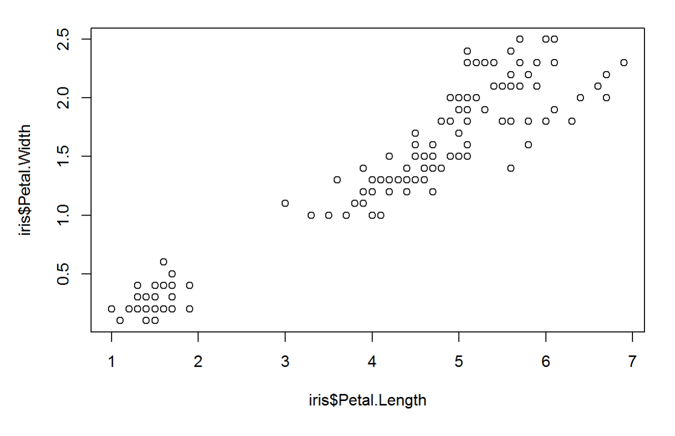
title: Synthetic Data for Social Scientists
output: HTML

Using the Iris data set first step is to plot the petal width and length. This will provide a benchmark for the sythetic data.



Use the Synthpop to model and synthese the dataset. Load the required packages

```
## Loading required package: synthpop
```

```
## Warning: package 'synthpop' was built under R version 3.3.3
```

```
## Loading required package: lattice
```

```
## Loading required package: MASS
```

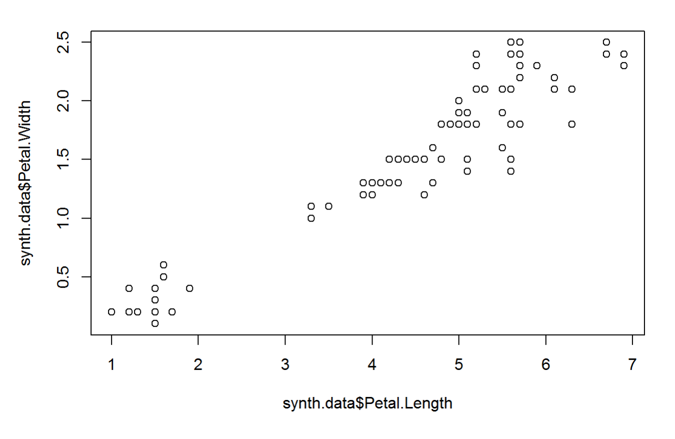
```
## Loading required package: nnet
```

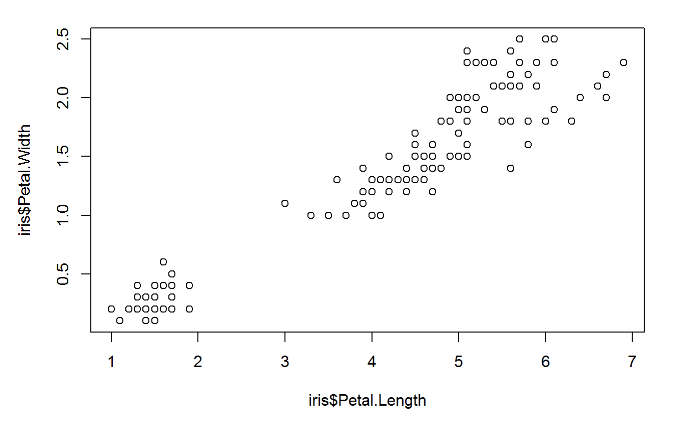
```
## Loading required package: ggplot2
```

```
## Warning: package 'ggplot2' was built under R version 3.3.3
```

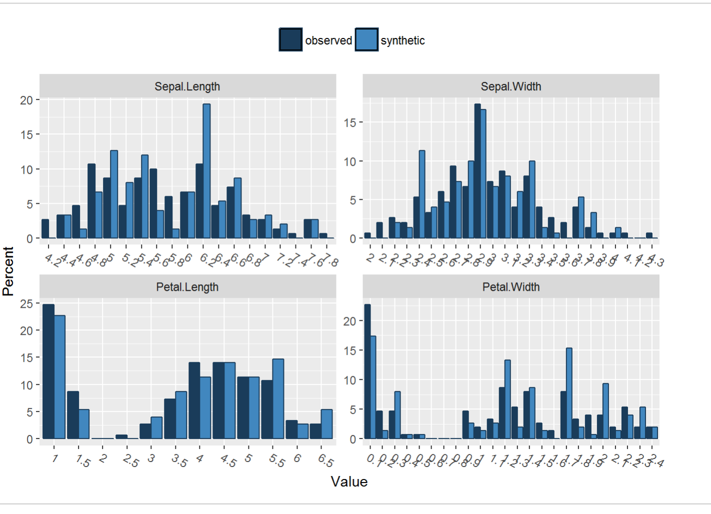
Use the Synthpop to model and synthese the dataset. Load the required dataset

```
## Element(s): {6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30,
31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58,
59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86,
87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111,
112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 1
34, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150} of the 'visit.sequence' rem
oved as not valid. No such column.
##
## syn  variables
## 1    Sepal.Length Sepal.Width Petal.Length Petal.Width Species
```

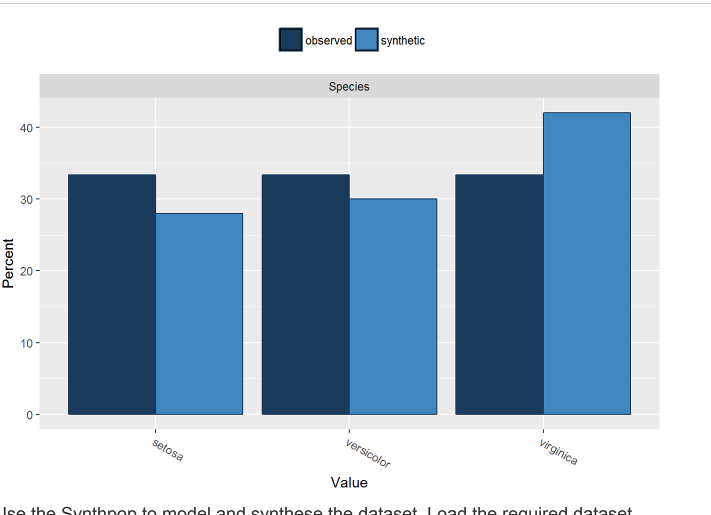





```
##
## Comparing percentages observed with synthetic
##
## $Sepal.Length
##                4.2      4.4      4.6       4.8         5      5.2
## observed  2.666667 3.333333 4.666667 10.666667  8.666667 4.666667
## synthetic 0.000000 3.333333 1.333333  6.666667 12.666667 8.000000
##                 5.4 5.6      5.8        6      6.2      6.4      6.6
## observed   8.666667  10 6.000000 6.666667 10.66667 4.666667 7.333333
## synthetic 12.000000   4 1.333333 6.666667 19.33333 5.333333 8.666667
##                6.8        7      7.2       7.4      7.6       7.8
## observed  3.333333 2.666667 1.333333 0.6666667 2.666667 0.6666667
## synthetic 2.666667 3.333333 2.000000 0.0000000 2.666667 0.0000000
##
## $Sepal.Width
##                   2 2.1      2.2      2.3       2.4      2.5      2.6
## observed  0.6666667   2 2.666667 2.000000  5.333333 3.333333 6.000000
## synthetic 0.0000000   0 2.000000 1.333333 11.333333 4.000000 4.666667
##                 2.7       2.8      2.9        3      3.1 3.2 3.3      3.4
## observed   9.333333  6.666667 17.33333 7.333333 8.666667   4   8 4.000000
## synthetic  7.333333 10.000000 16.66667 6.666667 8.000000   6  10 1.333333
##                 3.5 3.6      3.7      3.8       3.9         4       4.1
## observed  2.6666667   2 4.000000 1.333333 0.6666667 0.6666667 0.6666667
## synthetic 0.6666667   0 5.333333 3.333333 0.0000000 1.3333333 0.0000000
##           4.2       4.3
## observed    0 0.6666667
## synthetic   0 0.0000000
##
## $Petal.Length
##                  1      1.5 2       2.5        3      3.5        4 4.5
## observed  24.66667 8.666667 0 0.6666667 2.666667 7.333333 14.00000  14
## synthetic 22.66667 5.333333 0 0.0000000 4.000000 8.666667 11.33333  14
##                  5      5.5        6      6.5
## observed  11.33333 10.66667 3.333333 2.666667
## synthetic 11.33333 14.66667 2.666667 5.333333
##
## $Petal.Width
##                0.1      0.2      0.3       0.4       0.5 0.6 0.7 0.8
## observed  22.66667 4.666667 4.666667 0.6666667 0.6666667   0   0   0
## synthetic 17.33333 1.333333 8.000000 0.6666667 0.6666667   0   0   0
##                0.9        1      1.1       1.2      1.3      1.4      1.5
## observed  4.666667 2.000000 3.333333  8.666667 5.333333 8.000000 2.666667
## synthetic 2.666667 1.333333 2.666667 13.333333 2.000000 8.666667 1.333333
##                1.6      1.7      1.8       1.9        2      2.1      2.2
## observed  1.333333  8.00000 3.333333 4.0000000 4.000000 2.000000 5.333333
## synthetic 0.000000 15.33333 2.000000 0.6666667 9.333333 1.333333 4.000000
##                2.3 2.4
## observed  2.000000   2
## synthetic 5.333333   2
```

