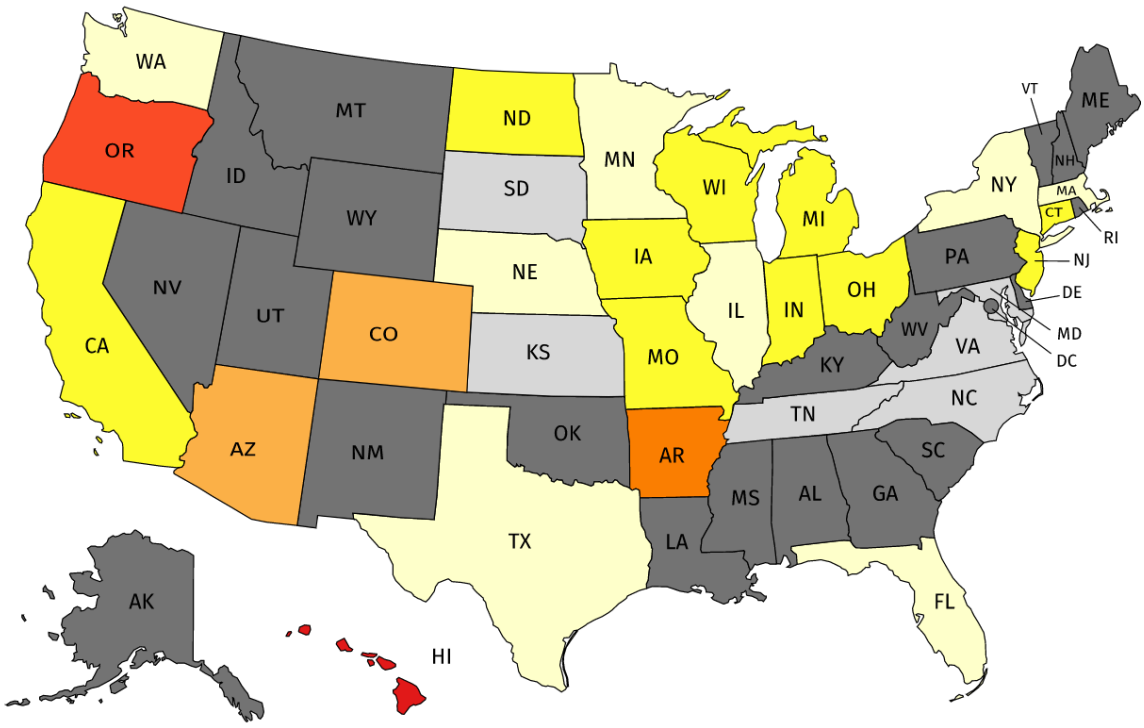
Only two of these (Question 1, Question 9) have a p-value of less than 0.05, which means that those have statistically significant differences in results by region. Unfortunately, even for those two questions, this p-value means that there is in fact a difference by region, but looking at individual p-values by region, for neither of those two questions does the Midwest specifically differ from the others. Unfortunately, at least based on my survey data, I must reject the null hypothesis that this is a Midwest-specific construction. I ran the same tests with Michigan even more specifically, and the results remained the same.

Map

I decided to make a map of the United States for this construction. The colors represent the average ability of each state to use this construction (In R: SECTION 4: MAP DATA). Since this is not a strict statistical analysis, I did decide to average the questions together for this map. The colors are as follows:

Dark gray	no data
Light gray	3.0 – 3.3
Light yellow	3.31 – 3.6
Dark yellow	3.61 – 3.9
Light orange	3.91 – 4.2
Dark orange	4.21 – 4.5
Light red	4.51 – 4.8
Dark red	4.81 +

(No state rated these questions an average of less than 3; in other words, no one found them unacceptable.)



As we can see, there isn't a particularly compelling evidence from the map about this construction either. I will mention here that one unfortunate problem I faced with this data was that about three-quarters of the total data was from the Midwest, and only about a quarter of the data from the rest of the country. Having more data would certainly help paint a clearer picture.

Hypothesis 2

Next, I will investigate my second hypothesis – that there is a difference in which stores can be made possessive; specifically, that stores that sound like someone's name can be made possessive. First, here are the average score for each question in the two groups. In the bar graph, the questions are in the same order as in the preceding tables. (In R: SECTION 5: AVERAGE SCORES).

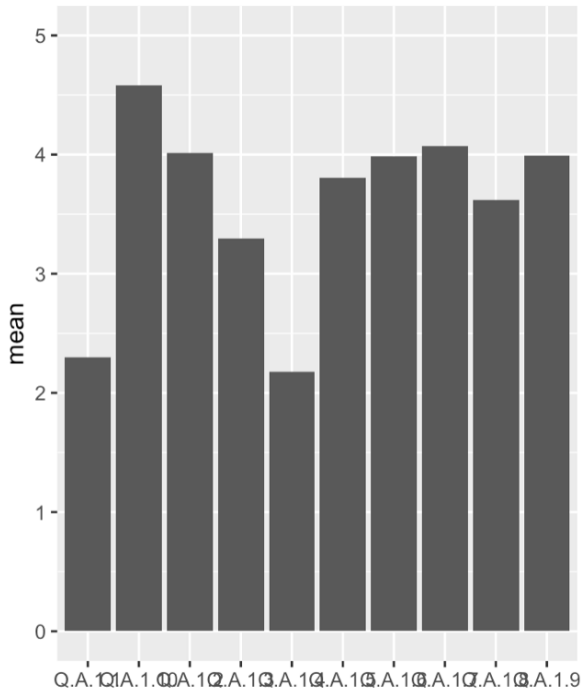
GROUP 1

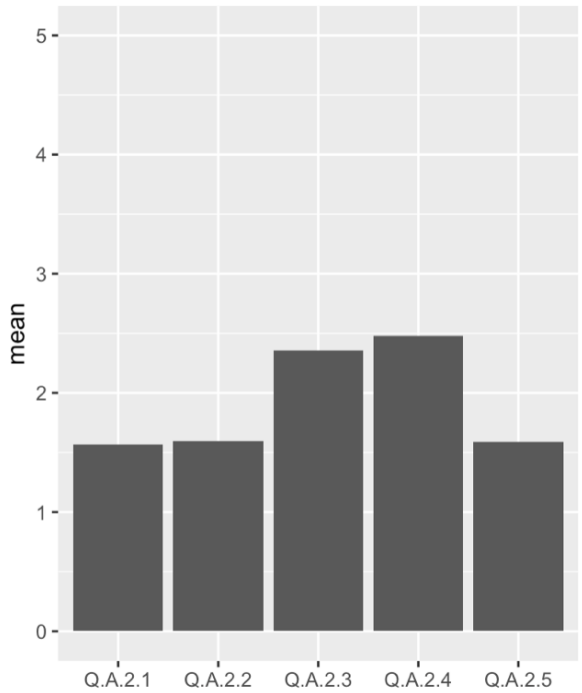
Question	**Average Score**
A.1.1 (Ford's)	2.30
A.1.2 (Nordstrom's)	4.01
A.1.3 (Barnes and Noble's)	3.30

A.1.4 (Ford's)	2.17
A.1.5 (Aldi's)	3.81
A.1.6 (Kroger's)	3.99
A.1.7 (Meijer's)	4.07
A.1.8 (Zara's)	3.62
A.1.9 (JC Penney's)	3.99
A.1.10 (Joann's)	4.58

GROUP 2

Question	**Average Score**
A.2.1 (Target's)	1.56
A.2.2 (T-mobile's)	1.59
A.2.3 (Stop & Shop's)	2.35
A.2.4 (Costco's)	2.47
A.2.5 (Walmart's)	1.59





As we can see, everything in the first category is rated above average (except the Ford questions, which I will discuss in the next section), and everything in the second section is rated below average. To see if these differences are statistically significant, I will do both pairwise and averaged t-tests for the difference in means.

Pairwise

Below is a matrix of the pairwise t-scores and corresponding p-values for the difference in means for each question. (Each corresponding row and column is a separate t-test.) (In R: SECTION 6: T-TESTS)

	A.1.1	A.1.2	A.1.3	A.1.4	A.1.5	A.1.6	A.1.7	A.1.8	A.1.9	A.1.10
A.2.1	t = 9.3 p = 0	t = 32 p = 0	t = 21 p = 0	t = 7.8 p = 1.5e14	t = 28 p = 0	t = 31 p = 0	t = 34 p = 0	t = 26 p = 0	t = 32 p = 0	t = 49 p = 0
A.2.2	t = 9.0 p = 0	t = 32 p = 0	t = 21 p = 0	t = 7.4 p = 2.4e13	t = 28 p = 0	t = 31 p = 0	t = 34 p = 0	t = 25 p = 0	t = 32 p = 0	t = 49 p = 0
A.2.3	t = -0.57 p = 0.57	t = 19 p = 0	t = 10 p = 0	t = -2.0 p = 0.04	t = 16 p = 0	t = 18 p = 0	t = 20 p = 0	t = 13 p = 0	t = 19 p = 0	t = 29 p = 0
A.2.4	t = -1.8 p = 0.06	t = 17 p = 0	t = 8.4 p = 0	t = -3.2 p = 0.001	t = 14 p = 0	t = 16 p = 0	t = 18 p = 0	t = 12 p = 0	t = 17 p = 0	t = 27 p = 0
A.2.5	t = 8.8 p = 0	t = 31 p = 0	t = 20 p = 0	t = 7.2 p = 6.39e13	t = 27 p = 0	t = 30 p = 0	t = 33 p = 0	t = 25 p = 0	t = 31 p = 0	t = 47 p = 0

The blue p-values are not significant. The green p-values are significant but greater than zero. Essentially, all of the pairwise t-tests are significant at the alpha = 0.05 level except for two. Both of these two are for the Ford question, and all of the higher p-values are from one of the two Ford questions as well. This leads me to believe that the Ford questions are outliers since every single other question has a statistically significant difference with an extremely low p-value. Additionally, the Ford questions also appear to be outliers based on a Principal Components test (In R: SECTION 7: PRIN COMP GRAPH).

