

A look at contractors

The ContractorCompanyName needed cleaning in some instances, so I took a version of the `data-processed/demolitions_full.csv` into Open Refine, created a new column `ContractorNameClean` and used facets to clean it. I exported it as `openrefine/contrator_clean.csv` and saved the steps as `openrefine/contractor_clean.json`.

```
In [1]: import pandas as pd
import altair as alt
```

```
In [2]: # Column type fixes
column_types = {
    'ApplicantPhone': pd.np.str,
    'ContractorPhone': pd.np.str,
    'CalendarYearIssued': pd.np.str,
    'OriginalZip': pd.np.str,
}

# import raw data
data_raw = pd.read_csv(
    '../openrefine/contractor_clean.csv',
    index_col=None,
    dtype=column_types,
    parse_dates=['IssuedDate']
)

demolitions = data_raw

demolitions.shape
```

```
Out[2]: (3165, 70)
```

Most permits by contractor

```
In [3]: demolitions.ContractorNameClean.value_counts().head(10)
```

```
Out[3]: DAR Construction                    569
Southwest Destructors****MAIN***          235
Building Abatement Demolition Co Inc      104
AAR Inc.                                   84
Absolute Demolition                        73
MX3 Investments                           72
Heart of Texas Demolition, LLC             40
Pecan Valley Homes LLC****MAIN***         35
Gossett and Jones Homes Inc.              31
Paradisa Homes, LLC                       31
Name: ContractorNameClean, dtype: int64
```

DAR Contruaction over time

```
In [4]: # filter to just dar permits
dar_permits = demolitions[demolitions['ContractorNameClean'] == 'DAR Construct
ion']
dar_permits.shape
```

Out[4]: (569, 70)

```
In [5]: dar_by_year = dar_permits.groupby(['CalendarYearIssued']).agg(dict(
    PermitNum='count'
)).reset_index()
dar_by_year.columns = ['Year', 'Count']
dar_by_year
```

Out[5]:

	Year	Count
0	2008	5
1	2009	5
2	2010	9
3	2011	15
4	2012	39
5	2013	66
6	2014	86
7	2015	97
8	2016	93
9	2017	108
10	2018	46

Explore contractor by year

DAR Construction has led demolitions each year since 2011, sometimes by twice the next closest competitor.

```
In [6]: # group by year, contractor to count them.
contractor_group = demolitions.groupby(['CalendarYearIssued', 'ContractorNameC
lean']).agg(dict(
    PermitNum='count'
)).reset_index()
# rename the columns
contractor_group.columns = ['Year', 'Contractor', 'Count']
# sort the columns so we can look at each year
contractor_sorted = contractor_group.sort_values(by=['Year', 'Count'], ascendi
ng=[1, 0])
```

```
In [7]: # Look at top contractors each year
# change the year value to see other years
# DAR has lead by far since 2011
contractor_sorted[contractor_sorted.Year == '2018'].head(10)
```

Out[7]:

	Year	Contractor	Count
1089	2018	DAR Construction	46
1062	2018	AAR Inc.	22
1153	2018	Southwest Destructors****MAIN***	16
1132	2018	Paradisa Homes, LLC	14
1066	2018	Absolute Demolition	7
1111	2018	Journeyman Construction	6
1121	2018	MX3 Investments	5
1070	2018	Austin Demolition DJ- Homes	4
1109	2018	Joseph Builders LLC	4
1134	2018	Pecan Valley Homes LLC****MAIN***	4

In []:

In []: