

[Static] Data Visualization



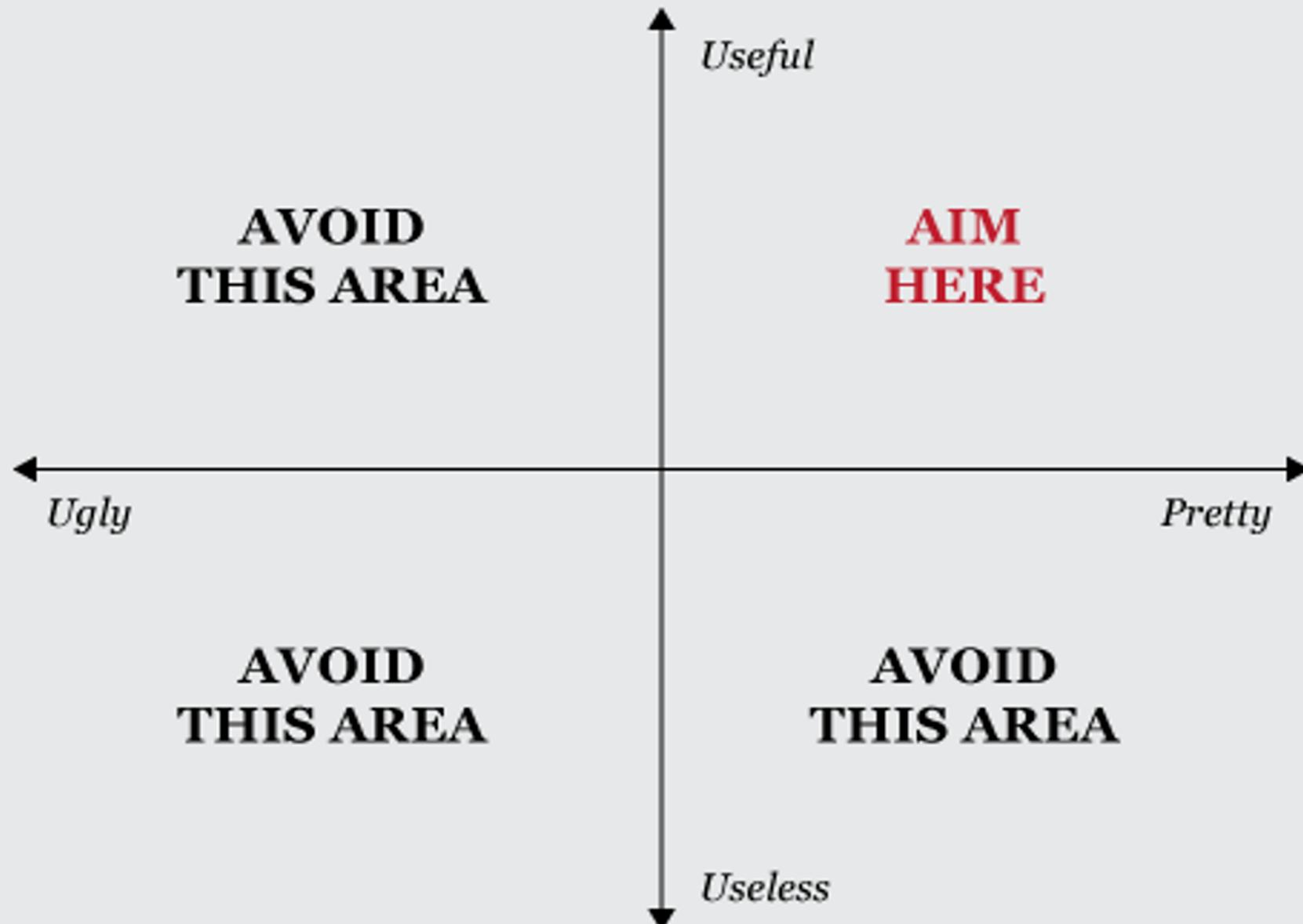
Introduction to Urban Data Analytics
Manuel Santana Palacios
June 03, 2020

Agenda for Today

- Four simple rules for effective visualization
- Lab 3 [Part 1]: Census Data [... cont.] + WordPress
- Lab 3 [Part 2]: MoE + Social Explorer

Rules for Effective Visualizations

Rules for Effective and Ethical Visualizations



Rule 1 The Chart Should Tell a Story

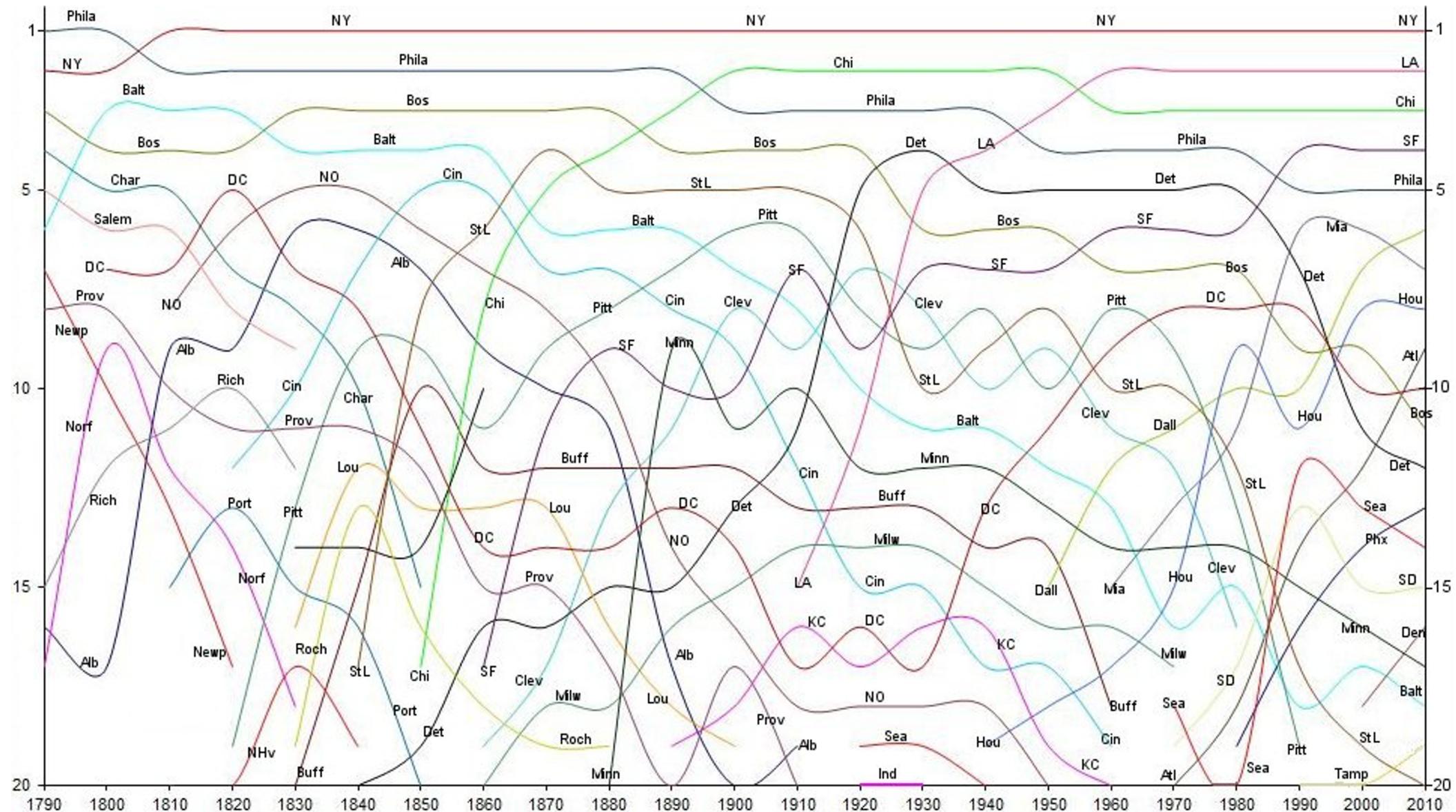
- Graphics should be **clear** on their own
- The depictions should **enable meaningful comparison**
- The chart should yield **insights beyond the text**
- **Labels** should be used to defeat ambiguity

CITY AND RURAL POPULATION.
1890.

W. E. B. Du Bois' Hand-Drawn Infographics of African-American Life (1900)

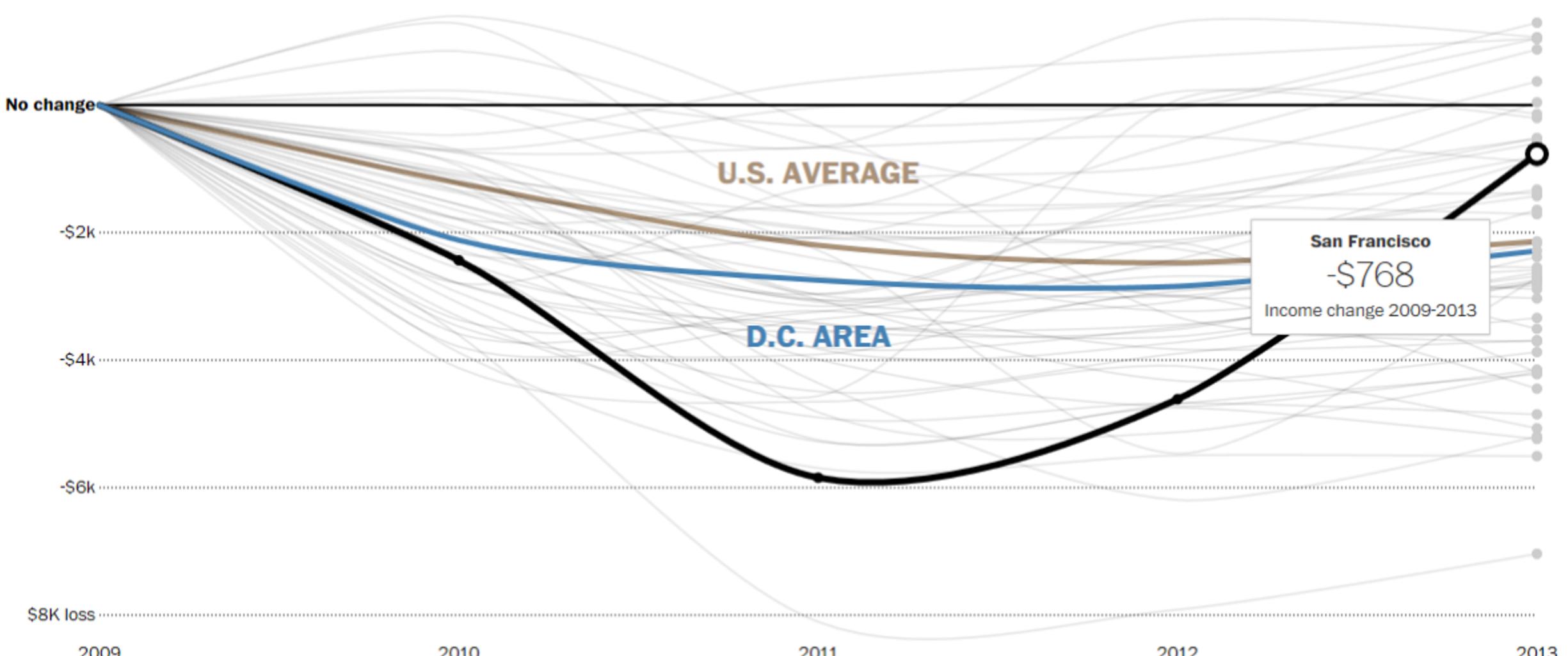
“Created by Du Bois and his students at Atlanta, the charts, many of which focus on economic life in Georgia, managed to condense an enormous amount of data into a set of aesthetically daring and **easily digestible** visualizations.”

Top 20 Metropolitan Areas, 1790 - 2010



Median Household Income

2K gain

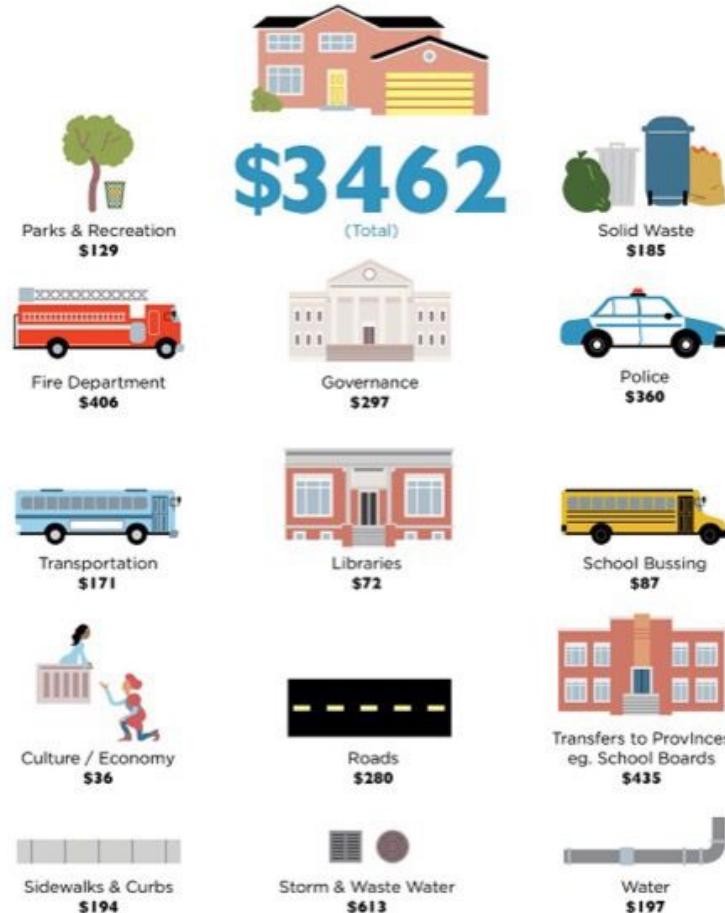


Source: ACS 1-Year Estimates

Comparing households

Suburban

City's Annual Cost, per Household



For more data and more reports, visit thecostofsprawl.com
Data based on Halifax Regional Municipality

