Stats 253 Hw 1

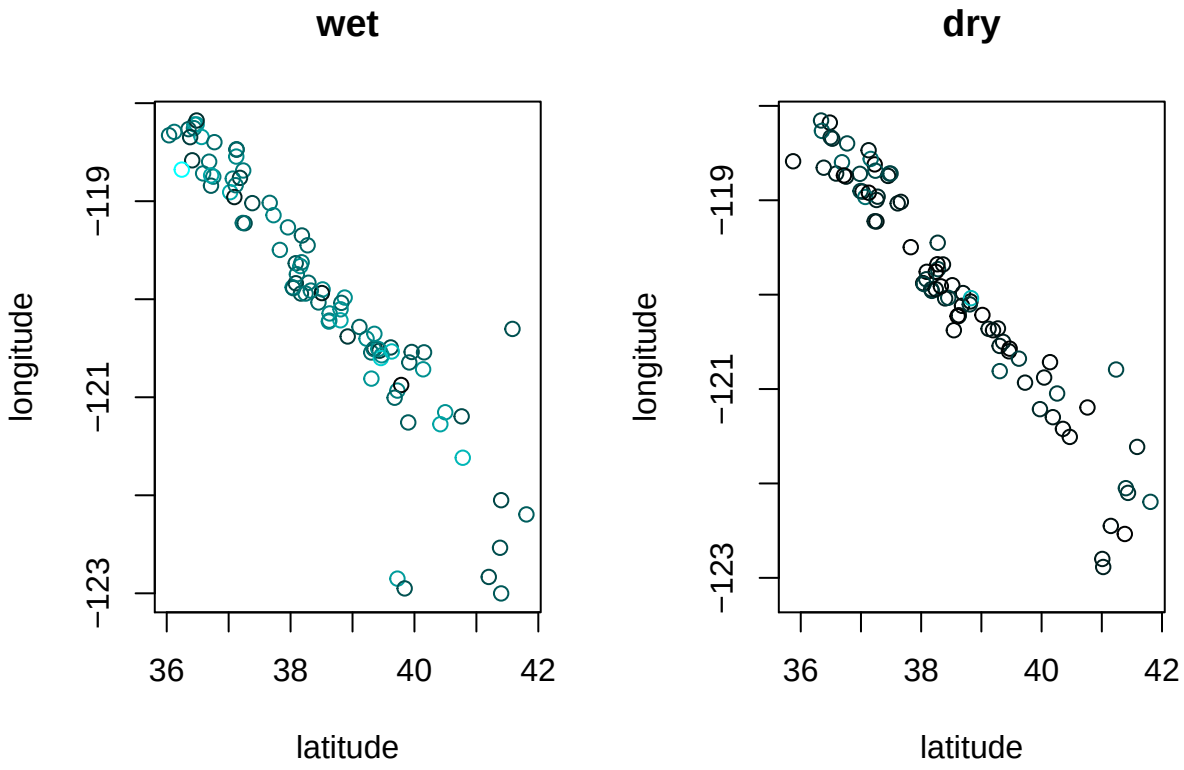
Charles Zheng

07/08/2015

Question 1.

Snowpack plots

```
snowpack <- read.table("snowpack.csv", sep = ",", header = TRUE)
elevations <- read.table("snowpack_elevations.csv", sep = ",", header = TRUE)
snowpack <- cbind(snowpack, elevation = elevations$elevation)

depthcol <- hsv(0.5, 1, 0:170/170)

layout(matrix(1:2, 1, 2))
plot(snowpack[!is.na(snowpack$snow_wet), 1:2], col = depthcol[snowpack$snow_wet + 1])
title("wet")
plot(snowpack[!is.na(snowpack$snow_dry), 1:2], col = depthcol[snowpack$snow_dry + 1])
title("dry")
```



