

EDA with categorical data

Comparing numerical data across
groups

Prof. Dr. Jan Kirenz
HdM Stuttgart

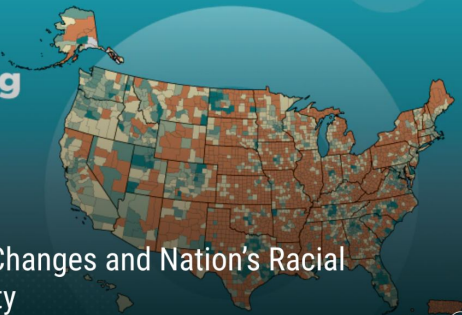
Data for 3142 counties in the United States



Search

[BROWSE BY TOPIC](#)[EXPLORE DATA](#)[LIBRARY](#)[SURVEYS/ PROGRAMS](#)[INFORMATION FOR...](#)[FIND A CODE](#)[ABOUT US](#)

2020 Census Redistricting Data



Local Population Changes and Nation's Racial and Ethnic Diversity

Read More

The U.S. Census Bureau today released additional 2020 Census results showing an increase in the population of U.S. metro areas compared to a decade ago.

SURVEYS

Help for Survey Participants

Verify that the survey you received is real and learn how to respond.

QUICKFACTS

Access Local Data

Learn about your community, county, state and the U.S. It's fast, easy and shareable.

POPULATION CLOCK

August 22, 2021

USA
332,659,369

World
7,784,236,581

U.S. CENSUS BUREAU ECONOMIC INDICATORS

Selected Services Revenue 2nd Quarter 2021 Report Released 10:00 AM EDT, 8/19/21	\$4,367.8 B Advance Report 4.0%
New Residential Construction July 2021 Report Released 8:30 AM EDT, 8/18/21	1,534,000 Housing starts -7.0%
Business Inventories June 2021 Report Released 10:00 AM EDT, 8/17/21	\$2,057.4 B 0.8%
Advance Monthly Retail Sales July 2021 Report Released 8:30 AM EDT, 8/17/21	\$617.7 B -1.1%

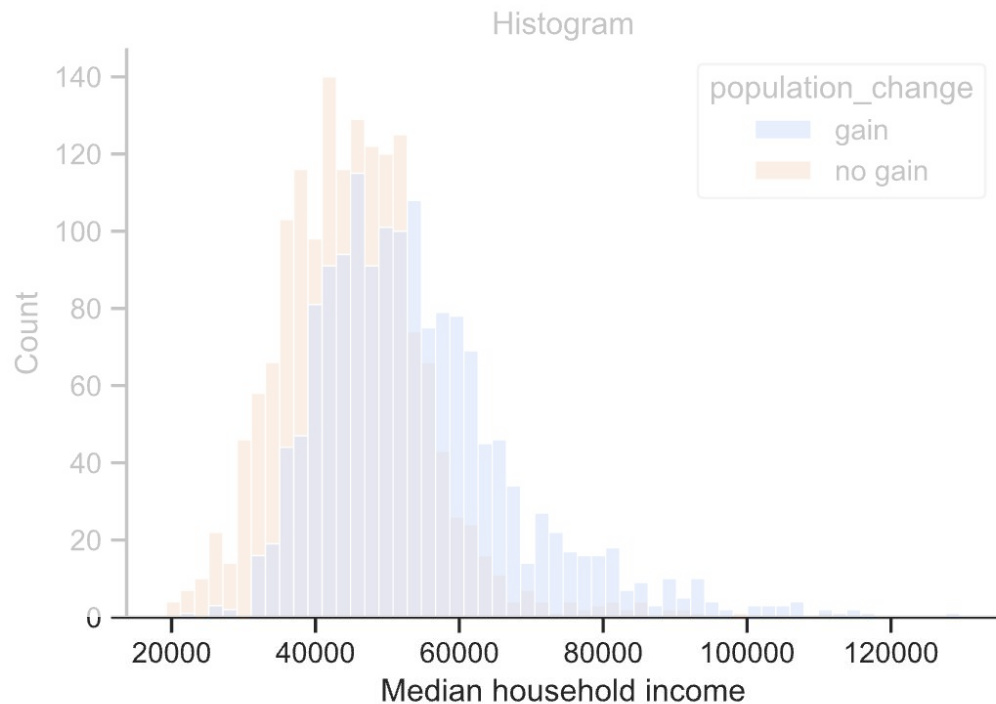
[All Economic Indicators](#)

* change not statistically significant
○ significance not reported / applicable

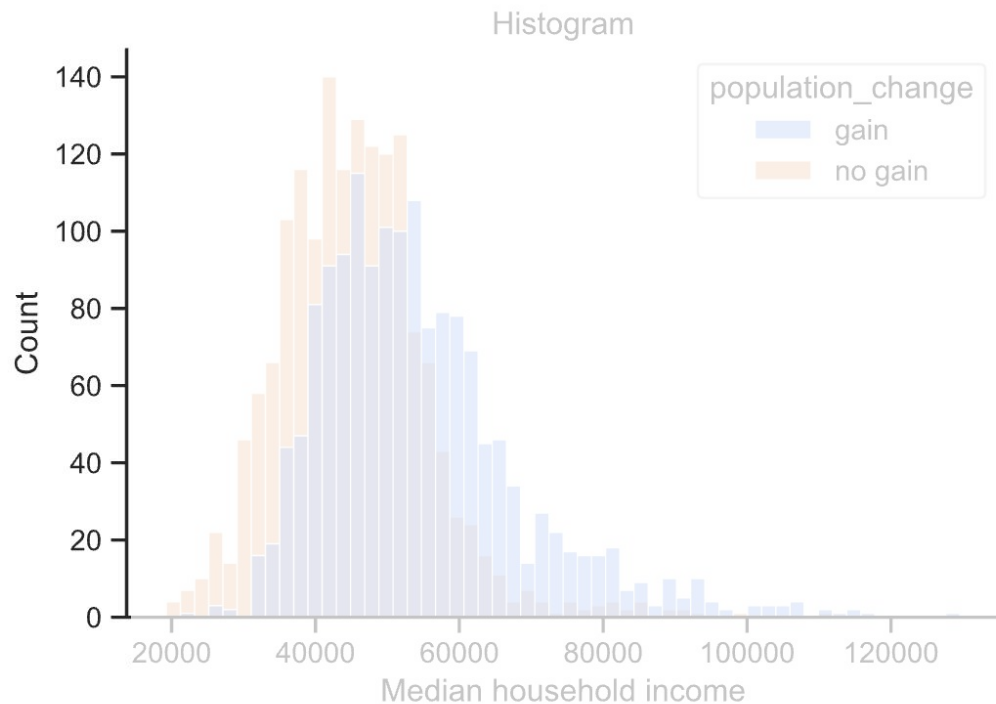
Overview about some counties with two groups: with and without population change

	state	name	pop_change	population_change	median_hh_income
0	Alabama	Autauga County	1.48	gain	55317.0
1	Alabama	Baldwin County	9.19	gain	52562.0
2	Alabama	Barbour County	-6.22	no gain	33368.0
3	Alabama	Bibb County	0.73	gain	43404.0
4	Alabama	Blount County	0.68	gain	47412.0

