Canonical Correlation Analysis

S18478 - Anusha Rajaguru

30,May 2024

1 Introduction

Canonical Correlation Analysis(CCA) is a statistical technique used to identify relationship between two multivariate data sets , all measured on the same individual.Canonical correlation analysis (CCA) is a method of correlating linear relationships between two multidimensional variables. CCA can be seen as using complex labels as a way of guiding feature selection toward the underlying semantics Hardoon et al., 2004.
This article aim to utilize the analysis knowledge acquired from Correlation Analysis and apply it to solve real world problem.In here we test for relationship between canonical variate pairs.We wish to test null hypothesis that these regression coeffients are all equal to zero.

2 Methodology

This data set about is the sleep Health and Lifestyle Dataset comprises 400 rows and 13 columns, covering a wide range of variables related to sleep and daily habits. It includes details such as gender, age, occupation, sleep duration, quality of sleep, physical activity level, stress levels, BMI category, blood pressure, heart rate, daily steps, and the presence or absence of sleep disorders.

For analysis these data set, Canonical correlation analysis statistical method usedWeenink, 2003.

variables	Description
Person ID	An identifier for each individual
Gender	The gender of the person (Male/Female)
Age	The age of the person in years
Occupation	The occupation or profession of the person
Sleep Duration	The number of hours the person sleeps per day
Quality of Sleep	A subjective rating of the quality of sleep
Physical Activity Level	The number of minutes the person engages in physical activity daily (n
Stress Level	A subjective rating of the stress level experienced by the pers
BMI Category	The BMI category of the person (
Blood Pressure (systolic/diastolic)	The blood pressure measurement
Heart Rate (bpm) Daily Steps	The resting heart rate of the person in beats per minute The number of steps the person takes per day
Sleep Disorder	The presence or absence of a sleep disorder in the person

Table 1: Description about data set

3 Results and Discussion

canonical correlation analysis is actually used to examine the relationship between two data sets, but today we will separate the data set we have into two separate data sets and examine the relationship between the two data sets. Two Clear Sets: We can clearly separate the variables into two meaningful sets: Health Metrics: Quality of Sleep, Blood Pressure, Heart Rate,BMI (converted to numerical) and Sleep Disorder Lifestyle Factors: Physical Activity Level, Daily steps,Stress Level and Sleep Duration. Create pairwise scatter plot with variables of data sets:

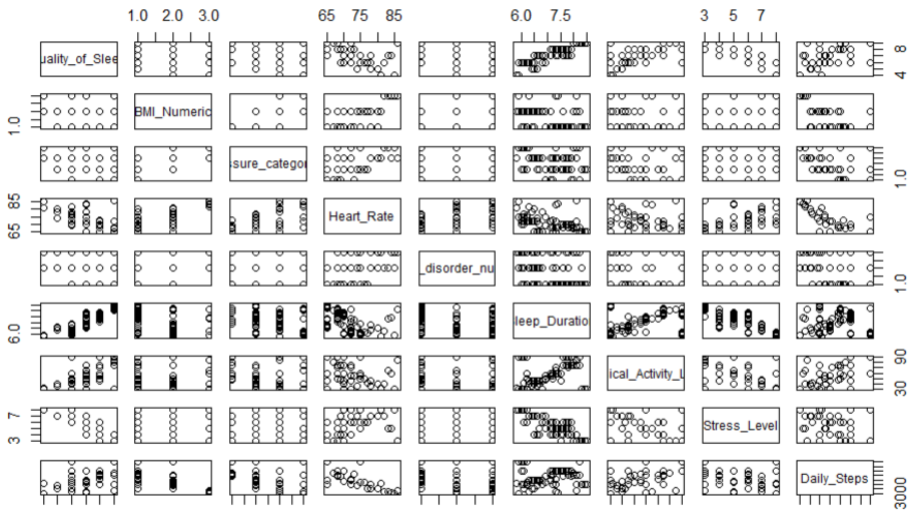


Figure 1: pairwise scatter plot

Estimating of canonical correlation,considering health matrices set there is positive strong correlation between sleeping disorder and BMI value and blood pressure and sleep disorder.considering life style factors set there is a strong positive correlation between physical activity level and daily steps.also strong negative correlation with stress level and sleeping duration. Considering Raw canonical coefficients for health variables:

$$U1 = -0.73821895X_{QSleep} - 0.29753716X_{BMI} + 0.14453297X_{Pressure} + 0.05150106X_{Rate} - 0.03736169X_{Disorder}$$

Considering Raw canonical coefficients for lifestyle variables:

$$U2 = -0.3733756533X_{SDuration} + 0.0090771566X_{Activity} + 0.4444873954X + Stress - 0.0001693747X_{Steps}$$

Also for each significant canonical correlation can be further tested using other multivariate test like Pillai's test,Roy's test.

4 Conclusions

Taking about limitation,in this data set include categorical variables, including gender, occupation, BMI category, and sleep disorder. Gender,Occupation like categorical variables were left out of the CCA while we translated the BMI Category,sleep disorder like variables convert in to a numerical format. If these variables include important information, this could restrict the scope of the study as a whole.Also considering subjective ratings,Due to individual biases in self-reporting, subjective variables such as Stress Level and Sleep Quality may add variability.also when using large sample sizes of data set it can increase the CCA's efficacy.

5 Appendices

References

Hardoon, D. R., Szedmak, S., & Shawe-Taylor, J. (2004). Canonical correlation analysis: An overview with application to learning methods. *Neural computation*, *16*(12), 2639–2664.

Weenink, D. (2003). Canonical correlation analysis. *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam*, *25*, 81–99.

Canonical Correlation Analysis

S18478-Anusha Rajaguru

2024-05-30

load the library

```
library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse
2.0.0 ──
## ✓ dplyr     1.1.3     ✓ readr     2.1.4
## ✓ forcats   1.0.0     ✓ stringr   1.5.0
## ✓ ggplot2   3.4.4     ✓ tibble    3.2.1
## ✓ lubridate 1.9.3     ✓ tidyr     1.3.0
## ✓ purrr     1.0.2
## ── Conflicts ──────────────────────────────────────────
tidyverse_conflicts() ──
## ✕ dplyr::filter() masks stats::filter()
## ✕ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all
conflicts to become errors

library(tidyr)
library(ggplot2)
library(janitor)

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

library(dplyr)
library(Matrix)

##
## Attaching package: 'Matrix'
##
## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack

library(CCA)

## Warning: package 'CCA' was built under R version 4.3.3

## Loading required package: fda
```

```
## Warning: package 'fda' was built under R version 4.3.3

## Loading required package: splines
## Loading required package: fds

## Warning: package 'fds' was built under R version 4.3.3

## Loading required package: rainbow

## Warning: package 'rainbow' was built under R version 4.3.3

## Loading required package: MASS
##
## Attaching package: 'MASS'
##
## The following object is masked from 'package:dplyr':
##
##     select
##
## Loading required package: pcaPP

## Warning: package 'pcaPP' was built under R version 4.3.3

## Loading required package: RCurl

## Warning: package 'RCurl' was built under R version 4.3.2

##
## Attaching package: 'RCurl'
##
## The following object is masked from 'package:tidyr':
##
##     complete
##
## Loading required package: deSolve

## Warning: package 'deSolve' was built under R version 4.3.3

##
## Attaching package: 'fda'
##
## The following object is masked from 'package:graphics':
##
##     matplot
##
## Loading required package: fields

## Warning: package 'fields' was built under R version 4.3.3

## Loading required package: spam

## Warning: package 'spam' was built under R version 4.3.3
```

```
## Spam version 2.10-0 (2023-10-23) is loaded.
## Type 'help( Spam)' or 'demo( spam)' for a short introduction
## and overview of this package.
## Help for individual functions is also obtained by adding the
## suffix '.spam' to the function name, e.g. 'help( chol.spam)'.
##
## Attaching package: 'spam'
##
## The following object is masked from 'package:Matrix':
##
##     det
##
## The following objects are masked from 'package:base':
##
##     backsolve, forwardsolve
##
## Loading required package: viridisLite
##
## Try help(fields) to get started.
```

```
library(psych)
```

```
## Warning: package 'psych' was built under R version 4.3.3
```

```
##
## Attaching package: 'psych'
##
## The following object is masked from 'package:fields':
##
##     describe
##
## The following objects are masked from 'package:ggplot2':
##
##     %+%, alpha
```

```
library(CCP)
library(GGally)
```

```
## Warning: package 'GGally' was built under R version 4.3.3
```

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

Load the data set

```
health_lifestyle_data <- read_csv(file = "../R studio
code/Data/Sleep_health_and_lifestyle_dataset.csv")
```

```
## Rows: 374 Columns: 13
## ── Column specification
────────────────────────────────────────────────────────
```

```
## Delimiter: ","
## chr (5): Gender, Occupation, BMI_Category, Blood_Pressure, Sleep_Disorder
## dbl (8): Person_ID, Age, Sleep_Duration, Quality_of_Sleep,
Physical_Activity...
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this
message.
```

```
view(health_lifestyle_data)
#diabetes_data
```

Remove empty rows and columns

```
health_lifestyle_data <- health_lifestyle_data %>%
remove_empty(c("cols","rows"))
```

Short summery for data datasets

```
glimpse(health_lifestyle_data)
```

```
## Rows: 374
## Columns: 13
## $ Person_ID               <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13,
14,…
## $ Gender                  <chr> "Male", "Male", "Male", "Male", "Male",
"Male"…
## $ Age                     <dbl> 27, 28, 28, 28, 28, 28, 29, 29, 29, 29,
29, 29…
## $ Occupation              <chr> "Software Engineer", "Doctor", "Doctor",
"Sale…
## $ Sleep_Duration          <dbl> 6.1, 6.2, 6.2, 5.9, 5.9, 5.9, 6.3, 7.8,
7.8, 7…
## $ Quality_of_Sleep        <dbl> 6, 6, 6, 4, 4, 4, 6, 7, 7, 7, 6, 7, 6, 6,
6, 6…
## $ Physical_Activity_Level <dbl> 42, 60, 60, 30, 30, 30, 40, 75, 75, 75,
30, 75…
## $ Stress_Level            <dbl> 6, 8, 8, 8, 8, 8, 7, 6, 6, 6, 8, 6, 8, 8,
8, 8…
## $ BMI_Category            <chr> "Overweight", "Normal", "Normal", "Obese",
"Ob…
## $ Blood_Pressure          <chr> "126/83", "125/80", "125/80", "140/90",
"140/9…
## $ Heart_Rate              <dbl> 77, 75, 75, 85, 85, 85, 82, 70, 70, 70,
70, 70…
## $ Daily_Steps             <dbl> 4200, 10000, 10000, 3000, 3000, 3000,
3500, 80…
## $ Sleep_Disorder          <chr> "None", "None", "None", "Sleep Apnea",
"Sleep …
```

checking missing values

```
sum(is.na(health_lifestyle_data))

## [1] 0
```

There is no missing values in the data set.

Dimension of data set

```
dim(health_lifestyle_data)

## [1] 374  13
```

summery of data set

```
summary(health_lifestyle_data)

##    Person_ID        Gender               Age         Occupation
##  Min.   :  1.00   Length:374         Min.   :27.00   Length:374
##  1st Qu.: 94.25   Class :character   1st Qu.:35.25   Class :character
##  Median :187.50   Mode  :character   Median :43.00   Mode  :character
##  Mean   :187.50                      Mean   :42.18
##  3rd Qu.:280.75                      3rd Qu.:50.00
##  Max.   :374.00                      Max.   :59.00
##  Sleep_Duration  Quality_of_Sleep Physical_Activity_Level  Stress_Level
##  Min.   :5.800   Min.   :4.000    Min.   :30.00           Min.   :3.000
##  1st Qu.:6.400   1st Qu.:6.000    1st Qu.:45.00           1st Qu.:4.000
##  Median :7.200   Median :7.000    Median :60.00           Median :5.000
##  Mean   :7.132   Mean   :7.313    Mean   :59.17           Mean   :5.385
##  3rd Qu.:7.800   3rd Qu.:8.000    3rd Qu.:75.00           3rd Qu.:7.000
##  Max.   :8.500   Max.   :9.000    Max.   :90.00           Max.   :8.000
##  BMI_Category       Blood_Pressure       Heart_Rate     Daily_Steps
##  Length:374         Length:374         Min.   :65.00   Min.   : 3000
##  Class :character   Class :character   1st Qu.:68.00   1st Qu.: 5600
##  Mode  :character   Mode  :character   Median :70.00   Median : 7000
##                                        Mean   :70.17   Mean   : 6817
##                                        3rd Qu.:72.00   3rd Qu.: 8000
##                                        Max.   :86.00   Max.   :10000
##  Sleep_Disorder
##  Length:374
##  Class :character
##  Mode  :character
##
##
##
```

categorical variable : Gender, occupation , BMI Category ,Sleep Disorder

```
describeBy(health_lifestyle_data)

## Warning in describeBy(health_lifestyle_data): no grouping variable
requested
```

```
##                          vars   n    mean      sd median trimmed     mad
min
## Person_ID                   1 374  187.50  108.11  187.5  187.50  138.62
1.0
## Gender*                     2 374    1.51    0.50    2.0    1.51    0.00
1.0
## Age                         3 374   42.18    8.67   43.0   41.84   10.38
27.0
## Occupation*                 4 374    4.77    3.06    4.0    4.47    2.97
1.0
## Sleep_Duration              5 374    7.13    0.80    7.2    7.12    1.04
5.8
## Quality_of_Sleep            6 374    7.31    1.20    7.0    7.32    1.48
4.0
## Physical_Activity_Level     7 374   59.17   20.83   60.0   58.97   22.24
30.0
## Stress_Level                8 374    5.39    1.77    5.0    5.36    2.97
3.0
## BMI_Category*               9 374    1.82    0.97    1.0    1.77    0.00
1.0
## Blood_Pressure*            10 374   14.11    7.10   16.0   14.48    8.90
1.0
## Heart_Rate                 11 374   70.17    4.14   70.0   69.74    2.97
65.0
## Daily_Steps                12 374 6816.84 1617.92 7000.0 6732.67 1482.60
3000.0
## Sleep_Disorder*            13 374    2.00    0.64    2.0    2.00    0.00
1.0
##                             max  range  skew kurtosis    se
## Person_ID                 374.0  373.0  0.00    -1.21  5.59
## Gender*                     2.0    1.0 -0.02    -2.00  0.03
## Age                        59.0   32.0  0.26    -0.92  0.45
## Occupation*                11.0   10.0  0.74    -0.53  0.16
## Sleep_Duration              8.5    2.7  0.04    -1.29  0.04
## Quality_of_Sleep            9.0    5.0 -0.21    -0.77  0.06
## Physical_Activity_Level    90.0   60.0  0.07    -1.27  1.08
## Stress_Level                8.0    5.0  0.15    -1.33  0.09
## BMI_Category*               3.0    2.0  0.37    -1.84  0.05
## Blood_Pressure*            25.0   24.0 -0.18    -0.96  0.37
## Heart_Rate                 86.0   21.0  1.22     2.21  0.21
## Daily_Steps             10000.0 7000.0  0.18    -0.42 83.66
## Sleep_Disorder*             3.0    2.0  0.00    -0.60  0.03
```

Convert categorical BMI Category to numerical

```
health_lifestyle_data <- health_lifestyle_data %>%
  mutate(BMI_Numeric = case_when(
    BMI_Category == "Underweight" ~ 0,
    BMI_Category == "Normal" ~ 1,
```

```
    BMI_Category == "Overweight" ~ 2,
    BMI_Category == "Obese" ~ 3
  ))

health_lifestyle_data %>% tabyl(BMI_Numeric)

##  BMI_Numeric   n    percent
##            1 216 0.57754011
##            2 148 0.39572193
##            3  10 0.02673797

health_lifestyle_data <- health_lifestyle_data %>%
relocate(BMI_Numeric,.after = BMI_Category)
#health_lifestyle_data
```

Convert categorical Sleep Disorder variable Category to numerical

```
health_lifestyle_data %>% tabyl(Sleep_Disorder)

##  Sleep_Disorder   n   percent
##        Insomnia  77 0.2058824
##            None 219 0.5855615
##     Sleep Apnea  78 0.2085561

health_lifestyle_data <- health_lifestyle_data %>%
  mutate(sleep_disorder_numeric = case_when(
    Sleep_Disorder == "None" ~ 1,
    Sleep_Disorder == "Insomnia" ~ 2,
    Sleep_Disorder == "Sleep Apnea" ~ 3
  ))
#health_lifestyle_data
```

Split the blood pressure values into systolic and diastolic

```
health_lifestyle_data <- health_lifestyle_data %>%
separate(Blood_Pressure, into = c("systolic", "diastolic"), sep = "/",
convert = TRUE)
```

Categorize blood pressure readings based on AHA guidelines

```
health_lifestyle_data <- health_lifestyle_data %>%
mutate(blood_pressure_category = case_when(
(systolic < 120 & diastolic < 80) ~ "Normal",
(systolic >= 120 & systolic <= 129) | diastolic < 80 ~  "Elevated",
(systolic >= 130 & systolic <= 139) | (diastolic >= 80 & diastolic < 89) ~
"Hypertension Stage 1",
systolic >= 140 | diastolic >= 90 ~  "Hypertension Stage 2",
systolic > 180 | diastolic > 120 ~  "Hypertensive Crisis",
TRUE ~ "Unknown"
))

health_lifestyle_data %>% tabyl(blood_pressure_category)
```

```
##   blood_pressure_category   n    percent
##                  Elevated 125 0.3342246
##      Hypertension Stage 1 137 0.3663102
##      Hypertension Stage 2  71 0.1898396
##                    Normal  41 0.1096257
```

```
health_lifestyle_data <- health_lifestyle_data %>%
relocate(blood_pressure_category,.after = diastolic)

health_lifestyle_data <- health_lifestyle_data %>%
  mutate(blood_pressure_category_numeric = case_when(
    blood_pressure_category == "Normal" ~ 1,
    blood_pressure_category == "Elevated" ~ 2,
    blood_pressure_category == "Hypertension Stage 1" ~ 3,
    blood_pressure_category == "Hypertension Stage 2" ~ 4,
    blood_pressure_category == "Hypertensive Crisis" ~ 5

  ))

health_lifestyle_data <- health_lifestyle_data %>%
relocate(blood_pressure_category_numeric,.after = blood_pressure_category)
#health_lifestyle_data
```

check whether there are outliers in out data set or not

```
#boxplot(health_lifestyle_data$sleep_disorder_numeric)
#boxplot(health_lifestyle_data$Quality_of_Sleep)
#boxplot(health_lifestyle_data$Physical_Activity_Level)
#boxplot(health_lifestyle_data$Age)

#boxplot(health_lifestyle_data$Heart_Rate)
#boxplot(health_lifestyle_data$Stress_Level)
#boxplot(health_lifestyle_data$Daily_Steps)
```

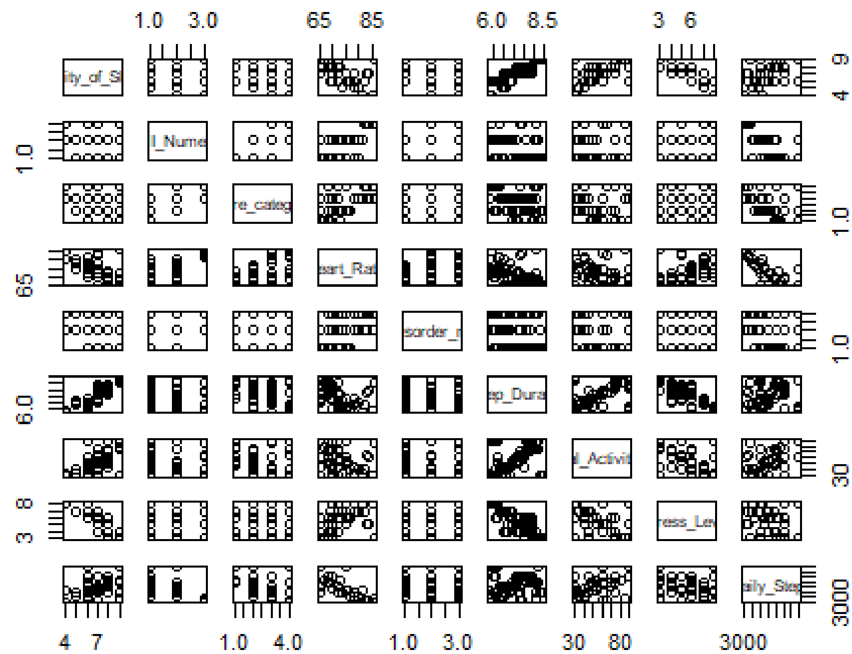
Canonical Correlation Analysis

canonical correlation analysis is actually used to examine the relationship between two data sets, but today we will separate the data set we have into two separate data sets and examine the relationship between the two data sets. Two Clear Sets: We can clearly separate the variables into two meaningful sets: Health Metrics: Quality of Sleep, Blood Pressure, Heart Rate,BMI (converted to numerical) and Sleep Disorder Lifestyle Factors: Physical Activity Level, Daily steps,Stress Level and Sleep Duration

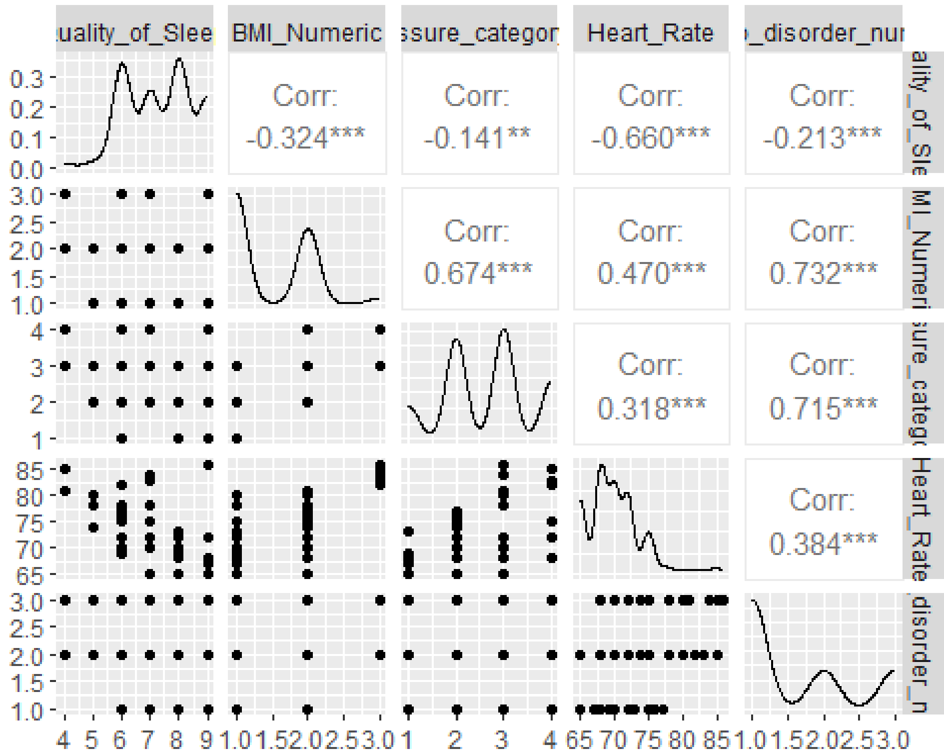
```
health_metrics <- subset(health_lifestyle_data ,select =
c(Quality_of_Sleep,BMI_Numeric,blood_pressure_category_numeric,Heart_Rate,sle
ep_disorder_numeric))

lifestyle_factors <- subset(health_lifestyle_data ,select =
c(Sleep_Duration,Physical_Activity_Level,Stress_Level,Daily_Steps))
```

Create pairwise scatterplot with variables of data sets

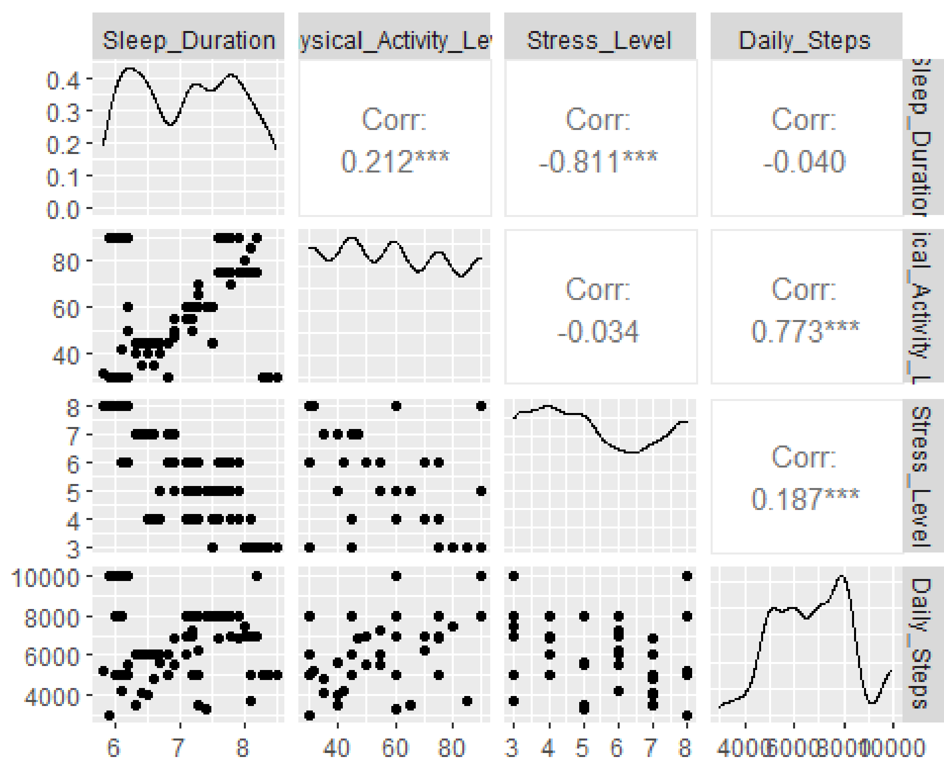
```r
# Combine two sets into a single dataset
combined_data <- cbind(health_metrics,lifestyle_factors)

# Create pairwise scatter plot matrix
scatterplot <- pairs(combined_data)
```




```r
scatterplot

## NULL

ggpairs(health_metrics)
```




```
ggpairs(lifestyle_factors)
```




```
matcor(health_metrics,lifestyle_factors)
```

```
## $Xcor
##                                 Quality_of_Sleep BMI_Numeric
## Quality_of_Sleep                       1.0000000  -0.3244127
## BMI_Numeric                           -0.3244127   1.0000000
## blood_pressure_category_numeric       -0.1410358   0.6744936
## Heart_Rate                            -0.6598647   0.4698965
## sleep_disorder_numeric                -0.2130026   0.7324788
##                                 blood_pressure_category_numeric Heart_Rate
## Quality_of_Sleep                                     -0.1410358 -0.6598647
## BMI_Numeric                                           0.6744936  0.4698965
## blood_pressure_category_numeric                       1.0000000  0.3179609
## Heart_Rate                                            0.3179609  1.0000000
## sleep_disorder_numeric                                0.7149036  0.3843434
##                                 sleep_disorder_numeric
## Quality_of_Sleep                            -0.2130026
## BMI_Numeric                                  0.7324788
## blood_pressure_category_numeric              0.7149036
## Heart_Rate                                   0.3843434
## sleep_disorder_numeric                       1.0000000
##
## $Ycor
##                         Sleep_Duration Physical_Activity_Level
Stress_Level
## Sleep_Duration              1.00000000              0.21236031  -
0.81102303
## Physical_Activity_Level     0.21236031              1.00000000  -
0.03413446
## Stress_Level               -0.81102303             -0.03413446
1.00000000
## Daily_Steps                -0.03953254              0.77272305
0.18682895
##                         Daily_Steps
## Sleep_Duration          -0.03953254
## Physical_Activity_Level  0.77272305
## Stress_Level             0.18682895
## Daily_Steps              1.00000000
##
## $XYcor
##                                 Quality_of_Sleep BMI_Numeric
## Quality_of_Sleep                      1.00000000 -0.32441269
## BMI_Numeric                          -0.32441269  1.00000000
## blood_pressure_category_numeric      -0.14103576  0.67449362
## Heart_Rate                           -0.65986473  0.46989650
## sleep_disorder_numeric               -0.21300261  0.73247882
## Sleep_Duration                        0.88321300 -0.34960530
## Physical_Activity_Level               0.19289645  0.05207636
## Stress_Level                         -0.89875203  0.15764925
## Daily_Steps                           0.01679141 -0.14278493
##                                 blood_pressure_category_numeric
Heart_Rate
```

```
## Quality_of_Sleep                                    -0.1410358 -
0.65986473
## BMI_Numeric                                          0.6744936
0.46989650
## blood_pressure_category_numeric                      1.0000000
0.31796088
## Heart_Rate                                           0.3179609
1.00000000
## sleep_disorder_numeric                               0.7149036
0.38434340
## Sleep_Duration                                      -0.1530167 -
0.51645489
## Physical_Activity_Level                              0.3736833
0.13697098
## Stress_Level                                         0.1678351
0.67002646
## Daily_Steps                                          0.2360696 -
0.03030858
##                                 sleep_disorder_numeric Sleep_Duration
## Quality_of_Sleep                            -0.2130026     0.88321300
## BMI_Numeric                                  0.7324788    -0.34960530
## blood_pressure_category_numeric              0.7149036    -0.15301672
## Heart_Rate                                   0.3843434    -0.51645489
## sleep_disorder_numeric                       1.0000000    -0.23910857
## Sleep_Duration                              -0.2391086     1.00000000
## Physical_Activity_Level                      0.2365600     0.21236031
## Stress_Level                                 0.1518935    -0.81102303
## Daily_Steps                                  0.1120534    -0.03953254
##                                 Physical_Activity_Level Stress_Level
## Quality_of_Sleep                             0.19289645  -0.89875203
## BMI_Numeric                                  0.05207636   0.15764925
## blood_pressure_category_numeric              0.37368326   0.16783507
## Heart_Rate                                   0.13697098   0.67002646
## sleep_disorder_numeric                       0.23656005   0.15189350
## Sleep_Duration                               0.21236031  -0.81102303
## Physical_Activity_Level                      1.00000000  -0.03413446
## Stress_Level                                -0.03413446   1.00000000
## Daily_Steps                                  0.77272305   0.18682895
##                                 Daily_Steps
## Quality_of_Sleep                 0.01679141
## BMI_Numeric                     -0.14278493
## blood_pressure_category_numeric  0.23606962
## Heart_Rate                      -0.03030858
## sleep_disorder_numeric           0.11205341
## Sleep_Duration                  -0.03953254
## Physical_Activity_Level          0.77272305
## Stress_Level                     0.18682895
## Daily_Steps                      1.00000000
```

Perform Canonical Correlation Analysis

```r
cca_result1 <- cc(health_metrics,lifestyle_factors)

# View the results
summary(cca_result1)

##        Length Class  Mode
## cor     4     -none- numeric
## names   3     -none- list
## xcoef  20     -none- numeric
## ycoef  16     -none- numeric
## scores  6     -none- list
```

Canonical correlations

```r
cca_result1$cor

## [1] 0.9553533 0.7009142 0.6138956 0.3753121
```

#Canonical coefficients

```r
cca_result1$xcoef

##                                        [,1]       [,2]        [,3]
[,4]
## Quality_of_Sleep                -0.73821895  0.6689135 -0.44469759 -
0.2434853
## BMI_Numeric                     -0.29753716 -1.3293501 -2.37175387 -
0.4980830
## blood_pressure_category_numeric  0.14453297  0.7265326  0.58023293
0.7682200
## Heart_Rate                       0.05150106  0.2867597 -0.09105226 -
0.1546095
## sleep_disorder_numeric          -0.03736169 -0.1001544  0.20174417
0.7915181

cca_result1$ycoef

##                                  [,1]          [,2]          [,3]
[,4]
## Sleep_Duration          -0.3733756533  1.2063667831  1.2840264494 -
1.3899048635
## Physical_Activity_Level  0.0090771566  0.0499729805 -0.0608361886
0.0209808726
## Stress_Level             0.4444873954  0.5722874271  0.4161509736 -
0.5497567842
## Daily_Steps             -0.0001693747 -0.0003888957  0.0008789486
0.0003556914
```

Structure correlations (canonical loadings)

```r
cca_result2 <- comput(health_metrics,lifestyle_factors, cca_result1)
```

Test canonical dimentions

```
rho <- cca_result1$cor
rho

## [1] 0.9553533 0.7009142 0.6138956 0.3753121
```

Define number of observations, number of variables in first set, and number of variables in the second set

```
n <- dim(health_metrics)[1]
p <- length(health_metrics)
q <- length(lifestyle_factors)
```

Calculate p-values using the F-approximations of different test statistics:

```
p.asym(rho, n, p, q, tstat = "Wilks")

## Wilks' Lambda, using F-approximation (Rao's F):
##                stat    approx df1       df2      p.value
## 1 to 4:  0.02377591 126.45481  20 1211.5179 0.000000e+00
## 2 to 4:  0.27234710  51.25269  12  968.6365 0.000000e+00
## 3 to 4:  0.53535826  44.86139   6  734.0000 0.000000e+00
## 4 to 4:  0.85914082  30.16745   2  368.0000 7.376322e-13

p.asym(rho, n, p, q, tstat = "Hotelling")

##  Hotelling-Lawley Trace, using F-approximation:
##                stat    approx df1  df2      p.value
## 1 to 4:  12.1892185 221.53905  20 1454 0.000000e+00
## 2 to 4:   1.7344700  52.82907  12 1462 0.000000e+00
## 3 to 4:   0.7687496  47.08591   6 1470 0.000000e+00
## 4 to 4:   0.1639535  30.29042   2 1478 1.281197e-13

p.asym(rho, n, p, q, tstat = "Pillai")

##  Pillai-Bartlett Trace, using F-approximation:
##               stat   approx df1  df2     p.value
## 1 to 4:  1.9217077 68.05476  20 1472 0.00000e+00
## 2 to 4:  1.0090077 41.60635  12 1480 0.00000e+00
## 3 to 4:  0.5177270 36.87141   6 1488 0.00000e+00
## 4 to 4:  0.1408592 27.30210   2 1496 2.25997e-12

p.asym(rho, n, p, q, tstat = "Roy")

##  Roy's Largest Root, using F-approximation:
##            stat   approx df1 df2 p.value
## 1 to 1:  0.9127 964.4505   4 369       0
##
##  F statistic for Roy's Greatest Root is an upper bound.
```

As shown in the table above, the first test,second test and third test of the canonical dimensions tests whether all three dimensions are significant. Therefore dimensions 1,2, and 3 must each be significant while dimension four is not.

standardized psych canonical coefficients diagonal matrix of health matrics sd's

```
sd1 <- diag(sqrt(diag(cov(health_metrics))))
sd1 %*% cca_result1$xcoef
```

```
##             [,1]        [,2]       [,3]       [,4]
## [1,] -0.88361555  0.80066000 -0.5322834 -0.2914412
## [2,] -0.16342938 -0.73017725 -1.3027424 -0.2735839
## [3,]  0.13187883  0.66292328  0.5294324  0.7009609
## [4,]  0.21299169  1.18594496 -0.3765626 -0.6394146
## [5,] -0.03020832 -0.08097855  0.1631177  0.6399719
```

standardized psych canonical coefficients diagonal matrix of lifestyle facors sd's

```
sd2 <- diag(sqrt(diag(cov(lifestyle_factors))))
sd2 %*% cca_result1$ycoef
```

```
##            [,1]       [,2]       [,3]       [,4]
## [1,] -0.2970789  0.9598539  1.0216443 -1.1058872
## [2,]  0.1890845  1.0409773 -1.2672667  0.4370484
## [3,]  0.7887546  1.0155392  0.7384709 -0.9755580
## [4,] -0.2740340 -0.6292004  1.4220648  0.5754786
```

download images

```
postscript(file = "scatterplot.eps",width = 6,height = 4,horizontal = FALSE)
dev.off()
```

```
## png
##   2
```