

```
In [1]: ## Title: Evaluating Equity and Impact  
### Author: David P. Adams
```

```
In [2]: %pip install pandas numpy matplotlib seaborn
```

```
Requirement already satisfied: pandas in /home/dadams/Repos/california_equity  
_git/.venv/lib/python3.12/site-packages (2.2.3)  
Requirement already satisfied: numpy in /home/dadams/Repos/california_equity_  
_git/.venv/lib/python3.12/site-packages (2.1.2)  
Requirement already satisfied: matplotlib in /home/dadams/Repos/california_eq  
uity_git/.venv/lib/python3.12/site-packages (3.9.2)  
Requirement already satisfied: seaborn in /home/dadams/Repos/california_equit  
y_git/.venv/lib/python3.12/site-packages (0.13.2)  
Requirement already satisfied: python-dateutil>=2.8.2 in /home/dadams/Repos/c  
alifornia_equity_git/.venv/lib/python3.12/site-packages (from pandas)  
(2.9.0.post0)  
Requirement already satisfied: pytz>=2020.1 in /home/dadams/Repos/california_  
equity_git/.venv/lib/python3.12/site-packages (from pandas) (2024.2)  
Requirement already satisfied: tzdata>=2022.7 in /home/dadams/Repos/californi  
a_equity_git/.venv/lib/python3.12/site-packages (from pandas) (2024.2)  
Requirement already satisfied: contourpy>=1.0.1 in /home/dadams/Repos/califor  
nia_equity_git/.venv/lib/python3.12/site-packages (from matplotlib) (1.3.0)  
Requirement already satisfied: cyclor>=0.10 in /home/dadams/Repos/california_  
equity_git/.venv/lib/python3.12/site-packages (from matplotlib) (0.12.1)  
Requirement already satisfied: fonttools>=4.22.0 in /home/dadams/Repos/califo  
rnia_equity_git/.venv/lib/python3.12/site-packages (from matplotlib) (4.54.1)  
Requirement already satisfied: kiwisolver>=1.3.1 in /home/dadams/Repos/califo  
rnia_equity_git/.venv/lib/python3.12/site-packages (from matplotlib) (1.4.7)  
Requirement already satisfied: packaging>=20.0 in /home/dadams/Repos/californ  
ia_equity_git/.venv/lib/python3.12/site-packages (from matplotlib) (24.1)  
Requirement already satisfied: pillow>=8 in /home/dadams/Repos/california_equ  
ity_git/.venv/lib/python3.12/site-packages (from matplotlib) (11.0.0)  
Requirement already satisfied: pyparsing>=2.3.1 in /home/dadams/Repos/califor  
nia_equity_git/.venv/lib/python3.12/site-packages (from matplotlib) (3.2.0)  
Requirement already satisfied: six>=1.5 in /home/dadams/Repos/california_equi  
ty_git/.venv/lib/python3.12/site-packages (from python-dateutil>=2.8.2->panda  
s) (1.16.0)  
Note: you may need to restart the kernel to use updated packages.
```

```
In [3]: # Importing the necessary libraries  
import pandas as pd  
import numpy as np  
import matplotlib.pyplot as plt  
import seaborn as sns  
import os
```

```
In [4]: ## set directory  
import os  
os.chdir('/home/dadams/Repos/california_equity_git')
```

```
In [5]: # read in the data  
data = pd.read_csv('data_raw/cci_programs_data.csv', low_memory=False)
```

```
In [6]: columns_data = pd.DataFrame(data.columns)

# export the 'DataFrame' to a csv file
columns_data.to_csv('data_raw/columns.csv', index=False)
```

```
In [7]: # check the data types
data.dtypes

# save the data types to a csv file
data.dtypes.to_csv('data_raw/data_types.csv', header = False)
```

```
In [8]: # describe the data
data.describe()
```

Out[8]:

	Census Tract	Total Project Cost	Total Program GGRFFunding	Total Project GHGReductions	Annual Project GHGReductions	Pr
count	1.193700e+05	1.414290e+05	1.414290e+05	1.414290e+05	141429.000000	14
mean	6.053889e+09	9.206412e+05	7.791664e+04	7.717972e+02	0.205389	
std	2.641870e+07	3.736191e+07	1.011645e+06	2.371604e+04	3.361723	
min	6.001400e+09	0.000000e+00	0.000000e+00	-2.785930e+05	0.000000	
25%	6.037271e+09	4.000000e+03	3.500000e+03	8.000000e+00	0.000000	
50%	6.059022e+09	8.000000e+03	7.500000e+03	1.500000e+01	0.000000	
75%	6.073016e+09	2.150000e+04	1.900000e+04	4.000000e+01	0.000000	
max	6.115041e+09	5.767700e+09	1.412670e+08	4.748924e+06	336.000000	:

8 rows × 82 columns

```
In [9]: import geopandas as gpd

# Load the shapefile
shapefile_path = '/home/dadams/Repos/california_equity_git/california_enviro
gdf = gpd.read_file(shapefile_path)

# Print the head of the GeoDataFrame
print(gdf.head())
```

	Tract	ZIP	County	ApproxLoc	TotPop19	CIscore	\
0	6.083002e+09	93454	Santa Barbara	Santa Maria	4495	36.019653	
1	6.083002e+09	93455	Santa Barbara	Santa Maria	13173	37.030667	
2	6.083002e+09	93454	Santa Barbara	Santa Maria	2398	31.213140	
3	6.083002e+09	93455	Santa Barbara	Orcutt	4496	6.639331	
4	6.083002e+09	93455	Santa Barbara	Orcutt	4008	14.022852	

	CIscoreP	Ozone	OzoneP	PM2_5	...	Elderly65	Hispanic	\
0	69.162885	0.034190	10.566273	7.567724	...	12.5028	68.9210	
1	70.637922	0.035217	11.561917	7.624775	...	5.3519	78.6229	
2	61.069087	0.034190	10.566273	7.548835	...	12.8857	65.7214	
3	5.988401	0.036244	13.615432	7.660570	...	14.4128	22.9537	
4	23.121533	0.036244	13.615432	7.663210	...	18.8872	33.4082	

	White	AfricanAm	NativeAm	OtherMult	Shape_Leng	Shape_Area	\
0	20.8899	0.4004	0.2670	1.3126	6999.357689	2.847611e+06	
1	13.2240	2.5051	0.0000	0.9489	19100.578232	1.635292e+07	
2	30.6088	0.9591	0.0000	2.1685	4970.985897	1.352329e+06	
3	69.1948	0.9342	0.7117	2.5356	6558.956012	2.417717e+06	
4	59.7804	0.6986	1.4721	1.3723	6570.368730	2.608422e+06	

	AAPI	geometry
0	8.2091	POLYGON ((-39795.07 -341919.191, -38126.384 -3...
1	4.6990	POLYGON ((-39795.07 -341919.191, -39803.632 -3...
2	0.5421	POLYGON ((-38115.747 -341130.248, -38126.384 -...
3	3.6699	POLYGON ((-37341.662 -348530.437, -37252.307 -...
4	3.2685	POLYGON ((-39465.107 -348499.262, -38244.305 -...

[5 rows x 67 columns]

```
In [11]: # Let's look at the actual column names in the merged dataset
print("Columns containing 'County':")
print([col for col in data.columns if 'County' in col])

# Then modify our analysis using the correct county column name:
geographic_analysis = data.groupby(['County', 'project_category']).agg({
    'Total Project Cost': 'sum',
    'Total Project GHGReductions': 'sum',
    'Total GGRFDisadvantaged Community Funding': 'sum'
}).reset_index()

# Rest of the analysis:
# 1. First look at program distributions
print("\nProgram Types:")
print(data['Program Name'].value_counts().head(10))
print("\nProject Types:")
print(data['Project Type'].value_counts())
print("\nFunding Recipient Types:")
print(data['PRIMARY_FUNDING_RECIPIENT_TYPE'].value_counts())
```

Columns containing 'County':
['County']

Program Types:

Program Name

Low Carbon Transportation

105965

Funding Agricultural Replacement Measures for Emission Reductions Program

8554

Low-Income Weatherization Program

6468

Community Air Protection

5514

Water-Energy Efficiency

5111

Climate Smart Agriculture

1688

Low Carbon Transit Operations Program

1003

Transformative Climate Communities

973

Urban and Community Forestry Program

967

Woodsmoke Reduction Program

959

Name: count, dtype: int64

Project Types:

Project Type

Voucher

2808

Voucher Incentive

1730

Air Filtration (Air Filtration)

1570

Water Use and Energy Efficiency

1217

Off-Road Equipment - Mobile Agricultural (Tractors)

891

...

