

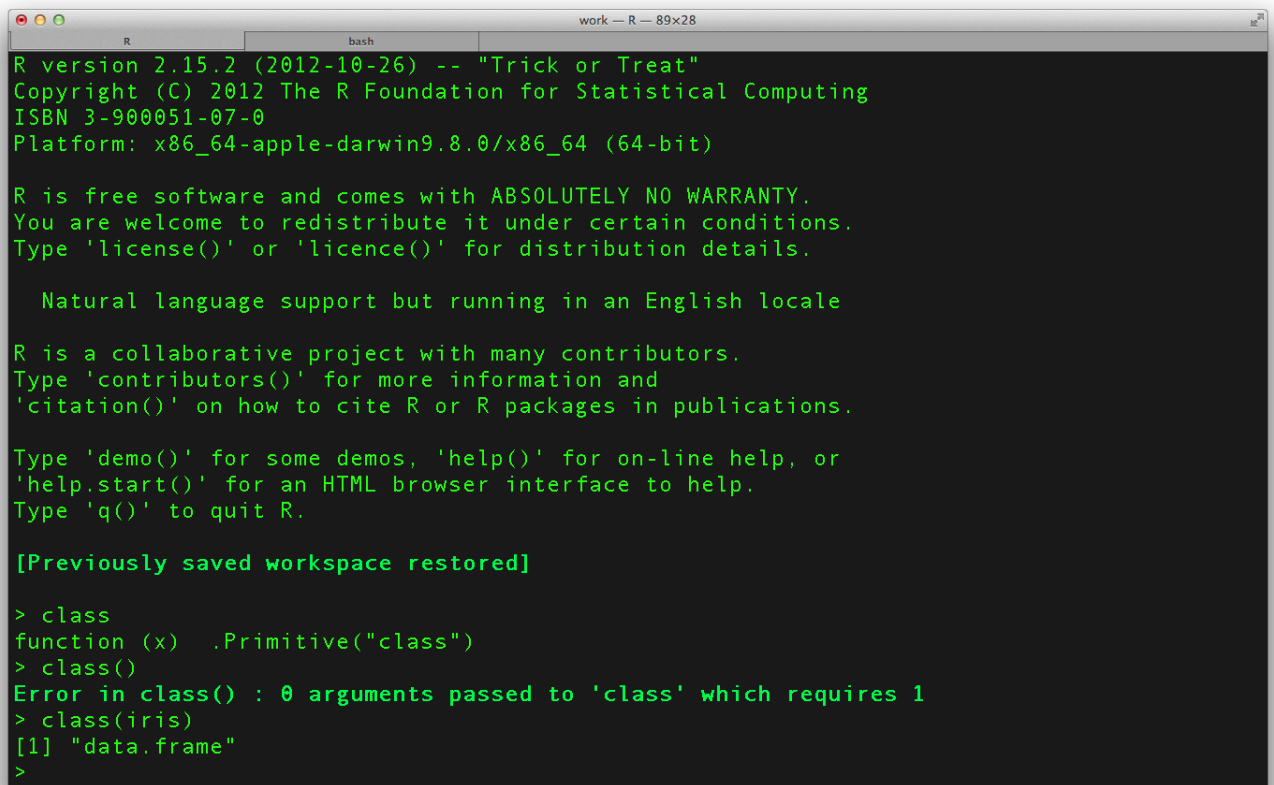
Project 2 - PCA

Blair Gemmer

CSCI 548 - Pattern Recognition

Spring 2013

1. `class(iris)` Command.



```
R version 2.15.2 (2012-10-26) -- "Trick or Treat"
Copyright (C) 2012 The R Foundation for Statistical Computing
ISBN 3-900051-07-0
Platform: x86_64-apple-darwin9.8.0/x86_64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

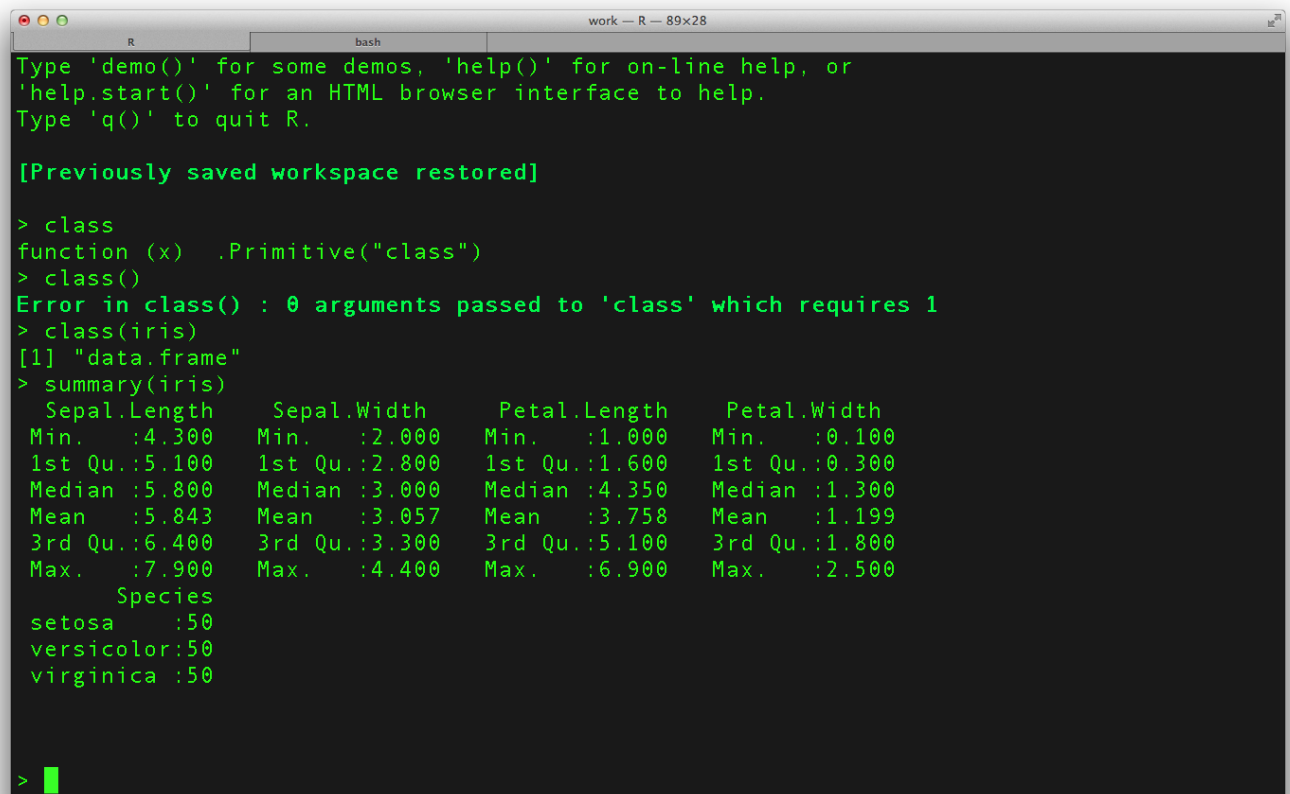
[Previously saved workspace restored]

> class
function (x) .Primitive("class")
> class()
Error in class() : 0 arguments passed to 'class' which requires 1
> class(iris)
[1] "data.frame"
>
```

Question Answer #1

Technically, a dataframe in R is a type of *object*. Less formally, a dataframe is a type of table where the typical use employs the rows as observations and the columns as variables. (<http://msenex.redwoods.edu/math/R/dataframe.php>)

2. `summary(iris)` Command.



```
work -- R -- 89x28
R
bash
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Previously saved workspace restored]

> class
function (x) .Primitive("class")
> class()
Error in class() : 0 arguments passed to 'class' which requires 1
> class(iris)
[1] "data.frame"
> 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

> █
```

3. labels(iris) Command.

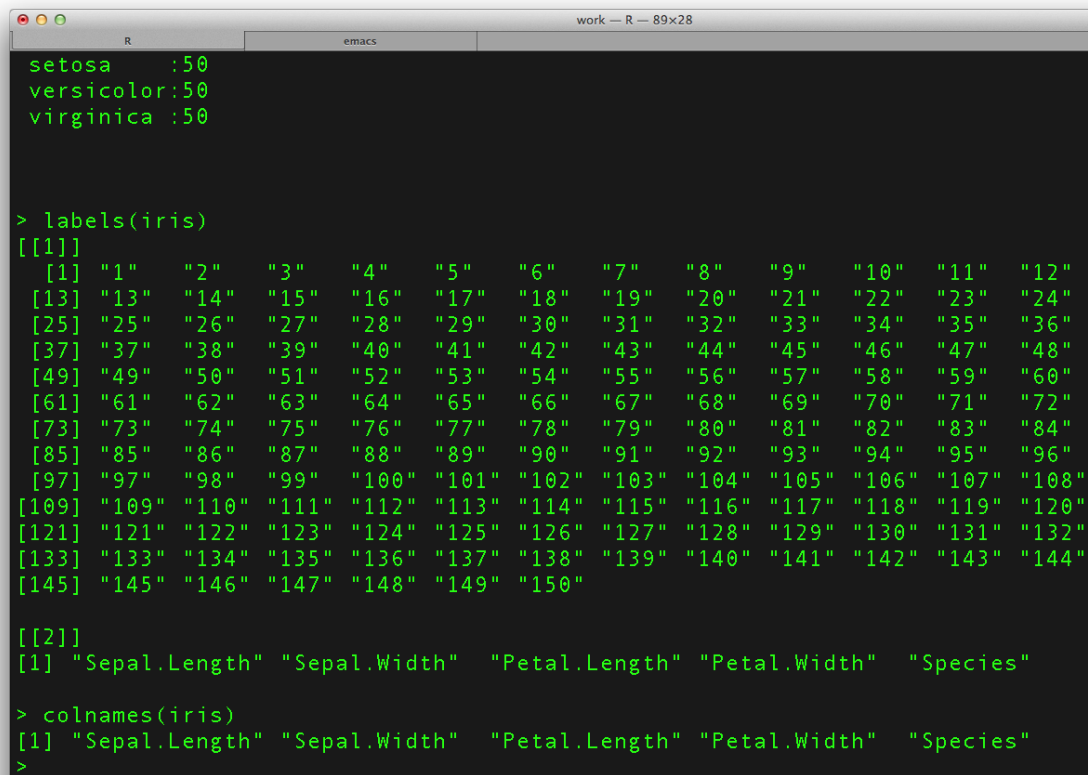
```
work -- R -- 89x28
R
Max. :7.900 Max. :4.400 Max. :6.900 Max. :2.500
Species
setosa :50
versicolor:50
virginica :50

> labels(iris)
[[1]]
 [1] "1" "2" "3" "4" "5" "6" "7" "8" "9" "10" "11" "12"
[13] "13" "14" "15" "16" "17" "18" "19" "20" "21" "22" "23" "24"
[25] "25" "26" "27" "28" "29" "30" "31" "32" "33" "34" "35" "36"
[37] "37" "38" "39" "40" "41" "42" "43" "44" "45" "46" "47" "48"
[49] "49" "50" "51" "52" "53" "54" "55" "56" "57" "58" "59" "60"
[61] "61" "62" "63" "64" "65" "66" "67" "68" "69" "70" "71" "72"
[73] "73" "74" "75" "76" "77" "78" "79" "80" "81" "82" "83" "84"
[85] "85" "86" "87" "88" "89" "90" "91" "92" "93" "94" "95" "96"
[97] "97" "98" "99" "100" "101" "102" "103" "104" "105" "106" "107" "108"
[109] "109" "110" "111" "112" "113" "114" "115" "116" "117" "118" "119" "120"
[121] "121" "122" "123" "124" "125" "126" "127" "128" "129" "130" "131" "132"
[133] "133" "134" "135" "136" "137" "138" "139" "140" "141" "142" "143" "144"
[145] "145" "146" "147" "148" "149" "150"

[[2]]
[1] "Sepal.Length" "Sepal.Width" "Petal.Length" "Petal.Width" "Species"

>
```

4. colnames(iris) Command.



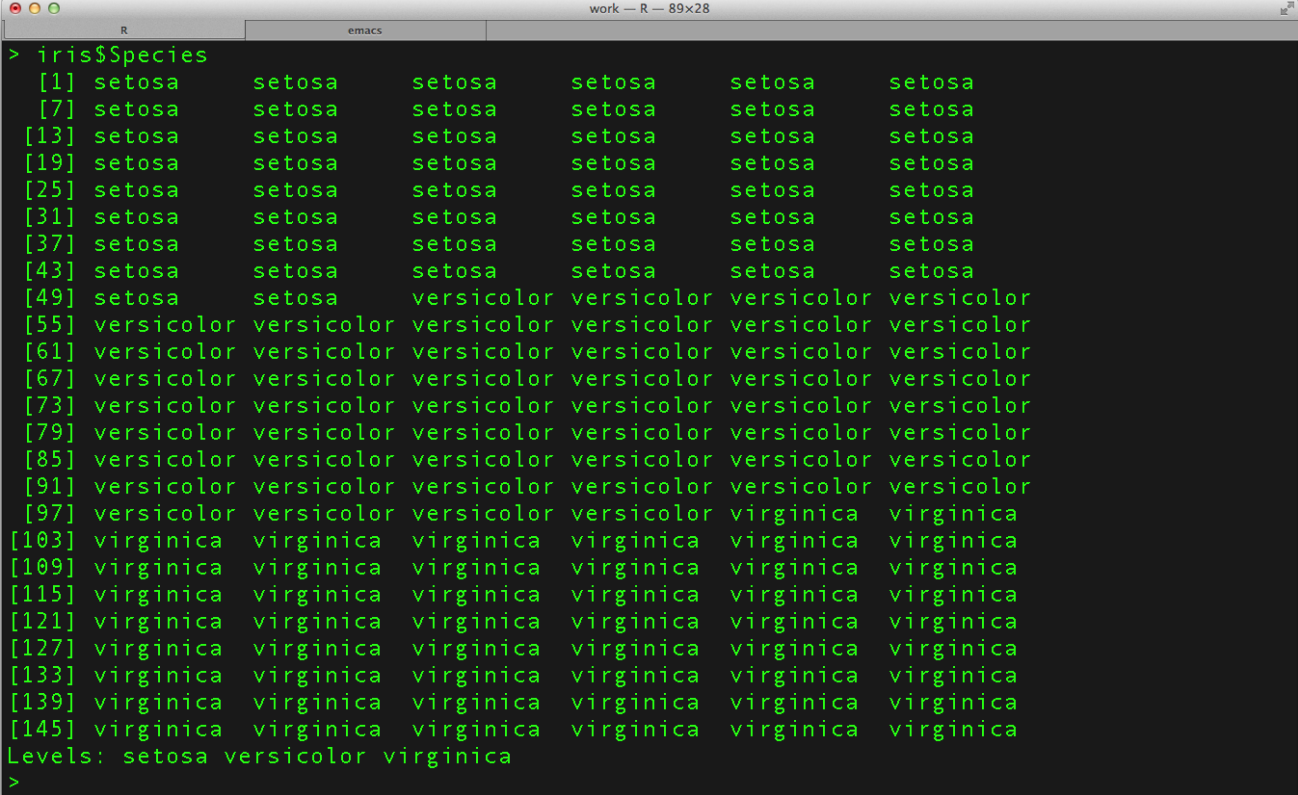
```
setosa :50
versicolor:50
virginica :50

> labels(iris)
[[1]]
 [1] "1" "2" "3" "4" "5" "6" "7" "8" "9" "10" "11" "12"
[13] "13" "14" "15" "16" "17" "18" "19" "20" "21" "22" "23" "24"
[25] "25" "26" "27" "28" "29" "30" "31" "32" "33" "34" "35" "36"
[37] "37" "38" "39" "40" "41" "42" "43" "44" "45" "46" "47" "48"
[49] "49" "50" "51" "52" "53" "54" "55" "56" "57" "58" "59" "60"
[61] "61" "62" "63" "64" "65" "66" "67" "68" "69" "70" "71" "72"
[73] "73" "74" "75" "76" "77" "78" "79" "80" "81" "82" "83" "84"
[85] "85" "86" "87" "88" "89" "90" "91" "92" "93" "94" "95" "96"
[97] "97" "98" "99" "100" "101" "102" "103" "104" "105" "106" "107" "108"
[109] "109" "110" "111" "112" "113" "114" "115" "116" "117" "118" "119" "120"
[121] "121" "122" "123" "124" "125" "126" "127" "128" "129" "130" "131" "132"
[133] "133" "134" "135" "136" "137" "138" "139" "140" "141" "142" "143" "144"
[145] "145" "146" "147" "148" "149" "150"

[[2]]
[1] "Sepal.Length" "Sepal.Width" "Petal.Length" "Petal.Width" "Species"

> colnames(iris)
[1] "Sepal.Length" "Sepal.Width" "Petal.Length" "Petal.Width" "Species"
>
```

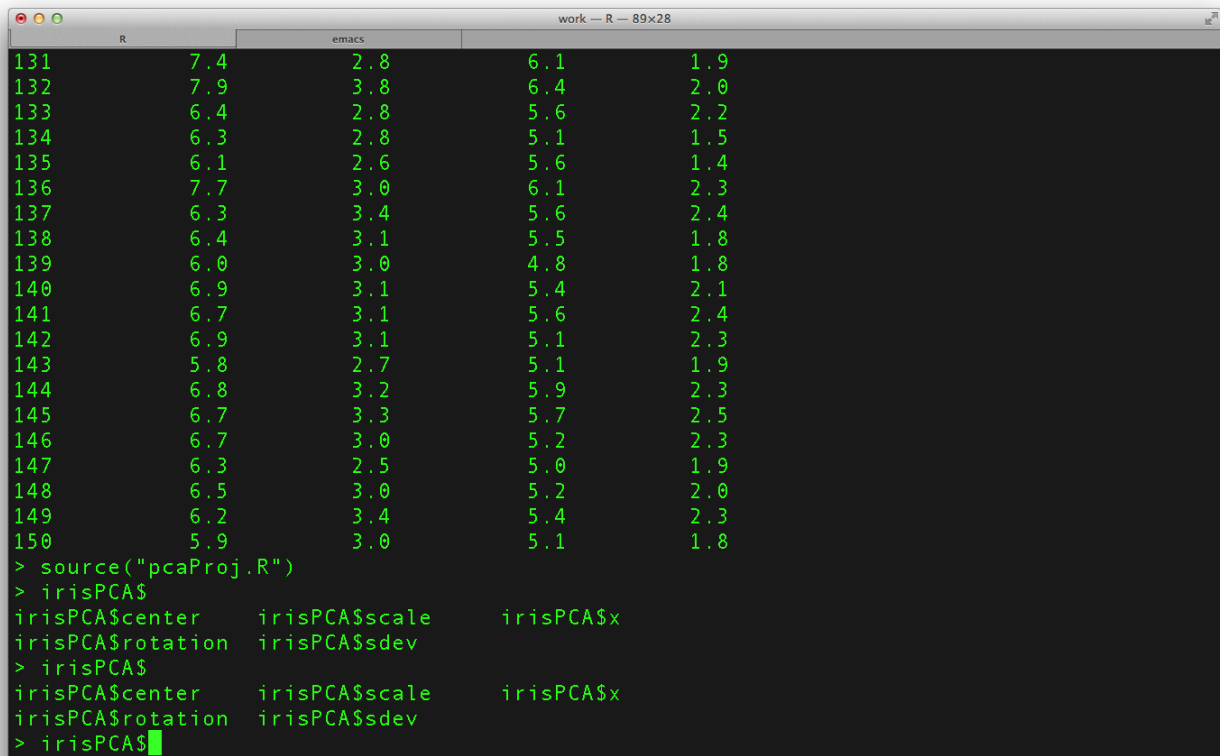
5. Tab Completion Command.



The screenshot shows an R console window titled "work - R - 89x28" with tabs for "R" and "emacs". The command `> iris$Species` has been entered, and the console displays the species names for each row of the iris dataset. The output is as follows:

```
> iris$Species
 [1] setosa      setosa      setosa      setosa      setosa      setosa
 [7] setosa      setosa      setosa      setosa      setosa      setosa
[13] setosa      setosa      setosa      setosa      setosa      setosa
[19] setosa      setosa      setosa      setosa      setosa      setosa
[25] setosa      setosa      setosa      setosa      setosa      setosa
[31] setosa      setosa      setosa      setosa      setosa      setosa
[37] setosa      setosa      setosa      setosa      setosa      setosa
[43] setosa      setosa      setosa      setosa      setosa      setosa
[49] setosa      setosa      versicolor  versicolor  versicolor  versicolor
[55] versicolor  versicolor  versicolor  versicolor  versicolor  versicolor
[61] versicolor  versicolor  versicolor  versicolor  versicolor  versicolor
[67] versicolor  versicolor  versicolor  versicolor  versicolor  versicolor
[73] versicolor  versicolor  versicolor  versicolor  versicolor  versicolor
[79] versicolor  versicolor  versicolor  versicolor  versicolor  versicolor
[85] versicolor  versicolor  versicolor  versicolor  versicolor  versicolor
[91] versicolor  versicolor  versicolor  versicolor  versicolor  versicolor
[97] versicolor  versicolor  versicolor  versicolor  virginica   virginica
[103] virginica   virginica   virginica   virginica   virginica   virginica
[109] virginica   virginica   virginica   virginica   virginica   virginica
[115] virginica   virginica   virginica   virginica   virginica   virginica
[121] virginica   virginica   virginica   virginica   virginica   virginica
[127] virginica   virginica   virginica   virginica   virginica   virginica
[133] virginica   virginica   virginica   virginica   virginica   virginica
[139] virginica   virginica   virginica   virginica   virginica   virginica
[145] virginica   virginica   virginica   virginica   virginica   virginica
Levels: setosa versicolor virginica
>
```

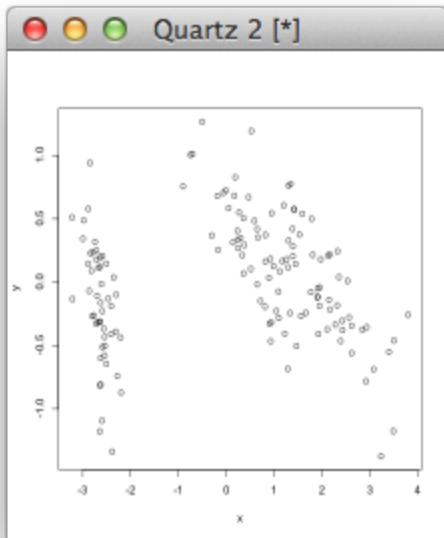
6. PCA Performed.



The screenshot shows an R console window titled "work - R - 89x28". The console displays the following output:

```
131      7.4      2.8      6.1      1.9
132      7.9      3.8      6.4      2.0
133      6.4      2.8      5.6      2.2
134      6.3      2.8      5.1      1.5
135      6.1      2.6      5.6      1.4
136      7.7      3.0      6.1      2.3
137      6.3      3.4      5.6      2.4
138      6.4      3.1      5.5      1.8
139      6.0      3.0      4.8      1.8
140      6.9      3.1      5.4      2.1
141      6.7      3.1      5.6      2.4
142      6.9      3.1      5.1      2.3
143      5.8      2.7      5.1      1.9
144      6.8      3.2      5.9      2.3
145      6.7      3.3      5.7      2.5
146      6.7      3.0      5.2      2.3
147      6.3      2.5      5.0      1.9
148      6.5      3.0      5.2      2.0
149      6.2      3.4      5.4      2.3
150      5.9      3.0      5.1      1.8
> source("pcaProj.R")
> irisPCA$
irisPCA$center  irisPCA$scale  irisPCA$x
irisPCA$rotation irisPCA$sdev
> irisPCA$
irisPCA$center  irisPCA$scale  irisPCA$x
irisPCA$rotation irisPCA$sdev
> irisPCA$
```

7. `plot(x,y)` Command.



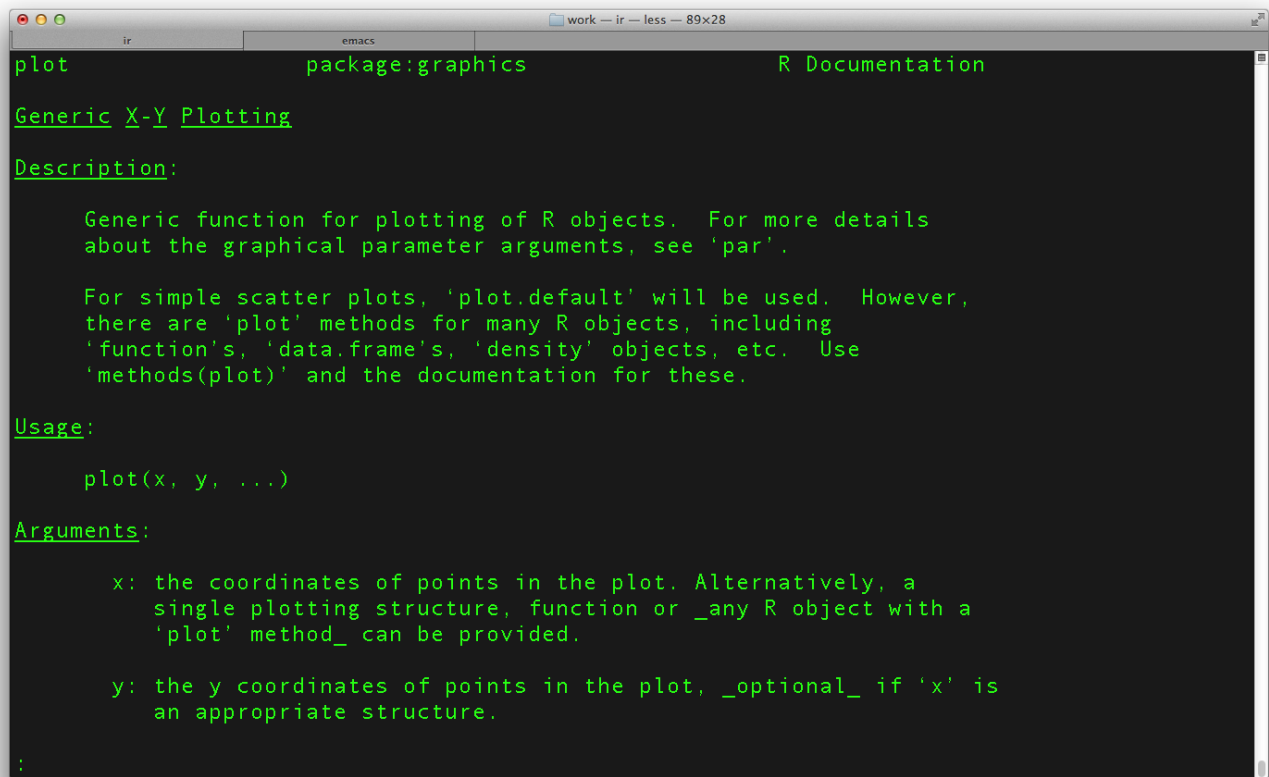
Question Answer #2

There are two distinct clusters which vary highly on the Y-axis and slightly on the X-axis.

Question Answer #3

It seems like you would see 3 different clusters since there were 3 different species of flowers sampled in the iris data.

8. `help(plot)` Command.



```
plot                                package:graphics                R Documentation

Generic X-Y Plotting

Description:

  Generic function for plotting of R objects.  For more details
  about the graphical parameter arguments, see 'par'.

  For simple scatter plots, 'plot.default' will be used.  However,
  there are 'plot' methods for many R objects, including
  'function's, 'data.frame's, 'density' objects, etc.  Use
  'methods(plot)' and the documentation for these.

Usage:

  plot(x, y, ...)

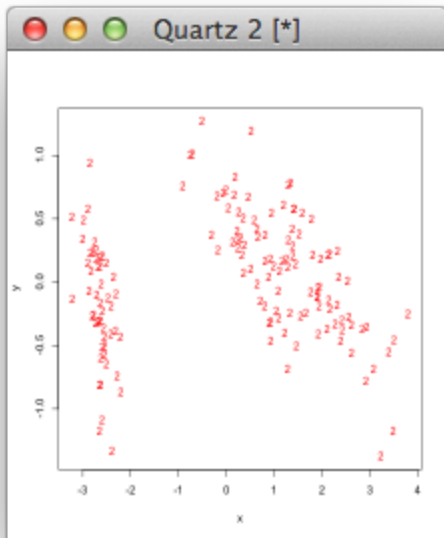
Arguments:

  x: the coordinates of points in the plot.  Alternatively, a
     single plotting structure, function or _any R object with a
     'plot' method_ can be provided.

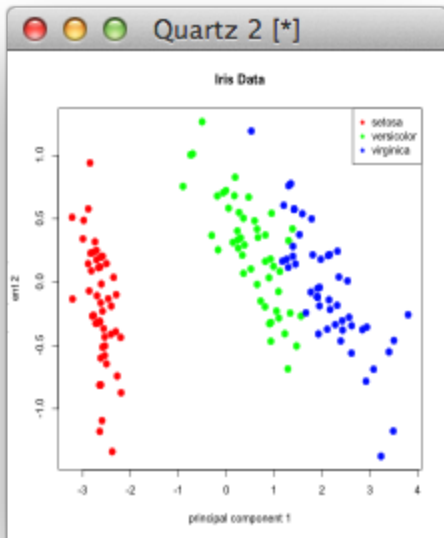
  y: the y coordinates of points in the plot, _optional_ if 'x' is
     an appropriate structure.

:
```

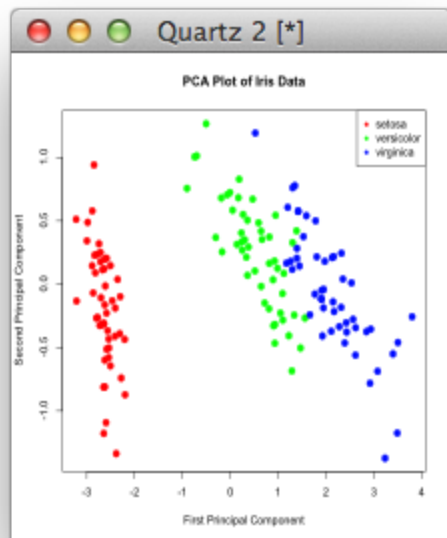
9. Red Filled Points.



10. Species-Colored Plot.



11. Filled and Updated Plot



Question Answer #4

It could be argued that the 3 clusters are separable enough to classify with a machine learning technique.

Question Answer #5

The versicolor and virginica classes would give you the most trouble because they are so close together on the plot.

Question Answer #6

I think if you made some function of x and y , you could linearly separate the 3 classes enough to do some machine learning on them, but it would be much more difficult because of how close the 3 classes are on the plot.

Question Answer #7

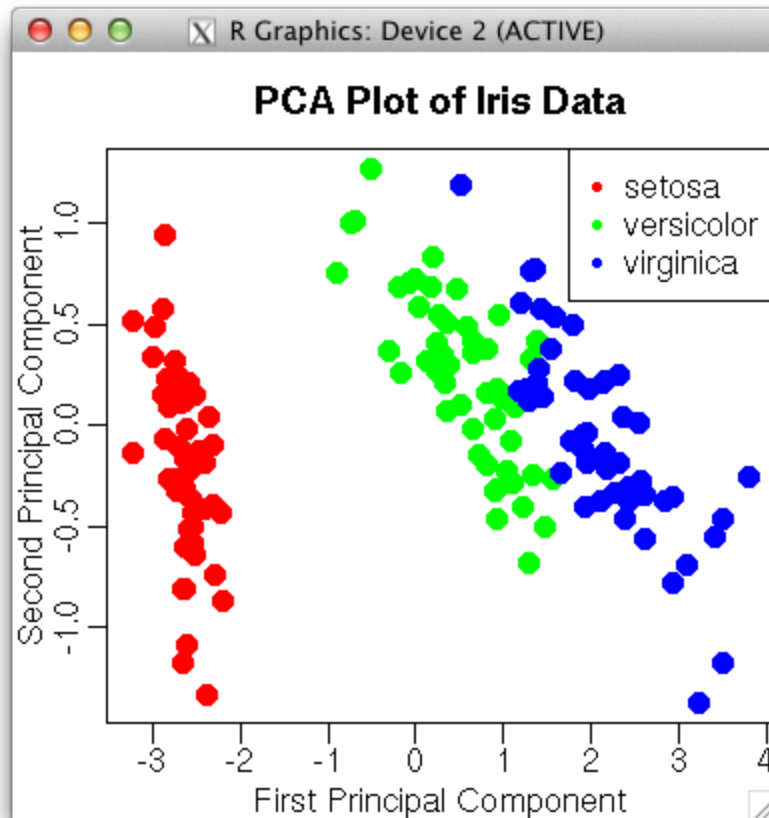
I think they definitely capture some variance, but I'm not certain that they capture anywhere close to 90% of the variance, since the plots seem so very close together.

Question Answer#8

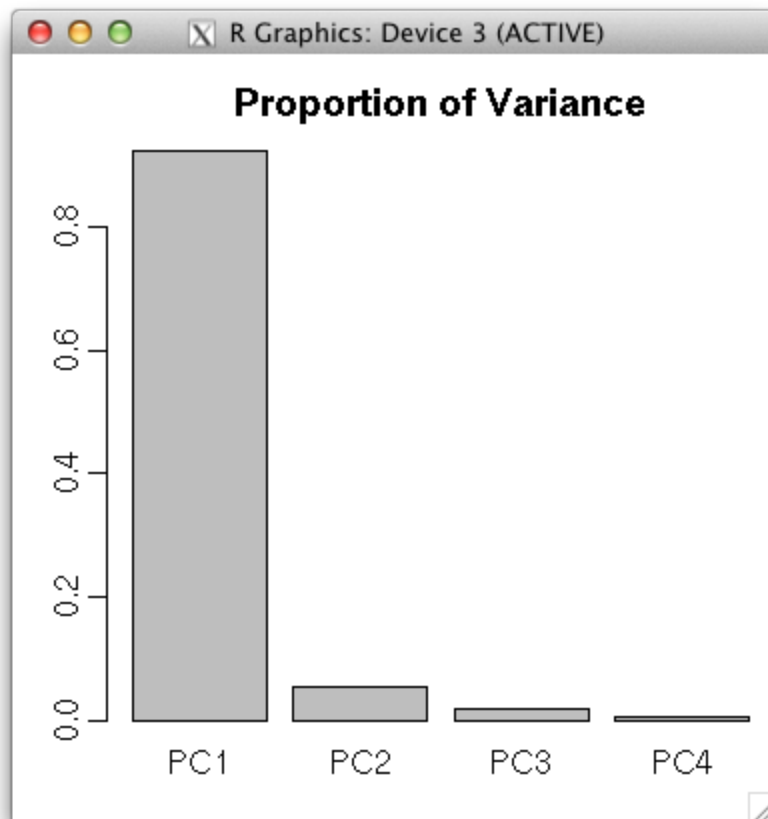
They are much more classifiable now! The variance is much easier to spot, now that I have changed the axes to the 2nd and 3rd principal components rather than the 1st and 2nd.

Script-Running Images:

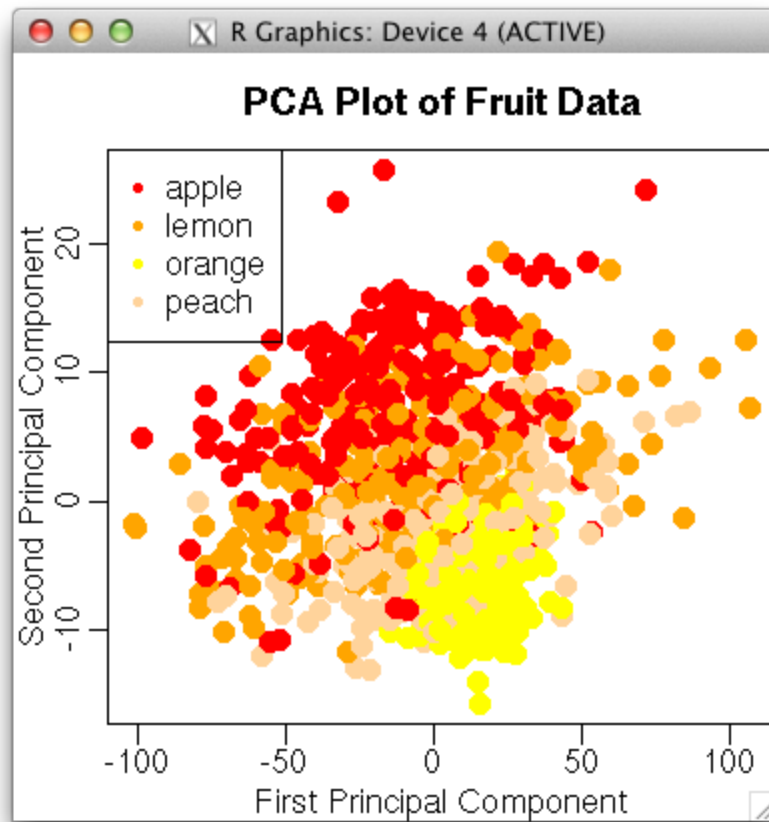
12. Iris Data.



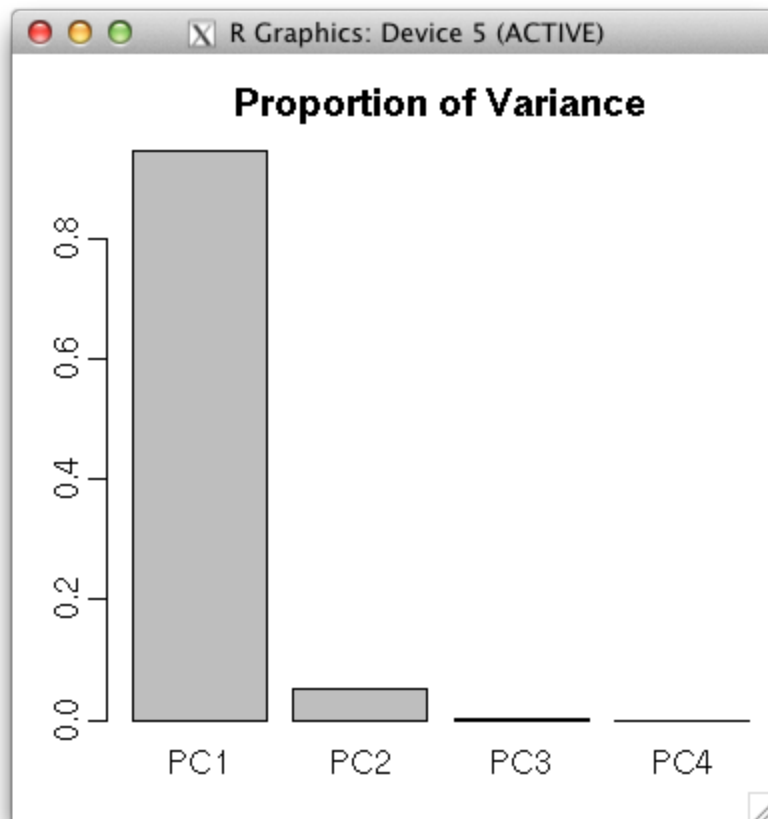
13. Proportions of Variance (iris).



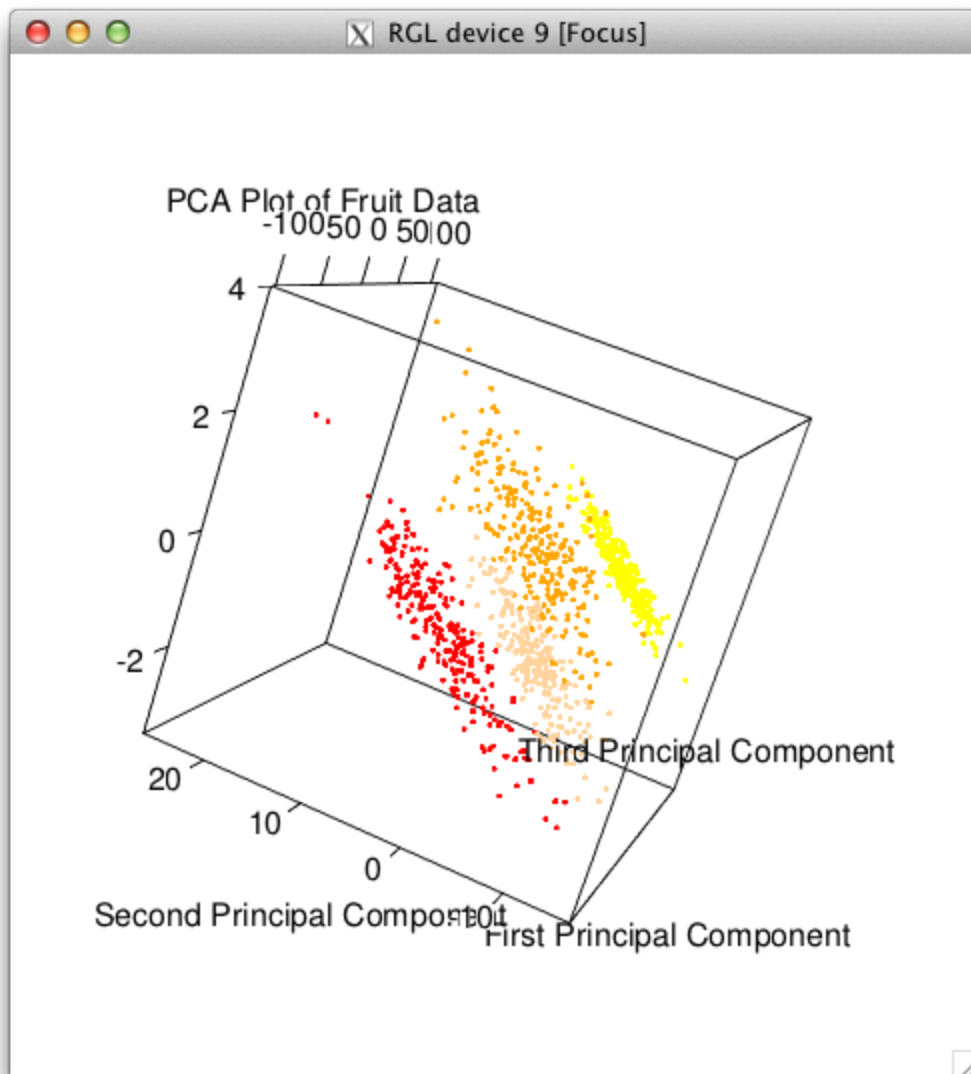
14. Fruit Data.



15. Proportions of Variance (fruit).



16. 3D Plot of Fruit Data.



17. 2D Plot of Fruit Data Using 2nd and 3rd Principal Components.

