# M6_L1_RomilShah

Romil Shah

July 7, 2016

## Read Data and additional packages

```
require(ggplot2)
```

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

```
## Warning: package 'ggplot2' was built under R version 3.2.5
```

```
require(class)
```

```
## Loading required package: class
```

```
## Warning: package 'class' was built under R version 3.2.5
```

```
library(class)
```

```
data_url <- 'https://archive.ics.uci.edu/ml/machine-learning-databases/pima-indians-diabetes/pima-indians-diabetes.data'
dataframe <- read.csv(url(data_url))
colnames(dataframe) <-
c("Npreg","plasmaGluc","bloodPress","skinFold","insulin","bmi","diabetes","age","class")
pima <- dataframe[1:8]
head(pima)
```

```
##   Npreg plasmaGluc bloodPress skinFold insulin  bmi diabetes age
## 1     1         85         66       29       0 26.6    0.351  31
## 2     8        183         64        0       0 23.3    0.672  32
## 3     1         89         66       23      94 28.1    0.167  21
## 4     0        137         40       35     168 43.1    2.288  33
## 5     5        116         74        0       0 25.6    0.201  30
## 6     3         78         50       32      88 31.0    0.248  26
```

```
length(pima)
```

```
## [1] 8
```

```
names(pima)
```

```
## [1] "Npreg"      "plasmaGluc" "bloodPress" "skinFold"   "insulin"
## [6] "bmi"        "diabetes"   "age"
```

```
table(pima$Npreg)
```

```
##
##   0   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  17
## 111 135 103  75  68  57  49  45  38  28  24  11   9  10   2   1   1
```

```
pima$diabetes
```

```
##   [1] 0.351 0.672 0.167 2.288 0.201 0.248 0.134 0.158 0.232 0.191 0.537
##  [12] 1.441 0.398 0.587 0.484 0.551 0.254 0.183 0.529 0.704 0.388 0.451
##  [23] 0.263 0.254 0.205 0.257 0.487 0.245 0.337 0.546 0.851 0.267 0.188
##  [34] 0.512 0.966 0.420 0.665 0.503 1.390 0.271 0.696 0.235 0.721 0.294
##  [45] 1.893 0.564 0.586 0.344 0.305 0.491 0.526 0.342 0.467 0.718 0.248
##  [56] 0.254 0.962 1.781 0.173 0.304 0.270 0.587 0.699 0.258 0.203 0.855
##  [67] 0.845 0.334 0.189 0.867 0.411 0.583 0.231 0.396 0.140 0.391 0.370
##  [78] 0.270 0.307 0.140 0.102 0.767 0.237 0.227 0.698 0.178 0.324 0.153
##  [89] 0.165 0.258 0.443 0.261 0.277 0.761 0.255 0.130 0.323 0.356 0.325
## [100] 1.222 0.179 0.262 0.283 0.930 0.801 0.207 0.287 0.336 0.247 0.199
## [111] 0.543 0.192 0.391 0.588 0.539 0.220 0.654 0.443 0.223 0.759 0.260
## [122] 0.404 0.186 0.278 0.496 0.452 0.261 0.403 0.741 0.361 1.114 0.356
## [133] 0.457 0.647 0.088 0.597 0.532 0.703 0.159 0.268 0.286 0.318 0.272
## [144] 0.237 0.572 0.096 1.400 0.218 0.085 0.399 0.432 1.189 0.687 0.137
## [155] 0.337 0.637 0.833 0.229 0.817 0.294 0.204 0.167 0.368 0.743 0.722
## [166] 0.256 0.709 0.471 0.495 0.180 0.542 0.773 0.678 0.370 0.719 0.382
## [177] 0.319 0.190 0.956 0.084 0.725 0.299 0.268 0.244 0.745 0.615 1.321
## [188] 0.640 0.361 0.142 0.374 0.383 0.578 0.136 0.395 0.187 0.678 0.905
## [199] 0.150 0.874 0.236 0.787 0.235 0.324 0.407 0.605 0.151 0.289 0.355
## [210] 0.290 0.375 0.164 0.431 0.260 0.742 0.514 0.464 1.224 0.261 1.072
## [221] 0.805 0.209 0.687 0.666 0.101 0.198 0.652 2.329 0.089 0.645 0.238
## [232] 0.583 0.394 0.293 0.479 0.586 0.686 0.831 0.582 0.192 0.446 0.402
## [243] 1.318 0.329 1.213 0.258 0.427 0.282 0.143 0.380 0.284 0.249 0.238
## [254] 0.926 0.543 0.557 0.092 0.655 1.353 0.299 0.761 0.612 0.200 0.226
## [265] 0.997 0.933 1.101 0.078 0.240 1.136 0.128 0.254 0.422 0.251 0.677
## [276] 0.296 0.454 0.744 0.881 0.334 0.280 0.262 0.165 0.259 0.647 0.619
## [287] 0.808 0.340 0.263 0.434 0.757 1.224 0.613 0.254 0.692 0.337 0.520
## [298] 0.412 0.840 0.839 0.422 0.156 0.209 0.207 0.215 0.326 0.143 1.391
## [309] 0.875 0.313 0.605 0.433 0.626 1.127 0.315 0.284 0.345 0.150 0.129
## [320] 0.527 0.197 0.254 0.731 0.148 0.123 0.692 0.200 0.127 0.122 1.476
## [331] 0.166 0.282 0.137 0.260 0.259 0.932 0.343 0.893 0.331 0.472 0.673
## [342] 0.389 0.290 0.485 0.349 0.654 0.187 0.279 0.346 0.237 0.252 0.243
## [353] 0.580 0.559 0.302 0.962 0.569 0.378 0.875 0.583 0.207 0.305 0.520
## [364] 0.385 0.499 0.368 0.252 0.306 0.234 2.137 1.731 0.545 0.225 0.816
## [375] 