COVID-19's impact on homelessness in California

Bryant Nguyen, Josh Nola, Mati Teklemariam, Yash Sharma

Research Question:

What was the **impact of COVID-19 on the homeless populations** in various counties of California, and **what factors influenced these effects**?

Data Cleaning Process:

- Manually transferred data from several datasets to XLSX format
- Removed dirty data, irrelevant data, and data that was skewing analysis
- Removed 2021’s dataset (data from the year 2020)

Data overview

2019

County Name	Homeless Total
San Jose/Santa Clara	9706
San Francisco	8035
Oakland	8022
Sacramento	5561
Santa Rosa, Petaluma/Sonoma	2951
Richmond/Contra Costa	2295
Salinas/Monsterey, San Benito	2704
Marin	1034
Watsonville/Santa Cruz	2167
Mendocino	785
Turlock, Modesto/Stanislaus	1923
Stockton/San Joaquin	2631
Daly City/ San Mateo	1512
Visalia/Kings, Tulare	1064
Fresno	2508
Roseville, Rocklin/Placer	617
Redding/Shasta, Siskiyou, Lassen, Plumas, Del Norte, Modoc, Sierra	1349
Napa	322
Vallejo/Solano	1151
Chico, Paradise/Butte	1266
Merced	608
Davis, Woodland/Yolo	655
Humboldt	1702
Colusa, Glenn, Trinity	192
Yuba, Sutter	721
El Dorado	613
Amador, Calaveras, Mariposa, Tuolumne	845
Tehama	288
Lake	408
Alpine, Inyo, Mono	214
Nevada	415
Los Angeles	56257
San Diego	8102
Santa Ana, Anaheim/Orange	6860
Santa Maria/Santa Barbara	1803
Bakersfield/Kern	1330
Long Beach	1894
Pasadena	542
Riverside	2811
San Bernardino	2607
Oxnard, San Buenaventura/Ventura	1669
Glendale	243
Imperial	1413
San Luis Obispo	1483
California	151278

2020

County Name	Homeless Total
San Jose/Santa Clara	9605
San Francisco	8124
Oakland	8137
Sacramento	5511
Santa Rosa, Petaluma/Sonoma	2745
Richmond/Contra Costa	2277
Salinas/Monsterey, San Benito	2683
Marin	1032
Watsonville/Santa Cruz	2256
Mendocino	751
Turlock, Modesto/Stanislaus	2107
Stockton/San Joaquin	2677
Daly City/ San Mateo	1572
Visalia/Kings, Tulare	1297
Fresno	3641
Roseville, Rocklin/Placer	744
Redding/Shasta, Siskiyou, Lassen, Plumas, Del Norte, Modoc, Sierra	1529
Napa	464
Vallejo/Solano	1162
Chico, Paradise/Butte	1274
Merced	636
Davis, Woodland/Yolo	641
Humboldt	1701
Colusa, Glenn, Trinity	261
Yuba, Sutter	721
El Dorado	663
Amador, Calaveras, Mariposa, Tuolumne	834
Tehama	300
Lake	357
Alpine, Inyo, Mono	184
Nevada	387
Los Angeles	63706
San Diego	7638
Santa Ana, Anaheim/Orange	6978
Santa Maria/Santa Barbara	1897
Bakersfield/Kern	1580
Long Beach	2034
Pasadena	527
Riverside	2884
San Bernardino	3125
Oxnard, San Buenaventura/Ventura	1787
Glendale	169
Imperial	1527
San Luis Obispo	1423
California	161548

2022

County Name	Homeless Total
San Jose/Santa Clara	10028
San Francisco	7754
Oakland	9747
Sacramento	9278
Santa Rosa, Petaluma/Sonoma	2893
Richmond/Contra Costa	3093
Salinas/Monsterey, San Benito	2404
Marin	1121
Watsonville/Santa Cruz	2299
Mendocino	830
Turlock, Modesto/Stanislaus	1857
Stockton/San Joaquin	2319
Daly City/ San Mateo	1808
Visalia/Kings, Tulare	1235
Fresno	4216
Roseville, Rocklin/Placer	750
Redding/Shasta, Siskiyou, Lassen, Plumas, Del Norte, Modoc, Sierra	1837
Napa	495
Vallejo/Solano	1179
Chico, Paradise/Butte	1006
Merced	855
Davis, Woodland/Yolo	746
Humboldt	1648
Colusa, Glenn, Trinity	340
Yuba, Sutter	1094
El Dorado	511
Amador, Calaveras, Mariposa, Tuolumne	625
Tehama	291
Lake	339
Alpine, Inyo, Mono	140
Nevada	527
Los Angeles	65111
San Diego	8427
Santa Ana, Anaheim/Orange	5718
Santa Maria/Santa Barbara	1962
Bakersfield/Kern	1603
Long Beach	3296
Pasadena	512
Riverside	3316
San Bernardino	3333
Oxnard, San Buenaventura/Ventura	2248
Glendale	225
Imperial	1057
San Luis Obispo	1448
California	171521

2023

County Name	Homeless Total
San Jose/Santa Clara	9903
San Francisco	7582
Oakland	9759
Sacramento	9281
Santa Rosa, Petaluma/Sonoma	2266
Richmond/Contra Costa	2372
Salinas/Monsterey, San Benito	2212
Marin	1118
Watsonville/Santa Cruz	1804
Mendocino	633
Turlock, Modesto/Stanislaus	2091
Stockton/San Joaquin	2454
Daly City/ San Mateo	1859
Visalia/Kings, Tulare	1470
Fresno	4493
Roseville, Rocklin/Placer	709
Redding/Shasta, Siskiyou, Lassen, Plumas, Del Norte, Modoc, Sierra	2521
Napa	506
Vallejo/Solano	1200
Chico, Paradise/Butte	1237
Merced	784
Davis, Woodland/Yolo	737
Humboldt	1656
Colusa, Glenn, Trinity	360
Yuba, Sutter	963
El Dorado	491
Amador, Calaveras, Mariposa, Tuolumne	731
Tehama	304
Lake	460
Alpine, Inyo, Mono	88
Nevada	492
Los Angeles	71320
San Diego	10264
Santa Ana, Anaheim/Orange	6050
Santa Maria/Santa Barbara	1887
Bakersfield/Kern	1948
Long Beach	3447
Pasadena	556
Riverside	3725
San Bernardino	4195
Oxnard, San Buenaventura/Ventura	2441
Glendale	195
Imperial	1303
San Luis Obispo	1532
California	181399

Data overview

County Incomes

	A	B
1	County Name	Median Income
2	San Jose/Santa Clara	153792
3	Daly City/San Mateo	149907
4	Marin	142019
5	San Francisco	136689
6	Oakland	122488
7	Richmond/Contra Costa	120020
8	Roseville, Rocklin/Placer	109375
9	Santa Ana, Anaheim/Orange	109361
10	Napa	105809
11	Watsonville/Santa Cruz	104409
12	Oxnard, San Buenaventura/Ventura	102141
13	Santa Rosa, Petaluma/Sonoma	99266
14	El Dorado	99246
15	Salinas/Monterey, San Benito	97747
16	Vallejo/Solano	97037
17	San Diego	96974
18	Santa Maria/Santa Barbara	92332
19	San Luis Obispo	90158
20	Davis, Woodland/Yolo	85097
21	Riverside	84505
22	Sacramento	84010
23	Los Angeles	83411
24	Long Beach	83411
25	Pasadena	83411
26	Glendale	83411
27	Stockton/San Joaquin	82837
28	Alpine, Inyo, Mono	82193
29	Nevada	79395
30	San Bernardino	77423
31	Turlock, Modesto/Stanislaus	74872
32	Amador, Calaveras, Mariposa, Tuolumne	70708
33	Yuba, Sutter	69673
34	Fresno	67756
35	Visalia/Kings, Tulare	66507
36	Chico, Paradise/Butte	66085
37	Merced	64772
38	Bakersfield/Kern	63883
39	Mendocino	61335
40	Redding/Shasta, Siskiyou, Lassen, Plumas, Del Norte, Modoc, Sierra	60980
41	Colusa, Glenn, Trinity	60323
42	Tehama	59029
43	Humboldt	57881
44	Lake	56259
45	Imperial	53847
46	California	83411
47		

