Lab 3 In-class assignment: Data summaries

brouwern@gmail.com

September 11, 2017

Fisher's Iris dataset

Load it

```
data(iris)
```

What "is" iris?

There are several things listed, but "data.frame" is the important one

```
is(iris)
```

```
## [1] "data.frame" "list"       "oldClass"    "vector"
```

How many rows?

```
#with dim: 1st number is rows
dim(iris)
```

```
## [1] 150   5
```

```
#with nrow
nrow(iris)
```

```
## [1] 150
```

How many columns

```
#with dim: 2nd number is columns
dim(iris)
```

```
## [1] 150   5
```

```
#with ncol
ncol(iris)
```

```
## [1] 5
```

How many rows are numeric? How many are categorical/grouping?

```
summary(iris)
```

```
##   Sepal.Length    Sepal.Width     Petal.Length    Petal.Width
##  Min.   :4.300   Min.   :2.000   Min.   :1.000   Min.   :0.100
##  1st Qu.:5.100   1st Qu.:2.800   1st Qu.:1.600   1st Qu.:0.300
##  Median :5.800   Median :3.000   Median :4.350   Median :1.300
```

```
##  Mean   :5.843   Mean   :3.057   Mean   :3.758   Mean   :1.199
##  3rd Qu.:6.400   3rd Qu.:3.300   3rd Qu.:5.100   3rd Qu.:1.800
##  Max.   :7.900   Max.   :4.400   Max.   :6.900   Max.   :2.500
##        Species
##  setosa    :50
##  versicolor:50
##  virginica :50
##
##
##
```

What "is" the species column?

```
is(iris$Species)
```

```
## [1] "factor"            "integer"           "oldClass"
## [4] "numeric"           "vector"            "data.frameRowLabels"
```

How many "levels" does this factor have?

```
levels(iris$Species)
```

```
## [1] "setosa"     "versicolor" "virginica"
```

Mammal sleep data

```
library(ggplot2)
data(msleep)
```

What "is" msleep?

Several things listed, but "data.frame" is the most important one.
```
is(msleep)
```

```
## [1] "tbl_df"     "tbl"        "data.frame" "list"       "oldClass"
## [6] "vector"
```

How many rows?

```
#with dim: 1st number is rows
dim(msleep)
```

```
## [1] 83 11
```

```
#with ncrow
nrow(msleep)
```

```
## [1] 83
```

How many columns

```r
#with dim: 2nd number is columns
dim(msleep)
```

```
## [1] 83 11
```

```r
#with ncol()
ncol(msleep)
```

```
## [1] 11
```

How many rows are numeric? How many are categorical/grouping?

For some reason the categorical variabls are in "character" (aka text) form. (Note: I'm just showing a few of the columns)

```r
summary(msleep)
```

```
##      name               genus               vore
##  Length:83          Length:83          Length:83
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##     order           conservation         sleep_total        sleep_rem
##  Length:83          Length:83          Min.   : 1.90    Min.    :0.100
##  Class :character   Class :character   1st Qu.: 7.85    1st Qu.:0.900
##  Mode  :character   Mode  :character   Median :10.10    Median :1.500
##                                        Mean   :10.43    Mean    :1.875
##                                        3rd Qu.:13.75    3rd Qu.:2.400
##                                        Max.   :19.90    Max.    :6.600
##                                                         NA's    :22
```

Make them explicitly categorical/factor

```r
msleep$genus <- factor(msleep$genus)
msleep$vore <- factor(msleep$vore)
msleep$order <- factor(msleep$order)
msleep$conservation <- factor(msleep$conservation)
```

Look at summary again

```r
summary(msleep)
```

```
##      name                   genus          vore              order
##  Length:83          Panthera   : 3    carni  :19    Rodentia    :22
##  Class :character   Spermophilus: 3   herbi  :32    Carnivora   :12
##  Mode  :character   Equus      : 2    insecti: 5    Primates    :12
##                     Vulpes     : 2    omni   :20    Artiodactyla: 6
##                     Acinonyx   : 1    NA's   : 7    Soricomorpha: 5
##                     Aotus      : 1                  Cetacea     : 3
```

```
##                            (Other)   :71                  (Other)   :23
```

Levels to genus variable

Output not shown; there are 77!

```r
levels(msleep$genus)
```