Week 11

Gillian McGinnis

11/20/2020

Cal Curve

```
lab_calcurve_x <- expression(paste("Concentration (", mu, "g/mL)"))
lab_calcurve_y <- ("Area of major chromatogram peak")

cal_curve <- ggplot(standard, aes(x = conc, y = area))+
  stat_smooth(method = "lm", se = FALSE, color = "red")+
  geom_point(shape = 1)+
  theme_few()+
  labs(x = lab_calcurve_x, y = lab_calcurve_y)

ggsave("cal_curve.png", plot = cal_curve, path = "figures/")
```

```
## Saving 6.5 x 4.5 in image
## `geom_smooth()` using formula 'y ~ x'
```

```
curve_results <- summary(lm(area ~ conc, data = standard))

model <- lm(area ~ conc, data = standard)

slope <- model$coefficients[2]
intercept <- model$coefficients[1]
slope_std <- summary(model)$coefficients[2,2]
intercept_std <- summary(model)$coefficients[1,2]
equation <- tibble(slope, slope_std, intercept, intercept_std)
```

Calculating levoglucosan concentrations in each sample with propogated error

```
all_conc <- lcms %>%
  mutate(conc = (area-intercept)/slope) %>%
    mutate(#rsd = (100*sd_conc)/mean_conc,
         #e_yb = sqrt(rsd)^2 + intercept_std,
         e_yb = intercept_std, #pretending that there is no error in y b/c minions took very few sample
         yb = conc-intercept,
         e_x = conc*sqrt((e_yb/yb)^2 + (slope_std/slope)^2)) %>%
  rename(conc_error = e_x) %>%
  select(loc_date, conc, conc_error)
```

Calculating airborne levoglucosan concentrations in each sample with propogated error

```
all_airborne <- all_conc %>%
  mutate(air_conc = conc*2*70/24,
         air_error = air_conc*sqrt(
           (conc_error/conc)^2+
             ((sqrt((0.1^2)+(0.1)^2))/(1+1))^2 +
             (0.1/70)^2+
             (0.0007/24)^2
         ))
```

Stat analysis: averages with standard deviation, and 95CI

```
ci95_alt <- all_airborne %>%
  group_by(loc_date) %>%
  summarise(mean = mean(air_conc),
         sd = sd(air_conc),
         n = n()) %>%
  mutate(se = qnorm(0.975)*sd/sqrt(n),
         lower_ci = mean - se,
         upper_ci = mean + se) %>%
  mutate(loc = case_when(
    str_detect(loc_date, "E") ~ "East",
    str_detect(loc_date, "W") ~ "West"
  )) %>%
  mutate(date = case_when(
    str_detect(loc_date, "1750") ~ "1750",
    str_detect(loc_date, "1950") ~ "1950",
    str_detect(loc_date, "2020") ~ "2020"
  ))
```

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

```
lab_conc_airborne <- expression(paste("Airborne concentration (", mu, "g/m"^3*")"))

ci95 <- ggplot(ci95_alt, aes(x = date, y = mean, color = loc))+
  geom_point(position = position_dodge(width=0.9))+
  geom_errorbar(ymin = ci95_alt$lower_ci, ymax = ci95_alt$upper_ci, position = position_dodge(width=0.9)
  expand_limits(ymin = 15, ymax = 60)+
  theme_few()+
  labs(x = "Date", y = lab_conc_airborne, color = "Location")

ggsave("ci95.png", plot = ci95, path = "figures/")
```

```
## Saving 6.5 x 4.5 in image
```