Monetary Relief

	A	B
1	County Name	Monetary Relief
2	San Jose/Santa Clara	$31,514,029.00
3	Daly City/San Mateo	$17,794,236.00
4	Marin	$26,594,123.00
5	San Francisco	$20,660,712.00
6	Oakland	n/a
7	Richmond/Contra Costa	$26,546,260.00
8	Roseville, Rocklin/Placer	$41,162,055.00
9	Santa Ana, Anaheim/Orange	$73,509,392.00
10	Napa	$14,181,303.00
11	Watsonville/Santa Cruz	$27,654,703.00
12	Oxnard, San Buenaventura/Ventura	$19,396,868.00
13	Santa Rosa, Petaluma/Sonoma	$50,263,852.00
14	El Dorado	$19,701,272.00
15	Salinas/Monterey, San Benito	$6,357,463.00
16	Vallejo/Solano	$44,884,892.00
17	San Diego	$53,702,472.00
18	Santa Maria/Santa Barbara	$46,069,250.00
19	San Luis Obispo	$28,269,109.00
20	Davis, Woodland/Yolo	$22,604,867.00
21	Riverside	$56,203,389.00
22	Sacramento	$24,982,972.00
23	Los Angeles	$163,402,515.00
24	Long Beach	n/a
25	Pasadena	n/a
26	Glendale	$2,997,601.00
27	Stockton/San Joaquin	$17,803,165.00
28	Alpine, Inyo, Mono	$1,894,810.00
29	Nevada	$10,003,626.00
30	San Bernardino	$50,179,490.00
31	Turlock, Modesto/Stanislaus	$12,834,248.00
32	Amador, Calaveras, Mariposa, Tuolumne	$5,599,294.00
33	Yuba, Sutter	$10,272,390.00
34	Fresno	$16,437,637.00
35	Visalia/Kings, Tulare	$15,661,750.00
36	Chico, Paradise/Butte	$21,441,104.00
37	Merced	$28,907,577.00
38	Bakersfield/Kern	$21,115,139.00
39	Mendocino	$8,966,905.00
40	Redding/Shasta, Siskiyou, Lassen, Plumas, Del Norte, Modoc, Sierra	$18,153,328.00
41	Colusa, Glenn, Trinity	$2,233,111.00
42	Tehama	$6,640,501.00
43	Humboldt	$13,591,367.00
44	Lake	$6,529,468.00
45	Imperial	$19,247,554.00
46	California	$1,105,965,799.00
47		

Hypothesis 1:

Less affluent counties were more impacted by COVID-19 than more affluent counties, correlating with higher homelessness.

[544]:
```python
# Load the datasets
df_countyincomes = pd.read_csv('HomelessnessbyCountyOverTime - County Incomes.csv', index_col=0)
df_monetaryrelief = pd.read_csv('HomelessnessbyCountyOverTime - Monetary Relief.csv', index_col=0)
```

[546]:
```python
# Merge the datasets on "County Name"
df = pd.merge(df_countyincomes, df_monetaryrelief, on="County Name", how="inner")
# Rename columns for clarity
df.rename(columns={"Median Income": "Median_Income", "Monetary Relief": "Monetary_Relief"}, inplace=True)
```

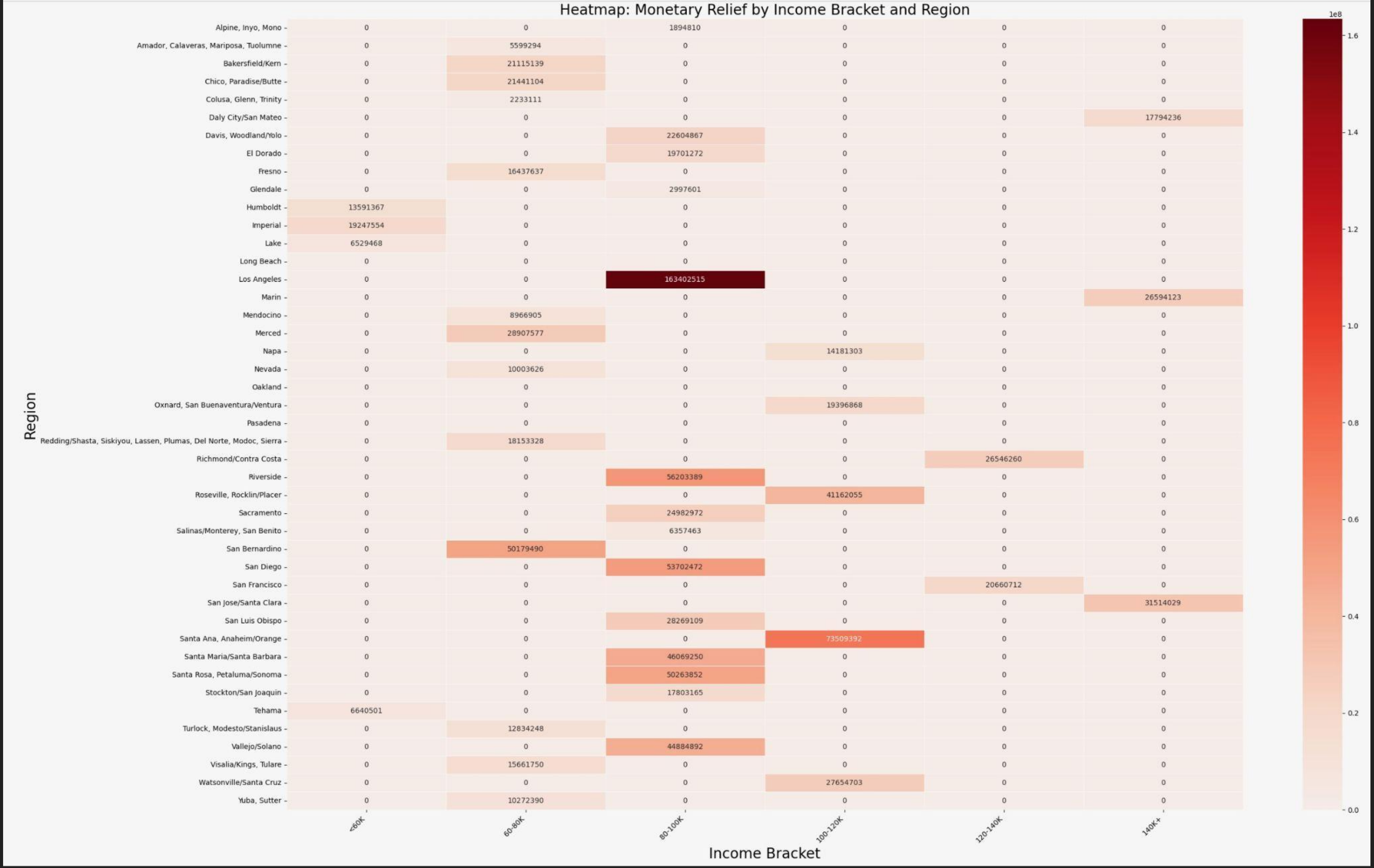
[548]:
```python
# Handle missing values
df["Monetary_Relief"] = pd.to_numeric(df["Monetary_Relief"], errors="coerce")
df.fillna(0, inplace=True)
```

