Booklet of Code and Output
for
STAD29/STA 1007 Final Exam

List of Figures in this document by page:

List of Figures

```r
library(ggbiplot)
library(MASS)
library(tidyverse)
library(car)
library(ggrepel)
```

Figure 1: Packages

```r
rats %>% sample_n(20)

## # A tibble: 20 x 3
##    dose    age resttime
##    <fct> <dbl>    <dbl>
##  1 0         8       59
##  2 0         7       65
##  3 0        15       52
##  4 20       13      175
##  5 10       10       98
##  6 10       10      102
##  7 30        6      144
##  8 30       15      248
##  9 10        6       91
## 10 20       16      200
## 11 20       10      153
## 12 10       14      122
## 13 0         6       53
## 14 0        11       62
## 15 30       13      219
## 16 30       14      234
## 17 0        16       53
## 18 0        12       55
## 19 20        9      126
## 20 20        7      110
```

Figure 2: Rat lethargy data (some)

```
rats.1=lm(resttime~dose*age, data=rats)
anova(rats.1)

## Analysis of Variance Table
##
## Response: resttime
##           Df Sum Sq Mean Sq F value    Pr(>F)
## dose       3 170643   56881 913.774 < 2.2e-16 ***
## age        1  36099   36099 579.921 < 2.2e-16 ***
## dose:age   3  15750    5250  84.339 < 2.2e-16 ***
## Residuals 52   3237      62
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 3: Rat lethargy analysis of covariance

```
## # A tibble: 27 x 4
##    treatment time  subject     y
##    <chr>     <chr> <chr>   <dbl>
##  1 A         T1    S1         10
##  2 A         T1    S2         12
##  3 A         T1    S3         13
##  4 A         T2    S1         16
##  5 A         T2    S2         19
##  6 A         T2    S3         20
##  7 A         T3    S1         25
##  8 A         T3    S2         27
##  9 A         T3    S3         28
## 10 B         T1    S4         12
## 11 B         T1    S5         11
## 12 B         T1    S6         10
## 13 B         T2    S4         18
## 14 B         T2    S5         20
## 15 B         T2    S6         22
## 16 B         T3    S4         25
## 17 B         T3    S5         26
## 18 B         T3    S6         27
## 19 C         T1    S7         10
## 20 C         T1    S8         12
## 21 C         T1    S9         13
## 22 C         T2    S7         22
## 23 C         T2    S8         23
## 24 C         T2    S9         22
## 25 C         T3    S7         31
## 26 C         T3    S8         34
## 27 C         T3    S9         33
```

Figure 4: Repeated measures data

```
##
## Type II Repeated Measures MANOVA Tests: Pillai test statistic
##                 Df test stat approx F num Df den Df    Pr(>F)
## (Intercept)      1   0.99751  2399.02      1      6 4.857e-09 ***
## treatment        2   0.70412     7.14      2      6  0.025902 *
## times            1   0.99876  2010.30      2      5 5.437e-08 ***
## treatment:times  2   1.34513     6.16      4     12  0.006206 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 5: Repeated measures MANOVA

```
## # A tibble: 20 x 6
##    vanadium  iron beryllium saturated aromatic zone
##       <dbl> <dbl>     <dbl>     <dbl>    <dbl> <chr>
##  1      8.4  18        0.2       4.38     7.98 Upper
##  2      6.7  52        0.5       4.8      3.2  Upper
##  3      3.5  46        0.1       7.81    12.6  Wilhelm
##  4      9    27        0.3       3.69     3.3  Upper
##  5      6.3  13        0.5       4.24     8.27 Upper
##  6      2.7  35        0         5.11     9    Wilhelm
##  7      1.7   5.6      1         5.69     4.64 Upper
##  8      5.6  20        0.5       5.07     6.7  Upper
##  9      9    17        0.2       4.39     8.33 Upper
## 10      7.3  15        0.05      3.76     6.84 Upper
## 11      3.4  32        0.2       5.82     4.69 SubMuli
## 12     11    20        0.5       4.27     8.4  Upper
## 13      6.2  27        0.3       3.97     2.97 Upper
## 14      4    12        0.5       5.71     6.32 Upper
## 15      2.7  49        0.07      7.14    12.2  Wilhelm
## 16      3.9  51        0.2       7.06    12.2  Wilhelm
## 17      9.5  19        0.5       3.72     7.37 Upper
## 18      5    47        0.07      7.06     6.1  SubMuli
## 19      2.8  36        0.3       7       11.3  Wilhelm
## 20      3.9  36        0.07      6.19     2.27 SubMuli
```

Figure 6: Crude oil data (random sample)

```
response=with(crude, cbind(iron, beryllium, saturated, aromatic))
crude.1=lm(response~zone, data=crude)
Manova(crude.1)

##
## Type II MANOVA Tests: Pillai test statistic
##      Df test stat approx F num Df den Df   Pr(>F)
## zone  2    1.1278   16.488      8    102 1.93e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 7: Crude oil MANOVA

