# Interpreting Regressions

Prof. Bisbee

Vanderbilt University

Lecture Date: 2023/11/28

Slides Updated: 2023-11-25

# Agenda

1. Derivatives
2. Continuous predictors
3. Categorical predictors
4. Interaction terms

# Derivatives

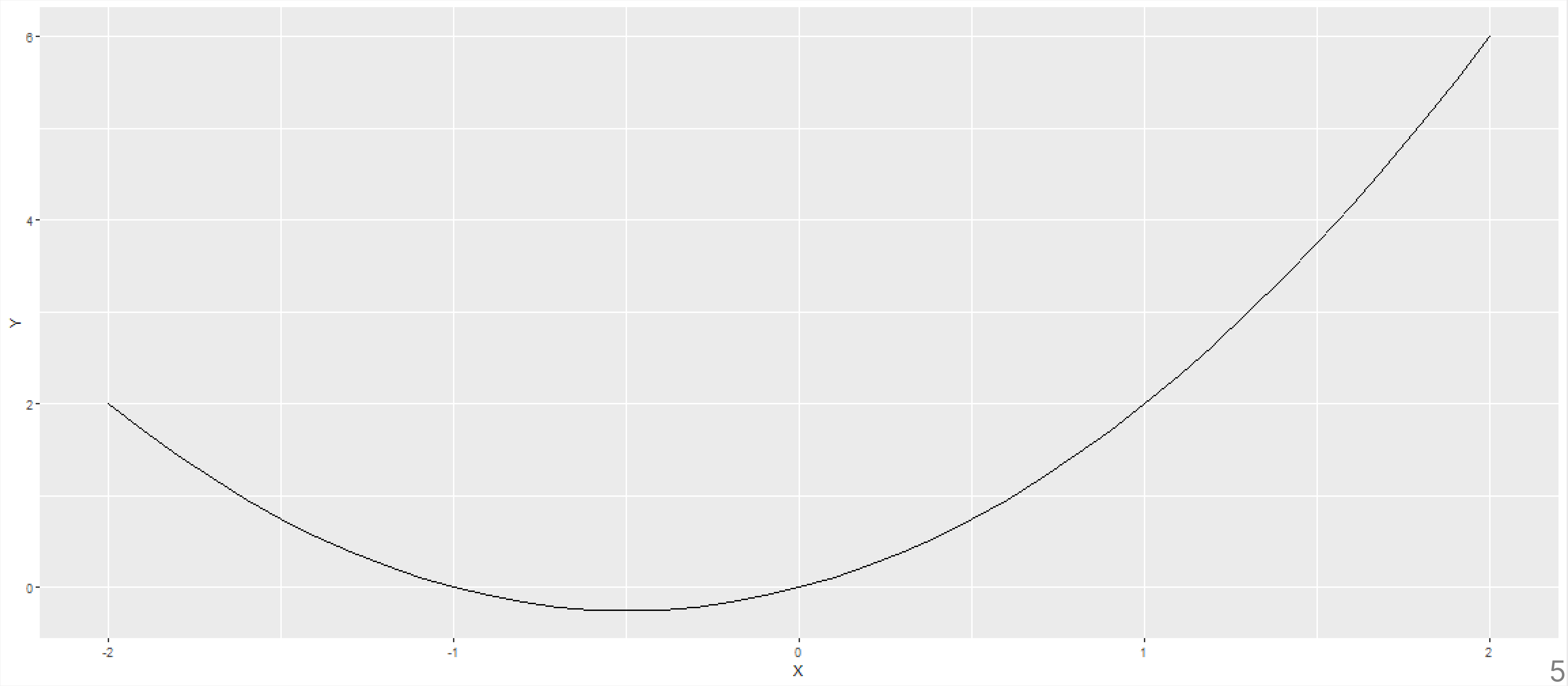
- Two standard ways to indicate a derivative:
  1. $\frac{d}{dx}$ stands for "the derivative with respect to x" using Leibniz notation
  2. $f'(x)$ stands for "the derivative of $f$ with respect to x" using Lagrange notation
- Regardless of notation, the underlying logic here is an attempt to fine the slope of a line
  - In a linear function of the form $y = mx + b$, we can find $m$ by plugging in the change in $y$ over the change in $x$, or $\frac{y_2 - y_1}{x_2 - x_1}$
  - In a nonlinear function denoted $y = f(x)$, the line might be curved. Denote $\Delta x = x_2 - x_1$ and $y_1 = f(x)$ and $y_2 = f(x + \Delta x)$ and plug in to yield $m = \frac{f(x+\Delta x) - f(x)}{\Delta x}$

# Derivatives

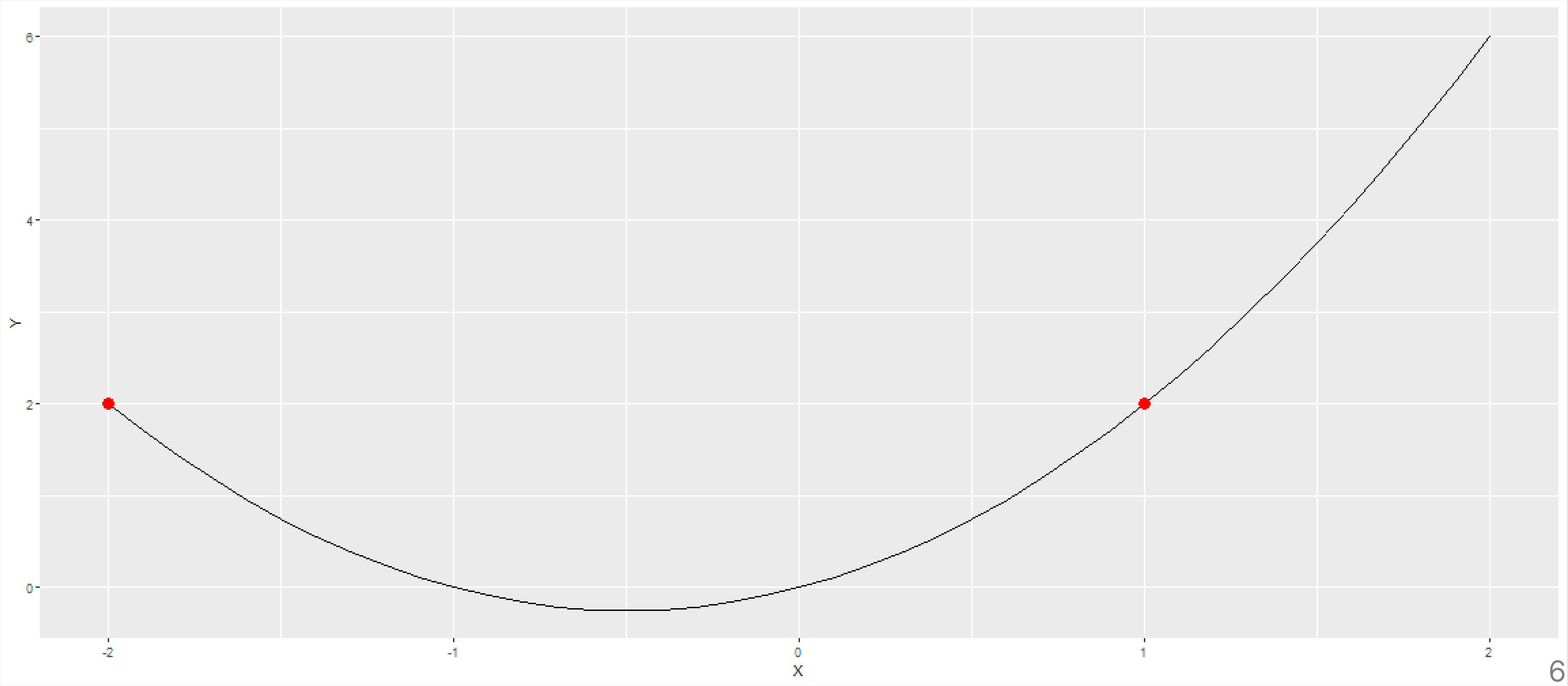
- Of course, if $\Delta x$ is large (meaning $x_2$ is far from $x_1$), we are just calculating the **secant line**, the line that connects the two points
  - Substantively, this is just the average rate of change between $y$ and $x$
  - Depending on the curve, this might be a *very* poor approximation of the slope
- As such, we want to calculate by taking the limit as $\Delta x \to 0$, which gives us the instantaneous rate of change at a given value of $x$, where the slope is now equal to the **tangent line** to the curve. Denote the derivative $f'(x)$

$$f'(x) = \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}$$

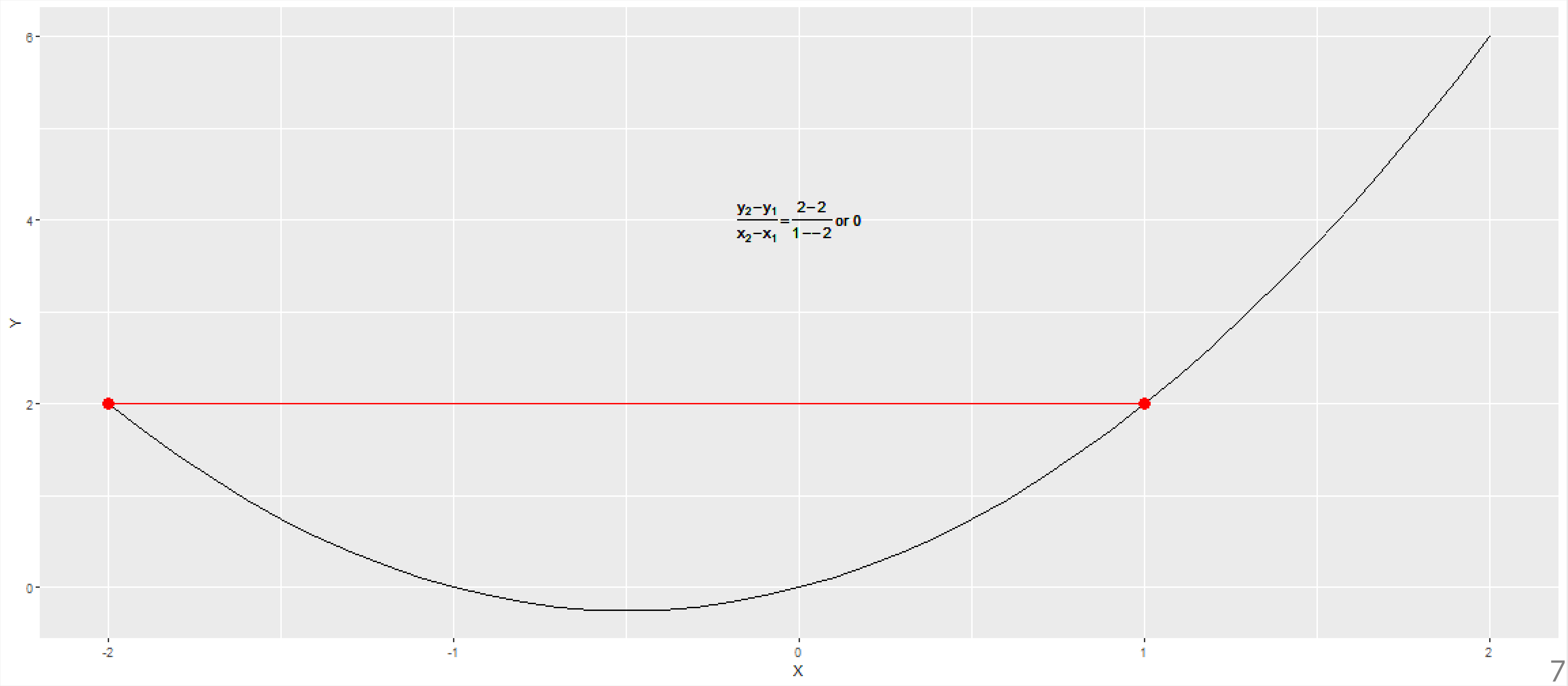
# Derivatives



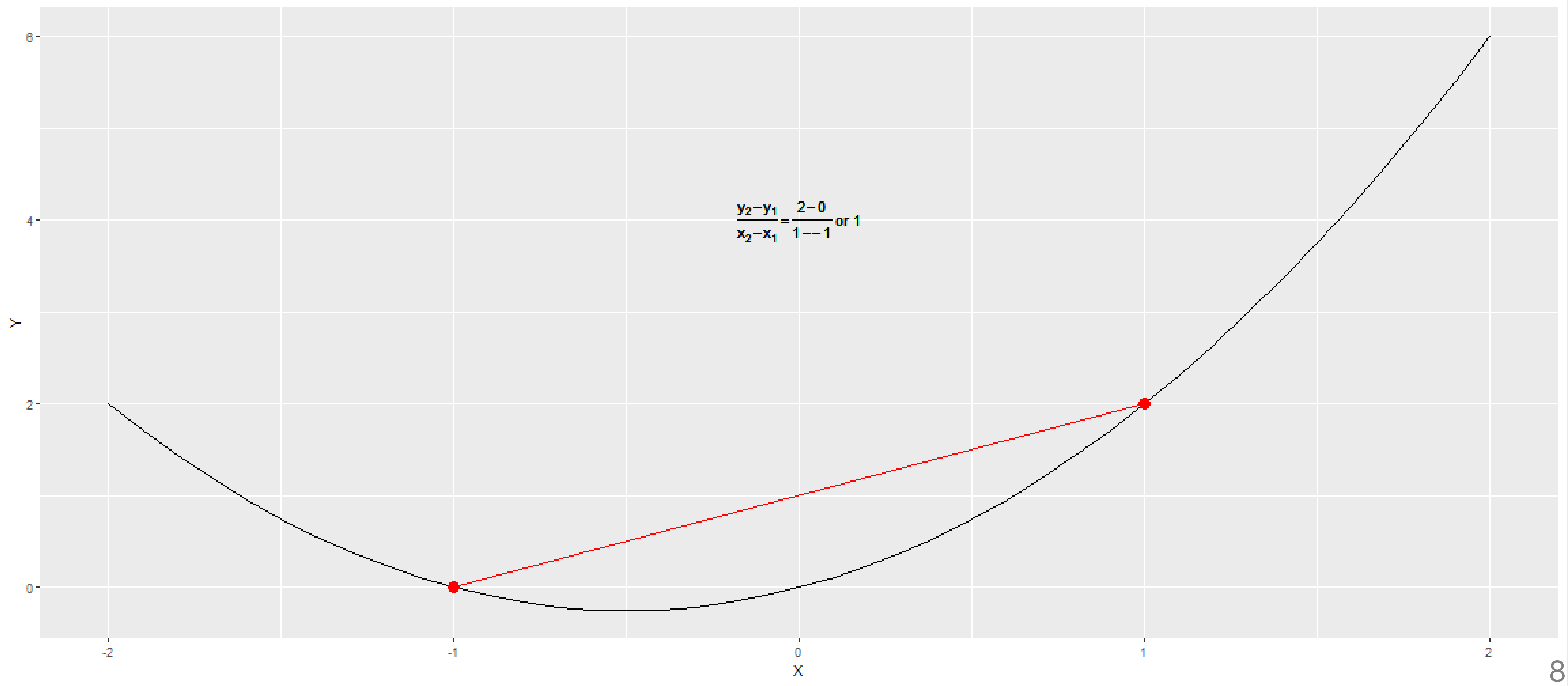
# Derivatives



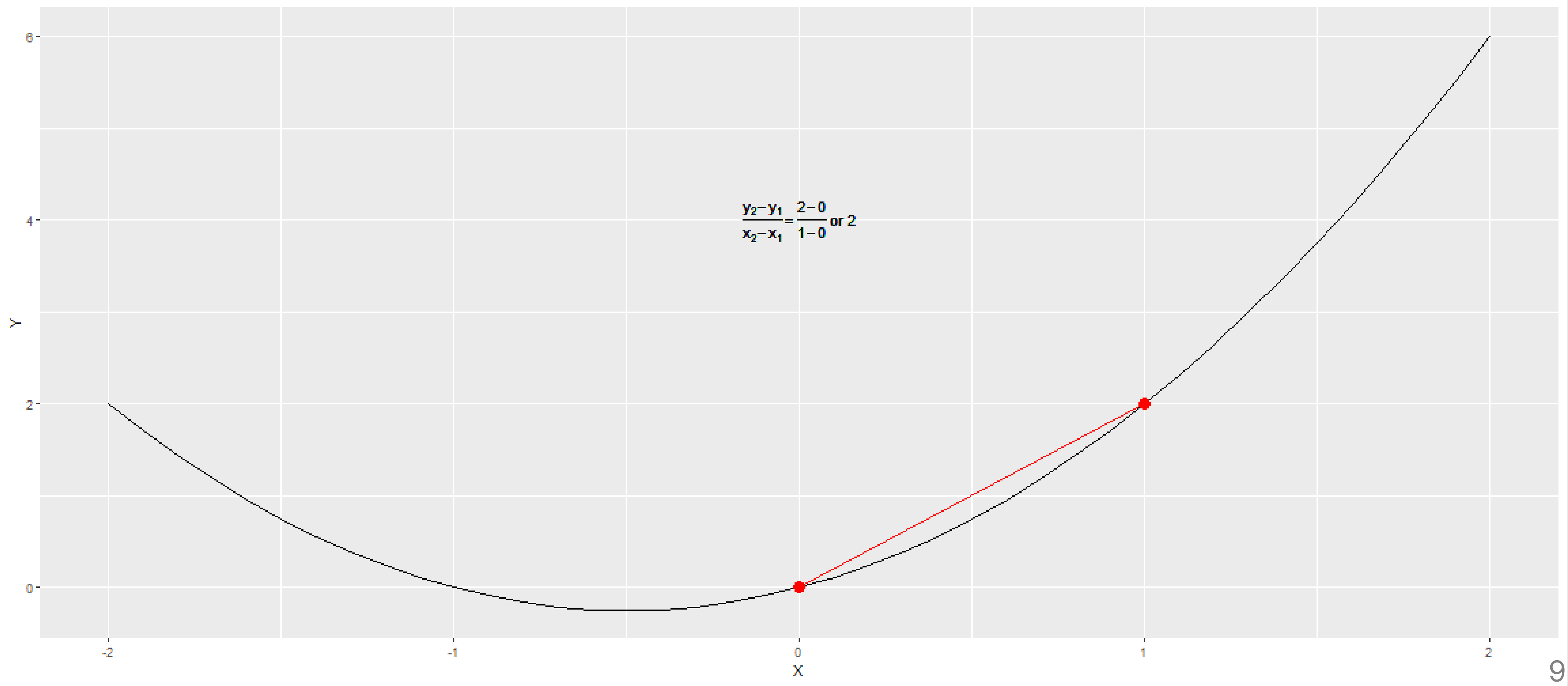
# Derivatives



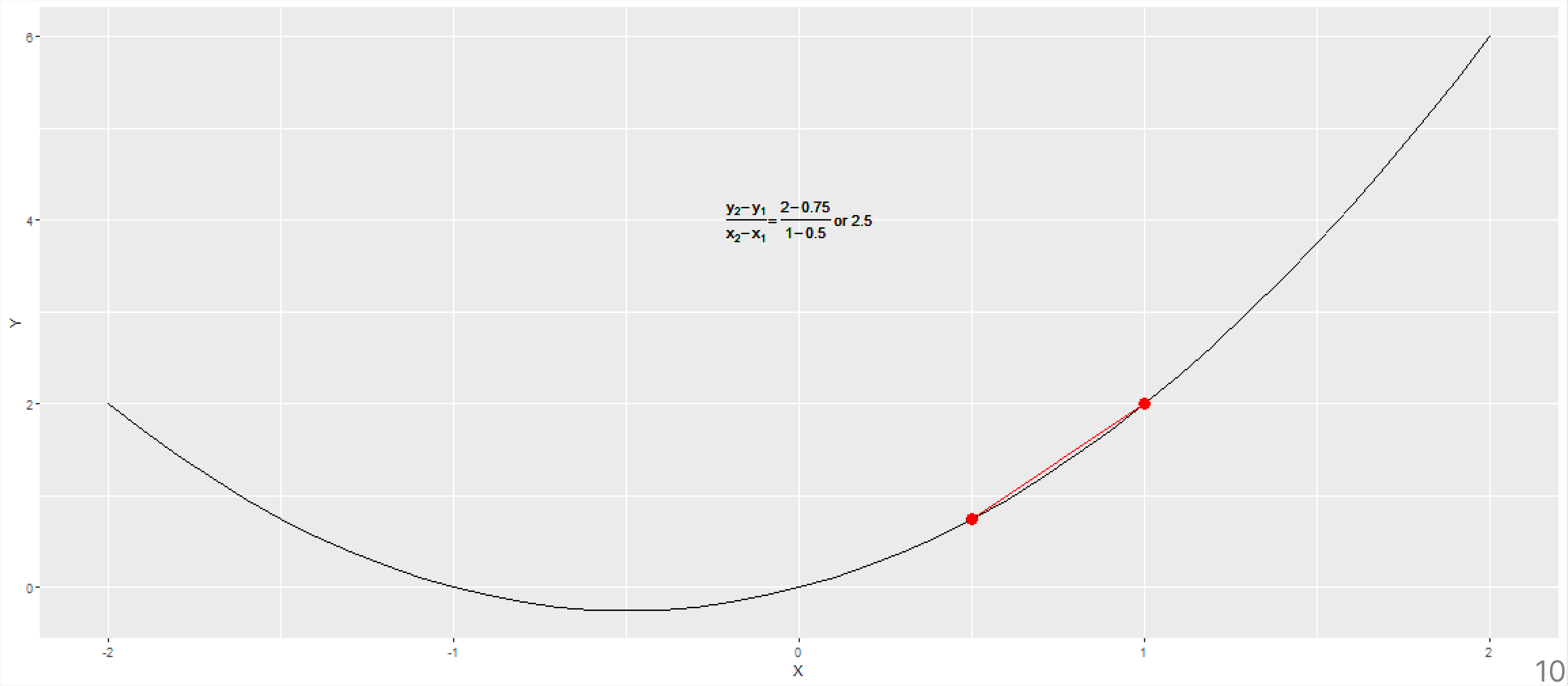
# Derivatives



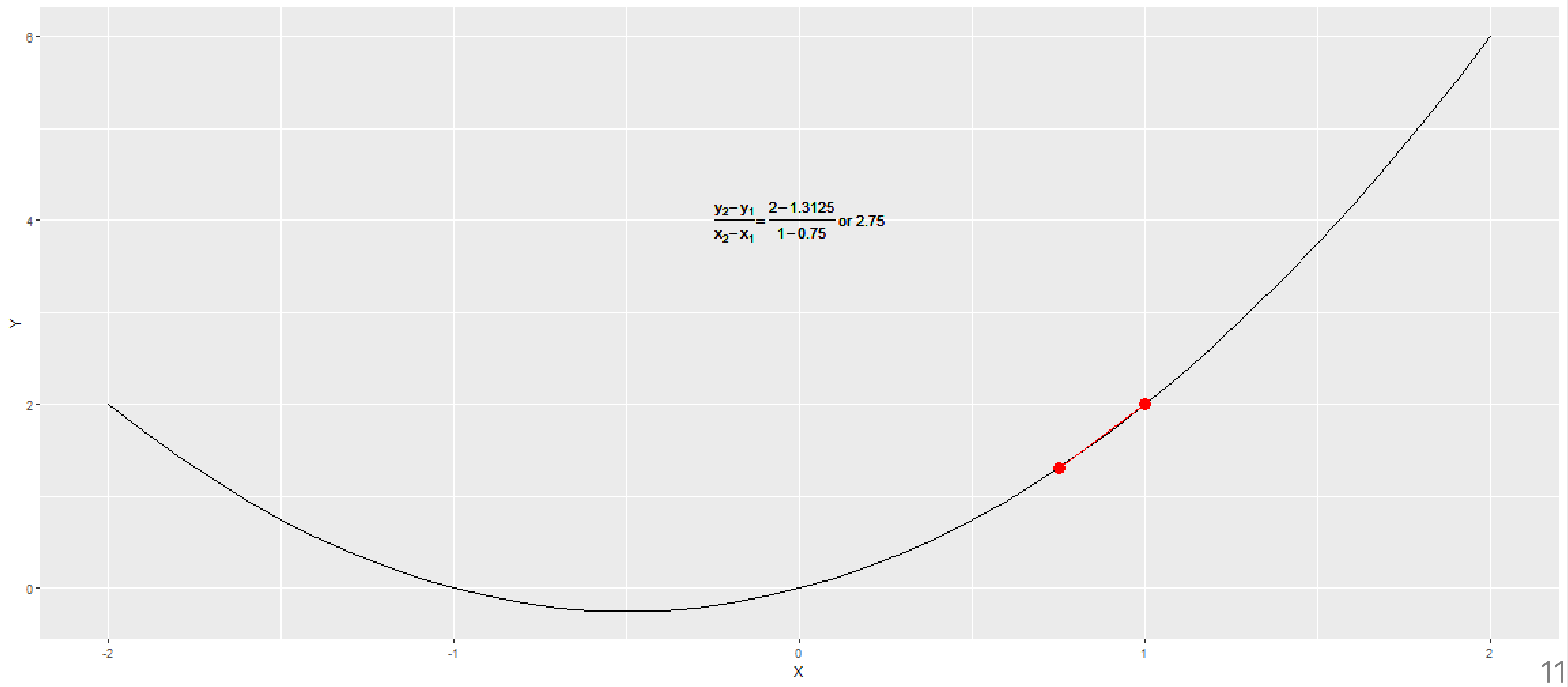
# Derivatives



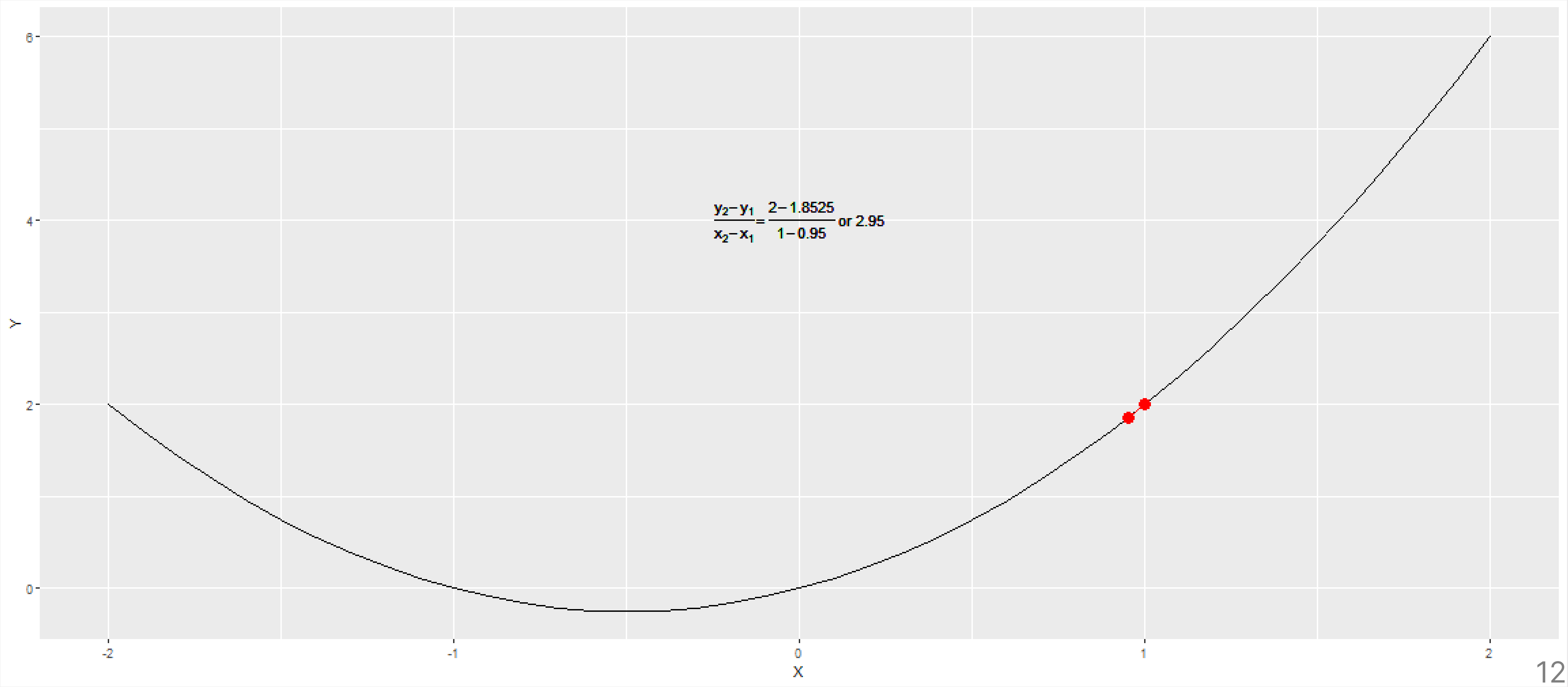
# Derivatives



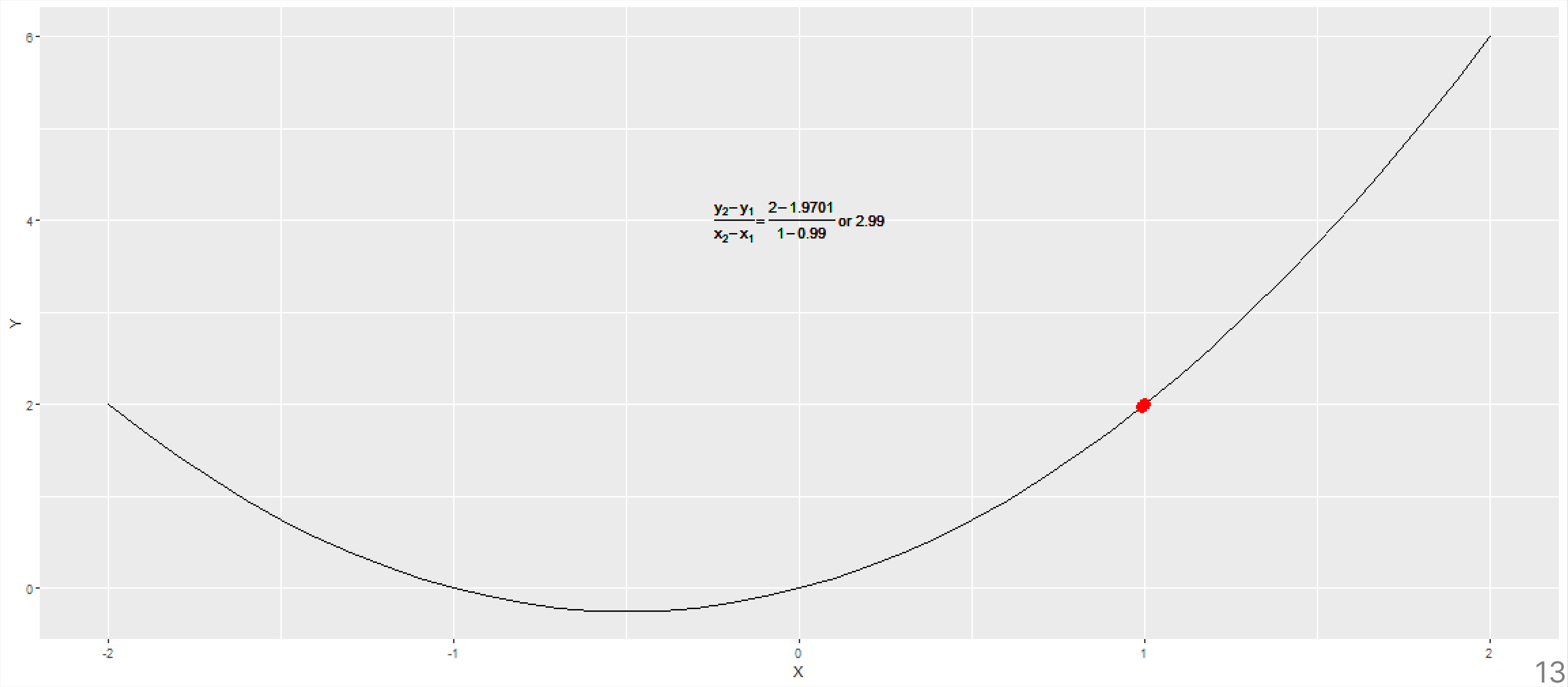
# Derivatives



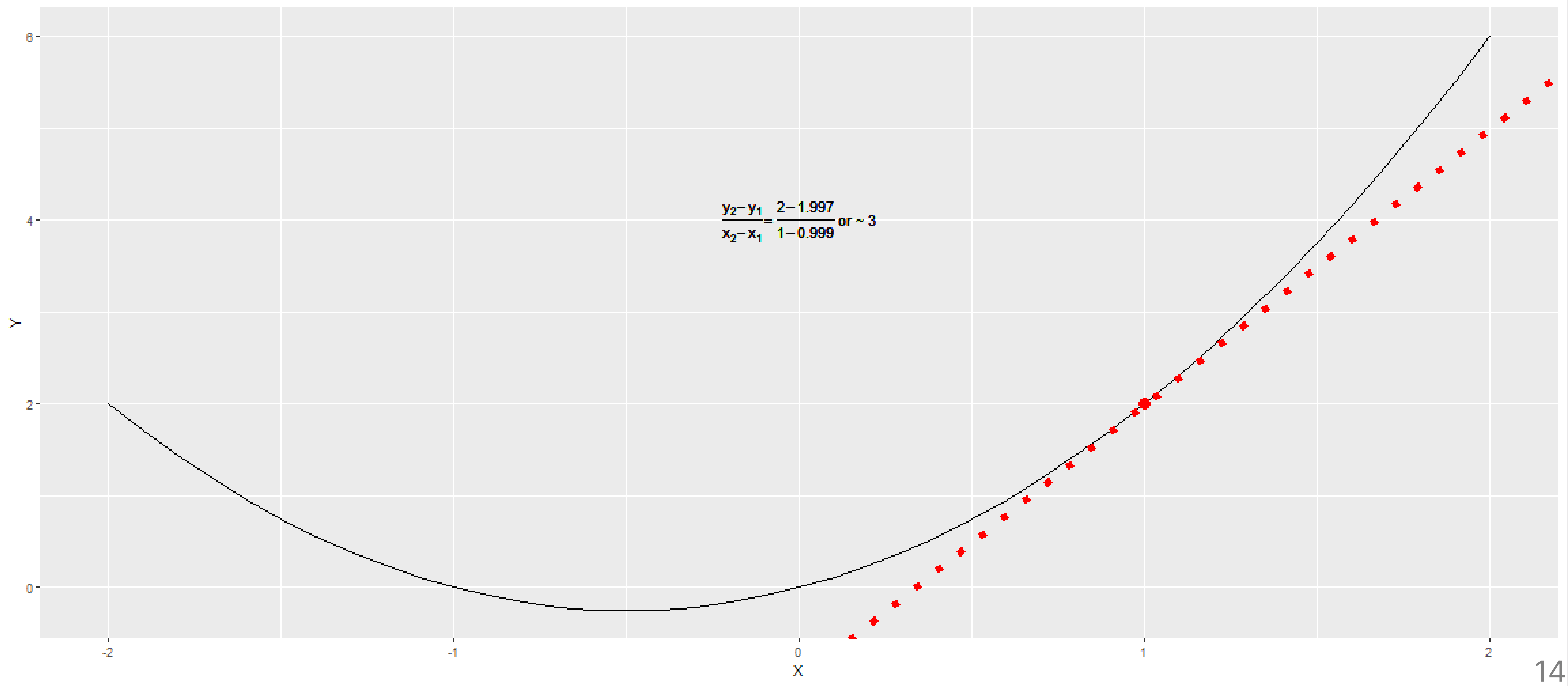
# Derivatives



# Derivatives



# Derivatives



# Derivates

$$f'(x) = \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}$$

- Given this equation, we *could* expand the numerator, cancel some things out, and then evaluate the limit
- But we often just rely on four **rules** that are easy to remember

# Derivatives: Rules

1. Derivative of a variable to a power:

$$\frac{d}{dx} ax^k = akx^{k-1}$$

2. Derivative of a sum of terms to a power:

$$\frac{d}{dx} \sum_{i=1}^{n} x_i^k = \sum_{i=1}^{n} \frac{d}{dx} x_i^k = \sum_{i=1}^{n} k x_i^{k-1}$$

- (the derivative of a sum is the sum of the derivatives).

3. Chain rule:

$$\frac{d}{dx} [f(g(x))] = f'(g(x))g'(x)$$

# Derivatives: Rules

4. Partial derivatives of a function of multiple variables $f(x, y)$

$$\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}$$

treat the other variables as constants.

- e.g. suppose $f(x, y) = 2x - y + 6$. Then

$$\begin{aligned} \frac{\partial f}{\partial x} &= \quad 2 \\ \frac{\partial f}{\partial y} &= \quad -1 \end{aligned}$$

- This winds up being the most important for interpreting regressions, as you'll see!

# Continuous Predictors

- Given a theoretical regression $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$, what is the relationship between $y$ and $x_1$?
- To answer, we can typically just reply on the **Power Rule** for calculating derivatives: $\frac{\partial x^n}{\partial x} = n * x^{n-1}$.
  - In our setting, we take the partial derivative of $y$ with respect to $x_1$:

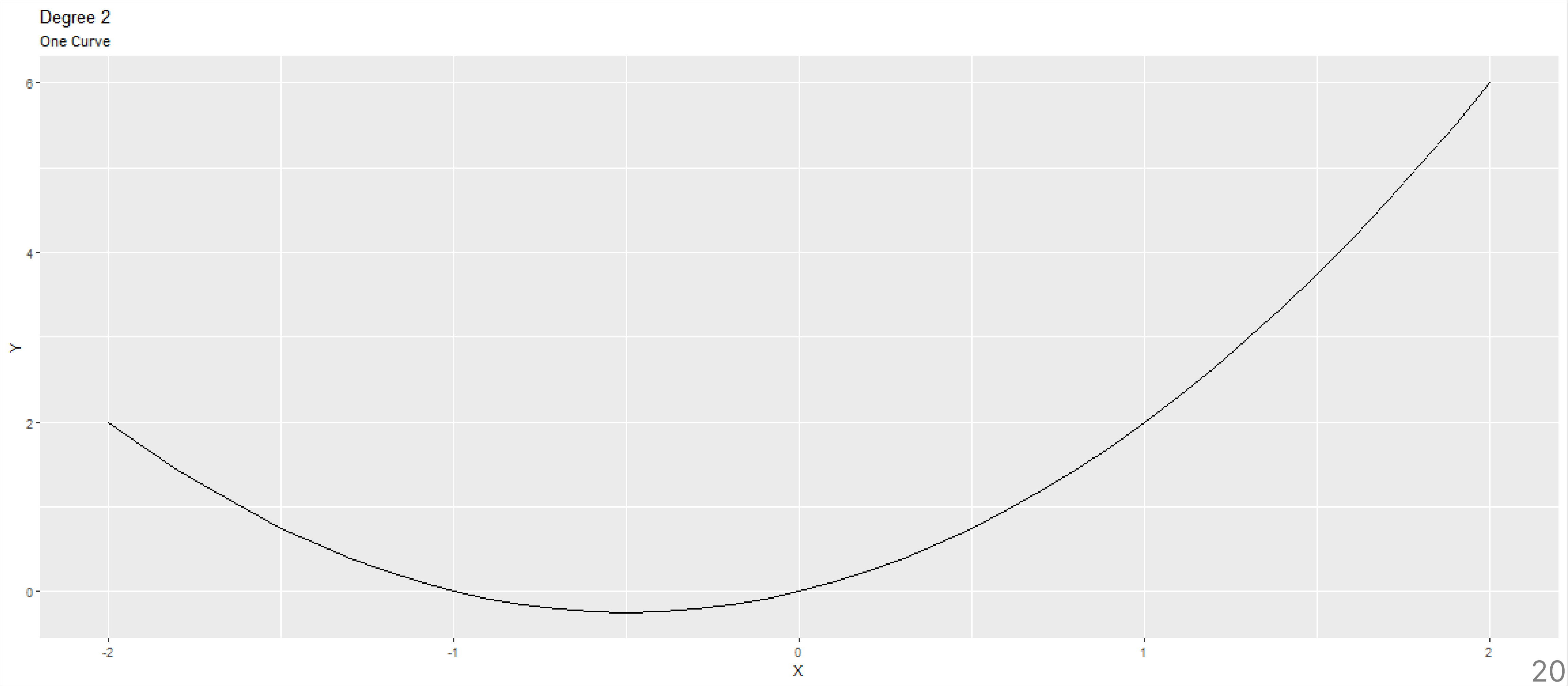
$$
\begin{aligned}
\frac{\partial y}{\partial x_1} &= \frac{\partial(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + u)}{\partial x_1} \\
&= 0 + \beta_1 * 1 * x_1^0 + 0 + 0 \\
&= \beta_1
\end{aligned}
$$

- Substantively, we say that "a one unit change in $x_1$ corresponds to a $\beta_1$ unit change in $y$"

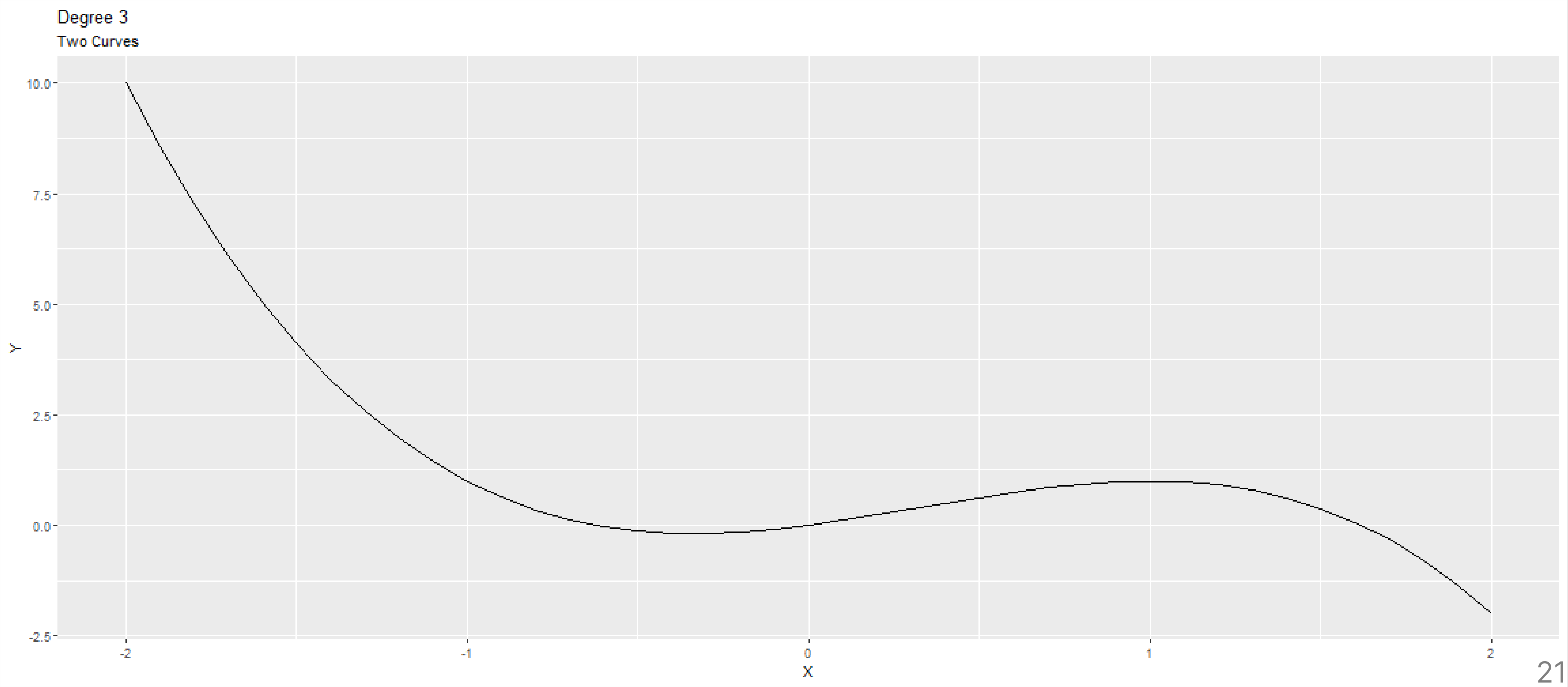
# Continuous Predictors: Polynomials

- Recall assumption 1?
  - True model is **linear**
- In some cases, we might have good reason to believe that the true relationship between $y$ and $x$ is **non-linear** (i.e., age and annual wage income)
  - In this setting, we can **transform** $x$ to make our model "linear in the parameters"
  - A very typical transformation is to add **polynomial terms** of $x$ as additional predictors
  - A polynomial regression model of degree $r$ is written: $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_r x^r + u$
  - **NB:** Each additional degree allows for $r - 1$ additional curves

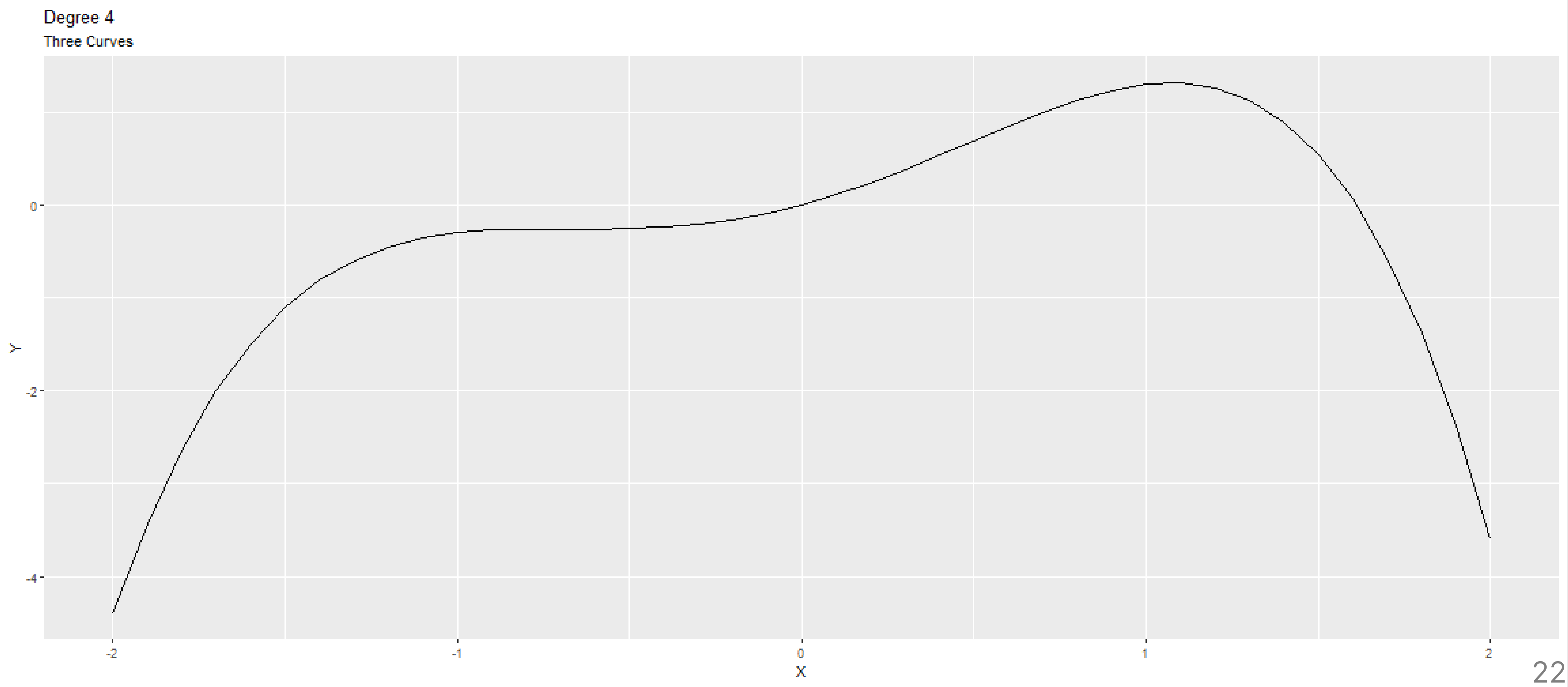
# Continuous Predictors: Polynomials



# Continuous Predictors: Polynomials



# Continuous Predictors: Polynomials



# Continuous Predictors: Polynomials

- One of the most commonly occurring polynomials used in social science is the **quadradic** model:

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

- How do we interpret this slope? Take the partial derivative with respect to $x$ as always!

$$\begin{aligned}
\frac{\partial y}{\partial x} &= \frac{\partial(\beta_0 + \beta_1 x + \beta_2 x^2 + u)}{\partial x} \\
&= 0 + \beta_1 * 1 * x^0 + \beta_2 * 2 * x^1 + 0 \\
&= \beta_1 + 2\beta_2 x
\end{aligned}$$

# Continuous Predictors: Polynomials

- How do we make inferential statements about $\frac{\partial y}{\partial x} = \beta_1 + 2\beta_2 x$? Answer one of two questions:

1. Is the quadratic model a **better fit** to the data than the linear model?

- To answer, just interpret the $p$-value and $t$-test associated with $\hat{\beta}_2$

2. At what values of $x$ is the relationship between $x$ and $y$ **statistically distinct**, and are these values **substantively meaningful**?

- This is often the much more important question. When $x$ is a predictor of interest (as opposed to simply being a control variable) and we fit it with a quadratic, it is usually because we want to make the claim that the effect of $x$ on $y$ is significantly higher (or lower) somewhere in the middle of the range of $x$ than at lower and higher values of $x$.

- This is distinct from (1) because either (a) the quadratic might provide a better fit, but the effect of $x$ doesn't attain an extremum in the empirical range of $x$; or (b) the quadratic might provide a better fit, and the effect of $x$ attains an extremum in the empirical range of $x$, but $\frac{\partial y}{\partial x}$ at this extremum is not significantly different from $\frac{\partial y}{\partial x}$ at other meaningful values of $x$

# Categorical Predictors

- Sometimes we are interested in a dichotomous predictor, such as whether an individual has a PhD or whether a village was given mosquito nets
    - We can represent this predictor as either 0 (meaning no PhD, or no mosquito nets) or a 1 (meaning a PhD or mosquito nets), often referred to as a **dummy** variable
    - Then we just add it to our regression as normal
- Consider this example where we predict income as a function of years in the labor market and whether the individual $i$ has a PhD (our dichotomous predictor)

$$Income_i = \beta_0 + \beta_1 Labor_i + \beta_2 PhD_i + u_i$$

- What is the partial derivative of income with respect to the PhD dummy variable?

# Categorical Predictors

- Here is an example using dummy data (see raw code for details)

```
m1 <- lm(inc ~ labor + phd, data = data)
summary(m1)
```

```
##
## Call:
## lm(formula = inc ~ labor + phd, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13.5052  -3.5249  -0.2899   3.1433  12.7515
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 13.64262    0.94620   14.42   <2e-16 ***
## labor        6.11087    0.04866  125.59   <2e-16 ***
## phd         40.08627    0.83957   47.75   <2e-16 ***
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

# Categorical Predictors

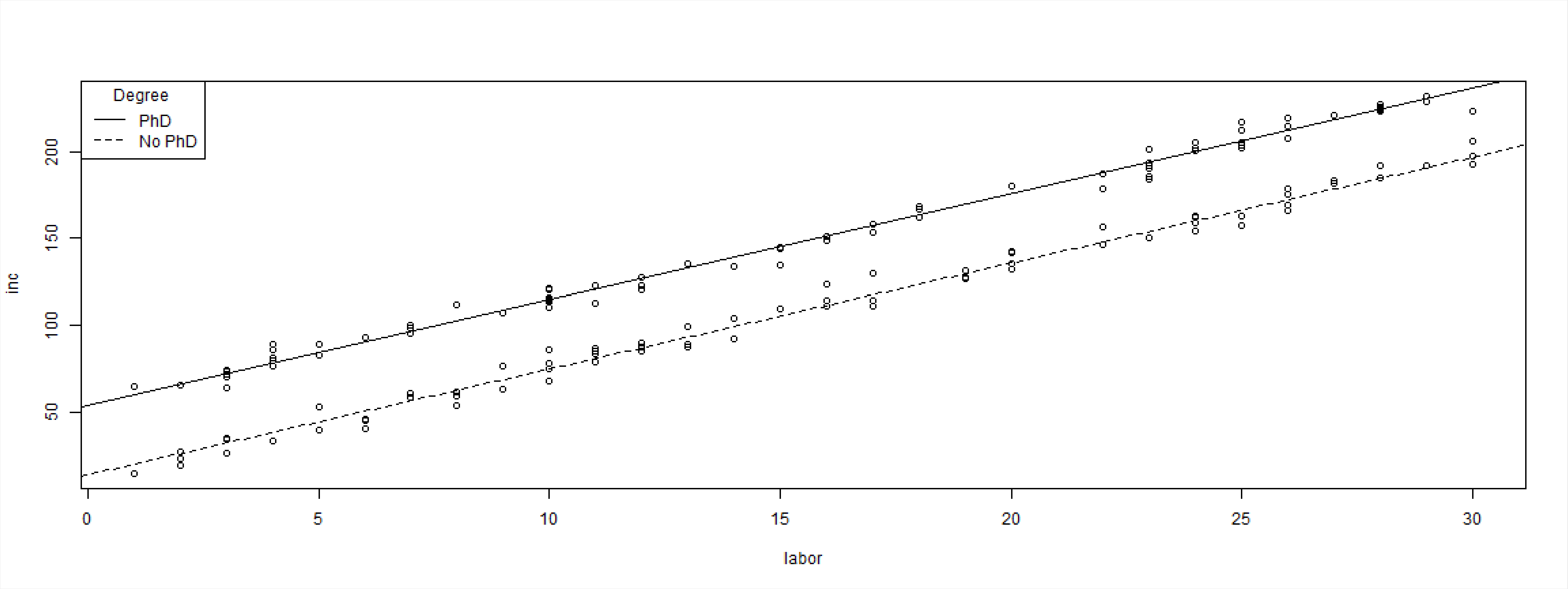
- How do we interpret this result? (NB: outcome is measured in thousands of dollars)

```
coef(m1)
```

```
## (Intercept)       labor         phd
##    13.64262     6.11087    40.08627
```

- A one unit change in $x$ corresponds to a $\beta$ unit change in $y$
- A one unit change in *PhD* corresponds to a 40.97 unit change in *Income*
  - And a one unit change in *PhD* just means going from 0 (no PhD) to 1 (has a PhD)

# Categorical Predictors



# Categorical Predictors

- What if our categorical measure is not dichotomous? (I.e., college degree or less, masters degree, PhD?)
- Any multi-level categorical variable can be "dummied out" by creating dichotomous versions of it's levels
  - NB: we **can't** include all levels though...we always need to drop one. Why?
  - Assumption 3: no perfect multicollinearity in $\mathbf{X}$!
  - If we include dummies of every level, then we can perfectly predict *PhD* with *College degree or less* and *MA degree*
  - $\mathbf{X}^\top\mathbf{X}$ is therefore not invertible, meaning that $(\mathbf{X}^\top\mathbf{X})^{-1}$ doesn't exist
- Thus, for a categorical predictor with $k$ levels, we only add $k-1$ dummies for each of its levels
  - In this example, we don't include the dummy for "College degree or less"

$$Income_i = \beta_0 + \beta_1 Labor_i + \beta_2 PhD_i + \beta_3 MA_i + u_i$$

# Categorical Predictors

```
m2 <- lm(inc ~ labor + phd + ma, data = data2)

summary(m2)
```

```
##
## Call:
## lm(formula = inc ~ labor + phd + ma, data = data2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.8649  -3.3441   0.0557   3.4478  10.2087
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 14.17746    1.09259   12.98   <2e-16 ***
## labor        6.03887    0.04708  128.26   <2e-16 ***
## phd         39.04398    1.08605   35.95   <2e-16 ***
## ma          11.96320    1.18987   10.05   <2e-16 ***
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

# Categorical Predictors

- How to interpret?

```
coef(m2)
```

```
## (Intercept)       labor         phd          ma
##   14.177459    6.038865   39.043975   11.963200
```

- What is 0 for the *PhD* dummy? What is 0 for the *MA* dummy?
  - Be mindful of the **reference category** when working with categorical variables!
- Note: `R` can dummy out categorical variables automatically as long as they are either stored as `factor` or `character` types. But you still need to pay attention to the **reference category** when it does so!

# Categorical Predictors

```
data2a <- data2 %>%
  mutate(educCat = ifelse(phd == 1,'PhD',
                         ifelse(ma == 1,'MA','<MA')))

m2a <- lm(inc ~ labor + educCat, data = data2a)
coef(m2a)
```

```
## (Intercept)       labor   educCatMA  educCatPhD
##   14.177459    6.038865   11.963200   39.043975
```

# Categorical Predictors

```
data2b <- data2 %>%
  mutate(educCat = ifelse(phd == 1,'1-PhD',
                          ifelse(ma == 1,'2-MA','3-<MA')))

m2b <- lm(inc ~ labor + educCat, data = data2b)
coef(m2b)
```

```
##  (Intercept)        labor  educCat2-MA educCat3-<MA
##    53.221434     6.038865   -27.080776   -39.043975
```

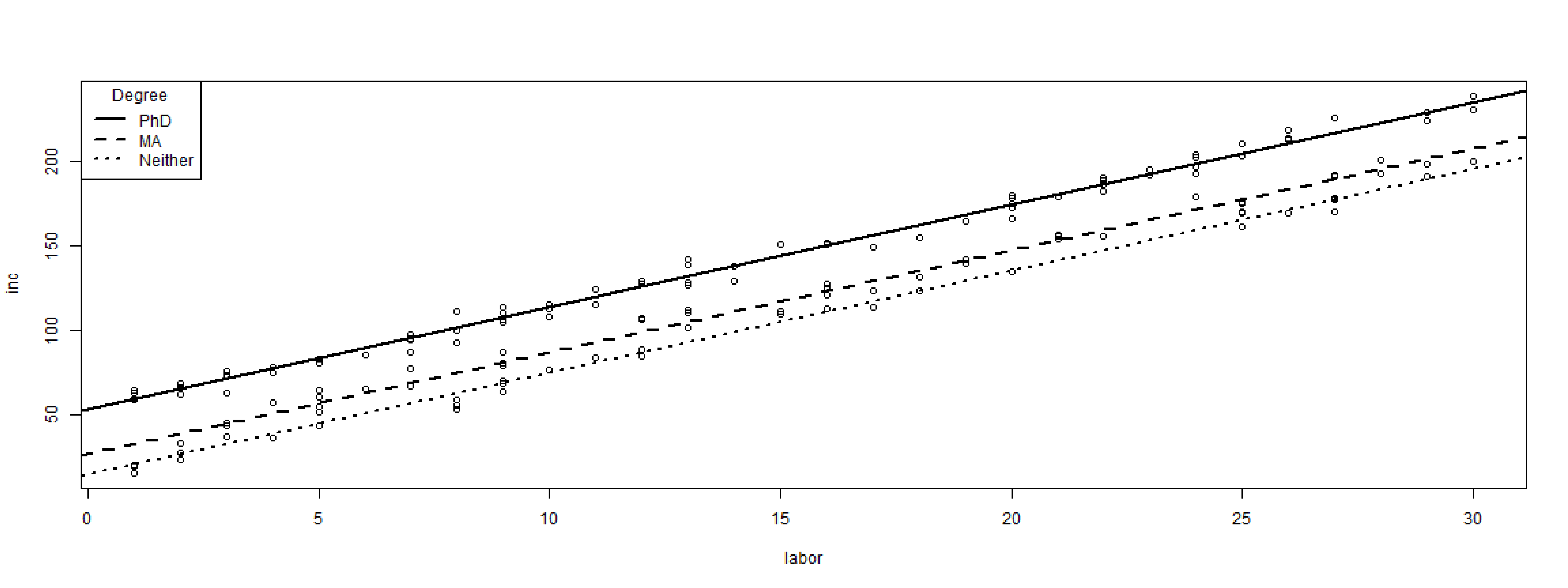
# Categorical Predictors

```
data2c <- data2a %>%
  mutate(educCat = factor(educCat,
                          levels = c('MA','PhD','<MA')))

m2c <- lm(inc ~ labor + educCat, data = data2c)
coef(m2c)
```

```
## (Intercept)       labor  educCatPhD  educCat<MA
##   26.140658    6.038865   27.080776  -11.963200
```

# Categorical Predictors



# Interactions

- Finally, what if we theorize that the relationship between $y$ and $x$ varies by some other predictor $z$?
  - I.e., we think that additional experience in the labor market increases income more for PhDs than non-PhDs
- To test this, we "interact" the two variables by multiplying them together

$$Income_i = \beta_0 + \beta_1 Labor_i + \beta_2 PhD_i + \beta_3 Labor_i * PhD_i + u_i$$

- What is the relationship between the outcome and *Labor* now?

$$\begin{aligned}\frac{\partial Income}{\partial Labor} &= \frac{\partial(\beta_0 + \beta_1 Labor + \beta_2 PhD + \beta_3 Labor * PhD + u)}{\partial x} \\ &= 0 + \beta_1 * 1 * Labor^0 + 0 + \beta_3 * PhD * Labor^0 \\ &= \beta_1 + \beta_3 PhD\end{aligned}$$

- On your own, calculate the relationship between the outcome and *PhD*

# Interactions

- **NB**: You must *always* include the "constitutive terms" of an interaction along with the interaction itself
  - I.e., if you interact $x_1$ with $x_2$, your regression cannot be written $y = \beta_0 + \beta_1 x_1 * x_2 + u$, nor can it be written $y = \beta_0 + \beta_1 x_1 + \beta_2 x_1 * x_2 + u$, nor can it be written $y = \beta_0 + \beta_1 x_2 + \beta_2 x_1 * x_2 + u$
  - It **must** be written as $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 * x_2 + u$
  - (However, there are instances in which the constitutive terms might drop out depending on the specification...see Brambor, Clark and Golder (2006))
- As with categorical variables, `R` will make your life easier by always including the constitutive terms for you
  - I.e., you can run `lm(y ~ x1*x2)` and `R` will automatically re-write as `lm(y ~ x1 + x2 + x1*x2)`

# Interactions

```
m3 <- lm(inc ~ labor + phd + labor*phd, data = data3)
summary(m3)
```

```
##
## Call:
## lm(formula = inc ~ labor + phd + labor * phd, data = data3)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.2402  -3.4522   0.0735   3.7365  11.1788
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 15.07763    1.09962   13.71   <2e-16 ***
## labor        6.00523    0.06555   91.62   <2e-16 ***
## phd         40.43731    1.59918   25.29   <2e-16 ***
## labor:phd    7.98499    0.09159   87.18   <2e-16 ***
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.189 on 146 degrees of freedom
```

# Interactions

- How do we interpret this?

```
coef(m3)
```

```
## (Intercept)       labor         phd   labor:phd
##   15.077627    6.005231   40.437307    7.984985
```

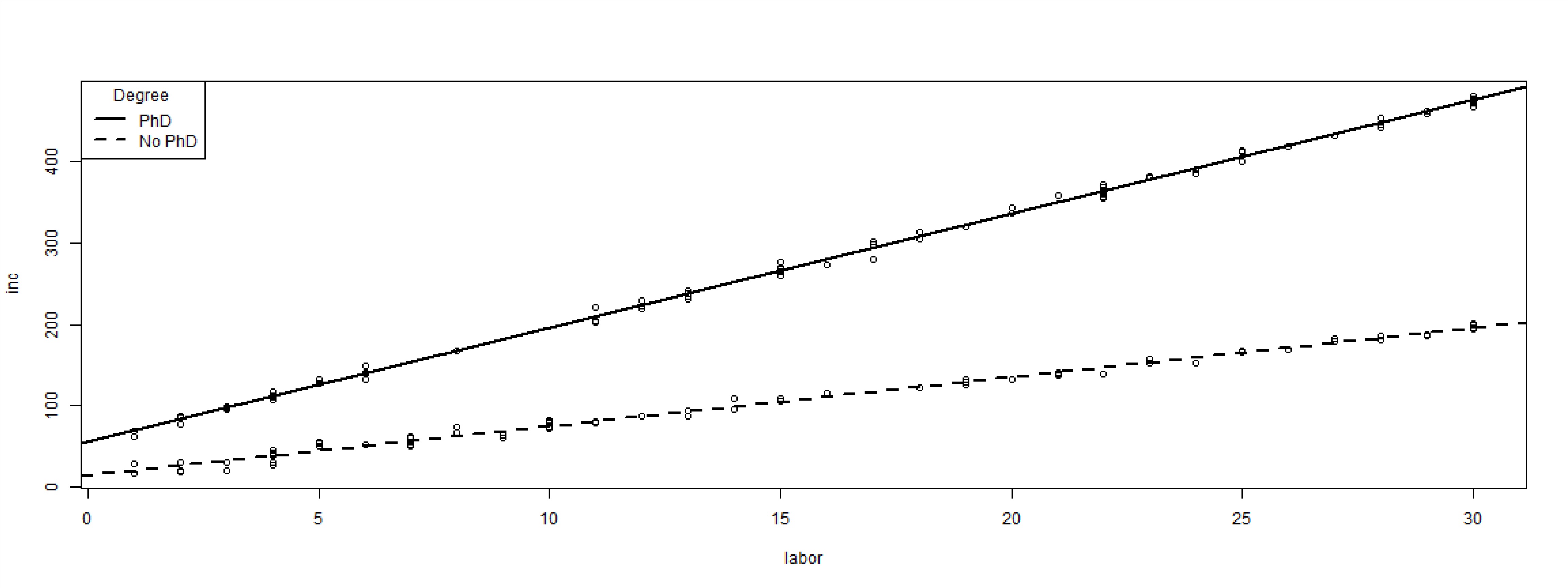
$$\frac{\partial Income}{\partial Labor} = \beta_1 + \beta_3 PhD$$

1. Among those **with** a PhD, the relationship between years in the labor market and income is 14k ( $\beta_1 + \beta_3 * 1$ )

2. Among those **without** a PhD, the relationship is 6k ( $\beta_1 + \beta_3 * 0$ )

- Often best practice is to visualize this

# Interactions



# Interactions

- What if we interact two continuous variables? Math stays the same!

```
m4 <- lm(inc ~ labor + age + labor*age,data4)
summary(m4)
```

```
##
## Call:
## lm(formula = inc ~ labor + age + labor * age, data = data4)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -33.459  -8.082  -1.084   8.929  32.162
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.80194    7.64010    0.89    0.375
## labor        9.17093    0.43578   21.05   <2e-16 ***
## age          3.14616    0.16891   18.63   <2e-16 ***
## labor:age   -0.20539    0.00953  -21.55   <2e-16 ***
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

# Interactions

- How to interpret?

```
coef(m4)
```

```
## (Intercept)       labor         age   labor:age
##   6.8019452   9.1709284   3.1461574  -0.2053922
```

$$\begin{aligned}\frac{\partial Income}{\partial Labor} &= \frac{\partial(\beta_0 + \beta_1 Labor + \beta_2 Age + \beta_3 Labor * Age + u)}{\partial x} \\ &= 0 + \beta_1 * 1 * Labor^0 + 0 + \beta_3 * Age * Labor^0 \\ &= \beta_1 + \beta_3 Age\end{aligned}$$

- As with the quadratic discussion, we want to evaluate this at different values of *Age*

2. At what values of *Age* is the relationship between *Income* and *Labor* **statistically distinct**, and are these values **substantively meaningful**?

# Interactions

- Marginal Effects plots are especially useful here
    - Visualize the $\beta_1$ slope ( $\hat{\beta}_1$ on the y-axis) at different values of *Age* (on the x-axis)
    - Literally answer, what is the additional income expected with an additional year experience for a 20 year old versus a 60 year old?
    - (What do you think this would be?)

# Interactions