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/m^2. Do this with indexing and with `subset()`.

```
> # your code here
```

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()`.

```
> # your code here
```

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

```
> # your code here
```

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

```
> # your code here
```

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 `unname()` or `as.numeric()` to remove any names in your final answer. Hint: you may `t.test()`'s `data` argument to restrict the analysis.

```
> # your code here
```

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.

```
> # your code here
```

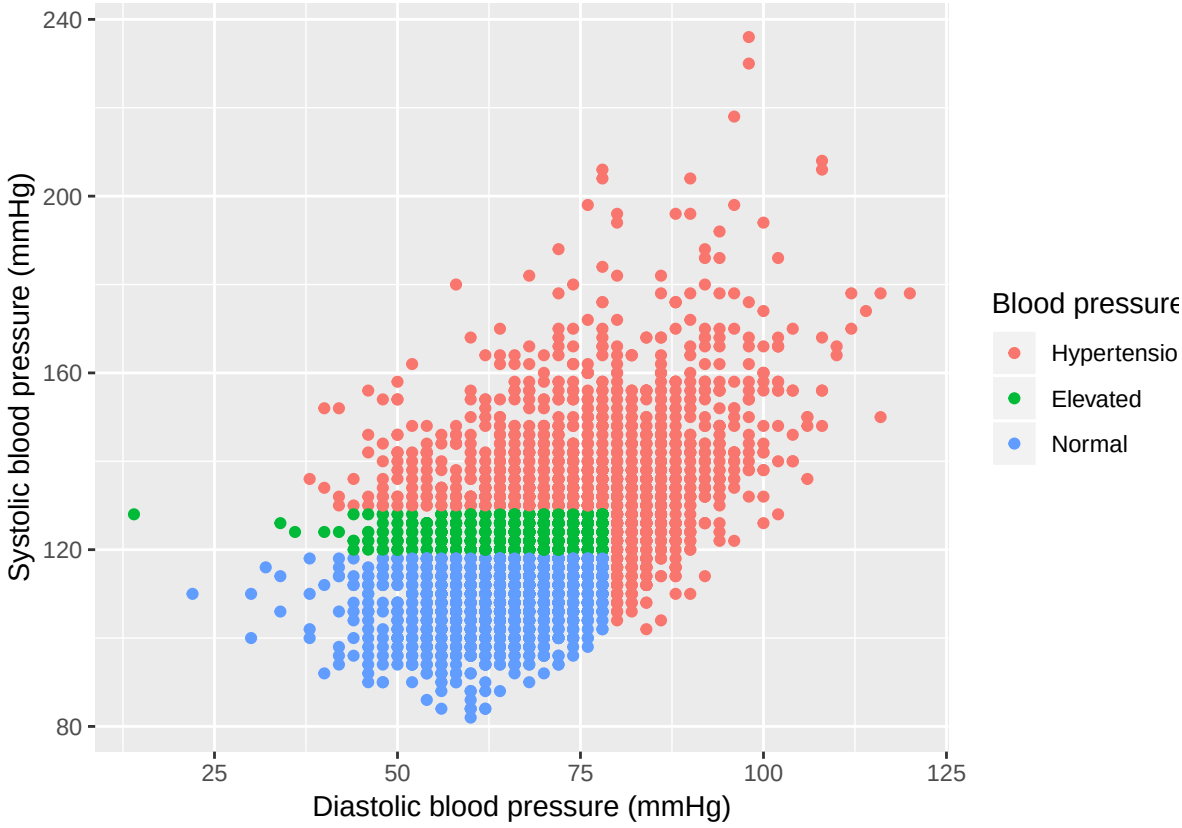
7. Create a new factor that indicates having an LDL cholesterol less than 100 mg/dL, yes or no. Do this using (a) addition of logical vectors, (b) `cut()`, (c) `ifelse()`, and (d) indexing and assignment. Use `table()` to check each variable, and verify that all of your solutions are equivalent.

```
> # your code here
```

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:




```
> # your code here
```