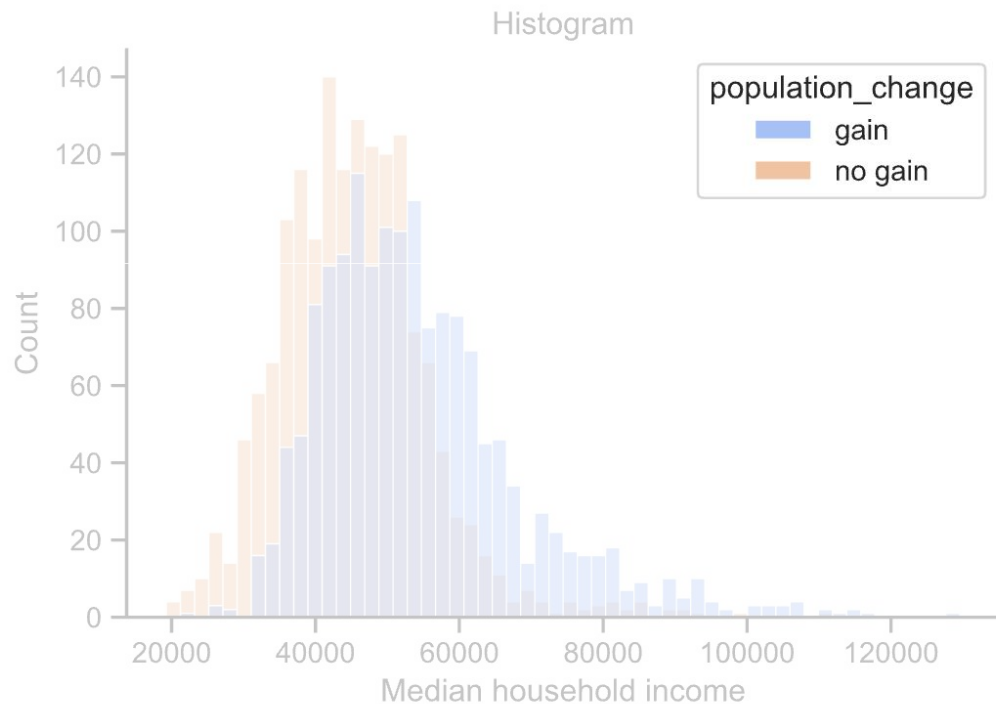
Median household income is on x-axis



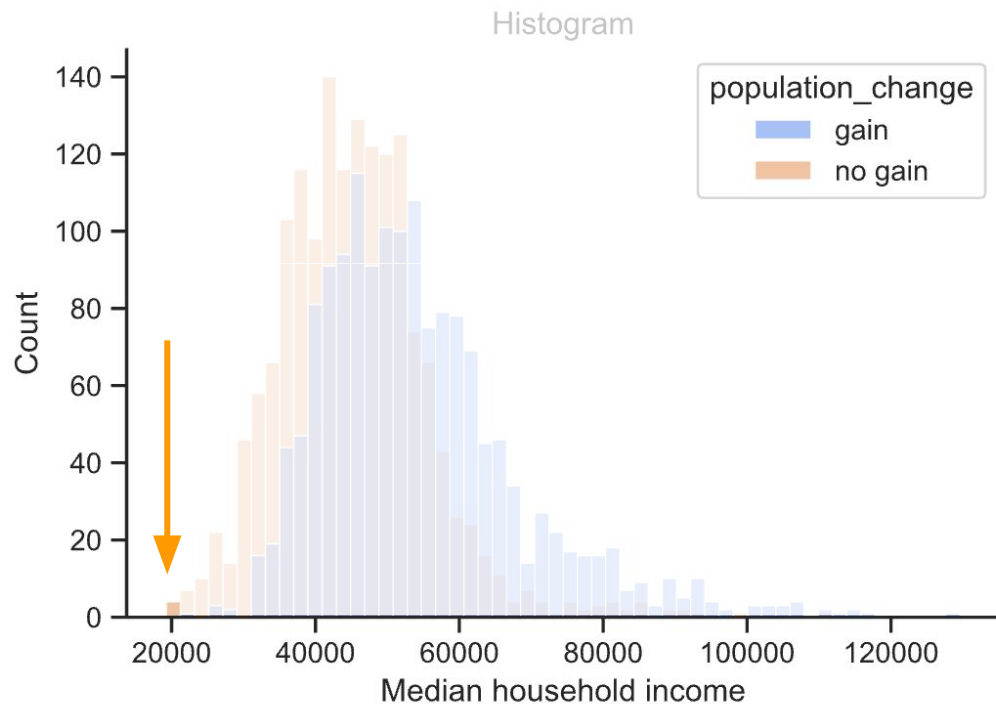
Count is on y-axis



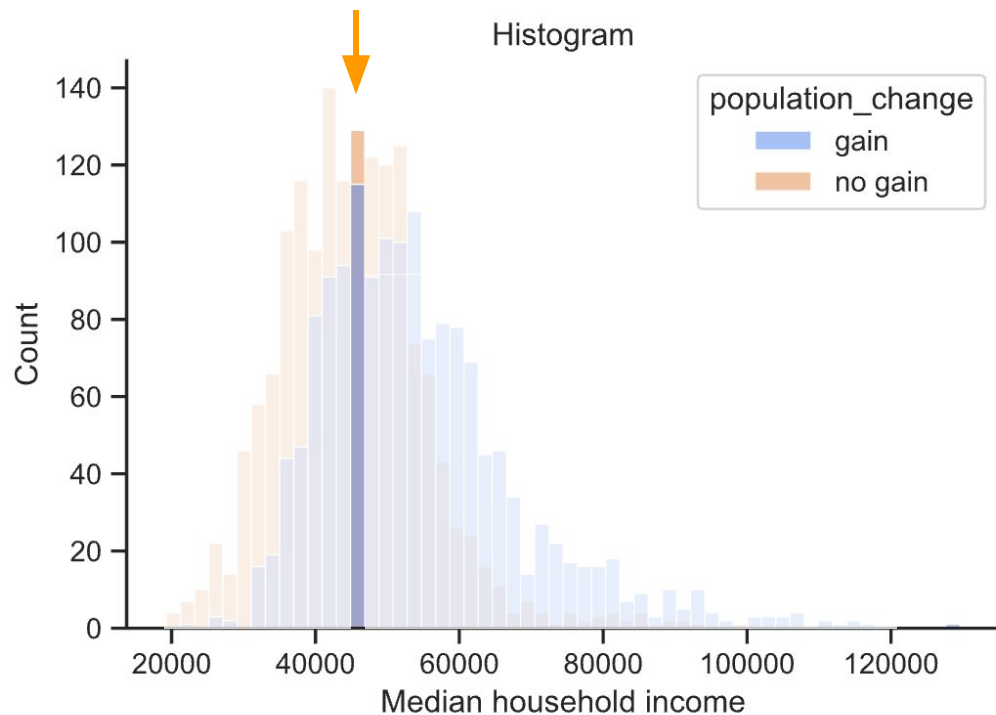
We compare two groups: **change in population**



