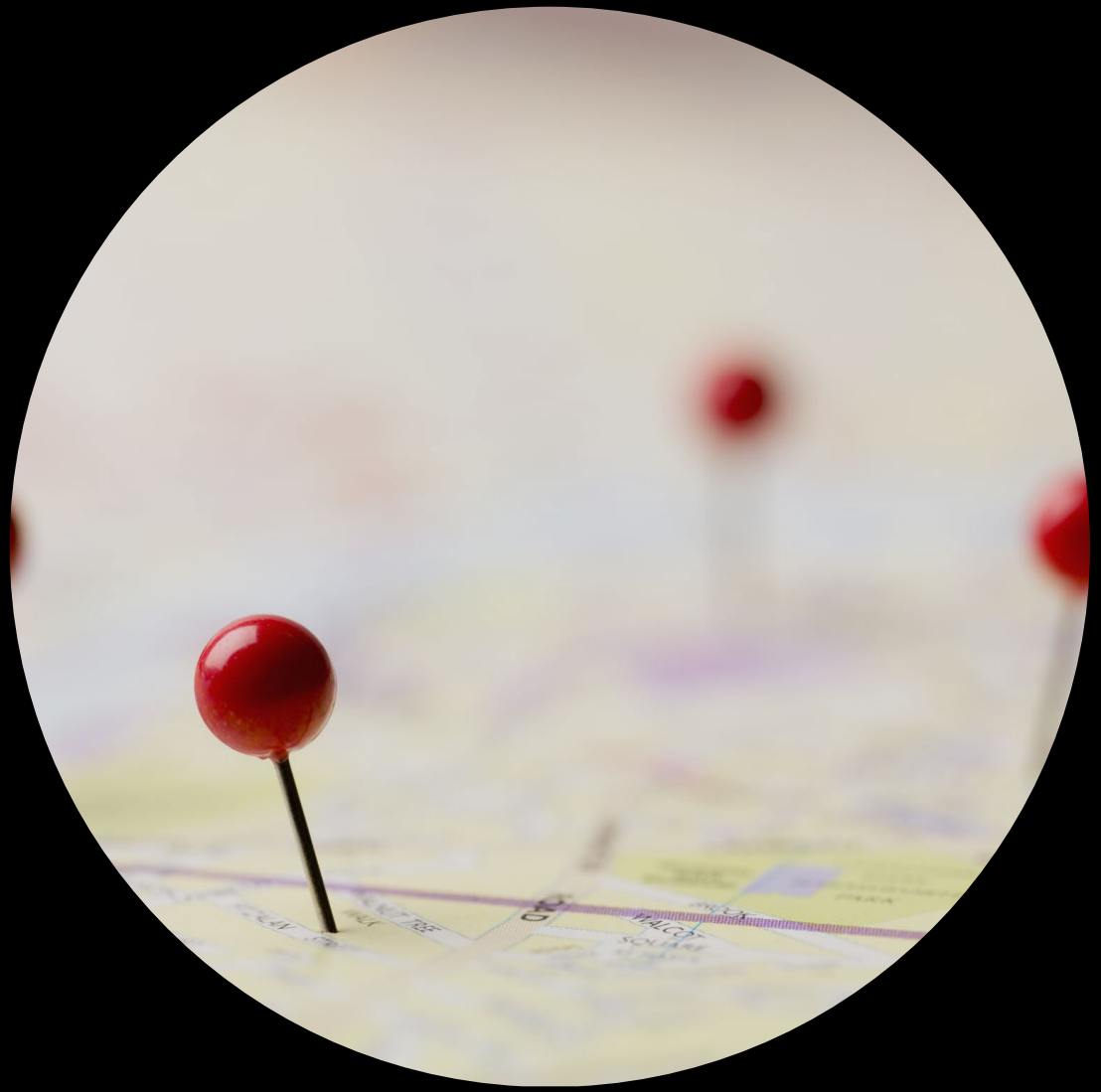


Once More, With Feeling: CSA's longtime collaboration with the OTC and a crash-course for sales-tax ESDA from first principles

OKSCAUG, September 16, 2025.

Leah Nash

Jim Anderson, GISP

University of Oklahoma Center for Spatial Analysis

We're CSA.

- The Center for Spatial Analysis at the University of Oklahoma supports Government, Private, University and Grant Research interests and partnerships by delivering **applied geospatial technology solutions and outreach.**
- CSA@OU is part of the Department of Geography and Environmental Sustainability in the College of Atmospheric and Geographic Sciences at OU.
- With CSA having assumed several forms since its inception in 1994, CSA in 2025 will celebrate its 22nd year of service to the Oklahoma Tax Commission on the Streamlined Sales Tax project, our 30th year of service to OTC overall. We'll address CSA's tradition of **sales and use tax support** shortly.
- Since 2019, our staff has grown, our leadership has been refreshed, and our contractual engagements with Oklahoma State Agencies have become more numerous.

CSA's Partnership with OTC: Background to SSUTA

- Oklahoma is a full member State of the Streamlined Sales and Use Tax Agreement (SSUTA), as-administered by the Streamlined Sales Tax Governing Board (SSTGB).
- What? Why? What are the Benefits?
  - The *Commerce Clause*, Article 1, Section 8, Clause 3 of the US Constitution grants congress the power to regulate interstate commerce. **Effectively, States themselves cannot collect taxes on interstate commerce.**
  - ***National Bellas Hess v. Department of Revenue of Illinois***, 386 U.S. 753 (1967)
    - Shows that mail-order sales delivered from out-of-state are not subject to sales/use tax.
  - ***Quill Corp. v. North Dakota***, 504 U.S. 298 (1992)
    - Affirmed *Bellas Hess v. Illinois*, this time with respect to software purchased over the internet.
    - For example, Amazon.com then avoids collection of sales tax on interstate sales largely between 1995 and 2012. **This was an anticompetitive period that objectively hampered local governments and in-state business operations.**
  - ***South Dakota v. Wayfair, Inc.***, 585 U.S. ___ (2018)
    - A 5-4 Justice majority overturns *Quill v. ND* and *Bellas Hess v. Illinois* at-once.
    - June 21, 2018 – States can require online retailers to collect sales taxes on out-of-state sales.

CSA's Partnership with OTC: Background to SSUTA

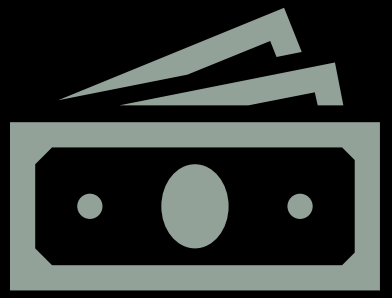
- Member States essentially agree to devolve some of their autonomy with respect to setting and collecting sales taxes, such that a simplified system **reducing the burden of compliance** can be established.
  - For example, only one tax rate may be applied per taxable jurisdiction; with the exception of the option to assess a separate rate for food and drugs. States also fund *Certified Service Providers* to act as tax processors for out-of-state businesses selling into a given SST state.
  - "Emerging Technologies" are central to supporting the SSUTA vision, particularly where - **One of the most difficult tax issues for a business is keeping up with the local taxes and knowing when a sale is inside or outside a local government jurisdiction.**
  - *States that have met the requirement of providing a Zip Code database may elect to certify **vendor provided address-based databases** for assigning tax rates and jurisdictions.*

- CSA serves as OTC's address-based database vendor for this purpose. Our goal is to associate a set of taxable jurisdictions to **every delivery address in Oklahoma.**

CSA's Partnership with OTC: Technical Assistance as *Certified Vendor of Address-Databases*

- Termed "SST Rate and Boundary Files", exports of these databases are publicly available:
  - https://www.streamlinedsalestax.org/Shared-Pages/rate-and-boundary-files
- The preparation of these data products is intensively spatial, in that a great deal of technical labor is invested in **tracking municipal boundary changes, and reflecting those changes in address-range records** at CSA.
  - While automation is very helpful throughout this process, **there is often no practical replacement** for interpretation of a given address-range's participation in a taxable jurisdiction, on-the-map.
- SST Deliverables (Rate and Boundary Files) are released to OTC quarterly. CSA maintains a website allowing the Public to perform searches across these data:
  - https://taxproject.csa.ou.edu/Rate_Locator/

Locating your Sales Tax Rate



Am I paying the correct rate?

https://taxproject.csa.ou.edu/Rate_Locator/

What search-related metadata does CSA Collect?

The *public IP* address of the client (the browser) sending the search request

The parameters of the search (varies by search type: *Street Names, Zip Codes, etc.* as-submitted by the User)

The *Lookup Number* associated with the search (for traceability and to provide the taxpayer with a means to identify their search, and therefore the result received at-the-time)

A Flag indicating *success or failure* of each search.

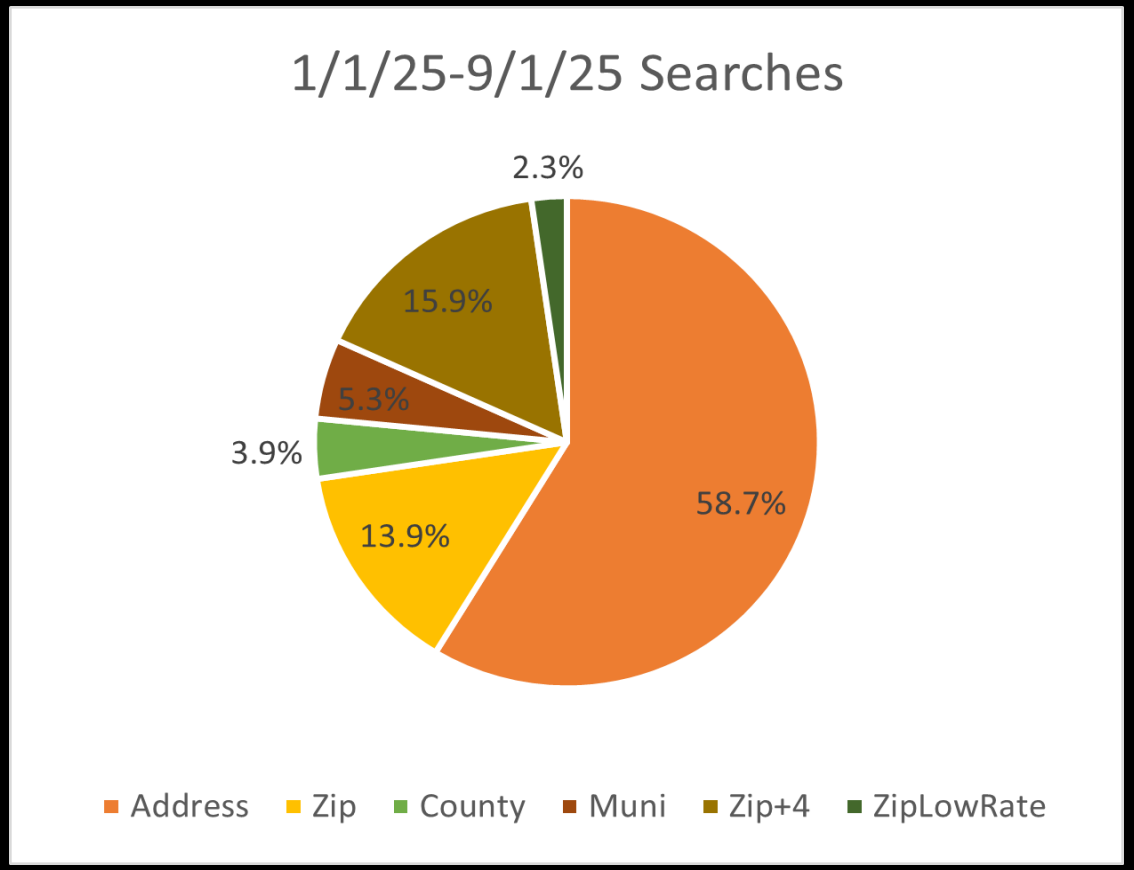
A *timestamp* capturing the submission time of the search

What can be drawn from this information?

- All manner of site usage statistics!
  - *Most searched for X*
  - *How many of X search type* in each of YYYY
  - *Trend in* success/failed searches over YYYY-YYYY
  - *Top X Users* by logged searches

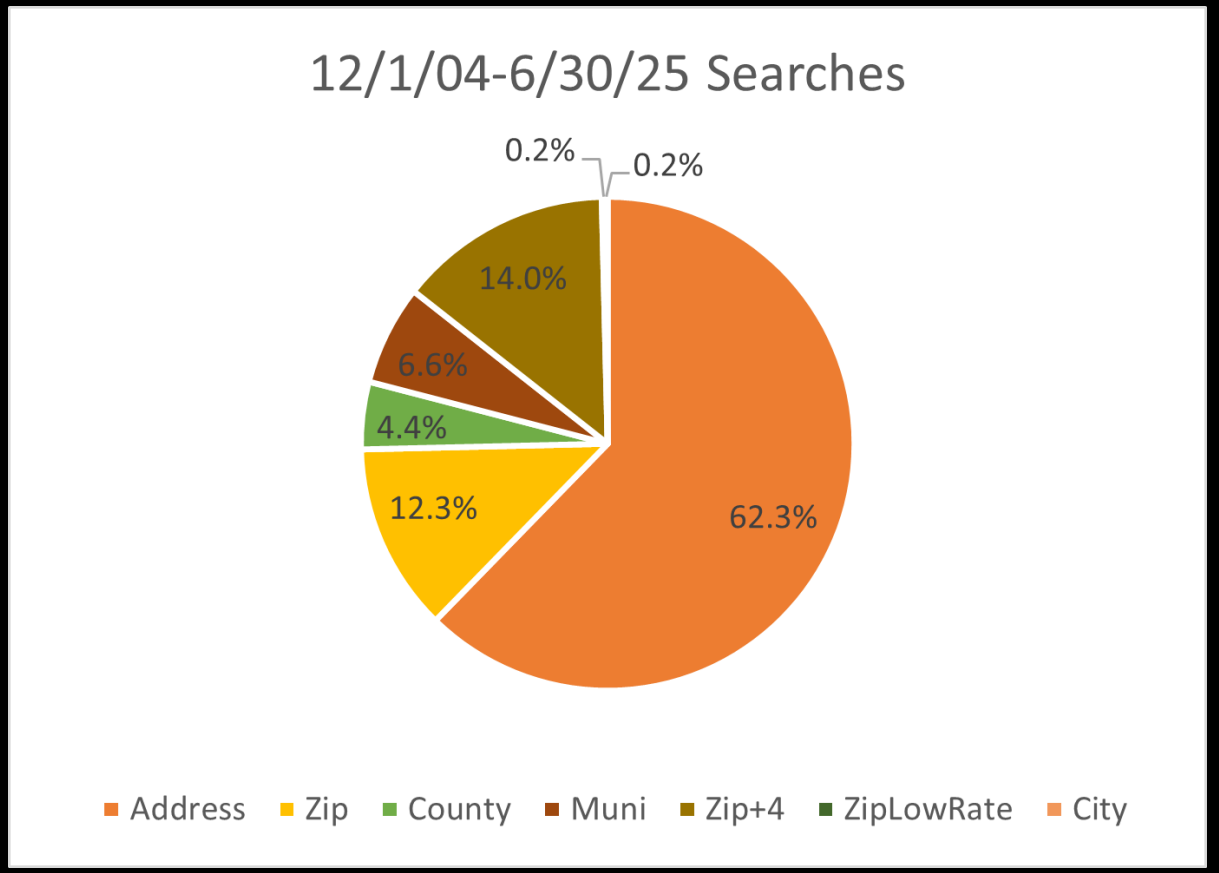
General Usage Stats for 2025 So Far

	1/1/25-9/1/25			
	Searches	Failed	% Failed	% of total searches
Address	114219	46416	40.6%	58.7%
Zip	27084	812	3.0%	13.9%
County	7651	0	0.0%	3.9%
Muni	10254	0	0.0%	5.3%
Zip+4	30986	5625	18.2%	15.9%
ZipLowRate	4404	0	0.0%	2.3%
Total	194598	52853	27.2%	

