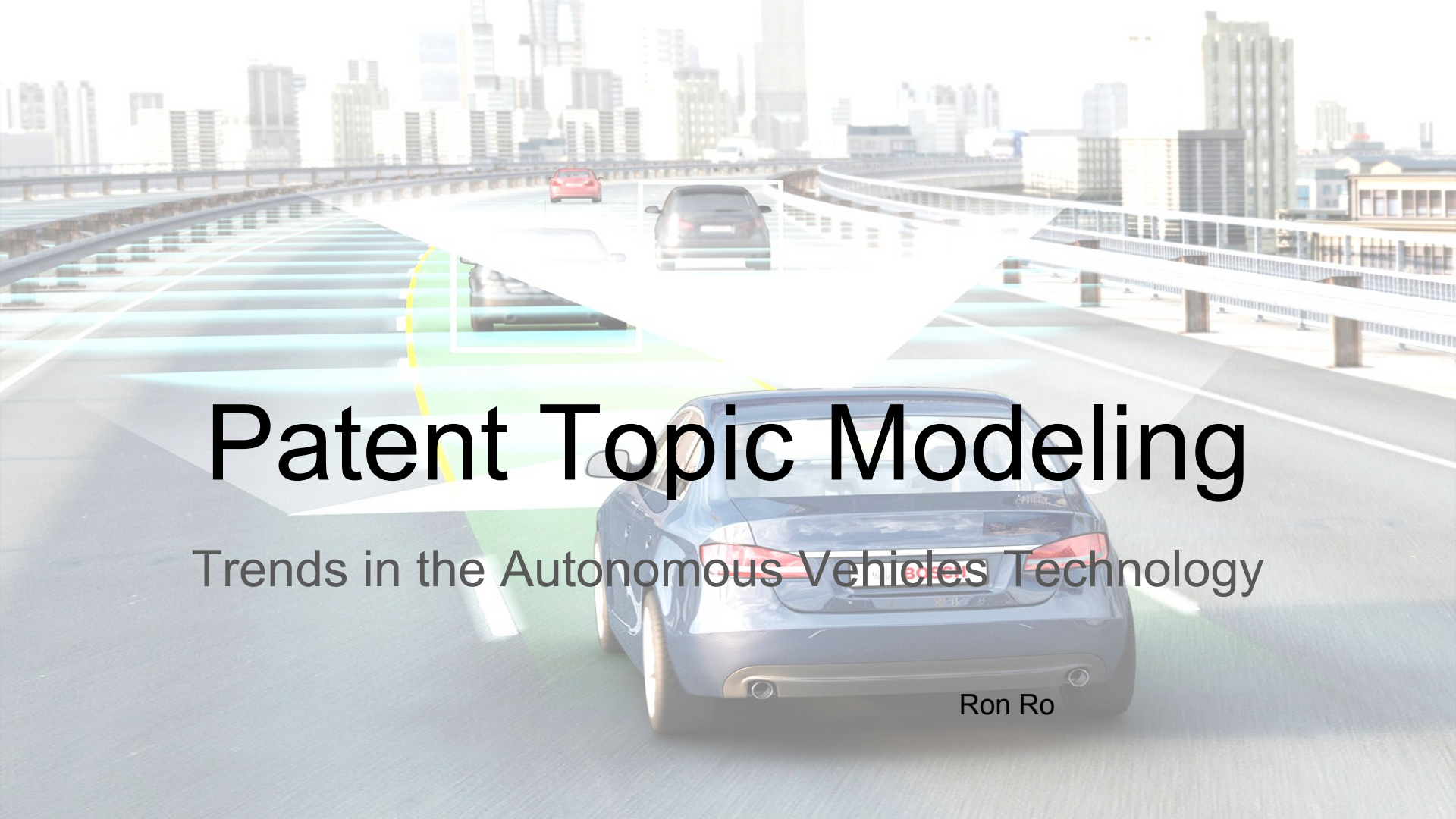
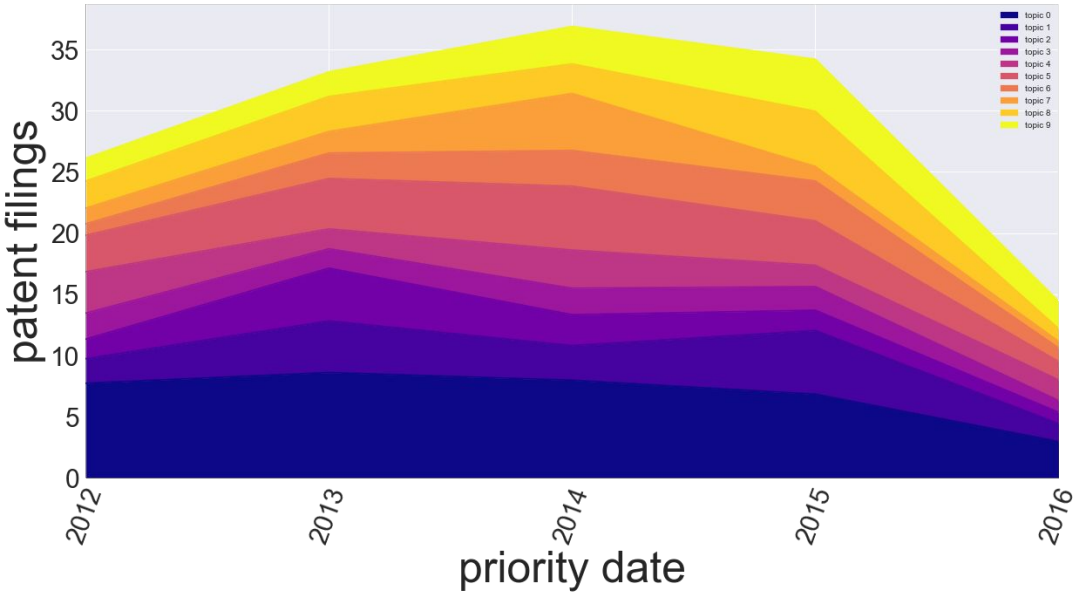
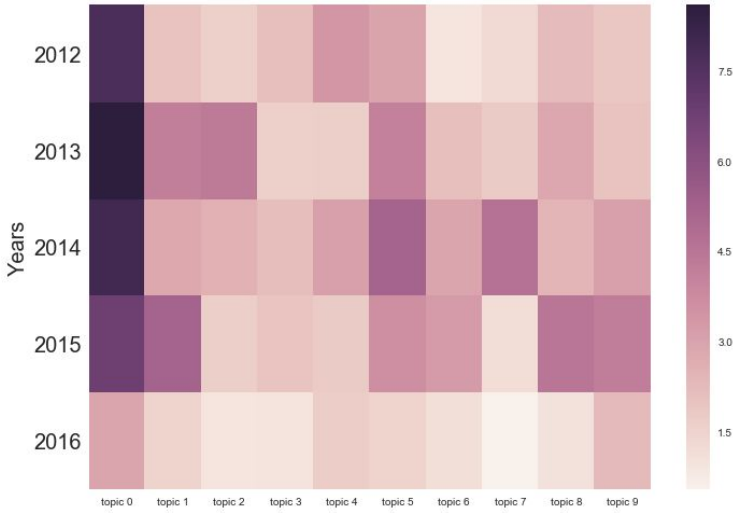


