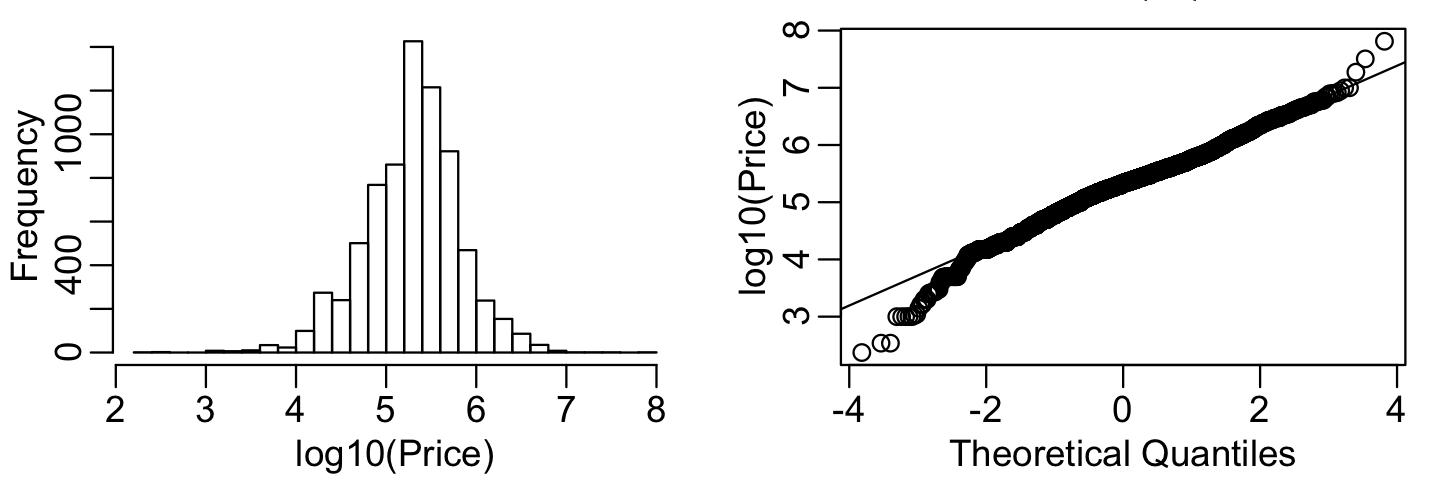
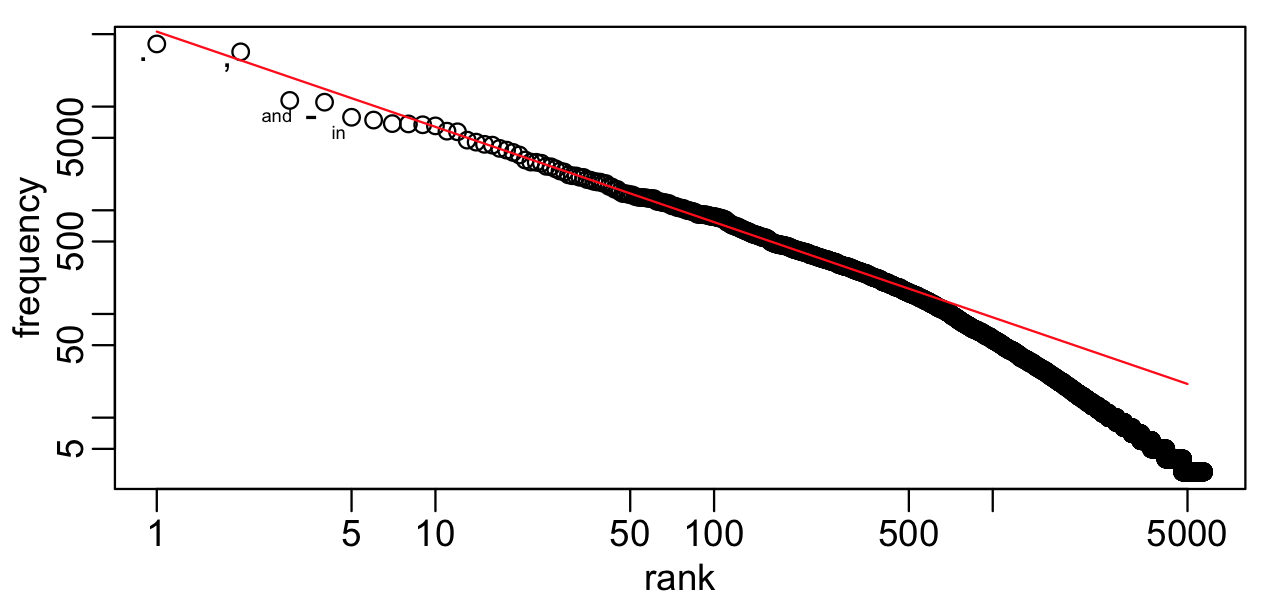
1. Log normal distribution for the prices

