AHS Census Analysis for Ollie Pets Application

Eli Groves

3/29/2020

Introduction

To accompany my application to the Data Analyst position at Ollie Pets, I thought it would be fun to include a short and simple analysis of data that pertains to dogs. Perhaps unsurprising to you all, dog/pet ownership data wasn't particularly easy to find free online! After some research, I learned that the American Housing Survey from the Census Bureau has a field labeled `DPEVACPETS` which corresponds to the prompt "Assistance needed Evacuating or Sheltering Pets". Possible reponses were as follows:

"1" Indicates "Need help evacuating or sheltering pets". "2" Indicates "Do not need help with pets". "3" Indicates "No pets". "-6" and "-9" Indicate "Not reported".

We'll begin by importing in the 2017 American Housing Survey from the Census Bureau.

```
census_r <- fread("ahs2017n.csv", keepLeadingZeros = T, stringsAsFactors = F)
```

Let's take a look at the `DPEVACPETS` field now

```
pets <- census_r %>% select(DPEVACPETS)
prop.table(table(pets$DPEVACPETS))
```

```
##
##       '-6'       '-9'        '1'        '2'        '3'
## 0.56450743 0.01728787 0.04501738 0.15072208 0.22246524
```

At quick glance, it appears this field was not reported in roughly 57% of cases. Additionally, about 5% of respondents need help evacuating pets, 15% do not need help (but have pets) and 22% have no pets.

We will filter out the cases where this question was left unreported, because they provide no insight on our primary variable. We will also recode the variable with a binary interpretation because we are interested in whether a respondent has a pet or does not have a pet, not so much whether they need help with evacuation or not.

```
census_c <- census_r %>% filter(DPEVACPETS %in% c("'1'", "'2'", "'3'"))

census_c1 <- census_c %>% mutate(DPEVACPETS_bin = case_when(DPEVACPETS == "'1'" ~ 1,
                                                            DPEVACPETS == "'2'" ~ 1,
                                                            DPEVACPETS == "'3'" ~ 0),
                                 row_tot = n())
```

Analysis

To start, we'll look at the relationship between number of people in a household and pet ownership.

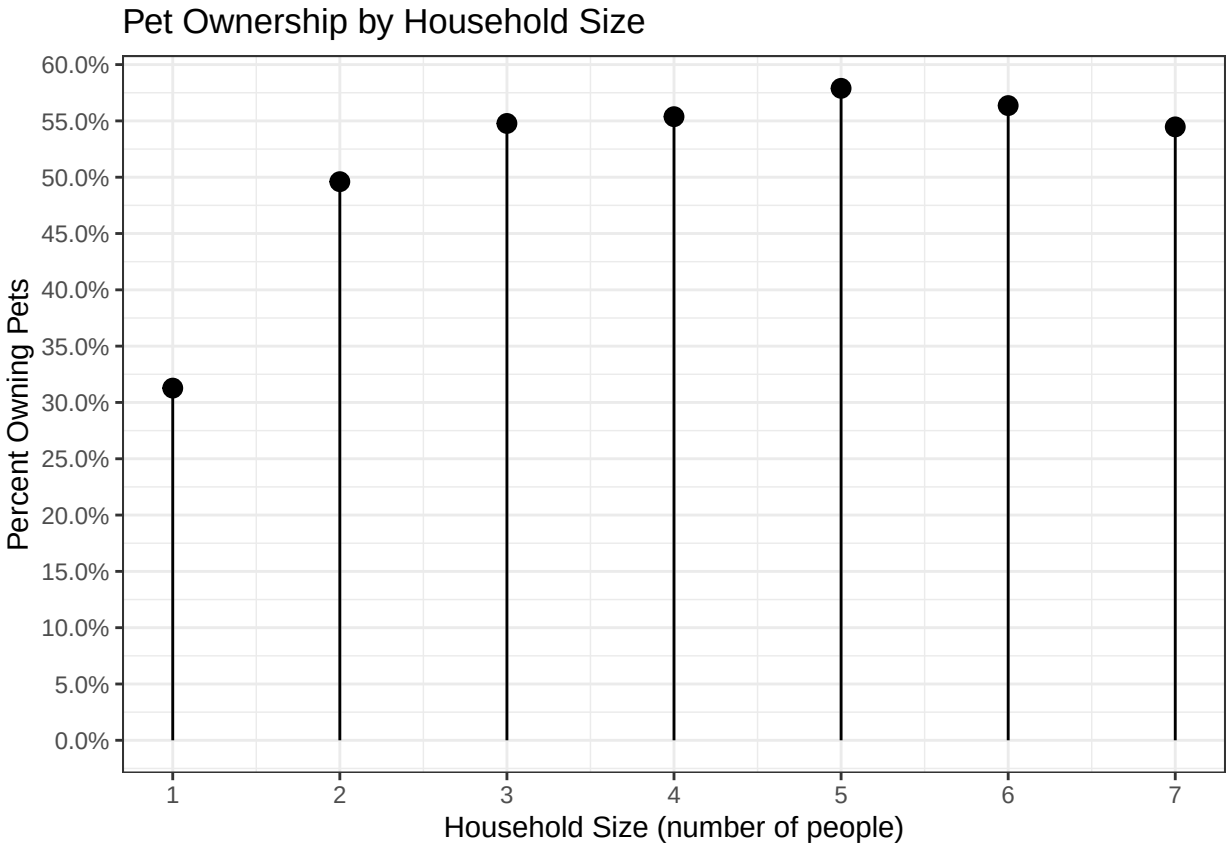
```
# Get the frequency of each NUMPEOPLE value to divide by later
num_people_freq <- census_c1 %>% group_by(NUMPEOPLE) %>% summarise(freq = n())

# Get the number of pets owned for each NUMPEOPLE value
num_people <- census_c1 %>% group_by(NUMPEOPLE) %>% summarise(pet_count = sum(DPEVACPETS_bin)) %>%
  left_join(num_people_freq, by = c("NUMPEOPLE")) %>% mutate(pet_ownership_perc = pet_count/freq)

# We'll filter out instances where n < 100 and arrange in descending order
num_people2 <- num_people %>% filter(freq > 100) %>% arrange(desc(pet_ownership_perc)) %>%
  mutate(pet_ownership_perc_f = paste0(round(pet_ownership_perc*100, 1), "%"))

num_people2 %>% kable() %>% kable_styling()
```

