Data Visualizations and Analysis

Grayson White

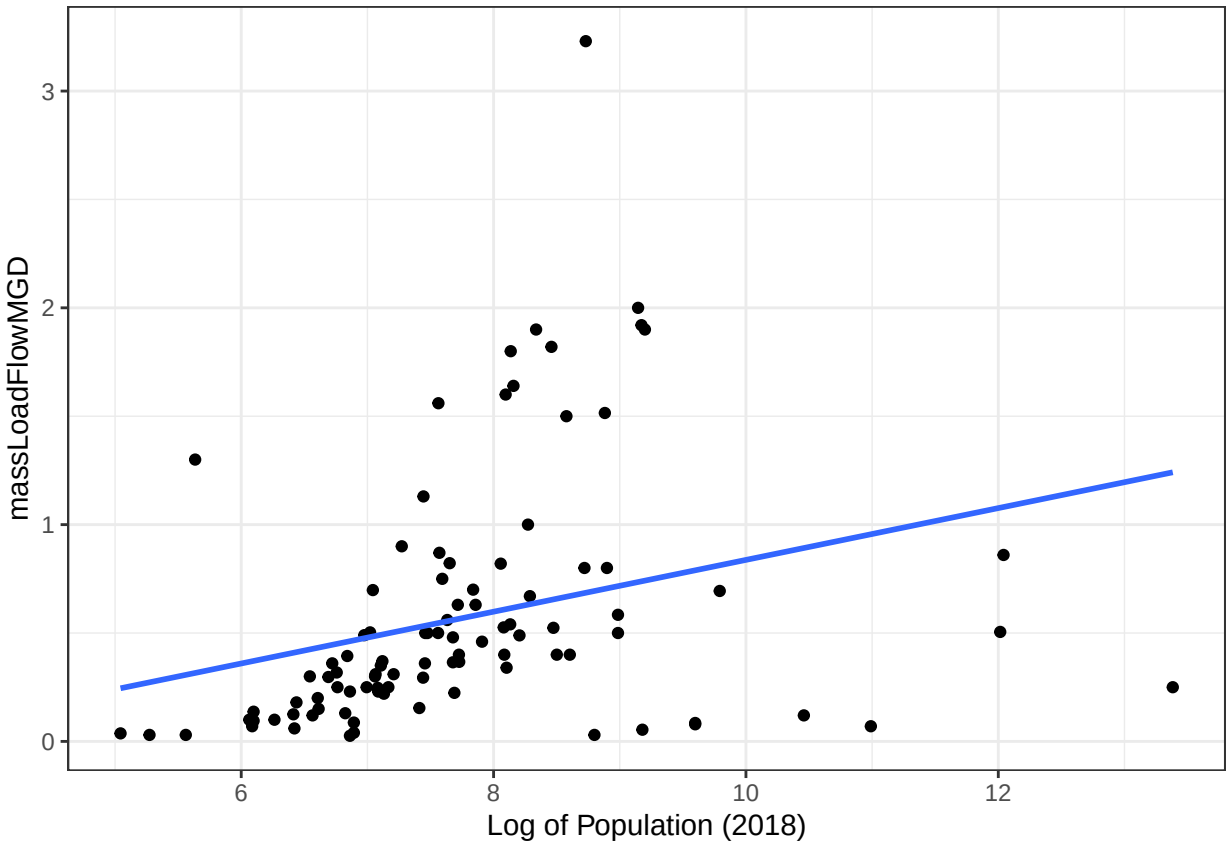
7/13/2020

```
ggplot(working_df, aes(x = log(pop_2018), y = massLoadFlowMGD)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  theme_bw() +
  labs(x = "Log of Population (2018)")
```

```
## `geom_smooth()` using formula 'y ~ x'
```

```
## Warning: Removed 17 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 17 rows containing missing values (geom_point).
```

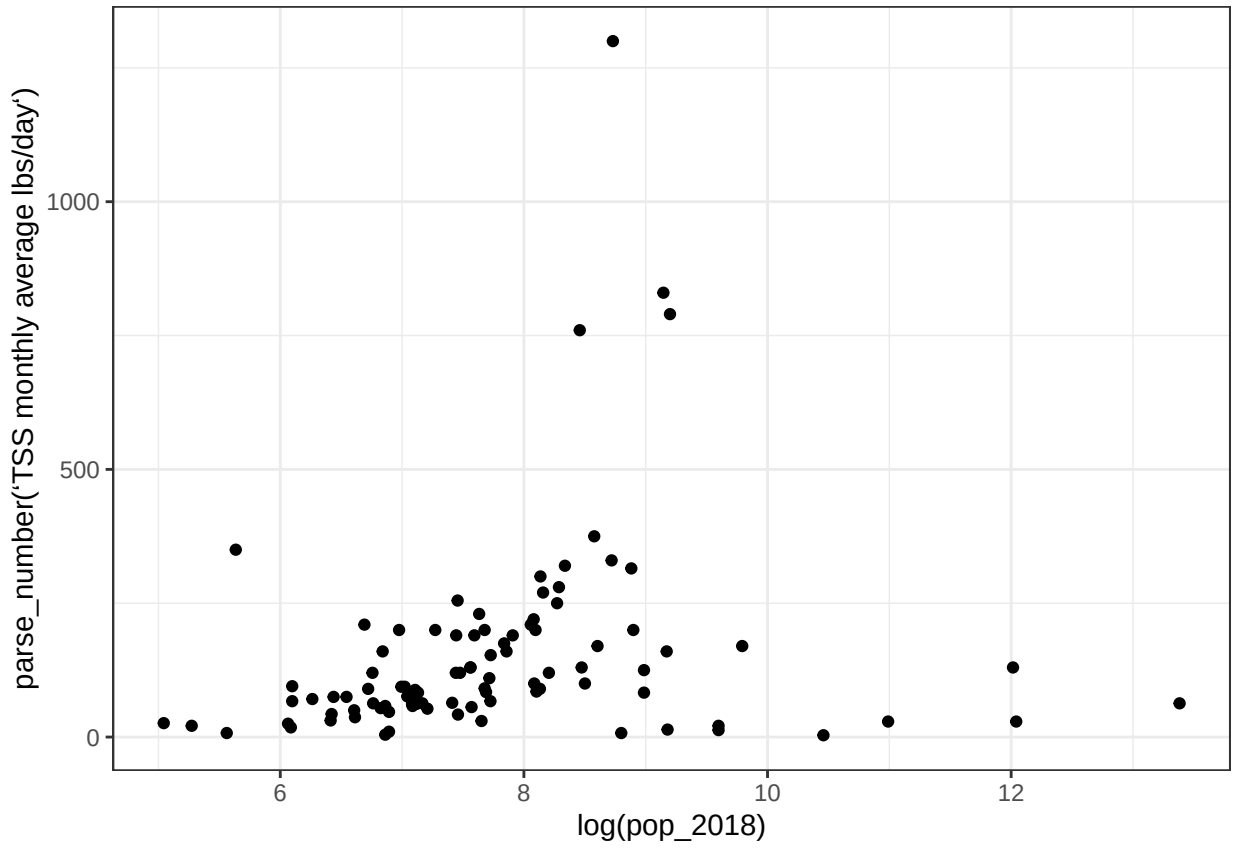


```
ggplot(working_df, aes(x = log(pop_2018),
                       y = parse_number(`TSS monthly average lbs/day`))) +
  geom_point() +
  theme_bw()
```

```
## Warning: 2 parsing failures.
## row col expected actual
##  13  -- a number     na
##  45  -- a number     na
```

```
## Warning: 2 parsing failures.
## row col expected actual
##  13  -- a number     na
##  45  -- a number     na
```

```
## Warning: Removed 19 rows containing missing values (geom_point).
```




```
working_df %>%
  filter(type1 %in% c("lagoons", "activated sludge")) %>%
  group_by(type1) %>%
  summarize(median = median(pop_2018, na.rm = TRUE))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
## # A tibble: 2 x 2
##   type1            median
##   <chr>             <dbl>
## 1 activated sludge  1962.
## 2 lagoons           1718.
```