

In [3]:

```
import pandas as pd
import numpy as np
```

In [4]:

```
entire_dataset=pd.read_csv("C:/Users/iGuest/Desktop/capstone/jul15-cps.csv")
statecode= pd.read_csv("C:/Users/iGuest/Desktop/capstone/statecode.csv")
```

In [5]:

```
#merge on state code file to get state name
entire_dataset = entire_dataset.merge(statecode, left_on="gestfips", right_
on = "gestfips")
```

In [6]:

```
#filter for valid responses
entire_dataset = entire_dataset[(entire_dataset['hufinal'].isin([1,201]))]
```

In [7]:

```
entire_dataset.loc[(entire_dataset.statename == 'Washington')][['peinhome',
'pwsswgt']]
```

Out[7]:

	peinhome	pwsswgt
138784	1	3071.1304
138785	1	3057.0532
138786	1	3354.5851
138787	1	2780.9683
138788	1	2931.8577
138789	1	3862.7681
138790	1	3885.3392
138791	1	2598.1616
138792	1	3407.1026
138793	2	3085.8550
138794	1	3071.7103
138795	1	2415.8610
138796	1	2838.8285
138797	1	3803.0397
138798	1	3014.5550
.....

138799	1	2564.6102
	peinhome	pwsswgt
138800	1	2694.8204
138801	1	2305.7409
138802	1	3363.4157
138803	2	3152.2795
138804	2	1834.2159
138805	1	2706.5193
138806	1	2442.3452
138807	1	2462.8431
138808	1	3668.9001
138809	1	2509.9832
138810	1	2509.2191
138811	1	3535.1789
138812	1	5095.2877
138813	1	2399.2974
...
141453	1	3101.7532
141454	1	2625.0206
141455	1	3091.9058
141456	1	4205.8786
141457	1	4294.1757
141478	1	2318.9475
141479	1	2898.7854
141480	1	2927.5993
141481	1	2643.2019
141482	1	2260.7668
141483	1	2670.0333
141484	1	2136.6029
141494	1	2788.0867
141495	1	2571.4411
141496	2	2597.7029
141508	2	3156.9772
141516	1	2565.5234
141517	1	2280.9371
141520	1	3857.2141
141521	1	3348.7149

	peinnome	peinwork	peinschl	peincare	peintrav	peinlico	peinemo	peinot
	peinhome	peinwork	peinschl	peincare	peintrav	peinlico	peinemo	peinot
Alabama	0.629927	0.535999	0.139548	0.112929	0.269916	0.136625	0.167080	0.0145
Connecticut	0.730460	0.602454	0.163063	0.144981	0.289031	0.178794	0.241344	0.0240
Delaware	0.710205	0.614612	0.175661	0.134346	0.274951	0.183985	0.206997	0.0149
District of Columbia	0.693995	0.730714	0.141291	0.296337	0.435275	0.289874	0.336317	0.0354
Florida	0.654104	0.524739	0.132852	0.124857	0.267825	0.150422	0.155587	0.0086
Georgia	0.660626	0.568667	0.178719	0.157263	0.321905	0.188771	0.213904	0.0240
Kentucky	0.690126	0.613805	0.152854	0.146349	0.318101	0.159568	0.245427	0.0122
Louisiana	0.662222	0.602156	0.164859	0.119264	0.308424	0.172627	0.226328	0.0094
Maine	0.733251	0.610817	0.157925	0.179939	0.360689	0.186992	0.289234	0.0161
Maryland	0.738112	0.623906	0.147954	0.163669	0.336915	0.195664	0.236005	0.0337
Massachusetts	0.718393	0.610173	0.143977	0.170839	0.332020	0.187322	0.270849	0.0233
Mississippi	0.580970	0.485789	0.180342	0.092739	0.288613	0.151409	0.143703	0.0065
New Hampshire	0.758157	0.562828	0.150410	0.153519	0.325188	0.176186	0.265363	0.0239
New Jersey	0.727526	0.606213	0.158808	0.142244	0.345197	0.191392	0.253173	0.0176
New York	0.662380	0.527307	0.133979	0.130499	0.270936	0.174636	0.202034	0.0210
North Carolina	0.626169	0.522758	0.152732	0.139399	0.301267	0.171511	0.195720	0.0167
Ohio	0.683280	0.555901	0.150687	0.162785	0.306242	0.165416	0.276809	0.0150
Pennsylvania	0.657448	0.504800	0.135021	0.127498	0.299476	0.147247	0.248104	0.0219
Rhode Island	0.742476	0.566198	0.162751	0.110731	0.293745	0.160393	0.232734	0.0169
South Carolina	0.656880	0.547408	0.147703	0.130451	0.299314	0.154383	0.211481	0.0229
Tennessee	0.623800	0.504293	0.120433	0.161330	0.282552	0.141470	0.204494	0.0090
Vermont	0.764492	0.607879	0.161848	0.179490	0.352794	0.159755	0.226935	0.0245
Virginia	0.744316	0.614434	0.158276	0.180420	0.374828	0.176301	0.249916	0.0304
Alaska	0.731942	0.596780	0.163647	0.177131	0.358015	0.179390	0.237144	0.0217
Arizona	0.633164	0.535762	0.146640	0.125453	0.299119	0.132426	0.185819	0.0146
Arkansas	0.634456	0.541187	0.151554	0.124503	0.257032	0.143005	0.182818	0.0152
California	0.667593	0.543021	0.149883	0.171736	0.322876	0.170521	0.212865	0.0183
Colorado	0.677894	0.587608	0.161383	0.220210	0.371008	0.222769	0.268372	0.0231
Hawaii	0.656322	0.536142	0.145581	0.151038	0.307856	0.175082	0.192073	0.0128
Idaho	0.758961	0.610194	0.177638	0.187666	0.370909	0.177534	0.268507	0.0222
Illinois	0.756225	0.613118	0.186239	0.196791	0.381674	0.210674	0.265699	0.0242
Indiana	0.710463	0.554247	0.165697	0.158659	0.314151	0.181813	0.220955	0.0246
Iowa	0.723697	0.594916	0.173239	0.183771	0.330843	0.208387	0.295459	0.0210

