Advocating for Affordable Childcare

November 16, 2024

- Sauda Haywood
- DSC640 Final Project

0.1 Introduction

The cost of childcare has been steadily rising across the United States, which creates financial burdens to families as many parents struggle to balance work with the high cost of early childhood care. In this project I analyzed childcare cost data and present findings to parents in an accessible way, encouraging them to vote for candidates who prioritize affordable childcare policies. I highlighted the growing of the cost of childcare and convinved parents that this rise can be prevented if they vote for the right candidates.

The analysis was conducted using national childcare price data across various years and states. The data covers toddler, preschool, and infant care costs. Using this data, I was able to examine trends over time, identify geographical disparities, and assess the affordability of childcare relative to household income. Outlier detection helped identify regions where childcare costs were exceptionally high, and affordability analyses showed the financial burden of childcare on low- and middle-income families.

Target Audience: Parents

- The primary target audience for this project is parents, especially those who are facing financial challenges due to the high costs of childcare. My goal is to help parents understand the extent of the childcare affordability crisis and motivate them to vote for candidates who prioritize affordable childcare policies.

Forms of Communication Mediums

- Facebook Post: I will make post on Facebook with pictures of visuals and statistics, and add a paragraph explaining the child care crisis and way to avoid. I choose facebook because many parents/families are present on facebook platform.
- Instagram Reel: I will make short, engaging video clip that will feature a quick breakdown of the key data points, focusing on the most shocking statistics for exemple how much childcare costs have increased. I will also emphasize on importance of voting for candidates who prioritize affordable childcare. I chose this midium because families are on social media most of the time, and video is better for those who do not like reading posts.
- Flyer: I will make handouts that will highlight key statistics on the rising costs of childcare and a message to vote for affordable childcare policies. Flyers will help me reach families who are not on social media.

0.2 Data Explortaion

[15]:
```
import pandas as pd
import warnings
warnings.filterwarnings('ignore')
```

[2]:
```
# Load the data
childcare_df = pd.read_excel('nationaldatabaseofchildcareprices.xlsx')

# Display the first 10 rows of the dataset
print("First 10 rows of the dataset:")
print(childcare_df.head(5))
```

```
First 10 rows of the dataset:
  State_Name State_Abbreviation     County_Name  County_FIPS_Code  StudyYear  \
0    Alabama                AL  Autauga County              1001       2008
1    Alabama                AL  Autauga County              1001       2009
2    Alabama                AL  Autauga County              1001       2010
3    Alabama                AL  Autauga County              1001       2011
4    Alabama                AL  Autauga County              1001       2012

   UNR_16  FUNR_16  MUNR_16  UNR_20to64  FUNR_20to64  …  MFCCToddler  \
0    5.42     4.41     6.32         4.6          3.5  …        83.45
1    5.93     5.72     6.11         4.8          4.6  …        87.39
2    6.21     5.57     6.78         5.1          4.6  …        91.33
3    7.55     8.13     7.03         6.2          6.3  …        95.28
4    8.60     8.88     8.29         6.7          6.4  …        99.22

   MFCCToddler_flag  MFCCPreschool  MFCCPreschool_flag  _75FCCInfant  \
0               3.0          81.40                 1.0          97.4
1               3.0          85.68                 1.0         102.0
2               3.0          89.96                 1.0         106.6
3               3.0          94.25                 1.0         111.2
4               3.0          98.53                 1.0         115.8

   _75FCCInfant_flag  _75FCCToddler  _75FCCToddler_flag  _75FCCPreschool  \
0                1.0           97.4                 3.0             95.0
1                1.0          102.0                 3.0            100.0
2                1.0          106.6                 3.0            105.0
3                1.0          111.2                 3.0            110.0
4                1.0          115.8                 3.0            115.0

   _75FCCPreschool_flag
0                   1.0
1                   1.0
2                   1.0
3                   1.0
4                   1.0
```

```
[5 rows x 227 columns]
```

