CalEnviroScreen describe

2/17/2022

Load data

```
dat<- st_read(dsn = "C:/Users/dnmed/OneDrive - cumc.columbia.edu/AAMEHS/OEHHA",
              layer = "CES4 Final Shapefile", geometry_column = "geometry")
```

```
## Reading layer 'CES4 Final Shapefile' from data source
##   'C:\Users\dnmed\OneDrive - cumc.columbia.edu\AAMEHS\OEHHA'
##   using driver 'ESRI Shapefile'
## Simple feature collection with 8035 features and 66 fields
## Geometry type: MULTIPOLYGON
## Dimension:     XY
## Bounding box:  xmin: -373976.1 ymin: -604512.6 xmax: 539719.6 ymax: 450022.5
## Projected CRS: NAD83 / California Albers
```

```
#replace -999 with NA
dat<- dat %>%
  mutate_if(is.numeric, ~na_if(., -999))

#summary(dat)

dat.dw<- read_csv("C:/Users/dnmed/OneDrive - cumc.columbia.edu/As Cancer Project/Data/ces20drinkingwate

#convert to numeric
dat.dw[,6:31]<- apply(dat.dw[,6:31], 2, as.numeric)


#colnames(dat)
#colnames(dat.dw)

dat_merge<- dat %>% full_join(dat.dw[,c(1,6:31)], by=c("Tract"="Census Tract"))

dat_merge_sf<- dat_merge

#drop geometry for summarizing bc it sometimes produces an error
dat_merge<- dat_merge %>% st_drop_geometry()
```

Summarize data

Drinking water contaminants

```
#summary(dat_merge)

#Metals in drinking water

summary(dat_merge$Arsenic)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   0.000   0.106   0.700   1.212   1.751  35.689     103
```

```
summary(dat_merge$Uranium)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   0.000   0.125   1.796   2.197   2.685 159.732     314
```

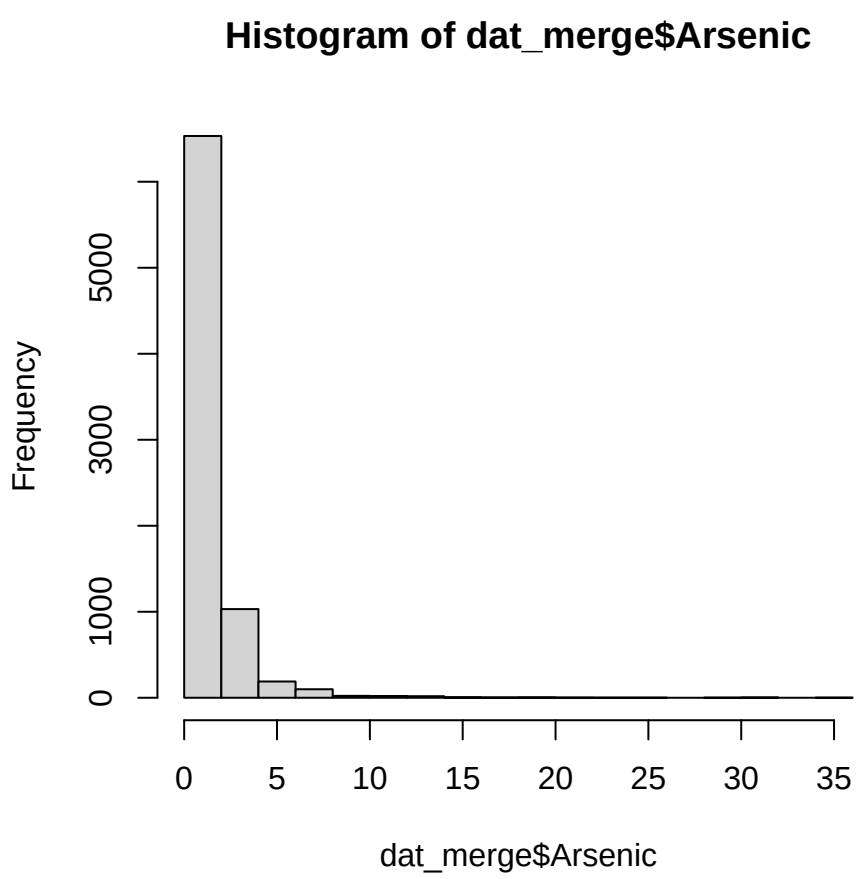
```
summary(dat_merge$Cadmium)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
## 0.00000 0.00000 0.00000 0.00721 0.00000 5.00000     177
```

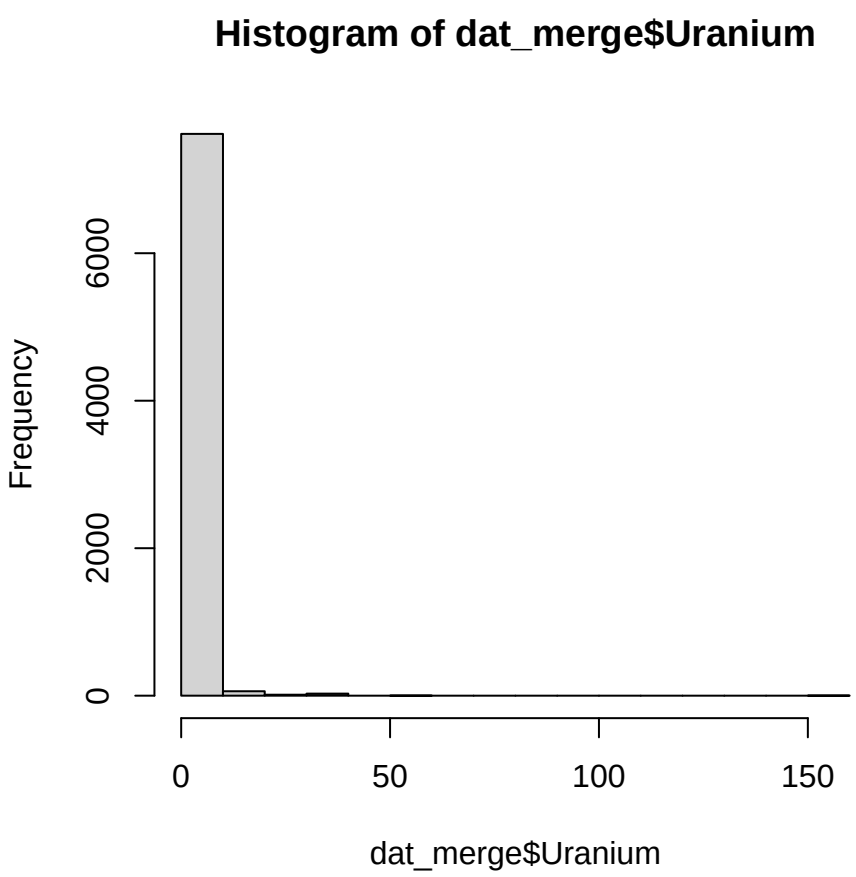
```
summary(dat_merge$Lead.y)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.      NA's
##    0.0000    0.0000    0.0000    0.4908    0.0060 1332.7940       219
```

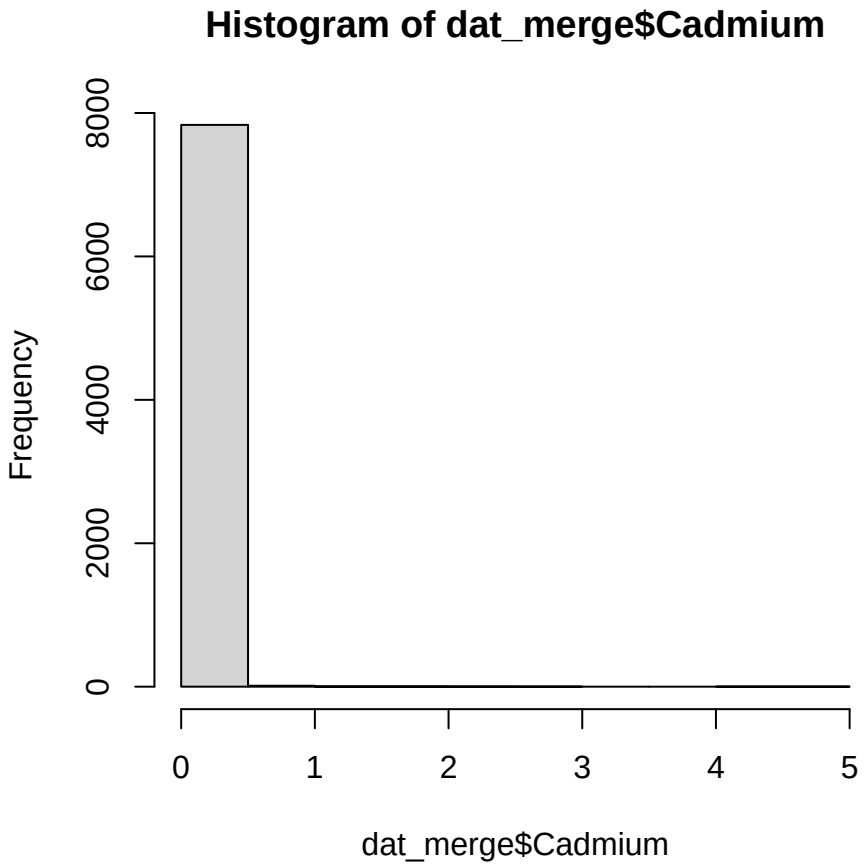
```
hist(dat_merge$Arsenic)
```



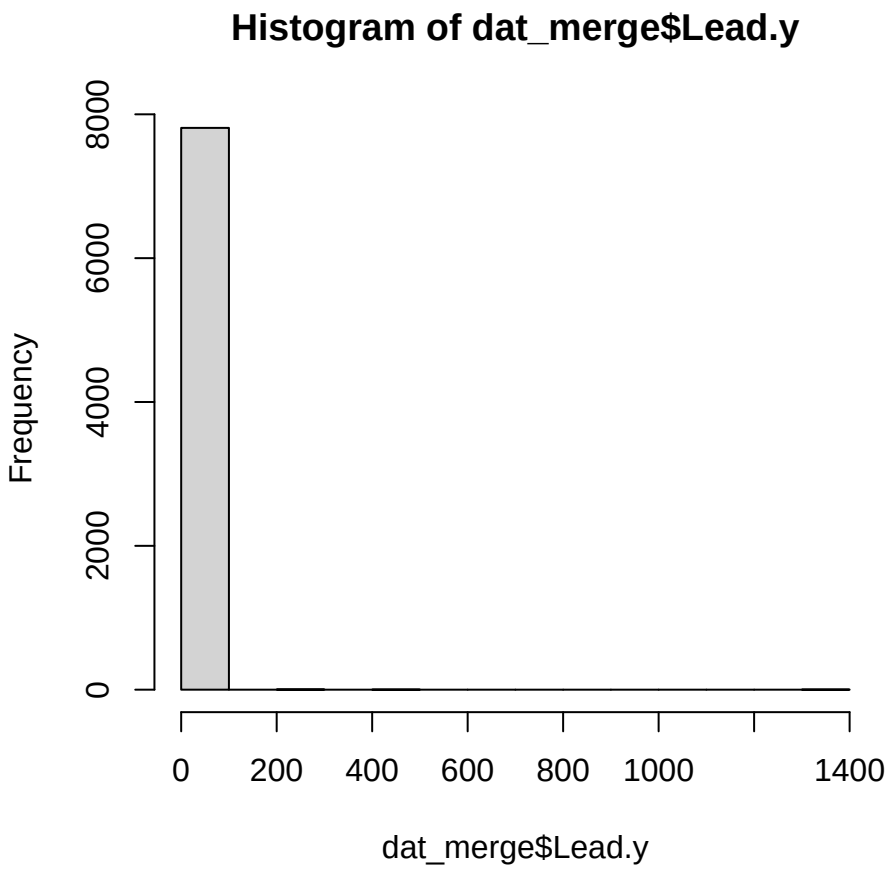

```
hist(dat_merge$Uranium)
```




```
hist(dat_merge$Cadmium)
```




```
hist(dat_merge$Lead.y)
```




```
#Fuller distribution

quantile(dat_merge$Arsenic, seq(0,1,by=0.05), na.rm = TRUE)
```

```
##       0%       5%      10%      15%      20%      25%      30%      35%
##  0.00000  0.00000  0.00000  0.00000  0.02000  0.10600  0.24860  0.41300
##      40%      45%      50%      55%      60%      65%      70%      75%
##  0.51500  0.61800  0.70000  0.86305  0.98360  1.26300  1.36100  1.75100
##      80%      85%      90%      95%     100%
##  2.00000  2.37000  2.40000  3.78890 35.68900
```

```
quantile(dat_merge$Uranium, seq(0,1,by=0.05), na.rm = TRUE)
```

```
##      0%      5%     10%     15%     20%     25%     30%     35%     40%     45%
##   0.000   0.000   0.000   0.000   0.000   0.125   0.609   0.988   1.276   1.422
##     50%     55%     60%     65%     70%     75%     80%     85%     90%     95%
##   1.796   1.900   2.299   2.300   2.400   2.685   3.120   3.724   4.771   6.382
##    100%
## 159.732
```

```
quantile(dat_merge$Cadmium, seq(0,1,by=0.05), na.rm = TRUE)
```

```
##    0%    5%   10%   15%   20%   25%   30%   35%   40%   45%   50%   55%   60%
## 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
##   65%   70%   75%   80%   85%   90%   95%  100%
```

```
## 0.000 0.000 0.000 0.000 0.000 0.000 0.002 5.000
```

