EDA on Childcare Prices Dataset

Kaylynn Mosier

5 September 2024

Summary of Analysis I began exploratory analysis by reducing granularity of the dataset. I found the medians of all interesting variables for each year by state. This allowed me to more easily find relationships within the data. In future analysis, I plan to bring granularity back into the analysis. The data revealed that unemployment rate and median household income were inversely related. Peak unemployment rates were in 2013, and the lowest unemployment rates were in 2018. The highest poverty rates for both families and individuals was the same year as peak unemployment. In all years, poverty rates were higher for individuals than families. Childcare rates for both center-based and in-home care facilities were highest in 2014. Throughout all years, in both center-based and in-home care facilities, infant care was the most expensive and preschool care was the least expensive. My plan, further outlined below, is to convince decision makers at a company that they should subsidize childcare within their company. I will focus on Walmart and employees that work at the corporate location, so I chose to research statistics about Arkansas versus the rest of the country. This analysis revealed that Arkansas has a higher poverty rate and unemployment rate than the rest of the country. Arkansas has a lower median household income and lower median earnings for males and females than the rest of the country. However, Arkansas does have lower childcare rates for all age groups in both center-based and family-based facilities than the rest of the country.

Message I intend to used this data as a call-to-action to employers to provide and/or subsidize childcare within their corporation.

Target Audience The target audience will be the decision-makers in the company. These will likely be executive level leaders with the power to make change within the company. These individuals will not be familiar with this specific data, but they are comfortable interacting with data and visualizations. As a subordinate employee in this company, it is crucial that I show I have credibility in this topic. It will also be important to show the audience that other employees want this service.

Proposed Mediums

1. Corporate email lobbying for time in the next board meeting for my proposal
2. Presentation, likely on PowerPoint to be presented during the board meeting

3. Interactive dashboard, likely on PowerBI to be sent to board members after presentation. This will allow me to highlight key metrics to drive the pitch home.

Analysis

In [2]:
```
# Load libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

In [3]:
```
childcare_prices = pd.read_excel("C:/Users/kayly/OneDrive/Desktop/MSDS/DSC640/Final Project/nationaldatabaseofchildca
childcare_prices.head()
```

Out[3]:

	State_Name	State_Abbreviation	County_Name	County_FIPS_Code	StudyYear	UNR_16	FUNR_16	MUNR_16	UNR_20to64	F
0	Alabama	AL	Autauga County	1001	2008	5.42	4.41	6.32	4.6	
1	Alabama	AL	Autauga County	1001	2009	5.93	5.72	6.11	4.8	
2	Alabama	AL	Autauga County	1001	2010	6.21	5.57	6.78	5.1	
3	Alabama	AL	Autauga County	1001	2011	7.55	8.13	7.03	6.2	
4	Alabama	AL	Autauga County	1001	2012	8.60	8.88	8.29	6.7	

5 rows × 227 columns

In [4]:
```
childcare_prices.shape
```

Out[4]:
```
(34567, 227)
```

Need to Calculate:

- Median unemployment rate of each year, by state
- Median poverty rate of all families of each year, by state
- Median poverty rate of all people of each year, by state
- Median household income of each year, by state
- Median female earnings of each year, by state
- Median male earnings of each year, by state
- Total population of each year, by state
- Median cost of childcare of each age group each year for center-based facility, by state
- Median cost of childcare of each age gorup each year for family-based facility, by state

Store all values in new dataframe indexed by state

In [5]:
```
# Create a list of state names and state abbreviations
states = childcare_prices['State_Name'].unique()
state_abbrev = childcare_prices['State_Abbreviation'].unique()

# Create datatable to store median values by state, add states name and stat abbreviations as columns
medians_by_state = pd.DataFrame()
medians_by_state['State_Name'] = states
medians_by_state['State_Abbreviation'] = state_abbrev
medians_by_state.set_index('State_Name', inplace=True)

medians_by_state.head()
```

Out[5]:

State_Name	State_Abbreviation
Alabama	AL
Alaska	AK
Arizona	AZ
Arkansas	AR
California	CA

In [6]:

```
# Find median unemployment rate of each year by state
median_unemployment = childcare_prices.groupby(['State_Name', 'StudyYear'])['UNR_16'].median()

# Create dataframe of median unemployment and set the index to state name
median_unemployment = pd.DataFrame(median_unemployment)
median_unemployment.reset_index(inplace=True)
median_unemployment.set_index('State_Name', inplace=True)

# Merge two dataframes on State Name
medians_by_state= pd.merge(medians_by_state, median_unemployment, on='State_Name', how='outer')
medians_by_state
```

Out[6]:

State_Name	State_Abbreviation	StudyYear	UNR_16
Alabama	AL	2008	7.06
Alabama	AL	2009	8.17
Alabama	AL	2010	9.51
Alabama	AL	2011	10.24
Alabama	AL	2012	11.00
...	...	...	...
Wyoming	WY	2014	4.30
Wyoming	WY	2015	4.40
Wyoming	WY	2016	4.51
Wyoming	WY	2017	4.60
Wyoming	WY	2018	4.34

561 rows × 3 columns

In [7]:
```
columns_needed = ['PR_F', 'PR_P', 'MHI', 'FME', 'MME', 'TotalPop', 'MCSA','MCInfant', 'MCToddler', 'MCPreschool',
                  'MFCCSA','MFCCInfant', 'MFCCToddler', 'MFCCPreschool']

for column in columns_needed:
    # Find median of each value for each year by state
    column_data = childcare_prices.groupby(['State_Name', 'StudyYear'])[column].median()

    # Create dataframe of column and set the index to state name
    column_df = pd.DataFrame(column_data)
    column_df.reset_index(inplace=True)
    column_df.set_index('State_Name', inplace=True)

    # Merge two dataframes on State Name
    medians_by_state = pd.merge(medians_by_state, column_df, on=['State_Name','StudyYear'], how='outer')
```

