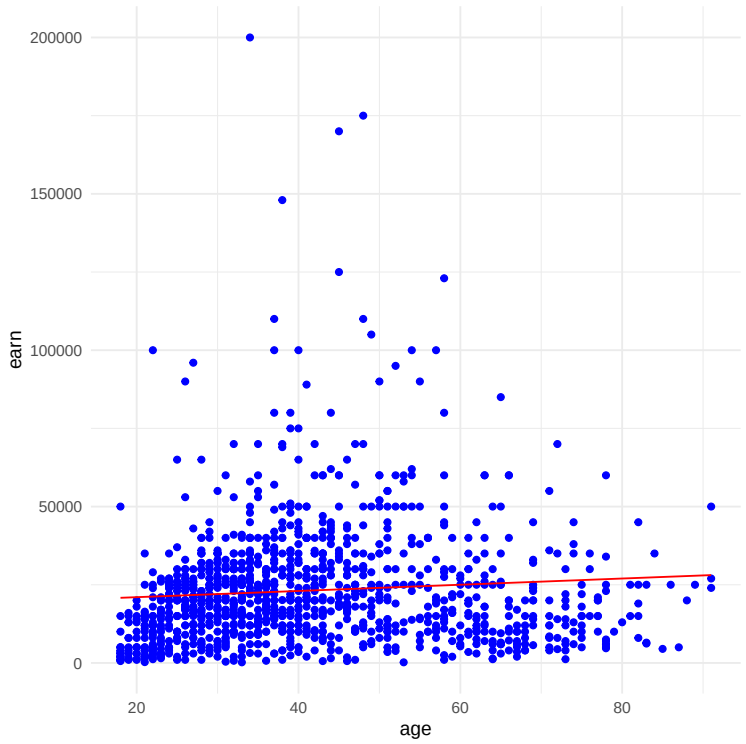
July 30, 2023

The results below are generated from an R script.

```
# Assignment: ASSIGNMENT 6
# Name: Reppeto, Brian
# Date: 2023-07-24

## Set the working directory to the root of your DSC 520 directory
setwd("~/DSC520/dsc520")

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

## Load the ggplot2 library
library(ggplot2)

## Fit a linear model using the `age` variable as the predictor and `earn` as the outcome
age_lm <-  lm(formula = earn ~ age, data = heights_df)

## View the summary of your model using `summary()`
summary(age_lm)

##
## Call:
## lm(formula = earn ~ age, data = heights_df)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -25098 -12622  -3667   6883 177579
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 19041.53    1571.26  12.119  < 2e-16 ***
## age            99.41      35.46   2.804  0.00514 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 19420 on 1190 degrees of freedom
## Multiple R-squared:  0.006561,Adjusted R-squared:  0.005727
## F-statistic:  7.86 on 1 and 1190 DF,  p-value: 0.005137

## Creating predictions using `predict()`
age_predict_df <-
  data.frame(earn = predict(age_lm, newdata = heights_df), age = heights_df$age)

## Plot the predictions against the original data
ggplot(data = heights_df, aes(y = earn, x = age)) +
```

```
  geom_point(color='blue') +
  geom_line(color='red',data = age_predict_df, aes(y=earn, x=age))
```




```
mean_earn <- mean(heights_df$earn)
## Corrected Sum of Squares Total
sst <- sum((mean_earn - heights_df$earn)^2)
## Corrected Sum of Squares for Model
ssm <- sum((mean_earn - age_predict_df$earn)^2)
## Residuals
residuals <- heights_df$earn - age_predict_df$earn
## Sum of Squares for Error
sse <- sum(residuals^2)
## R Squared R^2 = SSM\SST
r_squared <- ssm/sst

## Number of observations
n <- nrow(heights_df)
## Number of regression parameters
p <- 2
## Corrected Degrees of Freedom for Model (p-1)
dfm <- p-1
## Degrees of Freedom for Error (n-p)
dfe <- n-p
## Corrected Degrees of Freedom Total:   DFT = n - 1
dft <- n-1

## Mean of Squares for Model:   MSM = SSM / DFM
msm <- sse/dfm
## Mean of Squares for Error:   MSE = SSE / DFE
```

```r
mse <- sse/dfe
## Mean of Squares Total:   MST = SST / DFT
mst <- sst/dft
## F Statistic F = MSM/MSE
f_score <- msm/mse

## Adjusted R Squared R2 = 1 - (1 - R2)(n - 1) / (n - p)
adjusted_r_squared <- 1 - (1 - r_squared)*(n - 1) / (n - p)

## Calculate the p-value from the F distribution
p_value <- pf(f_score, dfm, dft, lower.tail=F)
```

The R session information (including the OS info, R version and all packages used):

```r
sessionInfo()

## R version 4.3.0 (2023-04-21)
## Platform: aarch64-apple-darwin20 (64-bit)
## Running under: macOS Ventura 13.4.1
##
## Matrix products: default
## BLAS:   /System/Library/Frameworks/Accelerate.framework/Versions/A/Frameworks/vecLib.framework/Versio
## LAPACK: /Library/Frameworks/R.framework/Versions/4.3-arm64/Resources/lib/libRlapack.dylib;  LAPACK ve
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## time zone: America/New_York
## tzcode source: internal
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
##  [1] knitr_1.43       lm.beta_1.7-2    corrplot_0.92    lmtest_0.9-40    zoo_1.8-12
##  [6] car_3.1-2        carData_3.0-5    conflicted_1.2.0 readxl_1.4.3     lubridate_1.9.2
## [11] forcats_1.0.0    stringr_1.5.0    dplyr_1.1.2      purrr_1.0.1      readr_2.1.4
## [16] tidyr_1.3.0      tibble_3.2.1     ggplot2_3.4.2    tidyverse_2.0.0
##
## loaded via a namespace (and not attached):
##  [1] utf8_1.2.3        generics_0.1.3    stringi_1.7.12    lattice_0.21-8
##  [5] hms_1.1.3         magrittr_2.0.3    evaluate_0.21     grid_4.3.0
##  [9] timechange_0.2.0  fastmap_1.1.1     cellranger_1.1.0  tinytex_0.45
## [13] fansi_1.0.4       scales_1.2.1      abind_1.4-5       cli_3.6.1
## [17] rlang_1.1.1       munsell_0.5.0     withr_2.5.0       cachem_1.0.8
## [21] tools_4.3.0       tzdb_0.4.0        memoise_2.0.1     colorspace_2.1-0
## [25] vctrs_0.6.3       R6_2.5.1          lifecycle_1.0.3   pkgconfig_2.0.3
## [29] pillar_1.9.0      gtable_0.3.3      glue_1.6.2        highr_0.10
## [33] xfun_0.39         tidyselect_1.2.0  rstudioapi_0.15.0 farver_2.1.1
## [37] xtable_1.8-4      labeling_0.4.2    compiler_4.3.0

Sys.time()

## [1] "2023-07-30 16:16:44 EDT"
```