Effect of 400 user main run: redo with RSE

John Blakkan, Andres Borrero, Jason Liu

1. Introduction and motivation

Travelers of Interstate Highway 5 (I-5) pass a memorable landmark in California's Central Valley- the Harris Ranch feedlot. Located near the intersection of I-5 and California Route 198, it is readily visible from I-5 to motorists. It is also well-known for the pungent smell of thousands of cattle, usually noticable for several miles.

For city-dwellers and other travelers unfamiliar with feedlots, the sight (and smell) may be shocking. An obvious speculation is that beef consumers may, upon viewing the conditions under which cattle are raised in their final weeks prior to slaughter, exhibit a reduced demand for beef.

2. Summary findings

We found blah blah blah. See figure XYZ. TOTO: use RSE

3. Methods

3.1 Design

The objective of our study was to determine if viewing a cattle feedlot video causes reduction of consumer demand for beef. Subjects participating in the study were asked to take a survey (descibed in Section 3.3) that randomly showed one of three videos depicting cattle feedlots (treatment), irrigation systems (control), or cattle grazing in open pasture (placebo).

3.2 Subjects

Subjects within the United States and Colombia were recruited via Amazon Mechanical Turk. Overall, the study enrolled **XXX** subjects from the United States and **XXX** subject from Colombia.

3.3 Survey and Outcome Measures

Study participants were asked to take a survey. A copy is avialable at (**link**). The first part of the survey asks for some background information that we are interested in as co-variates. These include age, gender, geographical area (rural, suburbs, farm, city), and co-habitating pets. We believe that geographical area is an important co-variate to take into consideration because there may be varying levels of beef consumption depending on where the participant is located. For example, someone who lives in a more suburban area where beef may be less expensive might consume more beef at baseline compared to a person living in the city where beef is more expensive. Another possibility is that people who live on farms may have stronger feelings toward treatment of animals and cattle feedlots. Likewise, people who own pets may be more sympathetic to animals and have stronger reactions to animals living in poor conditions like those in the feedlot video.

After demographic data was collected, participants were then asked to estimate about how many times in the last week they ate pork, dairy, eggs, fruit, beef, and vegetables. The purpose of this question is to obtain a baseline consumption level for various food groups. It also serves to identify any potential vegetarians that may be taking part in the study. Participants were then shown one of three videos at random as treatment, control, or placebo. Each video depicted cattle feedlots (treatment), irrigation systems (control), or cattle grazing in open pasture (placebo). After treatment with video, participants were then shown images of various food items (hamburger, chicken, eggs, grain, and fruit/vegetables) and asked to rank them by preference for their next meal. Finally, the last question asks the participants to estimate how many times in the next week they expect to eat pork, dairy, eggs, fruit, beef, and vegetables.

To ensure compliance to the treatment videos, each video was embedded with numbers that would appear and flash at certain time points during the video. Participants were then asked to enter the numbers that appeared in the video.

The primary outcome of this study is the difference in ranking of hamburger between groups. The secondary outcome is the difference in differences between groups between the number of times the participant plans to eat beef in the next week and the number of times the participant ate beef in the pervious week.