```
medians_by_state.reset_index(inplace=True)
medians_by_state
```

Out[7]:

	State_Name	State_Abbreviation	StudyYear	UNR_16	PR_F	PR_P	MHI	FME	MME	TotalPop	MCSA	MCInfant
0	Alabama	AL	2008	7.06	13.9	18.1	34952.33	17938.0	30016.11	34635.0	83.76	98.77
1	Alabama	AL	2009	8.17	14.0	18.1	35137.00	18225.0	29594.00	34704.0	86.17	100.11
2	Alabama	AL	2010	9.51	14.7	18.7	36077.00	18879.0	29589.00	34579.0	88.58	101.44
3	Alabama	AL	2011	10.24	15.4	19.7	37161.00	19989.0	30288.00	34375.0	90.99	102.78
4	Alabama	AL	2012	11.00	15.4	19.8	37059.00	20084.0	30829.00	34228.0	93.10	104.11
...	...	...	...	...	...	...	...	...	...	...	...	...
556	Wyoming	WY	2014	4.30	7.4	10.5	56318.00	22221.0	40721.00	15795.0	95.61	154.38
557	Wyoming	WY	2015	4.40	7.5	10.7	56569.00	21941.0	40851.00	15739.0	100.68	149.12
558	Wyoming	WY	2016	4.51	7.5	11.0	55640.00	21942.0	40254.00	15696.0	105.67	143.14
559	Wyoming	WY	2017	4.60	7.5	10.5	57644.00	23761.0	40395.00	15689.0	110.47	136.87
560	Wyoming	WY	2018	4.34	8.0	11.6	58521.00	24679.0	43136.00	15477.0	115.25	137.57

561 rows × 18 columns

In [8]:
```
# Find number of missing values
medians_by_state.isna().sum()
```

Out[8]:
```
State_Name              0
State_Abbreviation      0
StudyYear               0
UNR_16                  0
PR_F                    0
PR_P                    0
MHI                     0
FME                     0
MME                     0
TotalPop                0
MCSA                  173
MCInfant              173
MCToddler             173
MCPreschool           173
MFCCSA                180
MFCCInfant            180
MFCCToddler           180
MFCCPreschool         180
dtype: int64
```

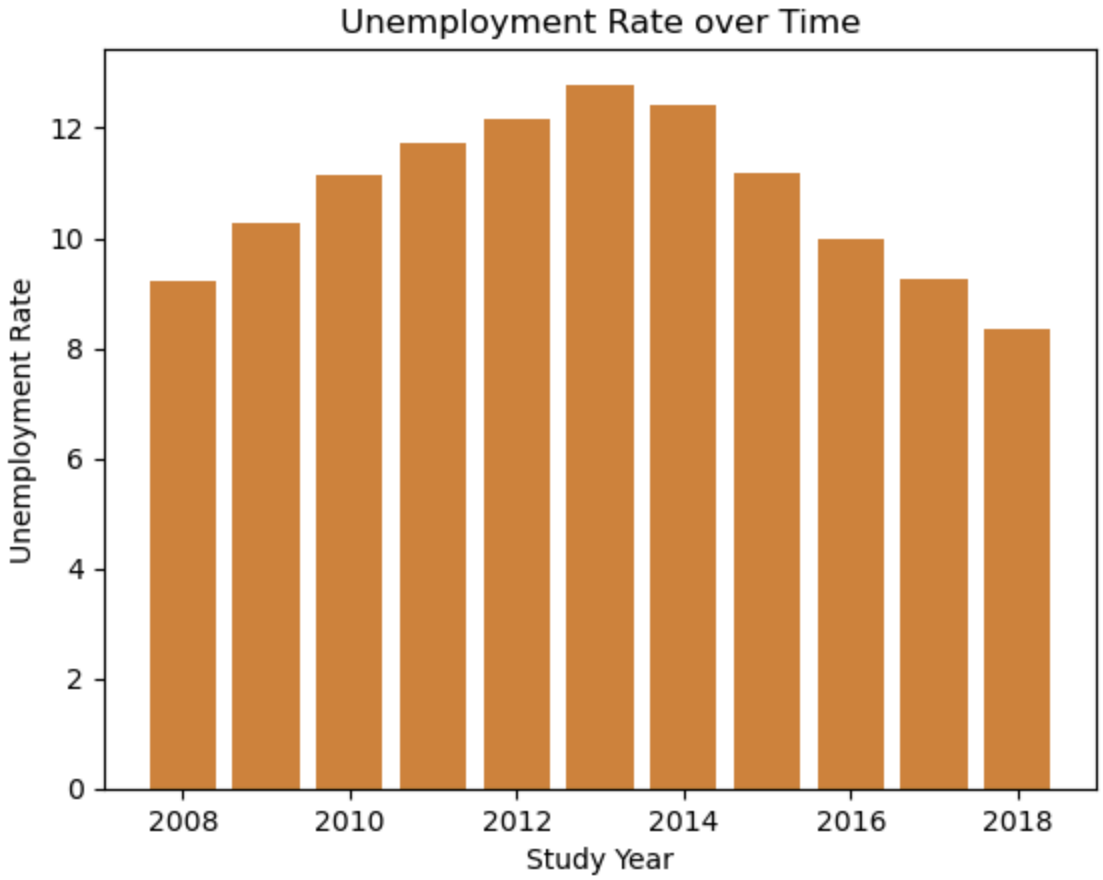
In [9]:
```
medians_by_state.to_csv('medians_by_state.csv', index=False)
```

Exploratory Visualizations

In [10]:
```
years = medians_by_state['StudyYear'].unique()
years
```

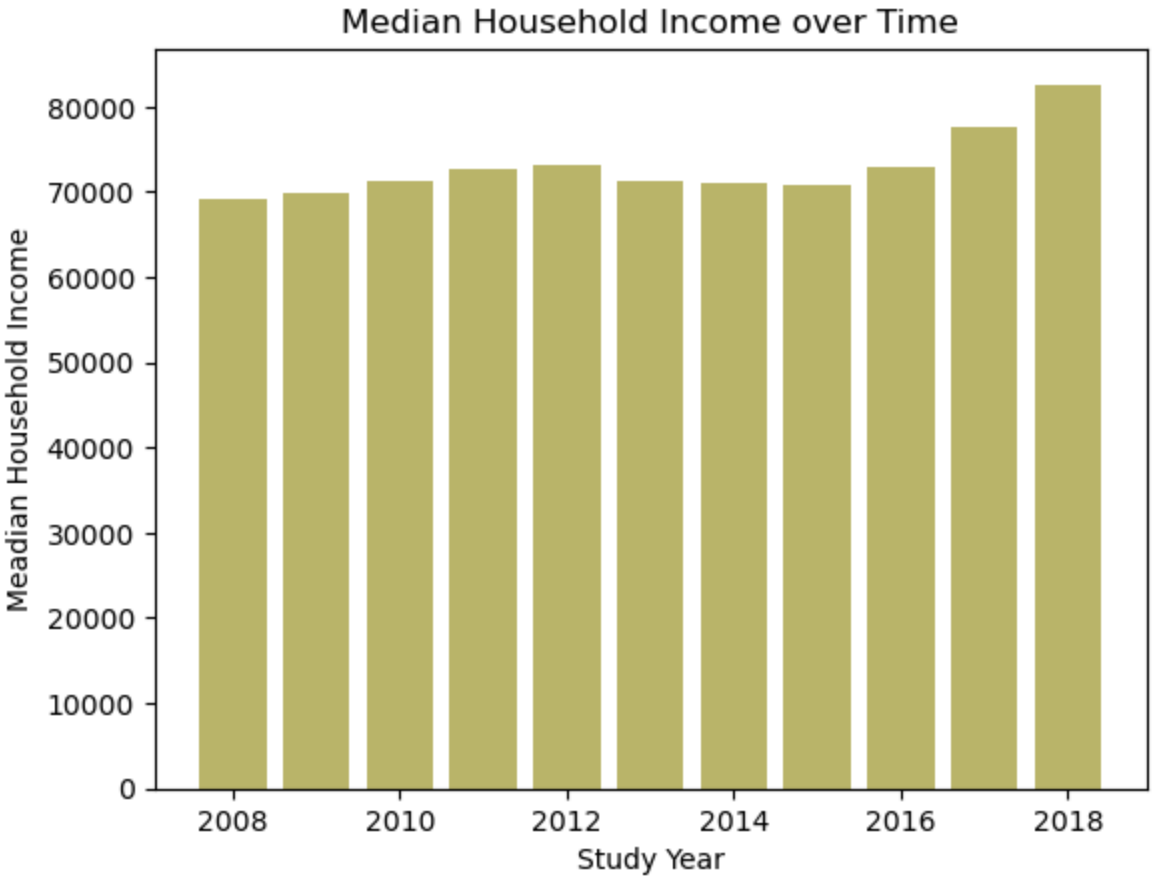
Out[10]:
```
array([2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018],
      dtype=int64)
```

In [11]:
```
# Plot unemployment rate for each year
plt.bar(x=medians_by_state['StudyYear'], height=medians_by_state['UNR_16'],color='peru')
plt.xlabel('Study Year')
plt.ylabel('Unemployment Rate')
plt.title('Unemployment Rate over Time')
plt.show()
```



The overall employment rate decreased from 2013 to 2018, with a significant reduction in unemployment rate by 2018. It would be interesting to see how childcare cost has changed since then.

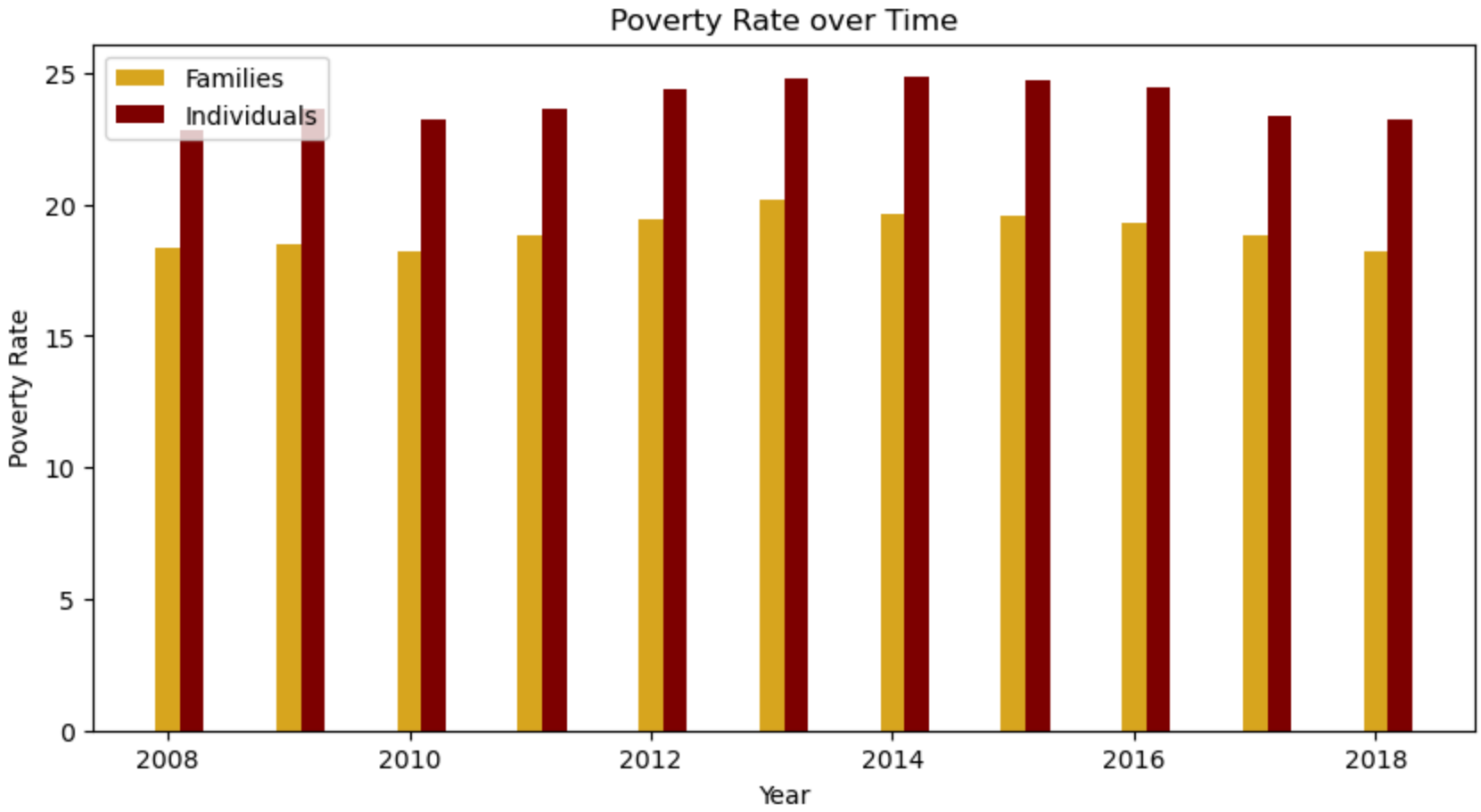
In [12]:
```
# Plot median household income
plt.bar(x=medians_by_state['StudyYear'], height=medians_by_state['MHI'], color='darkkhaki')
plt.xlabel('Study Year')
plt.ylabel('Meadian Household Income')
plt.title('Median Household Income over Time')
plt.show()
```



Over the 10 years of the study, median household income only increased about 10,000 dollars. Most of that increase was observed from 2015 to 2018; the same years unemployment dropped the most.

In [13]:
```
# Plot grouped bar plot of median poverty rates for individuals and families
x = medians_by_state['StudyYear']
y1 = medians_by_state['PR_F']
y2 = medians_by_state['PR_P']
width=0.2

