5.1 Data Introspection

Contents

This section is about getting familiar with our data. We will be using functions to know the size of our table or data frame, the names of the columns or variables, the staructure of the data and the type of data for each of the variables or colummns.

Read the `raw data` again

```
# load the library xlsx
library(xlsx)

# read the raw data
myXl <- read.xlsx("../extdata/oilfield_100w_raw_data.xlsx",
                  sheetIndex = 1, stringsAsFactors = FALSE)
```

Printing the `head` of the dataframe

Let's print 6 rows of data with the function head(). You will see a long printing. We will fix this in a minute. Read on.

```
# the function head() prints the first 6 rows
# to print the last 6 rows use tail()
print(head(myXl))
#:>        Wellname       Company Analyst Field Location Platform Fluid
#:> 1  PSCO-M005-TS Oil Gains Co.    Aida PISCO  M005-TS        M     0
#:> 2 PSCO-M0007-TS Oil Gains Co.    Aida PISCO  M007-TS        M     0
#:> 3  PSCO-M004-LS Oil Gains Co.    Aida PISCO  M004-LS        M     0
#:> 4  PSCO-M008-TS Oil Gains Co.    Aida PISCO  M008-TS        M     0
#:> 5  PSCO-M010-SS Oil Gains Co.    Aida PISCO  M010-SS        M     0
#:> 6  PSCO-M006-TS Oil Gains Co.    Aida PISCO  M006-TS        M     0
#:>   WellType AL_Method Completion SandControl WT_COUNT PVT_GOR PVT_API
#:> 1        0         1          0           0       27   445.7    36.0
#:> 2        0         1          0           0       22   473.0    36.0
#:> 3        0         1          0           0       11   280.0    36.0
#:> 4        0         1          0           0       14   414.0    36.0
#:> 5        0         1          0           0       13   420.0    35.2
#:> 6        0         0          0           0       20   416.0    36.0
#:>   PVT_SG_gas PVT_WaterSalinity PVT_H2S PVT_CO2 PVT_PB_CORR PVT_VISC_CORR
#:> 1        1.2             25000       0    65.5           3             2
#:> 2        1.2             25000       0    65.0           3             2
#:> 3        1.2             25000       0    65.0           3             2
#:> 4        1.2             25000       0    65.0           3             2
#:> 5        1.2             25000       0    65.0           3             2
#:> 6        1.2             25000       0    65.0           3             2
#:>   PVT_BPTEMP PVT_BPPRES VLP_CORR IPR_CORR IPR_RESPRES IPR_RESTEMP
#:> 1        209       1821       10        0         930         209
#:> 2        209       1921       10        0        1300         209
#:> 3        209       1753        1        1        1573         209
#:> 4        209       1698       10        1        1286         200
#:> 5        209       1722       10        0        1468         209
#:> 6        209       1753       10        0        1286         209
#:>   IPR_TOTGOR IPR_WC IPR_VOGELRT IPR_VOGELPRES IPR_PI   GEO_THMD GEO_THTEMP
#:> 1     1449.0     66       384.0       331.000   4.56  0|2289.5|    90|209|
#:> 2     1581.5     70       973.7       956.000   1.15    0|1744|    90|200|
#:> 3     1235.0      0      1327.0       941.436   0.71 0|1954.09|    80|200|
#:> 4     4867.0      5       150.8       418.464   0.25    0|1720|    90|200|
#:> 5      420.0     90      1290.3       430.877   1.35    0|2308|    90|200|
#:> 6     6000.0     80       559.0       902.000   7.80 0|1496.87|    90|200|
#:>   GL_method                                    GL_ArrayMandrels GL_Vdepth
#:> 1         0                             0|0|0|0|0|0|0|0|0|0|0|   1807.53
