Final Project Proposal

Max Wagner

October 10, 2015

Research question

I was originally working with climate data, but a more interesting / maybe not as serious example came up. A survey on what people ate with their thanksgiving dinner. I am curious how/what each US region tends to eat in comparison to the others.

Cases

Every case in the set is an individual response to a number of questions. There are 999 observations in the set, with each observation containing up to 65 answers to 65 unique questions.

Data collection

A SurveyMonkey poll was used to collect data from 1,058 respondents. How people found out about the survey was not clear in the description on the fivethirtyeight Github page.

Type of study

This is an observational study in every way possible.

Data Source

The data source is the fivethirtyeight Github page. A link to the exact repository can be found here.

Response

If the idea of the study is correct, then the majority of the variables could be considered responses to the single explanatory variable. A few examples from the long list are:

- Do you celebrate Thanksgiving?
- What is typically the main dish at your Thanksgiving dinner?
- How is the main dish typically cooked?
- What kind of stuffing/dressing do you typically have?
- What type of cranberry sauce do you typically have?
- Do you typically have gravy?
- Which of these side dishes are typically served at your Thanksgiving dinner?
- Which type of pie is typically served at your Thanksgiving dinner?
- Which of these desserts do you typically have at Thanksgiving dinner?
- Do you typically pray before or after the Thanksgiving meal?
- How far will you travel for Thanksgiving?
- Will you watch any of the following programs on Thanksgiving?
- What's the age cutoff at your "kids' table" at Thanksgiving?

- Have you ever tried to meet up with hometown friends on Thanksgiving night?
- Have you ever attended a "Friendsgiving?"
- Will you shop any Black Friday sales on Thanksgiving Day?
- Will your employer make you work on Black Friday?

Explanatory

Under the assumption that region affects what people do/eat on the holidays, there is the single categorical variable describing the region of the respondent.

Relevant summary statistics

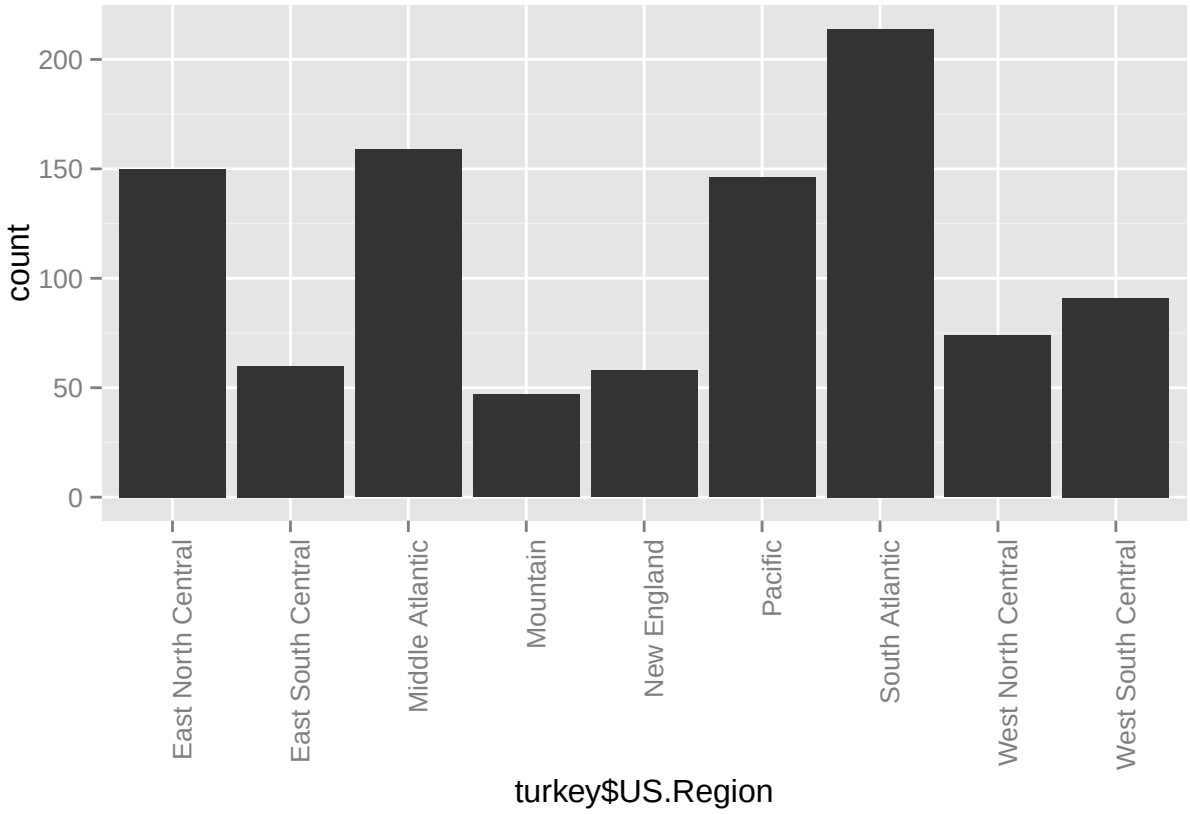
Provide summary statistics relevant to your research question. For example, if you're comparing means across groups provide means, SDs, sample sizes of each group. This step requires the use of R, hence a code chunk is provided below. Insert more code chunks as needed.

I think a good place to start is to which regions are represented, and how many respondents are from each. It's important to note that there are 59 responses that did not list a region.

```
turkey[turkey == ""] <- NA
summary(turkey$US.Region)
```

```
## East North Central East South Central    Middle Atlantic
##                150                 60                159
##           Mountain        New England             Pacific
##                 47                 58                146
##     South Atlantic West North Central West South Central
##                214                 74                 91
```

```
region.plot <- qplot(turkey$US.Region)
region.plot + theme(axis.text.x = element_text(angle = 90, hjust = 1))
```



The next piece to look at is the question asking respondents if they celebrated Thanksgiving. In this case 78 respondents answered no, which means 78 of the 1058 responses are empty.

```
summary(turkey$Do.you.celebrate.Thanksgiving.)
```

```
##  No Yes
##  68 931
```

If we include only responses that both celebrate Thanksgiving and listed a region, in narrows the size of the sample to 931 respondents.

```
turkey[turkey == ""] <- NA
nrow(subset(turkey, Do.you.celebrate.Thanksgiving. == "Yes" & !is.na(US.Region)))
```

```
## [1] 931
```

There are plenty more variables to look at, but I'll save those for the main project.