NUMPEOPLE	pet_count	freq	pet_ownership_perc	pet_ownership_perc_f
5	1005	1736	0.5789171	57.9%
6	381	676	0.5636095	56.4%
4	2091	3776	0.5537606	55.4%
3	2398	4378	0.5477387	54.8%
7	128	235	0.5446809	54.5%
2	4528	9128	0.4960561	49.6%
1	2448	7828	0.3127236	31.3%

To interpret this verbally, we can say that households with 5 people are the most likely to have a pet, at 58%. Households with 6 people follow, with 56% of them owning a pet. Households with 1 individual are far less likely to own a pet (31.3%).

We can visualize this with a simple lollipop chart.

```
theme_set(theme_bw())

# Plot
num_people_viz <- ggplot(num_people2, aes(x=NUMPEOPLE, y=pet_ownership_perc)) +
                          geom_point(size=3) +
                          geom_segment(aes(x=NUMPEOPLE,
                              xend=NUMPEOPLE,
                              y=0,
                              yend=pet_ownership_perc)) +
                labs(title="Pet Ownership by Household Size") +
  xlab("Household Size (number of people)") +
  ylab("Percent Owning Pets") +
  scale_y_continuous(breaks = seq(0,0.6,0.05), labels = scales::percent) +
  scale_x_continuous(breaks = seq(1,7,1))

num_people_viz
```



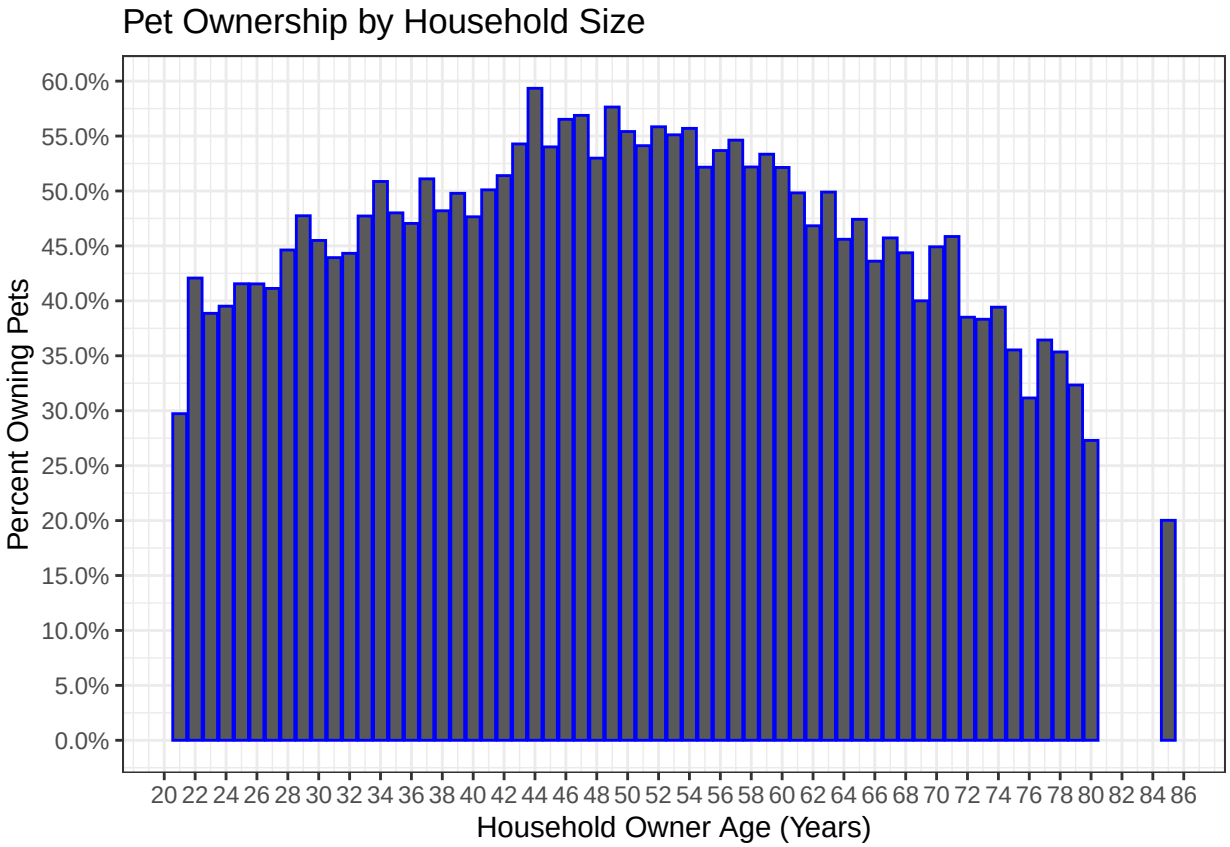
Although simple, this visualization gets the point across: households with 3-7 individuals are more likely to own a pet than households with 1-2 individuals, and ownership peaks in households of 5 individuals.

Another variable that might be interesting to explore is the age of the householder. What age householder is most likely to own a pet?

```
hh_age_freq <- census_c1 %>% group_by(HHAGE) %>% summarise(freq = n())

hh_age <- census_c1 %>% group_by(HHAGE) %>% summarise(pet_count = sum(DPEVACPETS_bin)) %>%
  left_join(hh_age_freq, by = c("HHAGE")) %>% mutate(pet_ownership_perc = pet_count/freq)

hh_age2 <- hh_age %>% filter(freq > 100) %>% arrange(desc(pet_ownership_perc)) %>%
  mutate(pet_ownership_perc_f = paste0(round(pet_ownership_perc*100, 1), "%"))
```

To plot this, we'll use a bar chart with the age of the house owner on the x-axis and the percent of owners at that age who own a pet on the y-axis.

```
# Plot
hh_age_viz <- ggplot(hh_age2, aes(x=HHAGE, y=pet_ownership_perc)) +
  geom_bar(stat = "identity", color = "blue") +
  labs(title="Pet Ownership by Household Size") +
  xlab("Household Owner Age (Years)") +
  ylab("Percent Owning Pets") +
  scale_x_continuous(breaks = seq(20,86,2)) +
  scale_y_continuous(breaks = seq(0,0.6, 0.05), labels = scales::percent)

hh_age_viz
```



Pet ownership appears to peak with householders who are 44, with 59.3% of them owning a pet. By examining the shape of the figure, we can see that pet ownership is highest with householders roughly 43-60 years old, with ownership beginning to decline with householders younger and older than that range.

Another interesting question to explore might be whether a male or female household owner is more likely to own a pet. In the HHSEX variable, a 1 corresponds to male, and a 2 corresponds to female.

```
hh_sex_freq <- census_c1 %>% group_by(HHSEX) %>% summarise(freq = n())

hh_sex <- census_c1 %>% group_by(HHSEX) %>% summarise(pet_count = sum(DPEVACPETS_bin)) %>%
  left_join(hh_sex_freq, by = c("HHSEX")) %>% mutate(pet_ownership_perc = pet_count/freq)

hh_sex2 <- hh_sex %>% filter(freq > 100) %>% arrange(desc(pet_ownership_perc)) %>%
  mutate(pet_ownership_perc_f = paste0(round(pet_ownership_perc*100, 1), "%"))

hh_sex2 %>% kable() %>% kable_styling()
```

HHSEX	pet_count	freq	pet_ownership_perc	pet_ownership_perc_f
'2'	6559	13811	0.4749113	47.5%
'1'	6507	14105	0.4613258	46.1%

It appears that females, by a small margin (1.4 percentage points), are more likely to own a pet than males.

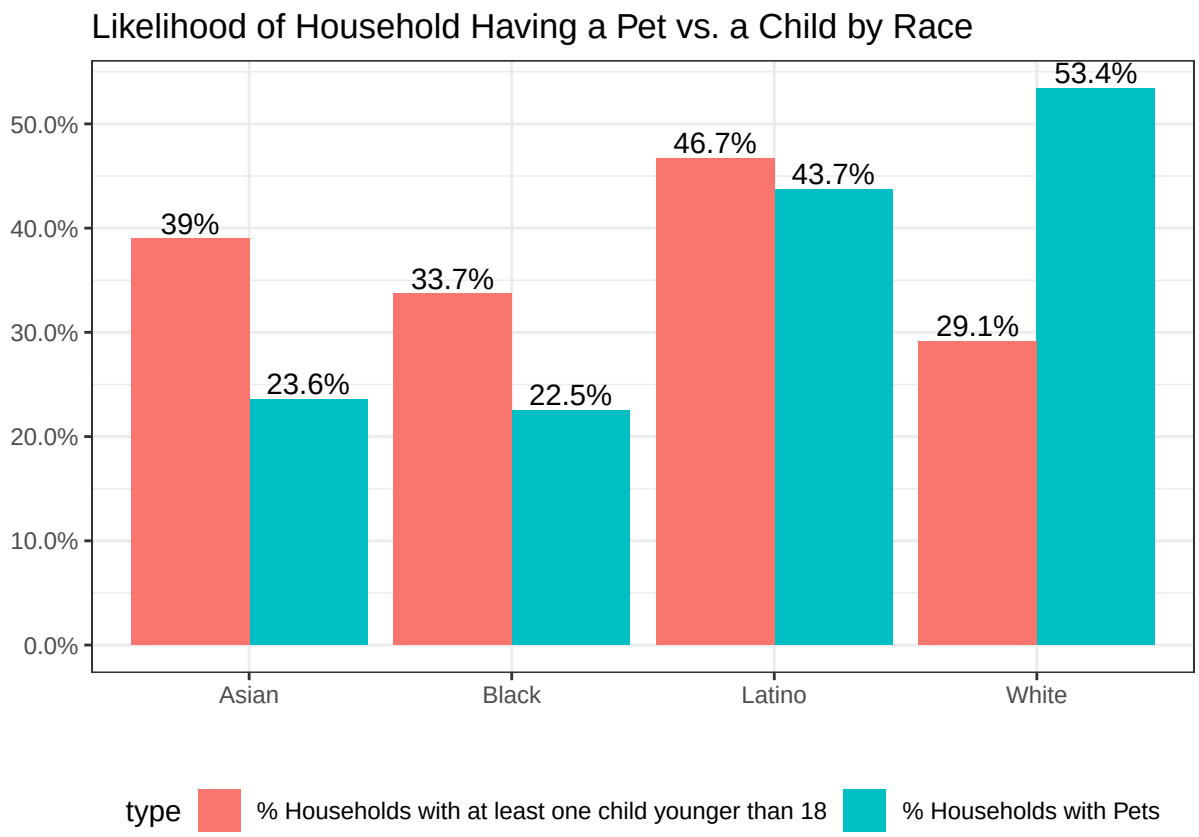
Finally, we'll compare by race, the likelihood of owning a pet versus likelihood of having children under 18 years of age.

```
census_c_race <- census_c1 %>% mutate(white_bin = ifelse(HHRACE == "'01'", 1, 0),
                                      black_bin = ifelse(HHRACE == "'02'", 1, 0),
                                      asian_bin = ifelse(HHRACE == "'04'", 1, 0),
                                      latino_bin = ifelse(HHSPAN == "'1'", 1, 0),
                                      children_bin = ifelse(NUMYNGKIDS != 0, 1,
                                                            ifelse(NUMOLDKIDS != 0, 1, 0)))

white <- census_c_race %>% filter(HHRACE == "'01'") %>%
  mutate(sum_pets = sum(DPEVACPETS_bin), sum_children = sum(children_bin), n_row = n(),
         `% Households with Pets` = sum_pets/n_row,
         `% Households with at least one child younger than 18` =
           sum_children/n_row, race = "White") %>%
  select(`% Households with Pets`, `% Households with at least one child younger than 18`, race) %>%
  unique()

black <- census_c_race %>% filter(HHRACE == "'02'") %>%
  mutate(sum_pets = sum(DPEVACPETS_bin), sum_children = sum(children_bin), n_row = n(),
         `% Households with Pets` = sum_pets/n_row,
         `% Households with at least one child younger than 18` =
           sum_children/n_row, race = "Black") %>%
  select(`% Households with Pets`, `% Households with at least one child younger than 18`, race) %>%
  unique()

asian <- census_c_race %>% filter(HHRACE == "'04'") %>%
  mutate(sum_pets = sum(DPEVACPETS_bin), sum_children = sum(children_bin), n_row = n(),
         `% Households with Pets` = sum_pets/n_row,
         `% Households with at least one child younger than 18` =
           sum_children/n_row, race = "Asian") %>%
  select(`% Households with Pets`, `% Households with at least one child younger than 18`, race) %>%
  unique()

latino <- census_c_race %>% filter(HHSPAN == "'1'") %>%
  mutate(sum_pets = sum(DPEVACPETS_bin), sum_children = sum(children_bin), n_row = n(),
         `% Households with Pets` = sum_pets/n_row,
         `% Households with at least one child younger than 18` =
           sum_children/n_row, race = "Latino") %>%
  select(`% Households with Pets`, `% Households with at least one child younger than 18`, race) %>%
  unique()


race_pets <- bind_rows(white, black, asian, latino) %>%
  pivot_longer(cols = c(`% Households with Pets`,
                        `% Households with at least one child younger than 18`),
               names_to = "type", values_to = "value" ) %>%
  mutate(label = paste0(round(value*100,1), "%"))
```

We'll then plot the results in a clustered bar chart

```
race_pets_viz <- ggplot(race_pets, aes(race, value )) +
  geom_bar(aes(fill = type), stat = "identity", position = "dodge") +
  geom_text(aes(label = label, group = type), vjust = -0.25, position = position_dodge(width = 0.9)) +
  theme(legend.position = "bottom") +
  scale_y_continuous( breaks = seq(0, 0.7, 0.1), labels = scales::percent) +
```

```
  labs(title = "Likelihood of Household Having a Pet vs. a Child by Race", x ="", y ="" )

race_pets_viz
```



There are a number of takeaways from this figure. One is that White households are not only the most likely of all races to have pets, but also the only race more likely to own at least one pet than have a child younger than 18. Specifically, White households are 24.3 percentage points more likely to have a pet than a child younger than 18. This differs starkly with Asian households which are 15.4 percentage points more likely to have a child younger than 18 than a pet. Latino households present the least difference in likelihood of having a pet (43.7%) versus having a child (46.7%).

Conclusion

Though there have certainly been some interesting insights to come from this analysis, there are a number of shortcomings worth acknolwedging. For one, the `DPEVACPETS` field refers to the presence of pets generally and is not specific to dogs, which would be more desirable given the context. Additionally, the data is from 2017 and thus a bit outdated. The AHS survey is conducted on a biennial basis and unfortunately the results from 2019 haven't yet been publicized. Going forward, the results could be updated with the most recent data, and figures could be put into a dashboard to allow for increased interactivity and accessibility. I hope you've enjoyed this short and simple analysis of AHS Census data as it pertains to pet ownership! Thank you for reading.