# Plot data
plt.figure(figsize=(10,5))
plt.bar(x, y1, width, color='goldenrod')
plt.bar(x+0.2, y2, width, color='maroon')
```

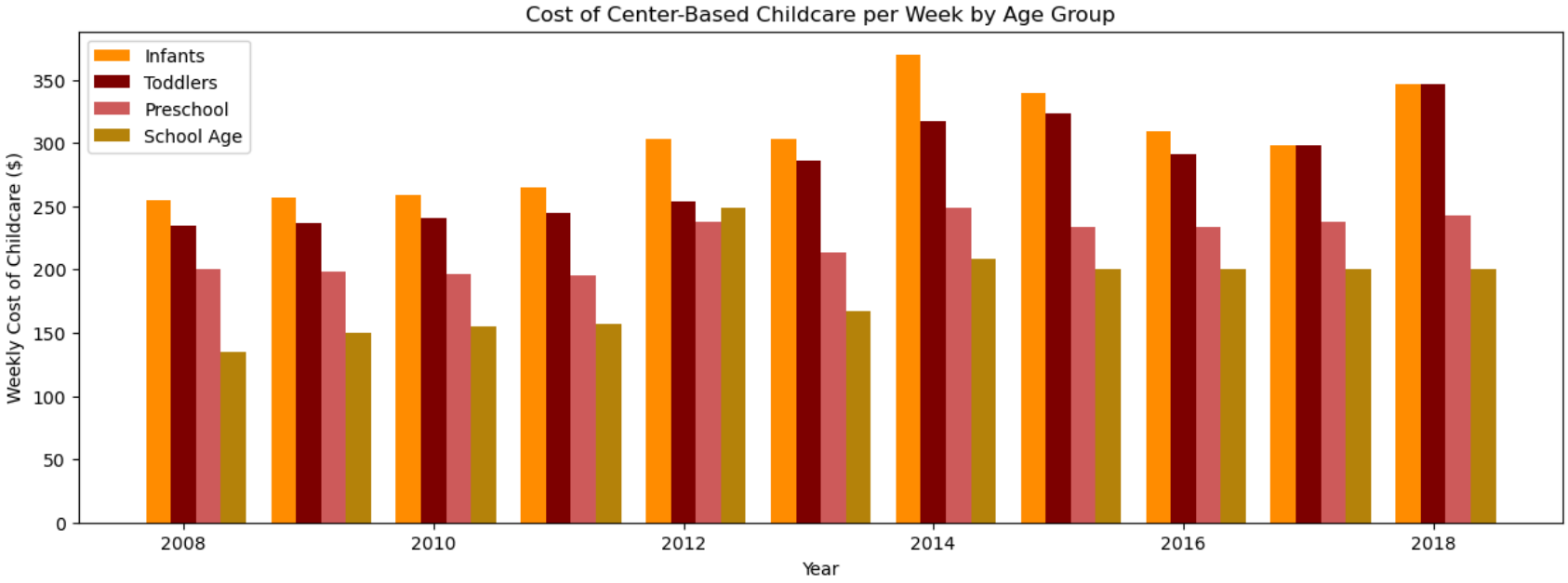
```
plt.legend(['Families', 'Individuals'], loc='upper left')
plt.xlabel('Year')
plt.ylabel('Poverty Rate')
plt.title('Poverty Rate over Time')
plt.show()
```



Throughout all years, poverty rate of individuals was higher than for families. For both groups, poverty rates were the highest in 2013- during peak unemployment.

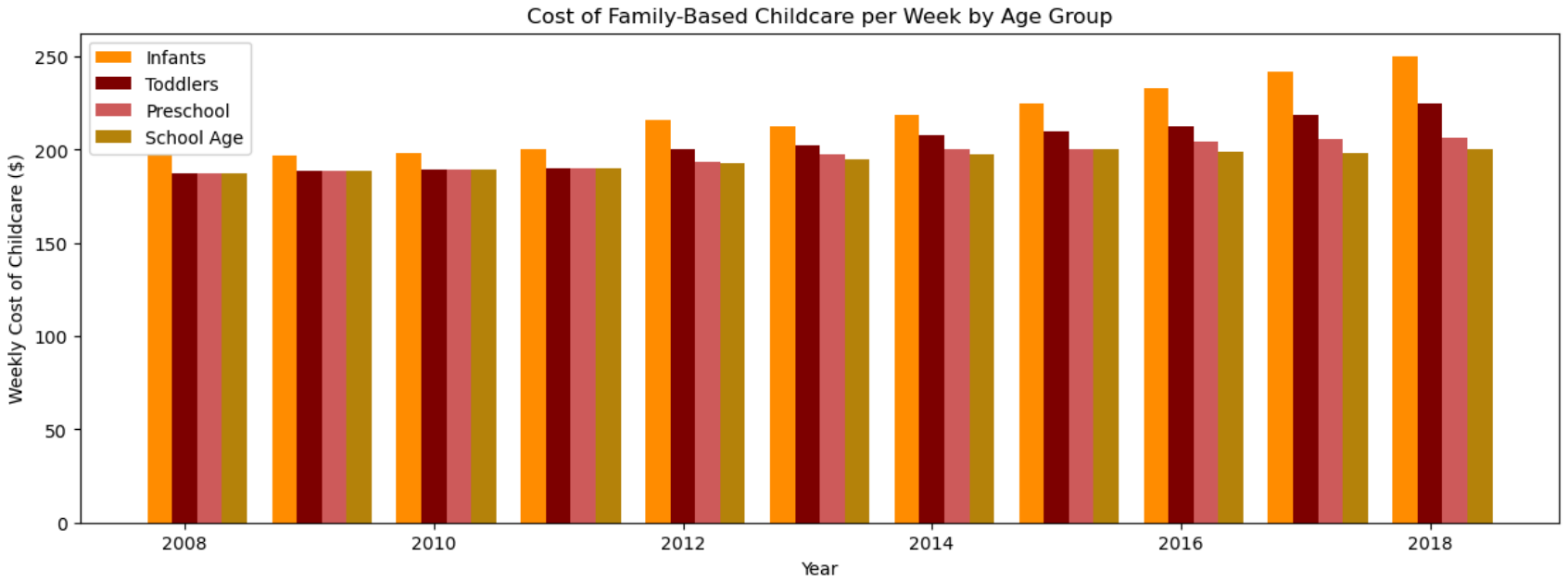
In [14]:
```
# Plot grouped bar plot of median cost of center-based childcare per week by age group
# create data
x = medians_by_state['StudyYear']
y1 = medians_by_state['MCInfant']
```

