Question 0

- Draw a concepts diagram that uses all the following Github terminology
  - Include any missing keywords that will simplify the concepts diagrams

`Push`, `Repository`, `Clone`, `Pull`, `Pull Request`, `Branch`, `Merging`, `Github Client`, `README file`, `Private or Public`

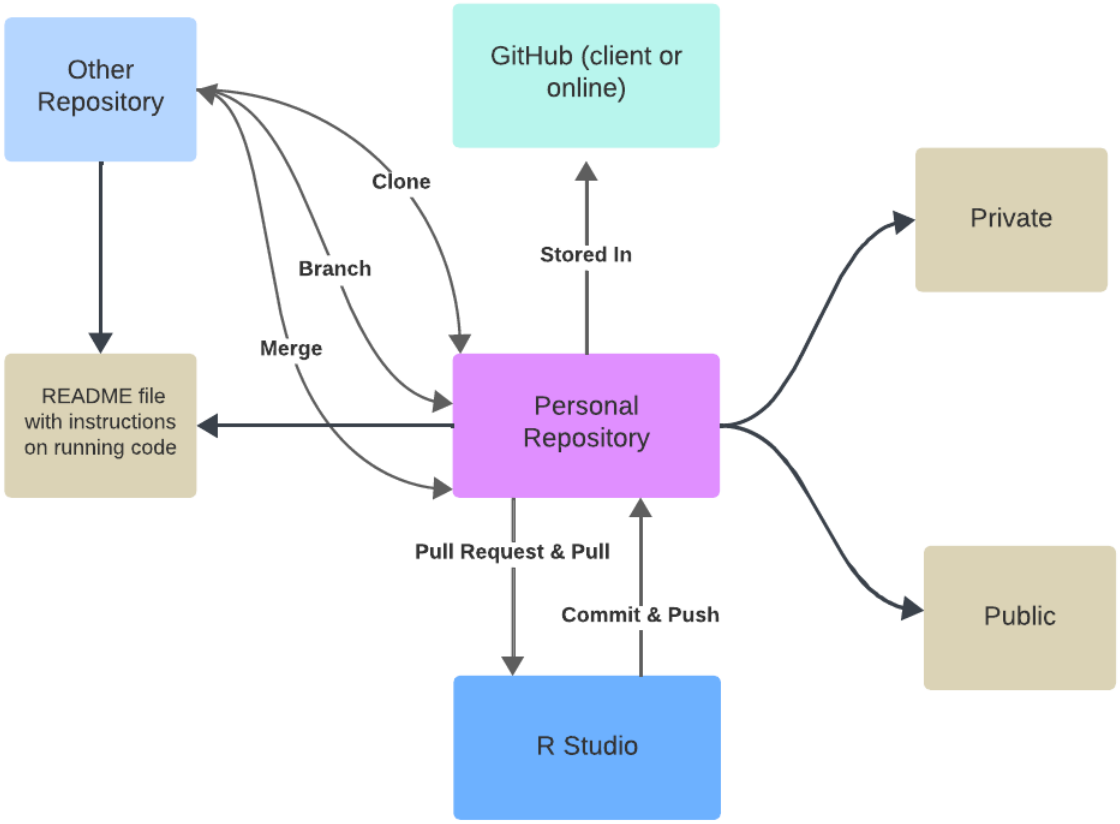

Figure 1: image

Question 1 `matrix(c(1,2,3,4,5,6) , nrow = 3)`

- Running the expression produces the following matrix

1	4
2	5
3	6

- How can you modify the call to `matrix()` to produce the following matrix instead?

	col_1	col_2
row_1	1	2
row_2	3	4
row_3	5	6

- Note that you need to name the columns (col_1 and col_2) and name the rows (row_1, row_2, row_3)

Hint: Use the `?` symbol to invoke the matrix documentation

```
matrix <- matrix(c(1,2,3,4,5,6), nrow = 3, byrow = TRUE)

colnames(matrix) <- c("col_1", "col_2")
rownames(matrix) <- c("row_1", "row_2", "row_3")

matrix
```

```
##       col_1 col_2
## row_1     1     2
## row_2     3     4
## row_3     5     6
```

