Figures

```
library(tidyverse)
library(MASS)
library(car)
library(ggbiplot)
library(conflicted)
```

Figure 1: Packages

```
## Rows: 72 Columns: 2
## -- Column specification ---------------------------------------------------------------------
## Delimiter: " "
## chr (2): Drug, Relapse
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
cocaine
```

```
## # A tibble: 72 x 2
##    Drug        Relapse
##    <chr>       <chr>
##  1 Desipramine no
##  2 Lithium     yes
##  3 Placebo     no
##  4 Placebo     no
##  5 Desipramine yes
##  6 Lithium     yes
##  7 Placebo     yes
##  8 Placebo     yes
##  9 Lithium     no
## 10 Lithium     yes
## # ... with 62 more rows
```

Figure 2: Cocaine treatment data (some)

```
cocaine.1 <- glm(factor(Relapse) ~ Drug, data = cocaine, family = "binomial")
summary(cocaine.1)
```

```
## 
## Call:
## glm(formula = factor(Relapse) ~ Drug, family = "binomial", data = cocaine)
## 
## Deviance Residuals: 
##     Min       1Q   Median       3Q      Max  
## -1.8930  -1.0383   0.6039   0.7585   1.3232  
## 
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)   
## (Intercept)  -0.3365     0.4140  -0.813   0.4164   
## DrugLithium   1.4351     0.6274   2.287   0.0222 * 
## DrugPlacebo   1.9459     0.6866   2.834   0.0046 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## (Dispersion parameter for binomial family taken to be 1)
## 
##     Null deviance: 91.658  on 71  degrees of freedom
## Residual deviance: 81.220  on 69  degrees of freedom
## AIC: 87.22
## 
## Number of Fisher Scoring iterations: 4
```

```
drop1(cocaine.1, test = "Chisq")
```

```
## Single term deletions
## 
## Model:
## factor(Relapse) ~ Drug
##        Df Deviance    AIC    LRT Pr(>Chi)   
## <none>      81.220 87.220                   
## Drug    2   91.658 93.658 10.438 0.005413 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 3: Cocaine treatment model

```
Wool
```

```
##    len amp load cycles
## 1  250   8   40    674
## 2  250   8   45    370
## 3  250   8   50    292
## 4  250   9   40    338
## 5  250   9   45    266
## 6  250   9   50    210
## 7  250  10   40    170
## 8  250  10   45    118
## 9  250  10   50     90
## 10 300   8   40   1414
## 11 300   8   45   1198
## 12 300   8   50    634
## 13 300   9   40   1022
## 14 300   9   45    620
## 15 300   9   50    438
## 16 300  10   40    443
## 17 300  10   45    332
## 18 300  10   50    220
## 19 350   8   40   3636
## 20 350   8   45   3184
## 21 350   8   50   2000
## 22 350   9   40   1568
## 23 350   9   45   1070
## 24 350   9   50    566
## 25 350  10   40   1140
## 26 350  10   45    884
## 27 350  10   50    360
```

Figure 4: Wool data

```
ggplot(Wool, aes(x = len, y = cycles, fill = amp)) + geom_boxplot()
```

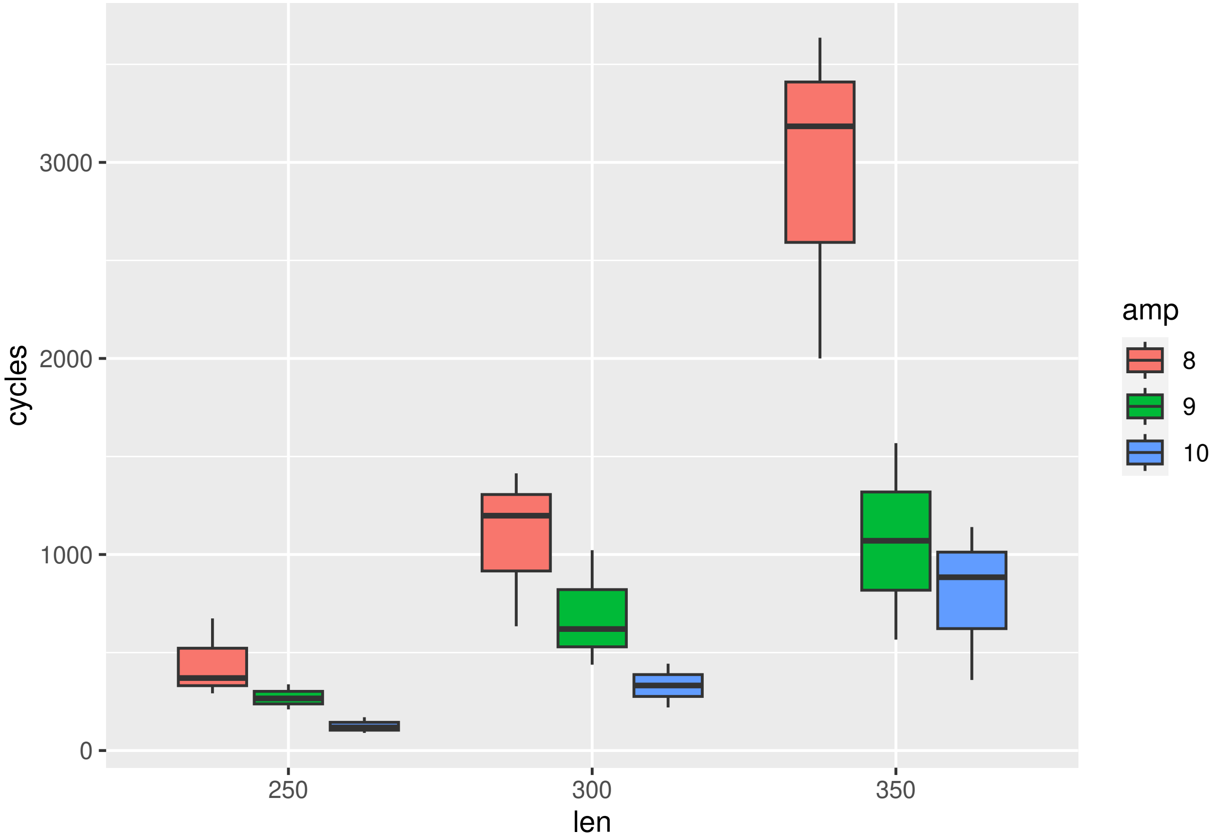


Figure 5: Wool boxplot

```
wool.1 <- aov(cycles ~ len * amp, data = Wool)
summary(wool.1)
```

```
##             Df  Sum Sq Mean Sq F value   Pr(>F)
## len          2 8182253 4091126  25.689 5.33e-06 ***
## amp          2 5624249 2812124  17.658 5.70e-05 ***
## len:amp      4 3555537  888884   5.582  0.00421 **
## Residuals   18 2866579  159254
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 6: Wool ANOVA

