Scaling

What is it? Why is it relevant? What can we do with it?
And why is it fundamental to Human Systems, Spatial Statistics, and Simulation Modelling

Michael Batty
University College London

m.batty@ucl.ac.uk
http://www.casa.ucl.ac.uk/

Outline

- The Forbes 400
- Self-Similarity, Universality
- The Mathematics
- City-Size Distributions
- Two Derivations: First, Random Dynamics Second, Entropy again; see next week – back to the future
- My Work on Skyscrapers

The Forbes 400

http://www.forbes.com/lists/

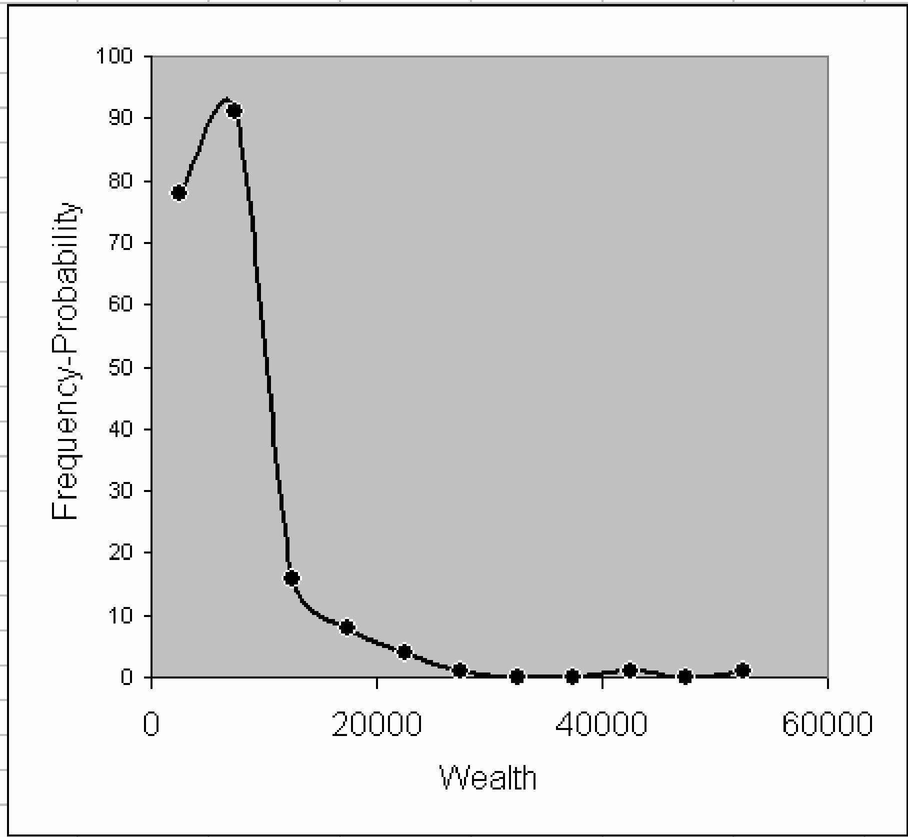
Top 400 most wealthy persons in the US, and the top 400 in the world, each year

We will look at 2009

Now this was first observed by Pareto in the late 19th century, it is the origin of the 80-20 rule, and it is the source of ideas about the long tail

Wealth in Billions	No of people in USA
2500	78
7500	91
12500	16
17500	8
22500	4
27500	1
32500	0
37500	0
42500	1
47500	0
52500	1

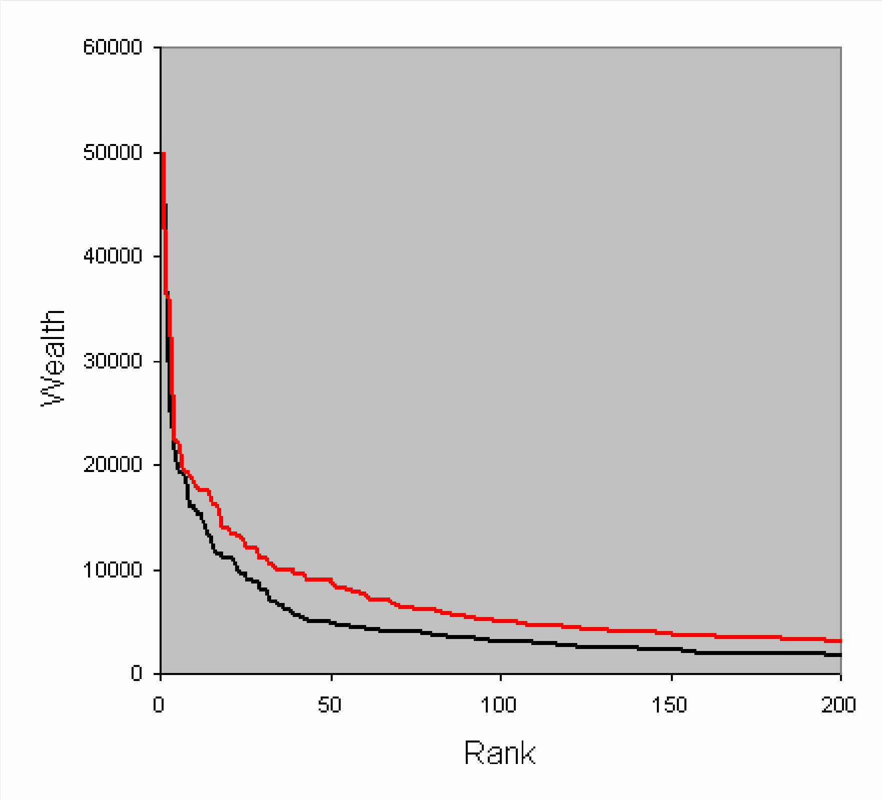
Actually these figures are in millions, I think

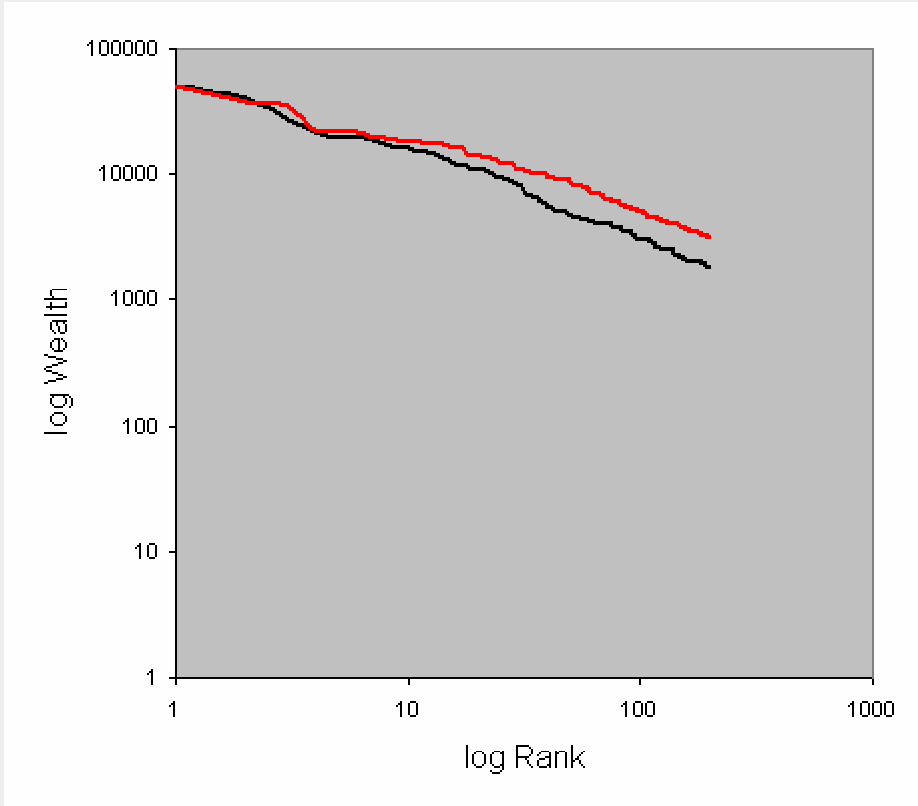


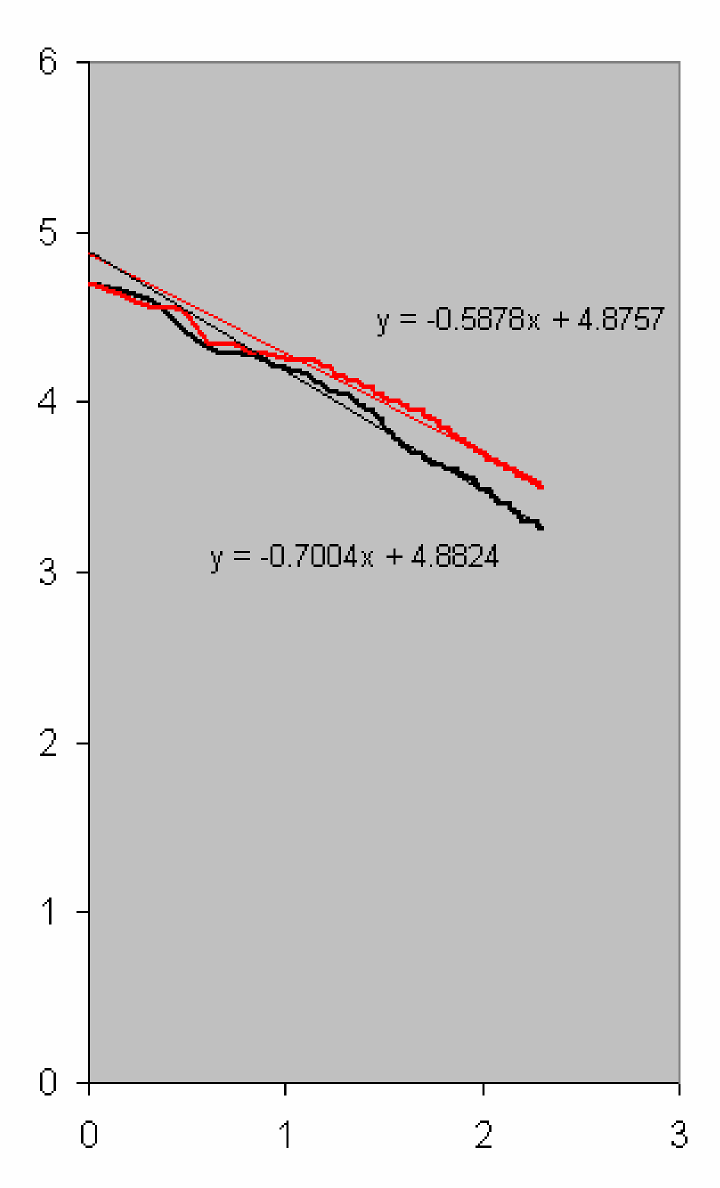
Note the concept of tails – heavy or upper tails, long tails

Ok I will show you how we do this in a minute but take on trust that we can manipulate this probability frequency distribution into what we call the counter-cumulative which will give us a relationship between wealth and rank –

The so-called rank size relationship where can graph Wealth against rank, so Bill Gates is number 1, Warren Buffet number 2 and so on


60000
50000
40000
30000
20000
10000
0
Wealth
0
50
100
150
200
Rank

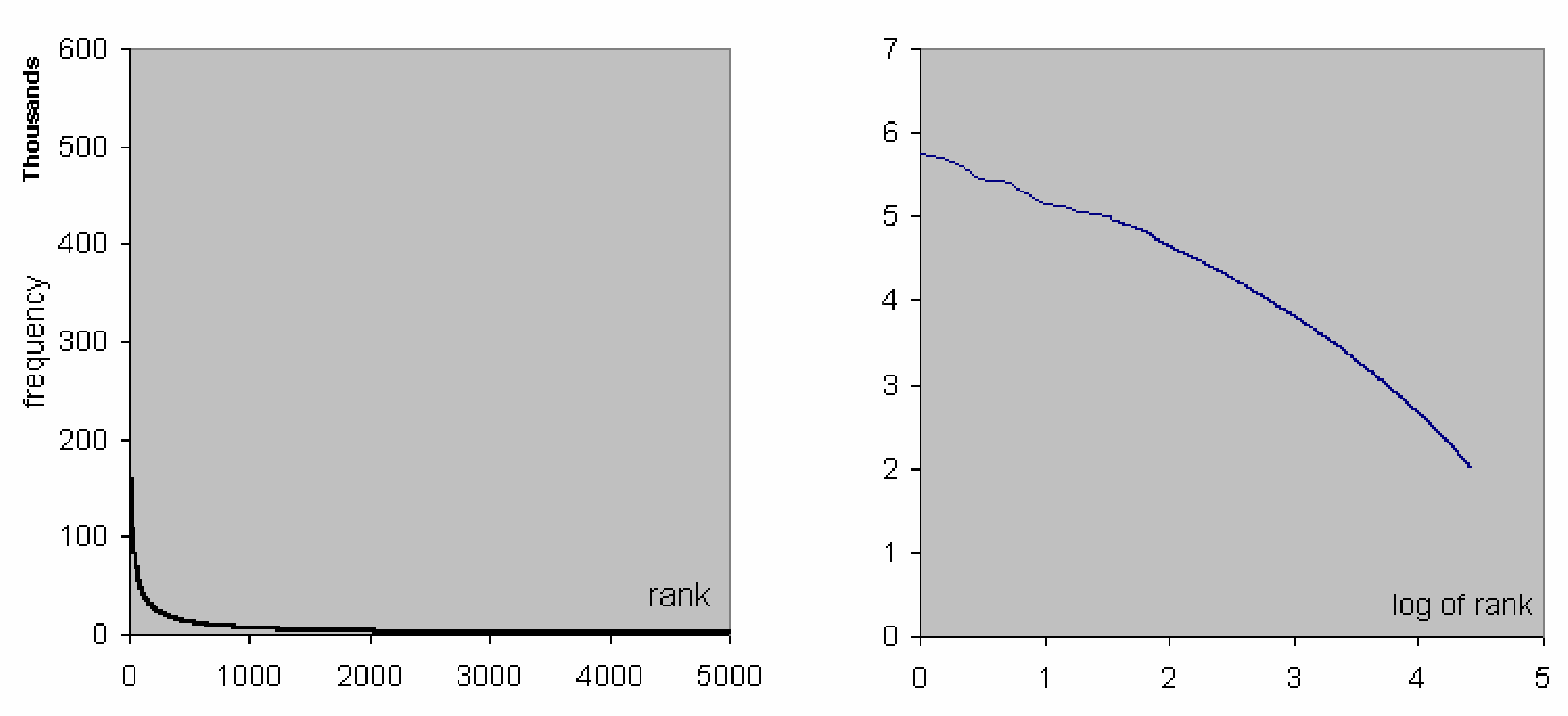

100000
10000
1000
100
10
1
log Wealth
1
10
100
1000
log Rank



Now for those who live in the dataland of surnames

From 1996 Electoral Register

1	SMITH	560122
2	JONES	431558
3	WILLIAMS	285836
4	BROWN	264869
5	TAYLOR	251567
6	DAVIES	216535
7	WILSON	192338
8	EVANS	173636
9	THOMAS	154557
10	JOHNSON	145459



There are 25000 or so distinct names in the register and here are the rank sizes but let us note the differences to incomes and city sizes

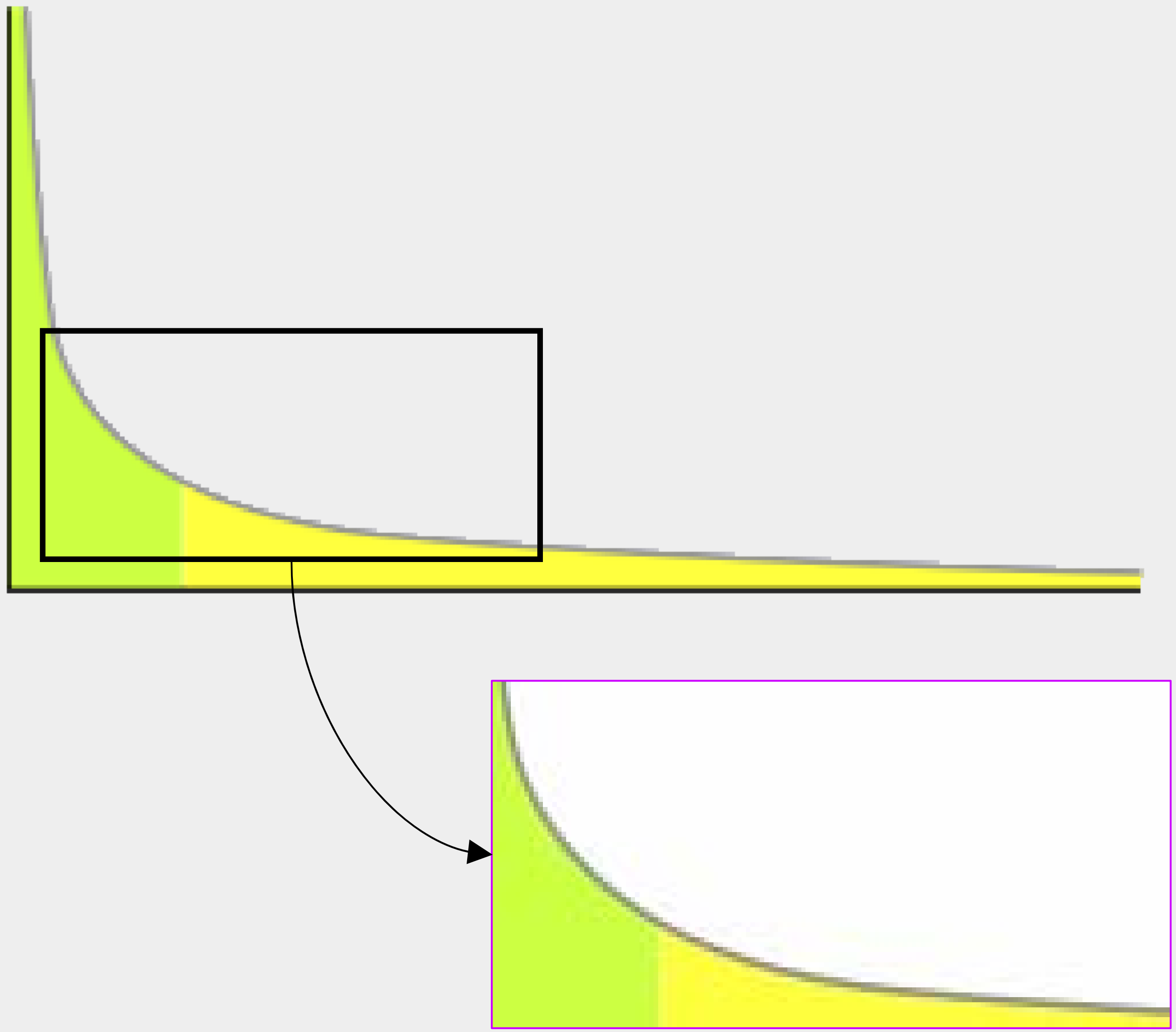
Self-Similarity, Universality

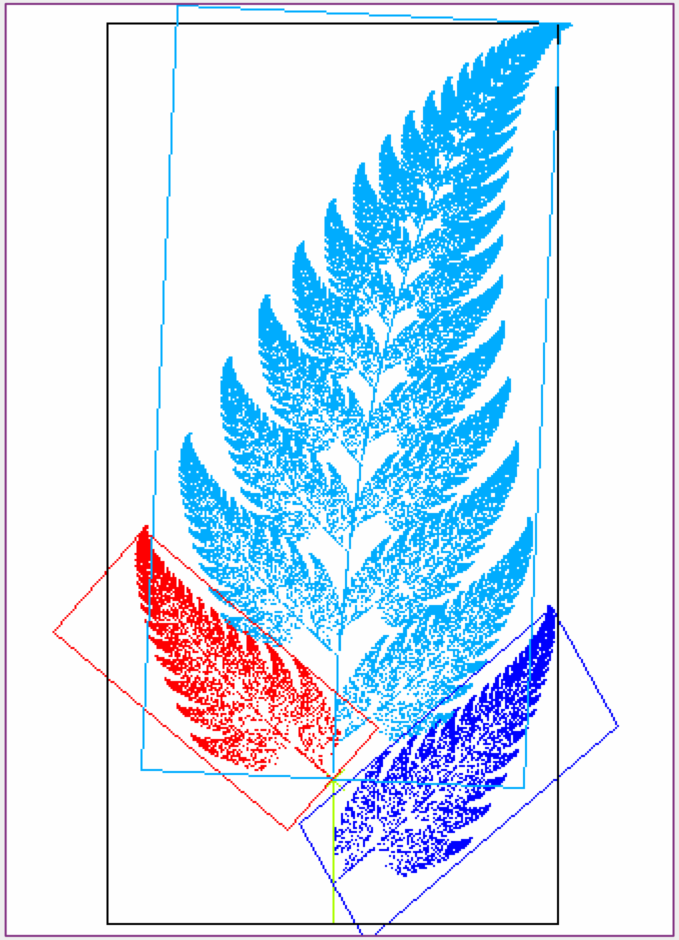
First if we examine any portion of the curve and scale it up, we get the same shape

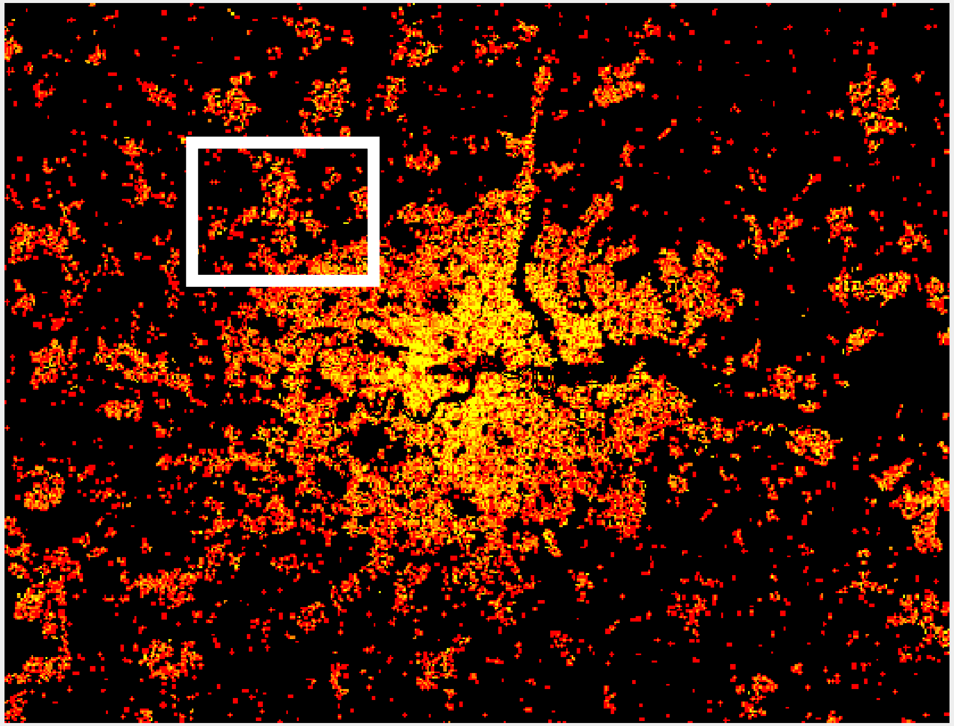
As we examine the phenomena at different scales it is the same

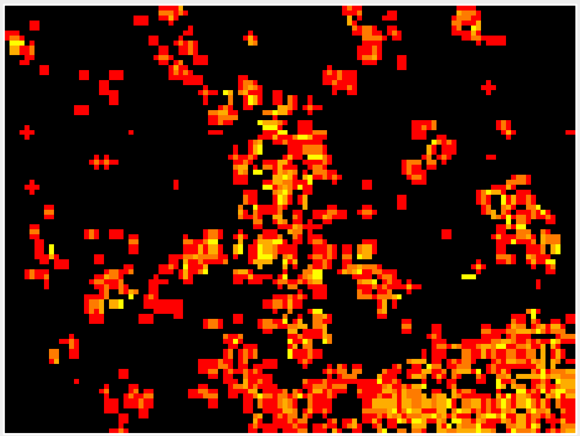
This is a characteristic of power laws, I will show you in a minute

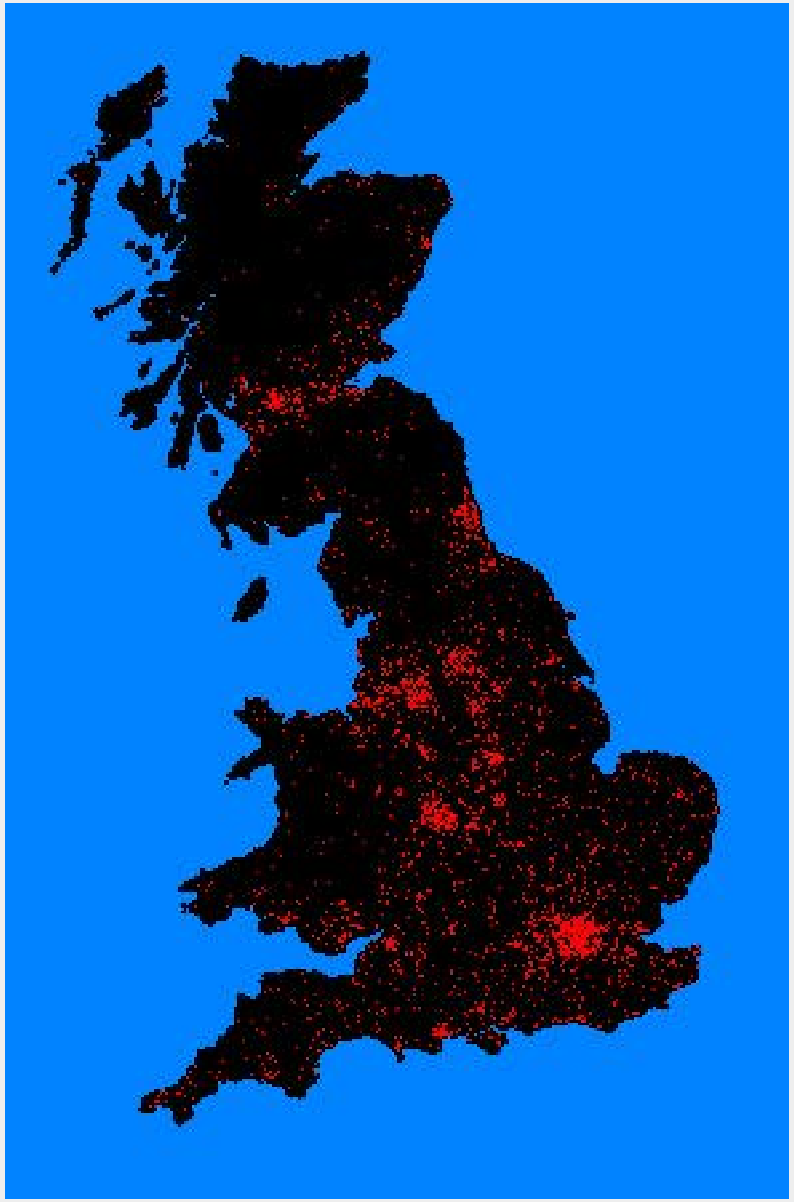
It is also a characteristic of many other shapes



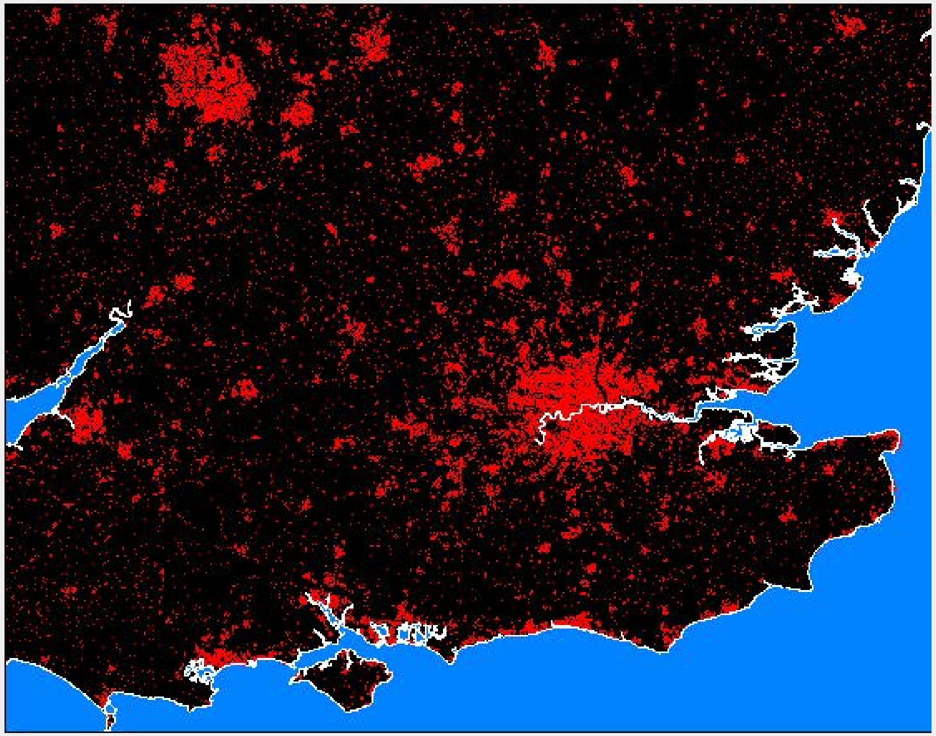








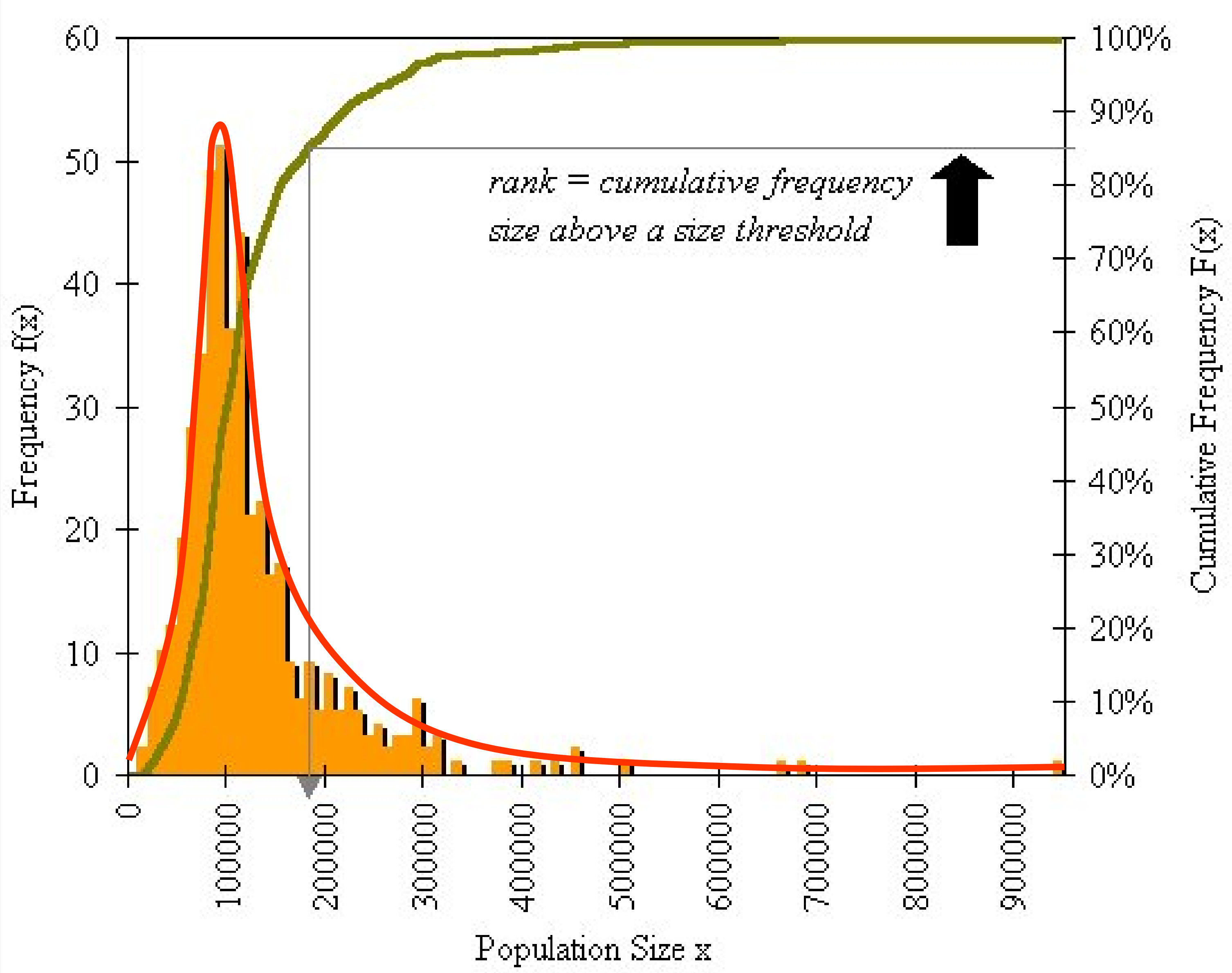
These clusters scale geometrically and their organization is ‘fractal’. This is fractal geometry where objects of the same shape exist at all scales



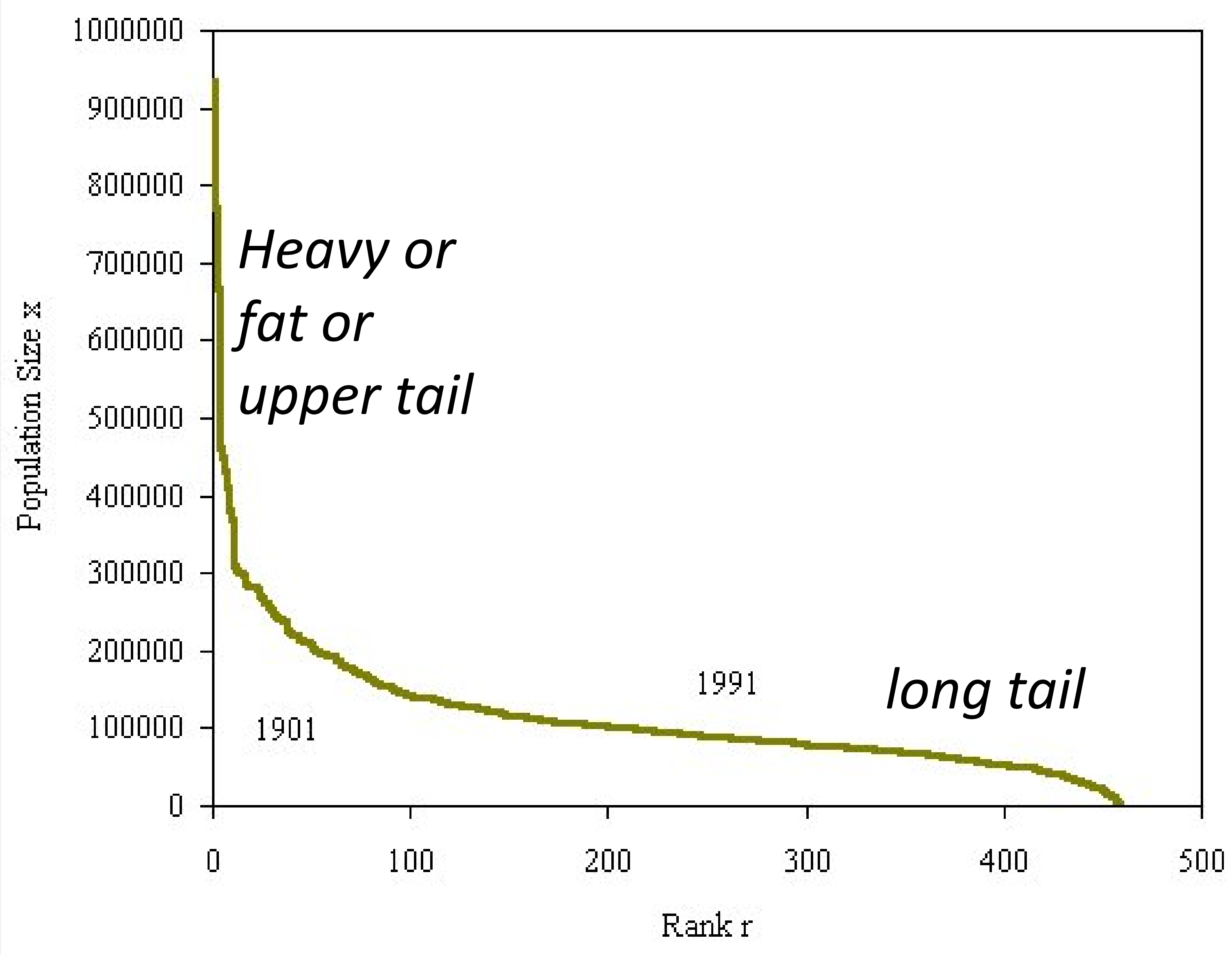
Universality – what does this mean?

Now The Mathematics

First let us look at how we can transform a frequency distribution into a rank size distribution – we start with city size distribution in the UK from the paper I handed out.



60
50
40
30
20
10
0
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Frequency f(x)
Cumulative Frequency F(x)
rank = cumulative frequency
size above a size threshold
0
100000
200000
300000
400000
500000
600000
700000
800000
900000
Population Size x


1000000
900000
800000
700000
600000
500000
400000
300000
200000
100000
0
Population Size x
Heavy or
fat or
upper tail
1991
long tail
1901
0
100
200
300
400
500
Rank r

The frequency distribution

$$f(x) \sim x^{-\alpha}$$

Scaling

$$f(\lambda x) \sim (\lambda x)^{-\alpha} = \lambda^{-\alpha} x^{-\alpha} \sim f(x)$$

The cumulative frequency

$$r(x) = F_{x \to x_{\max}}(x) = \int_{x_{\min}}^{x_{\max}} f(x) \sim x^{-\alpha+1}$$

The rank size

$$x \sim r(x)^{1/1-\alpha}$$

City-Size Distributions

From Zipf's (1949) book *Human Bhevior and the Principle of Least Effort* - the US urban system

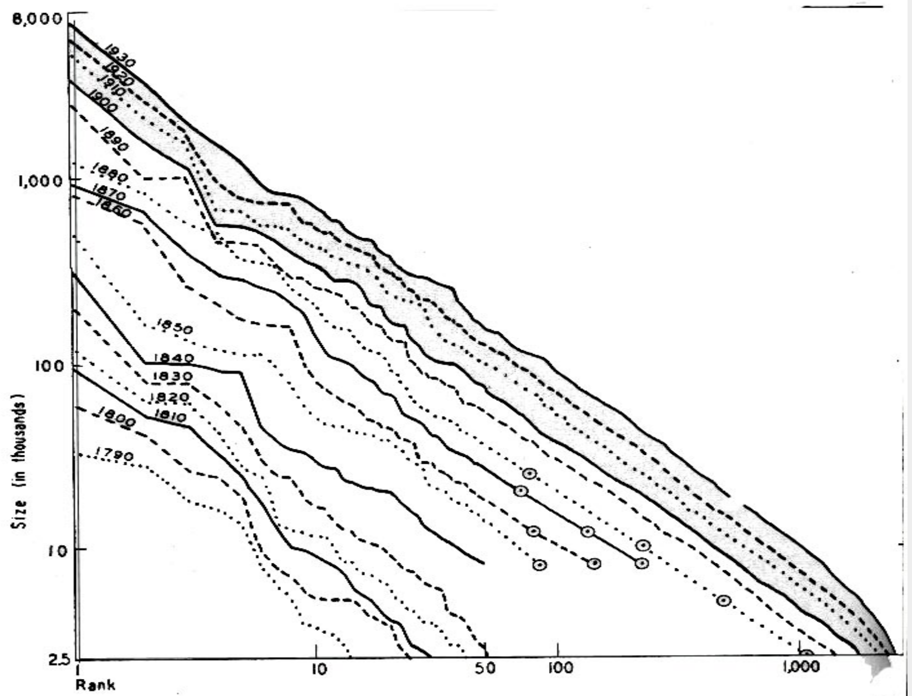
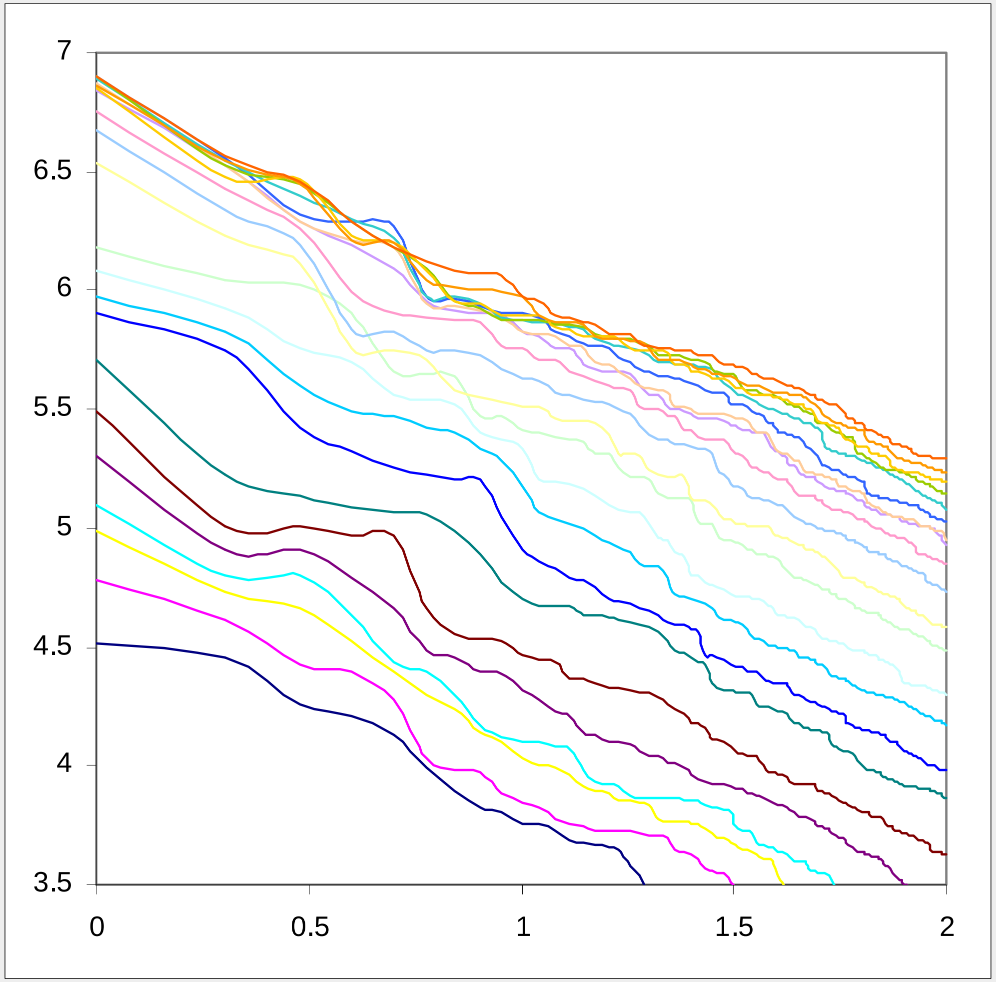
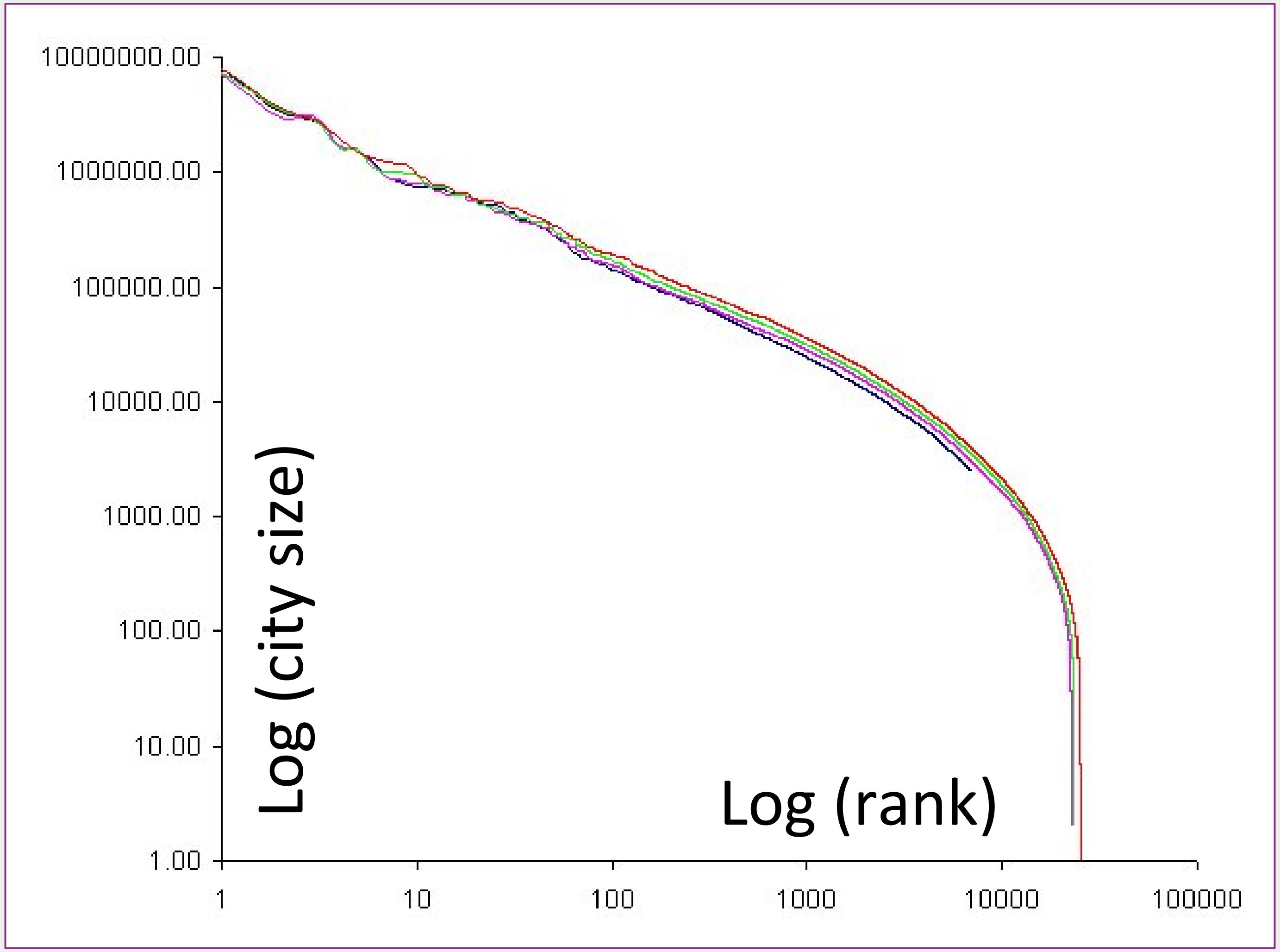


