

Final Project

Eric M

2024-12-05

```
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
library(plotly)
```

```
##
## Attaching package: 'plotly'

## The following object is masked from 'package:ggplot2':
##
##   last_plot

## The following object is masked from 'package:stats':
##
##   filter

## The following object is masked from 'package:graphics':
##
##   layout
```

```
library(readxl)
library(wordcloud)
```

```
## Loading required package: RColorBrewer
```

```
library(tm)
```

```
## Loading required package: NLP
```

```
##
```

```
## Attaching package: 'NLP'
```

```
## The following object is masked from 'package:ggplot2':
```

```
##
```

```
##      annotate
```

```
library(mdsr)
```

```
pepe_data=read_excel("C:/Users/magaw/OneDrive/Desktop/DATA 211/Pepe Data.xlsx")
```

Testing whether my dog, Penelope, barks significantly more at origins of noises or movement, based on whether they occur inside or our outside of my apartment.

Word Cloud

```
text_df2=as.data.frame(pepe_data$Reason)
vector_pepe=VectorSource(text_df2)
corp_pepe=Corpus(vector_pepe)
```

```
text_mining=corp_pepe%>%
  tm_map(content_transformer(tolower))%>%
  tm_map(removeNumbers)%>%
  tm_map(removePunctuation)%>%
  tm_map(stripWhitespace)%>%
  TermDocumentMatrix()%>%
  as.matrix()
```

```
## Warning in tm_map.SimpleCorpus(., content_transformer(tolower)): transformation
## drops documents
```

```
## Warning in tm_map.SimpleCorpus(., removeNumbers): transformation drops
## documents
```

```
## Warning in tm_map.SimpleCorpus(., removePunctuation): transformation drops
## documents
```

```
## Warning in tm_map.SimpleCorpus(., stripWhitespace): transformation drops
## documents
```

```
text_freq=sort(rowSums(text_mining),decreasing=TRUE)
text_freqDF=as.data.frame(text_freq)
text_freqDF
```

```
##          text_freq
## door          55
## opened        53
## roommate      30
## wife          23
## pancho        10
## eric          9
## neighbor      6
## let           5
## stood         5
## fed           4
## kennel        4
## out           4
## phone         3
## playing       3
## with          3
## answered      2
## attention     2
## car           2
## closed        2
## getting       2
## keys          2
## laughed       2
## room          2
## ate           1
## barked        1
## bed           1
## bedroom       1
## buzzer        1
## cupstairs     1
## dropped       1
## entered       1
## floor         1
## food          1
## front         1
## into          1
## jumped        1
## moved         1
## neighbors     1
## picked        1
## rang          1
## talking       1
## window        1
```

```
wordcloud(rownames(text_freqDF),text_freqDF$text_freq,colors=brewer.pal(11,'Dark2'))
```

```
## Warning in brewer.pal(11, "Dark2"): n too large, allowed maximum for palette Dark2 is 8
## Returning the palette you asked for with that many colors
```

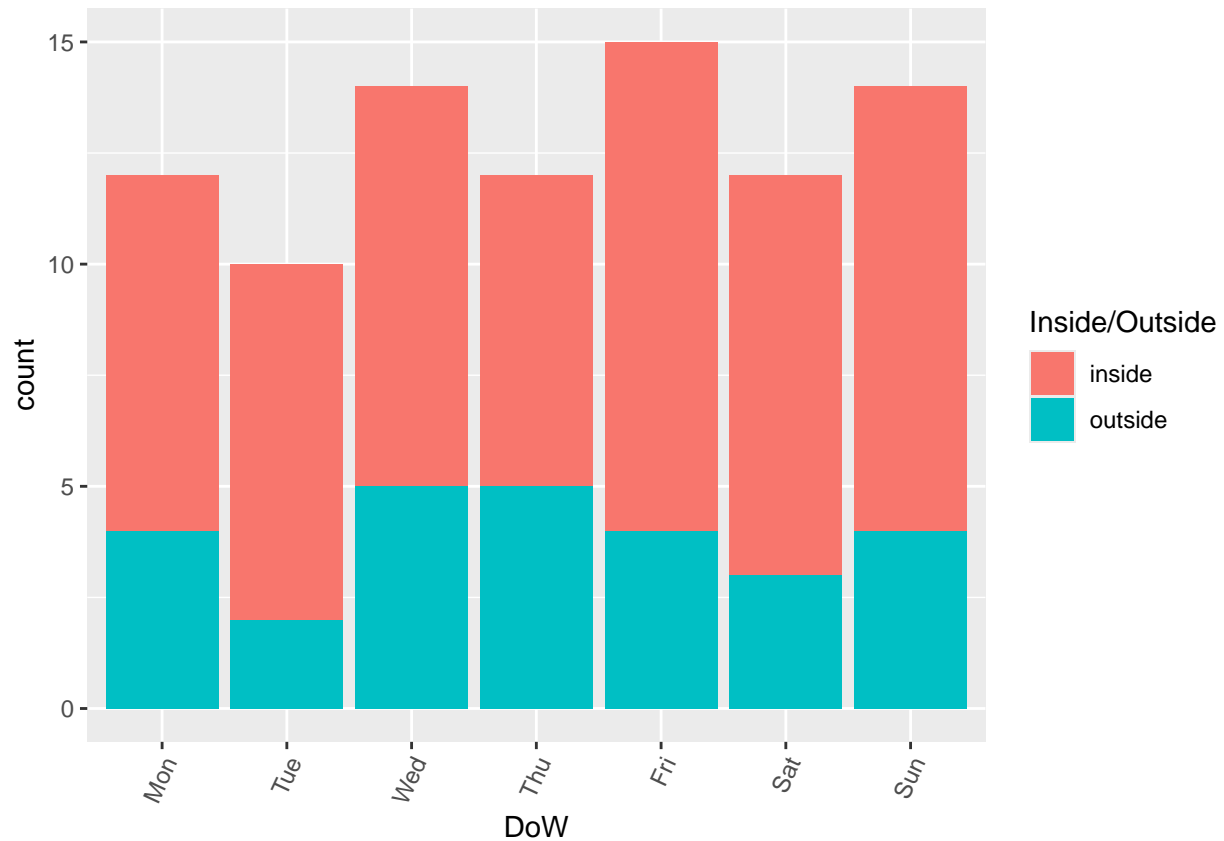
roommate
opened
wife pancho
fed out
let playing
with kenel
neighbor eric
phone stood
door

Bar Graph

```
positions<-c('Mon','Tue','Wed','Thu','Fri','Sat','Sun')

pepe_graph=pepe_data%>%
  ggplot(aes(fill=`Inside/Outside`,x=DoW))+
  geom_bar()+
  scale_x_discrete(labels=positions)+
  theme(axis.text.x=element_text(angle=65,hjust=1))

pepe_graph
```



Hypothesis Testing

$$H_0 : \mu_I = \mu_O \quad H_a : \mu_I \neq \mu_O$$

Where

μ_I = mean number of barks at inside sources μ_O = mean number of barks at outside sources

```
inside_count=c(3,5,5,7,5,4,6,4,3,4,3,6,4,3)
outside_count=c(2,1,2,3,1,2,1,3,1,0,1,3,1,1)

t.test(outside_count,inside_count,alternative="two.sided")
```

```
##
##  Welch Two Sample t-test
##
## data:  outside_count and inside_count
## t = -6.7244, df = 23.796, p-value = 6.173e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.734476 -1.979809
## sample estimates:
## mean of x mean of y
##  1.571429  4.428571
```

Conclusion: $p\text{-value}=6.29e-7 < 0.05 = \alpha$

Therefore we reject H_0 .

At 5% significance level, there is evidence to support the claim that the mean number of barks between inside and outside sources are significantly different.

##Final Thoughts During this research it was interesting to see the sources that cause Penelope to bark. Doors seem to be a large trigger, but only when they are opened and not closed. My wife and roommate being close in total numbers of barks was unexpected. My roommate is likely Penelope's favorite person, however he works overnight shifts so him opening the front door did not cause her to bark, but entering/exiting his room was a main contributor. My wife on the other hand caused more barks at the front door, which I presume is due to her work schedule having her return during time the dogs are not either in bed or in their kennels.

Also while inside sources had a much larger amount of barks, I believe a contributing factor was the time at which the research was conducted. Beginning at the end of November, about the time we began keeping our windows closed, may have skewed the data in one direction. My apartment is sub-ground as well, so being a chihuahua Penelope does not have easy access to seeing out the windows, causing her to react mostly to noises outside.