Understanding Experimental Data

Statistics Meets Experimental Science

- Conduct an experiment to gather data
  - Physical (e.g., in a biology lab)
  - Social (e.g., questionnaires)
- Use theory to generate some questions about data
  - Physical (e.g., gravitational fields)
  - Social (e.g., people give inconsistent answers)
- Design a computation to help answer questions about data

Net Gain on a missed jump shot

$$= P(\text{off reb}) \times E[\text{pts for}] - P(\text{def reb}) \times E[\text{pts against}]$$

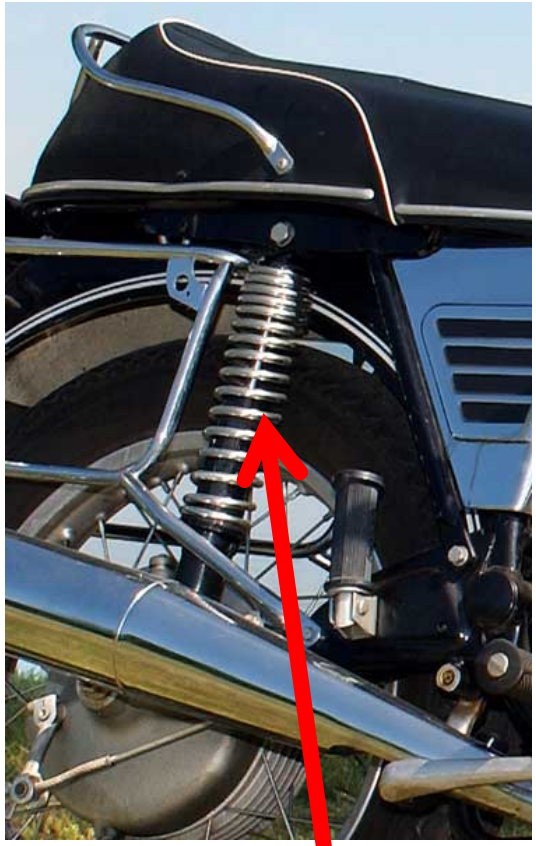
- Consider, for example, a spring

One Kind of Spring

Another Kind of Spring

Photo by Bachsteize

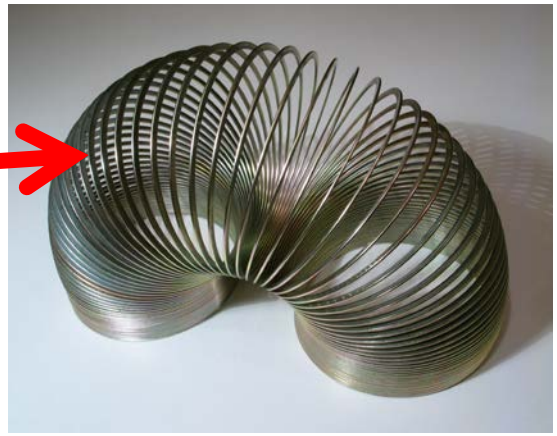
This Kind of Spring



$k \approx 35,000 N/m$

$k \approx 1 N/m$

$1N = 1kg\frac{m}{s^2}$



Linear spring: amount of force needed to stretch or compress spring is linear in the distance the spring is stretched or compressed

Each spring has a spring constant, k, that determines how much force is needed

Hooke's Law

- F = -kd
- How much does a rider have to weigh to compress spring 1cm?