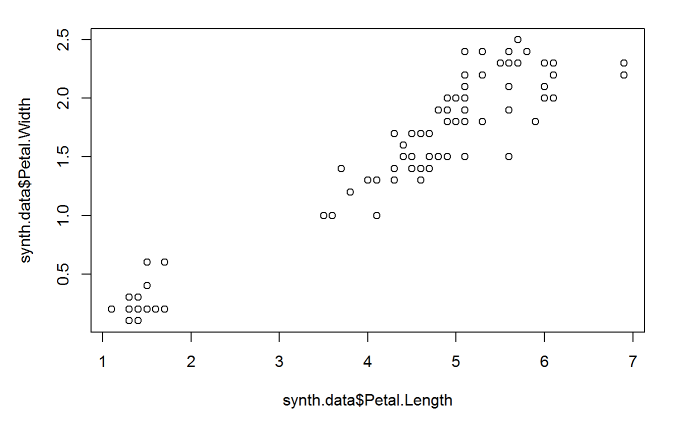


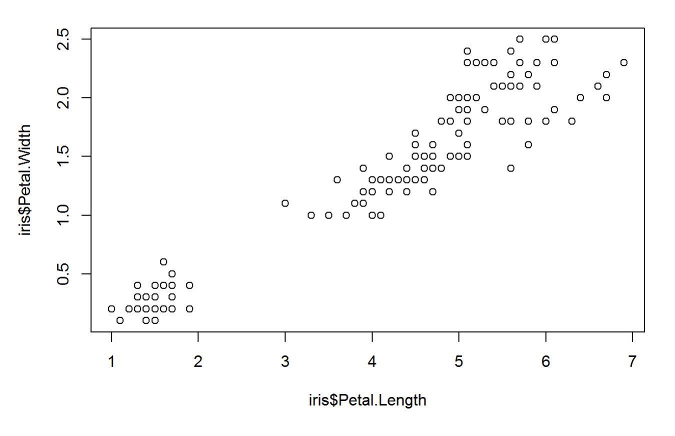

```
## Press return for next plot:
## $Species
##             setosa versicolor virginica
## observed  33.33333   33.33333  33.33333
## synthetic 28.00000   30.00000  42.00000
```



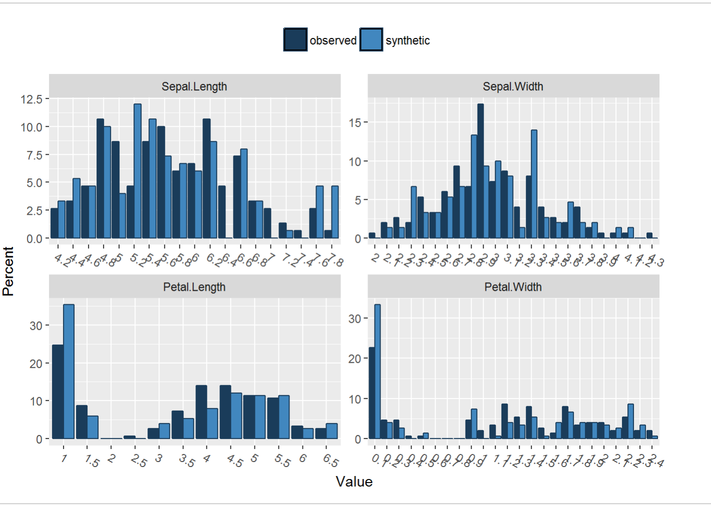
Use the Synthpop to model and synthese the dataset. Load the required dataset

```
## Element(s): {6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30,
31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58,
59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86,
87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111,
112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 1
34, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150} of the 'visit.sequence' rem
oved as not valid. No such column.
##
## syn  variables
## 1    Sepal.Length Sepal.Width Petal.Length Petal.Width Species
```

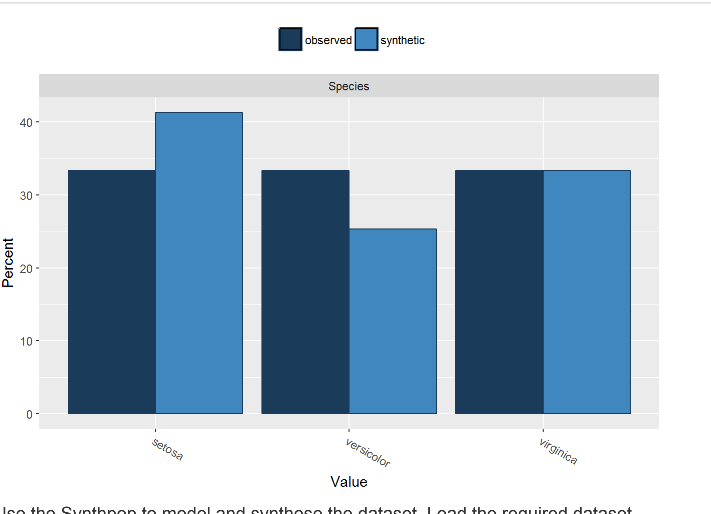





```
##
## Comparing percentages observed with synthetic
##
## $Sepal.Length
##                4.2      4.4      4.6      4.8        5       5.2       5.4
## observed  2.666667 3.333333 4.666667 10.66667 8.666667  4.666667  8.666667
## synthetic 3.333333 5.333333 4.666667 10.00000 4.000000 12.000000 10.666667
##                 5.6      5.8        6       6.2      6.4      6.6      6.8
## observed  10.000000 6.000000 6.666667 10.666667 4.666667 7.333333 3.333333
## synthetic  7.333333 6.666667 6.000000  8.666667 0.000000 8.000000 3.333333
##                  7       7.2       7.4      7.6       7.8
## observed  2.666667 1.3333333 0.6666667 2.666667 0.6666667
## synthetic 0.000000 0.6666667 0.0000000 4.666667 4.6666667
##
## $Sepal.Width
##                   2      2.1      2.2      2.3      2.4      2.5      2.6
## observed  0.6666667 2.000000 2.666667 2.000000 5.333333 3.333333 6.000000
## synthetic 0.0000000 1.333333 1.333333 6.666667 3.333333 3.333333 5.333333
##                2.7       2.8       2.9         3      3.1      3.2 3.3
## observed  9.333333  6.666667 17.333333  7.333333 8.666667 4.000000   8
## synthetic 6.666667 13.333333  9.333333 10.000000 8.000000 1.333333  14
##                3.4      3.5      3.6 3.7      3.8       3.9         4
## observed  4.000000 2.666667 2.000000   4 1.333333 0.6666667 0.6666667
## synthetic 2.666667 2.000000 4.666667   2 2.000000 0.0000000 1.3333333
##                 4.1 4.2       4.3
## observed  0.6666667   0 0.6666667
## synthetic 1.3333333   0 0.0000000
##
## $Petal.Length
##                  1      1.5 2       2.5        3      3.5  4 4.5        5
## observed  24.66667 8.666667 0 0.6666667 2.666667 7.333333 14  14 11.33333
## synthetic 35.33333 6.000000 0 0.0000000 4.000000 5.333333  8  12 11.33333
##                5.5        6      6.5
## observed  10.66667 3.333333 2.666667
## synthetic 11.33333 2.666667 4.000000
##
## $Petal.Width
##                0.1      0.2      0.3       0.4       0.5 0.6 0.7 0.8
## observed  22.66667 4.666667 4.666667 0.6666667 0.6666667   0   0   0
## synthetic 33.33333 4.000000 2.666667 0.0000000 1.3333333   0   0   0
##                0.9 1       1.1      1.2      1.3      1.4       1.5
## observed  4.666667 2 3.3333333 8.666667 5.333333 8.000000 2.6666667
## synthetic 7.333333 0 0.6666667 4.000000 3.333333 5.333333 0.6666667
##                1.6      1.7      1.8 1.9        2      2.1      2.2
## observed  1.333333 8.000000 3.333333   4 4.000000 2.000000 5.333333
## synthetic 4.000000 6.666667 4.000000   4 3.333333 2.666667 8.666667
##                2.3       2.4
## observed  2.000000 2.0000000
## synthetic 3.333333 0.6666667
```

