ISYE-6501
Fall 2024

Homework 10

Question 14.1

The breast cancer data set breast-cancer-wisconsin-data.txt has missing values.

1. Use the mean/mode imputation method to impute values for the missing data.

2. Use regression to impute values for missing data.

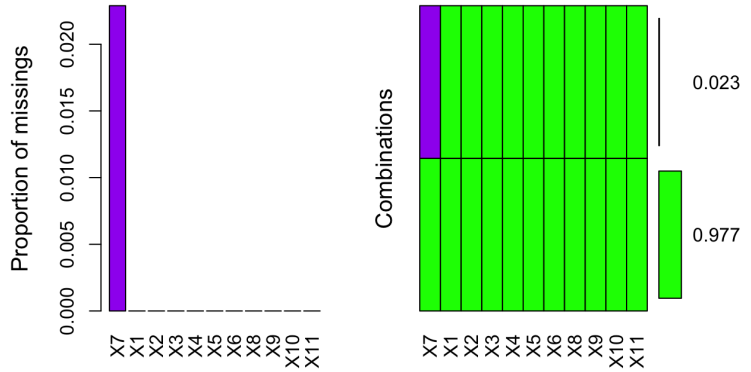
3. Use regression with perturbation to impute values for the missing data.

ANALYSIS:

We can see that the missing values are in column X7, you can see in the end of the code portion in blue the same imputations were made using the three methods.

Below is the code/ R markdown, find the input in black, the comments in green, and the output in blue.

```
> #load all the necessary libraries
> library(mice)
> library("tidyverse")
> library(magrittr)
> library(dplyr)
> library(VIM)
> #load the data breast cancer data into a table named data.
> cancer_df = read_delim('breast-cancer-wisconsin.data.txt', delim = ',',col_names = F, na = c('?')) %>% >
#check for missing values
> md.pattern(cancer_df[,-11])
    X1 X2 X3 X4 X5 X6 X8 X9 X10 X7
683  1  1  1  1  1  1  1  1   1  0
16   1  1  1  1  1  1  1  1   0  1
     0  0  0  0  0  0  0  0   0 16 16
> # plot the missing data
> plot_missing <- aggr(cancer_df, col = c('green', 'purple'), numbers = T, sortVars = T)
```



