14.582: International Trade II
— Lecture 16: Economic Geography (Empirics I)

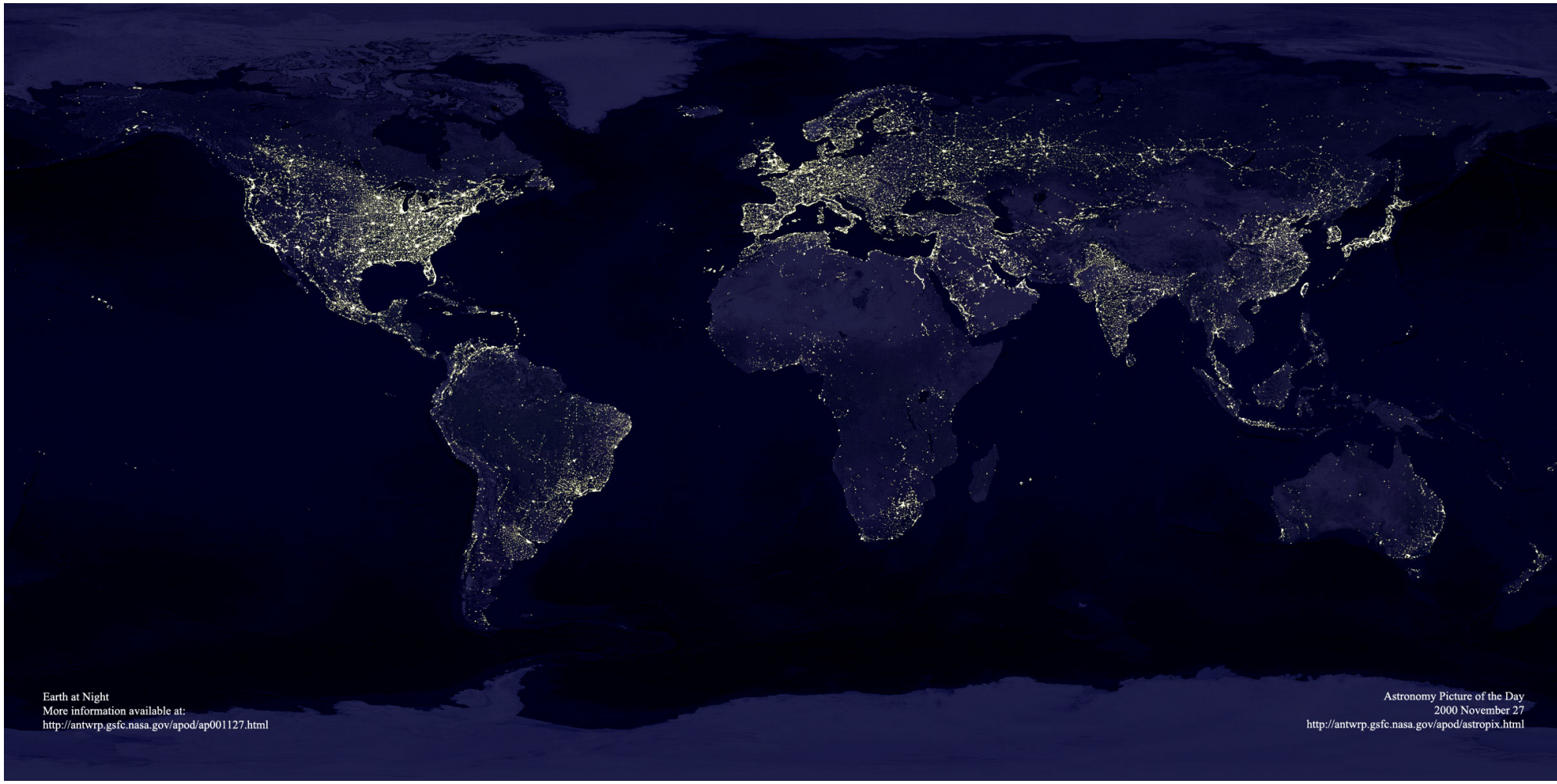
Plan for Today's Lecture

- Stylized facts about agglomeration of economic activity:
  1. Some pictures
  2. A systematic approach to measuring agglomeration
  3. Why agglomeration? A first look, via the patterns of co-agglomeration

Plan for Today's Lecture

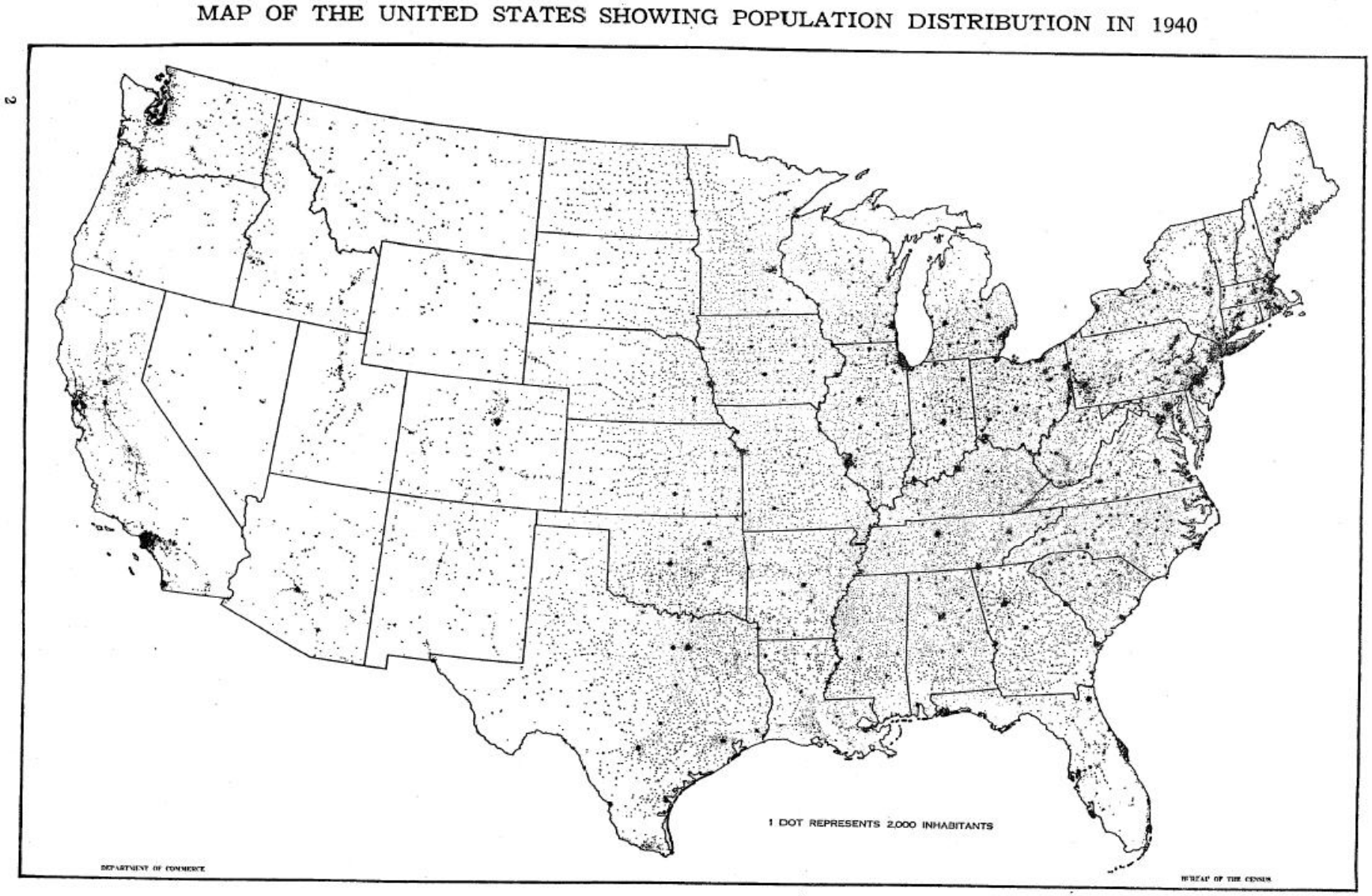
- **Stylized facts about agglomeration of economic activity:**
  1. **Some pictures**
  2. A systematic approach to measuring agglomeration
  3. Why agglomeration? A first look, via the patterns of co-agglomeration