Based on t-tests for difference in means, I believe that I can reject my second null hypothesis that there is no difference between these two groups, and accept an alternative hypothesis that there is in fact a difference in these two groups, the difference being that store names that sound like family names can be made possessive.

Averaged

The t-test for the average of all the A.1.x questions against the average of all the A.2.x questions gives a t-score of 61.424 and a p-value of zero, which is also highly statistically significant and would support my hypothesis that these two groups are different.

Out of Curiosity

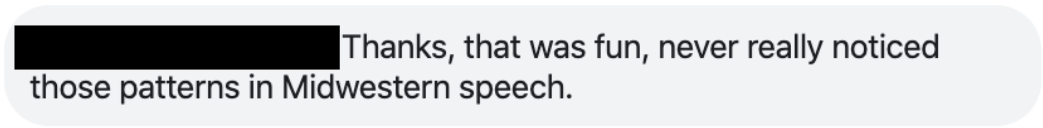
Out of curiosity, I decided to test the difference between the "work AT Ford's" question and the "work FOR Ford's" question. For this difference, the t-score is 1.4035 with a p-value of 0.1608, which is not statistically significant at an alpha = 0.05 level. I had included both questions to see if people would answer them differently, but in fact they are not significantly different. One other possible thing I would want to mention is that perhaps people rated these "Ford's" questions lower than the others because the question was in the survey twice, which might have been confusing. This could possibly be something to look into in the future. (In R: SECTION 8: AT VS FOR)

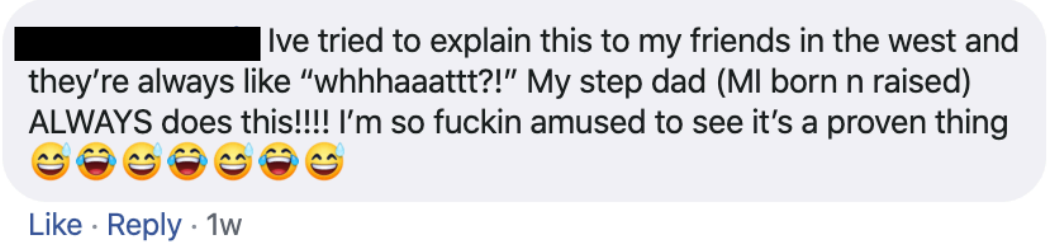
Social Media

People had a lot to say about this phenomenon on social media. I found this interesting due to our studies in class, especially about Black Twitter and how people speak on social media as opposed to more formally, as well as how they orthographically represent how they speak. I reached out a lot on social media for this survey, and the comments I got (unprompted comments, as many who finished the survey left comments afterward) fit mostly into four categories:

1. Restating my Hypothesis 1 (expressing excitement at this Midwestern phenomenon, even though now I don't necessarily believe this to be factually correct)

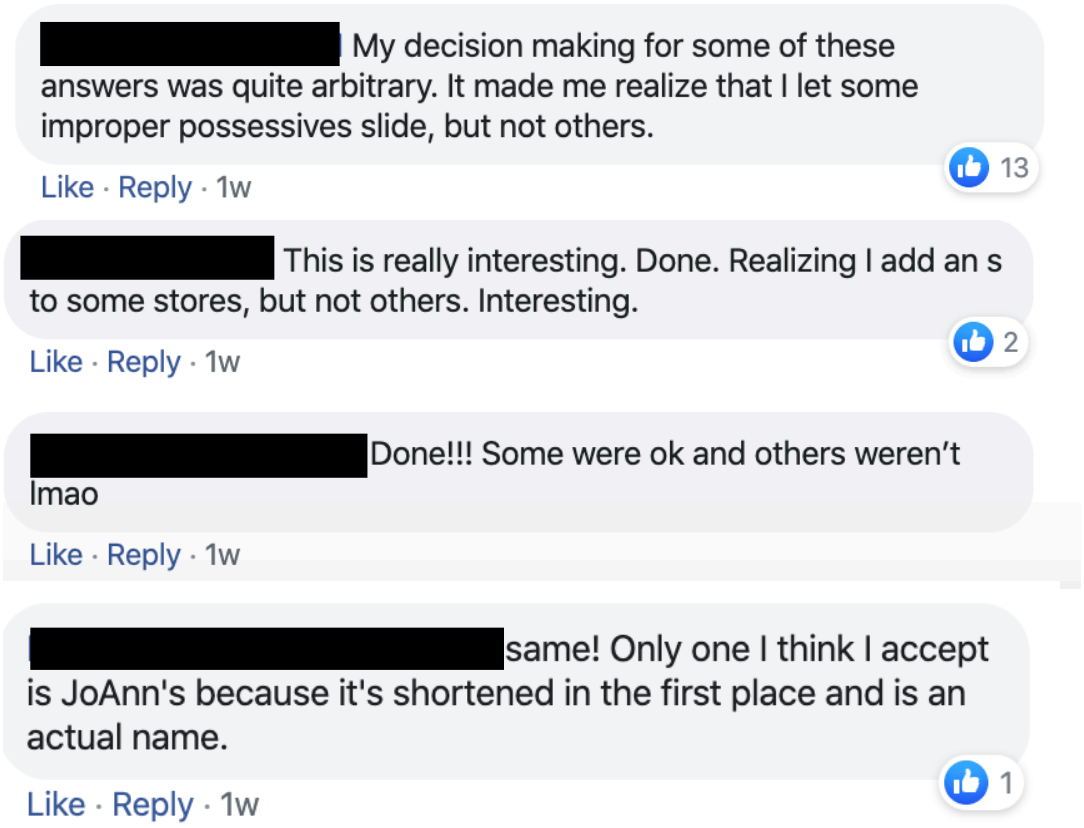
   Examples:





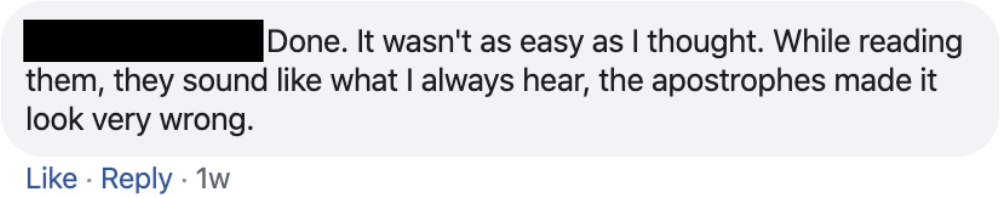
2. Restating my Hypothesis 2 (expressing that they could make some words possessive but not others)

   Examples:

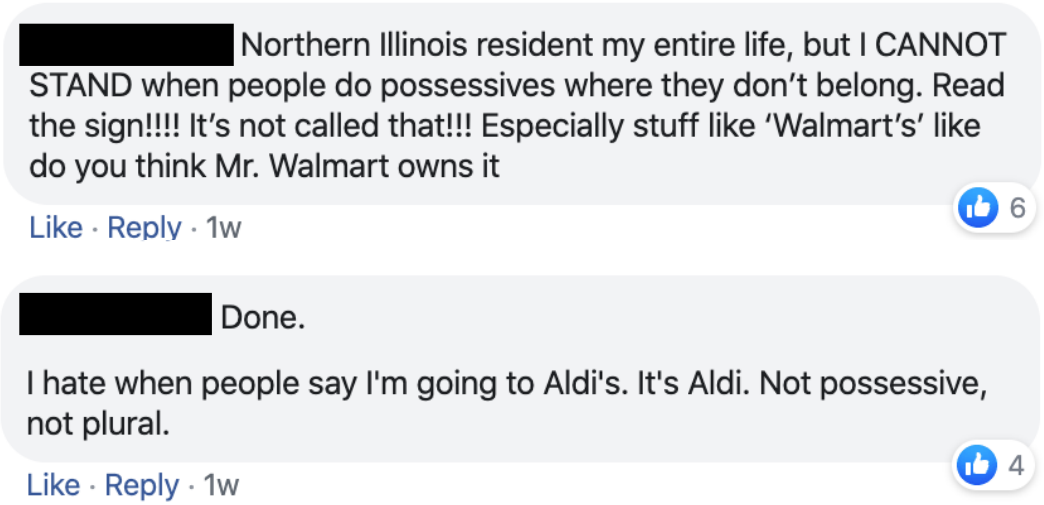


3. Expressing that they found the survey difficult because they would say aloud many of these words but believed that they would never write it down, and therefore it looked odd in survey form (orthographic concerns)

   Example:



4. Complaining about the phenomenon – there were many survey respondents who had an extreme negative response to this construction who included this in their comments as well. Examples:



Conclusion

Going back to my two hypotheses, it seems at least from my own survey work that it is very unlikely that my first hypothesis is true, and very likely that my second hypothesis is true. In other words, this phenomenon is not in fact a Midwestern phenomenon as I predicted, but the phenomenon does exist and does have a difference based on the type of name being tested. Perhaps all across the country, it is common or at least acceptable to make a business name possessive if it sounds like someone's name. The conjecture that this is a "folksy" or "rural" thing to do could still be somewhat true and still explain why only places that sound like names can use this construction.