Question 1 `matrix(c(1,2,3,4,5,6) , nrow = 3)`

- Running the expression produces the following matrix

1	4
2	5
3	6

- How can you modify the call to `matrix()` to produce the following matrix instead?

	col_1	col_2
row_1	1	2
row_2	3	4
row_3	5	6

- Note that you need to name the columns (col_1 and col_2) and name the rows (row_1, row_2, row_3)

Hint: Use the `?` symbol to invoke the matrix documentation

```
matrix <- matrix(c(1,2,3,4,5,6), nrow = 3, byrow = TRUE)

colnames(matrix) <- c("col_1", "col_2")
rownames(matrix) <- c("row_1", "row_2", "row_3")

matrix
```

```
##       col_1 col_2
## row_1     1     2
## row_2     3     4
## row_3     5     6
```

Question 2

- Load then sort the airquality data frame on its `Temp` and `Solar.R` columns in reverse order (largest to smallest values)
  - The function to sort a data frame is called order
- Display only the first 15 lines of your table

```
aq <- airquality

#sorting data frame by 'Temp', then `Solar.R`, in reverse order
aq_rev1 <- aq[order(-aq[, 4], -aq[, 2]), ]

#displaying first 15 lines
head(x = aq_rev1, n = 15)
```

```
##     Ozone Solar.R Wind Temp Month Day
## 120    76     203  9.7   97     8  28
## 122    84     237  6.3   96     8  30
## 121   118     225  2.3   94     8  29
## 123    85     188  6.3   94     8  31
## 42     NA     259 10.9   93     6  11
## 127    91     189  4.6   93     9   4
## 126    73     183  2.8   93     9   3
## 70     97     272  5.7   92     7   9
## 69     97     267  6.3   92     7   8
## 43     NA     250  9.2   92     6  12
## 102    NA     222  8.6   92     8  10
## 125    78     197  5.1   92     9   2
## 75     NA     291 14.9   91     7  14
## 124    96     167  6.9   91     9   1
## 40     71     291 13.8   90     6   9
```

```
#sorting data frame by 'Solar.R', then `Temp`, in reverse order
aq_rev2 <- aq[order(-aq[, 2], -aq[, 4]), ]

#displaying first 15 lines
head(x = aq_rev2, n = 15)
```

```
##    Ozone Solar.R Wind Temp Month Day
## 16    14     334 11.5   64     5  16
## 45    NA     332 13.8   80     6  14
## 41    39     323 11.5   87     6  10
## 46    NA     322 11.5   79     6  15
## 19    30     322 11.5   68     5  19
## 22    11     320 16.6   73     5  22
## 67    40     314 10.9   83     7   6
## 4     18     313 11.5   62     5   4
## 17    34     307 12.0   66     5  17
## 7     23     299  8.6   65     5   7
## 84    NA     295 11.5   82     7  23
## 85    80     294  8.6   86     7  24
## 75    NA     291 14.9   91     7  14
## 40    71     291 13.8   90     6   9
## 13    11     290  9.2   66     5  13
```

Question 3

- Sort the airquality data frame on its `Temp` in decreasing order and `Solar.R` in increasing order
- Display only the first 15 lines of your table

```
aq <- airquality

#sorting data frame by 'Temp' in decreasing order, then `Solar.R` in increasing order
aq_rev3 <- aq[order(-aq[, 4], aq[, 2]), ]

#displaying first 15 lines
head(x = aq_rev3, n = 15)
```

```
##     Ozone Solar.R Wind Temp Month Day
## 120    76     203  9.7   97     8  28
## 122    84     237  6.3   96     8  30
## 123    85     188  6.3   94     8  31
## 121   118     225  2.3   94     8  29
## 126    73     183  2.8   93     9   3
## 127    91     189  4.6   93     9   4
## 42     NA     259 10.9   93     6  11
## 125    78     197  5.1   92     9   2
## 102    NA     222  8.6   92     8  10
## 43     NA     250  9.2   92     6  12
## 69     97     267  6.3   92     7   8
## 70     97     272  5.7   92     7   9
## 124    96     167  6.9   91     9   1
## 75     NA     291 14.9   91     7  14
## 101   110     207  8.0   90     8   9
```

```
#sorting data frame by `Solar.R` in increasing order, then `Temp` in decreasing order
aq_rev4 <- aq[order(aq[, 2], -aq[, 4]), ]

#displaying first 15 lines
head(x = aq_rev4, n = 15)
```

```
##     Ozone Solar.R Wind Temp Month Day
## 82     16       7  6.9   74     7  21
## 21      1       8  9.7   59     5  21
## 28     23      13 12.0   67     5  28
## 145    23      14  9.2   71     9  22
## 9       8      19 20.1   61     5   9
## 148    14      20 16.6   63     9  25
## 94      9      24 13.8   81     8   2
## 137     9      24 10.9   71     9  14
## 23      4      25  9.7   61     5  23
## 141    13      27 10.3   76     9  18
## 60     NA      31 14.9   77     6  29
## 114     9      36 14.3   72     8  22
## 49     20      37  9.2   65     6  18
## 20     11      44  9.7   62     5  20
## 58     NA      47 10.3   73     6  27
```

Question 4

- There are various ways to select a subset of observations from a data frame.
- Consult your R Reference Card, see `Data Selection and Manipulation` section.
  - What operations can you use to select all observations where the temperature is 72. Give at least two different answers to this question

```
#selecting all observations where Temp = 72 with "which" function
T72.w <- aq[which(aq$Temp == 72), ]
head(T72.w)
```

```
##     Ozone Solar.R Wind Temp Month Day
## 2      36     118  8.0   72     5   2
## 48     37     284 20.7   72     6  17
## 114     9      36 14.3   72     8  22
```

```
#selecting all observations where Temp = 72 with "subset" function
T72.s <- subset(x = aq, Temp == 72)
head(T72.s)
```

```
##     Ozone Solar.R Wind Temp Month Day
## 2      36     118  8.0   72     5   2
## 48     37     284 20.7   72     6  17
## 114     9      36 14.3   72     8  22
```

Question 6

- You may have noticed when working with the `airqulity` data that some values show as `NA`
- `NA` stands for not available, or missing values.
- A major part of data wrangling consists of cleaning missing values by either:
  - Dropping the lines that have missing values
  - Sometimes we can drop the column with missing values if the column is made of predominantly missing values
  - Imputing the missing values, which uses educated guesses (or more complex algorithms) to fill the missing values
- Find and remove all rows that are missing values for the `Solar.R` or `Ozone` variables
- Save the cleaned data to a new data frame called airquality_no_na
  - How many lines have been removed?

```
#creating dataframe with "NA" values from Solar.R and Ozone columns removed using "which" function
airquality_no_na <- aq[which(aq$Solar.R != "NA", aq$Ozone != "NA" ), ]

#7 lines were removed
lines.removed = nrow(aq) - nrow(airquality_no_na)
lines.removed
```

```
## [1] 7
```

Question 7

- Let's use a different strategy and impute the missing value.
  - replace the missing values for Solar.R using that month's average.
  - Example:
    - The missing value for line 6 should be replaced with the average for month 5.
    - The missing value for line 97 should be replaced with the average for month 8.

```
#saving dataframe in variable
aq <- airquality

#Writing function that takes input "x" (month #) to subset aq into a smaller dataframe with data for on
na.rm <- function(x) {
  Mo <- aq[which(aq$Month == x), ]
  AvgMo <- mean(Mo$Solar.R, na.rm = TRUE)
  Mo[["Solar.R"]][is.na(Mo[["Solar.R"]])] <- AvgMo
  print(Mo)
}

#apply function across all month # in dataframe
Mo5 <- na.rm(x = 5)
```

```
##    Ozone  Solar.R Wind Temp Month Day
## 1     41 190.0000  7.4   67     5   1
## 2     36 118.0000  8.0   72     5   2
## 3     12 149.0000 12.6   74     5   3
## 4     18 313.0000 11.5   62     5   4
## 5     NA 181.2963 14.3   56     5   5
## 6     28 181.2963 14.9   66     5   6
## 7     23 299.0000  8.6   65     5   7
## 8     19  99.0000 13.8   59     5   8
## 9      8  19.0000 20.1   61     5   9
## 10    NA 194.0000  8.6   69     5  10
## 11     7 181.2963  6.9   74     5  11
## 12    16 256.0000  9.7   69     5  12
## 13    11 290.0000  9.2   66     5  13
## 14    14 274.0000 10.9   68     5  14
## 15    18  65.0000 13.2   58     5  15
## 16    14 334.0000 11.5   64     5  16
## 17    34 307.0000 12.0   66     5  17
## 18     6  78.0000 18.4   57     5  18
## 19    30 322.0000 11.5   68     5  19
## 20    11  44.0000  9.7   62     5  20
## 21     1   8.0000  9.7   59     5  21
## 22    11 320.0000 16.6   73     5  22
## 23     4  25.0000  9.7   61     5  23
## 24    32  92.0000 12.0   61     5  24
## 25    NA  66.0000 16.6   57     5  25
## 26    NA 266.0000 14.9   58     5  26
## 27    NA 181.2963  8.0   57     5  27
## 28    23  13.0000 12.0   67     5  28
## 29    45 252.0000 14.9   81     5  29
## 30   115 223.0000  5.7   79     5  30
## 31    37 279.0000  7.4   76     5  31
```

```
Mo6 <- na.rm(x = 6)
```

```
##    Ozone Solar.R Wind Temp Month Day
## 32    NA     286  8.6   78     6   1
## 33    NA     287  9.7   74     6   2
## 34    NA     242 16.1   67     6   3
## 35    NA     186  9.2   84     6   4
## 36    NA     220  8.6   85     6   5
## 37    NA     264 14.3   79     6   6
## 38    29     127  9.7   82     6   7
## 39    NA     273  6.9   87     6   8
## 40    71     291 13.8   90     6   9
## 41    39     323 11.5   87     6  10
## 42    NA     259 10.9   93     6  11
## 43    NA     250  9.2   92     6  12
## 44    23     148  8.0   82     6  13
## 45    NA     332 13.8   80     6  14
## 46    NA     322 11.5   79     6  15
## 47    21     191 14.9   77     6  16
## 48    37     284 20.7   72     6  17
## 49    20      37  9.2   65     6  18
## 50    12     120 11.5   73     6  19
## 51    13     137 10.3   76     6  20
## 52    NA     150  6.3   77     6  21
## 53    NA      59  1.7   76     6  22
## 54    NA      91  4.6   76     6  23
## 55    NA     250  6.3   76     6  24
## 56    NA     135  8.0   75     6  25
## 57    NA     127  8.0   78     6  26
## 58    NA      47 10.3   73     6  27
## 59    NA      98 11.5   80     6  28
## 60    NA      31 14.9   77     6  29
## 61    NA     138  8.0   83     6  30
```

```
Mo7 <- na.rm(x = 7)
```

```
##    Ozone Solar.R Wind Temp Month Day
## 62   135     269  4.1   84     7   1
## 63    49     248  9.2   85     7   2
## 64    32     236  9.2   81     7   3
## 65    NA     101 10.9   84     7   4
## 66    64     175  4.6   83     7   5
## 67    40     314 10.9   83     7   6
## 68    77     276  5.1   88     7   7
## 69    97     267  6.3   92     7   8
## 70    97     272  5.7   92     7   9
## 71    85     175  7.4   89     7  10
## 72    NA     139  8.6   82     7  11
## 73    10     264 14.3   73     7  12
## 74    27     175 14.9   81     7  13
## 75    NA     291 14.9   91     7  14
## 76     7      48 14.3   80     7  15
## 77    48     260  6.9   81     7  16
```

```
## 78    35     274 10.3   82     7  17
## 79    61     285  6.3   84     7  18
## 80    79     187  5.1   87     7  19
## 81    63     220 11.5   85     7  20
## 82    16       7  6.9   74     7  21
## 83    NA     258  9.7   81     7  22
## 84    NA     295 11.5   82     7  23
## 85    80     294  8.6   86     7  24
## 86   108     223  8.0   85     7  25
## 87    20      81  8.6   82     7  26
## 88    52      82 12.0   86     7  27
## 89    82     213  7.4   88     7  28
## 90    50     275  7.4   86     7  29
## 91    64     253  7.4   83     7  30
## 92    59     254  9.2   81     7  31
```

```
Mo8 <- na.rm(x = 8)
```

```
##     Ozone  Solar.R Wind Temp Month Day
## 93     39  83.0000  6.9   81     8   1
## 94      9  24.0000 13.8   81     8   2
## 95     16  77.0000  7.4   82     8   3
## 96     78 171.8571  6.9   86     8   4
## 97     35 171.8571  7.4   85     8   5
## 98     66 171.8571  4.6   87     8   6
## 99    122 255.0000  4.0   89     8   7
## 100    89 229.0000 10.3   90     8   8
## 101   110 207.0000  8.0   90     8   9
## 102    NA 222.0000  8.6   92     8  10
## 103    NA 137.0000 11.5   86     8  11
## 104    44 192.0000 11.5   86     8  12
## 105    28 273.0000 11.5   82     8  13
## 106    65 157.0000  9.7   80     8  14
## 107    NA  64.0000 11.5   79     8  15
## 108    22  71.0000 10.3   77     8  16
## 109    59  51.0000  6.3   79     8  17
## 110    23 115.0000  7.4   76     8  18
## 111    31 244.0000 10.9   78     8  19
## 112    44 190.0000 10.3   78     8  20
## 113    21 259.0000 15.5   77     8  21
## 114     9  36.0000 14.3   72     8  22
## 115    NA 255.0000 12.6   75     8  23
## 116    45 212.0000  9.7   79     8  24
## 117   168 238.0000  3.4   81     8  25
## 118    73 215.0000  8.0   86     8  26
## 119    NA 153.0000  5.7   88     8  27
## 120    76 203.0000  9.7   97     8  28
## 121   118 225.0000  2.3   94     8  29
## 122    84 237.0000  6.3   96     8  30
## 123    85 188.0000  6.3   94     8  31
```

```
Mo9 <- na.rm(x = 9)
```

```
##     Ozone Solar.R Wind Temp Month Day
```

```
## 124    96     167  6.9   91     9   1
## 125    78     197  5.1   92     9   2
## 126    73     183  2.8   93     9   3
## 127    91     189  4.6   93     9   4
## 128    47      95  7.4   87     9   5
## 129    32      92 15.5   84     9   6
## 130    20     252 10.9   80     9   7
## 131    23     220 10.3   78     9   8
## 132    21     230 10.9   75     9   9
## 133    24     259  9.7   73     9  10
## 134    44     236 14.9   81     9  11
## 135    21     259 15.5   76     9  12
## 136    28     238  6.3   77     9  13
## 137     9      24 10.9   71     9  14
## 138    13     112 11.5   71     9  15
## 139    46     237  6.9   78     9  16
## 140    18     224 13.8   67     9  17
## 141    13      27 10.3   76     9  18
## 142    24     238 10.3   68     9  19
## 143    16     201  8.0   82     9  20
## 144    13     238 12.6   64     9  21
## 145    23      14  9.2   71     9  22
## 146    36     139 10.3   81     9  23
## 147     7      49 10.3   69     9  24
## 148    14      20 16.6   63     9  25
## 149    30     193  6.9   70     9  26
## 150    NA     145 13.2   77     9  27
## 151    14     191 14.3   75     9  28
## 152    18     131  8.0   76     9  29
## 153    20     223 11.5   68     9  30
```