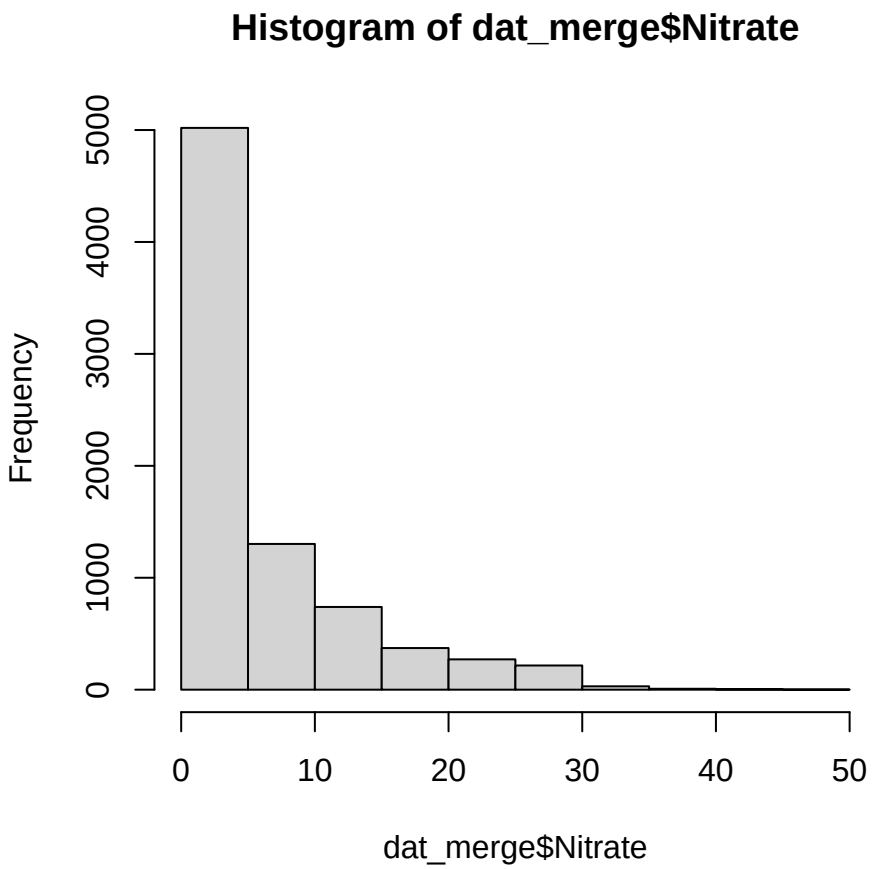
```
quantile(dat_merge$Lead.y, seq(0,1,by=0.05), na.rm = TRUE)
```

```
##          0%          5%         10%         15%         20%         25%         30%
##     0.00000     0.00000     0.00000     0.00000     0.00000     0.00000     0.00000
##         35%         40%         45%         50%         55%         60%         65%
##     0.00000     0.00000     0.00000     0.00000     0.00000     0.00000     0.00000
##         70%         75%         80%         85%         90%         95%        100%
##     0.00000     0.00600     0.03200     0.09900     0.18800     0.45875 1332.79400
```

```
#dat_merge<- dat_merge %>% mutate(ArsenicQ= cut2(Arsenic, g=4),
#                    UraniumQ=cut2(Uranium, g=4),
#                    CadmiumQ=cut2(Cadmium,g=4),
#                    LeadQ= cut2(Lead.y, g=4))

#table(dat_merge$ArsenicQ)
#table(dat_merge$UraniumQ)
#table(dat_merge$CadmiumQ)
#able(dat_merge$LeadQ)



#Nitrate better distribution

summary(dat_merge$Nitrate)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   0.000   0.415   3.500   5.702   8.355  49.897      73
```

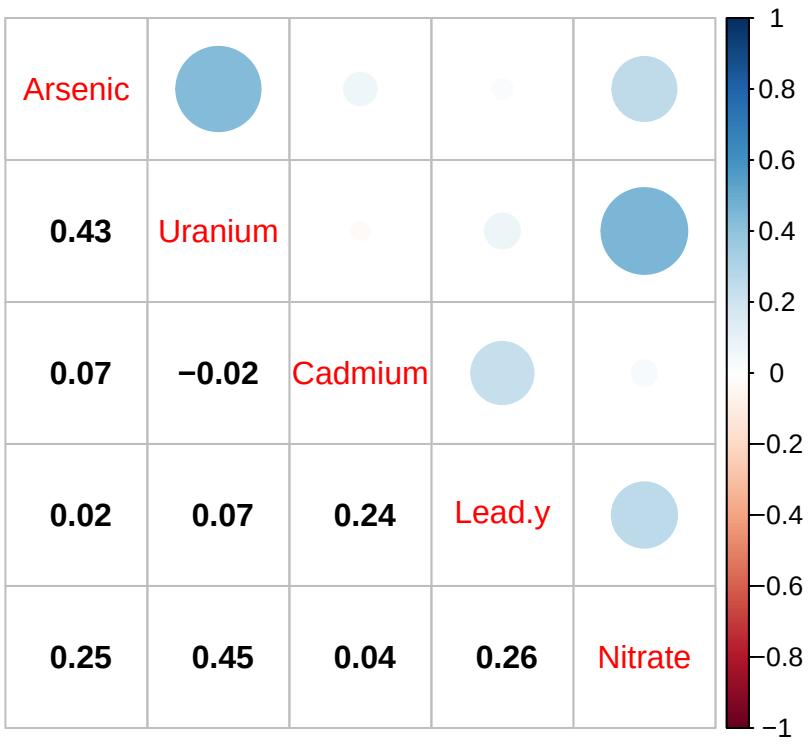
```
hist(dat_merge$Nitrate)
```




```
quantile(dat_merge$Nitrate, seq(0,1,by=0.05), na.rm = TRUE)
```

```
##       0%       5%      10%      15%      20%      25%      30%      35%
##  0.00000  0.00000  0.01800  0.08230  0.20200  0.41500  0.75500  1.10000
##      40%      45%      50%      55%      60%      65%      70%      75%
##  1.67620  2.86345  3.50000  3.80000  4.01260  5.38700  6.35620  8.35500
##      80%      85%      90%      95%     100%
## 10.16900 12.32600 16.54570 22.36100 49.89700
```

```
#correlation

metals<- dat_merge %>% select(Arsenic, Uranium, Cadmium, Lead.y, Nitrate)
cormetal<- cor(metals,use="complete.obs",method = "spearman")
corrplot.mixed(cormetal, lower.col = "black")
```



Air pollution

```
summary(dat_merge$Ozone)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.02655 0.04193 0.04716 0.04867 0.05680 0.07313
```

```
summary(dat_merge$PM2_5)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.875   8.575  10.119  10.153  11.938  16.395
```

```
summary(dat_merge$DieselPM)
```

```
##       Min.    1st Qu.     Median       Mean    3rd Qu.       Max.
##  0.000052  0.067683  0.144929  0.225442  0.286794 14.611221
```

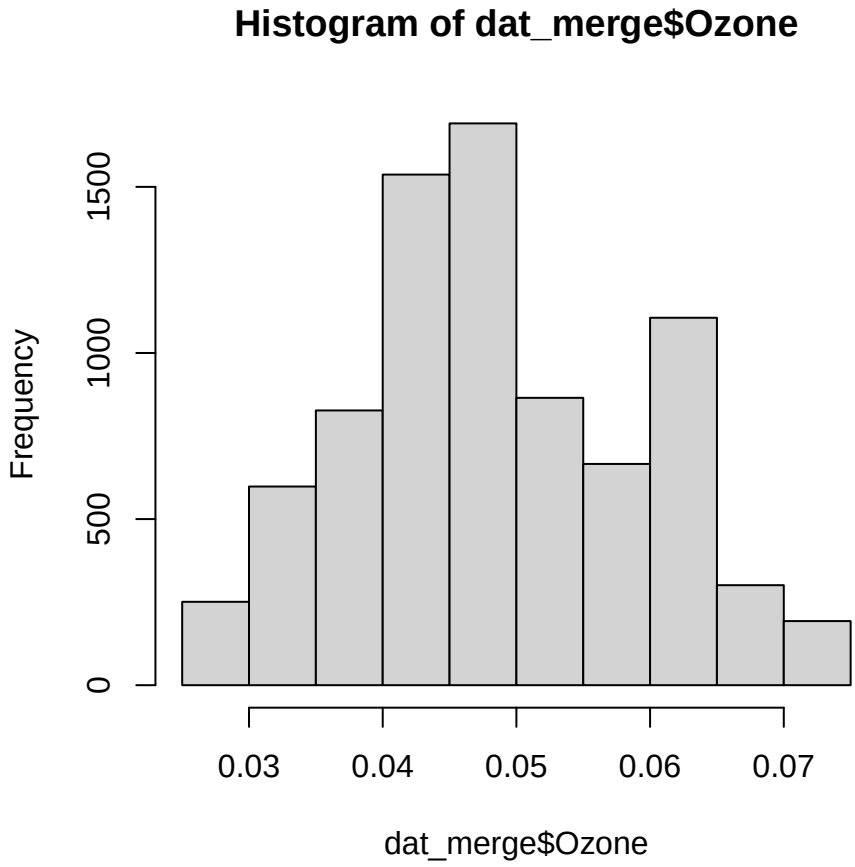
```
summary(dat_merge$Tox_Rel)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     0.0   111.5   456.3  1624.0  1625.5 96985.6
```

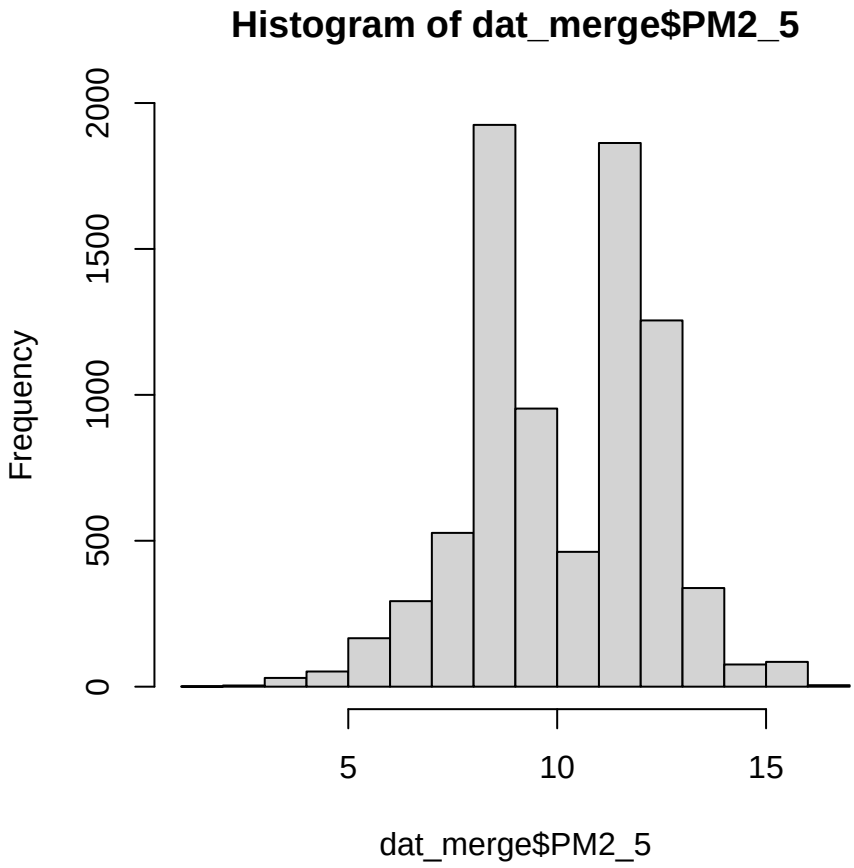
```
summary(dat_merge$Traffic)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.     NA's
##    20.75   553.80   881.04  1117.45  1386.57 45752.00       35
```

