

DS6003  
spm9r  
Assignment1

Motivation:

The motivation for this analysis was to examine the effects of various variables on the assessed value of properties in Charlottesville.

```
# View counts of categorical items
```

```
cols = ['UseCode'
        , 'Style'
        , 'Grade'
        , 'Roof'
        , 'Flooring'
        , 'Heating'
        , 'BasementType'
        , 'ExternalWalls']

for c in cols:
    temp = sqlc.sql("""SELECT {0}, count({0})
    FROM assessments ass \
    inner join detail det on ass.ParcelNumber = det.ParcelNumber \
    inner join resdetail rd on ass.ParcelNumber = rd.ParcelNumber
    group by {0} \
    order by count({0}) desc""".format(c))
    temp.show()
```

This code is used to display the frequencies of classes in categorical variables in the real estate data set. This is for data exploration that did not end up used in the final analysis.

This plot shows the weak linear relationship between square footage and assessed value.

