

This is an analysis of complaints data, munged [here \(https://github.com/TheOregonian/long-term-care-db/blob/master/notebooks/transformation/mung-3-29-scrape.ipynb\)](https://github.com/TheOregonian/long-term-care-db/blob/master/notebooks/transformation/mung-3-29-scrape.ipynb).

The fields are:

1. *abuse\_number*: A unique number assigned each complaint.
2. *facility\_id*: A unique number to each facility building. Stays if ownership changes.
3. *facility\_type*: NF: Nursing Facility; ALF: Assisted Living Facility; RCF: Residential Care Facility.
4. *facility\_name*: Name of facility as of January 2017, when DHS provided the facility data to The Oregonian.
5. *abuse\_type*: A – facility abuse; L – licensing. Note: This does not apply to nursing facilities. All their complaints are either blank in this field or licensing.
6. *fine*: Amount that state initially fined the facility. Not necessarily amount of final fine.
7. *action\_notes*: DHS determination of what general acts constituted the abuse or rule violation.
8. *incident\_date*: Date the incident occurred
9. *outcome*: A very brief description of the consequences of the abuse or rule violation to the resident
10. *outcome\_notes*: A detailed description of what happened.
11. *year*: year incident occurred
12. *online\_fac\_name*: If complaint is online, name listed for the facility
13. *public*: Whether or not complaint is online
14. *omg\_outcome*: Field we created to group some similar outcomes.

```
In [17]: import pandas as pd
import numpy as np
from IPython.core.display import display, HTML
display(HTML("<style>.container { width:100% !important; }</style>"))
pd.set_option('display.max_colwidth', -1)
```

```
In [18]: df = pd.read_csv('../data/processed/complaints-3-29-scrape.csv')
```

## How many total complaints are there?

```
In [19]: df.count()[0]
```

```
Out[19]: 13032
```

## How many complaints do not appear in the state's public database?

```
In [20]: df[df['public']=='offline'].count()[0]
```

```
Out[20]: 7846
```

## How many complaints do appear in the state's public database?

```
In [21]: df[df['public']=='online'].count()[0]
```

```
Out[21]: 5186
```

## What percent of complaints are missing?

```
In [22]: df[df['public']=='offline'].count()[0]/df.count()[0]*100
```

```
Out[22]: 60.205647636586868
```

## How many complaints were labelled 'Exposed to potential harm' or 'No negative outcome'?

```
In [23]: df[(df['outcome']=='Exposed to Potential Harm') | (df['outcome']=='No Negative Outcome')].count()[0]
```

```
Out[23]: 2509
```

## Of all missing complaints, what percent are in the above two categories?

```
In [24]: df[(df['outcome']=='Exposed to Potential Harm') | (df['outcome']=='No Negative Outcome')].count()[0]/df[df['public']=='offline'].count()[0]*100
```

```
Out[24]: 31.978078001529443
```

## What's the online/offline breakdown by outcome?

*This was used in graphics*

```
In [25]: totals = df.groupby(['omg_outcome','public']).count()['abuse_number'].unstack().reset_index()
```

```
In [26]: totals.fillna(0, inplace = True)
```

```
In [27]: totals['total'] = totals['online']+totals['offline']
```

```
In [28]: totals['pct_offline'] = round(totals['offline']/totals['total']*100)
```

```
In [29]: totals.sort_values('pct_offline', ascending=False)
```

```
Out[29]:
```

	public	omg_outcome	offline	online	total	pct_offline
16		Staffing issues	12.0	0.0	12.0	100.0
1		Denied readmission or moved improperly	35.0	2.0	37.0	95.0
14		Potential harm	2361.0	148.0	2509.0	94.0
3		Fall, no injury	150.0	13.0	163.0	92.0
8		Left facility without attendant, no injury	207.0	18.0	225.0	92.0
9		Loss of Dignity	884.0	97.0	981.0	90.0
12		Medication error	983.0	217.0	1200.0	82.0
5		Inadequate care	496.0	170.0	666.0	74.0
6		Inadequate hygiene	138.0	104.0	242.0	57.0
10		Loss of property, theft or financial exploitation	809.0	737.0	1546.0	52.0
13		Physical abuse	89.0	92.0	181.0	49.0
18		Verbal or emotional abuse	70.0	94.0	164.0	43.0
7		Involuntary seclusion	8.0	11.0	19.0	42.0
2		Failure to address resident aggression	395.0	622.0	1017.0	39.0
4		Fracture or other injury	680.0	1185.0	1865.0	36.0
11		Medical condition developed or worsened	370.0	1046.0	1416.0	26.0
0		Death	7.0	23.0	30.0	23.0
15		Sexual abuse	15.0	49.0	64.0	23.0
17		Unreasonable discomfort or continued pain	115.0	452.0	567.0	20.0
19		Weight loss	20.0	106.0	126.0	16.0

**How many offline complaints in the database were found to have "abuse," "neglect" or "exploitation?"**

```
In [30]: df['outcome_notes'].fillna('', inplace = True)
```

```
In [31]: df[(df['outcome_notes'].str.contains('constitute neglect|constitutes neglect|constitute abuse|constitutes abuse|constitutes exploitation|constitutes financial exploitation')) & (df['public']=='offline')].count()[0]
```

```
Out[31]: 483
```

"The state fined the facilities in hundreds of those cases."

**In how many 'potential harm' cases were facilities fined?**

```
In [32]: df[(df['omg_outcome']=='Potential harm') & (df['fine']>0) & (df['public']=='of  
fine')].count()[0]
```

Out[32]: 206