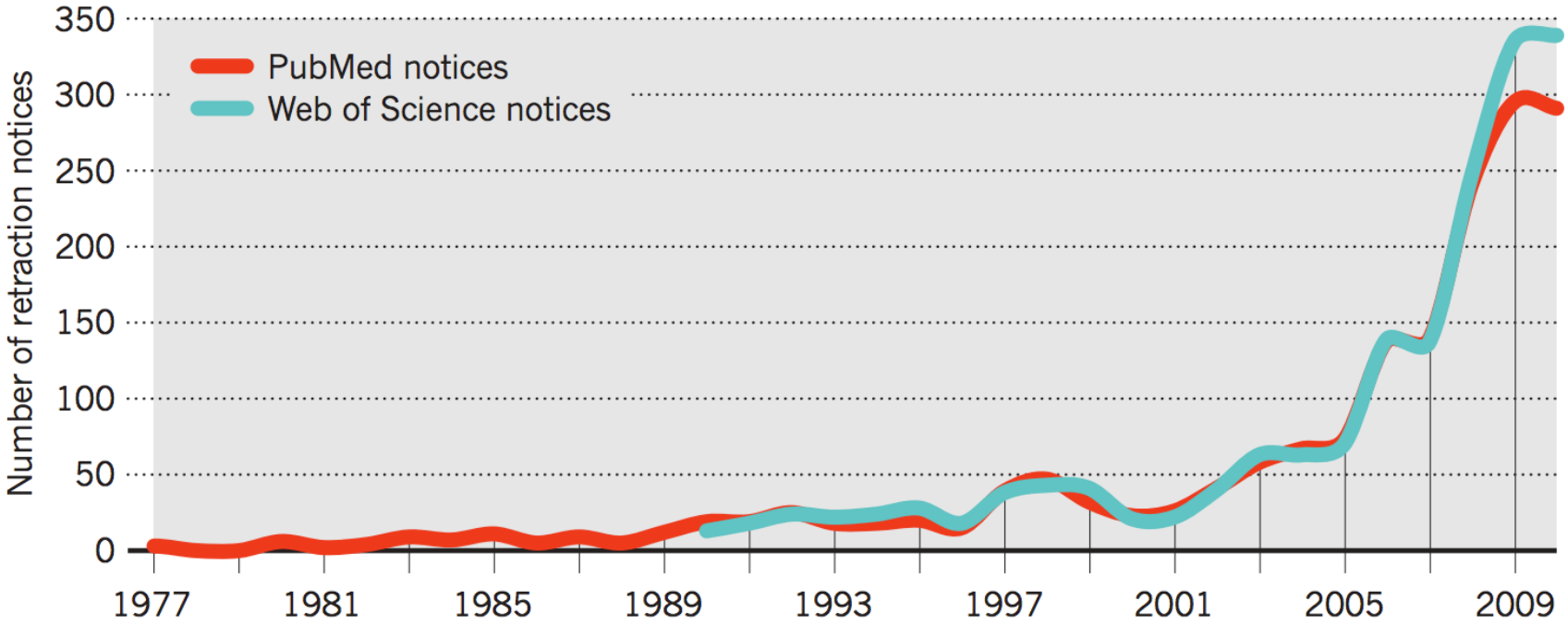
Reason 2: Mistakes and Fraud

2001 – 2011:

- 10X increase in retractions
- only 1.44X increase in papers

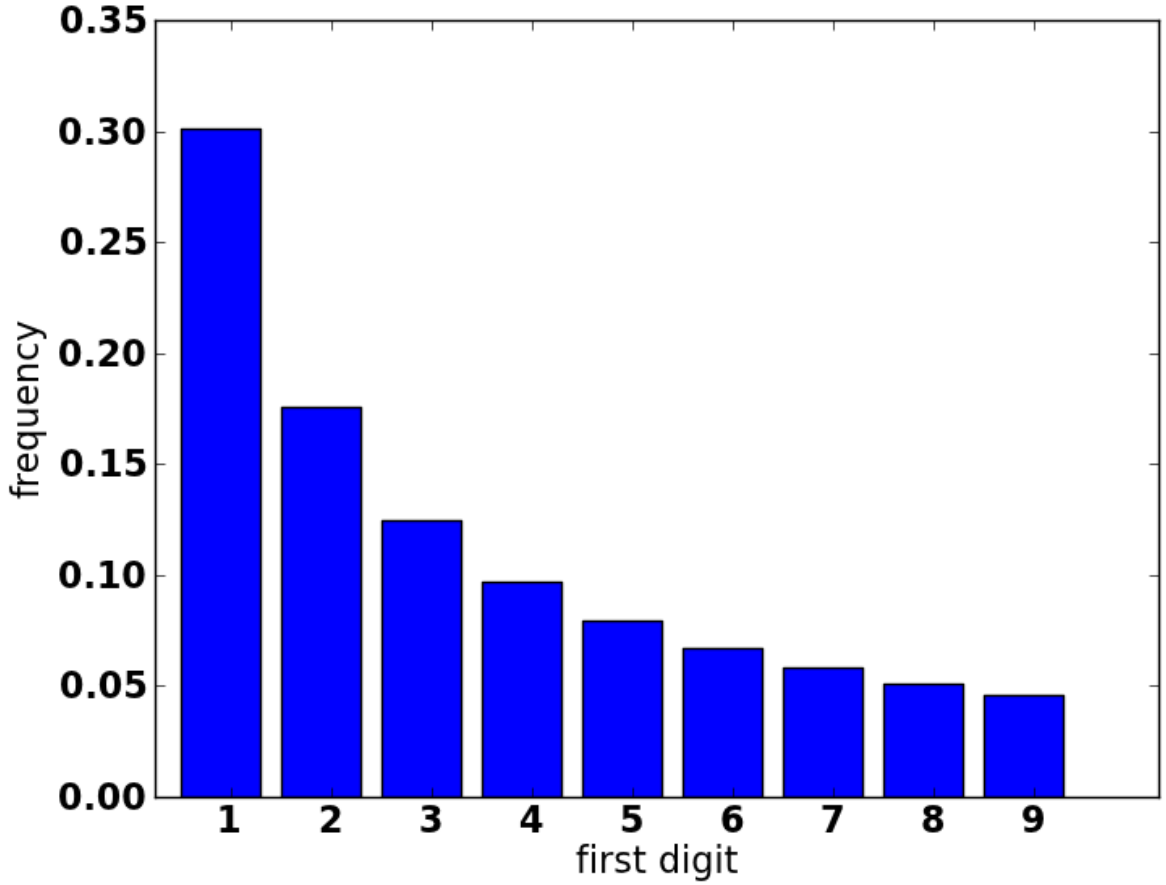