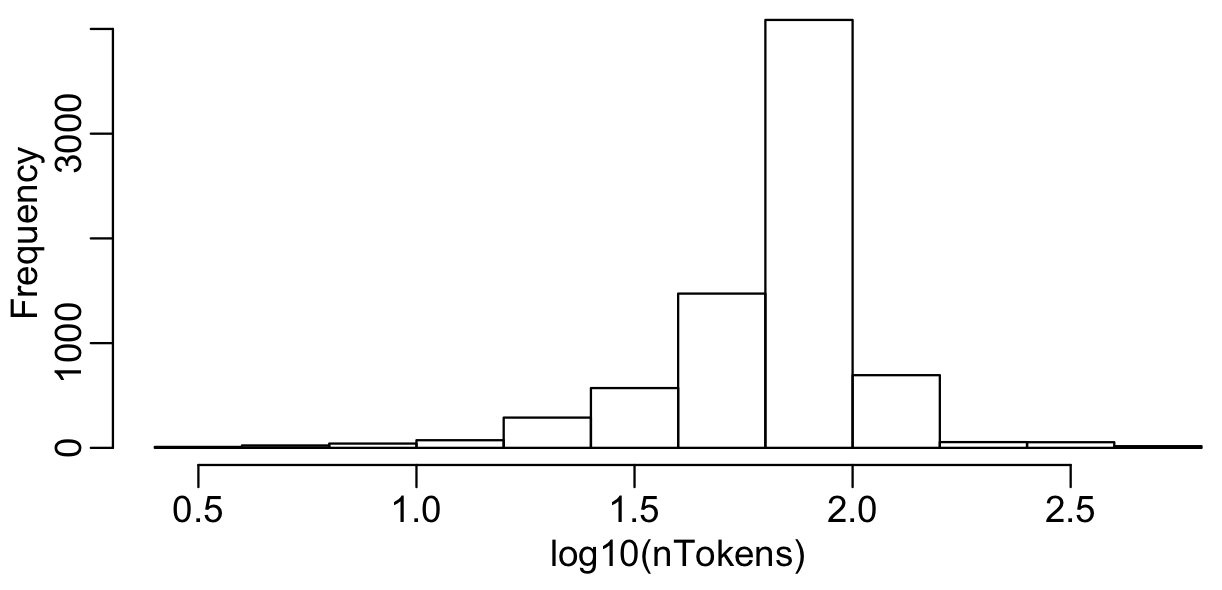


2. Near Zipf distribution for words (slope ≈ −0.92)



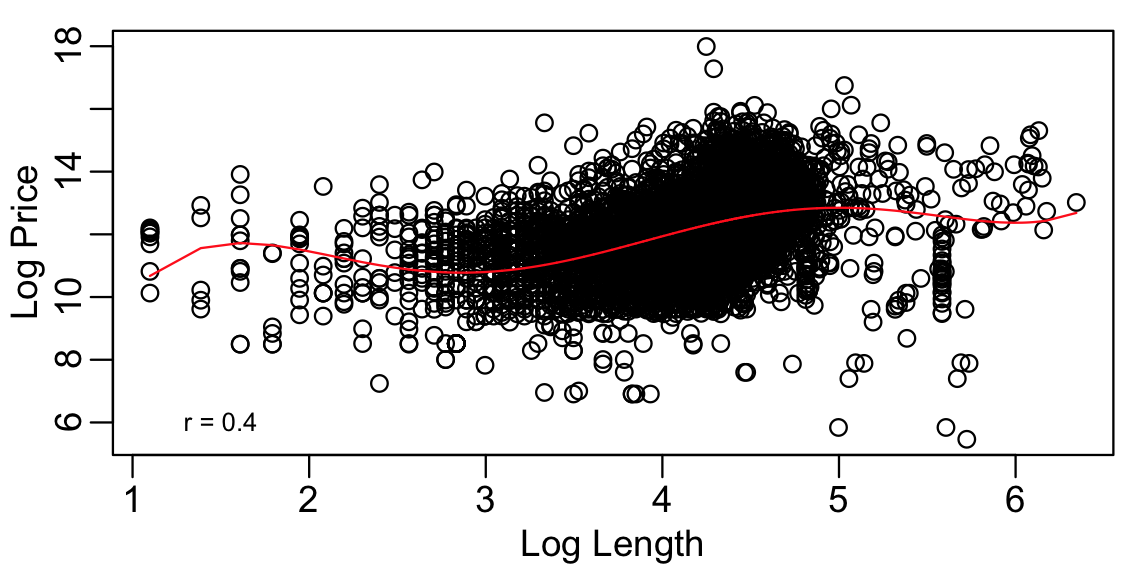
3. Skewed distribution of listing lengths (number of tokens). Log plot is fat tailed. Mean is 74 tokens with five number summary

3 55 75 86 569



4. Correlation between lengths and the log price (corr ≈ 0.4).

The association for listings is not linear. The effect initially seems near zero, up to lengths near 50. Then linear before right tail.



5. Signal concentrates in initial LSA coefficients, then gradually falls off (though signal remains)

