BaseR Programming and Data Manipulation

Makenna Meyer

2025-05-27

The following was completed to complete Homework 2 for ST558 and practice basic programming and data manipulation in R.

Task One: Basic Vector Practice

Question 1

```
#Creating two vectors called pre and post using provided data

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

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

Question 2

```
#Assigning patient IDs to the vectors
patient <- paste("Subject", 1:20, sep = "_")
names(pre) <- patient
names(post) <- patient
```

Question 3

```r
#Creating a vector of the difference in blood pressure pre vs post treatment
diff_bp <- pre-post
diff_bp
```

```
 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

```r
#Finding the mean difference in blood pressure
mean(diff_bp)
```

```
[1] 17
```

Question 5

```r
#Identifying patients that experienced a decrease in blood pressure after treatment
which(diff_bp > 0)
```

```
 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
#Subseting the difference vector to only return patients with a positive change
diff_bp[which(diff_bp > 0)]
```

```
 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

```
#Calculating the mean blood pressure difference of patients with a positive change
mean(diff_bp[which(diff_bp > 0)])
```

```
[1] 20.64706
```

Task Two

Question 1

```
#Combining vectors into a data frame
bp_df_trt <- data.frame(patient, pre, post, diff_bp)
colnames(bp_df_trt) <- c("patient", "pre_bp", "post_bp", "diff_bp")
```

Question 2

```
#Subseting the data frame to only include patients with a positive change
bp_df_trt[(bp_df_trt$diff_bp < 0),]
```

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

Question 3

```
#Adding a data column identifying if the patient's post bp is normal (less than 120)
bp_df_trt$normal <- bp_df_trt$post_bp < 120
```

Question 4

```
#Displaying the final treatment blood pressure data frame
knitr::kable(bp_df_trt)
```

```
Warning: 'xfun::attr()' is deprecated.
Use 'xfun::attr2()' instead.
See help("Deprecated")
Warning: 'xfun::attr()' is deprecated.
Use 'xfun::attr2()' instead.
See help("Deprecated")
```

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

Task Three

Question 1

```
#Creating a data frame for patients who received a placebo
#Only permanently saving the final dataset
bp_df_placebo <- data.frame(names_plac = paste("Subject", 1:10, sep = "_"),
                            pre_bp = c(138, 135, 147, 117, 152, 134, 114, 121, 131, 130),
                            post_bp = c(105, 136, 123, 130, 134, 143, 135, 139, 120, 124))
bp_df_placebo$diff_bp <- bp_df_placebo$pre_bp - bp_df_placebo$post_bp
bp_df_placebo$normal <- bp_df_placebo$post_bp < 120

bp_df_placebo
```

```
   names_plac 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
7   Subject_7    114     135     -21  FALSE
8   Subject_8    121     139     -18  FALSE
9   Subject_9    131     120      11  FALSE
10 Subject_10    130     124       6  FALSE
```

Question 2

```
#Creating a list with 2 elements, the treatment dataset and the placebo dataset
bp_list <- list(treatment=bp_df_trt, placebo=bp_df_placebo)
```

Question 3

```
#Accessing the treatment dataset from the list in 3 different ways
bp_list[1]
```

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

```
bp_list[[1]]
```

```
              patient pre_bp post_bp diff_bp normal
Subject_1   Subject_1    130     114      16   TRUE
Subject_2   Subject_2    128      98      30   TRUE
Subject_3   Subject_3    116     113       3   TRUE
Subject_4   Subject_4    124      99      25   TRUE
Subject_5   Subject_5    133     107      26   TRUE
Subject_6   Subject_6    134     116      18   TRUE
Subject_7   Subject_7    118     113       5   TRUE
Subject_8   Subject_8    126     111      15   TRUE
Subject_9   Subject_9    114     119      -5   TRUE
Subject_10 Subject_10    127     117      10   TRUE
Subject_11 Subject_11    141     101      40   TRUE
Subject_12 Subject_12    138     119      19   TRUE
Subject_13 Subject_13    128     130      -2  FALSE
Subject_14 Subject_14    140     122      18  FALSE
Subject_15 Subject_15    137     106      31   TRUE
Subject_16 Subject_16    131     106      25   TRUE
```

```
Subject_17 Subject_17    120     124      -4  FALSE
Subject_18 Subject_18    128     102      26   TRUE
Subject_19 Subject_19    139     117      22   TRUE
Subject_20 Subject_20    135     113      22   TRUE
```

```
bp_list$treatment
```

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

Question 4

```
#Displaying thepre_bp column from the placebo dataset
bp_list[[2]][2]
```

```
   pre_bp
1     138
2     135
3     147
4     117
5     152
```

6	134
7	114
8	121
9	131
10	130

Task Four

Question 1

```
#Initializing status columns in both treatment and placebo datasets
bp_list$treatment$status <- character(20)
bp_list$placebo$status <- character(10)
```

Question 2

```
#Using a for loop to iterate through the 20 treatment patients
#and classify the post-treatment bp as high, borderline, or optimal
for(i in 1:20){
  if(bp_list$treatment$post_bp[i] > 130) {
    bp_list$treatment$status[i] <- "high"
    } else if(bp_list$treatment$post_bp[i] > 120){
      bp_list$treatment$status[i] <- "borderline"
      } else {
        bp_list$treatment$status[i] <- "optimal"
      }
}

bp_list$treatment
```

```
              patient pre_bp post_bp diff_bp normal     status
Subject_1   Subject_1    130     114      16   TRUE    optimal
Subject_2   Subject_2    128      98      30   TRUE    optimal
Subject_3   Subject_3    116     113       3   TRUE    optimal
Subject_4   Subject_4    124      99      25   TRUE    optimal
Subject_5   Subject_5    133     107      26   TRUE    optimal
Subject_6   Subject_6    134     116      18   TRUE    optimal
Subject_7   Subject_7    118     113       5   TRUE    optimal
Subject_8   Subject_8    126     111      15   TRUE    optimal
Subject_9   Subject_9    114     119      -5   TRUE    optimal
Subject_10 Subject_10    127     117      10   TRUE    optimal
Subject_11 Subject_11    141     101      40   TRUE    optimal
Subject_12 Subject_12    138     119      19   TRUE    optimal
Subject_13 Subject_13    128     130      -2  FALSE borderline
Subject_14 Subject_14    140     122      18  FALSE borderline
```

```
Subject_15 Subject_15    137     106      31   TRUE    optimal
Subject_16 Subject_16    131     106      25   TRUE    optimal
Subject_17 Subject_17    120     124      -4  FALSE borderline
Subject_18 Subject_18    128     102      26   TRUE    optimal
Subject_19 Subject_19    139     117      22   TRUE    optimal
Subject_20 Subject_20    135     113      22   TRUE    optimal
```

Question 3

```
#Using a for loop to iterate through the 10 placebo patients
#and classify the post-placebo bp as high, borderline, or optimal
for(i in 1:10){
  if(bp_list$placebo$post_bp[i] > 130) {
    bp_list$placebo$status[i] <- "high"
    } else if(bp_list$placebo$post_bp[i] > 120){
      bp_list$placebo$status[i] <- "borderline"
      } else {
        bp_list$placebo$status[i] <- "optimal"
      }
  }

bp_list$placebo
```

```
   names_plac 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 Five

```
#Writing a function that takes in a list of two dataframes, and calculates a
#certain statistic for the pre, post, and difference bp columns of both dataframes
task5_func <- function(list, stat = "mean") {
  my_fun <- get(stat)
  vec_values <- c(my_fun(list[[1]]$pre_bp),
                  my_fun(list[[1]]$post_bp),
                  my_fun(list[[1]]$diff_bp),
                  my_fun(list[[2]]$pre_bp),
                  my_fun(list[[2]]$post_bp),
                  my_fun(list[[2]]$diff_bp)
                  )
  vec_names <- c(paste(attributes(list)$names[[1]], "pre_bp", stat, sep = " "),
                   paste(attributes(list)$names[[1]], "post_bp", stat, sep = " "),
                   paste(attributes(list)$names[[1]], "diff_bp", stat, sep = " "),
                   paste(attributes(list)$names[[2]], "pre_bp", stat, sep = " "),
                   paste(attributes(list)$names[[2]], "post_bp", stat, sep = " "),
                   paste(attributes(list)$names[[2]], "diff_bp", stat, sep = " ")
                   )
  names(vec_values) <- vec_names
  return(list(vec_values))
}
```

```
#Using the function to calculate mean, var, sd, min, and max
#from the bp treatment and placebo datasets
task5_func(bp_list)
```

```
[[1]]
 treatment pre_bp mean treatment post_bp mean treatment diff_bp mean
                129.35                 112.35                  17.00
   placebo pre_bp mean   placebo post_bp mean   placebo diff_bp mean
                131.90                 128.90                   3.00
```

```
task5_func(bp_list, "var")
```

```
[[1]]
 treatment pre_bp var treatment post_bp var treatment diff_bp var
             64.55526              74.76579             153.68421
   placebo pre_bp var   placebo post_bp var   placebo diff_bp var
            149.87778             124.98889             341.33333
```

```r
task5_func(bp_list, "sd")
```

```
[[1]]
 treatment pre_bp sd treatment post_bp sd treatment diff_bp sd 
            8.034629             8.646721            12.396944 
   placebo pre_bp sd   placebo post_bp sd   placebo diff_bp sd 
           12.242458            11.179843            18.475209 
```

```r
task5_func(bp_list, "min")
```

```
[[1]]
 treatment pre_bp min treatment post_bp min treatment diff_bp min 
                  114                    98                    -5 
   placebo pre_bp min   placebo post_bp min   placebo diff_bp min 
                  114                   105                   -21 
```

```r
task5_func(bp_list, "max")
```

```
[[1]]
 treatment pre_bp max treatment post_bp max treatment diff_bp max 
                  141                   130                    40 
   placebo pre_bp max   placebo post_bp max   placebo diff_bp max 
                  152                   143                    33 
```