

Exercise 1.2

Characterize a data frame

useful commands: `[row , col]`, `str()`, `dim()`, `nrow()`, `unique()`, `length()`, `rownames()`, `summary()`, `min()`, `max()`, `mean()`, `range()`, `levels()`, `factor()`, `as.factor()`, `class()`, `ncol()`, `nrow()`, `table()`, `sum()`, `quantile()`, `var()`

We'll use the built-in 'iris' dataset

```
data(iris) # this loads the 'iris' dataset
```

load help file for the 'iris' dataset

```
?iris  
help(iris)
```

this calls the object iris

```
head(iris) ## let's not print the entire thing to our screens
```

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species  
## 1         5.1         3.5          1.4          0.2  setosa  
## 2         4.9         3.0          1.4          0.2  setosa  
## 3         4.7         3.2          1.3          0.2  setosa  
## 4         4.6         3.1          1.5          0.2  setosa  
## 5         5.0         3.6          1.4          0.2  setosa  
## 6         5.4         3.9          1.7          0.4  setosa
```

If you're not a botanist, remind yourself what an iris looks like <http://suruchifialoke.com/img/ML/iris.png>

* 01. How many rows are in the dataset?

```
nrow(iris)
```

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

02. What are three distinct ways to figure this out?

```
nrow(iris)
```

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

```
dim(iris)[1]
```

```
## [1] 150
length(rownames(iris))

## [1] 150
str(iris)

## 'data.frame': 150 obs. of 5 variables:
## $ Sepal.Length: num 5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
## $ Sepal.Width : num 3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
## $ Petal.Length: num 1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
## $ Petal.Width : num 0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
## $ Species : Factor w/ 3 levels "setosa","versicolor",...: 1 1 1 1 1 1 1 1 1 1 ...
```

* 03. How many species of flowers are in the dataset?

```
length(unique(iris[, 'Species'])) ## this counts how long the unique() vector is

## [1] 3
length(table(iris[, 'Species'])) ## this asks how many entries are in the table

## [1] 3
```

04. What class is iris?

```
class(iris)

## [1] "data.frame"
```

* 05. How many columns does this data frame have? What are their names?

```
ncol(iris)

## [1] 5
colnames(iris)

## [1] "Sepal.Length" "Sepal.Width" "Petal.Length" "Petal.Width"
## [5] "Species"
```

* 06. What class did R assign to each column?

```
class(iris[,1])

## [1] "numeric"
```

```
class(iris[,2])
```

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

```
class(iris[,3])
```

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

```
class(iris[,4])
```

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

```
class(iris[,5])
```

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

Or, we can use `lapply()`, which we'll learn about later

```
## ssh, we're not supposed to learn this until Day 3
```

```
lapply(iris, class)
```

```
## $Sepal.Length
```

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

```
##
```

```
## $Sepal.Width
```

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

```
##
```

```
## $Petal.Length
```

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

```
##
```

```
## $Petal.Width
```

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

```
##
```

```
## $Species
```

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

07. Assign the first flower's petal's width and length to new objects called `setosa1.petalwidth` and `setosa1.petallength`

```
setosa1.petalwidth = iris[1,'Petal.Width']
```

```
setosa1.petallength = iris[1,'Petal.Length']
```

*** 08. Calculate the approximate area of the petal of the first flower, `setosa1` (assume petal area can be approximated by a rectangle).**

```
setosa1.petalwidth * setosa1.petallength
```

```
## [1] 0.28
```

09. Calculate the petal area of each flower in the iris dataset and assign this to a new column named PetalArea.

```
iris['PetalArea'] = iris['Petal.Width'] * iris['Petal.Length']
head(iris)
```

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species	PetalArea
## 1	5.1	3.5	1.4	0.2	setosa	0.28
## 2	4.9	3.0	1.4	0.2	setosa	0.28
## 3	4.7	3.2	1.3	0.2	setosa	0.26
## 4	4.6	3.1	1.5	0.2	setosa	0.30
## 5	5.0	3.6	1.4	0.2	setosa	0.28
## 6	5.4	3.9	1.7	0.4	setosa	0.68

10. What is the maximum sepal length of the irises?

```
max(iris[, 'Sepal.Length'])
```

```
## [1] 7.9
```

* 11. What is the average sepal length among flowers in the dataset?

```
mean(iris[, 'Sepal.Length'])
```

```
## [1] 5.843333
```

12. How about the minimum and median sepal length?

```
min(iris[, 'Sepal.Length'])
```

```
## [1] 4.3
```

```
median(iris[, 'Sepal.Length'])
```

```
## [1] 5.8
```