

q1

February 18, 2021

```
[1]: from pandas import read_excel
```

Part A: Filtering the data using 'query()'

```
[2]: city_data = read_excel("MA-Jan-14-2021-city-data.xlsx")
city_data
```

```
[2]:
```

	City/Town	Total Case Counts	Two Week Case Counts	Average Daily Rate	\
0	Abington	1073	224	87.795166	
1	Acton	596	90	27.093016	
2	Acushnet	808	241	164.741333	
3	Adams	171	46	39.93607	
4	Agawam	1699	288	72.190166	
..	...	...	...	...	
348	Worthington	20	<5	6.823981	
349	Wrentham	671	106	67.174173	
350	Yarmouth	772	140	41.559829	
351	Unknown	1290	256	*	
352	State	427752	76054	78	

	% Change in Last Week	Total Tests	Total Tests Last Two Weeks	\
0	Higher	18095	2064	
1	Higher	25882	2510	
2	Higher	12928	1893	
3	Higher	8095	991	
4	Higher	38404	3965	
..	...	...	...	
348	No Change	1158	98	
349	Higher	18750	1936	
350	Lower	23531	2862	
351	*	209782	16941	
352	Higher	11944985	1069689	

	Total Positive Tests	Percent Positivity	Change Since Last Week
0	240	0.116279	Lower
1	100	0.039841	No Change
2	272	0.143687	Higher
3	48	0.048436	Higher

4	319	0.080454	Higher
..	...	...	...
348	1	0.010204	Lower
349	114	0.058884	Higher
350	162	0.056604	Lower
351	259	*	*
352	85470	0.0799	Higher

[353 rows x 10 columns]

```
[3]: # city_data.columns = [column.replace(" ", "_") for column in city_data.columns]
```

```
[4]: filtered = city_data.query('`Total Case Counts` != "<5" and `Percent_
    ↳Positivity` != "*"')
filtered = filtered[['Total Case Counts', 'Percent Positivity']]
```

```
[5]: filtered
```

```
[5]:      Total Case Counts Percent Positivity
0           1073      0.116279
1           596      0.039841
2           808      0.143687
3           171      0.048436
4          1699      0.080454
..          ...          ...
347         16924      0.084394
348           20      0.010204
349          671      0.058884
350          772      0.056604
352        427752      0.0799
```

[346 rows x 2 columns]

There are 346 data instances in the new DataFrame.

**Part B: Determine how many cities or towns simultaneously had both more than 1000 cases and a positivity percentage less than .1**

```
[6]: filtered_partb = filtered.query('`Total Case Counts` > 1000 and `Percent_
    ↳Positivity` < .1')
```

```
[7]: filtered_partb
```

```
[7]:      Total Case Counts Percent Positivity
4           1699      0.080454
8           1748      0.071254
10          1194      0.034736
14          1025      0.086225
```

16	2732	0.078713
17	1082	0.09375
20	2128	0.088017
30	2342	0.074249
31	2574	0.099944
35	46455	0.059139
40	2664	0.083113
46	1438	0.027907
48	1261	0.074861
49	3346	0.018063
50	1117	0.074295
56	1866	0.082755
71	2171	0.075089
73	1407	0.070344
85	1035	0.075984
100	5640	0.075118
101	1038	0.071893
106	1182	0.082225
130	1133	0.053716
136	3514	0.081138
166	1009	0.082383
175	3632	0.067546
177	1330	0.045224
188	1522	0.056898
197	1256	0.046924
198	1045	0.03953
206	2692	0.0296
209	2006	0.070026
210	1349	0.089709
219	1846	0.072667
235	1777	0.05813
238	2762	0.092811
242	4757	0.074628
245	1316	0.081476
257	3225	0.08236
270	1969	0.091367
273	3727	0.042958
283	1629	0.075609
294	2299	0.0788
304	1513	0.085372
306	1171	0.083556
307	3716	0.072045
313	1563	0.049099
327	1007	0.040918
328	1949	0.091559
345	1626	0.062018
346	2941	0.088546

347	16924	0.084394
352	427752	0.0799

**Part C: Use pandas to print the mean and standard deviation of the positivity percentage**

```
[8]: mean = filtered['Percent Positivity'].mean()
      mean
```

```
[8]: 0.0766657688113855
```

```
[9]: stdev = filtered['Percent Positivity'].std()
      stdev
```

```
[9]: 0.038291167032190127
```

```
[10]: print("Mean and Standard Deviation (Percent Positivity_: {} +- {}".format(mean,
      ↪stdev))
```

```
Mean and Standard Deviation (Percent Positivity_: 0.0766657688113855 +-
0.038291167032190127
```

## q2 and q3

February 18, 2021

```
[1]: from pandas import read_csv
import matplotlib.pyplot as plt
from scipy.stats import median_test
```

**Q2 Part A: Filter data to only include Republicans and state/poll rating**

```
[2]: poll_data = read_csv('governor_polls.csv')
poll_data
```

```
[2]:
```

	question_id	poll_id	cycle	state	pollster_id	\
0	136573	72762	2020	Montana	1365	
1	136573	72762	2020	Montana	1365	
2	136573	72762	2020	Montana	1365	
3	136574	72770	2020	North Carolina	1522	
4	136574	72770	2020	North Carolina	1522	
...	...	...	...	...	...	
2246	83152	52626	2018	Nevada	1056	
2247	83155	52629	2018	New Mexico	468	
2248	83155	52629	2018	New Mexico	468	
2249	79648	36742	2018	New York	383	
2250	79648	36742	2018	New York	383	

		pollster	sponsor_ids	sponsors	display_name	\
0		Change Research	NaN	NaN	Change Research	
1		Change Research	NaN	NaN	Change Research	
2		Change Research	NaN	NaN	Change Research	
3		Swayable	NaN	NaN	Swayable	
4		Swayable	NaN	NaN	Swayable	
...		...	...	...	...	
2246	Remington Research Group		NaN	NaN	Remington Research Group	
2247	Tarrance Group		NaN	NaN	Tarrance Group	
2248	Tarrance Group		NaN	NaN	Tarrance Group	
2249	Public Policy Polling		NaN	NaN	Public Policy Polling	
2250	Public Policy Polling		NaN	NaN	Public Policy Polling	

	pollster_rating_id	...	created_at	notes	\
0	48.0	...	11/2/2020 22:15	NaN	
1	48.0	...	11/2/2020 22:15	NaN	

2	48.0	...	11/2/2020	22:15	NaN
3	543.0	...	11/2/2020	22:15	NaN
4	543.0	...	11/2/2020	22:15	NaN
...	...	...	...	...	...
2246	279.0	...	6/22/2018	14:58	NaN
2247	333.0	...	6/22/2018	15:00	NaN
2248	333.0	...	6/22/2018	15:00	NaN
2249	263.0	...	11/17/2017	11:23	NaN
2250	263.0	...	11/17/2017	11:23	NaN

	url	stage	race_id	\
0	https://docs.google.com/spreadsheets/d/1MPKy3A...	general	7821	
1	https://docs.google.com/spreadsheets/d/1MPKy3A...	general	7821	
2	https://docs.google.com/spreadsheets/d/1MPKy3A...	general	7821	
3	https://www.swayable.com/polls/2020-11-02-larg...	general	7824	
4	https://www.swayable.com/polls/2020-11-02-larg...	general	7824	
...	...	...	...	...
2246	https://www.reviewjournal.com/news/politics-an...	general	151	
2247	https://www.scribd.com/document/350187692/May-...	general	153	
2248	https://www.scribd.com/document/350187692/May-...	general	153	
2249	http://www.publicpolicypolling.com/pdf/2015/PP...	general	154	
2250	http://www.publicpolicypolling.com/pdf/2015/PP...	general	154	

	answer	candidate_id	candidate_name	candidate_party	\
0	Cooney	14688	Mike Cooney	DEM	
1	Gianforte	14689	Greg Gianforte	REP	
2	Bishop	16066	Lyman Bishop	LIB	
3	Cooper	13404	Roy A. Cooper	DEM	
4	Forest	13406	Dan Forest	REP	
...	...	...	...	...	...
2246	Laxalt	11120	Adam Laxalt	REP	
2247	Lujan Grisham	11122	Michelle Lujan Grisham	DEM	
2248	Pearce	11123	Stevan Pearce	REP	
2249	Cuomo	10430	Andrew Cuomo	DEM	
2250	Gibson	10431	Christopher P. Gibson	REP	

	pct
0	44.0
1	48.0
2	3.0
3	54.7
4	42.2
...	...
2246	46.0
2247	47.0
2248	43.0
2249	49.0

2250 26.0

[2251 rows x 38 columns]

```
[3]: republican_governors = poll_data.query('candidate_party == "REP"')
republican_governors = republican_governors[['state', 'pct']]
republican_governors
```

```
[3]:
```

	state	pct
1	Montana	48.0
4	North Carolina	42.2
8	Washington	40.9
10	North Carolina	45.0
14	North Carolina	44.8
...	...	...
2242	Ohio	55.0
2244	Arizona	42.0
2246	Nevada	46.0
2248	New Mexico	43.0
2250	New York	26.0

[868 rows x 2 columns]

**Q2 Part B: Group poll ratings by state and print median poll ratings by state in descending order**

```
[4]: republican_governors = republican_governors.groupby('state').median()
republican_governors = republican_governors.sort_values(by = 'pct',
↪ascending=False)
```

```
[5]: republican_governors
```

```
[5]:
```

state	pct
Massachusetts	64.000
Wyoming	61.000
Arkansas	60.000
North Dakota	59.000
New Hampshire	55.200
Alabama	54.500
Vermont	53.500
Utah	53.500
Tennessee	53.000
Texas	52.895
Indiana	52.500
South Carolina	52.000
Arizona	52.000
Maryland	51.800

West Virginia	50.500
Missouri	50.100
Montana	47.000
Oklahoma	46.400
Georgia	46.350
South Dakota	46.000
Nevada	44.150
Wisconsin	44.000
Louisiana	44.000
Florida	44.000
New Mexico	43.000
Ohio	43.000
Alaska	43.000
Mississippi	42.645
Iowa	41.500
Oregon	40.800
Idaho	40.500
Kentucky	40.500
Colorado	40.000
North Carolina	40.000
Minnesota	39.000
Kansas	39.000
Maine	38.500
California	38.000
Michigan	38.000
Connecticut	37.500
Pennsylvania	35.500
Rhode Island	34.800
Washington	32.000
New York	31.000
Hawaii	31.000
Illinois	29.800
Delaware	26.000

**Q3 Part A: draw side-by-side notched box plots of the New Hampshire and Maryland governors**

```
[6]: poll_data
```

```
[6]:
```

	question_id	poll_id	cycle	state	pollster_id \
0	136573	72762	2020	Montana	1365
1	136573	72762	2020	Montana	1365
2	136573	72762	2020	Montana	1365
3	136574	72770	2020	North Carolina	1522
4	136574	72770	2020	North Carolina	1522
...	...	...	...	...	...
2246	83152	52626	2018	Nevada	1056
2247	83155	52629	2018	New Mexico	468



2248	83155	52629	2018	New Mexico	468
2249	79648	36742	2018	New York	383
2250	79648	36742	2018	New York	383

		pollster	sponsor_ids	sponsors	display_name \
0		Change Research	NaN	NaN	Change Research
1		Change Research	NaN	NaN	Change Research
2		Change Research	NaN	NaN	Change Research
3		Swayable	NaN	NaN	Swayable
4		Swayable	NaN	NaN	Swayable
...		...	...	...	...
2246	Remington Research Group		NaN	NaN	Remington Research Group
2247	Tarrance Group		NaN	NaN	Tarrance Group
2248	Tarrance Group		NaN	NaN	Tarrance Group
2249	Public Policy Polling		NaN	NaN	Public Policy Polling
2250	Public Policy Polling		NaN	NaN	Public Policy Polling

	pollster_rating_id	...	created_at	notes \
0	48.0	...	11/2/2020 22:15	NaN
1	48.0	...	11/2/2020 22:15	NaN
2	48.0	...	11/2/2020 22:15	NaN
3	543.0	...	11/2/2020 22:15	NaN
4	543.0	...	11/2/2020 22:15	NaN
...	...	...	...	...
2246	279.0	...	6/22/2018 14:58	NaN
2247	333.0	...	6/22/2018 15:00	NaN
2248	333.0	...	6/22/2018 15:00	NaN
2249	263.0	...	11/17/2017 11:23	NaN
2250	263.0	...	11/17/2017 11:23	NaN

	url	stage	race_id \
0	<a href="https://docs.google.com/spreadsheets/d/1MPKy3A...">https://docs.google.com/spreadsheets/d/1MPKy3A...</a>	general	7821
1	<a href="https://docs.google.com/spreadsheets/d/1MPKy3A...">https://docs.google.com/spreadsheets/d/1MPKy3A...</a>	general	7821
2	<a href="https://docs.google.com/spreadsheets/d/1MPKy3A...">https://docs.google.com/spreadsheets/d/1MPKy3A...</a>	general	7821
3	<a href="https://www.swayable.com/polls/2020-11-02-larg...">https://www.swayable.com/polls/2020-11-02-larg...</a>	general	7824
4	<a href="https://www.swayable.com/polls/2020-11-02-larg...">https://www.swayable.com/polls/2020-11-02-larg...</a>	general	7824
...	...	...	...
2246	<a href="https://www.reviewjournal.com/news/politics-an...">https://www.reviewjournal.com/news/politics-an...</a>	general	151
2247	<a href="https://www.scribd.com/document/350187692/May-...">https://www.scribd.com/document/350187692/May-...</a>	general	153
2248	<a href="https://www.scribd.com/document/350187692/May-...">https://www.scribd.com/document/350187692/May-...</a>	general	153
2249	<a href="http://www.publicpolicypolling.com/pdf/2015/PP...">http://www.publicpolicypolling.com/pdf/2015/PP...</a>	general	154
2250	<a href="http://www.publicpolicypolling.com/pdf/2015/PP...">http://www.publicpolicypolling.com/pdf/2015/PP...</a>	general	154

	answer	candidate_id	candidate_name	candidate_party \
0	Cooney	14688	Mike Cooney	DEM
1	Gianforte	14689	Greg Gianforte	REP
2	Bishop	16066	Lyman Bishop	LIB

3	Cooper	13404	Roy A. Cooper	DEM
4	Forest	13406	Dan Forest	REP
...	...	...	...	...
2246	Laxalt	11120	Adam Laxalt	REP
2247	Lujan Grisham	11122	Michelle Lujan Grisham	DEM
2248	Pearce	11123	Stevan Pearce	REP
2249	Cuomo	10430	Andrew Cuomo	DEM
2250	Gibson	10431	Christopher P. Gibson	REP

	pct
0	44.0
1	48.0
2	3.0
3	54.7
4	42.2

...	...
2246	46.0
2247	47.0
2248	43.0
2249	49.0
2250	26.0

[2251 rows x 38 columns]

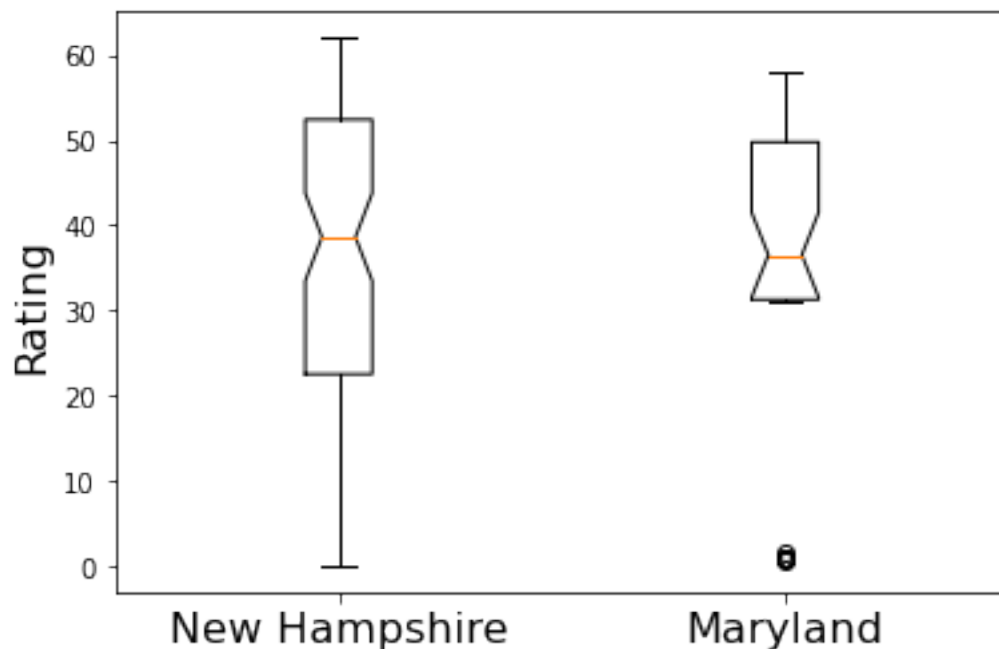
```
[7]: governors_nh = poll_data.query('state == "New Hampshire"')
governors_md = poll_data.query('state == "Maryland"')
```

```
[8]: nh_ratings = governors_nh['pct'].tolist()
md_ratings = governors_md['pct'].tolist()
```

```
[9]: box = plt.boxplot([nh_ratings, md_ratings], notch=True)
plt.suptitle('Poll Ratings of Two Governors', fontsize=22)
plt.xticks([1, 2], ['New Hampshire', 'Maryland'], fontsize=16)
plt.ylabel('Rating', fontsize=16)
```

```
[9]: Text(0, 0.5, 'Rating')
```

## Poll Ratings of Two Governors



### Q3 Part B: Which of the two governors have a higher median rating?

Based on the boxplot, the New Hampshire governor seems to have a slightly higher median rating. The governors' median ratings do not appear to differ significantly. The New Hampshire governor's median rating appears to be around 38, while the Maryland governor's rating appears to be about 35. The notching overlaps signify that the medians are similar and the

### Q3 Part C: Apply Mood's Median Test

```
[10]: # scipy.stats.mood returns 1. the z-score for the hypothesis test and 2. the p-value for the hypothesis test
stat, p, med, tbl = median_test(nh_ratings, md_ratings)
```

```
[11]: p
```

```
[11]: 0.8851438301402731
```

The null hypothesis of Mood's Median Test is that the samples come from the same distribution. The alternative hypothesis is that they come from different distributions. Because the p value returned from Mood's Median Test is far above the threshold of .05, we have strong evidence for the null hypothesis, which is that the data comes from the same distribution.

We conclude that we are fairly confident that the governor's median ratings do not vary significantly.