Patent Topic Modeling

Trends in the Autonomous Vehicles Technology

Ron Ro

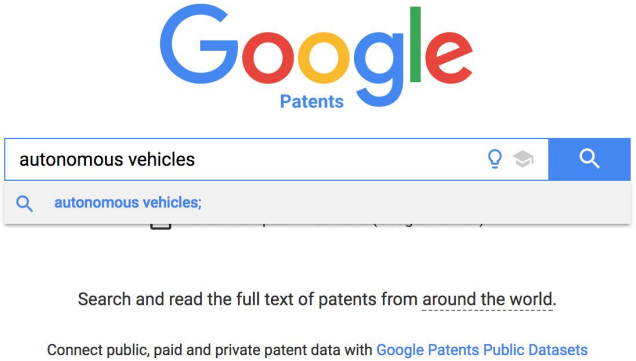

topic 0
topic 1
topic 2
topic 3
topic 4
topic 5
topic 6
topic 7
topic 8
topic 9
patent filings
0
5
10
15
20
25
30
35
2012
2013
2014
2015
2016
priority date

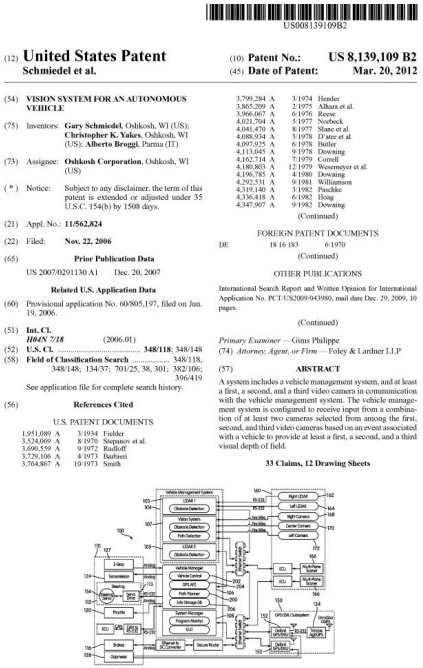

Years
2012
2013
2014
2015
2016
topic 0
topic 1
topic 2
topic 3
topic 4
topic 5
topic 6
topic 7
topic 8
topic 9
7.5
6.0
4.5
3.0
1.5


0.4
0.2
0.0
−0.2
−0.4
−0.4
−0.2
0.0
0.2
0.4
0.6

Problem statement

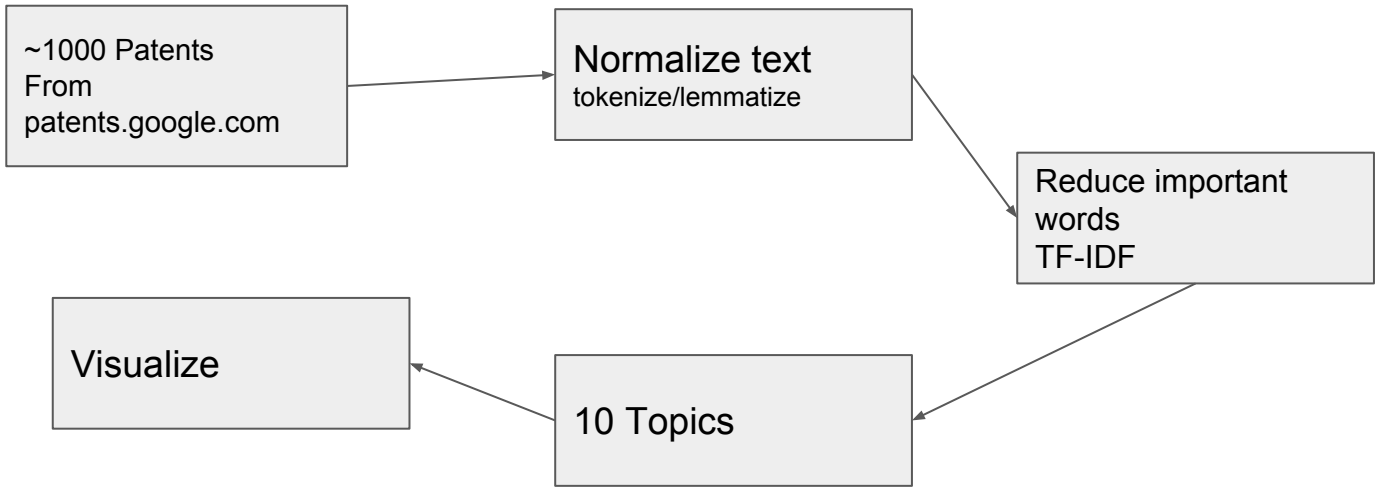
- VCs and companies want to know where to place their next bets
- Patents can be informative





Objective

- Obtain text data from patents.google.com
- Observe trends within the field using Natural Language Processing



Success Metrics

- Did I get actual topics, i.e. sub-classes of technologies within autonomous vehicles?
- Subject Matter Expertise
- Different models produce different results
  - LDA (Latent Dirichlet Allocation)
  - NMF(Non-negative Matrix Factorization)