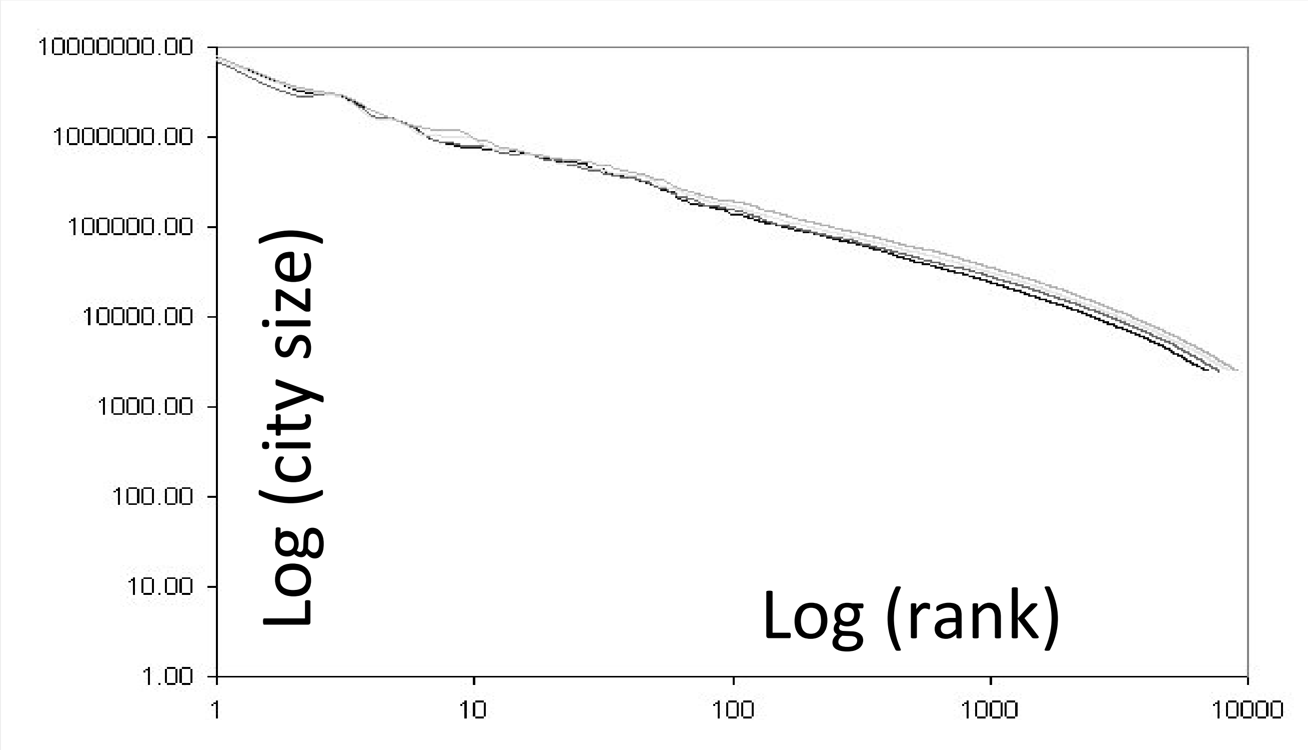
Fig. 10–2. U. S. A. 1790–1930. Communities of 2500 or more inhabitants ranked in the decreasing order of population size.

In this way, we have reworked Zipf's data (from 1790 to 1930)

