EOC PS2

```
beef <- read_xlsx('olderbeef.xlsx')
beef$Year <- as.numeric(as.character(beef$Year))
```

```
## Warning: NAs introduced by coercion
```

```
beef <- mutate(beef , lagyear = Year + 20) %>%
  mutate(normcons = (percapconsumption/33.67591811) * 100)


murder <- read_xlsx('murder.xlsx')
murder$Year <- as.numeric(as.character(murder$Year))
murder$murder <- as.numeric(as.character(murder$murder))
murder <-  mutate(murder , normmurder = (murder/5.4) * 100)
```

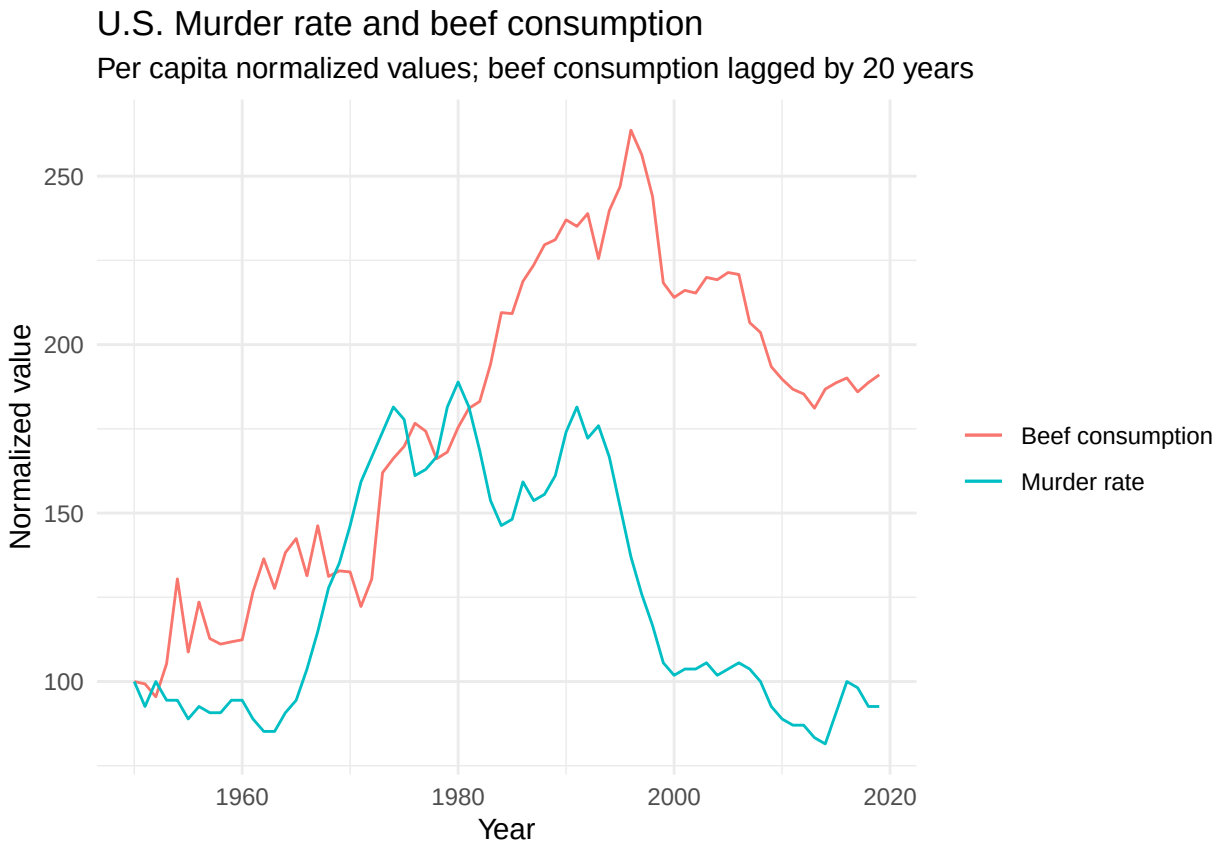
Question 1

Begin by collecting national data on annual U.S. beef consumption per capita from 1930 to the present (or something similar to that... the government collects a variety of related statistics) and annual U.S. homicide per capita from 1950 to the present. Make a graph that plots homicide rates against the 20 year lag of beef consumption. Compute the simple correlation between those two data series. Assuming the relationship is causal, how much could we reduce violent crime if we eliminated all beef consumption? Does this magnitude seem reasonable or unreasonable?

Response I used boneless beef availability data from the USDA as a proxy for beef consumption in the US. I used a combination of data from the FBI Unified Crime Reports and the U.S. Department of Health, Education, and Welfare (now Health and Human Services and Education) for the murder rate. Data from 1950-1960 is from the USHEW, 1960-2019 is from the FBI UCR.

```
ggplot() +
  geom_line(data = beef ,  mapping = aes(x = lagyear , y = normcons , color = 'Beef consumption' )) +
  geom_line(data = murder , mapping = aes(x = Year , y = normmurder , color = 'Murder rate')) +
  scale_x_continuous(name = 'Year' , limits = c(1950 , 2019)) +
  scale_y_continuous(name = 'Normalized value') +
  labs(colour = '' ,
       title = 'U.S. Murder rate and beef consumption' ,
       subtitle = 'Per capita normalized values; beef consumption lagged by 20 years') + theme_minimal()
```

```
## Warning: Removed 51 row(s) containing missing values (geom_path).
```




```
murder <- rename(murder , lagyear = Year)
data <- full_join(x = beef , y = murder, by = 'lagyear') %>%
  filter(Year > 1929)

model <- lm(murder ~ percapconsumption , data)
summary(model)
```

```
##
## Call:
## lm(formula = murder ~ percapconsumption, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.4220 -1.5501 -0.6474  1.4274  3.5355
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)         4.2367     0.8661   4.892 6.43e-06 ***
## percapconsumption   0.0411     0.0140   2.935  0.00455 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.786 on 68 degrees of freedom
##   (20 observations deleted due to missingness)
## Multiple R-squared:  0.1124, Adjusted R-squared:  0.09937
## F-statistic: 8.613 on 1 and 68 DF,  p-value: 0.004549
```

Question 2

If instead of national-level data, you had annual, state-level data from 1930-present on beef and homicide, describe how you would analyze it. Be specific. What would your regression look like? What covariates would you want to include?

Response

With state-level annual data I would run a regression controlling for state and time effects to use panel data. The use of panel data would control for a number of confounding factors, and allow for a more robust causal conclusion.

I would regress the murder rate on percent change in lagged, per capita beef consumption by state per year, setting 1930 (lagged to 1950) as the base level. I would include dummy variables for every year and every state to ensure the regression uses panel data and isolates the effect of increased beef consumption.

In addition to using state and year dummys, I would like to include controls for other major variables that effect crime. Ideally, this would include abortions per year and crack cocaine prevalence. By controlling for these two variables which have a significant effect on crime, the panel data could provide a better causal estimate of the effect of beef consumption on murder.

Question 3

Imagine you were able to run a randomized experiment, and that you could do whatever you wanted to, subject to a budget of $5 million. you want the experiment to be a good one. Describe the experiment you would run. Explain the choices you make. What are the weaknesses of your randomized experiment?

Response

I would run an RCT centering on a number of families in a single city. Ideally, I would randomize at the school level. Randomizing at the school level would limit

Question 4

Now, let's say you have zero budget. Professor Levitt wants you to find three different natural experiments that might shed light on this question. Give your three best ideas for possible natural experiments you might be able to exploit. I want these to be real things that you could actually explore. What are the strengths and weaknesses of your natural experiments.