Variables sorted by number of missings:

```
 Variable      Count
       X7 0.02288984
       X1 0.00000000
       X2 0.00000000
       X3 0.00000000
       X4 0.00000000
       X5 0.00000000
       X6 0.00000000
       X8 0.00000000
       X9 0.00000000
      X10 0.00000000
      X11 0.00000000
> # imputation using mean imputation
> mean_impute <- mice(cancer_df, m = 5, meth = 'mean' )
 iter imp variable
  1   1  X7
  1   2  X7
  1   3  X7
  1   4  X7
  1   5  X7
  2   1  X7
  2   2  X7
  2   3  X7
  2   4  X7
  2   5  X7
  3   1  X7
  3   2  X7
  3   3  X7
  3   4  X7
  3   5  X7
  4   1  X7
  4   2  X7
  4   3  X7
  4   4  X7
  4   5  X7
  5   1  X7
  5   2  X7
  5   3  X7
  5   4  X7
  5   5  X7
Warning message:
Number of logged events: 1
```

```
> # look at the values
> mean_impute$imp

$X7
        1       2       3       4       5
24  3.544656 3.544656 3.544656 3.544656 3.544656
41  3.544656 3.544656 3.544656 3.544656 3.544656
140 3.544656 3.544656 3.544656 3.544656 3.544656
146 3.544656 3.544656 3.544656 3.544656 3.544656
159 3.544656 3.544656 3.544656 3.544656 3.544656
165 3.544656 3.544656 3.544656 3.544656 3.544656
236 3.544656 3.544656 3.544656 3.544656 3.544656
250 3.544656 3.544656 3.544656 3.544656 3.544656
276 3.544656 3.544656 3.544656 3.544656 3.544656
293 3.544656 3.544656 3.544656 3.544656 3.544656
295 3.544656 3.544656 3.544656 3.544656 3.544656
298 3.544656 3.544656 3.544656 3.544656 3.544656
316 3.544656 3.544656 3.544656 3.544656 3.544656
322 3.544656 3.544656 3.544656 3.544656 3.544656
412 3.544656 3.544656 3.544656 3.544656 3.544656
618 3.544656 3.544656 3.544656 3.544656 3.544656

> # imputation using regression imputation
> regression_impute <- mice(cancer_df, m = 5, meth = 'norm.predict')
 iter imp variable
  1   1  X7
  1   2  X7
  1   3  X7
  1   4  X7
  1   5  X7
  2   1  X7
  2   2  X7
  2   3  X7
  2   4  X7
  2   5  X7
  3   1  X7
  3   2  X7
  3   3  X7
  3   4  X7
  3   5  X7
  4   1  X7
  4   2  X7
```

```
 4  3 X7
 4  4 X7
 4  5 X7
 5  1 X7
 5  2 X7
 5  3 X7
 5  4 X7
 5  5 X7
Warning message:
Number of logged events: 1
> # look at the values
> regression_impute$imp
$X7
        1       2       3       4       5
24  5.3669508 5.3669508 5.3669508 5.3669508 5.3669508
41  8.1907122 8.1907122 8.1907122 8.1907122 8.1907122
140 0.8738591 0.8738591 0.8738591 0.8738591 0.8738591
146 1.6463893 1.6463893 1.6463893 1.6463893 1.6463893
159 1.0731978 1.0731978 1.0731978 1.0731978 1.0731978
165 2.1870186 2.1870186 2.1870186 2.1870186 2.1870186
236 2.7459168 2.7459168 2.7459168 2.7459168 2.7459168
250 2.0127161 2.0127161 2.0127161 2.0127161 2.0127161
276 2.3072038 2.3072038 2.3072038 2.3072038 2.3072038
293 5.9989744 5.9989744 5.9989744 5.9989744 5.9989744
295 1.1204527 1.1204527 1.1204527 1.1204527 1.1204527
298 2.6839366 2.6839366 2.6839366 2.6839366 2.6839366
316 5.6353059 5.6353059 5.6353059 5.6353059 5.6353059
322 1.8585015 1.8585015 1.8585015 1.8585015 1.8585015
412 0.8587684 0.8587684 0.8587684 0.8587684 0.8587684
618 0.5907393 0.5907393 0.5907393 0.5907393 0.5907393

> # imputation using regression with perturbation
> pert_impute <- mice(cancer_df, m = 5, meth = 'norm.nob')
 iter imp variable
  1  1 X7
  1  2 X7
  1  3 X7
  1  4 X7
  1  5 X7
  2  1 X7
  2  2 X7
  2  3 X7
```

```
 2  4 X7
 2  5 X7
 3  1 X7
 3  2 X7
 3  3 X7
 3  4 X7
 3  5 X7
 4  1 X7
 4  2 X7
 4  3 X7
 4  4 X7
 4  5 X7
 5  1 X7
 5  2 X7
 5  3 X7
 5  4 X7
 5  5 X7
Warning message:
Number of logged events: 1
> # look at the values
> pert_impute$imp
$X7
         1        2        3       4        5
24   6.23358472  8.0271227  0.9337837  8.256625  7.1834263
41   8.81357783  5.3377516  6.8513655  6.812686 10.4767514
140  3.88406799  2.0909207  2.5526660 -2.412527  4.0841718
146  3.69481835  0.1574474  1.4538946  2.452247  1.6006730
159  0.02118125  3.1902911 -0.7885421  2.762870 -1.2017635
165  3.60465145  3.5975785 -0.9113188  6.374273  2.8073435
236 -2.14400409  1.5982281  7.3236024  3.317701  2.2737085
250  0.52755448  4.5951968  2.3537885  6.290421  2.0531766
276  0.61722386  0.4089054  3.2084575  2.090915  0.5328379
293  4.93585833  4.6880737  6.5444722  7.956978  4.5417857
295  1.40174202 -0.2426303  1.6503846 -3.856720  1.6431164
298  6.92900576  4.4402970  3.0282464  5.232050  3.7465408
316  4.32343591  4.9453924  7.4233530  5.657383  6.8217522
322  0.77094260 -0.1269052  1.2712673  3.069592 -3.2332647
412  1.62130833 -1.2096171 -0.9429011  2.038874 -0.5928264
618  5.26062704  1.8172556 -0.1356313  4.616169  5.4106782

> # cancer with mean impute
> cancer_mean_df <- complete(mean_impute)
```

