Final Project

March 3, 2025

1 Data Exploration and Analysis - Final Project

Author: Todd Estep

Date: 3/2/2025

Class: Data Exploration and Analysis

Prof: Metzger

1.0.1 Part 1 - Importing data and Libraries

[47]:
```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
```

[48]:
```
data = pd.read_excel('finaldata.xlsx')
```

[49]:
```
data
```

[49]:

	state abv.	county_fips	county	m_housing	m_food	\
0	AL	1001	Autauga County	944.584583	338.420067	
1	AL	1003	Baldwin County	1051.000000	382.602658	
2	AL	1005	Barbour County	652.000000	325.259275	
3	AL	1007	Bibb County	691.599283	309.278317	
4	AL	1009	Blount County	640.721025	304.578042	
…	…	…	…	…	…	
3138	WY	56037	Sweetwater County	896.000000	338.420067	
3139	WY	56039	Teton County	1317.000000	475.668175	
3140	WY	56041	Uinta County	684.000000	326.199342	
3141	WY	56043	Washakie County	684.000000	354.401000	
3142	WY	56045	Weston County	801.000000	350.640792	

	m_transport	m_health	m_other	m_tax	m_total	…	\
0	1124.548500	525.053017	431.675333	667.752767	4032.034167	…	
1	1128.868000	495.053017	482.345008	708.367350	4248.236000	…	
2	1072.752583	485.199867	328.805300	548.314408	3412.331417	…	
3	1190.644917	501.053017	336.751833	587.831750	3617.159167	…	

```
4     1165.808083  501.053017  318.052067  564.144975  3494.357083  …
…           …         …          …          …           …  …
3138  1218.016000  674.583208  415.328733  564.086425  4106.434583  …
3139  1252.413583  674.583208  603.154992  774.953450  5097.773417  …
3140  1191.231917  674.583208  339.888183  499.248617  3715.151333  …
3141  1140.456833  674.583208  349.376833  496.646892  3699.464833  …
3142  1298.514667  674.583208  387.477092  558.107058  4070.322917  …

      st_cost_rank  st_med_aff_rank  st_income_rank  min_wage     m_income  \
0                8             48.0            11.0      7.25  1256.666667
1                2             59.0             4.0      7.25  1256.666667
2               63             19.0            55.0      7.25  1256.666667
3               18             20.0            39.0      7.25  1256.666667
4               24             46.0            18.0      7.25  1256.666667
…              …                …               …        …           …
3138            12             19.0             6.0      7.25  1256.666667
3139             1             13.0             1.0      7.25  1256.666667
3140            19             20.0            10.0      7.25  1256.666667
3141            18              8.0            18.0      7.25  1256.666667
3142             7              9.0            12.0      7.25  1256.666667

      a_min_income   m_ratio  a_min_ratio  median_household_income  \
0          15080.0  0.311671     0.311671                    70148
1          15080.0  0.295809     0.295809                    71704
2          15080.0  0.368272     0.368272                    41151
3          15080.0  0.347418     0.347418                    54309
4          15080.0  0.359627     0.359627                    60553
…            …       …           …                          …
3138       15080.0  0.306024     0.306024                    75779
3139       15080.0  0.246513     0.246513                   127677
3140       15080.0  0.338255     0.338255                    73072
3141       15080.0  0.339689     0.339689                    60699
3142       15080.0  0.308739     0.308739                    67677

      median_ratio
0         1.449806
1         1.406545
2         1.004958
3         1.251189
4         1.444066
…            …
3138      1.537810
3139      2.087137
3140      1.639054
3141      1.367292
3142      1.385578
```

[3143 rows x 29 columns]

1.0.2 Part 2 - Calculating Descriptive Statistics and Outliers

[112]:
```
data.mean(numeric_only=True)
```

[112]:
```
county_fips               30374.576519
m_housing                   803.242426
m_food                      337.483889
m_transport                1144.573051
m_health                    490.475464
m_other                     383.804850
m_tax                       589.548597
m_total                    3749.128282
a_housing                  9638.909116
a_food                     4049.806665
a_transport               13734.876612
a_health                   5885.705567
a_other                    4605.658203
a_taxes                    7074.583165
a_total_cost              44989.539389
median_family_income      82714.867135
num_counties_in_st           96.394209
st_cost_rank                 48.697105
st_med_aff_rank              48.688279
st_income_rank               48.688279
min_wage                      9.711429
m_income                   1683.314286
a_min_income              20199.771429
m_ratio                       0.449504
a_min_ratio                   0.449504
median_household_income   63240.506841
median_ratio                  1.398116
dtype: float64
```

[113]:
```
data.std(numeric_only=True)
```

[113]:
```
county_fips               15168.624870
m_housing                   274.255576
m_food                       36.104752
m_transport                  90.963024
m_health                     87.105507
m_other                      99.454344
m_tax                       122.434463
m_total                     491.214497
a_housing                  3291.066914
a_food                      433.257020
```

```
a_transport                 1091.556288
a_health                    1045.266081
a_other                     1193.452125
a_taxes                     1469.213556
a_total_cost                5894.573965
median_family_income       20074.548603
num_counties_in_st            57.025363
st_cost_rank                  43.109713
st_med_aff_rank               42.814481
st_income_rank                42.814462
min_wage                       2.910342
m_income                     504.459227
a_min_income                6053.510718
m_ratio                        0.123340
a_min_ratio                    0.123340
median_household_income    16248.920157
median_ratio                   0.257360
dtype: float64
```

[112]:
```
data.mean(numeric_only=True)
```

[112]:
```
county_fips                30374.576519
m_housing                    803.242426
m_food                       337.483889
m_transport                 1144.573051
m_health                     490.475464
m_other                      383.804850
m_tax                        589.548597
m_total                     3749.128282
a_housing                   9638.909116
a_food                      4049.806665
a_transport                13734.876612
a_health                    5885.705567
a_other                     4605.658203
a_taxes                     7074.583165
a_total_cost               44989.539389
median_family_income       82714.867135
num_counties_in_st            96.394209
st_cost_rank                  48.697105
st_med_aff_rank               48.688279
st_income_rank                48.688279
min_wage                       9.711429
m_income                    1683.314286
a_min_income               20199.771429
m_ratio                        0.449504
a_min_ratio                    0.449504
median_household_income    63240.506841
```

```
median_ratio                    1.398116
dtype: float64
```

[113]:
```
data.std(numeric_only=True)
```

[113]:
```
county_fips                15168.624870
m_housing                    274.255576
m_food                        36.104752
m_transport                   90.963024
m_health                      87.105507
m_other                       99.454344
m_tax                        122.434463
m_total                      491.214497
a_housing                   3291.066914
a_food                       433.257020
a_transport                 1091.556288
a_health                    1045.266081
a_other                     1193.452125
a_taxes                     1469.213556
a_total_cost                5894.573965
median_family_income       20074.548603
num_counties_in_st            57.025363
st_cost_rank                  43.109713
st_med_aff_rank               42.814481
st_income_rank                42.814462
min_wage                       2.910342
m_income                     504.459227
a_min_income                6053.510718
m_ratio                        0.123340
a_min_ratio                    0.123340
median_household_income    16248.920157
median_ratio                   0.257360
dtype: float64
```

[224]:
```
data.min(numeric_only=True)
```

[224]:
```
county_fips                 1001.000000
m_housing                    412.534217
m_food                       256.635217
m_transport                  531.612633
m_health                     320.842792
m_other                      237.797950
m_tax                        397.818642
m_total                     2994.983750
a_housing                   4950.410600
a_food                      3079.622600
a_transport                 6379.351600
```

```
a_health                    3850.113500
a_other                     2853.575400
a_taxes                     4773.823700
a_total_cost               35939.805000
median_family_income       34894.000000
num_counties_in_st             1.000000
st_cost_rank                   1.000000
st_med_aff_rank                1.000000
st_income_rank                 1.000000
min_wage                       7.250000
m_income                    1256.666667
a_min_income               15080.000000
m_ratio                        0.209606
a_min_ratio                    0.209606
median_household_income    28972.000000
median_ratio                   0.652809
dtype: float64
```

[225]:
```
data.max(numeric_only=True)
```

[225]:
```
county_fips                56045.000000
m_housing                   2849.000000
m_food                       741.703942
m_transport                 1767.560583
m_health                     814.621342
m_other                     1130.125750
m_tax                       1669.890333
m_total                     7727.068333
a_housing                  34188.000000
a_food                      8900.447300
a_transport                21210.727000
a_health                    9775.456100
a_other                    13561.509000
a_taxes                    20038.684000
a_total_cost               92724.820000
median_family_income      211445.000000
num_counties_in_st           254.000000
st_cost_rank                 254.000000
st_med_aff_rank              252.000000
st_income_rank               252.000000
min_wage                      17.500000
m_income                    3033.333333
a_min_income               36400.000000
m_ratio                        0.852108
a_min_ratio                    0.852108
median_household_income    167605.000000
median_ratio                   2.711813
```

```
dtype: float64
```

[122]:
```
datadf = pd.DataFrame(data)
datadfoutlier = datadf[datadf['min_wage']>=17]
datadfoutlier
```

[122]:
```
     state abv.  county_fips                county    m_housing      m_food  \
205          CA         6037     Los Angeles County  1777.000000  364.741617
227          CA         6081       San Mateo County  2550.637333  436.185833
320          DC        11001  District of Columbia  1687.667833  445.586425
1208         MD        24031      Montgomery County  1803.139833  424.905192

      m_transport    m_health     m_other         m_tax      m_total  …  \
205    984.808167  373.777833   720.603025   908.883500  5129.814167  …
227   1169.506750  584.777833  1004.936167  1539.703000  7285.746750  …
320    843.867833  542.053017   717.747400   965.791083  5202.713583  …
1208  1001.162750  352.053017   749.640383   989.722417  5320.623667  …

      st_cost_rank  st_med_aff_rank  st_income_rank  min_wage     m_income  \
205             22             26.0            28.0     17.27  2993.466667
227              1             46.0             2.0     17.06  2957.066667
320              1              1.0             1.0     17.50  3033.333333
1208             2             19.0             2.0     17.15  2972.666667

      a_min_income   m_ratio  a_min_ratio  median_household_income  \
205        35921.6  0.583543     0.583543                    82455
227        35484.8  0.405870     0.405870                   143795
320        36400.0  0.583029     0.583029                    99897
1208       35672.0  0.558706     0.558706                   118020

      median_ratio
205       1.339473
227       1.644707
320       1.600078
1208      1.848468

[4 rows x 29 columns]
```

[126]:
```
datadfoutlier3 = datadf[datadf['a_total_cost']<=27305]
datadfoutlier3
```

[126]:
```
Empty DataFrame
Columns: [state abv., county_fips, county, m_housing, m_food, m_transport,
m_health, m_other , m_tax, m_total, a_housing, a_food, a_transport, a_health,
a_other, a_taxes, a_total_cost, median_family_income, num_counties_in_st,
st_cost_rank, st_med_aff_rank, st_income_rank, min_wage, m_income, a_min_income,
m_ratio, a_min_ratio, median_household_income, median_ratio]
```

```
Index: []

[0 rows x 29 columns]
```

[111]:

[111]: 2.909878666684371

[125]:
```
datadfoutlier2 = datadf[datadf['a_total_cost']>=62673.4]

datadfoutlier2
```