```
crude.2=lda(zone~iron+beryllium+saturated+aromatic, data=crude)
crude.2

## Call:
## lda(zone ~ iron + beryllium + saturated + aromatic, data = crude)
##
## Prior probabilities of groups:
##   SubMuli     Upper   Wilhelm
## 0.1964286 0.6785714 0.1250000
##
## Group means:
##             iron beryllium saturated  aromatic
## SubMuli 33.09091 0.1709091  6.560909  5.483636
## Upper   22.25263 0.4321053  4.658158  5.767895
## Wilhelm 43.57143 0.1171429  6.795714 11.540000
##
## Coefficients of linear discriminants:
##                  LD1         LD2
## iron       0.0611089  0.05039847
## beryllium -2.7160984  1.63910398
## saturated  0.7735772 -0.77701517
## aromatic   0.1025370  0.39908518
##
## Proportion of trace:
##    LD1    LD2
## 0.8246 0.1754
```

Figure 8: Crude oil discriminant analysis

```
##      r    zone   class p.SubMuli p.Upper p.Wilhelm
## 1   1 Wilhelm Wilhelm     0.001   0.000     0.999
## 2   2 Wilhelm Wilhelm     0.002   0.000     0.998
## 3   3 Wilhelm Wilhelm     0.101   0.008     0.891
## 4   4 Wilhelm Wilhelm     0.002   0.000     0.998
## 5   5 Wilhelm Wilhelm     0.004   0.000     0.996
## 6   6 Wilhelm Wilhelm     0.034   0.001     0.964
## 7   7 Wilhelm Wilhelm     0.239   0.281     0.480
## 8   8 SubMuli SubMuli     0.850   0.000     0.150
## 9   9 SubMuli SubMuli     0.764   0.234     0.002
## 10 10 SubMuli SubMuli     0.684   0.316     0.000
## 11 11 SubMuli SubMuli     0.937   0.063     0.000
## 12 12 SubMuli SubMuli     0.999   0.000     0.001
## 13 13 SubMuli   Upper     0.226   0.774     0.000
## 14 14 SubMuli SubMuli     0.948   0.049     0.003
## 15 15 SubMuli SubMuli     0.992   0.008     0.000
## 16 16 SubMuli Wilhelm     0.085   0.001     0.914
## 17 17 SubMuli SubMuli     0.942   0.000     0.058
## 18 18 SubMuli Wilhelm     0.103   0.326     0.571
## 19 19   Upper   Upper     0.000   1.000     0.000
## 20 20   Upper   Upper     0.000   1.000     0.000
## 21 21   Upper   Upper     0.120   0.880     0.000
## 22 22   Upper   Upper     0.000   1.000     0.000
## 23 23   Upper   Upper     0.002   0.998     0.000
## 24 24   Upper   Upper     0.000   1.000     0.000
## 25 25   Upper   Upper     0.001   0.999     0.000
## 26 26   Upper   Upper     0.001   0.999     0.000
## 27 27   Upper   Upper     0.000   1.000     0.000
## 28 28   Upper   Upper     0.001   0.999     0.000
## 29 29   Upper   Upper     0.003   0.997     0.000
## 30 30   Upper   Upper     0.000   1.000     0.000
## 31 31   Upper   Upper     0.002   0.998     0.000
## 32 32   Upper   Upper     0.001   0.999     0.000
## 33 33   Upper   Upper     0.008   0.991     0.001
## 34 34   Upper   Upper     0.002   0.997     0.000
## 35 35   Upper   Upper     0.001   0.999     0.000
## 36 36   Upper   Upper     0.000   1.000     0.000
## 37 37   Upper   Upper     0.010   0.990     0.000
## 38 38   Upper   Upper     0.056   0.938     0.006
## 39 39   Upper   Upper     0.001   0.999     0.000
## 40 40   Upper   Upper     0.000   1.000     0.000
## 41 41   Upper   Upper     0.000   1.000     0.000
## 42 42   Upper SubMuli     0.801   0.186     0.013
## 43 43   Upper   Upper     0.002   0.998     0.000
## 44 44   Upper   Upper     0.002   0.998     0.000
## 45 45   Upper   Upper     0.004   0.996     0.000
## 46 46   Upper   Upper     0.000   1.000     0.000
## 47 47   Upper   Upper     0.011   0.983     0.005
## 48 48   Upper   Upper     0.018   0.982     0.000
## 49 49   Upper   Upper     0.001   0.999     0.000
## 50 50   Upper   Upper     0.164   0.836     0.000
## 51 51   Upper SubMuli     0.531   0.468     0.000
## 52 52   Upper   Upper     0.057   0.943     0.000
## 53 53   Upper   Upper     0.006   0.994     0.000
## 54 54   Upper   Upper     0.082   0.918     0.000
## 55 55   Upper   Upper     0.000   1.000     0.000
## 56 56   Upper   Upper     0.003   0.997     0.000
```

Figure 9: Crude oil posterior probabilities

```r
speakers=read_delim("loudspeaker.txt", " ")

