water_data

Christopher Wetherill

May 15, 2015

Basic Data Munging

The following is an analysis of multidrug resistant samples of *E. coli* taken from the Chobe region of northern Botswana between 2011 and 2012.

Please note that you should edit the code in the block immediately below to reflect the working directory. There's no better way to do this in R (sorry...).

```
setwd("/Users/chris/Desktop/waterData")
```

We will start by reading in our diarrheal and *E. coli* data. We will then reshape the diarrheal data such that we have a count of diarrheal cases for each two-week sampling period.

```
# Read in our data
diarrhea.cases <- read.csv("diarrhea_cases.csv", header = TRUE)
diarrhea.concordance <- read.csv("diarrhea_concordance.csv", header = TRUE)

# Merge the two data frames
diarrhea <- merge(diarrhea.cases, diarrhea.concordance)
diarrhea$From.Date <- as.Date(diarrhea$From.Date, format='%d-%b-%y')

# Read in E. coli data
e.coli <- read.csv("e_coli_resistance.csv", header = TRUE)

e.coli.counts <- read.csv("e_coli_counts.csv", header = TRUE)
e.coli.counts$date <- as.Date(e.coli.counts$date, format='%e-%b-%y')

# Read in floodplain data
floodplains <- read.csv("floodplains.csv", header = TRUE)

# Reconcile dates of observation between E. coli
# and diarrhea datasets. N.B. this isn't a perfect
# concondance table; the observation dates don't
# line up perfectly. It's just a best approximation
# and should be treated as such.
#
# Note that this is also shitty code. I tried. :'(
diarrhea.counts = data.frame(date = c(
  unique(as.Date(e.coli$Date.Collected,
                 format='%m/%d/%y'))),
  cases = c(sum(diarrhea[diarrhea$From.Date == "2011-07-11", 4],
                diarrhea[diarrhea$From.Date == "2011-07-18", 4]),
            sum(diarrhea[diarrhea$From.Date == "2011-07-25", 4],
                diarrhea[diarrhea$From.Date == "2011-08-01", 4]),
            sum(diarrhea[diarrhea$From.Date == "2011-08-08", 4],
                diarrhea[diarrhea$From.Date == "2011-08-15", 4]),
            sum(diarrhea[diarrhea$From.Date == "2011-08-22", 4],
                diarrhea[diarrhea$From.Date == "2011-08-29", 4]),
            sum(diarrhea[diarrhea$From.Date == "2011-08-22", 4],
                diarrhea[diarrhea$From.Date == "2011-08-29", 4]),
            sum(diarrhea[diarrhea$From.Date == "2011-09-05", 4],
                diarrhea[diarrhea$From.Date == "2011-09-12", 4]),
            sum(diarrhea[diarrhea$From.Date == "2011-09-05", 4],
```

```
            diarrhea[diarrhea$From.Date == "2011-09-12", 4]),
          sum(diarrhea[diarrhea$From.Date == "2011-09-19", 4],
            diarrhea[diarrhea$From.Date == "2011-09-26", 4]),
          sum(diarrhea[diarrhea$From.Date == "2011-10-03", 4],
            diarrhea[diarrhea$From.Date == "2011-10-10", 4]),
          sum(diarrhea[diarrhea$From.Date == "2011-10-03", 4],
            diarrhea[diarrhea$From.Date == "2011-10-10", 4]),
          sum(diarrhea[diarrhea$From.Date == "2011-10-17", 4],
            diarrhea[diarrhea$From.Date == "2011-10-24", 4]),
          sum(diarrhea[diarrhea$From.Date == "2011-10-31", 4],
            diarrhea[diarrhea$From.Date == "2011-11-07", 4]),
          sum(diarrhea[diarrhea$From.Date == "2011-11-14", 4],
            diarrhea[diarrhea$From.Date == "2011-11-21", 4]),
          sum(diarrhea[diarrhea$From.Date == "2011-11-28", 4],
            diarrhea[diarrhea$From.Date == "2011-12-05", 4]),
          sum(diarrhea[diarrhea$From.Date == "2011-11-28", 4],
            diarrhea[diarrhea$From.Date == "2011-12-05", 4]),
          sum(diarrhea[diarrhea$From.Date == "2011-12-19", 4],
            diarrhea[diarrhea$From.Date == "2011-12-26", 4]),
          sum(diarrhea[diarrhea$From.Date == "2012-01-02", 4],
            diarrhea[diarrhea$From.Date == "2012-01-09", 4]),
          sum(diarrhea[diarrhea$From.Date == "2012-01-16", 4],
            diarrhea[diarrhea$From.Date == "2012-01-23", 4]),
          sum(diarrhea[diarrhea$From.Date == "2012-02-13", 4],
            diarrhea[diarrhea$From.Date == "2012-02-20", 4]),
          sum(diarrhea[diarrhea$From.Date == "2012-03-05", 4],
            diarrhea[diarrhea$From.Date == "2012-03-12", 4]),
          sum(diarrhea[diarrhea$From.Date == "2012-03-05", 4],
            diarrhea[diarrhea$From.Date == "2012-03-12", 4]),
          sum(diarrhea[diarrhea$From.Date == "2012-04-09", 4],
            diarrhea[diarrhea$From.Date == "2012-04-16", 4]),
          sum(diarrhea[diarrhea$From.Date == "2012-04-23", 4],
            diarrhea[diarrhea$From.Date == "2012-04-30", 4]),
          sum(diarrhea[diarrhea$From.Date == "2011-09-19", 4],
