# Week 3 Report

correlation matrix

leaps →reg subset, mallow cp,

- take out room_f, pump_f, sw_comp_purge - done
- put inverse distance for flare distance, 0 for when count is 0 -done
- subset model for only with flares -nope
- put smoothing for time -done

- changed na.rm to TRUE for trailer mean processing
- made datetime a Date variable
  - str caused datetime to be considered a categorical variable before
- Error in leaps.exhaustive(a, really.big) :
  Exhaustive search will be S L O W, must specify really.big=T
  - fixed

## Regsubsets results

Daily

1. using Adj_R^2: 21 variables
   a. Intercept, o3_mean, **temp_f_mean, pressure_altcorr_mean, wsp_mean**, wdr_mean, rain_mean, **co2_ppm_mean, ch4_mean, h2o_sync_mean, ethene_mean**, i.butane_mean, n.butane_mean, **acetylene_mean, cyclopentane_mean, i.pentane_mean, n.pentane_mean**, n.hexane_mean, benzene_mean, ethyl.benzene_mean, m.p.xylene_mean, o.xylene_mean
2. Using Cp, 14 variables

a. Intercept, **temp_f_mean, pressure_altcorr_mean, wsp_mean**, wdr_mean, rain_mean, **co2_ppm_mean, ch4_mean, h2o_sync_mean, ethene_mean, acetylene_mean, cyclopentane_mean, i.pentane_mean, n.pentane_mean**, benzene_mean

3. Using Bic, 11 variables
    a. Intercept, **temp_f_mean, pressure_altcorr_mean, wsp_mean**, **co2_ppm_mean, ch4_mean, h2o_sync_mean, ethene_mean**, **acetylene_mean, cyclopentane_mean, i.pentane_mean, n.pentane_mean**

Hourly

1. using Adj_R^2: 25 variables
    a. Intercept, datetime, co, no2, nox, temp_f, pressure_altcorr, wsp, wdr, relh, co2_ppm, h2o_sync, ethane, propene, X1_3.butadiene, i.butane, n.butane, acetylene, cyclopentane, n.pentane, n.hexane, benzene, ethyl.benzene, m.p.xylene, o.xylene, hour
2. Using Cp, 20 variables
    a. Intercept, datetime, no2, nox, temp_f, pressure_altcorr, wsp, wdr, relh, co2_ppm, ethane, propene, X1_3.butadiene, i.butane, n.butane, acetylene, cyclopentane, n.hexane, n.heptane, benzene, hour
3. Using bic, 16 variables
    a. Intercept, datetime, no2, temp_f, pressure_altcorr, wsp, wdr, relh, co2_ppm, ethane, propene, X1_3.butadiene, i.butane, n.butane, acetylene, benzene, hour

Variables shared across: temp_f, pressure, wsp, co2_ppm,

Fitting Gam on BIC selected variables:

1. Daily went from `R-sq.(adj) =  0.721   Deviance explained = 80.2%` to `R-sq.(adj) =  0.541 Deviance explained = 57.7%`
    a. 8 observations were added tho

b. adding s(datetime) increases to 70% deviance explained

2. Hourly went from `R-sq.(adj) =  0.631   Deviance explained = 63.8%` to `R-sq.(adj) = 0.608   Deviance explained = 61.2%`

    a. 2770 to 2977 obs