Descriptive Statistics

Jason Renn

September 19, 2016

Contents

The AMERICAN NATIONAL ELECTION STUDIES (ANES)

The ANES is a survey that covers voting behavior, public opinion, and political participation. While the primary mission of the ANES is to answer questions about voting behavior, the wealth of variables contained in the survey mean that we can use these data to answer other questions too.

Today's question: **How have the attitudes of Americans changed towards gays and lesbians over time?**

Tasks: Description and Comparison Across Groups

We will use some basic descriptive statistics in today's class. Remember that after generating a number or creating a figure, you must discuss and describe it. We will do the following:

- Using two American National Election Studies, from two different years, characterize the attitudes of all Americans toward gays and lesbians.
- Plot the distribution of attitudes for each year separately. Has the distribution changed from one year to the other?

- Use measures of central tendency and measures of dispersion to characterize each distribution. More importantly, use them to characterize any changes in the distribution from one time to the other.

- Then do the same for Whites and African Americans separately.

The Data

The ANES dataset contains hundred of variables. We are only interested in a few today. If you would like to know about the other variables contained inthe ANES, read the codebook.

This version of the ANES includes multiple waves or years. We will look at two sets of years, 2004 and 2008.

```
# There are many ways of importing data. Here is one.
setwd("~/uips-stat-share/workshops-tutorials/descriptive-statistics")
anes <- readRDS("anes-data.RData")

# write.csv(anes, file = 'anes.csv') Here is another anes <-
# read.csv('anes.csv')

anes <- anes[(which(anes$VCF0004 %in% c(1988, 2008))), ]
anes <- anes[, c("VCF0232", "VCF0004", "VCF0105B", "VCF0850")]
names(anes)[names(anes) == "VCF0004"] <- "year"
```

Variables of Interest

- Feeling thermometer for attitude toward Gays and Lesbians (*ft*)
- Year of survey (*year*)
- Race-ethnicity of the respondent (*black*)
- Religiosity (*religion*)

Feeling Thermometer Question for Gays and Lesbians

Some questions on the ANES are what we refer to as "feeling thermometer" questions. The answer to these questions produce a number from 0 to 100, with higher values corresponding to a "warmer" or more favorable view.

Question: *"I'd like to get your feelings toward some of our political leaders and other people who are in the news these days. I'll read the name of a person and I'd like you to rate that person using something we call the feeling thermometer. Ratings between 50 degrees and 100 degrees mean that you feel favorable and warm toward the person. Ratings between 0 degrees and 50 degrees mean that you don't feel favorable toward the person and that you don't care too much for that person. You would rate the person at the 50 degree mark if you don't feel particularly warm or cold toward the person. If we come to a person whose name you don't recognize, you don't need to rate that person. Just tell me and we'll move on to the next one."*

```
names(anes)[names(anes) == "VCF0232"] <- "ft"
```

Ethnicity-Race

Question: What racial or ethnic group or groups best describes you?

Responses:
1. White non-Hispanic
2. Black non-Hispanic
3. Hispanic
4. Other or multiple races, non-Hispanic

```
# We will recode so that this variable just indicates if the respondent is
# black
anes$black <- as.numeric(anes$VCF0105B == "(2) 2. Black non-Hispanic")

# Check table(anes$VCF0105B, exclude = NULL) table(anes$VCF0105B,
# anes$black, exclude = NULL)
```

Religiosity

Question: Which of these statements comes closest to describing your feelings about the Bible? You can just give me the number of your choice.

Responses:
1. The Bible is the actual Word of God and is to be taken literally, word for word
2. The Bible is the Word of God but not everything in it should be taken literally, word for word
3. The Bible is a book written by men and is not the Word of God

```
# We will recode so that this variable just indicates the highest level of
# religiosity Yes, this is Christianity-centric. You may want to look in the
# anes codebook and find another variable
anes$religion <- as.numeric(anes$VCF0850 == "(1) 1. The Bible is the actual Word of God and is to be")

# Check levels(anes$VCF0850) table(anes$VCF0850, anes$religion, exclude =
# NULL)
```

Description

Using two American National Election Studies, from two different years, characterize the attitudes of all Americans toward gays and lesbians.

How would you "characterize the attitues of all Americans?" We will begin with some measures of central tendency.

FT Attitudes towards Gays and Lesbians

Recall that the FT scores range from 0 to 100, with high values indicating "warmer" or more approving attitudes. How do the respondents in the ANES view gays and lesbians? I will produce some numbers in the code below. Your job is to discuss the data and explain in substantive terms how it describes the attitudes of Americans towards gays and lesbians.

```
# The Average of ft
mean(anes$ft, na.rm = T)
```

```
[1] 40.03928
```

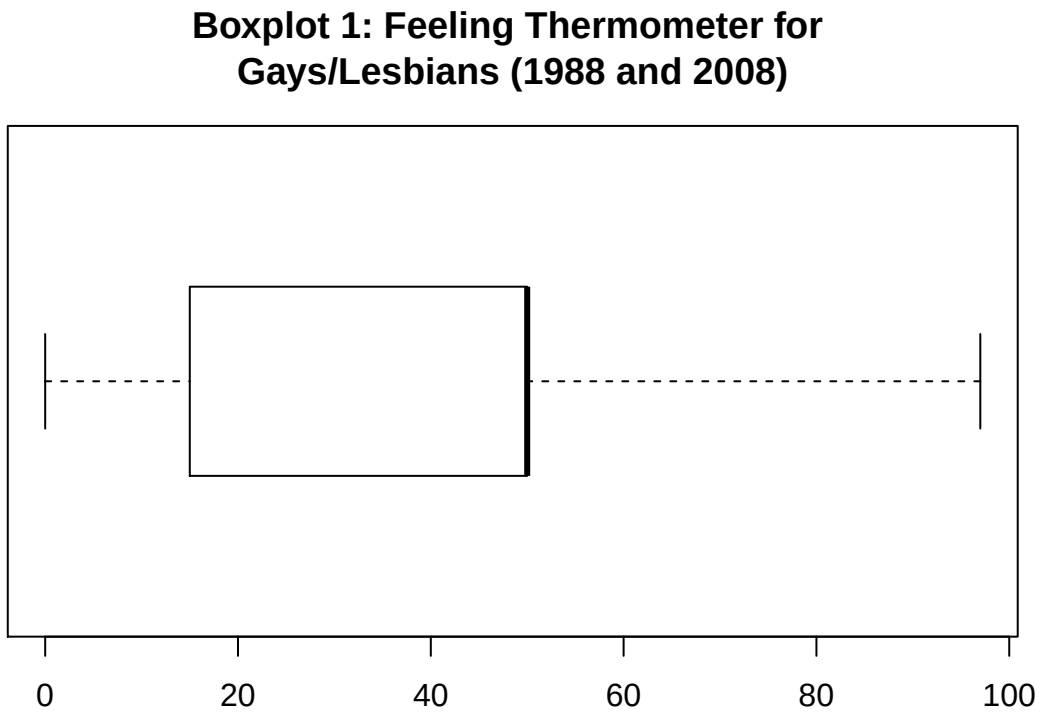
```r
# The Median
quantile(anes$ft, 0.5, na.rm = T)
```

```
50%
 50
```