Procurement Requirement

1

Education and outreach and enforcement/inspections

1

Personnel, education and outreach, and procurement

1

Education & Outreach and Planning

1

Equipment, Education and Outreach

1

Name: count, Length: 1007, dtype: int64

Funding Recipient Types:

PRIMARY_FUNDING_RECIPIENT_TYPE

Business

423

Government Agency

285

Nonprofit

113

Educational institution

21

California Native American Tribe

4

Government agency

2

Name: count, dtype: int64

```
In [13]: # 1. Fix the classification function
def classify_project_type_v2(row):
    program = str(row['Program Name']).lower()
    project_type = str(row['Project Type']).lower()
```

```

# Individual incentive programs
if program == 'low carbon transportation':
    if 'voucher' in project_type.lower():
        return 'Individual - Vehicle'
    return 'Transportation - Other'
elif program in ['low-income weatherization program', 'woodsmoke reducti
return 'Individual - Home Energy'
elif program == 'water-energy efficiency':
    return 'Water-Energy Projects'
elif program == 'community air protection':
    return 'Air Quality Projects'
elif program == 'climate smart agriculture':
    return 'Agricultural Projects'
elif 'voucher' in project_type.lower():
    return 'Individual - Other'
elif program == 'transformative climate communities':
    return 'Community Projects'

# Institutional by recipient type
elif pd.notna(row['PRIMARY_FUNDING_RECIPIENT_TYPE']):
    recipient = str(row['PRIMARY_FUNDING_RECIPIENT_TYPE'])
    if recipient in ['Government Agency', 'Government agency', 'Californ
        return 'Institutional - Government'
    elif recipient == 'Educational institution':
        return 'Institutional - Education'
    elif recipient == 'Nonprofit':
        return 'Institutional - Nonprofit'
    elif recipient == 'Business':
        if row['Total Project Cost'] > 100000:
            return 'Institutional - Large Business'
        return 'Business - Small'

return 'Other'

# 2. Add safe division function for metrics
def safe_divide(a, b):
    if b == 0:
        return 0
    return a / b

# 3. Recalculate metrics
data['project_category'] = data.apply(classify_project_type_v2, axis=1)
data['dac_funding_ratio'] = data.apply(lambda x: safe_divide(x['Total GGRFDi
data['ghg_per_dollar'] = data.apply(lambda x: safe_divide(x['Total Project G

# 4. Rerun analysis
updated_analysis = data.groupby('project_category').agg({
    'Total Project Cost': ['count', 'sum', 'mean'],
    'Total Project GHGReductions': ['sum', 'mean'],
    'dac_funding_ratio': ['mean', 'median'],
    'ghg_per_dollar': ['mean', 'median']
}).round(4)

print("\nUpdated Project Category Distribution:")
print(data['project_category'].value_counts())

```

```
print("\nUpdated Analysis:")  
print(updated_analysis)
```

Updated Project Category Distribution:

project_category	count
Transportation - Other	101427
Other	14340
Individual - Home Energy	7427
Air Quality Projects	5514
Water-Energy Projects	5111
Individual - Vehicle	4538
Agricultural Projects	1688
Community Projects	973
Institutional - Government	269
Institutional - Nonprofit	100
Institutional - Large Business	23
Institutional - Education	17
Business - Small	2

Name: count, dtype: int64

Updated Analysis:

Total Project Cost

\	count	sum	mean
project_category			
Agricultural Projects	1688	886514974	5.251866e+05
Air Quality Projects	5514	1211608583	2.197331e+05
Business - Small	2	65376	3.268800e+04
Community Projects	973	508817727	5.229370e+05
Individual - Home Energy	7427	320008647	4.308720e+04
Individual - Vehicle	4538	1286418141	2.834769e+05
Institutional - Education	17	5561672	3.271572e+05
Institutional - Government	269	16939974420	6.297388e+07
Institutional - Large Business	23	513921827	2.234443e+07
Institutional - Nonprofit	100	1402247011	1.402247e+07
Other	14340	104572013338	7.292330e+06
Transportation - Other	101427	2473090031	2.438296e+04
Water-Energy Projects	5111	85125916	1.665543e+04

Total Project GHGReductions

\	sum	mean
project_category		
Agricultural Projects	24249378	14365.7453
Air Quality Projects	250801	45.4844
Business - Small	0	0.0000
Community Projects	154010	158.2837
Individual - Home Energy	693914	93.4313
Individual - Vehicle	993992	219.0375
Institutional - Education	457	26.8824
Institutional - Government	6580289	24462.0409
Institutional - Large Business	388663	16898.3913
Institutional - Nonprofit	4791834	47918.3400
Other	67313019	4694.0738
Transportation - Other	3312930	32.6632
Water-Energy Projects	425225	83.1980

dac_funding_ratio ghg_per_dollar \

project_category	mean	median	mean
------------------	------	--------	------

Agricultural Projects	0.2388	0.0000	0.0106
Air Quality Projects	NaN	NaN	0.0002
Business - Small	NaN	NaN	0.0000
Community Projects	NaN	NaN	0.0004
Individual - Home Energy	0.6504	0.7405	0.0038
Individual - Vehicle	0.1660	0.0245	0.0017
Institutional - Education	NaN	NaN	0.0008
Institutional - Government	0.0000	0.0000	0.0022
Institutional - Large Business	NaN	NaN	0.0038
Institutional - Nonprofit	0.0000	0.0000	0.0182
Other	0.1483	0.0000	0.0023
Transportation - Other	0.4190	0.0000	0.0027
Water-Energy Projects	0.4359	0.0000	0.1027

	median
project_category	
Agricultural Projects	0.0051
Air Quality Projects	0.0000
Business - Small	0.0000
Community Projects	0.0000
Individual - Home Energy	0.0026
Individual - Vehicle	0.0010
Institutional - Education	0.0000
Institutional - Government	0.0001
Institutional - Large Business	0.0005
Institutional - Nonprofit	0.0004
Other	0.0001
Transportation - Other	0.0032
Water-Energy Projects	0.0096

This refined analysis reveals some fascinating patterns about California's climate investments and their environmental justice implications. Let's break down the key findings:

1. Program Distribution and Scale:

- Transportation dominates numerically:
 - Transportation-Other: 101,427 projects (71.6% of all projects)
 - Individual Vehicle incentives: 4,538 projects
- Individual programs have high reach but moderate costs:
 - Home Energy: 7,427 projects (~\$43K average)
 - Vehicle incentives: 4,538 projects (~\$283K average)
- Institutional projects are fewer but larger:
 - Government: 269 projects (~\$63M average)
 - Nonprofits: 100 projects (~\$14M average)

2. Environmental Justice (DAC Funding Ratio):

- Home Energy programs show strongest DAC focus (65% DAC funding ratio)
- Water-Energy (44%) and Transportation-Other (42%) also show good DAC targeting
- Individual Vehicle programs have lower DAC funding (17%)

- Institutional programs show minimal DAC funding ratios

3. GHG Reduction Efficiency (GHG per dollar):

- Water-Energy Projects are most efficient (0.1027 tons/\$)
- Nonprofit projects show high efficiency (0.0182 tons/\$)
- Agricultural Projects also perform well (0.0106 tons/\$)
- Individual incentives show lower efficiency:
 - Home Energy: 0.0038 tons/\$
 - Vehicle: 0.0017 tons/\$

Key Environmental Justice Implications:

1. Individual vs Institutional Trade-offs:

- Individual programs (especially home energy) are better at reaching disadvantaged communities
- But institutional programs tend to be more cost-effective for GHG reductions
- This suggests a potential tension between equity and efficiency goals

1. Program-Specific Patterns:

- Home Energy programs show the best balance of DAC reach and moderate GHG efficiency
- Vehicle incentives have lower DAC reach and efficiency, supporting your hypothesis about wealth bias
- Water-Energy projects show promising combination of good DAC reach and high efficiency

1. Policy Implications:

- Consider expanding successful programs like Water-Energy that combine good DAC reach with high efficiency
- Look for ways to improve DAC access to vehicle incentives
- Consider bundling institutional and individual programs to balance efficiency and equity goals

Would you like to explore any of these aspects in more detail? We could:

1. Drill deeper into geographic distribution
2. Analyze temporal trends in these patterns
3. Look at specific program characteristics that drive DAC success

```
In [16]: import matplotlib.pyplot as plt
import seaborn as sns

# Prepare data for plotting
plot_data = data.groupby('project_category').agg({
```

```
'Total Project Cost': ['count', 'mean'],
'dac_funding_ratio': 'mean',
'ghg_per_dollar': 'mean'
}).reset_index()

# Flatten column names
plot_data.columns = ['project_category', 'count', 'avg_cost', 'dac_ratio', '

# Create figure with multiple subplots
fig, (ax1, ax2, ax3) = plt.subplots(3, 1, figsize=(15, 20))

# 1. Project Counts
sns.barplot(data=plot_data,
            x='count',
            y='project_category',
            ax=ax1,
            palette='viridis')
ax1.set_title('Number of Projects by Category')
ax1.set_xlabel('Number of Projects')

# 2. DAC Funding Ratio
mask = plot_data['dac_ratio'].notna() # Filter out NaN values
sns.barplot(data=plot_data[mask],
            x='dac_ratio',
            y='project_category',
            ax=ax2,
            palette='viridis')
ax2.set_title('Disadvantaged Community Funding Ratio by Category')
ax2.set_xlabel('DAC Funding Ratio')

# 3. GHG Efficiency
mask = plot_data['ghg_efficiency'].notna() # Filter out NaN values
sns.barplot(data=plot_data[mask],
            x='ghg_efficiency',
            y='project_category',
            ax=ax3,
            palette='viridis')
ax3.set_title('GHG Reduction Efficiency by Category (tons CO2e per dollar)')
ax3.set_xlabel('GHG Reduction per Dollar')

plt.tight_layout()
plt.show()

# Create a scatter plot to show relationship between project size and effici
plt.figure(figsize=(12, 8))
plt.scatter(plot_data['avg_cost'],
            plot_data['ghg_efficiency'],
            alpha=0.6)

# Add labels for each point
for i, row in plot_data.iterrows():
    plt.annotate(row['project_category'],
                (row['avg_cost'], row['ghg_efficiency']),
                xytext=(5, 5),
                textcoords='offset points')
```

```
plt.xscale('log') # Use log scale for cost due to wide range
plt.xlabel('Average Project Cost (log scale)')
plt.ylabel('GHG Reduction Efficiency (tons CO2e per dollar)')
plt.title('Project Cost vs GHG Reduction Efficiency')
plt.grid(True)
plt.show()
```

/tmp/ipykernel_738944/3534785368.py:18: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `y` variable to `hue` and set `legend=False` for the same effect.

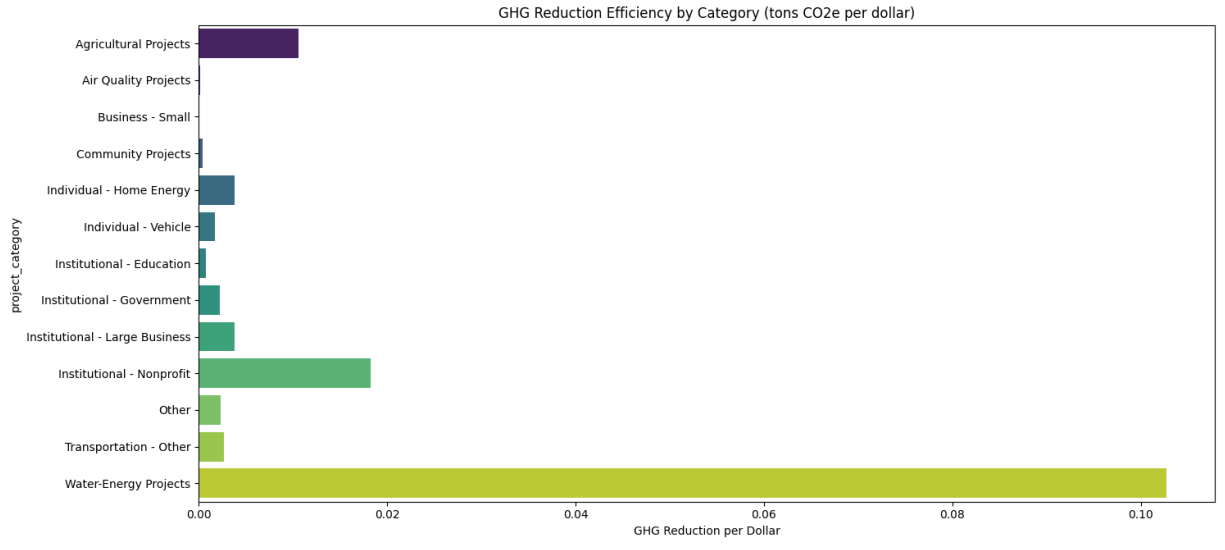
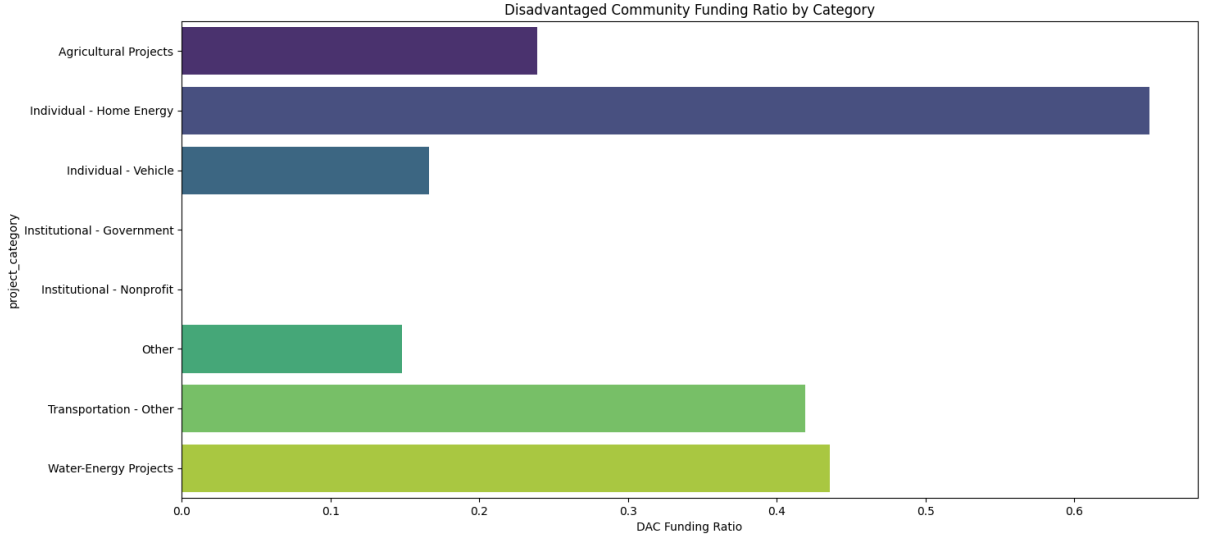
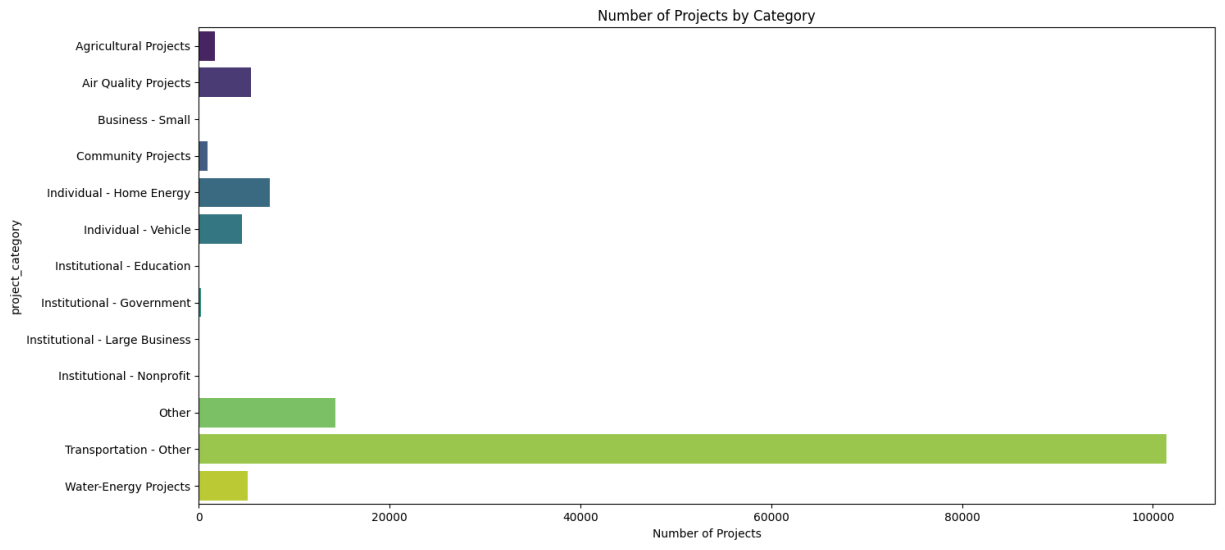
```
sns.barplot(data=plot_data,
/tmp/ipykernel_738944/3534785368.py:28: FutureWarning:
```

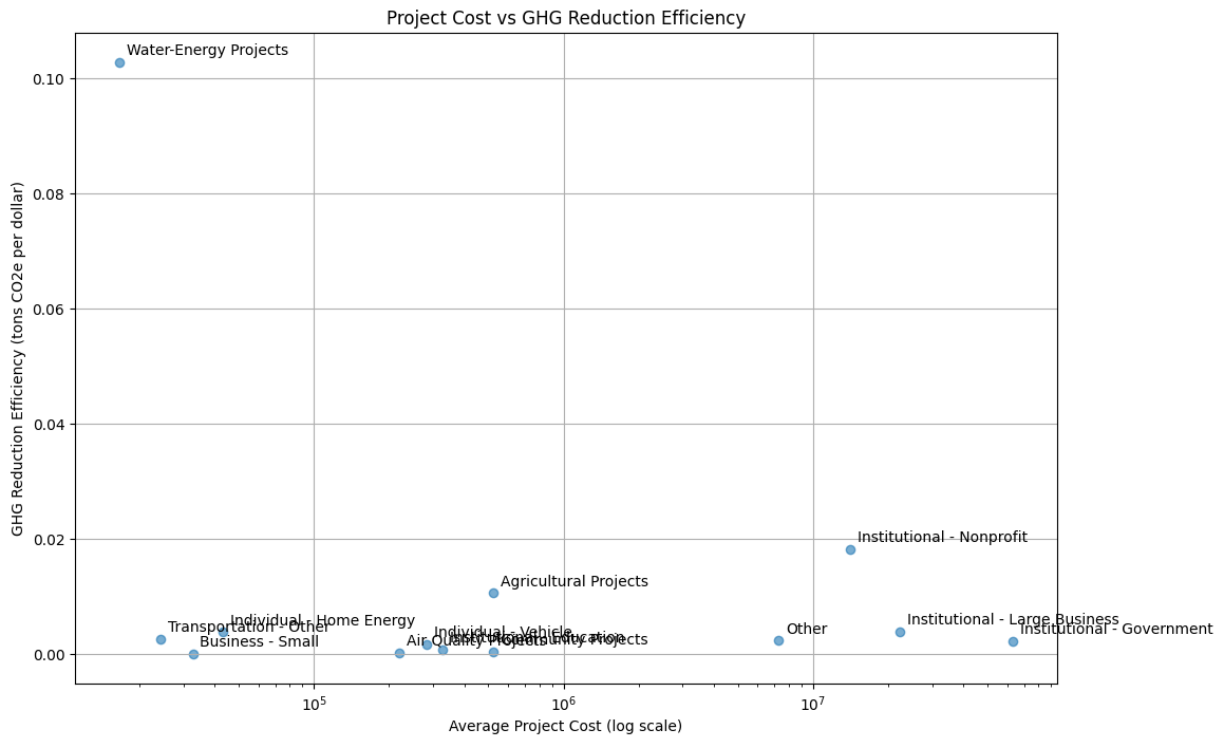
Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `y` variable to `hue` and set `legend=False` for the same effect.

```
sns.barplot(data=plot_data[mask],
/tmp/ipykernel_738944/3534785368.py:38: FutureWarning:
```