[125]:

	state abv.	county_fips	county	m_housing	m_food	\
187	CA	6001	Alameda County	1823.915167	420.204917	
193	CA	6013	Contra Costa County	1827.062667	418.324792	
207	CA	6041	Marin County	2278.204917	494.469317	
213	CA	6053	Monterey County	2340.000000	370.381958	
214	CA	6055	Napa County	1819.000000	440.886108	
…	…	…	…	…	…	
2925	VA	51600	Fairfax city	1884.858583	413.624508	
2926	VA	51610	Falls Church city	1958.583000	433.365683	
2936	VA	51685	Manassas Park city	1931.935583	358.161217	
2970	WA	53033	King County	2252.594083	423.965125	
2984	WA	53061	Snohomish County	2065.523583	397.643558	

	m_transport	m_health	m_other	m_tax	m_total	…	\
187	1063.247417	532.777833	755.048908	1057.811583	5653.005833	…	
193	1126.897750	537.777833	755.475342	1087.298500	5752.836583	…	
207	1096.417667	529.777833	932.884417	1372.521333	6704.275417	…	
213	1239.690083	510.777833	911.925750	1389.984250	6762.759750	…	
214	1194.407417	529.777833	760.353517	1121.111833	5865.536500	…	
…	…	…	…	…	…	…	
2925	1004.960750	366.053017	773.339683	1059.659083	5502.495417	…	
2926	870.210667	366.053017	804.786783	1056.068167	5489.067750	…	
2936	1244.308750	366.053017	770.518067	1143.825000	5814.801417	…	
2970	955.555167	376.777833	900.545917	956.448000	5865.886083	…	
2984	1124.091417	367.777833	828.748783	916.859250	5700.644500	…	

	st_cost_rank	st_med_aff_rank	st_income_rank	min_wage	m_income	\
187	6	49.0	5.0	16.00	2773.333333	
193	7	39.0	6.0	16.00	2773.333333	
207	2	53.0	1.0	16.00	2773.333333	
213	16	11.0	25.0	16.00	2773.333333	
214	10	41.0	9.0	16.00	2773.333333	
…	…	…	…	…	…	
2925	7	126.0	5.0	12.00	2080.000000	
2926	3	133.0	1.0	12.00	2080.000000	

```
2936             4              7.0            35.0     12.00  2080.000000
2970             1             38.0             1.0     16.28  2821.866667
2984             2             31.0             2.0     16.28  2821.866667

      a_min_income   m_ratio  a_min_ratio  median_household_income  \
187        33280.0  0.490594     0.490594                   121190
193        33280.0  0.482081     0.482081                   119667
207        33280.0  0.413666     0.413666                   135960
213        33280.0  0.410089     0.410089                    91450
214        33280.0  0.472818     0.472818                    98580
…              …         …            …                        …
2925       24960.0  0.378010     0.378010                   122790
2926       24960.0  0.378935     0.378935                   142513
2936       24960.0  0.357708     0.357708                    90816
2970       33862.4  0.481064     0.481064                   116044
2984       33862.4  0.495008     0.495008                   101440

      median_ratio
187       1.786513
193       1.733449
207       1.689966
213       1.126882
214       1.400554
…              …
2925      1.859611
2926      2.163588
2936      1.301506
2970      1.648572
2984      1.482873

[72 rows x 29 columns]
```

[128]:
```
datadfoutlier4 = datadf[datadf['a_min_ratio']>=.758]
datadfoutlier4
```

[128]:
```
     state abv.  county_fips              county   m_housing      m_food  \
1194         MD        24001     Allegany County  642.328408  336.539933
2954         WA        53001        Adams County  754.000000  305.518108
2955         WA        53003       Asotin County  819.946042  347.820600
2958         WA        53009      Clallam County  847.000000  381.662642
2960         WA        53013     Columbia County  735.000000  359.101275
2961         WA        53015      Cowlitz County  909.000000  356.281125
2963         WA        53019        Ferry County  640.000000  342.180300
2965         WA        53023     Garfield County  676.000000  347.820600
2967         WA        53027  Grays Harbor County  775.000000  355.341067
2973         WA        53039    Klickitat County  836.000000  354.401000
2974         WA        53041        Lewis County  843.000000  353.460942
```

2975	WA	53043	Lincoln County	731.000000	368.501833	
2976	WA	53045	Mason County	866.000000	363.801558	
2977	WA	53047	Okanogan County	712.000000	323.379167	
2978	WA	53049	Pacific County	754.000000	337.480000	
2979	WA	53051	Pend Oreille County	702.000000	343.120317	
2985	WA	53063	Spokane County	890.000000	368.501833	
2986	WA	53065	Stevens County	751.000000	342.180300	
2988	WA	53069	Wahkiakum County	727.000000	356.281125	
2991	WA	53075	Whitman County	794.000000	376.022300	

	m_transport	m_health	m_other	m_tax	m_total	…	\
1194	1045.776667	352.053017	329.346683	509.913333	3215.958000	…	
2954	1212.068583	392.583208	356.481808	460.397258	3481.049167	…	
2955	1080.206333	362.777833	392.902750	457.011967	3460.665333	…	
2958	1052.130083	359.583208	413.391600	466.991208	3520.758750	…	
2960	1198.619833	354.583208	368.117550	459.355675	3474.777667	…	
2961	1071.424417	455.777833	425.712117	499.704550	3717.900083	…	
2963	1209.159500	357.583208	330.461000	432.246458	3311.630500	…	
2965	1225.840083	354.583208	344.471150	446.067058	3394.782250	…	
2967	1090.622750	359.583208	380.310667	448.486408	3409.344083	…	
2973	1146.176750	336.583233	400.518233	470.954633	3544.633750	…	
2974	1149.899250	359.583208	402.557125	477.885133	3586.385750	…	
2975	1220.424417	357.583208	369.934608	465.732100	3513.176083	…	
2976	1160.333833	361.583208	413.774825	489.224125	3654.717417	…	
2977	1188.224000	392.583208	348.360108	449.221317	3413.767917	…	
2978	1158.775250	359.583208	367.235592	451.717200	3428.791333	…	
2979	1209.340917	357.583208	351.637575	449.049317	3412.731417	…	
2985	995.041333	365.777833	423.431192	464.797733	3507.549833	…	
2986	1248.310083	365.777833	367.807658	471.233117	3546.308917	…	
2988	1171.299667	359.583208	364.477050	452.029175	3430.670250	…	
2991	1056.086000	354.583208	393.661658	451.175533	3425.528667	…	

	st_cost_rank	st_med_aff_rank	st_income_rank	min_wage	m_income	\
1194	22	3.0	23.0	15.00	2600.000000	
2954	38	2.0	38.0	16.28	2821.866667	
2955	30	15.0	30.0	16.28	2821.866667	
2958	21	13.0	27.0	16.28	2821.866667	
2960	28	39.0	12.0	16.28	2821.866667	
2961	11	5.0	23.0	16.28	2821.866667	
2963	39	6.0	36.0	16.28	2821.866667	
2965	37	30.0	24.0	16.28	2821.866667	
2967	27	7.0	32.0	16.28	2821.866667	
2973	23	21.0	25.0	16.28	2821.866667	
2974	20	14.0	26.0	16.28	2821.866667	
2975	32	24.0	28.0	16.28	2821.866667	
2976	19	26.0	18.0	16.28	2821.866667	
2977	36	3.0	37.0	16.28	2821.866667	

2978	26	17.0	29.0	16.28	2821.866667
2979	35	8.0	34.0	16.28	2821.866667
2985	16	28.0	16.0	16.28	2821.866667
2986	31	16.0	31.0	16.28	2821.866667
2988	25	1.0	39.0	16.28	2821.866667
2991	34	35.0	19.0	16.28	2821.866667

	a_min_income	m_ratio	a_min_ratio	median_household_income	\
1194	31200.0	0.808468	0.808468	50582	
2954	33862.4	0.810637	0.810637	62410	
2955	33862.4	0.815412	0.815412	65625	
2958	33862.4	0.801494	0.801494	66693	
2960	33862.4	0.812100	0.812100	62935	
2961	33862.4	0.758995	0.758995	71847	
2963	33862.4	0.852108	0.852108	57064	
2965	33862.4	0.831236	0.831236	64413	
2967	33862.4	0.827686	0.827686	62258	
2973	33862.4	0.796095	0.796095	70371	
2974	33862.4	0.786827	0.786827	67954	
2975	33862.4	0.803224	0.803224	68318	
2976	33862.4	0.772116	0.772116	78029	
2977	33862.4	0.826614	0.826614	58661	
2978	33862.4	0.822992	0.822992	55615	
2979	33862.4	0.826865	0.826865	59332	
2985	33862.4	0.804512	0.804512	69079	
2986	33862.4	0.795719	0.795719	65748	
2988	33862.4	0.822541	0.822541	70099	
2991	33862.4	0.823776	0.823776	54412	

	median_ratio
1194	1.310703
2954	1.494042
2955	1.580260
2958	1.578566
2960	1.509329
2961	1.610385
2963	1.435949
2965	1.581177
2967	1.521749
2973	1.654402
2974	1.578981
2975	1.620518
2976	1.779185
2977	1.431971
2978	1.351667
2979	1.448791
2985	1.641198

```
2986      1.544987
2988      1.702753
2991      1.323689

[20 rows x 29 columns]
```

[129]:
```
datadfoutlier5 = datadf[datadf['a_min_ratio']<=.142]
datadfoutlier5
```

[129]:
```
Empty DataFrame
Columns: [state abv., county_fips, county, m_housing, m_food, m_transport,
m_health, m_other , m_tax, m_total, a_housing, a_food, a_transport, a_health,
a_other, a_taxes, a_total_cost, median_family_income, num_counties_in_st,
st_cost_rank, st_med_aff_rank, st_income_rank, min_wage, m_income, a_min_income,
m_ratio, a_min_ratio, median_household_income, median_ratio]
Index: []

[0 rows x 29 columns]
```

[131]:
```
datadfoutlier6 = datadf[datadf['median_household_income']>=111987]
datadfoutlier6
```

[131]:

	state abv.	county_fips	county	m_housing	\
68	AK	2016	Aleutians West Census Area	1349.000000	
187	CA	6001	Alameda County	1823.915167	
193	CA	6013	Contra Costa County	1827.062667	
207	CA	6041	Marin County	2278.204917	
224	CA	6075	San Francisco County	2132.731250	
227	CA	6081	San Mateo County	2550.637333	
229	CA	6085	Santa Clara County	2383.000000	
252	CO	8014	Broomfield County	1906.903000	
263	CO	8035	Douglas County	1926.211083	
265	CO	8039	Elbert County	1320.305000	
445	GA	13117	Forsyth County	1974.428583	
495	GA	13219	Oconee County	1190.128000	
726	IN	18057	Hamilton County	1230.887167	
1086	KY	21185	Oldham County	1033.193833	
1195	MD	24003	Anne Arundel County	1732.200167	
1197	MD	24009	Calvert County	1510.906917	
1201	MD	24017	Charles County	1685.003083	
1203	MD	24021	Frederick County	1515.348167	
1206	MD	24027	Howard County	1773.981917	
1208	MD	24031	Montgomery County	1803.139833	
1211	MD	24037	St. Mary's County	1470.000000	
1226	MA	25017	Middlesex County	2161.855333	
1228	MA	25021	Norfolk County	2376.973667	
1324	MN	27019	Carver County	1207.509583	