Urban

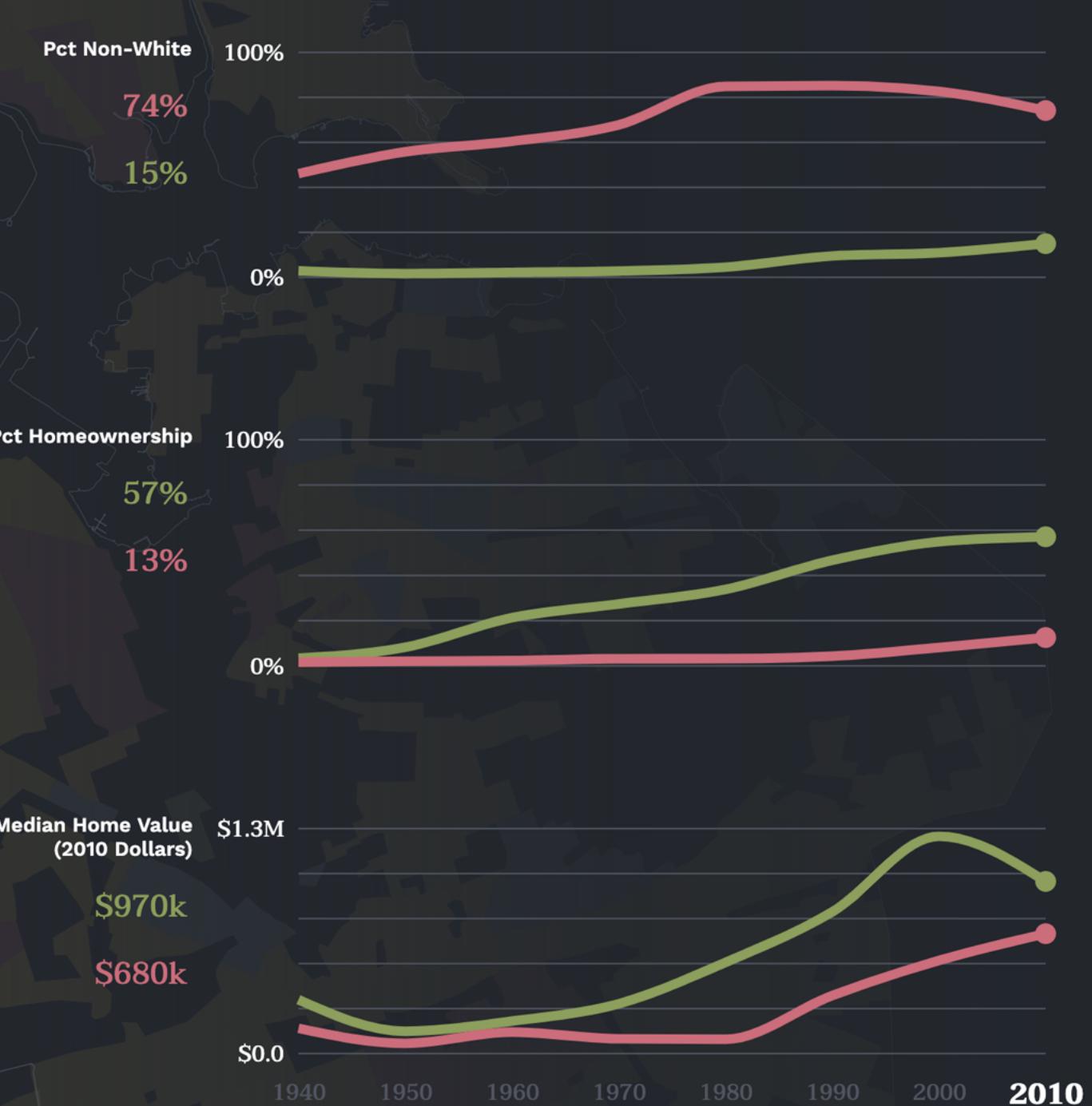
City's Annual Cost, per Household



For more data and more reports, visit thecostofsprawl.com
Data based on Halifax Regional Municipality

Homeownership rates in **Harlem** lag behind the **Upper East Side**

Homeownership has consistently grown in greenlined Upper East Side while staying stagnant in redlined, majority non-white Harlem



Rule 2 The Chart Should Have Graphical Integrity

- Basically, the chart shouldn't use visuals to lie about data

Comparative Annual Cost per Capita for care of Insane in
Pittsburgh City Homes and Pennsylvania State Hospitals.

\$147



South Mountain

\$172



Pittsburgh

\$198



Harrisburg

\$213



Norrtown

\$214

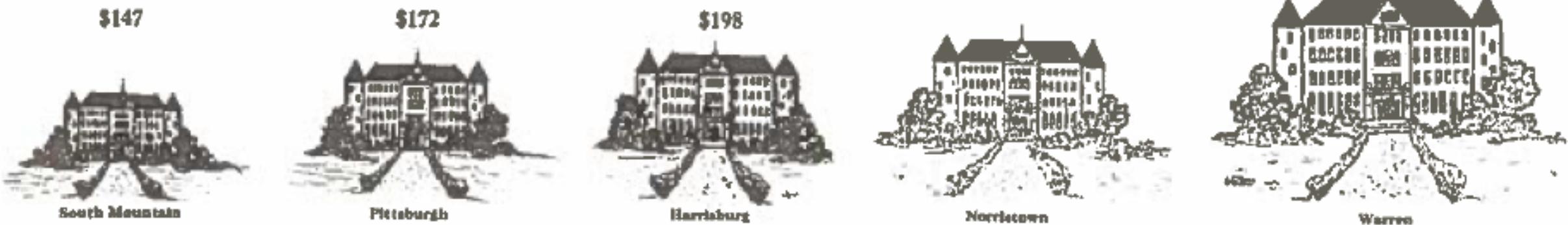


Warren

Pittsburgh Civic Commission, *Report on Expenditures of the Department of Charities* (Pittsburgh, 1911), p. 7.

Graphic Distortion of Data

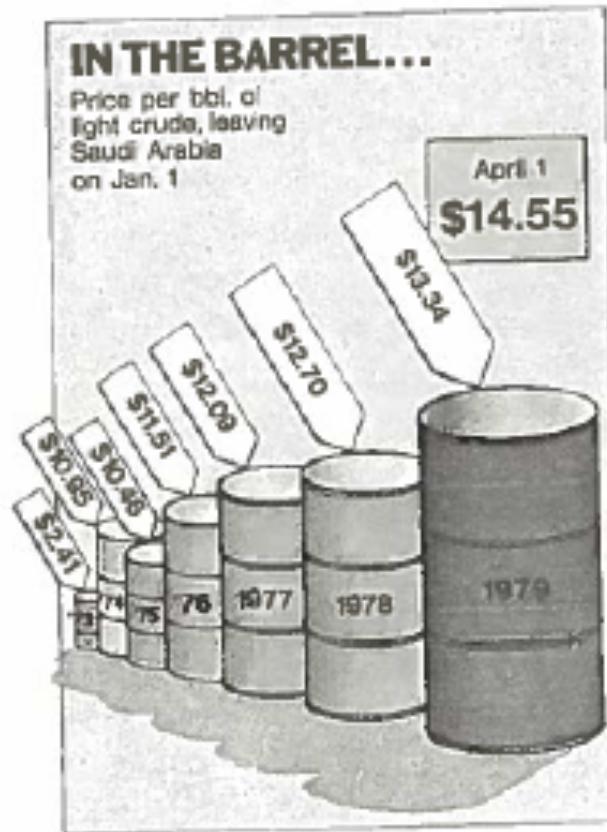
Comparative Annual Cost per Capita for care of Insane in Pittsburgh City Homes and Pennsylvania State Hospitals.



Pittsburgh Civic Commission, *Report on Expenditures of the Department of Charities* (Pittsburgh, 1911), p. 7.

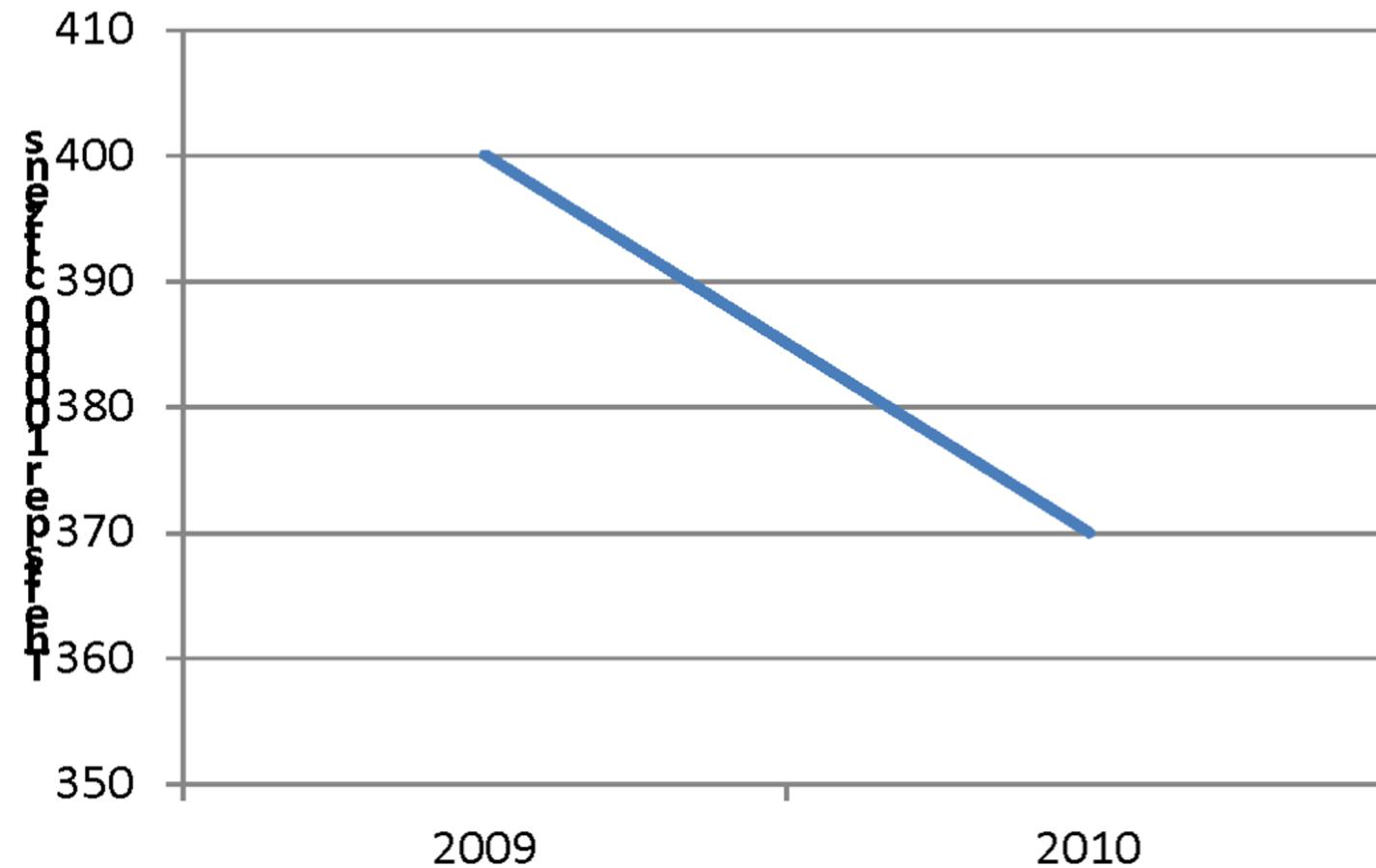
Graphic Distortion of Data and the Lie Factor

Design variation infected similar graphics in other publications. Here an increase of 454 percent is depicted as an increase of 4,280 percent, for a Lie Factor of 9.4:



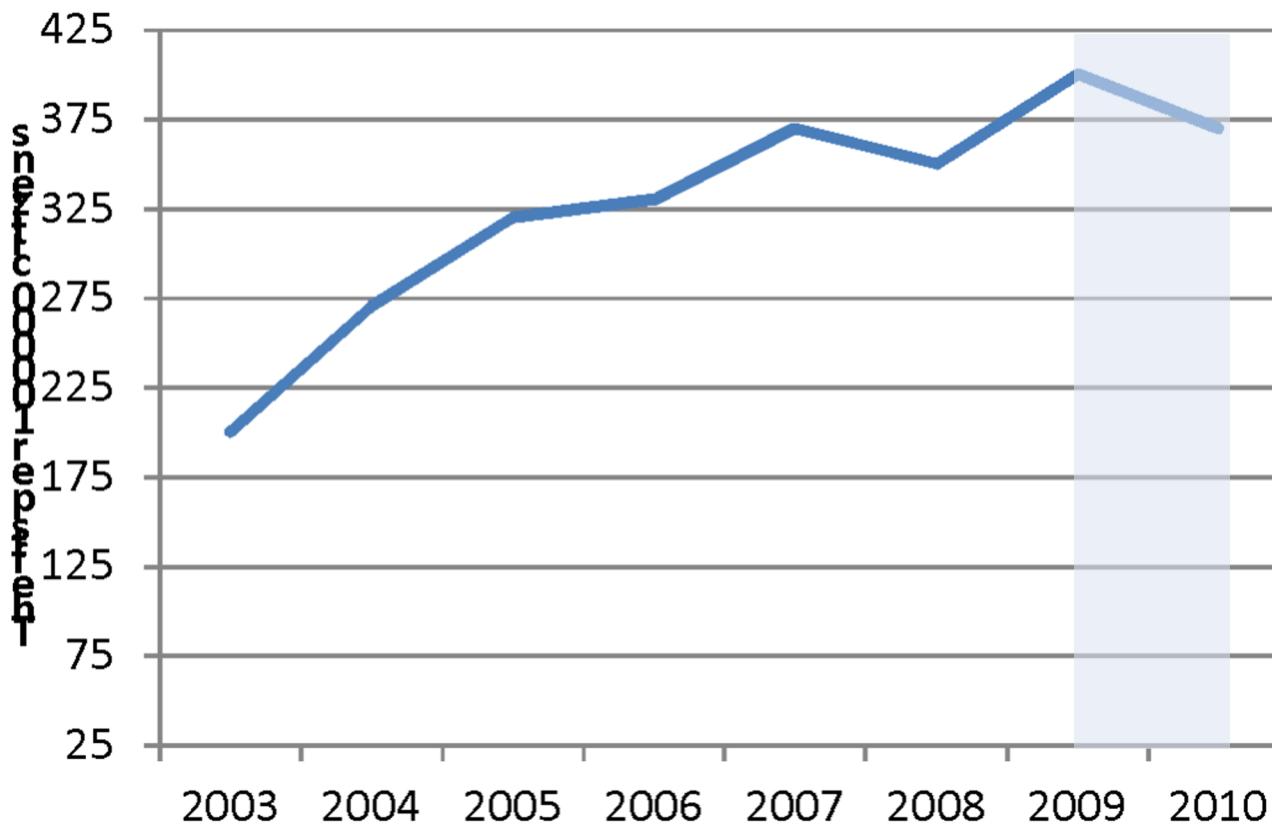
Time, April 9, 1979, p. 57.