The Earth at Night



The US at Night

The US "at night" (1940)



More Sophisticated Use of Satellite Data: Burchfield et al (2006, QJE)

For more examples of the use of such satellite ("remote sensing") data in economics, see Donaldson and Storeygard (JEP, 2016)

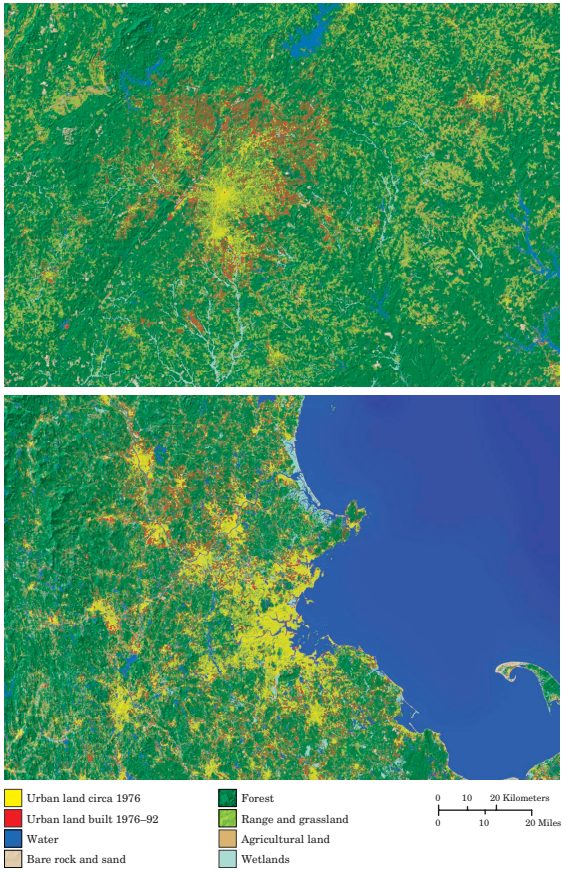


FIGURE IIa
Urban Land in Atlanta, GA (Top Panel) and Boston, MA (Bottom Panel)

More Sophisticated Use of Satellite Data: Burchfield et al (2006, QJE)

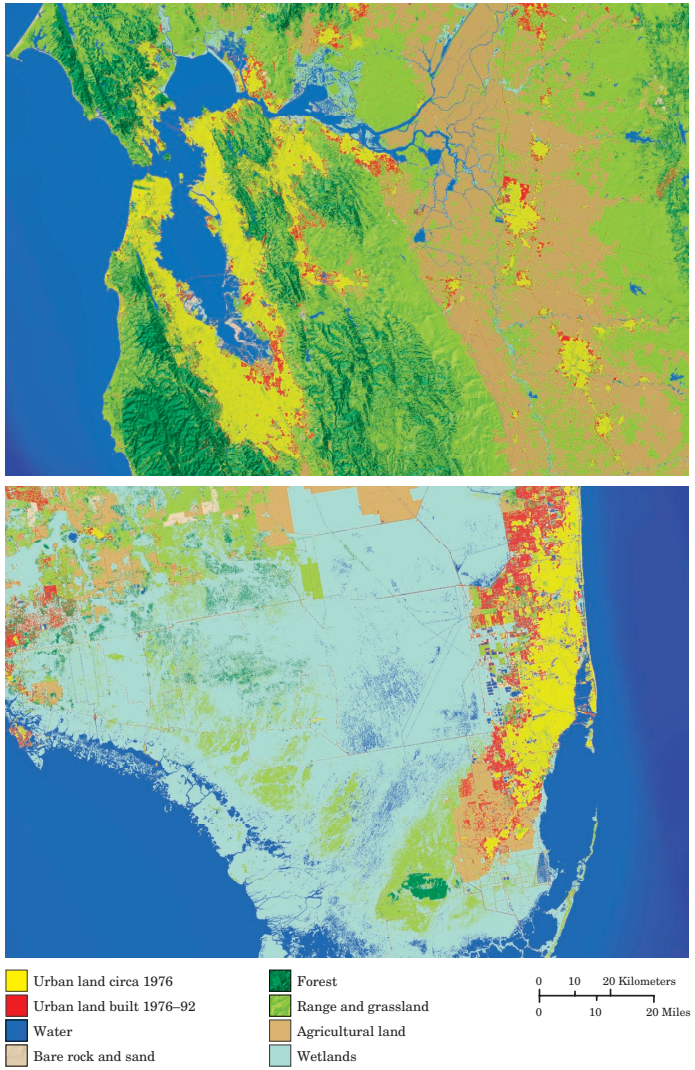


FIGURE IIb
Urban Land in San Francisco, CA (Top Panel) and Miami, FL (Bottom Panel)

More Sophisticated Use of Satellite Data: Burchfield et al (2006, QJE)

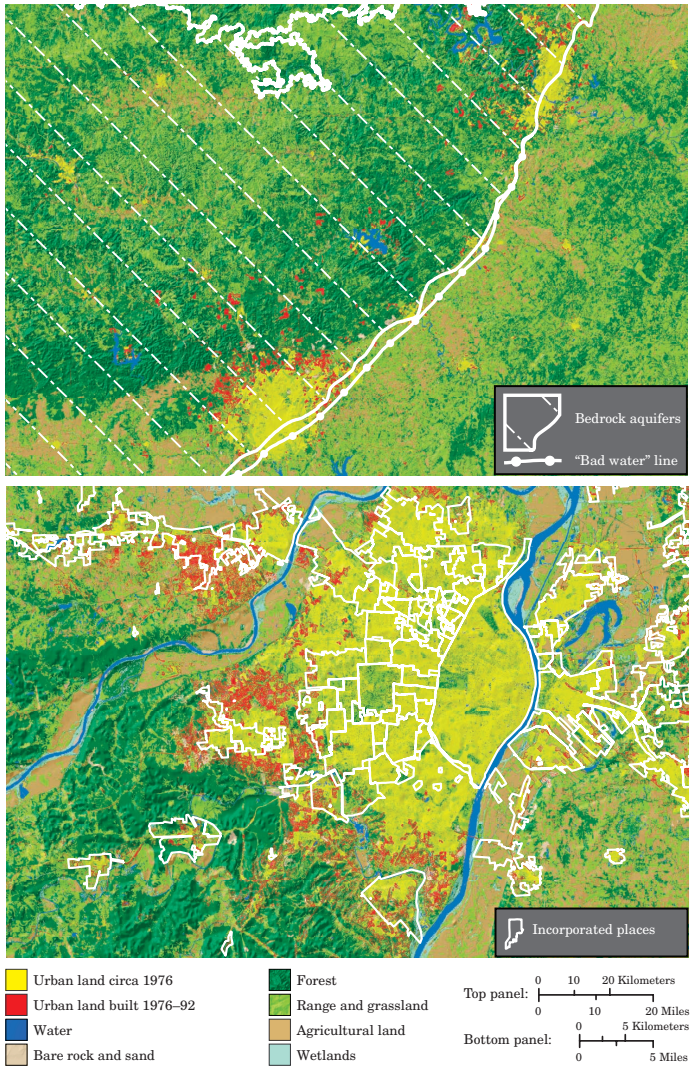


FIGURE III
Urban Land and Aquifers in San Antonio and Austin, TX (Top Panel), and Urban Land and Incorporated Places in Saint Louis, MO (Bottom Panel)

What about concentration of individual industries?

- No shortage of examples/anecdotes...
  - Marshall's original examples (cutlery in Sheffield; jewelry in Birmingham)
  - Silicon Valley, (Route 128?)
  - Detroit
  - Vegas
  - Dalton, GA

Holmes and Stevens (Hbk Urban/Regional, 2004)

Figure 2:
Location of Large Manufacturing Plants (1947)

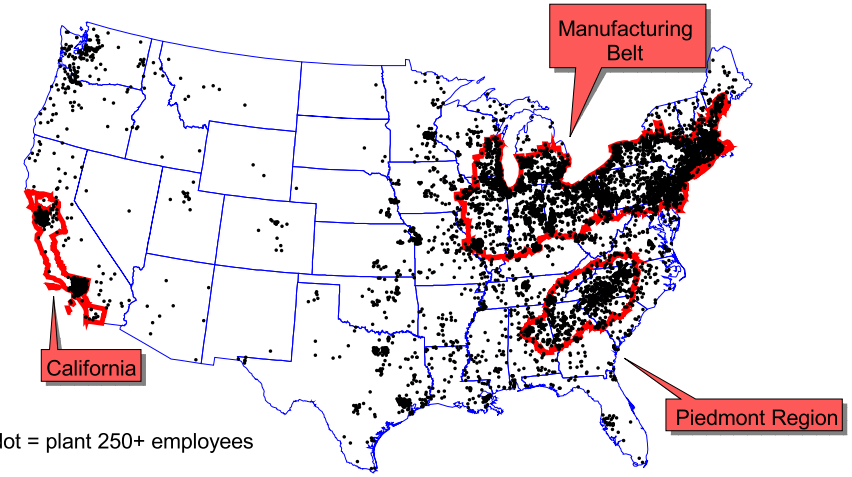
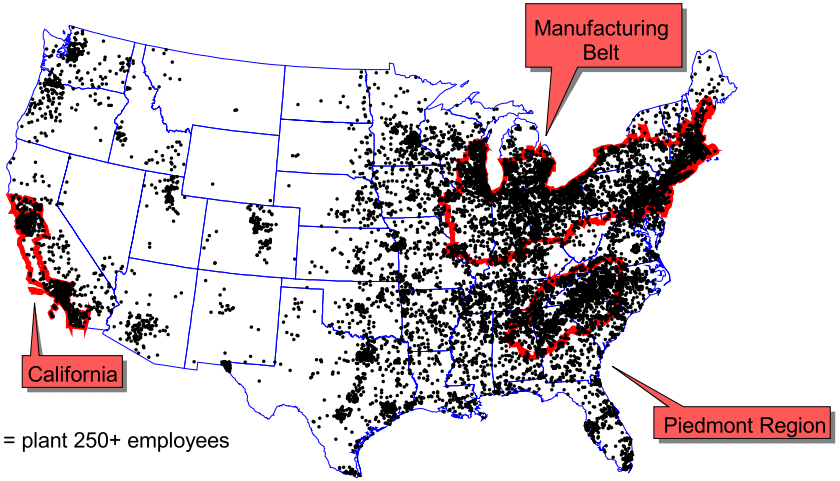
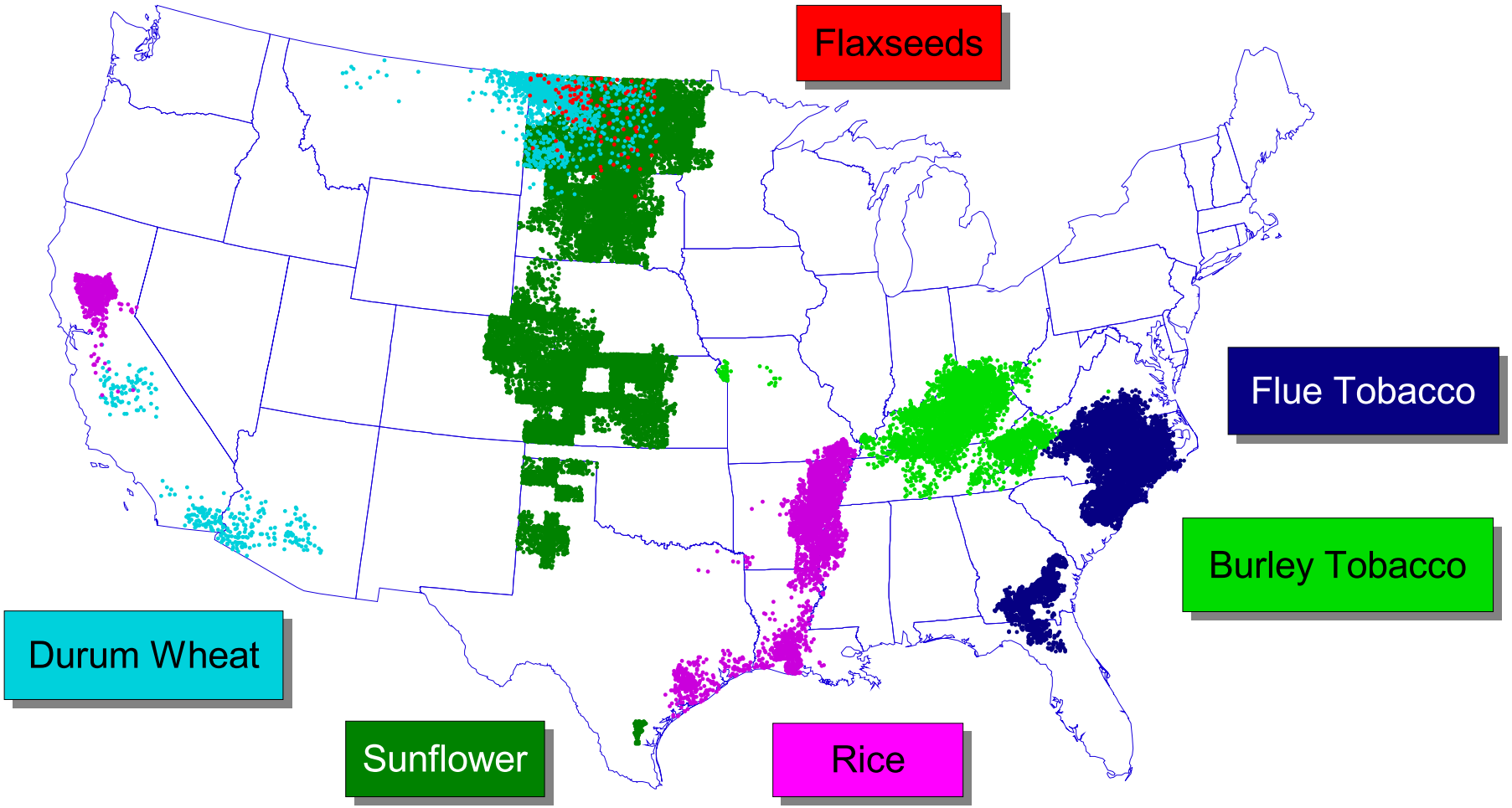


Figure 3:
Location of Large Manufacturing Plants (1999)



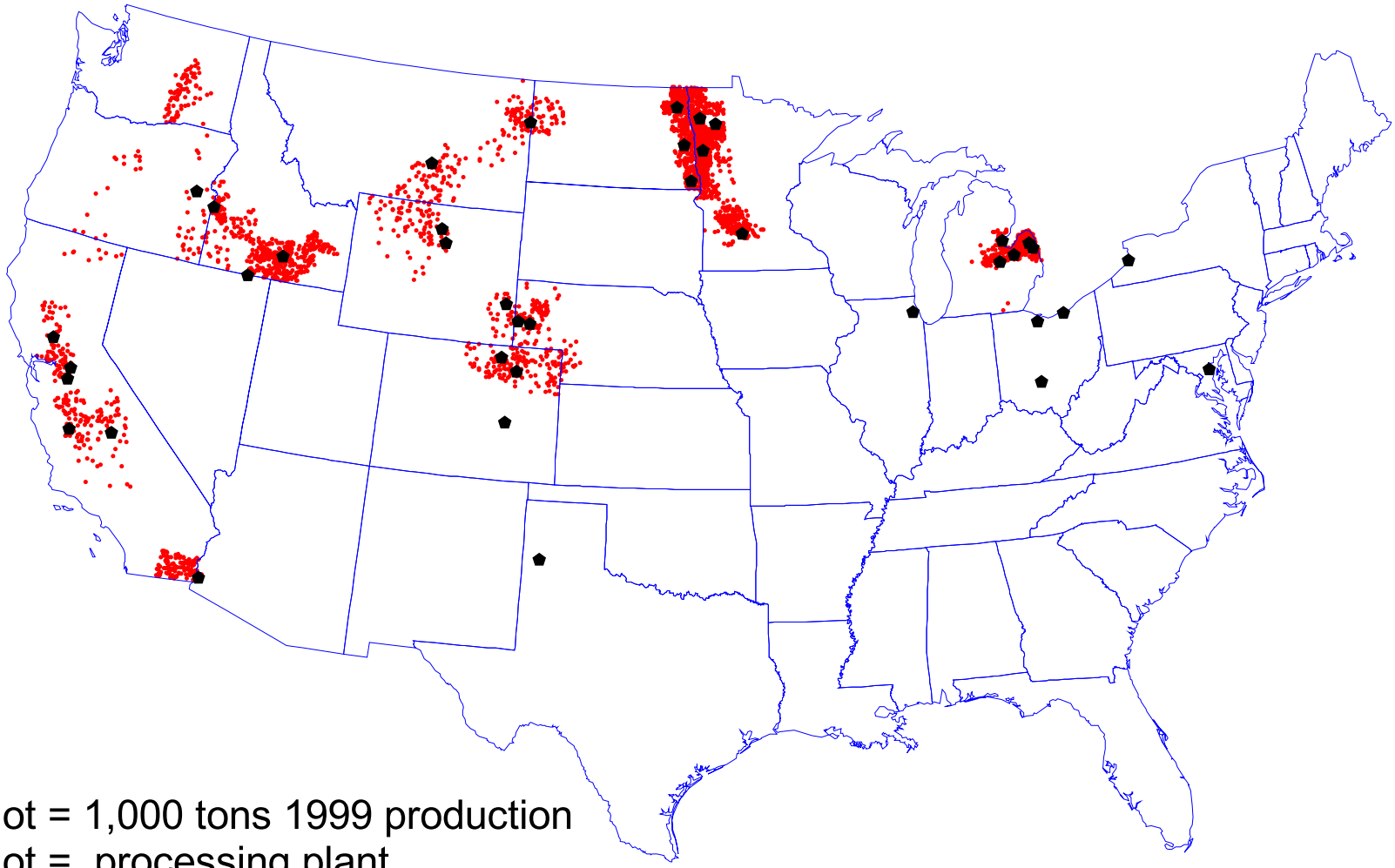
Holmes and Stevens (Hbk Urban/Regional, 2004)

Figure 4:
Location of Durum Wheat, Rice, Flue Tobacco, and Burley Tobacco



1 dot = 50,000 tons

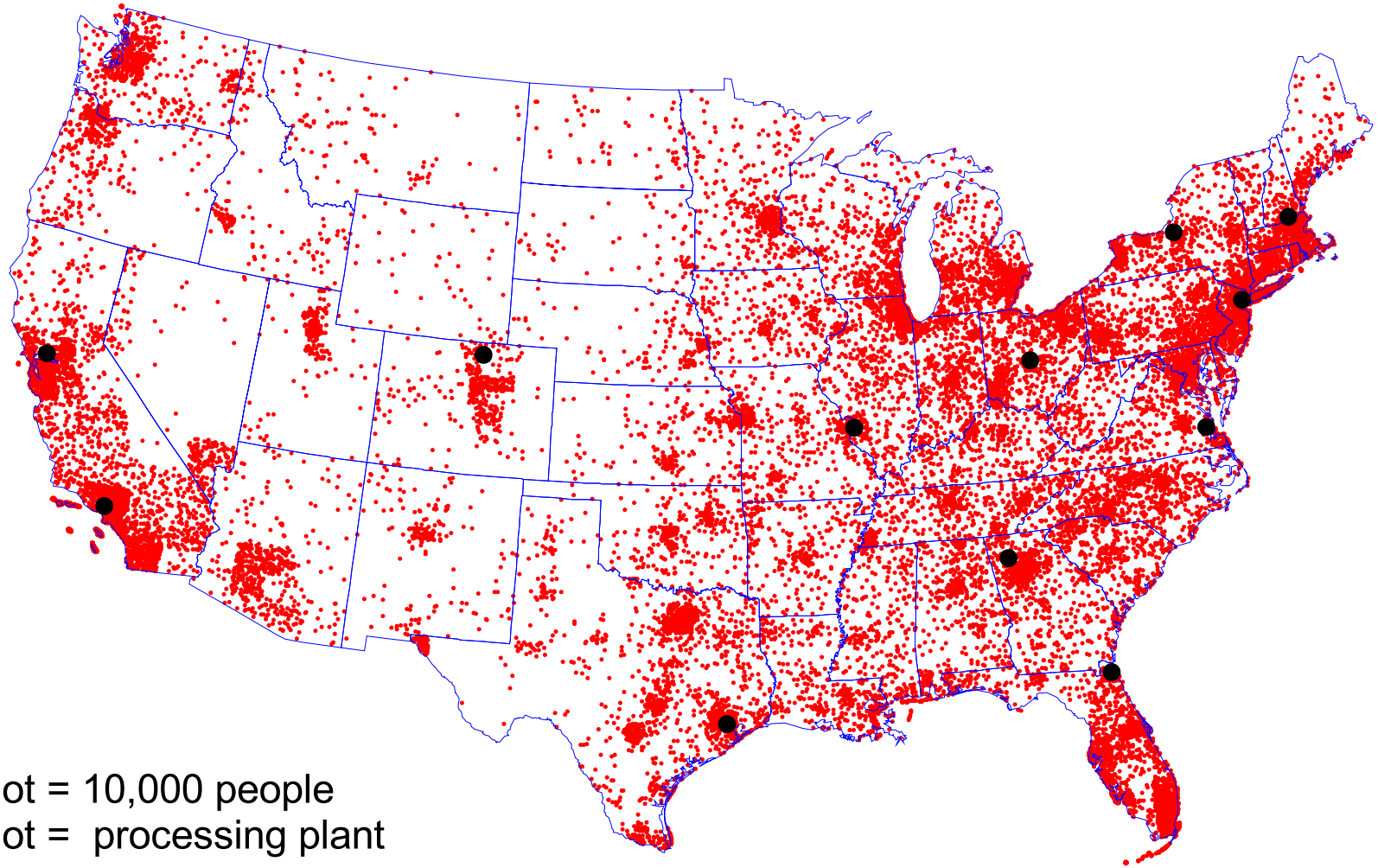
Figure 5:
Location of Sugar Beet Plants and Sugar Beet Crops



