Datacamp_Cleaning Data in R_Introduction and exploring raw data

dizhen

2019/4/4

Introduction to cleaning data in R

```
weather <- readRDS("data/weather.rds")

# View the first 6 rows of data
head(weather)
```

```
##   X year month           measure X1 X2 X3 X4 X5 X6 X7 X8 X9 X10 X11 X12
## 1 1 2014    12  Max.TemperatureF 64 42 51 43 42 45 38 29 49  48  39  39
## 2 2 2014    12 Mean.TemperatureF 52 38 44 37 34 42 30 24 39  43  36  35
## 3 3 2014    12  Min.TemperatureF 39 33 37 30 26 38 21 18 29  38  32  31
## 4 4 2014    12    Max.Dew.PointF 46 40 49 24 37 45 36 28 49  45  37  28
## 5 5 2014    12    MeanDew.PointF 40 27 42 21 25 40 20 16 41  39  31  27
## 6 6 2014    12     Min.DewpointF 26 17 24 13 12 36 -3  3 28  37  27  25
##   X13 X14 X15 X16 X17 X18 X19 X20 X21 X22 X23 X24 X25 X26 X27 X28 X29 X30
## 1  42  45  42  44  49  44  37  36  36  44  47  46  59  50  52  52  41  30
## 2  37  39  37  40  45  40  33  32  33  39  45  44  52  44  45  46  36  26
## 3  32  33  32  35  41  36  29  27  30  33  42  41  44  37  38  40  30  22
## 4  28  29  33  42  46  34  25  30  30  39  45  46  58  31  34  42  26  10
## 5  26  27  29  36  41  30  22  24  27  34  42  44  43  29  31  35  20   4
## 6  24  25  27  30  32  26  20  20  25  25  37  41  29  28  29  27  10  -6
##   X31
## 1  30
## 2  25
## 3  20
## 4   8
## 5   5
## 6   1
```

```
# View the last 6 rows of data
tail(weather)
```

```
##       X year month            measure   X1   X2   X3   X4   X5   X6   X7
## 281 281 2015    12 Mean.Wind.SpeedMPH    6 <NA> <NA> <NA> <NA> <NA> <NA>
## 282 282 2015    12  Max.Gust.SpeedMPH   17 <NA> <NA> <NA> <NA> <NA> <NA>
## 283 283 2015    12    PrecipitationIn 0.14 <NA> <NA> <NA> <NA> <NA> <NA>
## 284 284 2015    12         CloudCover    7 <NA> <NA> <NA> <NA> <NA> <NA>
## 285 285 2015    12             Events Rain <NA> <NA> <NA> <NA> <NA> <NA>
## 286 286 2015    12     WindDirDegrees  109 <NA> <NA> <NA> <NA> <NA> <NA>
##       X8   X9  X10  X11  X12  X13  X14  X15  X16  X17  X18  X19  X20  X21
## 281 <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
## 282 <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
## 283 <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
```

```
## 284 <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
## 285 <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
## 286 <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
##      X22  X23  X24  X25  X26  X27  X28  X29  X30  X31
## 281 <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
## 282 <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
## 283 <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
## 284 <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
## 285 <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
## 286 <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
```

```r
# View a condensed summary of the data
str(weather)
```

```
## 'data.frame':    286 obs. of  35 variables:
##  $ X      : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ year   : int  2014 2014 2014 2014 2014 2014 2014 2014 2014 2014 ...
##  $ month  : int  12 12 12 12 12 12 12 12 12 12 ...
##  $ measure: chr  "Max.TemperatureF" "Mean.TemperatureF" "Min.TemperatureF" "Max.Dew.PointF" ...
##  $ X1     : chr  "64" "52" "39" "46" ...
##  $ X2     : chr  "42" "38" "33" "40" ...
##  $ X3     : chr  "51" "44" "37" "49" ...
##  $ X4     : chr  "43" "37" "30" "24" ...
##  $ X5     : chr  "42" "34" "26" "37" ...
##  $ X6     : chr  "45" "42" "38" "45" ...
##  $ X7     : chr  "38" "30" "21" "36" ...
##  $ X8     : chr  "29" "24" "18" "28" ...
##  $ X9     : chr  "49" "39" "29" "49" ...
##  $ X10    : chr  "48" "43" "38" "45" ...
##  $ X11    : chr  "39" "36" "32" "37" ...
##  $ X12    : chr  "39" "35" "31" "28" ...
##  $ X13    : chr  "42" "37" "32" "28" ...
##  $ X14    : chr  "45" "39" "33" "29" ...
##  $ X15    : chr  "42" "37" "32" "33" ...
##  $ X16    : chr  "44" "40" "35" "42" ...
##  $ X17    : chr  "49" "45" "41" "46" ...
##  $ X18    : chr  "44" "40" "36" "34" ...
##  $ X19    : chr  "37" "33" "29" "25" ...
##  $ X20    : chr  "36" "32" "27" "30" ...
##  $ X21    : chr  "36" "33" "30" "30" ...
##  $ X22    : chr  "44" "39" "33" "39" ...
##  $ X23    : chr  "47" "45" "42" "45" ...
##  $ X24    : chr  "46" "44" "41" "46" ...
##  $ X25    : chr  "59" "52" "44" "58" ...
##  $ X26    : chr  "50" "44" "37" "31" ...
##  $ X27    : chr  "52" "45" "38" "34" ...
##  $ X28    : chr  "52" "46" "40" "42" ...
##  $ X29    : chr  "41" "36" "30" "26" ...
##  $ X30    : chr  "30" "26" "22" "10" ...
##  $ X31    : chr  "30" "25" "20" "8" ...
```