[3]:
```
# Display summary statistics of the dataset
print("\nSummary statistics of the dataset:")
print(childcare_df.describe())
```

```
Summary statistics of the dataset:
       County_FIPS_Code     StudyYear        UNR_16      FUNR_16  \
count      34567.000000  34567.000000  34567.000000  34567.00000
mean       30388.132786   2012.999711      7.465902      7.02902
std        15161.015383      3.162232      3.538619      3.56342
min         1001.000000   2008.000000      0.000000      0.00000
25%        18177.000000   2010.000000      5.100000      4.64000
50%        29177.000000   2013.000000      7.050000      6.59000
75%        45081.000000   2016.000000      9.350000      8.88000
max        56045.000000   2018.000000     36.110000     38.24000

            MUNR_16    UNR_20to64   FUNR_20to64   MUNR_20to64  FLFPR_20to64  \
count  34567.000000  34567.000000  34567.000000  34567.000000  34567.000000
mean       7.860291      6.900073      6.482007      7.275457     70.086125
std        4.037657      3.446199      3.477956      3.990758      7.696499
min        0.000000      0.000000      0.000000      0.000000     33.600000
25%        5.200000      4.600000      4.200000      4.700000     65.100000
50%        7.390000      6.500000      6.000000      6.800000     70.600000
75%        9.920000      8.700000      8.250000      9.200000     75.500000
max       39.740000     33.900000     44.500000     45.500000    100.000000

       FLFPR_20to64_Under6  …   MFCCToddler  MFCCToddler_flag  \
count         34567.000000  …  23383.000000      23383.000000
mean             68.821409  …    106.759749          1.153359
std              11.758088  …     29.982431          0.532176
min               0.000000  …     43.080000          1.000000
25%              62.600000  …     85.085000          1.000000
50%              69.600000  …    100.250000          1.000000
75%              76.100000  …    124.950000          1.000000
max             100.000000  …    376.320000          3.000000

       MFCCPreschool  MFCCPreschool_flag  _75FCCInfant  _75FCCInfant_flag  \
count   23383.000000        23383.000000  23383.000000       23383.000000
mean      104.189510            1.287859    128.909289           1.792841
std        28.961701            0.696762     38.543010           0.818080
min        40.030000            1.000000     50.000000           1.000000
25%        84.255000            1.000000    100.830000           1.000000
50%        99.650000            1.000000    123.150000           2.000000
75%       120.200000            1.000000    146.950000           3.000000
max       331.340000            3.000000    502.970000           3.000000
```

```
       _75FCCToddler  _75FCCToddler_flag  _75FCCPreschool  \
count   23383.000000         23383.00000     23383.000000
mean      120.784283             1.18800       117.897482
std        35.334666             0.58367        34.111188
min        50.000000             1.00000        46.450000
25%        95.850000             1.00000        95.000000
50%       115.000000             1.00000       112.500000
75%       136.270000             1.00000       132.760000
max       439.220000             3.00000       386.720000

       _75FCCPreschool_flag
count          23383.000000
mean               1.294316
std                0.708542
min                1.000000
25%                1.000000
50%                1.000000
75%                1.000000
max                3.000000

[8 rows x 224 columns]
```

- The dataset contains multiple columns, including state names, county details, and various numerical measures related to childcare pricing across years.

[4]:
```
# Select columns related to childcare costs for outlier detection
childcare_cost_columns = ['MFCCToddler', 'MFCCPreschool', '_75FCCInfant',␣
 ↪'_75FCCToddler', '_75FCCPreschool']

# Create a function to detect outliers using IQR
def detect_outliers_iqr(childcare_df, column):
    Q1 = childcare_df[column].quantile(0.25)
    Q3 = childcare_df[column].quantile(0.75)
    IQR = Q3 - Q1
    lower_bound = Q1 - 1.5 * IQR
    upper_bound = Q3 + 1.5 * IQR
    outliers = childcare_df[(childcare_df[column] < lower_bound) |␣
 ↪(childcare_df[column] > upper_bound)]
    return outliers

# Check for outliers in each childcare cost column
for column in childcare_cost_columns:
    outliers = detect_outliers_iqr(childcare_df, column)
    print(f"Outliers in {column}:")
    print(outliers[[column]])
    print(f"Number of outliers in {column}: {len(outliers)}\n")
```

```
Outliers in MFCCToddler:
       MFCCToddler
2049        186.46
2050        191.18
2051        195.89
2052        214.69
2053        233.48
…            …
34268       185.00
34269       200.00
34531       240.57
34532       308.45
34533       376.32

[420 rows x 1 columns]
Number of outliers in MFCCToddler: 420

Outliers in MFCCPreschool:
       MFCCPreschool
2045          178.26
2046          178.56
2047          181.19
2048          183.83
2049          186.46
…              …
34269         200.00
34280         235.00
34531         225.58
34532         278.46
34533         331.34

[571 rows x 1 columns]
Number of outliers in MFCCPreschool: 571

Outliers in _75FCCInfant:
       _75FCCInfant
2045         217.95
2046         226.66
2047         235.37
2048         244.08
2049         252.79
…             …
34268        250.00
34269        225.00
34531        305.78
34532        410.00
34533        502.97
```

```
[782 rows x 1 columns]
Number of outliers in _75FCCInfant: 782

Outliers in _75FCCToddler:
       _75FCCToddler
2045          196.92
2046          204.32
2047          211.72
2048          219.11
2049          226.51
…              …
34529         201.56
34530         201.56
34531         280.78
34532         360.00
34533         439.22

[752 rows x 1 columns]
Number of outliers in _75FCCToddler: 752

Outliers in _75FCCPreschool:
       _75FCCPreschool
854              207.69
2045             196.92
2046             204.32
2047             211.72
2048             219.11
…                 …
34529            201.56
34530            201.56
34531            263.28
34532            325.00
34533            386.72

[747 rows x 1 columns]
Number of outliers in _75FCCPreschool: 747
```