● 1 dot = 1,000 tons 1999 production
● 1 dot = processing plant

Holmes and Stevens (Hbk Urban/Regional, 2004)

Figure 6:
Location of Anheuser-Busch Breweries and Population (2000)



Plan for Today's Lecture

- **Stylized facts about agglomeration of economic activity:**
  1. Some pictures
  2. **A systematic approach to measuring agglomeration**
  3. Why agglomeration? A first look, via the patterns of co-agglomeration

Ellison and Glaeser (JPE, 1997)

- EG (1997) aims to go beyond the anecdotes and ask just how concentrated is economic activity within any given industry in the US?
- Key point: What is the right null hypothesis?
  - If output, within an industry, is highly concentrated in a small number of plants, then that industry will look very concentrated spatially, simply by nature of the small number of plants. (Consider extreme case of one plant.)
- EG develop an index of localization that considers as its null hypothesis the random location of plants within an industry. They call this a "dartboard approach".

A Model to Motivate the EG Index

- Suppose industry has N units that choose where to locate (among M regions), sequentially. - kth unit gets profits from location i of

$$\log \pi_{ki} = \log \bar{\pi}_i + g(v_1, ..., v_{k-1}) + \varepsilon_{ki}$$

- Where:
  - $\bar{\pi}_i$ is systematic payoff from location i
  - ε_{ki} is idiosyncratic payoff
  - $g(.)$ is a spillover term—effects of earlier firms' location choices v_k

A Model to Motivate the EG Index

- Suppose (for now) that $g(\cdot) = 0$. Let ε_{ki} be Gumbel-distributed so, conditional on realizations of $\bar{\pi}_1, ..., \bar{\pi}_M$, we have

$$Prob(v_k = i | \bar{\pi}_1, ..., \bar{\pi}_M) = \frac{\bar{\pi}_i}{\sum_j \bar{\pi}_j}$$

- Then restrict $\bar{\pi}_i$ to be drawn from distribution with:
  - $E\left[\frac{\bar{\pi}_i}{\sum_j \bar{\pi}_j}\right] = x_i$, where x_i is location i's share of overall manuf. employment (though very similar results if x_i is instead just location's population)
  - $Var\left[\frac{\bar{\pi}_i}{\sum_j \bar{\pi}_j}\right] = \gamma^{NA} x_i(1 - x_i)$, with $\gamma^{NA} \in [0, 1]$. This means that γ^{NA} (NA stands for "natural advantages") governs importance of systematic component $\bar{\pi}_i$ relative to idiosyncratic one.

A Model to Motivate the EG Index

- For the spillover term assume that

$$g(.) = \sum_{l \neq k} e_{kl}(1 - u_{li})(-\infty)$$

- Where:
  - e_{kl} is a Bernoulli random variable equal to one with probability γ^S. Don't need to specify dependence, but assume that e_{kl} is symmetric and transitive. (This will then imply that the sequential entry equilibrium is also the rational expectations equilibrium for any entry order).
  - u_{li} is indicator variable equal to one if plant l is located in location i. So spillovers restricted to take place only within locations, and not at all across locations. (More on that below.)

The EG Index

- Proposition 1:
  - The following is an unbiased estimator for $\gamma \equiv \gamma^{NA} + \gamma^{S} - \gamma^{NA}\gamma^{S}$:

$$\gamma = \frac{G - (1 - \sum_i x_i^2)H}{(1 - \sum_i x_i^2)(1 - H)}$$

  - Where:
    - $G \equiv \sum_i (s_i - x_i)^2$, with $s_i \equiv \sum_k z_k u_{ki}$, where z_k is the employment (taken as exogenous) of plant k and u_{ki} is (again) the indicator for whether plant k locates in location i.
    - $H \equiv \sum_k z_k^2$ is the industry's (employment-based) Herfindahl over plants.

- Note that γ can't separately identify the spillover (γ^{S}) and natural advantage (γ^{NA}) determinants of agglomeration. (But it would be worrying if it could!)

EG (1997): Results ("*Slight* concentration is remarkably widespread")

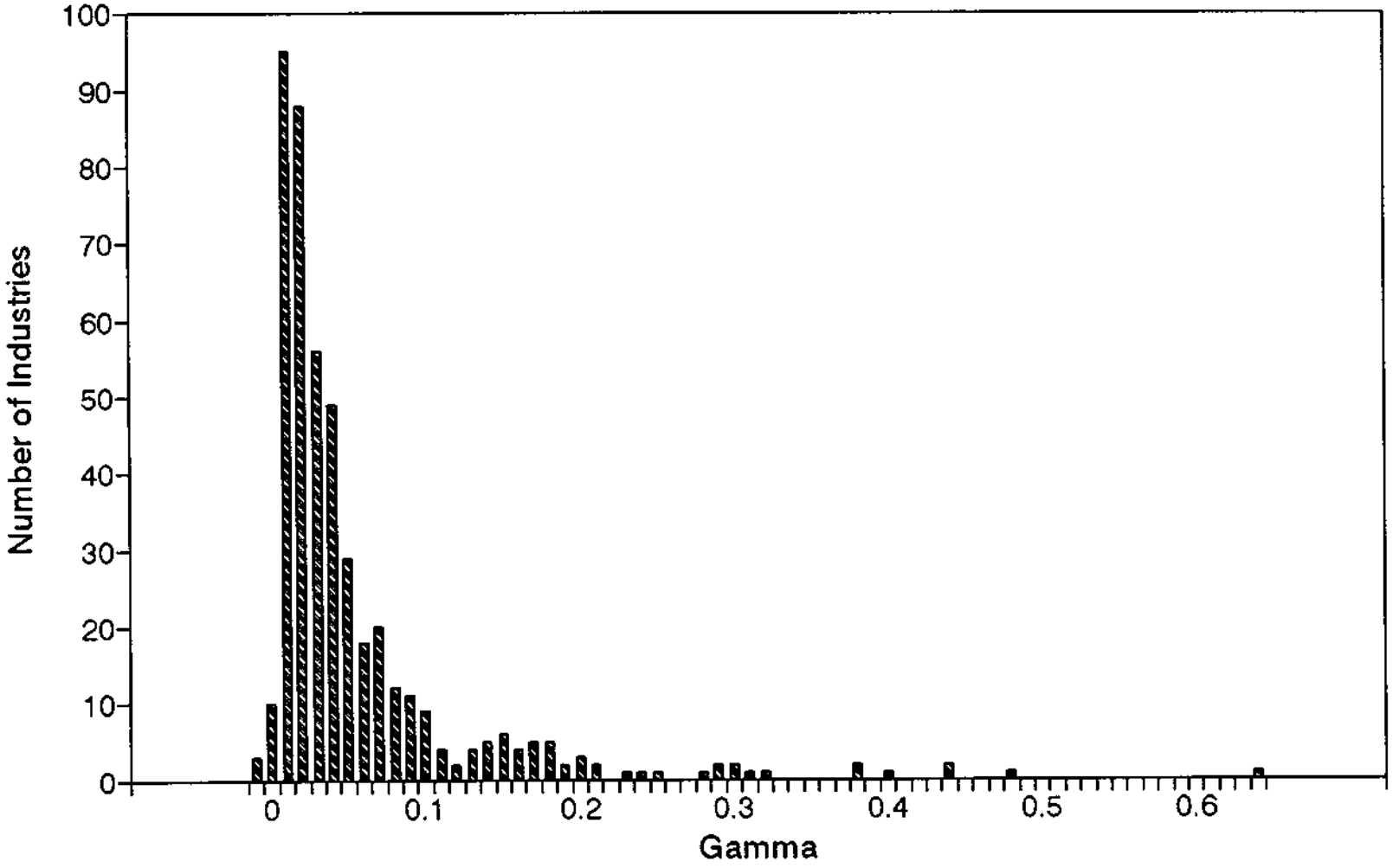


FIG. 1.—Histogram of γ (four-digit industries)

EG (1997): Results

- For industries that we might expect to be highly concentrated:
  - Autos: $\gamma = 0.127$
  - Auto parts: $\gamma = 0.089$
  - Carpets (i.e. Dalton, GA): $\gamma = 0.378$
  - Electronics (i.e. Silicon Valley): $\gamma = 0.059 - 0.142$
- For industries that we might expect to not be highly concentrated:
  - Bottled/canned soft drinks: $\gamma = 0.005$
  - Newspaper: $\gamma = 0.002$
  - Concrete: $\gamma = 0.012$
  - Ice: $\gamma = 0.012$

EG (1997): Results

TABLE 4

Most and Least Localized Industries

Four-Digit Industry	H	G	γ
	15 Most Localized Industries		
2371 Fur goods	.007	.60	.63
2084 Wines, brandy, brandy spirits	.041	.48	.48
2252 Hosiery not elsewhere classified	.008	.42	.44
3533 Oil and gas field machinery	.015	.42	.43
2251 Women's hosiery	.028	.40	.40
2273 Carpets and rugs	.013	.37	.38
2429 Special product sawmills not elsewhere classified	.009	.36	.37
3961 Costume jewelry	.017	.32	.32
2895 Carbon black	.054	.32	.30
3915 Jewelers' materials, lapidary	.025	.30	.30
2874 Phosphatic fertilizers	.066	.32	.29
2061 Raw cane sugar	.038	.30	.29
2281 Yarn mills, except wool	.005	.27	.28
2034 Dehydrated fruits, vegetables, soups	.030	.29	.28
3761 Guided missiles, space vehicles	.046	.27	.25
	15 Least Localized Industries		
3021 Rubber and plastics footwear	.06	.05	−.013
2032 Canned specialties	.03	.02	−.012
2082 Malt beverages	.04	.03	−.010
3635 Household vacuum cleaners	.18	.17	−.009
3652 Prerecorded records and tapes	.04	.03	−.008
3482 Small-arms ammunition	.18	.17	−.004
3324 Steel investment foundries	.04	.04	−.003
3534 Elevators and moving stairways	.03	.03	−.001
2052 Cookies and crackers	.03	.03	−.0009
2098 Macaroni and spaghetti	.03	.03	−.0008
3262 Vitreous china table, kitchenware	.13	.12	−.0006
2035 Pickles, sauces, salad dressings	.01	.01	−.0003
3821 Laboratory apparatus and furniture	.02	.02	−.0002
2062 Cane sugar refining	.11	.10	.0002
3433 Heating equipment except electric	.01	.01	.0002

Duranton and Overman (REStud, 2005)

- DO point out a few problems with the EG index:
  1. What if spillovers are truly smooth over space (as seems natural) rather than discrete at location borders? Then measurement error caused by discrete location borders will cause downward bias in concentration measures. (Special case of what geographers call the "modifiable unit area problem".)
  2. How to think of statistical significance of finding that an industry is "concentrated"?

Duranton and Overman (2005)

(Effectively) point-like location data from the UK

(a) Basic Pharmaceuticals (SIC2441)

(b) Pharmaceutical Preparations (SIC2442)

(c) Other Agricultural and Forestry Machinery (SIC2932)

(d) Machinery for Textile, Apparel and Leather Production (SIC2954)

FIGURE 1

EG (1997): By spatial unit

Lower values for county-level spatial unit may be consistent with DO's first point.

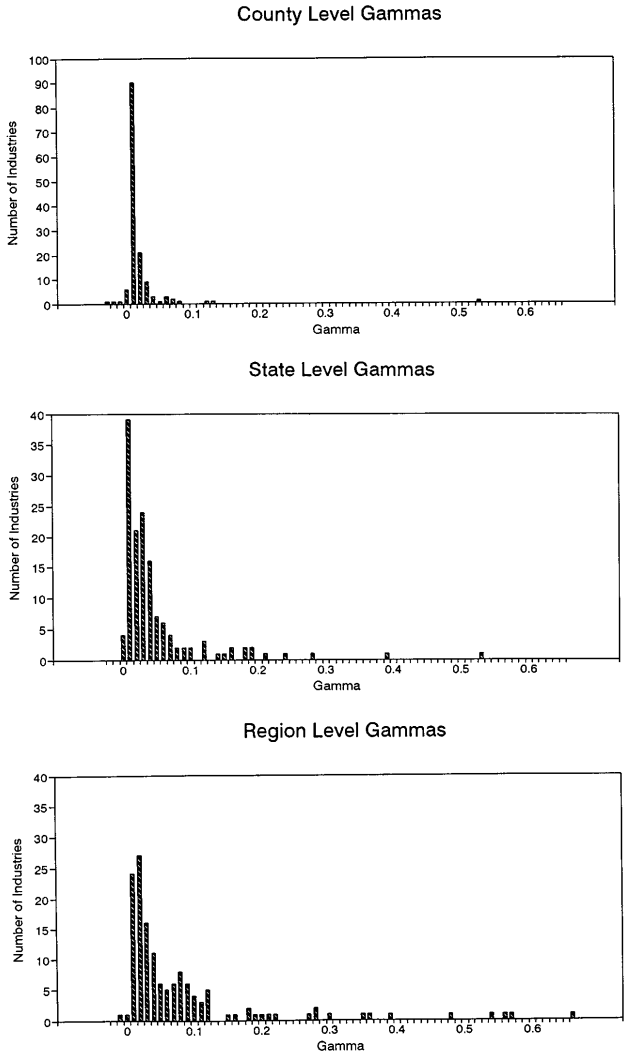


FIG. 2.—Concentration at the county, state, and regional levels

The DO Index

- DO propose the following index (for the employment-weighted concentration of any given industry, at distance band "d"):

$$K^{emp}(d) = \frac{1}{h\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} e(i)e(j)} \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} e(i)e(j) f\left(\frac{d-d_{ij}}{h}\right)$$

- Where:
  - d_{ij} is the distance between plant i and plant j
  - $f(\cdot)$ is a (Gaussian) kernel function (and h some bandwidth) to smooth out for measurement error in distances

- Hypothesis testing done via simulation (counterfactual distributions obtained from drawing plants with replacement from set of active locations)

DO Index: function of distance (for 4 sample industries)

Global confidence bands are a Bonferroni-like correction over all distances

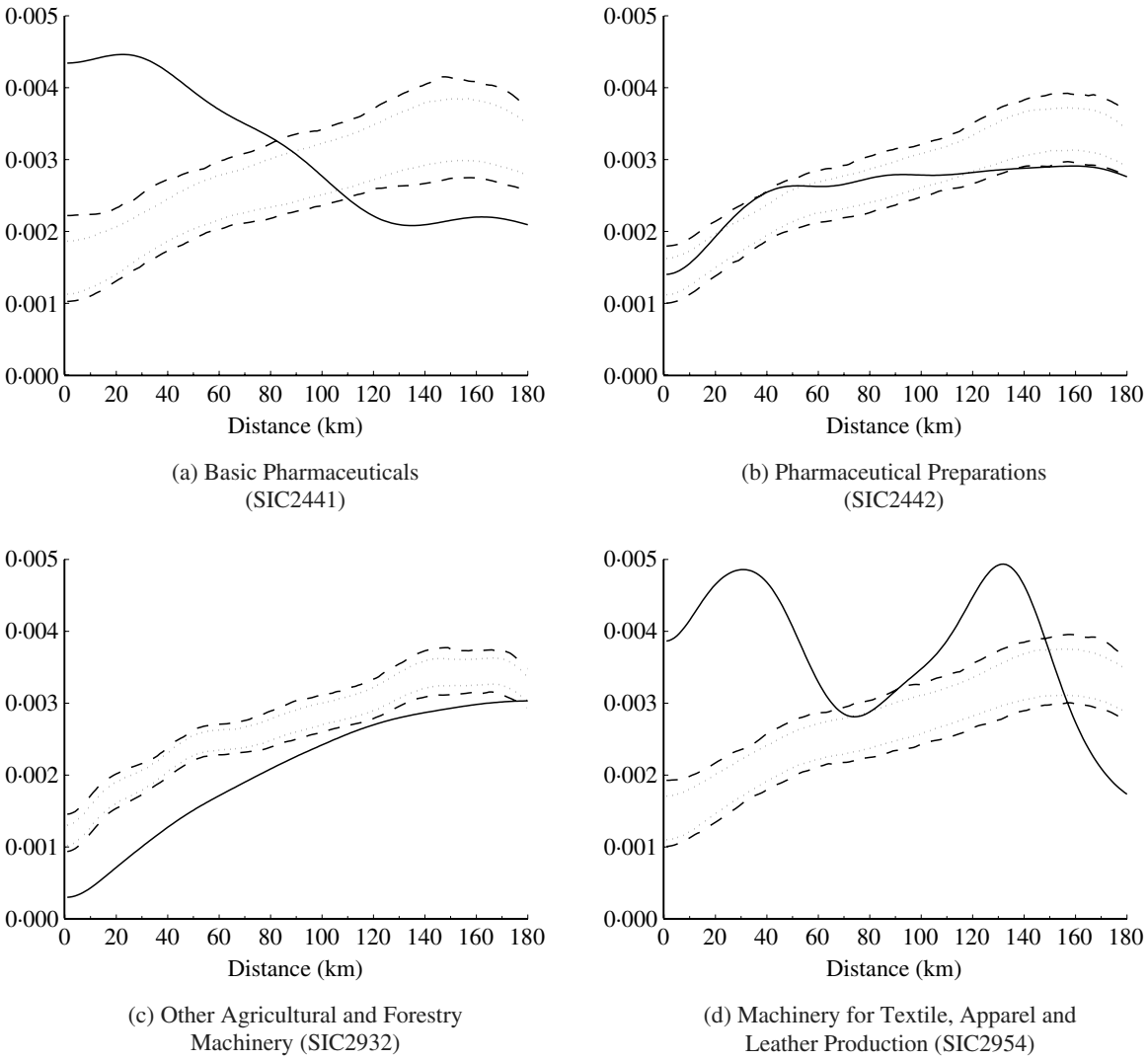


FIGURE 2

K-density, local confidence intervals and global confidence bands for four illustrative industries

Plan for Today's Lecture

- **Stylized facts about agglomeration of economic activity:**
  1. Some pictures
  2. A systematic approach to measuring agglomeration
  3. **Why agglomeration? A first look, via the patterns of co-agglomeration**

What Causes Agglomeration?

- Marshall postulated three different "transport" costs:
  1. *Goods:* Firms (with indivisible production processes) will locate near suppliers and customers to avoid shipping costs
  2. *People:* Benefits of thick labor markets ("labor pooling")
  3. *Ideas:* Knowledge spillovers that decay with distance ("Mysteries of the trade become no mystery, but are, as it were, in the air.")
- Ellison, Glaeser and Kerr (AER, 2010) seek to ask how well these explanations do, using evidence from industry "co-agglomeration" (as was actually defined in original EG 1997 paper)

Back to EG Index: Co-agglomeration

- Now suppose that:
  - $corr(u_{km}, u_{lm}) = \gamma_j$ if plants k and l both belong to industry j and locate in location m, where γ_j is the same γ as above (in the single-industry case) but the value of γ for industry j.
  - $corr(u_{km}, u_{lm}) = \gamma_{ij}$ if instead plant k is in industry i and plant l is in industry j.
  - So γ_{ij} measures tendency for industries i and j to co-agglomerate.

- Then Proposition 2: An unbiased estimator of γ_{ij} is given by

$$\gamma_{ij}^{C} \equiv \frac{\sum_m (s_{mi} - x_m)(s_{mj} - x_m)}{1 - \sum_m x_m^2}$$

EGK (2010): Co-agglomeration Index

TABLE 1—DESCRIPTIVE STATISTICS FOR PAIRWISE COAGGLOMERATION REGRESSIONS

	Mean	SD	Minimum	Maximum
Panel A. Pairwise EG coagglomeration measures				
EG state total employment coagglomeration	0.000	0.013	−0.065	0.207
EG PMSA total employment coagglomeration	0.000	0.006	−0.025	0.119
EG county total employment coagglomeration	0.000	0.003	−0.018	0.080
EG state firm birth employment coagglomeration	0.000	0.015	−0.082	0.259
EG expected coagglomeration due to natural advantages	0.000	0.001	−0.008	0.022

	Industry count	Relevant industries (nonzero) Mean	SD	Maximum
Panel B. Pairwise DO coagglomeration measures				
DO global localization coagglomeration, 1,000 mi.	7,371	0.133	0.073	0.454
DO global dispersion coagglomeration, 1,000 mi.	10	0.592	0.078	0.746
DO expected global localization coagglomeration, 1,000 mi.	7,381	0.181	0.027	0.256
DO global localization coagglomeration, 250 mi.	6,429	0.017	0.019	0.283
DO global dispersion coagglomeration, 250 mi.	952	0.042	0.029	0.307
DO expected global localization coagglomeration, 250 mi.	7,381	0.029	0.010	0.077

EGK (2010): Co-agglomeration Index

Table 2—Highest Pairwise Coagglomerations

Rank	Industry 1	Industry 2	Coagglomeration
Panel A. EG index using 1987 state total employments			
1	Broadwoven mills, cotton (221)	Yarn and thread mills (228)	0.207
2	Knitting mills (225)	Yarn and thread mills (228)	0.187
3	Broadwoven mills, fiber (222)	Textile finishing (226)	0.178
4	Broadwoven mills, cotton (221)	Broadwoven mills, fiber (222)	0.171
5	Broadwoven mills, fiber (222)	Yarn and thread mills (228)	0.164
6	Handbags (317)	Photographic equipment (386)	0.155
7	Broadwoven mills, wool (223)	Carpets and rugs (227)	0.149
8	Carpets and rugs (227)	Yarn and thread mills (228)	0.142
9	Photographic equipment (386)	Jewelry, silverware, plated ware (391)	0.139
10	Textile finishing (226)	Yarn and thread mills (228)	0.138
11	Broadwoven mills, cotton (221)	Textile finishing (226)	0.137
12	Broadwoven mills, cotton (221)	Carpets and rugs (227)	0.137
13	Broadwoven mills, cotton (221)	Knitting mills (225)	0.136
14	Carpets and rugs (227)	Pulp mills (261)	0.110
15	Jewelry, silverware, plated ware (391)	Costume jewelry and notions (396)	0.107

Ellison, Glaeser and Kerr (2010)

- Empirical proxies for 3 agglomeration mechanisms:
  1. *Goods:* from IO tables, max of input share or output share over the two industries in both directions
  2. *People:* Let $Share_{io}$ be the share of industry i's employees in occupation o, then compute $corr(Share_{io}, Share_{jo})$.
  3. *Ideas:* analogous to goods flows, but using patent citations (and Scherer (1984) "technology matrix" of R&D application relations)
- Also control for a measure of similarity in natural advantage—building on Ellison and Glaeser (AEA P&P, 1999) who used 16 locational characteristics multiplied by cost advantage of each
- Then simply regress γ_{ij}^{C} on these 4 variables (each standardized)

Ellison, Glaeser and Kerr (2010): OLS Results

TABLE 4—OLS MULTIVARIATE SPECIFICATIONS FOR PAIRWISE COAGGLOMERATION

	EG coaggl. index with state total emp.				DO coaggl. index, 250 mi.			
	Base estimation	Exclude natural advantages	Separate input & output	Exclude pairs in same SIC2	Base estimation	Exclude natural advantages	Separate input & output	Exclude pairs in same SIC2
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Natural advantages [DV specific]	0.163		0.162	0.172	0.251		0.252	0.253
	(0.017)		(0.017)	(0.016)	(0.012)		(0.012)	(0.013)
Labor correlation	0.118	0.146	0.114	0.085	0.069	0.098	0.066	0.029
	(0.011)	(0.012)	(0.011)	(0.012)	(0.012)	(0.013)	(0.012)	(0.012)
Input-output	0.146	0.149		0.110	0.162	0.150		0.177
	(0.032)	(0.032)		(0.022)	(0.035)	(0.035)		(0.032)
Input			0.106				0.097	
			(0.029)				(0.029)	
Output			0.093				0.107	
			(0.039)				(0.038)	
Technology flows Scherer R&D	0.096	0.112	0.079	0.046	0.076	0.075	0.065	0.033
	(0.035)	(0.035)	(0.035)	(0.019)	(0.033)	(0.034)	(0.032)	(0.020)
R^2	0.103	0.077	0.110	0.059	0.113	0.051	0.117	0.102
Observations	7,381	7,381	7,381	7,000	7,381	7,381	7,381	7,000

Notes: See Table 3. Regressions of pairwise coagglomeration on determinants of industrial co-location. Columns 4 and 8 exclude SIC3 pairwise combinations within the same SIC2. Online Appendix Table 6 provides additional robustness checks. Variables are transformed to have unit standard deviation for interpretation. Bootstrapped standard errors are reported in parentheses.

IV Approach

- Concern that the empirical proxies for agglomeration forces are themselves driven by agglomeration
  - E.g. is fact that shoe industry uses leather a purely technological feature, or partly the result of some other feature that caused shoes and leather to co-locate?

- Two IVs proposed:
  1. Measure the characteristics in the UK and IV with that. (Fine if random factors caused the unobserved reasons for coagglomeration in the US and the UK got different draws for those factors.)
  2. Find regions of the US where, say, industry j is located but i is not and (using Census microdata) measure the characteristics on that subset of plants

- NB: neither IV can be constructed for the knowledge spillovers variable so EGK just drop this regressor in the IV regressions

Ellison, Glaeser and Kerr (2010): IV Results

TABLE 5—IV MULTIVARIATE SPECIFICATIONS FOR PAIRWISE COAGGLOMERATION

	EG coaggl. index with state total emp.			DO coaggl. index, 250 mi.		
	Base OLS	UK IV	US spatial IV	Base OLS	UK IV	US spatial IV
	(1)	(2)	(3)	(4)	(5)	(6)
Natural advantages	0.173	0.173	0.171	0.254	0.210	0.233
[DV specific]	(0.016)	(0.019)	(0.016)	(0.013)	(0.016)	(0.012)
Labor correlation	0.083	0.079	0.091	0.027	0.501	0.248
	(0.012)	(0.060)	(0.023)	(0.012)	(0.060)	(0.023)
Input-output	0.122	0.191	0.185	0.186	0.164	0.213
	(0.023)	(0.048)	(0.036)	(0.031)	(0.054)	(0.049)
Observations	7,000	7,000	7,000	7,000	7,000	7,000

Notes: See Table 3. OLS and IV regressions of pairwise coagglomeration on determinants of industrial co-location. All estimations exclude SIC3 pairwise combinations within the same SIC2. Online Appendix Tables 7 and 8 report first stages and additional robustness checks. Variables are transformed to have unit standard deviation for interpretation. Bootstrapped standard errors are reported in parentheses.

Further Reading on Agglomeration Stylized Facts

- Other chapters in 2004 *Handbook of Urban Econ and Regional Science*:
  - Kim and Margo on historical facts from US
  - Chapters on European and Asian analogs of the Holmes and Stevens chapter
- City size distribution literature (e.g. Gabaix and Ioannides (2004) for survey)
- Dynamics of city sizes and specializations: Duranton (AER, 2007)