SE1:

```
Wool %>% filter(len == 350) -> d1
d1.1 <- aov(cycles ~ amp, data = d1)
summary(d1.1)
```

```
##             Df  Sum Sq Mean Sq F value  Pr(>F)
## amp          2 8181550 4090775   10.93 0.00999 **
## Residuals    6 2245731  374288
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
TukeyHSD(d1.1)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = cycles ~ amp, data = d1)
##
## $amp
##            diff       lwr       upr     p adj
## 9-8  -1872.0000 -3404.681 -339.3190 0.0222658
## 10-8 -2145.3333 -3678.014 -612.6523 0.0121259
## 10-9  -273.3333 -1806.014 1259.3477 0.8516592
```

SE2:

```
Wool %>% filter(len == 250) -> d2
d2.1 <- aov(cycles ~ amp, data = d2)
summary(d2.1)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## amp          2 153372   76686   4.947 0.0538 .
## Residuals    6  93005   15501
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
TukeyHSD(d2.1)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = cycles ~ amp, data = d2)
##
## $amp
##           diff       lwr        upr     p adj
## 9-8  -174.0000 -485.9083 137.908258 0.2761411
## 10-8 -319.3333 -631.2416  -7.425075 0.0457129
## 10-9 -145.3333 -457.2416 166.574925 0.3854304
```

Figure 7: Wool simple effects

```
salmon %>% slice_sample(n = 20)
```

```
##    Gender Freshwater Marine   Origin      combo
## 1       1        128    400 Canadian Canadian-1
## 2       2         92    404  Alaskan  Alaskan-2
## 3       1        154    390 Canadian Canadian-1
## 4       2        117    489  Alaskan  Alaskan-2
## 5       1        129    420 Canadian Canadian-1
## 6       1         95    433  Alaskan  Alaskan-1
## 7       2        109    325 Canadian Canadian-2
## 8       2        166    377 Canadian Canadian-2
## 9       1         76    442  Alaskan  Alaskan-1
## 10      2        115    354 Canadian Canadian-2
## 11      1        105    388  Alaskan  Alaskan-1
## 12      2         84    511  Alaskan  Alaskan-2
## 13      2        163    370 Canadian Canadian-2
## 14      2        111    422  Alaskan  Alaskan-2
## 15      2         84    399  Alaskan  Alaskan-2
## 16      1        121    403  Alaskan  Alaskan-1
## 17      1         95    411  Alaskan  Alaskan-1
## 18      2        144    345 Canadian Canadian-2
## 19      2        120    369 Canadian Canadian-2
## 20      2        140    388 Canadian Canadian-2
```

Figure 8: Salmon data (20 randomly chosen rows)

```
salmon.1 <- manova(response ~ Gender*Origin, data = salmon)
summary(salmon.1)
```

```
##               Df  Pillai approx F num Df den Df Pr(>F)
## Gender         1 0.00325    0.155      2     95 0.8568
## Origin         1 0.67939  100.657      2     95 <2e-16 ***
## Gender:Origin  1 0.03273    1.607      2     95 0.2059
## Residuals     96
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
salmon.2 <- manova(response ~ Gender + Origin, data = salmon)
summary(salmon.2)
```

```
##           Df  Pillai approx F num Df den Df Pr(>F)
## Gender     1 0.00320    0.154      2     96 0.8572
## Origin     1 0.67937  101.703      2     96 <2e-16 ***
## Residuals 97
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
salmon.3 <- manova(response ~ Origin, data = salmon)
summary(salmon.3)
```

```
##           Df Pillai approx F num Df den Df    Pr(>F)
## Origin     1  0.679   102.59      2     97 < 2.2e-16 ***
## Residuals 98
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 9: Salmon MANOVA

```
salmon.4 <- lda(combo ~ Freshwater + Marine, data = salmon)
salmon.4
```

```
## Call:
## lda(combo ~ Freshwater + Marine, data = salmon)
##
## Prior probabilities of groups:
##  Alaskan-1  Alaskan-2 Canadian-1 Canadian-2
##       0.26       0.24       0.26       0.24
##
## Group means:
##            Freshwater   Marine
## Alaskan-1    96.57692 423.6538
## Alaskan-2   100.33333 436.1667
## Canadian-1  139.53846 369.0000
## Canadian-2  135.20833 364.0417
##
## Coefficients of linear discriminants:
##                    LD1         LD2
## Freshwater  0.04419519 -0.03805305
## Marine     -0.01785288 -0.02360065
##
## Proportion of trace:
##    LD1    LD2
## 0.9836 0.0164
```

Figure 10: Salmon discriminant analysis

```
p <- predict(salmon.4)
d <- cbind(salmon, p)
```

Figure 11: Salmon discriminant analysis part 2

Note: the `cbbPalette` and the `scale_colour_manual` draw the points with (I am told) colour-blind-friendly colours. If it is still impossible for you to distinguish the colours, ask an invigilator for help identifying the colour of some points.