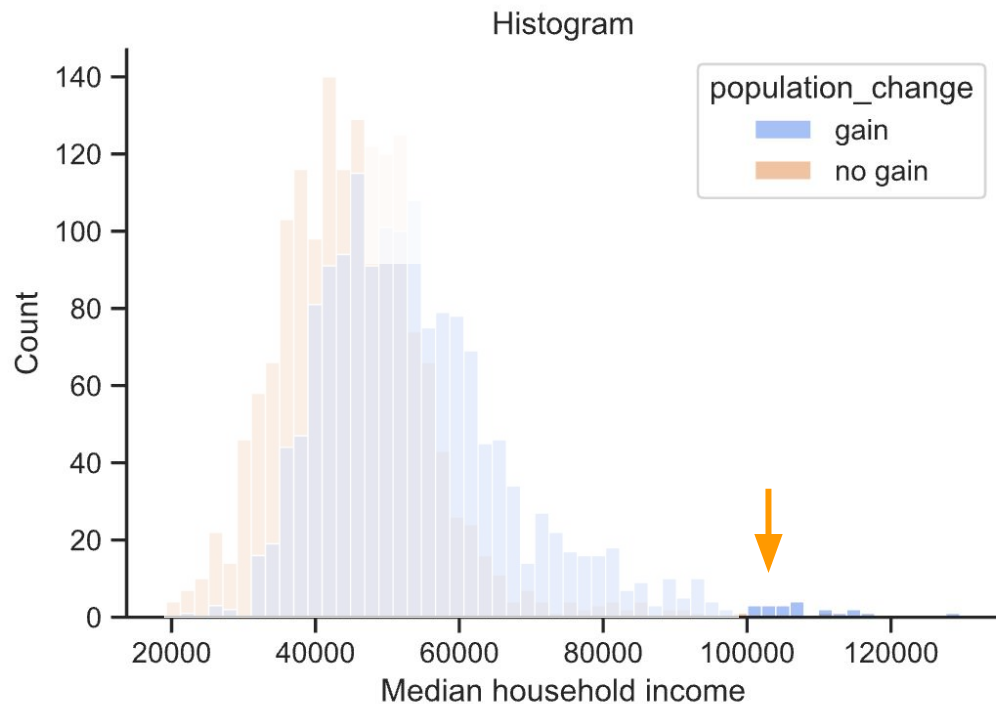
Only “no gain” in the first bin



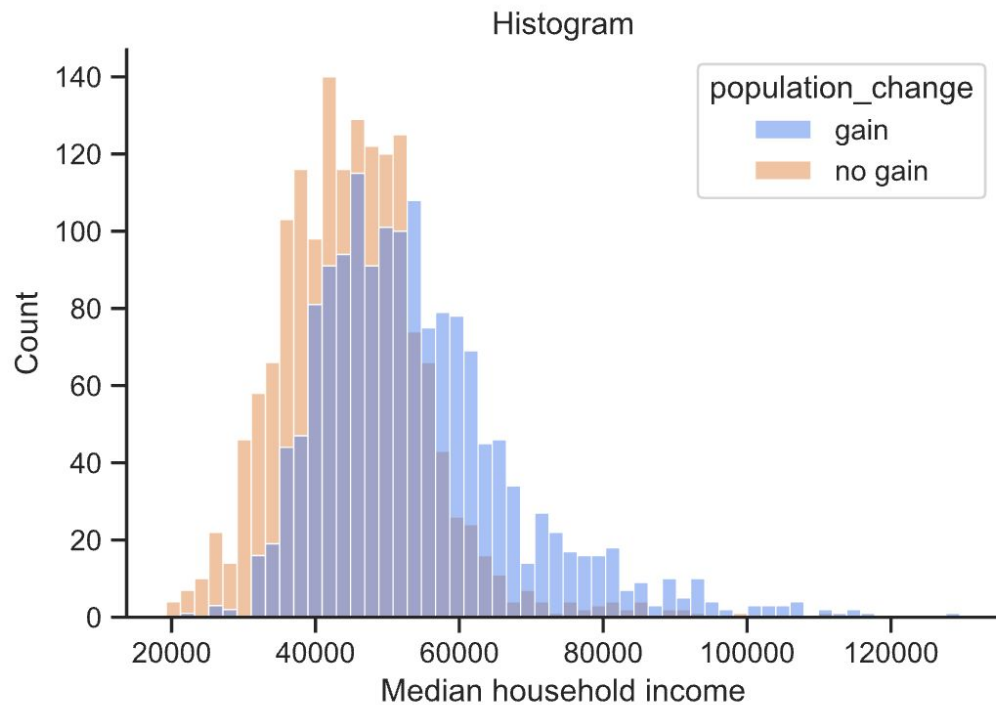
Now let's consider this bin



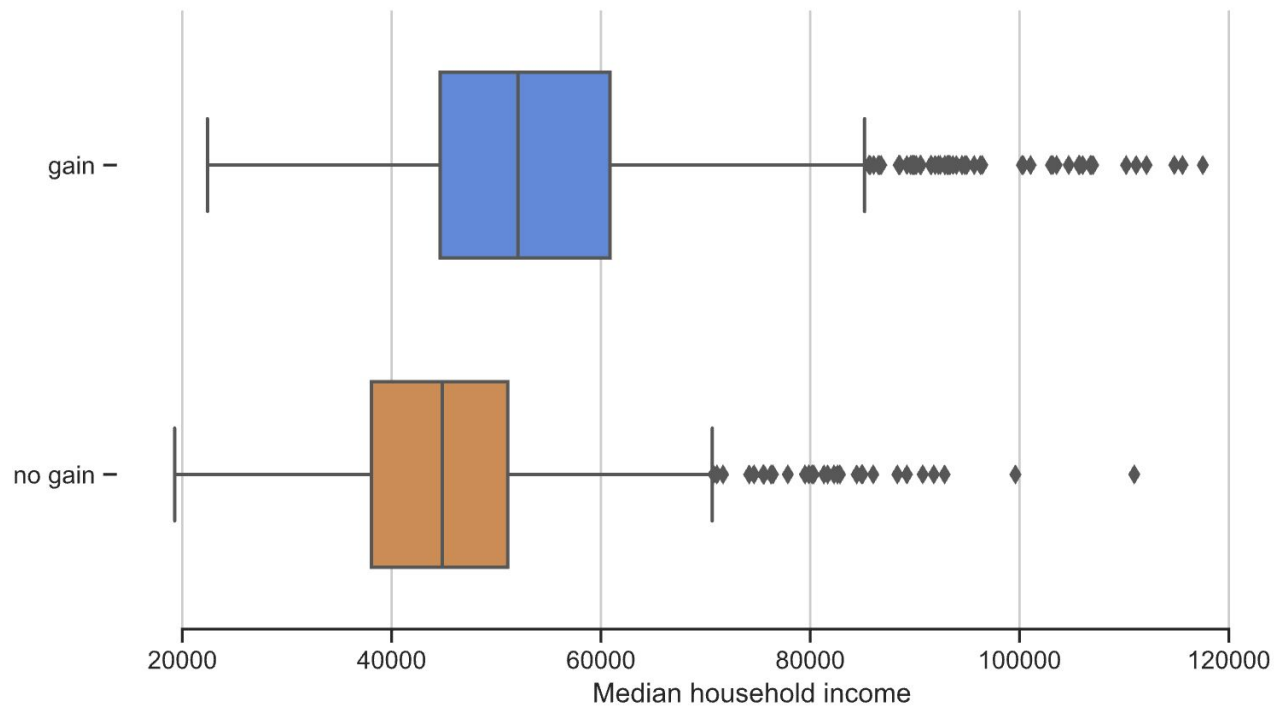
