Least Squares Best Fit Line

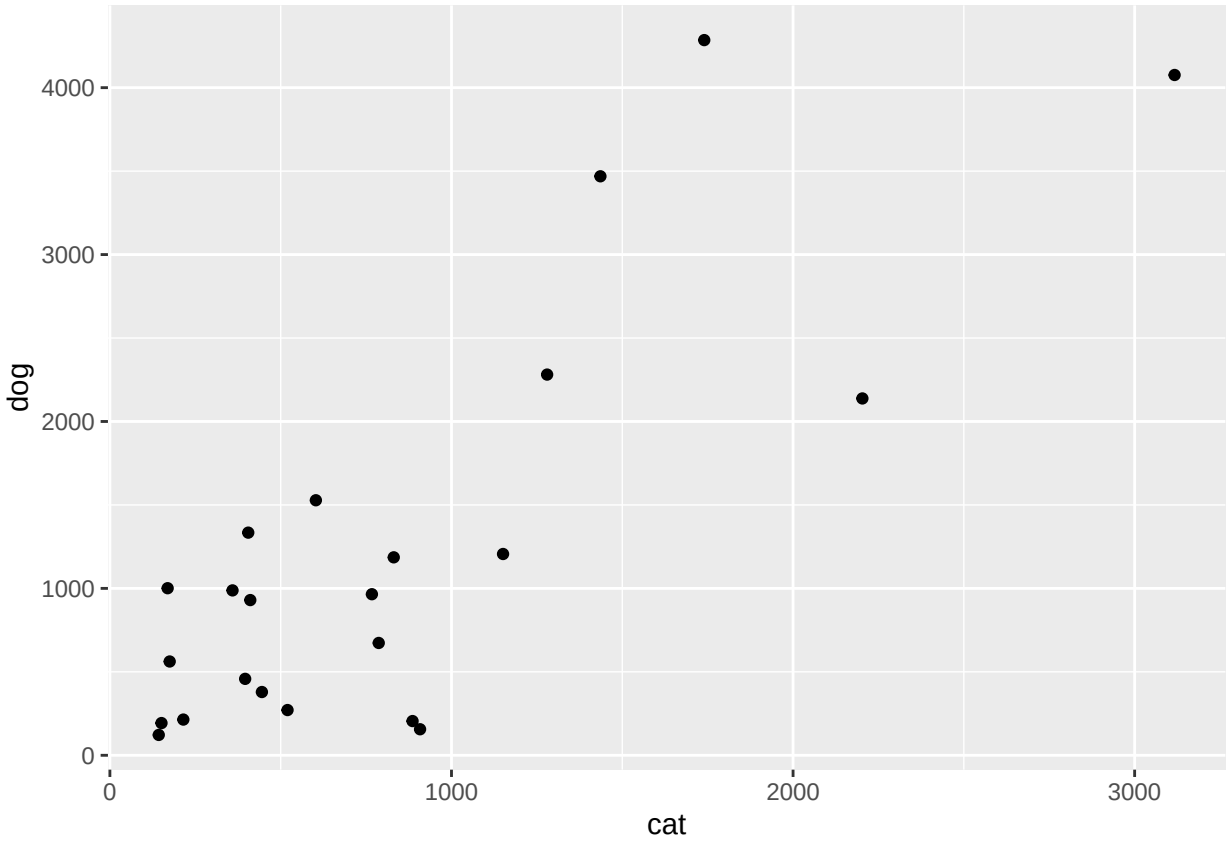
Michael Ghattas

Find Minimum Square Error Line

View the data. These are from a *Pets for Life* data set provided to the author.

```
dat<-read.csv("animal_stats_compact.csv")

(g<-ggplot(dat, aes(x=cat,y=dog))+geom_point())
```