Kansas	0.667587 peinhome	0.547179 peinwork	0.173728 peinschl	0.132574 peincafe	0.314975 peintrav	0.140691 peinlico	0.213402 peinelho	0.0148 peinot
Michigan	0.692647	0.532147	0.159604	0.182728	0.329059	0.201242	0.266528	0.0198
Minnesota	0.762292	0.562353	0.170665	0.188674	0.381237	0.164602	0.276534	0.0228
Missouri	0.685157	0.589794	0.183080	0.177337	0.356478	0.178265	0.255338	0.0300
Montana	0.686998	0.574962	0.150417	0.183396	0.345318	0.188454	0.266126	0.0136
Nebraska	0.689911	0.557749	0.199346	0.142707	0.329156	0.183732	0.232254	0.0219
Nevada	0.728277	0.575199	0.150090	0.158484	0.333308	0.158015	0.174120	0.0082
New Mexico	0.609734	0.523280	0.166398	0.166518	0.285727	0.176115	0.164343	0.0165
North Dakota	0.664474	0.532281	0.156016	0.147848	0.296399	0.144763	0.240966	0.0114
Oklahoma	0.608840	0.545356	0.155950	0.140589	0.277342	0.141050	0.197461	0.0058
Oregon	0.750196	0.599305	0.161619	0.243446	0.378818	0.201725	0.281408	0.0340
South Dakota	0.630944	0.505575	0.148655	0.116610	0.260813	0.108784	0.181143	0.0152
Texas	0.638100	0.548029	0.161536	0.161501	0.304964	0.170462	0.198628	0.0166
Utah	0.734308	0.590362	0.175005	0.144565	0.363999	0.194711	0.288839	0.0613
Washington	0.726934	0.608084	0.132525	0.217984	0.397176	0.214669	0.283313	0.0203
West Virginia	0.633607	0.534745	0.141520	0.132283	0.263294	0.107946	0.191940	0.0122
Wisconsin	0.773273	0.566364	0.183717	0.186267	0.349961	0.193458	0.281064	0.0125

For Section 2: Bar and Pie Chart

In [12]:

```
def ageband(row):
    if row['prtage'] in range(7,23):
        return '07-22'
    elif row['prtage'] in range(23,39):
        return '23-38'
    elif row['prtage'] in range(39,53):
        return '39-52'
    elif row['prtage'] in range(53,72):
        return '53-71'
    else:
        return '72+'
#add a column for age band using prtage
entire_dataset['ageband'] = entire_dataset.apply(ageband, axis=1)
```

In [13]:

```
pie_df = entire_dataset.groupby(['statename','ageband'])
pie_df = pd.DataFrame(pie_df.size())
pie_df = pie_df.reset_index()
pie_df = pie_df.rename(index=str, columns={0: "age_count"})
```

In [14]:

```
def count_age_loc(location,df,pie_df):
```

```

filtered_df = df.loc[df[location] == 1]
count_list = filtered_df.groupby(['statename', 'ageband']).count()[location].tolist()
if len(count_list) == pie_df.shape[0]:
    pie_df.insert(loc=3, column=location+'_count', value=count_list)

```

In [16]:

```

for i in internet_by_location:
    count_age_loc(i, entire_dataset, pie_df)

```

In [17]:

```

#find states with null values
filtered_df = entire_dataset.loc[entire_dataset['peinothr'] == 1]
count_list = pd.DataFrame(filtered_df.groupby(['statename', 'ageband'])['peinothr'].count())
count_list = count_list.reset_index()
for i in states:
    if count_list.loc[count_list.statename==i].ageband.count() != 5:
        print(i)