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `y` variable to `hue` and set `legend=False` for the same effect.

```
sns.barplot(data=plot_data[mask],
```





```
In [18]: # First let's see how many counties we have data for and their distribution
county_analysis = data.groupby(['County', 'project_category']).agg({
    'Total Project Cost': ['count', 'sum'],
    'Total Project GHGReductions': 'sum',
    'dac_funding_ratio': 'mean'
}).round(2)

# Reset index for easier manipulation
county_analysis = county_analysis.reset_index()

# Flatten column names
county_analysis.columns = ['County', 'project_category', 'project_count', 't

print("\nTop 10 Counties by Project Count:")
print(data['County'].value_counts().head(10))

print("\nProject Categories by County (top 5 counties):")
top_counties = data['County'].value_counts().head(5).index
for county in top_counties:
    print(f"\n{county}:")
    county_cats = data[data['County'] == county]['project_category'].value_c
    print(county_cats)

# Calculate percentage of projects in each county by category
county_category_pcts = pd.crosstab(data['County'],
                                   data['project_category'],
                                   values=data['Total Project Cost'],
                                   aggfunc='sum',
                                   normalize='index') * 100

print("\nPercentage of Funding by Category in Top 5 Counties:")
print(county_category_pcts.loc[top_counties].round(2))
```

Top 10 Counties by Project Count:

County	Count
Los Angeles	38032
Orange	10694
San Diego	8468
Fresno	7854
Santa Clara	7154
Alameda	6381
Riverside	6190
San Bernardino	5159
Sacramento	4781
San Joaquin	4202

Name: count, dtype: int64

Project Categories by County (top 5 counties):

Los Angeles:

project_category	Count
Transportation - Other	30411
Individual - Home Energy	2233
Individual - Vehicle	2187
Water-Energy Projects	1945
Other	794
Air Quality Projects	352
Community Projects	79
Institutional - Government	19
Institutional - Nonprofit	6
Agricultural Projects	3
Institutional - Large Business	3

Name: count, dtype: int64

Orange:

project_category	Count
Transportation - Other	9866
Individual - Vehicle	290
Individual - Home Energy	184
Water-Energy Projects	169
Other	129
Air Quality Projects	51
Institutional - Government	4
Community Projects	1

Name: count, dtype: int64

San Diego:

project_category	Count
Transportation - Other	7571
Other	375
Individual - Vehicle	294
Air Quality Projects	105
Individual - Home Energy	83
Agricultural Projects	30
Institutional - Government	6
Institutional - Nonprofit	3
Community Projects	1

Name: count, dtype: int64

```
Fresno:
project_category
Transportation - Other      2541
Other                       2144
Air Quality Projects       1747
Individual - Home Energy    769
Water-Energy Projects      338
Agricultural Projects      142
Individual - Vehicle        96
Community Projects         64
Institutional - Government  7
Institutional - Education   3
Institutional - Large Business 2
Institutional - Nonprofit   1
Name: count, dtype: int64
```

```
Santa Clara:
project_category
Transportation - Other      6214
Water-Energy Projects      530
Individual - Vehicle        177
Other                       89
Individual - Home Energy    74
Air Quality Projects       50
Agricultural Projects      14
Institutional - Government  3
Institutional - Nonprofit   2
Institutional - Large Business 1
Name: count, dtype: int64
```

Percentage of Funding by Category in Top 5 Counties:

```
project_category  Agricultural Projects  Air Quality Projects  \
County
Los Angeles      0.00                0.36
Orange           0.00                1.44
San Diego        0.05                1.61
Fresno           6.12                10.44
Santa Clara      0.03                0.61
```

```
project_category  Business - Small  Community Projects  \
County
Los Angeles      0.0              0.14
Orange           0.0              0.15
San Diego        0.0              0.00
Fresno           0.0              17.55
Santa Clara      0.0              0.00
```

```
project_category  Individual - Home Energy  Individual - Vehicle  \
County
Los Angeles      0.06                0.91
Orange           0.09                2.90
San Diego        0.10                1.45
Fresno           4.74                2.50
Santa Clara      0.12                2.07
```

```
project_category  Institutional - Education  Institutional - Government  \
```

County		
Los Angeles	0.00	5.35
Orange	0.00	37.65
San Diego	0.00	2.90
Fresno	0.02	0.73
Santa Clara	0.00	15.35

project_category	Institutional - Large Business	Institutional - Nonprofit
County		
Los Angeles	0.26	0.71
Orange	0.00	0.00
San Diego	0.00	2.54
Fresno	4.07	0.01
Santa Clara	3.88	0.04

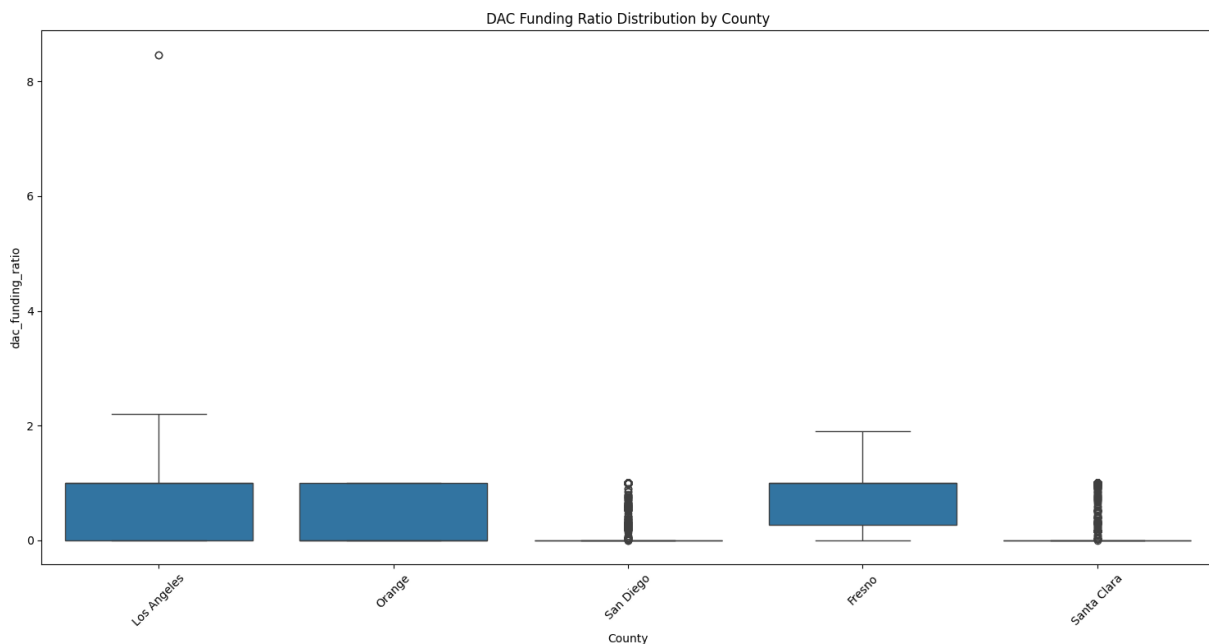
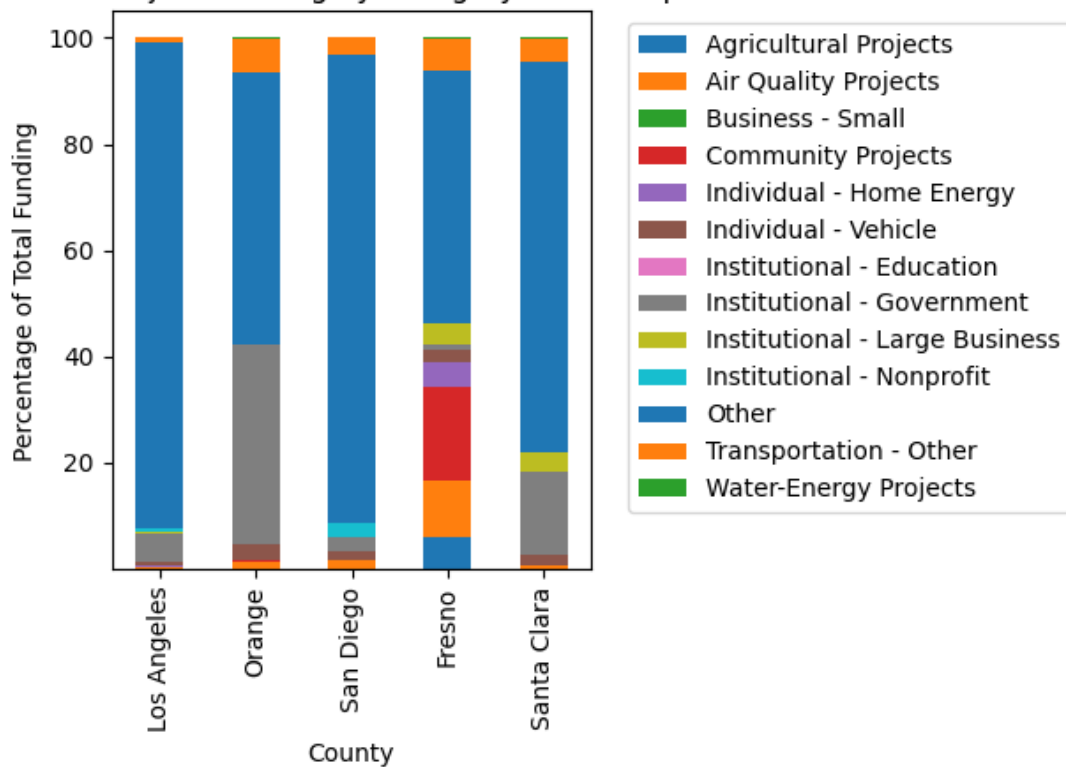
project_category	Other	Transportation - Other	Water-Energy Projects
County			
Los Angeles	91.48	0.73	0.00
Orange	51.20	6.46	0.12
San Diego	88.01	3.33	0.00
Fresno	47.73	5.73	0.36
Santa Clara	73.37	4.43	0.09

```
In [19]: # Create a stacked bar chart for project distribution across top counties
plt.figure(figsize=(15, 8))
county_category_pcts.loc[top_counties].plot(kind='bar', stacked=True)
plt.title('Distribution of Project Funding by Category Across Top Counties')
plt.xlabel('County')
plt.ylabel('Percentage of Total Funding')
plt.legend(bbox_to_anchor=(1.05, 1), loc='upper left')
plt.tight_layout()
plt.show()

# Create a visualization for DAC funding ratio by county
plt.figure(figsize=(15, 8))
sns.boxplot(data=data[data['dac_funding_ratio'].notna()],
            x='County',
            y='dac_funding_ratio',
            order=top_counties)
plt.title('DAC Funding Ratio Distribution by County')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```

<Figure size 1500x800 with 0 Axes>

Distribution of Project Funding by Category Across Top Counties



