Methodology

To measure the growth of home schooling during the pandemic, The Washington Post collected home-school student counts from 6,738 school districts. Together with students from The Washington Post Investigative Reporting Workshop practicum at American University, reporters trawled state websites, contacted education officials in all 50 states and the District of Columbia and submitted multiple public records requests for an annual count of home-schoolers from the 2017-18 school year through 2022-23. The Post ultimately collected data for all public school districts in 29 states and D.C. In all, The Post gathered data from states representing 61% of the American school-age population.

The Post made every effort to capture all legal ways to home-school, which vary by state. However, data on home schools established by certain methods, such as registering one's home-school as a private school, are tracked by some states but not others. That means The Post's tally is almost certainly an undercount, even in the states from which it gathered data. For instance, Wisconsin and Georgia only provided The Post with tallies of home-schoolers who had submitted required forms electronically. In Kentucky, some districts incorrectly reported zero or one home-schooled students in certain years, which a state education official attributed to an unclear form. The Post excluded those enrollment figures from its analysis. In California, which does not explicitly permit home schooling, many parents operate home-based private schools. The California Department of Education characterizes private schools with five or fewer students as home schools. In Louisiana, many home schools operate as nonpublic schools not seeking accreditation; The Post counted such schools with five or fewer students as home schools as well.

The statewide numbers are not always equivalent to the sum of all district totals in a state. Some states suppress district-level counts of home schoolers below a certain threshold. In Maine, the threshold is 5; in New Mexico, 6; in Mississippi, Ohio and Tennessee, 10; in Wisconsin before 2020-21, 5; and in Wisconsin from 2021-22 on, 20. The Post marked such suppressions as NA within its data. In addition, New Hampshire collects separate data on students who enter home schooling from schools run by the state department of education or from private schools; these additional students are reflected in state data but not district data.

The Post used a variety of methods to match each school district name to an NCES district id. However, this was not always possible. In Georgia, families self-report their school district on home-schooling forms; some report programs which are not school districts, and therefore have no corresponding NCES id. In California, families were only required to report county and school district beginning in 2020-21; in addition, district mergers and name changes mean that some districts could not be matched with NCES ids. In Arkansas, 12 home-schoolers reported no school district in 2022-23. The Post marked all such home-schoolers as being within NA districts.

The Rise of Home Schooling Data from The Post's analysis of home-schooling enrollment across the US

This repository shares data hand-collected by The Washington Post from individual school districts and states as a whole regarding home-school enrollment from 2017-18 through 2022-23

In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pylab as plt
import seaborn as sns
plt.style.use('ggplot')
```

home_school_state

In [2]:
```
data_state = pd.read_csv('home_school_state.csv')
data_state=data_state.sort_values(by=['state'],ascending=True)
data_state
```

Out[2]:

	state	year	homeschool_students
51	ARKANSAS	2018-19	21959.0
84	ARKANSAS	2019-20	22249.0
183	ARKANSAS	2022-23	29762.0
117	ARKANSAS	2020-21	30267.0
18	ARKANSAS	2017-18	20331.0
...	...	...	...
116	WYOMING	2020-21	3884.0
83	WYOMING	2019-20	2585.0
50	WYOMING	2018-19	1797.0
17	WYOMING	2017-18	2572.0
182	WYOMING	2022-23	3769.0

198 rows × 3 columns

In [3]:
```
data_state.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 198 entries, 51 to 182
Data columns (total 3 columns):
 #   Column               Non-Null Count  Dtype
---  ------               --------------  -----
 0   state                198 non-null    object
 1   year                 198 non-null    object
 2   homeschool_students  192 non-null    float64
dtypes: float64(1), object(2)
memory usage: 6.2+ KB
```

In [4]:
```
data_state.describe()
```

Out[4]:

	homeschool_students
count	192.000000
mean	23392.354167
std	26237.918687
min	389.000000
25%	4884.000000
50%	17765.000000
75%	30220.500000
max	154289.000000

In [5]:
```
data_state.shape
```

Out[5]:
```
(198, 3)
```

In [6]:
```
data_state=data_state.dropna()
data_state
```

Out[6]:

	state	year	homeschool_students
51	ARKANSAS	2018-19	21959.0
84	ARKANSAS	2019-20	22249.0
183	ARKANSAS	2022-23	29762.0
117	ARKANSAS	2020-21	30267.0
18	ARKANSAS	2017-18	20331.0
...	...	...	...
116	WYOMING	2020-21	3884.0
83	WYOMING	2019-20	2585.0
50	WYOMING	2018-19	1797.0
17	WYOMING	2017-18	2572.0
182	WYOMING	2022-23	3769.0

192 rows × 3 columns

In [7]:
```
data_state.shape
```

Out[7]:
```
(192, 3)
```

In [8]:
```
def get_state(state):
    return data_state[data_state['state']== str(state)]
```

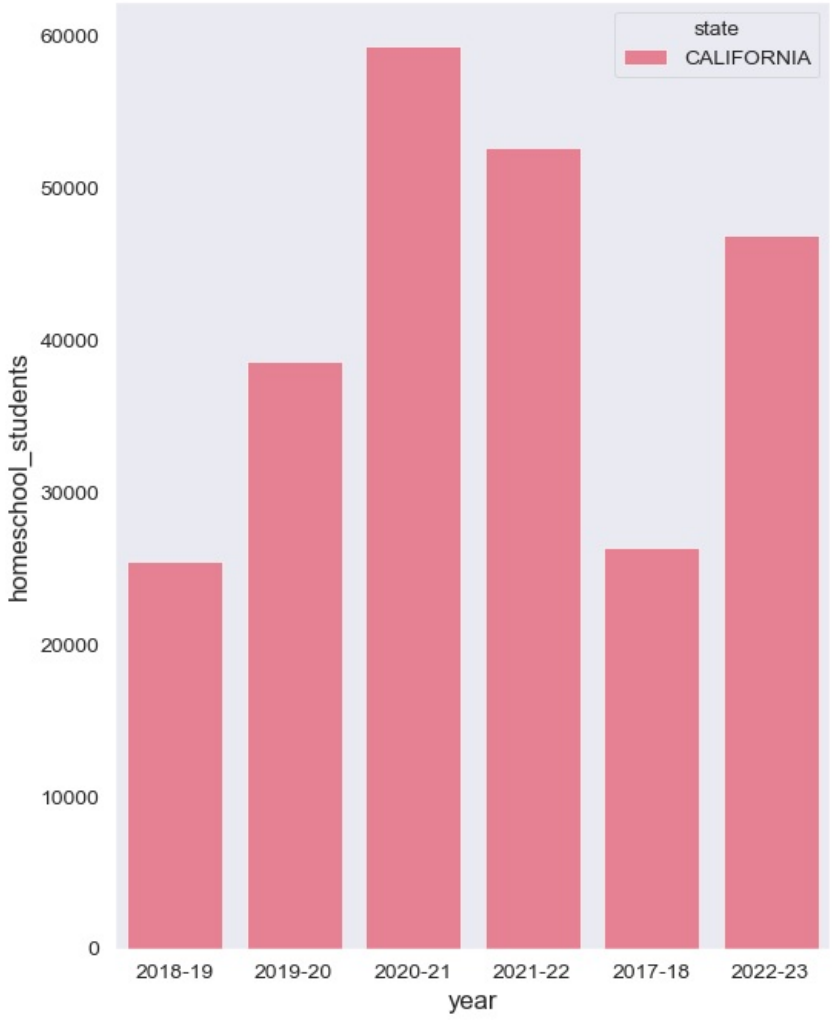
In [9]:
```
get_state('CALIFORNIA')
```

Out[9]:

	state	year	homeschool_students
33	CALIFORNIA	2018-19	25423.0
66	CALIFORNIA	2019-20	38532.0
99	CALIFORNIA	2020-21	59275.0
132	CALIFORNIA	2021-22	52623.0
0	CALIFORNIA	2017-18	26345.0
165	CALIFORNIA	2022-23	46814.0

In [10]:
```
sns.set_style("ticks")
sns.set_style("dark")
sns.set_palette("husl")
plt.figure(figsize=(6, 8))
sns.barplot(x="year", y="homeschool_students",hue="state" ,data=get_state('CALIFORNIA'))
```