Stat test: Grubbs test for outliers

```
grubbs_df <- all_airborne %>%
  filter(loc_date == "W2020")
```

```r
grubbs.test(grubbs_df$conc)
```

```
##
##  Grubbs test for one outlier
##
## data:  grubbs_df$conc
## G = 1.1536480, U = 0.0018223, p-value = 0.04078
## alternative hypothesis: highest value 10.2013934404882 is an outlier
```

```r
#p val = 0.04078, that's an outlier, fellers

#source: https://stackoverflow.com/questions/45486159/several-grubbs-tests-simultaneously-in-r
grubbs_all <- all_airborne %>%
  group_by(loc_date) %>%
  nest() %>%
  mutate(n = map_dbl(data, ~ nrow(.x)), # number of entries
         G = map(data, ~ grubbs.test(.x$conc)$statistic[[1]]), # G statistic
         U = map(data, ~ grubbs.test(.x$conc)$statistic[[2]]), # U statistic
         grubbs = map(data, ~ grubbs.test(.x$conc)$alternative), # Alternative hypotesis
         p_grubbs = map_dbl(data, ~ grubbs.test(.x$conc)$p.value)) %>% # p-value
  # Let's make the output more fancy
  mutate(G = signif(unlist(G), 3),
         U = signif(unlist(U), 3),
         grubbs = unlist(grubbs),
         p_grubbs = signif(p_grubbs, 3)) %>%
  select(-data) %>% # remove temporary column
  arrange(p_grubbs) %>%
  mutate(label = case_when(
    p_grubbs < 0.05 ~ "p < 0.05", # Reject null hypothesiss; diff is significant
    p_grubbs >= 0.05 ~ "Non-Sig" # Fail to reject null hyp; diff is not significant
  ))
grubbs_all
```

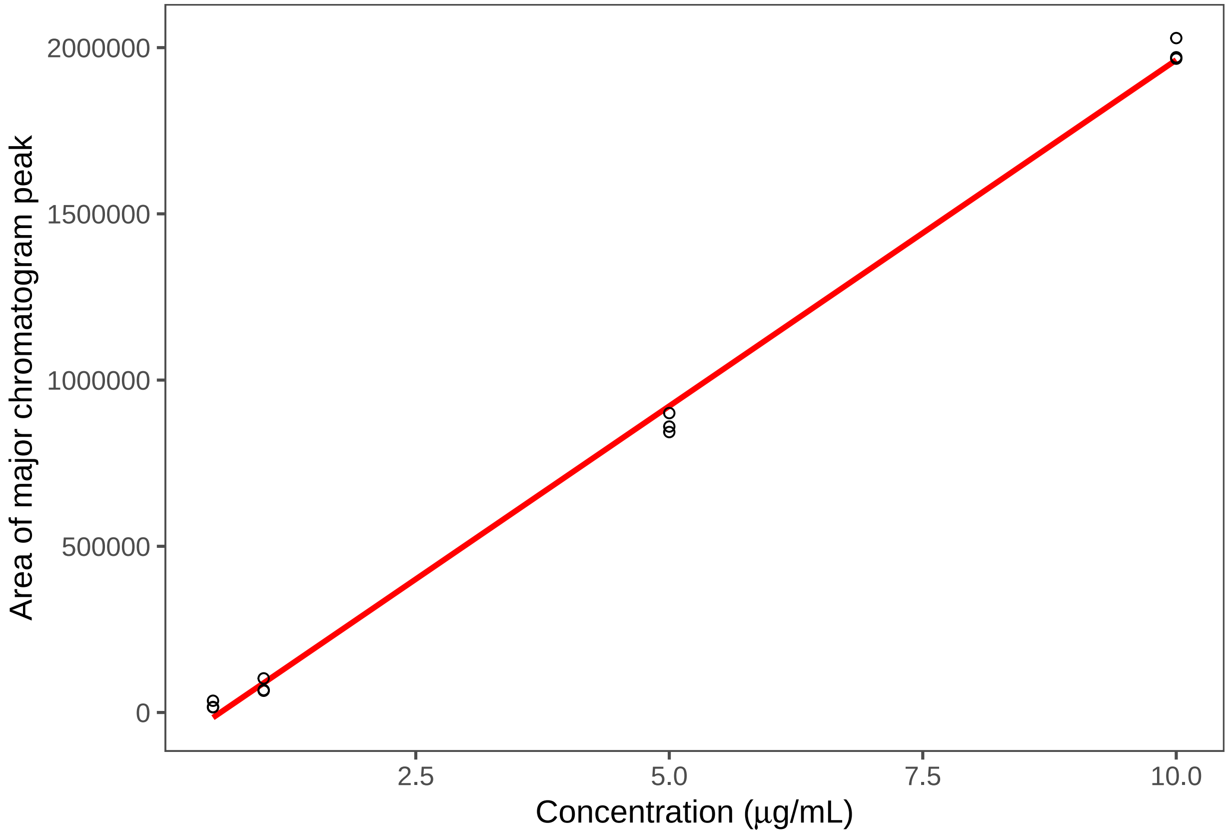
```
## # A tibble: 6 x 7
## # Groups:   loc_date [6]
##   loc_date     n     G       U grubbs                             p_grubbs label
##   <fct>    <dbl> <dbl>   <dbl> <chr>                                 <dbl> <chr>
## 1 W2020        3  1.15 0.00182 highest value 10.2013934404882 ~   0.0408 p < 0.~
## 2 E1750        3  1.15 0.00704 highest value 3.49098369328885 ~   0.0802 Non-Sig
## 3 W1950        3  1.13 0.0478  lowest value 3.39463644992568 i~   0.21   Non-Sig
## 4 E2020        3  1.07 0.147   highest value 4.84408046065269 ~   0.376  Non-Sig
## 5 W1750        3  1.06 0.157   highest value 4.33218771032802 ~   0.389  Non-Sig
## 6 E1950        3  1.01 0.231   lowest value 3.00349468552656 i~   0.479  Non-Sig
```

```r
#only W2020 has a significant outlier!
```

