

Assignment 3

BIOSTAT 213: Introduction to Computing in the R Software Environment

Due: Wednesday, November 6, 2019

Collaboration with other students in this course is permitted on this assignment. However, you must write all work—including R code—in your own voice, as much as possible, and acknowledge your collaborators as thoroughly as possible.

The assignment is due online, in the [Collaborative Learning Environment](#), by 11:59 pm, Wednesday, November 6, 2019. You must submit a Word-compatible document or a PDF file.

For each of the below, share your code as well as the output (if there is any output). If your answer to a question is unclear from your R output, or difficult to identify, include an answer in text form (in addition to the code and output).

You will be using the `nhanes.csv` data file for the following problems.

1. List the age (`ridageyr`), gender (`gender`), body mass index (`bmxbmi`), weight (`bmxwt`), and height (`bmxht`) variables for all individuals with a body mass index less or equal to 16 kg/m². Do this with indexing and with `subset()`.

```
> ii[ii$bmxbmi<=16,c('ridageyr','gender','bmxbmi','bmxwt','bmxht')]
  ridageyr gender bmxbmi bmxwt bmxht
1614      58 Female  14.5  39.6 165.0
2831      67 Female  15.5  44.6 169.4
3602      47  Male   16.0  52.5 181.0
> subset(ii,bmxbmi<=16,select=c(ridageyr,gender,bmxbmi,bmxwt,bmxht))
  ridageyr gender bmxbmi bmxwt bmxht
1614      58 Female  14.5  39.6 165.0
2831      67 Female  15.5  44.6 169.4
3602      47  Male   16.0  52.5 181.0
```

2. List the age (`ridageyr`), gender (`gender`), systolic blood pressure (`bpxsy1`), and blood pressure category (`bpcat`) for all individuals with a systolic blood pressure less or equal to 85 mmHg. Do this with indexing and with `subset()`.

```
> ii[!is.na(ii$bpxsy1)&ii$bpxsy1<=85,c('ridageyr','gender','bpxsy1','bpcat')]
  ridageyr gender bpxsy1 bpcat
280       28 Female    84 Normal
913       62  Male     84 Normal
2068      21 Female    82 Normal
2298      26  Male     84 Normal
3176      59 Female    84 Normal
3196      29 Female    84 Normal
> subset(ii,bpxsy1<=85,select=c(ridageyr,gender,bpxsy1,bpcat))
  ridageyr gender bpxsy1 bpcat
280       28 Female    84 Normal
913       62  Male     84 Normal
2068      21 Female    82 Normal
2298      26  Male     84 Normal
3176      59 Female    84 Normal
3196      29 Female    84 Normal
```

3. List the 1000th, 2000th, 3000th, and 4000th rows of the 100th, 200th, 300th, 400th, 500th, and 600th columns.

```
> ii[seq(1000,4000,1000),seq(100,600,100)]
      bpxdi4 dr1tm221 dr2tmfat mcq160b paq724i smq852q
1000      NA      0.003      11.044      2      NA      3
2000      NA      0.000      10.353      2      NA      NA
3000      98      0.023      9.482      2      NA      NA
4000      NA      0.015      27.072      2      NA      NA
```

4. List the 100th, 200th, 300th, 400th, 500th, and 600th rows for all columns that have a column name beginning with the letter l.

```
> ii[seq(100,600,100),substr(names(ii),0,1)=='l']
      lbdhdd lbdhddsi lbxtr lbdtrsi lbdldl lbdldlsi
100      52      1.34      NA      NA      NA      NA
200      35      0.91      146      1.648      118      3.051
300      43      1.11      NA      NA      NA      NA
400      NA      NA      NA      NA      NA      NA
500      64      1.66      92      1.039      63      1.629
600      40      1.03      NA      NA      NA      NA
```

5. Restricting analysis to individuals 30 years of age (`ridageyr`) or younger, conduct a two-sample *t*-test, using body mass index as the *y* variable (`bmxbmi`) and history of military service (`military`) as the *x* variable. Use object assignment to save the test. Without copying and pasting numeric values, use the test object to find the estimated military-minus-no mean difference in body mass index. Use `unnname()` or `as.numeric()` to remove any names in your final answer. Hint: you may `t.test()`'s `data` argument to restrict the analysis.

```
> tt<-t.test(ridageyr~military,data=subset(ii,ridageyr<=30))
> as.numeric(tt$estimate['mean in group History of military service']-
+           tt$estimate['mean in group No'])
[1] 2.202576
```

6. Use `with()` to cross tabulate history of gender (`gender`) and smoking (`hs`), among individuals 65 years of age (`ridageyr`) or older. Place gender along the rows and smoking along the columns. Hint: You may use `with()`'s `data` argument to restrict the analysis.

```
> with(data=subset(ii,ridageyr>=65),table(gender,hs))
      hs
gender History of smoking No
Female      85 111
Male      114 57
```

7. Create a new factor that indicates having an LDL cholesterol less than 100 mg/dL, yes or no. Do this using (a) addition of logical vectors, (b) `cut()`, (c) `ifelse()`, and (d) indexing and assignment. Use `table()` to check each variable, and verify that all of your solutions are equivalent.

```
> # (a)
> ii$ldl2<-factor((ii$lbdldl>=0)+(ii$lbdldl>=100),1:2)
> levels(ii$ldl2)<-c('Low','No')
> table(ii$ldl2,useNA='always')

Low  No <NA>
673 1101 2435

> # (b)
> ii$ldl2<-cut(ii$lbdldl,breaks=c(0,100,500),right=FALSE)
> levels(ii$ldl2)<-c('Low','No')
> table(ii$ldl2,useNA='always')
```

```

Low    No <NA>
673 1101 2435
> # (c)
> ii$ldl2<-as.factor(ifelse(ii$lbdld1<100,'Low','No'))
> table(ii$ldl2,useNA='always')

Low    No <NA>
673 1101 2435
> # (d)
> ii$ldl2<- 'No'
> ii$ldl2[ii$lbdld1<100]<- 'Low'
> ii$ldl2<-as.factor(ii$ldl2)
> table(ii$ldl2,useNA='always')

Low    No <NA>
673 3536    0

```

8. Using ggplot2, plot systolic blood pressure (**bpxsy1**) and diastolic blood pressure (**bpxdi1**), with systolic blood pressure on the *y* axis. Use color coding for categories of blood pressure (**bpcat**). Make sure your code includes the following elements:

- Reorder the levels of **bpcat**, using the following order: Hypertension, Elevated, Normal.
- If you have not already, recode values of diastolic blood pressure that are 10 mmHg of lower as missing.
- Using the **data** argument to **aes()**, restrict the plot to individuals not missing **bpcat**. The purpose of this is to prevent a missing value category from showing up in the legend.
- Add **labs()** with the **col** argument to change the label for the legend to **Blood pressure**.

Your plot should look like this:

```

> is.na(ii$bpxdi1)<-ii$bpxdi1<=10
> ii$bpcat<-factor(ii$bpcat,c('Hypertension','Elevated','Normal'))
> ggplot(aes(x=bpxdi1,y=bpxsy1,col=bpcat),data=subset(ii,!is.na(bpcat)))+
+   geom_point()+
+   xlab('Diastolic blood pressure (mmHg)')+
+   ylab('Systolic blood pressure (mmHg)')+
+   labs(col='Blood pressure')

```

Warning: Removed 11 rows containing missing values (geom_point).