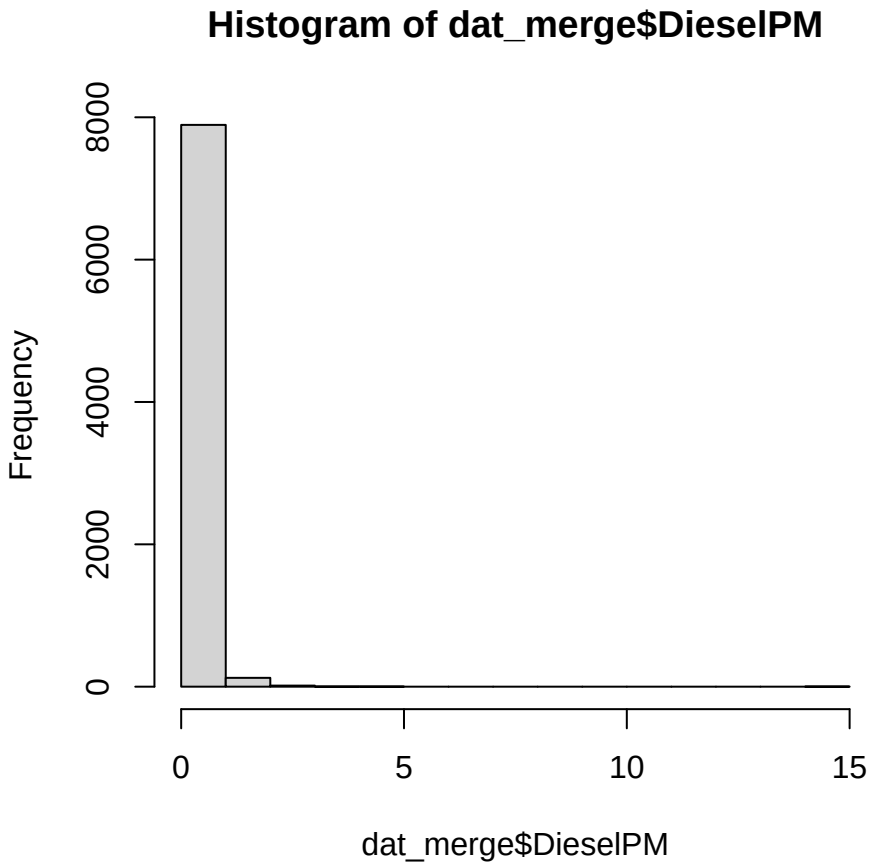
```
hist(dat_merge$Ozone)
```



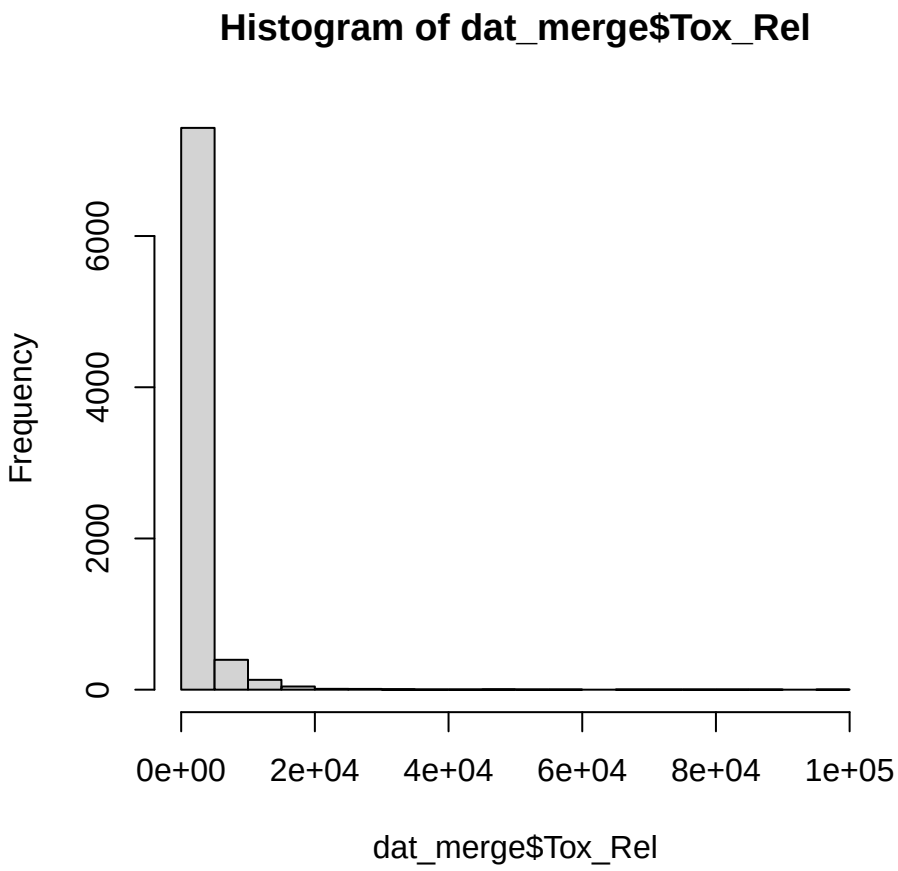

```
hist(dat_merge$PM2_5)
```



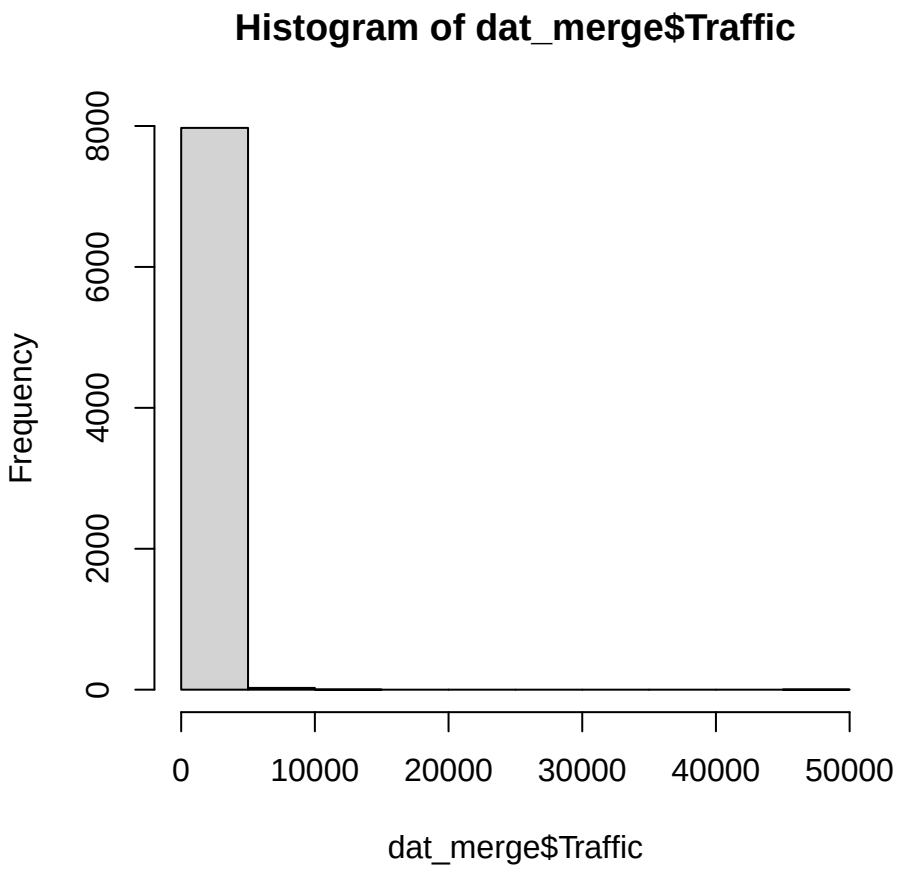

```
hist(dat_merge$DieselPM)
```




```
hist(dat_merge$Tox_Rel)
```




```
hist(dat_merge$Traffic)
```




```
#Fuller distribution

quantile(dat_merge$Ozone, seq(0,1,by=0.05), na.rm = TRUE)
```

```
##          0%          5%         10%         15%         20%         25%         30%
## 0.02655433 0.03190818 0.03419029 0.03803052 0.03975537 0.04192559 0.04320513
##         35%         40%         45%         50%         55%         60%         65%
## 0.04381120 0.04520525 0.04618857 0.04716458 0.04827750 0.04978792 0.05165298
##         70%         75%         80%         85%         90%         95%        100%
## 0.05395594 0.05679963 0.05948454 0.06236471 0.06350576 0.06591477 0.07313200
```

```
quantile(dat_merge$PM2_5, seq(0,1,by=0.05), na.rm = TRUE)
```

```
##         0%         5%        10%        15%        20%        25%        30%        35%
##  1.875092   6.510532   7.563873   8.112168   8.387227   8.575343   8.705661   8.891836
##        40%        45%        50%        55%        60%        65%        70%        75%
##  9.155859   9.561726  10.119433  11.011972  11.448526  11.665665  11.809253  11.937845
##        80%        85%        90%        95%       100%
## 12.030068  12.106934  12.336561  13.328635  16.394748
```

```
quantile(dat_merge$DieselPM, seq(0,1,by=0.05), na.rm = TRUE)
```

```
##           0%           5%          10%          15%          20%          25%
##  0.00005240   0.01350995   0.02832625   0.04177016   0.05500557   0.06768251
##          30%          35%          40%          45%          50%          55%
```

```
##  0.08031747  0.09337265  0.10873597  0.12638421  0.14492875  0.16507467
##         60%         65%         70%         75%         80%         85%
##  0.18903759  0.21661019  0.25011306  0.28679367  0.33665935  0.40191467
##         90%         95%        100%
##  0.50290676  0.68290748 14.61122094
```

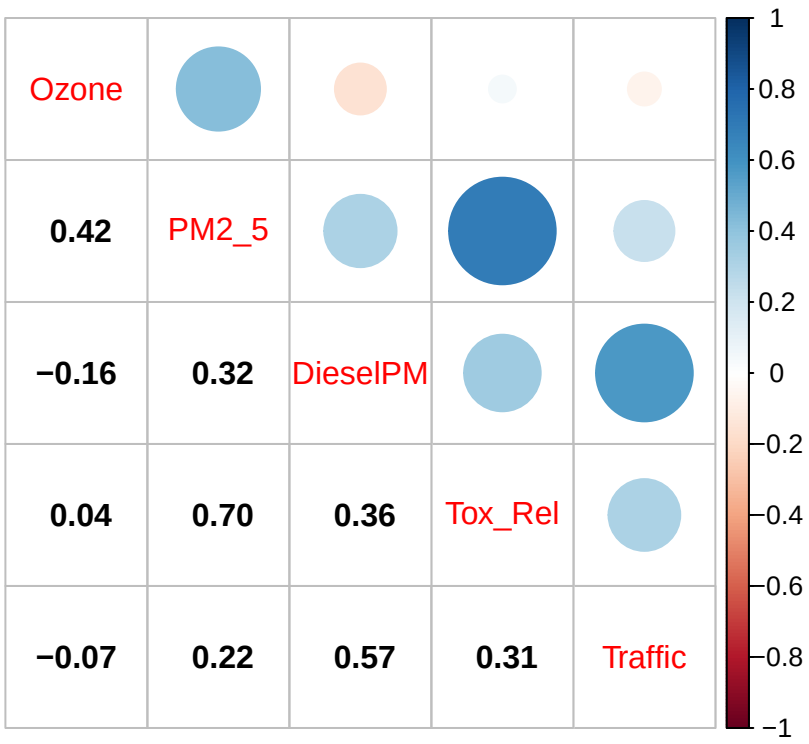
```
quantile(dat_merge$Tox_Rel, seq(0,1,by=0.05), na.rm = TRUE)
```

```
##           0%           5%          10%          15%          20%          25%
##     0.000000     1.182103    13.000241    31.260663    61.388579   111.523268
##          30%          35%          40%          45%          50%          55%
##   154.071003   200.221301   270.144875   351.967836   456.333669   588.785136
##          60%          65%          70%          75%          80%          85%
##   770.877734  1055.217123  1290.804244  1625.525848  2053.445483  3018.345659
##          90%          95%         100%
##  4144.665853  6716.583281 96985.629960
```

```
quantile(dat_merge$Traffic, seq(0,1,by=0.05), na.rm = TRUE)
```

```
##          0%          5%         10%         15%         20%         25%
##    20.74815   225.79496   334.15444   411.78363   488.36868   553.79709
##         30%         35%         40%         45%         50%         55%
##   623.04986   685.93118   746.01269   810.82985   881.03986   949.14531
##         60%         65%         70%         75%         80%         85%
##  1031.50732  1119.49896  1233.61149  1386.56621  1594.51271  1861.33719
##         90%         95%        100%
##  2219.10576  2835.37163 45752.00000
```

