

460_ModelSelection_Week2

Geyi Liu

2020/11/26

```
library(stringr)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
library(ggplot2)
library(lubridate)
```

```
## Warning: package 'lubridate' was built under R version 4.0.3

##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':
##
##   date, intersect, setdiff, union
```

```
library(car)
```

```
## Warning: package 'car' was built under R version 4.0.3

## Loading required package: carData

## Warning: package 'carData' was built under R version 4.0.3

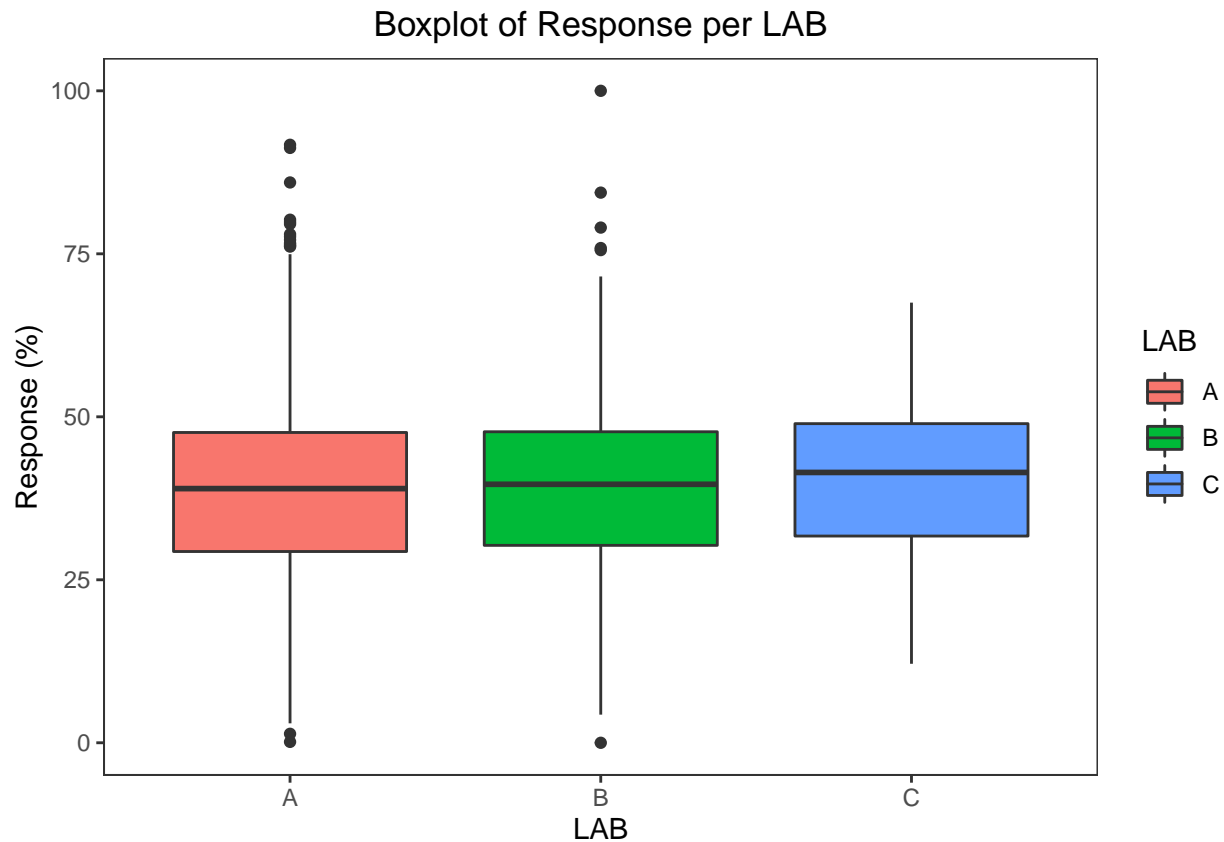
##
## Attaching package: 'car'

## The following object is masked from 'package:dplyr':
##
##   recode
```

```
data <- read.csv("../data/Data.csv", stringsAsFactors=FALSE)
data$LAB <- as.factor(data$LAB) # convert LAB to be a factor
data[is.na(data)] <- 0 # replace NAs with zero
```

```
data_model <- data
data_model$X <- NULL # drop identifier column
data_model$LAB <- NULL # drop non-numeric LAB column
data_model <- data_model[1:nrow(data_model),477:ncol(data_model)]
Data_remove <- data_model[which(colMeans(data_model) > 0.02,)] # Remove 0 majority columns, because they
```

```
ggplot(data=data) +
  geom_boxplot(aes(x=LAB,y=Response,fill = LAB)) +
  labs(title="Boxplot of Response per LAB",
       x="LAB", y="Response (%)") +
  theme_bw() +
  theme(panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        plot.title = element_text(hjust = 0.5))
```



```
# The Response distribution of three Labs are very similar
```

```
# A linear regression model of the relationship between remain group variables and Response was established.
# The difference in units between variables is not considered here (normalization)
model <- lm(Response~.,data = Data_remove)
summary(model)
```

```
##
## Call:
## lm(formula = Response ~ ., data = Data_remove)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -33.120  -4.460   0.365   4.636  43.062
##
## Coefficients: (8 not defined because of singularities)
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.598e+01  1.009e+01   3.564 0.000368 ***
## Group1_2     -1.064e+01  2.519e+01  -0.422 0.672745
## Group1_6     -6.154e+01  1.953e+01  -3.151 0.001634 **
## Group1_7      3.929e+00  1.463e+00   2.686 0.007254 **
## Group1_9     -1.786e+01  9.417e+00  -1.896 0.057988 .
## Group1_11     6.759e+00  1.014e+00   6.669 2.81e-11 ***
## Group1_13     7.464e+00  2.544e+01   0.293 0.769237
## Group1_17     5.246e-01  9.504e-01   0.552 0.580981
## Group1_19    -1.840e+00  1.847e+00  -0.997 0.318962
## Group1_20     8.469e-01  3.378e-01   2.507 0.012204 *
## Group1_22    -4.789e-02  3.762e-01  -0.127 0.898721
## Group1_25    -2.523e-01  2.806e-01  -0.899 0.368532
## Group1_26     1.159e-01  1.151e-01   1.007 0.314031
## Group1_29     3.553e-02  1.187e-01   0.299 0.764669
## Group1_30     5.799e-01  6.107e-01   0.949 0.342405
## Group13_1    -2.053e-01  1.039e-01  -1.975 0.048319 *
## Group2_1      8.124e+01  5.718e+01   1.421 0.155431
## Group2_2     -2.205e-01  2.730e+00  -0.081 0.935620
## Group2_3     -7.700e+01  1.174e+02  -0.656 0.511966
## Group2_4     -4.034e+01  1.917e+01  -2.105 0.035346 *
## Group2_5     -2.952e-01  5.798e+00  -0.051 0.959394
## Group2_6      3.610e+00  1.813e+01   0.199 0.842162
## Group2_8      2.178e+00  6.993e+00   0.311 0.755474
## Group2_9      1.947e+00  2.589e+00   0.752 0.451997
## Group2_10    -6.755e+01  1.059e+02  -0.638 0.523553
## Group2_11    -3.697e-01  7.948e-01  -0.465 0.641833
## Group2_12    -4.054e+01  7.201e+01  -0.563 0.573451
## Group2_13     8.303e-01  7.965e+00   0.104 0.916977
## Group2_14    -2.485e+01  1.058e+01  -2.348 0.018923 *
## Group2_15     1.328e+02  1.554e+02   0.855 0.392617
## Group2_18     2.837e+02  1.413e+02   2.007 0.044747 *
## Group2_19    -8.536e+00  9.729e+00  -0.877 0.380346
## Group2_20     1.159e+00  1.476e+00   0.785 0.432586
## Group2_21     3.359e-01  4.117e-01   0.816 0.414601
## Group2_22     1.835e+00  1.145e+00   1.602 0.109283
## Group2_23     7.675e-01  2.175e+00   0.353 0.724150
## Group2_25    -3.314e+00  1.120e+01  -0.296 0.767324
## Group2_26    -4.736e+00  7.225e+00  -0.656 0.512152
## Group2_27    -1.243e+00  1.347e+00  -0.923 0.356248
## Group2_28    -1.052e+00  1.604e+00  -0.656 0.512109
## Group2_29     7.126e-01  8.072e-01   0.883 0.377395
## Group2_30     6.243e-01  8.344e-01   0.748 0.454344
## Group2_32    -3.656e-01  9.355e-01  -0.391 0.695948
## Group2_33     5.555e+00  6.428e+00   0.864 0.387485
```

## Group2_34	5.487e-01	6.316e+00	0.087	0.930770	
## Group2_35	1.740e+01	2.659e+01	0.654	0.513005	
## Group2_36	-1.799e+01	2.616e+01	-0.687	0.491798	
## Group2_37	1.379e-01	4.531e+00	0.030	0.975717	
## Group2_38	-2.018e+01	2.628e+01	-0.768	0.442490	
## Group2_39	-4.322e-02	1.913e-01	-0.226	0.821276	
## Group2_40	4.640e-02	1.468e-01	0.316	0.752039	
## Group2_41	7.505e-02	1.975e-01	0.380	0.703979	
## Group2_43	1.808e+02	1.347e+02	1.342	0.179598	
## Group2_44	-6.383e+00	9.552e+00	-0.668	0.504015	
## Group2_45	-7.486e+00	6.892e+00	-1.086	0.277465	
## Group3_5	-2.505e+00	1.364e+00	-1.837	0.066225	.
## Group3_8	3.299e+00	2.892e+00	1.141	0.253956	
## Group3_9	2.090e+01	8.783e+01	0.238	0.811922	
## Group3_10	-1.753e+00	1.501e+00	-1.168	0.242777	
## Group3_11	3.623e+02	1.103e+02	3.284	0.001031	**
## Group3_12	-8.792e+00	3.870e+00	-2.272	0.023146	*
## Group3_15	-1.130e+02	8.804e+01	-1.284	0.199354	
## Group3_24	-2.073e+00	1.123e+00	-1.846	0.064995	.
## Group11_6	-2.011e+00	5.324e-01	-3.776	0.000161	***
## Group6_6	1.660e+00	2.523e-01	6.580	5.08e-11	***
## Group8_9	-2.986e+01	1.589e+01	-1.880	0.060185	.
## Group8_10	4.133e+00	2.120e+00	1.949	0.051288	.
## Group8_20	-1.091e-01	4.040e-01	-0.270	0.787145	
## Group9_4	1.103e-01	1.161e-01	0.950	0.342138	
## Group9_8	2.070e+00	2.321e+00	0.892	0.372515	
## Group9_9	1.205e-03	4.195e-03	0.287	0.773879	
## Group9_10	-1.317e-03	1.730e-03	-0.761	0.446428	
## Group9_12	2.091e+00	2.059e+00	1.016	0.309824	
## Group9_16	1.155e-02	1.805e-02	0.640	0.522313	
## Group9_17	NA	NA	NA	NA	
## Group9_21	NA	NA	NA	NA	
## Group9_24	-5.024e+03	1.533e+03	-3.276	0.001057	**
## Group9_25	1.183e+01	3.618e+00	3.270	0.001080	**
## Group9_26	-1.558e+00	4.794e-01	-3.250	0.001161	**
## Group9_28	1.671e+03	5.184e+02	3.223	0.001277	**
## Group9_30	NA	NA	NA	NA	
## Group9_32	NA	NA	NA	NA	
## Group9_33	NA	NA	NA	NA	
## Group9_35	-7.160e-01	6.096e-01	-1.175	0.240206	
## Group9_36	-2.129e-04	7.050e-04	-0.302	0.762665	
## Group9_37	3.734e-04	6.779e-04	0.551	0.581793	
## Group9_39	-6.487e-01	7.479e-01	-0.867	0.385835	
## Group9_43	6.569e-04	5.213e-03	0.126	0.899717	
## Group9_46	-5.523e-01	2.055e+01	-0.027	0.978563	
## Group9_48	-1.798e+00	5.348e+00	-0.336	0.736692	
## Group9_49	1.734e+00	2.087e+01	0.083	0.933763	
## Group9_52	2.957e+00	3.610e+00	0.819	0.412827	
## Group9_53	1.292e-03	3.097e-03	0.417	0.676618	
## Group9_54	-7.478e-04	8.225e-04	-0.909	0.363299	
## Group9_56	1.899e-02	3.980e+00	0.005	0.996194	
## Group9_58	-1.229e+00	3.468e+00	-0.355	0.722963	
## Group9_60	4.130e-04	1.043e-02	0.040	0.968416	
## Group9_61	-5.392e-01	1.138e+00	-0.474	0.635593	

## Group9_64	2.042e-01	2.777e-01	0.736	0.462021	
## Group9_65	-1.828e-01	2.473e-01	-0.739	0.459932	
## Group9_66	3.304e+00	2.174e+00	1.519	0.128710	
## Group9_69	-9.534e+00	2.302e+01	-0.414	0.678821	
## Group9_70	3.104e-03	8.107e-03	0.383	0.701848	
## Group9_71	NA	NA	NA	NA	
## Group9_77	1.422e-01	6.120e-01	0.232	0.816277	
## Group9_78	-3.782e-02	1.746e-01	-0.217	0.828512	
## Group10_1	-2.194e+02	1.468e+03	-0.149	0.881263	
## Group10_3	3.458e+01	5.057e+01	0.684	0.494082	
## Group10_4	-2.512e+02	7.588e+02	-0.331	0.740666	
## Group10_5	8.838e+02	1.842e+03	0.480	0.631410	
## Group10_6	5.943e+02	4.077e+03	0.146	0.884108	
## Group10_7	-8.418e+02	1.111e+03	-0.758	0.448590	
## Group10_11	-5.880e+01	2.689e+02	-0.219	0.826897	
## Group10_12	-1.030e+01	1.243e+01	-0.829	0.407108	
## Group10_13	1.014e+02	4.184e+02	0.242	0.808497	
## Group10_14	NA	NA	NA	NA	
## Group10_16	NA	NA	NA	NA	
## Group13_3	1.478e+01	1.582e+02	0.093	0.925581	
## Group13_4	-2.075e-02	1.310e-02	-1.583	0.113375	
## Group13_5	1.338e+01	5.299e+00	2.526	0.011572	*
## Group13_6	-2.003e+00	5.714e-01	-3.506	0.000458	***
## Group13_7	1.469e-01	1.894e-01	0.776	0.437944	
## Group13_8	-4.288e-01	6.861e-01	-0.625	0.531933	
## Group13_9	-1.968e+00	8.664e-01	-2.271	0.023151	*
## Group13_10	-1.152e-01	7.561e-01	-0.152	0.878903	
## Group13_11	-2.289e+00	2.828e+00	-0.809	0.418345	
## Group13_12	-1.560e+00	3.197e+00	-0.488	0.625670	
## Group13_13	-2.706e+00	2.960e+00	-0.914	0.360639	
## Group13_14	-2.586e-03	1.670e-02	-0.155	0.876973	
## Group13_15	3.690e-02	1.146e-01	0.322	0.747512	
## Group13_16	-5.341e-01	7.142e-01	-0.748	0.454585	
## Group13_17	1.259e+00	1.351e+00	0.932	0.351401	
## Group13_18	-5.054e-01	1.484e+00	-0.341	0.733431	
## Group13_19	-7.215e-01	9.872e-01	-0.731	0.464892	
## Group13_20	1.831e+00	6.596e-01	2.776	0.005528	**
## Group13_21	-4.138e+00	1.387e+00	-2.982	0.002871	**
## Group13_22	3.715e+00	1.963e+00	1.893	0.058436	.
## Group13_23	1.623e-01	2.677e+00	0.061	0.951651	
## Group13_24	-1.499e+00	1.758e+00	-0.853	0.393968	
## Group13_25	7.787e-01	1.084e+00	0.719	0.472466	
## Group13_26	-2.281e-01	1.237e+00	-0.184	0.853718	
## Group13_27	-6.532e-01	2.005e+00	-0.326	0.744615	
## Group13_28	6.618e-01	1.172e+00	0.565	0.572269	
## Group13_29	1.678e-02	2.995e-01	0.056	0.955315	
## Group13_30	-5.296e-01	4.398e-01	-1.204	0.228624	
## Group13_31	7.838e-01	5.174e-01	1.515	0.129875	
## Group13_32	-4.571e-01	3.975e-01	-1.150	0.250206	
## Group13_33	-3.118e-02	1.661e-01	-0.188	0.851076	
## Group13_34	1.010e-01	2.768e-02	3.648	0.000266	***
## Group13_36	3.180e+00	2.261e+00	1.406	0.159675	
## Group13_37	-8.836e-01	1.113e+00	-0.794	0.427219	
## Group13_38	-7.847e-01	7.828e-01	-1.002	0.316153	

```
## Group13_39 7.276e-01 4.400e-01 1.653 0.098291 .
## Group13_40 -5.060e-03 3.295e-01 -0.015 0.987746
## Group13_41 -2.131e-01 2.386e-01 -0.893 0.371836
## Group13_42 1.657e-01 2.924e-01 0.567 0.570884
## Group13_43 -4.023e-01 2.627e-01 -1.531 0.125795
## Group13_44 4.542e-01 3.763e-01 1.207 0.227453
## Group13_45 -7.396e-01 3.399e-01 -2.176 0.029609 *
## Group13_46 1.695e+00 6.044e-01 2.805 0.005045 **
## Group13_47 -2.114e+00 8.265e-01 -2.558 0.010540 *
## Group13_48 4.299e-01 3.042e-01 1.413 0.157698
## Group13_49 3.449e-01 1.244e-01 2.773 0.005569 **
## Group13_50 -5.573e+00 2.761e+00 -2.019 0.043572 *
## Group13_51 -9.564e-03 1.626e-02 -0.588 0.556514
## Group13_52 6.155e-05 3.222e-05 1.911 0.056107 .
## Group13_53 -1.714e-02 9.258e-02 -0.185 0.853123
## Group13_55 -7.766e-02 7.866e-02 -0.987 0.323555
## Group13_56 1.806e-01 8.258e-02 2.187 0.028769 *
## Group13_59 -4.867e-03 2.592e-03 -1.878 0.060455 .
## Group13_61 -8.634e-03 6.268e-03 -1.377 0.168418
## Group13_62 3.055e+02 2.402e+02 1.272 0.203423
## Group13_63 -2.960e+02 2.317e+02 -1.277 0.201496
## Group13_64 1.915e-02 2.959e-01 0.065 0.948396
## Group13_65 -1.331e-02 4.948e-02 -0.269 0.787870
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.454 on 6193 degrees of freedom
## Multiple R-squared: 0.6621, Adjusted R-squared: 0.653
## F-statistic: 73.09 on 166 and 6193 DF, p-value: < 2.2e-16
```

```
# Multiple R-squared: 0.6621, Adjusted R-squared: 0.653
# The model can only explain 65% of the data
```

```
# Too many variables cannot be processed with functions such as step()
# Here, we choose the statistically significant variables(p value<0.05) in the first model to continue
# A preliminary conclusion can only be reached by manual treatment
```

```
model_2 <- lm(Response~Group1_9+Group1_11+Group1_20+Group1_22+Group2_4+
  Group2_9+Group2_14+Group2_22+Group3_8+Group3_10+Group3_12+
  Group11_6+Group6_6+Group9_24+Group9_25+Group9_26+Group9_28+
  Group13_6+Group13_11+Group13_13+Group13_20+Group13_21+
  Group13_22+Group13_34+Group13_46+Group13_47+Group13_49+
  Group13_50+Group13_52+Group13_56,data = Data_remove)
summary(model_2)
```

```
##
## Call:
## lm(formula = Response ~ Group1_9 + Group1_11 + Group1_20 + Group1_22 +
##     Group2_4 + Group2_9 + Group2_14 + Group2_22 + Group3_8 +
##     Group3_10 + Group3_12 + Group11_6 + Group6_6 + Group9_24 +
##     Group9_25 + Group9_26 + Group9_28 + Group13_6 + Group13_11 +
##     Group13_13 + Group13_20 + Group13_21 + Group13_22 + Group13_34 +
##     Group13_46 + Group13_47 + Group13_49 + Group13_50 + Group13_52 +
##     Group13_56, data = Data_remove)
```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -51.025  -4.616   0.555   4.847  42.651
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.428e+01  2.851e+00   5.007 5.68e-07 ***
## Group1_9      1.015e+01  4.230e+00   2.400 0.016438 *
## Group1_11     3.688e+00  5.786e-01   6.373 1.98e-10 ***
## Group1_20     2.132e-02  8.999e-02   0.237 0.812719
## Group1_22    -2.642e-01  1.267e-01  -2.086 0.036992 *
## Group2_4     -1.121e+00  3.490e+00  -0.321 0.748068
## Group2_9      1.888e+00  2.565e-01   7.361 2.06e-13 ***
## Group2_14     2.267e+00  6.879e-01   3.295 0.000990 ***
## Group2_22     8.881e-01  2.199e-01   4.038 5.45e-05 ***
## Group3_8      6.658e+00  1.755e+00   3.793 0.000150 ***
## Group3_10    -4.271e+00  1.385e+00  -3.084 0.002053 **
## Group3_12    -4.037e+00  1.443e+00  -2.797 0.005174 **
## Group11_6    -1.120e+00  4.863e-01  -2.303 0.021295 *
## Group6_6      2.304e+00  1.973e-01  11.674 < 2e-16 ***
## Group9_24    -5.145e+03  1.281e+03  -4.018 5.94e-05 ***
## Group9_25     1.210e+01  3.021e+00   4.004 6.31e-05 ***
## Group9_26    -1.599e+00  3.996e-01  -4.001 6.38e-05 ***
## Group9_28     1.747e+03  4.307e+02   4.056 5.05e-05 ***
## Group13_6    -1.488e+00  3.173e-01  -4.690 2.79e-06 ***
## Group13_11   -1.518e-01  6.677e-02  -2.273 0.023076 *
## Group13_13   -9.538e-01  1.608e-01  -5.932 3.15e-09 ***
## Group13_20     1.716e-01  1.061e-01   1.618 0.105722
## Group13_21   -2.739e-01  2.941e-01  -0.931 0.351632
## Group13_22     1.519e-01  2.088e-01   0.728 0.466933
## Group13_34     5.063e-02  9.862e-03   5.134 2.92e-07 ***
## Group13_46     6.176e-01  3.204e-01   1.928 0.053937 .
## Group13_47   -1.257e+00  3.547e-01  -3.543 0.000398 ***
## Group13_49     2.677e-01  4.870e-02   5.497 4.00e-08 ***
## Group13_50   -6.110e+00  7.752e-01  -7.881 3.80e-15 ***
## Group13_52     3.718e-05  2.293e-05   1.621 0.105014
## Group13_56     2.088e-01  6.918e-02   3.019 0.002545 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.64 on 6329 degrees of freedom
## Multiple R-squared:  0.6373, Adjusted R-squared:  0.6355
## F-statistic: 370.6 on 30 and 6329 DF, p-value: < 2.2e-16
```

We need to reduce the number of variables as much as possible, and select the variables that have an

```
model_3 <- lm(Response~Group1_11+Group2_9+Group2_14+
  Group2_22+Group3_8+Group3_10+Group3_12+Group6_6+
  Group9_24+Group9_25+Group9_26+Group9_28+Group13_6+
  Group13_11+Group13_13+Group13_34+Group13_47+Group13_49+
  Group13_50+Group13_56,data = Data_remove)
# 20 variables
summary(model_3)
```

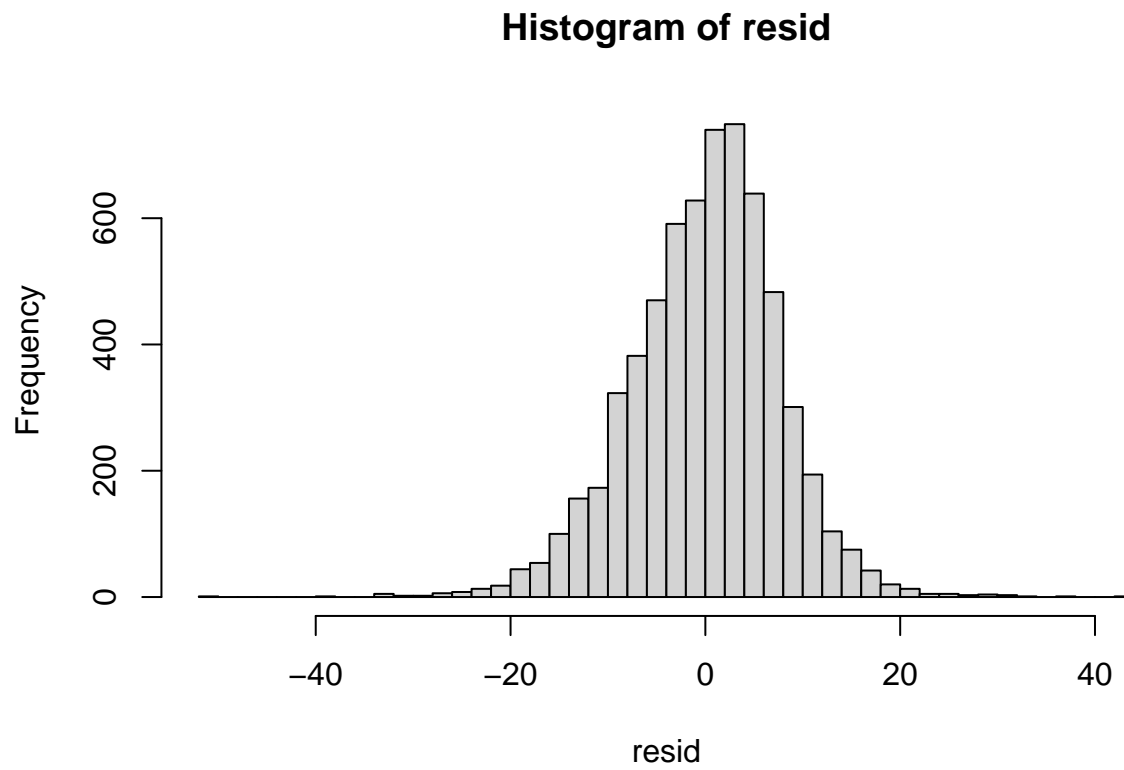
```
##
## Call:
## lm(formula = Response ~ Group1_11 + Group2_9 + Group2_14 + Group2_22 +
##      Group3_8 + Group3_10 + Group3_12 + Group6_6 + Group9_24 +
##      Group9_25 + Group9_26 + Group9_28 + Group13_6 + Group13_11 +
##      Group13_13 + Group13_34 + Group13_47 + Group13_49 + Group13_50 +
##      Group13_56, data = Data_remove)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -50.582  -4.595   0.586   4.836  42.748
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.272e+01  2.756e+00   4.614 4.03e-06 ***
## Group1_11     5.037e+00  6.786e-02  74.215 < 2e-16 ***
## Group2_9      1.830e+00  1.745e-01  10.490 < 2e-16 ***
## Group2_14     2.042e+00  5.379e-01   3.797 0.000148 ***
## Group2_22     9.045e-01  1.114e-01   8.119 5.60e-16 ***
## Group3_8      7.471e+00  1.738e+00   4.298 1.75e-05 ***
## Group3_10    -5.133e+00  1.345e+00  -3.815 0.000138 ***
## Group3_12    -4.226e+00  1.442e+00  -2.931 0.003388 **
## Group6_6      2.343e+00  1.880e-01  12.465 < 2e-16 ***
## Group9_24    -5.452e+03  1.281e+03  -4.258 2.10e-05 ***
## Group9_25     1.283e+01  3.021e+00   4.247 2.20e-05 ***
## Group9_26    -1.695e+00  3.996e-01  -4.241 2.25e-05 ***
## Group9_28     1.842e+03  4.307e+02   4.276 1.93e-05 ***
## Group13_6    -1.778e+00  2.870e-01  -6.194 6.23e-10 ***
## Group13_11   -2.001e-01  6.310e-02  -3.171 0.001529 **
## Group13_13   -9.329e-01  1.591e-01  -5.862 4.80e-09 ***
## Group13_34    5.601e-02  7.999e-03   7.002 2.79e-12 ***
## Group13_47   -3.164e-01  8.420e-02  -3.757 0.000173 ***
## Group13_49    1.310e-01  3.043e-02   4.304 1.70e-05 ***
## Group13_50   -3.802e+00  4.607e-01  -8.251 < 2e-16 ***
## Group13_56    2.307e-01  6.901e-02   3.343 0.000833 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.658 on 6339 degrees of freedom
## Multiple R-squared:  0.635, Adjusted R-squared:  0.6338
## F-statistic: 551.3 on 20 and 6339 DF, p-value: < 2.2e-16
```

```
vif(model_3)
```

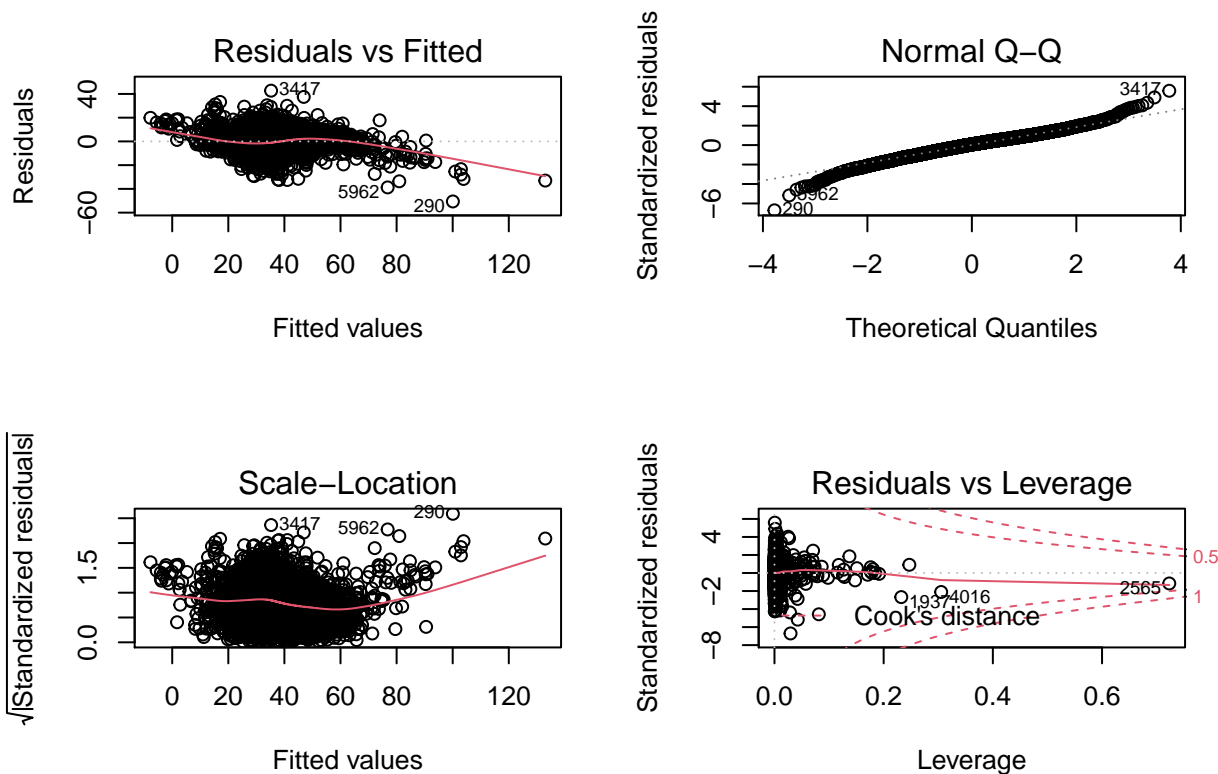
```
##      Group1_11      Group2_9      Group2_14      Group2_22      Group3_8      Group3_10
## 1.165492e+00 3.932537e+00 2.788751e+00 3.081065e+00 1.379437e+02 3.604530e+01
##      Group3_12      Group6_6      Group9_24      Group9_25      Group9_26      Group9_28
## 9.668301e+01 1.440010e+00 1.230999e+06 3.459065e+06 1.109405e+06 9.057347e+04
##      Group13_6      Group13_11      Group13_13      Group13_34      Group13_47      Group13_49
## 1.286216e+01 2.135932e+01 2.086930e+00 1.616362e+01 1.380179e+01 3.964368e+01
##      Group13_50      Group13_56
## 5.094612e+01 1.818830e+00
```



```
# # Residual histogram of model  
resid=residuals(model_3)  
hist(resid,breaks=40)
```



```
par(mfrow=c(2,2))  
plot(model_3)
```



```
# step function
stp_model2 <- step(model_2)
```

```
## Start: AIC=25895.48
## Response ~ Group1_9 + Group1_11 + Group1_20 + Group1_22 + Group2_4 +
##   Group2_9 + Group2_14 + Group2_22 + Group3_8 + Group3_10 +
##   Group3_12 + Group11_6 + Group6_6 + Group9_24 + Group9_25 +
##   Group9_26 + Group9_28 + Group13_6 + Group13_11 + Group13_13 +
##   Group13_20 + Group13_21 + Group13_22 + Group13_34 + Group13_46 +
##   Group13_47 + Group13_49 + Group13_50 + Group13_52 + Group13_56
##
##           Df Sum of Sq   RSS   AIC
## - Group1_20  1         3.3 369409 25894
## - Group2_4   1         6.0 369412 25894
## - Group13_22 1        30.9 369437 25894
## - Group13_21 1        50.6 369457 25894
## <none>                        369406 25896
## - Group13_20 1       152.8 369559 25896
## - Group13_52 1       153.4 369559 25896
## - Group13_46 1       216.9 369623 25897
## - Group1_22  1       254.1 369660 25898
## - Group13_11 1       301.5 369707 25899
## - Group11_6   1       309.6 369716 25899
## - Group1_9    1       336.1 369742 25899
## - Group3_12   1       456.6 369862 25901
## - Group13_56 1       532.0 369938 25903
```

```

## - Group3_10 1 555.1 369961 25903
## - Group2_14 1 633.7 370040 25904
## - Group13_47 1 732.7 370139 25906
## - Group3_8 1 839.8 370246 25908
## - Group9_26 1 934.3 370340 25910
## - Group9_25 1 935.5 370341 25910
## - Group9_24 1 942.3 370348 25910
## - Group2_22 1 951.8 370358 25910
## - Group9_28 1 960.3 370366 25910
## - Group13_6 1 1283.7 370690 25916
## - Group13_34 1 1538.7 370945 25920
## - Group13_49 1 1764.0 371170 25924
## - Group13_13 1 2053.8 371460 25929
## - Group1_11 1 2370.8 371777 25934
## - Group2_9 1 3162.5 372568 25948
## - Group13_50 1 3625.2 373031 25956
## - Group6_6 1 7954.6 377360 26029
##
## Step: AIC=25893.53
## Response ~ Group1_9 + Group1_11 + Group1_22 + Group2_4 + Group2_9 +
## Group2_14 + Group2_22 + Group3_8 + Group3_10 + Group3_12 +
## Group11_6 + Group6_6 + Group9_24 + Group9_25 + Group9_26 +
## Group9_28 + Group13_6 + Group13_11 + Group13_13 + Group13_20 +
## Group13_21 + Group13_22 + Group13_34 + Group13_46 + Group13_47 +
## Group13_49 + Group13_50 + Group13_52 + Group13_56
##
## Df Sum of Sq RSS AIC
## - Group2_4 1 4.3 369413 25892
## - Group13_22 1 29.3 369438 25892
## - Group13_21 1 48.7 369458 25892
## <none> 369409 25894
## - Group13_20 1 150.2 369559 25894
## - Group13_52 1 156.2 369565 25894
## - Group13_46 1 215.4 369625 25895
## - Group1_22 1 271.2 369680 25896
## - Group13_11 1 298.3 369707 25897
## - Group1_9 1 341.9 369751 25897
## - Group11_6 1 359.2 369768 25898
## - Group3_12 1 460.5 369870 25900
## - Group13_56 1 532.3 369941 25901
## - Group3_10 1 563.3 369972 25901
## - Group2_14 1 632.3 370041 25902
## - Group13_47 1 730.6 370140 25904
## - Group3_8 1 847.7 370257 25906
## - Group9_26 1 933.0 370342 25908
## - Group9_25 1 934.3 370343 25908
## - Group9_24 1 941.1 370350 25908
## - Group9_28 1 959.0 370368 25908
## - Group2_22 1 1013.9 370423 25909
## - Group13_6 1 1287.2 370696 25914
## - Group13_34 1 1535.5 370945 25918
## - Group13_49 1 1761.3 371170 25922
## - Group13_13 1 2050.8 371460 25927
## - Group1_11 1 2646.5 372056 25937

```

```

## - Group2_9      1      3249.9 372659 25947
## - Group13_50    1      3621.9 373031 25954
## - Group6_6      1      7957.4 377367 26027
##
## Step:  AIC=25891.61
## Response ~ Group1_9 + Group1_11 + Group1_22 + Group2_9 + Group2_14 +
##      Group2_22 + Group3_8 + Group3_10 + Group3_12 + Group11_6 +
##      Group6_6 + Group9_24 + Group9_25 + Group9_26 + Group9_28 +
##      Group13_6 + Group13_11 + Group13_13 + Group13_20 + Group13_21 +
##      Group13_22 + Group13_34 + Group13_46 + Group13_47 + Group13_49 +
##      Group13_50 + Group13_52 + Group13_56
##
##              Df Sum of Sq    RSS    AIC
## - Group13_22  1         32.8 369446 25890
## - Group13_21  1         53.5 369467 25891
## <none>                                369413 25892
## - Group13_20  1        159.4 369573 25892
## - Group13_52  1        160.1 369574 25892
## - Group13_46  1        215.1 369629 25893
## - Group1_22   1        268.2 369682 25894
## - Group13_11  1        296.6 369710 25895
## - Group1_9    1        346.6 369760 25896
## - Group11_6   1        361.8 369775 25896
## - Group3_12   1        457.2 369871 25898
## - Group13_56  1        531.5 369945 25899
## - Group3_10   1        560.6 369974 25899
## - Group13_47  1        733.2 370147 25902
## - Group3_8    1        844.8 370258 25904
## - Group9_26   1        940.8 370354 25906
## - Group9_25   1        941.9 370355 25906
## - Group9_24   1        948.7 370362 25906
## - Group9_28   1        967.1 370381 25906
## - Group2_14   1       1110.2 370524 25909
## - Group13_6   1       1289.8 370703 25912
## - Group13_34  1       1533.7 370947 25916
## - Group13_49  1       1795.5 371209 25920
## - Group13_13  1       2046.8 371460 25925
## - Group1_11   1       2642.2 372056 25935
## - Group2_22   1       2849.3 372263 25939
## - Group13_50  1       3703.1 373117 25953
## - Group2_9    1       5515.8 374929 25984
## - Group6_6    1      8632.6 378046 26037
##
## Step:  AIC=25890.17
## Response ~ Group1_9 + Group1_11 + Group1_22 + Group2_9 + Group2_14 +
##      Group2_22 + Group3_8 + Group3_10 + Group3_12 + Group11_6 +
##      Group6_6 + Group9_24 + Group9_25 + Group9_26 + Group9_28 +
##      Group13_6 + Group13_11 + Group13_13 + Group13_20 + Group13_21 +
##      Group13_34 + Group13_46 + Group13_47 + Group13_49 + Group13_50 +
##      Group13_52 + Group13_56
##
##              Df Sum of Sq    RSS    AIC
## - Group13_21  1         74.8 369521 25890
## <none>                                369446 25890

```

```

## - Group13_52 1      188.9 369635 25891
## - Group13_20 1      209.8 369656 25892
## - Group13_46 1      220.3 369667 25892
## - Group1_22  1      257.0 369703 25893
## - Group13_11 1      333.1 369779 25894
## - Group1_9   1      349.9 369796 25894
## - Group11_6  1      367.4 369814 25895
## - Group3_12  1      458.4 369905 25896
## - Group13_56 1      526.1 369972 25897
## - Group3_10  1      563.8 370010 25898
## - Group13_47 1      712.5 370159 25900
## - Group3_8   1      851.4 370298 25903
## - Group9_26  1      936.5 370383 25904
## - Group9_25  1      937.5 370384 25904
## - Group9_24  1      944.1 370390 25904
## - Group9_28  1      962.6 370409 25905
## - Group2_14  1     1116.8 370563 25907
## - Group13_6  1     1399.1 370845 25912
## - Group13_49 1     1950.5 371397 25922
## - Group13_13 1     2159.2 371605 25925
## - Group13_34 1     2227.4 371674 25926
## - Group1_11  1     2631.1 372077 25933
## - Group2_22  1     2830.9 372277 25937
## - Group13_50 1     4604.6 374051 25967
## - Group2_9   1     5565.7 375012 25983
## - Group6_6   1     8602.2 378049 26035
##
## Step:  AIC=25889.46
## Response ~ Group1_9 + Group1_11 + Group1_22 + Group2_9 + Group2_14 +
##      Group2_22 + Group3_8 + Group3_10 + Group3_12 + Group11_6 +
##      Group6_6 + Group9_24 + Group9_25 + Group9_26 + Group9_28 +
##      Group13_6 + Group13_11 + Group13_13 + Group13_20 + Group13_34 +
##      Group13_46 + Group13_47 + Group13_49 + Group13_50 + Group13_52 +
##      Group13_56
##
##           Df Sum of Sq    RSS    AIC
## <none>                369521 25890
## - Group13_52 1      184.6 369706 25891
## - Group13_46 1      215.1 369736 25891
## - Group1_22  1      266.3 369787 25892
## - Group1_9   1      322.4 369843 25893
## - Group11_6  1      361.7 369883 25894
## - Group13_11 1      381.5 369903 25894
## - Group3_12  1      455.5 369977 25895
## - Group13_56 1      514.5 370036 25896
## - Group3_10  1      563.5 370085 25897
## - Group13_47 1      678.6 370200 25899
## - Group3_8   1      852.5 370374 25902
## - Group9_26  1      943.4 370465 25904
## - Group9_25  1      944.7 370466 25904
## - Group9_24  1      951.4 370472 25904
## - Group9_28  1      969.6 370491 25904
## - Group2_14  1     1099.8 370621 25906
## - Group13_20 1     1104.8 370626 25906

```

```
## - Group13_6 1 1531.3 371052 25914
## - Group13_49 1 2094.1 371615 25923
## - Group13_34 1 2178.6 371700 25925
## - Group13_13 1 2183.1 371704 25925
## - Group1_11 1 2745.8 372267 25935
## - Group2_22 1 2876.2 372397 25937
## - Group13_50 1 4916.8 374438 25972
## - Group2_9 1 5647.1 375168 25984
## - Group6_6 1 8669.7 378191 26035
```

```
summary(stp_model2)
```

```
##
## Call:
## lm(formula = Response ~ Group1_9 + Group1_11 + Group1_22 + Group2_9 +
##      Group2_14 + Group2_22 + Group3_8 + Group3_10 + Group3_12 +
##      Group11_6 + Group6_6 + Group9_24 + Group9_25 + Group9_26 +
##      Group9_28 + Group13_6 + Group13_11 + Group13_13 + Group13_20 +
##      Group13_34 + Group13_46 + Group13_47 + Group13_49 + Group13_50 +
##      Group13_52 + Group13_56, data = Data_remove)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -50.704  -4.622   0.556   4.847  42.633
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.378e+01  2.769e+00  4.979 6.56e-07 ***
## Group1_9      9.548e+00  4.062e+00  2.351 0.018777 *
## Group1_11     3.775e+00  5.503e-01  6.860 7.54e-12 ***
## Group1_22    -2.656e-01  1.243e-01 -2.136 0.032697 *
## Group2_9      1.868e+00  1.899e-01  9.838 < 2e-16 ***
## Group2_14     2.365e+00  5.447e-01  4.342 1.44e-05 ***
## Group2_22     8.151e-01  1.161e-01  7.021 2.43e-12 ***
## Group3_8      6.694e+00  1.751e+00  3.822 0.000133 ***
## Group3_10    -4.291e+00  1.381e+00 -3.108 0.001894 **
## Group3_12    -4.023e+00  1.440e+00 -2.794 0.005221 **
## Group11_6    -1.157e+00  4.646e-01 -2.490 0.012812 *
## Group6_6      2.314e+00  1.899e-01 12.190 < 2e-16 ***
## Group9_24    -5.163e+03  1.279e+03 -4.038 5.45e-05 ***
## Group9_25     1.214e+01  3.017e+00  4.024 5.80e-05 ***
## Group9_26    -1.604e+00  3.990e-01 -4.021 5.86e-05 ***
## Group9_28     1.753e+03  4.300e+02  4.076 4.63e-05 ***
## Group13_6    -1.572e+00  3.069e-01 -5.123 3.10e-07 ***
## Group13_11   -1.668e-01  6.521e-02 -2.557 0.010578 *
## Group13_13   -9.727e-01  1.590e-01 -6.117 1.01e-09 ***
## Group13_20    4.418e-02  1.015e-02  4.351 1.37e-05 ***
## Group13_34    5.111e-02  8.364e-03  6.110 1.05e-09 ***
## Group13_46    6.143e-01  3.199e-01  1.920 0.054872 .
## Group13_47   -1.200e+00  3.519e-01 -3.410 0.000653 ***
## Group13_49    2.599e-01  4.339e-02  5.991 2.20e-09 ***
## Group13_50   -5.945e+00  6.477e-01 -9.180 < 2e-16 ***
## Group13_52    4.012e-05  2.255e-05  1.779 0.075344 .
## Group13_56    2.052e-01  6.911e-02  2.969 0.002995 **
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.639 on 6333 degrees of freedom
## Multiple R-squared:  0.6371, Adjusted R-squared:  0.6357
## F-statistic: 427.7 on 26 and 6333 DF,  p-value: < 2.2e-16

# According to the results of step (), we choose statistically significant variables to form a linear r
model_4 <- lm(Response~Group1_9+Group1_11+Group1_22+
              Group2_9+Group2_14+Group2_22+Group3_8+Group3_10+Group3_12+
              Group11_6+Group6_6+Group9_24+Group9_25+Group9_26+Group9_28+
              Group13_6+Group13_11+Group13_13+Group13_20+Group13_34+
              Group13_46+Group13_47+Group13_49+
              Group13_50+Group13_52+Group13_56,data = Data_remove)
# Backward selection
drop1(model_4,test='F')

## Single term deletions
##
## Model:
## Response ~ Group1_9 + Group1_11 + Group1_22 + Group2_9 + Group2_14 +
##      Group2_22 + Group3_8 + Group3_10 + Group3_12 + Group11_6 +
##      Group6_6 + Group9_24 + Group9_25 + Group9_26 + Group9_28 +
##      Group13_6 + Group13_11 + Group13_13 + Group13_20 + Group13_34 +
##      Group13_46 + Group13_47 + Group13_49 + Group13_50 + Group13_52 +
##      Group13_56
##      Df Sum of Sq    RSS   AIC  F value    Pr(>F)
## <none>                 369521 25890
## Group1_9      1      322.4 369843 25893   5.5250 0.0187768 *
## Group1_11     1     2745.8 372267 25935  47.0585 7.544e-12 ***
## Group1_22     1      266.3 369787 25892   4.5636 0.0326971 *
## Group2_9      1     5647.1 375168 25984  96.7815 < 2.2e-16 ***
## Group2_14     1     1099.8 370621 25906  18.8496 1.436e-05 ***
## Group2_22     1     2876.2 372397 25937  49.2934 2.434e-12 ***
## Group3_8      1      852.5 370374 25902  14.6111 0.0001334 ***
## Group3_10     1      563.5 370085 25897   9.6578 0.0018938 **
## Group3_12     1      455.5 369977 25895   7.8066 0.0052211 **
## Group11_6     1      361.7 369883 25894   6.1984 0.0128119 *
## Group6_6      1     8669.7 378191 26035 148.5854 < 2.2e-16 ***
## Group9_24     1      951.4 370472 25904  16.3053 5.455e-05 ***
## Group9_25     1      944.7 370466 25904  16.1899 5.797e-05 ***
## Group9_26     1      943.4 370465 25904  16.1690 5.861e-05 ***
## Group9_28     1      969.6 370491 25904  16.6176 4.628e-05 ***
## Group13_6     1     1531.3 371052 25914  26.2435 3.099e-07 ***
## Group13_11    1      381.5 369903 25894   6.5388 0.0105779 *
## Group13_13    1     2183.1 371704 25925  37.4152 1.012e-09 ***
## Group13_20    1     1104.8 370626 25906  18.9343 1.374e-05 ***
## Group13_34    1     2178.6 371700 25925  37.3372 1.053e-09 ***
## Group13_46    1      215.1 369736 25891   3.6873 0.0548723 .
## Group13_47    1      678.6 370200 25899  11.6308 0.0006527 ***
## Group13_49    1     2094.1 371615 25923  35.8899 2.203e-09 ***
## Group13_50    1     4916.8 374438 25972  84.2662 < 2.2e-16 ***
## Group13_52    1      184.6 369706 25891   3.1636 0.0753438 .
## Group13_56    1      514.5 370036 25896   8.8175 0.0029947 **
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# remove 13_46, 13_52
model_4_1 <- lm(Response~Group1_9+Group1_11+Group1_22+
  Group2_9+Group2_14+Group2_22+Group3_8+Group3_10+Group3_12+
  Group11_6+Group6_6+Group9_24+Group9_25+Group9_26+Group9_28+
  Group13_6+Group13_11+Group13_13+Group13_20+Group13_34+
  Group13_47+Group13_49+Group13_50+Group13_56
  ,data =Data_remove)
# 24 variables
drop1(model_4_1,test='F')
```

```
## Single term deletions
```

```
##
## Model:
## Response ~ Group1_9 + Group1_11 + Group1_22 + Group2_9 + Group2_14 +
##      Group2_22 + Group3_8 + Group3_10 + Group3_12 + Group11_6 +
##      Group6_6 + Group9_24 + Group9_25 + Group9_26 + Group9_28 +
##      Group13_6 + Group13_11 + Group13_13 + Group13_20 + Group13_34 +
##      Group13_47 + Group13_49 + Group13_50 + Group13_56
##      Df Sum of Sq    RSS   AIC  F value    Pr(>F)
## <none>                 369917 25892
## Group1_9      1      388.3 370306 25897    6.6506 0.0099346 **
## Group1_11     1     2587.3 372505 25935   44.3086 3.040e-11 ***
## Group1_22     1      249.3 370167 25895    4.2700 0.0388323 *
## Group2_9      1     5743.4 375661 25988   98.3579 < 2.2e-16 ***
## Group2_14     1     1081.8 370999 25909   18.5265 1.701e-05 ***
## Group2_22     1     2777.9 372695 25938   47.5734 5.812e-12 ***
## Group3_8      1      873.6 370791 25905   14.9614 0.0001108 ***
## Group3_10     1      588.8 370506 25900   10.0826 0.0015039 **
## Group3_12     1      471.6 370389 25898    8.0764 0.0044989 **
## Group11_6     1      364.1 370281 25897    6.2350 0.0125499 *
## Group6_6      1     8724.6 378642 26039  149.4130 < 2.2e-16 ***
## Group9_24     1      972.0 370889 25907   16.6461 4.560e-05 ***
## Group9_25     1      965.2 370883 25907   16.5298 4.847e-05 ***
## Group9_26     1      960.4 370878 25907   16.4477 5.061e-05 ***
## Group9_28     1      977.9 370895 25907   16.7465 4.325e-05 ***
## Group13_6     1     1370.5 371288 25914   23.4706 1.299e-06 ***
## Group13_11    1      328.7 370246 25896    5.6290 0.0176952 *
## Group13_13    1     2164.5 372082 25927   37.0683 1.207e-09 ***
## Group13_20    1      949.8 370867 25907   16.2663 5.568e-05 ***
## Group13_34    1     2099.8 372017 25926   35.9601 2.125e-09 ***
## Group13_47    1     1573.0 371490 25917   26.9386 2.166e-07 ***
## Group13_49    1     1926.1 371843 25923   32.9861 9.713e-09 ***
## Group13_50    1     4865.6 374783 25973   83.3263 < 2.2e-16 ***
## Group13_56    1      586.3 370504 25900   10.0401 0.0015390 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
vif(model_4_1)
```

```
##      Group1_9      Group1_11      Group1_22      Group2_9      Group2_14      Group2_22
```



```
## 7.602562e+01 7.619006e+01 1.114888e+00 4.677825e+00 2.867005e+00 3.346865e+00
##      Group3_8      Group3_10      Group3_12      Group11_6      Group6_6      Group9_24
## 1.406773e+02 3.810679e+01 9.690060e+01 1.166870e+00 1.473624e+00 1.232454e+06
##      Group9_25      Group9_26      Group9_28      Group13_6      Group13_11      Group13_13
## 3.463310e+06 1.110844e+06 9.068891e+04 1.455987e+01 2.276685e+01 2.094296e+00
##      Group13_20      Group13_34      Group13_47      Group13_49      Group13_50      Group13_56
## 1.864251e+01 1.706010e+01 2.362963e+01 5.148840e+01 7.715562e+01 1.823752e+00
```

```
# 1_9,1_22,11_6,13_11 can be removed
# The result of backward selection is similar to that of manual selection
# The current linear model can only provide reference
```

```
## PCA
# Next, data frame composed of the variables in the previous linear regression model is used to continue
Data_PCA <- select(Data_remove,Group1_11,Group2_9,Group2_14,
                  Group2_22,Group3_8,Group3_10,Group3_12,Group6_6,
                  Group9_24,Group9_25,Group9_26,Group9_28,Group13_6,
                  Group13_11,Group13_13,Group13_34,Group13_47,Group13_49,
                  Group13_50,Group13_56,Response)
# The Response variable is removed before PCA
PCA <- princomp(Data_PCA[, -21], cor = TRUE)
summary(PCA)
```

```
## Importance of components:
##              Comp.1    Comp.2    Comp.3    Comp.4    Comp.5
## Standard deviation  2.2416598 1.9567282 1.6133648 1.4866511 1.27266518
## Proportion of Variance 0.2512519 0.1914393 0.1301473 0.1105066 0.08098383
## Cumulative Proportion 0.2512519 0.4426912 0.5728385 0.6833451 0.76432889
##              Comp.6    Comp.7    Comp.8    Comp.9    Comp.10
## Standard deviation  1.23691906 0.95656108 0.86927865 0.78198331 0.57035850
## Proportion of Variance 0.07649844 0.04575045 0.03778227 0.03057489 0.01626544
## Cumulative Proportion 0.84082733 0.88657779 0.92436005 0.95493495 0.97120039
##              Comp.11    Comp.12    Comp.13    Comp.14
## Standard deviation  0.49240675 0.443214128 0.256405212 0.164394139
## Proportion of Variance 0.01212322 0.009821938 0.003287182 0.001351272
## Cumulative Proportion 0.98332361 0.993145548 0.996432730 0.997784001
##              Comp.15    Comp.16    Comp.17    Comp.18
## Standard deviation  0.144374779 0.1040051323 0.0824600888 0.0668232698
## Proportion of Variance 0.001042204 0.0005408534 0.0003399833 0.0002232675
## Cumulative Proportion 0.998826205 0.9993670586 0.9997070419 0.9999303094
##              Comp.19    Comp.20
## Standard deviation  3.733152e-02 4.120714e-04
## Proportion of Variance 6.968213e-05 8.490142e-09
## Cumulative Proportion 1.000000e+00 1.000000e+00
```

```
# Standard deviation: Standard deviation
# Proportion of Variance: Contribution rate of each principal component
# Cumulative Proportion: Cumulative contribution rate of principal components

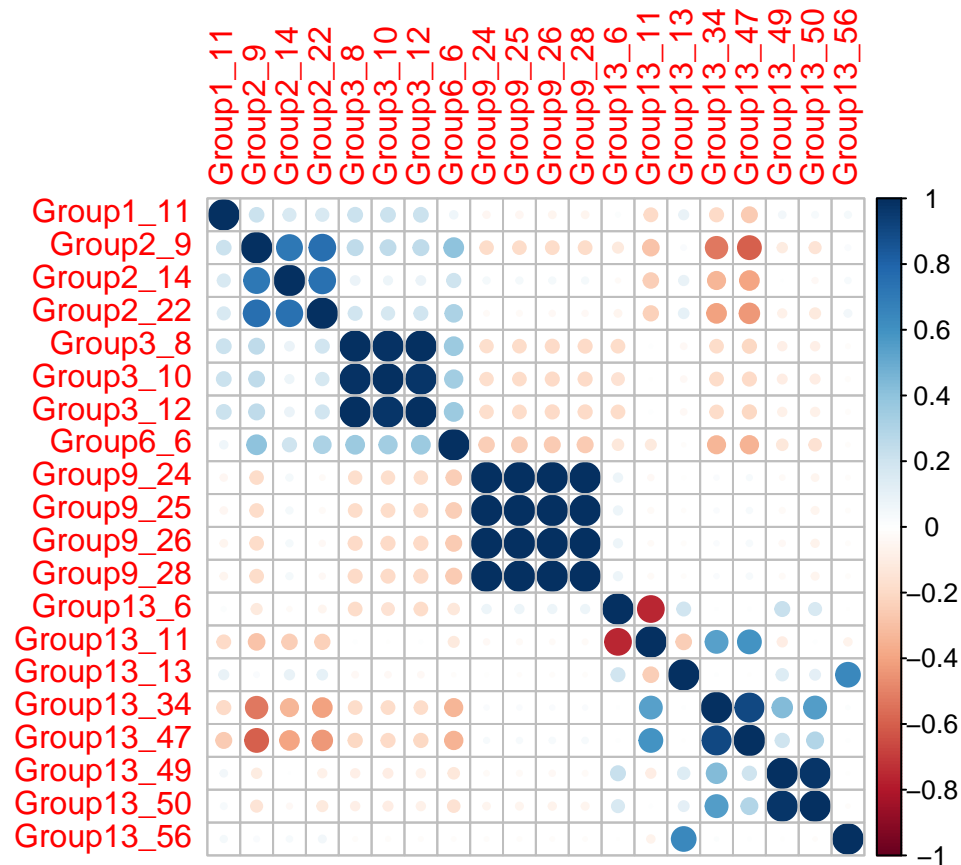
# Generally, the principal components with cumulative contribution rate over 85% are selected
# So we choose the first seven principal components
```

```
library(corrplot)
```

```
## Warning: package 'corrplot' was built under R version 4.0.3
```

```
## corrplot 0.84 loaded
```

```
# Correlation matrix
correlation <- cor(Data_PCA[,-21])
corrplot(correlation,)
```



```
# From this figure, we can see that the variables of Group 2, Group 3 and Group 9 are highly correlated
# The correlation between variables in different groups is not obvious
```

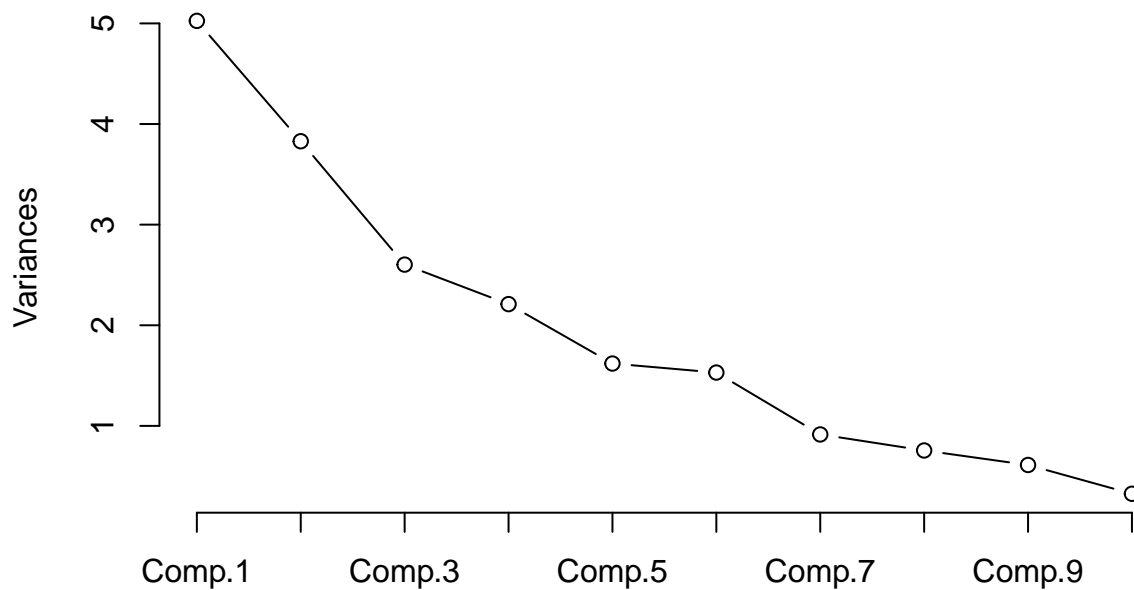
```
# Output load matrix
# The linear relationship between principal components and original variables can be obtained
loadings(PCA)
```

```
##
## Loadings:
##      Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8 Comp.9
## Group1_11  0.129      0.185      0.870      0.412
## Group2_9   0.299  0.211  0.151   -0.317
## Group2_14  0.171  0.252  0.187   -0.406 -0.115      0.306
```

## Group2_22	0.227	0.244	0.147		-0.356	-0.140	-0.127	0.261	
## Group3_8	0.291		-0.374	0.274	0.123				
## Group3_10	0.288		-0.370	0.268	0.144			0.102	
## Group3_12	0.291		-0.372	0.273	0.122				
## Group6_6	0.261						-0.317	-0.557	0.709
## Group9_24	-0.285	0.362	-0.150	0.109					
## Group9_25	-0.287	0.362	-0.148	0.103					
## Group9_26	-0.290	0.361	-0.145						
## Group9_28	-0.290	0.360	-0.144						
## Group13_6			0.326	0.197	0.380	0.311	-0.249	0.377	0.255
## Group13_11	-0.105	-0.219	-0.372	-0.143	-0.359	-0.279			
## Group13_13			0.237	0.233	0.277	-0.536			
## Group13_34	-0.246	-0.309	-0.107	0.218	-0.169	-0.107		0.238	0.251
## Group13_47	-0.252	-0.297	-0.157			-0.148	-0.143	0.371	0.339
## Group13_49		-0.156	0.193	0.510	-0.209	0.148		-0.279	-0.158
## Group13_50	-0.107	-0.197	0.156	0.499	-0.244	0.124		-0.235	-0.124
## Group13_56			0.136	0.120	0.252	-0.655		-0.114	
##	Comp.10	Comp.11	Comp.12	Comp.13	Comp.14	Comp.15	Comp.16	Comp.17	
## Group1_11									
## Group2_9	0.168	0.100	-0.827						
## Group2_14		-0.722	0.255						
## Group2_22		0.675	0.406						
## Group3_8					-0.105	0.228			
## Group3_10					0.302	-0.733			
## Group3_12					-0.190	0.512			
## Group6_6									
## Group9_24							-0.200	0.680	
## Group9_25								0.210	
## Group9_26							0.107	-0.469	
## Group9_28							0.156	-0.433	
## Group13_6	0.149			0.105	0.494	0.226			
## Group13_11				0.189	0.665	0.275			
## Group13_13	-0.702		-0.159						
## Group13_34			-0.162	-0.701			0.284		
## Group13_47				0.558	-0.410	-0.114	-0.133		
## Group13_49				0.342			0.585	0.183	
## Group13_50				-0.125			-0.683	-0.211	
## Group13_56	0.661		0.118						
##	Comp.18	Comp.19	Comp.20						
## Group1_11									
## Group2_9									
## Group2_14									
## Group2_22									
## Group3_8	-0.776								
## Group3_10	0.170								
## Group3_12	0.607								
## Group6_6									
## Group9_24		0.185	-0.457						
## Group9_25		-0.332	0.766						
## Group9_26		-0.575	-0.434						
## Group9_28		0.723	0.124						
## Group13_6									
## Group13_11									
## Group13_13									

```
## Group13_34
## Group13_47
## Group13_49
## Group13_50
## Group13_56
##
##          Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8 Comp.9
## SS loadings      1.00  1.00  1.00  1.00  1.00  1.00  1.00  1.00  1.00
## Proportion Var   0.05  0.05  0.05  0.05  0.05  0.05  0.05  0.05  0.05
## Cumulative Var   0.05  0.10  0.15  0.20  0.25  0.30  0.35  0.40  0.45
##          Comp.10 Comp.11 Comp.12 Comp.13 Comp.14 Comp.15 Comp.16 Comp.17
## SS loadings      1.00  1.00  1.00  1.00  1.00  1.00  1.00  1.00
## Proportion Var   0.05  0.05  0.05  0.05  0.05  0.05  0.05  0.05
## Cumulative Var   0.50  0.55  0.60  0.65  0.70  0.75  0.80  0.85
##          Comp.18 Comp.19 Comp.20
## SS loadings      1.00  1.00  1.00
## Proportion Var   0.05  0.05  0.05
## Cumulative Var   0.90  0.95  1.00
```

```
# Scree plot
screeplot(PCA,type = 'lines',main = '')
```



Ridge Regression, LASSO, Logistic regression can also be used for regression modeling.