```
#concatenate dataframes
new.df <- rbind(Mo5, Mo6, Mo7, Mo8, Mo9)
new.df
```

```
##     Ozone  Solar.R Wind Temp Month Day
## 1      41 190.0000  7.4   67     5   1
## 2      36 118.0000  8.0   72     5   2
## 3      12 149.0000 12.6   74     5   3
## 4      18 313.0000 11.5   62     5   4
## 5      NA 181.2963 14.3   56     5   5
## 6      28 181.2963 14.9   66     5   6
## 7      23 299.0000  8.6   65     5   7
## 8      19  99.0000 13.8   59     5   8
## 9       8  19.0000 20.1   61     5   9
## 10     NA 194.0000  8.6   69     5  10
## 11      7 181.2963  6.9   74     5  11
## 12     16 256.0000  9.7   69     5  12
## 13     11 290.0000  9.2   66     5  13
## 14     14 274.0000 10.9   68     5  14
## 15     18  65.0000 13.2   58     5  15
## 16     14 334.0000 11.5   64     5  16
## 17     34 307.0000 12.0   66     5  17
## 18      6  78.0000 18.4   57     5  18
```

```
## 19     30 322.0000 11.5   68     5  19
## 20     11  44.0000  9.7   62     5  20
## 21      1   8.0000  9.7   59     5  21
## 22     11 320.0000 16.6   73     5  22
## 23      4  25.0000  9.7   61     5  23
## 24     32  92.0000 12.0   61     5  24
## 25     NA  66.0000 16.6   57     5  25
## 26     NA 266.0000 14.9   58     5  26
## 27     NA 181.2963  8.0   57     5  27
## 28     23  13.0000 12.0   67     5  28
## 29     45 252.0000 14.9   81     5  29
## 30    115 223.0000  5.7   79     5  30
## 31     37 279.0000  7.4   76     5  31
## 32     NA 286.0000  8.6   78     6   1
## 33     NA 287.0000  9.7   74     6   2
## 34     NA 242.0000 16.1   67     6   3
## 35     NA 186.0000  9.2   84     6   4
## 36     NA 220.0000  8.6   85     6   5
## 37     NA 264.0000 14.3   79     6   6
## 38     29 127.0000  9.7   82     6   7
## 39     NA 273.0000  6.9   87     6   8
## 40     71 291.0000 13.8   90     6   9
## 41     39 323.0000 11.5   87     6  10
## 42     NA 259.0000 10.9   93     6  11
## 43     NA 250.0000  9.2   92     6  12
## 44     23 148.0000  8.0   82     6  13
## 45     NA 332.0000 13.8   80     6  14
## 46     NA 322.0000 11.5   79     6  15
## 47     21 191.0000 14.9   77     6  16
## 48     37 284.0000 20.7   72     6  17
## 49     20  37.0000  9.2   65     6  18
## 50     12 120.0000 11.5   73     6  19
## 51     13 137.0000 10.3   76     6  20
## 52     NA 150.0000  6.3   77     6  21
## 53     NA  59.0000  1.7   76     6  22
## 54     NA  91.0000  4.6   76     6  23
## 55     NA 250.0000  6.3   76     6  24
## 56     NA 135.0000  8.0   75     6  25
## 57     NA 127.0000  8.0   78     6  26
## 58     NA  47.0000 10.3   73     6  27
## 59     NA  98.0000 11.5   80     6  28
## 60     NA  31.0000 14.9   77     6  29
## 61     NA 138.0000  8.0   83     6  30
## 62    135 269.0000  4.1   84     7   1
## 63     49 248.0000  9.2   85     7   2
## 64     32 236.0000  9.2   81     7   3
## 65     NA 101.0000 10.9   84     7   4
## 66     64 175.0000  4.6   83     7   5
## 67     40 314.0000 10.9   83     7   6
## 68     77 276.0000  5.1   88     7   7
## 69     97 267.0000  6.3   92     7   8
## 70     97 272.0000  5.7   92     7   9
## 71     85 175.0000  7.4   89     7  10
## 72     NA 139.0000  8.6   82     7  11
```

```
## 73     10 264.0000 14.3   73     7  12
## 74     27 175.0000 14.9   81     7  13
## 75     NA 291.0000 14.9   91     7  14
## 76      7  48.0000 14.3   80     7  15
## 77     48 260.0000  6.9   81     7  16
## 78     35 274.0000 10.3   82     7  17
## 79     61 285.0000  6.3   84     7  18
## 80     79 187.0000  5.1   87     7  19
## 81     63 220.0000 11.5   85     7  20
## 82     16   7.0000  6.9   74     7  21
## 83     NA 258.0000  9.7   81     7  22
## 84     NA 295.0000 11.5   82     7  23
## 85     80 294.0000  8.6   86     7  24
## 86    108 223.0000  8.0   85     7  25
## 87     20  81.0000  8.6   82     7  26
## 88     52  82.0000 12.0   86     7  27
## 89     82 213.0000  7.4   88     7  28
## 90     50 275.0000  7.4   86     7  29
## 91     64 253.0000  7.4   83     7  30
## 92     59 254.0000  9.2   81     7  31
## 93     39  83.0000  6.9   81     8   1
## 94      9  24.0000 13.8   81     8   2
## 95     16  77.0000  7.4   82     8   3
## 96     78 171.8571  6.9   86     8   4
## 97     35 171.8571  7.4   85     8   5
## 98     66 171.8571  4.6   87     8   6
## 99    122 255.0000  4.0   89     8   7
## 100    89 229.0000 10.3   90     8   8
## 101   110 207.0000  8.0   90     8   9
## 102    NA 222.0000  8.6   92     8  10
## 103    NA 137.0000 11.5   86     8  11
## 104    44 192.0000 11.5   86     8  12
## 105    28 273.0000 11.5   82     8  13
## 106    65 157.0000  9.7   80     8  14
## 107    NA  64.0000 11.5   79     8  15
## 108    22  71.0000 10.3   77     8  16
## 109    59  51.0000  6.3   79     8  17
## 110    23 115.0000  7.4   76     8  18
## 111    31 244.0000 10.9   78     8  19
## 112    44 190.0000 10.3   78     8  20
## 113    21 259.0000 15.5   77     8  21
## 114     9  36.0000 14.3   72     8  22
## 115    NA 255.0000 12.6   75     8  23
## 116    45 212.0000  9.7   79     8  24
## 117   168 238.0000  3.4   81     8  25
## 118    73 215.0000  8.0   86     8  26
## 119    NA 153.0000  5.7   88     8  27
## 120    76 203.0000  9.7   97     8  28
## 121   118 225.0000  2.3   94     8  29
## 122    84 237.0000  6.3   96     8  30
## 123    85 188.0000  6.3   94     8  31
## 124    96 167.0000  6.9   91     9   1
## 125    78 197.0000  5.1   92     9   2
## 126    73 183.0000  2.8   93     9   3
```

```
## 127    91 189.0000  4.6   93     9   4
## 128    47  95.0000  7.4   87     9   5
## 129    32  92.0000 15.5   84     9   6
## 130    20 252.0000 10.9   80     9   7
## 131    23 220.0000 10.3   78     9   8
## 132    21 230.0000 10.9   75     9   9
## 133    24 259.0000  9.7   73     9  10
## 134    44 236.0000 14.9   81     9  11
## 135    21 259.0000 15.5   76     9  12
## 136    28 238.0000  6.3   77     9  13
## 137     9  24.0000 10.9   71     9  14
## 138    13 112.0000 11.5   71     9  15
## 139    46 237.0000  6.9   78     9  16
## 140    18 224.0000 13.8   67     9  17
## 141    13  27.0000 10.3   76     9  18
## 142    24 238.0000 10.3   68     9  19
## 143    16 201.0000  8.0   82     9  20
## 144    13 238.0000 12.6   64     9  21
## 145    23  14.0000  9.2   71     9  22
## 146    36 139.0000 10.3   81     9  23
## 147     7  49.0000 10.3   69     9  24
## 148    14  20.0000 16.6   63     9  25
## 149    30 193.0000  6.9   70     9  26
## 150    NA 145.0000 13.2   77     9  27
## 151    14 191.0000 14.3   75     9  28
## 152    18 131.0000  8.0   76     9  29
## 153    20 223.0000 11.5   68     9  30
```