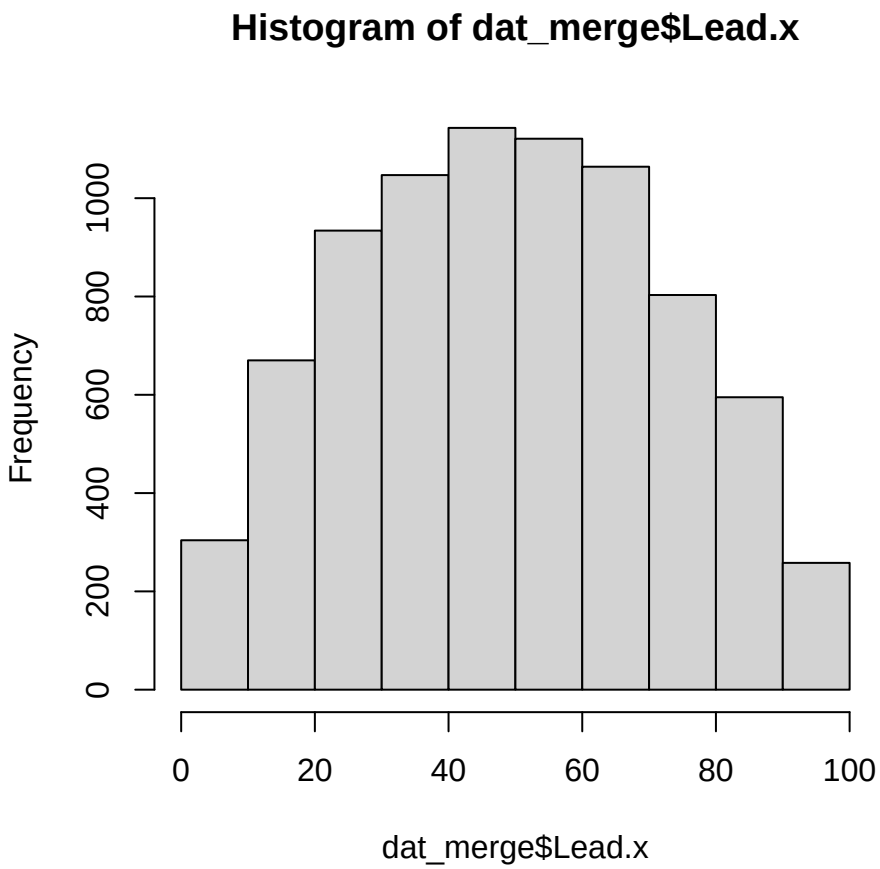
```
#correlation

AP<- dat_merge %>% select(Ozone, PM2_5, DieselPM, Tox_Rel, Traffic)
corAP<- cor(AP,use="complete.obs",method = "spearman")
corrplot.mixed(corAP, lower.col = "black")
```




```
#Lead in homes also a good distribution

summary(dat_merge$Lead.x)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    0.00   30.85   48.91   48.97   66.71   99.35      96
```

```
hist(dat_merge$Lead.x)
```




```
quantile(dat_merge$Lead.x, seq(0,1,by=0.05), na.rm = TRUE)
```

```
##       0%       5%      10%      15%      20%      25%      30%      35%
##  0.00000 11.60391 17.83103 22.38537 26.67219 30.84902 34.50882 38.51170
##      40%      45%      50%      55%      60%      65%      70%      75%
## 42.09256 45.45556 48.90805 52.43733 55.97328 59.42195 62.98696 66.70993
##      80%      85%      90%      95%     100%
## 70.66023 75.28880 80.80733 87.33777 99.35233
```

Maps

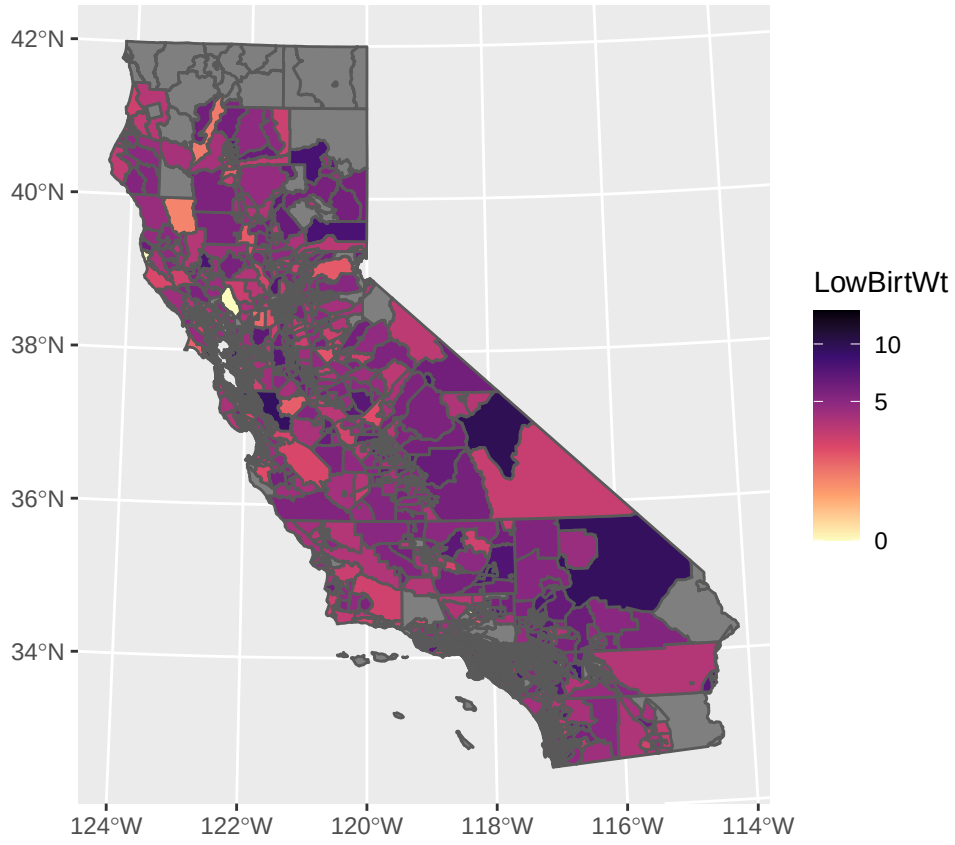
```
library(viridis)
```

```
## Loading required package: viridisLite
```

```
#ggplot(dat_merge_sf) +
#  geom_sf(aes(fill = Arsenic))+
#  scale_fill_gradientn(colours=rev(magma(6)), trans="sqrt")

#missing 227 tracts
summary(dat_merge$LowBirtWt)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   0.000   3.930   4.915   5.003   6.000  13.710     227
```

```
#CVD only missing 11 tracts
#summary(dat_merge$Cardiovas)


ggplot(dat_merge_sf) +
  geom_sf(aes(fill = LowBirtWt))+
  scale_fill_gradientn(colours=rev(magma(6)), trans="sqrt")
```

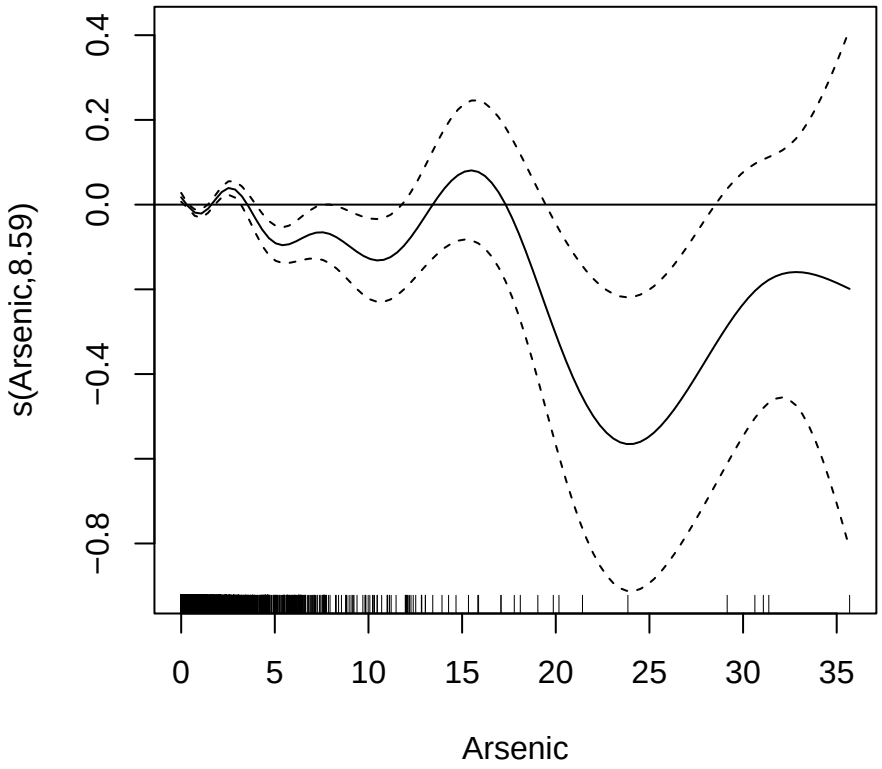


Quick models

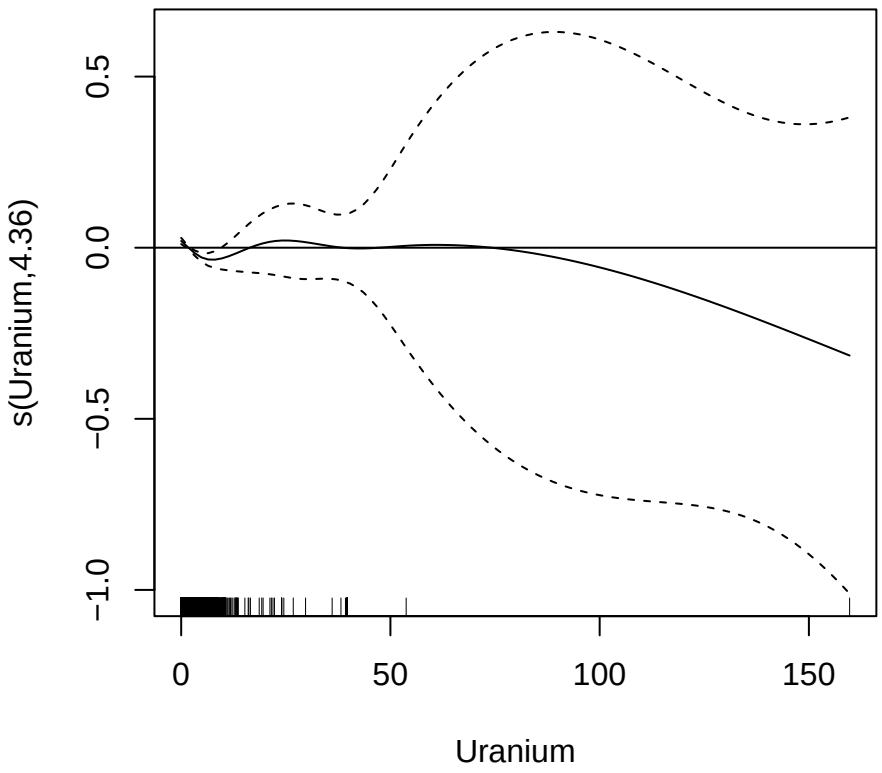
Low birth weight

```
#+Hispanic+White+AfricanAm+NativeAm+OtherMult (add in if want to adjust for race/eth)


#Water exposures

WaterExp<- c("Arsenic", "Uranium", "Cadmium", "Lead.y", "Nitrate")


WaterMod <- lapply(WaterExp, function(i) {
      # STRING INTERPOLATION WITH sprintf, THEN CONVERTED TO FORMULA OBJECT
      iformula <- as.formula(sprintf("LowBirtWt~s(%s)+EducatP+PovertyP+UnemplP+HousBurdP", i))
      gam(iformula, data=dat_merge, family = "quasipoisson")
```

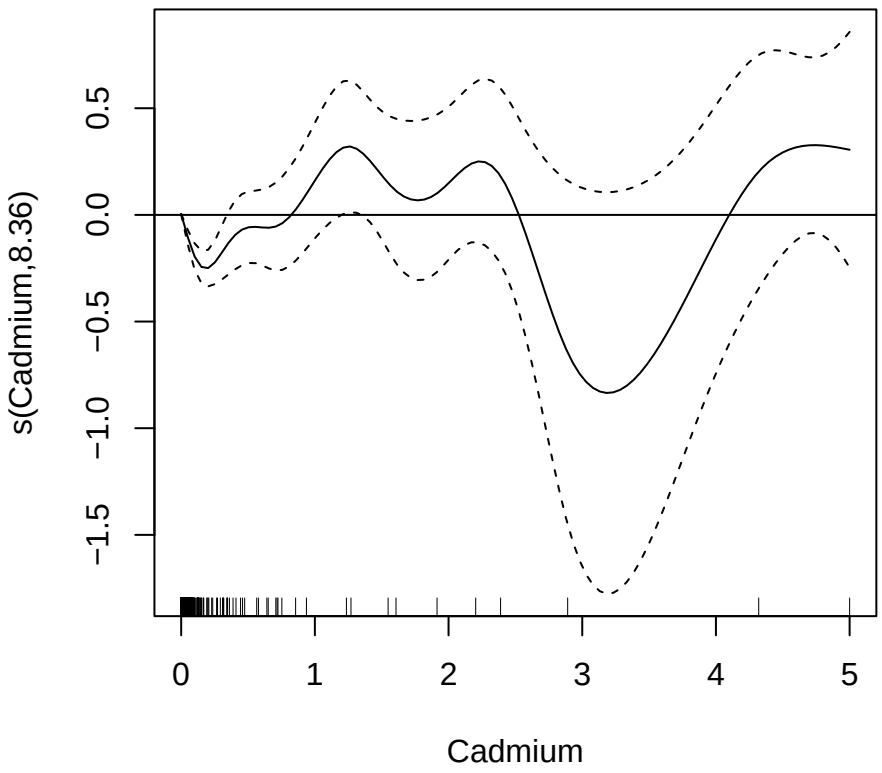
```
})


plot.gam(WaterMod[[1]])
abline(h=0)
```



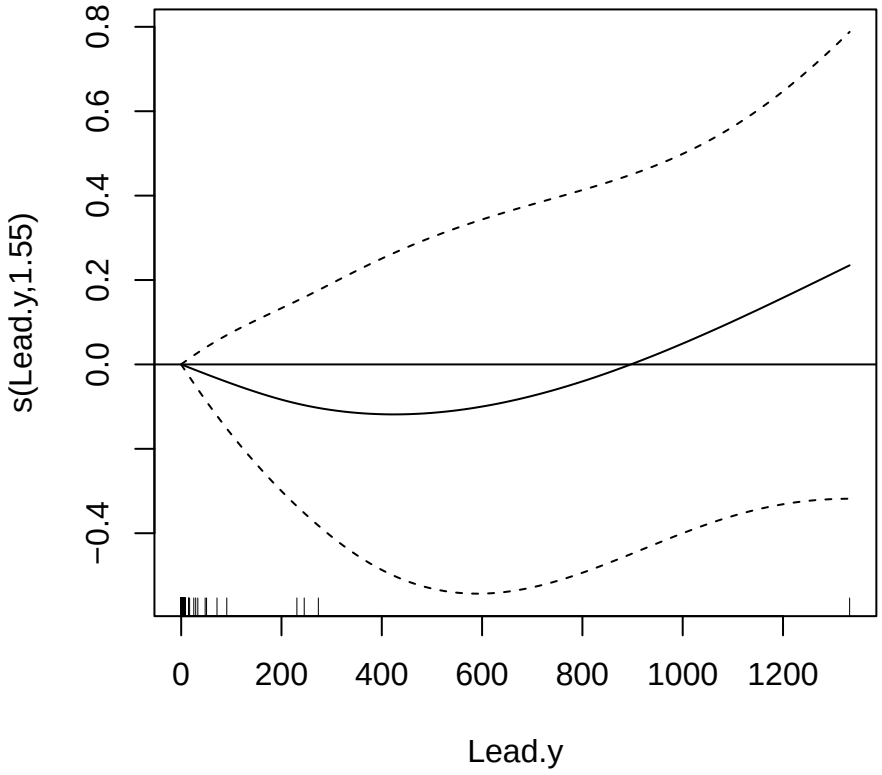

```
plot.gam(WaterMod[[2]])
abline(h=0)
```



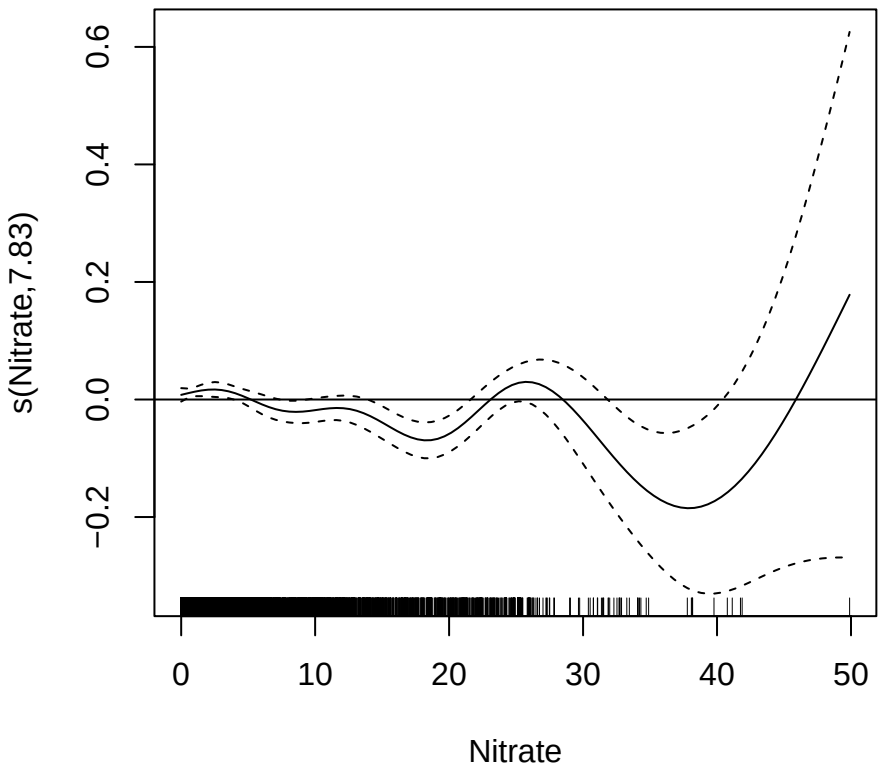

```
plot.gam(WaterMod[[3]])
abline(h=0)
```



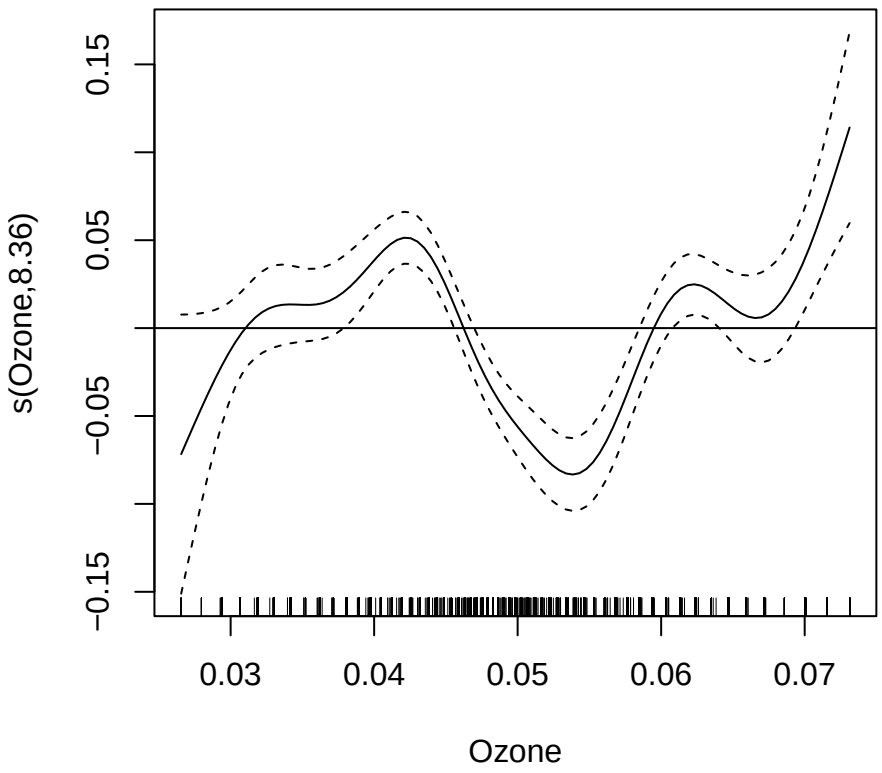

```
plot.gam(WaterMod[[4]])
abline(h=0)
```



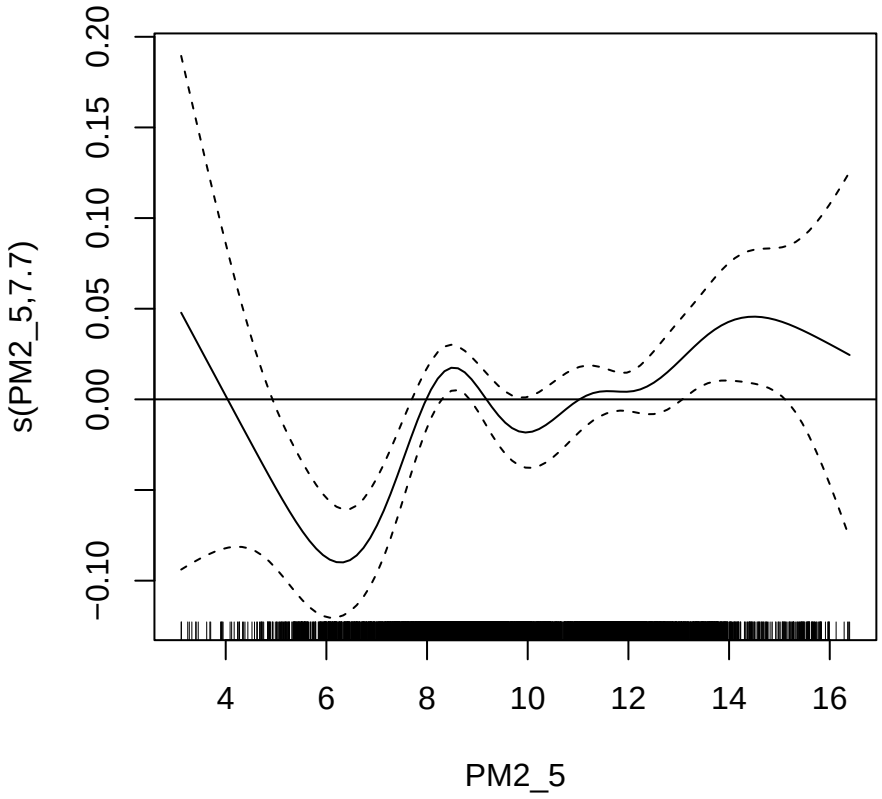

```
plot.gam(WaterMod[[5]])
abline(h=0)
```



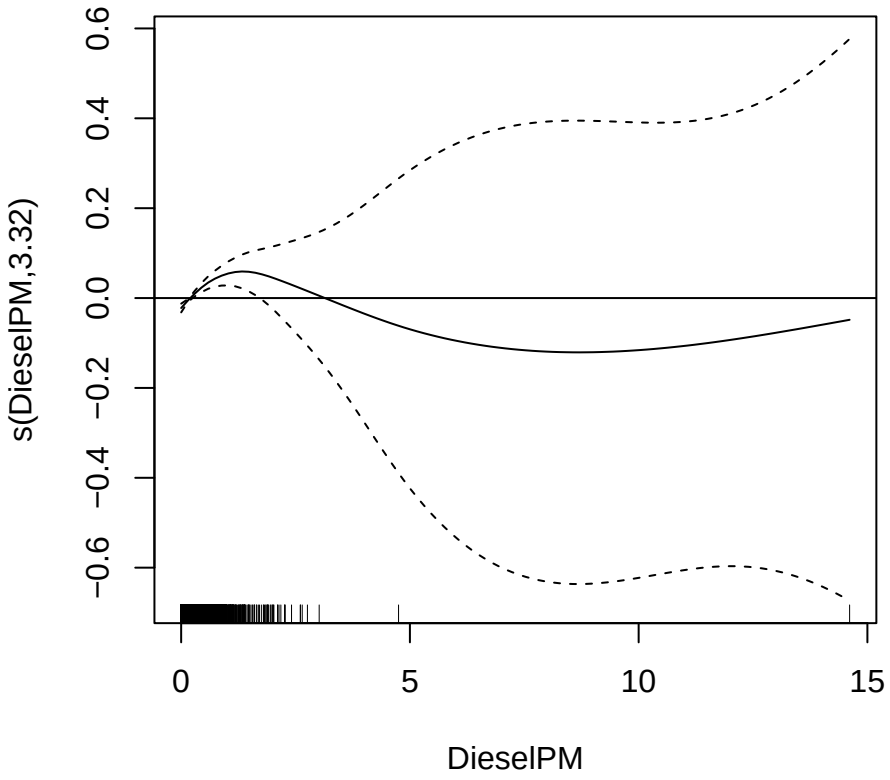

```
#Air pollution (and housing lead)

APExp<- c("Ozone", "PM2_5", "DieselPM", "Tox_Rel", "Traffic", "Lead.x")


APMod <- lapply(APExp, function(i) {
      # STRING INTERPOLATION WITH sprintf, THEN CONVERTED TO FORMULA OBJECT
      iformula <- as.formula(sprintf("LowBirtWt~s(%s)+EducatP+PovertyP+UnemplP+HousBurdP", i))
      gam(iformula, data=dat_merge, family = "quasipoisson")


})


plot.gam(APMod[[1]])
abline(h=0)
```