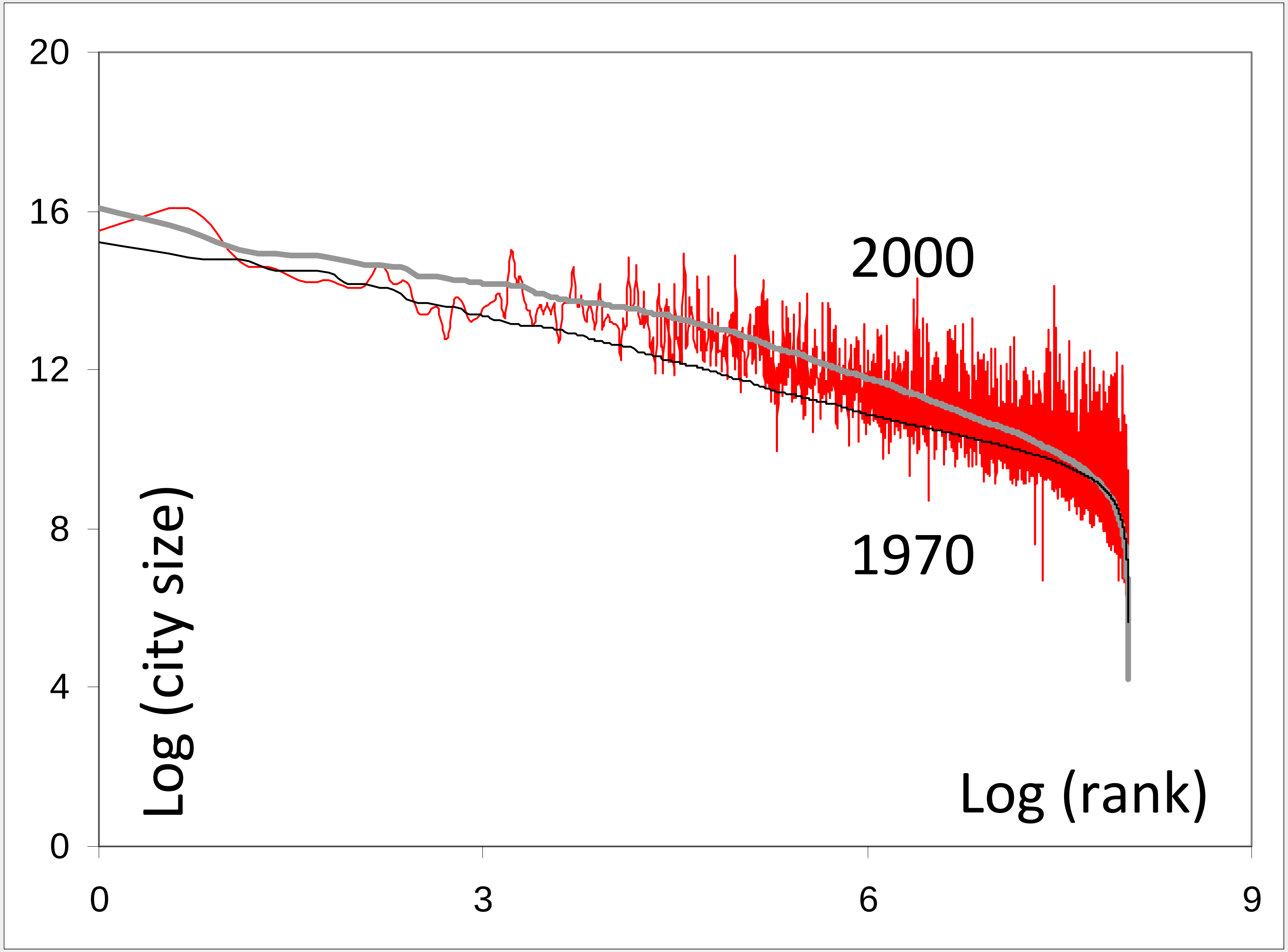


Year	r-squared	exponent
1790	0.975	0.876
1800	0.968	0.869
1810	0.989	0.909
1820	0.983	0.904
1830	0.990	0.899
1840	0.991	0.894
1850	0.989	0.917
1860	0.994	0.990
1870	0.992	0.978
1880	0.992	0.983
1890	0.992	0.951
1900	0.994	0.946
1910	0.991	0.912
1920	0.995	0.908
1930	0.995	0.903
1940	0.994	0.907
1950	0.990	0.900
1960	0.985	0.838
1970	0.980	0.808
1980	0.986	0.769
1990	0.987	0.744
2000	0.988	0.737


10000000.00
1000000.00
100000.00
10000.00
1000.00
100.00
10.00
1.00
1
10
100
1000
10000
100000
Log (city size)
Log (rank)



Parameter/Statistic	*1970*	*1980*	*1990*	*2000*
R Square	0.979	0.972	0.973	0.969
Intercept	16.790	16.891	17.090	17.360
Zipf-Exponent	*-0.986*	*-0.982*	*-0.995*	*-1.014*


20
16
12
8
4
0
0
3
6
9
Log (city size)
Log (rank)
2000
1970

Two Derivations: First, Random Dynamics

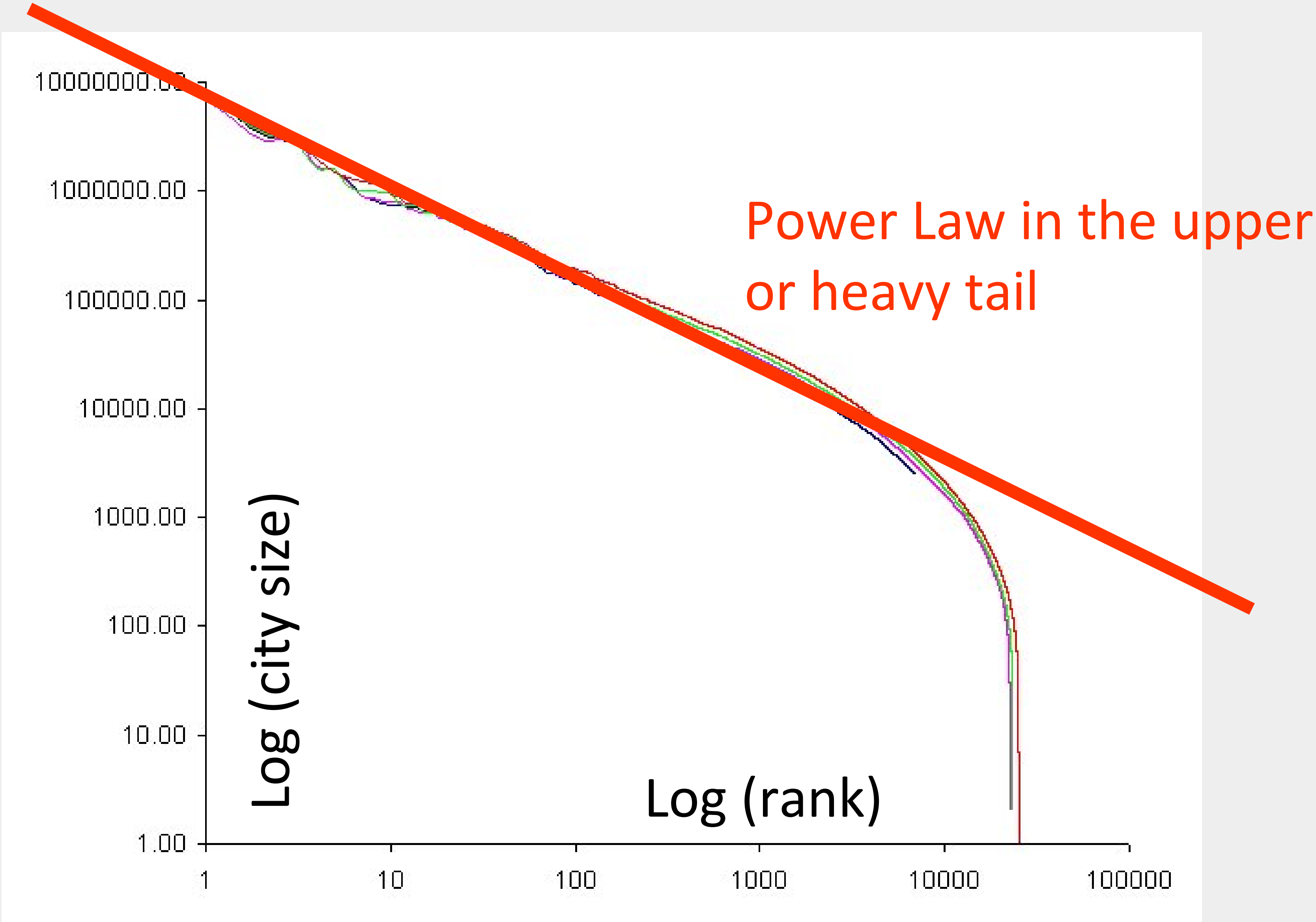
Assume all the cities or firms start evenly spaced. They can grow or decline randomly with a small randomly chosen rate of change which is applied proportionately to their size.

What happens is that it becomes increasingly unlikely for a place to grow big

This leads to a lognormal distribution

However if we don’t let the objects become zero – ie we establish a boundary condition, then what we get is a power law distribution.

Alternatively, if we look at the lognormal, with the largest objects and with a very large variance of the set of objects then in the heavy or upper tail, the power law is a good approximation (Gabaix, Solomon, Sornette)


Power Law in the upper or heavy tail
Log (city size)
Log (rank)
10000000.00
1000000.00
100000.00
10000.00
1000.00
100.00
10.00
1.00
1
10
100
1000
10000
100000

Second, Entropy again; the next set of slides

Well we will develop a worked example in the classroom where you are all given some money and then you exchange randomly fractions of this and we will show it leads to a negative exponential and from there a power law is quite easy to derive

My Work on Skyscrapers

Last thing if you look at the heights of skyscrapers then these are scaling – and I will develop this talk in the CASA conference on the 13th April prior to GISRUK which I hope you will all come to – it is free

www.casa.ucl.ac.uk/conference/