```
y2 = medians_by_state['MCToddler']
y3 = medians_by_state['MCPreschool']
y4 = medians_by_state['MCSA']
width = 0.2

# plot data in grouped manner of bar type
plt.figure(figsize=(15,5))
plt.bar(x-0.2, y1, width, color='darkorange')
plt.bar(x, y2, width, color='maroon')
plt.bar(x+0.2, y3, width, color='indianred')
plt.bar(x+0.4, y4, width, color='darkgoldenrod')
plt.legend(['Infants', 'Toddlers', 'Preschool', 'School Age'],loc='upper left')
plt.xlabel('Year')
plt.ylabel('Weekly Cost of Childcare ($)')
plt.title('Cost of Center-Based Childcare per Week by Age Group')
plt.show()
```



In all years except 2018, infant care was the most expensive. Care for school-aged children was the least expensive. Interestingly, in 2014 childcare for infants was more expensive than any other year. This is the first year unemployment began to decrease.

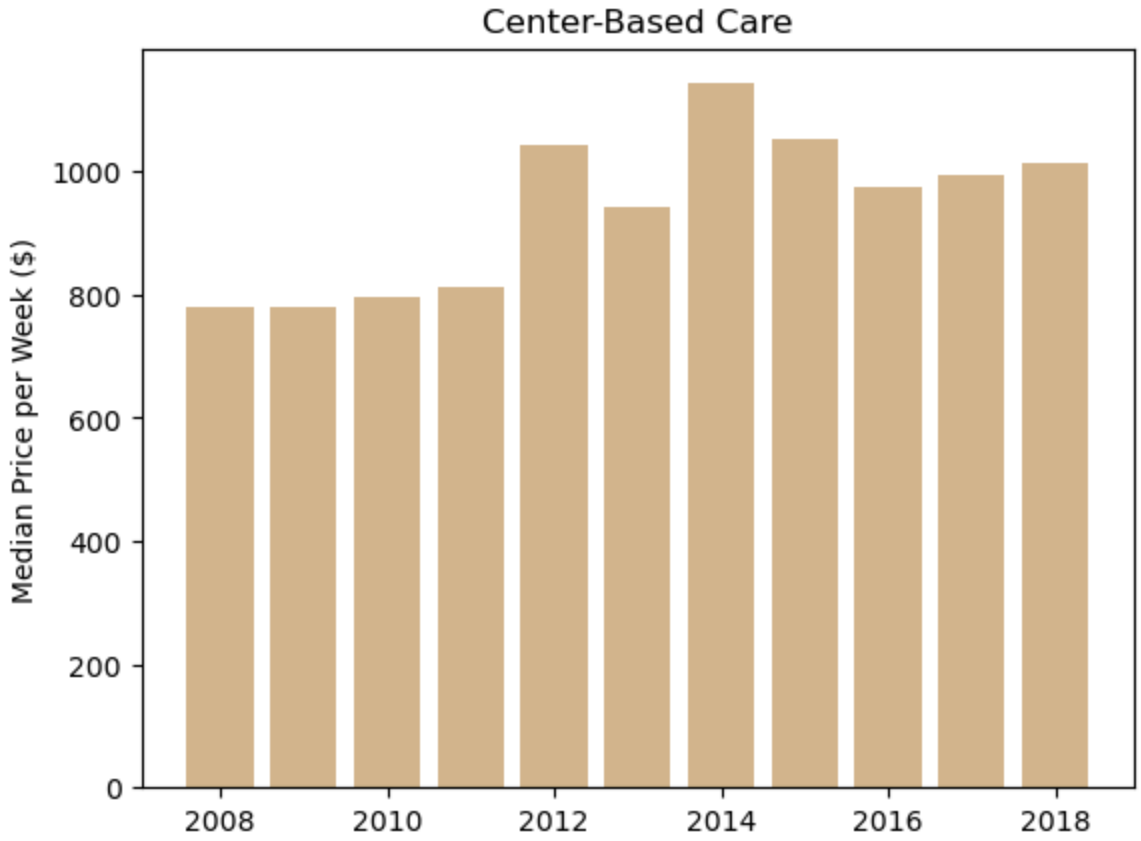
In [15]:
```
# Plot grouped bar plot of median cost of center-based childcare per week by age group
# create data
x = medians_by_state['StudyYear']
y1 = medians_by_state['MFCCInfant']
y2 = medians_by_state['MFCCToddler']
y3 = medians_by_state['MFCCPreschool']
y4 = medians_by_state['MFCCSA']
width = 0.2

# plot data in grouped manner of bar type
plt.figure(figsize=(15,5))
plt.bar(x-0.2, y1, width, color='darkorange')
plt.bar(x, y2, width, color='maroon')
plt.bar(x+0.2, y3, width, color='indianred')
plt.bar(x+0.4, y4, width, color='darkgoldenrod')
plt.legend(['Infants', 'Toddlers', 'Preschool', 'School Age'], loc='upper left')
plt.xlabel('Year')
plt.ylabel('Weekly Cost of Childcare ($)')
plt.title('Cost of Family-Based Childcare per Week by Age Group')
plt.show()
```

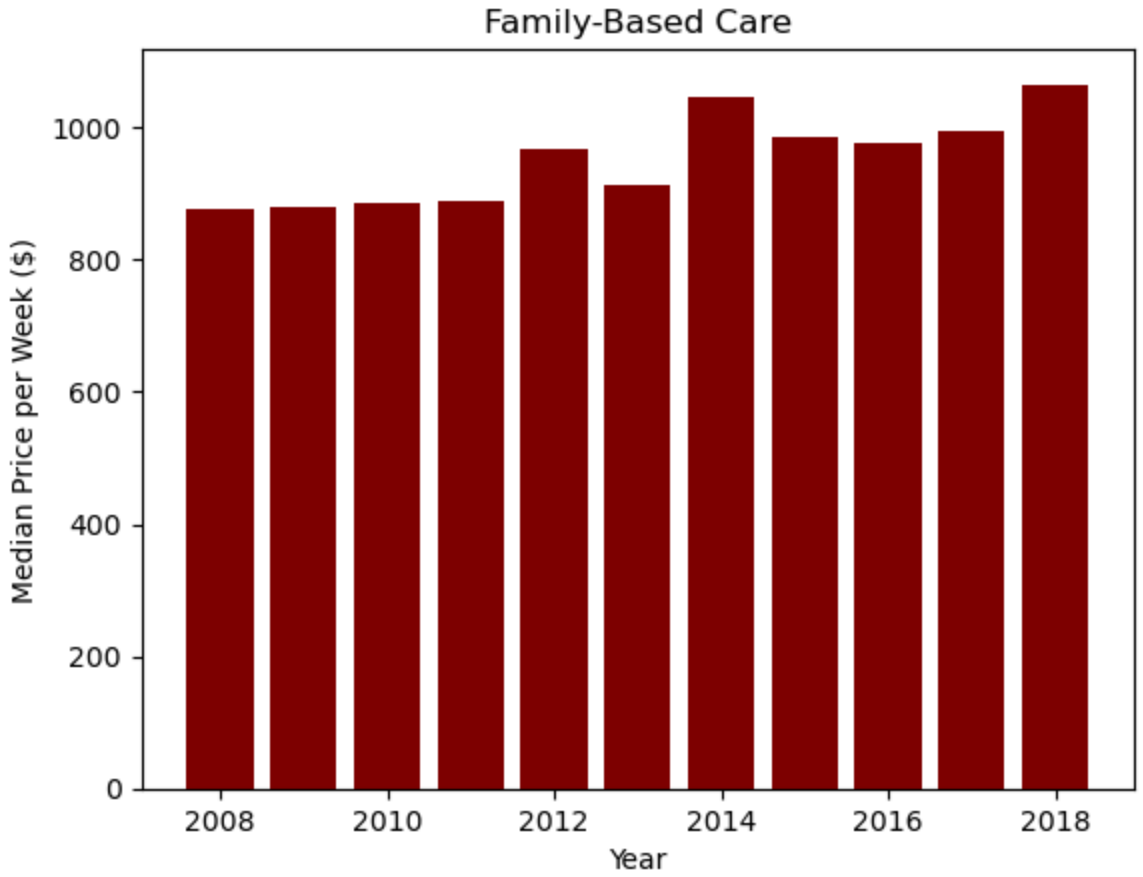


Family-based child care prices steadily increased across observation years. Like in center-based childcare, infant care is the most expensive.

In 2018, there was about a $100 a month price difference for infant childcare between center-based and family-based centers.

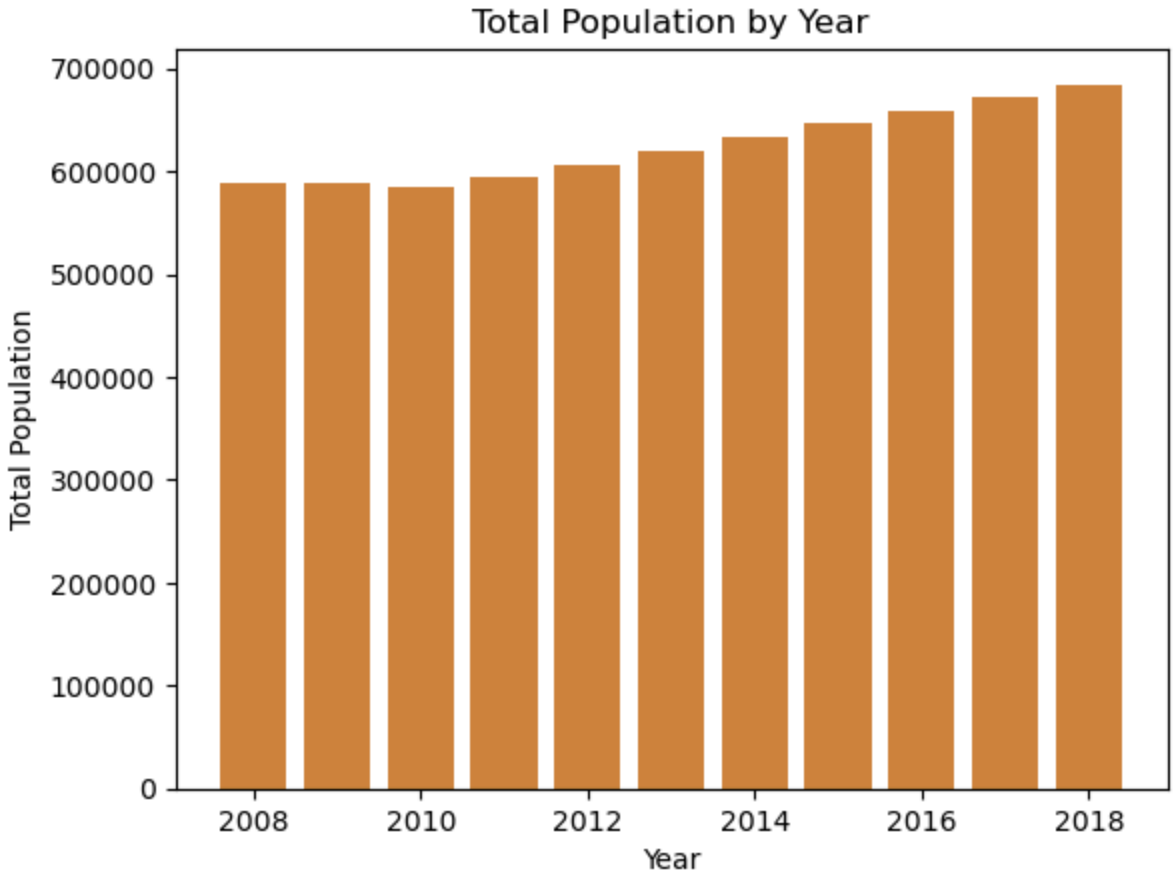
In [16]:
```
# create data for center based
x = medians_by_state['StudyYear']
y1_center = medians_by_state['MCInfant']
y2_center = medians_by_state['MCToddler']
y3_center = medians_by_state['MCPreschool']
y4_center = medians_by_state['MCSA']

# plot center-based facilites in stacked bar
plt.subplot(1,1,1)
plt.bar(x, y1_center, color='tan')
plt.bar(x, y2_center, bottom=y1_center, color='tan')
plt.bar(x, y3_center, bottom=(y1_center+y2_center), color='tan')
plt.bar(x, y4_center, bottom=(y1_center+y2_center+y3_center), color='tan')
plt.title('Year')
plt.ylabel('Median Price per Week ($)')
plt.title('Center-Based Care')
plt.show()

# Create data for family based
y1_family = medians_by_state['MFCCInfant']
y2_family = medians_by_state['MFCCToddler']
y3_family = medians_by_state['MFCCPreschool']
y4_family = medians_by_state['MFCCSA']

# plot family-based facilities in stacked bar
plt.subplot(1,1,1)
plt.bar(x, y1_family, color='maroon')
plt.bar(x, y2_family, bottom=y1_center, color='maroon')
plt.bar(x, y3_family, bottom=(y1_center+y2_center), color='maroon')
plt.bar(x, y4_family, bottom=(y1_center+y2_center+y3_center), color='maroon')
plt.xlabel('Year')
plt.ylabel('Median Price per Week ($)')
plt.title('Family-Based Care')
plt.show()
```


Center-Based Care
Median Price per Week ($)
0
200
400
600
800
1000
2008
2010
2012
2014
2016
2018



The total cost of childcare for center-based and family-based facilites is pretty similiar over time. In 2018, family-base childcare was more expensive than center-based care.

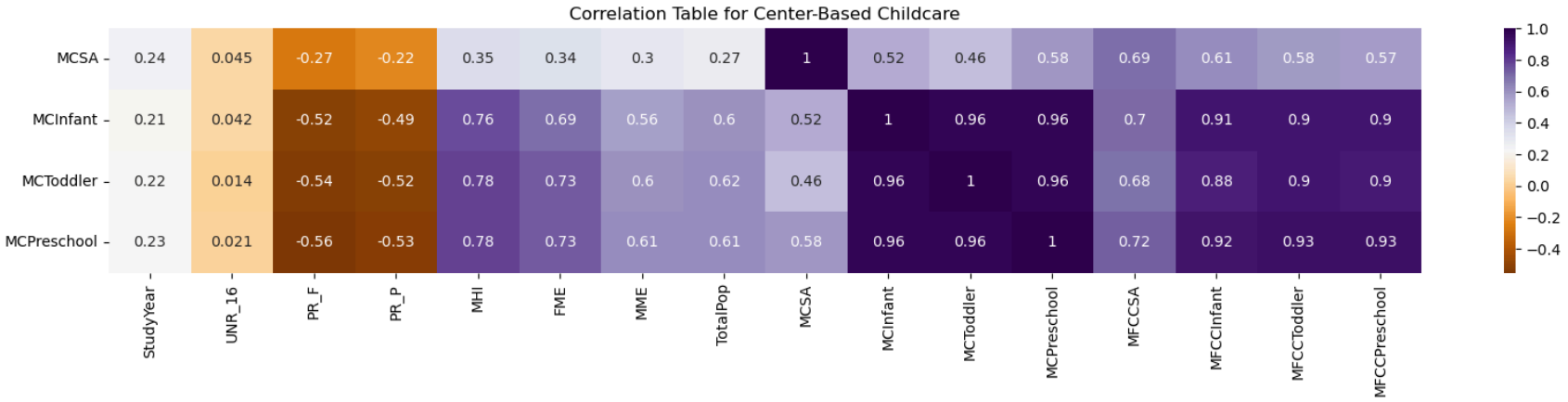
In [17]:
```
# Plot total popualation as a function of time
plt.bar(medians_by_state['StudyYear'], medians_by_state['TotalPop'],color='Peru')
plt.xlabel('Year')
plt.ylabel('Total Population')
plt.title('Total Population by Year')
plt.show()
```



Population steadily increased through observation years.

In [18]:
```
columns = medians_by_state.columns
columns

# Subset dataframe
medians_subset = medians_by_state.loc[:,['StudyYear', 'UNR_16', 'PR_F',
       'PR_P', 'MHI', 'FME', 'MME', 'TotalPop', 'MCSA', 'MCInfant',
       'MCToddler', 'MCPreschool', 'MFCCSA', 'MFCCInfant', 'MFCCToddler',
       'MFCCPreschool']]

# Plot correlation table for each feature variable with Diabetes_binary
plt.figure(figsize=(20,3))
sns.heatmap(medians_subset.corr().loc[['MCSA','MCInfant', 'MCToddler', 'MCPreschool'],:], annot=True, cmap='PuOr')
```