SUMMARY OF RESULTS

```r
#calibration curve; 2
cal_curve
```

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




```r
#concentration; 3
# units: µg/mL
all_conc %>%
  select(loc_date, conc, conc_error)
```

```
##    loc_date      conc conc_error
## 1     E1750  3.456683  0.5596459
## 2     E1750  3.490984  0.5651991
## 3     E1750  3.453178  0.5590784
## 4     E1950  3.075212  0.4978864
## 5     E1950  3.141744  0.5086578
## 6     E1950  3.003495  0.4862754
## 7     E2020  4.739858  0.7673871
## 8     E2020  4.780214  0.7739205
## 9     E2020  4.844080  0.7842602
## 10    W1750  4.102955  0.6642755
## 11    W1750  4.194380  0.6790768
## 12    W1750  4.332188  0.7013873
## 13    W1950  3.646516  0.5903794
## 14    W1950  3.721345  0.6024941
## 15    W1950  3.394636  0.5496007
## 16    W2020  5.605667  0.9075559
## 17    W2020  5.373193  0.8699201
## 18    W2020 10.201393  1.6515398
```

```r
#airborne conc; 4
# math: 2mL dilution factor (1mL water + 1mL ethanol), 70 b/c cut 1in^2, 24 hr
# units: µg/m^3
```

