hwk1 615

Handing Zhang

9/22/2021

hand_functions.R

We created a sum_special function in which we return the basic statistics of each column in a data frame, including their means, variances, covariances and correlations.

```
sum_special <- function(df_x){

  ## sum_special calculates data summary statistics
  ## the input param df_x is the data frame of input values

  #  browser()  # browser() will start the debugger
               # if the line is uncommented


  ## test the input data to assure that it is a data frame.
  try(if(!is.data.frame(df_x)) stop("Input data must be a data frame."))


  sp_means <- apply(df_x, MARGIN = 2, FUN = mean)
  sp_var <- apply(df_x, MARGIN = 2, FUN = var)
  sp_cov <- cov(df_x)
  sp_cor <- cor(df_x)

  ## Note that defining a list with the
  ## syntax list(list_name = list_content) produces
  ## named list items
  sp_outputs <- list(sp_means=sp_means,
                     sp_var = sp_var,
                     sp_cov = sp_cov,
                     sp_cor = sp_cor)

  return(sp_outputs)
}
```

car_viz.R In this exploration of mtcars, we did some data wrangling, in which we created a subset of rows of mtcars with only cars whose mpg is less than 20, as well as indexing certain columns. We then utilized the sum_special function we just definedt to extract some basic statistics from the data frame. We could also do some data visualizations such as scatter plot and boxplot to get a more intuitive and clear view of the trend and distribution of the data.

```r
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.3     v dplyr   1.0.7
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   2.0.1     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
# call built-in data mtcars.
data(mtcars)

# Select only car models where mpg<20
mtcars_mpg2 <- mtcars[mtcars$mpg < 20,]


# Reduce the variables to mpg, cyl, disp, hp, gears
mtcars_mpg2 <- mtcars_mpg2[, c(1,2,3,4,10)]

# read the R file hand_functions.R so that it can be used
# notice that with echo = TRUE
source(file = "hand_functions.R", echo = TRUE)
```

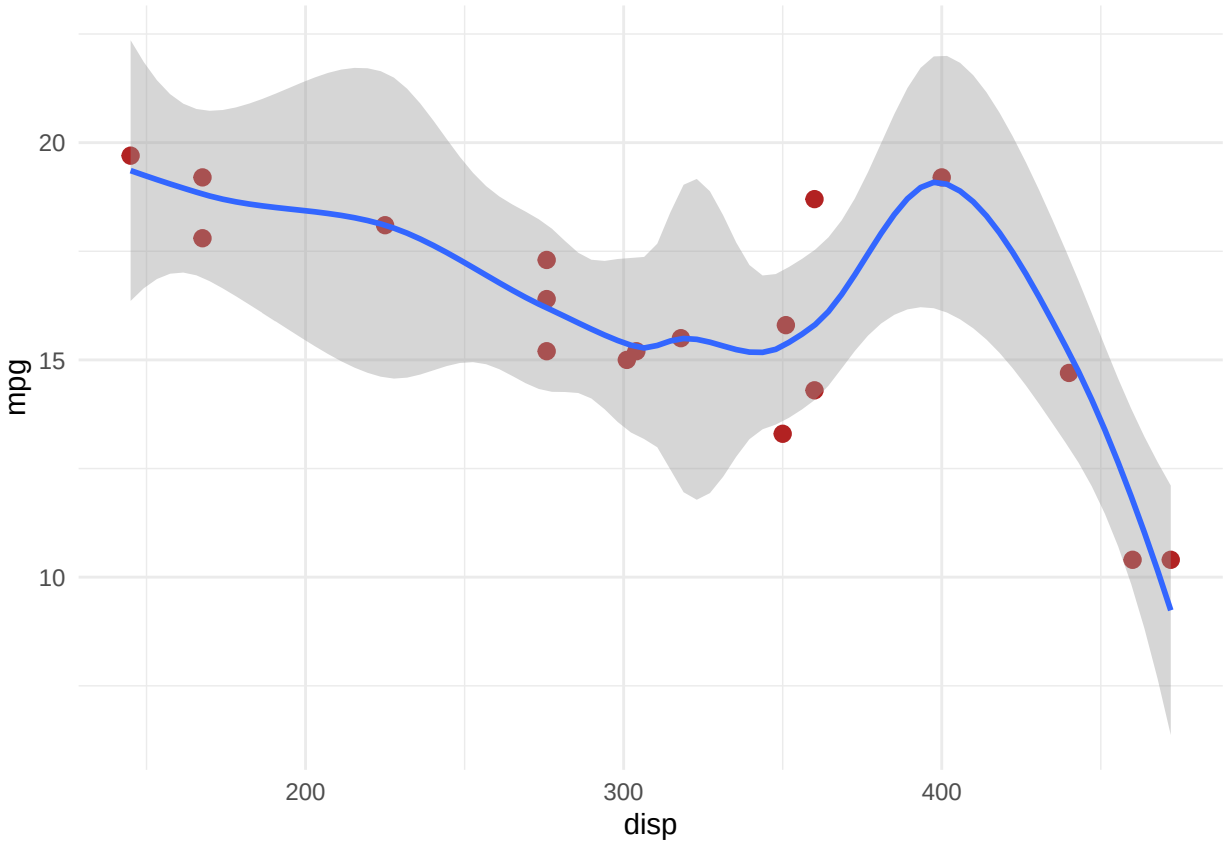
```
##
## > sum_special <- function(df_x) {
## +     try(if (!is.data.frame(df_x))
## +         stop("Input data must be a data frame."))
## +     `?`(try)
## +     `?`(ap .... [TRUNCATED]
##
## > head(mtcars)
##                   mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
##
## > sum_special(mtcars)
## $sp_means
##        mpg        cyl       disp         hp       drat         wt       qsec
##  20.090625   6.187500 230.721875 146.687500   3.596563   3.217250  17.848750
##         vs         am       gear       carb
##   0.437500   0.406250   3.687500   2.812500
##
## $sp_var
##          mpg          cyl         disp           hp         drat           wt
```

```
## 3.632410e+01 3.189516e+00 1.536080e+04 4.700867e+03 2.858814e-01 9.573790e-01
##         qsec           vs           am         gear         carb
## 3.193166e+00 2.540323e-01 2.489919e-01 5.443548e-01 2.608871e+00
##
## $sp_cov
##              mpg         cyl        disp          hp         drat          wt
## mpg    36.324103  -9.1723790  -633.09721 -320.732056   2.19506351  -5.1166847
## cyl    -9.172379   3.1895161   199.66028  101.931452  -0.66836694   1.3673710
## disp -633.097208 199.6602823 15360.79983 6721.158669 -47.06401915 107.6842040
## hp   -320.732056 101.9314516  6721.15867 4700.866935 -16.45110887  44.1926613
## drat    2.195064  -0.6683669   -47.06402  -16.451109   0.28588135  -0.3727207
## wt     -5.116685   1.3673710   107.68420   44.192661  -0.37272073   0.9573790
## qsec    4.509149  -1.8868548   -96.05168  -86.770081   0.08714073  -0.3054816
## vs      2.017137  -0.7298387   -44.37762  -24.987903   0.11864919  -0.2736613
## am      1.803931  -0.4657258   -36.56401   -8.320565   0.19015121  -0.3381048
## gear    2.135685  -0.6491935   -50.80262   -6.358871   0.27598790  -0.4210806
## carb   -5.363105   1.5201613    79.06875   83.036290  -0.07840726   0.6757903
##              qsec           vs           am        gear        carb
## mpg    4.50914919   2.01713710   1.80393145   2.1356855 -5.36310484
## cyl   -1.88685484  -0.72983871  -0.46572581  -0.6491935  1.52016129
## disp -96.05168145 -44.37762097 -36.56401210 -50.8026210 79.06875000
## hp   -86.77008065 -24.98790323  -8.32056452  -6.3588710 83.03629032
## drat   0.08714073   0.11864919   0.19015121   0.2759879 -0.07840726
## wt    -0.30548161  -0.27366129  -0.33810484  -0.4210806  0.67579032
## qsec   3.19316613   0.67056452  -0.20495968  -0.2804032 -1.89411290
## vs     0.67056452   0.25403226   0.04233871   0.0766129 -0.46370968
## am    -0.20495968   0.04233871   0.24899194   0.2923387  0.04637097
## gear  -0.28040323   0.07661290   0.29233871   0.5443548  0.32661290
## carb  -1.89411290  -0.46370968   0.04637097   0.3266129  2.60887097
##
## $sp_cor
##             mpg        cyl       disp         hp        drat         wt
## mpg   1.0000000 -0.8521620 -0.8475514 -0.7761684  0.68117191 -0.8676594
## cyl  -0.8521620  1.0000000  0.9020329  0.8324475 -0.69993811  0.7824958
## disp -0.8475514  0.9020329  1.0000000  0.7909486 -0.71021393  0.8879799
## hp   -0.7761684  0.8324475  0.7909486  1.0000000 -0.44875912  0.6587479
## drat  0.6811719 -0.6999381 -0.7102139 -0.4487591  1.00000000 -0.7124406
## wt   -0.8676594  0.7824958  0.8879799  0.6587479 -0.71244065  1.0000000
## qsec  0.4186840 -0.5912421 -0.4336979 -0.7082234  0.09120476 -0.1747159
## vs    0.6640389 -0.8108118 -0.7104159 -0.7230967  0.44027846 -0.5549157
## am    0.5998324 -0.5226070 -0.5912270 -0.2432043  0.71271113 -0.6924953
## gear  0.4802848 -0.4926866 -0.5555692 -0.1257043  0.69961013 -0.5832870
## carb -0.5509251  0.5269883  0.3949769  0.7498125 -0.09078980  0.4276059
##             qsec         vs          am       gear        carb
## mpg   0.41868403  0.6640389  0.59983243  0.4802848 -0.55092507
## cyl  -0.59124207 -0.8108118 -0.52260705 -0.4926866  0.52698829
## disp -0.43369788 -0.7104159 -0.59122704 -0.5555692  0.39497686
## hp   -0.70822339 -0.7230967 -0.24320426 -0.1257043  0.74981247
## drat  0.09120476  0.4402785  0.71271113  0.6996101 -0.09078980
## wt   -0.17471588 -0.5549157 -0.69249526 -0.5832870  0.42760594
## qsec  1.00000000  0.7445354 -0.22986086 -0.2126822 -0.65624923
## vs    0.74453544  1.0000000  0.16834512  0.2060233 -0.56960714
## am   -0.22986086  0.1683451  1.00000000  0.7940588  0.05753435
## gear -0.21268223  0.2060233  0.79405876  1.0000000  0.27407284
```

```
## carb -0.65624923 -0.5696071  0.05753435  0.2740728  1.00000000
##
##
## > mean(mtcars$mpg)
## [1] 20.09062
```

```
?source
# Now use the function from hand_functions.R

sp_out <- sum_special(mtcars_mpg2)


# library(esquisse)
#
# esquisser(data = mtcars_mpg2, viewer = "browser")



ggplot(mtcars_mpg2) +
  aes(x = disp, y = mpg) +
  geom_point(shape = "bullet", size = 4L, colour = "#B22222") +
  geom_smooth(span = 0.5) +
  theme_minimal()
```

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



```r
# note that this boxplot cannot be made with esquisse() unless
# the data is adjusted.  What adjustment is needed?


ggplot(mtcars_mpg2, aes(x=as.factor(cyl), y=mpg)) +
  geom_boxplot(fill="slateblue", alpha=0.2) +
  xlab("cyl")
```