MATH 3307
Lesson 21

Non-Linear Methods

Many times a scatter-plot reveals a curved pattern instead of a linear pattern.

We can **transform** the data by changing the scale of the measurement that was used when the data was collected. In order to find a good model we may need to transform our x value or our y value or both.

In this example from section 5.4, we saw that the linear model was not a good fit for this data:

Year	1790	1800	1810	1820	1830	1840	1850	1860	1870	1880
People per square mile	4.5	6.1	4.3	5.5	7.4	9.8	7.9	10.6	10.09	14.2
Year	1890	1900	1910	1920	1930	1940	1950	1960	1970	1980
People per square mile	17.8	21.5	26	29.9	34.7	37.2	42.6	50.6	57.5	64

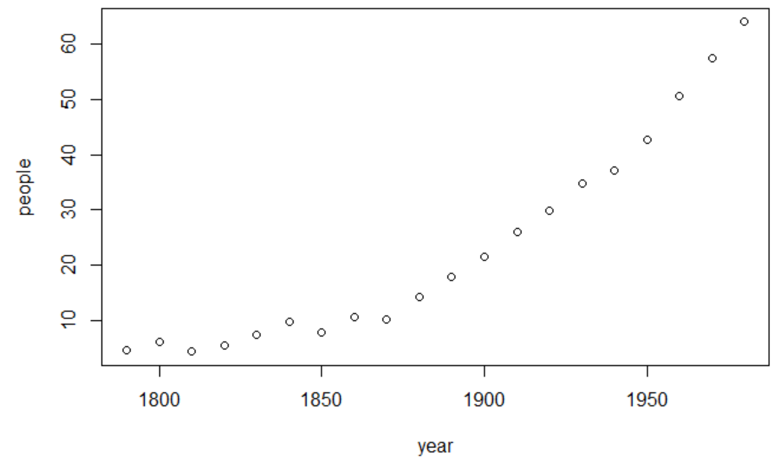
Data to copy

```
assign("year",c(1790,1800,1810,1820,1830,1840,1850,1860,1870,1880
,1890,1900,1910,1920,1930,1940,1950,1960,1970,1980))
```

```
assign("people",c(4.5,6.1,4.3,5.5,7.4,9.8,7.9,10.6,10.09,14.2,17.8,21.5,
26,29.9,34.7,37.2,42.6,50.6,57.5,64))
```

Linear Model

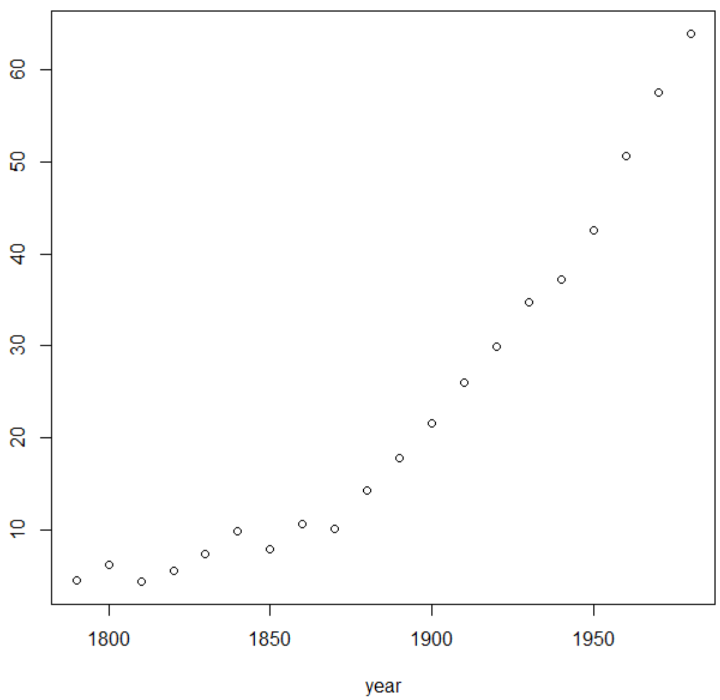
```
> plot(year,people)
> lm(people~year)

Call:
lm(formula = people ~ year)

Coefficients:
(Intercept)         year
  -544.0672       0.3009

> cor(year,people)^2
[1] 0.8894762
```

