description statitics

Pengfei Luo

1/22/2021

```r
rm(list=ls())
library(tidyverse)
library(data.table)
library(psych)
```

Here we try several ways (packages) to summary descriptive statistics of data.(link: https://www.youtube.com/watch?v=ZA28sOmq7nU)

load data

First, we load data from a local csv file.

```r
pokemon <- as.data.table(read.csv("Pokemon.csv",stringsAsFactors = TRUE))
pokemon
```
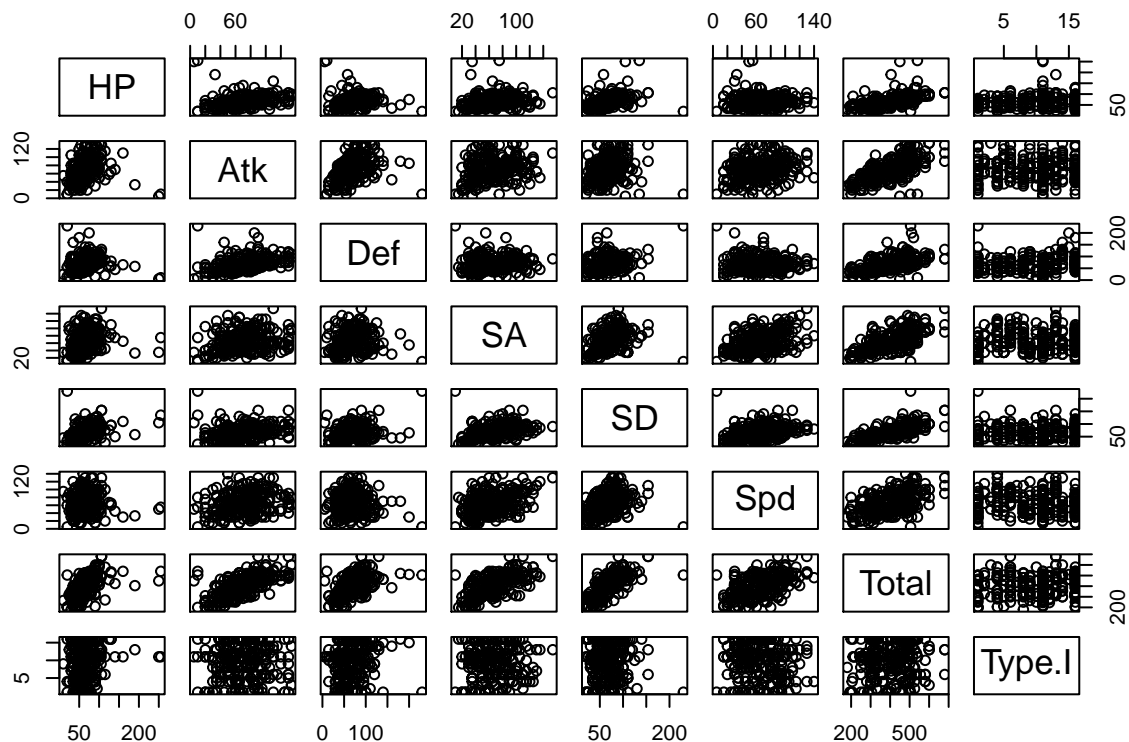
```
##       Nat    Pokemon  HP Atk Def  SA  SD Spd Total   Type.I Type.II   Gender
##   1:    1  Bulbasaur  45  49  49  65  65  45   318    Grass  Poison M (87.5%)
##   2:    2    Ivysaur  60  62  63  80  80  60   405    Grass  Poison M (87.5%)
##   3:    3   Venusaur  80  82  83 100 100  80   525    Grass  Poison M (87.5%)
##   4:    4 Charmander  39  52  43  60  50  65   309     Fire         M (87.5%)
##   5:    5  Charmeleon  58  64  58  80  65  80   405     Fire         M (87.5%)
##  ---
## 247:  247     Pupitar  70  84  70  65  70  51   410     Rock  Ground    50/50
## 248:  248   Tyranitar 100 134 110  95 100  61   600     Rock    Dark    50/50
## 249:  249       Lugia 106  90 130  90 154 110   680  Psychic  Flying     None
## 250:  250        Ho-oh 106 130  90 110 154  90   680     Fire  Flying     None
## 251:  251       Celebi 100 100 100 100 100 100   600  Psychic   Grass     None
##       Evolves.From Evolves.Into Captive
##   1:            --      Ivysaur Captive
##   2:     Bulbasaur     Venusaur    Wild
##   3:       Ivysaur           -- Captive
##   4:            --   Charmeleon Captive
##   5:     Charmander     Charizard Captive
##  ---
## 247:      Larvitar    Tyranitar Captive
## 248:       Pupitar           -- Captive
## 249:            --           --    Wild
## 250:            --           --    Wild
## 251:            --           -- Captive
```

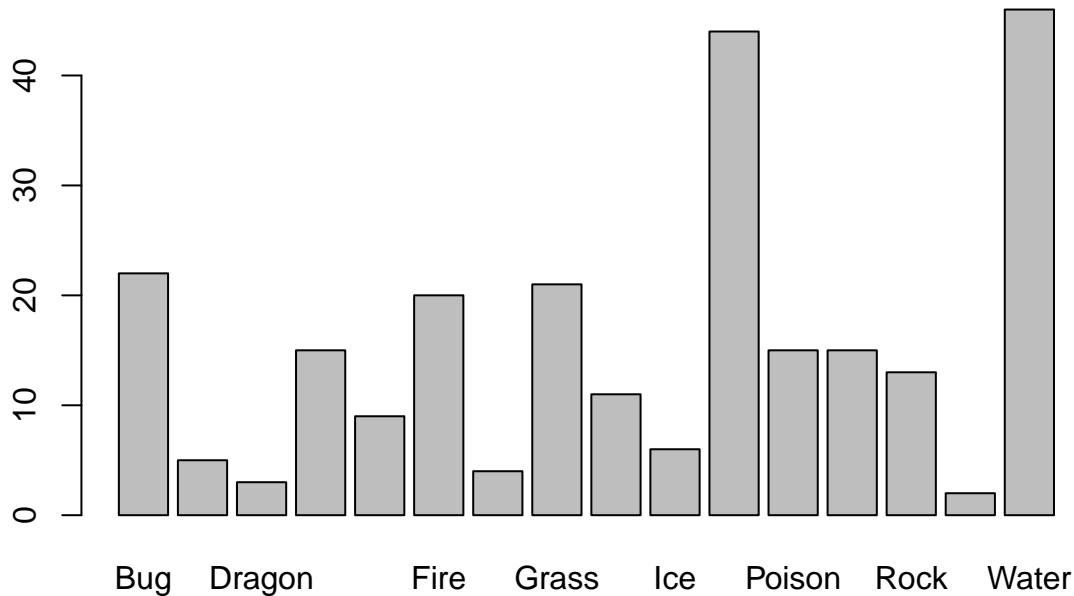```
str(as_tibble(pokemon))
```

```
## tibble [251 x 15] (S3: tbl_df/tbl/data.frame)
##  $ Nat         : int [1:251] 1 2 3 4 5 6 7 8 9 10 ...
##  $ Pokemon     : Factor w/ 251 levels "Abra","Aerodactyl",..: 17 93 236 23 24 22 209 241 15 19 ...
##  $ HP          : int [1:251] 45 60 80 39 58 78 44 59 79 45 ...
##  $ Atk         : int [1:251] 49 62 82 52 64 84 48 63 83 30 ...
##  $ Def         : int [1:251] 49 63 83 43 58 78 65 80 100 35 ...
##  $ SA          : int [1:251] 65 80 100 60 80 109 60 65 85 20 ...
##  $ SD          : int [1:251] 65 80 100 50 65 85 54 80 105 20 ...
##  $ Spd         : int [1:251] 45 60 80 65 80 100 43 58 78 45 ...
##  $ Total       : int [1:251] 318 405 525 309 405 534 314 405 530 195 ...
##  $ Type.I      : Factor w/ 16 levels "Bug","Dark","Dragon",..: 8 8 8 6 6 6 16 16 16 1 ...
##  $ Type.II     : Factor w/ 15 levels "","Dark","Dragon",..: 11 11 11 1 1 7 1 1 1 1 ...
##  $ Gender      : Factor w/ 7 levels "50/50","F (100%)",..: 6 6 6 6 6 6 6 6 6 1 ...
##  $ Evolves.From: Factor w/ 121 levels "--","Alakazam",..: 1 7 43 1 10 11 1 104 116 1 ...
##  $ Evolves.Into: Factor w/ 130 levels "--","Alakazam",..: 51 121 1 15 14 1 124 11 1 72 ...
##  $ Captive     : Factor w/ 2 levels "Captive","Wild": 1 2 1 1 1 1 1 2 1 2 ...
##  - attr(*, ".internal.selfref")=<externalptr>
```

plot variables in dataframe

```
# plot(pokemon)
plot(pokemon[,3:10])
```

```
plot(as_tibble(pokemon[,"Type.I"],pokemon[,'Atk']))
```


descriptive statistics for our data, pokemon

summary() function from basic package

```
summary(pokemon)
```

```