```
plt.title('Correlation Table for Center-Based Childcare')
plt.show()
```

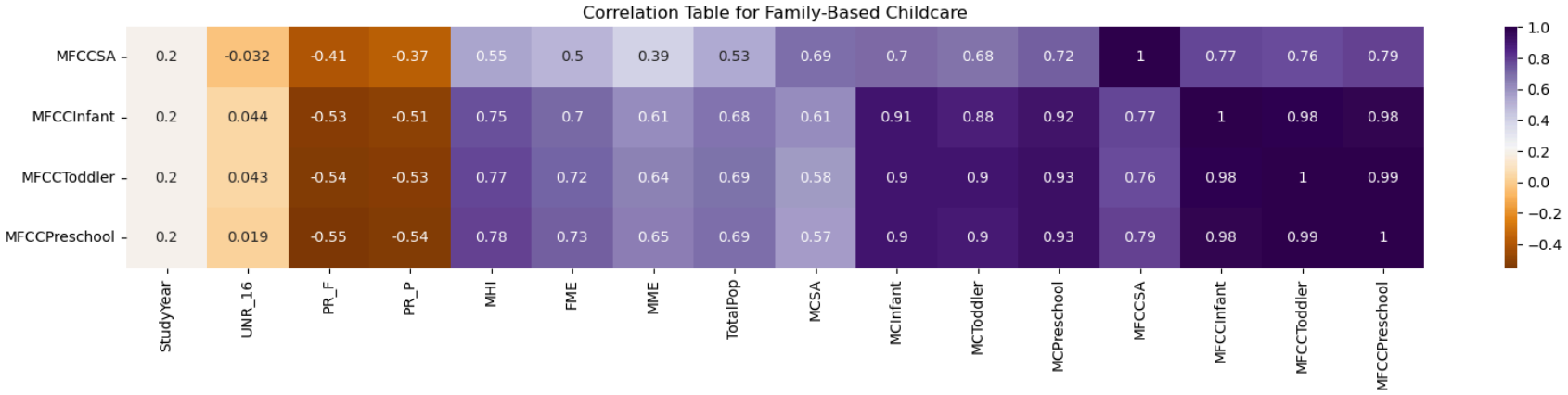


Center-Based Facilities: There are no strong relationships between cost of childcare for school aged children and any other variable. There is a strong relationship between mean houshold income and 3 of the child age care groups; infant, toddler, and preschool. There is a moderate relationship between female and male median earnings and 3 of the child age care groups: infant, toddler, and preschool.

There is also strong correlation between care-based center and family-based centers.

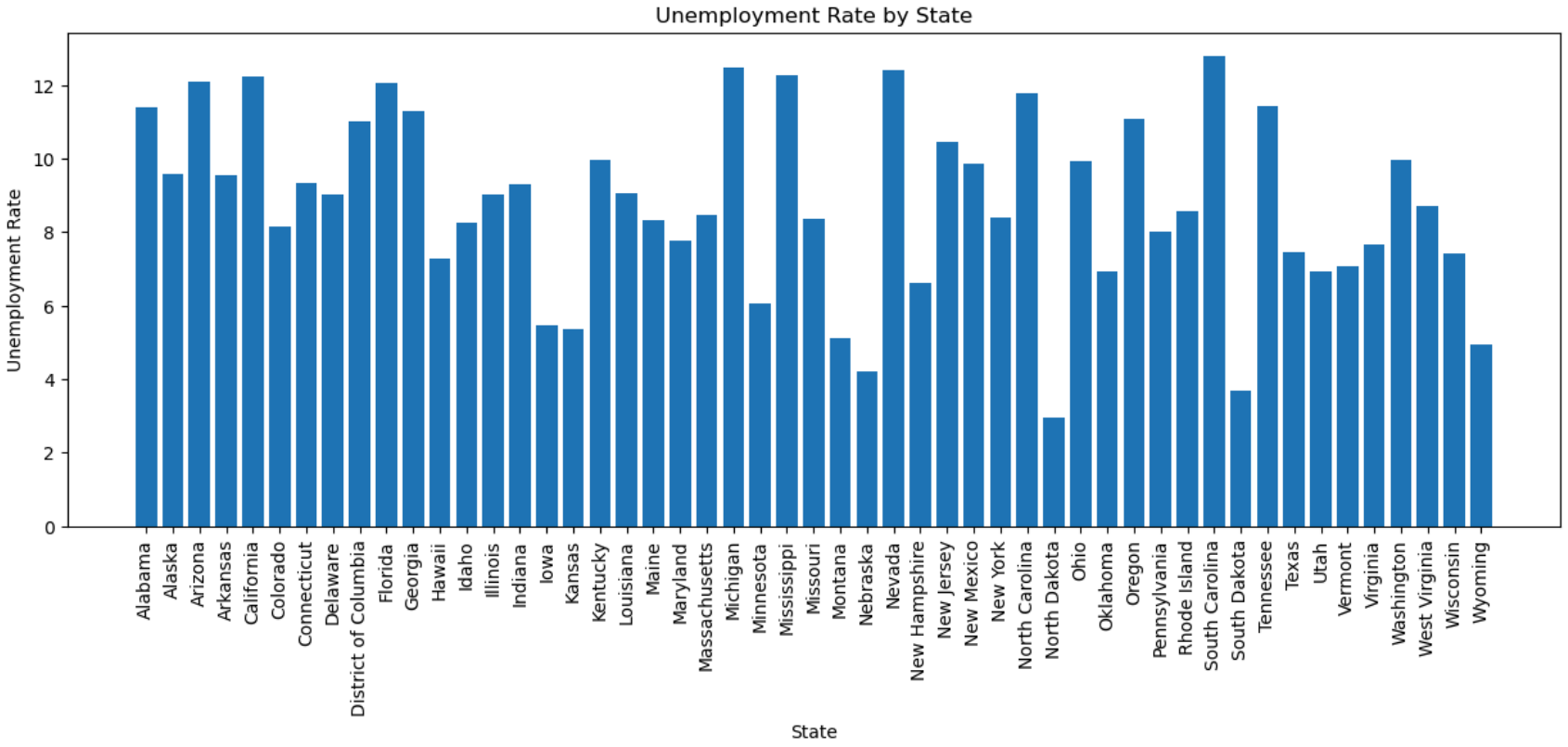
In [19]:
```
# Plot correlation table for each feature variable with Diabetes_binary
plt.figure(figsize=(20,3))
sns.heatmap(medians_subset.corr().loc[['MFCCSA','MFCCInfant', 'MFCCToddler', 'MFCCPreschool'],:], annot=True, cmap='
plt.title('Correlation Table for Family-Based Childcare')
```

Out[19]:
```
Text(0.5, 1.0, 'Correlation Table for Family-Based Childcare')
```

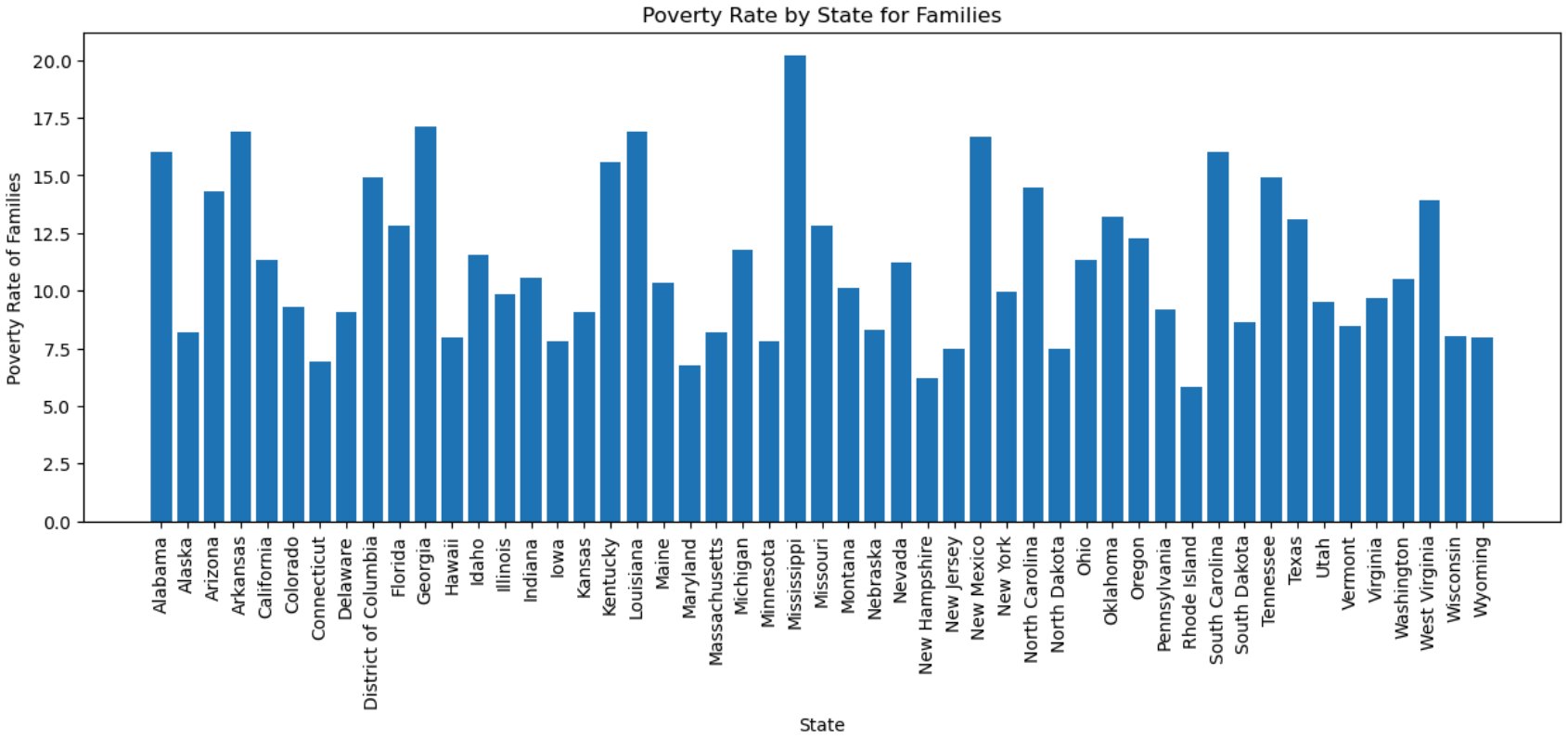


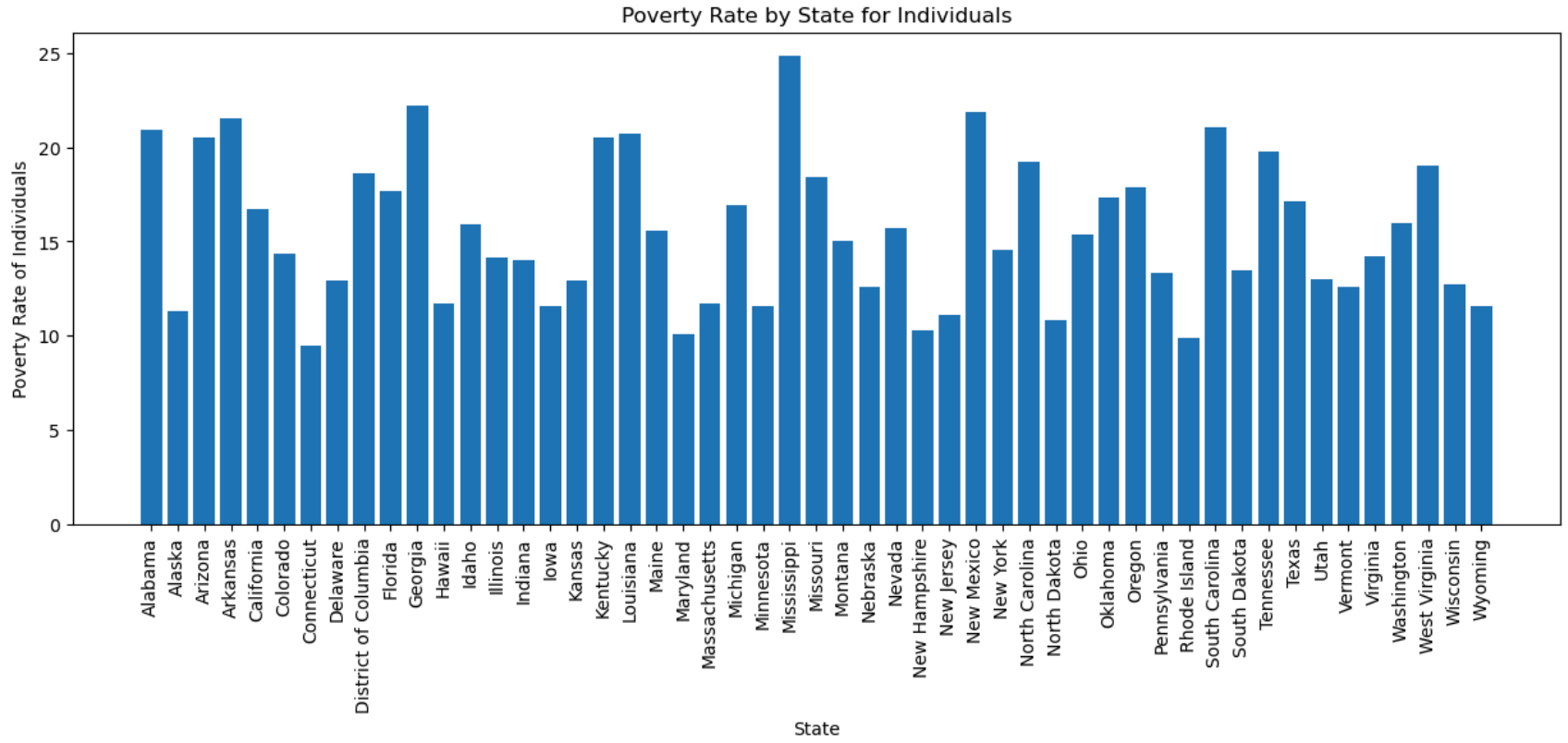
Family-Based Facilities: There are no strong relationships between cost of childcare for school aged children and any other variables. Like in center-based facilities, there is a strong relationship between mean household income and childcare prices for infants, toddlers, and preschoolers. There is also a moderate realtionship between childcare prices for those age brakets and female median earnings, male median earnings, and total population.

In [20]:
```
# Plot unemployment rate by state
plt.figure(figsize=(15,5))
plt.bar(medians_by_state['State_Name'], medians_by_state['UNR_16'])
plt.xticks(rotation='vertical')
plt.xlabel('State')
plt.ylabel('Unemployment Rate')
plt.title('Unemployment Rate by State')
plt.show()
```



In [21]:
```
# Plot poverty rate by state
plt.figure(figsize=(15,5))
plt.bar(medians_by_state['State_Name'], medians_by_state['PR_F'])
plt.xticks(rotation='vertical')
plt.xlabel('State')
plt.ylabel('Poverty Rate of Families')
plt.title('Poverty Rate by State for Families')
plt.show()