```
cbbPalette <- c("#000000", "#E69F00", "#56B4E9", "#009E73", "#F0E442", "#0072B2", "#D55E00", "#CC79A7")
ggplot(d, aes(x = x.LD1, y = x.LD2, colour = combo)) +
  geom_point() + scale_colour_manual(values = cbbPalette)
```

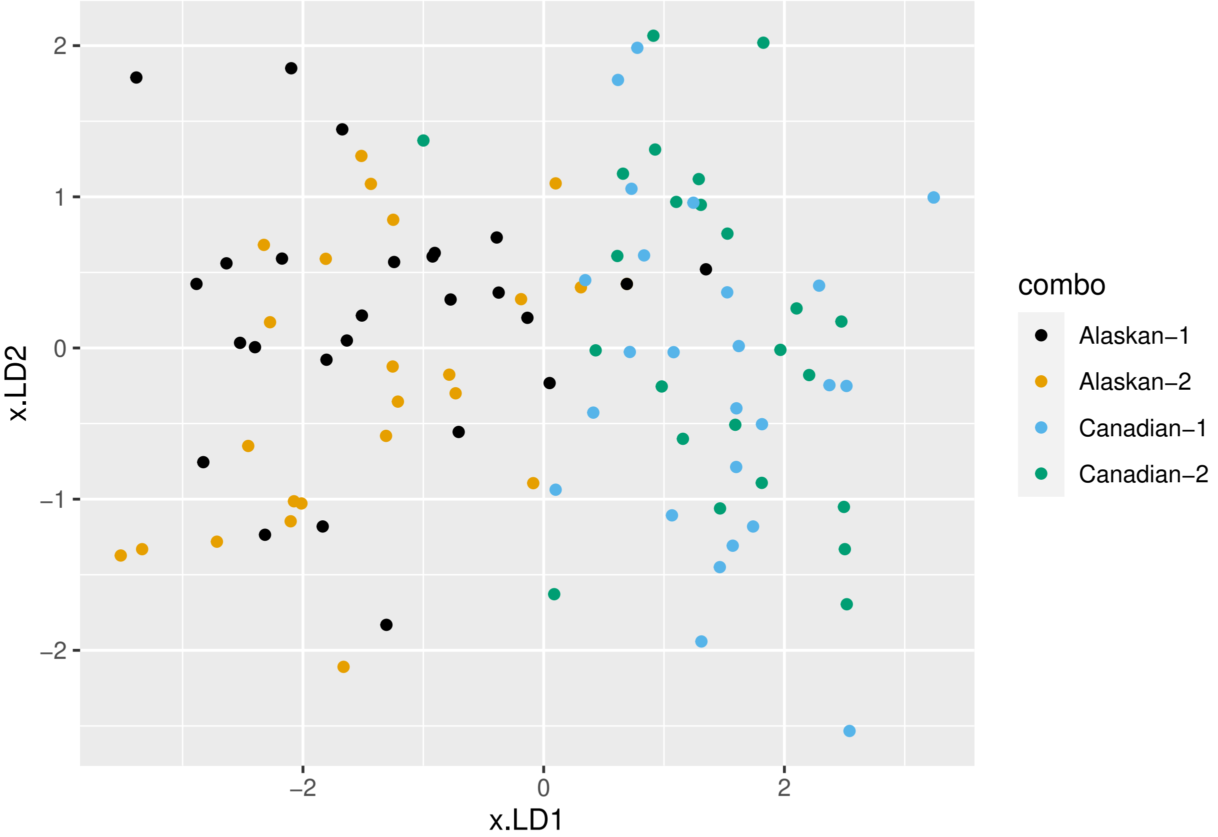


Figure 12: Salmon discriminant scores plot

```
with(d, table(combo, class))
```

```
##              class
## combo         Alaskan-1 Alaskan-2 Canadian-1 Canadian-2
##   Alaskan-1          18         5          1          2
##   Alaskan-2           9        12          0          3
##   Canadian-1          0         0         17          9
##   Canadian-2          1         0         13         10
```

Figure 13: Salmon discriminant misclassification table

```
se
```

```
##    subject treatment se1 se2 se3
## 1        1   Control  14  13  15
## 2        2   Control  13  14  17
## 3        3   Control  17  12  16
## 4        4   Control  11  11  12
## 5        5   Control  16  15  14
## 6        6   Control  17  18  18
## 7        7   Control  17  16  19
## 8        8   Control  13  15  15
## 9        9   Control  14  14  15
## 10      10   Control  14  15  13
## 11      11   Control  16  16  11
## 12      12   Control  15  13  16
## 13      13      Diet  12  11  14
## 14      14      Diet  13  14  15
## 15      15      Diet  17  11  18
## 16      16      Diet  16  15  18
## 17      17      Diet  16  17  15
## 18      18      Diet  13  11  18
## 19      19      Diet  12  11  14
## 20      20      Diet  12  11  11
## 21      21      Diet  17  16  19
## 22      22      Diet  19  19  19
## 23      23      Diet  15  15  15
## 24      24      Diet  16  14  18
## 25      25    DietEx  15  11  19
## 26      26    DietEx  16  12  18
## 27      27    DietEx  13  12  17
## 28      28    DietEx  16  13  17
## 29      29    DietEx  13  13  16
## 30      30    DietEx  15  12  18
## 31      31    DietEx  15  13  18
## 32      32    DietEx  16  14  17
## 33      33    DietEx  16  16  19
## 34      34    DietEx  17  17  17
```

Figure 14: Self esteem data

```
se %>%
  pivot_longer(starts_with("se"), names_to = "time", values_to = "self_esteem") -> xx
```

Figure 15: Self esteem interaction plot part 1

```
xx %>%
  group_by(treatment, time) %>%
  summarize(mean_se = mean(self_esteem)) %>%
  ggplot(aes(x = time, y = mean_se, colour = treatment, group = treatment)) +
  geom_point() + geom_line()
```

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

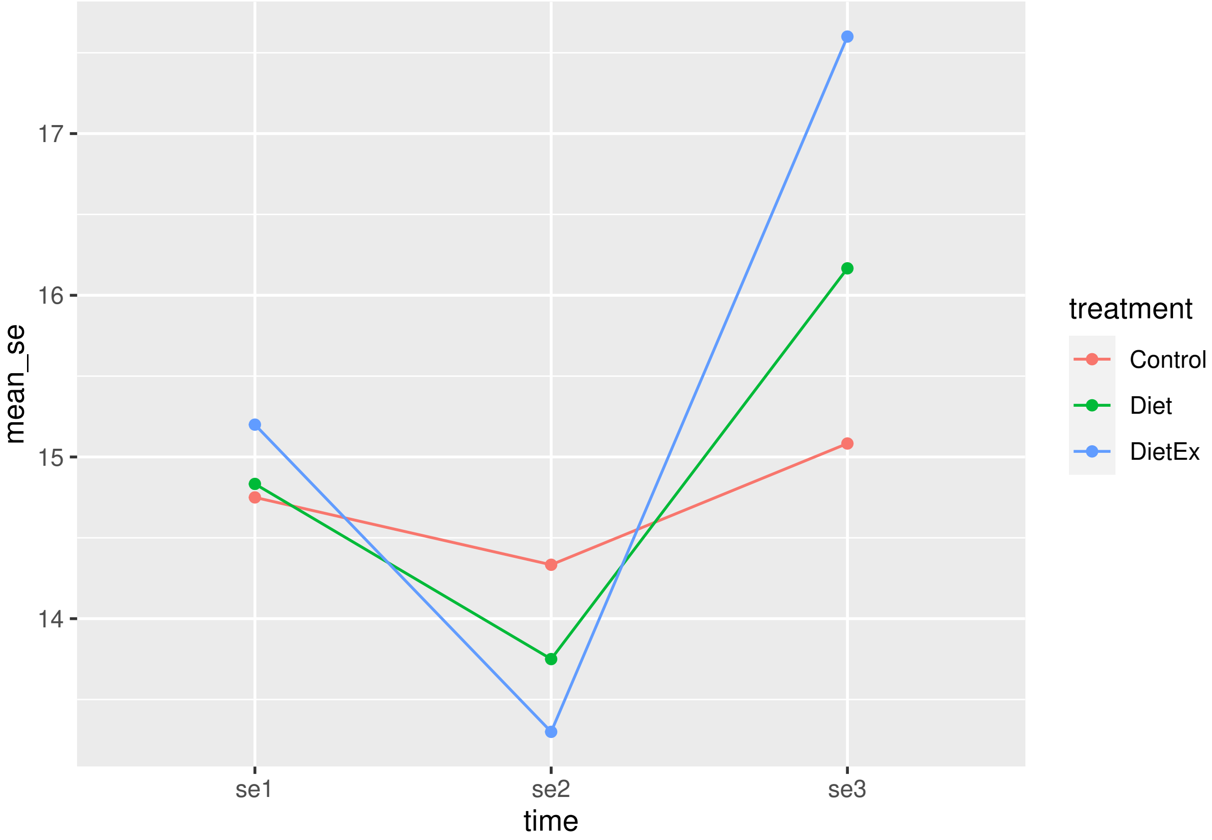


Figure 16: Self esteem interaction plot part 2

