Post Fire Water Quality. Analysis of water quality after the Camp Fire in Paradise, CA

Claire Morrison and Lauren Magliozzi

2/5/2022

Introduction:

The 2018 Camp Fire was the most destructive wildfire in California history, burning over 18,000 structures, largely residences, in the town of Paradise, CA in less than four hours. The fire was unprecedented due to its urban nature and upper-watershed setting. Because the fire was late in the fall, precipitation occurred prior to emergency clean up and erosion had already mobilized fire debris into downstream creeks and reservoirs. The transport of water contaminants including metals (e.g. zinc and lead) from ash of burned homes and cars by storm runoff was a major concern for the surrounding waterways. This study characterized the effects of watershed burning in a wildland-urban interface (WUI) on water quality during the first water year post-fire. The study measured metal concentrations and evaluated the fate and transport from the burn area to downstream locations during every major storm event of the wet 2019 water year.

Question of Interest:

This data analysis will focus on identifying the relationship between a common, easy to analyze metal (zinc) with a toxic metal that can be more expensive to measure in the lab (lead) as well as the effects of precipitation and extent of watershed burning.

All materials for this project can be found at https://github.com/clairemo22/watershed_metals

The data was initially stored in a single Excel sheet, in three tabs. The first contained metal concentration, the second had ion concentration and physical measurements like flow, the third tab had traits of each watershed.

```
metals_in<- read_csv("https://raw.githubusercontent.com/clairemo22/watershed_metals/main/CampFire_metal
```

```
## Rows: 5920 Columns: 19

## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (16): Chico State Project ID, ICP-MS ID, Burned, Urban, Watershed Full, ...
## dbl  (3): Original Chico State seq. no., Dilution Factor, Concentration (ppb)

##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
traits<- read_csv("https://raw.githubusercontent.com/clairemo22/watershed_metals/main/CampFire_watershe
```

```
## Rows: 8 Columns: 10

## -- Column specification --------------------------------------------------------
## Delimiter: ","
```

```
## chr (2): Watershed Full, Burn per A
## dbl (8): Area, Burned Structures, Urban Percent, Forest Percent, Impervious ...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
#the Time column wasn't read in with the correct format so it has to be
#extracted from the Date & Time Sampled column

metals<- metals_in %>%
  mutate(Date = parse_date(`Date Sampled`, format = "%m/%d/%y"),
  "Concentration (ppb)" = as.numeric(`Concentration (ppb)`)) %>%
  pivot_wider(names_from = Metal,
              values_from = "Concentration (ppb)")

#ions = ions_in %>%
 # mutate(Date = ymd(`Date Sampled`),
  #       Concentration = as.numeric(Concentration)) %>%
  ##pivot_wider(names_from = Ions,
             # values_from = Concentration)

dat = left_join(metals,traits)
```

```
## Joining, by = "Watershed Full"
```

```
dat = dat %>% mutate(Watershed = na_if(Watershed, "NA"))
```

```
dat<- dat %>% drop_na(Watershed) %>%
  rename(urban_percent= `Urban Percent`,
         burned_structures= `Burned Structures`) %>%
    mutate(burn_per_acre= as.numeric(`Burn per A`)) %>%
  select(Watershed, Date, Na, Mg, Al, K, Zn, Pb, burned_structures, urban_percent, burn_per_acre)
```

Include Precipitation data from California Department of Water Resources. https://cdec.water.ca.gov/dynamicapp/wsSensorData The station ID is PDE (Paradise), with Sensor Number 45. This data provides daily incrimental precipitation in inches.

```
precip<- read_csv("https://cdec.water.ca.gov/dynamicapp/req/CSVDataServlet?Stations=PDE&SensorNums=45&du
```

```
## Rows: 506 Columns: 9
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (5): STATION_ID, DURATION, SENSOR_TYPE, VALUE, UNITS
## dbl  (1): SENSOR_NUMBER
## lgl  (1): DATA_FLAG
## dttm (2): DATE TIME, OBS DATE
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

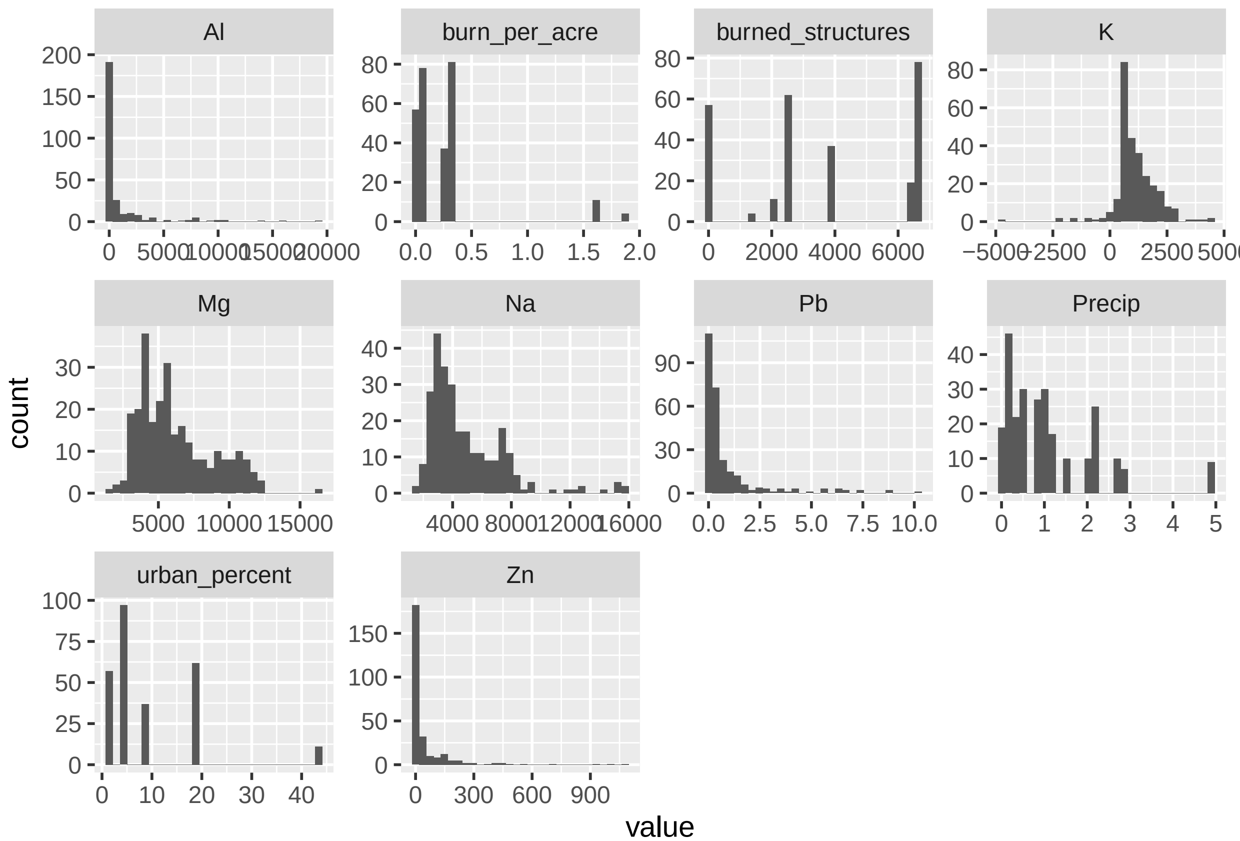
```
precip <- precip %>%
  mutate(Date = ymd(`OBS DATE`),
         Precip= as.numeric(VALUE)) %>%
  select(Date,Precip)
```

```r
dat <- dat %>%
  left_join(precip)
```

```
## Joining, by = "Date"
```

```r
dat %>%
  keep(is.numeric) %>%
  gather() %>%
  ggplot(aes(value)) +
    facet_wrap(~ key, scales = "free") +
    geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

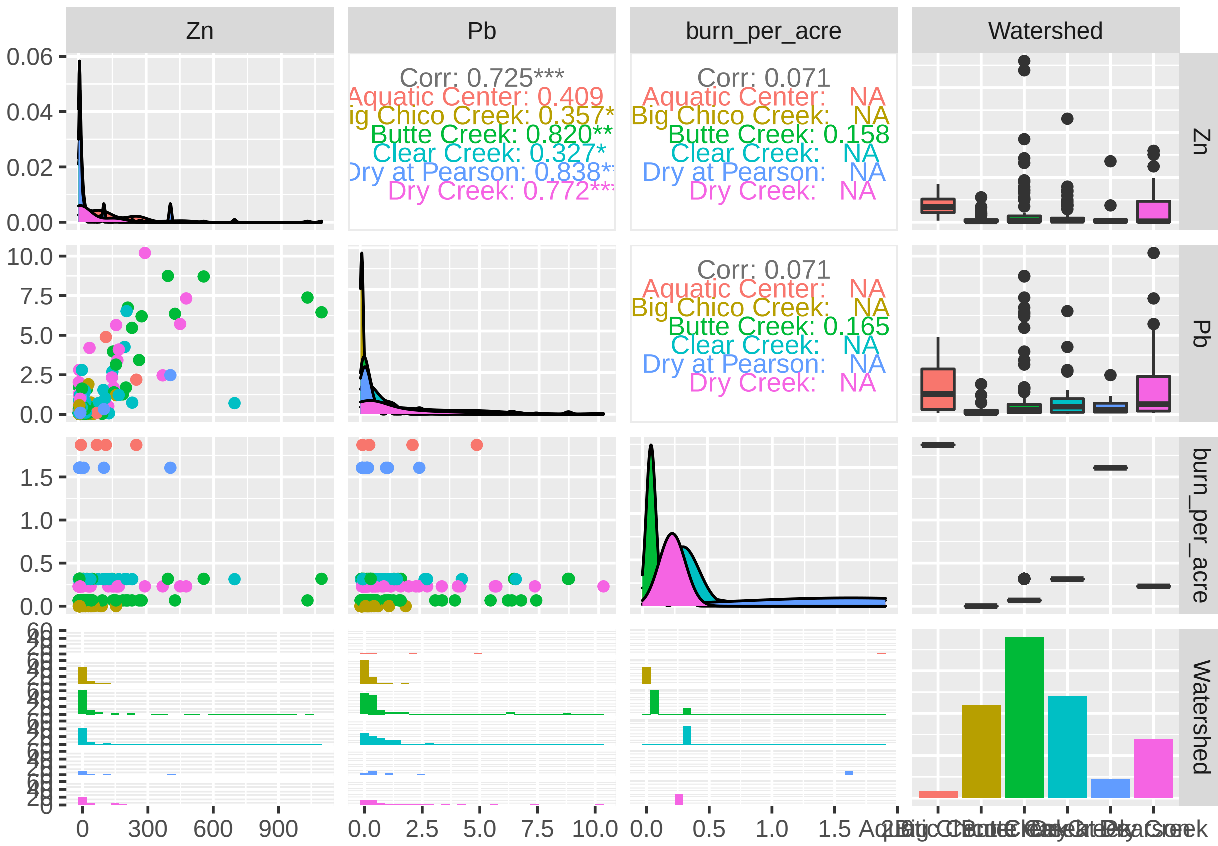


Zinc, lead burn per acre and precipitation are highly left skewed, so if we log transform and bin to be categorical they will be more normal.

Here are some descriptive plots and statistics for our main variables of interest.

```r
dat %>% select(Zn, Pb, burn_per_acre,Watershed ) %>%
  ggpairs(aes(colour=Watershed),
          upper=list(continuous=wrap('cor',size=3.5)))
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



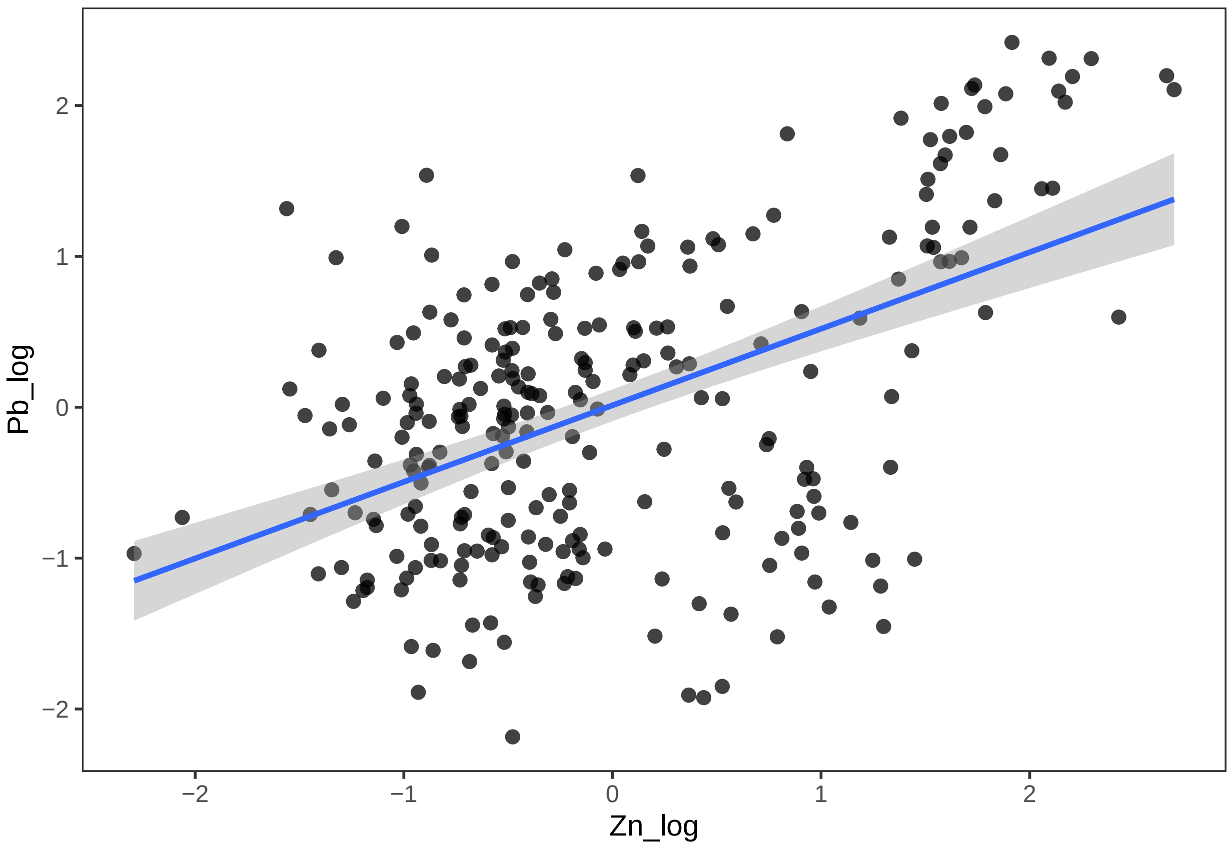
Model number 1: How do zinc and lead relate?

```
summary(lmer(Pb_log~Zn_log+(1|Watershed), data = dat))
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: Pb_log ~ Zn_log + (1 | Watershed)
##    Data: dat
##
## REML criterion at convergence: 629.8
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.73203 -0.70788  0.09882  0.71463  2.03204
##
## Random effects:
##  Groups    Name        Variance Std.Dev.
##  Watershed (Intercept) 0.1375   0.3708
##  Residual              0.6317   0.7948
## Number of obs: 258, groups:  Watershed, 6
##
## Fixed effects:
##             Estimate Std. Error        df t value Pr(>|t|)
## (Intercept)  0.05883    0.16832   5.40333   0.349     0.74
## Zn_log       0.45853    0.05052 254.59029   9.077   <2e-16 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##        (Intr)
## Zn_log -0.025
```

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



Model number 2: Does burn level predict zinc and lead levels?

```
summary(lmer(Zn_log~ burn_per_acre_cat+(1|Watershed), data = dat))
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: Zn_log ~ burn_per_acre_cat + (1 | Watershed)
##    Data: dat
##
## REML criterion at convergence: 731
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.2408 -0.6925 -0.2801  0.7447  2.7713
##
## Random effects:
##  Groups    Name        Variance Std.Dev.
##  Watershed (Intercept) 0.009985 0.09993
##  Residual              0.975073 0.98746
```

```
## Number of obs: 258, groups:  Watershed, 6
##
## Fixed effects:
##                            Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)                  0.1323     0.1029  6.7918   1.286   0.2406
## burn_per_acre_catLow Burn   -0.2785     0.1440  8.0530  -1.934   0.0889 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr)
## brn_pr_c_LB -0.649
```

```
summary(lmer(Pb_log~burn_per_acre_cat +(1|Watershed), data = dat))
```

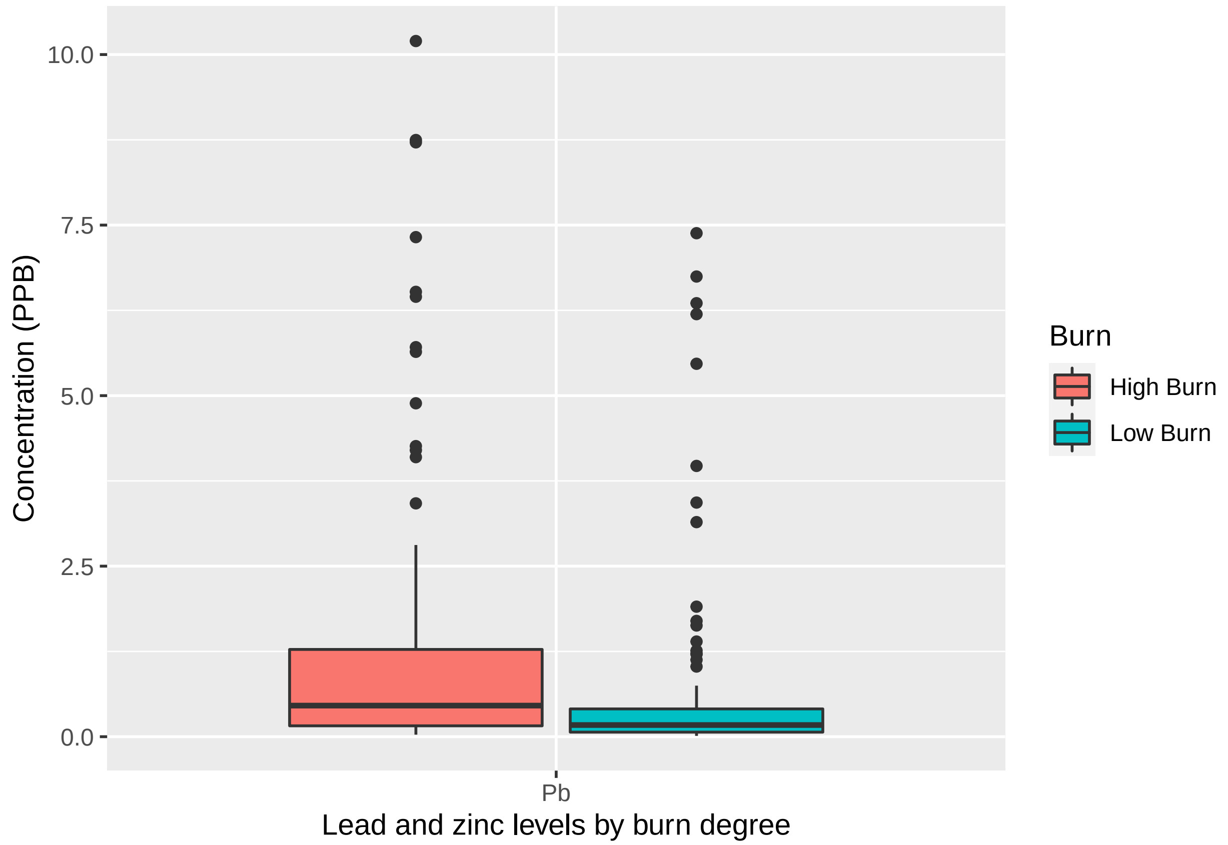
```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: Pb_log ~ burn_per_acre_cat + (1 | Watershed)
##    Data: dat
##
## REML criterion at convergence: 699.4
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.05604 -0.70856 -0.08606  0.69674  2.47445
##
## Random effects:
##  Groups    Name        Variance Std.Dev.
##  Watershed (Intercept) 0.09183  0.3030
##  Residual              0.82789  0.9099
## Number of obs: 260, groups:  Watershed, 6
##
## Fixed effects:
##                            Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)                  0.2614     0.1623  6.1648   1.611   0.1571
## burn_per_acre_catLow Burn   -0.5066     0.1901 32.9635  -2.665   0.0118 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr)
## brn_pr_c_LB -0.434
```

```
long<- dat %>% pivot_longer(
  cols = c(Pb, Zn),
  names_to = "var",
  values_to = "rate") %>%
  select(Watershed, var, rate, burn_per_acre_cat, precip_cat)

lead<- long %>% filter(var=="Pb")

ggplot(lead,aes(x=var,y=rate, fill=factor(burn_per_acre_cat)))+
  geom_boxplot()+
  xlab("Lead and zinc levels by burn degree")+
  ylab("Concentration (PPB)")+
```

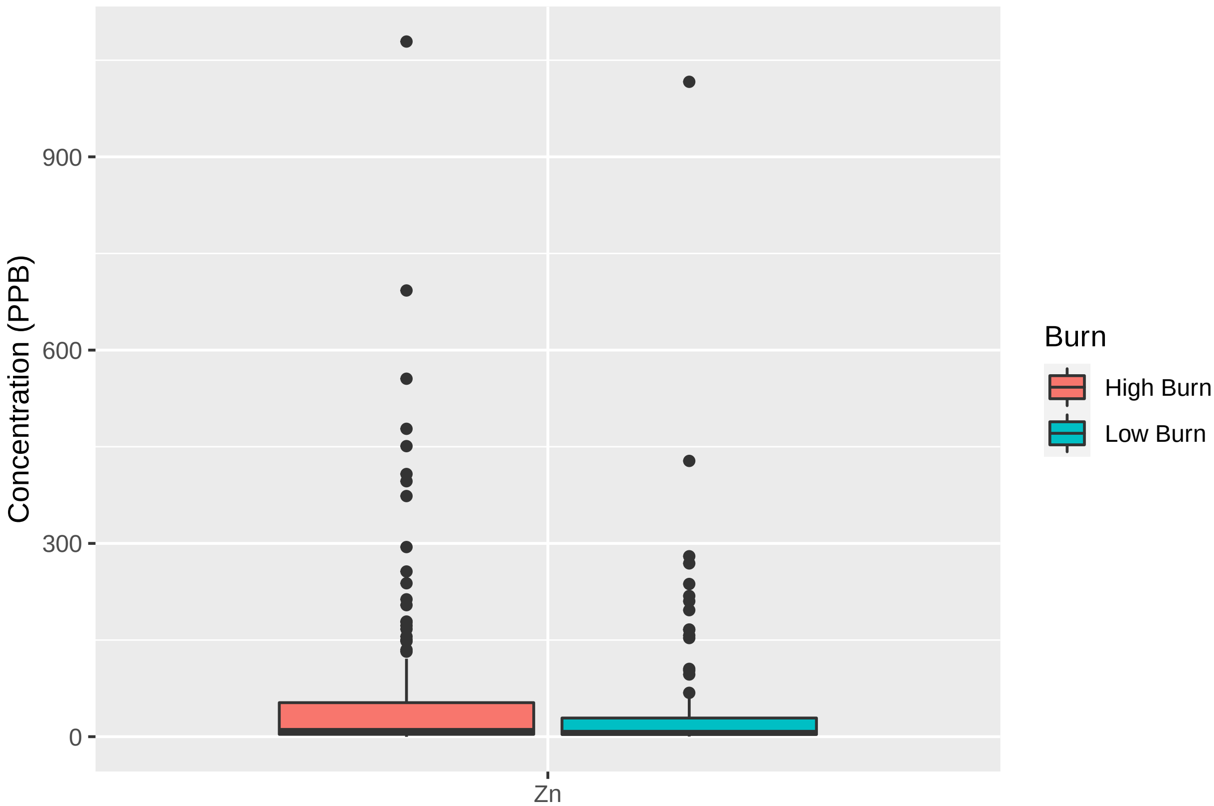
```
  scale_fill_discrete(name = "Burn", labels = c("High Burn", "Low Burn"))
```




```
zinc<- long %>% filter(var=="Zn")

ggplot(zinc,aes(x=var,y=rate, fill=factor(burn_per_acre_cat)))+
  geom_boxplot()+
  xlab("Lead and zinc levels by burn degree")+
  ylab("Concentration (PPB)")+
  scale_fill_discrete(name = "Burn", labels = c("High Burn", "Low Burn"))
```

```
## Warning: Removed 1 rows containing non-finite values (stat_boxplot).
```



Lead and zinc levels by burn degree

Model number 3: Does precipitation predict zinc and lead levels?

```
summary(lmer(Pb_log~Precip+(1|Watershed), data = dat))
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: Pb_log ~ Precip + (1 | Watershed)
##    Data: dat
##
## REML criterion at convergence: 695.7
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.20482 -0.78884 -0.05714  0.64684  2.52809
##
## Random effects:
##  Groups    Name        Variance Std.Dev.
##  Watershed (Intercept) 0.2086   0.4567
##  Residual              0.7977   0.8932
## Number of obs: 260, groups:  Watershed, 6
##
## Fixed effects:
##              Estimate Std. Error        df t value Pr(>|t|)
## (Intercept)  -0.12022    0.21373   6.19597  -0.562 0.593542
## Precip        0.18164    0.05203 255.47888   3.491 0.000567 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##        (Intr)
## Precip -0.292
```

```r
summary(lmer(Zn_log~Precip+(1|Watershed), data = dat))
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: Zn_log ~ Precip + (1 | Watershed)
##    Data: dat
##
## REML criterion at convergence: 733
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.3025 -0.7280 -0.2469  0.7260  2.7229
##
## Random effects:
##  Groups    Name        Variance Std.Dev.
##  Watershed (Intercept) 0.02507  0.1583
##  Residual              0.96991  0.9848
## Number of obs: 258, groups:  Watershed, 6
##
## Fixed effects:
##              Estimate Std. Error        df t value Pr(>|t|)
## (Intercept)  -0.10216    0.11589   6.14149  -0.881   0.4112
## Precip        0.10505    0.05741 254.43745   1.830   0.0684 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##        (Intr)
## Precip -0.552
```

Model number 4: does precipation interact with the relationship between zinc and lead?

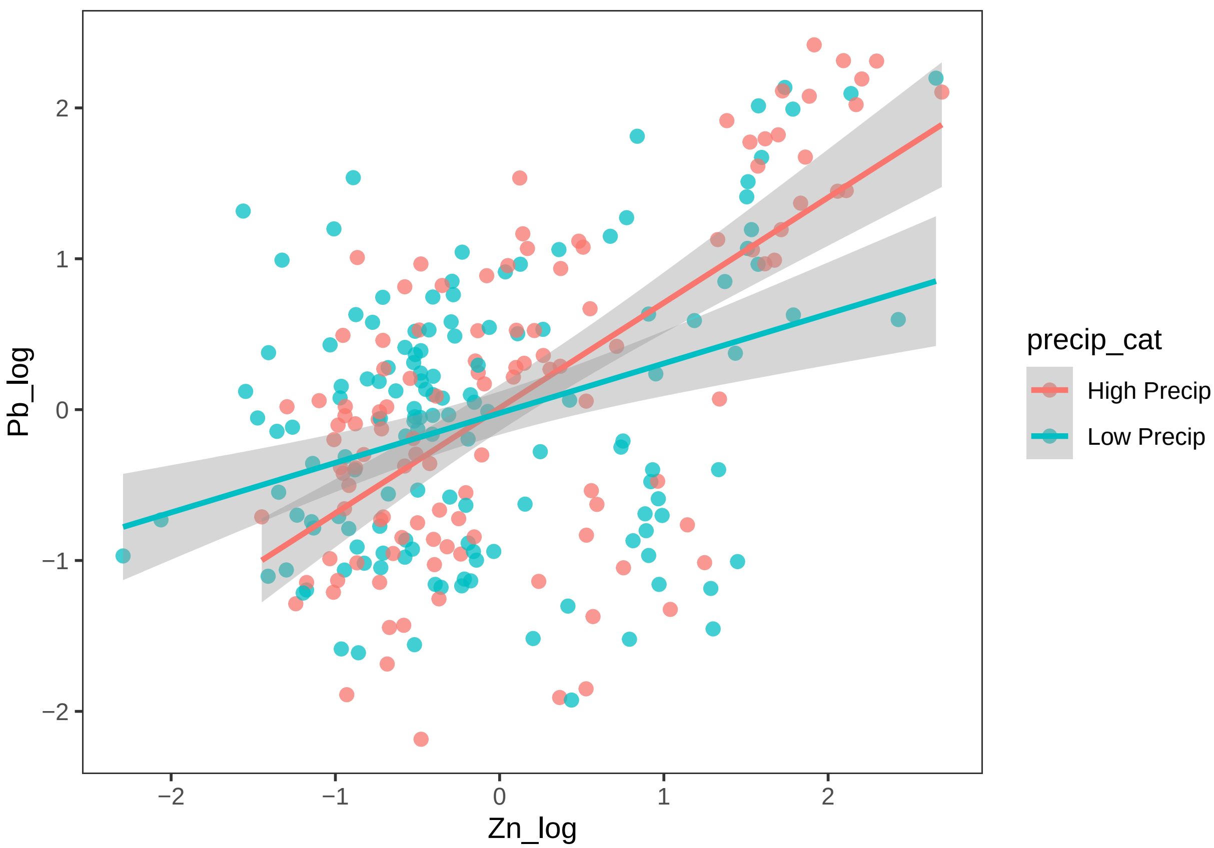
```r
summary(lmer(Pb_log~Zn_log+precip_cat+Zn_log*precip_cat+(1|Watershed), data = dat))
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: Pb_log ~ Zn_log + precip_cat + Zn_log * precip_cat + (1 | Watershed)
##    Data: dat
##
## REML criterion at convergence: 623.6
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.59543 -0.68927  0.06772  0.70593  2.10789
##
## Random effects:
##  Groups    Name        Variance Std.Dev.
##  Watershed (Intercept) 0.1350   0.3675
##  Residual              0.6079   0.7797
```

```
## Number of obs: 258, groups:  Watershed, 6
##
## Fixed effects:
##                              Estimate Std. Error        df t value Pr(>|t|)
## (Intercept)                   0.08780    0.17372   6.45084   0.505  0.63007
## Zn_log                        0.61896    0.07273 253.53525   8.510 1.54e-15 ***
## precip_catLow Precip         -0.12917    0.10025 253.11591  -1.289  0.19874
## Zn_log:precip_catLow Precip  -0.31595    0.09909 252.09130  -3.188  0.00161 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) Zn_log prc_LP
## Zn_log      -0.086
## prcp_ctLwPr -0.278  0.092
## Zn_lg:pr_LP  0.063 -0.727 -0.011
```

```
ggplot(dat,aes(Zn_log, Pb_log, na.rm = T, col=precip_cat)) +
  geom_point(size = 2, alpha = .75, position = "jitter", na.rm = T) +
  geom_smooth(na.rm = T, method = "lm", se = T, linetype = 1)+
  theme_bw()+
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank())
```

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



Conclusions:

Zinc and lead.

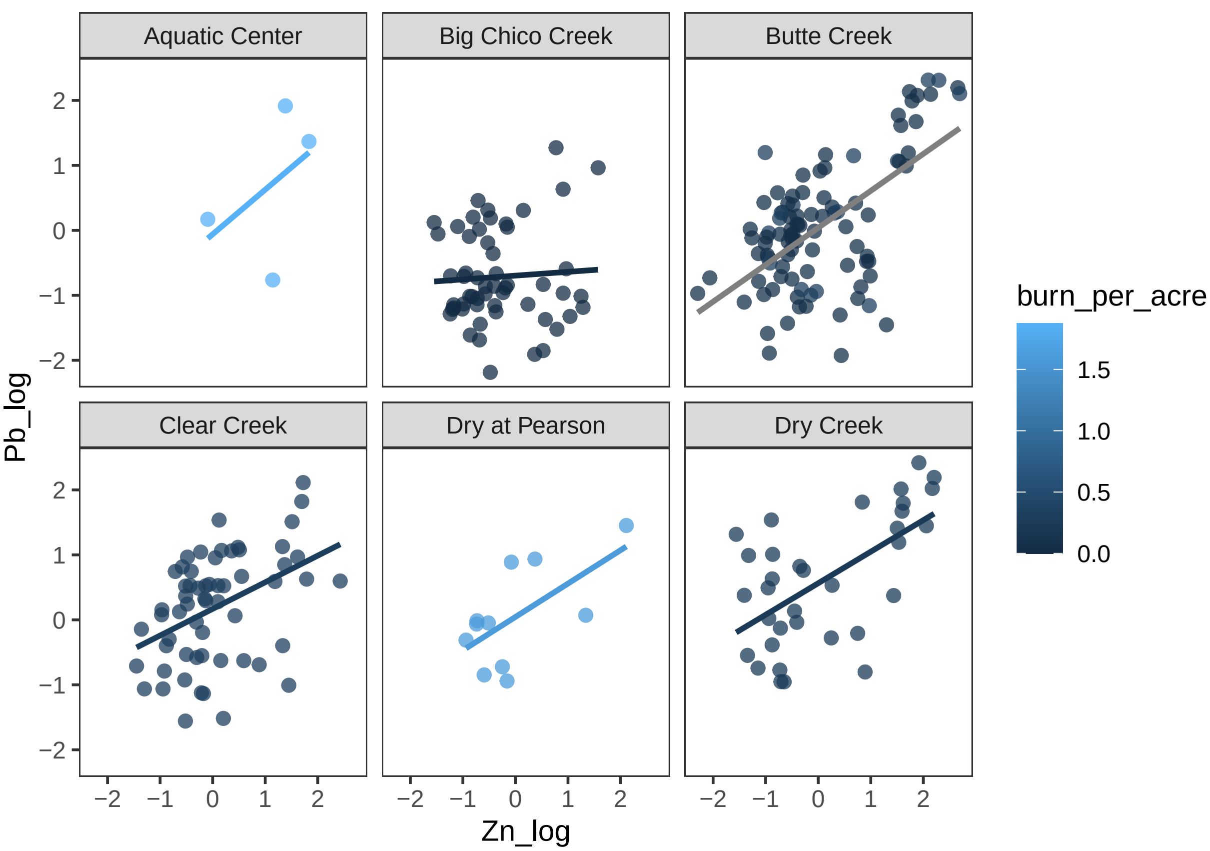
From these data, we found zinc levels significantly predicted lead levels in water samples taken after the Camp Fire. We saw a strong and linear relationship between zinc and lead. This is an important finding because zinc is a much easier and less toxic, metal to analyze than lead, so if they are predictive of each other it may be a good proxy for efficiently measuring lead in water after fires.

Burn severity.

Next, we wanted to see if burn per acre predicts the amount of lead and zinc in the water shed. Because burn per acre was highly left skewed, we performed a median split to transform it to a categorical variable. We then used that categorical variable to predict both zinc and lead, and found the difference between high and low burn severity did predict lead but not zinc.

The figure below shows the relationship between zinc and lead, in these data where burn occurred, are not significantly mediated by burn per acre. However, the plot shows that there is almost no relationship between lead and zinc in the Big Chico Creek watershed. This watershed was unaffected by the fire and is used here as a control, so this lack of relationship supports the idea that the lead and zinc relationship was caused by the fire, and should be investigated further.

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



Precipitation.

Given rain can have a large influence on watersheds and samples, we wanted to see how precipitation relates to both zinc and lead. Interestingly, we found precipitation only influenced lead, but not zinc.

Lastly, if future studies were to measure zinc as a proxy for lead, it would be useful to know how precipitation influences the zinc-lead relationship, because we found precipitation does significantly influence lead.

Interestingly, we found precipitation and zinc had a significant interaction on lead prediction, such that in areas with high precipitation, the relationship between lead and zinc was much more pronounced.

Limitations/ bias

Some limitations of this data are the sample size. From some of these plots, you can tell the number of samples are sparse, which significantly decreases our power to detect significant associations.

It is also important to recognize our biases in this study. A bias in this analysis is the assumption that major differences between Big Chico Creek and other watersheds are due to the fire. While the lack of fire in Big Chico Creek is a substantial difference, there are still other differences, such as geology, which could impact the different relationships we see in metal concentrations.