Summary Statistics

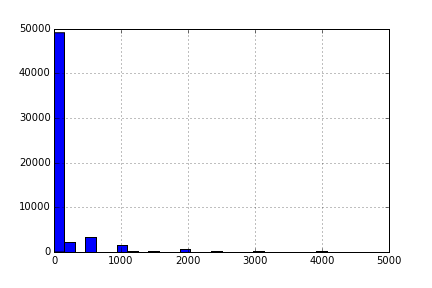
Barb & Andreas

# Five Number Summary for Numeric Columns

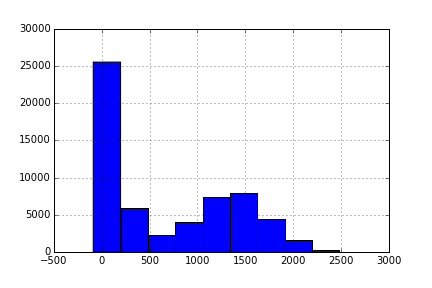
|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **amount\_tsh** | **gps\_height** | **longitude** | **latitude** | **num\_private** | **region\_code** | **district\_code** | **population** | **construction\_year** |
| **count** | 59400.00 | 59400.00 | 59400.00 | 59400.00 | 59400.00 | 59400.00 | 59400.00 | 59400.00 | 59400.00 |
| **mean** | 317.65 | 668.30 | 34.08 | -5.71 | 0.47 | 15.30 | 5.63 | 179.91 | 1300.65 |
| **std** | 2997.57 | 693.12 | 6.57 | 2.95 | 12.24 | 17.59 | 9.63 | 471.48 | 951.62 |
| **min** | 0.00 | -90.00 | 0.00 | -11.65 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 |
| **25%** | 0.00 | 0.00 | 33.09 | -8.54 | 0.00 | 5.00 | 2.00 | 0.00 | 0.00 |
| **50%** | 0.00 | 369.00 | 34.91 | -5.02 | 0.00 | 12.00 | 3.00 | 25.00 | 1986.00 |
| **75%** | 20.00 | 1319.25 | 37.18 | -3.33 | 0.00 | 17.00 | 5.00 | 215.00 | 2004.00 |
| **max** | 350000.00 | 2770.00 | 40.35 | 0.00 | 1776.00 | 99.00 | 80.00 | 30500.00 | 2013.00 |

# Histograms

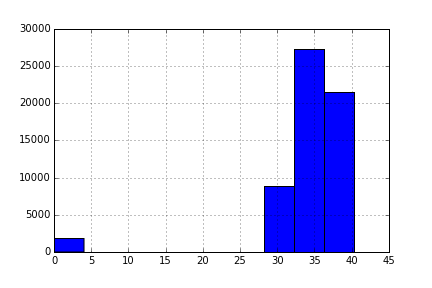
**Amount TSH**



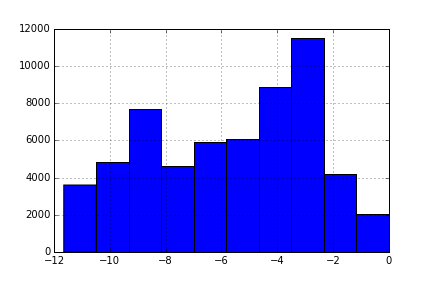
**GPS Height**



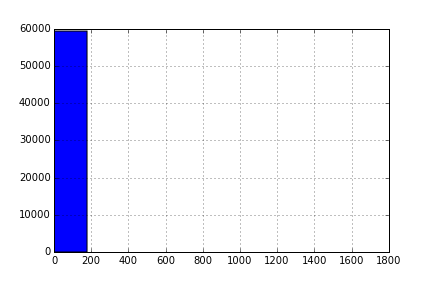
**Longitude**



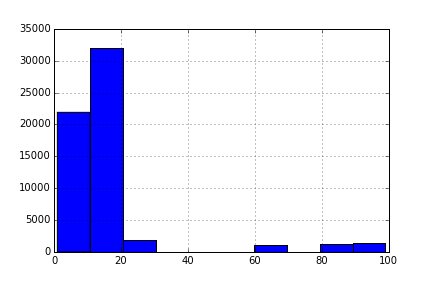
**Latitude**



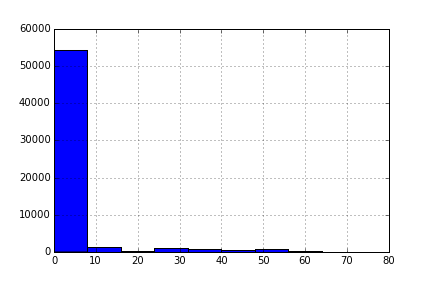
**Private num**



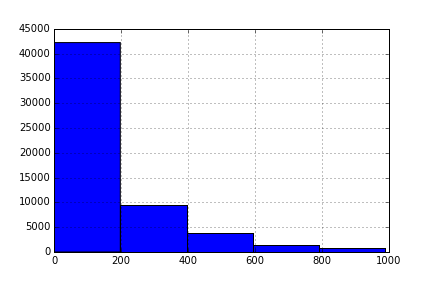
**Region Code**



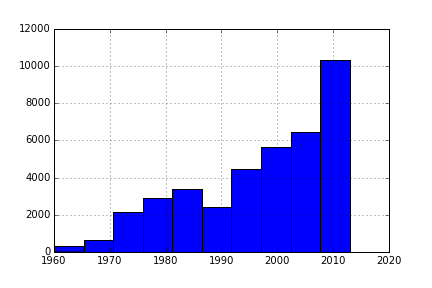
**District Code**



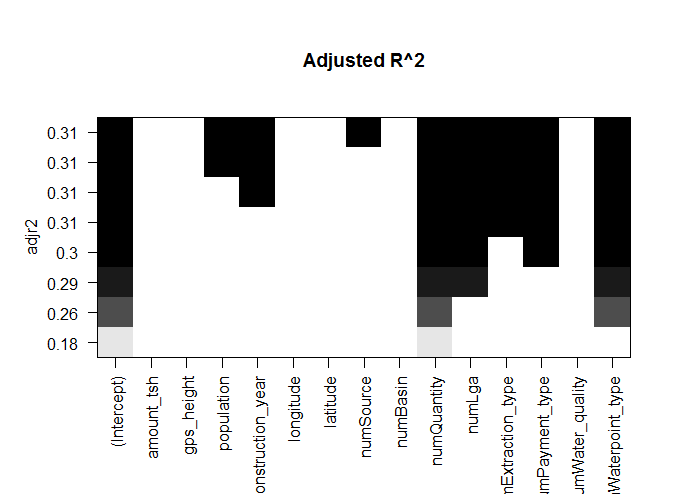
**Population (below 1000)**



**Construction Year (0’s excluded)**



# Regression Best Subsets



# Questions for Ted

1. Aggregate function so we can do a stacked bar plot of numStatus by construction\_Year <http://pandas.pydata.org/pandas-docs/version/0.15.0/visualization.html>
2. How to iterate over the columns so we can automate the process.