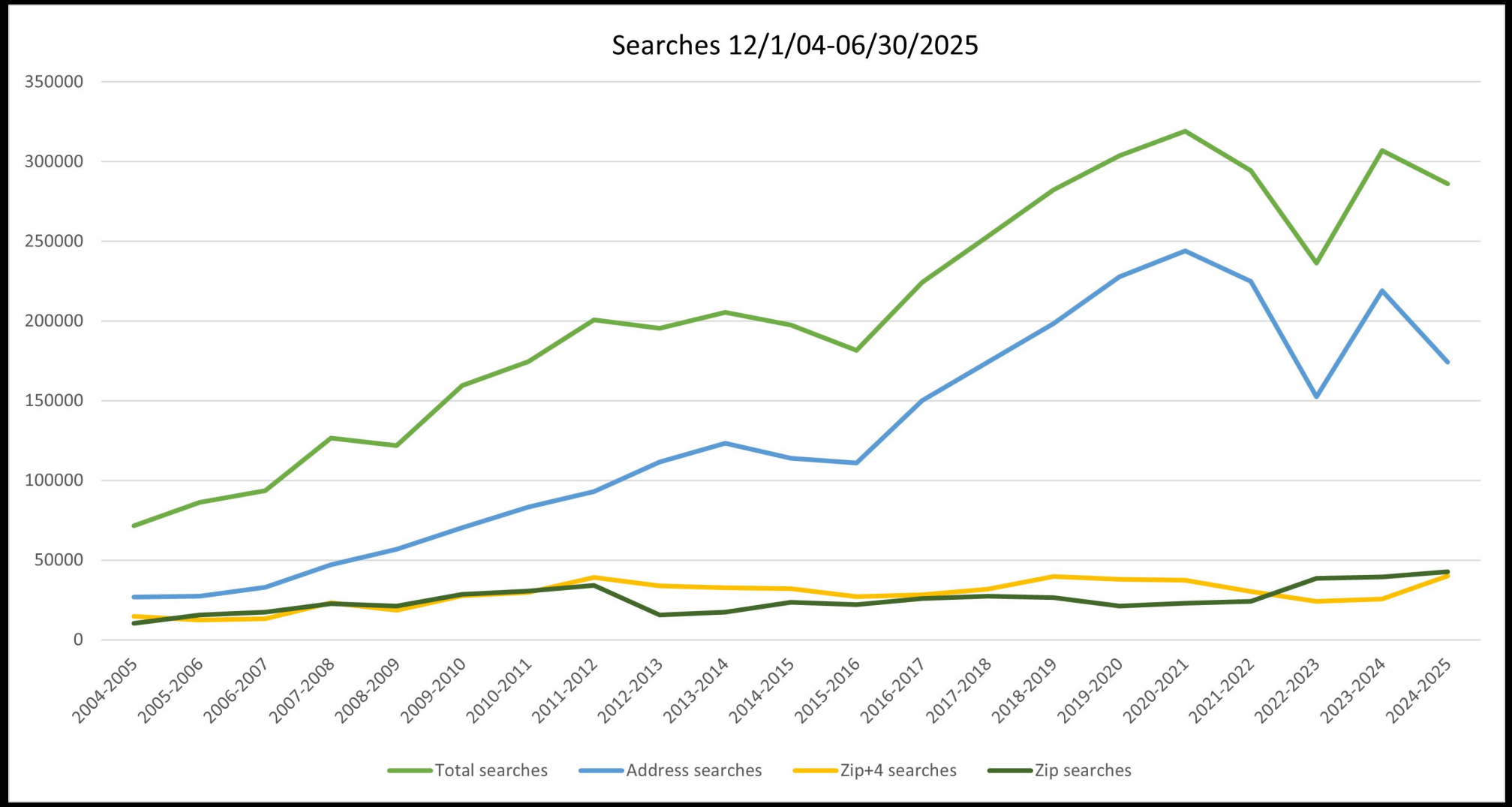


Historic usage data

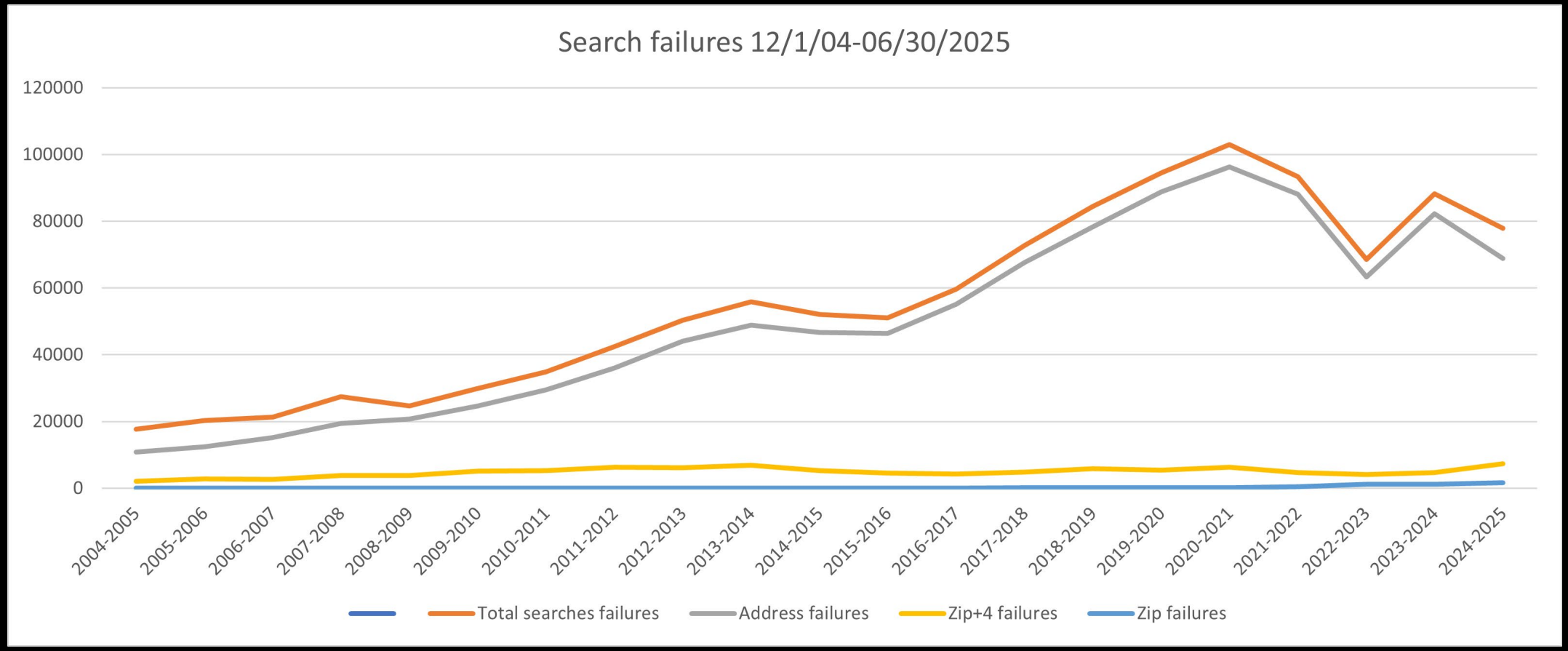
	12/1/04-6/30/25			
	Searches	Failed	% Failed	% of total searches
Address	2662416	1044005	83.5%	62.3%
Zip	527586	102760	8.2%	12.3%
County	188548	465	0.0%	4.4%
Muni	280343	670	0.1%	6.6%
Zip+4	598931	102760	8.2%	14.0%
ZipLowRate	8590	0	0.0%	0.2%
City	7735	11	0.0%	0.2%
Total	4274149	1250671		



Historic usage data



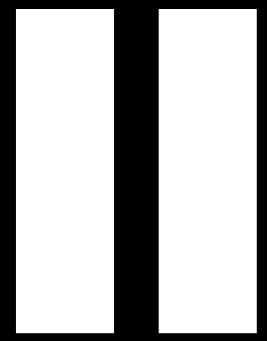
Historic usage data



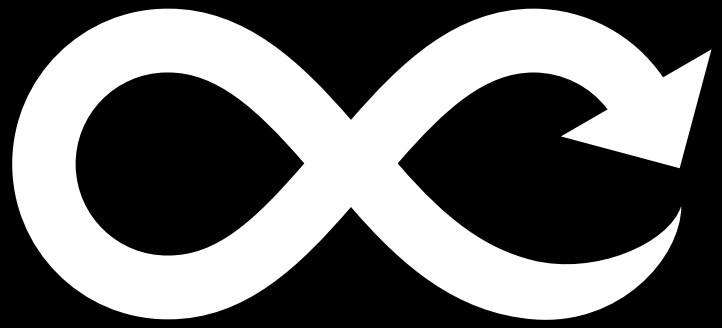
Top 5 Application Users in 2025

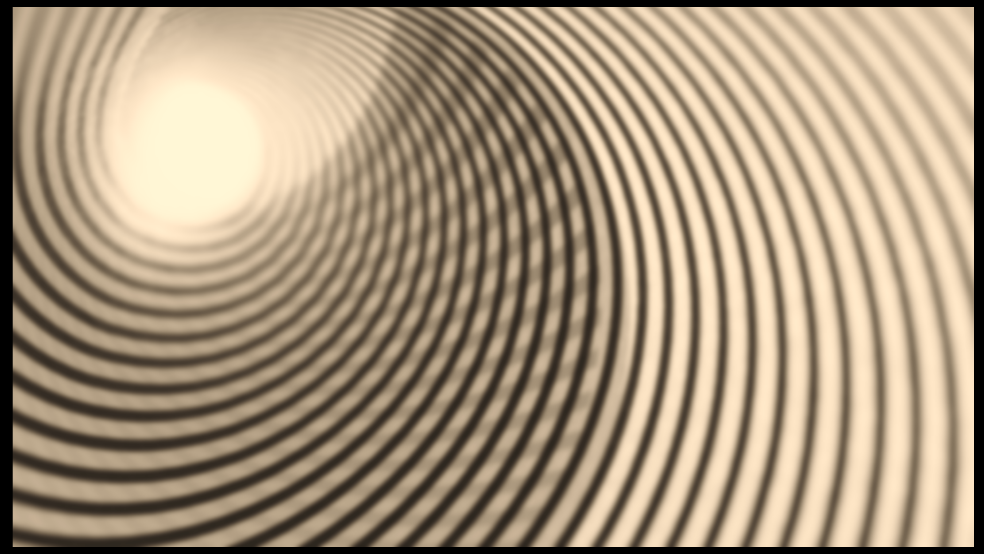
- WHOIS IP lookups for the top 5 **individual** IP addresses appearing in the web searches/hits data in 2025 (so far) are shown below. Assuming the IP addresses assigned to users are generally static (don't change on renewal), then we know we may have some particularly interested users.
- While a WHOIS lookup doesn't identify an individual user directly, it does identify the range of IP's from which the IP came – IP ranges per WHOIS report the entities responsible for the range in questions, and can be thought of as Users' internet service providers (ISP):
- **1** – 3,083 hits - ***Builders FirstSource, Inc.***
  - *One Builders FirstSource, Inc. user (by IP) has logged 3,083 searches in 2025. Thank you for your business.*
- **2** – 2,496 hits – ***Dobson Technologies***
- **3** – 2,266 hits – ***Mayfield Electric & Water Systems***
- **4** – 2,164 hits – ***AT&T***
- **5** – 2,046 hits - ***Cox Communications, LLC***

PAUSE



RESUME





What is a P-Value?


chance

is difficult to explain

france

How Might One Explain The Foundational Notions of

PROBABILITY

Marquis De Laplace

Pierre-Simon Laplace

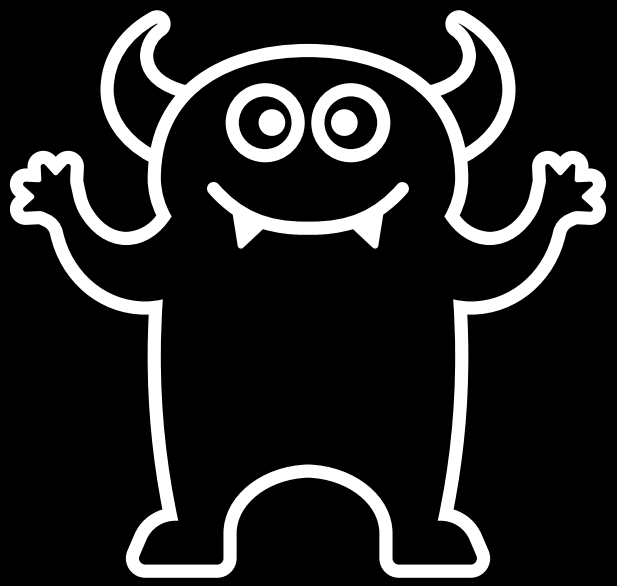
1749-1827

From the relatively safe perspective...

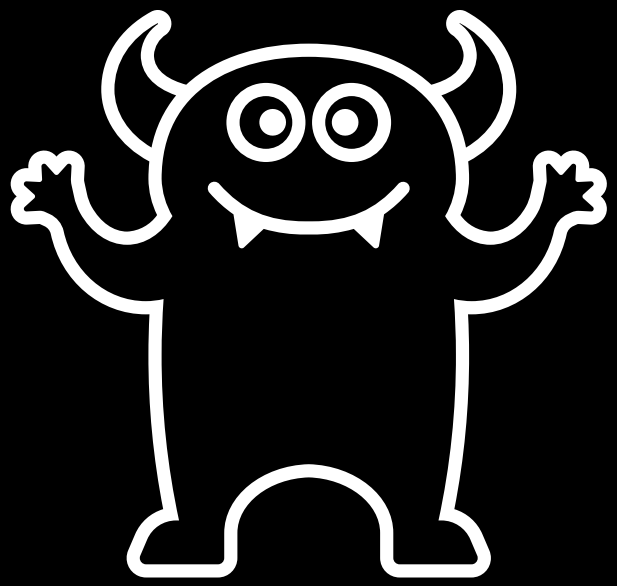
Of Causal Determinism.

Be Kind. Please Rewind.

Then **Now** **Next**

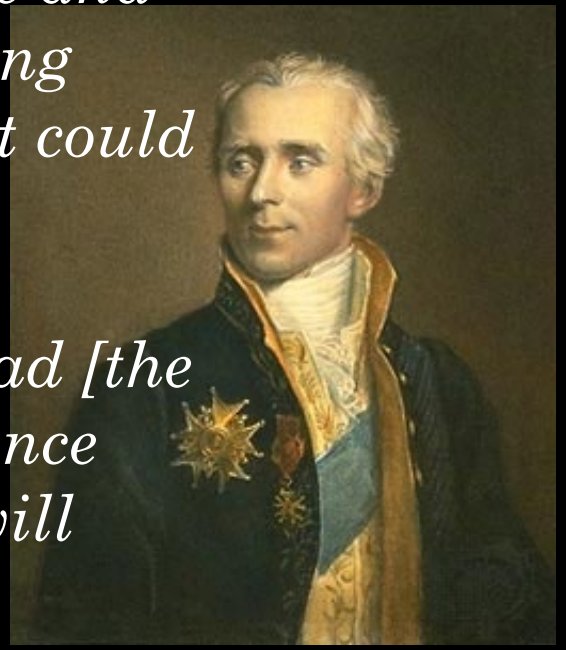


***Essai philosophique sur les probabilités* (1814):**
Une intelligence qui, pour un instant donné, connaîtrait toutes les forces dont la nature est animée, et la situation respective des êtres qui la composent, si d'ailleurs elle était assez vaste pour soumettre ces données à l'analyse, embrasserait dans la même formule les mouvements des plus grands corps de l'univers et ceux du plus léger atome : rien ne serait incertain pour elle, et l'avenir, comme le passé, serait présent à ses yeux.



We may regard the present state of the universe as the effect of its past and the cause of its future. An intellect which at a certain moment would know all forces that set nature in motion, and all positions of all items of which nature is composed, if this intellect were also vast enough to submit these data to analysis, it would embrace in a single formula the movements of the greatest bodies of the universe and those of the tiniest atom; for such an intellect nothing would be uncertain and the future just like the past could be present before its eyes.

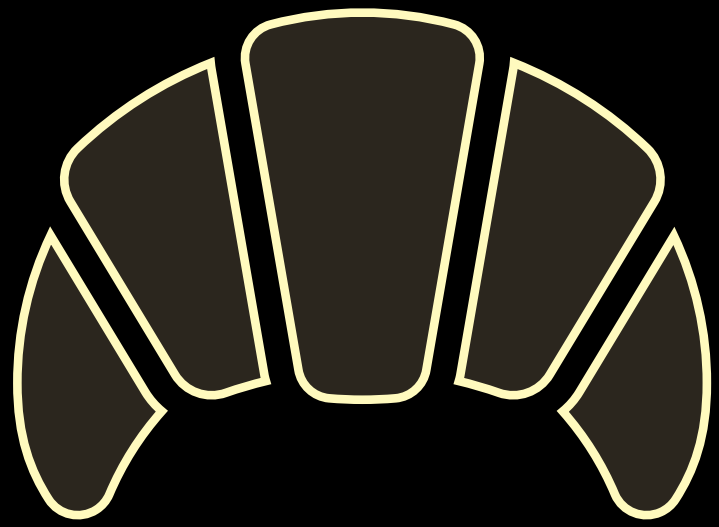
…All these efforts in the search for truth tend to lead [the human mind] back continually to the vast intelligence which we have just mentioned, but from which it will always remain infinitely removed…



- ***Obviously,*** *we can't reverse information loss, we can't un-fry an egg.*
- ***Theoretically,*** *the Demon can't predict it's own future, if the future is indeterminate, and the demon hasn't lived it yet.*
- ***Ostensibly,*** *One Demon couldn't predict for another Demon.*
- ***Potentially,*** *we live in outright chaos.*

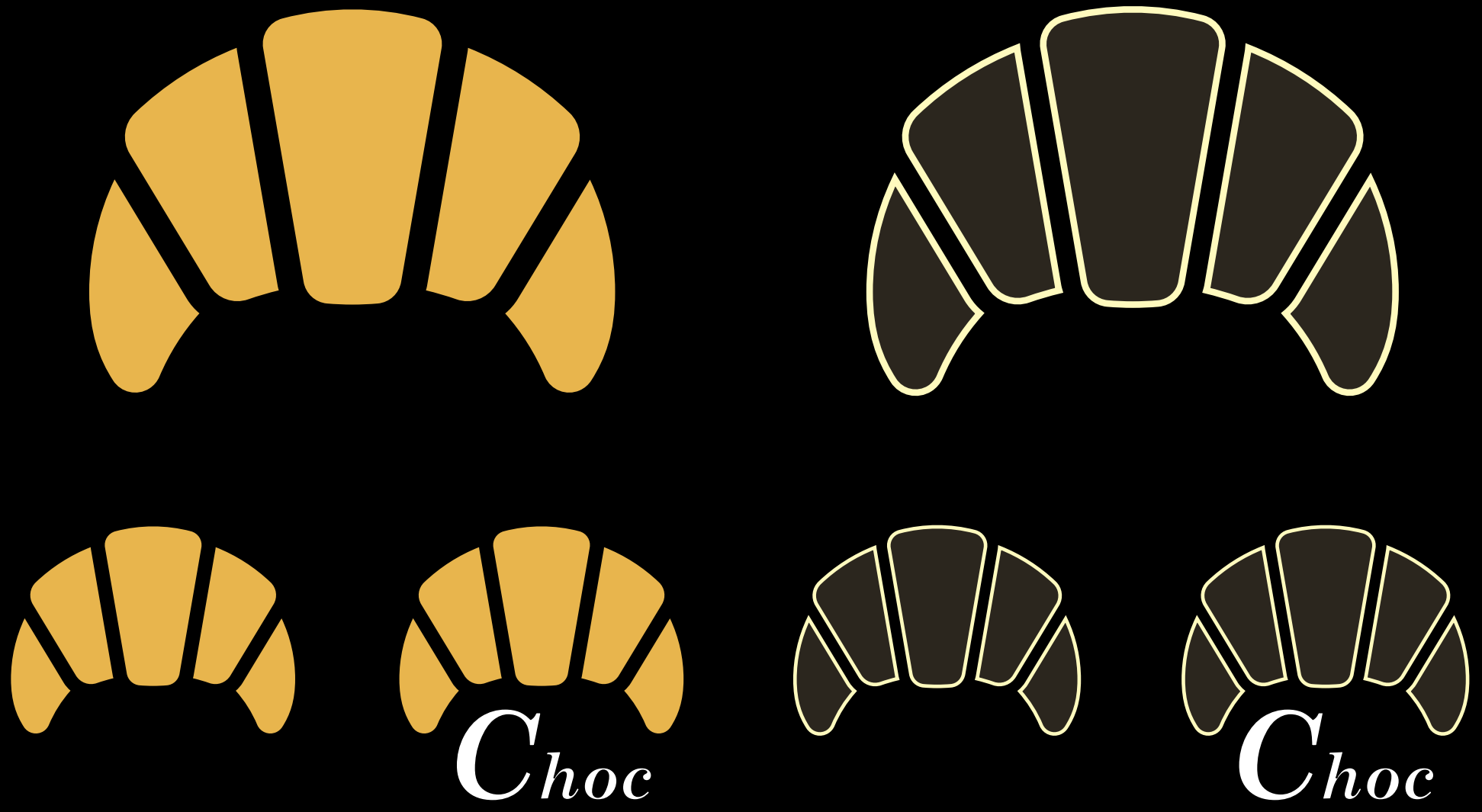
MARCEL

DES CROISSANTS



IL A LES BRÛLÉS TROP DES FOIS

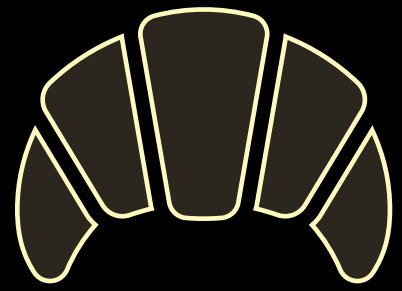
THE IDEA OF A RANDOM VARIABLE

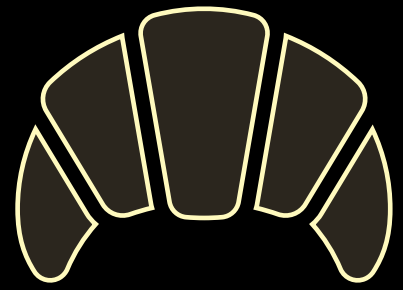
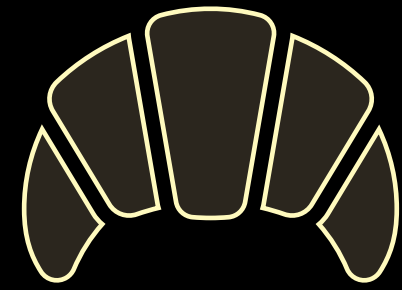


