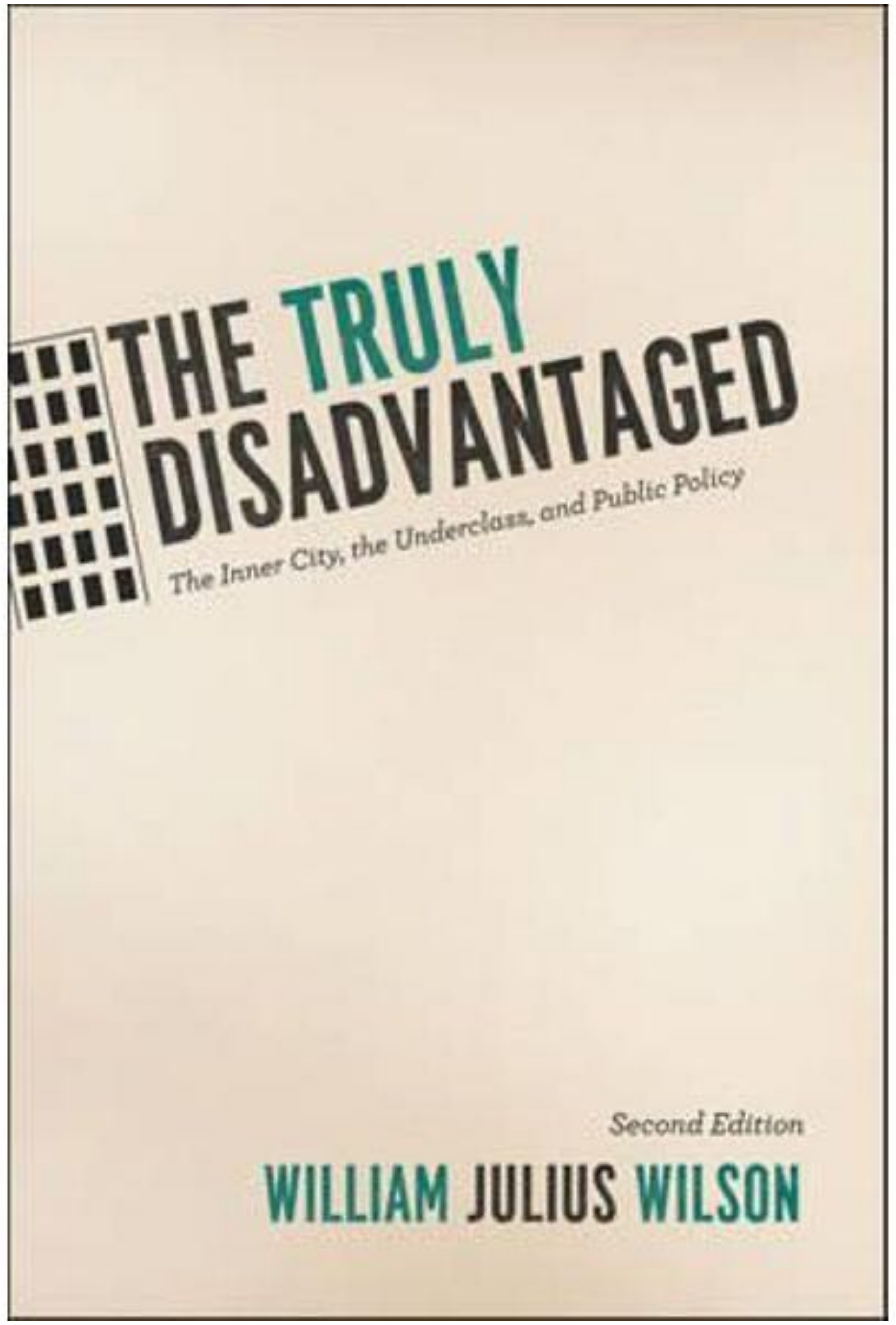
Lecture 11

Looking inside Cities: Spatial Structure and Neighborhoods

11.1 Information, Spatial Selection & the Statistics of Neighborhoods

IUS 6.3



William Julius Wilson, 1987

In "The Truly Disadvantaged," **Mr. Wilson takes on conservatives, liberals and civil rights leaders alike as he develops persuasive alternative explanations of what has gone wrong in the inner city and supports them with extensive data and research.** For example, he explores the sharpening of class lines among blacks; the proportions of black men with annual incomes over $25,000 and those with annual incomes under $5,000 have both increased. Inequality of income is greater now among black families than among white families.

NYtimes review

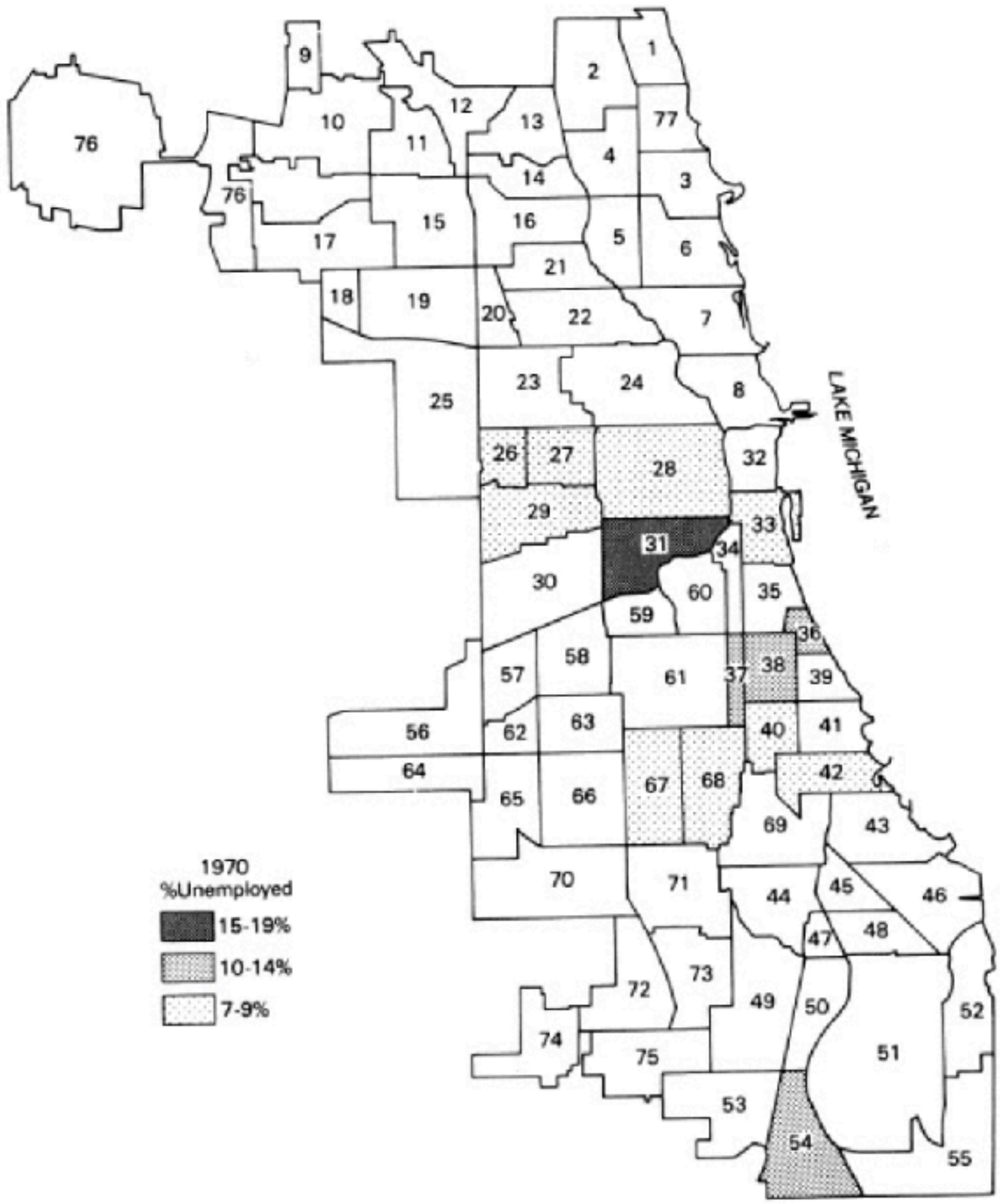


Figure 2.5. Unemployment rates in Chicago Community Areas, 1970.

Source: see fig. 2.3.