plt.figure(figsize=(15,5))
plt.bar(medians_by_state['State_Name'], medians_by_state['PR_P'])
plt.xticks(rotation='vertical')
plt.xlabel('State')
plt.ylabel('Poverty Rate of Individuals')
plt.title('Poverty Rate by State for Individuals')
plt.show()
```


Poverty Rate by State for Families
Poverty Rate of Families
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
Alabama
Alaska
Arizona
Arkansas
California
Colorado
Connecticut
Delaware
District of Columbia
Florida
Georgia
Hawaii
Idaho
Illinois
Indiana
Iowa
Kansas
Kentucky
Louisiana
Maine
Maryland
Massachusetts
Michigan
Minnesota
Mississippi
Missouri
Montana
Nebraska
Nevada
New Hampshire
New Jersey
New Mexico
New York
North Carolina
North Dakota
Ohio
Oklahoma
Oregon
Pennsylvania
Rhode Island
South Carolina
South Dakota
Tennessee
Texas
Utah
Vermont
Virginia
Washington
West Virginia
Wisconsin
Wyoming
State



In [22]:
```
medians_by_state.set_index(['State_Name'], inplace=True)
medians_by_state.sort_values('PR_F', ascending=False)
medians_by_state
```

Out[22]:

State_Name	State_Abbreviation	StudyYear	UNR_16	PR_F	PR_P	MHI	FME	MME	TotalPop	MCSA	MCInfant	MC
Alabama	AL	2008	7.06	13.9	18.1	34952.33	17938.0	30016.11	34635.0	83.76	98.77	
Alabama	AL	2009	8.17	14.0	18.1	35137.00	18225.0	29594.00	34704.0	86.17	100.11	
Alabama	AL	2010	9.51	14.7	18.7	36077.00	18879.0	29589.00	34579.0	88.58	101.44	
Alabama	AL	2011	10.24	15.4	19.7	37161.00	19989.0	30288.00	34375.0	90.99	102.78	
Alabama	AL	2012	11.00	15.4	19.8	37059.00	20084.0	30829.00	34228.0	93.10	104.11	
...	...	...	...	...	...	...	...	...	...	...	...	
Wyoming	WY	2014	4.30	7.4	10.5	56318.00	22221.0	40721.00	15795.0	95.61	154.38	
Wyoming	WY	2015	4.40	7.5	10.7	56569.00	21941.0	40851.00	15739.0	100.68	149.12	
Wyoming	WY	2016	4.51	7.5	11.0	55640.00	21942.0	40254.00	15696.0	105.67	143.14	
Wyoming	WY	2017	4.60	7.5	10.5	57644.00	23761.0	40395.00	15689.0	110.47	136.87	
Wyoming	WY	2018	4.34	8.0	11.6	58521.00	24679.0	43136.00	15477.0	115.25	137.57	

561 rows × 17 columns

Focus on Arkansas

In [23]:
```
arkansas_prices = childcare_prices.loc[childcare_prices['State_Name'] == 'Arkansas']
arkansas_prices
```

Out[23]:

	State_Name	State_Abbreviation	County_Name	County_FIPS_Code	StudyYear	UNR_16	FUNR_16	MUNR_16	UNR_20to64
1220	Arkansas	AR	Arkansas County	5001	2008	7.38	6.80	7.93	6.5
1221	Arkansas	AR	Arkansas County	5001	2009	7.95	7.31	8.51	7.2
1222	Arkansas	AR	Arkansas County	5001	2010	7.36	7.32	7.40	6.4
1223	Arkansas	AR	Arkansas County	5001	2011	7.59	7.81	7.39	6.8
1224	Arkansas	AR	Arkansas County	5001	2012	7.26	8.27	6.33	7.0
...	...	...	...	...	...	...	...	...	...
2040	Arkansas	AR	Yell County	5149	2014	8.10	5.64	10.23	7.3
2041	Arkansas	AR	Yell County	5149	2015	8.28	6.98	9.46	7.3
2042	Arkansas	AR	Yell County	5149	2016	6.75	5.62	7.73	6.4
2043	Arkansas	AR	Yell County	5149	2017	6.32	6.21	6.40	5.8
2044	Arkansas	AR	Yell County	5149	2018	5.55	5.64	5.48	5.1

825 rows × 227 columns

In [24]:
```
# Checking number of counties in Arkansas
arkansas_prices['County_Name'].nunique()
```

Out[24]: 75

In this scenario, I will be working for Walmart which has it's corporate headquarters in Bentonville, Arkansas.

It would be useful to investigate median household income, median earnings, etc of Arkansas versus the national averages.

In [25]:
```
# Find median key metrics of Arkansas vs the rest of the country
print('------Median Household Income------')
print('Country: ',medians_by_state['MHI'].median())
print('Arkansas: ',arkansas_prices['MHI'].median())