$F = 0.01m * 35{,}000N/m$

$F = 350N$

$mass * 9.8m/s^2 = 350N$

$$mass = \frac{350N}{9.81m/s^2}$$

$$mass = \frac{350kg}{9.81}$$

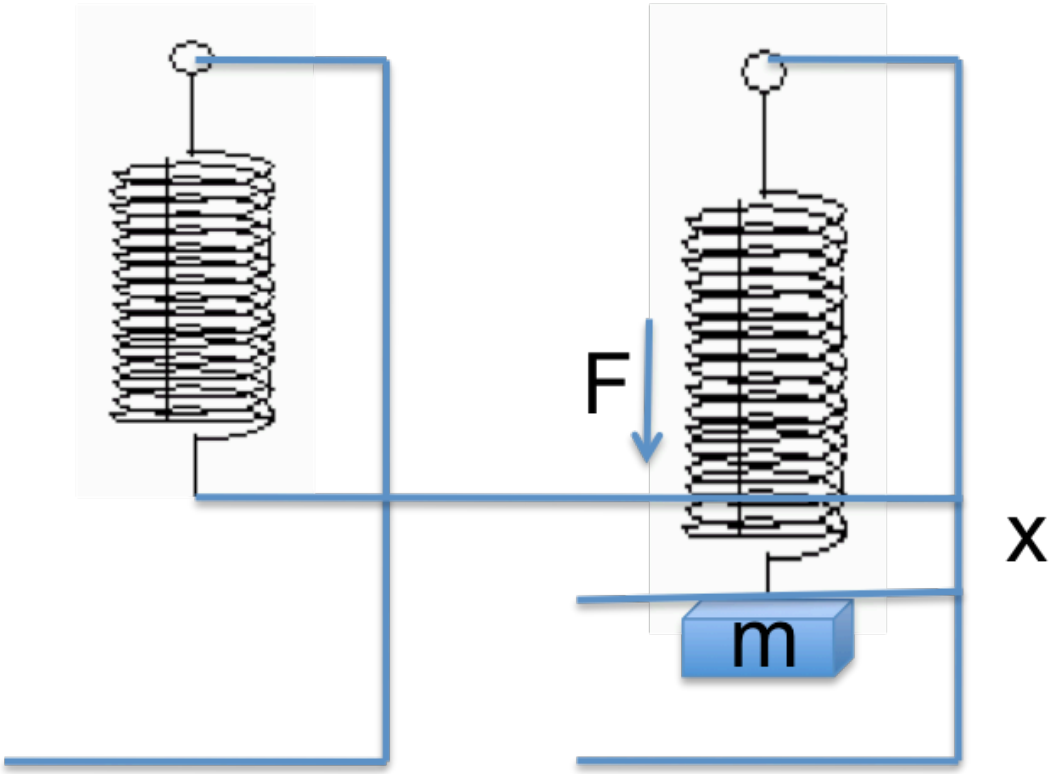
$mass \approx 35.68kg$

$F = mass * acc$

$F = mass * 9.8m/s^2$

Finding k

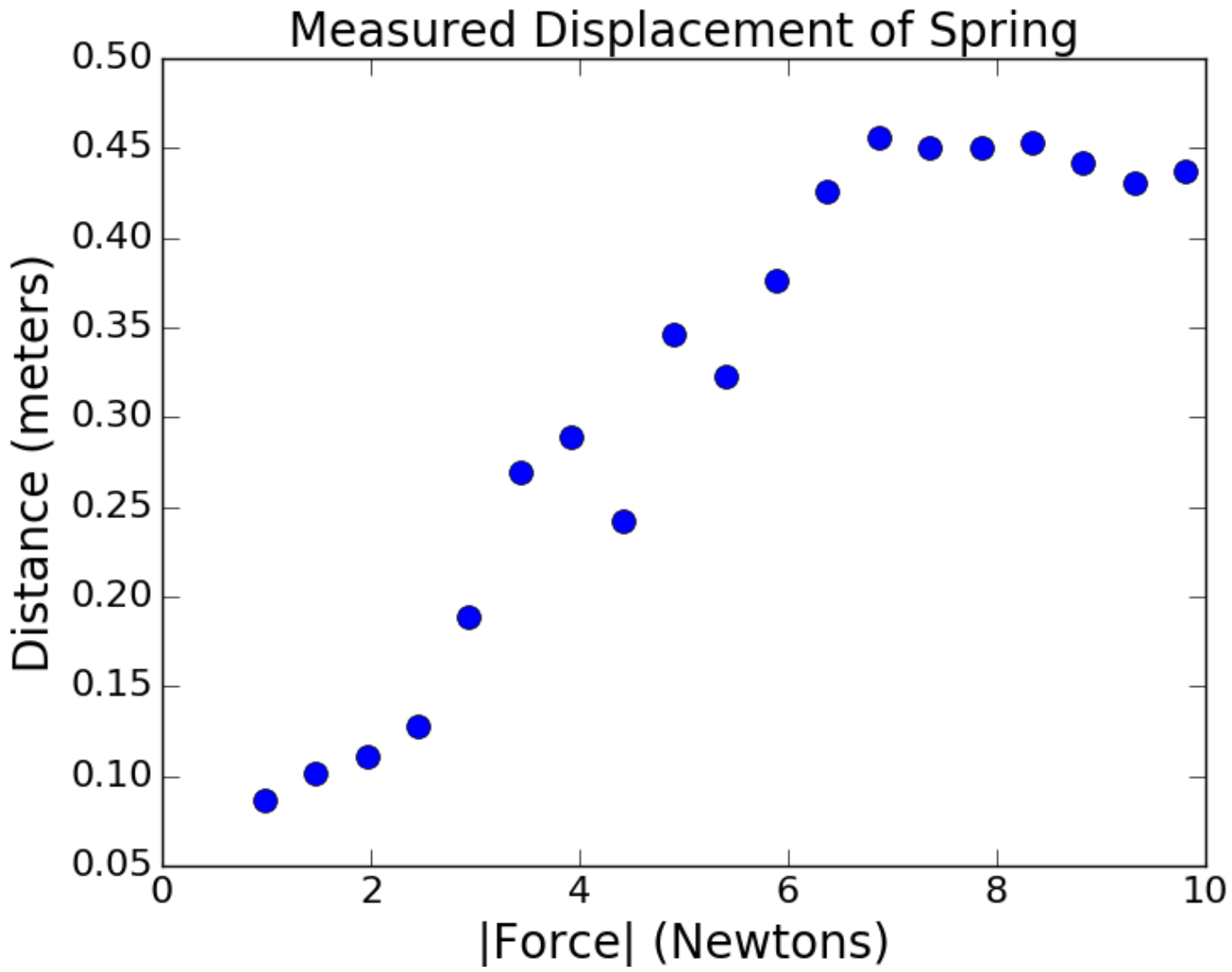
- F = -kx
- k = -F/x
- k =9.86*m/x



Some Data

Distance (m)	Mass (kg)
0.0865	0.1
0.1015	0.15
0.1106	0.2
0.1279	0.25
0.1892	0.3
0.2695	0.35
0.2888	0.4
0.2425	0.45
0.3465	0.5
0.3225	0.55
0.3764	0.6
0.4263	0.65
0.4562	0.7