> cancer_mean_df

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11
1	1000025	5	1	1	1	2	1.000000	3	1	1	Benign
2	1002945	5	4	4	5	7	10.000000	3	2	1	Benign
3	1015425	3	1	1	1	2	2.000000	3	1	1	Benign
4	1016277	6	8	8	1	3	4.000000	3	7	1	Benign
5	1017023	4	1	1	3	2	1.000000	3	1	1	Benign
6	1017122	8	10	10	8	7	10.000000	9	7	1	Malignant
7	1018099	1	1	1	1	2	10.000000	3	1	1	Benign
8	1018561	2	1	2	1	2	1.000000	3	1	1	Benign
9	1033078	2	1	1	1	2	1.000000	1	1	5	Benign
10	1033078	4	2	1	1	2	1.000000	2	1	1	Benign
11	1035283	1	1	1	1	1	1.000000	3	1	1	Benign
12	1036172	2	1	1	1	2	1.000000	2	1	1	Benign
13	1041801	5	3	3	3	2	3.000000	4	4	1	Malignant
14	1043999	1	1	1	1	2	3.000000	3	1	1	Benign
15	1044572	8	7	5	10	7	9.000000	5	5	4	Malignant
16	1047630	7	4	6	4	6	1.000000	4	3	1	Malignant
17	1048672	4	1	1	1	2	1.000000	2	1	1	Benign
18	1049815	4	1	1	1	2	1.000000	3	1	1	Benign
19	1050670	10	7	7	6	4	10.000000	4	1	2	Malignant
20	1050718	6	1	1	1	2	1.000000	3	1	1	Benign
21	1054590	7	3	2	10	5	10.000000	5	4	4	Malignant
22	1054593	10	5	5	3	6	7.000000	7	10	1	Malignant
23	1056784	3	1	1	1	2	1.000000	2	1	1	Benign
24	1057013	8	4	5	1	2	3.544656	7	3	1	Malignant
25	1059552	1	1	1	1	2	1.000000	3	1	1	Benign
26	1065726	5	2	3	4	2	7.000000	3	6	1	Malignant
27	1066373	3	2	1	1	1	1.000000	2	1	1	Benign
28	1066979	5	1	1	1	2	1.000000	2	1	1	Benign
29	1067444	2	1	1	1	2	1.000000	2	1	1	Benign
30	1070935	1	1	3	1	2	1.000000	1	1	1	Benign
31	1070935	3	1	1	1	1	1.000000	2	1	1	Benign
32	1071760	2	1	1	1	2	1.000000	3	1	1	Benign
33	1072179	10	7	7	3	8	5.000000	7	4	3	Malignant
34	1074610	2	1	1	2	2	1.000000	3	1	1	Benign
35	1075123	3	1	2	1	2	1.000000	2	1	1	Benign
36	1079304	2	1	1	1	2	1.000000	2	1	1	Benign
37	1080185	10	10	10	8	6	1.000000	8	9	1	Malignant
38	1081791	6	2	1	1	1	1.000000	7	1	1	Benign
39	1084584	5	4	4	9	2	10.000000	5	6	1	Malignant
40	1091262	2	5	3	3	6	7.000000	7	5	1	Malignant

41 1096800 6 6 6 9 6 3.544656 7 8 1 Benign
42 1099510 10 4 3 1 3 3.000000 6 5 2 Malignant
43 1100524 6 10 10 2 8 10.000000 7 3 3 Malignant
44 1102573 5 6 5 6 10 1.000000 3 1 1 Malignant
45 1103608 10 10 10 4 8 1.000000 8 10 1 Malignant
46 1103722 1 1 1 1 2 1.000000 2 1 2 Benign
47 1105257 3 7 7 4 4 9.000000 4 8 1 Malignant
48 1105524 1 1 1 1 2 1.000000 2 1 1 Benign
49 1106095 4 1 1 3 2 1.000000 3 1 1 Benign
50 1106829 7 8 7 2 4 8.000000 3 8 2 Malignant
51 1108370 9 5 8 1 2 3.000000 2 1 5 Malignant
52 1108449 5 3 3 4 2 4.000000 3 4 1 Malignant
53 1110102 10 3 6 2 3 5.000000 4 10 2 Malignant
54 1110503 5 5 5 8 10 8.000000 7 3 7 Malignant
55 1110524 10 5 5 6 8 8.000000 7 1 1 Malignant
56 1111249 10 6 6 3 4 5.000000 3 6 1 Malignant
57 1112209 8 10 10 1 3 6.000000 3 9 1 Malignant
58 1113038 8 2 4 1 5 1.000000 5 4 4 Malignant
59 1113483 5 2 3 1 6 10.000000 5 1 1 Malignant
60 1113906 9 5 5 2 2 2.000000 5 1 1 Malignant
61 1115282 5 3 5 5 3 3.000000 4 10 1 Malignant
62 1115293 1 1 1 1 2 2.000000 2 1 1 Benign
63 1116116 9 10 10 1 10 8.000000 3 3 1 Malignant
64 1116132 6 3 4 1 5 2.000000 3 9 1 Malignant
65 1116192 1 1 1 1 2 1.000000 2 1 1 Benign
66 1116998 10 4 2 1 3 2.000000 4 3 10 Malignant
67 1117152 4 1 1 1 2 1.000000 3 1 1 Benign
68 1118039 5 3 4 1 8 10.000000 4 9 1 Malignant
69 1120559 8 3 8 3 4 9.000000 8 9 8 Malignant
70 1121732 1 1 1 1 2 1.000000 3 2 1 Benign
71 1121919 5 1 3 1 2 1.000000 2 1 1 Benign
72 1123061 6 10 2 8 10 2.000000 7 8 10 Malignant
73 1124651 1 3 3 2 2 1.000000 7 2 1 Benign
74 1125035 9 4 5 10 6 10.000000 4 8 1 Malignant
75 1126417 10 6 4 1 3 4.000000 3 2 3 Malignant
76 1131294 1 1 2 1 2 2.000000 4 2 1 Benign
77 1132347 1 1 4 1 2 1.000000 2 1 1 Benign
78 1133041 5 3 1 2 2 1.000000 2 1 1 Benign
79 1133136 3 1 1 1 2 3.000000 3 1 1 Benign
80 1136142 2 1 1 1 3 1.000000 2 1 1 Benign
81 1137156 2 2 2 1 1 1.000000 7 1 1 Benign
82 1143978 4 1 1 2 2 1.000000 2 1 1 Benign

```
83 1143978  5  2  1  1  2  1.000000  3  1   1    Benign
84 1147044  3  1  1  1  2  2.000000  7  1   1    Benign
85 1147699  3  5  7  8  8  9.000000  7 10   7 Malignant
86 1147748  5 10  6  1 10  4.000000  4 10  10 Malignant
87 1148278  3  3  6  4  5  8.000000  4  4   1 Malignant
88 1148873  3  6  6  6  5 10.000000  6  8   3 Malignant
89 1152331  4  1  1  1  2  1.000000  3  1   1    Benign
90 1155546  2  1  1  2  3  1.000000  2  1   1    Benign
 [ reached 'max' / getOption("max.print") -- omitted 609 rows ]
> # cancer with regression prediction impute
> cancer_regression_df <- complete(regression_impute)
> cancer_regression_df
      X1 X2 X3 X4 X5 X6       X7 X8 X9 X10      X11
1  1000025  5  1  1  1  2  1.000000  3  1   1    Benign
2  1002945  5  4  4  5  7 10.000000  3  2   1    Benign
3  1015425  3  1  1  1  2  2.000000  3  1   1    Benign
4  1016277  6  8  8  1  3  4.000000  3  7   1    Benign
5  1017023  4  1  1  3  2  1.000000  3  1   1    Benign
6  1017122  8 10 10  8  7 10.000000  9  7   1 Malignant
7  1018099  1  1  1  1  2 10.000000  3  1   1    Benign
8  1018561  2  1  2  1  2  1.000000  3  1   1    Benign
9  1033078  2  1  1  1  2  1.000000  1  1   5    Benign
10 1033078  4  2  1  1  2  1.000000  2  1   1    Benign
11 1035283  1  1  1  1  1  1.000000  3  1   1    Benign
12 1036172  2  1  1  1  2  1.000000  2  1   1    Benign
13 1041801  5  3  3  3  2  3.000000  4  4   1 Malignant
14 1043999  1  1  1  1  2  3.000000  3  1   1    Benign
15 1044572  8  7  5 10  7  9.000000  5  5   4 Malignant
16 1047630  7  4  6  4  6  1.000000  4  3   1 Malignant
17 1048672  4  1  1  1  2  1.000000  2  1   1    Benign
18 1049815  4  1  1  1  2  1.000000  3  1   1    Benign
19 1050670 10  7  7  6  4 10.000000  4  1   2 Malignant
20 1050718  6  1  1  1  2  1.000000  3  1   1    Benign
21 1054590  7  3  2 10  5 10.000000  5  4   4 Malignant
22 1054593 10  5  5  3  6  7.000000  7 10   1 Malignant
23 1056784  3  1  1  1  2  1.000000  2  1   1    Benign
24 1057013  8  4  5  1  2  5.366951  7  3   1 Malignant
25 1059552  1  1  1  1  2  1.000000  3  1   1    Benign
26 1065726  5  2  3  4  2  7.000000  3  6   1 Malignant
27 1066373  3  2  1  1  1  1.000000  2  1   1    Benign
28 1066979  5  1  1  1  2  1.000000  2  1   1    Benign
29 1067444  2  1  1  1  2  1.000000  2  1   1    Benign
```

30 1070935 1 1 3 1 2 1.000000 1 1 1 Benign
31 1070935 3 1 1 1 1 1.000000 2 1 1 Benign
32 1071760 2 1 1 1 2 1.000000 3 1 1 Benign
33 1072179 10 7 7 3 8 5.000000 7 4 3 Malignant
34 1074610 2 1 1 2 2 1.000000 3 1 1 Benign
35 1075123 3 1 2 1 2 1.000000 2 1 1 Benign
36 1079304 2 1 1 1 2 1.000000 2 1 1 Benign
37 1080185 10 10 10 8 6 1.000000 8 9 1 Malignant
38 1081791 6 2 1 1 1 1.000000 7 1 1 Benign
39 1084584 5 4 4 9 2 10.000000 5 6 1 Malignant
40 1091262 2 5 3 3 6 7.000000 7 5 1 Malignant
41 1096800 6 6 6 9 6 8.190712 7 8 1 Benign
42 1099510 10 4 3 1 3 3.000000 6 5 2 Malignant
43 1100524 6 10 10 2 8 10.000000 7 3 3 Malignant
44 1102573 5 6 5 6 10 1.000000 3 1 1 Malignant
45 1103608 10 10 10 4 8 1.000000 8 10 1 Malignant
46 1103722 1 1 1 1 2 1.000000 2 1 2 Benign
47 1105257 3 7 7 4 4 9.000000 4 8 1 Malignant
48 1105524 1 1 1 1 2 1.000000 2 1 1 Benign
49 1106095 4 1 1 3 2 1.000000 3 1 1 Benign
50 1106829 7 8 7 2 4 8.000000 3 8 2 Malignant
51 1108370 9 5 8 1 2 3.000000 2 1 5 Malignant
52 1108449 5 3 3 4 2 4.000000 3 4 1 Malignant
53 1110102 10 3 6 2 3 5.000000 4 10 2 Malignant
54 1110503 5 5 5 8 10 8.000000 7 3 7 Malignant
55 1110524 10 5 5 6 8 8.000000 7 1 1 Malignant
56 1111249 10 6 6 3 4 5.000000 3 6 1 Malignant
57 1112209 8 10 10 1 3 6.000000 3 9 1 Malignant
58 1113038 8 2 4 1 5 1.000000 5 4 4 Malignant
59 1113483 5 2 3 1 6 10.000000 5 1 1 Malignant
60 1113906 9 5 5 2 2 2.000000 5 1 1 Malignant
61 1115282 5 3 5 5 3 3.000000 4 10 1 Malignant
62 1115293 1 1 1 1 2 2.000000 2 1 1 Benign
63 1116116 9 10 10 1 10 8.000000 3 3 1 Malignant
64 1116132 6 3 4 1 5 2.000000 3 9 1 Malignant
65 1116192 1 1 1 1 2 1.000000 2 1 1 Benign
66 1116998 10 4 2 1 3 2.000000 4 3 10 Malignant
67 1117152 4 1 1 1 2 1.000000 3 1 1 Benign
68 1118039 5 3 4 1 8 10.000000 4 9 1 Malignant
69 1120559 8 3 8 3 4 9.000000 8 9 8 Malignant
70 1121732 1 1 1 1 2 1.000000 3 2 1 Benign
71 1121919 5 1 3 1 2 1.000000 2 1 1 Benign

```
72 1123061  6 10  2  8 10  2.000000  7  8  10 Malignant
73 1124651  1  3  3  2  2  1.000000  7  2   1    Benign
74 1125035  9  4  5 10  6 10.000000  4  8   1 Malignant
75 1126417 10  6  4  1  3  4.000000  3  2   3 Malignant
76 1131294  1  1  2  1  2  2.000000  4  2   1    Benign
77 1132347  1  1  4  1  2  1.000000  2  1   1    Benign
78 1133041  5  3  1  2  2  1.000000  2  1   1    Benign
79 1133136  3  1  1  1  2  3.000000  3  1   1    Benign
80 1136142  2  1  1  1  3  1.000000  2  1   1    Benign
81 1137156  2  2  2  1  1  1.000000  7  1   1    Benign
82 1143978  4  1  1  2  2  1.000000  2  1   1    Benign
83 1143978  5  2  1  1  2  1.000000  3  1   1    Benign
84 1147044  3  1  1  1  2  2.000000  7  1   1    Benign
85 1147699  3  5  7  8  8  9.000000  7 10   7 Malignant
86 1147748  5 10  6  1 10  4.000000  4 10  10 Malignant
87 1148278  3  3  6  4  5  8.000000  4  4   1 Malignant
88 1148873  3  6  6  6  5 10.000000  6  8   3 Malignant
89 1152331  4  1  1  1  2  1.000000  3  1   1    Benign
90 1155546  2  1  1  2  3  1.000000  2  1   1    Benign
 [ reached 'max' / getOption("max.print") -- omitted 609 rows ]
> # cancer with pertubation impute
> cancer_pert_df <- complete(pert_impute)
> cancer_pert_df
      X1 X2 X3 X4 X5 X6       X7 X8 X9 X10      X11
1  1000025  5  1  1  1  2  1.000000  3  1   1    Benign
2  1002945  5  4  4  5  7 10.000000  3  2   1    Benign
3  1015425  3  1  1  1  2  2.000000  3  1   1    Benign
4  1016277  6  8  8  1  3  4.000000  3  7   1    Benign
5  1017023  4  1  1  3  2  1.000000  3  1   1    Benign
6  1017122  8 10 10  8  7 10.000000  9  7   1 Malignant
7  1018099  1  1  1  1  2 10.000000  3  1   1    Benign
8  1018561  2  1  2  1  2  1.000000  3  1   1    Benign
9  1033078  2  1  1  1  2  1.000000  1  1   5    Benign
10 1033078  4  2  1  1  2  1.000000  2  1   1    Benign
11 1035283  1  1  1  1  1  1.000000  3  1   1    Benign
12 1036172  2  1  1  1  2  1.000000  2  1   1    Benign
13 1041801  5  3  3  3  2  3.000000  4  4   1 Malignant
14 1043999  1  1  1  1  2  3.000000  3  1   1    Benign
15 1044572  8  7  5 10  7  9.000000  5  5   4 Malignant
16 1047630  7  4  6  4  6  1.000000  4  3   1 Malignant
17 1048672  4  1  1  1  2  1.000000  2  1   1    Benign
18 1049815  4  1  1  1  2  1.000000  3  1   1    Benign
```

