|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| R version 3.5.1 (2018-07-02) -- "Feather Spray"  Copyright (C) 2018 The R Foundation for Statistical Computing  Platform: x86\_64-w64-mingw32/x64 (64-bit)  R is free software and comes with ABSOLUTELY NO WARRANTY.  You are welcome to redistribute it under certain conditions.  Type 'license()' or 'licence()' for distribution details.  R is a collaborative project with many contributors.  Type 'contributors()' for more information and  'citation()' on how to cite R or R packages in publications.  Type 'demo()' for some demos, 'help()' for on-line help, or  'help.start()' for an HTML browser interface to help.  Type 'q()' to quit R.  [Workspace loaded from ~/.RData]  > install.packages(psych)  Error in install.packages : object 'psych' not found  > library(psych)  > hd <- housedata\_for\_final\_2018  Error: object 'housedata\_for\_final\_2018' not found  > library(readxl)  > housedata\_for\_final\_2018 <- read\_excel("C:/Users/VIP/Desktop/Final Exam/housedata for final 2018.xlsx")  > View(housedata\_for\_final\_2018)  > hd <- housedata\_for\_final\_2018  > cor(hd)  s\_p inv bath ltsz hssz bsemt a\_c f\_place  s\_p 1.00000000 0.075265157 0.4680243 0.48471784 0.40302121 0.160443217 0.26820048 0.48467559  inv 0.07526516 1.000000000 0.0517719 0.10366549 0.09023423 -0.050458023 0.11441777 0.11836174  bath 0.46802425 0.051771898 1.0000000 0.21429765 0.27978313 0.105962360 0.23309172 0.34917937  ltsz 0.48471784 0.103665486 0.2142976 1.00000000 0.26136144 0.078818751 0.08429919 0.22683247  hssz 0.40302121 0.090234233 0.2797831 0.26136144 1.00000000 0.017621117 0.18327055 0.25517864  bsemt 0.16044322 -0.050458023 0.1059624 0.07881875 0.01762112 1.000000000 0.05358194 0.05125987  a\_c 0.26820048 0.114417775 0.2330917 0.08429919 0.18327055 0.053581944 1.00000000 0.30114480  f\_place 0.48467559 0.118361737 0.3491794 0.22683247 0.25517864 0.051259874 0.30114480 1.00000000  garsz\_a 0.15953057 0.118327282 0.1337858 0.07423996 0.13978754 0.036196614 0.25977997 0.14119742  dw 0.34707740 0.234965231 0.3001625 0.14443338 0.26164452 0.069851977 0.40853420 0.37369511  dr 0.33738495 0.089345811 0.2510873 0.15648776 0.20381042 0.056018276 0.18540800 0.28434982  fr 0.28017849 0.104303707 0.2620055 0.12591320 0.16660137 0.187916294 0.31063105 0.33032655  age5 0.26594159 0.004636197 0.1503503 0.07435705 0.16892929 0.002652432 0.18142449 0.16577757  stl10 -0.22700986 -0.013066303 -0.1321319 -0.05363533 0.08870037 -0.117686819 -0.14163114 -0.22167356  bdrms 0.42344906 0.033068774 0.3730989 0.17175661 0.20040390 0.075778889 0.15776810 0.25943610  garsz\_a dw dr fr age5 stl10 bdrms  s\_p 0.15953057 0.34707740 0.33738495 0.28017849 0.265941586 -0.22700986 0.42344906  inv 0.11832728 0.23496523 0.08934581 0.10430371 0.004636197 -0.01306630 0.03306877  bath 0.13378577 0.30016246 0.25108732 0.26200551 0.150350283 -0.13213186 0.37309895  ltsz 0.07423996 0.14443338 0.15648776 0.12591320 0.074357052 -0.05363533 0.17175661  hssz 0.13978754 0.26164452 0.20381042 0.16660137 0.168929286 0.08870037 0.20040390  bsemt 0.03619661 0.06985198 0.05601828 0.18791629 0.002652432 -0.11768682 0.07577889  a\_c 0.25977997 0.40853420 0.18540800 0.31063105 0.181424493 -0.14163114 0.15776810  f\_place 0.14119742 0.37369511 0.28434982 0.33032655 0.165777565 -0.22167356 0.25943610  garsz\_a 1.00000000 0.26569611 0.10002019 0.15309227 0.054111950 -0.04743284 0.09016903  dw 0.26569611 1.00000000 0.30294577 0.32847429 0.171925207 -0.17520240 0.26670806  dr 0.10002019 0.30294577 1.00000000 0.14368573 0.042265191 -0.24903742 0.17570156  fr 0.15309227 0.32847429 0.14368573 1.00000000 -0.015465413 -0.19379137 0.32415282  age5 0.05411195 0.17192521 0.04226519 -0.01546541 1.000000000 -0.07106999 -0.03366809  stl10 -0.04743284 -0.17520240 -0.24903742 -0.19379137 -0.071069993 1.00000000 -0.20626502  bdrms 0.09016903 0.26670806 0.17570156 0.32415282 -0.033668087 -0.20626502 1.00000000  > summary(hd)  s\_p inv bath ltsz hssz bsemt  Min. : 29864 Min. : 61.0 Min. : 1.000 Min. :0.01561 Min. : 216 Min. :0.0000  1st Qu.: 59278 1st Qu.:101.0 1st Qu.: 1.500 1st Qu.:0.17880 1st Qu.: 875 1st Qu.:1.0000  Median : 70360 Median :135.0 Median : 2.000 Median :0.22495 Median :1014 Median :1.0000  Mean : 79037 Mean :140.8 Mean : 1.948 Mean :0.33333 Mean :1055 Mean :0.9368  3rd Qu.: 90741 3rd Qu.:154.0 3rd Qu.: 2.500 3rd Qu.:0.27548 3rd Qu.:1199 3rd Qu.:1.0000  Max. :222680 Max. :322.0 Max. :20.000 Max. :4.13223 Max. :3080 Max. :1.0000  a\_c f\_place garsz\_a dw dr fr  Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :0.0000  1st Qu.:1.0000 1st Qu.:0.0000 1st Qu.:1.0000 1st Qu.:1.0000 1st Qu.:1.0000 1st Qu.:1.0000  Median :1.0000 Median :1.0000 Median :1.0000 Median :1.0000 Median :1.0000 Median :1.0000  Mean :0.8137 Mean :0.5968 Mean :0.9611 Mean :0.8063 Mean :0.7589 Mean :0.7821  3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:1.0000  Max. :1.0000 Max. :1.0000 Max. :2.0000 Max. :1.0000 Max. :1.0000 Max. :1.0000  age5 stl10 bdrms  Min. :0.0000 Min. :0.0000 Min. :-4.000  1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.: 3.000  Median :0.0000 Median :0.0000 Median : 3.000  Mean :0.1537 Mean :0.4737 Mean : 3.276  3rd Qu.:0.0000 3rd Qu.:1.0000 3rd Qu.: 4.000  Max. :1.0000 Max. :1.0000 Max. : 6.000  > describe(hd)  vars n mean sd median trimmed mad min max range skew kurtosis  s\_p 1 950 79037.11 29169.78 70360.00 74918.59 20115.92 29864.00 222680.00 192816.00 1.71 4.05  inv 2 950 140.82 53.28 135.00 133.24 45.96 61.00 322.00 261.00 1.24 1.20  bath 3 950 1.95 0.87 2.00 1.89 0.74 1.00 20.00 19.00 9.38 188.84  ltsz 4 950 0.33 0.43 0.22 0.24 0.07 0.02 4.13 4.12 4.90 28.72  hssz 5 950 1055.49 296.61 1014.00 1034.73 222.39 216.00 3080.00 2864.00 1.36 5.36  bsemt 6 950 0.94 0.24 1.00 1.00 0.00 0.00 1.00 1.00 -3.59 10.87  a\_c 7 950 0.81 0.39 1.00 0.89 0.00 0.00 1.00 1.00 -1.61 0.59  f\_place 8 950 0.60 0.49 1.00 0.62 0.00 0.00 1.00 1.00 -0.39 -1.85  garsz\_a 9 950 0.96 0.20 1.00 1.00 0.00 0.00 2.00 2.00 -4.35 19.57  dw 10 950 0.81 0.40 1.00 0.88 0.00 0.00 1.00 1.00 -1.55 0.40  dr 11 950 0.76 0.43 1.00 0.82 0.00 0.00 1.00 1.00 -1.21 -0.54  fr 12 950 0.78 0.41 1.00 0.85 0.00 0.00 1.00 1.00 -1.36 -0.14  age5 13 950 0.15 0.36 0.00 0.07 0.00 0.00 1.00 1.00 1.92 1.68  stl10 14 950 0.47 0.50 0.00 0.47 0.00 0.00 1.00 1.00 0.11 -1.99  bdrms 15 950 3.28 0.72 3.00 3.25 0.00 -4.00 6.00 10.00 -0.60 11.49  se  s\_p 946.39  inv 1.73  bath 0.03  ltsz 0.01  hssz 9.62  bsemt 0.01  a\_c 0.01  f\_place 0.02  garsz\_a 0.01  dw 0.01  dr 0.01  fr 0.01  age5 0.01  stl10 0.02  bdrms 0.02  > hist(hd)  Error in hist.default(hd) : 'x' muss numerisch sein  > plot(hd)  Error in plot.new() : figure margins too large  > library(purrr)  Error in library(purrr) : es gibt kein Paket namens ‘purrr’  > install.packages(purrr)  Error in install.packages : object 'purrr' not found  > install.packages(purrr)  Error in install.packages : object 'purrr' not found  > library(purrr)  Error in library(purrr) : es gibt kein Paket namens ‘purrr’  > install.packages("purrr")  trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.5/purrr\_0.2.5.zip'  Content type 'application/zip' length 294282 bytes (287 KB)  downloaded 287 KB  package ‘purrr’ successfully unpacked and MD5 sums checked  The downloaded binary packages are in  C:\Users\VIP\AppData\Local\Temp\Rtmp00nVM9\downloaded\_packages  > install.packages("tidyr")  also installing the dependencies ‘bindr’, ‘bindrcpp’, ‘plogr’, ‘dplyr’, ‘tidyselect’  trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.5/bindr\_0.1.1.zip'  Content type 'application/zip' length 17704 bytes (17 KB)  downloaded 17 KB  trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.5/bindrcpp\_0.2.2.zip'  Content type 'application/zip' length 621596 bytes (607 KB)  downloaded 607 KB  trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.5/plogr\_0.2.0.zip'  Content type 'application/zip' length 18729 bytes (18 KB)  downloaded 18 KB  trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.5/dplyr\_0.7.6.zip'  Content type 'application/zip' length 3043056 bytes (2.9 MB)  downloaded 2.9 MB  trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.5/tidyselect\_0.2.4.zip'  Content type 'application/zip' length 621207 bytes (606 KB)  downloaded 606 KB  trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.5/tidyr\_0.8.1.zip'  Content type 'application/zip' length 943376 bytes (921 KB)  downloaded 921 KB  package ‘bindr’ successfully unpacked and MD5 sums checked  package ‘bindrcpp’ successfully unpacked and MD5 sums checked  package ‘plogr’ successfully unpacked and MD5 sums checked  package ‘dplyr’ successfully unpacked and MD5 sums checked  package ‘tidyselect’ successfully unpacked and MD5 sums checked  package ‘tidyr’ successfully unpacked and MD5 sums checked  The downloaded binary packages are in  C:\Users\VIP\AppData\Local\Temp\Rtmp00nVM9\downloaded\_packages  > install.packages("ggplot2")  trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.5/ggplot2\_3.0.0.zip'  Content type 'application/zip' length 3578992 bytes (3.4 MB)  downloaded 3.4 MB  package ‘ggplot2’ successfully unpacked and MD5 sums checked  The downloaded binary packages are in  C:\Users\VIP\AppData\Local\Temp\Rtmp00nVM9\downloaded\_packages  > library(ggplot2)  Attache Paket: ‘ggplot2’  The following objects are masked from ‘package:psych’:  %+%, alpha  > library(purrr)  > library(tidyr)  > attach(hd)  > hd <- housedata\_for\_final\_2018  > attach(hd)  The following objects are masked from hd (pos = 3):  a\_c, age5, bath, bdrms, bsemt, dr, dw, f\_place, fr, garsz\_a, hssz, inv, ltsz, s\_p, stl10  > plot(hssz)  > plot(ltsz)  > hd$bath[hd$bath < 0 ] <- NA  > hd$bdrms[hd$bdrms < 0 ] <- NA  plot(hssz, ltsz)  hd$ltsznew <- hd$ltsz\*43560  plot(hd$ltsznew, hd$hssz)  #I decided some are outliers or just wrong data and to replace them with median  hd$hssz[hd$hssz > 2500 ] <- median(hd$hssz)  hd$ltsznew[hd$ltsznew > 150000 ] <- median(hd$ltsznew)  hd$garsz\_a[hd$garsz\_a > 1 ] <- median(hd$garsz\_a)  hd$bdrms[hd$bdrms < 0 ] <- median(hd$bdrms)  hd$bath[hd$bath > 10 ] <- median(hd$bath)  plot(hd$ltsznew, hd$hssz)  hd$ratio <- hd$hssz/hd$ltsznew  summary(hd$ratio)  hd <- hd[hd$ratio<1, ]  #test if it worked  hd$ratio[hd$ratio>1]  # histogram  hist(s\_p)  par(mar = c(5, 4, 4, 5))  hist(hd$s\_p, nclass=50, type ="l", ylab = "# of Houses", main = "House Sale Prices and Distribution", xlab = "House Sale Price in $", col = "cyan4")  par(new = TRUE)  plot(density(hd$s\_p), type = "l", xaxt = "n", yaxt = "n", ylab = "", main="", xlab = "", col = "coral2", lty = 2, lwd = 3)  axis(side = 4)  mtext("House Sale Price Distribution", side = 4, line = 3)   |  | | --- | |  | |  | | |  | | --- | | > | | |
|  |
| |  | | --- | |  | |