0.528 0.299 0.509 0.238 1.021 0.821 0.236 0.947 1.268 0.221 0.205
## [386] 0.660 0.239 0.452 0.949 0.444 0.340 0.389 0.463 0.803 1.600 0.944
## [397] 0.196 0.389 0.241 0.161 0.151 0.286 0.280 0.135 0.520 0.376 0.336
## [408] 1.191 0.702 0.674 0.528 1.076 0.256 0.534 0.258 1.095 0.554 0.624
## [419] 0.219 0.507 0.561 0.496 0.421 0.516 0.264 0.256 0.328 0.284 0.233
## [430] 0.108 0.551 0.527 0.167 1.138 0.205 0.244 0.434 0.147 0.727 0.435
## [441] 0.497 0.230 0.955 0.380 2.420 0.658 0.330 0.510 0.285 0.415 0.542
## [452] 0.381 0.832 0.498 0.212 0.687 0.364 1.001 0.460 0.733 0.416 0.705
## [463] 0.258 1.022 0.452 0.269 0.600 0.183 0.571 0.607 0.170 0.259 0.210
## [474] 0.126 0.231 0.711 0.466 0.162 0.419 0.344 0.197 0.306 0.233 0.630
```

```
## [485] 0.365 0.536 1.159 0.294 0.551 0.629 0.292 0.145 1.144 0.174 0.304
## [496] 0.292 0.547 0.163 0.839 0.313 0.267 0.727 0.738 0.238 0.263 0.314
## [507] 0.692 0.968 0.409 0.297 0.207 0.200 0.525 0.154 0.268 0.771 0.304
## [518] 0.180 0.582 0.187 0.305 0.189 0.652 0.151 0.444 0.299 0.107 0.493
## [529] 0.660 0.717 0.686 0.917 0.501 1.251 0.302 0.197 0.735 0.804 0.968
## [540] 0.661 0.549 0.825 0.159 0.365 0.423 1.034 0.160 0.341 0.680 0.204
## [551] 0.591 0.247 0.422 0.471 0.161 0.218 0.237 0.126 0.300 0.121 0.502
## [562] 0.401 0.497 0.601 0.748 0.412 0.085 0.338 0.203 0.270 0.268 0.430
## [573] 0.198 0.892 0.280 0.813 0.693 0.245 0.575 0.371 0.206 0.259 0.190
## [584] 0.687 0.417 0.129 0.249 1.154 0.342 0.925 0.175 0.402 1.699 0.733
## [595] 0.682 0.194 0.559 0.088 0.407 0.400 0.190 0.100 0.692 0.212 0.514
## [606] 1.258 0.482 0.270 0.138 0.292 0.593 0.787 0.878 0.557 0.207 0.157
## [617] 0.257 1.282 0.141 0.246 1.698 1.461 0.347 0.158 0.362 0.206 0.393
## [628] 0.144 0.148 0.732 0.238 0.343 0.115 0.167 0.465 0.153 0.649 0.871
## [639] 0.149 0.695 0.303 0.178 0.610 0.730 0.134 0.447 0.455 0.260 0.133
## [650] 0.234 0.466 0.269 0.455 0.142 0.240 0.155 1.162 0.190 1.292 0.182
## [661] 1.394 0.165 0.637 0.245 0.217 0.235 0.141 0.430 0.164 0.631 0.551
## [672] 0.285 0.880 0.587 0.328 0.230 0.263 0.127 0.614 0.332 0.364 0.366
## [683] 0.536 0.640 0.591 0.314 0.181 0.828 0.335 0.856 0.257 0.886 0.439
## [694] 0.191 0.128 0.268 0.253 0.598 0.904 0.483 0.565 0.905 0.304 0.118
## [705] 0.177 0.261 0.176 0.148 0.674 0.295 0.439 0.441 0.352 0.121 0.826
## [716] 0.970 0.595 0.415 0.378 0.317 0.289 0.349 0.251 0.265 0.236 0.496
## [727] 0.433 0.326 0.141 0.323 0.259 0.646 0.426 0.560 0.284 0.515 0.600
## [738] 0.453 0.293 0.785 0.400 0.219 0.734 1.174 0.488 0.358 1.096 0.408
## [749] 0.178 1.182 0.261 0.223 0.222 0.443 1.057 0.391 0.258 0.197 0.278
## [760] 0.766 0.403 0.142 0.171 0.340 0.245 0.349 0.315
```

```
length(pima$Npreg)
```

```
## [1] 767
```

```
shuff <- runif(nrow(pima))
shuff
```

```
##   [1] 0.270674973 0.177951238 0.499809612 0.382276926 0.633782021
##   [6] 0.326668291 0.519428304 0.768980057 0.106852074 0.235425394
##  [11] 0.025432956 0.687705788 0.140341183 0.388677243 0.635836062
##  [16] 0.230557214 0.722230561 0.306075202 0.923716315 0.856244016
##  [21] 0.568131980 0.866280630 0.523177084 0.711119027 0.074558230
##  [26] 0.283650367 0.090685589 0.861134922 0.241146440 0.334849869
##  [31] 0.833224640 0.656017282 0.533237723 0.318739373 0.303482477
##  [36] 0.463509446 0.972989531 0.380257929 0.159432734 0.423768532
##  [41] 0.649297693 0.608438362 0.986025753 0.976432750 0.460770146
##  [46] 0.444826883 0.507094074 0.248609774 0.671746915 0.792892551
##  [51] 0.882687623 0.387333453 0.559083603 0.834202574 0.894376110
##  [56] 0.949239378 0.050745765 0.531493941 0.430598240 0.704845800
##  [61] 0.948392445 0.219795028 0.249843859 0.530757722 0.571075106
##  [66] 0.998015563 0.078791271 0.484607931 0.249646279 0.878652835
##  [71] 0.799077179 0.958609562 0.323852862 0.430929531 0.585406687
##  [76] 0.031969451 0.483772055 0.067798898 0.376883593 0.534235068
##  [81] 0.353010132 0.385505064 0.335316797 0.750399383 0.260854961
```

```
##  [86] 0.856415456 0.438047806 0.257605416 0.860515407 0.611912232
##  [91] 0.367325513 0.396970046 0.440231345 0.661441372 0.107771740
##  [96] 0.020428465 0.947860308 0.030781860 0.721292696 0.219014893
## [101] 0.275610993 0.277926044 0.763338278 0.750351384 0.353389408
## [106] 0.793221171 0.134114744 0.646873533 0.977283159 0.603391608
## [111] 0.542166549 0.637849282 0.825668881 0.732366951 0.731733648
## [116] 0.354684404 0.587818170 0.891687555 0.732893723 0.087911303
## [121] 0.778486799 0.554570421 0.949259529 0.928649618 0.426984912
## [126] 0.218105703 0.575483265 0.318784588 0.585888829 0.491293304
## [131] 0.037215400 0.180285048 0.144594142 0.895128918 0.496330530
## [136] 0.370831705 0.421616682 0.622692797 0.851072178 0.103879524
## [141] 0.909342947 0.409263829 0.437109719 0.237871907 0.286128460
## [146] 0.097889706 0.511814416 0.729043355 0.372284607 0.987162058
## [151] 0.641111029 0.219387798 0.880890076 0.378189020 0.017999123
## [156] 0.641131931 0.374697477 0.672112365 0.048281288 0.443937021
## [161] 0.467199357 0.089379506 0.364410518 0.832393534 0.319301159
## [166] 0.763694894 0.055187946 0.266376664 0.651731262 0.375362552
## [171] 0.864797767 0.841226000 0.342968574 0.401315324 0.782486306
## [176] 0.183238571 0.366238335 0.356197282 0.490243615 0.414070140
## [181] 0.737399935 0.825461418 0.875249160 0.763121339 0.562802617
## [186] 0.756578975 0.135156490 0.142422278 0.429882020 0.998193625
## [191] 0.686635610 0.973896974 0.455974906 0.579641059 0.854988286
## [196] 0.360982676 0.436232700 0.288533632 0.731702669 0.323023859
## [201] 0.023247084 0.830815928 0.422003824 0.933824959 0.804642475
## [206] 0.622457856 0.259662974 0.897570910 0.981462857 0.346884468
## [211] 0.273920765 0.264125727 0.818040791 0.090322453 0.987728896
## [216] 0.483744354 0.663001740 0.954704805 0.982599620 0.370113314
## [221] 0.351974638 0.028164590 0.825277753 0.297850498 0.218710567
## [226] 0.447002940 0.588508872 0.088809839 0.722923890 0.082375093
## [231] 0.543841367 0.506987900 0.382874427 0.275163011 0.851078733
## [236] 0.486885102 0.293069860 0.863461038 0.752517517 0.205474709
## [241] 0.965056503 0.118229863 0.059686696 0.233343639 0.312570449
## [246] 0.023839632 0.025047287 0.541962240 0.372281019 0.960839339
## [251] 0.853509359 0.588966531 0.586635466 0.189274601 0.460653288
## [256] 0.519271721 0.219171493 0.159996428 0.344543118 0.286944742
## [261] 0.919426024 0.893628073 0.179568283 0.074985379 0.459271111
## [266] 0.599660448 0.452320591 0.534843368 0.336016537 0.108969336
## [271] 0.002528643 0.493929912 0.064152717 0.725871279 0.743151726
## [276] 0.959807714 0.778800802 0.901328389 0.108274765 0.223303023
## [281] 0.155291500 0.267206452 0.368726659 0.365868954 0.030131972
## [286] 0.518710793 0.980456839 0.352980260 0.203681321 0.371393276
## [291] 0.640070276 0.082687737 0.489200881 0.265623467 0.899238213
## [296] 0.085442008 0.462996368 0.589141196 0.641498988 0.046769106
## [301] 0.199961977 0.742494137 0.674754295 0.634629409 0.178407559
## [306] 0.503224856 0.136367759 0.925382137 0.493957182 0.908272550
## [311] 0.255370980 0.082772127 0.439966182 0.899674069 0.875464407
## [316] 0.912499517 0.988886283 0.595213918 0.968740552 0.955259279
## [321] 0.190363795 0.967576629 0.909347964 0.718462338 0.904803471
## [326] 0.339576558 0.998848878 0.412509907 0.670974724 0.215495809
## [331] 0.686970876 0.818320565 0.014013665 0.803721442 0.889264057
```

```
## [336] 0.665100941 0.843704608 0.158857173 0.498549599 0.210866897
## [341] 0.462801459 0.980534928 0.333664184 0.217054545 0.220934652
## [346] 0.597623624 0.022290213 0.300313004 0.020133450 0.826528551
## [351] 0.812006671 0.011640979 0.605472320 0.957560137 0.337862505
## [356] 0.579775947 0.822036328 0.355806082 0.808081873 0.674665904
## [361] 0.376711447 0.867771583 0.183887082 0.237435219 0.963956327
## [366] 0.375170287 0.311083632 0.626440851 0.516440098 0.468826422
## [371] 0.264206921 0.421548418 0.041038045 0.080387948 0.426797864
## [376] 0.911067267 0.743225473 0.633563142 0.841823640 0.643828345
## [381] 0.518558334 0.329793501 0.548838827 0.670760981 0.461890625
## [386] 0.880028023 0.925184634 0.716372269 0.289069979 0.715799268
## [391] 0.706473493 0.921997962 0.771493991 0.328086992 0.088999926
## [396] 0.224631621 0.688877637 0.839945894 0.495639484 0.038849127
## [401] 0.861151811 0.840049658 0.119600981 0.911084764 0.554321170
## [406] 0.246728051 0.832714388 0.838361497 0.885161286 0.059279167
## [411] 0.523824133 0.406520542 0.949602298 0.971809302 0.010836672
## [416] 0.837820802 0.470415802 0.253135189 0.554002615 0.698237999
## [421] 0.988382705 0.663099222 0.730446751 0.373079120 0.129798417
## [426] 0.165266733 0.340198530 0.828703578 0.987399449 0.453425899
## [431] 0.491558523 0.907918191 0.521017098 0.966077595 0.559699062
## [436] 0.475428416 0.467287403 0.352499798 0.410150716 0.651437368
## [441] 0.209609460 0.134493174 0.739394485 0.358654697 0.931920222
## [446] 0.956804190 0.440569582 0.467026471 0.321022516 0.447193269
## [451] 0.400368120 0.206392426 0.688574494 0.606368739 0.464657934
## [456] 0.871352586 0.780032102 0.326336138 0.752807568 0.281578908
## [461] 0.840447326 0.462309491 0.887870681 0.104108577 0.601362747
## [466] 0.220254511 0.121286289 0.792397344 0.248049336 0.088820540
## [471] 0.753086073 0.734345580 0.102239018 0.338507035 0.297847428
## [476] 0.885203908 0.345684333 0.332530808 0.928222481 0.494724537
## [481] 0.782241985 0.999331812 0.134157675 0.929731841 0.149484087
## [486] 0.505937653 0.435445086 0.532656488 0.012111440 0.584609113
## [491] 0.209104112 0.198234994 0.103577056 0.404371321 0.683980264
## [496] 0.004231130 0.499158975 0.605663479 0.118380085 0.726739022
## [501] 0.588115835 0.986561796 0.284530697 0.219771018 0.131671715
## [506] 0.088534011 0.711384937 0.921489701 0.900286757 0.386626855
## [511] 0.951942104 0.050777514 0.006814915 0.008699968 0.391870557
## [516] 0.450989485 0.462714978 0.913462145 0.544098628 0.800430708
## [521] 0.368739144 0.634208471 0.807369689 0.760368623 0.982905303
## [526] 0.204512906 0.411167919 0.090057212 0.940223081 0.927100628
## [531] 0.918394159 0.393908898 0.583530485 0.507753937 0.694640200
## [536] 0.516996025 0.554435763 0.768786009 0.689175784 0.812058232
## [541] 0.853133051 0.295048989 0.939813332 0.424600971 0.157983920
## [546] 0.556142027 0.080593578 0.175695077 0.682855739 0.794175801
## [551] 0.722730709 0.710268325 0.215051728 0.513330555 0.532832441
## [556] 0.101794130 0.633183398 0.188854535 0.812133817 0.057735446
## [561] 0.378700252 0.198513043 0.328244017 0.212217483 0.305828776
## [566] 0.548387741 0.481235050 0.738875262 0.505896655 0.269450628
## [571] 0.277770656 0.270473904 0.085172922 0.701161620 0.946820496
## [576] 0.391379046 0.892809003 0.024175014 0.022543602 0.943920980
## [581] 0.818919676 0.369678697 0.658459911 0.997671628 0.439615517
```

```
## [586] 0.235829321 0.764107949 0.264939341 0.075140530 0.604137354
## [591] 0.756501241 0.971329555 0.430418518 0.923046902 0.805880693
## [596] 0.767274419 0.947075095 0.647937749 0.441377332 0.942995038
## [601] 0.246821713 0.535841689 0.794438082 0.066838652 0.929982645
## [606] 0.054625622 0.967911475 0.604888201 0.301564524 0.384728971
## [611] 0.150632010 0.051737066 0.040844050 0.926180224 0.166748435
## [616] 0.198278788 0.084355559 0.958136200 0.255359920 0.134014259
## [621] 0.048551660 0.216603061 0.596891514 0.981922208 0.588152379
## [626] 0.169908424 0.434684466 0.924926365 0.317577384 0.337097507
## [631] 0.498374466 0.811671766 0.920386641 0.056505230 0.308247832
## [636] 0.933919127 0.032579574 0.443336989 0.232457533 0.451160224
## [641] 0.877194010 0.230498137 0.890628624 0.471538314 0.037550909
## [646] 0.025297965 0.477925197 0.340192810 0.031476525 0.539991150
## [651] 0.548278350 0.235912025 0.032398006 0.776530538 0.369761863
## [656] 0.415439060 0.102255025 0.667183933 0.633610103 0.347180539
## [661] 0.540363113 0.099742538 0.983808873 0.482972481 0.580837161
## [666] 0.746373121 0.694248756 0.222764215 0.652860236 0.091571981
## [671] 0.113899986 0.859283415 0.835358828 0.642244019 0.684254018
## [676] 0.174834186 0.799206581 0.905039076 0.043495781 0.852537338
## [681] 0.384536258 0.444365759 0.914348981 0.230252693 0.933640281
## [686] 0.088220574 0.247331409 0.340644368 0.863747585 0.743806996
## [691] 0.933753373 0.809456720 0.238370359 0.947558406 0.919983970
