

# **DID YOU ACTUALLY SEE BIGFOOT?**

By: Julie Leung



# Business Proposition

Mr. David Attenborough is interested in featuring Bigfoot in his upcoming nature documentary. His camera crew would like to confidently know where to set up the cameras to capture a glimpse of this elusive creature.

# Defining the Metrics

Split into two classification:

Class A: clear sightings in circumstances where misinterpretation or misidentification of other animals can be ruled out with greater confidence.

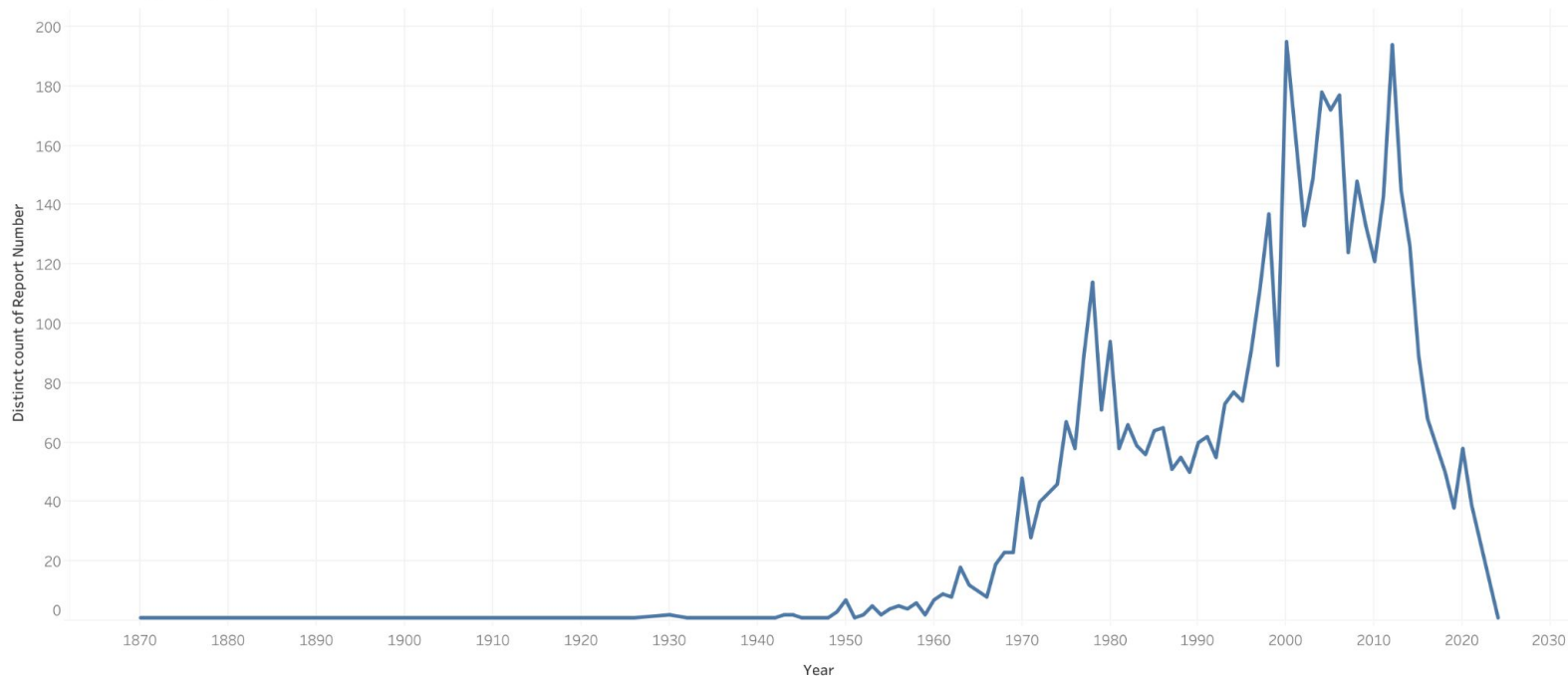
Class B: circumstances where a possible Sasquatch was observed but did not have a clear view of the subject. Any characteristic sounds are always considered in this class.



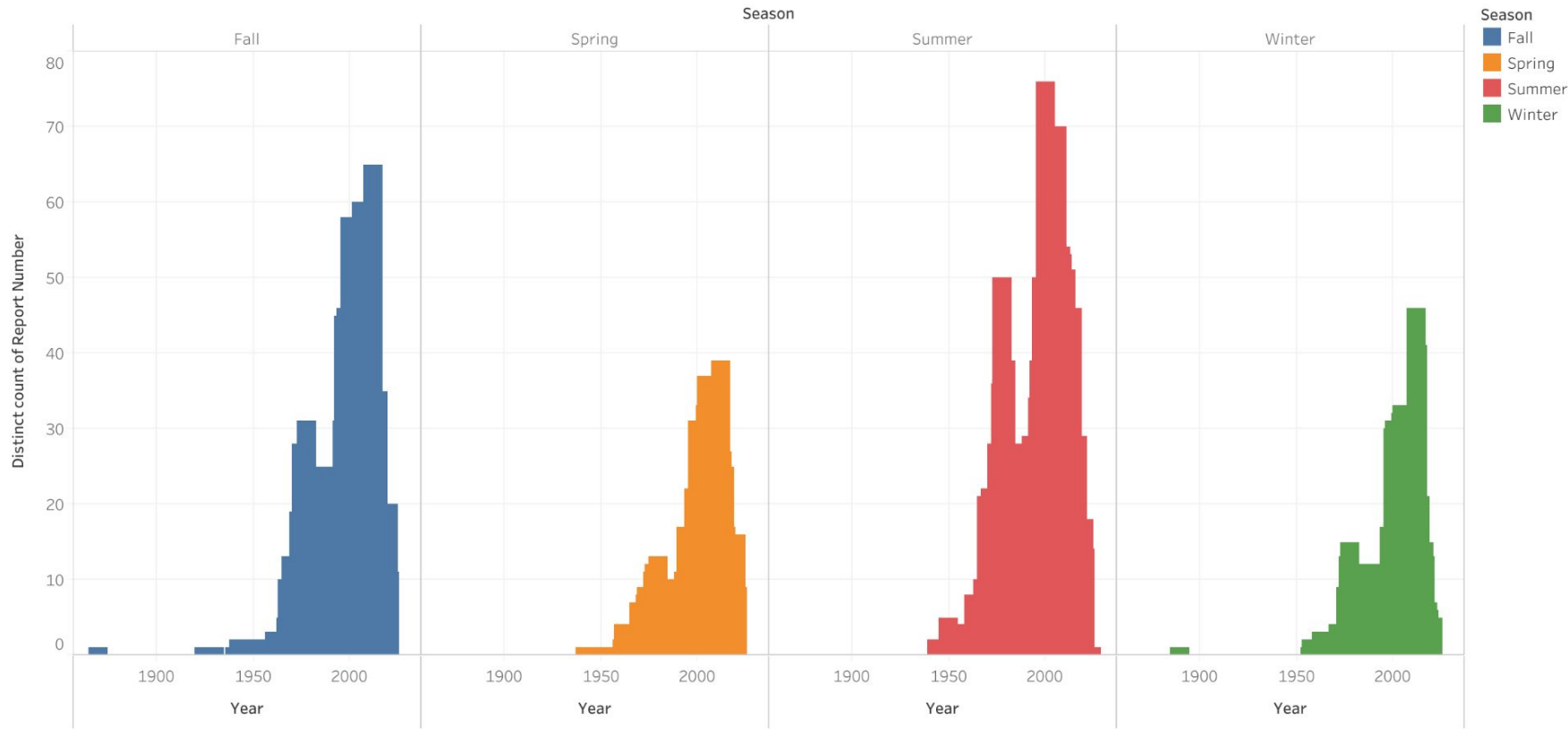
# Data Source

- Bigfoot Field Researchers Organization
- 4,983 reports, dating back to as early as the 1870's