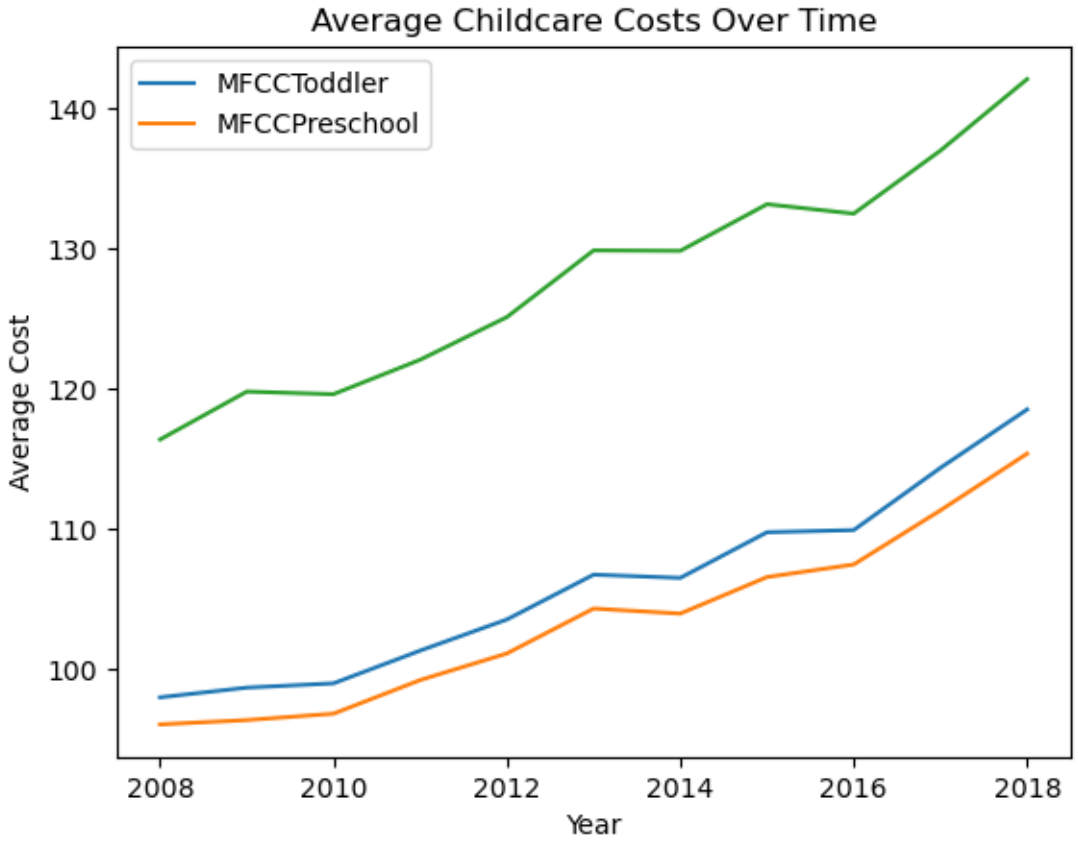
- There are many ouliers that need to be cleaned during the data cleaning step

0.2.1 Checking how childcare costs have changed over the years

[10]:
```
# Checking how childcare costs have changed over the years.

# Group data by year and calculate average childcare costs
avg_cost_by_year = childcare_df.groupby('StudyYear')[['MFCCToddler',␣
 ↪'MFCCPreschool', '_75FCCInfant']].mean()
```

```
# Plot the trend over time
avg_cost_by_year.plot(title='Average Childcare Costs Over Time', xlabel='Year',␣
↪ylabel='Average Cost')
```

[10]:
```
<Axes: title={'center': 'Average Childcare Costs Over Time'}, xlabel='Year',
ylabel='Average Cost'>
```



- These graphs shows that the price has significanlty increased.