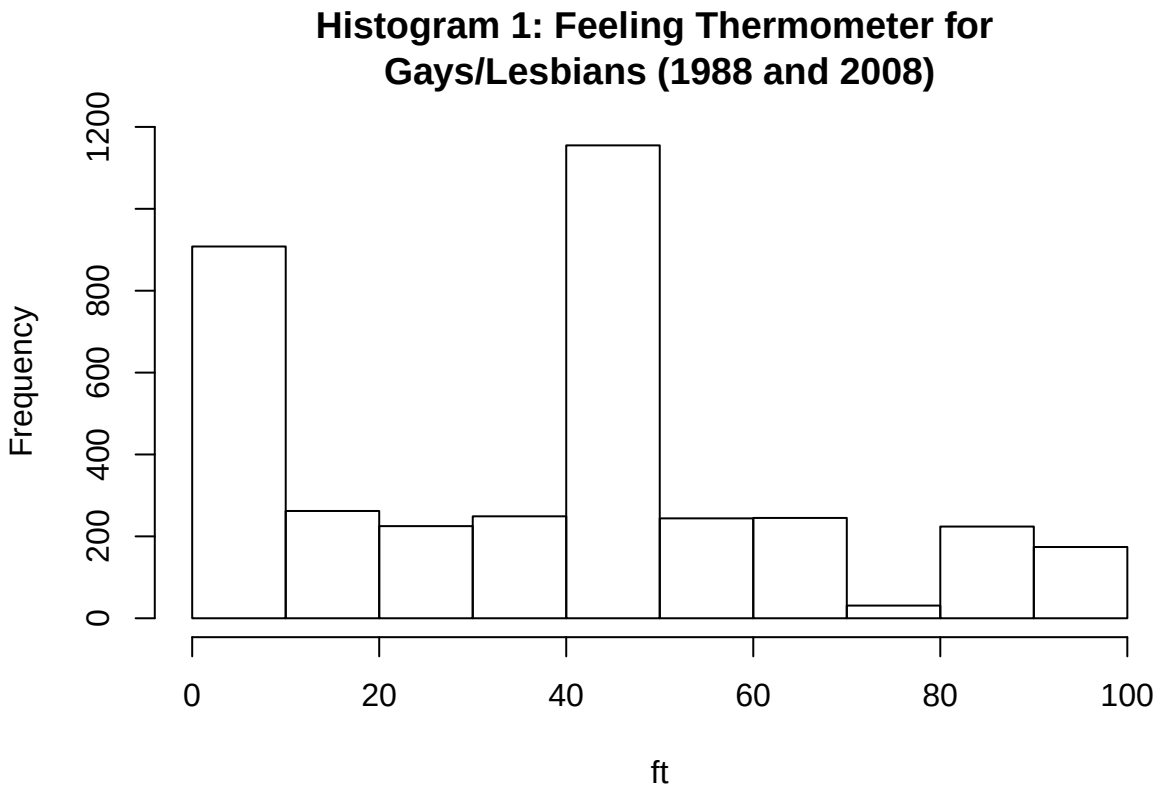
```r
# Alternatively, you could use summary summary(anes$ft)
```

We can visualize the distribution of attitudes towards gays and lesbians using a boxplot or a histogram. What does the distribution of attitudes tell you?

```r
with(anes, boxplot(ft, horizontal = T, main = "Boxplot 1: Feeling Thermometer for \nGays/Lesbians (1988
```




```r
with(anes, hist(ft, main = "Histogram 1: Feeling Thermometer for \nGays/Lesbians (1988 and 2008)"))
```



Disaggregating Attitudes Over Time

What is the point of looking at attitudes in two time periods? What do you expect to see when we do this? What do the descriptive statistics tell you?

```
# Start with the same summary statistics
with(subset(anes, year == 1988), summary(ft))
```

