case_study_2,通过智能设备数据分析帮助 Bellabeat 公司市场策略团队

qinggeli

2022-11-11

本次案例分析中，我们扮演一名 Bellabeat 公司的数据分析师，通过对用户的设备数据进行分析，并得出对市场部门有帮助的 insigh

加载必要的 function

```
library(ggplot2)
library(tibble)
library(tidyr)
library(readr)
library(purrr)
library(dplyr)
```

```
##
## 载入程辑包：'dplyr'
```

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

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

```
library(stringr)
library(forcats)
library(lubridate)
```

```
##
## 载入程辑包：'lubridate'
```

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

```
library(here)
```

```
## Warning: 程辑包'here'是用 R 版本 4.2.2 来建造的
```

```
## here() starts at D:/Data Analize/case_study_2

library(skimr)

## Warning: 程辑包'skimr'是用 R 版本 4.2.2 来建造的

library(janitor)

## Warning: 程辑包'janitor'是用 R 版本 4.2.2 来建造的

##
## 载入程辑包：'janitor'

## The following objects are masked from 'package:stats':
##
##     chisq.test, fisher.test
```

再加载原始数据

```
DAM <- read_csv("dailyActivity_merged.csv")#DAM 为每日活动数据的汇总

## Rows: 940 Columns: 15
## ── Column specification ────────────────────────────────────────
────────────
## Delimiter: ","
## chr  (1): ActivityDate
## dbl (14): Id, TotalSteps, TotalDistance, TrackerDistance, LoggedActi
vitiesDi...
##
## ℹ Use `spec()` to retrieve the full column specification for this d
ata.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet
 this message.

DCM <- read_csv("dailyCalories_merged.csv")

## Rows: 940 Columns: 3
## ── Column specification ────────────────────────────────────────
────────────
## Delimiter: ","
## chr (1): ActivityDay
## dbl (2): Id, Calories
##
## ℹ Use `spec()` to retrieve the full column specification for this d
ata.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet
 this message.

DIM <- read_csv("dailyIntensities_merged.csv")

## Rows: 940 Columns: 10
## ── Column specification ────────────────────────────────────────
────────────
```

```
## Delimiter: ","
## chr (1): ActivityDay
## dbl (9): Id, SedentaryMinutes, LightlyActiveMinutes, FairlyActiveMin
utes, Ve...
##
## ℹ Use `spec()` to retrieve the full column specification for this d
ata.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet
 this message.

DSM <- read_csv("dailySteps_merged.csv")

## Rows: 940 Columns: 3
## ── Column specification ────────────────────────────────────────────
────────────
## Delimiter: ","
## chr (1): ActivityDay
## dbl (2): Id, StepTotal
##
## ℹ Use `spec()` to retrieve the full column specification for this d
ata.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet
 this message.

DSLM <- read_csv("sleepDay_merged.csv")#DSLM 为每日睡眠数据的汇总

## Rows: 413 Columns: 5
## ── Column specification ────────────────────────────────────────────
────────────
## Delimiter: ","
## chr (1): SleepDay
## dbl (4): Id, TotalSleepRecords, TotalMinutesAsleep, TotalTimeInBed
##
## ℹ Use `spec()` to retrieve the full column specification for this d
ata.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet
 this message.

Weight <- read_csv("weightLogInfo_merged.csv")

## Rows: 67 Columns: 8
## ── Column specification ────────────────────────────────────────────
────────────
## Delimiter: ","
## chr (1): Date
## dbl (6): Id, WeightKg, WeightPounds, Fat, BMI, LogId
## lgl (1): IsManualReport
##
## ℹ Use `spec()` to retrieve the full column specification for this d
ata.
```

```
## ℹ  Specify the column types or set `show_col_types = FALSE` to quiet
 this message.

HeartRate<-read.csv("heartrate_seconds_merged.csv")
```