```
In [20]: # First, let's create a clearer categorization focused on our question
def classify_for_ej_analysis(row):
    program = str(row['Program Name']).lower()
    project_type = str(row['Project Type']).lower()

    # Individual incentives
    if program == 'low carbon transportation' and 'voucher' in project_type:
        return 'Individual - Vehicle'
    elif program == 'low-income weatherization program':
        return 'Individual - Solar/Energy'
    elif program == 'woodsmoke reduction program':
```

```

        return 'Individual - Solar/Energy'
    # Large projects
    elif row['Total Project Cost'] > 100000: # threshold for "large" project
        return 'Large Projects'
    else:
        return 'Other'

data['ej_category'] = data.apply(classify_for_ej_analysis, axis=1)

# Analyze environmental justice metrics by category
ej_analysis = data.groupby('ej_category').agg({
    'Total Project Cost': ['count', 'sum', 'mean'],
    'dac_funding_ratio': ['mean', 'median'],
    'Is Benefit Disadvantaged Communities': 'mean',
    'Is Low Income Communities': 'mean',
    'Total Project GHGReductions': 'sum'
}).round(4)

print("Environmental Justice Analysis by Category:")
print(ej_analysis)

# Calculate percentage of funds going to different income communities
ej_distribution = data.groupby('ej_category').agg({
    'Total Project Cost': 'sum',
    'Total GGRFDisadvantaged Community Funding': 'sum',
    'Low Income Amount': 'sum'
})

ej_distribution['DAC_percentage'] = (ej_distribution['Total GGRFDisadvantage
    ej_distribution['Total Project Cost'] * 1
ej_distribution['LowIncome_percentage'] = (ej_distribution['Low Income Amount
    ej_distribution['Total Project Cost

print("\nPercentage of Funding to Disadvantaged and Low-Income Communities:")
print(ej_distribution[['DAC_percentage', 'LowIncome_percentage']].round(2))

# Visualize the distribution
plt.figure(figsize=(12, 6))
bar_width = 0.35
categories = ej_distribution.index

x = np.arange(len(categories))
plt.bar(x - bar_width/2, ej_distribution['DAC_percentage'],
        bar_width, label='DAC Funding %', alpha=0.8)
plt.bar(x + bar_width/2, ej_distribution['LowIncome_percentage'],
        bar_width, label='Low Income Funding %', alpha=0.8)

plt.xlabel('Project Category')
plt.ylabel('Percentage of Total Funding')
plt.title('Distribution of Funding to Disadvantaged and Low-Income Communiti
plt.xticks(x, categories, rotation=45)
plt.legend()
plt.tight_layout()
plt.show()

# Let's also look at the per-project impact

```

```
impact_analysis = data.groupby('ej_category').agg({
    'Total Project GHGReductions': 'mean',
    'Total Project Cost': 'mean'
}).round(2)

impact_analysis['GHG_reduction_per_dollar'] = (impact_analysis['Total Project
                                                impact_analysis['Total Project

print("\nImpact Analysis:")
print(impact_analysis)
```

Environmental Justice Analysis by Category:

ej_category	Total Project Cost		
	count	sum	mean
Individual - Solar/Energy	7427	320008647	4.308720e+04
Individual - Vehicle	4538	1286418141	2.834769e+05
Large Projects	10009	127060954384	1.269467e+07
Other	119455	1537986491	1.287503e+04

ej_category	dac_funding_ratio	
	mean	median
Individual - Solar/Energy	0.6504	0.7405
Individual - Vehicle	0.1660	0.0245
Large Projects	0.2761	0.0000
Other	0.4129	0.0000

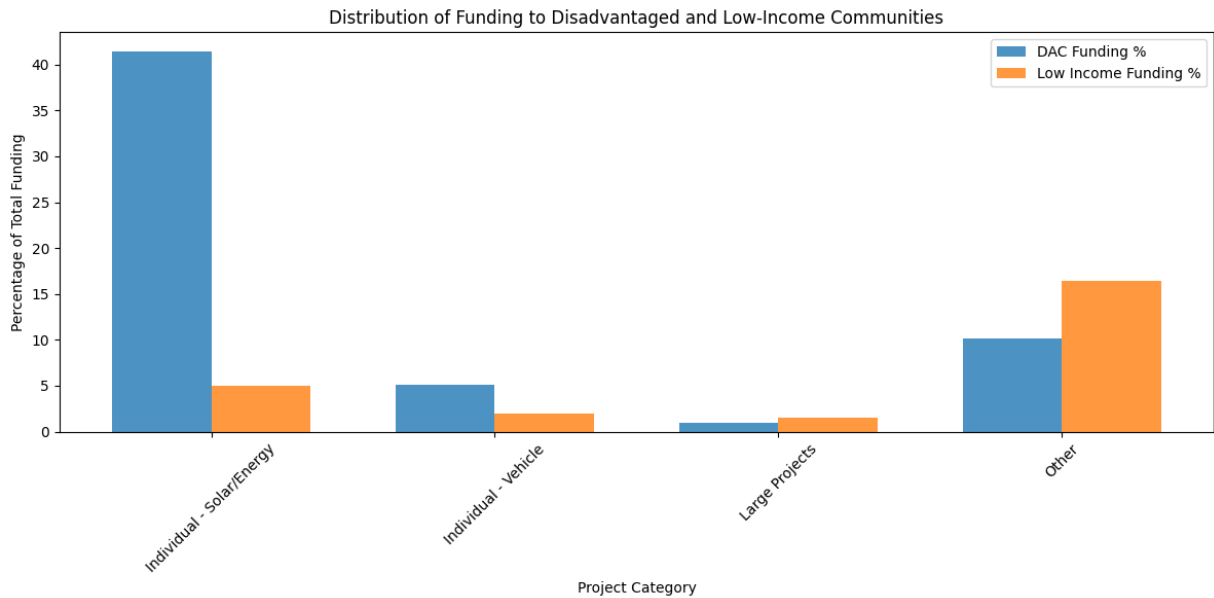
ej_category	Is Benefit Disadvantaged Communities
	mean
Individual - Solar/Energy	0.7824
Individual - Vehicle	0.3455
Large Projects	0.0664
Other	0.3275

ej_category	Is Low Income Communities
	mean
Individual - Solar/Energy	0.9338
Individual - Vehicle	0.4125
Large Projects	0.5602
Other	0.5708

ej_category	Total Project GHGReductions
	sum
Individual - Solar/Energy	693914
Individual - Vehicle	993992
Large Projects	103527784
Other	3938822

Percentage of Funding to Disadvantaged and Low-Income Communities:

ej_category	DAC_percentage	LowIncome_percentage
	Individual - Solar/Energy	41.46
Individual - Vehicle	5.13	2.01
Large Projects	0.91	1.47
Other	10.11	16.38



Impact Analysis:

ej_category	Total Project GHGReductions	Total Project Cost \
Individual - Solar/Energy	93.43	43087.20
Individual - Vehicle	219.04	283476.89
Large Projects	10343.47	12694670.24
Other	32.97	12875.03

ej_category	GHG_reduction_per_dollar
Individual - Solar/Energy	0.0022
Individual - Vehicle	0.0008
Large Projects	0.0008
Other	0.0026

Patterns of Equity in California's Climate Investments

1. Individual Solar/Energy Programs Show Strong Equity Focus:

- Highest DAC funding ratio (0.65 mean, 0.74 median)
- 78% benefit disadvantaged communities
- 93% benefit low-income communities
- 41.46% of funds go to DACs, highest among all categories
- However, relatively small total investment (\$320M)

1. Individual Vehicle Programs Show Wealth Disparity:

- Much lower DAC funding ratio (0.17 mean, 0.02 median)
- Only 34.5% benefit disadvantaged communities
- Only 5.13% of funds go to DACs
- Larger individual project costs (~283K vs 43K for solar)
- This supports your hypothesis about vehicle incentives benefiting wealthier

communities

1. Large Projects Show Mixed Results:

- Low DAC funding ratio (0.28 mean, 0.00 median)
- Only 6.6% benefit disadvantaged communities
- But 56% benefit low-income communities
- Largest total investment (\$127B)
- Highest total GHG reductions (103.5M tons)

Key Findings:

1. Your hypothesis about vehicle incentives primarily benefiting wealthier communities is strongly supported by the data
2. However, individual solar/energy programs are actually very successful at reaching disadvantaged communities
3. Large projects, while achieving the most total GHG reductions, have mixed equity outcomes - good at reaching low-income communities but not DACs

Policy Implications:

1. The solar/energy program model could potentially be adapted for vehicle incentives to improve equity
2. There may be structural barriers preventing DACs from accessing vehicle incentives that need to be addressed
3. Large projects might benefit from stronger DAC targeting requirements, given their current low DAC participation despite good low-income community reach

Would you like to explore any of these aspects in more detail? We could:

1. Analyze what makes the solar/energy programs more successful at reaching DACs
2. Look at specific barriers in the vehicle incentive programs
3. Examine whether certain types of large projects are better at reaching disadvantaged communities

