

HUGE

Hello.

March 15, 2016

# Will urban kale farmers invade your neighborhood?

Huge

## *Agenda*



- 1. What is gentrification?**
- 2. How is it measured?**
- 3. Is there a better way?**
- 4. What about open data sources?**
- 5. What next?**

*This talk is not:*

- 1. A debate on how to correctly define gentrification.**
- 2. A high level data science lesson.**
- 3. A discussion about when your neighborhood did or did not gentrify.**
- 4. A conversation about kale, urban farming, or bee keeping. Sorry. 😞**

*Because honestly ...*

**I only have one hour.**

**And over 500 slides to share.**

**Just kidding.**



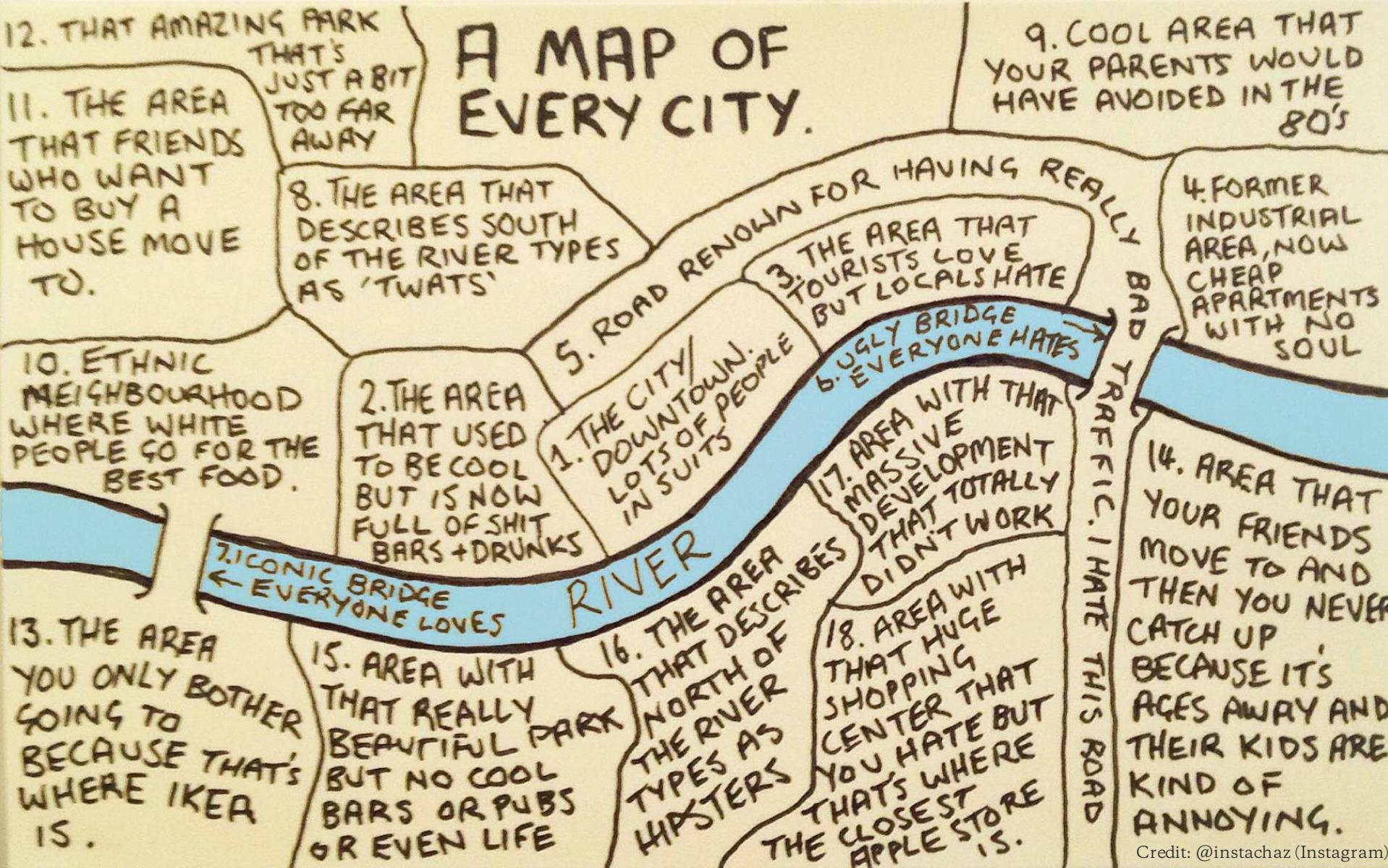
Here's an urban bee keeper.



And a rooftop farm.

**What is  
gentrification?**

# A MAP OF EVERY CITY.







ATELIER  
(STUDIO  
FOR  
RENT)

# *Denizens of 'Sesame Street' Adapt to an Upscale TV Neighborhood*

By EMILY STEEL JAN. 12, 2016



 SLIDE SHOW | 11 Photos

Inside the HBO Set of 'Sesame Street'

*What is gentrification?*

**An example.**

# Google Street View.

315 Bowery, New York, NY

A photograph of a dense forest scene. A rocky stream flows from the background towards the foreground. The water is clear and reflects the surrounding greenery. Large, mossy rocks are scattered along the stream bed. The forest is filled with tall trees and dense undergrowth, with sunlight filtering through the canopy.

1600s.



**Late 1700s.**

1800s.



Left of Samey Meyer & Sons, 440 Broadway N.Y.

112

New-York Historical Society Library

G. New York City - Printed & Engraved - 1800s.

PICTURE COLLECTION

1913.



1950s.



1985.



# CBGB

1993.

315

EL 776

# OMFUG

MILK  
COW 315

SURFING



# CBGB

315

# OMFUG

315

2006.

113



2006.



2006.



Today.

john varvatos  
**315 Bowery**

C  
HOME

RING  
ONCE  
FOR  
ENTRY

SPRINKLER  
THROUGHOUT  
BUILDING

SHUT OFF VALVE  
LOCATED ON  
ROOF TOP  
SHUT OFF VALVE  
LOCATED ON  
ROOF TOP  
SHUT OFF VALVE  
LOCATED ON  
ROOF TOP  
SHUT OFF VALVE  
LOCATED ON  
ROOF TOP

Vintage

CLOTHING  
RECORDS  
AUDIO

Today.



# Same thing down the block.



Mars Bar



TD Bank

25 East 1st Street, New York, NY 10003

*What is gentrification?*

# Definitions.



gentrification is



gentrification is **good**

gentrification is **bad**

gentrification is

gentrification is **a myth**

Press Enter to search.



## Ruth Glass, sociologist.

*“One by one, many of the working class quarters of London have been invaded by the middle-classes—upper and lower. Shabby, modest mews and cottages—two rooms up and two down—have been taken over, when their leases have expired, and have become elegant, expensive residences... Once this process of 'gentrification' starts in a district it goes on rapidly until all or most of the original working-class occupiers are displaced and the whole social character of the district is changed.”*

1964

# Gentrification.

*noun* gen·tri·fi·ca·tion \,jen-trə-fə-'kā-shən\

the process of renewal and rebuilding accompanying the influx of middle-class or affluent people into deteriorating areas that often displaces poorer residents.

**How is it  
measured?**

*How is it measured?*

**Decennial US  
Census.**

# The US Census collects a lot of data.

Standardized questions asked every 10 years on a number of topics and themes:

- Education
- Employment
- Families & living arrangements
- Health
- Housing
- Income & poverty
- Population
  - Age
  - Gender
  - Race
  - Migration
  - Ancestry & ethnicity
  - Language
  - Veteran status

# At various levels of granularity.

Aggregated by:

- 4 regions (Northeast, Midwest, South, & West)
- 50 states
- 3,000+ counties
- **74,000+ Census tracts**
- 220,000+ Census block groups
- 11,000,000+ Census blocks

# Decennial US Census datasets.

Decennial US Census datasets used:

**1970    1980    1990    2000**

Allowing analysis of change over time:

**1970 to 1980:** 10 year timeframe

**1980 to 1990:** 10 year timeframe

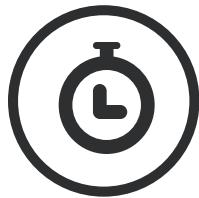
**1990 to 2000:** 10 year timeframe

*How is it measured?*

# American Community Survey.

# Adding ACS data.

The US Census Bureau started collecting annual sample data in 2005.



## Advantage

More frequent updates.



## Disadvantage

Smaller annual sample sizes; rolling 5-year aggregates for census tracts.

# Adding ACS data.

ACS datasets used:

**“2008”:** 2006-2010 five year estimates

**“2010”:** 2008-2012 five year estimates

**“2011”:** 2009-2013 five year estimates

Allowing for the following additional timeframes:

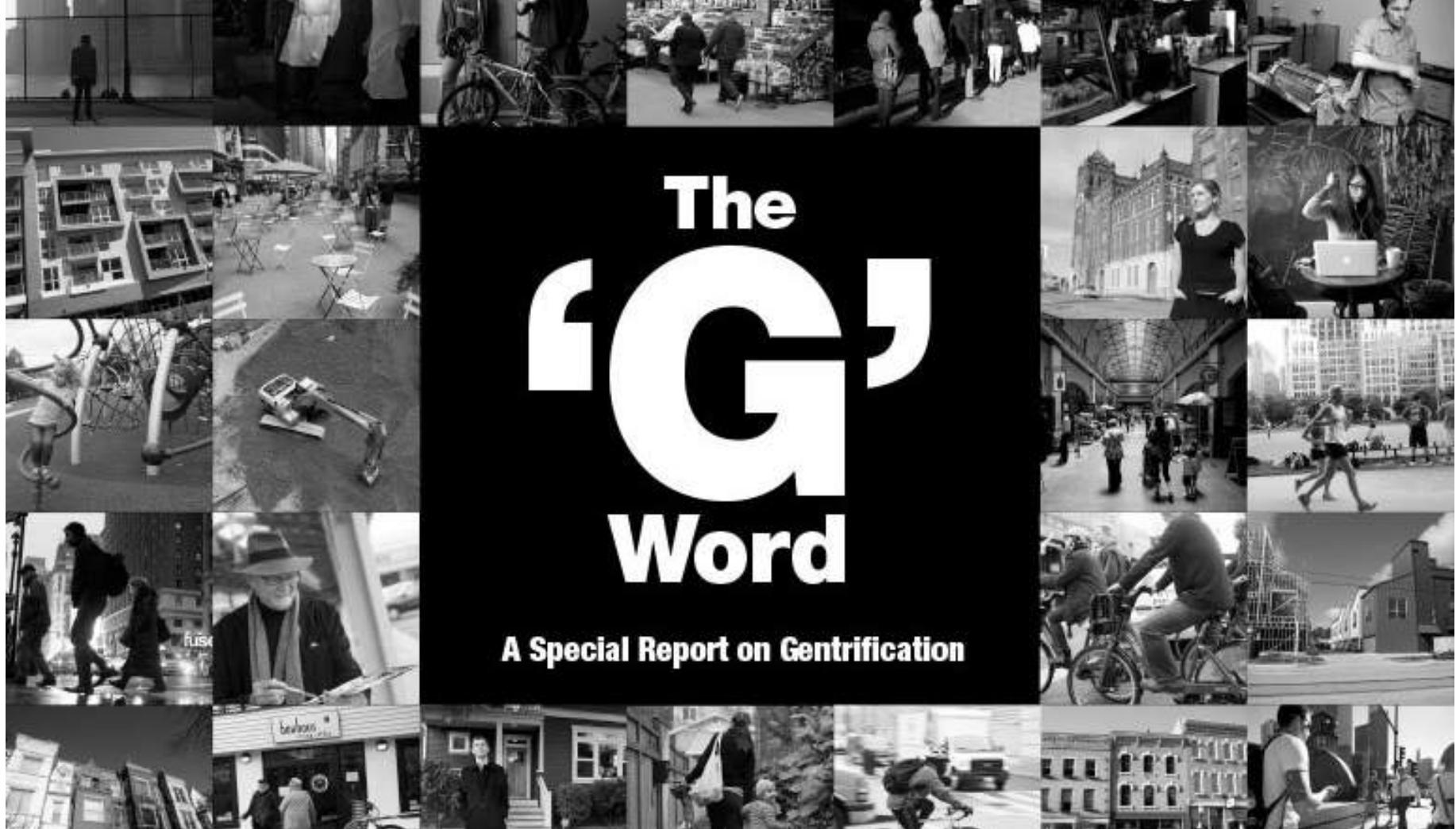
**2000 to 2008:** 8 year timeframe

**2008 to 2010:** 2 year timeframe

**2010 to 2011:** 1 year timeframe

*How is gentrification measured?*

# Governing Magazine's method.



# The ‘G’ Word

A Special Report on Gentrification

*Governing Magazine, February 2015.*

# Is a census tract eligible for gentrification?

Governing Magazine design.



Located  
in a  
central  
city?



Population at  
start and end  
 $\geq 500$ ?



Median  
household  
income  $\leq 40^{\text{th}}$   
percentile for  
metro area at  
start?



Median home  
value  $\leq 40^{\text{th}}$   
percentile for  
metro area at  
start?

**Yes**

**Continue.**

**Continue.**

**Continue.**

**Eligible.**

**No**

**Ineligible.**

**Ineligible.**

**Ineligible.**

**Ineligible.**

# Did an eligible tract gentrify?

Governing Magazine design.



% of population  
aged 25+ who hold  
a college degree  $\geq$   
67th percentile for  
metro at end?



Median  
inflation  
adjusted  
home value  
increased?



% increase in  
median home  
value  $\geq$  67th  
percentile for  
metro at end?

**Yes**

Continue.

Continue.

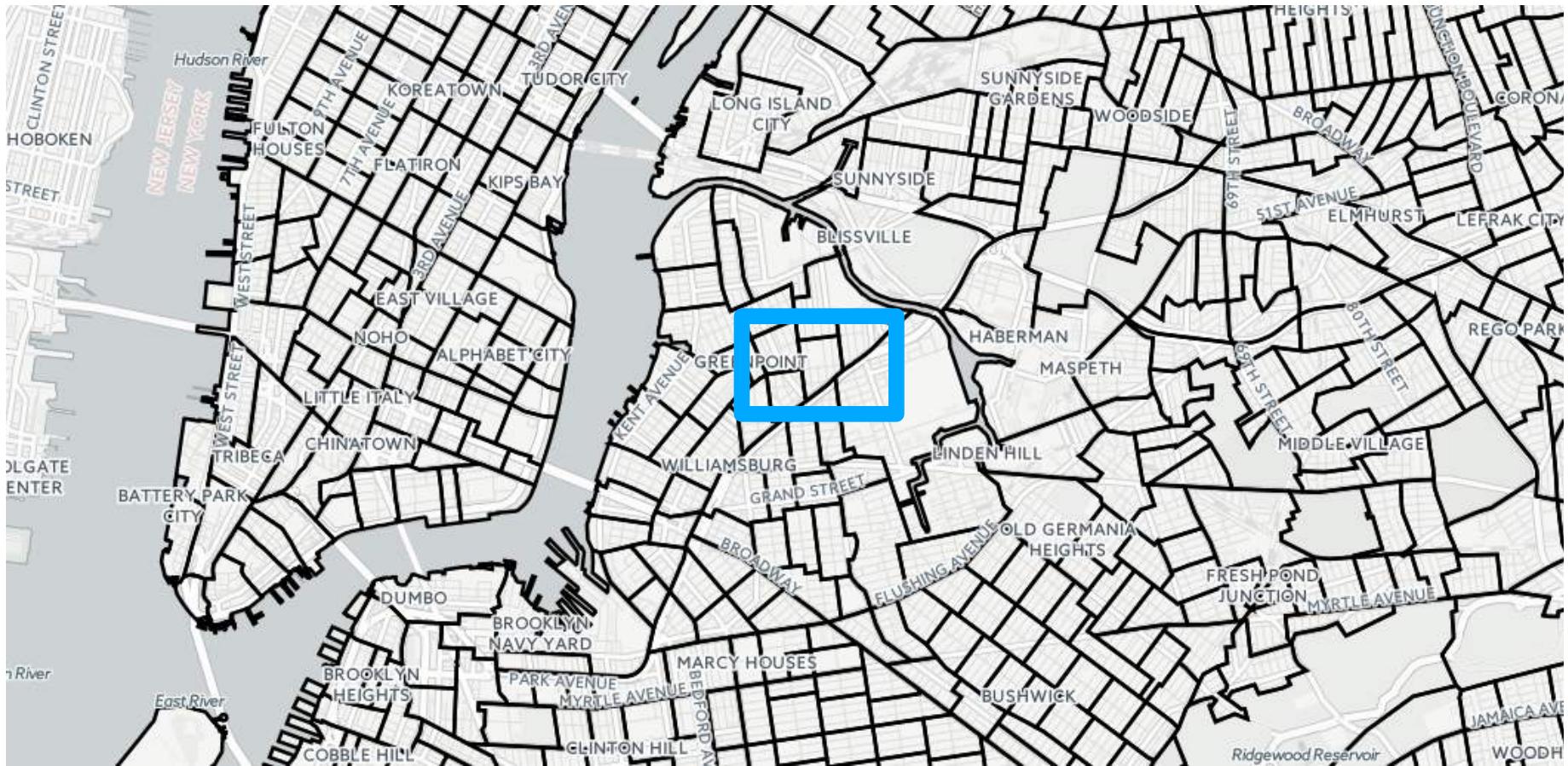
**Gentrified.**

**No**

Did not  
gentrify.

Did not  
gentrify.

Did not  
gentrify.



# Census tracts allow us to track change.

An example from North Brooklyn.

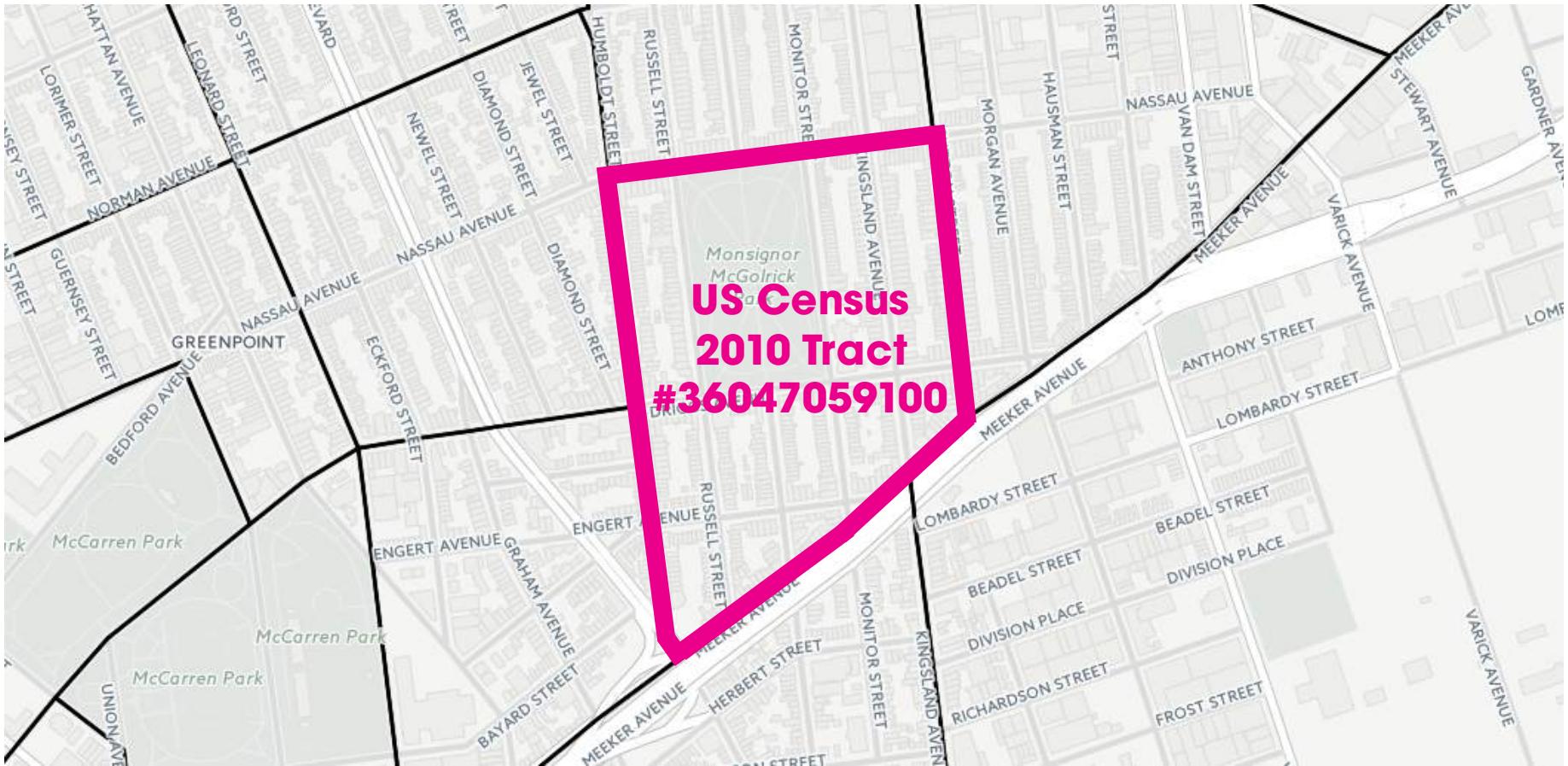


# Greenpoint.

# Williamsburg.

**Census tracts allow us to track change.**

An example from North Brooklyn.



**Census tracts allow us to track change.**

An example from North Brooklyn.

# Applying Governing Magazine's criteria.

An example from Greenpoint, Brooklyn.



Located  
in a  
central  
city?



Population at  
start and end  $\geq$   
500?



Median  
household  
income  $\leq$  40<sup>th</sup>  
percentile for  
metro area at  
start?



Median home  
value  $\leq$  40<sup>th</sup>  
percentile for  
metro area at  
start?



Eligible?

Decade	Located in a central city?	Population at start and end $\geq$ 500?	Median household income $\leq$ 40 <sup>th</sup> percentile for metro area at start?	Median home value $\leq$ 40 <sup>th</sup> percentile for metro area at start?	Eligible?
1970 to 1980	Yes	Yes	No	Yes	No
<b>1980 to 1990</b>	<b>Yes</b>	<b>Yes</b>	<b>Yes</b>	<b>Yes</b>	<b>Yes</b>
1990 to 2000	Yes	Yes	Yes	No	No
2000 to 2008	Yes	Yes	Yes	No	No
2008 to 2010	Yes	Yes	No	No	No
2010 to 2011	Yes	Yes	No	No	No

# Applying Governing Magazine's criteria.

An example from Greenpoint, Brooklyn.