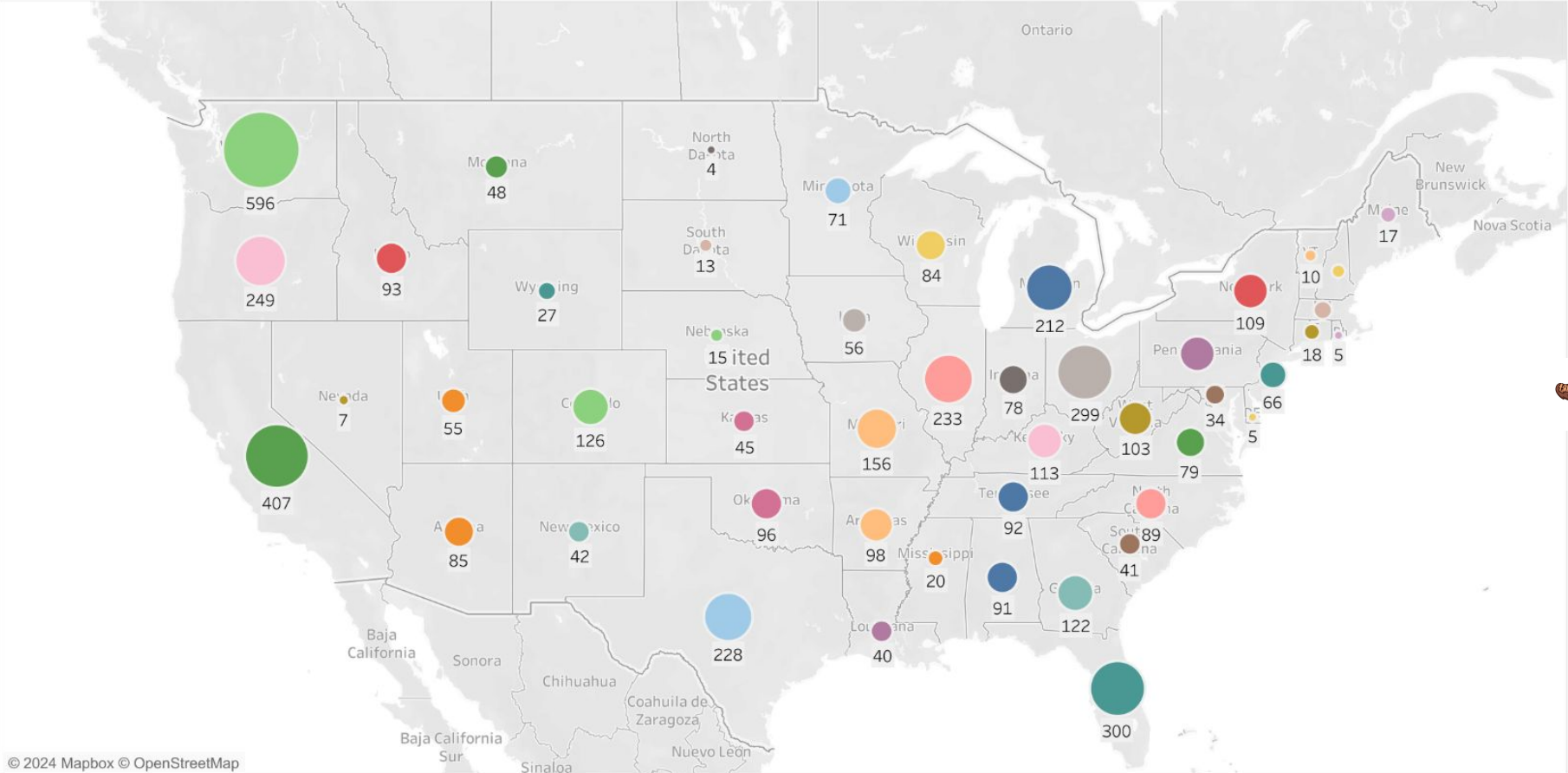
Number of Sightings over Time in USA



Number of Sightings per Season



Sightings across U.S.A.