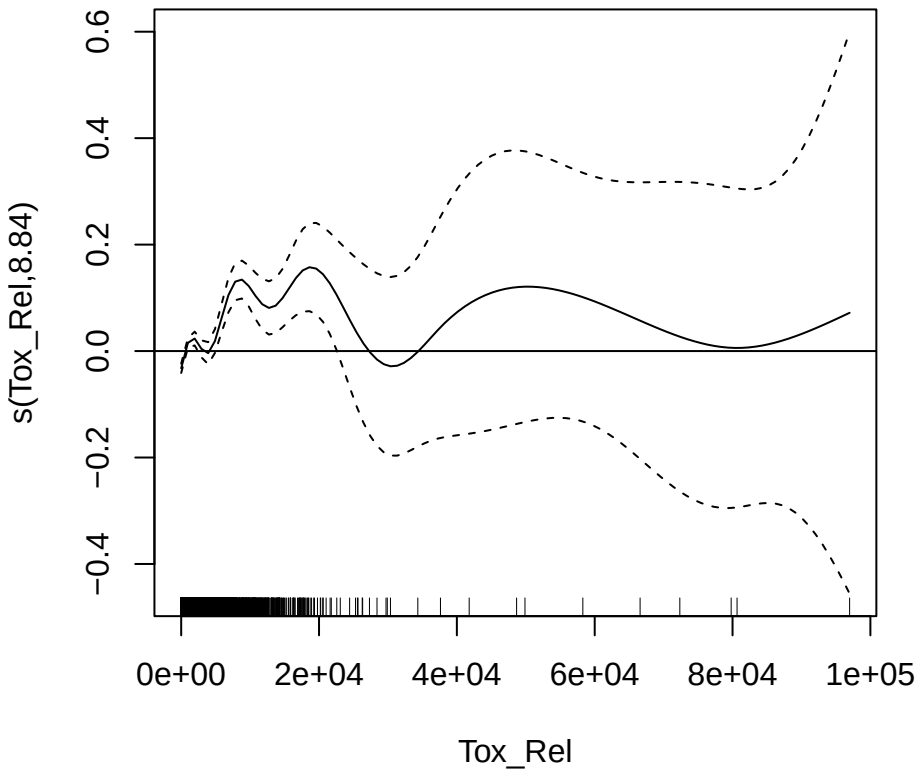

```
plot.gam(APMod[[2]])
abline(h=0)
```



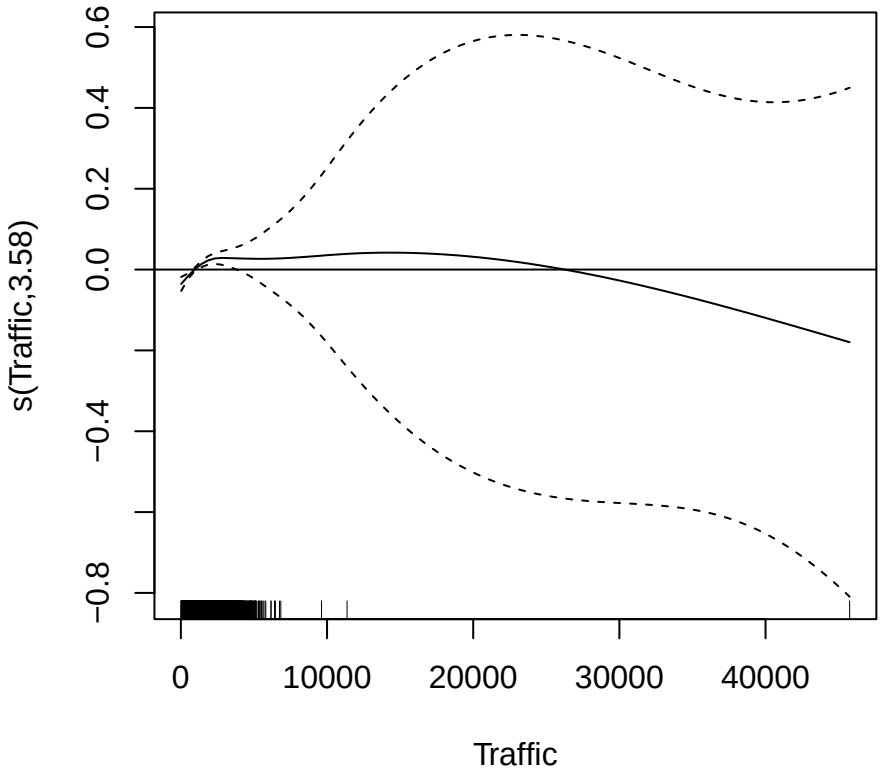

```
plot.gam(APMod[[3]])
abline(h=0)
```



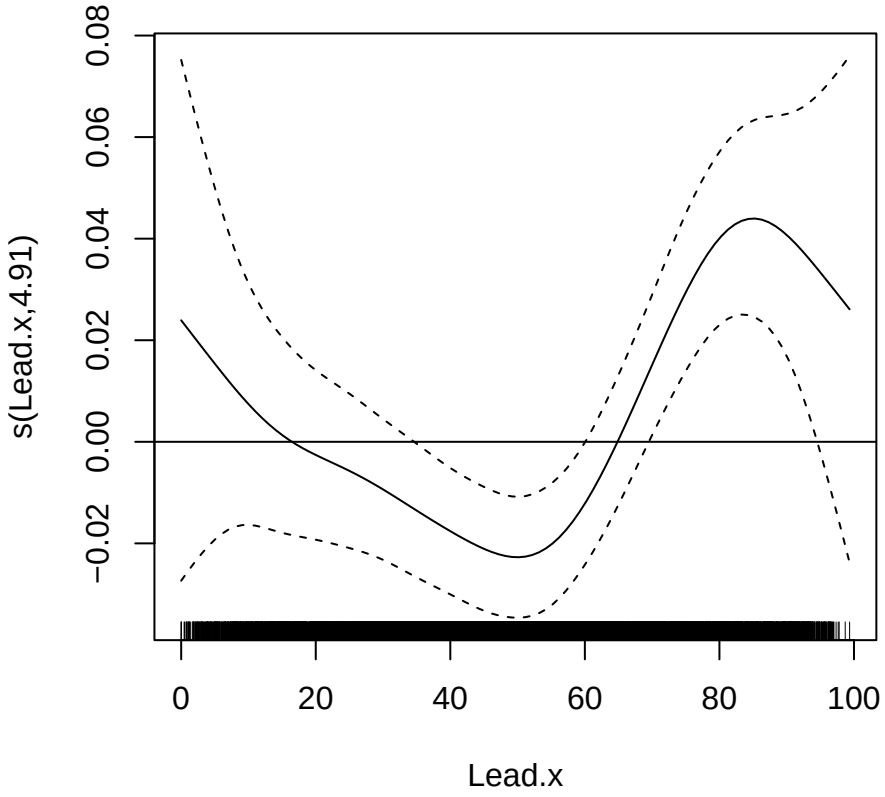

```
plot.gam(APMod[[4]])
abline(h=0)
```




```
plot.gam(APMod[[5]])
abline(h=0)
```

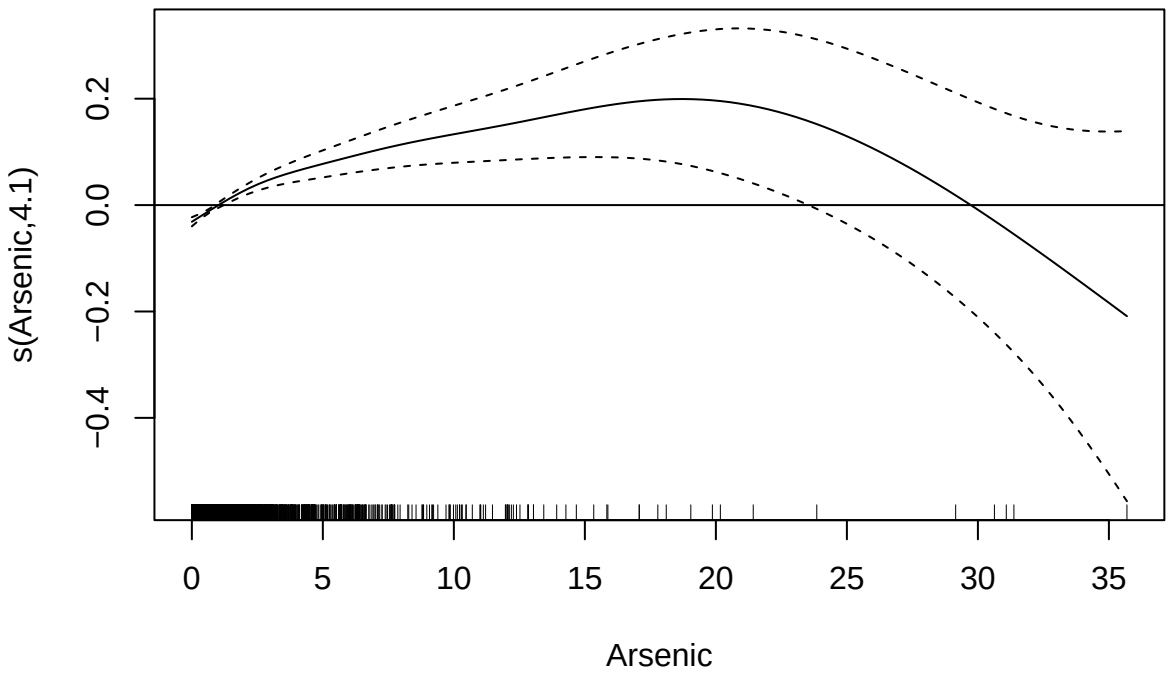



```
plot.gam(APMod[[6]])
abline(h=0)
```

