HW1

Consider the following soap experiment, run as a class project by Suyapa Silvia in 1985. The purpose of this experiment is to compare the extent to which three particular types of soap dissolve in water. It is expected that the experiment will answer the following questions: Are there any differences in weight loss due to dissolution among the three soaps when allowed to soak in water for the same length of time? What are these differences?

Four cubes of three different types of soap (Regular, Deodorant, and Moisturizing) were weighed and then placed in baking pans filled with water kept at room temperature. Note that there are 12 total cubes of soap used in the experiment. After sitting in the water for 24 hours, the cubes of soap were weighed again. The weight lost (weight before minus weight after) by the twelve cubes of soap are shown in Table 1.

Do the following. Turn in your completed homework as a PDF document created using R markdown. Include all R code used, and clearly label which R code corresponds to which problem. Make your final report clean and clear. Don't just slap your answers and code together. Take the time to write a sentence or two describing how everything comes together.

Q1

1. Type the soap experiment data into R. Format the data when you enter it as a data frame with two columns, one for the type of soap and another for the weight lost in the experiment. Each row of your data object should correspond to a single bar of soap (a single experimental unit), so you should end up with 12 rows.

```
stride = 4
types = c('Regular', 'Deodorant', 'Moisturizing')
soap_type = c(rep(types[1], 4),
              rep(types[2], 4),
              rep(types[3], 4))
weight_loss = c(-0.30, -0.10, -0.14, 0.40, 2.63, 2.61, 2.41, 3.15, 1.86, 2.03, 2.26, 1.82)
soap_data = data.frame(soap_type, weight_loss)
soap_data
```

```
##       soap_type weight_loss
## 1       Regular       -0.30
## 2       Regular       -0.10
## 3       Regular       -0.14
## 4       Regular        0.40
## 5     Deodorant        2.63
## 6     Deodorant        2.61
## 7     Deodorant        2.41
## 8     Deodorant        3.15
## 9  Moisturizing        1.86
## 10 Moisturizing        2.03
## 11 Moisturizing        2.26
## 12 Moisturizing        1.82
```

Q2

Convert the measurements of weight lost to kilograms.

```
to_kg = 1000
soap_data$weight_loss = soap_data$weight_loss/to_kg
soap_data
```

```
##       soap_type weight_loss
## 1       Regular    -0.00030
## 2       Regular    -0.00010
## 3       Regular    -0.00014
## 4       Regular     0.00040
## 5     Deodorant     0.00263
## 6     Deodorant     0.00261
## 7     Deodorant     0.00241
## 8     Deodorant     0.00315
## 9  Moisturizing     0.00186
## 10 Moisturizing     0.00203
## 11 Moisturizing     0.00226
## 12 Moisturizing     0.00182
```

Q3

3. Use R to compute the mean and standard deviation of the observed weight lost (in kilograms). Do this first for all 12 data points, and then compute the mean and standard deviation for each of the 3 types of soap.

- for all 12 data points:

```
total_mean = mean(soap_data$weight_loss)
total_sd = sd(soap_data$weight_loss)
cat('The total mean is: ', total_mean, 'kg\n')
```

```
## The total mean is:  0.0015525 kg
```

```
cat('The total standard deviation is: ', total_sd, 'kg\n\n')
```

```
## The total standard deviation is:  0.00123644 kg
```

- for each of the 3 types:

```
for (idx in seq(1,length(soap_data$weight_loss), by=stride)) {
  start = idx
  end = idx+stride-1
  type_idx = floor(idx/stride)
  batch = soap_data$weight_loss[start:end]
  cat('The mean of Type', types[type_idx+1], 'is: ', mean(batch), 'kg\n')
  cat('The standard deviation of Type', types[type_idx+1], 'is: ', sd(batch), 'kg\n\n')
}
```

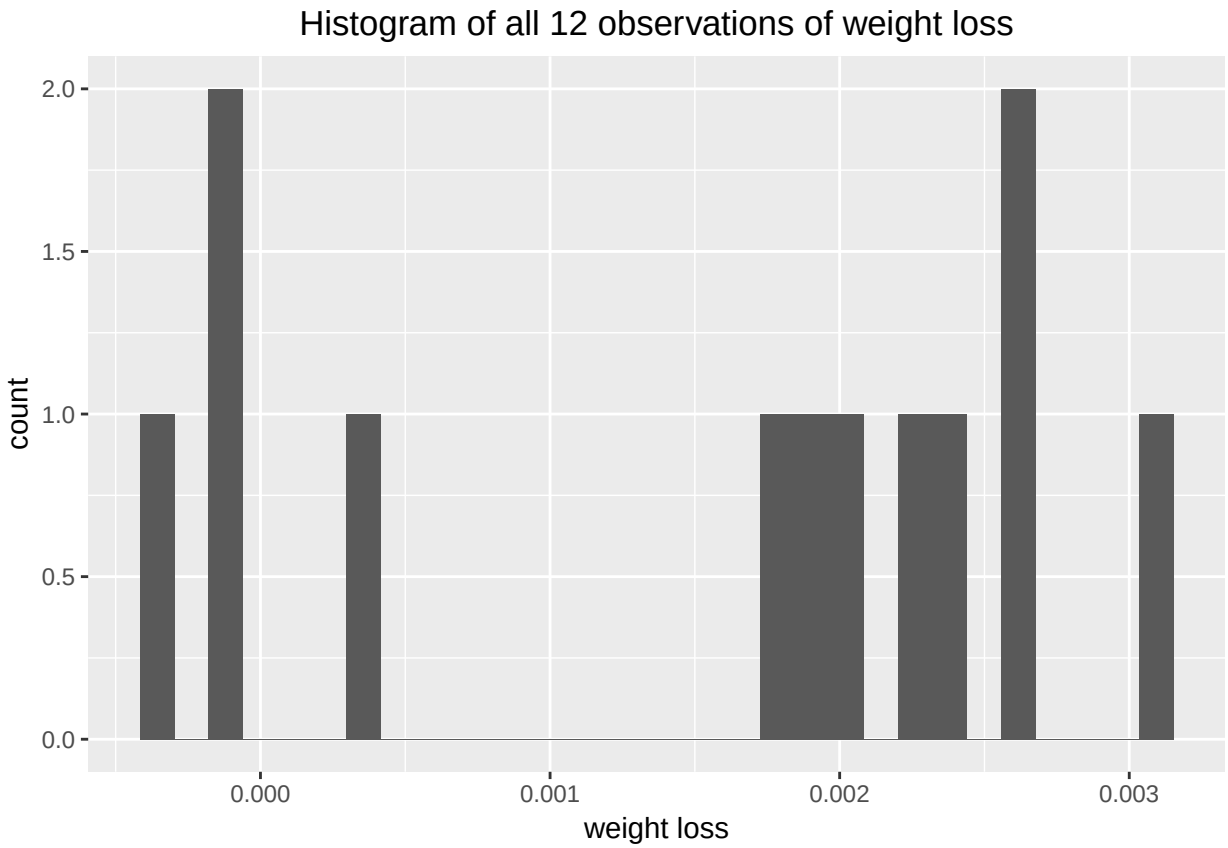
```
## The mean of Type Regular is:  -3.5e-05 kg
## The standard deviation of Type Regular is:  0.0003025998 kg
##
## The mean of Type Deodorant is:  0.0027 kg
## The standard deviation of Type Deodorant is:  0.0003160169 kg
##
## The mean of Type Moisturizing is:  0.0019925 kg
## The standard deviation of Type Moisturizing is:  0.000200229 kg
```

Q4

Plot a histogram of all of the observations from the experiment (in kilograms).

```
library(ggplot2)
ggplot(soap_data, aes(x=soap_data$weight_loss)) +
  geom_histogram() +
  ggtitle("Histogram of all 12 observations of weight loss") +
  theme(plot.title = element_text(hjust = 0.5)) +
  xlab('weight loss')
```

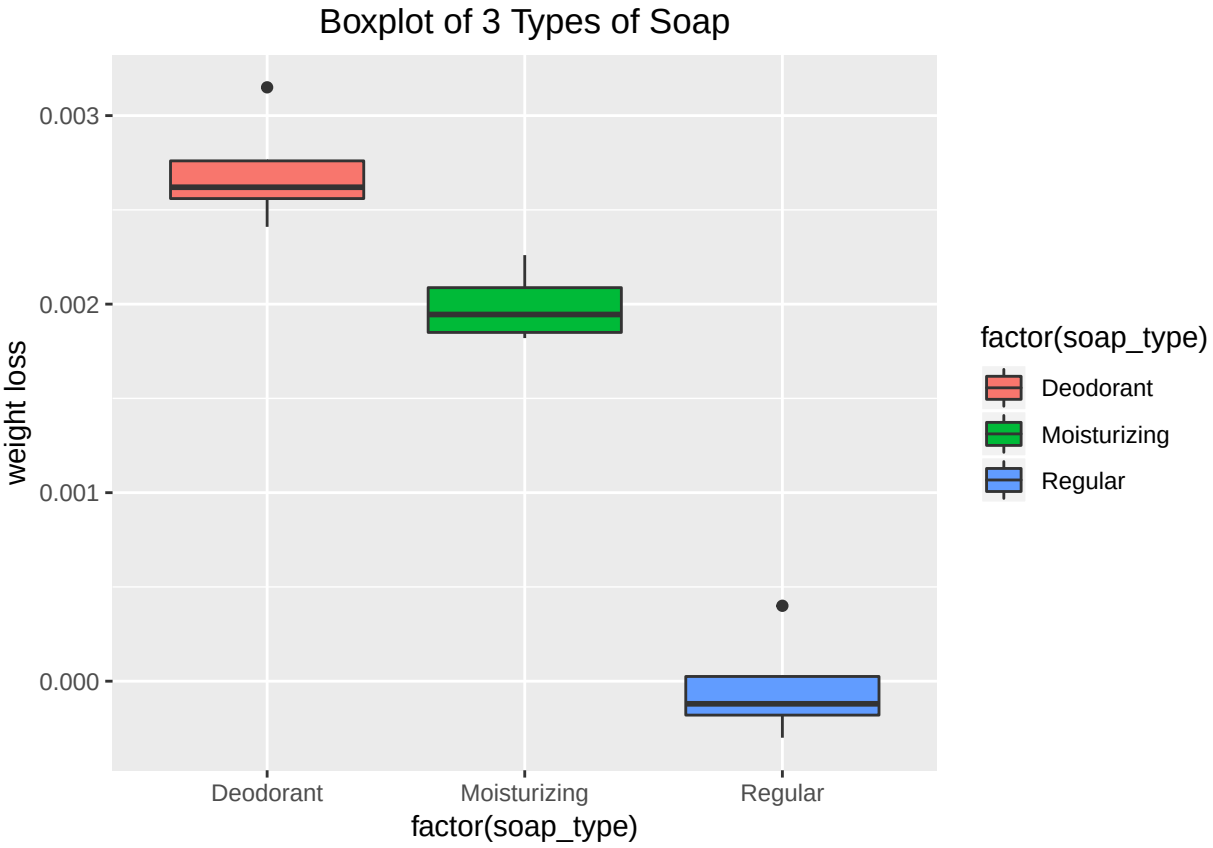
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Q5

5. Plot side-by-side boxplots of the weight lost (in kilograms), with one boxplot for each type of soap.

```
library(reshape2)
soap_data_melted <- melt(soap_data, id.vars=c("soap_type","weight_loss"))
box <- ggplot(soap_data_melted, aes(x=factor(soap_type),y=weight_loss,fill=factor(soap_type))) +
            geom_boxplot() +
            labs(title=paste("Boxplot of 3 Types of Soap")) +
            theme(plot.title = element_text(hjust = 0.5)) +
            ylab('weight loss')
box
```



Q6

6. Comment on any observed differences between the soap types that you think would be worth investigating further.

- Only regular type has the negative weight loss, the other two all have positive weight loss;
- There seems to exist an outliner in Regular and Deodorant type from observing the boxplot of the 3 types;
- The variances of weight loss among the 3 types appear to be similar