#:> 2         0             614.3|1118|1422.5|1564.6|0|0|0|0|0|0|0|   1564.60
#:> 3         2        167.152|245.913|327.69|373.99|0|0|0|0|0|0|0|   1227.00
#:> 4         2           560.9|1123.8|1427.6|1569.6|0|0|0|0|0|0|0|      0.00
#:> 5         0 193.054|380.384|487.893|584.93|649.657|0|0|0|0|0|0|   1911.98
#:> 6         2              543.1|969.3|1235.2|1358|0|0|0|0|0|0|0|    969.30
#:>   GL_GSG GL_CO2
#:> 1    1.2     65
#:> 2    1.2     65
#:> 3    1.2     65
#:> 4    1.2     65
#:> 5    1.2     65
#:> 6    1.2     65
#:>
#:> 1 09/09/2014|02/07/2012|08/08/2012|02/09/2012|03/10/2012|11/11/2012|08/12/2012|13/12/2012|02/01/201
#:> 2                                                        09/06/2014|21/08/2014|06/02/2012|17/03/201
#:> 3
#:> 4
#:> 5
#:> 6                                                                              09/07/2012|10/08/201
#:>
#:> 1 121|112|125|125|125|125|125|125|125|125|125|125|125|125|125|125|125|125|125|125|125|125|125|125|1
#:> 2                     125|125|125|125|135|125|125|127|125|125|125|125|125|125|125|125|125|125|125|1
#:> 3                                                                            96|99|99|99|99|99|99|99
#:> 4                                                     108|108|108|108|108|108|108|108|108|108|108|1
#:> 5                                                         190|192|190|190|190|188|145|190|190|190|1
#:> 6                             125|125|125|125|125|125|125|125|125|125|125|125|125|125|125|125|152|1
#:>
#:> 1 561.2|384.5|365.8|405.3|312.2|501.1|469.9|551.1|887.2|534.7|474.2|408.7|527.2|266.9|377.8|540.6|4
#:> 2                                   560|528|711.2|790.6|973.7|732.4|402.5|747.8|793.5|958.9|1190.5|
#:> 3
#:> 4                                                                                   150.8|93.9|257.
#:> 5                                                                           1369.3|1244.9|1035.3|10
#:> 6                                       1108.8|1440.4|1400.6|543.5|1417.3|676.6|1228.9|479.7|1050.2
#:>
#:> 1 65|66.9|71.08|71.09|75.96|71.1|71.09|68.66|71.1|63.42|71.09|71.08|71.08|26.04|71.09|71.09|71.08|3
#:> 2                         70|68|65.94|80.83|75.97|74.75|75.96|75.97|73.09|77.8|76.95|76.94|61.26|75
#:> 3
#:> 4
#:> 5                                                                        95.13|92|90.46|90.46|92.83
#:> 6                              80.64|71.09|85.66|80.83|99.85|90.46|97.66|85.66|80.82|95.24|84.89|90
#:>
#:> 1 246.5|232.1|246.6|217.6|246.6|217.6|217.6|203|203|203|232.1|232.1|232.1|232.1|261.1|217.6|246.6|2
#:> 2                          246.5|1189|246.6|203|232.1|232.1|246.6|246.6|203|290.1|290.1|261.1|246.6
#:> 3                                                                                               435
#:> 4                                                               362.6|464.1|507.6|507.6|319.1|304.6
#:> 5                                                                     261.1|217.6|246.6|232.1|232.1
#:> 6                                 304.6|362.6|304.6|348.1|319.1|290.1|246.6|261.1|290.1|333.6|362.6
#:>
#:> 1        3145|1449|2108|2496|4214|4672|3689|3688|4216|542.6|426|4215|4214|460.4|4216|4216|4215|2243
#:> 2 4160|3974|1624.7|336.5|1581.5|287.5|1581|1581.8|1053|1265.5|1265.9|1160.1|759.6|1160.2|1160.1|163
#:> 3                                                                                       453|573.5|4
#:> 4                                     4867|46172.6|17401.8|16889.5|13548.2|13622.8|16676.7|85862.6|
#:> 5                                                                          869|1800|43.9|2963|415.4
