```r
library(MASS)
library(ggbiplot)
library(tidyverse)
library(marginaleffects)
library(lme4)
library(car)
```

Figure 1: Packages loaded

low	lwt	smoke
no	113	no
no	95	yes
no	113	yes
no	120	no
no	190	no
no	131	no
yes	120	yes
no	121	yes
no	120	yes
no	130	no
no	90	yes
no	105	yes
no	113	no
no	160	no
no	121	yes

Figure 2: Low birth weight data (randomly chosen rows)

```r
birthwt.1 <- glm(factor(low) ~ lwt + smoke, data = birthwt, family = "binomial")
summary(birthwt.1)
```

```

Call:
glm(formula = factor(low) ~ lwt + smoke, family = "binomial",
    data = birthwt)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.62200    0.79592   0.781   0.4345
lwt         -0.01332    0.00609  -2.188   0.0287 *
smokeyes     0.67667    0.32470   2.084   0.0372 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 234.67  on 188  degrees of freedom
Residual deviance: 224.34  on 186  degrees of freedom
AIC: 230.34

Number of Fisher Scoring iterations: 4
```

Figure 3: Low birth weight logistic regression

```
new <- datagrid(model = birthwt.1, lwt = c(110, 125, 140))
new
```

low	smoke	lwt
no	no	110
no	no	125
no	no	140

```
cbind(predictions(birthwt.1, newdata = new)) %>%
  select(smoke, lwt, estimate)
```

smoke	lwt	estimate
no	110	0.3007605
no	125	0.2604667
no	140	0.2238432

```
new <- datagrid(model = birthwt.1, smoke = c("no", "yes"))
new
```

low	lwt	smoke
no	130	no
no	130	yes

```
cbind(predictions(birthwt.1, newdata = new)) %>%
  select(smoke, lwt, estimate)
```

smoke	lwt	estimate
no	130	0.2478400
yes	130	0.3932927

Figure 4: Low birth weight predictions

time	panel	emergenc
17	1	1
14	1	1
15	2	1
12	2	1
21	3	1
24	3	1
25	1	2
24	1	2
22	2	2
19	2	2
29	3	2
28	3	2
31	1	3
24	1	3
28	2	3
31	2	3
32	3	3
37	3	3
14	1	4
13	1	4
9	2	4
10	2	4
15	3	4
19	3	4

Figure 5: Display panels data

```
display.1 <- aov(time ~ panel * emergenc, data = display)
summary(display.1)
```

```
               Df Sum Sq Mean Sq F value   Pr(>F)    
panel           2  232.7   116.4  20.094 0.000148 ***
emergenc        3 1052.5   350.8  60.573 1.61e-07 ***
panel:emergenc  6   28.9     4.8   0.832 0.567501    
Residuals      12   69.5     5.8                     
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
display.2 <- aov(time ~ panel + emergenc, data = display)
summary(display.2)
```

```
            Df Sum Sq Mean Sq F value   Pr(>F)    
panel        2  232.7   116.4   21.29 1.81e-05 ***
emergenc     3 1052.5   350.8   64.16 8.28e-10 ***
Residuals   18   98.4     5.5                     
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 6: Display panels: two analyses

```
TukeyHSD(display.2)
```

```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = time ~ panel + emergenc, data = display)

$panel
      diff       lwr        upr     p adj
2-1 -2.000 -4.983847  0.9838467 0.2284740
3-1  5.375  2.391153  8.3588467 0.0006236
3-2  7.375  4.391153 10.3588467 0.0000173

$emergenc
          diff        lwr          upr     p adj
2-1   7.333333   3.517810  11.14885716 0.0001980
3-1  13.333333   9.517810  17.14885716 0.0000001
4-1  -3.833333  -7.648857  -0.01780951 0.0487057
3-2   6.000000   2.184476   9.81552383 0.0016255
4-2 -11.166667 -14.982190  -7.35114284 0.0000009
4-3 -17.166667 -20.982190 -13.35114284 0.0000000
```

Figure 7: Display panels: Tukey analysis

year	education	vocabulary
1974	8	3
1994	12	5
1974	12	7
1974	15	5
1994	12	3
1974	13	10
1994	20	10
2014	14	6
2014	12	6
1994	12	5
1994	12	7
2014	19	6
2014	16	7
2014	20	9
1994	12	5
1994	12	9
1994	16	7
1974	14	9
1974	12	1
1974	18	9

Figure 8: Education and vocabulary data (some randomly chosen rows)

```r
ggplot(vocab, aes(x = education, y = vocabulary, colour = factor(year))) +
  geom_jitter() + geom_smooth(method = "lm")
```

```
`geom_smooth()` using formula = 'y ~ x'
```

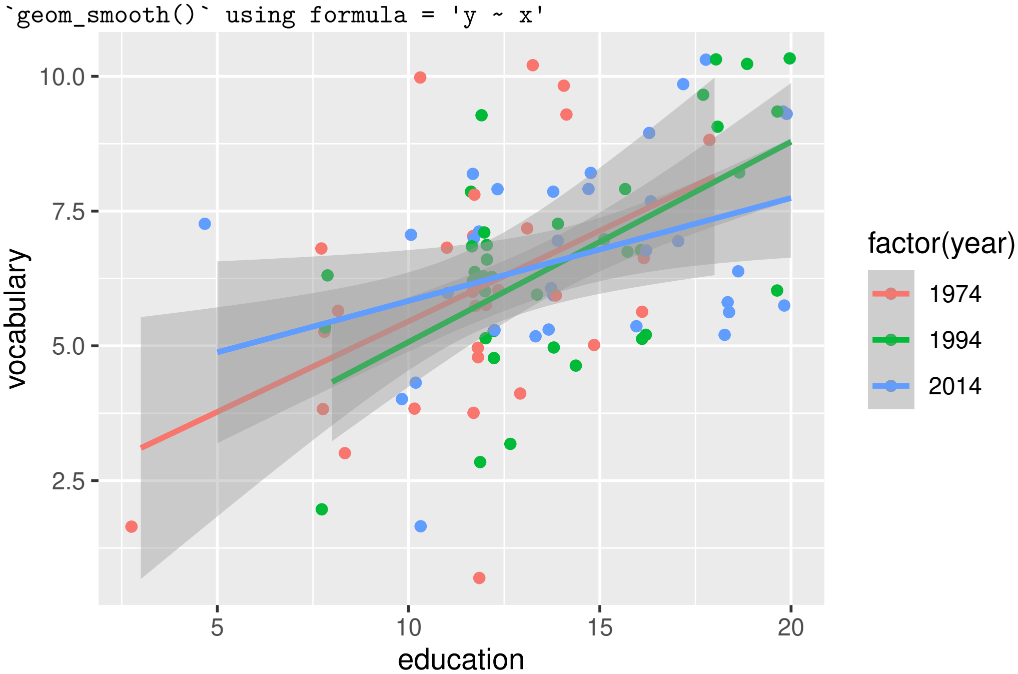


Figure 9: Scatterplot of education and vocabulary by year

```r
vocab.1 <- lm(vocabulary ~ education * factor(year), data = vocab)
summary(vocab.1)
```

```
Call:
lm(formula = vocabulary ~ education * factor(year), data = vocab)

Residuals:
    Min      1Q  Median      3Q     Max 
-5.1272 -1.3613  0.0457  1.2094  4.5447 

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)   
(Intercept)                 2.09583    1.32000   1.588  0.11570   
education                   0.33594    0.10909   3.080  0.00272 **
factor(year)1994           -0.73778    1.83257  -0.403  0.68816   
factor(year)2014            1.83017    1.84771   0.991  0.32447   
education:factor(year)1994  0.03545    0.14032   0.253  0.80111   
education:factor(year)2014 -0.14510    0.13968  -1.039  0.30158   
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.758 on 94 degrees of freedom
Multiple R-squared:  0.2675,    Adjusted R-squared:  0.2285 
F-statistic: 6.866 on 5 and 94 DF,  p-value: 1.679e-05
```

```r
drop1(vocab.1, test = "F")
```

	Df	Sum of Sq	RSS	AIC	F value	Pr(>F)
	NA	NA	290.6165	118.6834	NA	NA
education:factor(year)	2	7.157061	297.7735	117.1163	1.157477	0.3187163

Figure 10: Education and vocabulary analysis of covariance

```r
vocab.2 <- lm(vocabulary ~ education + factor(year), data = vocab)
summary(vocab.2)
```

```

Call:
lm(formula = vocabulary ~ education + factor(year), data = vocab)

Residuals:
    Min      1Q  Median      3Q     Max 
-5.1155 -1.3130  0.0278  1.1293  4.4720 

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)       2.59073    0.71278   3.635  0.00045 ***
education         0.29373    0.05402   5.438  4.1e-07 ***
factor(year)1994 -0.14335    0.45417  -0.316  0.75297    
factor(year)2014 -0.14745    0.46826  -0.315  0.75352    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.761 on 96 degrees of freedom
Multiple R-squared:  0.2495,	Adjusted R-squared:  0.226 
F-statistic: 10.64 on 3 and 96 DF,  p-value: 4.221e-06
```

Figure 11: Education and vocabulary analysis of covariance 2

animal	trt	day	weight
A1	A	0	233
A1	A	14	224
A1	A	28	245
A1	A	42	258
A1	A	56	271
A1	A	70	287
A1	A	84	287
A1	A	98	287
A1	A	112	290
A1	A	126	293
A1	A	133	297
A10	A	0	232
A10	A	14	240
A10	A	28	247
A10	A	42	263
A10	A	56	275
A10	A	70	286
A10	A	84	294
A10	A	98	302
A10	A	112	308
A10	A	126	319
A10	A	133	326
A11	A	0	234
A11	A	14	237
A11	A	28	259

Figure 12: Cattle data (some)

```
cattle %>%
  group_by(trt, day) %>%
  summarize(mean_weight = mean(weight)) %>%
  ggplot(aes(x = day, y = mean_weight, colour = trt, group = trt)) +
    geom_point() + geom_line()
```

```
`summarise()` has grouped output by 'trt'. You can override using the `.groups`
argument.
```

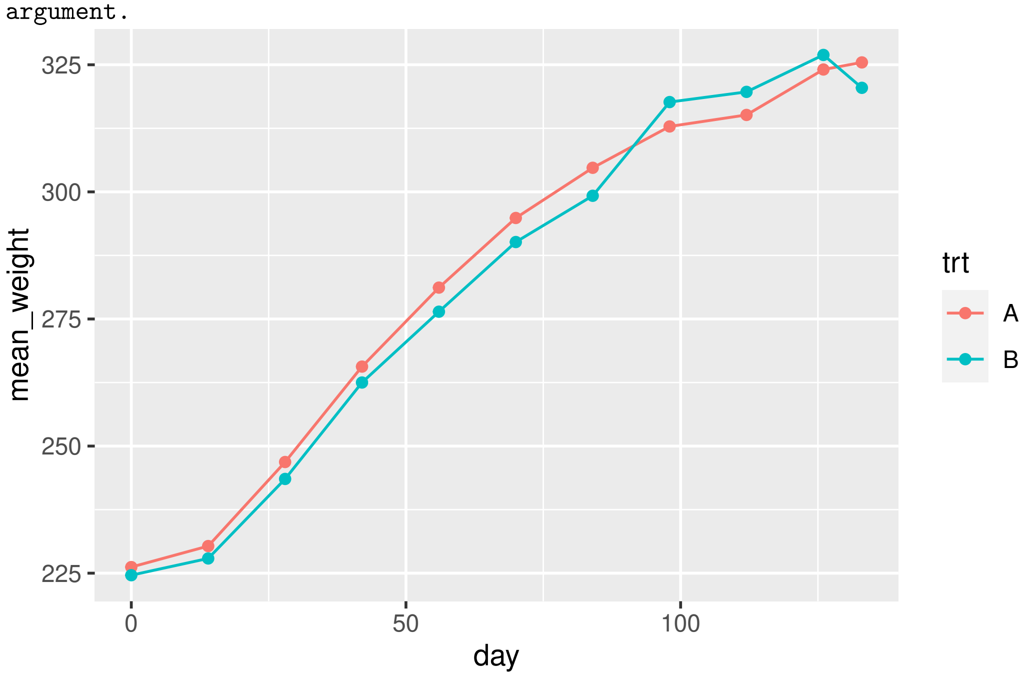


Figure 13: Cattle data interaction plot

```
cattle.1 <- lmer(weight ~ trt * factor(day) + (1 | animal), data = cattle)
drop1(cattle.1, test = "Chisq")
```

	npar	AIC	LRT	Pr(Chi)
	NA	4866.584	NA	NA
trt:factor(day)	10	4881.634	35.04987	0.0001224

```
cattle.1a <- lmer(weight ~ trt + factor(day) + (1 | animal), data = cattle)
drop1(cattle.1a, test = "Chisq")
```

	npar	AIC	LRT	Pr(Chi)
	NA	4881.634	NA	NA
trt	1	4879.839	0.2047196	0.650938
factor(day)	10	6721.837	1860.2032319	0.000000

Figure 14: Cattle data mixed model analysis

```
# pivot wider
cattle %>%
  pivot_wider(names_from = day, values_from = weight) -> cattle_wider
# set up for manova
cattle_wider %>% select(-animal, -trt) %>%
  as.matrix() -> response
cattle.2 <- lm(response ~ trt, data = cattle_wider)
times <- colnames(response)
times.df <- data.frame(times = factor(times))
cattle.3 <- Manova(cattle.2, idata = times.df, idesign = ~times)
```

Figure 15: Cattle repeated measures MANOVA code

```
summary(cattle.3)$univariate.tests
```

```
             Sum Sq num Df Error SS den Df    F value    Pr(>F)    
(Intercept) 53035479      1   133128     58 23106.1031 < 2.2e-16 ***
trt              455      1   133128     58     0.1982 0.6578077    
times         846142     10    37638    580  1303.9048 < 2.2e-16 ***
trt:times       2264     10    37638    580     3.4891 0.0001767 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(cattle.3)$sphericity
```

```
          Test statistic    p-value
times          3.399e-05 8.9223e-85
trt:times      3.399e-05 8.9223e-85
```

```
summary(cattle.3)$pval.adjustments
```

```
             GG eps   Pr(>F[GG])    HF eps    Pr(>F[HF])
times     0.2415572 2.496758e-96 0.2528023 1.100120e-100
trt:times 0.2415572 2.535322e-02 0.2528023  2.346705e-02
attr(,"na.action")
(Intercept)         trt 
          1           2 
attr(,"class")
[1] "omit"
```

Figure 16: Cattle repeated measures MANOVA output

species	nasal.length	nasal.width	ramus.height	zygomatic.width	mandible.depth	lacrymal.width
fuliginosus	552	205	751	919	194	454
giganteus	755	268	754	902	206	467
fuliginosus	574	212	641	822	191	405
giganteus	756	249	731	903	198	467
melanops	565	204	556	764	156	385
giganteus	687	223	688	873	205	432
giganteus	682	253	706	875	194	455
fuliginosus	522	190	629	799	179	374
fuliginosus	719	253	765	946	215	473
fuliginosus	554	195	657	837	188	392
melanops	893	260	824	994	216	499
fuliginosus	625	250	739	934	211	470
fuliginosus	497	167	648	807	178	390
giganteus	629	222	643	824	181	416
giganteus	616	220	652	805	180	412
melanops	800	245	813	939	240	492
fuliginosus	737	278	880	1090	271	535
melanops	690	242	708	855	210	451
giganteus	626	226	651	839	173	441
giganteus	734	245	724	920	193	462

Figure 17: Kangaroo skull data, variables of interest, randomly chosen rows

```
kanga %>%
  select(nasal.length:lacrymal.width) %>%
  as.matrix() -> response
kanga.0 <- manova(response ~ species, data = kanga)
summary(kanga.0)
```

```
          Df Pillai approx F num Df den Df    Pr(>F)
species    2 1.0065   15.873     12    188 < 2.2e-16 ***
Residuals 98
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

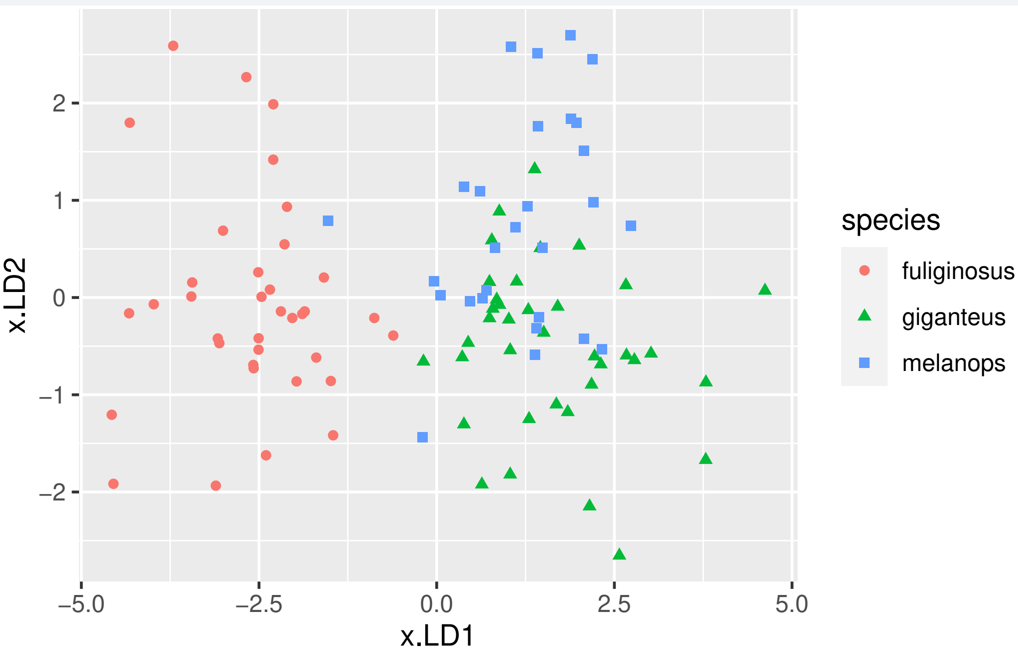
Figure 18: Kangaroo skull analysis 1

```r
kanga.1 <- lda(species ~ nasal.length + nasal.width + ramus.height +
                 zygomatic.width + mandible.depth + lacrymal.width, data = kanga)
kanga.1
```

```
Call:
lda(species ~ nasal.length + nasal.width + ramus.height + zygomatic.width +
    mandible.depth + lacrymal.width, data = kanga)

Prior probabilities of groups:
fuliginosus   giganteus    melanops
  0.3564356   0.3663366   0.2772277

Group means:
            nasal.length nasal.width ramus.height zygomatic.width
fuliginosus     614.1944    222.4722     728.5000        912.6944
giganteus       707.3243    246.8919     686.4595        869.8649
melanops        684.1071    233.2857     695.2143        860.5000
            mandible.depth lacrymal.width
fuliginosus       205.7500       442.3056
giganteus         194.3784       444.2973
melanops          194.7500       444.3214

Coefficients of linear discriminants:
                         LD1          LD2
nasal.length     0.027729200  0.003536769
nasal.width     -0.006006286 -0.067051346
ramus.height    -0.013524008  0.016312892
zygomatic.width -0.025345101 -0.032317300
mandible.depth  -0.008867705 -0.003186253
lacrymal.width   0.022430867  0.060543739

Proportion of trace:
   LD1    LD2
0.9359 0.0641
```

Figure 19: Kangaroo skulls, discriminant analysis

```
p <- predict(kanga.1)
d <- cbind(kanga, p)
ggplot(d, aes(x = x.LD1, y = x.LD2, colour = species, shape = species)) + geom_point()
```



Note that the points on the plot are distinguished by both colour and shape (plotting symbol).

Figure 20: Kangaroo skulls, further analysis

```
with(d, table(obs = species, pred = class))
```

```
             pred
obs           fuliginosus giganteus melanops
  fuliginosus          36         0        0
  giganteus             0        32        5
  melanops              1        10       17
```

Figure 21: Kangaroo skulls, a table

Baseball	Football	Basketball	Tennis	Cycling	Swimming	Jogging
7	5	1	6	4	3	2
7	5	6	3	4	1	2
7	5	6	3	4	1	2
4	1	5	7	3	2	6
6	5	7	1	2	3	4
2	5	4	1	6	7	3
3	1	5	4	6	7	2
3	4	1	7	6	5	2
2	1	3	5	4	7	6
1	3	5	7	4	6	2
4	2	3	5	6	7	1
7	4	3	2	1	5	6
3	1	4	7	6	2	5
4	7	5	6	2	3	1
6	7	5	3	2	1	4

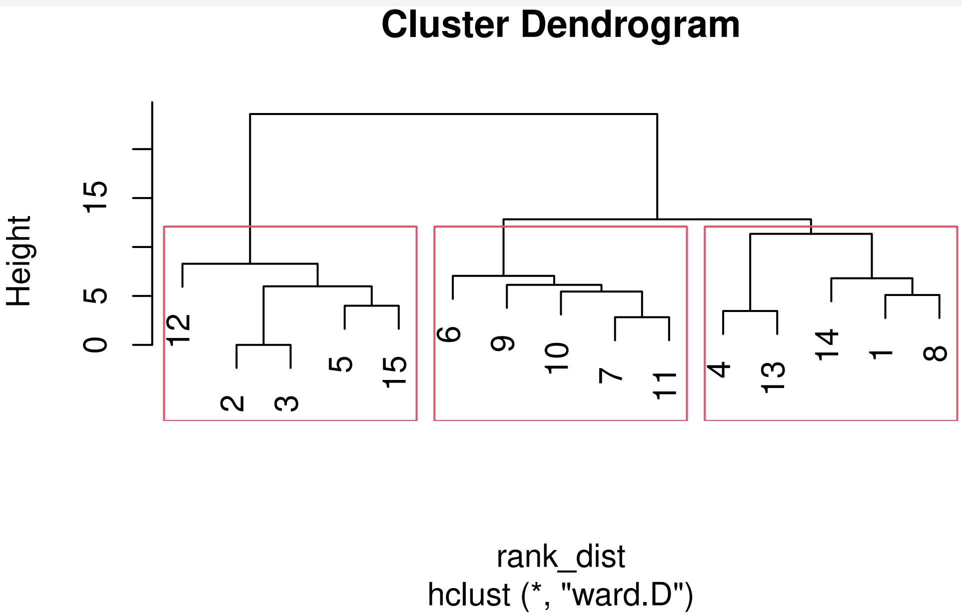
Figure 22: Sports preference data

```
my_dist <- function(d, i, j) {
  d %>% slice(i, j) %>%
    mutate(indiv = c("row1", "row2")) %>%
    pivot_longer(-indiv, names_to = "sport", values_to = "rating") %>%
    pivot_wider(names_from = indiv, values_from = rating) %>%
    summarize(diss = sqrt(sum((row1 - row2)^2))) %>%
    pull(diss)
}
```

Figure 23: Function to compute dissimilarity for sports ranking data

The data in **ranks** were first turned into a **dist** object, called **rank_dist**, and then the following code was run:

```
ranks.1 <- hclust(rank_dist, method =  "ward.D")
plot(ranks.1)
rect.hclust(ranks.1, 3)
```




```
ranks.1$merge
```

```
      [,1] [,2]
 [1,]   -2   -3
 [2,]   -7  -11
 [3,]   -4  -13
 [4,]   -5  -15
 [5,]   -1   -8
 [6,]  -10    2
 [7,]    1    4
 [8,]   -9    6
 [9,]  -14    5
[10,]   -6    8
[11,]  -12    7
[12,]    3    9
[13,]   10   12
[14,]   11   13
```

Figure 24: Cluster analysis for sports data

id	cluster	Baseball	Football	Basketball	Tennis	Cycling	Swimming	Jogging
1	1	7	5	1	6	4	3	2
4	1	4	1	5	7	3	2	6
8	1	3	4	1	7	6	5	2
13	1	3	1	4	7	6	2	5
14	1	4	7	5	6	2	3	1
2	2	7	5	6	3	4	1	2
3	2	7	5	6	3	4	1	2
5	2	6	5	7	1	2	3	4
12	2	7	4	3	2	1	5	6
15	2	6	7	5	3	2	1	4
6	3	2	5	4	1	6	7	3
7	3	3	1	5	4	6	7	2
9	3	2	1	3	5	4	7	6
10	3	1	3	5	7	4	6	2
11	3	4	2	3	5	6	7	1

Figure 25: Individuals (in column `id`) and cluster memberships for sports preference data

Carname	Length	Wheelbase	Width	Height	FrontHd	RearHd	FrtLegRoom
NissanPulsarNX	-1.3333333	-1.2	-1.0	-4.0	1	-2.0	-1
InfinitiQ45	2.3333333	2.2	2.0	0.5	-2	-0.5	0
HondaPrelude	-0.1111111	-0.2	-0.5	-3.0	-4	-2.0	2
HondaAccord	0.6666667	1.0	0.0	-0.5	2	0.0	1
FordAerostar	-0.4444444	3.4	2.0	16.5	3	1.5	1
ChevroletCavalier	0.0000000	-0.2	-1.0	0.0	1	1.0	-1
Porsche944	-1.1111111	-1.4	0.0	-4.0	0	-2.0	4
SubaruLoyale	-0.4444444	-1.0	-1.5	-1.0	-2	0.0	-1
GEOMetro	-3.2222222	-1.8	-2.5	-1.5	-2	-0.5	-1
Mazda626	0.0000000	-0.2	-0.5	-0.5	0	0.0	-2
BuickRiviera	2.1111111	1.2	2.5	-1.5	0	0.0	0
PontiacLeMans	-0.7777778	-0.6	-1.0	0.0	3	0.5	0
ChevroletLumina	2.1111111	1.2	1.5	1.0	2	1.5	1
ToyotaCelica	-0.5555556	-0.6	0.5	-3.0	-3	-2.0	1
PontiacBonneville	2.2222222	1.8	2.0	0.5	1	2.0	1

Figure 26: Cars data (some rows and columns)

```r
cars.1 <- princomp(cars_numeric, cor = TRUE)
summary(cars.1)
```

```
Importance of components:
                          Comp.1    Comp.2    Comp.3     Comp.4     Comp.5
Standard deviation     2.1859859 1.4764164 1.3821796 0.82691759 0.73804436
Proportion of Variance 0.4344122 0.1981641 0.1736746 0.06216297 0.04951904
Cumulative Proportion  0.4344122 0.6325763 0.8062509 0.86841391 0.91793295
                           Comp.6     Comp.7     Comp.8      Comp.9     Comp.10
Standard deviation     0.55143677 0.44712743 0.41450063 0.308496259 0.269616682
Proportion of Variance 0.02764386 0.01817481 0.01561916 0.008651813 0.006608469
Cumulative Proportion  0.94557681 0.96375163 0.97937079 0.988022601 0.994631070
                          Comp.11
Standard deviation     0.24301900
Proportion of Variance 0.00536893
Cumulative Proportion  1.00000000
```

```r
ggscreeplot(cars.1)
```

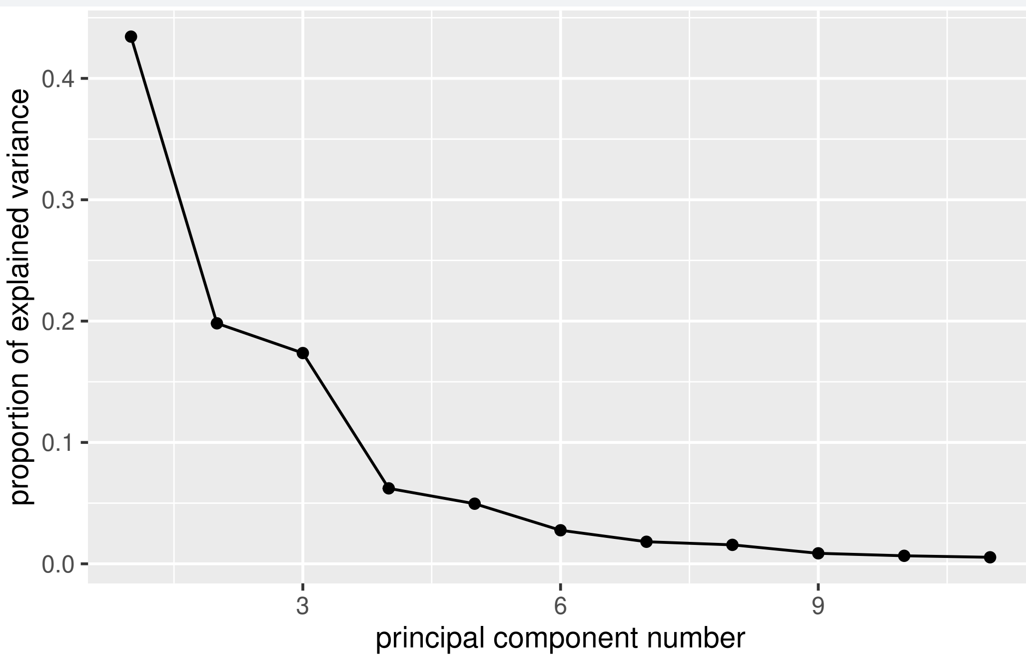


Figure 27: Cars principal components analysis and scree plot

```
              Comp.1      Comp.2      Comp.3
Length      0.36800447  0.26491429  0.15382216
Wheelbase   0.39303399  0.19208728  0.16487812
Width       0.36271313  0.08827474  0.37020532
Height      0.25231975 -0.48930844  0.04282917
FrontHd     0.27563231 -0.27414127 -0.05442509
RearHd      0.32033083 -0.37823667 -0.19691964
FrtLegRoom  0.03600793  0.41677436  0.32261159
RearSeating 0.28877687  0.13354004 -0.45373369
FrtShld     0.39776393 -0.04124462  0.27343479
RearShld    0.28585066  0.09344998 -0.46883122
Luggage     0.10803017  0.47489313 -0.40497703
```

Figure 28: Cars data: loadings of first three principal components

```
cars.2 <- factanal(cars_numeric, 3, scores = "r")
cars.2$loadings
```

```
Loadings:
            Factor1 Factor2 Factor3
Length       0.837           0.346
Wheelbase    0.810   0.172   0.309
Width        0.959   0.157
Height       0.193   0.935
FrontHd      0.318   0.481   0.150
RearHd       0.200   0.831   0.291
FrtLegRoom   0.341  -0.361
RearSeating  0.180   0.251   0.803
FrtShld      0.867   0.388
RearShld     0.123   0.343   0.822
Luggage      0.106  -0.400   0.799

               Factor1 Factor2 Factor3
SS loadings      3.382   2.472   2.295
Proportion Var   0.307   0.225   0.209
Cumulative Var   0.307   0.532   0.741
```

Figure 29: Cars data: factor analysis and factor loadings

```
cars.2$uniquenesses
```

```
     Length   Wheelbase       Width      Height     FrontHd      RearHd
 0.18046602  0.21837389  0.05494901  0.08052403  0.64476164  0.18494410
 FrtLegRoom RearSeating     FrtShld    RearShld     Luggage
 0.75282443  0.25953381  0.09399943  0.19096693  0.18980421
```

Figure 30: Cars data: uniquenesses