assignment_05_BrownLincoln.R

x

2021-07-25

```
# Assignment: ASSIGNMENT 5
# Name: Brown, Lincoln
# Date: 2021-07-24

## Set the working directory to the root of your DSC 520 directory
setwd("/media/x/disk/School/DSC520/git/dsc520/")

## Load the `data/r4ds/heights.csv` to
heights_df <- read.csv("data/r4ds/heights.csv")


## Using `cor()` compute correlation coefficients for
## 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
```

```
## Of the above three correlations, education vs earnings has the strongest correlation
## Spurious correlation
## The following is data on US spending on science, space, and technology in millions of today's dollar
## 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
```

```
## The spurious correlation has a strong correlation coeeficient of .99.
## It is important to remember that just because two things have correlation,
## the correlation does not imply causation.
```