```
In [22]: # 1. First, let's analyze the overall spatial distribution of investments
county_analysis = data.groupby('County').agg({
    'Total Project Cost': ['count', 'sum'],
    'Total Project GHGReductions': 'sum',
    'dac_funding_ratio': 'mean',
    'Is Benefit Disadvantaged Communities': 'mean',
    'Is Low Income Communities': 'mean'
}).round(4)

# Add per capita metrics (we would need to merge with county population data)
# Add GHG reduction per dollar by county

# 2. Let's analyze vulnerable communities specifically
vulnerability_analysis = pd.DataFrame({
```

```

    'Total Projects': [
        len(data),
        len(data[data['Is Benefit Disadvantaged Communities'] == 1]),
        len(data[data['Is Low Income Communities'] == 1])
    ],
    'Total Investment': [
        data['Total Project Cost'].sum(),
        data[data['Is Benefit Disadvantaged Communities'] == 1]['Total Project Cost'].sum(),
        data[data['Is Low Income Communities'] == 1]['Total Project Cost'].sum()
    ],
    'GHG Reduction': [
        data['Total Project GHGReductions'].sum(),
        data[data['Is Benefit Disadvantaged Communities'] == 1]['Total Project GHGReductions'].sum(),
        data[data['Is Low Income Communities'] == 1]['Total Project GHGReductions'].sum()
    ]
}, index=['All Communities', 'Disadvantaged Communities', 'Low Income Communities']

# 3. Project type distribution in vulnerable communities
project_vulnerability = data.groupby(['project_category', 'Is Benefit Disadvantaged Communities']).agg(
    'Total Project Cost': ['count', 'sum'],
    'Total Project GHGReductions': 'sum'
).round(4)

print("Overall Project Distribution by County:")
print(county_analysis.sort_values(('Total Project Cost', 'sum'), ascending=False))

print("\nVulnerable Communities Analysis:")
print(vulnerability_analysis)

print("\nProject Types in Vulnerable Communities:")
print(project_vulnerability)

# Visualizations
plt.figure(figsize=(15, 8))
# Plot investment distribution...
plt.subplot(1, 2, 1)
county_analysis[['Total Project Cost', 'sum']].sort_values(ascending=False).nlargest(10)
plt.title('Top 10 Counties by Total Investment')
plt.ylabel('Total Investment')
plt.xlabel('County')

# ...and GHG reductions
plt.subplot(1, 2, 2)
county_analysis[['Total Project GHGReductions', 'sum']].sort_values(ascending=False).nlargest(10)
plt.title('Top 10 Counties by GHG Reductions')
plt.ylabel('Total GHG Reductions')
plt.xlabel('County')

plt.tight_layout()
plt.show()

```

Overall Project Distribution by County:

County	Total Project Cost	\ count
Los Angeles	38032	
Alameda	6381	
Alameda, Contra Costa, Fresno, Kern, Kings, Mad...	9	
Sacramento	4781	
San Diego	8468	
San Francisco	2748	
Santa Clara	7154	
San Francisco, Santa Clara	6	
Orange	10694	
San Diego, Santa Barbara, Ventura	23	

County	sum
Los Angeles	69936060480
Alameda	10379456877
Alameda, Contra Costa, Fresno, Kern, Kings, Mad...	7333200000
Sacramento	5858784824
San Diego	4870580185
San Francisco	3813460158
Santa Clara	3178166479
San Francisco, Santa Clara	2998442533
Orange	2383485407
San Diego, Santa Barbara, Ventura	2235315130

County	Total Project GHGReduction
Los Angeles	1130770
Alameda	1051879
Alameda, Contra Costa, Fresno, Kern, Kings, Mad...	54640
Sacramento	197590
San Diego	223342
San Francisco	185098
Santa Clara	179384
San Francisco, Santa Clara	153806
Orange	118052
San Diego, Santa Barbara, Ventura	79143

dac_funding_ratio \

	mean
County	
Los Angeles	0.5751
Alameda	0.2983
Alameda, Contra Costa, Fresno, Kern, Kings, Mad...	NaN
Sacramento	0.4194
San Diego	0.1004
San Francisco	0.1057
Santa Clara	0.2247
San Francisco, Santa Clara	0.0101
Orange	0.3994
San Diego, Santa Barbara, Ventura	NaN

Is Benefit Disadvantaged C

ommunities \

mean

County	
Los Angeles	0.5420
Alameda	0.2612
Alameda, Contra Costa, Fresno, Kern, Kings, Mad...	0.0000
Sacramento	0.3403
San Diego	0.0993
San Francisco	0.1121
Santa Clara	0.2200
San Francisco, Santa Clara	0.1667
Orange	0.3571
San Diego, Santa Barbara, Ventura	0.0000

Is Low Income Communities

	mean
County	
Los Angeles	0.6534
Alameda	0.5557
Alameda, Contra Costa, Fresno, Kern, Kings, Mad...	1.0000
Sacramento	0.5162
San Diego	0.5265
San Francisco	0.5761
Santa Clara	0.4911
San Francisco, Santa Clara	0.8333
Orange	0.6461
San Diego, Santa Barbara, Ventura	1.0000

Vulnerable Communities Analysis:

	Total Projects	Total Investment	GHG Reduction
All Communities	141429	130205367663	109154512

Disadvantaged Communities	47167	14111685888	12626682
Low Income Communities	82595	103691488875	55739585

Project Types in Vulnerable Communities:

Project Cost \	count	Is Benefit Disadvantaged Communities	Total Pro
Agricultural Projects	1490	False	
Air Quality Projects	198	False	
Business - Small	5514	False	
Community Projects	2	False	
Individual - Home Energy	973	False	
Individual - Vehicle	1616	True	
Institutional - Education	5811	False	
Institutional - Government	2970	True	
Institutional - Large Business	1568	False	
Institutional - Nonprofit	17	False	
Other	269	False	
Transportation - Other	100	False	
Water-Energy Projects	13788	True	
	552	False	
	64674	True	
	36753	False	
	2826	True	
	2285		
\			
	sum		
Agricultural Projects	269	False	824640
		True	61874

705		
Air Quality Projects	False	1211608
583		
Business - Small	False	65
376		
Community Projects	False	508817
727		
Individual - Home Energy	False	117426
101		
	True	202582
546		
Individual - Vehicle	False	839347
111		
	True	447071
030		
Institutional - Education	False	5561
672		
Institutional - Government	False	16939974
420		
Institutional - Large Business	False	513921
827		
Institutional - Nonprofit	False	1402247
011		
Other	False	91784138
193		
	True	12787875
145		
Transportation - Other	False	1888149
934		
	True	584940
097		
Water-Energy Projects	False	57783
551		
	True	27342
365		

Total Pro

ject GHGReductions

sum

project_category Is Benefit Disadvantaged Communities

Agricultural Projects False

22903912

True

1345466

Air Quality Projects False

250801

Business - Small False

0

Community Projects False

154010

Individual - Home Energy False

203507

True

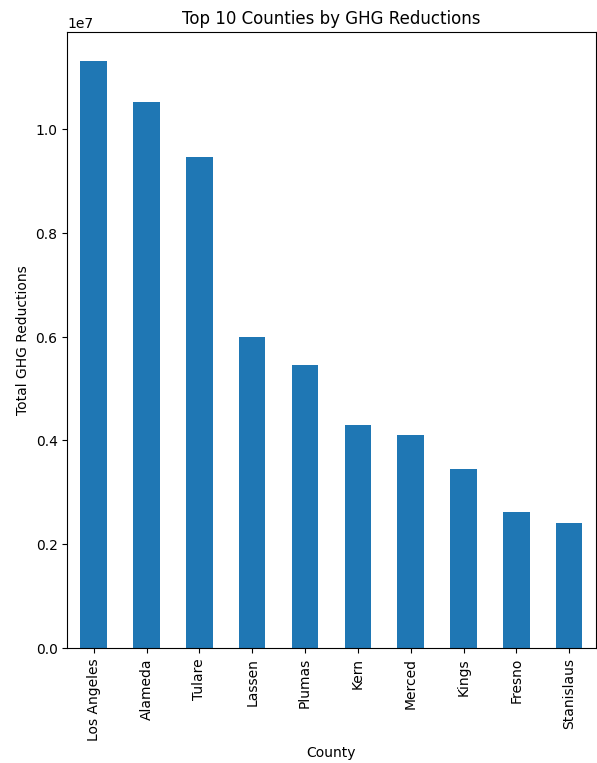
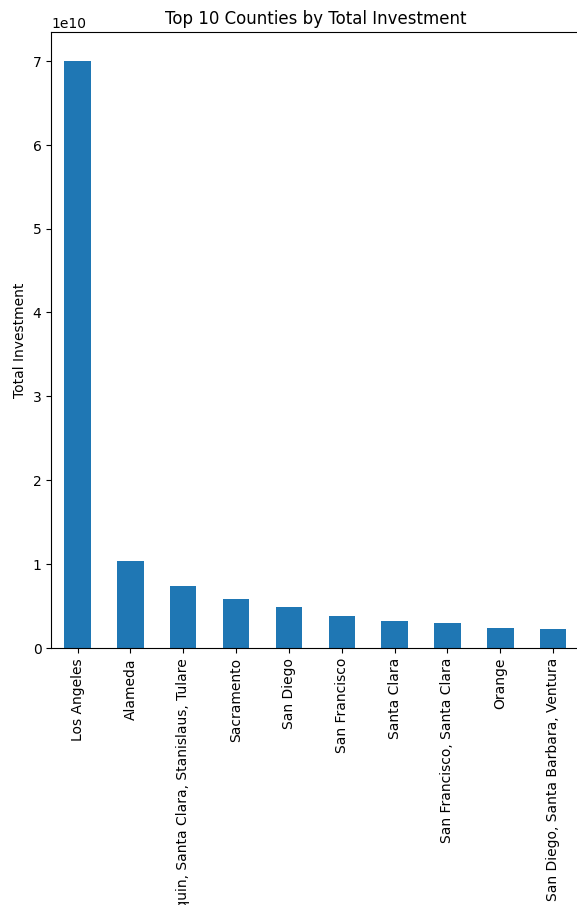
490407

Individual - Vehicle False

699758		True
294234		
Institutional - Education	457	False
Institutional - Government	6580289	False
Institutional - Large Business	388663	False
Institutional - Nonprofit	4791834	False
Other	57935224	False
		True
9377795		
Transportation - Other	2392628	False
		True
920302		
Water-Energy Projects	226747	False
		True
198478		

