Programming in Base R

Calista Harris

Task 1: Basic Vector Practice

Question 1

```r
#pre-treatment blood pressure
pre <- c(130, 128, 116, 124, 133, 134, 118, 126, 114, 127,
         141, 138, 128, 140, 137, 131, 120, 128, 139, 135)

#after treatment blood pressure
post <- c(114, 98, 113, 99, 107, 116, 113, 111, 119, 117,
          101, 119, 130, 122, 106, 106, 124, 102, 117, 113)
```

Question 2

```r
#create subject names
subject_names <- paste("Subject", 1:20, sep = "_")

#assign names to vectors
names(pre) <- subject_names
names(post) <- subject_names
```

Question 3

```r
#calculate the change in blood pressure
diff_op <- pre - post

#print the results
diff_op
```

```
 Subject_1  Subject_2  Subject_3  Subject_4  Subject_5  Subject_6  Subject_7
        16         30          3         25         26         18          5
 Subject_8  Subject_9 Subject_10 Subject_11 Subject_12 Subject_13 Subject_14
        15         -5         10         40         19         -2         18
Subject_15 Subject_16 Subject_17 Subject_18 Subject_19 Subject_20
        31         25         -4         26         22         22
```

Question 4

```
#calculate the average of all differences (pre - post)
mean_diff_op <- mean(diff_op)

#print the results
mean_diff_op
```

```
[1] 17
```

Question 5

```
#patient with decrease in blood pressure (positive change)
positive_change <- which(diff_op > 0)

#display results with names
positive_change
```

```
 Subject_1  Subject_2  Subject_3  Subject_4  Subject_5  Subject_6  Subject_7
         1          2          3          4          5          6          7
 Subject_8 Subject_10 Subject_11 Subject_12 Subject_14 Subject_15 Subject_16
         8         10         11         12         14         15         16
Subject_18 Subject_19 Subject_20
        18         19         20
```

Question 6

```r
#subset diff_op vector with positive change
subset_positive_change <- diff_op[diff_op > 0]

#display results
subset_positive_change
```

```
 Subject_1  Subject_2  Subject_3  Subject_4  Subject_5  Subject_6  Subject_7
        16         30          3         25         26         18          5
 Subject_8 Subject_10 Subject_11 Subject_12 Subject_14 Subject_15 Subject_16
        15         10         40         19         18         31         25
Subject_18 Subject_19 Subject_20
        26         22         22
```

Question 7

```r
#calculate average decrease in blood pressure that had a positive change
avg_positive_change <- mean(diff_op[diff_op > 0])

#display results
avg_positive_change
```

```
[1] 20.64706
```

Task 2: Basic Data Frame Practice

Question 1

```r
#reuse subject names in Question 2 from Task 1 for patient
patient <- paste("Subject", 1:20, sep = "_")

#create data frame
bp_df <- data.frame(
  patient = patient,
  pre_bp = pre,
  post_bp = post,
  diff_bp = diff_op,
  row.names = NULL #let the row names be label as 1,2,3,...
)

#display first 6 rows
head(bp_df)
```

```
   patient pre_bp post_bp diff_bp
1 Subject_1    130     114      16
2 Subject_2    128      98      30
3 Subject_3    116     113       3
4 Subject_4    124      99      25
5 Subject_5    133     107      26
6 Subject_6    134     116      18
```

Question 2

```r
#return rows where diff_bp is negative
subset(bp_df, diff_bp < 0)
```

```
      patient pre_bp post_bp diff_bp
9   Subject_9    114     119      -5
13 Subject_13    128     130      -2
17 Subject_17    120     124      -4
```

Question 3

```
# new column called normal blood pressure when post_bp < 120
bp_df$normal <- bp_df$post_bp < 120

#display data frame with new column
head(bp_df)
```

```
  patient pre_bp post_bp diff_bp normal
1 Subject_1    130     114      16   TRUE
2 Subject_2    128      98      30   TRUE
3 Subject_3    116     113       3   TRUE
4 Subject_4    124      99      25   TRUE
5 Subject_5    133     107      26   TRUE
6 Subject_6    134     116      18   TRUE
```

Question 4

```
#print table nicely
knitr::kable(bp_df,
             col.names = c("Patient", "Pre-Treatment", "Post-Treatment",
                           "Difference", "Normal BP (< 120)"),
             caption = "Patients' Blood Pressures")
```

