**Classifying Big Mines**

***Selected sample:*** Mines in the 90th percentile of average, maximum, *and* median hours worked (N = 132)

***Robustness Checks:***

* Mines in the 90th percentile of average hours worked (N = 159)
* Mines in the 90th percentile of maximum hours worked (N = 159)
* Mines in the 90th percentile of median hours worked (N = 159)
* Mines in the 90th percentile of average, maximum, *or* median hours worked (N = 186)

***Methodology:***

To identify big mines, we use production hours as a proxy for mine size.

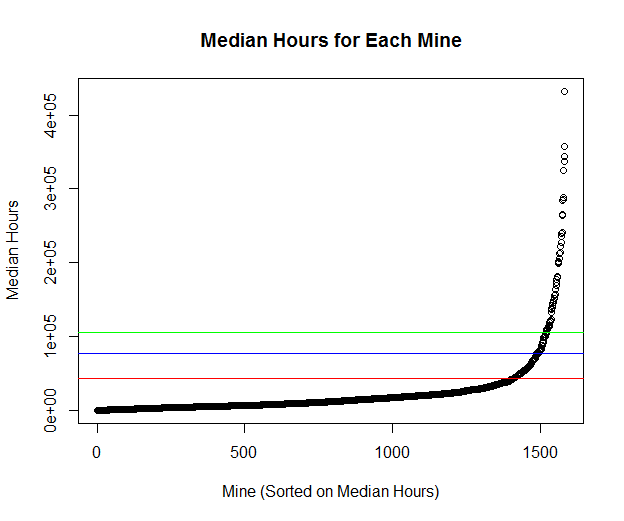
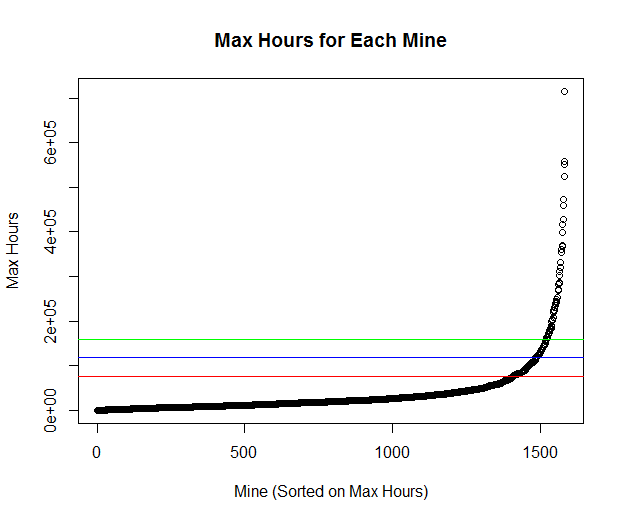
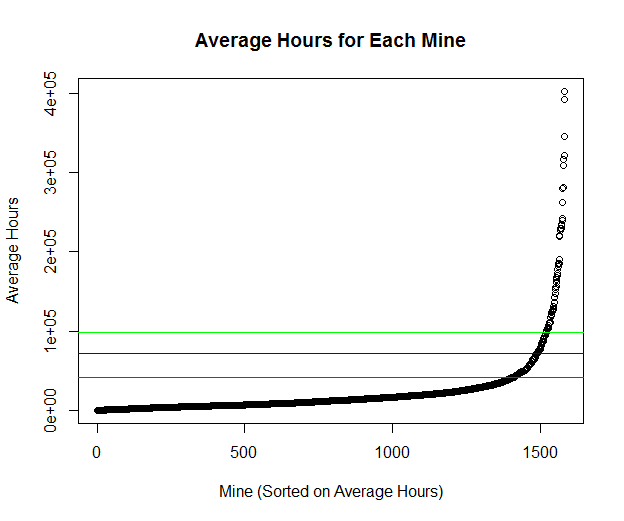
For every mine in the dataset, across all quarters in which the mine had non-zero production, we calculate:

* Average number of hours worked
* Maximum number of hours worked
* Median number of hours worked

For each of these measures, we select the mines in the 90th, 95th, and 97th percentile. These percentile cutoffs result in the following number of mines (for each measure of mine size):

* 90th percentile - 159 mines
* 95th percentile - 80 mines
* 97th percentile - 48 mines

We want an adequate sample size for subsequent analyses, so we are inclined to use the 90th percentile cutoff. There is a concern that this may include too-small mines. We therefore plot each measure of mine size with the 90th (red), 95th (blue), and 97th (green) percentile cutoff drawn:



These plots give us confidence that the mines included by the 90th percentile cutoff are reasonable.

We next assess the differences between the mines in the 90th percentile of size according to each measure of size. The number of mines that are not shared between the following groups of mines are as follows:

* Average & Maximum – 19
* Average & Median – 10
* Maximum & Median – 25
* Average & Maximum & Median – 37

(Note: we assess the amount of non-overlap at the 95th and 97th percentile cutoff, as well. Similarly, we see only a minimal number of mines that are not shared between the different specifications).

To be conservative, we select mines that are in the 90th percentile of size for *all* measures of size (i.e., average, maximum, *and* median). We use individual measures, as well as mines classified as large by *any* measure as robustness checks.