

Spatial Inequality

Sergio J. Rey¹

¹ San Diego State University, San Diego, USA

Received: August 28, 2024/Accepted: February 7, 2025

Abstract. This paper explores the concepts and computational methods used to measure spatial inequality, emphasizing a reproducible approach that social scientists can apply to their research. The analysis focuses on geographic income disparities at the sub-national level, using Mexico as a case study. By examining various a-spatial and spatially explicit approaches, the paper highlights the complexities of measuring inequality across places and over time. The discussion includes a review of traditional inequality measures and introduces spatial decomposition methods that account for the geographical distribution of income. The findings underscore the importance of integrating spatial considerations into inequality analysis to better understand the patterns and drivers of regional disparities, thereby informing more effective and equitable policy interventions.

Geographic income inequality has risen more than 40% between 1980 and 2021.
– [U.S. Department of Commerce \(2023\)](#)

1 Introduction

The study of spatial, or geographical, disparities is crucial for both scientific and policy-oriented reasons. Scientifically, understanding these disparities allows researchers to uncover patterns and correlations that are vital for advancing knowledge in various fields such as economics ([Kanbur, Venables 2005](#)), public health ([Deb Nath, Odoi 2024](#)), and environmental science ([Venter et al. 2023](#)). From a policy perspective, recognizing and addressing geographical disparities is essential for promoting social equity and economic development. A prime example is the European Union's Cohesion Policy, where the reduction of spatial disparities between member regions takes center stage ([Widuto 2019](#)). Additionally, understanding spatial disparities can inform urban planning and environmental policies to create more sustainable and resilient communities.

By bridging the gap between scientific research and policy implementation, the study of geographical disparities helps in crafting evidence-based strategies that promote balanced regional development, reduce inequalities, and improve the overall quality of life. Ultimately, this interdisciplinary approach fosters a deeper understanding of the complex dynamics at play and supports the creation of more inclusive and effective policies.

The field of spatial data science ([Rey et al. 2023](#)) provides tools to visualize and analyze spatial inequality. Methods such as local spatial autocorrelation analysis ([Anselin 1995](#)), spatial distribution dynamics ([Rey 2015](#)), and regionalization ([Wei et al. 2020](#)) offer robust frameworks to identify patterns, relationships, and variations across geographical contexts. Thus, spatial data science is well-positioned to support such interdisciplinary research. The goal of this paper is to introduce social scientists to the concepts and measurement of spatial inequality. The emphasis is on adopting a computationally focused

and reproducible treatment that would allow researchers to apply the methods introduced here to their own investigations.

The focus is on a case study of regional income inequality in Mexico, which provides a compelling example of spatial disparities and their implications for both scientific inquiry and policy-making. Mexico's diverse regional landscape, marked by significant economic, social, and cultural differences, offers a rich context for analyzing how inequality manifests across geographical spaces. By examining income inequality at the regional level, this case study highlights the spatial patterns and clusters of economic disparities, providing insight into the underlying processes that drive inequality.

The paper proceeds as follows. We first develop a conceptual understanding of the measurement of spatial inequality. Next, we describe the computational environment employed to analyze spatial disparities. The specific case study is then introduced. We then discuss different a-spatial approaches towards measuring inequality, followed by a detailed exploration of spatially explicit approaches for measuring geographical disparities. The paper concludes with the identification of future research areas in the field of spatial inequality.

2 Inequality Concepts

The growing concern with inequality brings to mind Peter Drucker's often-cited principle, "You can't improve what you don't measure." Before we can address the technical challenges of measuring spatial inequality, it is essential to first grapple with the conceptual issues surrounding what we are measuring.

It is important to distinguish between terms that frequently appear in the inequality literature: equality (inequality) and equity (inequity). Equality refers to the state of being equal, particularly in status, rights, and opportunities. In economics, this often means distributing resources and opportunities uniformly across all individuals or groups.

Equity, conversely, involves fairness and justice in the distribution of resources and opportunities. It considers individual needs and circumstances, aiming to level the playing field. Therefore, although inequality and inequity are interconnected within the context of social justice (Sen 2004), they are not synonymous.

For instance, an equal distribution of resources, such as uniform per capita expenditure on students, can create inequities by ignoring the challenges faced by students in different contexts, such as urban versus rural districts or advantaged versus disadvantaged neighborhoods (Tine 2017). Conversely, some distributions are intentionally unequal to achieve greater equity. Progressive income tax schemes, where the tax rate increases with income, are a prime example of this approach (Ledić et al. 2023).

A closely related distinction is between equality of outcomes and equality of opportunities. Inequality in outcomes refers to the unequal distribution of income, wealth, and resources among individuals in a society, which can result from factors such as luck, effort, and inherited wealth. In contrast, inequality in opportunities focuses on the unequal access to education, healthcare, and other essential services that enable individuals to achieve their potential, irrespective of their background (Roemer 1998). Studies may differ, then, in whether they measure spatial inequality in outcomes (Khedmati Morasae et al. 2024) or spatial disparities in opportunities (Knaap 2017).

In addition to the distinction between inequality and inequity, and between outcomes versus opportunities, there is much variation in the substantive variable under focus. Income studies dominate the literature quantitatively (Gaubert et al. 2021) and are sometimes contrasted with studies of the inequality of wealth (Suss et al. 2024).¹ More granular studies examine disparities in the sources of income, such as wages, as well as employment rates (Overman, Xu 2022). Outside of economics, topics such as disparities in educational outcomes (Graetz et al. 2020), health outcomes (Khedmati Morasae et al. 2024), voting patterns (Barber, Holbein 2022), among many others, are replete across the social and life sciences.

¹It is important to keep in mind that income is measured as a flow whereas wealth is a stock. This distinction matters in terms of the way disparities in the two variables are examined. See the discussions in (Saez, Zucman 2016) and (Piketty 2014).

The unit of measurement employed in inequality analysis is also an important consideration. [Sala-i-Martin \(2006\)](#) demonstrates that when using countries as the unit of analysis, the picture that emerges is one of large and static levels of international inequality. However, when the analysis used countries weighted by their populations, the view is one of declining inequality over time. Later, we shall see that this choice of weighting versus non-weighting of the units is an important issue in spatial inequality measurement.

There is much variation across the inequality literature in the unit of measure. Studies of personal income inequality often focus on data recorded for individuals ([Piketty, Saez 2003](#)). Other studies take the household or family as the unit of analysis ([Brandolini, Smeeding 2011](#)). In both cases, the focus is on income inequality across people. This is an essential point of departure for our study of spatial inequality, which is where the unit of measurement is a geographical area. In other words, in spatial inequality analysis, the focus is inequality across places.

Still further, some studies examine the spatial distribution of personal income inequality ([Partridge et al. 1998, Frank 2009](#))—that is, how inequality between individuals within a state varies across states. In the spatially oriented inequality studies, the geographical unit of analysis can range from countries ([Milanović 2018](#)), to intra-national regions ([Ganong, Shoag 2017](#)), to cities ([U.S. Department of Commerce 2023, Sarkar et al. 2024](#)), and down to neighborhoods ([Nijman, Wei 2020](#)).

A final inequality concept we need to consider is the role of time. One question is the time unit. Is income measured per person, per year, or is some life-time earnings, or permanent income ([Hall 1978](#)) measure employed? A second question pertains to whether the study of inequality is a snapshot at one point in time or focuses on the dynamics of income distribution.

Addressing all these issues is beyond the scope of any one study. We raise them here in order to situate the study of spatial inequality in a much broader context. For this paper, we will hone in on the question of measuring spatial income inequality at the sub-national scale.

3 Computational Environment

In the following section, we present the packages and computational environment used. The narrative following code cells explains the computational concepts.

3.1 Packages

```
[1]: import inequality as ineq ①
import numpy as np ②
import pandas as pd
import geopandas as gpd
import libpysal as lps
import seaborn as sns
import matplotlib.pyplot as plt
import watermark ③
%load_ext watermark
%watermark -a "Sergio Rey" -u -n -t -v \
-p numpy,pandas,scipy,matplotlib,inequality,seaborn,libpysal
```

Notes:

- ① We alias the package `inequality` as `ineq`
- ② We do the same for `numpy`
- ③ `watermark` reports the current versions of our packages to support reproducibility

```
[1]: Author: Sergio Rey
Last updated: Mon Dec 16 2024 09:17:39
Python implementation: CPython
Python version       : 3.10.16
IPython version      : 8.30.0
```

```

numpy      : 2.2.0
pandas     : 2.2.3
scipy      : 1.14.1
matplotlib: 3.9.4
inequality: 1.0.2.dev78+gbb07357
seaborn    : 0.13.2
libpysal   : 4.12.1

```

Our analysis of spatial inequality utilizes packages from the Python Spatial Analysis Library (`pysal`) (Rey et al. 2022), together with `geopandas` (Jordahl et al. 2019) and `seaborn` (Waskom 2021). The main package focusing on measuring spatial disparities is `inequality`, which implements analytics for measuring a-spatial inequality and spatially explicit inequality measures. Also, from `pysal`, we will depend on the `libpysal` package for constructing spatial weights that are central to the analysis of spatial inequality. `geopandas` provides for spatial data processing and producing maps of the spatial distribution of income, while we adopt `seaborn` for constructing a-spatial graphical views of inequality.

For purposes of reproducibility, we include the `watermark` package which reports the version numbers of each of the packages we use in our analysis.

4 Data

To illustrate the core concepts in spatial inequality measurement, we will rely on a data set for the states in Mexico (Rey, Sastré-Gutiérrez 2010). The variable of interest is state per capita gross domestic product (`pcgdp`) measured in 2000 USD measured for each decade from 1940-2000 for each of 32 areas consisting of the 31 federal states of Mexico plus Mexico City. This data-set is included as an example data-set in `libpysal`.

```
[2]: lps.examples.explain("mexico")
```

```

[2]: mexico
=====

Decennial per capita incomes of Mexican states 1940-2000
-----

* mexico.csv: attribute data. (n=32, k=13)
* mexico.gal: spatial weights in GAL format.
* mexicojoin.shp: Polygon shapefile. (n=32)

Data used in Rey, S.J. and M.L. Sastre Gutierrez. (2010) "Interregional inequality
dynamics in Mexico." Spatial Economic Analysis, 5: 277-298.

```

In addition to the income data contained in the `mexico.csv` file, there are two additional files available in this example: `mexico.gal` which stores information about the contiguity relationships between the states, and `mexicojoin.shp` which is a shapefile.

The following code block produces Figure 1 which lists the locations of the 32 Mexican states.

```

[3]: # Create a map to provide
      # context for the subsequent analysis

      # We use the library mapclassify for choropleth classifications
      import mapclassify

      # set the path to our shapefile and read the file into a geodataframe
      pth = lps.examples.get_path("mexicojoin.shp")
      gdf = gpd.read_file(pth)

      # we will use greedy from mapclassify
      # states to ensure contiguous states are of a different color
      sgdf = gdf.sort_values(by='NAME')
      sgdf.reset_index(inplace=True)
      sgdf['label'] = range(1, 33)
      sgdf['greedy'] = mapclassify.greedy(sgdf)

```

```

# set the font size and position
font_size = 9
outside = [9, 29]
oc = [(-103, 17.5), (-95, 22.5)]
oe = [(-102.55, 17.49), (-95.5, 22.1)]
oinfo = zip(outside, oc)

# we will use LineStrings from shapely
# for call-outs to state names
from shapely.geometry import LineString

# plot the map
import matplotlib.pyplot as plt
sgdf['centroid'] = sgdf.centroid
ax = sgdf.plot(
    figsize=(8, 12),
    column="greedy",
    categorical=True,
    cmap="Set3",
    #legend=True,
    edgecolor="w",
)

# build the table of state names
table = []
for idx, row in sgdf.iterrows():
    centroid = row['centroid']
    table.append(f'{idx+1:2d} {row["NAME"]}')
    if idx+1 not in outside:
        ax.text(centroid.x, centroid.y, str(idx+1), ha='center',
              va='center', fontsize=font_size, color='black')

# add the call-outs and number the polygons
i = 0
for out in oinfo:
    idx, coords = out
    ax.text(coords[0], coords[1], str(idx), ha='center',
          va='center', fontsize=font_size, color='black')
    start_point = coords
    end_point = sgdf.centroid[idx-1]

    # Create a LineString object
    start_point = oe[i]
    line = LineString([start_point, end_point])

    # Create a GeoSeries for the line
    line_gdf = gpd.GeoSeries([line])

    # Plot the line
    line_gdf.plot(ax=ax, color='red', linewidth=2)
    i+=1

for i, label in enumerate(table):
    if i < 16:
        ax.text(-120, 20-i*1, label, ha='left',
              va='center', fontsize=font_size, color='black');
    else:
        ax.text(-110, 20-(i-16)*1, label, ha='left',
              va='center', fontsize=font_size, color='black');
ax.set_axis_off()

```

[3]: Output in Figure 1

We can read the `mexico.csv` file using `pandas` to create a `DataFrame` that will hold the attributes of interest. This dataframe has the shape `(32,13)` indicating there are 32 observations on 13 variables. The 13 variables include the per capita gross domestic product for each decade, for example `pcgdp1990` for 1990, together with the state name, and five variables that define different regionalization schemes for the country. We will return to these regional variables in a later section.



Figure 1: States of Mexico

```
[4]: # Create a DataFrame from a csv file
pth = lps.examples.get_path("mexico.csv")
df = pd.read_csv(pth)
print(f"Shape of dataframe: {df.shape}")
print(f"First 5 rows of dataframe:\n {df.head()}")
df.head()
print(f"\nVariables: {df.columns}")
```

```
[4]: Shape of dataframe: (32, 13)
First 5 rows of dataframe:
   State  pcgdp1940  pcgdp1950  pcgdp1960  pcgdp1970  pcgdp1980
0  Aguascalientes  10384.0   6234.0   8714.0   16078.0   21022.0
1  Baja California  22361.0  20977.0  17865.0  25321.0  29283.0
2  Baja California Sur  9573.0  16013.0  16707.0  24384.0  29038.0
3    Campeche      3758.0   4929.0   5925.0  10274.0  12166.0
4    Chiapas      2934.0   4138.0   5280.0   7015.0  16200.0

   pcgdp1990  pcgdp2000  hanson03  hanson98  esquivel99  inegi  inegi2
0  20787.0   27782.0         2.0         2.0         3.0   4.0   4.0
1  26839.0   29855.0         1.0         1.0         5.0   1.0   1.0
2  25842.0   26103.0         2.0         2.0         6.0   1.0   1.0
3  51123.0   36163.0         6.0         5.0         4.0   5.0   5.0
4  8637.0    8684.0         5.0         5.0         7.0   5.0   5.0

Variables: Index(['State', 'pcgdp1940', 'pcgdp1950', 'pcgdp1960', 'pcgdp1970',
                  'pcgdp1980', 'pcgdp1990', 'pcgdp2000', 'hanson03', 'hanson98',
                  'esquivel99', 'inegi', 'inegi2'],
                dtype='object')
```

5 Measuring Spatial Inequality in Mexico

5.1 Visualizing Inequality in Distributions

We begin with different perspectives on the distribution of state incomes in Mexico. There are two different approaches towards visualizing the distribution: one focusing on the geographical distribution and the second on the attribute distribution. From a geographical perspective, Figure 2 shows the spatial distribution of incomes for the last year of the sample (2000).

```
[5]: # Create a choropleth map of per capita gross
# domestic product in 2000 using quintiles for the classification
pth = lps.examples.get_path("mexicojoin.shp")
gdf = gpd.read_file(pth)
ax = gdf.plot(column="PCGDP2000", k=5, scheme="Quantiles",
             legend=True,
             edgecolor='grey',
             legend_kwds={"loc": "center left",
                          "bbox_to_anchor": (1, 0.5),
                          "fmt": "{:.0f}"}))

ax.set_axis_off()
ax.set_title("PC GDP 2000");
```

[5]: Output in Figure 2

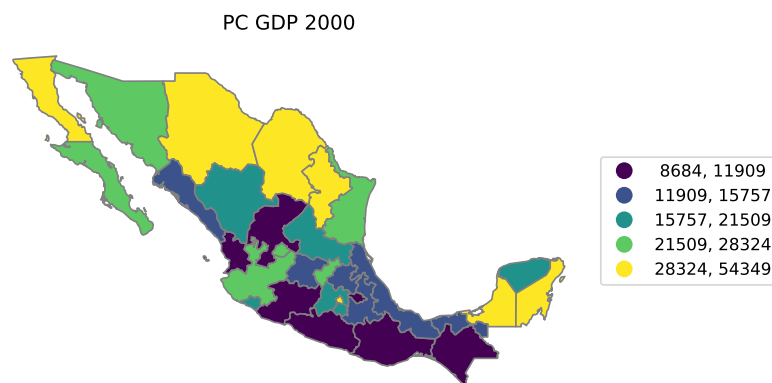


Figure 2: Per Capita Gross Domestic Product by State (Quintiles)

This is a choropleth map using quintiles to classify the incomes. The visual impression is that incomes are not randomly distributed in Mexico, as the states with incomes below the bottom quintile are more concentrated in the south, while in the north, the highest income states dominate. We will be able to make more quantitative evaluation of this spatial pattern later on in this paper.

Figure 3 presents a different perspective on income distribution based on the concept of Pen's Parade, introduced by Dutch economist Jan Pen (Pen 1971). The metaphor uses a Parade to illustrate economic inequality, with each person representing a state in the economy, and their height being proportional to the state's per capita income. The Parade starts with the shortest individuals depicting the poorest states, gradually increasing in height as income rises. At the end of the parade, the tallest individuals represent the most affluent states, highlighting the significant income disparities across states. This visual tool showcases the inequality in income distribution. The parade uses two different scales, with the x-axis showing the ordinal distances between states and the y-axis representing the interval distances in their per capita incomes.

```
[6]: # Produce the Pen's Parade for 2000
from inequality.pen import pen
f = pen(gdf, 'PCGDP2000', 'NAME')
```

[6]: Output in Figure 3

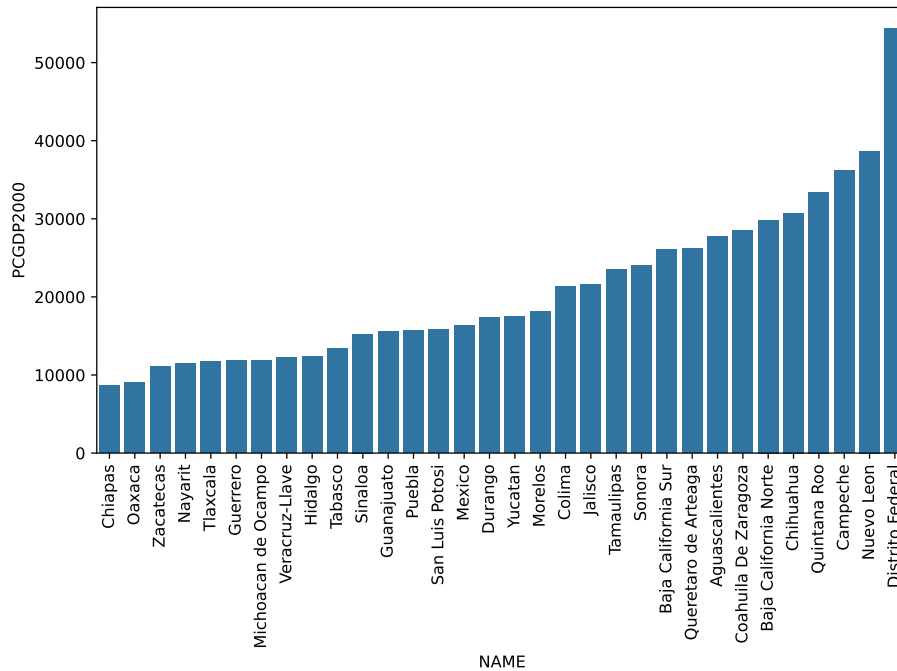


Figure 3: Pen's Parade Per Capita Gross Domestic Product by State 2000

We can start to integrate the spatial and attribute distributions together using a `pengram` from the `pysal-inequality` package:

```
[7]: # Produce pengram for 2000
from inequality.pen import pengram
f = pengram(gdf, 'PCGDP2000', 'NAME', xticks=False, leg_pos='lower left',
            fmt="{:.0f}")
```

[7]: Output in Figure 4

As shown in Figure 4, the `pengram` combines the Pen's Parade alongside the choropleth map. This affords a more granular view of the distribution than those offered by either view in isolation. For example, one of the well-known limitations of a quintile classed map is that the intra-class variation is obscured. In the `pengram`, the intra-class variation now becomes visible through the Pen's Parade, revealing the much larger absolute and relative variance above the upper quintile relative to the other classes.

A second feature of the `pengram` is the ability to query for specific observations. This makes it possible to locate the position of a state in both the Pen's Parade (attribute space) as well as on the map (geographical space). We do this for states occupying the two extremes of the attribute distribution: Chiapas and Distrito Federal in Figure 5. While the two reside in the extremes of the attribute distribution, the high income Distrito Federal is in the center of the geographic distribution while Chiapas is on the southern border of the country. Moreover, although Distrito Federal stands out in the Pen's Parade, its small geographic area makes it difficult to identify on the map without the query functionality of the `pengram`.

```
[8]: # Code snippet to produce the pengram
# with query for Chiapas and Distrito Federal
from inequality.pen import pengram
f = pengram(gdf, 'PCGDP2000', 'NAME', leg_pos='lower left',
            fmt="{:.0f}",
            xticks=False,
            query=['Chiapas', 'Distrito Federal'])
```

[8]: Output in Figure 5

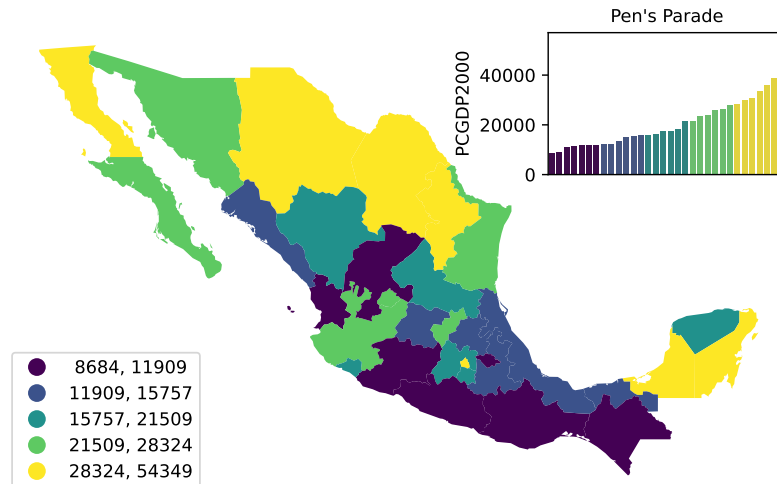


Figure 4: Pengram Per Capita Gross Domestic Product by State 2000

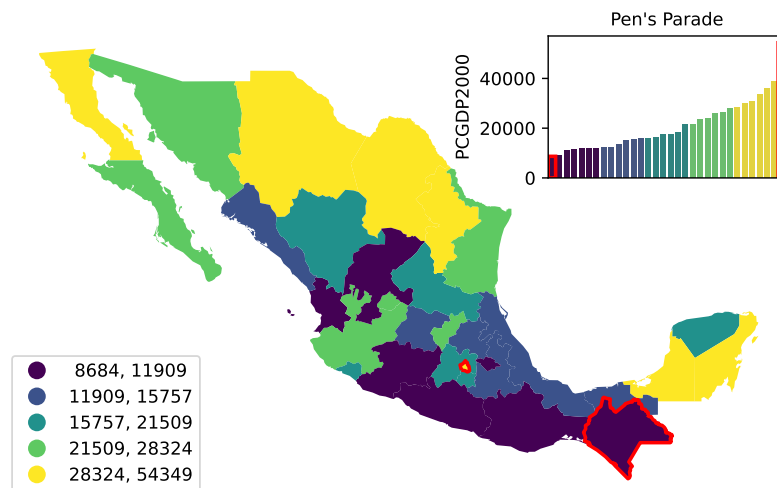


Figure 5: Querying the Pengram of Per Capita Gross Domestic Product by State 2000

Returning to a more granular view of the attribute distribution, Figure 6 combines a histogram of the distribution together with a kernel density estimate, and a rug plot. The kernel density estimate is a smooth curve that represents the probability density function of the data, providing a continuous approximation of the underlying distribution. The rug plot signifies the positions of each state as short ticks on the x-axis. The outlier nature of the Distrito Federal that we saw in the `pengram` is responsible for the positive (right) skew of the density function. There is some evidence of polarization in the distribution with the mode being at the poorest group of states and other, lower, peaks in the middle of the distribution.

```
[9]: # Produce histogram,
# kernel density and rug plot
years = range(1940, 2010, 10)
yvars = [f'pcgdp{year}' for year in years]
sns.histplot(df[yvars[-1]], kde=True, element="step",
             bins=10, edgecolor="black")
sns.rugplot(df[yvars[-1]], color='blue')
plt.show()
```

[9]: Output in Figure 6

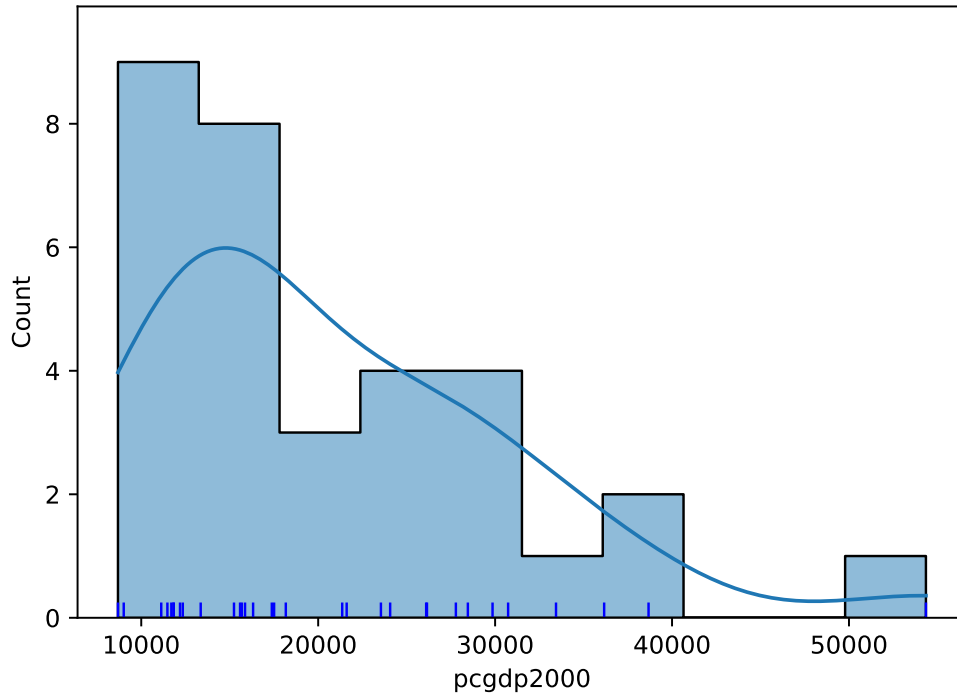


Figure 6: Distribution Per Capita Gross Domestic Product by State 2000 (pcgdp2000), histogram, kernel density, rug plot

Inequality in a distribution is often considered by an inspection of the shares of overall income belonging to units at different locations within the distribution. Here, we must keep in mind, the distinction between the different units under study in the analysis of spatial versus personal income inequality. In spatial inequality analysis, we essentially treat each state as an “individual” and set that individual’s level of income to the state’s per capita income. The share for the state is then derived as the ratio of its per capita income to the sum of the per capita income of all states.

These shares can be portrayed in a Lorenz curve, shown in Figure 7, which orders the states by their per capita incomes from lowest to highest. Then, against the cumulative proportion of states (x-axis) we plot the cumulative income share on the y-axis. Both scales have limits of $[0, 1]$. In the case of perfect equality, where all state per capita incomes are equal, this plot would be a 45-degree line, the so called line of perfect equality. Any departure from perfect equality will result in states with above average per capita income receiving more than $1/n$ share of $n\bar{y}$, while states with below average per capita incomes will have shares below $1/n$.²

```
[10]: # Produce Lorenz Curve and Schutz Line
from inequality.schutz import Schutz
s = Schutz(gdf, 'PCGDP2000')
s.plot(xlabel='Share of States', ylabel='Share of Per Capita Income')
```

[10]: Output in Figure 7

5.2 Measures of Inequality

A large number of measures of inequality exist, and choosing among the rich diversity of inequality measures has been the subject of a vast literature.³ To help guide that selection, there are five desirable properties of an inequality measure:

²States with shares below $1/n$ would also have location quotients of less than one for their relative per capita incomes, where their per capita income was expressed relative to national per capita income.

³See Cowell (2011) for an overview of inequality measures.

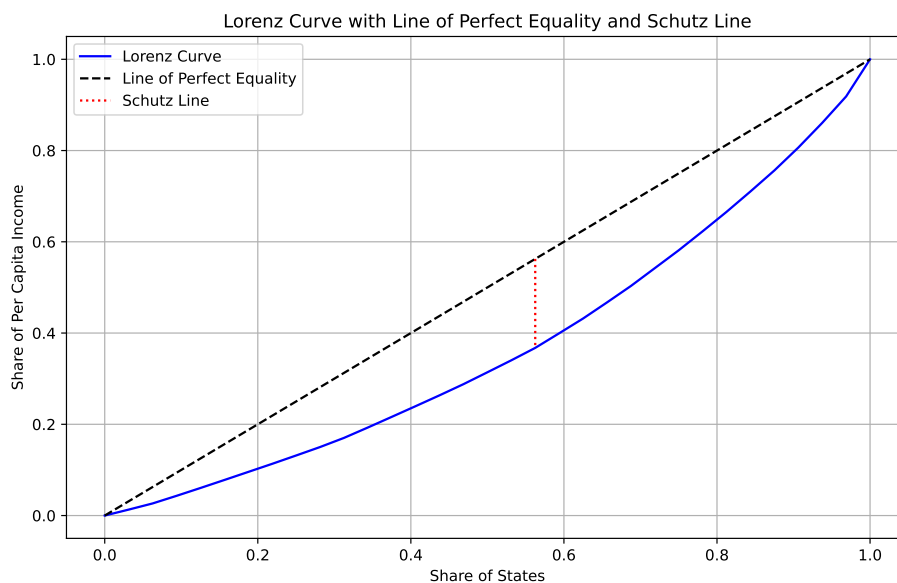


Figure 7: Lorenz Curve and Schutz Line Per Capita Gross Domestic Product by State 2000

1. Symmetry or anonymity
2. Principle of transfers
3. Scale invariance
4. Replication invariance
5. Zero normalization

Symmetry implies that the names of the income receiving unit should be immaterial. That is, if we swap the income of one geographical unit with that of another, the overall inequality measure should not change. The principle of transfers implies that the measure should reflect a reduction in inequality if income is transferred from a richer unit to a poorer unit, as long as the transfer does not reverse their income ranking. Scale invariance means that if all incomes are multiplied by the same constant, inequality remains unchanged. Replication invariance implies that an inequality measure should be unaffected if the population is replicated, meaning that duplicating the entire income distribution does not alter the measure of inequality. Zero normalization means that an inequality measure assigns a value of zero to a perfectly equal income distribution, serving as a baseline where all units have the same income.

Not all of the inequality measures satisfy all of these five properties. For example, the variance is not scale invariant, and both the logarithmic variance and variance of logarithms fail the principle of transfers property. Even for measures that respect these five properties, their use in comparing the levels of inequality between countries at one point in time, or the same country at different points of time, requires the researcher to specify a social welfare function (Atkinson 1970).

We can derive two indices from the Lorenz curve: the Gini coefficient and the Schutz coefficient. The Gini coefficient is an area measurement expressed as a ratio of the area of the lens above the Lorenz curve but below the diagonal to area under the line of perfect equality. The Schutz coefficient is a distance measure defined as the maximum vertical distance between the equality diagonal and the Lorenz curve. Both coefficients measure the extent to which the Lorenz curve departs from perfect equality. Moreover, both indices run from 0, perfect equality, to 1, maximal inequality.

For 2000, the Gini coefficient is 0.258, while the Schutz coefficient is 0.195. As they are both scaled to be within the unit interval, the temptation is to compare the two values. However, this would be misleading as we recall one is a measure of area, the second a measure of distance. Instead, each coefficient is often used to compare across distributions, either over time or space. We can do this for Mexico by asking what has

happened to inequality over time. Along the way, we add a third commonly used indicator of inequality, the coefficient of variation (*CV*). The coefficient of variation is a relative measure of variation as it is defined as the ratio of the standard deviation to the sample mean.

```
[11]: # Code snippet to calculate inequality measures for each decade
# and display in a table
yvars = [f'PCGDP{year}' for year in years]
ginis = [ineq.gini.Gini(gdf[yvar]).g for yvar in yvars]
res_df = pd.DataFrame(data=ginis, columns=['Gini'], index=years)
cv = gdf[yvars].std() / gdf[yvars].mean()
res_df['CV'] = cv.values
s = [ineq.schutz.Schutz(gdf, yvar).distance for yvar in yvars]
res_df['Schutz'] = s
res_df['Gini_rank'] = res_df['Gini'].rank()
res_df['CV_rank'] = res_df['CV'].rank()
res_df['Schutz_rank'] = res_df['Schutz'].rank()
res_df
```

[11]: Output in Table 1

Table 1 shows that the three indices agree that inequality was lowest in 1980, and the highest in 1940. However, while the Gini and Schutz coefficients agree in their rankings, the inclusion of the CV creates discordance. Part of the discordance reflects the sensitivity of the measures to different parts of the income distribution. The Gini coefficient puts more weight on the middle of the distribution, while the CV is more affected by the right tail of the distribution. The discordance also reflects the property that when the Lorenz curves do not intersect, the CV and Gini would agree on the rankings of inequality. However, Figure 8 shows that there are cases where the Lorenz curves intersect.

The discordance in rankings complicates whether income inequality between states in Mexico has increased or decreased. The answer now depends upon which temporal interval one chooses. Both the Gini and CV agree that inequality has declined since 1940, irrespective of the terminal year selected. However, they disagree on the answer when the question is whether inequality decreased from, say, 1960 to 1970 or between 1990 and 2000 (see Table 1).

```
[12]: # Plot Lorenz Curves for each decade
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import cumfreq

def lorenz_curve(incomes):
    sorted_incomes = np.sort(incomes)
    cumulative_incomes = np.cumsum(sorted_incomes)
    normalized = cumulative_incomes / cumulative_incomes[-1]
    lorenz_curve = np.insert(normalized, 0, 0)
    n = len(incomes)
    x = np.linspace(0.0, 1.0, n + 1)
    return x, lorenz_curve
```

Table 1: Inequality Index Rankings

	Gini	CV	Schutz	Gini_rank	CV_rank	Schutz_rank
1940	0.353724	0.719858	0.260037	7.0	7.0	7.0
1950	0.296446	0.624611	0.213920	6.0	6.0	6.0
1960	0.253718	0.492447	0.181549	3.0	3.0	3.0
1970	0.255134	0.472039	0.185266	4.0	2.0	4.0
1980	0.245053	0.462657	0.179702	1.0	1.0	1.0
1990	0.251818	0.497729	0.181363	2.0	5.0	2.0
2000	0.258113	0.492565	0.195043	5.0	4.0	5.0

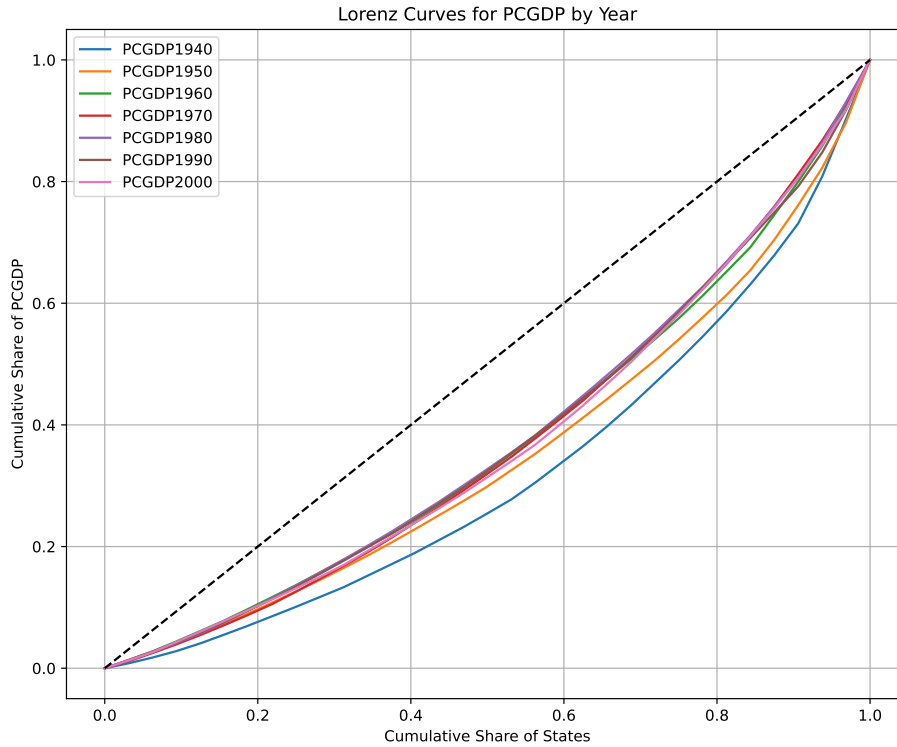


Figure 8: Lorenz Curves by Decade

```
plt.figure(figsize=(10, 8))

for yvar in yvars:
    incomes = gdf[yvar].values
    x, y = lorenz_curve(incomes)
    plt.plot(x, y, label=yvar)

# Plotting the line of equality
plt.plot([0, 1], [0, 1], color='black', linestyle='--')

# Adding titles and labels
plt.title('Lorenz Curves for PCGDP by Year')
plt.xlabel('Cumulative Share of States')
plt.ylabel('Cumulative Share of PCGDP')
plt.legend()
plt.grid(True)
plt.show()
```

[12]: Output in Figure 8

In the study of spatial inequality, it is important to note that all the measures mentioned above share a sixth property: *spatial invariance*. This means they are insensitive to the *geographical* distribution of the income values. The spatial invariance is demonstrated in Figure 9, where we compare two spatial distributions, one that is the actual distribution for 2000 and one where we artificially permute the incomes randomly. Despite markedly different spatial patterns, the income distributions summarized in the histograms in the bottom row are identical. Since all of the inequality measures mentioned only consider information about the statistical distribution, each measure will take on the same value whether applied to the spatial distribution on the left or right of the figure.

From a geographical perspective, spatial invariance is not a desirable property in an inequality measure. Let's now turn our attention to spatially explicit measures of inequality.

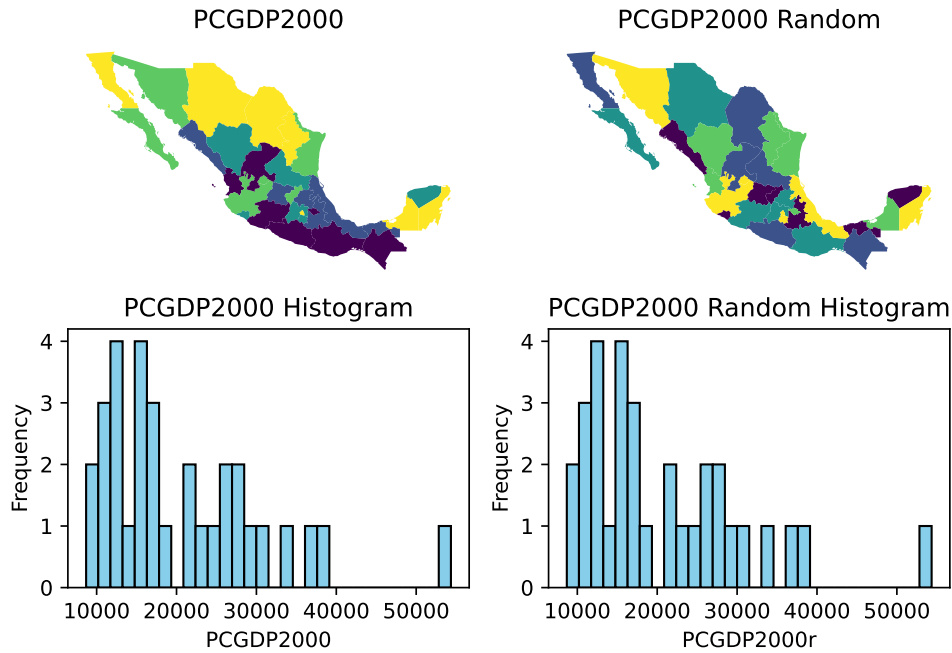


Figure 9: Spatial Invariance of Distributions

```
[13]: # Generate a 2x2 grid
# Choropleths in first row
# Income histograms in second row
fig, axs = plt.subplots(2, 2)

gdf.plot(column='PCGDP2000', ax=axs[0, 0], scheme='quantiles',
         cmap='viridis')
axs[0, 0].set_title('PCGDP2000')
axs[0, 0].axis('off')

gdf['PCGDP2000r'] = np.random.permutation(gdf.PCGDP2000)

gdf.plot(column='PCGDP2000r', ax=axs[0, 1], scheme='quantiles',
         cmap='viridis')
axs[0, 1].set_title('PCGDP2000 Random')
axs[0, 1].axis('off')

axs[1, 0].hist(gdf['PCGDP2000'], bins=30, color='skyblue',
              edgecolor='black')
axs[1, 0].set_title('PCGDP2000 Histogram')
axs[1, 0].set_xlabel('PCGDP2000')
axs[1, 0].set_ylabel('Frequency')

axs[1, 1].hist(gdf['PCGDP2000r'], bins=30, color='skyblue',
              edgecolor='black')
axs[1, 1].set_title('PCGDP2000 Random Histogram')
axs[1, 1].set_xlabel('PCGDP2000r')
axs[1, 1].set_ylabel('Frequency')

plt.tight_layout()

plt.show()
```

[13]: Output in Figure 9

5.3 Putting Space into the Measurement of Inequality

The inability of the inequality measures to capture any of the geographical dimensions of inequality stems from the treatment of the geographical units of measurement as individuals, and the desire to respect the principle of symmetry or anonymity in a classic inequality measure. This enables researchers to draw upon the wealth of knowledge about the properties of classic inequality measures, but at the cost of ignoring geography. In other words, these measures say a lot about inequality in the statistical distribution of incomes but they are silent on the spatial distribution of incomes.

In this section, we discuss the approaches used to integrate the spatial and statistical distributions in the study of spatial inequality. Given that these methods take the geographical distribution into account, they can be said to be *spatially explicit measures of inequality*.

These approaches are special cases of inequality decomposition. However, we highlight a key distinction between those approaches that use the concept of regional groupings to represent the geographical dimensions of inequality, and those that take a more comprehensive approach to introducing geography by considering the pair-wise spatial relationships between observations. We label the former category as regional inequality decomposition approaches, and the latter as spatial inequality decomposition methods.

5.3.1 Regional Inequality Decomposition

A common approach to introducing geography in the measurement of inequality leverages the fact that certain inequality measures can be *decomposed* if the individual income receiving units are placed into a set of mutually exclusive and exhaustive groups. The decomposition then identifies the overall inequality that is due to inequality within groups and between the groups.

To see how this works, we focus on the Theil inequality index (Theil 1967). The Theil inequality index is a measure of economic inequality derived from information theory, capturing the extent to which a distribution deviates from perfect equality. It is sensitive to differences across the distribution and can be decomposed into “within-group” and “between-group” components to analyze inequality at different levels.

More formally, let y_i be the income of unit i , and s_i represent the income share of unit i such that $s_i = \frac{y_i}{\sum_i y_i}$ and $\sum_i s_i = 1$. Then, consider the distribution of the shares. When all units have the same income $s_i = s_j = 1/n$. Then the entropy of the shares given as

$$H(y) = \sum_{i=1}^n s_i \ln \frac{1}{s_i}$$

will be maximized at $\ln n$. In the case of extreme inequality, all but one unit have $s_i = 0$ and a single unit has all income $s_j = 1$, and $H(y) = 0$. Thus, the entropy function can be viewed as an indicator of *income equality*.

To generate an indicator of *income inequality*, we can contrast an observed distribution’s equality against the maximum:

$$T(y) = \ln n - H(y) = \ln n - \sum_{i=1}^n s_i \ln \frac{1}{s_i} = \sum_{i=1}^n s_i \ln n s_i.$$

This can be viewed as a weighted average of the logarithmic deviations of the shares, with the weights defined as the shares. The logarithmic deviation of share i from perfect equality is $\ln \frac{s_i}{1/n} = \ln s_i n$.

Alternatively, the Theil index can be defined using relative incomes:

$$T = \frac{1}{n} \sum_i \frac{y_i}{\mu} \ln \frac{y_i}{\mu}$$

where y_i is the income of unit i and $\mu = \frac{1}{n} \sum_i y_i$.

Decomposition of the overall T measure requires assigning each unit to exactly one of G sets S_1, S_2, \dots, S_G , with the size of each set given as n_g so that:

$$\sum_{g=1}^G n_g = n.$$

Given this, we have:

$$Y_g = \sum_{i \in S_g} y_i \quad g = 1, \dots, G$$

Defining $\omega_g = \frac{n_g \bar{Y}_g}{n \mu}$, the Theil index can be rewritten as:

$$T = \sum_{g=1}^G \omega_g T_g + \sum_{g=1}^G \omega_g \ln \frac{\bar{Y}_g}{\mu}. \quad (1)$$

The first term is the within group inequality defined as a weighted average of the inequality within each group with the weights equal to the group's share of overall income, with:

$$T_g = \frac{1}{n_g} \sum_{i \in S_g} \frac{y_i}{\bar{Y}_g} \ln \frac{y_i}{\bar{Y}_g}.$$

The second term in Equation 1 is the between group inequality component, measuring the inequality that would exist if within each group there was no inequality (i.e., all members of the same set have the same income).

Decomposition of inequality had been widely applied in economics to study inequality between occupational groups, sexes, and races, where individuals would be placed into the mutually exclusive groups and overall individual inequality decomposed into that due to the differences between and within the groups. It was a short jump to adopt this to spatial inequality by using regions to define the groups, with individual units (in our case, states) being assigned to one and only one region.⁴

We will illustrate this for the case of Mexican states using a regional partition due to [Hanson \(1996\)](#) as shown in Figure 10. This regionalization scheme consists of 5 regions, with the size of the regions ranging from 2 states to 10 states.

We can apply the `theil` module from `pysal-inequality` to calculate the value of the overall level of inequality as measured by the global Theil index, and its decomposition into the between region inequality and within region inequality components:

```
[14]: # Produce 2x2 Grid
# First row: spatial distribution and random distribution
# Second row: another random distribution and density of
# polarization values
fig, axs = plt.subplots(2, 2)
from inequality.theil import TheilDSim
np.random.seed(12345)

gdf.plot('PCGDP2000', ax=axs[0, 0], scheme='quantiles',
         cmap='viridis')
axs[0, 0].set_title('PCGDP2000')
axs[0, 0].axis('off')

# Extract the per capita GDP and regimes
income = gdf['PCGDP2000']
regimes = gdf['HANSON98']

res = TheilDSim(income, regimes, 999)

gdf['PCGDP2000r'] = np.random.permutation(gdf.PCGDP2000)
gdf.plot(column='PCGDP2000r', ax=axs[0, 1], scheme='quantiles',
         cmap='viridis')
axs[0, 1].set_title('PCGDP2000 Random')
```

⁴One of the earliest applications of decomposition for regional inequality analysis is [Theil \(1967\)](#).

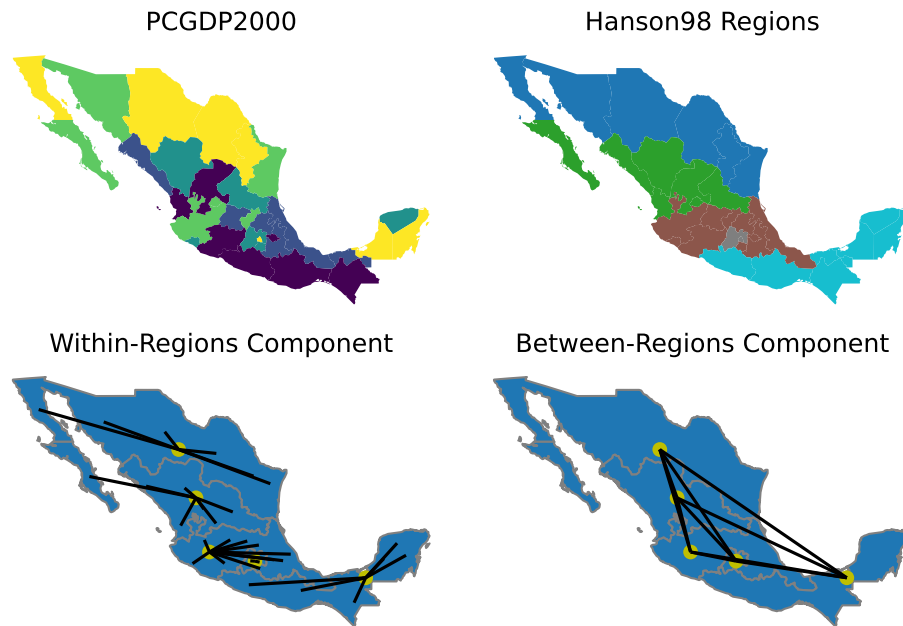


Figure 10: Regional Inequality Decomposition

```

axes[0, 1].axis('off')

gdf['PCGDP2000r'] = np.random.permutation(gdf.PCGDP2000)
gdf.plot(column='PCGDP2000r', ax=axes[1, 0], scheme='quantiles',
         cmap='viridis')
axes[1, 0].set_title('PCGDP2000 Random')
axes[1, 0].axis('off')

msg=f'Spatial polarization: {res.bg[0][0]/res.T:.3f}'
msg=f'{msg}, pseudo p-value: {res.bg_pvalue[0]}'
print(msg)
realizations = np.array([t.bg/t.T for t in res.results])
print(f'Ho mean: {realizations.mean():.3f}')

kde = sns.kdeplot(realizations, fill=False, color='blue', ax=axes[1,1])
x, y = kde.get_lines()[0].get_data()
plt.legend([], [], frameon=False)
# Fill the area to the right of the specified value
plt.fill_between(x, y, where=(x >= realizations[0]),
               interpolate=True, color='red', alpha=0.5)

# Add vertical line at the specified value
plt.axvline(x=realizations[0], color='red', linestyle='--')
plt.xlabel("Spatial Polarization")
print(f'Theil: {res.T}')
print(f'KB p-value: {(realizations >= realizations[0]).sum()/1000}')

```

```

[14]: Spatial polarization: 0.341, pseudo p-value: 0.036
      Ho mean: 0.139
      Theil: 0.10660832349588023
      KB p-value: 0.036

```

Graphical output in Figure 11

The overall level of regional inequality is 0.106. The between region component stands at 0.036, while the within region element is 0.070. In relative terms, inequality between the regions in Mexico accounts for 34 percent of state income inequality, while inequality between states from the same region is the larger share presenting 72 percent of spatial inequality.

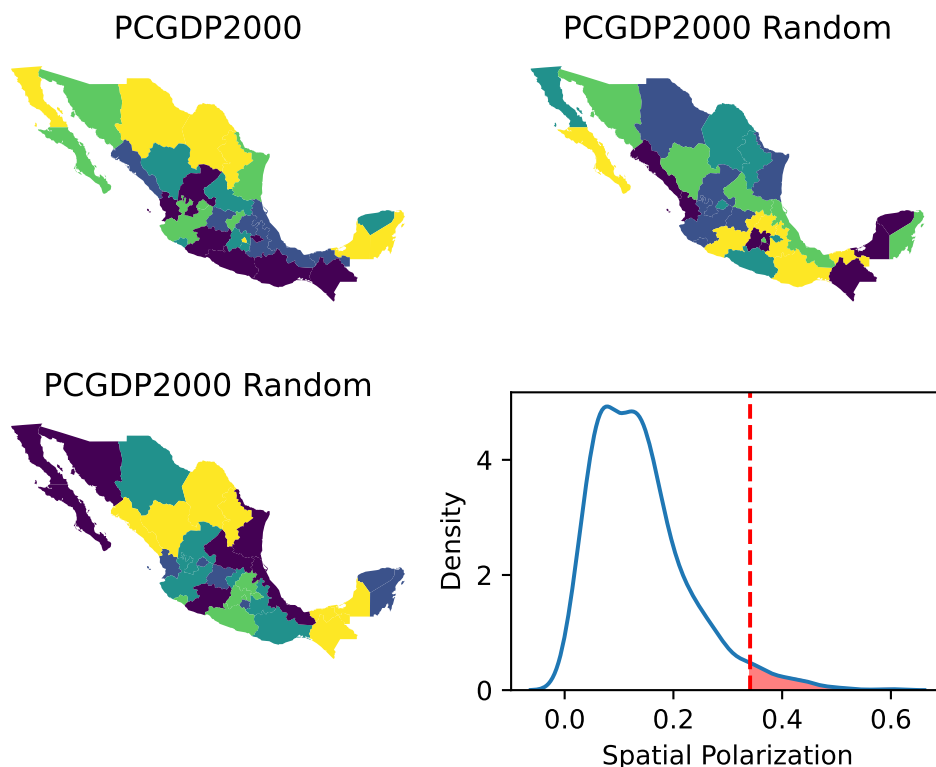


Figure 11: Spatial Polarization

The ratio of between to within region inequality has been suggested as a measure of *spatial polarization* (Zhang, Kanbur 2001). Since the two components sum to overall inequality, we can re-express the measure of spatial polarization as the ratio of between region to overall inequality. Thus, the level of spatial polarization of state incomes stood at 34 percent in 2000 in Mexico.

A relevant question here is whether this level of spatial polarization in 2000 should be considered high or not? Another way to express this, is to ask if the polarization is higher than we would expect if state incomes were randomly distributed in space in Mexico in 2000.

We can answer this question by developing counter-factual spatial distributions that reflect a null hypothesis of spatially random income distributions. Given the n observations, in our case states, there are $n!$ permutations of incomes that are equally likely. In Figure 11, the maps in the top-right and lower-left are two such realizations, where the observed incomes from 2000 (top-left) have been randomly reassigned to states.

Since it is not feasible to generate all $32!$ maps in our case,⁵ we take a sample of 999 such maps from the distribution of permutations. For each of these maps we calculate the spatial polarization measure, and collect all measures to develop a reference distribution for our index under the null of spatially random income distributions. We evaluate our observed spatial polarization index against this distribution and derive a pseudo p-value as the number of counterfactual distributions that generate polarization levels as large as the observed value over the number of permutations plus one.

The reference distribution for the polarization index is reported in the bottom right of Figure 11. The area to the right of our observed polarization index of 0.34 is 0.036 of the distribution. By convention, this p-value points to a significant departure from the null, and we would conclude that the level of spatial polarization of incomes in Mexico is significantly different from that expected were incomes generated by a spatially random process.

The regional decomposition is a spatially explicit measure of spatial inequality as it is

⁵ $32! = 2.6313084e + 35$.

indeed sensitive to the geographical distribution of the income values. It is important to note, however, that it is the spatial polarization index, and not the overall T that is spatially explicit. Each of the counterfactual spatial distributions used to construct the reference distribution for the spatial polarization index would still have the same attribute distribution, with similar means and variances, and levels of overall inequality.

While the regional inequality decomposition is spatially explicit, it treats geography in an aggregated fashion. There are, in essence, two scales of spatial inequality implied in this decomposition. As displayed in Figure 10, the within-regions component might be considered the “local” measure as it compares incomes belonging to the same region to the mean income of the group. The between-regions inequality component, by contrast, views spatial inequality in a more aggregate fashion considering only the differences in regional means.

A close inspection of the decomposition Equation 1 reveals both the within and between inequality components are functions of a T index. In the former case this is applied to states from the same region, and in the latter case, the T is applied to the means of regional incomes. For the within region component, the actual spatial distribution of the member states within each of their regions is ignored. By the same token, the geographical location of the regions is immaterial to the calculation of the between region component. In other words, once the states have been assigned to regions, geography no longer matters. Thus, we draw a distinction between *regional* inequality decomposition and *spatial* inequality decomposition.

5.3.2 Spatial Inequality Decomposition

To capture these ignored dimensions of the spatial distribution, Rey, Smith (2013) suggested a spatial decomposition of the Gini coefficient. Starting from the Gini coefficient expressed as half the relative mean absolute deviation:⁶

$$Gini = \frac{\sum_{i=1}^n \sum_{j=1}^n |y_i - y_j|}{2n^2 \bar{y}} \quad (2)$$

the sum of the absolute deviations can be decomposed as:

$$\sum_{i=1}^n \sum_{j=1}^n |y_i - y_j| = \sum_{i=1}^n \sum_{j=1}^n w_{i,j} |y_i - y_j| + (1 - w_{i,j}) \sum_{i=1}^n \sum_{j=1}^n |y_i - y_j| \quad (3)$$

where $w_{i,j} = 1$ if states i and j are geographical neighbors, $w_{i,j} = 0$ otherwise. Here we define neighbors as states that share a border.

Substituting Equation 3 into Equation 2 gives:

$$Gini = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{i,j} |y_i - y_j|}{2n^2 \bar{y}} + \frac{\sum_{i=1}^n \sum_{j=1}^n (1 - w_{i,j}) |y_i - y_j|}{2n^2 \bar{y}} \quad (4)$$

The first term represents a measure of inequality between neighboring states, while the second term captures inequality between “distant” pairs of states. For most spatial configurations, the number of neighboring pairs will be dwarfed by the number of pairs of states that are distant. So while, in the case of spatially clustered income values, the expectation would be for the average difference in incomes to be smaller for neighboring rather than distant states, our measure of spatial clustering here has to take into account the different cardinality of the two sets of pairs. As such, the relevant comparison is if the first term (second term) is smaller (larger) than what could be expected if state incomes were randomly distributed in space.

We apply the Spatial Gini Decomposition to Mexican State incomes in 2000 using `pysal-inequality` in Figure 12. The adjacency graph based on the criterion of Queen neighbors is shown in the top-right figure. An edge defines a pair of neighboring states.⁷

⁶It is sometimes stated that the maximum value of the Gini in this form is 1 (e.g., Wang et al. 2024). This is technically incorrect, as in this form, the Gini has a range of $[0, (n-1)/n]$. As n grows larger, the upper bound approaches 1. Moreover, because the upper bound is a function of n , care should be taken when using the Gini in this form when comparing distributions of different sizes.

⁷Two units are Queen neighbors if their borders share at least one vertex.

```
[15]: # Produce 2x2 Grid
# First row: spatial distribution and neighbor graph
# Second row: random distribution and density of
# Spatial Gini values
fig, axs = plt.subplots(2, 2)
from inequality.gini import Gini_Spatial
import libpysal
np.random.seed(12345)

gdf.plot('PCGDP2000', ax=axs[0, 0], scheme='quantiles', cmap='viridis')
axs[0, 0].set_title('PCGDP2000')
axs[0, 0].axis('off')

wq = libpysal.weights.Queen.from_dataframe(gdf)
wq.transform = 'B'
gs2000 = Gini_Spatial(gdf["PCGDP2000"], wq)

income = gdf['PCGDP1940']

gdf.plot(ax=axs[0, 1])
axs[0, 1].set_title('Neighbor Graph')
axs[0, 1].axis('off')

wq.plot(gdf, ax=axs[0,1])

axs[1, 0].set_title('Counterfactual')
axs[1, 0].axis('off')
gdf['PCGDP2000r'] = np.random.permutation(gdf.PCGDP2000)
gdf.plot(column='PCGDP2000r', ax=axs[1,0], scheme='quantiles',
         cmap='viridis')

adsum = gs2000.dtotal
realizations = gs2000.wcgp / adsum
kde = sns.kdeplot(realizations, fill=False, color='blue', ax=axs[1,1])
x, y = kde.get_lines()[0].get_data()
plt.legend([], [], frameon=False)
# Fill the area to the right of the specified value
plt.fill_between(x, y, where=(x >= gs2000.wcgp/adsum),
                interpolate=True, color='red', alpha=0.5)

# Add vertical line at the specified value
plt.axvline(x=gs2000.wcgp/adsum, color='red', linestyle='--')
plt.xlabel("Spatial Gini")

plt.tight_layout()
plt.show()
G = gs2000
msg = f'Expected Distant SADS/Total SADS: {G.e_wcg/adsum:.2f}'
msg = f'{msg}, Observed: {G.wcg/adsum:.2f}\n p-value: {G.p_sim}'
print(msg)
```

```
[15]: Expected Distant SADS/Total SADS: 0.86, Observed: 0.90
p-value: 0.01
```

Graphical output in Figure 12

For inference on the spatial Gini, the same computationally based approach that we used in the Theil decomposition is employed. One of the counterfactuals representing a random permutation of the incomes is shown in the lower-left figure. The reference distribution for the spatial Gini index is shown on the bottom right. Here the spatial Gini is expressed as the share of the overall absolute pair-wise differences due to inequality between distant (non-neighbor) pairs of states. The pseudo p-value (0.01) for the observed index is calculated as the area to its right under the distribution.

Under the null, the distant pairs should account for 86 percent of the absolute differences, however, the observed share is much higher at 90 percent. In fact, the p-value

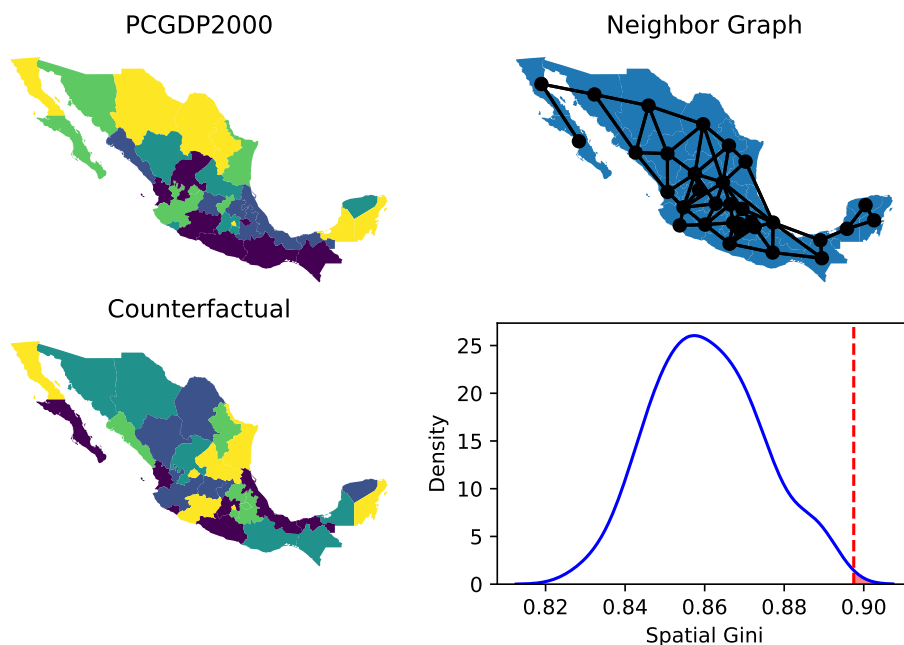


Figure 12: Spatial Inequality Decomposition

indicates that none of the counterfactual spatial distributions of income generated a spatial Gini as large as the one observed. In other words, the inequality between distant pairs of states is larger than the inequality between neighboring states. This pairwise orientation of the spatial Gini decomposition offers a useful complement over the regional decomposition of the Theil approach, as it introduces a more spatially explicit view of the income distribution that demonstrates how spatial autocorrelation affects overall inequality across the states.⁸

Spatial autocorrelation refers to the degree to which a spatial variable is correlated with itself across geographical space. It measures whether similar or dissimilar values of a variable tend to cluster together spatially. Positive spatial autocorrelation indicates that similar values (e.g., areas with high incomes) are located near each other, while negative spatial autocorrelation suggests that dissimilar values (e.g., areas with high and low incomes) are spatially proximate. Spatial autocorrelation is crucial in studying spatial inequality because it reveals the extent to which inequality is spatially patterned, reflecting processes like segregation, clustering of poverty or wealth, and the spatial diffusion of economic opportunities or disadvantages.

Identifying and quantifying spatial autocorrelation allows researchers to understand the spatial dimensions of inequality, assess the effectiveness of place-based policies, and model spatial processes more accurately (Anselin 1995). For example, regions with high positive spatial autocorrelation of income inequality may require targeted regional policies to address concentrated disadvantage. Without considering spatial autocorrelation, analyses of inequality risk overlooking critical spatial dependencies that shape social and economic outcomes. Moreover, the presence of spatial autocorrelation implies that inference on inequality measures that rely on the assumption of random sampling may be misleading. This is because the autocorrelation violates the assumption of independence underlying the random sampling assumption.

While the spatial Gini decomposition does take spatial autocorrelation into account, it does not allow for the exact additivity of within-group and between-group inequality components. This occurs because the Gini index depends on the degree of overlap in incomes among states from different regions. Any overlap complicates the decomposition, as part of the total inequality arises from the overlap itself, which cannot be clearly attributed to either within-group or between-group inequality. As a result, the decomposition requires

⁸For a recent extension of the spatial Gini decomposition see Panzera, Postiglione (2020).

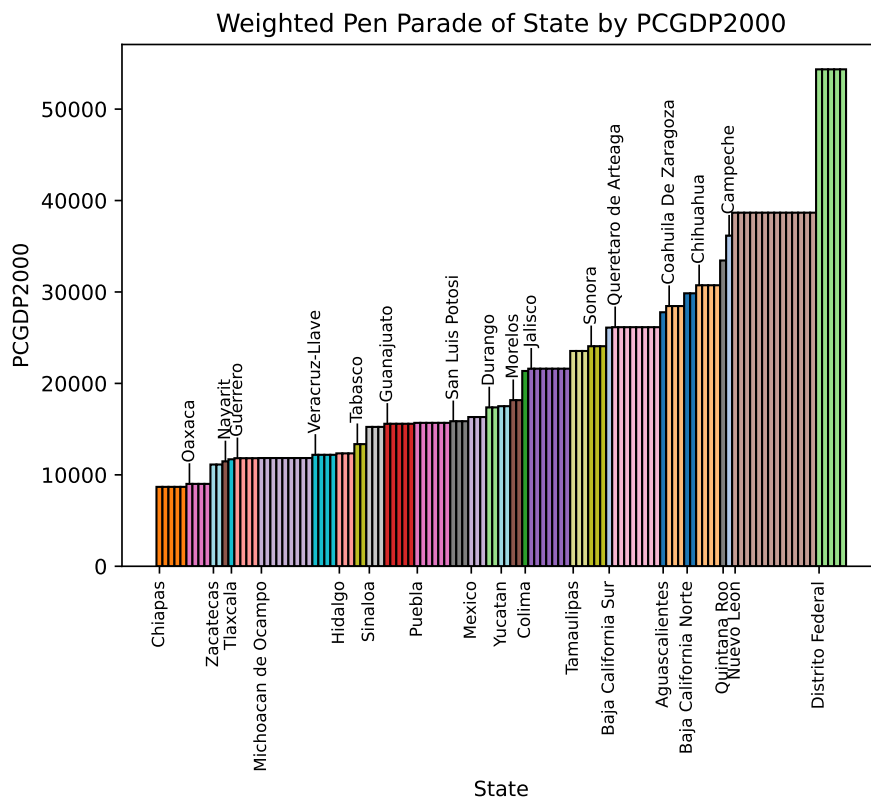


Figure 13: Populated Weighted Pen Parade

a residual term to account for the overlap. Since this residual term is difficult to interpret, researchers rarely use the Gini index for regional inequality decomposition.

5.4 Weighted or Unweighted Inequality: Places versus People

A final issue we explore is the question of whether the measure of spatial inequality should take into account the population sizes of the enumeration units. This was briefly mentioned earlier in the context of measuring international inequality. In the regional literature, a debate rages as to whether population weighted or unweighted approaches should be used to measure spatial inequality (Gluschenko 2018).

To frame the debate, it is helpful to consider three different concepts of inequality at the international scale suggested by Milanović (2007). Here we adapt them to the question of measuring spatial inequality at the sub-national scale. Concept 1 is unweighted spatial inequality, where each state is one unit of measurement, and we use its per capita income irrespective of the state's population. In addition to being the dominant approach in regional inequality analysis, this concept is at the core of the literature on regional convergence (Rey, Montouri 1999) where the focus is on whether the incomes of poor and rich states in a system are coming together or growing apart over time.

Concept 2 takes into account the population of the individual states, recognizing that a state like Nuevo Leon with a population of 8.6 million in 2000 having its average income increase by 10,000 USD is likely to have a larger impact than is seeing the per capita income of Colima, with a population of 500,000 change by the same amount. So here, the per capita incomes of the states are now weighted by the population of the states. This is weighted spatial inequality.

Finally, Concept 3, measures inequality between all the individuals in the country. Here we would require information on the individuals both in terms of their incomes and their state of residence. Were it available, such data would allow us to measure personal income inequality. However, in the literature on regional inequality, such data is scarce, and so this concept is not operationalized.

This means the debate in the regional literature is between proponents of Concept 1 and Concept 2.

Thus far in this paper, we have adopted the Concept 1 definition of spatial inequality, that is, unweighted spatial inequality. We can contrast this with the perspective offered by Concept 2, by developing the Weighted Pen Parade under concept 2 shown in Figure 13 and comparing it with the Unweighted Pen Parade from Figure 3. The number of bars for each state in the Weighted Pen Parade is proportional to the population of that state. The logic behind the weighted spatial inequality is that rather than having one representative observation (individual) from the state with their level of income equal to the state per capita income, now there are a proportional number of individuals from each state having their level of income equal to the state's per capita income.⁹

A fundamental issue with this approach is that it assumes that inequality among individuals within the state is zero, since all members have the same level of income. This also implies that for two states that have different levels of per capita income, the poorest members of the richer state will be richer than the richest member of the poorer state. This is at odds with empirical reality.

6 Conclusion

Spatial inequality continues to attract the attention of researchers and policy makers alike. This paper has presented the key methodological approaches available to researchers interested in analyzing spatial disparities. By highlighting the challenges posed in adapting classical inequality measures to the question of spatial inequality, the paper draws important distinctions between different types of inequality decomposition approaches based on their treatment of space. These methods reveal the spatial dimensions of inequality, particularly the extent to which income disparities are geographically clustered.

While the focus has been on regional inequality at the sub-national scale, these methods have a broad scope that researchers can apply at different spatial scales, from analyzing geographical disparities at the international level (Redding, Venables 2004), inter-regional (Bathelt et al. 2024), as well as intra-urban scale (OECD 2018). This scope is vital, as the mechanisms of inequality can differ at each scale. For example, trade policy likely plays a more significant role on an international scale. At the same time, industrial restructuring is more influential on an inter-regional scale, and residential sorting is operative on an intra-urban scale. Furthermore, the resulting patterns of spatial inequality may also vary across these scales. The work by Ganong, Shoag (2017) shows a substantial divergence of incomes across states but reports a more mixed pattern when measuring convergence using labor market areas.

Linking the spatial inequality measures to potential policy interventions offers some interesting possibilities. For example, the distinction between inter-regional (between-region) and intra-regional (within-region) inequality in the decomposition may help researchers and policymakers design such interventions in a tailored fashion. If between-region inequality dominates, there may be a strong case for place-based policies. Conversely, if the within-region inequality component accounts for the majority share, more nationally oriented policies, such as federal tax laws, may be more appropriate.

In future research, it would be beneficial to investigate the relationship between spatial inequality and place mobility. Place mobility refers to the economic trajectory of a location within the national context, similar to the concept of inter-generational income mobility (Rey, Casimiro Vieyra 2023). Exploring the link between place mobility and spatial polarization is crucial, especially how the movement of states within the income distribution impacts broader regional inequalities. Moreover, gaining insight into the interaction between place mobility and "place-based policies" could help in designing more effective regional development strategies that reduce inequality while fostering economic mobility.

⁹A referee suggested that if data on individual incomes were available one could use a pentagram for representing within region/state/city inequality to complement Figure 13. In some cases, such as the Netherlands, access to micro level data may be possible under strict conditions.

References

- Anselin L (1995) Local indicators of spatial association-LISA. *Geographical Analysis* 27: 93–115. [CrossRef](#)
- Atkinson AB (1970) On the measurement of inequality. *Journal of Economic Theory* 2: 244–263. [CrossRef](#)
- Barber M, Holbein JB (2022) 400 million voting records show profound racial and geographic disparities in voter turnout in the United States. *PLoS ONE* 17: e0268134. [CrossRef](#)
- Bathelt H, Buchholz M, Storper M (2024) The nature, causes, and consequences of inter-regional inequality. *Journal of Economic Geography* 24: 353–374. [CrossRef](#)
- Brandolini A, Smeeding TM (2011) *Income Inequality in Richer and OECD Countries*. Oxford University Press. [CrossRef](#)
- Cowell FA (2011) *Measuring Inequality* (3rd ed.). LSE Perspectives in Economic Analysis. Oxford University Press, Oxford
- Deb Nath N, Odoi A (2024) Geographic disparities and temporal changes of diabetes-related mortality risks in Florida: A retrospective study. *PeerJ* 12: e17408. [CrossRef](#)
- Frank MW (2009) Inequality and growth in the United States: Evidence from a new state-level panel of income inequality measures. *Economic Inquiry* 47: 55–68. [CrossRef](#)
- Ganong P, Shoag D (2017) Why has regional income convergence in the U.S. declined? *Journal of Urban Economics* 102: 76–90. [CrossRef](#)
- Gaubert C, Kline P, Vergara D, Yagan D (2021) Trends in US Spatial Inequality: Concentrating Affluence and a Democratization of Poverty. *AEA Papers and Proceedings* 111: 520–525. [CrossRef](#)
- Gluschenko K (2018) Measuring regional inequality: To weight or not to weight? *Spatial Economic Analysis* 13: 36–59. [CrossRef](#)
- Graetz N, Woyczynski L, Wilson KF, Hall JB, Abate KH, Abd-Allah F, Adebayo OM, Adekanmbi V, Afshari M, Ajumobi O, Akinyemiju T, Alahdab F, Al-Aly Z, Rabanal JEA, Alijanzadeh M, Alipour V, Altirkawi K, Amiresmaili M, Anber NH, Andrei CL, Anjomshoa M, Antonio CAT, Arabloo J, Aremu O, Aryal KK, Asadi-Aliabadi M, Atique S, Ausloos M, Awasthi A, Quintanilla BPA, Azari S, Badawi A, Banoub JAM, Barker-Collo SL, Barnett A, Bedi N, Bennett DA, Bhattacharjee NV, Bhattacharyya K, Bhattacharai S, Bhutta ZA, Bijani A, Bikbov B, Britton G, Burstein R, Butt ZA, Cárdenas R, Carvalho F, Castañeda-Orjuela CA, Castro F, Cerin E, Chang JC, Collison ML, Cooper C, Cork MA, Daoud F, Das Gupta R, Weaver ND, De Neve JW, Deribe K, Desalegn BB, Deshpande A, Desta M, Dhimal M, Diaz D, Dinberu MT, Djalalinia S, Dubey M, Dubljanin E, Durães AR, Dwyer-Lindgren L, Earl L, Kalan ME, El-Khatib Z, Eshrati B, Faramarzi M, Fareed M, Faro A, Fereshtehnejad SM, Fernandes E, Filip I, Fischer F, Fukumoto T, García JA, Gill PS, Gill TK, Gona PN, Gopalani SV, Grada A, Guo Y, Gupta R, Gupta V, Haj-Mirzaian A, Haj-Mirzaian A, Hamadeh RR, Hamidi S, Hasan M, Hassen HY, Hendrie D, Henok A, Henry NJ, Prado BH, Herteliu C, Hole MK, Hossain N, Hosseinzadeh M, Hu G, Ilesanmi OS, Irvani SSN, Islam SMS, Izadi N, Jakovljevic M, Jha RP, Ji JS, Jonas JB, Shushtari ZJ, Jozwiak JJ, Kanchan T, Kasaeian A, Karyani AK, Keiyoro PN, Kesavachandran CN, Khader YS, Khafaie MA, Khan EA, Khater MM, Kiadaliri AA, Kiirithio DN, Kim YJ, Kimokoti RW, Kinyoki DK, Kisa A, Kosen S, Koyanagi A, Krishan K, Defo BK, Kumar M, Kumar P, Lami FH, Lee PH, Levine AJ, Li S, Liao Y, Lim LL, Listl S, Lopez JCF, Majdan M, Majdzadeh R, Majeed A, Malekzadeh R, Mansournia MA, Martins-Melo FR, Masaka A, Massenbourg BB, Mayala BK, Mehta KM, Mendoza W, Mensah GA, Meretoja TJ, Mestrovic T, Miller TR, Mini GK, Mirrakhimov EM, Moazen B, Mohammad DK, Darwesh AM, Mohammed S, Mohebi F, Mokdad AH, Monasta L, Moodley Y, Moosazadeh M, Moradi G, Moradi-Lakeh M, Moraga P, Morawska L, Morrison SD, Mosser JF, Mousavi SM,

