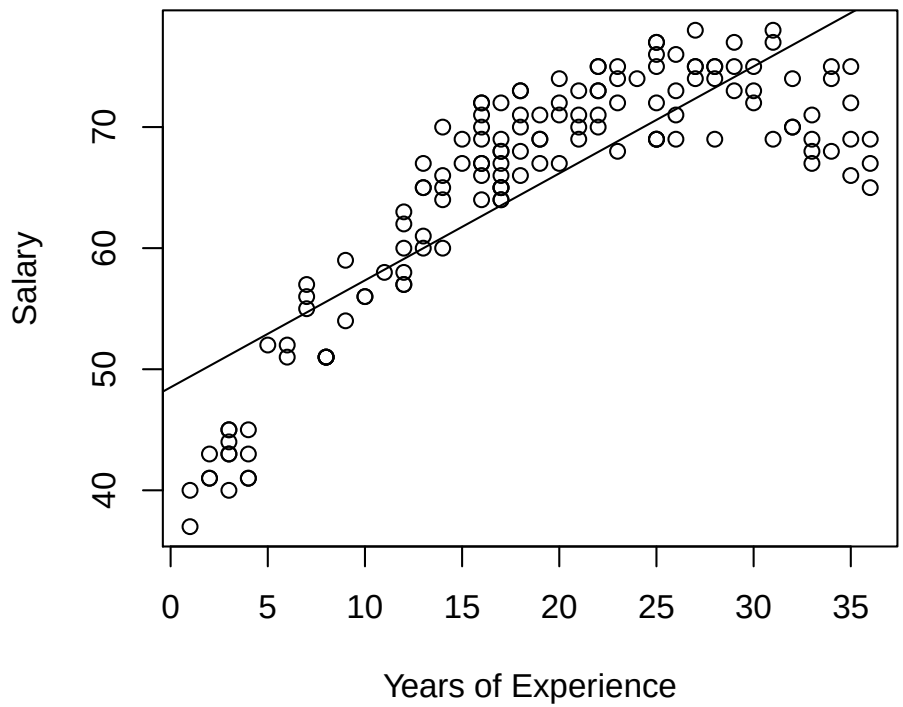
Lecture 14:
Quadratic Regression Model
STAT 310, Spring 2023

Salary Data Set

- For this example we consider a salary data set collected from a random sample $n = 143$ employees.
- We want to use this data to estimate a regression model that predicts salary (y), which is in thousands of dollars, from the number of years of work experience (x).

Fitting a straight line obviously does not capture the trend in the data.

```
> profsalary = read.csv("https://ericwfox.github.io/data/profsalary.csv")
> lm1 = lm(Salary ~ Experience, data = profsalary)
> plot(Salary ~ Experience, data = profsalary,
       xlab = "Years of Experience", ylab = "Salary")
> abline(lm1)
```



Quadratic Regression Model

Since a quadratic relationship is evident in the scatter plot, we consider the following model:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \epsilon$$

where y = salary, x = years of experience, and ϵ is the random error term.

```
> lm2 = lm(Salary ~ Experience + I(Experience^2), data = profsalary)
> summary(lm2)

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)     34.720498   0.828724   41.90   <2e-16 ***
Experience       2.872275   0.095697   30.01   <2e-16 ***
I(Experience^2) -0.053316   0.002477  -21.53   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.817 on 140 degrees of freedom
Multiple R-squared:  0.9247,Adjusted R-squared:  0.9236
F-statistic: 859.3 on 2 and 140 DF,  p-value: < 2.2e-16
```

Estimated quadratic regression model:

$$\hat{y} = 34.720 + 2.872x - 0.053x^2$$

Prediction when $x = 10$:

$$\hat{y} = 34.720 + 2.872(10) - 0.053(10^2) = 58.14$$

So 58,140 dollars is the predicted salary for an employee with 10 year of work experience.

The $R^2 = 0.9247$, so about 92% of the variability salary can be explained by the quadratic regression model with years of experience as a predictor.

The modern graphics package ggplot2 provides a convenient way to visualize the estimated quadratic regression curve.

```
library(ggplot2)
ggplot(profsalary, aes(x = Experience, y = Salary)) +
  geom_point() +
  geom_smooth(method = "lm", formula = y ~ poly(x, 2)) +
  xlab("Years of Experience") + ylab("Salary")
```