Now let's consider these bins



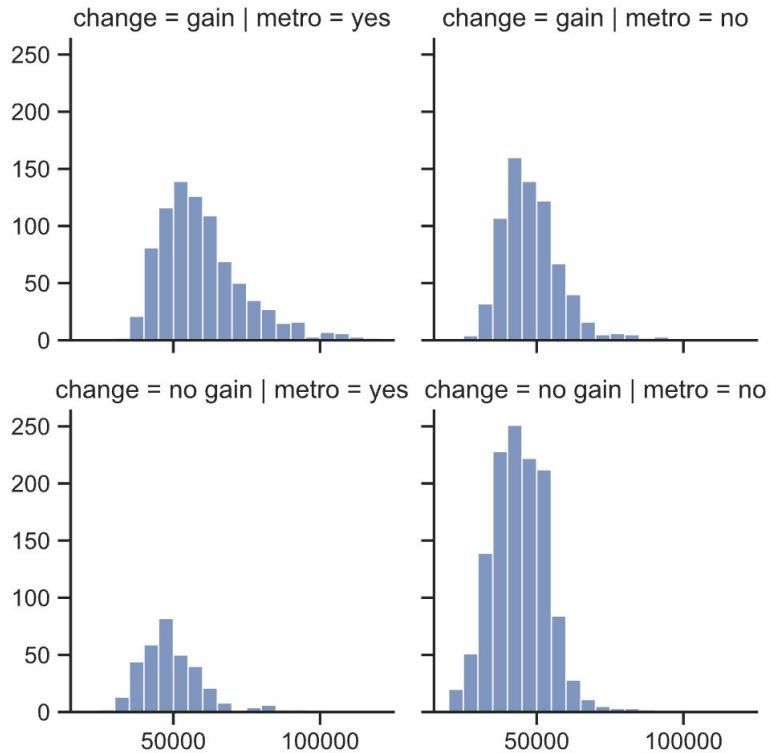
Histogram with two groups



Boxplot with two groups

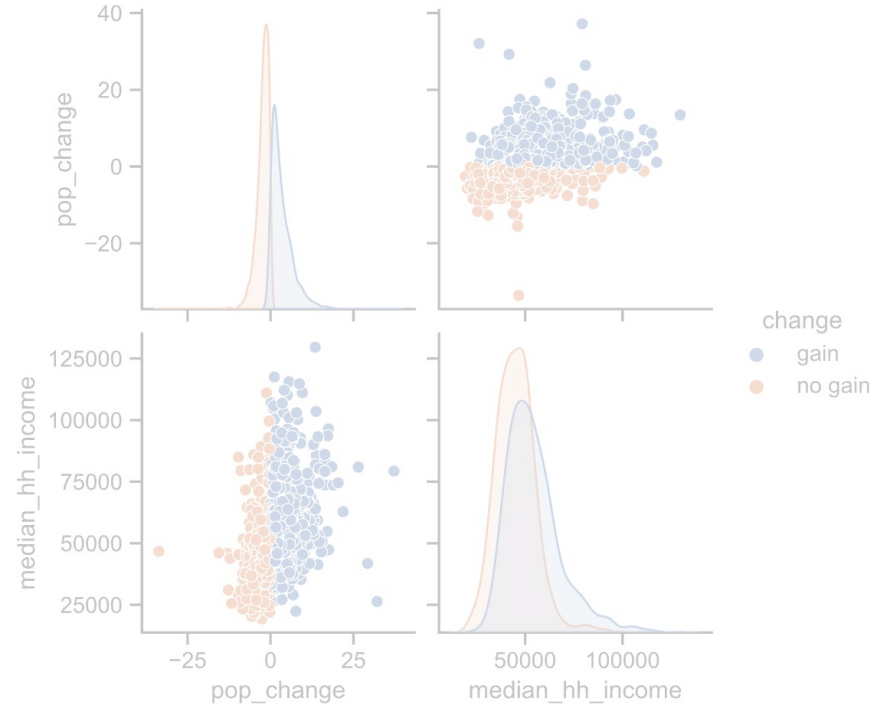
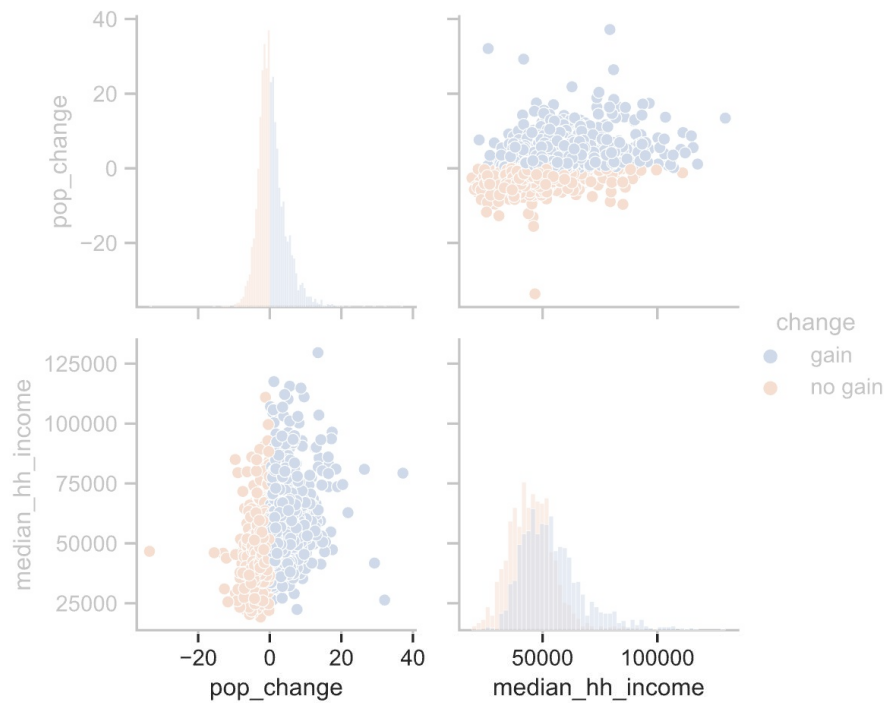