```
WeightLoss %>%
  select(starts_with("se")) %>%
  as.matrix() -> response
se.1 <- lm(response ~ group, data = WeightLoss)
times <- colnames(response)
times.df <- data.frame(times = factor(times))
se.2 <- Manova(se.1, idata = times.df, idesign = ~times)
summary(se.2)$univariate.tests
```

```
##               Sum Sq num Df Error SS den Df   F value    Pr(>F)
## (Intercept) 22890.0      1   278.94     31 2543.8949 < 2.2e-16 ***
## group           7.0      2   278.94     31    0.3902  0.680205
## times          96.7      2   134.58     62   22.2807 5.111e-08 ***
## group:times    34.7      4   134.58     62    3.9962  0.006003 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(se.2)$sphericity.tests # Mauchly's test
```

```
##             Test statistic  p-value
## times              0.75295 0.014173
## group:times        0.75295 0.014173
```

```
summary(se.2)$pval.adjustments
```

```
##               GG eps   Pr(>F[GG])    HF eps   Pr(>F[HF])
## times       0.801891 7.595864e-07 0.8389008 4.583464e-07
## group:times 0.801891 1.105415e-02 0.8389008 9.855788e-03
## attr(,"na.action")
## (Intercept)       group
##           1           2
## attr(,"class")
## [1] "omit"
```

Figure 17: Self esteem ANOVA: univariate tests, Mauchly's test, Greenhouse-Geisser and Huynh-Feldt adjustments

- `BR` brakes
- `FU` fuel system
- `EL` electrical
- EX exhaust
- `ST` steering
- `EM` engine, mechanical
- `RS` rattles and squeaks
- `RA` rear axle
- `RU` rust
- `SA` shock absorbers
- `TC` transmission or clutch
- `WA` wheel alignment
- `OT` other

Figure 18: Cars data column names

```
us.car.repair.1969
```

```
##                model BR FU EL EX ST EM RS RA RU SA TC WA OT
## 1  AMC Ambassador 8  +  -  -  -  -  -  -  +  -  -  -  -  -
## 2   Buick Special 6  -  -  -  -  -  -  +  -  +  -  -  -  +
## 3   Buick Special 8  -  -  -  -  -  -  +  -  -  +  -  +  +
## 4      Buick 8 Full  -  -  -  +  -  +  +  -  +  +  -  +  -
## 5     Buick Riviera  -  -  +  +  -  -  -  -  -  +  -  -  -
## 6          Chevy II  -  +  -  -  +  -  +  +  +  -  -  +  -
## 7        Chevelle 6  -  -  -  -  -  +  +  -  +  -  -  -  -
## 8        Chevelle 8  -  +  -  +  +  -  +  -  +  +  -  +  -
## 9    Chevrolet Full  -  +  +  +  +  -  +  +  +  +  +  +  -
## 10        Corvair 6  -  +  -  -  +  +  -  +  -  +  +  +  +
## 11         Corvette  -  -  -  +  -  -  +  +  -  -  +  -  -
## 12 Chrysler Newport  +  -  -  -  -  -  -  -  -  -  -  -  -
## 13       New Yorker  +  -  -  -  -  -  -  +  -  -  -  -  +
## 14  Dodge Full Size  +  -  -  -  -  -  +  -  -  +  -  -  -
## 15         Falcon 6  -  -  -  -  -  -  +  -  -  -  +  +  -
## 16       Fairlane 6  -  -  -  -  -  -  +  -  -  -  -  -  -
## 17       Fairlane 8  -  -  -  +  -  -  +  +  -  -  -  +  -
## 18   Ford Full Size  -  -  -  +  +  -  -  -  -  +  -  +  +
## 19      Thunderbird  -  -  +  -  +  +  -  -  -  -  -  +  +
## 20     Mercury Full  -  -  -  -  -  -  -  -  -  -  -  +  -
## 21        Olds Full  +  +  -  -  -  -  +  -  -  +  -  +  -
## 22    Plymouth Full  +  -  -  -  -  -  -  -  -  -  -  -  -
## 23  Pontiac Tempest  -  +  -  -  -  -  +  -  +  +  -  +  -
## 24     Pontiac Full  +  +  +  -  -  -  +  -  +  +  -  +  -
## 25  Rambler Rebel 6  -  -  +  -  -  -  -  +  -  -  +  -  +
## 26         Mercedes  -  -  -  -  -  -  -  -  -  +  -  -  -
## 27          MG 1100  -  -  +  +  -  -  -  -  -  -  -  -  -
## 28          Peugeot  -  -  -  -  -  -  -  -  -  -  +  -  -
## 29          Porsche  -  -  -  -  -  -  -  -  -  -  -  -  -
## 30          Renault  -  -  -  -  -  -  -  -  -  -  +  -  -
## 31            Volvo  -  -  -  +  -  -  -  +  -  -  -  -  -
## 32           VW bug  +  -  +  +  +  +  -  -  -  -  +  -  -
## 33           VW bus  -  -  +  -  -  +  -  -  +  -  +  -  -
```

Figure 19: Cars data