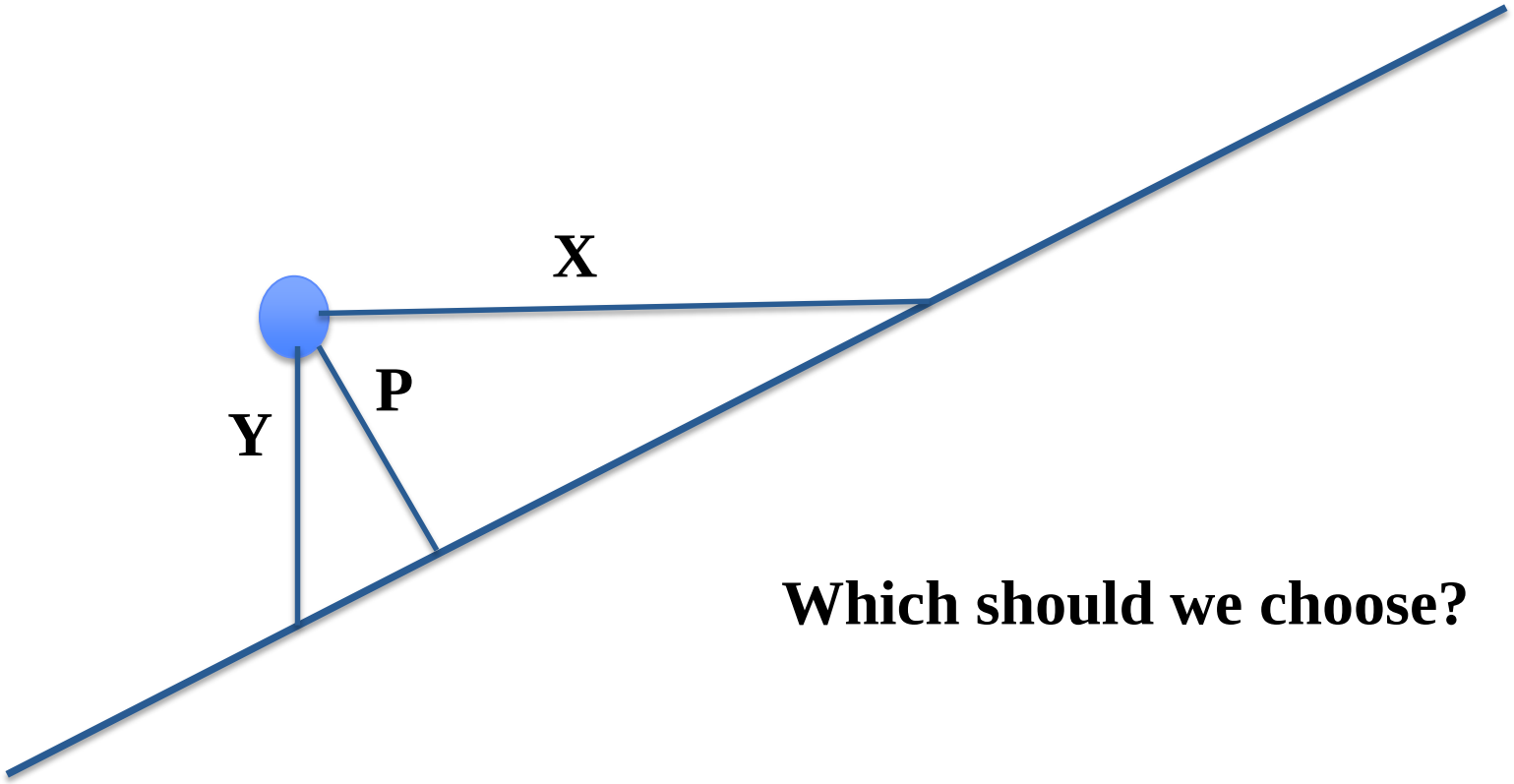
Taking a Look at the Data



Fitting Curves to Data

- When we fit a curve to a set of data, we are finding a fit that relates an independent variable (the mass) to an estimated value of a dependent variable (the distance)
- In this case, we want to find a line such that some function of the sum of the distances from the line to the measured points is minimized

Measuring Distance



Which should we choose?

Least Squares Objective Function

$$\sum_{i=0}^{len(observed)-1} (observed[i] - predicted[i])^2$$

- Look familiar?