# Models

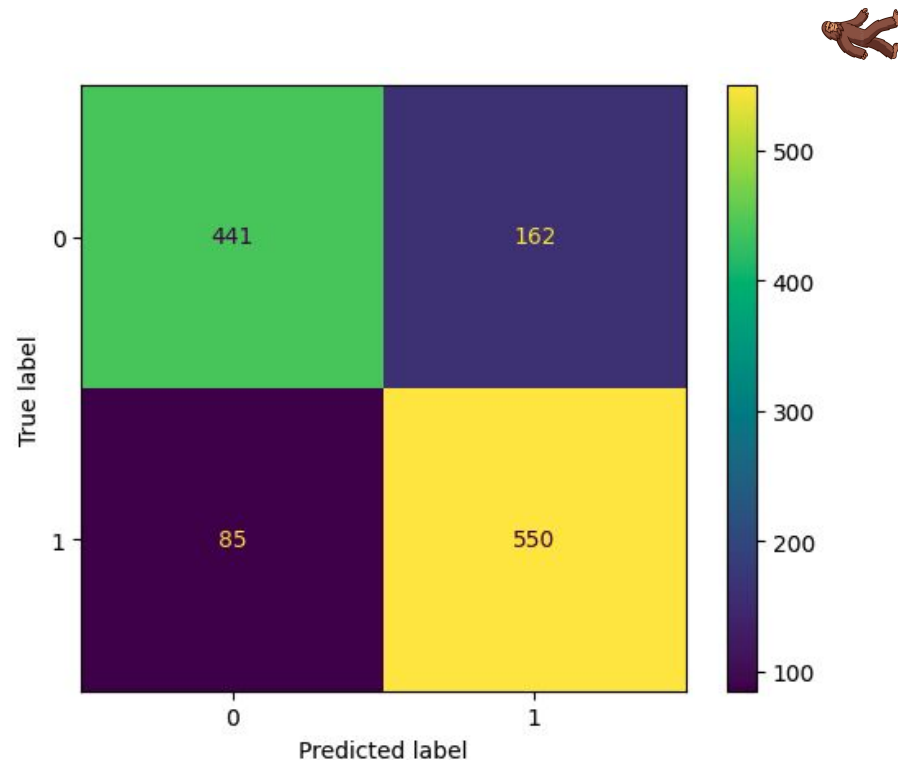
- Naive Bayes Classifier - Multinomial

- CountVectorizer: ~79.6%
- TfidfVectorizer: ~78.8%

- Tree-Based

- Decision Tree Classifier: ~72.4%
- Random Forest Classifier : ~78.1%
- Extra Trees Classifier: ~78.4%