19 1050670 10 7 7 6 4 10.000000 4 1 2 Malignant
20 1050718 6 1 1 1 2 1.000000 3 1 1 Benign
21 1054590 7 3 2 10 5 10.000000 5 4 4 Malignant
22 1054593 10 5 5 3 6 7.000000 7 10 1 Malignant
23 1056784 3 1 1 1 2 1.000000 2 1 1 Benign
24 1057013 8 4 5 1 2 6.233585 7 3 1 Malignant
25 1059552 1 1 1 1 2 1.000000 3 1 1 Benign
26 1065726 5 2 3 4 2 7.000000 3 6 1 Malignant
27 1066373 3 2 1 1 1 1.000000 2 1 1 Benign
28 1066979 5 1 1 1 2 1.000000 2 1 1 Benign
29 1067444 2 1 1 1 2 1.000000 2 1 1 Benign
30 1070935 1 1 3 1 2 1.000000 1 1 1 Benign
31 1070935 3 1 1 1 1 1.000000 2 1 1 Benign
32 1071760 2 1 1 1 2 1.000000 3 1 1 Benign
33 1072179 10 7 7 3 8 5.000000 7 4 3 Malignant
34 1074610 2 1 1 2 2 1.000000 3 1 1 Benign
35 1075123 3 1 2 1 2 1.000000 2 1 1 Benign
36 1079304 2 1 1 1 2 1.000000 2 1 1 Benign
37 1080185 10 10 10 8 6 1.000000 8 9 1 Malignant
38 1081791 6 2 1 1 1 1.000000 7 1 1 Benign
39 1084584 5 4 4 9 2 10.000000 5 6 1 Malignant
40 1091262 2 5 3 3 6 7.000000 7 5 1 Malignant
41 1096800 6 6 6 9 6 8.813578 7 8 1 Benign
42 1099510 10 4 3 1 3 3.000000 6 5 2 Malignant
43 1100524 6 10 10 2 8 10.000000 7 3 3 Malignant
44 1102573 5 6 5 6 10 1.000000 3 1 1 Malignant
45 1103608 10 10 10 4 8 1.000000 8 10 1 Malignant
46 1103722 1 1 1 1 2 1.000000 2 1 2 Benign
47 1105257 3 7 7 4 4 9.000000 4 8 1 Malignant
48 1105524 1 1 1 1 2 1.000000 2 1 1 Benign
49 1106095 4 1 1 3 2 1.000000 3 1 1 Benign
50 1106829 7 8 7 2 4 8.000000 3 8 2 Malignant
51 1108370 9 5 8 1 2 3.000000 2 1 5 Malignant
52 1108449 5 3 3 4 2 4.000000 3 4 1 Malignant
53 1110102 10 3 6 2 3 5.000000 4 10 2 Malignant
54 1110503 5 5 5 8 10 8.000000 7 3 7 Malignant
55 1110524 10 5 5 6 8 8.000000 7 1 1 Malignant
56 1111249 10 6 6 3 4 5.000000 3 6 1 Malignant
57 1112209 8 10 10 1 3 6.000000 3 9 1 Malignant
58 1113038 8 2 4 1 5 1.000000 5 4 4 Malignant
59 1113483 5 2 3 1 6 10.000000 5 1 1 Malignant
60 1113906 9 5 5 2 2 2.000000 5 1 1 Malignant

```
61 1115282  5  3  5  5  3  3.000000  4 10   1 Malignant
62 1115293  1  1  1  1  2  2.000000  2  1   1    Benign
63 1116116  9 10 10  1 10  8.000000  3  3   1 Malignant
64 1116132  6  3  4  1  5  2.000000  3  9   1 Malignant
65 1116192  1  1  1  1  2  1.000000  2  1   1    Benign
66 1116998 10  4  2  1  3  2.000000  4  3  10 Malignant
67 1117152  4  1  1  1  2  1.000000  3  1   1    Benign
68 1118039  5  3  4  1  8 10.000000  4  9   1 Malignant
69 1120559  8  3  8  3  4  9.000000  8  9   8 Malignant
70 1121732  1  1  1  1  2  1.000000  3  2   1    Benign
71 1121919  5  1  3  1  2  1.000000  2  1   1    Benign
72 1123061  6 10  2  8 10  2.000000  7  8  10 Malignant
73 1124651  1  3  3  2  2  1.000000  7  2   1    Benign
74 1125035  9  4  5 10  6 10.000000  4  8   1 Malignant
75 1126417 10  6  4  1  3  4.000000  3  2   3 Malignant
76 1131294  1  1  2  1  2  2.000000  4  2   1    Benign
77 1132347  1  1  4  1  2  1.000000  2  1   1    Benign
78 1133041  5  3  1  2  2  1.000000  2  1   1    Benign
79 1133136  3  1  1  1  2  3.000000  3  1   1    Benign
80 1136142  2  1  1  1  3  1.000000  2  1   1    Benign
81 1137156  2  2  2  1  1  1.000000  7  1   1    Benign
82 1143978  4  1  1  2  2  1.000000  2  1   1    Benign
83 1143978  5  2  1  1  2  1.000000  3  1   1    Benign
84 1147044  3  1  1  1  2  2.000000  7  1   1    Benign
85 1147699  3  5  7  8  8  9.000000  7 10   7 Malignant
86 1147748  5 10  6  1 10  4.000000  4 10  10 Malignant
87 1148278  3  3  6  4  5  8.000000  4  4   1 Malignant
88 1148873  3  6  6  6  5 10.000000  6  8   3 Malignant
89 1152331  4  1  1  1  2  1.000000  3  1   1    Benign
90 1155546  2  1  1  2  3  1.000000  2  1   1    Benign
 [ reached 'max' / getOption("max.print") -- omitted 609 rows ]
```

Question 15.1

Describe a situation or problem from your job, everyday life, current events, etc., for which optimization would be appropriate. What data would you need?

As a data analyst at a hospital, a situation for which optimization would be appropriate is patient scheduling for outpatient services. If the scheduling process is inefficient, it can lead to long wait times and overbooked appointments causing patient dissatisfaction. The data I will need include the following; provider schedules to see the availability and workload of the providers, average appointment time to see how long on average does a certain service take, patient flow data to see the amount of time spent during check-in, waiting, and consultation, and historical data on bookings, cancellations, and no-shows.