# STA 104 Midterm 2 Project Report

Radhika Kulkarni, Sameerah Helal, Shannon Meyer

05/01/2021

Class: STA 135 - Multivariate Data Analysis
Instructor: Maxime G. Pouokam
Email: rrkulkarni@ucdavis.edu
shelal@ucdavis.edu
shmmeyer@ucdavis.edu
Student IDs: 917212169, 917496794, 915419139

## I. Introduction.

The following paper addresses the question of _____. A claim we would like to test is _____. We are interested in the result of testing this claim because _____. We will use the statistical technique of _____ to determine _____.

## II. Summary of Data

This paper utilizes categorical and numeric data of newborn babies with features related to the newborns' own and parents' physical characteristics.

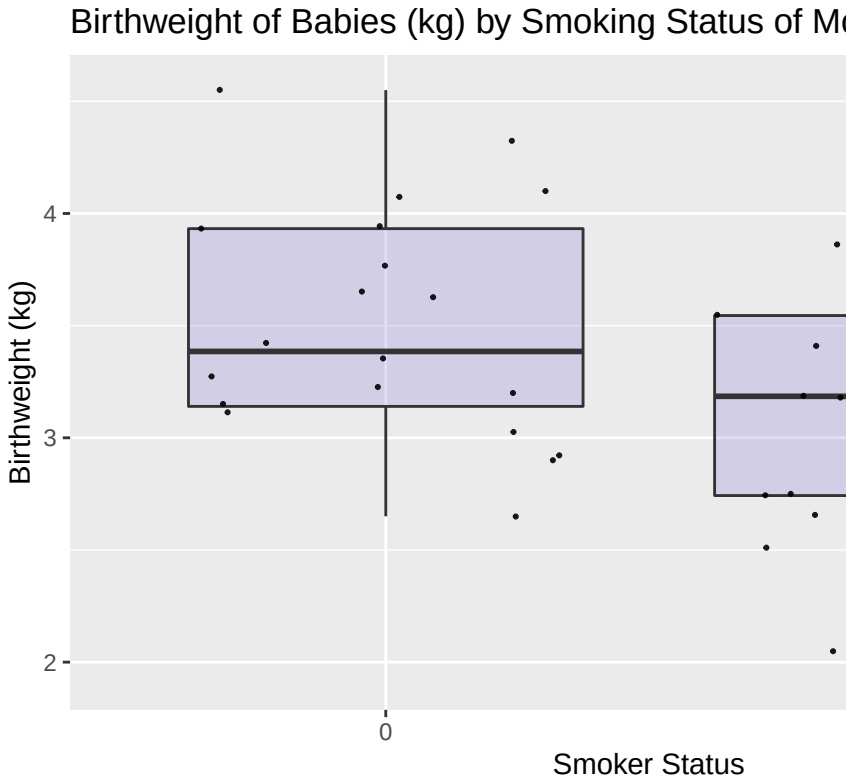



Figure 1: Birthweight of babies with nonsmoking mothers appears to be s
smoking mothers. Median weight regardless of smoking status is betwee

We begin by looking into our observation of interest:

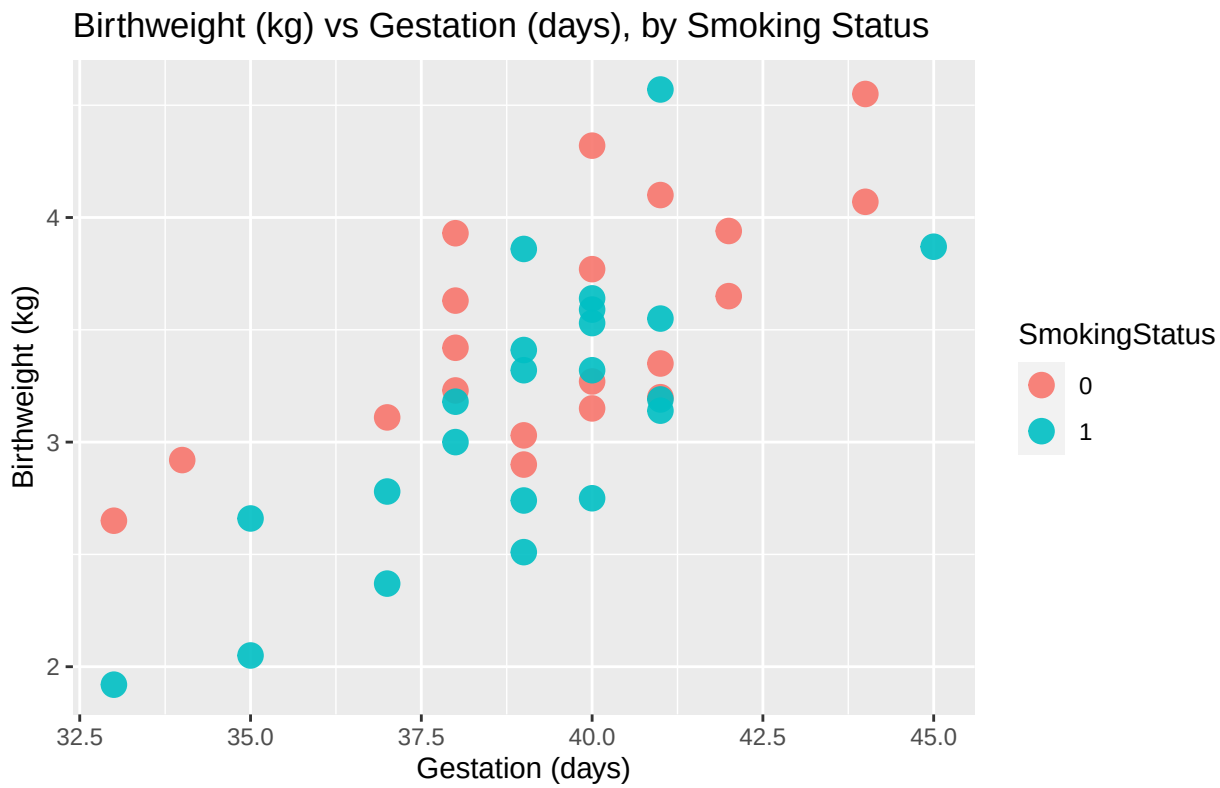
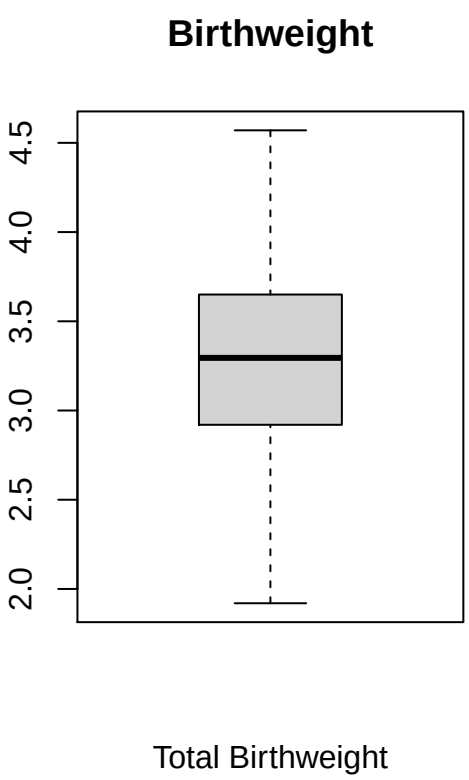
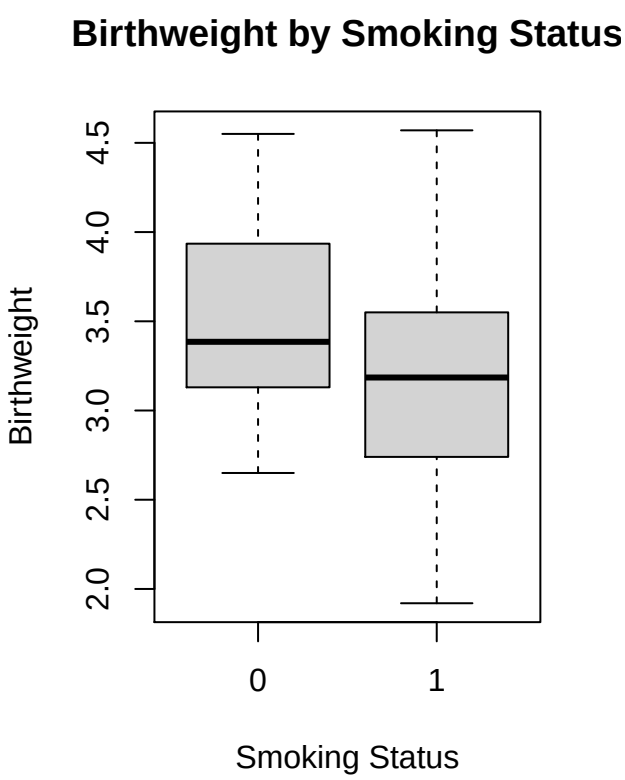
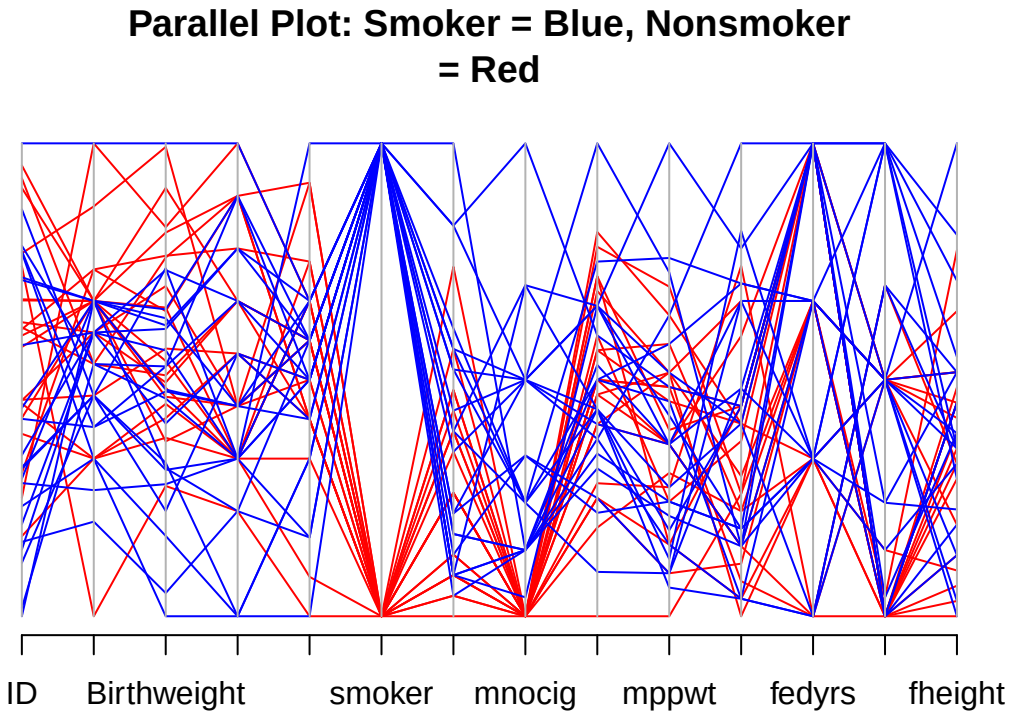



Figure 2: Smoking mothers appear to have lower days of gestation compared to nonsmoking mothers, but are more varied in range of days of gestation.




It also might make sense to see how of primary variable of interest, smoking status, is related to the other variables in the dataset. The parallel plot doesn't give an obvious trend, but we can see that in general, we would need more precise analysis to draw mroe conclusions about interactions between the variables, especially because our dimensionality here is far too high.



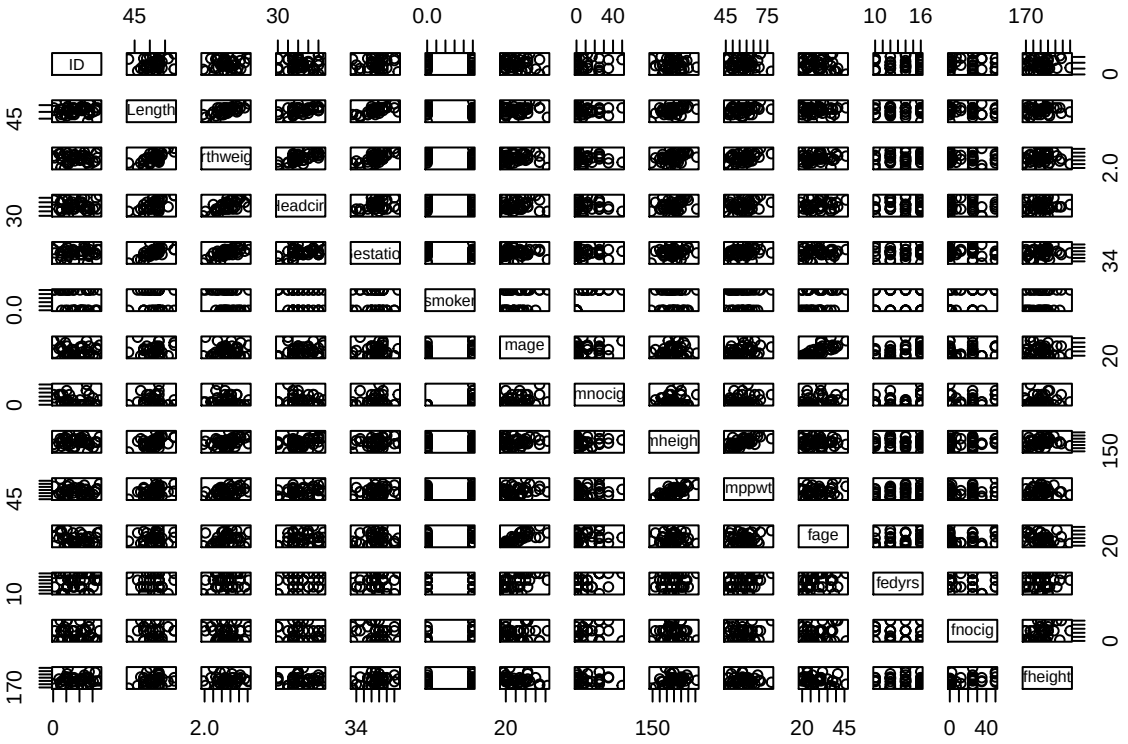

We look at the mean vector of the dataset.

```
##          ID      Length Birthweight    Headcirc   Gestation      smoker
## 894.0714286  51.3333333   3.3128571  34.5952381  39.1904762   0.5238095
##        mage      mnocig     mheight       mppwt        fage      fedyrs
##  25.5476190   9.4285714 164.4523810  57.5000000  28.9047619  13.6666667
##      fnocig     fheight
##  17.1904762 180.5000000
```

We note that not all the variables are on the same scale, so we use correlation information for these analyses. If we were to use the covariance matrix, the benefit would be the ability to use the original data values in the analysis, but we would end up with results that were very difficult to interpret. Instead, the correlation matrix standardizes our variables, so we can directly see the associated changes in variation between the different variables. Now, let's look at the correlation matrix for the data.

```
##                     ID       Length   Birthweight    Headcirc   Gestation
## ID           1.00000000  0.192635194  0.1112107085 -0.04489917  0.10271426
## Length       0.19263519  1.000000000  0.7268334848  0.56317161  0.70511059
## Birthweight  0.11121071  0.726833485  1.0000000000  0.68461562  0.70830289
## Headcirc    -0.04489917  0.563171606  0.6846156185  1.00000000  0.40463477
## Gestation    0.10271426  0.705110591  0.7083028937  0.40463477  1.00000000
## smoker      -0.27618220 -0.153406153 -0.3142339464 -0.18287185 -0.09474608
## mage        -0.18642559  0.075268357  0.0001731023  0.14584152  0.01077846
## mnocig      -0.19462487 -0.039842761 -0.1523351845 -0.13298772  0.04319486
## mheight     -0.01089325  0.484992403  0.3630551926  0.33704682  0.21050318
## mppwt       -0.02651624  0.398197402  0.4008856280  0.30285407  0.25508237
```

```
## fage        -0.23195803  0.137184365  0.1757099933  0.30115080  0.14217533
## fedyrs       0.11992711  0.079484567  0.0710452258  0.12389250  0.13098664
## fnocig      -0.11288857  0.008800476 -0.0931357623 -0.04683678 -0.11383061
## fheight      0.08901791  0.208358435  0.0310224972  0.04150923  0.20759684
##                    smoker          mage      mnocig       mheight       mppwt
## ID          -0.2761821977 -0.1864255867 -0.19462487 -0.0108932452 -0.02651624
## Length      -0.1534061527  0.0752683574 -0.03984276  0.4849924026  0.39819740
## Birthweight -0.3142339464  0.0001731023 -0.15233518  0.3630551926  0.40088563
## Headcirc    -0.1828718541  0.1458415206 -0.13298772  0.3370468163  0.30285407
## Gestation   -0.0947460785  0.0107784551  0.04319486  0.2105031792  0.25508237
## smoker       1.0000000000  0.2124787863  0.72721809  0.0003532676  0.00000000
## mage         0.2124787863  1.0000000000  0.34029438  0.0599563823  0.27416768
## mnocig       0.7272180924  0.3402943777  1.00000000  0.1264388844  0.14894461
## mheight      0.0003532676  0.0599563823  0.12643888  1.0000000000  0.68062174
## mppwt        0.0000000000  0.2741676755  0.14894461  0.6806217412  1.00000000
## fage         0.1975014481  0.8065844174  0.24842538 -0.0798698948  0.25570584
## fedyrs      -0.0148905839  0.4416826598  0.19852620  0.0352970193  0.18037409
## fnocig       0.4176329588  0.0909266363  0.25730739  0.0483980589  0.05716254
## fheight      0.1106327308 -0.1995468633  0.02067224  0.2743379320  0.09298347
##                     fage      fedyrs       fnocig     fheight
## ID          -0.23195803  0.11992711 -0.112888569  0.08901791
## Length       0.13718437  0.07948457  0.008800476  0.20835843
## Birthweight  0.17570999  0.07104523 -0.093135762  0.03102250
## Headcirc     0.30115080  0.12389250 -0.046836781  0.04150923
## Gestation    0.14217533  0.13098664 -0.113830614  0.20759684
## smoker       0.19750145 -0.01489058  0.417632959  0.11063273
## mage         0.80658442  0.44168266  0.090926636 -0.19954686
## mnocig       0.24842538  0.19852620  0.257307386  0.02067224
## mheight     -0.07986989  0.03529702  0.048398059  0.27433793
## mppwt        0.25570584  0.18037409  0.057162540  0.09298347
## fage         1.00000000  0.30047147  0.135862017 -0.26937685
## fedyrs       0.30047147  1.00000000 -0.263103019  0.01779765
## fnocig       0.13586202 -0.26310302  1.000000000  0.32936416
## fheight     -0.26937685  0.01779765  0.329364159  1.00000000
```

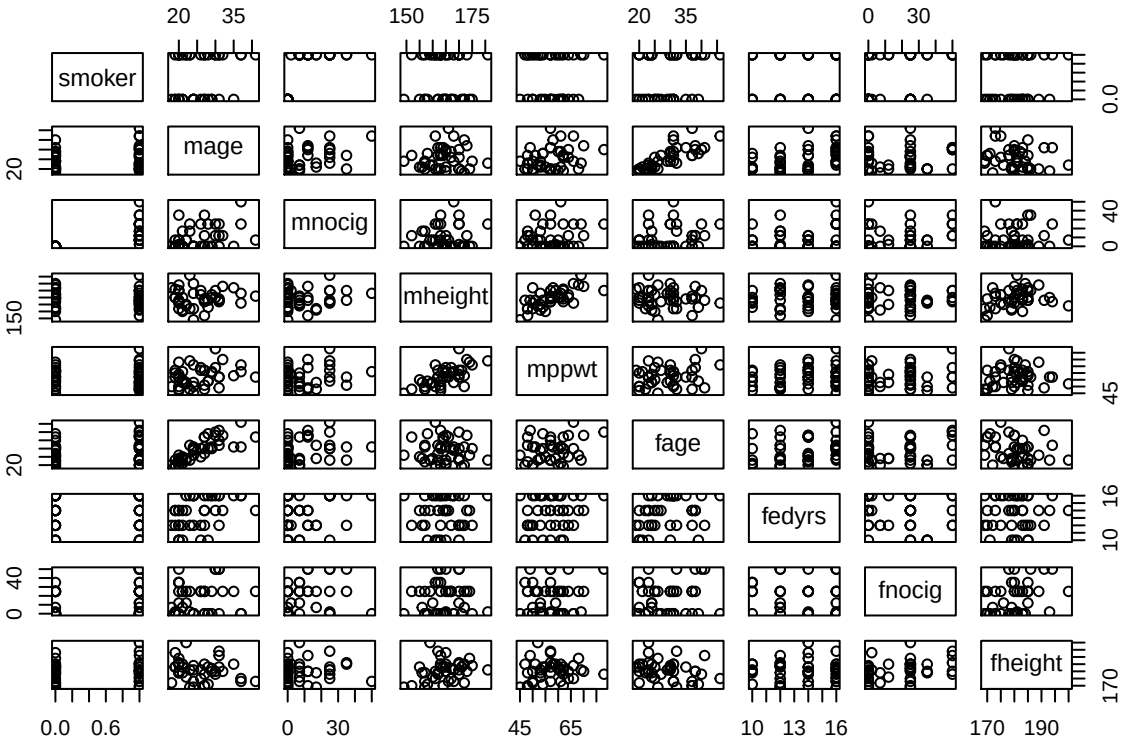



It is very difficult to understand what the important trends are using only a correlation matrix, so we use some manual analysis to get an idea of the data.

We might think that some types of variables will be more correlated with each other than other. For instance, we can divide the birthweight predictor variables into parental predictors and physical birth characteristics.

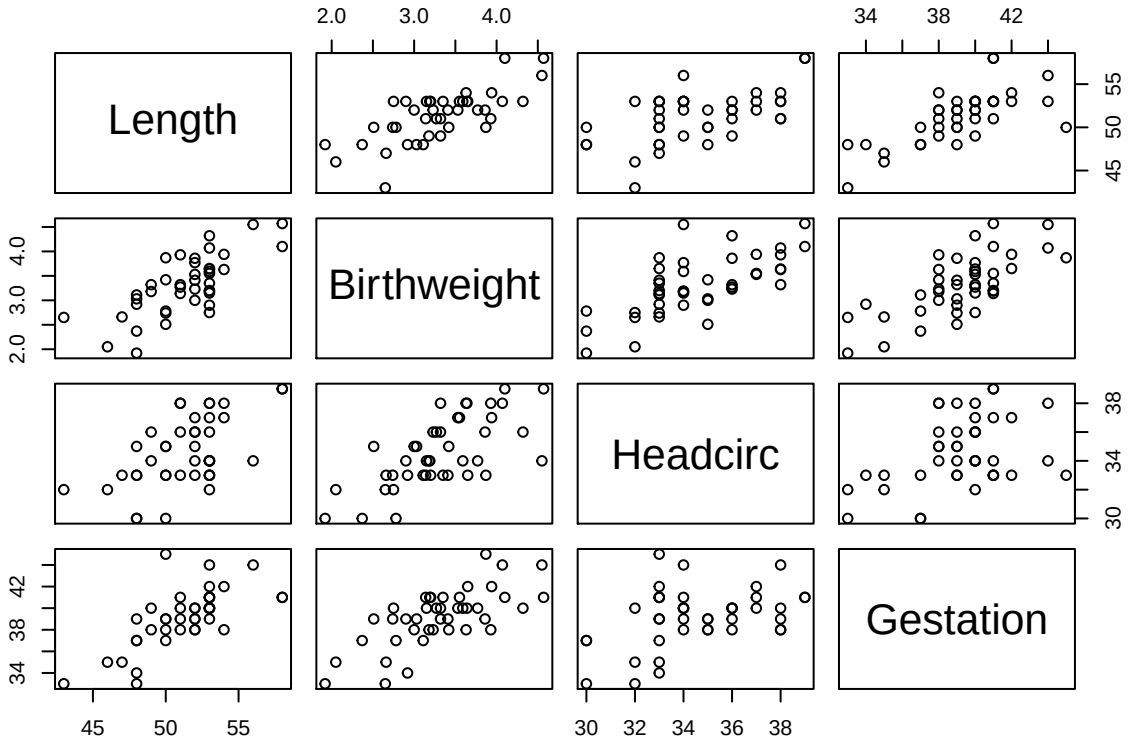
We can see the correlation matrix and plots for the parental characteristics. The strongest associations are between smoking status and the number of cigarettes smoked by the mother (a trivial observation) and between the age of the mother and the age of the father.

```
##          smoker    mage mnocig mheight  mppwt    fage  fedyrs  fnocig fheight
## smoker   1.0000  0.2125 0.7272  0.0004 0.0000  0.1975 -0.0149  0.4176  0.1106
## mage     0.2125  1.0000 0.3403  0.0600 0.2742  0.8066  0.4417  0.0909 -0.1995
## mnocig   0.7272  0.3403 1.0000  0.1264 0.1489  0.2484  0.1985  0.2573  0.0207
## mheight  0.0004  0.0600 0.1264  1.0000 0.6806 -0.0799  0.0353  0.0484  0.2743
## mppwt    0.0000  0.2742 0.1489  0.6806 1.0000  0.2557  0.1804  0.0572  0.0930
## fage     0.1975  0.8066 0.2484 -0.0799 0.2557  1.0000  0.3005  0.1359 -0.2694
## fedyrs  -0.0149  0.4417 0.1985  0.0353 0.1804  0.3005  1.0000 -0.2631  0.0178
## fnocig   0.4176  0.0909 0.2573  0.0484 0.0572  0.1359 -0.2631  1.0000  0.3294
## fheight  0.1106 -0.1995 0.0207  0.2743 0.0930 -0.2694  0.0178  0.3294  1.0000
```

