

```

---
title: "Homework4"
output: html_document
date: "2024-06-22"
---

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
```

```{r}
library(tidyverse)
library(moderndiver)
library(skimr)
library(gapminder)
```

### LC5.1
Selecting a subset of variables
```{r}
evals_ch5 <- evals %>%
 select(ID, score, bty_avg, age)
```

Computing summary statistics for age and teaching score
```{r}
evals_ch5 %>%
 summarize(mean_age = mean(age), mean_score = mean(score),
 median_age = median(age), median_score = median(score))

evals_ch5 %>%
 select(age, score) %>%
 skim()
```

Correlation coefficient of age and teaching score
```{r}
evals_ch5 %>%
 get_correlation(formula = age ~ score)
```

Data visualization: Scatterplot of the relationship between age and teaching scores

The negative slope of the regression line demonstrates a negative relationship, as the age decreases they also receive lower evaluations.
```{r}
ggplot(evals_ch5, aes(x = age, y = score)) +
 geom_point() +
 labs(x = "age",
 y = "score",
 title = "Scatterplot of relationship between age and teaching scores") +
 geom_smooth(method = "lm", se = FALSE)
```

### LC5.2
Fit regression model
```{r}
age_model <- lm(score ~ age, data = evals_ch5)
```

Get regression table

```

For every increase of 1 unit of age there is an associated increase of, on average, -0.006 units of score. This is similar to the correlation coefficient of age and score (-0.107) in that they are both negative, but the numbers are different.

```
```{r}
get_regression_table(age_model)
```
```

###LC5.3

Dataframe of the residuals of the model for age and teaching scores

```
```{r}
regression_points2 <- get_regression_points(age_model)
regression_points2
```
```

###LC5.4

###LC5.5

###LC5.6

###LC5.7

###LC5.8