HW3_Lian

Jiayi Lian

September 13, 2019

Homework3

p3

```
# Store url
# In the order of "sensory data, gold medal performance, brain weight vs body weight,
# triplicate measurements of tomato."
url_sensory<-"http://www2.isye.gatech.edu/~jeffwu/wuhamadabook/data/Sensory.dat"
url_gold<-"http://www2.isye.gatech.edu/~jeffwu/wuhamadabook/data/LongJumpData.dat"
url_brain<-"http://www2.isye.gatech.edu/~jeffwu/wuhamadabook/data/BrainandBodyWeight.dat"
url_tomato<-"http://www2.isye.gatech.edu/~jeffwu/wuhamadabook/data/tomato.dat"
```

```
#download into local memory
Sensory<-read.table(url_sensory, header=F, skip=1, fill=T, stringsAsFactors = F)
#manipulate data into a tidy format
#transfer data into a vector
Sen_2<-Sensory[-1,]
Sen_2_a<-Sen_2 %>% filter(V1 %in% 1:10) %>%
                   rename(Item=V1,V1=V2,V2=V3,V3=V4,V4=V5,V5=V6)
Sen_2_b<-Sen_2 %>% filter(!(V1 %in% 1:10)) %>%
                   mutate(Item=rep(as.character(1:10),each=2)) %>%
                   mutate(V1=as.numeric(V1)) %>%
                   select(c(Item,V1:V5))
Sen_final<-Sen_2_a %>% full_join(Sen_2_b)
```

```
## Joining, by = c("Item", "V1", "V2", "V3", "V4", "V5")
```

```
#Show table of Sensory data
Sen_final
```

```
##    Item  V1  V2  V3  V4  V5
## 1     1 4.3 4.9 3.3 5.3 4.4
## 2     2 6.0 5.3 4.5 5.9 4.7
## 3     3 2.4 2.5 2.3 3.1 2.4
## 4     4 7.4 8.2 6.4 6.8 6.0
## 5     5 5.7 6.3 5.4 6.1 5.9
## 6     6 2.2 2.4 1.7 3.4 1.7
## 7     7 1.2 1.5 1.2 0.9 0.7
## 8     8 4.2 4.8 4.5 4.6 3.2
## 9     9 8.0 8.6 9.0 9.4 8.8
## 10   10 5.0 4.8 3.9 5.5 3.8
## 11    1 4.3 4.5 4.0 5.5 3.3
## 12    1 4.1 5.3 3.4 5.7 4.7
## 13    2 4.9 6.3 4.2 5.5 4.9
## 14    2 6.0 5.9 4.7 6.3 4.6
## 15    3 3.9 3.0 2.8 2.7 1.3
## 16    3 1.9 3.9 2.6 4.6 2.2
## 17    4 7.1 7.9 5.9 7.3 6.1
```

```
## 18    4 6.4 7.1 6.9 7.0 6.7
## 19    5 5.8 5.7 5.4 6.2 6.5
## 20    5 5.8 6.0 6.1 7.0 4.9
## 21    6 3.0 1.8 2.1 4.0 1.7
## 22    6 2.1 3.3 1.1 3.3 2.1
## 23    7 1.3 2.4 0.8 1.2 1.3
## 24    7 0.9 3.1 1.1 1.9 1.6
## 25    8 3.0 4.5 4.7 4.9 4.6
## 26    8 4.8 4.8 4.7 4.8 4.3
## 27    9 9.0 7.7 6.7 9.0 7.9
## 28    9 8.9 9.2 8.1 9.1 7.6
## 29   10 5.4 5.0 3.4 4.9 4.6
## 30   10 2.8 5.2 4.1 3.9 5.5
```

```
str(Sen_final)
```

