

Crosstabs and Bivariate Viz

Preamble

```
setwd("C:/Users/19107/Desktop/R Stuff/2023 or Earlier")

library(ggplot2)
library(haven)
library(patchwork)
library(tidyr)
library(dplyr)
library(purrr)
library(tidyverse)
library(stargazer)
library(countrycode)
library(stringi)
```

Crosstabulation

We are moving toward the tools need for inference rather than “mere” description. One such tool is a Bivariate frequency distribution or Cross Tabulation (Crosstab). These are tables which calculate the frequency of each pair of categories across all possible pairs of categories for a pair variables. For example consider the simplest case, two dichotomous variables, their crosstab would be a 2x2 table.

Let’s use some real data though, the ANES is a good option here. Load the ANES

```
anes=read_dta("C:/Users/19107/Desktop/R Stuff/Data Locker/anes_cdf.dta")

# Give me some variables to examine
```

```

#Questions about trust in people and how much influence black Americans
↳ have
anes$trust=ifelse(anes$VCF9244>0,anes$VCF9244,NA) # high values = less
↳ trusting
anes$black_infl=ifelse(anes$VCF9275>0,anes$VCF9275,NA) # high values =
↳ less black influence

table(anes$trust,anes$black_infl, dnn=c("trust", "black infl"))

```

	black infl		
trust	1	2	3
1	13	80	48
2	632	3318	3415
3	486	2277	2013
4	463	1857	1844
5	79	315	361

```

# Questions about gubernatorial vote and the party that best handles the
↳ economy

```

```

anes$gov_vote=ifelse(anes$VCF9025>0,anes$VCF9025,NA)
anes$party_econ=ifelse(anes$VCF9205>0, anes$VCF9205, NA)

anes$gov_vote=ifelse(anes$gov_vote!=9,anes$gov_vote,NA)
anes$party_econ=ifelse(anes$party_econ!=7,anes$party_econ,NA)

```

```

# Note: 1 is democrat
table(anes$gov_vote, anes$party_econ, dnn=c("Gov Vote", "Econ Party"))

```

	Econ Party		
Gov Vote	1	2	3
1	828	634	173
2	97	488	936
3	4	24	26

```

anes$part_rep=ifelse(anes$VCF9203>0,anes$VCF9203,NA)
anes$voted=ifelse(anes$VCF0702>0,anes$VCF0702,NA)

```

```
table(anes$part_rep,anes$voted, dnn = c("Party Represent Me", "I
↪ Voted"))
```

```

                I Voted
Party Represent Me  1    2
                1 1110 7872
                2 1356 3666
```

```

anes$abortion=ifelse(anes$VCF0837<5, anes$VCF0837, NA)
anes$pres_choice=ifelse(anes$VCF0705<3, anes$VCF0705, NA)

anes$abortion=ifelse(anes$abortion==0,NA,anes$abortion)

anes$pres_choice=ifelse(anes$pres_choice==0,NA,anes$pres_choice)

table(anes$abortion,anes$pres_choice, dnn=c("abortion","pres choice"))
```

```

                pres choice
abortion      1    2
                1 168 193
                2 676 1000
                3 255 381
                4 469 526
```

Bivariate Viz

Just like with frequencies we can opt for a plot to show the relationship between two variables. Smooth is a useful tool for this purpose as it estimate a simple version of the relationship between two variables with a grey confidence band around the line to indicate uncertainty about the estimates.

Let's continue with the topic we were discussing ealier, race relations in the U.S., and we'll add in feelings towards the police.

Lastly, let's consider how these relationships might have changed over time.

```

anes$VCF0206 = ifelse(anes$VCF0206>97 , NA, anes$VCF0206)
anes$VCF0214 = ifelse(anes$VCF0214>97 ,NA, anes$VCF0214)
```

```
anes92=subset.data.frame(anes, anes$year==1992)

ggplot(anes92)+geom_smooth(aes(x=VCF0206,y=VCF0214))+
  xlab("Feeling Thermometer: Black Americans")+
  ylab("Feeling Thermometer: Police")
```