##       Nat              Pokemon         HP              Atk
##  Min.   :  1.0   Abra      :  1   Min.   : 10.00   Min.   :  5.00
##  1st Qu.: 63.5   Aerodactyl:  1   1st Qu.: 49.00   1st Qu.: 50.00
##  Median :126.0   Aipom     :  1   Median : 65.00   Median : 70.00
##  Mean   :126.0   Alakazam  :  1   Mean   : 66.73   Mean   : 70.62
##  3rd Qu.:188.5   Ampharos  :  1   3rd Qu.: 80.00   3rd Qu.: 87.50
##  Max.   :251.0   Arbok     :  1   Max.   :255.00   Max.   :134.00
##                  (Other)   :245
##       Def              SA              SD              Spd
##  Min.   :  5.00   Min.   : 10.00   Min.   : 20.00   Min.   :  5.00
##  1st Qu.: 49.50   1st Qu.: 44.50   1st Qu.: 50.00   1st Qu.: 45.00
##  Median : 65.00   Median : 65.00   Median : 65.00   Median : 65.00
##  Mean   : 68.61   Mean   : 65.83   Mean   : 68.26   Mean   : 65.89
##  3rd Qu.: 85.00   3rd Qu.: 85.00   3rd Qu.: 85.00   3rd Qu.: 85.00
##  Max.   :230.00   Max.   :154.00   Max.   :230.00   Max.   :140.00
##
##      Total           Type.I          Type.II           Gender        Evolves.From
##  Min.   :180.0   Water   :46             :141   50/50     :158   --        :124
##  1st Qu.:320.0   Normal  :44   Flying : 38   F (100%) :  9   Eevee   :  5
##  Median :410.0   Bug     :22   Poison : 22   F (75%)  : 11   Tyrogue :  3
##  Mean   :405.9   Grass   :21   Ground : 13   M (100%) :  8   Gloom   :  2
##  3rd Qu.:490.0   Fire    :20   Psychic:  9   M (75%)  : 12   Alakazam:  1
##  Max.   :680.0   Electric:15   Rock   :  5   M (87.5%): 32   Azurill :  1
```

```
##              (Other) :83   (Other): 23   None    : 21   (Other) :115
##     Evolves.Into    Captive
##   --        :122   Captive:133
##   Alakazam  :  1   Wild   :118
##   Ambipom   :  1
##   Ampharados:  1
##   Arbok     :  1
##   Arcanine  :  1
##   (Other)   :124
```

describe/describeBy functions from psych package

load *psych* package

```
library(psych)
describe(pokemon[,3:9])
```

```
##        vars   n   mean      sd median trimmed    mad min max range   skew kurtosis
## HP        1 251  66.73  29.82     65   64.03  22.24  10 255   245   2.38    11.53
## Atk       2 251  70.62  27.17     70   70.07  29.65   5 134   129   0.15    -0.32
## Def       3 251  68.61  30.40     65   66.21  25.20   5 230   225   1.30     4.03
## SA        4 251  65.83  27.20     65   64.39  29.65  10 154   144   0.44    -0.38
## SD        5 251  68.26  27.44     65   66.66  22.24  20 230   210   1.17     4.19
## Spd       6 251  65.89  27.04     65   65.29  29.65   5 140   135   0.20    -0.59
## Total     7 251 405.94 104.38    410  406.98 126.02 180 680   500  -0.02    -0.64
##          se
## HP     1.88
## Atk    1.71
## Def    1.92
## SA     1.72
## SD     1.73
## Spd    1.71
## Total 6.59
```

use 'data table' and 'psy::describe' for grouping describe

```
library(data.table)
as.data.table(pokemon)[,describe(Atk),by=Type.I]
```

```
##        Type.I vars  n     mean        sd median   trimmed     mad min max range
## 1:      Grass    1 21 64.80952 20.563606   62.0 64.47059 25.2042  30 105    75
## 2:       Fire    1 20 80.85000 27.562895   80.0 79.75000 29.6520  40 130    90
## 3:      Water    1 46 69.08696 24.728594   66.0 68.55263 24.4629  10 130   120
## 4:        Bug    1 22 65.68182 36.557048   65.0 64.44444 44.4780  10 130   120
## 5:     Normal    1 44 64.47727 28.680671   67.5 64.22222 27.4281   5 130   125
## 6:     Poison    1 15 74.13333 17.602218   80.0 74.00000 17.7912  45 105    60
## 7:   Electric    1 15 61.06667 20.126268   60.0 61.23077 29.6520  30  90    60
## 8:     Ground    1 11 82.72727 25.334131   80.0 81.11111 29.6520  50 130    80
## 9:   Fighting    1  9 94.44444 27.663654  100.0 94.44444 29.6520  35 130    95
## 10:   Psychic    1 15 64.40000 27.249640   65.0 64.30769 29.6520  20 110    90
```

```
## 11:     Rock    1 13 85.53846 27.972056   84.0 85.27273 31.1346  40 134    94
## 12:    Ghost    1  4 52.50000 13.228757   55.0 52.50000 11.1195  35  65    30
## 13:      Ice    1  6 61.66667 25.819889   52.5 61.66667 18.5325  30 100    70
## 14:   Dragon    1  3 94.00000 36.055513   84.0 94.00000 29.6520  64 134    70
## 15:     Dark    1  5 79.00000 15.572412   85.0 79.00000 14.8260  60  95    35
## 16:    Steel    1  2 82.50000  3.535534   82.5 82.50000  3.7065  80  85     5
##             skew    kurtosis          se
##  1:  0.12547639 -1.10185458   4.487347
##  2:  0.23657860 -1.10819469   6.163251
##  3:  0.16031101  0.07363318   3.646032
##  4:  0.26373032 -1.17100145   7.793989
##  5:  0.04361559 -0.43067986   4.323774
##  6: -0.12027162 -1.21273309   4.544873
##  7:  0.05952224 -1.45445074   5.196580
##  8:  0.52283912 -1.00820313   7.638528
##  9: -0.79733914 -0.22545978   9.221218
## 10:  0.14007534 -1.32023873   7.035827
## 11: -0.08007189 -1.18239047   7.758052
## 12: -0.32396955 -2.00892857   6.614378
## 13:  0.32732088 -1.68739583  10.540926
## 14:  0.25601548 -2.33333333  20.816660
## 15: -0.20972856 -2.14903603   6.964194
## 16:  0.00000000 -2.75000000   2.500000
```