- Logistic Regression: 80.0%



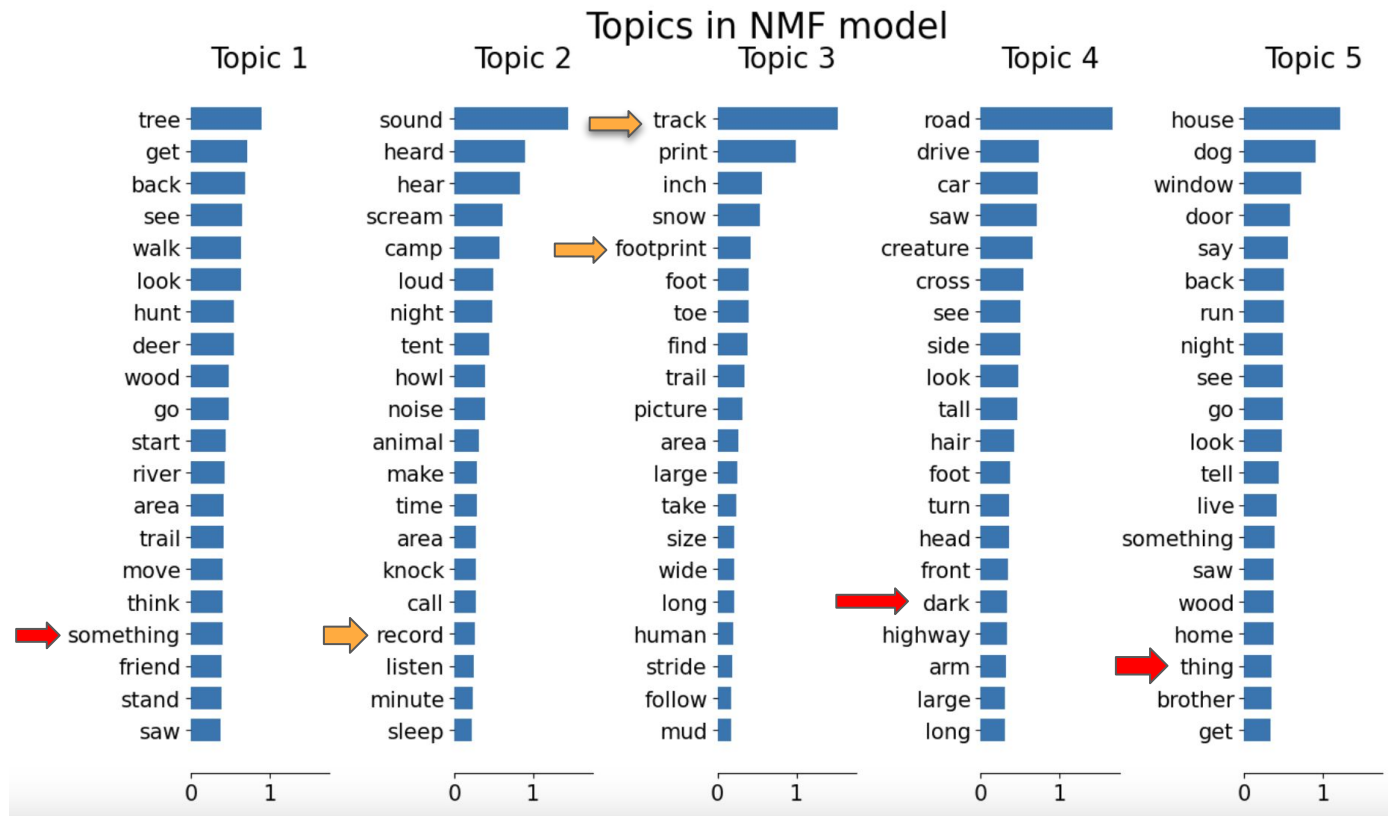
Logistic Regression Confusion Matrix



# Skeptic to Believer

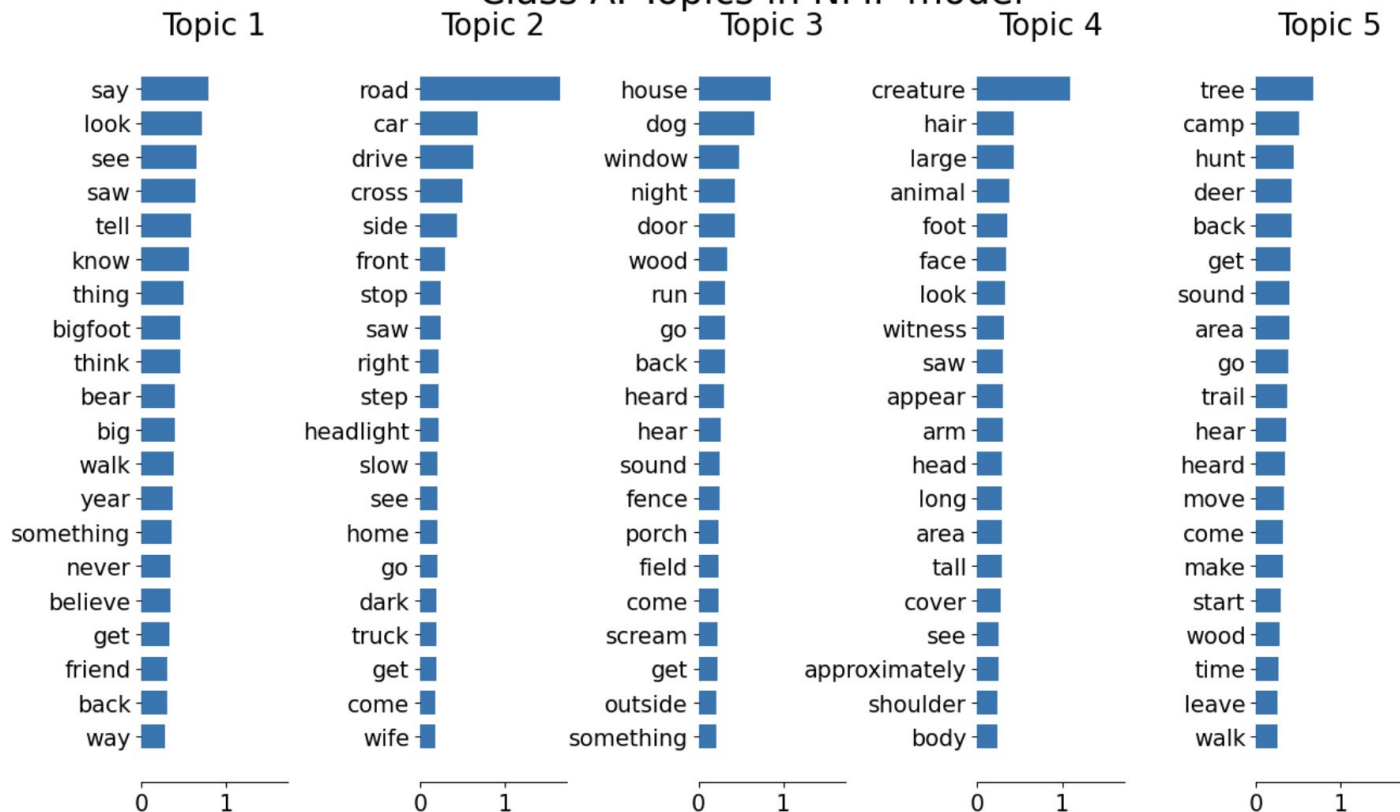
Topics are broken down into:

- Topic 1: Outdoors
  - Topic 2: Sounds
  - Topic 3: Prints
  - Topic 4: Roads
  - Topic 5: Indoors
- 
- Words of uncertainty = “something”, “thing”, or “dark”
  - As for “record” or “footprint”, they serve as evidence of the sightings.





## Class A: Topics in NMF model



Topic 1: ?

Topic 2:  
Roads

Topic 3:  
Indoors

Topic 4:  
Features

Topic 5:  
Outdoors





## Class B: Topics in NMF model

Topic 1:  
Outdoors

Topic 1

tree  
wood  
knock  
get  
hunt  
trail  
walk  
back  
area  
rock  
deer  
truck  
start  
hear  
come  
something  
see  
leave  
time  
creek

0 1

Topic 2:  
Roads

Topic 2

road  
saw  
see  
look  
drive  
back  
say  
house  
tall  
run  
go  
car  
figure  
foot  
walk  
creature  
get  
turn  
tell  
side

0 1

Topic 3:  
Prints

Topic 3

track  
print  
inch  
snow  
foot  
toe  
footprint  
picture  
trail  
find  
take  
long  
wide  
size  
human  
large  
stride  
area  
mud  
creek

0 1

Topic 4:  
Sounds

Topic 4

sound  
heard  
hear  
scream  
dog  
howl  
loud  
house  
night  
call  
record  
animal  
time  
coyote  
noise  
listen  
bigfoot  
make  
come  
never