#:> 6                1476.3|1371|1265.8|1266|1573.4|1686.4|1794.5|1792.9|600.5|1792|1792.3|3188.6|5614.
#:>
#:> 1 0|0|0.3|0.3|0.3|0.3|0.3|0.3|0.3|0.3|0.3|0.3|0.3|0.3|0.3|0.3|0.3|0.3|0.3|0.3|0.3|0.3|0.3|0.3|0.3|0
#:> 2              0.5|0.6|0.1|0.1|0.25|0.1|0.1|0.1|0.1|0.1|0.1|0.1|0.1|0.1|0.1|0.1|0.1|0.1|0.1|0.1|0.1
#:> 3                                                             0.2|0.1|0.1|0.1|0.1|0.1|0.1|0.1|0.1|0
#:> 4                                                   0|0.1|0.1|0.1|0.1|0.1|0.1|0.1|0.1|0.1|0.1|0.1|0
#:> 5                                                     0.3|0.5|0.3|0.3|0.3|0.1|0.4|0.2|0.2|0.3|0.2|0
#:> 6                                                                 0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0
#:>
#:> 1 1807.53|1807.53|1807.53|1807.53|1807.53|1807.53|1807.53|1807.53|1807.53|1807.53|1807.53|1807.53|1
#:> 2                                                               1564.6|1564.6|1564.6|1564.6|1564.6|
#:> 3
#:> 4
#:> 5
#:> 6
#:>                                                 WT_Enable
#:> 1 0|1|1|1|1|1|1|1|1|1|1|1|1|1|1|1|1|1|1|1|1|1|1|1|1|1|1|
#:> 2           1|1|1|1|1|1|1|1|1|1|1|1|1|1|1|1|1|1|1|1|1|0|
#:> 3                                 1|1|1|1|1|1|1|1|1|0|1|
#:> 4                           1|1|1|1|1|1|1|1|1|1|1|1|1|0|
#:> 5                             1|1|1|1|1|1|0|1|1|1|1|1|1|
#:> 6               1|1|1|1|1|1|1|1|1|1|1|1|1|1|1|1|0|1|1|1|
#:>                                                 WT_GDEPTH
#:> 1 0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|
#:> 2           0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|
#:> 3                                 0|0|0|0|0|0|0|0|0|0|0|
#:> 4                           0|0|0|0|0|0|0|0|0|0|0|0|0|0|
#:> 5                             0|0|0|0|0|0|0|0|0|0|0|0|0|
#:> 6               0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|
#:>
```

```
#:> 1 246.5|232.1|246.6|217.6|246.6|217.6|217.6|203|203|203|232.1|232.1|232.1|232.1|261.1|217.6|246.6|2
#:> 2                          246.5|1189|246.6|203|232.1|232.1|246.6|246.6|203|290.1|290.1|261.1|246.6
#:> 3                                                                                               435
#:> 4                                                               362.6|464.1|507.6|507.6|319.1|304.6
#:> 5                                                                     261.1|217.6|246.6|232.1|232.1
#:> 6                                 304.6|362.6|304.6|348.1|319.1|290.1|246.6|261.1|290.1|333.6|362.6
#:>
#:> 1   930|930|930|930|930|930|930|930|930|930|930|930|930|930|930|930|930|930|930|930|930|930|930|930
#:> 2 1300|1300|1300|1300|1300|1300|1300|1300|1300|1300|1300|1300|1300|1300|1300|1300|1300|1300|1300|13
#:> 3                                                        1573|1573|1573|1573|1573|1573|1573|1573|15
#:> 4                                         1286|1025|1025|1025|1025|1025|1025|1025|1025|1025|1025|10
#:> 5                                               1468|900|1468|1468|1468|1468|1468|1468|1468|1468|14
#:> 6           1286|1286|1286|1286|1286|1286|1286|1286|1286|1286|1286|1286|1286|1286|1286|1286|1275|12
#:>                                      ProsperFilename
#:> 1 \\\\network\\piscis\\well_models\\PISC-M005-TS.Out
#:> 2 \\\\network\\piscis\\well_models\\PISC-M007-TS.Out
#:> 3 \\\\network\\piscis\\well_models\\PISC-M004-LL.Out
#:> 4 \\\\network\\piscis\\well_models\\PISC-M008-TS.Out
#:> 5 \\\\network\\piscis\\well_models\\PISC-M010-SS.Out
#:> 6 \\\\network\\piscis\\well_models\\PISC-M006-TS.Out
```

It looks pretty long. Let's try with a package that adds better printing capabilities: `tibble`.

Install `tibble`

Install it with `install.packages("tibble")`

```
library(tibble)      # load the package
myXl <- as_tibble(myXl)      # convert the data frame to a tibble
```

```
head(myXl)
#:> # A tibble: 6 x 51
#:>        Wellname       Company Analyst Field Location Platform Fluid
#:>           <chr>         <chr>   <chr> <chr>    <chr>    <chr> <dbl>
#:> 1  PSCO-M005-TS Oil Gains Co.    Aida PISCO  M005-TS        M     0
#:> 2 PSCO-M0007-TS Oil Gains Co.    Aida PISCO  M007-TS        M     0
#:> 3  PSCO-M004-LS Oil Gains Co.    Aida PISCO  M004-LS        M     0
#:> 4  PSCO-M008-TS Oil Gains Co.    Aida PISCO  M008-TS        M     0
#:> 5  PSCO-M010-SS Oil Gains Co.    Aida PISCO  M010-SS        M     0
#:> 6  PSCO-M006-TS Oil Gains Co.    Aida PISCO  M006-TS        M     0
#:> # ... with 44 more variables: WellType <dbl>, AL_Method <dbl>,
#:> #   Completion <dbl>, SandControl <dbl>, WT_COUNT <dbl>, PVT_GOR <dbl>,
#:> #   PVT_API <dbl>, PVT_SG_gas <dbl>, PVT_WaterSalinity <dbl>,
#:> #   PVT_H2S <dbl>, PVT_CO2 <dbl>, PVT_PB_CORR <dbl>, PVT_VISC_CORR <dbl>,
#:> #   PVT_BPTEMP <dbl>, PVT_BPPRES <dbl>, VLP_CORR <dbl>, IPR_CORR <dbl>,
#:> #   IPR_RESPRES <dbl>, IPR_RESTEMP <dbl>, IPR_TOTGOR <dbl>, IPR_WC <dbl>,
#:> #   IPR_VOGELRT <dbl>, IPR_VOGELPRES <dbl>, IPR_PI <dbl>, GEO_THMD <chr>,
#:> #   GEO_THTEMP <chr>, GL_method <dbl>, GL_ArrayMandrels <chr>,
#:> #   GL_Vdepth <dbl>, GL_GSG <dbl>, GL_CO2 <dbl>, WT_DATE <chr>,
#:> #   WT_THT <chr>, WT_LIQRT <chr>, WT_WC <chr>, WT_THP <chr>, WT_GOR <chr>,
#:> #   WT_GLIR <chr>, WT_DEPTH <chr>, WT_Enable <chr>, WT_GDEPTH <chr>,
#:> #   WT_GPRES <chr>, WT_RESPRES <chr>, ProsperFilename <chr>
```

```
tail(myXl)
#:> # A tibble: 6 x 51
#:>        Wellname       Company  Analyst Field Location Platform Fluid
#:>           <chr>         <chr>    <chr> <chr>    <chr>    <chr> <dbl>
#:> 1  PSCO-S021-TS Oil Gains Co.   Camden PISCO  S021-TS        S     0
#:> 2  PSCO-S016-SS Oil Gains Co.   Camden PISCO  S016-SS        S     0
#:> 3  PSCO-S015-SS Oil Gains Co.   Camden PISCO  S015-SS        S     0
#:> 4  PSCO-S012-LS Oil Gains Co.     <NA> PISCO  S012-LS        S     0
#:> 5  PSCO-M001-TS Oil Gains Co.     Aida PISCO  M001-TS     <NA>     0
#:> 6 PSCO-M0026-TS Oil Gains Co. Ibironke PISCO  M026-TS     <NA>     0
#:> # ... with 44 more variables: WellType <dbl>, AL_Method <dbl>,
#:> #   Completion <dbl>, SandControl <dbl>, WT_COUNT <dbl>, PVT_GOR <dbl>,
#:> #   PVT_API <dbl>, PVT_SG_gas <dbl>, PVT_WaterSalinity <dbl>,
#:> #   PVT_H2S <dbl>, PVT_CO2 <dbl>, PVT_PB_CORR <dbl>, PVT_VISC_CORR <dbl>,
#:> #   PVT_BPTEMP <dbl>, PVT_BPPRES <dbl>, VLP_CORR <dbl>, IPR_CORR <dbl>,
#:> #   IPR_RESPRES <dbl>, IPR_RESTEMP <dbl>, IPR_TOTGOR <dbl>, IPR_WC <dbl>,
#:> #   IPR_VOGELRT <dbl>, IPR_VOGELPRES <dbl>, IPR_PI <dbl>, GEO_THMD <chr>,
#:> #   GEO_THTEMP <chr>, GL_method <dbl>, GL_ArrayMandrels <chr>,
#:> #   GL_Vdepth <dbl>, GL_GSG <dbl>, GL_CO2 <dbl>, WT_DATE <chr>,
#:> #   WT_THT <chr>, WT_LIQRT <chr>, WT_WC <chr>, WT_THP <chr>, WT_GOR <chr>,
#:> #   WT_GLIR <chr>, WT_DEPTH <chr>, WT_Enable <chr>, WT_GDEPTH <chr>,
#:> #   WT_GPRES <chr>, WT_RESPRES <chr>, ProsperFilename <chr>
```

Now it looks much better.

dimensions of the data frame: `dim`

Let's use some R functions to find out more about our data.

```
# get the dimensions of the table.
dim(myXl)
#:> [1] 100  51
```

Our table has 100 rows and 51 columns.