Table 1: Patients' Blood Pressures

Patient	Pre-Treatment	Post-Treatment	Difference	Normal BP (< 120)
Subject_1	130	114	16	TRUE
Subject_2	128	98	30	TRUE
Subject_3	116	113	3	TRUE
Subject_4	124	99	25	TRUE
Subject_5	133	107	26	TRUE
Subject_6	134	116	18	TRUE
Subject_7	118	113	5	TRUE
Subject_8	126	111	15	TRUE
Subject_9	114	119	-5	TRUE
Subject_10	127	117	10	TRUE
Subject_11	141	101	40	TRUE

Patient	Pre-Treatment	Post-Treatment	Difference	Normal BP (< 120)
Subject_12	138	119	19	TRUE
Subject_13	128	130	-2	FALSE
Subject_14	140	122	18	FALSE
Subject_15	137	106	31	TRUE
Subject_16	131	106	25	TRUE
Subject_17	120	124	-4	FALSE
Subject_18	128	102	26	TRUE
Subject_19	139	117	22	TRUE
Subject_20	135	113	22	TRUE

Task 3: List Practice

Question 1

```r
#pre and post measurements for placebo group
pre <- c(138, 135, 147, 117, 152, 134, 114, 121, 131, 130)
post <- c(105, 136, 123, 130, 134, 143, 135, 139, 120, 124)

#subject names
patient <- paste("Subject", 1:10, sep = "_")

#compute differences
diff_op <- pre - post

#create data frame
bp_df_placebo <- data.frame(
  patient = patient,
  pre_bp = pre,
  post_bp = post,
  diff_bp = diff_op,
  row.names = NULL
)

#new column for normal blood pressure (post < 120)
bp_df_placebo$normal <- post < 120

#display first 6 rows of new data frame
head(bp_df_placebo)
```

```
    patient pre_bp post_bp diff_bp normal
1 Subject_1    138     105      33   TRUE
2 Subject_2    135     136      -1  FALSE
3 Subject_3    147     123      24  FALSE
4 Subject_4    117     130     -13  FALSE
5 Subject_5    152     134      18  FALSE
6 Subject_6    134     143      -9  FALSE
```

Question 2

```
#combine the two data frames into a named list
bp_list <- list(
  treatment = bp_df,
  placebo = bp_df_placebo
)

#structure of the new list
str(bp_list)
```

```
List of 2
 $ treatment:'data.frame':  20 obs. of  5 variables:
  ..$ patient: chr [1:20] "Subject_1" "Subject_2" "Subject_3" "Subject_4" ...
  ..$ pre_bp : num [1:20] 130 128 116 124 133 134 118 126 114 127 ...
  ..$ post_bp: num [1:20] 114 98 113 99 107 116 113 111 119 117 ...
  ..$ diff_bp: num [1:20] 16 30 3 25 26 18 5 15 -5 10 ...
  ..$ normal : logi [1:20] TRUE TRUE TRUE TRUE TRUE TRUE ...
 $ placebo  :'data.frame':  10 obs. of  5 variables:
  ..$ patient: chr [1:10] "Subject_1" "Subject_2" "Subject_3" "Subject_4" ...
  ..$ pre_bp : num [1:10] 138 135 147 117 152 134 114 121 131 130
  ..$ post_bp: num [1:10] 105 136 123 130 134 143 135 139 120 124
  ..$ diff_bp: num [1:10] 33 -1 24 -13 18 -9 -21 -18 11 6
  ..$ normal : logi [1:10] TRUE FALSE FALSE FALSE FALSE FALSE ...
```

Question 3

```
#access the first element with 3 different types syntax

#type 1: Dollar sign syntax
bp_list$treatment
```

```
      patient pre_bp post_bp diff_bp normal
1   Subject_1    130     114      16   TRUE
2   Subject_2    128      98      30   TRUE
3   Subject_3    116     113       3   TRUE
4   Subject_4    124      99      25   TRUE
5   Subject_5    133     107      26   TRUE
6   Subject_6    134     116      18   TRUE
```

```
7    Subject_7    118    113     5    TRUE
8    Subject_8    126    111    15    TRUE
9    Subject_9    114    119    -5    TRUE
10 Subject_10    127    117    10    TRUE
11 Subject_11    141    101    40    TRUE
12 Subject_12    138    119    19    TRUE
13 Subject_13    128    130    -2    FALSE
14 Subject_14    140    122    18    FALSE
15 Subject_15    137    106    31    TRUE
16 Subject_16    131    106    25    TRUE
17 Subject_17    120    124    -4    FALSE
18 Subject_18    128    102    26    TRUE
19 Subject_19    139    117    22    TRUE
20 Subject_20    135    113    22    TRUE
```

```r
#type 2: Double brackets with variable name syntax
bp_list[["treatment"]]
```

```
     patient pre_bp post_bp diff_bp normal
1    Subject_1    130    114    16    TRUE
2    Subject_2    128     98    30    TRUE
3    Subject_3    116    113     3    TRUE
4    Subject_4    124     99    25    TRUE
5    Subject_5    133    107    26    TRUE
6    Subject_6    134    116    18    TRUE
7    Subject_7    118    113     5    TRUE
8    Subject_8    126    111    15    TRUE
9    Subject_9    114    119    -5    TRUE
10 Subject_10    127    117    10    TRUE
11 Subject_11    141    101    40    TRUE
12 Subject_12    138    119    19    TRUE
13 Subject_13    128    130    -2    FALSE
14 Subject_14    140    122    18    FALSE
15 Subject_15    137    106    31    TRUE
16 Subject_16    131    106    25    TRUE
17 Subject_17    120    124    -4    FALSE
18 Subject_18    128    102    26    TRUE
19 Subject_19    139    117    22    TRUE
20 Subject_20    135    113    22    TRUE
```

```
#type 3: Double brackets with column numeric index syntax
bp_list[[1]]
```

```
   patient pre_bp post_bp diff_bp normal
1   Subject_1    130     114      16   TRUE
2   Subject_2    128      98      30   TRUE
3   Subject_3    116     113       3   TRUE
4   Subject_4    124      99      25   TRUE
5   Subject_5    133     107      26   TRUE
6   Subject_6    134     116      18   TRUE
7   Subject_7    118     113       5   TRUE
8   Subject_8    126     111      15   TRUE
9   Subject_9    114     119      -5   TRUE
10 Subject_10    127     117      10   TRUE
11 Subject_11    141     101      40   TRUE
12 Subject_12    138     119      19   TRUE
13 Subject_13    128     130      -2  FALSE
14 Subject_14    140     122      18  FALSE
15 Subject_15    137     106      31   TRUE
16 Subject_16    131     106      25   TRUE
17 Subject_17    120     124      -4  FALSE
18 Subject_18    128     102      26   TRUE
19 Subject_19    139     117      22   TRUE
20 Subject_20    135     113      22   TRUE
```

Question 4

```
#access the pre_bp column from the placebo data frame
bp_list$placebo$pre_bp
```

```
 [1] 138 135 147 117 152 134 114 121 131 130
```

Task 4: Control Flow Practice

Question 1

```
#create status column for both data frames
bp_list$treatment$status <- character(nrow(bp_list$treatment))
bp_list$placebo$status <- character(nrow(bp_list$placebo))

#structure with the new column in each data frame
str(bp_list$treatment)
```

```
'data.frame':    20 obs. of  6 variables:
 $ patient: chr  "Subject_1" "Subject_2" "Subject_3" "Subject_4" ...
 $ pre_bp : num  130 128 116 124 133 134 118 126 114 127 ...
 $ post_bp: num  114 98 113 99 107 116 113 111 119 117 ...
 $ diff_bp: num  16 30 3 25 26 18 5 15 -5 10 ...
 $ normal : logi  TRUE TRUE TRUE TRUE TRUE TRUE ...
 $ status : chr  "" "" "" "" ...
```

```
str(bp_list$placebo)
```

```
'data.frame':    10 obs. of  6 variables:
 $ patient: chr  "Subject_1" "Subject_2" "Subject_3" "Subject_4" ...
 $ pre_bp : num  138 135 147 117 152 134 114 121 131 130
 $ post_bp: num  105 136 123 130 134 143 135 139 120 124
 $ diff_bp: num  33 -1 24 -13 18 -9 -21 -18 11 6
 $ normal : logi  TRUE FALSE FALSE FALSE FALSE FALSE ...
 $ status : chr  "" "" "" "" ...
```

Question 2

```
#for non-placebo data frame
#use for loop with if/then/else to assign status values
for (i in 1:nrow(bp_list$treatment)) {
  post <- bp_list$treatment$post_bp[i]

  if (post <= 120) {
    bp_list$treatment$status[i] <- "optimal"
```

```r
  } else if (post <= 130) {
    bp_list$treatment$status[i] <- "borderline"
  } else {
    bp_list$treatment$status[i] <- "high"
  }
}

#display first 6 rows of updated treatment data frame
head(bp_list$treatment)
```

```
  patient pre_bp post_bp diff_bp normal  status
1 Subject_1    130     114      16   TRUE optimal
2 Subject_2    128      98      30   TRUE optimal
3 Subject_3    116     113       3   TRUE optimal
4 Subject_4    124      99      25   TRUE optimal
5 Subject_5    133     107      26   TRUE optimal
6 Subject_6    134     116      18   TRUE optimal
```

Question 3

```r
#repeat the same steps from question 2 for placebo data frame
for (i in 1:nrow(bp_list$placebo)) {
  post <- bp_list$placebo$post_bp[i]

  if (post <= 120) {
    bp_list$placebo$status[i] <- "optimal"
  } else if (post <= 130) {
    bp_list$placebo$status[i] <- "borderline"
  } else {
    bp_list$placebo$status[i] <- "high"
  }
}

#display first 6 rows of updated placebo data frame
bp_list$placebo
```

```
  patient pre_bp post_bp diff_bp normal    status
1 Subject_1    138     105      33   TRUE   optimal
2 Subject_2    135     136      -1  FALSE      high
```

```
3   Subject_3    147    123      24  FALSE borderline
4   Subject_4    117    130     -13  FALSE borderline
5   Subject_5    152    134      18  FALSE       high
6   Subject_6    134    143      -9  FALSE       high
7   Subject_7    114    135     -21  FALSE       high
8   Subject_8    121    139     -18  FALSE       high
9   Subject_9    131    120      11  FALSE    optimal
10 Subject_10    130    124       6  FALSE borderline
```

Task 5: Function Writing

```r
#create a function to retrieve a statistic from a list with 2 data frames
calculate_stat <- function(df_list, stat = "mean") {
  #access the actual function from the string
  my_fun <- get(stat)

  #get each column of interest
  #treatment data frame
  t_pre  <- my_fun(df_list$treatment$pre_bp)
  t_post <- my_fun(df_list$treatment$post_bp)
  t_diff <- my_fun(df_list$treatment$diff_bp)

  #placebo data frame
  p_pre  <- my_fun(df_list$placebo$pre_bp)
  p_post <- my_fun(df_list$placebo$post_bp)
  p_diff <- my_fun(df_list$placebo$diff_bp)

  #create a vector of names
  stat_names <- c(
    treatment_pre  = t_pre,
    treatment_post = t_post,
    treatment_diff = t_diff,
    placebo_pre    = p_pre,
    placebo_post   = p_post,
    placebo_diff   = p_diff
  )
  return(stat_names)
}

#test it
#use it without specifying your statistic
#mean
calculate_stat(bp_list)
```

```
 treatment_pre treatment_post treatment_diff    placebo_pre   placebo_post
        129.35         112.35          17.00         131.90         128.90
  placebo_diff
          3.00
```

14

```
#use it with specifying your statistic as "var", "sd", "min", and "max".
#variance
calculate_stat(bp_list, "var")
```

treatment_pre	treatment_post	treatment_diff	placebo_pre	placebo_post
64.55526	74.76579	153.68421	149.87778	124.98889

placebo_diff
341.33333

```
#standard deviation
calculate_stat(bp_list, "sd")
```

treatment_pre	treatment_post	treatment_diff	placebo_pre	placebo_post
8.034629	8.646721	12.396944	12.242458	11.179843

placebo_diff
18.475209

```
#minimum
calculate_stat(bp_list, "min")
```

treatment_pre	treatment_post	treatment_diff	placebo_pre	placebo_post
114	98	-5	114	105

placebo_diff
-21

```
#maximum
calculate_stat(bp_list, "max")
```

treatment_pre	treatment_post	treatment_diff	placebo_pre	placebo_post
141	130	40	152	143

placebo_diff
33