What I would like to do in the future is obtain more data, especially from areas outside of the Midwest where I did not have as much data. I would also like to study in more detail exactly which names can take this construction, such as looking into why something like "Meijer's" scores twice as well as something like "Ford's." Ideally, I would do this survey again, obtain at least twice as many responses (n = 1000), and include more variety of questions to get a more specific answer.

References

"Create Custom Map." *MapChart*, mapchart.net/.

Culgan, Rossilynne. "Why Pittsburghers Add an 'S' to the End of Words." *The York Daily Record*, USA TODAY, 2 Apr. 2018, www.ydr.com/story/news/local/pennsylvania/2018/04/02/why-pittsburghers-add-s-end-words/477820002/.

Holdship, Deborah. "Michigan English." *Michigan Today*, umich.edu, 14 July 2010, michigantoday.umich.edu/2010/07/14/a7793/.

"How Y'all, Youse and You Guys Talk." *The New York Times*, The New York Times, 21 Dec. 2013, archive.nytimes.com/www.nytimes.com/interactive/2013/12/20/sunday-review/dialect-quiz-map.mobile.html?r=0011484808040060101200400100j100040k00802000l0200.

RStudio version 1.2.1335

Staff, Stateside. "The Story behind the Michigan 's.'" *Michigan Radio*, NPR, www.michiganradio.org/post/story-behind-michigan-s.

r code for ling project

Kelsey Evans

4/9/2020

SECTION 1: CLEANING & LIBRARIES

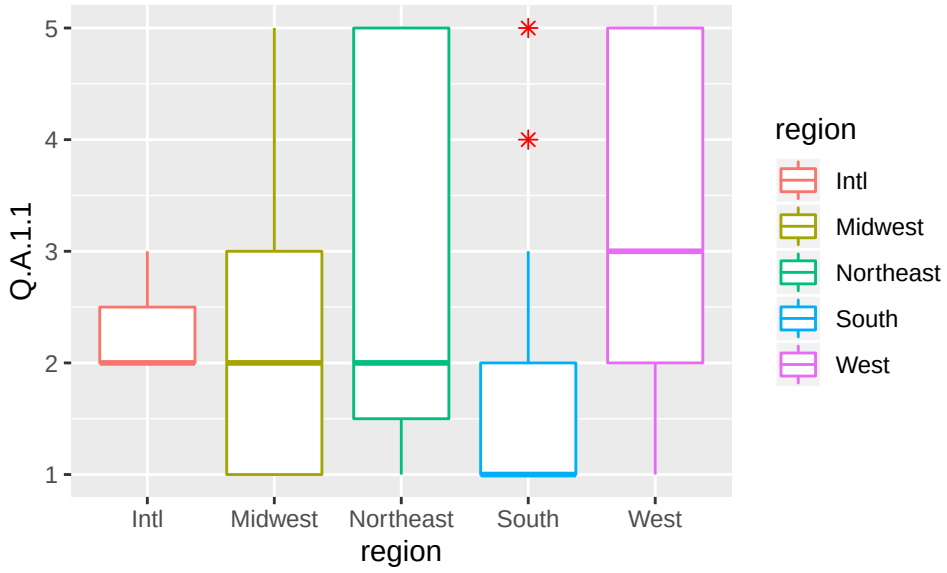
```
#ALL LIBRARIES HERE
library(dplyr)
library(ggplot2)
library(data.table)

#read it in
initial <- read.csv('initial.csv')

#take out extra stuff
initial <- initial[c(3:514), c(18, 21, 22, 25:54)]

#rename columns
names(initial) <- c("age", "state", "years", "Q.NA.1.1", "Q.A.1.1", "Q.A.2.1", "Q.A.1.2", "Q.NA.3.1", "(



#columns we want to change
change.vec = c("Q.NA.1.1", "Q.A.1.1", "Q.A.2.1", "Q.A.1.2", "Q.NA.3.1", "Q.NA.1.2", "Q.NA.2.1", "Q.A.1.3

#turn -99's to NA's
initial <- data.frame(lapply(initial, function(x) {is.na(x[x=="-99"]) <- TRUE; droplevels(x)}))

initial <- initial[complete.cases(initial), ]

#rename factors
initial[] <- lapply(initial, function(x) if(is.factor(x)) factor(x) else x)
for(i in 1:length(change.vec)) {
    levels(initial[,change.vec[i]]) = c("1", "2", "3", "4", "5")
}
for(i in 1:length(change.vec)) {
    levels(initial[,change.vec[i]]) = c("1", "2", "3", "4", "5")
}

#state column
initial$state <- as.character(toupper(initial$state))

#create region column
regionconvert3 <- function(x) {
  out <- x[NA]
  ind <- x %in% c("TX", "OK", "AR", "LA", "MS", "AL", "TN", "KY",
           "GA", "FL", "SC", "NC", "VA", "WV")
  out[ind] <- "South"
  ind <- x %in% c("ME", "NH", "NY", "MA", "RI", "VT", "PA",
              "NJ", "CT", "DE", "MD", "DC")
  out[ind] <- "Northeast"
```

```
  ind <- x %in% c("KS", "NE", "NE ", "SD", "ND", "MN", "MO", "IA", "IL",
            "IN", "MI", "WI", "OH")
  out[ind] <- "Midwest"
  ind <- x %in% c("WA", "OR", "CA", "NV", "AZ", "ID", "MT", "WY",
          "CO", "NM", "UT", "HI")
  out[ind] <- "West"
  ind <- x %in% c("INDIA", "PAKISTAN", "SWITZERLAND")
  out[ind] <- "Intl"
  return(out)
}
initial$region <- as.factor(regionconvert3(initial$state))
initial$state <- as.factor(initial$state)

#our df
survey <- initial

#change to numeric
survey[, 4:33] <- sapply(survey[, 4:33], as.numeric)
survey$age <- as.numeric(paste(survey$age))
survey$ years <- as.numeric(paste(survey$years))

#length if we ever need it
n <- length(survey$age)
```

SECTION 2: PLOTS

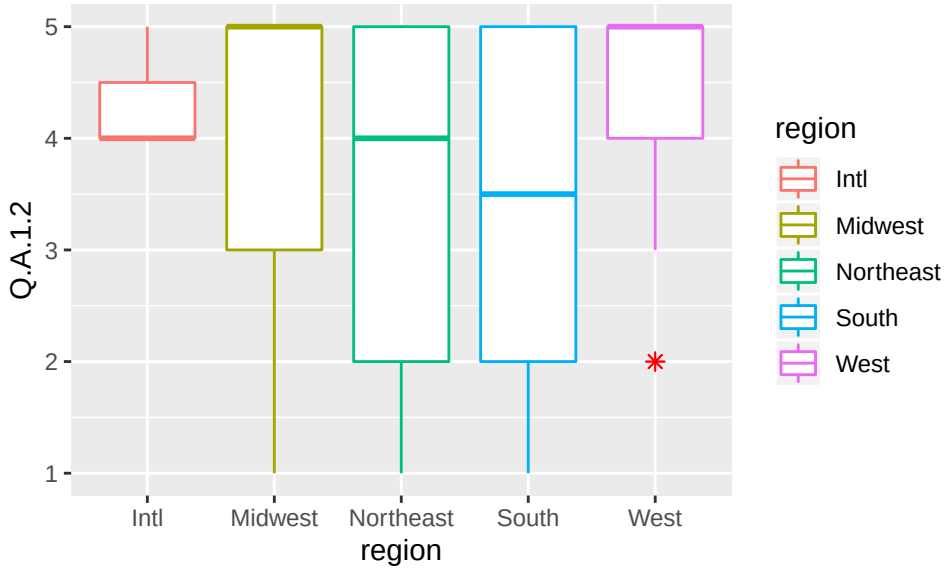
Boxplots:

```
p1 <- ggplot(survey, aes(region, Q.A.1.1, color = region)) +
  geom_boxplot(outlier.colour="red", outlier.shape=8,
               outlier.size=2)
p1 + theme(plot.margin = unit(c(2,2,2,2),"cm"))
```

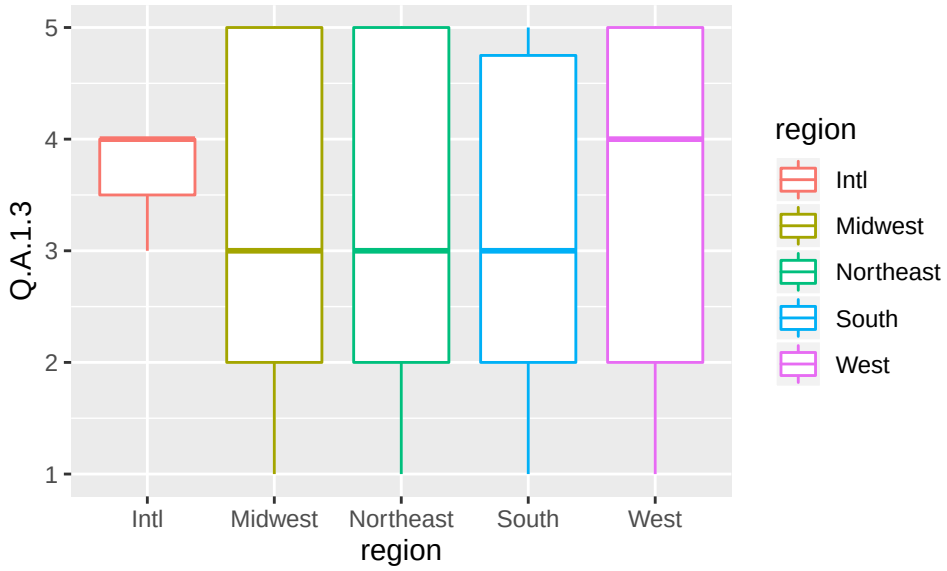



```
p2 <- ggplot(survey, aes(region, Q.A.1.2, color = region)) +
  geom_boxplot(outlier.colour="red", outlier.shape=8,
               outlier.size=2)
```

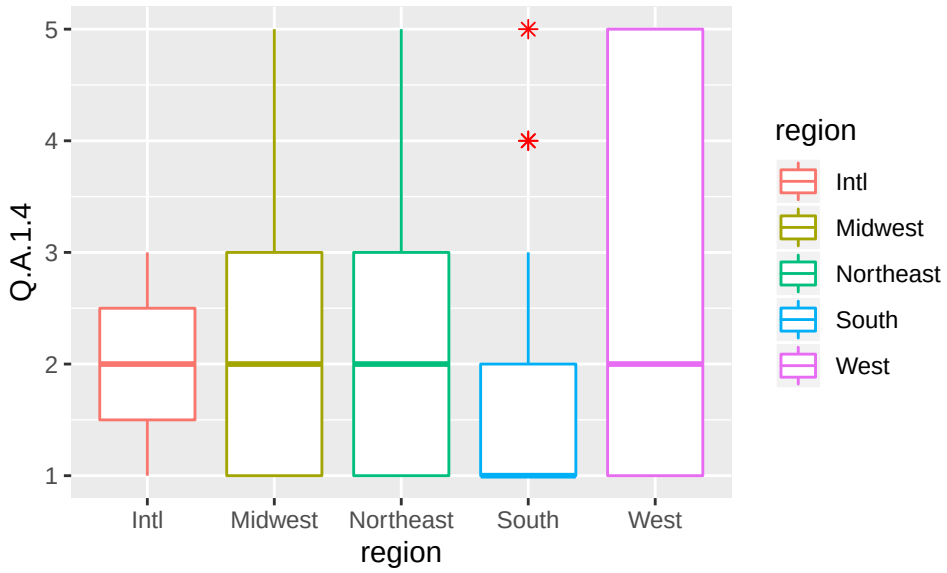
```
p2 + theme(plot.margin = unit(c(2,2,2,2),"cm"))
```



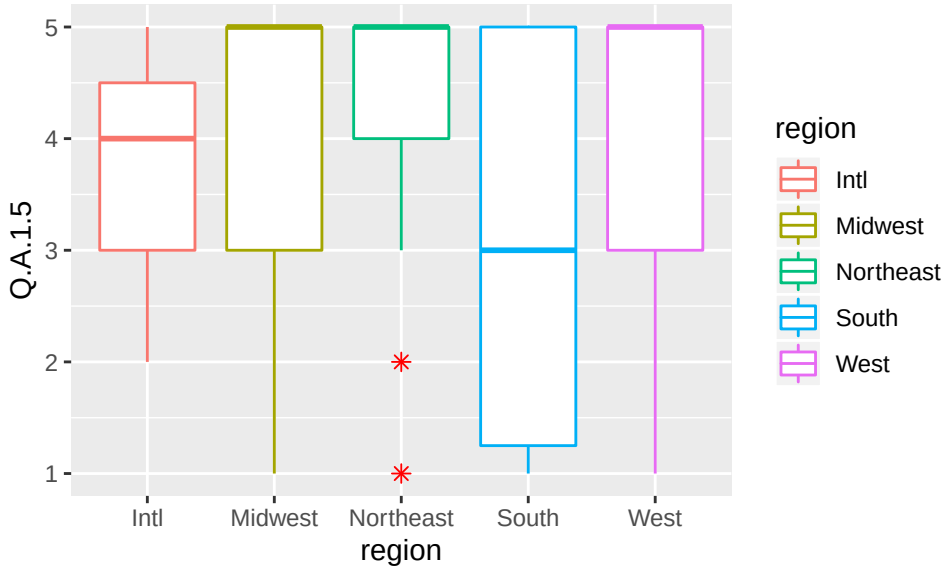

```
p3 <- ggplot(survey, aes(region, Q.A.1.3, color = region)) +
  geom_boxplot(outlier.colour="red", outlier.shape=8,
               outlier.size=2)
p3 + theme(plot.margin = unit(c(2,2,2,2),"cm"))
```



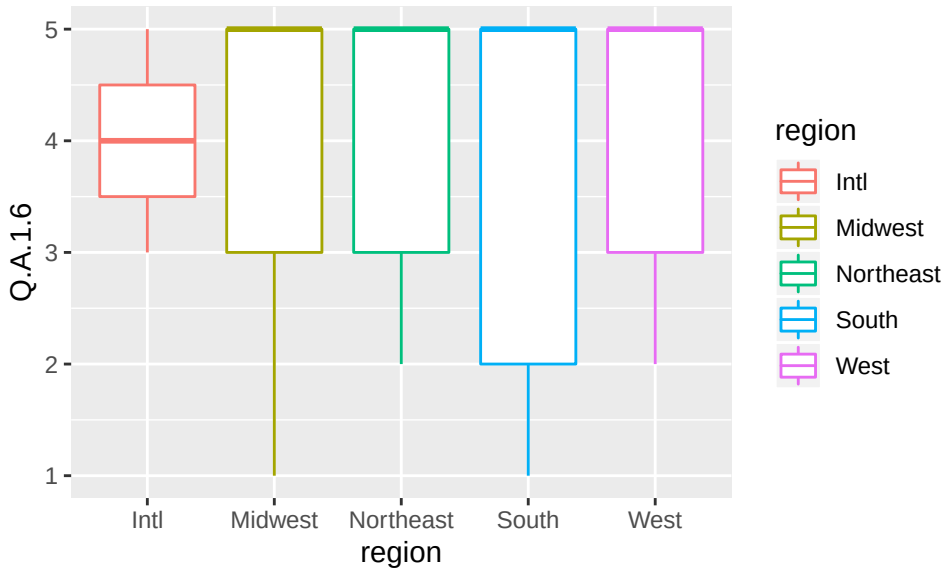

```
p4 <- ggplot(survey, aes(region, Q.A.1.4, color = region)) +
  geom_boxplot(outlier.colour="red", outlier.shape=8,
               outlier.size=2)
p4 + theme(plot.margin = unit(c(2,2,2,2),"cm"))
```



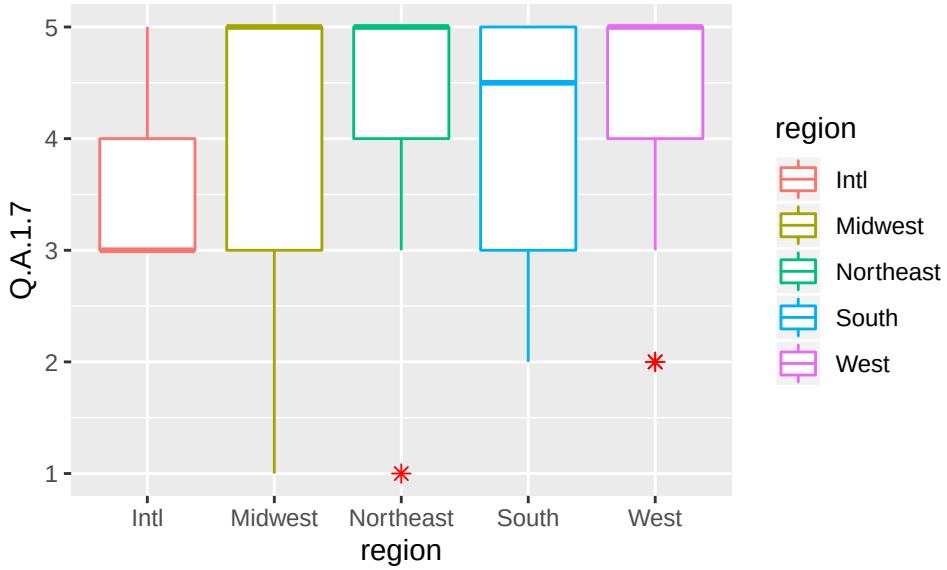

```
p5 <- ggplot(survey, aes(region, Q.A.1.5, color = region)) +
  geom_boxplot(outlier.colour="red", outlier.shape=8,
               outlier.size=2)
p5 + theme(plot.margin = unit(c(2,2,2,2),"cm"))
```



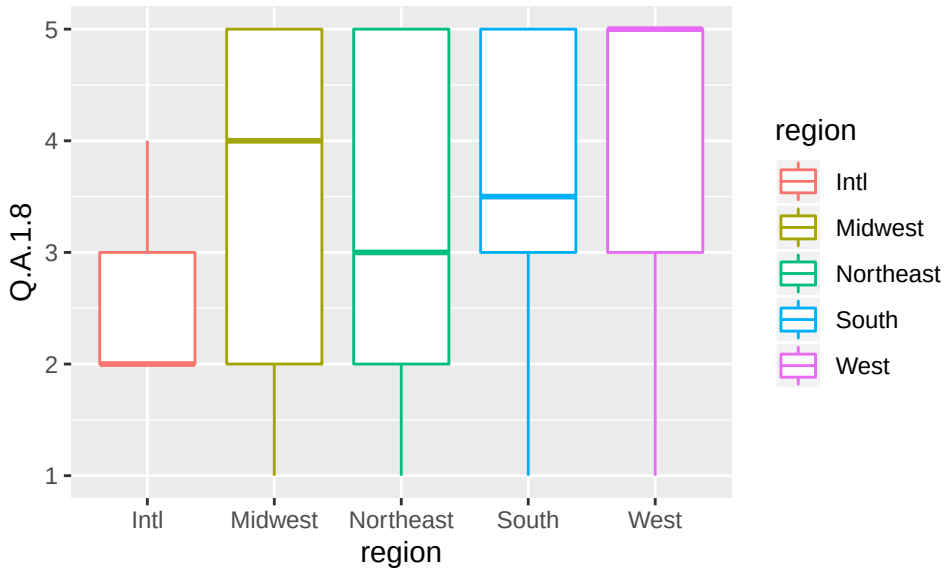

```
p6 <- ggplot(survey, aes(region, Q.A.1.6, color = region)) +
  geom_boxplot(outlier.colour="red", outlier.shape=8,
               outlier.size=2)
p6 + theme(plot.margin = unit(c(2,2,2,2),"cm"))
```



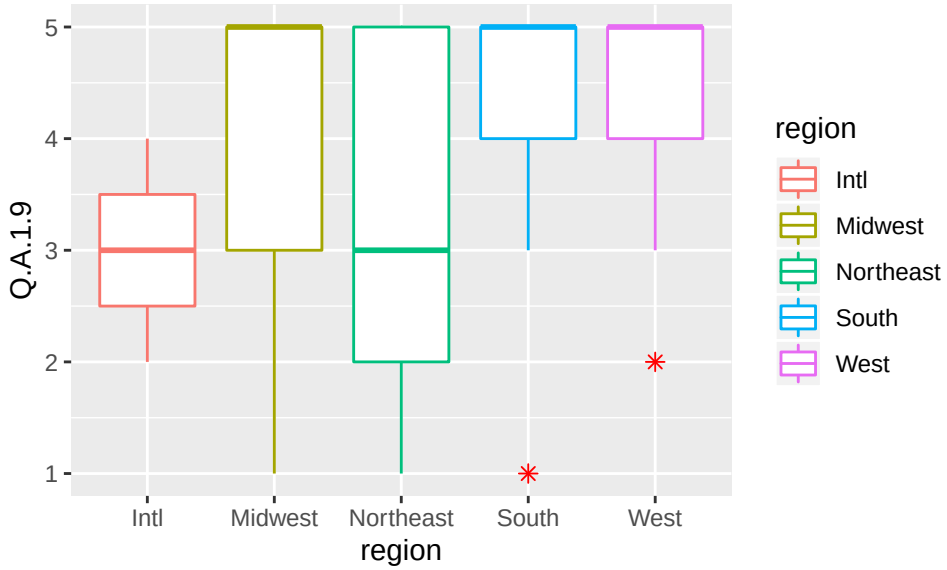

```
p7 <- ggplot(survey, aes(region, Q.A.1.7, color = region)) +
  geom_boxplot(outlier.colour="red", outlier.shape=8,
               outlier.size=2)
p7 + theme(plot.margin = unit(c(2,2,2,2),"cm"))
```



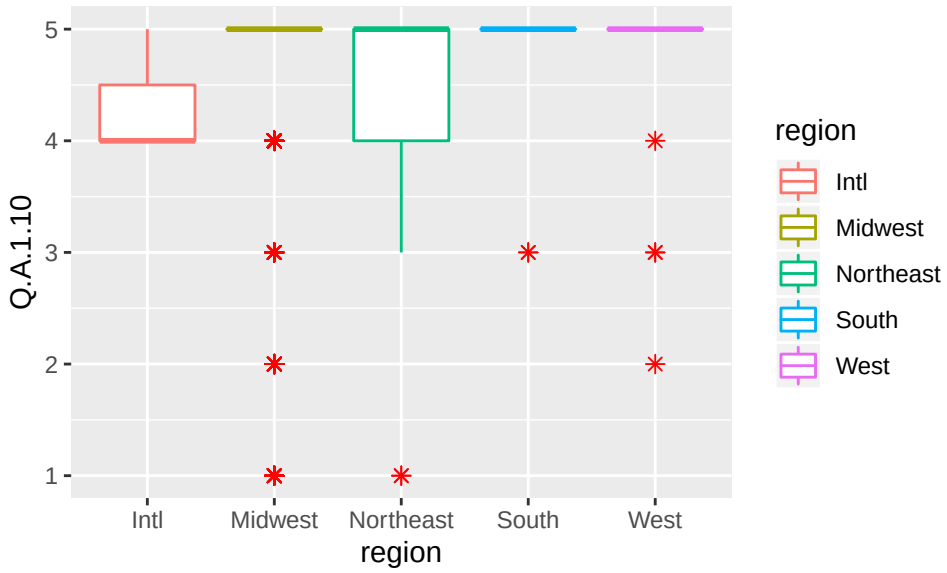

```
p8 <- ggplot(survey, aes(region, Q.A.1.8, color = region)) +
  geom_boxplot(outlier.colour="red", outlier.shape=8,
               outlier.size=2)
p8 + theme(plot.margin = unit(c(2,2,2,2),"cm"))
```




```
p9 <- ggplot(survey, aes(region, Q.A.1.9, color = region)) +
  geom_boxplot(outlier.colour="red", outlier.shape=8,
               outlier.size=2)
p9 + theme(plot.margin = unit(c(2,2,2,2),"cm"))
```

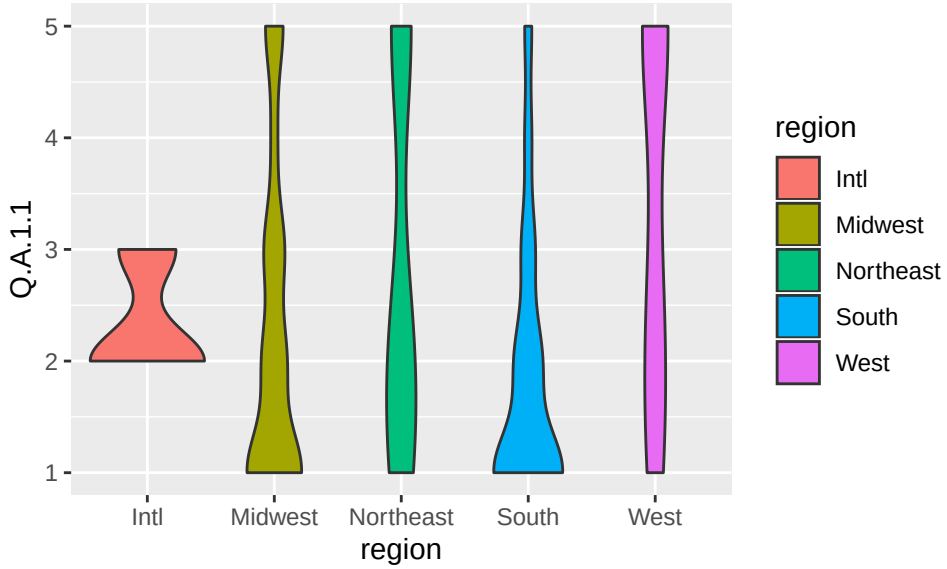



```
p10 <- ggplot(survey, aes(region, Q.A.1.10, color = region)) +
  geom_boxplot(outlier.colour="red", outlier.shape=8,
               outlier.size=2)
p10 + theme(plot.margin = unit(c(2,2,2,2),"cm"))
```