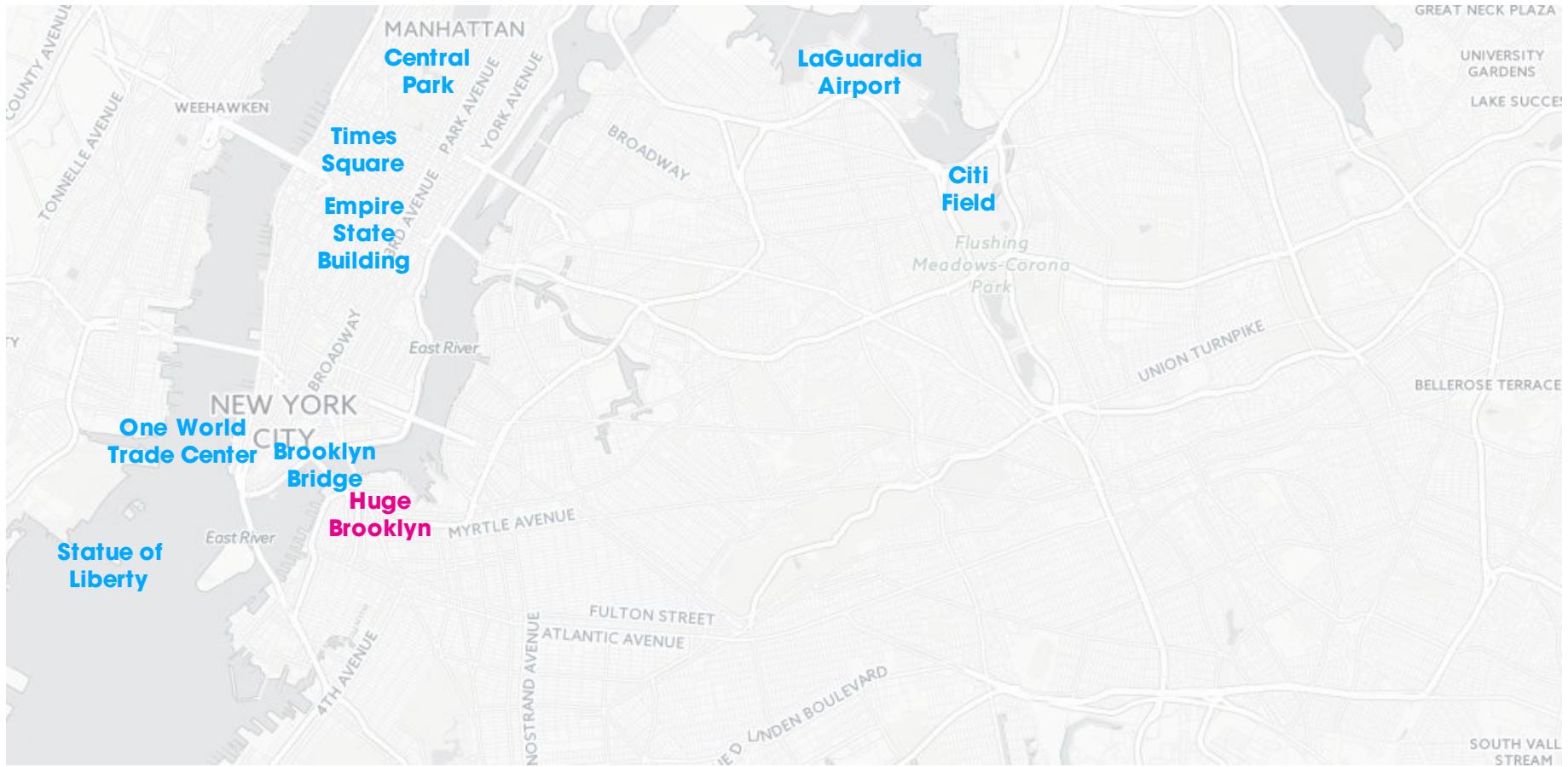
Eligible?	% of population aged 25+ who hold a college degree $\geq$ 67th percentile for metro at end?	Median inflation adjusted home value increased?	% increase in median home value $\geq$ 67th percentile for metro at end?	Gentrified?
1970 to 1980	No	--	--	--
<b>1980 to 1990</b>	<b>Yes</b>	<b>No</b>	<b>No</b>	<b>-</b>
1990 to 2000	No	--	--	--
2000 to 2008	No	--	--	--
2008 to 2010	No	--	--	--
2010 to 2011	No	--	--	--

*Results.*

# Governing Magazine replication.

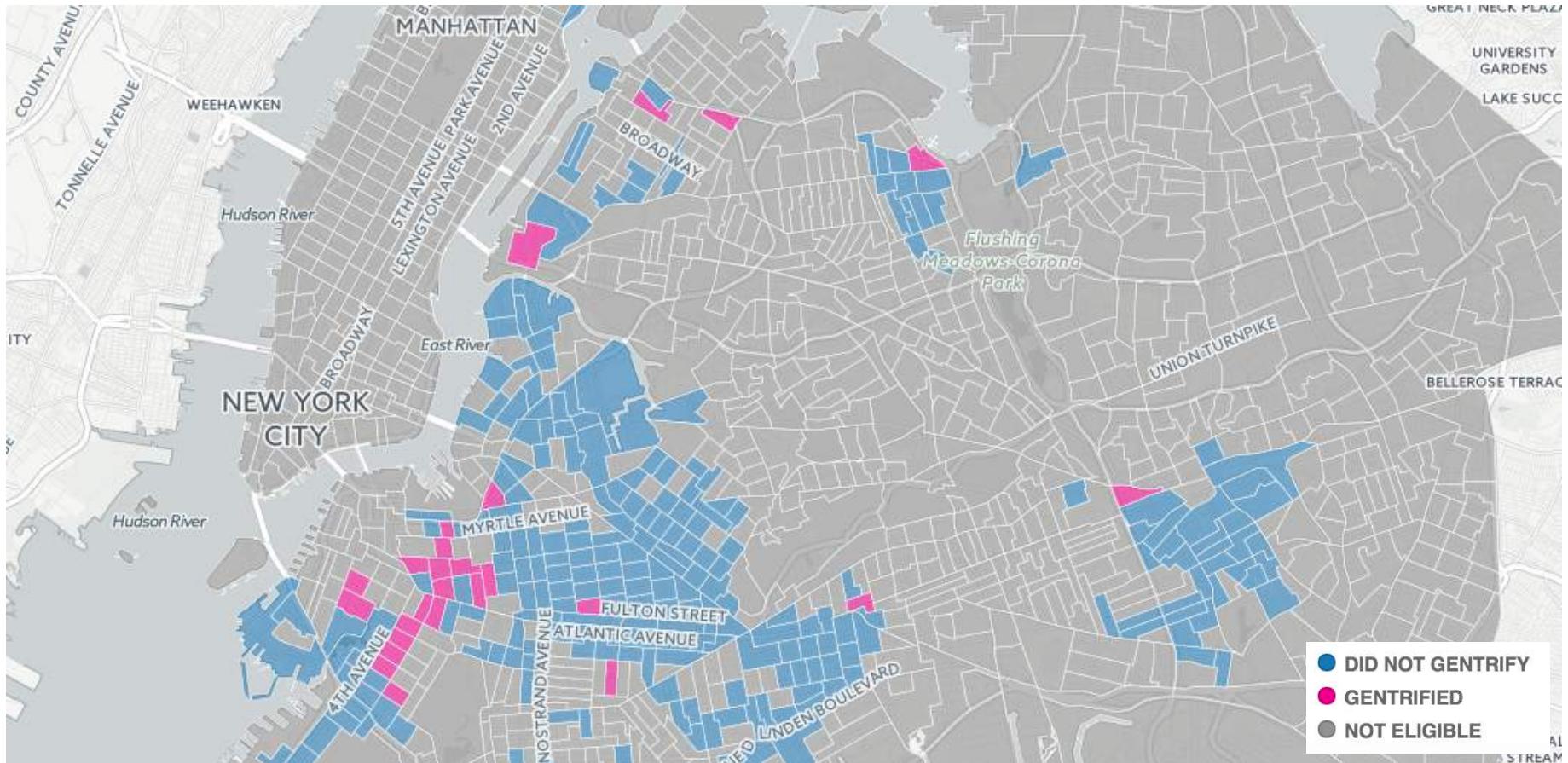
An aerial photograph of the New York City skyline, featuring the One World Trade Center and the Brooklyn Bridge. The city is densely packed with skyscrapers, and the East River flows in the foreground.

New York City.



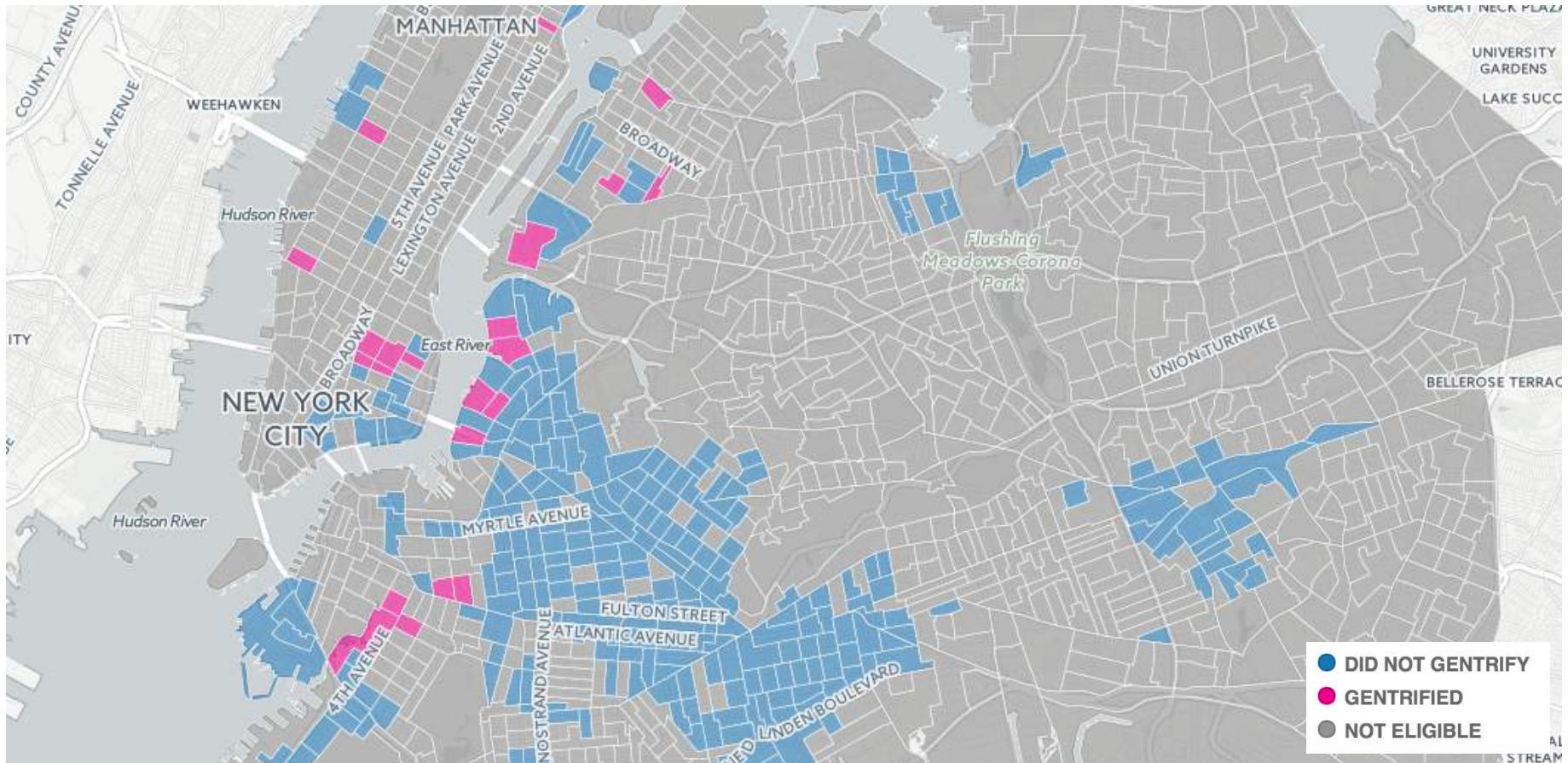
# New York City.

Landmarks.



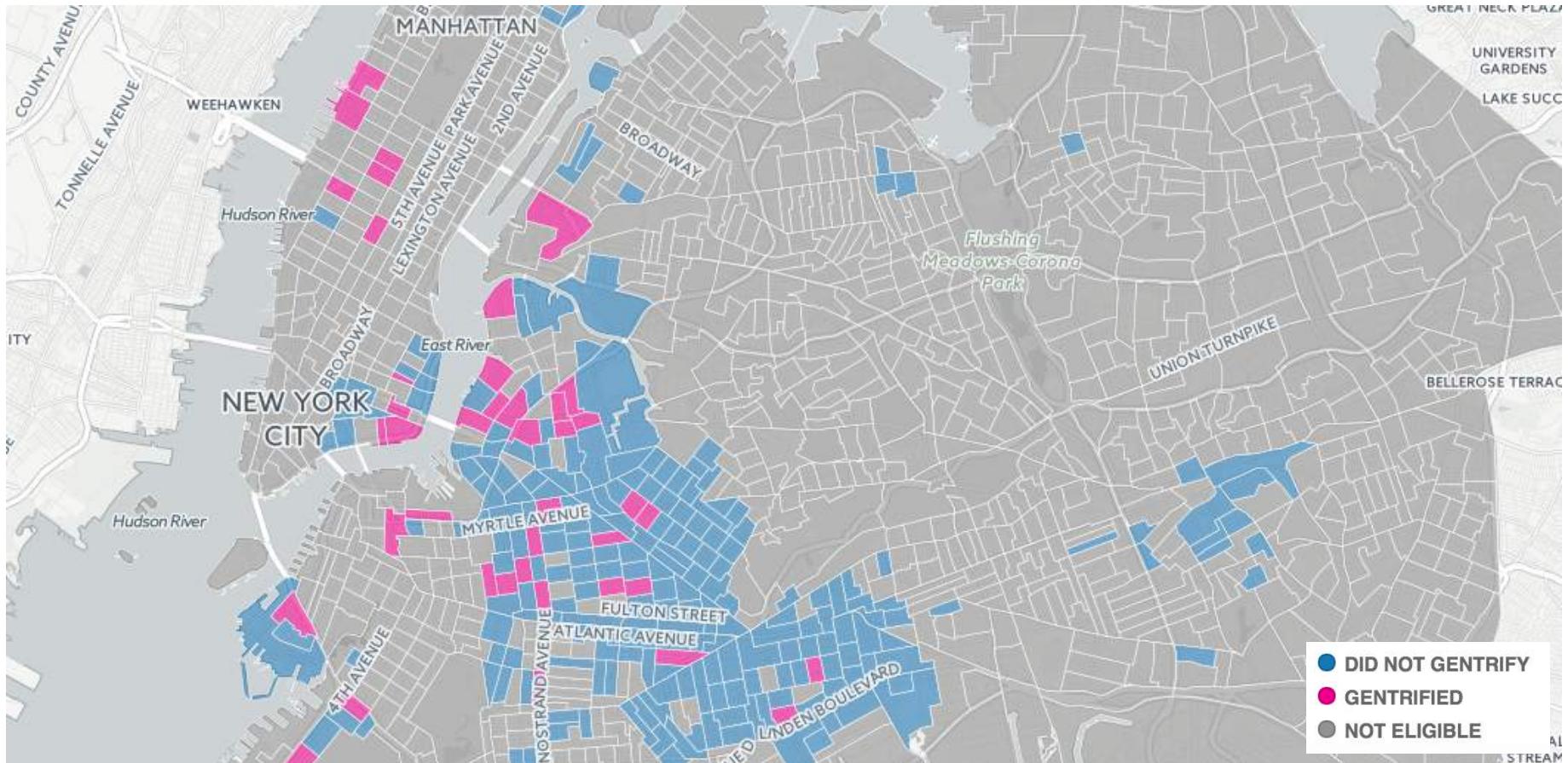
# New York City.

1970 - 1980.



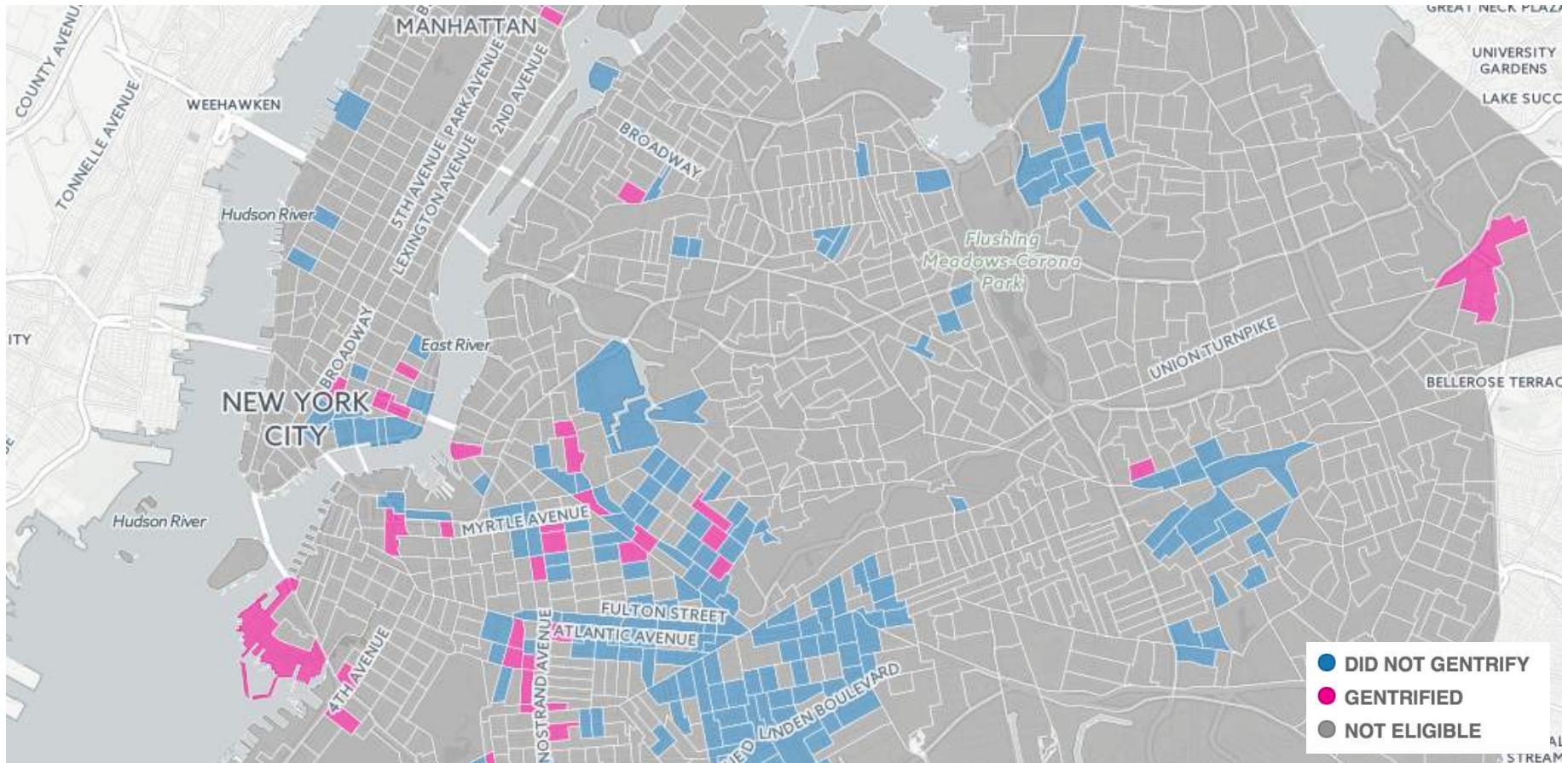
# New York City.

1980 - 1990.



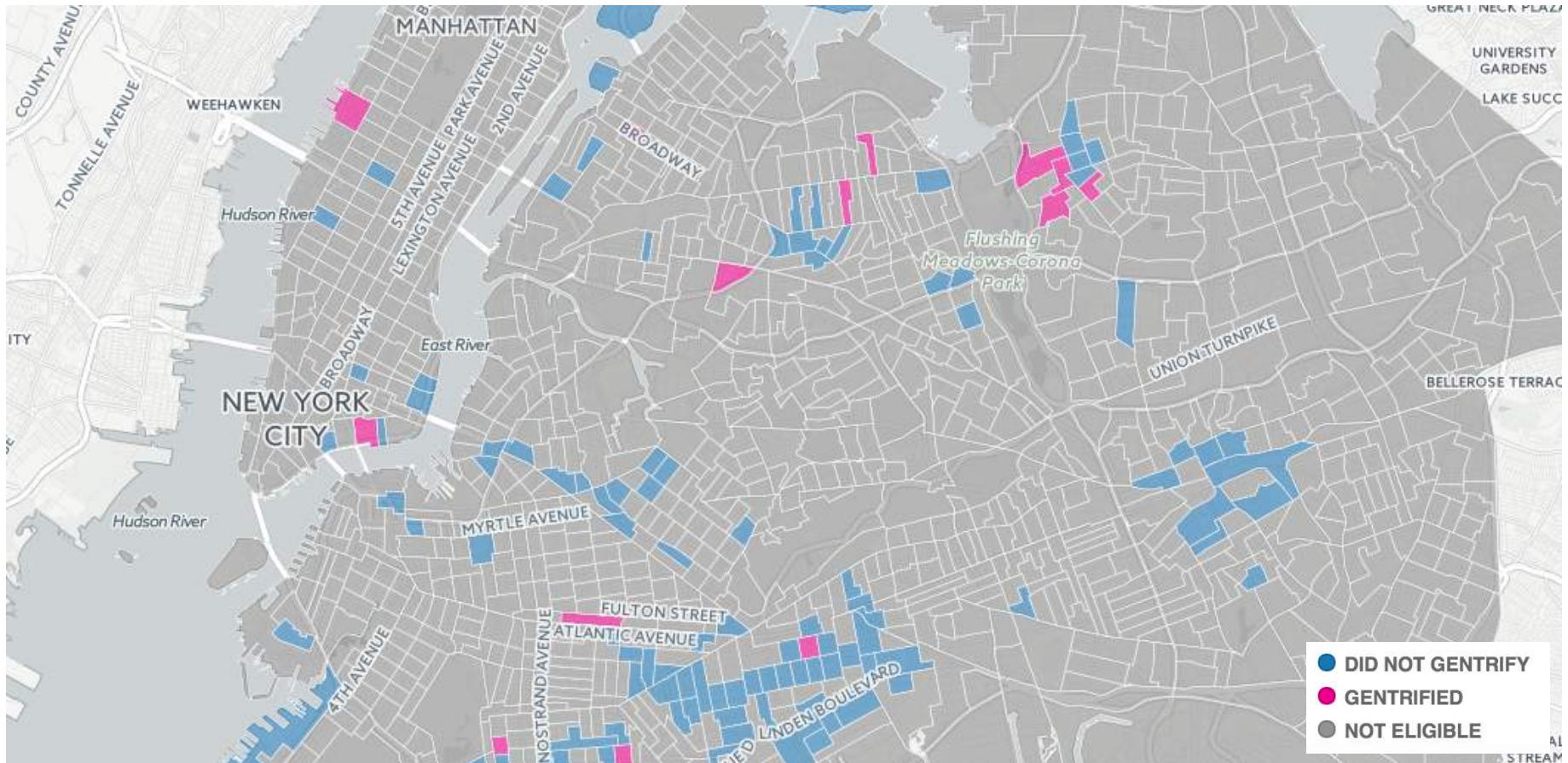
# New York City.

1990 - 2000.



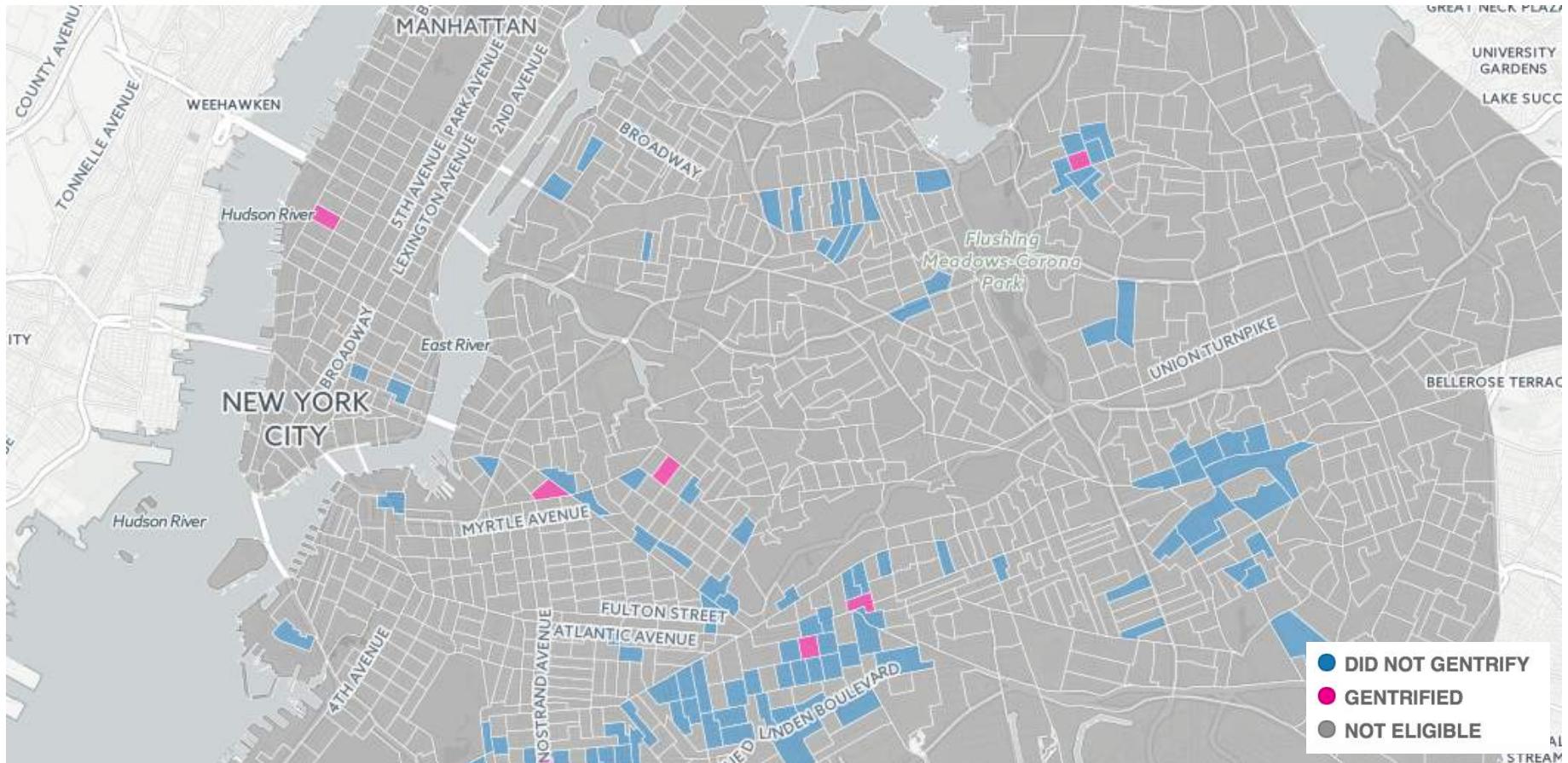
# New York City.

2000 - 2008.



# New York City.

2008 - 2010.



# New York City.

2010 - 2011.

# Applying Governing Magazine's criteria.

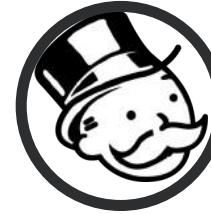
New York City.



Not  
eligible.



Did not  
gentrify.

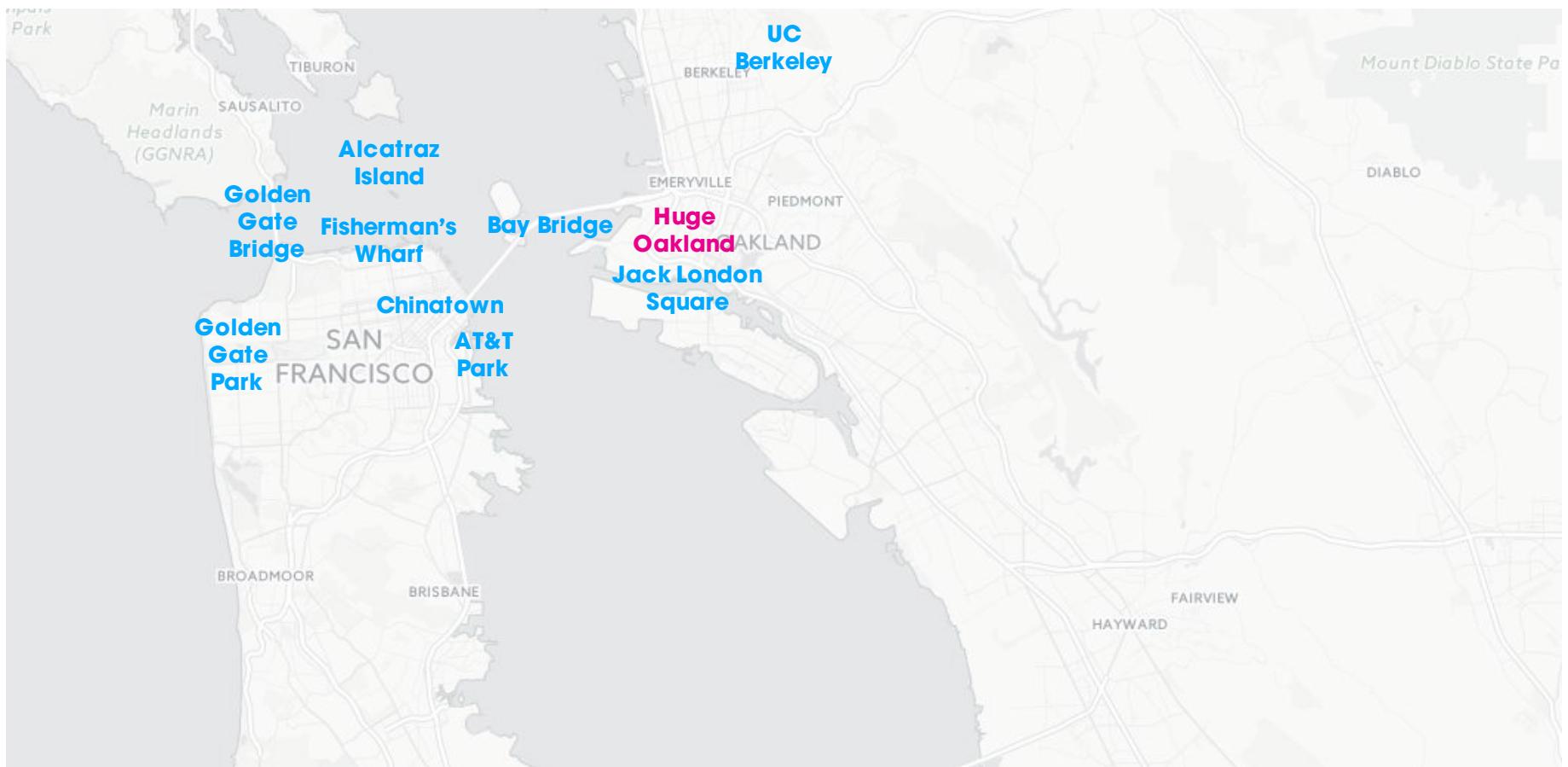


**Gentrified.**

1970 to 1980	78%	20%	<b>2%</b>
1980 to 1990	74%	24%	<b>2%</b>
1990 to 2000	75%	22%	<b>3%</b>
2000 to 2008	79%	19%	<b>2%</b>
2008 to 2010	83%	15%	<b>2%</b>
2010 to 2011	83%	15%	<b>2%</b>

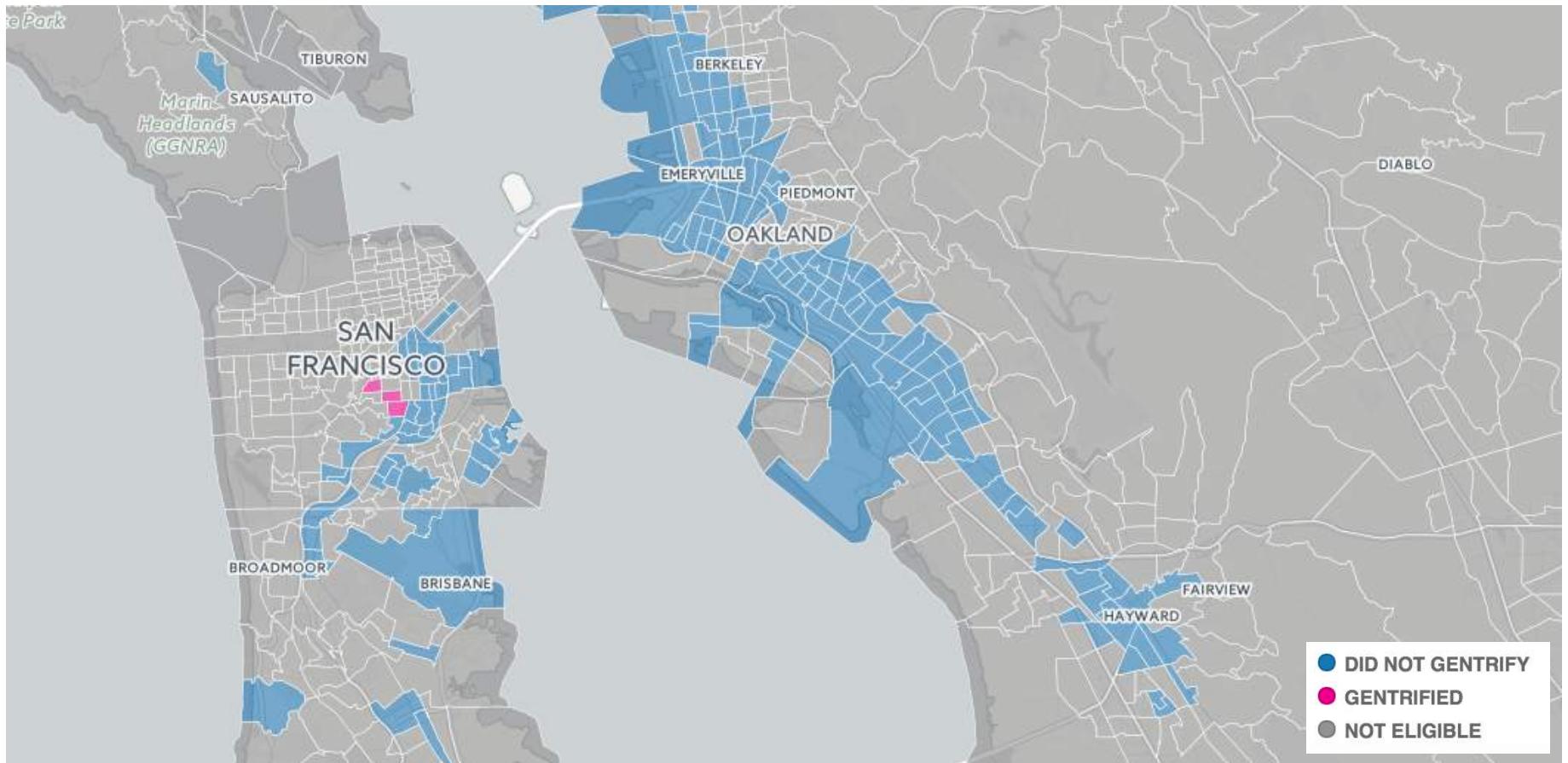


Bay Area.



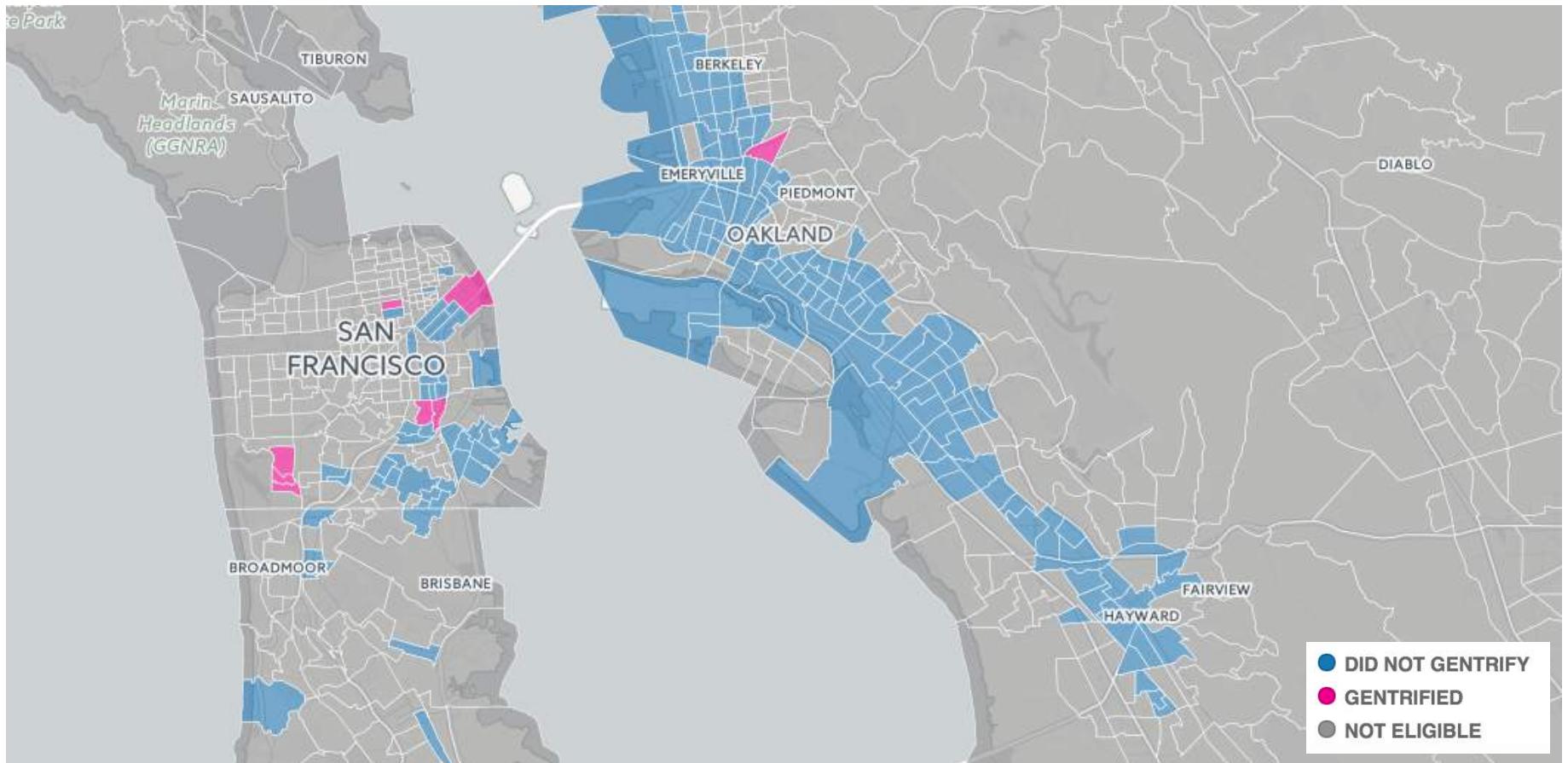
# Bay Area.

Landmarks.



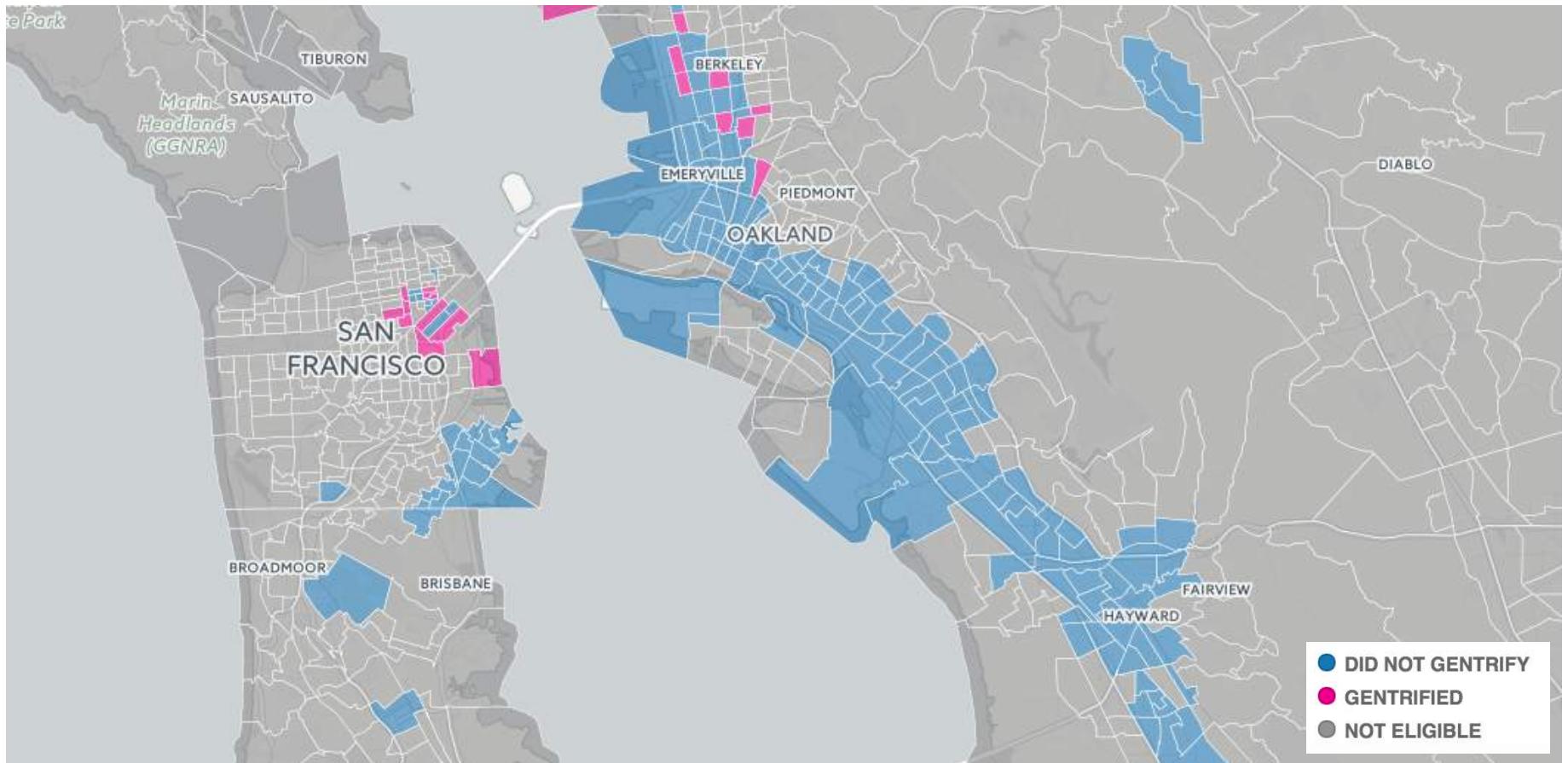
# Bay Area.

1970 - 1980.



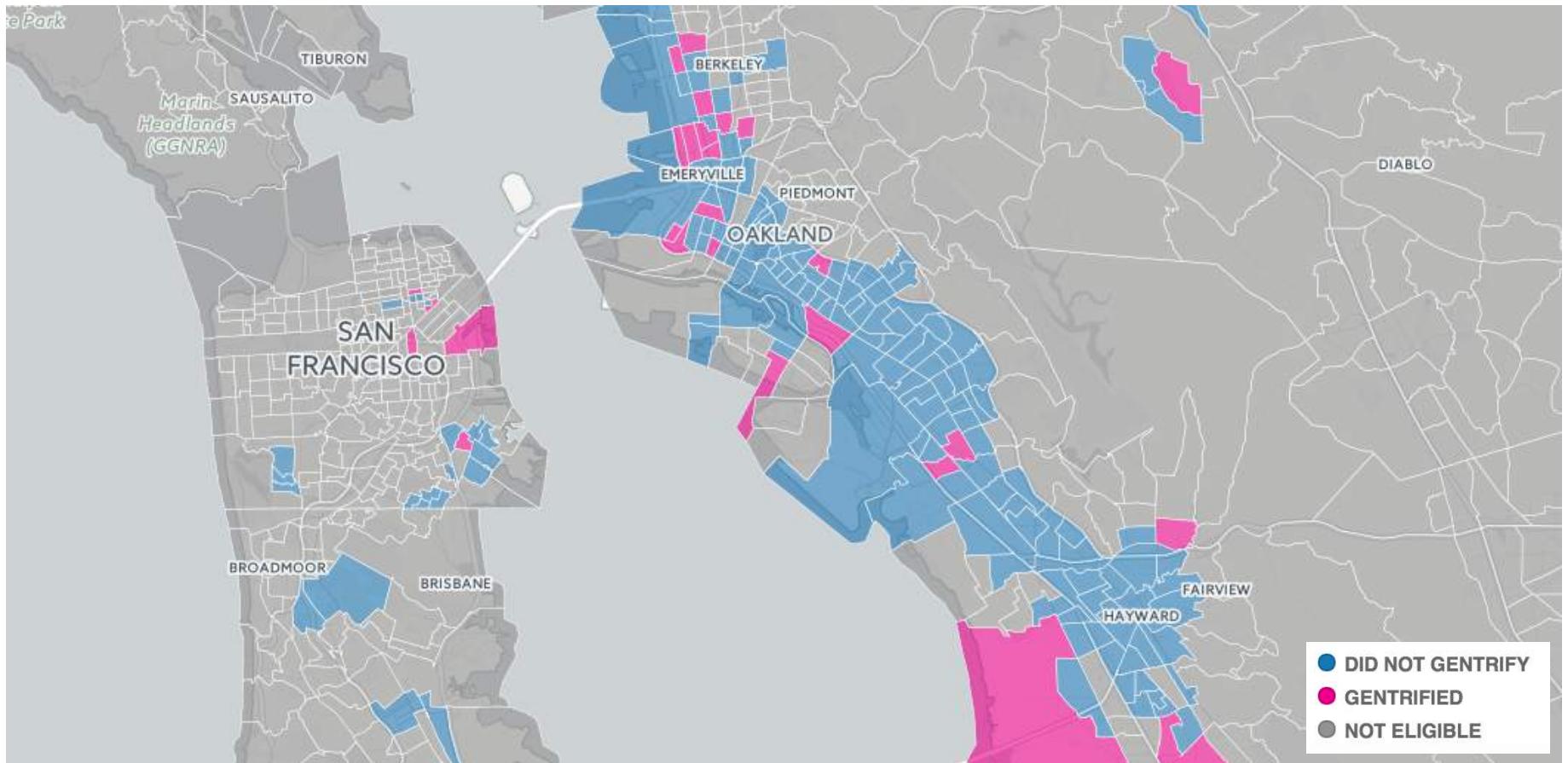
# Bay Area.

1980 - 1990.



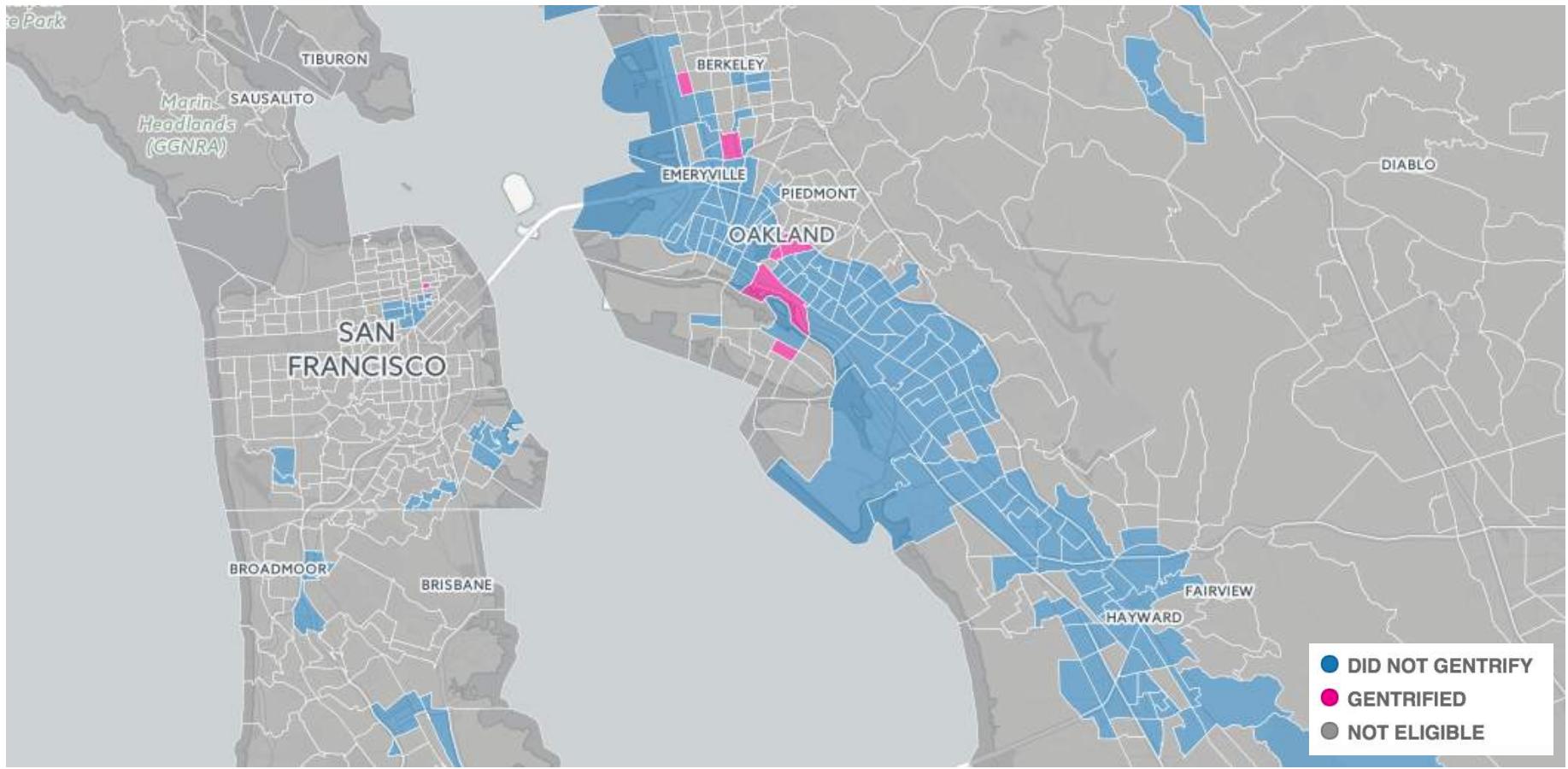
# Bay Area.

1990 - 2000.



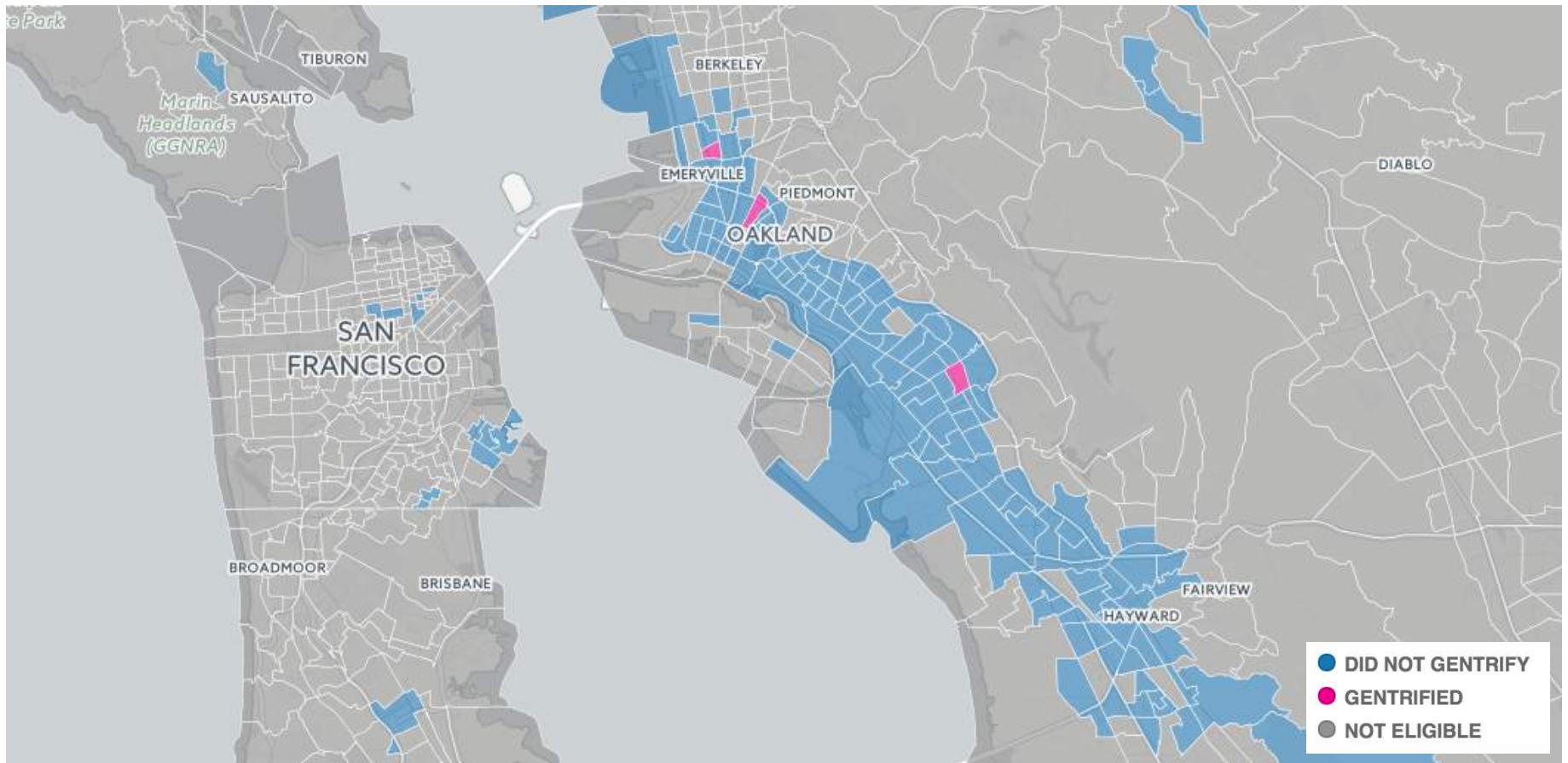
# Bay Area.

2000 - 2008.



# Bay Area.

2008 - 2010.



# Bay Area.

2010 - 2011.

# Applying Governing Magazine's criteria.

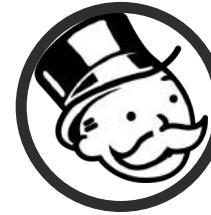
Bay Area.



Not  
eligible.



Did not  
gentrify.



**Gentrified.**

1970 to 1980	77%	23%	< 1%
1980 to 1990	75%	25%	< 1%
1990 to 2000	74%	23%	<b>2%</b>
2000 to 2008	72%	24%	<b>3%</b>
2008 to 2010	73%	26%	< 1%
2010 to 2011	74%	26%	< 1%

This is really  
arbitrary.

**It excludes too many  
neighborhoods.**

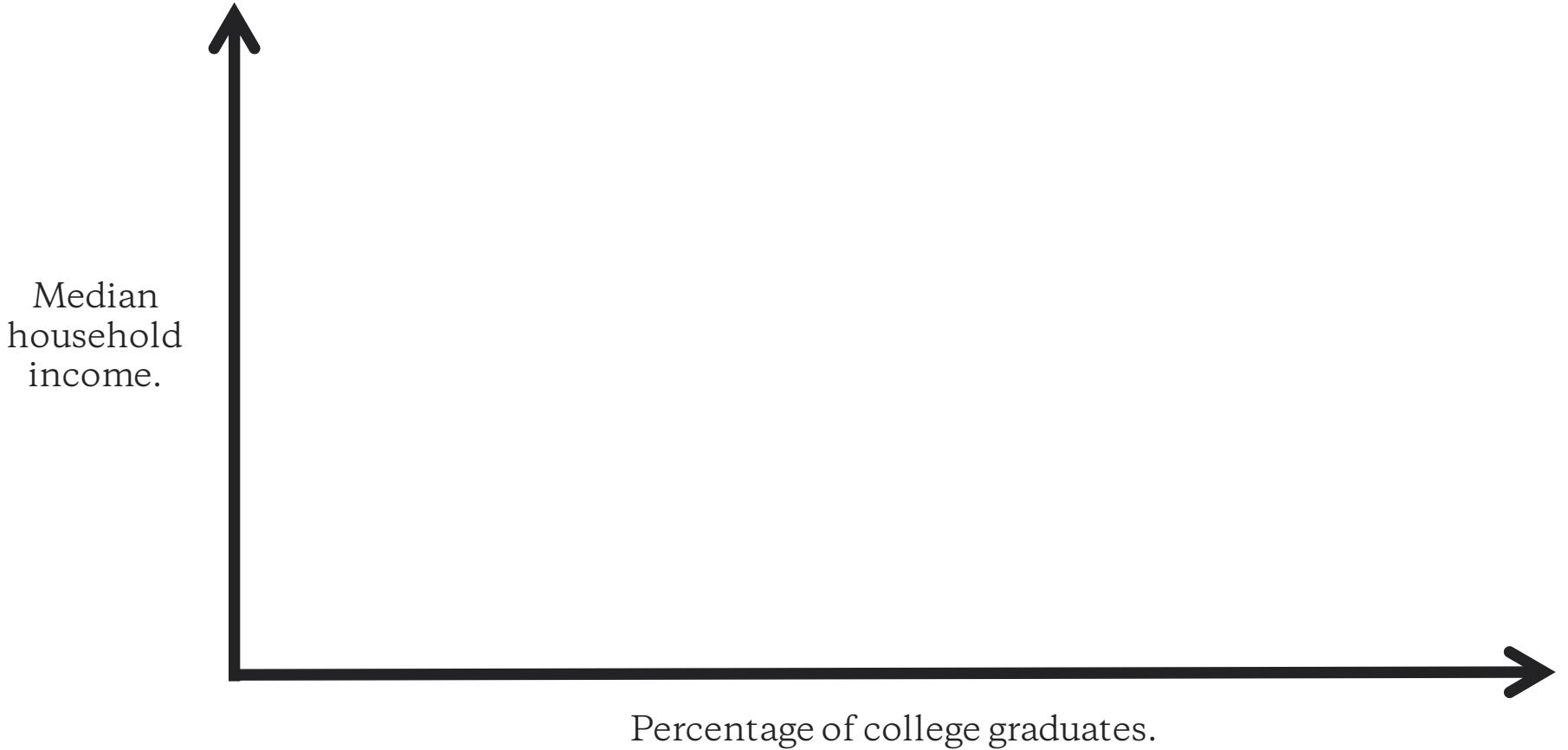
**And provides no insight on  
the full cycle of the economic rise  
and decline of neighborhoods.**

**Is there a  
better way?**

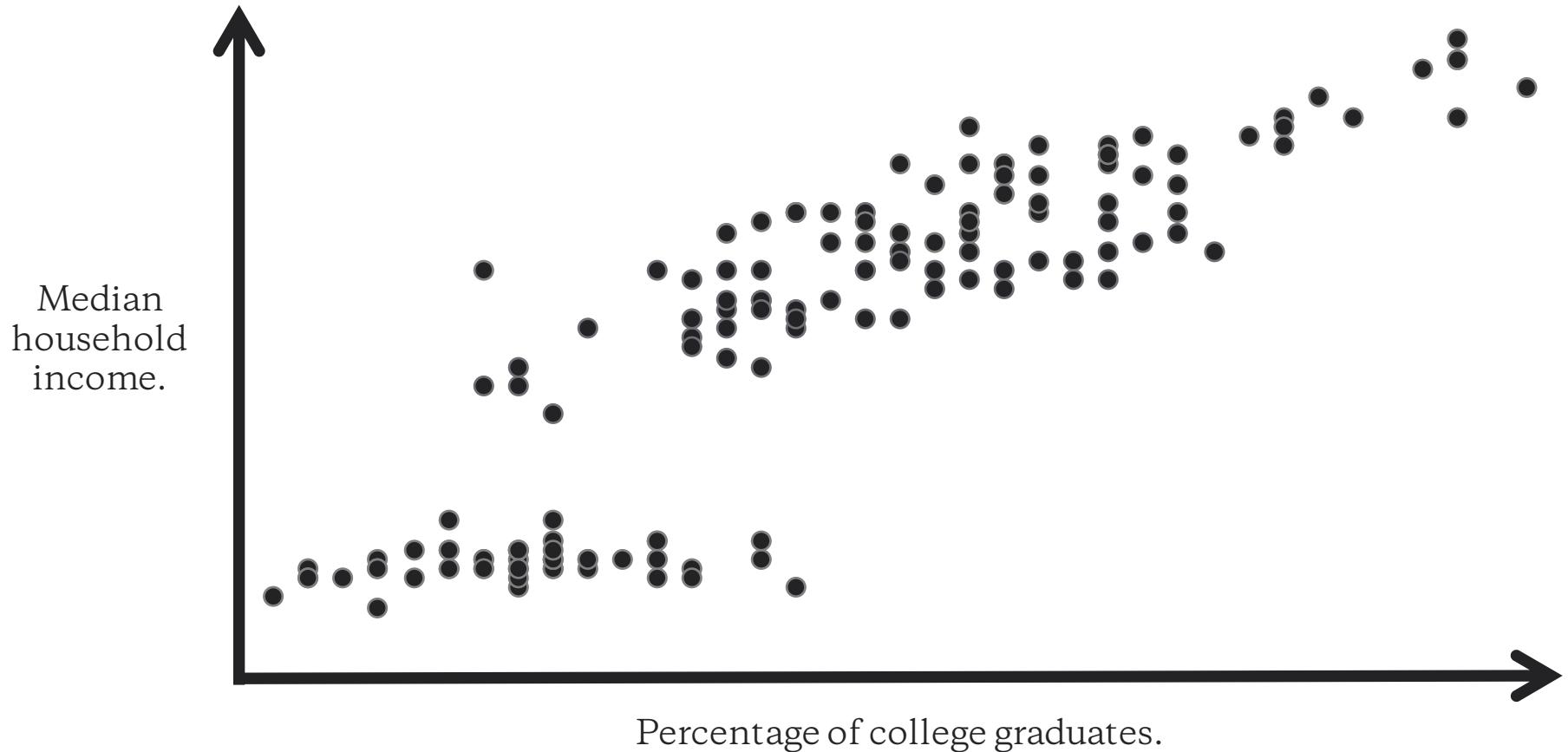
*Yes, and it's called*  
**cluster analysis.**

# **How does cluster analysis work?**

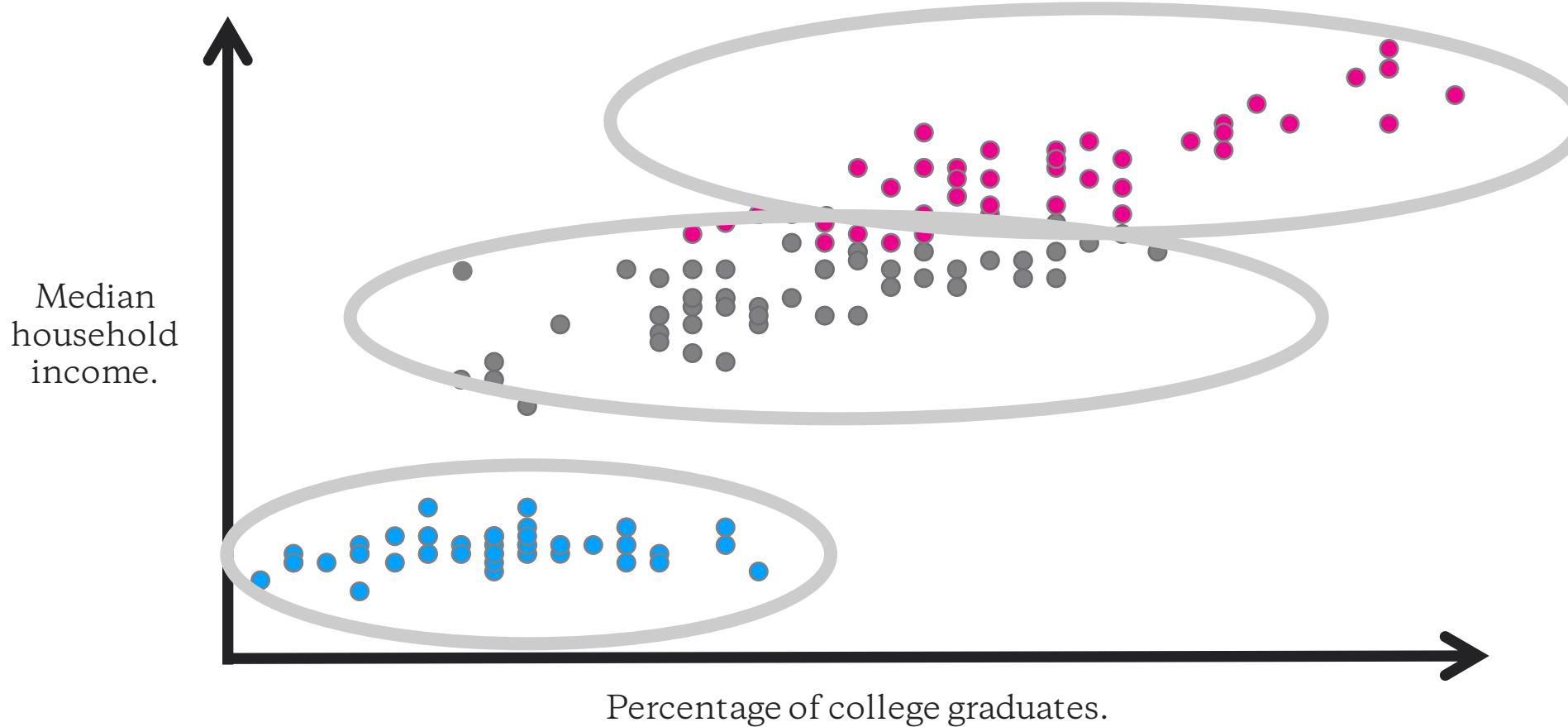
It determines the natural groupings of census tracts by analyzing how they agglomerate within a point cloud.



**Simple two variable segmentation example.**



**Each point is a Census Tract.**



**Three groups.**

*Cluster analysis.*

**But first, let's get  
our data ready.**

# Which variables to use?

Taking a cue from Governing Magazine.



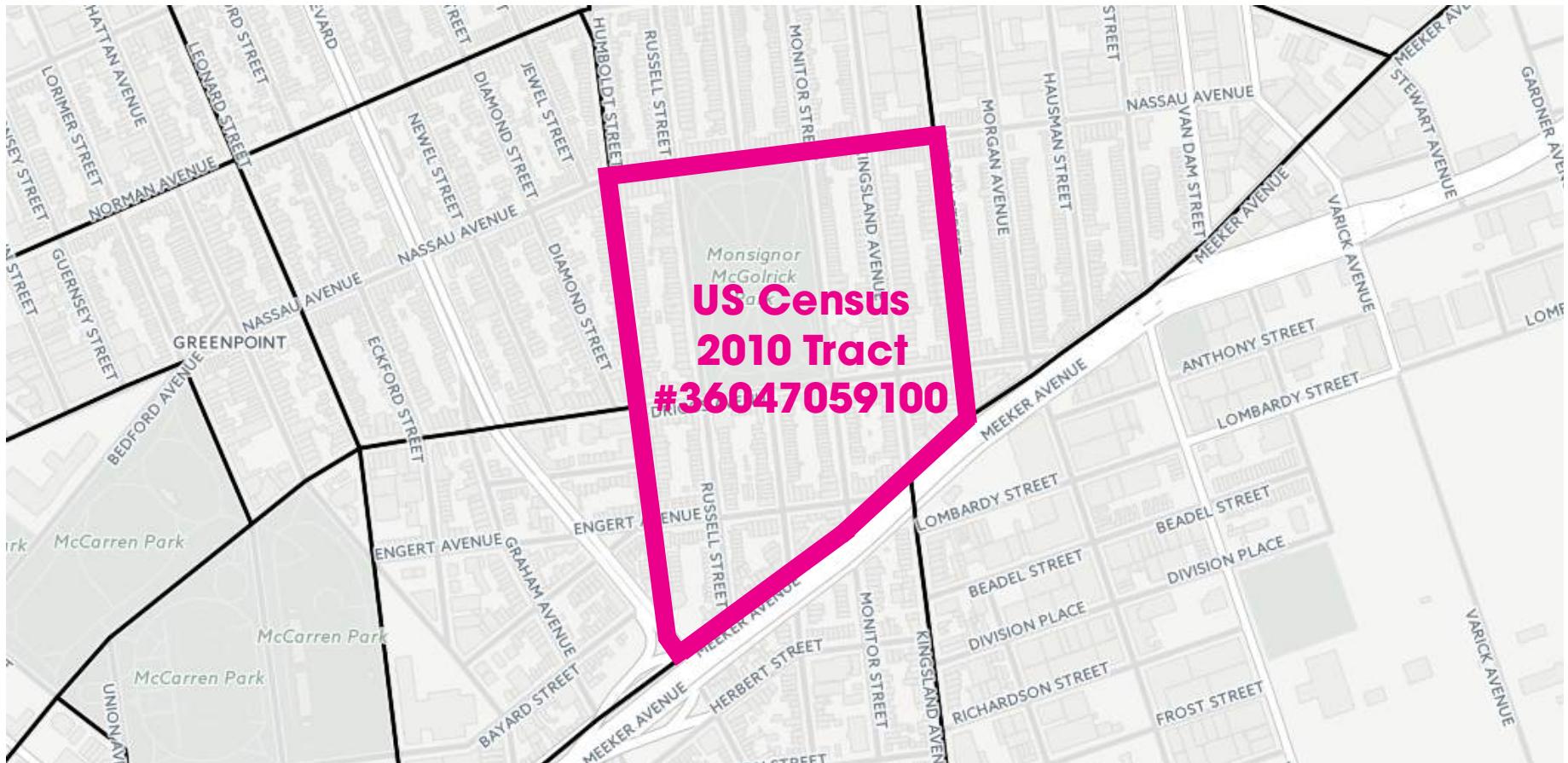
Percent  
change of  
population  
aged 25+ who  
hold a 4 year  
college degree.



Change in  
median  
household  
income,  
controlling for  
inflation.



Change in  
median home  
value,  
controlling for  
inflation.



# Back to Greenpoint.

An example from North Brooklyn.

# Tracking change in US Census tracts.

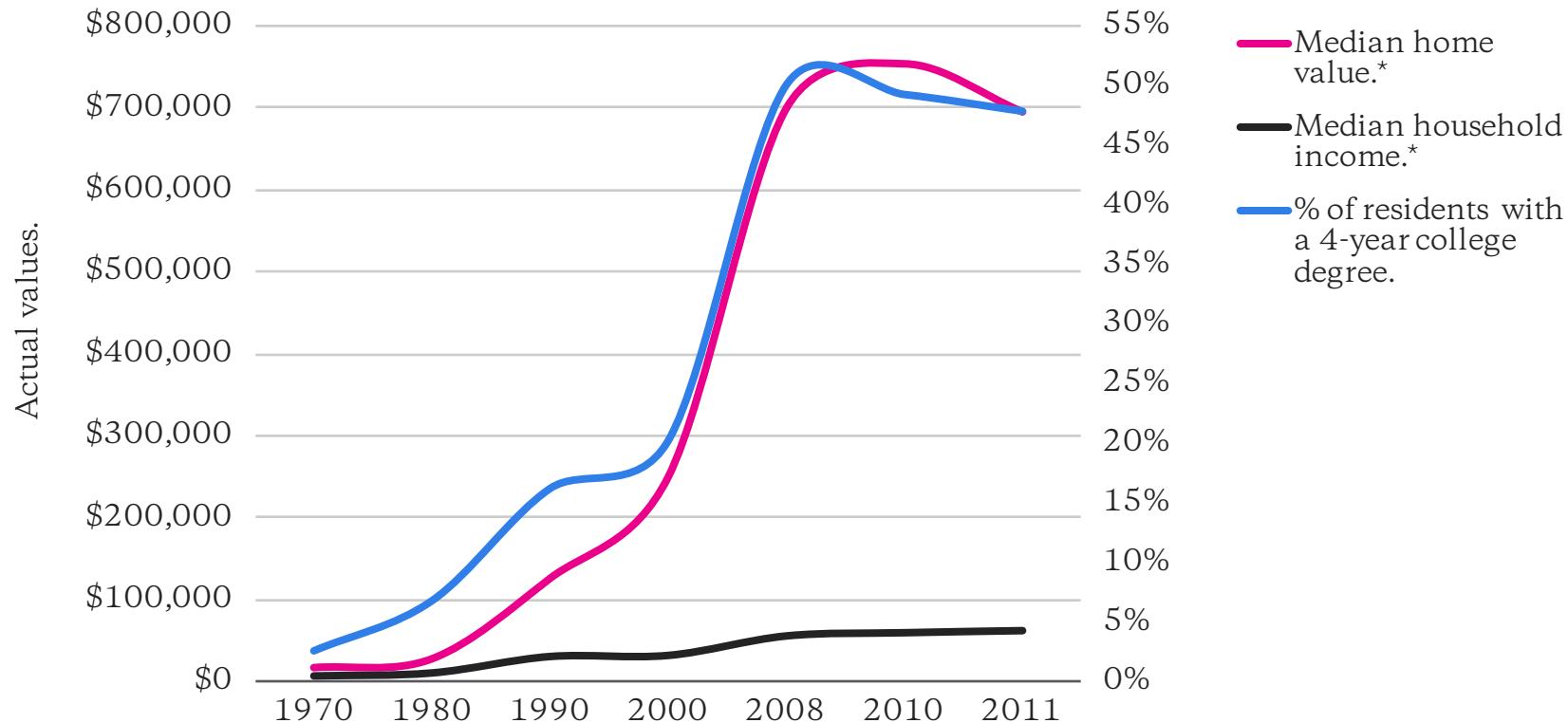
An example from Greenpoint, Brooklyn.

	1970 Census	1980 Census	1990 Census	2000 Census	2008 ACS	2010 ACS	2011 ACS
Median home value*	\$15,875	\$26,667	\$125,000	\$249,000	\$699,900	\$754,400	\$695,600
Median household income*	\$8,274	\$11,882	\$31,734	\$33,047	\$56,573	\$60,478	\$62,986
% of residents with a 4 year college degree	2.6%	6.8%	16.2%	20.2%	50.0%	49.2%	47.8%

Note: \*absolute values, not adjusted for inflation.

# Tracking change in US Census tracts.

An example from Greenpoint, Brooklyn.



Note: \*absolute values, not adjusted for inflation.

# Tracking change in US Census tracts.

An example from Greenpoint, Brooklyn.

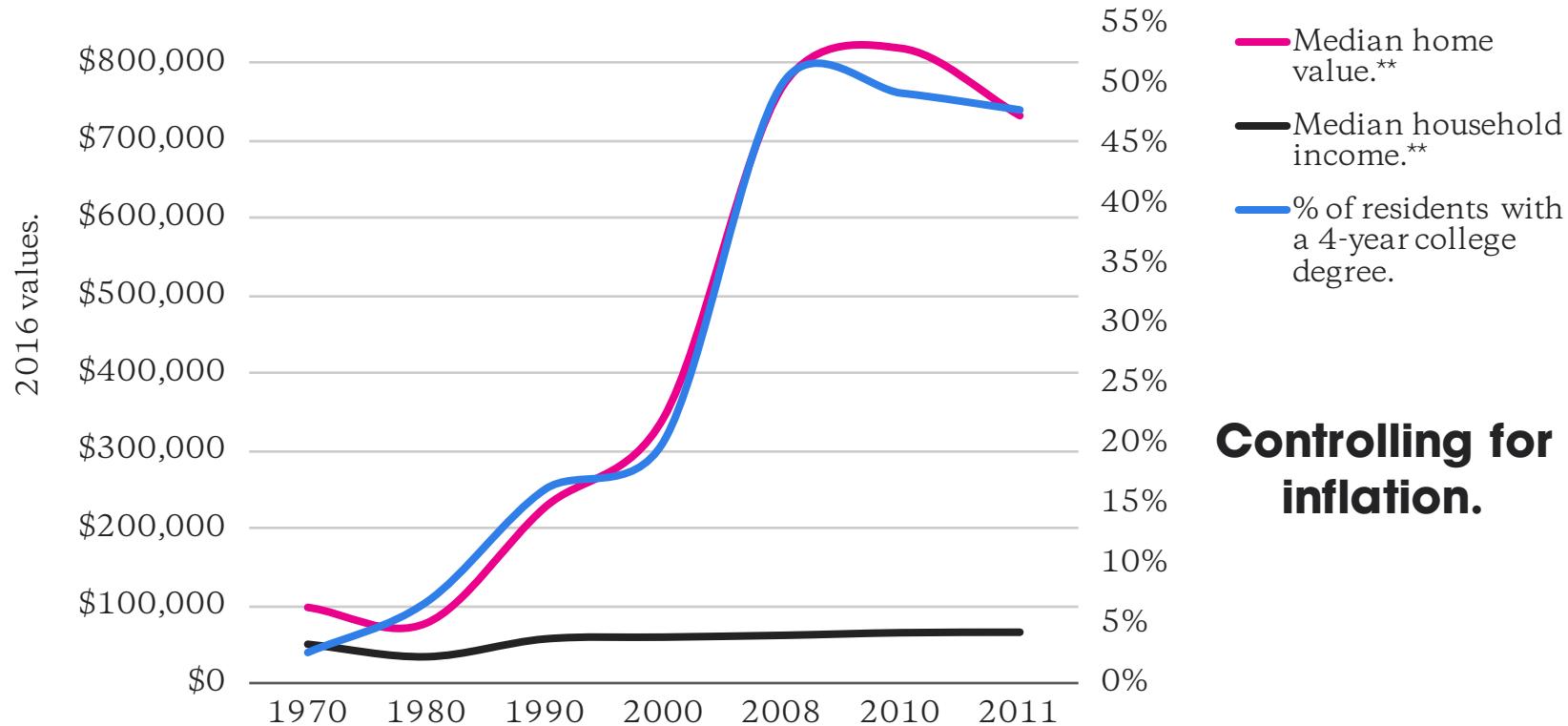
	1970 Census	1980 Census	1990 Census	2000 Census	2008 ACS	2010 ACS	2011 ACS
Median home value**	\$96,934	\$76,673	\$226,584	\$342,579	770,159	819,649	732,638
Median household income**	\$50,522	\$34,163	\$57,523	\$59,903	62,252	65,709	66,340
% of residents with a 4 year college degree	2.6%	6.8%	16.2%	20.2%	50.0%	49.2%	47.8%

**Controlling for inflation.**

Note: \*\*adjusted for inflation, 2016 US dollar values (US Dept. of Labor Bureau of Labor Statistics CPI calculator, <http://data.bls.gov/cgi-bin/cpicalc.pl>)

# Tracking change in US Census tracts.

An example from Greenpoint, Brooklyn.



**Controlling for inflation.**

## **OK, but ...**

we need to account for the trends of the entire metropolitan area, not just the absolute changes in each tract.

# Standardization across NYC tracts.

An example from Greenpoint, Brooklyn from 1970 to 1980.

	1970 Census	1980 Census	End minus start.	Metro mean change (1970- 1980).	Metro standard deviation (1970- 1980).	<b>z-score.</b>
Median home value**	\$96,934	\$76,673	-\$20,261	-\$3,347	\$39,728	<b>-.59</b>
Median household income**	\$50,522	\$34,163	-\$16,359	-\$8,843	\$9,160	<b>-2.75</b>
% of residents with a 4 year college degree	2.6%	6.8%	+4.2%	+5.0%	6.5%	<b>.12</b>

**1970 to 1980**

**Then do this for each time period  
and each tract in the metro area.**

**These three standardized scores  
determine the clustering solutions.**

**This is a very different  
way of exploring  
neighborhood  
dynamics.**

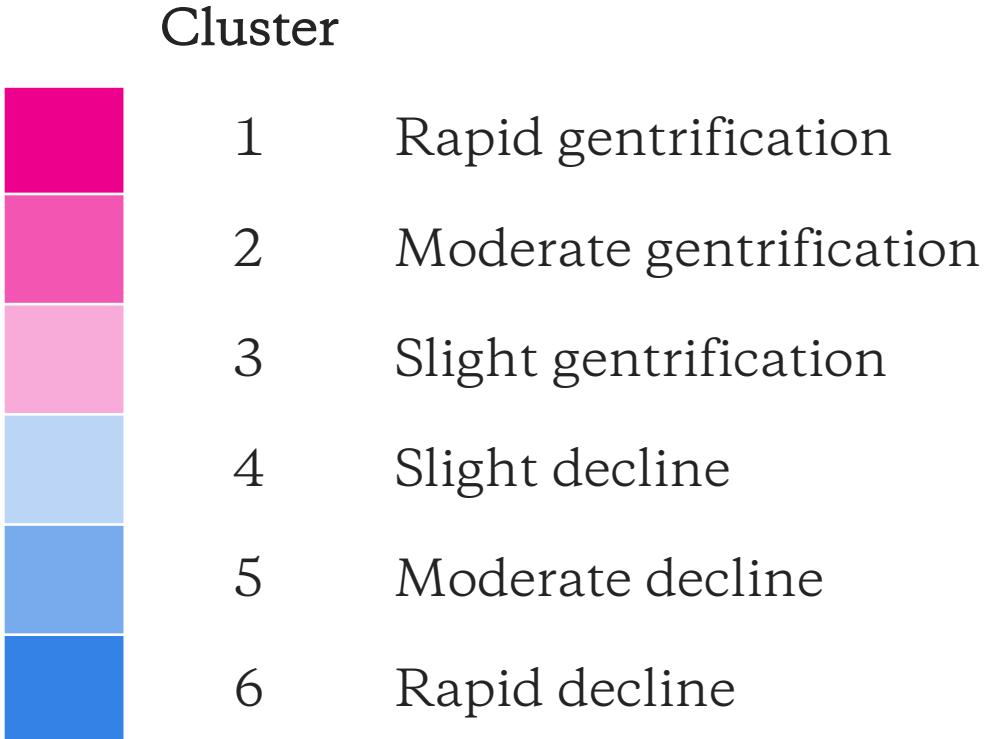
*Results.*

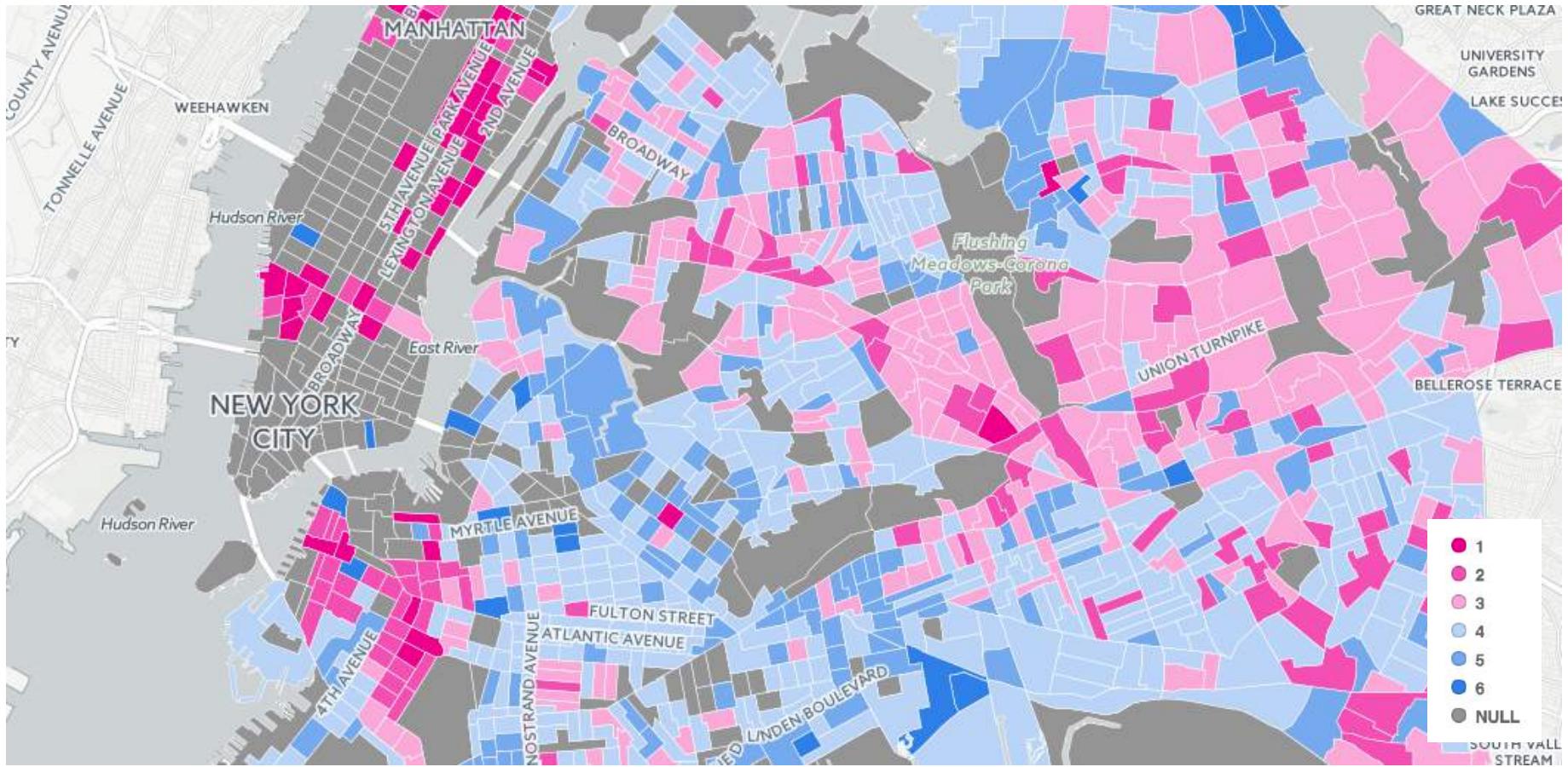
# Clustering solution.

An aerial photograph of the New York City skyline, featuring the One World Trade Center and the Brooklyn Bridge. The city is densely packed with skyscrapers, and the East River flows in the foreground.

New York City.

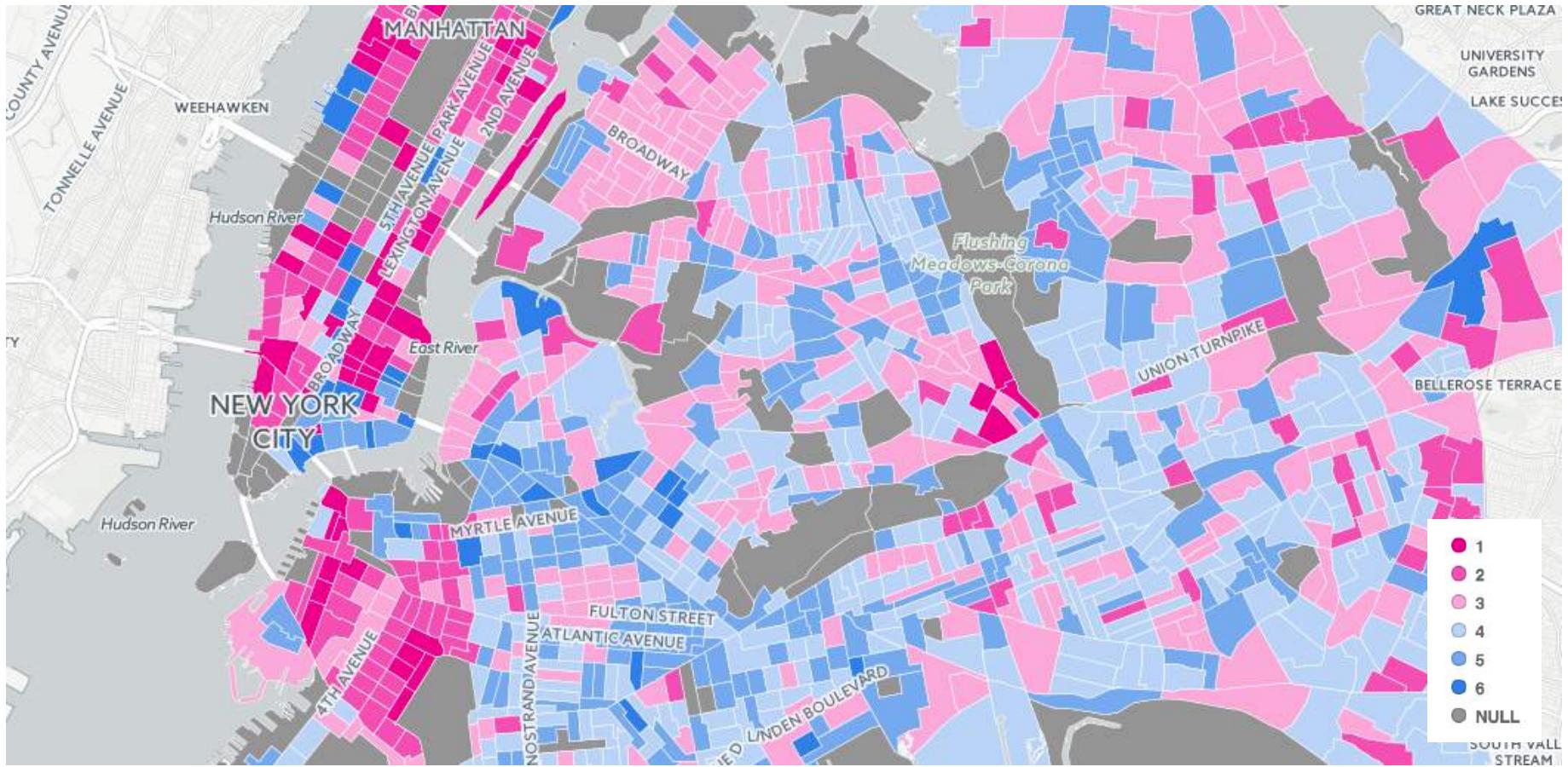
# Six clusters in NYC.





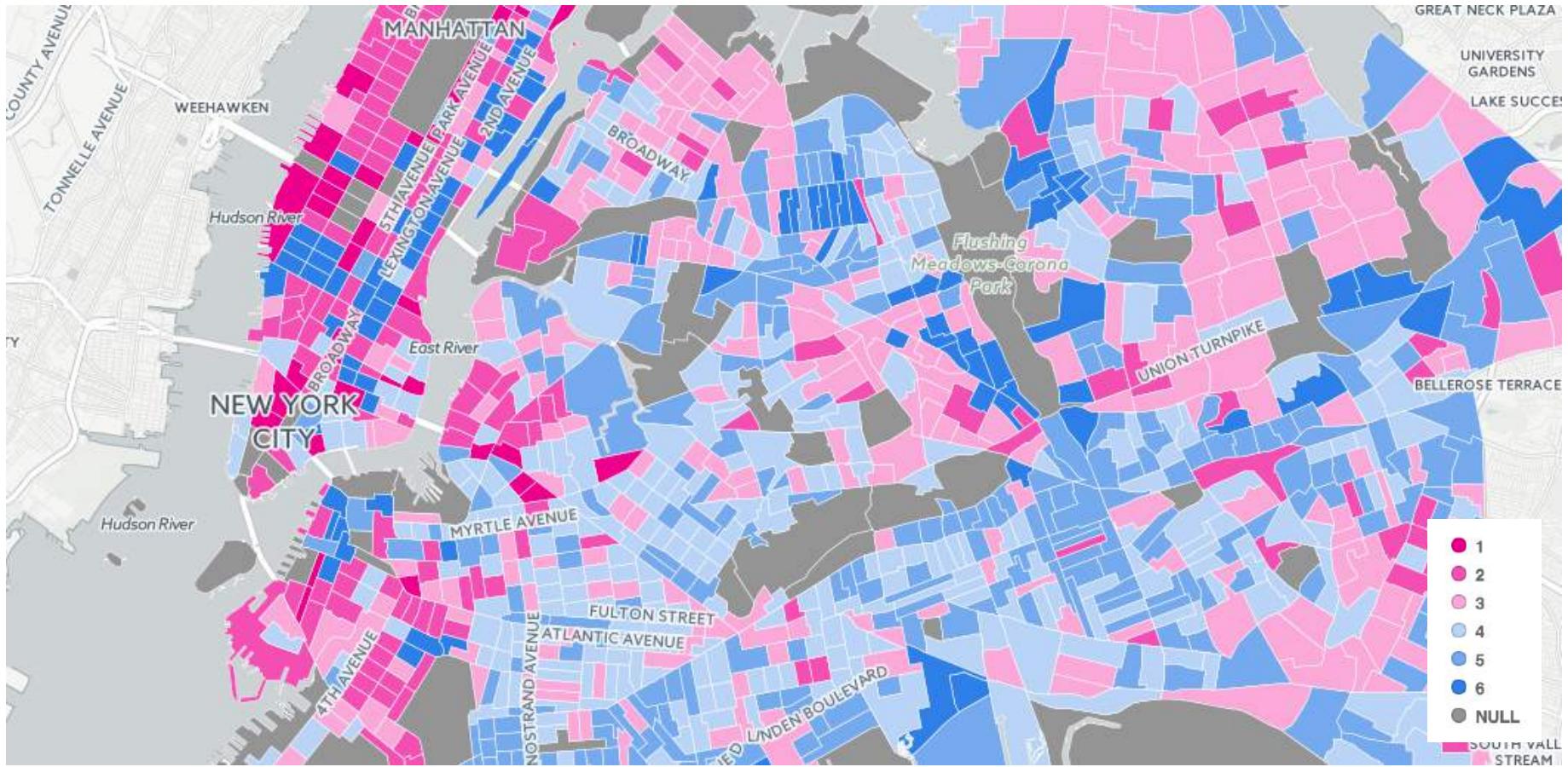
# New York City.

1970 - 1980.



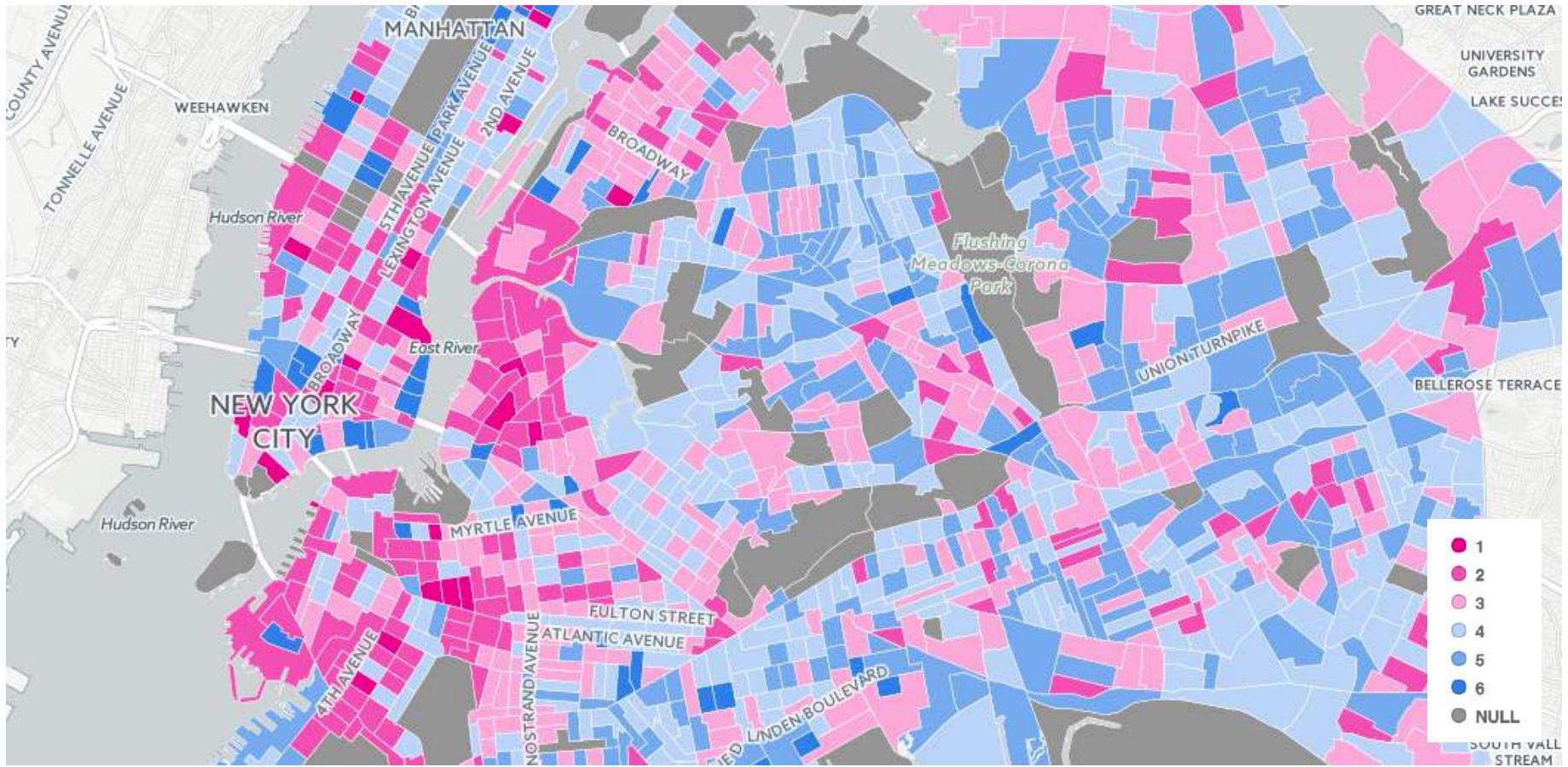
# New York City.

1980 - 1990.



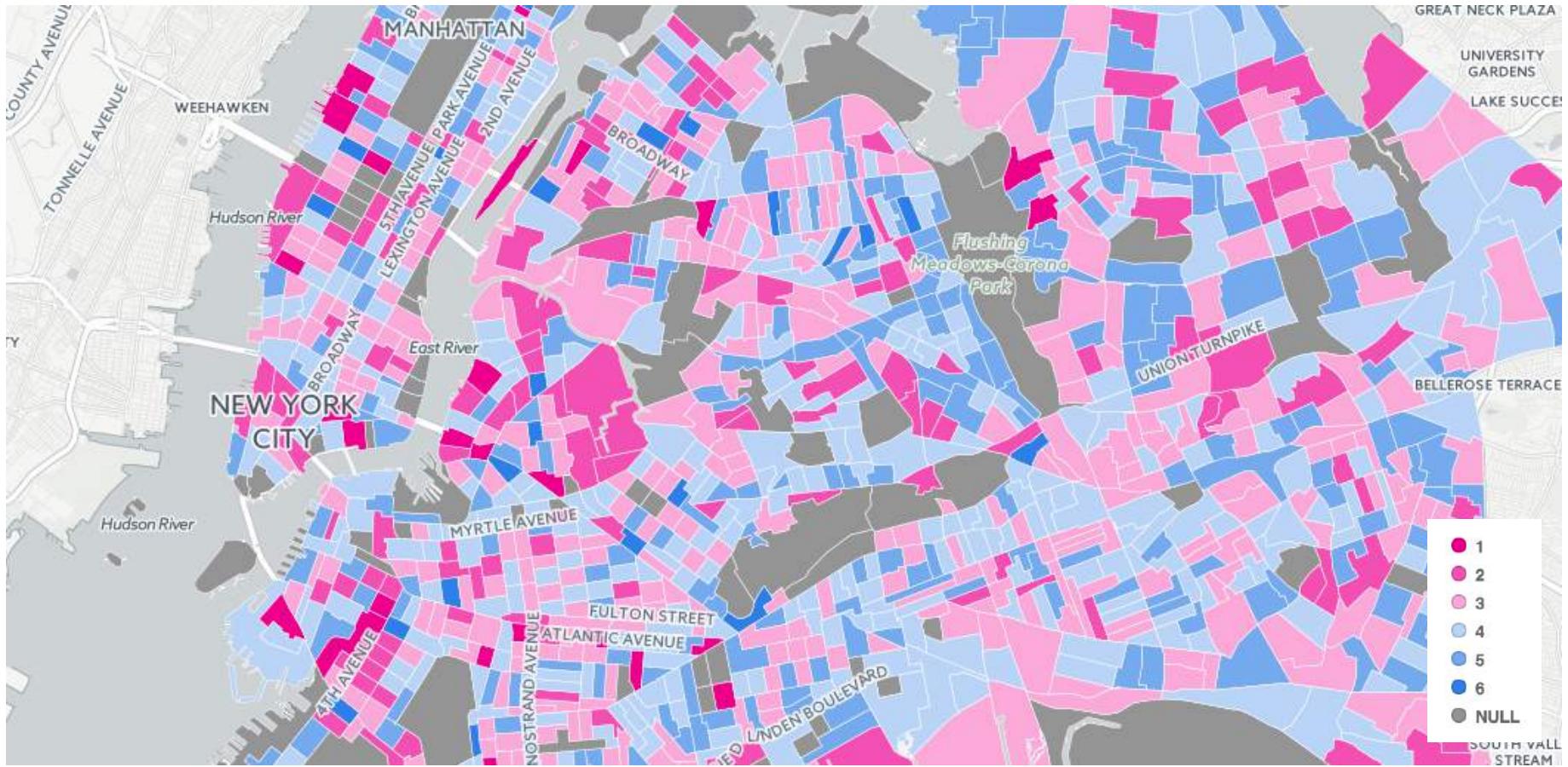
# New York City.

1990 - 2000.



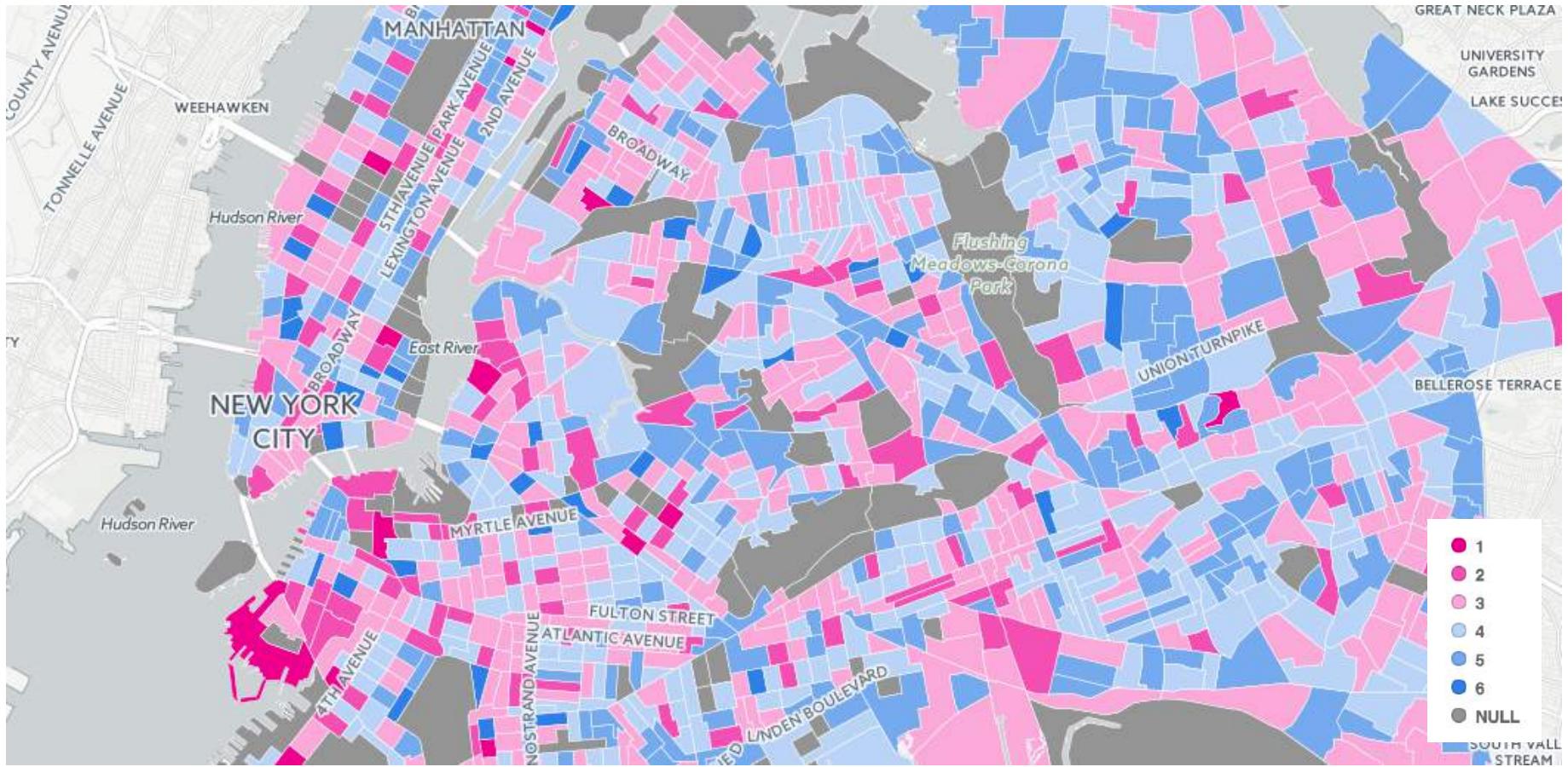
# New York City.

2000 - 2008.



# New York City.

2008 - 2010.



# New York City.

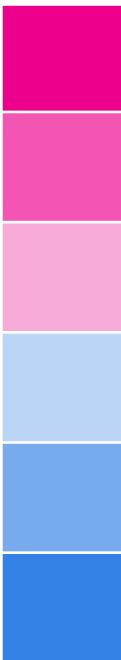
2010 - 2011.



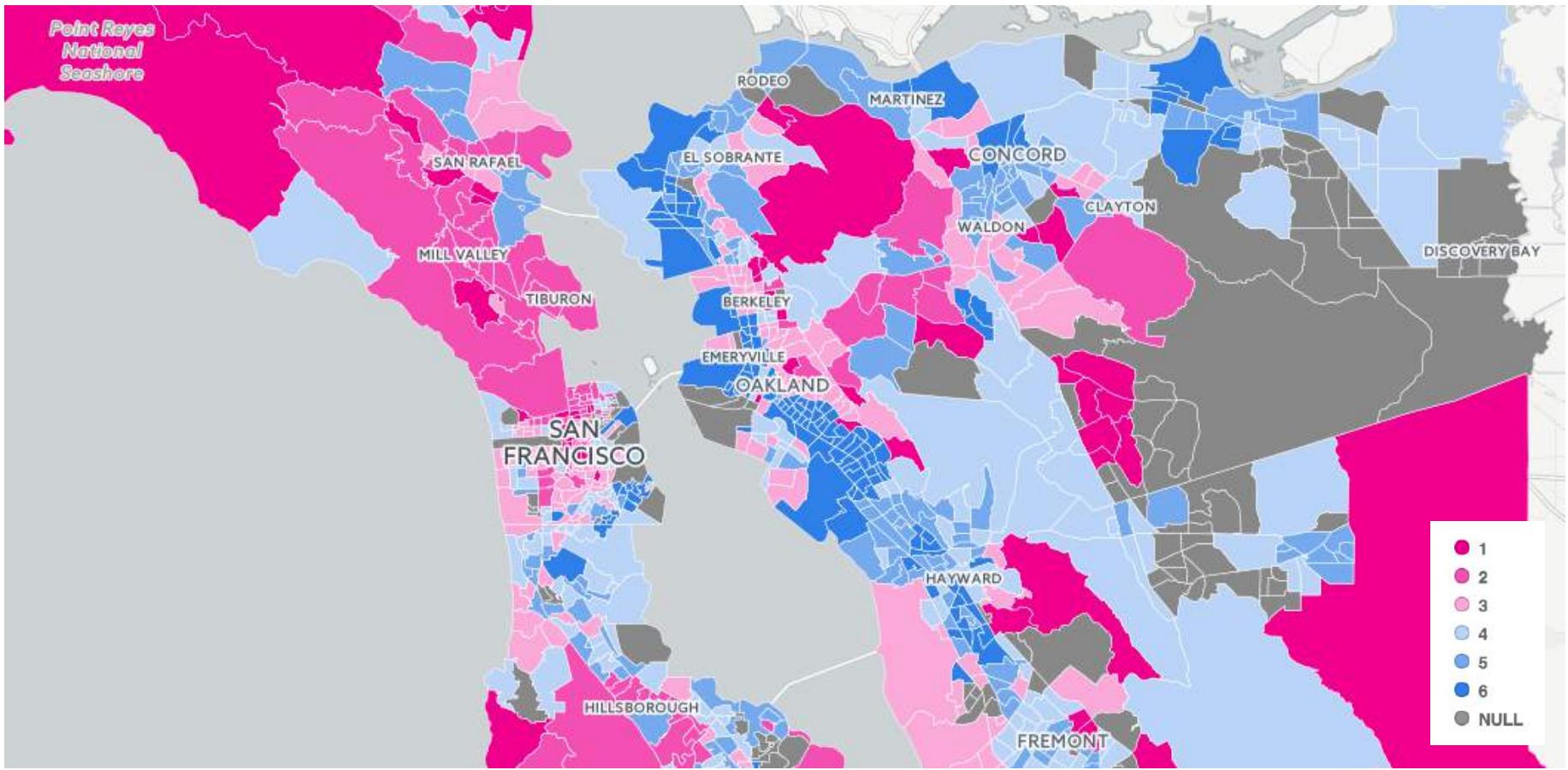
Bay Area.

# Six clusters in the Bay Area.

## Cluster

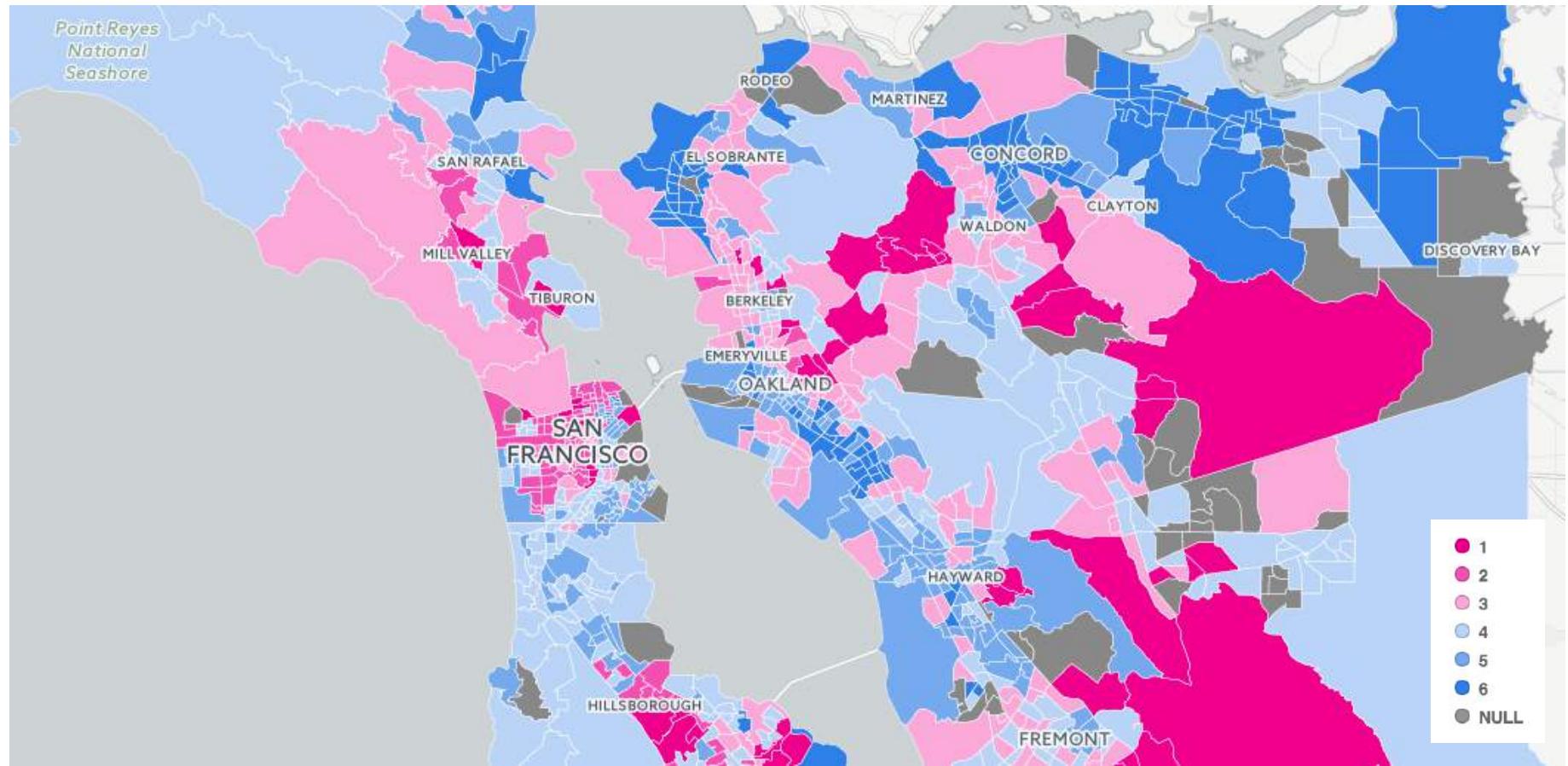


- 1      Rapid gentrification
- 2      Moderate gentrification
- 3      Slight gentrification
- 4      Slight decline
- 5      Moderate decline
- 6      Rapid decline



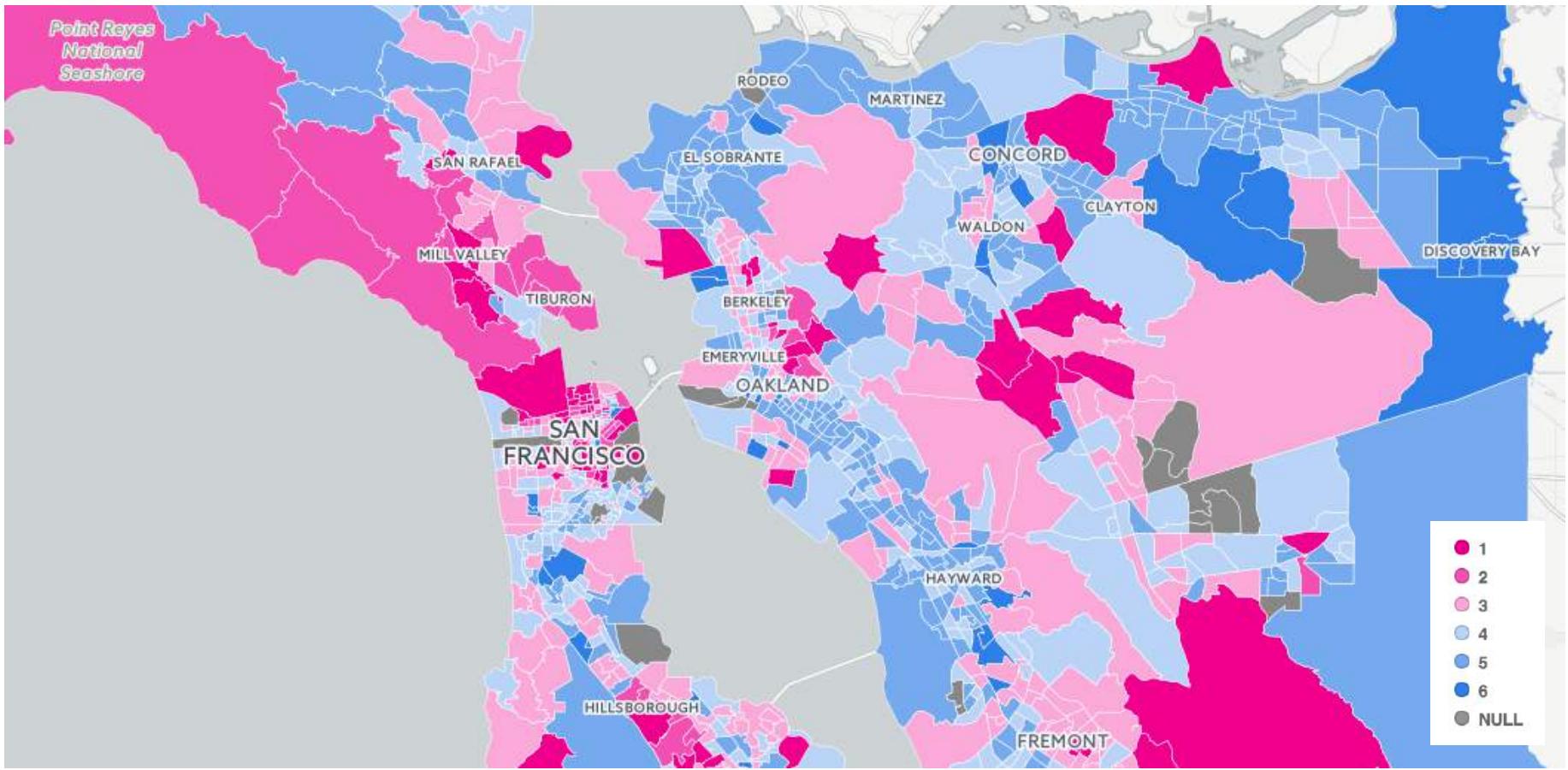
# Bay Area.

1970 - 1980.



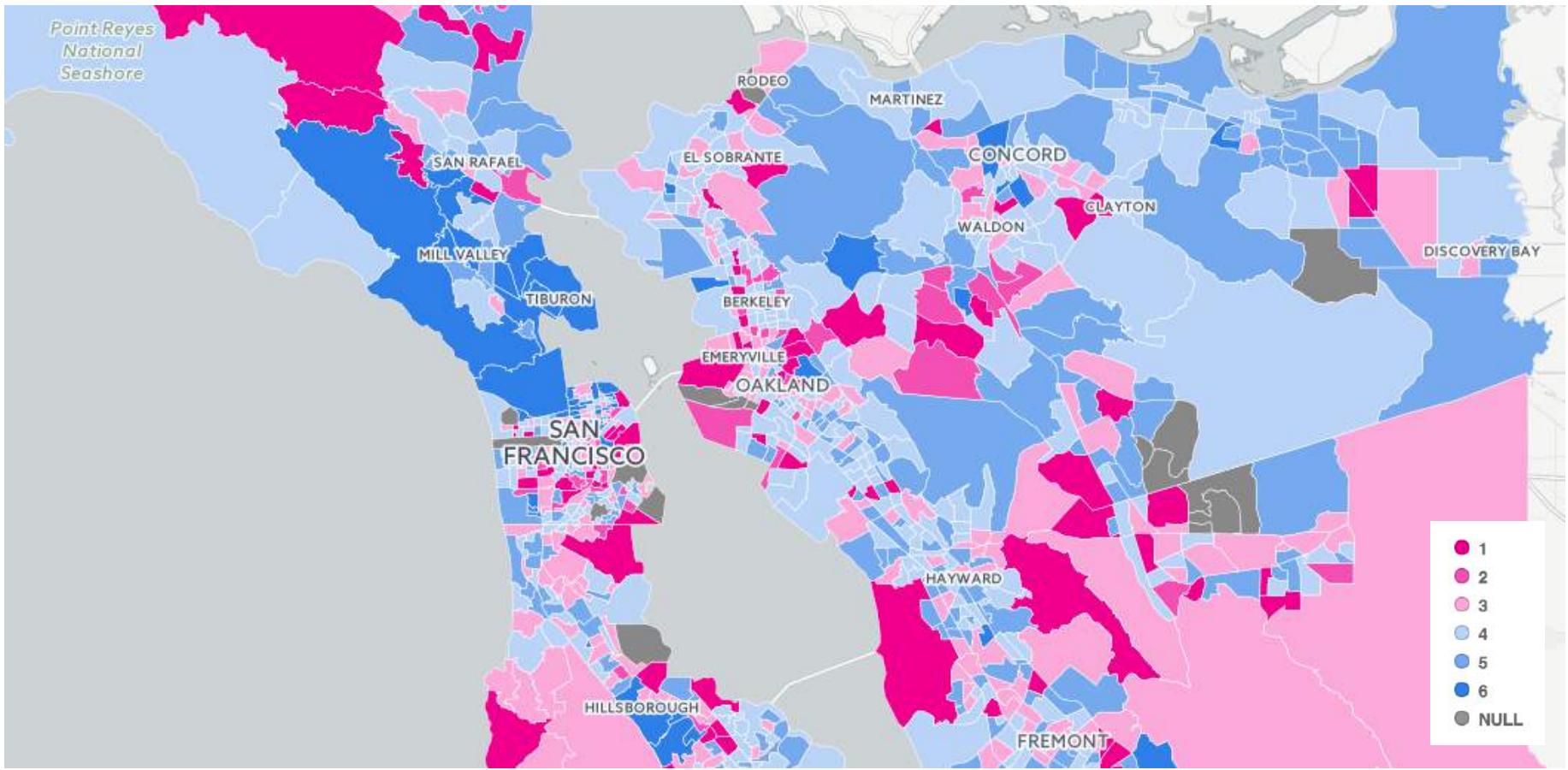
# Bay Area.

1980 - 1990.



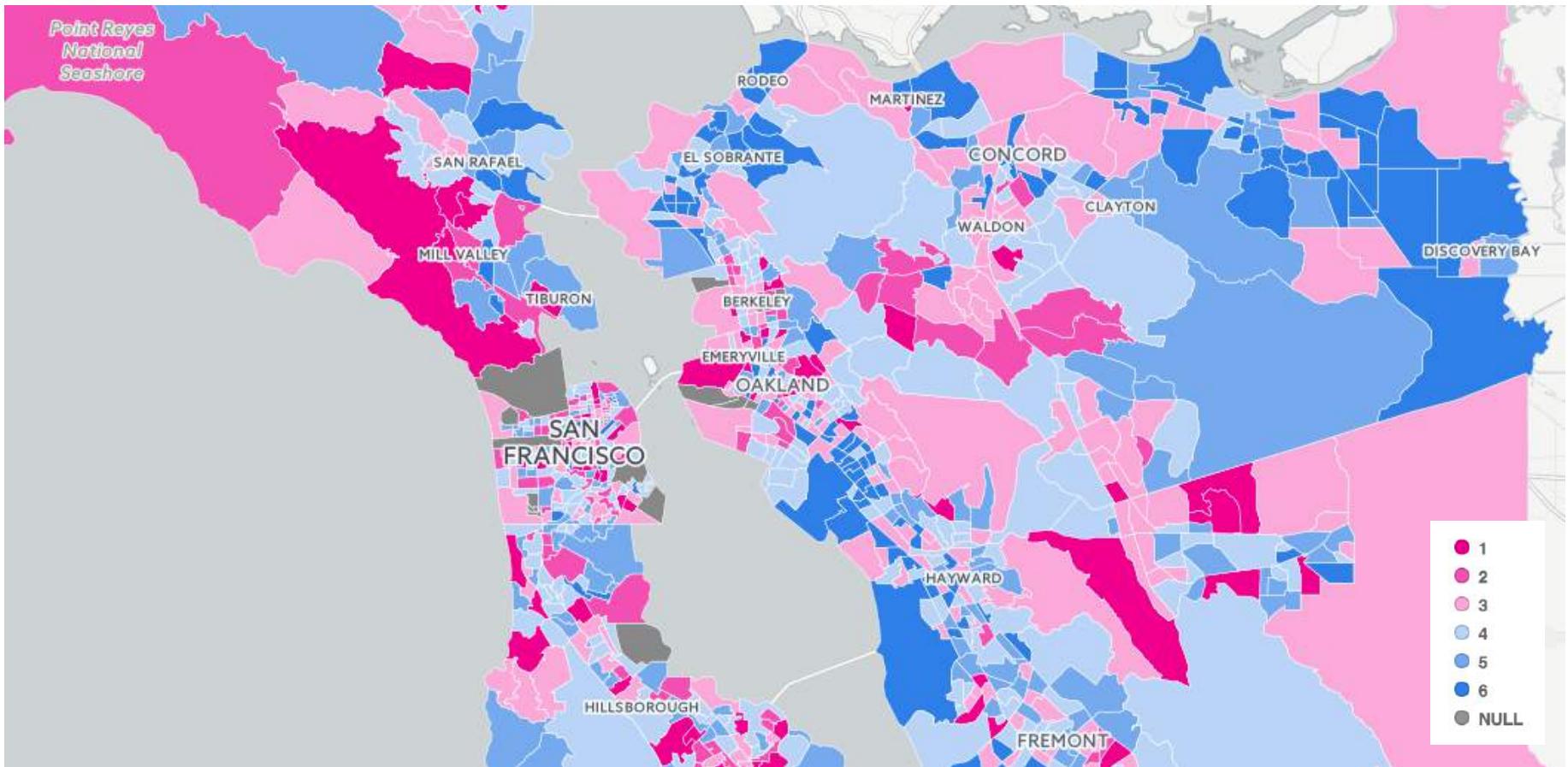
# Bay Area.

1990 - 2000.



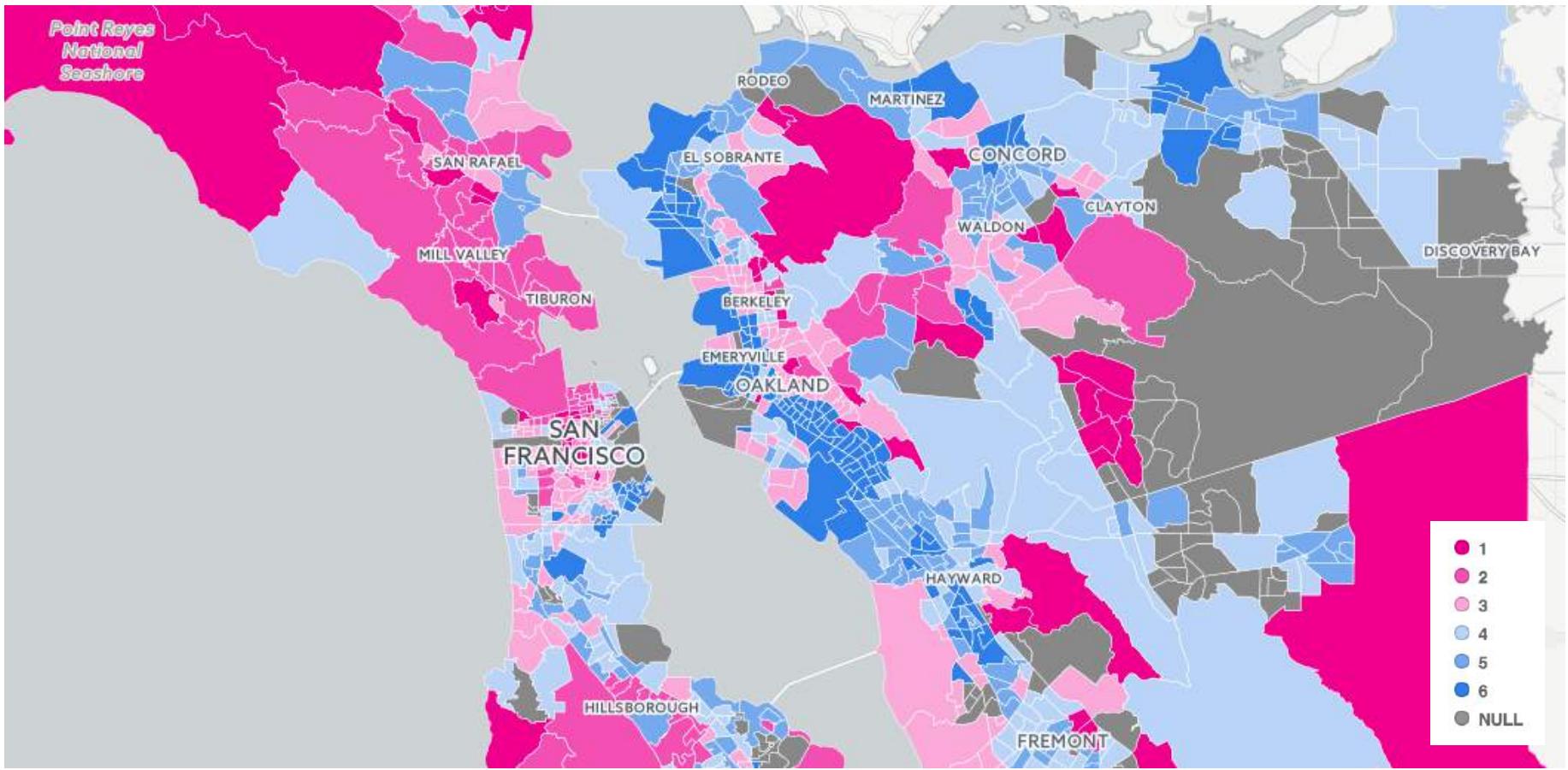
# Bay Area.

2000 - 2008.



# Bay Area.

2008 - 2010.



# Bay Area.

2010 - 2011.

**What about  
other sources of  
open data?**

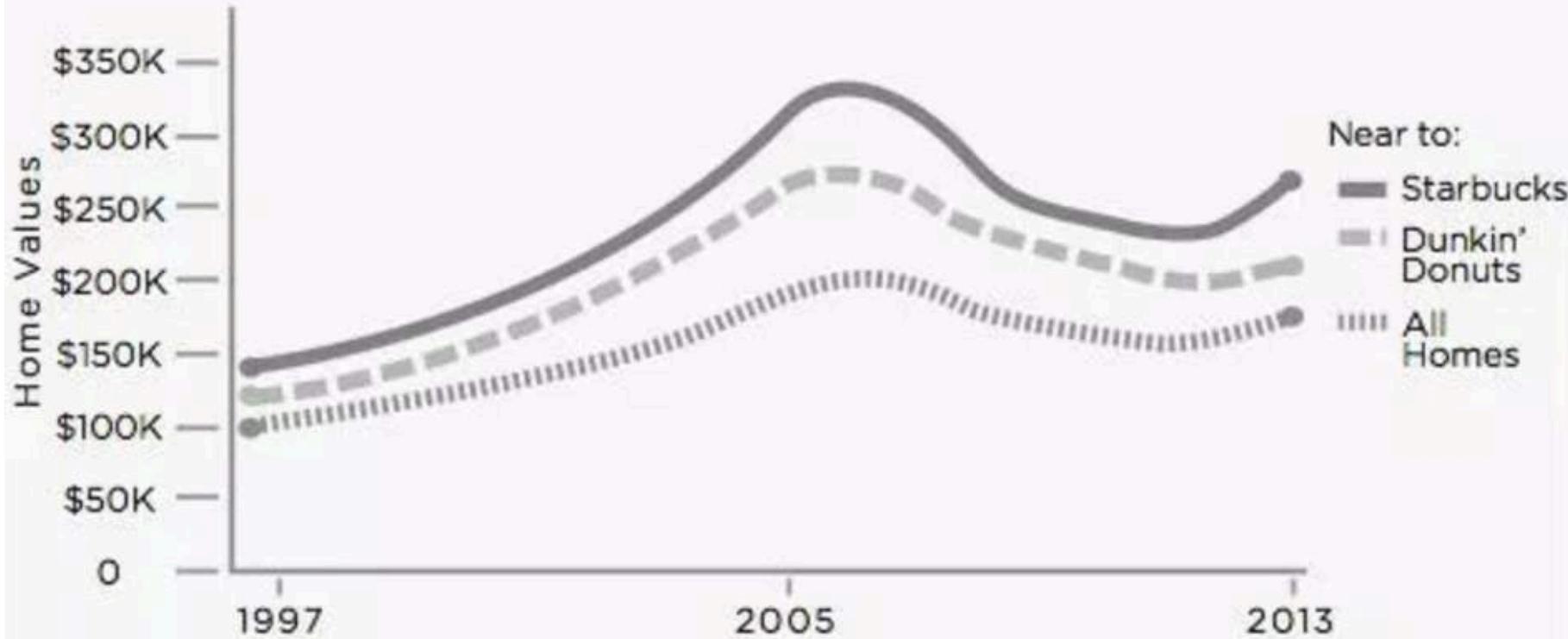
VENTI VALUE

## Confirmed: Starbucks knows the next hot neighborhood before everybody else does

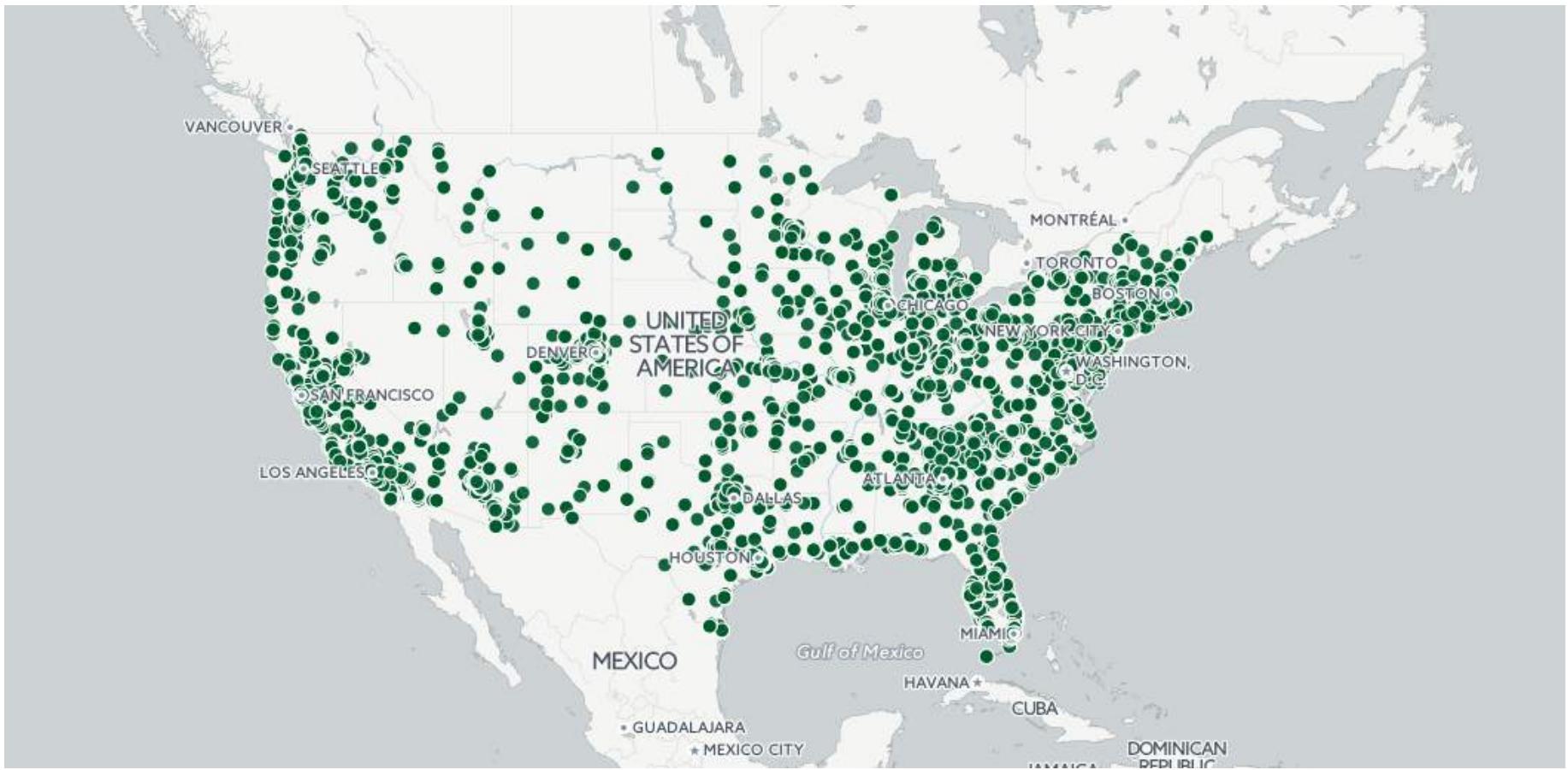




# The Frappuccino® Effect

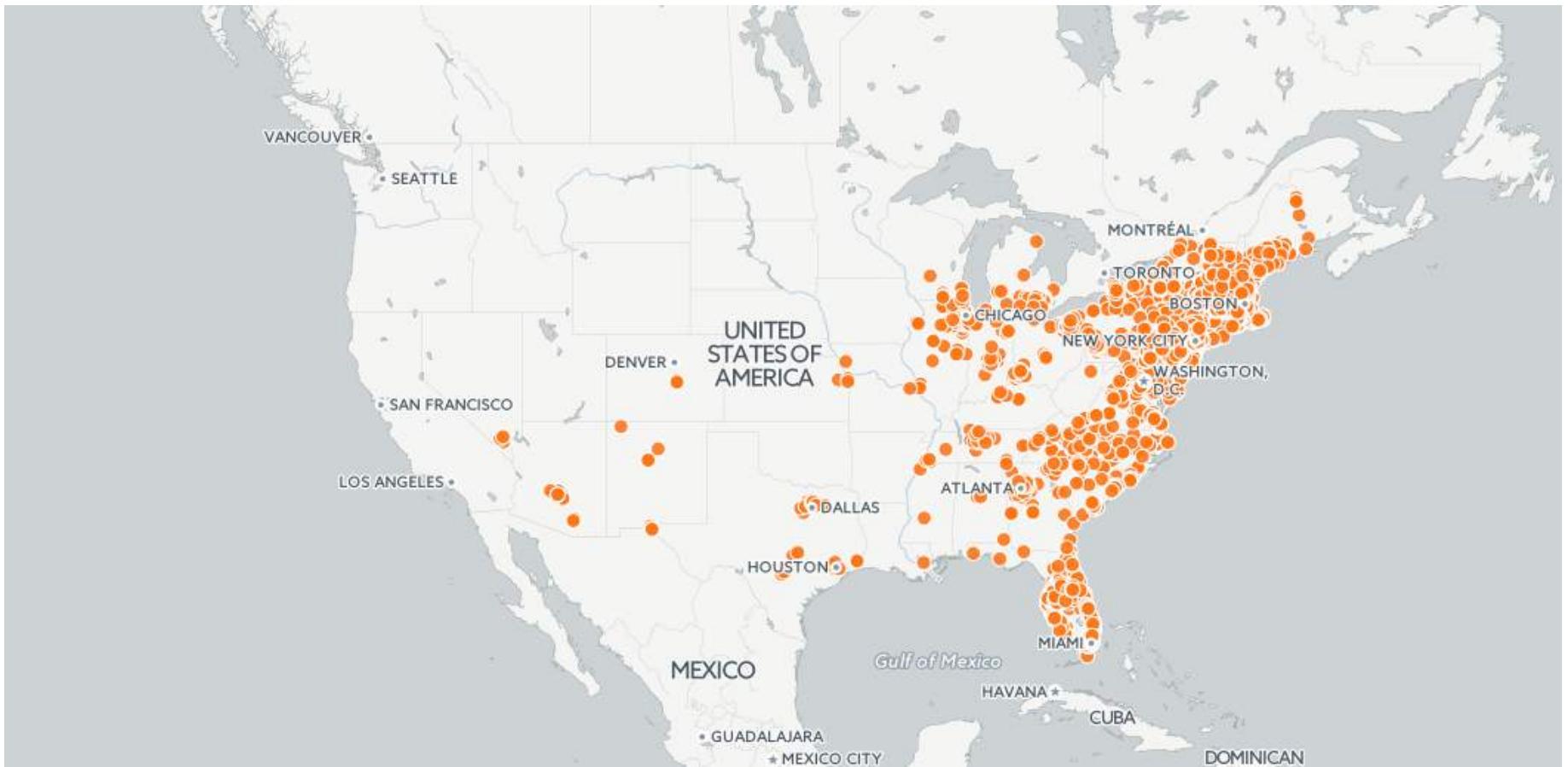


Starbucks targets affluent neighborhoods and makes them even more affluent.



# Starbuck locations.

Circa 2014.



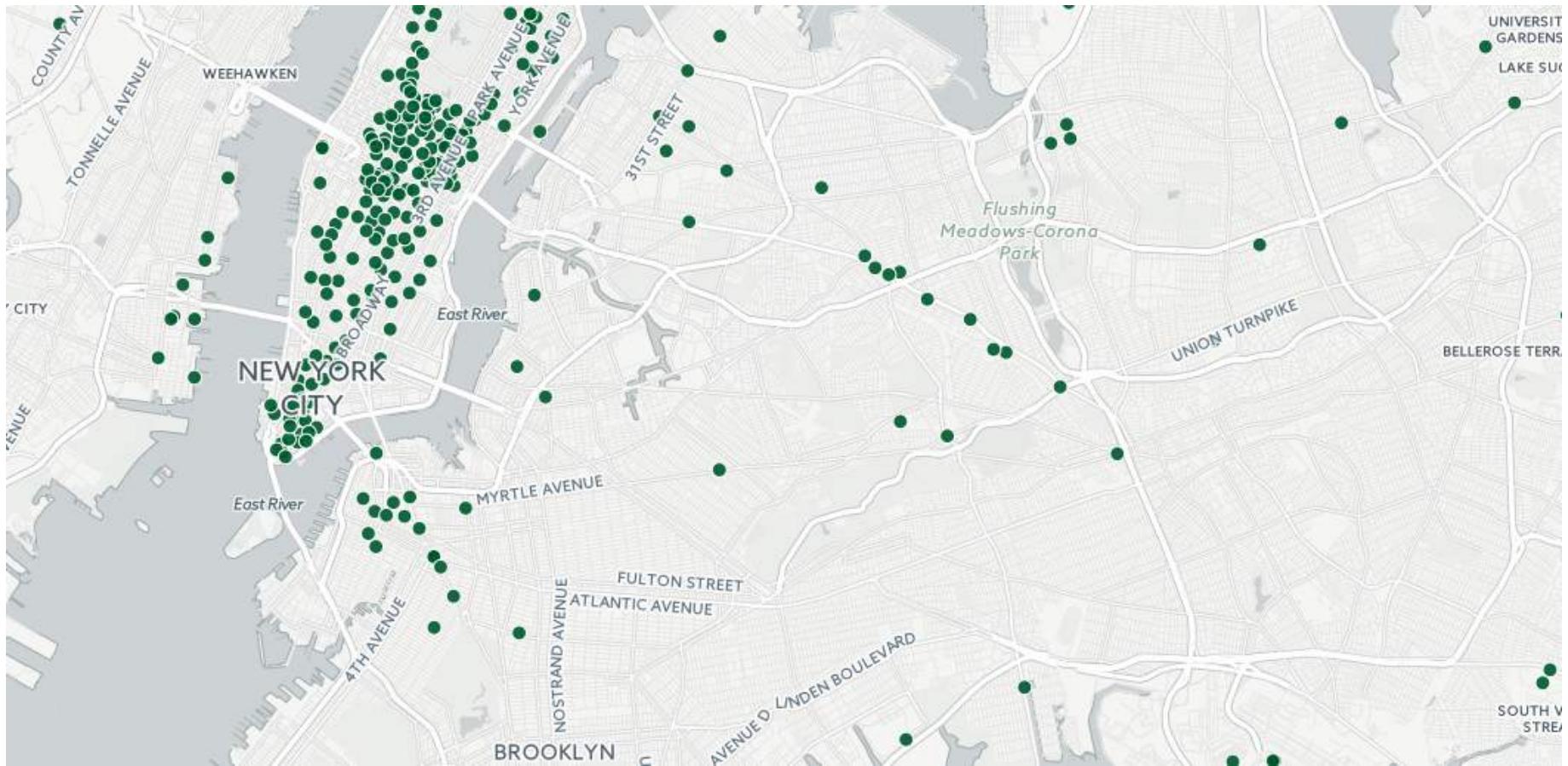
# Dunkin' Donuts locations.

Circa 2014.

An aerial photograph of the New York City skyline, featuring the One World Trade Center and the Brooklyn Bridge. The city is densely packed with skyscrapers, and the East River flows in the foreground.

New York City.





# Starbuck locations in NYC.

Circa 2014.

*Lots of Starbucks locations in NYC.*

321

locations  
(circa 2014).

*However.*

**91%**

of NYC census tracts have  
no Starbucks locations.

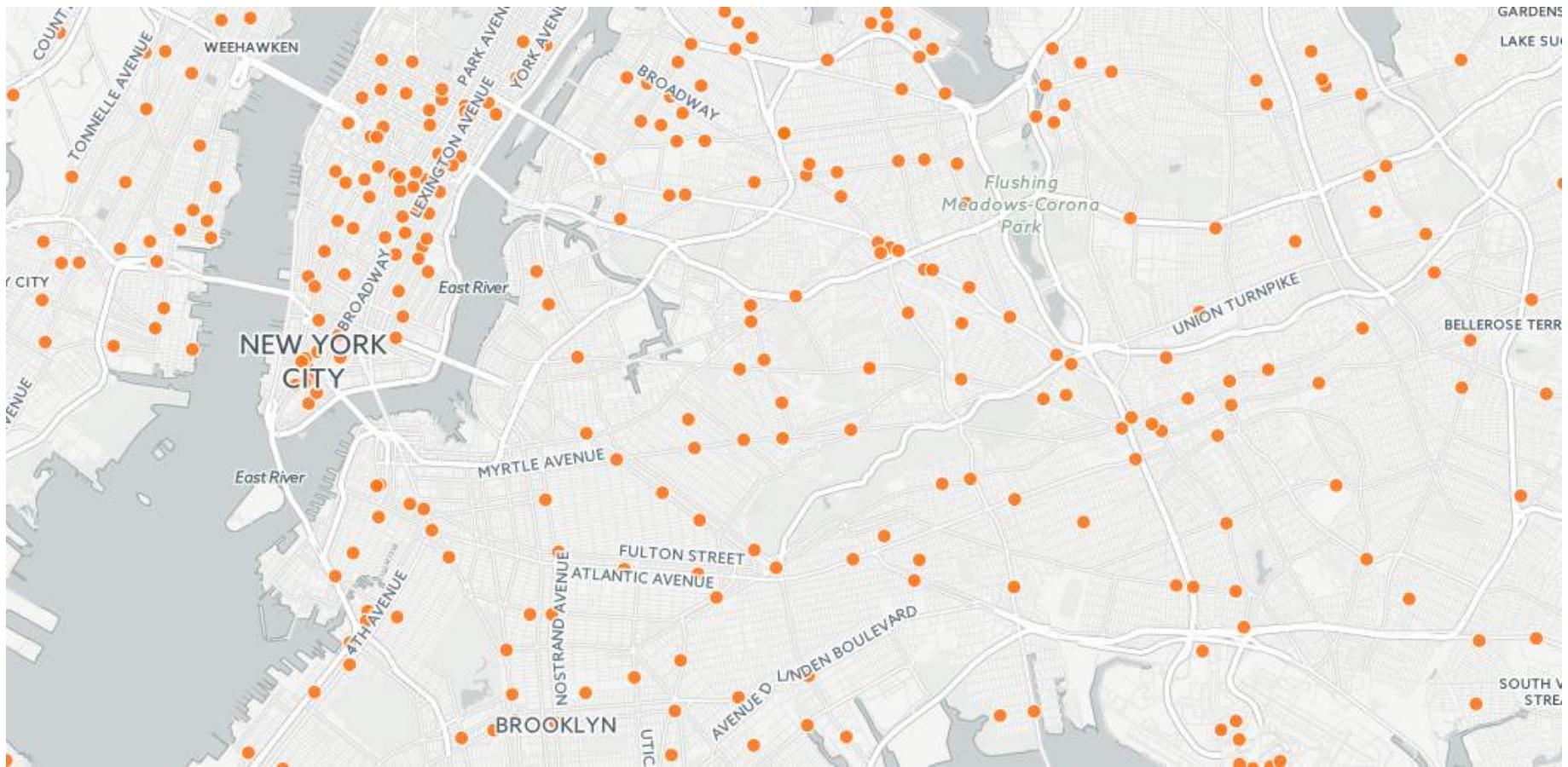
# Gentrification indicators in NYC.

Tracts with 2+ Starbucks locations tend to be gentrifying.



Number of locations (circa 2014)	Number of 2010 Census Tracts	Change in % of college graduates (2010 to 2011)	Change in median home value (2010 to 2011)	Change in household income (2010 to 2011)
0	1,981	+ .49%	-\$23,738	-\$1,236
1	129	+ .42%	-\$23,955	+\$149
2+	61	+1.99%	+\$5,674	+\$215

DUNKIN'  
DONUTS®



# Dunkin' Donuts locations in NYC.

Circa 2014.

*Even more Dunkin' Donuts in NYC!*

**449**

locations  
(circa 2014).

*But alas.*

**82%**

of NYC census tracts have  
no Dunkin' Donuts locations.

# Gentrification indicators in NYC.

Results are mixed with Dunkin' Donuts ...