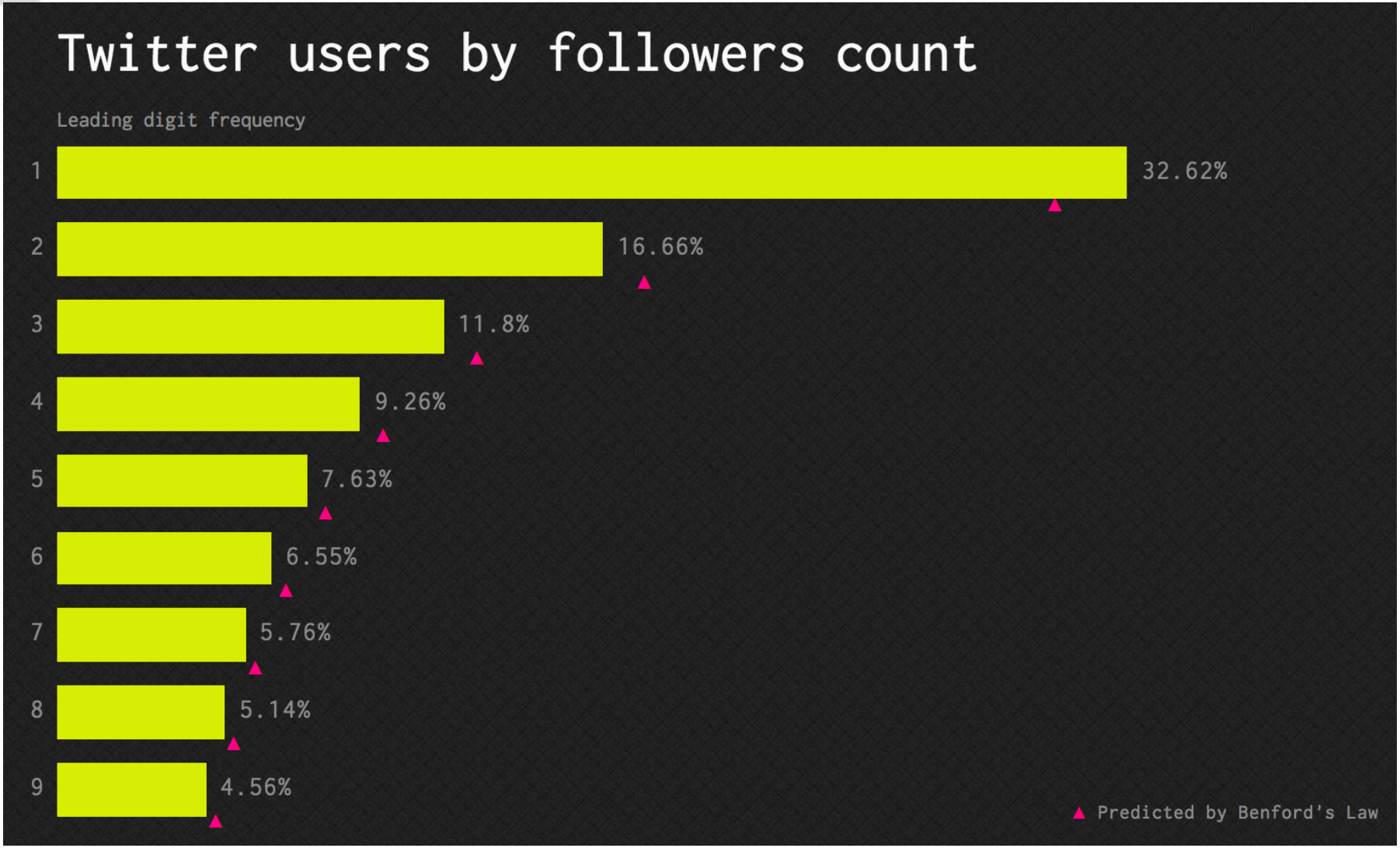
Richard Van Noorden, 2011, Nature 478
The Rise of the Retractions