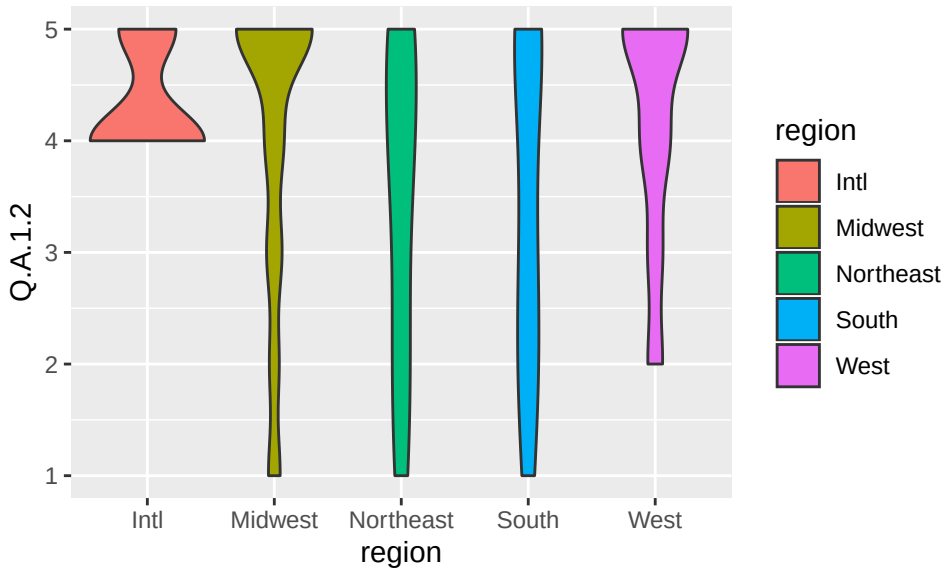


Violin Plots:

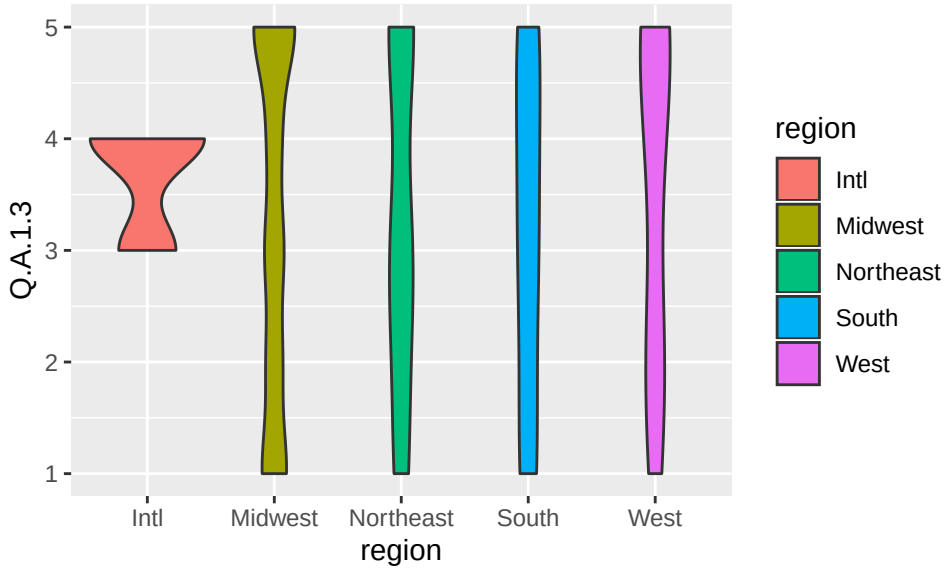
```
p11 <- ggplot(survey, aes(region, Q.A.1.1, fill = region)) +
  geom_violin()
p11 + theme(plot.margin = unit(c(2,2,2,2),"cm"))
```



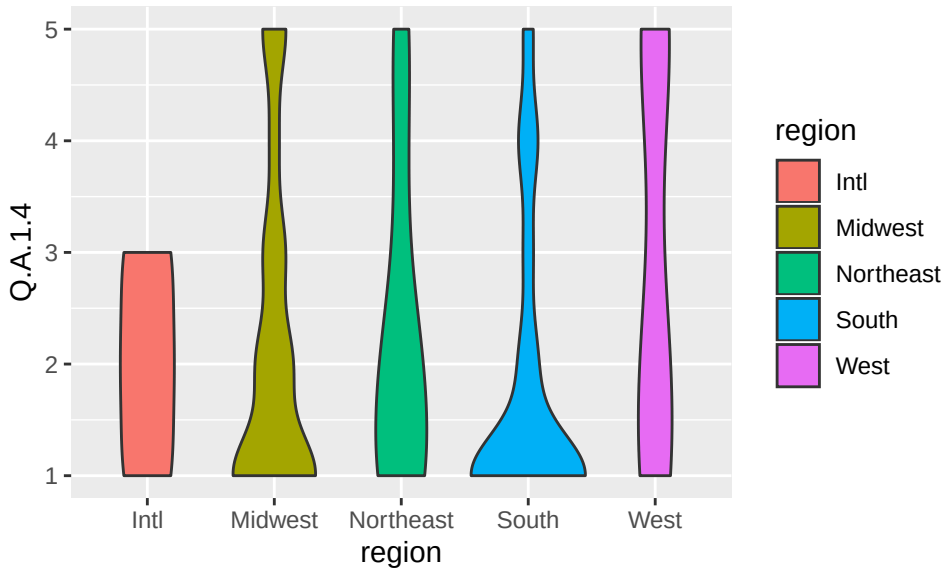

```
p12 <- ggplot(survey, aes(region, Q.A.1.2, fill = region)) +
  geom_violin()
p12 + theme(plot.margin = unit(c(2,2,2,2),"cm"))
```



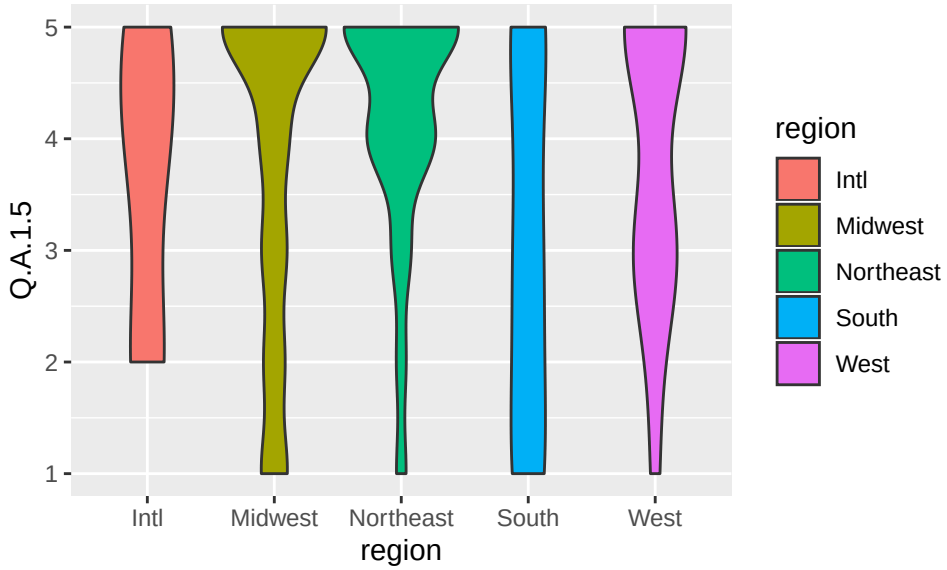

```
p13 <- ggplot(survey, aes(region, Q.A.1.3, fill = region)) +
  geom_violin()
p13 + theme(plot.margin = unit(c(2,2,2,2),"cm"))
```



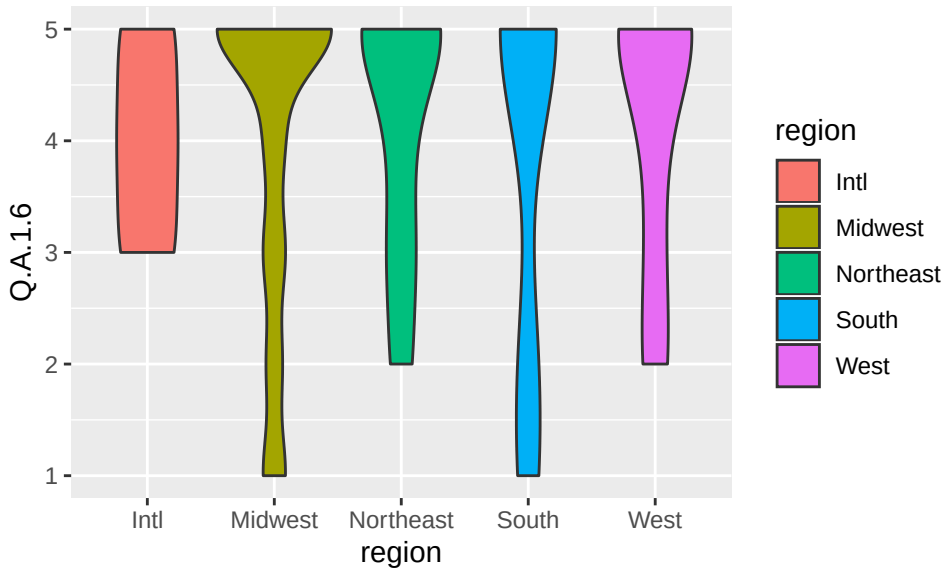

```
p14 <- ggplot(survey, aes(region, Q.A.1.4, fill = region)) +
  geom_violin()
p14 + theme(plot.margin = unit(c(2,2,2,2),"cm"))
```