IL N'EST PAS UN BON BOULANGER

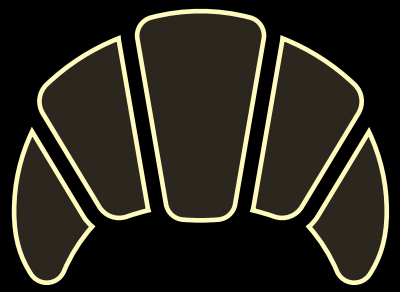
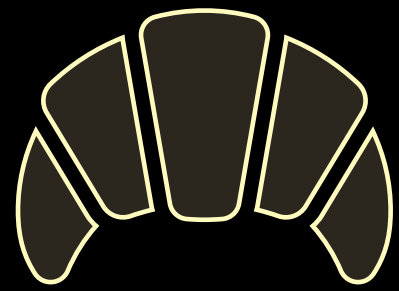
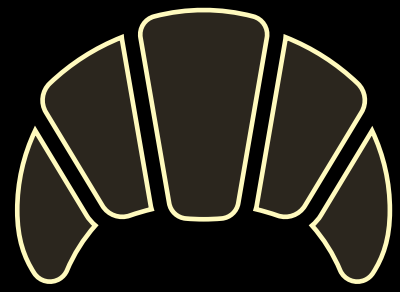
	 Burn		Total
 Choc	10	2	12
	30	58	88
Total	40	60	100

COUNT OF OBSERVATIONS FOR CROISSANT-EVENTS

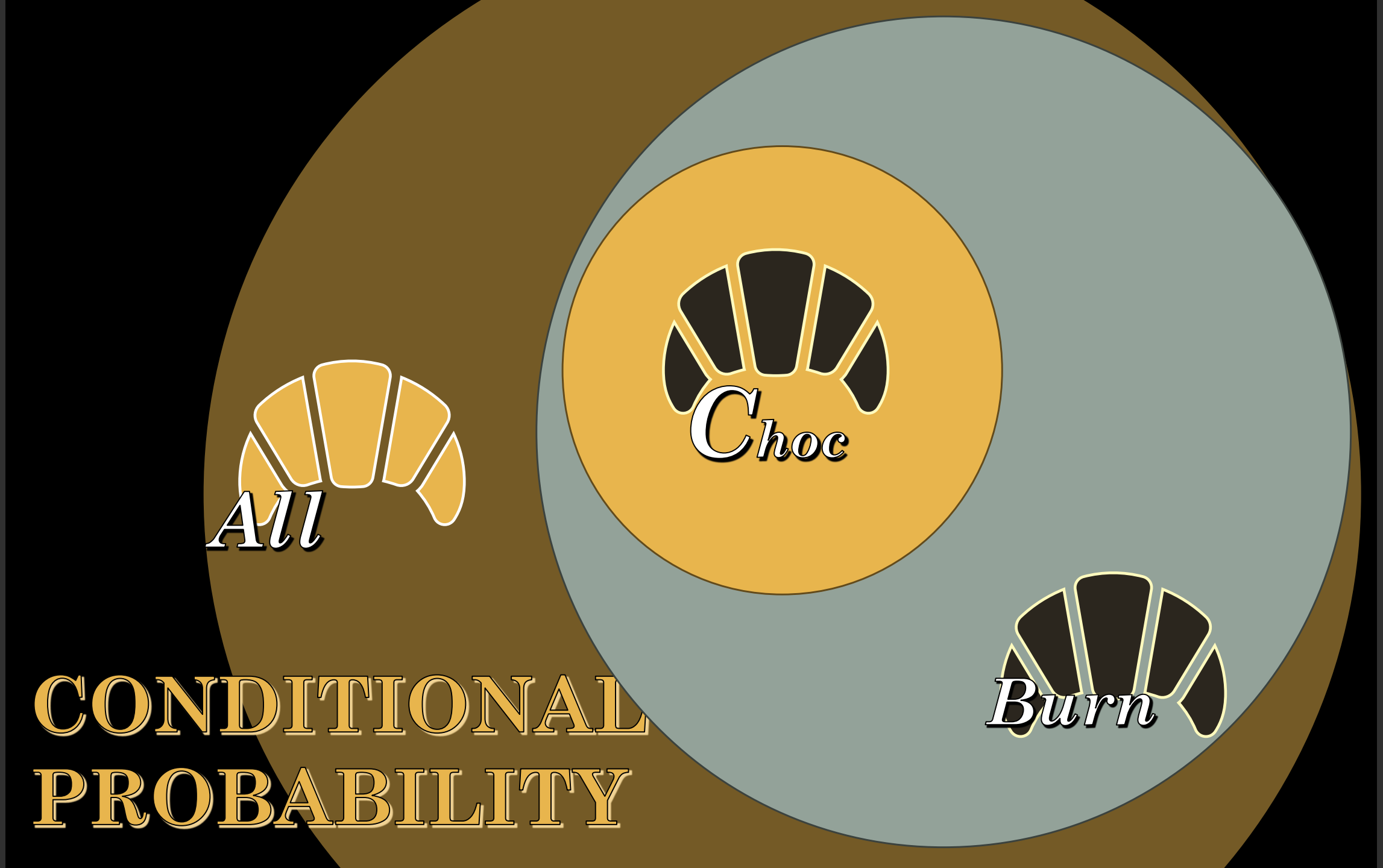
$C_{hoc} \cap$  $=$

$C_{hoc} \mid$  $\cdot$ 

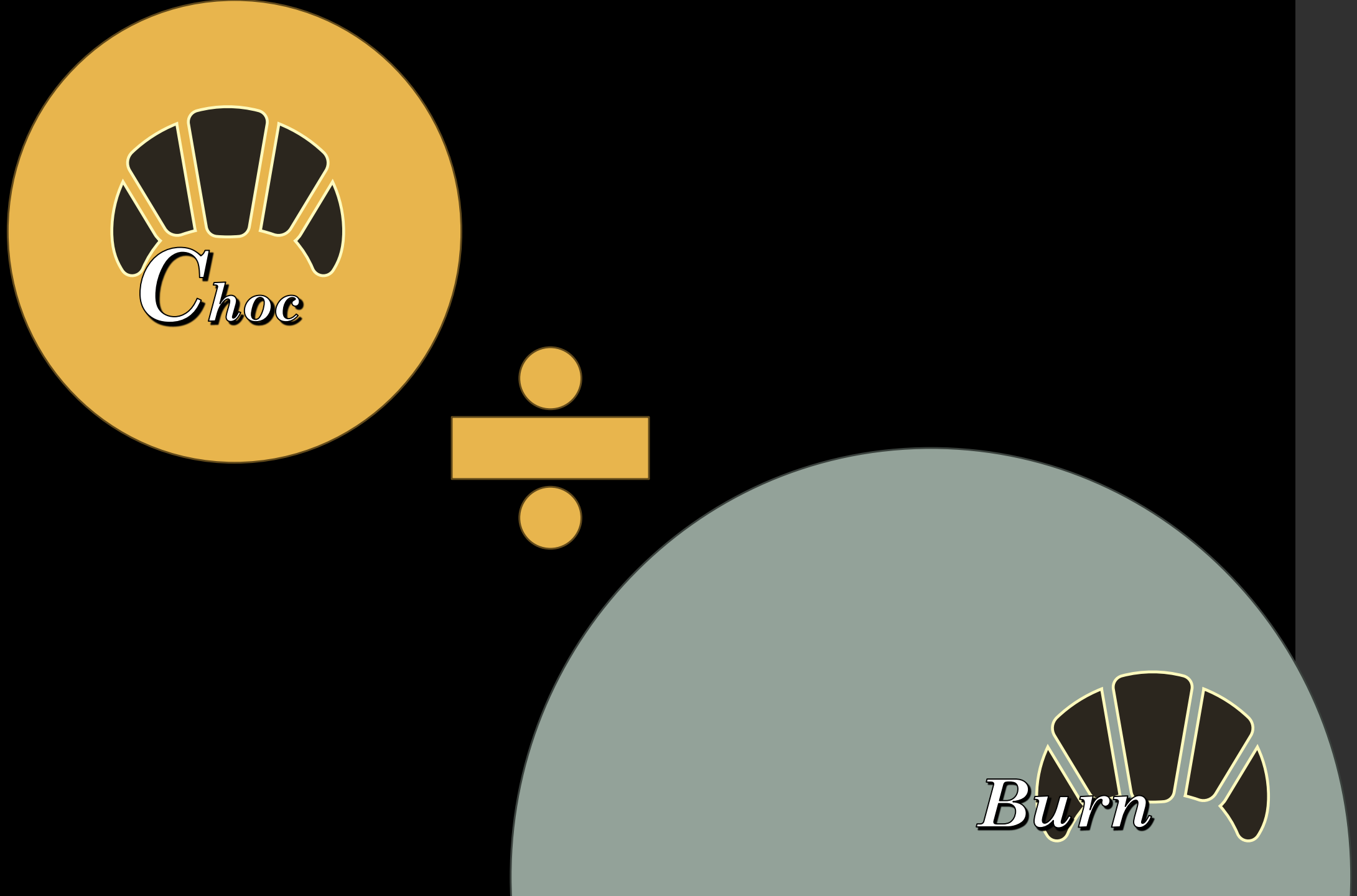
CALCULATION OF JOINT PROBABILITY

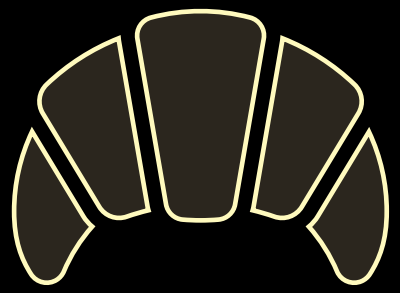
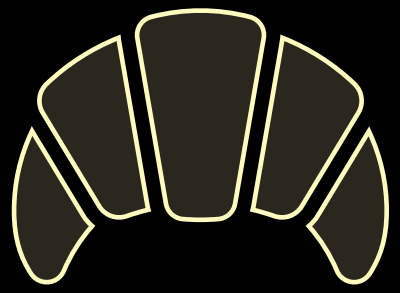
$$C_{hoc} \mid \text{  } = \frac{C_{hoc} \cap \text{  }}{\text{  }}$$

CONDITIONAL PROBABILITY



CONDITIONAL PROBABILITY


Choc
÷
Burn

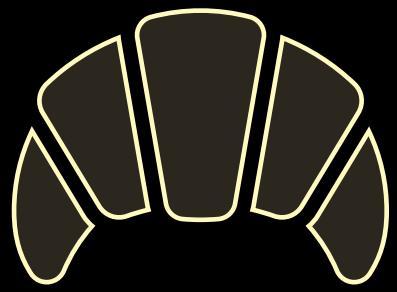
$$C_{choc} \mid \text{  } = \frac{0.1 \quad \text{  }}{0.4 \quad \text{  }}$$

CONDITIONAL PROBABILITY

0.1

0.25

2.5 x WORSE!

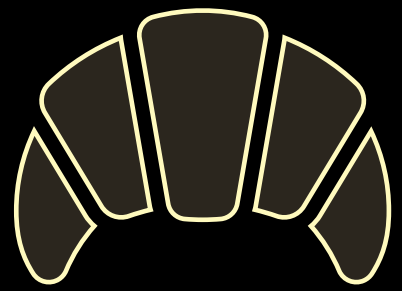


CONDITIONAL PROBABILITY

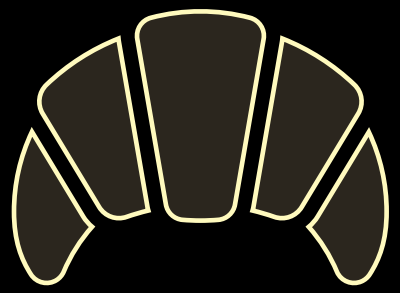
			T_{otal}
	0.1	0.02	0.12
	0.3	0.58	0.88
T_{otal}	0.4	0.6	1.0

PROBABILITY MODEL FOR CROISSANT-EVENTS

L_{Choc}



LIKELIHOOD

L  $\mid C_{choc} = 0.25$

LIKELIHOOD

...is not a probability, special case here

	Burn		Total
Choc	0.1	0.02	0.12
	0.3	0.58	0.88
Total	0.4	0.6	1.0

JOINT PROBABILITY OF CROISSANTS

	Burn		Total
Choc	0.1	0.02	0.12
	0.3	0.58	0.88
Total	0.4	0.6	1.0

MARGINAL PROBABILITY OF CROISSANTS

	Burn 		Total
Choc 	0.1	0.02	0.12
	0.3	0.58	0.88
Total	0.4	0.6	1.0

LAW OF TOTAL PROBABILITY


belief

$$P(Burn|Choc) = \frac{P(Burn) \cdot L(Burn|Choc)}{P(Choc)}$$

$$P(Burn|Choc) \propto P(Burn) \cdot L(Burn|Choc)$$

BAYES

$$0.833 > 0.4$$

$$P(Burn|Choc) = \frac{0.4 \cdot 0.25}{0.12} = 0.833$$

He's Probably Gonna Burn It.

BAYES

$$Y|\theta \sim Bin(n, \theta)$$

$$Y \equiv data$$

$$\theta \sim Beta(\alpha, \beta)$$

$$\theta = parameter\ for\ a\ binomial\ dist$$

BONUS: CROISSANTS ON A CONTINUUM

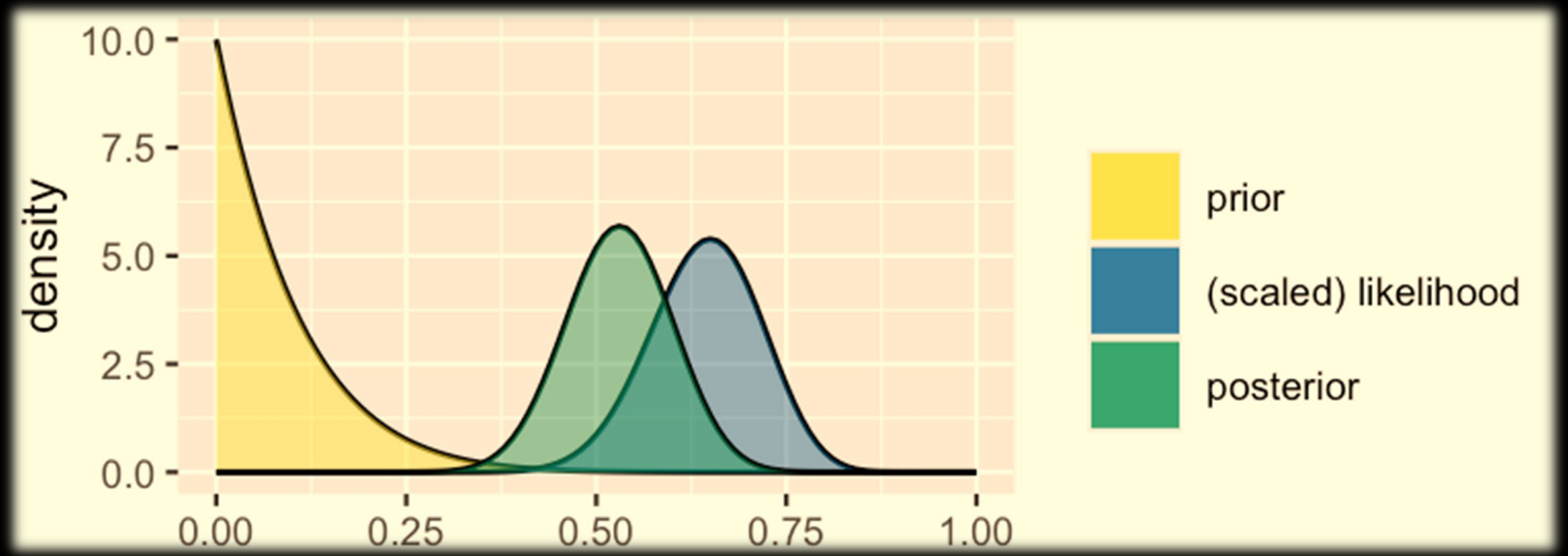
$$Y|\theta \sim Bin(n, \theta) \qquad Y|\theta \sim Bin(40, \theta)$$
$$\theta \sim Beta(a, \beta) \qquad \theta \sim Beta(1, 10)$$

$$\theta|\,(Y = y) \sim Beta(a + y\,, \beta + n\, - y)$$

$$\theta|\,Y = 26 \sim Beta(27{,}24)$$

26/40 Burned!

BONUS: [illegible]SSANTS ON A CONTINUUM



He's getting worse at this...

https://www.bayesrulesbook.com/chapter-3#ch3-bbmodel

The Posterior Distribution is proportional to the Prior Distribution times the Likelihood of having observed your new data.

In all but the most basic models, or those for which conjugate priors are available, numeric solutions for "the evidence" (the marginal in the denominator) ***become necessary.***

This strategy operationalizes scientific thought. It provides an accounting for all sources of knowledge we might configure it to consider; it then returns to us a juxtaposition of what we have seen, and what we once thought we knew.

