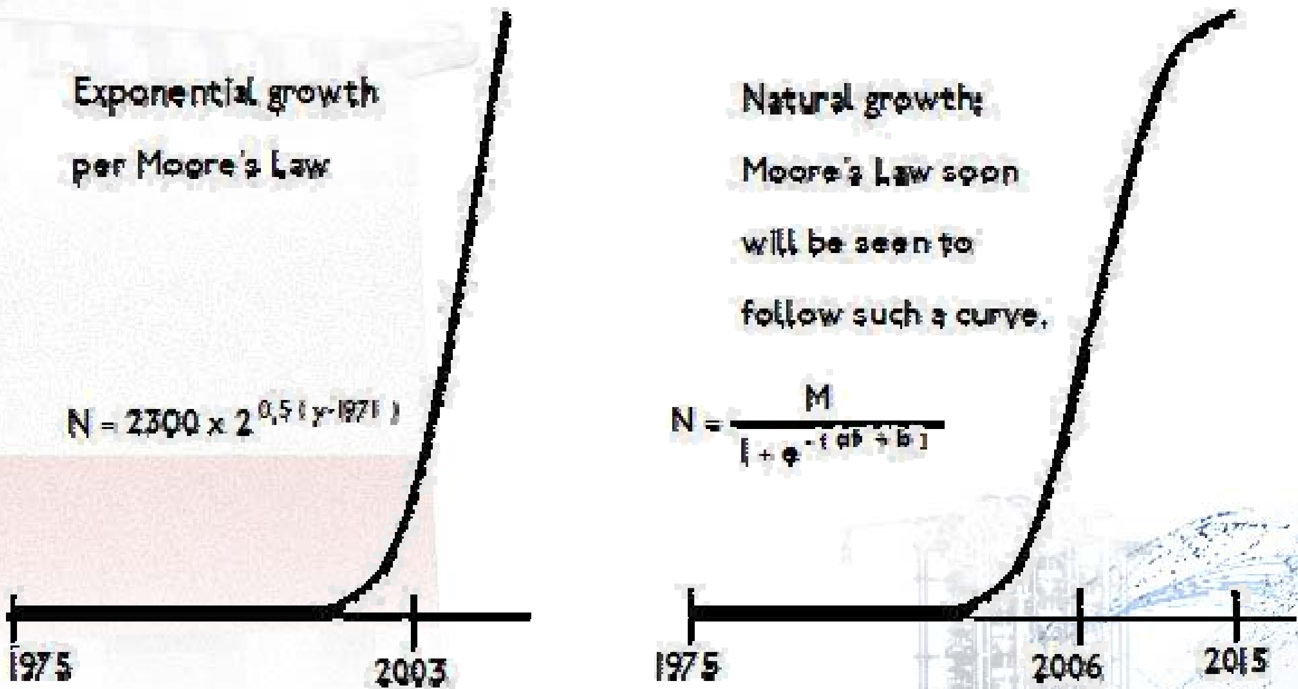
Introduction to Informatics

Lecture 25:

Computing Models – Limits of Computation

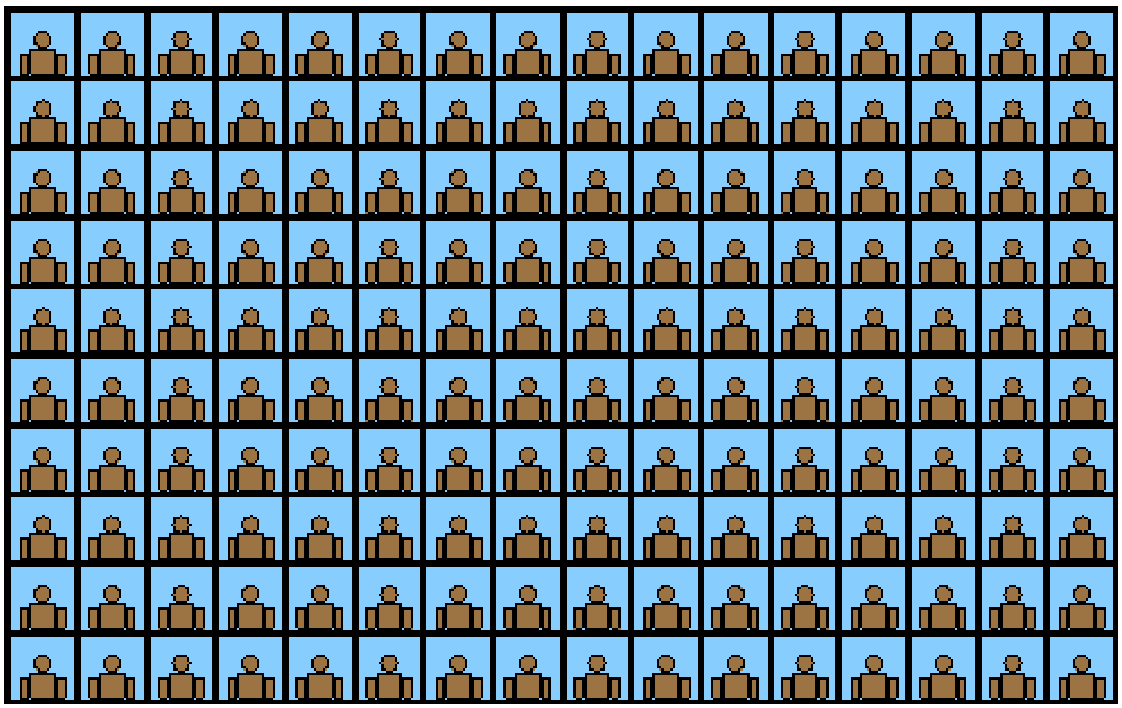
Group Assignment 2 Results



EXPONENTIAL GROWTH VERSUS NATURAL GROWTH

Luis M.Rocha and Santiago Schnell

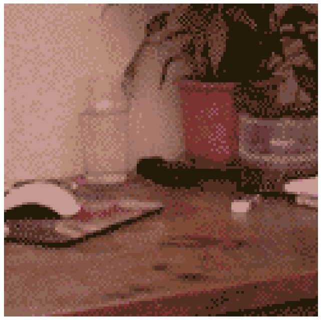
NO MORE LABS !!!



Exam Schedule

- 11595
  - Midterm
    - March 1st (Thursday)
      - Regular Class time
  - Final Exam
    - May 3rd (Thursday)
      - 7:15-9:15 p.m.

Readings until now



- Lecture notes
  - Posted online
    - http://informatics.indiana.edu/rocha/i101
      - *The Nature of Information*
      - *Technology*
      - *Modeling the World*
  - *@ infoport*
    - *http://infoport.blogspot.com*
  - From course package
    - Von Baeyer, H.C. [2004]. *Information: The New Language of Science*. Harvard University Press.
      - Chapters 1, 4 (pages 1-12)
      - Chapter 10 (pages 13-17)
    - From Andy Clark's book "*Natural-Born Cyborgs*"
      - Chapters 2 and 6 (pages 19 - 67)
    - From Irv Englander's book "*The Architecture of Computer Hardware and Systems Software*"
      - Chapter 3: Data Formats (pp. 70-86)
    - Klir, J.G., U. St. Clair, and B.Yuan [1997]. Fuzzy Set Theory: foundations and Applications. Prentice Hall
      - Chapter 2: Classical Logic (pp. 87-97)
      - Chapter 3: Classical Set Theory (pp. 98-103)
    - Norman, G.R. and D.L. Streinrt [2000]. *Biostatistics: The Bare Essentials*.
      - Chapters 1-3 (pages 105-129)
      - OPTIONAL: Chapter 4 (pages 131-136)
      - Chapter 13 (pages 147-155)
      - Chapter 5 (pages 141-144)
    - Igor Aleksander, "Understanding Information Bit by Bit"
      - Pages 157-166
    - Ellen Ullman, "Dining with Robots"
      - Pages 167-172

Assignment Situation

- Labs
  - Past
    - Lab 1: Blogs
      - Closed (Friday, January 19): Grades Posted
    - Lab 2: Basic HTML
      - Closed (Wednesday, January 31): Grades Posted
    - Lab 3: Advanced HTML: Cascading Style Sheets
      - Closed (Friday, February 2): Grades Posted
    - Lab 4: More HTML and CSS
      - Closed (Friday, February 9): Grades Posted
    - Lab 5: Introduction to Operating Systems: Unix
      - Closed (Friday, February 16): Grades Posted
    - Lab 6: More Unix and FTP
      - Closed (Friday, February 23): Grades Posted
    - Lab 7: Logic Gates
      - Closed (Friday, March 9): Grades Posted
    - Lab 8: Intro to Statistical Analysis using Excel
      - Closed (Friday, March 30): Grades Posted
    - Lab 9: Data analysis with Excel (linear regression)
      - Closed (Friday, April 6): Grades Posted
    - Lab 10: Simple programming in Excel and Measuring Uncertainty
      - April 12 and 13, Due April 20
- Assignments
  - Individual
    - First installment
      - Closed: February 9: Grades Posted
    - Second Installment
      - Past: March 2: Grades Posted
    - Third installment
      - Past: Grades Posted
    - Fourth Installment
      - Presented April 10th, Due April 20th
  - Group
    - First Installment
      - Past: March 9th, graded
    - Second Installment
      - Past: April 6th Graded
    - Third Installment
      - Presented Thursday, April 12; Due Friday, April 27