Hypothetical City Crime

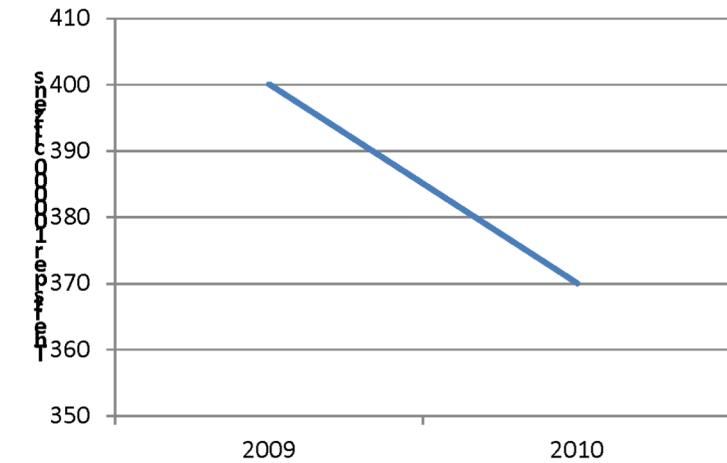


Design and Data Variation

Hypothetical City Crime



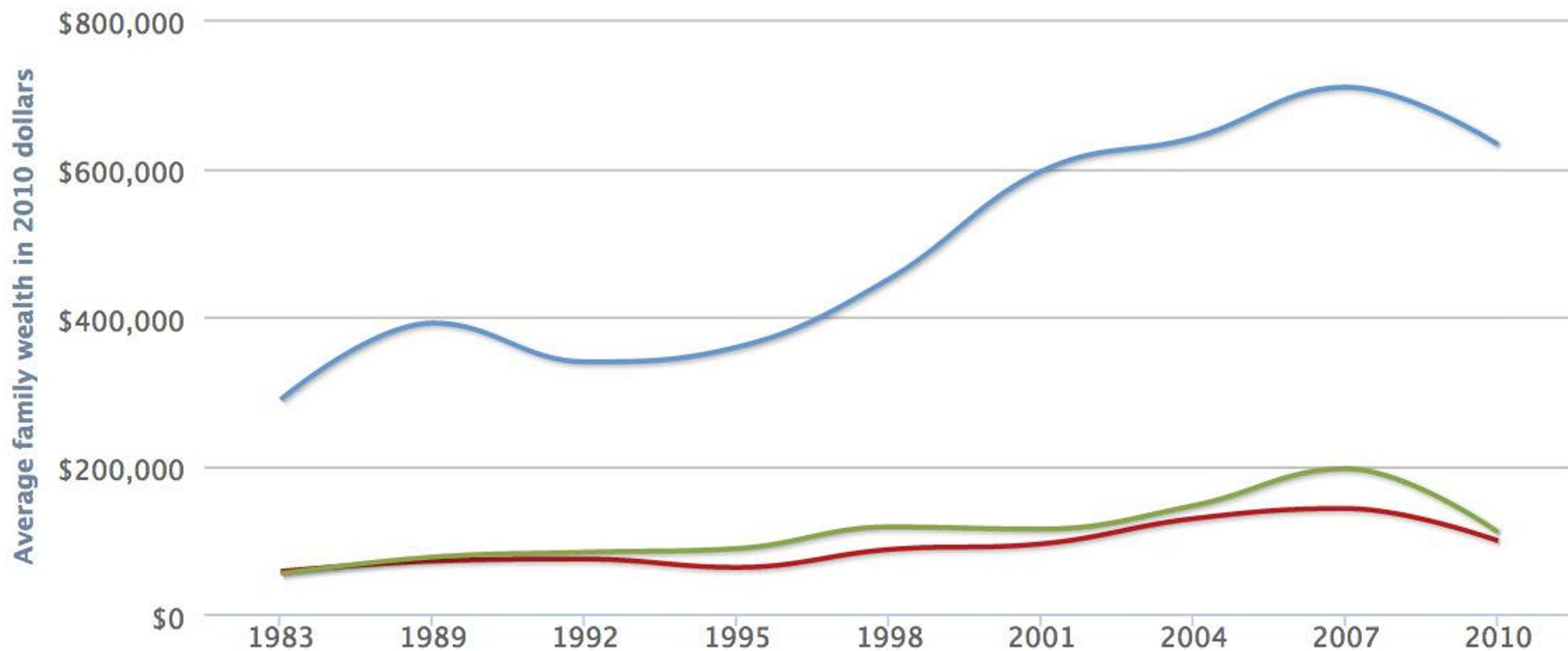
Hypothetical City Crime



The Racial Wealth Gap Is Not Improving

AVERAGE FAMILY WEALTH BY RACE AND ETHNICITY, 1983–2010

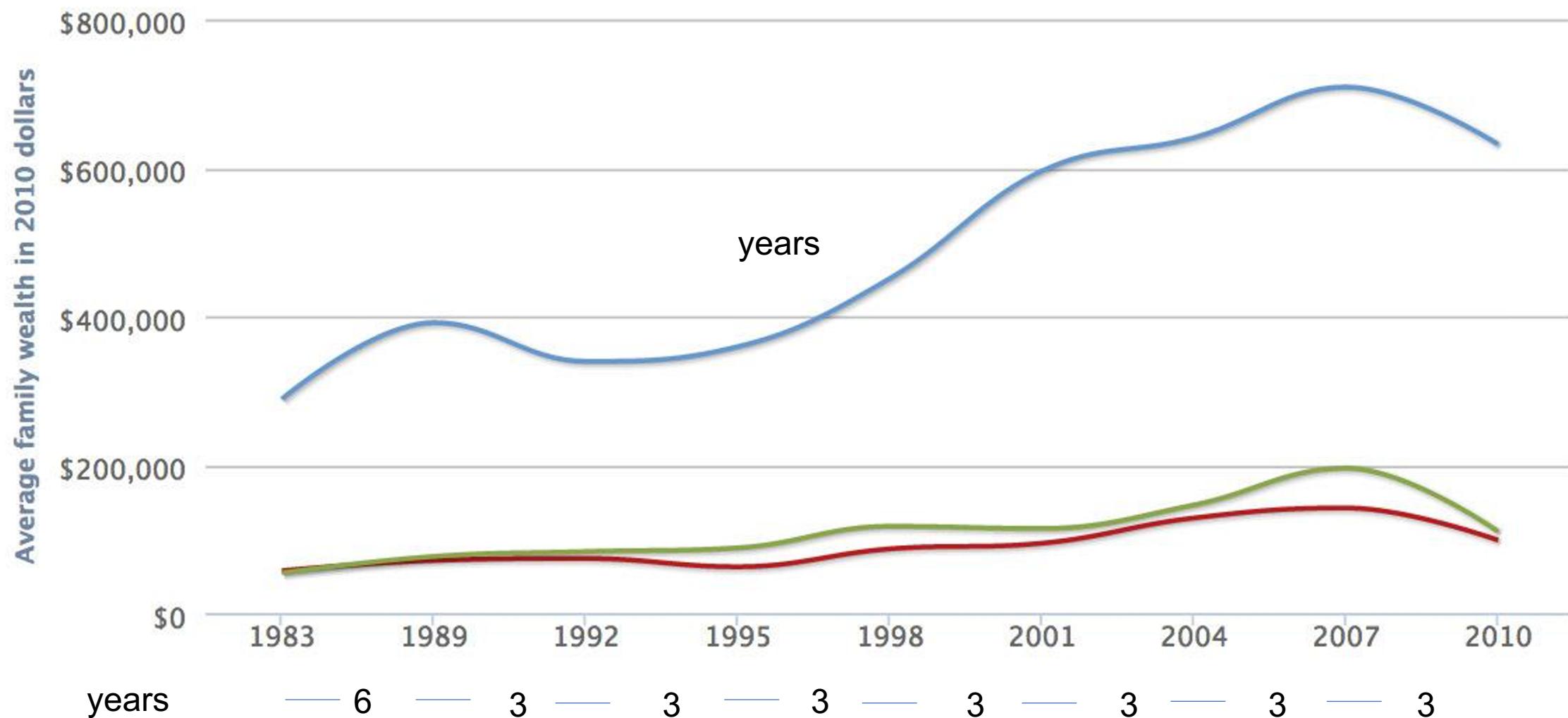
—White non-Hispanic —Black non-Hispanic —Hispanic



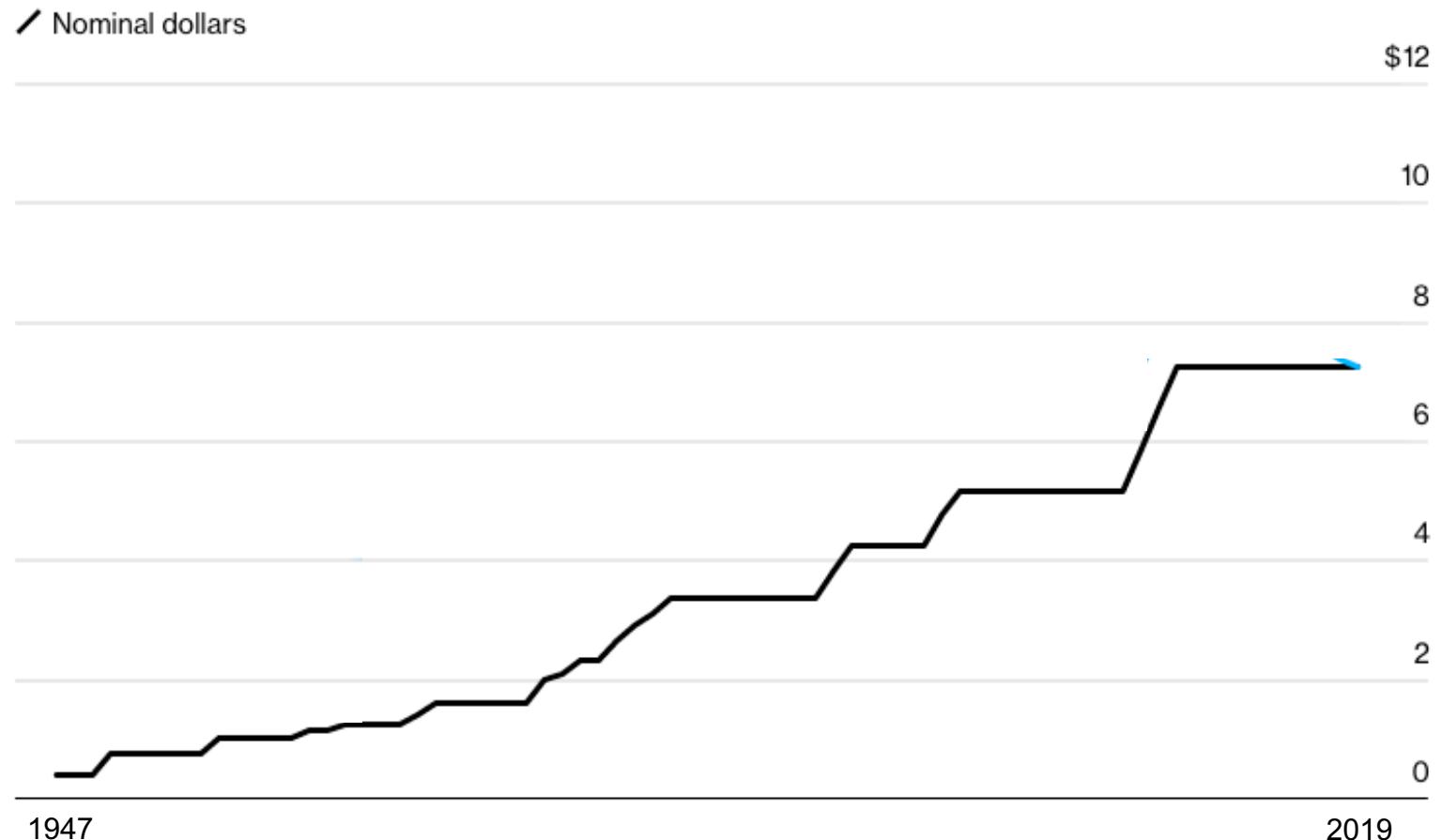
The Racial Wealth Gap Is Not Improving

AVERAGE FAMILY WEALTH BY RACE AND ETHNICITY, 1983–2010

—White non-Hispanic —Black non-Hispanic —Hispanic

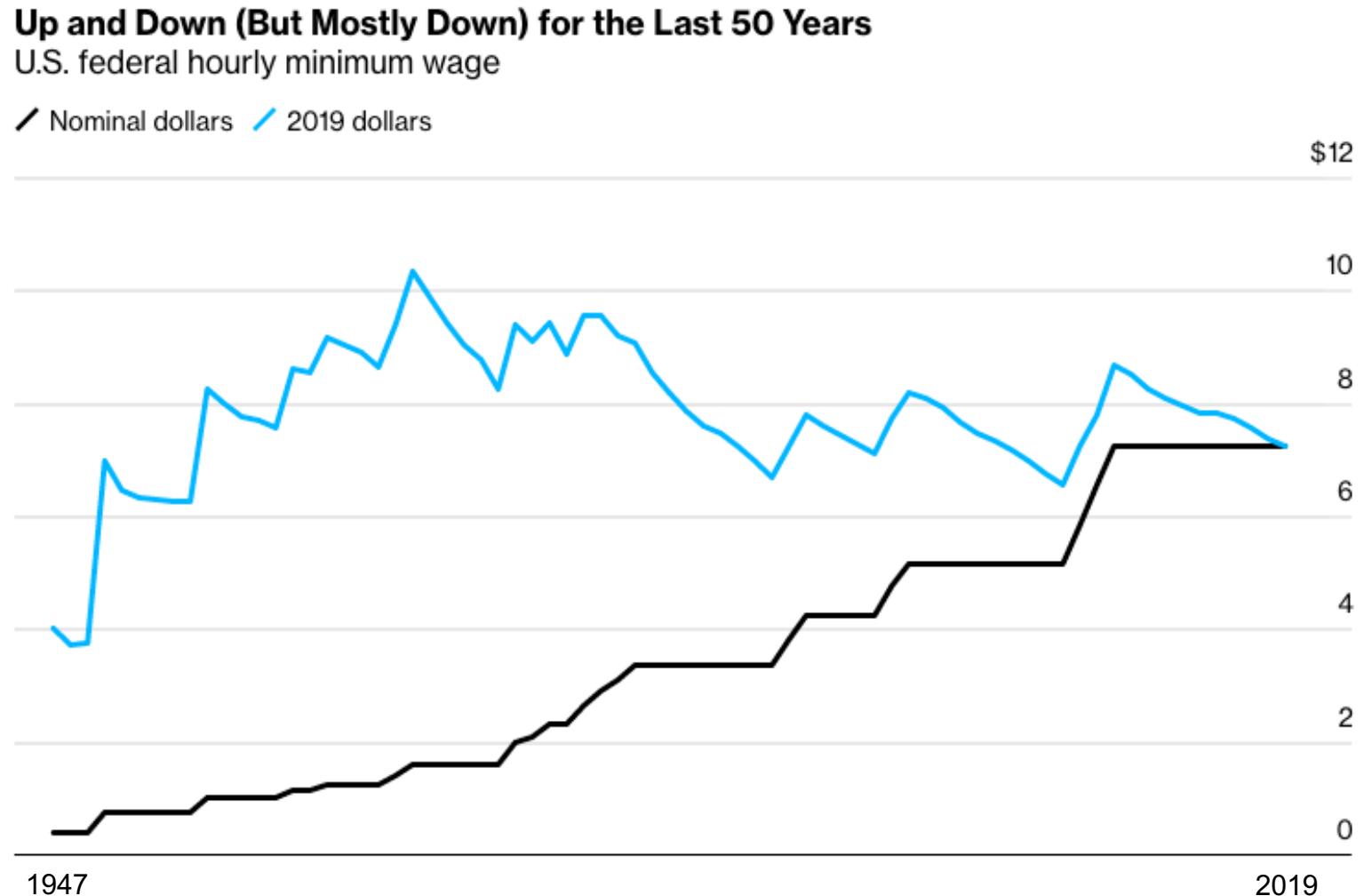


The Dream of a \$ Minimum Wage Gets a Reality Check



Sources: Bureau of Labor Statistics, U.S. Census, Bloomberg

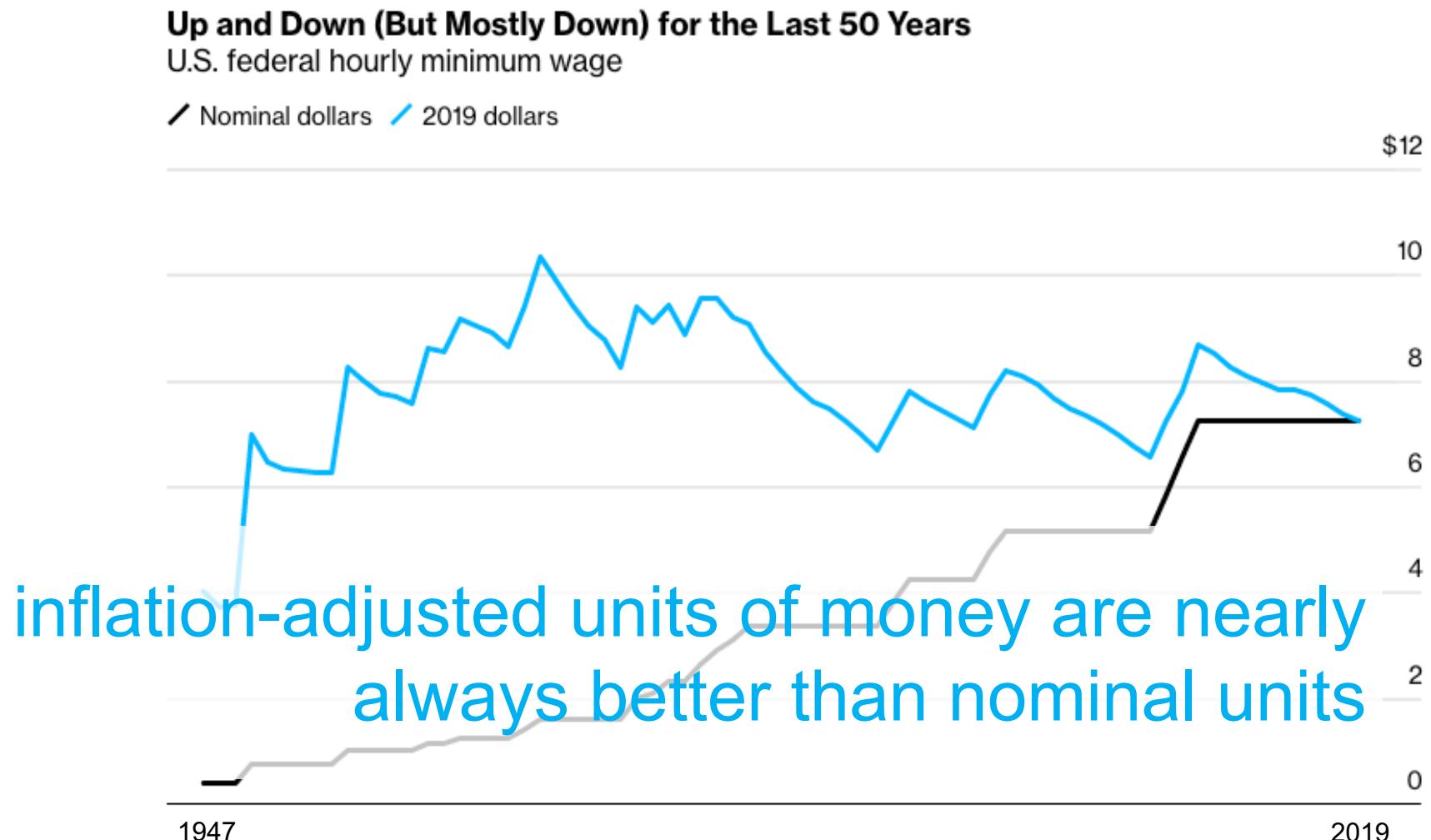
The Dream of a \$ Minimum Wage Gets a Reality Check From Inflation



Note: Inflation adjustment uses CPI-U-RS through 2018, Bloomberg consensus forecast for 2019. Earliest available CPI-U-RS figures are for 1947.

Sources: Bureau of Labor Statistics, U.S. Census, Bloomberg

The Dream of a \$ Minimum Wage Gets a Reality Check **From Inflation**



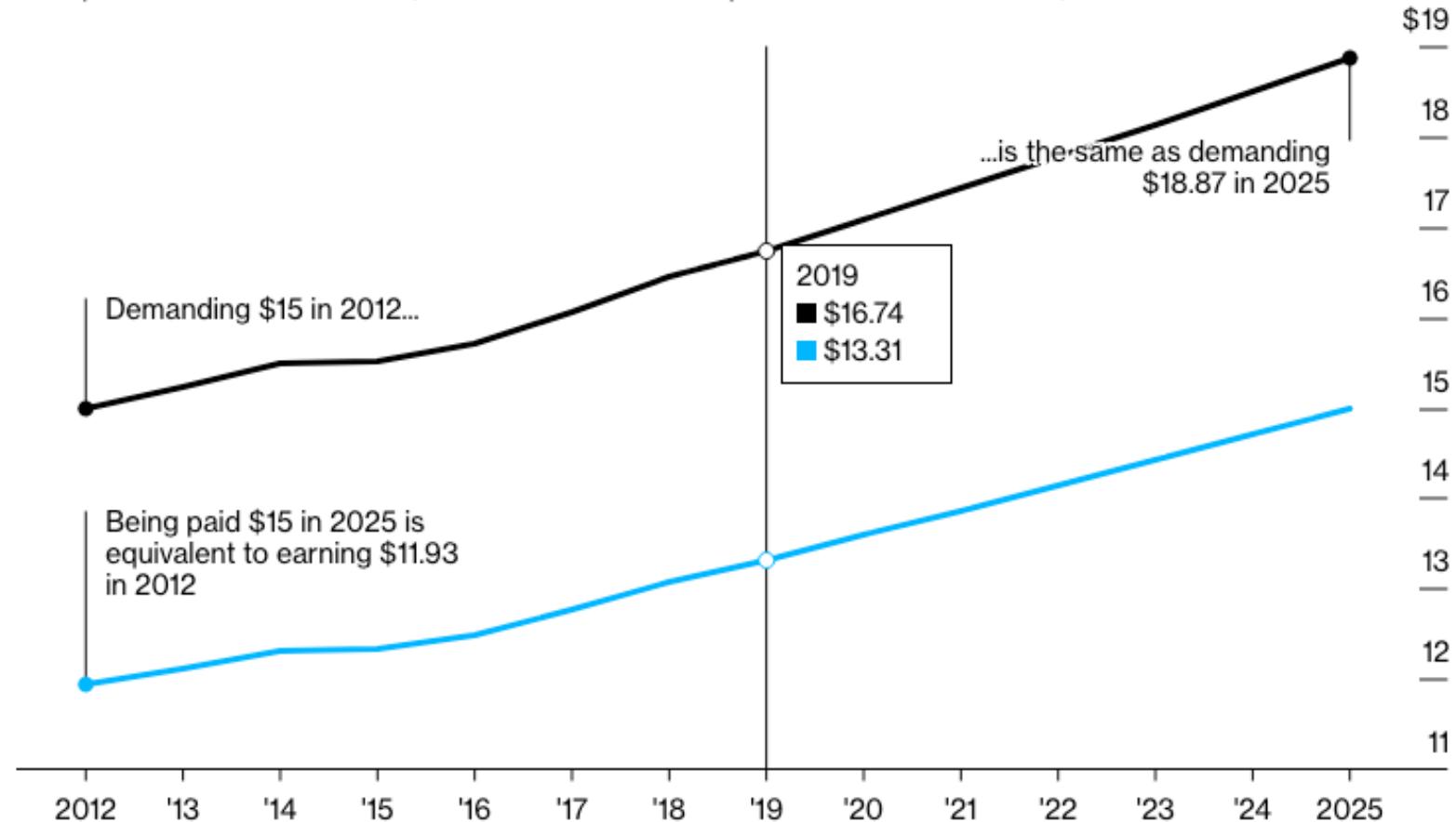
Note: Inflation adjustment uses CPI-U-RS through 2018, Bloomberg consensus forecast for 2019. Earliest available CPI-U-RS figures are for 1947.

Sources: Bureau of Labor Statistics, U.S. Census, Bloomberg

The Dream of a \$ Minimum Wage Gets a Reality Check From Inflation

How Much Is \$15? It Depends When You Ask

↗ Equivalent nominal value of \$15 in 2012 dollars ↘ Equivalent nominal value of \$15 in 2025 dollars



Note: Inflation adjustment uses CPI-U-RS from 2012-2018, Bloomberg consensus forecast for CPI-U for 2019-2021, and assumes 2% for 2022-2025.

Sources: Bureau of Labor Statistics, U.S. Census, Bloomberg

Rule 3 Minimize Graphical Complexity

- Pick the right visualization for your data
- Avoid chart junk

Charts vs. Tables

When to Use Tables vs. Graphs

Use Tables When

1. The display will be used to look up individual values
2. The display will be only be used to compare individual values rather than whole series of values
3. Precise values are required
4. Values involve more than one unit of measure
5. Values must be presented at various levels of aggregation (i.e., summary and detail)

Use Graphs When

1. The message is contained in patterns, trends, and exceptions
2. Entire series of values must be seen as a whole and/or compared

Charts vs. Tables

Primary Function	Relationship Type	Relationship
Look-up	Quantitative-to-Categorical	Between a single set of quantitative values and a single set of categorical items

Salesperson	QTD Sales
Robert Jones	13,803
Mandy Rodriguez	20,374
Terri Moore	28,520
John Donnelly	34,786
Jennifer Taylor	36,973
Total	\$134,456

Charts vs. Tables

Primary Function	Relationship Type	Relationship
Look-up	Quantitative-to-Categorical	
		Between a single set of quantitative values and the intersection of multiple categories

Salesperson	Jan	Feb	Mar
Robert Jones	2,834	4,838	6,131
Mandy Rodriguez	5,890	6,482	8,002
Terri Moore	7,398	9,374	11,748
John Donnelly	9,375	12,387	13,024
Jennifer Taylor	10,393	12,383	14,197
Total	\$35,890	\$45,464	\$53,102

Charts vs. Tables

Primary Function	Relationship Type	Relationship
Look-up	Quantitative-to-Categorical	
		Between a single set of quantitative values and the intersection of multiple hierarchical categories

Product Line	Product Family	Product	Sales
Hardware	Printer	PPS	6,131
		PXT	8,002
		PQT	11,748
	Router	RRZ	13,024
		RTS	14,197
		RQZ	23,293
Software	Business	ACT	12,393
		SPR	9,393
		DBM	5,392
	Game	ZAP	10,363
		ZAM	15,709
		ZOW	13,881
Total			\$143,526

Charts vs. Tables

Primary Function	Relationship Type	Relationship
Look-up	Quantitative-to-Categorical	
Comparison	Quantitative-to-Quantitative	Among a single set of quantitative values associated with multiple categorical items

Salesperson	Jan	Feb	Mar
Robert Jones	2,834	4,838	6,131
Mandy Rodriguez	5,890	6,482	8,002
Terri Moore	7,398	9,374	11,748
John Donnelly	9,375	12,387	13,024
Jennifer Taylor	10,393	12,383	14,197
Total	\$35,890	\$45,464	\$53,102

Charts vs. Tables

Primary Function	Relationship Type	Relationship
Look-up	Quantitative-to-Categorical	

Primary Function	Relationship Type	Relationship
Comparison	Quantitative-to-Quantitative	

Primary Function	Relationship Type	Relationship
Comparison	Quantitative-to-Quantitative	

Among distinct sets of quantitative values associated with a single categorical item

Salesperson	Sales	Returns	Net Sales
Robert Jones	13,803	593	13,210
Mandy Rodriguez	20,374	1,203	19,171
Terri Moore	28,520	10,393	18,127
John Donnelly	34,786	483	34,303
Jennifer Taylor	36,973	0	36,973
Total	\$134,456	\$12,672	\$121,784

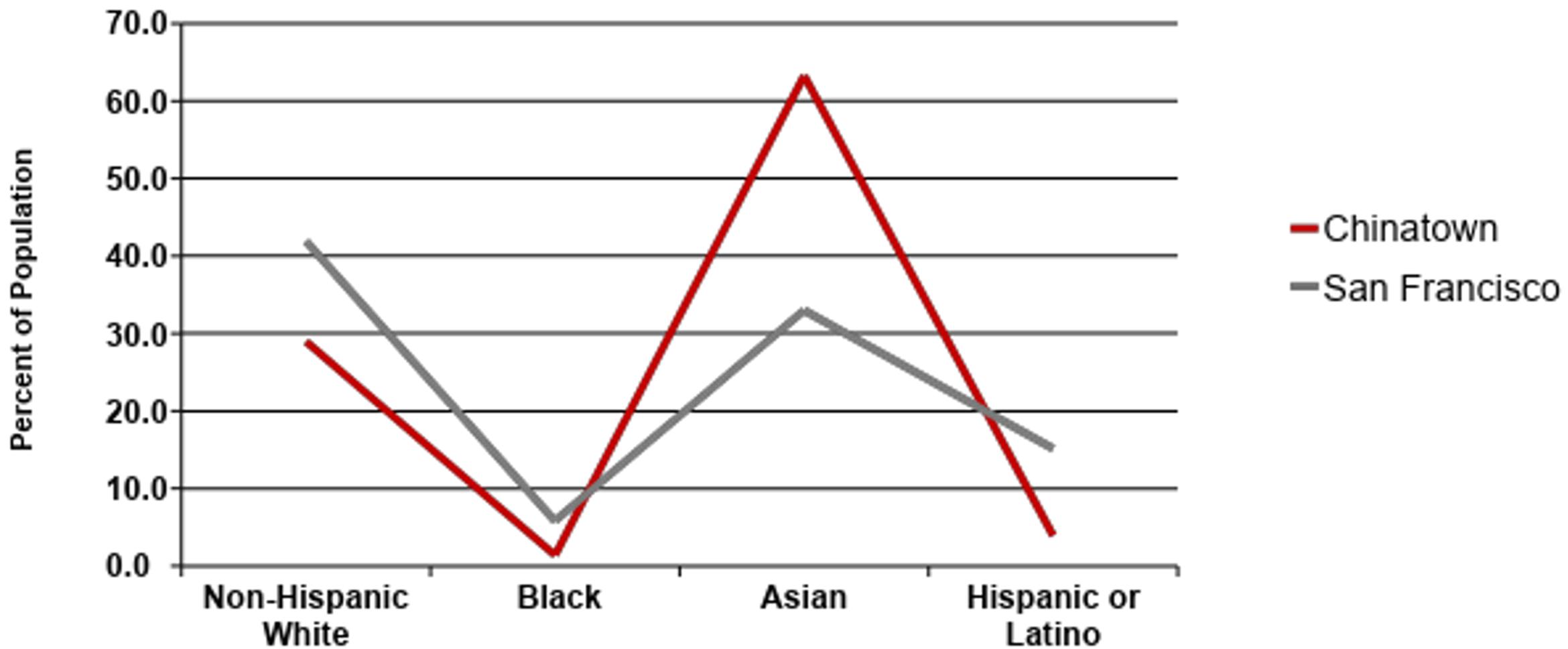
Chinatown Demographics

	Total	Non-Hispanic White	Black	American Indian	Asian	Native Hawaiian	Other	Two or More	Hispanic or Latino
San Francisco									
Population	805,235	337,451	46,781	1,828	265,700	3,128	2,494	26,079	121,774
Percent		41.9	5.8	0.2	33.0	0.4	0.3	3.2	15.1
Chinatown									
Population	13,470	3,900	186	13	8,524	25	19	274	529
Percent		29.0	1.4	0.1	63.3	0.2	0.1	2.0	3.9