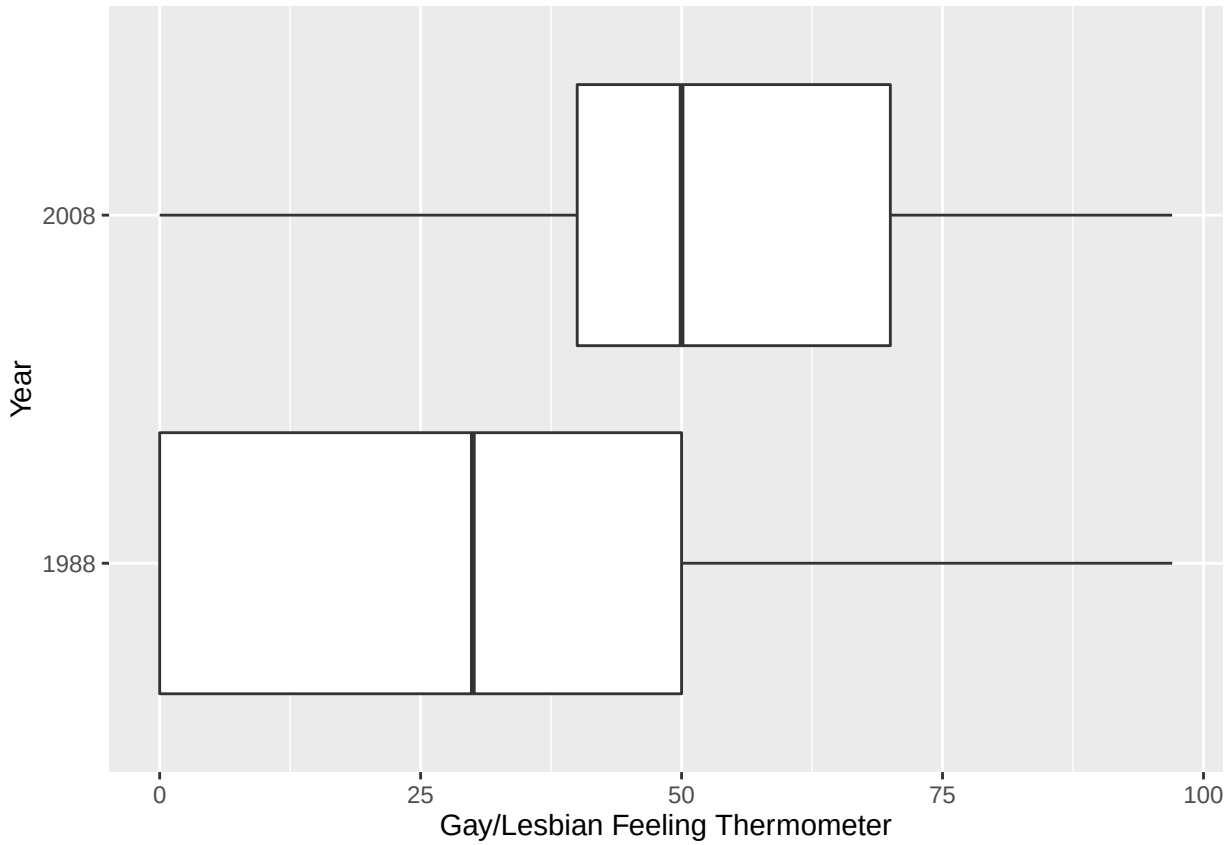
```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
    0.0     0.0    30.0    28.5    50.0    97.0     364
```

```
with(subset(anes, year == 2008), summary(ft))
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
   0.00   40.00   50.00   49.51   70.00   97.00     281
```

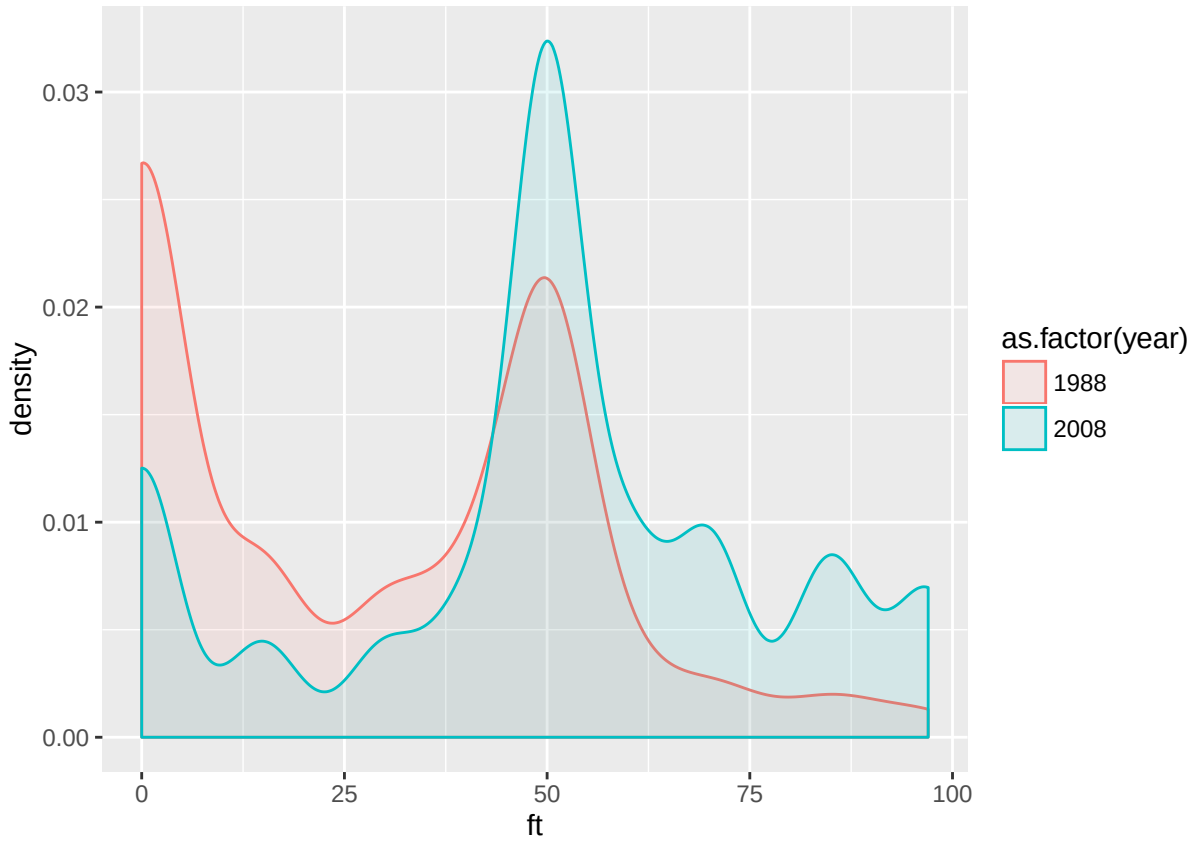
Let's use the same visual representations used above to depict the differences between attitudes in 1988 and 2008. What differences between the top and bottom boxplot would you highlight when talking about changes in attitudes towards gays and lesbians?

```
# Boxplot install.packages(ggplot2)
require(ggplot2)
ggplot(anes, aes(as.factor(year), ft)) + geom_boxplot() + coord_flip() + ylab("Gay/Lesbian Feeling Therm
    xlab("Year")
```

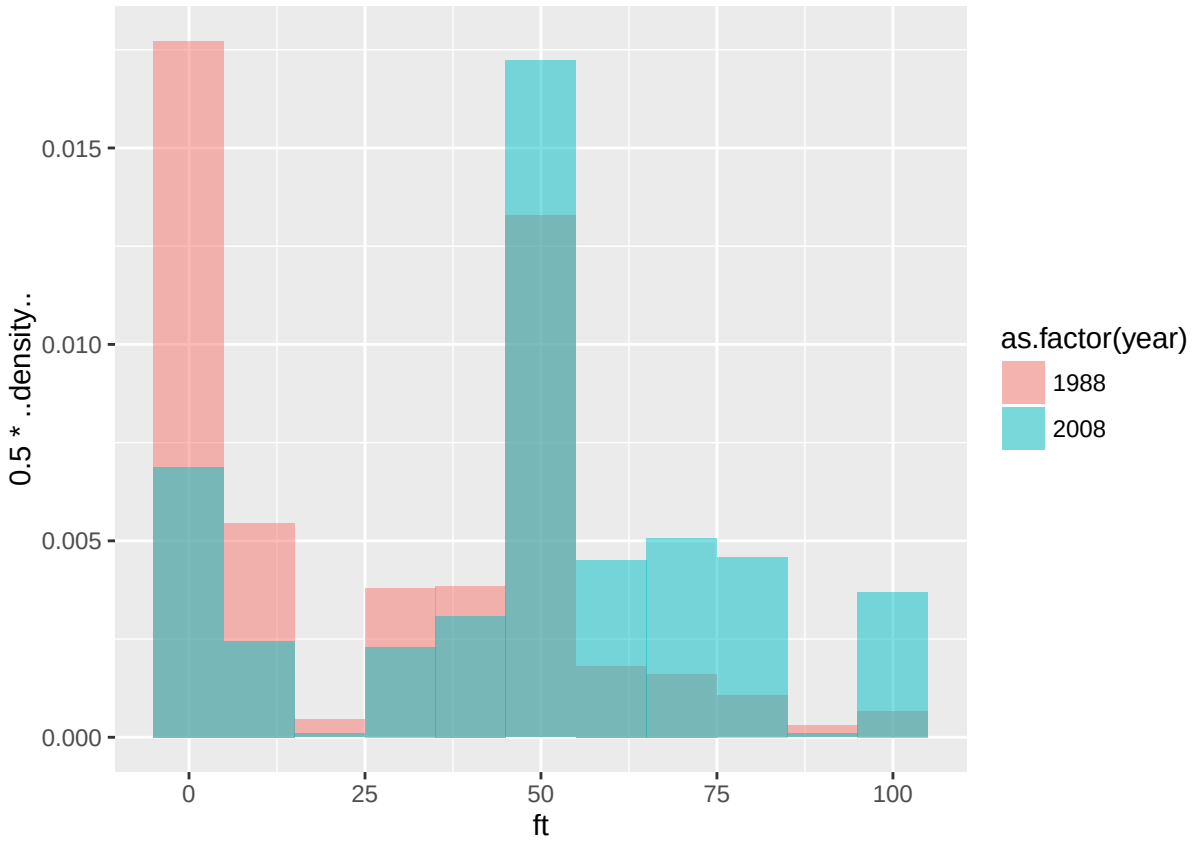


Here's one more way to look at the differences across our two survey periods. What strikes you about these figures?

```
# Hist
ggplot(anes, aes(color = as.factor(year), fill = as.factor(year), ft)) + geom_density(alpha = 0.1)
```




```
ggplot(anes, aes(x = ft, fill = as.factor(year))) + geom_histogram(aes(y = 0.5 *
    ..density..), alpha = 0.5, position = "identity", binwidth = 10)
```



Description: Attitudes by Ethnicity

```
# All Ethnic Groups
with(anes, summary(VCF0232))
with(anes, boxplot(VCF0232))
with(anes, hist(VCF0232))

# Separate Ethnic Groups
with(subset(anes, anes$VCF0105B == "(1) 1. White non-Hispanic"), summary(VCF0232))
with(subset(anes, anes$VCF0105B == "(2) 2. Black non-Hispanic"), summary(VCF0232))

# Boxplot
require(ggplot2)
ggplot(anes, aes(as.factor(VCF0105B), VCF0232)) + geom_boxplot() + ylab("Gay/Lesbian Feeling Thermometer
    xlab("Year")

# Histogram
with(subset(anes, VCF0105B == "(1) 1. White non-Hispanic"), hist(VCF0232))
with(subset(anes, VCF0105B == "(2) 2. Black non-Hispanic"), hist(VCF0232))

ggplot(anes, aes(x = VCF0232, fill = as.factor(VCF0105B))) + geom_histogram(aes(y = 0.5 *
    ..density..), alpha = 0.5, position = "identity", binwidth = 10)

# ggplot(anes,aes(x=VCF0232,fill=as.factor(VCF0105B)))+
```

```r
# geom_histogram(aes(y=0.5*..density..),binwidth=0.5)+
# facet_wrap(~as.factor(VCF0105B),nrow=4)

# Density
ggplot(anes, aes(color = as.factor(VCF0105B), fill = as.factor(VCF0105B), VCF0232)) +
    geom_density(alpha = 0.1)
```