1384	MN	27139	Scott County	1241.794833
1776	NJ	34003	Bergen County	1513.822083
1784	NJ	34019	Hunterdon County	1500.925583
1787	NJ	34025	Monmouth County	1413.447750
1788	NJ	34027	Morris County	1628.298833
1792	NJ	34035	Somerset County	1689.090833
1811	NM	35028	Los Alamos County	995.000000
1858	NY	36059	Nassau County	1710.035000
1868	NY	36079	Putnam County	2394.327000
1880	NY	36103	Suffolk County	1706.139667
2064	OH	39041	Delaware County	1145.778167
2259	PA	42029	Chester County	1550.335917
2522	TN	47187	Williamson County	1832.267750
2566	TX	48085	Collin County	1648.510750
2653	TX	48259	Kendall County	1229.000000
2722	TX	48397	Rockwall County	1746.943000
2792	UT	49029	Morgan County	1079.666250
2799	UT	49043	Summit County	1383.000000
2803	UT	49051	Wasatch County	981.000000
2827	VA	51013	Arlington County	2020.760250
2849	VA	51059	Fairfax County	1980.789083
2850	VA	51061	Fauquier County	1386.552417
2857	VA	51075	Goochland County	1488.633667
2873	VA	51107	Loudoun County	2058.066583
2882	VA	51127	New Kent County	1570.332833
2893	VA	51153	Prince William County	1778.269083
2906	VA	51179	Stafford County	1709.874000
2925	VA	51600	Fairfax city	1884.858583
2926	VA	51610	Falls Church city	1958.583000
2970	WA	53033	King County	2252.594083
3139	WY	56039	Teton County	1317.000000

	m_food	m_transport	m_health	m_other	m_tax	\
68	449.346642	1025.811917	728.583250	605.065508	725.969400	
187	420.204917	1063.247417	532.777833	755.048908	1057.811583	
193	418.324792	1126.897750	537.777833	755.475342	1087.298500	
207	494.469317	1096.417667	529.777833	932.884417	1372.521333	
224	483.188642	711.081950	574.777833	880.143417	1137.178833	
227	436.185833	1169.506750	584.777833	1004.936167	1539.703000	
229	413.624508	1117.580167	453.777833	940.942667	1362.795917	
252	388.243000	1037.334500	404.777833	772.216958	1016.534667	
263	401.403808	1149.467500	404.777833	783.141275	1073.220333	
265	365.681683	1354.404333	404.777833	567.261275	840.198000	
445	390.123125	1189.301583	454.053017	795.568925	1191.887750	
495	368.501833	1209.641833	486.053017	524.411217	813.644775	
726	368.501833	1091.121250	439.677733	538.124917	703.860592	
1086	334.659833	1240.897583	436.053017	460.223350	730.045083	

```
1195  381.662642  1114.850750  352.053017   711.222983   975.627917
1197  392.943275  1297.733750  352.053017   640.562867   940.403917
1201  360.981408  1202.138750  352.053017   688.384925   974.388000
1203  376.962325  1156.409000  352.053017   636.680258   884.557667
1206  409.864300  1095.976750  352.053017   734.769367  1002.710250
1208  424.905192  1001.162750  352.053017   749.640383   989.722417
1211  362.861492  1234.711667  352.053017   616.678267   884.151583
1226  420.204917  1041.497583  393.842775   868.751083  1189.015250
1228  432.425617  1091.623250  389.842775   945.240833  1317.078083
1324  438.065958  1207.299667  346.677733   553.664717   764.247150
1384  392.943275  1235.801500  346.677733   550.018392   752.868900
1776  403.283892  1020.270833  462.842775   645.022908   800.355717
1784  383.542725  1187.359417  462.842775   634.041708   850.352417
1787  403.283892  1083.700833  455.842775   611.251300   772.246908
1788  406.104083  1142.305083  462.842775   684.488200   906.075083
1792  396.703492  1140.378167  455.842775   701.779133   927.602500
1811  386.362917  1141.841167  601.583208   464.768592   688.548017
1858  394.823408  1117.987667  656.842775   708.193358  1112.916750
1868  406.104083  1238.034583  649.842775   942.223417  1499.033333
1880  396.703492  1182.969250  656.842775   707.515383  1136.226917
2064  361.921467  1141.394667  450.677733   507.275433   650.373575
2259  412.684450  1124.296333  346.842775   660.471067   852.803917
2522  423.965125  1144.049833  469.053017   759.124350   867.807167
2566  354.401000  1032.091167  433.053017   673.892825   721.306400
2653  328.079450  1227.222583  453.053017   523.889608   611.830117
2722  346.880575  1139.969417  433.053017   704.480717   789.509117
2792  344.060383  1399.862667  443.777833   479.022175   794.006675
2799  480.368450  1261.528250  418.583208   626.942542   952.193250
2803  363.801558  1233.820250  418.583208   452.467283   698.090658
2827  471.907958   864.354750  366.053017   838.674583  1103.238833
2849  433.365683  1032.460750  366.053017   812.258217  1126.515000
2850  377.902383  1244.002833  366.053017   593.662392   884.338000
2857  356.281125  1247.267500  379.053017   620.733683   930.045000
2873  417.384725  1155.098250  366.053017   832.881833  1203.570750
2882  355.341067  1294.734417  379.053017   647.905600   988.713667
2893  384.482792  1174.547917  366.053017   727.672042  1055.347583
2906  373.202108  1225.552583  366.053017   700.864667  1035.137500
2925  413.624508  1004.960750  366.053017   773.339683  1059.659083
2926  433.365683   870.210667  366.053017   804.786783  1056.068167
2970  423.965125   955.555167  376.777833   900.545917   956.448000
3139  475.668175  1252.413583  674.583208   603.154992   774.953450

          m_total  …  st_cost_rank  st_med_aff_rank  st_income_rank  \
68    4883.776667  …             5             27.0             2.0
187   5653.005833  …             6             49.0             5.0
193   5752.836583  …             7             39.0             6.0
207   6704.275417  …             2             53.0             1.0
```

224	5919.102250	…	3	43.0	4.0
227	7285.746750	…	1	46.0	2.0
229	6671.721333	…	4	55.0	3.0
252	5526.009750	…	11	62.0	2.0
263	5738.222000	…	7	61.0	1.0
265	4852.628250	…	15	59.0	5.0
445	5995.363250	…	1	157.0	1.0
495	4592.380833	…	19	159.0	2.0
726	4372.173500	…	1	92.0	1.0
1086	4235.072583	…	7	120.0	1.0
1195	5267.617500	…	3	12.0	4.0
1197	5134.603500	…	7	24.0	3.0
1201	5262.949250	…	5	17.0	7.0
1203	4922.010417	…	8	21.0	5.0
1206	5369.355500	…	1	20.0	1.0
1208	5320.623667	…	2	19.0	2.0
1211	4920.456083	…	12	23.0	8.0
1226	6075.166667	…	5	14.0	2.0
1228	6553.184250	…	1	13.0	1.0
1324	4517.464833	…	2	87.0	1.0
1384	4520.104833	…	3	76.0	2.0
1776	4845.598333	…	6	17.0	5.0
1784	5019.064417	…	3	21.0	3.0
1787	4739.773417	…	7	19.0	4.0
1788	5230.114250	…	2	20.0	1.0
1792	5311.396833	…	1	18.0	2.0
1811	4278.103833	…	1	33.0	1.0
1858	5700.798833	…	5	62.0	1.0
1868	7129.565083	…	1	44.0	4.0
1880	5786.397167	…	4	56.0	3.0
2064	4257.420917	…	1	88.0	1.0
2259	4947.434583	…	1	67.0	1.0
2522	5496.266917	…	1	95.0	1.0
2566	4863.255250	…	4	249.0	1.0
2653	4373.074833	…	9	248.0	4.0
2722	5160.835917	…	1	243.0	3.0
2792	4540.396167	…	4	28.0	2.0
2799	5122.615917	…	1	29.0	1.0
2803	4147.763000	…	9	27.0	3.0
2827	5664.989583	…	1	120.0	3.0
2849	5751.442083	…	6	128.0	4.0
2850	4852.511083	…	22	131.0	7.0
2857	5022.014000	…	15	125.0	10.0
2873	6033.055333	…	5	132.0	2.0
2882	5236.080417	…	19	117.0	18.0
2893	5486.372417	…	8	114.0	9.0
2906	5410.683917	…	10	127.0	8.0

```
2925  5502.495417  …             7            126.0             5.0
2926  5489.067750  …             3            133.0             1.0
2970  5865.886083  …             1             38.0             1.0
3139  5097.773417  …             1             13.0             1.0

      min_wage     m_income  a_min_income   m_ratio  a_min_ratio  \
68       11.73  2033.200000       24398.4  0.416317     0.416317
187      16.00  2773.333333       33280.0  0.490594     0.490594
193      16.00  2773.333333       33280.0  0.482081     0.482081
207      16.00  2773.333333       33280.0  0.413666     0.413666
224      16.00  2773.333333       33280.0  0.468540     0.468540
227      17.06  2957.066667       35484.8  0.405870     0.405870
229      16.00  2773.333333       33280.0  0.415685     0.415685
252      14.42  2499.466667       29993.6  0.452309     0.452309
263      14.42  2499.466667       29993.6  0.435582     0.435582
265      14.42  2499.466667       29993.6  0.515075     0.515075
445       7.25  1256.666667       15080.0  0.209606     0.209606
495       7.25  1256.666667       15080.0  0.273642     0.273642
726       7.25  1256.666667       15080.0  0.287424     0.287424
1086      7.25  1256.666667       15080.0  0.296728     0.296728
1195     15.00  2600.000000       31200.0  0.493582     0.493582
1197     15.00  2600.000000       31200.0  0.506368     0.506368
1201     15.00  2600.000000       31200.0  0.494020     0.494020
1203     15.00  2600.000000       31200.0  0.528239     0.528239
1206     15.00  2600.000000       31200.0  0.484229     0.484229
1208     17.15  2972.666667       35672.0  0.558706     0.558706
1211     15.00  2600.000000       31200.0  0.528406     0.528406
1226     15.00  2600.000000       31200.0  0.427972     0.427972
1228     15.00  2600.000000       31200.0  0.396754     0.396754
1324     10.85  1880.666667       22568.0  0.416310     0.416310
1384     10.85  1880.666667       22568.0  0.416067     0.416067
1776     15.13  2622.533333       31470.4  0.541220     0.541220
1784     15.13  2622.533333       31470.4  0.522514     0.522514
1787     15.13  2622.533333       31470.4  0.553304     0.553304
1788     15.13  2622.533333       31470.4  0.501429     0.501429
1792     15.13  2622.533333       31470.4  0.493756     0.493756
1811     12.00  2080.000000       24960.0  0.486197     0.486197
1858     16.00  2773.333333       33280.0  0.486482     0.486482
1868     15.00  2600.000000       31200.0  0.364679     0.364679
1880     16.00  2773.333333       33280.0  0.479285     0.479285
2064     10.45  1811.333333       21736.0  0.425453     0.425453
2259      7.25  1256.666667       15080.0  0.254004     0.254004
2522      7.25  1256.666667       15080.0  0.228640     0.228640
2566      7.25  1256.666667       15080.0  0.258400     0.258400
2653      7.25  1256.666667       15080.0  0.287365     0.287365
2722      7.25  1256.666667       15080.0  0.243501     0.243501
2792      7.25  1256.666667       15080.0  0.276775     0.276775
```

