Data Forensics via:

Digital Development Pattern

Benford's Law is a statement about the digital configuration of the **entire data set**.

Could we then naively conclude that any **sub-set of the data** on any segment of the x-axis is also Benford?

Does the whole endow its Benford property to its parts?

NO!

Local digit configuration changes and develops as we move from the left to the right on the x-axis!

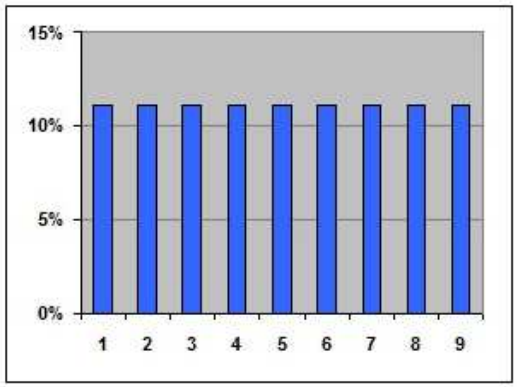
By some amazing coincidence…

Digital Behavior = Political Narratives

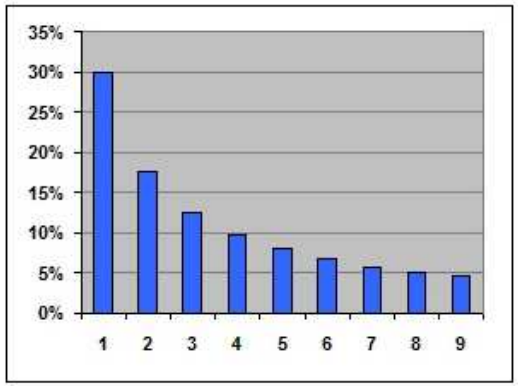
LEFT	CENTER	FAR RIGHT
Material equality	Some inequality	Extreme inequality

LEFT | **CENTER** | **FAR RIGHT**

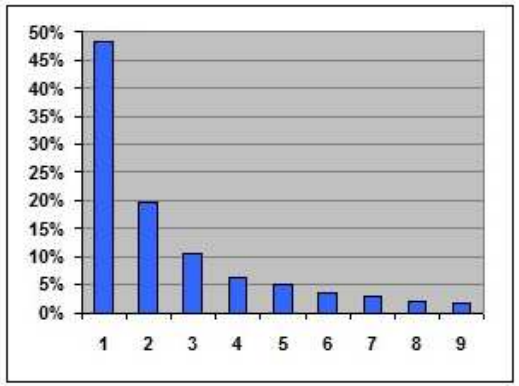
Digital Equality

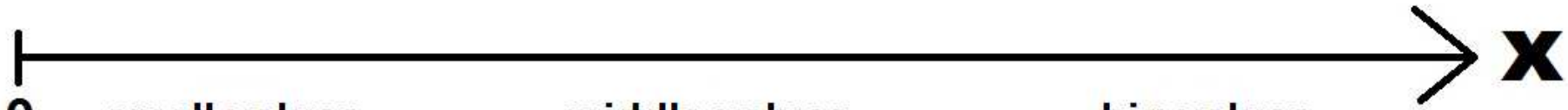


Benford



Extreme digital inequality





0 **small values** **middle values** **big values**

This can be seen on sub-intervals between

… 0.001, 0.01, 0.1, 1, 10, 100, 1000, 10000, 100000, …

All honest data should come with Digital Development Pattern!

Let us check this **fact of life** empirically:

Canford Audio PLC

(Honest Data)

The entire catalog of Canford Audio PLC in the U.K. which manufactures and retails 15,194 electronic items.

http://www.canford.co.uk/

Canford Audio PLC - UK

Catalog

Price: $20.79

Price: $15.64

Price: $31.99

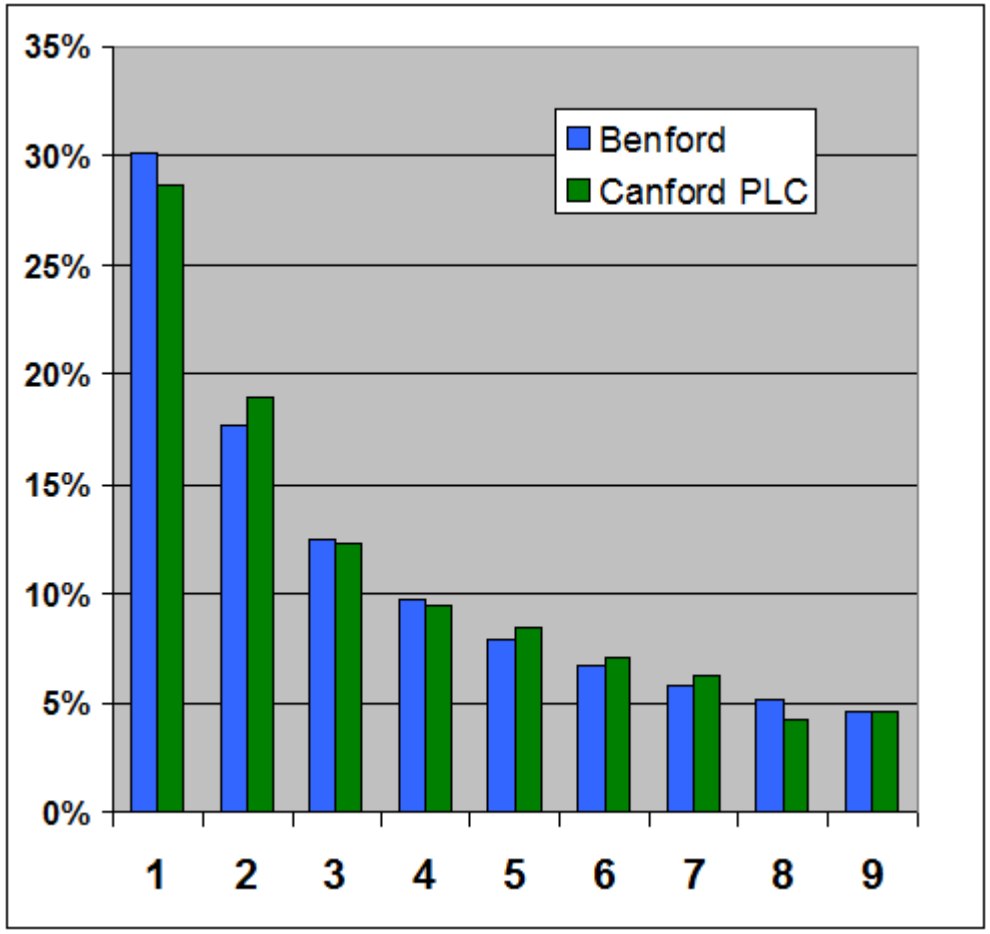
Price: $6.38

Price: $25.50

Canford Audio PLC

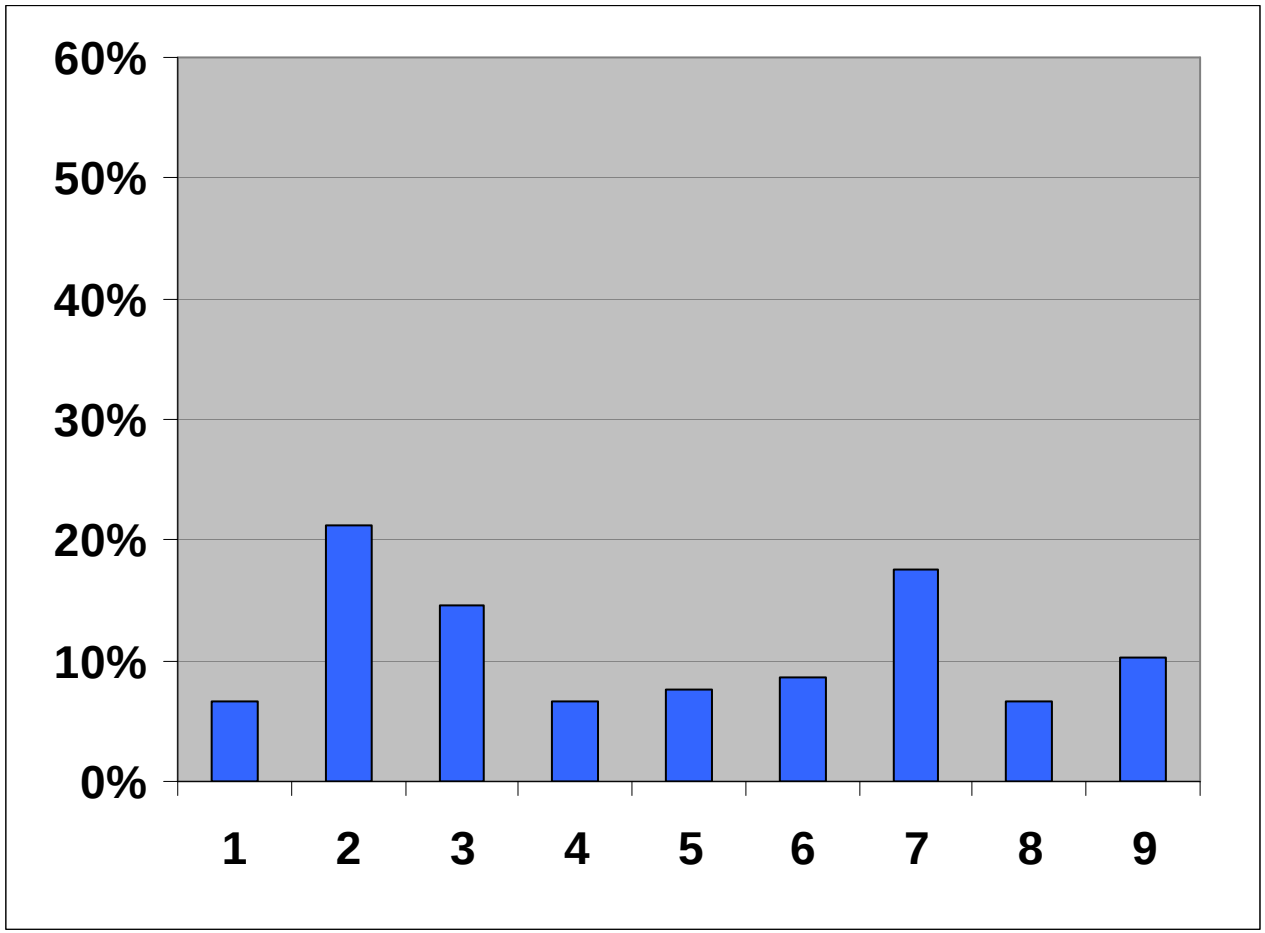
Benford	*Canford*
30.1%	28.6%
17.6%	19.0%
12.5%	12.3%
9.7%	9.4%
7.9%	8.5%
6.7%	7.1%
5.8%	6.3%
5.1%	4.2%
4.6%	4.6%

Canford Audio PLC

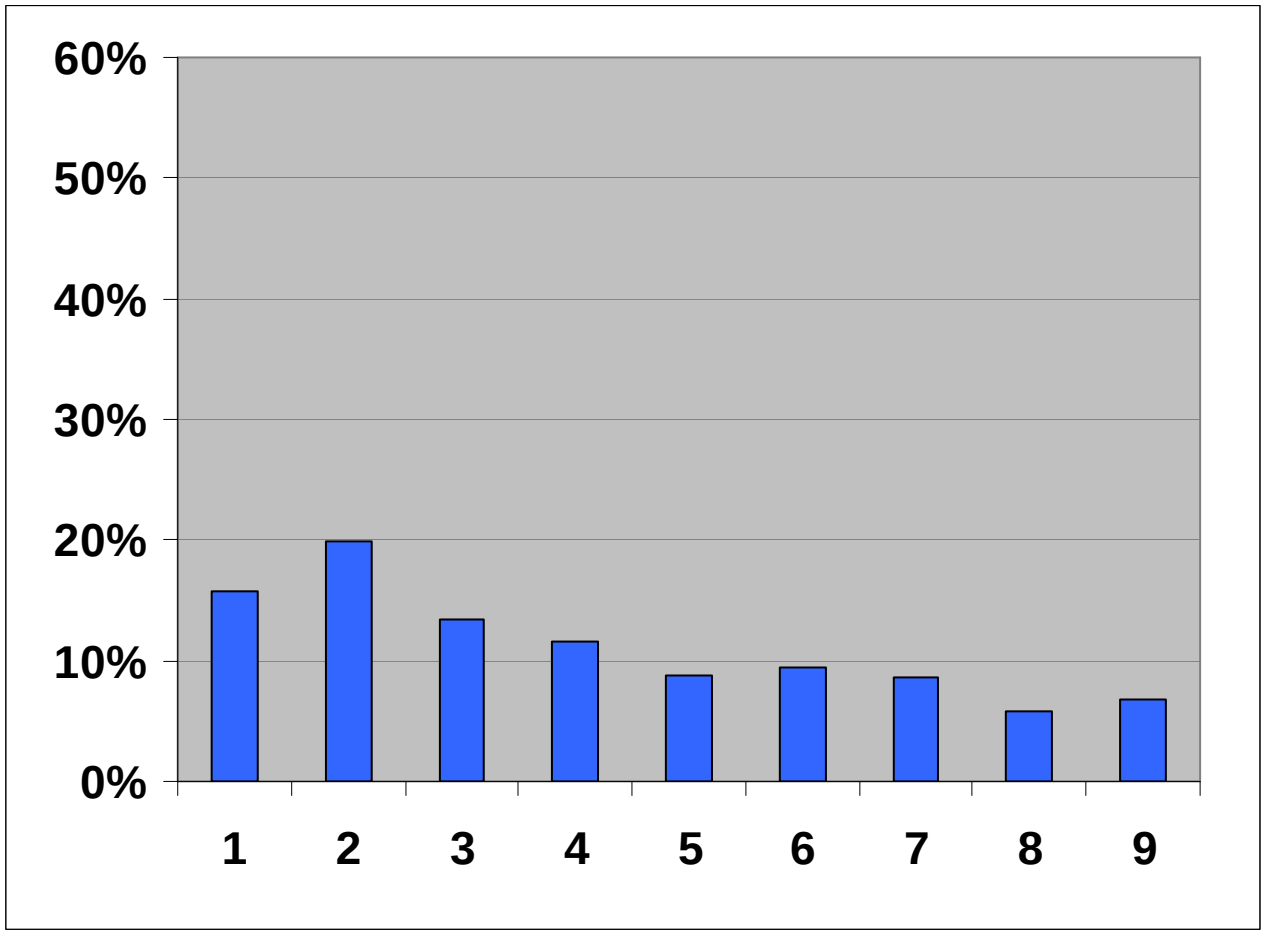


Yet… on local sub-intervals
it develops from left to right:

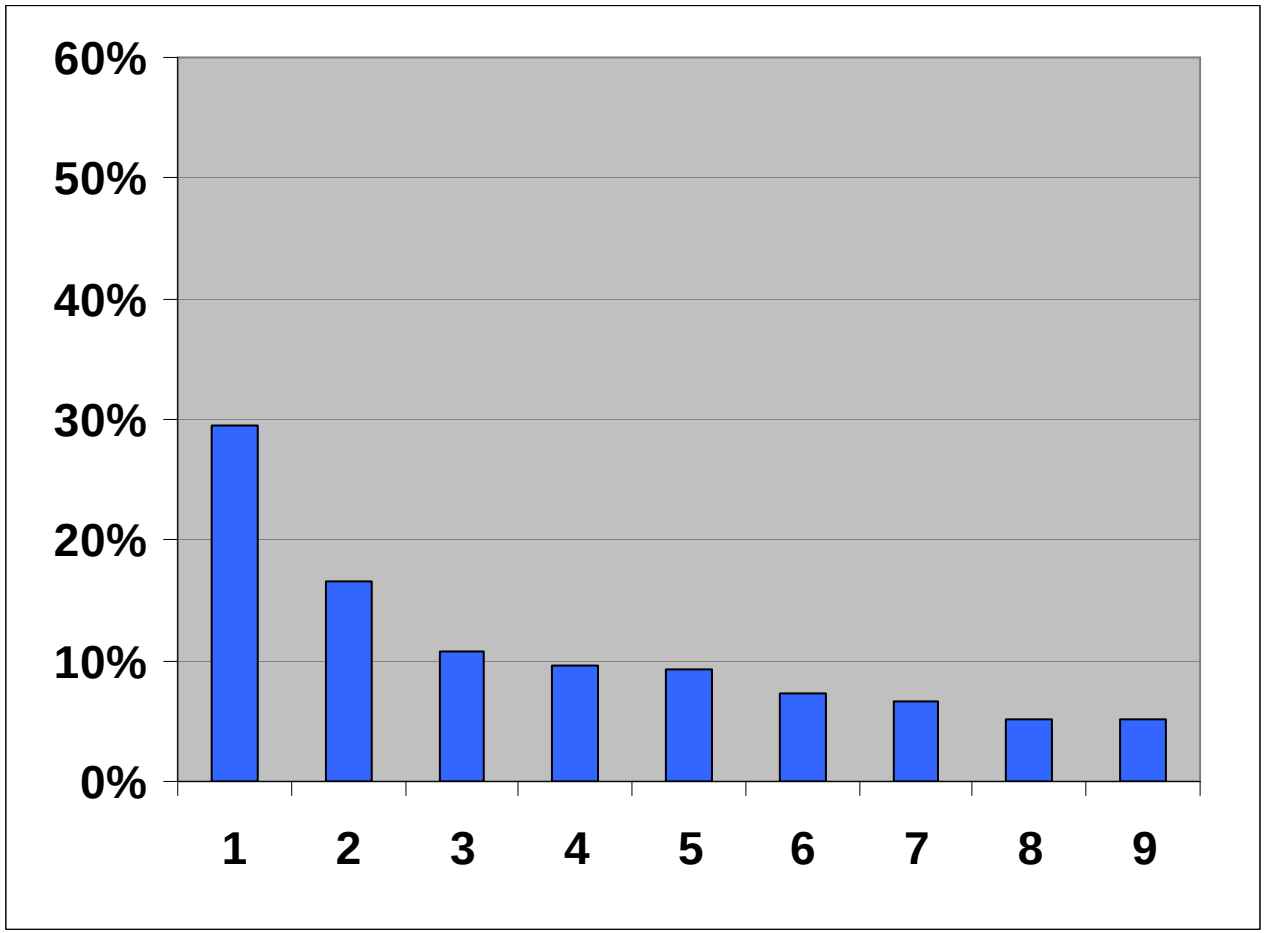
0.1 – 1



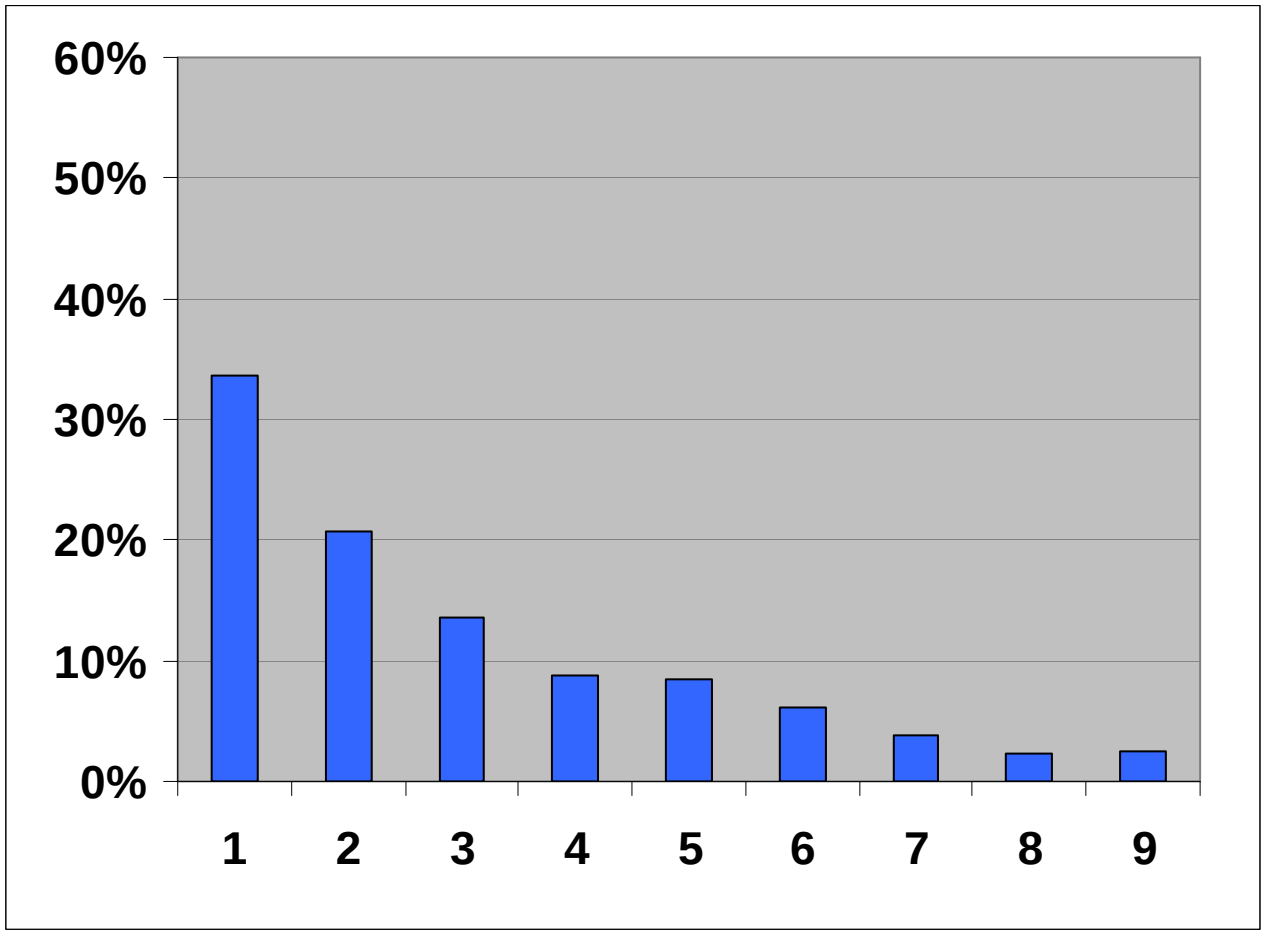
1 – 10



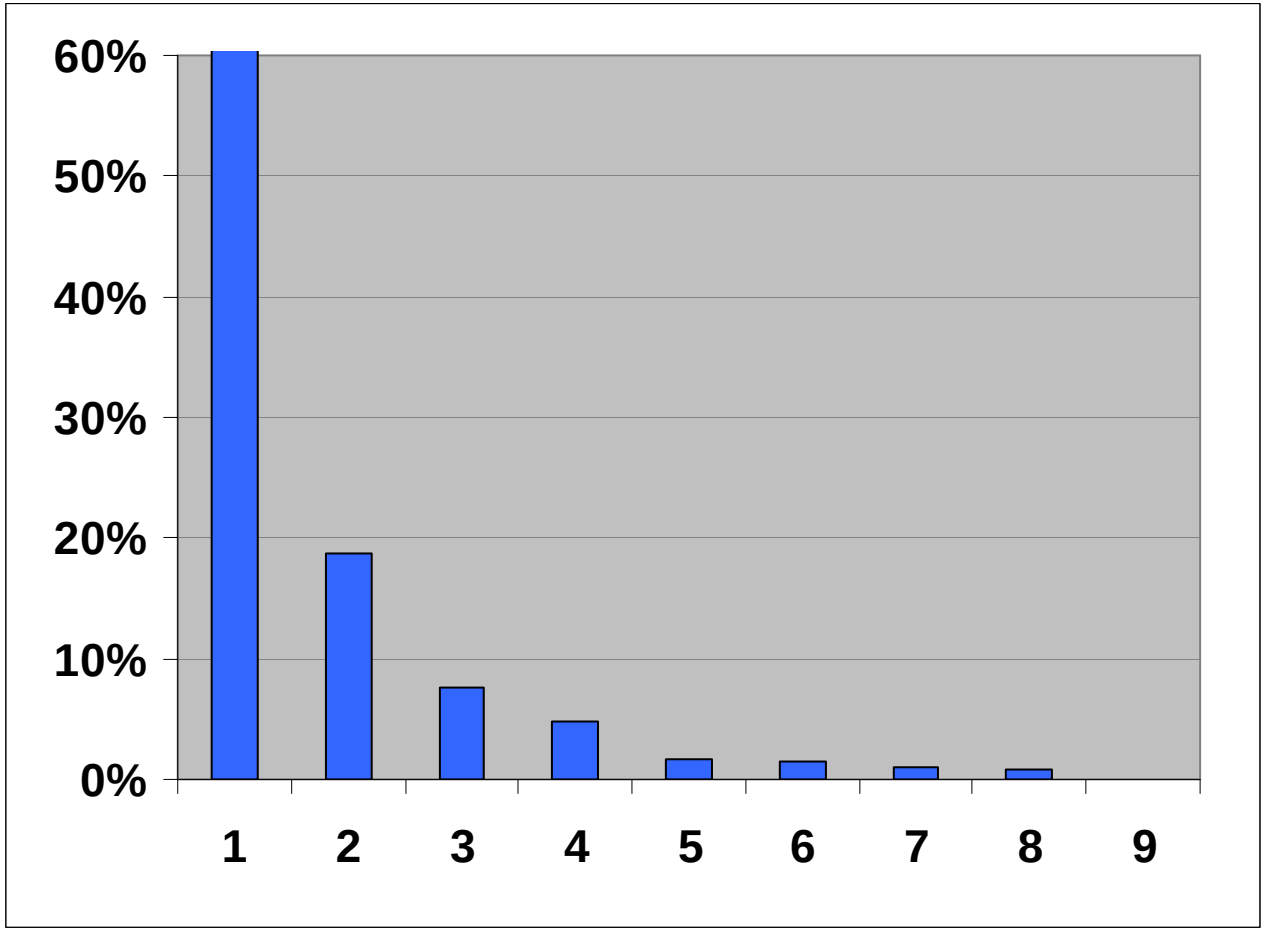
10 – 100



100 – 1000



1000 – 10000



NASDAQ USA Market Capitalization

(Honest Data)

Oct 9, 2016

2,889 companies

http://www.nasdaq.com/screening/companies-by-industry.aspx?exchange=NASDAQ

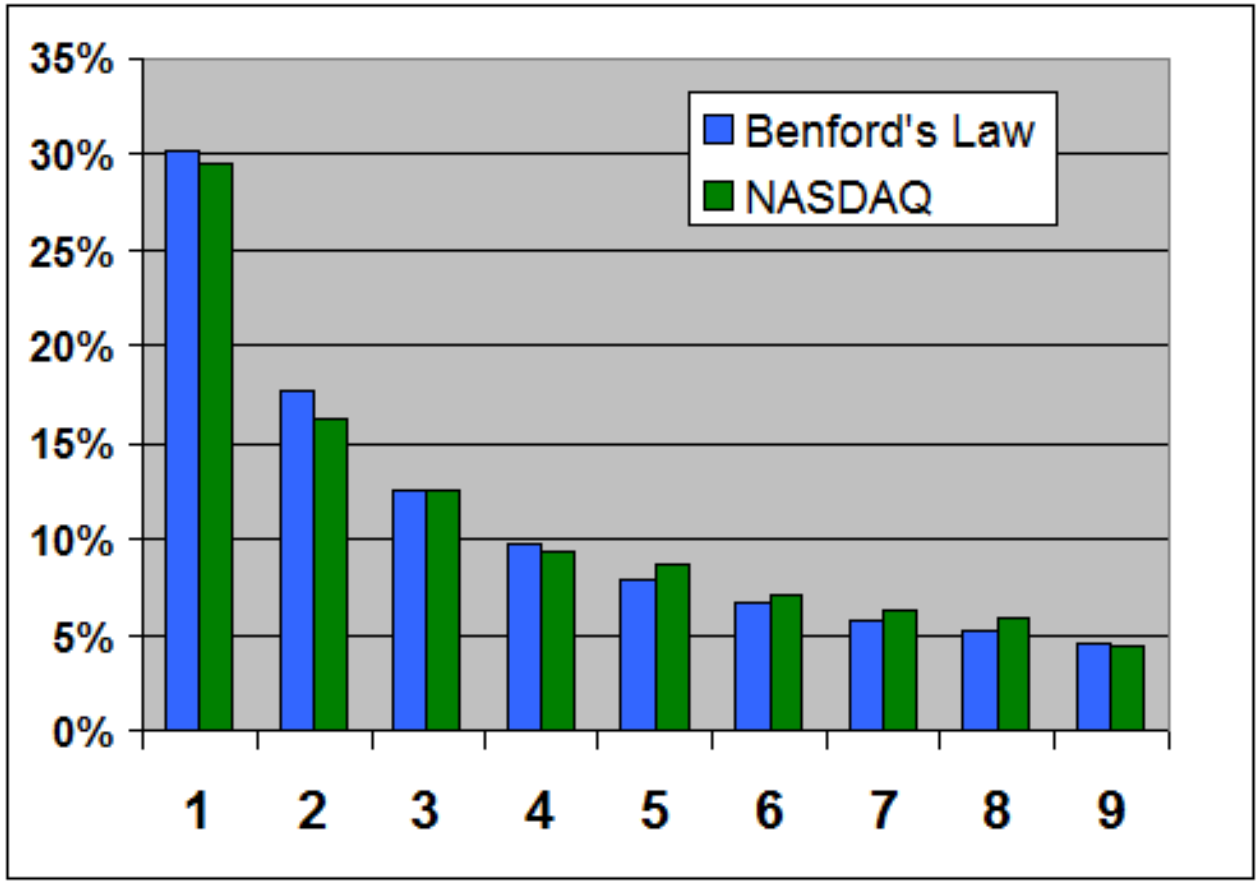

NASDAQ
NASDAQ COMPOSITE
+0.4%
2,677.01 +10.65
+1.3%
44.45 +0.56
+1.3%
17.17 +0.22
CITRIX
DELL
-0.1%
-0.08
NASDAQ


WE'RE CHANGING THE WAY THE WORLD TRADES.
NASDAQ OMX
© 2008 The NASDAQ OMX Group, Inc. All Rights Reserved.
NASDAQ
DO NOT ENTER

NASDAQ Market Capitalization

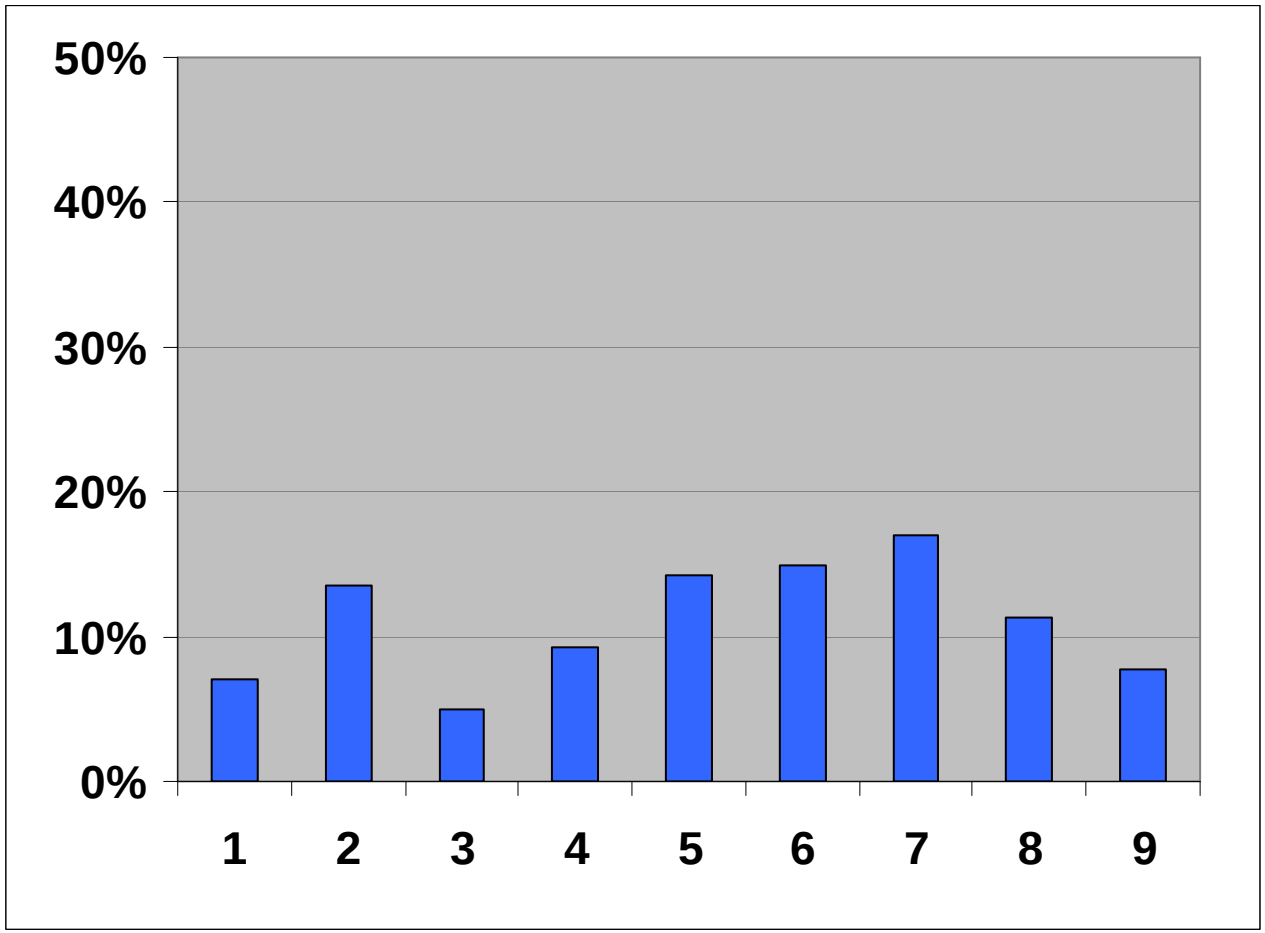
Benford	*NASDAQ*
30.1%	29.6%
17.6%	16.3%
12.5%	12.5%
9.7%	9.3%
7.9%	8.6%
6.7%	7.2%
5.8%	6.3%
5.1%	5.9%
4.6%	4.4%

NASDAQ Market Capitalization

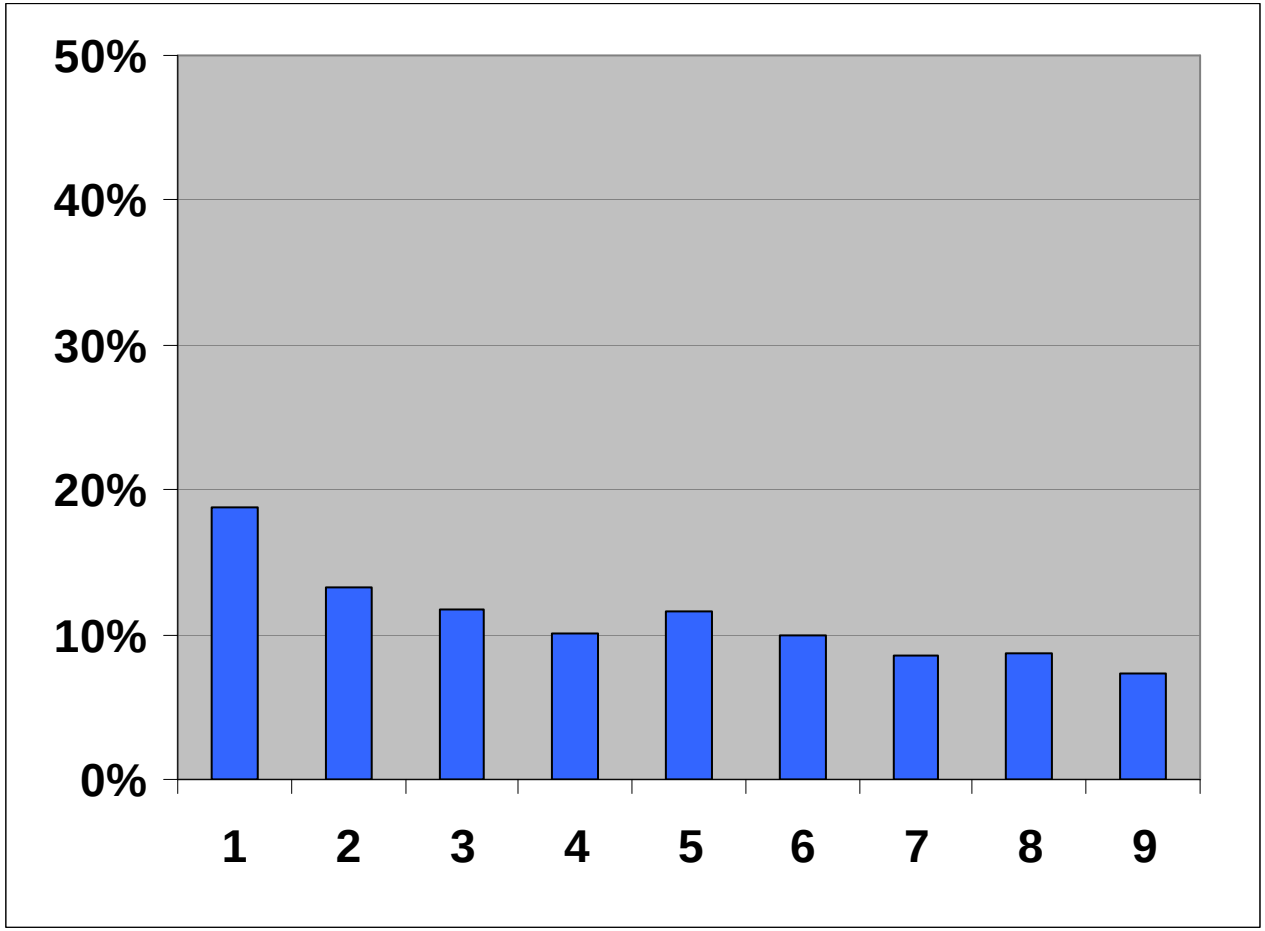


Yet… on local sub-intervals
it develops from left to right:

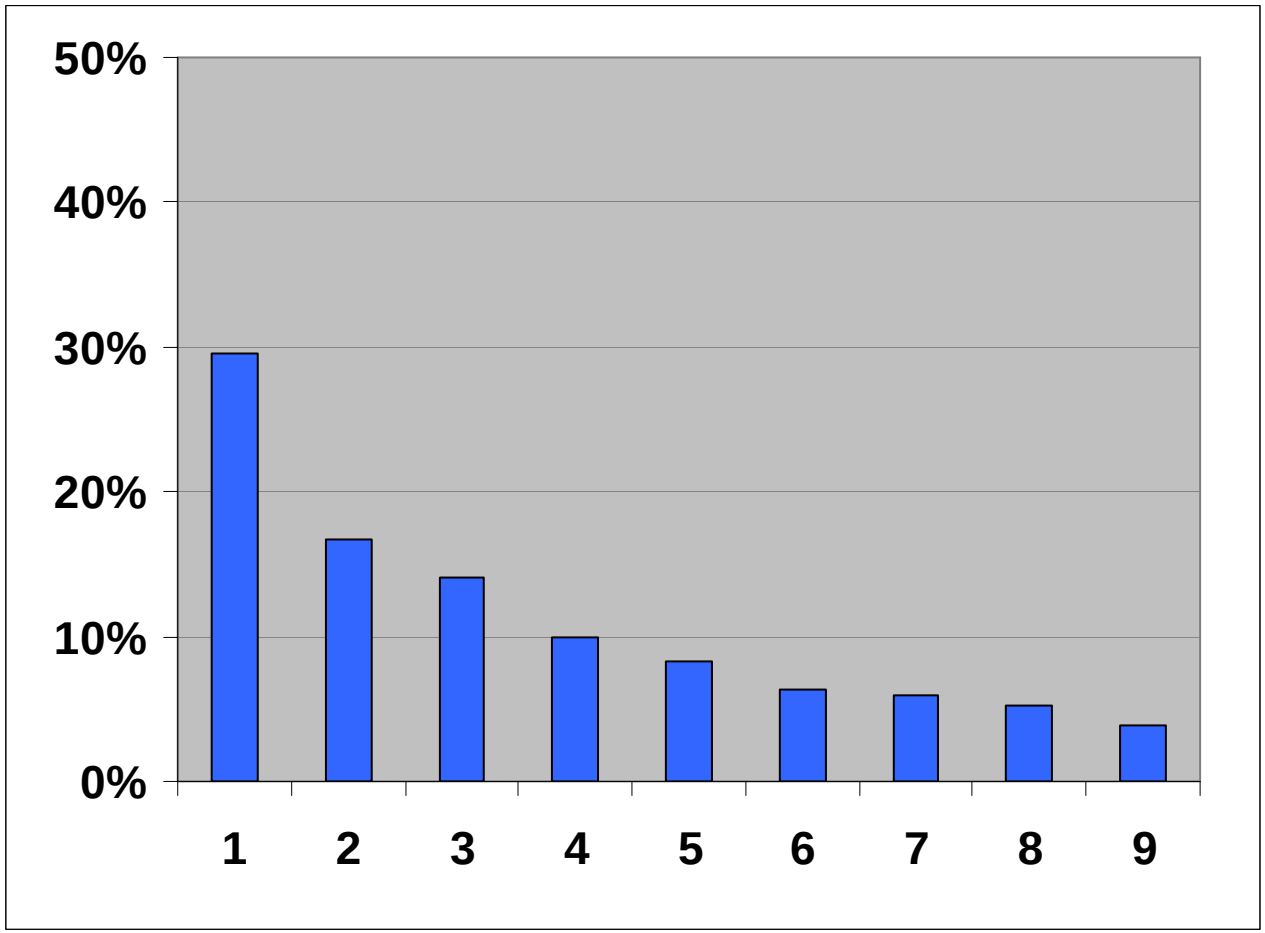
$1 Million – $10 Million



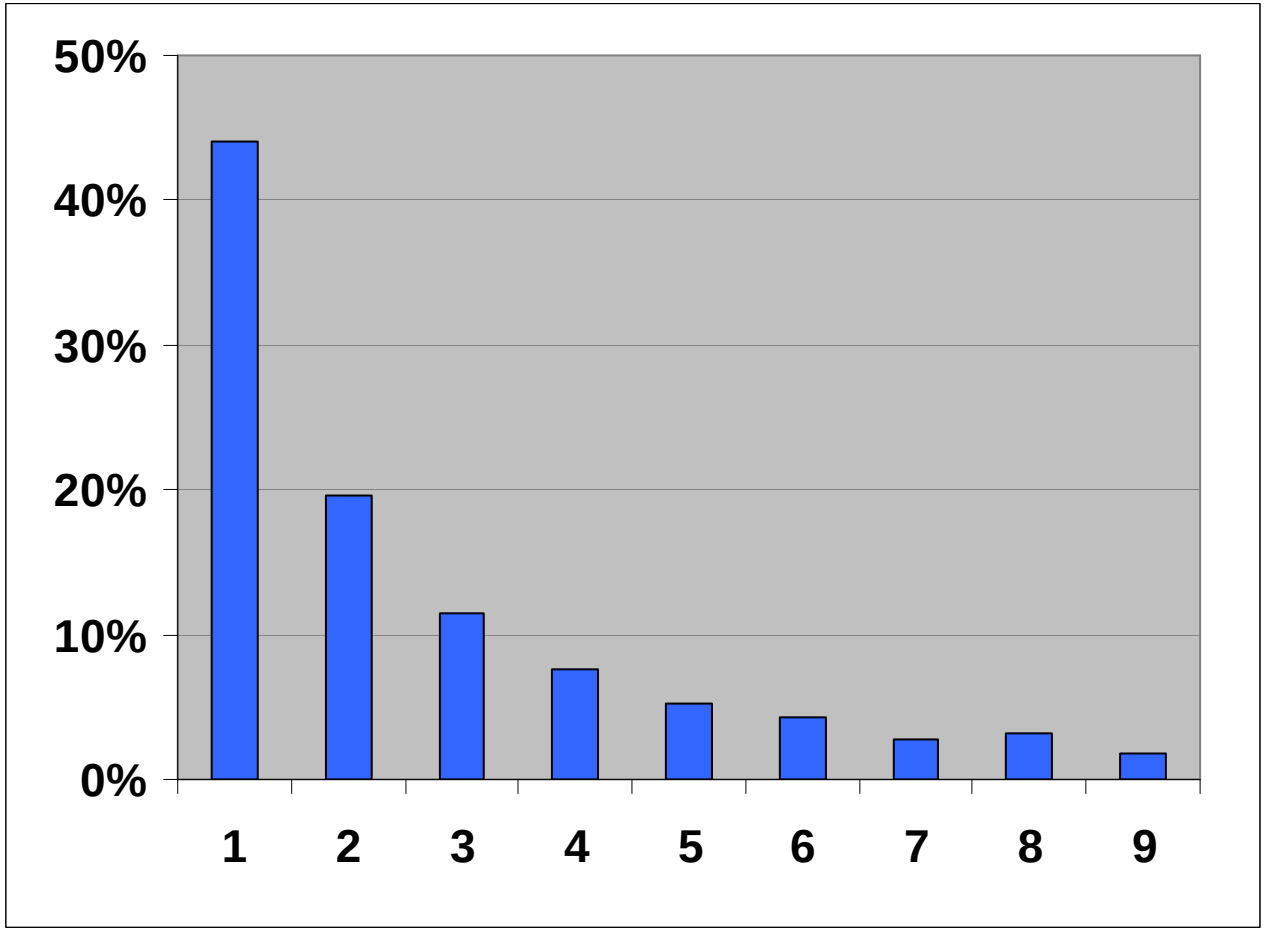
$10 Million – $100 Million



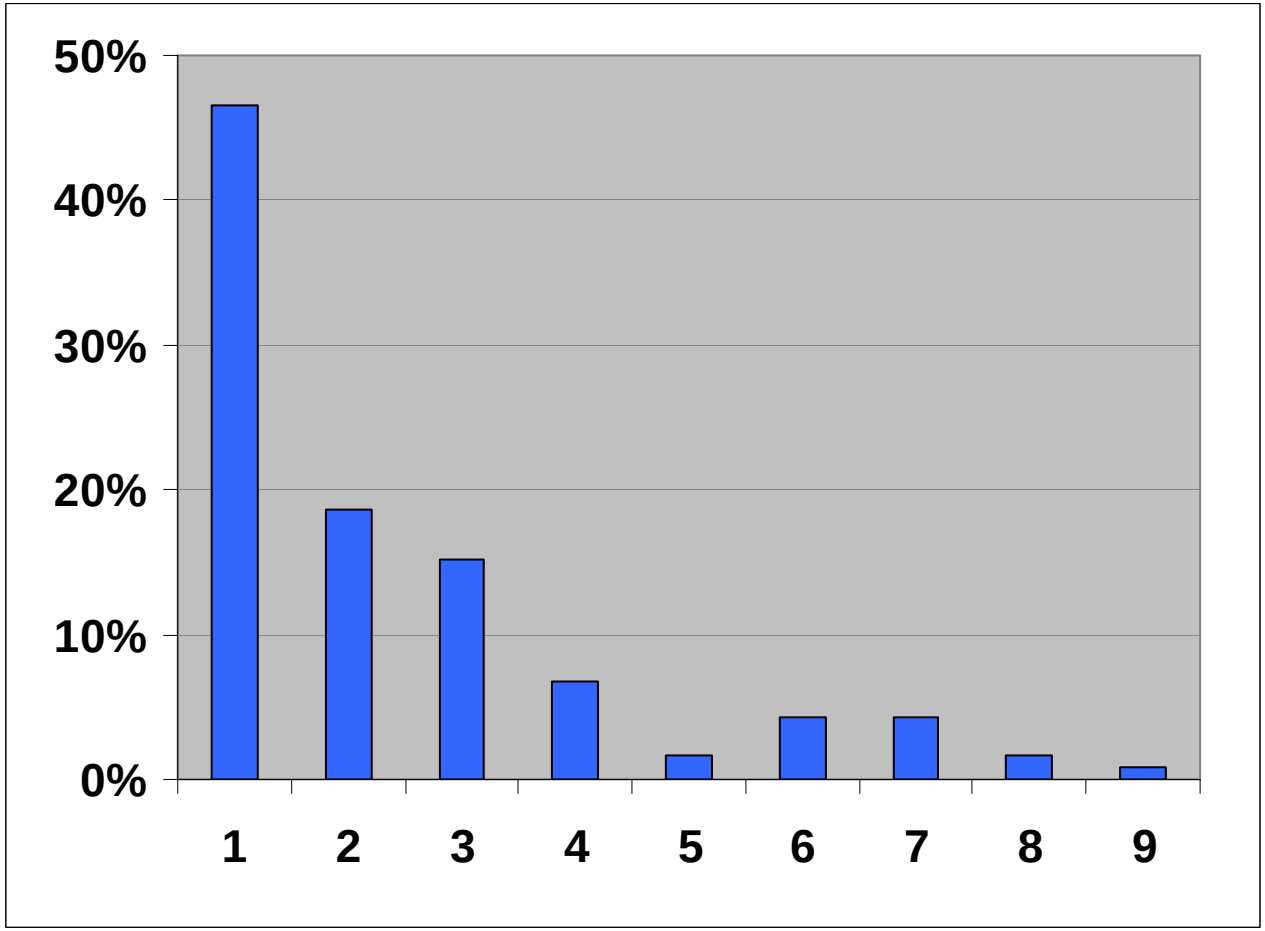
$100 Million – $1 Billion



$1 Billion– $10 Billion



$10 Billion – $100 Billion



Oklahoma State Expenses

(Honest Data)

The State Of Oklahoma in the USA
966,990 payments in 2011.

https://data.ok.gov/dataset/state-oklahoma-vendor-payments-fiscal-year-2011

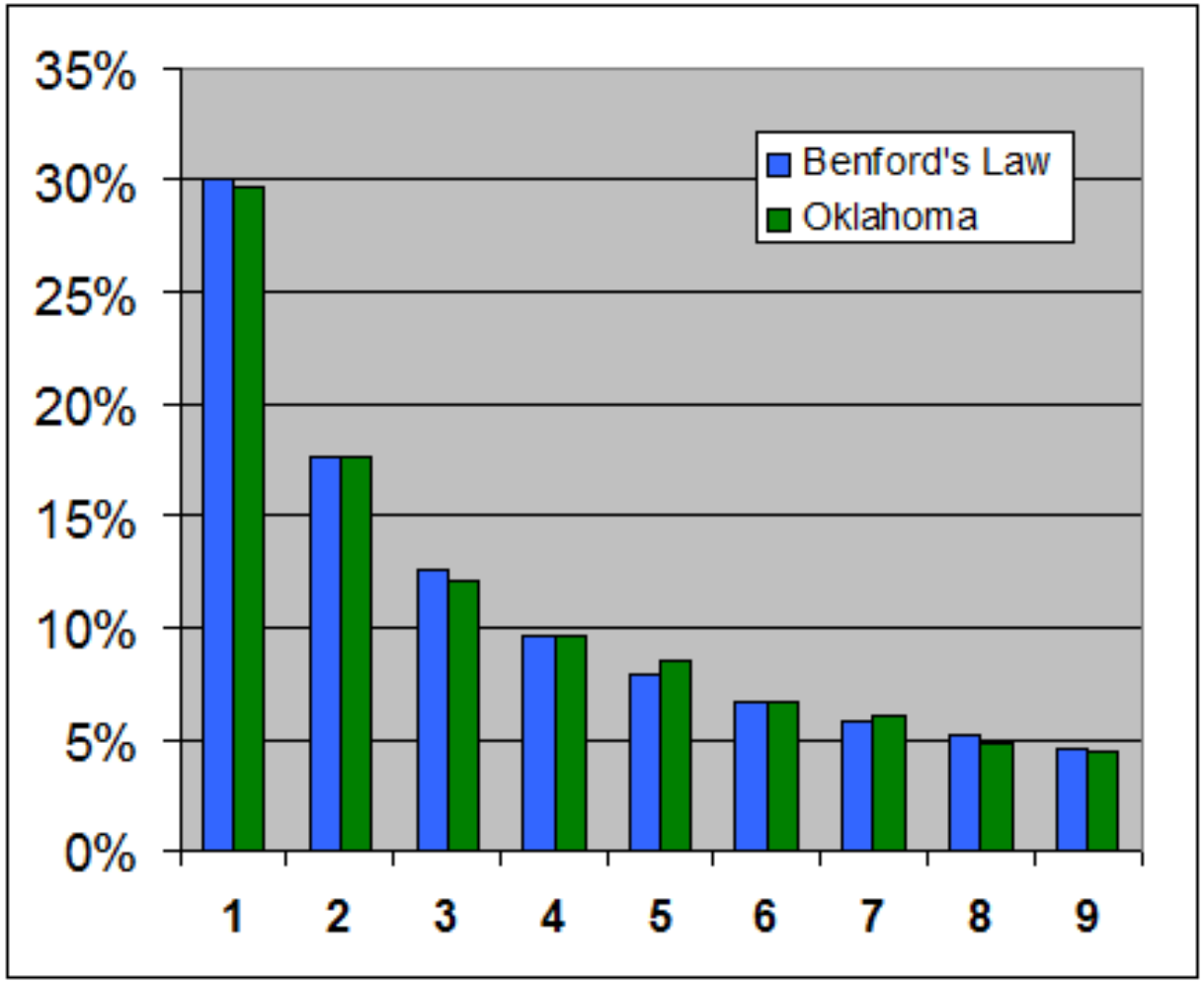
Oklahoma

Oklahoma overall 1st digit configuration:

Superb compliance with Benford!

