SIOB 296 Introduction to Programming with R

Eric Archer (eric.archer@noaa.gov)

Week 5: February 5, 2019

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

1. Load the `ctd.csv` data. Add three columns to the data frame for year, month, and day extracted from the `sample_date` column.

```
'data.frame':   77641 obs. of  12 variables:
 $ station    : chr  "Station.1" "Station.1" "Station.1" "Station.1" ...
 $ sample_date: chr  "2012-11-08" "2012-04-19" "2010-01-06" "2014-02-06" ...
 $ 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 ...
 $ year       : num  2012 2012 2010 2014 2011 ...
 $ month      : num  11 4 1 2 1 2 4 10 3 2 ...
 $ day        : num  8 19 6 6 5 3 19 4 3 6 ...
```

2. Create a column representing the month name

```
'data.frame':   77641 obs. of  13 variables:
 $ station    : chr  "Station.1" "Station.1" "Station.1" "Station.1" ...
 $ sample_date: chr  "2012-11-08" "2012-04-19" "2010-01-06" "2014-02-06" ...
 $ 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 ...
 $ year       : num  2012 2012 2010 2014 2011 ...
 $ month      : num  11 4 1 2 1 2 4 10 3 2 ...
 $ day        : num  8 19 6 6 5 3 19 4 3 6 ...
 $ month.name : chr  "November" "April" "January" "February" ...
```

3. Create a frequency table of the number of casts by year and month

```
  1  2  3  4  5  6  7  8  9 10 11 12
```

```
2010 52 52 52 15 52 52 52 52 52 52 52 52
2011 52 52 52 51 51 51 52 52 52 52 52 49
2012 52 52 52 52 52 51 51 52 51 52 49 52
2013 50 52 52 52 52 52 52 51 52 52 52 51
2014 18 49 17 18 52 18 18 52 15 21 52 18
2015 35 68 35 35 68 27 35 64 35 35 68 35
2016 25 68 27 35 68 35 35 68 35 35 68 34
```

4. Create a frequency table of the number of casts by unique combinations of year and month

```
# first 6 values...
head(year.month.freq)
```

```
year.month
 2010_1 2010_10 2010_11 2010_12  2010_2  2010_3
   1154    1163    1165    1153    1136    1157
```

5. Create a data frame with casts from five random dates. How many times was each station sampled on each date?

```
'data.frame':   742 obs. of  13 variables:
 $ station    : chr  "Station.12" "Station.12" "Station.12" "Station.12" ...
 $ sample_date: chr  "2012-08-13" "2012-08-13" "2012-08-13" "2012-08-13" ...
 $ temp       : num  13.8 15 16.8 15.7 14.7 ...
 $ salinity   : num  33.4 33.4 33.4 33.4 33.3 ...
 $ dox        : num  8.85 9.25 9.24 9.38 9 9.12 8.88 8.74 8.71 9.17 ...
 $ ph         : num  8.5 8.45 8.4 8.43 8.46 8.54 8.47 8.48 8.35 8.57 ...
 $ pct_light  : num  82 87.3 88.4 88.5 85.3 ...
 $ density    : num  24.9 24.7 24.3 24.6 24.7 ...
 $ depth      : int  13 9 7 8 10 15 11 12 6 17 ...
 $ year       : num  2012 2012 2012 2012 2012 ...
 $ month      : num  8 8 8 8 8 8 8 8 8 8 ...
 $ day        : num  13 13 13 13 13 13 13 13 13 13 ...
 $ month.name : chr  "August" "August" "August" "August" ...
```

	2010-11-15	2012-08-13	2013-05-28	2015-06-22	2016-07-07
Station.12	0	1	0	0	0
Station.14	0	1	0	0	0
Station.15	0	1	0	0	0
Station.16	0	1	0	0	0
Station.17	0	1	0	0	0
Station.18	0	1	0	0	0
Station.19	0	1	0	1	1
Station.22	0	1	0	0	0
Station.23	0	1	0	0	0
Station.24	0	1	0	1	1
Station.25	0	1	1	1	1
Station.26	0	1	1	1	1
Station.27	0	1	0	0	0
Station.28	1	0	0	0	0

Station.29	1	0	0	0	0
Station.30	1	0	0	0	0
Station.31	1	0	0	0	0
Station.32	1	0	0	0	1
Station.33	1	0	0	0	0
Station.34	1	0	0	0	0
Station.35	1	0	0	0	0
Station.36	1	0	0	0	0
Station.37	1	0	0	0	0
Station.38	1	0	0	0	0
Station.39	0	1	1	1	1
Station.40	0	1	0	1	1

6. What is the mean number of days between sampling dates in the original ctd data frame?

```
Time difference of 4.758427 days
```

7. Sort the ctd data frame by station, date, and depth. Make sure the stations are correctly sorted numerically.

```
       station sample_date  temp salinity  dox   ph pct_light density depth
2554 Station.1  2010-01-06 15.40   33.441 7.34 8.15     88.67  24.683     1
2310 Station.1  2010-01-06 15.38   33.441 7.34 8.15     88.65  24.686     2
2911 Station.1  2010-01-06 15.38   33.440 7.34 8.15     88.54  24.687     3
2059 Station.1  2010-01-06 15.37   33.440 7.35 8.15     88.43  24.688     4
2910 Station.1  2010-01-06 15.37   33.440 7.35 8.15     88.36  24.688     5
2665 Station.1  2010-01-06 15.37   33.440 7.35 8.15     88.37  24.689     6
     year month day month.name
2554 2010     1   6    January
2310 2010     1   6    January
2911 2010     1   6    January
2059 2010     1   6    January
2910 2010     1   6    January
2665 2010     1   6    January
```

```
         station sample_date  temp salinity  dox   ph pct_light density depth
67052 Station.40  2016-12-20 14.28   33.232 7.95 8.11     71.47  24.763     5
66941 Station.40  2016-12-20 14.19   33.246 7.93 8.11     70.61  24.792     6
67091 Station.40  2016-12-20 14.17   33.280 7.83 8.11     68.64  24.823     7
67090 Station.40  2016-12-20 14.18   33.287 7.73 8.11     65.03  24.825     8
67333 Station.40  2016-12-20 14.19   33.294 7.70 8.10     62.37  24.830     9
67576 Station.40  2016-12-20 14.18   33.294 7.77 8.10     51.10  24.831    10
      year month day month.name
67052 2016    12  20   December
66941 2016    12  20   December
67091 2016    12  20   December
67090 2016    12  20   December
67333 2016    12  20   December
67576 2016    12  20   December
```