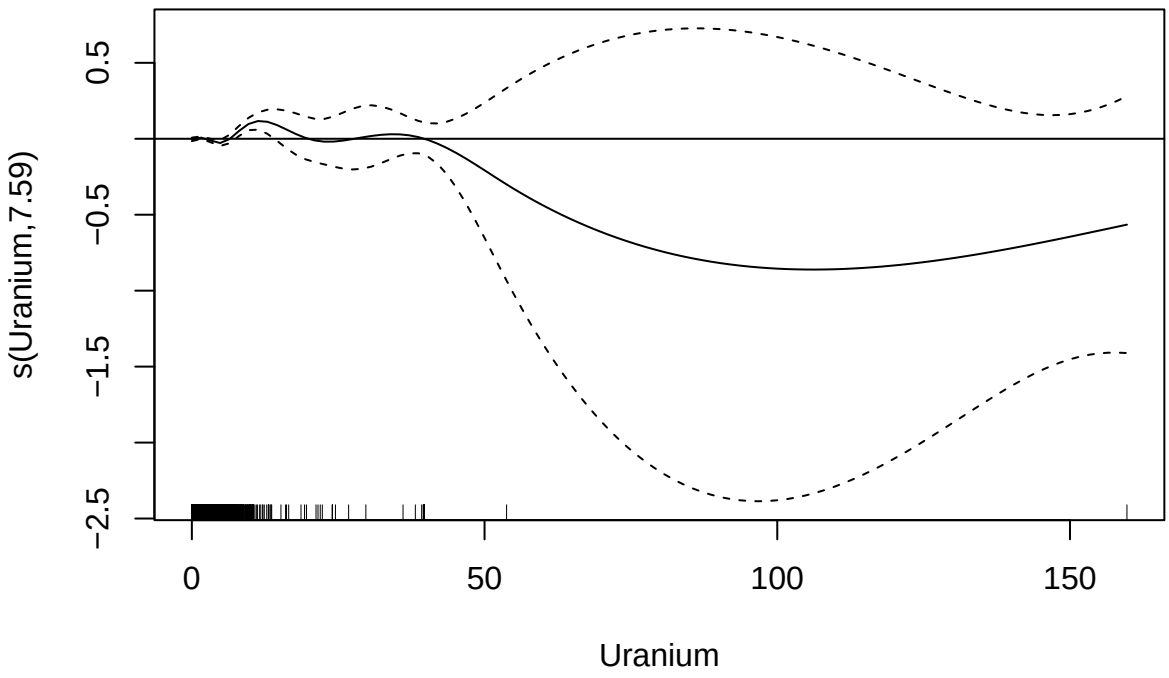


CVD

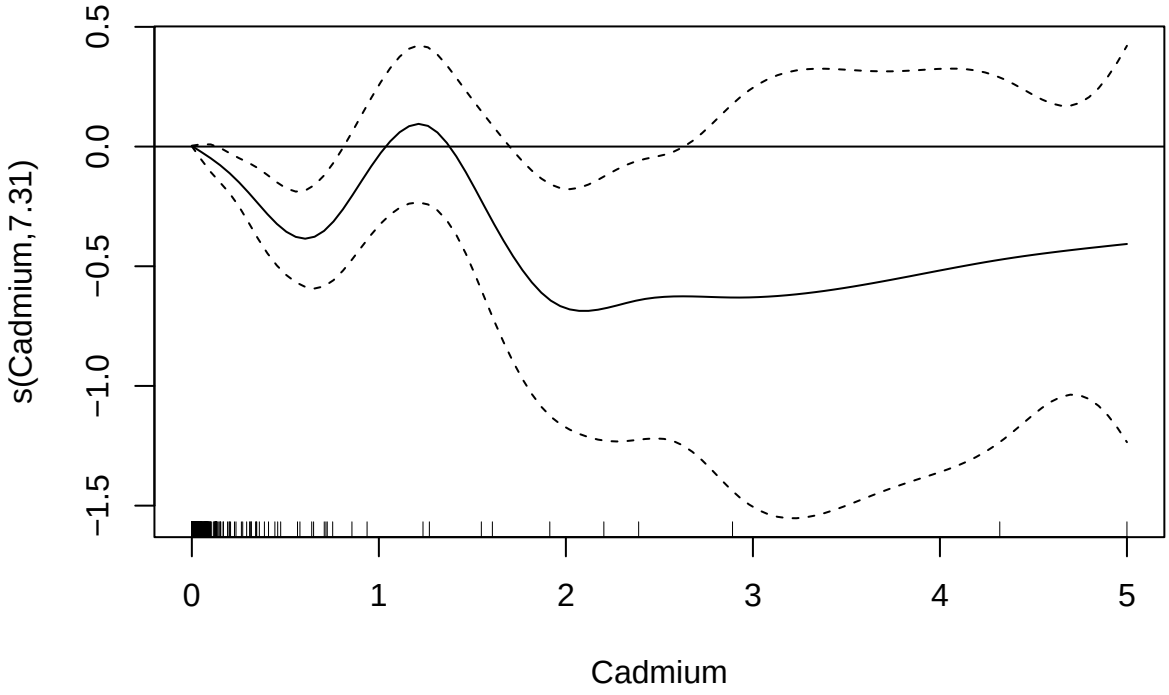
```
#+Hispanic+White+AfricanAm+NativeAm+OtherMult (add in if want to adjust for race/eth)


#Water exposures

WaterExp<- c("Arsenic", "Uranium", "Cadmium", "Lead.y", "Nitrate")


WaterMod <- lapply(WaterExp, function(i) {
      # STRING INTERPOLATION WITH sprintf, THEN CONVERTED TO FORMULA OBJECT
      iformula <- as.formula(sprintf("Cardiovas~s(%s)+EducatP+PovertyP+UnemplP+HousBurdP", i))
      gam(iformula, data=dat_merge, family = "quasipoisson")


})


plot.gam(WaterMod[[1]])
abline(h=0)
```



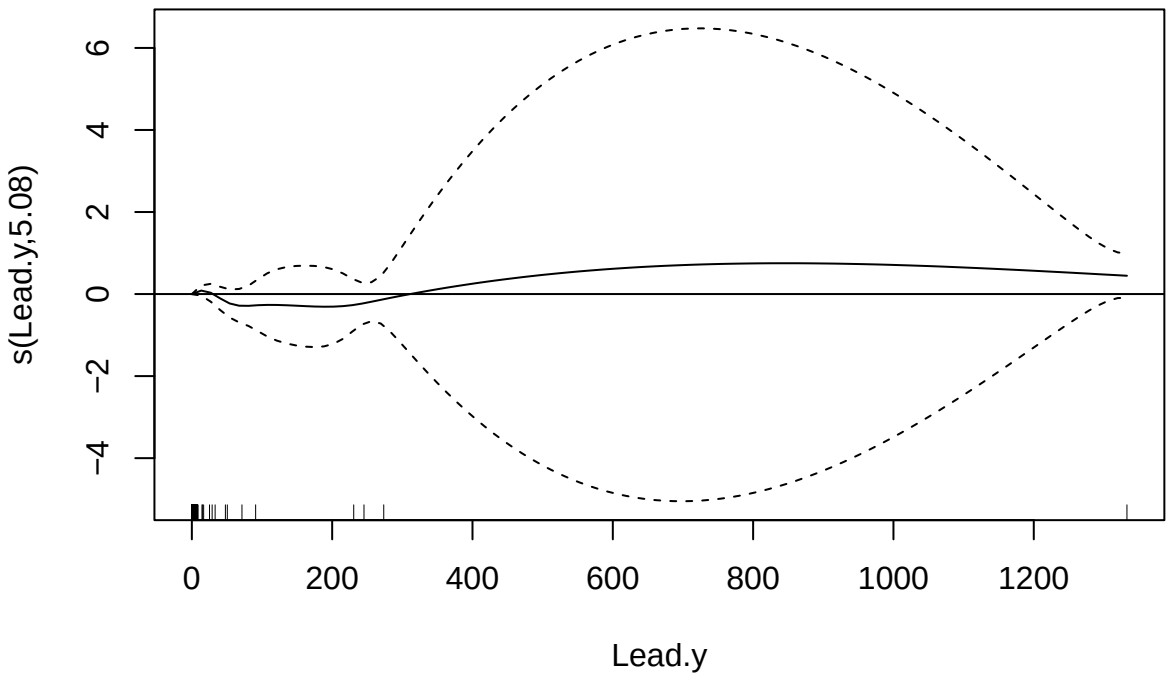

```
plot.gam(WaterMod[[2]])
abline(h=0)
```



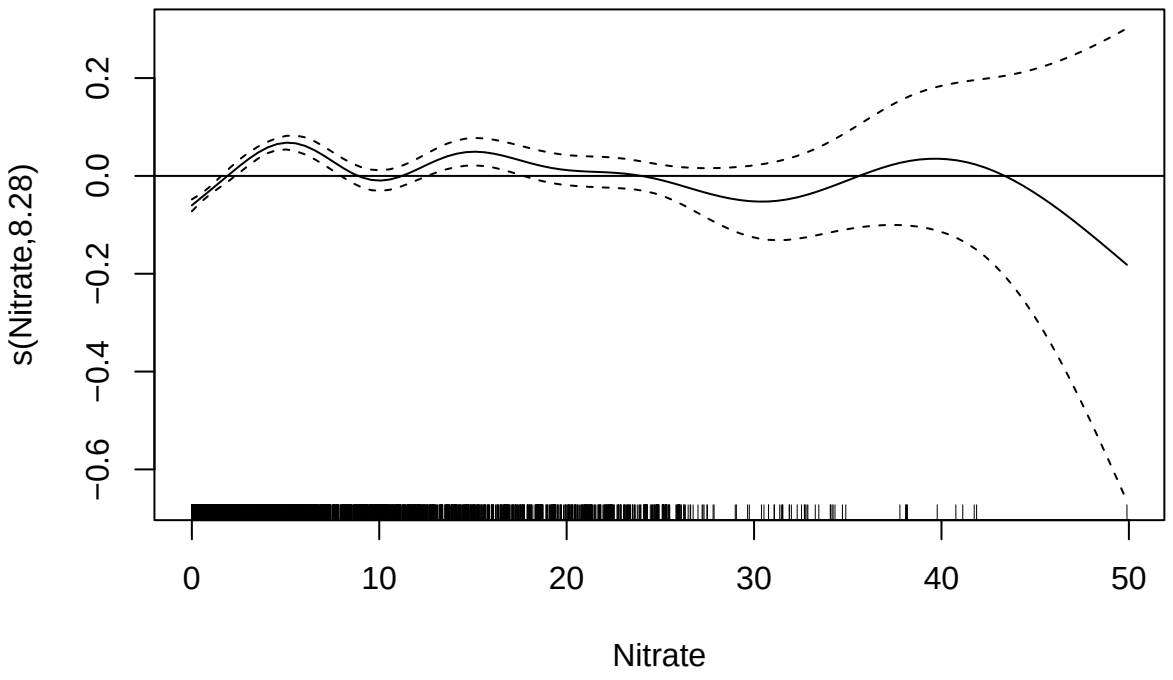

```
plot.gam(WaterMod[[3]])
abline(h=0)
```



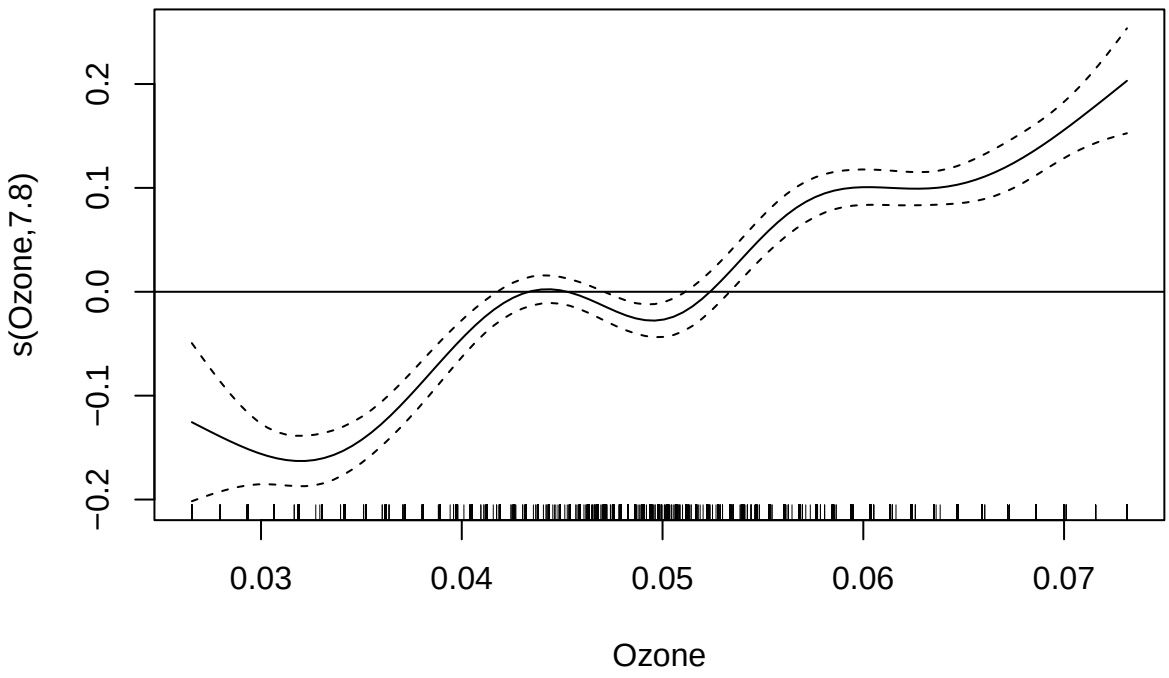

```
plot.gam(WaterMod[[4]])
abline(h=0)
```



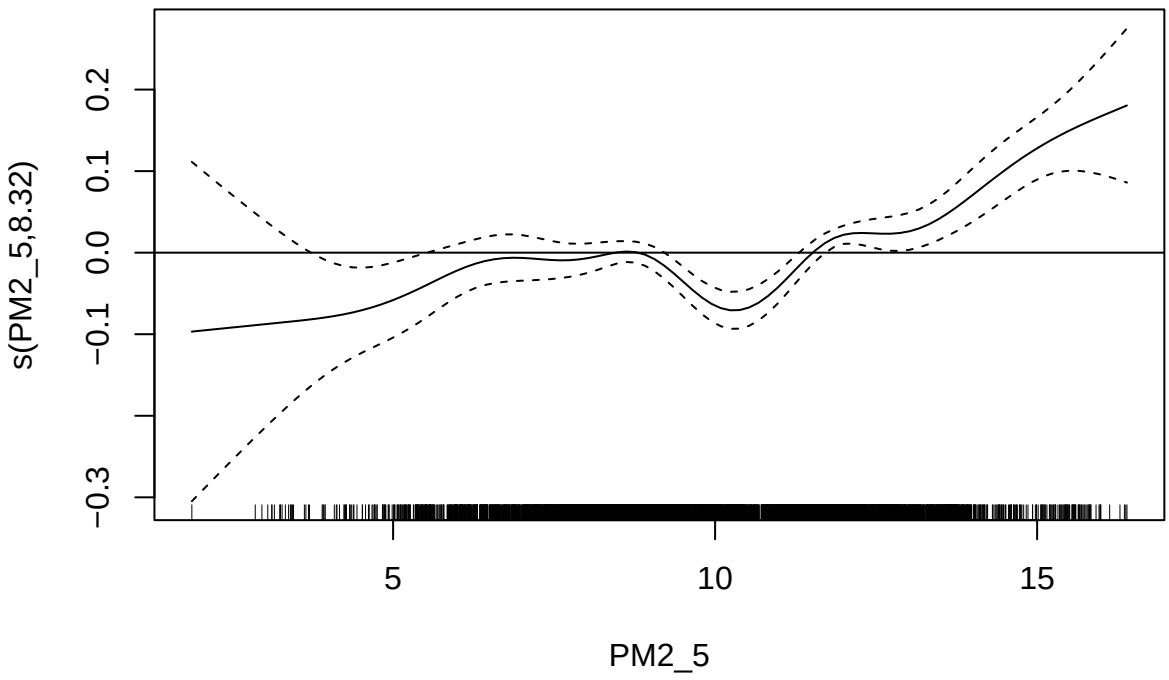

```
plot.gam(WaterMod[[5]])
abline(h=0)
```