## [696] 0.220845574 0.359508091 0.523226958 0.746728534 0.725286673
## [701] 0.938098685 0.884132923 0.312518165 0.268736780 0.381593128
## [706] 0.617963716 0.556035494 0.963128706 0.872842932 0.922076823
## [711] 0.985375277 0.651056516 0.749602308 0.901431165 0.508632452
## [716] 0.190887612 0.480363028 0.235964982 0.072762106 0.512802262
## [721] 0.354363891 0.379816553 0.281463252 0.321286829 0.616745618
## [726] 0.663093884 0.955266339 0.497973599 0.187315247 0.211401375
## [731] 0.072514750 0.110042139 0.092002085 0.215827771 0.620801940
## [736] 0.914946633 0.787461532 0.898676320 0.487236168 0.220312014
## [741] 0.167424328 0.514786095 0.524413092 0.551974465 0.007009527
## [746] 0.335031430 0.401502178 0.786021811 0.695986275 0.786592212
## [751] 0.287721761 0.242476191 0.657373350 0.874510155 0.113073478
## [756] 0.685324726 0.171129700 0.261375620 0.305686465 0.411909958
## [761] 0.959848570 0.349312279 0.429144432 0.293529131 0.931356327
## [766] 0.284001777 0.134778974

pimaID <- pima[order(shuff),]
pimaID$Npreg

##   [1]  2  5  2 12  3  3  3  8 12  7  5  2  3  2  1 10 10  0  1 10  7  7  6
##  [24]  0  7  2  2  9  2  4  6  2  2  0 17  2  0  9  7  1  4 10  6  6  6  1
##  [47]  4  0  8  5 10  4  0  2  2  4  4  2  2  2  2  2  0  3  0  4  1  2  0
##  [70]  0  9  1  6  2  9  8  1  7  1  4  3 10  8  6  2 10  2  1  1  3  6  9
##  [93]  0  4 10  2  4  0  4  1  1  0  1  8  8  0  3 10  8  9  4  1  0  3  3
## [116]  0  0  9  1  8  2  3  3  6  4  2 11 12  3  3  4  6  1  2  5  1  1  0
## [139]  2  2  1  3  0  1  8  2  6  8  3  1  1  2  9  3  5  0 11  3  8  6  0
## [162]  3  5  6  0  1  2  4  8  5  1  4  4  7  5  3  4  6  1  6  7  4  2  1
## [185]  0  0 15  5  2  1  7  0  3  0  4  7  4  3  3  1  0  3  1  2  0  5  9
```

```
## [208]  7  1  7  0  3  1  4  3  0  2 10  0  1  3  1  4  6  2  1 13  0  2  9
## [231]  4 10  1  6  0  1  0  4 10  3  4  6  1  8  5  5  1  0  2  7  9  4  1
## [254] 11  1  1  1 11  7  2 10  9  2  1  4  2  1  1  5 12  5  4  0  1  2  2
## [277]  0  4  7  3 12  2  0  0  0  1  2  8  1  6  6  5  2  8  0  1  2  6  0
## [300]  4  0  3  7 12  5  5  6  3  1  7  2  2  1  3  1  2  6  3  2  2  6  2
## [323]  0  0  2  3  1 12  1 10  5  2  0  1  0  0  3 10  2  1  3  4  0  1  7
## [346]  4  0  1  0  1  9  0  2  2 11  5  1  0  1  8  7  1  0 11 14  0  7  5
## [369]  3  4  3 12  0 10  6  6  5  5  1  7  1  1  5  4  3  3  2  3  3  2  2
## [392]  0  7  2  1 10  0  1  1  2  1  7  2  4  1  1  1  0  0  5  3 10  2  9
## [415]  4  1  9  7  0  4  7  6  3  0  1  1  1  9  8  6  6  1  1  1 13  3  2
## [438]  0  2  2  5  8  0  7  8  5  1  8  1  1  6  2  1  0  0  5  3  4  3  2
## [461] 14  3  0  1  0  0  3 11  0  1  7  2  7  1  4 10  4  8  0  3  8  4  3
## [484]  5  6  3  7  1  0  4  2  8  8  1  3  1  7 10  0  3  9  3  0  8  2  6
## [507]  1  0  0 11  1  6  7  2  5  5  4  6  6  7  9  2 10  2  0  3 10  4  6
## [530]  1  1  2  5  6 11  1  1  5  2  1  7  3  0  2 13  2  5  2  4  4  7  4
## [553]  0  0  4  8  5  2  1  8  4  4  0  2  5  0  9  0  2  8  3  4  1  3  6
## [576]  0  0  2  4  1  6  0  5  0  8  3  4  8  8  1  1  1  7  5  0  2  1  5
## [599]  0  9  1  2  2  4  8 11  0  1  6 13  7  1  4  4  0  0  0  0  3  7  3
## [622]  1  8  3  5  1  2  0  5  5  4  2  3  2  5  3 13 10  1 13  6  9  1  6
## [645]  7  5  1  2  8  5  2  4  2  5  1  1  1  1  2  5  0  4  4  2  4  1  2
## [668]  1  2  6  7  8  5  3  1  3  1  6  5 13  0  5  3 13  4  0  0  3  7  1
## [691]  2  1  3  6  1  5  8  0 11  2  4  0  0  1  5  0  2 13  6  5  6  4  0
## [714]  1  0  1  1  2  1  8  7  5  1  0  5  4  0  1  3  9 13  7  9  9  9  5
## [737]  4  1  0  1  6  3  0  9  7  7  0  1  1  7  2  5  3  9  5  9  6  1  1
## [760] 12  2  3  8  0  3 10  4
```

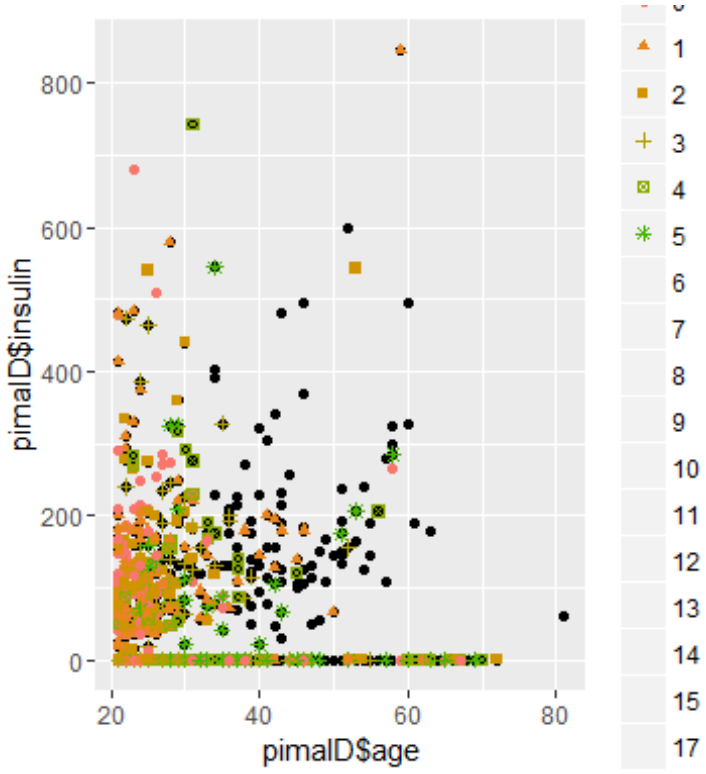
Plotting

```
qplot(pimaID$age,pimaID$insulin,data=pimaID)+geom_point(aes(colour=factor(pim
aID$Npreg),shape=factor(pimaID$Npreg)))
```

```
## Warning: The shape palette can deal with a maximum of 6 discrete values
## because more than 6 becomes difficult to discriminate; you have
## 17. Consider specifying shapes manually if you must have them.
```

```
## Warning: Removed 218 rows containing missing values (geom_point).
```

```
## Warning: The shape palette can deal with a maximum of 6 discrete values
## because more than 6 becomes difficult to discriminate; you have
## 17. Consider specifying shapes manually if you must have them.
```

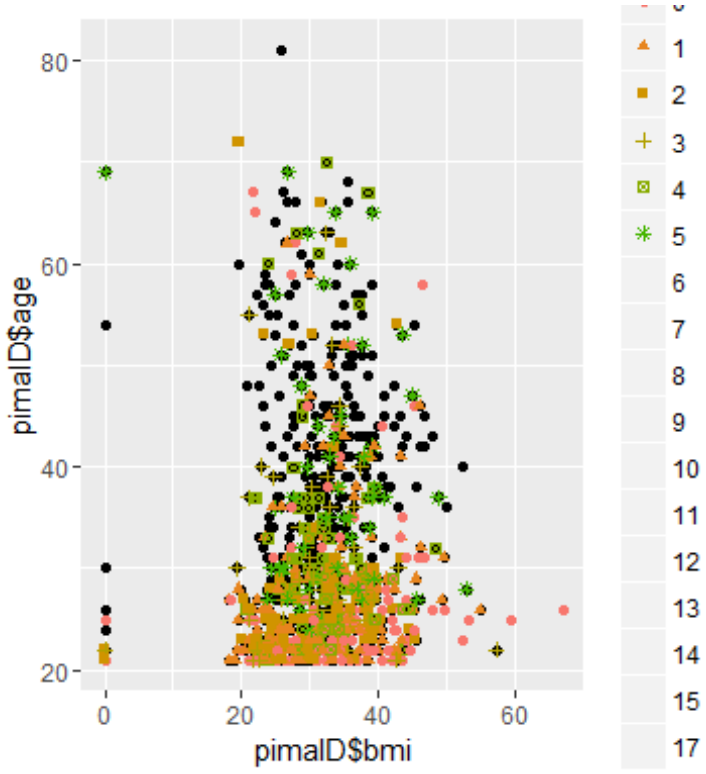



```
qplot(pimaID$bmi,pimaID$age,data=pimaID)+geom_point(aes(colour=factor(pimaID$
Npreg),shape=factor(pimaID$Npreg)))

## Warning: The shape palette can deal with a maximum of 6 discrete values
## because more than 6 becomes difficult to discriminate; you have
## 17. Consider specifying shapes manually if you must have them.

## Warning: Removed 218 rows containing missing values (geom_point).

## Warning: The shape palette can deal with a maximum of 6 discrete values
## because more than 6 becomes difficult to discriminate; you have
## 17. Consider specifying shapes manually if you must have them.
```




```
summary(pimaID)
```

```
##      Npreg          plasmaGluc      bloodPress       skinFold
##  Min.   : 0.000   Min.   :  0.0   Min.   :  0.0   Min.   : 0.00
##  1st Qu.: 1.000   1st Qu.: 99.0   1st Qu.: 62.0   1st Qu.: 0.00
##  Median : 3.000   Median :117.0   Median : 72.0   Median :23.00
##  Mean   : 3.842   Mean   :120.9   Mean   : 69.1   Mean   :20.52
##  3rd Qu.: 6.000   3rd Qu.:140.0   3rd Qu.: 80.0   3rd Qu.:32.00
##  Max.   :17.000   Max.   :199.0   Max.   :122.0   Max.   :99.00
##     insulin           bmi           diabetes           age
##  Min.   :  0.0   Min.   : 0.00   Min.   :0.0780   Min.   :21.00
##  1st Qu.:  0.0   1st Qu.:27.30   1st Qu.:0.2435   1st Qu.:24.00
##  Median : 32.0   Median :32.00   Median :0.3710   Median :29.00
##  Mean   : 79.9   Mean   :31.99   Mean   :0.4717   Mean   :33.22
##  3rd Qu.:127.5   3rd Qu.:36.60   3rd Qu.:0.6250   3rd Qu.:41.00
##  Max.   :846.0   Max.   :67.10   Max.   :2.4200   Max.   :81.00
```

```
pimaID.scaled <- as.data.frame(lapply(pimaID[,c(2:8)], scale))
head(pimaID.scaled)
```

```
##   plasmaGluc  bloodPress    skinFold     insulin        bmi   diabetes
## 1 -0.4021203 -0.36666864  0.71971646 -0.20734626 -0.8607433 -1.0367324
## 2 -0.3395782 -0.05688177 -1.28604269 -0.69310694 -0.7593375 -0.5420075
## 3 -0.9337280 -0.36666864 -1.28604269 -0.69310694 -0.5945530  0.1608639
## 4 -0.6522886  0.76921655  0.78239644  0.21769434 -0.2523082  0.0492491
## 5 -0.6835597 -0.77971780 -0.09512319  0.05288268 -0.8100404 -0.9583004
## 6  1.6304974  0.76921655  0.78239644  3.41851028  0.4702085 -0.6445724
```

```
##          age
## 1 -1.0397148
## 2 -0.2739069
## 3 -0.9546250
## 4  1.0875292
## 5 -0.7844455
## 6 -0.9546250
```

```
summary(pimaID.scaled)
```

```
##    plasmaGluc        bloodPress         skinFold          insulin
##  Min.   :-3.7794   Min.   :-3.5678   Min.   :-1.2860   Min.   :-0.6931
##  1st Qu.:-0.6836   1st Qu.:-0.3667   1st Qu.:-1.2860   1st Qu.:-0.6931
##  Median :-0.1207   Median : 0.1496   Median : 0.1556   Median :-0.4155
##  Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.0000   Mean   : 0.0000
##  3rd Qu.: 0.5986   3rd Qu.: 0.5627   3rd Qu.: 0.7197   3rd Qu.: 0.4129
##  Max.   : 2.4435   Max.   : 2.7312   Max.   : 4.9193   Max.   : 6.6453
##       bmi               diabetes            age
##  Min.   :-4.055028   Min.   :-1.1876   Min.   :-1.0397
##  1st Qu.:-0.594553   1st Qu.:-0.6883   1st Qu.:-0.7844
##  Median : 0.001206   Median :-0.3037   Median :-0.3590
##  Mean   : 0.000000   Mean   : 0.0000   Mean   : 0.0000
##  3rd Qu.: 0.584290   3rd Qu.: 0.4625   3rd Qu.: 0.6621
##  Max.   : 4.450388   Max.   : 5.8773   Max.   : 4.0657
```

```
normalize <- function (x) {
  return((x-min(x))/(max(x)-min(x)))
}

pimaID.normalized <- as.data.frame(lapply(pimaID[,c(2:8)],normalize))
head(pimaID.normalized)
```

```
##   plasmaGluc bloodPress  skinFold    insulin       bmi   diabetes
## 1  0.5427136  0.5081967 0.3232323 0.06619385 0.3755589 0.02134927
## 2  0.5527638  0.5573770 0.0000000 0.00000000 0.3874814 0.09137489
## 3  0.4572864  0.5081967 0.0000000 0.00000000 0.4068554 0.19086251
## 4  0.5025126  0.6885246 0.3333333 0.12411348 0.4470939 0.17506405
## 5  0.4974874  0.4426230 0.1919192 0.10165485 0.3815201 0.03245090
## 6  0.8693467  0.6885246 0.3333333 0.56028369 0.5320417 0.07685739
##          age
## 1 0.00000000
## 2 0.15000000
## 3 0.01666667
## 4 0.41666667
## 5 0.05000000
## 6 0.01666667
```

```
summary(pimaID.normalized)
```

```