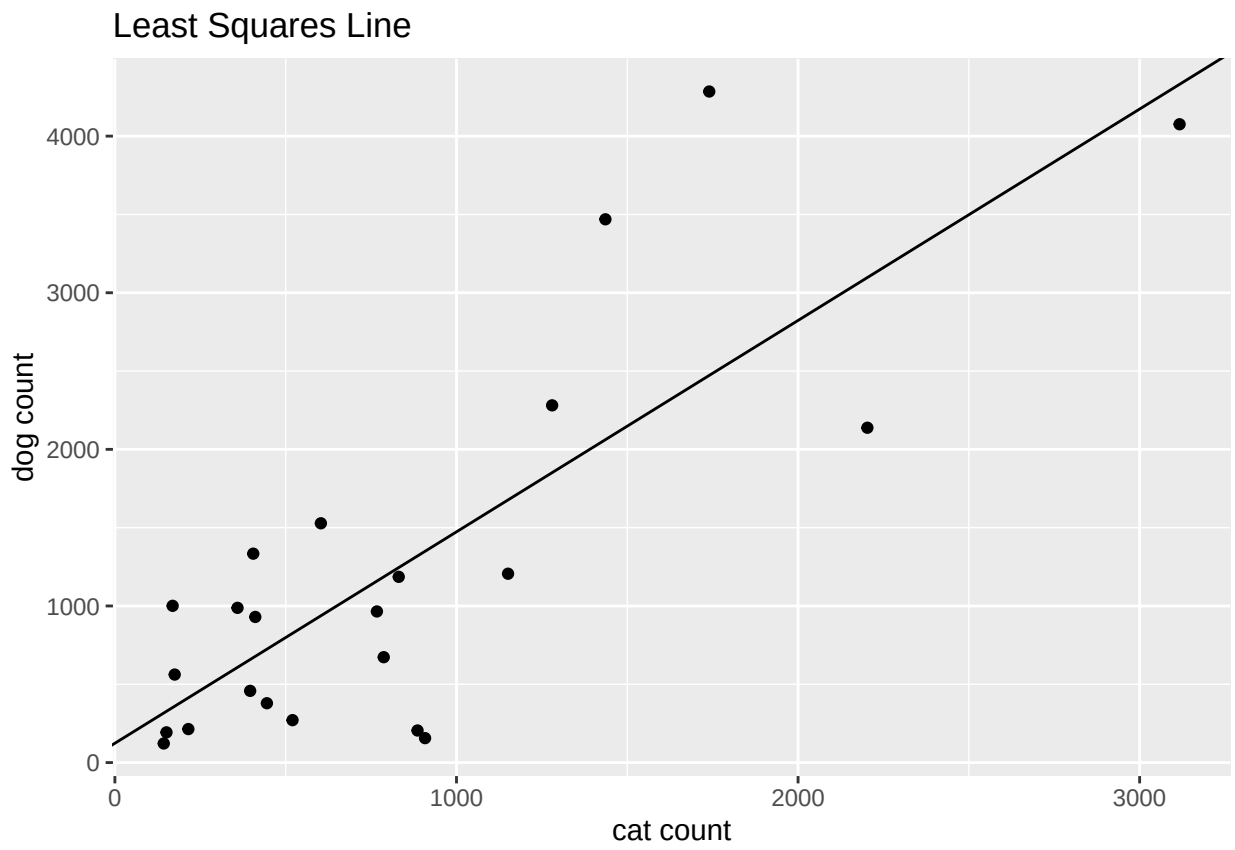


Calculate the slope and intercept according to the formula.

- Recall that the slope in the $y = mx + b$ model according to the least squares criterion is $\frac{\frac{1}{n}\sum x_i y_i - \bar{x}\bar{y}}{\frac{1}{n}\sum x_i^2 - \bar{x}^2}$ and the intercept is $\bar{y} - m\bar{x}$.*

```
m<-(mean(dat$cat*dat$dog)-mean(dat$cat)*mean(dat$dog))/(mean(dat$cat^2)-mean(dat$cat)^2)
b<-mean(dat$dog-m*dat$cat)

g<-g+geom_abline(slope=m,intercept = b)
g<-g+labs(title="Least Squares Line",x="cat count",y="dog count")
g
```



Or use minimization directly. (This is just for illustration. It is not a recommended way to do this calculation.)

```
sq_error<-function(x){
  return(sum((dat$dog-x[2]*dat$cat-x[1])^2))
}
nlm(sq_error,p=c(0,1))$estimate
```

```
## [1] 123.506089   1.349846
```

There is a built-in function for this. The formula $dog = m\,(cat) + b$ is represented by "dog~cat".

```
lm(dog~cat,data=dat)$coefficients
```

```
## (Intercept)         cat
##  123.504954    1.349847
```

Practice

- In the code block below, please calculate the estimated slope and intercept from the formulas $\frac{\frac{1}{n}\sum x_i y_i - \bar{x}\bar{y}}{\frac{1}{n}\sum x_i^2 - \bar{x}^2}$ and $\bar{y} - m\bar{x}$, respectively, but this time modeling "dog" as the x-variable and "cat" as the y-variable.Output your values for "m" and "b". Check your formulas using the appropriate call to "lm".*

```
sq_error<-function(x){
  return(sum((dat$cat - x[2] * dat$dog - x[1])^2))
}
nlm(sq_error,p=c(0,1))$estimate
```

```
## [1] 231.200597   0.481495
```

```
lm(cat~dog, data = dat)$coefficients
```

```
## (Intercept)         dog
##  231.200579    0.481495
```