



# **LAB 03**

PH142 Fall 2025

# Announcements

- **Lab03:** due 9/12 at 11:59pm
- **Quiz02:** due 9/12 at 11:59pm



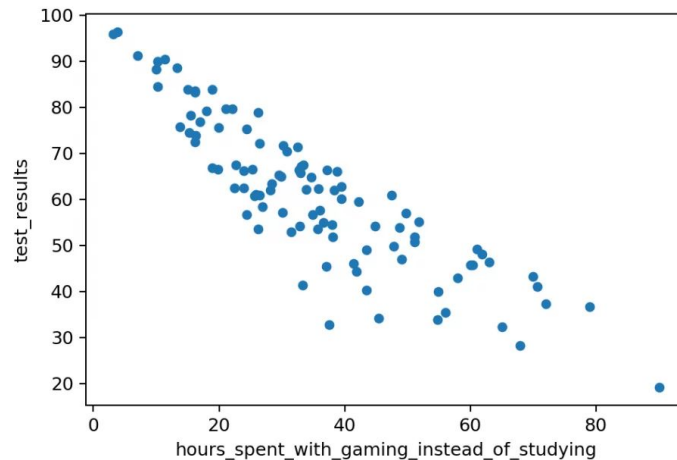
# Week 3 Lecture Review

## Scatterplots

1. Direction - Is it positive or negative?
2. Form - Is it linear or nonlinear?
3. Strength - Is it strong or weak?
4. Outliers - Are there any obvious ones?

How would you interpret this scatterplot?

Can you think of any confounding factors?



# Week 3 Lecture Review

## Scatterplots

`geom_point` is used for scatter plots:

Ex. `name of plot <- ggplot(data = dataset, aes(x=var, y=var)) +  
 geom_point(na.rm = TRUE) +  
 theme_minimal(base_size = 15) +  
 labs (x= "", y= "", title= "")`

- To color the points by a variable include `col=variable`
- To create separate plots for combinations of levels of 2 vars i.e. (gender) use `facet_wrap` ex. `facet_wrap ( ~ gender)`

# Week 3 Lecture Review

## Correlation

- Correlation measures the **strength** and **direction** of a linear relationship between two quantitative variables
- Also written as **r**
- Takes values between -1 to 1 inclusively

# Week 3 Lecture Review

## Intro to Linear Regression

- **Regression:** Straight line fitted to data to minimize distance between the data and fitted line
  - “Line of best fit” =  $a + bx$
  - $a$  = **intercept** (Predictive Value of  $y$  when  $x=0$ )
  - $b$  = **slope**  $r^*(s_y/s_x)$
- Interpretation
  - Intercept : the value of the outcome when  $x = 0$
  - Slope: For a one-unit change in  $x$ , the outcome changes by [number] [units]

# Week 3 Lecture Review

## Intro to Linear Regression

- `lm()` is the function for a linear model
  - `lm(formula = y ~ x, data = your_dataset)`
  - Add regression line to a scatterplot using `geom_abline()`
  - `glance(data_lm)` will give r squared
- **Interpretation of `lm()`:** A one unit change in **X** is associated with a \_\_\_\_ increase/decrease of **Y**.
- **Interpretation of r-squared:** the fraction of the variation in the values of y that is explained by the line of best fit (the regression of y on x)
- Find **correlation** in data using `summarize(corr_variable = cor(var1, var2))`



---

# **LAB 03 Walkthrough**



# Lab Submission

- Follow the directions on the LAB02 file
- Submit using the **Terminal Tab** (next to the console in the bottom left pane)
  - Copy and paste the given line into the terminal
  - Follow prompts (NOTE: the terminal will **not** show your password being typed out!)
- **CHECK IN GRADESCOPE THAT ALL YOUR TESTS PASSED**