SIOB 296 Introduction to Programming with R

Eric Archer (eric.archer@noaa.gov)

Week 04: January 27, 2020

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

1. Create the following list of state information (call it `states`) using the `built-in data objectsstate.abb,state.area,state.center,state.name`:

```
str(states)
```

```
List of 4
 $ abb   : chr [1:50] "AL" "AK" "AZ" "AR" ...
 $ area  : num [1:50] 51609 589757 113909 53104 158693 ...
 $ center:List of 2
  ..$ x: num [1:50] -86.8 -127.2 -111.6 -92.3 -119.8 ...
  ..$ y: num [1:50] 32.6 49.2 34.2 34.7 36.5 ...
 $ name  : chr [1:50] "Alabama" "Alaska" "Arizona" "Arkansas" ...
```

2. Add the `state.division` factor to the `states` list.

```
List of 5
 $ abb     : chr [1:50] "AL" "AK" "AZ" "AR" ...
 $ area    : num [1:50] 51609 589757 113909 53104 158693 ...
 $ center  :List of 2
  ..$ x: num [1:50] -86.8 -127.2 -111.6 -92.3 -119.8 ...
  ..$ y: num [1:50] 32.6 49.2 34.2 34.7 36.5 ...
 $ name    : chr [1:50] "Alabama" "Alaska" "Arizona" "Arkansas" ...
 $ division: Factor w/ 9 levels "New England",..: 4 9 8 5 9 8 1 3 3 3 ...
```

3. Create a two-column data frame from the `abb` and `name` elements, and put them into an element called `abb.name`. Delete `abb` and `name`.

```
List of 4
 $ area    : num [1:50] 51609 589757 113909 53104 158693 ...
 $ center  :List of 2
  ..$ x: num [1:50] -86.8 -127.2 -111.6 -92.3 -119.8 ...
  ..$ y: num [1:50] 32.6 49.2 34.2 34.7 36.5 ...
 $ division: Factor w/ 9 levels "New England",..: 4 9 8 5 9 8 1 3 3 3 ...
 $ abb.name:'data.frame':   50 obs. of  2 variables:
  ..$ abb : chr [1:50] "AL" "AK" "AZ" "AR" ...
  ..$ name: chr [1:50] "Alabama" "Alaska" "Arizona" "Arkansas" ...
```

4. Add a list to the `states` list that summarizes the mean, median, standard deviation and range of state areas:

```
List of 5
 $ area     : num [1:50] 51609 589757 113909 53104 158693 ...
 $ center   :List of 2
  ..$ x: num [1:50] -86.8 -127.2 -111.6 -92.3 -119.8 ...
  ..$ y: num [1:50] 32.6 49.2 34.2 34.7 36.5 ...
 $ division : Factor w/ 9 levels "New England",..: 4 9 8 5 9 8 1 3 3 3 ...
 $ abb.name :'data.frame':  50 obs. of  2 variables:
  ..$ abb : chr [1:50] "AL" "AK" "AZ" "AR" ...
  ..$ name: chr [1:50] "Alabama" "Alaska" "Arizona" "Arkansas" ...
 $ area.smry:List of 4
  ..$ mean  : num 72368
  ..$ median: num 56222
  ..$ sd    : num 88278
  ..$ range : num [1:2] 1214 589757
```

5. Extract the `abb.name` data frame and add the `division` element to it. Call the new data frame "`states.df`":

```
'data.frame':   50 obs. of  3 variables:
 $ abb     : chr  "AL" "AK" "AZ" "AR" ...
 $ name    : chr  "Alabama" "Alaska" "Arizona" "Arkansas" ...
 $ division: Factor w/ 9 levels "New England",..: 4 9 8 5 9 8 1 3 3 3 ...
```

6. Use `states.df` to extract and create a character vector of the abbreviations of the New England states:

```
[1] "CT" "ME" "MA" "NH" "RI" "VT"
```

7. Using this vector and the `states` list, what is the mean area of New England states?

```
[1] 11101.33
```

8. Convert the built-in `state.x77` matrix to a data frame called `state77.df`

```
'data.frame':   50 obs. of  8 variables:
 $ Population: num  3615 365 2212 2110 21198 ...
 $ Income    : num  3624 6315 4530 3378 5114 ...
 $ Illiteracy: num  2.1 1.5 1.8 1.9 1.1 0.7 1.1 0.9 1.3 2 ...
 $ Life Exp  : num  69 69.3 70.5 70.7 71.7 ...
 $ Murder    : num  15.1 11.3 7.8 10.1 10.3 6.8 3.1 6.2 10.7 13.9 ...
 $ HS Grad   : num  41.3 66.7 58.1 39.9 62.6 63.9 56 54.6 52.6 40.6 ...
 $ Frost     : num  20 152 15 65 20 166 139 103 11 60 ...
 $ Area      : num  50708 566432 113417 51945 156361 ...
```

9. Add columns that are the per-capita income rate (Income/Population) and per-capita life expectancy rate (Life Exp/Population):

```
'data.frame':   50 obs. of  10 variables:
 $ Population : num  3615 365 2212 2110 21198 ...
 $ Income     : num  3624 6315 4530 3378 5114 ...
 $ Illiteracy : num  2.1 1.5 1.8 1.9 1.1 0.7 1.1 0.9 1.3 2 ...
 $ Life Exp   : num  69 69.3 70.5 70.7 71.7 ...
 $ Murder     : num  15.1 11.3 7.8 10.1 10.3 6.8 3.1 6.2 10.7 13.9 ...
 $ HS Grad    : num  41.3 66.7 58.1 39.9 62.6 63.9 56 54.6 52.6 40.6 ...
 $ Frost      : num  20 152 15 65 20 166 139 103 11 60 ...
 $ Area       : num  50708 566432 113417 51945 156361 ...
 $ pc.income  : num  1.002 17.301 2.048 1.601 0.241 ...
 $ pc.life.exp: num  0.0191 0.18989 0.03189 0.03349 0.00338 ...
```

10. Which states have both per-capita income and life expectancy that are greater than the mean?

```
 [1] "Alaska"        "Delaware"      "Hawaii"        "Idaho"
 [5] "Maine"         "Montana"       "Nebraska"      "Nevada"
 [9] "New Hampshire" "New Mexico"    "North Dakota"  "Rhode Island"
[13] "South Dakota"  "Utah"          "Vermont"       "Wyoming"
```

11. Use the species code table "tblCodeSpecies.csv". Create a data frame of Order, Suborder, and Family that has one row per unique family.

```
'data.frame':   26 obs. of  3 variables:
 $ ORDER   : chr  "CETACEA" "CETACEA" "CETACEA" "CETACEA" ...
 $ SUBORDER: chr  "ODONTOCETI" "ODONTOCETI" "ODONTOCETI" "ODONTOCETI" ...
 $ FAMILY  : chr  "ZIPHIIDAE" "DELPHINIDAE" "PHOCOENIDAE" "MONODONTIDAE" ...
```

```
     ORDER   SUBORDER       FAMILY
1  CETACEA ODONTOCETI    ZIPHIIDAE
2  CETACEA ODONTOCETI  DELPHINIDAE
9  CETACEA ODONTOCETI  PHOCOENIDAE
45 CETACEA ODONTOCETI MONODONTIDAE
46 CETACEA ODONTOCETI PHYSETERIDAE
47 CETACEA ODONTOCETI     KOGIIDAE
```

12. Add a column to the data frame from 1 with the number of entries in the species code table for each family.

```
'data.frame':   26 obs. of  4 variables:
 $ ORDER   : chr  "CETACEA" "CETACEA" "CETACEA" "CETACEA" ...
 $ SUBORDER: chr  "ODONTOCETI" "ODONTOCETI" "ODONTOCETI" "ODONTOCETI" ...
 $ FAMILY  : chr  "ZIPHIIDAE" "DELPHINIDAE" "PHOCOENIDAE" "MONODONTIDAE" ...
 $ n       : int  24 56 7 2 1 3 4 1 1 9 ...
```

13. Extract a data frame containing all entries of the families Monodontidae, Phocoenidae, and Delphinidae from the original species code table. How many genera are there in each family?

```
'data.frame':   65 obs. of  8 variables:
 $ SPCODE      : chr  "002" "003" "004" "005" ...
 $ ORDER       : chr  "CETACEA" "CETACEA" "CETACEA" "CETACEA" ...
 $ SUBORDER    : chr  "ODONTOCETI" "ODONTOCETI" "ODONTOCETI" "ODONTOCETI" ...
 $ FAMILY      : chr  "DELPHINIDAE" "DELPHINIDAE" "DELPHINIDAE" "DELPHINIDAE" ...
 $ FAMILY.NAMES: chr  "DOLPHINS" "DOLPHINS" "DOLPHINS" "DOLPHINS" ...
 $ GENUS       : chr  "Stenella" "Stenella" "Stenella" "Delphinus" ...
 $ SPECIES     : chr  "attenuata" "longirostris subsp." "clymene" "sp." ...
 $ COMMON.NAME : chr  "Pantropical spotted dolphin" "unidentified spinner dolphin" "Clymene dolphin" "U
```

```
 DELPHINIDAE MONODONTIDAE  PHOCOENIDAE
          20            2            5
```

14. Sort the species code table by genus, species, and species code. Put these three columns first.

```
'data.frame':   165 obs. of  8 variables:
 $ GENUS       : chr  "" "" "Arctocephalus" "Arctocephalus" ...
 $ SPECIES     : chr  "" "" "australis" "forsteri" ...
 $ SPCODE      : chr  "800" "801" "AA" "AF" ...
 $ ORDER       : chr  "" "" "CARNIVORA" "CARNIVORA" ...
 $ SUBORDER    : chr  "" "" "PINNIPEDIA" "PINNIPEDIA" ...
 $ FAMILY      : chr  "" "" "OTARIIDAE" "OTARIIDAE" ...
 $ FAMILY.NAMES: chr  "" "" "SEA LIONS & FUR SEALS" "SEA LIONS & FUR SEALS" ...
 $ COMMON.NAME : chr  "Wooly mammoth (in temporary use only, 9/20/2004)" "Bison species (in temporary us
```

```
            GENUS       SPECIES SPCODE     ORDER   SUBORDER    FAMILY
120                                800
121                                801
122 Arctocephalus     australis     AA CARNIVORA PINNIPEDIA OTARIIDAE
123 Arctocephalus      forsteri     AF CARNIVORA PINNIPEDIA OTARIIDAE
124 Arctocephalus galapagoensis     AG CARNIVORA PINNIPEDIA OTARIIDAE
126 Arctocephalus       gazella     AZ CARNIVORA PINNIPEDIA OTARIIDAE
             FAMILY.NAMES                                      COMMON.NAME
120                         Wooly mammoth (in temporary use only, 9/20/2004)
121                         Bison species (in temporary use only, 9/20/2004)
122 SEA LIONS & FUR SEALS                            South American fur seal
123 SEA LIONS & FUR SEALS                               New Zealand fur seal
124 SEA LIONS & FUR SEALS                                 Galapagos fur seal
126 SEA LIONS & FUR SEALS                                 Antarctic fur seal
```

15. Use the CTD cast data "ctd.csv". Extract the full casts for 10 random dates from Station 1.

```
'data.frame':   600 obs. of  9 variables:
 $ station    : chr  "Station.1" "Station.1" "Station.1" "Station.1" ...
 $ sample_date: chr  "2011-01-05" "2012-03-06" "2012-03-06" "2011-01-05" ...
```

```
 $ temp       : num  14.2 13.2 13.3 14.2 13.9 ...
 $ salinity   : num  33.3 33.4 33.4 33.3 33.4 ...
 $ dox        : num  7.91 7.19 7.45 7.93 8.17 6.96 8.13 7.97 7.9 6.95 ...
 $ ph         : num  8.16 8.05 8.07 8.16 8.13 8.18 8.23 8.16 8.16 8.17 ...
 $ pct_light  : num  86.2 87.6 87 86.2 77.7 ...
 $ density    : num  24.8 25.1 25.1 24.8 25 ...
 $ depth      : int  3 11 9 4 6 17 4 7 1 30 ...
```

16. Sort the data frame from 5 by date and depth.

```
       station sample_date  temp salinity  dox   ph pct_light density depth
1261 Station.1  2010-02-25 15.47   33.421 8.01 8.17     88.43  24.652     1
1624 Station.1  2010-02-25 15.50   33.420 8.01 8.17     88.85  24.644     2
1744 Station.1  2010-02-25 15.49   33.420 8.02 8.17     88.98  24.646     3
1013 Station.1  2010-02-25 15.49   33.420 8.03 8.17     89.00  24.646     4
1376 Station.1  2010-02-25 15.50   33.420 7.98 8.17     89.00  24.645     5
1616 Station.1  2010-02-25 15.43   33.420 8.01 8.17     88.95  24.660     6
```

```
       station sample_date  temp salinity  dox   ph pct_light density depth
1492 Station.1  2015-11-09 15.76   33.414 6.52 8.04     88.83  24.582    55
1372 Station.1  2015-11-09 15.71   33.417 6.52 8.04     88.90  24.596    56
653  Station.1  2015-11-09 15.66   33.418 6.48 8.04     88.93  24.608    57
1016 Station.1  2015-11-09 15.54   33.419 6.41 8.04     88.90  24.634    58
1495 Station.1  2015-11-09 15.37   33.427 6.32 8.03     88.95  24.678    59
1493 Station.1  2015-11-09 15.27   33.438 6.29 8.02     88.82  24.709    60
```

17. Which date of these 10 casts has the smallest surface temperature?

```
     sample_date  temp
1261  2010-02-25 15.47
103   2011-01-05 14.20
3421  2011-04-05 16.47
868   2011-08-24 18.47
132   2012-03-06 14.49
2027  2012-07-10 18.58
2286  2013-05-08 17.27
2015  2013-12-06 16.44
2975  2014-11-17 19.52
634   2015-11-09 21.72
```

```
      station sample_date temp salinity dox   ph pct_light density depth
103 Station.1  2011-01-05 14.2   33.279 7.9 8.16     84.04  24.817     1
```