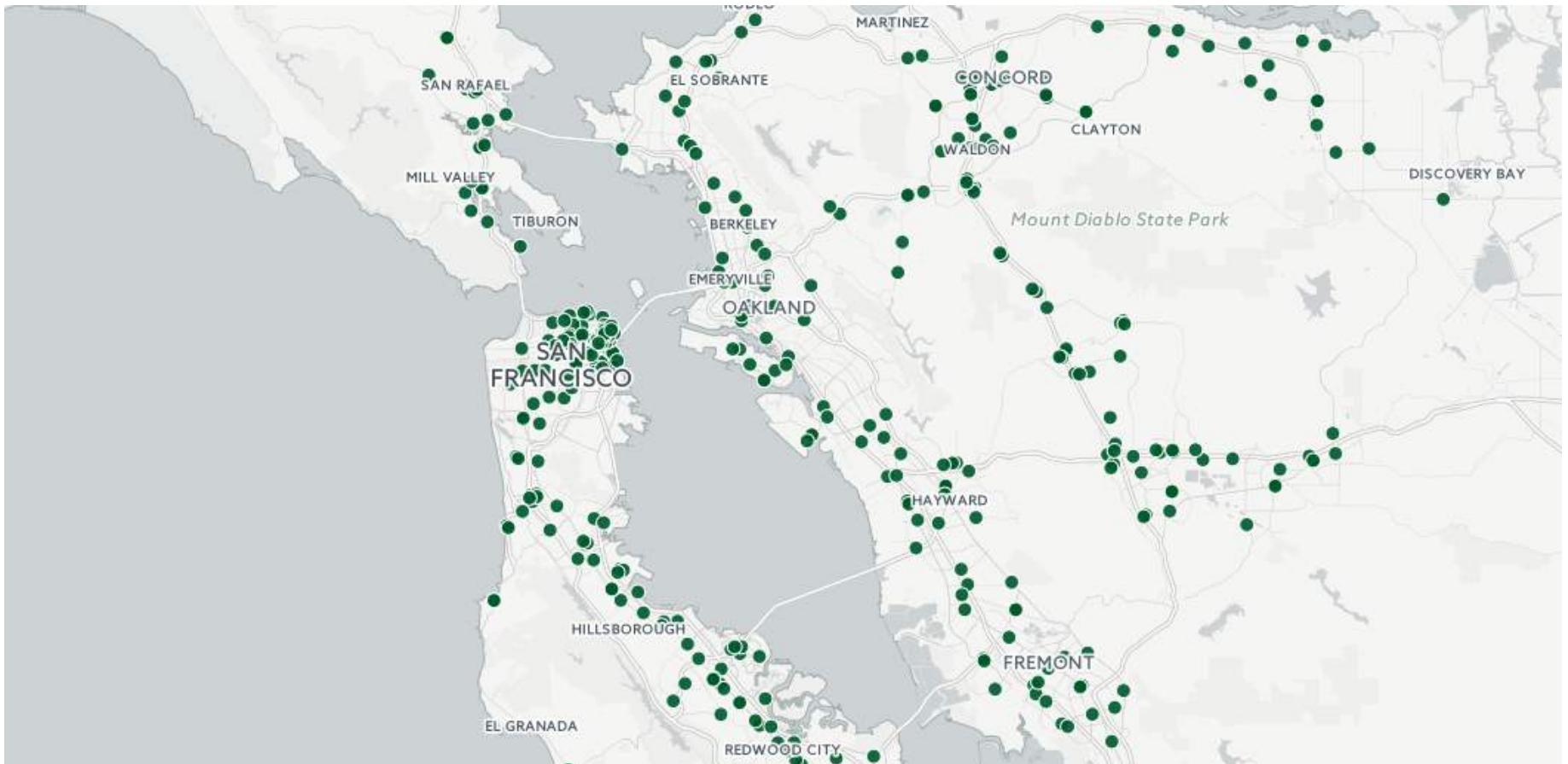


Number of locations (circa 2014)	Number of 2010 Census Tracts	Change in % of college graduates (2010 to 2011)	Change in median home value (2010 to 2011)	Change in household income (2010 to 2011)
0	1772	+.47%	-\$23,052	-\$1,213
1	361	+.81%	-\$18,805	-\$943
2+	38	+.79%	-\$13,830	+\$2,233



Bay Area.





# Starbucks locations in the Bay Area.

Circa 2014.

*Lots of Starbucks locations in the Bay Area.*

**385**

locations  
(circa 2014).

*However.*

**75%**

of Bay Area census tracts have  
no Starbucks locations.

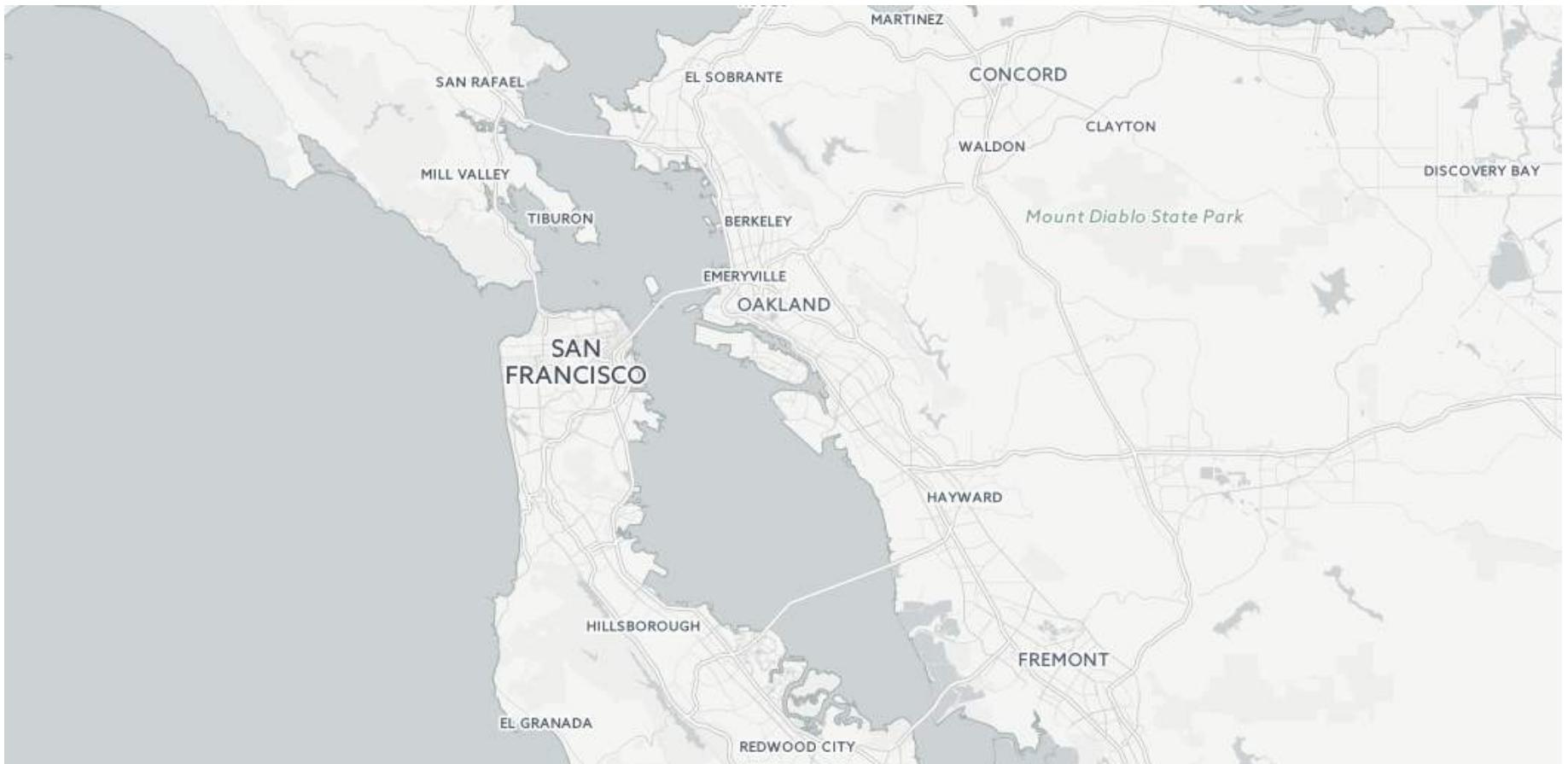
# Gentrification indicators in the Bay Area.

Decline is somewhat slower in tracts with 2+ Starbucks locations.



Number of locations (circa 2014)	Number of 2010 Census Tracts	Change in % of college graduates (2010 to 2011)	Change in median home value (2010 to 2011)	Change in household income (2010 to 2011)
0	734	+.39%	-\$28,708	-\$1,528
1	168	+.62%	-\$29,175	-\$2,396
2+	78	+.59%	-\$26,030	-\$1,031

DUNKIN'  
DONUTS®



# Dunkin' Donuts locations in the Bay Area.

Circa 2014.

*Lots of Dunkin Donuts in the Bay Area!*

0

locations  
(circa 2014).

*Therefore.*

**100%**

of Bay Area census tracts have  
no Dunkin' Donuts locations.

# Gentrification indicators in the Bay Area.

Can't do much with Dunkin' Donuts locations since they don't exist.



Number of locations (circa 2014)	Number of 2010 Census Tracts	Change in % of college graduates (2010 to 2011)	Change in median home value (2010 to 2011)	Change in household income (2010 to 2011)
0	980	+.45%	-\$28,581	-\$1,639
1	0	—	—	—
2+	0	—	—	—

**What's next?**

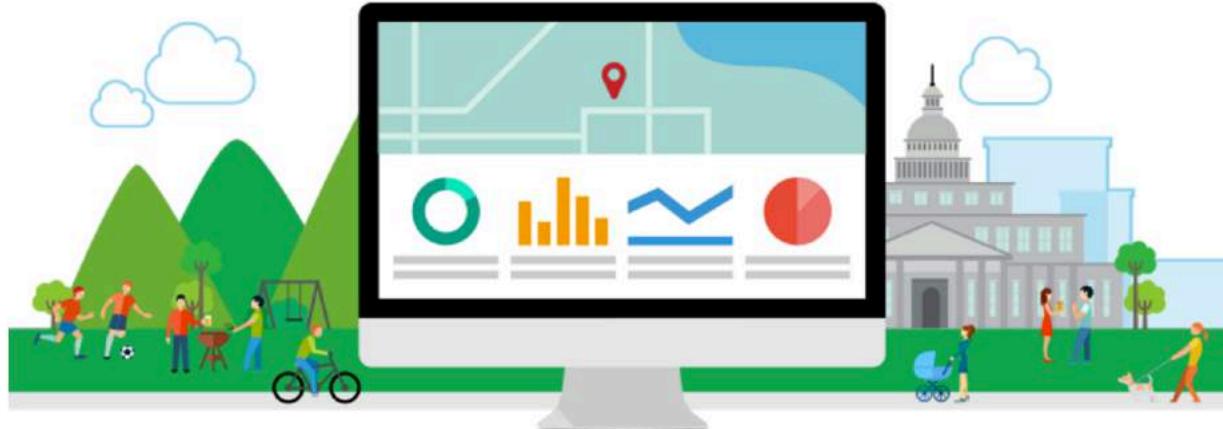
**Tons of open data sources and ways  
to scrape the web, but ...**

not much standardization across time.

# **We need a dynamic and open data repository**

that collects data across time in a standardized way.

**Time is ignored in many open data sources, but we're making progress.**



# Socrata Open Data

Putting insights into the hands of the public



# City Record Online

The newly-expanded City Record Online (CROL) is a fully searchable database for all the notices contained in the City Record newspaper – including schedules for public hearings, land-sales, and contract awards. All this data is now also available for download on the Open Data Portal, increasing New Yorkers access to important government information. [Click here to view the City Record Online data set.](#)

View More Stories

● o o o

Search



[Click here to view the NYC OpenData dashboard](#)

# SF OpenData

[About](#)[Data](#) ▾[Developers](#)[Showcase](#)[Help](#) ▾

Welcome to SF OpenData! SF OpenData is the central clearinghouse for data published by the City and County of San Francisco and is part of the broader open data program, [DataSF](#). Explore, view, and download our data. Developers - check out our developer page for tips on API access and use. Read more on our About page.





## OPEN GOVERNMENT

Since his first full day in office, President Obama has prioritized making government more open and accountable and has [taken substantial steps](#) to increase citizen participation, collaboration, and transparency in government.

Data.gov, the central site for U.S. Government data, is an important part of the Administration's overall [effort](#) to open government.

## Open Data in the United States

A large number of cities, counties, and states have open data sites.

- [Cities.Data.gov](#)
- [Counties.Data.gov](#)
- [States.Data.gov](#)

**Let's do this!**

A portrait photograph of Scott Wolf, a man with dark hair and a beard, wearing a black button-down shirt, smiling at the camera.

# Thank you! Questions?

## Scott Wolf

- ✉️ [swolf@hugeinc.com](mailto:swolf@hugeinc.com)
- 🐦 [@scott\\_t\\_wolf](https://twitter.com/scott_t_wolf)
- linkedin [linkedin.com/in/scottwolf1](https://linkedin.com/in/scottwolf1)
- github [github.com/scottwolf](https://github.com/scottwolf)

## Huge

- 🔗 [hugeinc.com](https://hugeinc.com)
- 🐦 [@hugeinc](https://twitter.com/hugeinc)

HUGE

Done.

# Appendix.

*Appendix.*

# **z-scores.**

# Time to learn about z-scores!

z-scores produce a standardized metric with an average of 0 and a standard deviation of 1.

$$z = \frac{\text{score} - \text{average}}{\text{standard deviation}}$$

**Positive** values are above average.

**Negative** values are below average.

**Average** scores are around zero.

*In a normal, bell-shaped distribution ...*

**68%**

of z-scores are between  
-1.0 and +1.0.

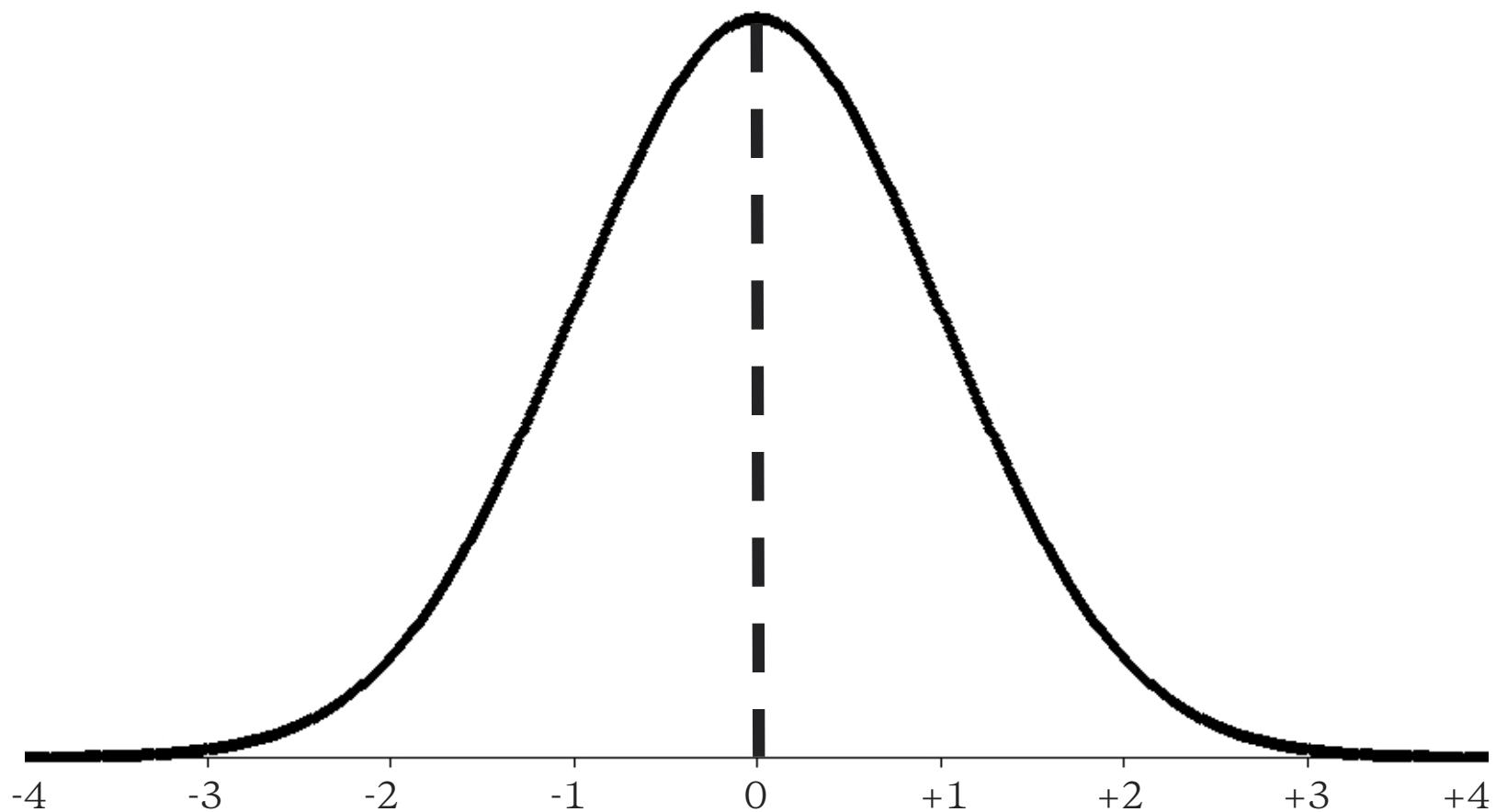
**95%**

of z-scores are between  
-2.0 and +2.0.

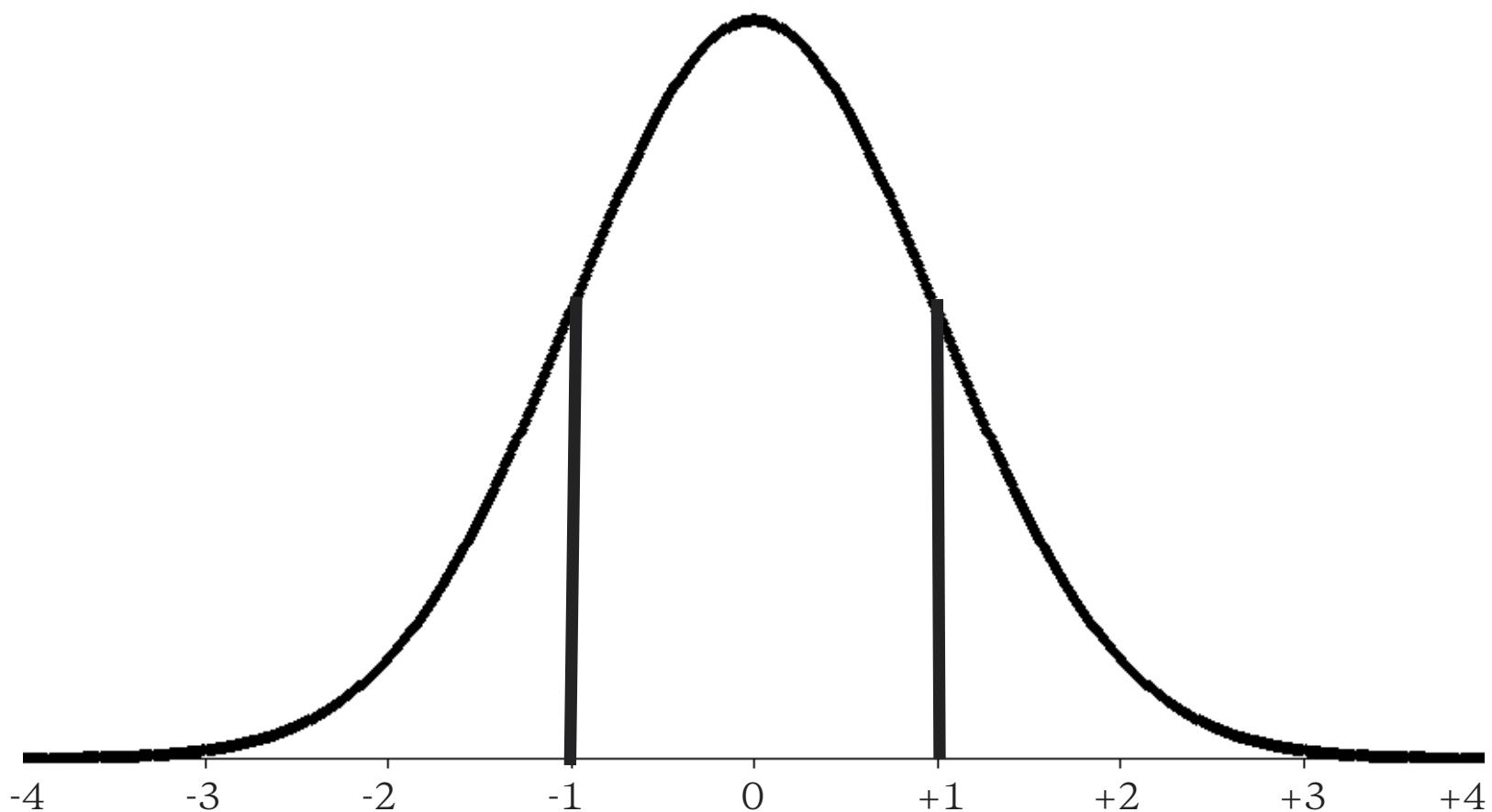
**99.7%**

of z-scores are between  
-3.0 and +3.0.

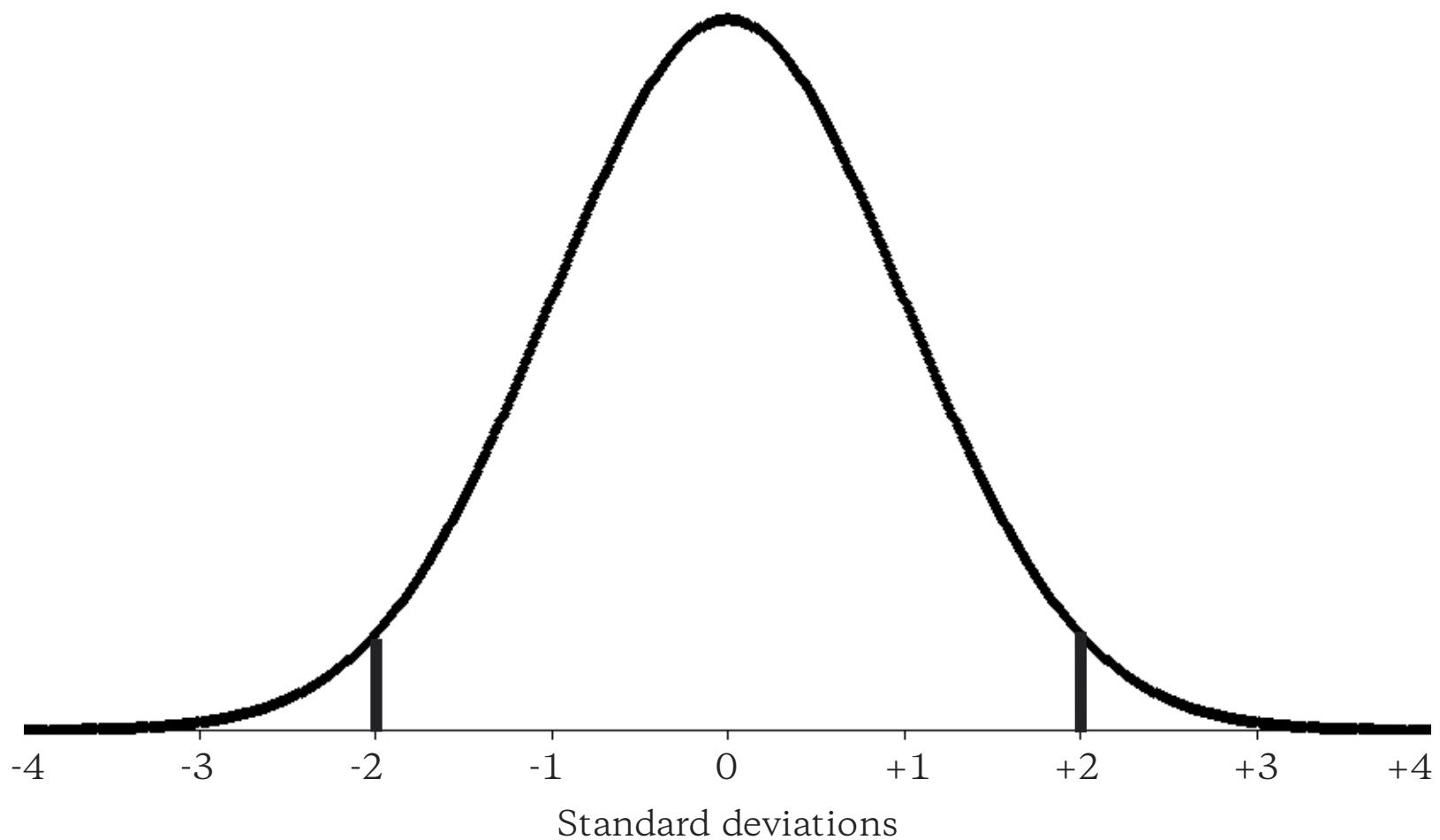
**Mean  
(average)**



**68% of z-scores fall within  
1 standard deviation  
of the mean**



**95% of z-scores fall within  
2 standard deviations  
of the mean**



**99.7% of z-scores fall within  
3 standard deviations  
of the mean**