names of the columns: `names`

These are the names of the variables or columns:

```
names(myXl)
#:>  [1] "Wellname"          "Company"           "Analyst"
#:>  [4] "Field"             "Location"          "Platform"
#:>  [7] "Fluid"             "WellType"          "AL_Method"
#:> [10] "Completion"        "SandControl"       "WT_COUNT"
#:> [13] "PVT_GOR"           "PVT_API"           "PVT_SG_gas"
#:> [16] "PVT_WaterSalinity" "PVT_H2S"           "PVT_CO2"
#:> [19] "PVT_PB_CORR"       "PVT_VISC_CORR"     "PVT_BPTEMP"
#:> [22] "PVT_BPPRES"        "VLP_CORR"          "IPR_CORR"
#:> [25] "IPR_RESPRES"       "IPR_RESTEMP"       "IPR_TOTGOR"
#:> [28] "IPR_WC"            "IPR_VOGELRT"       "IPR_VOGELPRES"
#:> [31] "IPR_PI"            "GEO_THMD"          "GEO_THTEMP"
#:> [34] "GL_method"         "GL_ArrayMandrels"  "GL_Vdepth"
#:> [37] "GL_GSG"            "GL_CO2"            "WT_DATE"
#:> [40] "WT_THT"            "WT_LIQRT"          "WT_WC"
#:> [43] "WT_THP"            "WT_GOR"            "WT_GLIR"
#:> [46] "WT_DEPTH"          "WT_Enable"         "WT_GDEPTH"
#:> [49] "WT_GPRES"          "WT_RESPRES"        "ProsperFilename"
```

summary of the data: `summary`

```
# A summary of all the variables.
# Notice the difference between numerical and non-numerical variables
summary(myXl)
#:>    Wellname           Company             Analyst
#:>  Length:100         Length:100         Length:100
#:>  Class :character   Class :character   Class :character
#:>  Mode  :character   Mode  :character   Mode  :character
#:>
#:>
#:>
#:>
#:>     Field             Location           Platform             Fluid
#:>  Length:100         Length:100         Length:100         Min.   :0
#:>  Class :character   Class :character   Class :character   1st Qu.:0
#:>  Mode  :character   Mode  :character   Mode  :character   Median :0
#:>                                                           Mean   :0
#:>                                                           3rd Qu.:0
#:>                                                           Max.   :0
#:>
#:>     WellType   AL_Method      Completion    SandControl      WT_COUNT
#:>  Min.   :0   Min.   :0.00   Min.   :0.00   Min.   :0.00   Min.   : 1.00
#:>  1st Qu.:0   1st Qu.:1.00   1st Qu.:0.00   1st Qu.:0.00   1st Qu.: 1.00
#:>  Median :0   Median :1.00   Median :0.00   Median :0.00   Median : 3.00
#:>  Mean   :0   Mean   :0.98   Mean   :0.07   Mean   :0.24   Mean   : 4.82
#:>  3rd Qu.:0   3rd Qu.:1.00   3rd Qu.:0.00   3rd Qu.:0.00   3rd Qu.: 7.00
#:>  Max.   :0   Max.   :1.00   Max.   :1.00   Max.   :3.00   Max.   :27.00
#:>
#:>     PVT_GOR         PVT_API        PVT_SG_gas    PVT_WaterSalinity
#:>  Min.   :280.0   Min.   :35.00   Min.   :0.800   Min.   : 1000
#:>  1st Qu.:416.0   1st Qu.:36.00   1st Qu.:1.200   1st Qu.:15000
#:>  Median :423.0   Median :36.00   Median :1.200   Median :15000
#:>  Mean   :431.2   Mean   :36.15   Mean   :1.221   Mean   :15247
#:>  3rd Qu.:455.2   3rd Qu.:36.00   3rd Qu.:1.237   3rd Qu.:15125
#:>  Max.   :473.0   Max.   :46.15   Max.   :1.300   Max.   :30000
#:>
#:>     PVT_H2S     PVT_CO2       PVT_PB_CORR   PVT_VISC_CORR
#:>  Min.   :0   Min.   :29.00   Min.   :0.00   Min.   :0.00
#:>  1st Qu.:0   1st Qu.:65.00   1st Qu.:3.00   1st Qu.:1.00
#:>  Median :0   Median :65.00   Median :3.00   Median :2.00
#:>  Mean   :0   Mean   :66.58   Mean   :2.78   Mean   :1.77
#:>  3rd Qu.:0   3rd Qu.:69.25   3rd Qu.:3.00   3rd Qu.:2.00
#:>  Max.   :0   Max.   :74.28   Max.   :3.00   Max.   :4.00
#:>
#:>    PVT_BPTEMP       PVT_BPPRES      VLP_CORR        IPR_CORR
#:>  Min.   : 97.78   Min.   :1683   Min.   : 0.00   Min.   :0.00
#:>  1st Qu.:208.00   1st Qu.:1722   1st Qu.:10.00   1st Qu.:0.00
#:>  Median :209.00   Median :1753   Median :10.00   Median :1.00
#:>  Mean   :215.58   Mean   :1779   Mean   :10.07   Mean   :0.53
#:>  3rd Qu.:209.00   3rd Qu.:1836   3rd Qu.:10.00   3rd Qu.:1.00
#:>  Max.   :408.20   Max.   :1936   Max.   :18.00   Max.   :3.00
#:>  NA's   :1        NA's   :1
#:>   IPR_RESPRES    IPR_RESTEMP      IPR_TOTGOR        IPR_WC
#:>  Min.   : 658   Min.   :107.6   Min.   :  404   Min.   : 0.00
#:>  1st Qu.:1246   1st Qu.:206.0   1st Qu.:  595   1st Qu.:51.75
#:>  Median :1404   Median :209.0   Median : 1247   Median :70.00
#:>  Mean   :1386   Mean   :207.3   Mean   : 2028   Mean   :64.73
#:>  3rd Qu.:1565   3rd Qu.:211.0   3rd Qu.: 2348   3rd Qu.:87.53
#:>  Max.   :2727   Max.   :226.0   Max.   :11229   Max.   :96.00
#:>
#:>   IPR_VOGELRT     IPR_VOGELPRES        IPR_PI          GEO_THMD
#:>  Min.   :   0.0   Min.   :   0.0   Min.   : 0.0000   Length:100
#:>  1st Qu.:   0.0   1st Qu.:   0.0   1st Qu.: 0.8261   Class :character
#:>  Median : 559.4   Median : 782.2   Median : 1.7362   Mode  :character
#:>  Mean   : 670.3   Mean   : 659.0   Mean   : 2.6829
#:>  3rd Qu.:1145.0   3rd Qu.: 982.0   3rd Qu.: 3.4625
#:>  Max.   :2420.8   Max.   :1381.1   Max.   :12.0000
#:>
#:>   GEO_THTEMP          GL_method    GL_ArrayMandrels     GL_Vdepth
#:>  Length:100         Min.   :0.00   Length:100         Min.   :   0
#:>  Class :character   1st Qu.:0.00   Class :character   1st Qu.:1220
#:>  Mode  :character   Median :2.00   Mode  :character   Median :1601
#:>                     Mean   :1.06                      Mean   :2143
#:>                     3rd Qu.:2.00                      3rd Qu.:2304
#:>                     Max.   :2.00                      Max.   :8852
#:>
#:>      GL_GSG          GL_CO2       WT_DATE             WT_THT
#:>  Min.   :0.800   Min.   :65.0   Length:100         Length:100
#:>  1st Qu.:1.200   1st Qu.:65.0   Class :character   Class :character
#:>  Median :1.200   Median :65.0   Mode  :character   Mode  :character
#:>  Mean   :1.196   Mean   :65.1
#:>  3rd Qu.:1.200   3rd Qu.:65.0
#:>  Max.   :1.200   Max.   :70.0
#:>
#:>    WT_LIQRT            WT_WC              WT_THP
#:>  Length:100         Length:100         Length:100
#:>  Class :character   Class :character   Class :character
#:>  Mode  :character   Mode  :character   Mode  :character
#:>
#:>
#:>
#:>
#:>     WT_GOR            WT_GLIR            WT_DEPTH
#:>  Length:100         Length:100         Length:100
#:>  Class :character   Class :character   Class :character
#:>  Mode  :character   Mode  :character   Mode  :character
#:>
#:>
#:>
```

```
#:>
#:>   WT_Enable          WT_GDEPTH           WT_GPRES
#:>  Length:100         Length:100         Length:100
#:>  Class :character   Class :character   Class :character
#:>  Mode  :character   Mode  :character   Mode  :character
#:>
#:>
#:>
#:>
#:>   WT_RESPRES        ProsperFilename
#:>  Length:100         Length:100
#:>  Class :character   Class :character
#:>  Mode  :character   Mode  :character
#:>
#:>
#:>
#:>
```

structure of the data: str

```
# show the data type structure of the table
str(myXl)
#:> Classes 'tbl_df', 'tbl' and 'data.frame':   100 obs. of  51 variables:
#:>  $ Wellname         : chr  "PSCO-M005-TS" "PSCO-M0007-TS" "PSCO-M004-LS" "PSCO-M008-TS" ...
#:>  $ Company          : chr  "Oil Gains Co." "Oil Gains Co." "Oil Gains Co." "Oil Gains Co." ...
#:>  $ Analyst          : chr  "Aida" "Aida" "Aida" "Aida" ...
#:>  $ Field            : chr  "PISCO" "PISCO" "PISCO" "PISCO" ...
#:>  $ Location         : chr  "M005-TS" "M007-TS" "M004-LS" "M008-TS" ...
#:>  $ Platform         : chr  "M" "M" "M" "M" ...
#:>  $ Fluid            : num  0 0 0 0 0 0 0 0 0 0 ...
#:>  $ WellType         : num  0 0 0 0 0 0 0 0 0 0 ...
#:>  $ AL_Method        : num  1 1 1 1 1 0 1 1 1 1 ...
#:>  $ Completion       : num  0 0 0 0 0 0 0 0 0 0 ...
#:>  $ SandControl      : num  0 0 0 0 0 0 0 0 0 0 ...
#:>  $ WT_COUNT         : num  27 22 11 14 13 20 3 2 2 2 ...
#:>  $ PVT_GOR          : num  446 473 280 414 420 ...
#:>  $ PVT_API          : num  36 36 36 36 35.2 ...
#:>  $ PVT_SG_gas       : num  1.2 1.2 1.2 1.2 1.2 ...
#:>  $ PVT_WaterSalinity: num  25000 25000 25000 25000 25000 25000 15000 15000 15000 15000 ...
#:>  $ PVT_H2S          : num  0 0 0 0 0 0 0 0 0 0 ...
#:>  $ PVT_CO2          : num  65.5 65 65 65 65 65 65 65 65 65 ...
#:>  $ PVT_PB_CORR      : num  3 3 3 3 3 3 3 3 3 3 ...
#:>  $ PVT_VISC_CORR    : num  2 2 2 2 2 2 4 2 0 2 ...
#:>  $ PVT_BPTEMP       : num  209 209 209 209 209 209 209 209 209 209 ...
#:>  $ PVT_BPPRES       : num  1821 1921 1753 1698 1722 ...
#:>  $ VLP_CORR         : num  10 10 1 10 10 10 10 10 10 10 ...
#:>  $ IPR_CORR         : num  0 0 1 1 0 0 0 1 1 0 ...
#:>  $ IPR_RESPRES      : num  930 1300 1573 1286 1468 ...
#:>  $ IPR_RESTEMP      : num  209 209 209 200 209 209 214 211 202 216 ...
#:>  $ IPR_TOTGOR       : num  1449 1582 1235 4867 420 ...
#:>  $ IPR_WC           : num  66 70 0 5 90 80 90 95 90 90 ...
#:>  $ IPR_VOGELRT      : num  384 974 1327 151 1290 ...
#:>  $ IPR_VOGELPRES    : num  331 956 941 418 431 ...
#:>  $ IPR_PI           : num  4.56 1.15 0.71 0.25 1.35 ...
#:>  $ GEO_THMD         : chr  "0|2289.5|" "0|1744|" "0|1954.09|" "0|1720|" ...
#:>  $ GEO_THTEMP       : chr  "90|209|" "90|200|" "80|200|" "90|200|" ...
#:>  $ GL_method        : num  0 0 2 2 0 2 0 0 0 0 ...
#:>  $ GL_ArrayMandrels : chr  "0|0|0|0|0|0|0|0|0|0|" "614.3|1118|1422.5|1564.6|0|0|0|0|0|0|" "167.152|
#:>  $ GL_Vdepth        : num  1808 1565 1227 0 1912 ...
#:>  $ GL_GSG           : num  1.2 1.2 1.2 1.2 1.2 ...
#:>  $ GL_CO2           : num  65 65 65 65 65 65 65 65 65 65 ...
#:>  $ WT_DATE          : chr  "09/09/2014|02/07/2012|08/08/2012|02/09/2012|03/10/2012|11/11/2012|08/12
#:>  $ WT_THT           : chr  "121|112|125|125|125|125|125|125|125|125|125|125|125|125|125|125|125|125
#:>  $ WT_LIQRT         : chr  "561.2|384.5|365.8|405.3|312.2|501.1|469.9|551.1|887.2|534.7|474.2|408.7
#:>  $ WT_WC            : chr  "65|66.9|71.08|71.09|75.96|71.1|71.09|68.66|71.1|63.42|71.09|71.08|71.08
#:>  $ WT_THP           : chr  "246.5|232.1|246.6|217.6|246.6|217.6|217.6|203|203|203|232.1|232.1|232.1
#:>  $ WT_GOR           : chr  "3145|1449|2108|2496|4214|4672|3689|3688|4216|542.6|426|4215|4214|460.4|
#:>  $ WT_GLIR          : chr  "0|0|0.3|0.3|0.3|0.3|0.3|0.3|0.3|0.3|0.3|0.3|0.3|0.3|0.3|0.3|0.3|0.3|0.3
#:>  $ WT_DEPTH         : chr  "1807.53|1807.53|1807.53|1807.53|1807.53|1807.53|1807.53|1807.53|1807.53
#:>  $ WT_Enable        : chr  "0|1|1|1|1|1|1|1|1|1|1|1|1|1|1|1|1|1|1|1|1|1|1|1|1|1|1|" "1|1|1|1|1|1|1|
#:>  $ WT_GDEPTH        : chr  "0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|0|" "0|0|0|0|0|0|0|
#:>  $ WT_GPRES         : chr  "246.5|232.1|246.6|217.6|246.6|217.6|217.6|203|203|203|232.1|232.1|232.1
#:>  $ WT_RESPRES       : chr  "930|930|930|930|930|930|930|930|930|930|930|930|930|930|930|930|930|930
#:>  $ ProsperFilename  : chr  "\\\\network\\piscis\\well_models\\PISC-M005-TS.Out" "\\\\network\\pisci
```

```
# outr table is one of R data structures along with vectors, matrices, arrays and lists.
class(myXl)
#:> [1] "tbl_df"     "tbl"        "data.frame"
```