```

Delaware
Kansas
West Virginia

In [18]:

```

#insert zero for null value for 'peinothr' column
filtered_df = entire_dataset.loc[entire_dataset['peinothr'] == 1]
count_list = filtered_df.groupby(['statename', 'ageband']).count()['peinothr'].tolist()
count_list.insert(35, 0)
count_list.insert(84, 0)
count_list.insert(244, 0)
pie_df.insert(loc=3, column='peinothr_count', value=count_list)

```

In [19]:

```
pie_df
```

Out[19]:

	statename	ageband	age_count	peinothr_count	peinelho_count	peinlico_count	peintr
0	Alabama	07-22	515	7	113	97	123
1	Alabama	23-38	450	3	105	82	176
2	Alabama	39-52	474	10	103	75	168
3	Alabama	53-71	636	11	66	49	150
4	Alabama	72+	440	5	11	15	32
5	Alaska	07-22	359	7	86	67	96
6	Alaska	23-38	384	4	113	84	176
7	Alaska	39-52	272	9	76	54	124
8	Alaska	53-71	312	10	54	49	106

9	Alaska	72+	266	4	26	16	27
	state name	age band	age count	peinothr count	peinelho count	peinlico count	peintr
10	Arizona	07-22	483	5	97	93	102
11	Arizona	23-38	425	7	112	75	177
12	Arizona	39-52	332	3	50	29	115
13	Arizona	53-71	449	10	62	39	134
14	Arizona	72+	369	1	32	15	40
15	Arkansas	07-22	496	7	124	98	122
16	Arkansas	23-38	467	11	128	89	172
17	Arkansas	39-52	439	7	76	65	124
18	Arkansas	53-71	555	8	54	34	127
19	Arkansas	72+	395	1	28	25	35
20	California	07-22	2454	33	548	563	590
21	California	23-38	2409	66	690	515	1039
22	California	39-52	2093	40	465	329	851
23	California	53-71	2314	43	408	290	753
24	California	72+	1861	16	156	107	196
25	Colorado	07-22	331	3	105	85	85
26	Colorado	23-38	360	13	126	96	188
27	Colorado	39-52	331	9	98	88	155
28	Colorado	53-71	351	11	72	60	119
29	Colorado	72+	261	1	17	20	26
...
220	Utah	07-22	520	26	151	97	142
221	Utah	23-38	486	46	196	132	245
222	Utah	39-52	300	20	85	60	140
223	Utah	53-71	335	21	57	44	116
224	Utah	72+	319	4	50	34	36
225	Vermont	07-22	297	6	89	51	84
226	Vermont	23-38	296	9	86	55	144
227	Vermont	39-52	305	8	76	49	132
228	Vermont	53-71	535	13	92	84	177
229	Vermont	72+	274	6	23	20	40
230	Virginia	07-22	458	16	134	84	132
231	Virginia	23-38	467	10	155	93	235
232	Virginia	39-52	430	15	111	86	202
233	Virginia	53-71	500	24	100	80	204

233	Virginia	55-71	560	21	106	88	204
	statename	ageband	age_count	peinothr_count	peinelho_count	peinlico_count	peintr
234	Virginia	72+	415	5	38	32	46
235	Washington	07-22	449	4	140	115	123
236	Washington	23-38	532	16	202	137	295
237	Washington	39-52	433	10	130	111	241
238	Washington	53-71	587	16	130	91	217
239	Washington	72+	412	2	43	31	41
240	West Virginia	07-22	576	8	171	80	177
241	West Virginia	23-38	527	6	142	84	211
242	West Virginia	39-52	562	10	110	63	178
243	West Virginia	53-71	827	8	84	66	151
244	West Virginia	72+	538	0	35	14	32
245	Wisconsin	07-22	384	4	134	93	111
246	Wisconsin	23-38	307	4	135	79	168
247	Wisconsin	39-52	324	4	88	70	158
248	Wisconsin	53-71	481	9	89	65	130
249	Wisconsin	72+	286	2	25	18	27

250 rows × 11 columns



In [20]:

```
pie_df.loc[pie_df.statename == 'Washington']
```

Out[20]:

	statename	ageband	age_count	peinothr_count	peinelho_count	peinlico_count	peintr
235	Washington	07-22	449	4	140	115	123
236	Washington	23-38	532	16	202	137	295
237	Washington	39-52	433	10	130	111	241
238	Washington	53-71	587	16	130	91	217
239	Washington	72+	412	2	43	31	41



In [50]:

```
#pie_df.set_index(['statename','ageband'], inplace=True)
```


In [21]:

```
import json
import pprint
#pprint.pprint(pie_df.to_json(orient='table'))
```