            diarrhea[diarrhea$From.Date == "2011-09-26", 4])
),
kasaneCases = c(sum(diarrhea[diarrhea$From.Date == "2011-07-11" &
                               (diarrhea$Health.Facility ==
                                  "KASANE H.POST" |
                                  diarrhea$Health.Facility ==
                                  "KASANE PRIMARY HOSPITAL"), 4],
                    diarrhea[diarrhea$From.Date == "2011-07-18" &
                               (diarrhea$Health.Facility ==
                                  "KASANE H.POST" |
                                  diarrhea$Health.Facility ==
                                  "KASANE PRIMARY HOSPITAL"), 4]),
                sum(diarrhea[diarrhea$From.Date == "2011-07-25" &
                               (diarrhea$Health.Facility ==
                                  "KASANE H.POST" |
                                  diarrhea$Health.Facility ==
                                  "KASANE PRIMARY HOSPITAL"), 4],
                    diarrhea[diarrhea$From.Date == "2011-08-01" &
                               (diarrhea$Health.Facility ==
                                  "KASANE H.POST" |
                                  diarrhea$Health.Facility ==
                                  "KASANE PRIMARY HOSPITAL"), 4]),
                sum(diarrhea[diarrhea$From.Date == "2011-08-08" &
                               (diarrhea$Health.Facility ==
                                  "KASANE H.POST" |
```

```
                    diarrhea$Health.Facility ==
                    "KASANE PRIMARY HOSPITAL"), 4],
    diarrhea[diarrhea$From.Date == "2011-08-15" &
                (diarrhea$Health.Facility ==
                    "KASANE H.POST" |
                    diarrhea$Health.Facility ==
                    "KASANE PRIMARY HOSPITAL"), 4]),
sum(diarrhea[diarrhea$From.Date == "2011-08-22" &
                (diarrhea$Health.Facility ==
                    "KASANE H.POST" |
                    diarrhea$Health.Facility ==
                    "KASANE PRIMARY HOSPITAL"), 4],
    diarrhea[diarrhea$From.Date == "2011-08-29" &
                (diarrhea$Health.Facility ==
                    "KASANE H.POST" |
                    diarrhea$Health.Facility ==
                    "KASANE PRIMARY HOSPITAL"), 4]),
sum(diarrhea[diarrhea$From.Date == "2011-08-22" &
                (diarrhea$Health.Facility ==
                    "KASANE H.POST" |
                    diarrhea$Health.Facility ==
                    "KASANE PRIMARY HOSPITAL"), 4],
    diarrhea[diarrhea$From.Date == "2011-08-29" &
                (diarrhea$Health.Facility ==
                    "KASANE H.POST" |
                    diarrhea$Health.Facility ==
                    "KASANE PRIMARY HOSPITAL"), 4]),
sum(diarrhea[diarrhea$From.Date == "2011-09-05" &
                (diarrhea$Health.Facility ==
                    "KASANE H.POST" |
                    diarrhea$Health.Facility ==
                    "KASANE PRIMARY HOSPITAL"), 4],
    diarrhea[diarrhea$From.Date == "2011-09-12" &
                (diarrhea$Health.Facility ==
                    "KASANE H.POST" |
                    diarrhea$Health.Facility ==
                    "KASANE PRIMARY HOSPITAL"), 4]),
sum(diarrhea[diarrhea$From.Date == "2011-09-05" &
                (diarrhea$Health.Facility ==
                    "KASANE H.POST" |
                    diarrhea$Health.Facility ==
                    "KASANE PRIMARY HOSPITAL"), 4],
    diarrhea[diarrhea$From.Date == "2011-09-12" &
                (diarrhea$Health.Facility ==
                    "KASANE H.POST" |
                    diarrhea$Health.Facility ==
                    "KASANE PRIMARY HOSPITAL"), 4]),
sum(diarrhea[diarrhea$From.Date == "2011-09-19" &
                (diarrhea$Health.Facility ==
                    "KASANE H.POST" |
                    diarrhea$Health.Facility ==
                    "KASANE PRIMARY HOSPITAL"), 4],
    diarrhea[diarrhea$From.Date == "2011-09-26" &
                (diarrhea$Health.Facility ==
                    "KASANE H.POST" |
                    diarrhea$Health.Facility ==
                    "KASANE PRIMARY HOSPITAL"), 4]),
sum(diarrhea[diarrhea$From.Date == "2011-10-03" &
                (diarrhea$Health.Facility ==
```

```
                    "KASANE H.POST" |
                    diarrhea$Health.Facility ==
                    "KASANE PRIMARY HOSPITAL"), 4],
    diarrhea[diarrhea$From.Date == "2011-10-10" &
                (diarrhea$Health.Facility ==
                    "KASANE H.POST" |
                    diarrhea$Health.Facility ==
                    "KASANE PRIMARY HOSPITAL"), 4]),