2799	7.25	1256.666667	15080.0	0.245317	0.245317
2803	7.25	1256.666667	15080.0	0.302975	0.302975
2827	12.00	2080.000000	24960.0	0.367167	0.367167
2849	12.00	2080.000000	24960.0	0.361648	0.361648
2850	12.00	2080.000000	24960.0	0.428644	0.428644
2857	12.00	2080.000000	24960.0	0.414176	0.414176
2873	12.00	2080.000000	24960.0	0.344767	0.344767
2882	12.00	2080.000000	24960.0	0.397244	0.397244
2893	12.00	2080.000000	24960.0	0.379121	0.379121
2906	12.00	2080.000000	24960.0	0.384425	0.384425
2925	12.00	2080.000000	24960.0	0.378010	0.378010
2926	12.00	2080.000000	24960.0	0.378935	0.378935
2970	16.28	2821.866667	33862.4	0.481064	0.481064
3139	7.25	1256.666667	15080.0	0.246513	0.246513

	median_household_income	median_ratio
68	112218	1.914809
187	121190	1.786513
193	119667	1.733449
207	135960	1.689966
224	135366	1.905779
227	143795	1.644707
229	150502	1.879850
252	114746	1.730393
263	140768	2.044303
265	120394	2.067505
445	130909	1.819587
495	114758	2.082398
726	117068	2.231308
1086	116173	2.285931
1195	112525	1.780137
1197	118110	1.916896
1201	112932	1.788161
1203	116796	1.977444
1206	133068	2.065239
1208	118020	1.848468
1211	112154	1.899451
1226	118494	1.625388
1228	115770	1.472185
1324	117602	2.169395
1384	115687	2.132823
1776	113885	1.958564
1784	137334	2.280206
1787	117699	2.069350
1788	131562	2.096226
1792	134735	2.113929
1811	139217	2.711813

```
1858                    135528      1.981126
1868                    112167      1.311055
1880                    119253      1.717433
2064                    121528      2.378748
2259                    117326      1.976209
2522                    133486      2.023889
2566                    113943      1.952447
2653                    120238      2.291256
2722                    124291      2.006958
2792                    119482      2.192944
2799                    132358      2.153164
2803                    115793      2.326415
2827                    131020      1.927335
2849                    144632      2.095590
2850                    120301      2.065958
2857                    113617      1.885316
2873                    167605      2.315093
2882                    115627      1.840228
2893                    119051      1.808283
2906                    134456      2.070841
2925                    122790      1.859611
2926                    142513      2.163588
2970                    116044      1.648572
3139                    127677      2.087137

[55 rows x 29 columns]
```

[132]:
```
datadfoutlier6 = datadf[datadf['median_household_income']<=14500]
datadfoutlier6
```

[132]:
```
Empty DataFrame
Columns: [state abv., county_fips, county, m_housing, m_food, m_transport,
m_health, m_other , m_tax, m_total, a_housing, a_food, a_transport, a_health,
a_other, a_taxes, a_total_cost, median_family_income, num_counties_in_st,
st_cost_rank, st_med_aff_rank, st_income_rank, min_wage, m_income, a_min_income,
m_ratio, a_min_ratio, median_household_income, median_ratio]
Index: []

[0 rows x 29 columns]
```

[133]:
```
datadfoutlier7 = datadf[datadf['median_ratio']>=2.18]
datadfoutlier7
```

[133]:
```
     state abv.  county_fips             county    m_housing      m_food  \
726          IN        18057    Hamilton County  1230.887167  368.501833
814          IA        19049      Dallas County   921.501833  380.722575
1086         KY        21185      Oldham County  1033.193833  334.659833
```

```
1784         NJ        34019   Hunterdon County  1500.925583  383.542725
1811         NM        35028  Los Alamos County   995.000000  386.362917
2064         OH        39041    Delaware County  1145.778167  361.921467
2071         OH        39055      Geauga County   822.186442  397.643558
2126         OH        39165      Warren County   994.050333  342.180300
2653         TX        48259     Kendall County  1229.000000  328.079450
2792         UT        49029      Morgan County  1079.666250  344.060383
2803         UT        49051     Wasatch County   981.000000  363.801558
2873         VA        51107     Loudoun County  2058.066583  417.384725

      m_transport    m_health    m_other         m_tax      m_total  …  \
726   1091.121250  439.677733  538.124917   703.860592  4372.173500  …
814   1142.270833  405.677733  438.141967   653.790850  3942.105833  …
1086  1240.897583  436.053017  460.223350   730.045083  4235.072583  …
1784  1187.359417  462.842775  634.041708   850.352417  5019.064417  …
1811  1141.841167  601.583208  464.768592   688.548017  4278.103833  …
2064  1141.394667  450.677733  507.275433   650.373575  4257.420917  …
2071  1194.671833  442.677733  410.419800   572.328692  3839.928083  …
2126  1112.372917  435.677733  449.583575   587.420250  3921.285167  …
2653  1227.222583  453.053017  523.889608   611.830117  4373.074833  …
2792  1399.862667  443.777833  479.022175   794.006675  4540.396167  …
2803  1233.820250  418.583208  452.467283   698.090658  4147.763000  …
2873  1155.098250  366.053017  832.881833  1203.570750  6033.055333  …

      st_cost_rank  st_med_aff_rank  st_income_rank  min_wage     m_income  \
726              1             92.0             1.0      7.25  1256.666667
814              1             99.0             1.0      7.25  1256.666667
1086             7            120.0             1.0      7.25  1256.666667
1784             3             21.0             3.0     15.13  2622.533333
1811             1             33.0             1.0     12.00  2080.000000
2064             1             88.0             1.0     10.45  1811.333333
2071             3             84.0             4.0     10.45  1811.333333
2126             2             86.0             2.0     10.45  1811.333333
2653             9            248.0             4.0      7.25  1256.666667
2792             4             28.0             2.0      7.25  1256.666667
2803             9             27.0             3.0      7.25  1256.666667
2873             5            132.0             2.0     12.00  2080.000000

      a_min_income   m_ratio  a_min_ratio  median_household_income  \
726        15080.0  0.287424     0.287424                   117068
814        15080.0  0.318781     0.318781                   109138
1086       15080.0  0.296728     0.296728                   116173
1784       31470.4  0.522514     0.522514                   137334
1811       24960.0  0.486197     0.486197                   139217
2064       21736.0  0.425453     0.425453                   121528
2071       21736.0  0.471710     0.471710                   101636
2126       21736.0  0.461923     0.461923                   104523
```

```
2653       15080.0  0.287365     0.287365                  120238
2792       15080.0  0.276775     0.276775                  119482
2803       15080.0  0.302975     0.302975                  115793
2873       24960.0  0.344767     0.344767                  167605

      median_ratio
726       2.231308
814       2.307100
1086      2.285931
1784      2.280206
1811      2.711813
2064      2.378748
2071      2.205684
2126      2.221274
2653      2.291256
2792      2.192944
2803      2.326415
2873      2.315093

[12 rows x 29 columns]
```

[135]:
```
datadfoutlier8 = datadf[datadf['median_ratio']<=.77]
datadfoutlier8
```

[135]:
```
     state abv.  county_fips            county  m_housing      m_food  \
65           AL         1131     Wilcox County   779.0000  320.559000
1427         MS        28051     Holmes County   704.0000  322.439108
1428         MS        28053  Humphreys County   775.0000  322.439108
1831         NY        36005      Bronx County  1885.0075  378.842450
2370         SD        46017    Buffalo County   720.0000  344.060383
3016         WV        54047   McDowell County   668.0000  313.978617

      m_transport    m_health    m_other        m_tax      m_total  …  \
65    1133.738833  502.199867  369.953817  606.010942  3711.462583  …
1427  1076.569083  532.053017  345.352175  556.212242  3536.625667  …
1428  1091.783250  522.199867  369.240600  581.478067  3662.140917  …
1831   531.612633  695.842775  761.687175  989.748250  5242.740583  …
2370  1171.214917  585.727500  358.010092  491.912642  3670.925417  …
3016  1030.745167  755.199867  330.393150  600.056967  3698.373667  …

      st_cost_rank  st_med_aff_rank  st_income_rank  min_wage     m_income  \
65              15              6.0            60.0      7.25  1256.666667
1427            20              2.0            80.0      7.25  1256.666667
1428            68              3.0            81.0      7.25  1256.666667
1831            14              1.0            62.0     15.00  2600.000000
2370            35              8.0            59.0     11.20  1941.333333
3016            40              1.0            55.0      8.75  1516.666667
```

```
      a_min_income   m_ratio  a_min_ratio  median_household_income  \
65         15080.0  0.338591     0.338591                    33682
1427       15080.0  0.355329     0.355329                    31972
1428       15080.0  0.343151     0.343151                    33148
1831       31200.0  0.495924     0.495924                    45864
2370       23296.0  0.528840     0.528840                    30659
3016       18200.0  0.410090     0.410090                    28972

      median_ratio
65        0.756261
1427      0.753355
1428      0.754295
1831      0.729008
2370      0.695987
3016      0.652809

[6 rows x 29 columns]
```

[140]:
```
datadf['a_min_income'].mode
```

[140]:
```
<bound method Series.mode of 0       15080.0
1       15080.0
2       15080.0
3       15080.0
4       15080.0
         ...
3138    15080.0
3139    15080.0
3140    15080.0
3141    15080.0
3142    15080.0
Name: a_min_income, Length: 3143, dtype: float64>
```

[141]:
```
datadf['a_total_cost'].mode
```

[141]:
```
<bound method Series.mode of 0       48384.410
1       50978.832
2       40947.977
3       43405.910
4       41932.285
          ...
3138    49277.215
3139    61173.281
3140    44581.816
3141    44393.578
3142    48843.875
```

```
Name: a_total_cost, Length: 3143, dtype: float64>
```

[142]:
```
datadf['a_min_ratio'].mode
```

[142]:
```
<bound method Series.mode of 0       0.311671
1       0.295809
2       0.368272
3       0.347418
4       0.359627
          …
3138    0.306024
3139    0.246513
3140    0.338255
3141    0.339689
3142    0.308739
Name: a_min_ratio, Length: 3143, dtype: float64>
```

[143]:
```
datadf['median_household_income'].mode
```

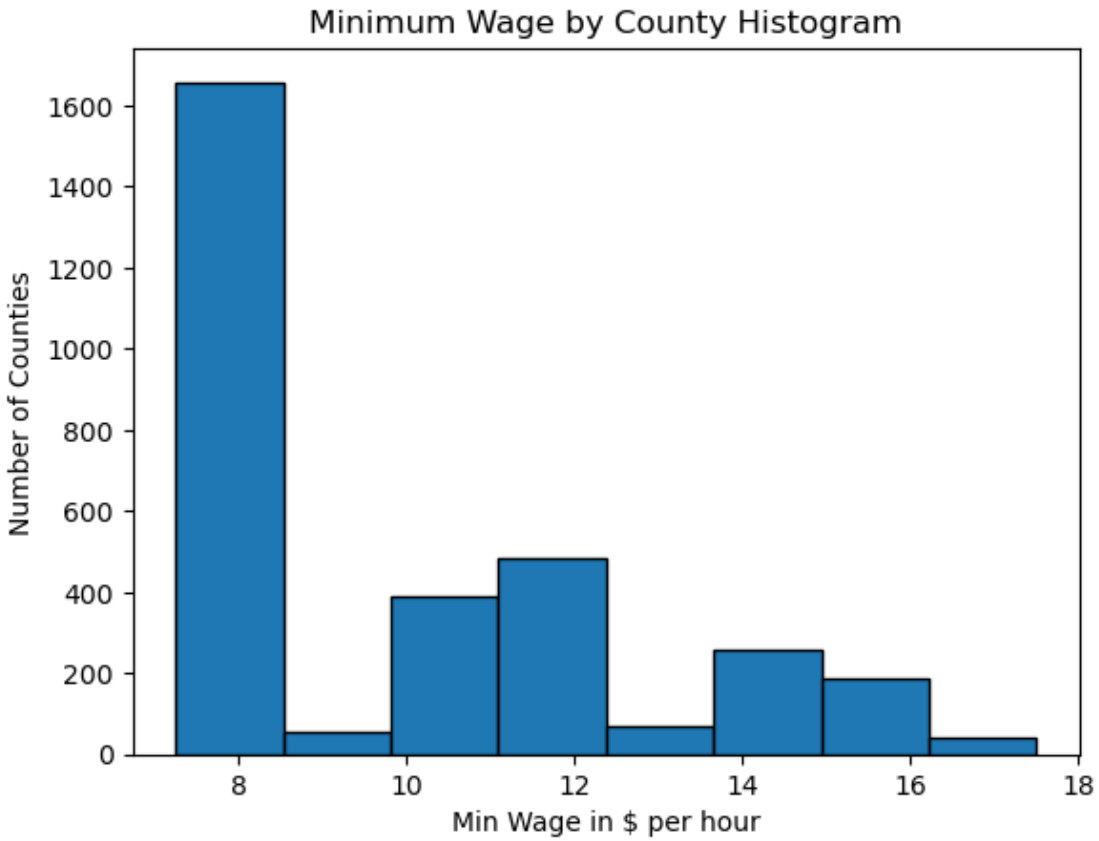
[143]:
```
<bound method Series.mode of 0        70148
1        71704
2        41151
3        54309
4        60553
         …
3138     75779
3139    127677
3140     73072
3141     60699
3142     67677
Name: median_household_income, Length: 3143, dtype: int64>
```

[144]:
```
datadf['median_ratio'].mode
```

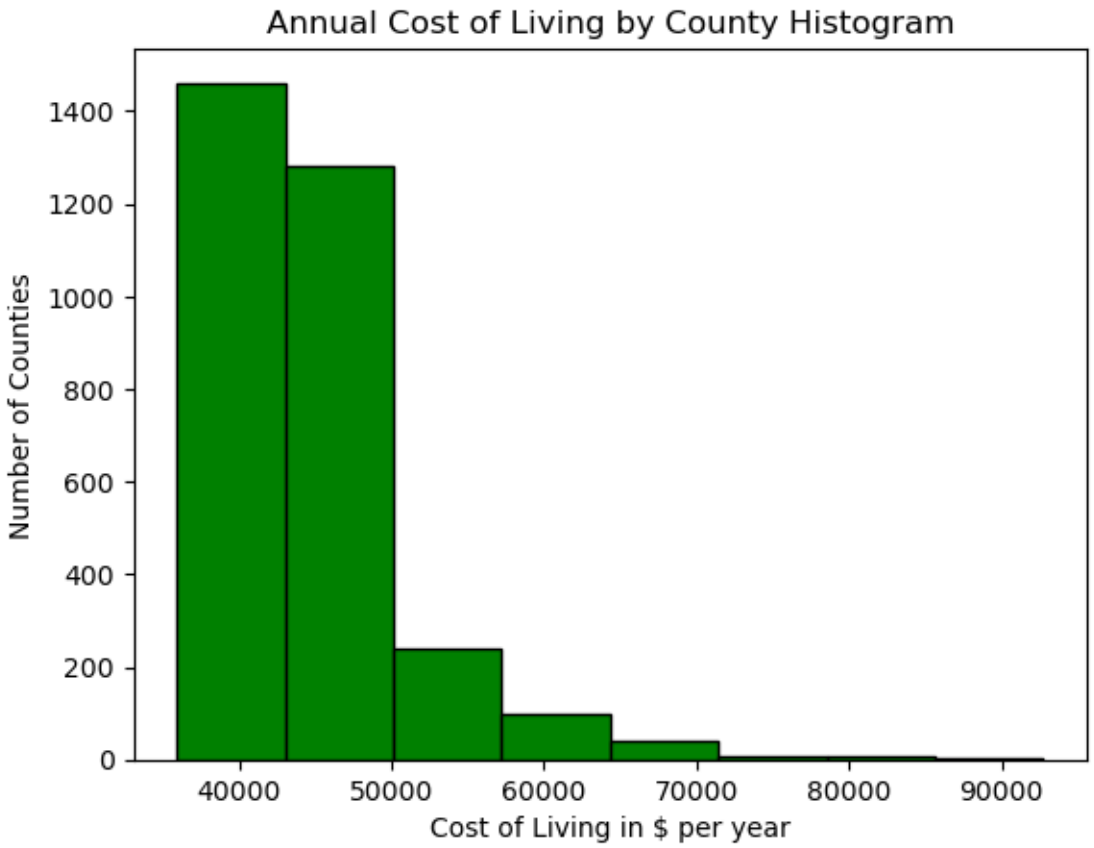
[144]:
```
<bound method Series.mode of 0       1.449806
1       1.406545
2       1.004958
3       1.251189
4       1.444066
          …
3138    1.537810
3139    2.087137
3140    1.639054
3141    1.367292
3142    1.385578
Name: median_ratio, Length: 3143, dtype: float64>
```

1.0.3 Part 4 - Creating Histograms

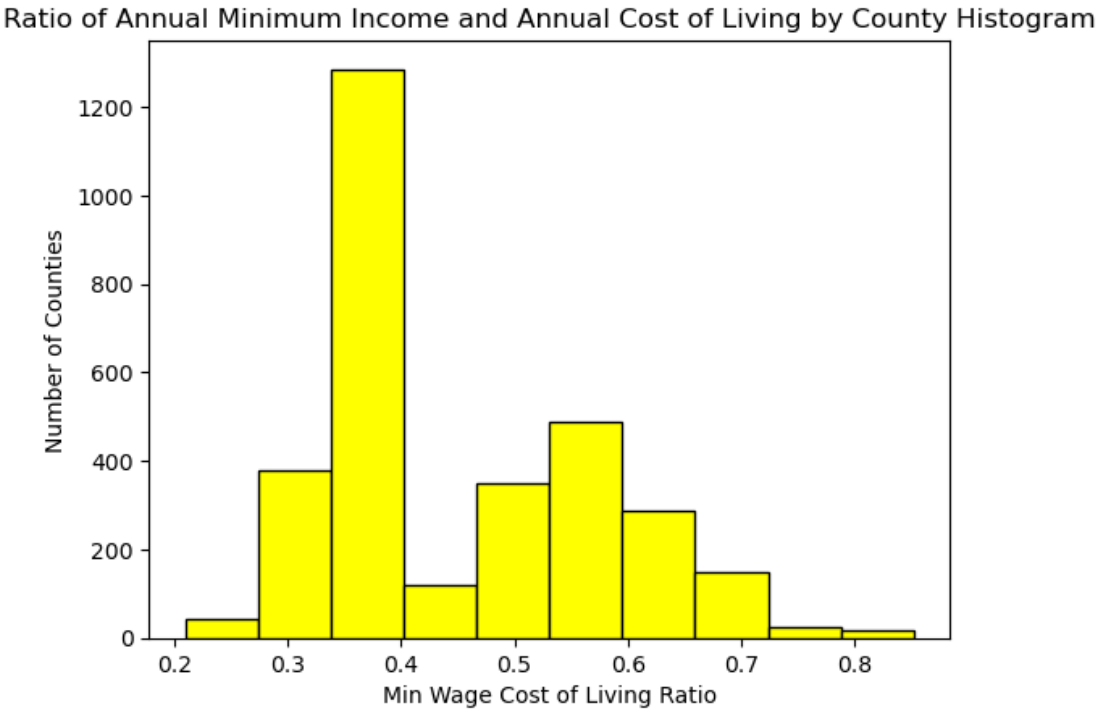
[77]:
```
plt.hist(data['min_wage'], bins = 8, edgecolor='black')
plt.xlabel('Min Wage in $ per hour')
plt.ylabel('Number of Counties')
plt.title('Minimum Wage by County Histogram')

min_wage_hist = plt
min_wage_hist.show()
```



[108]:
```
plt.hist(data['a_total_cost'], bins = 8, color='green', edgecolor='black')
plt.xlabel('Cost of Living in $ per year')
plt.ylabel('Number of Counties')
plt.title('Annual Cost of Living by County Histogram')

a_total_cost_hist = plt
a_total_cost_hist.show()
```

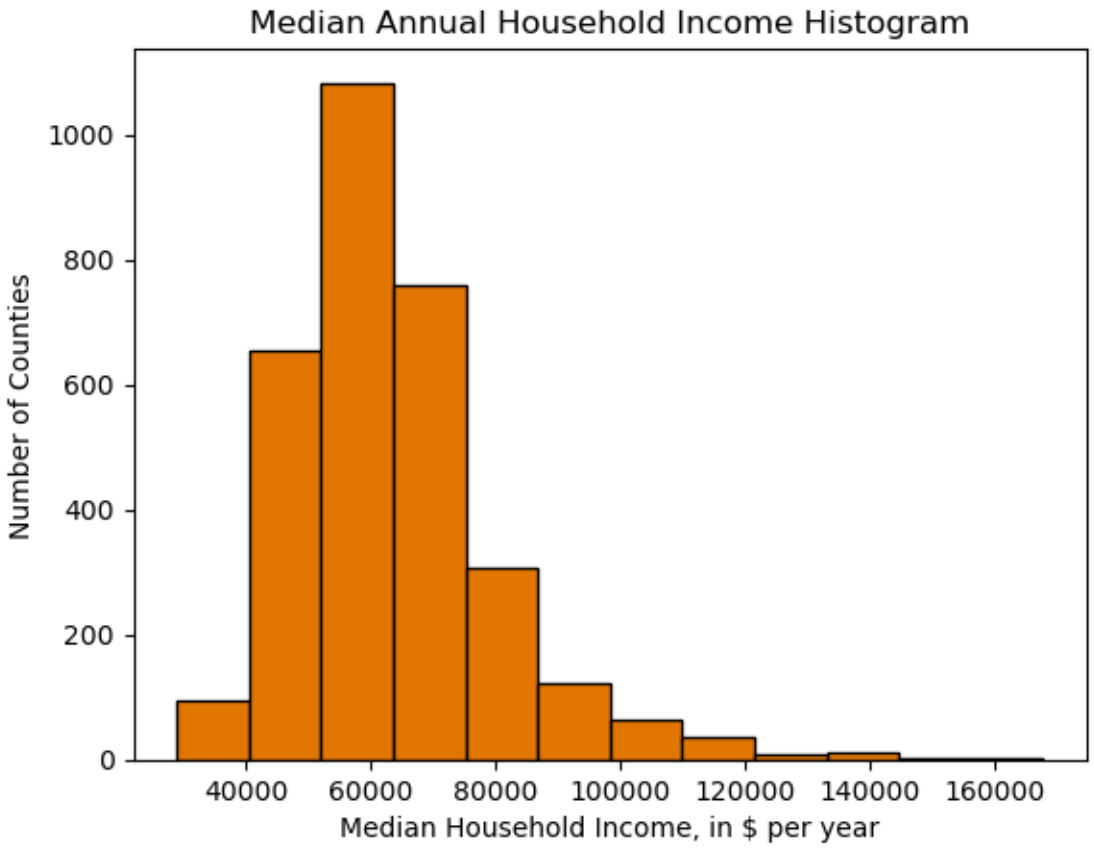


[105]:
```
plt.hist(data['a_min_ratio'], bins = 10, color='yellow', edgecolor='black')
plt.xlabel('Min Wage Cost of Living Ratio')
plt.ylabel('Number of Counties')
plt.title('Ratio of Annual Minimum Income and Annual Cost of Living by County␣
 ↪Histogram')

a_min_ratio_hist = plt
a_min_ratio_hist.show()
```