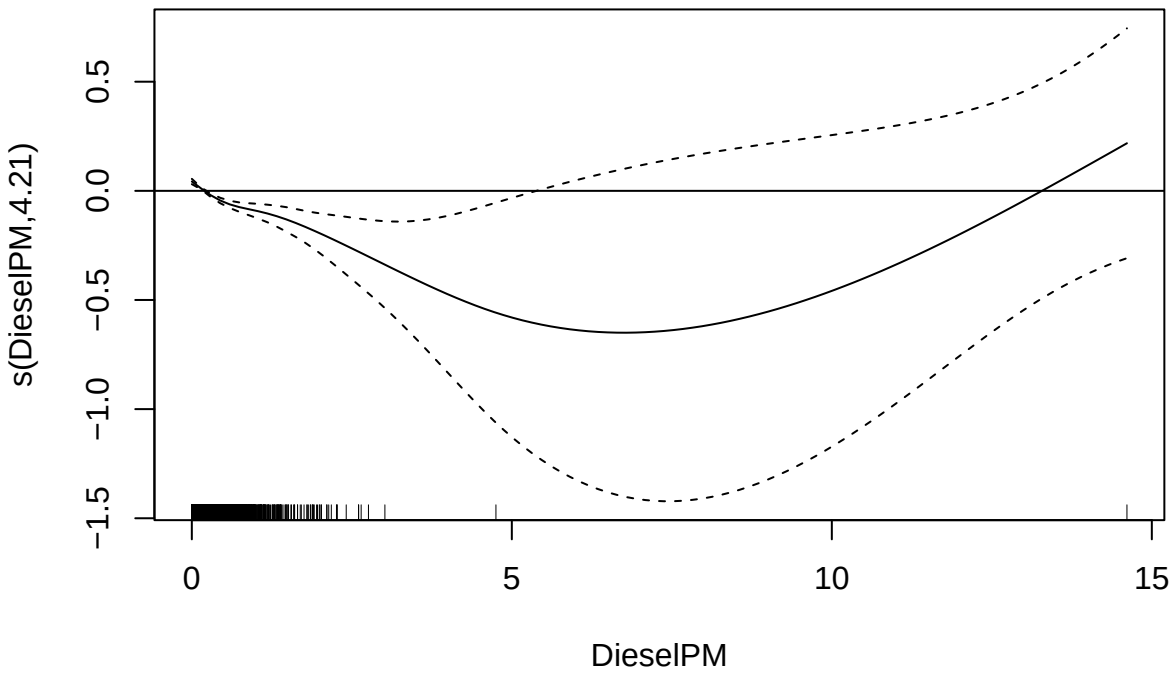

```
#Air pollution (and housing lead)

APExp<- c("Ozone", "PM2_5", "DieselPM", "Tox_Rel", "Traffic", "Lead.x")


APMod <- lapply(APExp, function(i) {
      # STRING INTERPOLATION WITH sprintf, THEN CONVERTED TO FORMULA OBJECT
      iformula <- as.formula(sprintf("Cardiovas~s(%s)+EducatP+PovertyP+UnemplP+HousBurdP", i))
      gam(iformula, data=dat_merge,  family = "quasipoisson")


})


plot.gam(APMod[[1]])
abline(h=0)
```



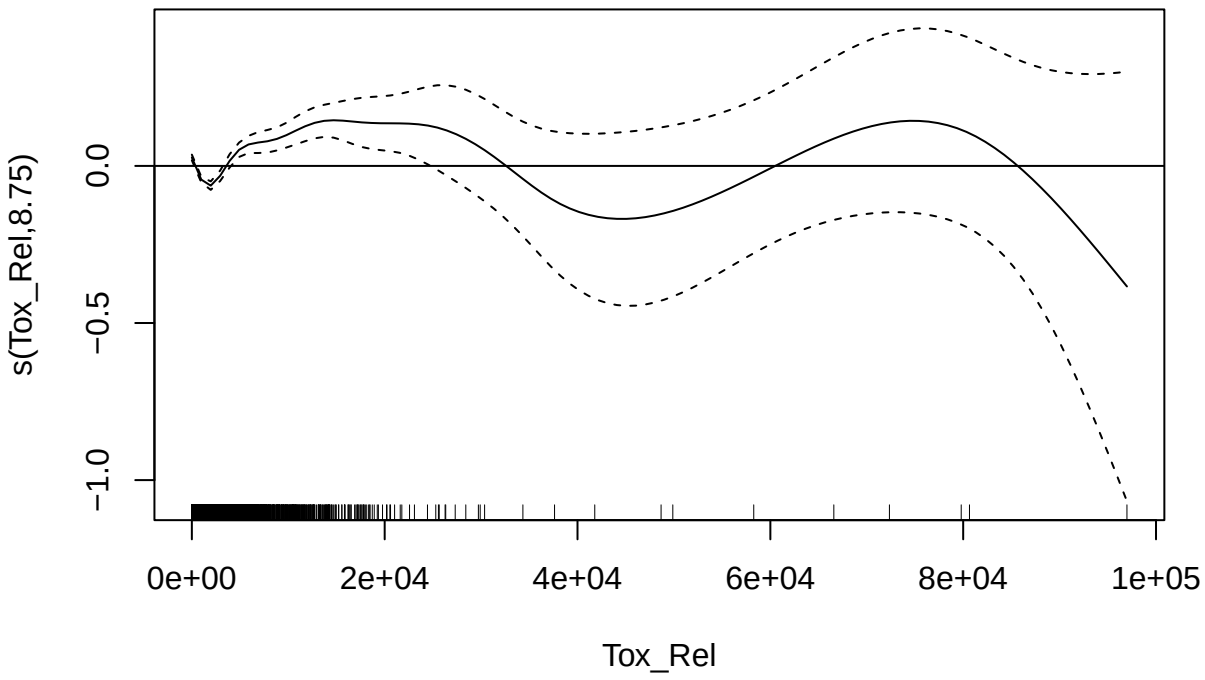

```
plot.gam(APMod[[2]])
abline(h=0)
```



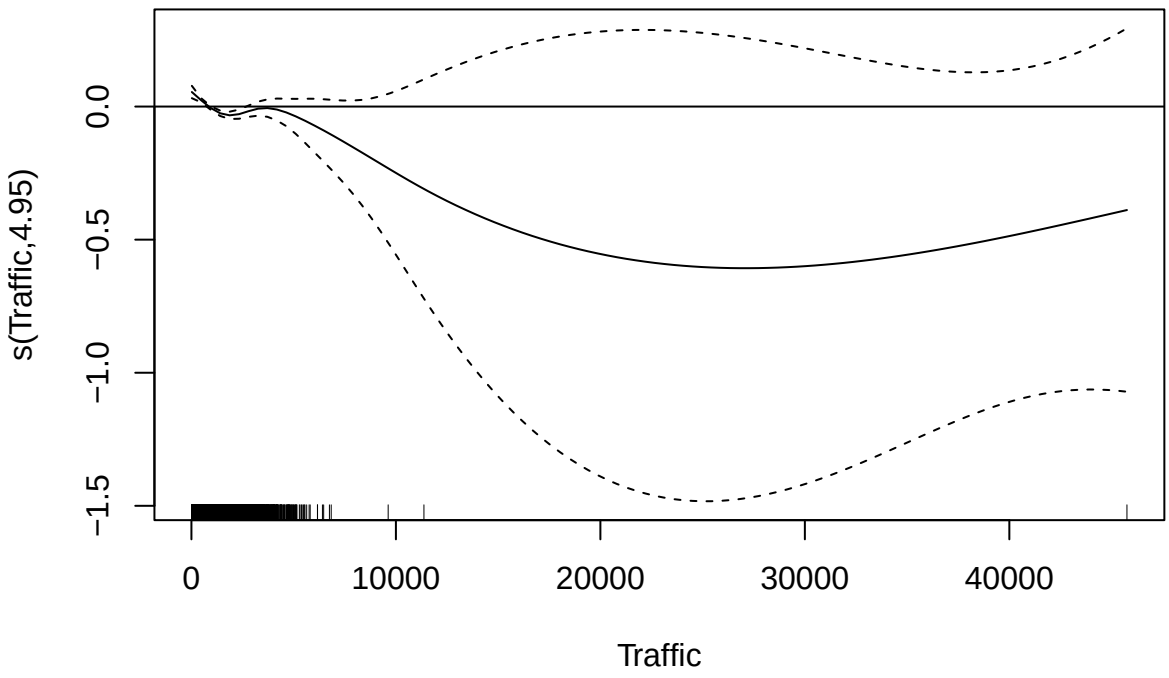

```
plot.gam(APMod[[3]])
abline(h=0)
```



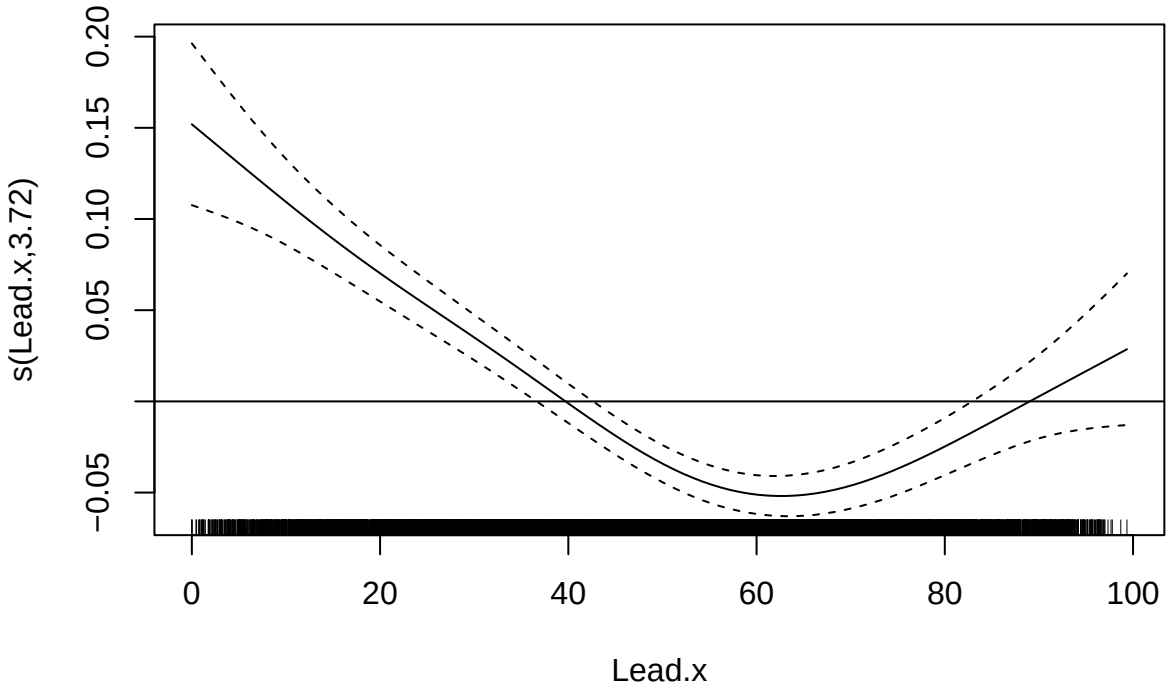

```
plot.gam(APMod[[4]])
abline(h=0)
```




```
plot.gam(APMod[[5]])
abline(h=0)
```




```
plot.gam(APMod[[6]])
abline(h=0)
```


s(Lead.x,3.72)
0.20
0.15
0.10
0.05
−0.05
0
20
40
60
80
100
Lead.x