[550]:
```python
# Step 1: Create income brackets
bins = [0, 60000, 80000, 100000, 120000, 140000, 160000]
labels = ["<60K", "60-80K", "80-100K", "100-120K", "120-140K", "140K+"]
df["Income Bracket"] = pd.cut(df["Median_Income"], bins=bins, labels=labels, include_lowest=True)
df = df[df.index != "California"]
```

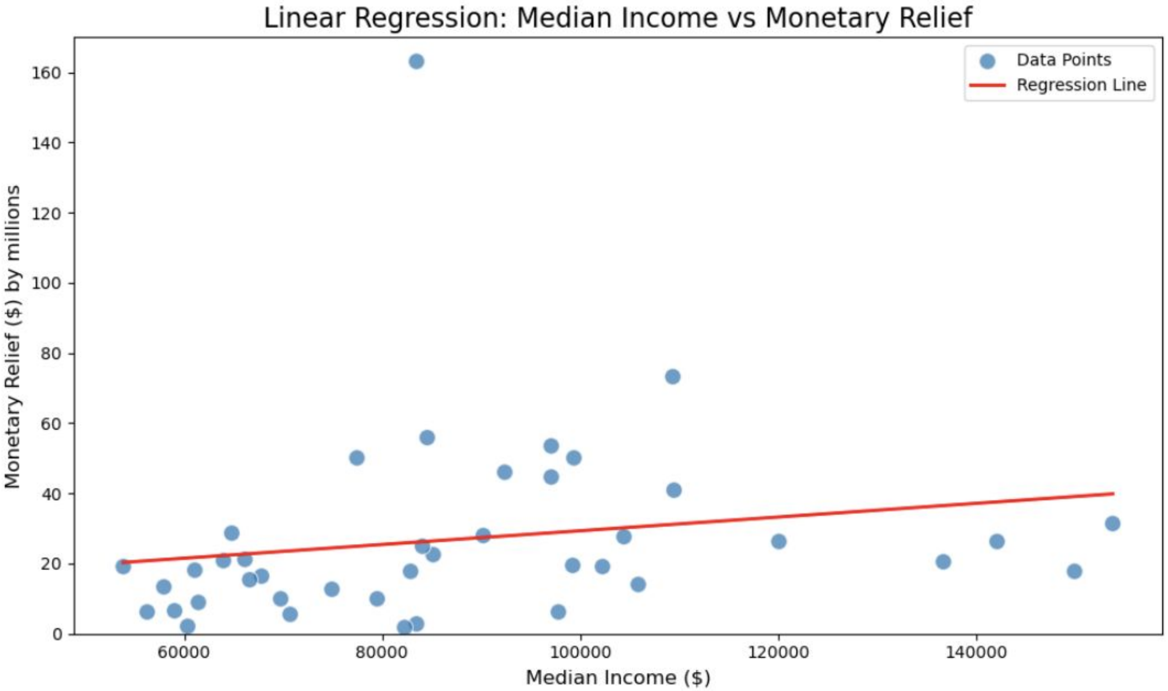
[552]:
```python
# Step 2: Create a pivot table for the heatmap
heatmap_data = df.pivot_table(values="Monetary_Relief", index="Income Bracket", columns="County Name", aggfunc="sum")
# Step 3: Transpose the pivot table to have regions on the x-axis and income brackets on the y-axis
heatmap_data = heatmap_data.transpose()
```

```
/var/folders/s_/b8ftbst93f787fwsbl3mjbf40000gn/T/ipykernel_14647/1156688371.py:2: FutureWarning: The default value of observed=False is deprecated and will change to observed=True in a future version of pandas. Specify observed=False to silence this warning and retain the current behavior
  heatmap_data = df.pivot_table(values="Monetary_Relief", index="Income Bracket", columns="County Name", aggfunc="sum")
```

[554]:
```python
# Step 4: Plot the heatmap
plt.figure(figsize=(30, 20))
sns.heatmap(heatmap_data, cmap="Reds", annot=True, fmt=".0f", cbar=True, linewidths=0.5)
plt.title("Heatmap: Monetary Relief by Income Bracket and Region", fontsize=20)
plt.xlabel("Income Bracket", fontsize=20)
plt.ylabel("Region", fontsize=20)
plt.xticks(rotation=45, ha="right", fontsize=10)
plt.yticks(fontsize=10)
plt.show()
```

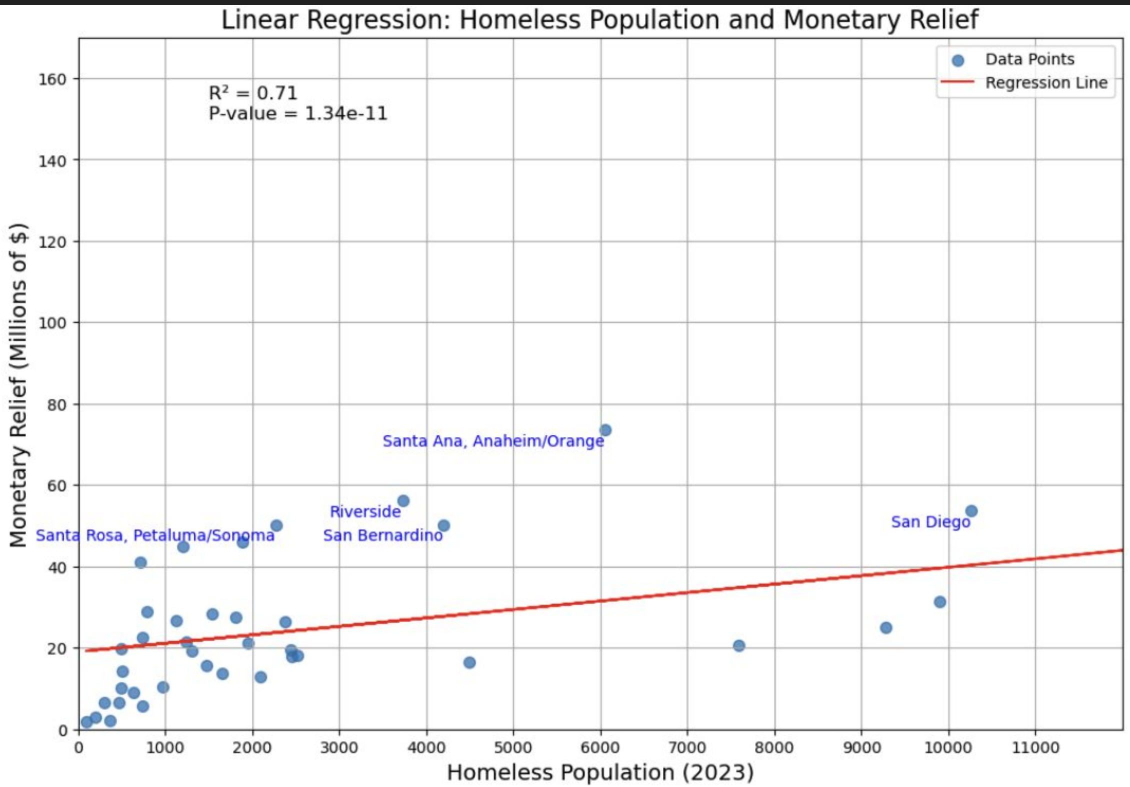

Heatmap: Monetary Relief by Income Bracket and Region
Region
Income Bracket
<60K
60-80K
80-100K
100-120K
120-140K
140K+
Alpine, Inyo, Mono 0 0 1894810 0 0 0
Amador, Calaveras, Mariposa, Tuolumne 0 5599294 0 0 0 0
Bakersfield/Kern 0 21115139 0 0 0 0
Chico, Paradise/Butte 0 21441104 0 0 0 0
Colusa, Glenn, Trinity 0 2233111 0 0 0 0
Daly City/San Mateo 0 0 0 0 0 17794236
Davis, Woodland/Yolo 0 0 22604867 0 0 0
El Dorado 0 0 19701272 0 0 0
Fresno 0 16437637 0 0 0 0
Glendale 0 0 2997601 0 0 0
Humboldt 13591367 0 0 0 0 0
Imperial 19247554 0 0 0 0 0
Lake 6529468 0 0 0 0 0
Long Beach 0 0 0 0 0 0
Los Angeles 0 0 163402515 0 0 0
Marin 0 0 0 0 0 26594123
Mendocino 0 8966905 0 0 0 0
Merced 0 28907577 0 0 0 0
Napa 0 0 0 14181303 0 0
Nevada 0 10003626 0 0 0 0
Oakland 0 0 0 0 0 0
Oxnard, San Buenaventura/Ventura 0 0 0 19396868 0 0
Pasadena 0 0 0 0 0 0
Redding/Shasta, Siskiyou, Lassen, Plumas, Del Norte, Modoc, Sierra 0 18153328 0 0 0 0
Richmond/Contra Costa 0 0 0 0 26546260 0
Riverside 0 0 56203389 0 0 0
Roseville, Rocklin/Placer 0 0 0 41162055 0 0
Sacramento 0 0 24982972 0 0 0
Salinas/Monterey, San Benito 0 0 6357463 0 0 0
San Bernardino 0 50179490 0 0 0 0
San Diego 0 0 53702472 0 0 0
San Francisco 0 0 0 0 20660712 0
San Jose/Santa Clara 0 0 0 0 0 31514029
San Luis Obispo 0 0 28269109 0 0 0
Santa Ana, Anaheim/Orange 0 0 0 73509392 0 0
Santa Maria/Santa Barbara 0 0 46069250 0 0 0
Santa Rosa, Petaluma/Sonoma 0 0 50263852 0 0 0
Stockton/San Joaquin 0 0 17803165 0 0 0
Tehama 6640501 0 0 0 0 0
Turlock, Modesto/Stanislaus 0 12834248 0 0 0 0
Vallejo/Solano 0 0 44884892 0 0 0
Visalia/Kings, Tulare 0 15661750 0 0 0 0
Watsonville/Santa Cruz 0 0 0 27654703 0 0
Yuba, Sutter 0 10272390 0 0 0 0
1e8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0

```python
# Remove the California dataset
df_merged = df_merged[df_merged.index != "California"]

# Prepare the data
X = df_merged[['MedianIncome']]  # Median Income as independent variable
y = df_merged['MonetaryRelief'] / 1_000_000 # Monetary Relief as dependent variable

# Calculate correlation and regression statistics
slope, intercept, r_value, p_value, std_err = linregress(df_merged['MedianIncome'], y)

# Create a linear regression model
model = LinearRegression()
model.fit(X, y)

# Extract R-squared and p-value
r_squared = regression.rvalue**2
p_value = regression.pvalue

# Predict values
y_pred = model.predict(X)


# Visualize the results
plt.figure(figsize=(10, 6))
sns.scatterplot(x=X['MedianIncome'], y=y, label='Data Points', s=100, alpha=0.7)
plt.plot(X, y_pred, color='red', label='Regression Line', linewidth=2)
plt.title('Linear Regression: Median Income vs Monetary Relief', fontsize=16)
plt.xlabel('Median Income ($)', fontsize=12)
plt.ylabel('Monetary Relief ($) by millions', fontsize=12)
plt.ylim(0, 170)  # Set y-axis limits
plt.yticks(ticks=range(0, 171, 20))  # Set y-axis ticks from 20 to 170 with steps of 20
plt.legend()
plt.tight_layout()
plt.show()


# Print model statistics
print(f"Intercept: {model.intercept_:.2f}")
print(f"Coefficient: {model.coef_[0]:.2f}")
print(f"R-squared: {model.score(X, y):.2f}")
print(f"R-value (Correlation Coefficient): {r_value:.2f}")
print(f"P-value: {p_value:.2e}")

# Identify outliers where counties received more than 50 million in monetary relief
outliers = df_merged[df_merged['MonetaryRelief'] > 50_000_000]
print("\nOutliers (Counties with more than 50 million in Monetary Relief):")
print(outliers[['MedianIncome','MonetaryRelief']])
```




```
Intercept: 9.80
R-squared: 0.03
R-value (Correlation Coefficient): 0.18
P-value: 1.34e-11

Outliers (Counties with more than 50 million in Monetary Relief):
                          MedianIncome  MonetaryRelief
County Name
Santa Ana, Anaheim/Orange       109361     73509392.00
Santa Rosa, Petaluma/Sonoma      99266     50263852.00
San Diego                        96974     53702472.00
Riverside                        84505     56203389.00
Los Angeles                      83411    163402515.00
San Bernardino                   77423     50179490.00
```

```python
df = df[df.index != "California"]

# Extract variables for regression
x = df["Homeless_Total_2023"]
y = df["Monetary_Relief"] / 1_000_000  # Convert monetary relief to millions

# Perform linear regression
regression = linregress(x, y)

# Extract R-squared and p-value
r_squared = regression.rvalue**2
p_value = regression.pvalue

# Plotting the results
plt.figure(figsize=(12, 8))
plt.scatter(x, y, alpha=0.7, label="Data Points", s=50)
plt.plot(x, regression.intercept + regression.slope * x, color="red", label="Regression Line")
plt.title("Linear Regression: Homeless Population and Monetary Relief", fontsize=16)
plt.xlabel("Homeless Population (2023)", fontsize=14)
plt.ylabel("Monetary Relief (Millions of $)", fontsize=14)  # Update y-axis label
plt.legend()
plt.grid(True)

# Adjust x-axis and y-axis range
plt.xlim(0, 12000)  # Set x-axis limits from 0 to 80,000
plt.ylim(0, 170)  # Set y-axis limits to 0 to 170 million to focus on the majority of data

# Set x-axis ticks separated by 1000
plt.xticks(ticks=range(0, 12000, 1000))

# Display the R-squared and p-value on the plot
plt.text(
    1500,  # Adjust x-coordinate of the text position
    150,  # Adjust y-coordinate to fit in the reduced range
    f"R² = {r_squared:.2f}\nP-value = {p_value:.2e}",
    fontsize=12,
    color="black",
)
# Annotate outliers like Los Angeles
outliers = df[(df["Homeless_Total_2023"] > 10000) | (df["Monetary_Relief"] > 50_000_000)]
for idx, row in outliers.iterrows():
    plt.annotate(
        idx,
        (row["Homeless_Total_2023"], row["Monetary_Relief"] / 1_000_000),
        fontsize=10,
        color="blue",
        ha="right",
        xytext=(0, -10),
        textcoords="offset points"
    )

plt.show()

# Print R-squared and p-value to console
print(f"R-squared: {r_squared:.2f}")
print(f"P-value: {p_value:.2e}")
# Identify and print outliers where counties have more than 10,000 homeless and received more than 50 million
print("\nOutliers (Counties with more more than 50 million in Monetary Relief):")
outliers = df[(df["Homeless_Total_2023"] > 10000) | (df["Monetary_Relief"] > 50_000_000)]
print(outliers[["Homeless_Total_2023", "Monetary_Relief"]])
```




```
R-squared: 0.71
P-value: 1.34e-11

Outliers (Counties with more more than 50 million in Monetary Relief):
                              Homeless_Total_2023  Monetary_Relief
County Name
Santa Rosa, Petaluma/Sonoma                  2266      50263852.00
Los Angeles                                 71320     163402515.00
San Diego                                   10264      53702472.00
Santa Ana, Anaheim/Orange                    6050      73509392.00
Riverside                                    3725      56203389.00
San Bernardino                               4195      50179490.00
```