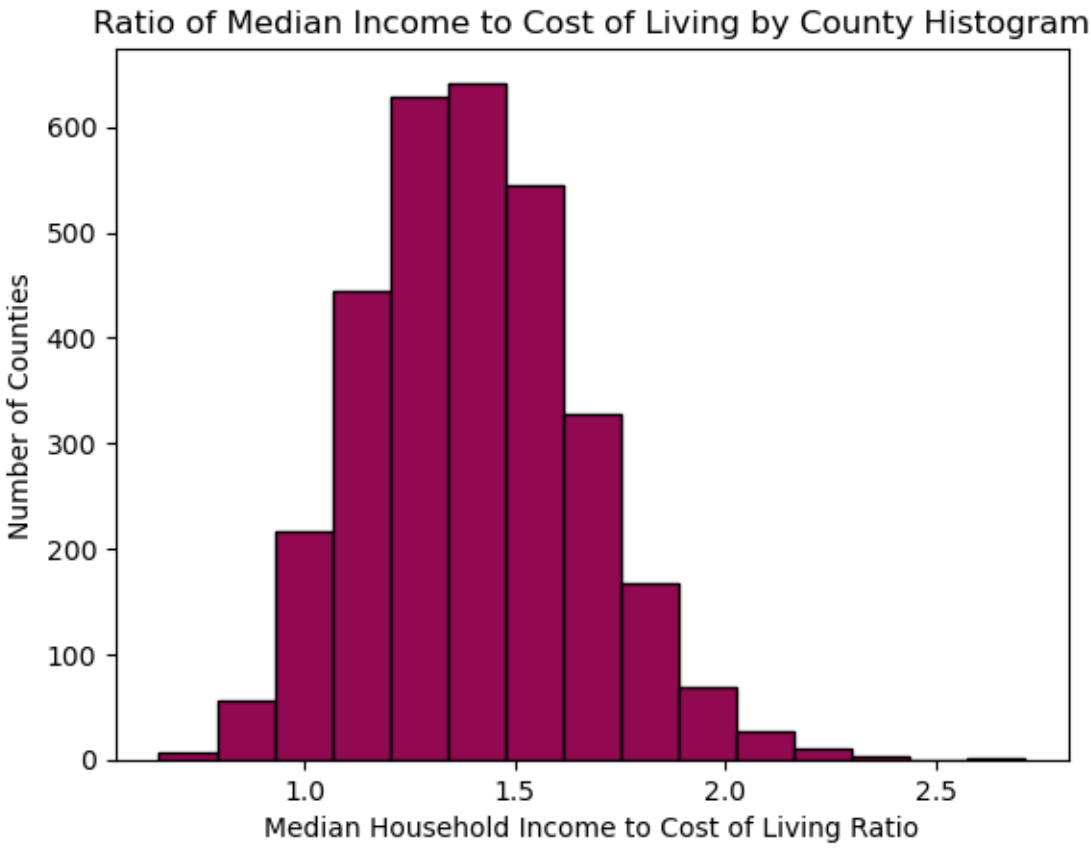


[ ]:

[89]:
```
plt.hist(data['median_household_income'], bins = 10, color='orange',␣
  ↪edgecolor='black')
plt.xlabel('Median Household Income, in $ per year')
plt.ylabel('Number of Counties')
plt.title('Median Annual Household Income Histogram')

median_household_income_hist = plt
median_household_income_hist.show()
```



[106]:
```
plt.hist(data['median_ratio'], bins = 15, color='#910951', edgecolor='black')
plt.xlabel('Median Household Income to Cost of Living Ratio')
plt.ylabel('Number of Counties')
plt.title('Ratio of Median Income to Cost of Living by County Histogram')

median_household_income_hist = plt
median_household_income_hist.show()
```



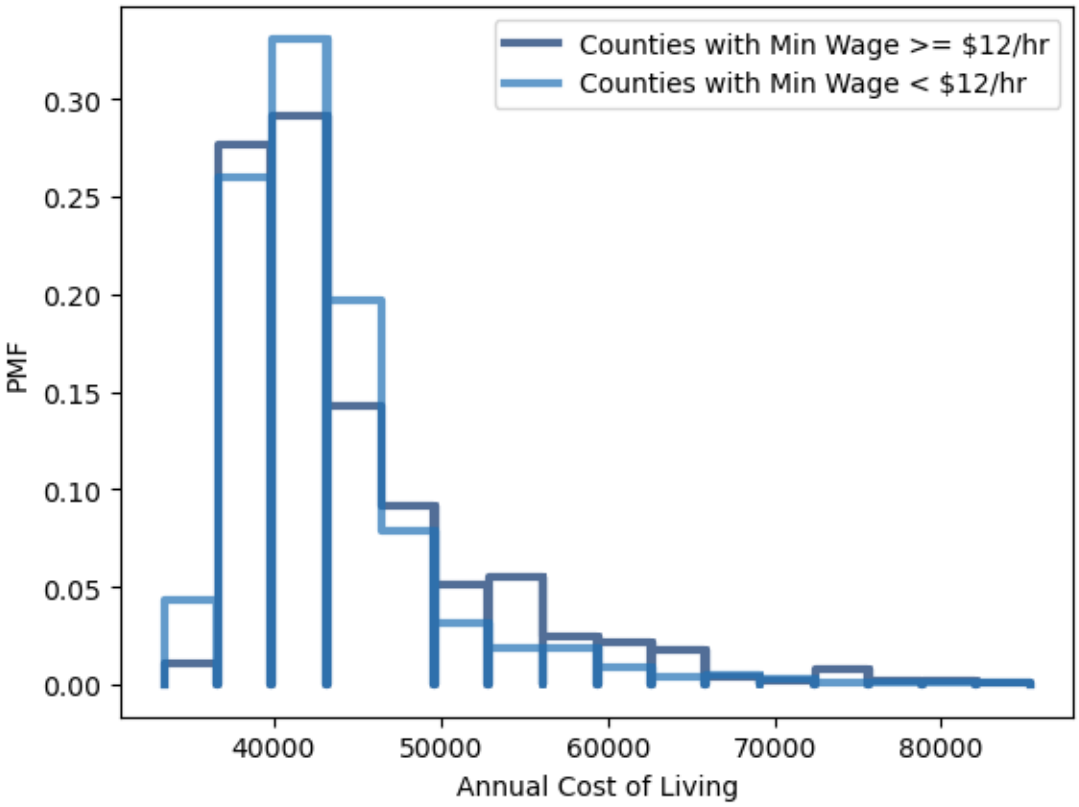
1.0.4 Part 5 - PMF

[244]:
```
import thinkplot
import thinkstats2
pmf1data = datadf[datadf['min_wage']>=12]
pmf2data = datadf[datadf['min_wage']<12]
```

[ ]:
```
```

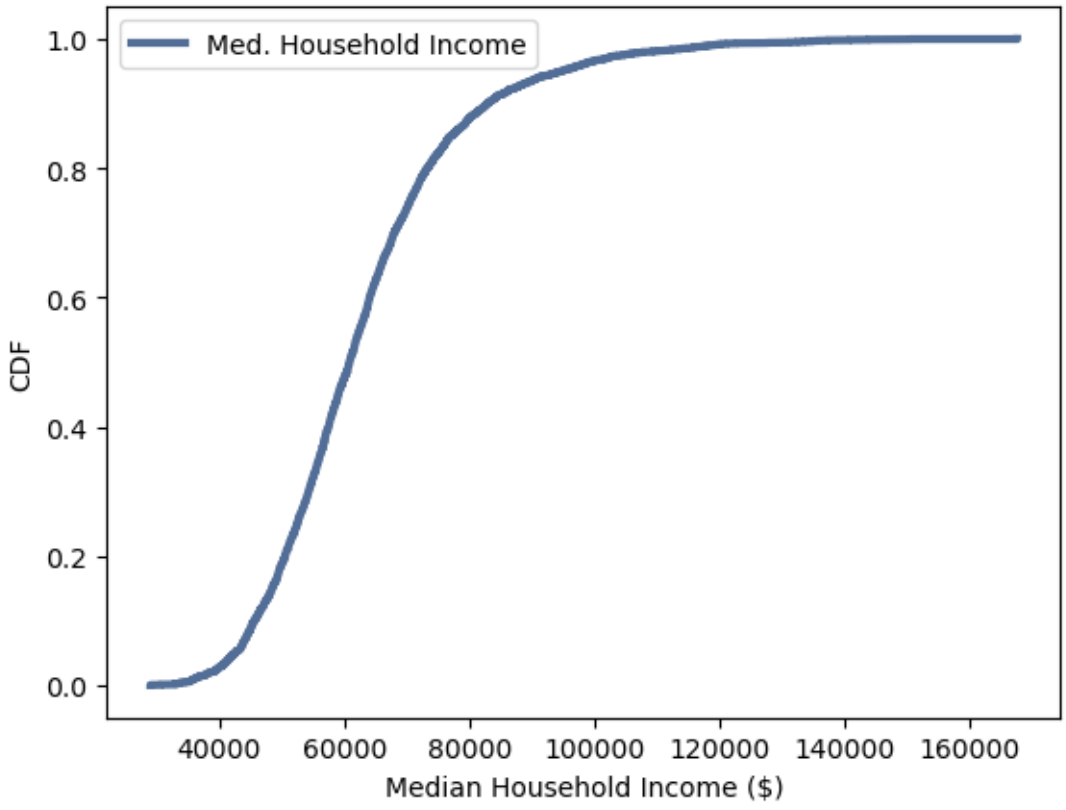
[245]:
```
bins = 16
binlabels = [35000, 38250, 41500, 44850, 48000, 51250, 54500, 57750, 61000,␣
 ↪64250, 67500, 70750, 74000, 77250, 80500, 83750]
pmf1bin = pd.cut(pmf1data.a_total_cost, bins=bins, labels=binlabels)
pmf2bin = pd.cut(pmf2data.a_total_cost, bins=bins, labels=binlabels)

cost1 = thinkstats2.Pmf(pmf1bin, label = "Counties with Min Wage >= $12/hr")
cost2 = thinkstats2.Pmf(pmf2bin, label = "Counties with Min Wage < $12/hr")


axis = [35000, 93000, 0, 0.4]
```

```
first_pmf = thinkplot.Pmf(cost1)
other_pmf = thinkplot.Pmf(cost2)
thinkplot.Config(xlabel="Annual Cost of Living", ylabel="PMF")
```