Exploring raw data

1. Understanding the structure of your data

class() - Class of data object

dim() - Dimensions of data

names() - Column names

str() - Preview of data with helpful details

glimpse() - Be!er version of str() from dplyr

summary() - Summary of data

2. Looking at your data

head() - View top of dataset

tail() - View bo!om of dataset

print() - View entire dataset (not recommended!)

3. Visualizing your data

hist() - View histogram of a single variable

plot() - View plot of two variables

Practice

```
library(readr)
bmi <- read_csv("data/bmi_clean.csv")
```

```
## Parsed with column specification:
## cols(
##   .default = col_double(),
##   Country = col_character()
## )
```

```
## See spec(...) for full column specifications.
```

```
# Check the class of bmi
class(bmi)
```

```
## [1] "spec_tbl_df" "tbl_df"      "tbl"         "data.frame"
```

```
# Check the dimensions of bmi
dim(bmi)
```

```
## [1] 199  30
```

```r
# View the column names of bmi
names(bmi)
```

```
##  [1] "Country" "Y1980"   "Y1981"   "Y1982"   "Y1983"   "Y1984"   "Y1985"
##  [8] "Y1986"   "Y1987"   "Y1988"   "Y1989"   "Y1990"   "Y1991"   "Y1992"
## [15] "Y1993"   "Y1994"   "Y1995"   "Y1996"   "Y1997"   "Y1998"   "Y1999"
## [22] "Y2000"   "Y2001"   "Y2002"   "Y2003"   "Y2004"   "Y2005"   "Y2006"
## [29] "Y2007"   "Y2008"
```

```r
# Check the structure of bmi
str(bmi)
```

```
## Classes 'spec_tbl_df', 'tbl_df', 'tbl' and 'data.frame': 199 obs. of  30 variables:
##  $ Country: chr  "Afghanistan" "Albania" "Algeria" "Andorra" ...
##  $ Y1980  : num  21.5 25.2 22.3 25.7 20.9 ...
##  $ Y1981  : num  21.5 25.2 22.3 25.7 20.9 ...
##  $ Y1982  : num  21.5 25.3 22.4 25.7 20.9 ...
##  $ Y1983  : num  21.4 25.3 22.5 25.8 20.9 ...
##  $ Y1984  : num  21.4 25.3 22.6 25.8 20.9 ...
##  $ Y1985  : num  21.4 25.3 22.7 25.9 20.9 ...
##  $ Y1986  : num  21.4 25.3 22.8 25.9 21 ...
##  $ Y1987  : num  21.4 25.3 22.8 25.9 21 ...
##  $ Y1988  : num  21.3 25.3 22.9 26 21 ...
##  $ Y1989  : num  21.3 25.3 23 26 21.1 ...
##  $ Y1990  : num  21.2 25.3 23 26.1 21.1 ...
##  $ Y1991  : num  21.2 25.3 23.1 26.2 21.1 ...
##  $ Y1992  : num  21.1 25.2 23.2 26.2 21.1 ...
##  $ Y1993  : num  21.1 25.2 23.3 26.3 21.1 ...
##  $ Y1994  : num  21 25.2 23.3 26.4 21.1 ...
##  $ Y1995  : num  20.9 25.3 23.4 26.4 21.2 ...
##  $ Y1996  : num  20.9 25.3 23.5 26.5 21.2 ...
##  $ Y1997  : num  20.8 25.3 23.5 26.6 21.2 ...
##  $ Y1998  : num  20.8 25.4 23.6 26.7 21.3 ...
##  $ Y1999  : num  20.8 25.5 23.7 26.8 21.3 ...
##  $ Y2000  : num  20.7 25.6 23.8 26.8 21.4 ...
##  $ Y2001  : num  20.6 25.7 23.9 26.9 21.4 ...
##  $ Y2002  : num  20.6 25.8 24 27 21.5 ...
##  $ Y2003  : num  20.6 25.9 24.1 27.1 21.6 ...
##  $ Y2004  : num  20.6 26 24.2 27.2 21.7 ...
##  $ Y2005  : num  20.6 26.1 24.3 27.3 21.8 ...
##  $ Y2006  : num  20.6 26.2 24.4 27.4 21.9 ...
##  $ Y2007  : num  20.6 26.3 24.5 27.5 22.1 ...
##  $ Y2008  : num  20.6 26.4 24.6 27.6 22.3 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   Country = col_character(),
##   ..   Y1980 = col_double(),
##   ..   Y1981 = col_double(),
##   ..   Y1982 = col_double(),
##   ..   Y1983 = col_double(),
##   ..   Y1984 = col_double(),
##   ..   Y1985 = col_double(),
##   ..   Y1986 = col_double(),
```

```
##   ..   Y1987 = col_double(),
##   ..   Y1988 = col_double(),
##   ..   Y1989 = col_double(),
##   ..   Y1990 = col_double(),
##   ..   Y1991 = col_double(),
##   ..   Y1992 = col_double(),
##   ..   Y1993 = col_double(),
##   ..   Y1994 = col_double(),
##   ..   Y1995 = col_double(),
##   ..   Y1996 = col_double(),
##   ..   Y1997 = col_double(),
##   ..   Y1998 = col_double(),
##   ..   Y1999 = col_double(),
##   ..   Y2000 = col_double(),
##   ..   Y2001 = col_double(),
##   ..   Y2002 = col_double(),
##   ..   Y2003 = col_double(),
##   ..   Y2004 = col_double(),
##   ..   Y2005 = col_double(),
##   ..   Y2006 = col_double(),
##   ..   Y2007 = col_double(),
##   ..   Y2008 = col_double()
##   .. )
```

```
# Load dplyr
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
# Check the structure of bmi, the dplyr way
glimpse(bmi)
```

```
## Observations: 199
## Variables: 30
## $ Country <chr> "Afghanistan", "Albania", "Algeria", "Andorra", "Angol...
## $ Y1980   <dbl> 21.48678, 25.22533, 22.25703, 25.66652, 20.94876, 23.3...
## $ Y1981   <dbl> 21.46552, 25.23981, 22.34745, 25.70868, 20.94371, 23.3...
## $ Y1982   <dbl> 21.45145, 25.25636, 22.43647, 25.74681, 20.93754, 23.4...
## $ Y1983   <dbl> 21.43822, 25.27176, 22.52105, 25.78250, 20.93187, 23.5...
## $ Y1984   <dbl> 21.42734, 25.27901, 22.60633, 25.81874, 20.93569, 23.6...
## $ Y1985   <dbl> 21.41222, 25.28669, 22.69501, 25.85236, 20.94857, 23.7...
## $ Y1986   <dbl> 21.40132, 25.29451, 22.76979, 25.89089, 20.96030, 23.8...
## $ Y1987   <dbl> 21.37679, 25.30217, 22.84096, 25.93414, 20.98025, 23.9...
```

```
## $ Y1988   <dbl> 21.34018, 25.30450, 22.90644, 25.98477, 21.01375, 24.0...
## $ Y1989   <dbl> 21.29845, 25.31944, 22.97931, 26.04450, 21.05269, 24.1...
## $ Y1990   <dbl> 21.24818, 25.32357, 23.04600, 26.10936, 21.09007, 24.2...
## $ Y1991   <dbl> 21.20269, 25.28452, 23.11333, 26.17912, 21.12136, 24.3...
## $ Y1992   <dbl> 21.14238, 25.23077, 23.18776, 26.24017, 21.14987, 24.4...
## $ Y1993   <dbl> 21.06376, 25.21192, 23.25764, 26.30356, 21.13938, 24.5...
## $ Y1994   <dbl> 20.97987, 25.22115, 23.32273, 26.36793, 21.14186, 24.6...
## $ Y1995   <dbl> 20.91132, 25.25874, 23.39526, 26.43569, 21.16022, 24.6...
## $ Y1996   <dbl> 20.85155, 25.31097, 23.46811, 26.50769, 21.19076, 24.7...
## $ Y1997   <dbl> 20.81307, 25.33988, 23.54160, 26.58255, 21.22621, 24.7...
## $ Y1998   <dbl> 20.78591, 25.39116, 23.61592, 26.66337, 21.27082, 24.8...
## $ Y1999   <dbl> 20.75469, 25.46555, 23.69486, 26.75078, 21.31954, 24.9...
## $ Y2000   <dbl> 20.69521, 25.55835, 23.77659, 26.83179, 21.37480, 24.9...
## $ Y2001   <dbl> 20.62643, 25.66701, 23.86256, 26.92373, 21.43664, 25.0...
## $ Y2002   <dbl> 20.59848, 25.77167, 23.95294, 27.02525, 21.51765, 25.1...
## $ Y2003   <dbl> 20.58706, 25.87274, 24.05243, 27.12481, 21.59924, 25.2...
## $ Y2004   <dbl> 20.57759, 25.98136, 24.15957, 27.23107, 21.69218, 25.2...
## $ Y2005   <dbl> 20.58084, 26.08939, 24.27001, 27.32827, 21.80564, 25.3...
## $ Y2006   <dbl> 20.58749, 26.20867, 24.38270, 27.43588, 21.93881, 25.5...
## $ Y2007   <dbl> 20.60246, 26.32753, 24.48846, 27.53363, 22.08962, 25.6...
## $ Y2008   <dbl> 20.62058, 26.44657, 24.59620, 27.63048, 22.25083, 25.7...
```

```
# View a summary of bmi
summary(bmi)
```

```
##    Country              Y1980           Y1981           Y1982
##  Length:199         Min.   :19.01   Min.   :19.04   Min.   :19.07
##  Class :character   1st Qu.:21.27   1st Qu.:21.31   1st Qu.:21.36
##  Mode  :character   Median :23.31   Median :23.39   Median :23.46
##                     Mean   :23.15   Mean   :23.21   Mean   :23.26
##                     3rd Qu.:24.82   3rd Qu.:24.89   3rd Qu.:24.94
##                     Max.   :28.12   Max.   :28.36   Max.   :28.58
##      Y1983           Y1984           Y1985           Y1986
##  Min.   :19.10   Min.   :19.13   Min.   :19.16   Min.   :19.20
##  1st Qu.:21.42   1st Qu.:21.45   1st Qu.:21.47   1st Qu.:21.49
##  Median :23.57   Median :23.64   Median :23.73   Median :23.82
##  Mean   :23.32   Mean   :23.37   Mean   :23.42   Mean   :23.48
##  3rd Qu.:25.02   3rd Qu.:25.06   3rd Qu.:25.11   3rd Qu.:25.20
##  Max.   :28.82   Max.   :29.05   Max.   :29.28   Max.   :29.52
##      Y1987           Y1988           Y1989           Y1990
##  Min.   :19.23   Min.   :19.27   Min.   :19.31   Min.   :19.35
##  1st Qu.:21.50   1st Qu.:21.52   1st Qu.:21.55   1st Qu.:21.57
##  Median :23.87   Median :23.93   Median :24.03   Median :24.14
##  Mean   :23.53   Mean   :23.59   Mean   :23.65   Mean   :23.71
##  3rd Qu.:25.27   3rd Qu.:25.34   3rd Qu.:25.37   3rd Qu.:25.39
##  Max.   :29.75   Max.   :29.98   Max.   :30.20   Max.   :30.42
##      Y1991           Y1992           Y1993           Y1994
##  Min.   :19.40   Min.   :19.45   Min.   :19.51   Min.   :19.59
##  1st Qu.:21.60   1st Qu.:21.65   1st Qu.:21.74   1st Qu.:21.76
##  Median :24.20   Median :24.19   Median :24.27   Median :24.36
##  Mean   :23.76   Mean   :23.82   Mean   :23.88   Mean   :23.94
##  3rd Qu.:25.42   3rd Qu.:25.48   3rd Qu.:25.54   3rd Qu.:25.62
##  Max.   :30.64   Max.   :30.85   Max.   :31.04   Max.   :31.23
##      Y1995           Y1996           Y1997           Y1998
```

```
##  Min.   :19.67   Min.   :19.71   Min.   :19.74   Min.   :19.77
##  1st Qu.:21.83   1st Qu.:21.89   1st Qu.:21.94   1st Qu.:22.00
##  Median :24.41   Median :24.42   Median :24.50   Median :24.49
##  Mean   :24.00   Mean   :24.07   Mean   :24.14   Mean   :24.21
##  3rd Qu.:25.70   3rd Qu.:25.78   3rd Qu.:25.85   3rd Qu.:25.94
##  Max.   :31.41   Max.   :31.59   Max.   :31.77   Max.   :31.95
##      Y1999           Y2000           Y2001           Y2002
##  Min.   :19.80   Min.   :19.83   Min.   :19.86   Min.   :19.84
##  1st Qu.:22.04   1st Qu.:22.12   1st Qu.:22.22   1st Qu.:22.29
##  Median :24.61   Median :24.66   Median :24.73   Median :24.81
##  Mean   :24.29   Mean   :24.36   Mean   :24.44   Mean   :24.52
##  3rd Qu.:26.01   3rd Qu.:26.09   3rd Qu.:26.19   3rd Qu.:26.30
##  Max.   :32.13   Max.   :32.32   Max.   :32.51   Max.   :32.70
##      Y2003           Y2004           Y2005           Y2006
##  Min.   :19.81   Min.   :19.79   Min.   :19.79   Min.   :19.80
##  1st Qu.:22.37   1st Qu.:22.45   1st Qu.:22.54   1st Qu.:22.63
##  Median :24.89   Median :25.00   Median :25.11   Median :25.24
##  Mean   :24.61   Mean   :24.70   Mean   :24.79   Mean   :24.89
##  3rd Qu.:26.38   3rd Qu.:26.47   3rd Qu.:26.53   3rd Qu.:26.59
##  Max.   :32.90   Max.   :33.10   Max.   :33.30   Max.   :33.49
##      Y2007           Y2008
##  Min.   :19.83   Min.   :19.87
##  1st Qu.:22.73   1st Qu.:22.83
##  Median :25.36   Median :25.50
##  Mean   :24.99   Mean   :25.10
##  3rd Qu.:26.66   3rd Qu.:26.82
##  Max.   :33.69   Max.   :33.90
```

```
# Print bmi to the console
print(bmi)
```

```
## # A tibble: 199 x 30
##    Country Y1980 Y1981 Y1982 Y1983 Y1984 Y1985 Y1986 Y1987 Y1988 Y1989
##    <chr>   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
##  1 Afghan~  21.5  21.5  21.5  21.4  21.4  21.4  21.4  21.4  21.3  21.3
##  2 Albania  25.2  25.2  25.3  25.3  25.3  25.3  25.3  25.3  25.3  25.3
##  3 Algeria  22.3  22.3  22.4  22.5  22.6  22.7  22.8  22.8  22.9  23.0
##  4 Andorra  25.7  25.7  25.7  25.8  25.8  25.9  25.9  25.9  26.0  26.0
##  5 Angola   20.9  20.9  20.9  20.9  20.9  20.9  21.0  21.0  21.0  21.1
##  6 Antigu~  23.3  23.4  23.5  23.5  23.6  23.7  23.8  23.9  24.1  24.2
##  7 Argent~  25.4  25.4  25.5  25.6  25.6  25.7  25.7  25.8  25.8  25.9
##  8 Armenia  23.8  23.9  23.9  24.0  24.0  24.0  24.1  24.1  24.2  24.2
##  9 Austra~  24.9  25.0  25.1  25.1  25.2  25.3  25.4  25.5  25.6  25.7
## 10 Austria  24.8  24.9  24.9  25.0  25.0  25.1  25.1  25.1  25.2  25.2
## # ... with 189 more rows, and 19 more variables: Y1990 <dbl>, Y1991 <dbl>,
## #   Y1992 <dbl>, Y1993 <dbl>, Y1994 <dbl>, Y1995 <dbl>, Y1996 <dbl>,
## #   Y1997 <dbl>, Y1998 <dbl>, Y1999 <dbl>, Y2000 <dbl>, Y2001 <dbl>,
## #   Y2002 <dbl>, Y2003 <dbl>, Y2004 <dbl>, Y2005 <dbl>, Y2006 <dbl>,
## #   Y2007 <dbl>, Y2008 <dbl>
```

```
# View the first 6 rows
head(bmi,n=6)
```

```
## # A tibble: 6 x 30
```

```
##   Country Y1980 Y1981 Y1982 Y1983 Y1984 Y1985 Y1986 Y1987 Y1988 Y1989 Y1990
##   <chr>   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 Afghan~  21.5  21.5  21.5  21.4  21.4  21.4  21.4  21.4  21.3  21.3  21.2
## 2 Albania  25.2  25.2  25.3  25.3  25.3  25.3  25.3  25.3  25.3  25.3  25.3
## 3 Algeria  22.3  22.3  22.4  22.5  22.6  22.7  22.8  22.8  22.9  23.0  23.0
## 4 Andorra  25.7  25.7  25.7  25.8  25.8  25.9  25.9  25.9  26.0  26.0  26.1
## 5 Angola   20.9  20.9  20.9  20.9  20.9  20.9  21.0  21.0  21.0  21.1  21.1
## 6 Antigu~  23.3  23.4  23.5  23.5  23.6  23.7  23.8  23.9  24.1  24.2  24.3
## # ... with 18 more variables: Y1991 <dbl>, Y1992 <dbl>, Y1993 <dbl>,
## #   Y1994 <dbl>, Y1995 <dbl>, Y1996 <dbl>, Y1997 <dbl>, Y1998 <dbl>,
## #   Y1999 <dbl>, Y2000 <dbl>, Y2001 <dbl>, Y2002 <dbl>, Y2003 <dbl>,
## #   Y2004 <dbl>, Y2005 <dbl>, Y2006 <dbl>, Y2007 <dbl>, Y2008 <dbl>
```

```
# View the first 15 rows
head(bmi,n=15)
```

```
## # A tibble: 15 x 30
##    Country Y1980 Y1981 Y1982 Y1983 Y1984 Y1985 Y1986 Y1987 Y1988 Y1989
##    <chr>   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
##  1 Afghan~  21.5  21.5  21.5  21.4  21.4  21.4  21.4  21.4  21.3  21.3
##  2 Albania  25.2  25.2  25.3  25.3  25.3  25.3  25.3  25.3  25.3  25.3
##  3 Algeria  22.3  22.3  22.4  22.5  22.6  22.7  22.8  22.8  22.9  23.0
##  4 Andorra  25.7  25.7  25.7  25.8  25.8  25.9  25.9  25.9  26.0  26.0
##  5 Angola   20.9  20.9  20.9  20.9  20.9  20.9  21.0  21.0  21.0  21.1
##  6 Antigu~  23.3  23.4  23.5  23.5  23.6  23.7  23.8  23.9  24.1  24.2
##  7 Argent~  25.4  25.4  25.5  25.6  25.6  25.7  25.7  25.8  25.8  25.9
##  8 Armenia  23.8  23.9  23.9  24.0  24.0  24.0  24.1  24.1  24.2  24.2
##  9 Austra~  24.9  25.0  25.1  25.1  25.2  25.3  25.4  25.5  25.6  25.7
## 10 Austria  24.8  24.9  24.9  25.0  25.0  25.1  25.1  25.1  25.2  25.2
## 11 Azerba~  24.5  24.5  24.6  24.6  24.6  24.7  24.7  24.8  24.8  24.8
## 12 Bahamas  24.2  24.3  24.4  24.5  24.7  24.8  24.9  25.0  25.1  25.3
## 13 Bahrain  24.0  24.1  24.2  24.3  24.4  24.5  24.6  24.7  24.9  25.0
## 14 Bangla~  20.5  20.5  20.4  20.4  20.4  20.3  20.3  20.3  20.2  20.2
## 15 Barbad~  24.4  24.4  24.5  24.5  24.6  24.6  24.7  24.8  24.8  24.9
## # ... with 19 more variables: Y1990 <dbl>, Y1991 <dbl>, Y1992 <dbl>,
## #   Y1993 <dbl>, Y1994 <dbl>, Y1995 <dbl>, Y1996 <dbl>, Y1997 <dbl>,
## #   Y1998 <dbl>, Y1999 <dbl>, Y2000 <dbl>, Y2001 <dbl>, Y2002 <dbl>,
## #   Y2003 <dbl>, Y2004 <dbl>, Y2005 <dbl>, Y2006 <dbl>, Y2007 <dbl>,
## #   Y2008 <dbl>
```

```
# View the last 6 rows
tail(bmi,n = 6)
```

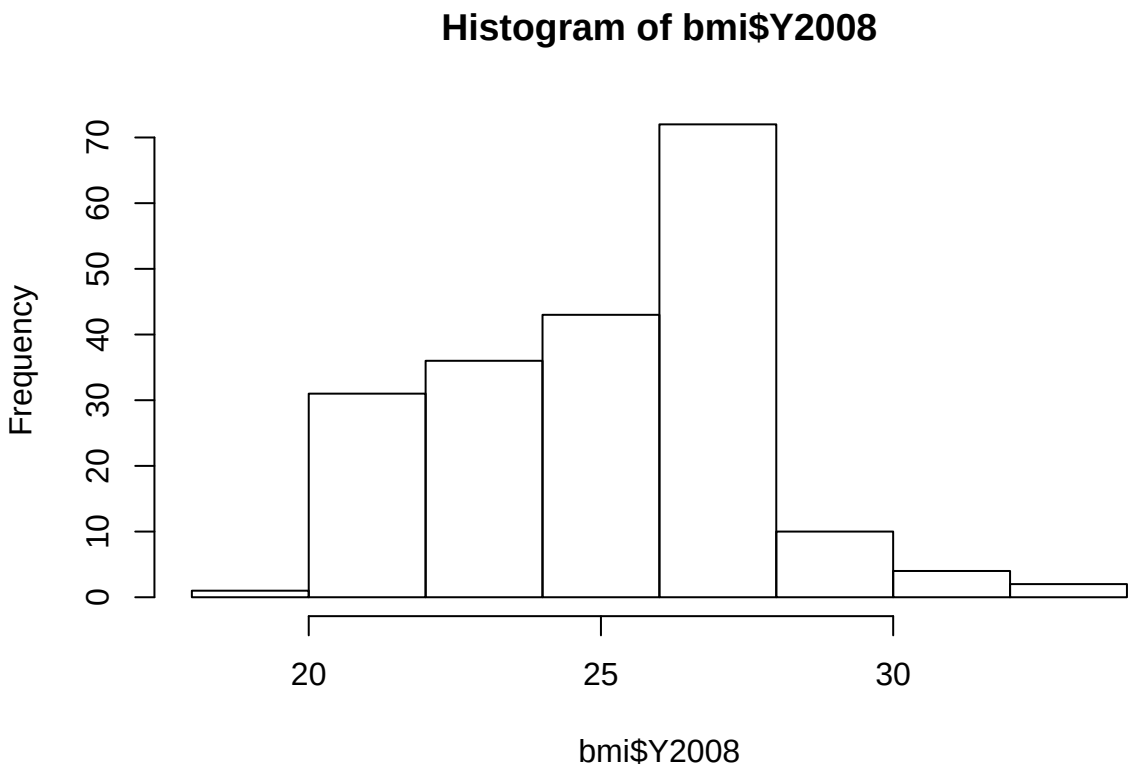
```
## # A tibble: 6 x 30
##   Country Y1980 Y1981 Y1982 Y1983 Y1984 Y1985 Y1986 Y1987 Y1988 Y1989 Y1990
##   <chr>   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 Venezu~  24.6  24.7  24.8  24.9  25.0  25.1  25.2  25.2  25.4  25.4  25.5
## 2 Vietnam  19.0  19.0  19.1  19.1  19.1  19.2  19.2  19.2  19.3  19.3  19.4
## 3 West B~  24.3  24.4  24.5  24.6  24.7  24.7  24.8  24.9  25.0  25.1  25.2
## 4 Yemen,~  22.9  23.0  23.0  23.1  23.1  23.2  23.2  23.3  23.3  23.3  23.4
## 5 Zambia   19.7  19.7  19.7  19.8  19.8  19.8  19.8  19.9  19.9  19.9  20.0
## 6 Zimbab~  21.5  21.5  21.5  21.5  21.5  21.5  21.5  21.5  21.6  21.6  21.6
## # ... with 18 more variables: Y1991 <dbl>, Y1992 <dbl>, Y1993 <dbl>,
```

```
## #   Y1994 <dbl>, Y1995 <dbl>, Y1996 <dbl>, Y1997 <dbl>, Y1998 <dbl>,
## #   Y1999 <dbl>, Y2000 <dbl>, Y2001 <dbl>, Y2002 <dbl>, Y2003 <dbl>,
## #   Y2004 <dbl>, Y2005 <dbl>, Y2006 <dbl>, Y2007 <dbl>, Y2008 <dbl>
```

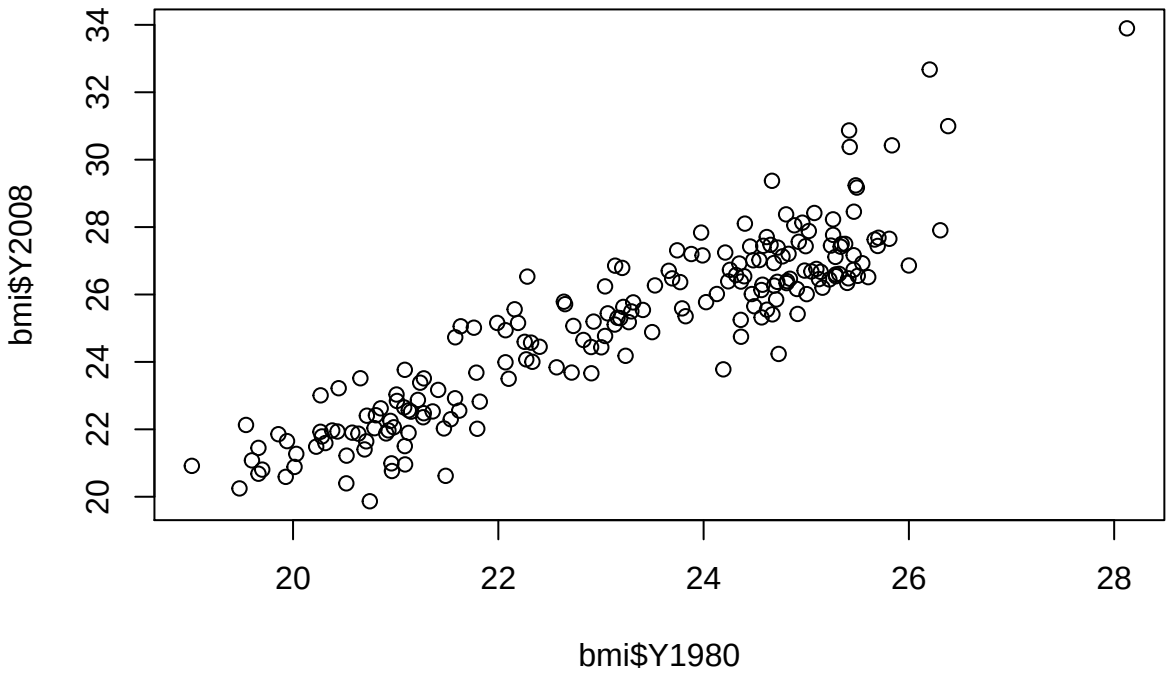
```
# View the last 10 rows
tail(bmi, n=10)
```

```
## # A tibble: 10 x 30
##    Country Y1980 Y1981 Y1982 Y1983 Y1984 Y1985 Y1986 Y1987 Y1988 Y1989
##    <chr>   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
##  1 United~  25.5  25.6  25.7  25.8  25.9  26.0  26.1  26.3  26.4  26.5
##  2 Uruguay  24.2  24.3  24.4  24.4  24.5  24.5  24.6  24.7  24.7  24.8
##  3 Uzbeki~  24.6  24.6  24.6  24.6  24.7  24.7  24.7  24.7  24.8  24.8
##  4 Vanuatu  23.2  23.3  23.5  23.6  23.8  23.9  24.0  24.2  24.3  24.4
##  5 Venezu~  24.6  24.7  24.8  24.9  25.0  25.1  25.2  25.2  25.4  25.4
##  6 Vietnam  19.0  19.0  19.1  19.1  19.1  19.2  19.2  19.2  19.3  19.3
##  7 West B~  24.3  24.4  24.5  24.6  24.7  24.7  24.8  24.9  25.0  25.1
##  8 Yemen,~  22.9  23.0  23.0  23.1  23.1  23.2  23.2  23.3  23.3  23.3
##  9 Zambia   19.7  19.7  19.7  19.8  19.8  19.8  19.8  19.9  19.9  19.9
## 10 Zimbab~  21.5  21.5  21.5  21.5  21.5  21.5  21.5  21.5  21.6  21.6
## # ... with 19 more variables: Y1990 <dbl>, Y1991 <dbl>, Y1992 <dbl>,
## #   Y1993 <dbl>, Y1994 <dbl>, Y1995 <dbl>, Y1996 <dbl>, Y1997 <dbl>,
## #   Y1998 <dbl>, Y1999 <dbl>, Y2000 <dbl>, Y2001 <dbl>, Y2002 <dbl>,
## #   Y2003 <dbl>, Y2004 <dbl>, Y2005 <dbl>, Y2006 <dbl>, Y2007 <dbl>,
## #   Y2008 <dbl>
```

```
# Histogram of BMIs from 2008
hist(bmi$Y2008)
```




```
# Scatter plot comparing BMIs from 1980 to those from 2008
plot(bmi$Y1980,bmi$Y2008)
```


bmi$Y2008
20
22
24
26
28
30
32
34
bmi$Y1980
20
22
24
26
28