Faceting

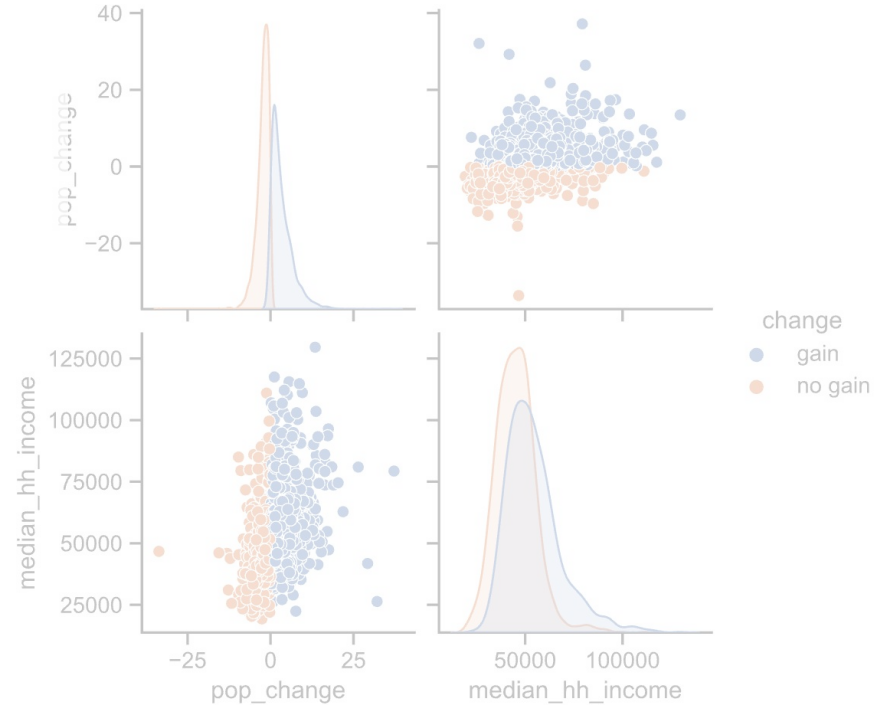
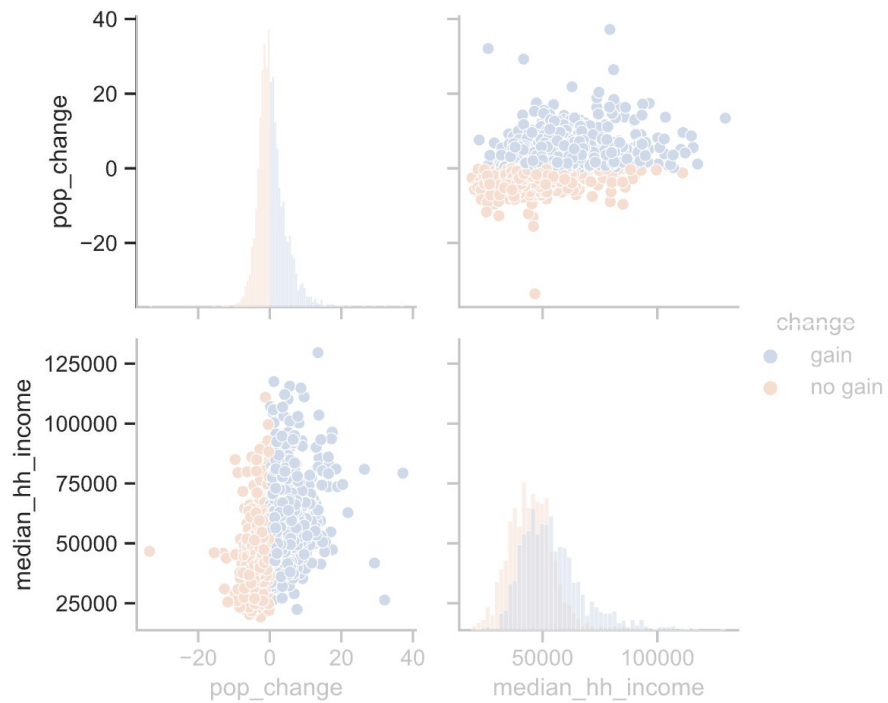


Additional graphs

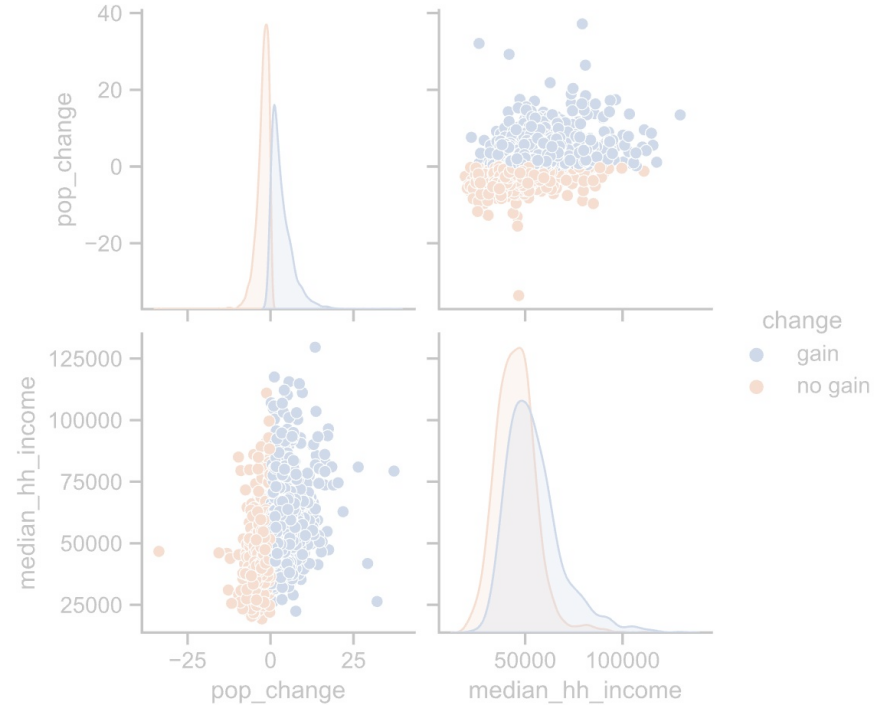
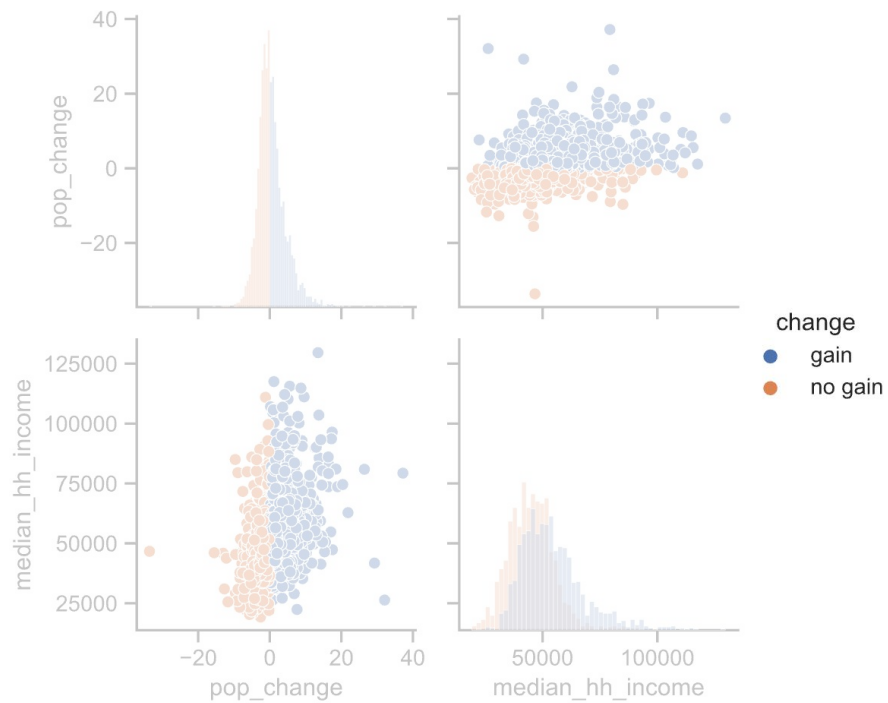
Plotting pairwise data relationships



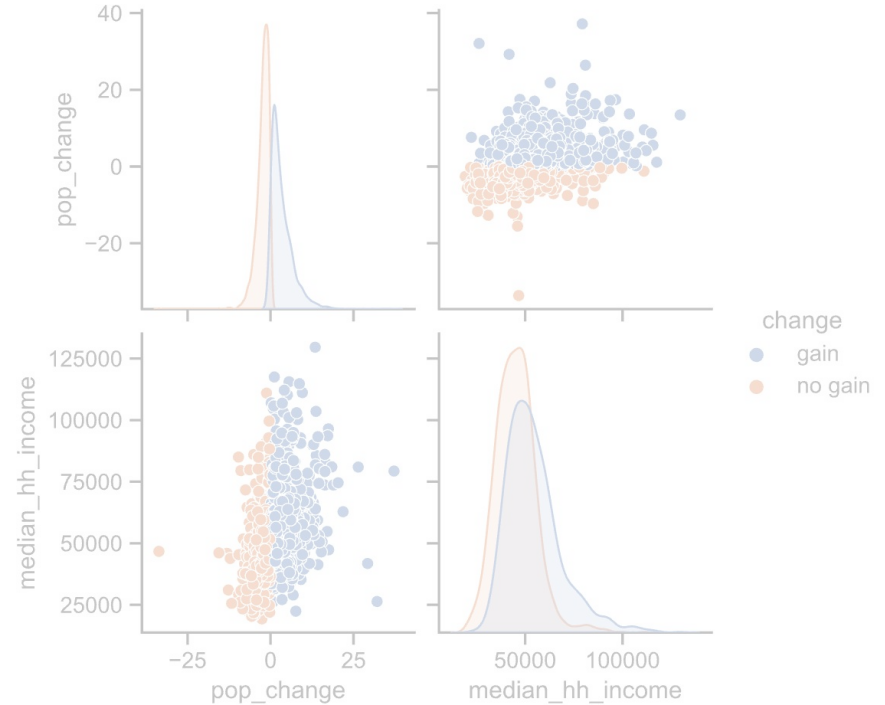
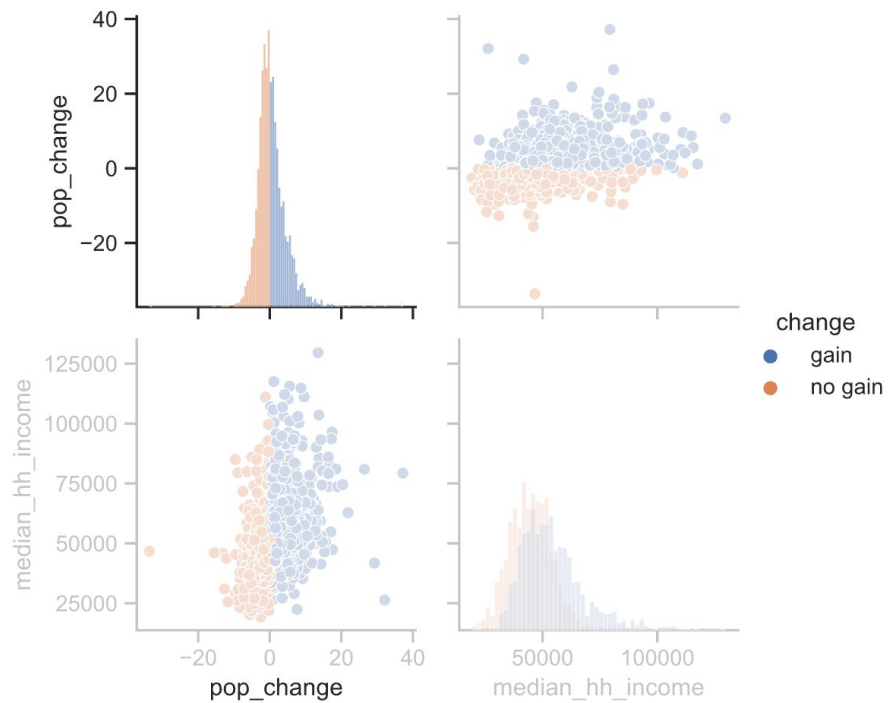
Plotting pairwise data relationships



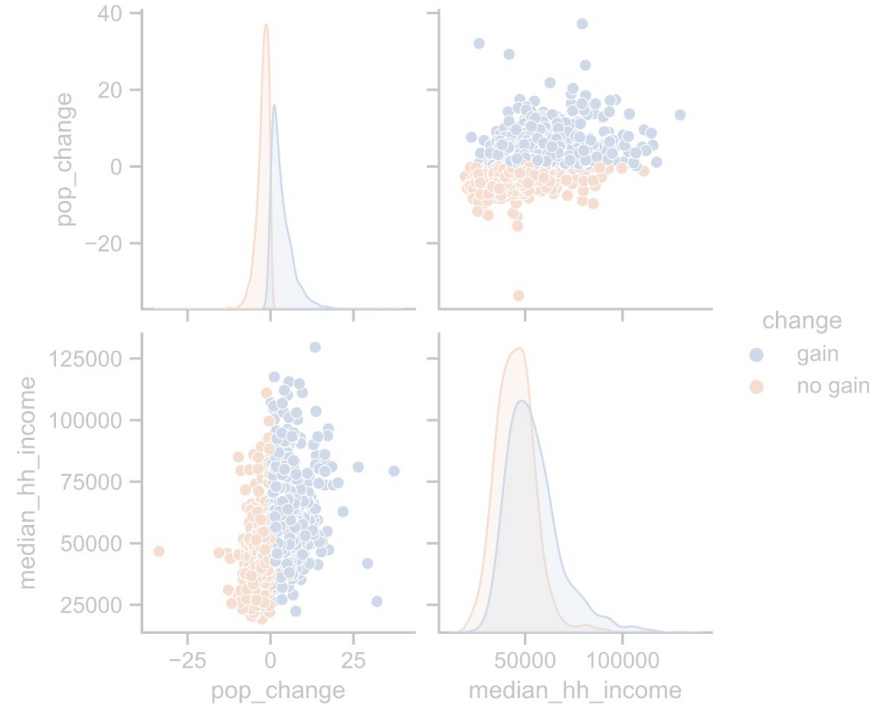
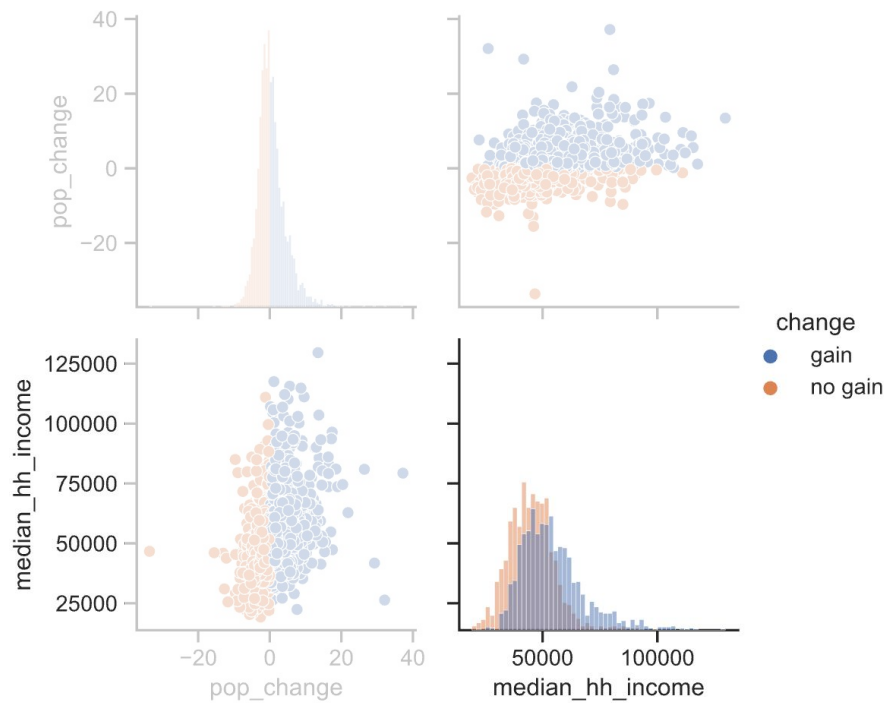
Plotting pairwise data relationships



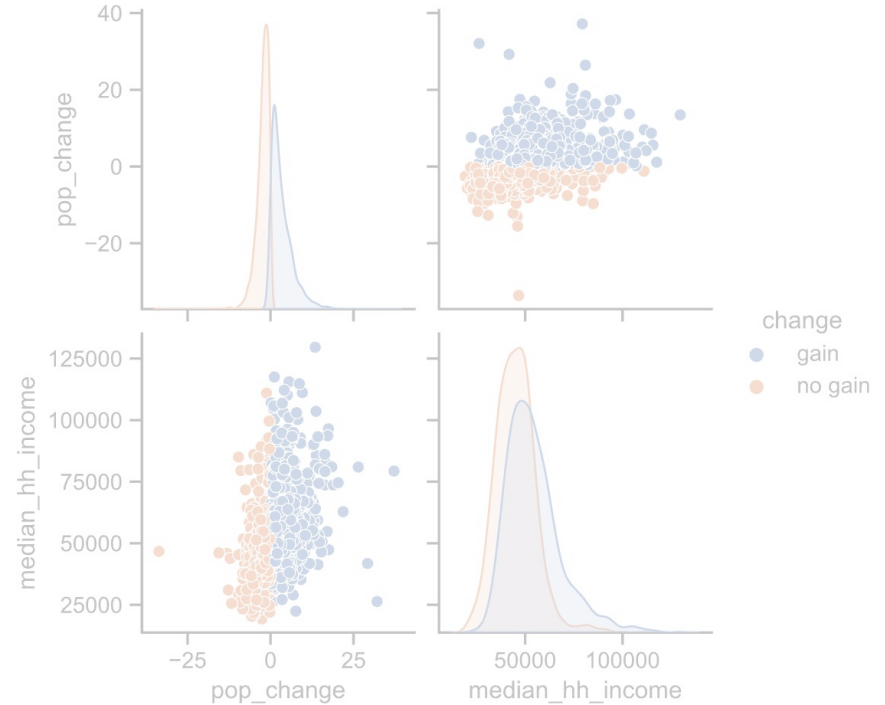
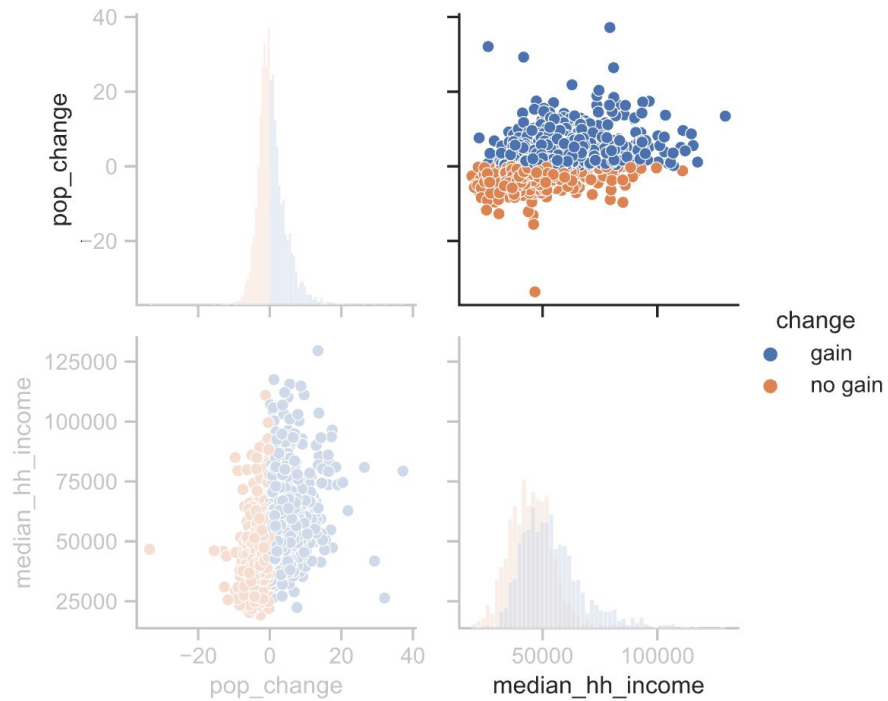
Plotting pairwise data relationships