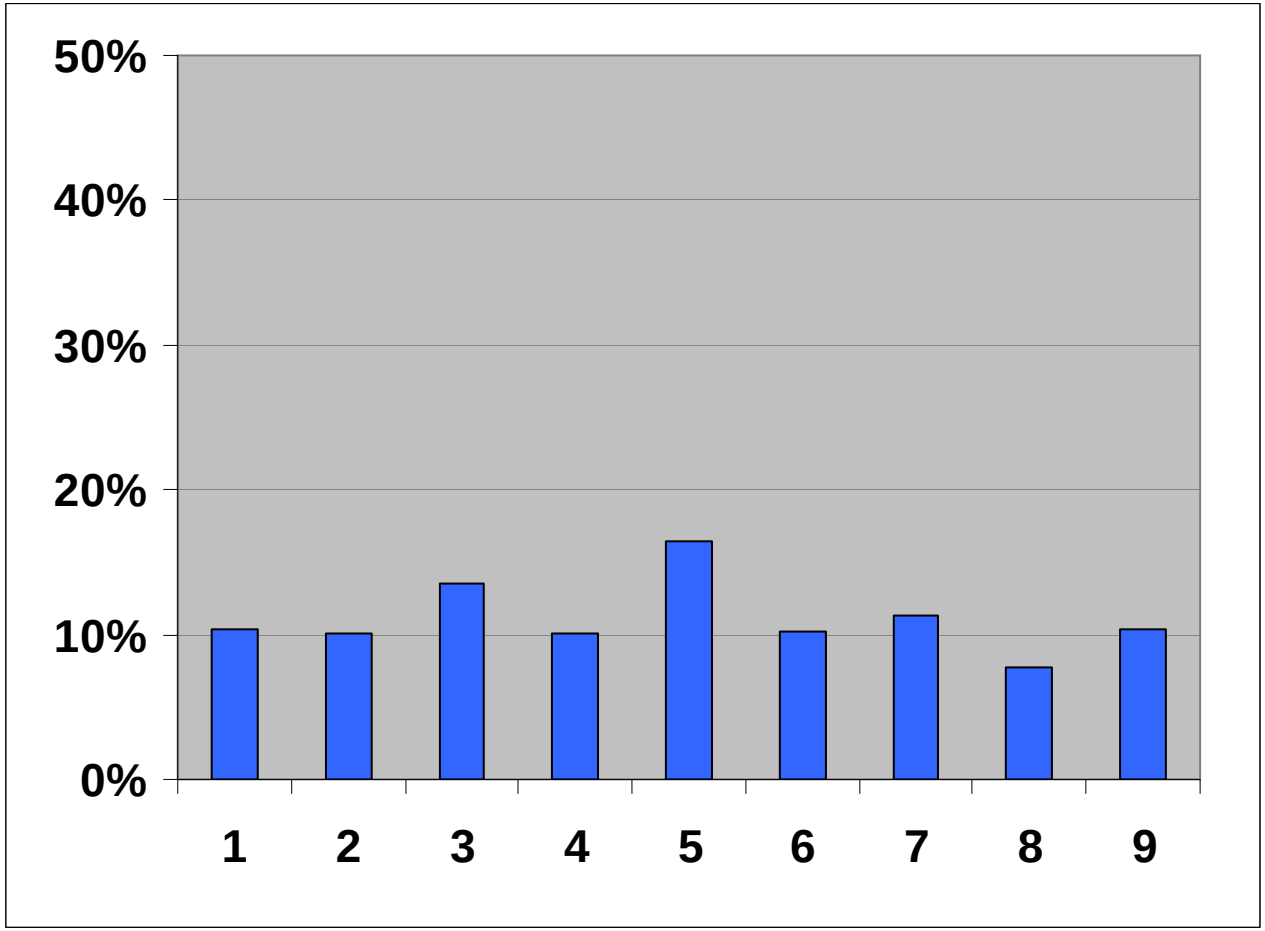
Benford	Oklahoma
30.1%	**29.7%**
17.6%	**17.7%**
12.5%	**12.1%**
9.7%	**9.7%**
7.9%	**8.5%**
6.7%	**6.6%**
5.8%	**6.1%**
5.1%	**4.9%**
4.6%	**4.5%**

Oklahoma overall 1st digit configuration:

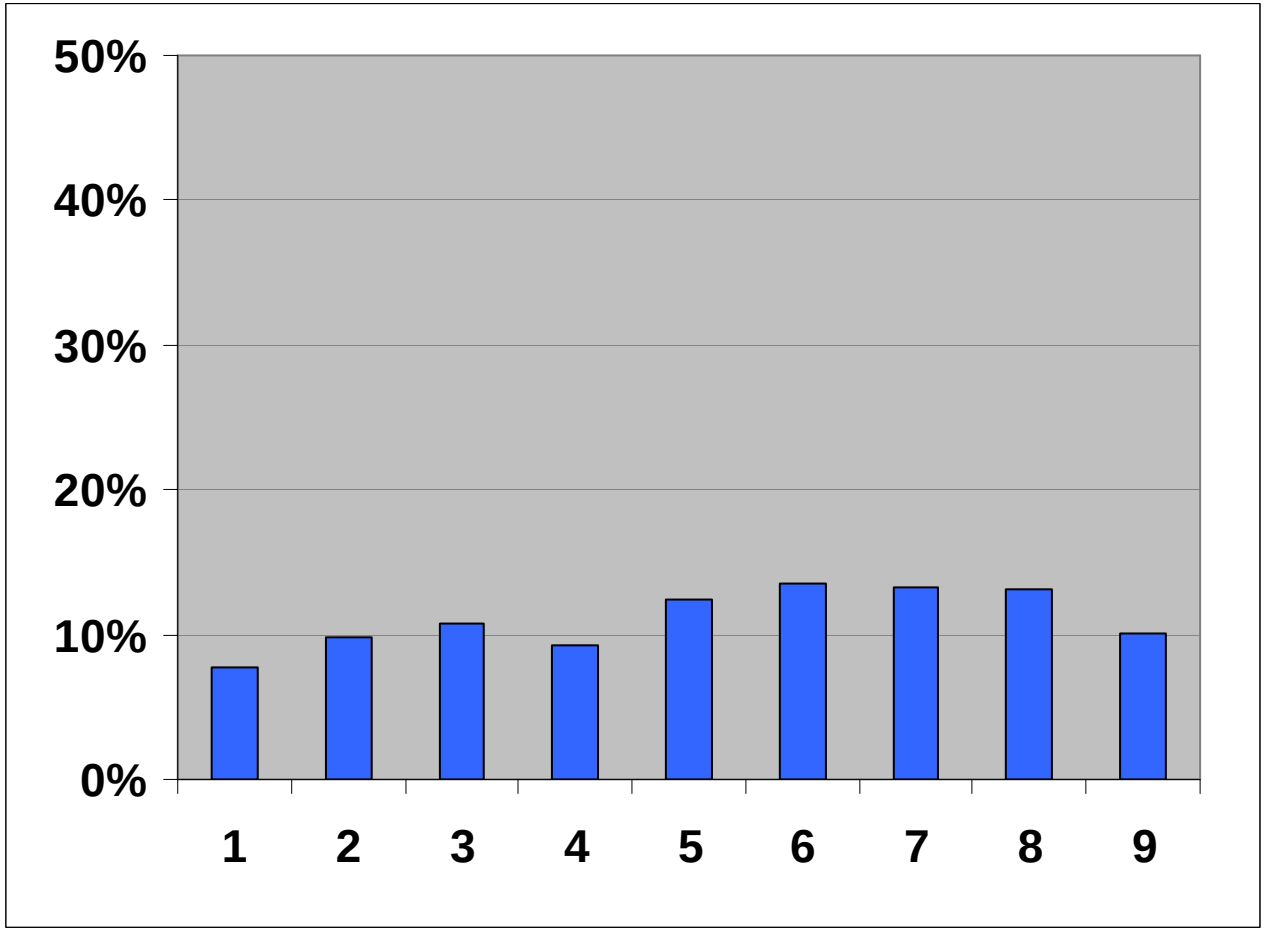


Yet… on local sub-intervals
it develops from left to right:

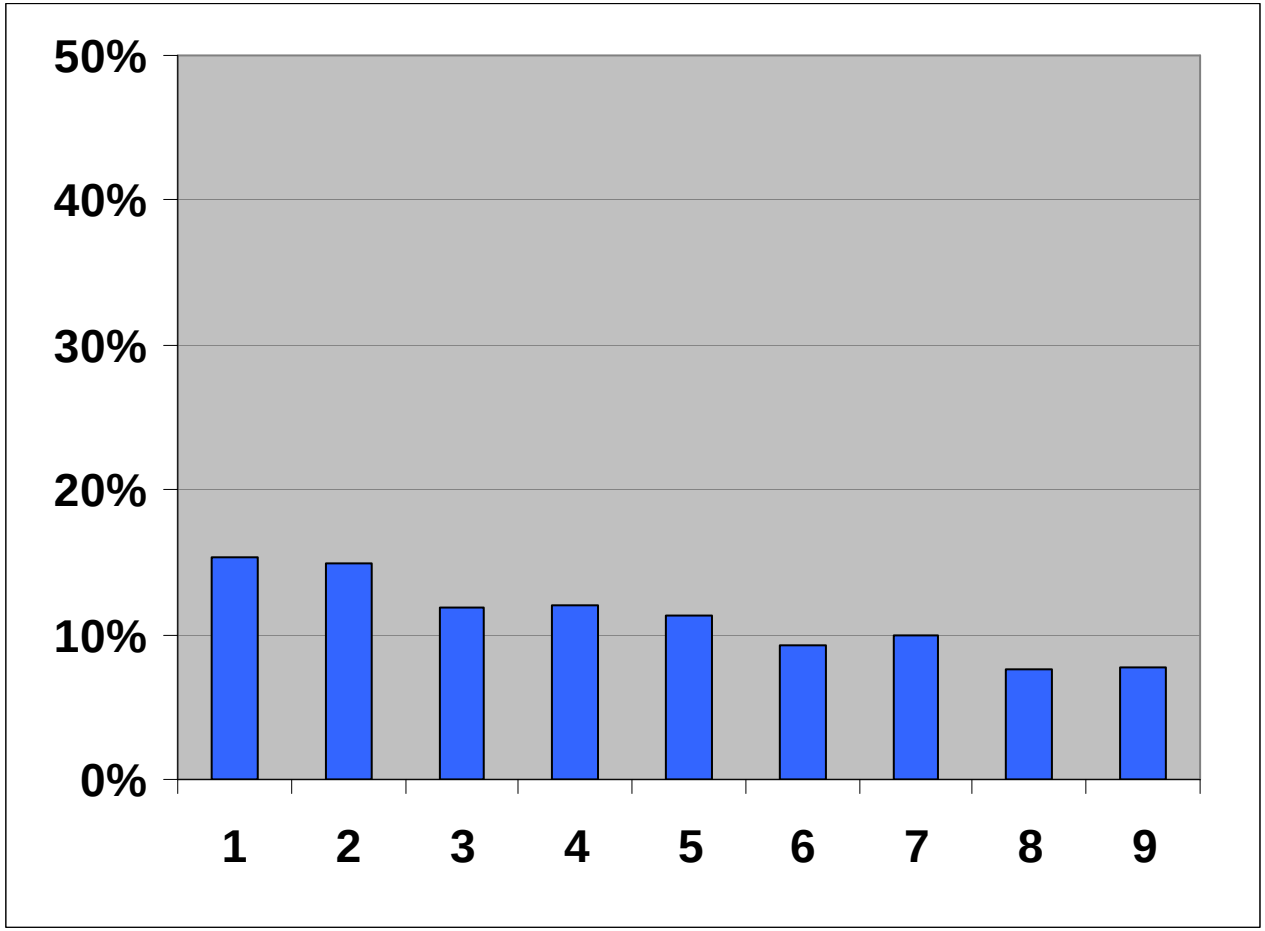
$0.1 – $1



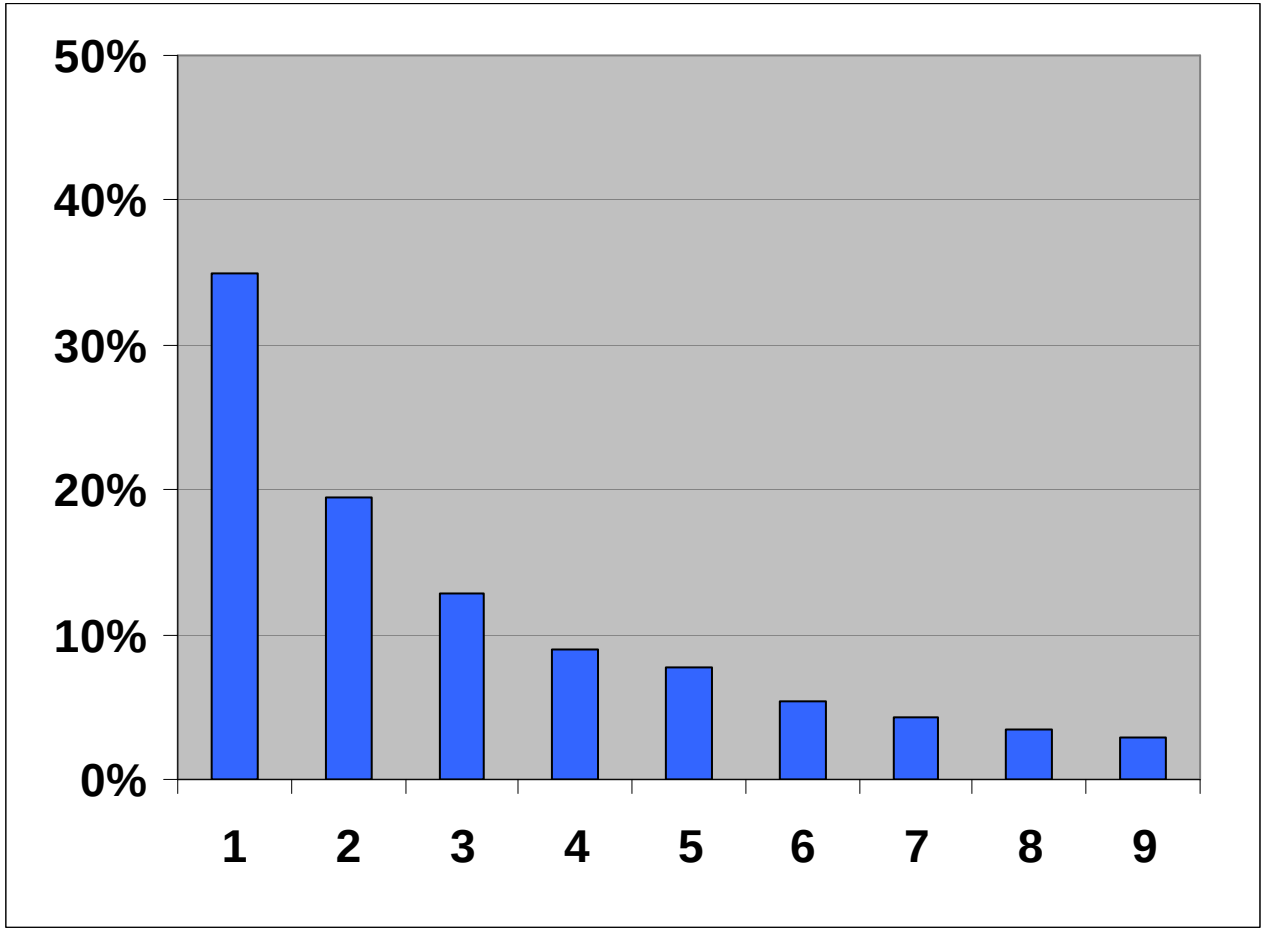
$1 – $10



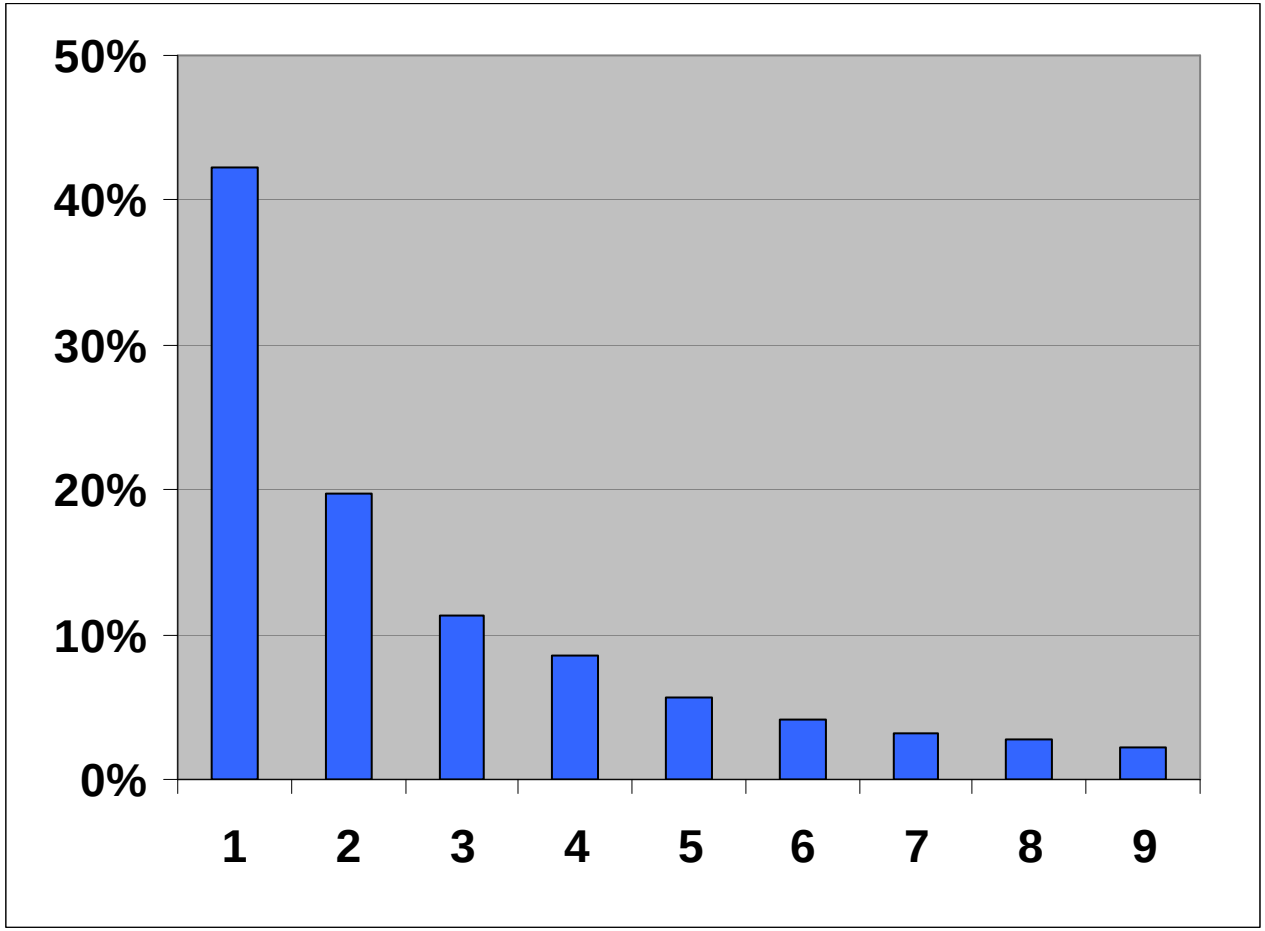
$10 – $100



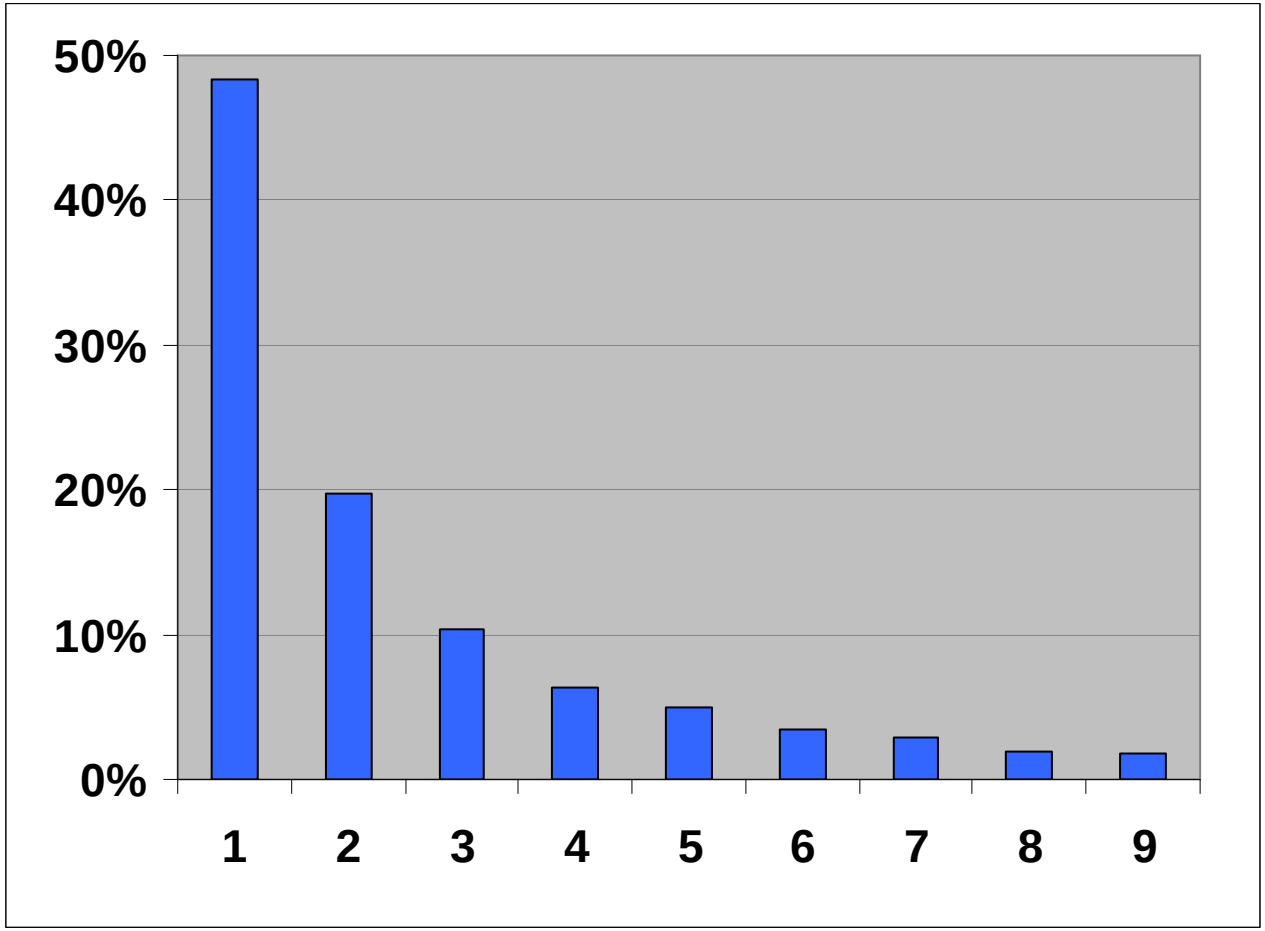
$100 – $1000



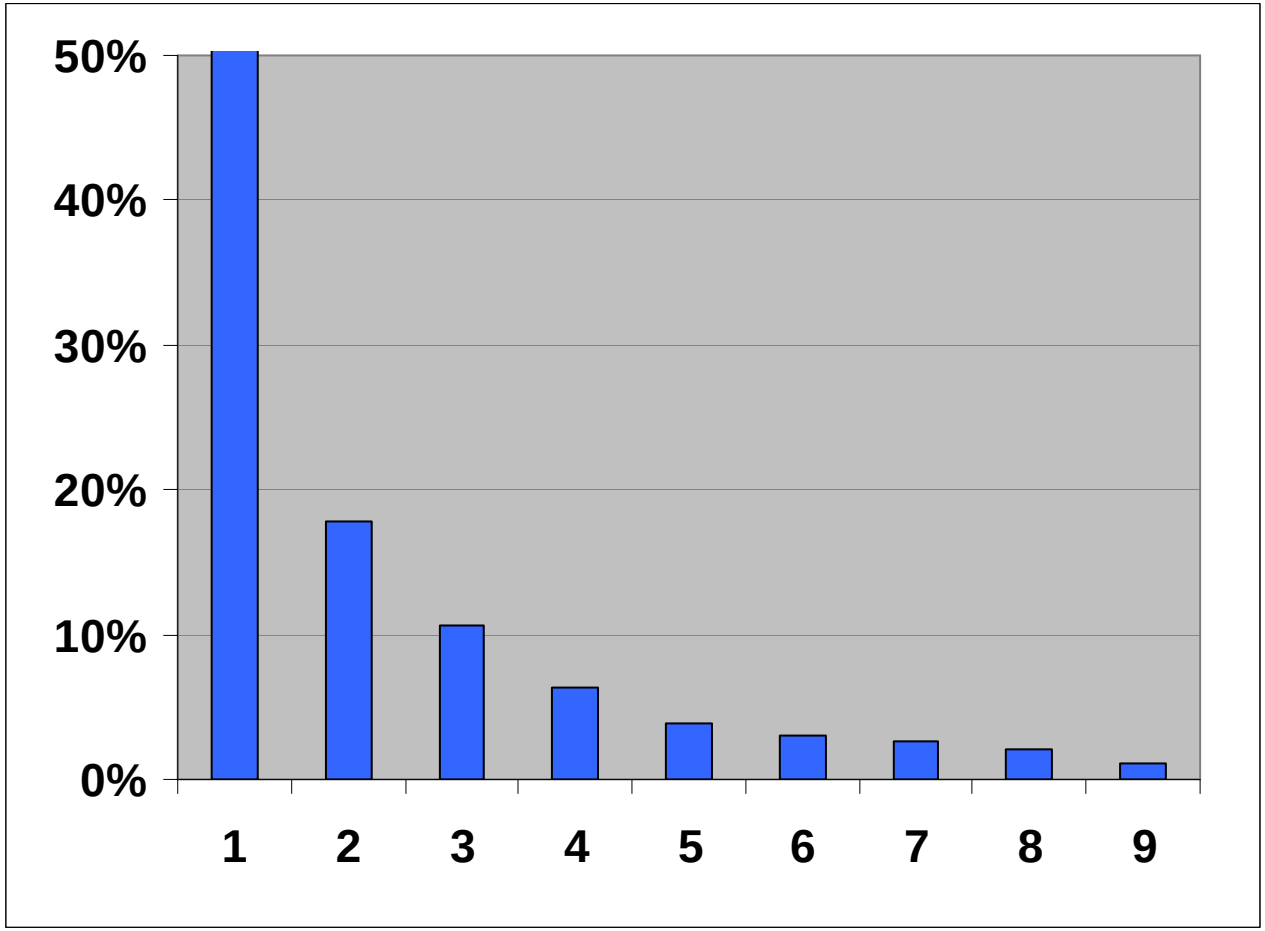
$1000 – $10,000



$10,000 – $100,000



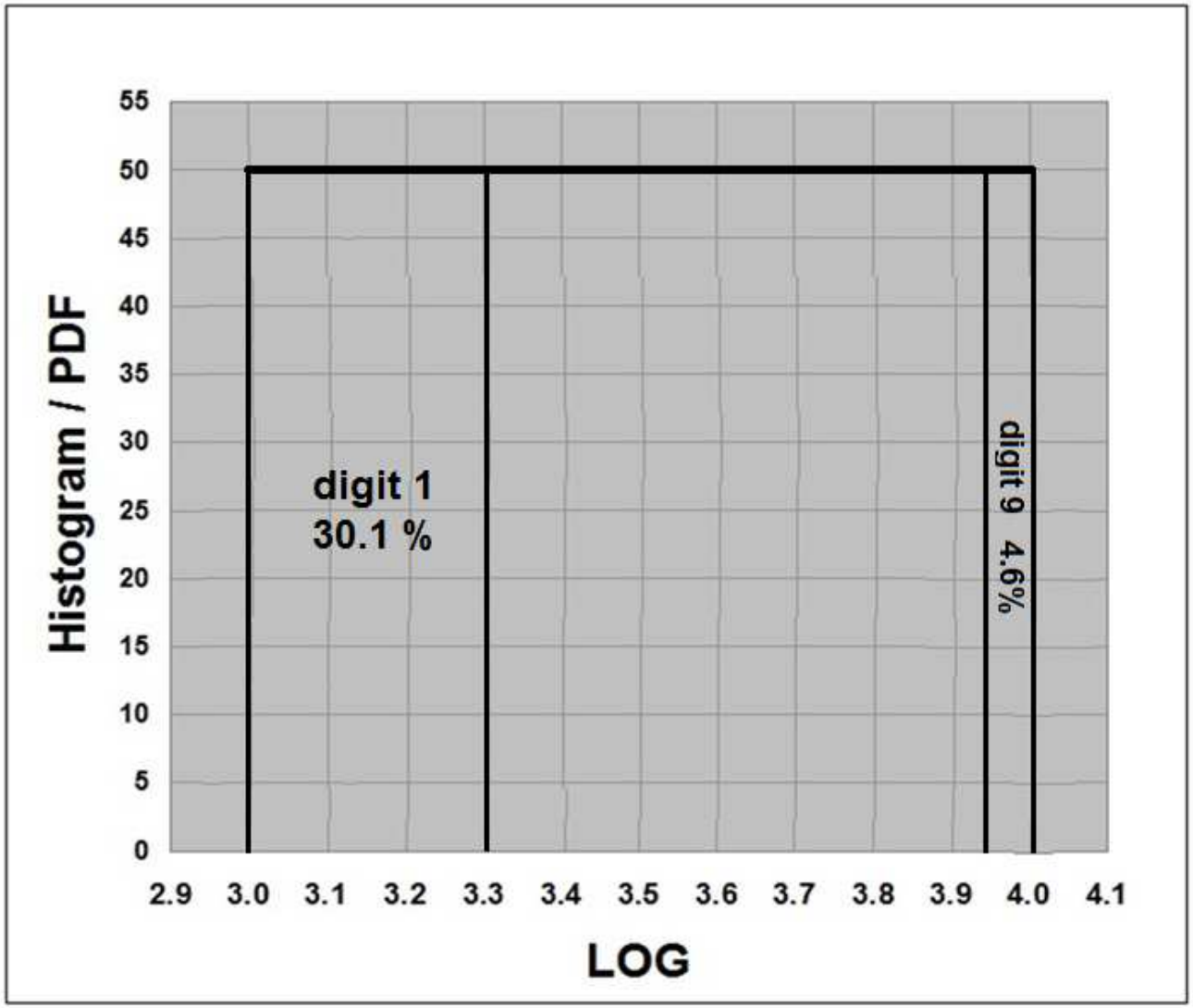
$100,000 – $1,000,000



The **mathematical** Basis of Digital Development Pattern using **mantissa** consideration:

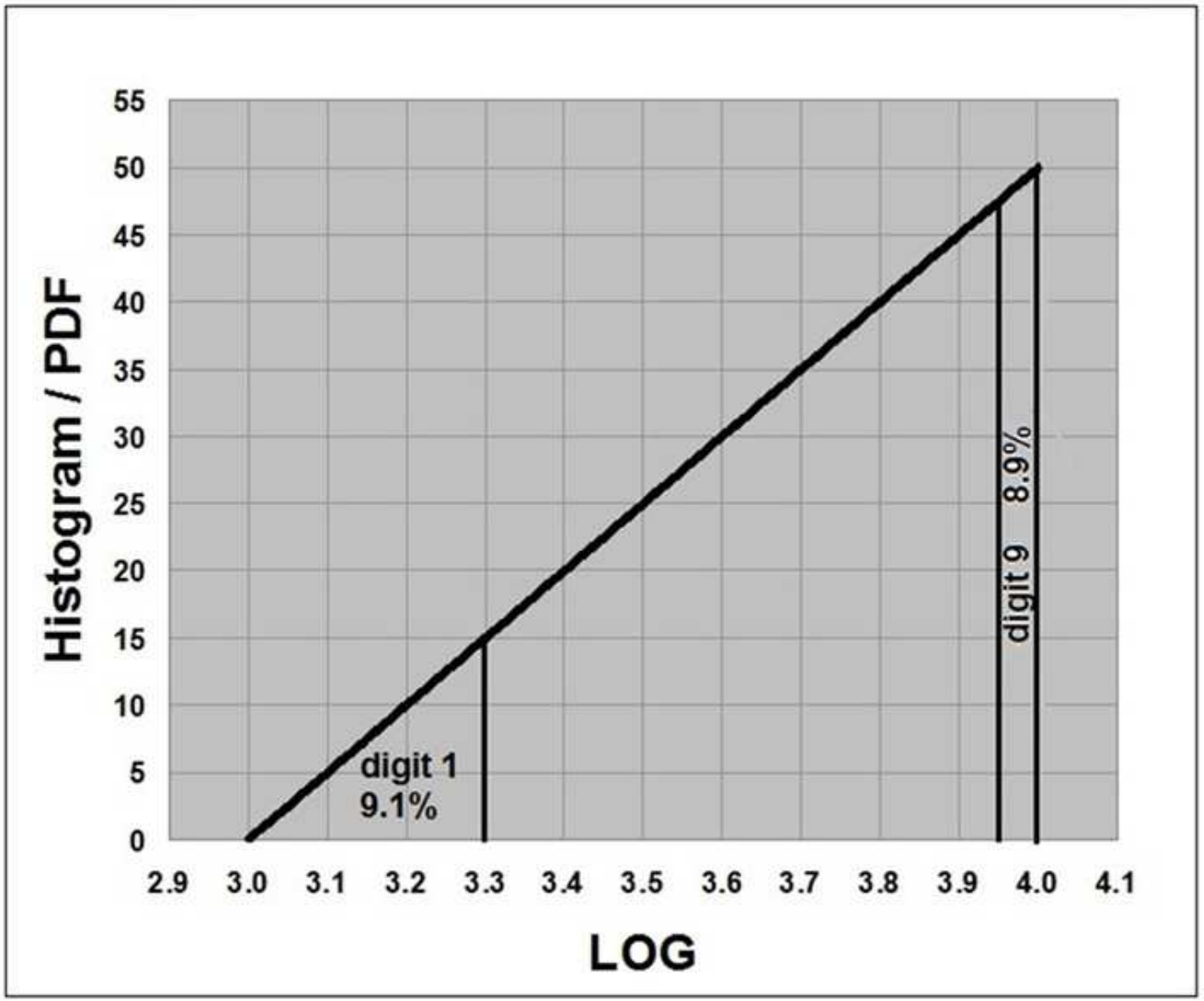
Uniform LOG Yields **Benford**

30.1	17.6	12.5	9.7	7.9	6.7	5.8	5.1	4.6



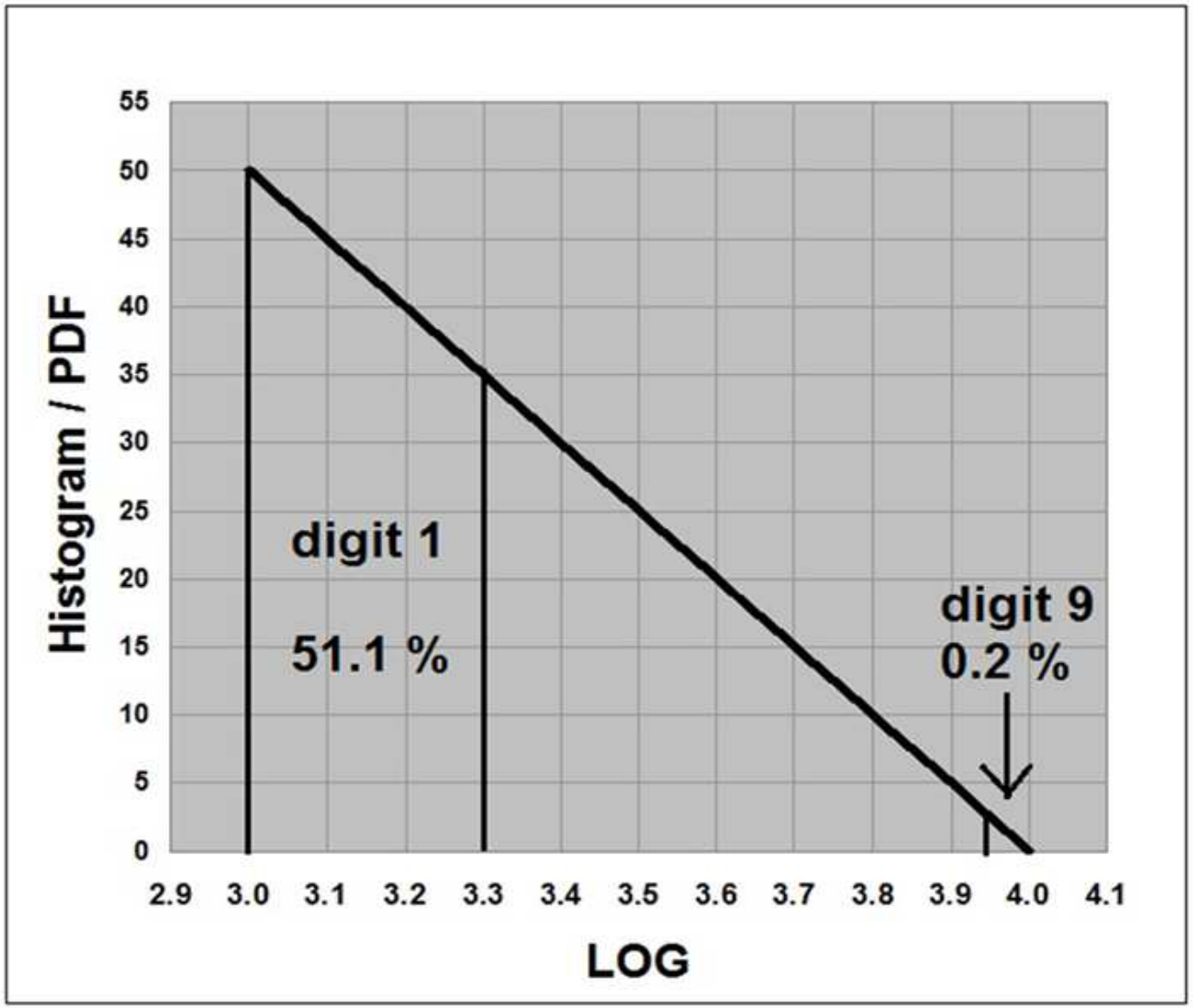
Rising LOG Yields **Near Digital Equality**