查看数据结构（1）

```
str(DAM)

## spc_tbl_ [940 × 15] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ Id                       : num [1:940] 1.5e+09 1.5e+09 1.5e+09 1.5
e+09 1.5e+09 ...
##  $ ActivityDate             : chr [1:940] "4/12/2016" "4/13/2016" "4/
14/2016" "4/15/2016" ...
##  $ TotalSteps               : num [1:940] 13162 10735 10460 9762 1266
9 ...
##  $ TotalDistance            : num [1:940] 8.5 6.97 6.74 6.28 8.16 ...
##  $ TrackerDistance          : num [1:940] 8.5 6.97 6.74 6.28 8.16 ...
##  $ LoggedActivitiesDistance: num [1:940] 0 0 0 0 0 0 0 0 0 0 ...
##  $ VeryActiveDistance       : num [1:940] 1.88 1.57 2.44 2.14 2.71 ...
##  $ ModeratelyActiveDistance: num [1:940] 0.55 0.69 0.4 1.26 0.41 ...
##  $ LightActiveDistance      : num [1:940] 6.06 4.71 3.91 2.83 5.04 ...
##  $ SedentaryActiveDistance : num [1:940] 0 0 0 0 0 0 0 0 0 0 ...
##  $ VeryActiveMinutes        : num [1:940] 25 21 30 29 36 38 42 50 28
19 ...
##  $ FairlyActiveMinutes      : num [1:940] 13 19 11 34 10 20 16 31 12
8 ...
##  $ LightlyActiveMinutes     : num [1:940] 328 217 181 209 221 164 233
 264 205 211 ...
##  $ SedentaryMinutes         : num [1:940] 728 776 1218 726 773 ...
##  $ Calories                 : num [1:940] 1985 1797 1776 1745 1863 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   Id = col_double(),
##   ..   ActivityDate = col_character(),
##   ..   TotalSteps = col_double(),
##   ..   TotalDistance = col_double(),
##   ..   TrackerDistance = col_double(),
##   ..   LoggedActivitiesDistance = col_double(),
##   ..   VeryActiveDistance = col_double(),
##   ..   ModeratelyActiveDistance = col_double(),
##   ..   LightActiveDistance = col_double(),
##   ..   SedentaryActiveDistance = col_double(),
##   ..   VeryActiveMinutes = col_double(),
##   ..   FairlyActiveMinutes = col_double(),
##   ..   LightlyActiveMinutes = col_double(),
##   ..   SedentaryMinutes = col_double(),
##   ..   Calories = col_double()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
str(DSLM)
```

```
## spc_tbl_ [413 × 5] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ Id                : num [1:413] 1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.
5e+09 ...
##  $ SleepDay          : chr [1:413] "4/12/2016 12:00:00 AM" "4/13/201
6 12:00:00 AM" "4/15/2016 12:00:00 AM" "4/16/2016 12:00:00 AM" ...
##  $ TotalSleepRecords : num [1:413] 1 2 1 2 1 1 1 1 1 1 ...
##  $ TotalMinutesAsleep: num [1:413] 327 384 412 340 700 304 360 325 3
61 430 ...
##  $ TotalTimeInBed    : num [1:413] 346 407 442 367 712 320 377 364 3
84 449 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   Id = col_double(),
##   ..   SleepDay = col_character(),
##   ..   TotalSleepRecords = col_double(),
##   ..   TotalMinutesAsleep = col_double(),
##   ..   TotalTimeInBed = col_double()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
str(Weight)
```

```
## spc_tbl_ [67 × 8] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ Id            : num [1:67] 1.50e+09 1.50e+09 1.93e+09 2.87e+09 2.
87e+09 ...
##  $ Date          : chr [1:67] "5/2/2016 11:59:59 PM" "5/3/2016 11:59:
59 PM" "4/13/2016 1:08:52 AM" "4/21/2016 11:59:59 PM" ...
##  $ WeightKg      : num [1:67] 52.6 52.6 133.5 56.7 57.3 ...
##  $ WeightPounds  : num [1:67] 116 116 294 125 126 ...
##  $ Fat           : num [1:67] 22 NA NA NA NA 25 NA NA NA NA ...
##  $ BMI           : num [1:67] 22.6 22.6 47.5 21.5 21.7 ...
##  $ IsManualReport: logi [1:67] TRUE TRUE FALSE TRUE TRUE TRUE ...
##  $ LogId         : num [1:67] 1.46e+12 1.46e+12 1.46e+12 1.46e+12 1.
46e+12 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   Id = col_double(),
##   ..   Date = col_character(),
##   ..   WeightKg = col_double(),
##   ..   WeightPounds = col_double(),
##   ..   Fat = col_double(),
##   ..   BMI = col_double(),
##   ..   IsManualReport = col_logical(),
##   ..   LogId = col_double()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
str(HeartRate)
```

```
## 'data.frame':    2483658 obs. of  3 variables:
##  $ Id   : num  2.02e+09 2.02e+09 2.02e+09 2.02e+09 2.02e+09 ...
##  $ Time : chr  "4/12/2016 7:21:00 AM" "4/12/2016 7:21:05 AM" "4/12/2
016 7:21:10 AM" "4/12/2016 7:21:20 AM" ...
##  $ Value: int  97 102 105 103 101 95 91 93 94 93 ...

colnames(DAM)

##  [1] "Id"                       "ActivityDate"
##  [3] "TotalSteps"               "TotalDistance"
##  [5] "TrackerDistance"          "LoggedActivitiesDistance"
##  [7] "VeryActiveDistance"       "ModeratelyActiveDistance"
##  [9] "LightActiveDistance"      "SedentaryActiveDistance"
## [11] "VeryActiveMinutes"        "FairlyActiveMinutes"
## [13] "LightlyActiveMinutes"     "SedentaryMinutes"
## [15] "Calories"

colnames(DSLM)

## [1] "Id"                 "SleepDay"           "TotalSleepRecords"
## [4] "TotalMinutesAsleep" "TotalTimeInBed"

colnames(Weight)

## [1] "Id"             "Date"           "WeightKg"       "WeightPounds
"
## [5] "Fat"            "BMI"            "IsManualReport" "LogId"

colnames(HeartRate)

## [1] "Id"    "Time"  "Value"
```