```
## 'data.frame':    30 obs. of  6 variables:
##  $ Item: chr  "1" "2" "3" "4" ...
##  $ V1  : num  4.3 6 2.4 7.4 5.7 2.2 1.2 4.2 8 5 ...
##  $ V2  : num  4.9 5.3 2.5 8.2 6.3 2.4 1.5 4.8 8.6 4.8 ...
##  $ V3  : num  3.3 4.5 2.3 6.4 5.4 1.7 1.2 4.5 9 3.9 ...
##  $ V4  : num  5.3 5.9 3.1 6.8 6.1 3.4 0.9 4.6 9.4 5.5 ...
##  $ V5  : num  4.4 4.7 2.4 6 5.9 1.7 0.7 3.2 8.8 3.8 ...
```

```
summary(Sen_final)
```

```
##      Item                V1              V2              V3
##  Length:30          Min.   :0.900   Min.   :1.500   Min.   :0.800
##  Class :character   1st Qu.:2.850   1st Qu.:3.450   1st Qu.:2.650
##  Mode  :character   Median :4.550   Median :4.950   Median :4.150
##                     Mean   :4.593   Mean   :5.063   Mean   :4.167
##                     3rd Qu.:5.950   3rd Qu.:6.225   3rd Qu.:5.400
##                     Max.   :9.000   Max.   :9.200   Max.   :9.000
##        V4              V5
##  Min.   :0.900   Min.   :0.700
##  1st Qu.:3.925   1st Qu.:2.250
##  Median :5.400   Median :4.600
##  Mean   :5.193   Mean   :4.267
##  3rd Qu.:6.275   3rd Qu.:5.800
##  Max.   :9.400   Max.   :8.800
```

```
#load gold medal performance dat
gold<-read.table(url_gold, header =  F,skip = 1, fill=T)
#mung
gold_a<-gold %>% select(c(V1,V2)) %>% mutate(Year= V1+1900, Performance=V2) %>% select(c(Year,Performan
gold_b<-gold %>% select(c(V3,V4)) %>% mutate(Year= V3+1900, Performance=V4) %>% select(c(Year,Performan
gold_c<-gold %>% select(c(V5,V6)) %>% mutate(Year= V5+1900, Performance=V6) %>% select(c(Year,Performan
gold_d<-gold %>% select(c(V7,V8)) %>% mutate(Year= V7+1900, Performance=V8) %>% select(c(Year,Performan
gold_final<-gold_a %>% full_join(gold_b) %>% full_join(gold_c) %>% full_join(gold_d)
```

```
## Joining, by = c("Year", "Performance")
## Joining, by = c("Year", "Performance")
## Joining, by = c("Year", "Performance")
```

```
#show
gold_final
```

```
##    Year Performance
## 1  1896      249.75
## 2  1900      282.88
## 3  1904      289.00
## 4  1908      294.50
## 5  1912      299.25
## 6  1920      281.50
## 7  1924      293.13
## 8  1928      304.75
## 9  1932      300.75
## 10 1936      317.31
## 11 1948      308.00
## 12 1952      298.00
## 13 1956      308.25
## 14 1960      319.75
## 15 1964      317.75
## 16 1968      350.50
## 17 1972      324.50
## 18 1976      328.50
## 19 1980      336.25
## 20 1984      336.25
## 21 1988      343.25
## 22 1992      342.50
```

```
str(gold_final)
```

```
## 'data.frame':    22 obs. of  2 variables:
##  $ Year       : num  1896 1900 1904 1908 1912 ...
##  $ Performance: num  250 283 289 294 299 ...
```

```
summary(gold_final)
```

```
##       Year       Performance
##  Min.   :1896   Min.   :249.8
##  1st Qu.:1921   1st Qu.:295.4
##  Median :1950   Median :308.1
##  Mean   :1945   Mean   :310.3
##  3rd Qu.:1971   3rd Qu.:327.5
##  Max.   :1992   Max.   :350.5
```

```
#load brain vs body weight dat
brain<-read.table(url_brain,header = F,skip = 1,fill = T)
#munge
brain_a<-brain %>% select(V1,V2) %>% mutate(Brain_weight=V1,Body_weight=V2) %>% select(Brain_weight,Body
brain_b<-brain %>% select(V3,V4) %>% mutate(Brain_weight=V3,Body_weight=V4) %>% select(Brain_weight,Body
brain_c<-brain %>% select(V5,V6) %>% mutate(Brain_weight=V5,Body_weight=V6) %>% select(Brain_weight,Body
brain_final<-brain_a %>% full_join(brain_b) %>% full_join(brain_c)
```

```
## Joining, by = c("Brain_weight", "Body_weight")
## Joining, by = c("Brain_weight", "Body_weight")
```

```
#show
print(brain_final)
```

```
##    Brain_weight Body_weight
## 1         3.385       44.50
## 2         0.480       15.50
```

```
## 3         1.350       8.10
## 4       465.000     423.00
## 5        36.330     119.50
## 6        27.660     115.00
## 7        14.830      98.20
## 8         1.040       5.50
## 9         4.190      58.00
## 10        0.425       6.40
## 11        0.101       4.00
## 12        0.920       5.70
## 13        1.000       6.60
## 14        0.005       0.10
## 15        0.060       1.00
## 16        3.500      10.80
## 17        2.000      12.30
## 18        1.700       6.30
## 19     2547.000    4603.00
## 20        0.023       0.30
## 21      187.100     419.00
## 22      521.000     655.00
## 23        0.785       3.50
## 24       10.000     115.00
## 25        3.300      25.60
## 26        0.200       5.00
## 27        1.410      17.50
## 28      529.000     680.00
## 29      207.000     406.00
## 30       85.000     325.00
## 31        0.750      12.30
## 32       62.000    1320.00
## 33     6654.000    5712.00
## 34        3.500       3.90
## 35        6.800     179.00
## 36       35.000      56.00
## 37        4.050      17.00
## 38        0.120       1.00
## 39        0.023       0.40
## 40        0.010       0.30
## 41        1.400      12.50
## 42      250.000     490.00
## 43        2.500      12.10
## 44       55.500     175.00
## 45      100.000     157.00
## 46       52.160     440.00
## 47       10.550     179.50
## 48        0.550       2.40
## 49       60.000      81.00
## 50        3.600      21.00
## 51        4.288      39.20
## 52        0.280       1.90
## 53        0.075       1.20
## 54        0.122       3.00
## 55        0.048       0.33
## 56      192.000     180.00
```

```
## 57        3.000       25.00
## 58      160.000      169.00
## 59        0.900        2.60
## 60        1.620       11.40
## 61        0.104        2.50
## 62        4.235       50.40
```

```
str(brain_final)
```

```
## 'data.frame':    62 obs. of  2 variables:
##  $ Brain_weight: num  3.38 0.48 1.35 465 36.33 ...
##  $ Body_weight : num  44.5 15.5 8.1 423 119.5 ...
```

```
summary(brain_final)
```

```
##   Brain_weight       Body_weight
##  Min.   :   0.005   Min.   :   0.10
##  1st Qu.:   0.600   1st Qu.:   4.25
##  Median :   3.342   Median :  17.25
##  Mean   : 198.790   Mean   : 283.13
##  3rd Qu.:  48.203   3rd Qu.: 166.00
##  Max.   :6654.000   Max.   :5712.00
```

```
#download tomato dat
tomato<-read.table(url_tomato,fill=T, ,skip=1,header = F, stringsAsFactors = F,comment.char="*")
#munge
t_meas<-tomato[-1,]

t_meas<-t_meas %>% mutate(V2=as.character(V2),S_20000=as.character(V2),V3=as.character(V3))%>% mutate(Sp
t_meas<-t_meas %>% separate(V2, into = c('first_10000','Second_10000',"Thrid_10000"), sep=",", extra =
#show and summary
t_meas
```

```
##          Spicies first_10000 Second_10000 Thrid_10000 first_20000
## 1        Ife\\#1        16.1         15.3        17.5        16.6
## 2 PusaEarlyDwarf         8.1          8.6        10.1        12.7
##   Second_20000 Thrid_20000 first_30000 Second_30000 Thrid_30000
## 1         19.2        18.5        20.8         18.0        21.0
## 2         13.7        11.5        14.4         15.4        13.7
```

```
str(t_meas)
```

```
## 'data.frame':    2 obs. of  10 variables:
##  $ Spicies     : chr  "Ife\\#1" "PusaEarlyDwarf"
##  $ first_10000 : chr  "16.1" "8.1"
##  $ Second_10000: chr  "15.3" "8.6"
##  $ Thrid_10000 : chr  "17.5" "10.1"
##  $ first_20000 : chr  "16.6" "12.7"
##  $ Second_20000: chr  "19.2" "13.7"
##  $ Thrid_20000 : chr  "18.5" "11.5"
##  $ first_30000 : chr  "20.8" "14.4"
##  $ Second_30000: chr  "18.0" "15.4"
##  $ Thrid_30000 : chr  "21.0" "13.7"
```

```
summary(t_meas)
```

```
##    Spicies          first_10000        Second_10000
##  Length:2           Length:2           Length:2
```

```
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##  Thrid_10000        first_20000        Second_20000
##  Length:2           Length:2           Length:2
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##  Thrid_20000        first_30000        Second_30000
##  Length:2           Length:2           Length:2
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##  Thrid_30000
##  Length:2
##  Class :character
##  Mode  :character
```