Group Assignment

- Second Installment: Given the text of “Lottery of Babylon” by Jorge Luis Borges
  - Measures of central tendency and dispersion of letter frequency
  - Probability of a letter being a vowel
  - Probability of a letter being a consonant
  - Conditional probability of letters ‘e’ and ‘u’
    - P(e|♥) where ♥ is the letter occurring before ‘e’
    - P(u|♥) where ♥ is the letter occurring before ‘u’
    - Compute for all letters (not space)
    - Produce histogram of P(e|♥), for all ♥.
    - Produce histogram of P(u|♥), for all ♥.
    - Discuss the independence of ‘e’ and ‘u’ from other letters
  - Upload to Oncourse

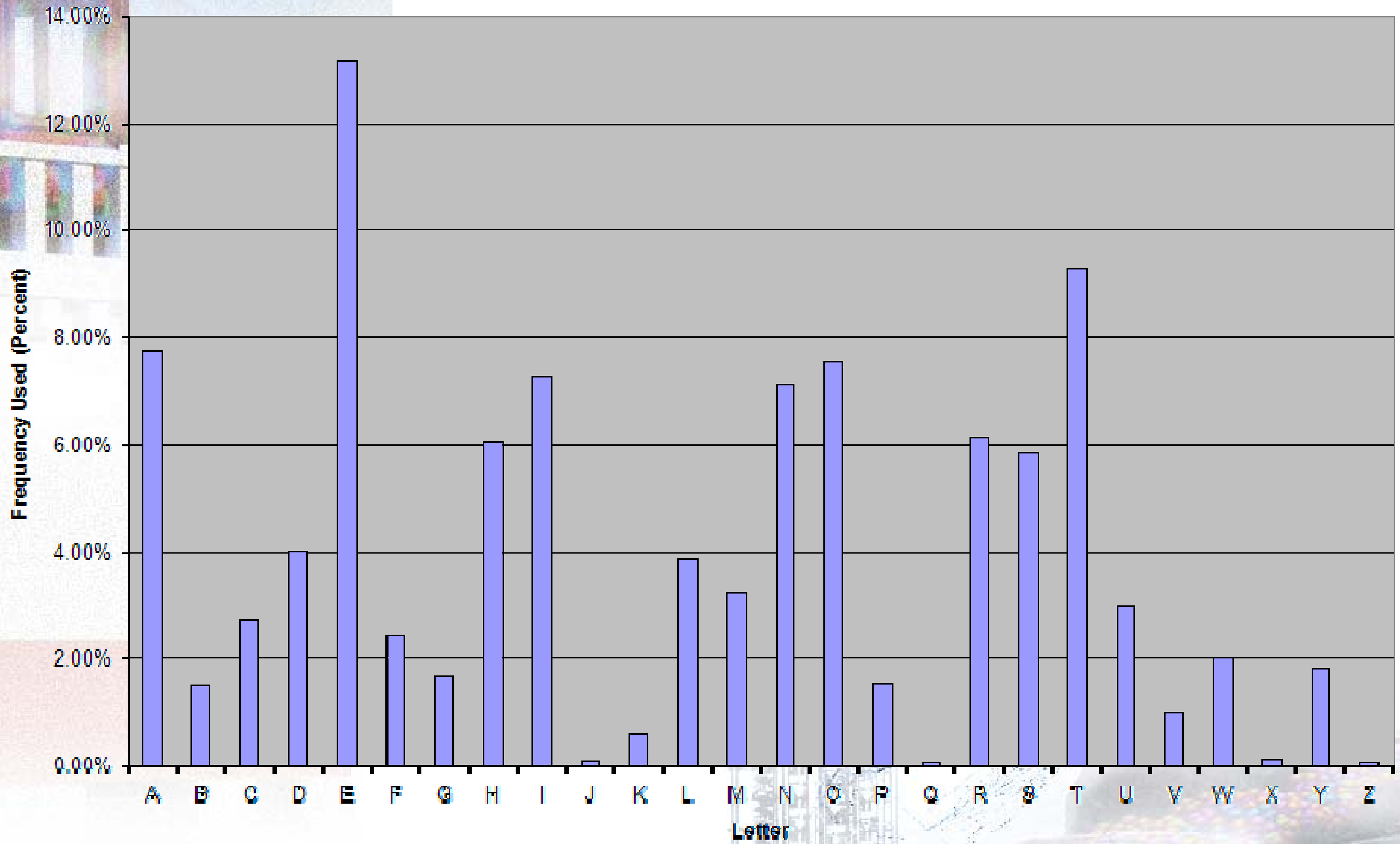
$$P(e \mid h) = \frac{|h \wedge e|}{|h|} = \frac{|'he'|}{|h|}$$

$$P(e) = \frac{|e|}{N}$$

Conditional Probability

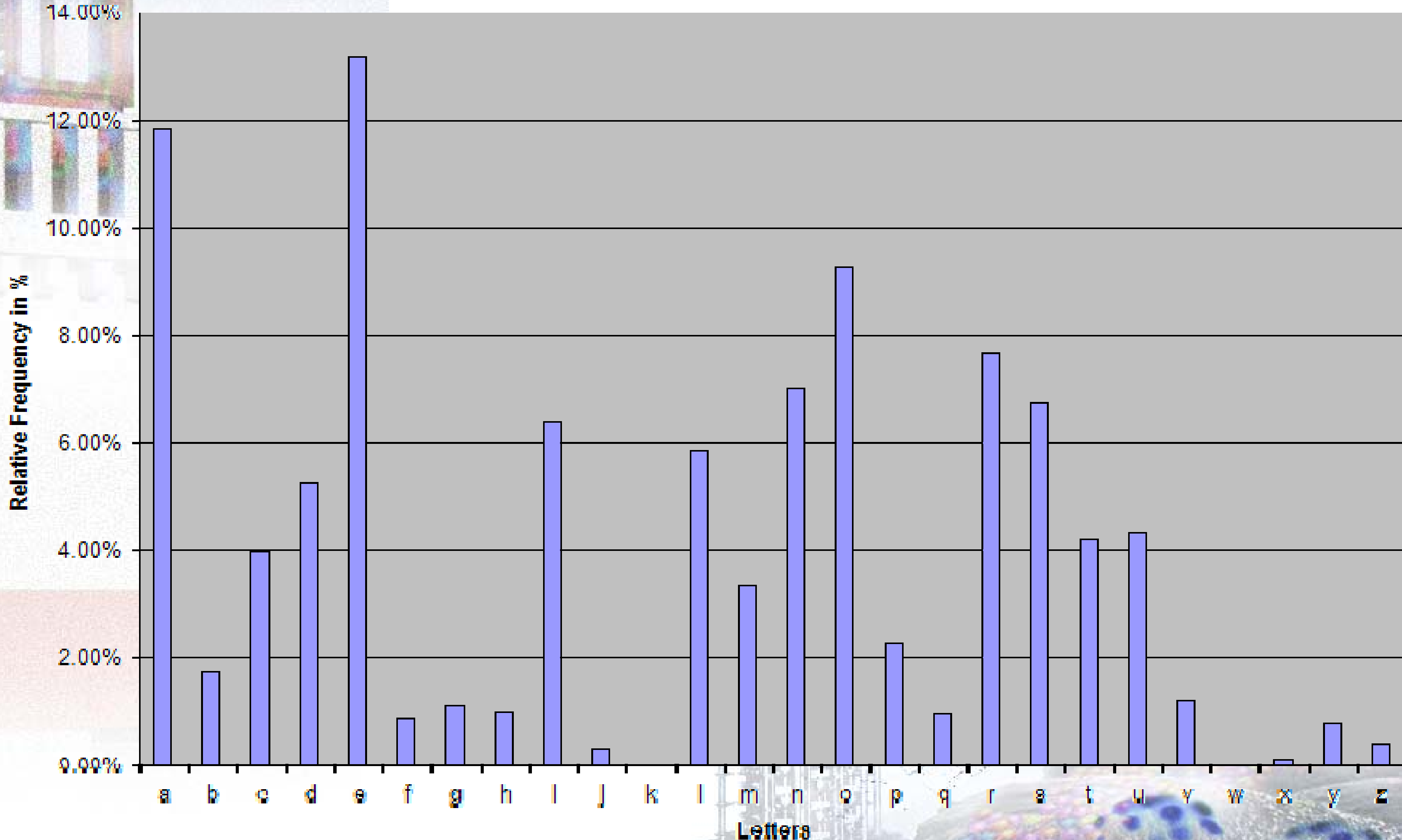
- $P(B|A) = |A \wedge B|/|A|$
- Multiplication Rule
  - $P(A \wedge B) = P(A) . P(B|A)$
- Two events A, B are ***independent*** if the occurrence of one has ***no effect on the probability*** of the occurrence of the other
  - $P(B|A) = P(B)$
  - Multiplication Rule
    - $P(A \wedge B) = P(A).P(B)$

Lottery of Babylon (English)



Luis M.Rocha and Santiago Schnell

La Loteria En Babilonia (Spanish)



