

# Lab3

Merlise Clyde

February 3, 2017

# Topics

- ▶ wercker
- ▶ termplot
- ▶ more on interactions

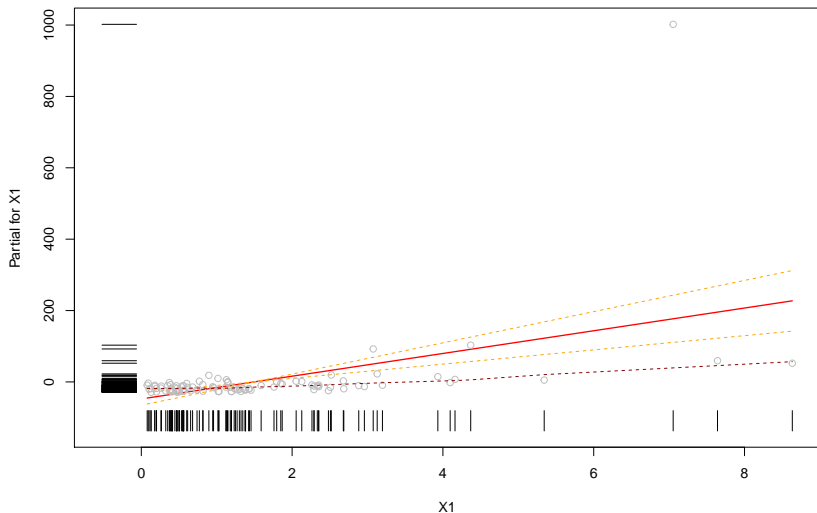
- ▶ Check your builds to see where things are breaking
- ▶ If there are missing packages let us know so that we may update the `wercker.yml`
- ▶ Should have passing badge at time of submission

## Termplot Example

```
n = 100
logx1 = rnorm(n)
x1 = exp(logx1)
x2 = abs(rnorm(n))
logy = .5 + 3*logx1 - 2*sqrt(x2) + rnorm(n)
simdat = data.frame(X1=x1, X2 = x2, Y=exp(logy))
```

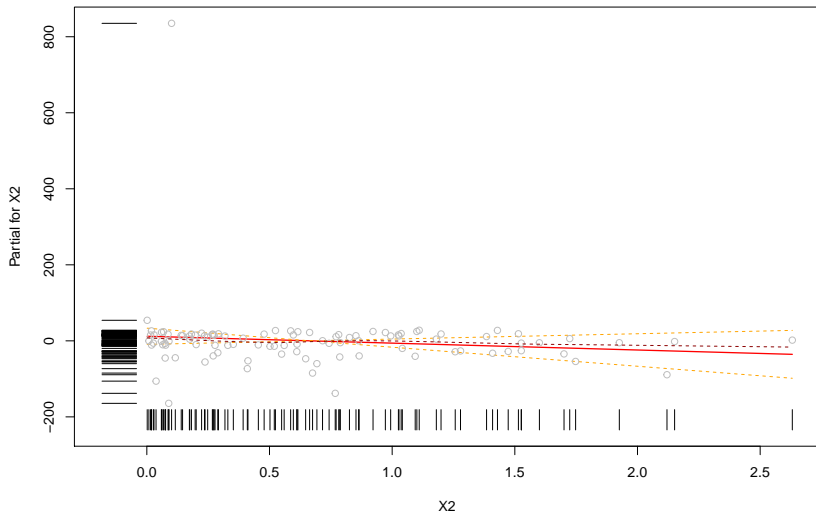
# Termplot output

```
termplot(lm(Y ~ X1 + X2, data=simdat), terms = "X1",  
         partial.resid = T, se=T, rug=T,  
         smooth = panel.smooth)
```



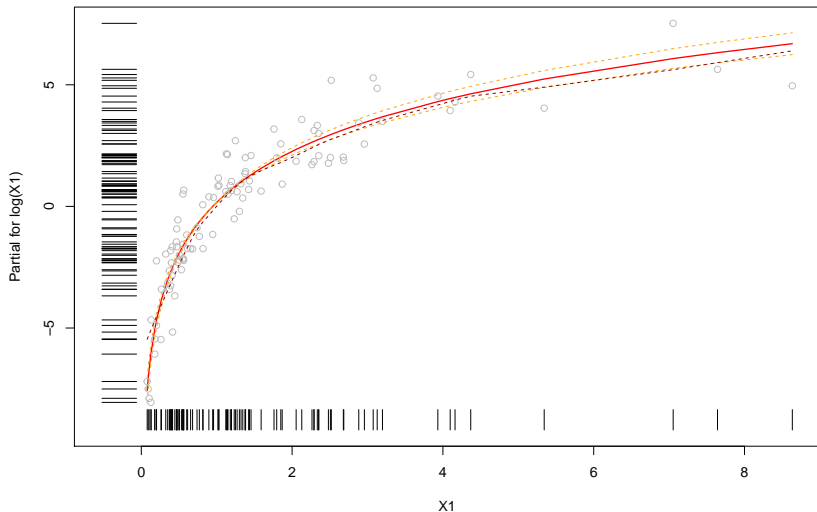
# Termplot output

```
termplot(lm(Y ~ X1 + X2, data=simdat), terms = "X2",  
         partial.resid = T, se=T, rug=T,  
         smooth = panel.smooth)
```



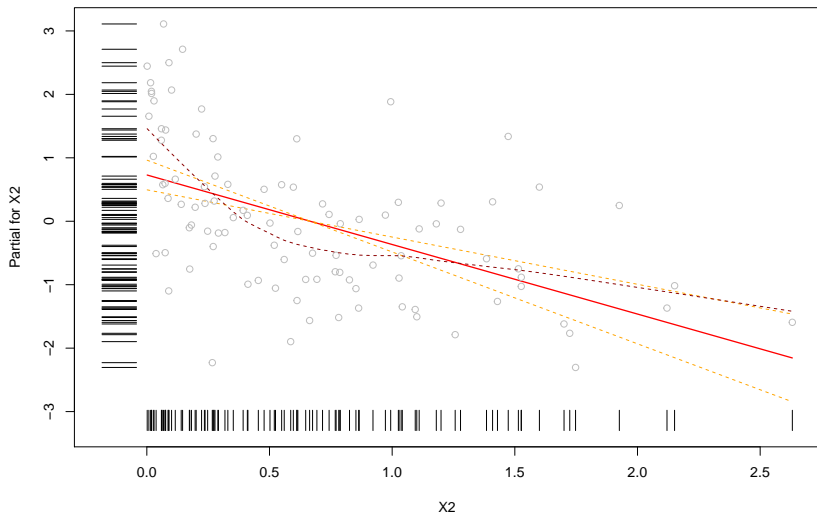
## termplot with transformation of Y and X1

```
termplot(lm(log(Y) ~ log(X1) + X2, data=simdat),  
         terms = "log(X1)", partial.resid = T, se=T, rug=T,  
         smooth = panel.smooth)
```



## termplot with transformation of Y and X1

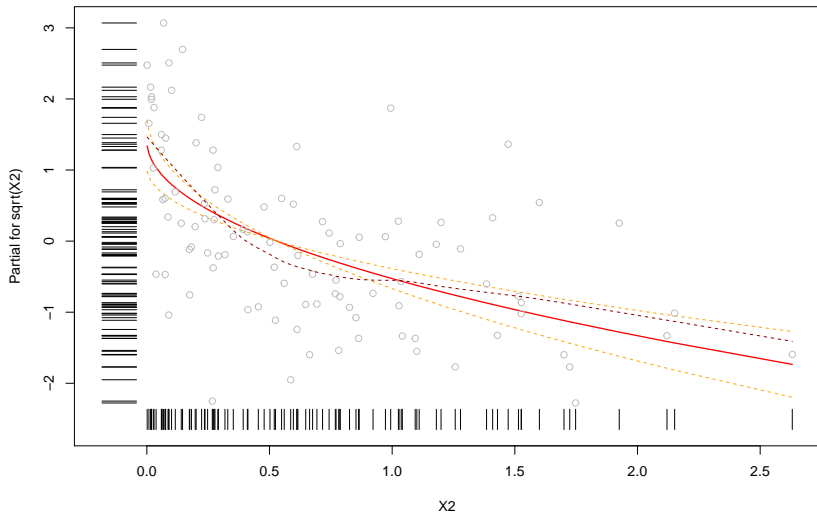
```
termplot(lm(log(Y) ~ log(X1) + X2, data=simdat), terms = "X2",  
         partial.resid = T, se=T, rug=T, smooth = panel.smooth)
```