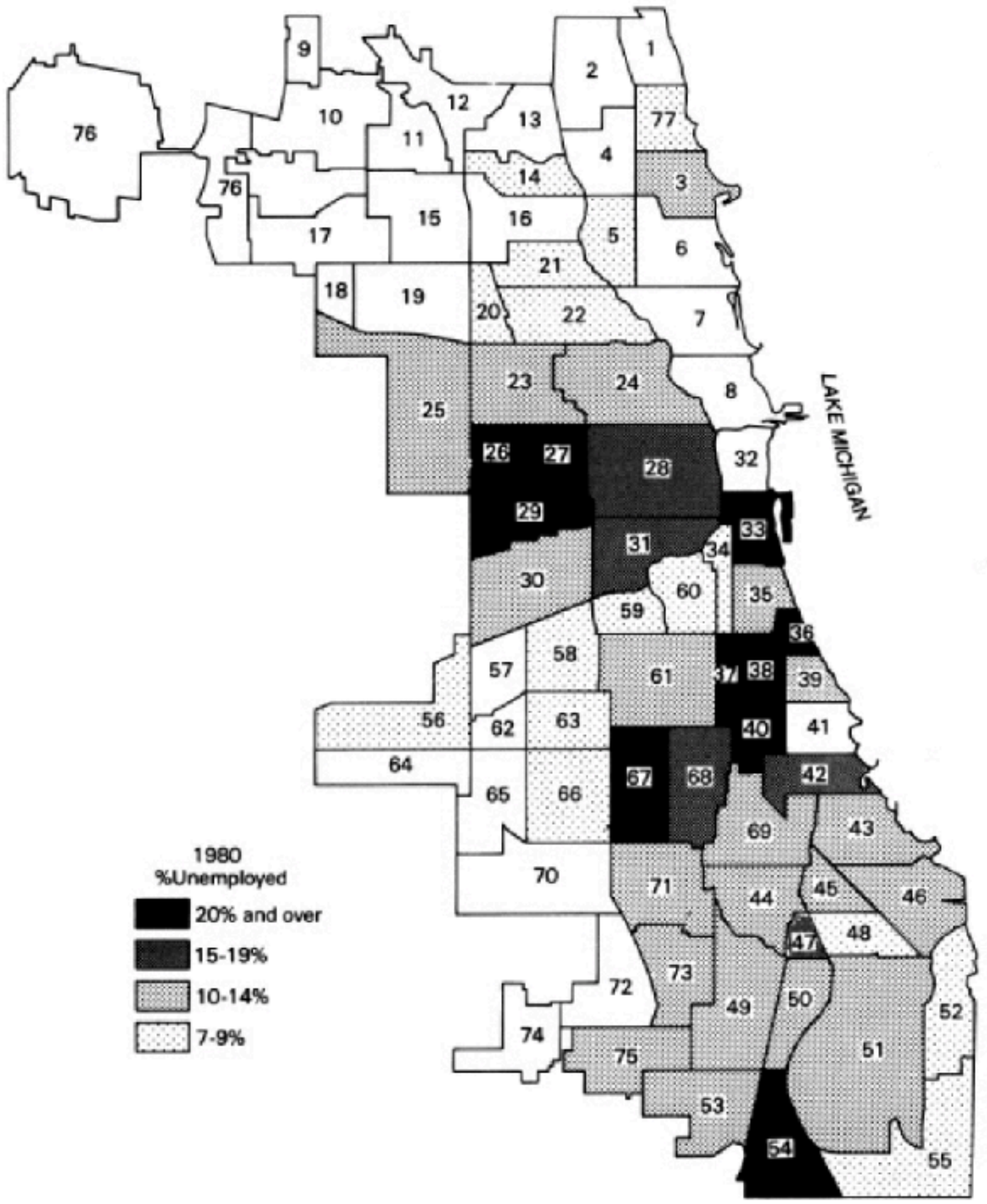
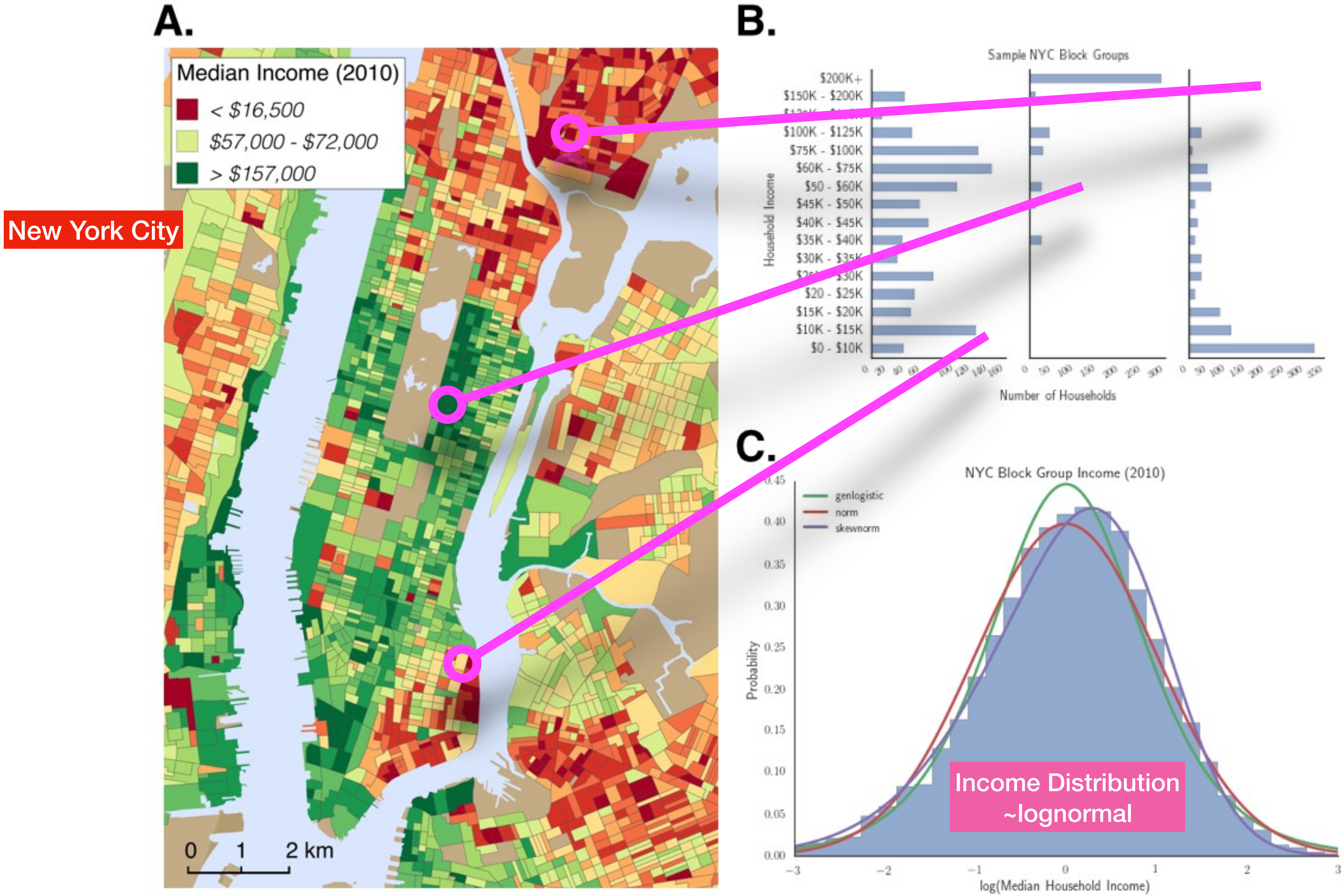


Figure 2.6. Unemployment rates in Chicago Community Areas, 1980.

Source: see fig. 2.3.



New York City
A.
Median Income (2010)
< $16,500
$57,000 - $72,000
> $157,000
0 1 2 km
B.
Sample NYC Block Groups
Household Income
$200K+
$150K - $200K
$100K - $125K
$75K - $100K
$60K - $75K
$50 - $60K
$45K - $50K
$40K - $45K
$35K - $40K
$20 - $25K
$15K - $20K
$10K - $15K
$0 - $10K
Number of Households
C.
NYC Block Group Income (2010)
genlogistic
norm
skewnorm
Probability
log(Median Household Income)
Income Distribution
~lognormal

Choice and Information

necessary to get us closer to people and their urban dilemmas

What is “Information” ?

Dictionary

Search for a word

in·for·ma·tion

/ˌinfərˈmāSH(ə)n/

See definitions in:

All | Law | Computing | Mathematics

noun

1. facts provided or learned about something or someone.
   "a vital piece of information"
   Similar: details | particulars | facts | figures | statistics | data | knowledge
2. what is conveyed or represented by a particular arrangement or sequence of things.
   "genetically transmitted information"

