My project involves examining data on US power plants using the North American Electric Reliability Corporation (NERC) Generating Availability Data System (GADS) public database. US power generating plants have to report planned and unplanned outages using this database.

I am using data for three types of power plants.

* Fossil fuel fired utility boilers with steam turbine generation
* Combined Cycle gas and steam turbine generation
* Simple Cycle gas turbine generation

My goal is determine which types of equipment failures are most responsible for unplanned/forced outages for each of these three types of plants. To do this, I will have to do further examine the field Cause\_Code, using the metadata associated with each four character code and chunking it into higher level or less granular categories.

I am starting by looking at some statistics by outage Type.

Since I am focusing on unplanned outages, I used the following to filter each data set (fossil fuel steam shown as an example) to include only unplanned derates: D1, D2, D3, start up failures: SF, and unplanned full outages: U1, U2, U3.

# Filter for only forced, not planned or reserve shutdowns, outages and derates

type\_filter <- c("D1", "D2", "D3", "D4", "SF", "U1", "U2", "U3")

ff\_steam\_df <- ff\_steam\_df %>%

filter(ff\_steam\_df$Type %in% type\_filter)

I calculated the mean by Type per below. It’s not surprising that U1 has the highest mean as a U1 is an outage for which the generating unit must be shut down immediately.

# Calculate mean Time\_To\_Repair for each outage Type

ff\_steam\_df %>% group\_by(Type) %>%

summarise(Mean\_TTR\_Type\_ff = mean(Time\_To\_Repair, na.rm = TRUE)) %>%

arrange(desc(Mean\_TTR\_Type\_ff))

|  |
| --- |
| # A tibble: 8 x 2  Type Mean\_TTR\_Type\_ff  <chr> <dbl>  1 U1 62.2  2 SF 58.5  3 U2 58.1  4 U3 53.9  5 D1 37.4  6 D2 33.2  7 D4 11.1  8 D3 10.7 |
|  |
| |  | | --- | | > | |

Per below, I created an box plot by type. This was a bit hard to read and I may need to adjust the range of the Y axis, Time\_To\_Repair. The graphs did show that each outage Type had a distribution with a long tail to the right.

# Create a Box Plot graph, x = Type, y = Time\_To\_Repair

ggplot(data = ff\_steam\_df) +

geom\_boxplot(mapping = aes(x = Type, y = Time\_To\_Repair))

I also created a stat summary graph by year for each (faceted by) each outage type. The confirmed the right tail distribution as the mean is close to the minimum value while the max values appear as outliers. This graph also showed that there was very little difference in these statistics year to year.

# create a stat summary plot, x = year(start\_st), y = time to repair

ggplot(data = ff\_steam\_df) +

stat\_summary(

mapping = aes(x = year(start\_dt), y = Time\_To\_Repair),

fun.ymin = min,

fun.ymax = max,

fun.y = mean

) +

facet\_wrap(~ Type, nrow = 2)

As mentioned above, in order to get a better idea of good predictors of outages, I will need to categorize or chunk the Cause\_Code field into larger buckets.

> glimpse(ff\_steam\_df)

Observations: 23,089

Variables: 12

$ X1 <int> 16, 17, 18, 20, 21, 22, 23, 24, 26, 27, 28, 29, 31, 32, 33, 3...

$ Util\_Unit <int> 1424654553, 1424654553, 1424654553, 1424654553, 1424654553, 1...

$ start\_dt <dttm> 2015-01-12 10:51:00, 2015-01-12 11:51:00, 2015-01-12 12:12:0...

$ Type <chr> "D4", "U1", "D4", "D4", "D4", "D2", "D2", "D4", "U1", "D4", "...

$ Cause\_Code <chr> "1999", "1799", "1100", "1100", "1100", "0380", "0380", "1100...

$ end\_dt <dttm> 2015-01-12 11:51:00, 2015-01-12 12:12:00, 2015-01-12 20:30:0...

$ Time\_To\_Repair <dbl> 1.00, 0.35, 8.30, 1.13, 2.80, 0.40, 0.82, 5.97, 1.02, 0.98, 5...

$ TBF\_Period\_Hrs <dbl> 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0...

$ IA\_PH <dbl> 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.000...

$ TBF\_Service\_Hrs <dbl> 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0.00, 0...

$ IA\_SH <dbl> 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.0000, 0.000...

$ X\_Derate <dbl> 67.27, 100.00, 78.79, 78.79, 54.55, 92.73, 78.79, 54.55, 100....