Question 2

Model trend

```r
res_wet <- lm(snow_wet ~ latitude + longitude + elevation, data = snowpack)
summary(res_wet)
```

```
##
## Call:
## lm(formula = snow_wet ~ latitude + longitude + elevation, data = snowpack)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -61.983 -13.711   1.096  14.635  75.894
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.295e+03  5.422e+02  -4.233 4.01e-05 ***
## latitude    -8.593e+00  4.048e+00  -2.123 0.035418 *
## longitude   -2.197e+01  5.579e+00  -3.938 0.000126 ***
## elevation    2.690e-02  5.766e-03   4.666 6.80e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 23.63 on 149 degrees of freedom
##   (107 observations deleted due to missingness)
## Multiple R-squared:  0.1564, Adjusted R-squared:  0.1394
## F-statistic: 9.207 on 3 and 149 DF,  p-value: 1.259e-05
```

```r
length(res_wet$residuals) # 153
```

```
## [1] 153
```

```r
res_dry <- lm(snow_dry ~ latitude + longitude + elevation, data = snowpack)
summary(res_dry)
```

```
##
## Call:
## lm(formula = snow_dry ~ latitude + longitude + elevation, data = snowpack)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -32.891 -11.995  -2.812   9.473  83.867
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -5.967e+02  3.276e+02  -1.822   0.0706 .
## latitude     8.966e-01  2.662e+00   0.337   0.7368
## longitude   -4.327e+00  3.405e+00  -1.271   0.2058
## elevation    2.908e-02  4.329e-03   6.717 3.85e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 17.72 on 146 degrees of freedom
##   (110 observations deleted due to missingness)
## Multiple R-squared:  0.2578, Adjusted R-squared:  0.2426
## F-statistic: 16.91 on 3 and 146 DF,  p-value: 1.764e-09
```

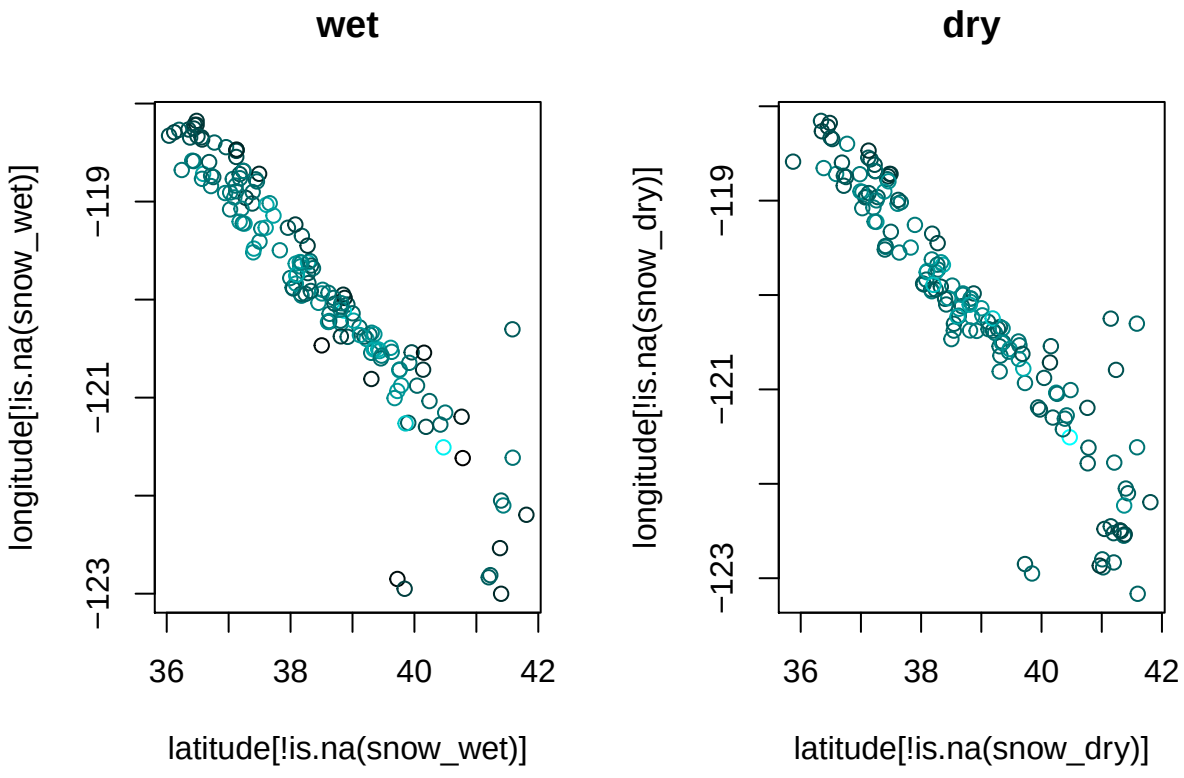
```r
length(res_dry$residuals) # 150
```

```
## [1] 150
```