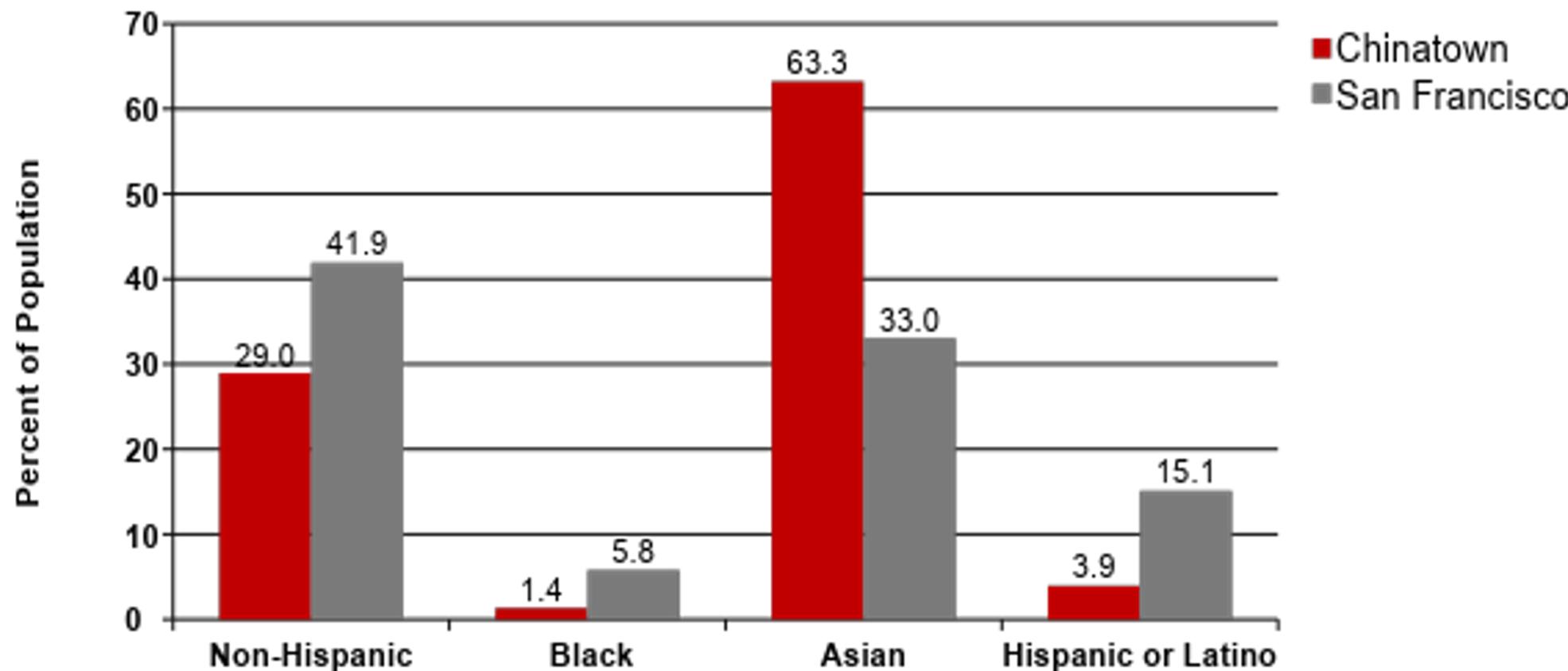
Source: Census, 2010 SF1, Table P5.

Good if you want to present detail, and also give a sense of the size as well as the distribution of the population. But, doesn't quickly tell a story if you're giving a presentation

Better?

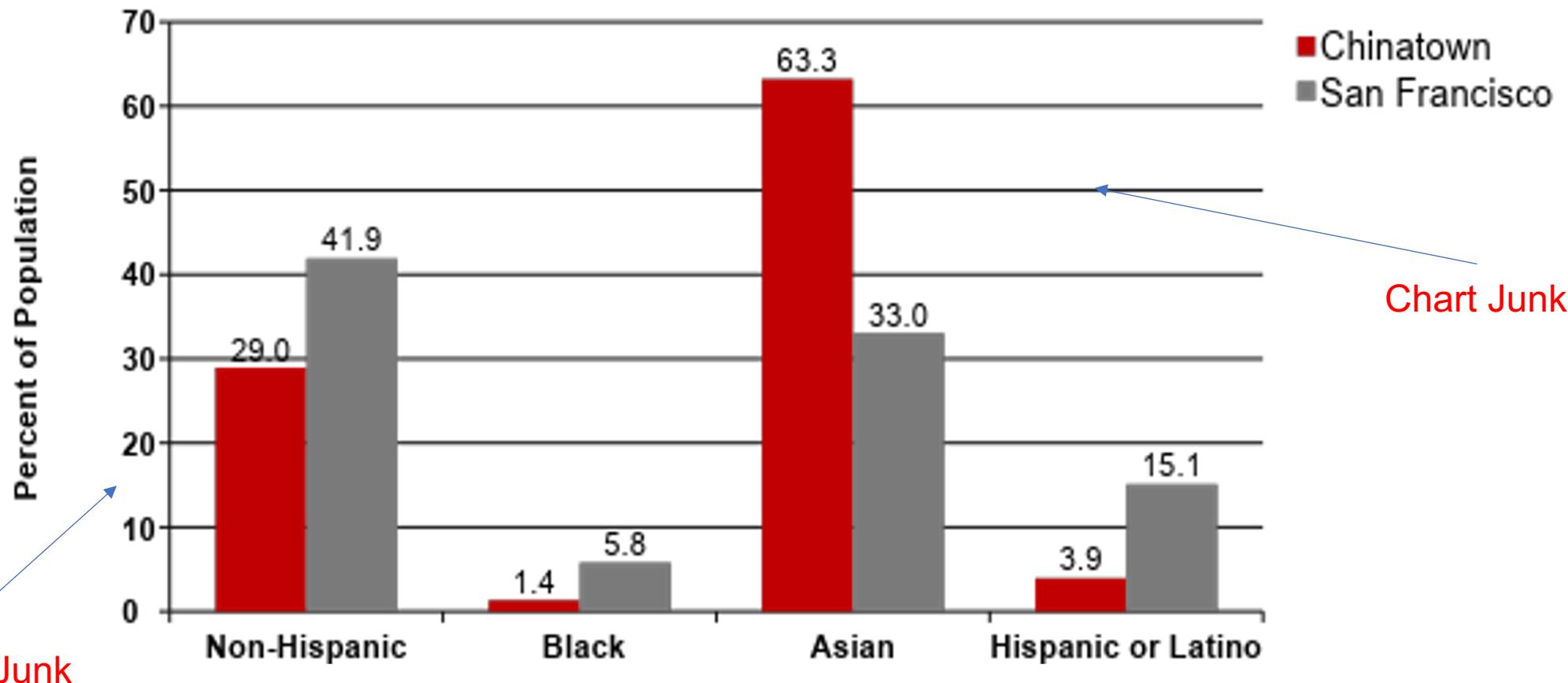


Better?

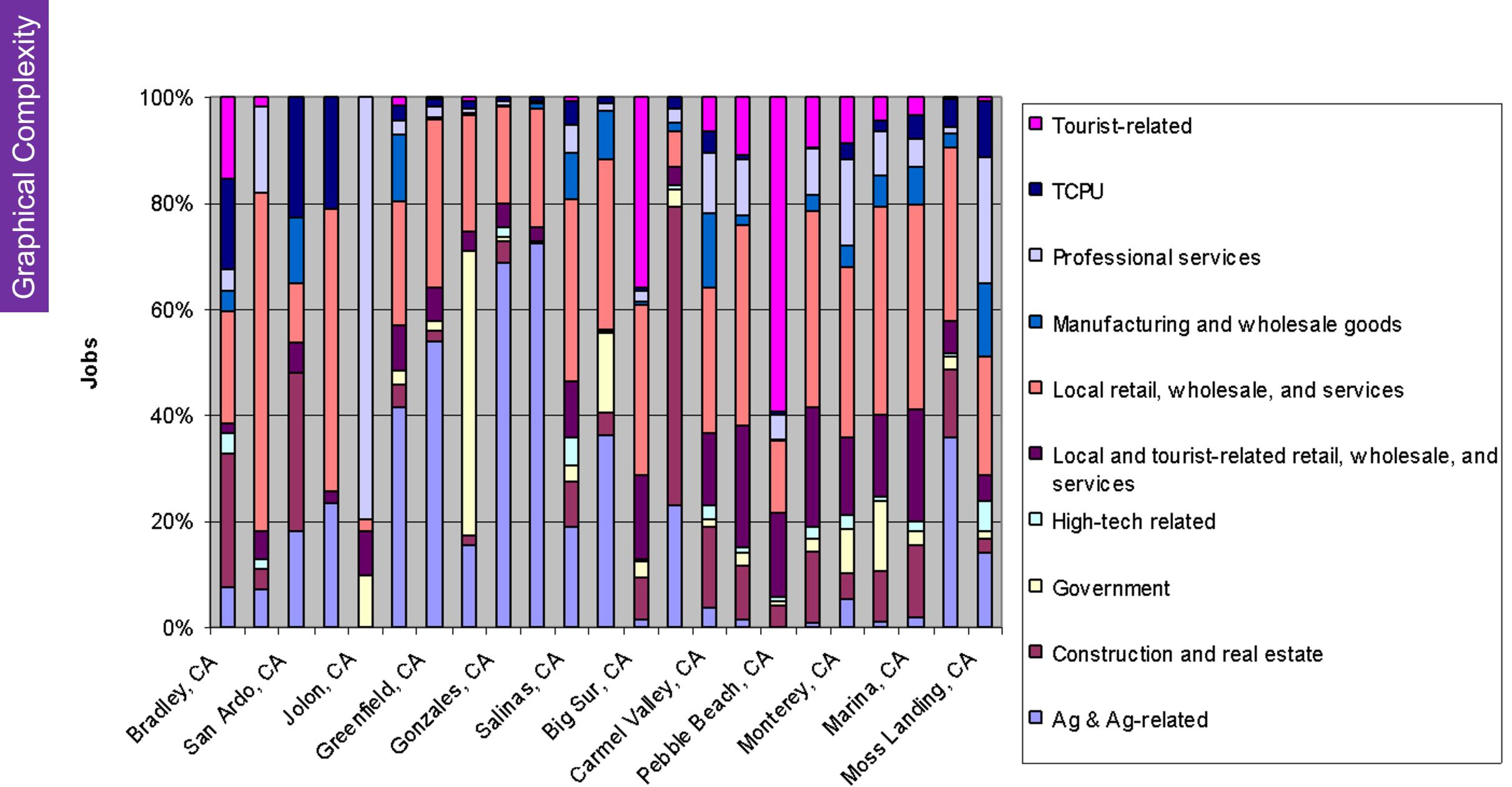


Source: Census, 2010 SF1, Table P5. Chart does not include data for Native Hawaiian, Pacific Islander, or Native American population, or for individuals who reported as belonging to "other" or two or more races. Blacks and Asians are designated as being of non-Hispanic ethnic origin. Chinatown is defined as census tracts 107, 113, 118 and 601 in San Francisco County.

Graphical Complexity

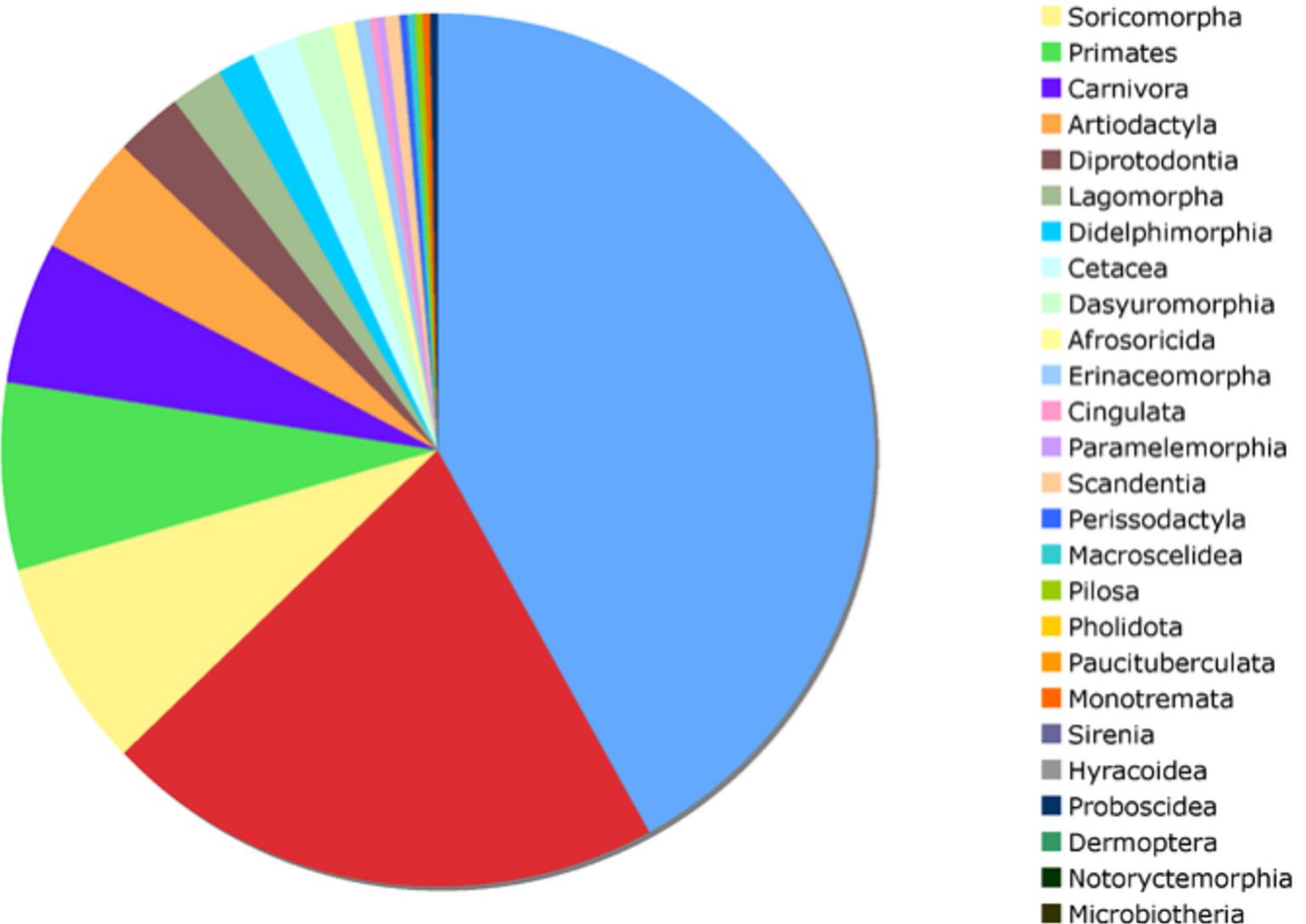


Source: Census, 2010 SF1, Table P5. Chart does not include data for Native Hawaiian, Pacific Islander, or Native American population, or for individuals who reported as belonging to "other" or two or more races. Blacks and Asians are designated as being of non-Hispanic ethnic origin. Chinatown is defined as census tracts 107, 113, 118 and 601 in San Francisco County.



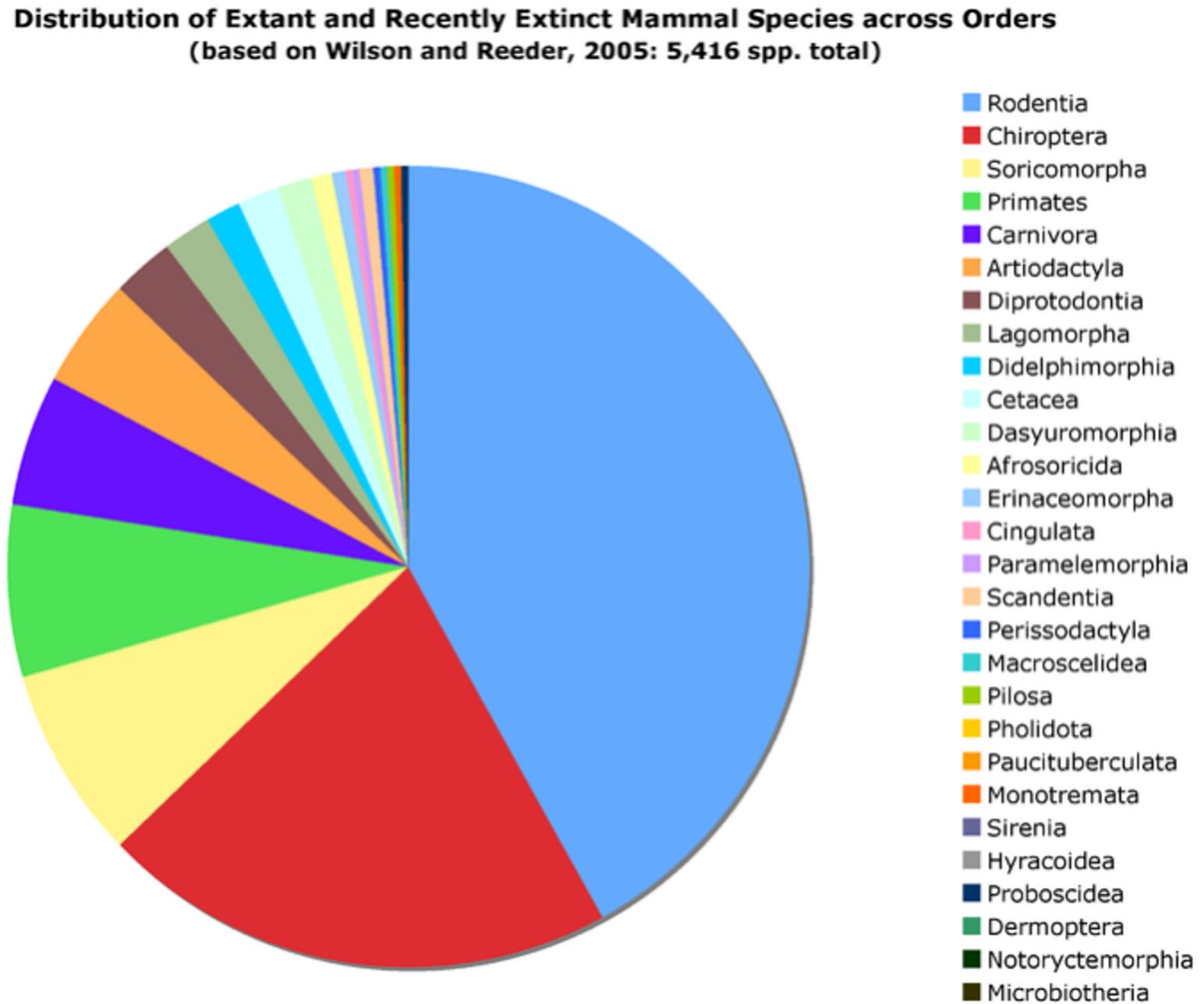
Pie Charts

Distribution of Extant and Recently Extinct Mammal Species across Orders
(based on Wilson and Reeder, 2005: 5,416 spp. total)



Pie Charts

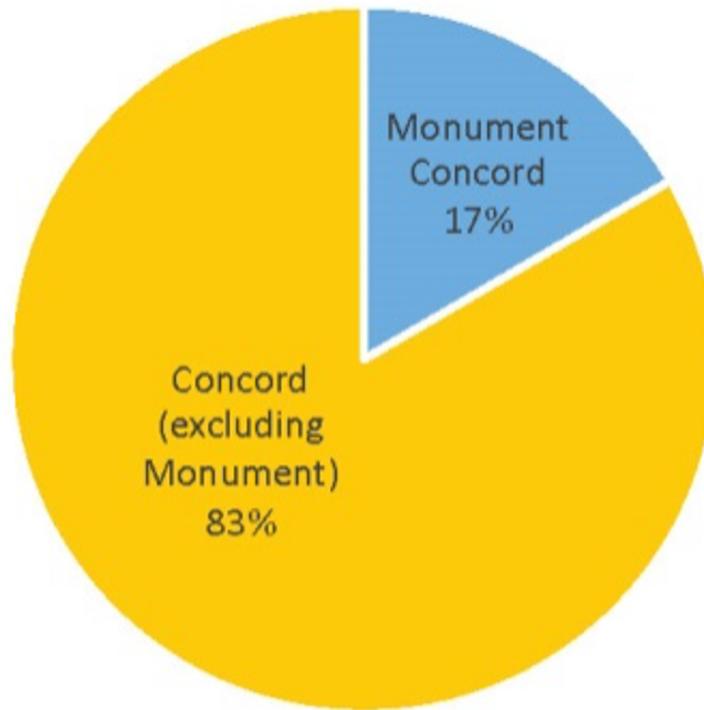
- Limit number of pie slices (8 max)
- Place labels in or alongside each slice of pie
- Provide the base total in your source notes



Pie Charts

- Limit number of pie slices (8 max)
- Place labels in or alongside each slice of pie
- Provide the base total in your source notes

Neighborhood Breakdown of Concord

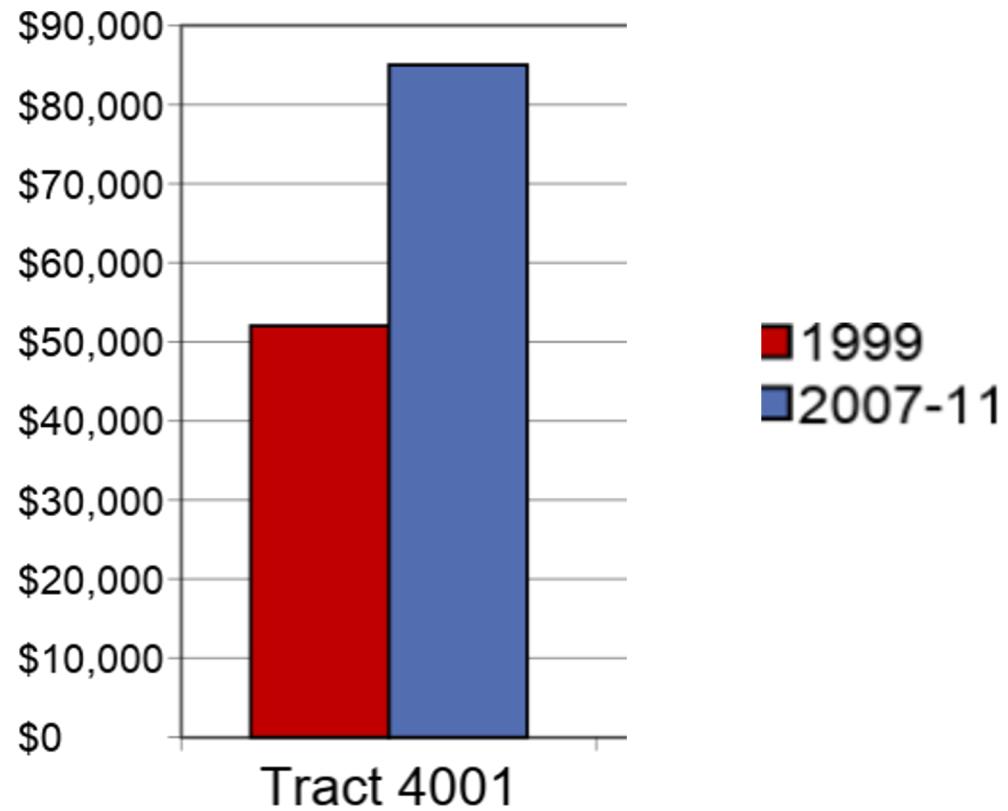


Note: 2010 Decennial Census

Base: Total Population (N=147,145)

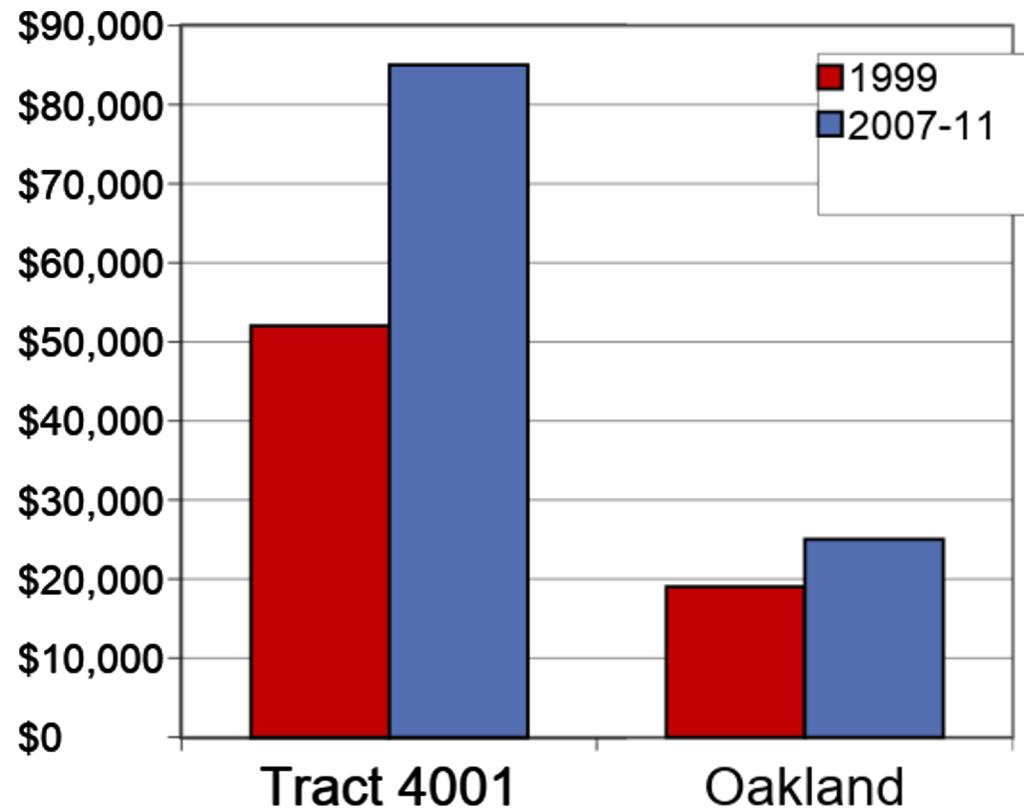
Rule 4 Context is Essential

Median Household Income

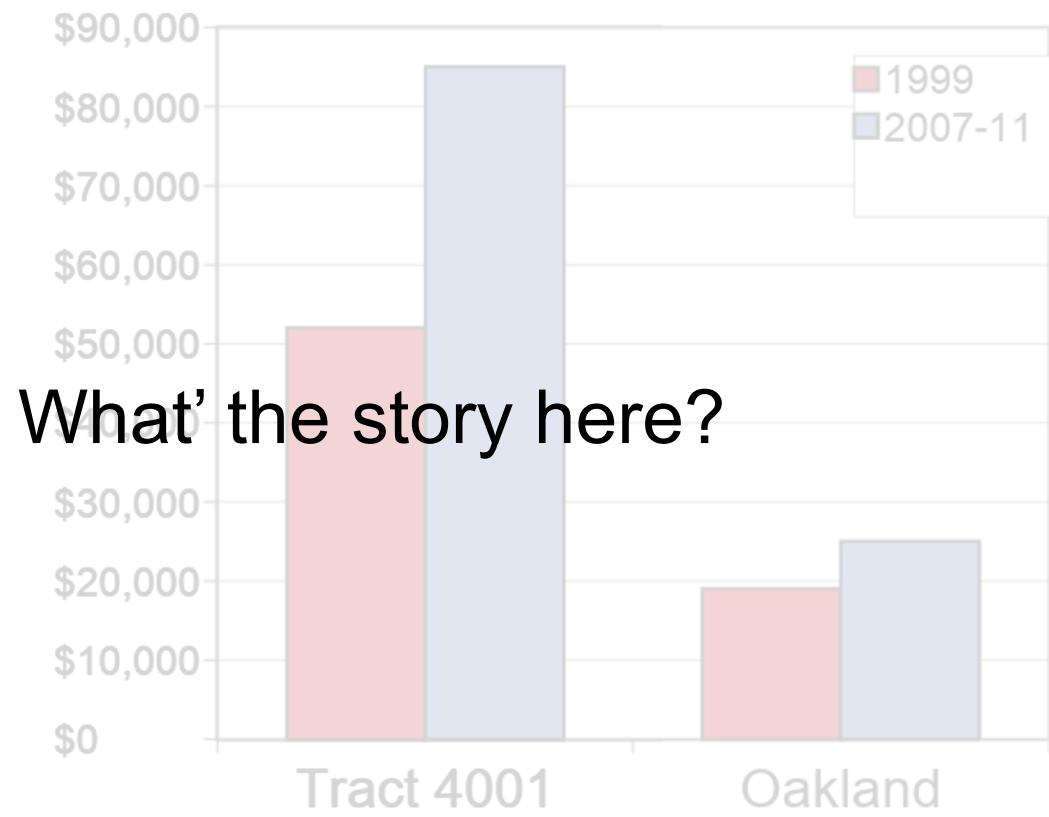


What' the story here?

Median Household Income

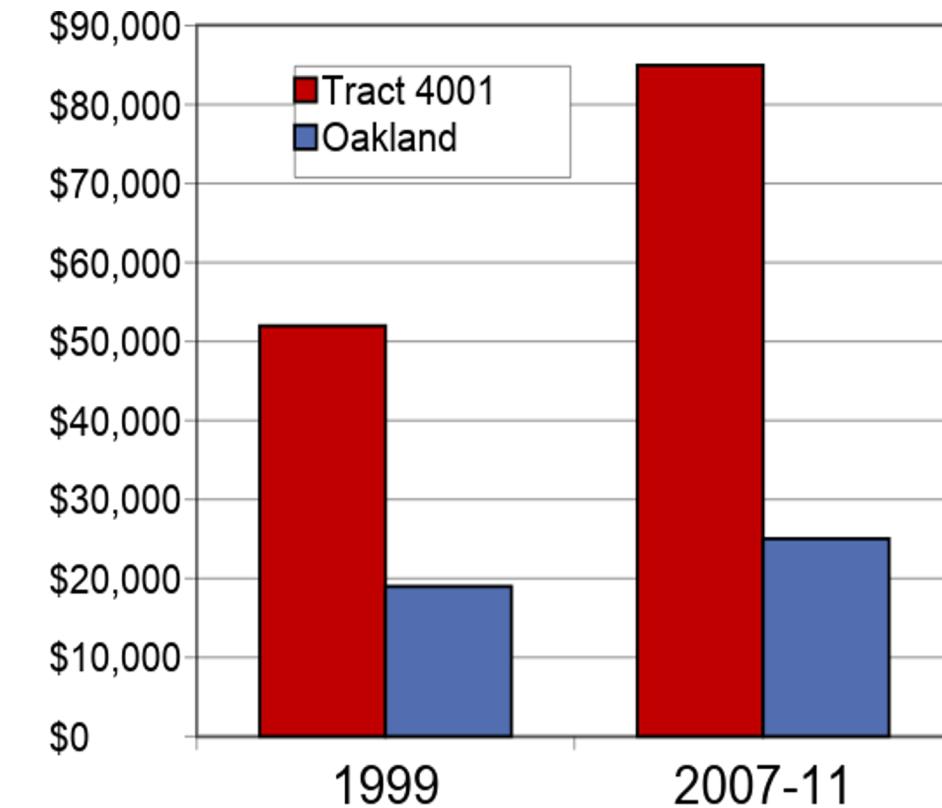


What' the story here?



What's the story here?

Median Household Income

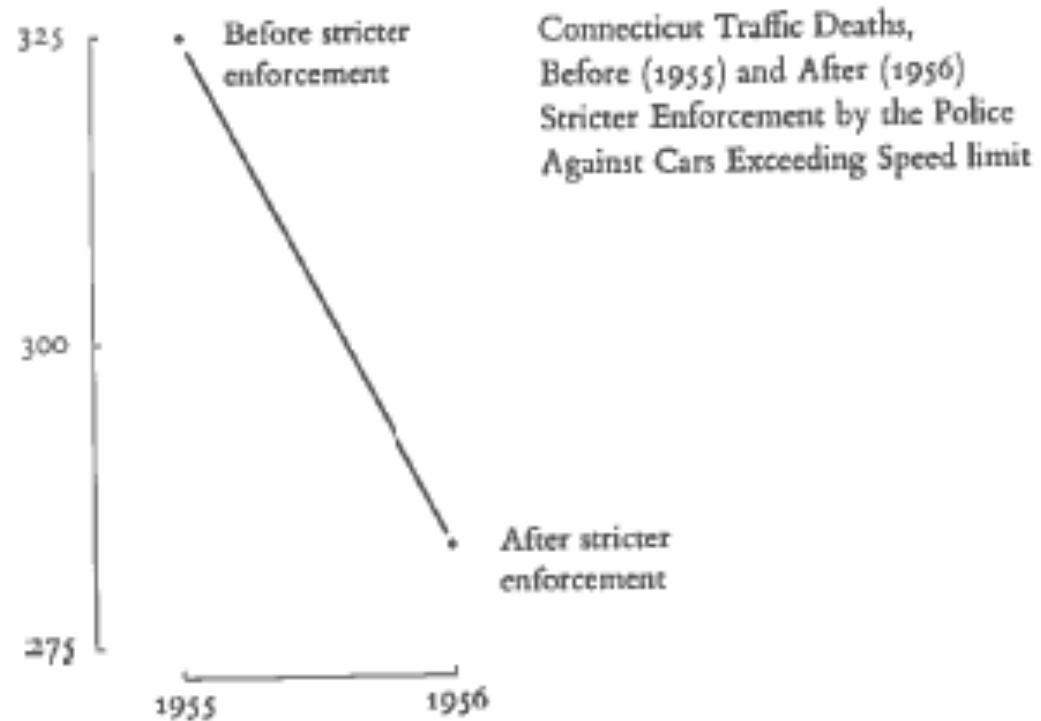


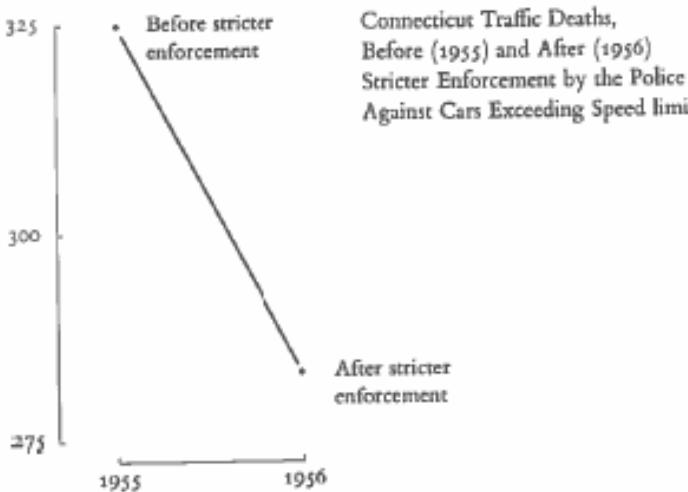
According to Tufte:

Context is Essential for Graphical Integrity

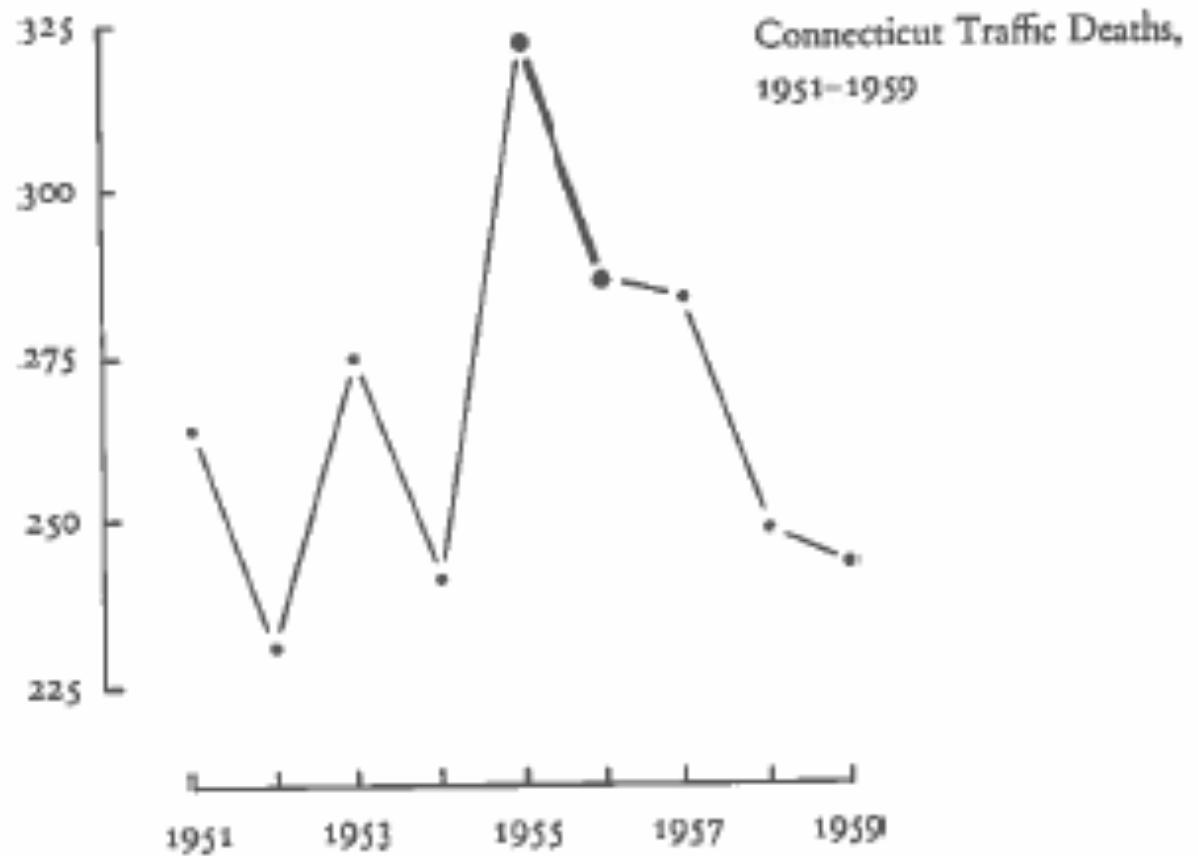
To be truthful and revealing, data graphics must bear on the question at the heart of quantitative thinking: "Compared to what?" The emaciated, data-thin design should always provoke suspicion, for graphics often lie by omission, leaving out data sufficient for comparisons. The principle:

Graphics must not quote data out of context.

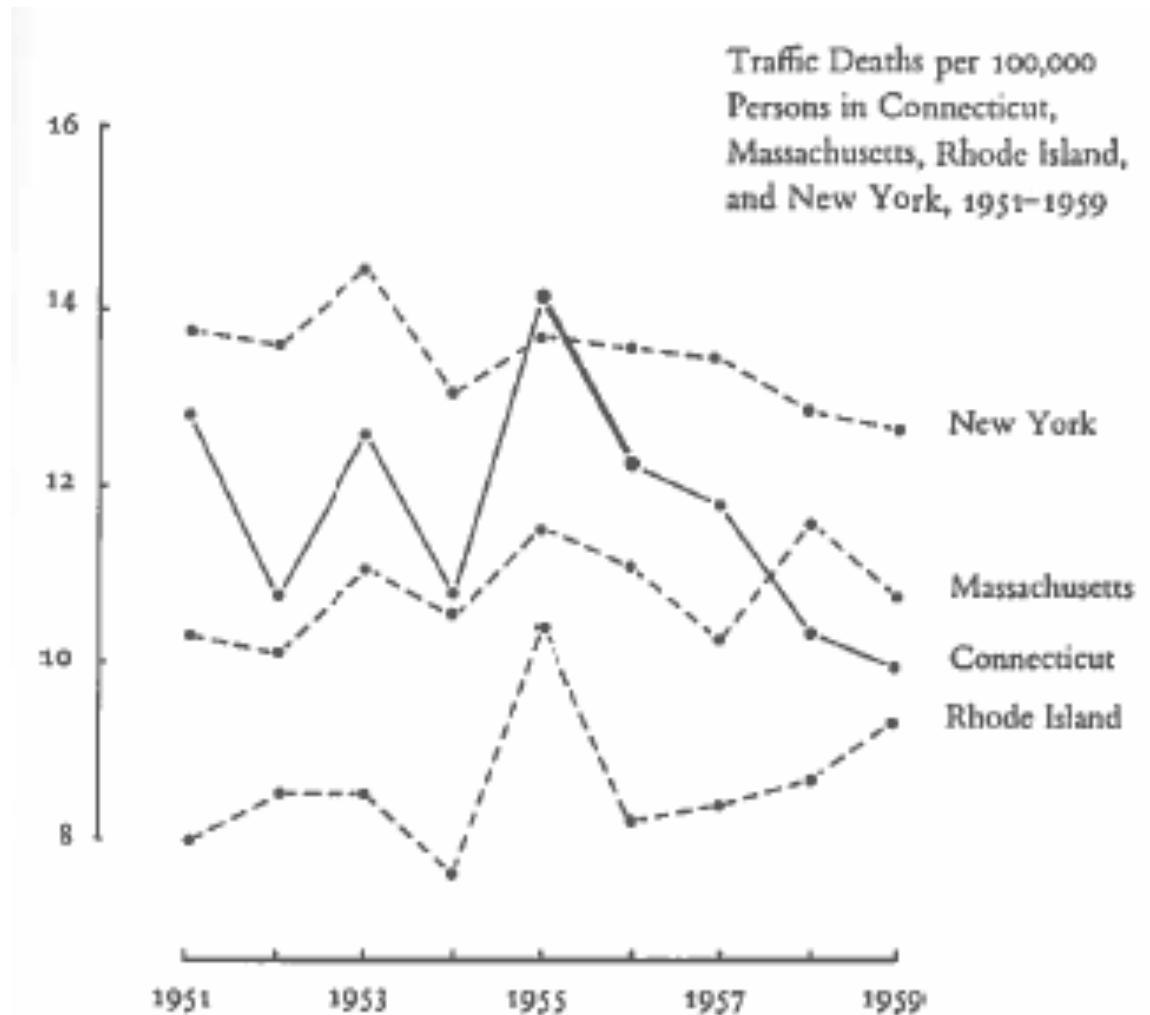
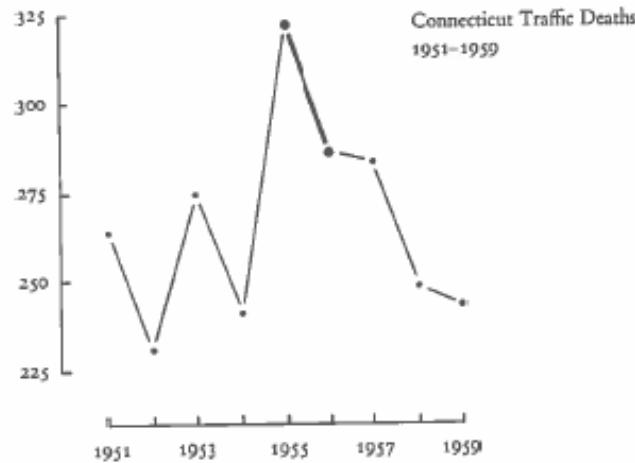




A few more data points add immensely to the account:



A few more data points add immensely to the account:



The Adjacency Principle in Visual Perception

A basic assumption underlying the measurement of adjacency effects and well supported by the experimental data is that the visual system is able to combine information from sources that are not in agreement. Suppose there are two sources, one source that by itself would lead to one perception and another that by itself would lead to a different perception. The assumption is that when the two sources are presented together, they will give rise to yet a third perception, to which both sources contribute.