Hypothesis 2:

Following the start of the data, California’s homeless population has increased but has since decreased to a level between where it started and its peak.

Hypothesis #2 - Bryant

Following the start of the data (4/14/20), California's homeless population has increased but has since decreased to a level between where it started and its peak. 2019-2023 (line graph over time for California (bolded in some way), smaller line graphs for 6-7 other counties)

This question will be answered by comparing homeless populations throughout the years 2019-2023 for the entire state as well as prominent counties

58]:
```
# print all the outputs in a cell
from IPython.core.interactiveshell import InteractiveShell
InteractiveShell.ast_node_interactivity = "all"
```

97]:
```
import seaborn as sns
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

91]:
```
#First importing the different CSV files needed to answer Hypothesis #2
df1 = pd.read_csv('HomelessnessbyCountyOverTime - 2019.csv')
df2 = pd.read_csv('HomelessnessbyCountyOverTime - 2020 (1).csv')
df3 = pd.read_csv('HomelessnessbyCountyOverTime - 2022.csv')
df4 = pd.read_csv('HomelessnessbyCountyOverTime - 2023.csv')
```

07]:
```
#Reviewing what is inside the files
df1.head()
```

07]:

	County Name	Homeless Total
0	San Jose/Santa Clara	9706
1	San Francisco	8035
2	Oakland	8022
3	Sacramento	5561
4	Santa Rosa, Petaluma/Sonoma	2951

11]:
```
#Reviewing what is inside the files
df2.head()
```

11]:

	County Name	Homeless Total
0	San Jose/Santa Clara	9605
1	San Francisco	8124
2	Oakland	8137
3	Sacramento	5511
4	Santa Rosa, Petaluma/Sonoma	2745

[227]:

```python
# Add a new column for the total homeless populations for all the years in the set
merged_df['Homeless Total_Sum'] = merged_df['Homeless Total_2019'] + merged_df['Homeless Total_2020'] \
+ merged_df['Homeless Total_2022'] + merged_df['Homeless Total_2023']

merged_df.head()
```

[227]:

	County Name	Homeless Total_2019	Homeless Total_2020	Homeless Total_2022	Homeless Total_2023	Homeless Total_Sum
0	Alpine, Inyo, Mono	214	184	140	88	626
1	Amador, Calaveras, Mariposa, Tuolumne	845	834	625	731	3035
2	Bakersfield/Kern	1330	1580	1603	1948	6461
3	California	151278	161548	171521	181399	665746
4	Chico, Paradise/Butte	1266	1274	1006	1237	4783

We are interested in the largest areas/counties only, since they make up the greatest percentage of California and therefore give us the biggest and most whole picture of the homeless trends

[235]:

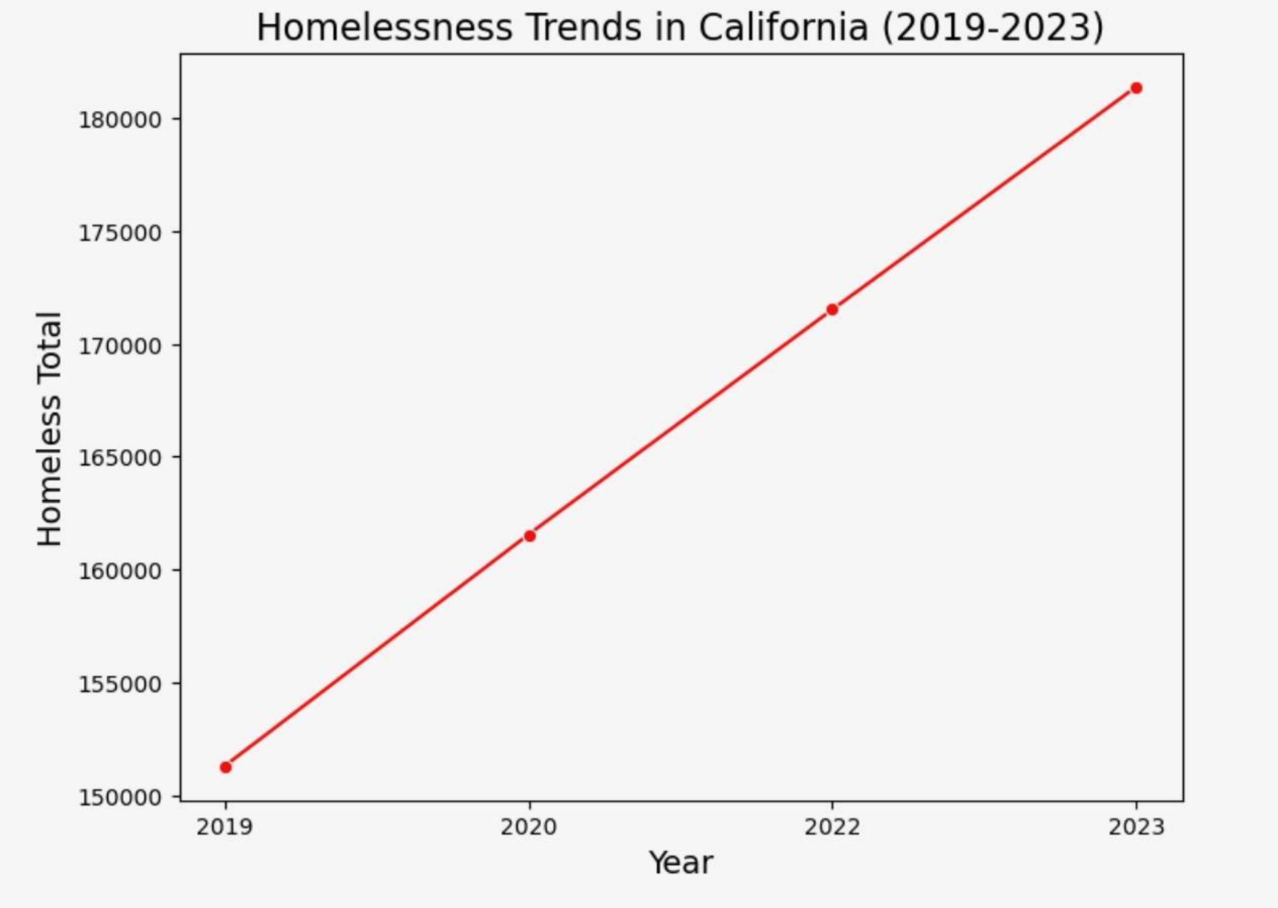
```python
top_5_largest = merged_df.nlargest(6, 'Homeless Total_Sum')

# Print the top 5 largest rows by the total
top_5_largest
```

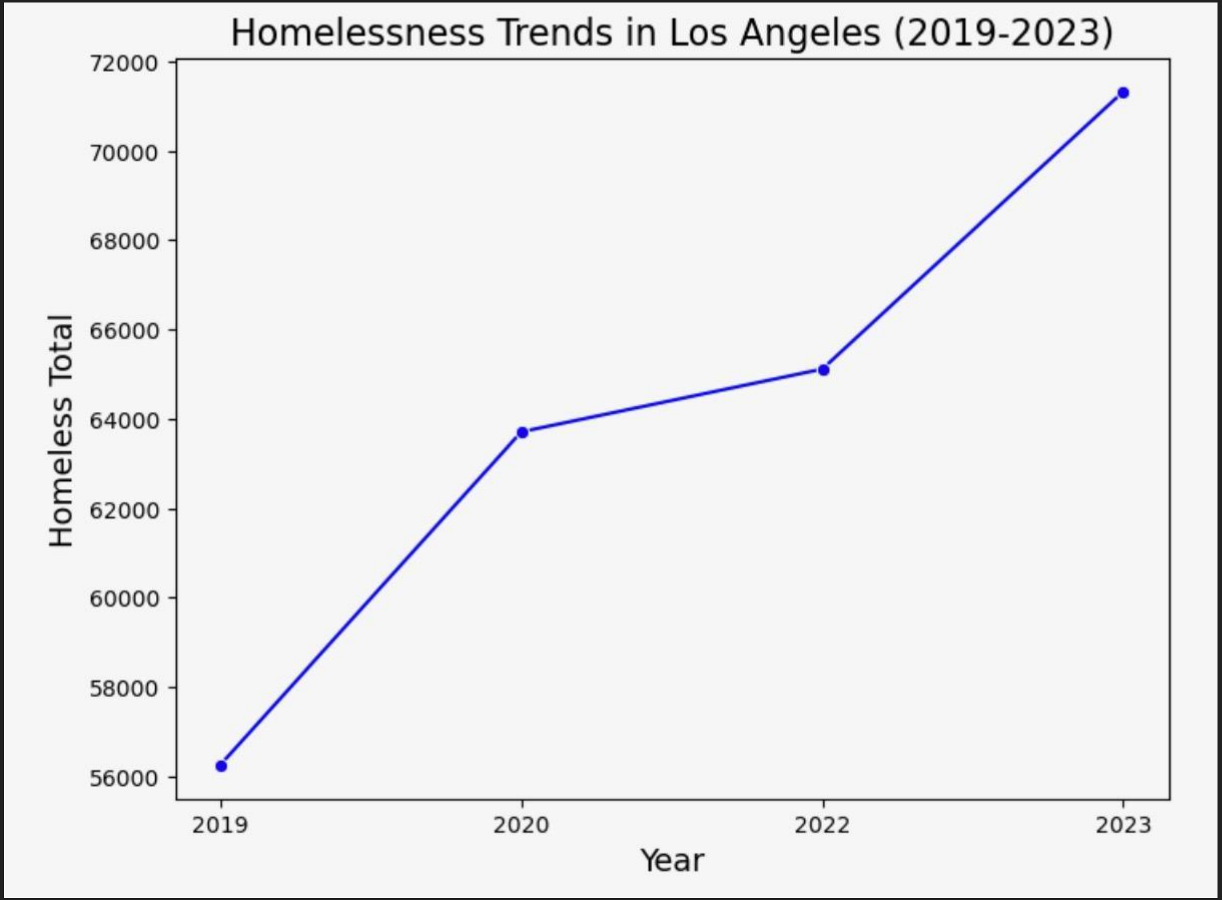
[235]:

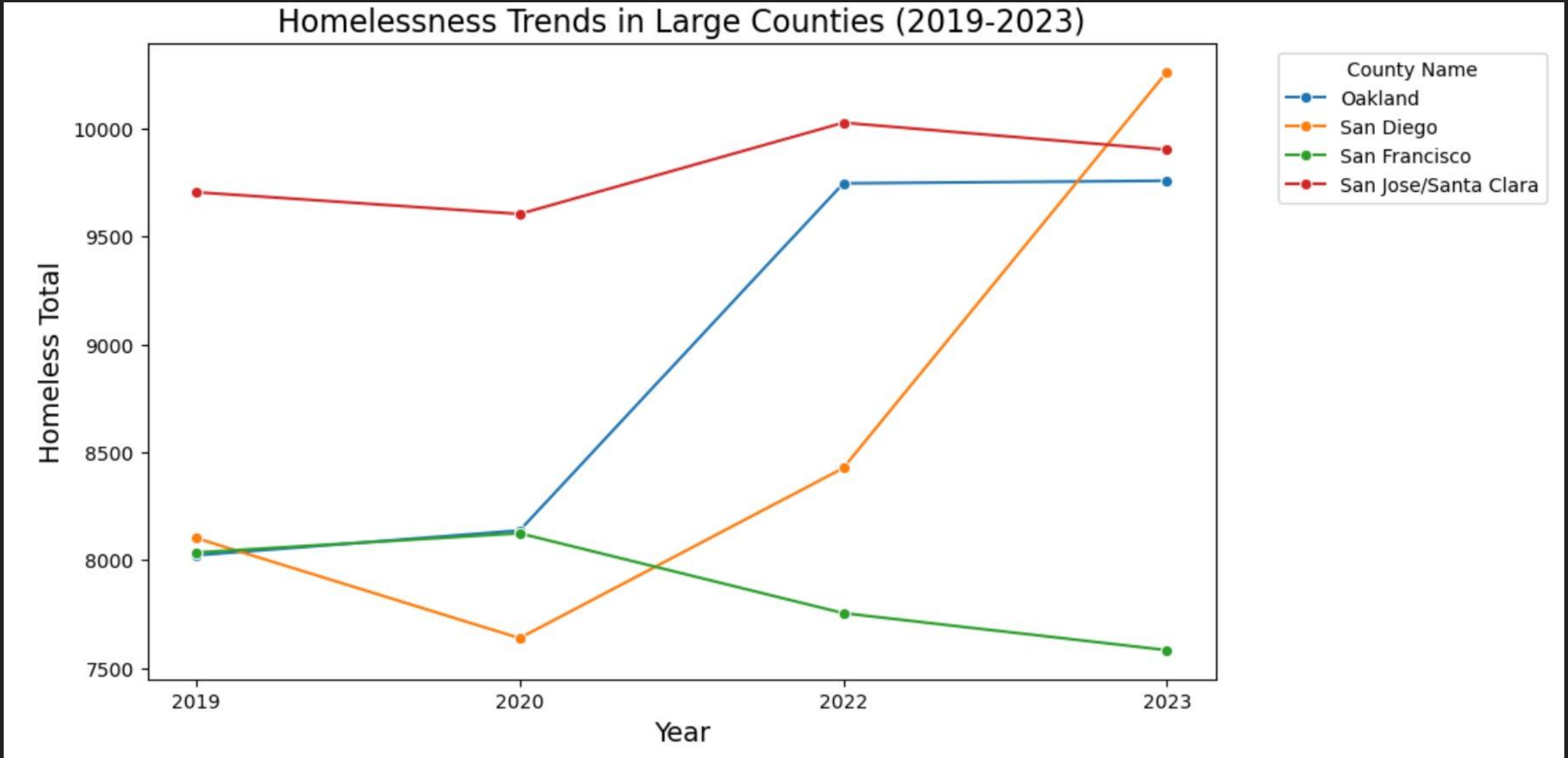
	County Name	Homeless Total_2019	Homeless Total_2020	Homeless Total_2022	Homeless Total_2023	Homeless Total_Sum
3	California	151278	161548	171521	181399	665746
15	Los Angeles	56257	63706	65111	71320	256394
33	San Jose/Santa Clara	9706	9605	10028	9903	39242
21	Oakland	8022	8137	9747	9759	35665
31	San Diego	8102	7638	8427	10264	34431
32	San Francisco	8035	8124	7754	7582	31495

Graph 1: California Homeless Population Over Time