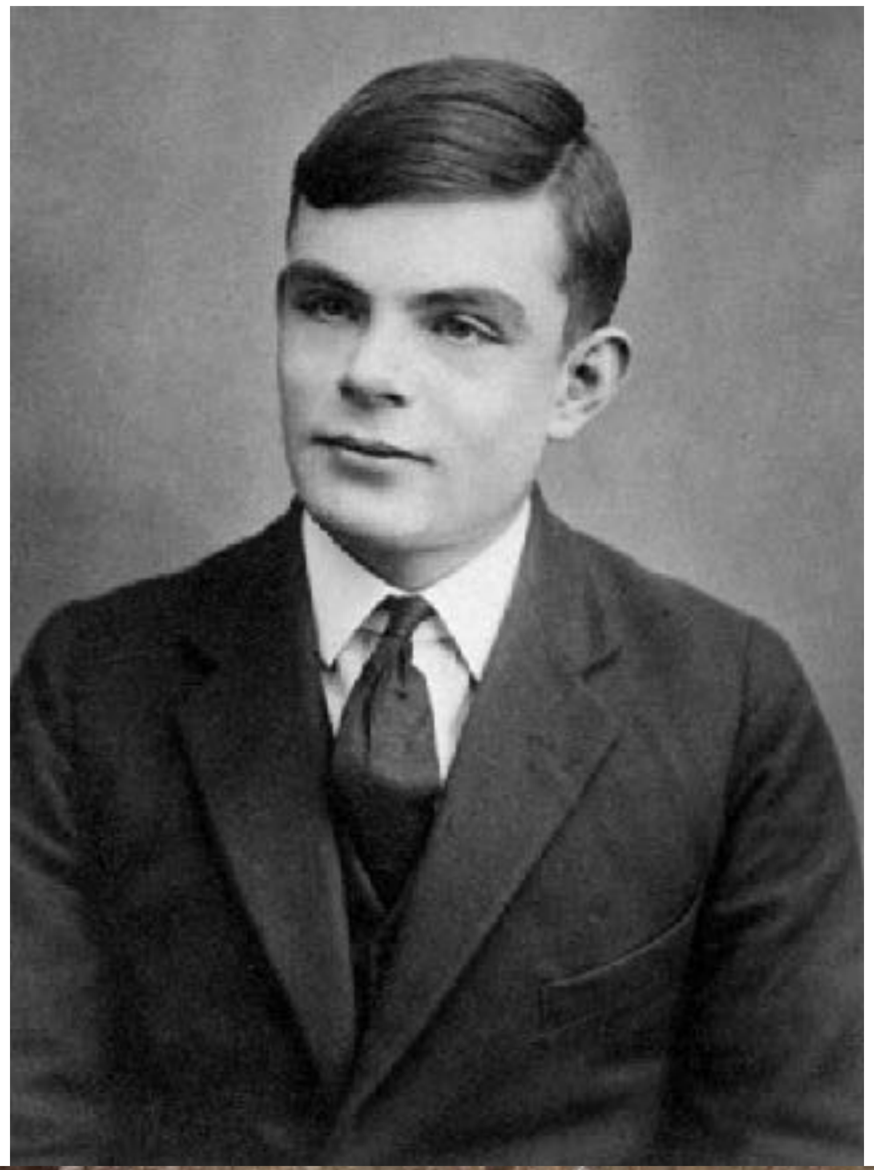
Definitions from Oxford Languages

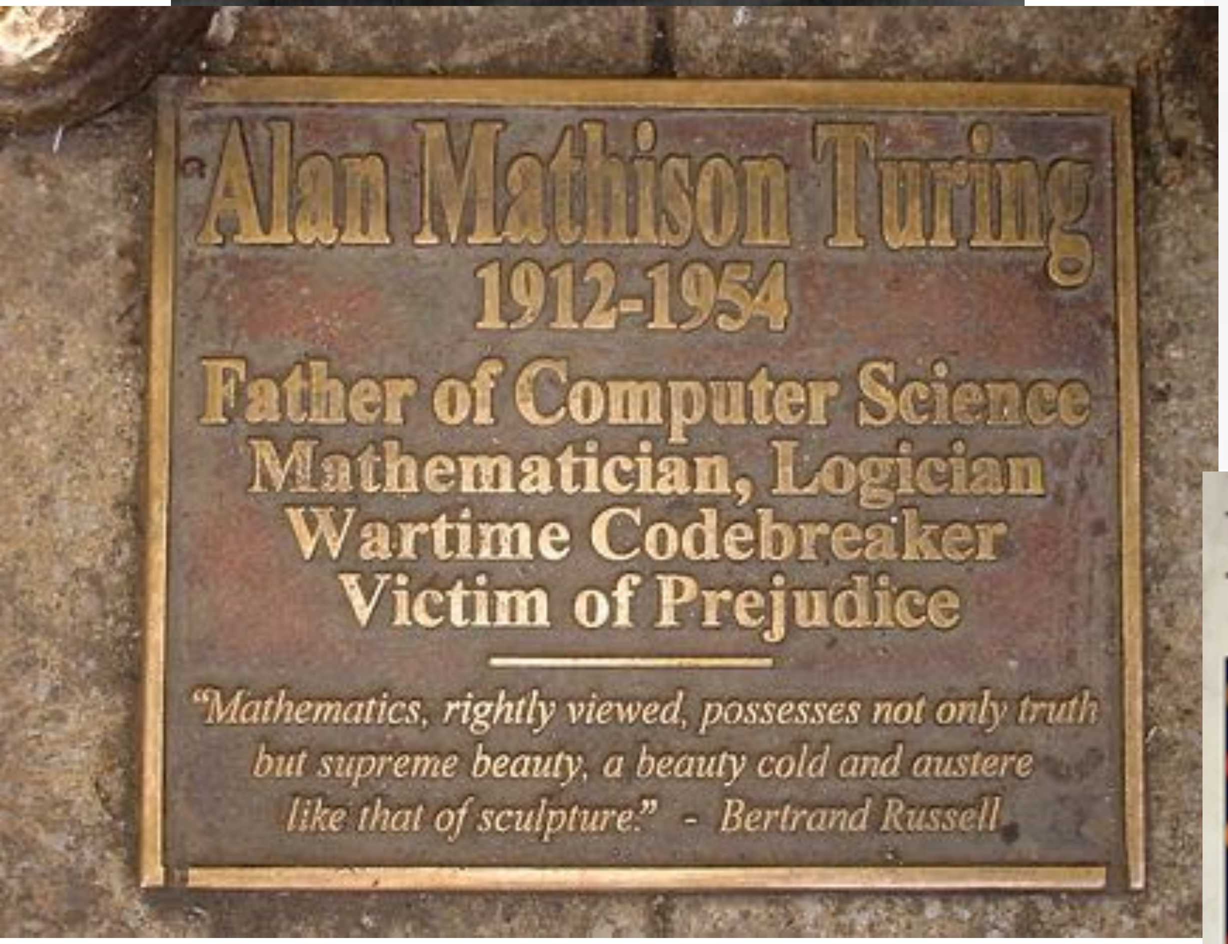
Feedback

Information is a relative quantity

it is reciprocal:
you change the city; it changes you

patterns convey information





Cryptography: Bletchley Park

Bell Labs: inventing the telephone

Claude Shannon

The breakthrough paper:

Reprinted with corrections from *The Bell System Technical Journal*,
Vol. 27, pp. 379–423, 623–656, July, October, 1948.

A Mathematical Theory of Communication

By C. E. SHANNON → Warren Weaver → Jane Jacobs

INTRODUCTION

THE recent development of various methods of modulation such as PCM and PPM which exchange bandwidth for signal-to-noise ratio has intensified the interest in a general theory of communication. A basis for such a theory is contained in the important papers of Nyquist[1] and Hartley[2] on this subject. In the present paper we will extend the theory to include a number of new factors, in particular the effect of noise in the channel, and the savings possible due to the statistical structure of the original message and due to the nature of the final destination of the information.

The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have *meaning*; that is they refer to or are correlated according to some system with certain physical or conceptual entities.

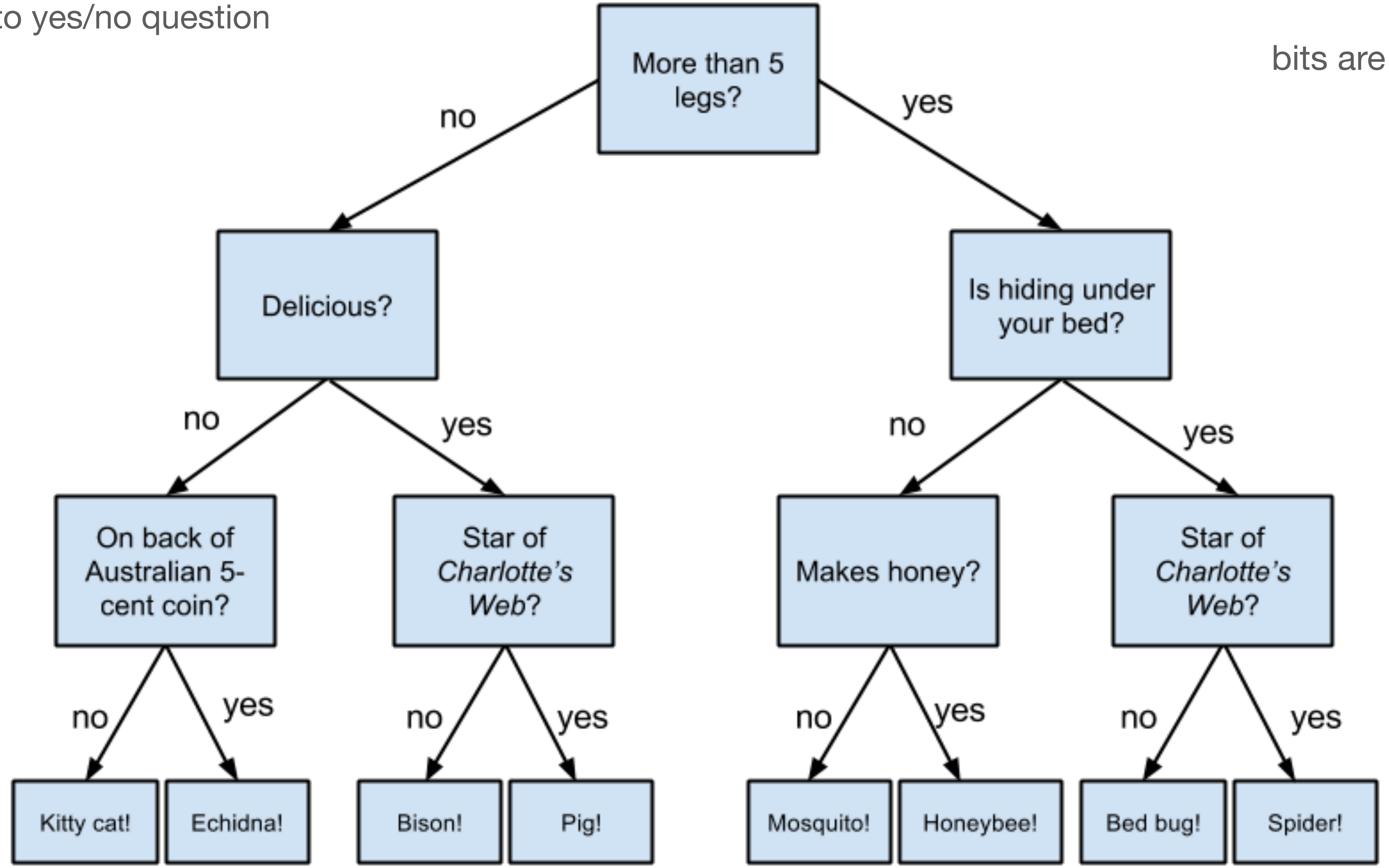
The invention of the “bit”

Information and the Game of 20 Questions

what is the “bit”?
“binary unit”: answer to yes/no question

Each question is a decision… = 1bit

bits are units of choice



credit : o’reilly.com

Information between two variables

$$H(y) = -\sum_i p(y_i) \ln_2 p(y_i)$$ **entropy**

$$H(y|n) = -\sum_{i,j} p(n_j)p(y_i|n_j) \ln_2 p(y_i|n_j)$$ **conditional entropy**

$$I(n;y) = H(y) - H(y|n) = \sum_{i,j} p(y_i, n_j) \; \ln_2 \frac{p(y_i, n_j)}{p(y_i)p(n_j)}$$ **mutual information**

example:

(fair coin)

$y =$ **0 1 0 1 0 0 1 1 0 1 0**

$$p(y=0) = p(y=1) = \frac{1}{2}$$

$$H(y) = 2\frac{1}{2}\ln_2 2 = 1$$ bit

A tip from a friend

$x =$ **0 1 0 1 0 0 1 1 0 1 0**

perfect prediction!

$$p(y=1|x=1) = p(y=0|x=0) = 1$$

$$H(y|x) = 0$$

$$p(x=1) = p(x=0) = \frac{1}{2}$$

$$I(y;x) = H(y) - H(y|x) = H(y) = 1$$

Information between two variables

$$H(y) = -\sum_i p(y_i) \ln_2 p(y_i)$$

$$H(y|n) = -\sum_{i,j} p(n_j) p(y_i|n_j) \ln_2 p(y_i|n_j)$$

$$I(n;y) = H(y) - H(y|n) = \sum_{i,j} p(y_i, n_j) \ \ln_2 \frac{p(y_i, n_j)}{p(y_i)p(n_j)}$$

example:

(fair coin)

$y =$ **0 1 0 1 0 0 1 1 0 1 0**

$$p(y=0) = p(y=1) = \frac{1}{2}$$

$$H(y) = 2\frac{1}{2}\ln_2 2 = 1 \ \text{bit}$$

A tip from a friend

$x =$ **a b a b a a b b a b a**

perfect prediction!

$$p(y=1|x=b) = p(y=0|x=a) = 1$$

$$p(x=a) = p(x=b) = \frac{1}{2}$$

$$H(y|x) = 0$$

$$I(y;x) = H(y) - H(y|x) = H(y) = 1$$

Information between two variables

example:

(fair coin)

$y =$ **0 1 0 1 0 0 1 1 0 1 0**

$$p(y=0) = p(y=1) = \frac{1}{2}$$

$$H(y) = 2\frac{1}{2}\ln_2 2 = 1 \text{ bit}$$

A noisy tip from a friend

$x =$ **a a a b b a b b a b a**

imperfect prediction!