## termplot with transformation of Y, X1, and X2

```
termplot(lm(log(Y) ~ log(X1) + sqrt(X2), data=simdat), term  
         partial.resid = T, se=T, rug=T, smooth = panel.smo
```



## What is in a term plot?

- ▶ x-axis is the (untransformed) variable in your dataframe ( $X_1, X_2$ )
- ▶ line is the “term” of that variable’s contribution to  $f(x)$
- ▶ y-axis is partial residuals for term
- ▶ `partial.resid = T` adds the partial residuals to the plot
- ▶ `rug = T` shows location of data on axes
- ▶ `se = T` adds the SE of the term’s contribution to  $f(x)$
- ▶ `smooth = panel.smooth` adds “smoothed” means to plot

## Terms

$$Y = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + e$$

Equivalent to centered model

$$Y = \bar{Y} + \hat{\beta}_1 (X_1 - \bar{X}_1) + \hat{\beta}_2 (X_2 - \bar{X}_2) + e$$

Terms are coefficient estimates times centered predictors

$$\hat{\beta}_1 (X_1 - \bar{X}_1)$$

$$\hat{\beta}_2 (X_2 - \bar{X}_2)$$

## Terms with transformations

$$\log(Y) = \hat{\beta}_0 + \hat{\beta}_1 \log(X_1) + \hat{\beta}_2 X_2 + e$$

Equivalent to centered model

$$\log(Y) = \log(\bar{Y}) + \hat{\beta}_1 (\log(X_1) - \log(\bar{X}_1)) + \hat{\beta}_2 (X_2 - \bar{X}_2) + e$$

Terms are coefficient estimates times centered “predictors”

$$\hat{\beta}_1 (\log(X_1) - \log(\bar{X}_1))$$

## partial residuals for a term

$$\log(Y) = \log(\bar{Y}) + \hat{\beta}_1(\log(X1) - \log(\bar{X1})) + \hat{\beta}_2(X2 - \bar{X2}) + e$$

$$\log(Y) - (\log(\bar{Y}) + \hat{\beta}_1(\log(X1) - \log(\bar{X1}))) = \hat{\beta}_2(X2 - \bar{X2}) + e$$

- ▶ Lefthand side takes response and removes the part of the response that is explained by  $X1$
- ▶ Equal to the term for  $X2$  plus the residual  $e$
- ▶ part of residual variation that is not explained by the other terms that potentially can be explained by  $X2 =$  partial residual for  $X2$
- ▶ partial residual for  $X1$

$$\hat{\beta}_1(\log(X1) - \log(\bar{X1})) + e$$

## All possible interactions

```
lm(log(Y) ~ (log(X1) + sqrt(X2))^2, data=simdat)
```

```
##
```

```
## Call:
```

```
## lm(formula = log(Y) ~ (log(X1) + sqrt(X2))^2, data = simdat)
```

```
##
```

```
## Coefficients:
```

##	(Intercept)	log(X1)	sqrt(X2)	log(X1):sqrt(X2)
##	0.3916	2.7542	-1.9643	0.0000

Equivalent to model formula  $\log(X1) + \sqrt{X2} + \log(X1):\sqrt{X2}$

## Selective proposal Interactions

- ▶ add interactions if main effects are “significant”
- ▶ think about possible reason for interaction (science)