0.2.2 Regional Comparisons

[11]:
```
# Group data by state and calculate average childcare costs
avg_cost_by_state = childcare_df.groupby('State_Name')[['MFCCToddler',␣
↪'MFCCPreschool', '_75FCCInfant']].mean().sort_values(by='MFCCToddler',␣
↪ascending=False)

# Display the top 10 states with the highest average toddler childcare costs
print(avg_cost_by_state.head(10))
```

```
                MFCCToddler  MFCCPreschool  _75FCCInfant
State_Name
```

```
Massachusetts           203.701299     199.581169     249.749870
District of Columbia    200.340000     192.800000     250.400000
Rhode Island            188.872000     174.130000     214.432000
Connecticut             184.761364     177.213182     205.767045
New York                166.966613     163.864194     204.085161
California              160.905408     160.905408     203.755392
New Hampshire           159.241000     159.241000     170.691000
Washington              158.543869     145.856807     204.249814
New Jersey              158.157976     145.939048     185.584167
Colorado                157.000000     149.145312     203.050000
```

0.2.3 Affordability Analysis

[12]:
```
# Calculate the percentage of household income spent on childcare for toddlers,␣
 ↪preschool, and infants
childcare_df['Toddler_Cost_Percentage_Income'] = (childcare_df['MFCCToddler'] /␣
 ↪childcare_df['MHI']) * 100
childcare_df['Preschool_Cost_Percentage_Income'] =␣
 ↪(childcare_df['MFCCPreschool'] / childcare_df['MHI']) * 100
childcare_df['Infant_Cost_Percentage_Income'] = (childcare_df['_75FCCInfant'] /␣
 ↪childcare_df['MHI']) * 100

# Display the first few rows to check the new columns
print(childcare_df[['State_Name', 'County_Name', 'MFCCToddler',␣
 ↪'MFCCPreschool', '_75FCCInfant', 'MHI',
             'Toddler_Cost_Percentage_Income',␣
 ↪'Preschool_Cost_Percentage_Income', 'Infant_Cost_Percentage_Income']].head())

# Summary statistics of the percentage of income spent on childcare
print("\nSummary statistics for percentage of income spent on childcare:")
print(childcare_df[['Toddler_Cost_Percentage_Income',␣
 ↪'Preschool_Cost_Percentage_Income', 'Infant_Cost_Percentage_Income']].
 ↪describe())
```

```
  State_Name     County_Name  MFCCToddler  MFCCPreschool  _75FCCInfant  \
0    Alabama  Autauga County        83.45          81.40          97.4
1    Alabama  Autauga County        87.39          85.68         102.0
2    Alabama  Autauga County        91.33          89.96         106.6
3    Alabama  Autauga County        95.28          94.25         111.2
4    Alabama  Autauga County        99.22          98.53         115.8

       MHI  Toddler_Cost_Percentage_Income  Preschool_Cost_Percentage_Income  \
0  50837.0                        0.164152                          0.160120
1  51463.0                        0.169811                          0.166489
2  53255.0                        0.171496                          0.168923
3  53899.0                        0.176775                          0.174864
4  53773.0                        0.184516                          0.183233
```

```
   Infant_Cost_Percentage_Income
0                       0.191593
1                       0.198201
2                       0.200169
3                       0.206312
4                       0.215350

Summary statistics for percentage of income spent on childcare:
       Toddler_Cost_Percentage_Income  Preschool_Cost_Percentage_Income  \
count                    23383.000000                      23383.000000
mean                         0.233908                          0.228371
std                          0.055186                          0.053541
min                          0.088024                          0.088024
25%                          0.196073                          0.191958
50%                          0.229205                          0.223836
75%                          0.263953                          0.257642
max                          0.637731                          0.613737

       Infant_Cost_Percentage_Income
count                    23383.000000
mean                         0.282003
std                          0.069792
min                          0.100968
25%                          0.234134
50%                          0.272682
75%                          0.318117
max                          0.811554
```

0.2.4 Demographic Analysis

[13]:
```
# Calculate the proportion of different races in each county
childcare_df['White_Proportion'] = childcare_df['OneRace_W'] /␣
 ↪childcare_df['TotalPop']
childcare_df['Black_Proportion'] = childcare_df['OneRace_B'] /␣
 ↪childcare_df['TotalPop']

# Check correlation between race and childcare costs
race_correlation = childcare_df[['White_Proportion', 'Black_Proportion',␣
 ↪'MFCCToddler', 'MFCCPreschool']].corr()
print("Race and Childcare Cost Correlation:")
print(race_correlation)
```

```
Race and Childcare Cost Correlation:
                  White_Proportion  Black_Proportion  MFCCToddler  \
White_Proportion          1.000000          0.022957    -0.109490
Black_Proportion          0.022957          1.000000    -0.207226
MFCCToddler              -0.109490         -0.207226     1.000000
MFCCPreschool            -0.108050         -0.210821     0.990089
```