```
/tmp/ipykernel_738944/3838036285.py:63: UserWarning: Tight layout not applied
. The bottom and top margins cannot be made large enough to accommodate all Axes decorations.
```

```
plt.tight_layout()
```



```
In [27]: import geopandas as gpd
import matplotlib.pyplot as plt
import numpy as np

# First, summarize the data at tract level
tract_summary = data.groupby('Census Tract').agg({
    'Total Project Cost': ['count', 'sum'],
    'Total Project GHGReductions': 'sum',
    'Is Benefit Disadvantaged Communities': 'mean',
    'Is Low Income Communities': 'mean'
})
```

```
}).round(4)

# Reset the index to make Census Tract a column
tract_summary = tract_summary.reset_index()

# Create a few different maps to show different aspects
fig, ((ax1, ax2), (ax3, ax4)) = plt.subplots(2, 2, figsize=(20, 20))

# Function to create a map with proper legend
def create_map(data, column, title, ax, cmap):
    im = ax.scatter(data['Census Tract'],
                    data[column],
                    c=data[column],
                    cmap=cmap)
    plt.colorbar(im, ax=ax, label=title)
    ax.set_title(title)
    ax.axis('off')

# 1. Project Count Map
create_map(tract_summary,
           ('Total Project Cost', 'count'),
           'Number of CCI Projects by Census Tract',
           ax1,
           'YlOrRd')

# 2. Total Investment Map
create_map(tract_summary,
           ('Total Project Cost', 'sum'),
           'Total CCI Investment by Census Tract',
           ax2,
           'viridis')

# 3. DAC Benefits Map
create_map(tract_summary,
           ('Is Benefit Disadvantaged Communities', 'mean'),
           'Proportion of Projects Benefiting DACs',
           ax3,
           'RdYlBu')

# 4. GHG Reductions Map
create_map(tract_summary,
           ('Total Project GHGReductions', 'sum'),
           'Total GHG Reductions by Census Tract',
           ax4,
           'YlGn')

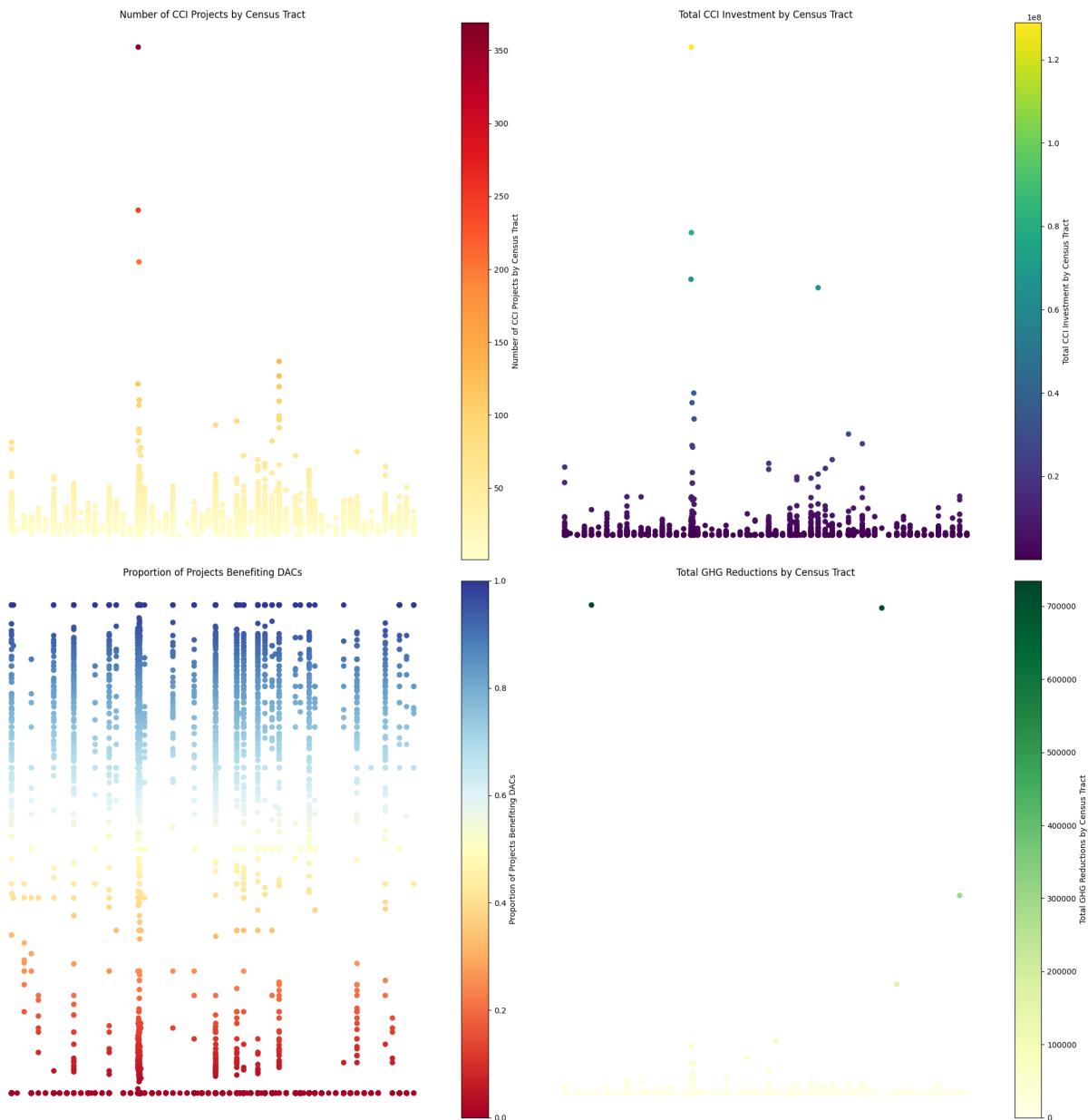
plt.tight_layout()
plt.show()

# Let's also look at the basic statistics
print("\nSpatial Distribution Statistics:")
print("\nTop 10 Census Tracts by Total Investment:")
print(tract_summary.nlargest(10, ('Total Project Cost', 'sum')))

print("\nTop 10 Census Tracts by Project Count:")
print(tract_summary.nlargest(10, ('Total Project Cost', 'count')))
```

```
print("\nTop 10 Census Tracts by GHG Reductions:")
print(tract_summary.nlargest(10, ('Total Project GHGReductions', 'sum')))