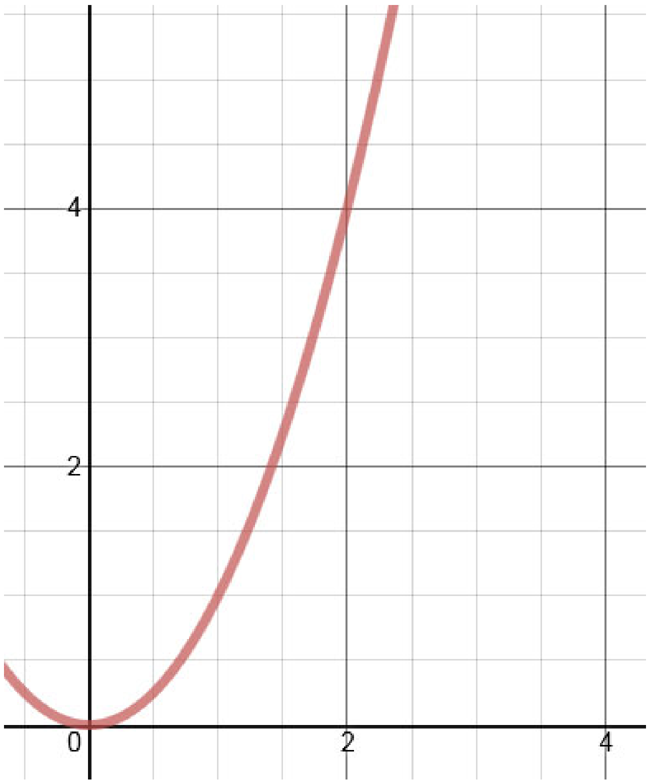


$\hat{y} = .3009x - 544.0672$

Linear equations should have no exponents (larger than one) on either the x or y variables.

Obviously, this does not look like a straight line. It looks more like a parabola.

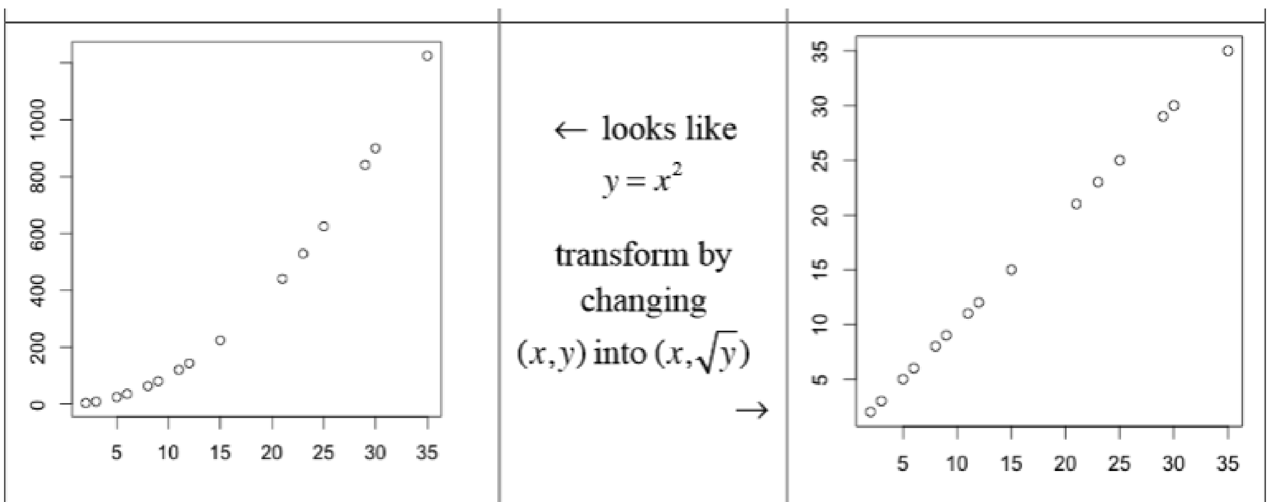




How to change the graph.

$y = x^2$ $(x, y) \to (x, \sqrt{y})$

U shaped



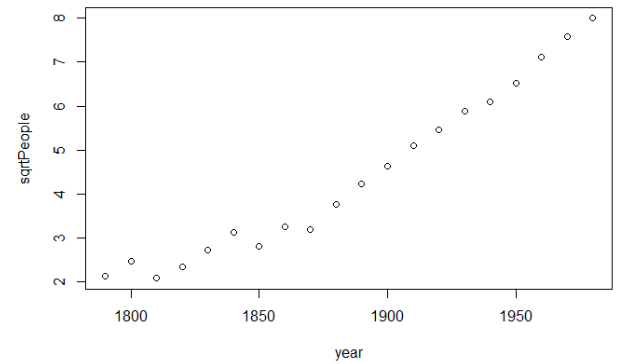
in R Studio:

```
sqrtpeople=sqrt(people)
plot(year,sqrtpeople)
```

Illustrations from the textbook.

Parabolic Model

```
> sqrtPeople=sqrt(people)
> plot(year,sqrtPeople)
> lm(sqrtPeople~year)

Call:
lm(formula = sqrtPeople ~ year)

Coefficients:
(Intercept)          year
  -55.55919       0.03182
> cor(year,sqrtPeople)^2
[1] 0.9550843
```



$$\left(\sqrt{\hat{y}}\right)^2 = \left(.03182x - 55.559\right)^2$$

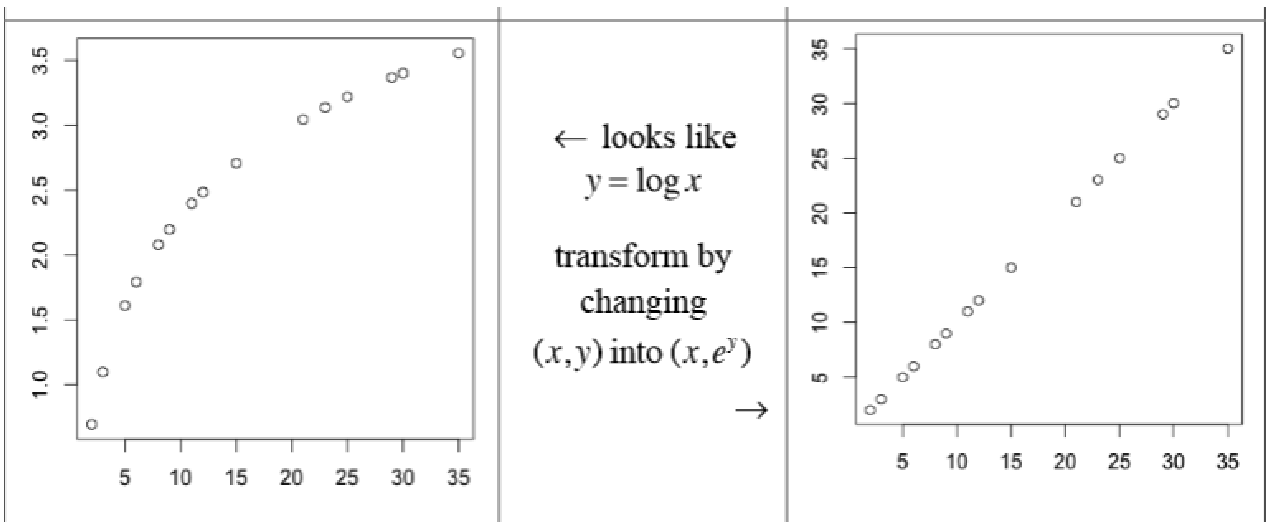
$$\hat{y} = (.03182x - 55.559)^2$$

In a parabolic (or quadratic) model, there should be an exponent of two on either the x or y variables (typically on the x-variable)

How to change the graph.

$y = \ln x$ $(\ln x, y) \to (x, e^y)$

Fast change followed by slow change



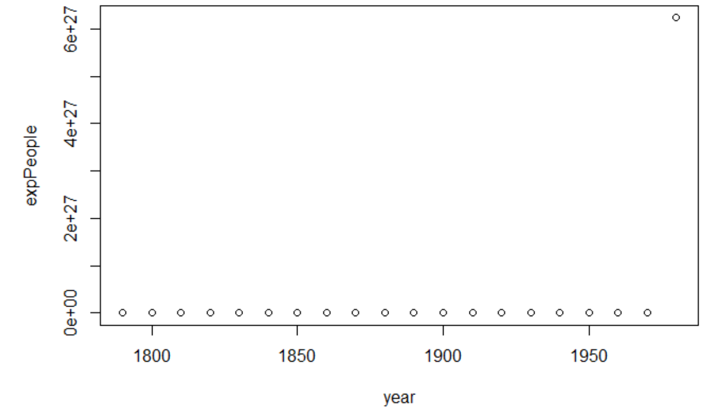
in R Studio:

```
exppeople=exp (people)
plot(year,exppeople)
```

Illustrations from the textbook.

Logarithmic Model

```
> expPeople=exp(people)
> plot(year,expPeople)
> lm(expPeople~year)

Call:
lm(formula = expPeople ~ year)

Coefficients:
(Intercept)         year
 -1.650e+28    8.919e+24

> cor(year,expPeople)^2
[1] 0.1432645
```



$$e^{\hat{y}} = 8,919\times10^{24}x - 1,65\times10^{28}$$

$$\hat{y} = \ln(8,919\times10^{24}x - 1,65\times10^{28})$$

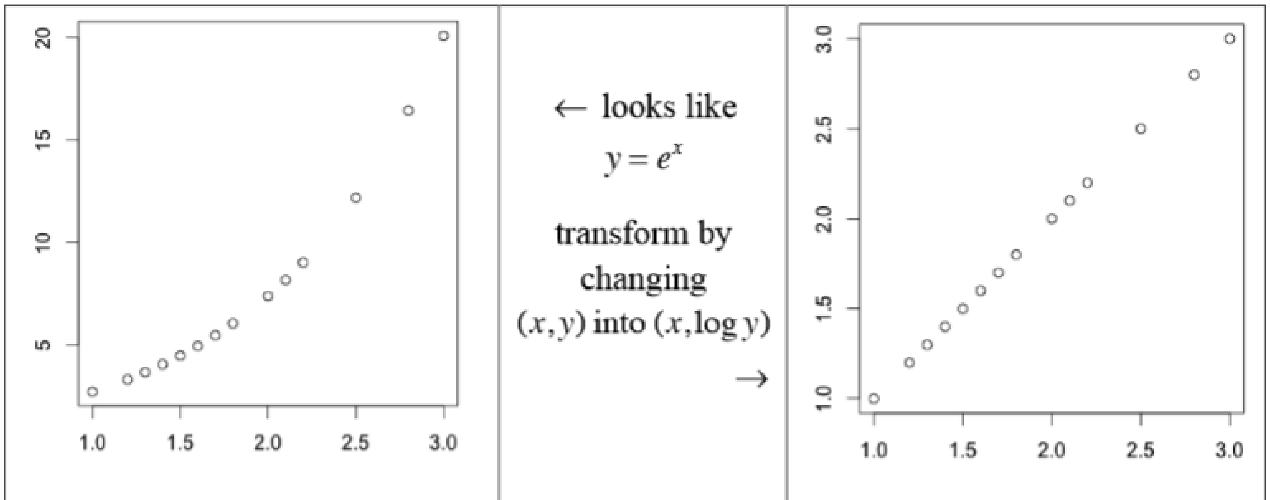
Logarithmic Models will have an ln(something) or log(something) in their equation.

How to change the graph.

$(e^x, y) \to (x, \ln y)$

$y = e^x$

slow change then fast change

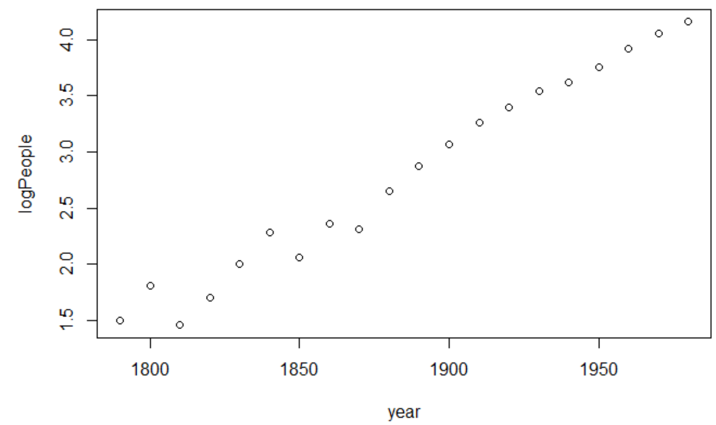


in R Studio:
logpeople=log(people)
plot(year,logpeople)

Illustrations from the textbook.

Exponential Model

```
> logPeople=log(people)
> plot(year,logPeople)
> lm(logPeople~year)

Call:
lm(formula = logPeople ~ year)

Coefficients:
(Intercept)         year
  -25.21748      0.01486

> cor(year,logPeople)^2
[1] 0.9782003
```



$\ln \hat{y} = 0.01486x - 25.217$

$\hat{y} = e^{0.01486x - 25.217}$

In the equation of the exponential model, the x-variable will appear in the exponent of a constant.

Summarize:

Linear: r^2: 0.889

Parabolic: r^2: 0.955

Logarithmic: r^2: 0.143

Exponential: r^2: 0.978

If we had to chose which of these four models is the best fit to this set of data, we would have to chose the exponential model, since the r^2 value is closest to 1.

Compare the scatterplots for linear regression and quadratic regression.

Popper 16:

Find the r^2 values for each to determine the best curve of fit.

1. For the Linear Model B
2. For the Quadratic Model D
3. For the Logarithmic Model C
4. For the Exponential Model A

a. 0.9782 b. 0.8895 c. 0.1433 d. 0.9551

5. Which is the best model of this data?

a. Linear b. Quadratic c. Logarithmic d. Exponential