Final Project: First Draft

Kan Luo, Shih-Ni Prim

10/20/2020

Contents

Questions for Kan

- Why are there 7 treatment groups now?
-

Exploratory Data Analysis

Data preparation

```
## Response [https://raw.githubusercontent.com/luokan1227/537P1/master/Data.xlsx]
##   Date: 2020-10-22 05:47
##   Status: 200
##   Content-Type: application/octet-stream
##   Size: 341 kB
## <ON DISK>  C:\Users\shihn\AppData\Local\Temp\RtmpQ1uFXd\file108228a7b21.xlsx

## Response [https://raw.githubusercontent.com/luokan1227/537P1/master/MonkeyID.xlsx]
##   Date: 2020-10-22 05:47
##   Status: 200
##   Content-Type: application/octet-stream
##   Size: 50.1 kB
## <ON DISK>  C:\Users\shihn\AppData\Local\Temp\RtmpQ1uFXd\file10867bd2df3.xlsx

##
##    1    2    3    4    5    6    7
##  117  142 1045 1104   25   24    8
```

```
## 
##       1   2   3   4   5   6   7
##    3 127 917 352 303 148 474 138

## 
##     1    2    3    4    5    6
##   130   64   90 1368  468  345

## 
##    1  10  11   2   3   4   5   6   7   9
## 890  43  15 800 450  39  84  47  93   4

## 
##     1    2    3    4    5    6
##   674 1099  220  385    1   86
```

Two datasets for analysis:

- `Data`: Kan has been using this one.
- `Data2`: Shih-Ni created this subset, which removed some ID info that we won't use and added extracted information from antibodies.

Contingency Tables

```
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
```

```
table(Data2$Time_Point)
```

```
## 
##    0    1    2    3
##  273 1004  823  365
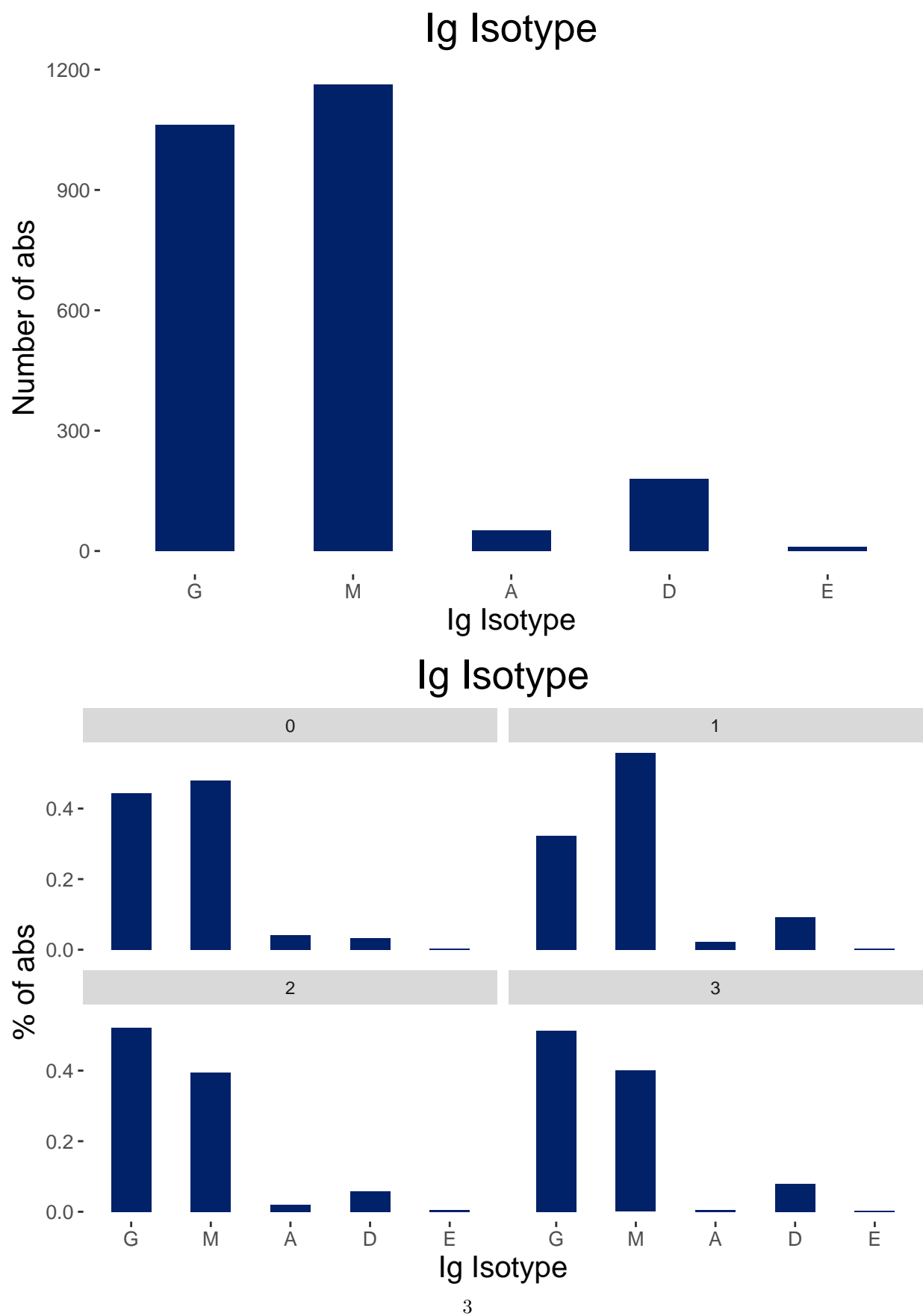```

```
table(Data2$Treatment)
```

```
## 
## group 1 group 2 group 3 group 4 group 5 group 6 group 7
##     582     170      96     444     374     131     668
```
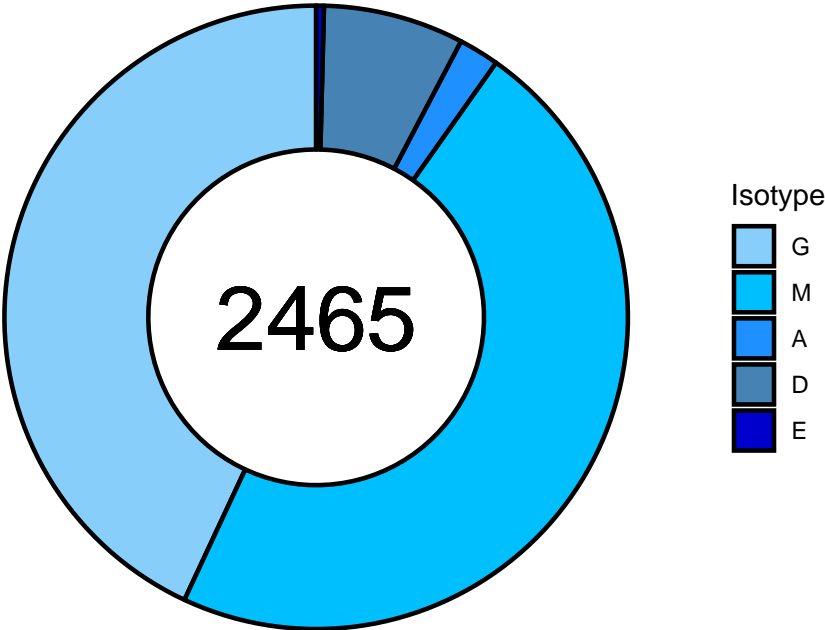
```
table(Data2$Time_Point, Data2$Treatment)
```

```
## 
##     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
```
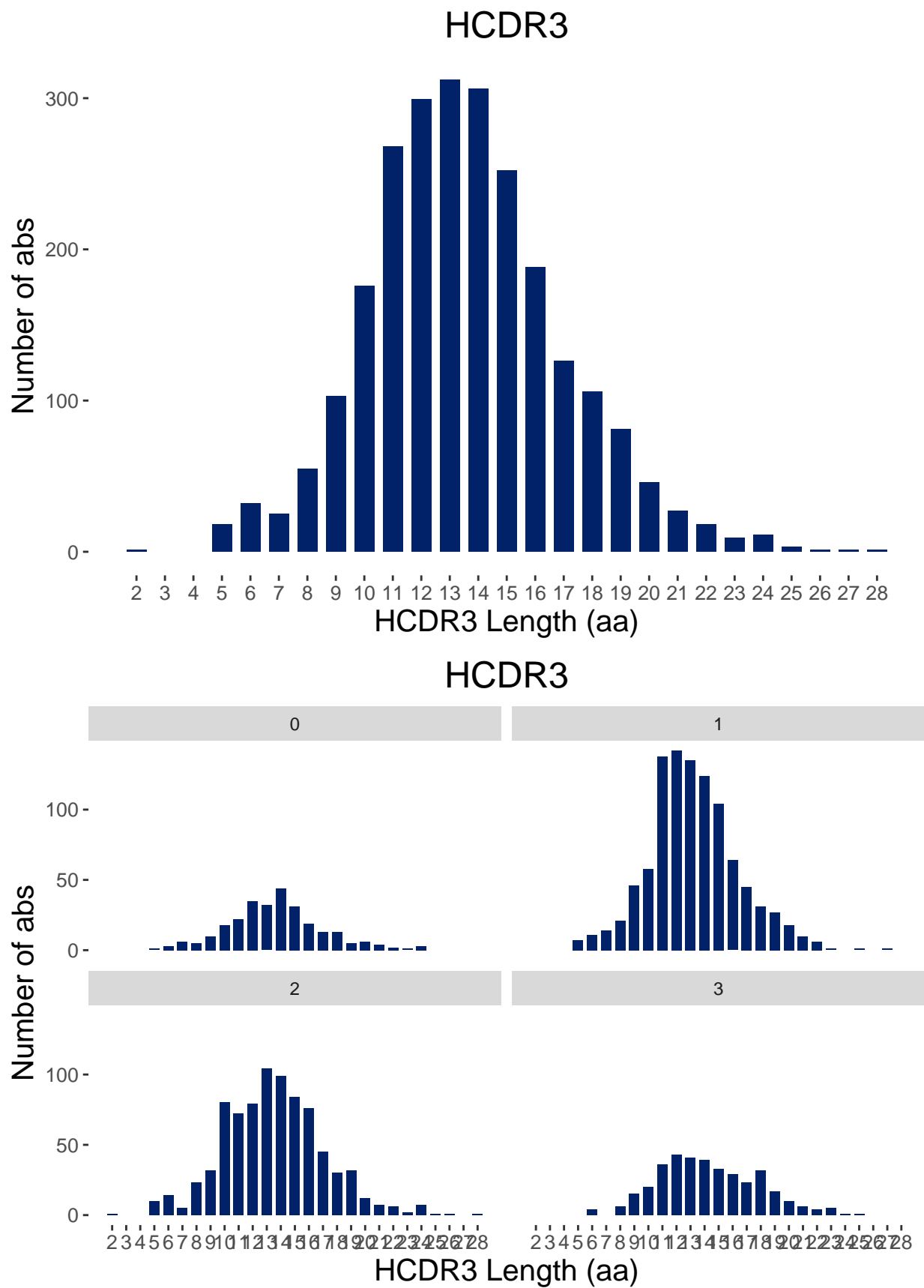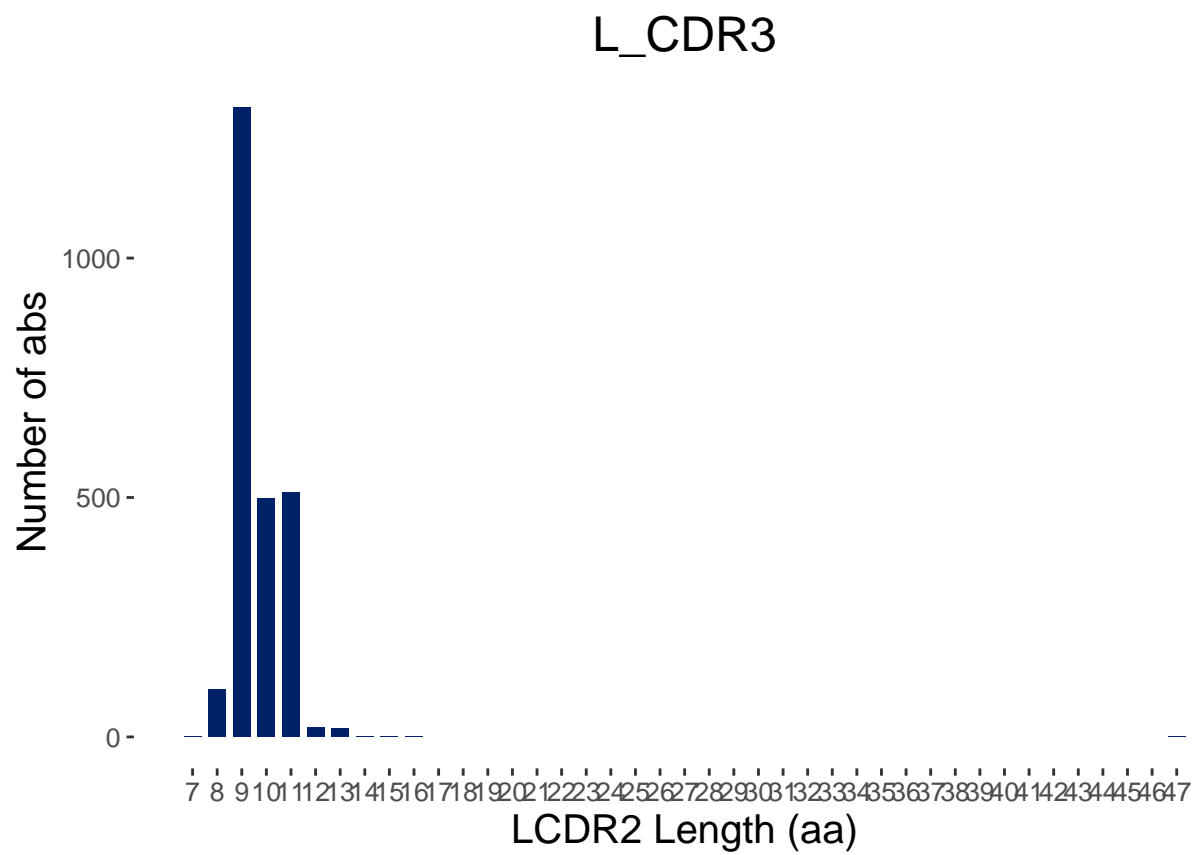
Isotype Plots and Table

Ig Isotype

Ig Isotype

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
```
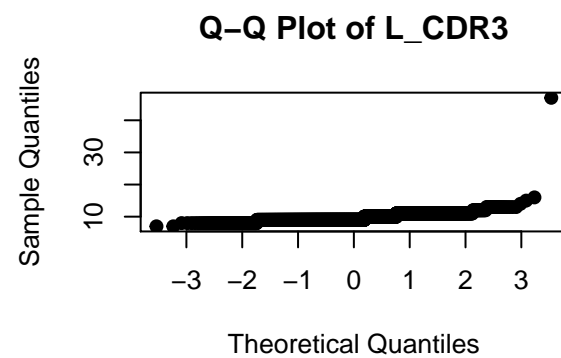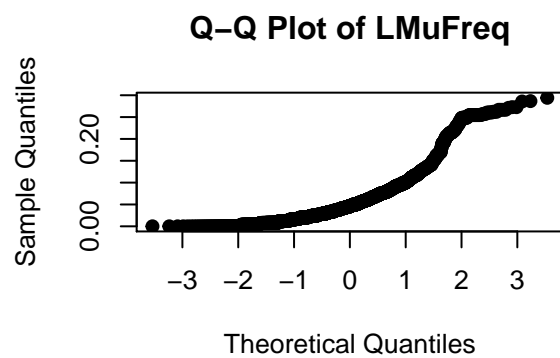
HCDR3

HCDR3

L_CDR3

Q–Q Plot of HMuFreq

Q–Q Plot of H_CDR3

Q–Q Plot of LMuFreq

Q–Q Plot of L_CDR3
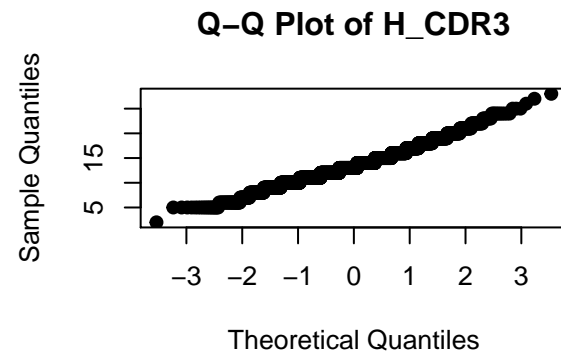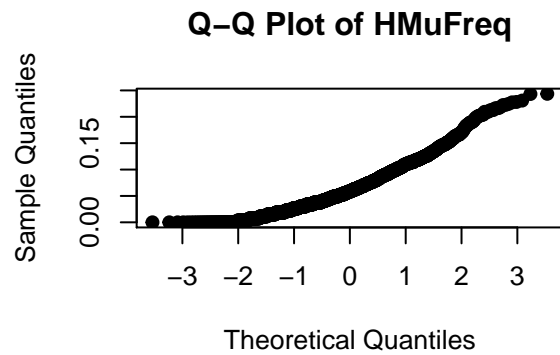
Outlier detection

[Need to add more]
Notice may have outlier in LCDR3 variable:
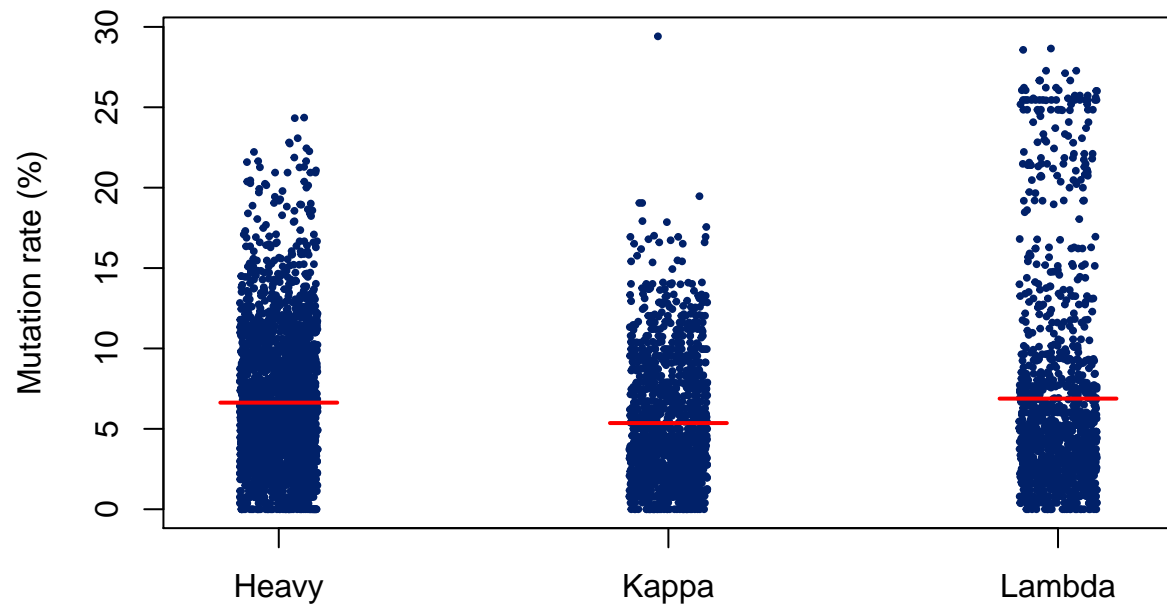
```
summary(Data$L_CDR3)
```

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

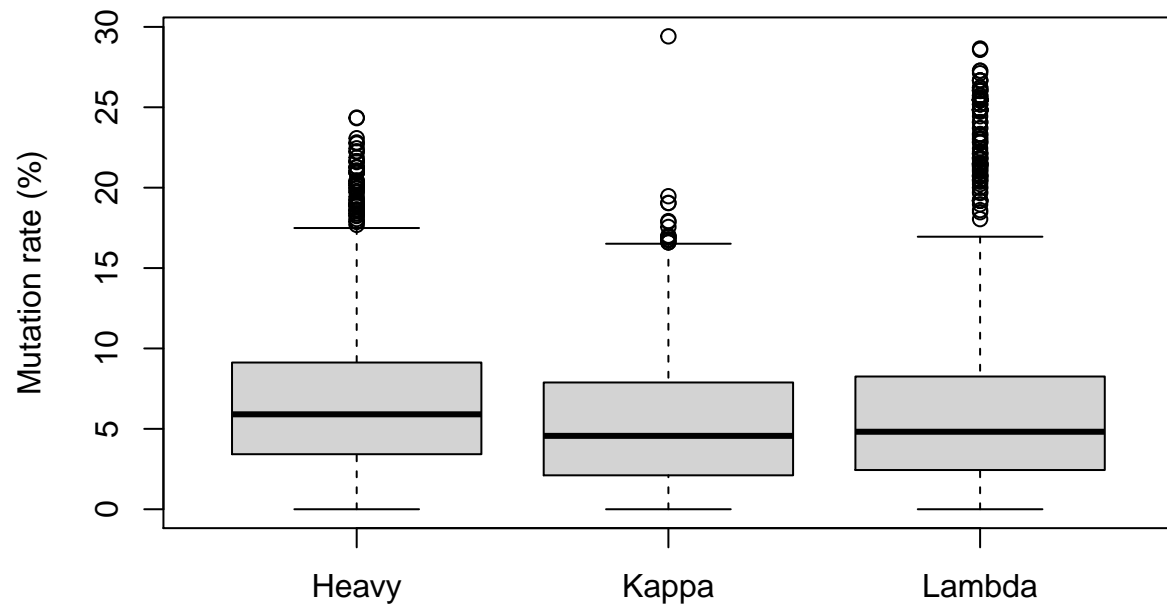## Mutation Rate

```
##         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

H/K/L mutation rate

H_Mutation
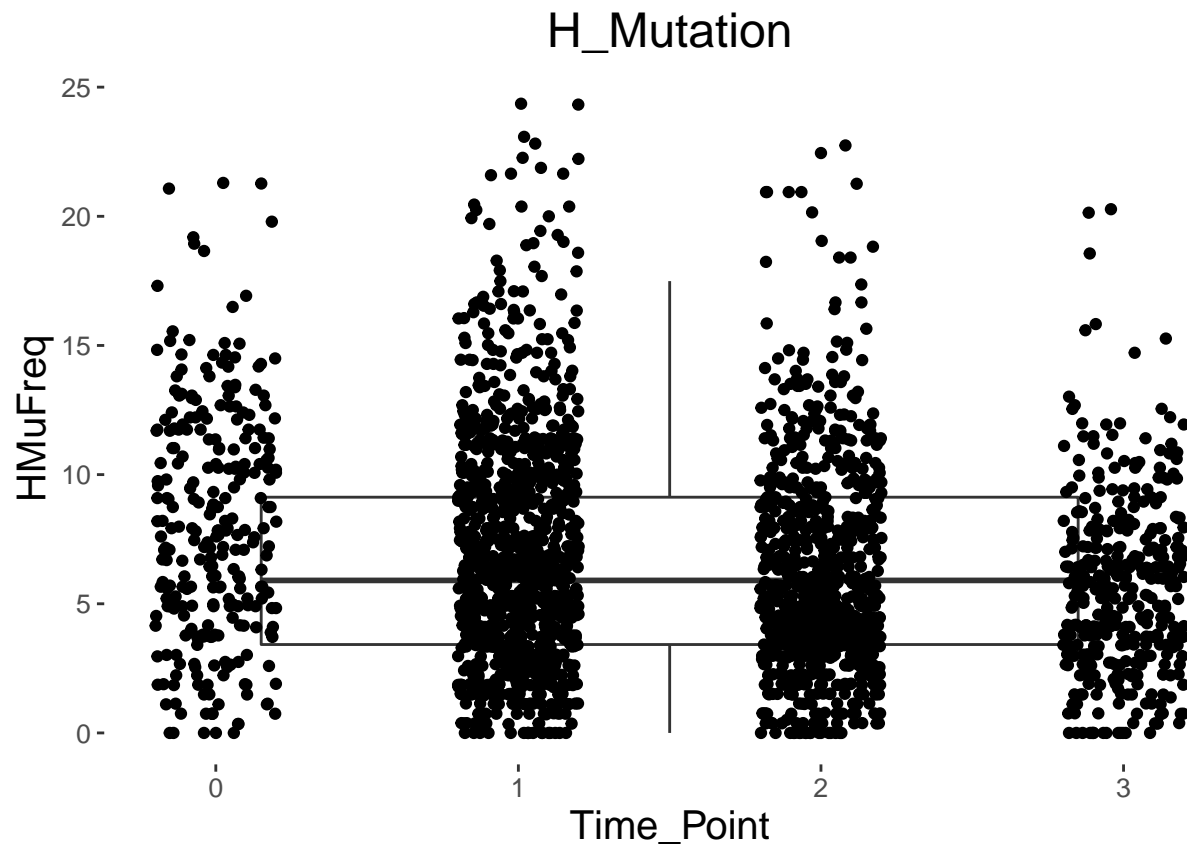
Test independence assumption

```
subset1 <- Data2 %>% filter(Time_Point == 1) %>% filter(MonkeyID == 6132) %>% select(HMuFreq, LMuFreq, 
```

```
subset1 <- Data2 %>% filter(Time_Point == 1) %>% select(HMuFreq, LMuFreq, H_CDR3, L_CDR3) %>% pairs()
```
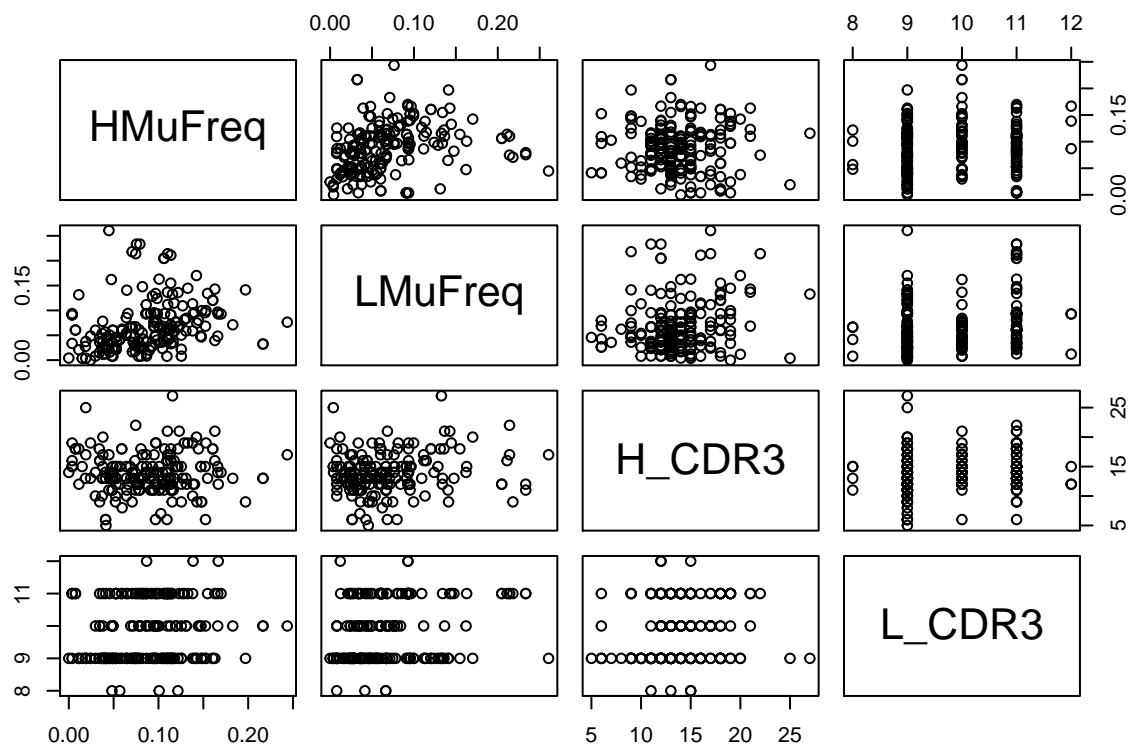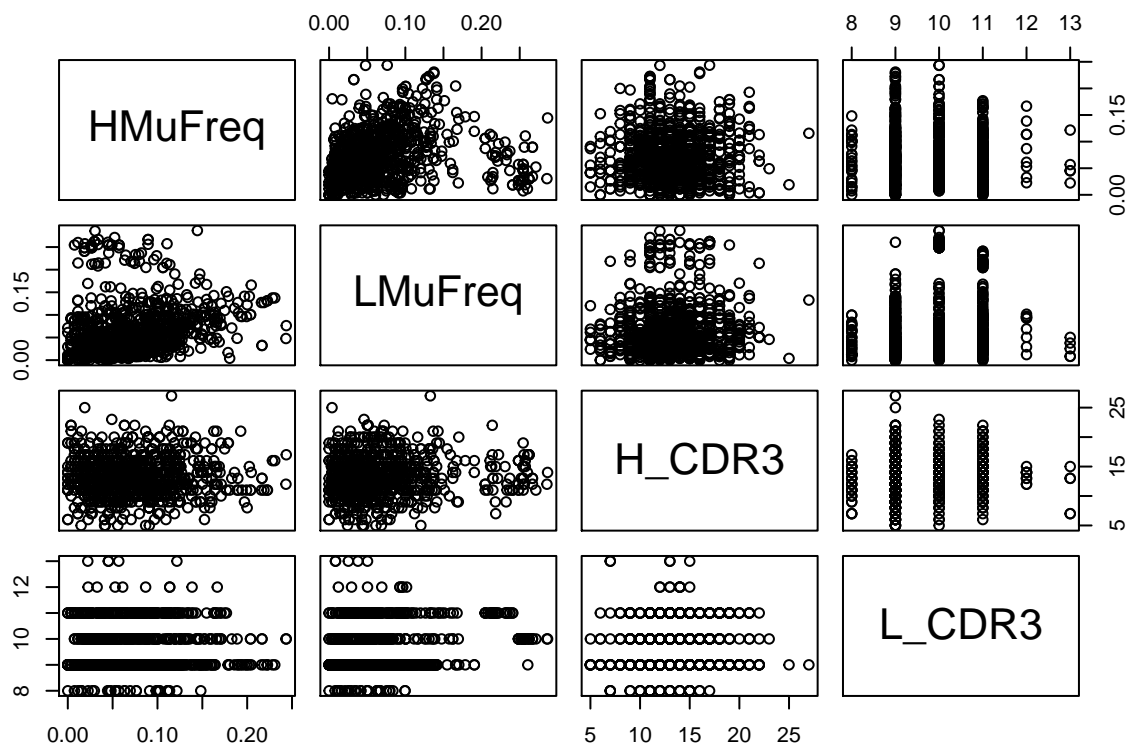
Multivariate Data Analysis

```
ID <- as.factor(Data2$MonkeyID)
trt <- as.factor(Data2$Treatment)
tp <- as.factor(Data2$Time_Point)
it <- as.factor(Data2$Isotype)
fit.manova <- manova(cbind(Data2$L_CDR3, Data2$LMuFreq, Data2$H_CDR3, Data2$HMuFreq) ~ ID + trt+ tp + i
summary(fit.manova)
```

```
##              Df   Pillai approx F num Df den Df    Pr(>F)
## ID           19 0.126901   4.1991     76   9740 < 2.2e-16 ***
## tp            3 0.045931   9.4611     12   7302 < 2.2e-16 ***
## it            4 0.087622  13.6336     16   9740 < 2.2e-16 ***
## Residuals  2435
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Longitudinal Data Analysis

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

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

```
meanTrend <- lm(Data2$Binding ~ Data2$Time_Point)
summary(meanTrend)

##
## Call:
## lm(formula = Data2$Binding ~ Data2$Time_Point)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.1567 -0.6566 -0.2724  0.1697  3.2404
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       -0.16971    0.04233  -4.009 6.28e-05 ***
## Data2$Time_Point   0.44214    0.02414  18.316  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.05 on 2463 degrees of freedom
## Multiple R-squared:  0.1199, Adjusted R-squared:  0.1195
## F-statistic: 335.5 on 1 and 2463 DF,  p-value: < 2.2e-16
```

```
# simply connects the mean of each time point
ggplot(Data2, aes(x = Time_Point, y = Binding)) + geom_point(aes(color = Treatment)) + geom_jitter() + s
```
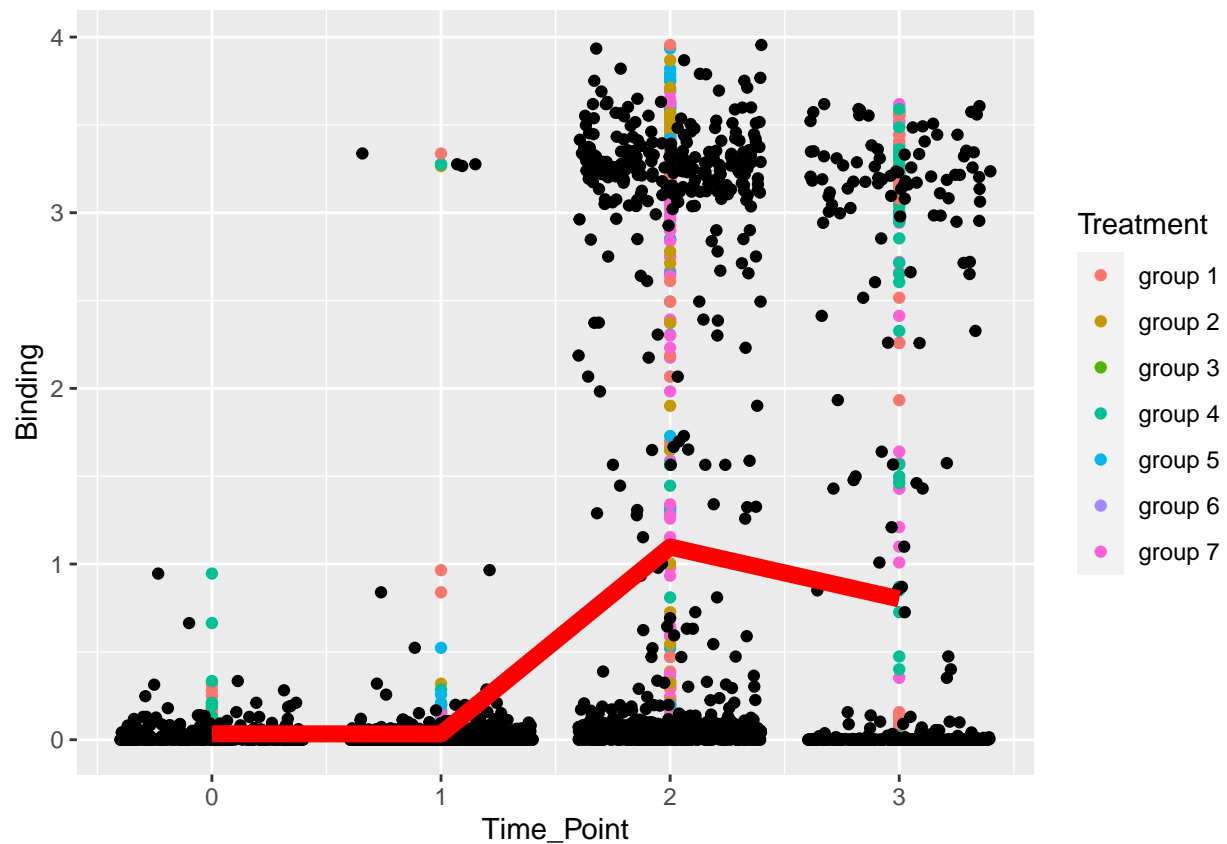

Here we use Binding as the only response. Prdictors: `Treatment`.
Random effect for both intercept and slope.

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

```
lda <- lme(fixed = Binding ~ Time_Point + Treatment,
           random = ~ Time_Point | MonkeyID, data = Data2, method = "REML")
summary(lda)
```

```
## Linear mixed-effects model fit by REML
##   Data: Data2
##        AIC      BIC    logLik
##   6753.155 6822.836 -3364.578
##
## Random effects:
##  Formula: ~Time_Point | MonkeyID
##  Structure: General positive-definite, Log-Cholesky parametrization
##             StdDev    Corr
## (Intercept) 0.4643973 (Intr)
## Time_Point  0.5545189 -0.981
## Residual    0.9335139
##
## Fixed effects: Binding ~ Time_Point + Treatment
##                     Value Std.Error   DF   t-value p-value
## (Intercept)     -0.3652114 0.1578008 2444 -2.314383  0.0207
## Time_Point       0.6640931 0.1709633 2444  3.884418  0.0001
## Treatmentgroup 2 -0.1880652 0.1727988   13 -1.088348  0.2962
## Treatmentgroup 3 -0.2745892 0.1500821   13 -1.829593  0.0903
## Treatmentgroup 4 -0.0112400 0.1089163   13 -0.103199  0.9194
## Treatmentgroup 5 -0.3893418 0.1134100   13 -3.433045  0.0045
## Treatmentgroup 6 -0.2971236 0.1430497   13 -2.077066  0.0582
## Treatmentgroup 7 -0.0215799 0.1045159   13 -0.206475  0.8396
##   Correlation:
##                  (Intr) Tm_Pnt Trtmn2 Trtmn3 Trtmn4 Trtmn5 Trtmn6
## Time_Point       -0.899
## Treatmentgroup 2  0.002 -0.197
## Treatmentgroup 3 -0.027 -0.194  0.222
## Treatmentgroup 4 -0.277 -0.001  0.254  0.292
## Treatmentgroup 5 -0.169 -0.109  0.265  0.302  0.387
## Treatmentgroup 6 -0.029 -0.203  0.233  0.262  0.307  0.317
## Treatmentgroup 7 -0.272 -0.020  0.268  0.308  0.420  0.405  0.323
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -3.52596967 -0.52101840 -0.10300523  0.03925278  3.58753831
##
## Number of Observations: 2465
## Number of Groups: 20
```