sum(diarrhea[diarrhea$From.Date == "2011-10-03" &
                (diarrhea$Health.Facility ==
                    "KASANE H.POST" |
                    diarrhea$Health.Facility ==
                    "KASANE PRIMARY HOSPITAL"), 4],
    diarrhea[diarrhea$From.Date == "2011-10-10" &
                (diarrhea$Health.Facility ==
                    "KASANE H.POST" |
                    diarrhea$Health.Facility ==
                    "KASANE PRIMARY HOSPITAL"), 4]),
sum(diarrhea[diarrhea$From.Date == "2011-10-17" &
                (diarrhea$Health.Facility ==
                    "KASANE H.POST" |
                    diarrhea$Health.Facility ==
                    "KASANE PRIMARY HOSPITAL"), 4],
    diarrhea[diarrhea$From.Date == "2011-10-24" &
                (diarrhea$Health.Facility ==
                    "KASANE H.POST" |
                    diarrhea$Health.Facility ==
                    "KASANE PRIMARY HOSPITAL"), 4]),
sum(diarrhea[diarrhea$From.Date == "2011-10-31" &
                (diarrhea$Health.Facility ==
                    "KASANE H.POST" |
                    diarrhea$Health.Facility ==
                    "KASANE PRIMARY HOSPITAL"), 4],
    diarrhea[diarrhea$From.Date == "2011-11-07" &
                (diarrhea$Health.Facility ==
                    "KASANE H.POST" |
                    diarrhea$Health.Facility ==
                    "KASANE PRIMARY HOSPITAL"), 4]),
sum(diarrhea[diarrhea$From.Date == "2011-11-14" &
                (diarrhea$Health.Facility ==
                    "KASANE H.POST" |
                    diarrhea$Health.Facility ==
                    "KASANE PRIMARY HOSPITAL"), 4],
    diarrhea[diarrhea$From.Date == "2011-11-21" &
                (diarrhea$Health.Facility ==
                    "KASANE H.POST" |
                    diarrhea$Health.Facility ==
                    "KASANE PRIMARY HOSPITAL"), 4]),
sum(diarrhea[diarrhea$From.Date == "2011-11-28" &
                (diarrhea$Health.Facility ==
                    "KASANE H.POST" |
                    diarrhea$Health.Facility ==
                    "KASANE PRIMARY HOSPITAL"), 4],
    diarrhea[diarrhea$From.Date == "2011-12-05" &
                (diarrhea$Health.Facility ==
                    "KASANE H.POST" |
                    diarrhea$Health.Facility ==
                    "KASANE PRIMARY HOSPITAL"), 4]),
sum(diarrhea[diarrhea$From.Date == "2011-11-28" &
```

```
                (diarrhea$Health.Facility ==
                   "KASANE H.POST" |
                   diarrhea$Health.Facility ==
                   "KASANE PRIMARY HOSPITAL"), 4],
    diarrhea[diarrhea$From.Date == "2011-12-05" &
                (diarrhea$Health.Facility ==
                   "KASANE H.POST" |
                   diarrhea$Health.Facility ==
                   "KASANE PRIMARY HOSPITAL"), 4]),
  sum(diarrhea[diarrhea$From.Date == "2011-12-19" &
                (diarrhea$Health.Facility ==
                   "KASANE H.POST" |
                   diarrhea$Health.Facility ==
                   "KASANE PRIMARY HOSPITAL"), 4],
    diarrhea[diarrhea$From.Date == "2011-12-26" &
                (diarrhea$Health.Facility ==
                   "KASANE H.POST" |
                   diarrhea$Health.Facility ==
                   "KASANE PRIMARY HOSPITAL"), 4]),
  sum(diarrhea[diarrhea$From.Date == "2012-01-02" &
                (diarrhea$Health.Facility ==
                   "KASANE H.POST" |
                   diarrhea$Health.Facility ==
                   "KASANE PRIMARY HOSPITAL"), 4],
    diarrhea[diarrhea$From.Date == "2012-01-09" &
                (diarrhea$Health.Facility ==
                   "KASANE H.POST" |
                   diarrhea$Health.Facility ==
                   "KASANE PRIMARY HOSPITAL"), 4]),
  sum(diarrhea[diarrhea$From.Date == "2012-01-16" &
                (diarrhea$Health.Facility ==
                   "KASANE H.POST" |
                   diarrhea$Health.Facility ==
                   "KASANE PRIMARY HOSPITAL"), 4],
    diarrhea[diarrhea$From.Date == "2012-01-23" &
                (diarrhea$Health.Facility ==
                   "KASANE H.POST" |
                   diarrhea$Health.Facility ==
                   "KASANE PRIMARY HOSPITAL"), 4]),
  sum(diarrhea[diarrhea$From.Date == "2012-02-13" &
                (diarrhea$Health.Facility ==
                   "KASANE H.POST" |
                   diarrhea$Health.Facility ==
                   "KASANE PRIMARY HOSPITAL"), 4],
    diarrhea[diarrhea$From.Date == "2012-02-20" &
                (diarrhea$Health.Facility ==
                   "KASANE H.POST" |
                   diarrhea$Health.Facility ==