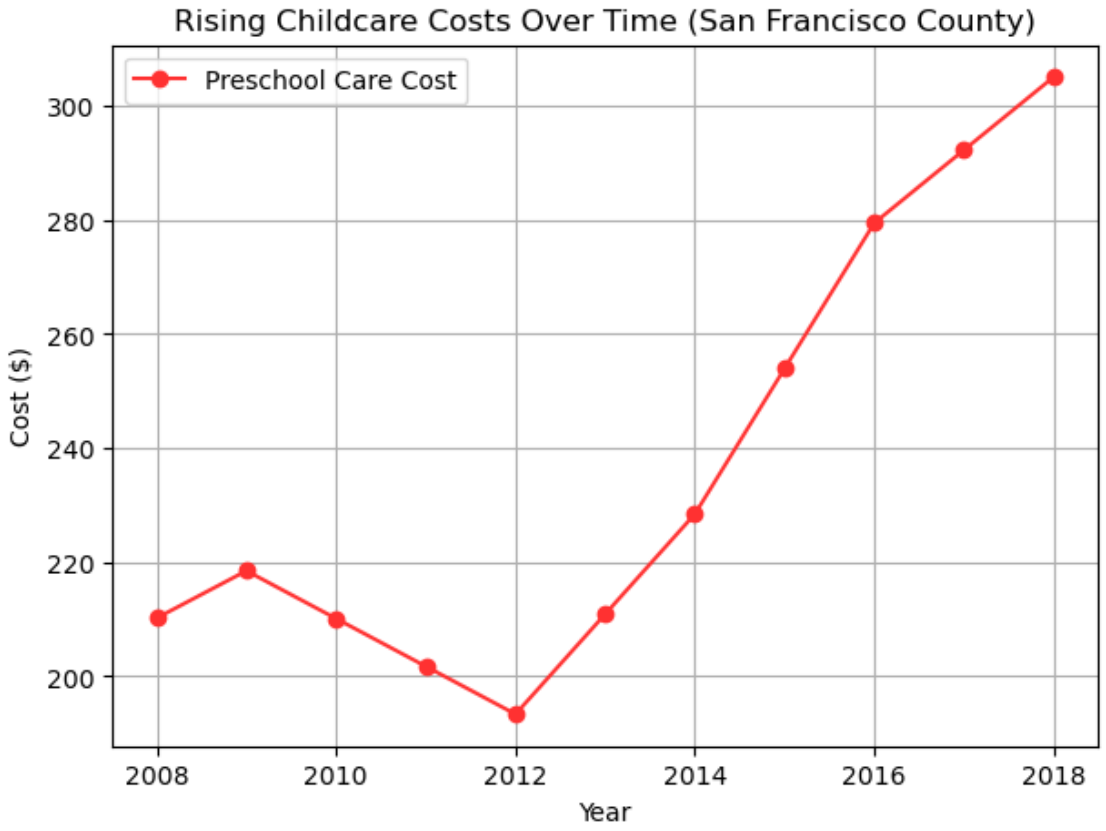
```
                    MFCCPreschool
White_Proportion        -0.108050
Black_Proportion        -0.210821
MFCCToddler              0.990089
MFCCPreschool            1.000000
```

Interesting Findings from the Data

- Rising Childcare Costs: Across different counties and states, the cost of toddler and preschool care has been steadily increasing. This trend places a growing financial burden on families, particularly those with limited incomes.
- Geographic Disparities: There are significant differences in childcare costs across regions. Some states have much higher costs compared to others,
- Income and Childcare Affordability: For many families, a large percentage of their income is spent on childcare, often leaving them with insufficient money for other essential expenses like food and housing.

