

# EEEB UN3005/GR5005

## Homework - Week 02 - Due 12 Feb 2019

*Xun Zhao, xz2827*

**Homework Instructions:** Complete this assignment by writing code in the code chunks provided. If required, provide written explanations below the relevant code chunks. Replace “USE YOUR NAME HERE” with your name in the document header. When complete, knit this document within RStudio to generate a pdf. Please review the resulting pdf to ensure that all content relevant for grading (i.e., code, code output, and written explanations) appears in the document. Rename your pdf document according to the following format: hw\_week\_02\_firstname\_lastname.pdf. Upload this final homework document to CourseWorks by 5 pm on the due date.

### Problem 1 (1 point)

On the class CourseWorks and GitHub site you will find a dataset called `orangutans.csv`. This file contains hypothetical data on weight, sex, and observed tool use for orangutans from Sumatra and Borneo. Import this data into R, assigning the data to an object called `orangutans`. Use a summary function of your choice to get an idea of the data structure you’re working with. Which variable(s) in the data represent numeric data?

```
orangutans = read.csv('orangutans.csv')
summary(orangutans)
```

```
##      individual      location  weight_kg      sex      tool_use
## A          :1      Borneo :7  Min.    : 30.00  female:9  Mode :logical
## B          :1      Sumatra:8  1st Qu.: 37.50  male  :6  FALSE:7
## C          :1                      Median : 42.00                      TRUE :8
## D          :1                      Mean    : 55.93
## E          :1                      3rd Qu.: 74.50
## F          :1                      Max.    :110.00
## (Other):9
```

```
structure(orangutans)
```

```
##      individual location weight_kg      sex tool_use
## 1            A   Borneo      105  male     TRUE
## 2            B   Borneo       72  male    FALSE
## 3            C   Borneo       60  male    FALSE
## 4            D   Borneo       43 female    TRUE
## 5            E   Borneo       41 female    TRUE
## 6            F   Borneo       38 female    TRUE
```

```
## 7      G Borneo      33 female FALSE
## 8      H Sumatra    110  male  FALSE
## 9      I Sumatra     81  male  FALSE
## 10     J Sumatra     77  male   TRUE
## 11     K Sumatra     42 female  TRUE
## 12     L Sumatra     38 female  TRUE
## 13     M Sumatra     37 female  FALSE
## 14     N Sumatra     32 female  TRUE
## 15     O Sumatra     30 female  FALSE
```

**Answer:** Based on the summaries given above, we can know that the variable `weight_kg` is numeric.

## Problem 2 (2 points)

Using `dplyr`, return the row of data for individual J. Perform the same operation using bracket subsetting (i.e., the base R solution).

```
filter(orangutans, individual == 'J') # 'dplyr' operation
```

```
## individual location weight_kg sex tool_use
## 1      J Sumatra      77 male   TRUE
```

```
orangutans[orangutans$individual == 'J',] # base R operation
```

```
## individual location weight_kg sex tool_use
## 10      J Sumatra      77 male   TRUE
```

## Problem 3 (2 points)

Using `dplyr`, return all rows of data that correspond to orangutans sampled from Sumatra that are greater than 40 kg in weight.

```
filter(orangutans, location == 'Sumatra' & weight_kg > 40)
```

```
## individual location weight_kg sex tool_use
## 1      H Sumatra    110  male  FALSE
## 2      I Sumatra     81  male  FALSE
## 3      J Sumatra     77  male   TRUE
## 4      K Sumatra     42 female  TRUE
```

## Problem 4 (2 points)

Using `dplyr`, sort the `orangutans` data frame according to weight, from high to low, and return only the `individual` and `weight_kg` variables.

```
select(arrange(orangutans, desc(weight_kg)), individual, weight_kg)
```

```
##      individual weight_kg
## 1           H       110
## 2           A       105
## 3           I        81
## 4           J        77
## 5           B        72
## 6           C        60
## 7           D        43
## 8           K        42
## 9           E        41
## 10          F        38
## 11          L        38
## 12          M        37
## 13          G        33
## 14          N        32
## 15          O        30
```

## Problem 5 (3 points)

Using `dplyr`, calculate the mean weight for orangutans of each sex in each in location. In other words, you should be returning one mean weight value for each of the following groups: females from Borneo, females from Sumatra, males from Borneo, and males from Sumatra.

```
summarize(group_by(orangutans, sex, location), mean(weight_kg))
```

```
## # A tibble: 4 x 3
## # Groups:   sex [?]
##   sex    location `mean(weight_kg)`
##   <fct> <fct>          <dbl>
## 1 female Borneo          38.8
## 2 female Sumatra         35.8
## 3 male   Borneo          79
## 4 male   Sumatra        89.3
```