                   "KASANE PRIMARY HOSPITAL"), 4]),
  sum(diarrhea[diarrhea$From.Date == "2012-03-05" &
                (diarrhea$Health.Facility ==
                   "KASANE H.POST" |
                   diarrhea$Health.Facility ==
                   "KASANE PRIMARY HOSPITAL"), 4],
    diarrhea[diarrhea$From.Date == "2012-03-12" &
                (diarrhea$Health.Facility ==
                   "KASANE H.POST" |
                   diarrhea$Health.Facility ==
                   "KASANE PRIMARY HOSPITAL"), 4]),
```

```
                  sum(diarrhea[diarrhea$From.Date == "2012-03-05" &
                                 (diarrhea$Health.Facility ==
                                    "KASANE H.POST" |
                                    diarrhea$Health.Facility ==
                                    "KASANE PRIMARY HOSPITAL"), 4],
                      diarrhea[diarrhea$From.Date == "2012-03-12" &
                                 (diarrhea$Health.Facility ==
                                    "KASANE H.POST" |
                                    diarrhea$Health.Facility ==
                                    "KASANE PRIMARY HOSPITAL"), 4]),
                  sum(diarrhea[diarrhea$From.Date == "2012-04-09" &
                                 (diarrhea$Health.Facility ==
                                    "KASANE H.POST" |
                                    diarrhea$Health.Facility ==
                                    "KASANE PRIMARY HOSPITAL"), 4],
                      diarrhea[diarrhea$From.Date == "2012-04-16" &
                                 (diarrhea$Health.Facility ==
                                    "KASANE H.POST" |
                                    diarrhea$Health.Facility ==
                                    "KASANE PRIMARY HOSPITAL"), 4]),
                  sum(diarrhea[diarrhea$From.Date == "2012-04-23" &
                                 (diarrhea$Health.Facility ==
                                    "KASANE H.POST" |
                                    diarrhea$Health.Facility ==
                                    "KASANE PRIMARY HOSPITAL"), 4],
                      diarrhea[diarrhea$From.Date == "2012-04-30" &
                                 (diarrhea$Health.Facility ==
                                    "KASANE H.POST" |
                                    diarrhea$Health.Facility ==
                                    "KASANE PRIMARY HOSPITAL"), 4]),
                  sum(diarrhea[diarrhea$From.Date == "2011-09-19" &
                                 (diarrhea$Health.Facility ==
                                    "KASANE H.POST" |
                                    diarrhea$Health.Facility ==
                                    "KASANE PRIMARY HOSPITAL"), 4],
                      diarrhea[diarrhea$From.Date == "2011-09-26" &
                                 (diarrhea$Health.Facility ==
                                    "KASANE H.POST" |
                                    diarrhea$Health.Facility ==
                                    "KASANE PRIMARY HOSPITAL"), 4])
  )
)

# Create a new data frame containing the relevant
# variables from each dataset
water.data <- data.frame(sample = e.coli$Sample.ID,
                         date = as.Date(e.coli$Date.Collected,
                                        format='%m/%d/%y'),
                         isolate = e.coli$Isolate.No.,
                         transect = e.coli$Transect.No.,
                         resistance = e.coli$TOTAL
                         )

# Add in diarrheal incidence data
water.data <- merge(water.data, diarrhea.counts)

# Add in floodplain data
water.data <- merge(water.data, floodplains)
```

```
# Specify wet or dry season based on sampling date
water.data$season[water.data$date >= "2011-07-13" &
                    water.data$date <= "2011-10-21"] <- "Dry"
water.data$season[water.data$date >= "2011-11-03" &
                    water.data$date <= "2012-03-07"] <- "Wet"
water.data$season[water.data$date >= "2012-04-11" &
                    water.data$date <= "2012-04-25"] <- "Dry"

# Specify landuse by transect number
water.data$landuse[water.data$transect >= 31] <- "Park"
water.data$landuse[water.data$transect >= 19 &
                     water.data$transect <= 29] <- "Town"
water.data$landuse[water.data$transect <= 17] <- "Mixed"

# Add in E. coli count (i.e., (#E. coli / vol mL) x 100)
water.data <- merge(water.data, e.coli.counts)

# Convert to factors
water.data <- within(water.data, {
  isolate <- factor(isolate)
  floodplain <- factor(floodplain)
  season <- factor(season)
  landuse <- factor(landuse)
})
```

This ultimately (if not very prettily) produces a final dataset for analysis that takes the general form:

```
##   transect isolate resistance cases kasaneCases floodplain
## 1        1       1          1    28          16          0
## 2        1       4          4    97          38          0
## 3        1       3          2    97          38          0
## 4        1       6          1    97          38          0
## 5        1       2          3    97          38          0
## 6        1       5          1    97          38          0
```

Temporal / Spatial MDR differences

Now we can begin analysing the data. Here, we will construct a linear mixed model, taking the general form $\vec{y} = X\vec{\beta} + Z\vec{u} + \vec{\epsilon}$ where $\vec{y}$ is a vector of known observations, $\vec{\beta}$ is a vector of unknown fixed effects, $\vec{u}$ is a vector of unknown random effects, $\vec{\epsilon}$ is a vector of unknown random errors, and X and Z are known design matrices.

Here, we treat our resistance observations (in terms of number of drugs to which a given isolate of a given sample was found resistant) as the outcome vector; landuse (park, mixed, or town), season (wet or dry), and floodplain (boolean value) as our fixed-factor design matrix; and isolate as a random effect (in our model, we assign this as a random slope).

```
mdr.model <- lmer(resistance ~ landuse + season + floodplain +
                    (1 | isolate), water.data)
summary(mdr.model)
```

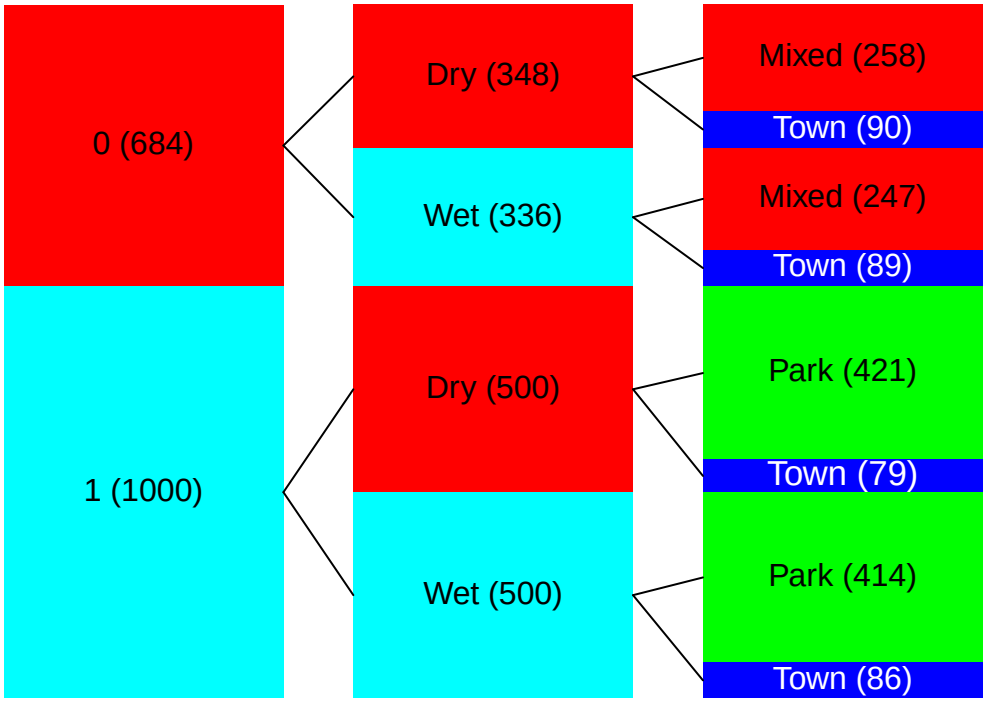
```
## Linear mixed model fit by REML ['lmerMod']
## Formula: resistance ~ landuse + season + floodplain + (1 | isolate)
##    Data: water.data
##
## REML criterion at convergence: 6599.1
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.5306 -0.6381 -0.2297  0.3544  3.7021
```

```
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  isolate  (Intercept) 2.93e-14 1.712e-07
##  Residual             2.93e+00 1.712e+00
## Number of obs: 1684, groups:  isolate, 6
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  1.82258    0.08642  21.091
## landusePark -0.72939    0.20843  -3.500
## landuseTown  0.22772    0.14890   1.529
## seasonWet    0.26962    0.08344   3.231
## floodplain1  0.30008    0.18475   1.624
##
## Correlation of Fixed Effects:
##             (Intr) lndsPr lndsTw sesnWt
## landusePark -0.325
## landuseTown -0.449  0.714
## seasonWet   -0.472  0.007 -0.005
## floodplain1  0.005 -0.886 -0.595 -0.011
```

```
anova(mdr.model)
```

```
## Analysis of Variance Table
##            Df  Sum Sq Mean Sq F value
## landuse     2 170.804  85.402 29.1463
## season      1  30.931  30.931 10.5562
## floodplain  1   7.730   7.730  2.6382
```

These results appear encouraging (landuse and season equate to statistically significant factors), but we should still probably take a look at a couple diagnostic measures. For starters, let's figure out if we have a balanced design:

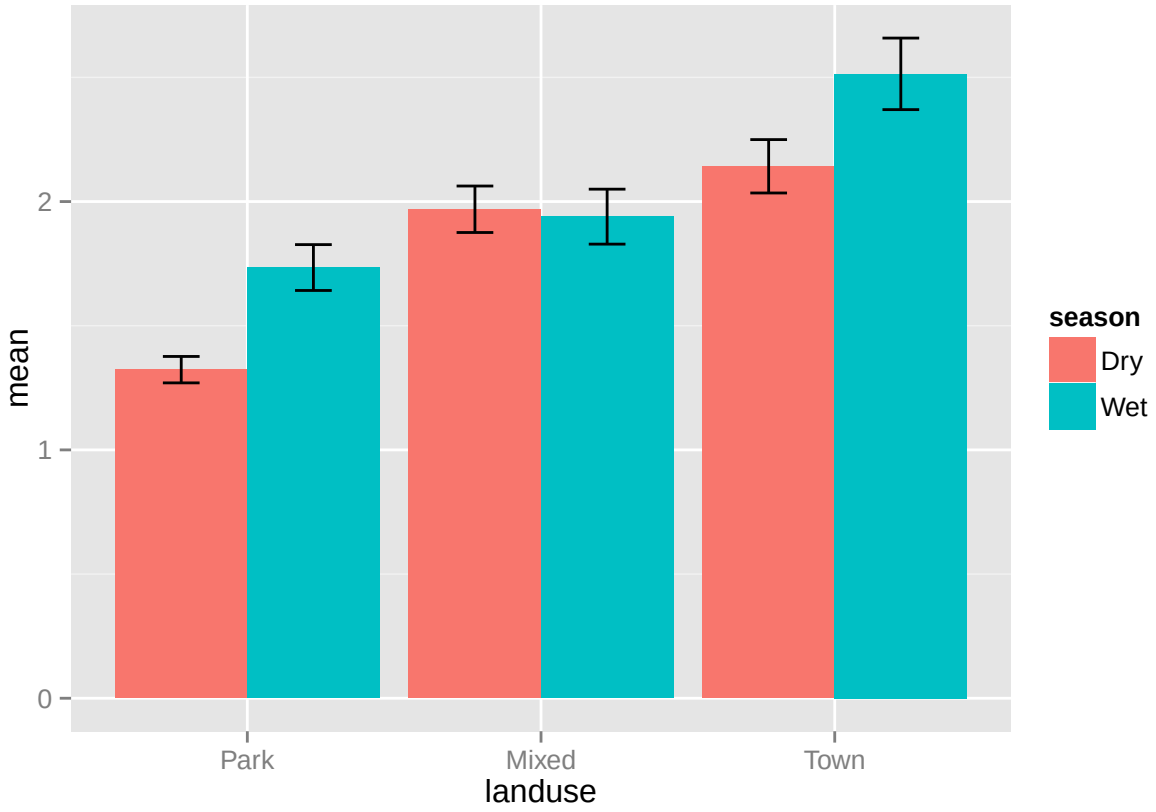


This looks a bit problematic. It appears that we don't have a balanced design: specifically, parks are only found in floodplains and mixed-use land is only found on non-flooding lands. However, since floodplain status does not contribute significantly to our model, we can drop it, giving us instead:

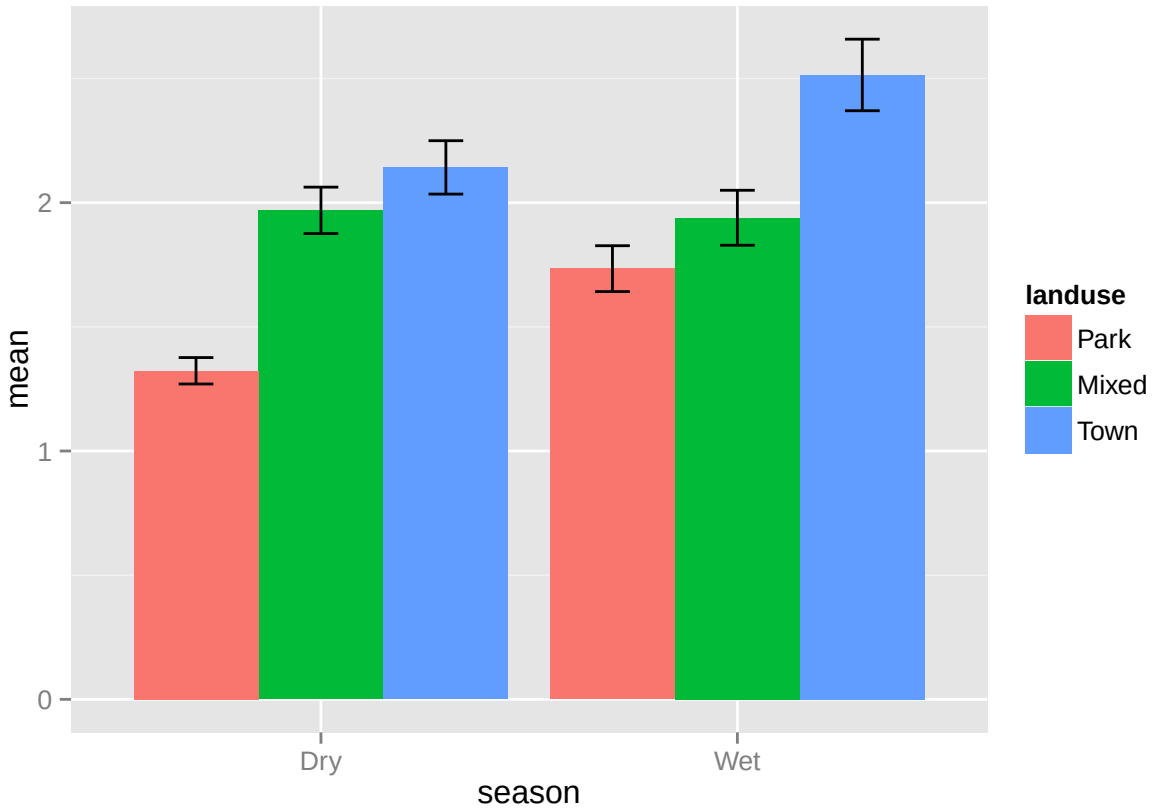
```
## Linear mixed model fit by REML ['lmerMod']
## Formula: resistance ~ landuse + season + (1 | isolate)
##    Data: water.data
##
## REML criterion at convergence: 6600.2
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.4391 -0.6382 -0.2713  0.3547  3.6999
##
## Random effects:
##  Groups   Name        Variance  Std.Dev.
##  isolate  (Intercept) 2.239e-14 1.496e-07
##  Residual             2.933e+00 1.713e+00
## Number of obs: 1684, groups:  isolate, 6
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)  1.82187    0.08646  21.072
## landusePark -0.42933    0.09654  -4.447
## landuseTown  0.37162    0.11974   3.104
## seasonWet    0.27109    0.08348   3.247
##
## Correlation of Fixed Effects:
##             (Intr) lndsPr lndsTw
## landusePark -0.693
## landuseTown -0.555  0.503
## seasonWet   -0.472 -0.006 -0.014


## Analysis of Variance Table
##         Df  Sum Sq Mean Sq F value
## landuse  2 170.804  85.402  29.118
## season   1  30.931  30.931  10.546
```

To visualize this:

```
test <- ddply(water.data, .(season, landuse), summarize,
              mean = mean(resistance), sd = sd(resistance))
test$se <- test$sd / sqrt(c(380,609,257,247,414,175))

dodge <- position_dodge(width = 0.9)
limits <- aes(ymax = test$mean + test$se, ymin=test$mean - test$se)
test$landuse <- with(test, factor(landuse,
                                  levels = c("Park", "Mixed", "Town")))

p <- ggplot(data = test, aes(x = landuse, y = mean, fill = season))
p + geom_bar(stat = "identity", position = dodge) +
  geom_errorbar(limits, position=dodge, width=0.25)
```




```
p <- ggplot(data = test, aes(x = season, y = mean, fill = landuse))
p + geom_bar(stat = "identity", position = dodge) +
  geom_errorbar(limits, position=dodge, width=0.25)
```



MDR Profile — Diarrhea Interplay

Convert Data to Wide

```
water <- water.data[, c(1:2, 4:5)]
water <- water[order(water$date, water$transect, water$isolate), ]
water.wide <- reshape(water,
                      timevar = "isolate",
                      idvar = c("transect", "date"),
                      direction = "wide")
colnames(water.wide) <- c("transect", "date", "isolate1", "isolate2",
                          "isolate3", "isolate4", "isolate5", "isolate6")

water.wide$mdrCount <- rowSums(water.wide[, 3:8] >= 3, na.rm = TRUE)
water.wide$numIsolates <- rowSums(!is.na(water.wide[, 3:8]))
water.wide$propMDR <- water.wide$mdrCount / water.wide$numIsolates

# Add diarrheal data
water.wide <- merge(water.wide, diarrhea.counts)

# Specify wet or dry season based on sampling date
water.wide$season[water.wide$date >= "2011-07-13" &
                    water.wide$date <= "2011-10-21"] <- "Dry"
water.wide$season[water.wide$date >= "2011-11-03" &
                    water.wide$date <= "2012-03-07"] <- "Wet"
water.wide$season[water.wide$date >= "2012-04-11" &
                    water.wide$date <= "2012-04-25"] <- "Dry"

# Specify landuse by transect number
water.wide$landuse[water.wide$transect >= 31] <- "Park"
water.wide$landuse[water.wide$transect >= 19 &
                     water.wide$transect <= 29] <- "Town"
water.wide$landuse[water.wide$transect <= 17] <- "Mixed"

# Add in E. coli count (i.e., (#E. coli / vol mL) x 100)
water.wide <- merge(water.wide, e.coli.counts)

# Convert to factors
water.wide <- within(water.wide, {
  season <- factor(season)
  landuse <- factor(landuse)
  count <- as.numeric(count)
})
```

This gives us data taking the general form:

```
##         date transect isolate1 isolate2 mdrCount numIsolates   propMDR
## 1 2011-07-13        1        1       NA        0           1 0.0000000
## 2 2011-07-13       11        2        0        2           5 0.4000000
## 3 2011-07-13       13        2        0        1           6 0.1666667
## 4 2011-07-13       15        2        2        1           5 0.2000000
## 5 2011-07-13       17       NA        3        4           4 1.0000000
## 6 2011-07-13       19        6        3        4           4 1.0000000
```

ANCOVA

```
fit1 <- aov(propMDR ~ season * landuse + kasaneCases, data = water.wide)
fit2 <- aov(propMDR ~ season * landuse, data = water.wide)

summary(fit1)
```

```
##                 Df Sum Sq Mean Sq F value   Pr(>F)
## season           1  0.136  0.1359   2.048   0.1533
## landuse          2  3.562  1.7810  26.844 1.61e-11 ***
## kasaneCases      1  0.606  0.6062   9.137   0.0027 **
## season:landuse   2  0.143  0.0715   1.078   0.3414
## Residuals      327 21.696  0.0663
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(fit2)
```

```
##                 Df Sum Sq Mean Sq F value  Pr(>F)
## season           1  0.136  0.1359   1.996   0.159
## landuse          2  3.562  1.7810  26.152 2.9e-11 ***
## season:landuse   2  0.108  0.0538   0.790   0.455
## Residuals      328 22.337  0.0681
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(fit1, fit2)
```

```
## Analysis of Variance Table
##
## Model 1: propMDR ~ season * landuse + kasaneCases
## Model 2: propMDR ~ season * landuse
##   Res.Df    RSS Df Sum of Sq      F   Pr(>F)
## 1    327 21.696
## 2    328 22.337 -1   -0.6417 9.6718 0.002036 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

MANOVA

```
fit3 <- aov(kasaneCases ~ season * landuse, data = water.wide)

y <- cbind(water.wide$propMDR, water.wide$kasaneCases)
fit4 <- manova(y ~ water.wide$season + water.wide$landuse)
summary(fit3, test = "Pillai")
```

```
##                 Df Sum Sq Mean Sq F value   Pr(>F)
## season           1   1028  1028.2  11.638 0.000727 ***
## landuse          2    219   109.6   1.240 0.290632
## season:landuse   2    142    71.0   0.804 0.448635
## Residuals      328  28976    88.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary.aov(fit4)
```

```
##  Response 1 :
##                      Df  Sum Sq Mean Sq F value    Pr(>F)
## water.wide$season     1  0.1359 0.13590  1.9982    0.1584
## water.wide$landuse    2  3.5620 1.78101 26.1855 2.785e-11 ***
## Residuals           330 22.4450 0.06802
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##  Response 2 :
##                      Df  Sum Sq Mean Sq F value    Pr(>F)
## water.wide$season     1  1028.2 1028.15 11.6523 0.0007211 ***
## water.wide$landuse    2   219.2  109.58  1.2418 0.2901973
## Residuals           330 29118.0   88.24
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

MDR / E. coli count relationship

```
fit5 <- lm(propMDR ~ count, data = water.wide)
summary(fit5)
```

```
##
## Call:
## lm(formula = propMDR ~ count, data = water.wide)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.25889 -0.25087 -0.08041  0.14688  0.75340
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.591e-01  3.244e-02   7.988 2.28e-14 ***
## count       -2.192e-05  9.285e-05  -0.236    0.814
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2806 on 332 degrees of freedom
## Multiple R-squared:  0.0001678,  Adjusted R-squared:  -0.002844
## F-statistic: 0.05572 on 1 and 332 DF,  p-value: 0.8135
```