Homework 2 solutions

1. (1 points) Which of the following names are invalid for objects in R? Why?
   a. `.object`: **this is a valid name.**
   b. `.2object`: **names start with a dot cannot followed by a number.**
   c. `2object`: **object names cannot start with a number.**
   d. `repeat`: **object names cannot use reserved words.**
   e. `_2object`: **object names cannot start with an underscore.**
2. (1 point) Read a file about light speed measurements (https://raw.githubusercontent.com/vincentarelbundock/Rdatasets/master/csv/gamclass/cvalues.csv) into R as a data frame using `read.table()` or `read.csv()`. Assign it a valid and meaningful name. Note: the first column is row names.

```
light_speed <- read.table("https://raw.githubusercontent.com/vincentarelbundock/Rdatasets/master/csv/ga
                header = T, sep = ",", row.names = 1)
```

3. (1 point) How many rows does it have? How many columns? What are the types of columns?

```
nrow(light_speed) # number of rows
```

```
## [1] 9
```

```
ncol(light_speed) # number of columns
```

```
## [1] 3
```

```
class(light_speed[, 1]) # type of first column
```

```
## [1] "integer"
```

```
class(light_speed[, 2]) # type of second column
```

```
## [1] "numeric"
```

```
class(light_speed[, 3]) # type of third column
```

```
## [1] "numeric"
```

4. (1 point) Remove the rows that have missing data (i.e. `NA`).

```
# by looking at the data, it is the error column that has NAs
light_speed_no_NA <- light_speed[!is.na(light_speed$error), ]
# or use the na.omit()
# light_speed_no_NA <- na.omit(light_speed)
```

5. (1 point) Save the data from question 4 as an external file with an meaningful name.

```
write.csv(light_speed_no_NA, file = "light_speed_without_NAs.csv")
```

6. (1 point) Read a file about Hawaian island chain hotspot Potassium-Argon ages (https://raw.githubusercontent.com/vincentarelbundock/Rdatasets/master/csv/DAAG/hotspots.csv) into R as a data frame using `read.table()` or `read.csv()`. Assign it a valid and meaningful name. Note: the first column is row names.

```
hotspot_ages <- read.table("https://raw.githubusercontent.com/vincentarelbundock/Rdatasets/master/csv/D
           header = TRUE, sep = ",", row.names = 1)
```

7. (1 point) How many rows does it have? How many columns? What are the types of columns?

```r
nrow(hotspot_ages) # number of rows
```

```
## [1] 35
```

```r
ncol(hotspot_ages) # number of columns
```

```
## [1] 6
```

```r
class(hotspot_ages[, 1]) # type of first column
```

```
## [1] "factor"
```

```r
class(hotspot_ages[, 2]) # type of second column
```

```
## [1] "factor"
```

```r
class(hotspot_ages[, 3]) # type of third column
```

```
## [1] "integer"
```

```r
class(hotspot_ages[, 4]) # type of fourth column
```

```
## [1] "numeric"
```

```r
class(hotspot_ages[, 5]) # type of fifth column
```

```
## [1] "numeric"
```

```r
class(hotspot_ages[, 6]) # type of sixth column
```

```
## [1] "factor"
```

8. (1 point) What is the minimal value of the `distance` variable? What is the average of the `error` variable?

```r
min(hotspot_ages$distance) # minimal
```

```
## [1] 54
```

```r
mean(hotspot_ages$error, na.rm = TRUE) # use na.rm = T when it has NAs.
```

```
## [1] 0.6485714
```

9. (1 point) Create a data frame that only has `name` and `source` variables.

```r
hotspot_ages_source <- hotspot_ages[, c("name", "source")]
```

10. (1 point) Save the data frame from question 9 as an external file with a meaningful name. The save file should NOT have row names.

```r
write.table(hotspot_ages_source, file = "hotspot_ages_source.txt",
            sep = ";", row.names = FALSE)
```