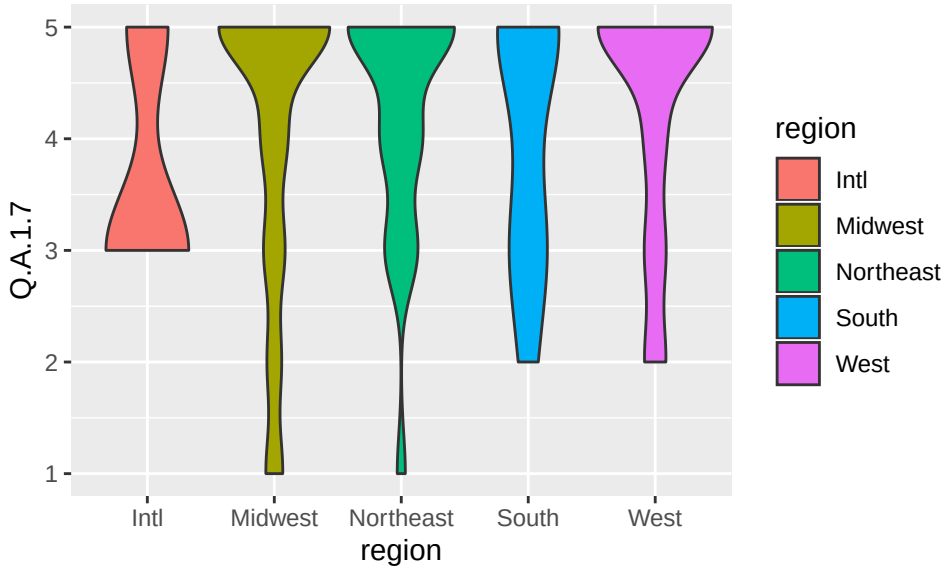

```
p15 <- ggplot(survey, aes(region, Q.A.1.5, fill = region)) +
  geom_violin()
p15 + theme(plot.margin = unit(c(2,2,2,2),"cm"))
```



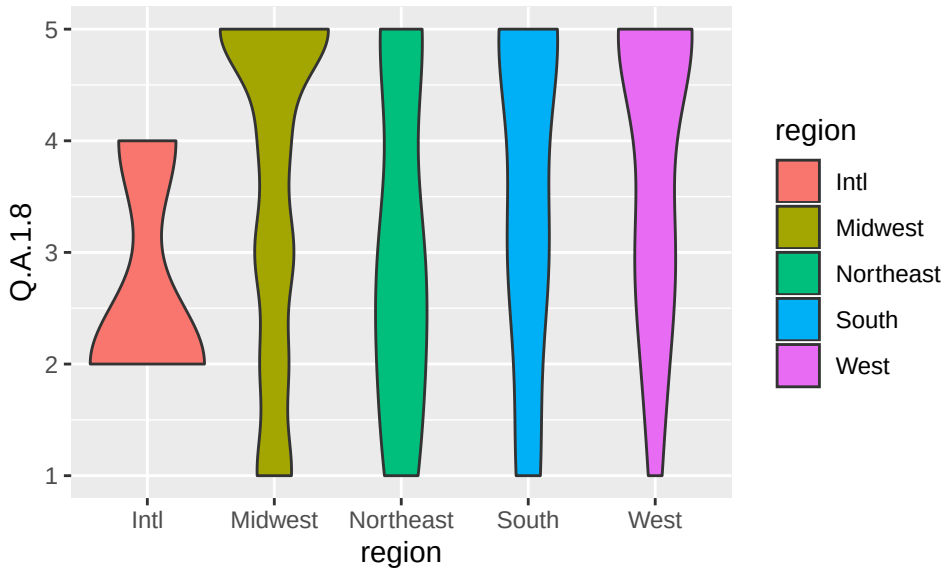

```
p16 <- ggplot(survey, aes(region, Q.A.1.6, fill = region)) +
  geom_violin()
p16 + theme(plot.margin = unit(c(2,2,2,2),"cm"))
```



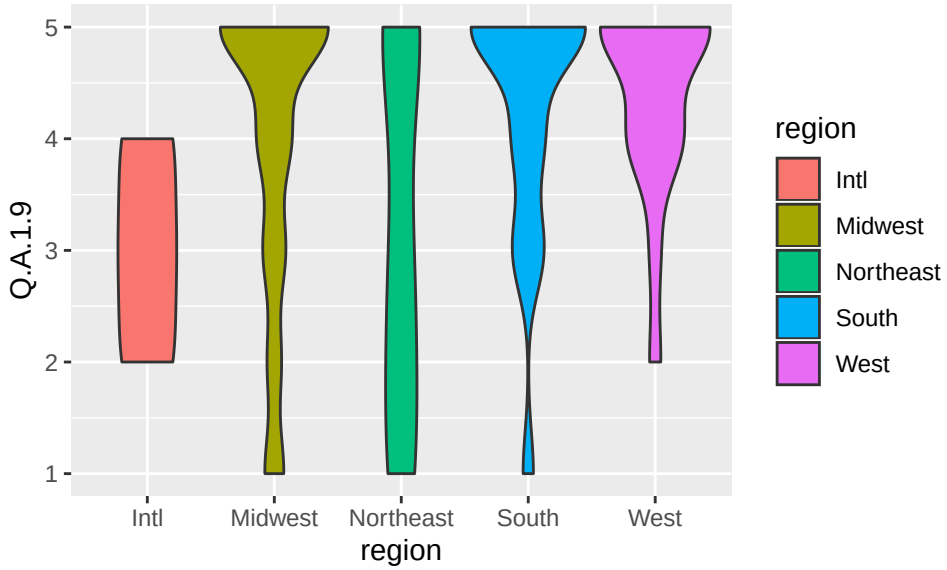

```
p17 <- ggplot(survey, aes(region, Q.A.1.7, fill = region)) +
  geom_violin()
p17 + theme(plot.margin = unit(c(2,2,2,2),"cm"))
```



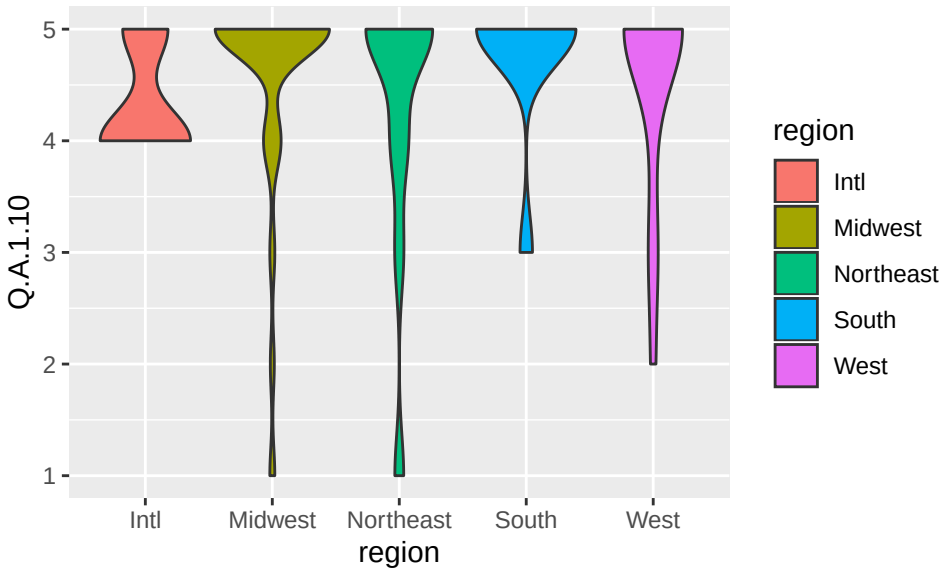

```
p18 <- ggplot(survey, aes(region, Q.A.1.8, fill = region)) +
  geom_violin()
p18 + theme(plot.margin = unit(c(2,2,2,2),"cm"))
```




```
p19 <- ggplot(survey, aes(region, Q.A.1.9, fill = region)) +
  geom_violin()
p19 + theme(plot.margin = unit(c(2,2,2,2),"cm"))
```




```
p20 <- ggplot(survey, aes(region, Q.A.1.10, fill = region)) +
  geom_violin()
p20 + theme(plot.margin = unit(c(2,2,2,2),"cm"))
```



SECTION 3: ANOVA

```
a <- summary(lm(survey$Q.A.1.1 ~ survey$region))
b <- summary(lm(survey$Q.A.1.2 ~ survey$region))
c <- summary(lm(survey$Q.A.1.3 ~ survey$region))
d <- summary(lm(survey$Q.A.1.4 ~ survey$region))
e <- summary(lm(survey$Q.A.1.5 ~ survey$region))
f <- summary(lm(survey$Q.A.1.6 ~ survey$region))
g <- summary(lm(survey$Q.A.1.7 ~ survey$region))
h <- summary(lm(survey$Q.A.1.8 ~ survey$region))
i <- summary(lm(survey$Q.A.1.9 ~ survey$region))
j <- summary(lm(survey$Q.A.1.10 ~ survey$region))
```