9.1	13.7	13.5	12.6	11.7	10.9	10.1	9.5	8.9

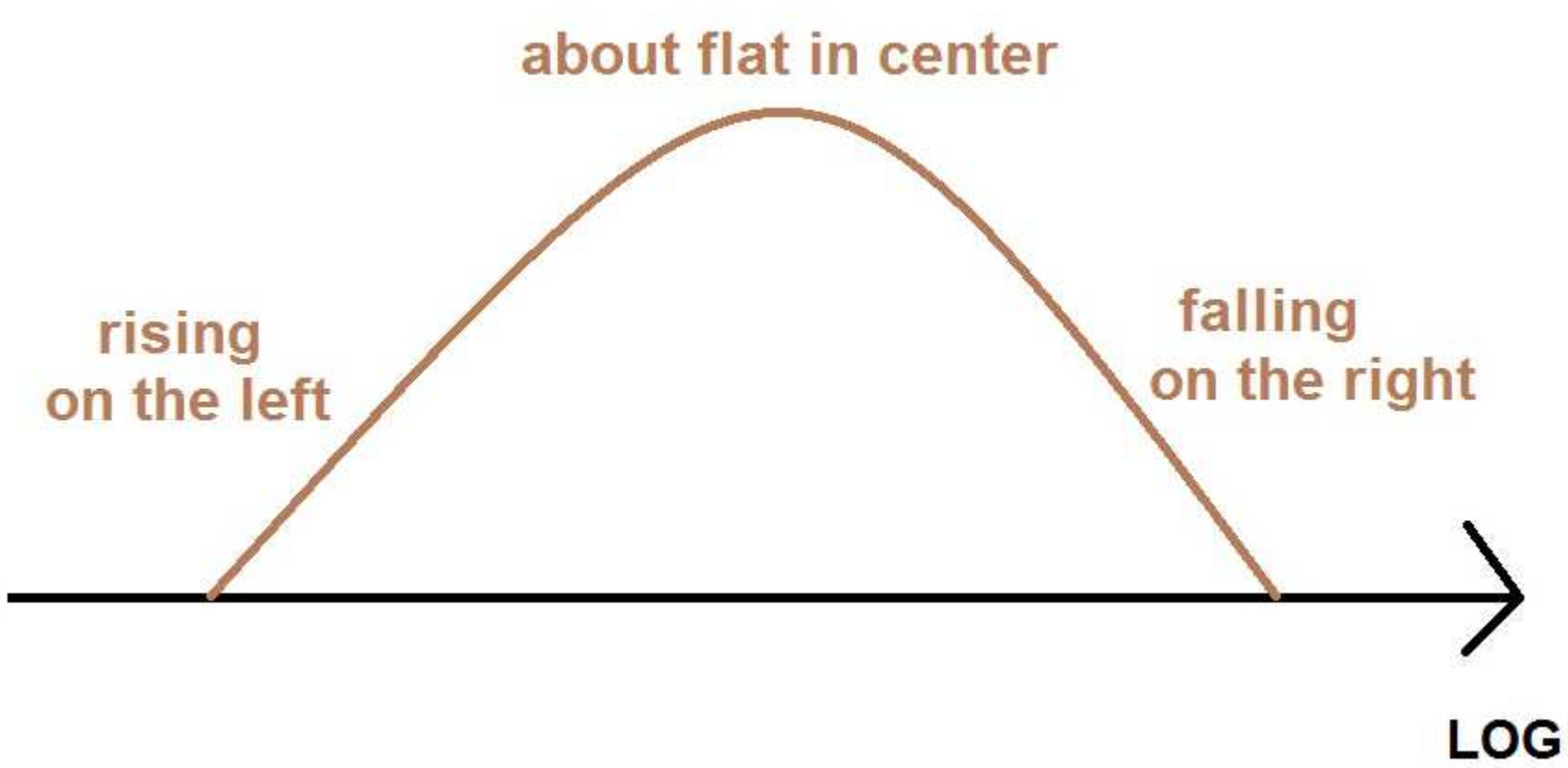


Falling Log Yields **Extreme Digital Skewness**

51.1	21.5	11.5	6.8	4.1	2.5	1.5	0.7	0.2

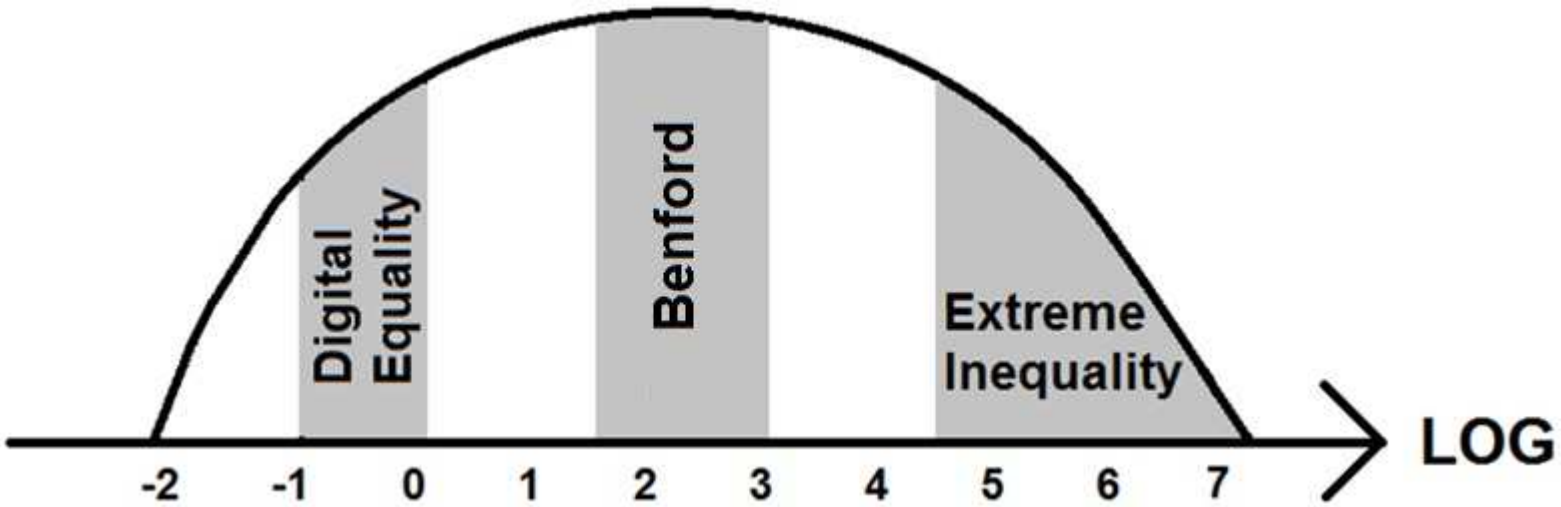


Typical LOG Histograms



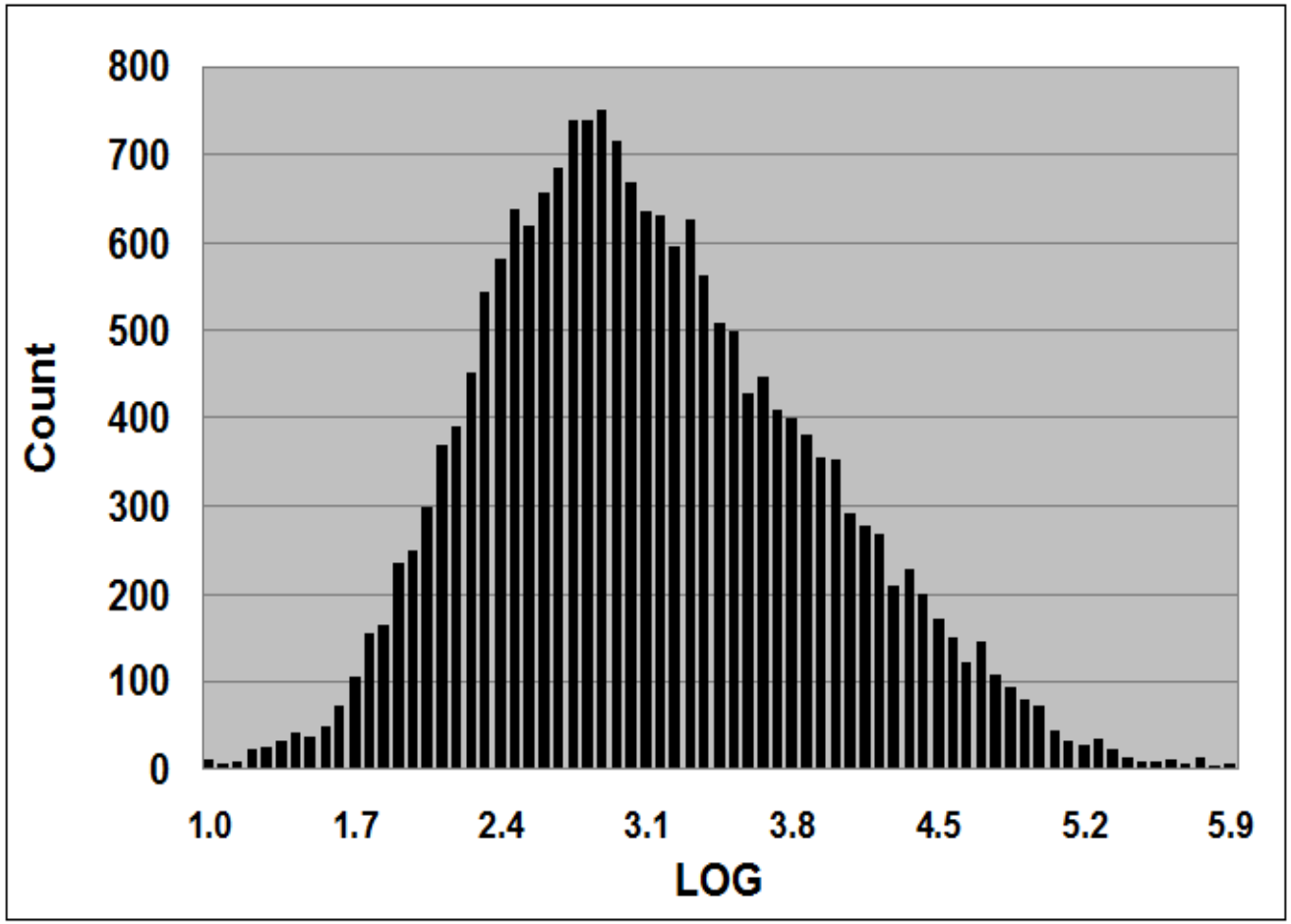
Such differentiated LOG histogram leads to differentiated digital structure.

Log of data and its digital status throughout:

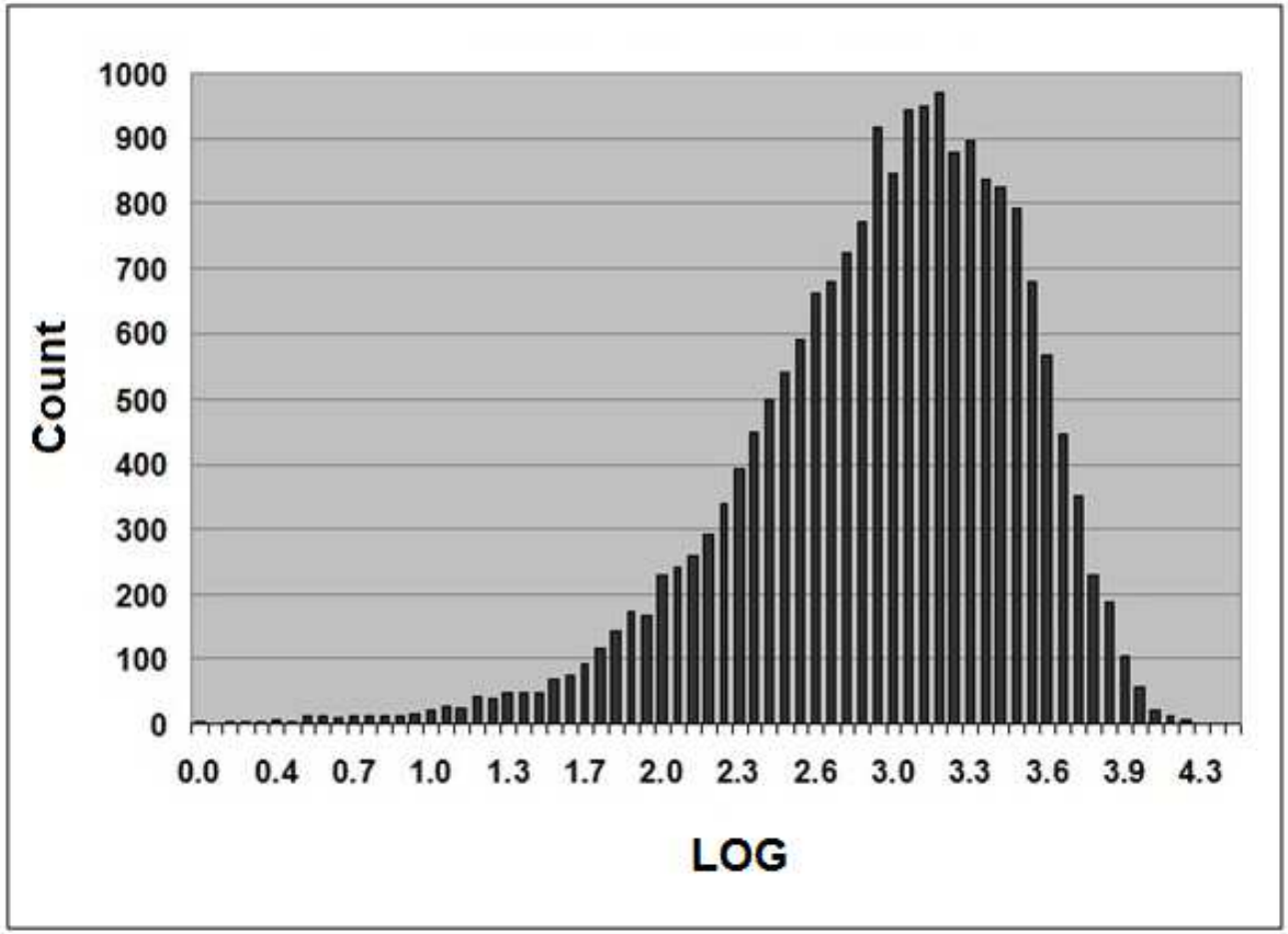


Real-Life examples
of LOG histograms:

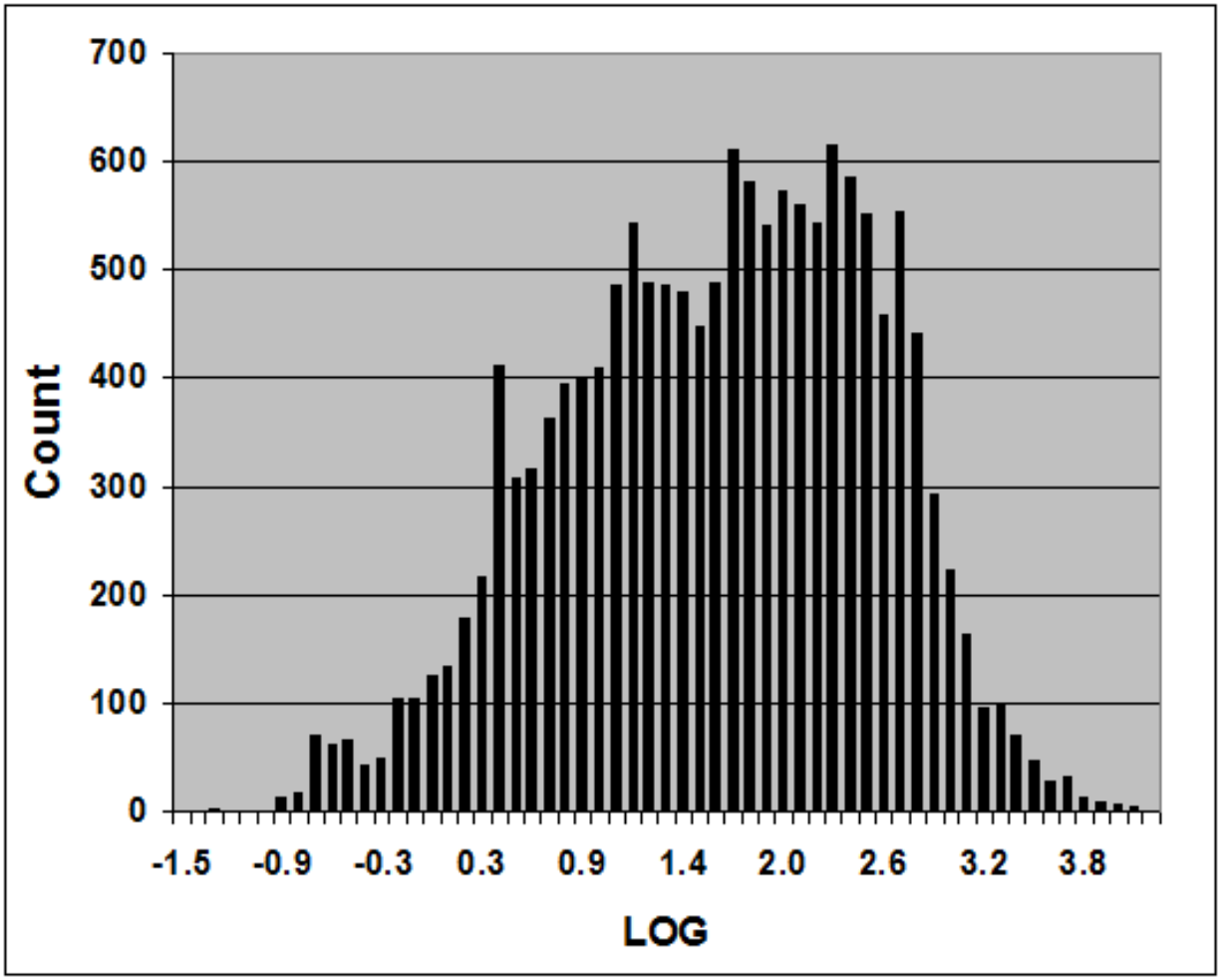
Population of USA Cities and Towns – 2009 Census



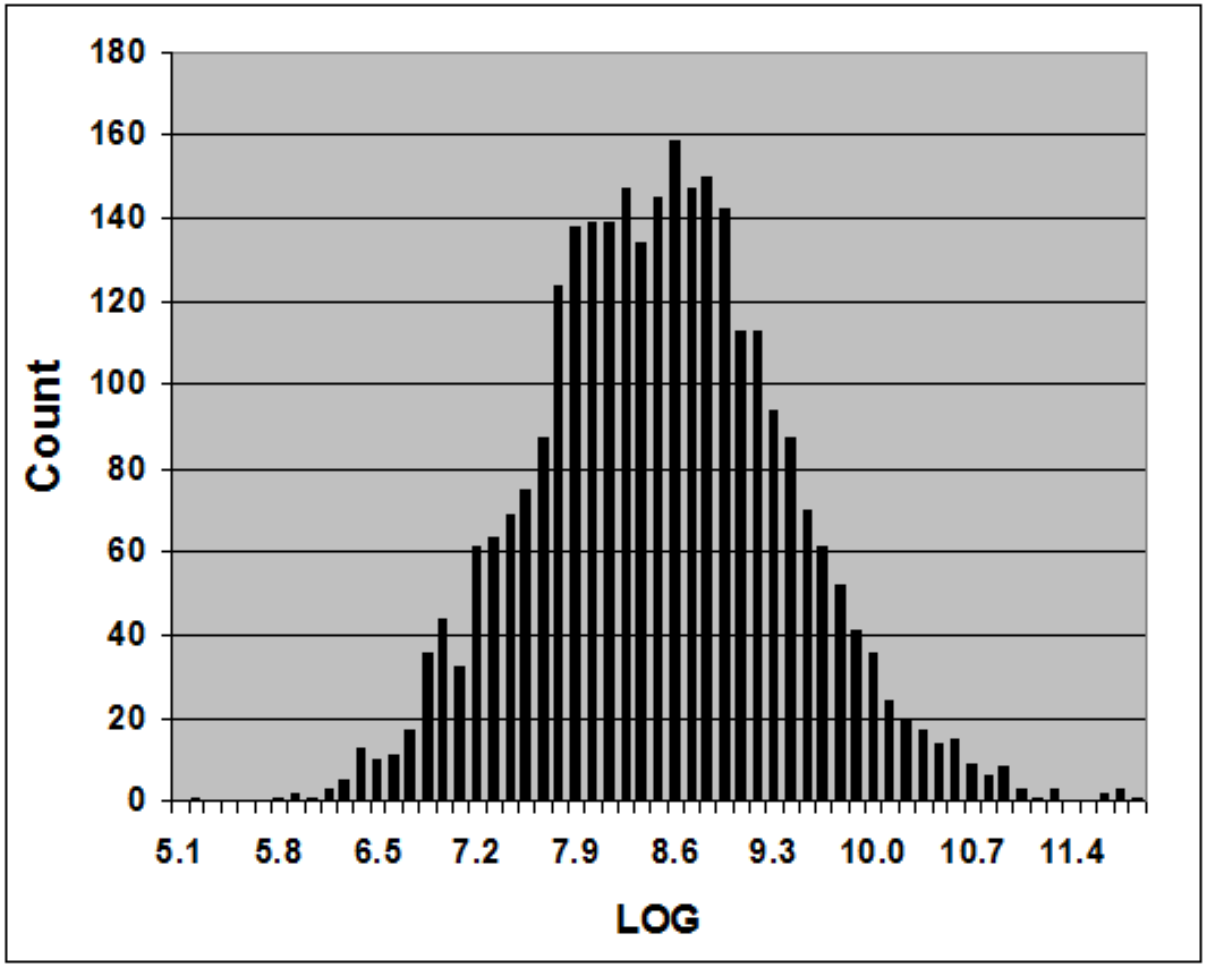
Time between all Earthquake – 2012 Global Data



Canford Audio UK - Catalog



NASDAQ Market Capitalization



Orthodox Data Forensics via Benford's Law

Two Pitfalls

Data itself
is not Benford

Fraudster knows about
Benford's Law and adjusts

Question: How could we detect fraud in these 2 cases?

Answer: Via Digital Development Pattern!

Fake Data Concocted to Appear as if Benford – Unusually Steady, No Develop

From: To:	0.1 1	1 10	10 100	100 1,000	Entire Range
1	29.2%	31.2%	30.6%	29.7%	30.6%
2	16.8%	16.5%	18.5%	19.4%	17.8%
3	12.9%	13.5%	11.7%	13.3%	12.6%
4	9.1%	9.6%	9.3%	9.2%	9.4%
5	8.3%	7.7%	8.3%	7.4%	8.0%
6	7.3%	6.8%	7.1%	6.2%	6.9%
7	5.8%	5.3%	5.5%	5.5%	5.4%
8	6.4%	4.9%	5.1%	5.2%	5.1%
9	4.2%	4.5%	3.9%	4.1%	4.2%

Fake Data Concocted to Appear as if Benford – Zigzag and Chaotic 'Develop'

From: To:	10 100	100 1,000	1,000 10,000	10,000 100,000	Entire Range
1	21.3%	44.1%	16.7%	33.6%	30.8%
2	18.4%	20.1%	15.2%	21.8%	18.6%
3	11.1%	11.3%	14.1%	13.7%	12.6%
4	10.7%	6.2%	12.0%	8.8%	9.1%
5	9.6%	4.5%	11.1%	8.3%	7.9%
6	8.6%	4.5%	8.4%	6.1%	6.5%
7	7.9%	3.9%	7.3%	3.1%	5.3%
8	6.3%	2.9%	8.2%	2.5%	5.0%
9	6.1%	2.7%	7.0%	2.1%	4.4%

Fake Data Concocted to Appear as if Benford – Inverse Development

From: To:	1 10	10 100	100 1,000	1,000 10,000	Entire Range
1	38.9%	31.5%	26.2%	12.3%	29.1%
2	22.4%	18.5%	14.7%	11.8%	17.4%
3	12.3%	14.1%	13.5%	11.3%	13.2%
4	9.3%	8.5%	12.1%	10.5%	9.9%
5	6.2%	7.3%	9.1%	11.4%	8.1%
6	4.1%	6.8%	8.2%	11.2%	7.2%
7	3.3%	5.9%	6.0%	11.1%	6.0%
8	2.3%	4.2%	5.2%	10.1%	4.8%
9	1.2%	3.2%	5.0%	10.3%	4.2%

Development Compliance Test

An Exact Measure of Development

Via

A Measure of ‘Over-Skewness’

ES12 ≡ A measure of skewness over and above the normal Benford skewed configuration

ES12 ≡ Excess Sum of digits 1 and 2

ES12 ≡ [observed digits 1 & 2] **–** [Benford allocation for 1 and 2]

ES12 ≡ [observed digits 1 & 2] **–** [**30.1% + 17.6%**]

ES12 ≡ [observed digits 1 & 2] **–** [**47.7%**]

For example, if digits within sub-interval are:

{44.1%, 19.6%, 11.5%, 7.6%, 5.3%, 4.3%, 2.7%, 3.2%, 1.8%}

ES12 ≡ [observed digits 1 & 2] – [47.7%]

ES12 = [44.1% + 19.6%] – [47.7%]

ES12 = [63.7%] – [47.7%]

ES12 = +16.0%

- **ES12 is negative** $\Rightarrow$ **≈ digital equality**

0 **ES12 is ≈ zero** $\Rightarrow$ **≈ Benford**

\+ **ES12 is positive** $\Rightarrow$ **extreme inequality**

Canford Catalog

From	0.1	1	10	100	1,000
To	1	10	100	1,000	10,000
Digit 1 :	6.7%	15.7%	29.5%	33.6%	63.9%
Digit 2 :	21.2%	19.9%	16.6%	20.6%	18.8%
Digit 3 :	14.5%	13.4%	10.7%	13.7%	7.6%
Digit 4 :	6.7%	11.6%	9.6%	8.8%	4.8%
Digit 5 :	7.7%	8.8%	9.3%	8.5%	1.7%
Digit 6 :	8.7%	9.4%	7.3%	6.1%	1.5%
Digit 7 :	17.6%	8.6%	6.7%	3.8%	0.9%
Digit 8 :	6.7%	5.8%	5.2%	2.4%	0.8%
Digit 9 :	10.2%	6.8%	5.2%	2.5%	0.0%
Data Proportion	4.0%	21.0%	35.7%	34.8%	4.4%
ES12:	-19.8%	-12.1%	-1.6%	6.6%	35.0%

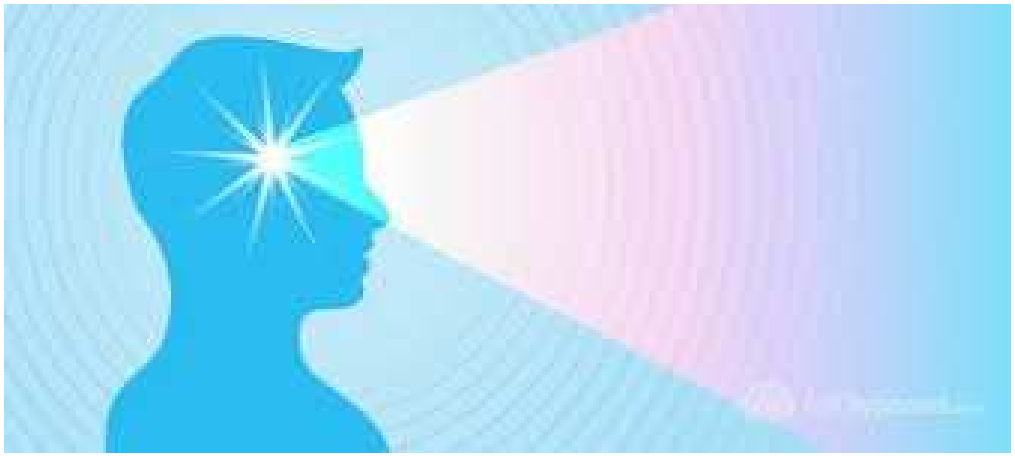
Nasdaq Market Capitalization

From	1,000,000	10,000,000	100,000,000	1,000,000,000	10,000,000,000
To	10,000,000	100,000,000	1,000,000,000	10,000,000,000	100,000,000,000
Digit 1 :	7.1%	18.7%	29.6%	44.1%	46.6%
Digit 2 :	13.5%	13.3%	16.7%	19.6%	18.6%
Digit 3 :	5.0%	11.8%	14.1%	11.5%	15.3%
Digit 4 :	9.2%	10.0%	9.9%	7.6%	6.8%
Digit 5 :	14.2%	11.6%	8.3%	5.3%	1.7%
Digit 6 :	14.9%	9.9%	6.3%	4.3%	4.2%
Digit 7 :	17.0%	8.6%	5.9%	2.7%	4.2%
Digit 8 :	11.3%	8.7%	5.3%	3.2%	1.7%
Digit 9 :	7.8%	7.4%	3.8%	1.8%	0.8%
Data Proportion	4.9%	25.9%	42.8%	21.7%	4.1%
ES12:	-27.1%	-15.7%	-1.4%	16.0%	17.5%

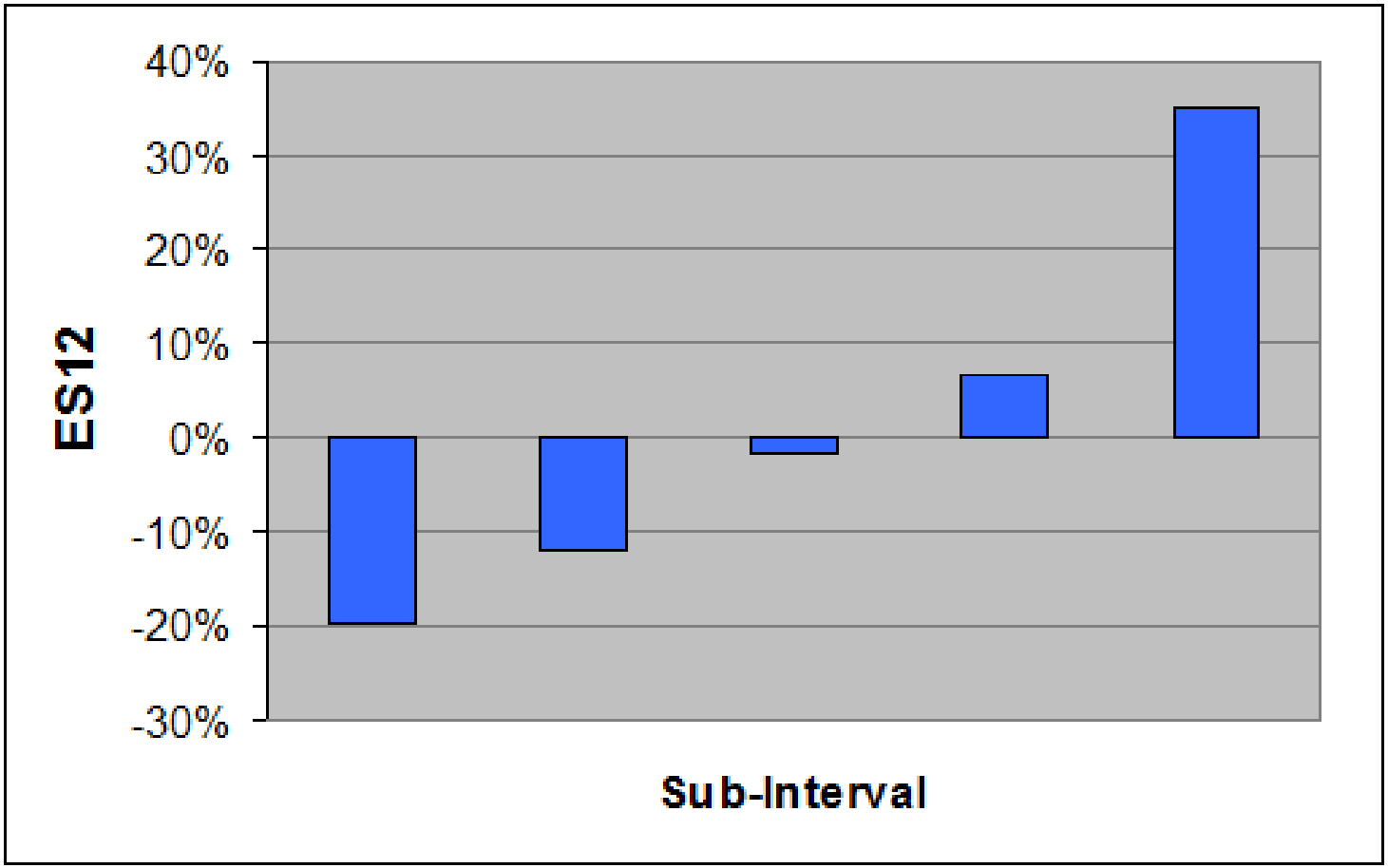
Oklahoma State Expenses

From	0.1	1	10	100	1,000	10,000	100,000
To	1	10	100	1,000	10,000	100,000	1,000,000
Digit 1 :	10.3%	7.7%	15.3%	35.0%	42.3%	48.4%	52.5%
Digit 2 :	10.1%	9.9%	14.9%	19.5%	19.7%	19.8%	17.8%
Digit 3 :	13.5%	10.8%	11.9%	12.8%	11.3%	10.4%	10.6%
Digit 4 :	10.0%	9.3%	12.0%	9.0%	8.6%	6.4%	6.3%
Digit 5 :	16.4%	12.5%	11.3%	7.7%	5.7%	5.0%	3.9%
Digit 6 :	10.2%	13.5%	9.2%	5.4%	4.1%	3.4%	3.1%
Digit 7 :	11.4%	13.3%	9.9%	4.3%	3.2%	2.9%	2.6%
Digit 8 :	7.8%	13.1%	7.7%	3.4%	2.8%	2.0%	2.0%
Digit 9 :	10.3%	10.0%	7.8%	3.0%	2.2%	1.7%	1.1%
Data Proportion	0.2%	4.0%	30.8%	43.5%	16.8%	4.1%	0.6%
ES12:	-27.2%	-30.1%	-17.5%	6.8%	14.3%	20.5%	22.6%

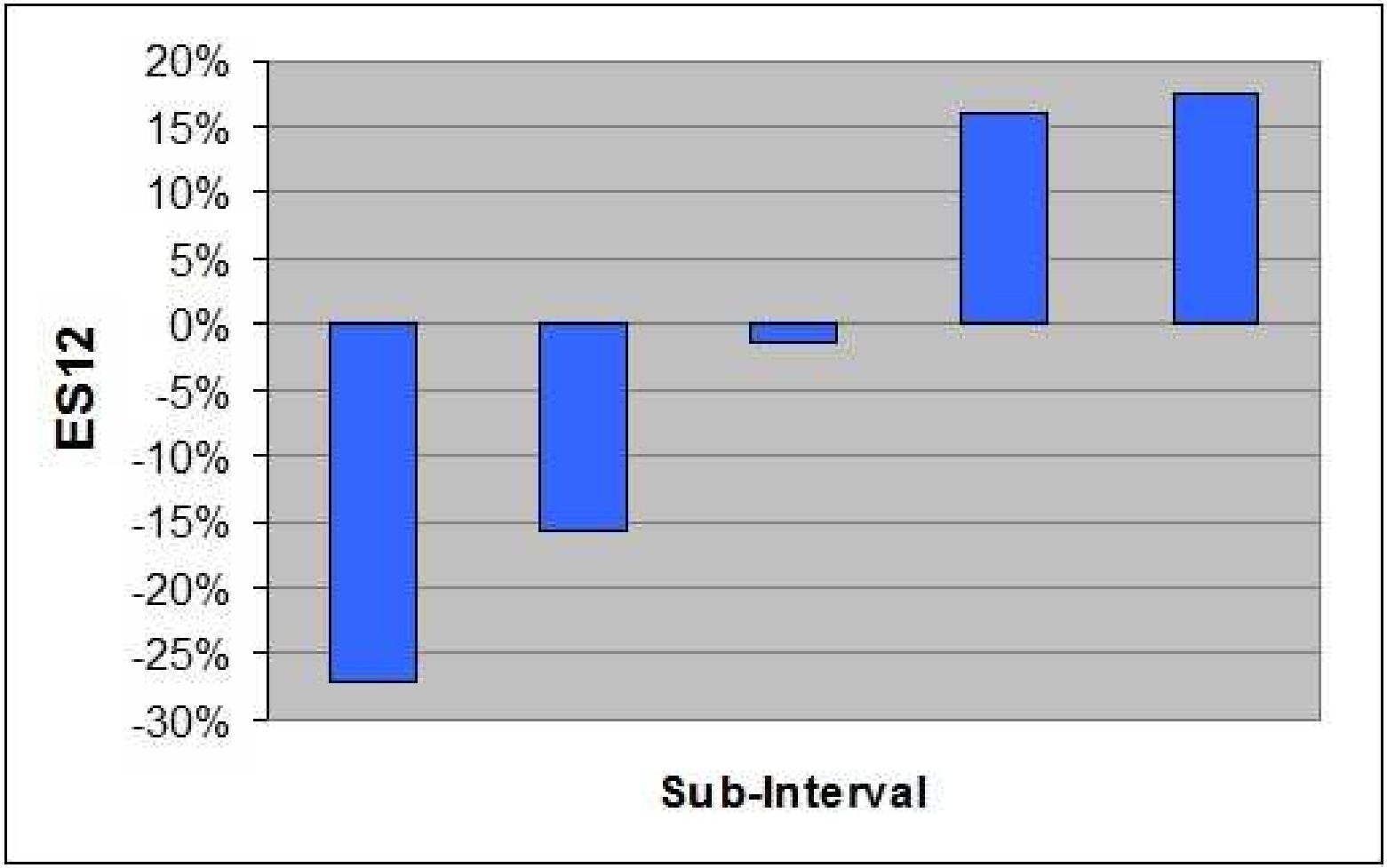
Now we can **visualize** development clearly!



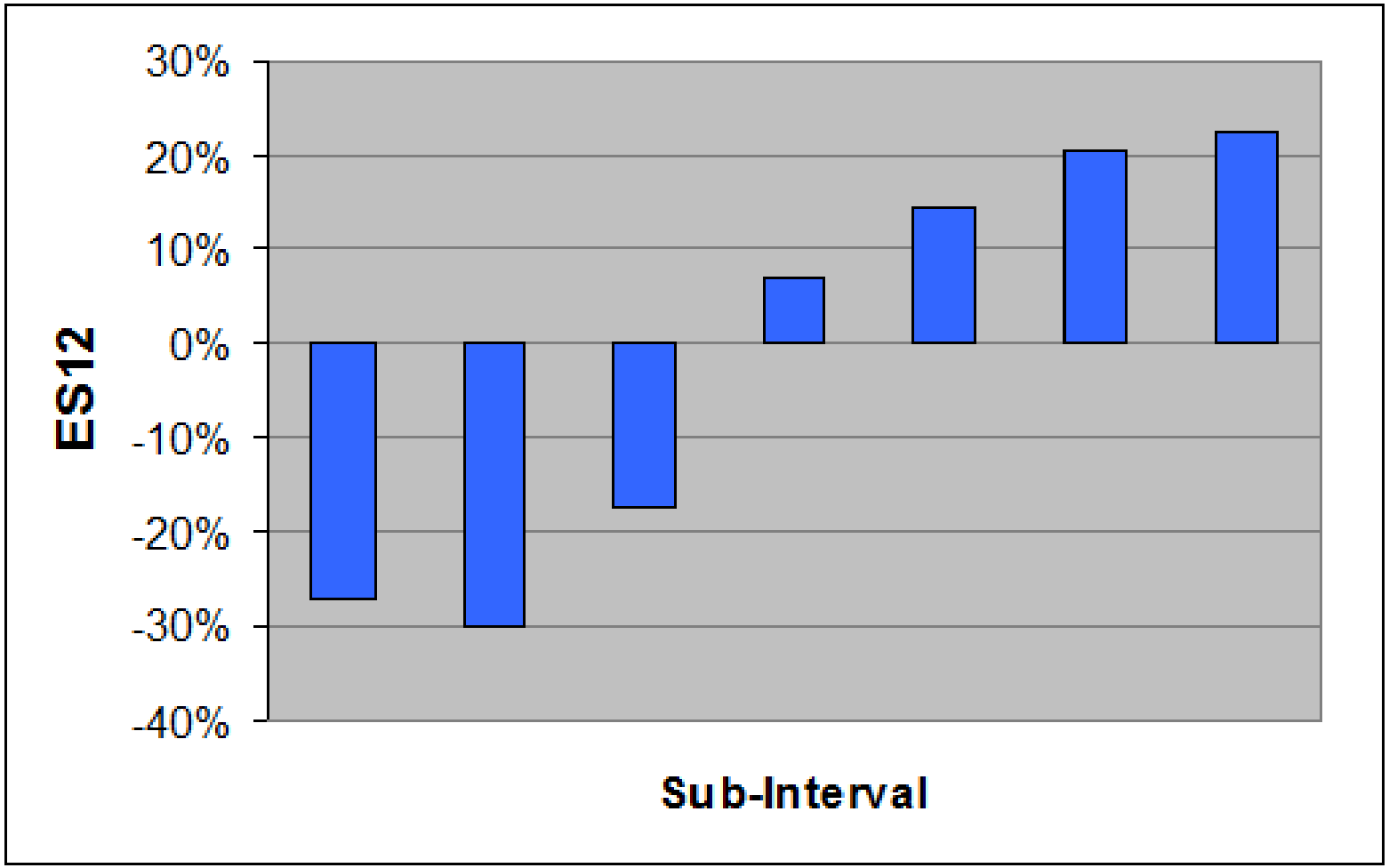
Canford Catalog



Nasdaq Market Capitalization



Oklahoma State Expenses



Question:

What is the minimum level of development so that data does not raise suspicion?

Namely: What is the threshold or cutoff point differentiating ***'development-compliance'*** from ***'development-non-compliance'***?

Answer:

Simple Linear Regression is performed on **ES12**, requiring left-most point to be **below -15%** and right-most point to be **above +15%**, in order to declare data as development-compliance.

Firstly, any sub-interval with less than **0.1%** of overall data is eliminated, together with any sub-interval further out.

$\mathbf{N_j}$ ≡ Rank number of sub-interval
J = 1, 2, 3, … , L

L being the last on the right.
1 being the first on the left.

$ES12_j$ being the dependent variable.

N_j being the independent variable.

Regression is modeled on:

$$ES12_j = mN_j + b$$

where **m** is the slope and **b** is the intercept.

ES12_left ≡ m**1** + b

ES12_right ≡ m**L** + b

<u>Development Test:</u>

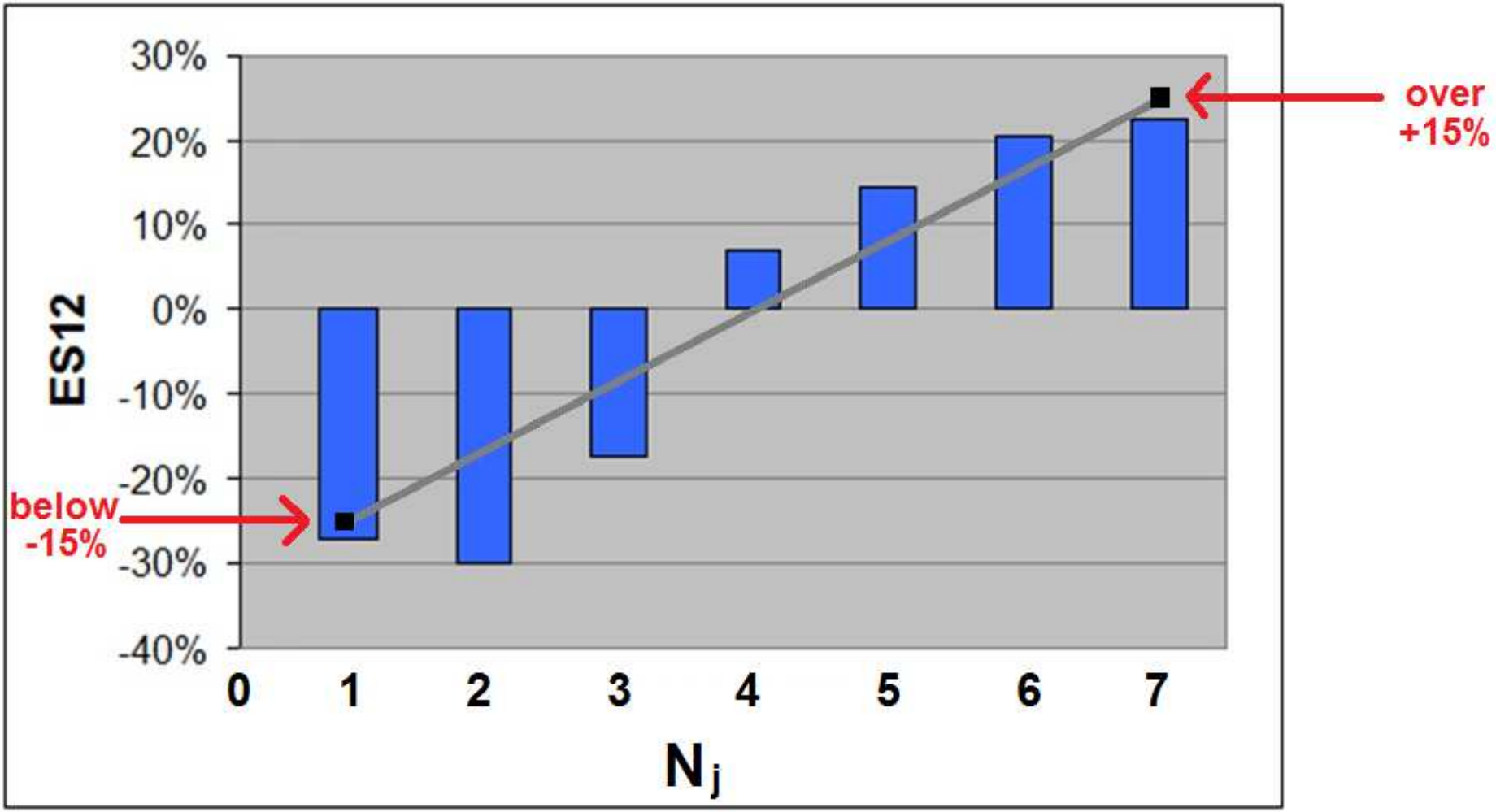
Reported data passes honesty test if the following two conditions are true:

ES12_left < -15%

AND

ES12_right > +15%

Oklahoma State Expenses easily passes Development Test!



The constraint of this test constitutes the **EMPIRICAL LAW OF DEVELOPMENT**, true across practically all random and statistical data types, Benford as well as non-Benford ones.

Failure to pass the test strongly hints at the possibility of fraudulent activity in data reporting.

Remarkably, **Digital Development Pattern** is found even in non-Benford data!

Remarkably, **Digital Development Pattern** is found even in data with low order of magnitude!

Hence…

Digital Development Pattern is **even more prevalent** than Benford's Law itself!

!!!

Hence digital forensics can still be perform on data sets which are not Benford to begin with!

Just about the ONLY EXCEPTIONS are exponential growth series and the k/x distribution which come with flat and horizontal LOG histogram and steady digital configuration of the Benford type throughout.

funnily,
strangely,
surprisingly,
shockingly,

k/x IS WITHOUT DEVELOPMENT!

It’s steady!

Limited Applicability of k/x Distribution:

In spite of its Benfordness & uniqueness, the k/x distribution is NOT the appropriate model for typical everyday data sets!

!!!

k/x distribution has very **limited relevance** to real life empirical application of Benford's Law.

Theoretically though k/x plays a crucial role.

Patent has been granted for the above **forensic algorithm** by the US Patent Office in 2015.

Patent # 9,058,285.

The United States of America

The Director of the United States Patent and Trademark Office

Has received an application for a patent for a new and useful invention. The title and description of the invention are enclosed. The requirements of law have been complied with, and it has been determined that a patent on the invention shall be granted under the law.

Therefore, this

United States Patent

Grants to the person(s) having title to this patent the right to exclude others from making, using, offering for sale, or selling the invention throughout the United States of America or importing the invention into the United States of America, and if the invention is a process, of the right to exclude others from using, offering for sale or selling throughout the United States of America, or importing into the United States of America, products made by that process, for the term set forth in 35 U.S.C. 154(a)(2) or (c)(1), subject to the payment of maintenance fees as provided by 35 U.S.C. 41(b). See the Maintenance Fee Notice on the inside of the cover.

Michelle K. Lee

Director of the United States Patent and Trademark Office

US009058285B2

(12) **United States Patent**
Kossovsky

(10) **Patent No.: US 9,058,285 B2**
(45) **Date of Patent: Jun. 16, 2015**

(54) **METHOD AND SYSTEM FOR FORENSIC DATA ANALYSIS IN FRAUD DETECTION EMPLOYING A DIGITAL PATTERN MORE PREVALENT THAN BENFORD'S LAW**

(76) Inventor: **Alex Ely Kossovsky**, New York City, NY (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 236 days.

(21) Appl. No.: **13/535,269**

(22) Filed: **Jun. 27, 2012**

(65) **Prior Publication Data**

US 2014/0006468 A1 Jan. 2, 2014

(51) **Int. Cl.**

G06F 17/00	(2006.01)
G06F 17/18	(2006.01)
G06Q 20/40	(2012.01)
G06F 17/10	(2006.01)
G06Q 40/00	(2012.01)
[illegible]	(2006.01)

7,940,989 B2	5/2011	Shi et al.	
8,175,844 B2	5/2012	Nobuyoshi et al.	
2010/0205076 A1 *	8/2010	Parson et al.	705/30

OTHER PUBLICATIONS

Alex E. Kossovsky: "Towards a Better Understanding of the Leading Digits Phenomena (Benford's Law)", Article,2012,Cornell University Library, http://arxiv.org/abs/math/0612627.

Mark J. Nigrini:"I've Got Your Number", Article, May 1999, Journal of Accountancy, Article.

Mark J. Nigrini: "Benford's Law—Applications for Forensic Accounting, Auditing, and Fraud Detection" , Book, Apr. 2012, Wiley Publishing.

Cindy Durtschi: "The Effective Use of Benford's Law to Assist in Detecting Fraud in Accounting Data", Article, Jul. 2003, DOA Utah State Pb.

* cited by examiner

Primary Examiner — Chuong D Ngo

(57) **ABSTRACT**

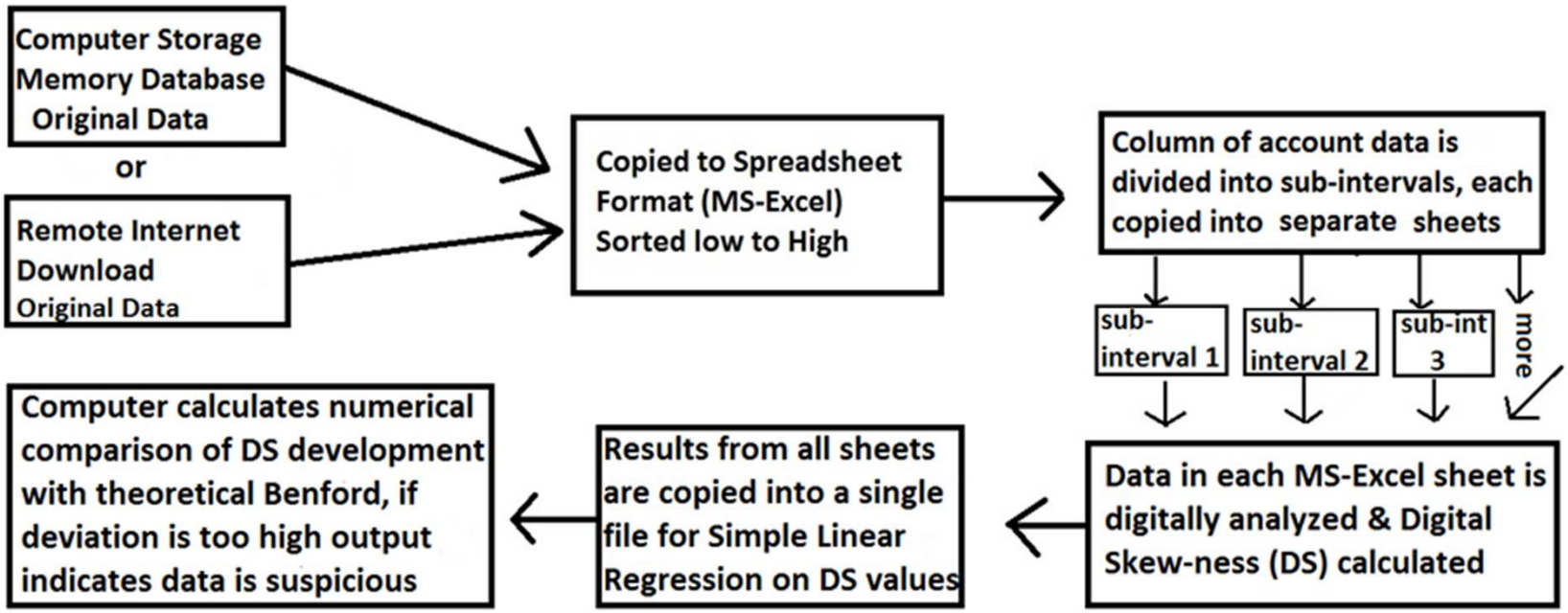
A computerized system for a digital method for the detection of fraud and/or anomalous transactions is disclosed based on a novel statistical interpretation of Benford's Law and a unique set of computer implementations outlining the development of digital distributions from the low-value region on [illegible] data set to the high-value region on the right.

Patent's Title:

"***Method and system for Forensic Data Analysis in fraud detection employing a digital pattern more prevalent than Benford's Law***".

Link on US Patent Office website:

http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&p=1&u=%2Fnetahtml%2FPTO%2Fsearch-bool.html&r=1&f=G&l=50&co1=AND&d=PTXT&s1=9058285&OS=9058285&RS=9058285


Computer Storage
Memory Database
Original Data
or
Remote Internet
Download
Original Data
Copied to Spreadsheet
Format (MS-Excel)
Sorted low to High
Column of account data is
divided into sub-intervals, each
copied into separate sheets
sub-
interval 1
sub-
interval 2
sub-int
3
more
Data in each MS-Excel sheet is
digitally analyzed & Digital
Skew-ness (DS) calculated
Results from all sheets
are copied into a single
file for Simple Linear
Regression on DS values
Computer calculates numerical
comparison of DS development
with theoretical Benford, if
deviation is too high output
indicates data is suspicious

END