Question 3

Plot residuals

```r
resid_wet <- res_wet$residuals
resid_dry <- res_dry$residuals
resid <- c(resid_wet, resid_dry)

st_resid <- floor((resid - min(resid)) * 170/(max(resid) - min(resid))) + 1
st_wet <- st_resid[1:153]
st_dry <- st_resid[-(1:153)]

library(magrittr)


layout(matrix(1:2, 1, 2))
snowpack %$% plot(latitude[!is.na(snow_wet)], longitude[!is.na(snow_wet)],
                  col = depthcol[st_wet])
title("wet")
snowpack %$% plot(latitude[!is.na(snow_dry)], longitude[!is.na(snow_dry)],
                  col = depthcol[st_dry])
title("dry")
```



Calculate distances

```
library(geosphere)
```

```
## Loading required package: sp
```

```
library(pracma)
```

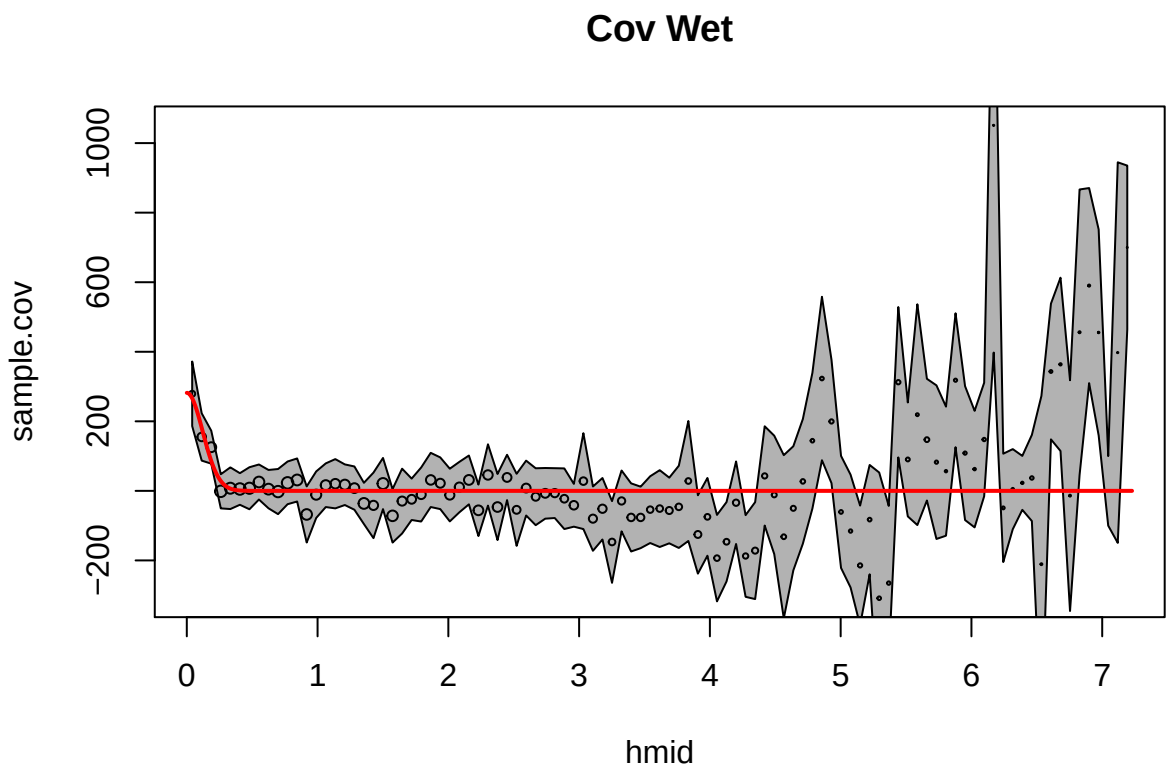
```
##
## Attaching package: 'pracma'
##
## The following object is masked from 'package:geosphere':
##
##     geomean
##
## The following objects are masked from 'package:magrittr':
##
##     and, mod, or
```

```
n <- dim(snowpack)[1]
Dall <- zeros(n, n)
for (i in 1:n) {
  Dall[i, ] <- distGeo(as.matrix(snowpack[i, 2:1]), as.matrix(snowpack[, 2:1]))/1e5
}

e_wet <- resid_wet
D_wet <- Dall[which(!is.na(snowpack$snow_wet)), which(!is.na(snowpack$snow_wet))]
e_dry <- resid_dry
D_dry <- Dall[which(!is.na(snowpack$snow_dry)), which(!is.na(snowpack$snow_dry))]
```

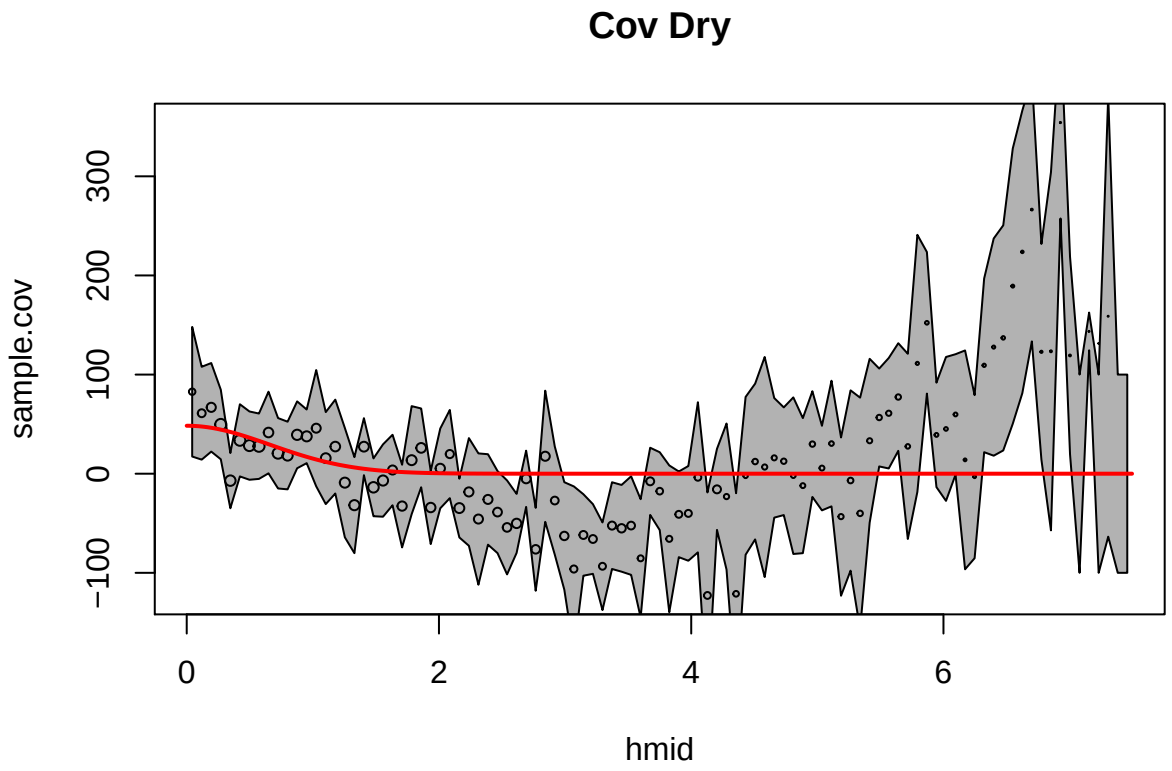
Question 4

Estimate covariances

```
source("covariance.R")
cov.class <- exp2.cov.class

cov_wet <- estimate.cov.fun(e_wet, D_wet, cov.class)
title("Cov Wet")
```




```
cov_dry <- estimate.cov.fun(e_dry, D_dry, cov.class)
title("Cov Dry")
```



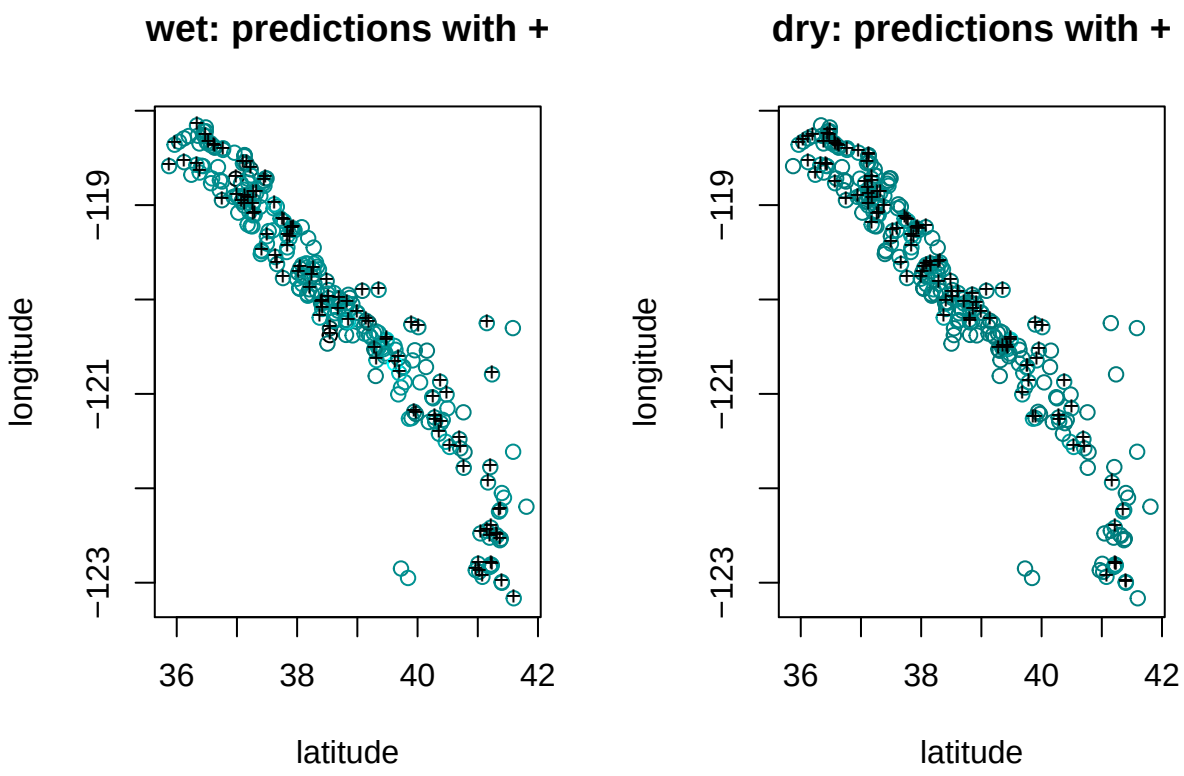
Question 5

Make predictions

```
X_wet <- model.matrix(res_wet)
X_dry <- model.matrix(res_dry)
temp <- 1:dim(snowpack)[1]
X_all <- model.matrix(temp ~ latitude + longitude + elevation, data = snowpack)

pre_wet <- with(snowpack, gls(y = snow_wet[!is.na(snow_wet)], X = X_wet,
                              Sigma = cov_wet(D_wet),
                              SigmaX0_X = cov_wet(Dall[, !is.na(snow_wet)]),
                              X0 = X_all))

pre_dry <- with(snowpack, gls(y = snow_dry[!is.na(snow_dry)], X = X_dry,
                              Sigma = cov_dry(D_dry) + 0.01 * eye(150),
                              SigmaX0_X = cov_dry(Dall[, !is.na(snow_dry)]),
                              X0 = X_all))

predictions <- snowpack
predictions$snow_wet[is.na(snowpack$snow_wet)] <- pre_wet[is.na(snowpack$snow_wet)]
predictions$snow_dry[is.na(snowpack$snow_dry)] <- pre_wet[is.na(snowpack$snow_dry)]
write.table(predictions, "predictions.csv", sep = ",", row.names = FALSE)

rr <- c(predictions$snow_wet, predictions$snow_dry)
st_rr <- floor((rr - min(rr))/(max(rr) - min(rr)) * 170) + 1
st_wet <- st_rr[1:260]
st_dry <- st_rr[-(1:260)]
length(depthcol)
```

```
## [1] 171
```

```
layout(matrix(1:2, 1, 2))
snowpack %$% plot(latitude, longitude,
                  col = depthcol[st_wet])
snowpack %$% points(latitude[is.na(snow_wet)], longitude[is.na(snow_wet)],
                  cex = .8, pch = "+")
title("wet: predictions with +")
snowpack %$% plot(latitude, longitude,
                  col = depthcol[st_dry])
snowpack %$% points(latitude[is.na(snow_dry)], longitude[is.na(snow_dry)],
                  cex = .8, pch = "+")
title("dry: predictions with +")
```



Question 6

Cross-validation

```
pre_snowpack <- function(snowpack) {
  res_wet <- lm(snow_wet ~ latitude + longitude + elevation, data = snowpack)
  res_dry <- lm(snow_dry ~ latitude + longitude + elevation, data = snowpack)
  resid_wet <- res_wet$residuals
  resid_dry <- res_dry$residuals
  e_wet <- resid_wet
  D_wet <- Dall[which(!is.na(snowpack$snow_wet)), which(!is.na(snowpack$snow_wet))]
  e_dry <- resid_dry
  D_dry <- Dall[which(!is.na(snowpack$snow_dry)), which(!is.na(snowpack$snow_dry))]
  cov_wet <- estimate.cov.fun(e_wet, D_wet, cov.class, plot = FALSE)
  cov_dry <- estimate.cov.fun(e_dry, D_dry, cov.class, plot = FALSE)
  X_wet <- model.matrix(res_wet)
  X_dry <- model.matrix(res_dry)
  pre_wet <- with(snowpack, gls(y = snow_wet[!is.na(snow_wet)], X = X_wet,
                                Sigma = cov_wet(D_wet) + 0.001 * eye(dim(D_wet)[1]),
                                SigmaX0_X = cov_wet(Dall[, !is.na(snow_wet)]),
                                X0 = X_all))
  pre_dry <- with(snowpack, gls(y = snow_dry[!is.na(snow_dry)], X = X_dry,
                                Sigma = cov_dry(D_dry) + 0.001 * eye(dim(D_dry)[1]),
                                SigmaX0_X = cov_dry(Dall[, !is.na(snow_dry)]),
                                X0 = X_all))
  predictions <- snowpack
  predictions$snow_wet[is.na(snowpack$snow_wet)] <- pre_wet[is.na(snowpack$snow_wet)]
  predictions$snow_dry[is.na(snowpack$snow_dry)] <- pre_wet[is.na(snowpack$snow_dry)]
```

```
  predictions
}


errs_wet <- numeric()
errs_dry <- numeric()

for(i in 1:10) {
  te_wet <- sample(which(!is.na(snowpack$snow_wet)), 5)
  te_dry <- sample(which(!is.na(snowpack$snow_dry)), 5)
  snowpack2 <- snowpack
  snowpack2$snow_wet[te_wet] <- NA
  snowpack2$snow_dry[te_dry] <- NA
  predictions2 <- pre_snowpack(snowpack2)
  err_wet <- mean((predictions2$snow_wet[te_wet] - snowpack$snow_wet[te_wet])^2)
  err_dry <- mean((predictions2$snow_dry[te_dry] - snowpack$snow_dry[te_dry])^2)
  errs_wet <- c(errs_wet, err_wet)
  errs_dry <- c(errs_wet, err_dry)
}

mean(errs_wet)
```

```
## [1] 14287.83
```

```
mean(errs_dry)
```

```
## [1] 13182.05
```