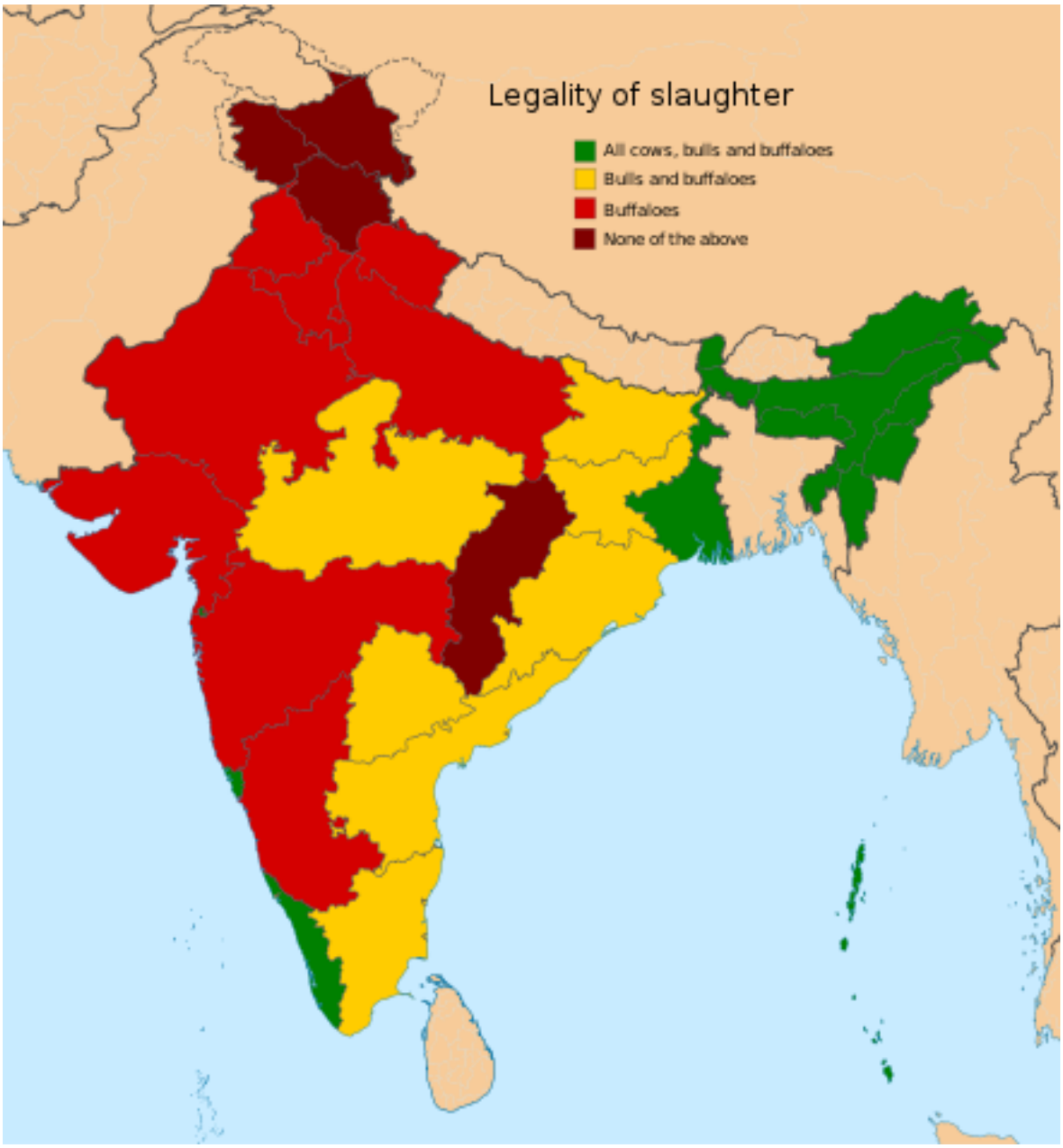
Response

1) As a result of various religious and cultural factors, the legality of slaughtering cows varies greatly across India. Exploiting these differences, I would want to either run a panel data regression if annual state murder data is sufficient, or run a Difference-in-Differences regression using legality as the treatment. The laws are quite variable and have changed significantly over time. An example difference-in-differences event could be the 1994 passage of a cattle slaughter ban in Delhi.\

This is a strong natural experiment because of this legislation's clear impact and reflection of beef eating habits on a national scale by region. I believe the fixed-effect regression would be more revealing than the DiD, as the passage of legislation that bans cattle-slaughter is likely reflective of cultural and religious factors

that imply limited beef consumption both before and after treatment. However, with the number of regions where beef consumption is more common, fixed effects regression could reveal a strong causal effect.\

This is a weak natural experiment because it establishes treatment and control groups that are fundamentally different. Because the consumption of beef is a reflection of the cultural and religious situation across India, comparing states which do consume beef to those which don't could allow for the introduction of numerous confounders that are difficult to control for. I am in no way an expert on Indian culture and beef consumption, and have no way of knowing or predicting the potential confounders that might arise as a result of this type of fixed effects regression. \



2)

Question 5

For one of your natural experiments in question 4, gather a little bit of data. . . just enough to get some idea of how big an impact your natural experiment had on beef consumption. If you absolutely can't find relevant data, try to make an educated guess at the impact. (In either case, don't worry about trying to figure out the actual impact on crime.)

Response