3.4 Randomization

3.5 Data pipeline

3.6 Internationalization

4 Analysis

```
###df = read.csv("main_run.csv")   #This file has some dummy data in which femail participants who see a
                                # reduce their beef consumption next week by one meal, wiht probablity of

df = read.csv("https://thawing-shore-85209.herokuapp.com/get_csv/MainRunUS")
#rename some columns
column_names = c("mturkcode", "age", "sex", "living_status", "has_dog", "has_cat", "has_bird", "has_fish
                 "veg_last_week", "fruit_last_week", "dairy_last_week", "eggs_last_week", "beef_last_we
                 "pork_last_week", "hamburger_rank", "chicken_rank", "eggs_rank", "grain_rank",
                 "fruit_veg_rank", "veg_next_week", "fruit_next_week", "dairy_next_week", "eggs_next_we
                 "beef_next_week", "pork_next_week", "video_type", "attention_check")
colnames(df) <- column_names

#Remove anyone failing attention check
nrow(df)
```

```
## [1] 405
```

```
df <- df[df$attention_check == "true",]
nrow(df)
```

```
## [1] 395
```

```
# Refactor pet variables to 1/0; should change to look for non-blank, as Spanish-lanugage version
# may differ.
#df$has_dog <- df$has_dog == 'Dog'
#df$has_cat <- df$has_cat == 'Cat'
#df$has_bird <- df$has_bird == 'Bird'
#df$has_fish <- df$has_fish == 'Fish'
df$has_dog <- as.integer(df$has_dog != '')
df$has_cat <- as.integer(df$has_cat != '')
df$has_bird <- as.integer(df$has_bird != '')
```

```r
df$has_fish <- as.integer(df$has_fish != '')



#column for number of pets
df$num_pets <- df$has_dog + df$has_cat + df$has_bird + df$has_fish


# pre_ is weekly consumption before they watch the video
# post_ is weekly plan for next week


head(df)
```

```
##    mturkcode age    sex                 living_status has_dog has_cat
## 1 9465313139  26   Male A small town or suburban area       1       0
## 2  680067078  27   Male A small town or suburban area       0       1
## 3 9779571387  30 Female                  A rural area       0       1
## 4 1743430299  33 Female                  A rural area       1       1
## 5 9906804100  41 Female                        A city       0       1
## 6 7893580021  29 Female A small town or suburban area       1       1
##   has_bird has_fish veg_last_week fruit_last_week dairy_last_week
## 1        0        0             5               6               7
## 2        0        0             7               7               7
## 3        1        0             2               3               5
## 4        0        0             7               7               0
## 5        1        0             8              15               5
## 6        0        0             6               5              10
##   eggs_last_week beef_last_week pork_last_week hamburger_rank chicken_rank
## 1              2              3              3              1            2
## 2              0              3              0              3            5
## 3              2              0              0              5            2
## 4              0              0              0              4            5
## 5              1              0              0              5            4
## 6              4              3              0              3            1
##   eggs_rank grain_rank fruit_veg_rank veg_next_week fruit_next_week
## 1         5          4              3             5               5
## 2         4          2              1             7               7
## 3         4          1              3             3               4
## 4         3          2              1             7               7
## 5         3          2              1            12              15
## 6         5          2              4             5               5
##   dairy_next_week eggs_next_week beef_next_week pork_next_week video_type
## 1               7              3              5              3          F
## 2               7              0              3              0          F
## 3               5              2              1              0          P
## 4               0              0              0              0          F
## 5               4              1              0              0          I
## 6               5              2              3              0          F
##   attention_check num_pets
## 1            true        1
## 2            true        1
## 3            true        2
## 4            true        2
```

```
## 5             true        2
## 6             true        2
```

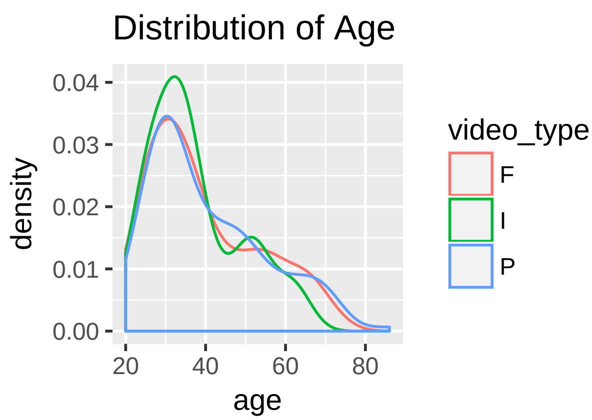
4.1 Baseline Characteristics

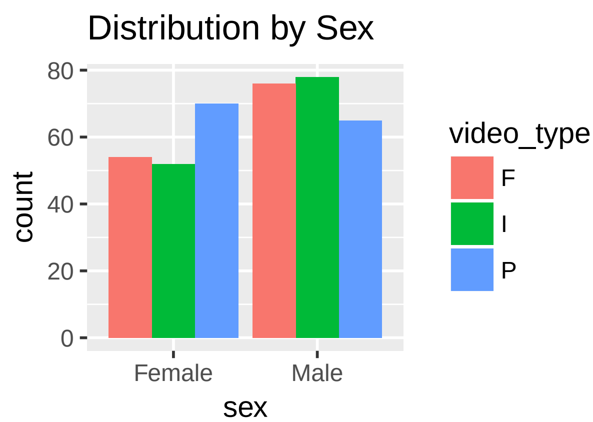
We plot various baseline characteristics that were collected such as age, sex, living status, and pet ownership.