print('\n------Unemployment Rate------')
print('Country: ',medians_by_state['UNR_16'].median())
print('Arkansas: ',arkansas_prices['UNR_16'].median())

print('\n------Povert Rate of Families------')
print('Country: ',medians_by_state['PR_F'].median())
print('Arkansas: ',arkansas_prices['PR_F'].median())

print('\n------Poverty Rate of Individuals------')
print('Country: ',medians_by_state['PR_P'].median())
print('Arkansas: ',arkansas_prices['PR_P'].median())

print('\n------Female Mean Earnings------')
print('Country: ',medians_by_state['FME'].median())
print('Arkansas: ',arkansas_prices['FME'].median())

print('\n------Male Mean Earnings------')
print('Country: ',medians_by_state['MME'].median())
print('Arkansas: ',arkansas_prices['MME'].median())
```

```
------Median Household Income------
Country:  46872.0
Arkansas:  35732.0

------Unemployment Rate------
Country:  7.04
Arkansas:  8.04

------Povert Rate of Families------
Country:  9.5
Arkansas:  15.2

------Poverty Rate of Individuals------
Country:  13.81
Arkansas:  20.1

------Female Mean Earnings------
Country:  21562.0
Arkansas:  19160.0

------Male Mean Earnings------
Country:  32671.5
Arkansas:  28568.0
```

Arkansas has a lower household income and median earnings that the rest of the country. Additionally, Arkansas has a higher unemployment rate and poverty rate than the rest of the country.

In [26]:
```
# Calculate median prices of center-based childcare in Arkansas versus the rest of the country
print('\n------Center-Based Childcare of School-Aged Children------')
print('Country: ',medians_by_state['MCSA'].median())
print('Arkansas: ',arkansas_prices['MCSA'].median())

print('\n------Center-Based Childcare of Infants------')
print('Country: ',medians_by_state['MCInfant'].median())
print('Arkansas: ',arkansas_prices['MCInfant'].median())

print('\n------Center-Based Childcare of Toddlers------')
print('Country: ',medians_by_state['MCToddler'].median())
print('Arkansas: ',arkansas_prices['MCToddler'].median())

print('\n------Center-Based Childcare of Preschoolers------')
```

```
print('Country: ',medians_by_state['MCPreschool'].median())
print('Arkansas: ',arkansas_prices['MCPreschool'].median())
```

```
------Center-Based Childcare of School-Aged Children------
Country:  98.325
Arkansas:  77.5

------Center-Based Childcare of Infants------
Country:  148.765
Arkansas:  97.5

------Center-Based Childcare of Toddlers------
Country:  133.235
Arkansas:  95.75

------Center-Based Childcare of Preschoolers------
Country:  123.33
Arkansas:  87.5
```

In [27]:
```
# Calculate median prices of family-based childcare in Arkansas versus the rest of the country
print('\n------Family-Based Childcare of School-Aged Children------')
print('Country: ',medians_by_state['MFCCSA'].median())
print('Arkansas: ',arkansas_prices['MFCCSA'].median())

print('\n------Family-Based Childcare of Infants------')
print('Country: ',medians_by_state['MFCCInfant'].median())
print('Arkansas: ',arkansas_prices['MFCCInfant'].median())

print('\n------Family-Based Childcare of Toddlers------')
print('Country: ',medians_by_state['MFCCToddler'].median())
print('Arkansas: ',arkansas_prices['MFCCToddler'].median())

print('\n------Family-Based Childcare of Preschoolers------')
print('Country: ',medians_by_state['MFCCPreschool'].median())
print('Arkansas: ',arkansas_prices['MFCCPreschool'].median())
```

```
------Family-Based Childcare of School-Aged Children------
Country:  93.33500000000001
Arkansas:  78.75

------Family-Based Childcare of Infants------
Country:  115.38
Arkansas:  87.5

------Family-Based Childcare of Toddlers------
Country:  109.95
Arkansas:  85.0

------Family-Based Childcare of Preschoolers------
Country:  105.63
Arkansas:  81.25
```

Childcare for both center-based and family-based facilities is less expensive in Arkansas than the rest of the country.

Focus on County

There is a large wealth gap in Arkansas. To correctly represent metrics it is important to focus on the county level

In [34]:
```
# Filter for Benton and Washington counties
arkansas_counties = arkansas_prices[(arkansas_prices['County_Name'] == 'Benton County') |
                                    (arkansas_prices['County_Name'] == 'Washington County')]
arkansas_counties.head()
```

Out[34]:

	State_Name	State_Abbreviation	County_Name	County_FIPS_Code	StudyYear	UNR_16	FUNR_16	MUNR_16	UNR_20to64
1253	Arkansas	AR	Benton County	5007	2008	4.64	3.96	5.21	4.1
1254	Arkansas	AR	Benton County	5007	2009	5.46	5.54	5.39	5.1
1255	Arkansas	AR	Benton County	5007	2010	5.33	4.54	6.00	4.8
1256	Arkansas	AR	Benton County	5007	2011	5.53	4.37	6.52	4.9
1257	Arkansas	AR	Benton County	5007	2012	5.25	3.99	6.30	4.6

5 rows × 227 columns

In [75]:
```
# Find average male and female earnings
print('\n------Female Median Earnings------')
print('Country: ',arkansas_counties['FME'].median())

print('\n------Male Median Earnings------')
print('Country: ',arkansas_counties['MME'].median())
```

```
------Female Median Earnings------
Country:  23518.0

------Male Median Earnings------
Country:  31659.0
```

In [78]:
```
median_earnings = ((23518+31659)/2)/12
median_earnings
```

Out[78]: 2299.0416666666665

In [58]:
```
# Calculate median prices of center-based childcare in Benton & Washington County
center_school_aged = arkansas_counties['MCSA'].median()
```

```
center_infant =arkansas_counties['MCInfant'].median()
center_toddler = arkansas_counties['MCToddler'].median()
center_preschool = arkansas_counties['MCPreschool'].median()

print('\n------Center-Based Childcare------')
print('School-Aged:', center_school_aged)
print('Infant:', center_infant)
print('Toddler:', center_toddler)
print('Preschool:', center_preschool)

# Calculate median prices of family-based childcare in Arkansas versus the rest of the country
family_school_aged = arkansas_counties['MFCCSA'].median()
family_infant = arkansas_counties['MFCCInfant'].median()
family_toddler = arkansas_counties['MFCCToddler'].median()
family_preschool = arkansas_counties['MFCCPreschool'].median()

print('\n------Family-Based Childcare------')
print('School-Aged:', family_school_aged)
print('Infant:', family_infant)
print('Toddler:', family_toddler)
print('Preschool:', family_preschool)
```

```
------Center-Based Childcare------
School-Aged: 100.75
Infant: 128.75
Toddler: 119.59
Preschool: 108.75

------Family-Based Childcare------
School-Aged: 86.25
Infant: 102.5
Toddler: 96.25
Preschool: 91.25
```

In [77]:
```
# Calulcate median
median_price_care = sum([center_school_aged+center_infant,center_toddler+center_preschool])/4
median_price_care
```

Out[77]: 114.46000000000001

In [79]:
```
# Calculate average % spent on childcare
(median_price_care / median_earnings)*100
```

Out[79]: 4.978596154194683

In [ ]:
```
# Find average cost of childcare for infants per state
```