Data Analysis Report

Birthweights

Todo

Cholesterol

TODO

Diet

Dataset peak:

```
head(diet)
```

```
##   person gender age height preweight diet weight6weeks weight_loss
## 1      1      0  22    159        58    1         54.2         3.8
## 2      2      0  46    192        60    1         54.0         6.0
## 3      3      0  55    170        64    1         63.3         0.7
## 4      4      0  33    171        64    1         61.1         2.9
## 5      5      0  50    170        65    1         62.2         2.8
## 6      6      0  50    201        66    1         64.0         2.0
```

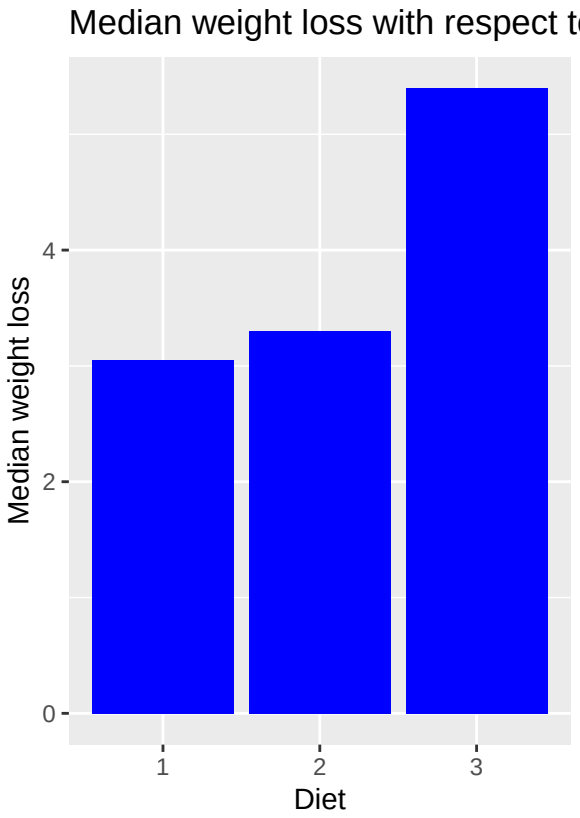
a) Informative graphical summary

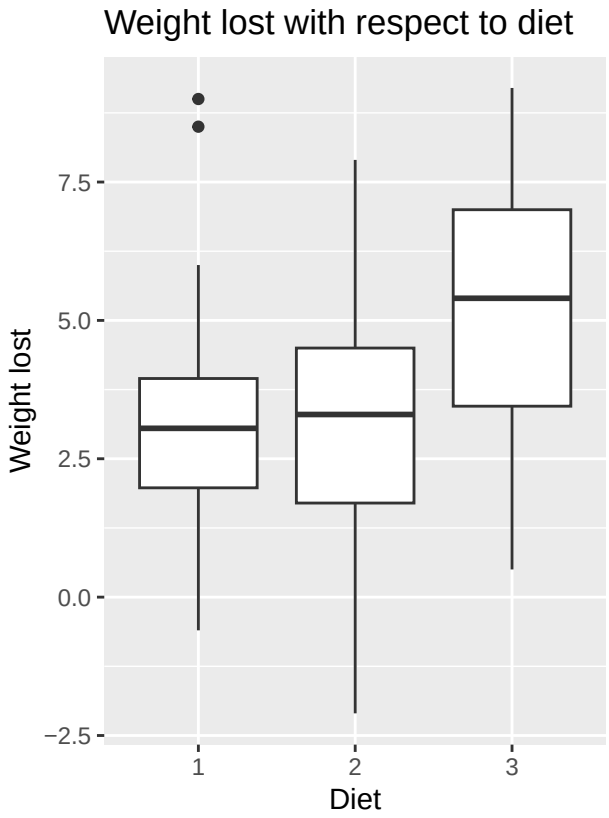
We will visualise the weight loss and the average weight loss with respect to each diet.

```
box_plot <- diet %>%
  ggplot(aes(x = factor(diet), y = weight_loss)) +
  geom_boxplot() +
  labs(x = "Diet", y = "Weight lost", title = "Weight lost with respect to diet")

median_plot <- diet %>%
  group_by(diet) %>%
  summarize(median_weight_loss = median(weight_loss)) %>%
  ggplot(aes(x=factor(diet), y=median_weight_loss)) +
  geom_bar(stat="identity", fill="blue") +
  labs(x="Diet", y ="Median weight loss", title="Median weight loss with respect to diet")

median_plot + box_plot
```





Based on the visual information of weight loss and medium weight loss we argue that the third diet was the most efficient!

b) One-way Anova to test whether the the diet has an effect on the weight loss

We assume that the weight loss is significantly different across all types of diet.

```
fit <- aov(weight_loss ~ diet , data = diet)
summary(fit)
```

```
##             Df Sum Sq Mean Sq F value  Pr(>F)
## diet         1   45.8   45.78   7.639 0.00716 **
## Residuals   76  455.5    5.99
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We observe a very low value for P(F) « 1%, thus we will reject our hypothesis and conclude that at least one of the diets must have an effect over the weight loss.

Given the high mean values for weight loss with respect to diet we conclude that the best diets were: 3, 2 and 1:

```
diet %>%
  group_by(diet) %>%
  summarize(median_weight_loss = median(weight_loss)) %>%
  arrange(desc(median_weight_loss))
```

```
## # A tibble: 3 x 2
##    diet median_weight_loss
##   <int>              <dbl>
## 1     3                5.4
```

```
## 2     2               3.30
## 3     1               3.05
```

b) Two-way ANOVA to investigate effect of the diet and gender

```
fit <- aov(weight_loss ~ diet * gender, data = diet)
summary(fit)
```

```
##             Df Sum Sq Mean Sq F value  Pr(>F)
## diet         1   45.2   45.21   7.957 0.00619 **
## gender       1    0.1    0.14   0.025 0.87521
## diet:gender  1   16.5   16.47   2.898 0.09300 .
## Residuals   72  409.1    5.68
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 2 observations deleted due to missingness
```

We observe:

- A very low value P(F) « 1% for diet
  - We conclude diet has an effect on weight loss
- A very high value P(F) » 85% for gender
  - We conclude gender has no significant effect on weight loss
- A low value P(F) « 5% for the factorized effect of diet & gender on weight loss.
  - We conclude that there must be a causal effect between the gender and a particular diet.

e) Preferred ANOVA

The results of the Two-way ANOVA suggest that the effectiveness of a diet is significantly associated with the gender of the participant.

Yield of peas

TODO
