

# Data from our man Zipf

## Principles of Complex Systems

CSYS/MATH 300, Spring, 2013 | #SpringPoCS2013

Zipf in brief

Zipfian empirics

Yet more Zipfian  
Empirics

References

Prof. Peter Dodds  
@peterdodds

Department of Mathematics & Statistics | Center for Complex Systems |  
Vermont Advanced Computing Center | University of Vermont



# Outline

Data from our man  
Zipf

Zipf in brief

Zipf in brief

Zipfian empirics

Zipfian empirics

Yet more Zipfian Empirics

Yet more Zipfian  
Empirics

References

References



# George Kingsley Zipf:

Data from our man  
Zipf

## In brief:

- ▶ Zipf (田) (1902–1950) was a linguist at Harvard, specializing in Chinese languages.
- ▶ Unusual passion for statistical analysis of texts.
- ▶ Studied human behavior much more generally...

## Zipf's masterwork:

- ▶ “Human Behavior and the Principle of Least Effort”  
Addison-Wesley, 1949  
Cambridge, MA [3]

- ▶ Bonus field of study: Glottometrics. (田)
- ▶ Bonus ‘word’ word: Glossolalia. (田)

Zipf in brief

Zipfian empirics

Yet more Zipfian  
Empirics

References



# George Kingsley Zipf:

Data from our man  
Zipf

## In brief:

- ▶ Zipf (田) (1902–1950) was a linguist at Harvard, specializing in Chinese languages.
- ▶ Unusual passion for statistical analysis of texts.
- ▶ Studied human behavior much more generally...

## Zipf's masterwork:

- ▶ “Human Behavior and the Principle of Least Effort”  
Addison-Wesley, 1949  
Cambridge, MA [3]

- ▶ Bonus field of study: Glottometrics. (田)
- ▶ Bonus ‘word’ word: Glossolalia. (田)

Zipf in brief

Zipfian empirics

Yet more Zipfian  
Empirics

References



# George Kingsley Zipf:

Data from our man  
Zipf

## In brief:

- ▶ Zipf (田) (1902–1950) was a linguist at Harvard, specializing in Chinese languages.
- ▶ Unusual passion for statistical analysis of texts.
- ▶ Studied human behavior much more generally...

Zipf in brief

Zipfian empirics

Yet more Zipfian  
Empirics

References

## Zipf's masterwork:

- ▶ “Human Behavior and the Principle of Least Effort”  
Addison-Wesley, 1949  
Cambridge, MA [3]

- ▶ Bonus field of study: Glottometrics. (田)
- ▶ Bonus ‘word’ word: Glossolalia. (田)



# George Kingsley Zipf:

Data from our man  
Zipf

## In brief:

- ▶ Zipf (田) (1902–1950) was a linguist at Harvard, specializing in Chinese languages.
- ▶ Unusual passion for statistical analysis of texts.
- ▶ Studied human behavior much more generally...

## Zipf's masterwork:

- ▶ “Human Behavior and the Principle of Least Effort”  
Addison-Wesley, 1949  
Cambridge, MA [3]

- ▶ Bonus field of study: Glottometrics. (田)
- ▶ Bonus 'word' word: Glossolalia. (田)

Zipf in brief

Zipfian empirics

Yet more Zipfian  
Empirics

References



# George Kingsley Zipf:

Data from our man  
Zipf

## In brief:

- ▶ Zipf (田) (1902–1950) was a linguist at Harvard, specializing in Chinese languages.
- ▶ Unusual passion for statistical analysis of texts.
- ▶ Studied human behavior much more generally...

## Zipf's masterwork:

- ▶ “Human Behavior and the Principle of Least Effort”  
Addison-Wesley, 1949  
Cambridge, MA [3]

- ▶ Bonus field of study: Glottometrics. (田)
- ▶ Bonus 'word' word: Glossolalia. (田)

Zipf in brief

Zipfian empirics

Yet more Zipfian  
Empirics

References



# George Kingsley Zipf:

Data from our man  
Zipf

## In brief:

- ▶ Zipf (田) (1902–1950) was a linguist at Harvard, specializing in Chinese languages.
- ▶ Unusual passion for statistical analysis of texts.
- ▶ Studied human behavior much more generally...

## Zipf's masterwork:

- ▶ “Human Behavior and the Principle of Least Effort”  
Addison-Wesley, 1949  
Cambridge, MA [3]
- ▶ Bonus field of study: Glottometrics. (田)
- ▶ Bonus ‘word’ word: Glossolalia. (田)

Zipf in brief

Zipfian empirics

Yet more Zipfian  
Empirics

References





# Human Behavior/Principle of Least Effort:

Data from our man  
Zipf

## From the Preface—

Nearly twenty-five years ago it occurred to me that we might gain considerable insight into the mainsprings of human behavior if we viewed it purely as a natural phenomenon like everything else in the universe, ...

## And—

... the expressed purpose of this book is to establish The Principle of Least Effort as the primary principle that governs our entire individual and collective behavior ...

Zipf in brief

Zipfian empirics

Yet more Zipfian  
Empirics

References



# Human Behavior/Principle of Least Effort:

Data from our man  
Zipf

Zipf in brief

Zipfian empirics

Yet more Zipfian  
Empirics

References

## From the Preface—

Nearly twenty-five years ago it occurred to me that we might gain considerable insight into the mainsprings of human behavior if we viewed it purely as a natural phenomenon like everything else in the universe, ...

## And—

... the expressed purpose of this book is to establish **The Principle of Least Effort** as the primary principle that governs our entire individual and collective behavior ...



# The Principle of Least Effort:

Data from our man  
Zipf

Zipf in brief

Zipfian empirics

Yet more Zipfian  
Empirics

References

## Zipf's framing (p. 1):

“... a person in solving his immediate problems will view these against the background of his probable future problems *as estimated by himself.*”

“... he will strive ... to minimize the *total work* that he must expend in solving *both* his immediate problems *and* his probable future problems.”

“[he will strive to] minimize the *probable average rate of his work-expenditure...*”



# The Principle of Least Effort:

Data from our man  
Zipf

Zipf in brief

Zipfian empirics

Yet more Zipfian  
Empirics

References

## Zipf's framing (p. 1):

“... a person in solving his immediate problems will view these against the background of his probable future problems *as estimated by himself.*”

“... he will strive ... to minimize the *total work* that he must expend in solving *both* his immediate problems *and* his probable future problems.”

“[he will strive to] minimize the *probable average rate of his work-expenditure...*”



# The Principle of Least Effort:

Data from our man  
Zipf

Zipf in brief

Zipfian empirics

Yet more Zipfian  
Empirics

References

## Zipf's framing (p. 1):

“... a person in solving his immediate problems will view these against the background of his probable future problems *as estimated by himself.*”

“... he will strive ... to minimize the *total work* that he must expend in solving *both* his immediate problems *and* his probable future problems.”

“[he will strive to] minimize the *probable average rate of his work-expenditure...*”



# Rampaging research

## Within Human Behavior and the Principle of Least Effort:

- ▶ City sizes
- ▶ # retail stores in cities
- ▶ # services (barber shops, beauty parlors, cleaning, ...)
- ▶ # people in occupations
- ▶ # one-way trips in cars and trucks vs. distance
- ▶ # new items by dateline
- ▶ weight moved between cities by rail
- ▶ # telephone messages between cities
- ▶ # people moving vs. distance
- ▶ # marriages vs. distance

- ▶ Observed general dependency of 'interactions' between cities  $A$  and  $B$  on  $P_A P_B / D_{AB}$  where  $P_A$  and  $P_B$  are population size and  $D_{AB}$  is distance between  $A$  and  $B$ .

Zipf in brief

Zipfian empirics

Yet more Zipfian  
Empirics

References



# Rampaging research

Data from our man  
Zipf

## Within Human Behavior and the Principle of Least Effort:

- ▶ City sizes
- ▶ # retail stores in cities
- ▶ # services (barber shops, beauty parlors, cleaning, ...)
- ▶ # people in occupations
- ▶ # one-way trips in cars and trucks vs. distance
- ▶ # new items by dateline
- ▶ weight moved between cities by rail
- ▶ # telephone messages between cities
- ▶ # people moving vs. distance
- ▶ # marriages vs. distance

- ▶ Observed **general dependency** of 'interactions' between **cities A and B** on  $P_A P_B / D_{AB}$  where  $P_A$  and  $P_B$  are population size and  $D_{AB}$  is distance between A and B.

Zipf in brief

Zipfian empirics

Yet more Zipfian  
Empirics

References



# Rampaging research

Data from our man  
Zipf

## Within Human Behavior and the Principle of Least Effort:

- ▶ City sizes
- ▶ # retail stores in cities
- ▶ # services (barber shops, beauty parlors, cleaning, ...)
- ▶ # people in occupations
- ▶ # one-way trips in cars and trucks vs. distance
- ▶ # new items by dateline
- ▶ weight moved between cities by rail
- ▶ # telephone messages between cities
- ▶ # people moving vs. distance
- ▶ # marriages vs. distance

- ▶ Observed **general dependency** of 'interactions' between **cities A and B** on  $P_A P_B / D_{AB}$  where  $P_A$  and  $P_B$  are population size and  $D_{AB}$  is distance between A and B.  $\Rightarrow$  'Gravity Law.'

Zipf in brief

Zipfian empirics

Yet more Zipfian  
Empirics

References





# Zipfian empirics:

- ▶ vocabulary balance:  $f \sim r^{-1} \rightarrow r \cdot f \sim \text{constant}$  ( $f$  = frequency,  $r$  = rank).

Data from our man  
Zipf

Zipf in brief

Zipfian empirics

Yet more Zipfian  
Empirics

References

TABLE 2-1

Arbitrary Ranks with Frequencies  
in James Joyce's *Ulysses*  
(Hanley Index)

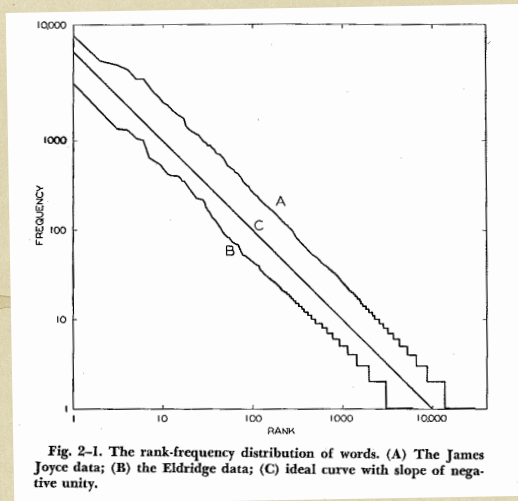
| I<br>Rank<br>( $r$ ) | II<br>Frequency<br>( $f$ ) | III<br>Product of<br>I and II<br>( $r \times f = C$ ) | IV<br>Theoretical<br>Length of <i>Ulysses</i><br>( $C \times 10$ ) |
|----------------------|----------------------------|---|--|
| 10                   | 2,653                      | 26,530  | 265,500  |
| 20                   | 1,311                      | 26,220  | 262,200  |
| 30                   | 926                        | 27,780  | 277,800  |
| 40                   | 717                        | 28,680  | 286,800  |
| 50                   | 556                        | 27,800  | 278,800  |
| 100                  | 265                        | 26,500  | 265,000  |
| 200                  | 133                        | 26,600  | 266,000  |
| 300                  | 84                         | 25,200  | 252,000  |
| 400                  | 62                         | 24,800  | 248,000  |
| 500                  | 50                         | 25,000  | 250,000  |
| 1,000                | 26                         | 26,000  | 260,000  |
| 2,000                | 12                         | 24,000  | 240,000  |
| 3,000                | 8                          | 24,000  | 240,000  |
| 4,000                | 6                          | 24,000  | 240,000  |
| 5,000                | 5                          | 25,000  | 250,000  |
| 10,000               | 2                          | 20,000  | 200,000  |
| 20,000               | 1                          | 20,000  | 200,000  |
| 29,899               | 1                          | 29,899  | 298,990  |



# Zipfian empirics:

Data from our man  
Zipf

►  $f \sim r^{-1}$  for word frequency:



Zipf in brief

Zipfian empirics

Yet more Zipfian  
Empirics

References



# Zipf's basic idea:

Data from our man  
Zipf

## Forces of Unification and Diversification:

- ▶ Easiest for the speaker to use just one word.
- ▶ Zipf uses the analogy of tools: one tool for all tasks.
- ▶ Optimal for listener if all pieces of information correspond to different words (or morphemes).
- ▶ Analogy: a specialized tool for every task.
- ▶ Zipf thereby argues for a tension that should lead to an uneven distribution of word usage.
- ▶ No formal theory beyond this... (more later<sup>[1]</sup>)

Zipf in brief

Zipfian empirics

Yet more Zipfian  
Empirics

References



# Zipf's basic idea:

Data from our man  
Zipf

## Forces of Unification and Diversification:

- ▶ Easiest for the speaker to use just one word.
- ▶ Zipf uses the analogy of tools: **one tool for all tasks.**
- ▶ Optimal for listener if all pieces of information correspond to different words (or morphemes).
- ▶ Analogy: a specialized tool for every task.
- ▶ Zipf thereby argues for a tension that should lead to an uneven distribution of word usage.
- ▶ No formal theory beyond this... (more later<sup>[1]</sup>)

Zipf in brief

Zipfian empirics

Yet more Zipfian  
Empirics

References



# Zipf's basic idea:

Data from our man  
Zipf

## Forces of Unification and Diversification:

- ▶ Easiest for the speaker to use just one word.
- ▶ Zipf uses the analogy of tools: **one tool for all tasks.**
- ▶ Optimal for listener if all pieces of information correspond to different words (or morphemes).
- ▶ Analogy: a specialized tool for every task.
- ▶ Zipf thereby argues for a tension that should lead to an uneven distribution of word usage.
- ▶ No formal theory beyond this... (more later<sup>[1]</sup>)

Zipf in brief

Zipfian empirics

Yet more Zipfian  
Empirics

References



# Zipf's basic idea:

Data from our man  
Zipf

## Forces of Unification and Diversification:

- ▶ Easiest for the speaker to use just one word.
  - ▶ **Encoding is simple** but decoding is hard
- ▶ Zipf uses the analogy of tools: **one tool for all tasks.**
- ▶ Optimal for listener if all pieces of information correspond to different words (or morphemes).
- ▶ Analogy: a specialized tool for every task.
- ▶ Zipf thereby argues for a tension that should lead to an uneven distribution of word usage.
- ▶ No formal theory beyond this... (more later<sup>[1]</sup>)

Zipf in brief

Zipfian empirics

Yet more Zipfian  
Empirics

References



# Zipf's basic idea:

Data from our man  
Zipf

## Forces of Unification and Diversification:

- ▶ Easiest for the speaker to use just one word.
  - ▶ **Encoding is simple** but decoding is hard
- ▶ Zipf uses the analogy of tools: **one tool for all tasks.**
  
- ▶ Optimal for listener if all pieces of information correspond to different words (or morphemes).
- ▶ Analogy: a specialized tool for every task.
  - ▶ **Decoding is simple** but encoding is hard
  
- ▶ Zipf thereby argues for a tension that should lead to an uneven distribution of word usage.
- ▶ No formal theory beyond this... (more later<sup>[1]</sup>)

Zipf in brief

Zipfian empirics

Yet more Zipfian  
Empirics

References



# Zipf's basic idea:

Data from our man  
Zipf

## Forces of Unification and Diversification:

- ▶ Easiest for the speaker to use just one word.
  - ▶ **Encoding is simple** but **decoding is hard**
- ▶ Zipf uses the analogy of tools: **one tool for all tasks.**
  
- ▶ Optimal for listener if all pieces of information correspond to different words (or morphemes).
- ▶ Analogy: a specialized tool for every task.
  - ▶ **Decoding is simple** but **encoding is hard**
  
- ▶ Zipf thereby argues for a tension that should lead to an uneven distribution of word usage.
- ▶ No formal theory beyond this... (more later<sup>[1]</sup>)

Zipf in brief

Zipfian empirics

Yet more Zipfian  
Empirics

References





# Zipf's basic idea:

Data from our man  
Zipf

## Forces of Unification and Diversification:

- ▶ Easiest for the speaker to use just one word.
  - ▶ **Encoding is simple** but **decoding is hard**
- ▶ Zipf uses the analogy of tools: **one tool for all tasks.**
  
- ▶ Optimal for listener if all pieces of information correspond to different words (or morphemes).
- ▶ Analogy: a specialized tool for every task.
  - ▶ **Decoding is simple** but **encoding is hard**
  
- ▶ Zipf thereby argues for a tension that should lead to an uneven distribution of word usage.
- ▶ No formal theory beyond this... (more later<sup>[1]</sup>)

Zipf in brief

Zipfian empirics

Yet more Zipfian  
Empirics

References



# Zipf's basic idea:

Data from our man  
Zipf

## Forces of Unification and Diversification:

- ▶ Easiest for the speaker to use just one word.
  - ▶ **Encoding is simple** but **decoding is hard**
- ▶ Zipf uses the analogy of tools: **one tool for all tasks.**
  
- ▶ Optimal for listener if all pieces of information correspond to different words (or morphemes).
- ▶ Analogy: a specialized tool for every task.
  - ▶ **Decoding is simple** but **encoding is hard**
  
- ▶ Zipf thereby argues for a tension that should lead to an uneven distribution of word usage.
- ▶ No formal theory beyond this... (more later <sup>[1]</sup>)

Zipf in brief

Zipfian empirics

Yet more Zipfian  
Empirics

References



# Zipfian empirics:

Data from our man  
Zipf

- ▶ Number of meanings  $m_r \propto f_r^{1/2}$  where  $r$  is rank and  $f_r$  is frequency.

Zipf in brief

Zipfian empirics

Yet more Zipfian  
Empirics

References

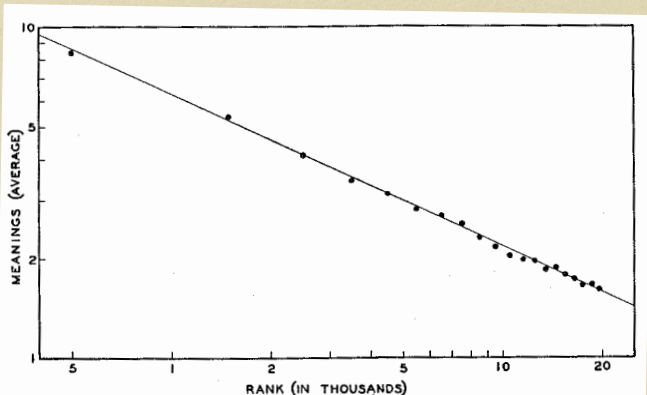


Fig. 2-2. The meaning-frequency distribution of words.

# Zipfian empirics:

Data from our man  
Zipf

- ▶ Article length in the Encyclopedia Britannica:

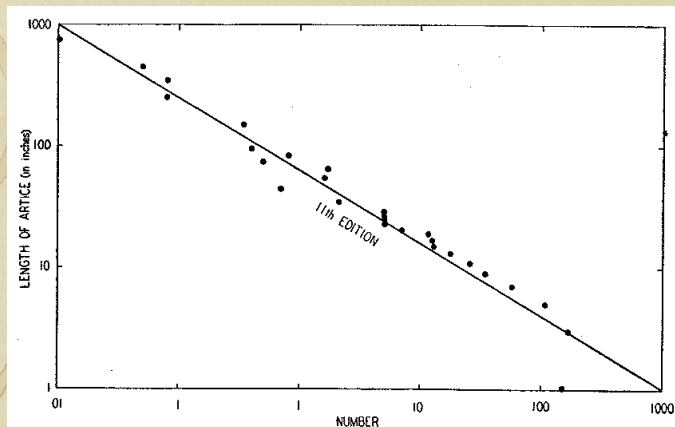


Fig. 5-3. The number of different articles of like length in samples of the 11th edition of the *Encyclopaedia Britannica*. Lengths in inches.

- ▶ (?) slope of  $-3/5$  corresponds to  $\gamma = 5/3$ .

Zipf in brief

Zipfian empirics

Yet more Zipfian  
Empirics

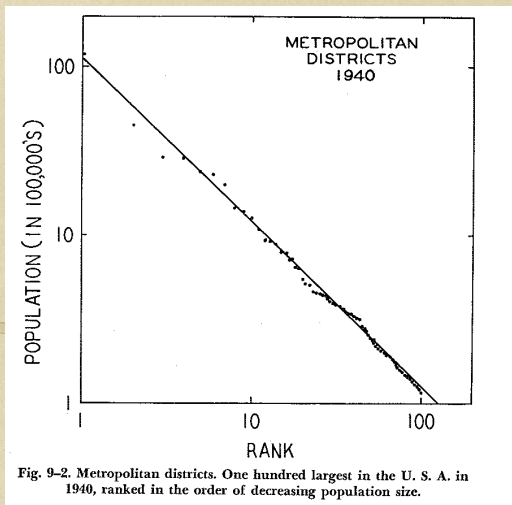
References



# Zipfian empirics:

Data from our man  
Zipf

- ▶ Population size of districts:



- ▶  $\alpha = 1$  corresponds to  $\gamma = 1 + 1/\alpha = 2$ .

Zipf in brief

Zipfian empirics

Yet more Zipfian  
Empirics

References



# Zipfian empirics:

- ▶ Number of employees in organizations

Data from our man  
Zipf

Zipf in brief

Zipfian empirics

Yet more Zipfian  
Empirics

References

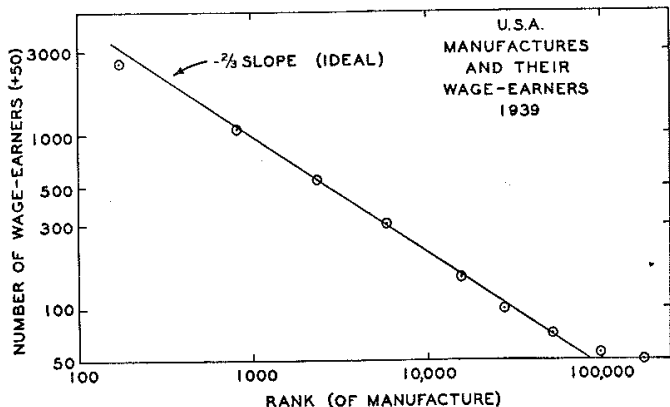


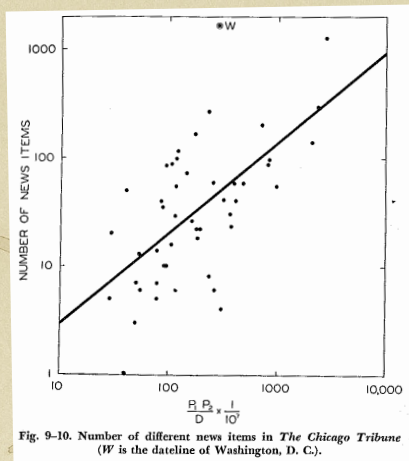
Fig. 9-8. Manufactures and their wage earners in the U. S. A. in 1939, with the manufactures ranked in the order of their decreasing number of wage earners.

- ▶  $\alpha = 2/3$  corresponds to  $\gamma = 1 + 1/\alpha = 5/2$ .



# Zipfian empirics:

- ▶ # news items as a function of population  $P_2$  of location in the Chicago Tribune
- ▶  $D$  = distance,  $P_1$  = Chicago's population
- ▶ Solid line = +1 exponent.



Data from our man  
Zipf

[Zipf in brief](#)

[Zipfian empirics](#)

[Yet more Zipfian  
Empirics](#)

[References](#)



# Zipfian empirics:

- ▶ # obituaries in the New York Times for locations with population  $P_2$ .
- ▶  $D$  = distance,  $P_1$  = New York's population
- ▶ Solid line = +1 exponent.

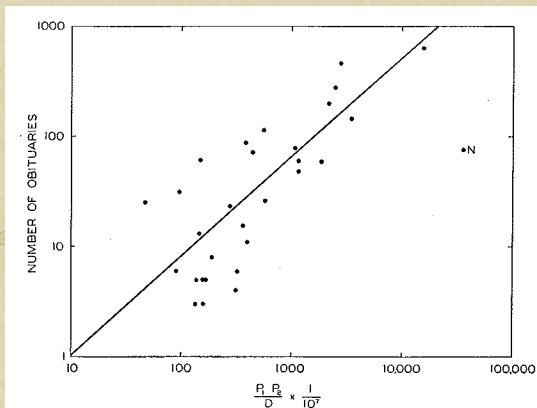


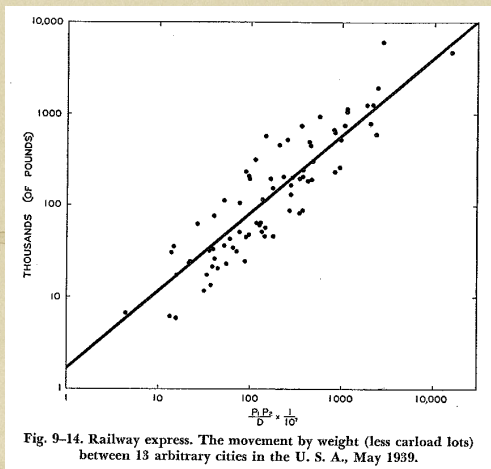
Fig. 9-11. Number of obituaries in *The New York Times* ( $N$  represents Newark, New Jersey).





# Zipfian empirics:

- ▶ Movement of stuff between cities
- ▶  $D$  = distance,  $P_1$  and  $P_2$  = city populations.
- ▶ Solid line = +1 exponent.



Data from our man  
Zipf

Zipf in brief

Zipfian empirics

Yet more Zipfian  
Empirics

References



# Zipfian empirics:

Data from our man  
Zipf

- ▶ Length of trip versus frequency of trip.
- ▶ Solid line =  $-1/2$  exponent corresponds to  $\gamma = 2$ .

Zipf in brief

Zipfian empirics

Yet more Zipfian  
Empirics

References

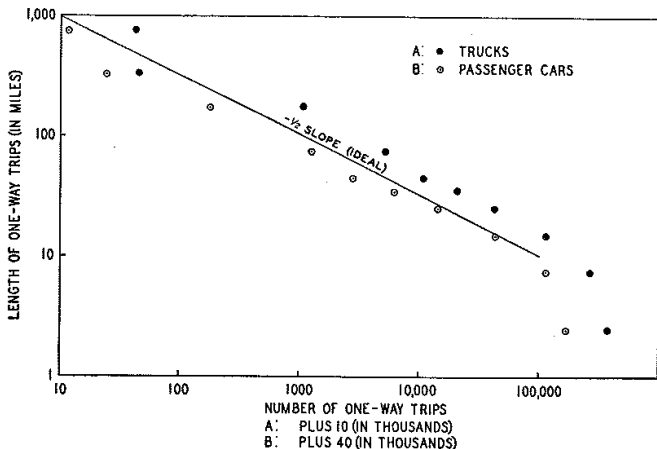


Fig. 9-19. Trucks and passenger cars: the number of one-way trips of like length.



# Zipfian empirics:

- ▶ The probability of marriage?
- ▶  $\gamma = 1$ ?

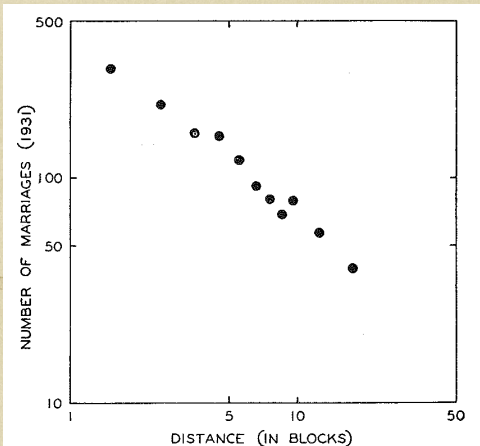


Fig. 9-22. Number of marriage licenses issued to 5,000 pairs of applicants living within Philadelphia in 1931 and separated by varying distances (the data of J. H. S. Bossard).



## Comment #60 in Math and the City (田) by Strogatz, NYT:

60. May 20, 2009  
9:26 am

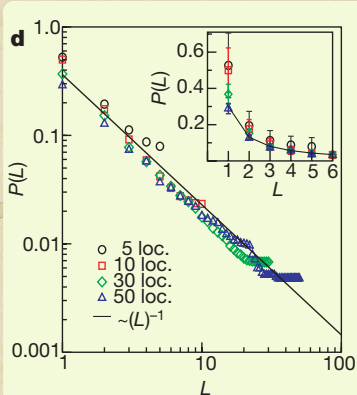
[Link](#)

George Kingsley Zipf was my teacher at Harvard...He had given a class project where we were to see if Chemical Companies when ranked by the number of different chemicals they produced, followed his Law of Least Effort. I missed turning in my assignment due to the accidental death of my father....When I returned from the funeral I was given a message to call Dr. Zipf immediately. I did and when I explained why I was late turning in the data. He said, "Well, your father's gone and I (Zipf) have no pipeline to God. I expect the data will be on my desk tomorrow morning!".....My mother, sister and extended family spread huge books of trade magazines on the kitchen and dining room tables and furiously went to work....We worked until late in the night and finished the project....I drove to Harvard the next morning and angrily gave the hundreds of 'three by five cards' to Zipf. All he said was, "Thank you." Years later, I wondered whether his 'meanness' had really been his way of helping me and my family to take our minds off our grief that day and concentrate on finishing my assignment. In my youth I thought not, but now as I approach 80, I like to think his seemingly hurtful attitude was really an act of kindness,,,,,

— *Jim Terry*

# Recent Zipf action:

Data from our man  
Zipf



- ▶ Probability of people being in certain locations follows a Zipfish law...
- ▶ From Gonzàlez et al., Nature (2008)  
“Understanding individual human mobility patterns” [2]

Bonus: Marta's talk (田) at  
UVM's 2011 TEDx event “Big Data, Big Stories.” (田)

Zipf in brief

Zipfian empirics

Yet more Zipfian  
Empirics

References



# Zipfian empirics:

Data from our man  
Zipf

TABLE 2-2

The Number-Frequency Relationship,  $N(f^2 - 1/4) = C$ , of (I) some Arbitrary Lower Frequencies of (II) Joyce's *Ulysses* and (III) four Latin plays of Plautus.

| I<br>Frequency<br>( $f$ ) | Calculated $N(f^2 - 1/4)$ |                |
|---------------------------|---------------------------|----------------|
|                           | II<br>Ulysses             | III<br>Plautus |
| 1                         | 12,324                    | 4,075          |
| 2                         | 15,410                    | 4,490          |
| 3                         | 19,193                    | 4,280          |
| 4                         | 20,239                    | 4,750          |
| 5                         | 22,424                    | 3,985          |
| 6                         | 22,773                    | 4,504          |
| 7                         | 23,546                    | 4,241          |
| 8                         | 23,651                    | 4,399          |
| 9                         | 24,063                    | 4,366          |
| 10                        | 22,145                    | 4,289          |
| 15                        | 21,576                    | 2,922          |
| 20                        | 27,844                    | 5,996          |
| 30                        | 18,000                    | 3,600          |
| 40                        | 25,600                    | 4,800          |
| 50                        | 22,500                    | 5,000          |

Zipf in brief

Zipfian empirics

Yet more Zipfian  
Empirics

References



# Zipfian empirics:

Data from our man  
Zipf

Zipf in brief

Zipfian empirics

Yet more Zipfian  
Empirics

References

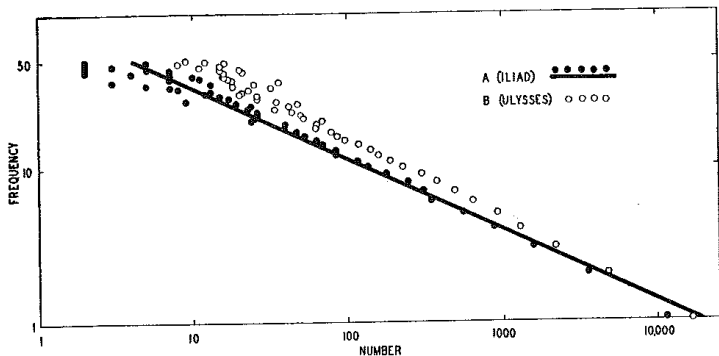


Fig. 2-3. The number-frequency relationship of words. (A) Homer's *Iliad*;  
(B) James Joyce's *Ulysses*.



# Zipfian empirics:

Data from our man  
Zipf

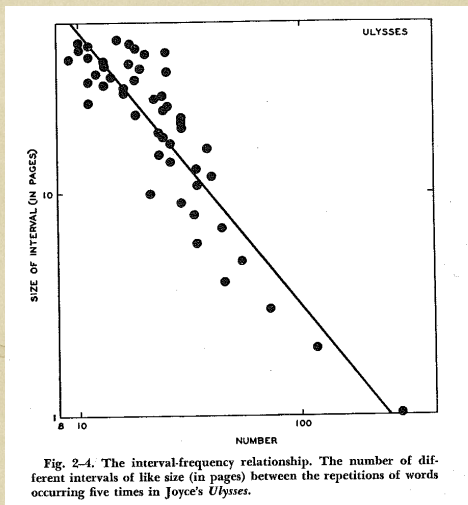


Fig. 2-4. The interval-frequency relationship. The number of different intervals of like size (in pages) between the repetitions of words occurring five times in Joyce's *Ulysses*.

Zipf in brief

Zipfian empirics

Yet more Zipfian  
Empirics

References





# Zipfian empirics:

Data from our man  
Zipf

TABLE 2-3

Calculated values of negative slopes, errors, and Y-intercepts of the number,  $N$ , of interval-sizes,  $I_p$ , between the repetition of words in 14 frequency-classes,  $f$ , as fitted to the equation  $aX + Y = C$  where  $X = \log N$  and  $Y = \log I_p$  and where  $I_p$  has integral values from 1 through 21 inclusive.

| I<br>No. of<br>Analysis | II<br>Frequency<br>of Occur.<br>( $f$ ) | III<br>No. of<br>Different<br>Words of<br>like $f$ | IV<br>Slope of<br>Best Line<br>of $Y$ 's<br>(negative)<br>( $Y = \log I_t$ ) | V<br>Error<br>(root-mean-<br>square) | VI<br>$Y$ -intercept<br>(antilog<br>thereof) |
|-------------------------|---|--|--|--------------------------------------|--|
| 1                       | 5                                       | 906  | 1.21   | .151                                 | 716  |
| 2                       | 6                                       | 637  | 1.20   | .169                                 | 666  |
| 3                       | 10                                      | 222  | 1.27   | .106                                 | 677  |
| 4                       | 12                                      | 155  | 1.24   | .111                                 | 491  |
| 5                       | 15                                      | 96   | 1.15   | .096                                 | 328  |
| 6                       | 16                                      | 86   | .96  | .124                                 | 153  |
| 7                       | 17                                      | 79   | 1.22   | .174                                 | 422  |
| 8                       | 18                                      | 62   | 1.20   | .120                                 | 264  |
| 9                       | 19                                      | 63   | 1.21   | .148                                 | 350  |
| 10                      | 20                                      | 69   | 1.29   | .124                                 | 944  |
| 11                      | 21                                      | 52   | 1.05   | .138                                 | 212  |
| 12                      | 22                                      | 50   | 1.10   | .117                                 | 264  |
| 13                      | 23                                      | 44   | 1.24   | .113                                 | 352  |
| 14F                     | 24                                      | 34   | 1.01   | .158                                 | 136  |
| 15Z                     | 24                                      | 34   | 1.05   | .147                                 | 153  |

Zipf in brief

Zipfian empirics

Yet more Zipfian  
Empirics

References



Zipf in brief

Zipfian empirics

Yet more Zipfian  
Empirics

References

TABLE 2-4

The dispersion of single-page intervals between the  $f - 1$  repetitions of all words that occur with ten arbitrarily selected frequencies of occurrence,  $f$ , in Joyce's *Ulysses* (Hanley's *Index*).

A

The First 12 Intervals between Repetitions

| No. of Sample | $f$ | $f - 1$ | Intervals between Repetitions in Order of Appearance |    |    |    |    |    |    |    |    |    |    |    |  |  |  |
|---------------|-----|---------|--|----|----|----|----|----|----|----|----|----|----|----|--|--|--|
|               |     |         | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | 10 | 11 | 12 |  |  |  |
| 1             | 6   | 5       | 62   | 55 | 62 | 58 | 52 |    |    |    |    |    |    |    |  |  |  |
| 2             | 12  | 11      | 7  | 19 | 15 | 16 | 9  | 12 | 18 | 16 | 12 | 15 | 14 |    |  |  |  |
| 3             | 16  | 15      | 6  | 10 | 10 | 13 | 18 | 11 | 16 | 11 | 11 | 9  | 11 | 9  |  |  |  |
| 4             | 17  | 16      | 4  | 3  | 5  | 6  | 4  | 8  | 5  | 10 | 11 | 9  | 14 | 5  |  |  |  |
| 5             | 18  | 17      | 9  | 11 | 6  | 5  | 6  | 7  | 7  | 6  | 9  | 6  | 2  | 6  |  |  |  |
| 6             | 19  | 18      | 3  | 8  | 5  | 11 | 5  | 6  | 13 | 9  | 6  | 5  | 6  | 8  |  |  |  |
| 7             | 21  | 20      | 3  | 4  | 10 | 5  | 8  | 9  | 3  | 10 | 8  | 11 | 7  | 7  |  |  |  |
| 8             | 22  | 21      | 7  | 5  | 8  | 12 | 5  | 9  | 5  | 9  | 6  | 7  | 5  | 8  |  |  |  |
| 9             | 23  | 22      | 3  | 5  | 6  | 4  | 8  | 4  | 3  | 2  | 7  | 3  | 4  | 4  |  |  |  |
| 10            | 24  | 23      | 3  | 5  | 2  | 1  | 3  | 3  | 3  | 3  | 4  | 5  | 2  | 3  |  |  |  |

B

The Intervals from 13 through 23

| No. of Sample | $f$ | $f - 1$ | Intervals between Repetitions in Order of Appearance |    |    |    |    |    |    |    |    |    |    |  |  |  |  |
|---------------|-----|---------|--|----|----|----|----|----|----|----|----|----|----|--|--|--|--|
|               |     |         | 13   | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 |  |  |  |  |
| 3             | 16  | 15      | 6  | 8  | 12 |    |    |    |    |    |    |    |    |  |  |  |  |
| 4             | 17  | 16      | 8  | 6  | 7  | 8  |    |    |    |    |    |    |    |  |  |  |  |
| 5             | 18  | 17      | 5  | 6  | 6  | 5  | 4  |    |    |    |    |    |    |  |  |  |  |
| 6             | 19  | 18      | 2  | 7  | 10 | 5  | 7  | 4  |    |    |    |    |    |  |  |  |  |
| 7             | 21  | 20      | 6  | 6  | 2  | 1  | 7  | 8  | 4  | 2  |    |    |    |  |  |  |  |
| 8             | 22  | 21      | 6  | 6  | 7  | 10 | 7  | 10 | 9  | 5  | 2  |    |    |  |  |  |  |
| 9             | 23  | 22      | 5  | 7  | 3  | 6  | 2  | 7  | 2  | 3  | 1  | 3  |    |  |  |  |  |
| 10            | 24  | 23      | 7  | 3  | 2  | 2  | 0  | 1  | 2  | 2  | 2  | 8  | 3  |  |  |  |  |



# Zipfian empirics:

Data from our man  
Zipf

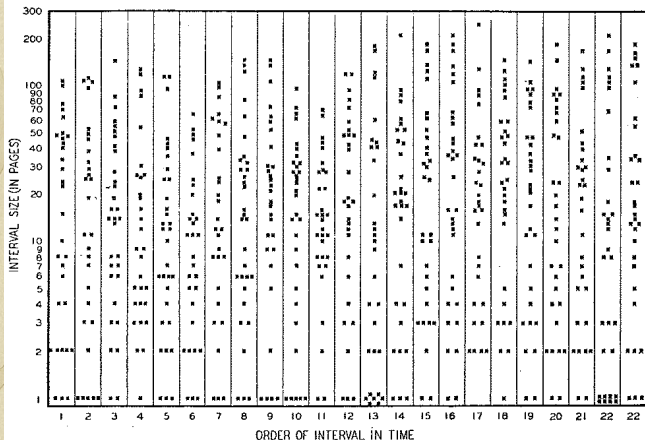


Fig. 2-5. The distribution of intervals between repetitions among the words occurring twenty-four times in James Joyce's *Ulysses*.

Zipf in brief

Zipfian empirics

Yet more Zipfian  
Empirics

References



# Zipfian empirics:

Data from our man  
Zipf

Zipf in brief

Zipfian empirics

Yet more Zipfian  
Empirics

References

TABLE 3-1

The Frequencies and Average Lengths of Words (A) in terms of the number of phonemes, and (B) in terms of the number of syllables in (A) American newspaper English and in (B) the Latin of Plautus.

(A) AMERICAN NEWSPAPER ENGLISH

(According to R. C. Eldridge)

(B) LATIN OF PLAUTUS

| Number of Occurrences | Number of Words | Average Number of Phonemes | Number of Occurrences | Number of Words | Average Number of Phonemes |
|-----------------------|-----------------|----------------------------|-----------------------|-----------------|----------------------------|
| 1                     | 2976            | (6.656)                    | 31                    | 6               | (3.905)                    |
| 2                     | 1079            | (6.151)                    | 32                    | 4               |                            |
| 3                     | 516             | (6.015)                    | 33                    | 6               |                            |
| 4                     | 294             | (6.081)                    | 34                    | 2               |                            |
| 5                     | 212             | (5.589)                    | 35                    | 5               |                            |
| 6                     | 151             | (5.768)                    | 36                    | 3               |                            |
| 7                     | 105             | (5.333)                    | 37                    | 2               |                            |
| 8                     | 84              | (5.654)                    | 39                    | 2               |                            |
| 9                     | 86              | (5.174)                    | 40                    | 4               |                            |
| 10                    | 45              | (5.377)                    | 41                    | 1               |                            |
| 11                    | 40              | (4.825)                    | 42                    | 7               |                            |
| 12                    | 37              | (5.459)                    | 43                    | 1               |                            |
| 13                    | 25              | (5.560)                    | 44                    | 4               |                            |
| 14                    | 28              | (5.00)                     | 45                    | 1               |                            |
| 15                    | 26              | (4.807)                    | 46                    | 2               |                            |
| 16                    | 17              | (5.058)                    | 47                    | 5               |                            |
| 17                    | 18              | (4.166)                    | 48                    | 1               |                            |
| 18                    | 10              | (6.100)                    | 49                    | 3               |                            |
| 19                    | 15              | (4.733)                    | 50                    | 3               |                            |
| 20                    | 16              | (4.687)                    | 51                    | 1               |                            |
| 21                    | 13              | (3.455)                    | 52                    | 3               |                            |
| 22                    | 11              |                            | 54                    | 1               |                            |
| 23                    | 6               |                            | 55                    | 1               |                            |
| 24                    | 8               |                            | 56                    | 1               |                            |
| 25                    | 6               |                            | 58                    | 2               |                            |
| 26                    | 10              |                            | 60                    | 1               |                            |
| 27                    | 9               |                            | 61-4290               | 71              |                            |
| 28                    | 6               |                            | (2.666)               |                 |                            |
| 29                    | 5               |                            |                       |                 |                            |
| 30                    | 4               |                            |                       |                 |                            |

| Number of Occurrences | Number of Words | Average Number of Syllables | Number of Occurrences | Number of Words | Average Number of Syllables |
|-----------------------|-----------------|-----------------------------|-----------------------|-----------------|-----------------------------|
| 1                     | 5429            | (3.23)                      | 31                    | 8               | (2.05)                      |
| 2                     | 1198            | (2.92)                      | 32                    | 3               |                             |
| 3                     | 492             | (2.77)                      | 33                    | 4               |                             |
| 4                     | 299             | (2.05)                      | 34                    | 6               |                             |
| 5                     | 161             | (2.60)                      | 35                    | 3               |                             |
| 6                     | 126             | (2.53)                      | 36                    | 5               |                             |
| 7                     | 87              | (2.39)                      | 37                    | 7               |                             |
| 8                     | 69              | (2.44)                      | 38                    | 2               |                             |
| 9                     | 54              | (2.35)                      | 39                    | 4               |                             |
| 10                    | 43              | (2.32)                      | 40                    | 3               |                             |
| 11                    | 44              | (2.29)                      | 41                    | 5               |                             |
| 12                    | 36              | (2.30)                      | 43                    | 4               |                             |
| 13                    | 33              | (2.30)                      | 44                    | 1               |                             |
| 14                    | 31              | (2.09)                      | 45                    | 1               |                             |
| 15                    | 13              | (2.07)                      | 46                    | 1               |                             |
| 16                    | 25              | (2.40)                      | 47                    | 3               |                             |
| 17                    | 21              | (2.09)                      | 48                    | 1               |                             |
| 18                    | 21              | (2.04)                      | 49                    | 1               |                             |
| 19                    | 11              | (2.18)                      | 50                    | 2               |                             |
| 20                    | 15              | (2.08)                      | 51                    | 2               |                             |
| 21                    | 10              |                             | 52                    | 4               |                             |
| 22                    | 8               |                             | 54                    | 1               |                             |
| 23                    | 8               |                             | 55                    | 1               |                             |
| 24                    | 9               |                             | 56                    | 2               |                             |
| 25                    | 11              |                             | 58                    | 1               |                             |
| 26                    | 7               |                             | 61                    | 3               |                             |
| 27                    | 9               |                             | 62-514                | 71              |                             |
| 28                    | 12              |                             | (2.00)                | 33,094          | 8,437                       |
| 29                    | 4               |                             |                       |                 |                             |
| 30                    | 4               |                             |                       |                 |                             |



# Zipfian empirics:

Data from our man  
Zipf

Zipf in brief

Zipfian empirics

Yet more Zipfian  
Empirics

References

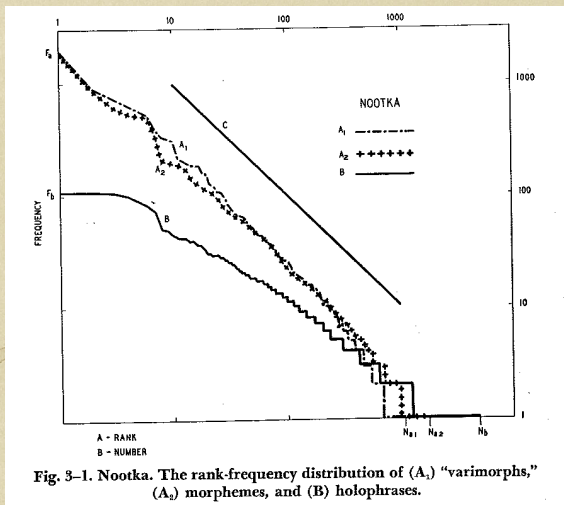


Fig. 3-1. Nootka. The rank-frequency distribution of (A<sub>1</sub>) "varimorphs," (A<sub>2</sub>) morphemes, and (B) holophrases.



# Zipfian empirics:

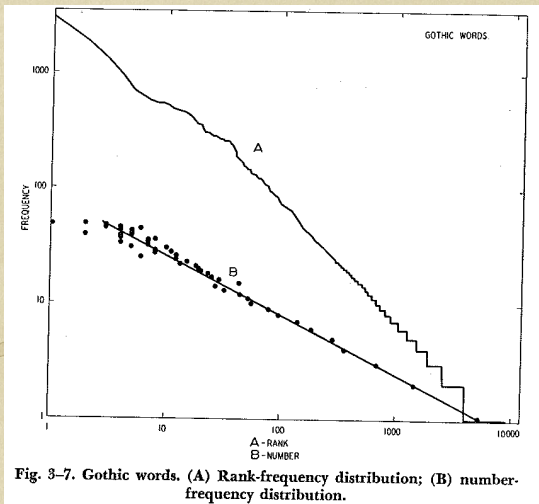
Data from our man  
Zipf

Zipf in brief

Zipfian empirics

Yet more Zipfian  
Empirics

References



# Zipfian empirics:

Data from our man  
Zipf

Zipf in brief

Zipfian empirics

Yet more Zipfian  
Empirics

References

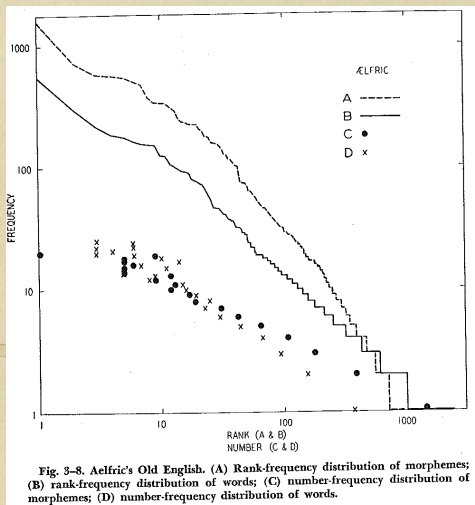


Fig. 3-8. Aelfric's Old English. (A) Rank-frequency distribution of morphemes; (B) rank-frequency distribution of words; (C) number-frequency distribution of morphemes; (D) number-frequency distribution of words.



# Zipfian empirics:

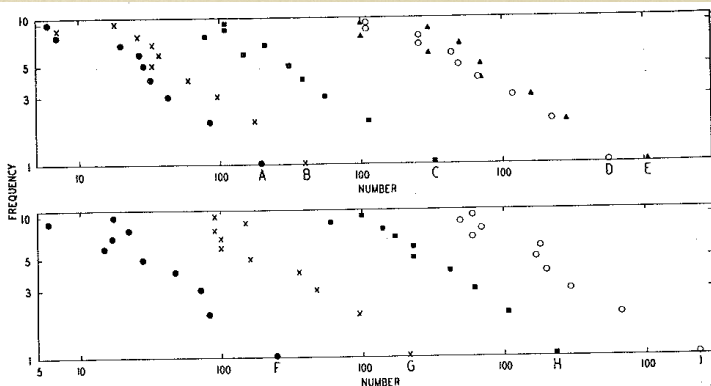
Data from our man  
Zipf

[Zipf in brief](#)

[Zipfian empirics](#)

[Yet more Zipfian  
Empirics](#)

[References](#)



**Fig. 3-9. English and German morphemes. The number-frequency distributions of nine different authors.**





# Zipfian empirics:

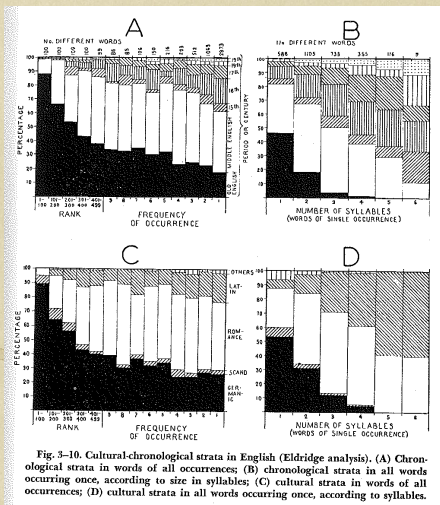
Data from our man  
Zipf

Zipf in brief

Zipfian empirics

Yet more Zipfian  
Empirics

References



# Zipfian empirics (p. 176):

Data from our man  
Zipf

## ► Article length in the Encyclopedia Britannica

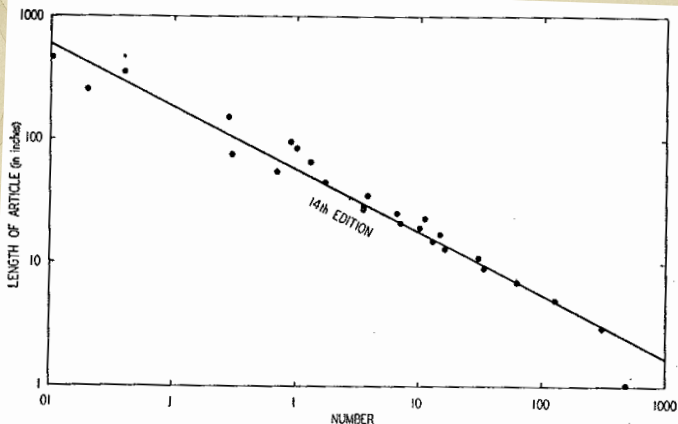


Fig. 5-2. The number of different articles of like length in samples of the 14th edition of the *Encyclopaedia Britannica*. Lengths in inches.

Zipf in brief

Zipfian empirics

Yet more Zipfian  
Empirics

References



# Zipfian empirics:

Data from our man  
Zipf

[Zipf in brief](#)

[Zipfian empirics](#)

[Yet more Zipfian  
Empirics](#)

[References](#)

TABLE 6-1

The  $X$  Number of Different  
Genera of Like  $Y$  Number of  
Different Species of the Flora of  
Ceylon (After J. C. Willis).

| No. of Genera<br>$X$ | No. of Species<br>$Y$ |
|----------------------|-----------------------|
| 573                  | 1                     |
| 176                  | 2                     |
| 85                   | 3                     |
| 49                   | 4                     |
| 36                   | 5                     |
| 20                   | 6                     |
| etc.                 |                       |



# Zipfian empirics:

Data from our man  
Zipf

Zipf in brief

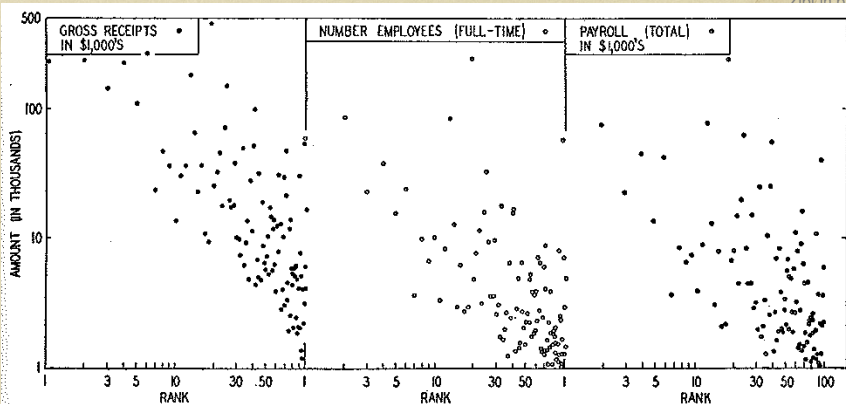


Fig. 9-9. Gross receipts, number of full-time employees, and total payroll of service establishments in the U. S. A. in 1939 when the service establishments are ranked in the order of their decreasing number of members as in Fig. 9-4 *supra*.

# Zipfian empirics:

Data from our man  
Zipf

[Zipf in brief](#)

[Zipfian empirics](#)

[Yet more Zipfian  
Empirics](#)

[References](#)

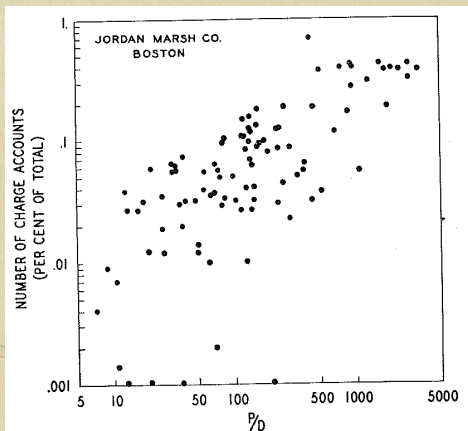


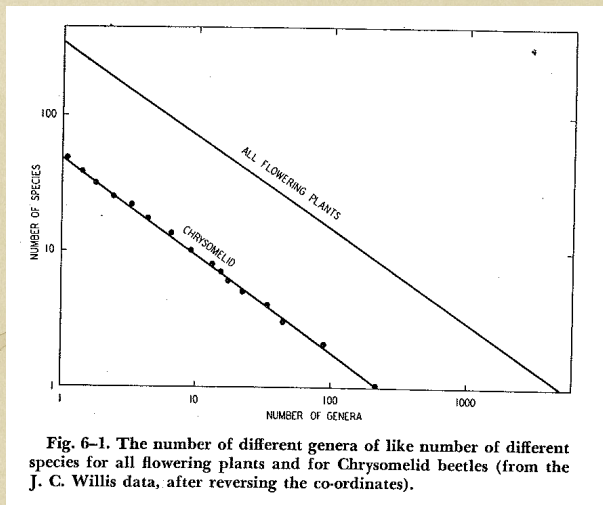
Fig. 9-13. Charge accounts of Jordan Marsh Co., Boston, in 96 cities and towns in Massachusetts, New Hampshire, and Maine, with their percentages of total charge accounts plotted against the communities' values of  $P/D$ .



# Zipfian empirics:

Data from our man  
Zipf

- ▶ # species per genera:



- ▶  $\alpha = 1$  corresponds to  $\gamma = 1 + 1/\alpha = 2$ .

Zipf in brief

Zipfian empirics

Yet more Zipfian  
Empirics

References



# Zipfian empirics:

Data from our man  
Zipf

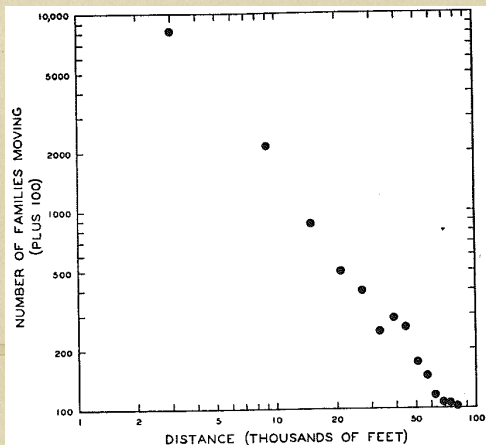


Fig. 9-23. Number of families (plus 100) moving varying distances within or between separated areas in Cleveland during 1933-1935 (adapted from the data of S. A. Stouffer).

Zipf in brief

Zipfian empirics

Yet more Zipfian  
Empirics

References



# References I

- [1] R. Ferrer i Cancho and R. V. Solé.  
Least effort and the origins of scaling in human  
language.  
[Proc. Natl. Acad. Sci, 100:788–791, 2003. pdf \(田\)](#)
- [2] M. C. González, C. A. Hidalgo, and A.-L. Barabási.  
Understanding individual human mobility patterns.  
[Nature, 453:779–782, 2008. pdf \(田\)](#)
- [3] G. K. Zipf.  
Human Behaviour and the Principle of Least-Effort.  
Addison-Wesley, Cambridge, MA, 1949.

[Zipf in brief](#)

[Zipfian empirics](#)

[Yet more Zipfian  
Empirics](#)

[References](#)