```
# Filter data for only California
california_data = merged_df[merged_df['County Name'] == 'California']

# Reshape California data for plotting (use of Github Co-pilot for this code)
melted_california_df = california_data.melt(
    id_vars=['County Name'],
    value_vars=['Homeless Total_2019', 'Homeless Total_2020', 'Homeless Total_2022', 'Homeless Total_2023'],
    var_name='Year',
    value_name='Homeless Total'
)

# Clean the 'Year' column (use of Github Co-pilot for this code)
melted_california_df['Year'] = melted_california_df['Year'].str.extract(r'(\d{4})')

# Plot for California (use of Github Co-pilot for this code)
plt.figure(figsize=(8, 6))
sns.lineplot(
    data=melted_california_df,
    x='Year',
    y='Homeless Total',
    marker='o',
    color='red'
)

# Customizing the plot
plt.title('Homelessness Trends in California (2019-2023)', fontsize=16)
plt.xlabel('Year', fontsize=14)
plt.ylabel('Homeless Total', fontsize=14)

plt.show()
```

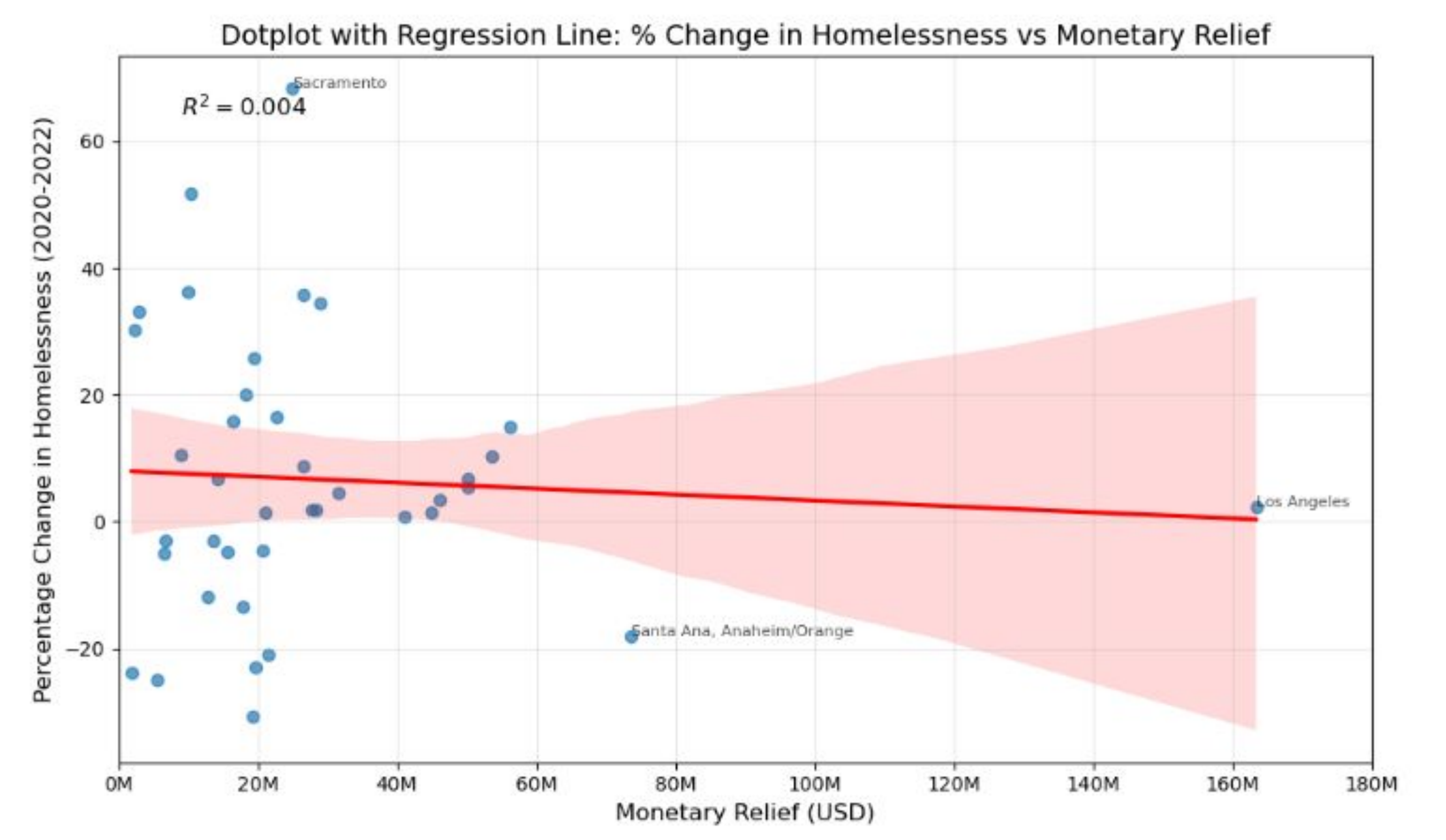

Homelessness Trends in California (2019-2023)
Homeless Total
180000
175000
170000
165000
160000
155000
150000
2019
2020
2022
2023
Year


Homelessness Trends in Los Angeles (2019-2023)
72000
70000
68000
66000
64000
62000
60000
58000
56000
Homeless Total
2019
2020
2022
2023
Year


Homelessness Trends in Large Counties (2019-2023)
Homeless Total
10000
9500
9000
8500
8000
7500
2019
2020
2022
2023
Year
County Name
Oakland
San Diego
San Francisco
San Jose/Santa Clara

Hypothesis 3:
Counties that received higher COVID-19 relief funding per capita **experienced a smaller rise in homelessness rates** compared to those with lower relief funding.

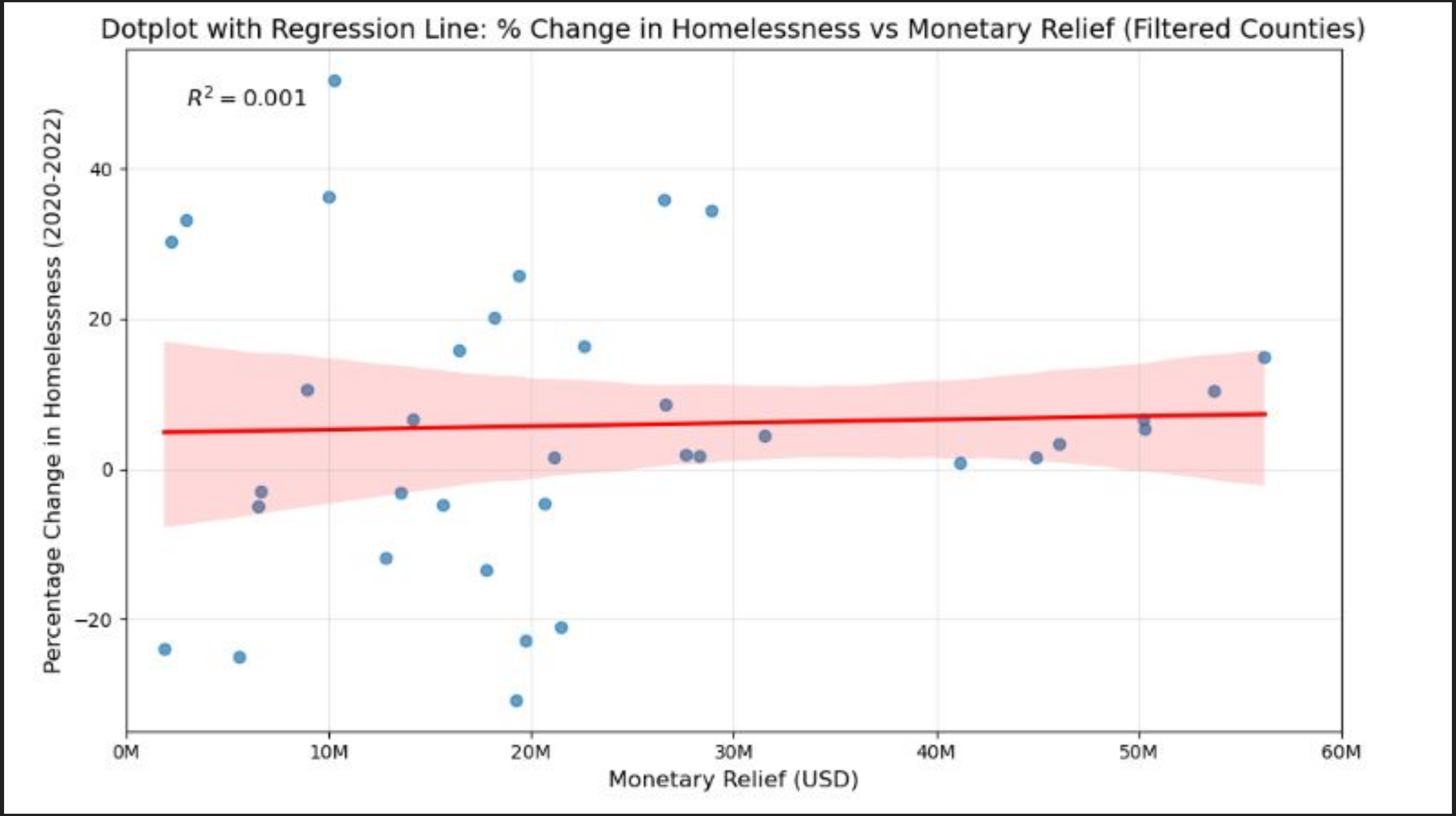
```python
import seaborn as sns
import matplotlib.pyplot as plt
from scipy.stats import linregress