data types: `typeof`

Let’s find out what type of variable we’ve got in our table:

```
typeof(myXl$Wellname)
#:> [1] "character"
typeof(myXl$Fluid)
#:> [1] "double"
typeof(myXl$IPR_RESTEMP)
#:> [1] "double"
```

using `sapply`, `length`, `sort`

We can do all the column names in one shot with `sapply`.

```
dataTypes <- sapply(myXl, typeof, simplify = "array")
typeof(dataTypes)
#:> [1] "character"
length(dataTypes)
#:> [1] 51
sort(dataTypes)
#:>          Wellname          Company          Analyst            Field
#:>       "character"      "character"      "character"      "character"
#:>          Location         Platform         GEO_THMD       GEO_THTEMP
#:>       "character"      "character"      "character"      "character"
#:>  GL_ArrayMandrels          WT_DATE           WT_THT         WT_LIQRT
#:>       "character"       "character"       "character"       "character" 
#:>             WT_WC            WT_THP            WT_GOR           WT_GLIR 
#:>       "character"       "character"       "character"       "character" 
#:>          WT_DEPTH         WT_Enable         WT_GDEPTH          WT_GPRES 
#:>       "character"       "character"       "character"       "character" 
#:>        WT_RESPRES   ProsperFilename             Fluid          WellType 
#:>       "character"       "character"          "double"          "double" 
#:>         AL_Method        Completion       SandControl          WT_COUNT 
#:>          "double"          "double"          "double"          "double" 
#:>           PVT_GOR           PVT_API        PVT_SG_gas PVT_WaterSalinity 
#:>          "double"          "double"          "double"          "double" 
#:>           PVT_H2S           PVT_CO2       PVT_PB_CORR     PVT_VISC_CORR 
#:>          "double"          "double"          "double"          "double" 
#:>        PVT_BPTEMP        PVT_BPPRES          VLP_CORR          IPR_CORR 
#:>          "double"          "double"          "double"          "double" 
#:>       IPR_RESPRES       IPR_RESTEMP        IPR_TOTGOR            IPR_WC 
#:>          "double"          "double"          "double"          "double" 
#:>       IPR_VOGELRT     IPR_VOGELPRES            IPR_PI         GL_method 
#:>          "double"          "double"          "double"          "double" 
#:>         GL_Vdepth            GL_GSG            GL_CO2 
#:>          "double"          "double"          "double"
```

An inventory of the kind of data we have: `table`

```
table(dataTypes)
#:> dataTypes
#:> character    double 
#:>        22        29
```