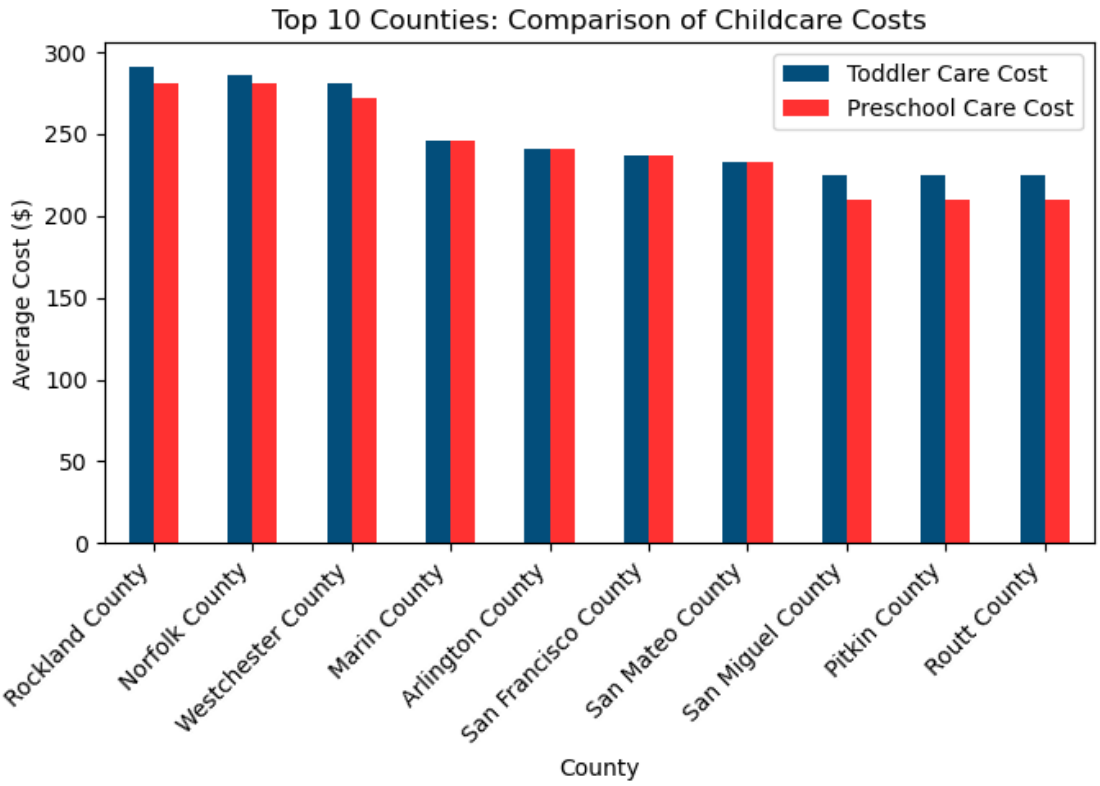
0.3 Visualization

[17]:
```
# Trend of Childcare Costs Over Time
import matplotlib.pyplot as plt


# Convert relevant cost columns to numeric after removing the dollar sign
childcare_df['MFCCPreschool'] = childcare_df['MFCCPreschool'].replace('[\$,]',␣
 ↪'', regex=True).astype(float)

# Filter data for a specific county (e.g., Autauga County)
county_data = childcare_df[childcare_df['County_Name'] == 'San Francisco␣
 ↪County']

# Create a plot for the trend of childcare costs over time
plt.figure(figsize=(7,5))
plt.plot(county_data['StudyYear'], county_data['MFCCPreschool'],␣
 ↪label='Preschool Care Cost', marker='o', color='#ff3131')

plt.title('Rising Childcare Costs Over Time (San Francisco County)')
plt.xlabel('Year')
plt.ylabel('Cost ($)')
plt.legend()
plt.grid(True)
plt.savefig('childcare_cost_trend.png')
plt.show()
```

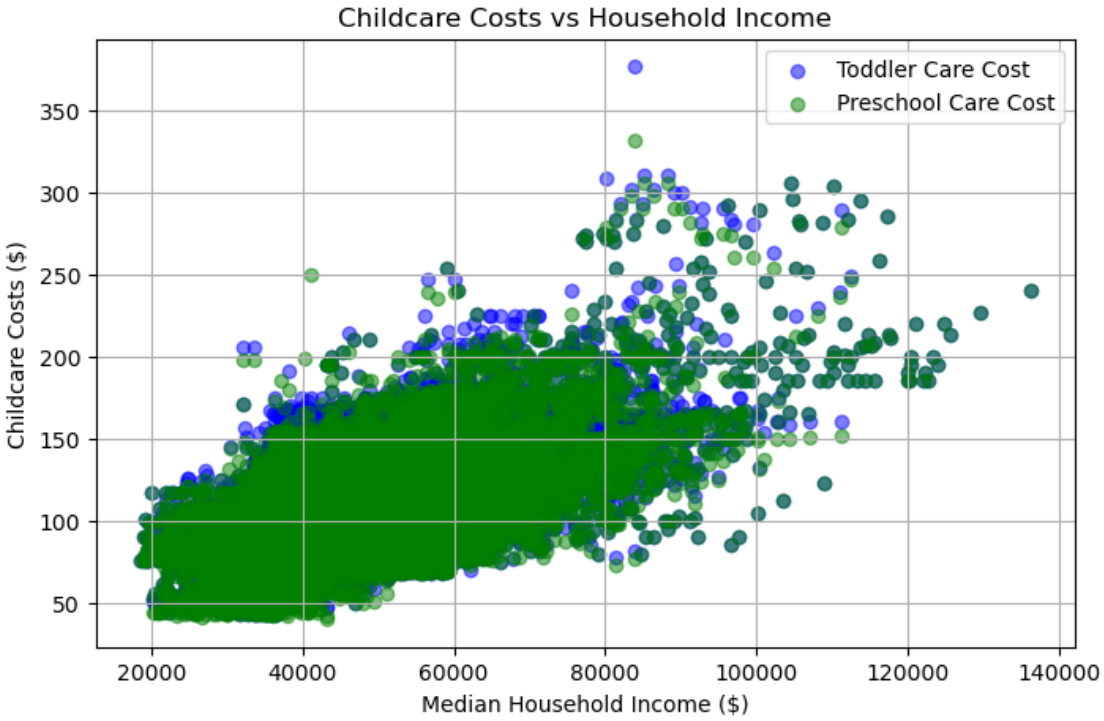


[32]:
```
# Group data by county and calculate the average childcare costs
average_costs = childcare_df.groupby('County_Name')[['MFCCToddler',␣
 ↪'MFCCPreschool']].mean()

# Sort by toddler care cost for better visualization and select the top 10␣
 ↪counties
average_costs_top10 = average_costs.sort_values(by='MFCCToddler',␣
 ↪ascending=False).head(10)

# Plot a bar chart comparing costs across the top 10 counties
average_costs_top10.plot(kind='bar', figsize=(7, 5), color=['#034e7b',␣
 ↪'#ff3131'])

plt.title('Top 10 Counties: Comparison of Childcare Costs')
plt.xlabel('County')
plt.ylabel('Average Cost ($)')
plt.xticks(rotation=45, ha='right')
plt.legend(['Toddler Care Cost', 'Preschool Care Cost'])
plt.tight_layout()
```

```
plt.savefig('childcare_cost_comparison_top10.png')  # Save visualization to␣
 ↪include in the flyer
plt.show()
```