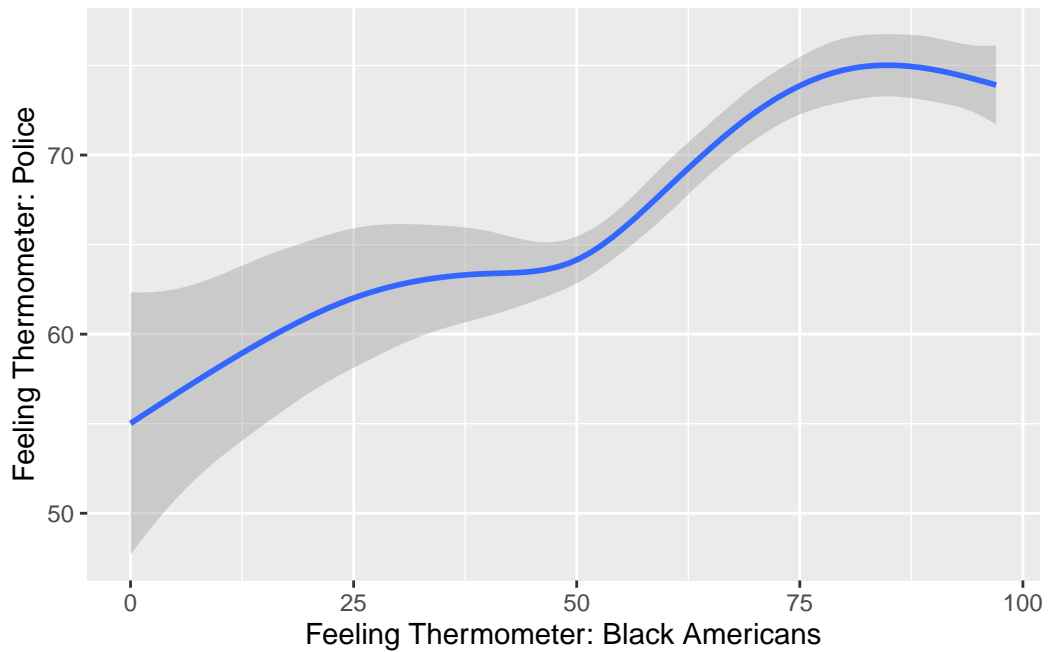


Figure 1: Rudimentary Estimation of the Relationship Between Feelings about Black Americans and Feelings about Police – 1992

```
anes20=subset.data.frame(anes, anes$year==2020)

ggplot(anes20)+geom_smooth(aes(x=VCF0206,y=VCF0214))+
  xlab("Feeling Thermometer: Black Americans")+
  ylab("Feeling Thermometer: Police")
```

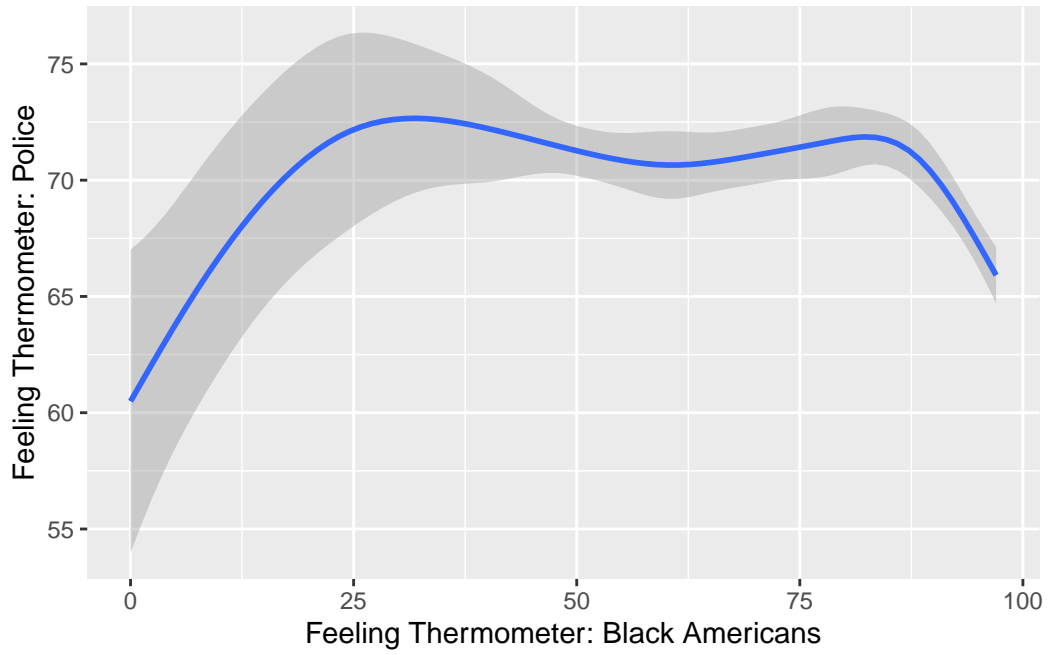


Figure 2: Rudimentary Estimation of the Relationship Between Feelings about Black Americans and Feelings about Police – 2020