- Murray CJL, Mustafa G, Nahvijou A, Najafi F, Nangia V, Ndwandwe DE, Negoi I, Negoi RI, Ngunjiri JW, Nguyen CT, Nguyen LH, Ningrum DNA, Noubiap JJ, Shiadeh MN, Nyasulu PS, Ogbo FA, Olagunju AT, Olusanya BO, Olusanya JO, Onwujekwe OE, Ortega-Altamirano DDV, Ortiz-Panozo E, Øverland S, P. A. M, Pana A, Panda-Jonas S, Pati S, Patton GC, Perico N, Pigott DM, Pirsahab M, Postma MJ, Poursahms A, Prakash S, Puri P, Qorbani M, Radfar A, Rahim F, Rahimi-Movaghar V, Rahman MHU, Rajati F, Ranabhat CL, Rawaf DL, Rawaf S, Reiner RC, Remuzzi G, Renzaho AMN, Rezaei S, Rezapour A, Rios-González C, Roever L, Ronfani L, Roshandel G, Rostami A, Rubagotti E, Sadat N, Sadeghi E, Safari Y, Sagar R, Salam N, Salamati P, Salimi Y, Salimzadeh H, Samy AM, Sanabria J, Santric Milicevic MM, Sartorius B, Sathian B, Sawant AR, Schaeffer LE, Schipp MF, Schwebel DC, Senbeta AM, Sepanlou SG, Shaikh MA, Shams-Beyranvand M, Shamsizadeh M, Sharafi K, Sharma R, She J, Sheikh A, Shigematsu M, Siabani S, Silveira DGA, Singh JA, Sinha DN, Skirbekk V, Sligar A, Sobaih BH, Soofi M, Soriano JB, Soyiri IN, Sreeramareddy CT, Sudaryanto A, Babale Sufiyan M, Sutradhar I, Sylaja PN, Tabarés-Seisdedos R, Tadesse BT, Temsah MH, Terkawi AS, Tessema B, Tessema ZT, Thankappan KR, Topor-Madry R, Tovani-Palone MR, Tran BX, Car LT, Ullah I, Uthman OA, Valdez PR, Veisani Y, Violante FS, Vlassov V, Vollmer S, Thu Vu G, Waheed Y, Wang YP, Wilkinson JC, Winkler AS, Local Burden of Disease Educational Attainment Collaborators (2020) Mapping disparities in education across low- and middle-income countries. *Nature* 577: 235–238. [CrossRef](#)
- Hall RE (1978) Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence. *Journal of Political Economy* 86: 971–987. [CrossRef](#)
- Hanson GH (1996) *U.S.-Mexico Integration and Regional Economies*. National Bureau of Economic Research, Cambridge. [CrossRef](#)
- Jordahl K, Van Den Bossche J, Wasserman J, McBride J, Gerard J, Tratner J, Perry M, Farmer C, Cochran M, Gillies S, Bartos M, Culbertson L, Eubank N, Maxalbert, Fleischmann M, Hjelle GA, Arribas-Bel D, Ren C, Rey S, Journois M, Wolf LJ, Bilogur A, Grue N, Wilson J, YuichiNotoya, Wasser L, Filipe, Holdgraf C, Greenhall A, Trengrove J (2019) Geopandas/geopandas: V0.4.1. ZENODO software repository. [CrossRef](#)
- Kanbur S, Venables A (2005) *Spatial Inequality and Development*. Oxford University Press, Oxford, UK. [CrossRef](#)
- Khedmati Morasae E, Derbyshire DW, Amini P, Ebrahimi T (2024) Social determinants of spatial inequalities in COVID-19 outcomes across England: A multiscale geographically weighted regression analysis. *SSM - Population Health* 25: 101621. [CrossRef](#)
- Knaap E (2017) The Cartography of Opportunity: Spatial Data Science for Equitable Urban Policy. *Housing Policy Debate* 27: 913–940. [CrossRef](#)
- Ledić M, Rubil I, Urban I (2023) Tax progressivity and social welfare with a continuum of inequality views. *International Tax and Public Finance* 30: 1266–1296. [CrossRef](#)
- Milanović B (2007) *Worlds Apart: Measuring International and Global Inequality*. Princeton University Press, Princeton (N.J.). [CrossRef](#)
- Milanović B (2018) *Global Inequality: A New Approach for the Age of Globalization* (First Harvard University Press paperback ed.). The Belknap Press of Harvard University Press, Cambridge, Massachusetts London, England
- Nijman J, Wei YD (2020) Urban inequalities in the 21st century economy. *Applied Geography* 117: 102188. [CrossRef](#)
- OECD (2018) *Divided Cities: Understanding Intra-urban Inequalities*. OECD. [CrossRef](#)
- Overman HG, Xu X (2022) Spatial disparities across labour markets. Institute for Fiscal Studies, London, UK. <https://ifs.org.uk/books/spatial-disparities-across-labour-markets>

- Panzer D, Postiglione P (2020) Measuring the Spatial Dimension of Regional Inequality: An Approach Based on the Gini Correlation Measure. *Social Indicators Research* 148: 379–394. [CrossRef](#)
- Partridge JS, Partridge MD, Rickman DS (1998) State patterns in family income inequality. *Contemporary Economic Policy* 16: 277–294. [CrossRef](#)
- Pen J (1971) *Income distribution*. Allen Lane, London
- Piketty T (2014) *Capital in the Twenty-first Century*. Harvard University Press. [CrossRef](#)
- Piketty T, Saez E (2003) Income Inequality in the United States, 1913–1998. *The Quarterly Journal of Economics* 118: 1–39. [CrossRef](#)
- Redding S, Venables AJ (2004) Economic geography and international inequality. *Journal of International Economics* 62: 53–82. [CrossRef](#)
- Rey SJ (2015) Discrete regional distributional dynamics revisited. *Revue d'Économie Régionale & Urbaine* mai: 83–104. [CrossRef](#)
- Rey SJ, Anselin L, Amaral P, Arribas-Bel D, Cortes RX, Gaboardi JD, Kang W, Knaap E, Li Z, Lumnitz S, Oshan TM, Shao H, Wolf LJ (2022) The PySAL Ecosystem: Philosophy and Implementation. *Geographical Analysis* 54: 467–487. [CrossRef](#)
- Rey SJ, Arribas-Bel D, Wolf LJ (2023) *Geographic Data Science with Python*. Chapman & Hall/CRC Texts in Statistical Science. CRC Press, Boca Raton. [CrossRef](#)
- Rey SJ, Casimiro Vieyra E (2023) Spatial inequality and place mobility in Mexico: 2000–2015. *Applied Geography* 152: 102871. [CrossRef](#)
- Rey SJ, Montouri BD (1999) US regional income convergence: A spatial econometric perspective. *Regional Studies* 33: 143–156. [CrossRef](#)
- Rey SJ, Sastré-Gutiérrez ML (2010) Interregional inequality dynamics in Mexico. *Spatial Economic Analysis* 5: 277–298. [CrossRef](#)
- Rey SJ, Smith RJ (2013) A spatial decomposition of the Gini coefficient. *Letters in Spatial and Resource Sciences* 6: 55–70. [CrossRef](#)
- Roemer JE (1998) *Equality of Opportunity*. Harvard Univ. Press, Cambridge, Mass. [CrossRef](#)
- Saez E, Zucman G (2016) Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data *. *The Quarterly Journal of Economics* 131: 519–578. [CrossRef](#)
- Sala-i-Martin X (2006) The World Distribution of Income: Falling Poverty and... Convergence, Period. *The Quarterly Journal of Economics* 121: 351–397. [CrossRef](#)
- Sarkar S, Cottineau-Mugadza C, Wolf LJ (2024) Spatial inequalities and cities: A review. *Environment and Planning B: Urban Analytics and City Science* 51: 1391–1407. [CrossRef](#)
- Sen A (2004) *Inequality Reexamined* (Reprint ed.). Oxford Univ. Press, New York. [CrossRef](#)
- Suss J, Kemeny T, Connor DS (2024) GEOWEALTH-US: Spatial wealth inequality data for the United States, 1960–2020. *Scientific Data* 11: 253. [CrossRef](#)
- Theil H (1967) *Economics and Information Theory*. North Holland, Amsterdam
- Tine M (2017) Growing up in Rural vs. Urban Poverty: Contextual, Academic, and Cognitive Differences. In: *Poverty, Inequality and Policy*. IntechOpen. [CrossRef](#)
- U.S. Department of Commerce (2023) Geographic Inequality on the Rise in the U.S. Regional economic research initiative blog

- Venter ZS, Figari H, Krange O, Gundersen V (2023) Environmental justice in a very green city: Spatial inequality in exposure to urban nature, air pollution and heat in Oslo, Norway. *Science of The Total Environment* 858: 160193. [CrossRef](#)
- Wang J, Kwan MP, Xiu G, Deng F (2024) A robust method for evaluating the potentials of 15-minute cities: Implications for sustainable urban futures. *Geography and Sustainability*: S2666683924000646. [CrossRef](#)
- Waskom M (2021) Seaborn: Statistical data visualization. *Journal of Open Source Software* 6: 3021. [CrossRef](#)
- Wei R, Rey S, Knaap E (2020) Efficient regionalization for spatially explicit neighborhood delineation. *International Journal of Geographical Information Science*: 1–17. [CrossRef](#)
- Widuto A (2019) Regional inequalities in the EU. European parliamentary research service, [https://www.europarl.europa.eu/RegData/etudes/BRIE/2019/637951/EPRS_BRI\(2019\)637951_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2019/637951/EPRS_BRI(2019)637951_EN.pdf)
- Zhang X, Kanbur R (2001) What difference do polarisation measures make?: An application to China. *Journal of Development Studies* 37: 85–98. [CrossRef](#)