```
plt.show()
```



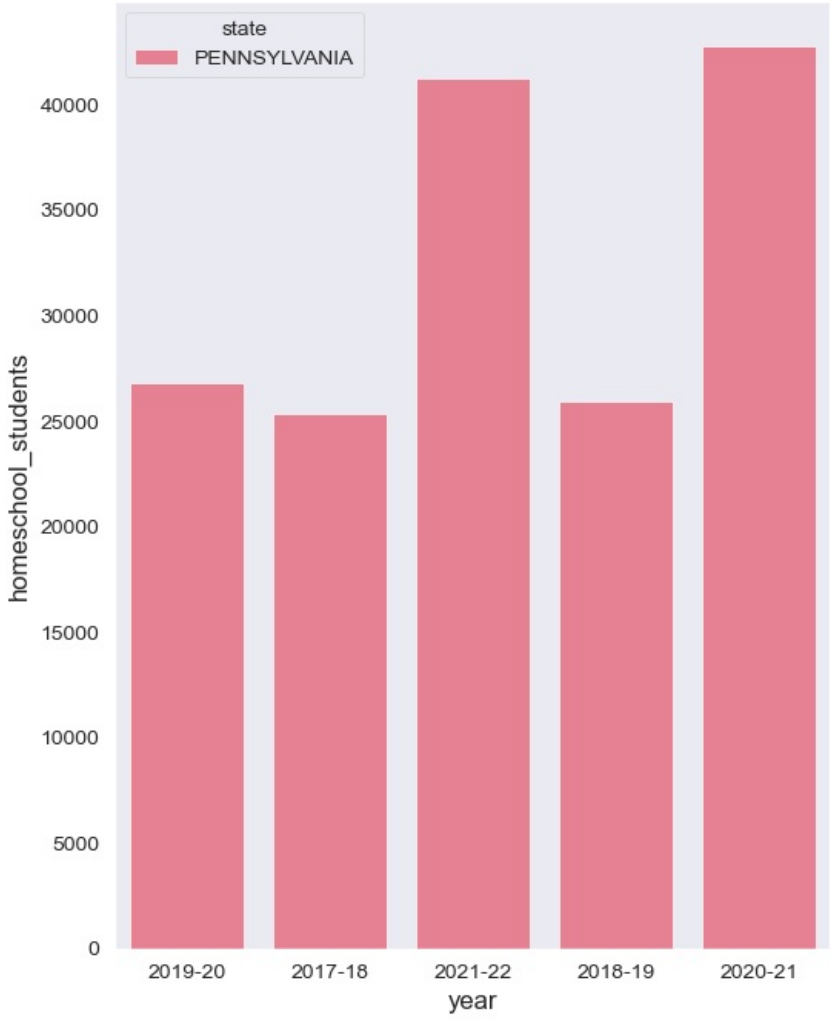
Three states — Pennsylvania, Rhode Island and Tennessee — have not published the number of home-schoolers in 2022-23, and Maine only shared district-level data starting with the 2020-21 school year. In seven states, The Post was unable to obtain usable home-school enrollment figures: In Arizona, Nevada and Oregon, only new home-school registrations are tracked annually at the district level; in North Carolina, home-school registration rolls are not regularly purged as students age out of the system; and in West Virginia, Utah and Alabama, annual enrollment data is unavailable. Eleven additional states do not require any notice when families decide to home-school their children, so enrollment figures in those states are also unavailable. Finally, Montana, Vermont and Nebraska collect data at a county level, not a district level, so there is no district data available - only statewide figures.

In [11]:
```
get_state('PENNSYLVANIA')
```

Out[11]:

	state	year	homeschool_students
95	PENNSYLVANIA	2019-20	26810.0
29	PENNSYLVANIA	2017-18	25378.0
161	PENNSYLVANIA	2021-22	41217.0
62	PENNSYLVANIA	2018-19	25909.0
128	PENNSYLVANIA	2020-21	42766.0

In [12]:
```
sns.set_style("ticks")
sns.set_style("dark")
sns.set_palette("husl")
plt.figure(figsize=(6, 8))
sns.barplot(x="year", y="homeschool_students",hue="state" ,data=get_state('PENNSYLVANIA'))
plt.show()
```

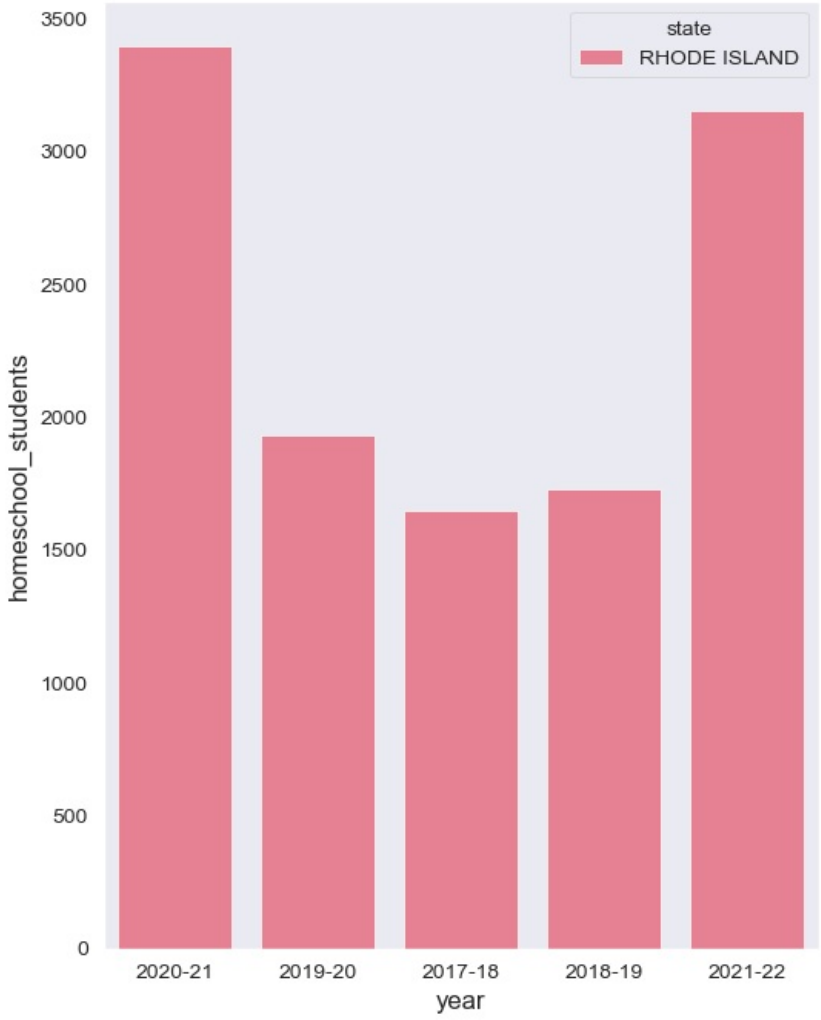


In [13]:
```
get_state('RHODE ISLAND')
```

Out[13]:

	state	year	homeschool_students
111	RHODE ISLAND	2020-21	3396.0
78	RHODE ISLAND	2019-20	1931.0
12	RHODE ISLAND	2017-18	1648.0
45	RHODE ISLAND	2018-19	1731.0
144	RHODE ISLAND	2021-22	3154.0

In [14]:
```
sns.set_style("ticks")
sns.set_style("dark")
sns.set_palette("husl")
plt.figure(figsize=(6, 8))
sns.barplot(x="year", y="homeschool_students",hue="state" ,data=get_state('RHODE ISLAND'))
plt.show()
```

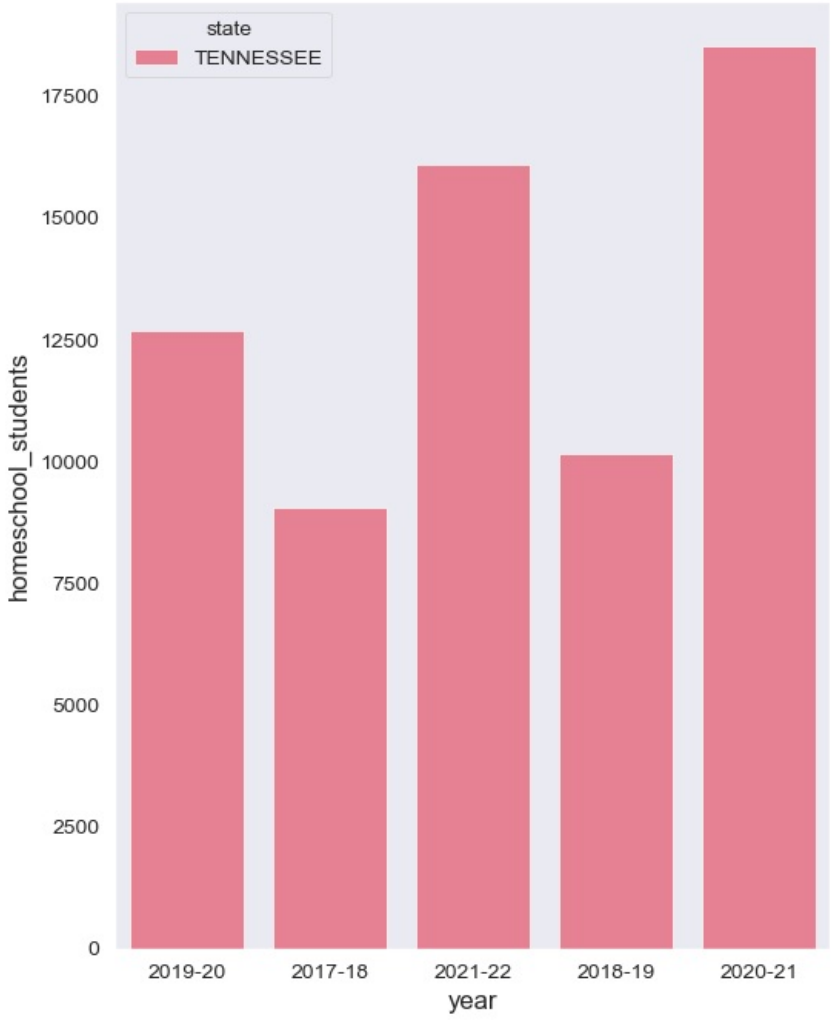


In [15]:
```
get_state('TENNESSEE')
```

Out[15]:

	state	year	homeschool_students
96	TENNESSEE	2019-20	12685.0
30	TENNESSEE	2017-18	9069.0
162	TENNESSEE	2021-22	16080.0
63	TENNESSEE	2018-19	10161.0
129	TENNESSEE	2020-21	18525.0

In [16]:
```
sns.set_style("ticks")
sns.set_style("dark")
sns.set_palette("husl")
plt.figure(figsize=(6, 8))
sns.barplot(x="year", y="homeschool_students",hue="state" ,data=get_state('TENNESSEE'))
plt.show()
```



In [17]:
```
def total_homeschool_students_per_state(state):
    statelocated=data_state[data_state['state']== str(state)]
    print("The total sum of homeschool_students for {} is of {}".format(state,statelocated['homeschool_students
```

In [18]:
```
total_homeschool_students_per_state('CALIFORNIA')
```