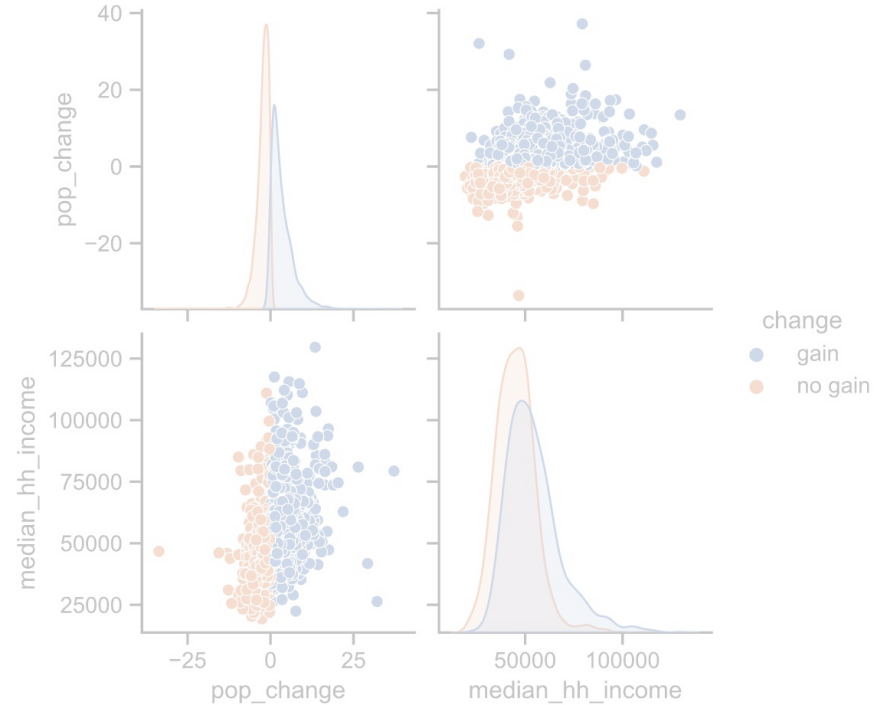
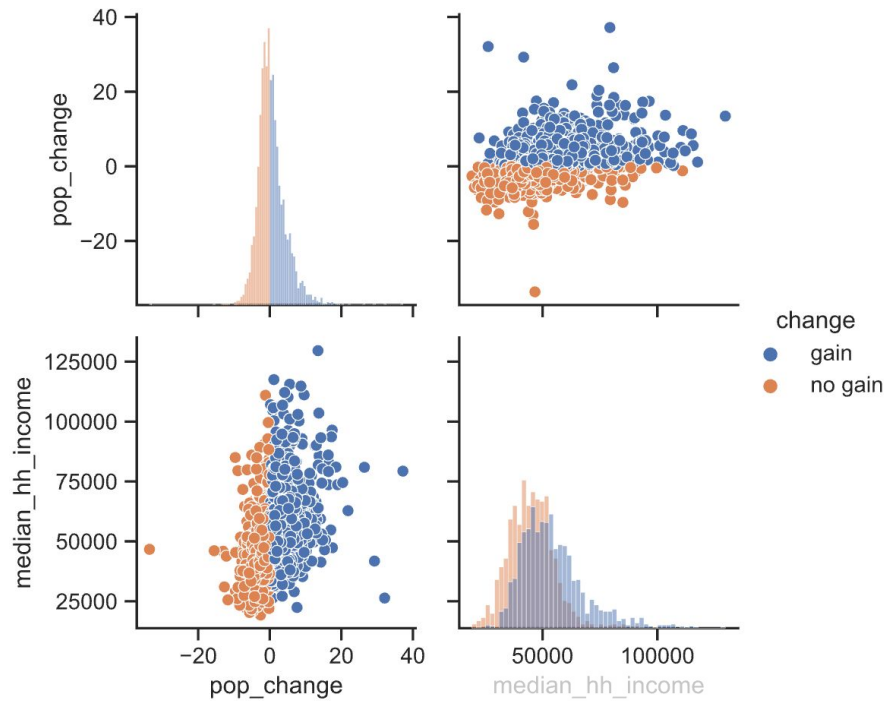
Plotting pairwise data relationships



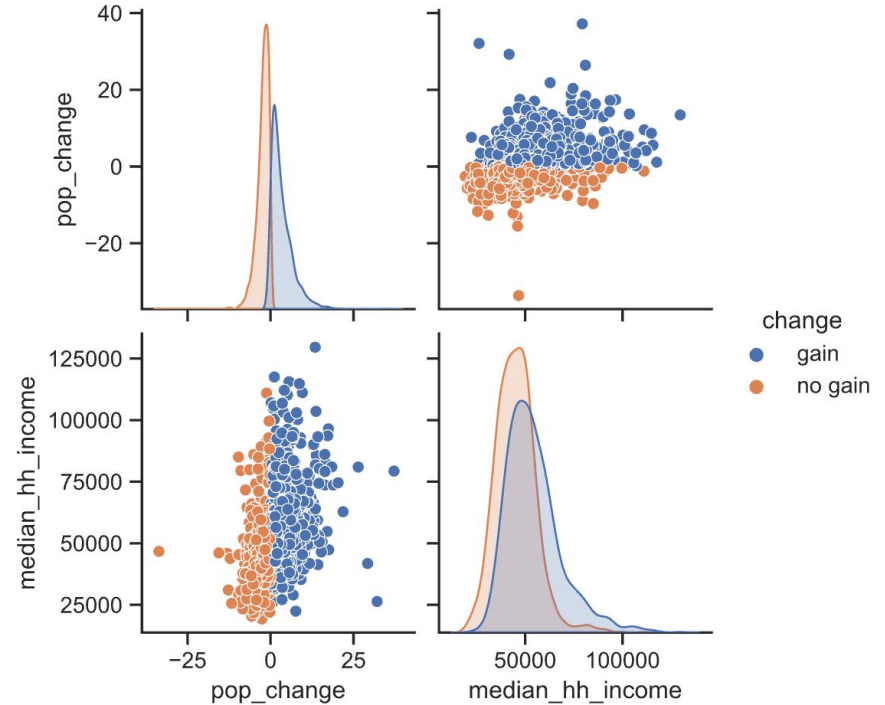
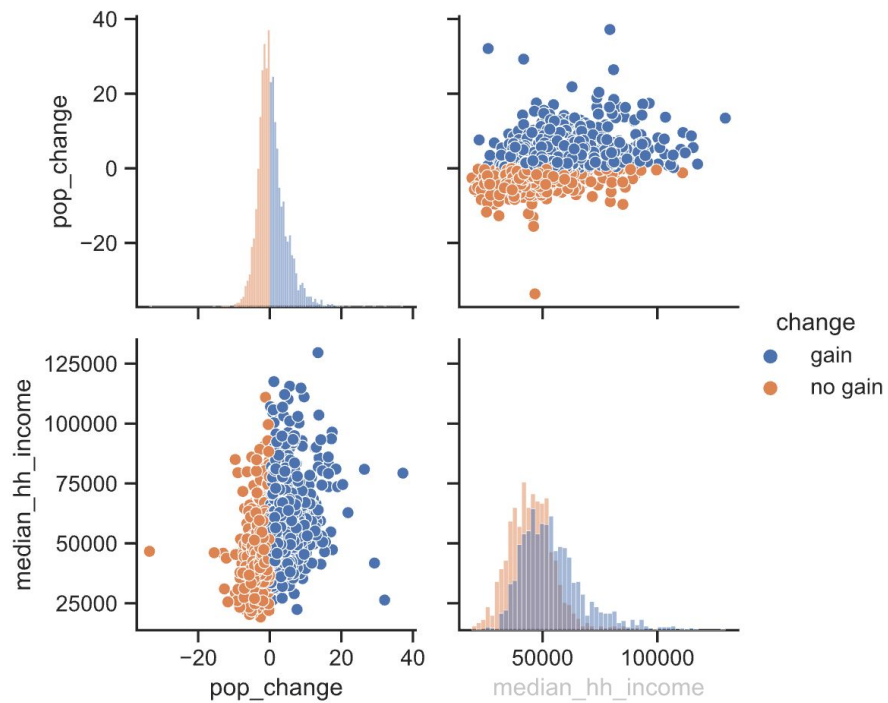
Plotting pairwise data relationships



Plotting pairwise data relationships



Plotting pairwise data relationships



Terms you should know

column proportions

faceted plot

row totals

column totals

filled bar plot

side-by-side box plot

conditional proportions

mosaic plot

stacked bar plot

contingency table

(ridge plot)

standardized bar plot

dodged bar plot

row proportions

Resources

The slides are based on the excellent book “Introduction to Modern Statistics” by Mine Çetinkaya-Rundel and Johanna Hardin.

The online version can be **accessed for free**:

<https://openintro-ims.netlify.app/explore-categorical.html>