0 1

Topic 5:  
Camping

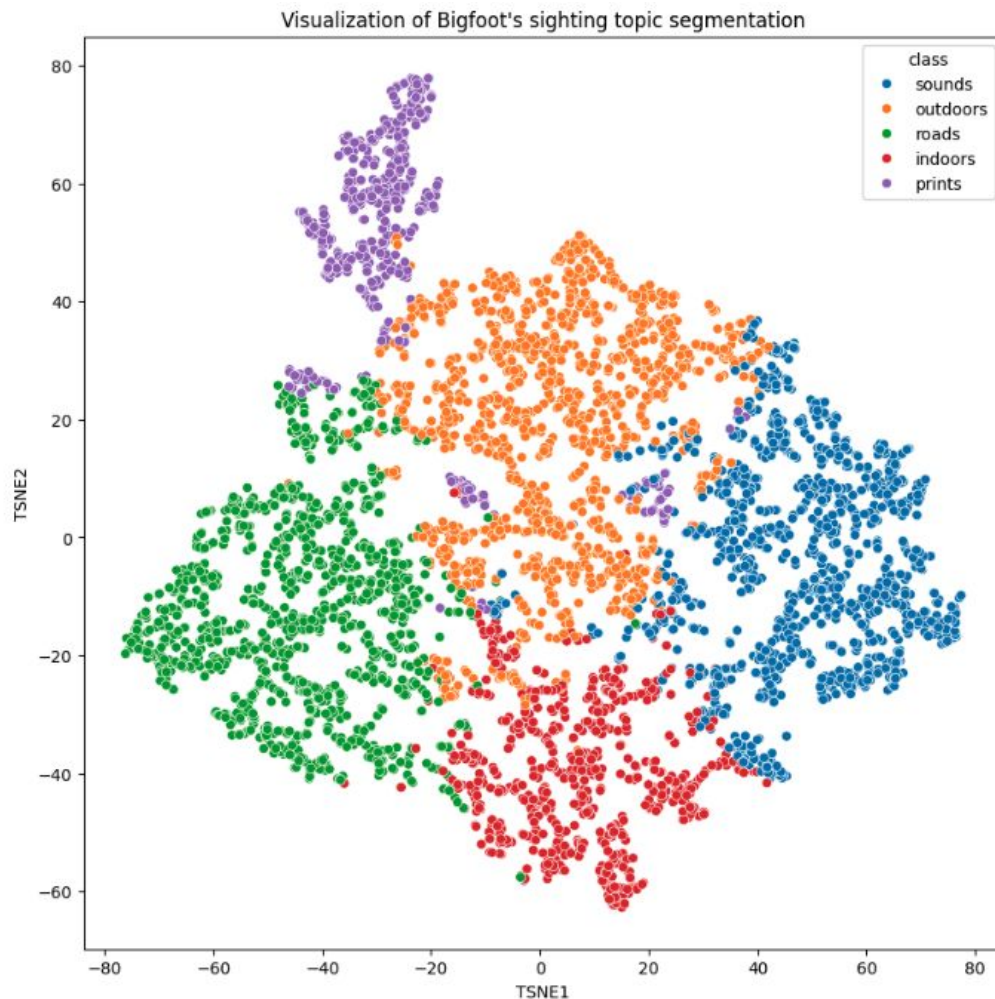
Topic 5

camp  
tent  
night  
sleep  
fire  
campfire  
go  
get  
campground  
morning  
friend  
heard  
site  
wake  
set  
time  
next  
bag  
awake  
day

0 1



## Visualizing the topics with TSNE





# Conclusion

With the highest probability of Class A given a sighting, I would recommend Mr. Attenborough's film crew to set up cameras near roads in:

- Arkansas: 68.3%
- Alabama: 63.7%
- Oklahoma: 61.4%
- Kentucky: 60.1%
- Pennsylvania: 59.2%

# Next Steps



- Find him
- Time series analysis on the topics

 THANK YOU!