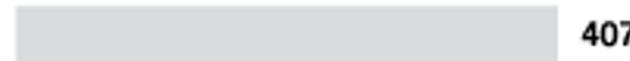
Walter C. Gogel
Scientific American
Vol. 238, No. 5 (May 1978), pp. 126

Most dangerous cities

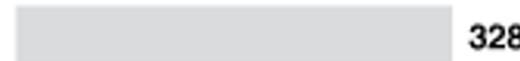
Total murders in 2014

WRONG

Chicago



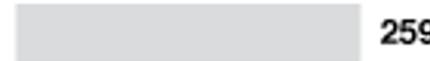
New York



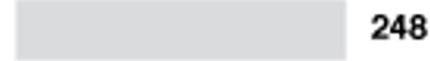
Detroit



Los Angeles



Philadelphia



Most dangerous cities

Total murders in 2014

WRONG



New York

Detroit

Los Angeles

Philadelphia

407

328

304

259

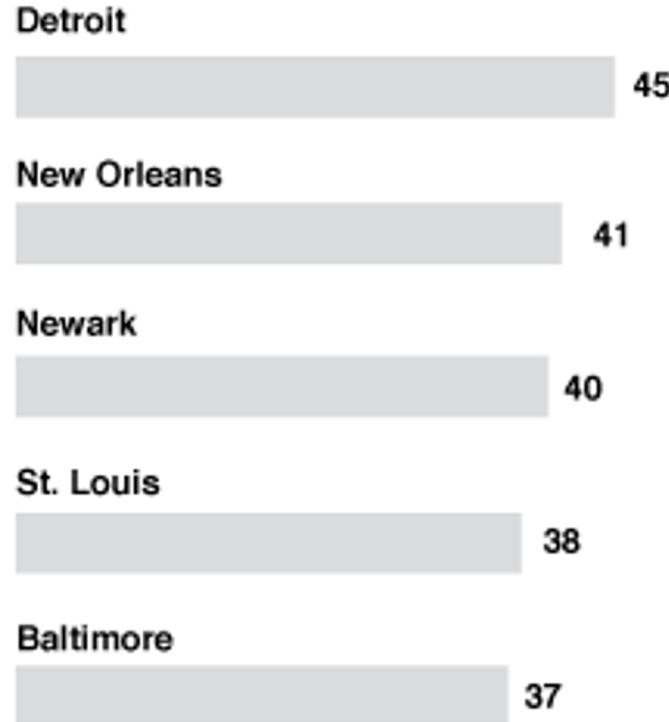
248

Always account for population size

Most dangerous cities

Murder rate in major US cities in 2014,
per 100,000 people

RIGHT



Detroit

New Orleans

Newark

St. Louis

Baltimore

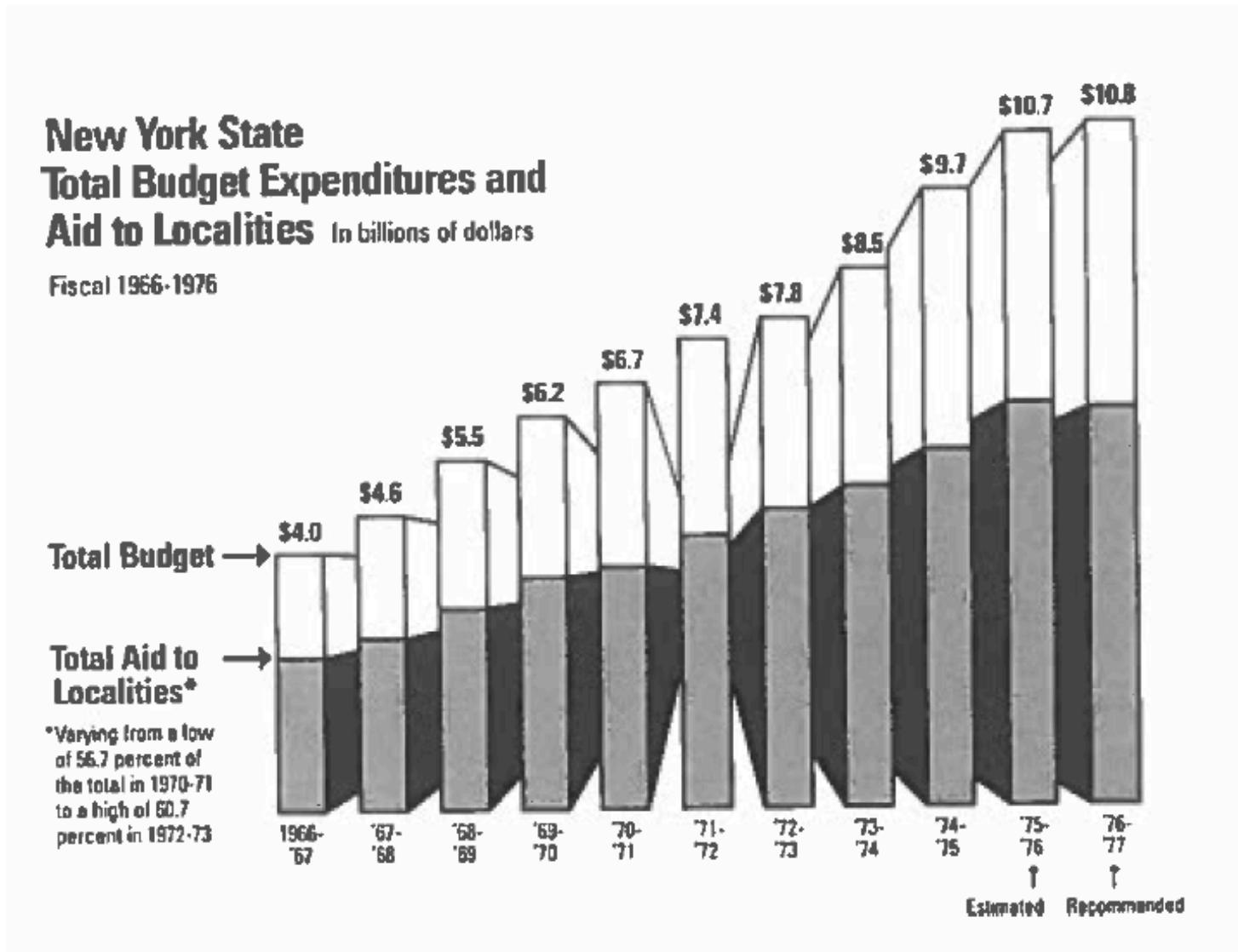
45

41

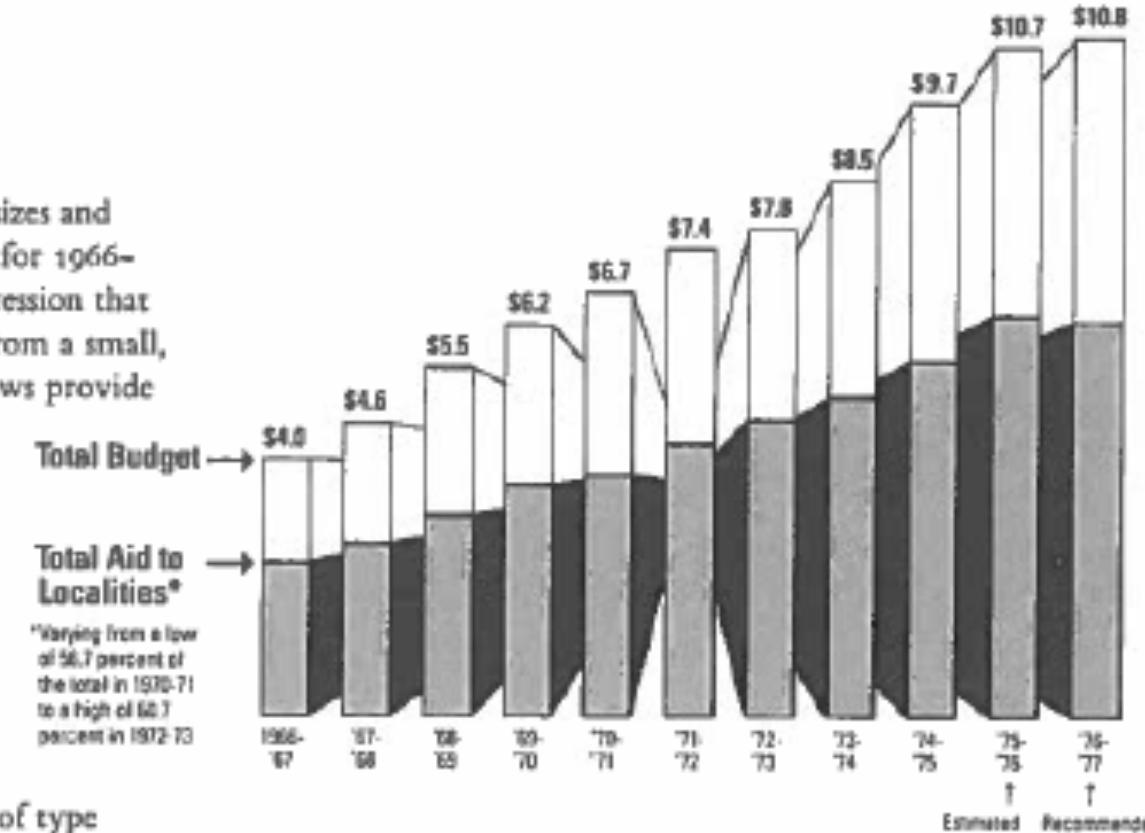
40

38

37



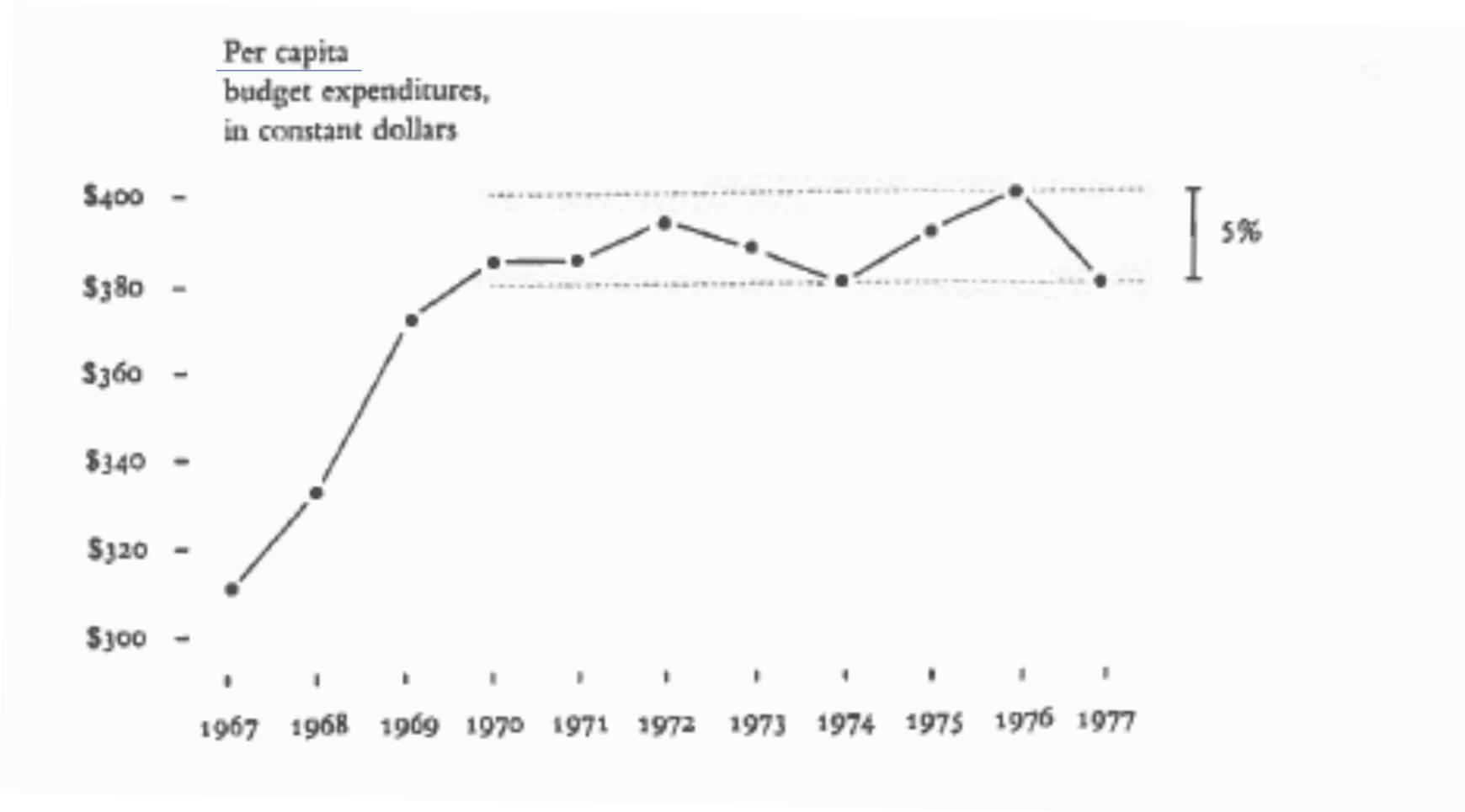
This cluster of type emphasizes and stretches out the low value for 1966-1967, encouraging the impression that recent years have shot up from a small, stable base. Horizontal arrows provide similar emphasis.



This squeezed-down block of type contributes to an image of small, squeezed-down budgets back in the good old days.

These three parallelepipeds have been placed on an optical plane *in front* of the other eight, creating the image that the newer budgets tower over the older ones.

Arrows pointing straight up emphasize recent growth. Compare with horizontal arrows at left.



Reading Response Question

Real Schein

Tufte, 1983. Graphical Integrity.

In this chapter, Tufte illustrates different methods in which visualized data can be skewed and analyzed. Assuming the role of a city planner or analyst, **what commonplace administrative standards are often utilized to preserve graphical integrity and visualization besides the six general principles listed in the conclusion?** A standardized principle is to emphasize data variation rather than design variation (as mentioned in the chapter). It is also stated that it is an "inefficient" technique to visualize one-dimension data in two or three dimensions. **As technology has advanced since this was published, to what extent is this concept commonly seen in the realm of city planning? In addition, as this was written more than a decade ago, how has the progression of technology embarked in combatting or advancing graphical integrity?**

Real Schein

Tufte, 1983. Graphical Integrity.

In this chapter, Tufte illustrates different methods in which visualized data can be skewed and analyzed. Assuming the role of a city planner or analyst, what commonplace administrative standards are often utilized to preserve graphical integrity and visualization besides the six general principles listed in the conclusion? A standardized principle is to emphasize data variation rather than design variation (as mentioned in the chapter). It is also stated that it is an "inefficient" technique to visualize one-dimension data in two or three dimensions. As technology has advanced since this was published, to what extent is this concept commonly seen in the realm of city planning? In addition, as this was written more than a decade ago, how has the progression of technology embarked in combatting or advancing graphical integrity?

Kevin Taylor

Few, 2012. Show Me the Numbers: Designing Tables and Graphs to Enlighten

On p. 309, the chart explaining the use of tables and graphs says to use graphs when an entire series of values is seen as compared, and to use tables when precise values are required. In Assignment 1 there are many variables that can be presented for our census tracts, and both tables and graphs could be helpful. For example, we could present the racial makeup of the census tract simply left as a table, or create a bar plot. Is there are preference for tables vs. graphs in the assignment? Should we use both? Are tables required for checking the accuracy of our reporting in the assignment?

Lab 3
data.census.gov
WordPress
MoE and Error Bars on Charts
Social Explorer