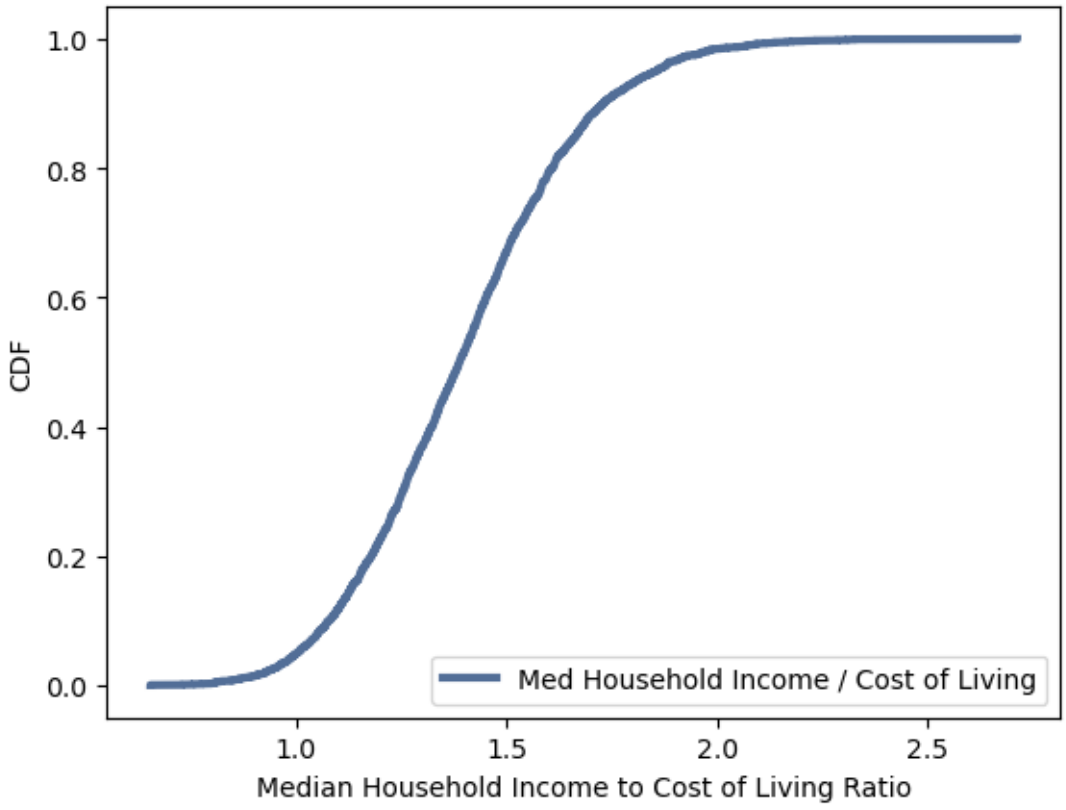
1.0.5 Part 6 - CDF

[259]:
```
cdf = thinkstats2.Cdf(datadf.median_household_income, label='Med. Household␣
↪Income')
thinkplot.Cdf(cdf)
thinkplot.Show(xlabel='Median Household Income ($)', ylabel="CDF")
```




```
<Figure size 800x600 with 0 Axes>
```

[256]:
```
cdf2 = thinkstats2.Cdf(datadf.median_ratio, label='Med Household Income / ')
thinkplot.Cdf(cdf2)
thinkplot.Show(xlabel='Median Household Income to Cost of Living Ratio',␣
↪ylabel="CDF")
```



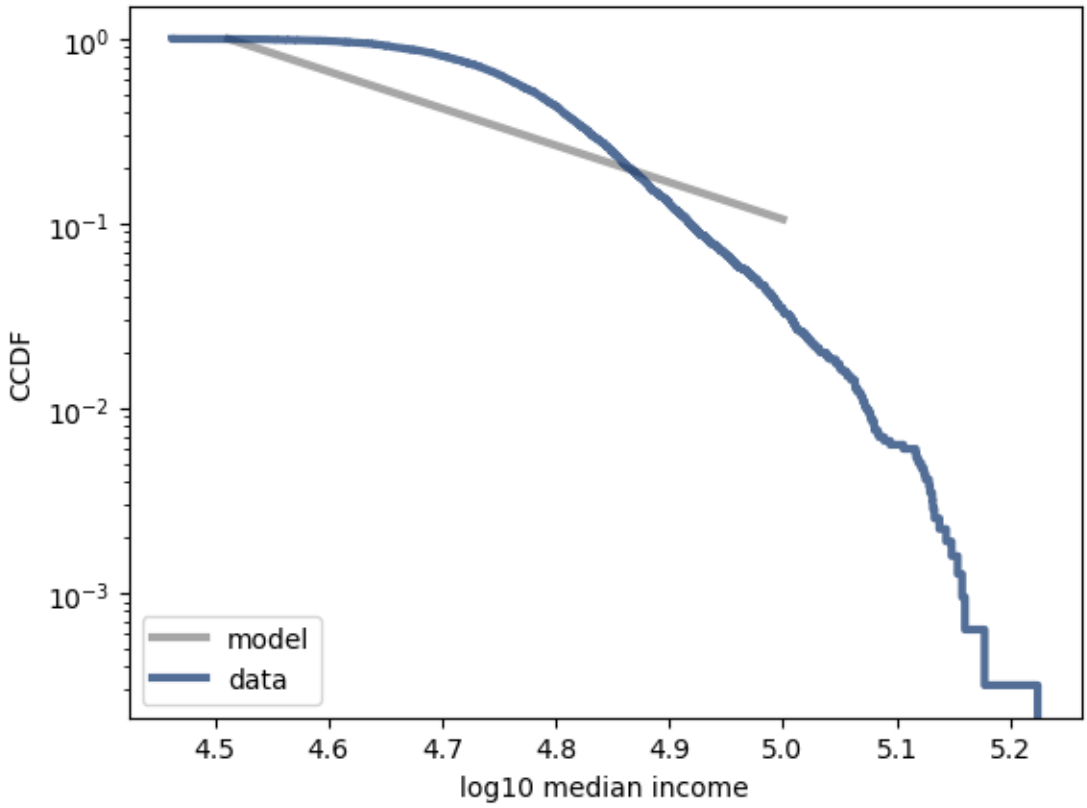

```
<Figure size 800x600 with 0 Axes>
```

1.0.6 Part 7 - Analytical Distribution

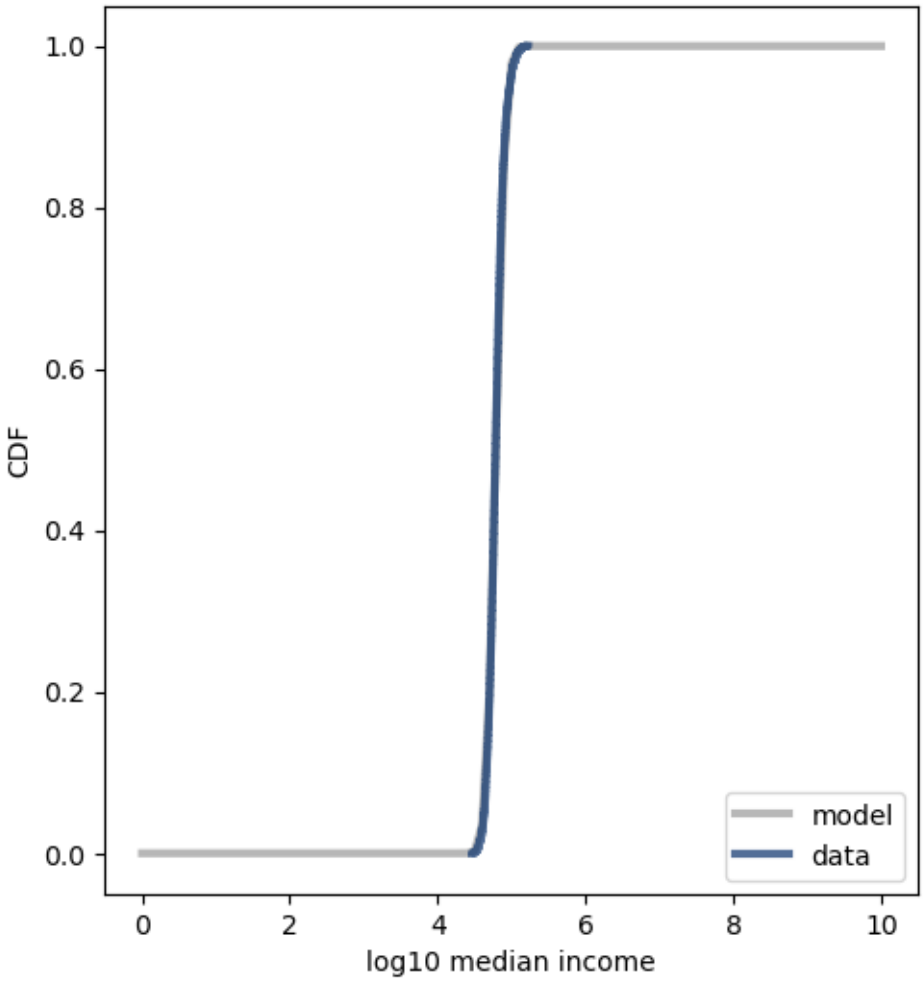
[276]:
```
import scipy.stats
```

[282]:
```
pareto_data = datadf

log_income = np.log10(pareto_data.median_household_income)
cdf = thinkstats2.Cdf(pareto_data.median_household_income)
cdf_log = thinkstats2.Cdf(log_income, label="data")
```

[319]:
```
xs, ys = thinkstats2.RenderParetoCdf(xmin=32500, alpha=2, low=0, high=1e5)
thinkplot.Plot(np.log10(xs), 1 - ys, label="model", color="0.5")

thinkplot.Cdf(cdf_log, complement=True)
thinkplot.Config(
    xlabel="log10 median income", ylabel="CCDF", yscale="log", loc="lower left")
```

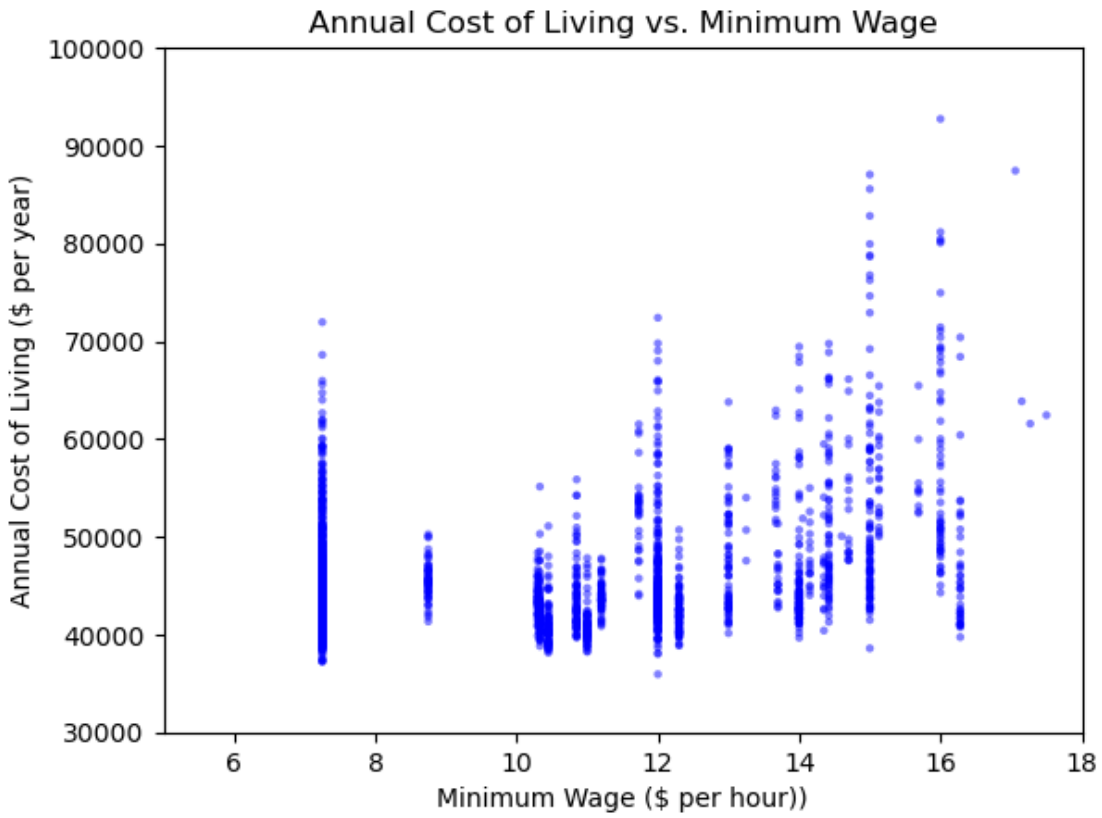


[325]:
```
thinkplot.PrePlot(cols=2)

mu, sigma = log_income.mean(), log_income.std()
xs, ps = thinkstats2.RenderNormalCdf(mu, sigma, low=0, high=10)
thinkplot.Plot(xs, ps, label="model", color="0.6")

thinkplot.Cdf(cdf_log)
thinkplot.Config(xlabel="log10 median income", ylabel="CDF", loc="lower right")
```