$$p(y=1\,|\,x=b) = p(y=0\,|\,x=a) = \frac{4}{5}$$

$$p(y=1\,|\,x=a) = p(y=0\,|\,x=b) = \frac{1}{5}$$

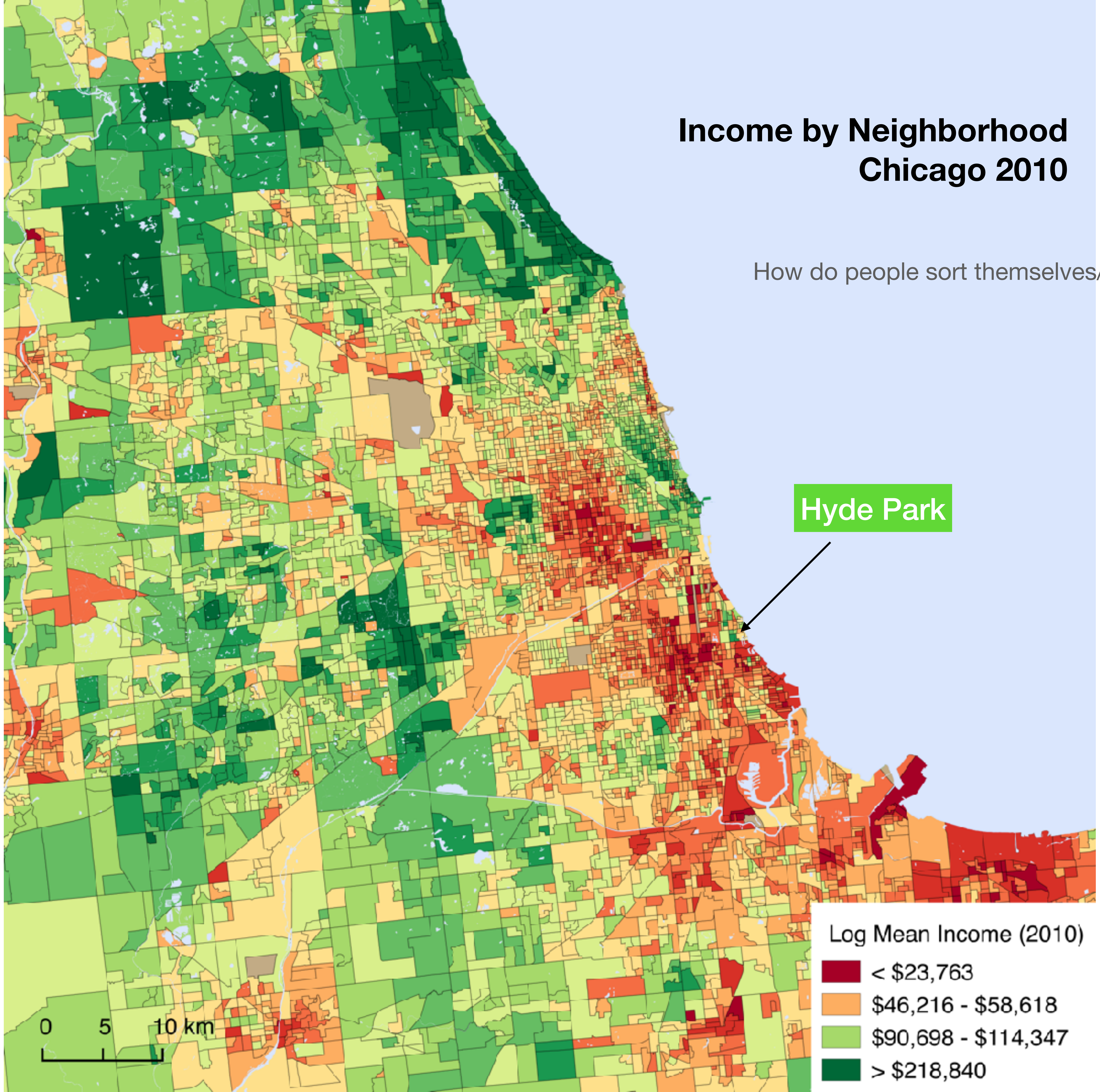
$$p(x=a) = p(x=b) = \frac{1}{2}$$

$$H(y\,|\,x) = 0.26$$

$$I(y;x) = H(y) - H(y\,|\,x) = H(y) = 1 - 0.26 = 0.74 < H(y)$$

The tip contains predictive information, but not enough for *perfect* prediction!

This is true of most situations in real life!



How do people sort themselves/ are sorted?

Back to cities: Pattern

If I know where you live, how much do I know about your income?

$$p(y_i \mid n_j)$$

conditional probability

income level

neighborhood

If I know your income, can I predict where you live?

$$p(n_j \mid y_i)$$

conditional probability

neighborhood

income level

Information between two variables

$$H(y) = -\sum_i p(y_i)\ln_2 p(y_i)$$

entropy

$$H(y|n) = -\sum_{i,j} p(n_j)p(y_i|n_j)\ln_2 p(y_i|n_j)$$

conditional entropy

$$I(n;y) = H(y) - H(y|n) = \sum_{i,j} p(y_i, n_j)\ \ln_2 \frac{p(y_i, n_j)}{p(y_i)p(n_j)}$$

mutual information

example:

(fair coin)

$y =$ 0 1 0 1 0 0 1 1 0 1 0

$$p(y=0) = p(y=1) = \frac{1}{2}$$

$$H(y) = 2\frac{1}{2}\ln_2 2 = 1$$ bit

A tip from a friend

$x =$ 0 1 0 1 0 0 1 1 0 1 0

perfect prediction!

$$p(y=1|x=1) = p(y=0|x=0) = 1$$

$$H(y|x) = 0$$

$$p(x=1) = p(x=0) = \frac{1}{2}$$

$$I(y;x) = H(y) - H(y|x) = H(y) = 1$$

Information between two variables

$$H(y) = -\sum_i p(y_i) \ln_2 p(y_i)$$

$$H(y|n) = -\sum_{i,j} p(n_j) p(y_i|n_j) \ln_2 p(y_i|n_j)$$

$$I(n;y) = H(y) - H(y|n) = \sum_{i,j} p(y_i, n_j) \ \ln_2 \frac{p(y_i, n_j)}{p(y_i)p(n_j)}$$

example:

(fair coin)

$y =$ **0 1 0 1 0 0 1 1 0 1 0**

$$p(y=0) = p(y=1) = \frac{1}{2}$$

$$H(y) = 2\frac{1}{2}\ln_2 2 = 1$$ bit

A tip from a friend

$x =$ **a b a b a a b b a b a**

perfect prediction!

$$p(y=1|x=b) = p(y=0|x=a) = 1$$

$$p(x=a) = p(x=b) = \frac{1}{2}$$

$$H(y|x) = 0$$

$$I(y;x) = H(y) - H(y|x) = H(y) = 1$$

Information between two variables

example:

(fair coin)

$y =$ **0 1 0 1 0 0 1 1 0 1 0**

$$p(y=0) = p(y=1) = \frac{1}{2}$$

$$H(y) = 2\frac{1}{2}\ln_2 2 = 1 \text{ bit}$$

A noisy tip from a friend

$x =$ **a a a b b a b b a b a**

imperfect prediction!

$$p(y=1\,|\,x=b) = p(y=0\,|\,x=a) = \frac{4}{5}$$

$$p(y=1\,|\,x=a) = p(y=0\,|\,x=b) = \frac{1}{5}$$

$$p(x=a) = p(x=b) = \frac{1}{2}$$

$$H(y\,|\,x) = 0.26$$

$$I(y;x) = H(y) - H(y\,|\,x) = H(y) = 1 - 0.26 = 0.74 < H(y)$$

The tip contains predictive information, but not enough for *perfect* prediction!

This is true of most situations in real life!

Bayes Theorem

Remember that probabilities combine as :

$$p(A, B) = p(A \mid B)p(B)$$

$$p(A \mid B) = \frac{p(B \mid A)}{p(B)} p(A)$$

optimal way of learning !!

All theories of learning are Bayesian: Evolution, Cognition, Child Development, Statistical Estimation, Machine Learning, AI

For neighborhoods:

probability of income level given neighborhood

$$p(y_i \mid n_j) = \frac{p(n_j \mid y_i)}{p(n_j)} p(y_i)$$

probability of income level

$$w_{ij} = \frac{p(n_j \mid y_i)}{p(n_j)} = \frac{p(y_i \mid n_j)}{p(y_i)} = \frac{p(n_j, y_i)}{p(y_i)p(n_j)}$$

strength of neighborhood selection

Where did Bayes Theorem come from?

Does God exist??

Hypothesis:

G : God exists vs. ~G : God does not exist

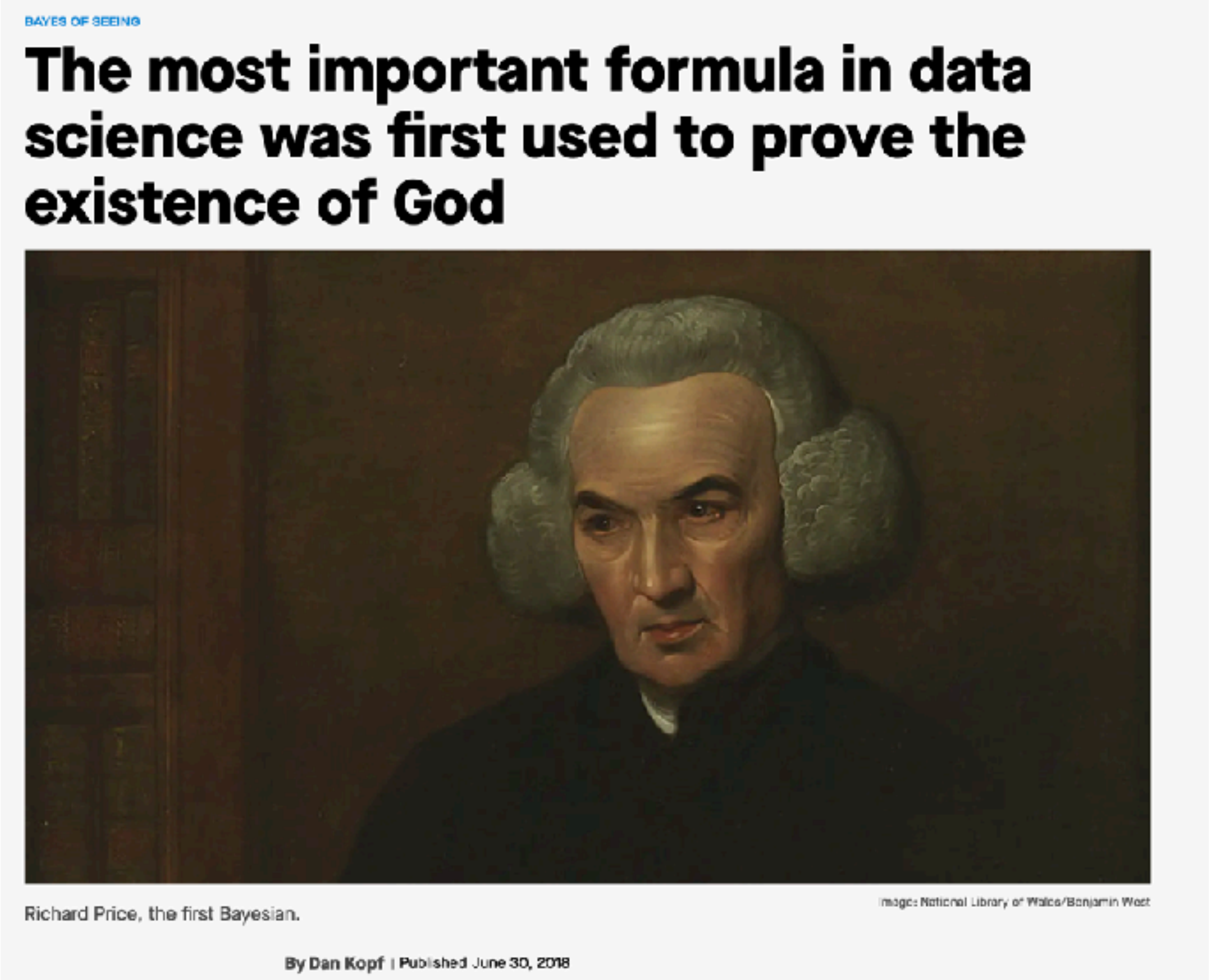
Initially: $p(G) = 0.5 = p(\sim G)$ agnostic

Bayesian inference: Observe events in the world for signs of God

M : a miracle happened vs. ~M : a miracle did not happen

$p(M) = 0.01, \quad p(\sim M) = 0.99$

Thomas Bayes



BAYES OF SEEING

The most important formula in data science was first used to prove the existence of God

Richard Price, the first Bayesian.

Image: National Library of Wales/Benjamin West

By Dan Kopf | Published June 30, 2018

https://qz.com/1315731/the-most-important-formula-in-data-science-was-first-used-to-prove-the-existence-of-god

What if we observe a miracle??

Compare: $p(G|M)$ versus $p(\sim G|M)$

$$p(G|M) = \frac{p(M|G)}{p(M)}p(G) = \frac{0.02}{0.01}0.5 = 1.$$

$$p(\sim G|M) = \frac{p(M|\sim G)}{p(M)}p(\sim G) = \frac{0.}{0.01}0.5 = 0.$$

There must be a God !

Model of God

$p(m\|g)$	G	$\sim G$
M low	high 0.02	low 0
$\sim M$ high	low 0.98	high 1
$= 1$		

Where did Bayes Theorem com from?

Does God exist??

Hypothesis:

G : God exists vs. ~G : God does not exist

Initially: $p(G) = 0.5 = p(\sim G)$ agnostic

Bayesian inference: Observe events in the world for signs of God

M : a miracle happened vs. ~M : a miracle did not happen

$p(M) = 0.01, \quad p(\sim M) = 0.99$

What if we do NOT observe a miracle??

Compare: $p(G \mid \sim M)$ versus $p(\sim G \mid \sim M)$

$$p(G \mid \sim M) = \frac{p(\sim M \mid G)}{p(\sim M)} p(G) = \frac{0.98}{0.99} 0.5 = 0.49$$

Likely God does not exist!

$$p(\sim G \mid \sim M) = \frac{p(\sim M \mid \sim G)}{p(\sim M)} p(\sim G) = \frac{1.}{0.99} 0.5 = 0.51$$

BAYES OF SEEING

The most important formula in data science was first used to prove the existence of God

Richard Price, the first Bayesian.

Image: National Library of Wales/Benjamin West

By Dan Kopf | Published June 30, 2018

https://qz.com/1315731/the-most-important-formula-in-data-science-was-first-used-to-prove-the-existence-of-god

Model of God

$p(m \mid g)$	G	$\sim G$
M low	high 0.02	low 0
$\sim M$ high	low 0.98	high 1
$= 1$		

Negative observations will reduce the believe (posterior probability) that God exists.

Information between two variables (Neighborhood vs. income level)

$$I(n; y) = H(y) - H(y \,|\, n) = \sum_{i,j} p(y_i, n_j) \; \ln_2 \frac{p(y_i, n_j)}{p(y_i)p(n_j)}$$

This is called the "mutual information"

The information in each specific neighborhood:

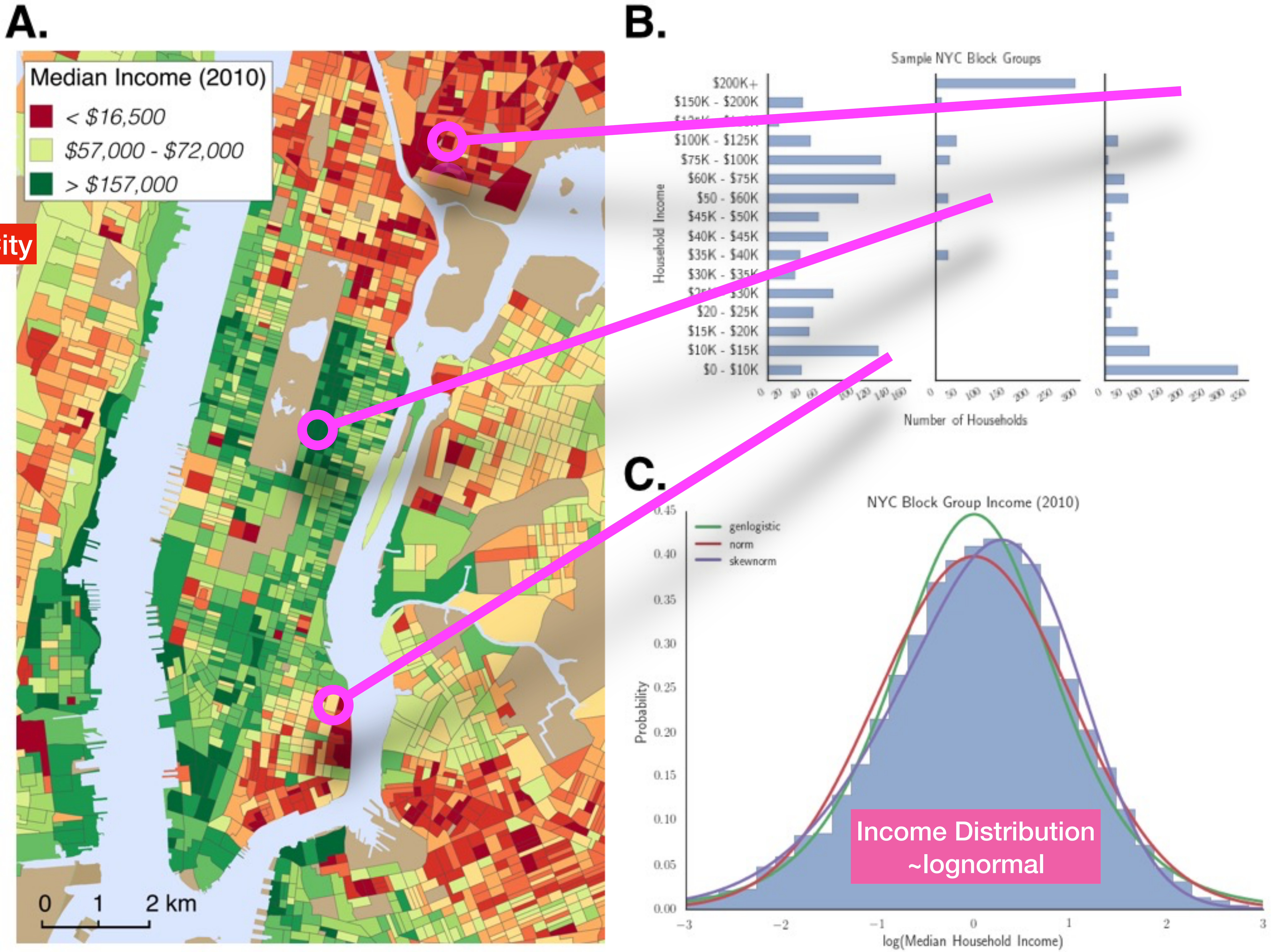
$$D(n_j) = \sum_{i} p(y_i \,|\, n_j) \; \ln_2 \frac{p(y_i \,|\, n_j)}{p(y_i)}$$