联合 DAM,DCM,DIM,DSM 为 combine_data1,该表中记录了 30 名用户的活动数据，该表格也用于后续分析

```
combine_data1<-bind_cols(DAM,DCM,DIM,DSM)

## New names:
## • `Id` -> `Id...1`
## • `VeryActiveDistance` -> `VeryActiveDistance...7`
## • `ModeratelyActiveDistance` -> `ModeratelyActiveDistance...8`
## • `LightActiveDistance` -> `LightActiveDistance...9`
## • `SedentaryActiveDistance` -> `SedentaryActiveDistance...10`
## • `VeryActiveMinutes` -> `VeryActiveMinutes...11`
## • `FairlyActiveMinutes` -> `FairlyActiveMinutes...12`
## • `LightlyActiveMinutes` -> `LightlyActiveMinutes...13`
## • `SedentaryMinutes` -> `SedentaryMinutes...14`
## • `Calories` -> `Calories...15`
## • `Id` -> `Id...16`
## • `ActivityDay` -> `ActivityDay...17`
## • `Calories` -> `Calories...18`
## • `Id` -> `Id...19`
```

```
## • `ActivityDay` -> `ActivityDay...20`
## • `SedentaryMinutes` -> `SedentaryMinutes...21`
## • `LightlyActiveMinutes` -> `LightlyActiveMinutes...22`
## • `FairlyActiveMinutes` -> `FairlyActiveMinutes...23`
## • `VeryActiveMinutes` -> `VeryActiveMinutes...24`
## • `SedentaryActiveDistance` -> `SedentaryActiveDistance...25`
## • `LightActiveDistance` -> `LightActiveDistance...26`
## • `ModeratelyActiveDistance` -> `ModeratelyActiveDistance...27`
## • `VeryActiveDistance` -> `VeryActiveDistance...28`
## • `Id` -> `Id...29`
## • `ActivityDay` -> `ActivityDay...30`
```

```
colnames(combine_data1)
```

```
##  [1] "Id...1"                       "ActivityDate"
##  [3] "TotalSteps"                   "TotalDistance"
##  [5] "TrackerDistance"              "LoggedActivitiesDistance"
##  [7] "VeryActiveDistance...7"       "ModeratelyActiveDistance...8"
##  [9] "LightActiveDistance...9"      "SedentaryActiveDistance...10"
## [11] "VeryActiveMinutes...11"       "FairlyActiveMinutes...12"
## [13] "LightlyActiveMinutes...13"    "SedentaryMinutes...14"
## [15] "Calories...15"                "Id...16"
## [17] "ActivityDay...17"             "Calories...18"
## [19] "Id...19"                      "ActivityDay...20"
## [21] "SedentaryMinutes...21"        "LightlyActiveMinutes...22"
## [23] "FairlyActiveMinutes...23"     "VeryActiveMinutes...24"
## [25] "SedentaryActiveDistance...25" "LightActiveDistance...26"
## [27] "ModeratelyActiveDistance...27" "VeryActiveDistance...28"
## [29] "Id...29"                      "ActivityDay...30"
## [31] "StepTotal"
```

```
row.names(combine_data1)
```

```
##   [1] "1"   "2"   "3"   "4"   "5"   "6"   "7"   "8"   "9"   "10"  "1
1"  "12"
##  [13] "13"  "14"  "15"  "16"  "17"  "18"  "19"  "20"  "21"  "22"  "2
3"  "24"
##  [25] "25"  "26"  "27"  "28"  "29"  "30"  "31"  "32"  "33"  "34"  "3
5"  "36"
##  [37] "37"  "38"  "39"  "40"  "41"  "42"  "43"  "44"  "45"  "46"  "4
7"  "48"
##  [49] "49"  "50"  "51"  "52"  "53"  "54"  "55"  "56"  "57"  "58"  "5
9"  "60"
##  [61] "61"  "62"  "63"  "64"  "65"  "66"  "67"  "68"  "69"  "70"  "7
1"  "72"
##  [73] "73"  "74"  "75"  "76"  "77"  "78"  "79"  "80"  "81"  "82"  "8
3"  "84"
##  [85] "85"  "86"  "87"  "88"  "89"  "90"  "91"  "92"  "93"  "94"  "9
5"  "96"
##  [97] "97"  "98"  "99"  "100" "101" "102" "103" "104" "105" "106" "1
07" "108"
```

```
## [109] "109" "110" "111" "112" "113" "114" "115" "116" "117" "118" "1
19" "120"
## [121] "121" "122" "123" "124" "125" "126" "127" "128" "129" "130" "1
31" "132"
## [133] "133" "134" "135" "136" "137" "138" "139" "140" "141" "142" "1
43" "144"
## [145] "145" "146" "147" "148" "149" "150" "151" "152" "153" "154" "1
55" "156"
## [157] "157" "158" "159" "160" "161" "162" "163" "164" "165" "166" "1
67" "168"
## [169] "169" "170" "171" "172" "173" "174" "175" "176" "177" "178" "1
79" "180"
## [181] "181" "182" "183" "184" "185" "186" "187" "188" "189" "190" "1
91" "192"
## [193] "193" "194" "195" "196" "197" "198" "199" "200" "201" "202" "2
03" "204"
## [205] "205" "206" "207" "208" "209" "210" "211" "212" "213" "214" "2
15" "216"
## [217] "217" "218" "219" "220" "221" "222" "223" "224" "225" "226" "2
27" "228"
## [229] "229" "230" "231" "232" "233" "234" "235" "236" "237" "238" "2
39" "240"
## [241] "241" "242" "243" "244" "245" "246" "247" "248" "249" "250" "2
51" "252"
## [253] "253" "254" "255" "256" "257" "258" "259" "260" "261" "262" "2
63" "264"
## [265] "265" "266" "267" "268" "269" "270" "271" "272" "273" "274" "2
75" "276"
## [277] "277" "278" "279" "280" "281" "282" "283" "284" "285" "286" "2
87" "288"
## [289] "289" "290" "291" "292" "293" "294" "295" "296" "297" "298" "2
99" "300"
## [301] "301" "302" "303" "304" "305" "306" "307" "308" "309" "310" "3
11" "312"
## [313] "313" "314" "315" "316" "317" "318" "319" "320" "321" "322" "3
23" "324"
## [325] "325" "326" "327" "328" "329" "330" "331" "332" "333" "334" "3
35" "336"
## [337] "337" "338" "339" "340" "341" "342" "343" "344" "345" "346" "3
47" "348"
## [349] "349" "350" "351" "352" "353" "354" "355" "356" "357" "358" "3
59" "360"
## [361] "361" "362" "363" "364" "365" "366" "367" "368" "369" "370" "3
71" "372"
## [373] "373" "374" "375" "376" "377" "378" "379" "380" "381" "382" "3
83" "384"
## [385] "385" "386" "387" "388" "389" "390" "391" "392" "393" "394" "3
95" "396"
## [397] "397" "398" "399" "400" "401" "402" "403" "404" "405" "406" "4
07" "408"
```

```
## [409] "409" "410" "411" "412" "413" "414" "415" "416" "417" "418" "4
19" "420"
