Final Project – Second Draft

Kan Luo, Shih-Ni Prim

10/24/2020

Contents

Questions/next steps

- Questions for Kan:
  - Samples from different organs?
  - Differences among isotypes?
- Why are there 7 treatment groups now?
  - Treatment 1-3: different doses for 1st drug
  - Treatment 4-6: different doses for 2nd drug
  - Treatment 7: control
- Instead of looking at average of Binding, look at average of reactivity to see the percentage of reactive
- Variances can be very different in different cells
- summaries of data points: time points, treatment groups, response variables, etc.

Data Preparation

The resulting dataset from this section is called `Data2`, and one outlier will be removed in a later section, which results in a final dataset `Data3`. So `Data3` is used for analysis. [This section won't be in the final report.]

```
## Response [https://raw.githubusercontent.com/luokan1227/537P1/master/Data.xlsx]
##   Date: 2020-10-25 02:54
##   Status: 200
##   Content-Type: application/octet-stream
##   Size: 341 kB
## <ON DISK>  C:\Users\shihn\AppData\Local\Temp\RtmpAbhvc2\file3bec418676a0.xlsx
```

```
## Response [https://raw.githubusercontent.com/luokan1227/537P1/master/MonkeyID.xlsx]
##   Date: 2020-10-25 02:54
##   Status: 200
##   Content-Type: application/octet-stream
##   Size: 50.1 kB
## <ON DISK>  C:\Users\shihn\AppData\Local\Temp\RtmpAbhvc2\file3bec6dc5487f.xlsx
```

Abstract

Introduction

Methodologies

Data Summaries

Our data came from a vaccine study, in which 20 rhesus marcaque was given HIV vaccines as well as immunosuppressing treatments that can inhibit human body's mechanism to suppress antibodies and, in theory, enhance the effect of vaccines. The dataset has 2465 data points, 20 rhesus monkeys.

```
table(Data2$MonkeyID)
```

```
##
## 6104 6105 6107 6117 6118 6119 6125 6132 6160 6193 6199 6200 6201 6202 6203 6204
##   35  228  239  243    7   55  216  251  183  117   48  191   73   78  238  156
## 6205 6209 6210 6214
##    5   46   50    6
```

In the current analysis, each row represents one antibody and its measurements. While it is possible to treat the 20 rhesus macaques as the observational units, the analysis will become quite complex. We would have to use information from gene segments of the heavy and light chains of the antibodies to classify each antibody. There are five such segments, and each has 6 to 11 subtypes. (In total, there can be $7 * 7 * 6 * 11 * 6 = 1.9404 \times 10^4$ combinations of gene segments.) Thus, we decided to follow the convention of vaccine studies and treat each antibody as independent.

The main goal of the study is to test whether any procedures can incease the diversity of antibodies and thus enhance the effectiveness of vaccines. Our predictors are time points, which indicates time and the number of vaccines given up to that point, treatment or drug, and Isotype. We first take a look at these variables.

There are four time points; one before any procedure was done, and three after vaccine shots were administered to the macaques. In the treatment groups, groups 1-3 represent different doses of drug 1, groups 4-6 represent different doses of drug 2, and group 7 represents the control group. Later we'll look at the effect made by different drugs first and then different doses.

```
table(Data2$Time_Point, Data2$Treatment)
```

```
##
##     group 1 group 2 group 3 group 4 group 5 group 6 group 7
##   0     129       0       0      90       0       0      54
##   1     190      60      96     105     297     131     125
```

```
##   2     141     110       0     148      77       0     347
##   3     122       0       0     101       0       0     142
```

```
table(Data2$Drug, Data2$Treatment)
```

```
##
##     group 1 group 2 group 3 group 4 group 5 group 6 group 7
##   1     582     170      96       0       0       0       0
##   2       0       0       0     444     374     131       0
##   3       0       0       0       0       0       0     668
```

```
table(Data2$Drug, Data2$Treatment)
```

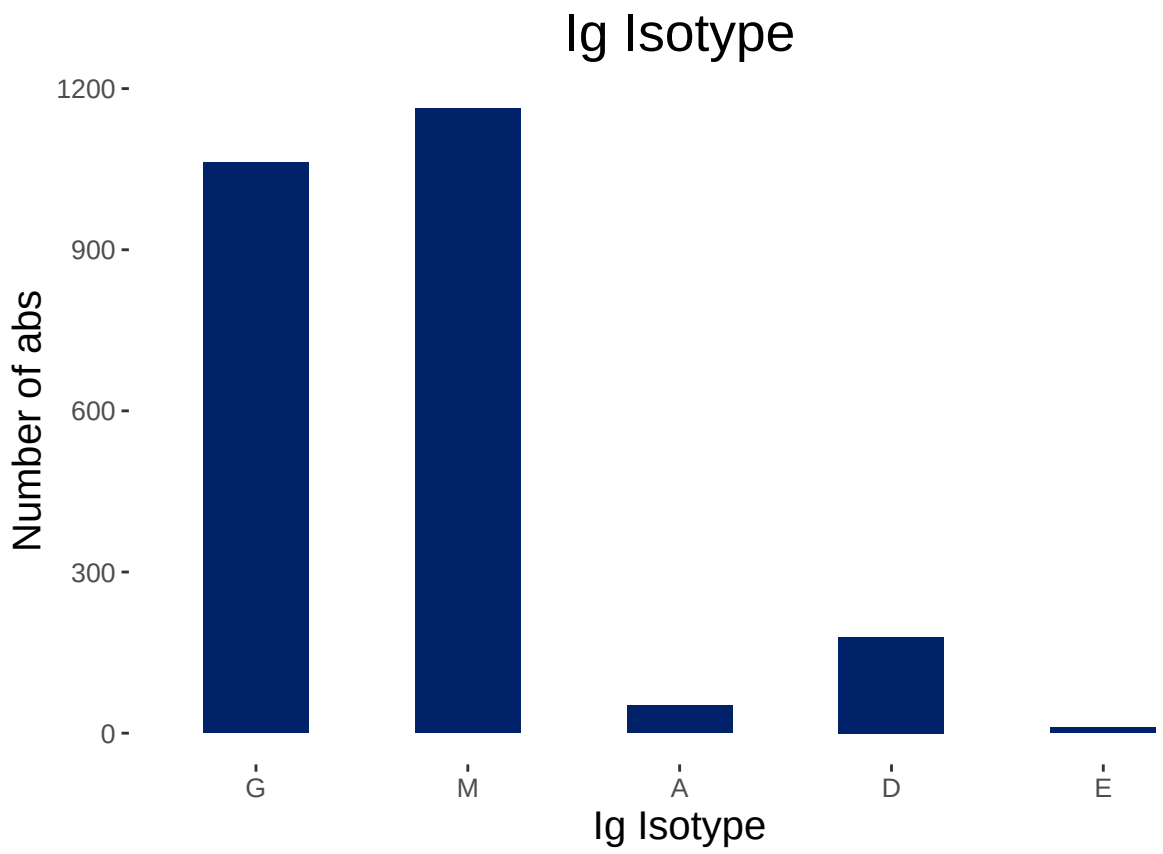
```
##
##     group 1 group 2 group 3 group 4 group 5 group 6 group 7
##   1     582     170      96       0       0       0       0
##   2       0       0       0     444     374     131       0
##   3       0       0       0       0       0       0     668
```

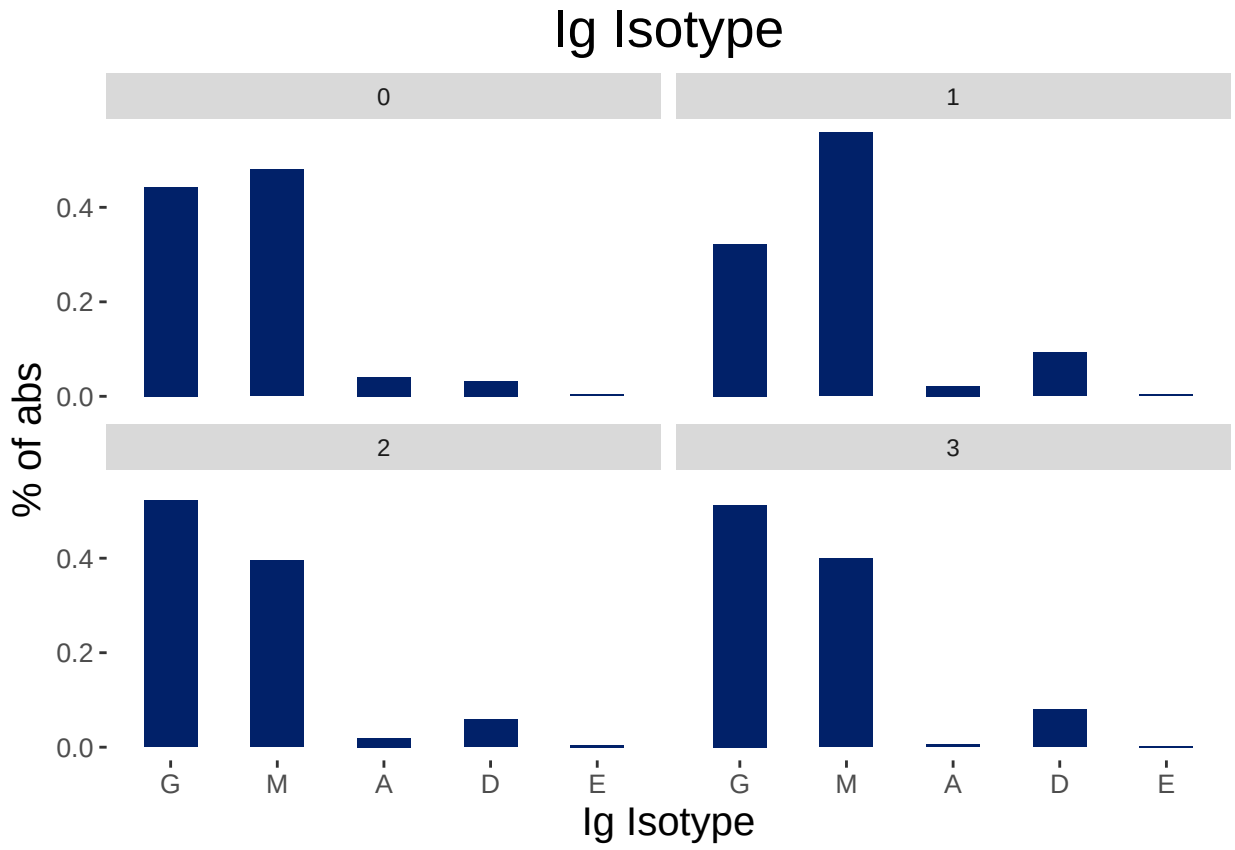
Each marcaque only received one kind of treatment.

```
table(Data2$MonkeyID, Data2$Treatment)
```

```
##
##        group 1 group 2 group 3 group 4 group 5 group 6 group 7
##   6104       0       0      35       0       0       0       0
##   6105       0       0       0     228       0       0       0
##   6107       0       0       0       0       0       0     239
##   6117     243       0       0       0       0       0       0
##   6118       0       7       0       0       0       0       0
##   6119       0       0      55       0       0       0       0
##   6125       0       0       0     216       0       0       0
##   6132       0       0       0       0     251       0       0
##   6160     183       0       0       0       0       0       0
##   6193       0     117       0       0       0       0       0
##   6199       0       0       0       0       0      48       0
##   6200       0       0       0       0       0       0     191
##   6201       0       0       0       0      73       0       0
##   6202       0       0       0       0       0      78       0
##   6203       0       0       0       0       0       0     238
##   6204     156       0       0       0       0       0       0
##   6205       0       0       0       0       0       5       0
##   6209       0      46       0       0       0       0       0
##   6210       0       0       0       0      50       0       0
##   6214       0       0       6       0       0       0       0
```

Next, we'll take a look at the variable `Isotype`. There are 5 kinds of heavy chain for antibodies: IgG, IgA, IgM, IgE, IgD. The two most important kinds are IgG and IgM. IgM mostly occurs in the acute stage of infection, and IgM appears later in blood with better neutralizing potentials.


Ig Isotype
1200
900
600
300
0
Number of abs
G
M
A
D
E
Ig Isotype

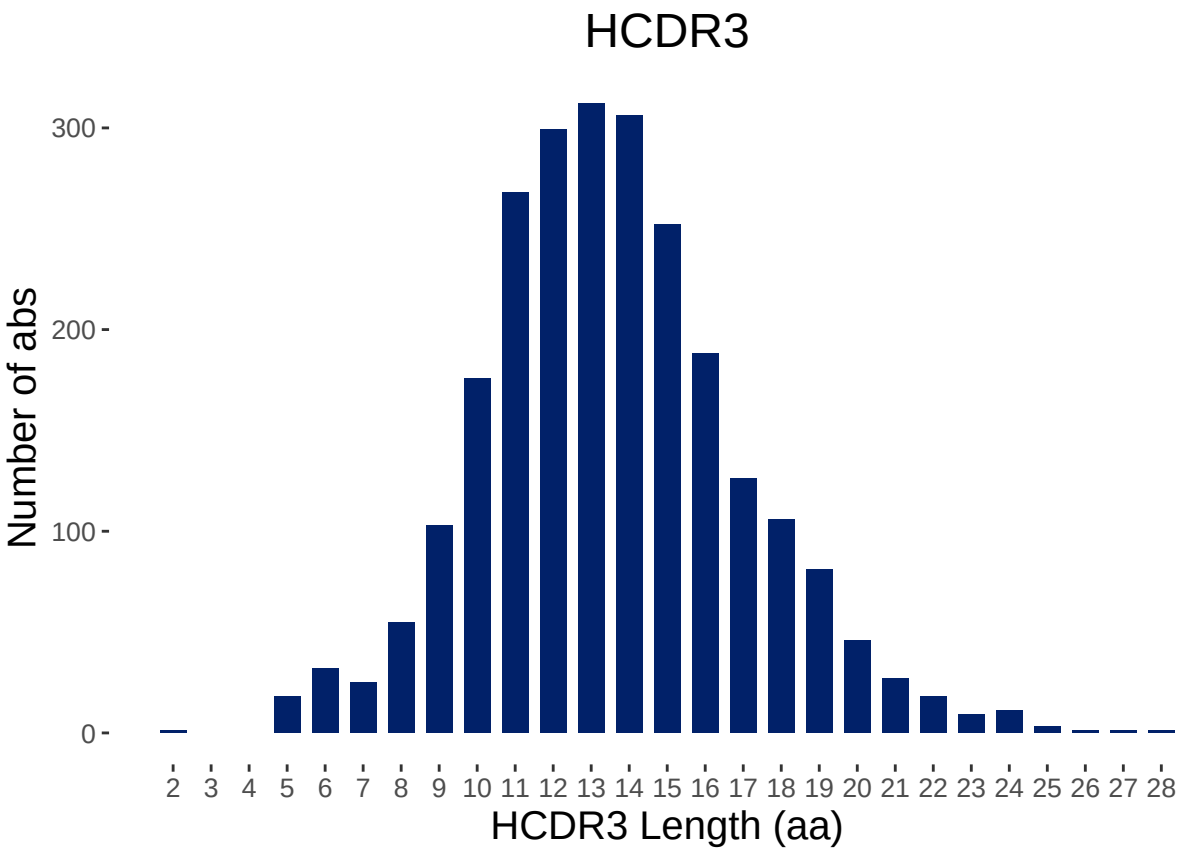


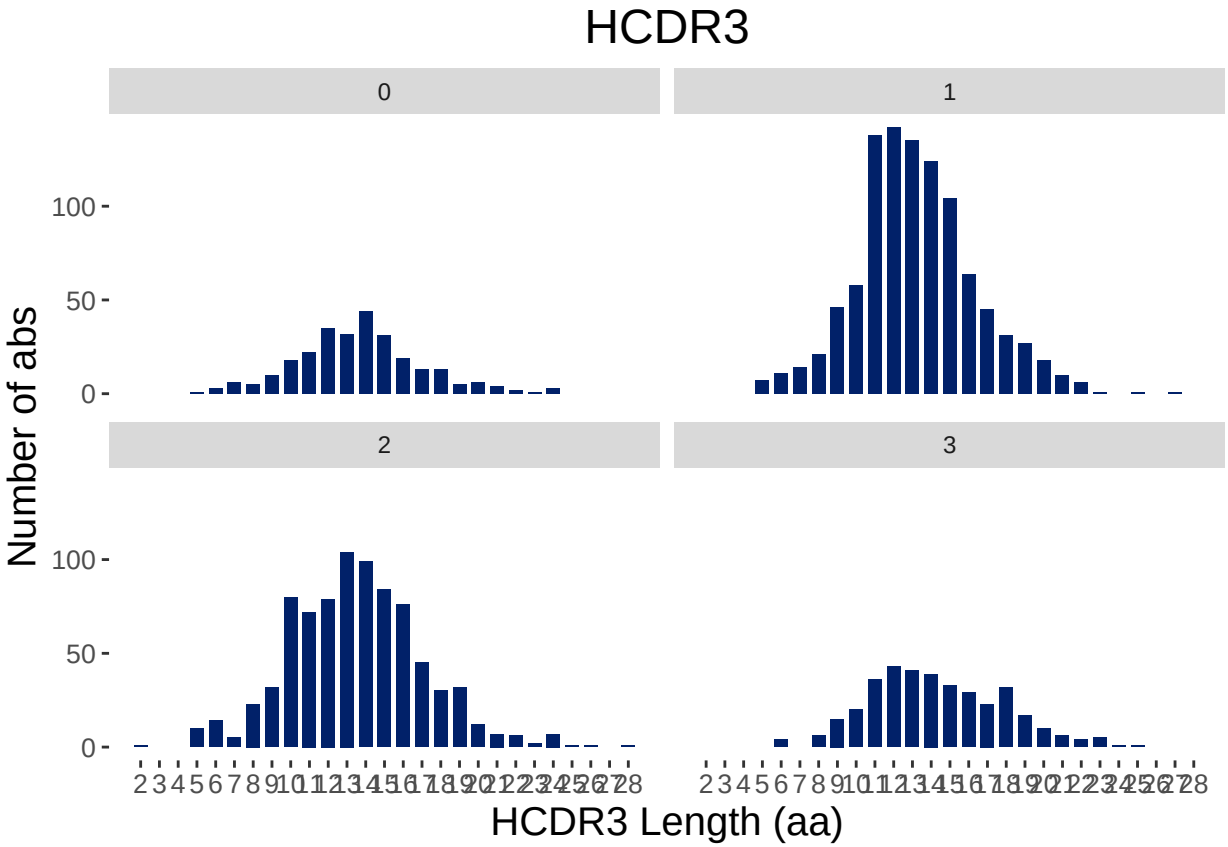

```
##   Isotype Ab # Ab %
## 1       A   51  2.1
## 2       D  179  7.3
## 3       E   10  0.4
## 4       G 1062 43.1
## 5       M 1163 47.2
```

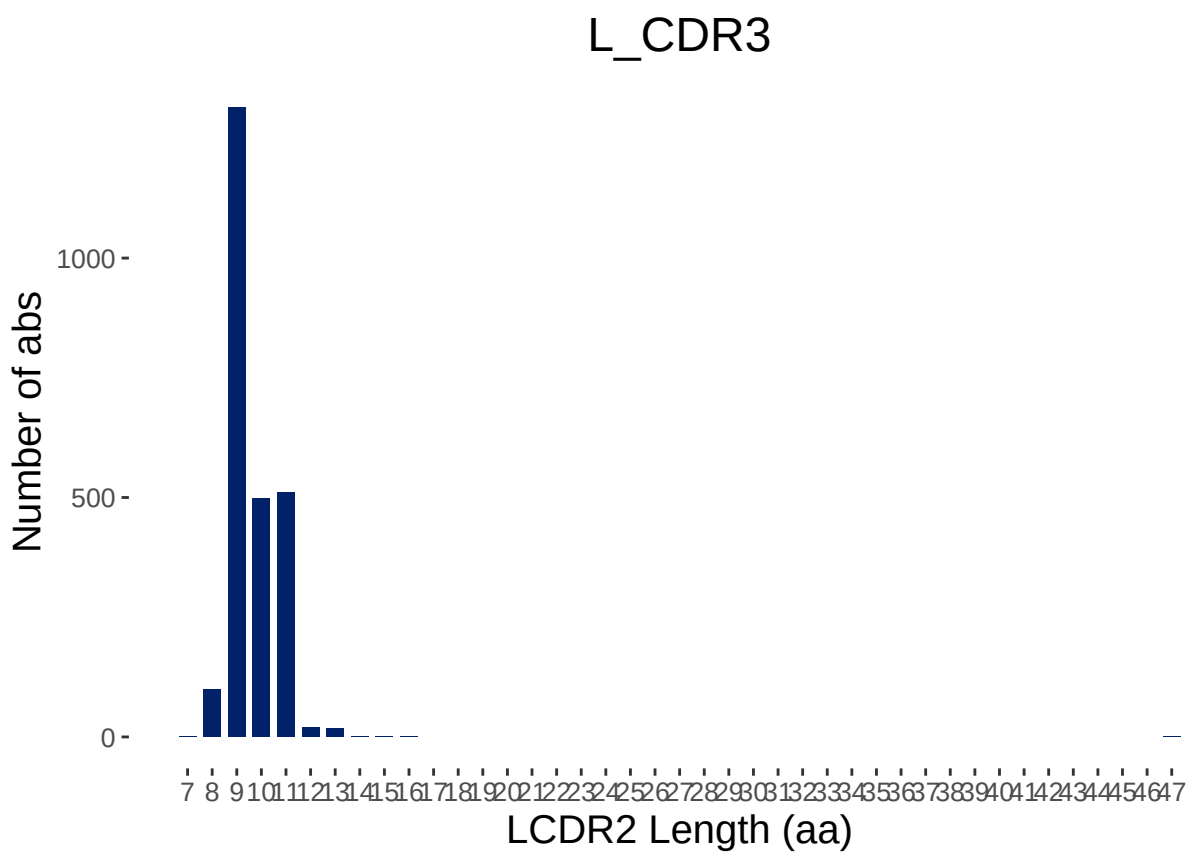
As expected, we see that IgG and IgM occupy the biggest proportion of all antibodies in all time points. We'll use the variable `Isotype` as a grouping covariate later.

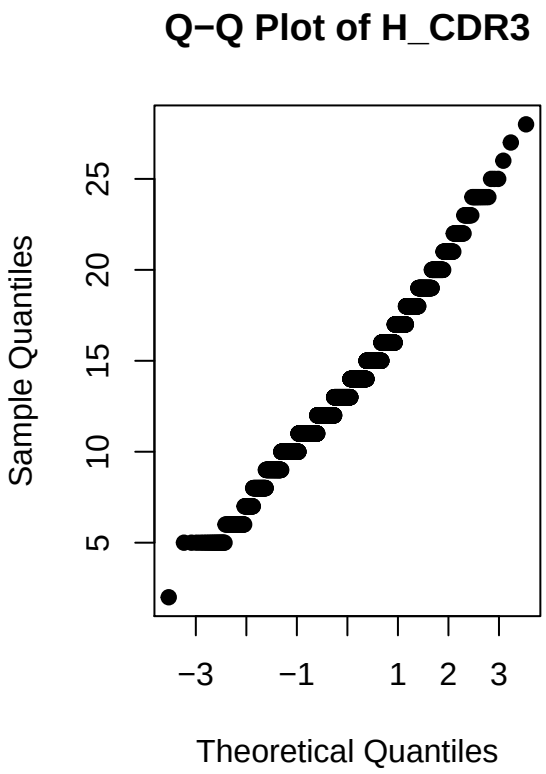
Next we'll examine our responses: `H_CDR3`, `HMuFreq`, `L_CDR3`, `LMuFreq`, `Binding`, and `Reactivity`. In each antibody, there are two sets of heavy chain and light chain, all of which forming a Y-shape immunoglobulin. Thus many of the variables start with H or L, indicating which chain the information comes from.