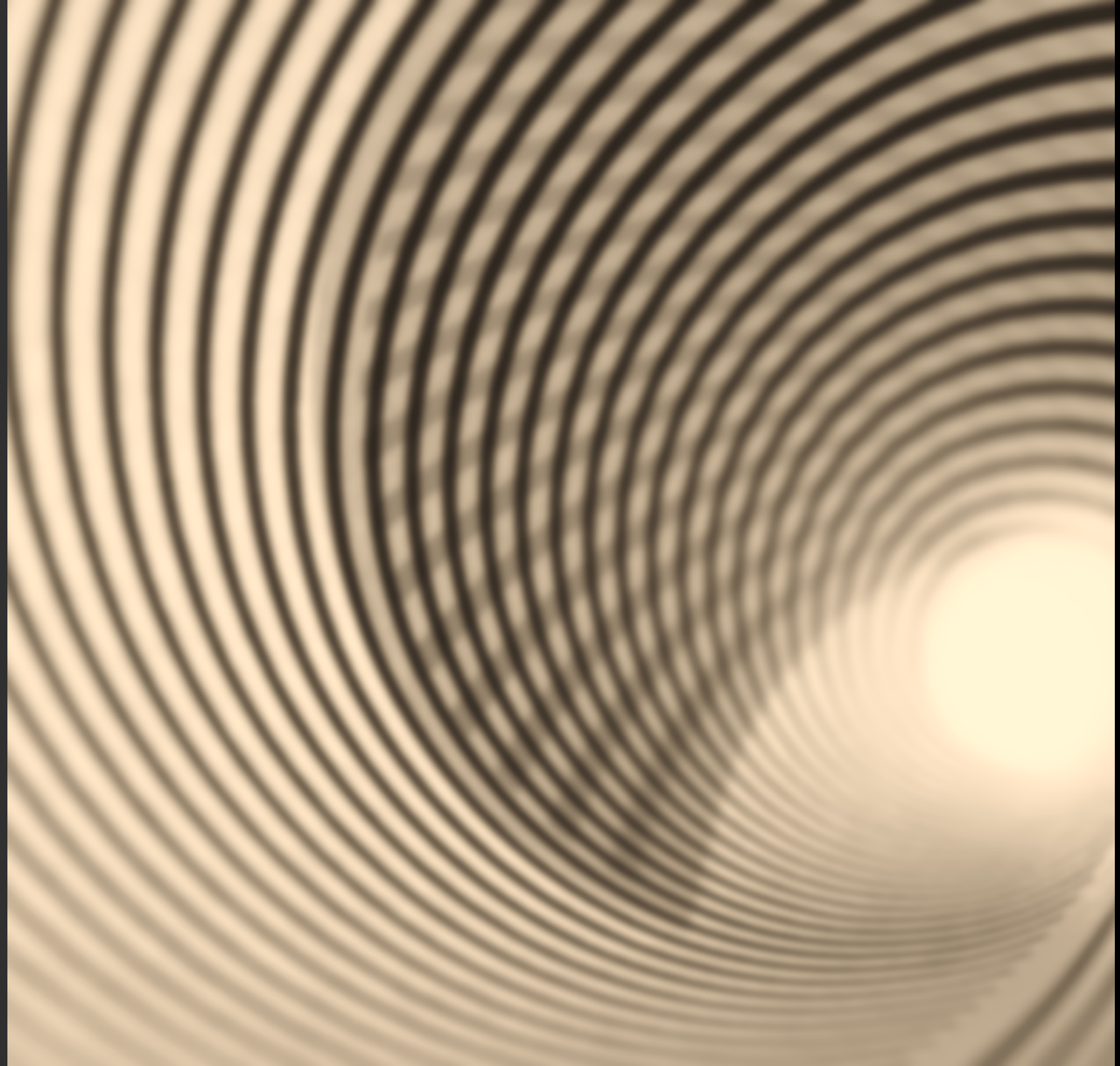
BAYES


complexity

REALITY

ABSTRACTION

- ***The mind,*** *like the sense of sight, has its illusions; and just as touch corrects those of the latter, so thought and calculation correct the former. (Laplace, 1825).*

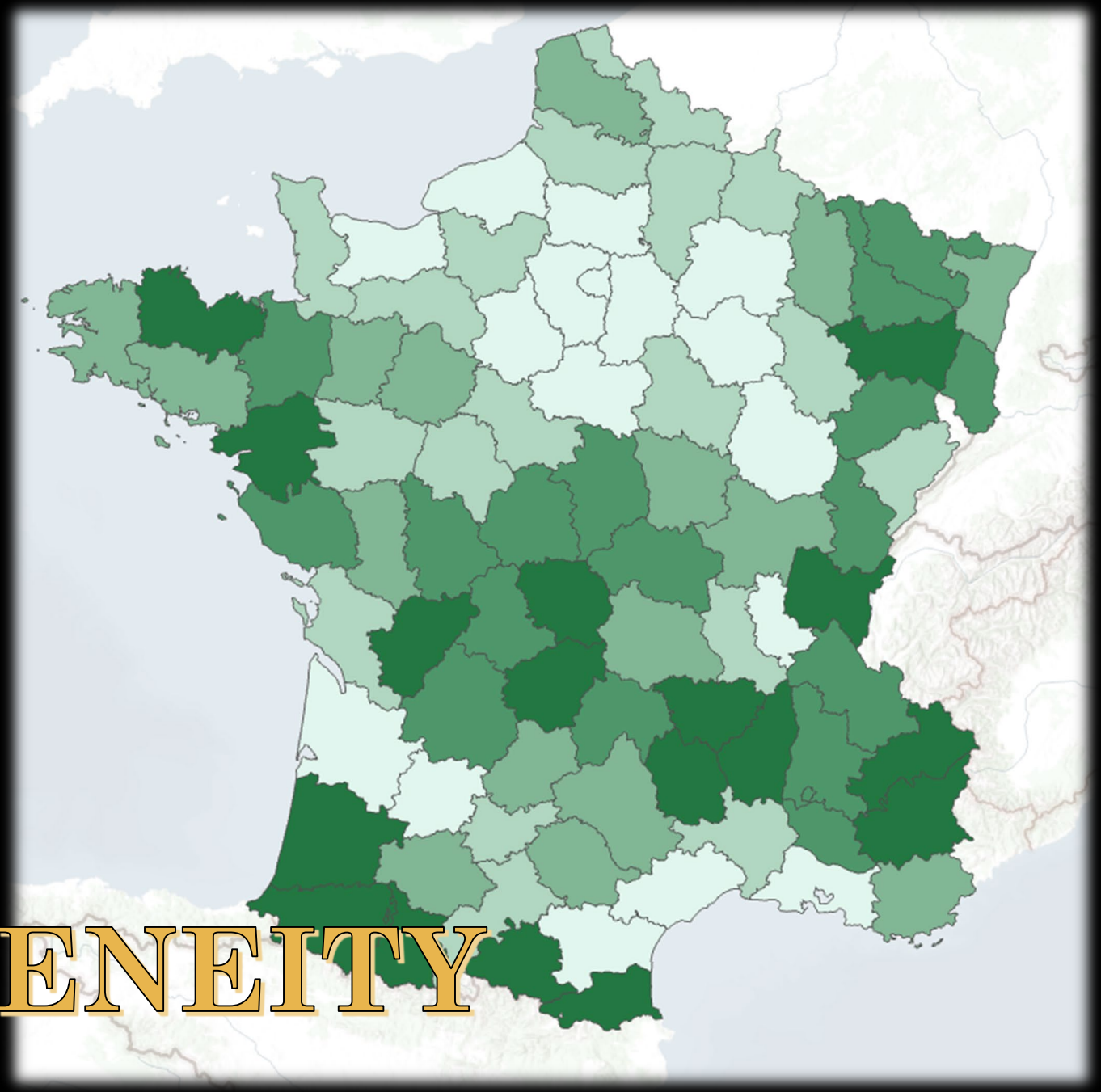


SPACE

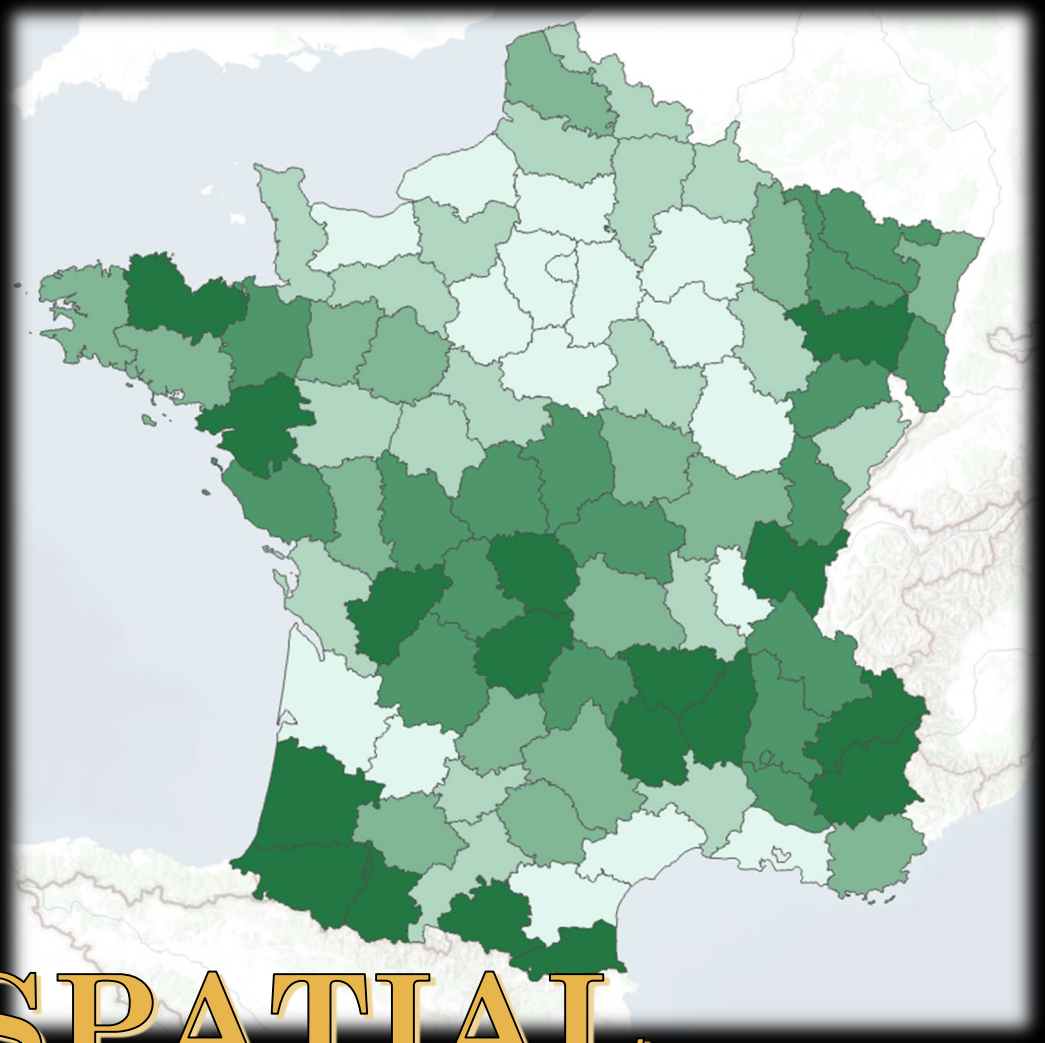
SPATIAL HETEROGENEITY

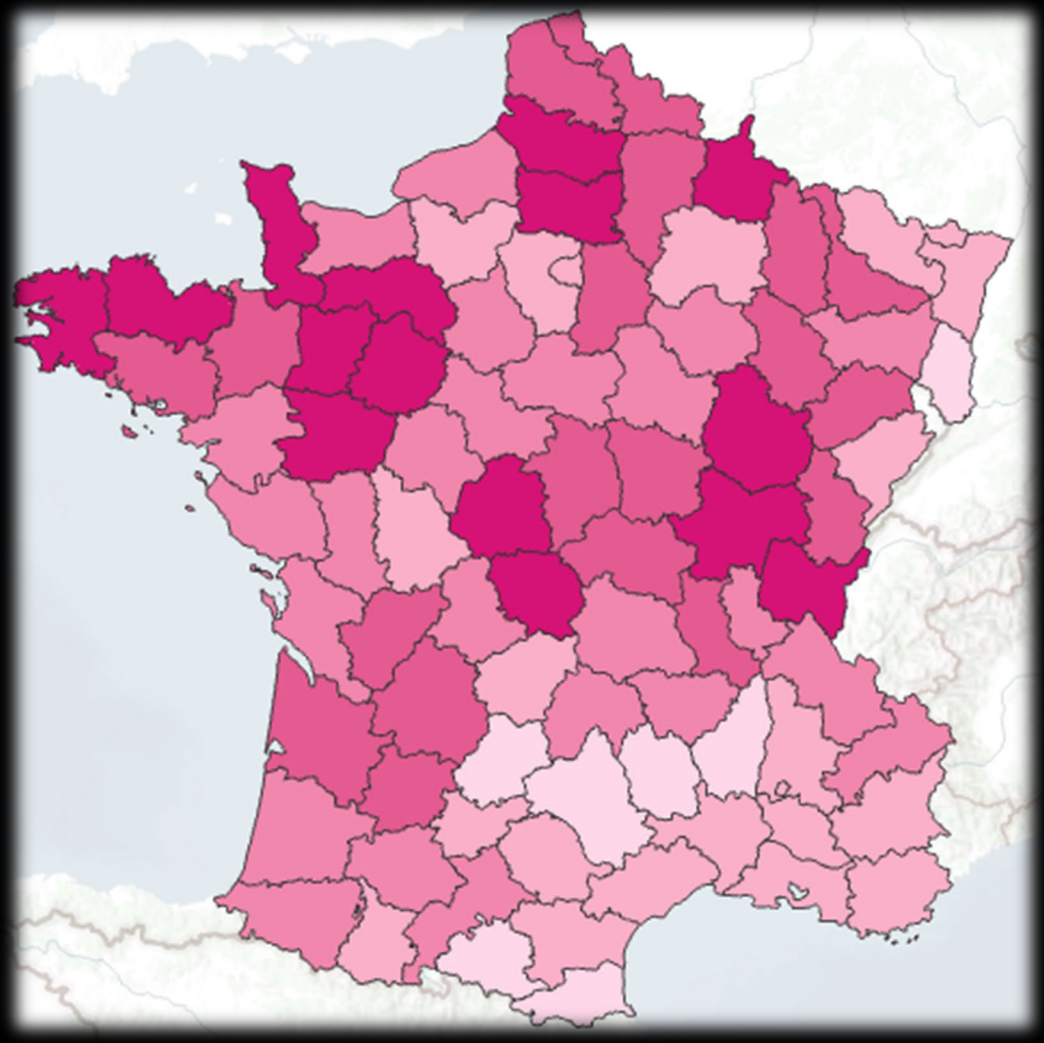
Per Capita Personal Property Tax Collected, 1833.

(Friendly, 2007).



SPATIAL HETEROGENEITY

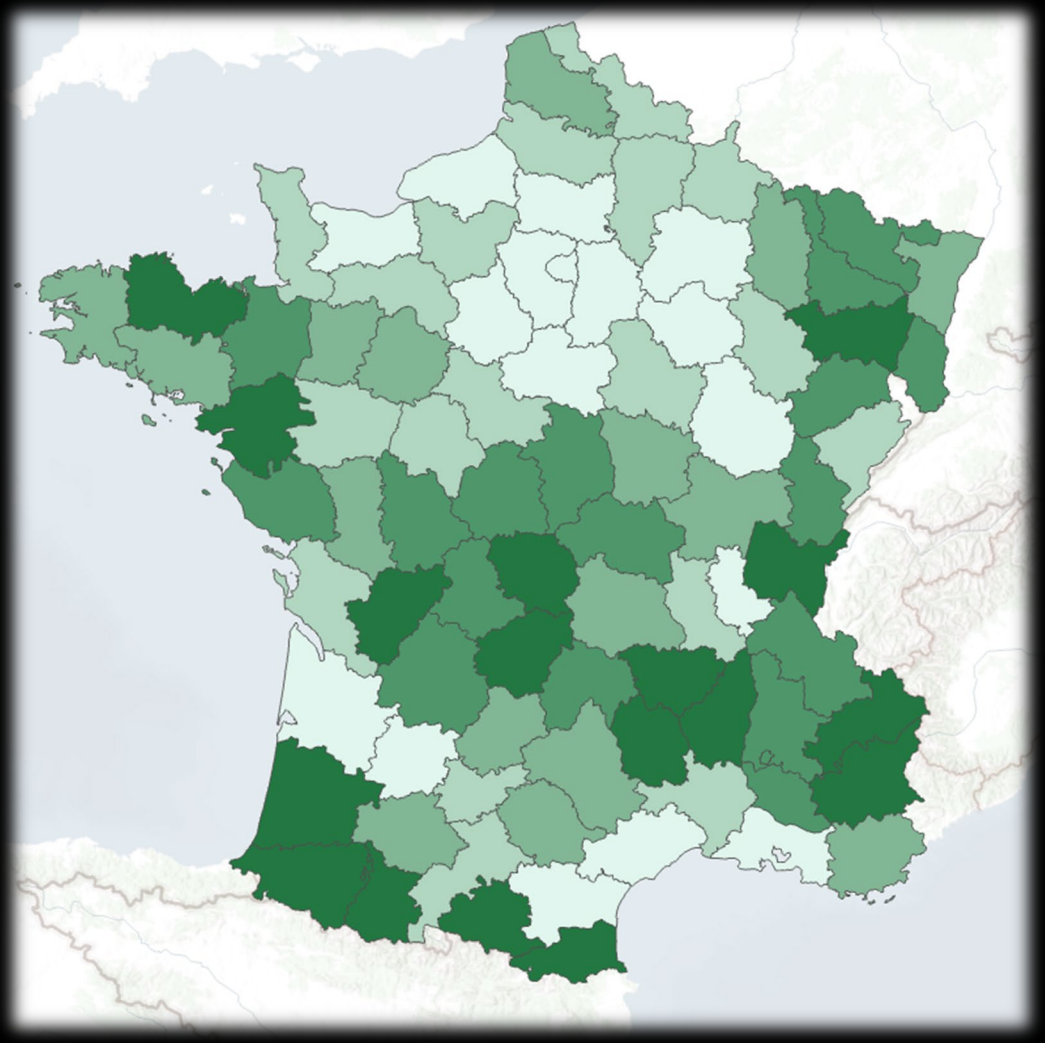


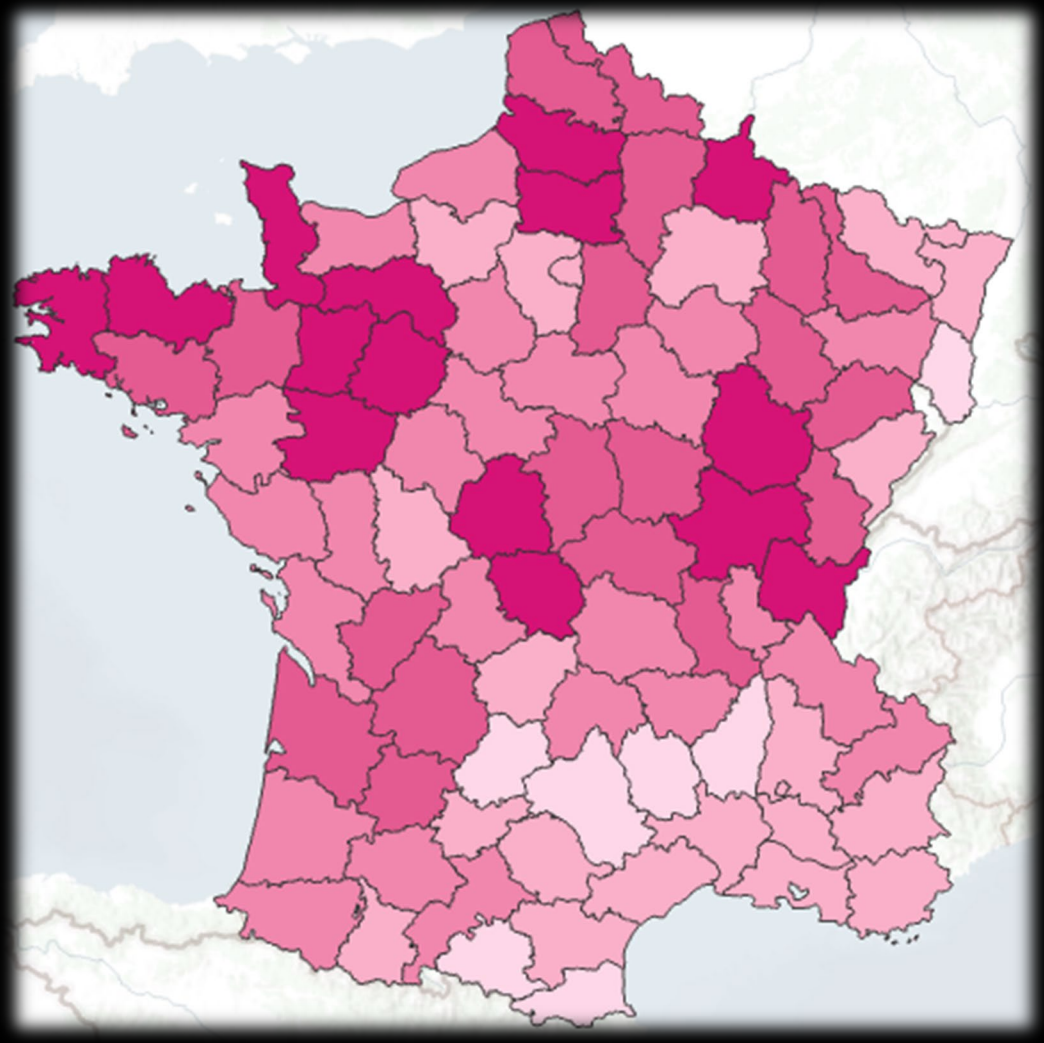


SPATIAL DEPENDENCE

Per Capita Crime Against Persons Collected, 1833.