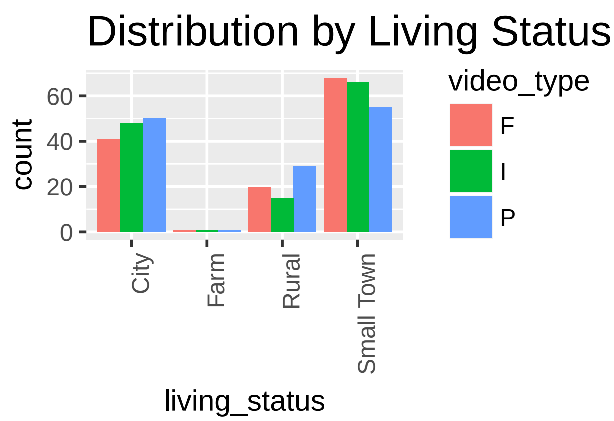
```
p1 <- ggplot(df, aes(age, color = video_type)) +
      geom_density() +
      ggtitle("Distribution of Age")
p2 <- ggplot(df, aes(x = sex, fill = video_type)) +
      geom_bar(stat = "count", position = position_dodge()) +
      ggtitle("Distribution by Sex")
p3 <- ggplot(df, aes(x = living_status, fill = video_type)) +
      geom_bar(stat = "count", position = position_dodge()) +
      ggtitle("Distribution by Living Status") +
      theme(axis.text.x = element_text(angle = 90, hjust = 1), plot.title = element_text(size = 16)) +
      scale_x_discrete(labels = c("City", "Farm", "Rural", "Small Town"))

pets <- df[c("has_dog", "has_cat", "has_bird", "has_fish", "video_type")]
pets_melt <- melt(pets)
```

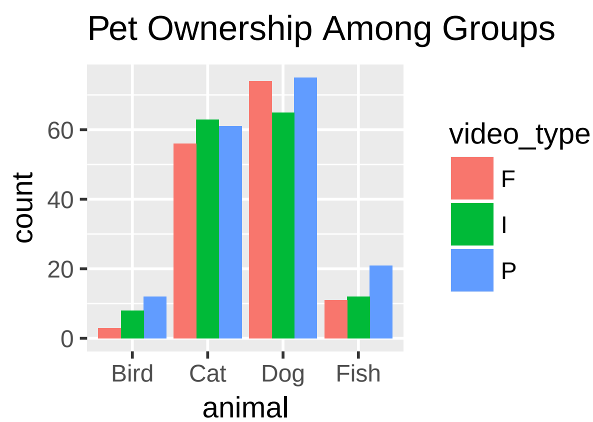
```
## Using video_type as id variables
```

```
pets_melt <- pets_melt[pets_melt$value == 1, ]

get_pet <- function(x) {
  if (x == "has_dog") {
    return("Dog")
  } else if (x == "has_cat") {
    return("Cat")
  } else if (x == "has_bird") {
    return("Bird")
  } else if (x == "has_fish") {
    return("Fish")
  }
}

pets_melt$animal <- sapply(pets_melt$variable, get_pet)

p4 <- ggplot(pets_melt, aes(x = animal, fill = video_type)) +
      geom_bar(stat = "count", position = position_dodge()) +
      ggtitle("Pet Ownership Among Groups")
layout = rbind(c())
grid.arrange(p1, p2, p3, p4, nrow = 2)
```









Examining the plots, there does not appear to be any systematic imbalance between baseline characteristics that would suggest poor randomization. There appears to be more females and fish owners in the pasture group, but it is likely that this observation is due to random variation. To know for sure, we will need to perform some statistical testing.

TODO: Statistical test for covariate balance.

```r
summary(df)
```

```
##    mturkcode              age             sex     
##  Min.   :2.483e+07   Min.   :20.00   Female:176  
##  1st Qu.:2.003e+09   1st Qu.:29.00   Male  :219  
##  Median :4.667e+09   Median :35.00               
##  Mean   :4.761e+09   Mean   :39.09               
##  3rd Qu.:7.383e+09   3rd Qu.:48.00               
##  Max.   :9.940e+09   Max.   :86.00               
##                       living_status    has_dog          has_cat      
##  A city                      :139   Min.   :0.0000   Min.   :0.0000  
##  A farm                      :  3   1st Qu.:0.0000   1st Qu.:0.0000  
##  A rural area                : 64   Median :1.0000   Median :0.0000  
##  A small town or suburban area:189   Mean   :0.5418   Mean   :0.4557  
##                                     3rd Qu.:1.0000   3rd Qu.:1.0000  
##                                     Max.   :1.0000   Max.   :1.0000  
##     has_bird           has_fish        veg_last_week     fruit_last_week 
##  Min.   :0.00000   Min.   :0.0000   Min.   : 0.000   Min.   : 0.000  
##  1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.: 4.000   1st Qu.: 3.000  
##  Median :0.00000   Median :0.0000   Median : 7.000   Median : 5.000  
```

```
##  Mean   :0.05823   Mean   :0.1114   Mean   :  7.268   Mean   :  5.684
##  3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.:  7.000   3rd Qu.:  7.000
##  Max.   :1.00000   Max.   :1.0000   Max.   :100.000   Max.   :100.000
##  dairy_last_week  eggs_last_week   beef_last_week   pork_last_week
##  Min.   : 0.000   Min.   : 0.000   Min.   : 0.000   Min.   : 0.000
##  1st Qu.: 3.000   1st Qu.: 1.000   1st Qu.: 1.000   1st Qu.: 0.000
##  Median : 5.000   Median : 3.000   Median : 2.000   Median : 1.000
##  Mean   : 5.737   Mean   : 3.071   Mean   : 2.734   Mean   : 1.132
##  3rd Qu.: 7.000   3rd Qu.: 5.000   3rd Qu.: 4.000   3rd Qu.: 2.000
##  Max.   :35.000   Max.   :14.000   Max.   :18.000   Max.   :10.000
##  hamburger_rank   chicken_rank     eggs_rank       grain_rank
##  Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000
##  1st Qu.:1.000   1st Qu.:1.000   1st Qu.:3.000   1st Qu.:2.500
##  Median :2.000   Median :2.000   Median :4.000   Median :4.000
##  Mean   :2.716   Mean   :2.311   Mean   :3.549   Mean   :3.587
##  3rd Qu.:4.000   3rd Qu.:3.000   3rd Qu.:5.000   3rd Qu.:5.000
##  Max.   :5.000   Max.   :5.000   Max.   :5.000   Max.   :5.000
##  fruit_veg_rank  veg_next_week    fruit_next_week   dairy_next_week
##  Min.   :1.000   Min.   :  0.00   Min.   :  0.000   Min.   : 0.000
##  1st Qu.:2.000   1st Qu.:  5.00   1st Qu.:  3.000   1st Qu.: 3.000
##  Median :3.000   Median :  7.00   Median :  5.000   Median : 5.000
##  Mean   :2.835   Mean   :  7.43   Mean   :  6.094   Mean   : 5.537
##  3rd Qu.:4.000   3rd Qu.:  7.00   3rd Qu.:  7.000   3rd Qu.: 7.000
##  Max.   :5.000   Max.   :100.00   Max.   :100.000   Max.   :35.000
##  eggs_next_week   beef_next_week  pork_next_week   video_type
##  Min.   : 0.000   Min.   : 0.00   Min.   : 0.000    :  0
##  1st Qu.: 2.000   1st Qu.: 1.00   1st Qu.: 0.000   F:130
##  Median : 3.000   Median : 2.00   Median : 1.000   I:130
##  Mean   : 3.256   Mean   : 2.77   Mean   : 1.309   P:135
##  3rd Qu.: 5.000   3rd Qu.: 4.00   3rd Qu.: 2.000
##  Max.   :14.000   Max.   :20.00   Max.   :12.000
##  attention_check    num_pets
##       :  0       Min.   :0.000
##  false:  0       1st Qu.:1.000
##  true :395       Median :1.000
##                  Mean   :1.167
##                  3rd Qu.:2.000
##                  Max.   :4.000
```

```r
# remember, video_type:  "F" => feedlot, "P" => Pasture, "I" => Irrigation
# Create a new column "vegetarian" for those who never eat meat before treatment
#mean(df$pre_beef)
#mean(df$post_beef)
#mean(df$post_beef[df$sex=="Male"])
#mean(df$post_beef[df$sex=="Female"])
#mean(df$post_beef[df$sex=="Female" & df$video_type=="F"])
#mean(df$post_beef[df$sex=="Female" & df$video_type=="P"])
#mean(df$post_beef[df$sex=="Female" & df$video_type=="I"])


# try a simple regression; set male and Irrigation video as reference levels for those factors
df$sex <-relevel(df$sex, ref = "Male")
df$video_type <-relevel(df$video_type, ref = "I")
df$vegetarian <- (df$beef_last_week == 0) & (df$pork_last_week == 0)
```

```r
model1 = lm( beef_next_week ~ beef_last_week*vegetarian + factor(sex) + factor(video_type) + factor(sex
#summary(model1)
```

Nicer output courtesy of stargazer

```r
stargazer(model1, type="latex", header=FALSE, no.space=TRUE)
```

Table 1:

	Dependent variable:
	beef_next_week
beef_last_week	0.878***
	(0.032)
vegetarian	−0.008
	(0.225)
factor(sex)Female	−0.162
	(0.239)
factor(video_type)F	−0.027
	(0.215)
factor(video_type)P	0.447**
	(0.224)
beef_last_week:vegetarian	
factor(sex)Female:factor(video_type)F	−0.245
	(0.337)
factor(sex)Female:factor(video_type)P	−0.530
	(0.331)
Constant	0.426**
	(0.183)
Observations	395
R^2	0.723
Adjusted R^2	0.718
Residual Std. Error	1.332 (df = 387)
F Statistic	144.520*** (df = 7; 387)
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

```r
# Now try it with standarized beef scores

df$standardized_beef_last_week <- scale(df$beef_last_week)
df$standardized_beef_next_week <- scale(df$beef_next_week)
model2 = lm( standardized_beef_next_week ~ standardized_beef_last_week*vegetarian + factor(sex) + facto


#summary(model2)
```

```r
stargazer(model2, type="latex", header=FALSE, no.space=TRUE)
```

```r
# Now try it with hamburger rank (1 = most desired, 5 = least desired)
# field is q16_1_rank

##df$standardized_pre_beef <- scale(df$pre_beef)
##df$standardized_post_beef <- scale(df$post_beef)
```

Table 2:

	Dependent variable:
	standardized_beef_next_week
standardized_beef_last_week	0.833***
	(0.030)
vegetarian	−0.003
	(0.090)
factor(sex)Female	−0.065
	(0.095)
factor(video_type)F	−0.011
	(0.086)
factor(video_type)P	0.178**
	(0.089)
standardized_beef_last_week:vegetarian	
factor(sex)Female:factor(video_type)F	−0.098
	(0.134)
factor(sex)Female:factor(video_type)P	−0.211
	(0.132)
Constant	0.023
	(0.061)
Observations	395
R^2	0.723
Adjusted R^2	0.718
Residual Std. Error	0.531 (df = 387)
F Statistic	144.520*** (df = 7; 387)
Note:	*p<0.1; **p<0.05; ***p<0.01

```
##model3 = lm( q16_1_rank ~ standardized_pre_beef + factor(video_type), data=df)
#summary(model3)
```

```
##stargazer(model3, type="latex", header=FALSE, no.space=FALSE)
```

5. Conclusions and directions for further investigations

6. Appendix: Notes on mehods

6.1 Qualtrics

6.2 Amazon Mechanical Turk

6.3 Production of Treatment and Control Videos

Field trips Editing and rendering Hosting

6.4 Support Scripts

Pulling results from qualtrics Paying subjects Automated test/validation generation