Homework 8
MTH 3270 Data Science
Due Mon., Apr. 20

Read These Chapters of the Book	Then Do These Exercises
Appendix E 8	E.2* 8.1 (just do *decision tree*, *random forest*, and *k-NN*), 8.3 (just *decision tree*, *random forest*, and *k-NN*)

* For **Problem E.2**:

- Look at the help page for the `HELPrct` data:

```
library(mdsr)            # Contains the HELPrct data set
? HELPrct
```

- The response variable, `homeless`, is **dichotomous** (takes only two values), so a (multiple-explanatory variable) ***logistic regression*** analysis is appropriate.

- **Don't** use any of the *categorical* variables (*factors*) as explanatory variables in the model.

  To find out which ones are are *numeric* (or *integer*) and which are categorical (*factors*), type:

```
str(HELPrct)
```

  To select just the numerical variables from `HELPrct` (and the `homeless` variable) and store them as, say, `my.data`, type:

```
library(dplyr)          # Contains the select_if() function

my.data <- HELPrct %>% select_if(is.numeric) %>%
  mutate(homeless = HELPrct$homeless) %>%
  select(homeless, everything())
```

An example of fitting a (two-explanatory variable) **logistic regression model** is:

```r
my.log.reg <- glm(homeless ~ age + indtot,
                  data = my.data,
                  family = "binomial")

summary(my.log.reg)
```