http://www.nature.com/news/2011/111005/pdf/478026a.pdf

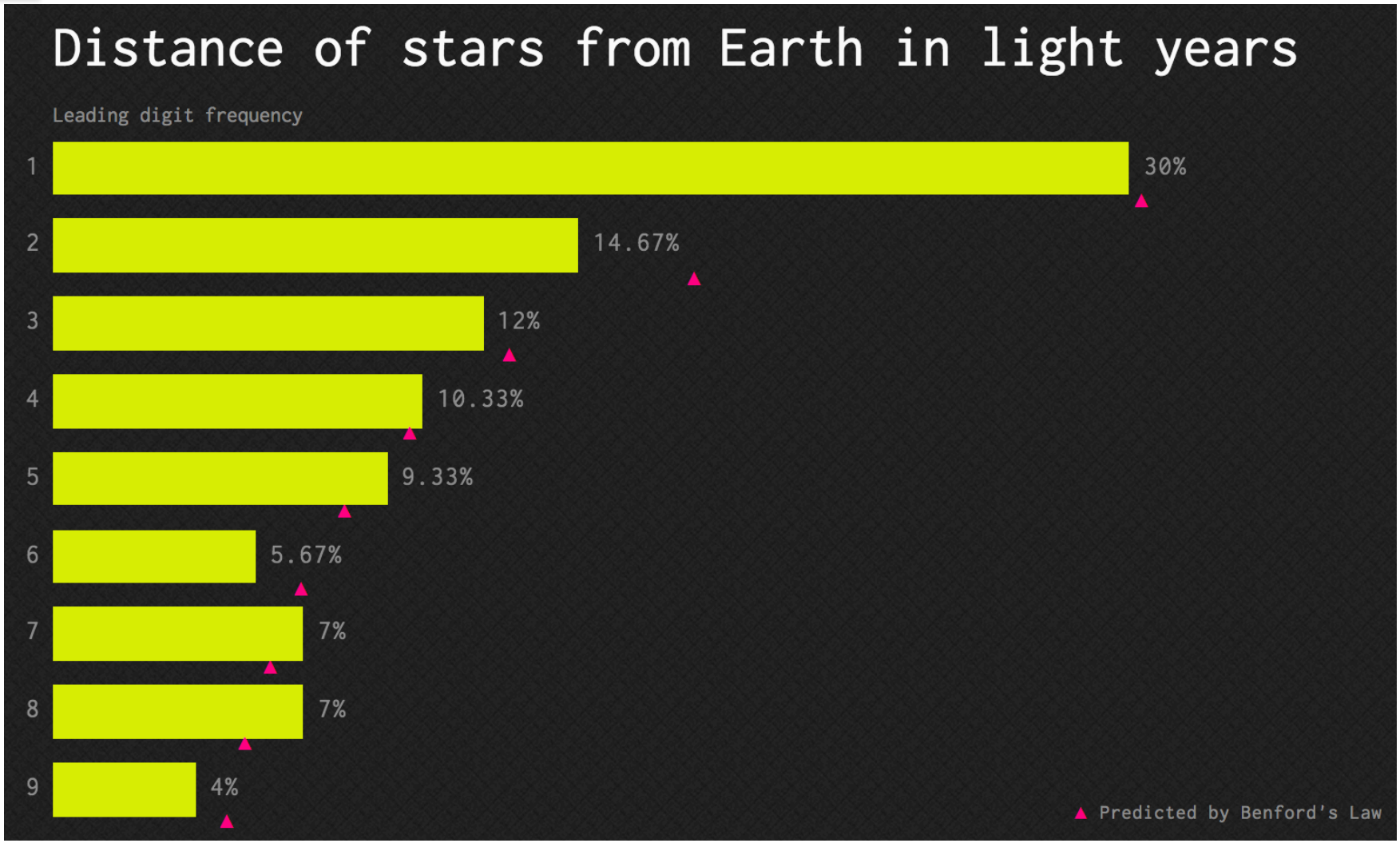
Benford's Law: potential tool for fraud detection

New York	8,336,697
Los Angeles	3,857,799
Chicago	2,714,856
Houston	2,160,821
Philadelphia	1,547,607
Phoenix	1,488,750
San Antonio	1,382,951
San Diego	1,338,348
Dallas	1,241,162