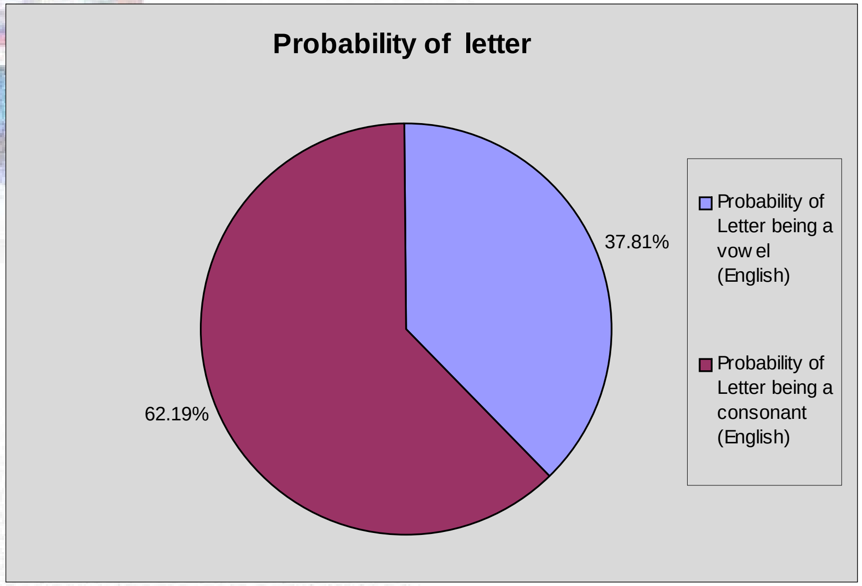
Measures of central tendency and dispersion of letter frequency

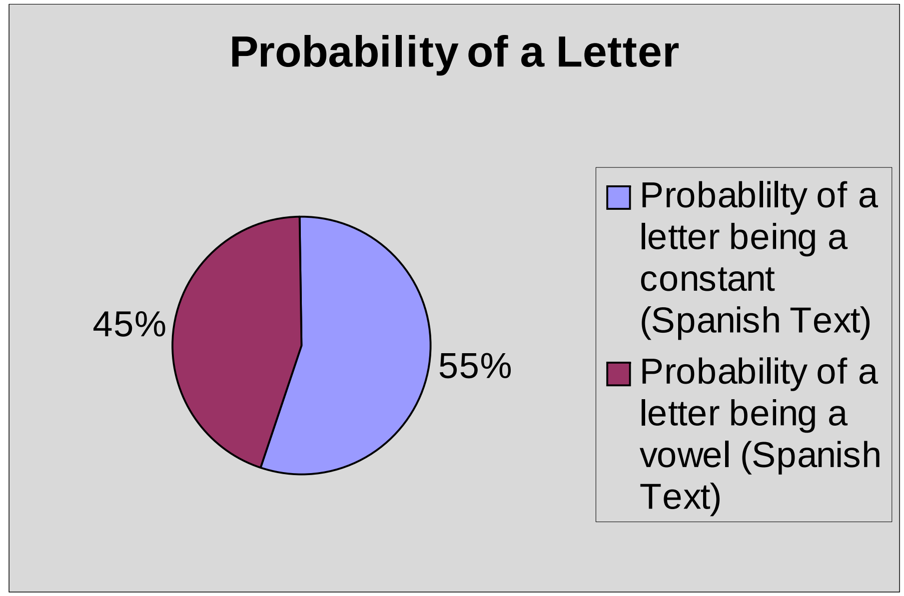
Max Cutler and Marc Epstein

	English	Spanish
Central Tendencies		
Mean	431.61538	323.40625
Median	5611.5 (M)	5175 (L)
Mode	1436 (E)	1348 (E)
Dispersion		
Range	1431	1348
Standard Deviation	377.1014799	380.1801661
Variance	142205.5262	144536.9587

Probability of letter being vowel or consonant

John Oglesby and Sarah Kepa

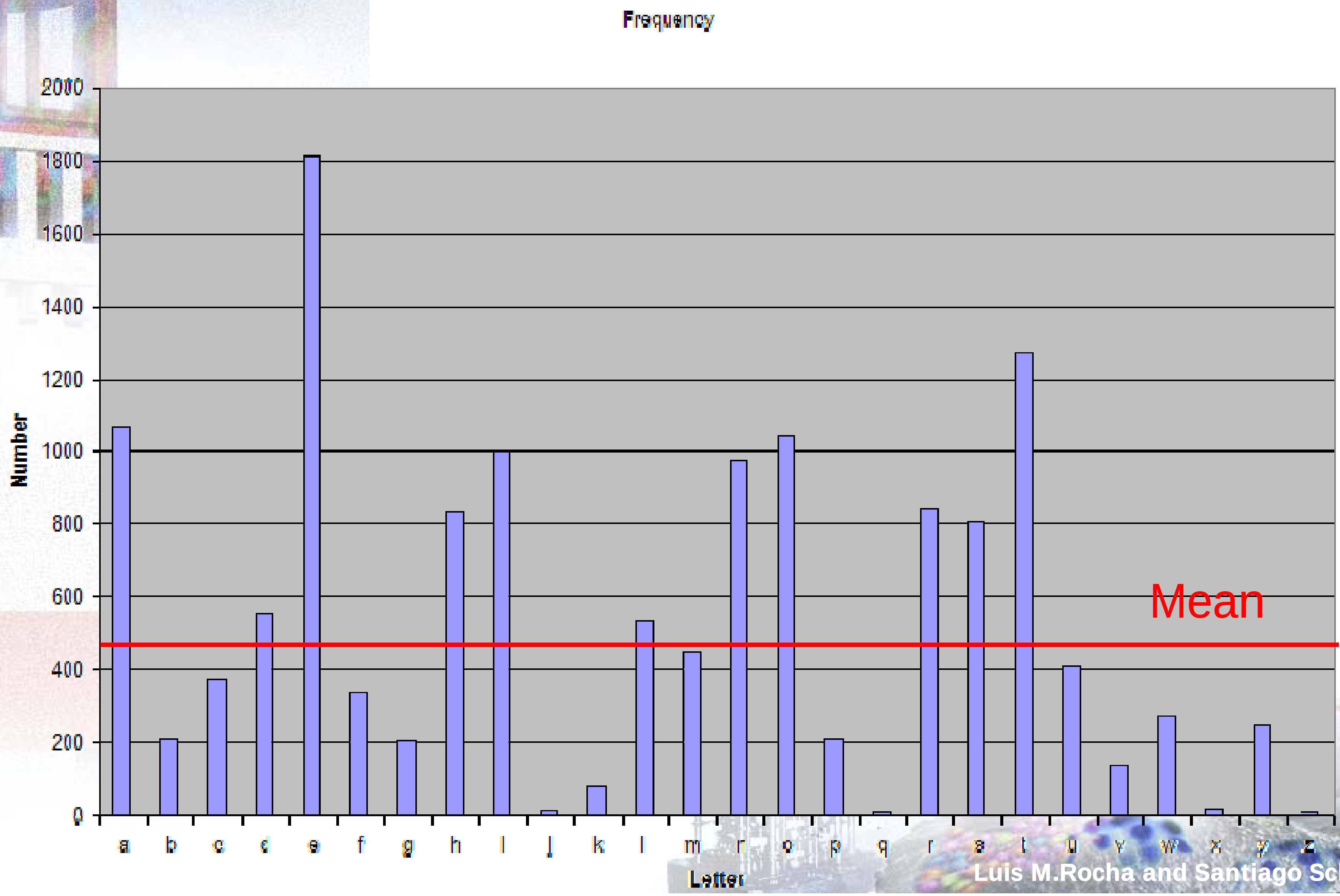




English

Spanish

Lottery of Babylon (English)



Max Cutler and Marc Epstein

English

P(e|a)	0.002814815	0.088888889
P(e|b)	0.323782678	0.085427136
P(e|c)	0.212041885	0.098078759
P(e|d)	0.193286715	0.480842912
P(e|e)	0.027810412	0.082585088
P(e|f)	0.055885968	0.149425287
P(e|g)	0.104046243	0.076888889
P(e|h)	0.443867089	0.290697674
P(e|i)	0.084482759	0.088718851
P(e|j)	0.4375	0.888888888
P(e|k)	0.383988051	0
P(e|l)	0.164556962	0.118556701
P(e|m)	0.273846154	0.222972973
P(e|n)	0.111656442	0.099087858
P(e|o)	0.003512679	0.049792531
P(e|p)	0.17826087	0.142857143
P(e|q)	0	0
P(e|r)	0.252288806	0.196261682
P(e|s)	0.103524684	0.148564856
P(e|t)	0.093874889	0.266514806
P(e|u)	0.083928077	0.216258852
P(e|v)	0.737864078	0.582022472
P(e|w)	0.147619048	0
P(e|x)	0.230769231	0
P(e|y)	0.052173913	0.120689655
P(e|z)	0.6	0.02173913

P_E('e') = 0.13

Spanish

P_E('e') = 0.13

Max Cutler and Marc Epstein

English

P(u|a)	0.006944444	0.024608175
P(u|b)	0.079207921	0.015075377
P(u|c)	0.044502618	0
P(u|d)	0	0.024904215
P(u|e)	0.006954103	0.015960485
P(u|f)	0.059288538	0.108448276
P(u|g)	0.046242775	0.201388889
P(u|h)	0.011075949	0.058139535
P(u|i)	0.002554278	0.00152207
P(u|j)	0.25	0.888888889
P(u|k)	0	0
P(u|l)	0.040084388	0.02288677
P(u|m)	0.012307692	0.084745763
P(u|n)	0.020858896	0.081290748
P(u|o)	0.101488297	0.006224066
P(u|p)	0.060869565	0.069498069
P(u|q)	0.888888889	0.980952381
P(u|r)	0.018482836	0.018851135
P(u|s)	0.087288622	0.068118812
P(u|t)	0.026525199	0.072892988
P(u|u)	0	0
P(u|v)	0.019417476	0.011235955
P(u|w)	0.004761905	0
P(u|x)	0	0
P(u|y)	0.013043478	0.034482759
P(u|z)	0	

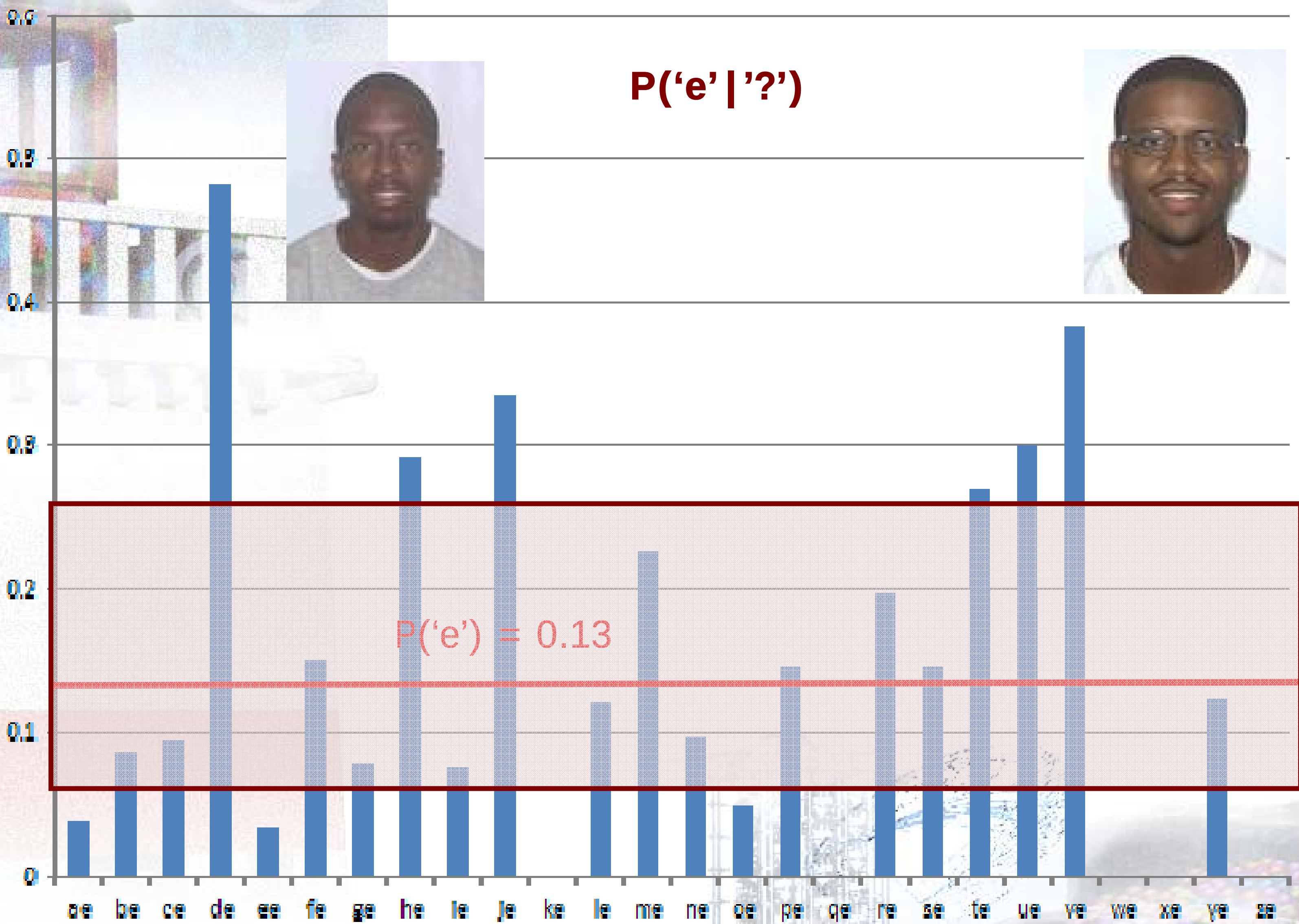
P_E('u') = 0.03

Spanish

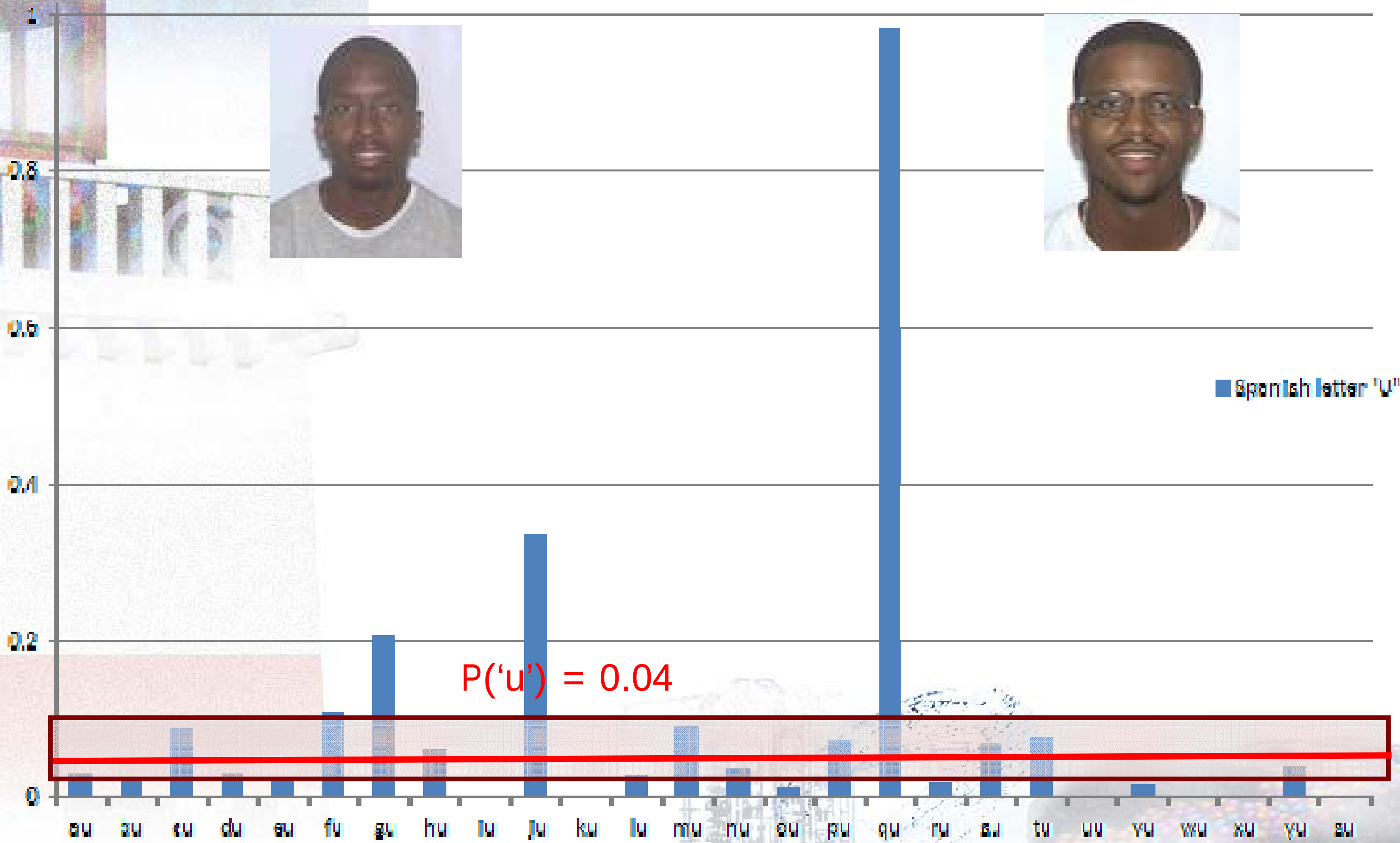
Andrew Dempsey and JT Waugh

P_S('u') = 0.04

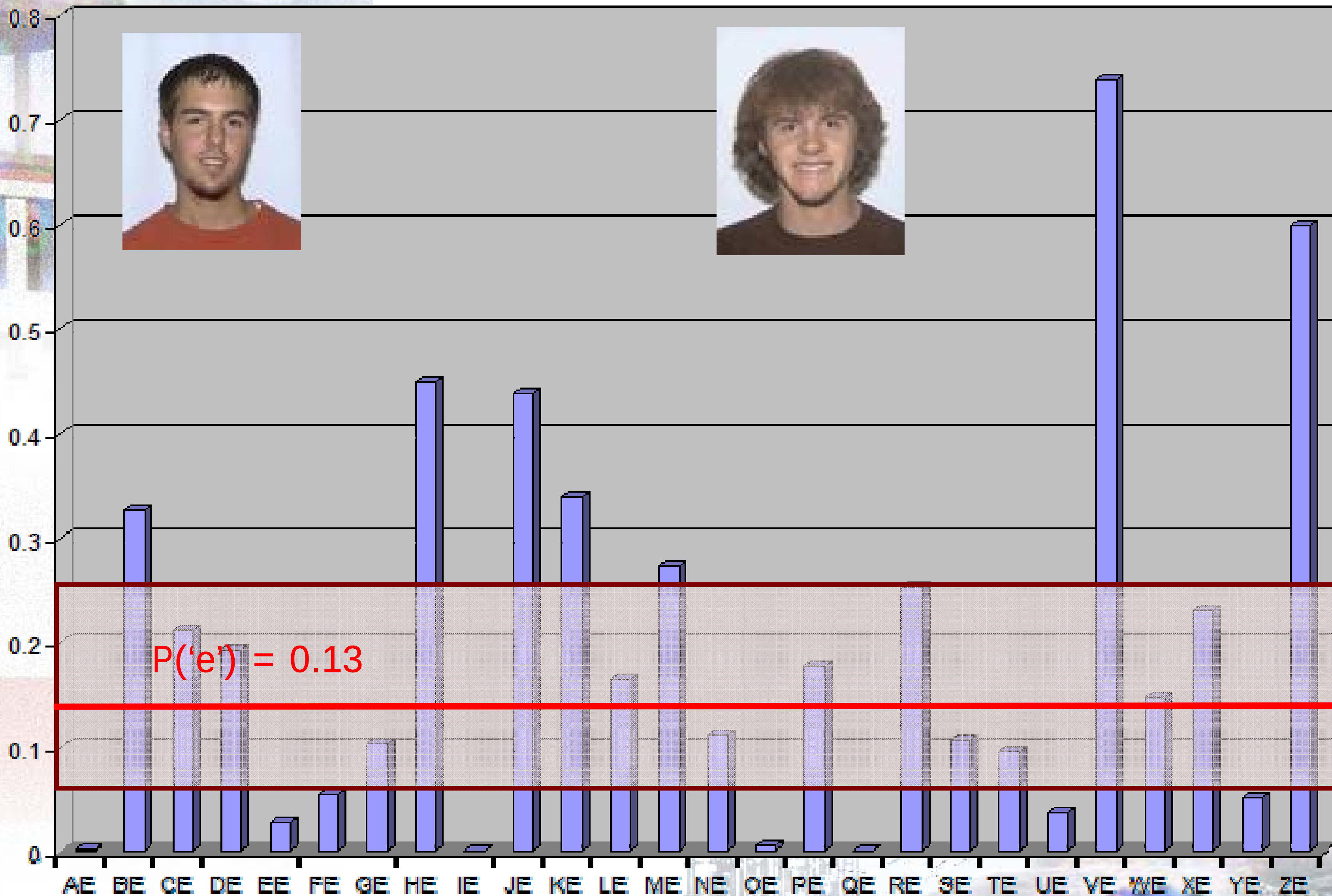
Marcus Bigbee & Brandon Smith ('e' Spanish)