##    plasmaGluc       bloodPress        skinFold         insulin
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.00000
##  1st Qu.:0.4975   1st Qu.:0.5082   1st Qu.:0.0000   1st Qu.:0.00000
```

```
##  Median :0.5879   Median :0.5902   Median :0.2323   Median :0.03783
##  Mean   :0.6073   Mean   :0.5664   Mean   :0.2072   Mean   :0.09445
##  3rd Qu.:0.7035   3rd Qu.:0.6557   3rd Qu.:0.3232   3rd Qu.:0.15071
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.00000
##       bmi            diabetes            age
##  Min.   :0.0000   Min.   :0.00000   Min.   :0.0000
##  1st Qu.:0.4069   1st Qu.:0.07067   1st Qu.:0.0500
##  Median :0.4769   Median :0.12511   Median :0.1333
##  Mean   :0.4768   Mean   :0.16809   Mean   :0.2037
##  3rd Qu.:0.5455   3rd Qu.:0.23356   3rd Qu.:0.3333
##  Max.   :1.0000   Max.   :1.00000   Max.   :1.0000
```

```
nrow(pimaID)
```

```
## [1] 767
```

```
pimaID.normalized.train <- pimaID.normalized[1:600,]
pimaID.normalized.test <- pimaID.normalized[601:767,]
pimaID.normalized.train.target <- pimaID[1:600,c(1)]
pimaID.normalized.test.target <- pimaID[601:767,c(1)]
pimaID.normalized.test.target
```

```
##   [1]  1  2  2  4  8 11  0  1  6 13  7  1  4  4  0  0  0  0  3  7  3  1  8
##  [24]  3  5  1  2  0  5  5  4  2  3  2  5  3 13 10  1 13  6  9  1  6  7  5
##  [47]  1  2  8  5  2  4  2  5  1  1  1  1  2  5  0  4  4  2  4  1  2  1  2
##  [70]  6  7  8  5  3  1  3  1  6  5 13  0  5  3 13  4  0  0  3  7  1  2  1
##  [93]  3  6  1  5  8  0 11  2  4  0  0  1  5  0  2 13  6  5  6  4  0  1  0
## [116]  1  1  2  1  8  7  5  1  0  5  4  0  1  3  9 13  7  9  9  9  5  4  1
## [139]  0  1  6  3  0  9  7  7  0  1  1  7  2  5  3  9  5  9  6  1  1 12  2
## [162]  3  8  0  3 10  4
```

```
#(1) K = 3
k<-3
knn.m1 <- class::knn(train=pimaID.normalized.train,
test=pimaID.normalized.test, pimaID.normalized.train.target,k)
knn.m1
```

```
##   [1] 5  0  1  5  1  3  2  0  0  2  7  5  2  1  2  1  7  3  9  7  1  3  9
##  [24] 6  1  1  6  1  7  0  0  2  1  2  2  0  4  2  2  12 6  4  7  1  7  10
##  [47] 4  0  12 10 1  0  3  6  1  1  7  0  8  9  0  6  2  4  0  2  1  1  2
##  [70] 2  10 6  2  7  3  3  2  12 3  4  3  4  2  3  0  5  4  0  3  3  2  2
##  [93] 4  2  2  7  8  2  7  1  0  0  0  1  1  0  1  5  8  12 10 0  4  3  1
## [116] 1  1  6  0  5  4  6  3  0  3  4  2  1  0  5  5  2  12 10 9  1  2  8
## [139] 4  4  6  9  2  12 11 10 0  0  5  0  3  10 1  17 10 5  8  2  9  9  1
## [162] 5  8  7  0  6  1
## Levels: 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 17
```

```
length(knn.m1)
```

```
## [1] 167
```

```
cm1 <- table(pimaID.normalized.test.target,knn.m1)
msetrain1<-mean((as.numeric(knn.m1)-
as.numeric(pimaID.normalized.train.target)))^2
```

```
## Warning in as.numeric(knn.m1) -
as.numeric(pimaID.normalized.train.target):
## longer object length is not a multiple of shorter object length
```

```
msetest1<-mean((as.numeric(knn.m1)-
as.numeric(pimaID.normalized.test.target)))^2
```

```
#(2) K = 5
k<-5
knn.m2 <-  class::knn(train=pimaID.normalized.train,
test=pimaID.normalized.test, pimaID.normalized.train.target,k)
knn.m2
```

```
##   [1] 5  0  3  5  7  6  2  0  0  0  7  2  2  0  1  1  7  3  3  7  1  3  1
##  [24] 1  4  1  6  1  10 0  4  2  1  5  4  0  6  9  2  10 6  4  8  1  1  5
##  [47] 0  0  1  2  2  0  0  6  1  1  9  0  0  4  0  6  2  0  0  1  1  1  2
##  [70] 2  11 0  8  6  3  4  2  12 4  4  2  6  2  10 0  2  4  0  3  4  2  2
##  [93] 4  2  1  7  3  2  7  1  5  0  0  0  3  0  0  5  14 13 7  1  4  2  1
## [116] 1  2  1  1  11 3  6  1  0  3  2  0  1  0  9  7  1  7  10 7  1  5  8
## [139] 10 1  10 5  2  12 10 10 1  0  2  0  3  10 1  8  10 8  8  2  1  6  1
## [162] 5  8  6  0  8  2
## Levels: 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 17
```

```
length(knn.m2)
```

```
## [1] 167
```

```
cm2 <- table(pimaID.normalized.test.target,knn.m2)
msetrain2<-mean((as.numeric(knn.m2)-
as.numeric(pimaID.normalized.train.target)))^2
```

```
## Warning in as.numeric(knn.m2) -
as.numeric(pimaID.normalized.train.target):
## longer object length is not a multiple of shorter object length
```

```
msetest2<-mean((as.numeric(knn.m2)-
as.numeric(pimaID.normalized.test.target)))^2
```

```
#(3) K = 7
k<-7
knn.m3 <-  class::knn(train=pimaID.normalized.train,
test=pimaID.normalized.test, pimaID.normalized.train.target,k)
knn.m3
```

```
##   [1] 8  0  2  6  7  6  2  0  0  7  5  2  0  4  1  6  4  1  1  8  1  1  10
##  [24] 1  9  1  0  1  10 0  4  2  1  6  8  0  7  9  2  3  6  2  8  1  7  5
##  [47] 0  0  1  0  2  4  2  6  1  1  9  0  0  5  0  6  2  4  1  1  1  1  2
##  [70] 2  7  4  8  5  3  4  2  8  8  7  2  7  1  10 4  2  4  0  6  4  2  2
```

```
##  [93] 4  2  1  7  5  2  7  1  7  0  0  2  3  0  1  7  7  13 5  1  4  2  0
## [116] 1  2  1  1  10 3  6  1  1  1  4  3  1  0  10 7  5  7  4  7  1  0  1
## [139] 4  1  6  4  1  12 10 10 1  0  1  0  0  10 1  9  10 8  8  2  1  3  1
## [162] 2  5  6  0  7  1
## Levels: 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 17
```

```
length(knn.m3)
```

```
## [1] 167
```

```
cm3 <- table(pimaID.normalized.test.target,knn.m3)
msetrain3<-mean((as.numeric(knn.m3)-
as.numeric(pimaID.normalized.train.target)))^2
```

```
## Warning in as.numeric(knn.m3) -
as.numeric(pimaID.normalized.train.target):
## longer object length is not a multiple of shorter object length
```

```
msetest3<-mean((as.numeric(knn.m3)-
as.numeric(pimaID.normalized.test.target)))^2
```

```
#(4) K = 10
k<-10
knn.m4 <-  class::knn(train=pimaID.normalized.train,
test=pimaID.normalized.test, pimaID.normalized.train.target,k)
knn.m4
```

```
##   [1] 8  0  3  5  12 6  2  0  0  0  5  1  0  0  2  1  4  1  1  7  0  1  9
##  [24] 1  9  1  0  1  1  1  4  2  1  5  2  0  7  9  0  10 6  4  11 1  7  10
##  [47] 4  0  4  0  2  4  1  7  1  1  7  0  3  4  0  0  2  0  1  1  1  2  2
##  [70] 2  7  4  8  6  3  4  2  8  1  7  2  7  1  10 4  2  4  0  2  4  2  4
##  [93] 4  2  1  7  7  1  11 1  7  4  0  2  3  0  0  7  6  10 7  1  4  2  0
## [116] 1  2  3  1  10 3  4  1  1  1  1  3  1  4  10 7  1  7  4  7  6  2  1
## [139] 4  1  6  7  1  12 10 4  1  0  2  4  0  4  1  7  10 7  8  2  1  3  1
## [162] 2  5  7  3  7  2
## Levels: 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 17
```

```
length(knn.m4)
```

```
## [1] 167
```

```
cm4 <- table(pimaID.normalized.test.target,knn.m4)
msetrain4<-mean((as.numeric(knn.m4)-
as.numeric(pimaID.normalized.train.target)))^2
```

```
## Warning in as.numeric(knn.m4) -
as.numeric(pimaID.normalized.train.target):
## longer object length is not a multiple of shorter object length
```

```
msetest4<-mean((as.numeric(knn.m4)-
as.numeric(pimaID.normalized.test.target)))^2
```

```
#(5) K = 25
```

```
k<-25
knn.m5 <-  class::knn(train=pimaID.normalized.train,
test=pimaID.normalized.test, pimaID.normalized.train.target,k)
knn.m5
```

```
##   [1] 3  0  3  5  1  6  2  0  2  0  7  2  1  5  0  3  4  1  1  6  0  1  9
##  [24] 1  4  1  0  1  10 0  4  1  2  0  0  1  7  8  2  7  6  4  7  2  7  7
##  [47] 4  0  10 1  2  0  1  6  0  1  7  0  1  3  0  0  2  4  1  1  1  2  2
##  [70] 2  1  8  8  10 4  4  1  6  5  9  1  7  1  6  4  2  0  0  8  4  1  2
##  [93] 3  1  2  7  12 1  10 1  10 0  0  0  3  1  2  7  12 1  10 1  4  2  0
## [116] 1  1  1  1  5  4  0  3  1  2  2  2  1  0  10 5  2  7  4  7  3  0  1
## [139] 4  1  6  7  1  8  4  6  1  1  2  3  0  4  1  1  4  8  1  1  6  3  1
## [162] 1  8  1  1  3  1
## Levels: 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 17
```

```
length(knn.m5)
```

```
## [1] 167
```

```
cm5 <- table(pimaID.normalized.test.target,knn.m5)
msetrain5<-mean((as.numeric(knn.m5)-
as.numeric(pimaID.normalized.train.target)))^2
```

```
## Warning in as.numeric(knn.m5) -
as.numeric(pimaID.normalized.train.target):
## longer object length is not a multiple of shorter object length
```

```
msetest5<-mean((as.numeric(knn.m5)-
as.numeric(pimaID.normalized.test.target)))^2
```

## Answers:

### A(1)

I have tried 5 different values of 'k' i.e. 5, 3, 7, 10 and 25. It is clearly observed from 'MSE test' and 'MSE train' that as k increases from 3 to 25, the MSE decreases for both training and testing. Thus higher the value of 'k', lessser is the error. At some point, the MSE value becomes almost constant. Thus the 'k' in knn makes a huge difference in reducing the error for the training and testing dataset.

### A(2)

Normalization of the data is important so that the data is consistent. Hence when we can see that the data is not consistent based upon the summary, we can normalize it in order to make it more consistent. Scaling of the data improves the performace in terms of accuracy and lesser MSE. Scaling of the data removes the high differences between the data and hence improve model performace. In certain data the scaling is not required as the data is close to each other. But in cases where the data variation is high, scaling makes it easier to apply knn and improve performace.