# Additional visualization of investment intensity
plt.figure(figsize=(15, 10))
tract_summary['investment_per_project'] = tract_summary[['Total Project Cost
plt.scatter(tract_summary[['Is Benefit Disadvantaged Communities', 'mean']],
            tract_summary['investment_per_project'],
            alpha=0.5)
plt.xlabel('Proportion of Projects Benefiting DACs')
plt.ylabel('Investment per Project ($)')
plt.title('Investment Intensity vs DAC Benefits')
plt.show()
```



Spatial Distribution Statistics:

Top 10 Census Tracts by Total Investment:

	Census Tract	Total Project Cost	Total Project GHGReductions		
\			count	sum	sum
1686	6.037206e+09		369	128838282	71147
1701	6.037208e+09		246	79854037	21955
1250	6.037113e+09		115	67551314	45625
5715	6.073006e+09		38	65309556	1546
3340	6.037901e+09		61	37506825	44030
2391	6.037408e+09		207	34961176	31158
3482	6.037980e+09		67	30672161	23886
6642	6.081602e+09		61	26686653	14972
7164	6.085512e+09		49	24118370	17130
2268	6.037403e+09		81	23674333	11183

Is Benefit Disadvantaged Communities Is Low Income Communities

	mean	mean
1686	0.5176	0.0108
1701	0.7886	0.0447
1250	0.0087	0.0435
5715	0.0000	0.5789
3340	0.0000	0.9672
2391	0.8164	0.7923
3482	0.0000	0.0149
6642	0.0000	0.9508
7164	0.0000	0.0204
2268	0.3333	0.8148

Top 10 Census Tracts by Project Count:

	Census Tract	Total Project Cost	Total Project GHGReductions		
\			count	sum	sum
1686	6.037206e+09		369	128838282	71147
1701	6.037208e+09		246	79854037	21955
2391	6.037408e+09		207	34961176	31158
6443	6.077002e+09		132	616605	650
6448	6.077002e+09		121	401633	284
1250	6.037113e+09		115	67551314	45625
6452	6.077003e+09		113	438571	580
2670	6.037504e+09		103	13744994	12914
6449	6.077002e+09		102	534104	625
2276	6.037403e+09		99	5585697	12839

Is Benefit Disadvantaged Communities Is Low Income Communities

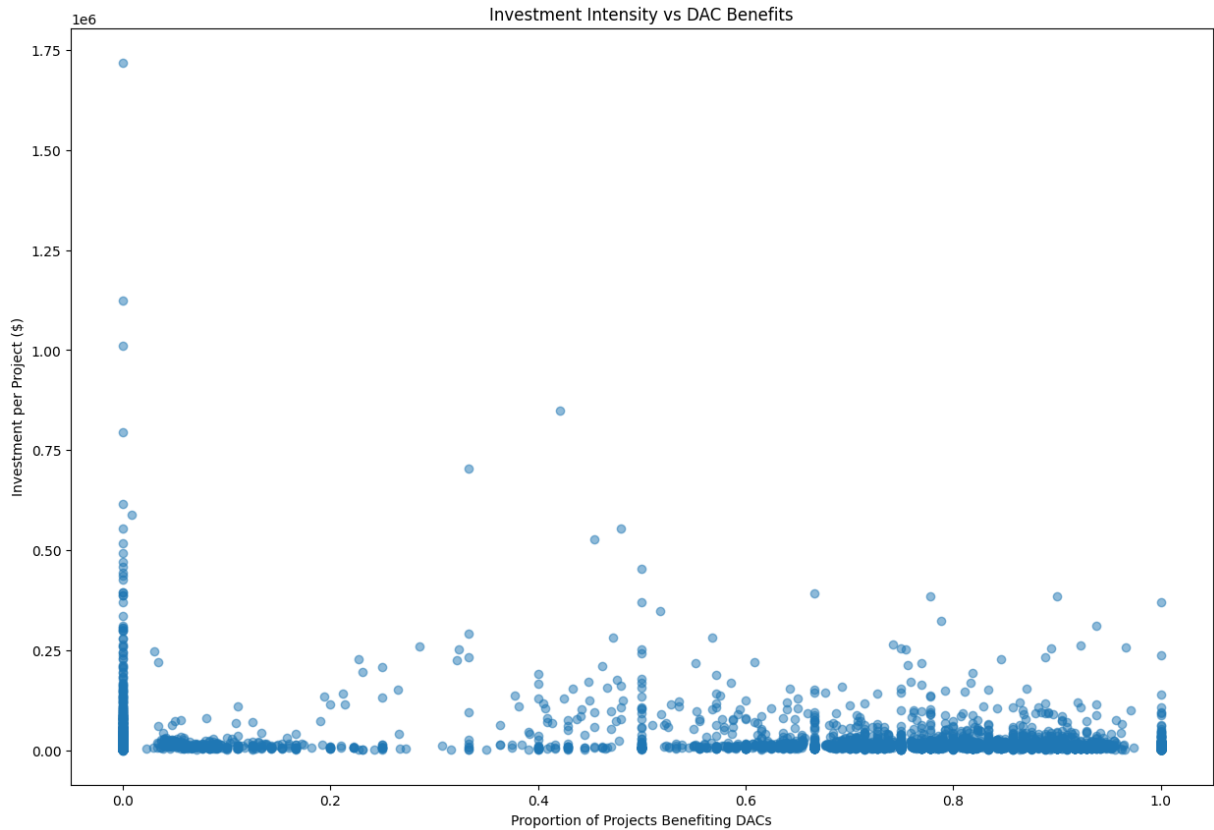
	mean	mean
1686	0.5176	0.0108
1701	0.7886	0.0447
2391	0.8164	0.7923
6443	0.0985	0.9848
6448	0.0661	0.9835
1250	0.0087	0.0435
6452	0.0885	0.9469
2670	0.1942	0.0583
6449	0.1275	0.9804

2276 0.0000 0.1818

Top 10 Census Tracts by GHG Reductions:

Census Tract	Total Project Cost	Total Project GHGReductions
419	6.009000e+09	6718556
7330	6.091010e+09	1842014
7963	6.113011e+09	9324110
7411	6.095253e+09	2631219
4432	6.061022e+09	1199979
1686	6.037206e+09	128838282
3677	6.053011e+09	5634176
1250	6.037113e+09	67551314
3340	6.037901e+09	37506825
4166	6.059086e+09	18951369

Census Tract	Is Benefit Disadvantaged Communities	Is Low Income Communities
419	0.0000	0.1538
7330	0.0000	0.7500
7963	0.0000	0.1429
7411	0.0000	0.2353
4432	0.0000	0.0000
1686	0.5176	0.0108
3677	0.7692	0.3462
1250	0.0087	0.0435
3340	0.0000	0.9672
4166	0.3214	0.2262



Analysis of Spatial Distribution and Environmental Justice Implications

Key Findings:

1. Individual vs. Large Project Analysis Key Findings:

- Individual Program Distribution:
 - Home Energy/Solar Programs:
 - Strong equity performance (78% benefit DACs)
 - High DAC funding ratio (0.65)
 - 93% benefit low-income communities
 - Lower total investment (\$320M)
 - Vehicle Programs:
 - Lower equity performance (35% benefit DACs)
 - Low DAC funding ratio (0.17)
 - Only 41% benefit low-income communities
 - Higher individual costs (~283K vs 43K for solar)
- Large Project Performance:
 - Higher total investment (\$127B)
 - Highest total GHG reductions (103.5M tons)
 - Mixed equity outcomes:
 - Low DAC funding ratio (0.28)
 - Only 6.6% benefit DACs
 - But 56% benefit low-income communities

2. Spatial Distribution Analysis Key Findings:

- Geographic Concentration:
 - Urban dominance (LA: 27% of projects, 54% of funds)
 - Top 3 counties (LA, Alameda, San Diego) = 37% of projects
 - Clear urban-rural divide in investment distribution
- Environmental Justice Implications:
 - DACs receive 33% of projects but only 11% of funding
 - Low-income communities show better performance (58% of projects, 80% of funding)
 - Significant variation in DAC benefit rates across regions

Next Steps:

1. Individual vs. Large Project Analysis

- Detailed program design analysis:
 - What makes home energy programs more successful at reaching DACs?
 - What barriers exist in vehicle programs?
 - How can large project DAC benefits be improved?
- Cost-effectiveness comparison:
 - Calculate and compare GHG reduction per dollar across programs
 - Analyze trade-offs between equity and efficiency
- Policy recommendations:
 - Identify best practices from successful programs
 - Suggest modifications for underperforming programs

2. Spatial Distribution Analysis

- Enhanced geographic analysis:
 - Create detailed maps showing investment patterns
 - Analyze urban-rural disparities
 - Examine relationship between CalEnviroScreen scores and investment patterns
- Environmental justice metrics:
 - Develop composite indicators of investment equity
 - Analyze temporal changes in distribution
 - Compare to demographic and socioeconomic patterns

3. Integration and Synthesis

- Combine findings from both approaches to:
 - Identify overlapping patterns
 - Develop comprehensive policy recommendations
 - Create framework for evaluating climate investment equity

4. Specific Data Analysis Tasks:

```
# 1. Create program efficiency metrics
program_efficiency = data.groupby('project_category').agg({
    'Total Project GHGReductions': 'sum',
    'Total Project Cost': 'sum'
}).assign(
    efficiency = lambda x: x['Total Project GHGReductions'] /
x['Total Project Cost']
)
```

```
# 2. Analyze spatial equity
spatial_equity = merged_data.groupby('CES_quartile').agg({
    'Total Project Cost': ['sum', 'mean', 'count'],
    'Total Project GHGReductions': ['sum', 'mean']
})
```

```
})  
  
# 3. Create temporal analysis  
temporal_patterns = data.groupby(['project_category', 'Date  
Operational']).agg({  
    'Total Project Cost': 'sum',  
    'Is Benefit Disadvantaged Communities': 'mean'  
})
```

Would you like me to focus on developing any of these next steps in more detail? We could:

1. Create specific visualizations for the paper
2. Develop detailed policy recommendations
3. Conduct additional statistical analyses
4. Create a framework for program evaluation