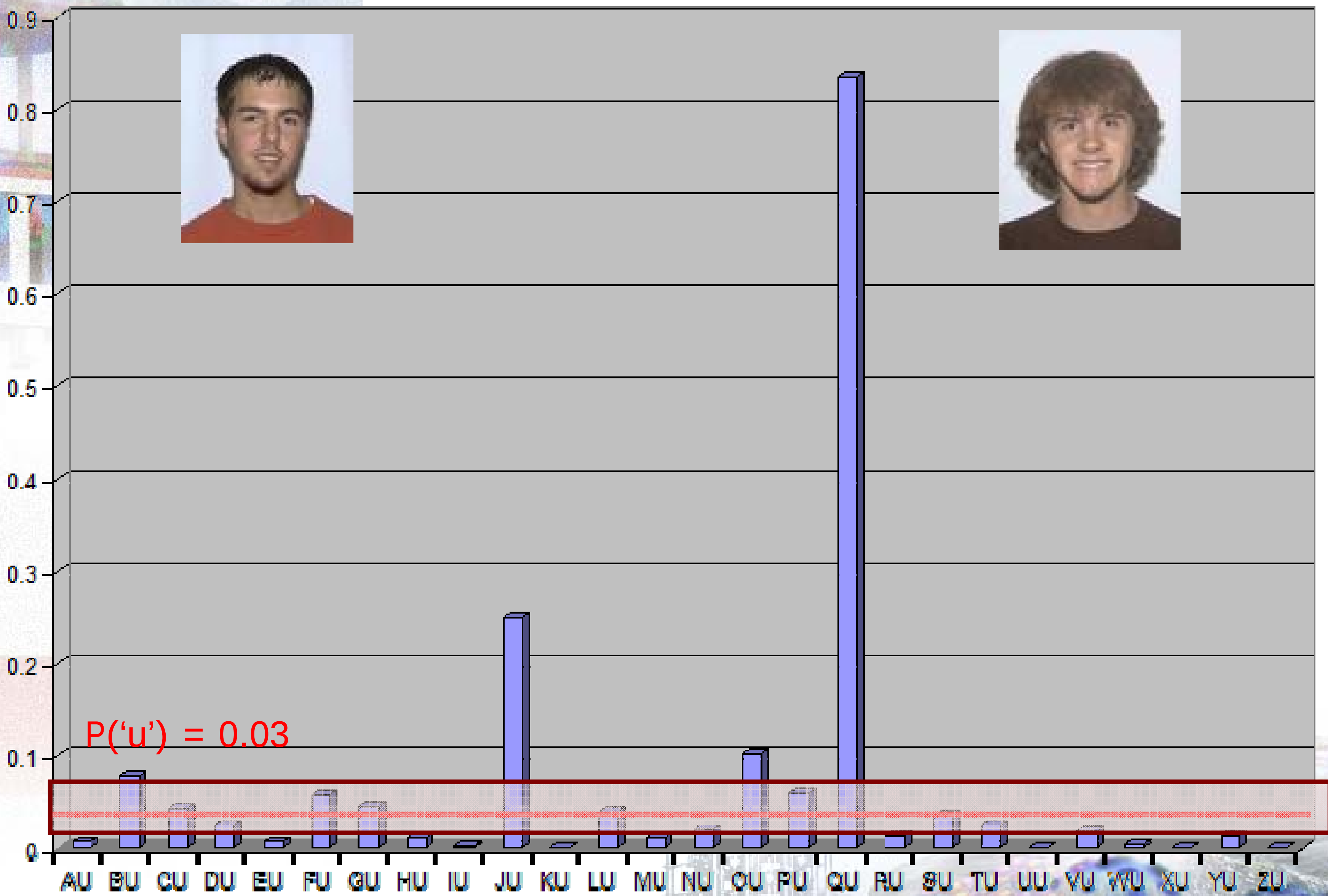
Marcus Bigbee & Brandon Smith('u' Spanish)



Craig Bauer & Chris Kremser ('e' English)



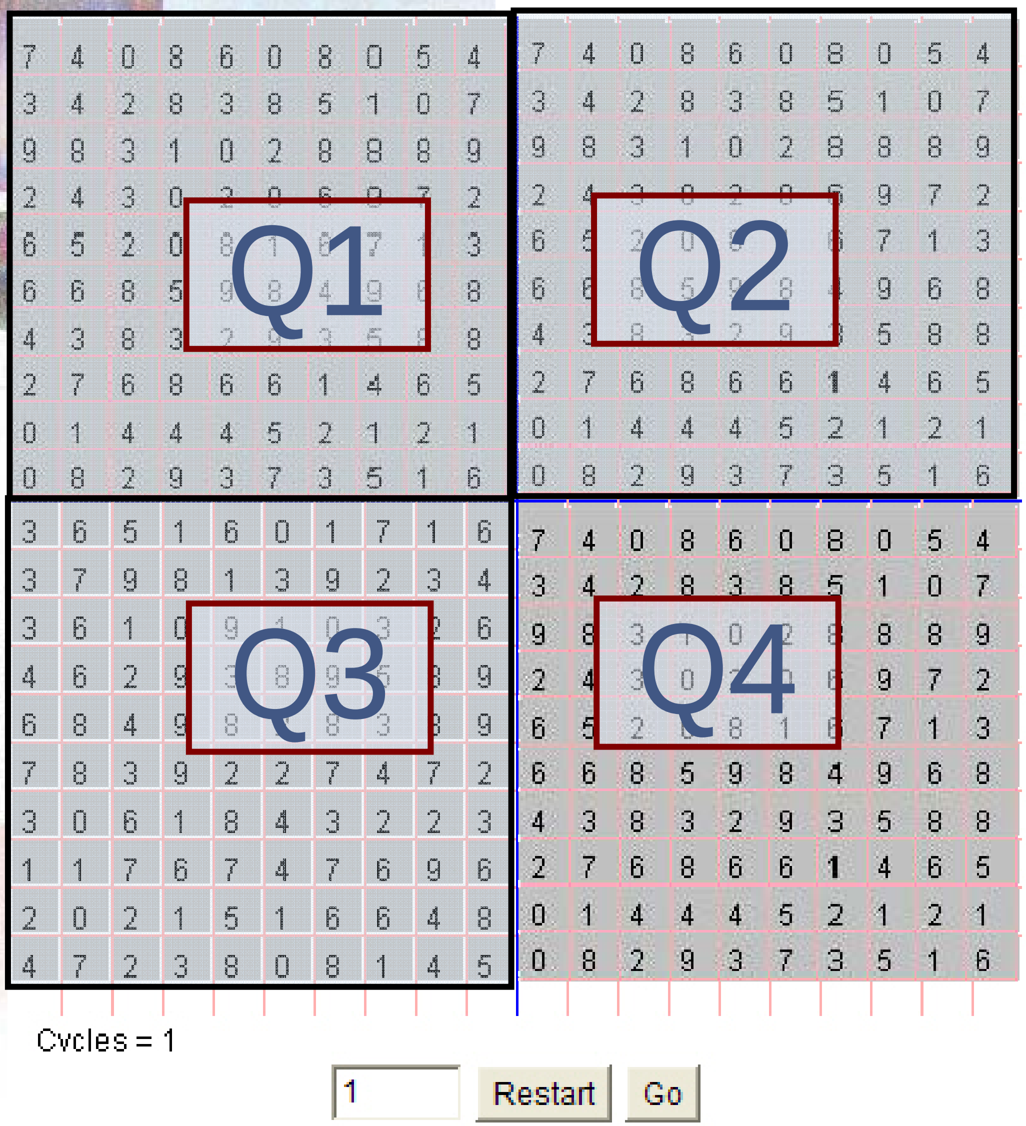
Craig Bauer & Chris Kremser('u' English)



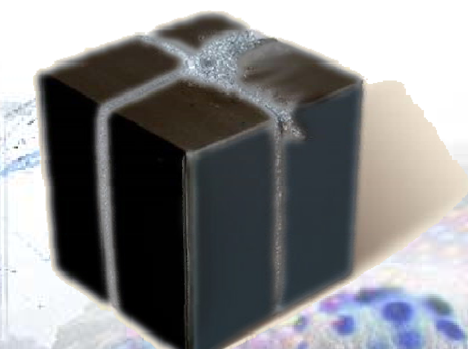
Group Assignment

- Third Installment
  - Given any text such as the *library of babylon* or *Funes, the memorious*
    - Create a ***database model*** and a ***relational database instance*** using *Microsoft Access* to store the data and conclusions from previous installments
      - Use the entity-relationship model
    - Examples of items that should appear
      - Title, author, language, publication date
      - Frequency/probability of each letter
      - Conditional probabilities for letters ‘e’ and ‘u’ (as produced in installment 2)
      - Positively and negatively dependent letters
    - Use at least 4 texts
  - Due on April 27th, 2005
  - Upload to Oncourse

Individual Assignment – Part IV

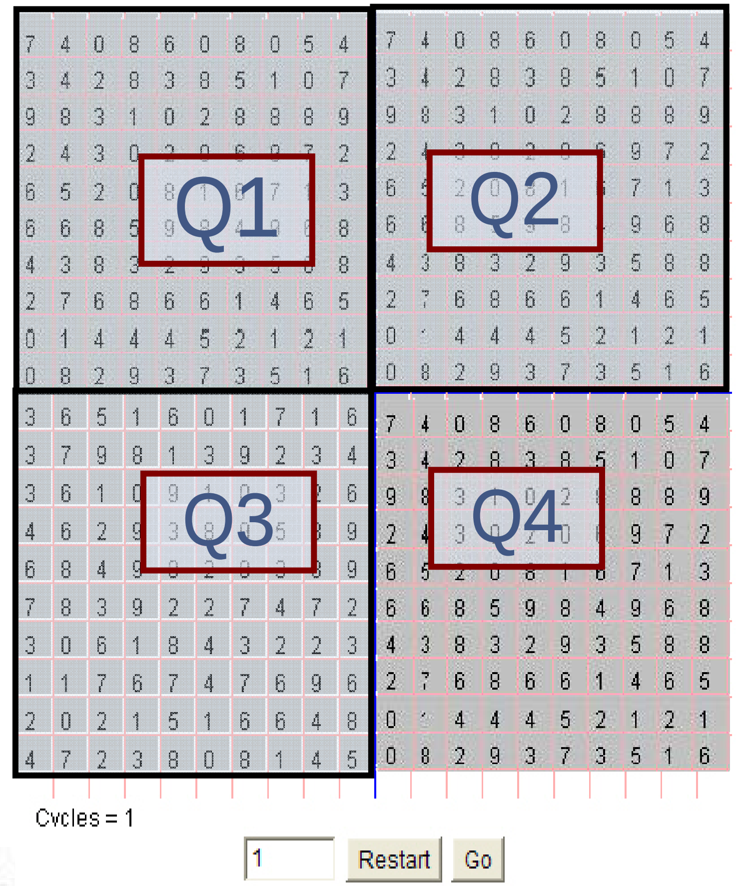


- Step by step analysis of “dying” squares
  - 4th Installment
    - Presented: April 10th
    - Due: April 20th
- Use inductive and deductive reasoning
  - To uncover the algorithm in each quadrant
    - Build from inductive knowledge accumulated so far



Summary of Black Box

- Quadrant 1
  - At the random initial state
    - All numbers have equal probability of being initially present
    - But the probability of changes are different
- In Any State
    - Any number changes depending on its neighbors
    - It ‘gravitates’ towards the smallest number that it ‘sees’ most often.
    - Odd and Even numbers do not show different behavior
- What is the Algorithm?



Summary of Black Box

- Quadrant 3
  - At the random initial state
    - All numbers have equal probability of being initially present
    - But the probability of changes are different
- In Any State
  - 0 can only change to 0
  - 5 can only change to 5 or 0
  - Even digits always change to even digits
  - Odd digits could change to any other digit
- What is the Algorithm?

	n(i)	p(i)
0	27	0.27
1	4	0.04
2	12	0.12
3	4	0.04
4	12	0.12
5	9	0.09
6	12	0.12
7	4	0.04
8	12	0.12
9	4	0.04

1. $0 \rightarrow 0$
2. $\{5\} \rightarrow \{0, 5\}$
3. $\{2, 4, 6, 8\} \rightarrow \{0, 2, 4, 6, 8\}$
4. $\{1, 3, 7, 9\} \rightarrow$ $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$

Summary of Black Box

- Quadrant 2
  - At the random initial state
    - All numbers have equal probability of being initially present
    - But the probability of changes are different
- In Any State
    - 0 can only change to 0
    - 5 can only change to 5 or 0
    - Even digits always change to even digits
    - Odd digits could change to any other digit
- What is the Algorithm?

1. $0 \rightarrow 0$
2. $\{5\} \rightarrow \{0, 5\}$
3. $\{2, 4, 6, 8\} \rightarrow \{0, 2, 4, 6, 8\}$
4. $\{1, 3, 7, 9\} \rightarrow$
   $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$

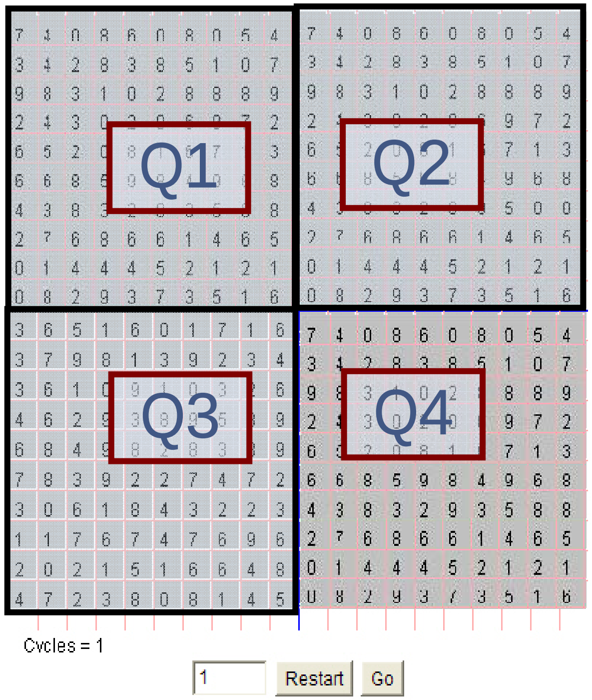
Luis M.Rocha and Santiago Schnell

Possible Operations Q2 and Q3

Operator	Meaning	Excel	Example
()	Brackets, grouping	()	y = (a + b) * (c + d)
*	Multiplication	*	i=j*k
+	Add	+	i = i+1
-	Subtract	-	i=j-3.2
/	Real division	/	i=8/5 = 1.6
div	Integer division	Quotient (a,b)	i=8/5 = 1
Mod, %	remainder	Mod (a, b)	i=8 mod 5 = 3
ROUND	Rounds	ROUND (a, d)	i=ROUND(3.67,0) = 4
INT	Integer Part	INT	i=INT(3.67) = 3
rand	Random number	Rand() RandBetween(a,b)	i=rand(n)

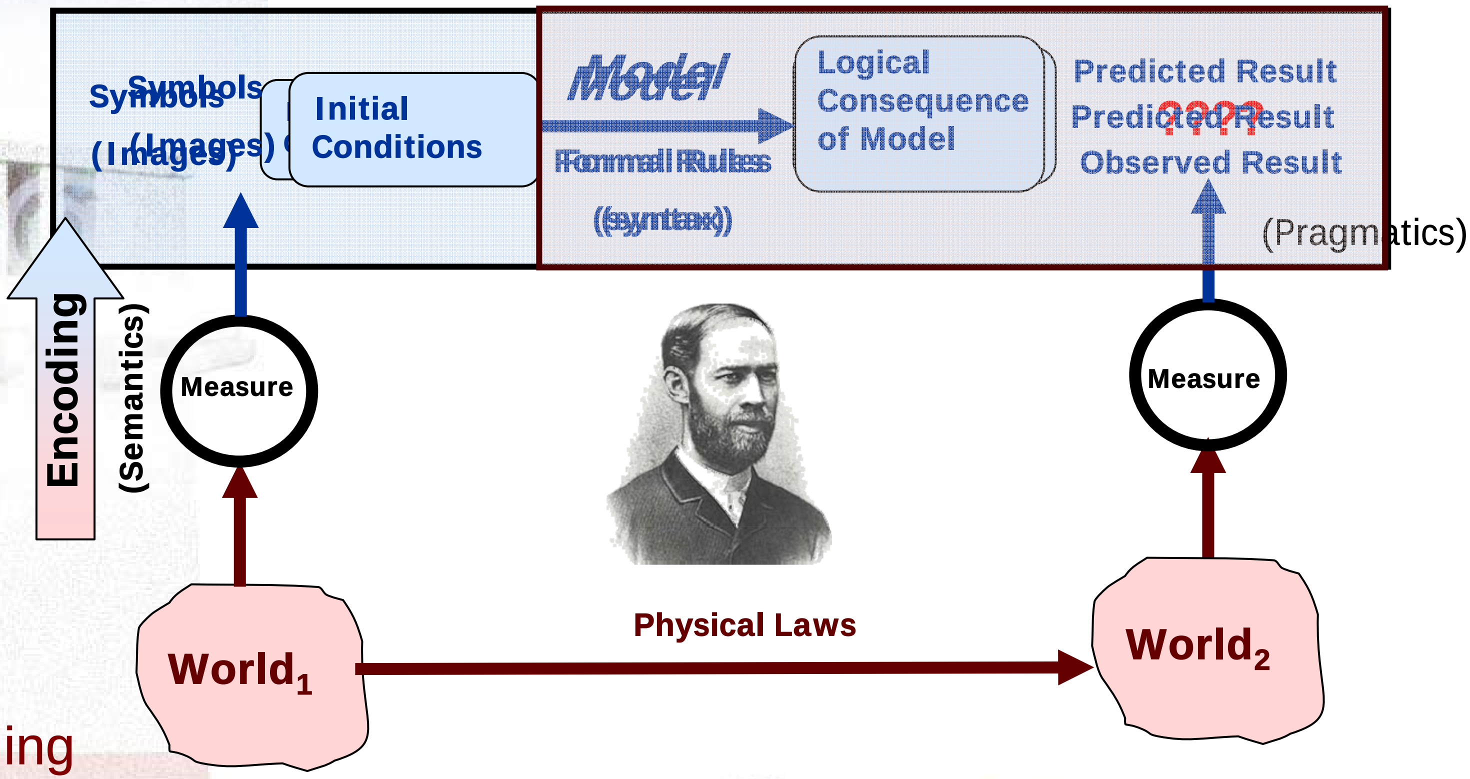
Tip for Individual Assignment

- Quadrant Q
  - There are 100 cells in each 10x10 quadrant
    - C = 1...100
  - Each cell can take one of 10 colors
    - V(C)=0..9
      - is the value of the cell
      - This is the state cell C is in
  - Random initialization of quadrant Q at cycle 1
    - For c=1 to 100 do
    - V(C) ← randbetween(0,9) {random number 0 to 9}
    - EndFor
    - Cycle ← 1
  - Run for Number of cycles
    - n ← Input dialog
    - For k=1 to n do
    - Cycle ← cycle+1
    - {Pick random cell}
    - C ← randbetween(1,100)
    - {Update the value of the cell (NOT THE REAL THING)}
    - V(C) ← ((V(C) * randbetween(0,9)) div 2) – 5***x**
    - EndFor
  - X may be a hidden variable
    - X ← ???



The Modeling Relation

Hertz' Modeling Paradigm



- Modeling
  - Compute hypothesis
- Rules from Inductive and Deductive Analysis
  - From Data analysis
  - Produce Conclusions

Types of Problems

- Algorithms are for Solving Problems
- Types of Problems
  - Search
    - Find an X in input satisfying property Y
      - Find a prime number in a random sequence of numbers
  - Structuring Problems
    - Transform the input to satisfy property Y
      - Sort a random sequence of numbers
  - Construction Problems
    - Build an X satisfying property Y
      - Generate a random sequence of numbers with a given mean and standard deviation
  - Optimization Problems
    - Find the best X satisfying property Y
      - Find the largest prime number in a given sequence
  - Decision Problems
    - Decide whether the input satisfies property Y
      - Is the input number a prime number
  - Adaptive Problems
    - Maintain property Y over time
      - Grow a sequence of numbers such that there are always m prime numbers with a given mean and standard deviation

Problem Difficulty

- Conceptually Hard Problem
  - No algorithm exists to solve the problem
- Analytically Hard Problem
  - An algorithm exists to solve the problem, but we don't know how long it will take to solve every instance of the problem
- Computationally Hard Problem
  - An algorithm exists to solve the problem, but relatively few instances take millions of years to solve
    - Problems we know to be
    - problems we suspect to be
- Computationally unsolvable Problem
  - No algorithm can exist to solve the problem

Artificial Intelligence

Algorithmic Complexity Theory

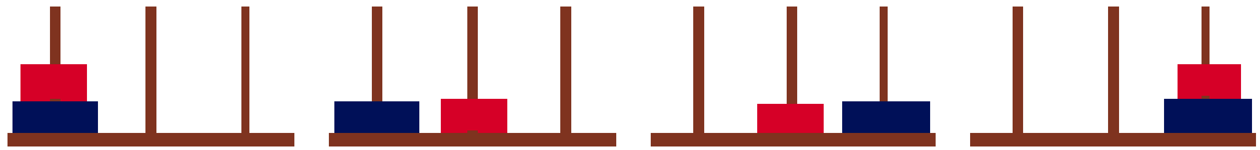
Computability Theory

Hanoi Problem

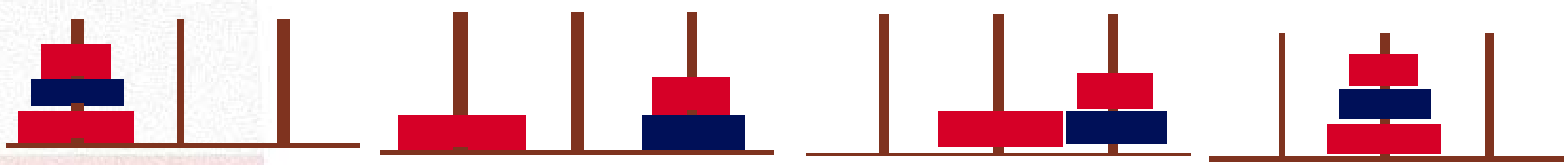
- Invented by French Mathematician Édouard Lucas in 1883
  - *At the Tower of Brahma in India, there are three diamond pegs and sixty-four gold disks. When the temple priests have moved all the disks, one at a time preserving size order, to another peg the world will come to an end*.
    - If the priests can move a disk from one peg to another in one second, how long does the World have to exist?

Solving the Hanoi Problem

- Solve for the smallest instances and then try to generalize
  - N=2



  - N=3



0 3 4 7

Use Hanoi_2 (H2) as building block (of 3 moves)
H3 uses H2 twice, plus 1 move of the largest disk

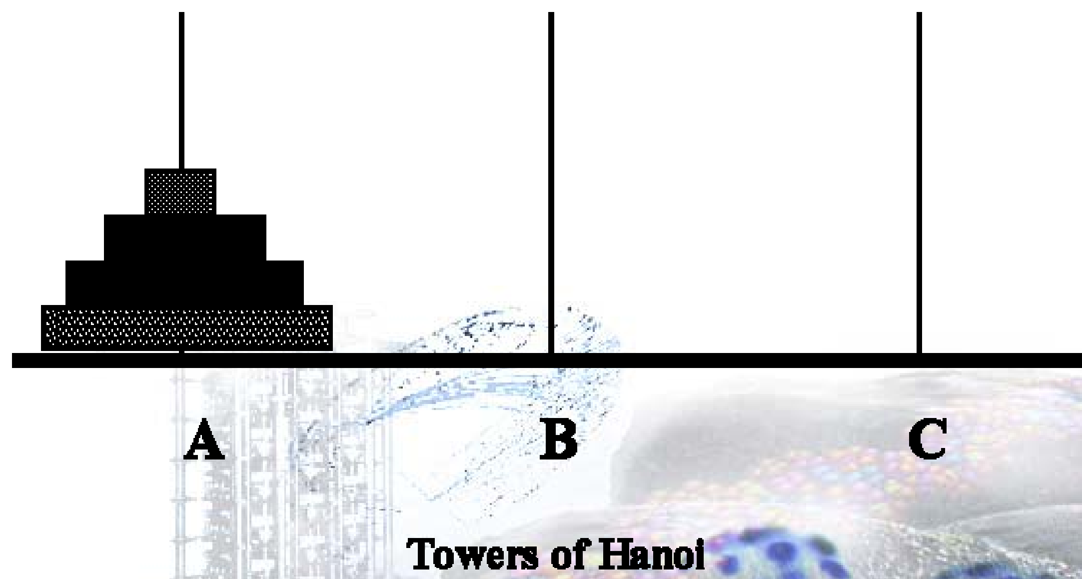
Hanoi Problem for *n* disks

Use Hanoi_2 (H2) as building block (of 3 moves)

H3 uses H2 twice, plus 1 move of the largest disk

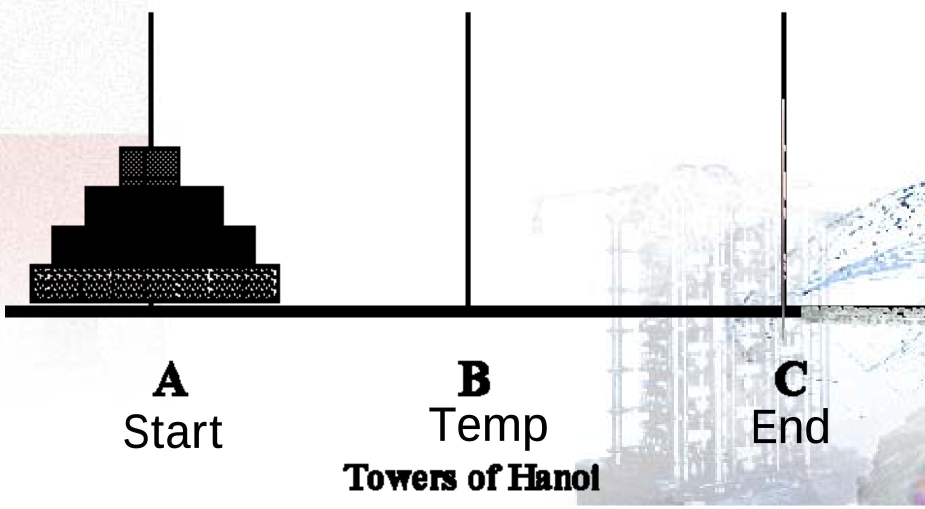
- Algorithm to move n disks from A to C
  - Move top n-1 disks from A to B
  - Move biggest disk to C
  - Move n-1 disks on B to C
- Recursion
  - Until H2

An Algorithm that uses itself to solve a problem



Pseudocode for Hanoi Problem

- Hanoi (*Start*, *Temp*, *End*, n)
  - If n = 1 then
    - Move *Start*'s top disk to *End*
  - Else
    - Hanoi (*Start*, *End*, *Temp*, n-1)
    - Move *Start*'s top disk to *End*
    - Hanoi (*Temp*, *Start*, *End*, n-1)



Computational Complexity

$$2^{10}(KILO) = 1{,}024$$
$$2^{20}(MEGA) = 1{,}048{,}576$$
$$2^{30}(GIGA) = 1{,}073{,}741{,}824$$
$$2^{40}(TERA) = 1{,}099{,}511{,}627{,}776$$
$$2^{64} = 18{,}446{,}744{,}073{,}709{,}551{,}616$$

585 billion years in seconds!!!!!!!!

Earth: 5 billion years

Universe: 15 billion years

Fastest Computer: 135.5 teraflops - 135.5 trillion calculations a second (aprox 2^{47} moves a second)

2^{17} s needed = 36 hours

- Resources required during computation of an algorithm to solve a given problem
  - Time
    - how many steps does it take to solve a problem?
  - Space
    - how much memory does it take to solve a problem?
- The Hanoi Towers Problem
  - $f(n)$ is the number of times the HANOI algorithm moves a disk for a problem of n disks
    - $f(1)=1, f(2)=3, f(3)=7$
    - $f(n)= f(n-1) + 1 + f(n-1) = 2 \times f(n-1) + 1$
  - Every time we add a disk, the time to compute is at least double
    - $f(n) = 2^n - 1$

IBM Blue Gene/L

MDGRAPE-3

Fastest Computer (june 2006): 1 petaflop !!! – 1 quadrillion calculations per second --- MDGRAPE-3 @ Riken, Japan --- aprox 2^{14} s needed = 4.6 hours for Hanoi problem (assuming one disk change per operation)

Fastest Computer (late 2005): 280.6 teraflops - 280.6 trillion calculations a second --- Approaching petaflops: 3 petaflops in late 2006????

Fastest Computer (early 2005): 135.5 teraflops - 135.5 trillion calculations a second --- Approaching petaflops: 2^{50}

Bremermann's Limit

- Physical Limit of Computation
  - Hans Bremmermann in 1962
  - "no data processing system, whether artificial or living, can process more than 2×10^{47} bits per second per gram of its mass."
    - Based on the idea that information could be stored in the energy levels of matter
    - Calculated using Heisenberg's uncertainty principle, the Hartley measure, Planck's constant, and Einstein's famous $E = mc^2$ formula
  - A computer with the mass of the entire Earth and a time period equal to the estimated age of the Earth
    - **would not be able to process more than about 10^{93} bits**
  - ***transcomputational problems***

Transcomputational Problems

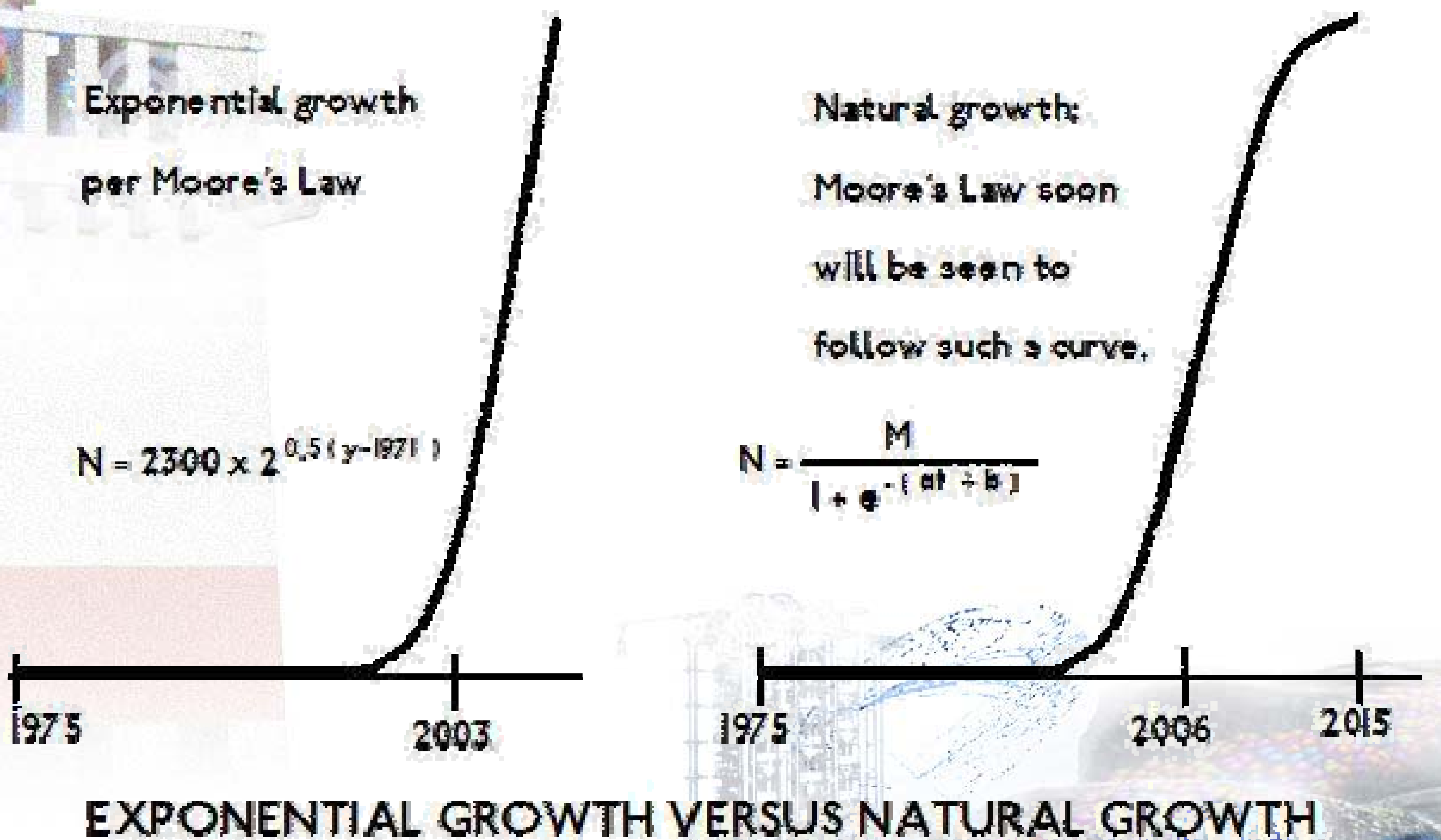
- A system of *n* variables, each of which can take *k* different states
  - k^n possible system states
  - When is it larger than 10^{93}?

k	2	3	4	5	6	7	8	9	10
n	308	194	154	133	119	110	102	97	93

7	4	0	8	6	0	8	0	5	4
3	4	2	8	3	8	5	1	0	7
9	8	3	1	0	2	8	8	8	9
2	4	3	0	2	0	6	9	7	2
6	5	2	0	8	1	6	7	1	3
6	6	8	5	9	8	4	9	6	8
4	3	8	3	2	9	3	5	8	8
2	7	6	8	6	6	1	4	6	5
0	1	4	4	4	5	2	1	2	1
0	8	2	9	3	7	3	5	1	6

  - Pattern Recognition
    - Grid of $n = q^2$ squares of k colors
    - Blackbox: 10^{100} possible states!
    - The human retina contains a million light-sensitive cells
  - Large scale integrated digital circuits
    - K= 2 (bits): a circuit with 308 inputs and one output!
- Complex Problems need simplification!

What happens to Moore's law ?



Luis M.Rocha and Santiago Schnell

Next Class!

- Topics
  - Databases and SQL
- Readings for Next week
  - @ *infoport*
  - From course package
    - Igor Aleksander, "Understanding Information Bit by Bit"
      - Resources tab in onCourse.
    - Ellen Ullman, "Dining with Robots"
      - Resources tab in onCourse.
- No More Labs!!!!!!!