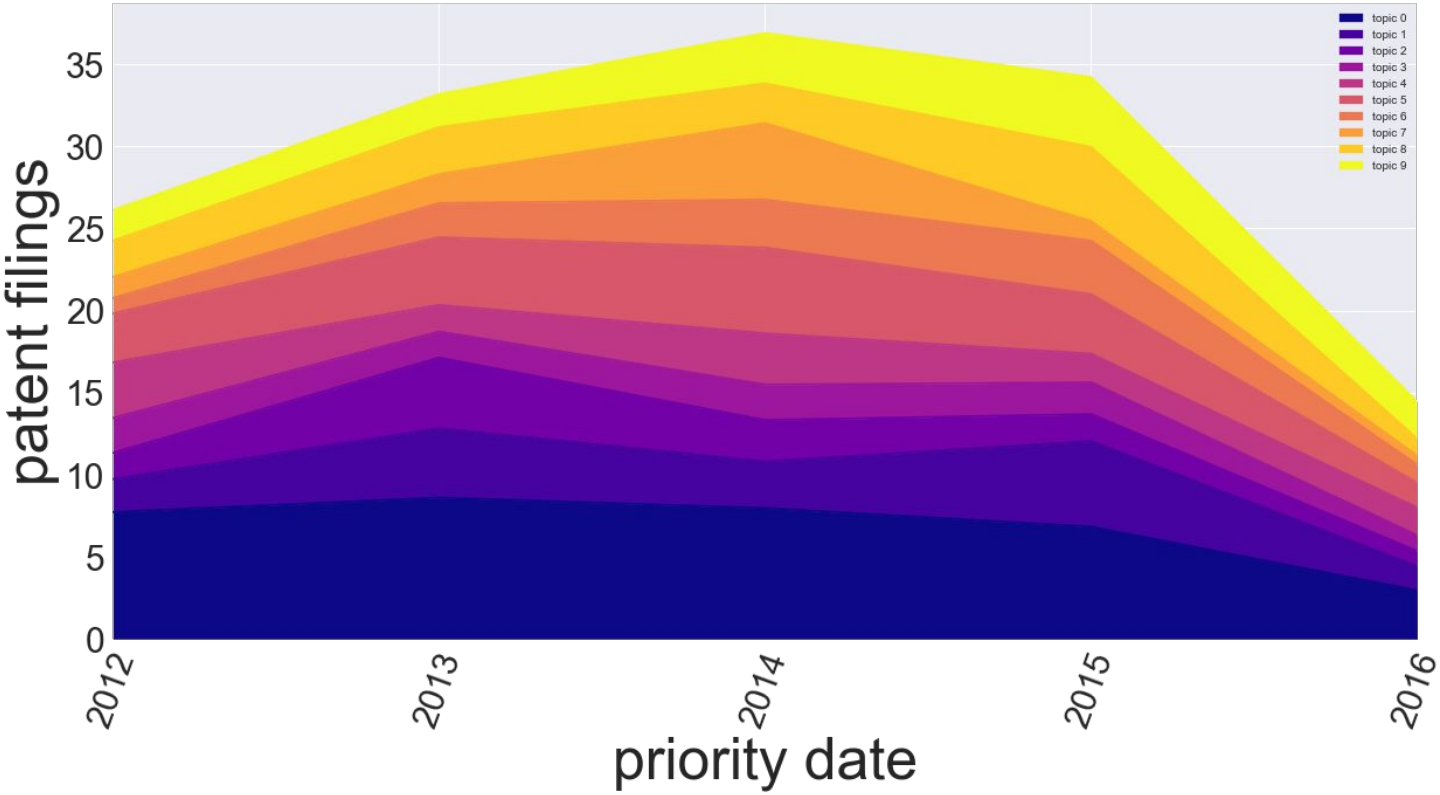
Dataset

- Abstracts
- Token count
  - Min 7
  - Max 167
  - Mean 76.49

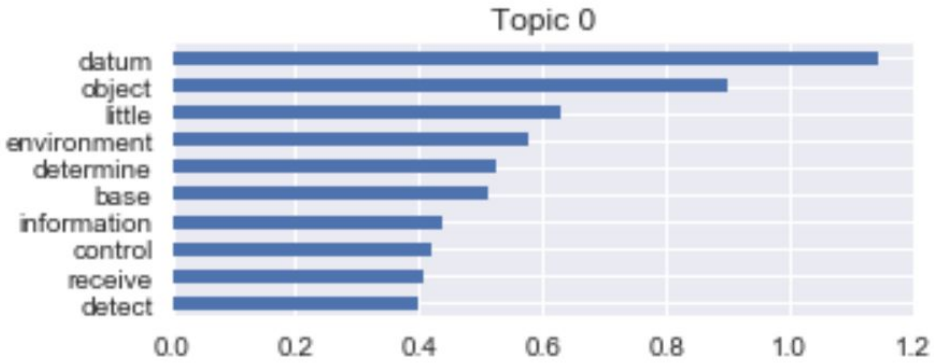
TFIDF Parameters

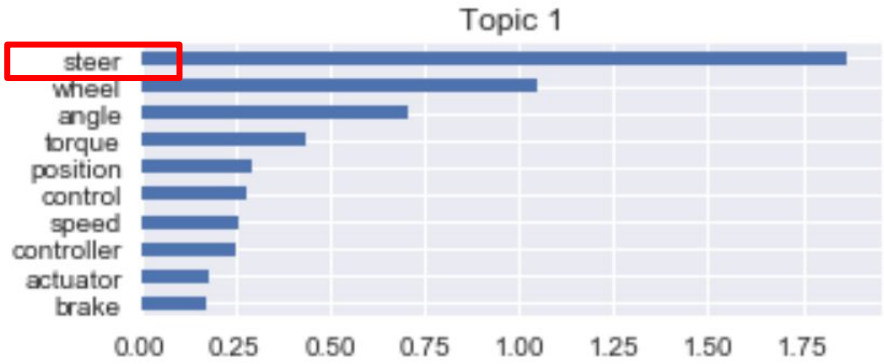
- Max_DF = 0.98
- Min_DF = 2

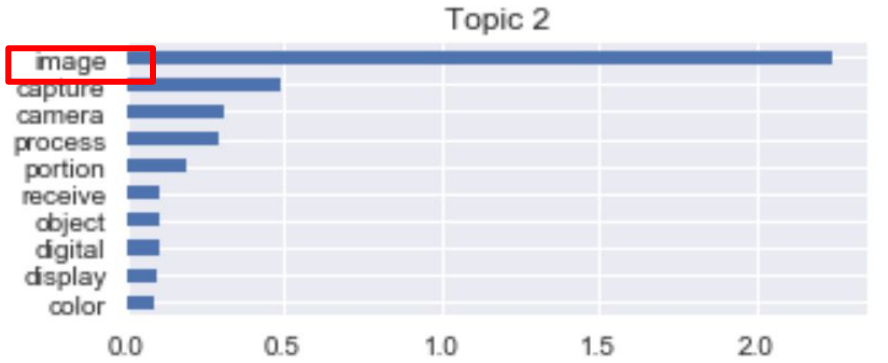
Results

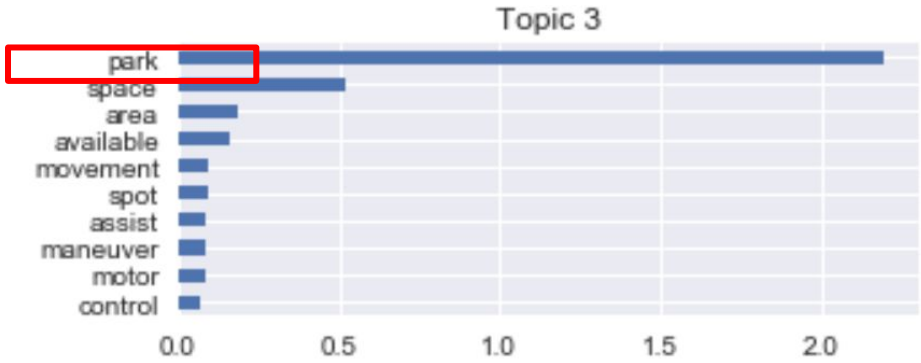

topic 0
topic 1
topic 2
topic 3
topic 4
topic 5
topic 6
topic 7
topic 8
topic 9
patent filings
0
5
10
15
20
25
30
35
2012
2013
2014
2015
2016
priority date

Results


Topic 0
datum
object
little
environment
determine
base
information
control
receive
detect
0.0
0.2
0.4
0.6
0.8
1.0
1.2


Topic 1
steer
wheel
angle
torque
position
control
speed
controller
actuator
brake
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75


Topic 2
image
capture
camera
process
portion
receive
object
digital
display
color
0.0
0.5
1.0
1.5
2.0


Topic 3
park
space
area
available
movement
spot
assist
maneuver
motor
control
0.0
0.5
1.0
1.5
2.0

Way Ahead

- TFIDF
- Predictive Analytics
  - Can patents in previous timeframe predict what will be in later timeframe?
  - PCA, K-Means Clustering

- Automated Patent Drafting

Questions?