## Parsed with column specification:
## cols(
##   id = col_character(),
##   price = col_double(),
##   accuracy = col_double(),
##   bass = col_double(),
##   power = col_double()
## )

speakers

## # A tibble: 19 x 5
##    id    price accuracy  bass power
##    <chr> <dbl>    <dbl> <dbl> <dbl>
##  1 A       600       91     5    38
##  2 B       598       92     4    18
##  3 C       550       90     4    36
##  4 D       500       90     4    29
##  5 E       630       90     4    15
##  6 F       580       87     5     5
##  7 G       460       87     5    15
##  8 H       600       88     4    29
##  9 I       590       88     3    15
## 10 J       599       89     3    23
## 11 K       598       85     2    23
## 12 L       618       84     2    12
## 13 M       600       88     3    46
## 14 N       600       82     3    29
## 15 O       600       85     2    36
## 16 P       500       83     2    45
## 17 Q       539       80     1    23
## 18 R       569       86     1    21
## 19 S       680       79     2    36
```

Figure 10: Loudspeakers data

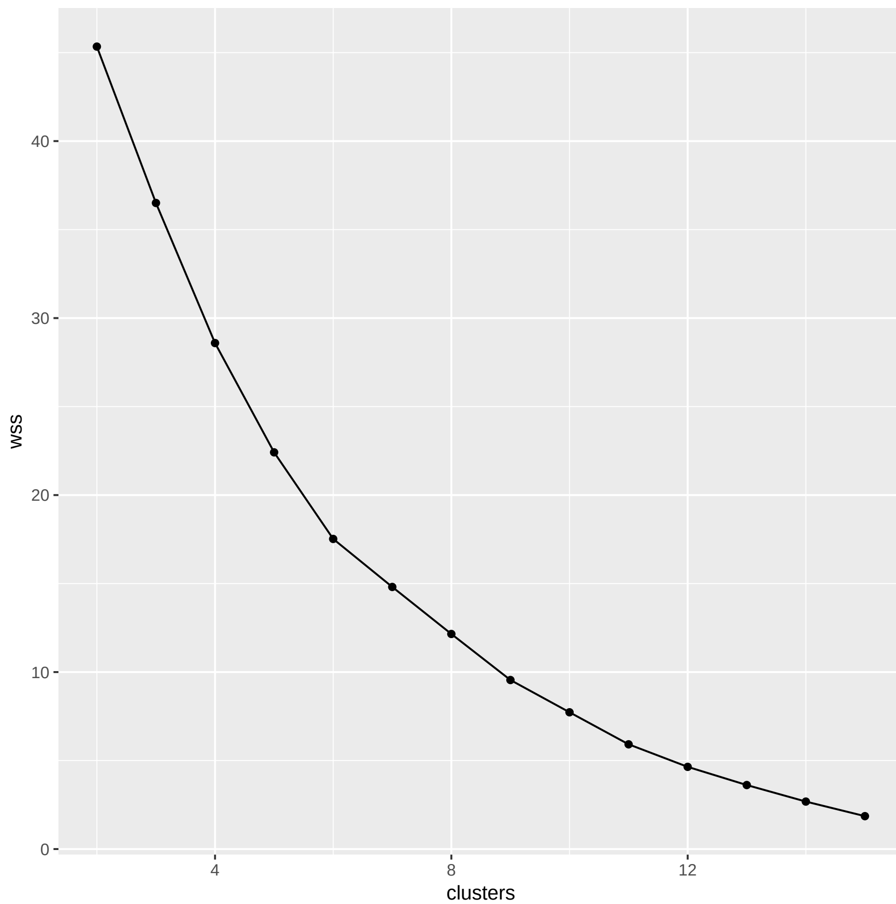

Figure 11: Loudspeakers scree plot

```
##    sp sex index   FL   RW   CL   CW   BD
## 1   B   M     4  9.6  7.9 20.1 23.1  8.2
## 2   B   M    12 12.3 11.0 26.8 31.5 11.4
## 3   B   M    17 13.1 10.6 28.2 32.3 11.0
## 4   B   M    23 15.0 10.9 31.4 36.4 13.2
## 5   B   M    24 15.0 11.5 32.4 37.0 13.4
## 6   B   M    25 15.0 11.9 32.5 37.2 13.6
## 7   B   M    34 16.4 13.0 35.7 41.8 15.2
## 8   B   M    36 16.8 12.8 36.2 41.8 14.9
## 9   B   M    39 17.1 12.7 36.7 41.9 15.6
## 10  B   M    45 19.3 13.5 41.6 47.4 17.8
## 11  B   M    48 19.8 14.2 43.2 49.7 18.6
## 12  B   M    50 21.3 15.7 47.1 54.6 20.0
## 13  B   F    53  9.1  8.1 18.5 21.6  7.7
## 14  B   F    57 10.1  9.3 20.9 24.4  8.4
## 15  B   F    64 11.6 11.0 24.6 28.5 10.4
## 16  B   F    71 12.8 11.7 27.1 31.2 11.9
## 17  B   F    79 13.9 13.0 30.0 34.9 13.1
## 18  B   F    95 16.2 15.2 34.5 40.1 13.9
## 19  B   F   100 19.2 16.5 40.9 47.9 18.1
## 20  O   M   101  9.1  6.9 16.7 18.6  7.4
## 21  O   M   102 10.2  8.2 20.2 22.2  9.0
## 22  O   M   103 10.7  8.6 20.7 22.7  9.2
## 23  O   M   111 14.0 11.5 29.2 32.2 13.1
## 24  O   M   113 14.1 10.5 29.1 31.6 13.1
## 25  O   M   114 14.1 10.7 28.7 31.9 13.3
## 26  O   M   116 14.2 10.7 27.8 30.9 12.7
## 27  O   M   128 17.5 12.0 34.4 37.3 15.3
## 28  O   M   131 17.9 12.9 36.9 40.9 16.5
## 29  O   M   134 18.4 13.4 37.9 42.2 17.7
## 30  O   M   145 21.6 15.4 45.7 49.7 20.6
## 31  O   M   148 22.1 15.8 44.6 49.6 20.5
## 32  O   M   149 23.0 16.8 47.2 52.1 21.5
## 33  O   M   150 23.1 15.7 47.6 52.8 21.6
## 34  O   F   151 10.7  9.7 21.4 24.0  9.8
## 35  O   F   152 11.4  9.2 21.7 24.1  9.7
## 36  O   F   156 14.0 11.9 27.0 31.4 12.6
## 37  O   F   157 14.0 12.8 28.8 32.4 12.7
## 38  O   F   158 14.3 12.2 28.1 31.8 12.5
## 39  O   F   161 15.0 12.3 30.1 33.3 14.0
## 40  O   F   183 18.9 16.7 36.3 41.7 15.3
```

Figure 12: Crabs data (sample)