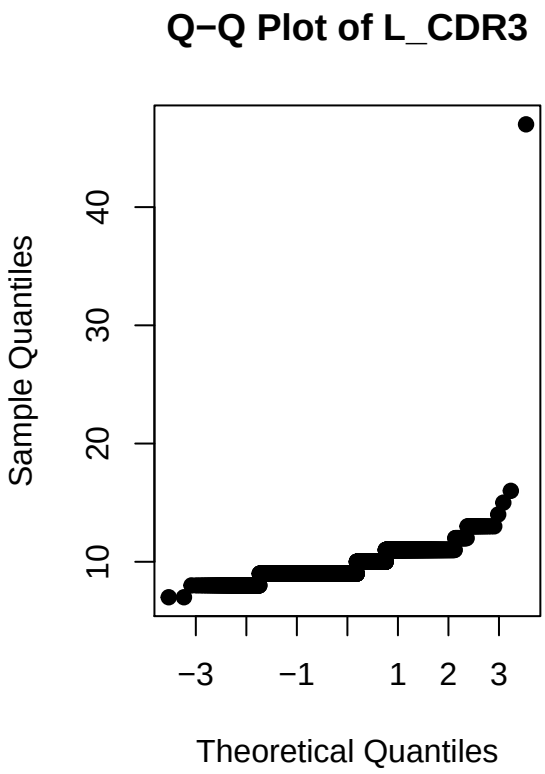
`H_CDR3` and `L_CDR3` indicates the length of the third complementarity-determining region on the variable heavy chain and light chain. The longer they are, the more potential there is to produce diverse antibodies. [Kan, could you check to see if this is correct?] In other words, we want the values to be higher. Below we see that the distributions are roughly normal with the center around 13 for `H_CDR3`, with all data points, and slightly centers for different time points. For the Q-Q plot for `L_CDR3`, we can see that there is one outlier. Without it, the distribution is likely normal. (We'll get to this soon.)


HCDR3
Number of abs
300
200
100
0
2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28
HCDR3 Length (aa)

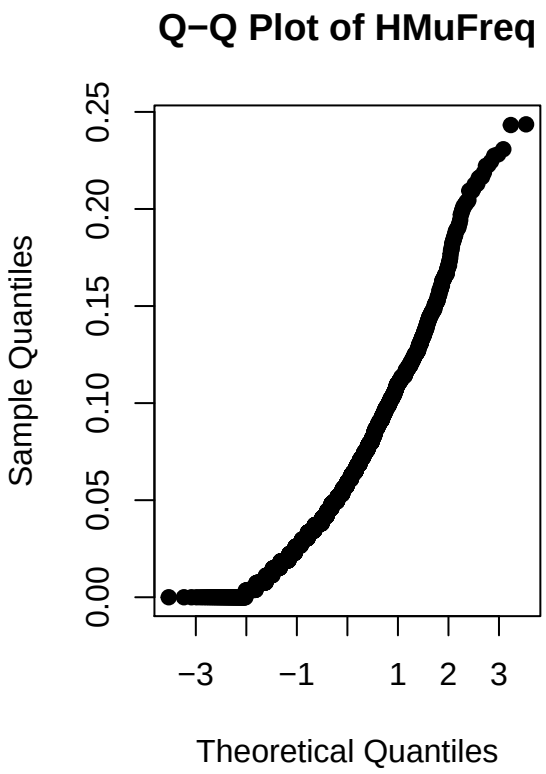

HCDR3
0
1
2
3
Number of abs
100
50
0
HCDR3 Length (aa)

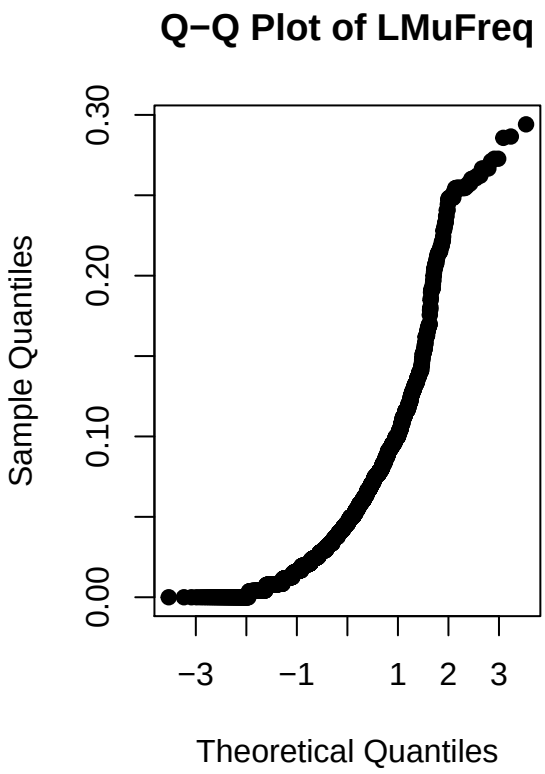

L_CDR3
Number of abs
1000
500
0
7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47
LCDR2 Length (aa)



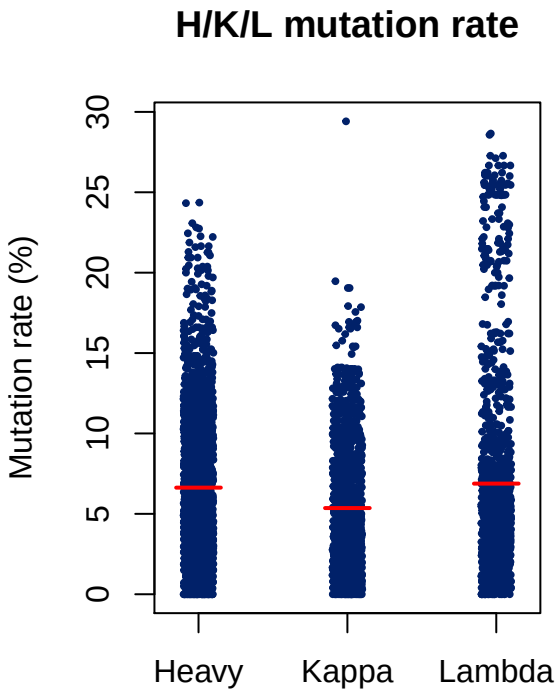


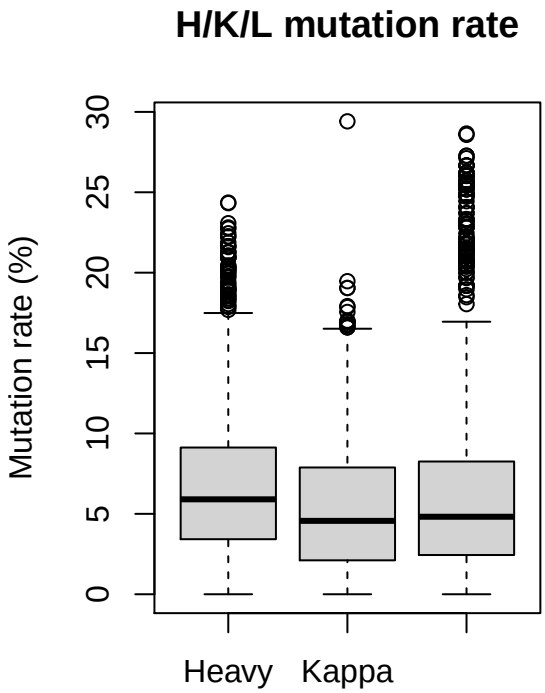
`HMuFreq` and `LMuFreq` are calculated by dividing `H_Substitution` by `H_VBase` for heavy chain and similarly for light chain. These two variables show how much the antibodies mutate. For the purpose of the study, the higher the mutation rate is, the better. Below we see some comparison of mutation rate between heavy chain and light chain. (Kappa and Lambda are two kinds of light chain.) [Kan, is there a reason to split up light chain into Kappa and Lambda? Could we simply plot heavy chain vs. light chain?]

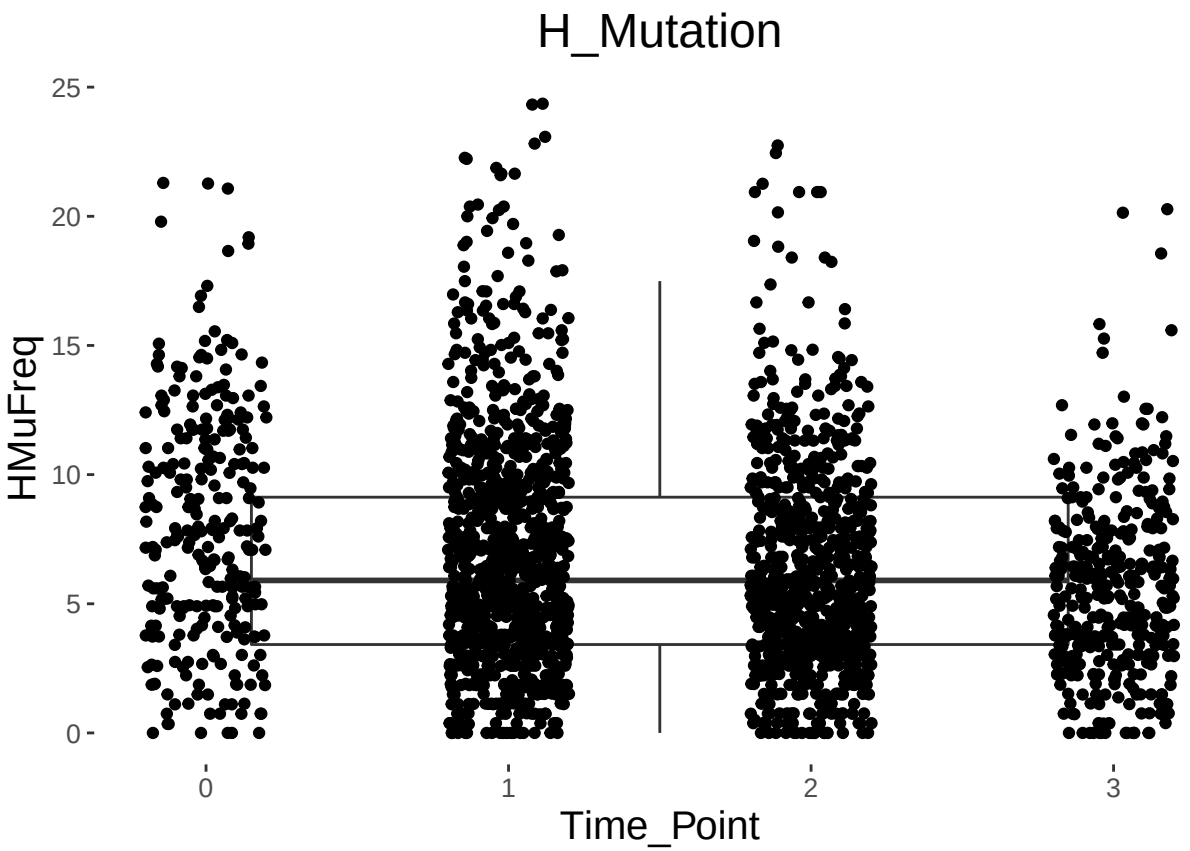





```
##         H_Mutation% K_Mutation% L_Mutation%
## Min.           0.00        0.00        0.00
## 1st Qu.        3.42        2.11        2.44
## Median         5.90        4.56        4.82
## Mean           6.63        5.36        6.88
## 3rd Qu.        9.13        7.88        8.25
## Max.          24.36       29.41       28.65
```

