Latent Dirichlet Allocation
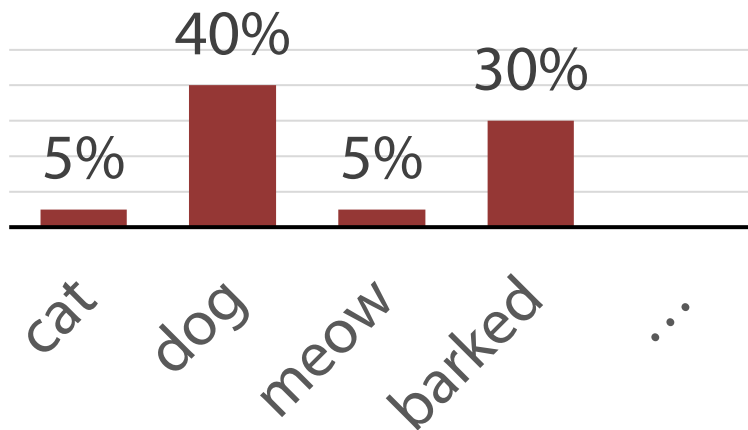
Document is a distribution over topics

Document →

80%

20%

"Cats" "Dogs"
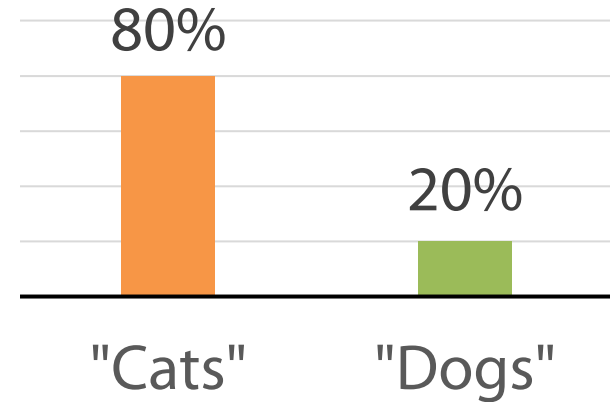
Topic is a distribution over words

Topic "Dogs"

5% 40% 5% 30%

cat dog meow barked ...

Topic "Cats"

40% 5% 30% 5%

cat dog meow barked ...