```
cars.1 <- hclust(dissim, method = "ward.D")
plot(cars.1)
rect.hclust(cars.1, 2)
```

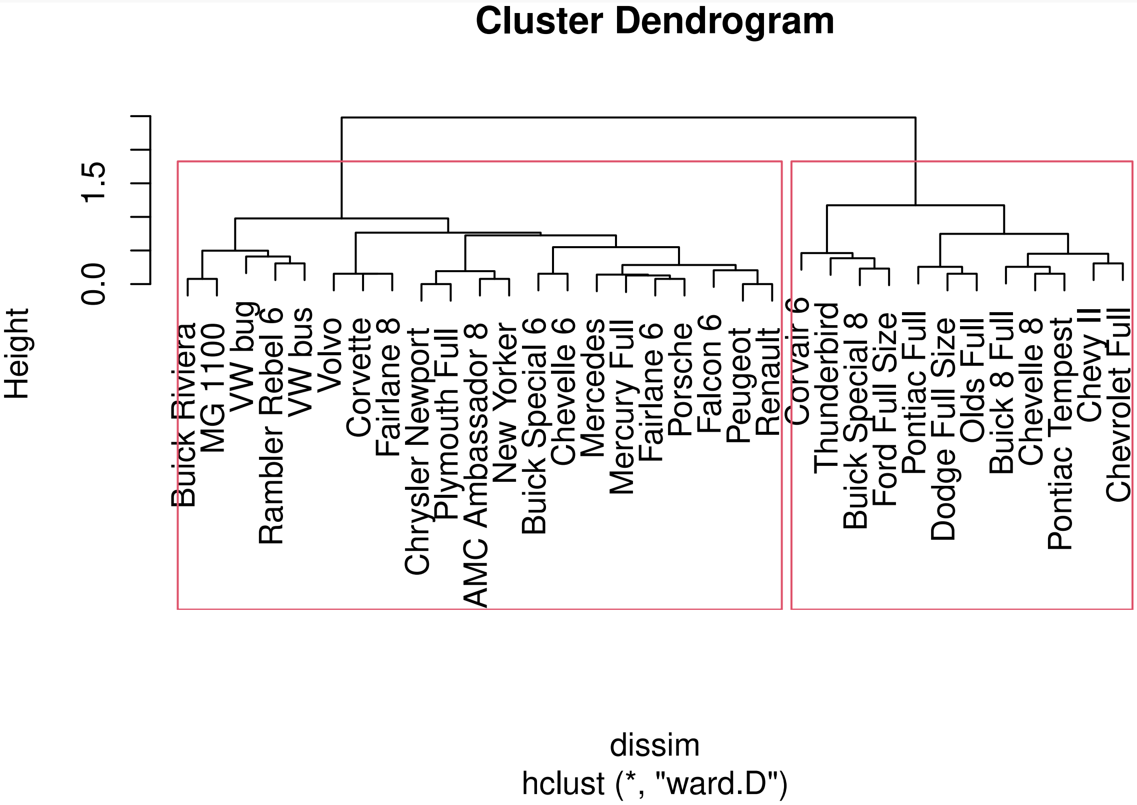


Figure 20: Cars cluster analysis

```
us.car.repair.1969 %>% pivot_longer(-model, names_to = "repair", values_to = "compare_avg") %>%
  group_by(model) %>%
  count(compare_avg) %>%
  filter(compare_avg == "+") %>%
  select(-compare_avg) -> problems
problems
```

```
## # A tibble: 32 x 2
## # Groups:   model [32]
##    model                n
##    <chr>            <int>
##  1 AMC Ambassador 8     2
##  2 Buick 8 Full         6
##  3 Buick Riviera        3
##  4 Buick Special 6      3
##  5 Buick Special 8      4
##  6 Chevelle 6           3
##  7 Chevelle 8           7
##  8 Chevrolet Full      10
##  9 Chevy II             6
## 10 Chrysler Newport     1
## # ... with 22 more rows
```

Figure 21: Cars: total number of above-average items by model

```
cutree(cars.1, 2) %>% enframe(name = "model", value = "cluster") -> clusters
clusters
```

```
## # A tibble: 33 x 2
##    model            cluster
##    <chr>              <int>
##  1 AMC Ambassador 8       1
##  2 Buick Special 6        1
##  3 Buick Special 8        2
##  4 Buick 8 Full           2
##  5 Buick Riviera          1
##  6 Chevy II               2
##  7 Chevelle 6             1
##  8 Chevelle 8             2
##  9 Chevrolet Full         2
## 10 Corvair 6              2
## # ... with 23 more rows
```

Figure 22: Cars: cluster membership

```
problems %>% left_join(clusters) %>%
  mutate(cluster = factor(cluster)) %>%
  ggplot(aes(x = cluster, y = n)) + geom_boxplot()
```

```
## Joining with `by = join_by(model)`
```

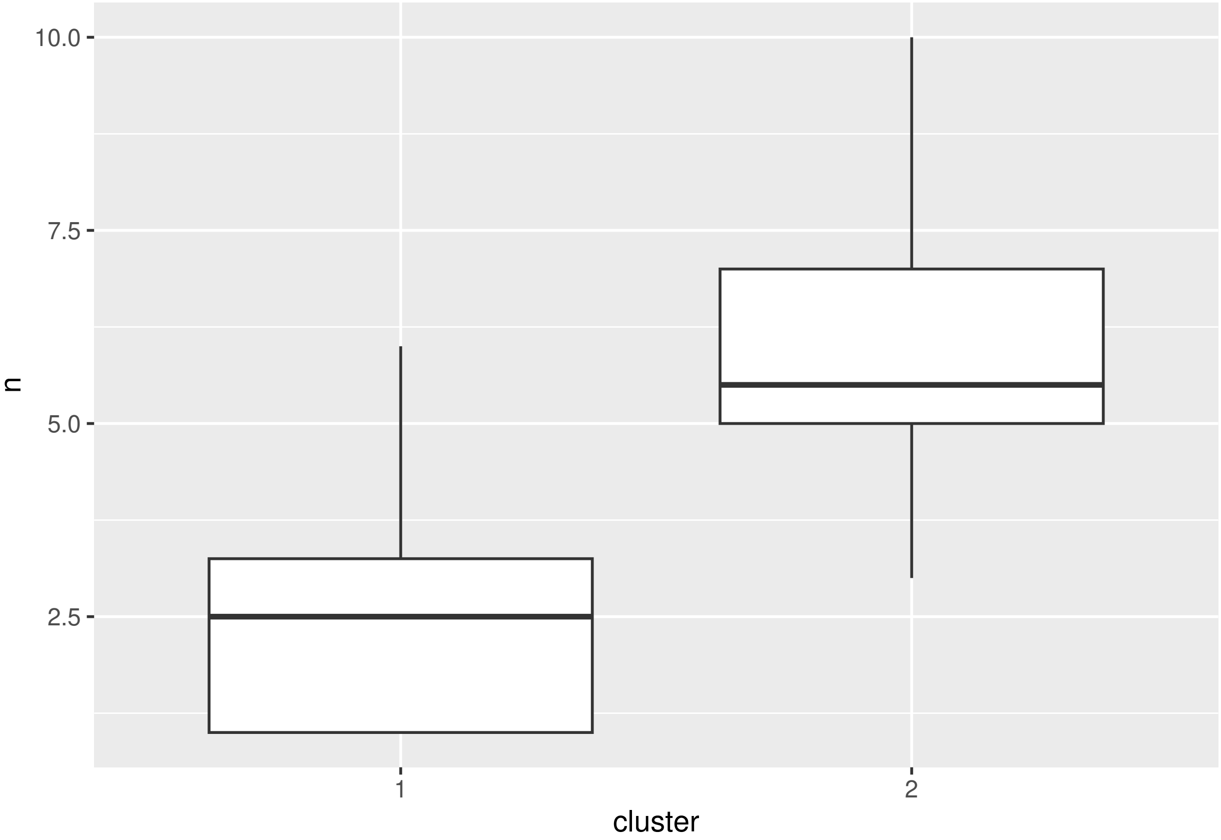


Figure 23: Cars: more computation and plot

- `crim` per capita crime rate.
- `zn` proportion of residential land zoned for lots over 25,000 sq.ft.
- `indus` proportion of tract occupied by non-retail business (ie. businesses that are not stores).
- `chas` Charles River dummy variable (= 1 if tract bounds river; 0 otherwise).
- `nox` nitrogen oxides concentration (parts per 10 million).
- `rm` average number of rooms per dwelling.
- `age` proportion of owner-occupied units built prior to 1940.
- `dis` weighted mean of distances to five Boston employment centres.
- `rad` index of accessibility to radial highways.
- `tax` full-value property-tax rate per $10,000.
- `ptratio` pupil-teacher ratio for schools in that tract.
- `black` A formula that is close to zero if the tract has proportion of Black people close to average (for Boston), and is high if that proportion is much higher *or lower* than average.
- `lstat` percent of the population that is of lower socio-economic status.
- `medv` median value of owner-occupied homes in $1000s.

Figure 24: Variables measured on Boston census tracts

```
Boston %>% slice(1:20)
```

```
## # A tibble: 20 x 14
##          crim    zn indus  chas   nox    rm   age   dis   rad   tax ptratio black lstat  medv
##         <dbl> <dbl> <dbl> <int> <dbl> <dbl> <dbl> <dbl> <int> <dbl>   <dbl> <dbl> <dbl> <dbl>
##  1 0.00632 18     2.31     0 0.538  6.58  65.2  4.09     1   296    15.3  397.  4.98  24
##  2 0.0273    0     7.07     0 0.469  6.42  78.9  4.97     2   242    17.8  397.  9.14  21.6
##  3 0.0273    0     7.07     0 0.469  7.18  61.1  4.97     2   242    17.8  393.  4.03  34.7
##  4 0.0324    0     2.18     0 0.458  7.00  45.8  6.06     3   222    18.7  395.  2.94  33.4
##  5 0.0690    0     2.18     0 0.458  7.15  54.2  6.06     3   222    18.7  397.  5.33  36.2
##  6 0.0298    0     2.18     0 0.458  6.43  58.7  6.06     3   222    18.7  394.  5.21  28.7
##  7 0.0883  12.5   7.87     0 0.524  6.01  66.6  5.56     5   311    15.2  396. 12.4   22.9
##  8 0.145   12.5   7.87     0 0.524  6.17  96.1  5.95     5   311    15.2  397. 19.2   27.1
##  9 0.211   12.5   7.87     0 0.524  5.63 100    6.08     5   311    15.2  387. 29.9   16.5
## 10 0.170   12.5   7.87     0 0.524  6.00  85.9  6.59     5   311    15.2  387. 17.1   18.9
## 11 0.225   12.5   7.87     0 0.524  6.38  94.3  6.35     5   311    15.2  393. 20.4   15
## 12 0.117   12.5   7.87     0 0.524  6.01  82.9  6.23     5   311    15.2  397. 13.3   18.9
## 13 0.0938  12.5   7.87     0 0.524  5.89  39    5.45     5   311    15.2  390. 15.7   21.7
## 14 0.630     0     8.14     0 0.538  5.95  61.8  4.71     4   307    21     397.  8.26  20.4
## 15 0.638     0     8.14     0 0.538  6.10  84.5  4.46     4   307    21     380. 10.3   18.2
## 16 0.627     0     8.14     0 0.538  5.83  56.5  4.50     4   307    21     396.  8.47  19.9
## 17 1.05      0     8.14     0 0.538  5.94  29.3  4.50     4   307    21     387.  6.58  23.1
## 18 0.784     0     8.14     0 0.538  5.99  81.7  4.26     4   307    21     387. 14.7   17.5
## 19 0.803     0     8.14     0 0.538  5.46  36.6  3.80     4   307    21     289. 11.7   20.2
## 20 0.726     0     8.14     0 0.538  5.73  69.5  3.80     4   307    21     391. 11.3   18.2
```

Figure 25: Boston census tract data (some)

```
boston.1 <- princomp(Boston, cor = TRUE)
ggscreeplot(boston.1)
```

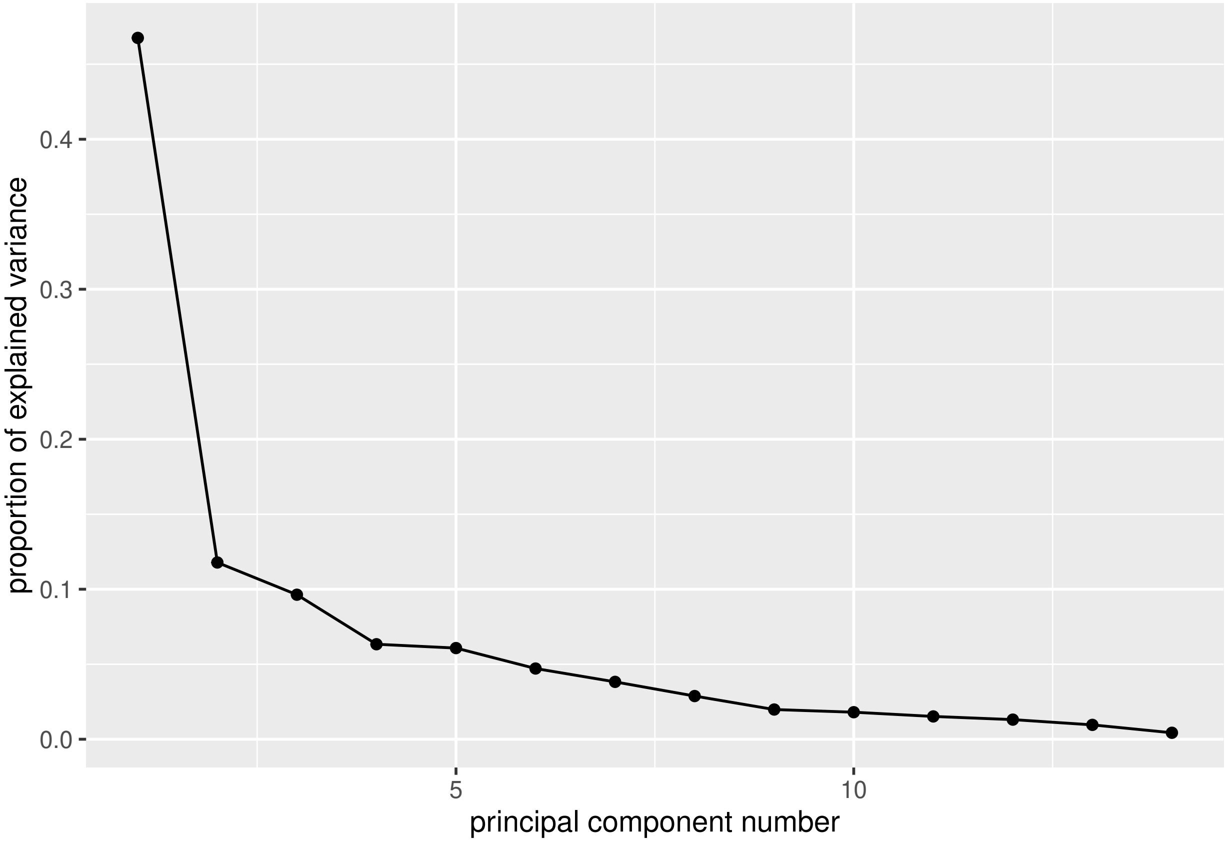


Figure 26: Boston scree plot