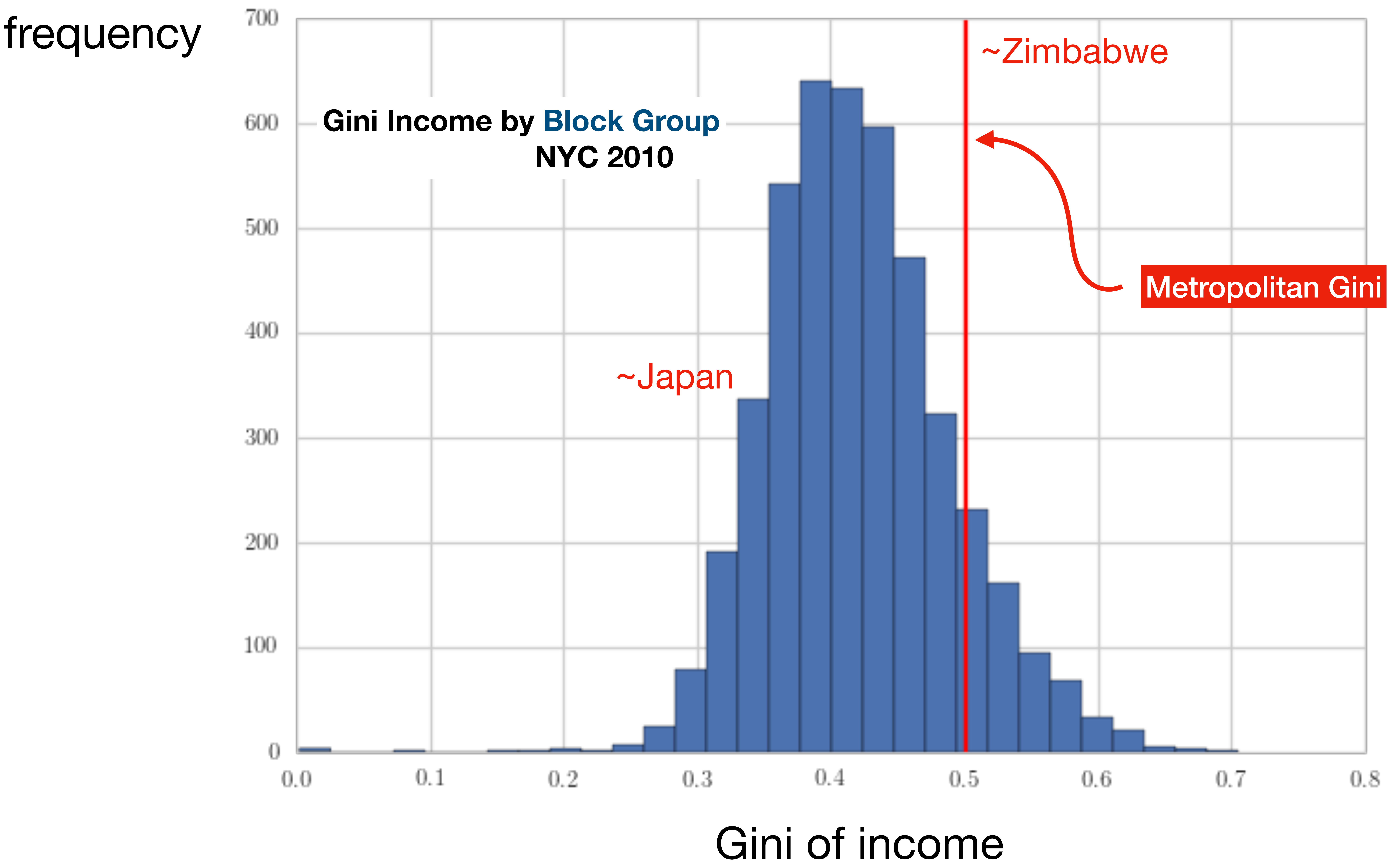
This is called the "Kullback-Leibler (KL) divergence"

The information in each specific income group:

$$D(y_i) = \sum_{j} p(n_j \,|\, y_i) \; \ln_2 \frac{p(n_j \,|\, y_i)}{p(n_j)}$$

New York City




frequency
Gini Income by Block Group
NYC 2010
~Zimbabwe
~Japan
Metropolitan Gini
700
600
500
400
300
200
100
0
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
Gini of income

Information & Spatial Selection

$$p(y_\ell|n_j) = w_{\ell,j}\, p(y_\ell)$$

Income distribution for neighborhood i

Income distribution for whole city population

$$p(y_\ell|n_j) = \frac{p(n_j|y_\ell)p(y_\ell)}{p(n_j)} \rightarrow w_{\ell,j} = \frac{p(n_j|y_\ell)}{p(n_j)} = \frac{p(y_\ell|n_j)}{p(y_\ell)} = \frac{p(n_j, y_\ell)}{p(y_\ell)\, p(n_j)}$$

Bayes Theorem

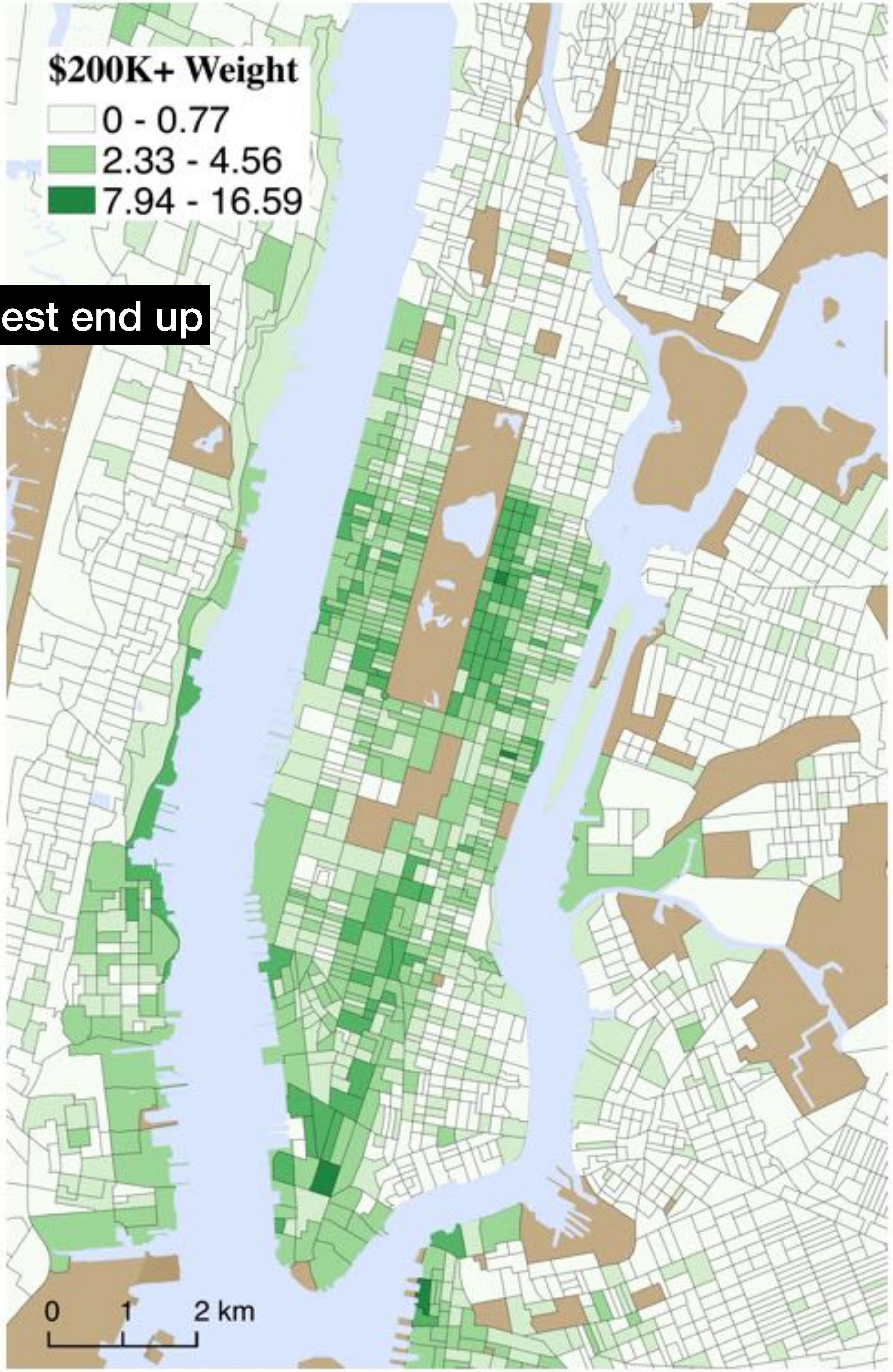
$$\langle \ln w_j \rangle = \sum_\ell p(y_\ell|n_j) \ln \frac{p(y_\ell|n_j)}{p(y_\ell)} = D_{\mathrm{KL}}\left[p(y_\ell|n_j)||p(y_\ell)\right]$$

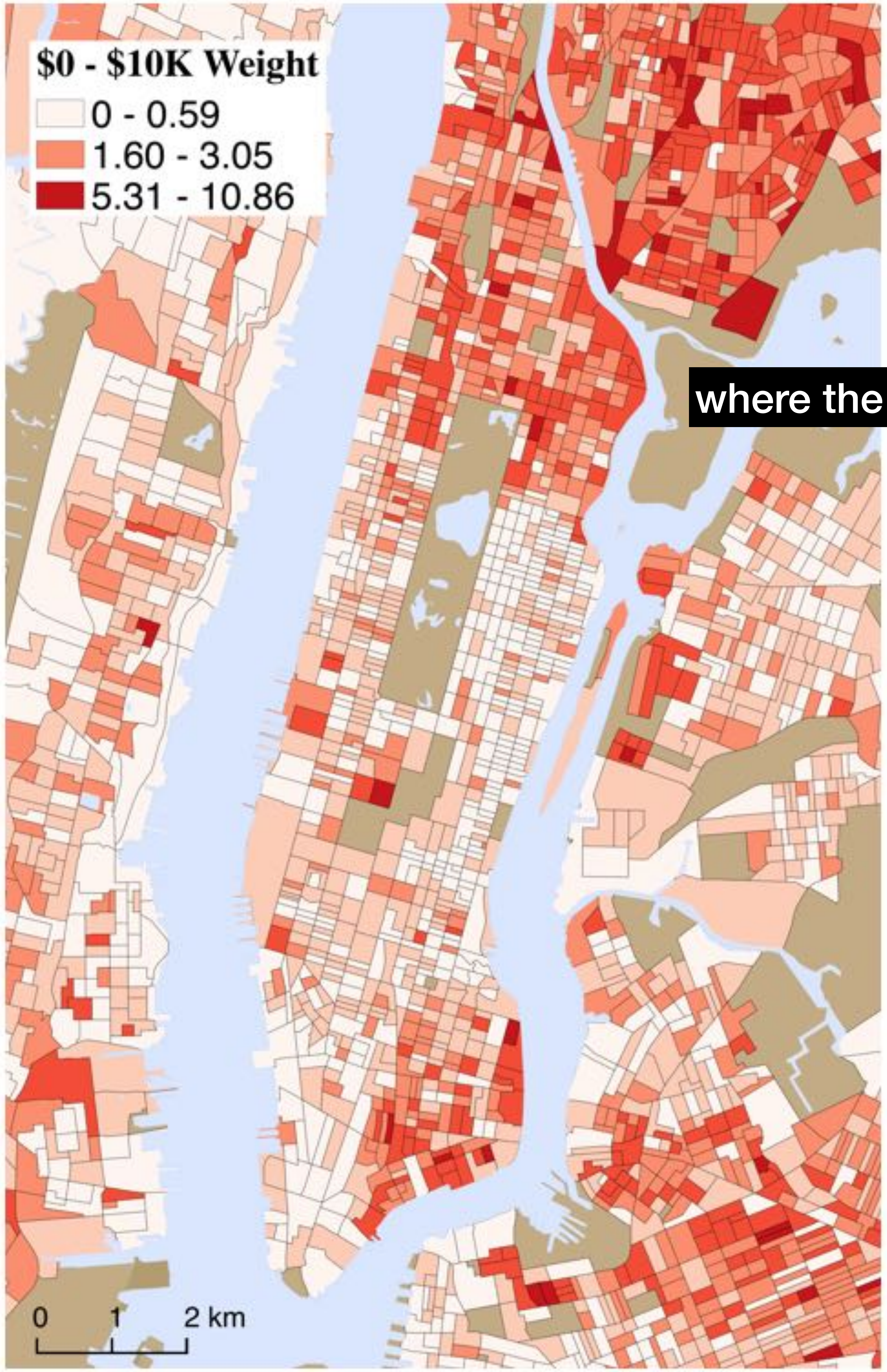
Information
to explain neighborhood j

$$\langle \ln w \rangle = \sum_{i,l} p(n_i, y_\ell) \ln w_{i,l} = I(n; y).$$

Information
between neighborhoods and income
for a city

where the richest end up





where the poorest end up

For 15 years, David Koch lived at the world's "richest building"

Brokerage sources speculate that Koch's unit may list for sale after construction is complete at his recently purchased UES townhouse

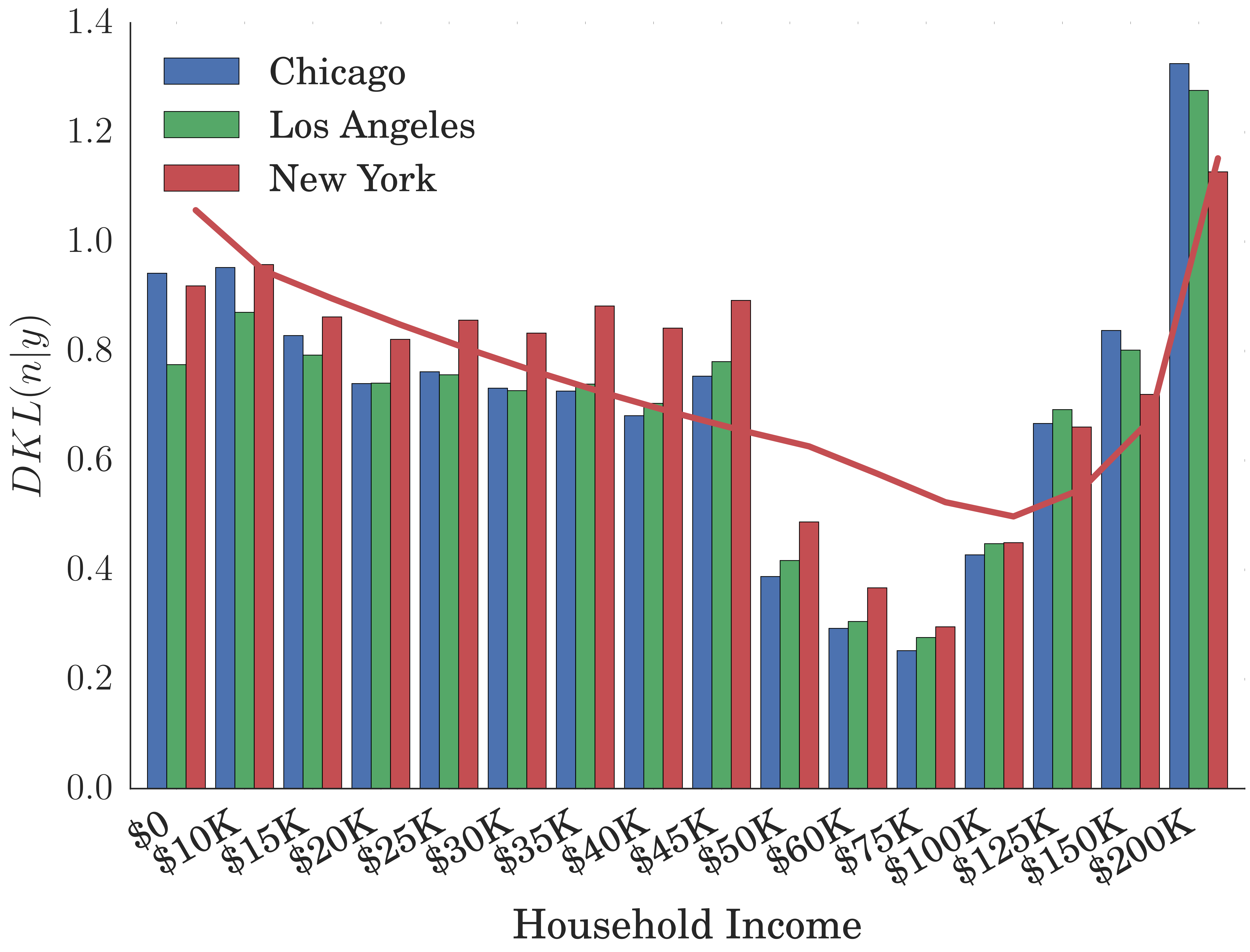
TRD New York / By E.B. Solomont

August 26, 2019 07:00 AM

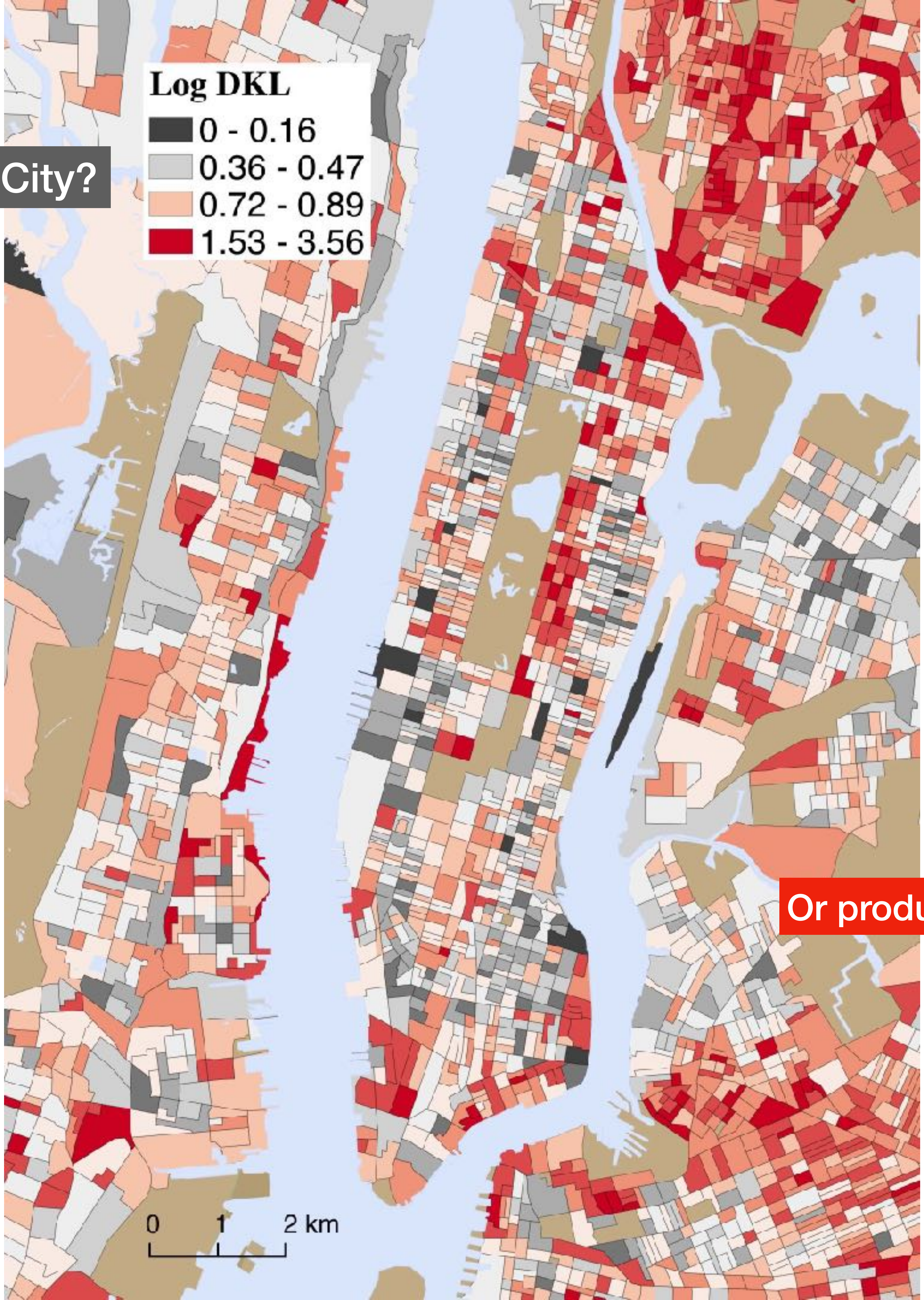
Clockwise from left: John D. Rockefeller, Izzy Englander, Steven Mnuchin, David Koch, Jacqueline Bouvier, and William Zeckendorf (Credit: Getty Images and StreetEasy)

Before paying $40 million for an Upper East Side townhouse last year, David Koch was among a handful of billionaires living at 740 Park Avenue.

which income group is most segregated spatially?

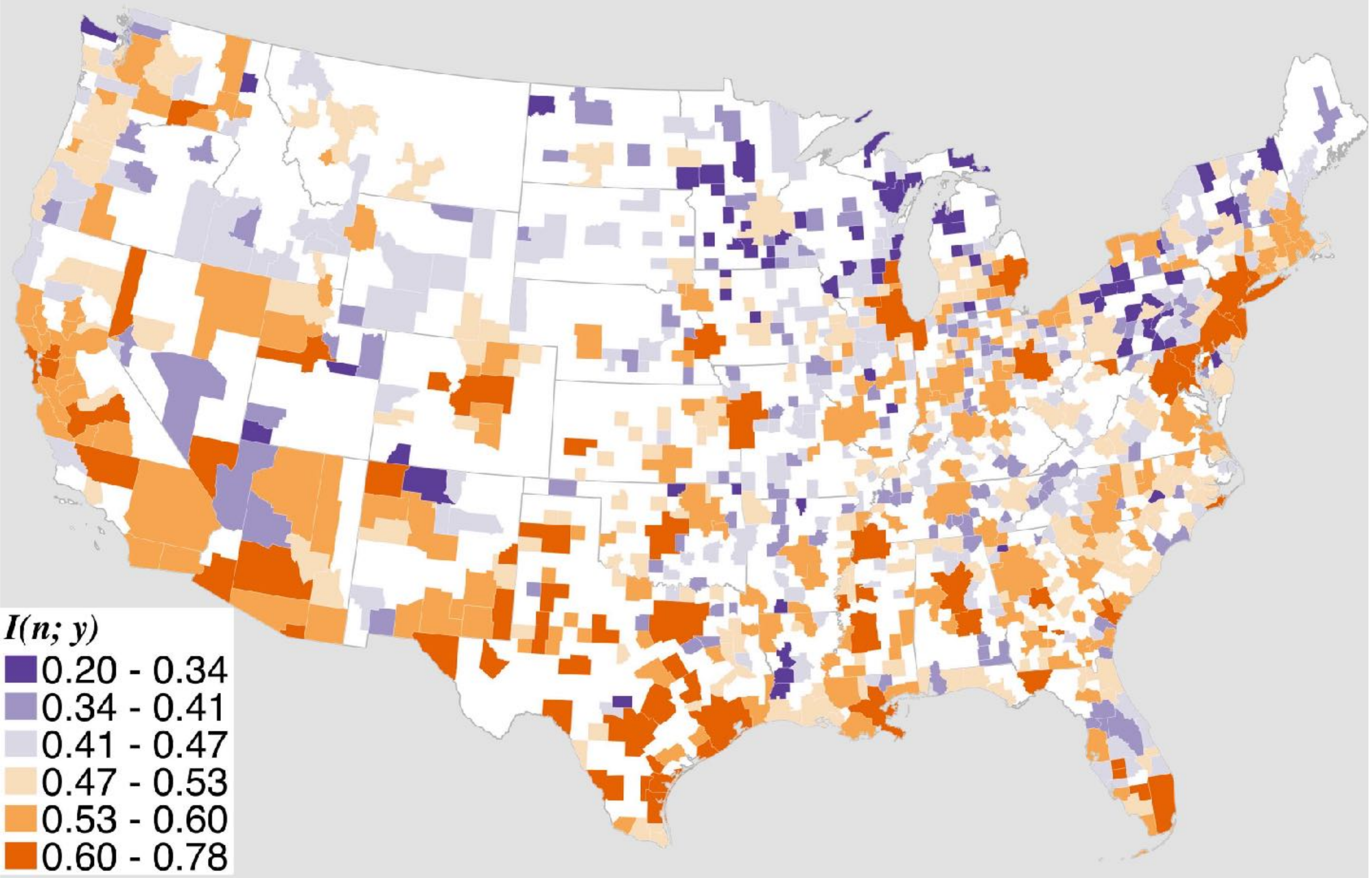


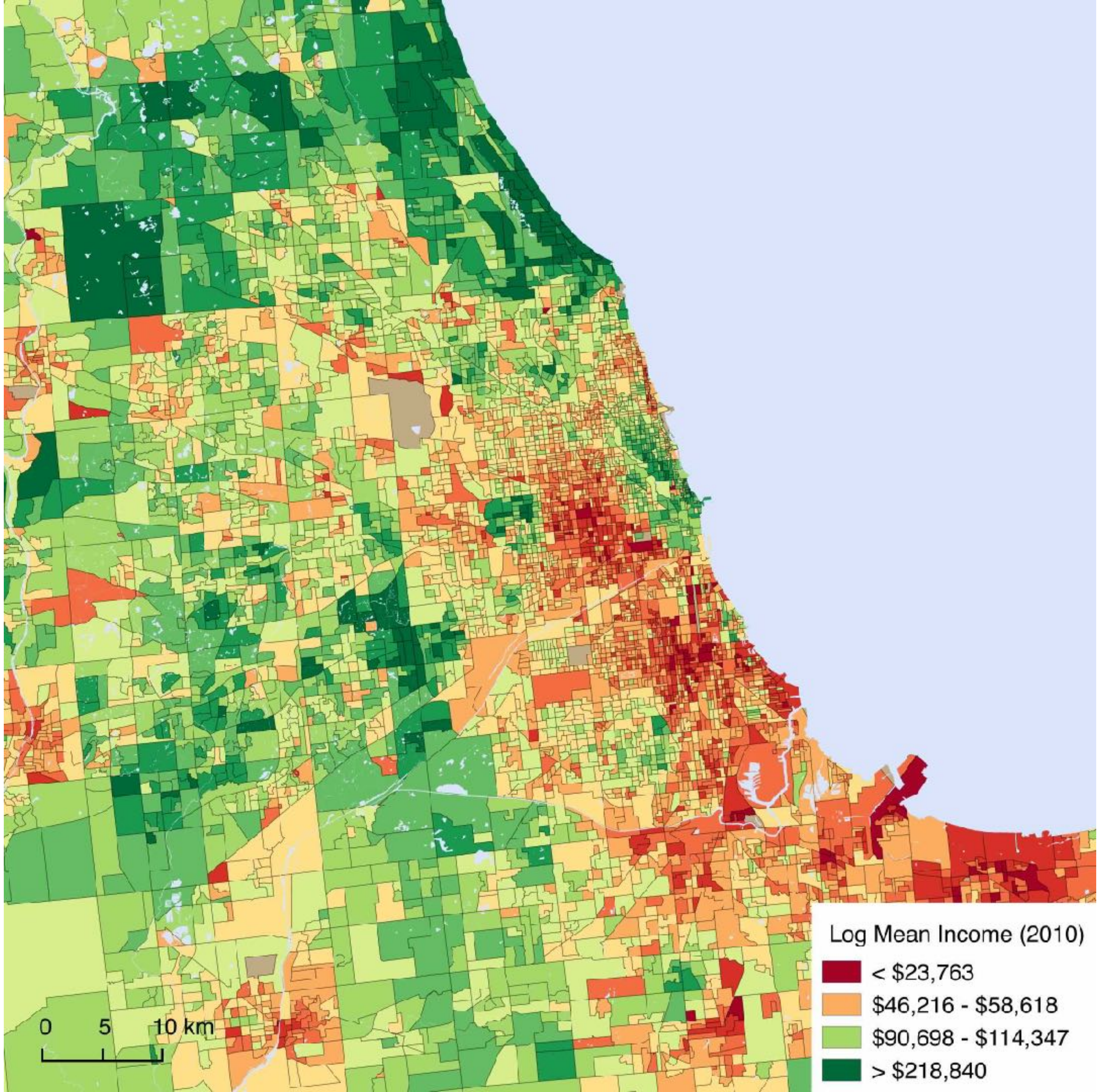
Microcosms of the City?



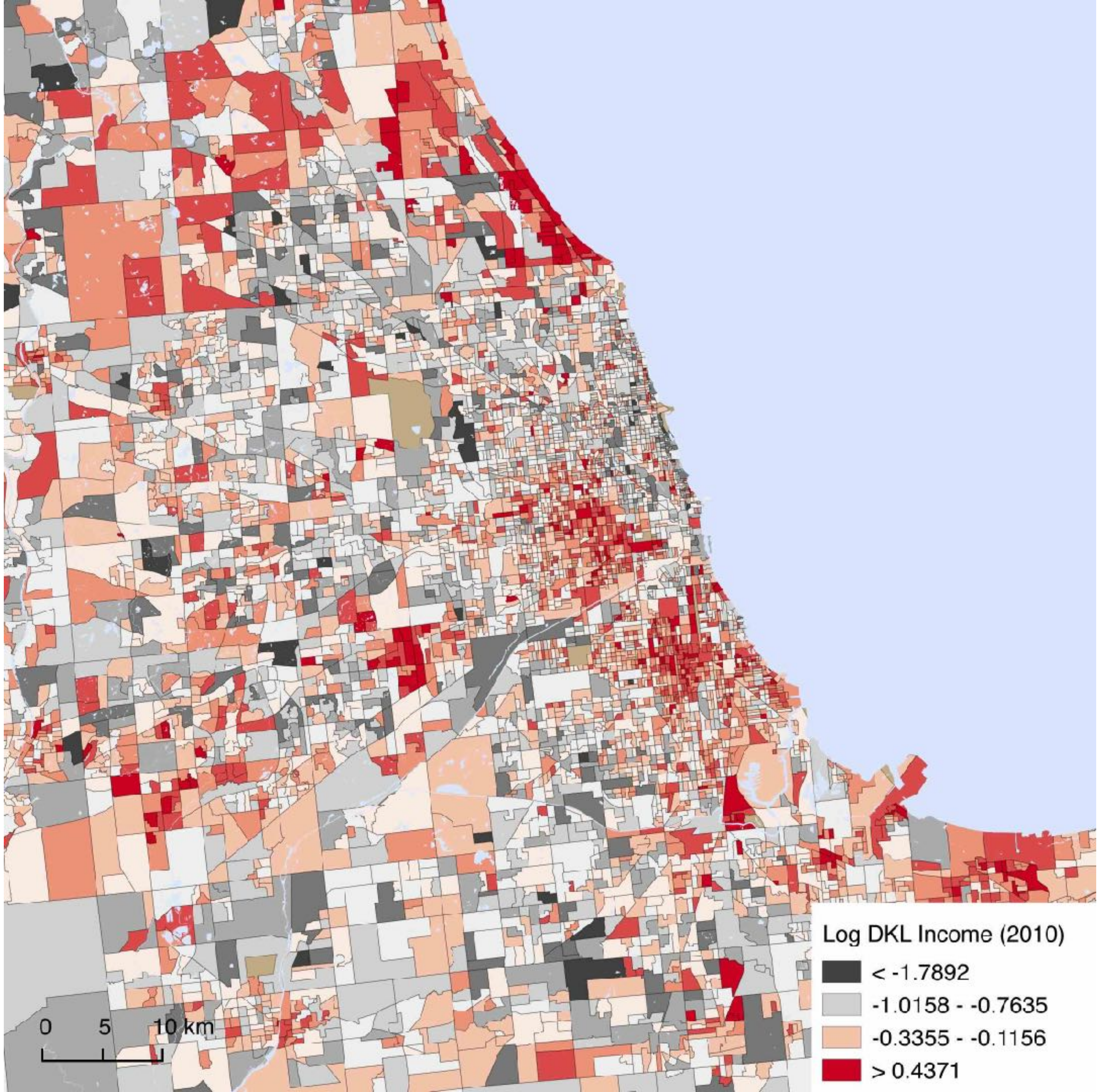
Or products of sorting & selection?

In which cities do neighborhoods tell me more about income ?





Log Mean Income (2010)
< $23,763
$46,216 - $58,618
$90,698 - $114,347
> $218,840
0 5 10 km


Log DKL Income (2010)
< -1.7892
-1.0158 - -0.7635
-0.3355 - -0.1156
> 0.4371
0
5
10 km

General features of household neighborhood sorting by income in US Metros

-There are many types of neighborhoods:

Segregated, Integrated, Rich, Poor,… this is often a simplistic objectification

-The greater the city-wide inequality, the more polarized neighborhoods tend to be

-Amounts of segregation and neighborhood effects vary considerably

from place to place and city to city