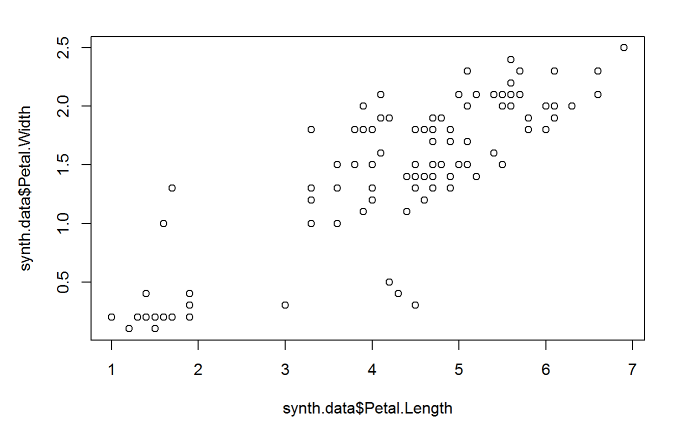


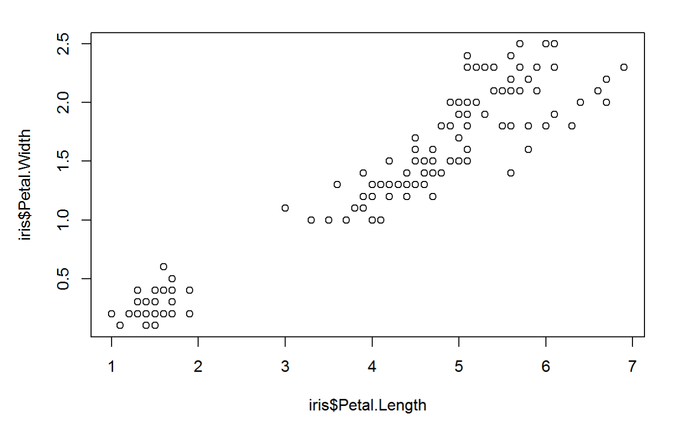

```
## Press return for next plot:
## $Species
##             setosa versicolor virginica
## observed  33.33333   33.33333  33.33333
## synthetic 41.33333   25.33333  33.33333
```



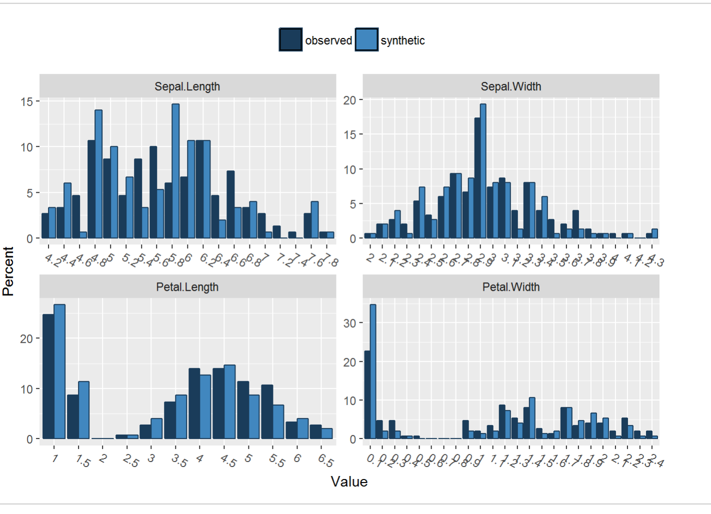
Use the Synthpop to model and synthese the dataset. Load the required dataset

```
## Element(s): {6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30,
31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58,
59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86,
87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111,
112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 1
34, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150} of the 'visit.sequence' rem
oved as not valid. No such column.
##
## syn  variables
## 1    Sepal.Length Sepal.Width Petal.Length Petal.Width Species
```

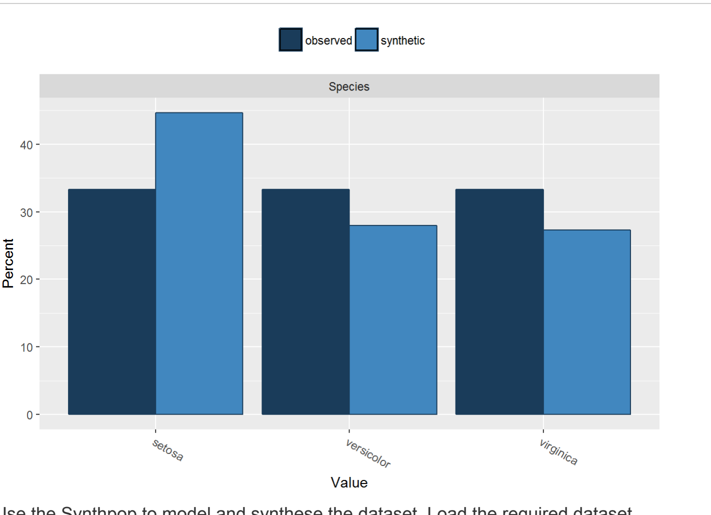





```
##
## Comparing percentages observed with synthetic
##
## $Sepal.Length
##                4.2      4.4       4.6      4.8         5      5.2      5.4
## observed  2.666667 3.333333 4.6666667 10.66667  8.666667 4.666667 8.666667
## synthetic 3.333333 6.000000 0.6666667 14.00000 10.000000 6.666667 3.333333
##                 5.6      5.8         6      6.2      6.4      6.6      6.8
## observed  10.000000  6.00000  6.666667 10.66667 4.666667 7.333333 3.333333
## synthetic  5.333333 14.66667 10.666667 10.66667 2.000000 3.333333 4.000000
##                   7      7.2       7.4      7.6       7.8
## observed  2.6666667 1.333333 0.6666667 2.666667 0.6666667
## synthetic 0.6666667 0.000000 0.0000000 4.000000 0.6666667
##
## $Sepal.Width
##                   2 2.1      2.2       2.3      2.4      2.5      2.6
## observed  0.6666667   2 2.666667 2.0000000 5.333333 3.333333 6.000000
## synthetic 0.6666667   2 4.000000 0.6666667 7.333333 2.666667 7.333333
##                2.7      2.8      2.9        3      3.1      3.2 3.3 3.4
## observed  9.333333 6.666667 17.33333 7.333333 8.666667 4.000000   8   4
## synthetic 9.333333 8.666667 19.33333 8.000000 8.000000 1.333333   8   6
##                 3.5      3.6      3.7       3.8       3.9         4
## observed  2.6666667 2.000000 4.000000 1.3333333 0.6666667 0.6666667
## synthetic 0.6666667 1.333333 1.333333 0.6666667 0.6666667 0.0000000
##                 4.1 4.2       4.3
## observed  0.6666667   0 0.6666667
## synthetic 0.6666667   0 1.3333333
##
## $Petal.Length
##                  1       1.5 2       2.5        3      3.5        4
## observed  24.66667  8.666667 0 0.6666667 2.666667 7.333333 14.00000
## synthetic 26.66667 11.333333 0 0.6666667 4.000000 8.666667 12.66667
##                4.5         5       5.5        6      6.5
## observed  14.00000 11.333333 10.666667 3.333333 2.666667
## synthetic 14.66667  8.666667  6.666667 4.000000 2.000000
##
## $Petal.Width
##                0.1      0.2      0.3       0.4       0.5 0.6 0.7 0.8
## observed  22.66667 4.666667 4.666667 0.6666667 0.6666667   0   0   0
## synthetic 34.66667 2.000000 2.000000 0.6666667 0.0000000   0   0   0
##                0.9        1      1.1      1.2      1.3      1.4      1.5
## observed  4.666667 2.000000 3.333333 8.666667 5.333333  8.00000 2.666667
## synthetic 2.000000 1.333333 2.000000 7.333333 4.000000 10.66667 1.333333
##                1.6 1.7      1.8      1.9        2       2.1      2.2
## observed  1.333333   8 3.333333 4.000000 4.000000 2.0000000 5.333333
## synthetic 2.000000   8 4.666667 6.666667 5.333333 0.6666667 3.333333
##                 2.3       2.4
## observed  2.0000000 2.0000000
## synthetic 0.6666667 0.6666667
```

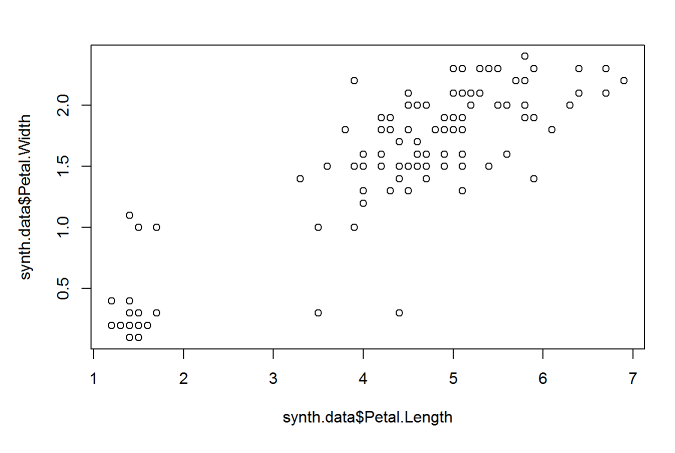


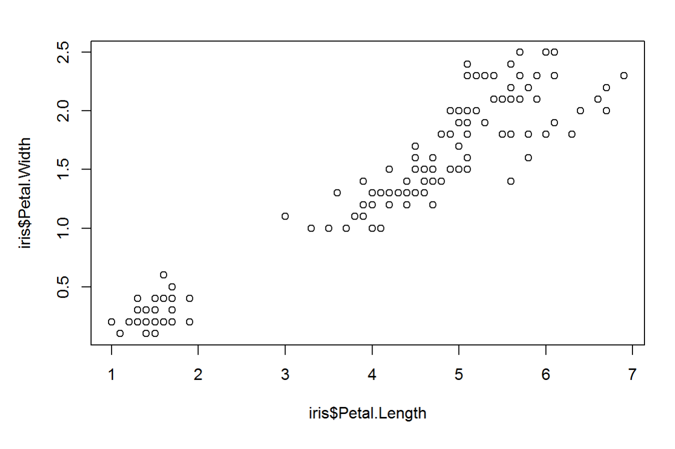

```
## Press return for next plot:
## $Species
##             setosa versicolor virginica
## observed  33.33333   33.33333  33.33333
## synthetic 44.66667   28.00000  27.33333
```



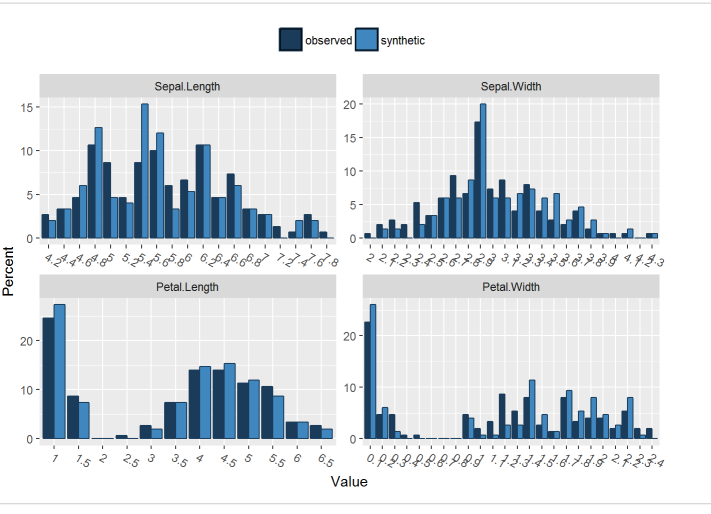
Use the Synthpop to model and synthese the dataset. Load the required dataset

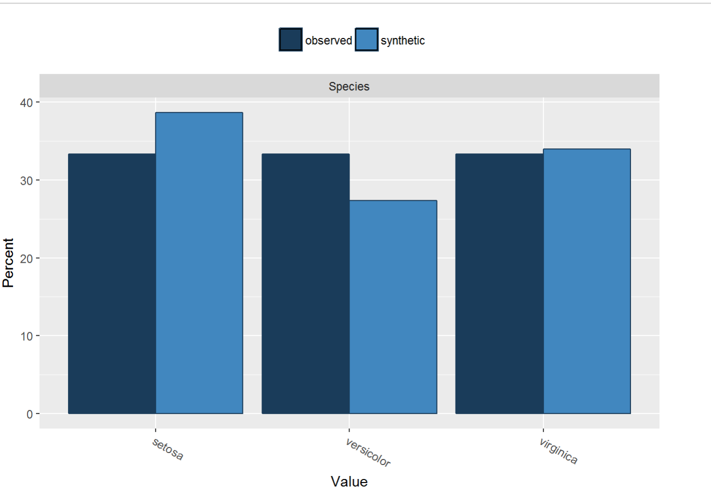
```
## Element(s): {6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30,
31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58,
59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86,
87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111,
112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 1
34, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150} of the 'visit.sequence' rem
oved as not valid. No such column.
##
## syn  variables
## 1    Sepal.Length Sepal.Width Petal.Length Petal.Width Species
```






```
##
## Comparing percentages observed with synthetic
##
## $Sepal.Length
##                4.2      4.4      4.6      4.8        5      5.2       5.4
## observed  2.666667 3.333333 4.666667 10.66667 8.666667 4.666667  8.666667
## synthetic 2.000000 3.333333 6.000000 12.66667 4.666667 4.000000 15.333333
##           5.6      5.8        6      6.2      6.4      6.6      6.8
## observed   10 6.000000 6.666667 10.66667 4.666667 7.333333 3.333333
## synthetic  12 3.333333 5.333333 10.66667 4.666667 6.000000 3.333333
##                  7      7.2       7.4      7.6       7.8
## observed  2.666667 1.333333 0.6666667 2.666667 0.6666667
## synthetic 2.666667 0.000000 2.0000000 2.000000 0.0000000
##
## $Sepal.Width
##                   2      2.1      2.2 2.3      2.4      2.5 2.6      2.7
## observed  0.6666667 2.000000 2.666667   2 5.333333 3.333333   6 9.333333
## synthetic 0.0000000 1.333333 1.333333   0 2.000000 3.333333   6 6.000000
##                2.8      2.9        3      3.1      3.2      3.3 3.4
## observed  6.666667 17.33333 7.333333 8.666667 4.000000 8.000000   4
## synthetic 8.666667 20.00000 6.000000 6.000000 6.666667 7.333333   6
##                3.5      3.6      3.7      3.8       3.9         4
## observed  2.666667 2.000000 4.000000 1.333333 0.6666667 0.6666667
## synthetic 6.666667 2.666667 4.666667 2.666667 0.6666667 0.0000000
##                 4.1 4.2       4.3
## observed  0.6666667   0 0.6666667
## synthetic 1.3333333   0 0.6666667
##
## $Petal.Length
##                  1      1.5 2       2.5        3      3.5        4
## observed  24.66667 8.666667 0 0.6666667 2.666667 7.333333 14.00000
## synthetic 27.33333 7.333333 0 0.0000000 2.000000 7.333333 14.66667
##                4.5        5       5.5        6      6.5
## observed  14.00000 11.33333 10.666667 3.333333 2.666667
## synthetic 15.33333 12.00000  8.666667 3.333333 2.000000
##
## $Petal.Width
##                0.1      0.2      0.3       0.4       0.5 0.6 0.7 0.8
## observed  22.66667 4.666667 4.666667 0.6666667 0.6666667   0   0   0
## synthetic 26.00000 6.000000 1.333333 0.0000000 0.0000000   0   0   0
##                0.9         1       1.1      1.2      1.3      1.4      1.5
## observed  4.666667 2.0000000 3.3333333 8.666667 5.333333  8.00000 2.666667
## synthetic 4.000000 0.6666667 0.6666667 2.666667 2.666667 11.33333 4.666667
##                1.6      1.7      1.8 1.9        2      2.1      2.2
## observed  1.333333 8.000000 3.333333   4 4.000000 2.000000 5.333333
## synthetic 1.333333 9.333333 5.333333   8 4.666667 2.666667 8.000000
##                 2.3 2.4
## observed  2.0000000   2
## synthetic 0.6666667   0
```




```
## Press return for next plot:
## $Species
##             setosa versicolor virginica
## observed  33.33333   33.33333  33.33333
## synthetic 38.66667   27.33333  34.00000
```