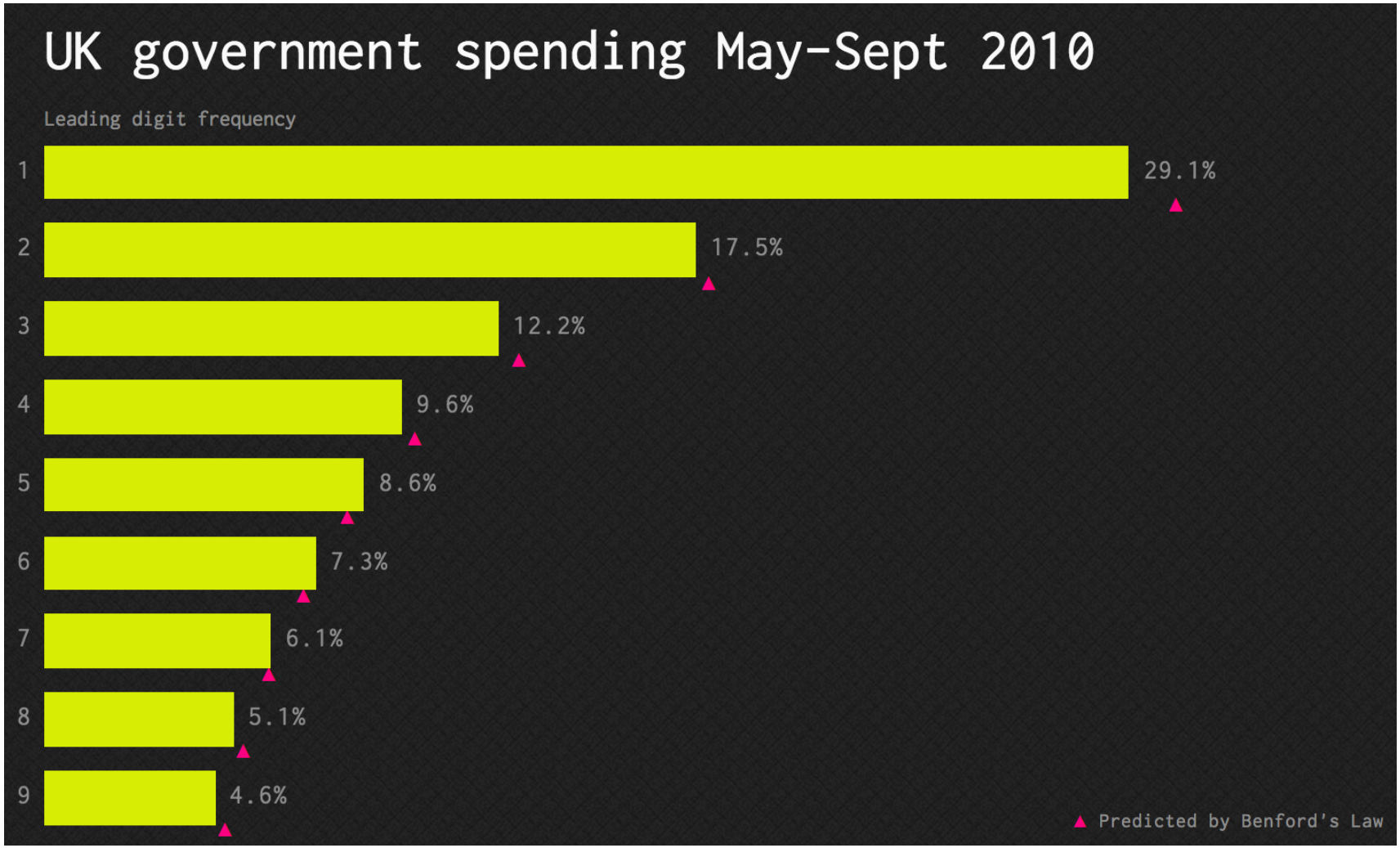




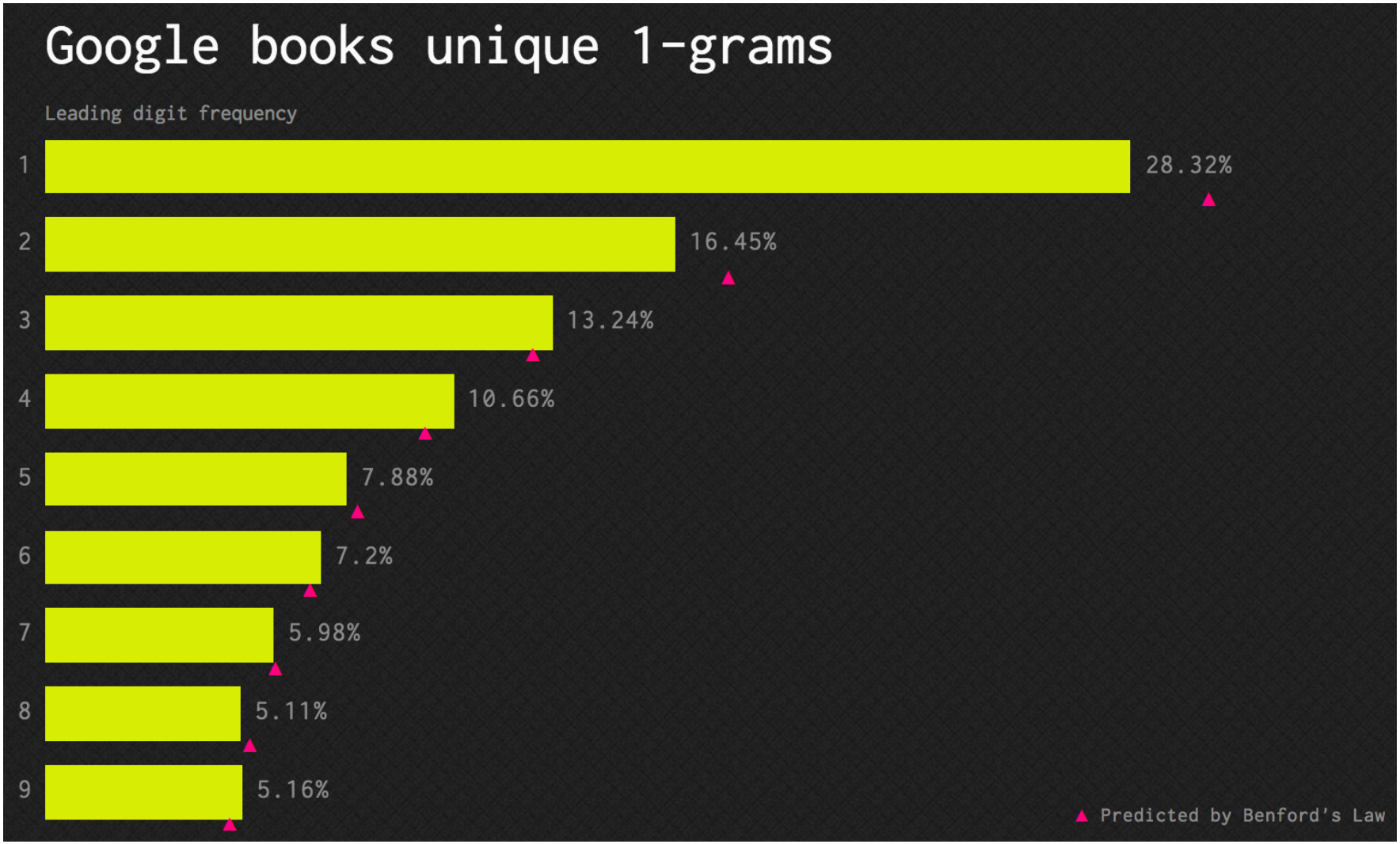
src: http://testingbenfordslaw.com/x



src: http://testingbenfordslaw.com/x



src: http://testingbenfordslaw.com/x



src: http://testingbenfordslaw.com/x

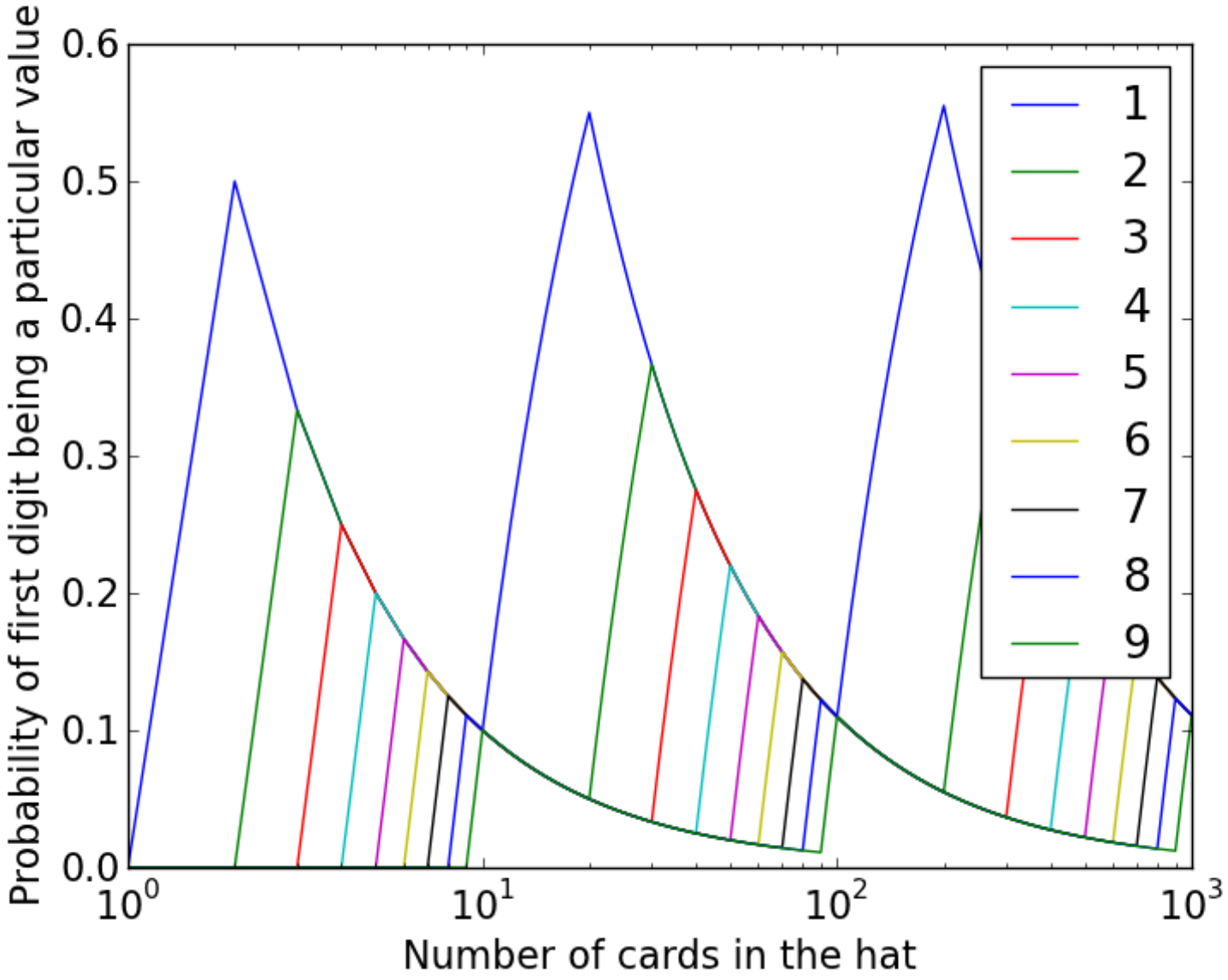
Benford's Law to Detect Fraud

- Diekmann, 2007
  - Found that first and second digits of published statistical estimates were approximately Benford distributed
  - Asked subjects to manufacture regression coefficients, and found that the first digits were hard to detect as anomolous, but the second and third digits deviated from expected distributions significantly.

Andreas Diekmann, 2007, Journal of Applied Statistics, 34(3)
Not the First Digit! Using Benford's Law to Detect Fraudulent Scientific Data

Benford's Law Intuition

- Given a sequence of cards labeled 1, 2, 3, … 999999
- Put them in a hat, one by one, in order
- After each card, ask "What is the probability of drawing the number 1?"


Probability of first digit being a particular value
0.0
0.1
0.2
0.3
0.4
0.5
0.6
10^0
10^1
10^2
10^3
Number of cards in the hat
1
2
3
4
5
6
7
8
9

Benford's Law Explanation

Limitations

- Data must span several orders of magnitude
- No min/max cutoffs