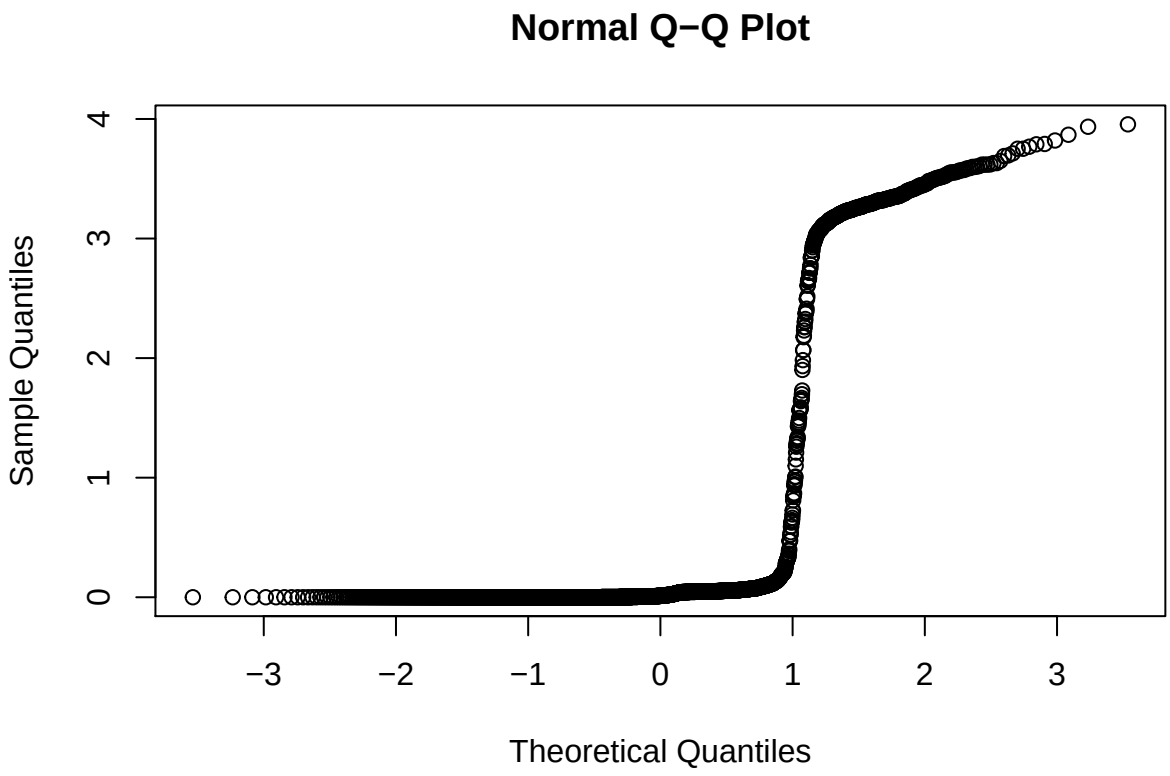

H/K/L mutation rate
Mutation rate (%)
0
5
10
15
20
25
30
Heavy
Kappa
Lambda


H/K/L mutation rate
Mutation rate (%)
0
5
10
15
20
25
30
Heavy
Kappa



Lastly, `Binding` indicates the rate of neutralizing, meaning how much the antibodies bind with the virus and thus make the virus ineffective. This is the most important measure of the study. `Reactivity` turns `Binding` into a binary variable; `Binding` rate above 0.1 is considered reactive. In the Q-Q plot of `Binding`, we can see that it is not normally distributed. Thus the `Reactivity` measure might be a better response to use. However, since our sample size is larger than 2000, we can use the Central Limit Theorem and assume normality.

```
qqnorm(Data2$Binding)
```

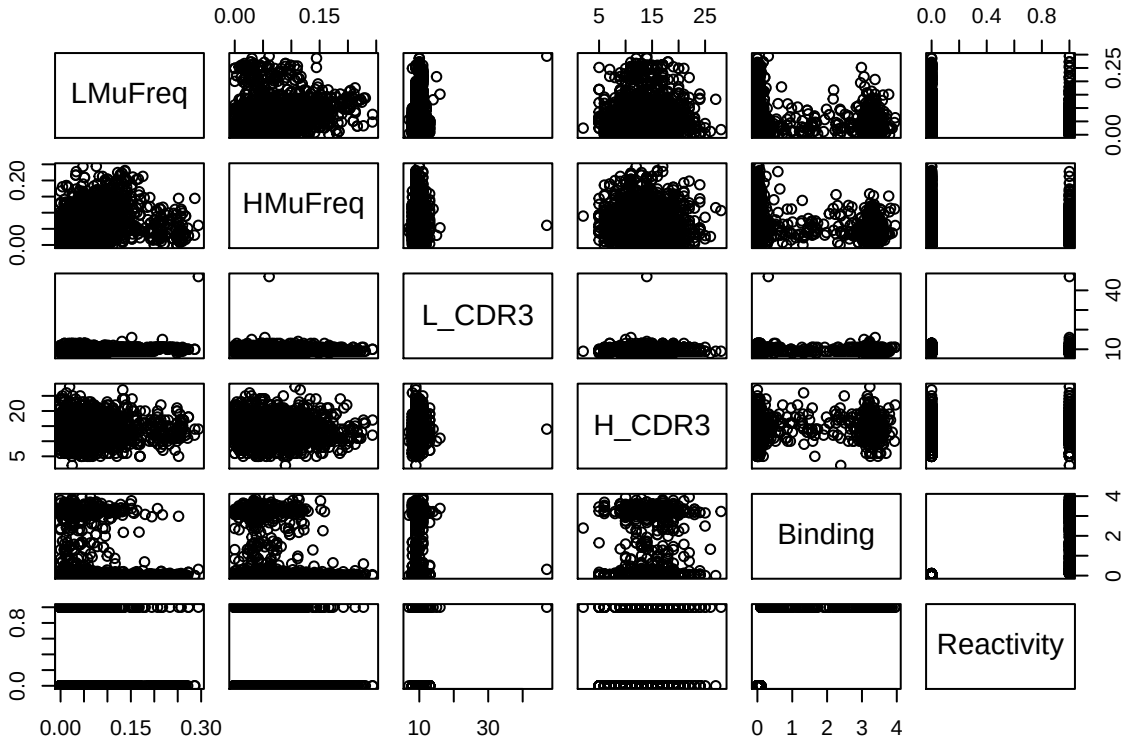



```
table(Data2$Reactivity)
```

```
##
##    0    1
## 1951  514
```

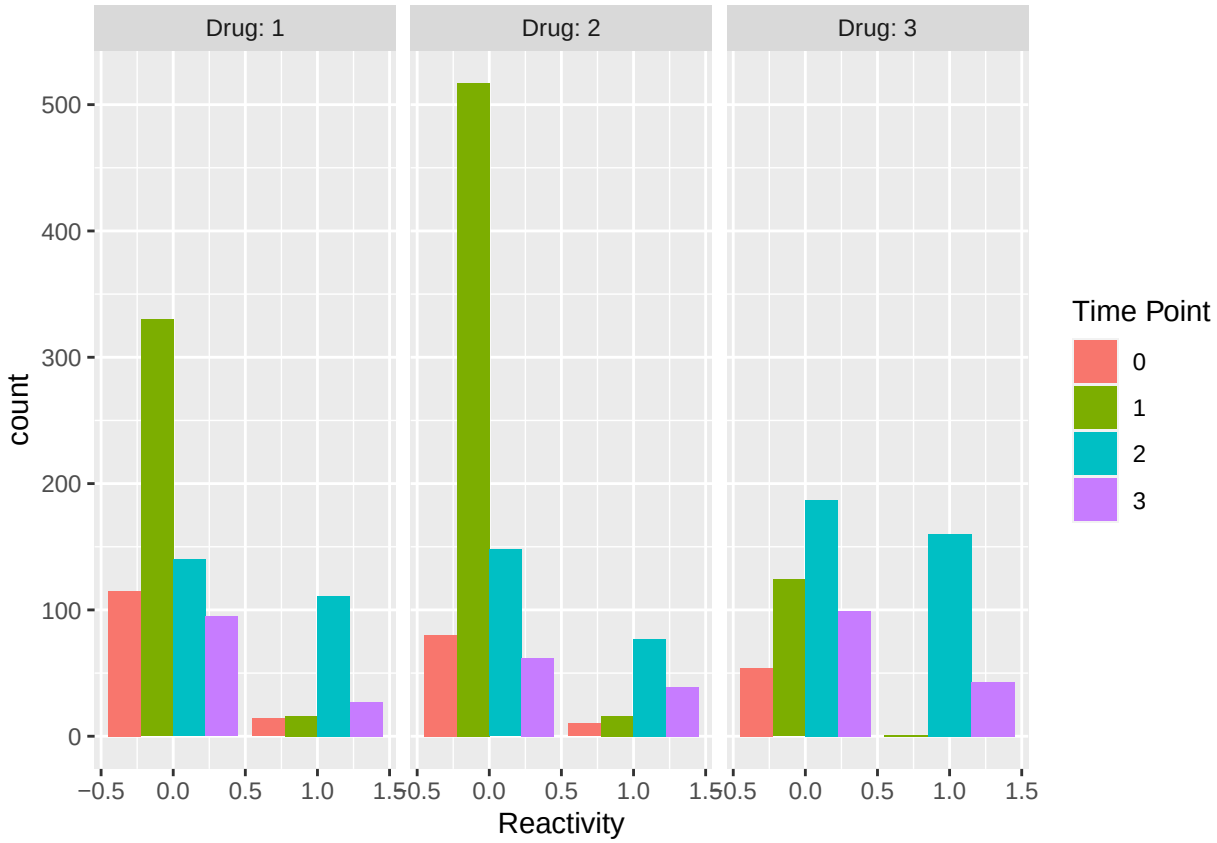
Let's take a look at these response variables and check whether they might be correlated.

```
Data2 %>% select(LMuFreq, HMuFreq, L_CDR3, H_CDR3, Binding, Reactivity) %>% pairs()
```

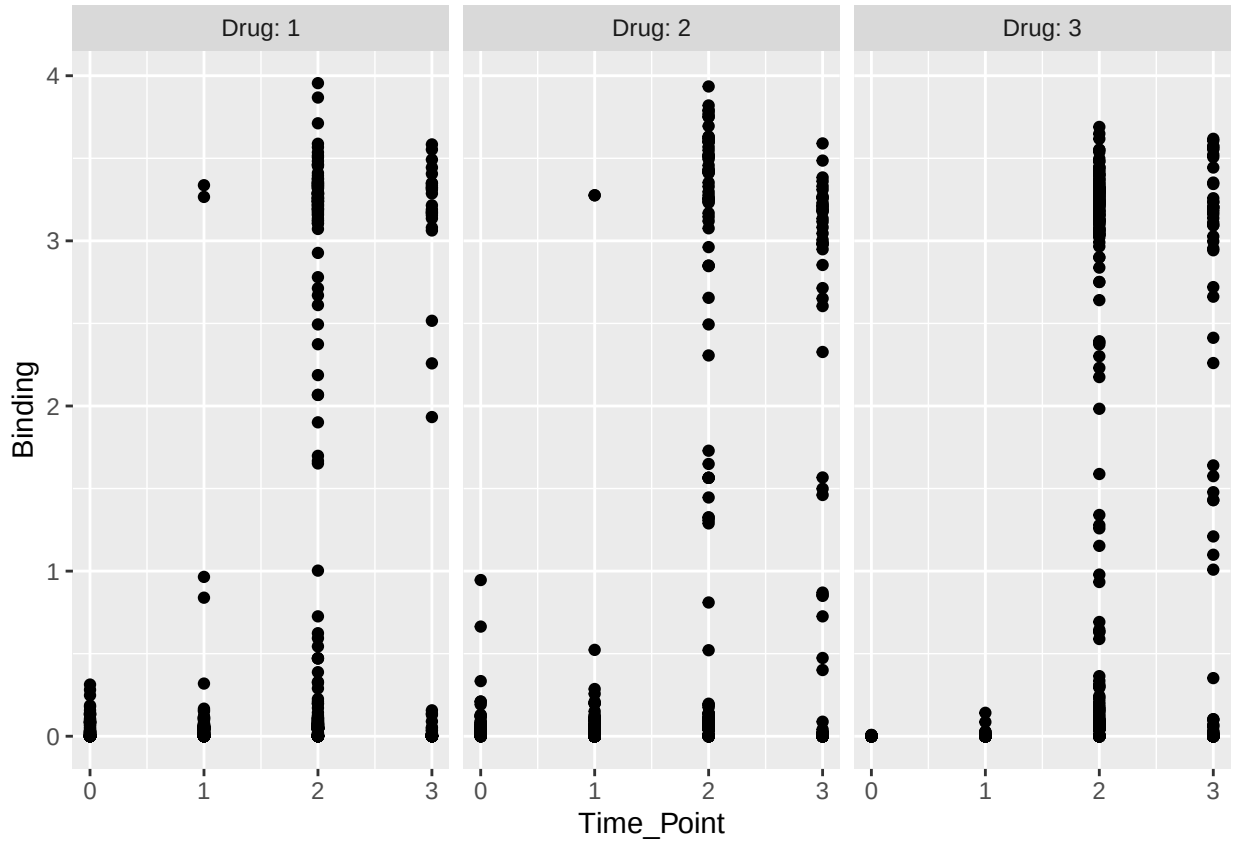


Now we use some plots to see whether the response variables might be different for different time points, treatment groups or drugs, and grouping covariate (`Isotype`).

```
ggplot(Data2, aes(x = Reactivity)) + geom_bar(aes(fill = as.factor(Time_Point)), position = "dodge") + 
```




```
ggplot(Data2, aes(x = Time_Point, y = Binding)) + geom_point() + facet_wrap( ~ Drug, labeller = label_b
```




```
table(Data$Drug, Data$Reactivity)
```

```
##
##       0   1
##   1 680 168
##   2 807 142
##   3 464 204
```

```
table(Data$Time_Point, Data$Reactivity)
```

```
##
##       0   1
##   0 249  24
##   1 971  33
##   2 475 348
##   3 256 109
```

`Binding` or `Reactivity` do seem to be affected by various predictors.
Different drugs appear to have different boxplots in terms of the `Binding` variable.

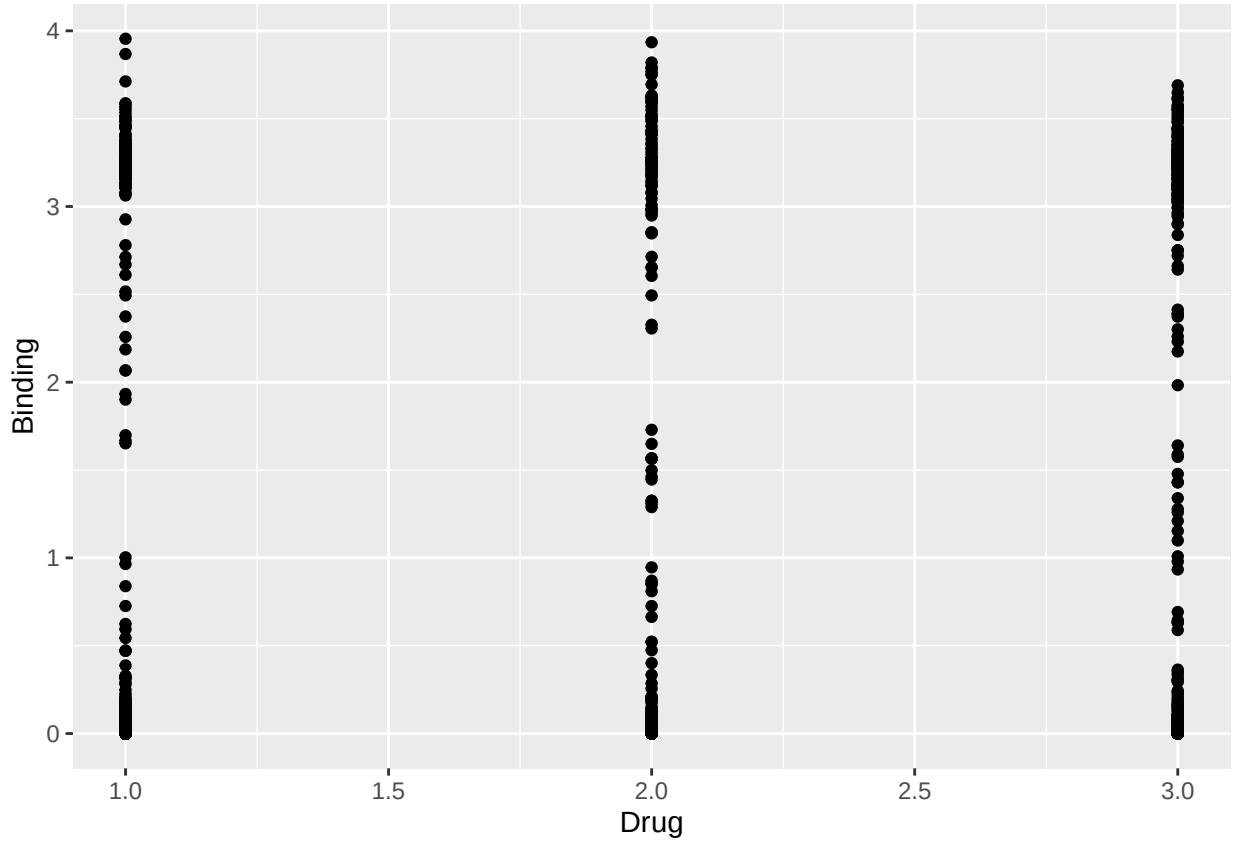
```
Data2 %>% group_by(Drug) %>% summarize(avgLMuFreq = mean(LMuFreq), avgHMuFreq = mean(HMuFreq), avgBindi
```

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

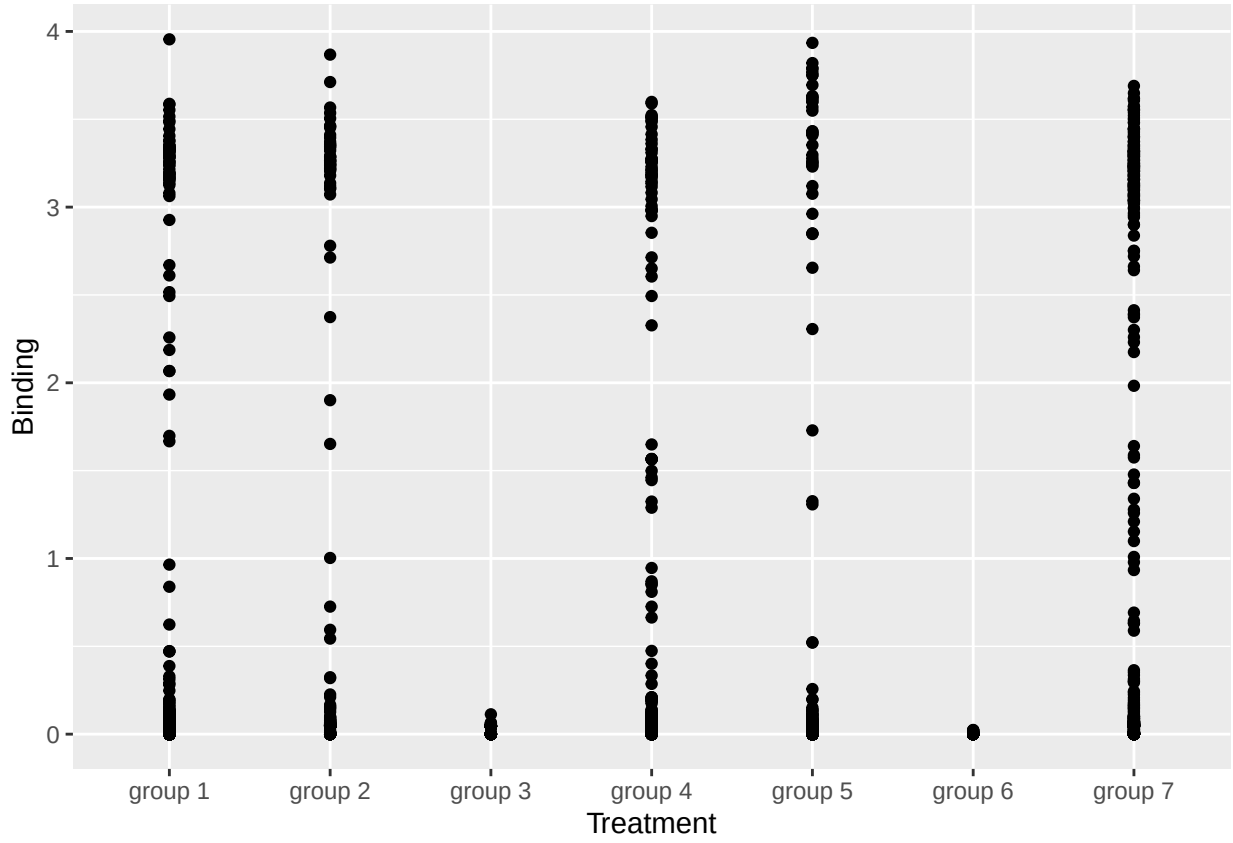
```
## # A tibble: 3 x 6
##    Drug avgLMuFreq avgHMuFreq avgBinding varBinding avgReact
##   <dbl>      <dbl>      <dbl>      <dbl>      <dbl>    <dbl>
## 1     1     0.0616    NA           0.450      1.14     0.198
## 2     2     0.0616     0.0730      0.334      0.864    0.150
```

```
## 3      3      0.0594      0.0559       0.807       1.81      0.305
```

```
ggplot(Data2, aes(x = Drug, y = Binding)) + geom_point()
```




```
ggplot(Data2, aes(x = Treatment, y = Binding)) + geom_point()
```



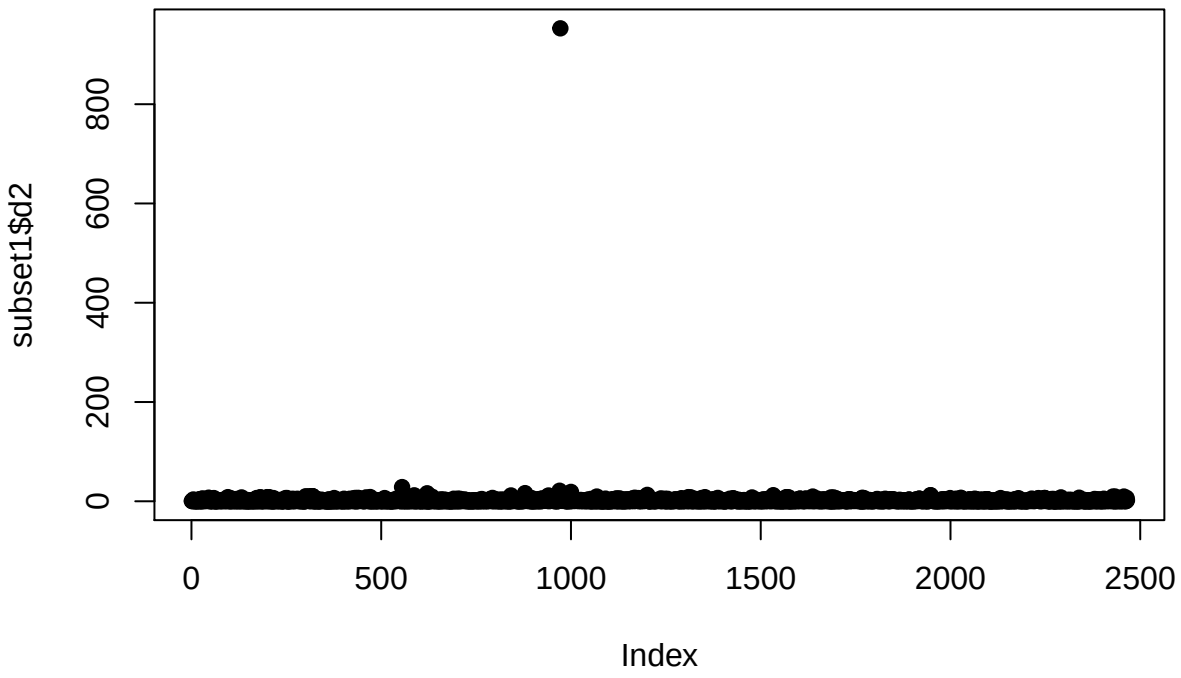
Outlier detection

Before we go into analyses, notice may have outlier in LCDR3 variable.

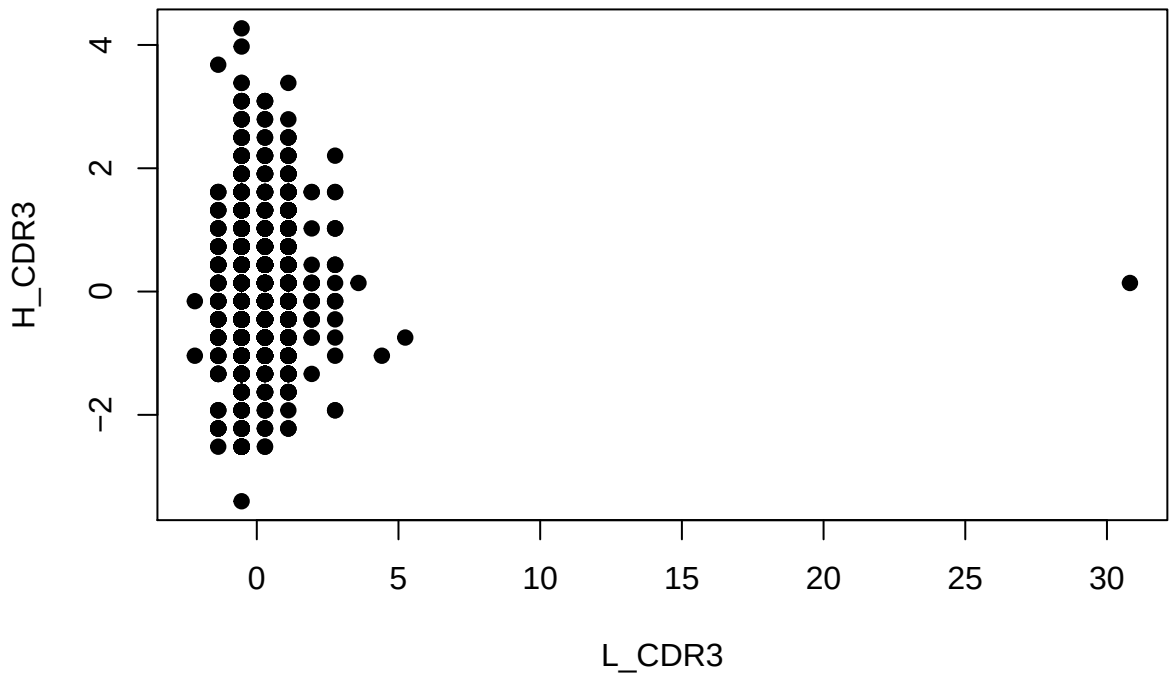
```
summary(Data2$L_CDR3)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    7.00    9.00    9.00    9.65   10.00   47.00
```

```
subset1 <- Data2 %>% select(L_CDR3, H_CDR3)
subset1$d2 <- mahalanobis(subset1, colMeans(subset1), cov(subset1))
subset1$Z <- scale(subset1)
plot(subset1$d2, pch = 19)
```




```
plot(subset1$Z, pch = 19)
```




```
subset2 <- subset1 %>% arrange(desc(d2), desc(Z))
subset2[1,]
```

```
## # A tibble: 1 x 4
##   L_CDR3 H_CDR3    d2 Z[,"L_CDR3"] [,"H_CDR3"] [,"d2"]
##    <dbl>  <dbl> <dbl>        <dbl>       <dbl>   <dbl>
## 1     47     14  953.         30.8       0.139    49.4
```

```
which(subset1$L_CDR3 == 47)
```

```
## [1] 972
```

Row 972 from `Data2` is in fact an outlier, as shown in the summary and plots above. The value for `L_CDR3` is quite unlikely. Since we can't go back to the original data, we remove the data point and will use the new dataset `Data3`.

```
Data2[972,]
```

```
## # A tibble: 1 x 19
##   MonkeyID  Drug Treatment Time_Point Isotype H_VBase H_Substitutions
##      <dbl> <dbl> <chr>          <dbl> <chr>     <dbl>           <dbl>
## 1     6107     3 group 7            2 G           263              16
## # ... with 12 more variables: H_Insertions <dbl>, H_Deletions <dbl>,
## #   HMuFreq <dbl>, H_CDR3 <dbl>, L_VBase <dbl>, L_Substitutions <dbl>,
## #   L_Insertions <dbl>, L_Deletions <dbl>, LMuFreq <dbl>, L_CDR3 <dbl>,
## #   Binding <dbl>, Reactivity <dbl>
```

```
Data3 <- Data2[-972,]
```

Multivariate Data Analysis

```
ID <- as.factor(Data3$MonkeyID)
trt <- as.factor(Data3$Treatment)
drug <- as.factor(Data3$Drug)
tp <- as.factor(Data3$Time_Point)
it <- as.factor(Data3$Isotype)
# four-way manova


fit.manova <- manova(cbind(Data3$L_CDR3, Data3$LMuFreq, Data3$H_CDR3, Data3$HMuFreq, Data3$Binding) ~ d
summary(fit.manova)
```

```
##             Df  Pillai approx F num Df den Df    Pr(>F)
## drug         2 0.06176   15.626     10   4904 < 2.2e-16 ***
## tp           3 0.19800   34.667     15   7359 < 2.2e-16 ***
## Residuals 2455
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
fit.gls <- lm(cbind(Data3$L_CDR3, Data3$LMuFreq, Data3$H_CDR3, Data3$HMuFreq, Data3$Binding) ~ drug + t
summary(fit.gls)
```

```
## Response Data3$L_CDR3 :
##
## Call:
## lm(formula = `Data3$L_CDR3` ~ drug + tp)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.5951 -0.6553 -0.5665  0.4049  6.2942
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  9.65526    0.06178 156.281   <2e-16 ***
## drug2       -0.04384    0.04543  -0.965    0.335
## drug3       -0.11063    0.05064  -2.184    0.029 *
## tp1          0.01725    0.06568   0.263    0.793
## tp2          0.05051    0.06748   0.749    0.454
## tp3          0.02183    0.07682   0.284    0.776
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9508 on 2455 degrees of freedom
##   (3 observations deleted due to missingness)
## Multiple R-squared:  0.002056,   Adjusted R-squared:  2.318e-05
## F-statistic: 1.011 on 5 and 2455 DF,  p-value: 0.4092
##
##
## Response Data3$LMuFreq :
##
## Call:
## lm(formula = `Data3$LMuFreq` ~ drug + tp)
##
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
## -0.07233 -0.03662 -0.01504  0.01987  0.22320
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.0719614  0.0035016  20.551  < 2e-16 ***
## drug2       -0.0001371  0.0025748  -0.053  0.95752
## drug3        0.0003640  0.0028703   0.127  0.89910
## tp1         -0.0084782  0.0037224  -2.278  0.02284 *
## tp2         -0.0147355  0.0038247  -3.853  0.00012 ***
## tp3         -0.0189684  0.0043541  -4.356 1.38e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.05389 on 2455 degrees of freedom
##   (3 observations deleted due to missingness)
## Multiple R-squared:  0.01006,    Adjusted R-squared:  0.008045
## F-statistic:  4.99 on 5 and 2455 DF,  p-value: 0.0001496
##
##
## Response Data3$H_CDR3 :
##
## Call:
## lm(formula = `Data3$H_CDR3` ~ drug + tp)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.2921  -2.1012  -0.1012   1.8988  14.7079
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 13.95536    0.21869  63.813  < 2e-16 ***
## drug2       -0.40087    0.16080  -2.493  0.01273 *
## drug3       -0.57162    0.17926  -3.189  0.00145 **
## tp1         -0.45325    0.23248  -1.950  0.05134 .
## tp2         -0.09164    0.23887  -0.384  0.70128
## tp3          0.69302    0.27193   2.549  0.01088 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.366 on 2455 degrees of freedom
##   (3 observations deleted due to missingness)
## Multiple R-squared:  0.01637,    Adjusted R-squared:  0.01436
## F-statistic:  8.17 on 5 and 2455 DF,  p-value: 1.162e-07
##
##
## Response Data3$HMuFreq :
##
## Call:
## lm(formula = `Data3$HMuFreq` ~ drug + tp)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.085072 -0.029604 -0.006174  0.024701  0.174404
```

```
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.079295   0.002713  29.230  < 2e-16 ***
## drug2        0.005776   0.001995   2.896 0.003816 **
## drug3       -0.008182   0.002224  -3.679 0.000239 ***
## tp1         -0.010457   0.002884  -3.626 0.000294 ***
## tp2         -0.016569   0.002963  -5.592 2.50e-08 ***
## tp3         -0.021684   0.003373  -6.428 1.55e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.04175 on 2455 degrees of freedom
##   (3 observations deleted due to missingness)
## Multiple R-squared:  0.04463,    Adjusted R-squared:  0.04268
## F-statistic: 22.94 on 5 and 2455 DF,  p-value: < 2.2e-16
##
##
## Response Data3$Binding :
##
## Call:
## lm(formula = `Data3$Binding` ~ drug + tp)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.1488 -0.8238 -0.0266  0.0164  3.2987
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.028715   0.065234   0.440    0.660
## drug2       -0.028075   0.047967  -0.585    0.558
## drug3        0.074364   0.053473   1.391    0.164
## tp1          0.009551   0.069347   0.138    0.890
## tp2          1.045709   0.071253  14.676   <2e-16 ***
## tp3          0.751695   0.081115   9.267   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.004 on 2455 degrees of freedom
##   (3 observations deleted due to missingness)
## Multiple R-squared:  0.1975, Adjusted R-squared:  0.1958
## F-statistic: 120.8 on 5 and 2455 DF,  p-value: < 2.2e-16
```

```
fit.logit <- lm(Data3$Reactivity ~ drug*tp)
summary(fit.logit)
```

```
##
## Call:
## lm(formula = Data3$Reactivity ~ drug * tp)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.45954 -0.30282 -0.04624 -0.00800  0.99200
##
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.108527   0.032132   3.378 0.000743 ***
## drug2        0.002584   0.050123   0.052 0.958890
## drug3       -0.108527   0.059152  -1.835 0.066668 .
## tp1         -0.062284   0.037649  -1.654 0.098182 .
## tp2          0.333704   0.039536   8.440  < 2e-16 ***
## tp3          0.112784   0.046089   2.447 0.014471 *
## drug2:tp1   -0.018808   0.056100  -0.335 0.737458
## drug3:tp1    0.070284   0.070352   0.999 0.317874
## drug2:tp2   -0.102593   0.060290  -1.702 0.088949 .
## drug3:tp2    0.125834   0.066442   1.894 0.058357 .
## drug2:tp3    0.162243   0.070163   2.312 0.020839 *
## drug3:tp3    0.190033   0.074355   2.556 0.010655 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.365 on 2452 degrees of freedom
## Multiple R-squared:  0.196,  Adjusted R-squared:  0.1924
## F-statistic: 54.34 on 11 and 2452 DF,  p-value: < 2.2e-16
```

Pairwise comparison

Now we take a look at the pairwise comparison for each treatment group.

```
respMat <- as.matrix(Data3[,c("L_CDR3", "LMuFreq", "H_CDR3", "HMuFreq", "Binding")])
# pairwise comparison among treatment groups
fit1 <- manova(respMat[,1:5] ~ trt)
# summary(fit1)

vars <- c("L_CDR3", "LMuFreq", "H_CDR3", "HMuFreq", "Binding")

p <- 5
q1 <- length(unique(trt))
alpha.old <- 0.05
nc1 <- p*q1*(q1-1)/2
alpha.new1 <- alpha.old/nc1

for (i in 1:5){
  w <- c(0, 0, 0, 0, 0)
  w[i] <- 1
  print(paste(vars[i], " pairwise CI's"))
  cont <- contrast(emmeans(fit1, "trt", weights = w), "pairwise")
  bb <- confint(cont, level = 1 - alpha.new1, adj = "none")
  print(bb)
}
```

```
## [1] "L_CDR3  pairwise CI's"
```

```
## Note: Use 'contrast(regrid(object), ...)' to obtain contrasts of back-transformed estimates
```

```
##  contrast          estimate     SE   df lower.CL upper.CL
##  group 1 - group 2  0.08842 0.0829 2454  -0.2016    0.378
##  group 1 - group 3 -0.03487 0.1047 2454  -0.4012    0.331
##  group 1 - group 4  0.11097 0.0599 2454  -0.0987    0.321
##  group 1 - group 5 -0.01158 0.0630 2454  -0.2321    0.209
##  group 1 - group 6  0.08361 0.0919 2454  -0.2380    0.405
```

```
##  group 1 - group 7  0.11559 0.0540 2454  -0.0732    0.304
##  group 2 - group 3 -0.12328 0.1213 2454  -0.5477    0.301
##  group 2 - group 4  0.02255 0.0857 2454  -0.2773    0.322
##  group 2 - group 5 -0.10000 0.0879 2454  -0.4075    0.208
##  group 2 - group 6 -0.00480 0.1105 2454  -0.3913    0.382
##  group 2 - group 7  0.02717 0.0816 2454  -0.2584    0.313
##  group 3 - group 4  0.14583 0.1070 2454  -0.2283    0.520
##  group 3 - group 5  0.02328 0.1087 2454  -0.3571    0.404
##  group 3 - group 6  0.11848 0.1277 2454  -0.3282    0.565
##  group 3 - group 7  0.15046 0.1037 2454  -0.2124    0.513
##  group 4 - group 5 -0.12255 0.0667 2454  -0.3559    0.111
##  group 4 - group 6 -0.02735 0.0945 2454  -0.3579    0.303
##  group 4 - group 7  0.00462 0.0582 2454  -0.1990    0.208
##  group 5 - group 6  0.09520 0.0965 2454  -0.2423    0.433
##  group 5 - group 7  0.12717 0.0614 2454  -0.0876    0.342
##  group 6 - group 7  0.03198 0.0908 2454  -0.2857    0.350
##
## Results are averaged over the levels of: rep.meas
## Note: contrasts are still on the [.: scale
## Confidence level used: 0.999523809523809
## [1] "LMuFreq  pairwise CI's"

## Note: Use 'contrast(regrid(object), ...)' to obtain contrasts of back-transformed estimates

##  contrast           estimate      SE   df lower.CL upper.CL
##  group 1 - group 2  0.007328 0.00472 2454 -0.00918  0.02384
##  group 1 - group 3 -0.006001 0.00596 2454 -0.02686  0.01486
##  group 1 - group 4  0.002551 0.00341 2454 -0.00939  0.01449
##  group 1 - group 5 -0.000993 0.00359 2454 -0.01355  0.01156
##  group 1 - group 6 -0.001342 0.00523 2454 -0.01966  0.01697
##  group 1 - group 7  0.003112 0.00307 2454 -0.00764  0.01386
##  group 2 - group 3 -0.013330 0.00691 2454 -0.03750  0.01084
##  group 2 - group 4 -0.004778 0.00488 2454 -0.02185  0.01230
##  group 2 - group 5 -0.008322 0.00500 2454 -0.02583  0.00919
##  group 2 - group 6 -0.008670 0.00629 2454 -0.03068  0.01334
##  group 2 - group 7 -0.004216 0.00465 2454 -0.02048  0.01205
##  group 3 - group 4  0.008552 0.00609 2454 -0.01275  0.02986
##  group 3 - group 5  0.005008 0.00619 2454 -0.01665  0.02667
##  group 3 - group 6  0.004659 0.00727 2454 -0.02077  0.03009
##  group 3 - group 7  0.009114 0.00591 2454 -0.01155  0.02978
##  group 4 - group 5 -0.003544 0.00380 2454 -0.01683  0.00974
##  group 4 - group 6 -0.003893 0.00538 2454 -0.02271  0.01493
##  group 4 - group 7  0.000562 0.00331 2454 -0.01103  0.01216
##  group 5 - group 6 -0.000349 0.00549 2454 -0.01957  0.01887
##  group 5 - group 7  0.004106 0.00350 2454 -0.00812  0.01633
##  group 6 - group 7  0.004454 0.00517 2454 -0.01364  0.02254
##
## Results are averaged over the levels of: rep.meas
## Note: contrasts are still on the [.: scale
## Confidence level used: 0.999523809523809
## [1] "H_CDR3  pairwise CI's"

## Note: Use 'contrast(regrid(object), ...)' to obtain contrasts of back-transformed estimates

##  contrast          estimate    SE   df lower.CL upper.CL
##  group 1 - group 2   0.5757 0.295 2454 -0.45488    1.606
```

```
##  group 1 - group 3   0.9285 0.372 2454 -0.37339    2.230
##  group 1 - group 4   0.9693 0.213 2454  0.22406    1.715
##  group 1 - group 5   0.3928 0.224 2454 -0.39095    1.177
##  group 1 - group 6   0.7509 0.327 2454 -0.39208    1.894
##  group 1 - group 7   0.6651 0.192 2454 -0.00594    1.336
##  group 2 - group 3   0.3528 0.431 2454 -1.15543    1.861
##  group 2 - group 4   0.3936 0.305 2454 -0.67188    1.459
##  group 2 - group 5  -0.1829 0.312 2454 -1.27567    0.910
##  group 2 - group 6   0.1753 0.393 2454 -1.19820    1.549
##  group 2 - group 7   0.0894 0.290 2454 -0.92557    1.104
##  group 3 - group 4   0.0408 0.380 2454 -1.28891    1.371
##  group 3 - group 5  -0.5357 0.386 2454 -1.88738    0.816
##  group 3 - group 6  -0.1776 0.454 2454 -1.76477    1.410
##  group 3 - group 7  -0.2634 0.369 2454 -1.55299    1.026
##  group 4 - group 5  -0.5765 0.237 2454 -1.40570    0.253
##  group 4 - group 6  -0.2184 0.336 2454 -1.39301    0.956
##  group 4 - group 7  -0.3042 0.207 2454 -1.02780    0.419
##  group 5 - group 6   0.3581 0.343 2454 -0.84126    1.558
##  group 5 - group 7   0.2723 0.218 2454 -0.49084    1.035
##  group 6 - group 7  -0.0858 0.323 2454 -1.21483    1.043
##
## Results are averaged over the levels of: rep.meas
## Note: contrasts are still on the [.: scale
## Confidence level used: 0.999523809523809
## [1] "HMuFreq  pairwise CI's"

## Note: Use 'contrast(regrid(object), ...)' to obtain contrasts of back-transformed estimates

##  contrast           estimate      SE   df lower.CL  upper.CL
##  group 1 - group 2  0.001782 0.00367 2454 -0.01105  0.014613
##  group 1 - group 3  0.014768 0.00463 2454 -0.00144  0.030977
##  group 1 - group 4 -0.000904 0.00265 2454 -0.01018  0.008375
##  group 1 - group 5 -0.005563 0.00279 2454 -0.01532  0.004195
##  group 1 - group 6 -0.009779 0.00407 2454 -0.02401  0.004452
##  group 1 - group 7  0.013055 0.00239 2454  0.00470  0.021410
##  group 2 - group 3  0.012986 0.00537 2454 -0.00579  0.031765
##  group 2 - group 4 -0.002686 0.00379 2454 -0.01595  0.010580
##  group 2 - group 5 -0.007345 0.00389 2454 -0.02095  0.006261
##  group 2 - group 6 -0.011561 0.00489 2454 -0.02866  0.005539
##  group 2 - group 7  0.011273 0.00361 2454 -0.00136  0.023910
##  group 3 - group 4 -0.015672 0.00473 2454 -0.03223  0.000884
##  group 3 - group 5 -0.020331 0.00481 2454 -0.03716 -0.003502
##  group 3 - group 6 -0.024547 0.00565 2454 -0.04431 -0.004785
##  group 3 - group 7 -0.001713 0.00459 2454 -0.01777  0.014343
##  group 4 - group 5 -0.004659 0.00295 2454 -0.01498  0.005665
##  group 4 - group 6 -0.008875 0.00418 2454 -0.02350  0.005750
##  group 4 - group 7  0.013959 0.00258 2454  0.00495  0.022968
##  group 5 - group 6 -0.004216 0.00427 2454 -0.01915  0.010717
##  group 5 - group 7  0.018618 0.00272 2454  0.00912  0.028120
##  group 6 - group 7  0.022834 0.00402 2454  0.00878  0.036891
##
## Results are averaged over the levels of: rep.meas
## Note: contrasts are still on the [.: scale
## Confidence level used: 0.999523809523809
## [1] "Binding  pairwise CI's"
```

```
## Note: Use 'contrast(regrid(object), ...)' to obtain contrasts of back-transformed estimates
```

```
##  contrast          estimate     SE   df lower.CL upper.CL
##  group 1 - group 2  -0.5394 0.0951 2454  -0.8721  -0.2066
##  group 1 - group 3   0.3560 0.1202 2454  -0.0644   0.7763
##  group 1 - group 4  -0.0201 0.0688 2454  -0.2608   0.2205
##  group 1 - group 5   0.0164 0.0723 2454  -0.2366   0.2695
##  group 1 - group 6   0.3819 0.1055 2454   0.0129   0.7510
##  group 1 - group 7  -0.4236 0.0619 2454  -0.6403  -0.2070
##  group 2 - group 3   0.8953 0.1392 2454   0.4083   1.3823
##  group 2 - group 4   0.5192 0.0983 2454   0.1752   0.8633
##  group 2 - group 5   0.5558 0.1009 2454   0.2029   0.9086
##  group 2 - group 6   0.9213 0.1268 2454   0.4778   1.3648
##  group 2 - group 7   0.1157 0.0937 2454  -0.2120   0.4434
##  group 3 - group 4  -0.3761 0.1227 2454  -0.8055   0.0532
##  group 3 - group 5  -0.3395 0.1248 2454  -0.7760   0.0969
##  group 3 - group 6   0.0259 0.1465 2454  -0.4866   0.5384
##  group 3 - group 7  -0.7796 0.1190 2454  -1.1960  -0.3632
##  group 4 - group 5   0.0366 0.0765 2454  -0.2311   0.3043
##  group 4 - group 6   0.4021 0.1084 2454   0.0228   0.7813
##  group 4 - group 7  -0.4035 0.0668 2454  -0.6371  -0.1699
##  group 5 - group 6   0.3655 0.1107 2454  -0.0218   0.7528
##  group 5 - group 7  -0.4401 0.0704 2454  -0.6865  -0.1937
##  group 6 - group 7  -0.8056 0.1042 2454  -1.1701  -0.4410
##
## Results are averaged over the levels of: rep.meas
## Note: contrasts are still on the [.: scale
## Confidence level used: 0.999523809523809
```

```
# pairwise comparison among drug groups
fit2 <- manova(respMat[,1:5] ~ drug)
# summary(fit2)
p <- 5
q2 <- length(unique(drug))
alpha.old <- 0.05
nc2 <- p*q2*(q2-1)/2
alpha.new2 <- alpha.old/nc2

for (i in 1:5){
  w <- c(0, 0, 0, 0, 0)
  w[i] <- 1
  print(paste(vars[i], " pairwise CI's"))
  cont <- contrast(emmeans(fit2, "drug", weights = w), "pairwise")
  bb <- confint(cont, level = 1 - alpha.new2, adj = "none")
  print(bb)
}
```

```
## [1] "L_CDR3  pairwise CI's"
```

```
## Note: Use 'contrast(regrid(object), ...)' to obtain contrasts of back-transformed estimates
```

```
##  contrast estimate     SE   df lower.CL upper.CL
##  1 - 2      0.0451 0.0450 2458  -0.0870    0.177
##  1 - 3      0.1018 0.0492 2458  -0.0429    0.246
##  2 - 3      0.0567 0.0480 2458  -0.0844    0.198
##
```

```
## Results are averaged over the levels of: rep.meas
## Note: contrasts are still on the [.: scale
## Confidence level used: 0.996666666666667
## [1] "LMuFreq  pairwise CI's"
```

```
## Note: Use 'contrast(regrid(object), ...)' to obtain contrasts of back-transformed estimates
```

```
##  contrast  estimate      SE   df lower.CL upper.CL
##  1 - 2    -0.000176 0.00256 2458 -0.00770  0.00735
##  1 - 3     0.002320 0.00280 2458 -0.00592  0.01056
##  2 - 3     0.002496 0.00273 2458 -0.00554  0.01053
##
## Results are averaged over the levels of: rep.meas
## Note: contrasts are still on the [.: scale
## Confidence level used: 0.996666666666667
## [1] "H_CDR3  pairwise CI's"
```

```
## Note: Use 'contrast(regrid(object), ...)' to obtain contrasts of back-transformed estimates
```

```
##  contrast estimate    SE   df lower.CL upper.CL
##  1 - 2      0.4907 0.160 2458   0.0204    0.961
##  1 - 3      0.4438 0.175 2458  -0.0712    0.959
##  2 - 3     -0.0469 0.171 2458  -0.5492    0.456
##
## Results are averaged over the levels of: rep.meas
## Note: contrasts are still on the [.: scale
## Confidence level used: 0.996666666666667
## [1] "HMuFreq  pairwise CI's"
```

```
## Note: Use 'contrast(regrid(object), ...)' to obtain contrasts of back-transformed estimates
```

```
##  contrast estimate      SE   df lower.CL  upper.CL
##  1 - 2      -0.006 0.00199 2458 -0.01186 -0.000145
##  1 - 3       0.011 0.00218 2458  0.00461  0.017432
##  2 - 3       0.017 0.00213 2458  0.01076  0.023276
##
## Results are averaged over the levels of: rep.meas
## Note: contrasts are still on the [.: scale
## Confidence level used: 0.996666666666667
## [1] "Binding  pairwise CI's"
```

```
## Note: Use 'contrast(regrid(object), ...)' to obtain contrasts of back-transformed estimates
```

```
##  contrast estimate     SE   df lower.CL upper.CL
##  1 - 2       0.118 0.0522 2458  -0.0355    0.271
##  1 - 3      -0.356 0.0571 2458  -0.5235   -0.188
##  2 - 3      -0.473 0.0557 2458  -0.6372   -0.310
##
## Results are averaged over the levels of: rep.meas
## Note: contrasts are still on the [.: scale
## Confidence level used: 0.996666666666667
```

```
# pairwise comparison among time point
fit3 <- manova(respMat[,1:5] ~ tp)
# summary(fit3)
p <- 5
q3 <- length(unique(tp))
alpha.old <- 0.05
```

```r
nc3 <- p*q3*(q3-1)/2
alpha.new3 <- alpha.old/nc3

for (i in 1:5){
  w <- c(0, 0, 0, 0, 0)
  w[i] <- 1
  print(paste(vars[i], " pairwise CI's"))
  cont <- contrast(emmeans(fit3, "tp", weights = w), "pairwise")
  bb <- confint(cont, level = 1 - alpha.new3, adj = "none")
  print(bb)
}
```

```
## [1] "L_CDR3  pairwise CI's"
```

```
## Note: Use 'contrast(regrid(object), ...)' to obtain contrasts of back-transformed estimates
```

```
##  contrast estimate     SE   df lower.CL upper.CL
##  0 - 1     -0.0169 0.0652 2457   -0.222    0.188
##  0 - 2     -0.0287 0.0667 2457   -0.239    0.181
##  0 - 3     -0.0034 0.0764 2457   -0.244    0.237
##  1 - 2     -0.0117 0.0447 2457   -0.153    0.129
##  1 - 3      0.0135 0.0581 2457   -0.169    0.197
##  2 - 3      0.0253 0.0598 2457   -0.163    0.214
##
## Results are averaged over the levels of: rep.meas
## Note: contrasts are still on the [.: scale
## Confidence level used: 0.998333333333333
## [1] "LMuFreq  pairwise CI's"
```

```
## Note: Use 'contrast(regrid(object), ...)' to obtain contrasts of back-transformed estimates
```

```
##  contrast estimate      SE   df  lower.CL upper.CL
##  0 - 1     0.00853 0.00369 2457 -3.09e-03   0.0202
##  0 - 2     0.01465 0.00378 2457  2.75e-03   0.0265
##  0 - 3     0.01889 0.00432 2457  5.28e-03   0.0325
##  1 - 2     0.00611 0.00253 2457 -1.86e-03   0.0141
##  1 - 3     0.01036 0.00329 2457 -4.12e-06   0.0207
##  2 - 3     0.00424 0.00339 2457 -6.42e-03   0.0149
##
## Results are averaged over the levels of: rep.meas
## Note: contrasts are still on the [.: scale
## Confidence level used: 0.998333333333333
## [1] "H_CDR3  pairwise CI's"
```

```
## Note: Use 'contrast(regrid(object), ...)' to obtain contrasts of back-transformed estimates
```

```
##  contrast estimate    SE   df lower.CL upper.CL
##  0 - 1       0.489 0.231 2457   -0.238    1.217
##  0 - 2       0.194 0.237 2457   -0.550    0.939
##  0 - 3      -0.608 0.271 2457   -1.460    0.244
##  1 - 2      -0.295 0.159 2457   -0.795    0.204
##  1 - 3      -1.097 0.206 2457   -1.746   -0.448
##  2 - 3      -0.802 0.212 2457   -1.469   -0.134
##
## Results are averaged over the levels of: rep.meas
## Note: contrasts are still on the [.: scale
## Confidence level used: 0.998333333333333
```

```
## [1] "HMuFreq  pairwise CI's"
```

```
## Note: Use 'contrast(regrid(object), ...)' to obtain contrasts of back-transformed estimates
```

```
##  contrast estimate      SE   df  lower.CL upper.CL
##  0 - 1     0.00870 0.00288 2457 -0.000379   0.0178
##  0 - 2     0.01872 0.00295 2457  0.009434   0.0280
##  0 - 3     0.02356 0.00338 2457  0.012930   0.0342
##  1 - 2     0.01002 0.00198 2457  0.003796   0.0163
##  1 - 3     0.01486 0.00257 2457  0.006768   0.0230
##  2 - 3     0.00484 0.00265 2457 -0.003491   0.0132
##
## Results are averaged over the levels of: rep.meas
## Note: contrasts are still on the [.: scale
## Confidence level used: 0.998333333333333
## [1] "Binding  pairwise CI's"
```

```
## Note: Use 'contrast(regrid(object), ...)' to obtain contrasts of back-transformed estimates
```

```
##  contrast estimate     SE   df lower.CL upper.CL
##  0 - 1     0.00161 0.0689 2457  -0.2151    0.218
##  0 - 2    -1.06381 0.0704 2457  -1.2855   -0.842
##  0 - 3    -0.76734 0.0806 2457  -1.0211   -0.514
##  1 - 2    -1.06542 0.0472 2457  -1.2141   -0.917
##  1 - 3    -0.76895 0.0614 2457  -0.9622   -0.576
##  2 - 3     0.29647 0.0632 2457   0.0976    0.495
##
## Results are averaged over the levels of: rep.meas
## Note: contrasts are still on the [.: scale
## Confidence level used: 0.998333333333333
```

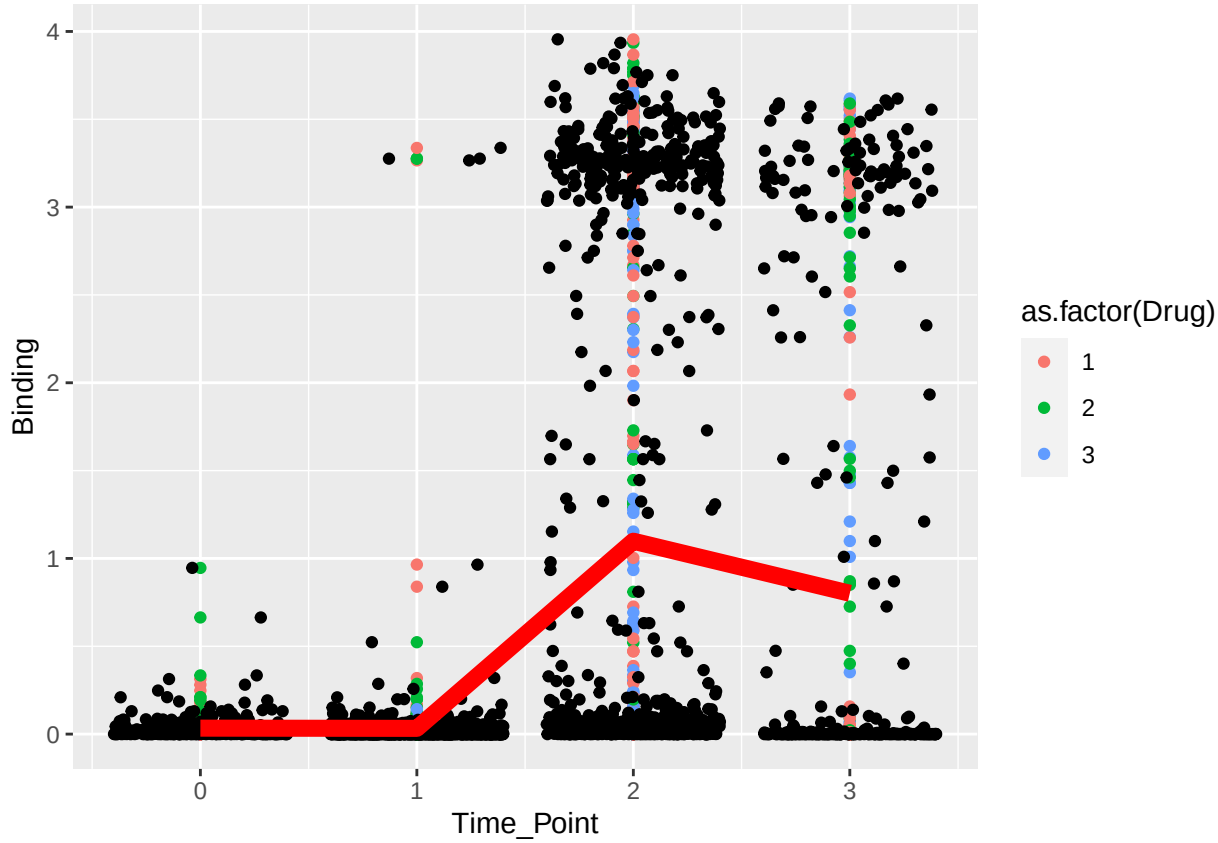
Longitudinal Data Analysis

- check assumption of compound symmetry or equal variance

First we don't consider treatments but only plot the mean trend over time.

$$Y_{ij} = \beta_0 + \beta_1 Time_{ij} + e_{ij}$$

```
# simply connects the mean of each time point
ggplot(Data3, aes(x = Time_Point, y = Binding)) + geom_point(aes(color = as.factor(Drug))) + geom_jitte
```



Here we use Binding as the only response. Prdictors: `Drug`.
Random effect for both intercept and slope.

$$Y_{ij} = \beta_0 + \beta_1 t_{ij} + b_{0i} + b_{1i} + e_{ij}$$

```
lda <- lme(fixed = Binding ~ Time_Point + Drug,
           random = ~ Time_Point | MonkeyID, data = Data3, method = "REML")
summary(lda)
```

```
## Linear mixed-effects model fit by REML
##  Data: Data3
##        AIC      BIC    logLik
##   6738.894 6779.552 -3362.447
##
## Random effects:
##  Formula: ~Time_Point | MonkeyID
##  Structure: General positive-definite, Log-Cholesky parametrization
##             StdDev    Corr
## (Intercept) 0.6043473 (Intr)
## Time_Point  0.5938004 -0.97
## Residual    0.9330025
##
## Fixed effects: Binding ~ Time_Point + Drug
##                  Value  Std.Error   DF   t-value p-value
## (Intercept) -0.5671081 0.21736533 2443 -2.609009  0.0091
## Time_Point   0.6404275 0.18256794 2443  3.507886  0.0005
```

```
## Drug         0.0414165 0.05945489   18  0.696604  0.4949
##  Correlation:
##            (Intr) Tm_Pnt
## Time_Point -0.847
## Drug       -0.499  0.007
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -3.52231270 -0.51818636 -0.09479562  0.01991047  3.61431329
##
## Number of Observations: 2464
## Number of Groups: 20
```

List of variables

- Treatment: Treatment A is the mock control, and treatment B and C are two different kinds of Treg inhibitor treatments.
- Time Points: 0 represents before immunization; 1 represents 2 weeks post 1st immunization; 2 represents 2 weeks post 2nd immunization; and 3 represents 2 weeks post 3rd immunization, respectively.
- Isotype: the category of antibody type
- H_ID and L_ID: heavy chain and light chain IDs for the particular observation
- H_V, H_D and H_J: the gene segments used in heavy chain VDJ recombination in that antibody. The same applies to L_V and L_J
- H_VBase: the number of nucleotide of the heavy chain variable region
- H_Substitutions, H_Insertions, H_Deletions: the number of relative nucleotide mutations.
- HMuFreq: calculated by H_Substitutions / H_VBase
- H_CDR3: the number of amino acid of the heavy chain's third complementarity determining region
- Binding: affinity of antibodies against a selected HIV glycoprotein. The larger value indicates stronger binding

Reference

The dataset, which can be found here, was provided by Kan Luo, as he was one of authors for the following four publications that used the dataset:

1. Luo K, Liao HX, Zhang R, et al. Tissue memory B cell repertoire analysis after ALVAC/AIDSVAX B/E gp120 immunization of rhesus macaques. *JCI Insight*. 2016;1(20):e88522. Published 2016 Dec 8. doi:10.1172/jci.insight.88522
2. Bradley, T., Kuraoka, M., Yeh, C.-H., Tian, M., Chen, H., Cain, D. W., . . . Haynes, B. F. (2020). Immune checkpoint modulation enhances HIV-1 antibody induction. *Nature Communications*, 11(1), 948. doi:10.1038/s41467-020-14670-w
3. Easterhoff, D., Pollara, J., Luo, K., Tolbert, W. D., Young, B., Mielke, D., . . . Ferrari, G. (2020). Boosting with AIDSVAX B/E Enhances Env Constant Region 1 and 2 Antibody-Dependent Cellular Cytotoxicity Breadth and Potency. *Journal of Virology*, 94(4), e01120-01119. doi:10.1128/jvi.01120-19
4. Wiehe, K., Easterhoff, D., Luo, K., Nicely, N. I., Bradley, T., Jaeger, F. H., Dennison, S. M., Zhang, R., Lloyd, K. E., Stolarchuk, C., Parks, R., Sutherland, L. L., Scearce, R. M., Morris, L., Kaewkungwal, J., Nitayaphan, S., Pitisuttithum, P., Rerks-Ngarm, S., Sinangil, F., Phogat, S., ... Haynes, B. F. (2014). Antibody light-chain-restricted recognition of the site of immune pressure in the RV144 HIV-1 vaccine trial is phylogenetically conserved. *Immunity*, 41(6), 909–918. https://doi.org/10.1016/j.immuni.2014.11.014