```
crabs %>% select_if(is.double) %>%
  princomp(cor=T) -> crabs.1
summary(crabs.1)

## Importance of components:
##                           Comp.1     Comp.2      Comp.3      Comp.4
## Standard deviation     2.188341 0.38946785 0.215946693 0.105524202
## Proportion of Variance 0.957767 0.03033704 0.009326595 0.002227071
## Cumulative Proportion  0.957767 0.98810400 0.997430593 0.999657664
##                              Comp.5
## Standard deviation     0.0413724263
## Proportion of Variance 0.0003423355
## Cumulative Proportion  1.0000000000
```

Figure 13: Crabs principal components analysis

```
ggscreeplot(crabs.1)
```

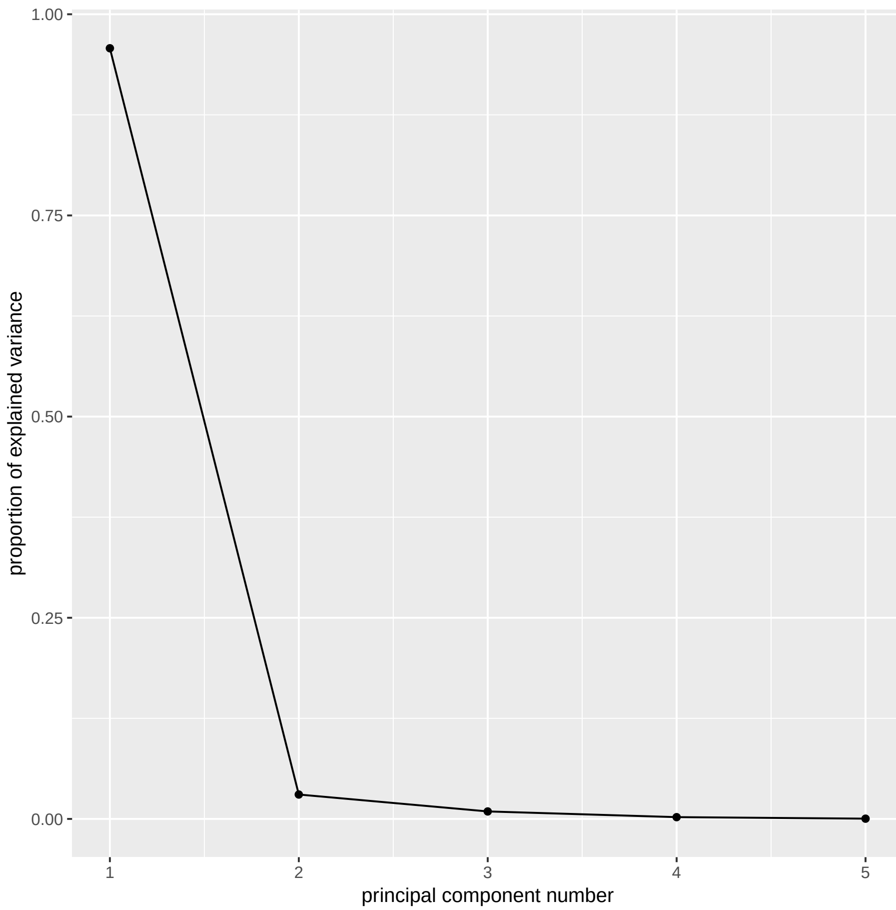

Figure 14: Crabs scree plot

```
crabs.1$loadings

##
## Loadings:
##    Comp.1 Comp.2 Comp.3 Comp.4 Comp.5
## FL  0.452  0.138  0.531  0.697
## RW  0.428 -0.898
## CL  0.453  0.268 -0.310        -0.792
## CW  0.451  0.181 -0.653         0.575
## BD  0.451  0.264  0.443 -0.707  0.176
##
##                Comp.1 Comp.2 Comp.3 Comp.4 Comp.5
## SS loadings       1.0    1.0    1.0    1.0    1.0
## Proportion Var    0.2    0.2    0.2    0.2    0.2
## Cumulative Var    0.2    0.4    0.6    0.8    1.0
```

Figure 15: Crabs principal component loadings

```r
d_crabs=cbind(crabs, crabs.1$scores)
ggplot(d_crabs, aes(x=Comp.1, y=Comp.2, label=index))+ geom_text()
```

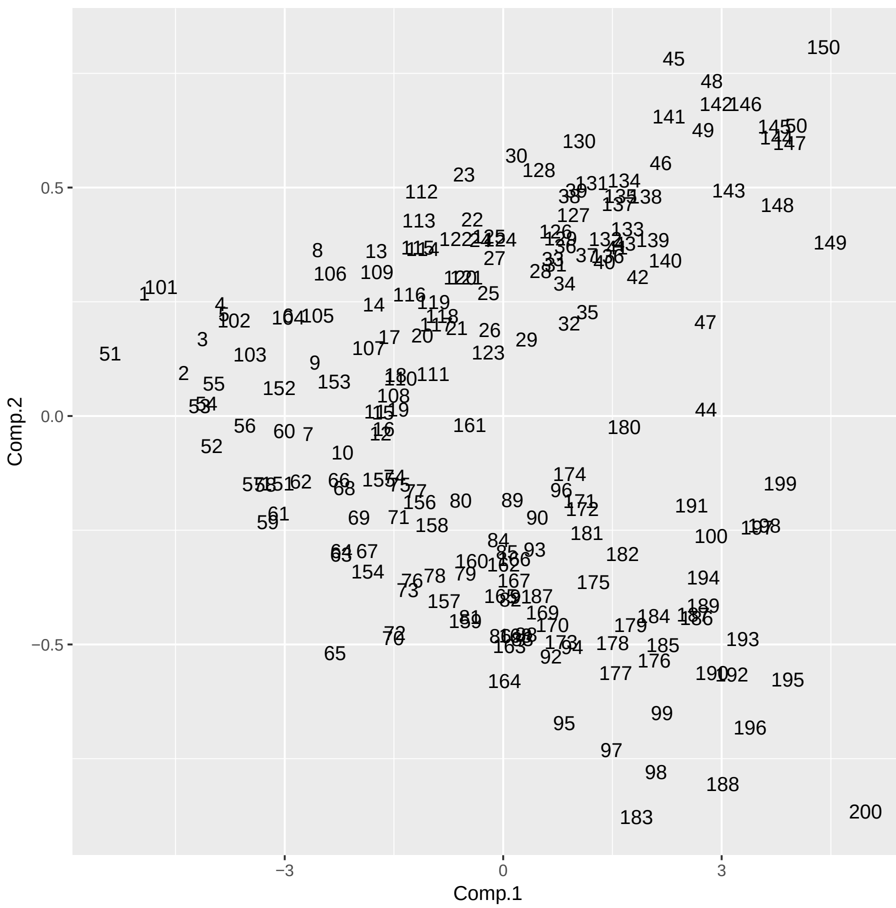

Figure 16: Crabs plot of component scores

```
hothand=read_csv("hothand.csv")
hothand %>% print(n=Inf)

## # A tibble: 36 x 4
##    Player           first_shot second_shot frequency
##    <chr>            <chr>      <chr>           <dbl>
##  1 Larry Bird       hit        hit               251
##  2 Larry Bird       hit        miss               34
##  3 Larry Bird       miss       hit                48
##  4 Larry Bird       miss       miss                5
##  5 Cedric Maxwell   hit        hit               245
##  6 Cedric Maxwell   hit        miss               57
##  7 Cedric Maxwell   miss       hit                97
##  8 Cedric Maxwell   miss       miss               31
##  9 Robert Parish    hit        hit               164
## 10 Robert Parish    hit        miss               49
## 11 Robert Parish    miss       hit                76
## 12 Robert Parish    miss       miss               29
## 13 Tiny Archibald   hit        hit               203
## 14 Tiny Archibald   hit        miss               42
## 15 Tiny Archibald   miss       hit                62
## 16 Tiny Archibald   miss       miss               14
## 17 Chris Ford       hit        hit                36
## 18 Chris Ford       hit        miss               15
## 19 Chris Ford       miss       hit                17
## 20 Chris Ford       miss       miss                5
## 21 Kevin McHale     hit        hit                93
## 22 Kevin McHale     hit        miss               35
## 23 Kevin McHale     miss       hit                29
## 24 Kevin McHale     miss       miss               20
## 25 ML Carr          hit        hit                39
## 26 ML Carr          hit        miss               18
## 27 ML Carr          miss       hit                21
## 28 ML Carr          miss       miss                5
## 29 Rick Robey       hit        hit                54
## 30 Rick Robey       hit        miss               37
## 31 Rick Robey       miss       hit                49
## 32 Rick Robey       miss       miss               31
## 33 Gerald Henderson hit        hit                77
## 34 Gerald Henderson hit        miss               24
## 35 Gerald Henderson miss       hit                29
## 36 Gerald Henderson miss       miss                8
```

Figure 17: Hot hand data

The columns of the output from the first two of these code chunks refer to the *second* shot: whether it is hit or missed.

```
hothand %>% count(first_shot, second_shot, wt=frequency) %>%
  group_by(first_shot) %>%
  mutate(proportion=n/sum(n)) %>%
  select(-n) %>%
  spread(second_shot, proportion)

## # A tibble: 2 x 3
## # Groups:   first_shot [2]
##   first_shot   hit  miss
##   <chr>      <dbl> <dbl>
## 1 hit        0.789 0.211
## 2 miss       0.743 0.257
```

```
hothand %>% count(first_shot, second_shot, wt=frequency) %>%
  spread(second_shot, n) -> d
d

## # A tibble: 2 x 3
##   first_shot   hit  miss
##   <chr>      <dbl> <dbl>
## 1 hit         1162   311
## 2 miss         428   148
```

```
d %>% select(-first_shot) %>%
  chisq.test()

##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  .
## X-squared = 4.739, df = 1, p-value = 0.02949
```

Figure 18: Hot hand chi-squared test

```
hothand %>% group_by(Player, first_shot) %>%
    count(second_shot, wt=frequency) %>%
    mutate(proportion=n/sum(n)) %>% filter(second_shot=="hit") %>%
    select(-n) %>% select(-second_shot) %>%
    spread(first_shot, proportion)

## # A tibble: 9 x 3
## # Groups:   Player [9]
##   Player              hit  miss
##   <chr>             <dbl> <dbl>
## 1 Cedric Maxwell    0.811 0.758
## 2 Chris Ford        0.706 0.773
## 3 Gerald Henderson 0.762 0.784
## 4 Kevin McHale      0.727 0.592
## 5 Larry Bird        0.881 0.906
## 6 ML Carr           0.684 0.808
## 7 Rick Robey        0.593 0.612
## 8 Robert Parish     0.770 0.724
## 9 Tiny Archibald    0.829 0.816
```

Figure 19: Proportion of second shots made for each player when first shot is hit or missed

```
hothand.1=glm(frequency~Player*first_shot*second_shot,
              family="poisson", data=hothand)
drop1(hothand.1, test="Chisq")

## Single term deletions
##
## Model:
## frequency ~ Player * first_shot * second_shot
##                              Df Deviance    AIC    LRT Pr(>Chi)
## <none>                            0.0000 267.31
## Player:first_shot:second_shot  8   6.6502 257.96 6.6502   0.5748
```

Figure 20: Log-linear analysis part 1

```r
hothand.2=update(hothand.1, .~.-Player:first_shot:second_shot)
drop1(hothand.2, test="Chisq")

## Single term deletions
##
## Model:
## frequency ~ Player + first_shot + second_shot + Player:first_shot +
##     Player:second_shot + first_shot:second_shot
##                        Df Deviance    AIC    LRT  Pr(>Chi)
## <none>                       6.650 257.96
## Player:first_shot       8   66.587 301.90 59.937 4.795e-10 ***
## Player:second_shot      8   71.056 306.37 64.405 6.326e-11 ***
## first_shot:second_shot  1    7.521 256.83  0.870    0.3508
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 21: Log-linear analysis part 2

```r
hothand.3=update(hothand.2, .~.-first_shot:second_shot)
drop1(hothand.3, test="Chisq")

## Single term deletions
##
## Model:
## frequency ~ Player + first_shot + second_shot + Player:first_shot +
##     Player:second_shot
##                    Df Deviance    AIC    LRT  Pr(>Chi)
## <none>                   7.521 256.83
## Player:first_shot   8   71.490 304.81 63.970 7.712e-11 ***
## Player:second_shot  8   75.959 309.27 68.438 1.005e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 22: Log-linear analysis part 3

```r
ggplot(rats, aes(x=age, y=resttime, colour=dose)) +
    geom_point() + geom_smooth(method="lm", se=F)
```

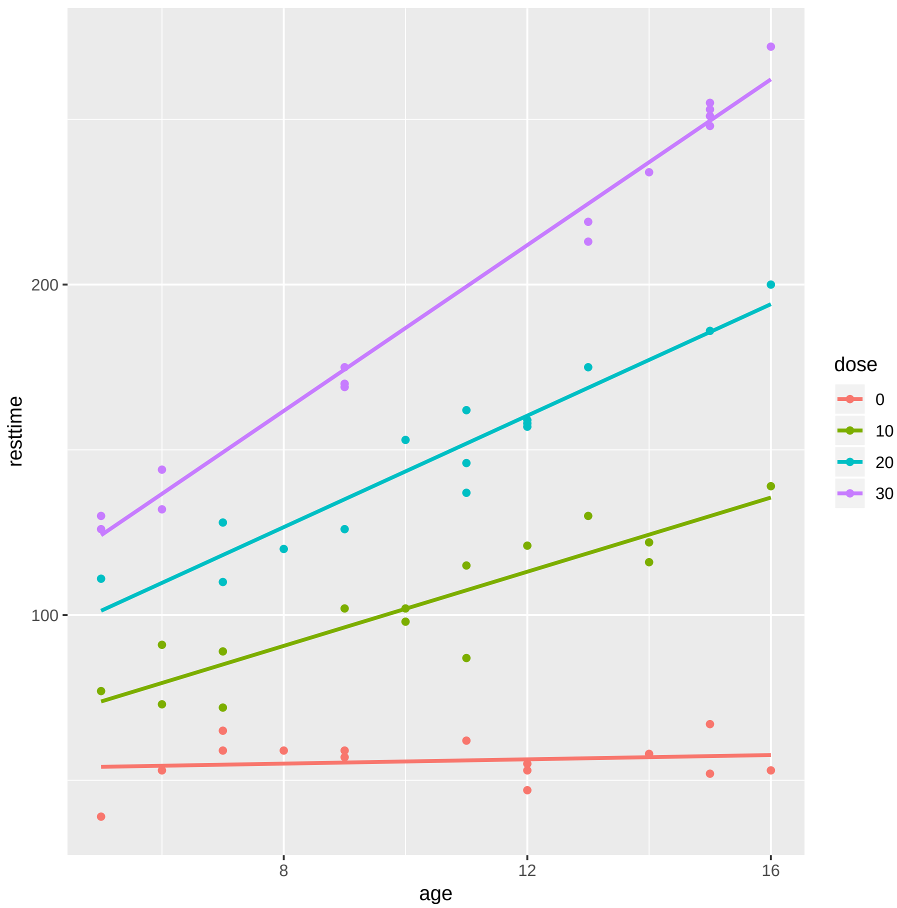

Figure 23: Rat lethargy data scatterplot

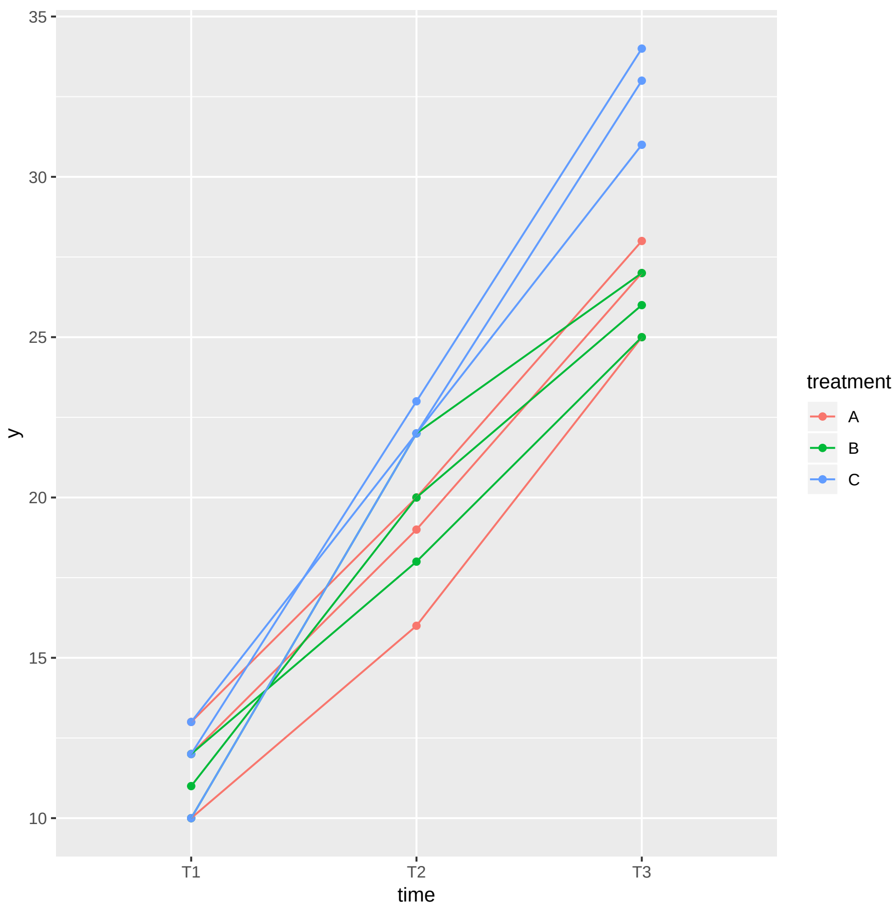

Figure 24: Repeated measures spaghetti plot

