SIOB 296 Introduction to Programming with R

Eric Archer (eric.archer@noaa.gov)

Week 06: February 10, 2020

Answer all questions in a script (.R) file. Use comments (`#` or `#'`).

1. Write a function that returns the mean, median, and standard deviation of a vector of numbers.

```
x <- sample(1:1000, 100, replace = TRUE)
vecSmry(x)
```

```
    mean   median       sd
534.3800 529.0000 289.6076
```

2. Use the function from #1 to write a second function that returns a summary of the temperature for a CTD station. Assume the data is being input in the same form as the data in `ctd.csv`.

```
ctd <- read.csv("ctd.csv", stringsAsFactors = FALSE)
stationTempSmry("Station.1", ctd)
```

```
     mean    median        sd
13.567723 13.070000  2.585194
```

```
stationTempSmry("Station.5", ctd)
```

```
     mean    median        sd
14.650087 14.530000  2.265715
```

3. Change the function in #2 to include the amount of missing temperature data at the station.

```
stationTempSmry2("Station.6", ctd)
```

```
     mean    median        sd   num.NAs
14.231704 14.115000  2.438763  0.000000
```

```
stationTempSmry2("Station.3", ctd)
```

```
     mean    median        sd   num.NAs
14.238296 14.135000  2.424914  0.000000
```

4. Change the function in #3 to produce an error if the station doesn't exist.

```
stationTempSmry3("Station.5", ctd)
```

```
     mean    median        sd   num.NAs
14.650087 14.530000  2.265715  0.000000
```

```
stationTempSmry3("Station.50", ctd)
```

```
Error in stationTempSmry3("Station.50", ctd): 'Station.50' does not exist
```

5. Create a function that returns a new data frame with separate columns for year, month, day, and one POSIXct date column.

```
new.ctd <- parseCTDdate(ctd)
str(new.ctd)
```

```
'data.frame':   77641 obs. of  12 variables:
 $ station  : chr  "Station.1" "Station.1" "Station.1" "Station.1" ...
 $ date     : POSIXct, format: "2012-11-08" "2012-04-19" ...
 $ year     : num  2012 2012 2010 2014 2011 ...
 $ month    : num  11 4 1 2 1 2 4 10 3 2 ...
 $ day      : int  8 19 6 6 5 3 19 4 3 6 ...
 $ temp     : num  16.8 10.5 15.1 14 14.2 ...
 $ salinity : num  33.4 33.8 33.4 33.4 33.3 ...
 $ dox      : num  8.07 3.16 7.22 7.31 7.91 6.45 3.32 6.14 8.82 6.98 ...
 $ ph       : num  8.2 7.73 8.13 NA 8.16 8.05 7.75 7.94 8.22 NA ...
 $ pct_light: num  90.3 88.1 89 88 86.2 ...
 $ density  : num  24.3 25.9 24.7 25 24.8 ...
 $ depth    : int  16 18 32 41 3 51 16 48 7 45 ...
```

6. Change the function in #5 to accept either a data frame or the name of the file and produce the same output.

```
new.ctd <- parseCTDdate(ctd)
str(new.ctd)
```

```
'data.frame':   77641 obs. of  12 variables:
 $ station  : chr  "Station.1" "Station.1" "Station.1" "Station.1" ...
 $ date     : POSIXct, format: "2012-11-08" "2012-04-19" ...
 $ year     : num  2012 2012 2010 2014 2011 ...
 $ month    : num  11 4 1 2 1 2 4 10 3 2 ...
 $ day      : int  8 19 6 6 5 3 19 4 3 6 ...
 $ temp     : num  16.8 10.5 15.1 14 14.2 ...
 $ salinity : num  33.4 33.8 33.4 33.4 33.3 ...
 $ dox      : num  8.07 3.16 7.22 7.31 7.91 6.45 3.32 6.14 8.82 6.98 ...
 $ ph       : num  8.2 7.73 8.13 NA 8.16 8.05 7.75 7.94 8.22 NA ...
 $ pct_light: num  90.3 88.1 89 88 86.2 ...
 $ density  : num  24.3 25.9 24.7 25 24.8 ...
 $ depth    : int  16 18 32 41 3 51 16 48 7 45 ...
```

```r
new.ctd2 <- parseCTDdate("ctd.csv")
str(new.ctd2)
```

```
'data.frame':   77641 obs. of  12 variables:
 $ station  : chr  "Station.1" "Station.1" "Station.1" "Station.1" ...
 $ date     : POSIXct, format: "2012-11-08" "2012-04-19" ...
 $ year     : num  2012 2012 2010 2014 2011 ...
 $ month    : num  11 4 1 2 1 2 4 10 3 2 ...
 $ day      : int  8 19 6 6 5 3 19 4 3 6 ...
 $ temp     : num  16.8 10.5 15.1 14 14.2 ...
 $ salinity : num  33.4 33.8 33.4 33.4 33.3 ...
 $ dox      : num  8.07 3.16 7.22 7.31 7.91 6.45 3.32 6.14 8.82 6.98 ...
 $ ph       : num  8.2 7.73 8.13 NA 8.16 8.05 7.75 7.94 8.22 NA ...
 $ pct_light: num  90.3 88.1 89 88 86.2 ...
 $ density  : num  24.3 25.9 24.7 25 24.8 ...
 $ depth    : int  16 18 32 41 3 51 16 48 7 45 ...
```