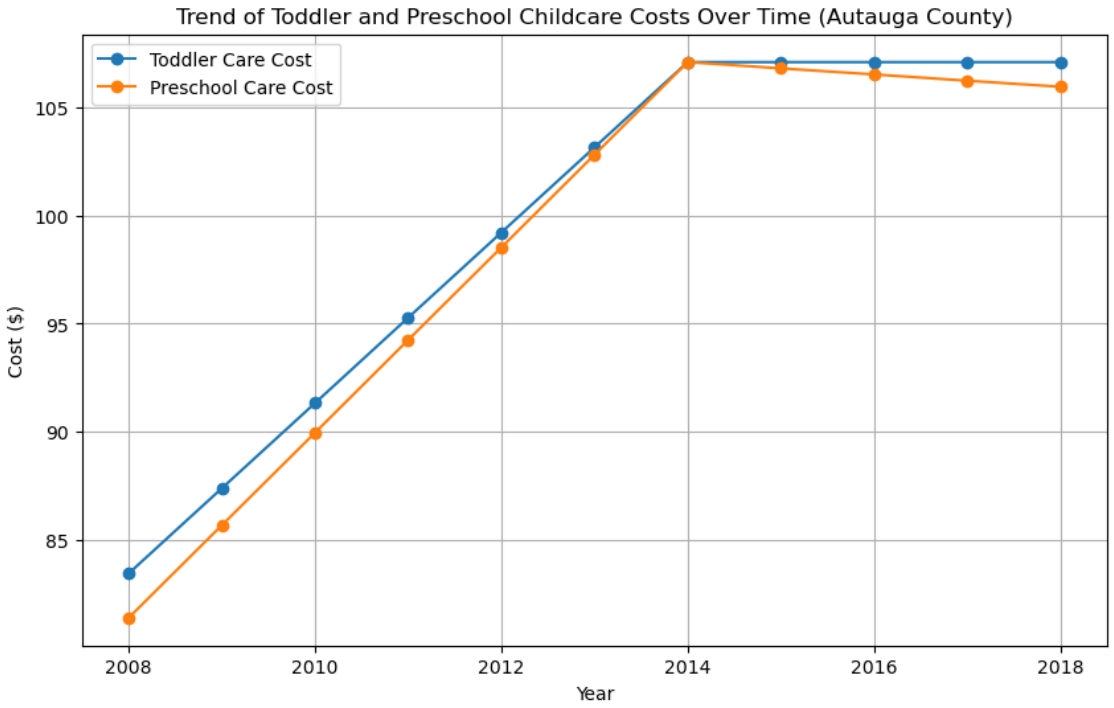


[22]:
```
# Scatter plot of household income vs childcare costs
plt.figure(figsize=(8,5))
plt.scatter(childcare_df['MHI'], childcare_df['MFCCToddler'], alpha=0.5,␣
 ↪label='Toddler Care Cost', color='blue')
plt.scatter(childcare_df['MHI'], childcare_df['MFCCPreschool'], alpha=0.5,␣
 ↪label='Preschool Care Cost', color='green')

plt.title('Childcare Costs vs Household Income')
plt.xlabel('Median Household Income ($)')
plt.ylabel('Childcare Costs ($)')
plt.legend()
plt.grid(True)
plt.savefig('childcare_income_correlation.png')
plt.show()
```



[10]:
```
import pandas as pd
import matplotlib.pyplot as plt

# Convert relevant cost columns to numeric after removing the dollar sign
childcare_df['MFCCToddler'] = childcare_df['MFCCToddler'].replace('[\$,]', '',␣
 ↪regex=True).astype(float)
childcare_df['MFCCPreschool'] = childcare_df['MFCCPreschool'].replace('[\$,]',␣
 ↪'', regex=True).astype(float)

# Filter data for a specific county (e.g., Autauga County) or state (e.g.,␣
 ↪Alabama)
county_data = childcare_df[childcare_df['County_Name'] == 'Autauga County']

# Plot the trend of childcare costs over time
plt.figure(figsize=(10,6))
plt.plot(county_data['StudyYear'], county_data['MFCCToddler'], label='Toddler␣
 ↪Care Cost', marker='o')
plt.plot(county_data['StudyYear'], county_data['MFCCPreschool'],␣
 ↪label='Preschool Care Cost', marker='o')

plt.title('Trend of Toddler and Preschool Childcare Costs Over Time (Autauga␣
 ↪County)')
plt.xlabel('Year')
```

```
plt.ylabel('Cost ($)')
plt.legend()
plt.grid(True)
plt.show()
```



0.4 Conclusion

The project yielded several significant insights:

- Rising Childcare Costs: The cost of childcare for toddlers, preschoolers, and infants has been increasing steadily across all states. This places a growing financial burden on families, particularly those in lower-income brackets.
- Geographic Disparities: There are significant differences in childcare costs across states. For example, Massachusetts and the District of Columbia consistently have some of the highest costs, while states like Alabama have lower rates.
- Income and Childcare Affordability: For many families, a large percentage of household income is spent on childcare, which can lead to financial strain. This often leaves insufficient money for other essential expenses like housing and food.

Assumptions:

Several assumptions were made throughout the analysis: * It was assumed that the reported data is accurate and consistent across counties and states. * The assumption was made that income data is a sufficient indicator of affordability without accounting for local subsidies, discounts, or childcare assistance programs. * Missing or incomplete data points were assumed to not significantly impact the overall analysis and were excluded.

Items That Still Need Clarification:

- Column Names: Several columns in the dataset are abbreviated, which required additional research and interpretation to understand their meaning.
- Local Factors: The dataset does not account for local policies or subsidies that may affect the overall childcare costs for certain families, which could provide more context on affordability.

Design Decisions:

Design decisions were made to ensure that the data is communicated in a clear, concise, and visually appealing way. For example, bar charts and pie charts are used to illustrate rising costs and regional disparities. They are easy to understand by non-technical parents.

Ethical Considerations:

- Changes to Data: Data cleaning included removing dollar signs and handling outliers. No major transformations were applied that would mislead the analysis.
- Legal and Regulatory Guidelines: As the data used is public and pertains to financial matters, there are no specific privacy concerns. It is essential to ensure accuracy in how the data is presented to avoid misrepresentation.
- Assumptions in Data Cleaning: Outliers were flagged but not removed, and all filtered data points were clearly documented.
- Data Sourcing and Credibility: The data was sourced from a national database, ensuring credibility.
- Mitigating Ethical Implications: Clear explanations of the data and visualizations will be provided to avoid misleading conclusions. Additionally, all assumptions and transformations will be documented to maintain transparency.

Lessons Learned:

- What would I do differently?: In the future, I would like to explore other factors influencing childcare affordability, such as local assistance programs and the impact of minimum wage policies.
- What did I enjoy the most?: The visual storytelling process was the most enjoyable part of the project. Being able to take complex data and come up with insights and visualizations for creating content was enjoyable.