```
ggplot(d, aes(x=x.LD1, y=x.LD2, colour=zone, label=r)) +
    geom_point() + geom_text_repel()
```

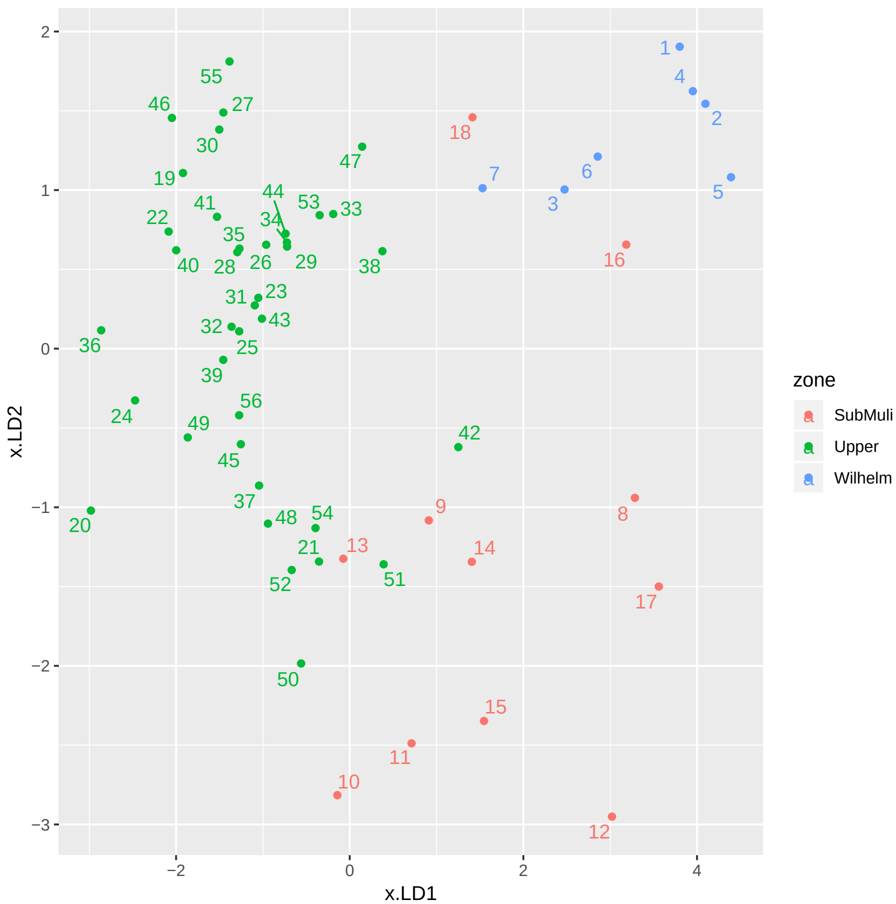

Figure 25: Crude oil LD plot

```r
ggplot(d_crabs, aes(x=Comp.1, y=Comp.2, colour=sex))+geom_point()
```

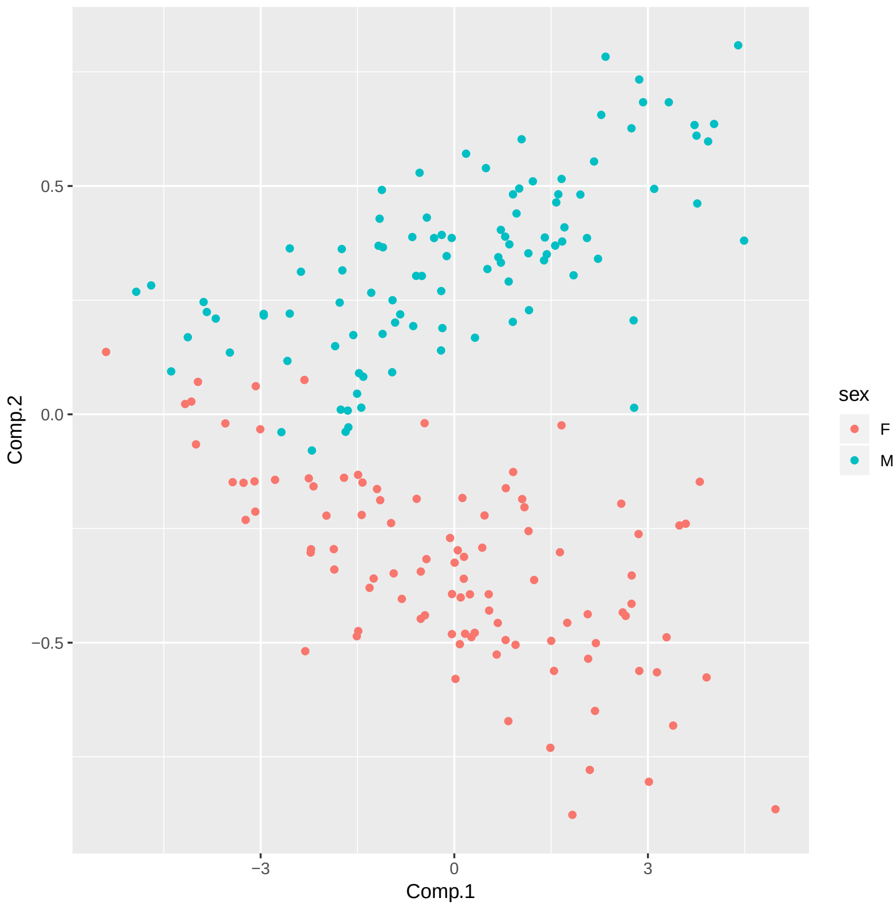

Figure 26: Another plot of component scores