F-stats & p-values

```
##    value
## 3.630543
```

```
##       value
## 0.006278761
```

```
##    value
## 2.098308
```

```
##      value
## 0.07989633
```

```
##    value
## 0.168016
```

```
##    value
## 0.954621
```

```
##    value
## 1.635613
```

```
##     value
## 0.1639885
```

```
##    value
## 1.677974
```

```
##     value
## 0.1537785
```

```
##     value
## 0.1481241
```

```
##     value
## 0.9638014
```

```
##     value
## 0.3314603
```

```
##     value
## 0.8568229
```

```
##    value
## 1.208299
```

```
##     value
## 0.3063265
```

```
##    value
## 2.975182
```

```
##      value
## 0.01903319
```

```
##     value
## 0.7696976
```

```
##     value
## 0.5452789
```

SECTION 4: MAP DATA

```
survey$Q.A.1.avg <- (survey$Q.A.1.1 + survey$Q.A.1.2 + survey$Q.A.1.3 + survey$Q.A.1.4 + survey$Q.A.1.5

states <- setDT(survey)[, mean(survey$Q.A.1.avg), by=state]
states
```

```
##           state       V1
##  1:          MI 3.584232
##  2:          NY 3.584232
##  3:          TX 3.584232
##  4:          CA 3.584232
##  5:          IL 3.584232
##  6:          WI 3.584232
##  7:          IN 3.584232
##  8:          MO 3.584232
##  9:          OH 3.584232
## 10:          CO 3.584232
## 11:          CT 3.584232
## 12:          IA 3.584232
## 13:          SD 3.584232
## 14:          NJ 3.584232
## 15:          MN 3.584232
## 16:          KS 3.584232
```

```
## 17:          VA 3.584232
## 18:          NE 3.584232
## 19:          ND 3.584232
## 20:          MA 3.584232
## 21:          FL 3.584232
## 22:          TN 3.584232
## 23:          WA 3.584232
## 24:          OR 3.584232
## 25:          AR 3.584232
## 26:         NE  3.584232
## 27:          MD 3.584232
## 28:          NC 3.584232
## 29:          AZ 3.584232
## 30:          HI 3.584232
## 31: SWITZERLAND 3.584232
## 32:       INDIA 3.584232
## 33:    PAKISTAN 3.584232
##           state       V1
```

SECTION 5: AVERAGE SCORES

```
mean(survey$Q.A.1.1)
```

```
## [1] 2.301397
```

```
mean(survey$Q.A.1.2)
```

```
## [1] 4.00998
```

```
mean(survey$Q.A.1.3)
```

```
## [1] 3.295409
```

```
mean(survey$Q.A.1.4)
```

```
## [1] 2.173653
```

```
mean(survey$Q.A.1.5)
```

```
## [1] 3.808383
```

```
mean(survey$Q.A.1.6)
```

```
## [1] 3.988024
```

```
mean(survey$Q.A.1.7)
```

```
## [1] 4.073852
```

```
mean(survey$Q.A.1.8)
```

```
## [1] 3.618762
```

```
mean(survey$Q.A.1.9)
```

```
## [1] 3.99002
```

```
mean(survey$Q.A.1.10)
```

```
## [1] 4.582834
```

```r
mean(survey$Q.A.2.1)
```

```
## [1] 1.56487
```

```r
mean(survey$Q.A.2.2)
```

```
## [1] 1.598802
```

```r
mean(survey$Q.A.2.3)
```

```
## [1] 2.353293
```

```r
mean(survey$Q.A.2.4)
```

```
## [1] 2.47505
```

```r
mean(survey$Q.A.2.5)
```

```
## [1] 1.590818
```

SECTION 6: T-TESTS

Pairwise:

```r
surveysubi <- survey[, c(5, 7, 11, 15, 18, 20, 25, 27, 29, 32)]
surveysubj <- survey[, c(6, 16, 22, 30, 33)]

mat.t.tests <- matrix(NA, nrow = 10, ncol = 5)
for (i in 1:10){
  for (j in 1:5){
    mat.t.tests[i, j] <- t.test(surveysubi[, ..i], surveysubj[, ..j])$statistic
  }
}

mat.p.vals <- matrix(NA, nrow = 10, ncol = 5)
for (i in 1:10){
  for (j in 1:5){
    mat.p.vals[i, j] <- t.test(surveysubi[, ..i], surveysubj[, ..j])$p.value
  }
}

(mat.t.tests <- round(mat.t.tests, 2))
```

```
##        [,1]  [,2]  [,3]  [,4]  [,5]
##  [1,]  9.33  8.96 -0.57 -1.85  8.78
##  [2,] 32.64 32.45 19.06 17.01 31.41
##  [3,] 20.80 20.53 10.01  8.44 20.03
##  [4,]  7.82  7.44 -2.01 -3.25  7.29
##  [5,] 28.35 28.13 16.06 14.22 27.33
##  [6,] 31.22 31.02 18.31 16.36 30.10
##  [7,] 34.01 33.83 20.02 17.91 32.71
##  [8,] 25.53 25.29 13.79 12.05 24.61
##  [9,] 32.28 32.09 18.79 16.76 31.07
## [10,] 49.03 49.08 29.44 26.52 46.68
```

```r
(mat.p.vals <- round(mat.p.vals, 15))
```

```
##          [,1]     [,2]       [,3]        [,4]     [,5]
```

```
##  [1,] 0.0e+00 0.00e+00 0.56631574 0.063962158 0.00e+00
##  [2,] 0.0e+00 0.00e+00 0.00000000 0.000000000 0.00e+00
##  [3,] 0.0e+00 0.00e+00 0.00000000 0.000000000 0.00e+00
##  [4,] 1.5e-14 2.41e-13 0.04499041 0.001187869 6.39e-13
##  [5,] 0.0e+00 0.00e+00 0.00000000 0.000000000 0.00e+00
##  [6,] 0.0e+00 0.00e+00 0.00000000 0.000000000 0.00e+00
##  [7,] 0.0e+00 0.00e+00 0.00000000 0.000000000 0.00e+00
##  [8,] 0.0e+00 0.00e+00 0.00000000 0.000000000 0.00e+00
##  [9,] 0.0e+00 0.00e+00 0.00000000 0.000000000 0.00e+00
## [10,] 0.0e+00 0.00e+00 0.00000000 0.000000000 0.00e+00
```

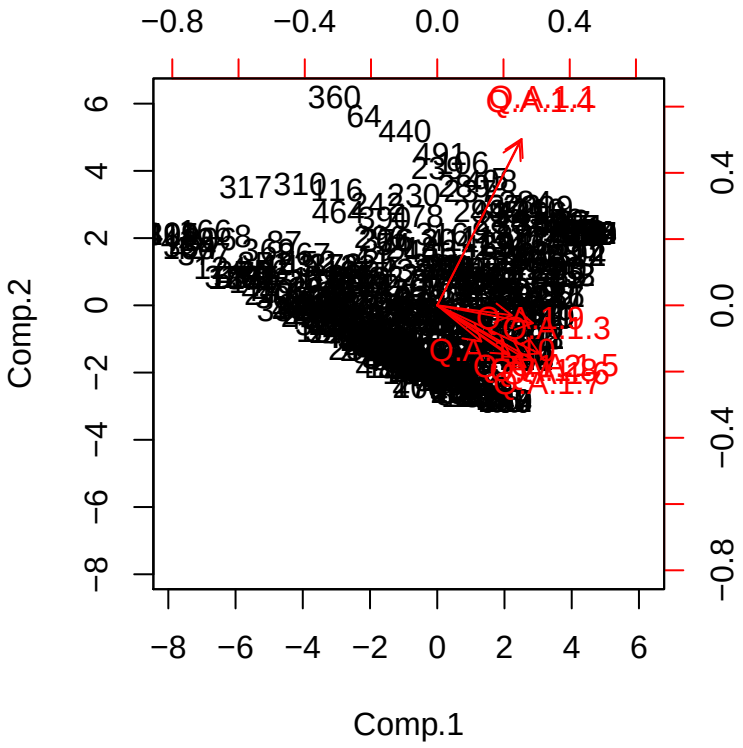
Averaged:

```
survey$Q.A.1.avg <- (survey$Q.A.1.1 + survey$Q.A.1.2 + survey$Q.A.1.3 + survey$Q.A.1.4 + survey$Q.A.1.5

survey$Q.A.2.avg <- (survey$Q.A.2.1 + survey$Q.A.2.2 + survey$Q.A.2.3 + survey$Q.A.2.4 + survey$Q.A.2.5

t.test(survey$Q.A.1.avg, survey$Q.A.2.avg)
```

```
##
##  Welch Two Sample t-test
##
## data:  survey$Q.A.1.avg and survey$Q.A.2.avg
## t = 61.424, df = 726.95, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.542018 2.709878
## sample estimates:
## mean of x mean of y
## 3.5842315 0.9582834
```

SECTION 7: PRIN COMP GRAPH

```
princomp <- princomp(surveysubi)
biplot(princomp, scale = 0)
```



SECTION 8: AT VS FOR

```
t.test(survey$Q.A.1.1, survey$Q.A.1.4)
```

```
##
##  Welch Two Sample t-test
##
## data:  survey$Q.A.1.1 and survey$Q.A.1.4
## t = 1.4035, df = 999.56, p-value = 0.1608
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.05087108  0.30636010
## sample estimates:
## mean of x mean of y
##  2.301397  2.173653
```