```
boston.1
```

```
## Call:
## princomp(x = Boston, cor = TRUE)
##
## Standard deviations:
##    Comp.1    Comp.2    Comp.3    Comp.4    Comp.5    Comp.6    Comp.7    Comp.8    Comp.9
## 2.5585132 1.2843410 1.1614241 0.9415625 0.9224421 0.8124105 0.7317177 0.6348831 0.5265582
##   Comp.10   Comp.11   Comp.12   Comp.13   Comp.14
## 0.5022524 0.4612919 0.4277704 0.3660733 0.2456149
##
##  14  variables and  506 observations.
```

Figure 27: Boston SD explained

```
boston.1$loadings
```

```
##
## Loadings:
##         Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8 Comp.9 Comp.10 Comp.11
## crim     0.242         0.395  0.100         0.225  0.777  0.157  0.254               
## zn      -0.245  0.148  0.395  0.343  0.114  0.336 -0.274 -0.380  0.383 -0.246   0.128
## indus    0.332 -0.127                             -0.340  0.172  0.627  0.255  -0.274
## chas           -0.411 -0.125  0.700 -0.535 -0.163                                    
## nox      0.325 -0.254                0.195  0.149 -0.198                0.212   0.437
## rm      -0.203 -0.434  0.353 -0.293        -0.131        -0.438         0.526  -0.224
## age      0.297 -0.260 -0.201         0.150         0.119 -0.588        -0.246   0.330
## dis     -0.298  0.359  0.157  0.185 -0.106        -0.104 -0.128 -0.176  0.299   0.115
## rad      0.303         0.419        -0.230  0.135 -0.137        -0.463 -0.116        
## tax      0.324         0.343        -0.163  0.188 -0.314        -0.179                
## ptratio  0.208  0.315        -0.342 -0.616 -0.279        -0.283  0.275 -0.160        
## black   -0.197        -0.361 -0.202 -0.367  0.786                       0.146        
## lstat    0.311  0.201 -0.161  0.243  0.178               -0.357 -0.172         -0.683
## medv    -0.267 -0.445  0.163 -0.180                       0.152        -0.576  -0.242
##         Comp.12 Comp.13 Comp.14
## crim                           
## zn      -0.221  -0.132         
## indus    0.348           -0.235
## chas                           
## nox     -0.449   0.525         
## rm      -0.126                 
## age      0.486                 
## dis      0.494   0.552         
## rad                      -0.635
## tax      0.170  -0.243   0.699 
## ptratio -0.232   0.188         
## black                          
## lstat   -0.182   0.249         
## medv             0.470   0.134 
##
##                Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8 Comp.9 Comp.10
## SS loadings     1.000  1.000  1.000  1.000  1.000  1.000  1.000  1.000  1.000   1.000
## Proportion Var  0.071  0.071  0.071  0.071  0.071  0.071  0.071  0.071  0.071   0.071
## Cumulative Var  0.071  0.143  0.214  0.286  0.357  0.429  0.500  0.571  0.643   0.714
##                Comp.11 Comp.12 Comp.13 Comp.14
## SS loadings      1.000   1.000   1.000   1.000
## Proportion Var   0.071   0.071   0.071   0.071
## Cumulative Var   0.786   0.857   0.929   1.000
```

Figure 28: Boston principal component loadings

```
boston.1$scores %>% as_tibble() %>%
  mutate(r = row_number()) %>%
  ggplot(aes(x = Comp.1, y = Comp.2, label = r)) + geom_text()
```

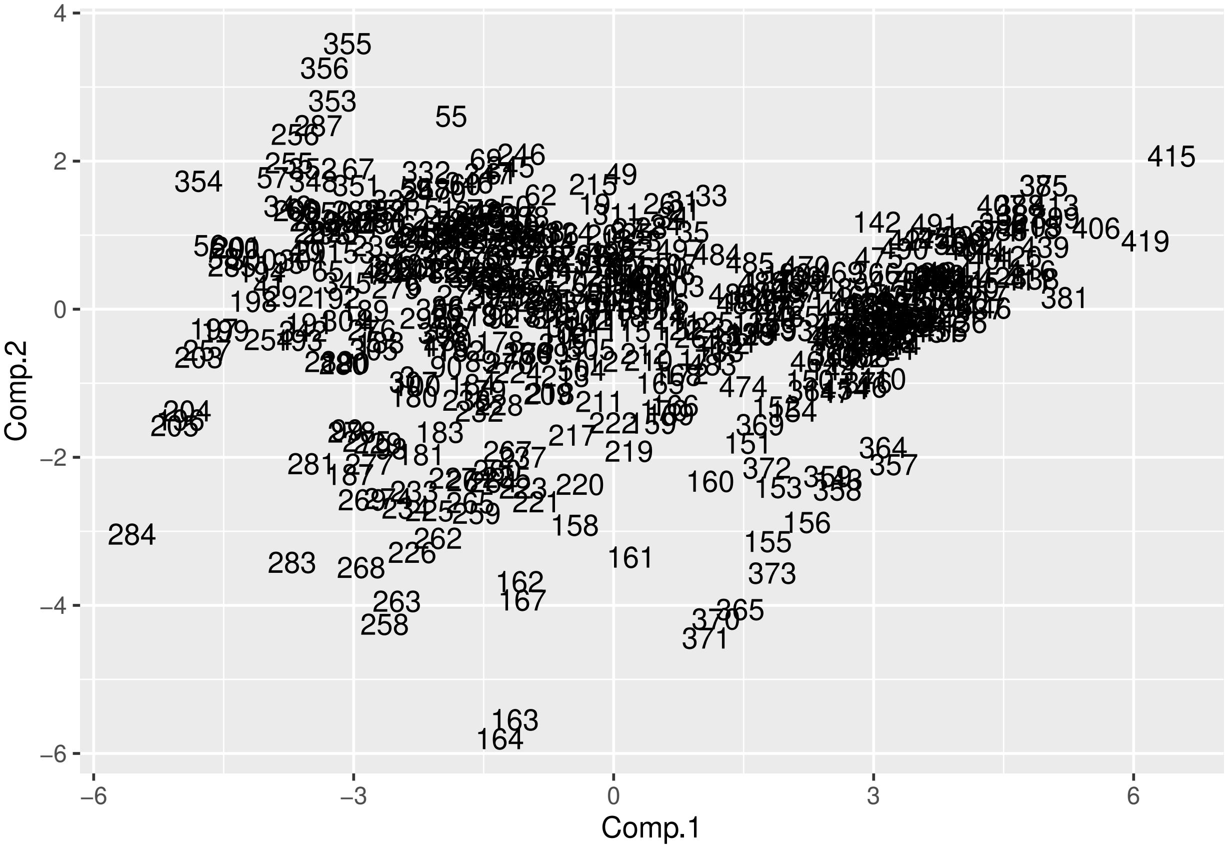


Figure 29: Boston principal component scores plot

```
summary(Boston)
```

```
##       crim                zn             indus            chas              nox
##  Min.   : 0.00632   Min.   :  0.00   Min.   : 0.46   Min.   :0.00000   Min.   :0.3850
##  1st Qu.: 0.08205   1st Qu.:  0.00   1st Qu.: 5.19   1st Qu.:0.00000   1st Qu.:0.4490
##  Median : 0.25651   Median :  0.00   Median : 9.69   Median :0.00000   Median :0.5380
##  Mean   : 3.61352   Mean   : 11.36   Mean   :11.14   Mean   :0.06917   Mean   :0.5547
##  3rd Qu.: 3.67708   3rd Qu.: 12.50   3rd Qu.:18.10   3rd Qu.:0.00000   3rd Qu.:0.6240
##  Max.   :88.97620   Max.   :100.00   Max.   :27.74   Max.   :1.00000   Max.   :0.8710
##        rm             age              dis              rad              tax
##  Min.   :3.561   Min.   :  2.90   Min.   : 1.130   Min.   : 1.000   Min.   :187.0
##  1st Qu.:5.886   1st Qu.: 45.02   1st Qu.: 2.100   1st Qu.: 4.000   1st Qu.:279.0
##  Median :6.208   Median : 77.50   Median : 3.207   Median : 5.000   Median :330.0
##  Mean   :6.285   Mean   : 68.57   Mean   : 3.795   Mean   : 9.549   Mean   :408.2
##  3rd Qu.:6.623   3rd Qu.: 94.08   3rd Qu.: 5.188   3rd Qu.:24.000   3rd Qu.:666.0
##  Max.   :8.780   Max.   :100.00   Max.   :12.127   Max.   :24.000   Max.   :711.0
##     ptratio          black            lstat            medv
##  Min.   :12.60   Min.   :  0.32   Min.   : 1.73   Min.   : 5.00
##  1st Qu.:17.40   1st Qu.:375.38   1st Qu.: 6.95   1st Qu.:17.02
##  Median :19.05   Median :391.44   Median :11.36   Median :21.20
##  Mean   :18.46   Mean   :356.67   Mean   :12.65   Mean   :22.53
##  3rd Qu.:20.20   3rd Qu.:396.23   3rd Qu.:16.95   3rd Qu.:25.00
##  Max.   :22.00   Max.   :396.90   Max.   :37.97   Max.   :50.00
```

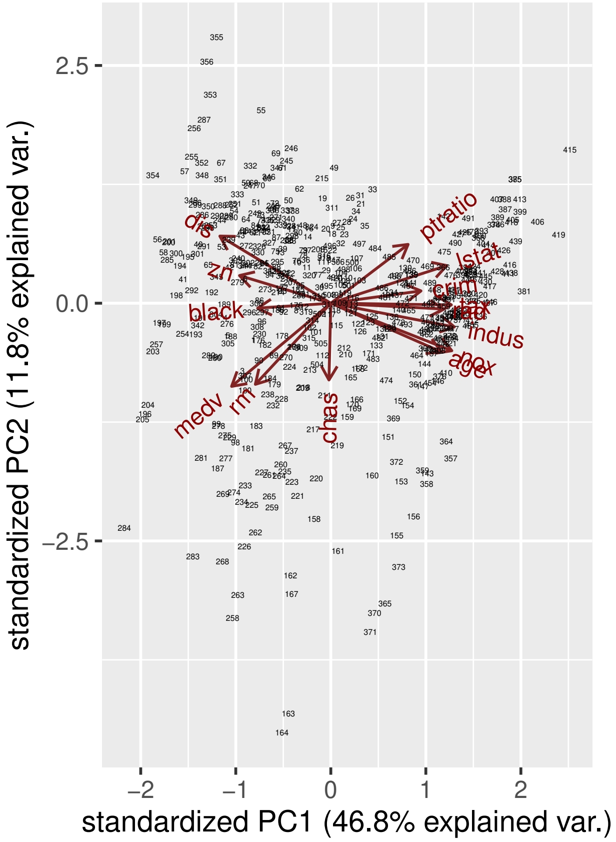
Figure 30: Boston summary

```
Boston %>% slice(163, 164)
```

```
## # A tibble: 2 x 14
##     crim    zn indus  chas   nox    rm   age   dis   rad   tax ptratio black lstat  medv
##    <dbl> <dbl> <dbl> <int> <dbl> <dbl> <dbl> <dbl> <int> <dbl>   <dbl> <dbl> <dbl> <dbl>
## 1  1.83     0  19.6     1 0.605  7.80  98.2  2.04     5   403    14.7  390.  1.92    50
## 2  1.52     0  19.6     1 0.605  8.38  93.9  2.16     5   403    14.7  388.  3.32    50
```

Figure 31: Boston data values for observations 163 and 164

```
ggbiplot(boston.1, labels = 1:506, labels.size = 1, alpha = 0.5)
```



I tried to make this come out bigger, but it wouldn't.

Figure 32: Boston data biplot