# Calculate the percentage change in homelessness from 2020 to 2022
df['Homeless Change (%)'] = ((df['Homeless Total_2022'] - df['Homeless Total_2020']) /
                              df['Homeless Total_2020']) * 100

# Perform linear regression
slope, intercept, r_value, p_value, std_err = linregress(df['Monetary Relief'], df['Homeless Change (%)'])

# Create the dotplot with a regression line
plt.figure(figsize=(10, 6))
sns.regplot(
    x='Monetary Relief',
    y='Homeless Change (%)',
    data=df,
    scatter_kws={'alpha': 0.7},
    line_kws={'color': 'red'}
)

# Annotate only the specified counties
for county in ["Sacramento", "Los Angeles", "Santa Ana, Anaheim/Orange"]:
    plt.annotate(county,
                 (df.loc[county, 'Monetary Relief'], df.loc[county, 'Homeless Change (%)']),
                 fontsize=8, alpha=0.7)

# Annotate the R-squared value
r_squared = r_value**2
plt.text(
    0.05, 0.95, f'$R^2 = {r_squared:.3f}$',
    fontsize=12, transform=plt.gca().transAxes,
    verticalalignment='top'
)

# Add labels and title
plt.xlabel('Monetary Relief (USD)', fontsize=12)
plt.ylabel('Percentage Change in Homelessness (2020-2022)', fontsize=12)
plt.title('Dotplot with Regression Line: % Change in Homelessness vs Monetary Relief', fontsize=14)

# Adjust x-axis to display values in millions
plt.xticks(ticks=plt.gca().get_xticks(), labels=[f'{int(x / 1_000_000)}M' for x in plt.gca().get_xticks()])
plt.xlim(left=0)

plt.grid(alpha=0.3)
plt.tight_layout()
plt.show()
```



Dotplot with Regression Line: % Change in Homelessness vs Monetary Relief
$R^2 = 0.004$
Sacramento
Los Angeles
Santa Ana, Anaheim/Orange
Percentage Change in Homelessness (2020-2022)
60
40
20
0
−20
0M
20M
40M
60M
80M
100M
120M
140M
160M
180M
Monetary Relief (USD)

[35]:

```python
import seaborn as sns
import matplotlib.pyplot as plt
from scipy.stats import linregress

# Drop the specified counties
df_filtered = df.drop(index=["Sacramento", "Los Angeles", "Santa Ana, Anaheim/Orange"])

# Calculate the percentage change in homelessness from 2020 to 2022
df_filtered['Homeless Change (%)'] = ((df_filtered['Homeless Total_2022'] - df_filtered['Homeless Total_2020']) /
                                      df_filtered['Homeless Total_2020']) * 100

# Perform linear regression
slope, intercept, r_value, p_value, std_err = linregress(df_filtered['Monetary Relief'], df_filtered['Homeless Change (%)'])

# Create the dotplot with a regression line
plt.figure(figsize=(10, 6))
sns.regplot(
    x='Monetary Relief',
    y='Homeless Change (%)',
    data=df_filtered,
    scatter_kws={'alpha': 0.7},
    line_kws={'color': 'red'}
)

# Annotate the R-squared value
r_squared = r_value**2
plt.text(
    0.05, 0.95, f'$R^2 = {r_squared:.3f}$',
    fontsize=12, transform=plt.gca().transAxes,
    verticalalignment='top'
)

# Add labels and title
plt.xlabel('Monetary Relief (USD)', fontsize=12)
plt.ylabel('Percentage Change in Homelessness (2020-2022)', fontsize=12)
plt.title('Dotplot with Regression Line: % Change in Homelessness vs Monetary Relief (Filtered Counties)', fontsize=14)

# Adjust x-axis to display values in millions
plt.xticks(ticks=plt.gca().get_xticks(), labels=[f'{int(x / 1_000_000)}M' for x in plt.gca().get_xticks()])
plt.xlim(left=0)

plt.grid(alpha=0.3)
plt.tight_layout()
plt.show()
```



Dotplot with Regression Line: % Change in Homelessness vs Monetary Relief (Filtered Counties)
$R^2 = 0.001$
Percentage Change in Homelessness (2020-2022)
40
20
0
−20
0M
10M
20M
30M
40M
50M
60M
Monetary Relief (USD)

Conclusion

Theory: Inflation