1.0.7 Part 8 Scatterplots and Correlation

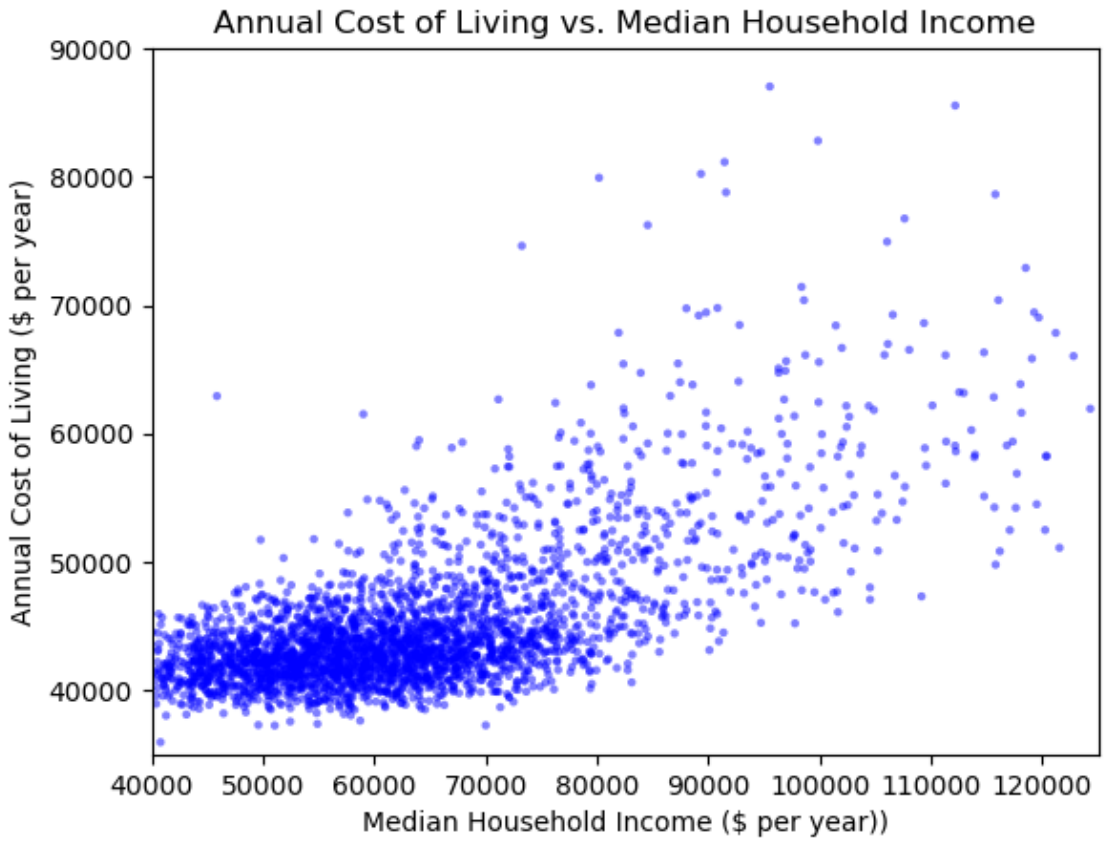
[357]:
```
thinkplot.Scatter(datadf.min_wage, datadf.a_total_cost, alpha=.5, s=10)
thinkplot.Config(xlabel='Minimum Wage ($ per hour))',
                     title = "Annual Cost of Living vs. Minimum Wage",
                     ylabel='Annual Cost of Living ($ per year)',
                     xlim=[5, 18],
                     ylim=[30000, 100000],
                     legend=False)
```



[367]:
```
print('Corr:  ', thinkstats2.Corr(datadf.min_wage, datadf.a_total_cost))
print('SpearmanCorr:  ', thinkstats2.SpearmanCorr(datadf.min_wage, datadf.
 ↪a_total_cost))
```

```
Corr:   0.3918515877955295
SpearmanCorr:   0.30506633739900596
```

[364]:
```
thinkplot.Scatter(datadf.median_household_income, datadf.a_total_cost, alpha=.
 ↪5, s=10)
thinkplot.Config(xlabel='Median Household Income ($ per year))',
                    title = "Annual Cost of Living vs. Median Household␣
 ↪Income",
                    ylabel='Annual Cost of Living ($ per year)',
                    xlim=[40000, 125000],
                    ylim=[35000, 90000],
                    legend=False)
```



[368]:
```
print('Corr:  ', thinkstats2.Corr(datadf.median_household_income, datadf.
 ↪a_total_cost))
print('SpearmanCorr:  ', thinkstats2.SpearmanCorr(datadf.
 ↪median_household_income, datadf.a_total_cost))
```

```
Corr:   0.6967097054309237
SpearmanCorr:   0.5506802578331028
```

1.0.8 Part 8 - Hypothesis Testing

[374]:
```
class CorrelationPermute(thinkstats2.HypothesisTest):
    def TestStatistic(self, data):
        xs, ys = data
        test_stat = abs(thinkstats2.Corr(xs, ys))
        return test_stat
    def RunModel(self):
        xs, ys = self.data
        xs = np.random.permutation(xs)
        return xs, ys
```

```
data = datadf.min_wage, datadf.a_total_cost
ht = CorrelationPermute(data)
pvalue = ht.PValue()
pvalue
```

[374]: 0.0

[375]:
```
data = datadf.median_household_income, datadf.a_total_cost
ht = CorrelationPermute(data)
pvalue = ht.PValue()
pvalue
```

[375]: 0.0

1.0.9 Part 9 - Regression Analysis

[380]:
```
import first
```

[381]:
```
median_income = datadf.median_household_income
a_total_cost = datadf.a_total_cost
min_wage = datadf.min_wage
```

[382]:
```
from thinkstats2 import Mean, MeanVar, Var, Std, Cov

def LeastSquares(xs, ys):
    meanx, varx = MeanVar(xs)
    meany = Mean(ys)

    slope = Cov(xs, ys, meanx, meany) / varx
    inter = meany - slope * meanx

    return inter, slope
```

[402]:
```
inter, slope = LeastSquares(median_income, a_total_cost)
inter, slope
```

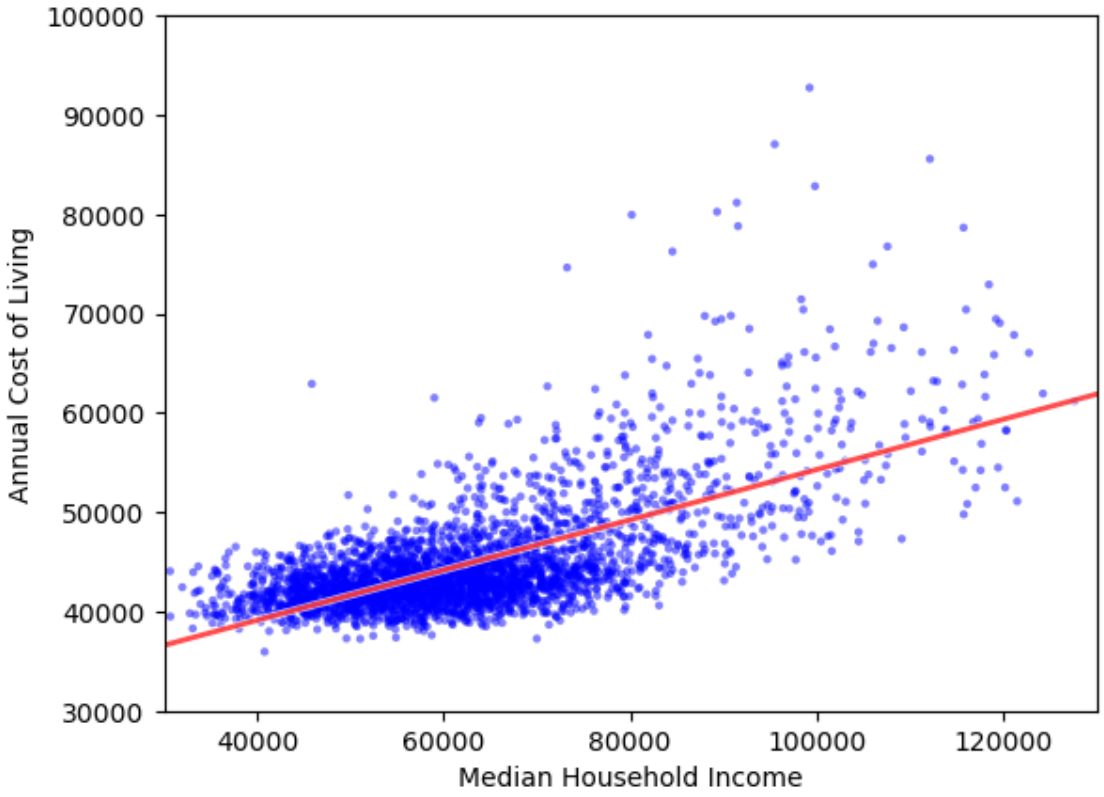
[402]: (29005.920369609048, 0.2527433731692815)

[403]:
```
def FitLine(xs, inter, slope):
    fit_xs = np.sort(xs)
    fit_ys = inter + slope * fit_xs
    return fit_xs, fit_ys
```

[404]:
```
fit_xs, fit_ys = FitLine(median_income, inter, slope)
```

[409]:
```
thinkplot.Scatter(median_income, a_total_cost, color='blue', alpha=0.5, s=10)
thinkplot.Plot(fit_xs, fit_ys, color='white', linewidth=3)
thinkplot.Plot(fit_xs, fit_ys, color='red', linewidth=2)
```

```
thinkplot.Config(xlabel="Median Household Income",
                 ylabel="Annual Cost of Living",
                 axis=[30000, 130000, 30000, 100000],
                 legend=False)
```



[ ]: