Exercise 8: Correlation

> # Assignment: ASSIGNMENT 5

> # Name: Gunasekaran, Ragunath

> # Date: 2020-09-27

>

> ## Set the working directory to the root of your DSC 520 directory

> setwd("C:/Users/ragun/Documents/GitHub/dsc520-master/DSC520-new")

>

>

> ## Load the `data/r4ds/heights.csv` to

> heights\_df <- read.csv("data/r4ds/heights.csv")

>

> ## Using `cor()` compute correclation coefficients for

> library("ggplot2")

> ## height vs. earn

> cor(heights\_df$height, heights\_df$earn)

[1] 0.2418481

> ### age vs. earn

> cor(heights\_df$age, heights\_df$earn)

[1] 0.08100297

> ### ed vs. earn

> cor(heights\_df$ed, heights\_df$earn)

[1] 0.3399765

>

> ## Spurious correlation

> ## The following is data on US spending on science, space, and technology in millions of today's dollars

> ## and Suicides by hanging strangulation and suffocation for the years 1999 to 2009

> ## Compute the correlation between these variables

> tech\_spending <- c(18079, 18594, 19753, 20734, 20831, 23029, 23597, 23584, 25525, 27731, 29449)

> suicides <- c(5427, 5688, 6198, 6462, 6635, 7336, 7248, 7491, 8161, 8578, 9000)

> cor(tech\_spending, suicides)

[1] 0.9920817

>

> # Various Other methods

> "kendall"

[1] "kendall"

> cor(tech\_spending, suicides,method="kendall")

[1] 0.9272727

> "pearson"

[1] "pearson"

> cor(tech\_spending, suicides,method="pearson")

[1] 0.9920817

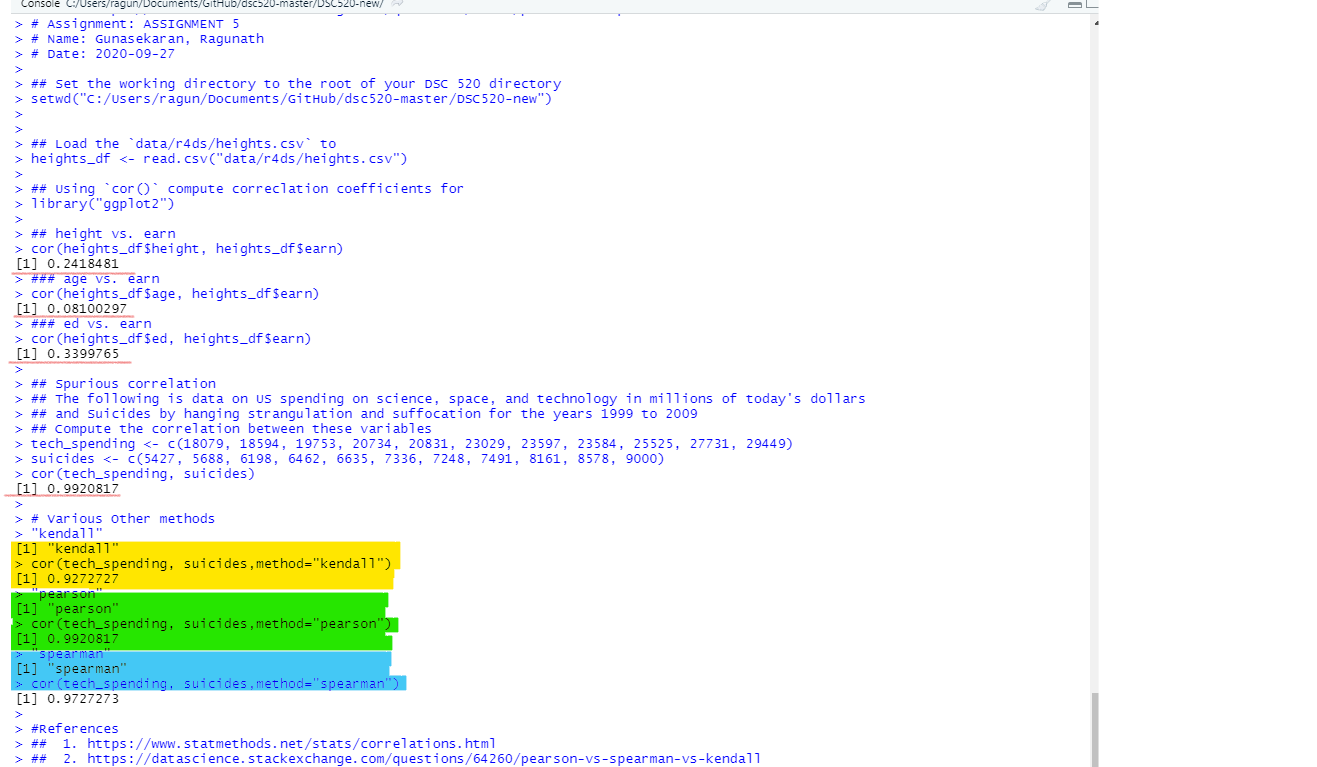
> "spearman"

[1] "spearman"

> cor(tech\_spending, suicides,method="spearman")

[1] 0.9727273

>



> #References

> ## 1. https://www.statmethods.net/stats/correlations.html

> ## 2. https://datascience.stackexchange.com/questions/64260/pearson-vs-spearman-vs-kendall