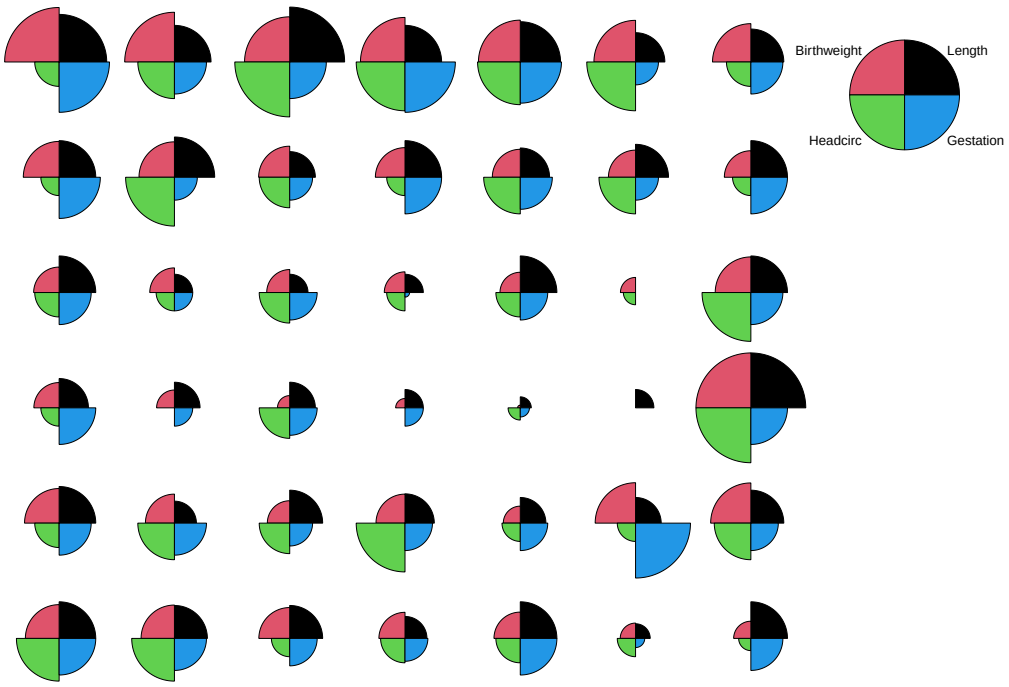



Similarly, we can look at these same charts for physical characteristics of the infant. We see high correlation between all the size-related infant variables.

```
##             Length Birthweight Headcirc Gestation
## Length      1.0000      0.7268   0.5632    0.7051
## Birthweight 0.7268      1.0000   0.6846    0.7083
## Headcirc    0.5632      0.6846   1.0000    0.4046
## Gestation   0.7051      0.7083   0.4046    1.0000
```




We can also show this with a star plot. Notice again that length and birthweight tend to move together. There are a few outliers with very low head circumference and gestation time; these may be prematurely born babies.




This gives us an intuitive sense of the correlation structures. But, in order to computationally reduce the number of relevant variables, we turn to PCA.

## III. Analysis.

We use the technique of PCA to analyze the data and perform feature reduction. This will make it easier to see the patterns in the data, which currently has 14 columns and is difficult to parse. We hope to reduce the number of features and find which combinations of features have the greatest influence on the data.

We separate the dataset into those samples for which 'smoker' was marked as 1, or True, and a second dataset where 'smoker' was marked False. We then conduct PCA on each of the datasets.

```
## Importance of components:
##                            Comp.1     Comp.2    Comp.3     Comp.4     Comp.5
## Standard deviation     18.2010014 13.1693540 9.3739499 8.52654435 5.97088377
## Proportion of Variance  0.4523284  0.2368058 0.1199799 0.09926803 0.04867888
## Cumulative Proportion   0.4523284  0.6891342 0.8091141 0.90838213 0.95706101
##                            Comp.6     Comp.7      Comp.8      Comp.9
## Standard deviation     3.82939145 3.05975421 1.645299140 1.411144588
## Proportion of Variance 0.02002271 0.01278311 0.003696179 0.002718982
## Cumulative Proportion  0.97708371 0.98986682 0.993563000 0.996281982
##                           Comp.10     Comp.11      Comp.12
## Standard deviation     1.311482522 0.985957764 1.757964e-01
## Proportion of Variance 0.002348488 0.001327333 4.219716e-05
## Cumulative Proportion  0.998630470 0.999957803 1.000000e+00
##
## Loadings:
```

```
##             Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8 Comp.9
## Length                                    0.107  0.310  0.330  0.160  0.588
## Birthweight
## Headcirc                   0.124                0.413  0.274 -0.364  0.323
## Gestation                                 0.190  0.308  0.221  0.509
## mage                0.308  0.352 -0.158  0.286 -0.643                0.448
## mnocig       0.107  0.757 -0.573 -0.271
## mheight             0.203         0.583 -0.140 -0.355  0.632        -0.224
## mppwt               0.371  0.258  0.641 -0.125  0.222 -0.532
## fage                0.334  0.560 -0.219  0.312  0.183  0.145 -0.113 -0.531
## fedyrs                                    0.120        -0.215  0.720
## fnocig      -0.975                       -0.148
## fheight     -0.149 -0.122 -0.372  0.303  0.834        -0.100 -0.153
##             Comp.10 Comp.11 Comp.12
## Length       0.610   0.139
## Birthweight                 -0.985
## Headcirc    -0.663   0.232
## Gestation   -0.235  -0.685   0.153
## mage        -0.145  -0.190
## mnocig
## mheight
## mppwt               -0.160
## fage         0.197   0.182
## fedyrs      -0.256   0.588
## fnocig
## fheight

## Importance of components:
##                             Comp.1    Comp.2    Comp.3     Comp.4     Comp.5
## Standard deviation     13.4236349 8.6098894 6.7799060 4.60464940 3.87261789
## Proportion of Variance  0.5106122 0.2100613 0.1302562 0.06008196 0.04249719
## Cumulative Proportion   0.5106122 0.7206735 0.8509297 0.91101164 0.95350883
##                             Comp.6      Comp.7      Comp.8      Comp.9
## Standard deviation     3.04425920 1.741297942 1.501975557 1.075009822
## Proportion of Variance 0.02626118 0.008592056 0.006392587 0.003274732
## Cumulative Proportion  0.97977001 0.988362066 0.994754654 0.998029386
##                            Comp.10      Comp.11 Comp.12
## Standard deviation     0.803655947 0.2226267921       0
## Proportion of Variance 0.001830169 0.0001404448       0
## Cumulative Proportion  0.999859555 1.0000000000       1
##
## Loadings:
##             Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8 Comp.9
## Length       0.135  0.153                0.388  0.438  0.347  0.360  0.408
## Birthweight                                                0.106        -0.155
## Headcirc            0.155                0.284         0.202 -0.188 -0.754
## Gestation    0.111  0.119                       0.593  0.342 -0.160 -0.264
## mage                      -0.588  0.217  0.193 -0.443  0.328  0.415
## mnocig
## mheight      0.201  0.628        -0.448  0.411 -0.167 -0.310         0.123
## mppwt        0.216  0.464 -0.232 -0.192 -0.723         0.297
## fage        -0.142        -0.701  0.186         0.397 -0.468 -0.213
## fedyrs                           0.224  0.169 -0.229  0.371 -0.755  0.386
## fnocig       0.896 -0.390 -0.157
```

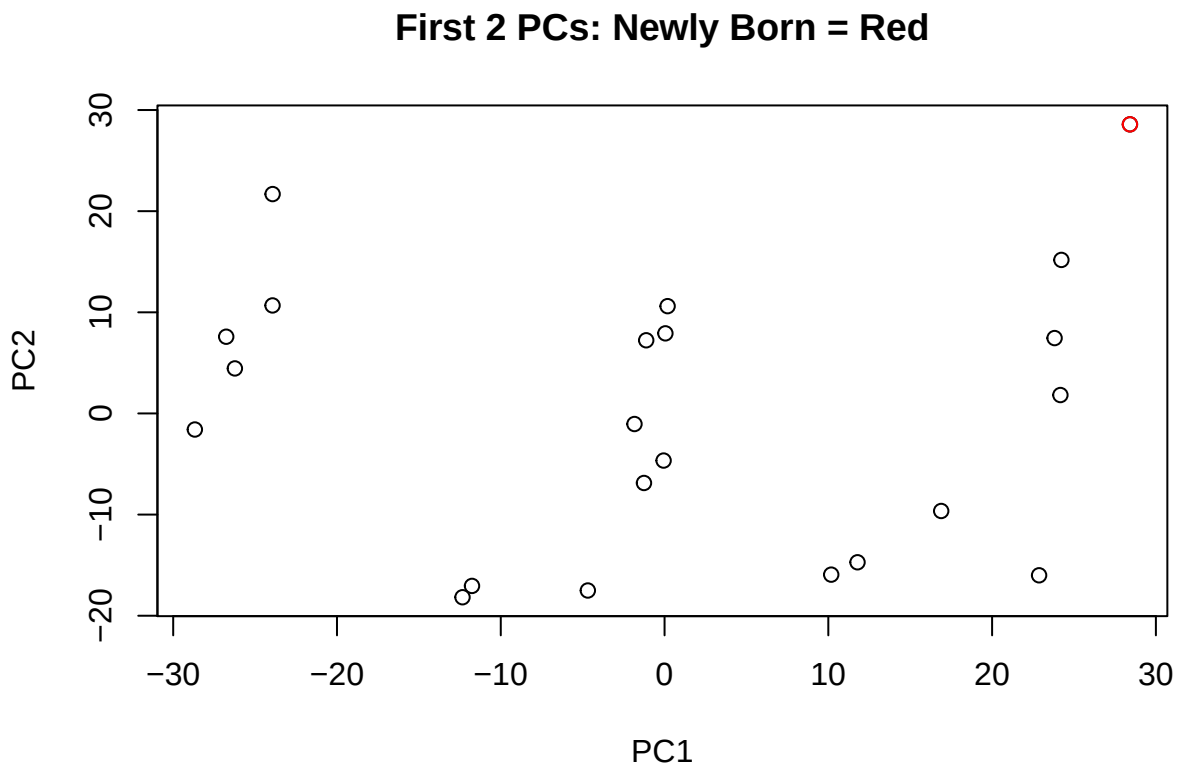
```
## fheight      0.234  0.413  0.281  0.788               -0.234
##             Comp.10 Comp.11 Comp.12
## Length       0.441
## Birthweight         -0.977
## Headcirc     0.460   0.172
## Gestation   -0.628
## mage        -0.294
## mnocig                      1.000
## mheight     -0.217
## mppwt        0.185
## fage         0.163
## fedyrs
## fnocig
## fheight
```

The details of the first four principal components for the smoker data are.

We see that we require the first four principal components to capture at least 90% of the variance in the original data. From the loadings, the most important features appear to be whether the parents smoked cigarettes, and the height of the mother. Notice that the features of length, birthweight, gestation, and fedyrs are dropped in all four. Similarly, the details of the first four principal components for the nonsmoker data are.

Again, we require the first four principal components in order to capture 90% of the variance in the data. The loadings tell us that, for the nonsmoker dataset, the mother's smoking status is most important in the first principal component, followed by some physical characteristics. The birthweight and father's smoking status are dropped.

We now consider an additional newborn and apply PCA to the single sample. We then plot the first two principal components of the entire dataset against each other.




Since the newly born has smoker set to 1, we use the first PCA result with similar data. Since this data point, mapped using PCA from the main data and marked in red, is distant from the others, it appears that this newly born is unusual compared to the other babies.

## IV. Interpretation.

## V. Conclusion.

The goal of this paper was to determine . . .

## Appendix Code

```
knitr::opts_chunk$set(echo = FALSE, message = FALSE, warning = FALSE)
# load
library(readr)
birth <- read_csv("Birthweight_reduced_kg_R.csv")

# drop last two columns
birth <- birth[, -c(15,16)]
attach(birth)

# view df
View(birth)
# Boxplot
library(ggplot2)
ggplot(birth, aes(x=as.factor(smoker), y=Birthweight)) +
    geom_boxplot(fill="slateblue", alpha=0.2) +
    xlab("Smoker Status") + ylab("Birthweight (kg)") +
    geom_jitter(color="black", size=0.4,alpha=0.9) + ggtitle("Birthweight of Babies (kg) by Smoking Sta
```

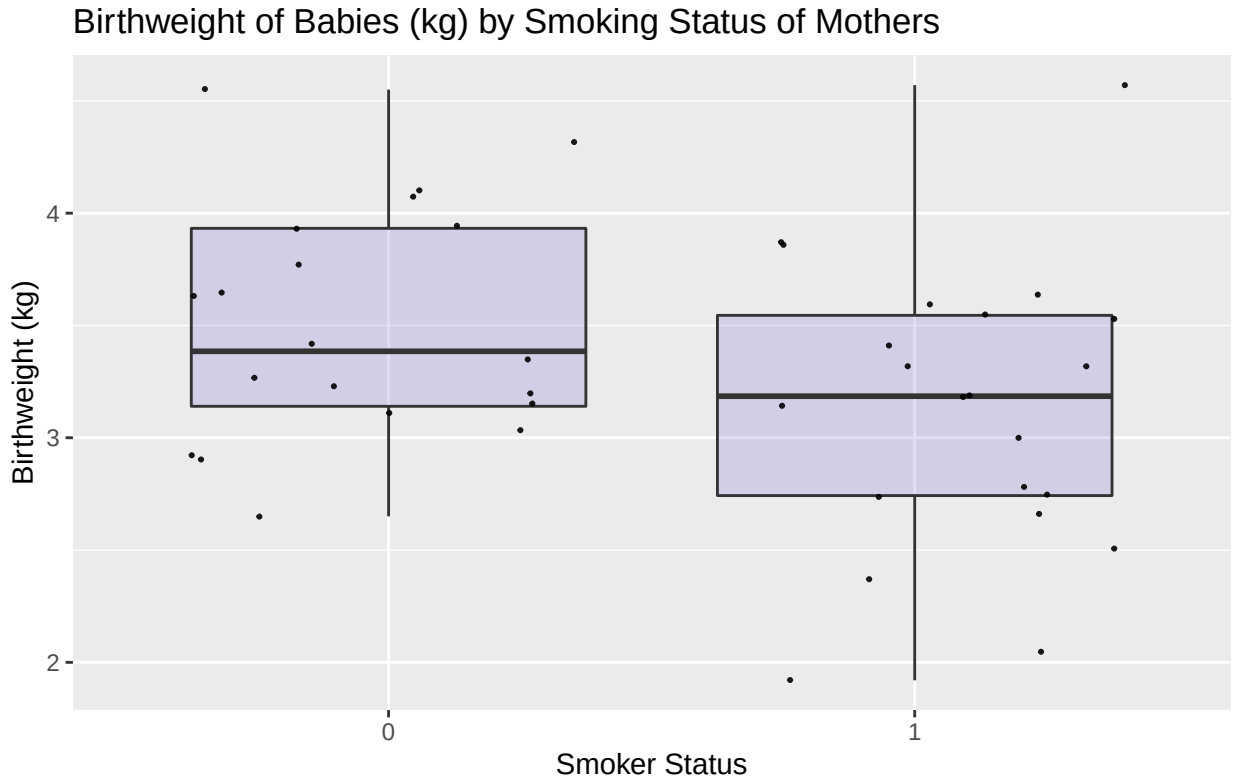



Figure 1: Birthweight of babies with nonsmoking mothers appears to be significantly higher than babies with smoking mothers. Median weight regardless of smoking status is between 3.0 kg and 3.5 kg.

```
#histograms
library(hrbrthemes)
SmokingStatus = as.factor(smoker)
# scatterplot of birthweight vs gestation with color depending on smoker
ggplot(birth, aes(x=Gestation, y=Birthweight, color=SmokingStatus
#                  ,lty=SmokingStatus
```

```
                ))+ geom_point(size=4, alpha = 0.9) +
#  geom_smooth(method=lm , color= "red", se=FALSE)+
  xlab("Gestation (days)")+ylab("Birthweight (kg)")+ggtitle("Birthweight (kg) vs Gestation (days), by Sm
```

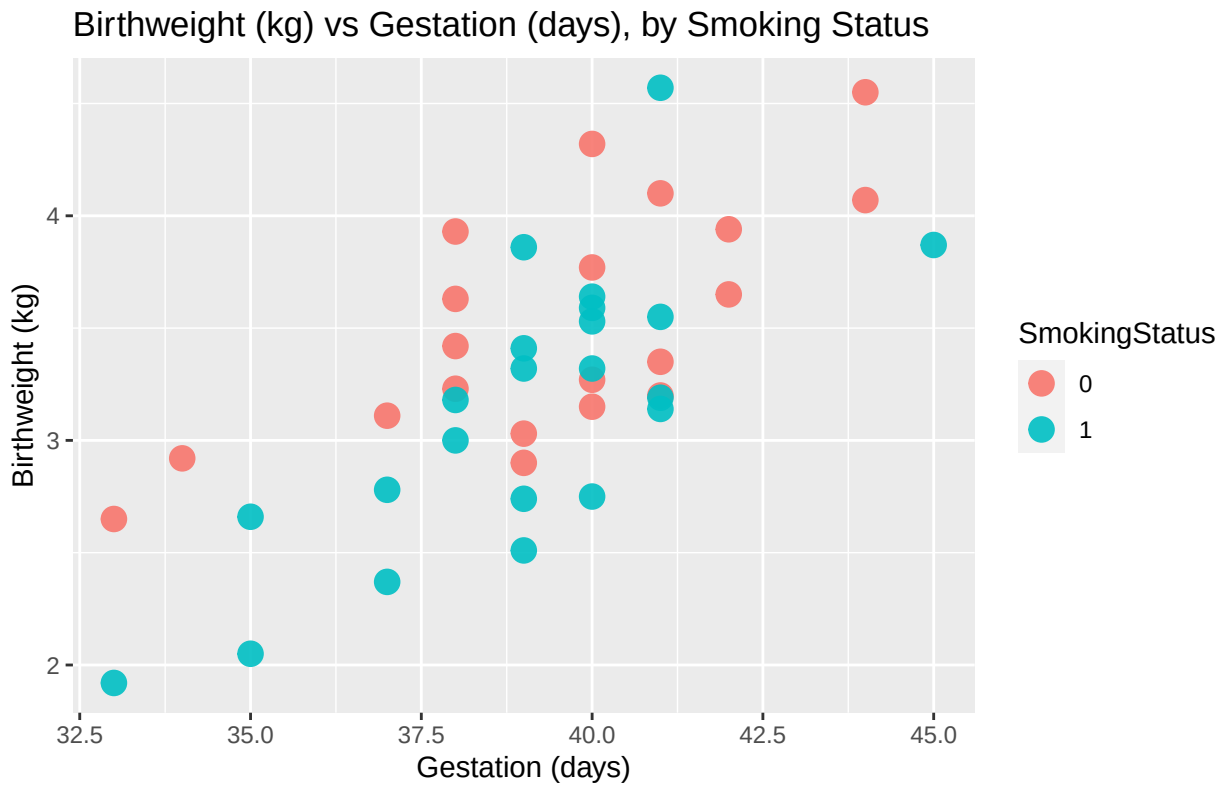
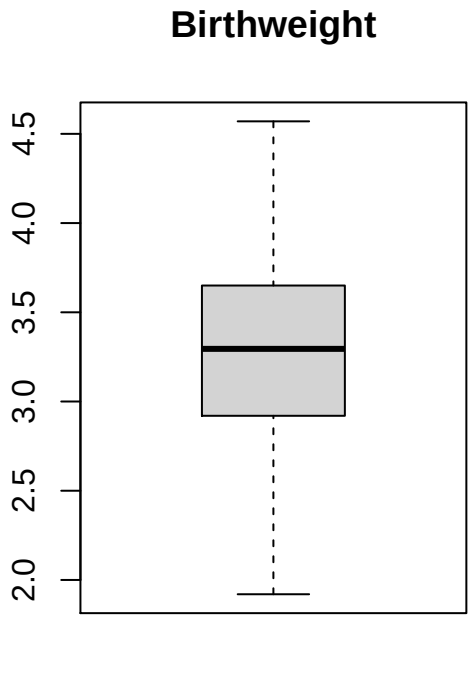
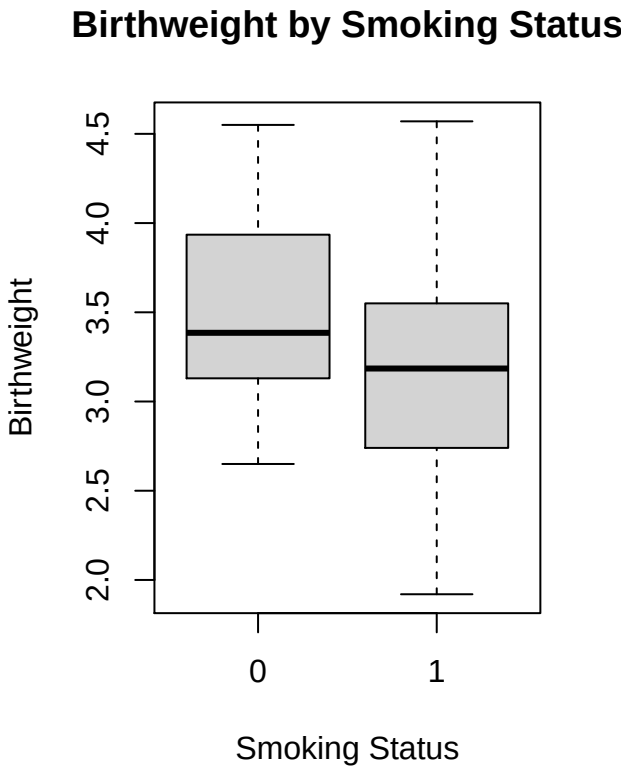



Figure 2: Smoking mothers appear to have lower days of gestation compared to nonsmoking mothers, but are more varied in range of days of gestation.

```
par(mfrow=c(1,2))
boxplot(Birthweight, main = "Birthweight", xlab = "Total Birthweight")
boxplot(Birthweight~smoker, main = "Birthweight by Smoking Status", xlab = "Smoking Status")
```

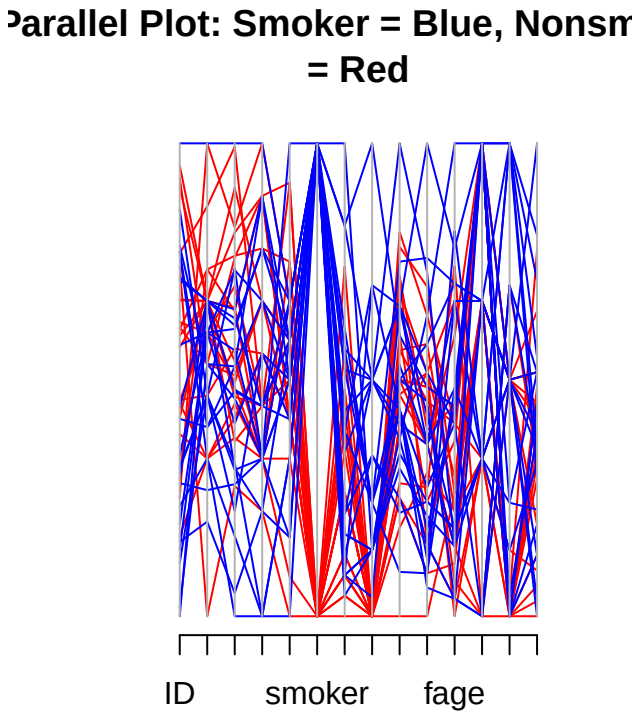



```
library(MASS)
# get colors
parallel.col <- ifelse(birth$smoker == 0, c("Red"), c("Blue"))
parcoord(x= birth, col = parallel.col, main= "Parallel Plot: Smoker = Blue, Nonsmoker
= Red")
colMeans(birth)
```

```
##          ID      Length Birthweight    Headcirc   Gestation      smoker
## 894.0714286  51.3333333   3.3128571  34.5952381  39.1904762   0.5238095
##        mage      mnocig     mheight       mppwt        fage      fedyrs
##  25.5476190   9.4285714 164.4523810  57.5000000  28.9047619  13.6666667
##      fnocig     fheight
##  17.1904762 180.5000000
```

```
par(mfrow=c(1,1))
```

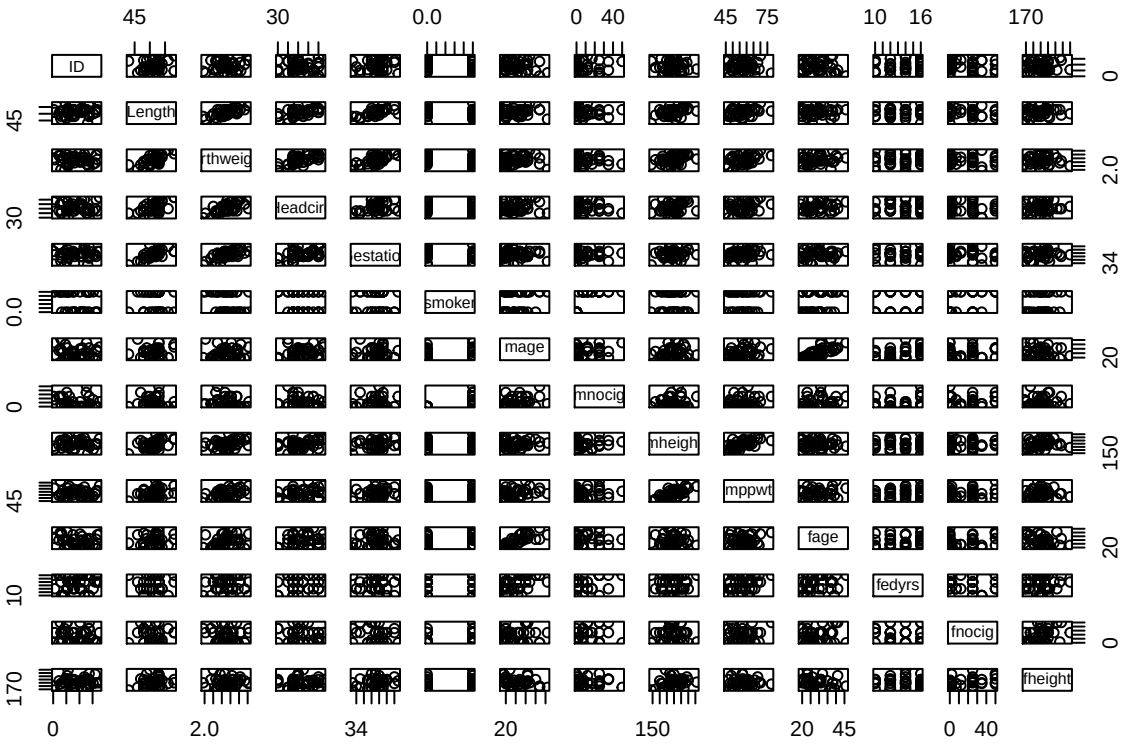



```
cor(birth)
```

```
##                      ID       Length   Birthweight    Headcirc   Gestation
## ID           1.00000000  0.192635194  0.1112107085 -0.04489917  0.10271426
## Length       0.19263519  1.000000000  0.7268334848  0.56317161  0.70511059
## Birthweight  0.11121071  0.726833485  1.0000000000  0.68461562  0.70830289
## Headcirc    -0.04489917  0.563171606  0.6846156185  1.00000000  0.40463477
## Gestation    0.10271426  0.705110591  0.7083028937  0.40463477  1.00000000
## smoker      -0.27618220 -0.153406153 -0.3142339464 -0.18287185 -0.09474608
## mage        -0.18642559  0.075268357  0.0001731023  0.14584152  0.01077846
## mnocig      -0.19462487 -0.039842761 -0.1523351845 -0.13298772  0.04319486
## mheight     -0.01089325  0.484992403  0.3630551926  0.33704682  0.21050318
## mppwt       -0.02651624  0.398197402  0.4008856280  0.30285407  0.25508237
## fage        -0.23195803  0.137184365  0.1757099933  0.30115080  0.14217533
## fedyrs       0.11992711  0.079484567  0.0710452258  0.12389250  0.13098664
## fnocig      -0.11288857  0.008800476 -0.0931357623 -0.04683678 -0.11383061
## fheight      0.08901791  0.208358435  0.0310224972  0.04150923  0.20759684
##                    smoker          mage      mnocig       mheight       mppwt
## ID          -0.2761821977 -0.1864255867 -0.19462487 -0.0108932452 -0.02651624
## Length      -0.1534061527  0.0752683574 -0.03984276  0.4849924026  0.39819740
## Birthweight -0.3142339464  0.0001731023 -0.15233518  0.3630551926  0.40088563
## Headcirc    -0.1828718541  0.1458415206 -0.13298772  0.3370468163  0.30285407
## Gestation   -0.0947460785  0.0107784551  0.04319486  0.2105031792  0.25508237
## smoker       1.0000000000  0.2124787863  0.72721809  0.0003532676  0.00000000
## mage         0.2124787863  1.0000000000  0.34029438  0.0599563823  0.27416768
## mnocig       0.7272180924  0.3402943777  1.00000000  0.1264388844  0.14894461
```

```
## mheight      0.0003532676  0.0599563823  0.12643888  1.0000000000  0.68062174
## mppwt        0.0000000000  0.2741676755  0.14894461  0.6806217412  1.00000000
## fage         0.1975014481  0.8065844174  0.24842538 -0.0798698948  0.25570584
## fedyrs      -0.0148905839  0.4416826598  0.19852620  0.0352970193  0.18037409
## fnocig       0.4176329588  0.0909266363  0.25730739  0.0483980589  0.05716254
## fheight      0.1106327308 -0.1995468633  0.02067224  0.2743379320  0.09298347
##                    fage      fedyrs       fnocig     fheight
## ID          -0.23195803  0.11992711 -0.112888569  0.08901791
## Length       0.13718437  0.07948457  0.008800476  0.20835843
## Birthweight  0.17570999  0.07104523 -0.093135762  0.03102250
## Headcirc     0.30115080  0.12389250 -0.046836781  0.04150923
## Gestation    0.14217533  0.13098664 -0.113830614  0.20759684
## smoker       0.19750145 -0.01489058  0.417632959  0.11063273
## mage         0.80658442  0.44168266  0.090926636 -0.19954686
## mnocig       0.24842538  0.19852620  0.257307386  0.02067224
## mheight     -0.07986989  0.03529702  0.048398059  0.27433793
## mppwt        0.25570584  0.18037409  0.057162540  0.09298347
## fage         1.00000000  0.30047147  0.135862017 -0.26937685
## fedyrs       0.30047147  1.00000000 -0.263103019  0.01779765
## fnocig       0.13586202 -0.26310302  1.000000000  0.32936416
## fheight     -0.26937685  0.01779765  0.329364159  1.00000000
```

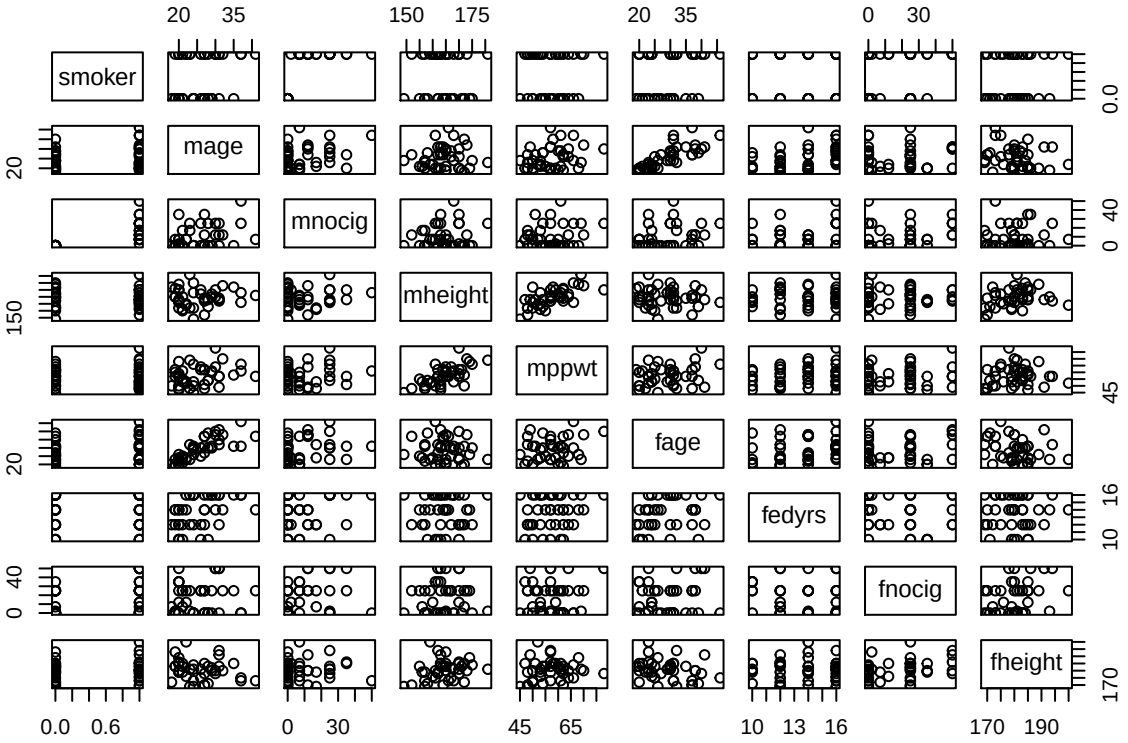
```
pairs(birth)
```




```
parents <- data.frame(smoker, mage, mnocig, mheight, mppwt, fage, fedyrs, fnocig, fheight)
round(cor(parents),4)
```

```
##           smoker    mage mnocig mheight  mppwt    fage  fedyrs  fnocig fheight
## smoker    1.0000  0.2125 0.7272  0.0004 0.0000  0.1975 -0.0149  0.4176  0.1106
## mage      0.2125  1.0000 0.3403  0.0600 0.2742  0.8066  0.4417  0.0909 -0.1995
## mnocig    0.7272  0.3403 1.0000  0.1264 0.1489  0.2484  0.1985  0.2573  0.0207
## mheight   0.0004  0.0600 0.1264  1.0000 0.6806 -0.0799  0.0353  0.0484  0.2743
## mppwt     0.0000  0.2742 0.1489  0.6806 1.0000  0.2557  0.1804  0.0572  0.0930
## fage      0.1975  0.8066 0.2484 -0.0799 0.2557  1.0000  0.3005  0.1359 -0.2694
## fedyrs   -0.0149  0.4417 0.1985  0.0353 0.1804  0.3005  1.0000 -0.2631  0.0178
## fnocig    0.4176  0.0909 0.2573  0.0484 0.0572  0.1359 -0.2631  1.0000  0.3294
## fheight   0.1106 -0.1995 0.0207  0.2743 0.0930 -0.2694  0.0178  0.3294  1.0000
```

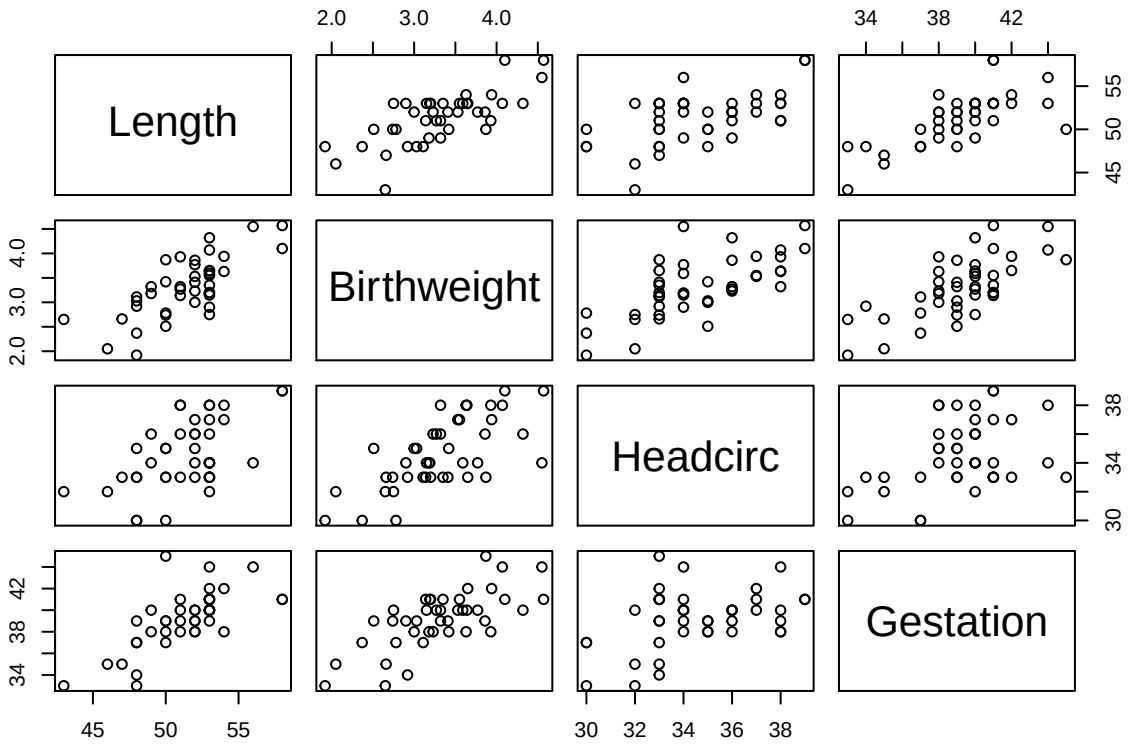
```
pairs(parents)
```



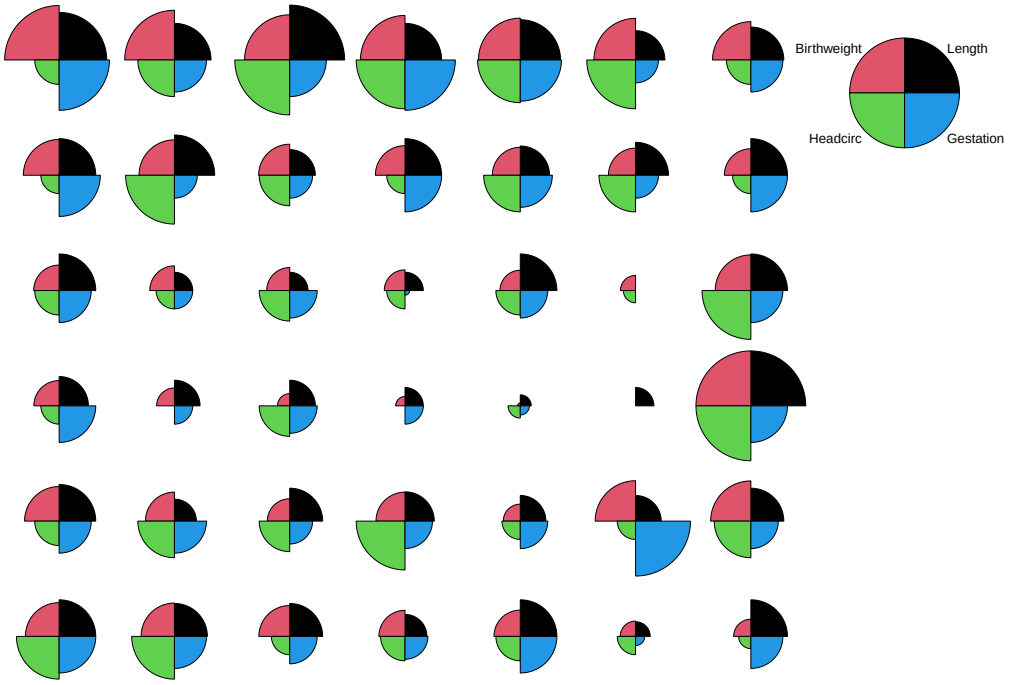

```
phys <- data.frame(Length, Birthweight, Headcirc, Gestation)
round(cor(phys),4)
```

```
##             Length Birthweight Headcirc Gestation
## Length      1.0000      0.7268   0.5632    0.7051
## Birthweight 0.7268      1.0000   0.6846    0.7083
## Headcirc    0.5632      0.6846   1.0000    0.4046
## Gestation   0.7051      0.7083   0.4046    1.0000
```

```
pairs(phys)
```




```
stars(x = phys, ncol = 7, draw.segments = T, key.loc = c(17.5,12), cex = 0.4)
```




```r
# separate by mother's smoking status
# Nonsmoker
pca_nonsmoker <- princomp((birth[birth$smoker == 0,-c(1,6)]))

# Smoker
pca_smoker <- princomp(birth[birth$smoker == 1,-c(1,6)])

# get pca summary
print(summary(pca_smoker, loadings = TRUE))
```

```
## Importance of components:
##                            Comp.1     Comp.2    Comp.3     Comp.4     Comp.5
## Standard deviation     18.2010014 13.1693540 9.3739499 8.52654435 5.97088377
## Proportion of Variance  0.4523284  0.2368058 0.1199799 0.09926803 0.04867888
## Cumulative Proportion   0.4523284  0.6891342 0.8091141 0.90838213 0.95706101
##                            Comp.6     Comp.7      Comp.8      Comp.9
## Standard deviation     3.82939145 3.05975421 1.645299140 1.411144588
## Proportion of Variance 0.02002271 0.01278311 0.003696179 0.002718982
## Cumulative Proportion  0.97708371 0.98986682 0.993563000 0.996281982
##                           Comp.10     Comp.11      Comp.12
## Standard deviation     1.311482522 0.985957764 1.757964e-01
## Proportion of Variance 0.002348488 0.001327333 4.219716e-05
## Cumulative Proportion  0.998630470 0.999957803 1.000000e+00
##
## Loadings:
##             Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8 Comp.9
```

```
## Length                                0.107  0.310  0.330  0.160  0.588
## Birthweight
## Headcirc                  0.124               0.413  0.274 -0.364  0.323
## Gestation                                     0.190  0.308  0.221  0.509
## mage                0.308  0.352 -0.158  0.286 -0.643                0.448
## mnocig       0.107  0.757 -0.573 -0.271
## mheight             0.203         0.583 -0.140 -0.355  0.632        -0.224
## mppwt               0.371  0.258  0.641 -0.125  0.222 -0.532
## fage                0.334  0.560 -0.219  0.312  0.183  0.145 -0.113 -0.531
## fedyrs                                        0.120        -0.215  0.720
## fnocig      -0.975                        -0.148
## fheight     -0.149 -0.122 -0.372  0.303  0.834        -0.100 -0.153
##             Comp.10 Comp.11 Comp.12
## Length       0.610   0.139
## Birthweight                 -0.985
## Headcirc    -0.663   0.232
## Gestation   -0.235  -0.685   0.153
## mage        -0.145  -0.190
## mnocig
## mheight
## mppwt               -0.160
## fage         0.197   0.182
## fedyrs      -0.256   0.588
## fnocig
## fheight
```

```
print(summary(pca_nonsmoker, loadings = TRUE))
```

```
## Importance of components:
##                            Comp.1    Comp.2    Comp.3     Comp.4     Comp.5
## Standard deviation     13.4236349 8.6098894 6.7799060 4.60464940 3.87261789
## Proportion of Variance  0.5106122 0.2100613 0.1302562 0.06008196 0.04249719
## Cumulative Proportion   0.5106122 0.7206735 0.8509297 0.91101164 0.95350883
##                            Comp.6      Comp.7      Comp.8      Comp.9
## Standard deviation     3.04425920 1.741297942 1.501975557 1.075009822
## Proportion of Variance 0.02626118 0.008592056 0.006392587 0.003274732
## Cumulative Proportion  0.97977001 0.988362066 0.994754654 0.998029386
##                           Comp.10      Comp.11 Comp.12
## Standard deviation     0.803655947 0.2226267921       0
## Proportion of Variance 0.001830169 0.0001404448       0
## Cumulative Proportion  0.999859555 1.0000000000       1
##
## Loadings:
##             Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8 Comp.9
## Length       0.135  0.153                0.388  0.438  0.347  0.360  0.408
## Birthweight                                                0.106        -0.155
## Headcirc            0.155                0.284         0.202 -0.188 -0.754
## Gestation    0.111  0.119                       0.593  0.342 -0.160 -0.264
## mage                      -0.588  0.217  0.193 -0.443  0.328  0.415
## mnocig
## mheight      0.201  0.628        -0.448  0.411 -0.167 -0.310         0.123
## mppwt        0.216  0.464 -0.232 -0.192 -0.723         0.297
## fage        -0.142        -0.701  0.186         0.397 -0.468 -0.213
## fedyrs                            0.224  0.169 -0.229  0.371 -0.755  0.386
## fnocig       0.896 -0.390 -0.157
```

```
## fheight      0.234  0.413  0.281  0.788               -0.234
##             Comp.10 Comp.11 Comp.12
## Length       0.441
## Birthweight         -0.977
## Headcirc     0.460   0.172
## Gestation   -0.628
## mage        -0.294
## mnocig                      1.000
## mheight     -0.217
## mppwt        0.185
## fage         0.163
## fedyrs
## fnocig
## fheight
```

```r
# add newly born data
newly_born <- c(61,5.1,36,43,0,43,7,165,64,38,19,45,189)[-6] # last two columns already dropped
birth_new <- rbind(birth, newly_born)

# get newly born PCs
newly_born_pcs <- predict(pca_smoker, newdata = birth_new[dim(birth)[1],])
plot(pca_smoker$scores[,1],pca_smoker$scores[,2],xlab="PC1",ylab="PC2",main="First 2 PCs: Newly Born = R
points(x=newly_born_pcs[1], y=newly_born_pcs[2],col='red')
```