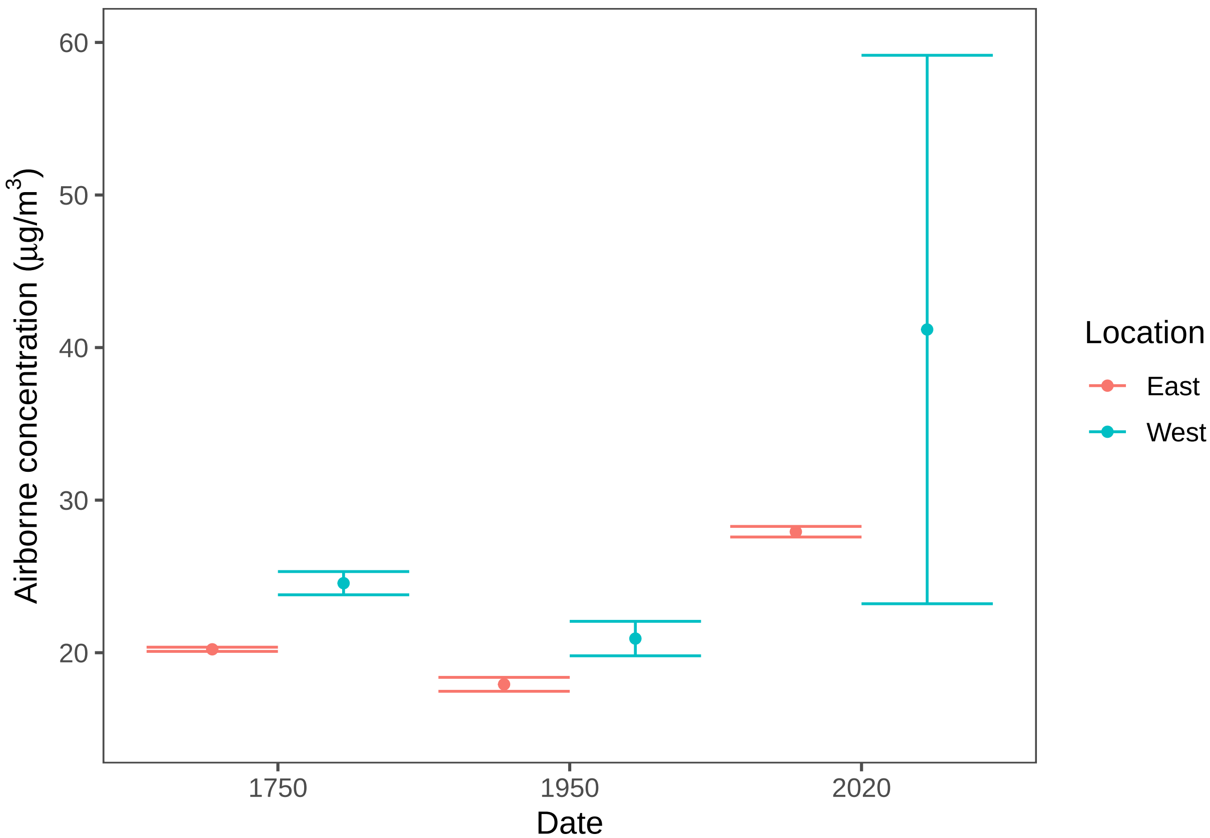
```r
all_airborne %>%
  select(loc_date, air_conc, air_error)
```

```
##    loc_date air_conc air_error
## 1     E1750 20.16398  3.562497
## 2     E1750 20.36407  3.597847
## 3     E1750 20.14354  3.558884
## 4     E1950 17.93874  3.169357
## 5     E1950 18.32684  3.237924
## 6     E1950 17.52039  3.095445
## 7     E2020 27.64917  4.884908
## 8     E2020 27.88458  4.926497
## 9     E2020 28.25714  4.992316
## 10    W1750 23.93390  4.228533
## 11    W1750 24.46722  4.322753
## 12    W1750 25.27109  4.464775
## 13    W1950 21.27134  3.758135
## 14    W1950 21.70785  3.835253
## 15    W1950 19.80205  3.498552
## 16    W2020 32.69973  5.777178
## 17    W2020 31.34363  5.537600
## 18    W2020 59.50813 10.513179
```

```r
#stat tests; 5
# 95ci results
# units: µg/m^3
ci95_alt %>%
  select(loc_date, mean, sd, lower_ci, upper_ci)
```

```
## # A tibble: 6 x 5
##   loc_date  mean     sd lower_ci upper_ci
##   <fct>    <dbl>  <dbl>    <dbl>    <dbl>
## 1 E1750     20.2  0.122     20.1     20.4
## 2 E1950     17.9  0.403     17.5     18.4
## 3 E2020     27.9  0.307     27.6     28.3
## 4 W1750     24.6  0.673     23.8     25.3
## 5 W1950     20.9  0.998     19.8     22.1
## 6 W2020     41.2 15.9       23.2     59.2
```

```r
# 95ci plot (avg conc w 95CI, units: µg/m^3)
ci95
```




```r
# grubbs test for outliers (only W2020 is significant)
grubbs.test(grubbs_df$conc)
```

```
##
##  Grubbs test for one outlier
##
## data:  grubbs_df$conc
## G = 1.1536480, U = 0.0018223, p-value = 0.04078
## alternative hypothesis: highest value 10.2013934404882 is an outlier
```