## [421] "421" "422" "423" "424" "425" "426" "427" "428" "429" "430" "4
31" "432"
## [433] "433" "434" "435" "436" "437" "438" "439" "440" "441" "442" "4
43" "444"
## [445] "445" "446" "447" "448" "449" "450" "451" "452" "453" "454" "4
55" "456"
## [457] "457" "458" "459" "460" "461" "462" "463" "464" "465" "466" "4
67" "468"
## [469] "469" "470" "471" "472" "473" "474" "475" "476" "477" "478" "4
79" "480"
## [481] "481" "482" "483" "484" "485" "486" "487" "488" "489" "490" "4
91" "492"
## [493] "493" "494" "495" "496" "497" "498" "499" "500" "501" "502" "5
03" "504"
## [505] "505" "506" "507" "508" "509" "510" "511" "512" "513" "514" "5
15" "516"
## [517] "517" "518" "519" "520" "521" "522" "523" "524" "525" "526" "5
27" "528"
## [529] "529" "530" "531" "532" "533" "534" "535" "536" "537" "538" "5
39" "540"
## [541] "541" "542" "543" "544" "545" "546" "547" "548" "549" "550" "5
51" "552"
## [553] "553" "554" "555" "556" "557" "558" "559" "560" "561" "562" "5
63" "564"
## [565] "565" "566" "567" "568" "569" "570" "571" "572" "573" "574" "5
75" "576"
## [577] "577" "578" "579" "580" "581" "582" "583" "584" "585" "586" "5
87" "588"
## [589] "589" "590" "591" "592" "593" "594" "595" "596" "597" "598" "5
99" "600"
## [601] "601" "602" "603" "604" "605" "606" "607" "608" "609" "610" "6
11" "612"
## [613] "613" "614" "615" "616" "617" "618" "619" "620" "621" "622" "6
23" "624"
## [625] "625" "626" "627" "628" "629" "630" "631" "632" "633" "634" "6
35" "636"
## [637] "637" "638" "639" "640" "641" "642" "643" "644" "645" "646" "6
47" "648"
## [649] "649" "650" "651" "652" "653" "654" "655" "656" "657" "658" "6
59" "660"
## [661] "661" "662" "663" "664" "665" "666" "667" "668" "669" "670" "6
71" "672"
## [673] "673" "674" "675" "676" "677" "678" "679" "680" "681" "682" "6
83" "684"
## [685] "685" "686" "687" "688" "689" "690" "691" "692" "693" "694" "6
95" "696"
## [697] "697" "698" "699" "700" "701" "702" "703" "704" "705" "706" "7
07" "708"
```

```
## [709] "709" "710" "711" "712" "713" "714" "715" "716" "717" "718" "7
19" "720"
## [721] "721" "722" "723" "724" "725" "726" "727" "728" "729" "730" "7
31" "732"
## [733] "733" "734" "735" "736" "737" "738" "739" "740" "741" "742" "7
43" "744"
## [745] "745" "746" "747" "748" "749" "750" "751" "752" "753" "754" "7
55" "756"
## [757] "757" "758" "759" "760" "761" "762" "763" "764" "765" "766" "7
67" "768"
## [769] "769" "770" "771" "772" "773" "774" "775" "776" "777" "778" "7
79" "780"
## [781] "781" "782" "783" "784" "785" "786" "787" "788" "789" "790" "7
91" "792"
## [793] "793" "794" "795" "796" "797" "798" "799" "800" "801" "802" "8
03" "804"
## [805] "805" "806" "807" "808" "809" "810" "811" "812" "813" "814" "8
15" "816"
## [817] "817" "818" "819" "820" "821" "822" "823" "824" "825" "826" "8
27" "828"
## [829] "829" "830" "831" "832" "833" "834" "835" "836" "837" "838" "8
39" "840"
## [841] "841" "842" "843" "844" "845" "846" "847" "848" "849" "850" "8
51" "852"
## [853] "853" "854" "855" "856" "857" "858" "859" "860" "861" "862" "8
63" "864"
## [865] "865" "866" "867" "868" "869" "870" "871" "872" "873" "874" "8
75" "876"
## [877] "877" "878" "879" "880" "881" "882" "883" "884" "885" "886" "8
87" "888"
## [889] "889" "890" "891" "892" "893" "894" "895" "896" "897" "898" "8
99" "900"
## [901] "901" "902" "903" "904" "905" "906" "907" "908" "909" "910" "9
11" "912"
## [913] "913" "914" "915" "916" "917" "918" "919" "920" "921" "922" "9
23" "924"
## [925] "925" "926" "927" "928" "929" "930" "931" "932" "933" "934" "9
35" "936"
## [937] "937" "938" "939" "940"
```

```
distinct(combine_data1)
```

```
## # A tibble: 940 × 31
##        Id...1 Activity…¹ Total…² Total…³ Track…⁴ Logge…⁵ VeryA…⁶ Mod
er…⁷ Light…⁸
##         <dbl> <chr>        <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <
dbl>   <dbl>
##  1 1503960366 4/12/2016    13162    8.5     8.5        0    1.88   0.
550    6.06
##  2 1503960366 4/13/2016    10735    6.97    6.97       0    1.57   0.
```

```
690    4.71
##  3 1503960366 4/14/2016    10460    6.74    6.74       0    2.44   0.
400    3.91
##  4 1503960366 4/15/2016     9762    6.28    6.28       0    2.14   1.
26     2.83
##  5 1503960366 4/16/2016    12669    8.16    8.16       0    2.71   0.
410    5.04
##  6 1503960366 4/17/2016     9705    6.48    6.48       0    3.19   0.
780    2.51
##  7 1503960366 4/18/2016    13019    8.59    8.59       0    3.25   0.
640    4.71
##  8 1503960366 4/19/2016    15506    9.88    9.88       0    3.53   1.
32     5.03
##  9 1503960366 4/20/2016    10544    6.68    6.68       0    1.96   0.
480    4.24
## 10 1503960366 4/21/2016     9819    6.34    6.34       0    1.34   0.
350    4.65
## # … with 930 more rows, 22 more variables: SedentaryActiveDistance...
10 <dbl>,
## #   VeryActiveMinutes...11 <dbl>, FairlyActiveMinutes...12 <dbl>,
## #   LightlyActiveMinutes...13 <dbl>, SedentaryMinutes...14 <dbl>,
## #   Calories...15 <dbl>, Id...16 <dbl>, ActivityDay...17 <chr>,
## #   Calories...18 <dbl>, Id...19 <dbl>, ActivityDay...20 <chr>,
## #   SedentaryMinutes...21 <dbl>, LightlyActiveMinutes...22 <dbl>,
## #   FairlyActiveMinutes...23 <dbl>, VeryActiveMinutes...24 <dbl>, …
```

联合每日活动数据和每日睡眠数据

```
str(DSLM)
```

```
## spc_tbl_ [413 × 5] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ Id                : num [1:413] 1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.
5e+09 ...
##  $ SleepDay          : chr [1:413] "4/12/2016 12:00:00 AM" "4/13/201
6 12:00:00 AM" "4/15/2016 12:00:00 AM" "4/16/2016 12:00:00 AM" ...
##  $ TotalSleepRecords : num [1:413] 1 2 1 2 1 1 1 1 1 1 ...
##  $ TotalMinutesAsleep: num [1:413] 327 384 412 340 700 304 360 325 3
61 430 ...
##  $ TotalTimeInBed    : num [1:413] 346 407 442 367 712 320 377 364 3
84 449 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   Id = col_double(),
##   ..   SleepDay = col_character(),
##   ..   TotalSleepRecords = col_double(),
##   ..   TotalMinutesAsleep = col_double(),
##   ..   TotalTimeInBed = col_double()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
DSLM<-rename(DSLM,ActivityDate=SleepDay)#改成与DAM 表格中要匹配的列同名
DSLM<-separate(DSLM,ActivityDate,into = c("ActivityDate","hour"),sep="
")#将原有的列分离成可匹配的型态

## Warning: Expected 2 pieces. Additional pieces discarded in 413 rows
[1, 2, 3, 4,
## 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, ...].

DSLM<-mutate(DSLM,Id=as.character(DSLM$Id))#将 Id 列转化为字串，这样才能匹
配
DAM<-mutate(DAM,Id=as.character(DAM$Id))#将 Id 列转化为字串，这样才能匹配
combine_data2<-merge(DAM,DSLM,by=c("Id","ActivityDate"))#根据两个匹配条
件合并表格

str(combine_data1)

## spc_tbl_ [940 × 31] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ Id...1                       : num [1:940] 1.5e+09 1.5e+09 1.5e+0
9 1.5e+09 1.5e+09 ...
##  $ ActivityDate                 : chr [1:940] "4/12/2016" "4/13/2016
" "4/14/2016" "4/15/2016" ...
##  $ TotalSteps                   : num [1:940] 13162 10735 10460 9762
 12669 ...
##  $ TotalDistance                : num [1:940] 8.5 6.97 6.74 6.28 8.1
6 ...
##  $ TrackerDistance              : num [1:940] 8.5 6.97 6.74 6.28 8.1
6 ...
##  $ LoggedActivitiesDistance     : num [1:940] 0 0 0 0 0 0 0 0 0 0 ...
##  $ VeryActiveDistance...7       : num [1:940] 1.88 1.57 2.44 2.14 2.
71 ...
##  $ ModeratelyActiveDistance...8 : num [1:940] 0.55 0.69 0.4 1.26 0.4
1 ...
##  $ LightActiveDistance...9      : num [1:940] 6.06 4.71 3.91 2.83 5.
04 ...
##  $ SedentaryActiveDistance...10 : num [1:940] 0 0 0 0 0 0 0 0 0 0 ...
##  $ VeryActiveMinutes...11       : num [1:940] 25 21 30 29 36 38 42 5
0 28 19 ...
##  $ FairlyActiveMinutes...12     : num [1:940] 13 19 11 34 10 20 16 3
1 12 8 ...
##  $ LightlyActiveMinutes...13    : num [1:940] 328 217 181 209 221 16
4 233 264 205 211 ...
##  $ SedentaryMinutes...14        : num [1:940] 728 776 1218 726 773
 ...
##  $ Calories...15                : num [1:940] 1985 1797 1776 1745 18
63 ...
##  $ Id...16                      : num [1:940] 1.5e+09 1.5e+09 1.5e+0
9 1.5e+09 1.5e+09 ...
##  $ ActivityDay...17             : chr [1:940] "4/12/2016" "4/13/2016
" "4/14/2016" "4/15/2016" ...
##  $ Calories...18                : num [1:940] 1985 1797 1776 1745 18
```

```
63 ...
##  $ Id...19                     : num [1:940] 1.5e+09 1.5e+09 1.5e+0
9 1.5e+09 1.5e+09 ...
##  $ ActivityDay...20            : chr [1:940] "4/12/2016" "4/13/2016
" "4/14/2016" "4/15/2016" ...
##  $ SedentaryMinutes...21       : num [1:940] 728 776 1218 726 773
 ...
##  $ LightlyActiveMinutes...22   : num [1:940] 328 217 181 209 221 16
4 233 264 205 211 ...
##  $ FairlyActiveMinutes...23    : num [1:940] 13 19 11 34 10 20 16 3
1 12 8 ...
##  $ VeryActiveMinutes...24      : num [1:940] 25 21 30 29 36 38 42 5
0 28 19 ...
##  $ SedentaryActiveDistance...25 : num [1:940] 0 0 0 0 0 0 0 0 0 0 ...
##  $ LightActiveDistance...26    : num [1:940] 6.06 4.71 3.91 2.83 5.
04 ...
##  $ ModeratelyActiveDistance...27: num [1:940] 0.55 0.69 0.4 1.26 0.4
1 ...
##  $ VeryActiveDistance...28     : num [1:940] 1.88 1.57 2.44 2.14 2.
71 ...
##  $ Id...29                     : num [1:940] 1.5e+09 1.5e+09 1.5e+0
9 1.5e+09 1.5e+09 ...
##  $ ActivityDay...30            : chr [1:940] "4/12/2016" "4/13/2016
" "4/14/2016" "4/15/2016" ...
##  $ StepTotal                   : num [1:940] 13162 10735 10460 9762
 12669 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   Id = col_double(),
##   ..   ActivityDate = col_character(),
##   ..   TotalSteps = col_double(),
##   ..   TotalDistance = col_double(),
##   ..   TrackerDistance = col_double(),
##   ..   LoggedActivitiesDistance = col_double(),
##   ..   VeryActiveDistance = col_double(),
##   ..   ModeratelyActiveDistance = col_double(),
##   ..   LightActiveDistance = col_double(),
##   ..   SedentaryActiveDistance = col_double(),
##   ..   VeryActiveMinutes = col_double(),
##   ..   FairlyActiveMinutes = col_double(),
##   ..   LightlyActiveMinutes = col_double(),
##   ..   SedentaryMinutes = col_double(),
##   ..   Calories = col_double()
##   .. )
##  - attr(*, "problems")=<externalptr>

str(combine_data2)

## 'data.frame':    413 obs. of  19 variables:
##  $ Id                       : chr  "1503960366" "1503960366" "1503960
```

```
366" "1503960366" ...
##  $ ActivityDate            : chr  "4/12/2016" "4/13/2016" "4/15/2016
" "4/16/2016" ...
##  $ TotalSteps              : num  13162 10735 9762 12669 9705 ...
##  $ TotalDistance           : num  8.5 6.97 6.28 8.16 6.48 ...
##  $ TrackerDistance         : num  8.5 6.97 6.28 8.16 6.48 ...
##  $ LoggedActivitiesDistance: num  0 0 0 0 0 0 0 0 0 0 ...
##  $ VeryActiveDistance      : num  1.88 1.57 2.14 2.71 3.19 ...
##  $ ModeratelyActiveDistance: num  0.55 0.69 1.26 0.41 0.78 ...
##  $ LightActiveDistance     : num  6.06 4.71 2.83 5.04 2.51 ...
##  $ SedentaryActiveDistance : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ VeryActiveMinutes       : num  25 21 29 36 38 50 28 19 41 39 ...
##  $ FairlyActiveMinutes     : num  13 19 34 10 20 31 12 8 21 5 ...
##  $ LightlyActiveMinutes    : num  328 217 209 221 164 264 205 211 26
2 238 ...
##  $ SedentaryMinutes        : num  728 776 726 773 539 775 818 838 73
2 709 ...
##  $ Calories                : num  1985 1797 1745 1863 1728 ...
##  $ hour                    : chr  "12:00:00" "12:00:00" "12:00:00" "
12:00:00" ...
##  $ TotalSleepRecords       : num  1 2 1 2 1 1 1 1 1 1 ...
##  $ TotalMinutesAsleep      : num  327 384 412 340 700 304 360 325 36
1 430 ...
##  $ TotalTimeInBed          : num  346 407 442 367 712 320 377 364 38
4 449 ...
```

```
combine_data1<- drop_na(combine_data1)
```

转换数据类型，使其可 manipulate，将原日期转换为星期天数

```
combine_data1$ActivityDate<-as.Date(combine_data1$ActivityDate,format=
"%m/%d/%Y")
combine_data1<-mutate(combine_data1,weekday = wday(ActivityDate, label
= TRUE))
```

combine_data1 描述性统计分析

```
combine_data1 %>%
  select(TotalSteps,
         TotalDistance,
         SedentaryMinutes...14,LightlyActiveMinutes...13,FairlyActiveMi
nutes...12,Calories...18) %>%
  summary()
```

```
##    TotalSteps    TotalDistance    SedentaryMinutes...14
##  Min.   :    0   Min.   : 0.000   Min.   :   0.0
##  1st Qu.: 3790   1st Qu.: 2.620   1st Qu.: 729.8
##  Median : 7406   Median : 5.245   Median :1057.5
```

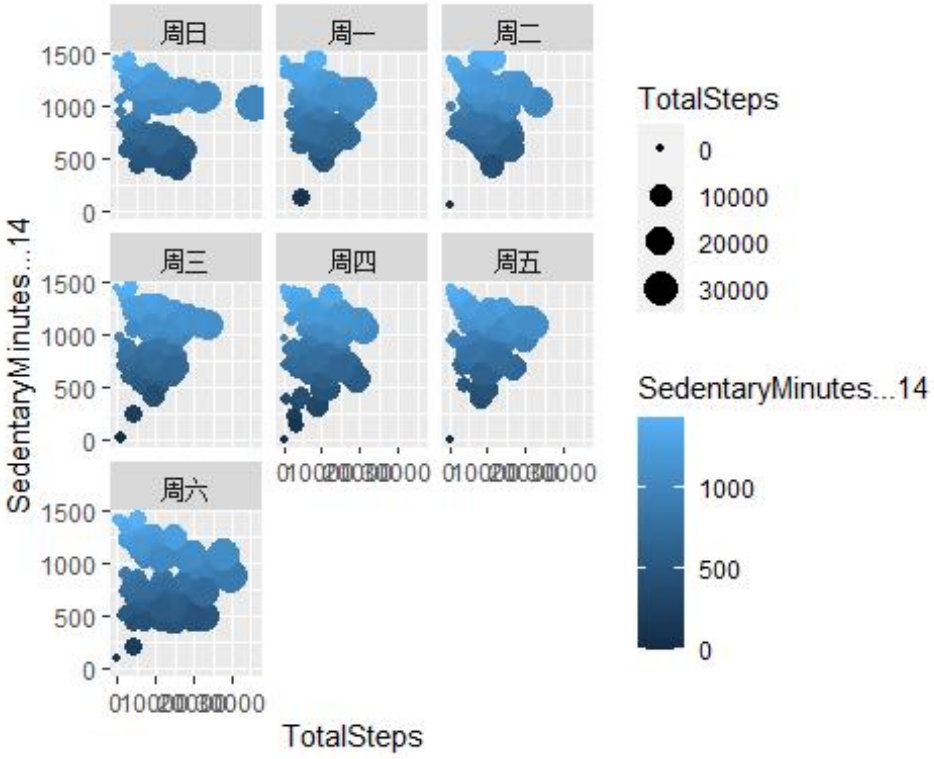
```
##  Mean   : 7638   Mean   : 5.490   Mean   : 991.2
##  3rd Qu.:10727   3rd Qu.: 7.713   3rd Qu.:1229.5
##  Max.   :36019   Max.   :28.030   Max.   :1440.0
##  LightlyActiveMinutes...13 FairlyActiveMinutes...12 Calories...18
##  Min.   :  0.0             Min.   :  0.00           Min.   :   0
##  1st Qu.:127.0             1st Qu.:  0.00           1st Qu.:1828
##  Median :199.0             Median :  6.00           Median :2134
##  Mean   :192.8             Mean   : 13.56           Mean   :2304
##  3rd Qu.:264.0             3rd Qu.: 19.00           3rd Qu.:2793
##  Max.   :518.0             Max.   :143.00           Max.   :4900

combine_data2 %>%
  select(TotalTimeInBed,TotalMinutesAsleep) %>%
  summary()

##  TotalTimeInBed  TotalMinutesAsleep
##  Min.   : 61.0   Min.   : 58.0
##  1st Qu.:403.0   1st Qu.:361.0
##  Median :463.0   Median :433.0
##  Mean   :458.6   Mean   :419.5
##  3rd Qu.:526.0   3rd Qu.:490.0
##  Max.   :961.0   Max.   :796.0
```

智能运动设备使用者坐着的时间和活跃时间的比较分析

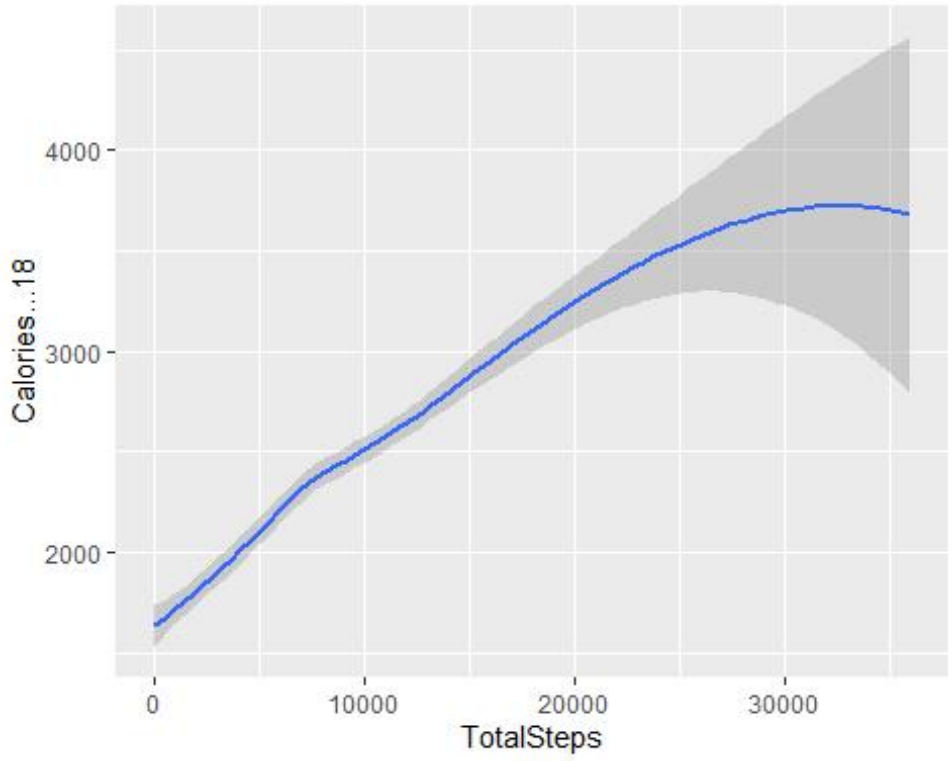
```
ggplot(data=combine_data1)+geom_jitter(mapping=aes(x=TotalSteps,y=Seden
taryMinutes...14,size=TotalSteps,color=SedentaryMinutes...14))+facet_wr
ap(combine_data1$weekday)
```



智能运动设备使用者每日卡路里消耗和总步数的比较分析

```
ggplot(data=combine_data1)+geom_smooth(mapping=aes(x=TotalSteps, y=Calo
ries...18,size=TotalSteps,color=Calories...18))
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



智能设备用户总步数与睡眠时间的比较分析

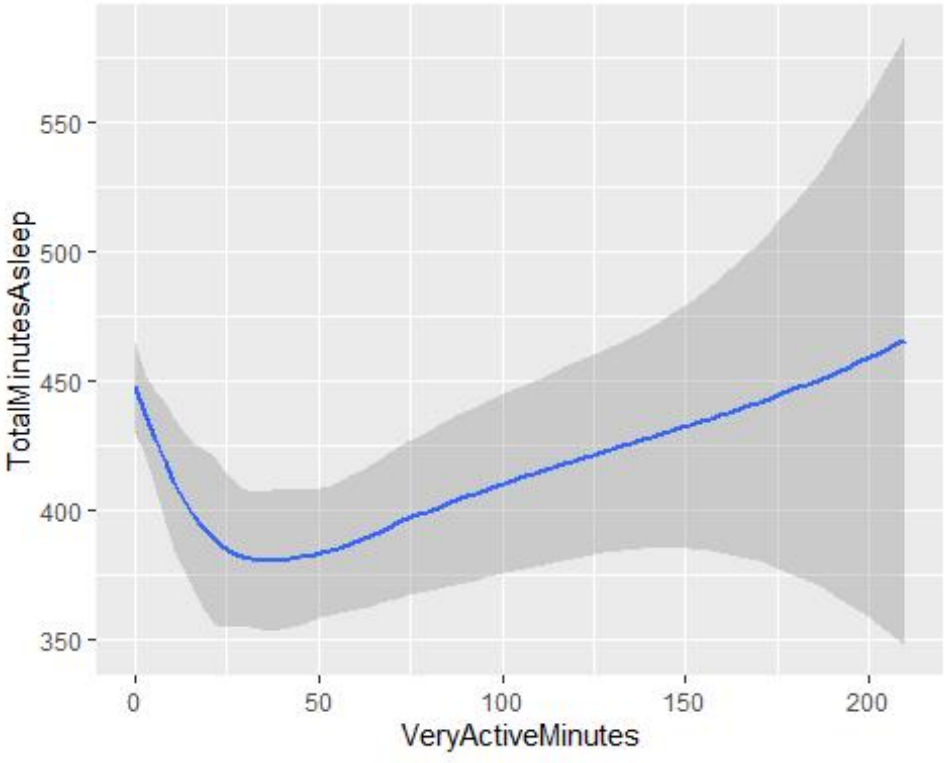
首先计算用户在床时间和睡眠时间的差值,并于总步数表格合并

```
combine_data2<-mutate(combine_data2,TimetofullAsllep=DSLM$TotalTimeInBe
d-DSLM$TotalMinutesAsleep)
combine_data2$ActivityDate<-as.Date(combine_data2$ActivityDate,format=
"%m/%d/%Y")
combine_data2<-mutate(combine_data2,weekday = wday(ActivityDate, label
= TRUE))
```

睡眠时间和极度活跃时间的关系分析

```
ggplot(data=combine_data2)+geom_smooth(mapping=aes(x=VeryActiveMinutes,
 y=TotalMinutesAsleep,size=VeryActiveMinutes,color=TotalMinutesAsleep))
```

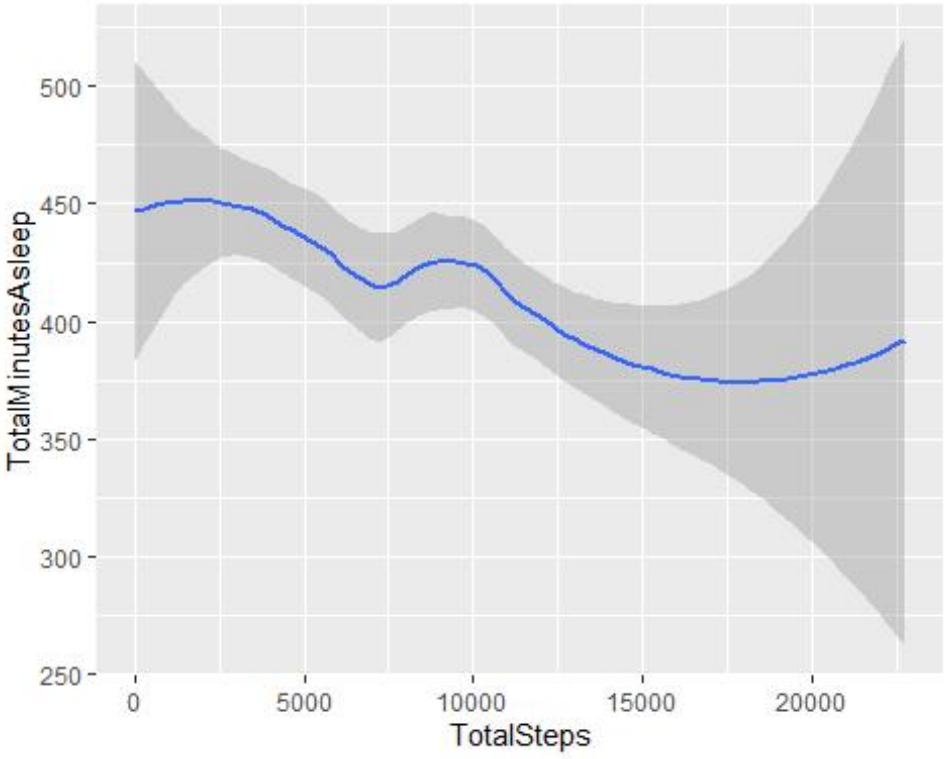
```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



睡眠时间和总步数的关系分析

```
ggplot(data=combine_data2)+geom_smooth(mapping=aes(x=TotalSteps, y=Tota
lMinutesAsleep,size=TotalSteps,color=TotalMinutesAsleep))

## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```


500
450
400
350
300
250
TotalMinutesAsleep
0
5000
10000
15000
20000
TotalSteps