```
The total sum of homeschool_students for CALIFORNIA is of 249012.0
```

In [19]:
```
total_homeschool_students_per_state('PENNSYLVANIA')
```

```
The total sum of homeschool_students for PENNSYLVANIA is of 162080.0
```

In [20]:
```
total_homeschool_students_per_state('RHODE ISLAND')
```

```
The total sum of homeschool_students for RHODE ISLAND is of 11860.0
```

In [21]:
```
total_homeschool_students_per_state('TENNESSEE')
```

```
The total sum of homeschool_students for TENNESSEE is of 66520.0
```

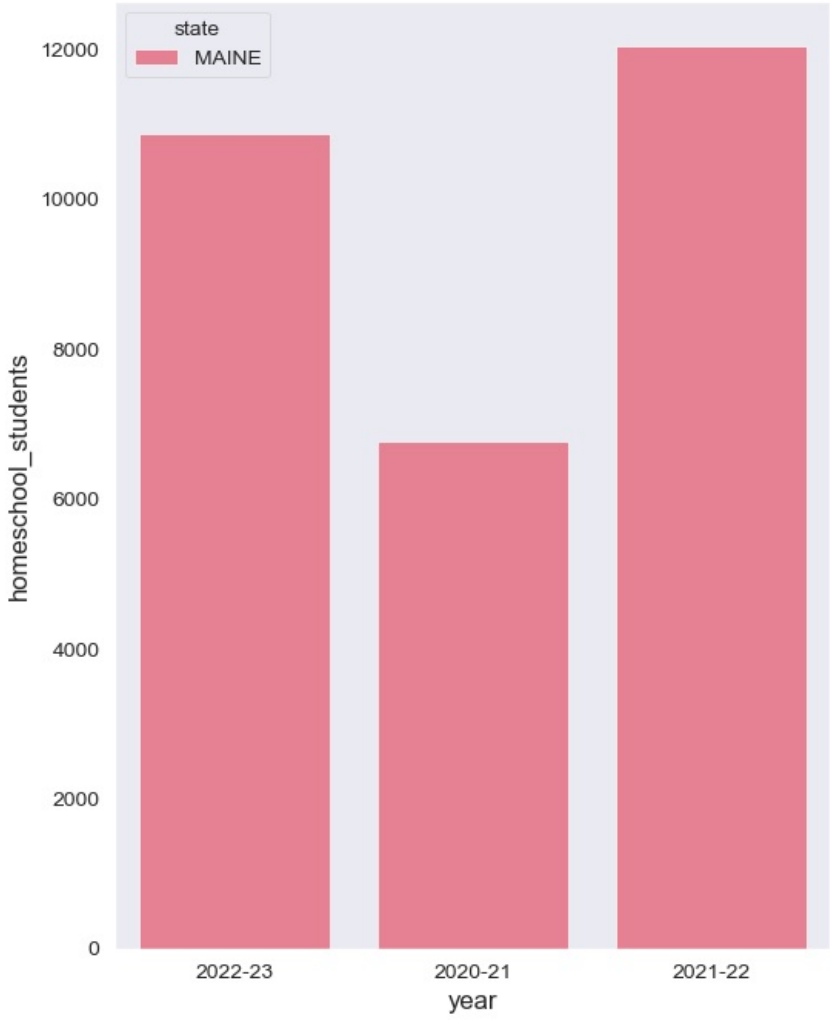
Maine only shared district-level data starting with the 2020-21 school year

In [22]:
```
get_state('MAINE')
```

Out[22]:

	state	year	homeschool_students
187	MAINE	2022-23	10871.0
121	MAINE	2020-21	6775.0
154	MAINE	2021-22	12044.0

In [23]:
```
sns.set_style("ticks")
sns.set_style("dark")
sns.set_palette("husl")
plt.figure(figsize=(6, 8))
sns.barplot(x="year", y="homeschool_students",hue="state" ,data=get_state('MAINE'))
plt.show()
```



In [24]:
```
total_homeschool_students_per_state('MAINE')
```

```
The total sum of homeschool_students for MAINE is of 29690.0
```

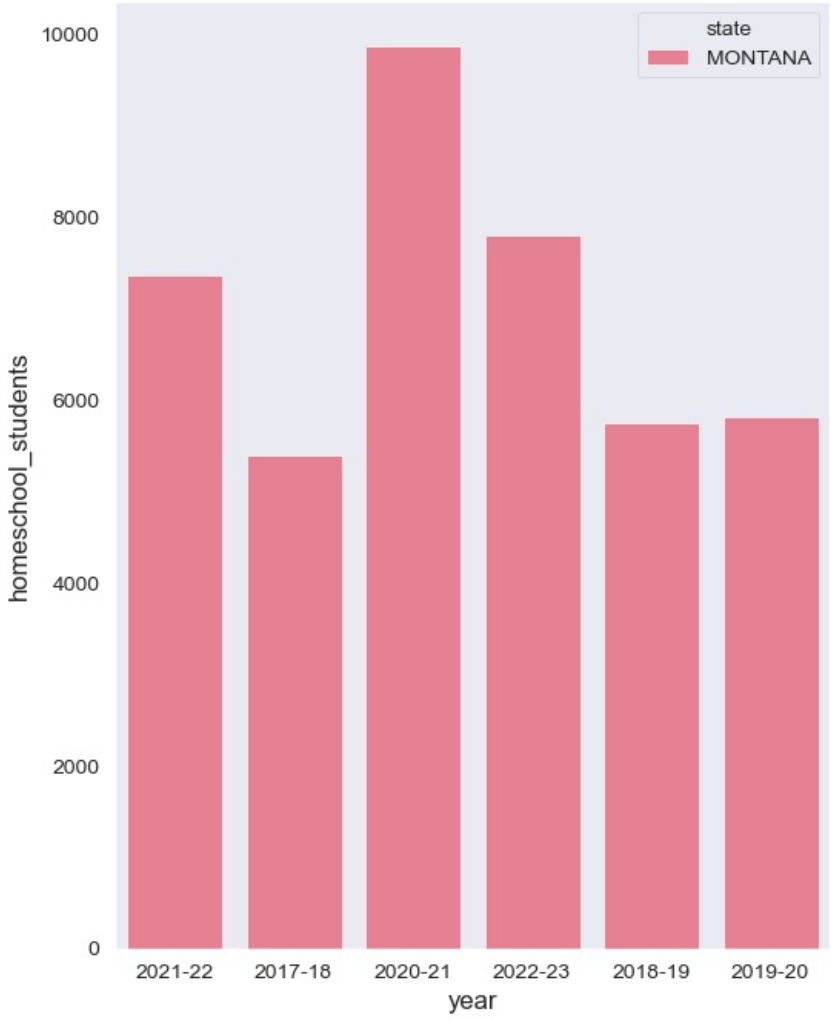
In [25]:
```
# MONTANA, VERMONT and NEBRASKA
```

In [26]:
```
get_state('MONTANA')
```

Out[26]:

	state	year	homeschool_students
156	MONTANA	2021-22	7368.0
24	MONTANA	2017-18	5390.0
123	MONTANA	2020-21	9868.0
189	MONTANA	2022-23	7799.0
57	MONTANA	2018-19	5743.0
90	MONTANA	2019-20	5815.0

In [27]:
```
sns.set_style("ticks")
sns.set_style("dark")
sns.set_palette("husl")
plt.figure(figsize=(6, 8))
sns.barplot(x="year", y="homeschool_students",hue="state" ,data=get_state('MONTANA'))
plt.show()
```



In [28]:
```
total_homeschool_students_per_state('MONTANA')
```

```
The total sum of homeschool_students for MONTANA is of 41983.0
```

In [29]:
```
total_homeschool_students_per_state('VERMONT')
```

```
The total sum of homeschool_students for VERMONT is of 20437.0
```

In [30]:
```
total_homeschool_students_per_state('NEBRASKA')
```

```
The total sum of homeschool_students for NEBRASKA is of 67593.0
```

Grouped by state to know the highest state with enrollment and the lowest state

In [31]:
```
DATA_STATE_AREA=(data_state
 .groupby(['state'])
 .agg(SUM_ESTIMATE_HOMESCHOOL_STUDENTS_STATE_VALUE=('homeschool_students','sum'))
).sort_values(by=['SUM_ESTIMATE_HOMESCHOOL_STUDENTS_STATE_VALUE'],ascending=True)
DATA_STATE_AREA
```

Out[31]:

state	SUM_ESTIMATE_HOMESCHOOL_STUDENTS_STATE_VALUE
DISTRICT OF COLUMBIA	4235.0
RHODE ISLAND	11860.0
WYOMING	18082.0
KANSAS	20260.0
VERMONT	20437.0
NORTH DAKOTA	21725.0
DELAWARE	22026.0
NEW HAMPSHIRE	23621.0
HAWAII	25797.0
MAINE	29690.0
SOUTH DAKOTA	38316.0
MONTANA	41983.0
COLORADO	59500.0
MASSACHUSETTS	64765.0
TENNESSEE	66520.0
NEW MEXICO	67315.0
NEBRASKA	67593.0
MISSISSIPPI	134301.0
MINNESOTA	144096.0
ARKANSAS	154773.0
WISCONSIN	154987.0
SOUTH CAROLINA	155294.0
LOUISIANA	160000.0
PENNSYLVANIA	162080.0
WASHINGTON	163713.0
KENTUCKY	188024.0
MARYLAND	196616.0
OHIO	243599.0
NEW YORK	243683.0
CALIFORNIA	249012.0
VIRGINIA	315343.0
GEORGIA	479064.0
FLORIDA	743022.0

In [ ]:

home_school_district

- lea_name: Official name of the local education agency (school district) as recorded by the National Center for Education Statistics
- lea_id: official id for the local education agency (school district) created by the National Center for Education Statistics
- state: state name
- year: Reflects the academic school year. Homeschooling counts are sometimes collected as point-in-time figures at different moments in the year, depending by state and district.
- homeschool_students: The number of home-schooling students reported by the state, under as many legal ways to home-school as it was possible for the Post to collect

In [32]:
```
import pandas as pd
data_district = pd.read_csv('home_school_district.csv')
data_district=data_district.sort_values(by=['state'],ascending=True)
data_district
```

Out[32]:

	lea_name	lea_id	state	year	homeschool_students
0	Alma School District	502250	AR	2017-18	142.0
945	Mulberry/Pleasant View Bi-County Schools	510290	AR	2021-22	71.0
944	Mulberry School District	510290	AR	2020-21	73.0
943	Mulberry School District	510290	AR	2019-20	54.0
942	Mulberry School District	510290	AR	2018-19	37.0
...	...	...	...	...	...
22986	Platte County School District #2	5603180	WY	2018-19	9.0
22987	Platte County School District #2	5603180	WY	2019-20	0.0
22988	Platte County School District #2	5603180	WY	2020-21	21.0
23009	Sublette County School District #1	5604860	WY	2017-18	56.0
23063	Washakie County School District #2	5605820	WY	2017-18	0.0

37674 rows × 5 columns

In [33]:
```
data_district.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 37674 entries, 0 to 23063
Data columns (total 5 columns):
 #   Column               Non-Null Count  Dtype
---  ------               --------------  -----
 0   lea_name             37637 non-null  object
 1   lea_id               37639 non-null  object
 2   state                37674 non-null  object
 3   year                 37674 non-null  object
 4   homeschool_students  37192 non-null  float64
dtypes: float64(1), object(4)
memory usage: 1.7+ MB
```

In [34]:
```
data_district.describe()
```

Out[34]:

	homeschool_students
count	37192.000000
mean	117.036701
std	402.135777
min	0.000000
25%	12.000000
50%	35.000000
75%	89.000000
max	13641.000000

In [35]:
```
data_district.describe(include='O')
```

Out[35]:

	lea_name	lea_id	state	year
count	37637	37639	37674	37674
unique	6660	6738	30	6
top	Perry Local	502250	CA	2021-22
freq	18	6	5355	6518

In [36]:
```
data_district.shape
```

Out[36]:
```
(37674, 5)
```

In [37]:
```
data_district=data_district.dropna()
data_district
```

Out[37]:

	lea_name	lea_id	state	year	homeschool_students
0	Alma School District	502250	AR	2017-18	142.0
945	Mulberry/Pleasant View Bi-County Schools	510290	AR	2021-22	71.0
944	Mulberry School District	510290	AR	2020-21	73.0
943	Mulberry School District	510290	AR	2019-20	54.0
942	Mulberry School District	510290	AR	2018-19	37.0
...	...	...	...	...	...
22986	Platte County School District #2	5603180	WY	2018-19	9.0
22987	Platte County School District #2	5603180	WY	2019-20	0.0
22988	Platte County School District #2	5603180	WY	2020-21	21.0
23009	Sublette County School District #1	5604860	WY	2017-18	56.0
23063	Washakie County School District #2	5605820	WY	2017-18	0.0

37167 rows × 5 columns

In [38]:
```
data_district.shape
```

Out[38]:
```
(37167, 5)
```

Data Transformation

In [39]:
```
data_district['state'].unique()
```

Out[39]:
```
array(['AR', 'CA', 'CO', 'DC', 'DE', 'FL', 'GA', 'HI', 'KS', 'KY', 'LA',
       'MA', 'MD', 'ME', 'MN', 'MS', 'ND', 'NH', 'NM', 'NY', 'OH', 'PA',
       'RI', 'SC', 'SD', 'TN', 'VA', 'WA', 'WI', 'WY'], dtype=object)
```

In [40]:
```
district_state={'AR':'ARKANSAS',
                        'CA':'CALIFORNIA',
                        'CO':'COLORADO',
                        'DC':'DISTRICT OF COLUMBIA',
                        'DE':'DELAWARE',
                        'FL':'FLORIDA',
                        'GA':'GEORGIA',
                        'HI':'HAWAII',
                        'KS':'KANSAS',
                        'KY':'KENTUCKY',
                        'LA':'LOUISIANA',
                        'MA':'MAINE',
                        'MD':'MARYLAND',
                        'ME':'MASSACHUSETTS',
                        'MN':'MINNESOTA',
                        'MS':'MISSISSIPPI',
                        'ND':'NORTH DAKOTA',
                        'NH':'NEW HAMPSHIRE',
                        'NM':'NEW MEXICO',
                        'NY':'NEW YORK',
                        'OH':'OHIO',
                        'PA':'PENNSYLVANIA',
                        'RI':'RHODE ISLAND',
                        'SC':'SOUTH CAROLINA',
                        'SD':'SOUTH DAKOTA',
                        'TN':'TENNESSEE',
                        'VA':'VIRGINIA',
                        'WA':'WASHINGTON',
                        'WI':'WISCONSIN',
                        'WY':'WYOMING'
                }
data_district['state'] = data_district['state'].replace(district_state)
```

```
C:\Users\Admin\AppData\Local\Temp\ipykernel_13696\1806540588.py:32: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  data_district['state'] = data_district['state'].replace(district_state)
```

In [41]:
```
data_district
```

Out[41]:

	lea_name	lea_id	state	year	homeschool_students
0	Alma School District	502250	ARKANSAS	2017-18	142.0
945	Mulberry/Pleasant View Bi-County Schools	510290	ARKANSAS	2021-22	71.0
944	Mulberry School District	510290	ARKANSAS	2020-21	73.0
943	Mulberry School District	510290	ARKANSAS	2019-20	54.0
942	Mulberry School District	510290	ARKANSAS	2018-19	37.0
...	...	...	...	...	...
22986	Platte County School District #2	5603180	WYOMING	2018-19	9.0
22987	Platte County School District #2	5603180	WYOMING	2019-20	0.0
22988	Platte County School District #2	5603180	WYOMING	2020-21	21.0
23009	Sublette County School District #1	5604860	WYOMING	2017-18	56.0
23063	Washakie County School District #2	5605820	WYOMING	2017-18	0.0

37167 rows × 5 columns

In [42]:
```
data_district.to_csv("data_district.csv",index=False)
```

In [43]:
```
data_state['state'].unique()
```

Out[43]:
```
array(['ARKANSAS', 'CALIFORNIA', 'COLORADO', 'DELAWARE',
       'DISTRICT OF COLUMBIA', 'FLORIDA', 'GEORGIA', 'HAWAII', 'KANSAS',
       'KENTUCKY', 'LOUISIANA', 'MAINE', 'MARYLAND', 'MASSACHUSETTS',
       'MINNESOTA', 'MISSISSIPPI', 'MONTANA', 'NEBRASKA', 'NEW HAMPSHIRE',
       'NEW MEXICO', 'NEW YORK', 'NORTH DAKOTA', 'OHIO', 'PENNSYLVANIA',
       'RHODE ISLAND', 'SOUTH CAROLINA', 'SOUTH DAKOTA', 'TENNESSEE',
       'VERMONT', 'VIRGINIA', 'WASHINGTON', 'WISCONSIN', 'WYOMING'],
      dtype=object)
```

In [44]:
```
data_state['state'].nunique()
```

Out[44]:
```
33
```

In [45]:
```
data_district['state'].unique()
```

Out[45]:
```
array(['ARKANSAS', 'CALIFORNIA', 'COLORADO', 'DISTRICT OF COLUMBIA',
       'DELAWARE', 'FLORIDA', 'GEORGIA', 'HAWAII', 'KANSAS', 'KENTUCKY',
       'LOUISIANA', 'MAINE', 'MARYLAND', 'MASSACHUSETTS', 'MINNESOTA',
       'MISSISSIPPI', 'NORTH DAKOTA', 'NEW HAMPSHIRE', 'NEW MEXICO',
       'NEW YORK', 'OHIO', 'PENNSYLVANIA', 'RHODE ISLAND',
       'SOUTH CAROLINA', 'SOUTH DAKOTA', 'TENNESSEE', 'VIRGINIA',
       'WASHINGTON', 'WISCONSIN', 'WYOMING'], dtype=object)
```

Absence of MONTANA, VERMONT and NEBRASKA in home_school_district confirmed

In [46]:
```
data_district['state'].nunique()
```

Out[46]:
```
30
```

Grouped by state to know the highest district with enrollment and the lowest state

In [47]:
```
STATE_AREA=(data_district
 .groupby(['state'])
 .agg(SUM_ESTIMATE_HOMESCHOOL_STUDENTS_DISTRICT_VALUE=('homeschool_students','sum'))
).sort_values(by=['SUM_ESTIMATE_HOMESCHOOL_STUDENTS_DISTRICT_VALUE'],ascending=True)
STATE_AREA
```

Out[47]:

state	SUM_ESTIMATE_HOMESCHOOL_STUDENTS_DISTRICT_VALUE
DISTRICT OF COLUMBIA	4235.0
RHODE ISLAND	11860.0
WYOMING	18082.0
KANSAS	20260.0
NEW HAMPSHIRE	21518.0
NORTH DAKOTA	21725.0
DELAWARE	22026.0
HAWAII	25797.0
MASSACHUSETTS	29423.0
SOUTH DAKOTA	38314.0
COLORADO	59500.0
NEW MEXICO	63884.0
MAINE	64765.0
TENNESSEE	65883.0
MISSISSIPPI	134129.0
MINNESOTA	144096.0
WISCONSIN	149976.0
ARKANSAS	154733.0
SOUTH CAROLINA	155294.0
LOUISIANA	159993.0
PENNSYLVANIA	162070.0
WASHINGTON	163713.0
KENTUCKY	188024.0
MARYLAND	196616.0
NEW YORK	243683.0
OHIO	246460.0
CALIFORNIA	248183.0
VIRGINIA	315343.0
GEORGIA	479288.0
FLORIDA	743022.0

In [48]:
```
def get_district_state(state):
    return data_district[data_district['state']== str(state)]
```

State with the highest enrollment is Florida

In [49]:
```
Florida_district = get_district_state('FLORIDA')
Florida_district
```

Out[49]:

	lea_name	lea_id	state	year	homeschool_students
1624	Leon	1201110	FLORIDA	2020-21	2007.0
1619	Lee	1201080	FLORIDA	2021-22	3698.0
1623	Leon	1201110	FLORIDA	2019-20	1650.0
1622	Leon	1201110	FLORIDA	2018-19	1801.0
1621	Leon	1201110	FLORIDA	2017-18	2026.0
...	...	...	...	...	...
1754	St. Johns	1201740	FLORIDA	2019-20	2224.0
1753	St. Johns	1201740	FLORIDA	2018-19	1974.0
1752	St. Johns	1201740	FLORIDA	2017-18	1745.0
1751	Seminole	1201710	FLORIDA	2022-23	4202.0
1758	St. Lucie	1201770	FLORIDA	2017-18	1414.0

402 rows × 5 columns

In [50]:
```
Florida_district['lea_name'].unique()
```

Out[50]:
```
array(['Leon', 'Lee', 'Lake', 'Liberty', 'Levy', 'Madison', 'Lafayette',
       'Indian River', 'Manatee', 'Holmes', 'Hillsborough', 'Highlands',
       'Jefferson', 'Jackson', 'Monroe', 'Palm Beach', 'Osceola',
       'Orange', 'Okeechobee', 'Polk', 'Pinellas', 'Pasco', 'Okaloosa',
       'Miami-Dade', 'Martin', 'Marion', 'Nassau', 'Hamilton', 'Hernando',
       'Clay', 'Citrus', 'Charlotte', 'Calhoun', 'Desoto', 'Dade',
       'Columbia', 'Collier', 'Broward', 'Bay', 'Baker', 'Alachua',
       'Bradford', 'Brevard', 'Gulf', 'Glades', 'Gilchrist', 'Hardee',
       'Hendry', 'Escambia', 'Duval', 'Dixie', 'Flagler', 'Gadsden',
       'Franklin', 'Putnam', 'Volusia', 'Union', 'Taylor', 'Suwannee',
       'Wakulla', 'Washington', 'Walton', 'Sumter', 'Seminole',
       'Sarasota', 'Santa Rosa', 'St. Lucie', 'St. Johns'], dtype=object)
```

In [51]:
```
Florida_district['lea_name'].nunique()
```

Out[51]:
```
68
```

The highest enrollment in Florida district and the local education agency name (lea_name)

In [52]:
```
FLORIDA_AREA=(Florida_district
 .groupby(['lea_name'])
 .agg(SUM_ESTIMATE_HOMESCHOOL_STUDENTS_DISTRICT_VALUE=('homeschool_students','sum'))
).sort_values(by=['SUM_ESTIMATE_HOMESCHOOL_STUDENTS_DISTRICT_VALUE'],ascending=True)
FLORIDA_AREA
```

Out[52]:

lea_name	SUM_ESTIMATE_HOMESCHOOL_STUDENTS_DISTRICT_VALUE
Franklin	169.0
Glades	398.0
Liberty	454.0
Lafayette	558.0
Hamilton	600.0
...	...
Broward	45036.0
Palm Beach	45531.0
Orange	46317.0
Duval	48072.0
Hillsborough	55389.0

68 rows × 1 columns

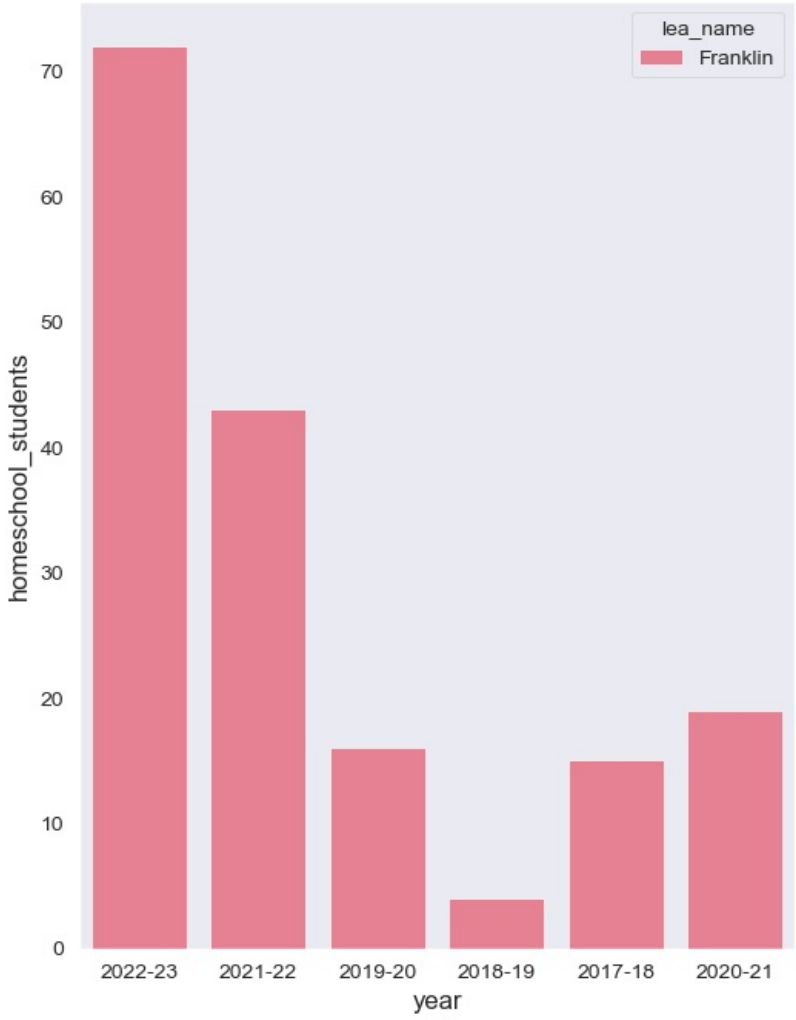
local education agency (lea) with the lowest enrollment

In [53]:
```
Franklin_lea=Florida_district[Florida_district['lea_name'] == 'Franklin']
Franklin_lea
```

Out[53]:

	lea_name	lea_id	state	year	homeschool_students
1518	Franklin	1200570	FLORIDA	2022-23	72.0
1517	Franklin	1200570	FLORIDA	2021-22	43.0
1515	Franklin	1200570	FLORIDA	2019-20	16.0
1514	Franklin	1200570	FLORIDA	2018-19	4.0
1513	Franklin	1200570	FLORIDA	2017-18	15.0
1516	Franklin	1200570	FLORIDA	2020-21	19.0

In [54]:
```
sns.set_style("ticks")
sns.set_style("dark")
sns.set_palette("husl")
plt.figure(figsize=(6, 8))
sns.barplot(x="year", y="homeschool_students",hue="lea_name" ,data=Franklin_lea)
plt.show()
```



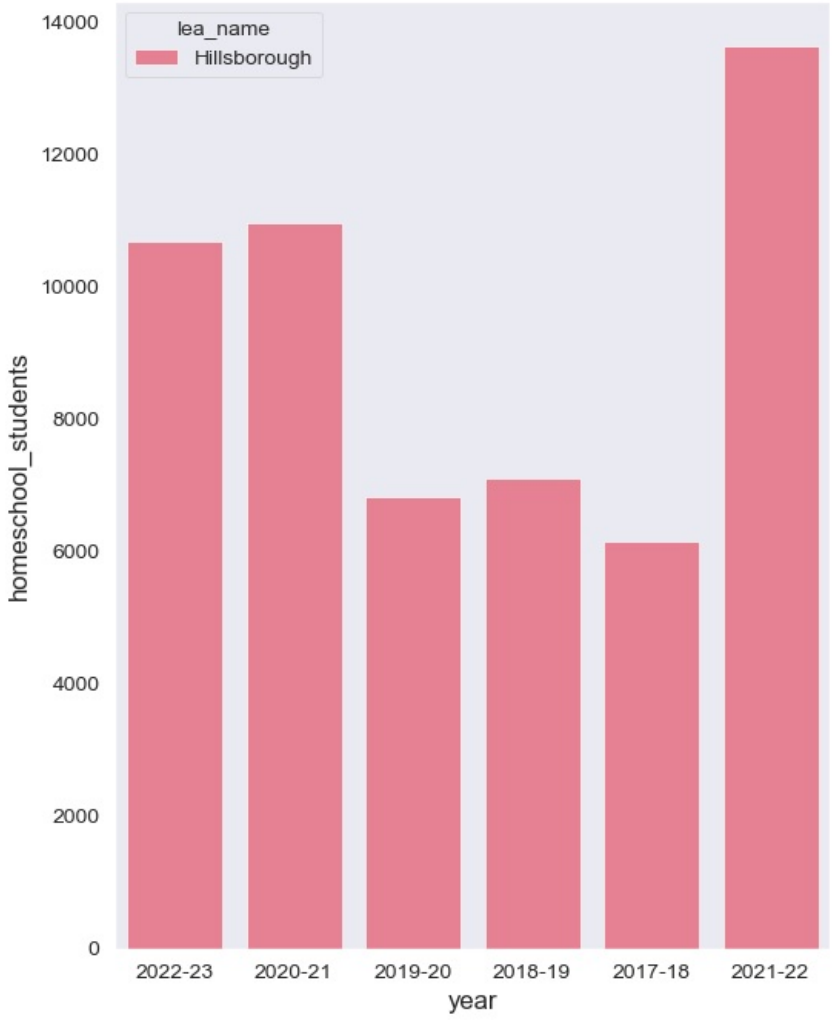
local education agency (lea) with the highest enrollment

In [55]:
```
Hillsborough_lea=Florida_district[Florida_district['lea_name'] == 'Hillsborough']
Hillsborough_lea
```

Out[55]:

	lea_name	lea_id	state	year	homeschool_students
1578	Hillsborough	1200870	FLORIDA	2022-23	10680.0
1576	Hillsborough	1200870	FLORIDA	2020-21	10964.0
1575	Hillsborough	1200870	FLORIDA	2019-20	6837.0
1574	Hillsborough	1200870	FLORIDA	2018-19	7117.0
1573	Hillsborough	1200870	FLORIDA	2017-18	6150.0
1577	Hillsborough	1200870	FLORIDA	2021-22	13641.0

In [56]:
```
sns.set_style("ticks")
sns.set_style("dark")
sns.set_palette("husl")
plt.figure(figsize=(6, 8))
sns.barplot(x="year", y="homeschool_students",hue="lea_name" ,data=Hillsborough_lea)
plt.show()
```



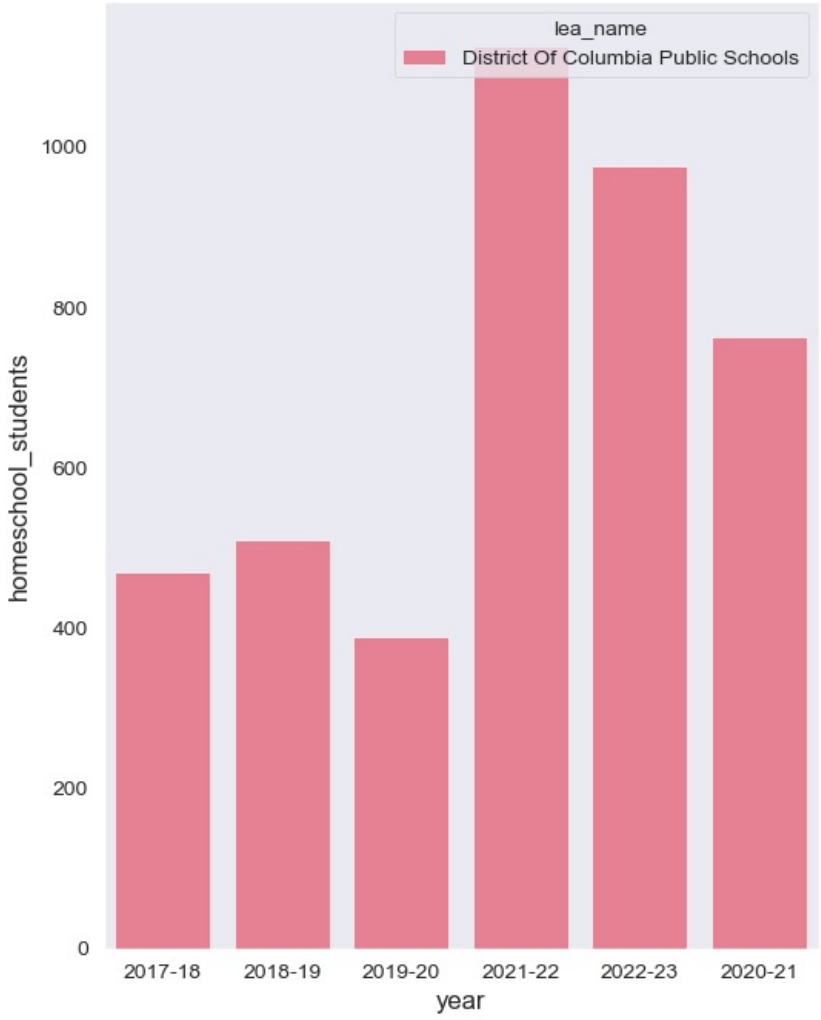
State with the lowest enrollment is District of Columbia

In [57]:
```
District_of_Columbia_district = get_district_state('DISTRICT OF COLUMBIA')
District_of_Columbia_district
```

Out[57]:

	lea_name	lea_id	state	year	homeschool_students
31233	District Of Columbia Public Schools	1100030	DISTRICT OF COLUMBIA	2017-18	469.0
31234	District Of Columbia Public Schools	1100030	DISTRICT OF COLUMBIA	2018-19	510.0
31235	District Of Columbia Public Schools	1100030	DISTRICT OF COLUMBIA	2019-20	389.0
31237	District Of Columbia Public Schools	1100030	DISTRICT OF COLUMBIA	2021-22	1126.0
31238	District Of Columbia Public Schools	1100030	DISTRICT OF COLUMBIA	2022-23	977.0
31236	District Of Columbia Public Schools	1100030	DISTRICT OF COLUMBIA	2020-21	764.0

In [58]:
```
sns.set_style("ticks")
sns.set_style("dark")
sns.set_palette("husl")
plt.figure(figsize=(6, 8))
sns.barplot(x="year", y="homeschool_students",hue="lea_name" ,data=District_of_Columbia_district)
plt.show()
```



In [ ]: