

This is a dataset of Assisted Living, Nursing and Residential Care facilities in Oregon, open as of September, 2016. For each, we have:

Data were 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>).

1. *facility\_id*: Unique ID used to join to complaints
2. *fac\_ccmunumber*: Unique ID used to join to ownership history
3. *facility\_type*: NF - Nursing Facility; RCF - Residential Care Facility; ALF - Assisted Living Facility
4. *fac\_capacity*: Number of beds facility is licensed to have. Not necessarily the number of beds facility does have.
5. *facility\_name*: Facility name at time of September extract.
6. *offline*: created in munging notebook, a count of complaints that DO NOT appear when facility is searched on state's [complaint search website](https://apps.state.or.us/cf2/spd/facility_complaints/) ([https://apps.state.or.us/cf2/spd/facility\\_complaints/](https://apps.state.or.us/cf2/spd/facility_complaints/)).
7. *online*: created in munging notebook, a count of complaints that DO appear when facility is searched on state's [complaint search website](https://apps.state.or.us/cf2/spd/facility_complaints/) ([https://apps.state.or.us/cf2/spd/facility\\_complaints/](https://apps.state.or.us/cf2/spd/facility_complaints/)).

```
In [91]: import pandas as pd
import numpy as np
from IPython.core.display import display, HTML
display(HTML("<style>.container { width:100% !important; }</style>"))
```

```
In [92]: df = pd.read_csv('../data/processed/facilities-3-29-scrape.csv')
```

## How many facilities are there?

```
In [93]: df.count()[0]
```

Out[93]: 642

## How many facilities have accurate records online?

Those that have no offline records.

```
In [94]: df[(df['offline'].isnull())].count()[0]
```

Out[94]: 59

## How many facilities have inaccurate records online?

Those that have offline records.

```
In [95]: df[(df['offline'].notnull())].count()[0]
```

```
Out[95]: 583
```

**How many facilities had more than double the number of complaints shown online?**

```
In [96]: df[(df['offline']>df['online']) & (df['online'].notnull())].count()[0]
```

```
Out[96]: 358
```

**How many facilities show zero complaints online but have complaints offline?**

```
In [97]: df[(df['online'].isnull()) & (df['offline'].notnull())].count()[0]
```

```
Out[97]: 59
```

**How many facilities have complaints and are accurate online?**

```
In [98]: df[(df['online'].notnull()) & (df['offline'].isnull())].count()[0]
```

```
Out[98]: 16
```

**How many facilities have complaints?**

```
In [99]: df[(df['online'].notnull()) | df['offline'].notnull()].count()[0]
```

```
Out[99]: 599
```

**What percent of facilities have accurate records online?**

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

```
Out[100]: 9.1900311526479754
```

**What is the total capacity of all facilities with inaccurate records?**

```
In [101]: df[df['offline'].notnull()].sum()['fac_capacity']
```

```
Out[101]: 35129.0
```

## How many facilities appear to have no complaints, whether or not they do?

```
In [102]: df[df['online'].isnull()].count()[0]
```

```
Out[102]: 102
```

## What are the ten facilities with >50 complaints that have the highest disparities?

For graphics

```
In [114]: over_50 = df[((df['offline']+df['online'])>50)]
```

```
In [115]: over_50['total'] = over_50['online']+over_50['offline']
```

```
/Users/fzarkhin/anaconda/lib/python3.5/site-packages/ipykernel/__main__.py:1:
SettingWithCopyWarning:
```

A value is trying to be set on a copy of a slice from a DataFrame.  
Try using `.loc[row_indexer,col_indexer] = value` instead

See the caveats in the documentation: <http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy>

```
if __name__ == '__main__':
```

```
In [116]: over_50['pct_offline'] = over_50['offline']/over_50['total']*100
```

```
/Users/fzarkhin/anaconda/lib/python3.5/site-packages/ipykernel/__main__.py:1:
SettingWithCopyWarning:
```

A value is trying to be set on a copy of a slice from a DataFrame.  
Try using `.loc[row_indexer,col_indexer] = value` instead

See the caveats in the documentation: <http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy>

```
if __name__ == '__main__':
```

```
In [117]: over_50[over_50['facility_name']=='Avamere Health Services of Rogue Valley']
```

```
Out[117]:
```

	facility_id	fac_ccmunumber	facility_type	fac_capacity	facility_name	offline	online	total	p
4	385024	385024	NF	91.0	Avamere Health Services of Rogue Valley	67.0	27.0	94.0	7

```
In [118]: over_50.sort_values('pct_offline',ascending = False).head(10)
```

```
Out[118]:
```

	facility_id	fac_ccmunumber	facility_type	fac_capacity	facility_name	offline	online	total
<b>50</b>	385166	385166	NF	165.0	Maryville Nursing Home	53.0	12.0	65.0
<b>78</b>	385219	385219	NF	93.0	Care Center East Health & Specialty Care Center	63.0	16.0	79.0
<b>45</b>	385157	385157	NF	114.0	Life Care Center Of Coos Bay	74.0	21.0	95.0
<b>63</b>	385190	385190	NF	78.0	Prestige Post- Acute and Rehabilitation Center-...	50.0	15.0	65.0
<b>34</b>	385143	385143	NF	118.0	Umpqua Valley Nursing & Rehabilitation Center	55.0	17.0	72.0
<b>144</b>	50A263	50A263	RCF	59.0	Brookdale Bend	40.0	13.0	53.0
<b>23</b>	385120	385120	NF	121.0	Valley West Health Care Center	55.0	20.0	75.0
<b>113</b>	385270	385270	NF	96.0	Prestige Post- Acute and Rehabilitation Center ...	50.0	19.0	69.0
<b>4</b>	385024	385024	NF	91.0	Avamere Health Services of Rogue Valley	67.0	27.0	94.0
<b>27</b>	385132	385132	NF	148.0	Avamere Rehabilitation of King City	36.0	15.0	51.0

```
In [ ]:
```