SI 206 - Final Project Report

By Jade Stoler & Dylan Zgodny

Goals of Our Project

Initial Goal: Determining if there is a correlation between a state's obesity rate and the number of Planet Fitness gyms

- Find a Planet Fitness Gym near me
- Obesity rate by state
- Possible API's:
  - BMI
  - Exercise/Gym API

Visualizations through Plotly and code in Jupyter Notebooks

Goals That Were Achieved

Determine if there is a correlation between the number of Planet Fitness Gyms and obesity/leisure rates around the country within certain age groups

- Find a Planet Fitness Gym near me
- State Abbreviations API
- Obesity and Leisure Questions
  - Questions: Percent of adults who engage in no leisure-time physical activity and Percent of adults aged 18 years and older who have obesity

Visualizations through Plotly and code in Jupyter Notebooks

Problems We Faced

- Finding an API
  - Gym/BMI API → API for abbreviations of states
- Working with an API
  - Sifting through a lot of information
  - Creating and editing our queries
- Finding a correlation
  - Tested age, income, education, race, and gender
  - No clear correlation (as seen through most of the regression lines being horizontal)
- Testing with the obesity vs. leisure data
  - They were very similar

Calculations From The Data in The Database

Table with State, Population, and Abbreviations

```
population['abbr'] = population['Name'].map(us_state_to_abbrev)
```

	Name	Pop	State	abbr
0	Mississippi	2976149	28	MS
1	Missouri	6137428	29	MO
2	Montana	1068778	30	MT
3	Nebraska	1934408	31	NE
4	Nevada	3080156	32	NV
5	New Hampshire	1359711	33	NH
6	New Jersey	8882190	34	NJ
7	New Mexico	2096829	35	NM
8	New York	19453561	36	NY
9	North Carolina	10488084	37	NC
10	North Dakota	762062	38	ND
11	Ohio	11689100	39	OH
12	Oklahoma	3956971	40	OK
13	Oregon	4217737	41	OR
14	Pennsylvania	12801989	42	PA
15	Rhode Island	1059361	44	RI
16	South Carolina	5148714	45	SC
17	South Dakota	884659	46	SD
18	Tennessee	6829174	47	TN
19	Texas	28995881	48	TX
20	Vermont	623989	50	VT
21	Utah	3205958	49	UT
22	Virginia	8535519	51	VA
23	Washington	7614893	53	WA
24	West Virginia	1792147	54	WV
25	Wisconsin	5822434	55	WI
26	Wyoming	578759	56	WY

	Name	Pop	State	abbr
27	Puerto Rico	3193694	72	PR
28	Alabama	4903185	01	AL
29	Alaska	731545	02	AK
30	Arizona	7278717	04	AZ
31	Arkansas	3017804	05	AR
32	California	39512223	06	CA
33	Colorado	5758736	08	CO
34	Delaware	973764	10	DE
35	District of Columbia	705749	11	DC
36	Connecticut	3565287	09	CT
37	Florida	21477737	12	FL
38	Georgia	10617423	13	GA
39	Idaho	1787065	16	ID
40	Hawaii	1415872	15	HI
41	Illinois	12671821	17	IL
42	Indiana	6732219	18	IN
43	Iowa	3155070	19	IA
44	Kansas	2913314	20	KS
45	Kentucky	4467673	21	KY
46	Louisiana	4648794	22	LA
47	Maine	1344212	23	ME
48	Maryland	6045680	24	MD
49	Massachusetts	6892503	25	MA
50	Michigan	9986857	26	MI
51	Minnesota	5639632	27	MN

Leisure Data

```python
leisure = pd.DataFrame(results)
```

[86]:

	yearstart	yearend	locationabbr	locationdesc	datasource	class	topic	question	data_value_type	data_value	...	stratification1	stratificationcategoryid1	stratificationid1	income	age_years	gender	data_value_footnote_symbol	data_value_footnote	race_ethnicity	total
0	2021	2021	WA	Washington	Behavioral Risk Factor Surveillance System	Physical Activity	Physical Activity - Behavior	Percent of adults who engage in no leisure-tim...	Value	17.2	...	Some college or technical school	EDU	EDUCOTEC	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	2021	2021	UT	Utah	Behavioral Risk Factor Surveillance System	Physical Activity	Physical Activity - Behavior	Percent of adults who engage in no leisure-tim...	Value	25.4	...	25,000–34,999	INC	INC2535	25,000–34,999	NaN	NaN	NaN	NaN	NaN	NaN
2	2021	2021	WA	Washington	Behavioral Risk Factor Surveillance System	Physical Activity	Physical Activity - Behavior	Percent of adults who engage in no leisure-tim...	Value	37.0	...	Less than high school	EDU	EDUHS	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	2021	2021	NE	Nebraska	Behavioral Risk Factor Surveillance System	Physical Activity	Physical Activity - Behavior	Percent of adults who engage in no leisure-tim...	Value	12.6	...	College graduate	EDU	EDUCOGRAD	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	2021	2021	WI	Wisconsin	Behavioral Risk Factor Surveillance System	Physical Activity	Physical Activity - Behavior	Percent of adults who engage in no leisure-tim...	Value	35.0	...	Less than high school	EDU	EDUHS	NaN	NaN	NaN	NaN	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1535	2021	2021	NH	New Hampshire	Behavioral Risk Factor Surveillance System	Physical Activity	Physical Activity - Behavior	Percent of adults who engage in no leisure-tim...	Value	21.5	...	Female	GEN	FEMALE	NaN	NaN	Female	NaN	NaN	NaN	NaN
1536	2021	2021	NJ	New Jersey	Behavioral Risk Factor Surveillance System	Physical Activity	Physical Activity - Behavior	Percent of adults who engage in no leisure-tim...	Value	34.1	...	Hispanic	RACE	RACEHIS	NaN	NaN	NaN	NaN	NaN	Hispanic	NaN
1537	2021	2021	PR	Puerto Rico	Behavioral Risk Factor Surveillance System	Physical Activity	Physical Activity - Behavior	Percent of adults who engage in no leisure-tim...	Value	37.4	...	35,000–49,999	INC	INC3550	35,000–49,999	NaN	NaN	NaN	NaN	NaN	NaN
1538	2021	2021	WI	Wisconsin	Behavioral Risk Factor Surveillance System	Physical Activity	Physical Activity - Behavior	Percent of adults who engage in no leisure-tim...	Value	11.8	...	$75,000 or greater	INC	INC75PLUS	$75,000 or greater	NaN	NaN	NaN	NaN	NaN	NaN
1539	2021	2021	US	National	Behavioral Risk Factor Surveillance System	Physical Activity	Physical Activity - Behavior	Percent of adults who engage in no leisure-tim...	Value	19.0	...	Data not reported	INC	INCNR	Data not reported	NaN	NaN	NaN	NaN	NaN	NaN

1540 rows × 32 columns

Obesity Weight Classification Table

```
obesity = pd.DataFrame(results)
```

[9]:

	yearstart	yearend	locationabbr	locationdesc	datasource	class	topic	question	data_value_type	data_value	...	stratification1	stratificationcategoryid1	stratificationid1	education	income	race_ethnicity	gender	data_value_footnote_symbol	data_value_footnote	total
0	2021	2021	NH	New Hampshire	Behavioral Risk Factor Surveillance System	Obesity / Weight Status	Obesity / Weight Status	Percent of adults aged 18 years and older who ...	Value	41.1	...	65 or older	AGEYR	AGEYR65PLUS	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	2021	2021	SD	South Dakota	Behavioral Risk Factor Surveillance System	Obesity / Weight Status	Obesity / Weight Status	Percent of adults aged 18 years and older who ...	Value	27.8	...	Less than high school	EDU	EDUHS	Less than high school	NaN	NaN	NaN	NaN	NaN	NaN
2	2021	2021	VI	Virgin Islands	Behavioral Risk Factor Surveillance System	Obesity / Weight Status	Obesity / Weight Status	Percent of adults aged 18 years and older who ...	Value	27.5	...	50,000–74,999	INC	INC5075	NaN	50,000 – 74,999	NaN	NaN	NaN	NaN	NaN
3	2021	2021	SD	South Dakota	Behavioral Risk Factor Surveillance System	Obesity / Weight Status	Obesity / Weight Status	Percent of adults aged 18 years and older who ...	Value	34.6	...	35,000–49,999	INC	INC3550	NaN	35,000 – 49,999	NaN	NaN	NaN	NaN	NaN
4	2021	2021	WY	Wyoming	Behavioral Risk Factor Surveillance System	Obesity / Weight Status	Obesity / Weight Status	Percent of adults aged 18 years and older who ...	Value	37.7	...	Data not reported	INC	INCNR	NaN	Data not reported	NaN	NaN	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1535	2021	2021	US	National	Behavioral Risk Factor Surveillance System	Obesity / Weight Status	Obesity / Weight Status	Percent of adults aged 18 years and older who ...	Value	35.9	...	45 - 54	AGEYR	AGEYR4554	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1536	2021	2021	LA	Louisiana	Behavioral Risk Factor	Obesity /	Obesity /	Percent of adults aged 18 years	Value	31.7	...	55 - 64	AGEYR	AGEYR5564	NaN	NaN	NaN	NaN	NaN	NaN	NaN

Merged Table with Primary Key (State_ID_fk)

```
df = df.merge(population[['abbr','State']],left_on = 'state',right_on = 'abbr').rename(columns = {'State':'State_ID_fk'})
```

```
df
```

	cc	state	city	zip	a1	abbr	State_ID_fk
0	US	RI	Pawtucket	02860	50 Ann Mary St	RI	44
1	US	RI	Warwick	02888	1222 Warwick Ave	RI	44
2	US	RI	Woonsocket	02895	1500 Diamond Hill Rd	RI	44
3	US	RI	East Providence	02914	585 Taunton Ave	RI	44
4	US	RI	Cumberland	02864	1800 Mendon Rd	RI	44
...	...	...	...	...	...	...	...
2369	US	MT	Helena	59601	3035 N Sanders St	MT	30
2370	US	MT	Missoula	59808	2640 N Reserve St	MT	30
2371	US	MT	Great Falls	59405	726 10th Ave S	MT	30
2372	US	MT	Billings	59102	640 S 24th Street W	MT	30
2373	US	MT	Billings	59105	1321 Main St	MT	30

2374 rows × 7 columns

Number of Gyms Table

	n_gyms	State_ID_fk	abbr	locationdesc	obesity_data	leisure_data	stratificationcategoryid1	stratificationid1	index	Name	Pop	State_ID_fk	abbr	gyms_per_100k
0	20	33	NH	New Hampshire	35.4	20.0	OVR	OVERALL	5	New Hampshire	1359711	33	NH	1.470901
1	3	46	SD	South Dakota	33.7	23.2	OVR	OVERALL	17	South Dakota	884659	46	SD	0.339114
2	4	56	WY	Wyoming	37.0	22.9	OVR	OVERALL	26	Wyoming	578759	56	WY	0.691134
3	137	36	NY	New York	34.5	25.8	OVR	OVERALL	8	New York	19453561	36	NY	0.704241
4	24	41	OR	Oregon	34.7	19.9	OVR	OVERALL	13	Oregon	4217737	41	OR	0.569026
5	38	01	AL	Alabama	31.9	31.5	OVR	OVERALL	28	Alabama	4903185	01	AL	0.775006
6	3	11	DC	District of Columbia	30.7	15.8	OVR	OVERALL	35	District of Columbia	705749	11	DC	0.425080
7	4	38	ND	North Dakota	34.8	25.5	OVR	OVERALL	10	North Dakota	762062	38	ND	0.524892
8	156	12	FL	Florida	NaN	NaN	OVR	OVERALL	37	Florida	21477737	12	FL	0.726334
9	173	48	TX	Texas	33.7	25.1	OVR	OVERALL	19	Texas	28995881	48	TX	0.596636
10	82	25	MA	Massachusetts	33.9	21.1	OVR	OVERALL	49	Massachusetts	6892503	25	MA	1.189698
11	23	40	OK	Oklahoma	32.5	28.3	OVR	OVERALL	12	Oklahoma	3956971	40	OK	0.581253
12	5	15	HI	Hawaii	35.3	20.0	OVR	OVERALL	40	Hawaii	1415872	15	HI	0.353139
13	20	35	NM	New Mexico	35.4	23.0	OVR	OVERALL	7	New Mexico	2096829	35	NM	0.953821
14	43	45	SC	South Carolina	34.5	25.0	OVR	OVERALL	16	South Carolina	5148714	45	SC	0.835160
15	5	30	MT	Montana	35.0	21.2	OVR	OVERALL	2	Montana	1068778	30	MT	0.467824
16	94	17	IL	Illinois	34.5	27.2	OVR	OVERALL	41	Illinois	12671821	17	IL	0.741803
17	55	04	AZ	Arizona	36.6	22.7	OVR	OVERALL	30	Arizona	7278717	04	AZ	0.755628

Visualizations We Created

Gyms vs Obesity Scatterplot (by age category)
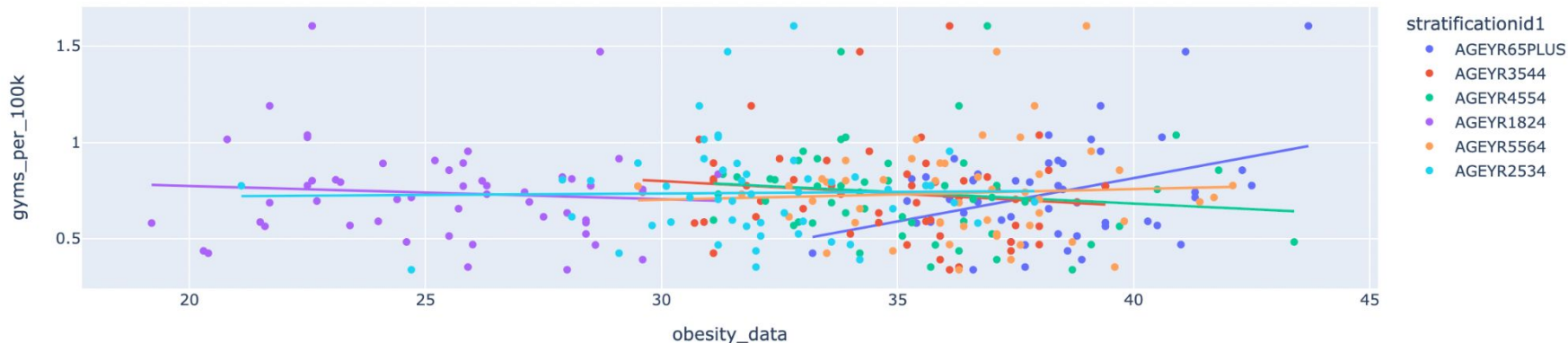
Gyms vs Leisure Scatterplot (by age category)
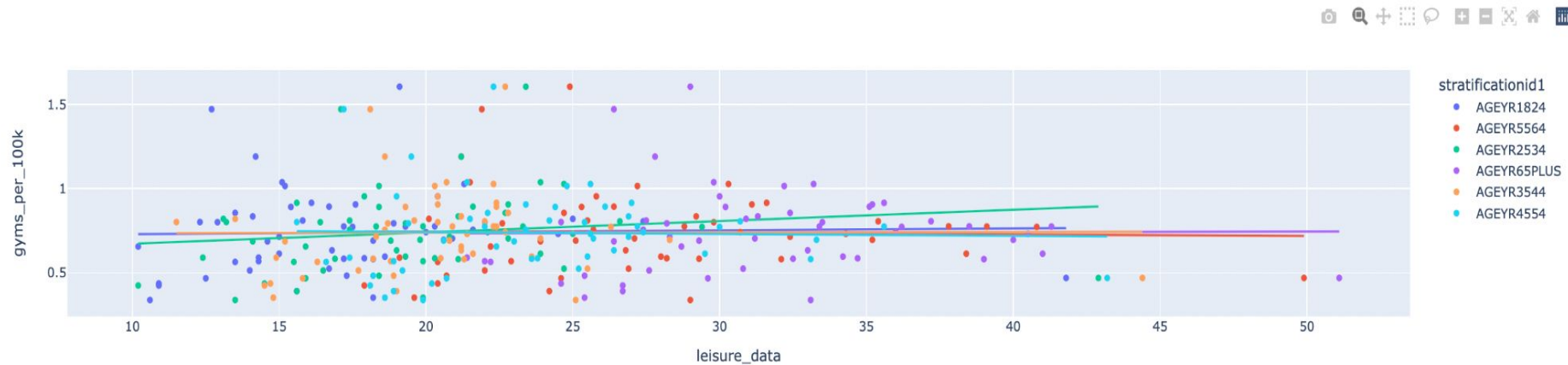
Thank you! Let us know if there are any questions!