(Friendly, 2007).

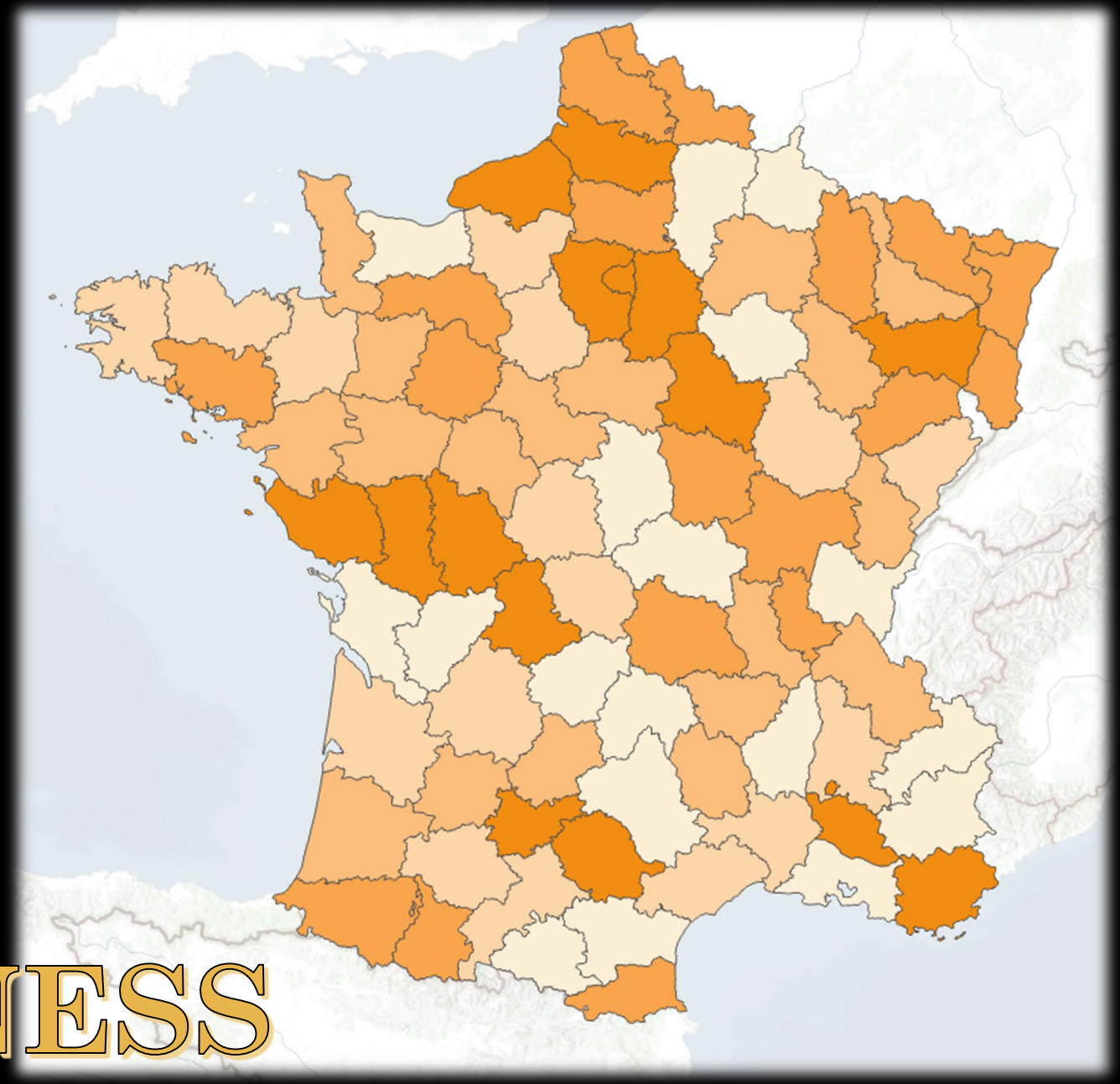




Qu'ils mangent de la brioche, n'est pas?

Random Value from 1-100, Generated, Last Week.

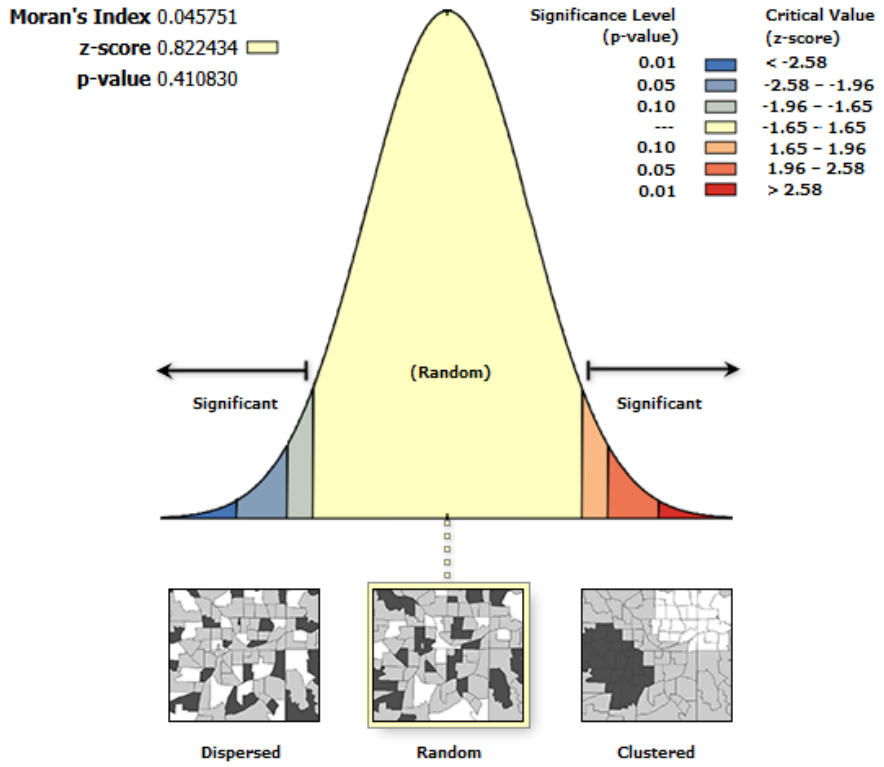
(Jim, 2025).



SPATIAL RANDOMNESS

SPATIAL RANDOMNESS

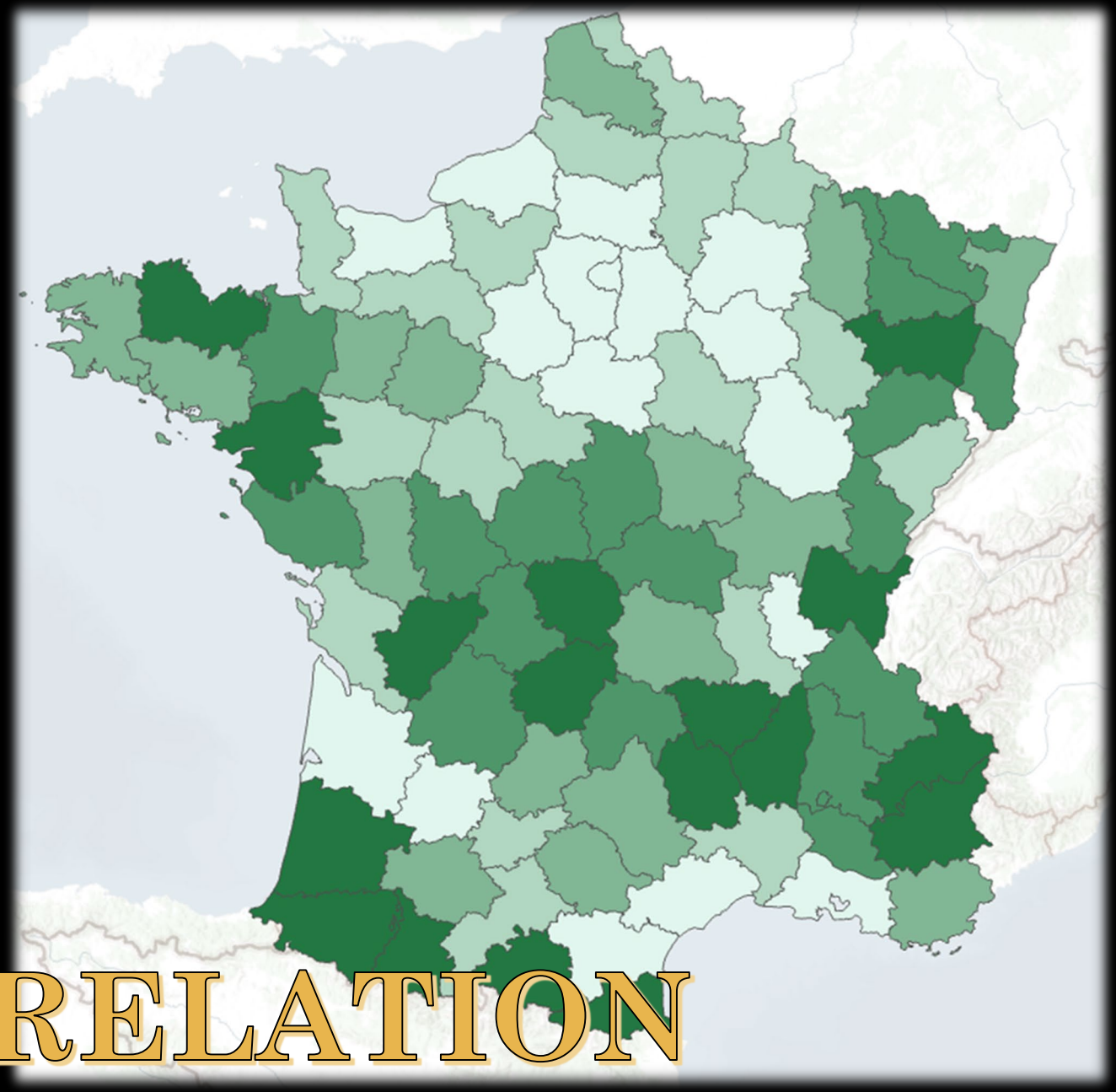
Spatial Autocorrelation Report



Given the z-score of 0.822434, the pattern does not appear to be significantly different than random.

Global Moran's I Summary

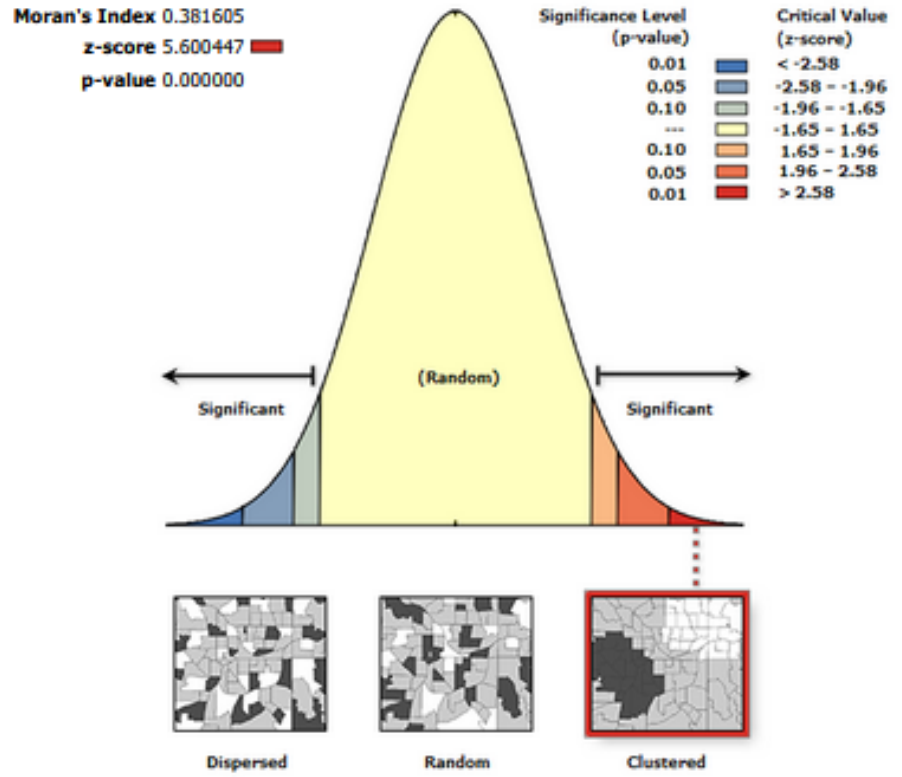
Moran's Index	0.045751
Expected Index	-0.011905
Variance	0.004915
z-score	0.822434
p-value	0.410830



SPATIAL AUTOCORRELATION

SPATIAL AUTOCORRELATION

Spatial Autocorrelation Report

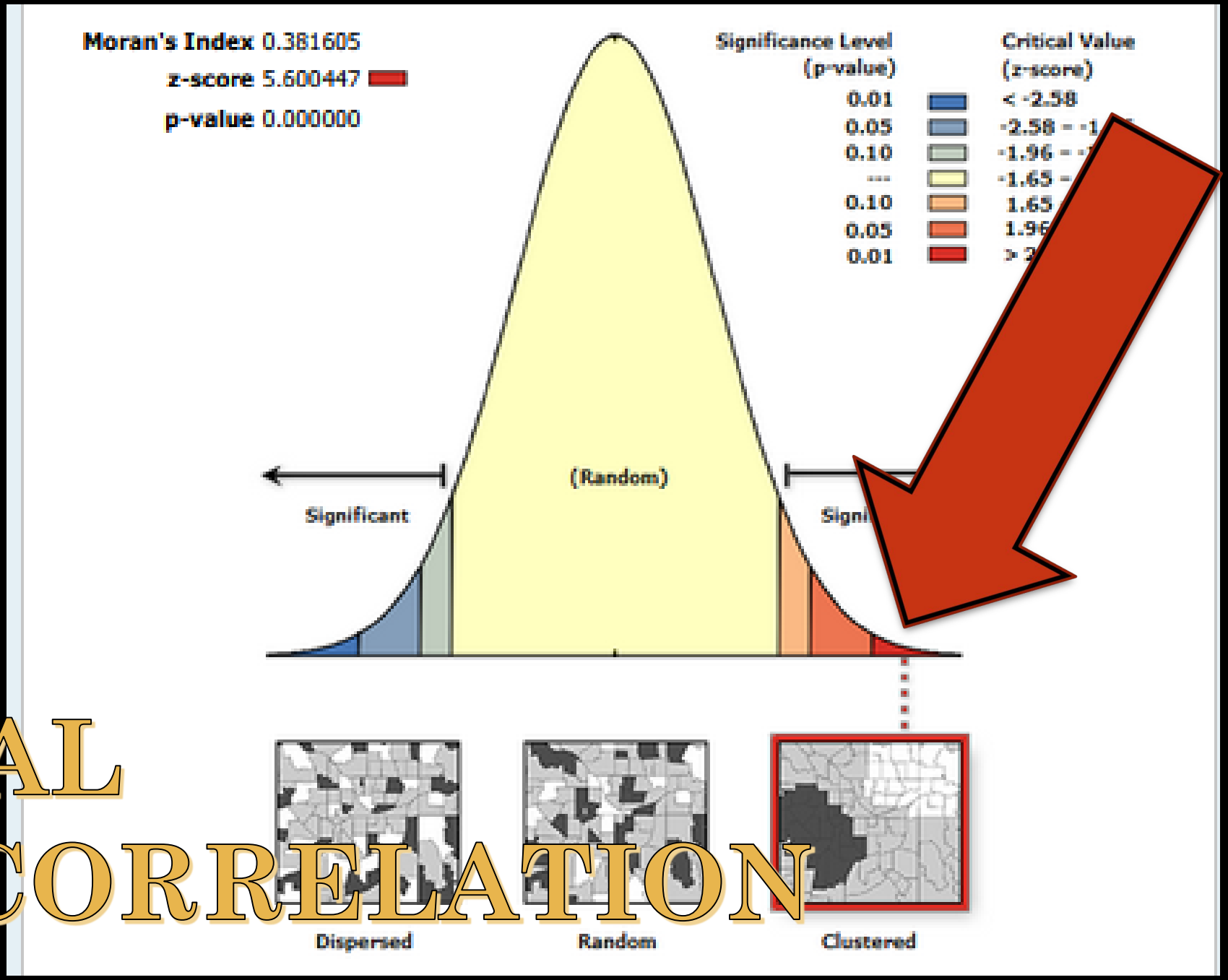


Given the z-score of 5.600447, there is a less than 1% likelihood that this clustered pattern could be the result of random chance.

Global Moran's I Summary

Moran's Index	0.381605
Expected Index	-0.011905
Variance	0.004937
z-score	5.600447
p-value	0.000000

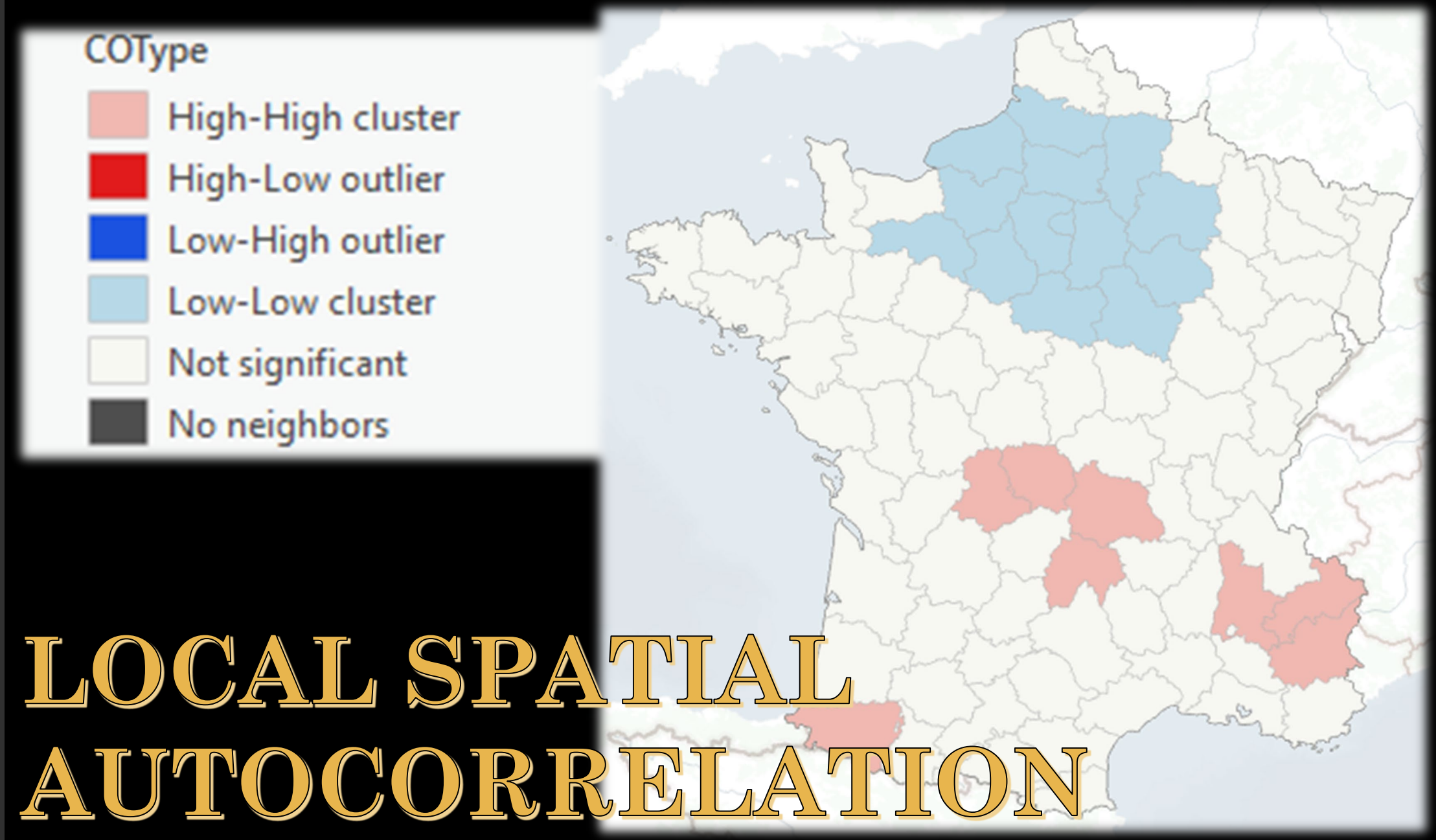
SPATIAL AUTOCORRELATION



$$I = \frac{\sum_j w_{ij} z_i z_j / S_0}{\sum_i z_i^2 / n}$$

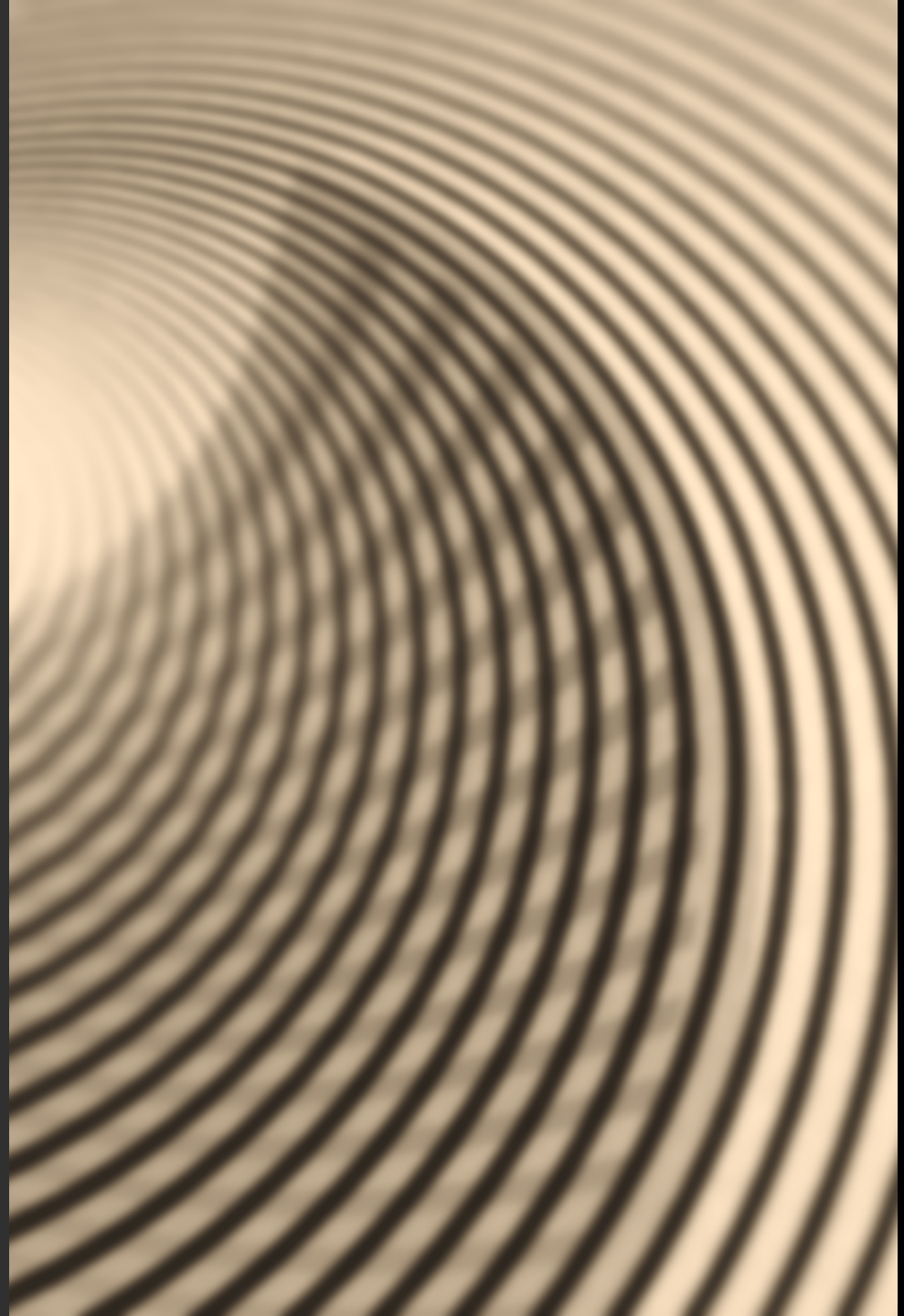
$$S_0 = \sum_i \sum_j w_{ij} \qquad z_i = x_i - \bar{x}$$

(GLOBAL MORAN'S) SPATIAL AUTOCORRELATION

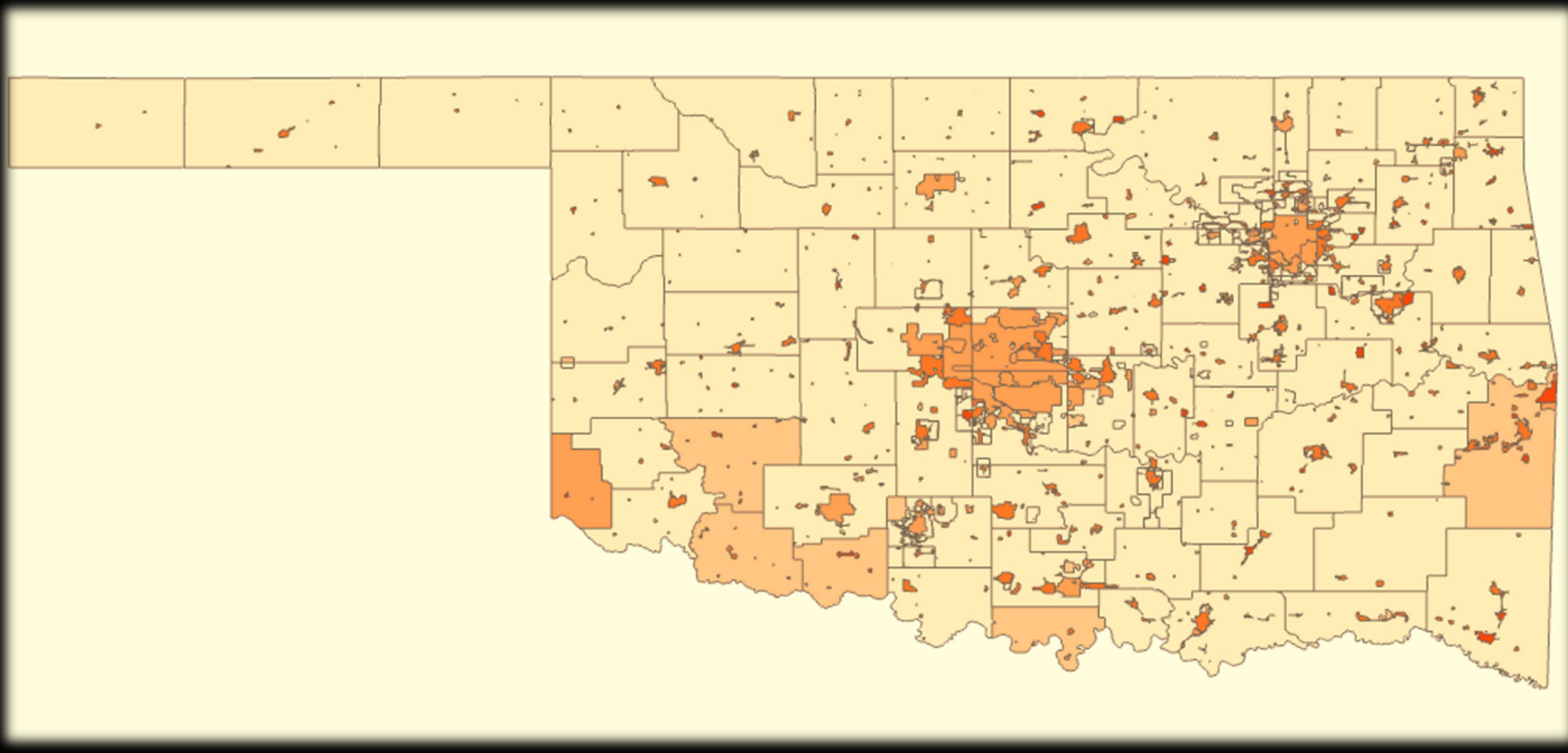


LOCAL SPATIAL AUTOCORRELATION

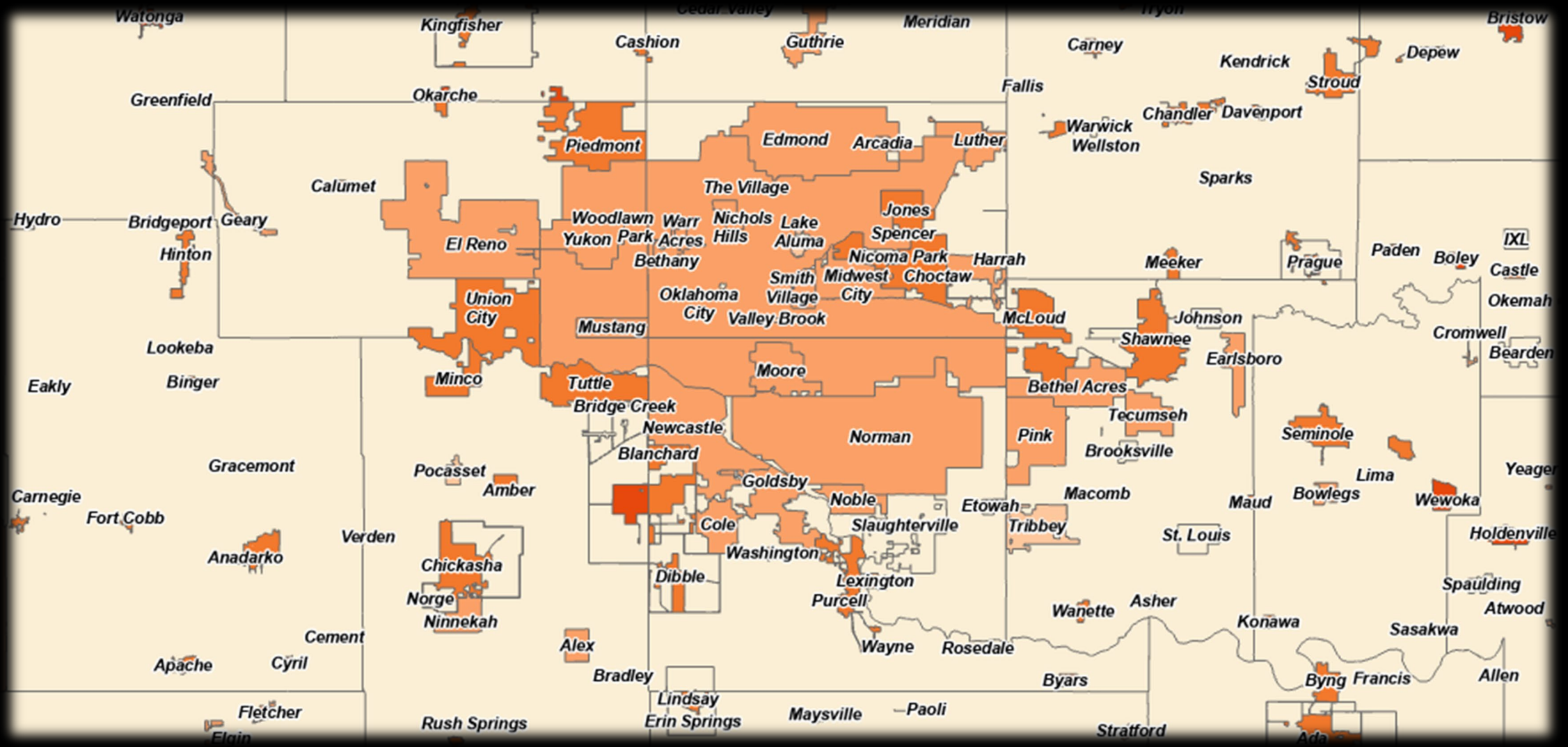
C’EST SUPER.



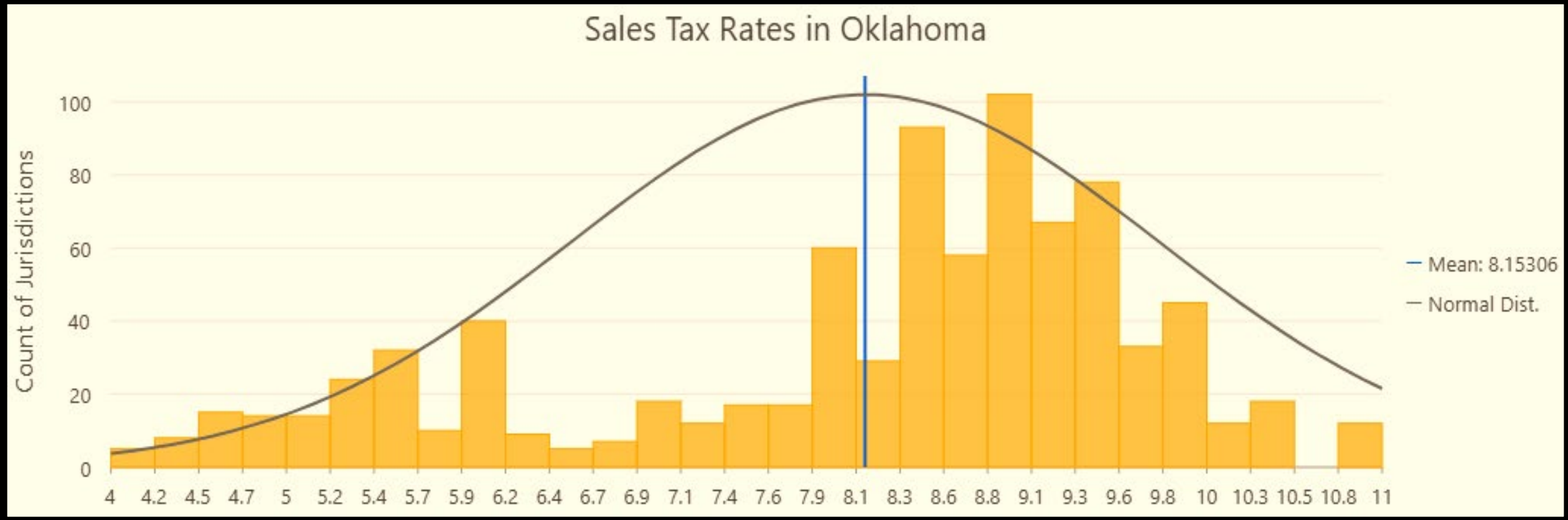
TAXES



SALES TAX JURISDICTIONS



SALES TAX JURISDICTIONS



SALES TAX RATE DISTRIBUTION

Are Sales Tax Rates Clustering in Oklahoma?

- How can we account for both *attribute* and *spatial* similarity, simultaneously?
- How can we classify extreme values *within a local neighborhood?*
- While complex and latent social influences result in the setting of tax rates, will these observations exhibit any informative pattern over space?
- Should I drive to a different Walmart? Can I save 2%? Is that worth the Drive?
- Is that ethical? Do I care?

MATH

Local Moran's I

- $I_i = \frac{\sum_j w_{ij} z_i z_j}{\sum_i z_i^2}$

- Where:
- I_i is the Local Moran's I statistic result, indicating positive spatial autocorrelation if I is positive, and vice versa.
- w_{ij} is the weights matrix element suggesting the influence on the variable of interest between geographies i and j.
- $z_i z_j$ is the product of deviations from the sample mean at locations i and j, expanding to $(x_i - \bar{x})(x_j - \bar{x})$
- w_{ij} is the weights matrix element suggesting the influence on the variable of interest between geographies i and j.
- x_j is the value of the variable of interest at location j.

Weights and Inference

- W_{ij}

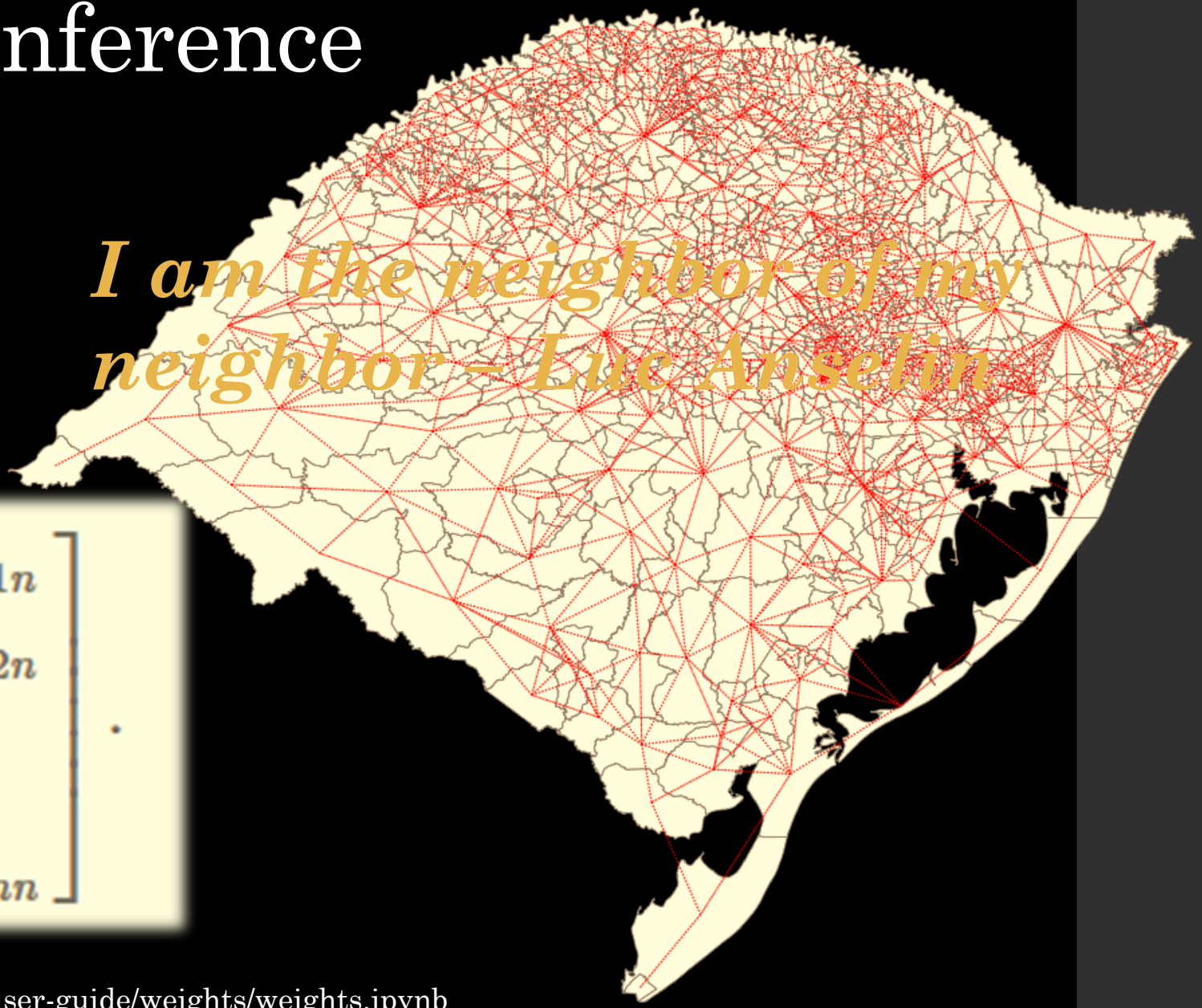
$$\mathbf{w} = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nn} \end{bmatrix}.$$

- Where:
- W is a 2D array or table which can be thought of as having essentially feature ID's as both the row numbers and column names. It usually has 0's on the diagonal.
- W_{ij} is the weights matrix element suggesting the influence on the variable of interest between geographies *i (row)* and *j (column)*.
- We scramble this matrix, or permute on the available values, for each of 100+ iterations of running I_i to establish a significance value for the focal geography.

Weights and Inference

- W_{ij}

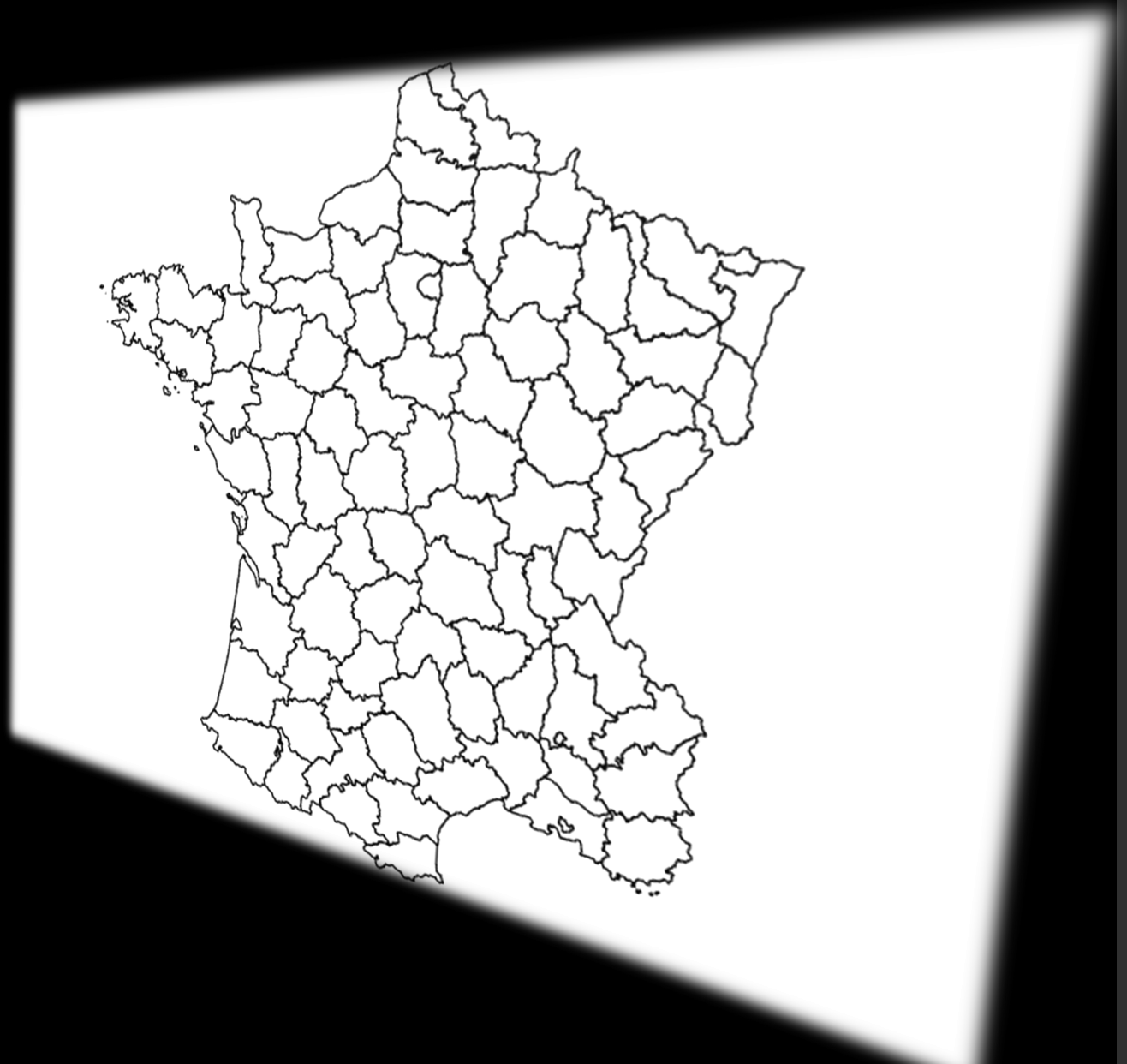
I am the neighbor of my neighbor – Luc Anselin



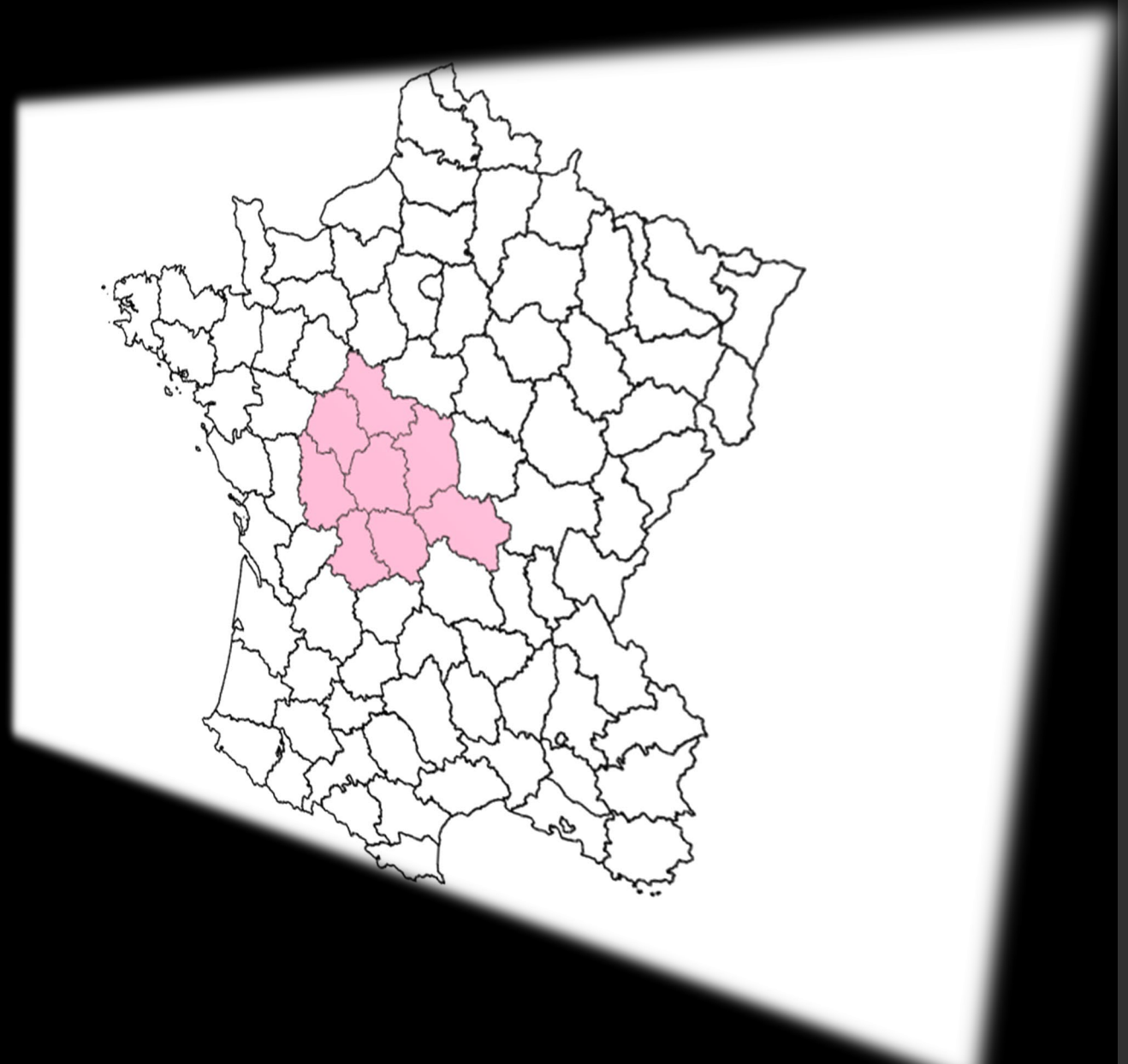
$$\mathbf{W} = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1n} \\ w_{21} & w_{22} & \dots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \dots & w_{nn} \end{bmatrix}.$$

https://nbviewer.org/github/pysal/libpysal/blob/main/docs/user-guide/weights/weights.ipynb

Inference and Conditional Randomization

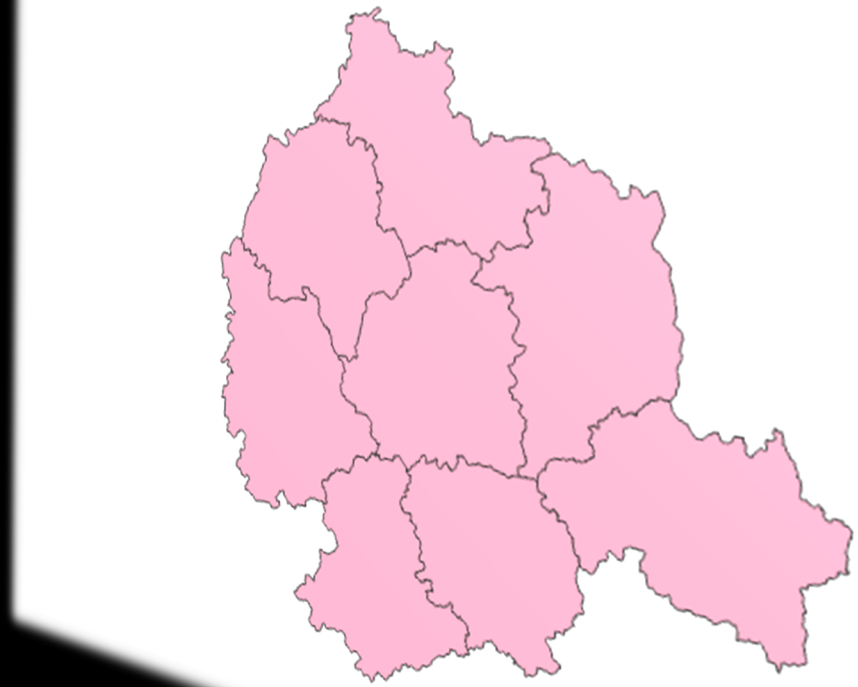


Inference and Conditional Randomization

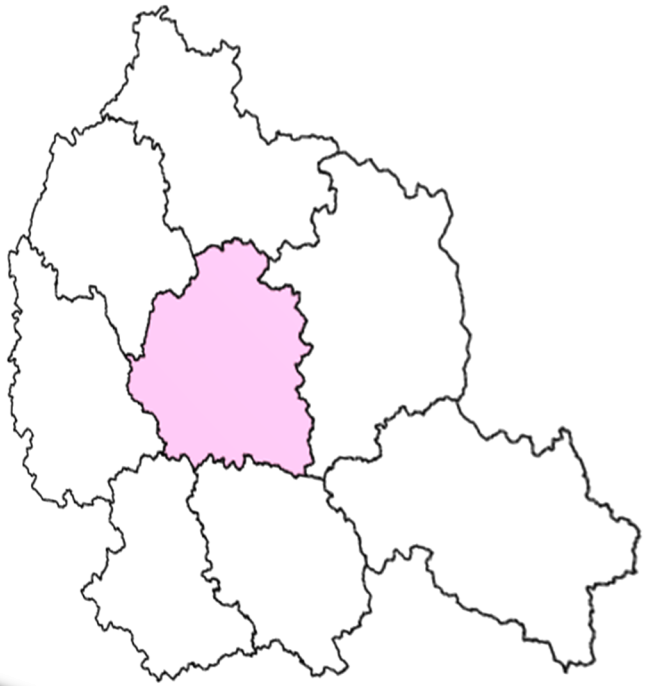


Inference and Conditional Randomization

Inference and Conditional Randomization

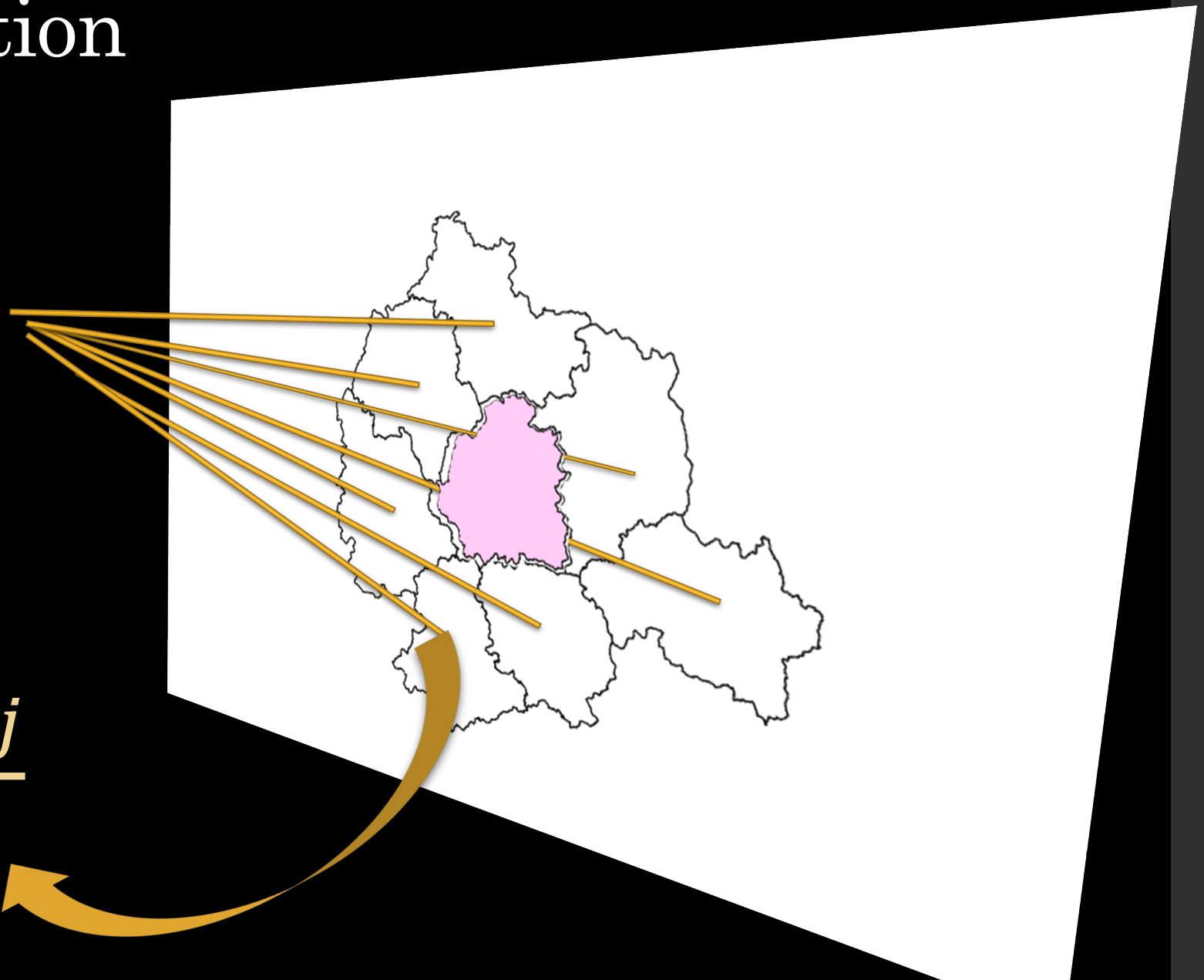


Inference and Conditional Randomization

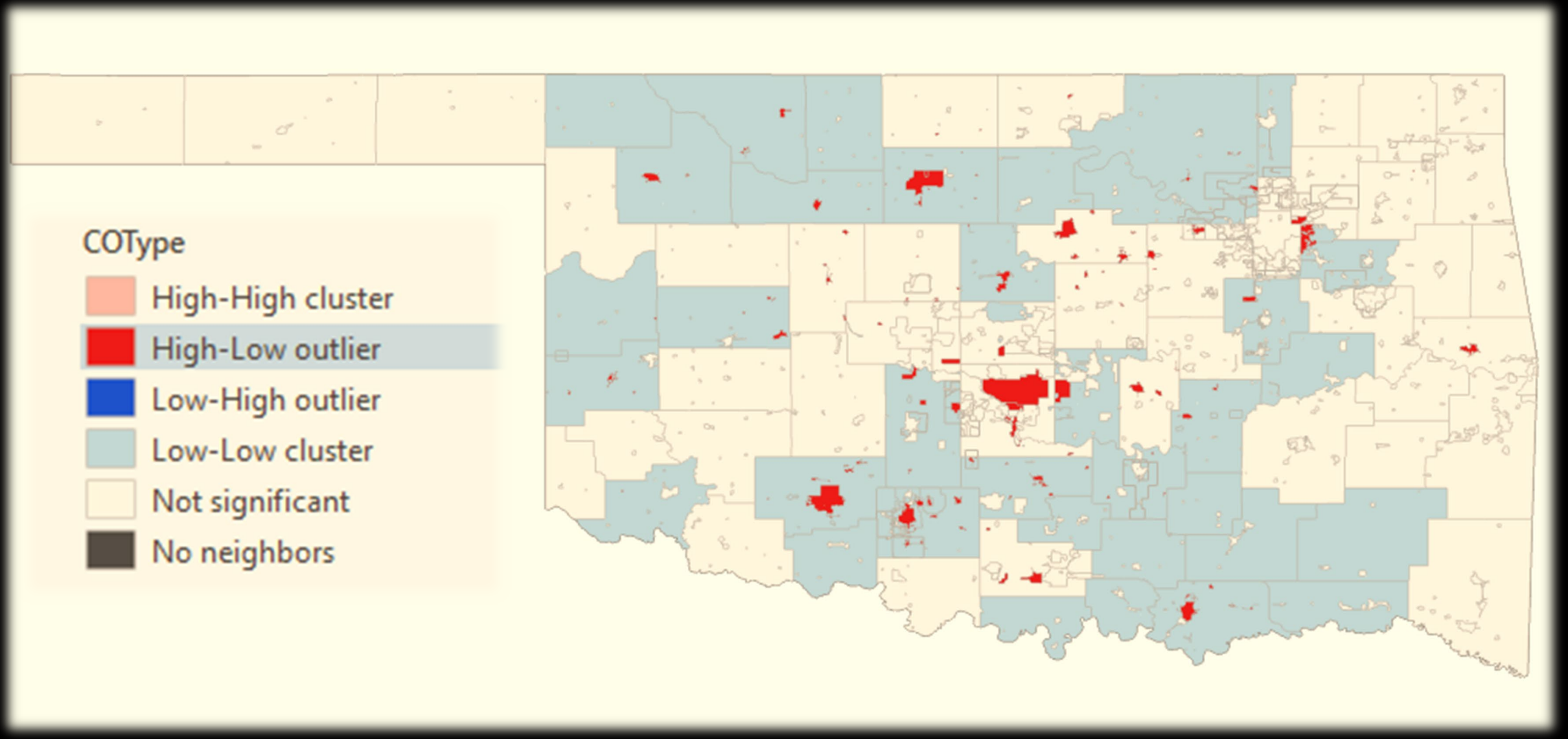


Inference and Conditional Randomization

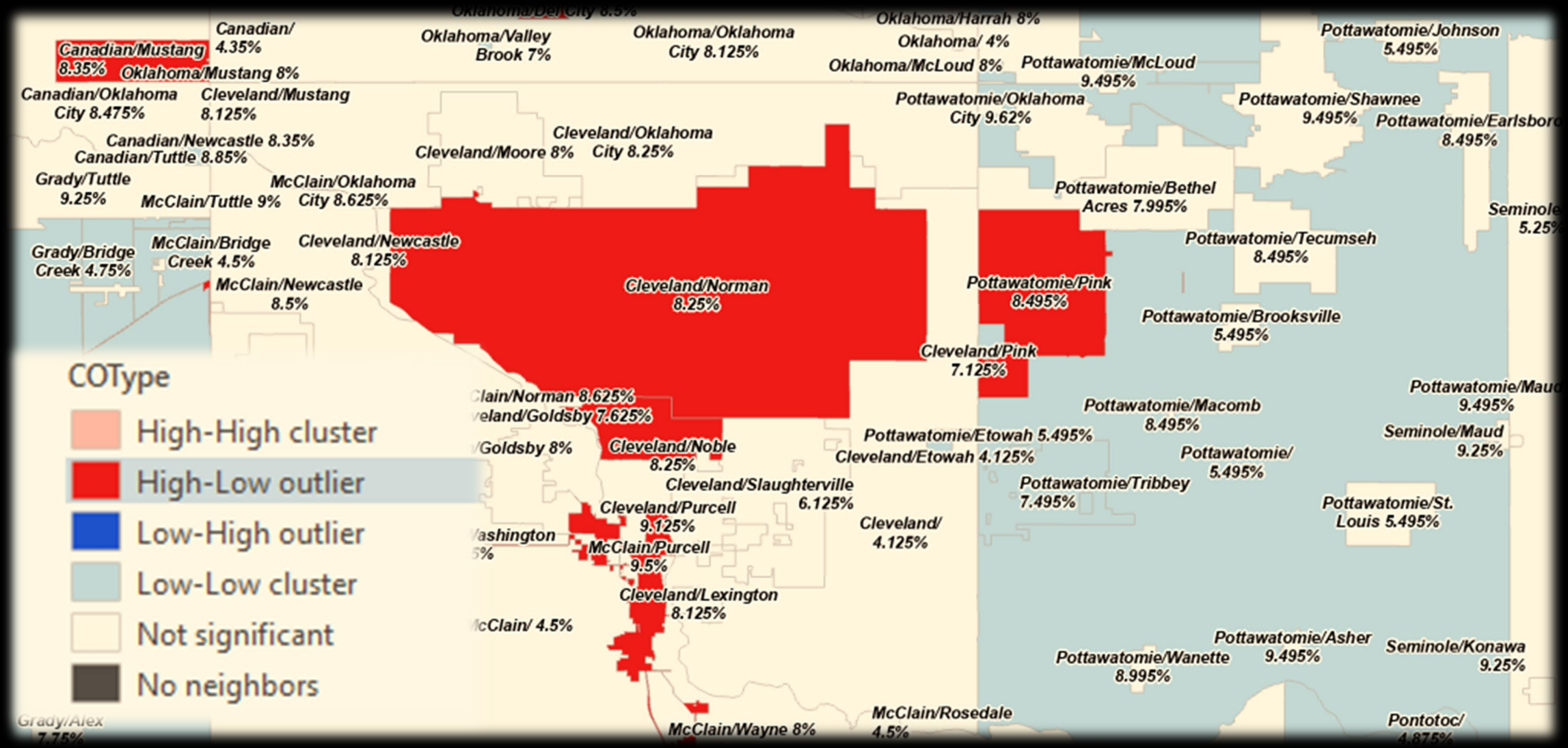
$$I_i = \frac{\sum_j w_{ij} z_i z_j}{\sum_i z_i^2}$$



Local Moran's I Results for Oklahoma Sales Taxes



Local Moran's I Results for Oklahoma Sales Taxes

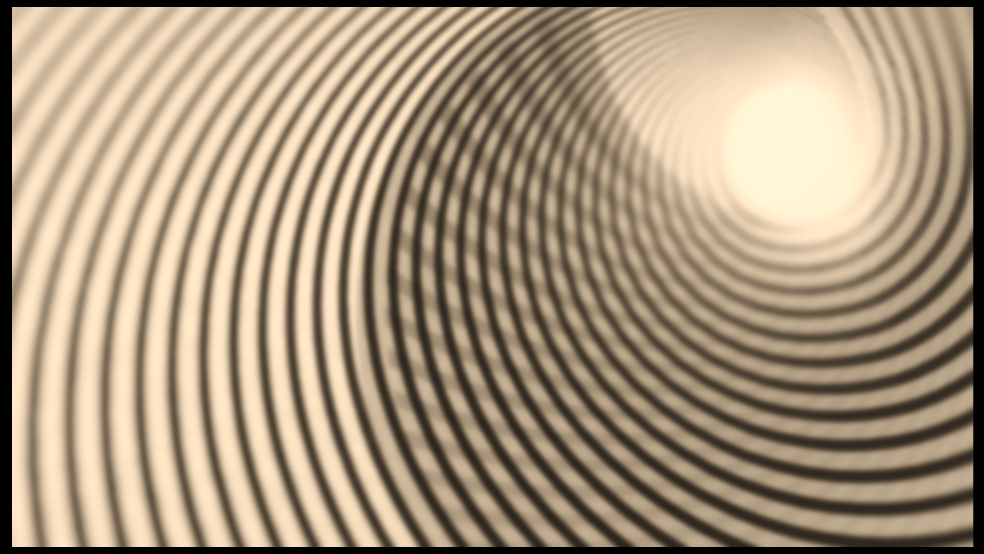


What are we looking at? What can we Conclude?

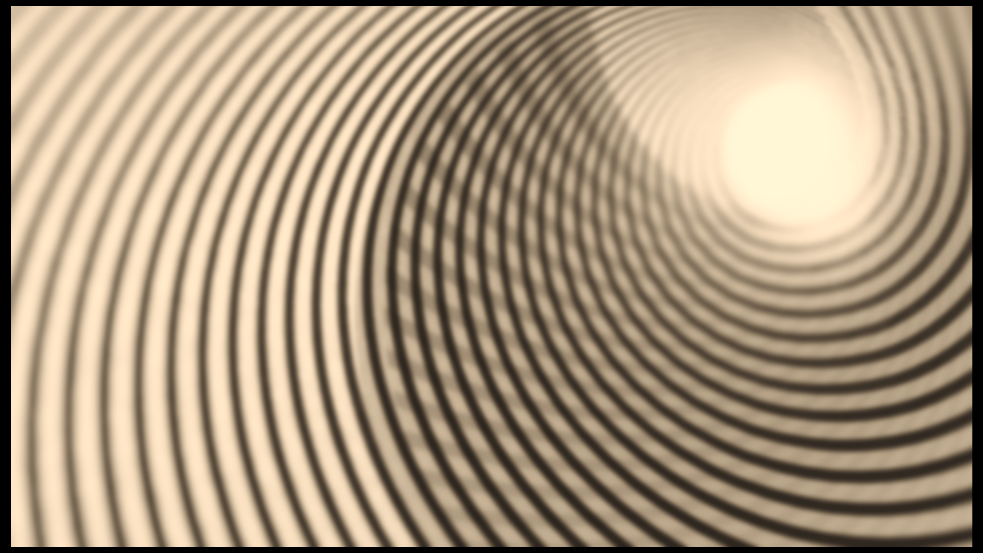
- This is a map identifying clustering locations (local spatial autocorrelation) in a single variable (sales tax rates) for Oklahoma taxable jurisdictions, with rates effective Q4 2025.
- *We have not* normalized by population, median incomes, or any other socioeconomic indicator incorporating *sales activity* along with the rates themselves.
- *We are treating rates as ratio data.*
- Where:
  - Results are significant at the $p = 0.01$ *level.*
  - The edges and corners of features establish contiguity(weights matrix, "*Queen's Contiguity*")
  - Conditional permutation for the *Local Moran's I* is calculated with draws from the entire set of available sales tax rates $n=854$.

What are we looking at? What can we Conclude?

- *Ultimately,* the idea that Oklahoma taxpayers could routinely shop at a national department store chain, in a neighboring tax jurisdiction, at a significantly lower sales tax rate *doesn't appear feasible given the pattern of Moran's I clustering for sales tax rates* current for Q4 2025.
- It's possible that available city services, population densities/market potential, *Municipal Incentives For New Businesses,* and labor availability **all influence business site selection, regardless of prevailing sales tax rates at the site.**
- In parallel, cities set sales tax rates democratically; rate increases or decreases might be positioned to offset business-attracting tax incentives themselves.
- The analysis as presented here suggests latent effects are present.



So, What is a P-Value?



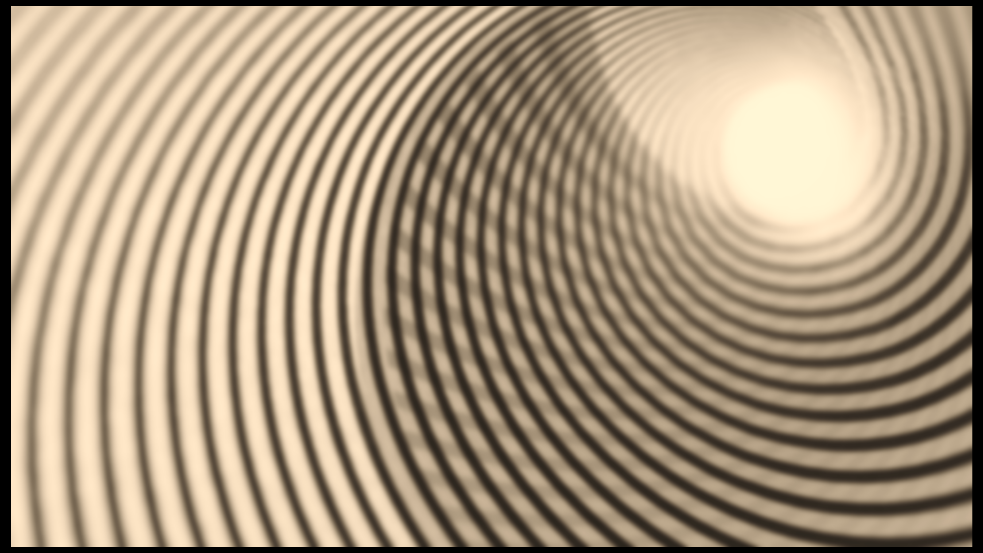
MATERIÉL

Johnson, Alicia A., Miles Q. Ott, and Mine Dogucu. *Bayes Rules! An Introduction to Bayesian Modeling with R*. Chapman & Hall/CRC Texts in Statistical Science. CRC Press, 2022.

Hobbs, N. Thompson, and Mevin B. Hooten. *Bayesian Models: A Statistical Primer for Ecologists*. Princeton University Press, 2015.

Anselin, Luc. *Spatial Econometrics: Methods and Models*. Kluwer Academic, 2010.

Cliff, A. D., and J. K. Ord. *Spatial Autocorrelation*. Monographs in Spatial and Environmental Systems Analysis 5. Pion, 1973. https://archive.org/details/spatialautocorre0000clif.



LOGICIEL

https://geodacenter.github.io/

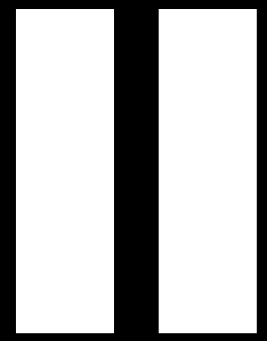
https://pysal.org/esda/index.html

https://cran.r-project.org/web/packages/spdep/index.html

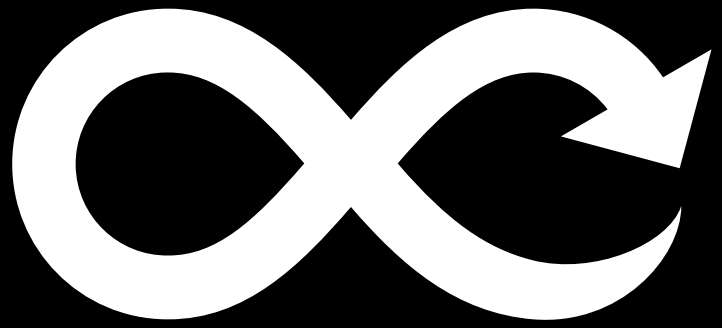
https://geopandas.org/en/stable/

https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/cluster-and-outlier-analysis-anselin-local-moran-s.htm

PAUSE



RESUME



What Other Activities is CSA Engaged In?

- CSA staff and academic leadership represent a great deal of technical and scientific expertise in GIS.
- CSA hosts, populates and maintains turn-key GIS infrastructure.
- CSA consults on the implementation and enculturation of GIS within spatially-concerned organizations.
- CSA may consult on a temporary or hourly basis for smaller needs and can provide commensurate temporary hosting and software up to delivery of any data product.

CSA's Prior Art in Oklahoma

- The CSA Story Map (A Brochure):
  - https://storymaps.arcgis.com/stories/10aedd9e282944c6845e36b832212f6f
  - The CSA Data Warehouse (Open Data)
  - https://csagis-uok.opendata.arcgis.com/
- The Oklahoma Water Resources Board
  - https://owrb.csa.ou.edu/portal/home/
- The Oklahoma Boll Weevil Eradication Organization
  - https://obweo.org/portal/apps/sites/#/obweo

Questions?

Contact

- Leah Nash
- Jim Anderson, GISP
- OU Center for Spatial Analysis - CSA (DGES)
  - ldnash@ou.edu
  - (405) 325-3131
  - janderson@ou.edu
  - (405) 325-4871
  - https://www.ou.edu/ags/csa

- Happy to Help.