

Regularization and Big Data

John Myles White

June 21, 2011

Data, data, data! I can't make bricks without clay!

According to reputable sources, we're in the Age of Big Data.

But what makes data big?

A diagram illustrating a 4x4 matrix. To the left of the matrix, the letter **N** is written vertically above a downward-pointing arrow. Below the matrix, a horizontal double-headed arrow is labeled **D**, indicating the dimension of the matrix.

11	12	91	87
19	21	17	92
84	71	29	35
73	67	53	48

What happens as $N \rightarrow \infty$?

Traditionally, good things:

- ▶ Law of Large Numbers
- ▶ Central Limit Theorem
- ▶ Consistent estimators
- ▶ Asymptotic guarantees
- ▶ Large sample theory

What happens as $D \rightarrow \infty$?

Traditionally, bad things:

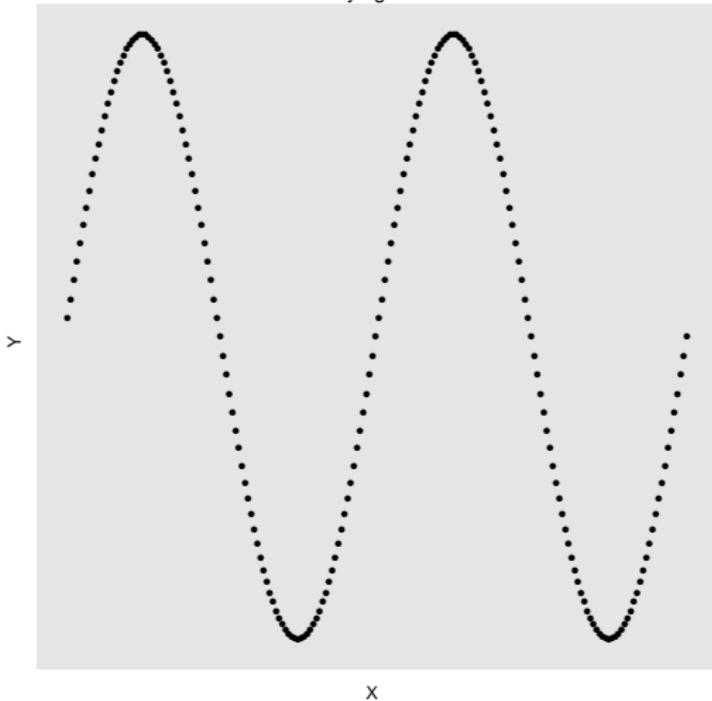
- ▶ Underdetermined systems
- ▶ Infinitely many perfect models
- ▶ Massive overfitting

What leads to overfitting?

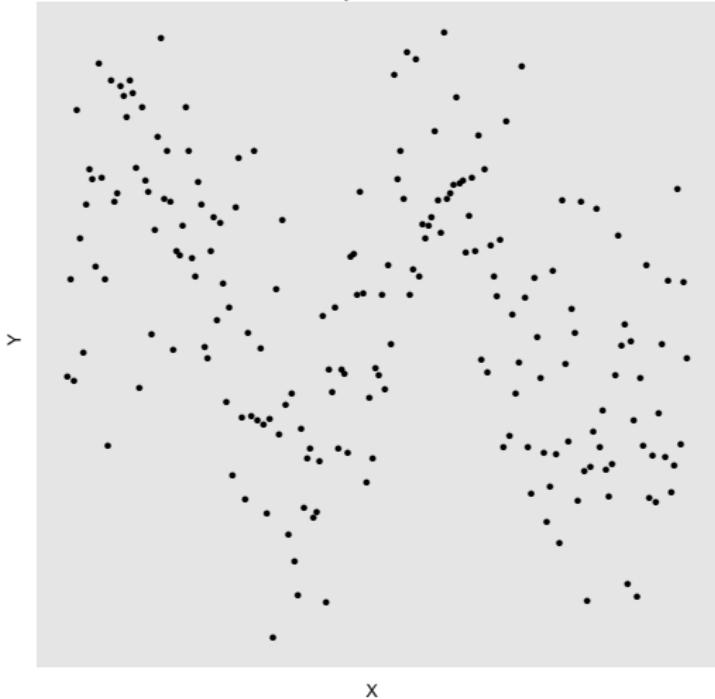
Example data:

- ▶ $N = 200$
- ▶ $D = 1$
- ▶ $y = \sin(4\pi x) + \epsilon$
- ▶ $\epsilon \sim N(0, 0.75)$

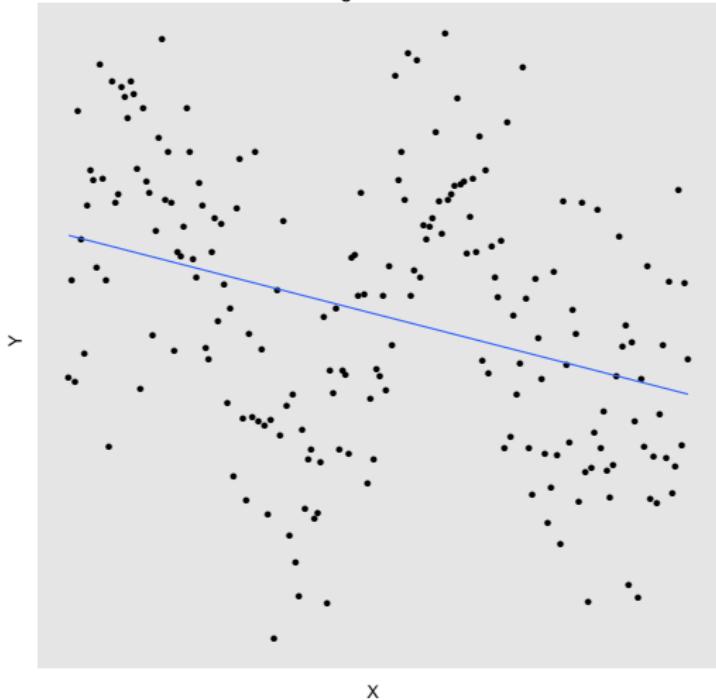
Underlying Pattern



Noisy Data Set



Linear Regression Results



Linear regression:

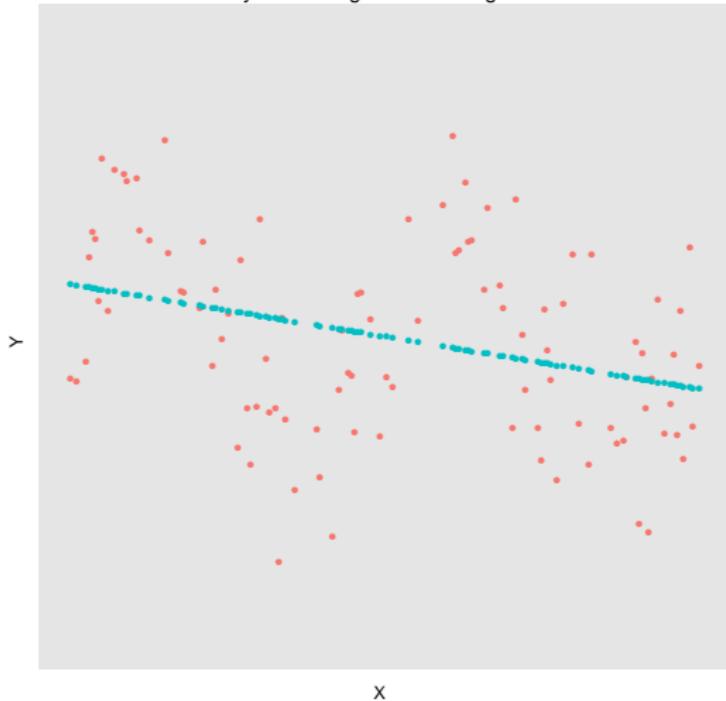
- ▶ $Y = \beta_0 + \beta_1 X$

Linear regression isn't very expressive. We need a stronger model.

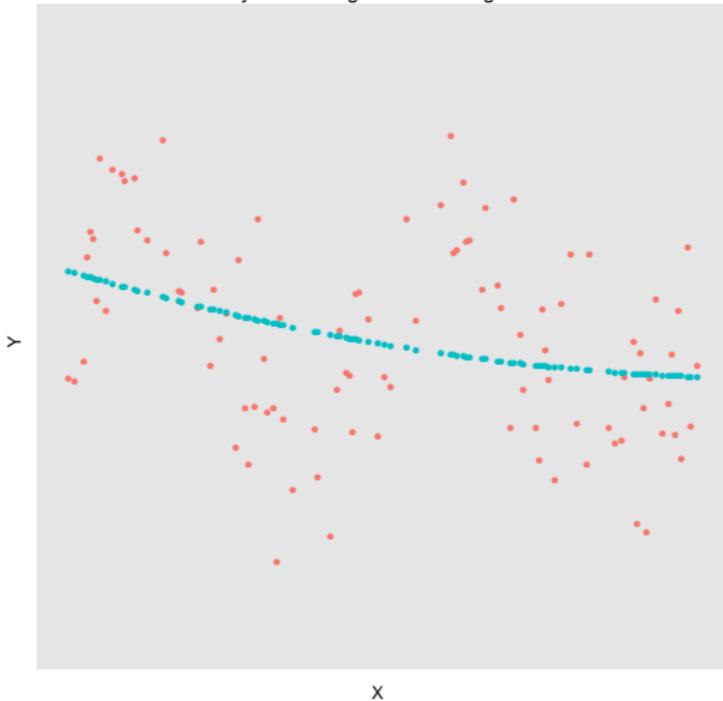
Polynomial regression:

- ▶ $Y = \beta_0 + \beta_1 X + \beta_2 X^2$
- ▶ $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3$
- ▶ ...
- ▶ $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \dots + \beta_{20} X^{20}$

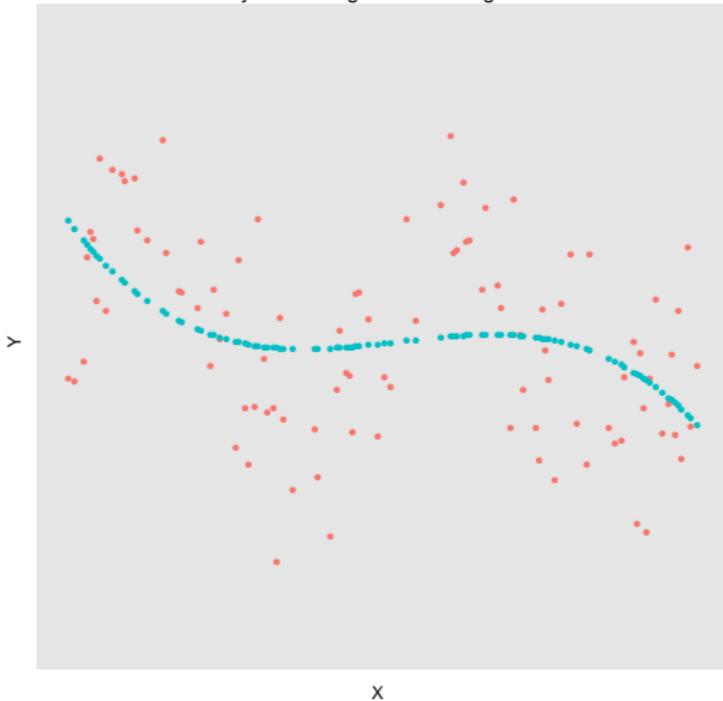
Polynomial Regression of Degree 1



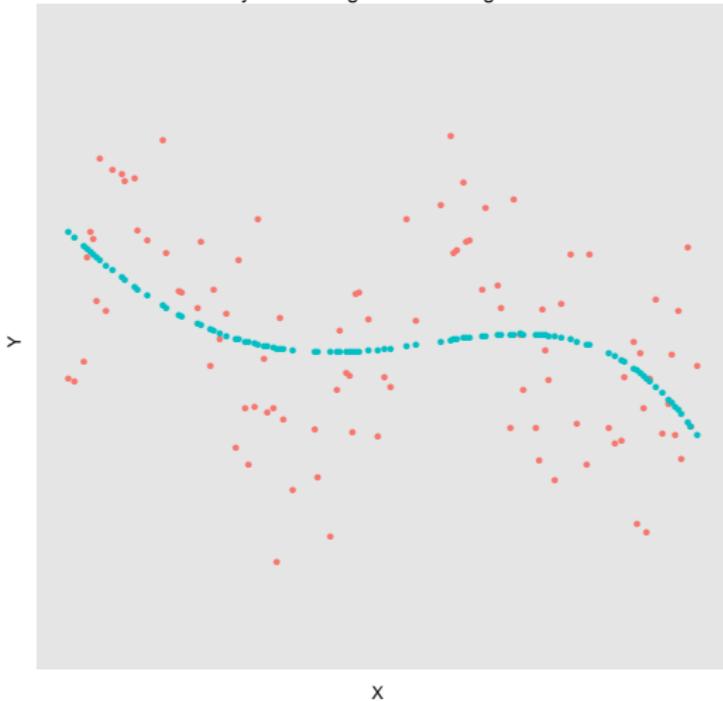
Polynomial Regression of Degree 2



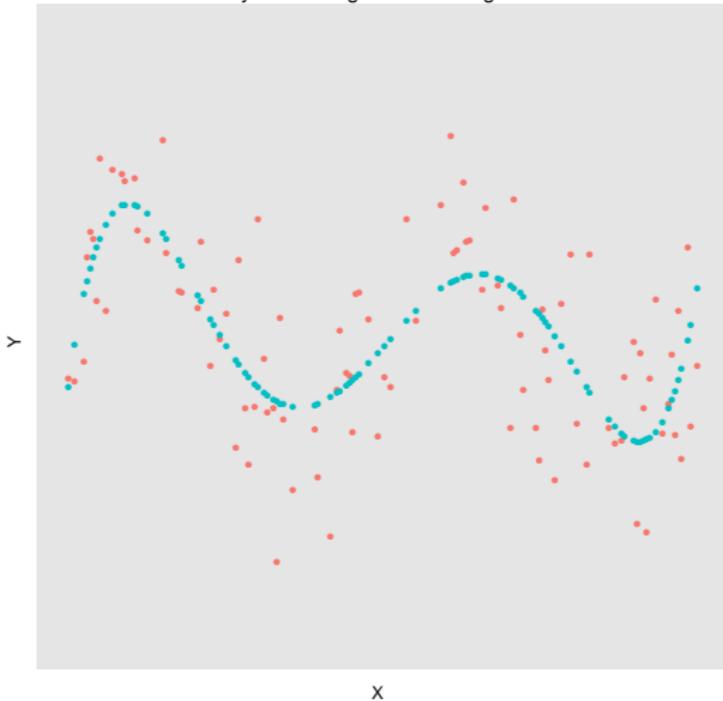
Polynomial Regression of Degree 3



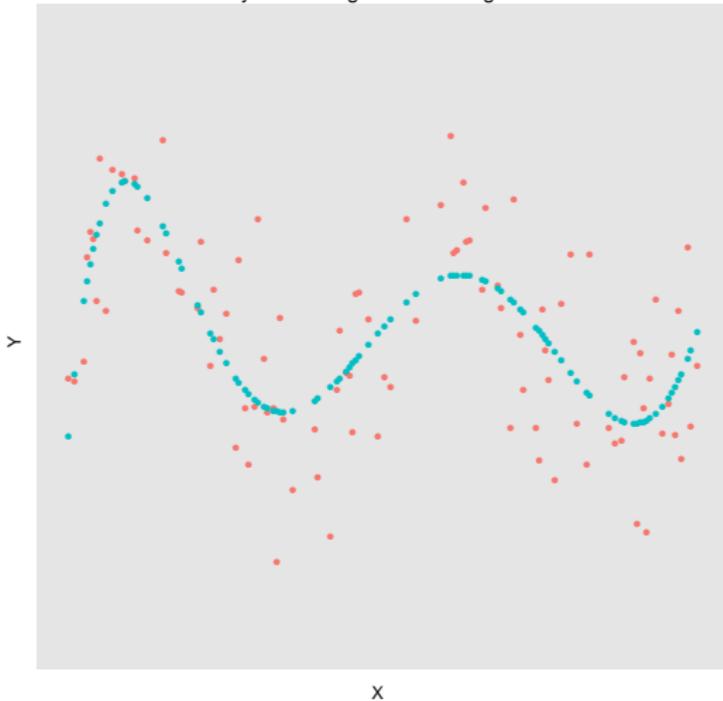
Polynomial Regression of Degree 4



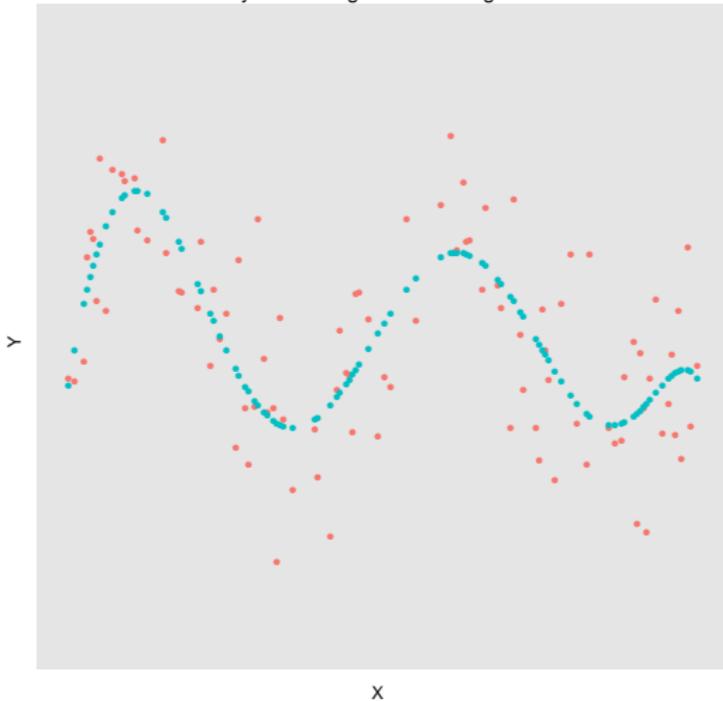
Polynomial Regression of Degree 5



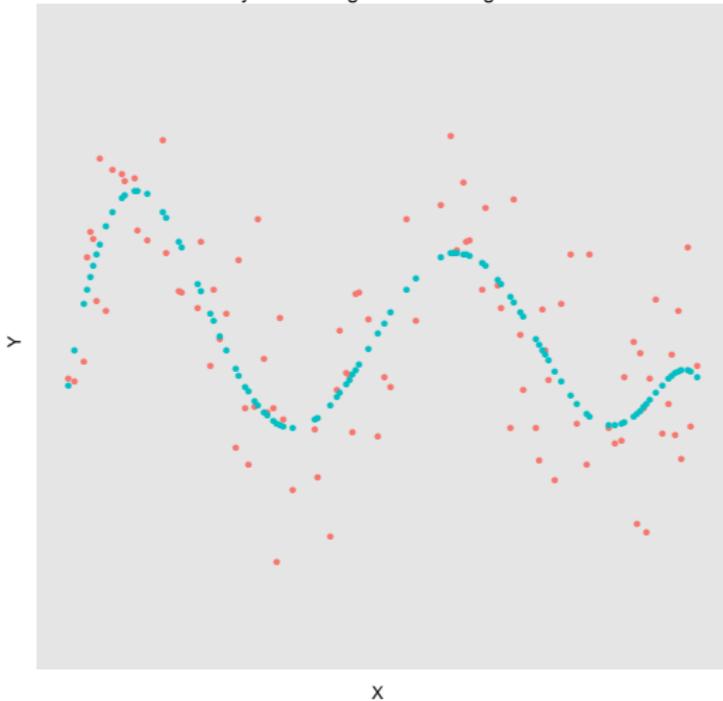
Polynomial Regression of Degree 6



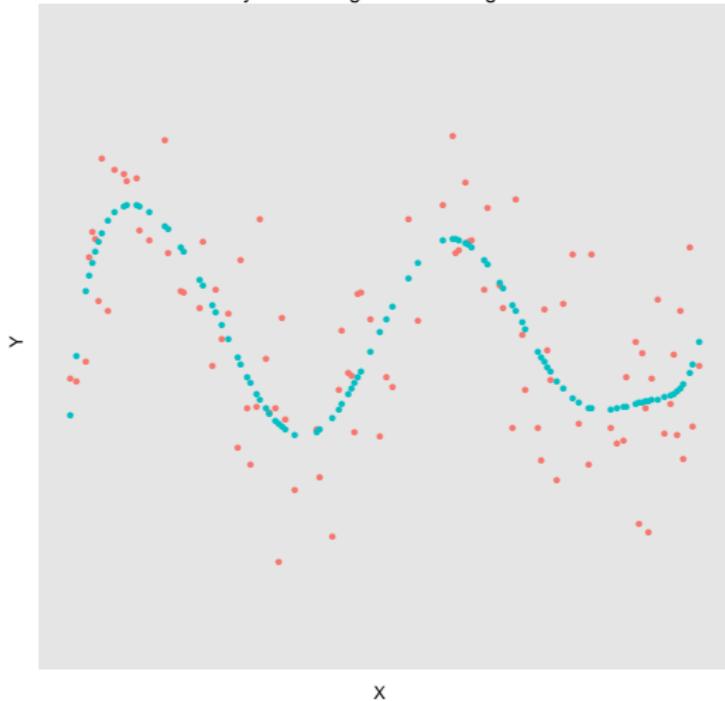
Polynomial Regression of Degree 7



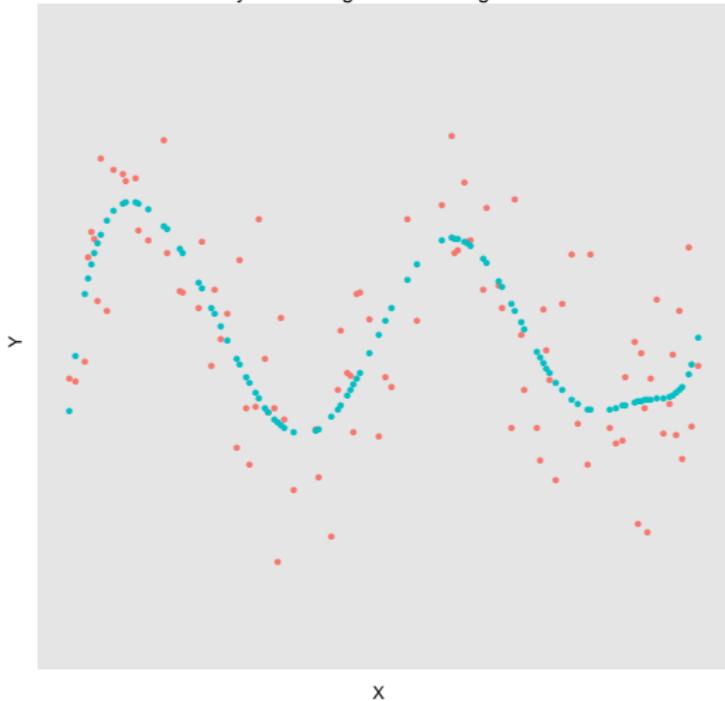
Polynomial Regression of Degree 8



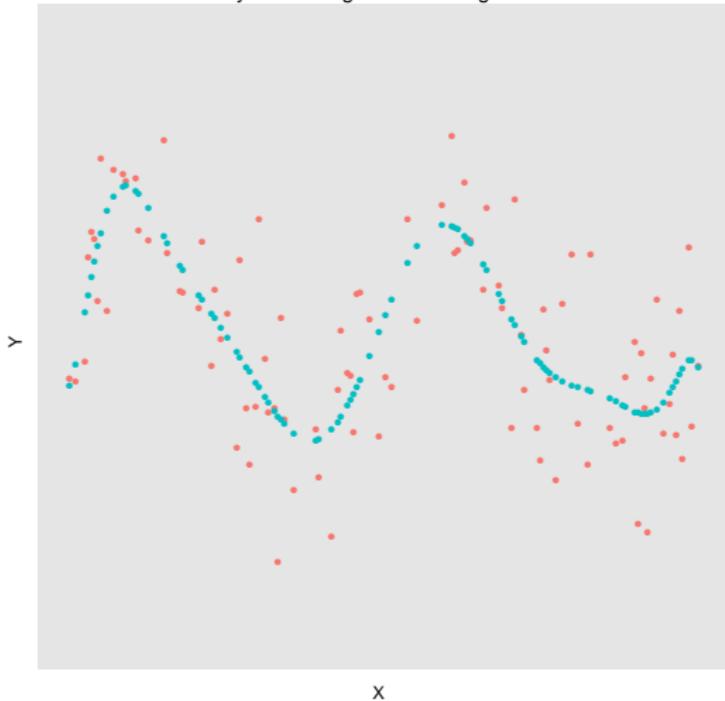
Polynomial Regression of Degree 9



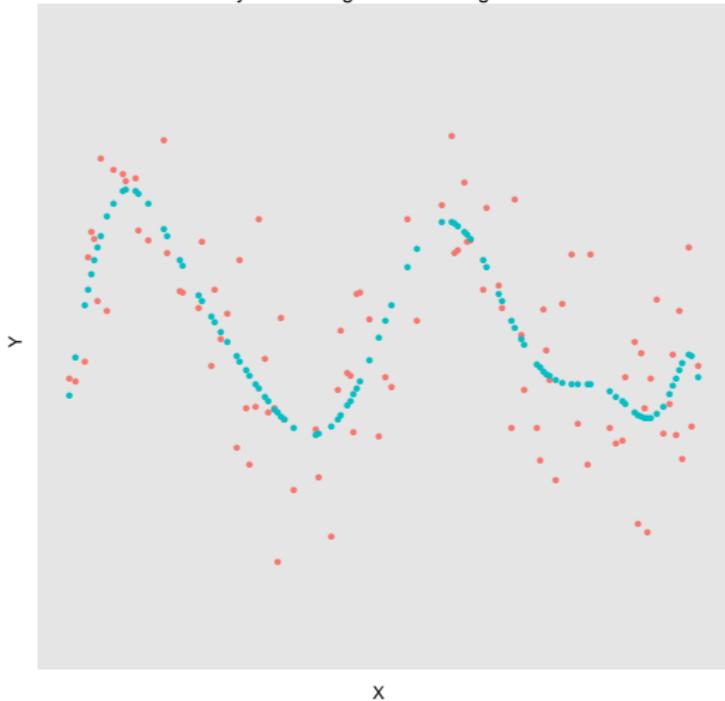
Polynomial Regression of Degree 10



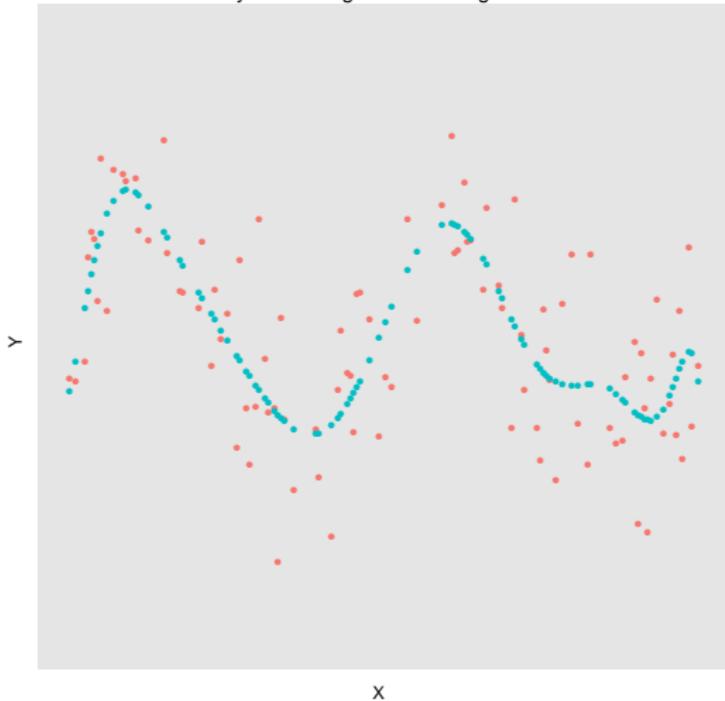
Polynomial Regression of Degree 11



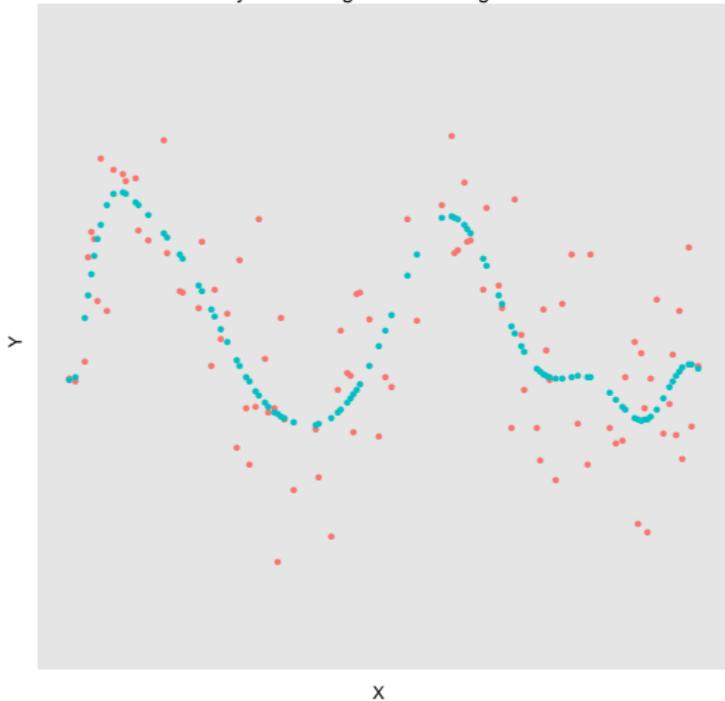
Polynomial Regression of Degree 12



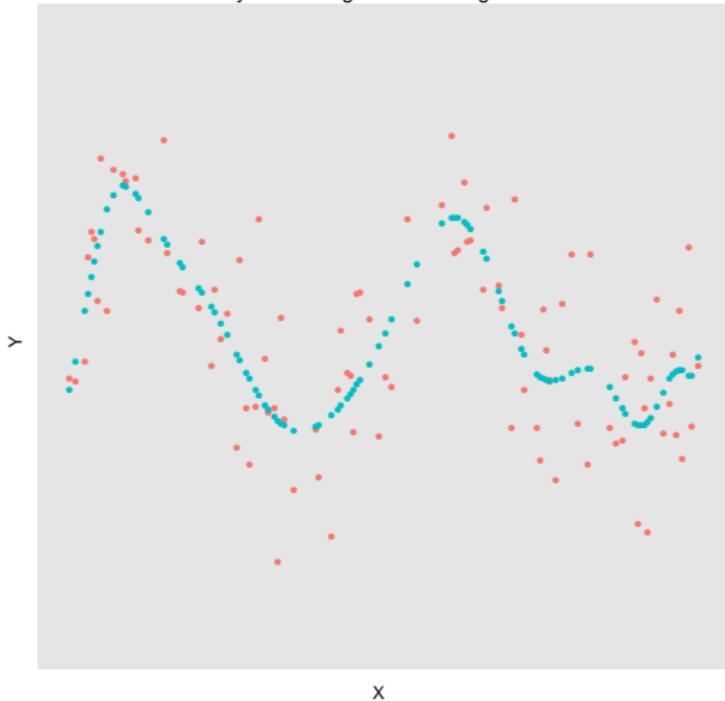
Polynomial Regression of Degree 13



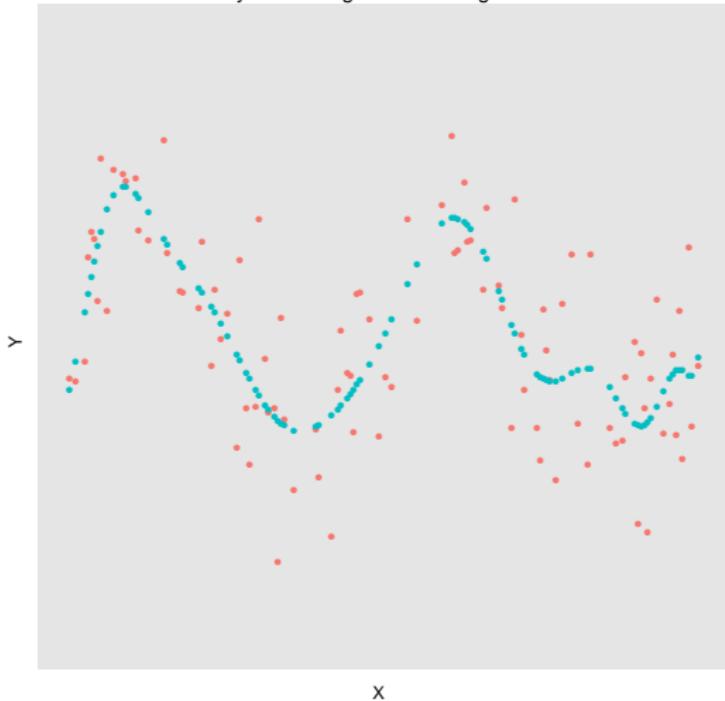
Polynomial Regression of Degree 14



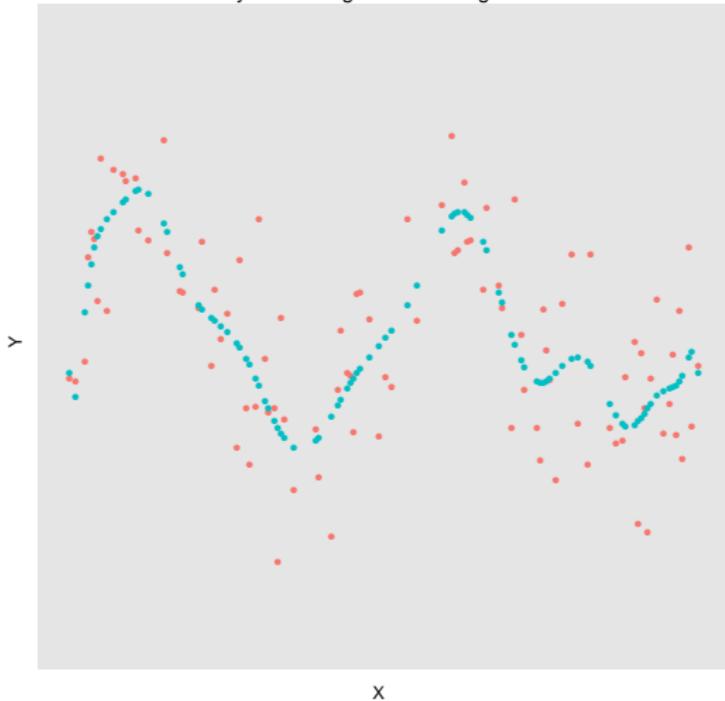
Polynomial Regression of Degree 15



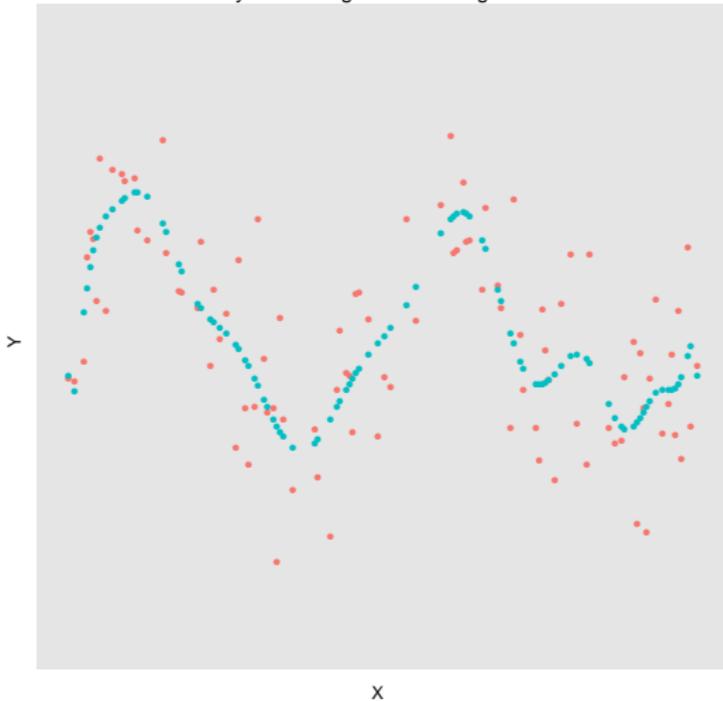
Polynomial Regression of Degree 16



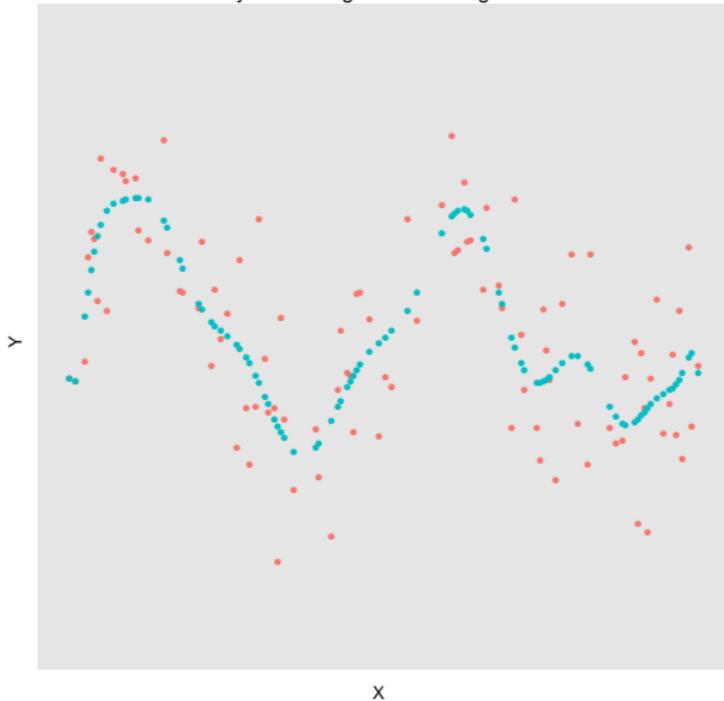
Polynomial Regression of Degree 17



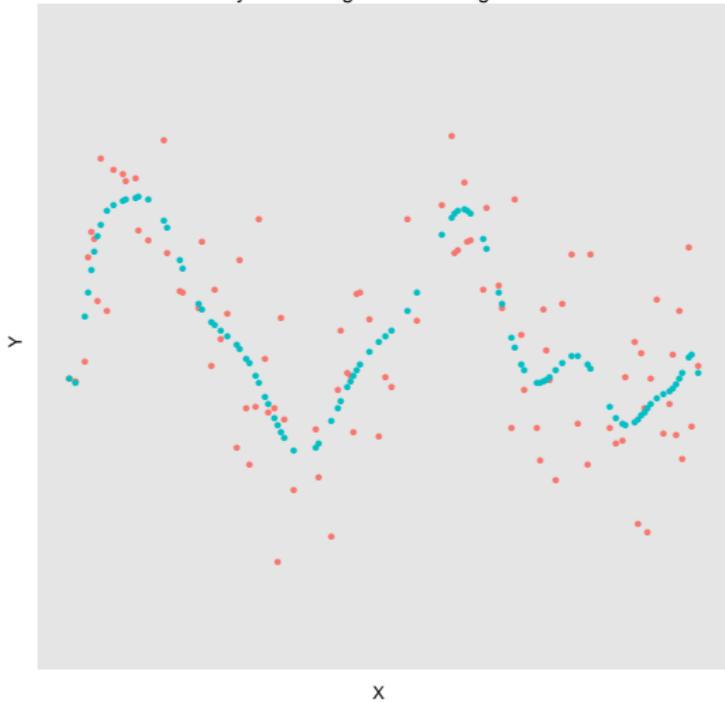
Polynomial Regression of Degree 18



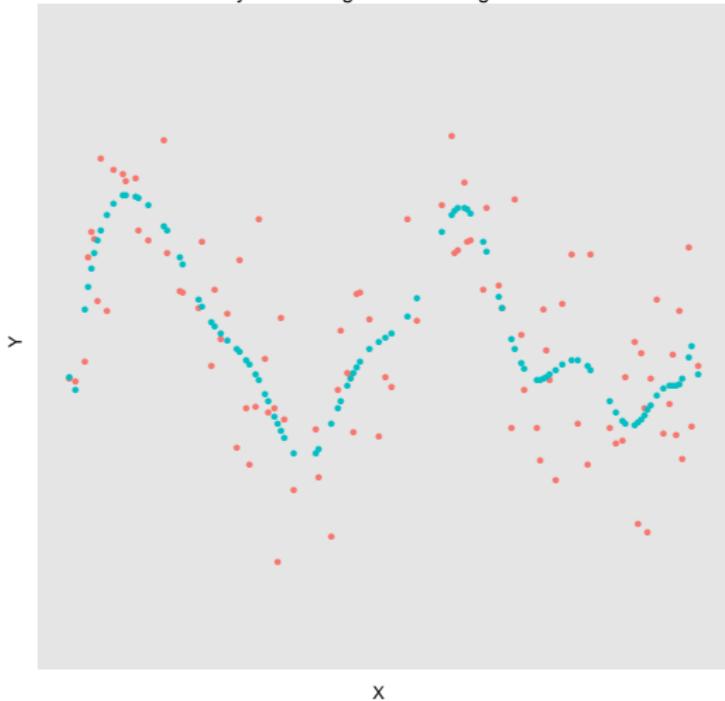
Polynomial Regression of Degree 19



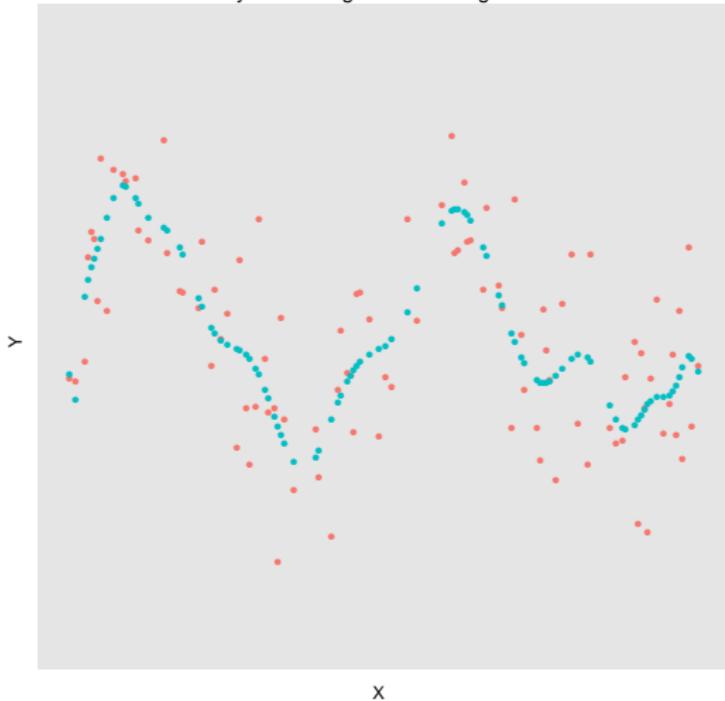
Polynomial Regression of Degree 20



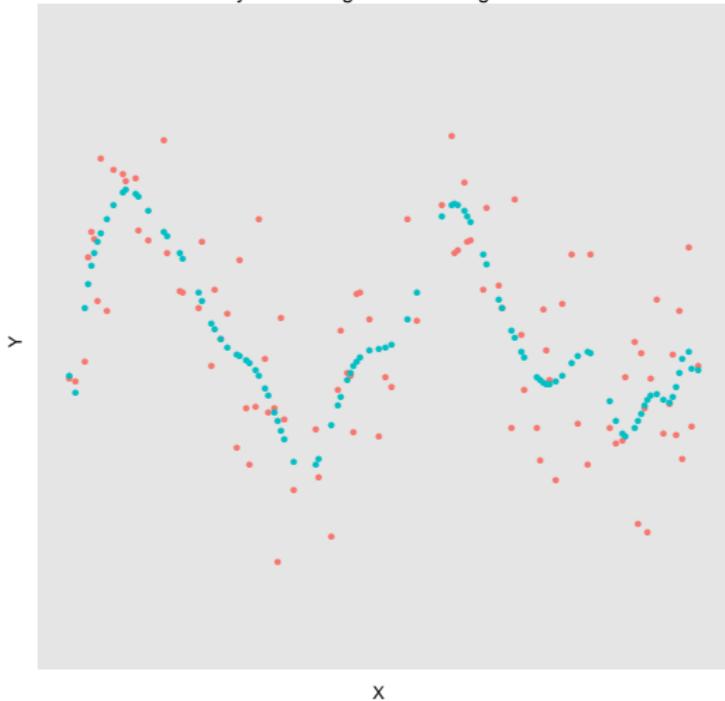
Polynomial Regression of Degree 21



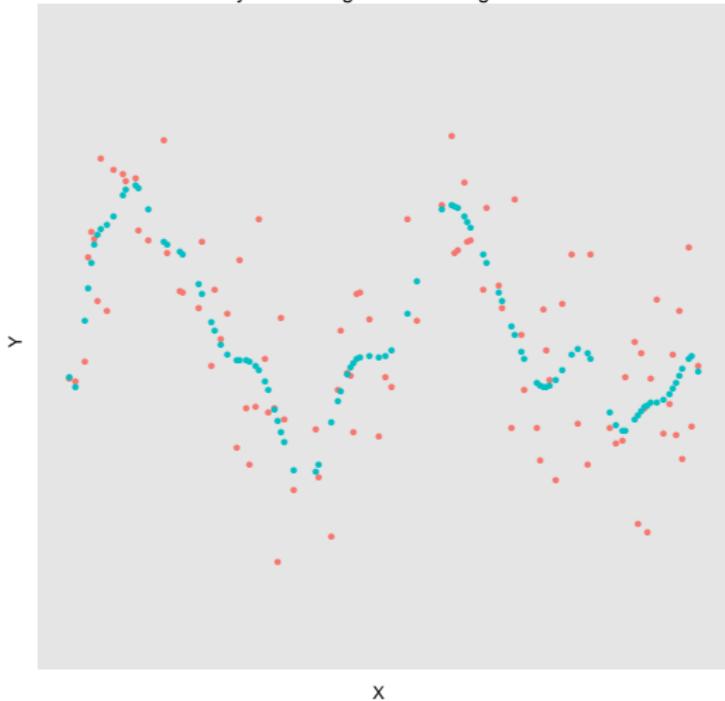
Polynomial Regression of Degree 22



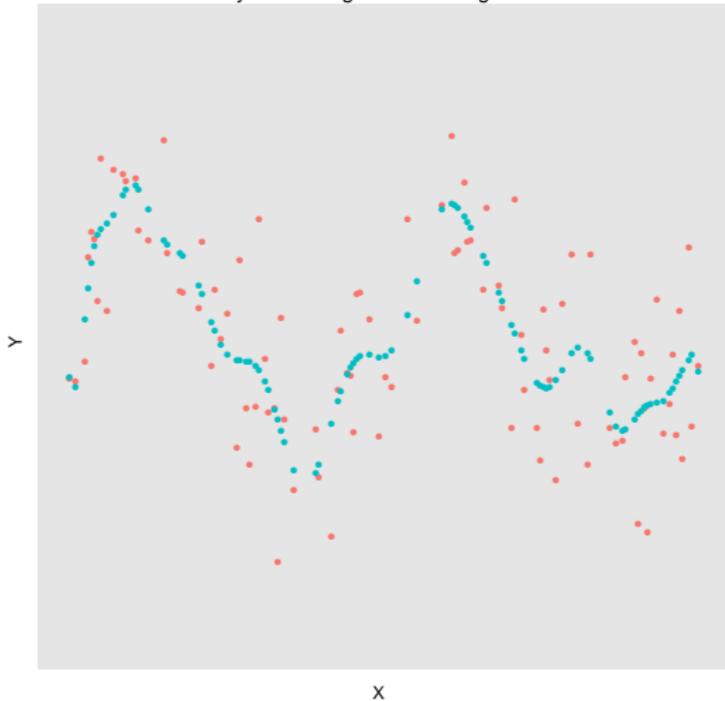
Polynomial Regression of Degree 23



Polynomial Regression of Degree 24



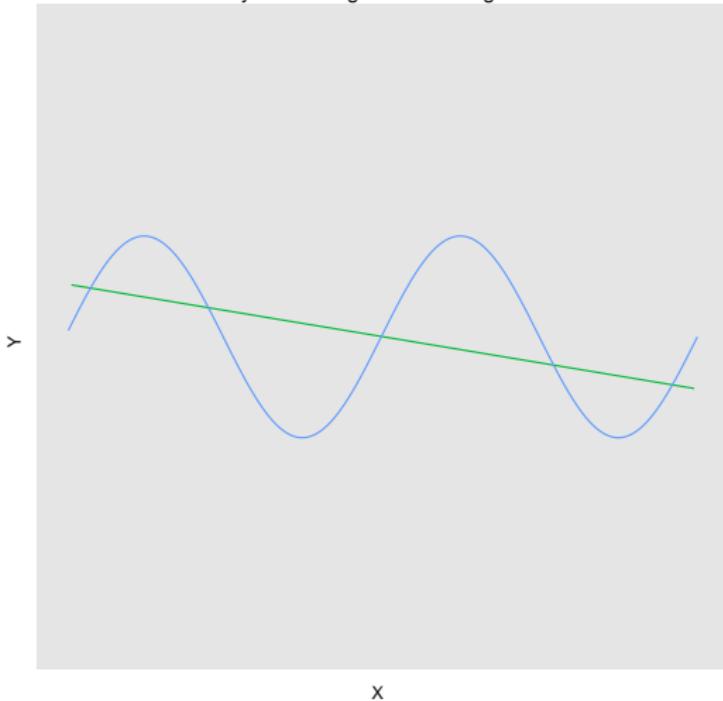
Polynomial Regression of Degree 25



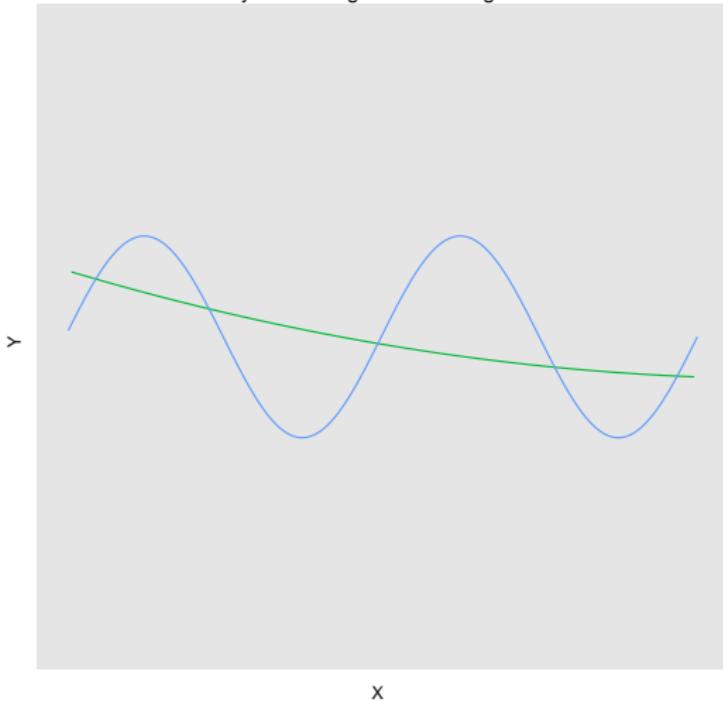
As $D \rightarrow N$, we fit the data better and better.

But our model gets further and further from the true pattern...

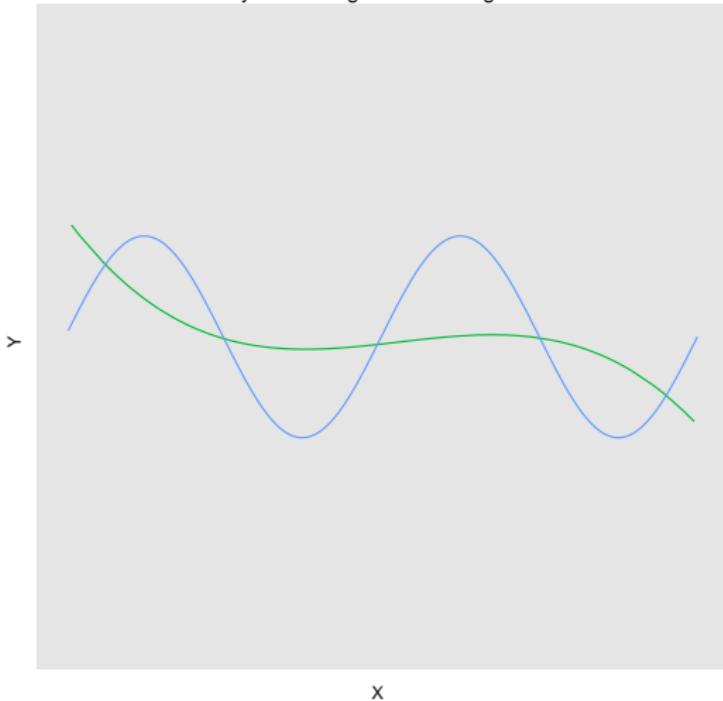
Polynomial Regression of Degree 1



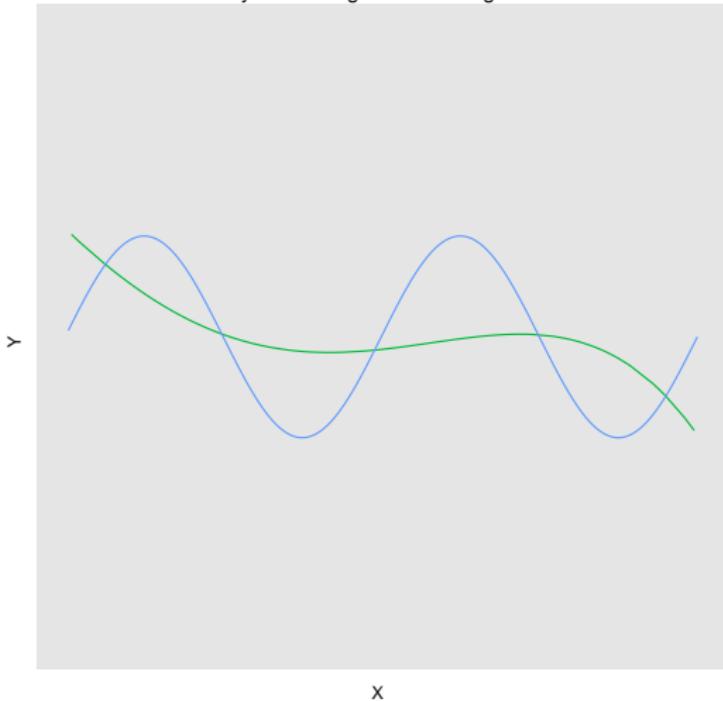
Polynomial Regression of Degree 2



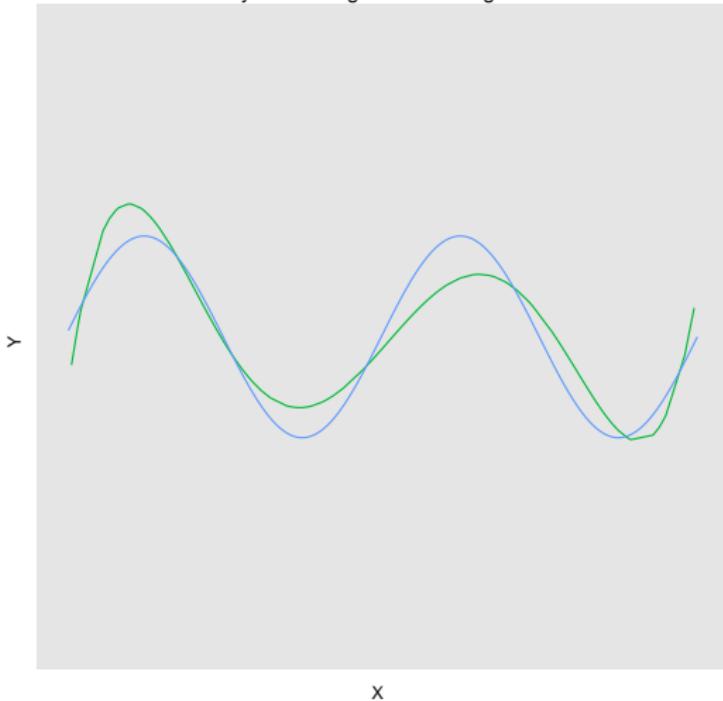
Polynomial Regression of Degree 3



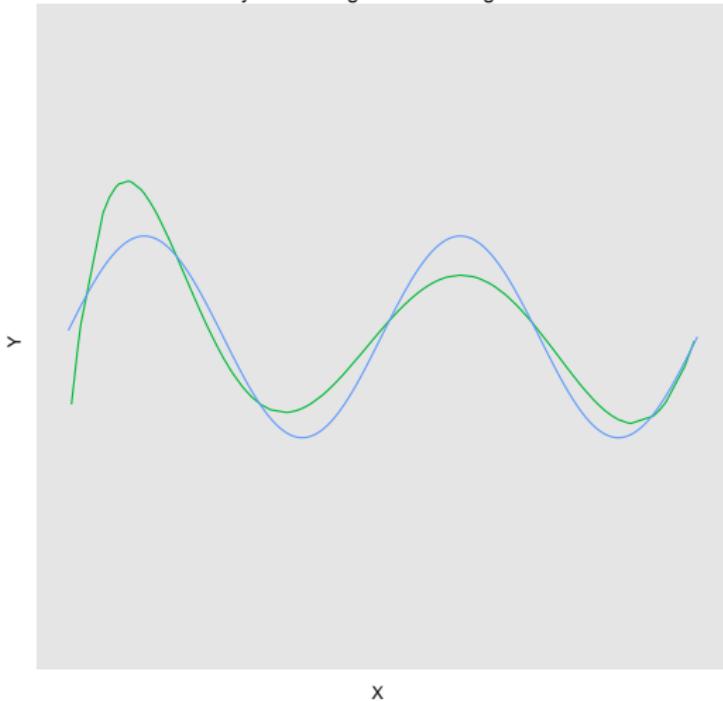
Polynomial Regression of Degree 4



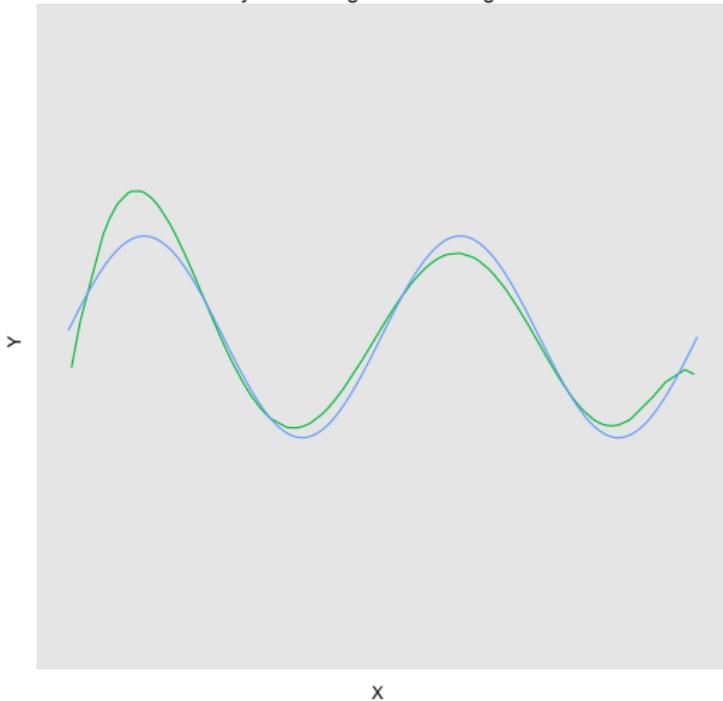
Polynomial Regression of Degree 5



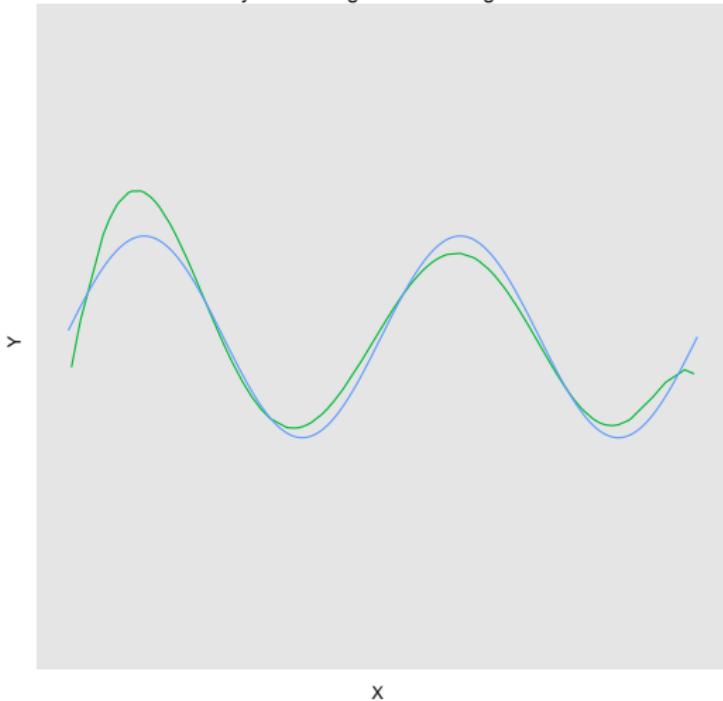
Polynomial Regression of Degree 6



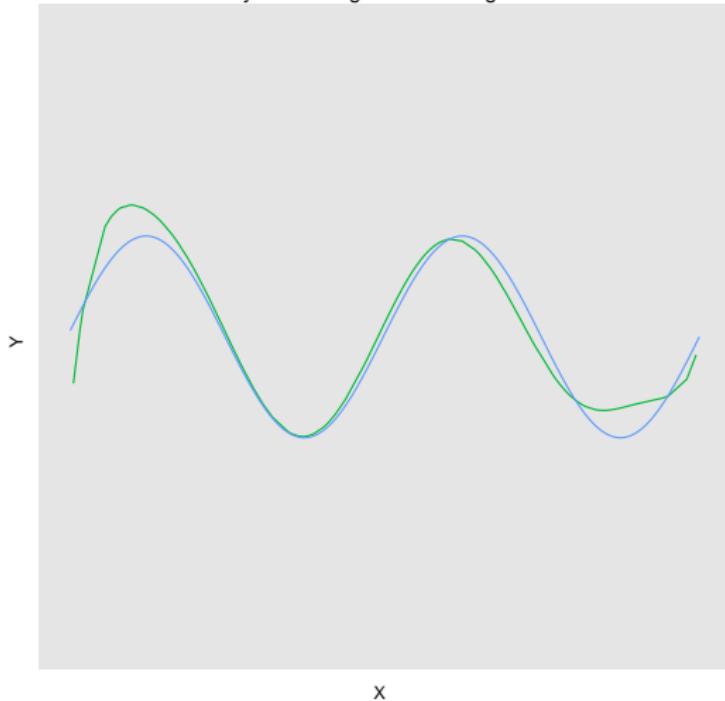
Polynomial Regression of Degree 7



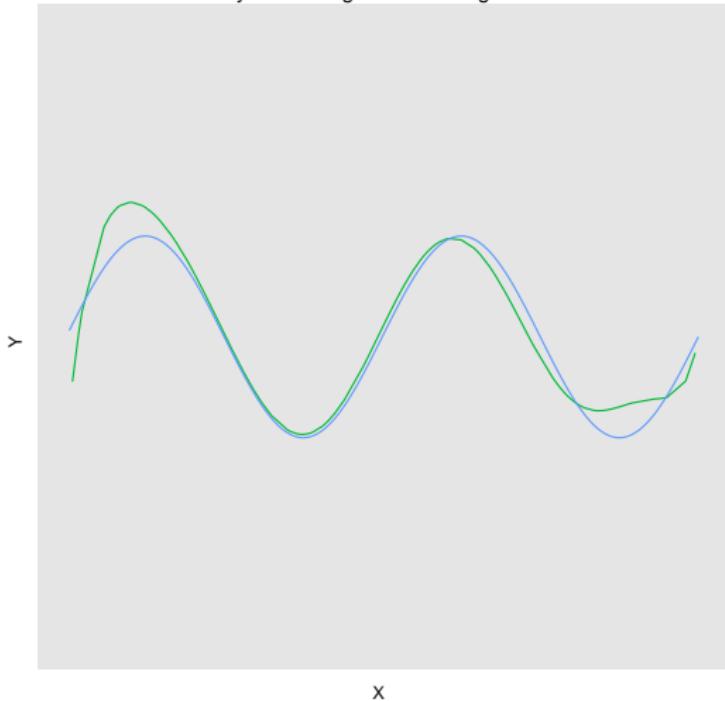
Polynomial Regression of Degree 8



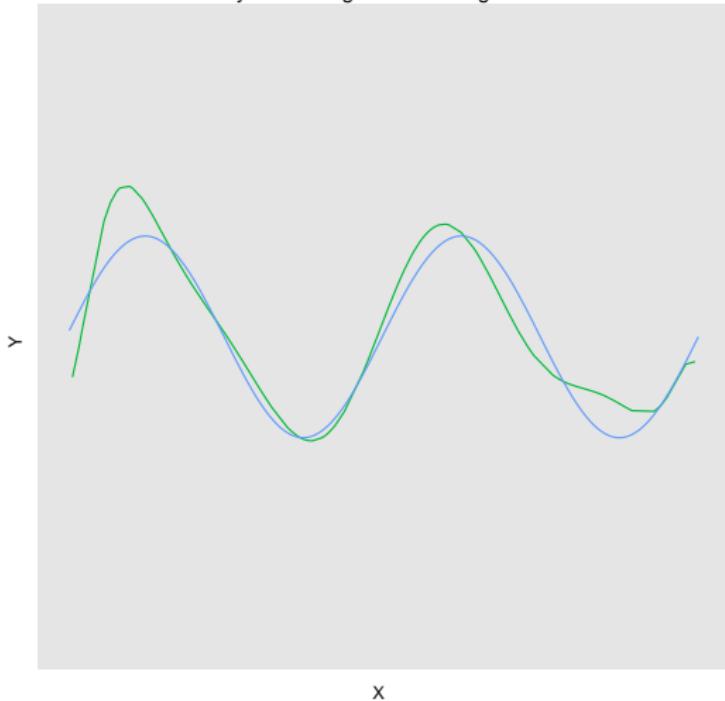
Polynomial Regression of Degree 9



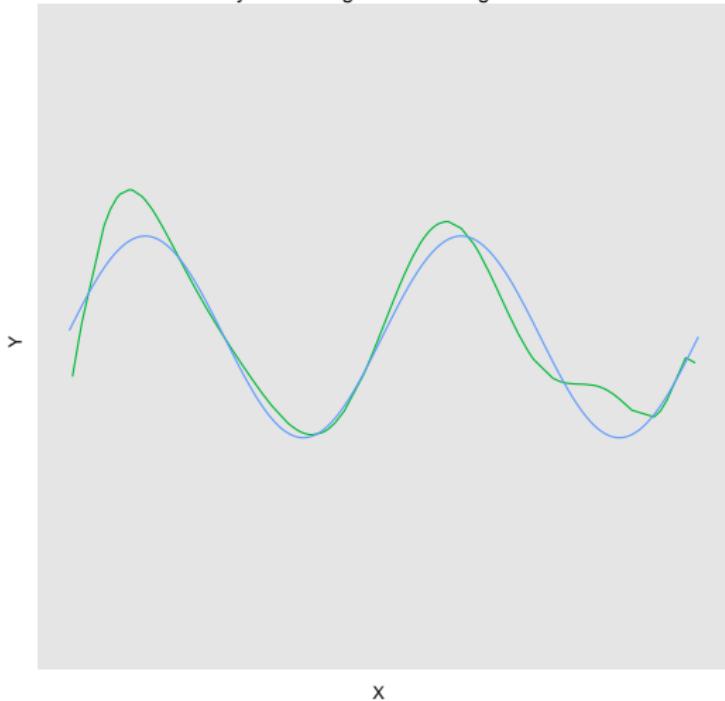
Polynomial Regression of Degree 10



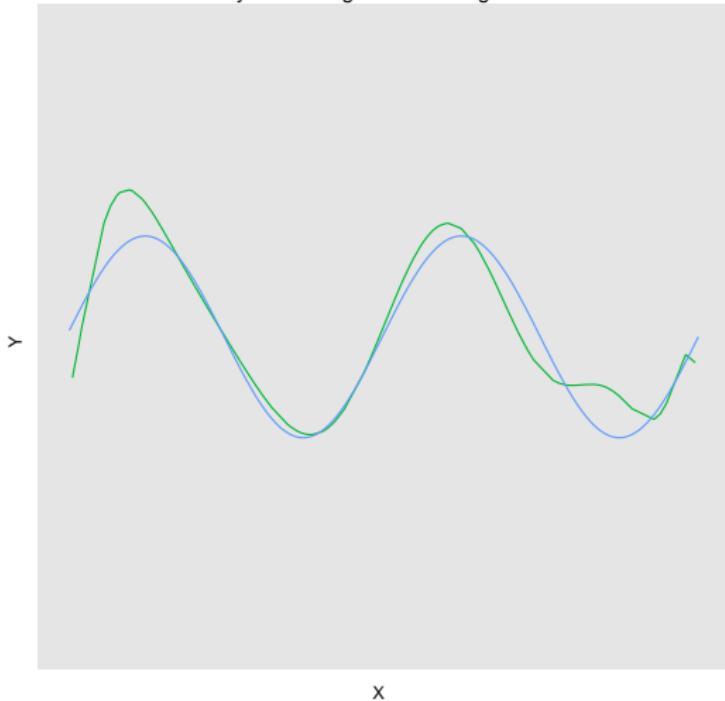
Polynomial Regression of Degree 11



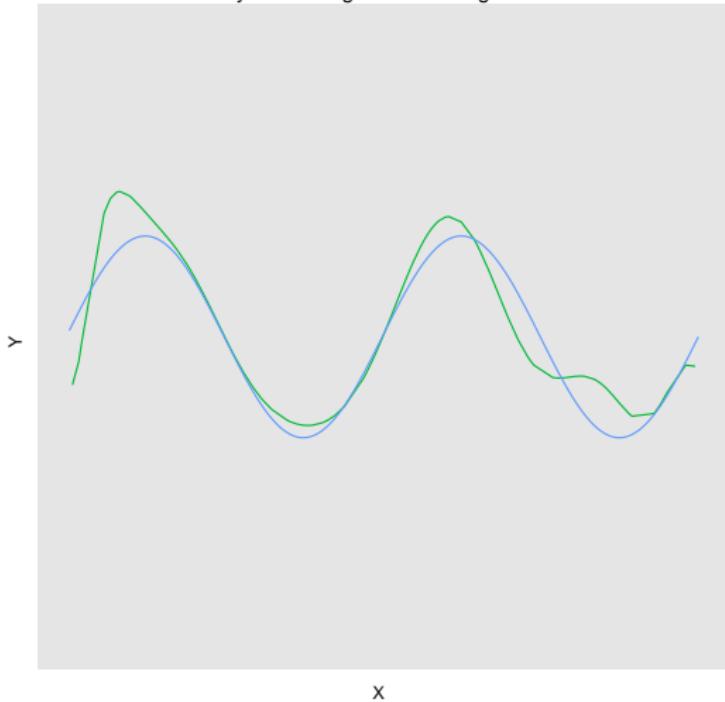
Polynomial Regression of Degree 12



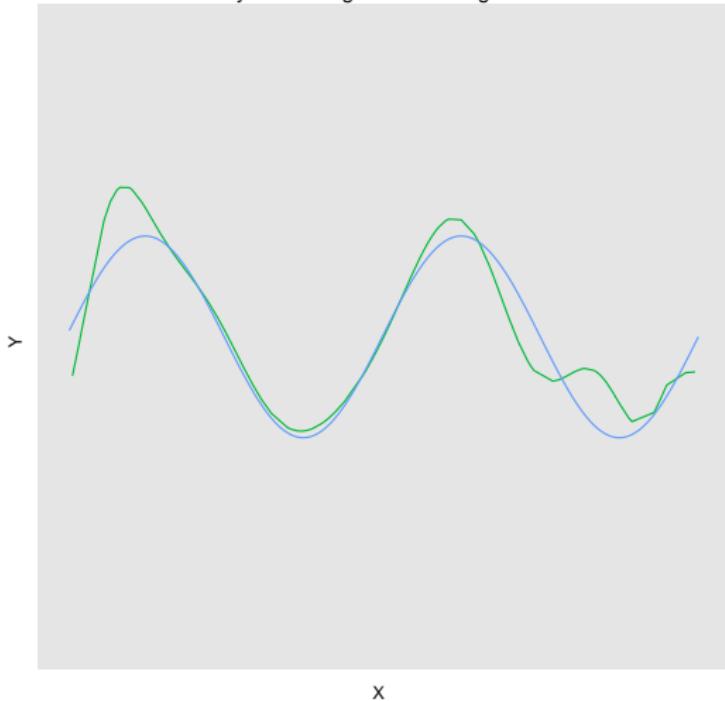
Polynomial Regression of Degree 13



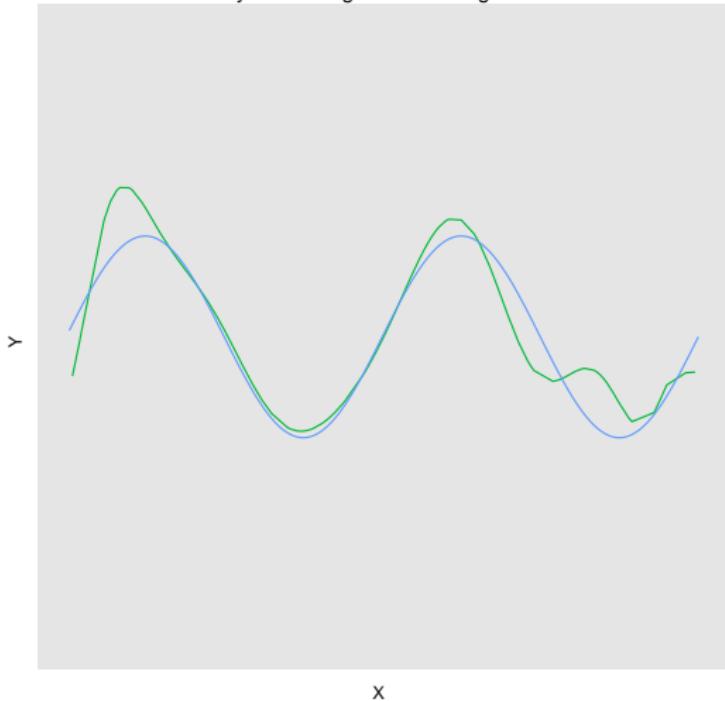
Polynomial Regression of Degree 14



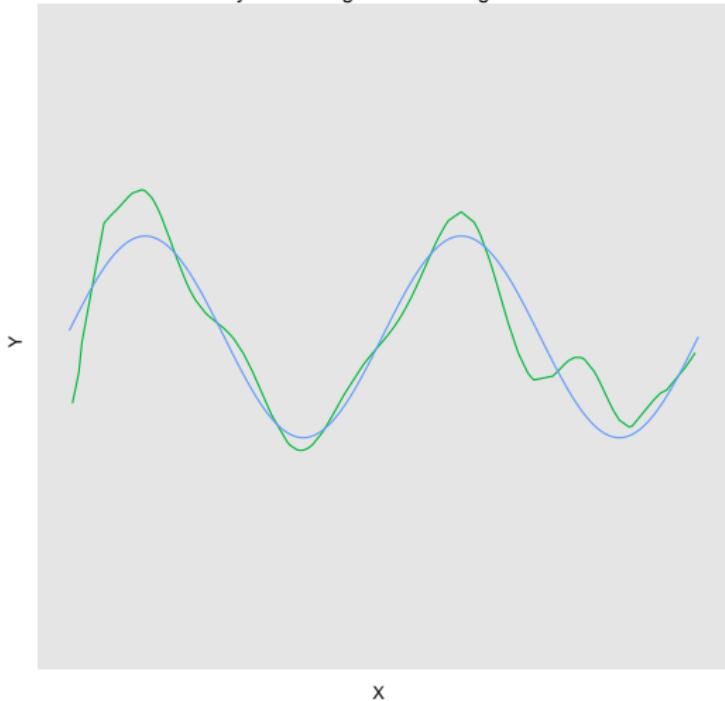
Polynomial Regression of Degree 15



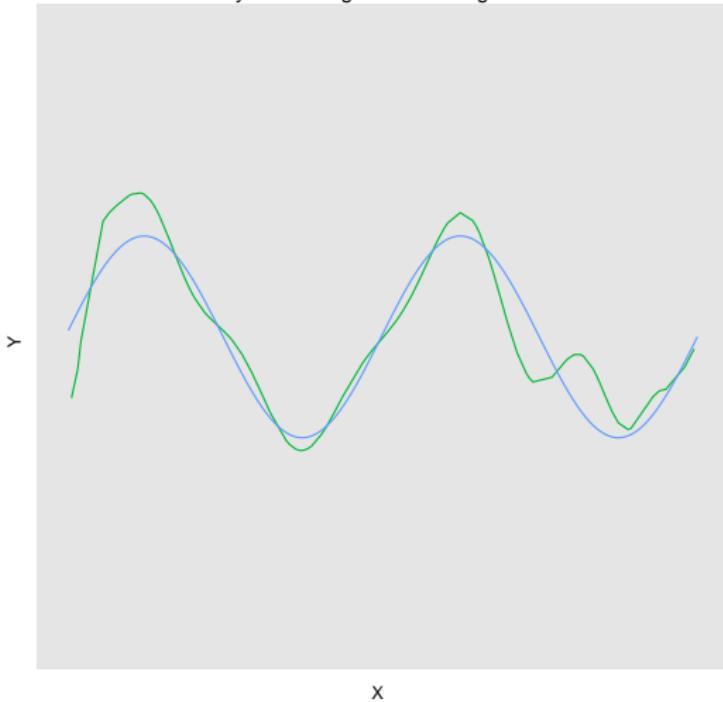
Polynomial Regression of Degree 16



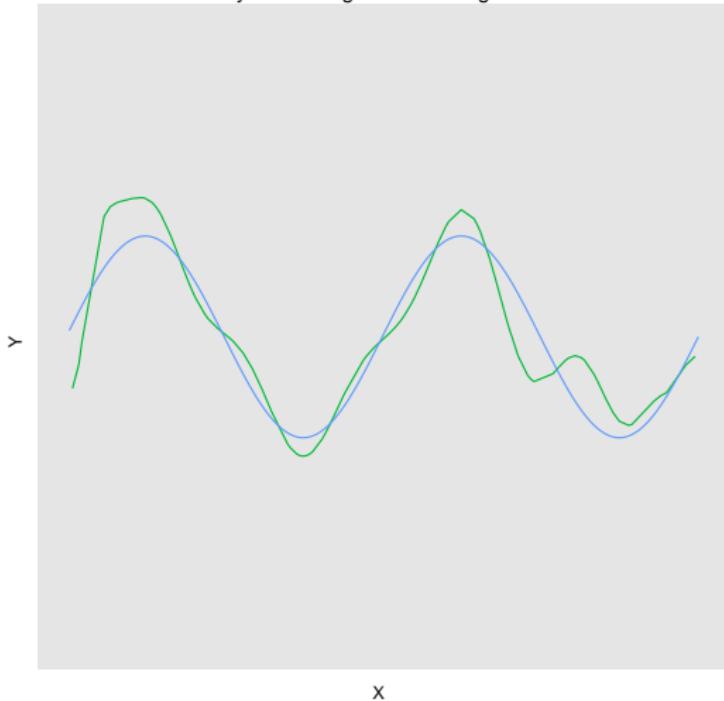
Polynomial Regression of Degree 17



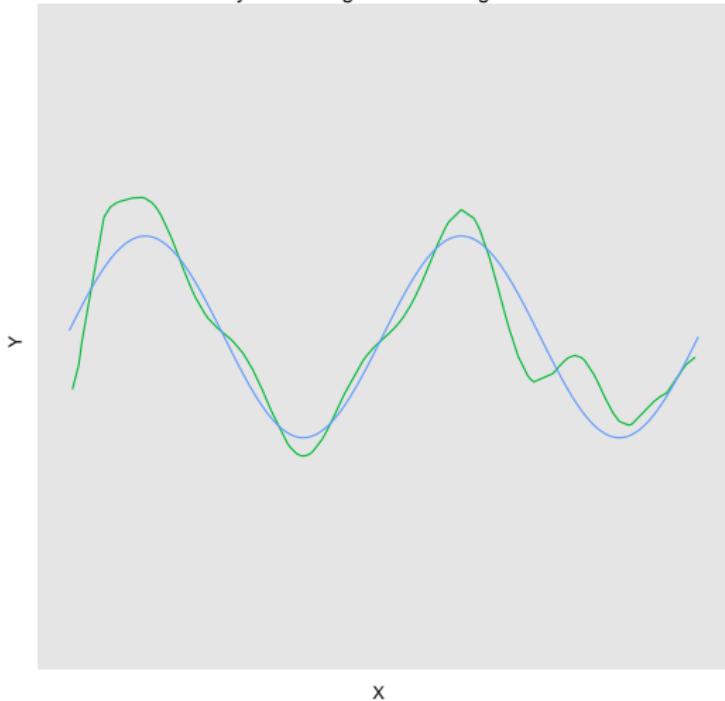
Polynomial Regression of Degree 18



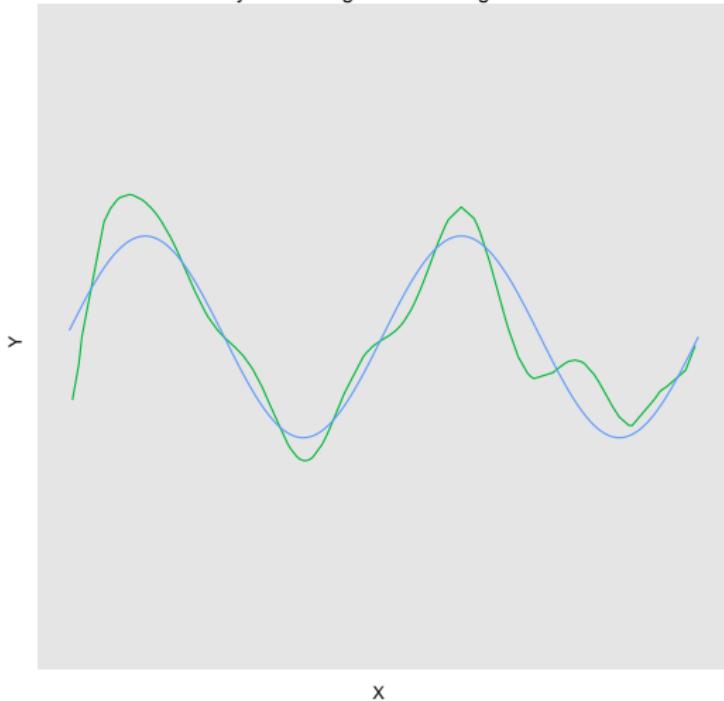
Polynomial Regression of Degree 19



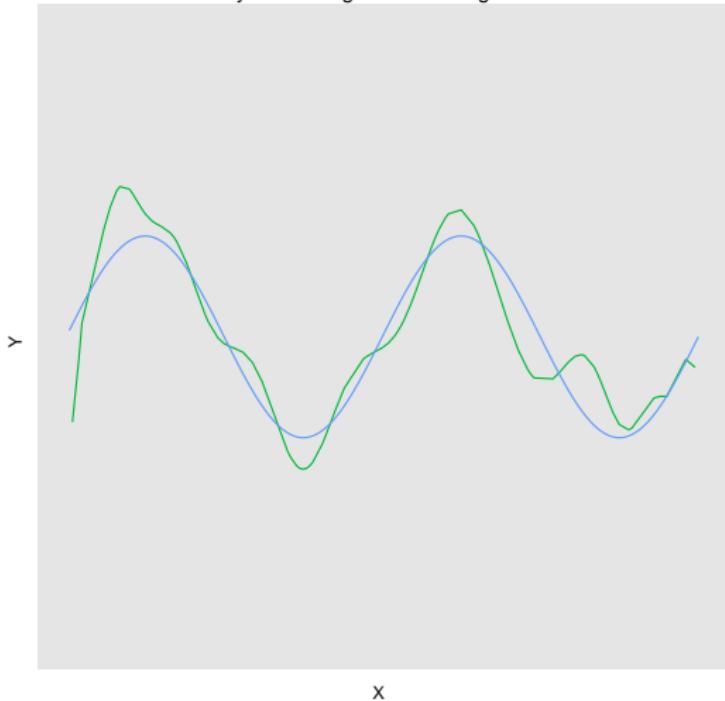
Polynomial Regression of Degree 20



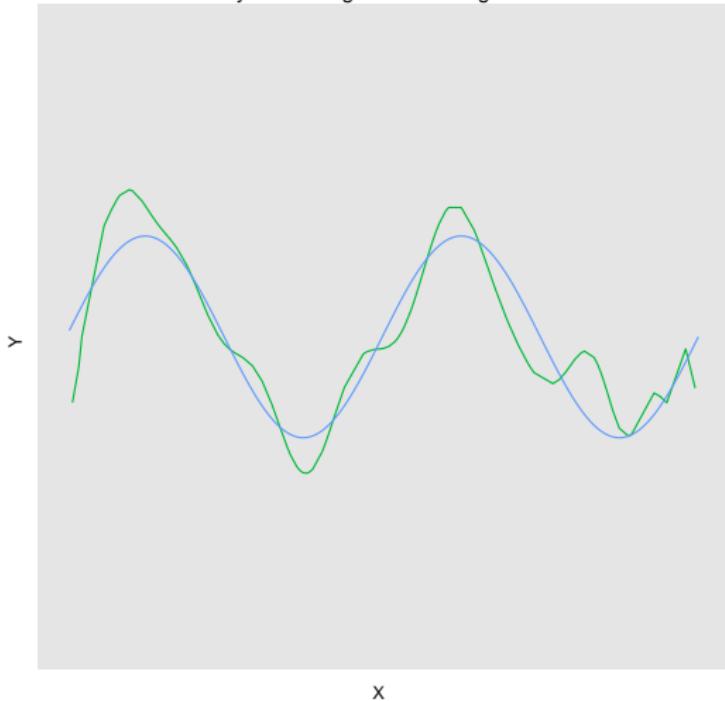
Polynomial Regression of Degree 21



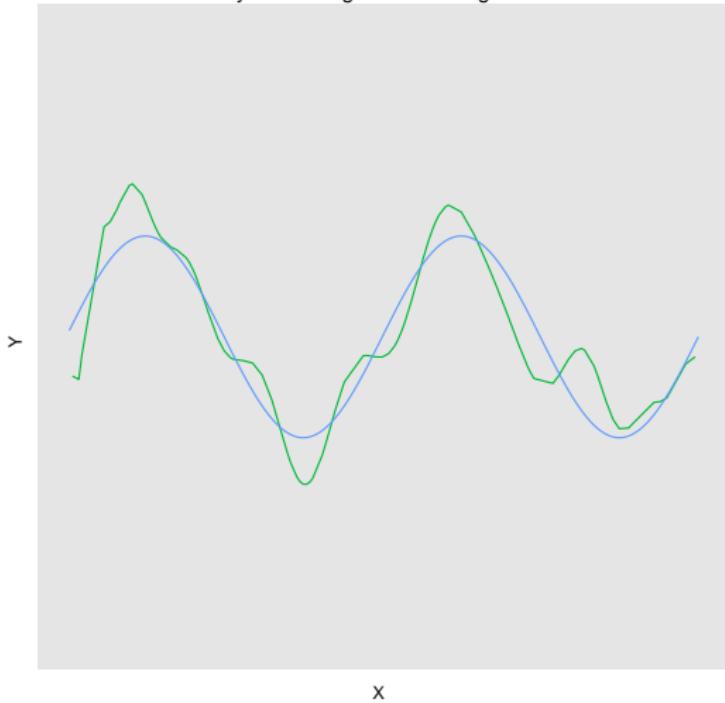
Polynomial Regression of Degree 22



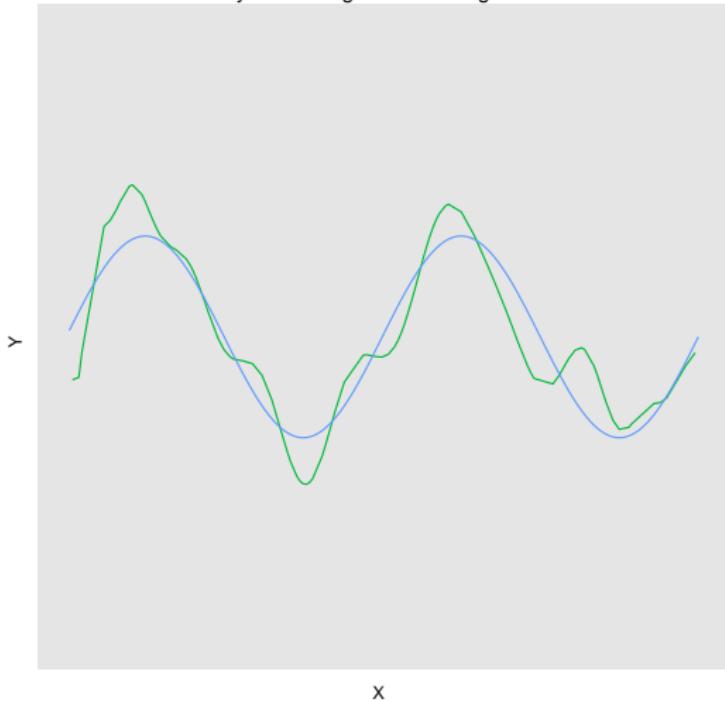
Polynomial Regression of Degree 23



Polynomial Regression of Degree 24

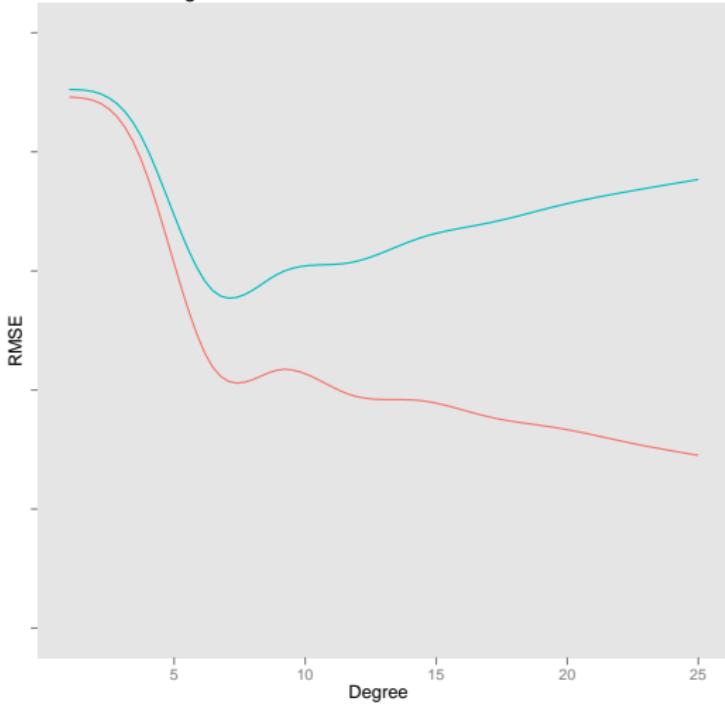


Polynomial Regression of Degree 25



Overfitting always occur when our models become too expressive

Training Set Performance versus Test Set Performance

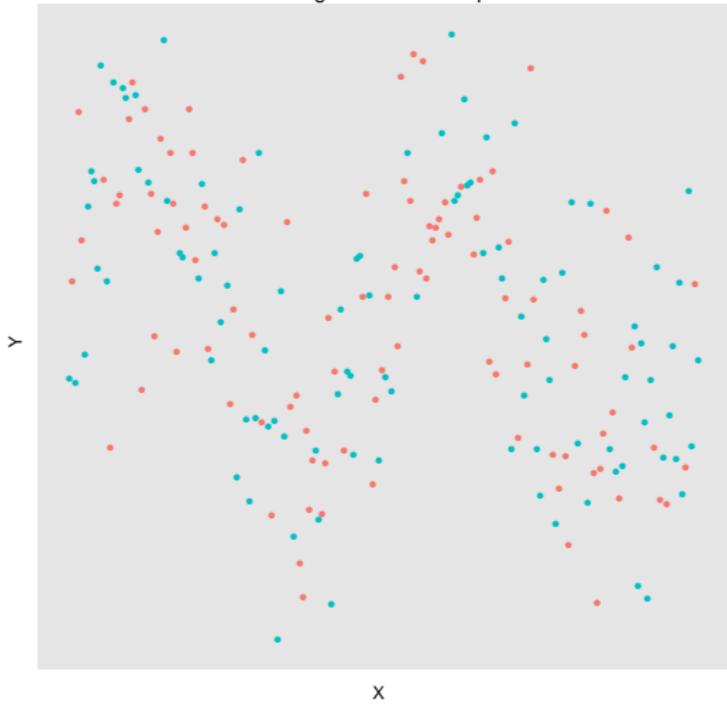


How do we prevent overfitting?

One approach:

- ▶ Split data into training set and test set
- ▶ Fit model to training set
- ▶ Assess performance on test set
- ▶ Pick model that does best on held out test set

Training Set / Test Set Split



This underestimates the strongest model we can safely use.

Another approach:

- ▶ Regularize our model

Unregularized models minimize prediction error:

- ▶ OLS regression
- ▶ Logistic regression

$$\beta^* = \arg \min_{\beta} (Y - X\beta)^2$$

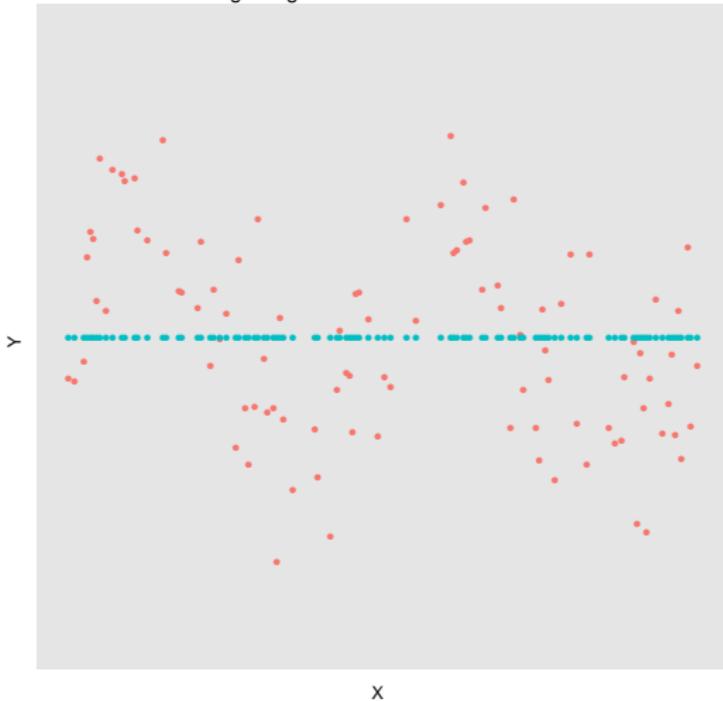
Regularized models minimize prediction error and model size:

- ▶ Ridge regression
- ▶ Lasso regression

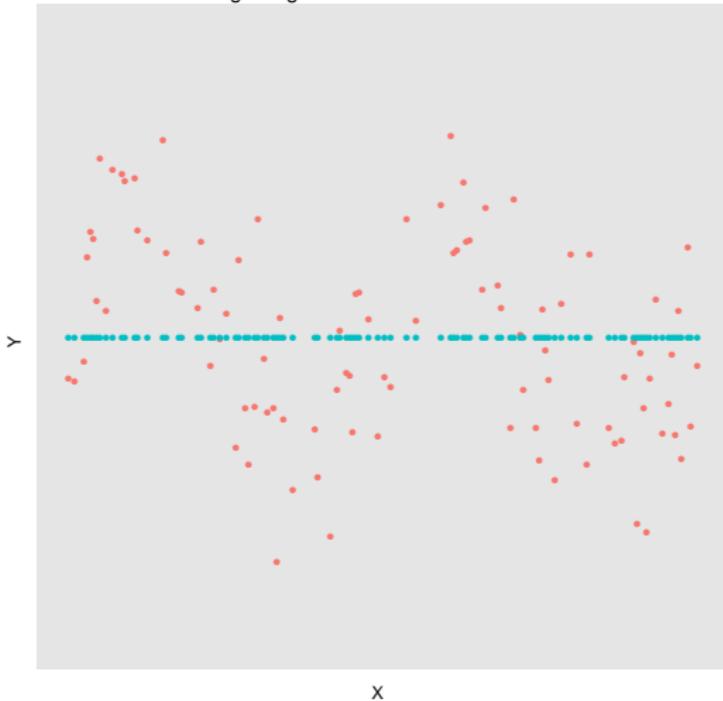
$$\beta^* = \arg \min_{\beta} (Y - X\beta)^2 + \beta^2$$

$$\beta^* = \arg \min_{\beta} (Y - X\beta)^2 + |\beta|$$

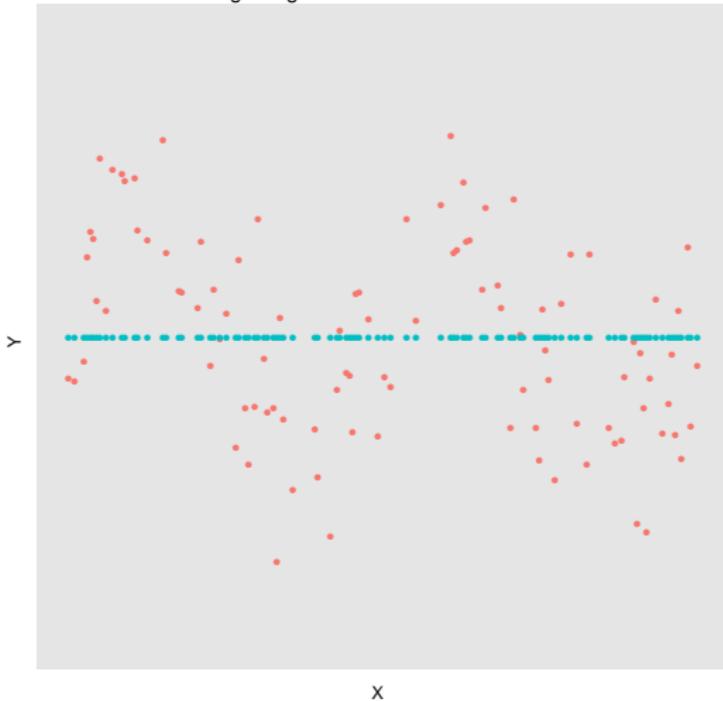
Ridge Regression with Lambda = 524



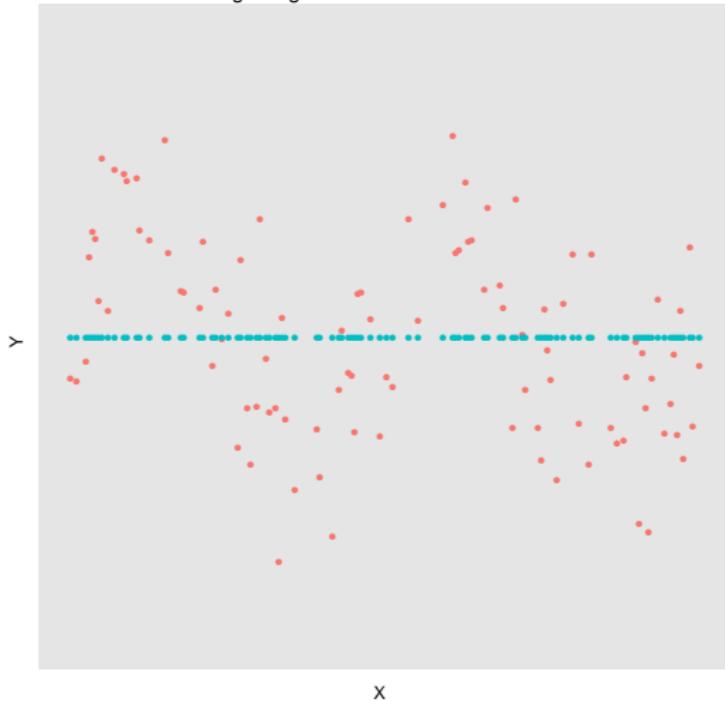
Ridge Regression with Lambda = 478



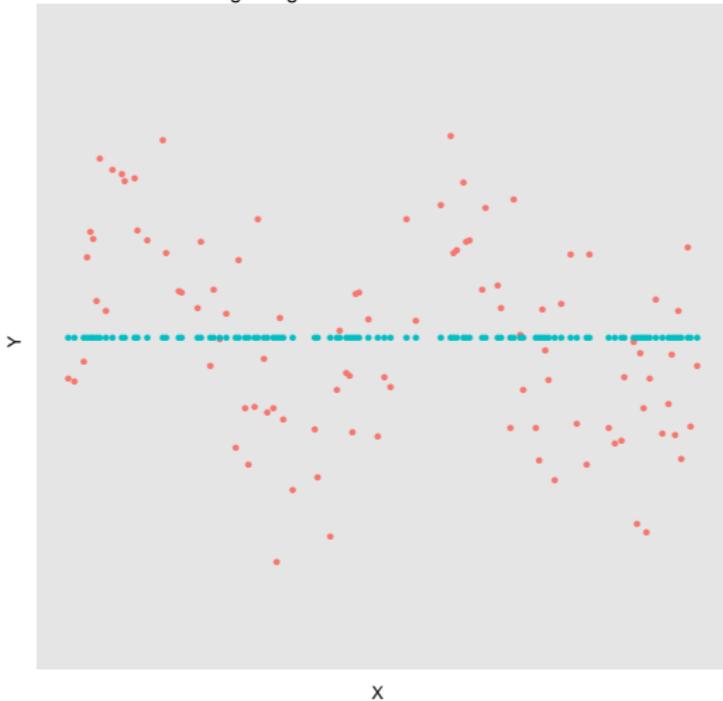
Ridge Regression with Lambda = 435



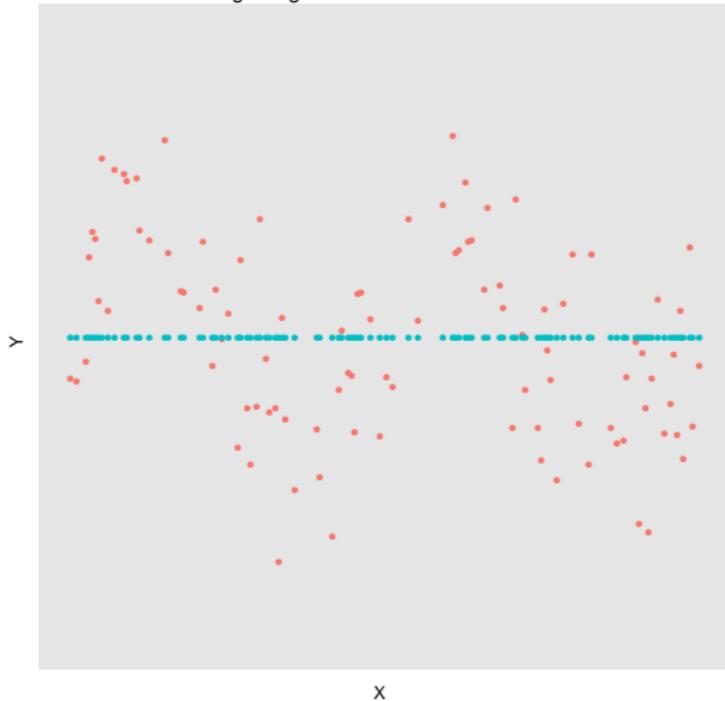
Ridge Regression with Lambda = 396



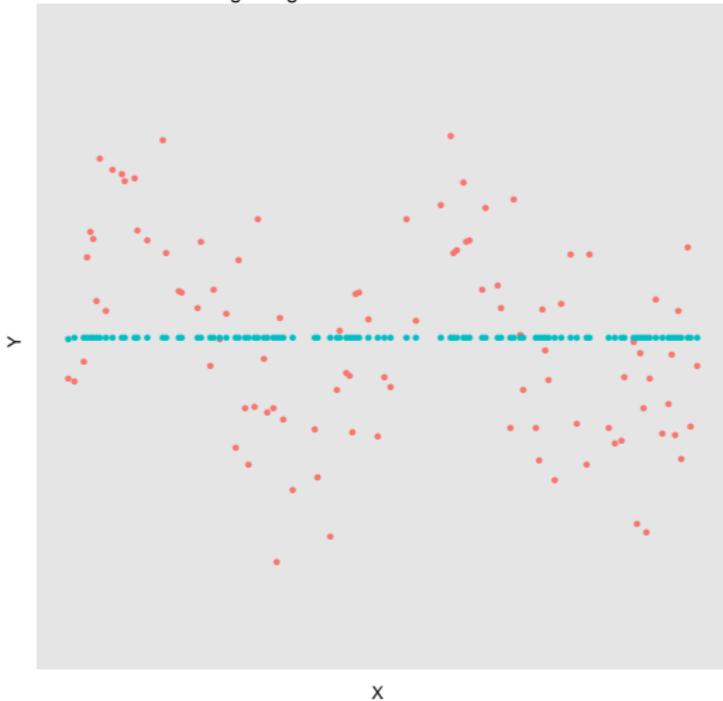
Ridge Regression with Lambda = 361



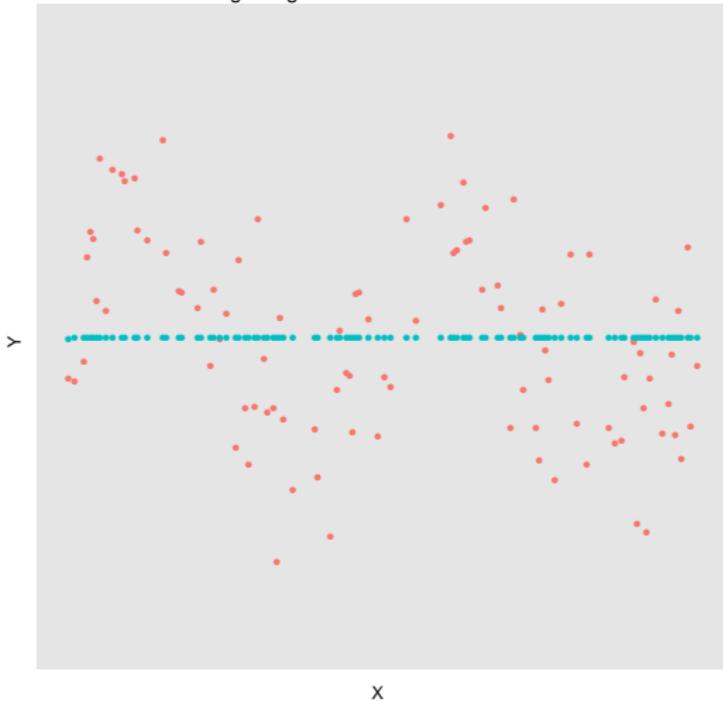
Ridge Regression with Lambda = 329



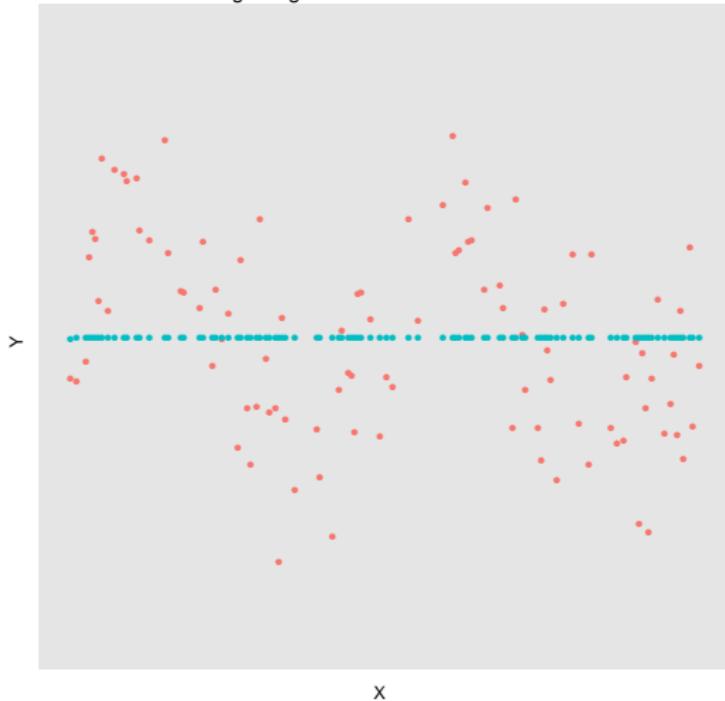
Ridge Regression with Lambda = 300



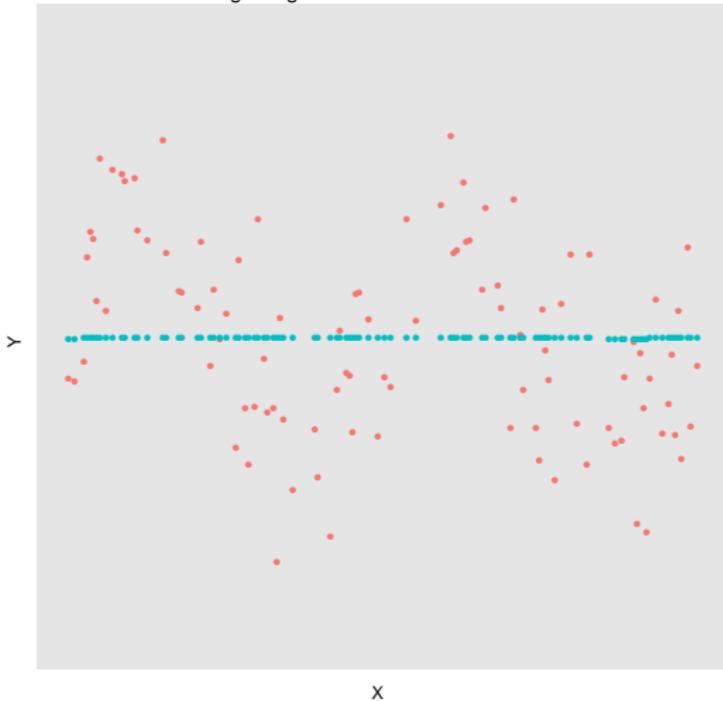
Ridge Regression with Lambda = 273



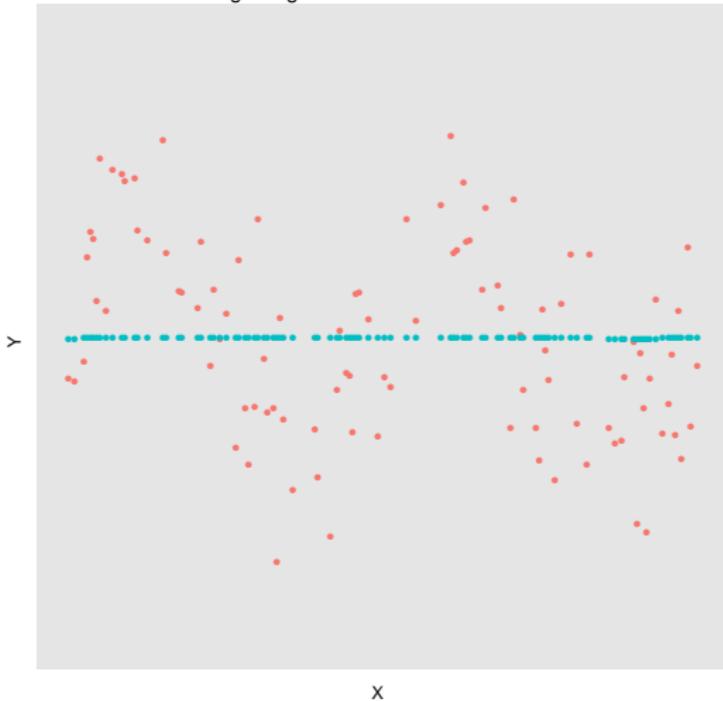
Ridge Regression with Lambda = 249



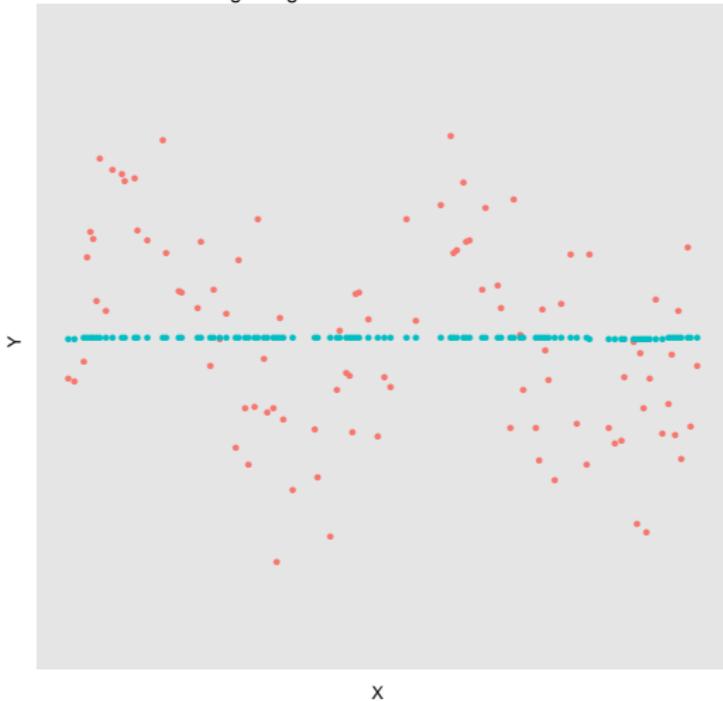
Ridge Regression with Lambda = 227



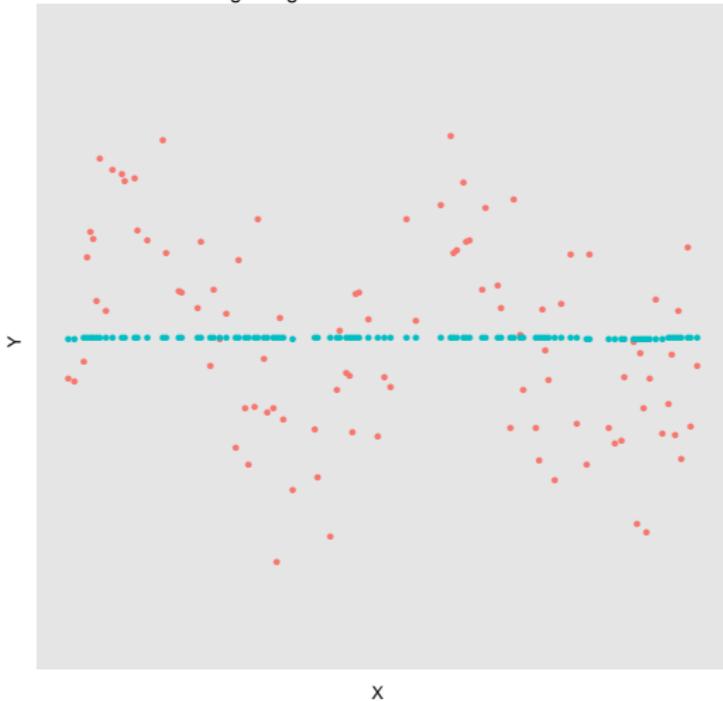
Ridge Regression with Lambda = 207



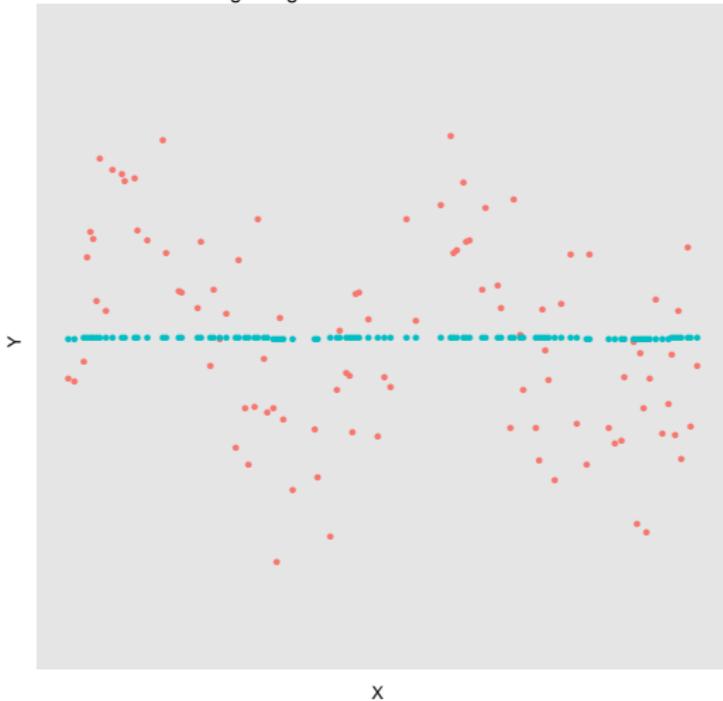
Ridge Regression with Lambda = 188



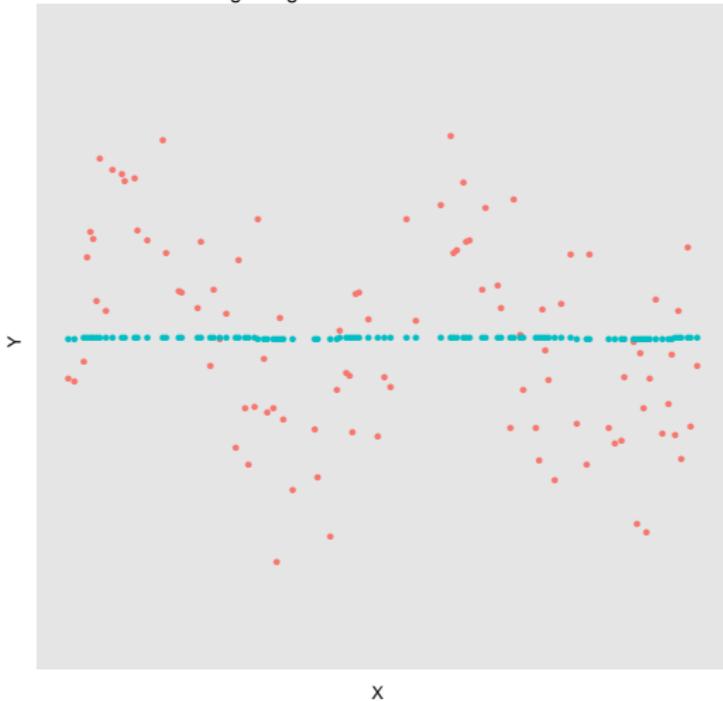
Ridge Regression with Lambda = 172



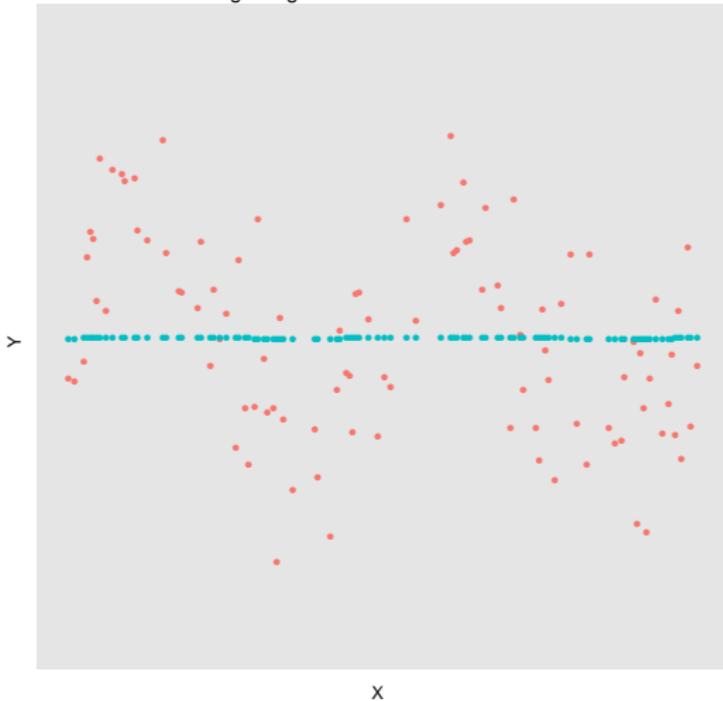
Ridge Regression with Lambda = 156



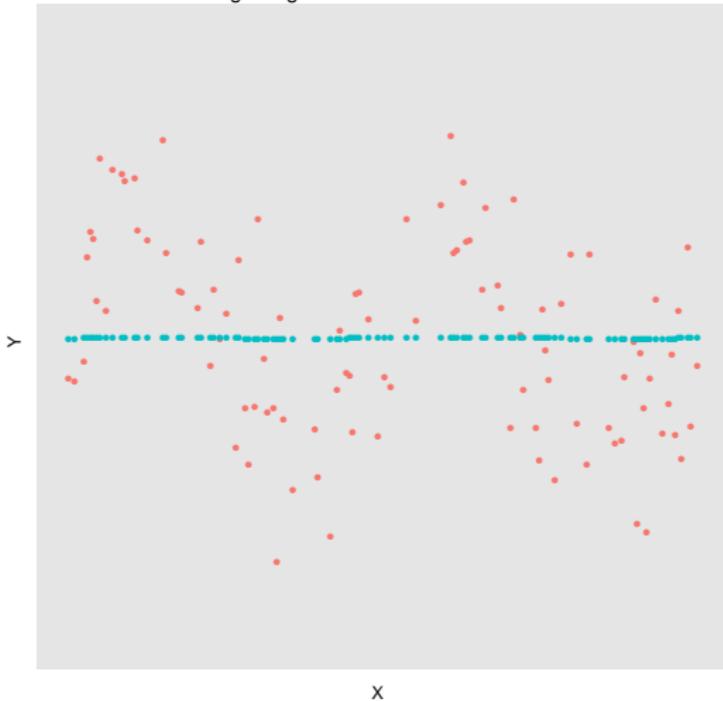
Ridge Regression with Lambda = 142



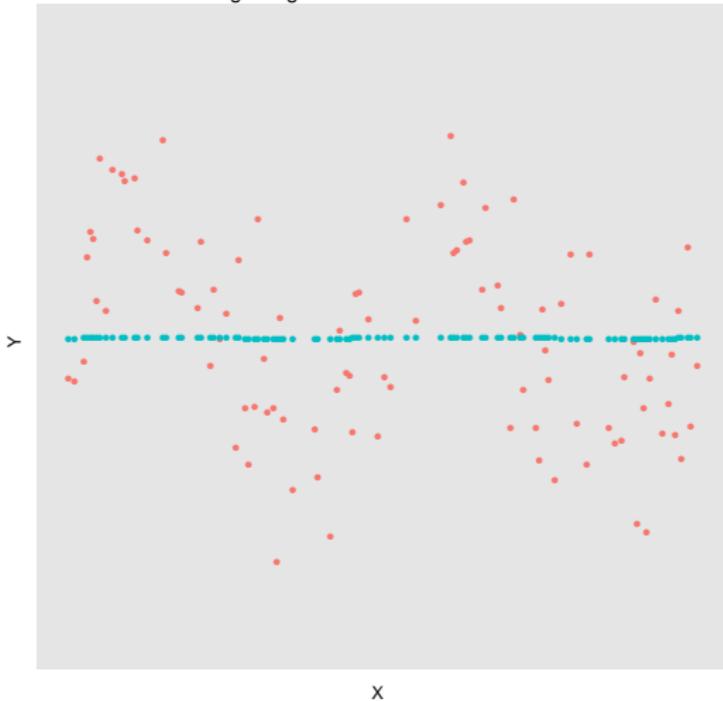
Ridge Regression with Lambda = 130



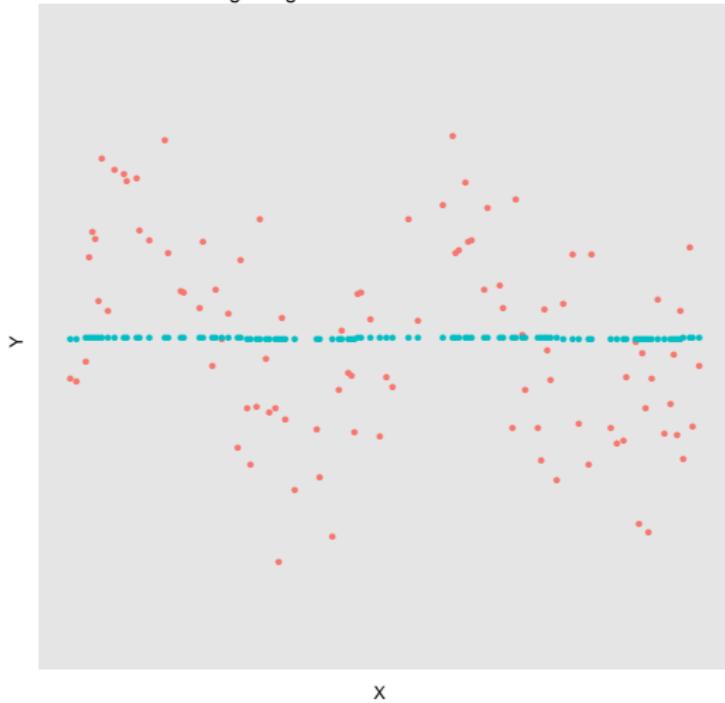
Ridge Regression with Lambda = 118



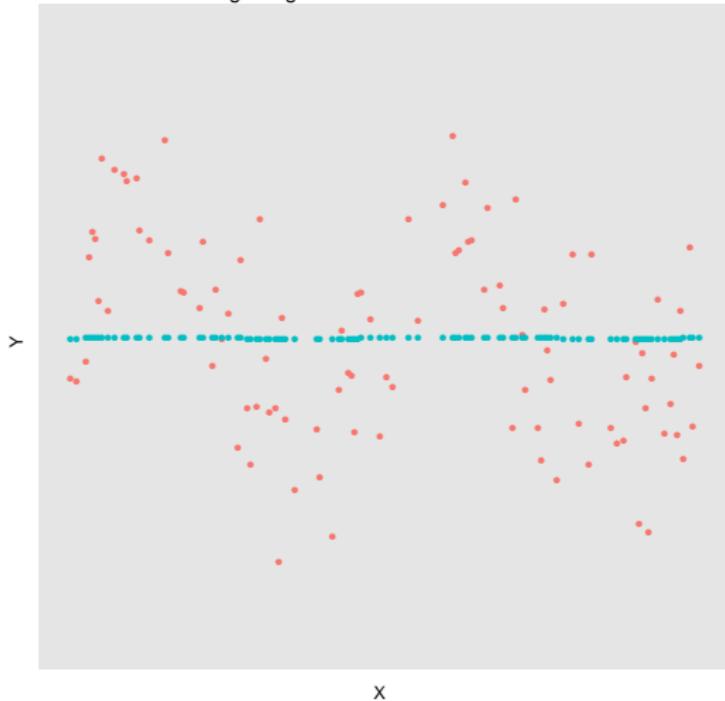
Ridge Regression with Lambda = 108



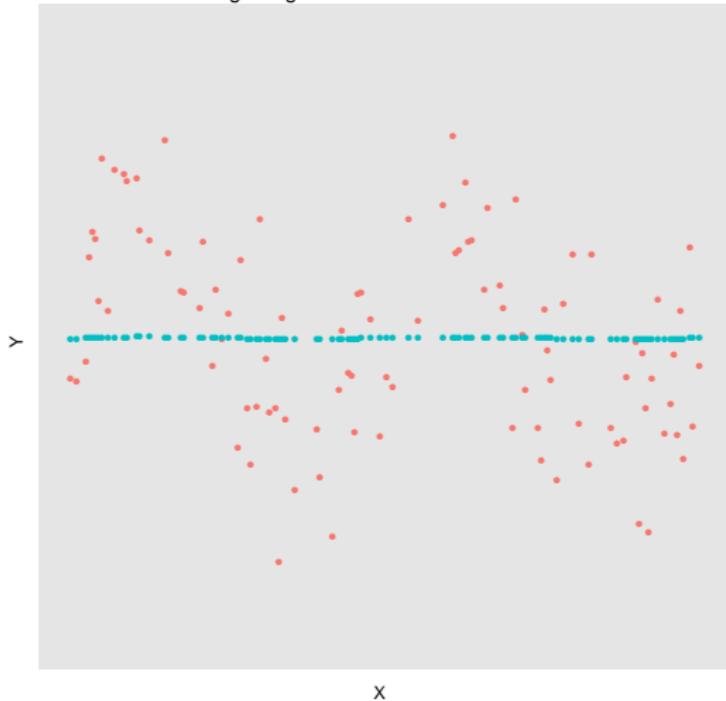
Ridge Regression with Lambda = 98.2



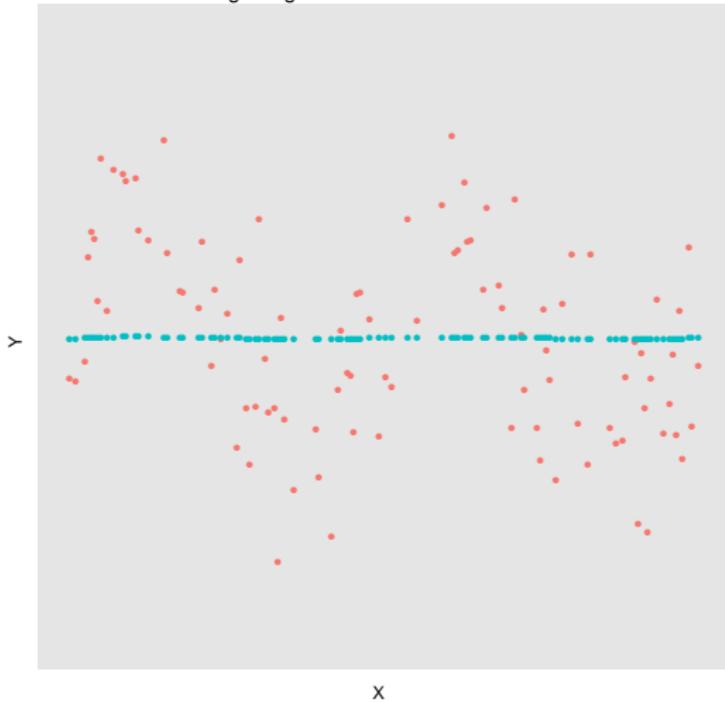
Ridge Regression with Lambda = 89.5



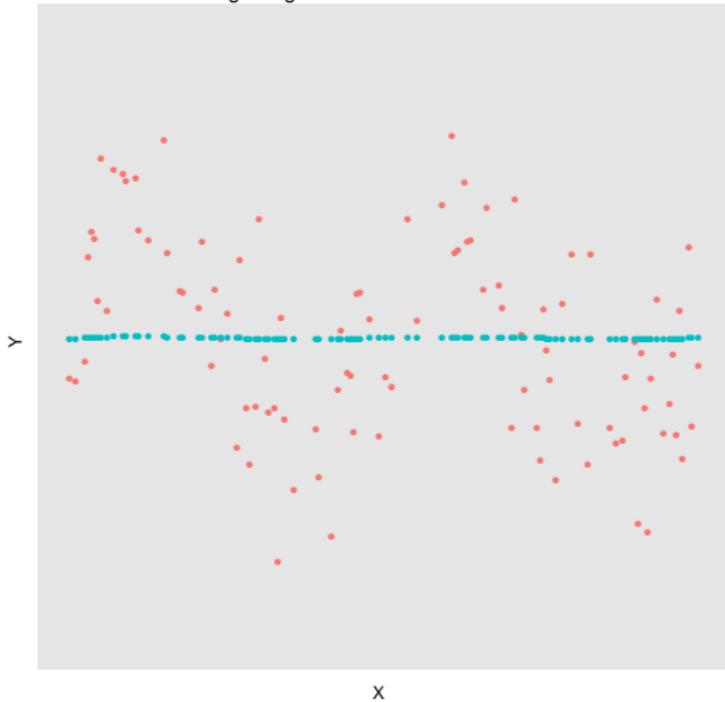
Ridge Regression with Lambda = 81.5



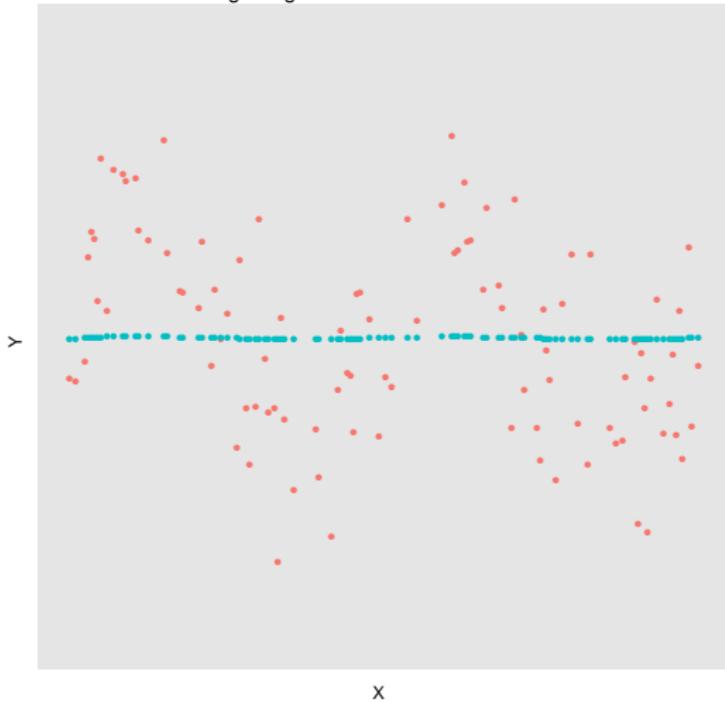
Ridge Regression with Lambda = 74.3



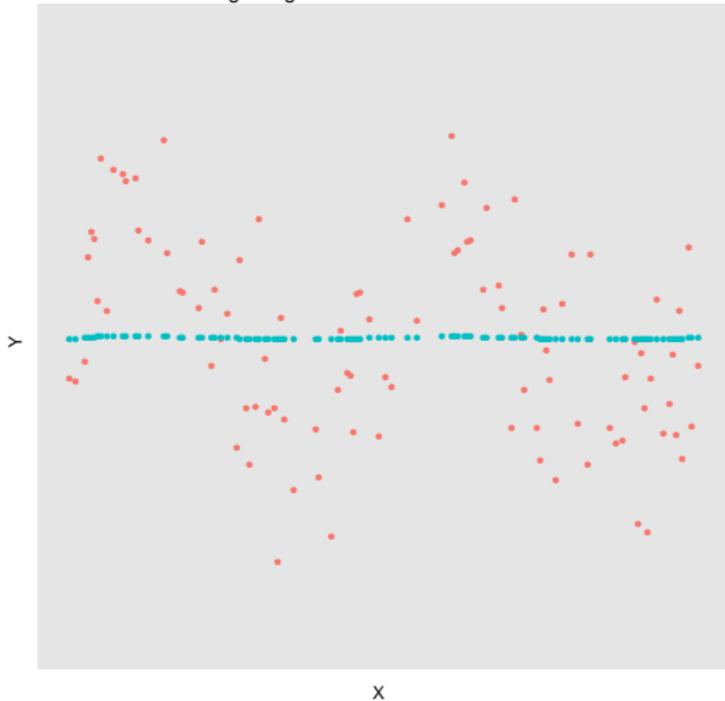
Ridge Regression with Lambda = 67.7



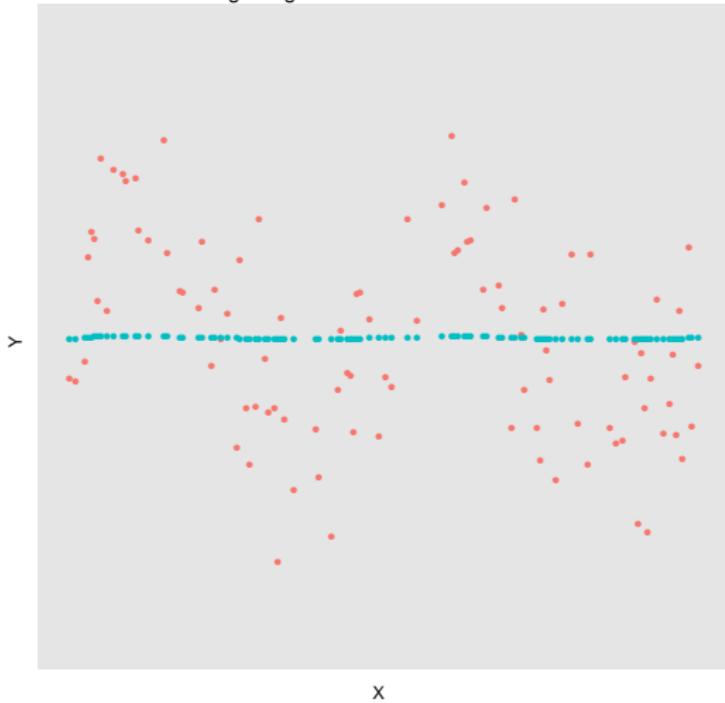
Ridge Regression with Lambda = 61.7



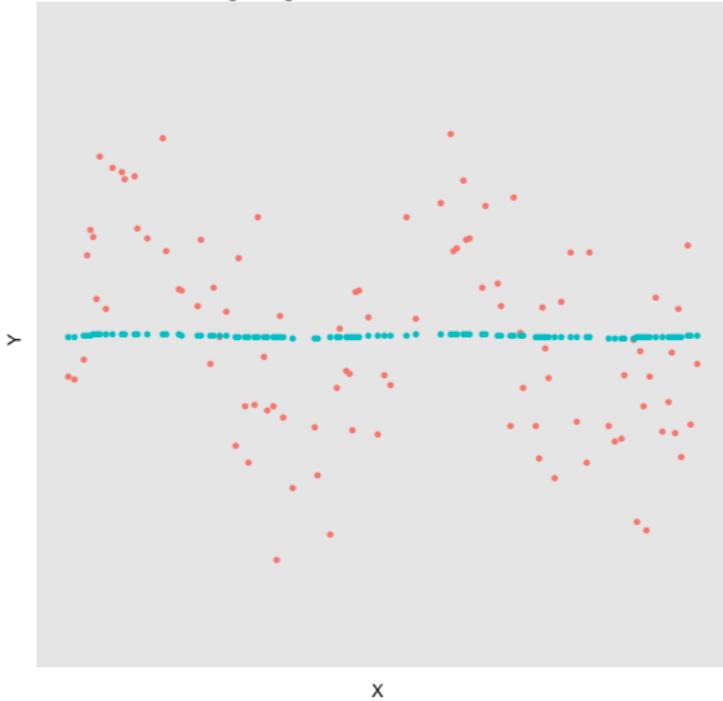
Ridge Regression with Lambda = 56.2



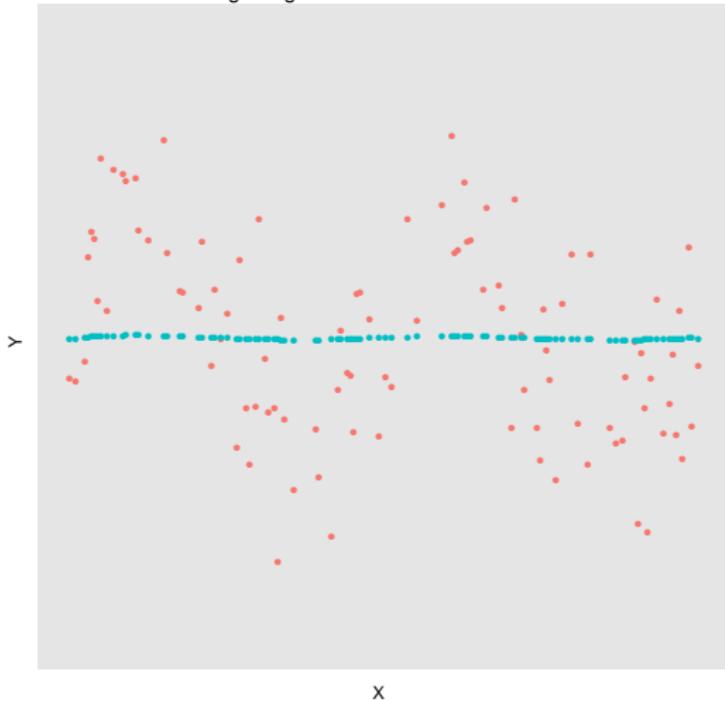
Ridge Regression with Lambda = 51.2



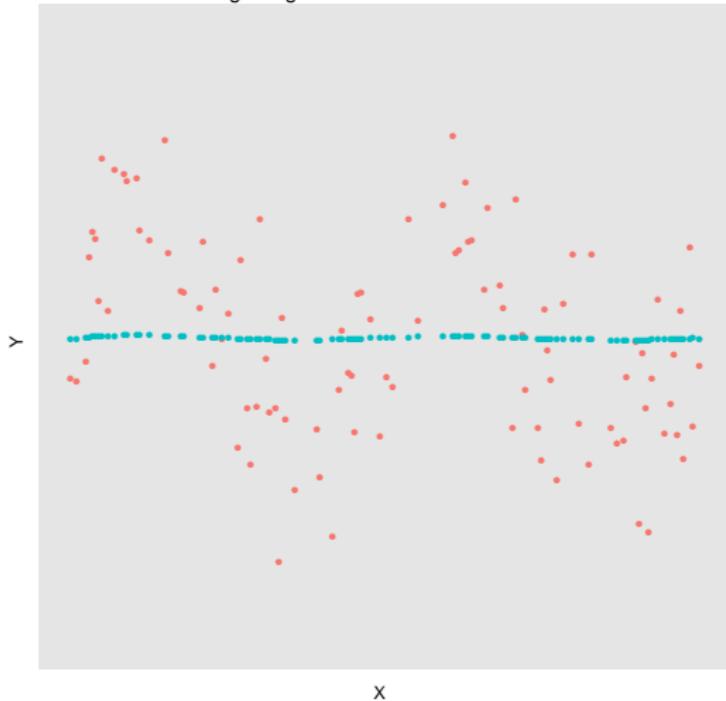
Ridge Regression with Lambda = 46.7



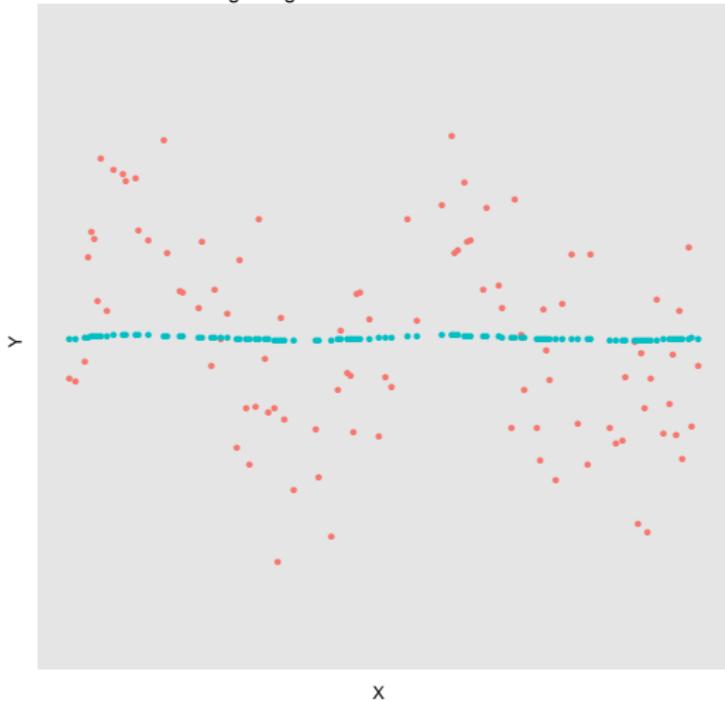
Ridge Regression with Lambda = 42.5



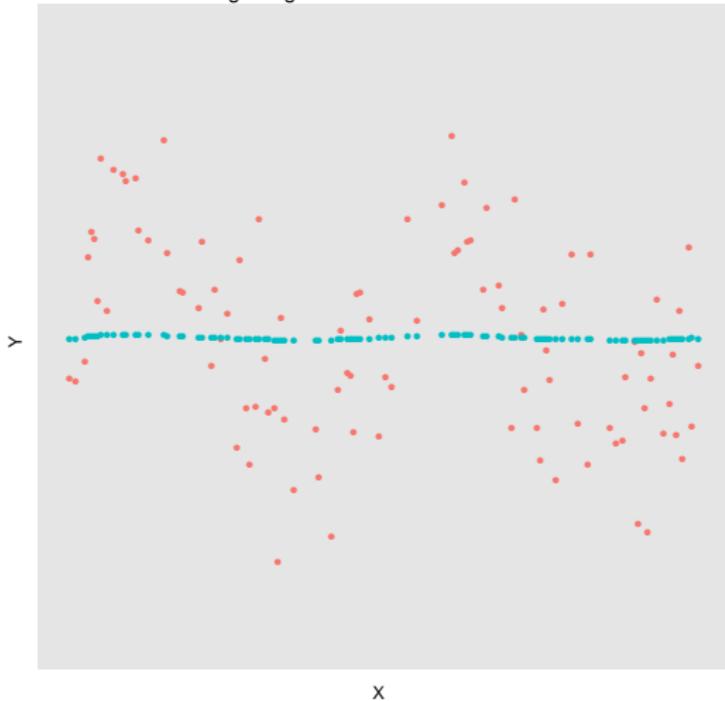
Ridge Regression with Lambda = 38.7



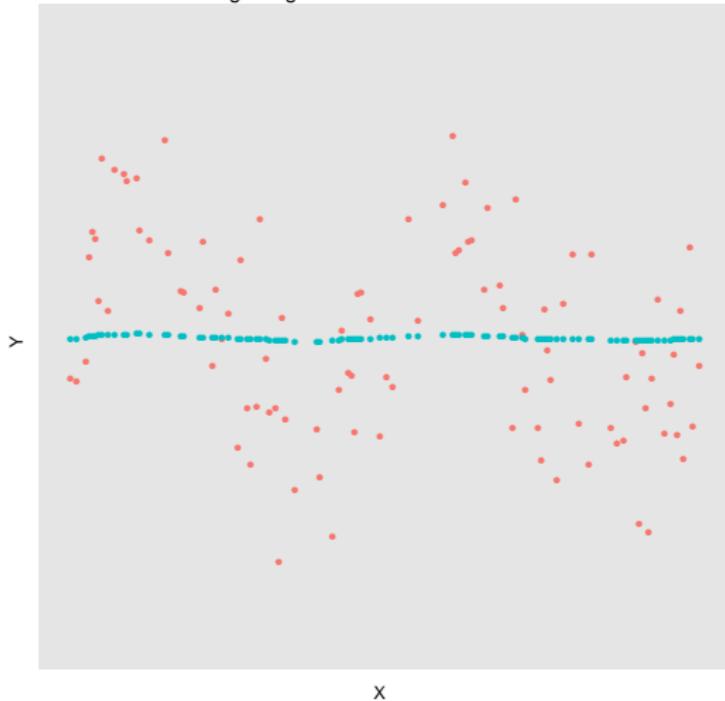
Ridge Regression with Lambda = 35.3



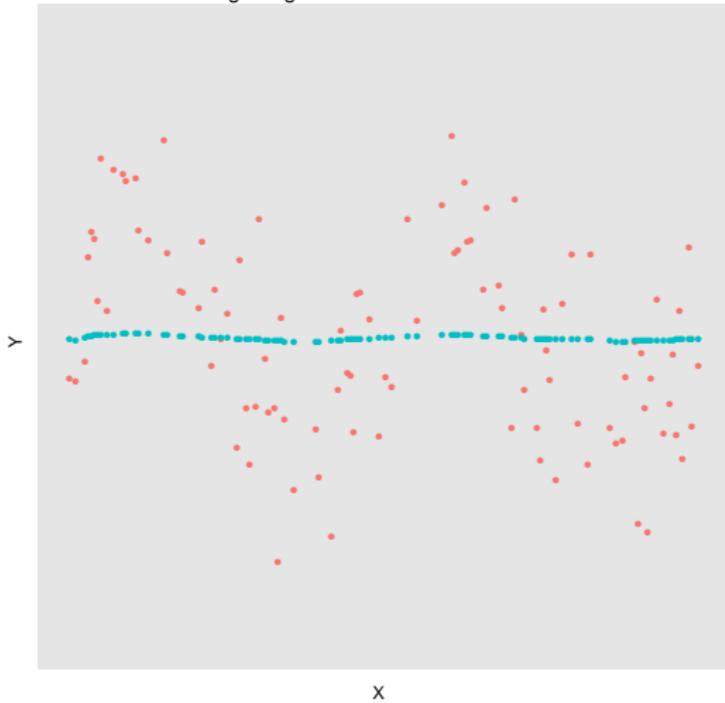
Ridge Regression with Lambda = 32.2



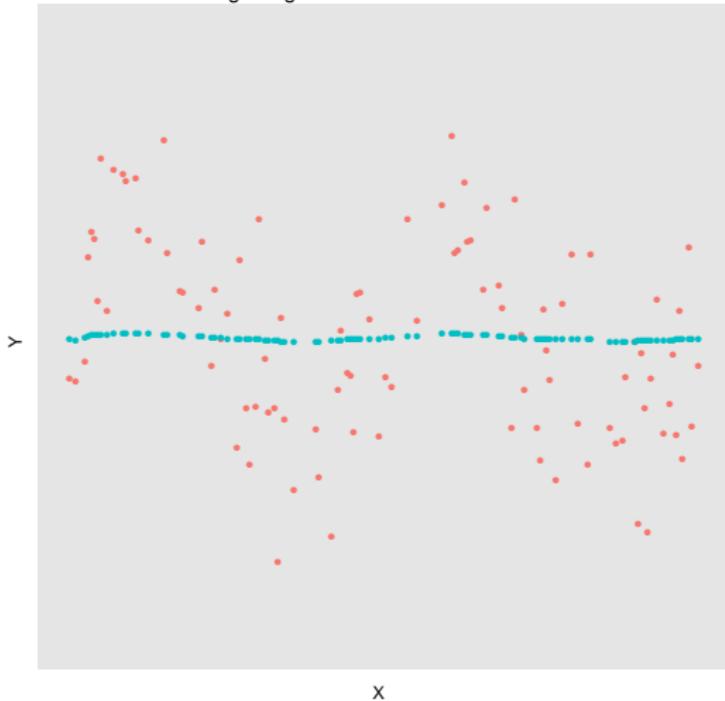
Ridge Regression with Lambda = 29.3



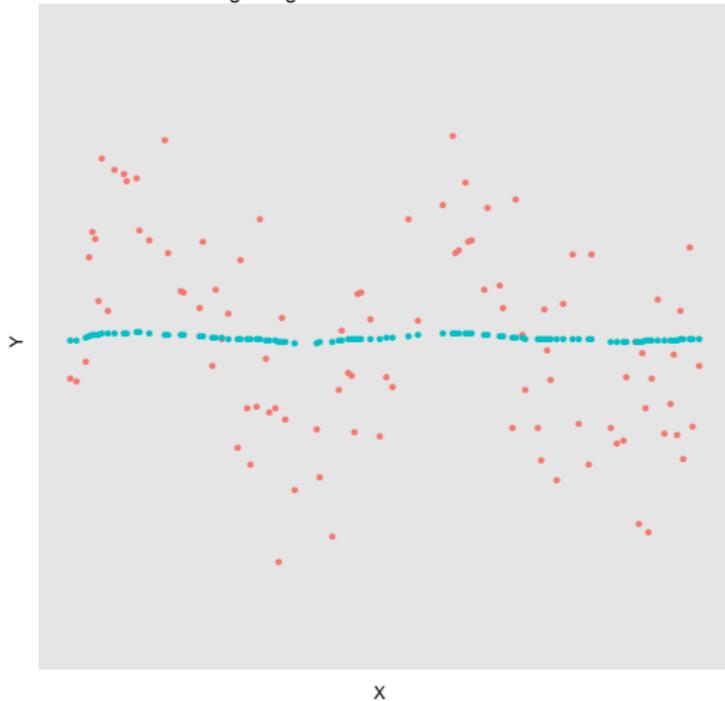
Ridge Regression with Lambda = 26.7



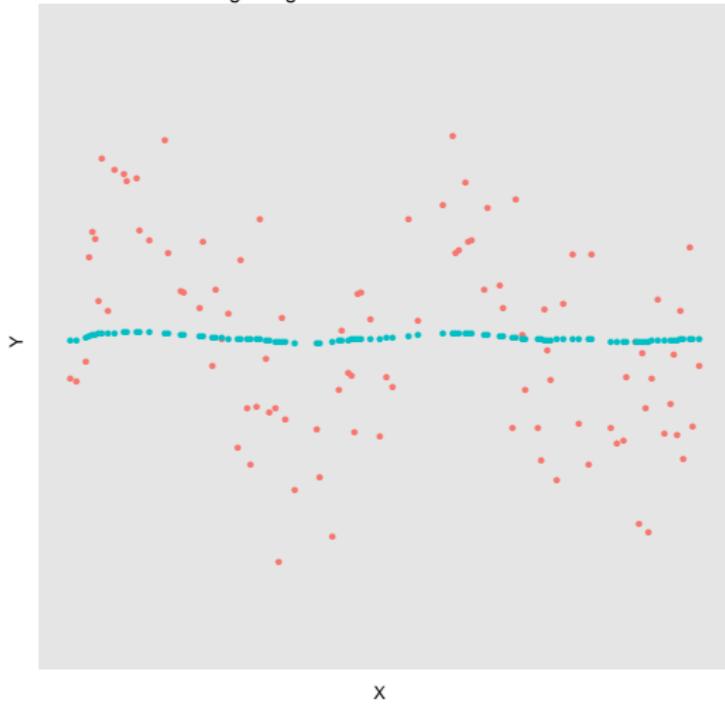
Ridge Regression with Lambda = 24.3



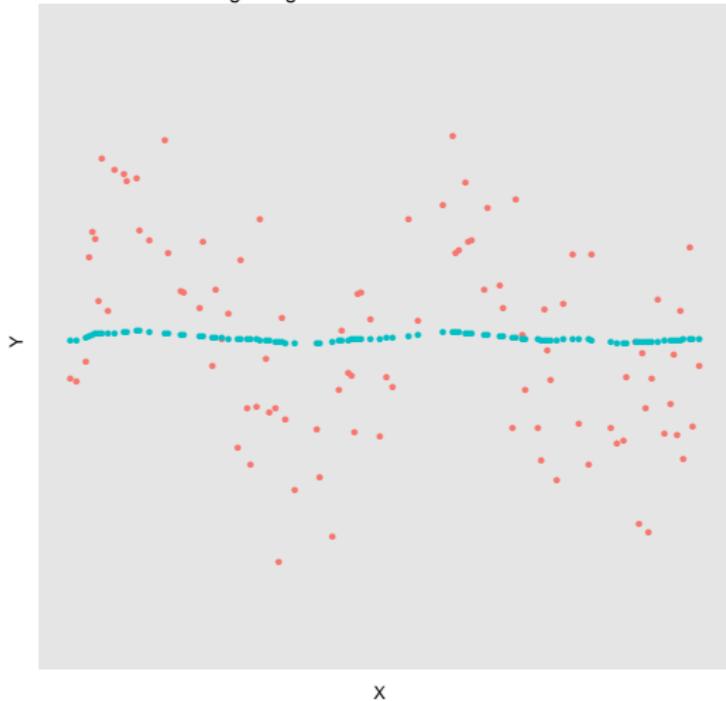
Ridge Regression with Lambda = 22.2



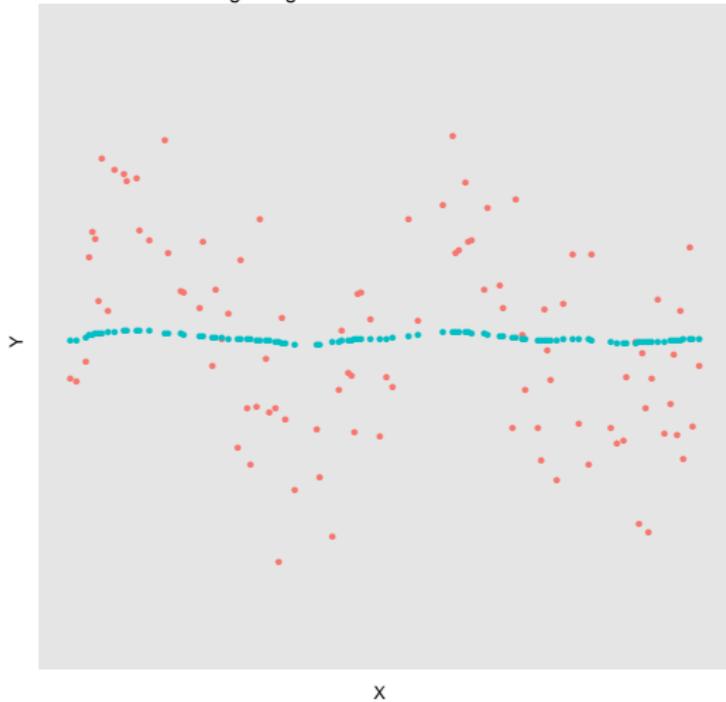
Ridge Regression with Lambda = 20.2



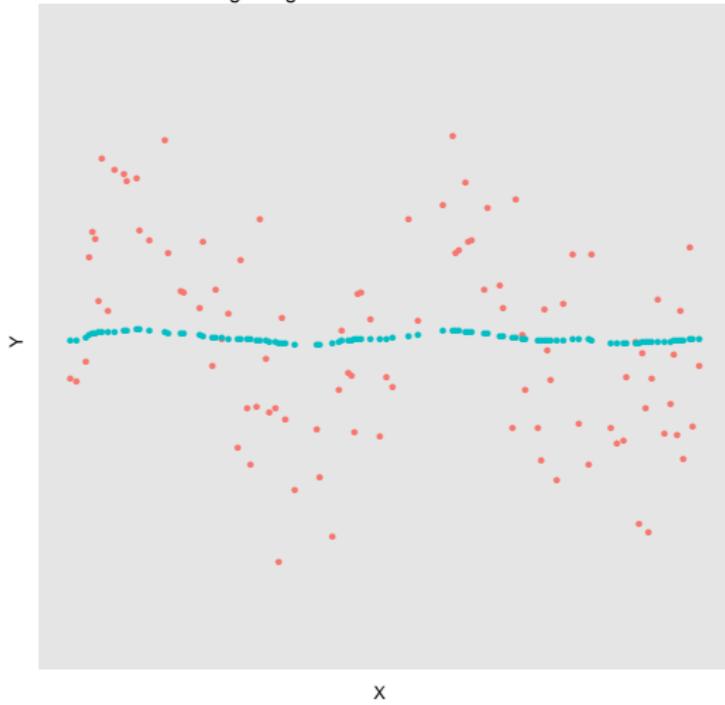
Ridge Regression with Lambda = 18.4



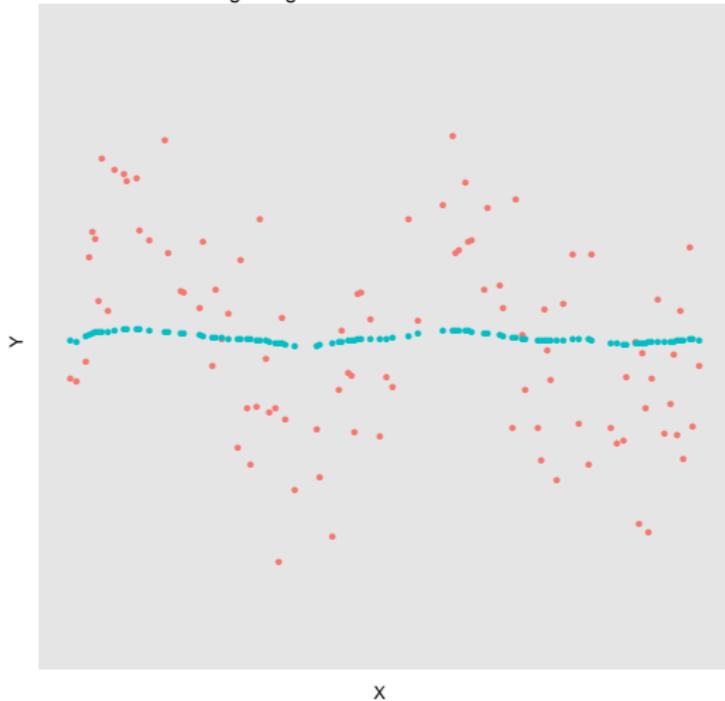
Ridge Regression with Lambda = 16.8



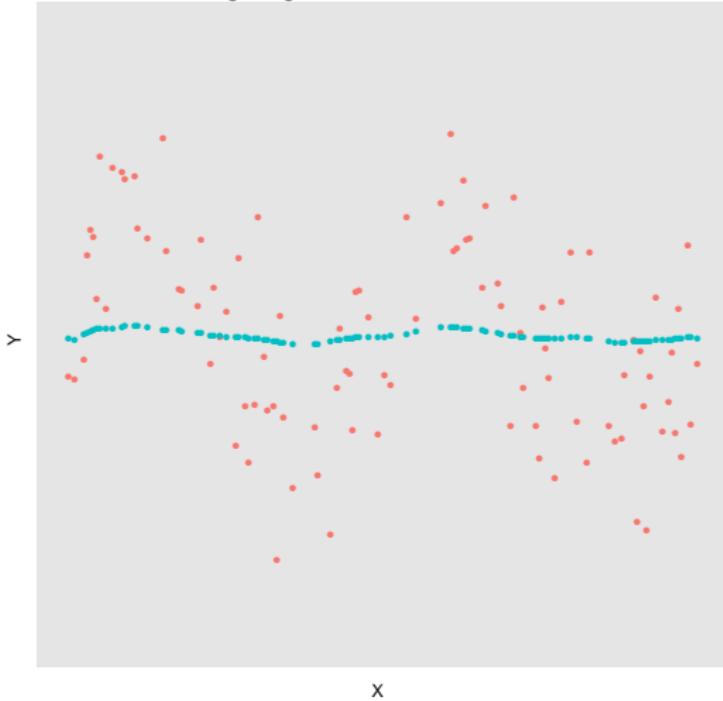
Ridge Regression with Lambda = 15.3



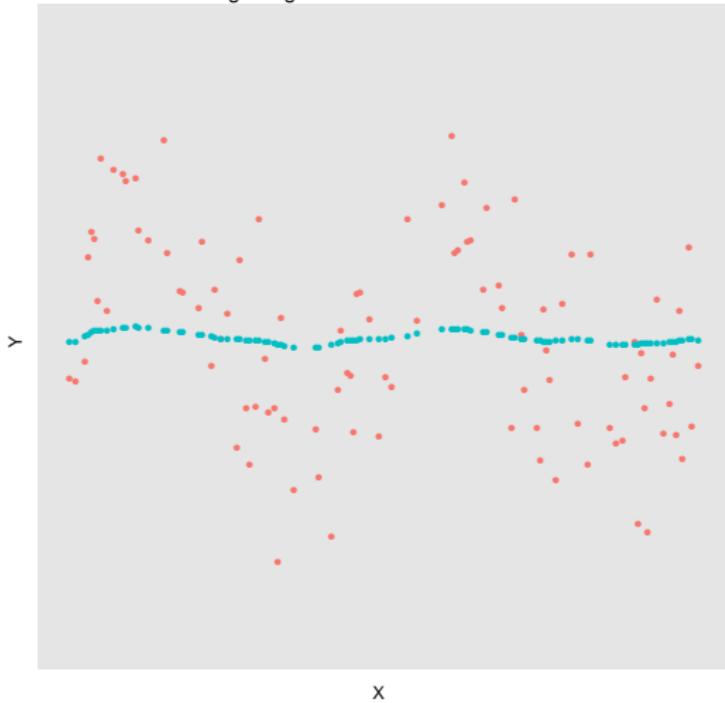
Ridge Regression with Lambda = 13.9



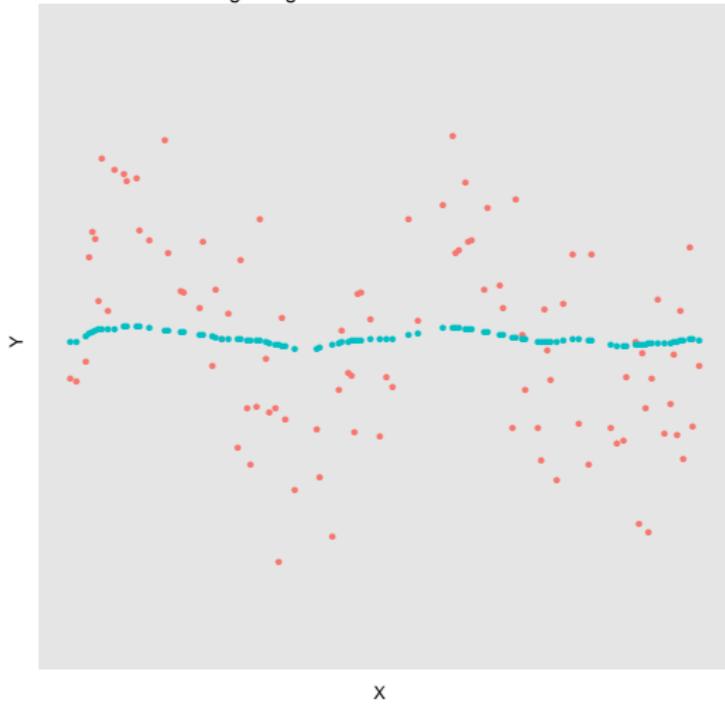
Ridge Regression with Lambda = 12.7



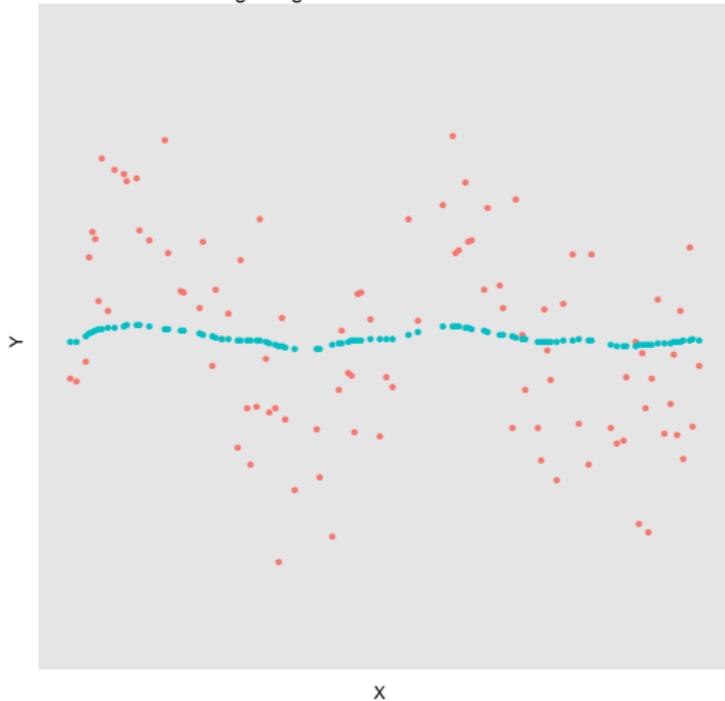
Ridge Regression with Lambda = 11.6



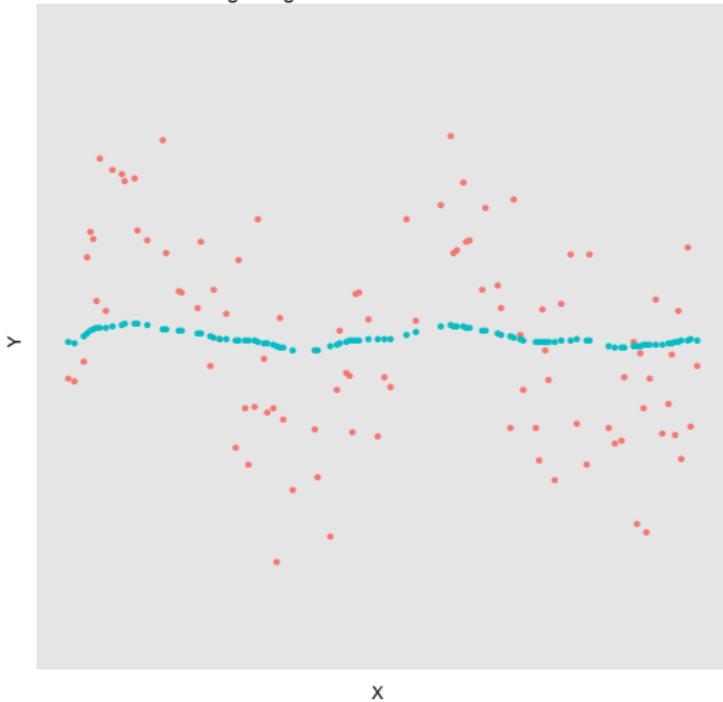
Ridge Regression with Lambda = 10.5



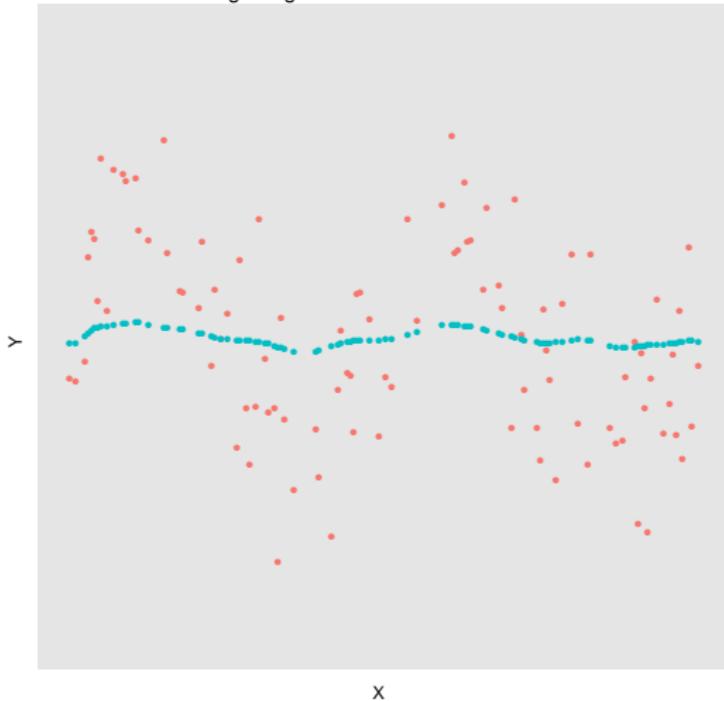
Ridge Regression with Lambda = 9.6



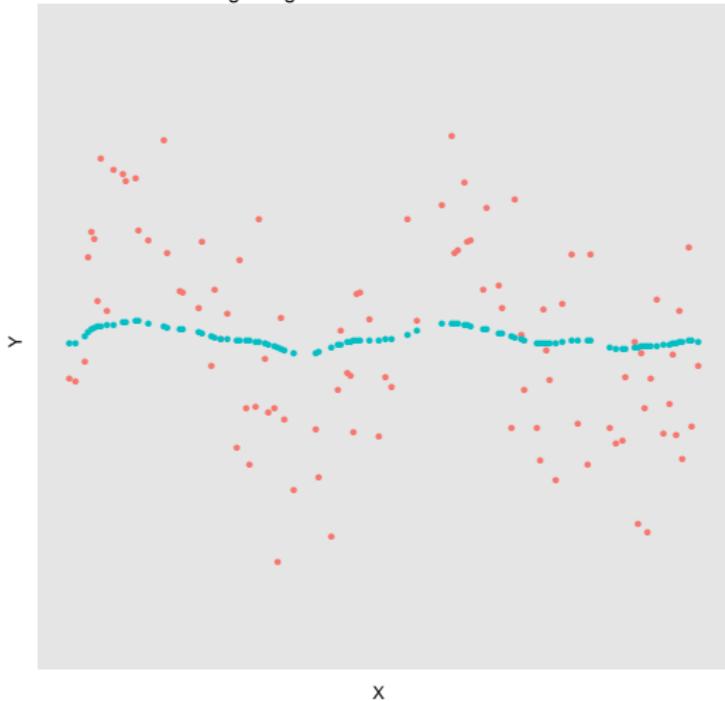
Ridge Regression with Lambda = 8.74



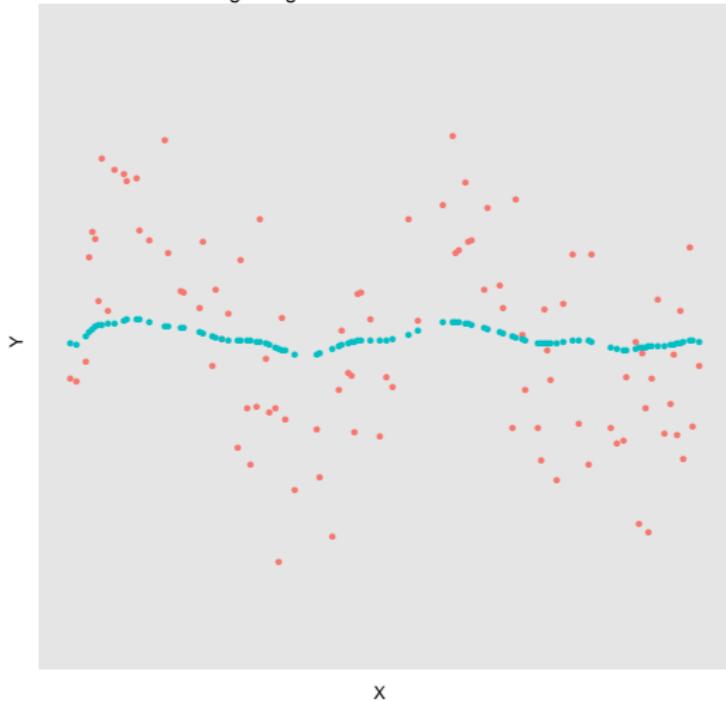
Ridge Regression with Lambda = 7.97



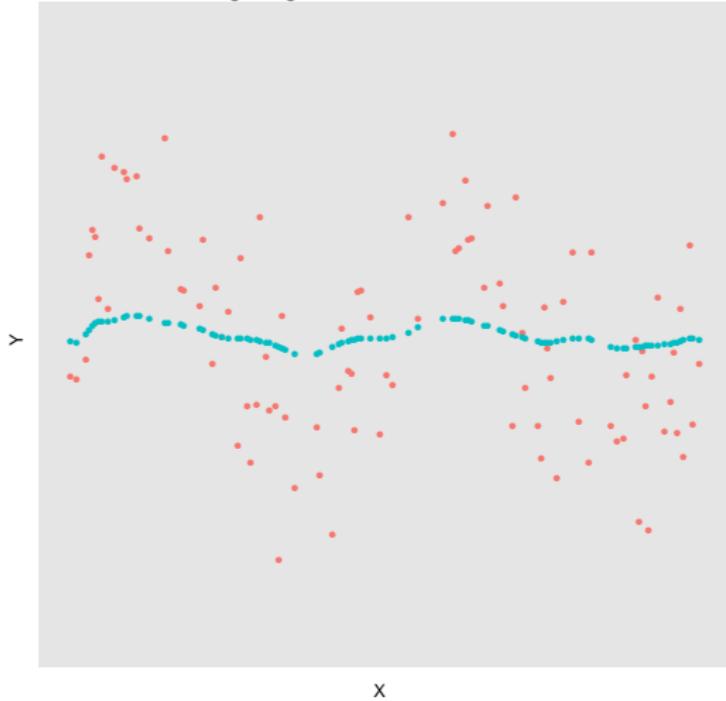
Ridge Regression with Lambda = 7.26



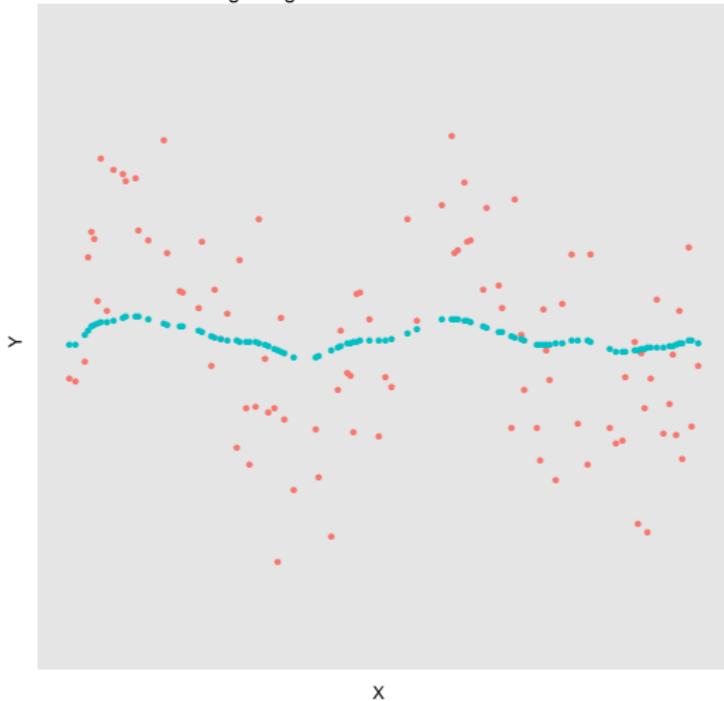
Ridge Regression with Lambda = 6.61



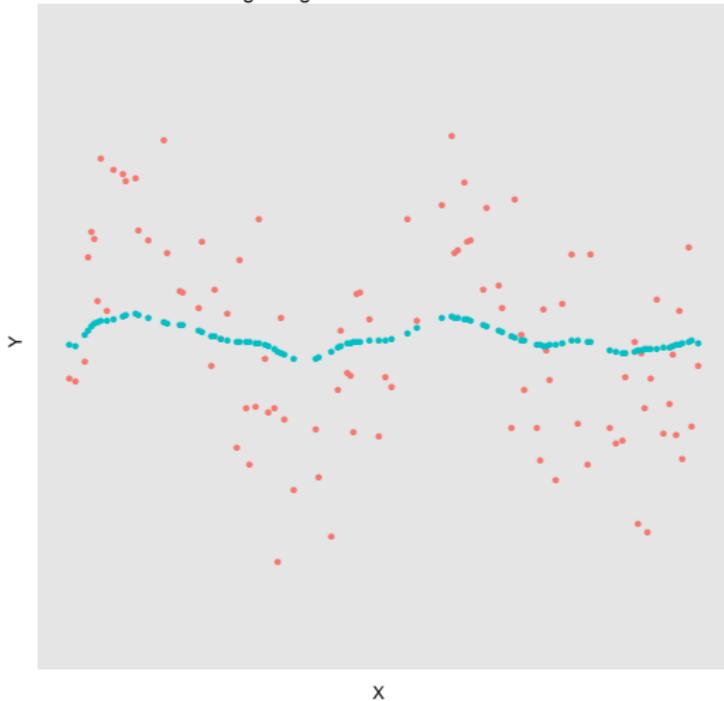
Ridge Regression with Lambda = 6.03



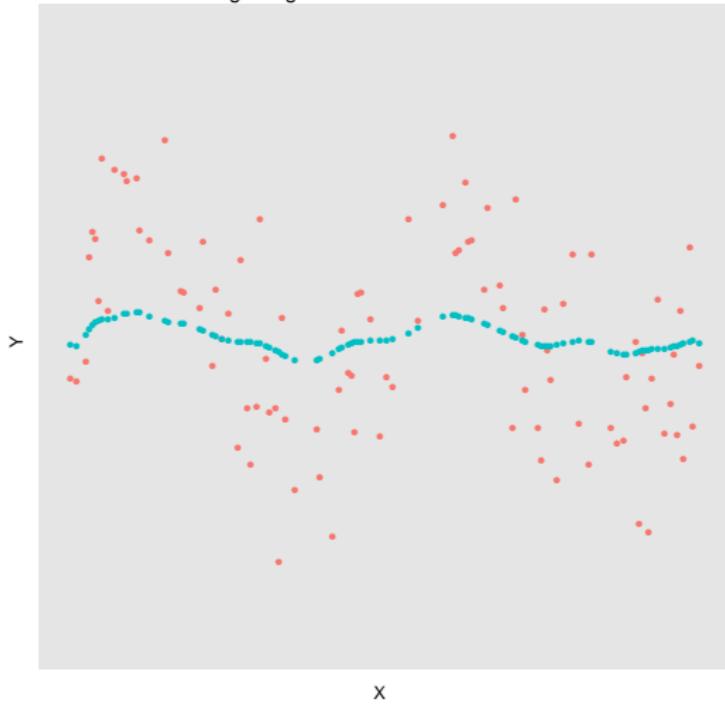
Ridge Regression with Lambda = 5.49



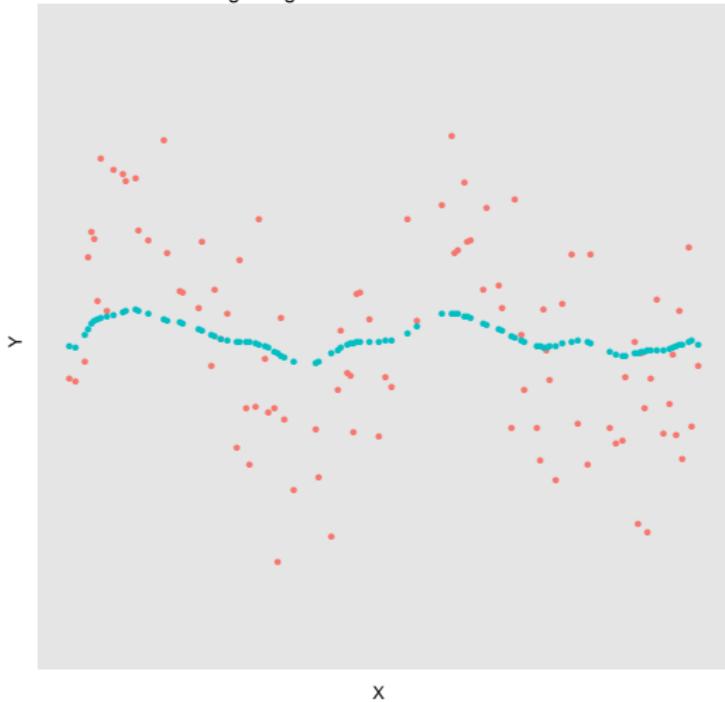
Ridge Regression with Lambda = 5



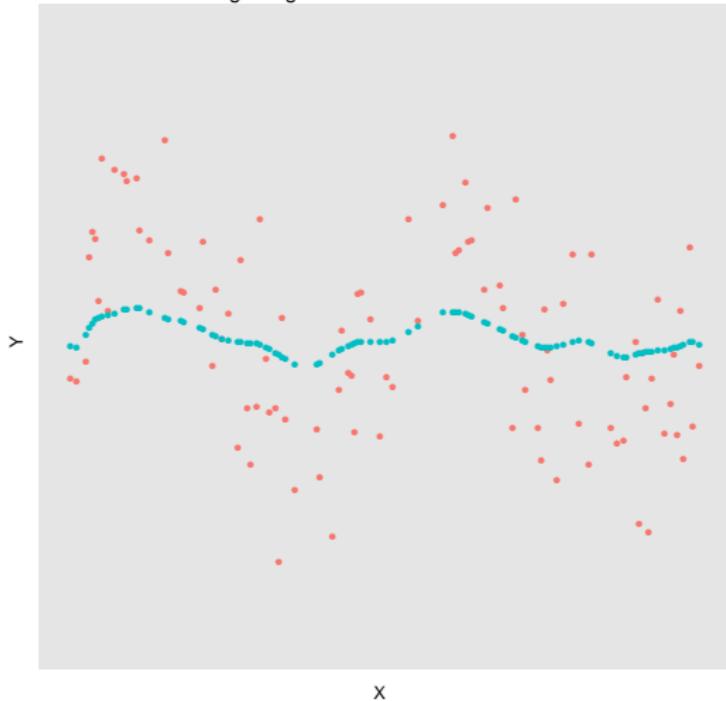
Ridge Regression with Lambda = 4.56



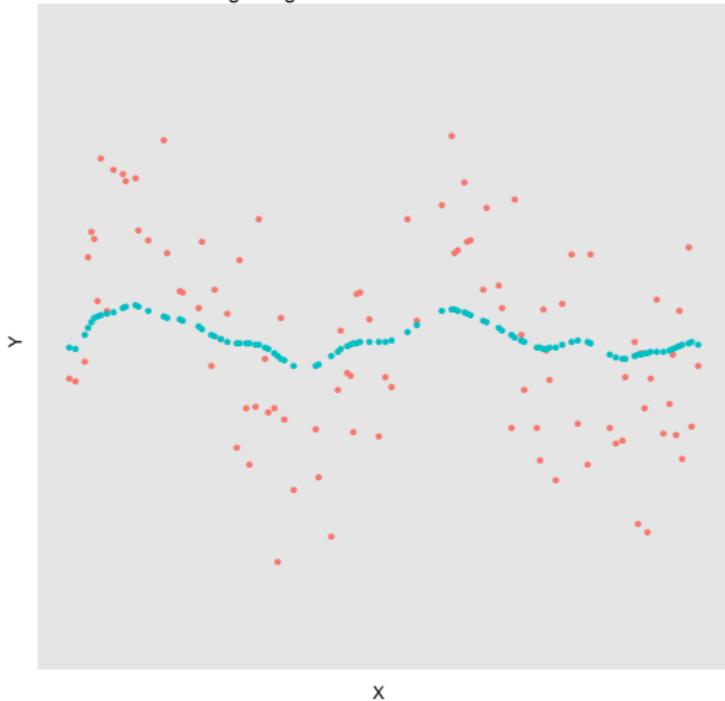
Ridge Regression with Lambda = 4.15



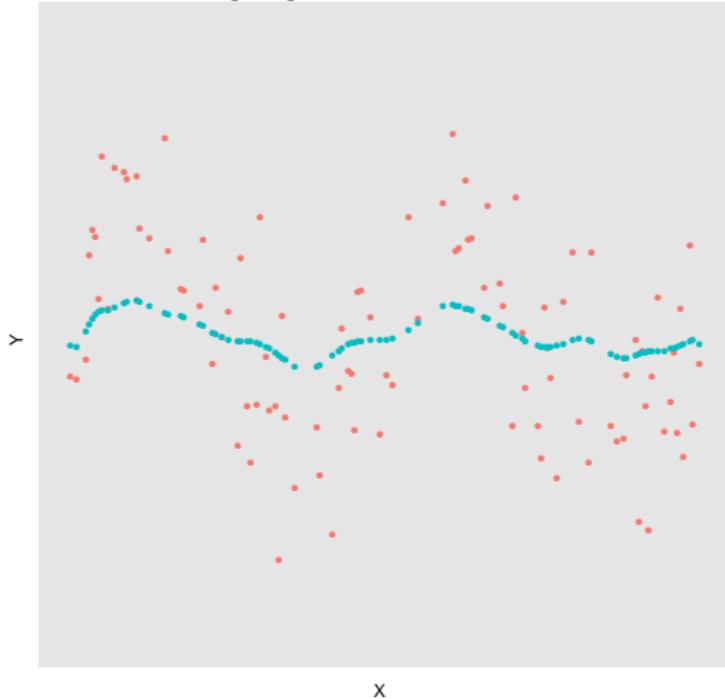
Ridge Regression with Lambda = 3.78



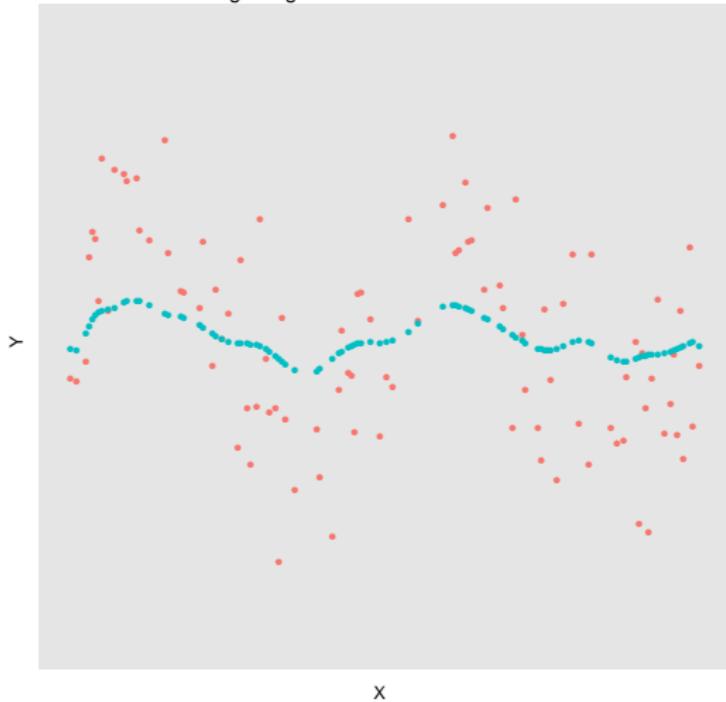
Ridge Regression with Lambda = 3.45



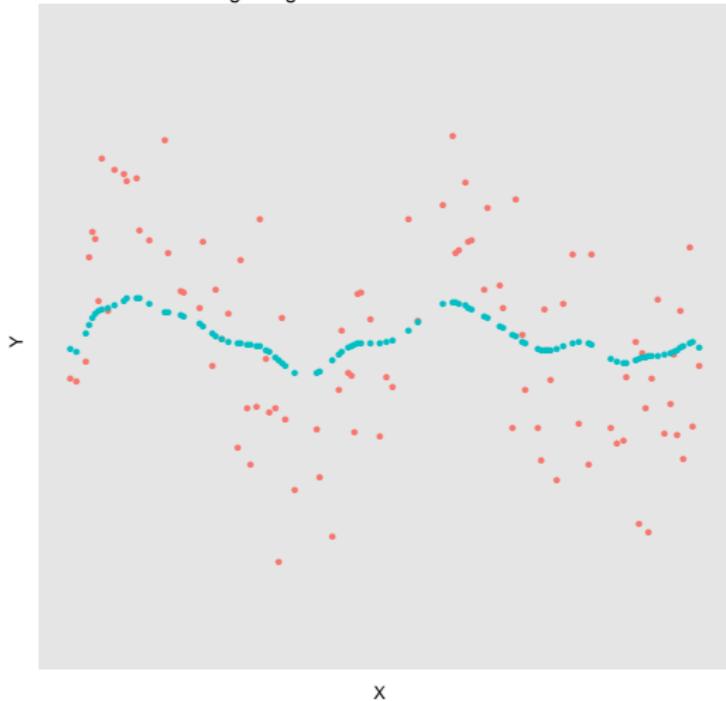
Ridge Regression with Lambda = 3.14



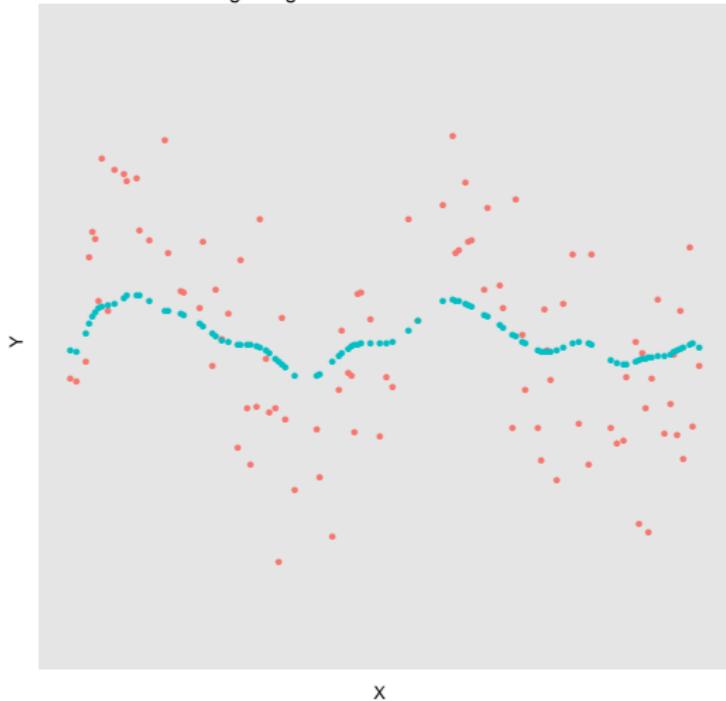
Ridge Regression with Lambda = 2.86



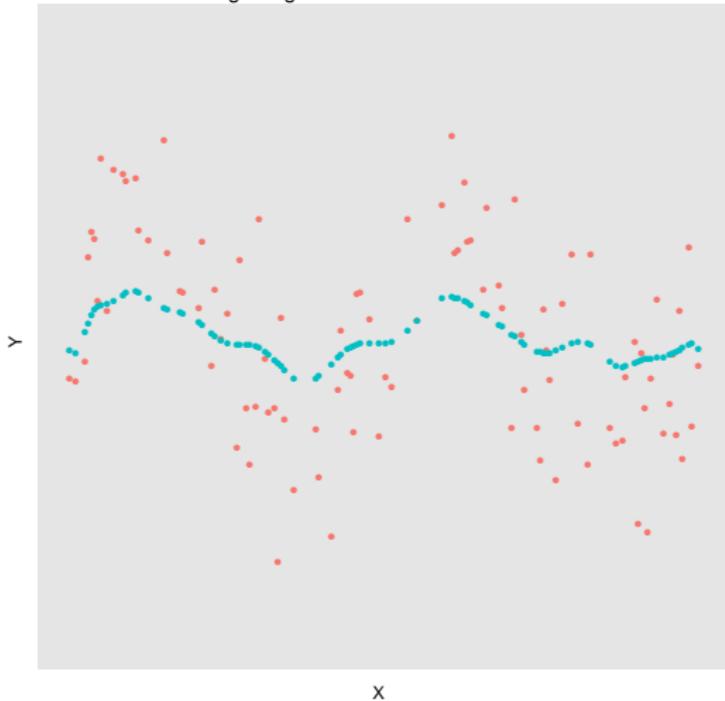
Ridge Regression with Lambda = 2.61



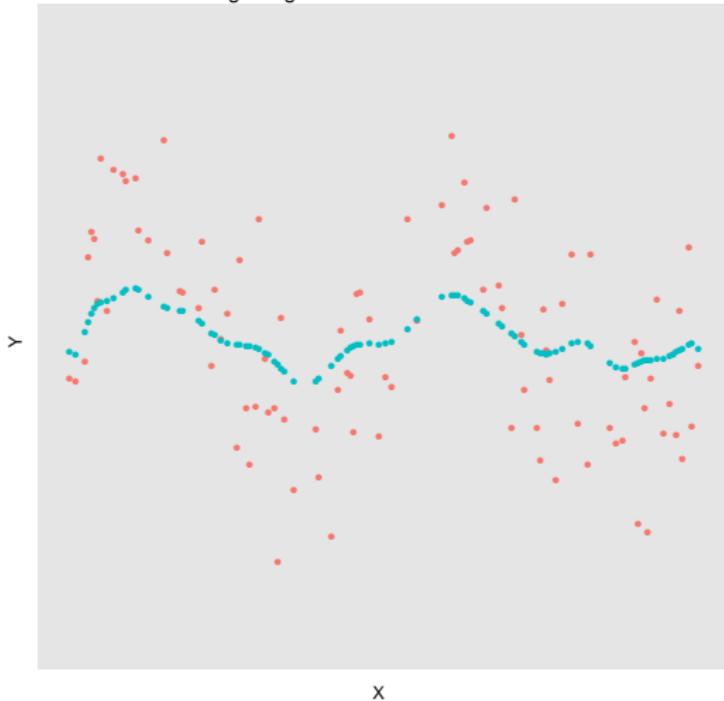
Ridge Regression with Lambda = 2.38



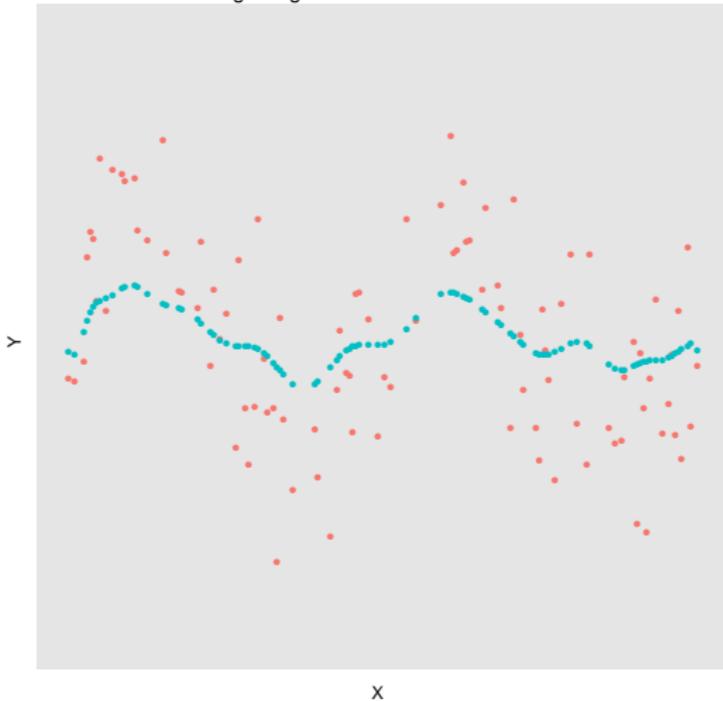
Ridge Regression with Lambda = 2.17



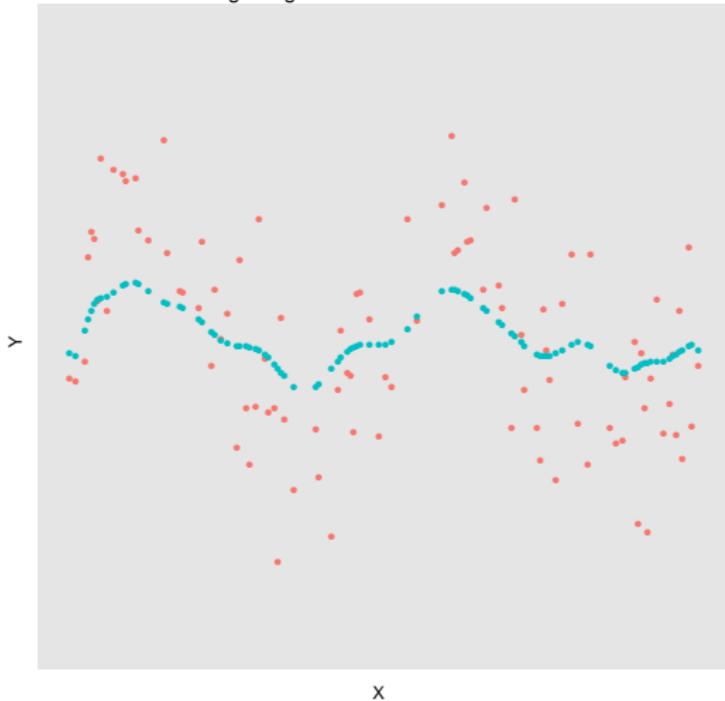
Ridge Regression with Lambda = 1.97



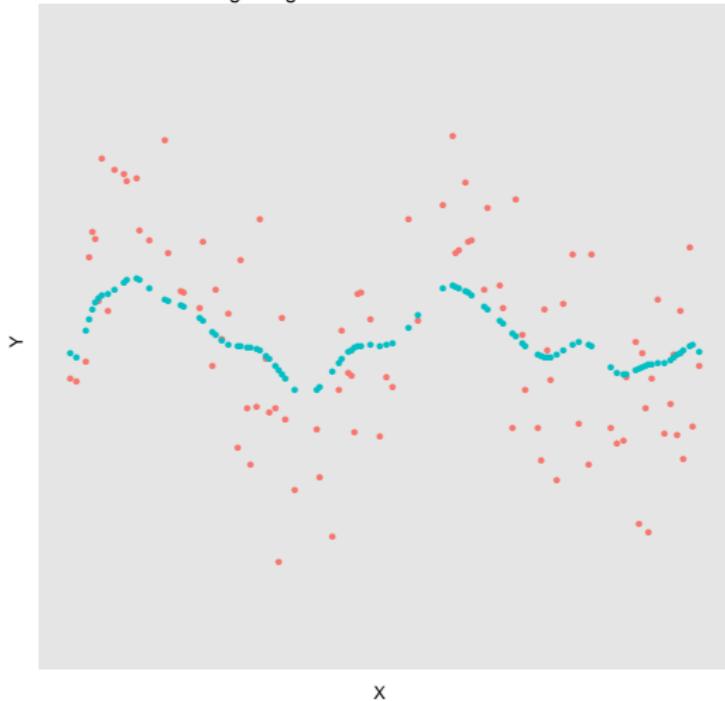
Ridge Regression with Lambda = 1.8



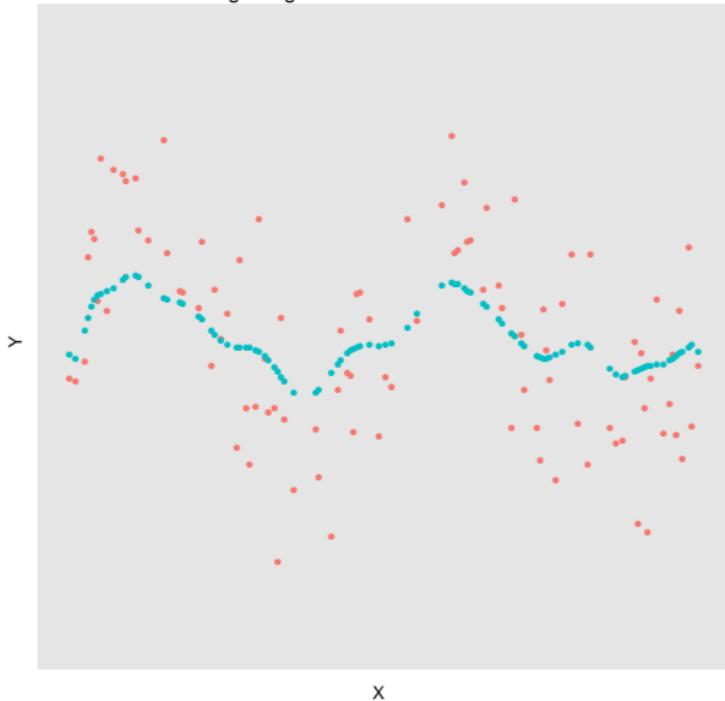
Ridge Regression with Lambda = 1.64



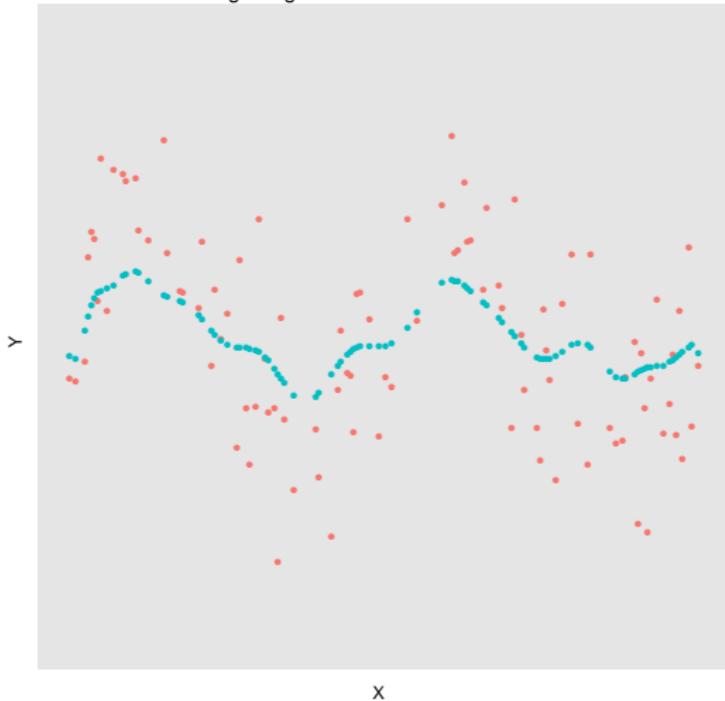
Ridge Regression with Lambda = 1.49



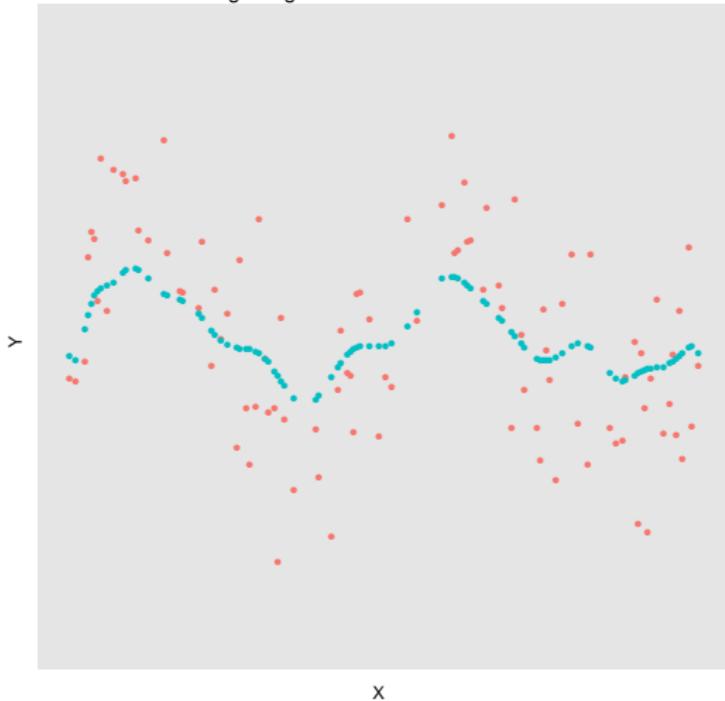
Ridge Regression with Lambda = 1.36



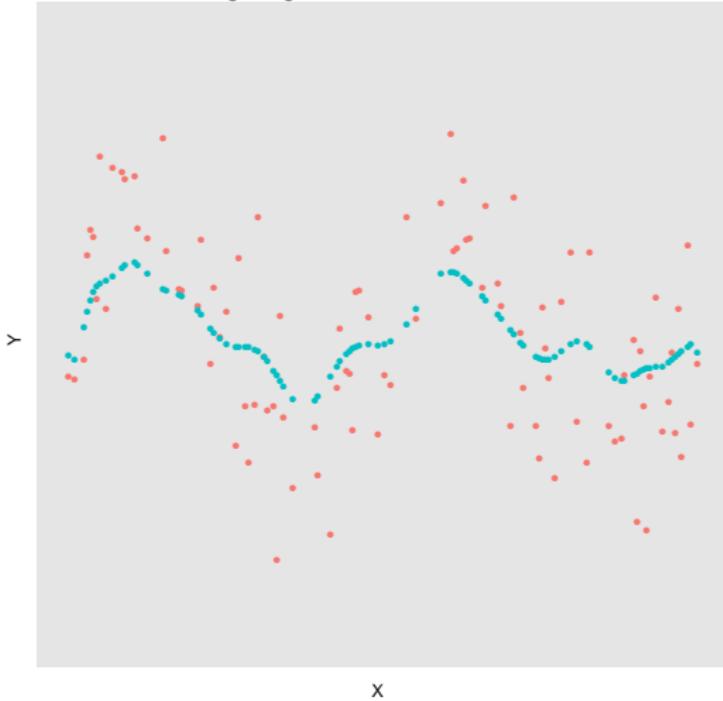
Ridge Regression with Lambda = 1.24



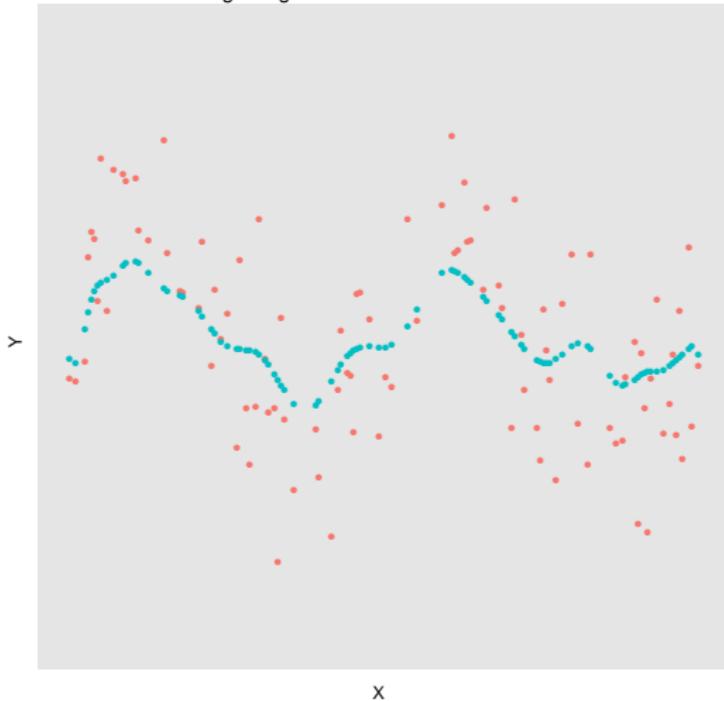
Ridge Regression with Lambda = 1.13



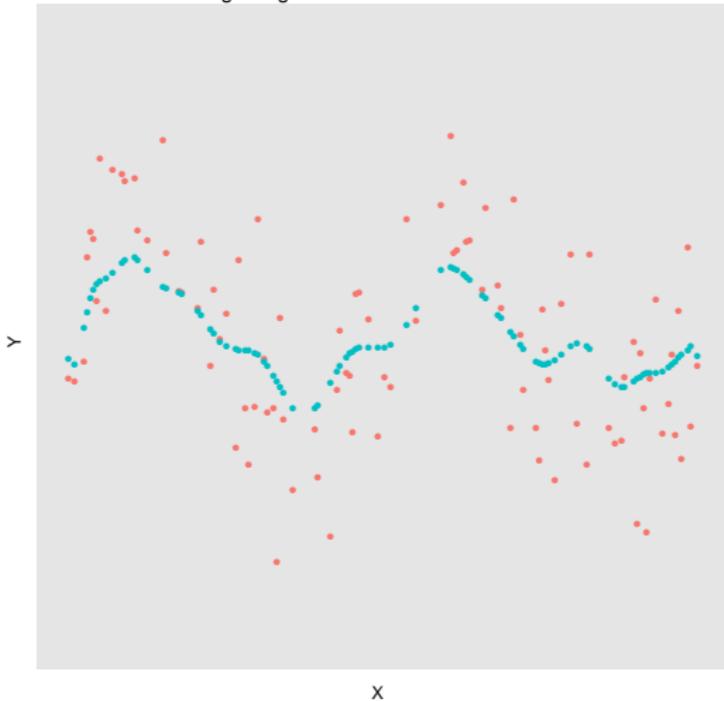
Ridge Regression with Lambda = 1.03



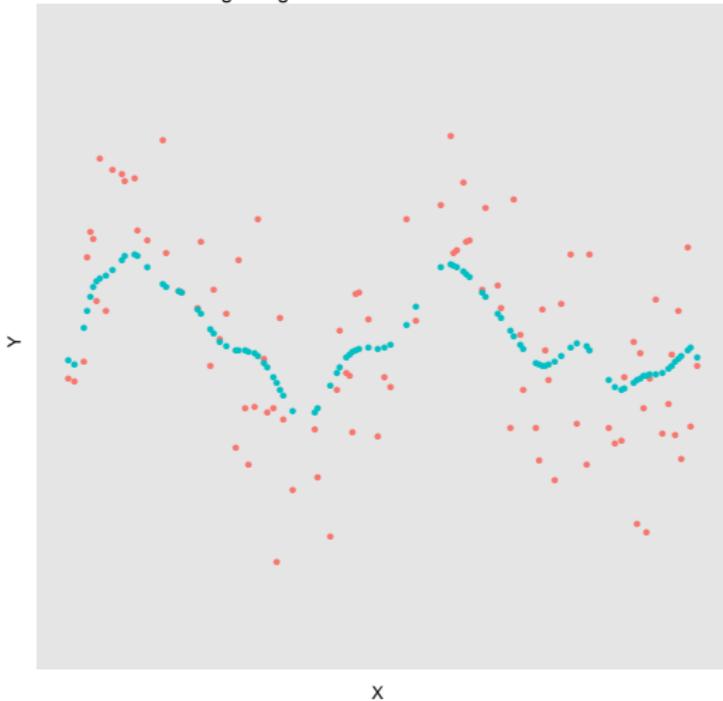
Ridge Regression with Lambda = 0.937



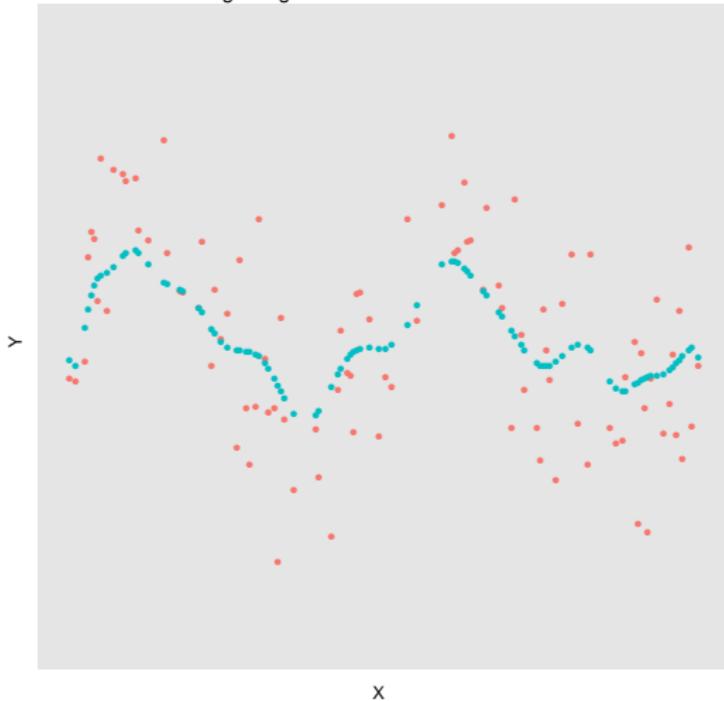
Ridge Regression with Lambda = 0.854



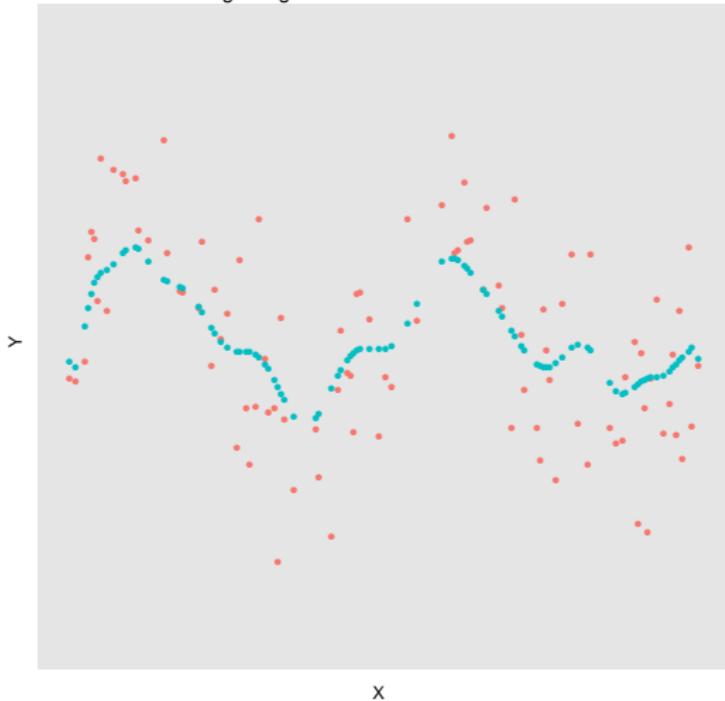
Ridge Regression with Lambda = 0.778



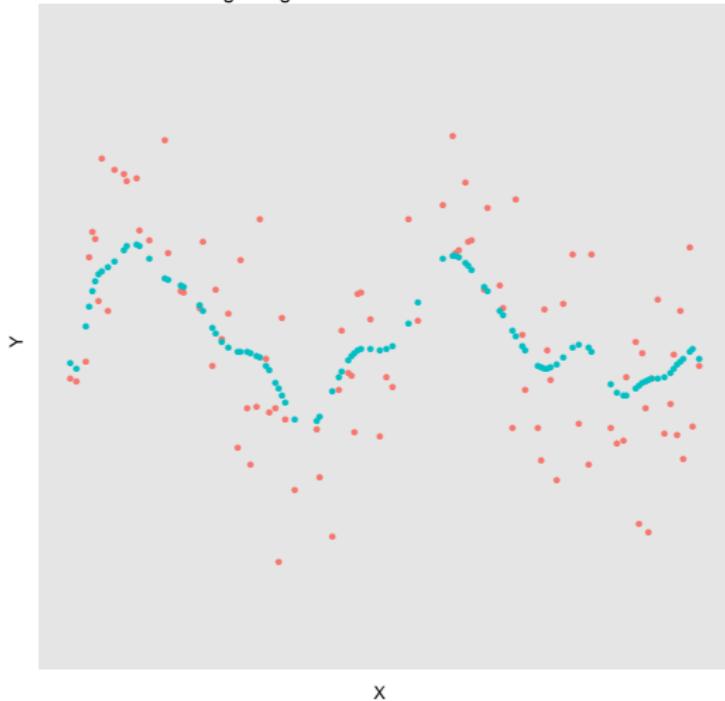
Ridge Regression with Lambda = 0.709



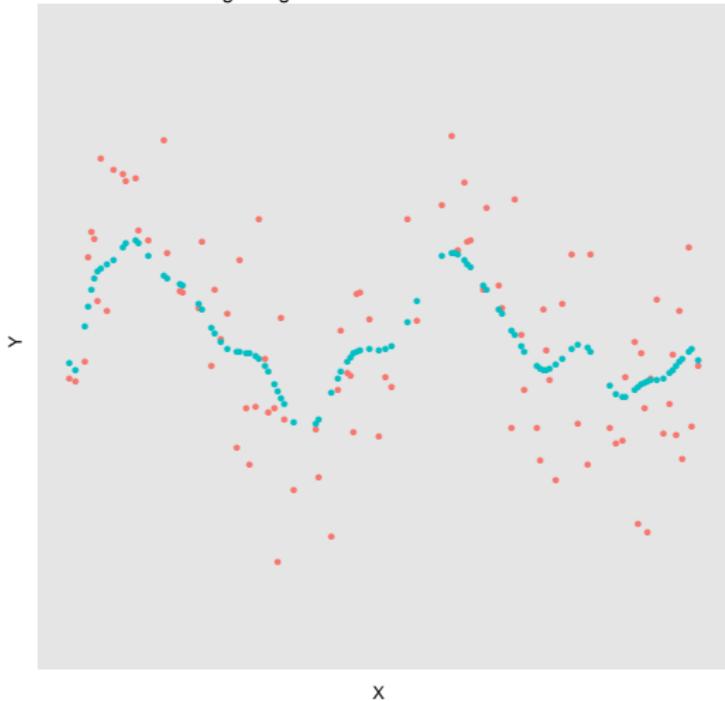
Ridge Regression with Lambda = 0.646



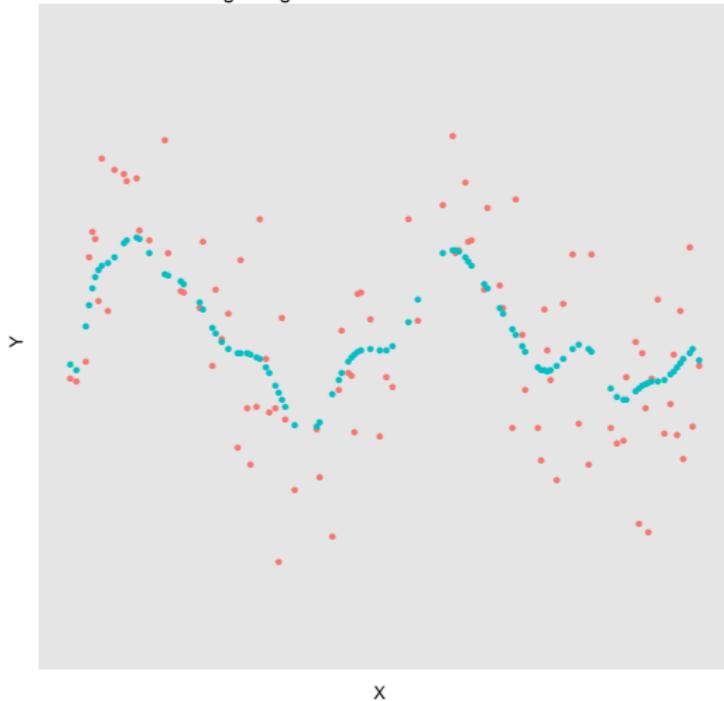
Ridge Regression with Lambda = 0.589



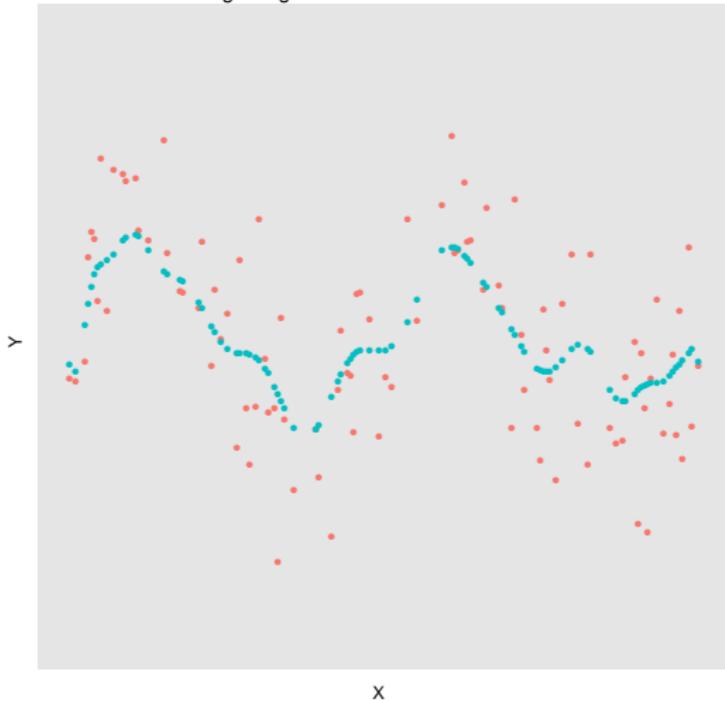
Ridge Regression with Lambda = 0.536



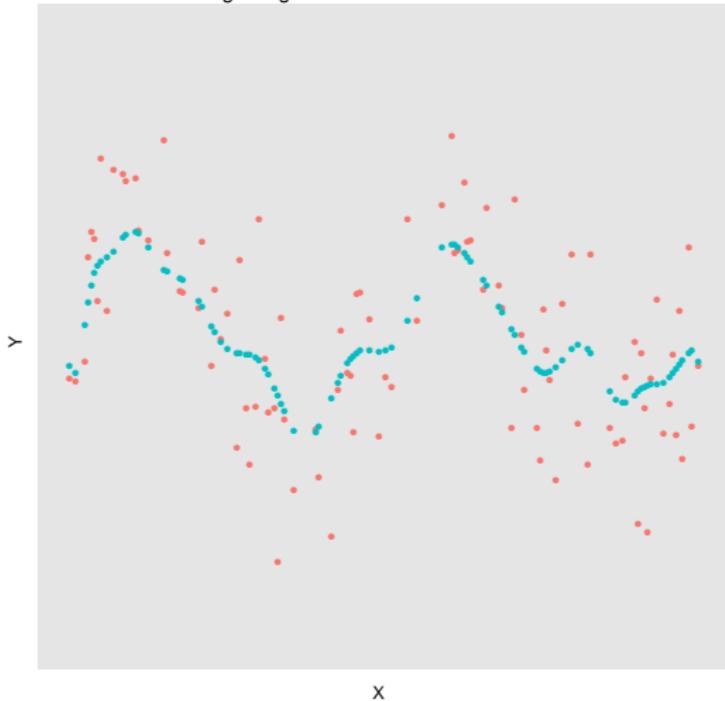
Ridge Regression with Lambda = 0.489



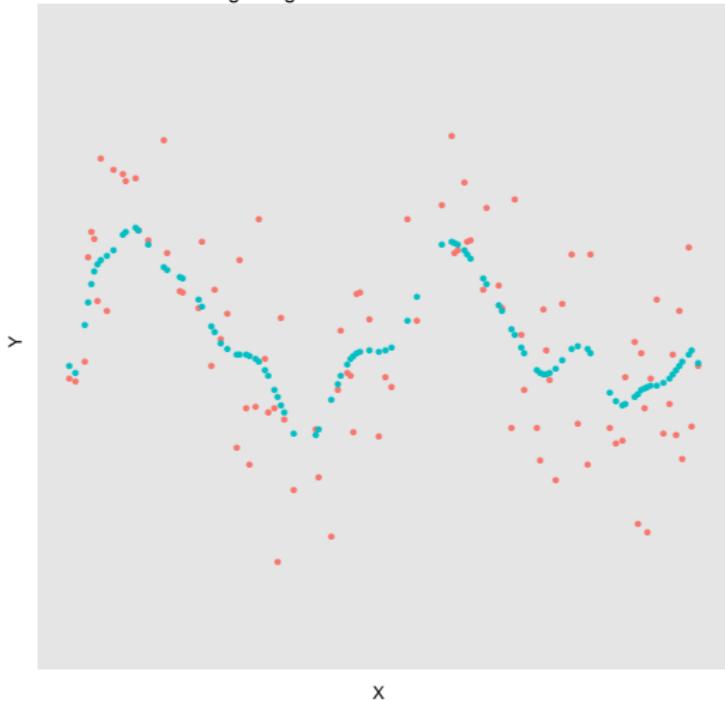
Ridge Regression with Lambda = 0.445



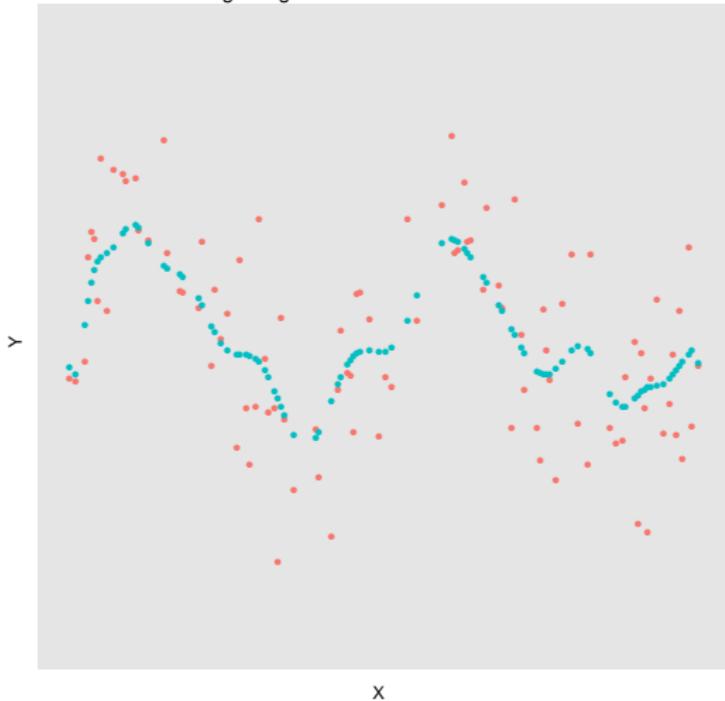
Ridge Regression with Lambda = 0.406



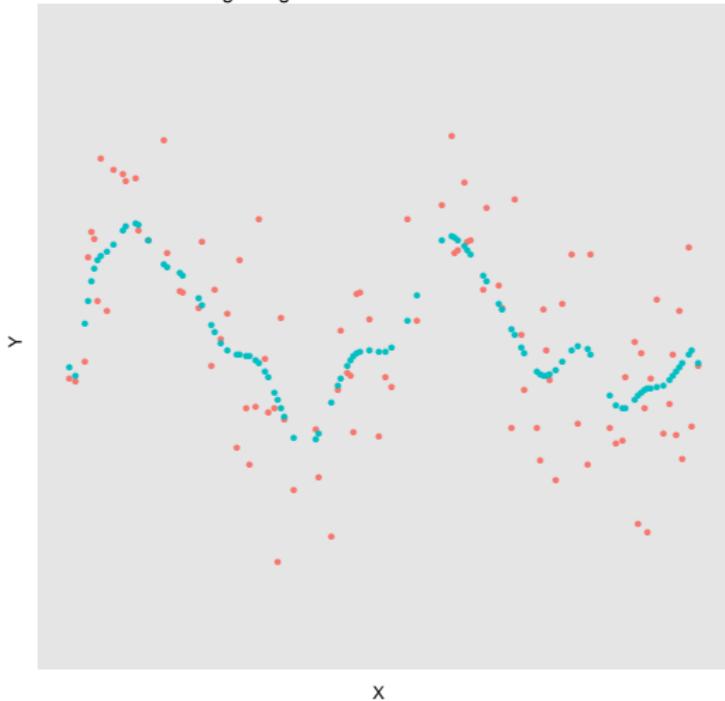
Ridge Regression with Lambda = 0.37



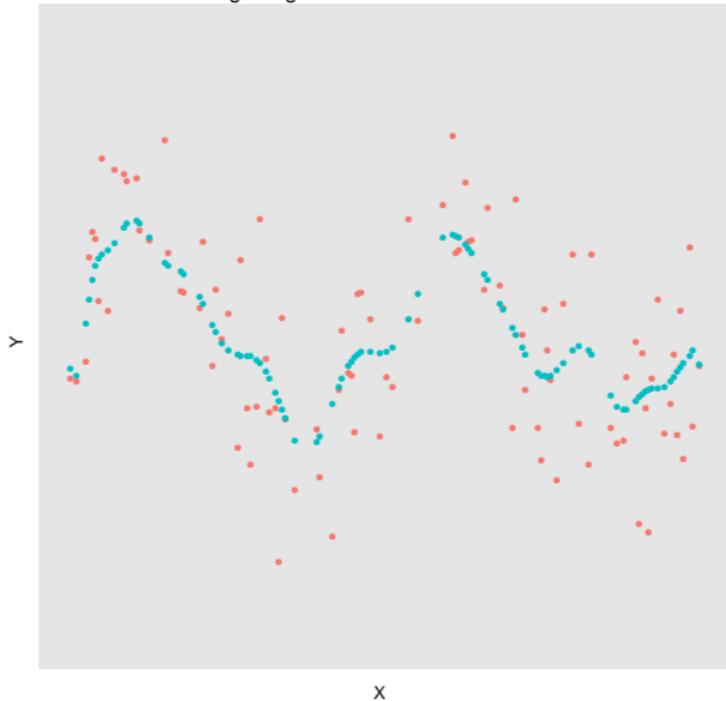
Ridge Regression with Lambda = 0.337



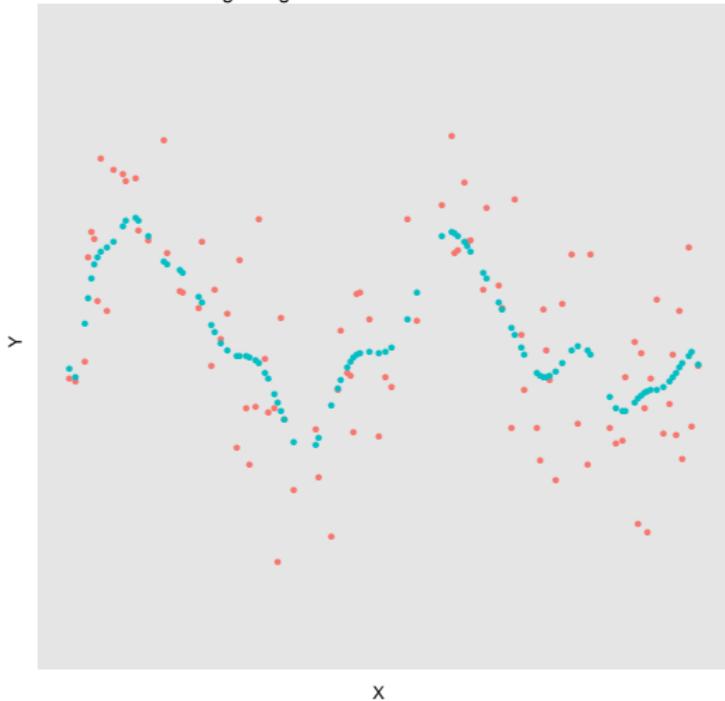
Ridge Regression with Lambda = 0.307



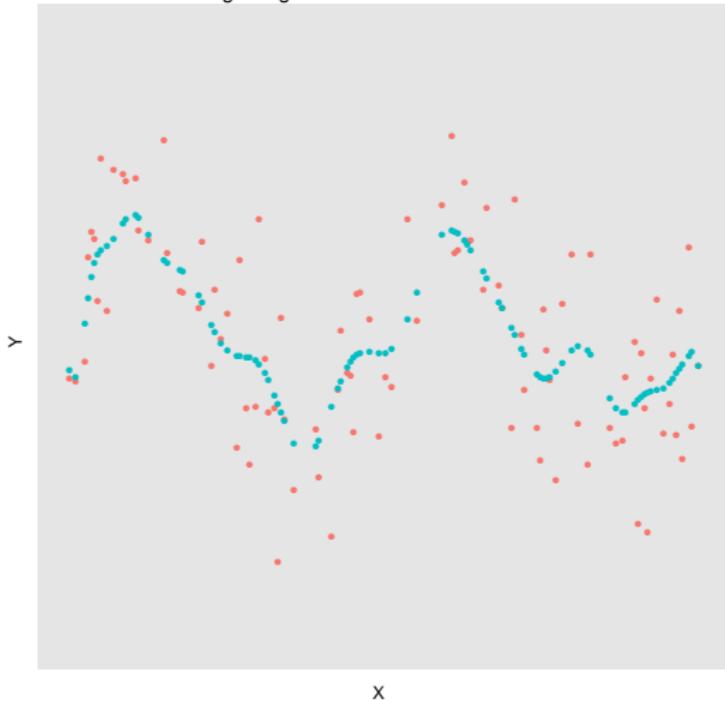
Ridge Regression with Lambda = 0.28



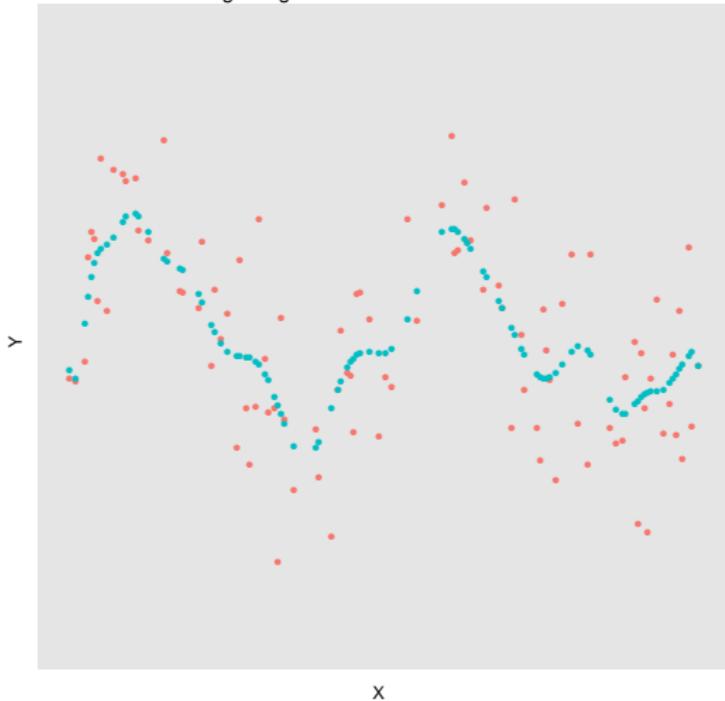
Ridge Regression with Lambda = 0.255



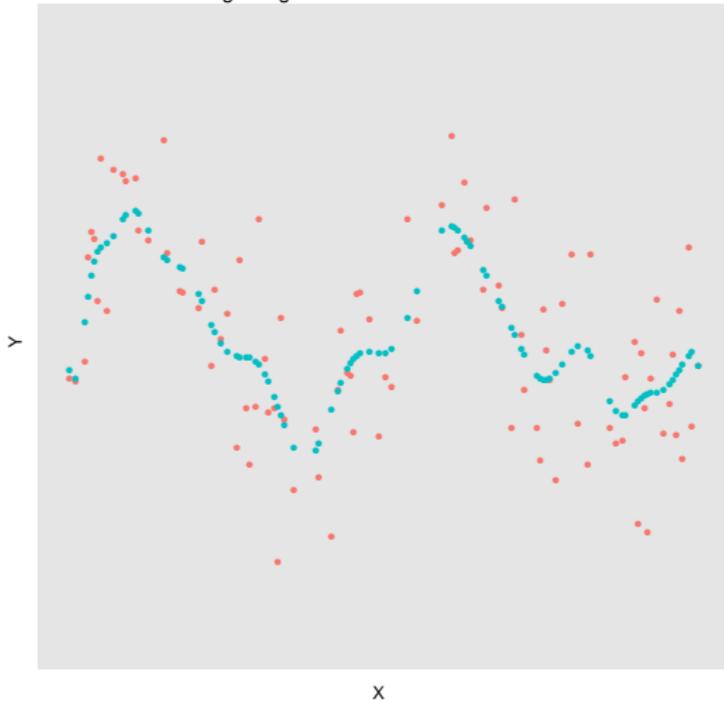
Ridge Regression with Lambda = 0.232



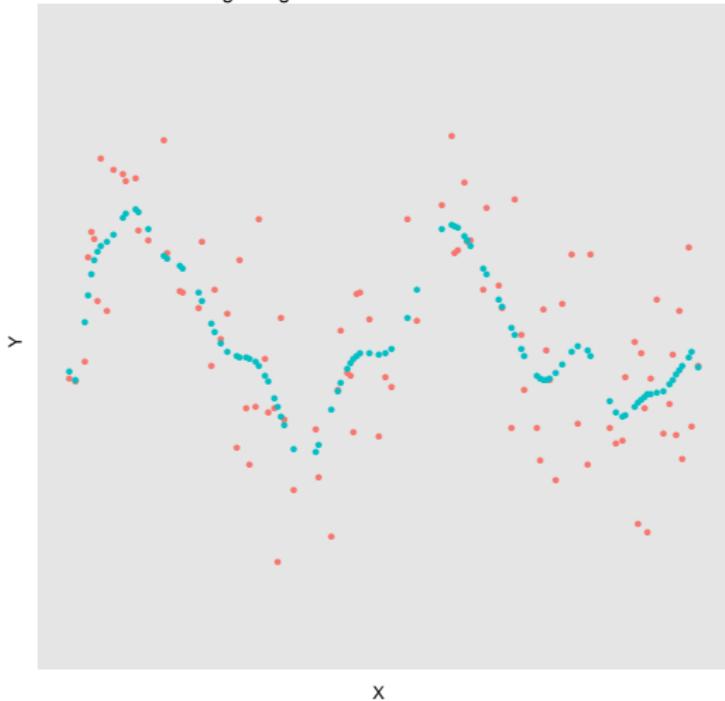
Ridge Regression with Lambda = 0.212



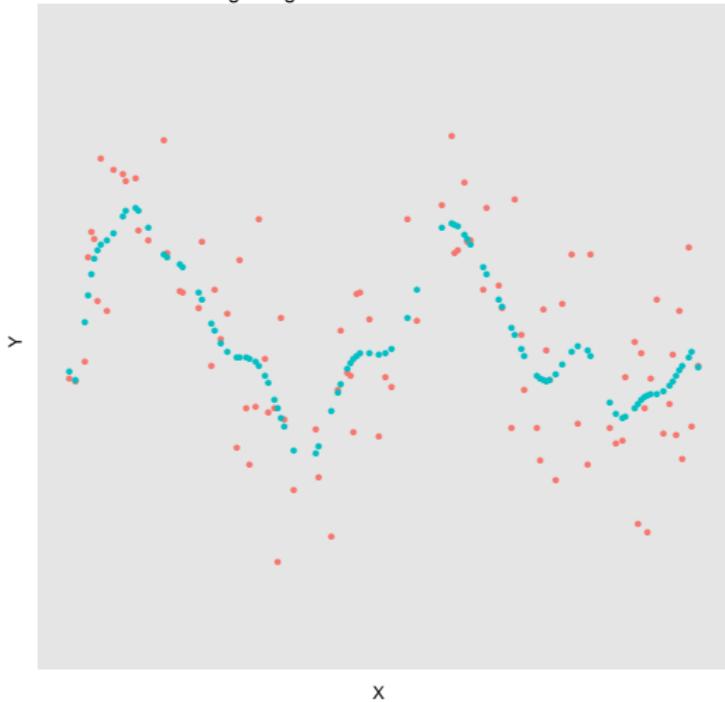
Ridge Regression with Lambda = 0.193



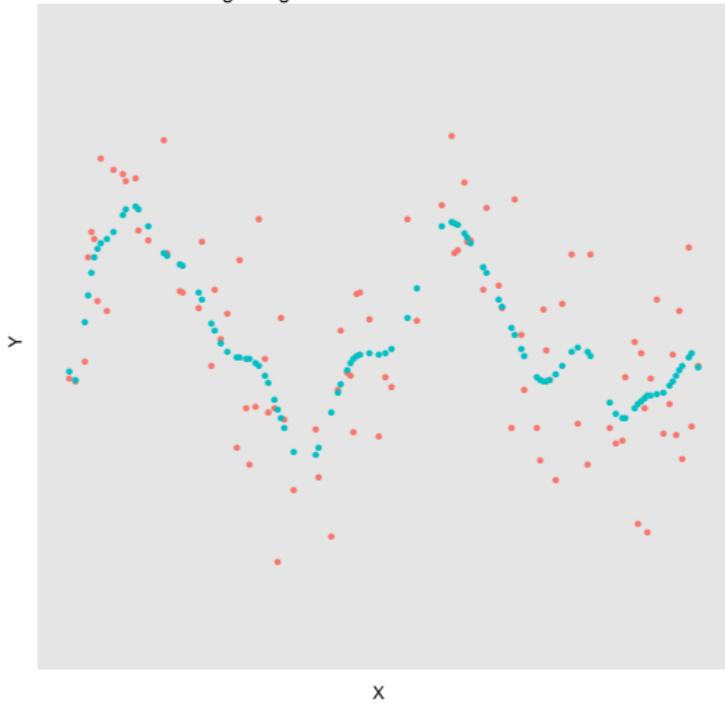
Ridge Regression with Lambda = 0.176



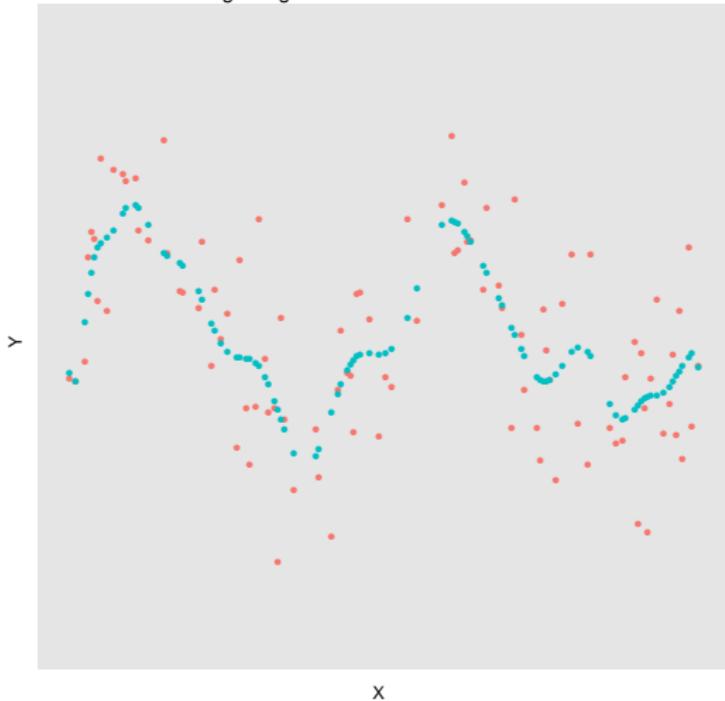
Ridge Regression with Lambda = 0.16



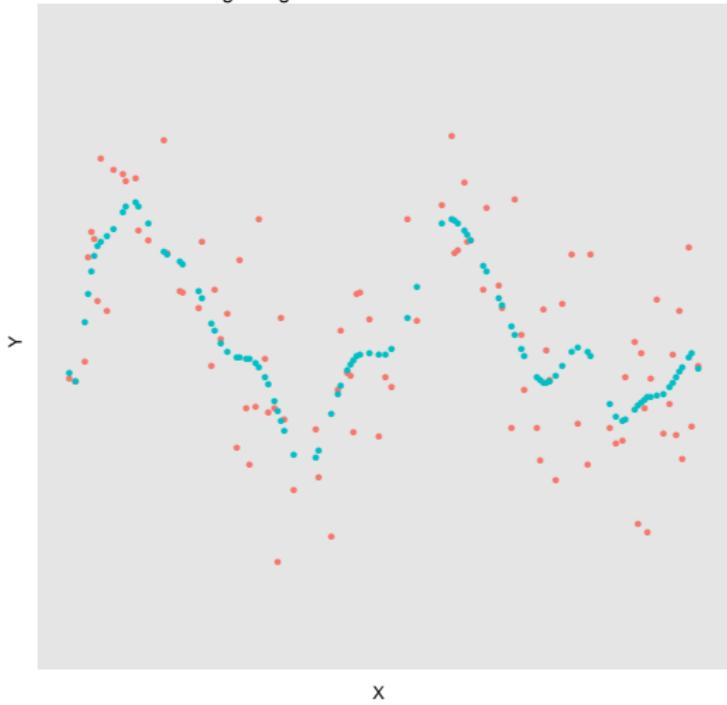
Ridge Regression with Lambda = 0.146



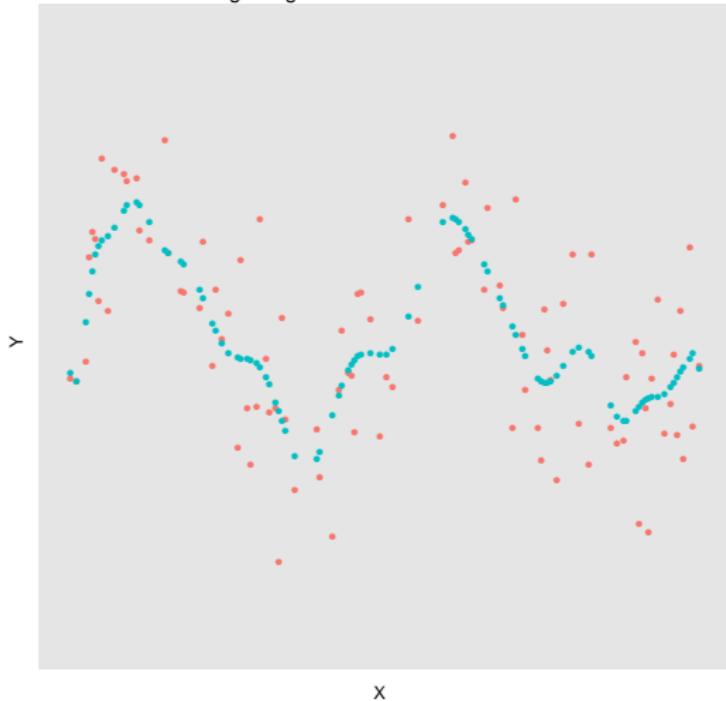
Ridge Regression with Lambda = 0.133



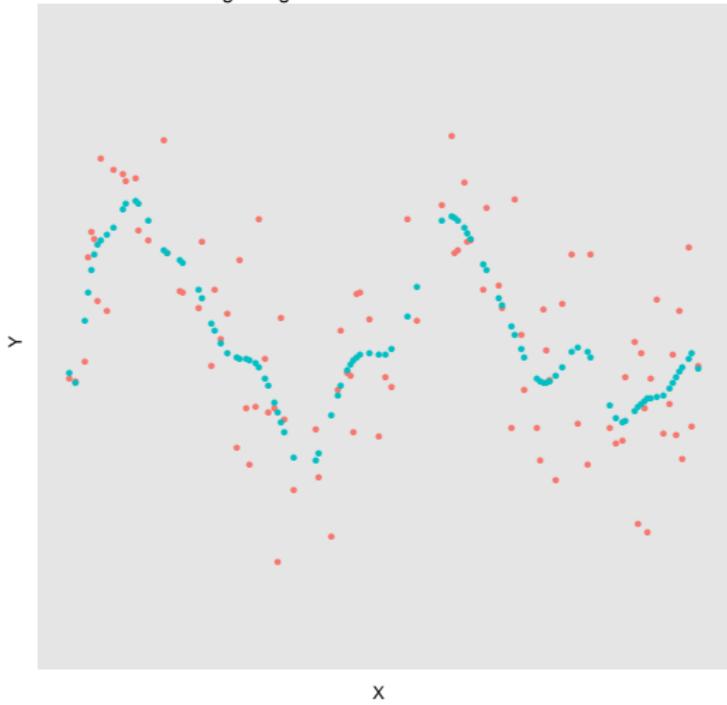
Ridge Regression with Lambda = 0.121



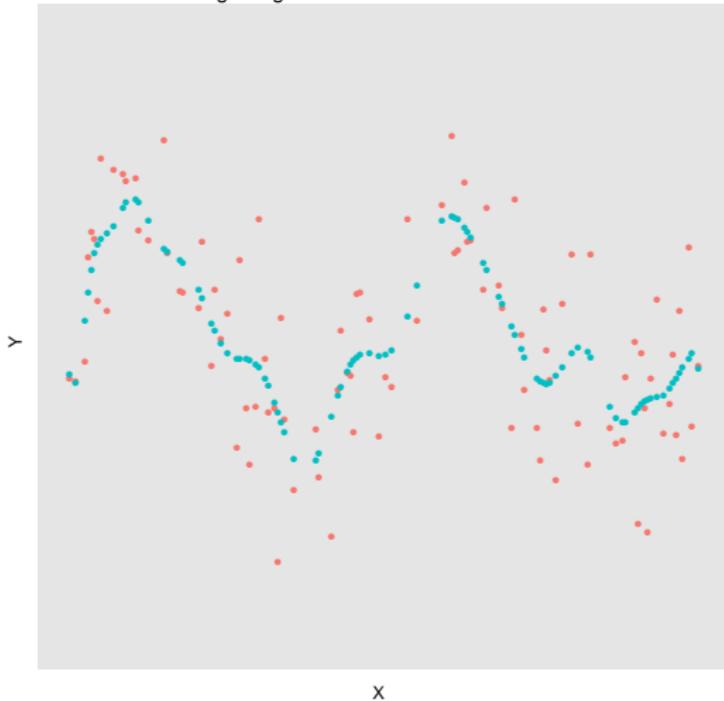
Ridge Regression with Lambda = 0.11



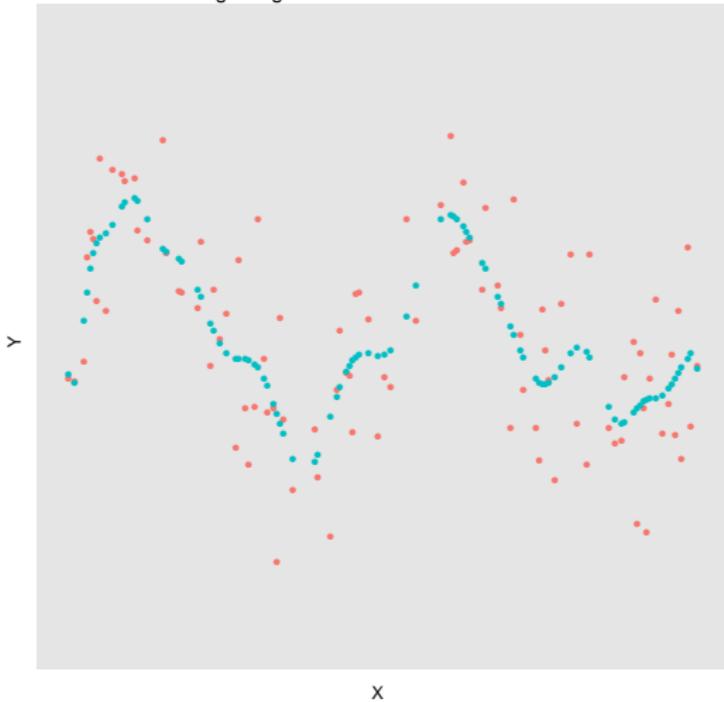
Ridge Regression with Lambda = 0.101



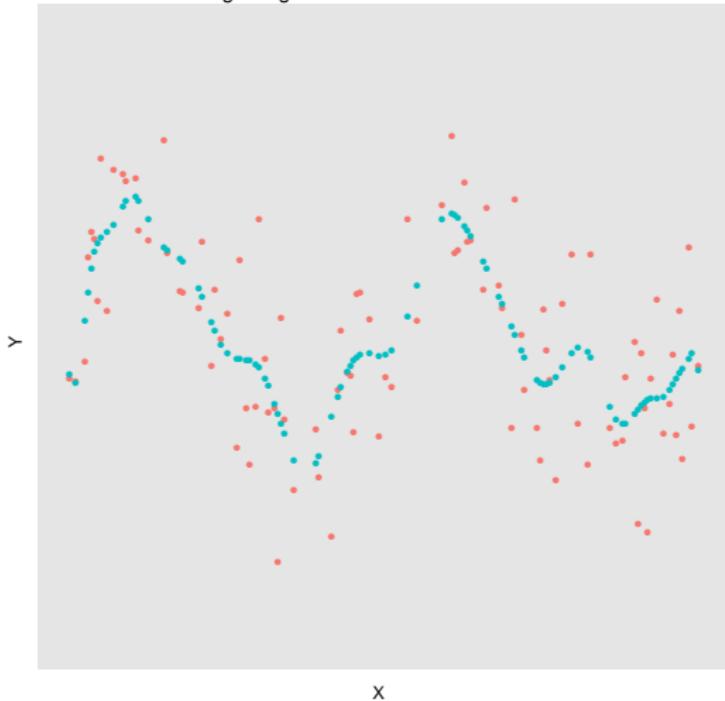
Ridge Regression with Lambda = 0.0916



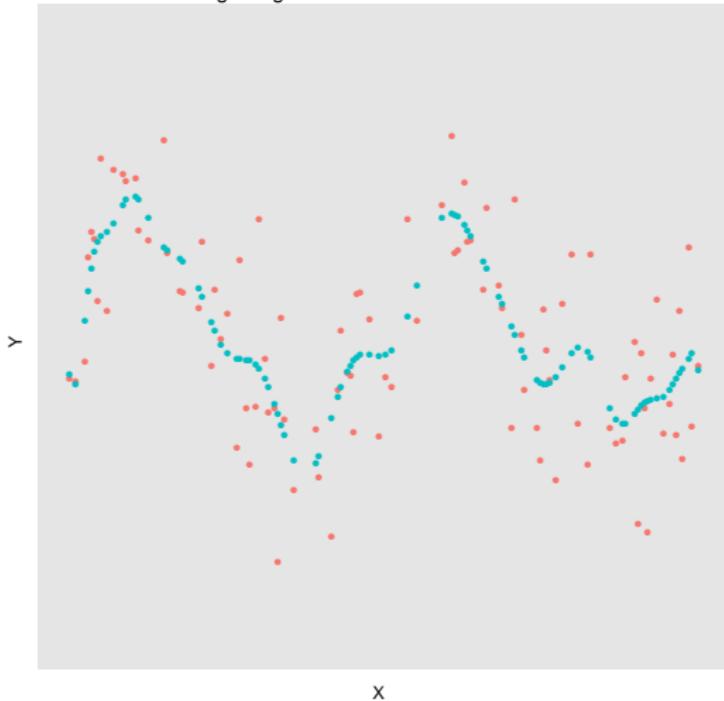
Ridge Regression with Lambda = 0.0835



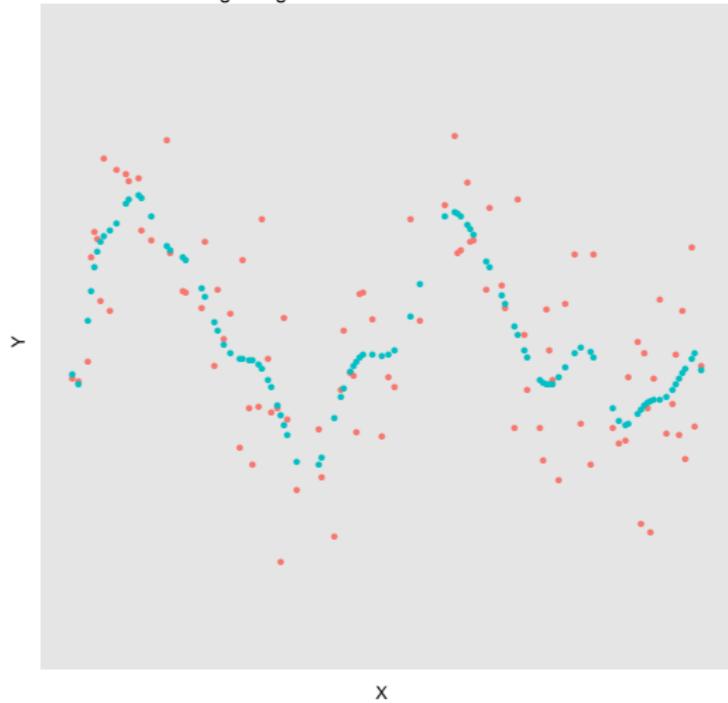
Ridge Regression with Lambda = 0.076



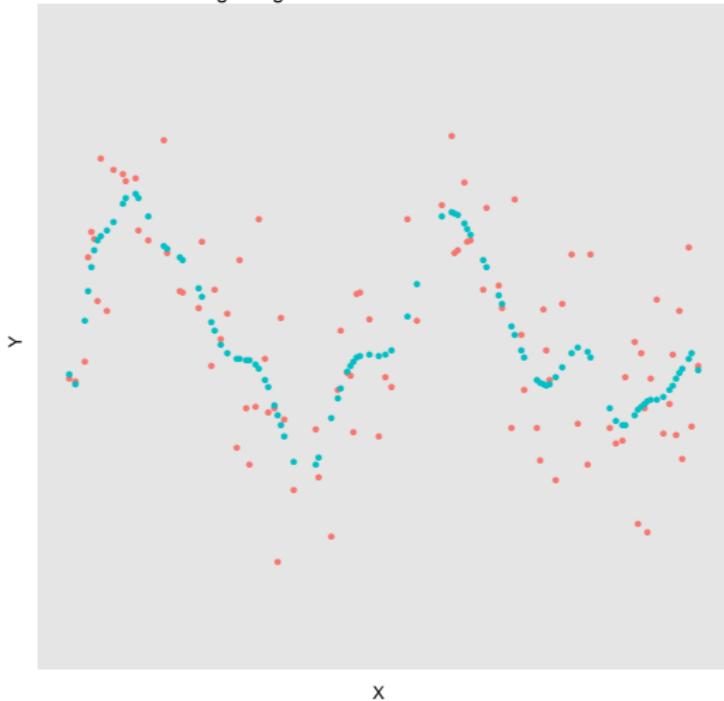
Ridge Regression with Lambda = 0.0693



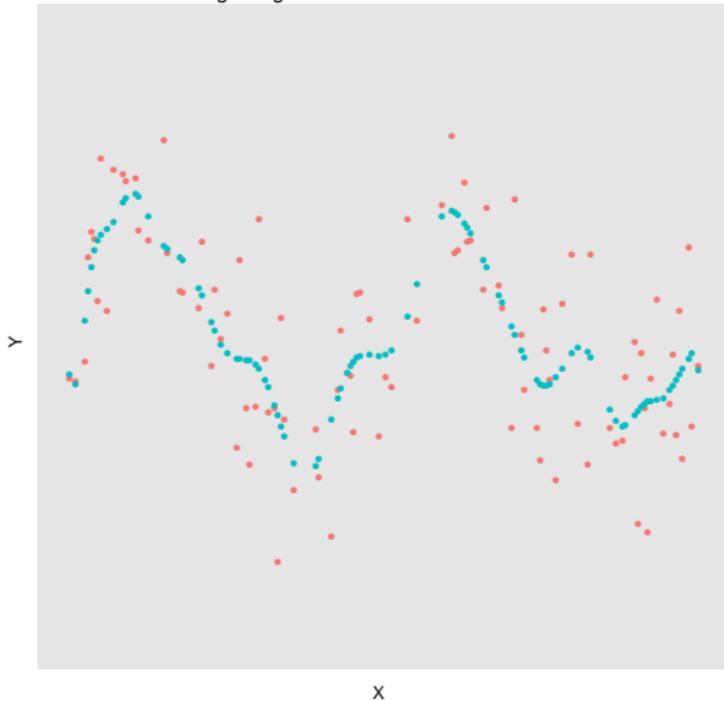
Ridge Regression with Lambda = 0.0631



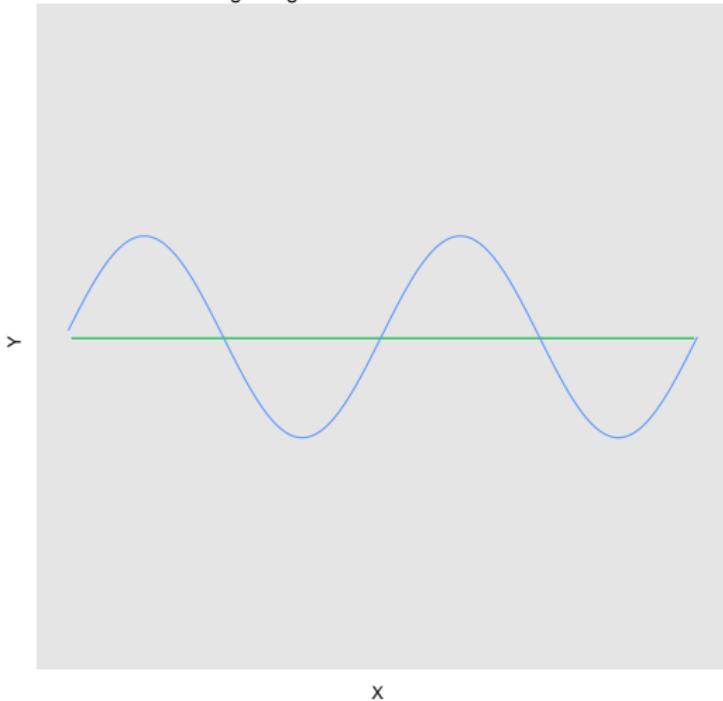
Ridge Regression with Lambda = 0.0575



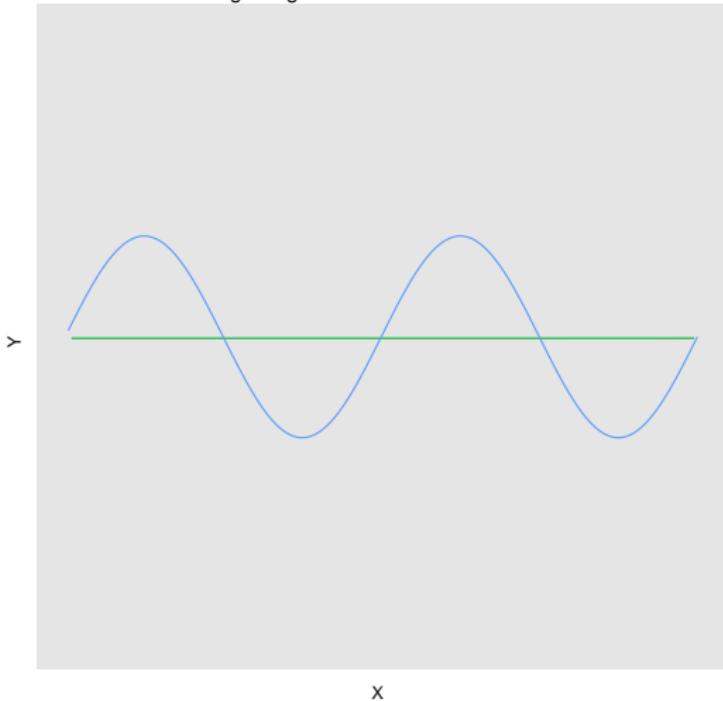
Ridge Regression with Lambda = 0.0524



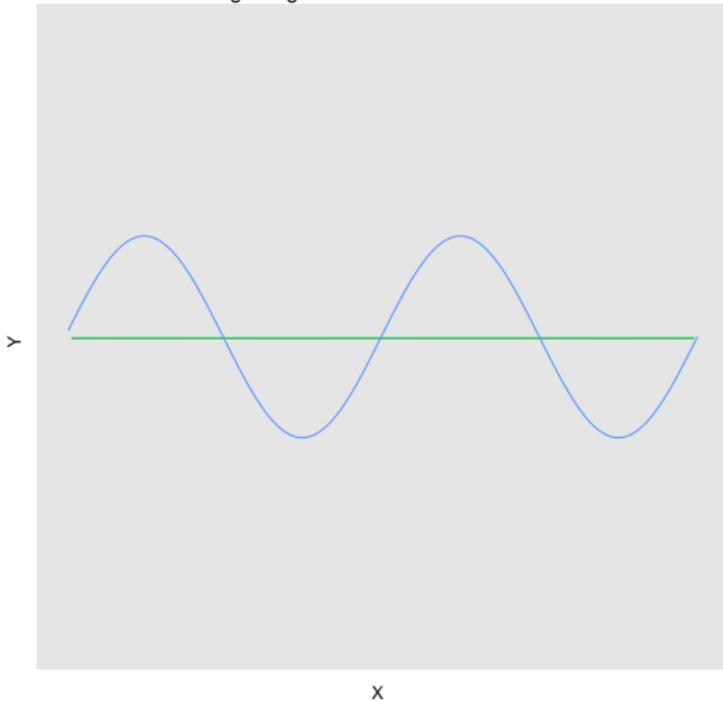
Ridge Regression with Lambda = 524



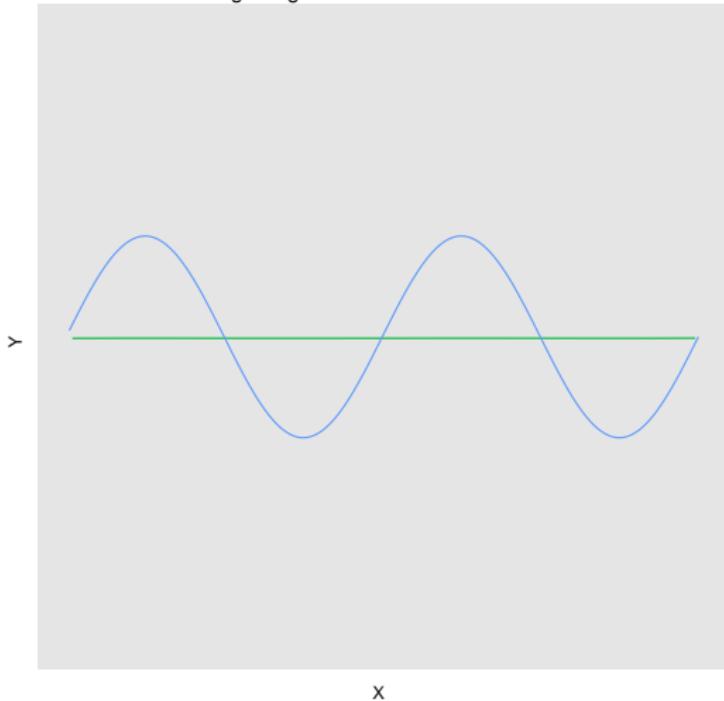
Ridge Regression with Lambda = 478



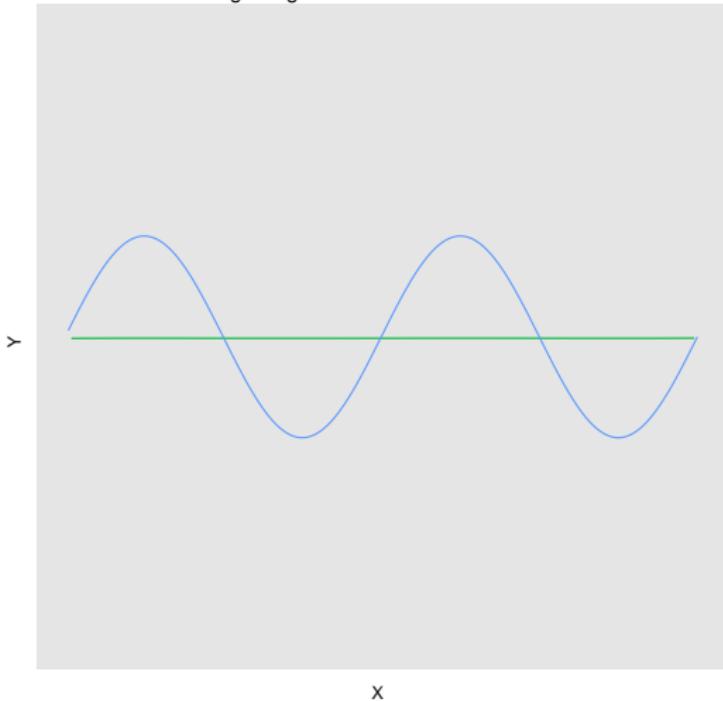
Ridge Regression with Lambda = 435



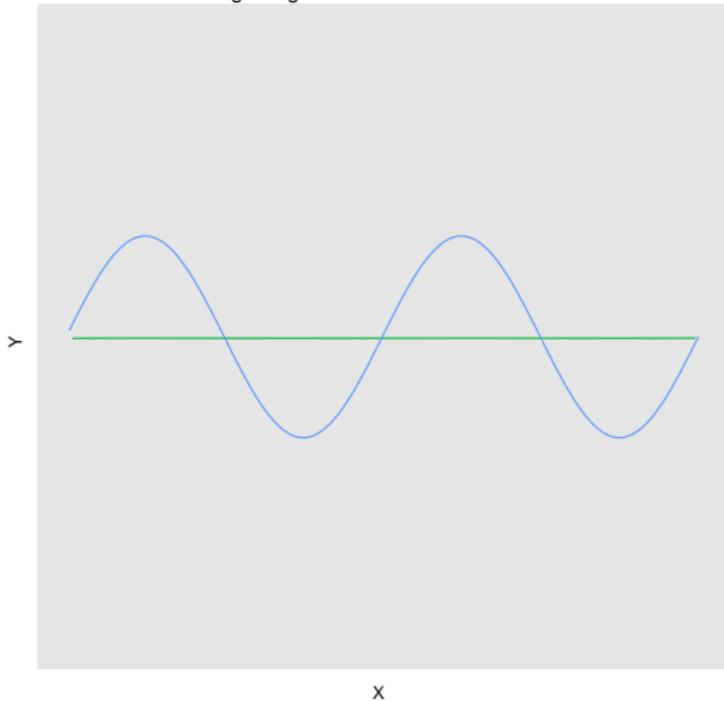
Ridge Regression with Lambda = 396



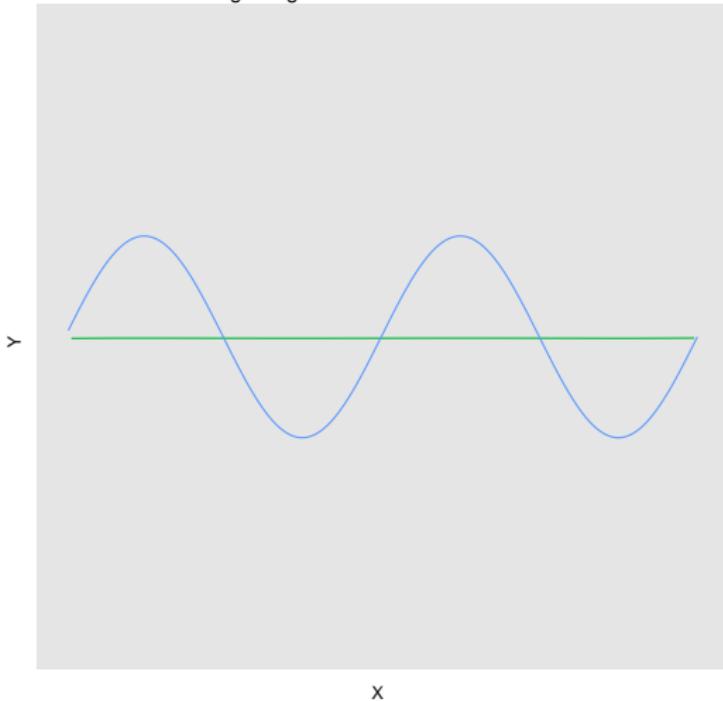
Ridge Regression with Lambda = 361



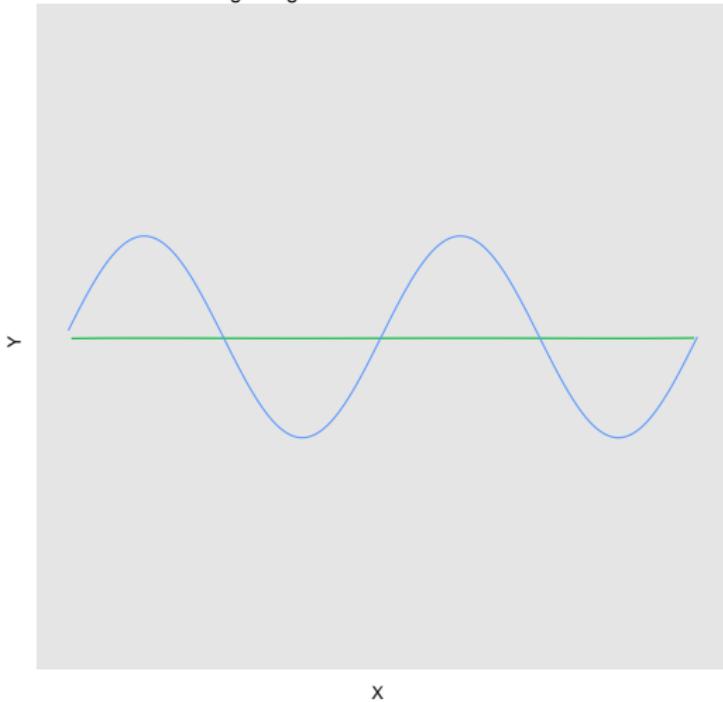
Ridge Regression with Lambda = 329



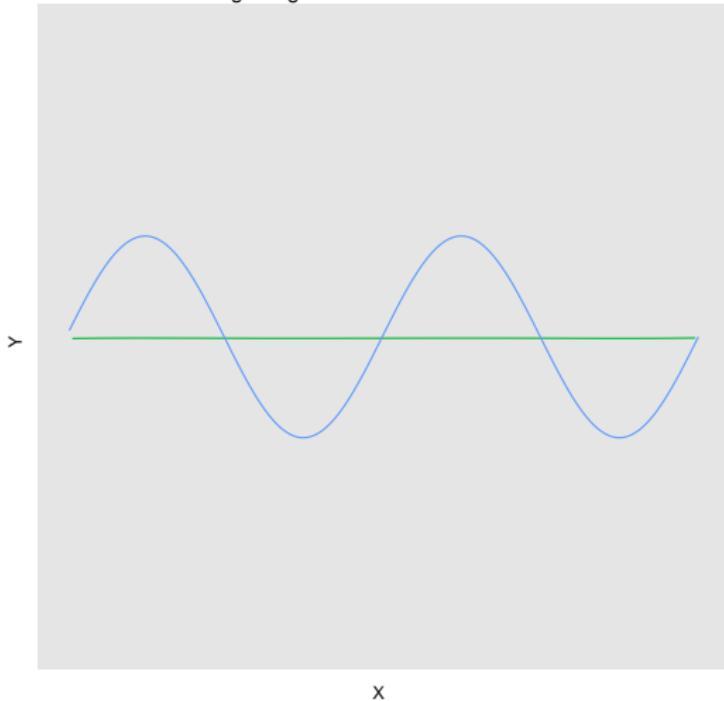
Ridge Regression with Lambda = 300



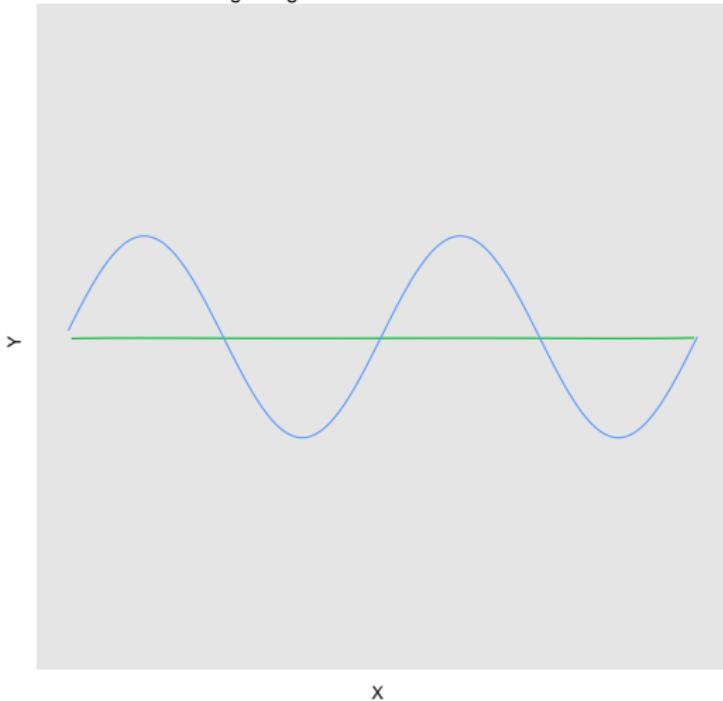
Ridge Regression with Lambda = 273



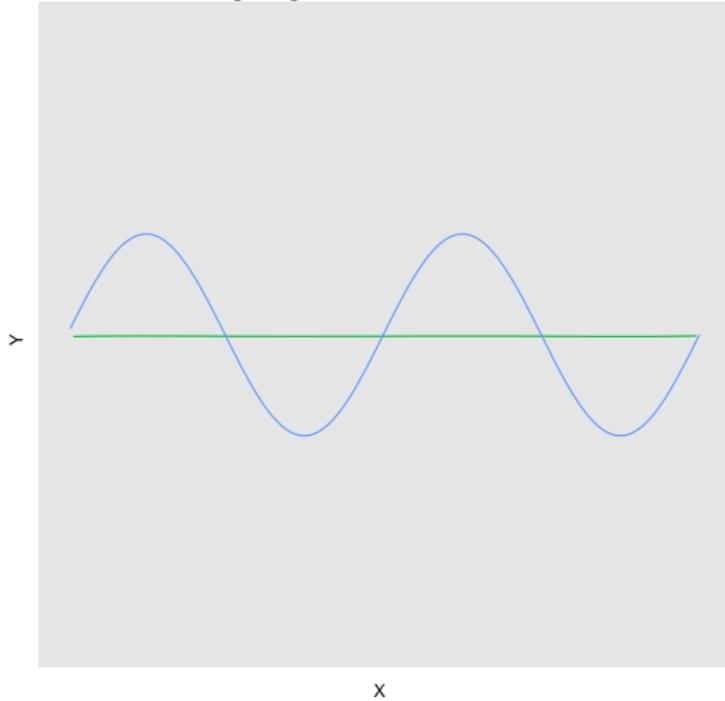
Ridge Regression with Lambda = 249



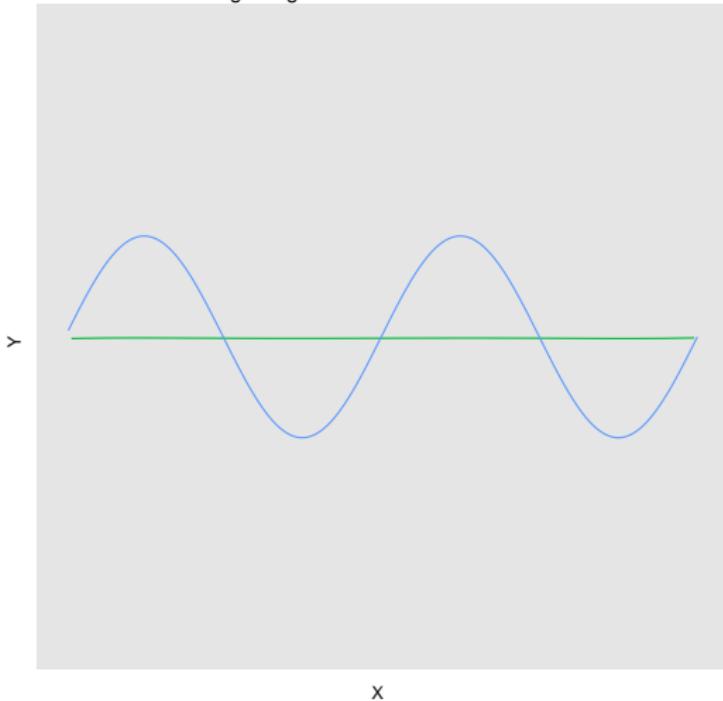
Ridge Regression with Lambda = 227



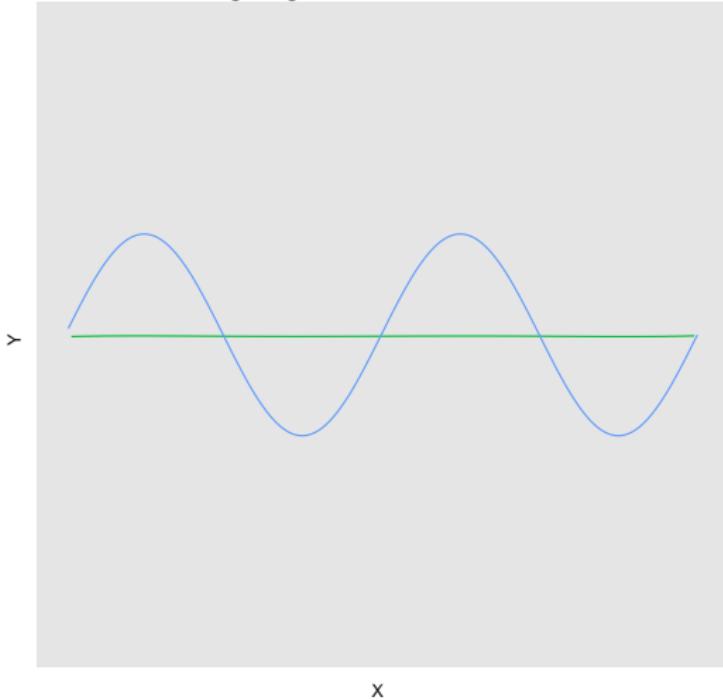
Ridge Regression with Lambda = 207



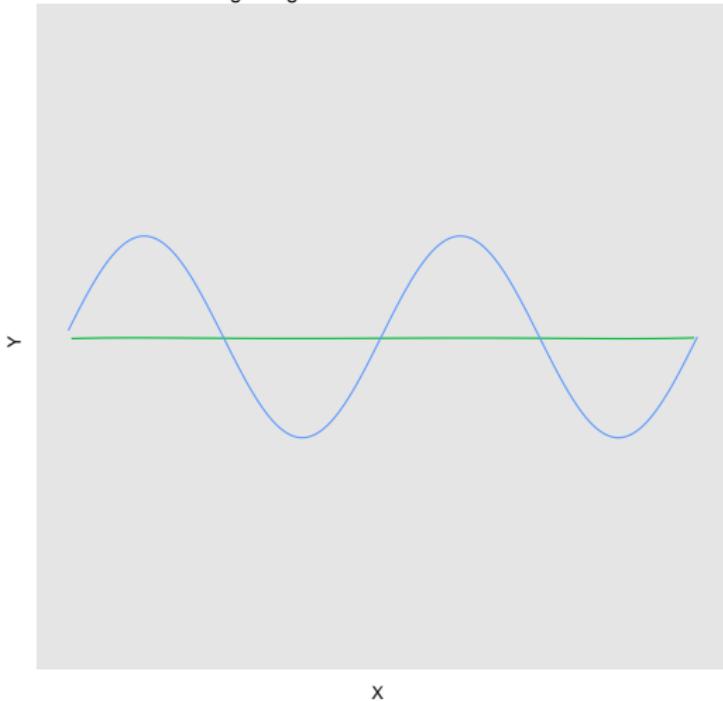
Ridge Regression with Lambda = 188



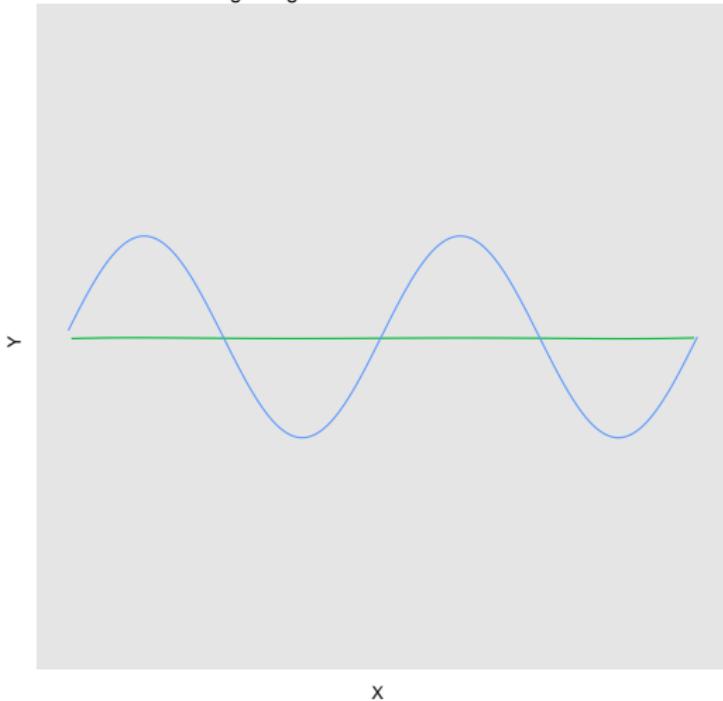
Ridge Regression with Lambda = 172



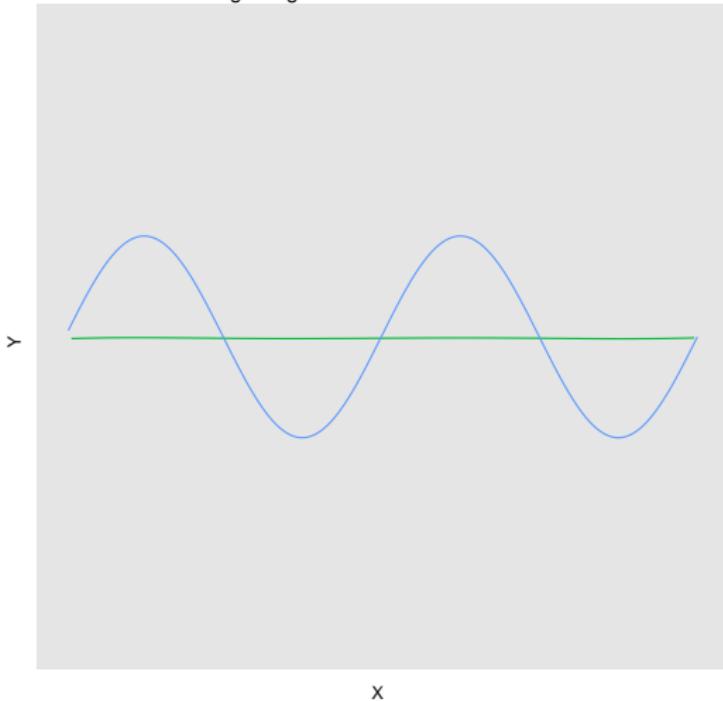
Ridge Regression with Lambda = 156



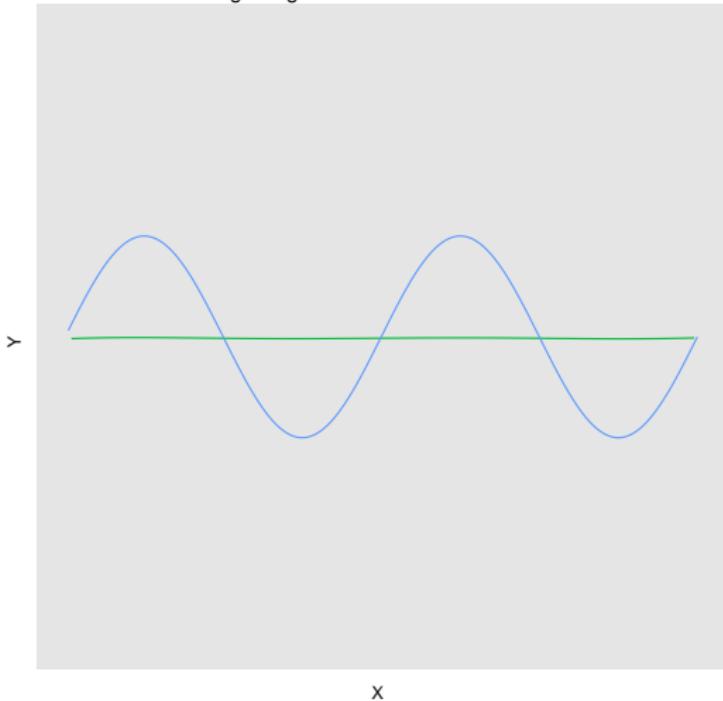
Ridge Regression with Lambda = 142



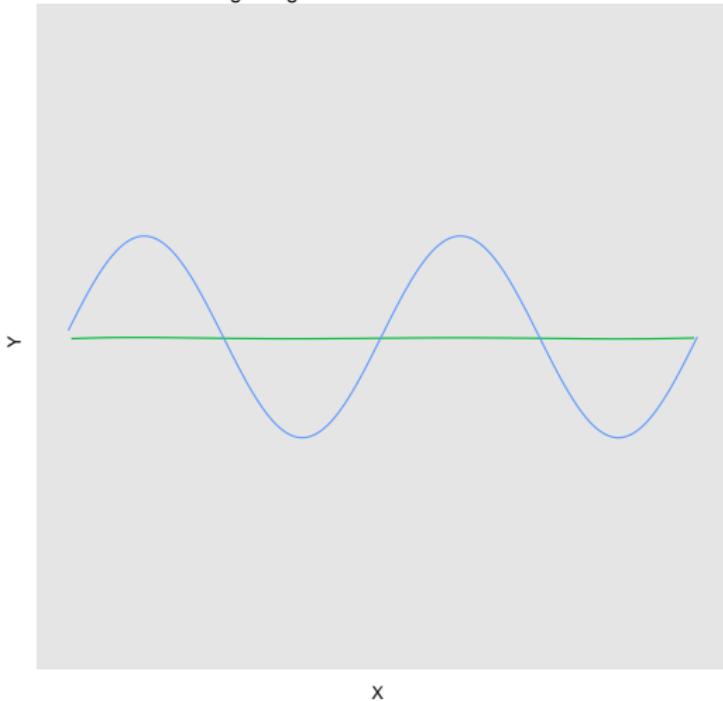
Ridge Regression with Lambda = 130



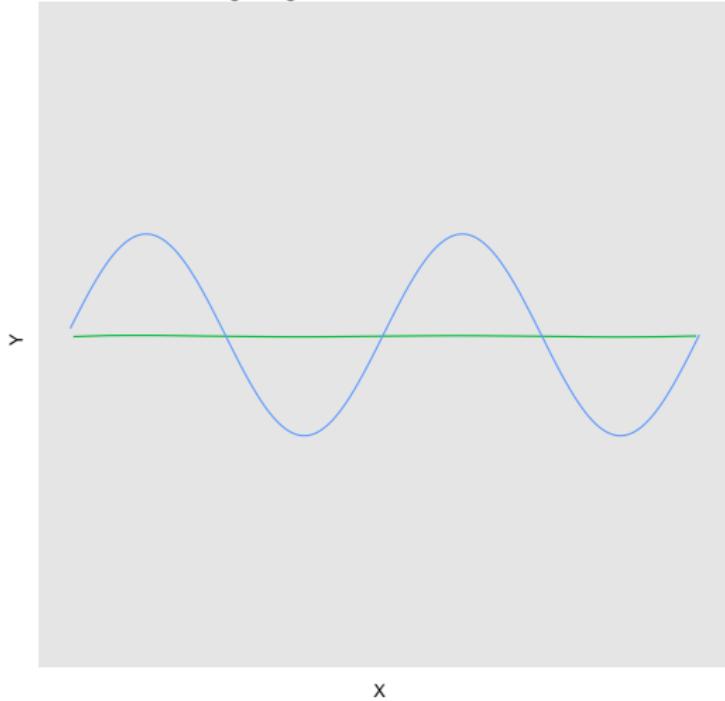
Ridge Regression with Lambda = 118



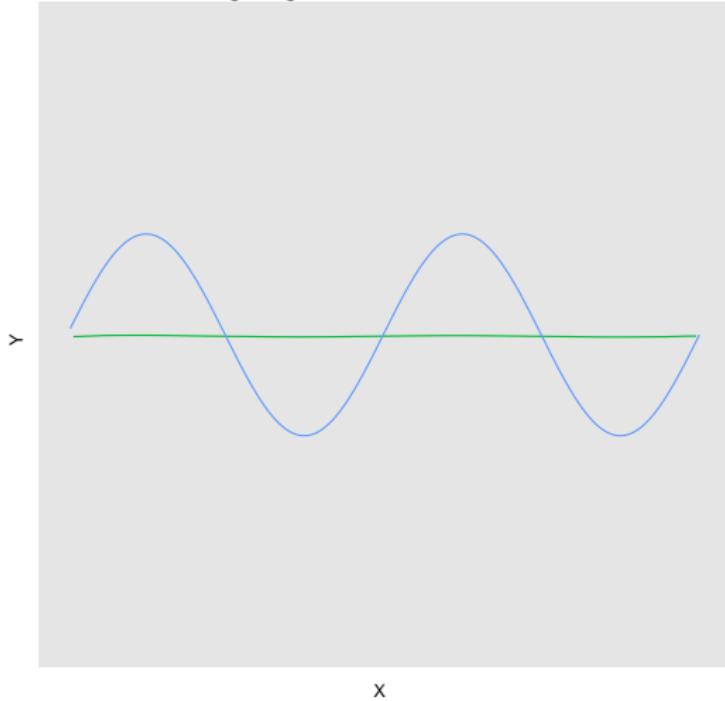
Ridge Regression with Lambda = 108



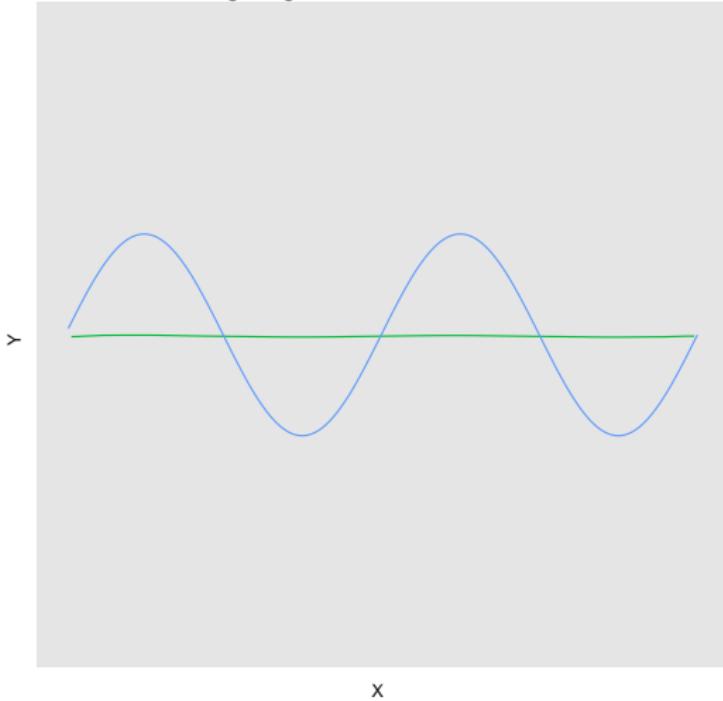
Ridge Regression with Lambda = 98.2



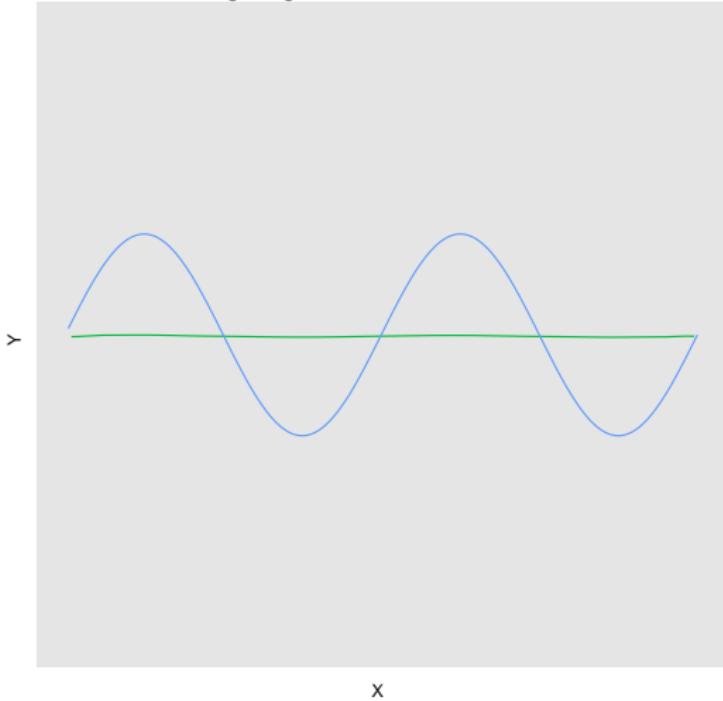
Ridge Regression with Lambda = 89.5



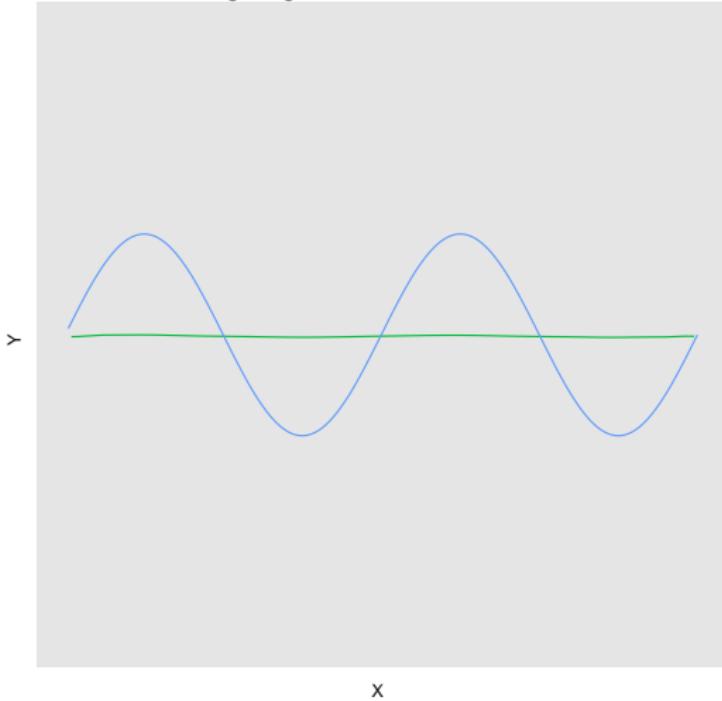
Ridge Regression with Lambda = 81.5



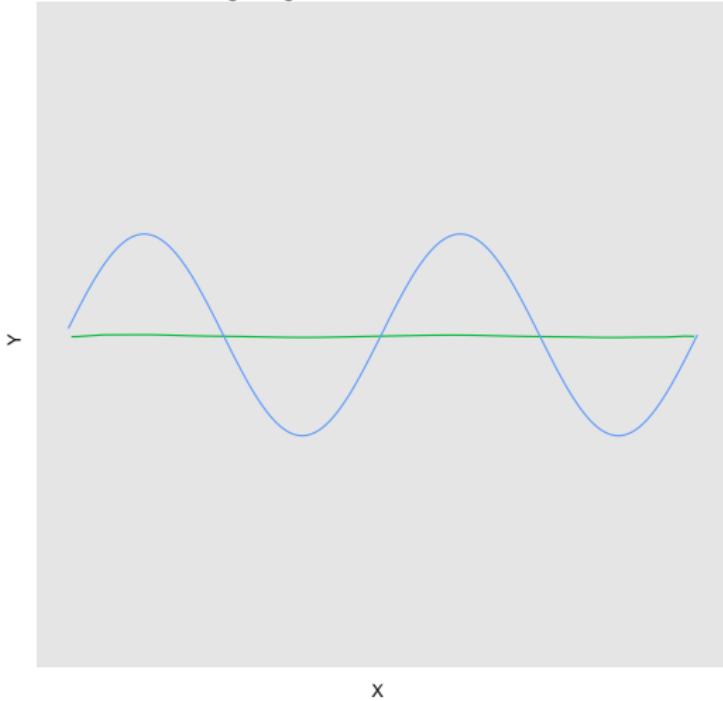
Ridge Regression with Lambda = 74.3



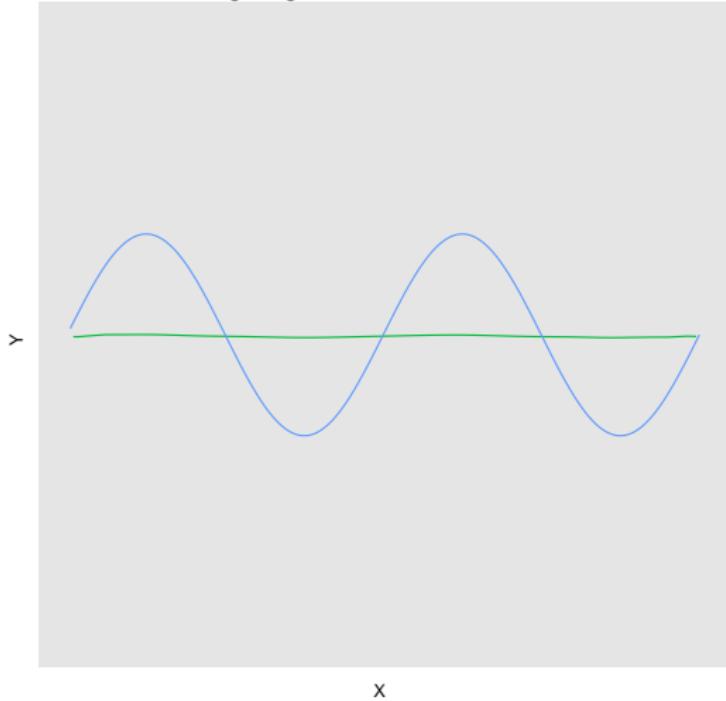
Ridge Regression with Lambda = 67.7



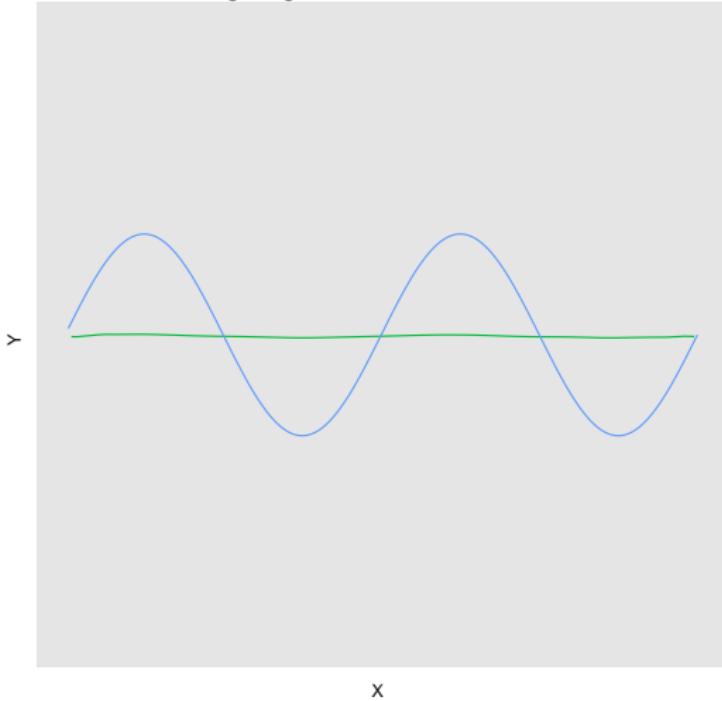
Ridge Regression with Lambda = 61.7



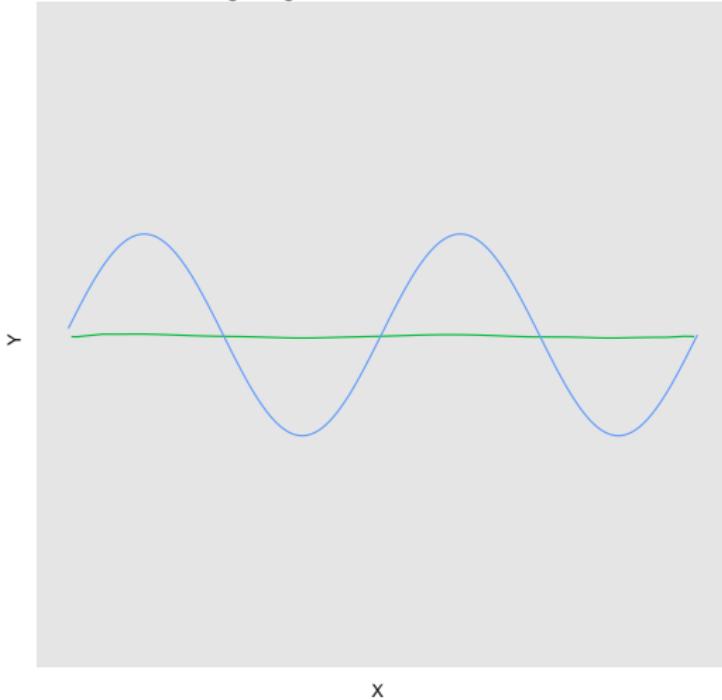
Ridge Regression with Lambda = 56.2



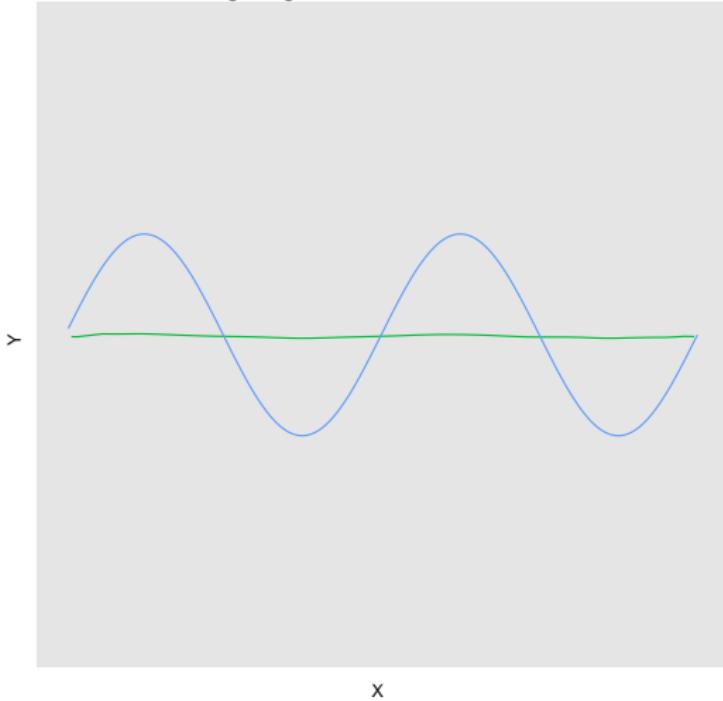
Ridge Regression with Lambda = 51.2



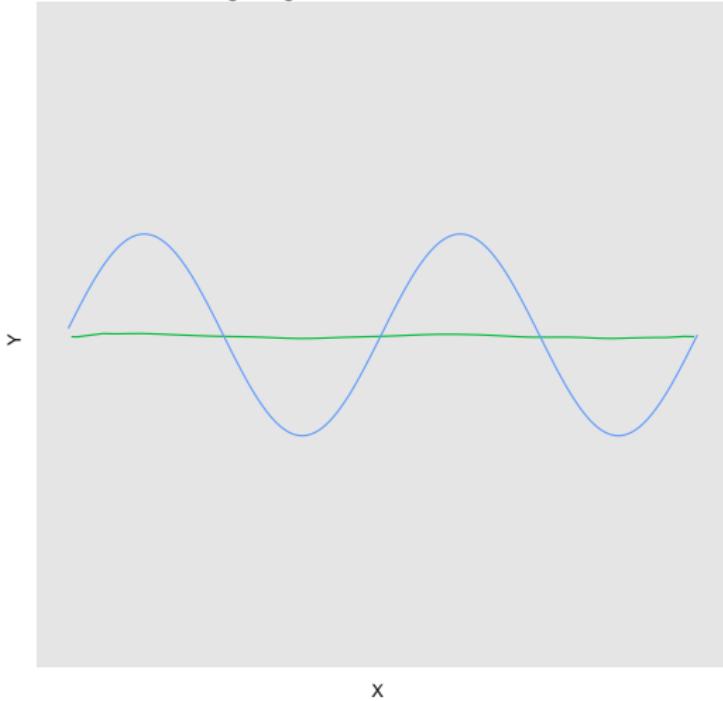
Ridge Regression with Lambda = 46.7



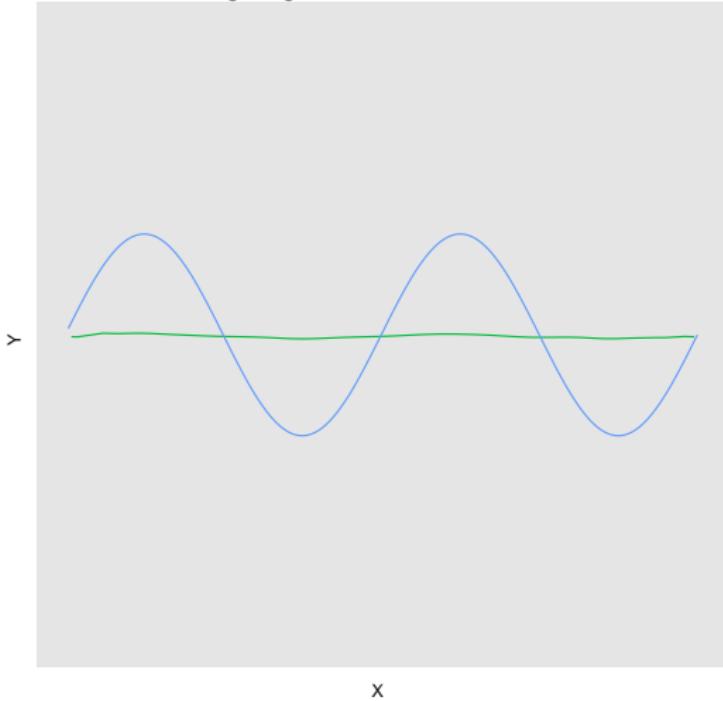
Ridge Regression with Lambda = 42.5



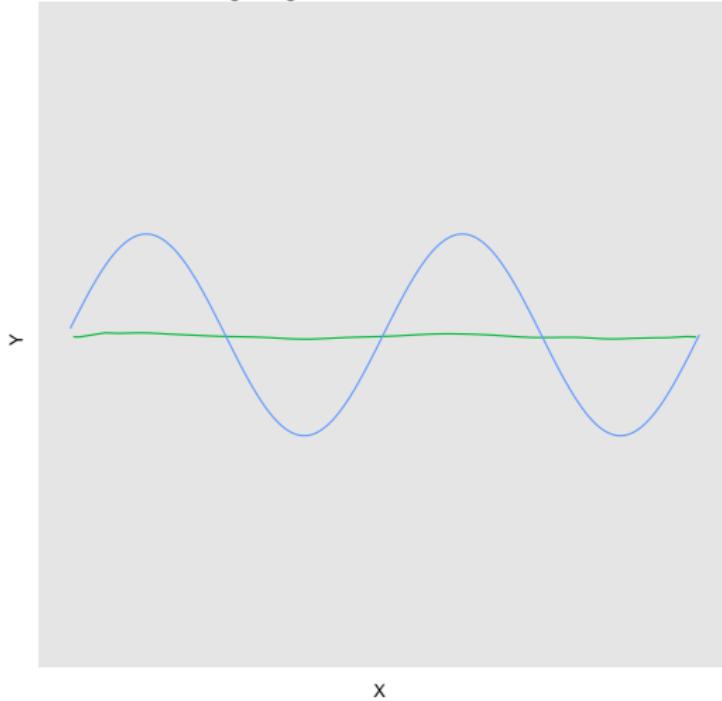
Ridge Regression with Lambda = 38.7



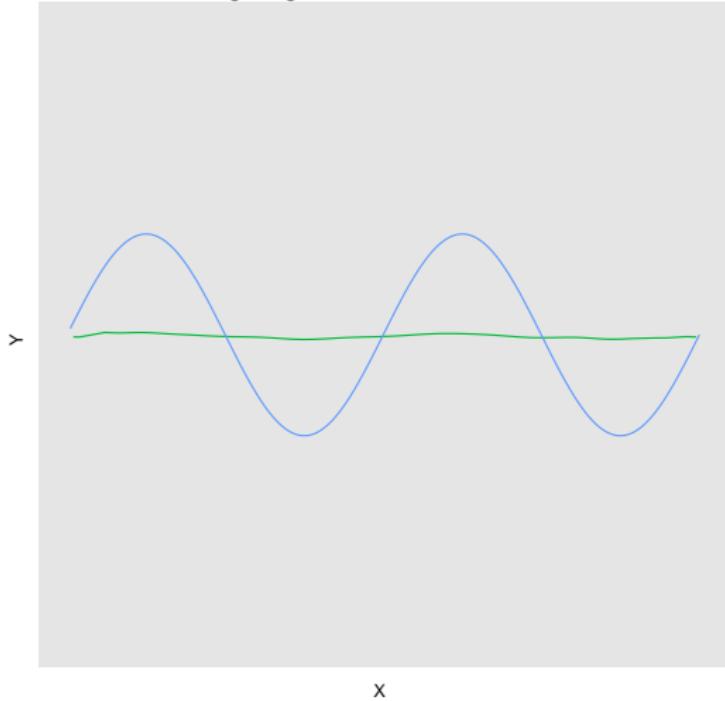
Ridge Regression with Lambda = 35.3



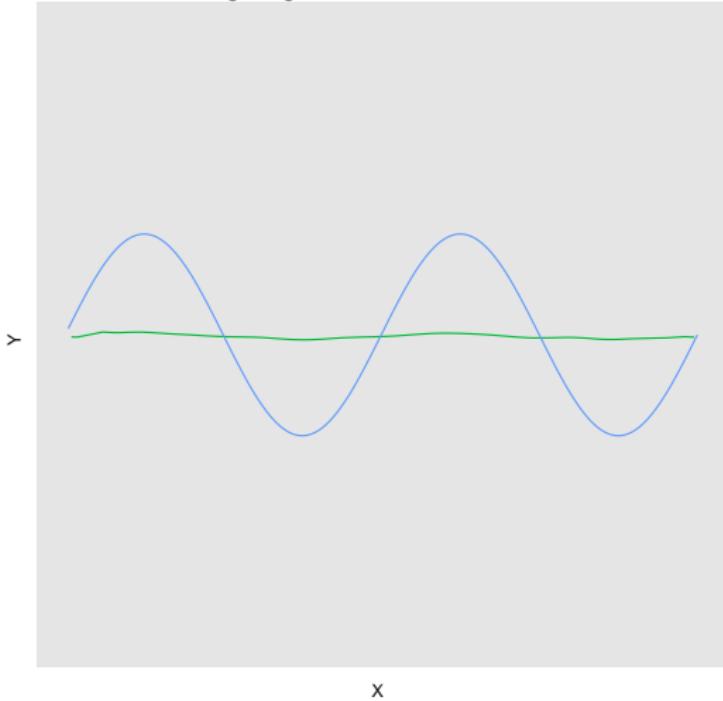
Ridge Regression with Lambda = 32.2



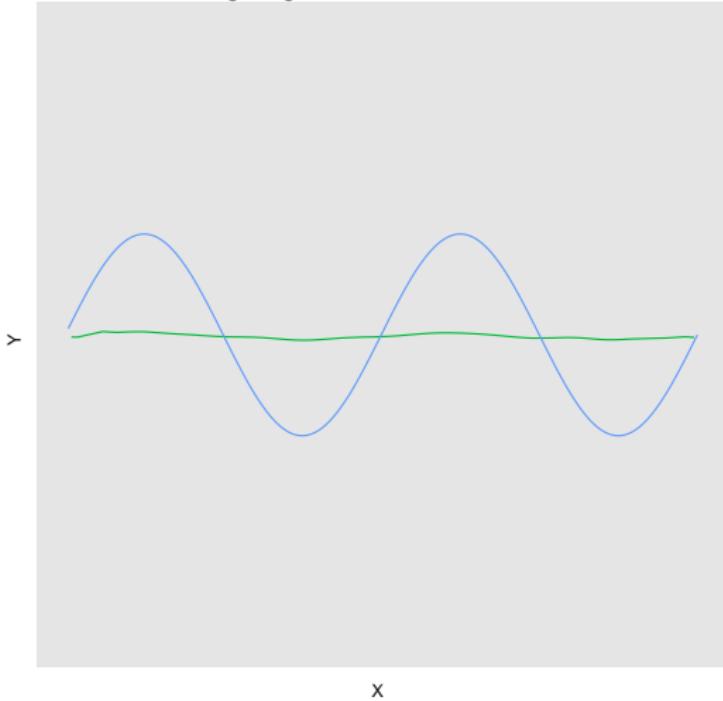
Ridge Regression with Lambda = 29.3



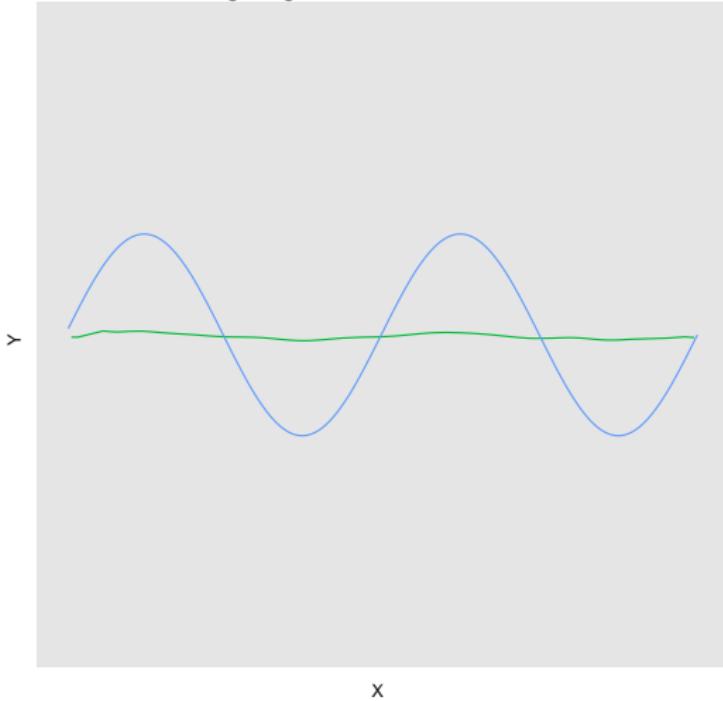
Ridge Regression with Lambda = 26.7



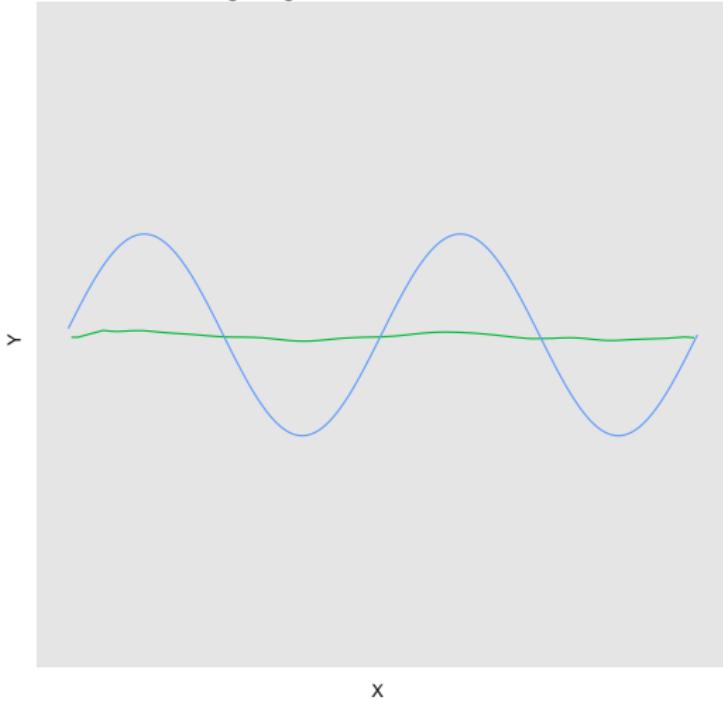
Ridge Regression with Lambda = 24.3



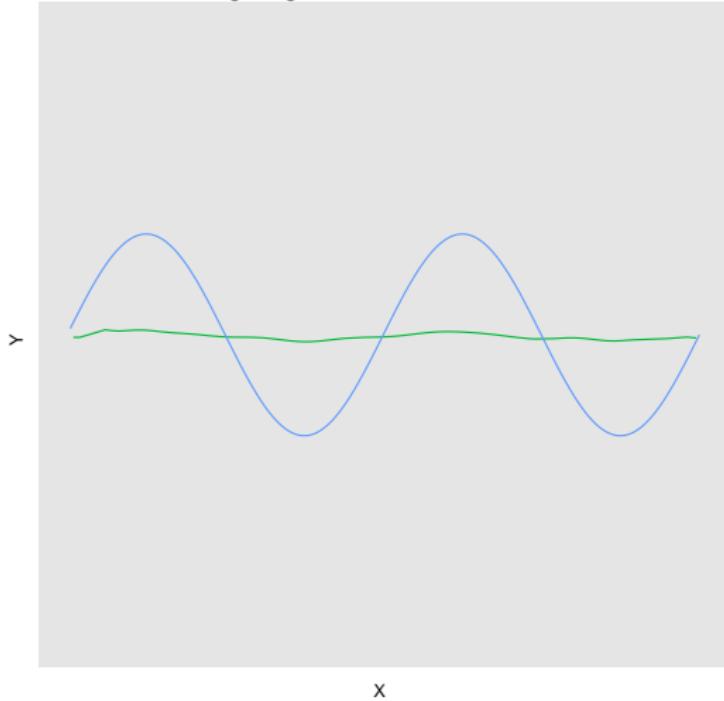
Ridge Regression with Lambda = 22.2



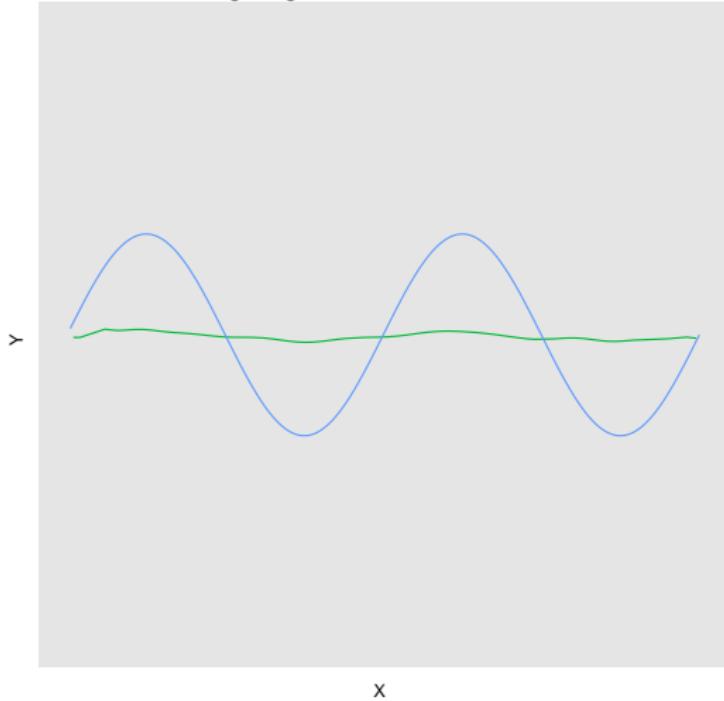
Ridge Regression with Lambda = 20.2



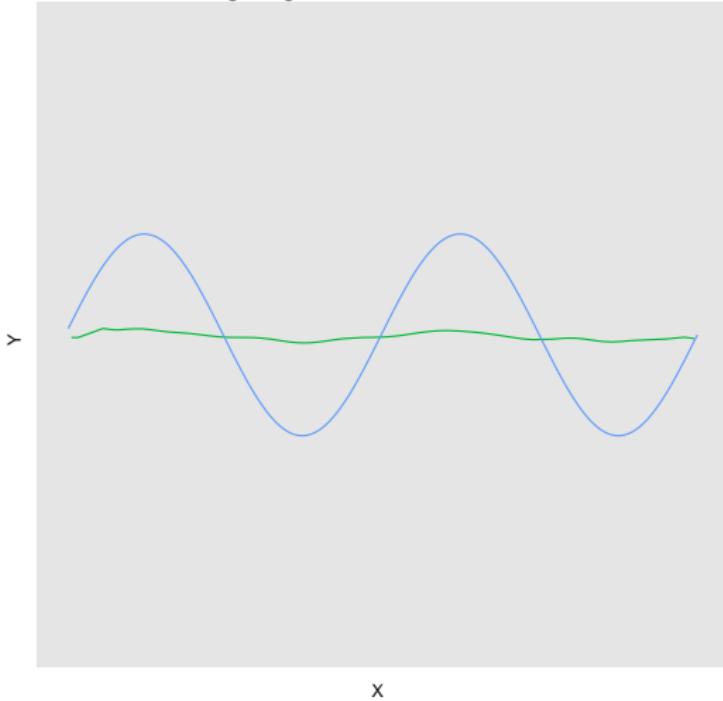
Ridge Regression with Lambda = 18.4



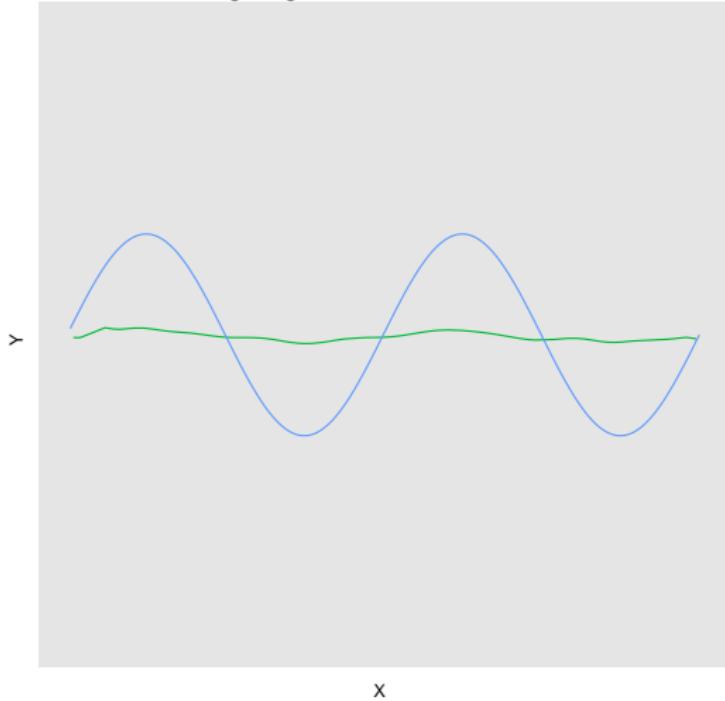
Ridge Regression with Lambda = 16.8



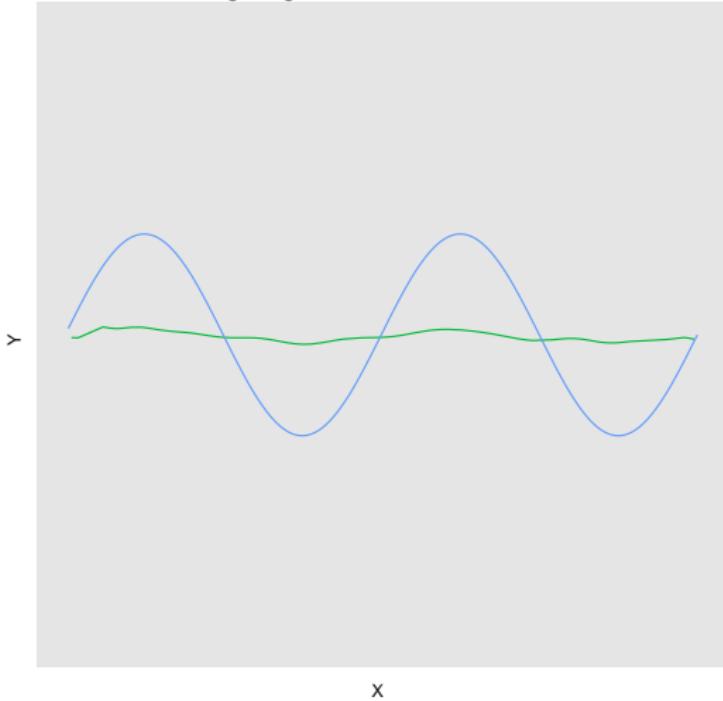
Ridge Regression with Lambda = 15.3



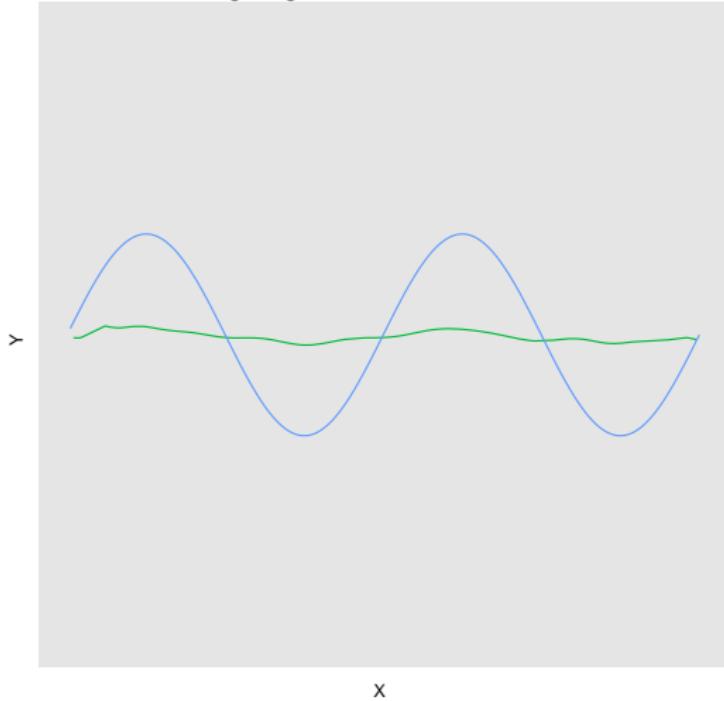
Ridge Regression with Lambda = 13.9



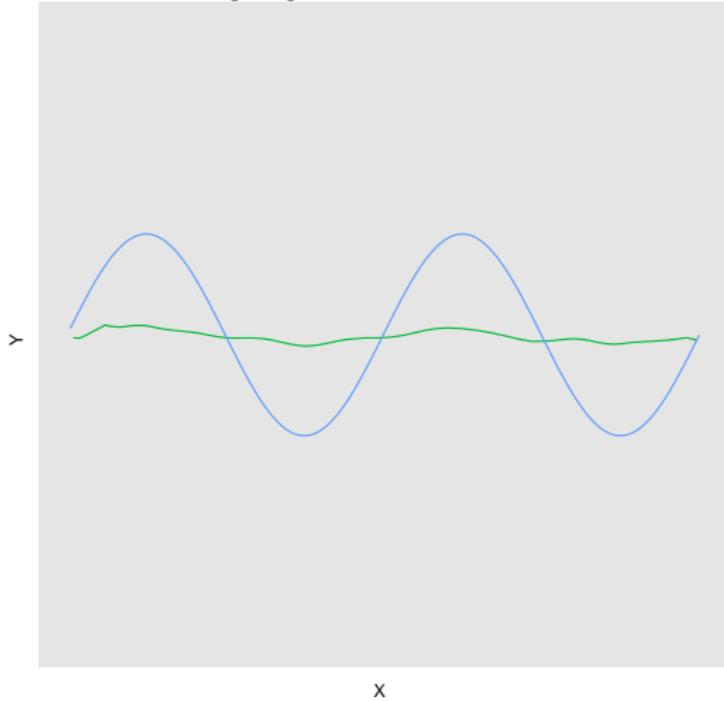
Ridge Regression with Lambda = 12.7



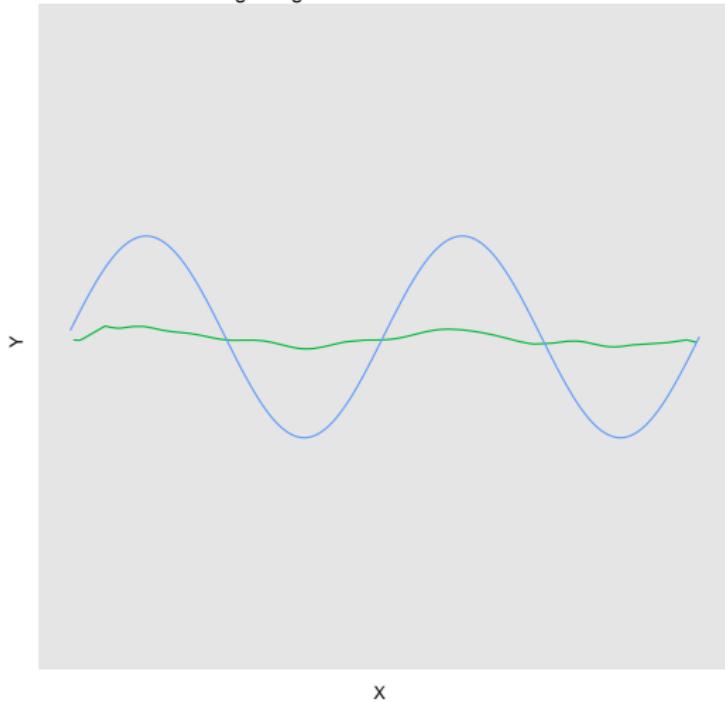
Ridge Regression with Lambda = 11.6



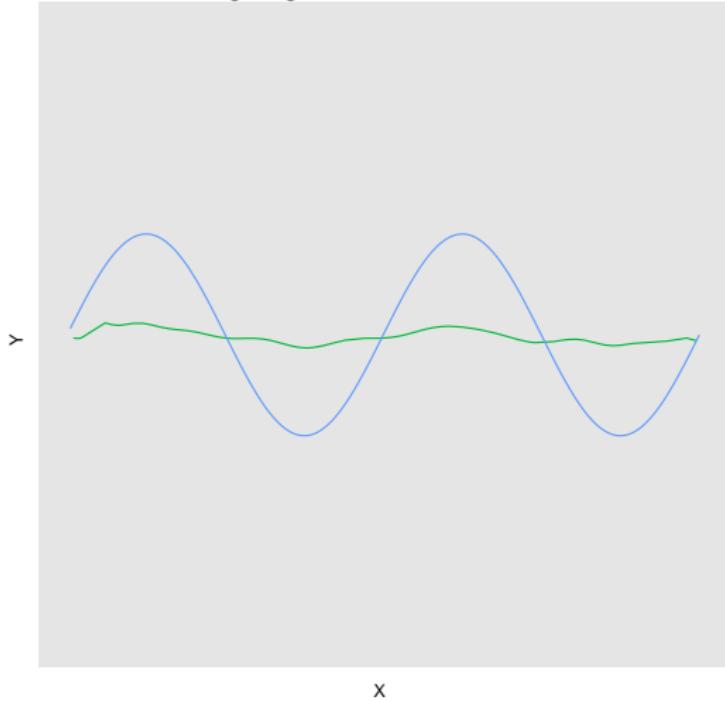
Ridge Regression with Lambda = 10.5



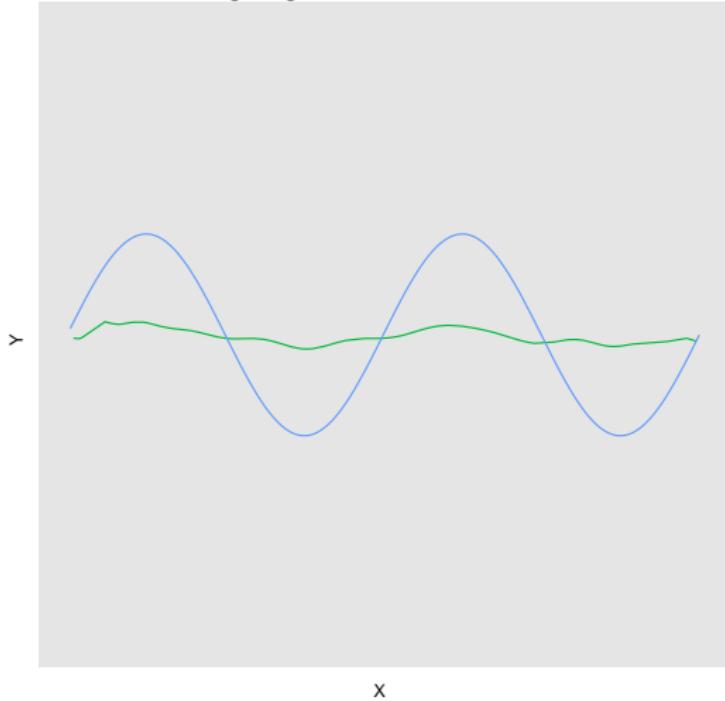
Ridge Regression with Lambda = 9.6



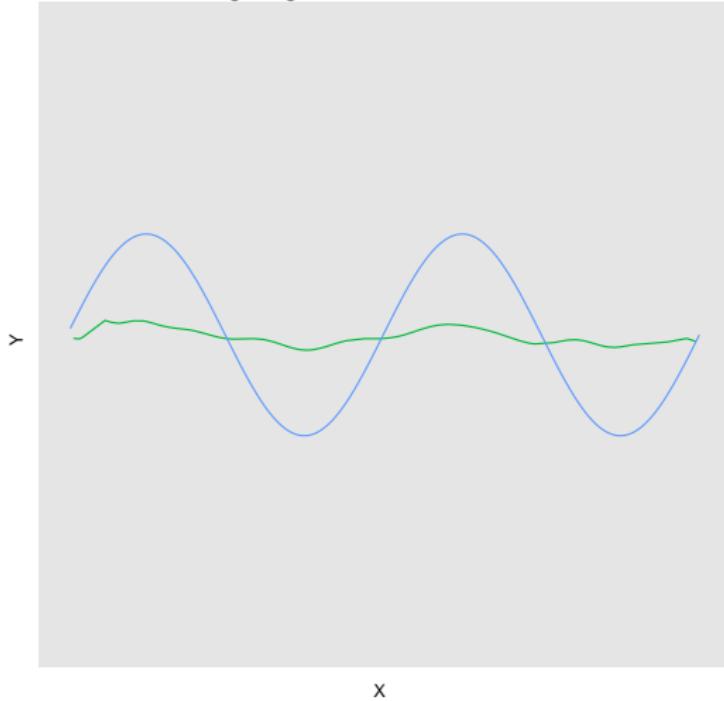
Ridge Regression with Lambda = 8.74



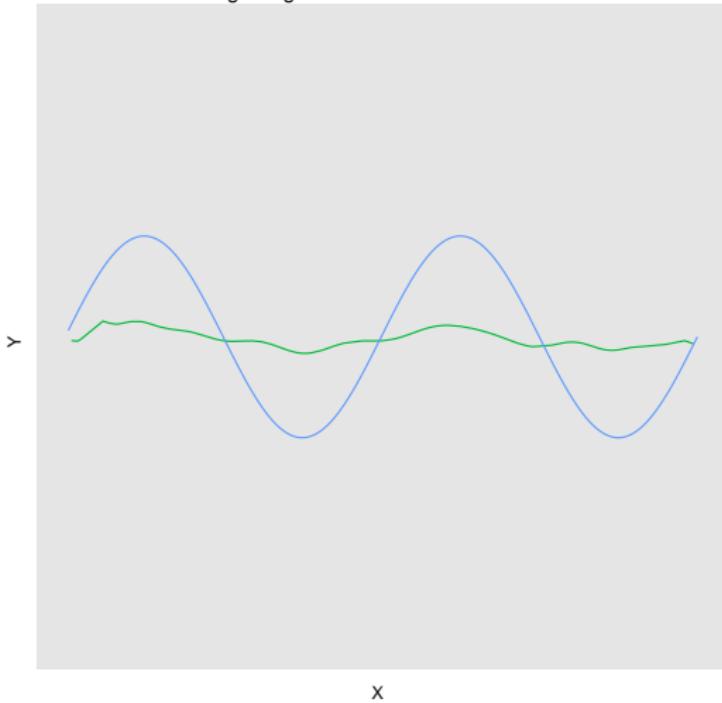
Ridge Regression with Lambda = 7.97



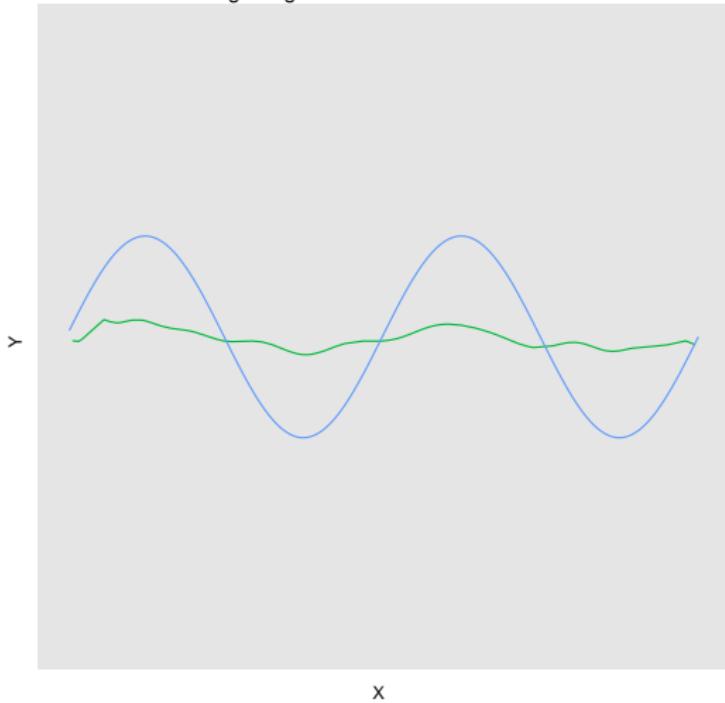
Ridge Regression with Lambda = 7.26



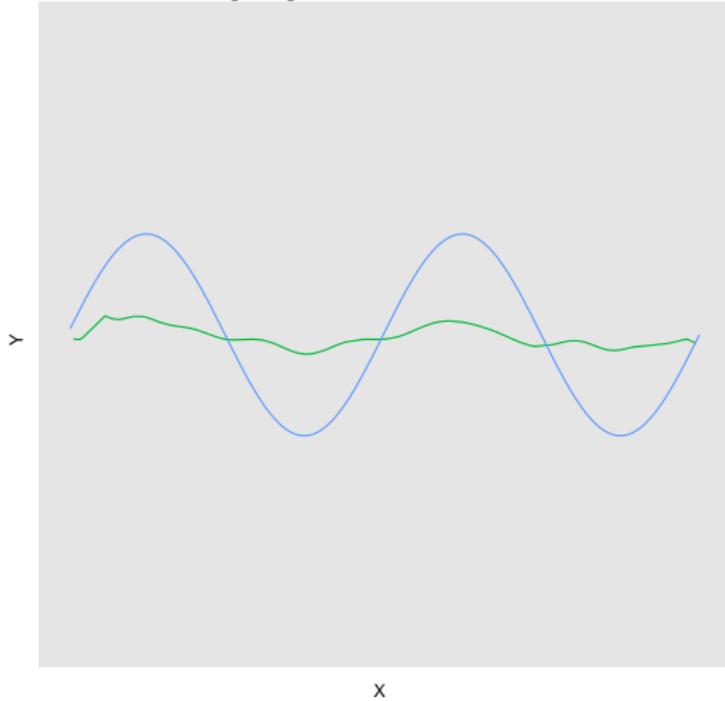
Ridge Regression with Lambda = 6.61



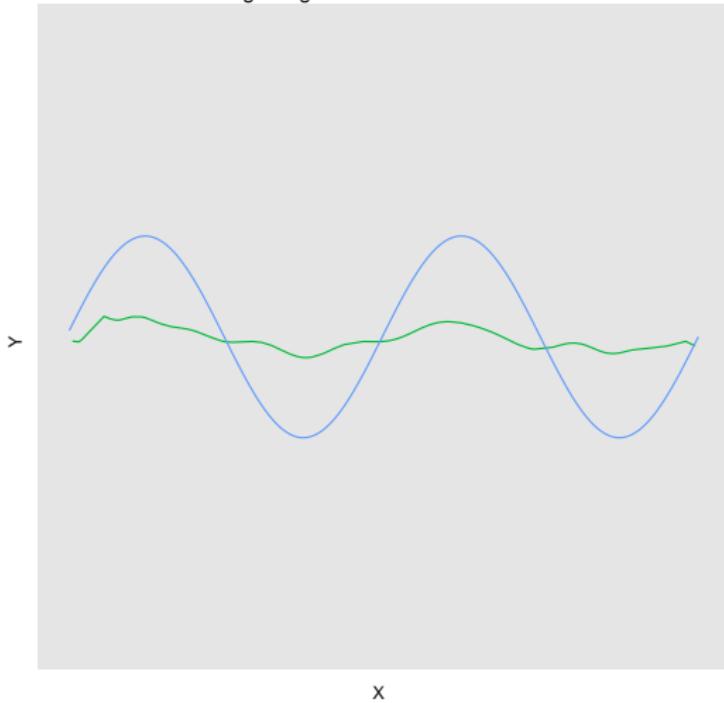
Ridge Regression with Lambda = 6.03



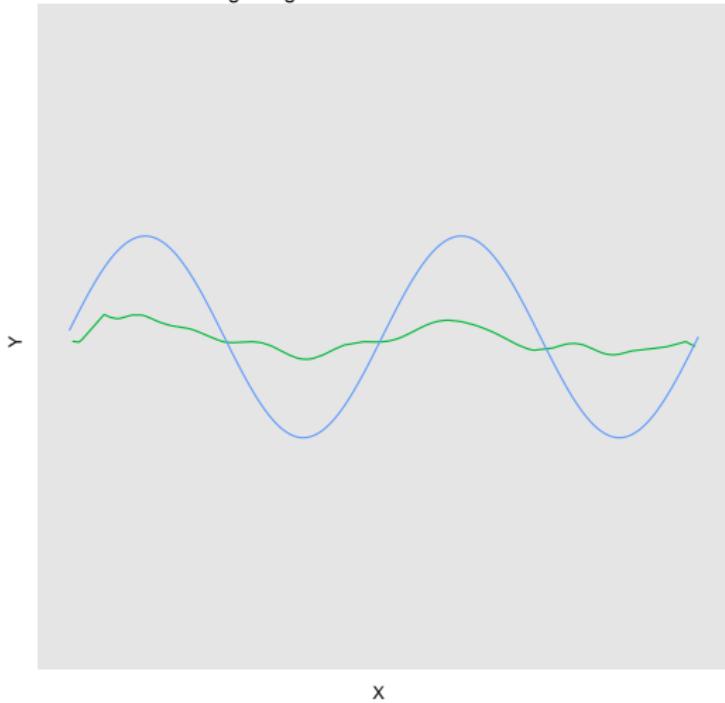
Ridge Regression with Lambda = 5.49



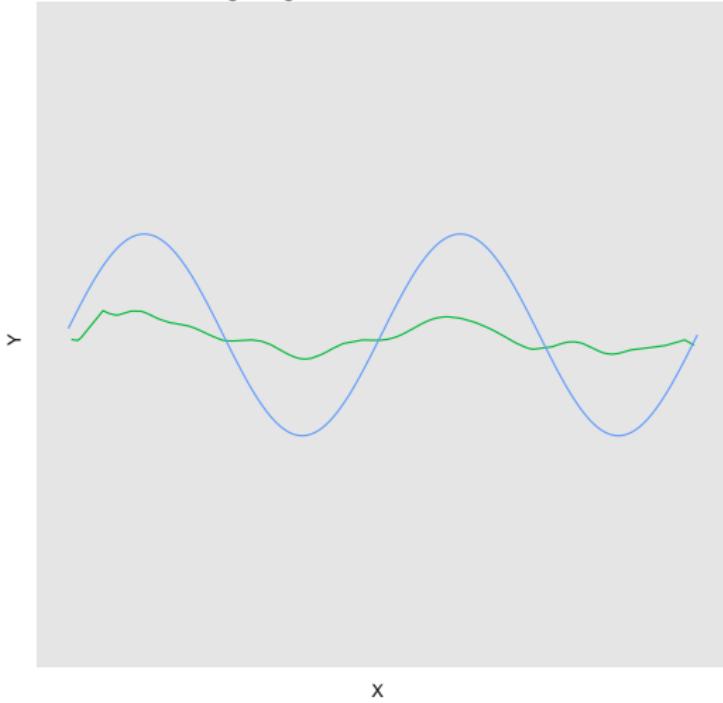
Ridge Regression with Lambda = 5



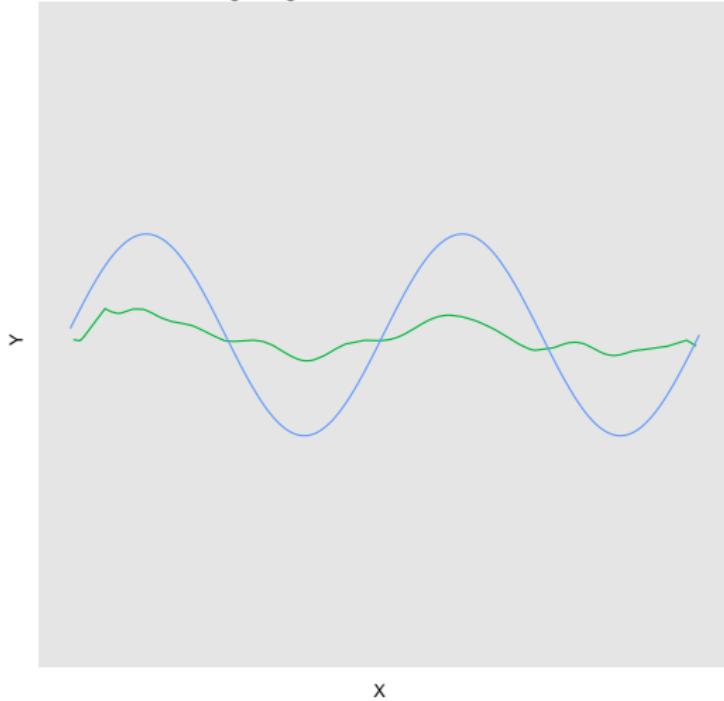
Ridge Regression with Lambda = 4.56



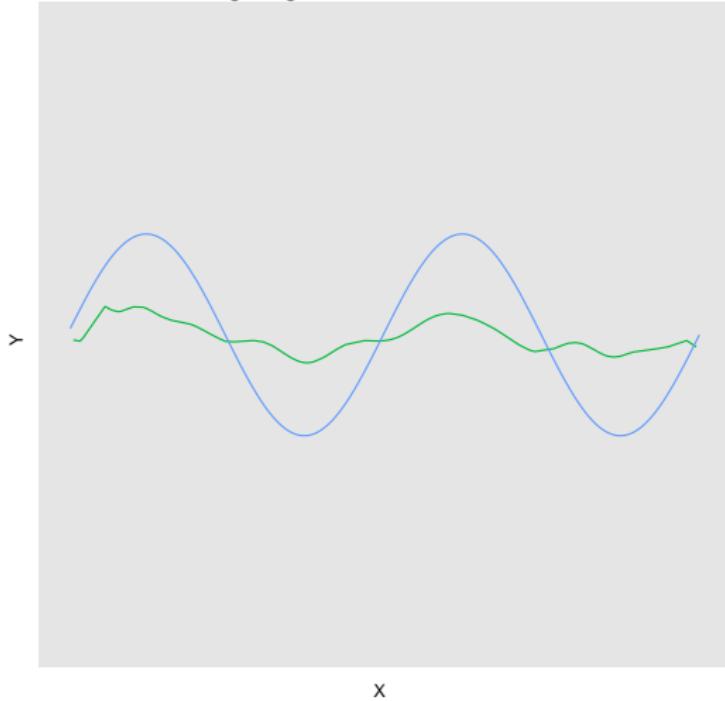
Ridge Regression with Lambda = 4.15



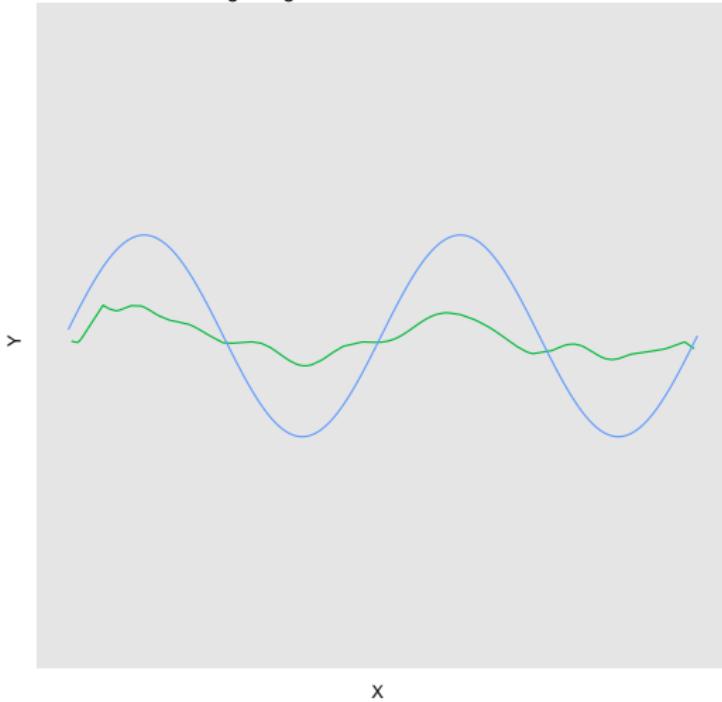
Ridge Regression with Lambda = 3.78



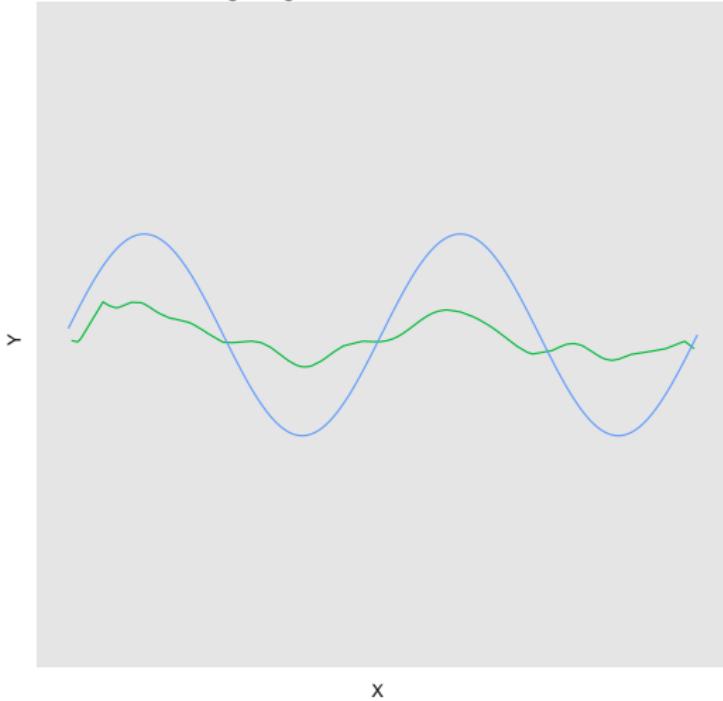
Ridge Regression with Lambda = 3.45



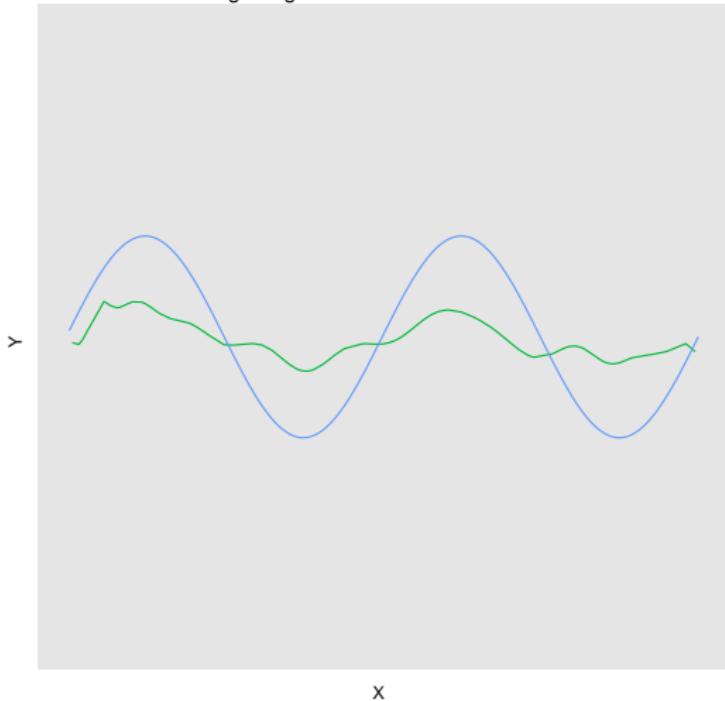
Ridge Regression with Lambda = 3.14



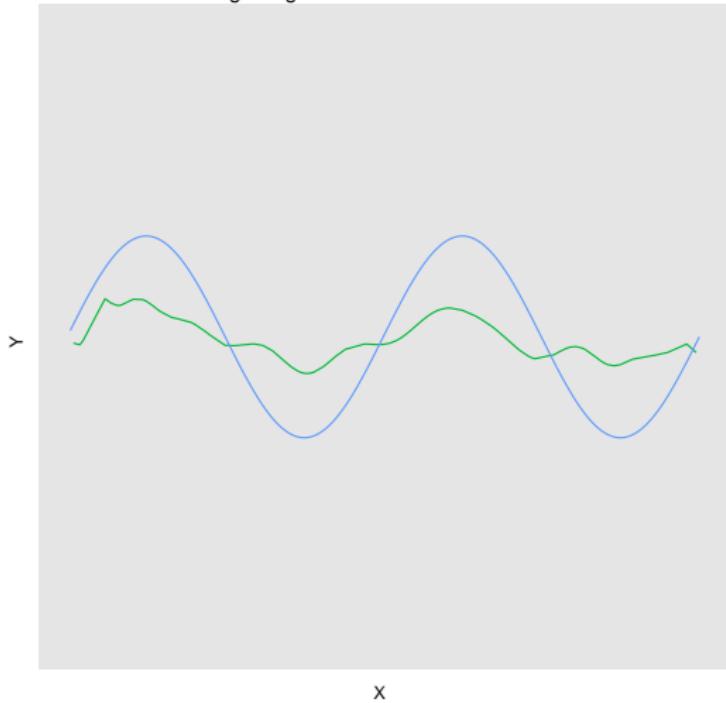
Ridge Regression with Lambda = 2.86



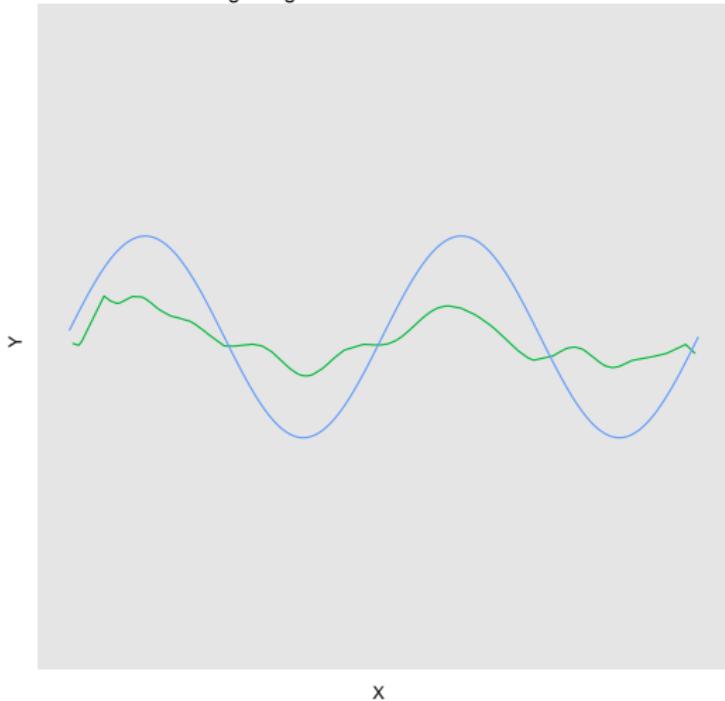
Ridge Regression with Lambda = 2.61



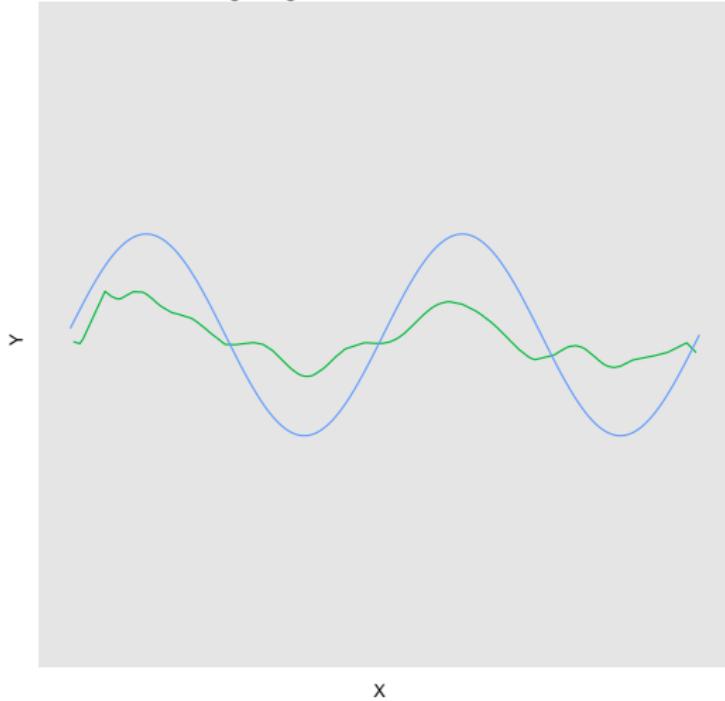
Ridge Regression with Lambda = 2.38



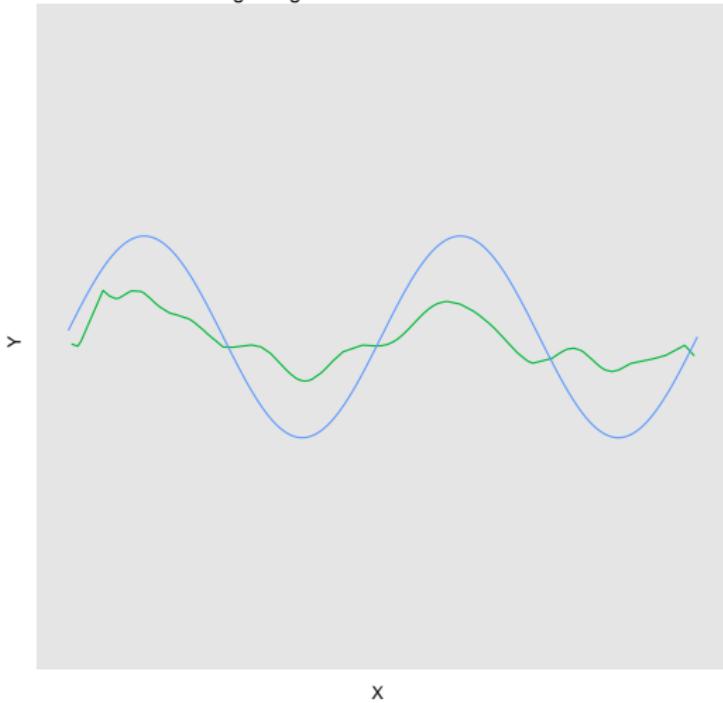
Ridge Regression with Lambda = 2.17



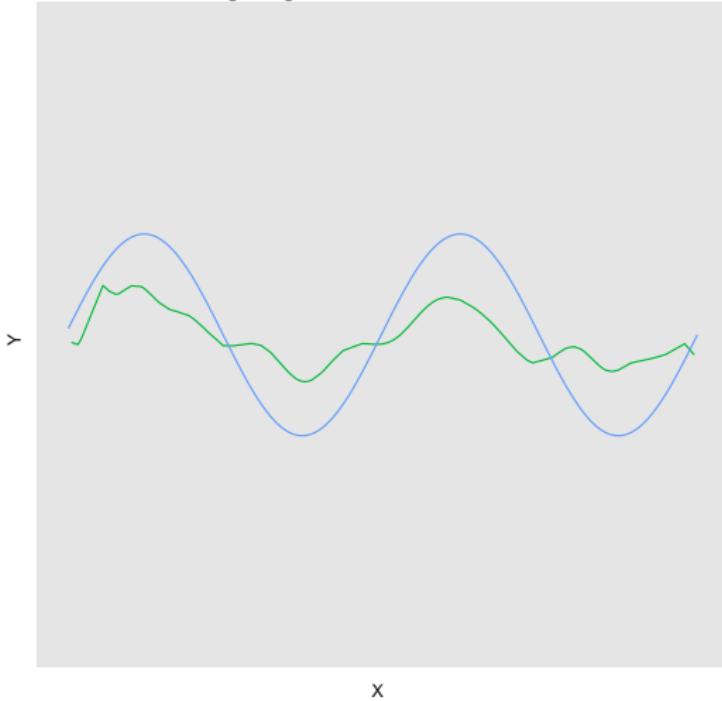
Ridge Regression with Lambda = 1.97



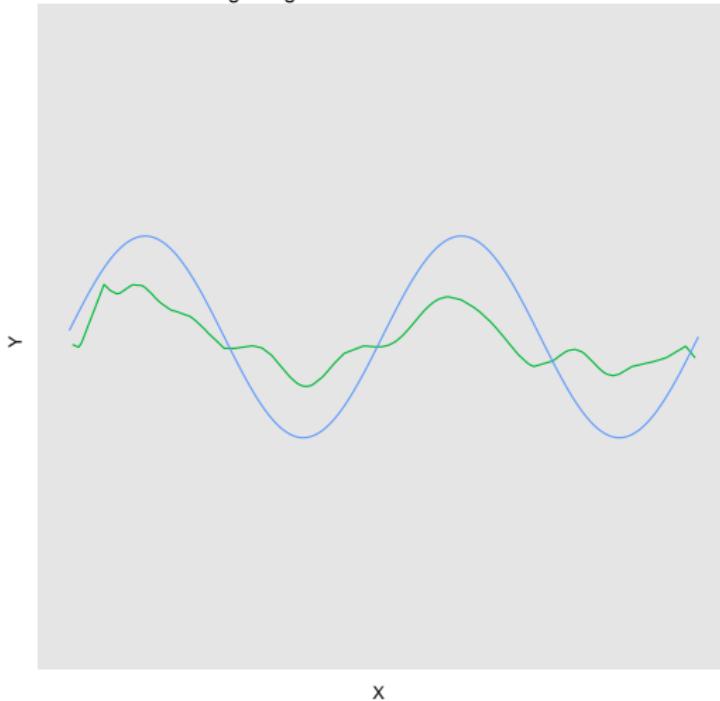
Ridge Regression with Lambda = 1.8



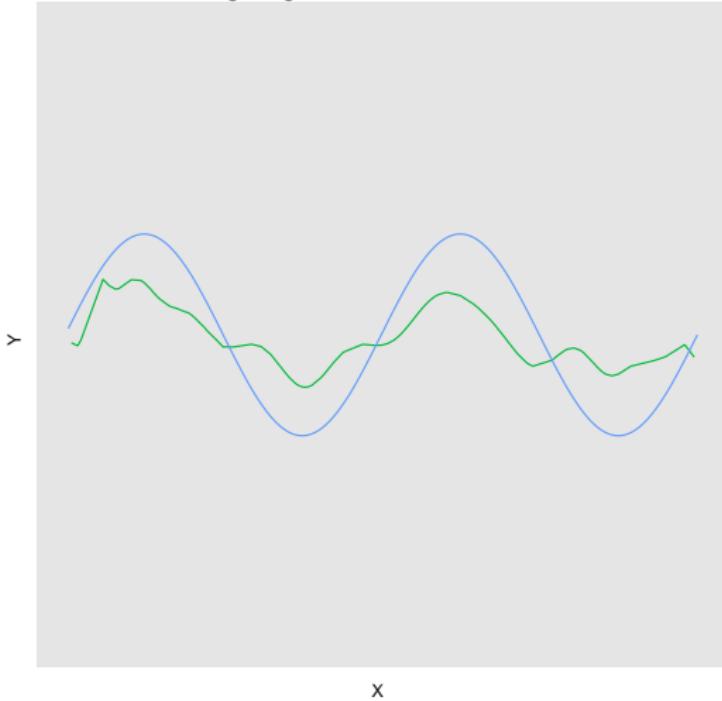
Ridge Regression with Lambda = 1.64



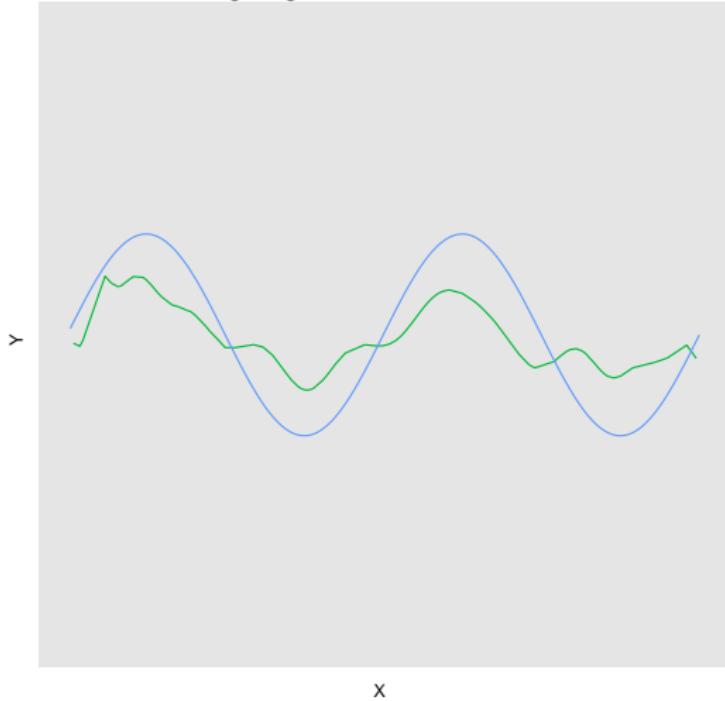
Ridge Regression with Lambda = 1.49



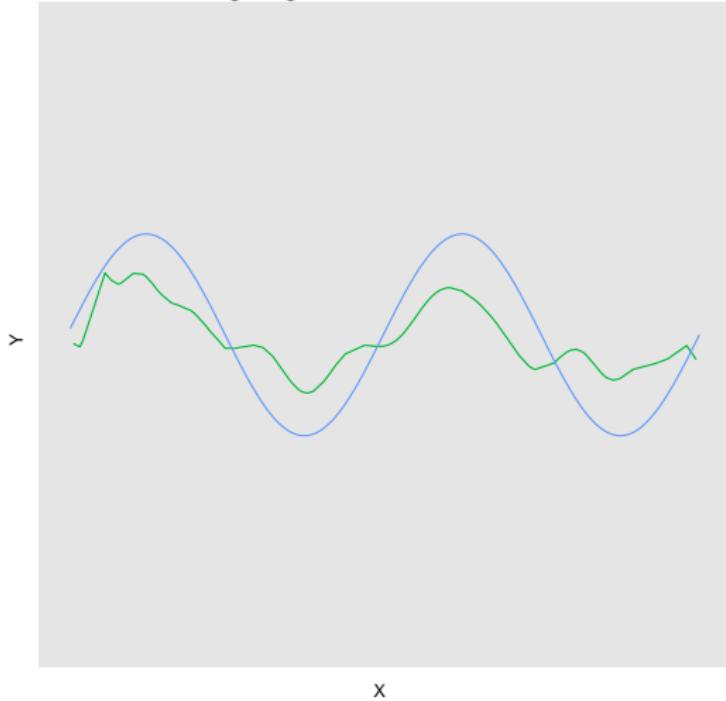
Ridge Regression with Lambda = 1.36



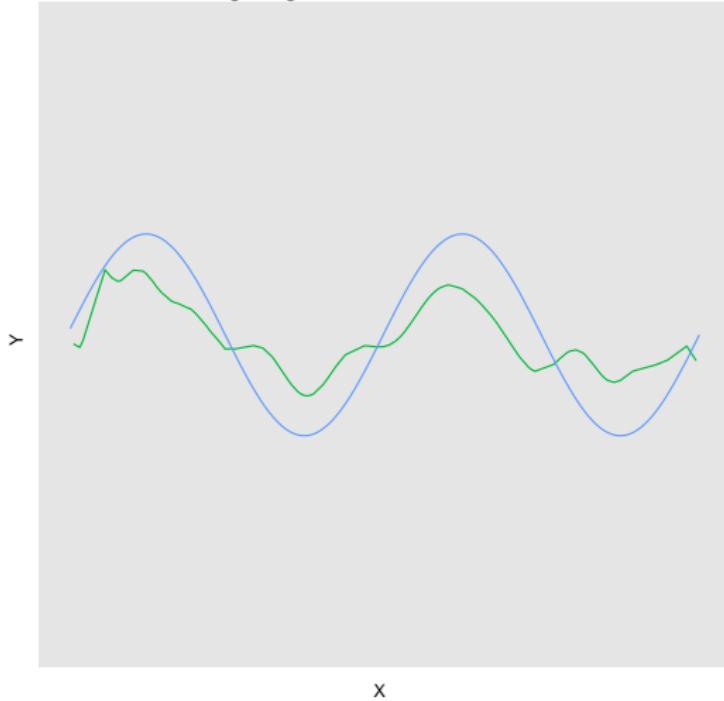
Ridge Regression with Lambda = 1.24



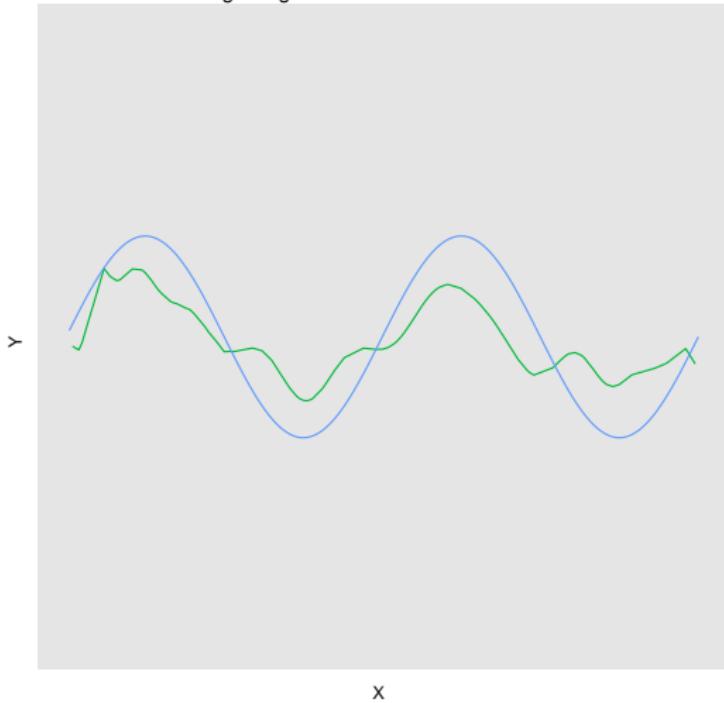
Ridge Regression with Lambda = 1.13



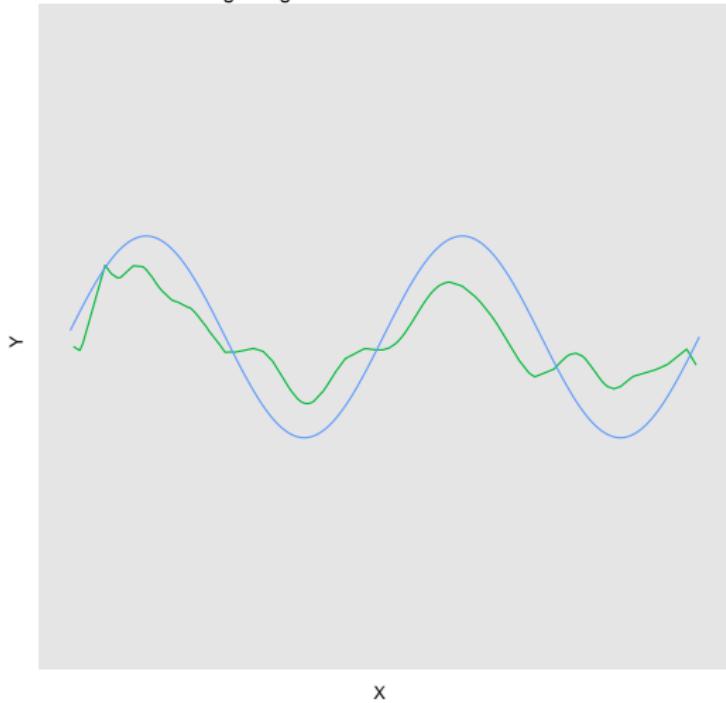
Ridge Regression with Lambda = 1.03



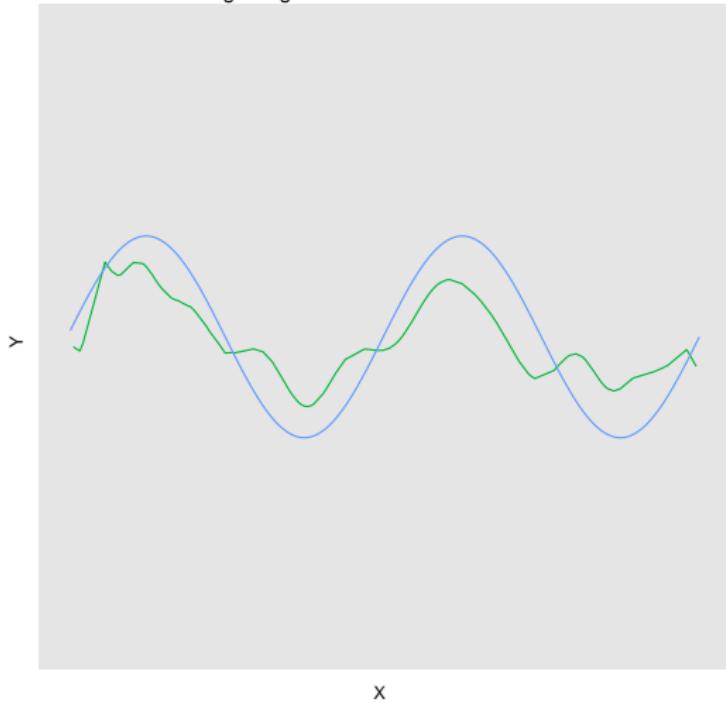
Ridge Regression with Lambda = 0.937



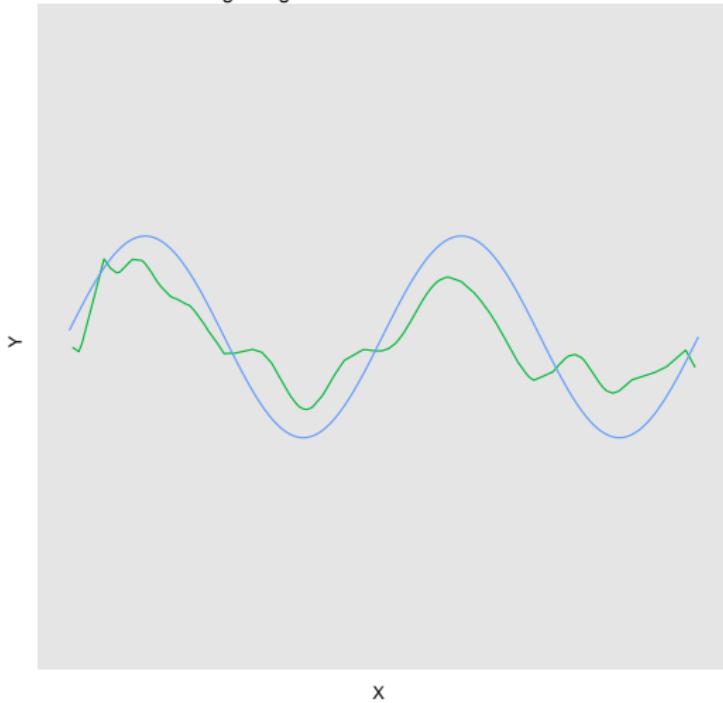
Ridge Regression with Lambda = 0.854



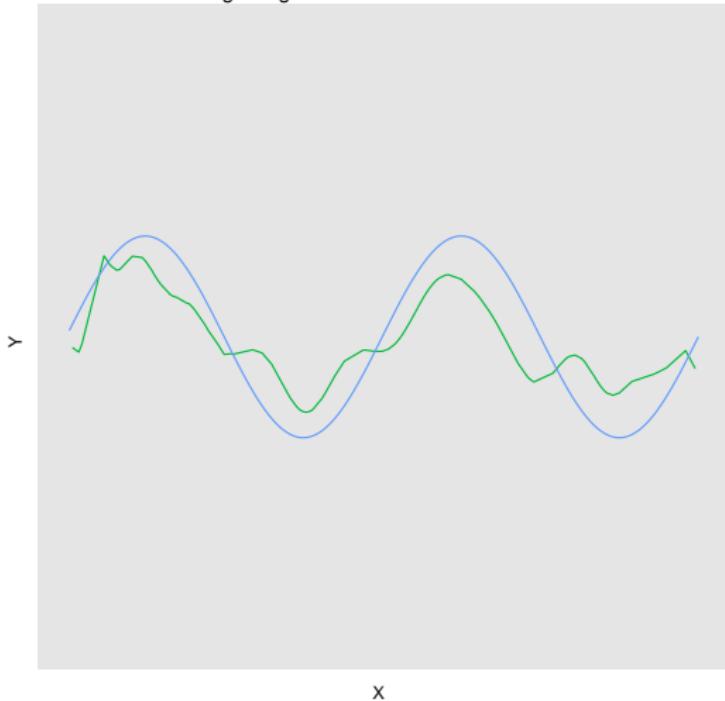
Ridge Regression with Lambda = 0.778



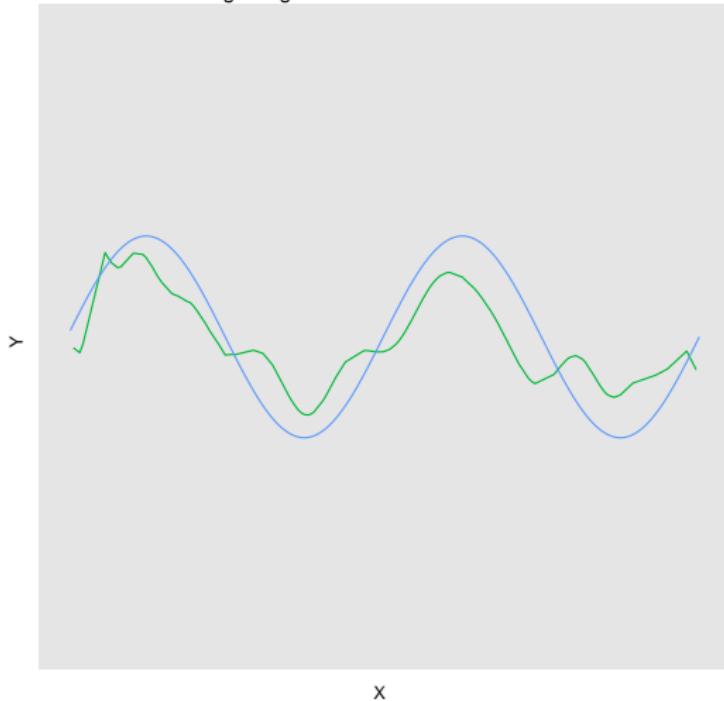
Ridge Regression with Lambda = 0.709



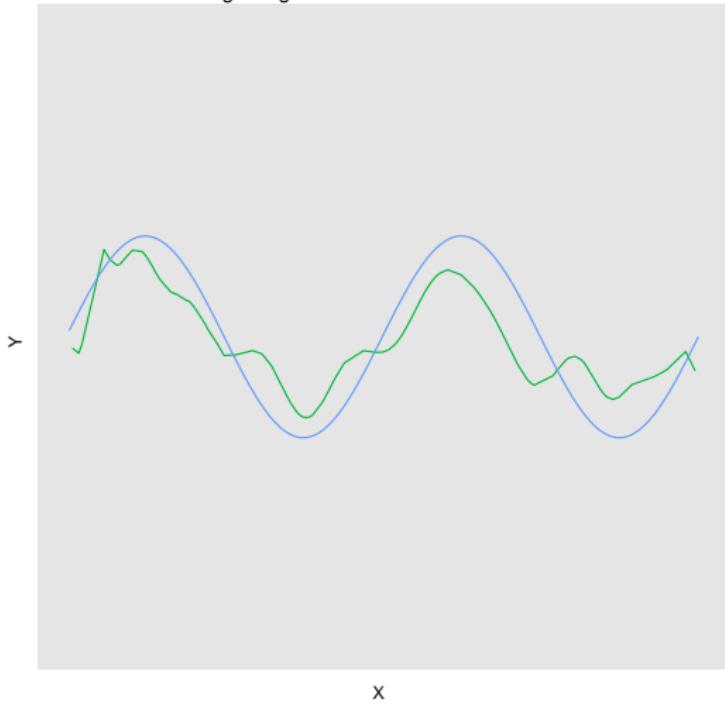
Ridge Regression with Lambda = 0.646



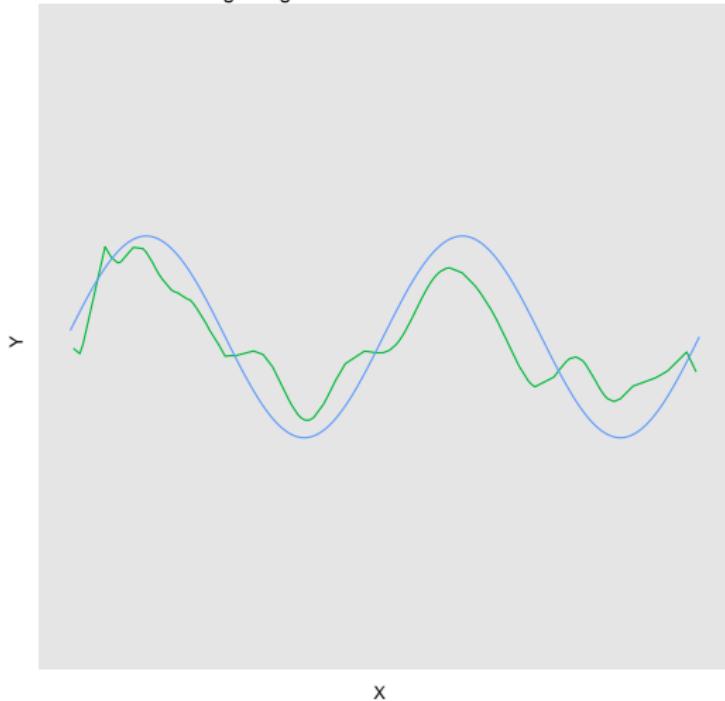
Ridge Regression with Lambda = 0.589



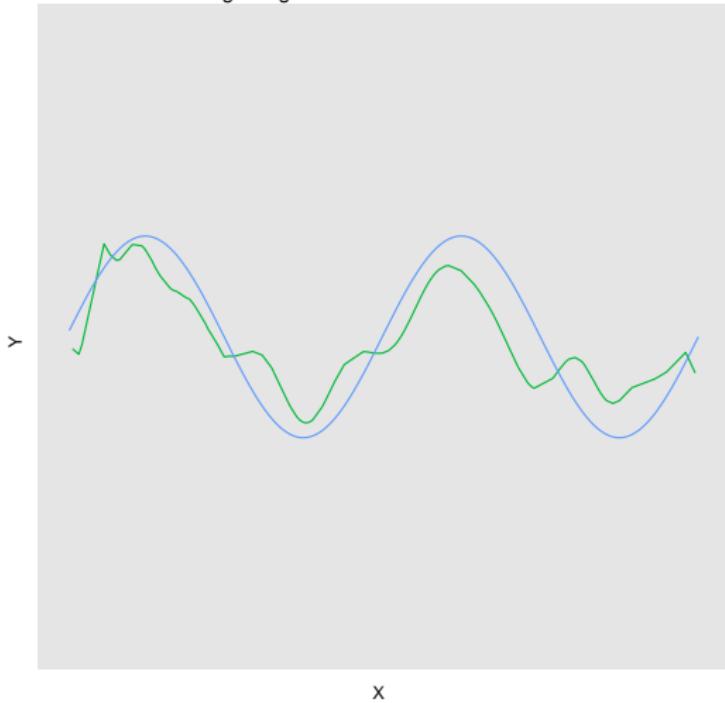
Ridge Regression with Lambda = 0.536



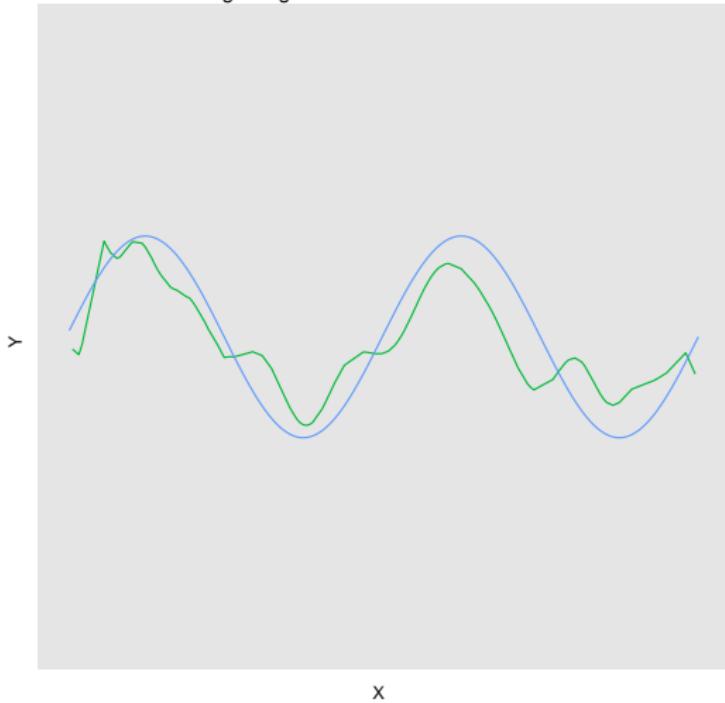
Ridge Regression with Lambda = 0.489



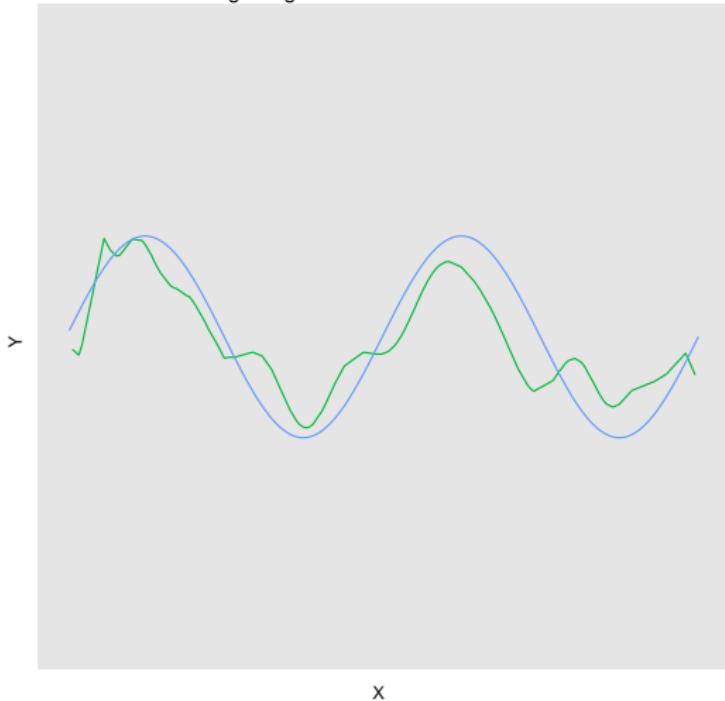
Ridge Regression with Lambda = 0.445



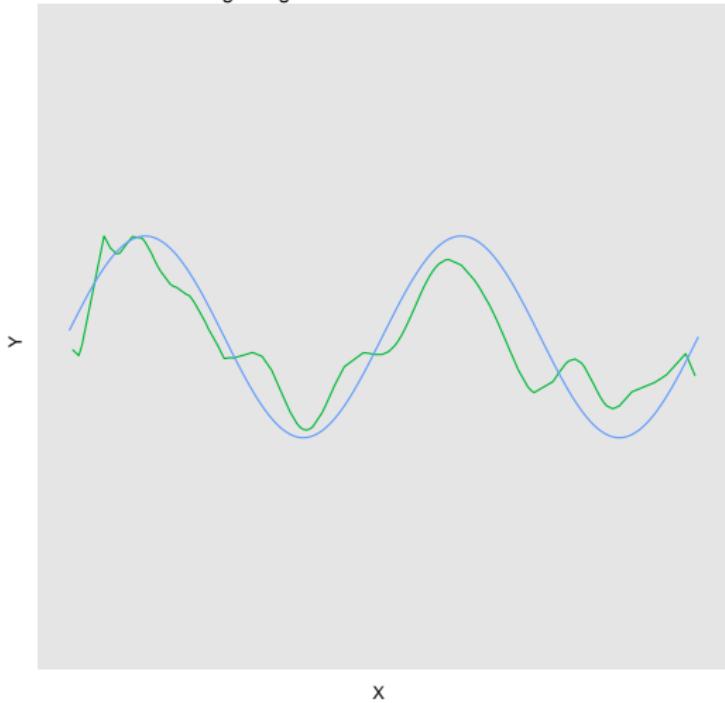
Ridge Regression with Lambda = 0.406



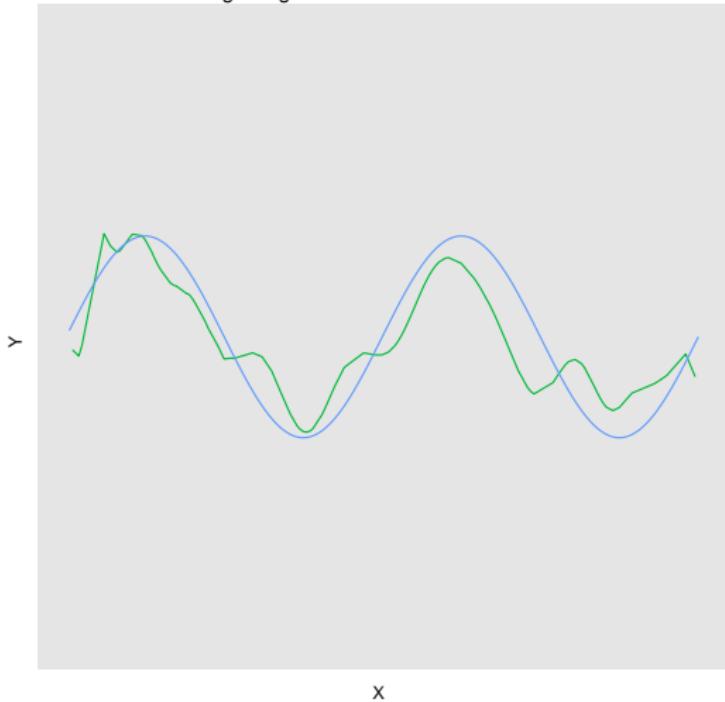
Ridge Regression with Lambda = 0.37



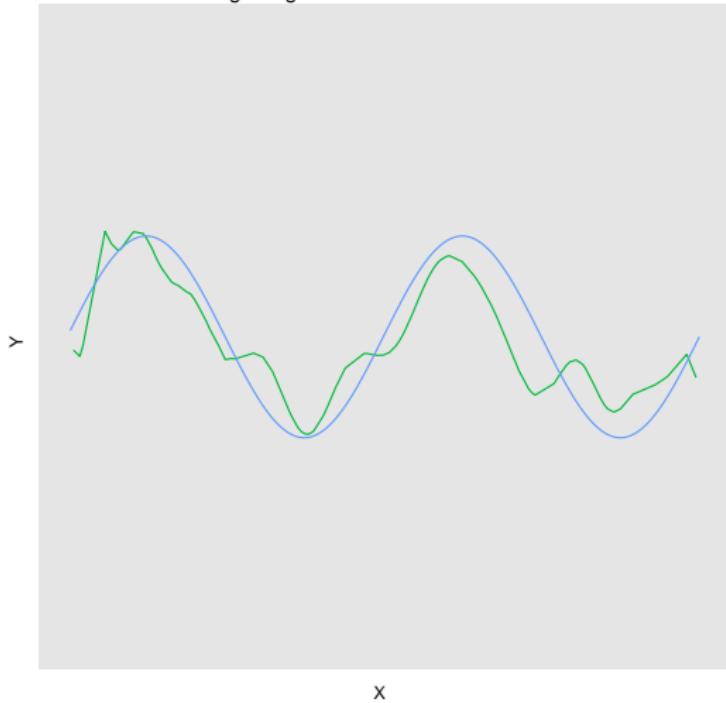
Ridge Regression with Lambda = 0.337



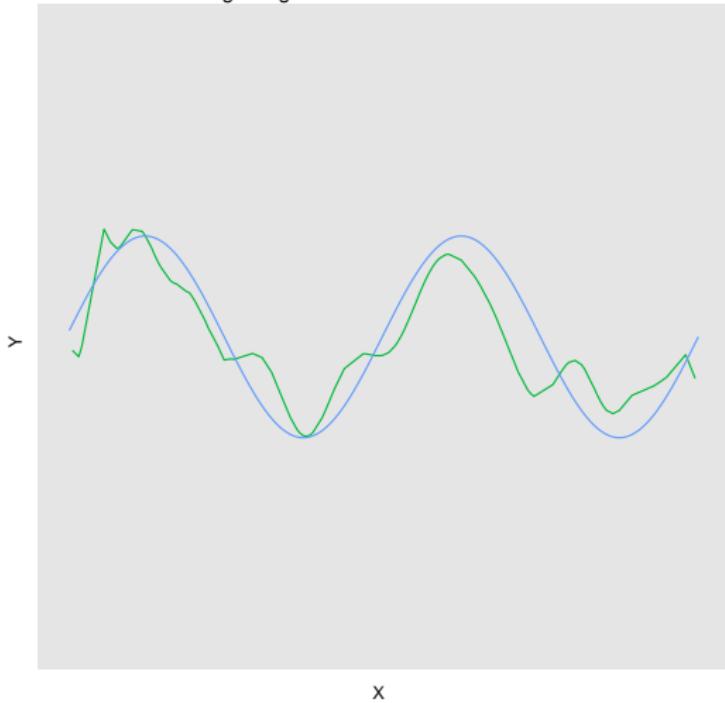
Ridge Regression with Lambda = 0.307



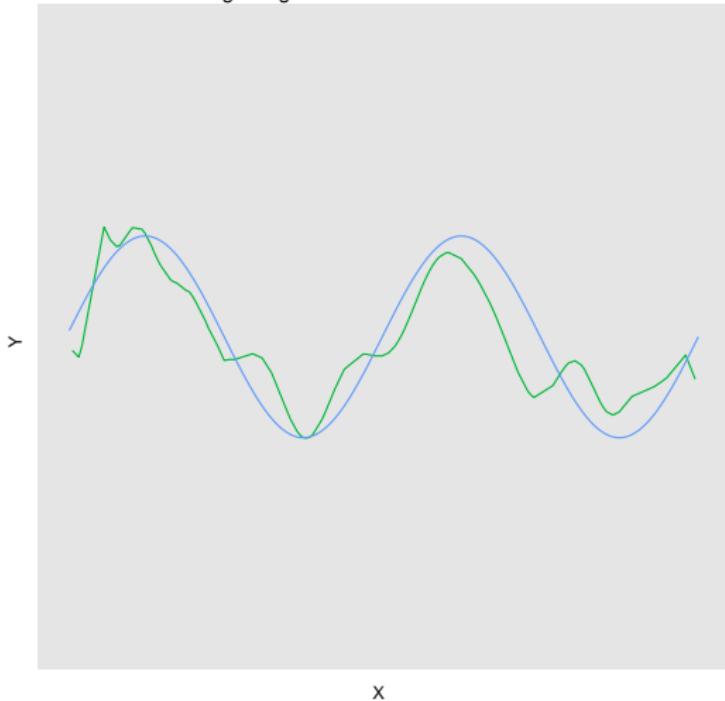
Ridge Regression with Lambda = 0.28



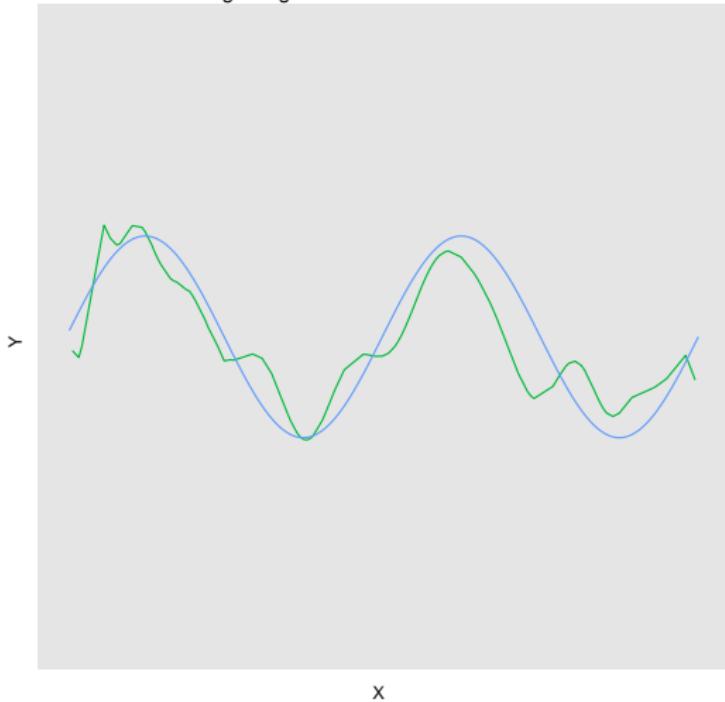
Ridge Regression with Lambda = 0.255



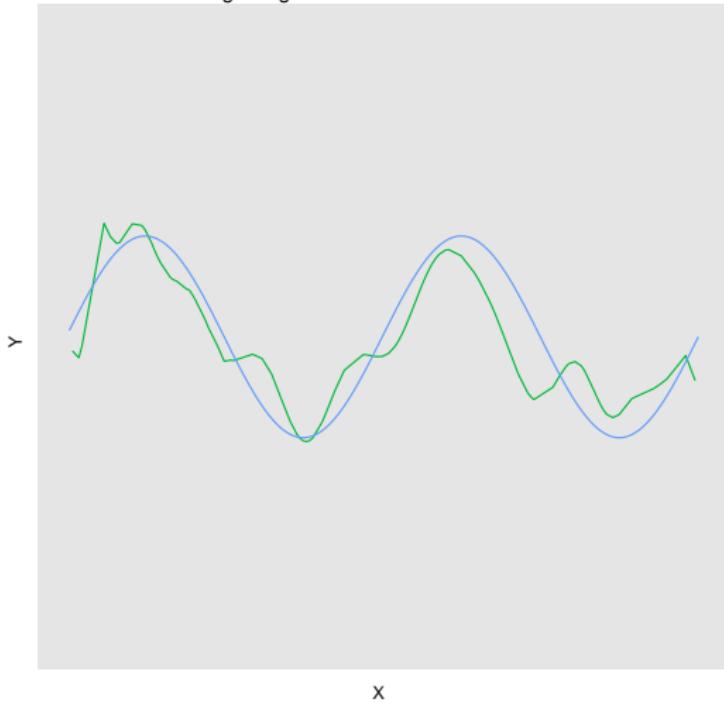
Ridge Regression with Lambda = 0.232



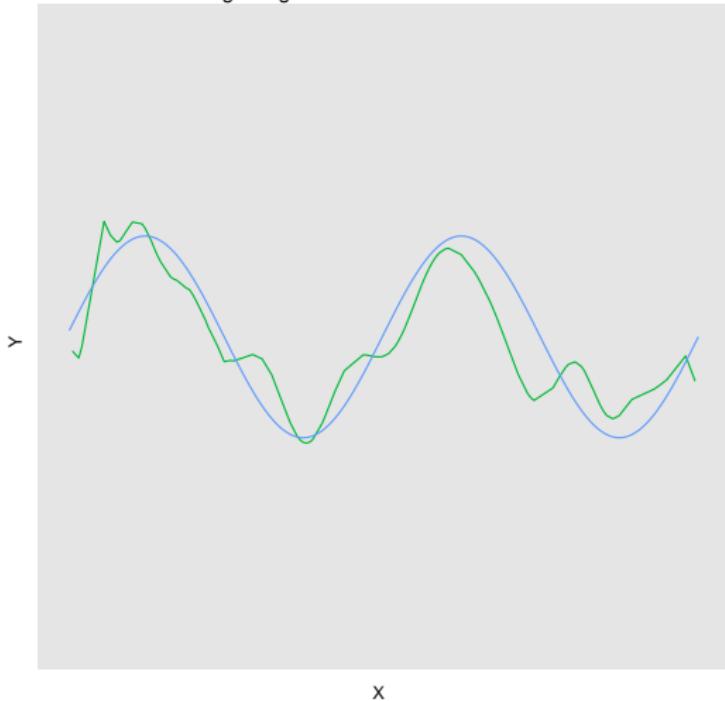
Ridge Regression with Lambda = 0.212



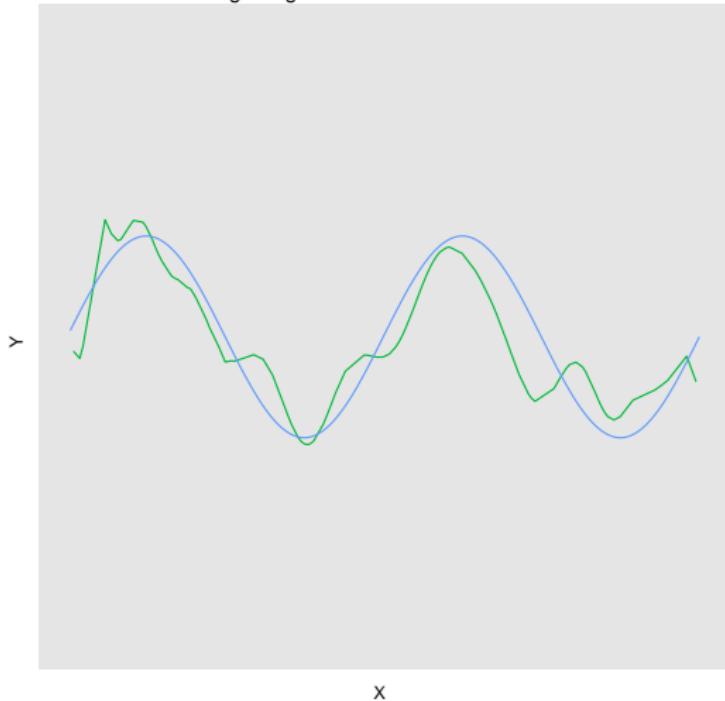
Ridge Regression with Lambda = 0.193



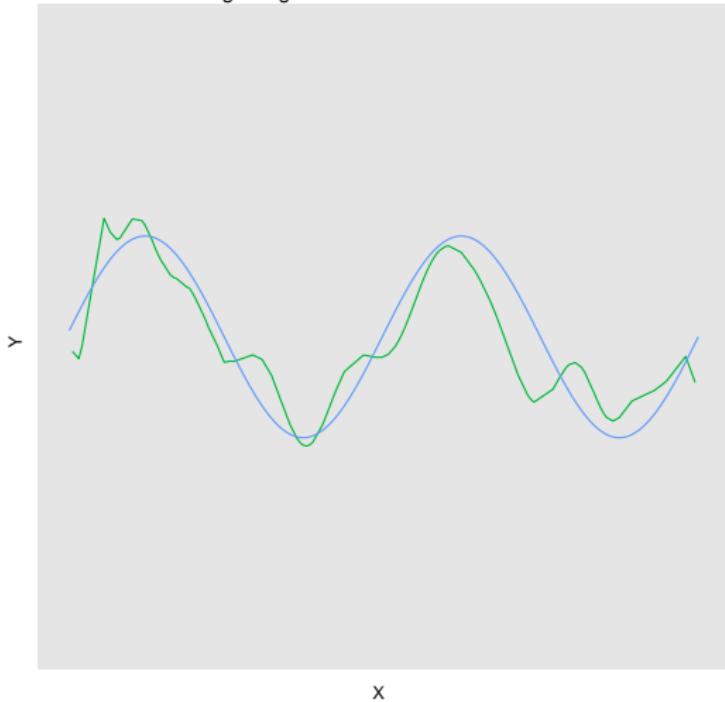
Ridge Regression with Lambda = 0.176



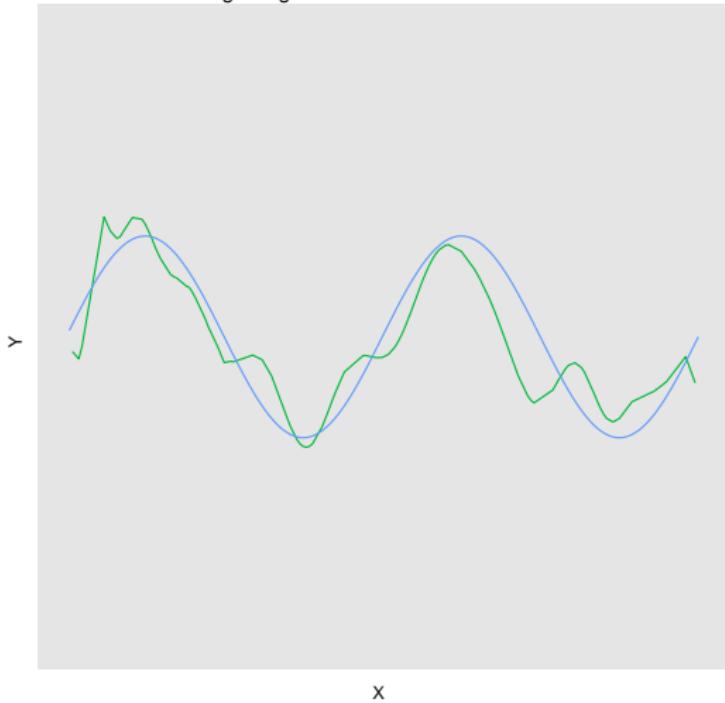
Ridge Regression with Lambda = 0.16



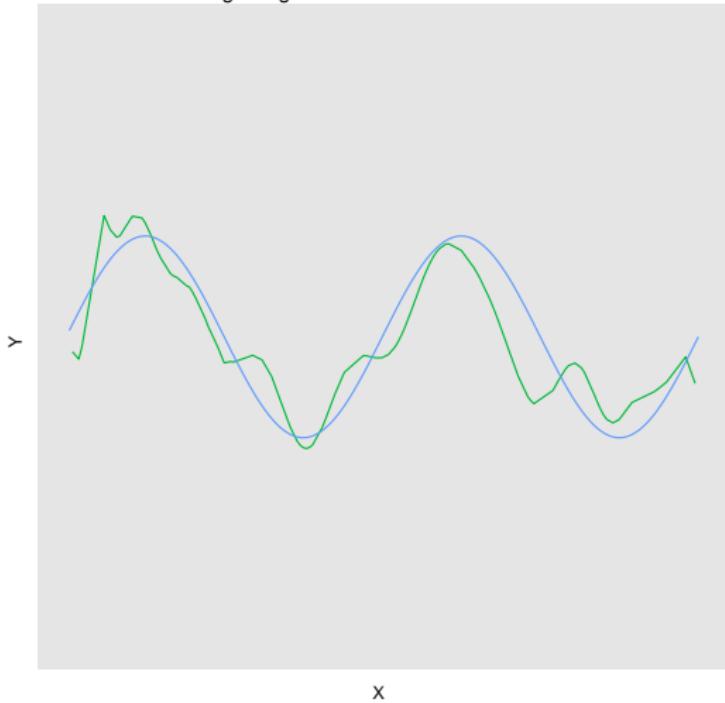
Ridge Regression with Lambda = 0.146



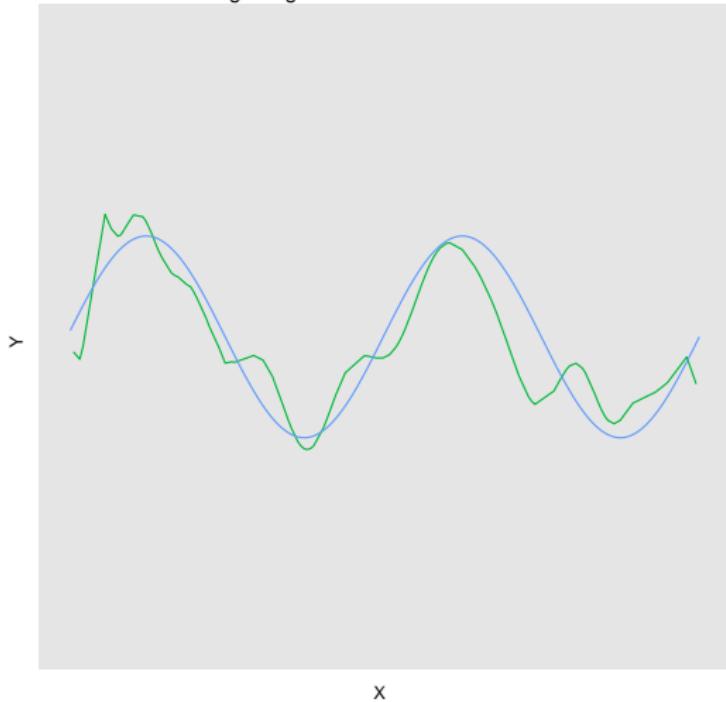
Ridge Regression with Lambda = 0.133



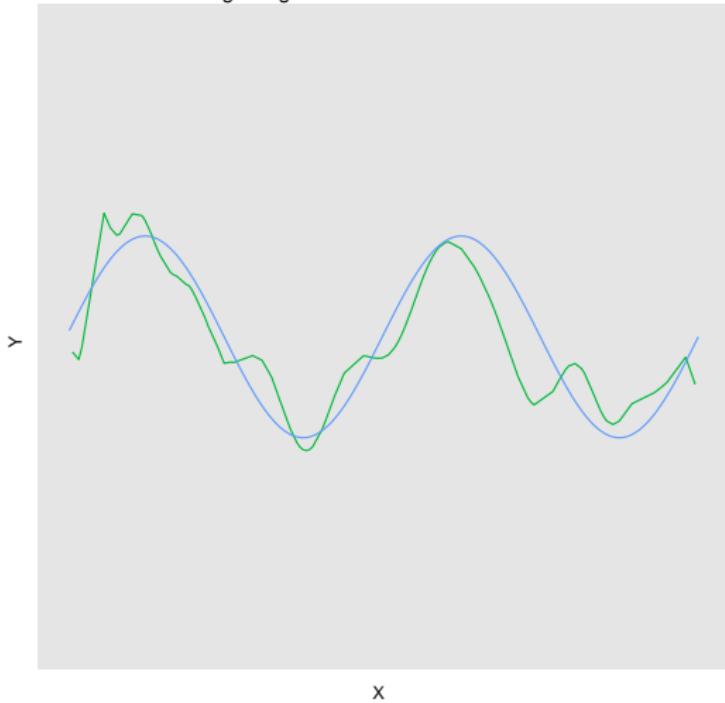
Ridge Regression with Lambda = 0.121



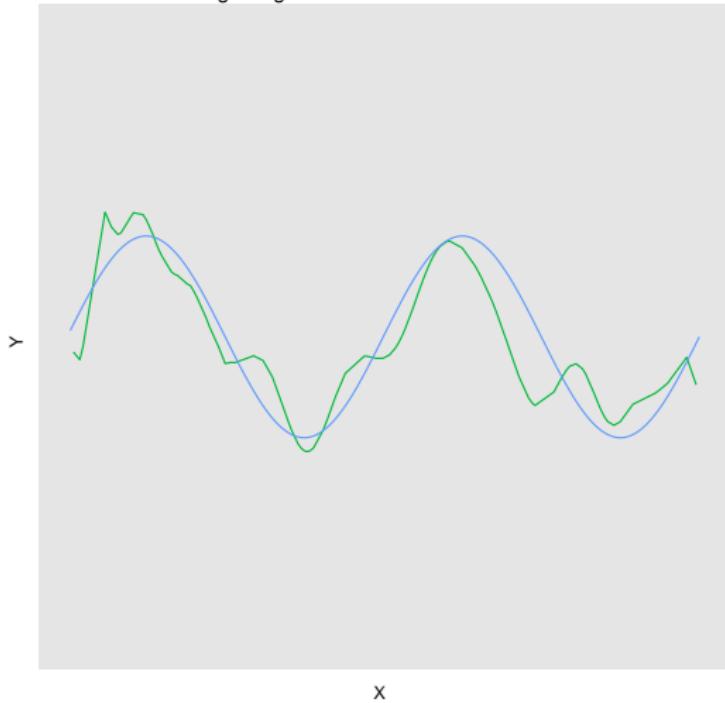
Ridge Regression with Lambda = 0.11



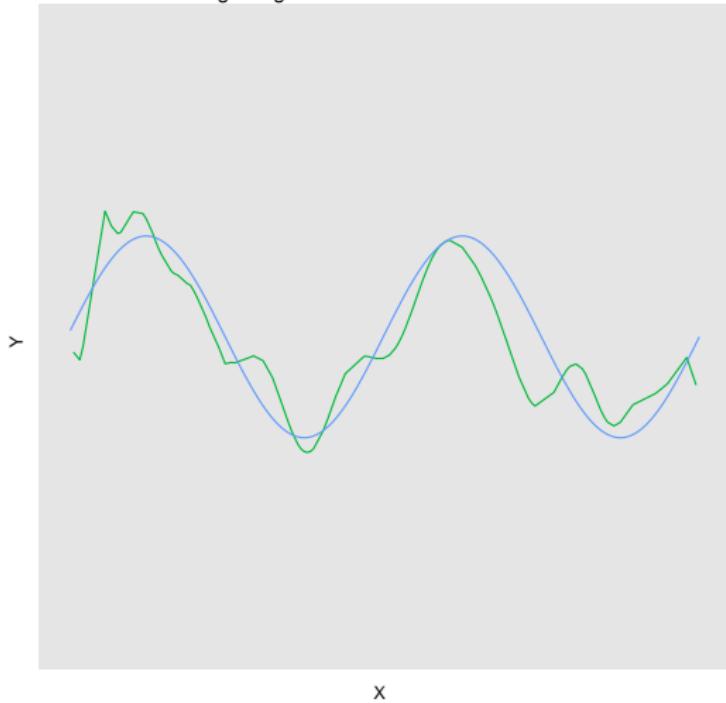
Ridge Regression with Lambda = 0.101



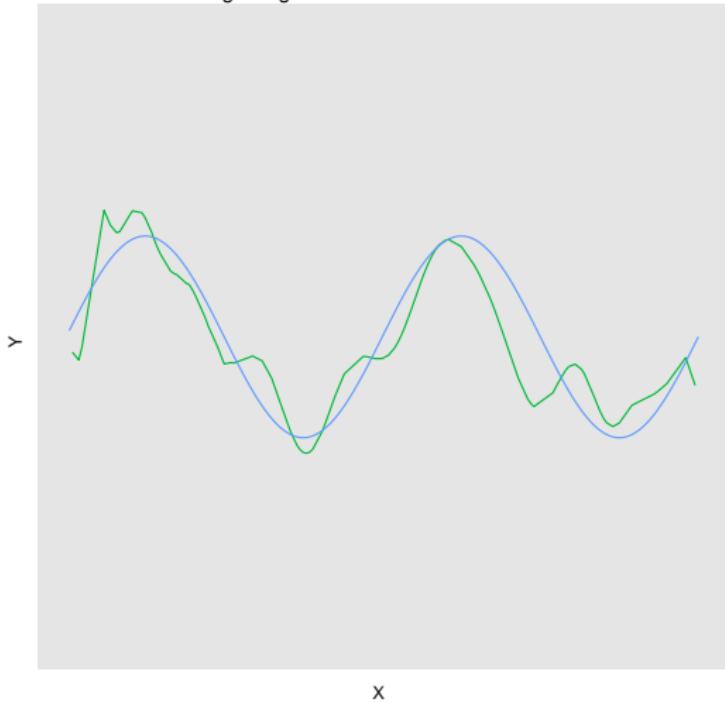
Ridge Regression with Lambda = 0.0916



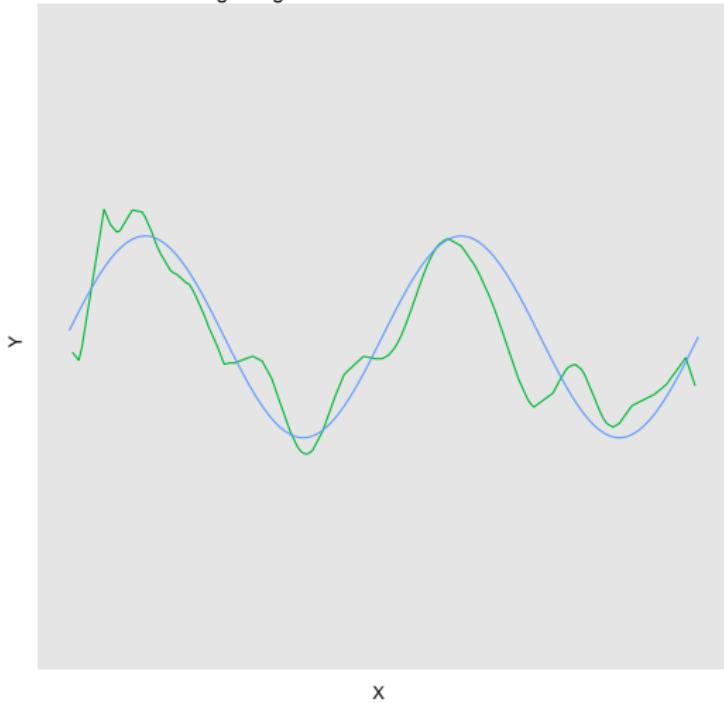
Ridge Regression with Lambda = 0.0835



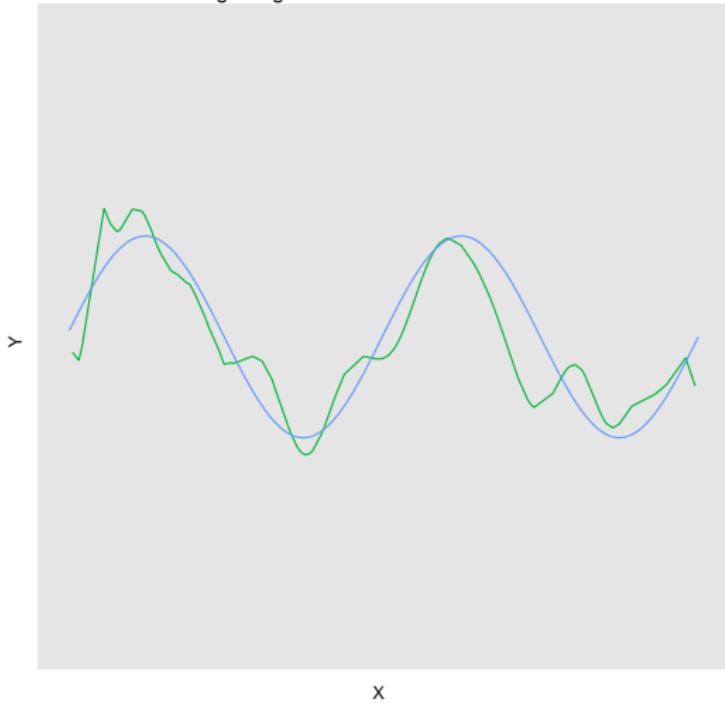
Ridge Regression with Lambda = 0.076



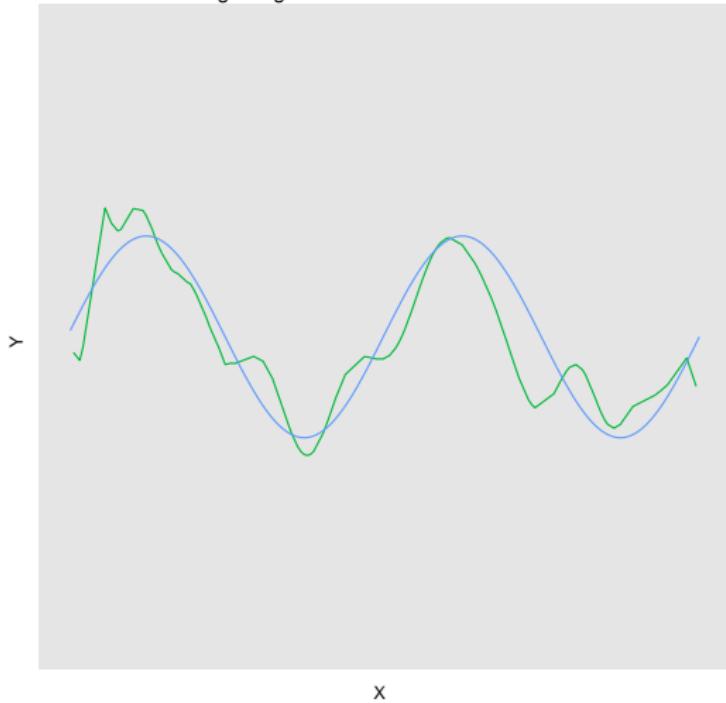
Ridge Regression with Lambda = 0.0693



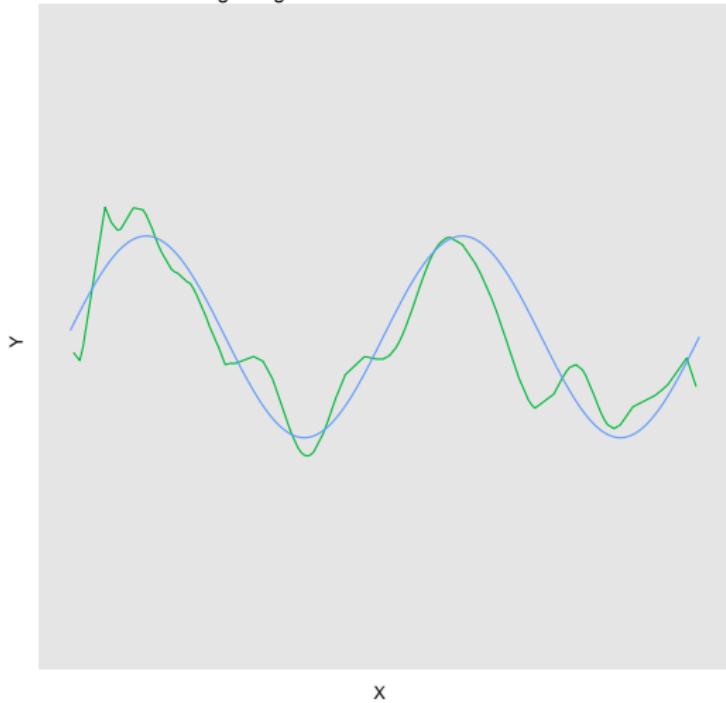
Ridge Regression with Lambda = 0.0631



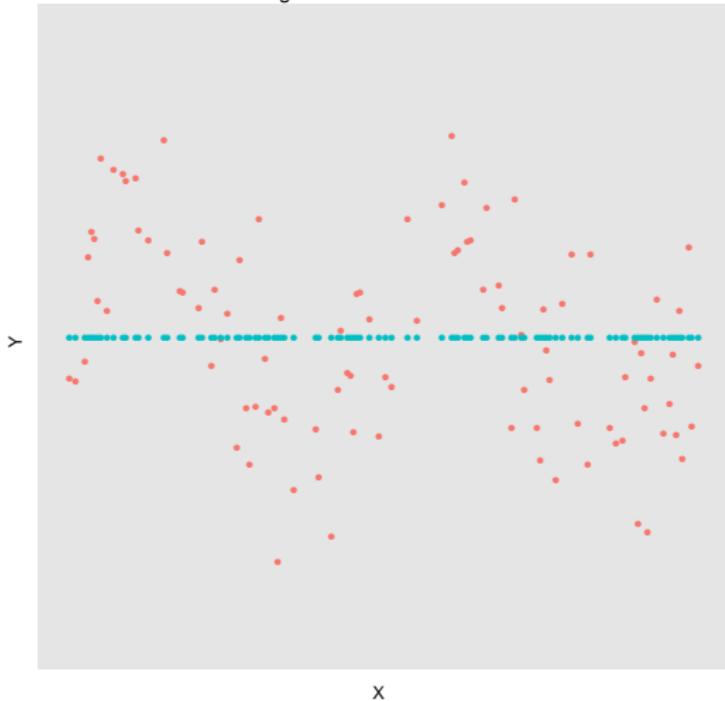
Ridge Regression with Lambda = 0.0575



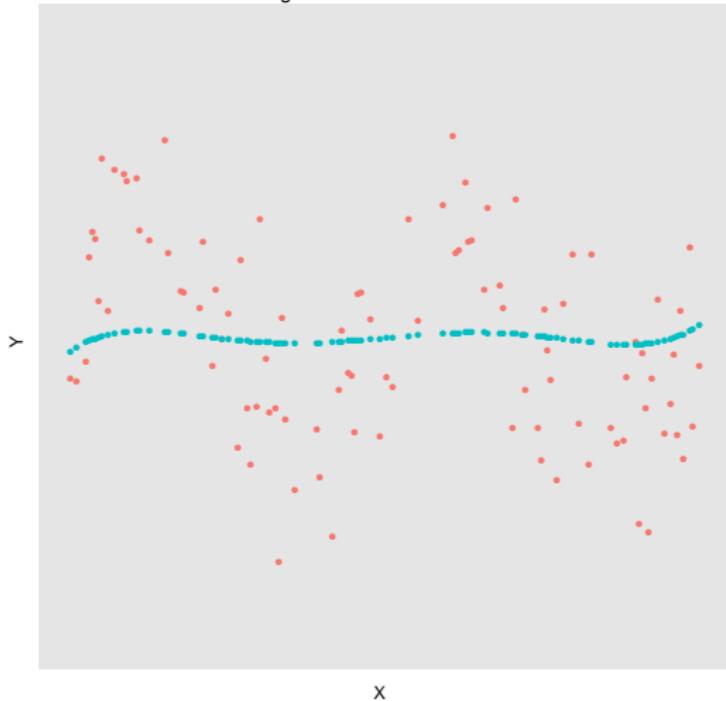
Ridge Regression with Lambda = 0.0524



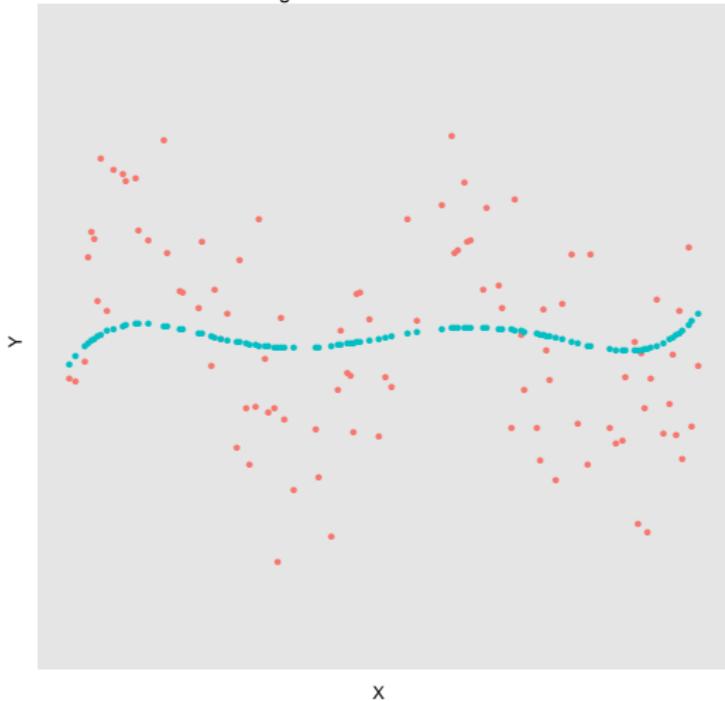
Lasso Regression with Lambda = 0.524



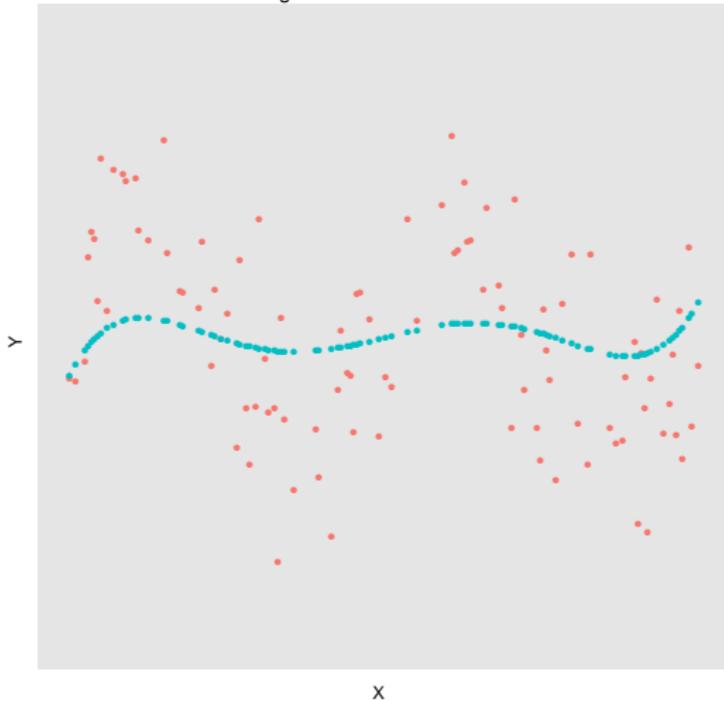
Lasso Regression with Lambda = 0.478



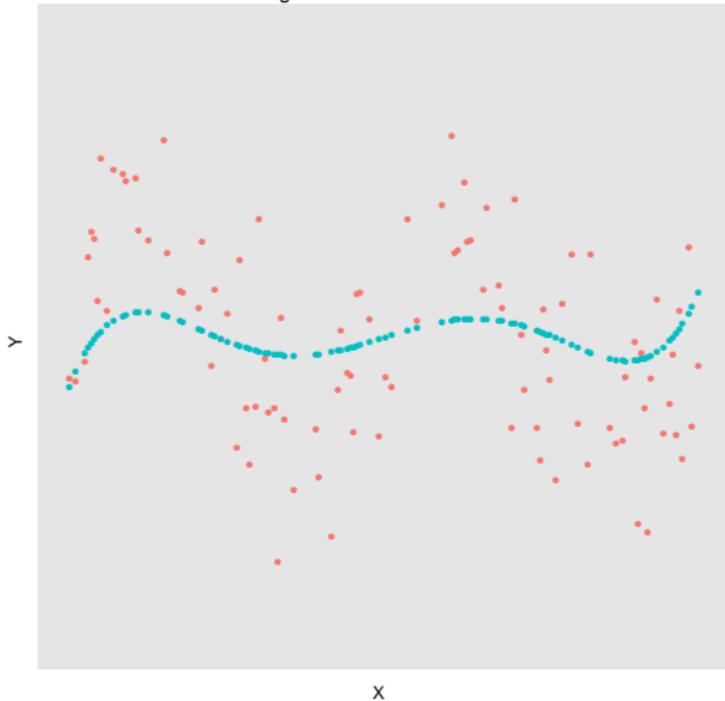
Lasso Regression with Lambda = 0.435



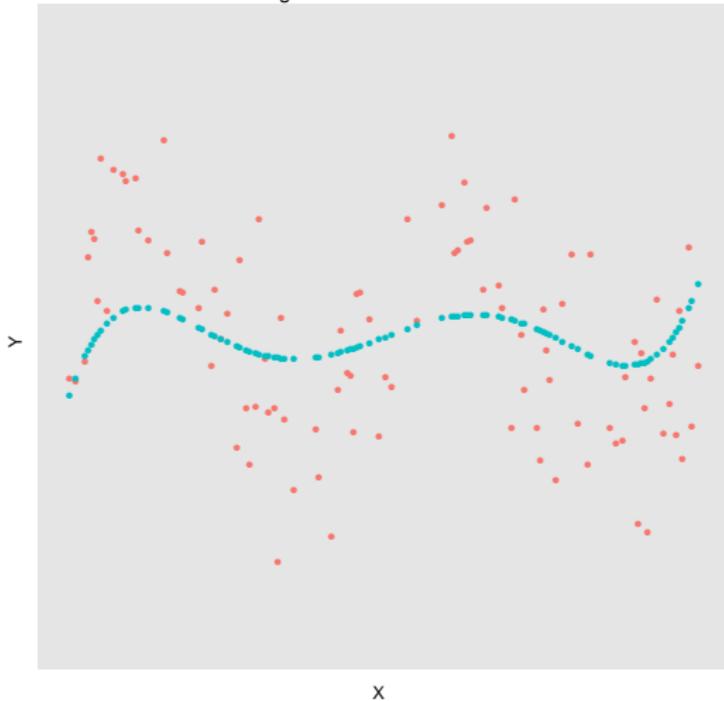
Lasso Regression with Lambda = 0.396



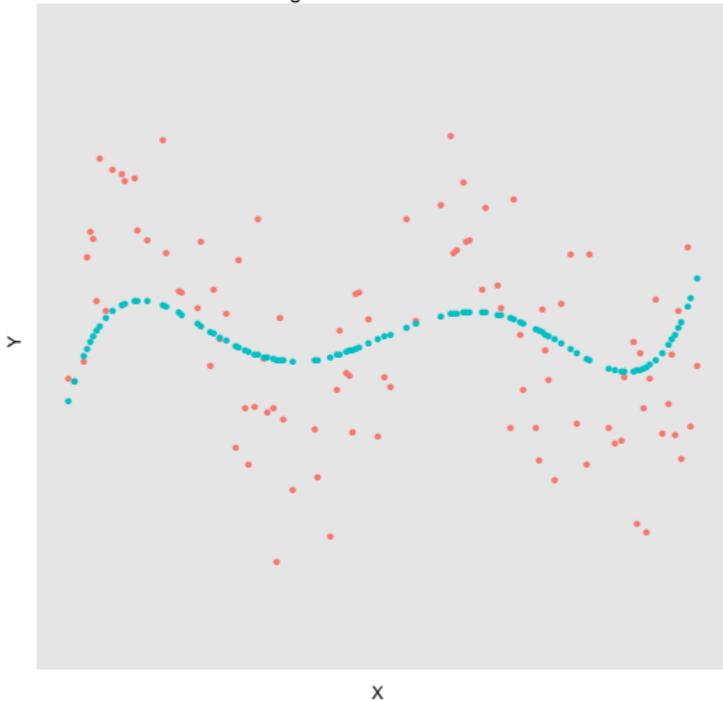
Lasso Regression with Lambda = 0.361



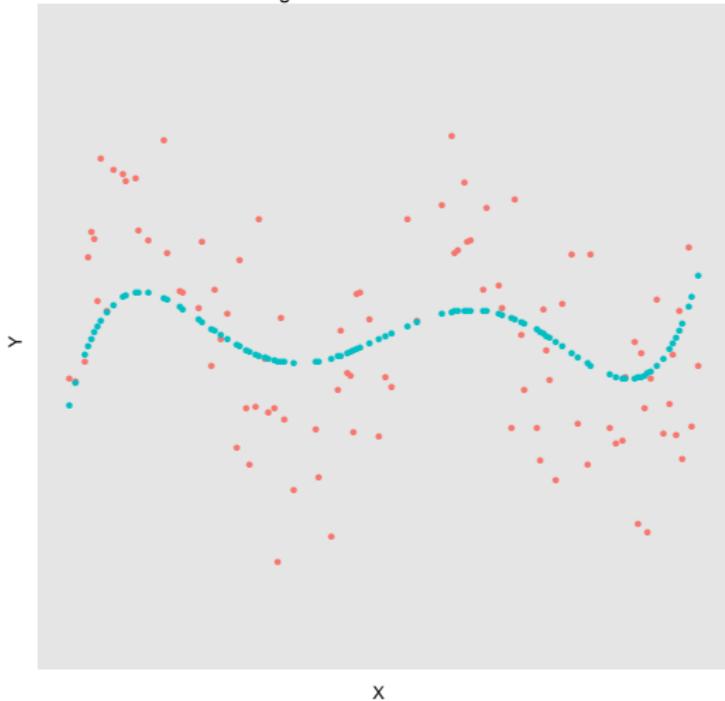
Lasso Regression with Lambda = 0.329



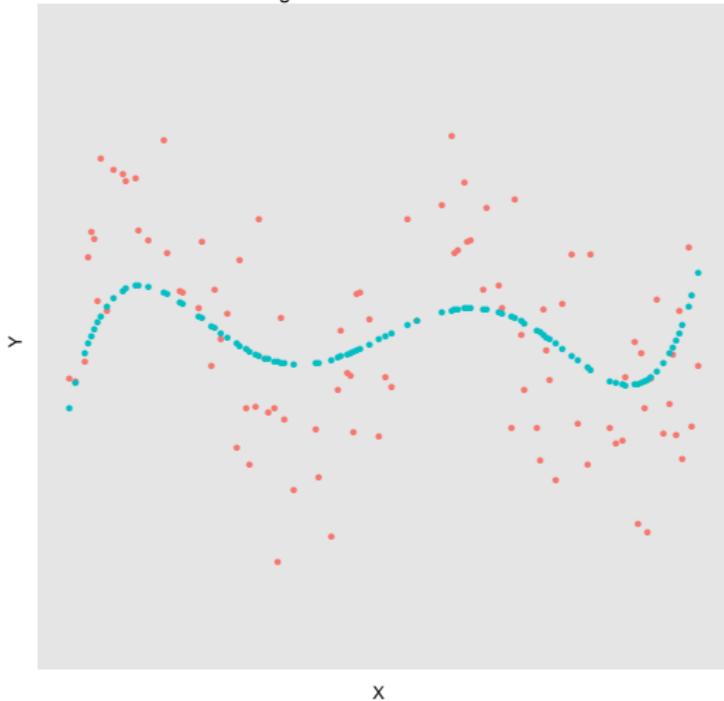
Lasso Regression with Lambda = 0.3



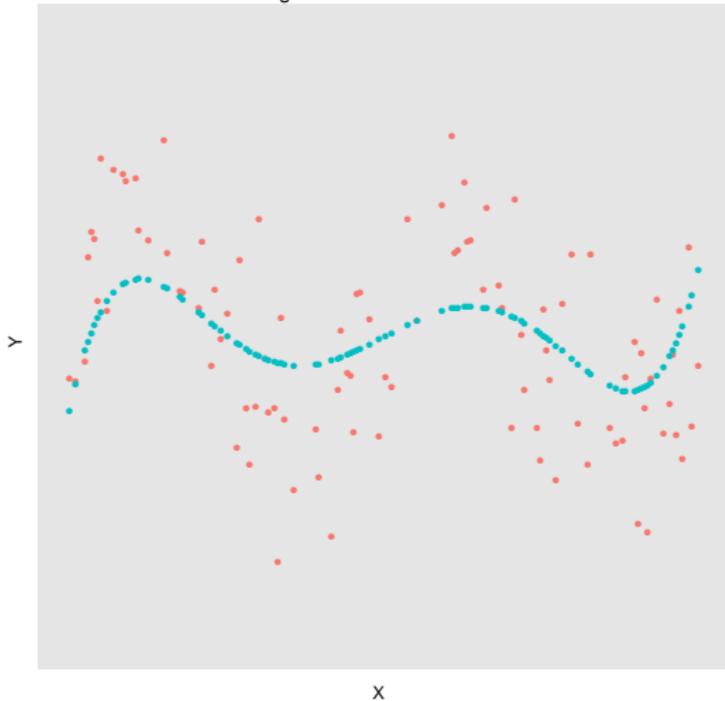
Lasso Regression with Lambda = 0.273



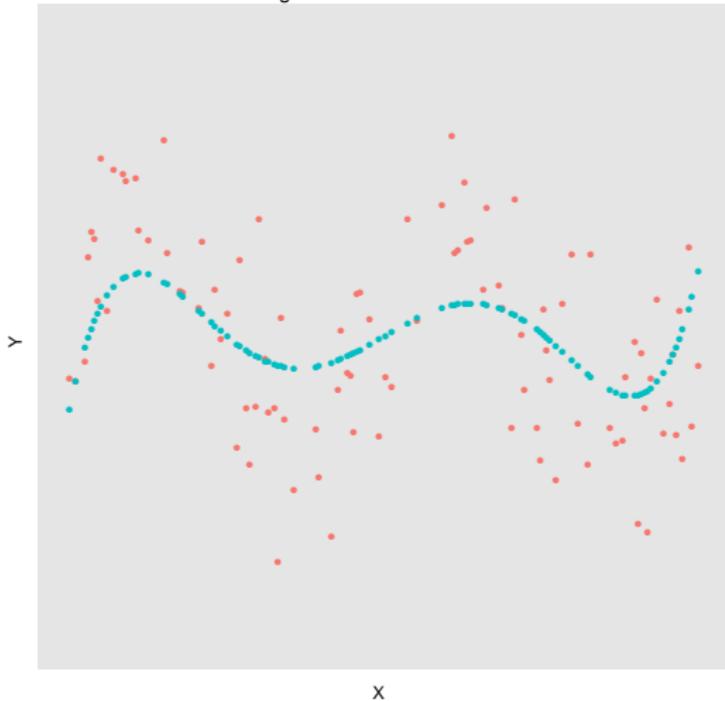
Lasso Regression with Lambda = 0.249



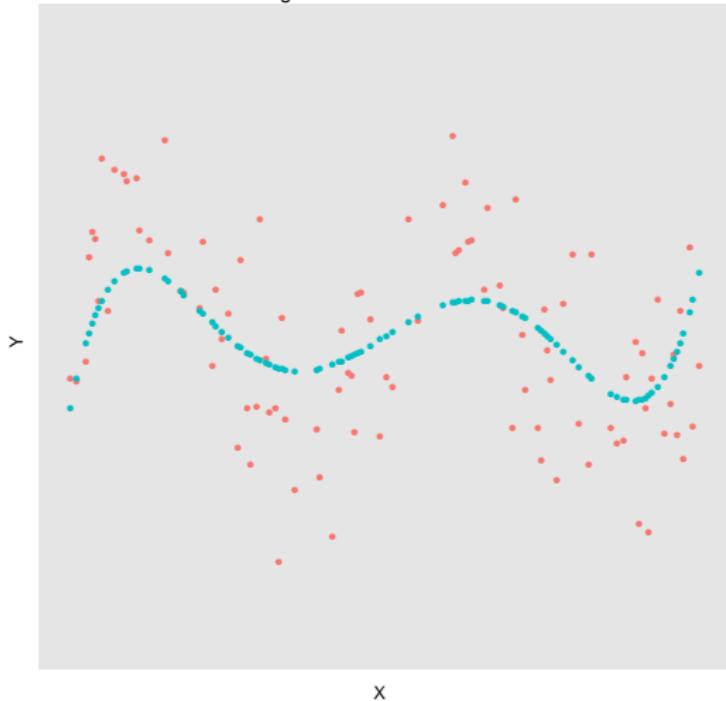
Lasso Regression with Lambda = 0.227



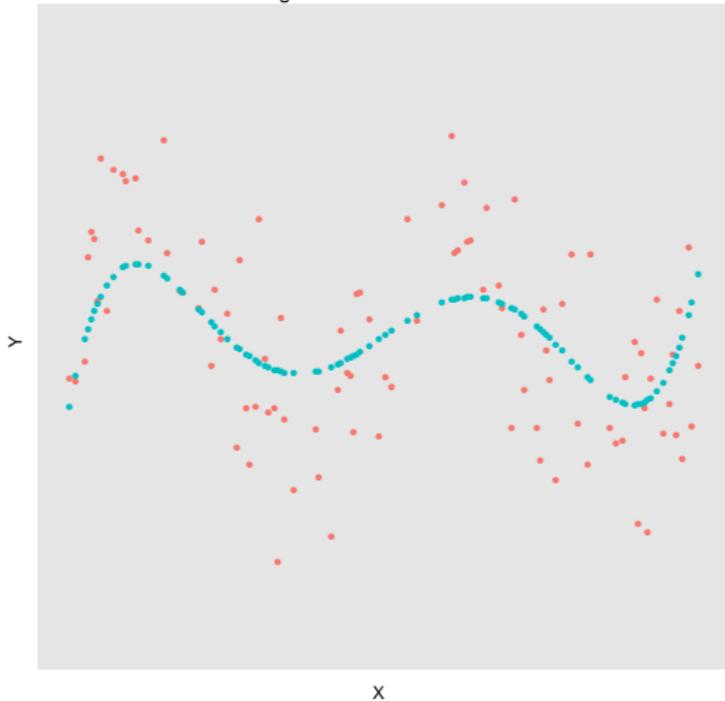
Lasso Regression with Lambda = 0.207



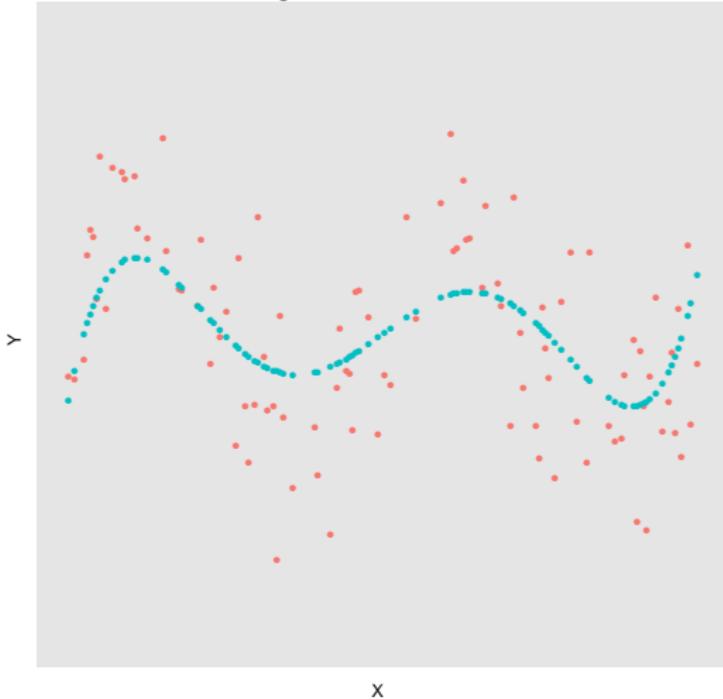
Lasso Regression with Lambda = 0.188



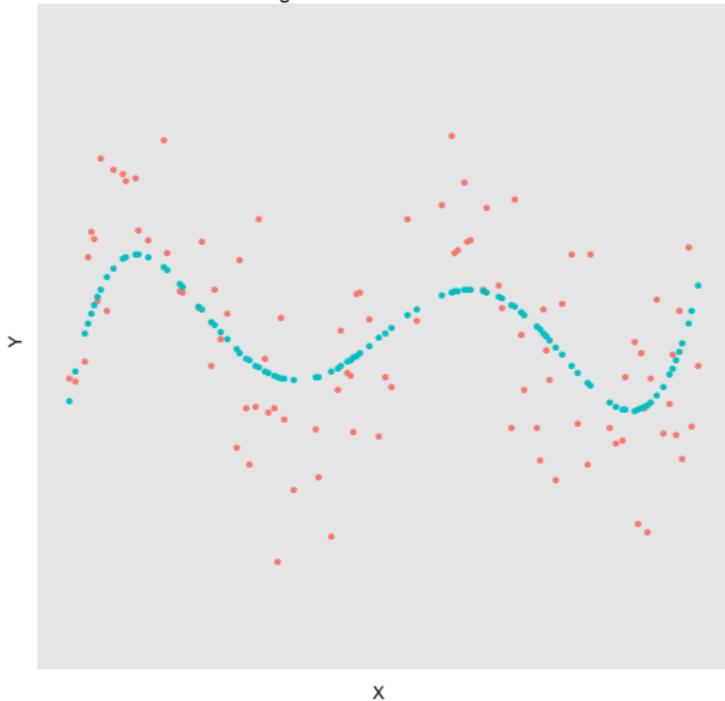
Lasso Regression with Lambda = 0.172



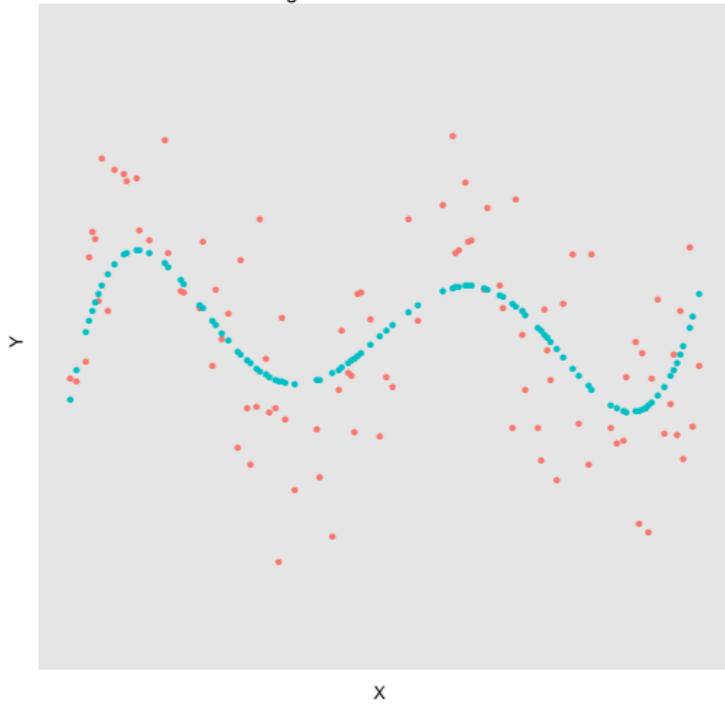
Lasso Regression with Lambda = 0.156



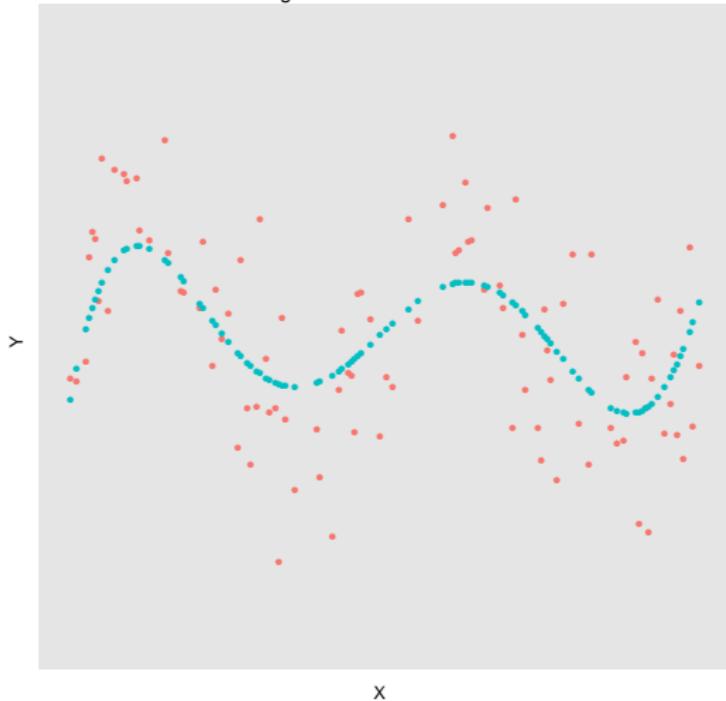
Lasso Regression with Lambda = 0.142



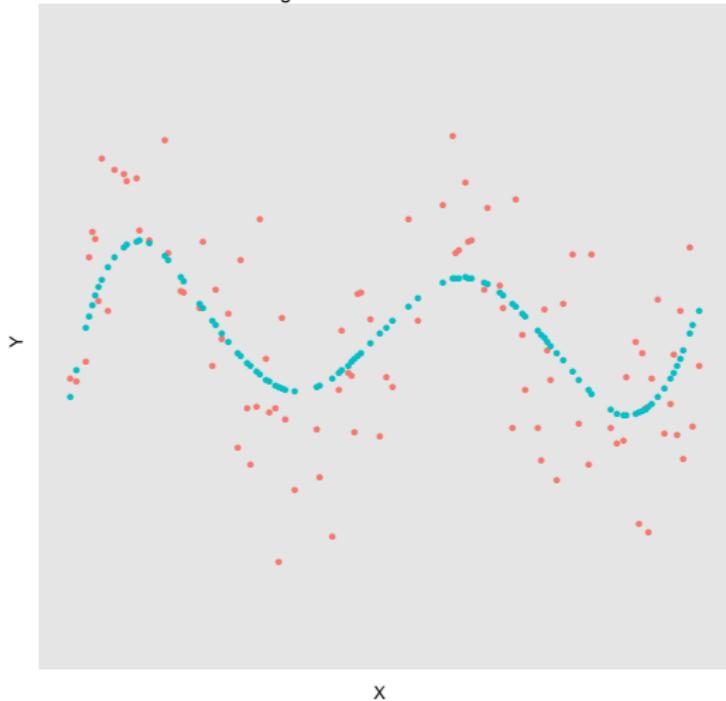
Lasso Regression with Lambda = 0.13



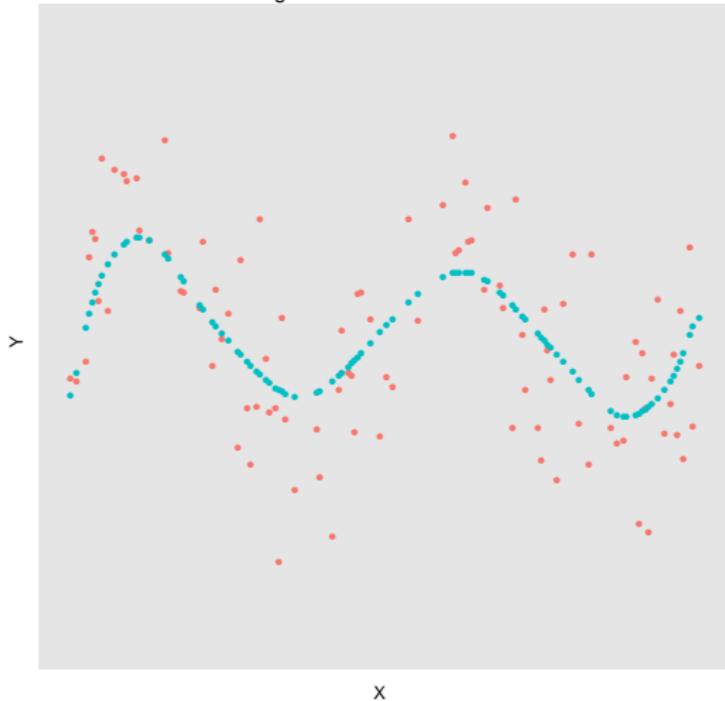
Lasso Regression with Lambda = 0.118



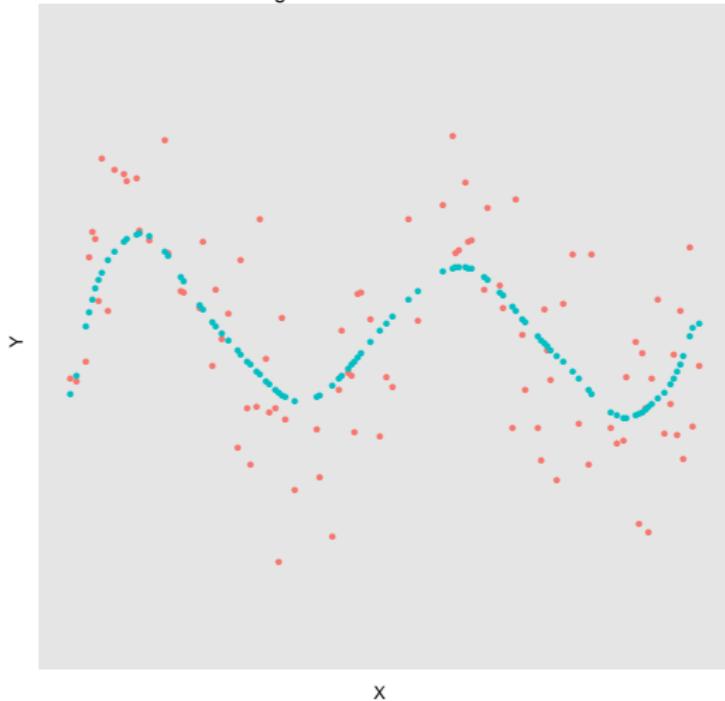
Lasso Regression with Lambda = 0.108



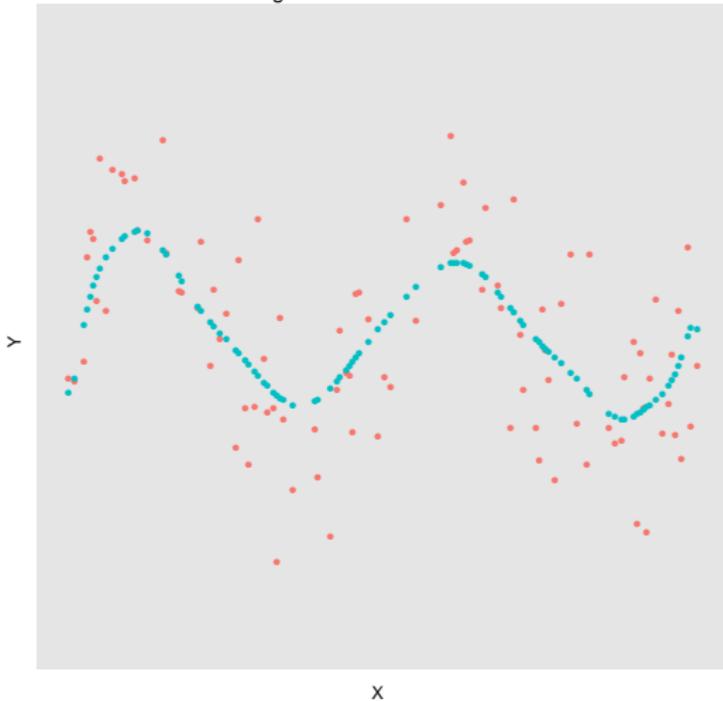
Lasso Regression with Lambda = 0.0982



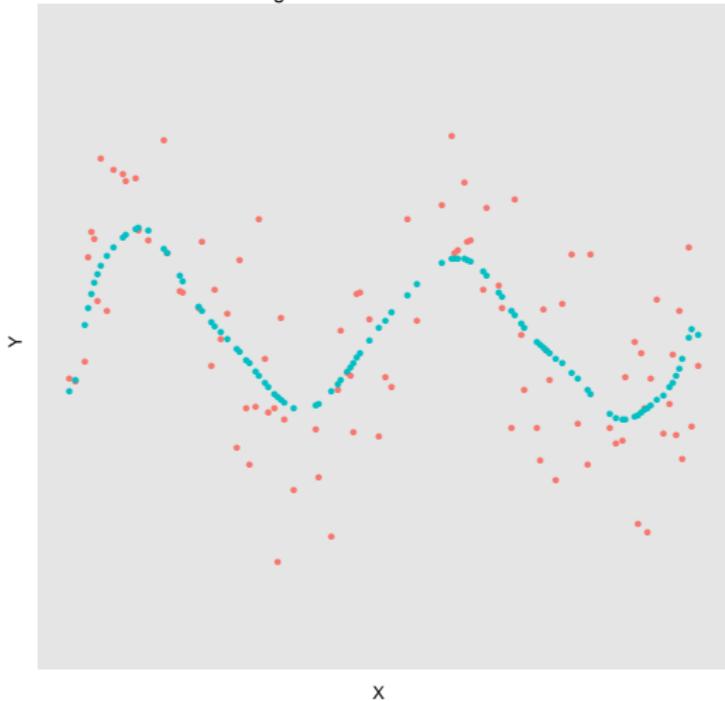
Lasso Regression with Lambda = 0.0895



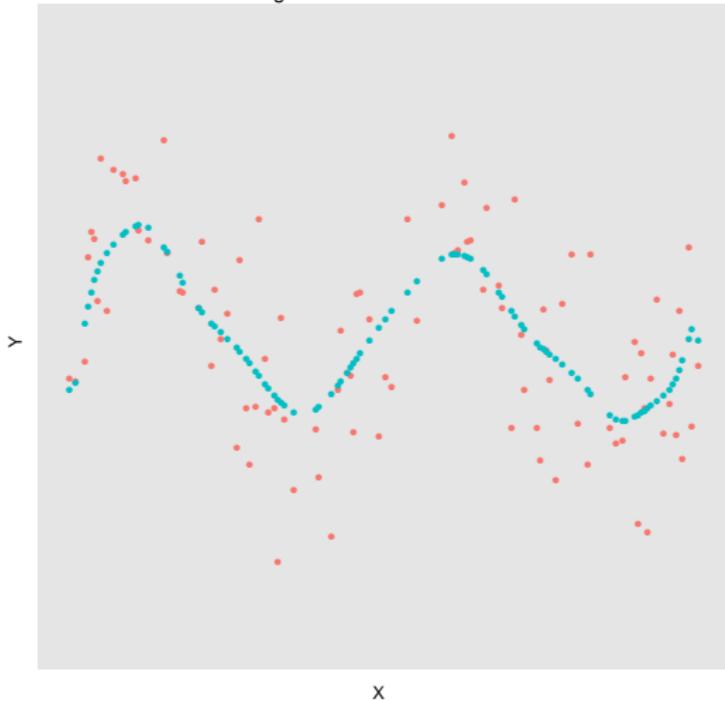
Lasso Regression with Lambda = 0.0815



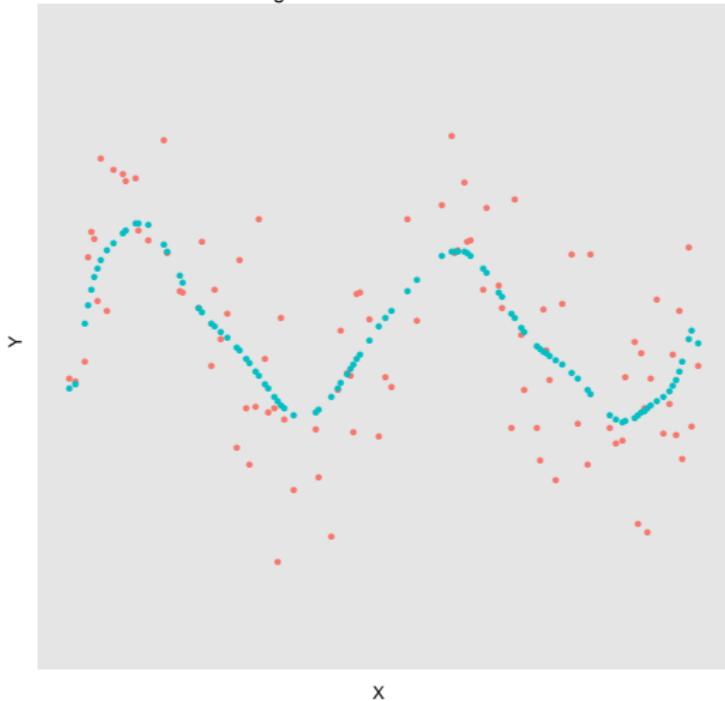
Lasso Regression with Lambda = 0.0743



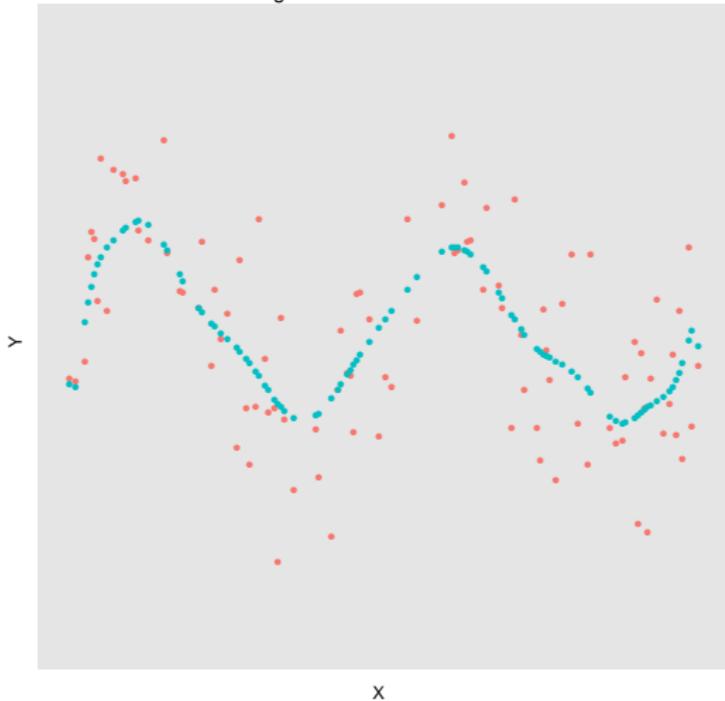
Lasso Regression with Lambda = 0.0677



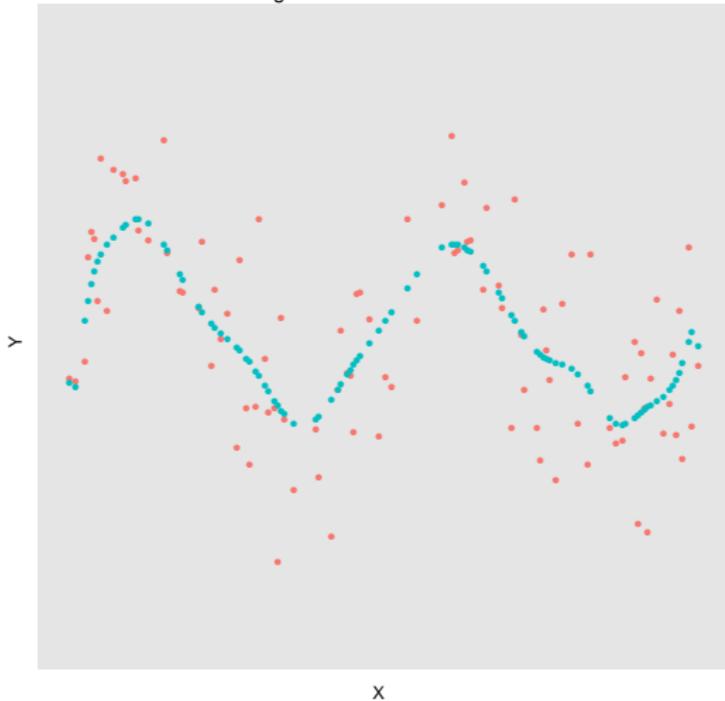
Lasso Regression with Lambda = 0.0617



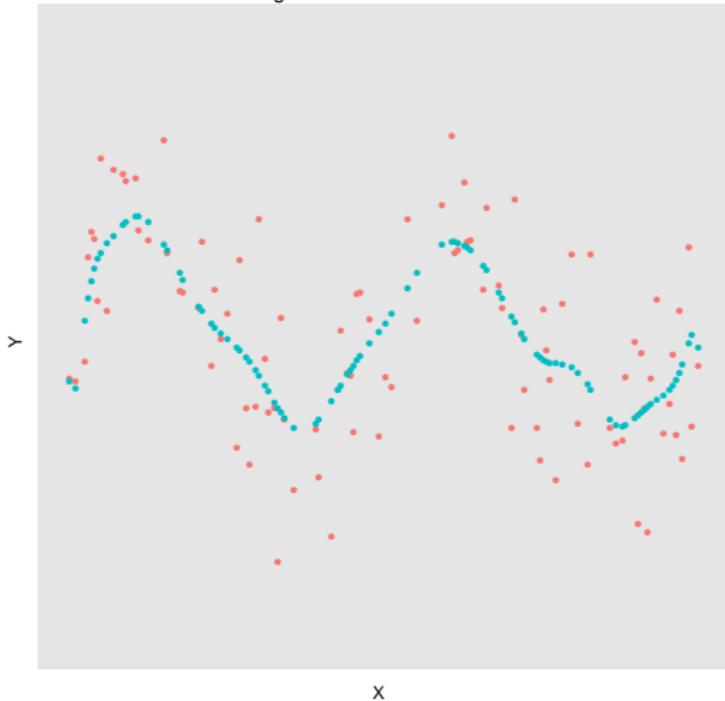
Lasso Regression with Lambda = 0.0562



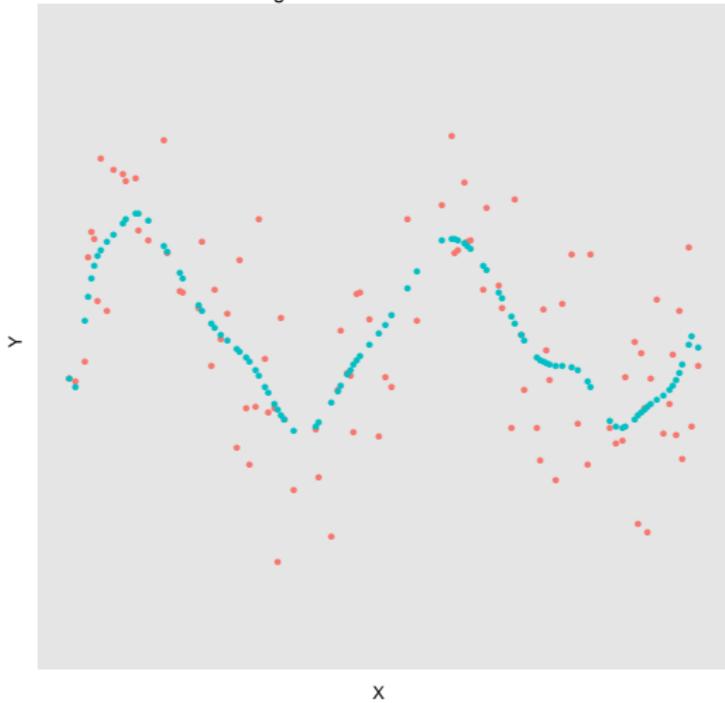
Lasso Regression with Lambda = 0.0512



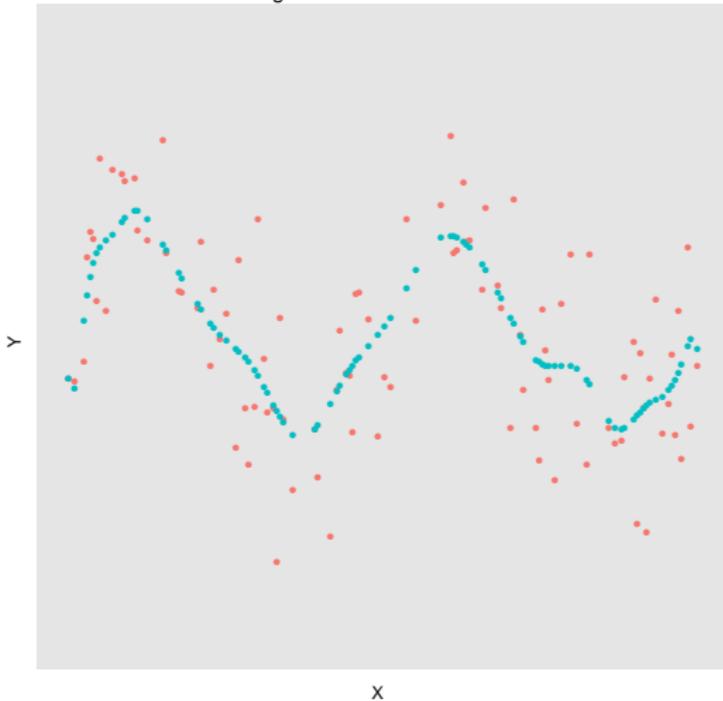
Lasso Regression with Lambda = 0.0467



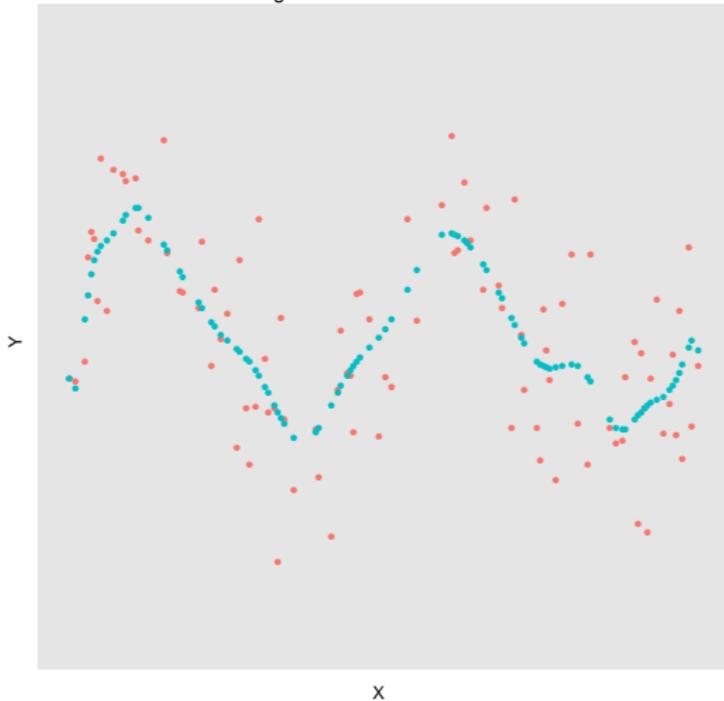
Lasso Regression with Lambda = 0.0425



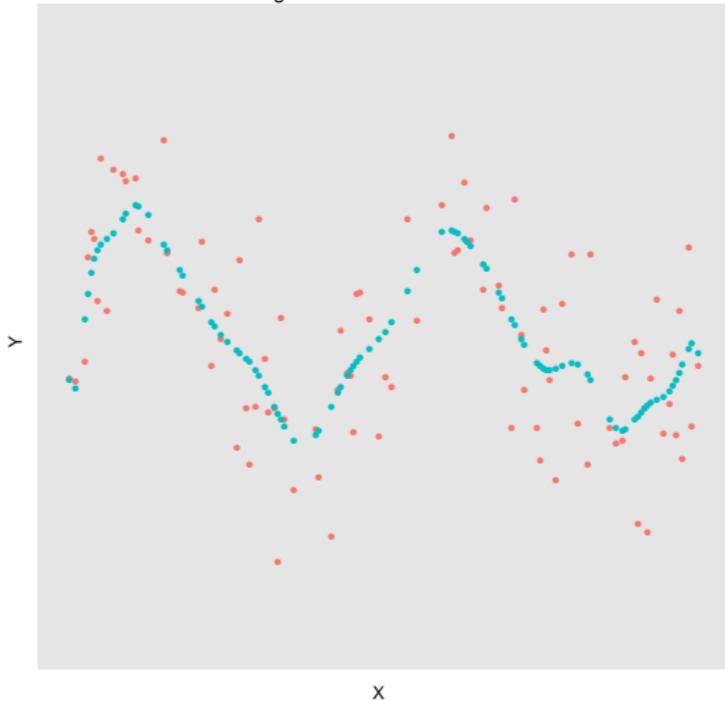
Lasso Regression with Lambda = 0.0387



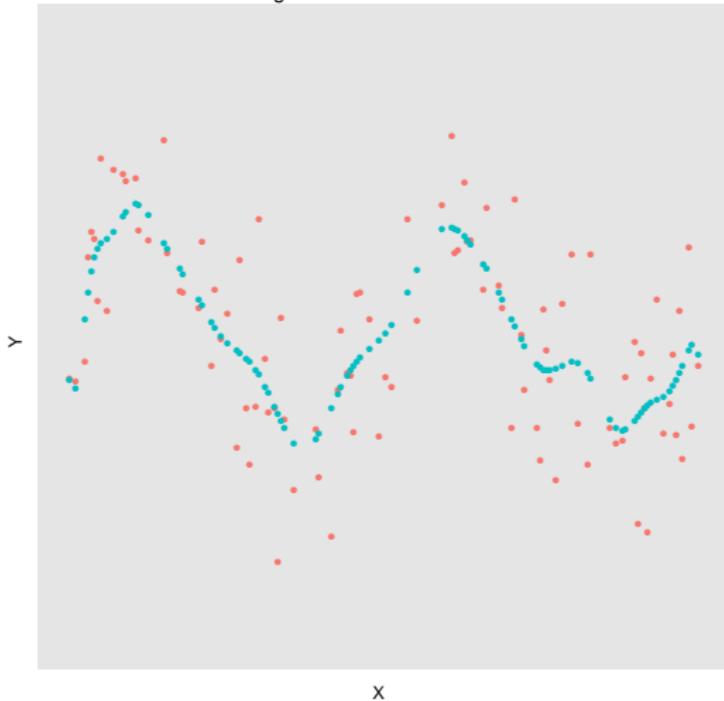
Lasso Regression with Lambda = 0.0353



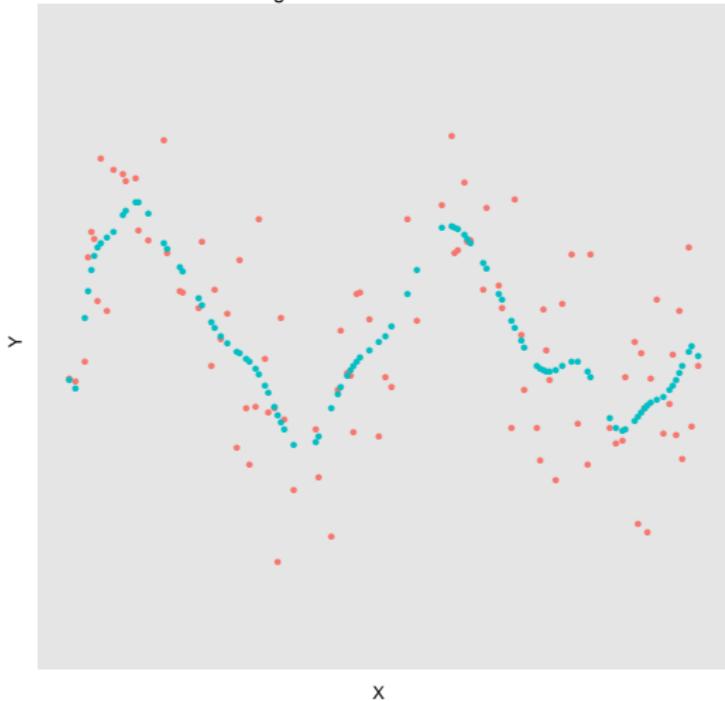
Lasso Regression with Lambda = 0.0322



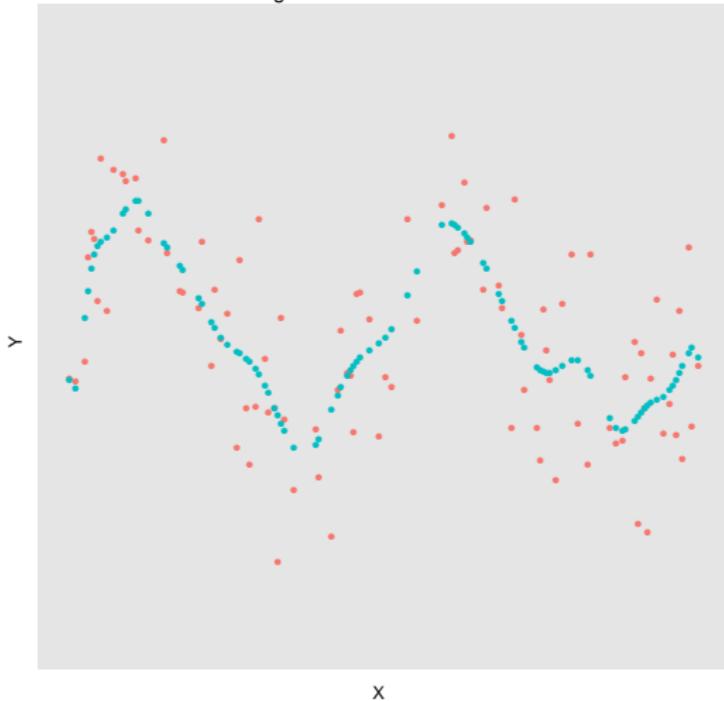
Lasso Regression with Lambda = 0.0293



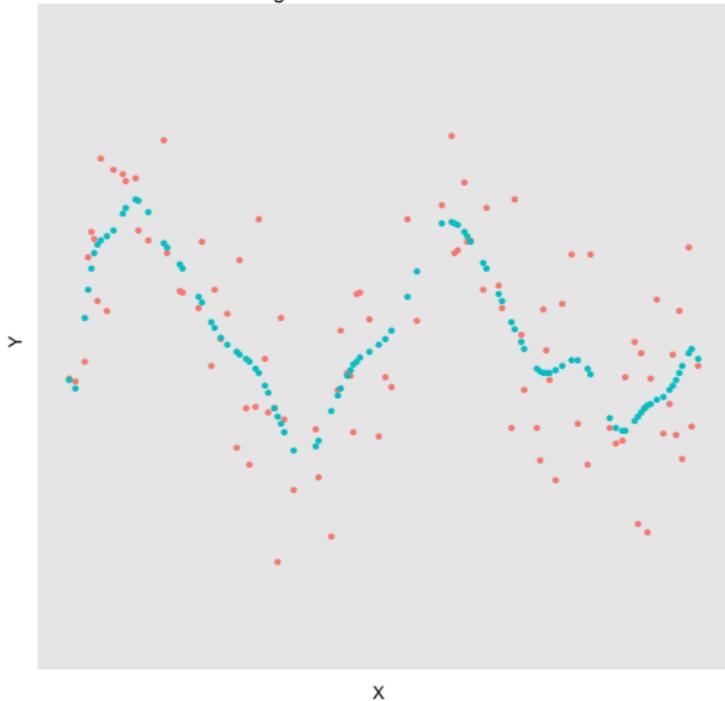
Lasso Regression with Lambda = 0.0267



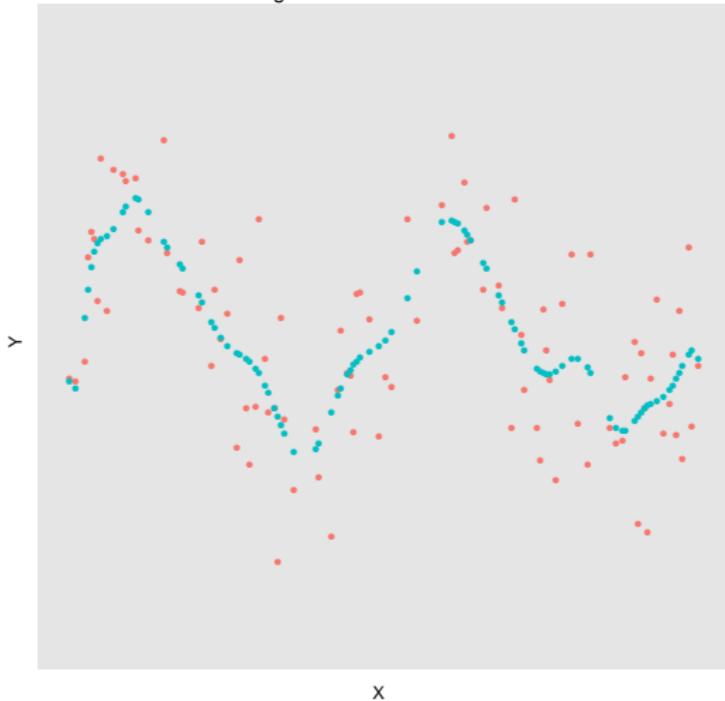
Lasso Regression with Lambda = 0.0243



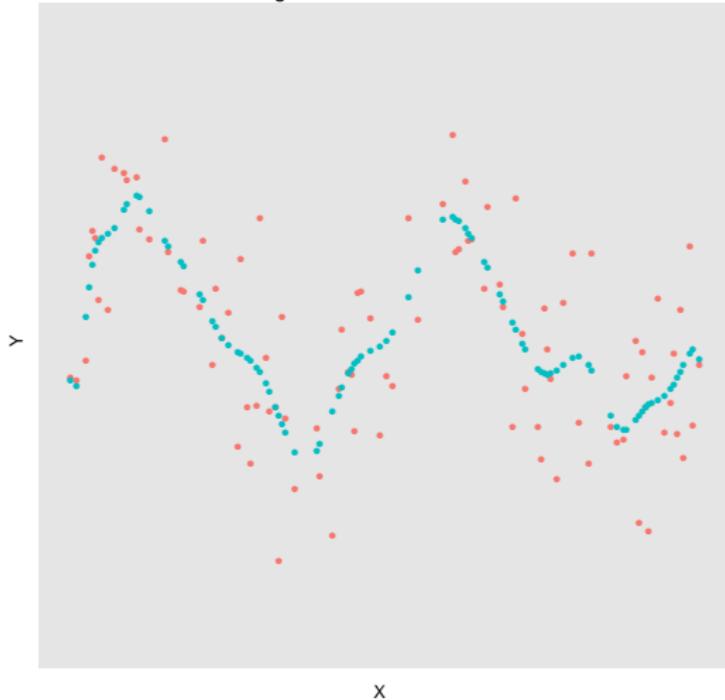
Lasso Regression with Lambda = 0.0222



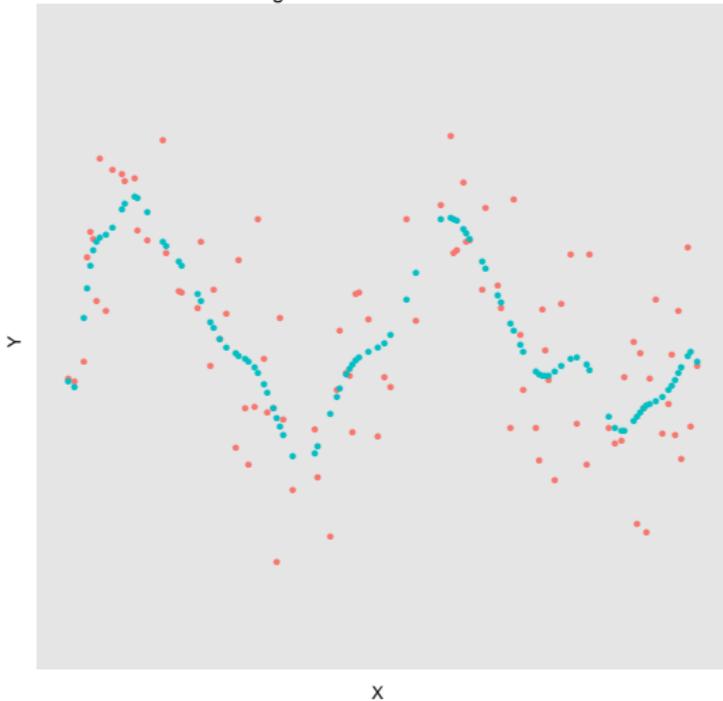
Lasso Regression with Lambda = 0.0202



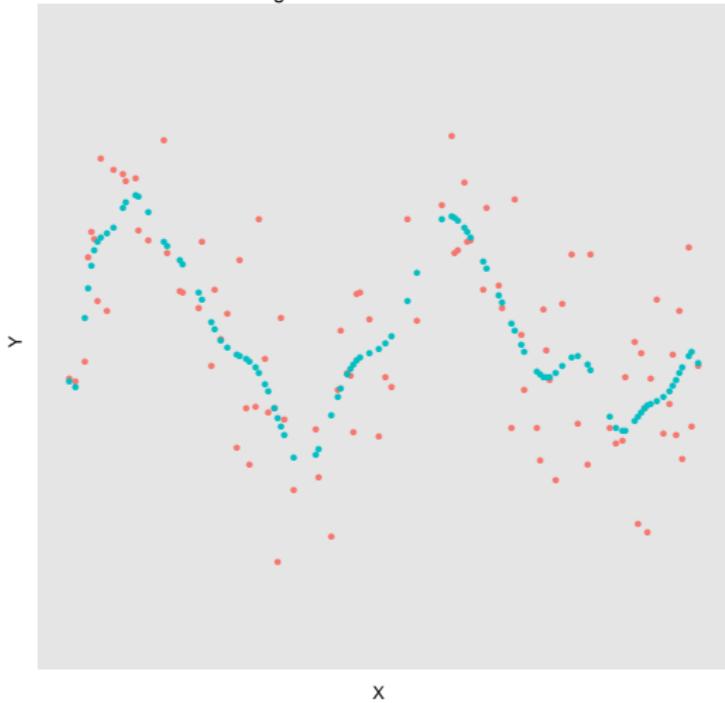
Lasso Regression with Lambda = 0.0184



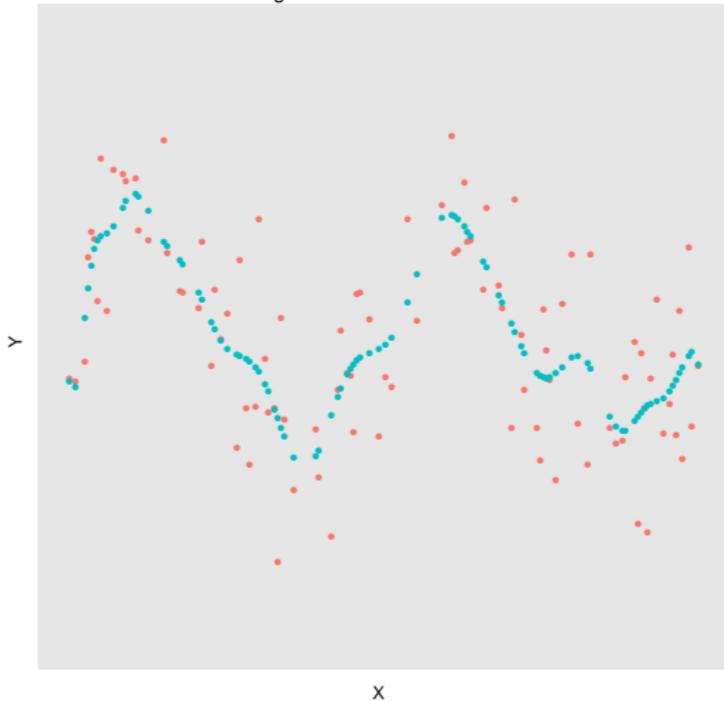
Lasso Regression with Lambda = 0.0168



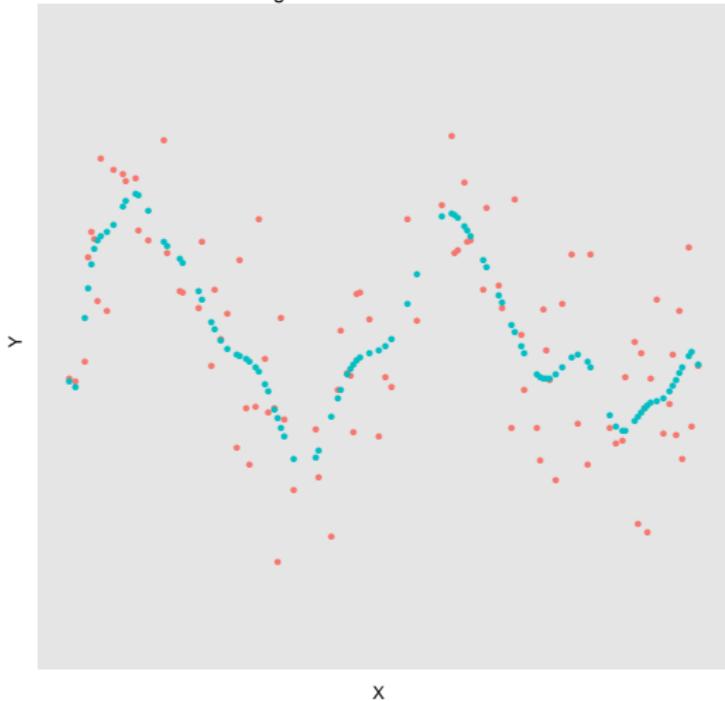
Lasso Regression with Lambda = 0.0153



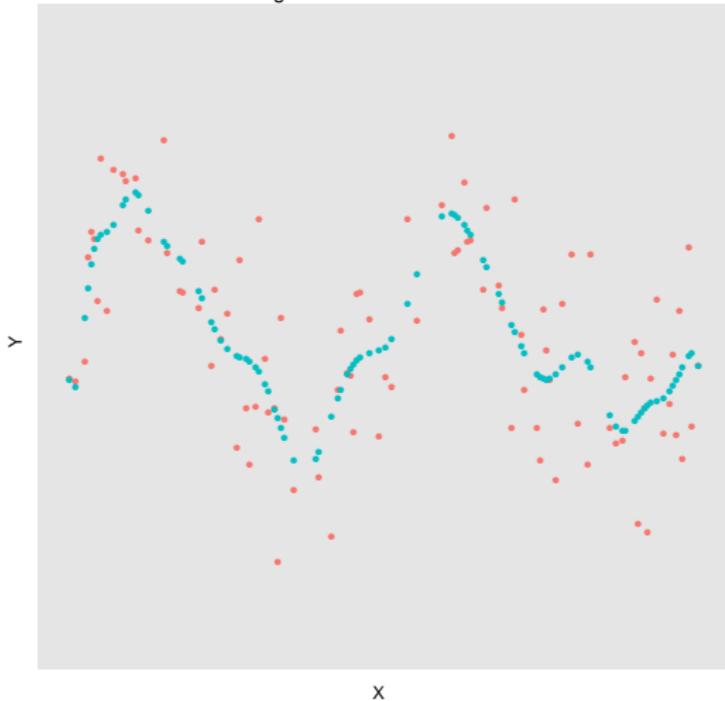
Lasso Regression with Lambda = 0.0139



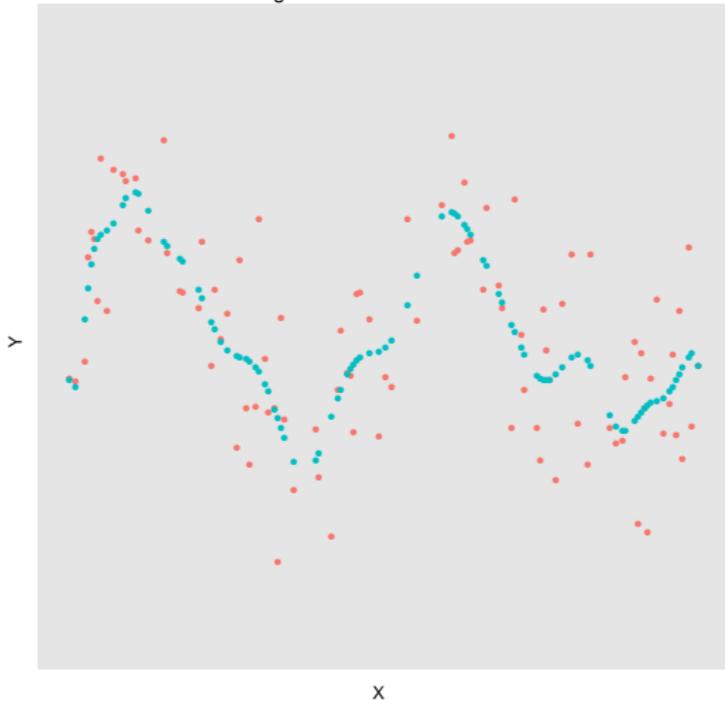
Lasso Regression with Lambda = 0.0127



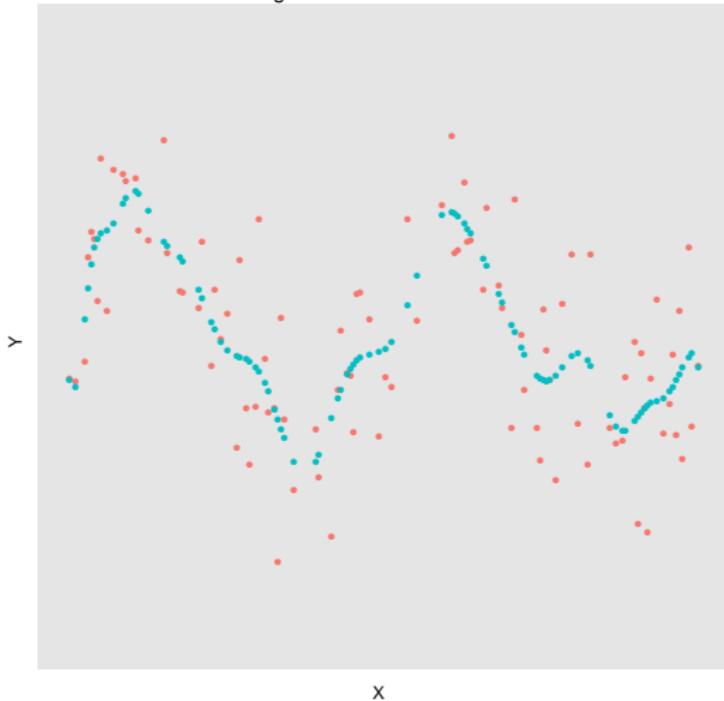
Lasso Regression with Lambda = 0.0116



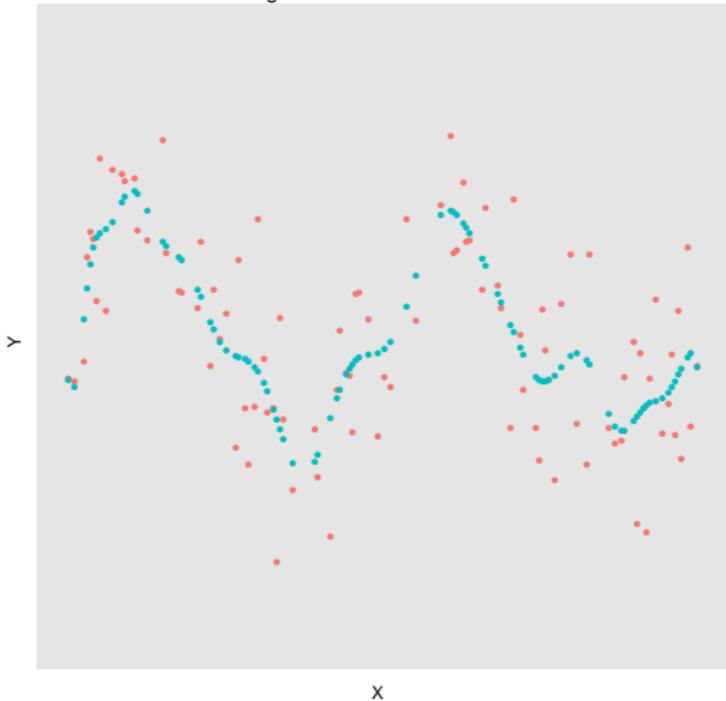
Lasso Regression with Lambda = 0.0105



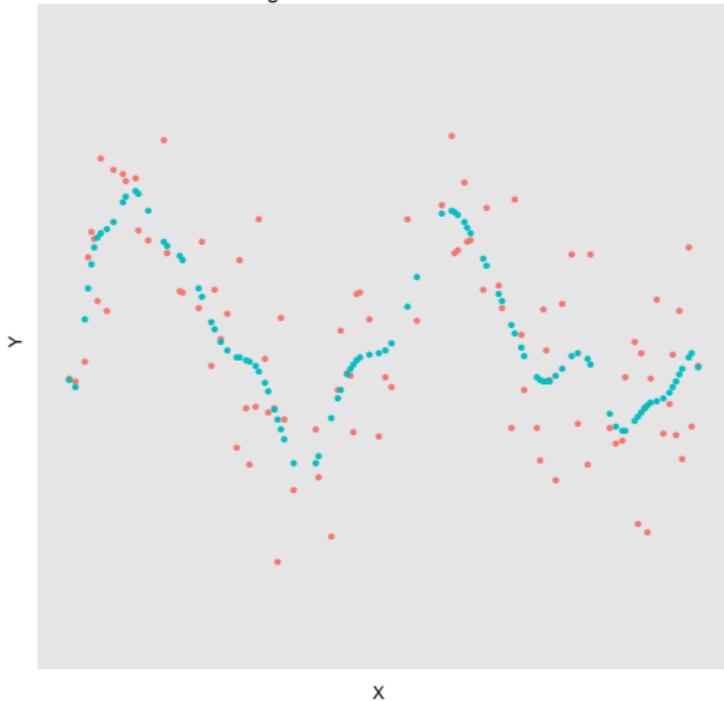
Lasso Regression with Lambda = 0.0096



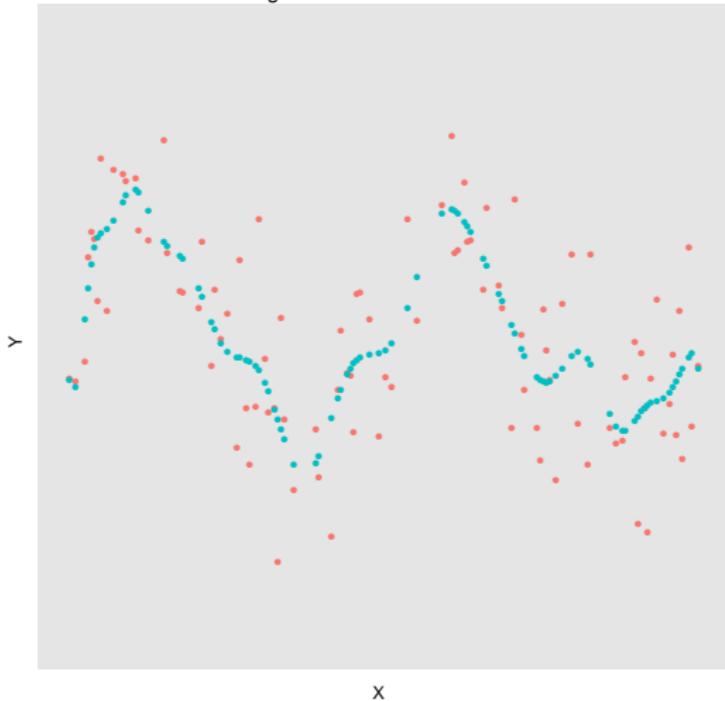
Lasso Regression with Lambda = 0.00874



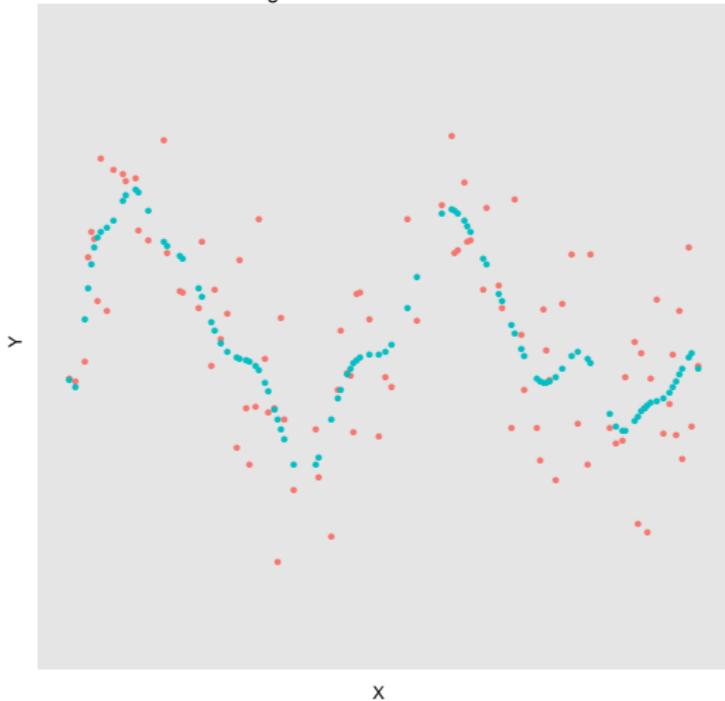
Lasso Regression with Lambda = 0.00797



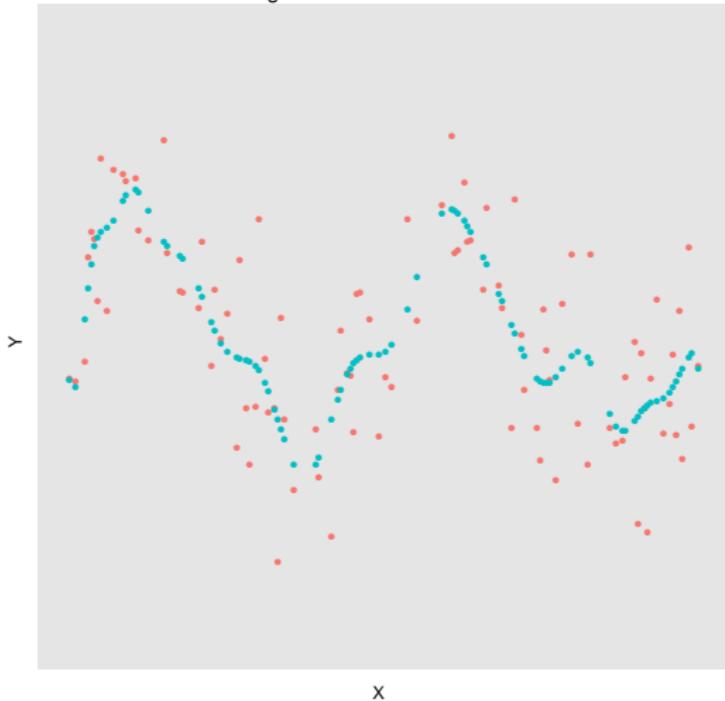
Lasso Regression with Lambda = 0.00726



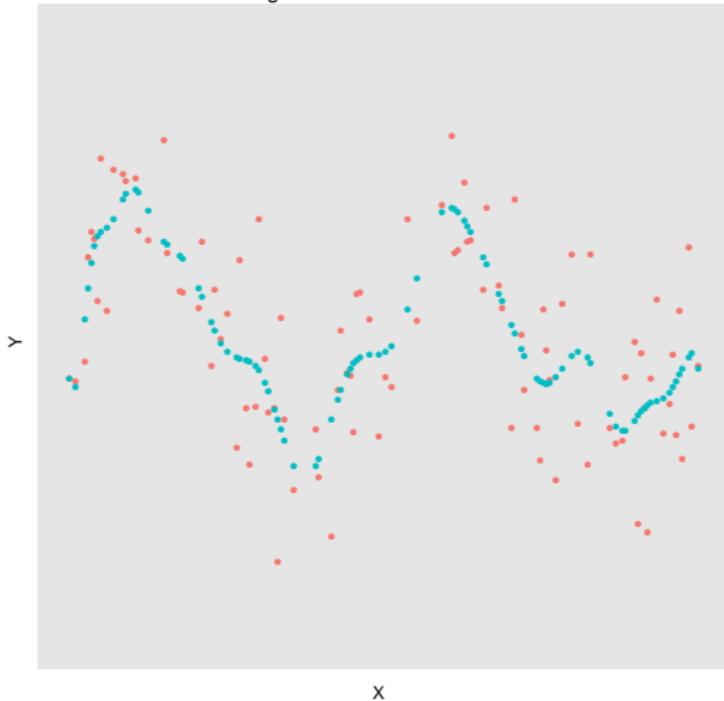
Lasso Regression with Lambda = 0.00661



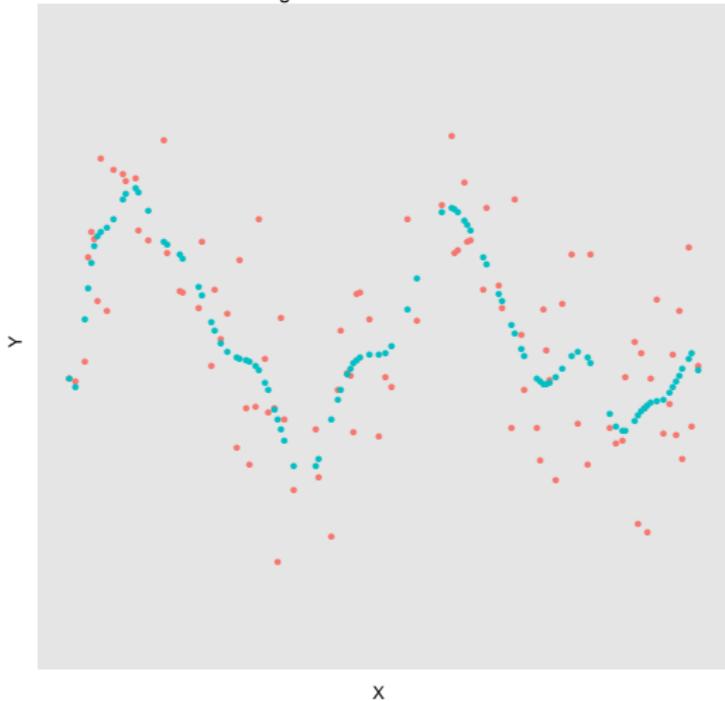
Lasso Regression with Lambda = 0.00603



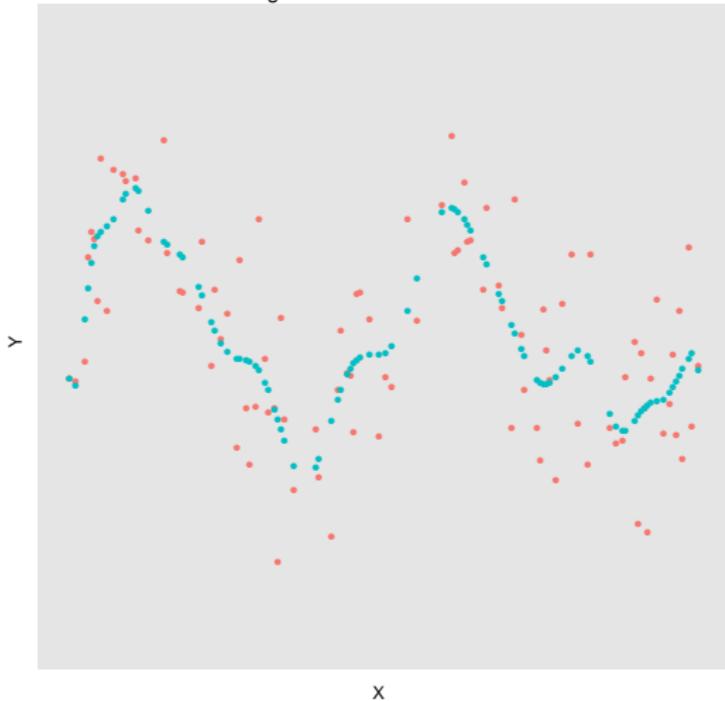
Lasso Regression with Lambda = 0.00549



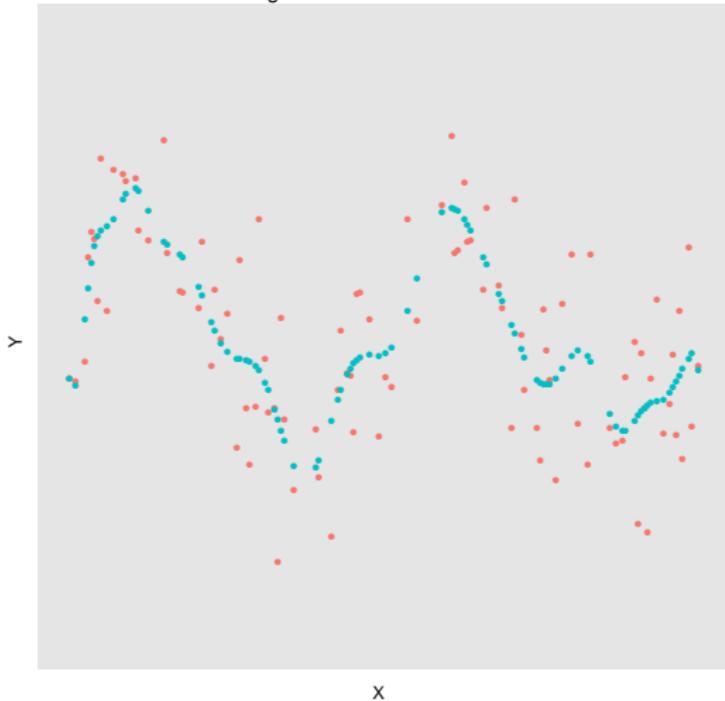
Lasso Regression with Lambda = 0.005



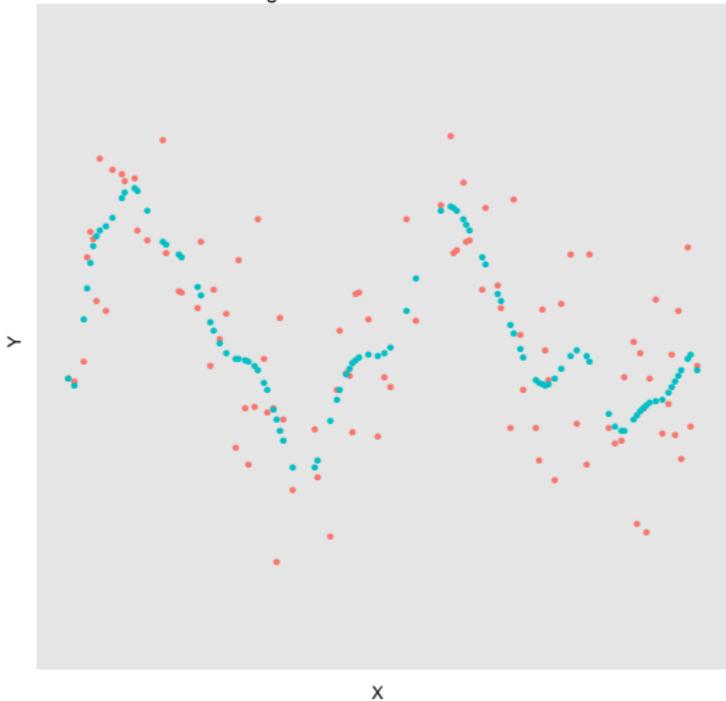
Lasso Regression with Lambda = 0.00456



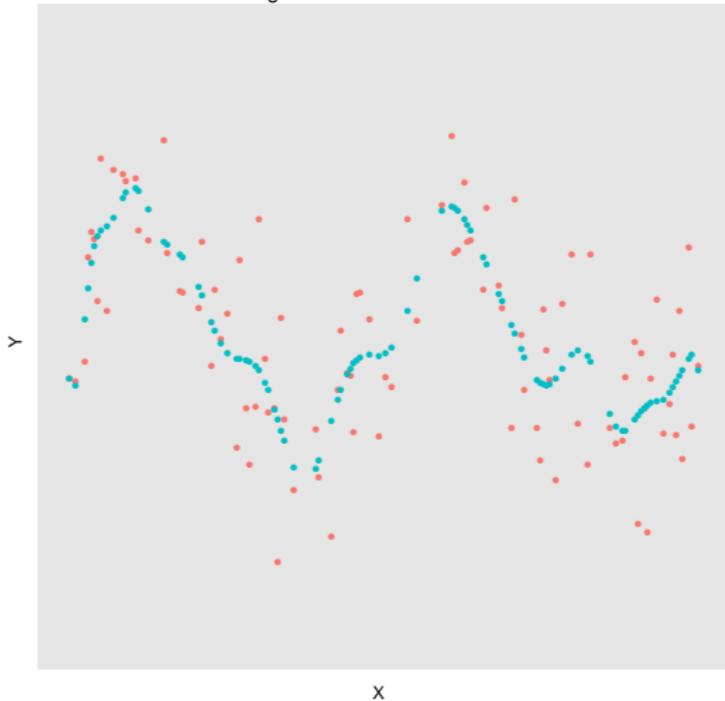
Lasso Regression with Lambda = 0.00415



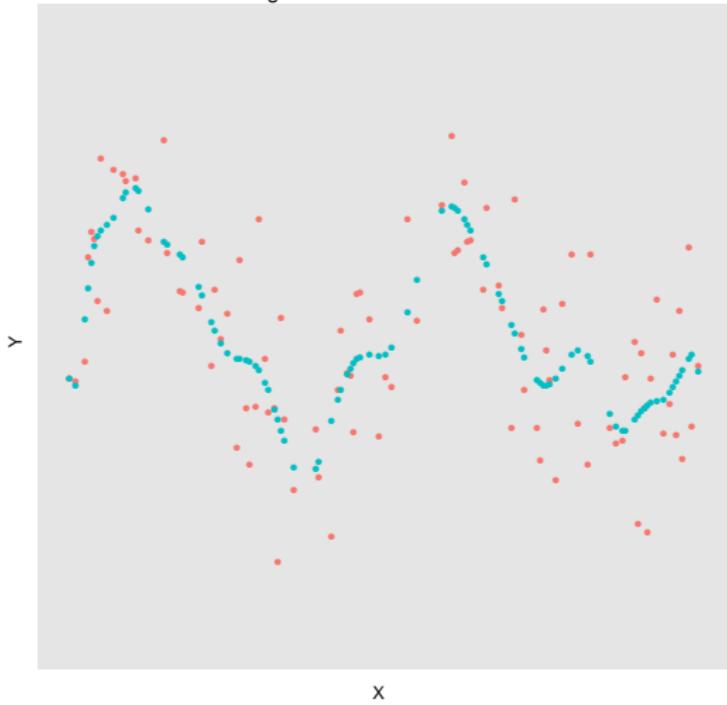
Lasso Regression with Lambda = 0.00378



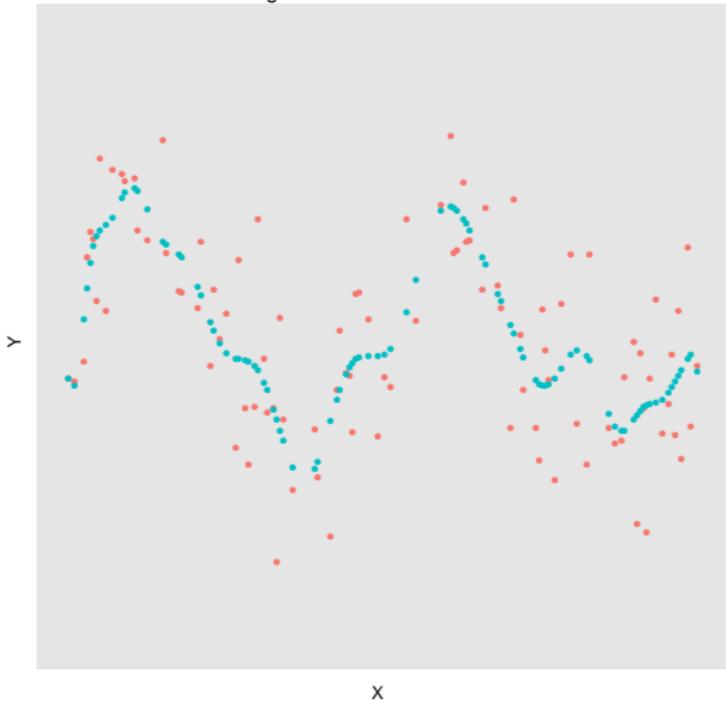
Lasso Regression with Lambda = 0.00345



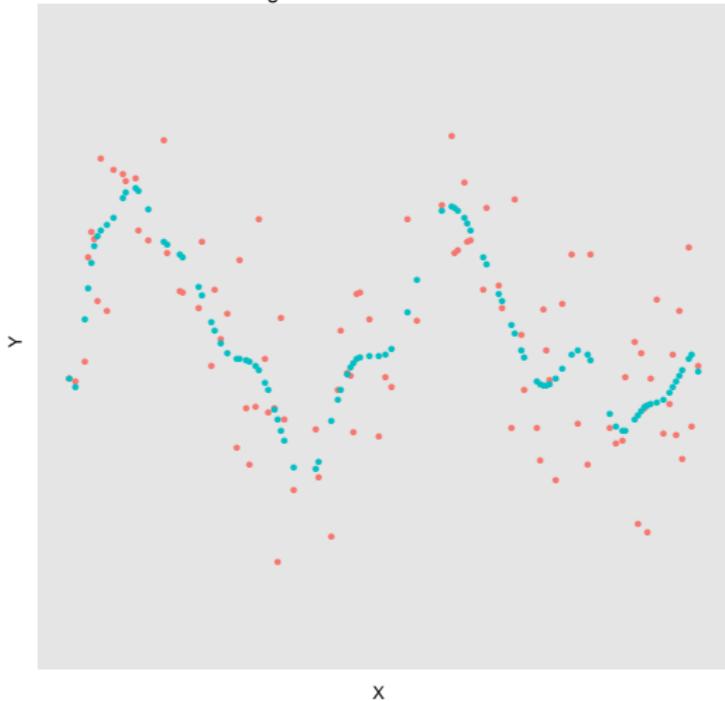
Lasso Regression with Lambda = 0.00314



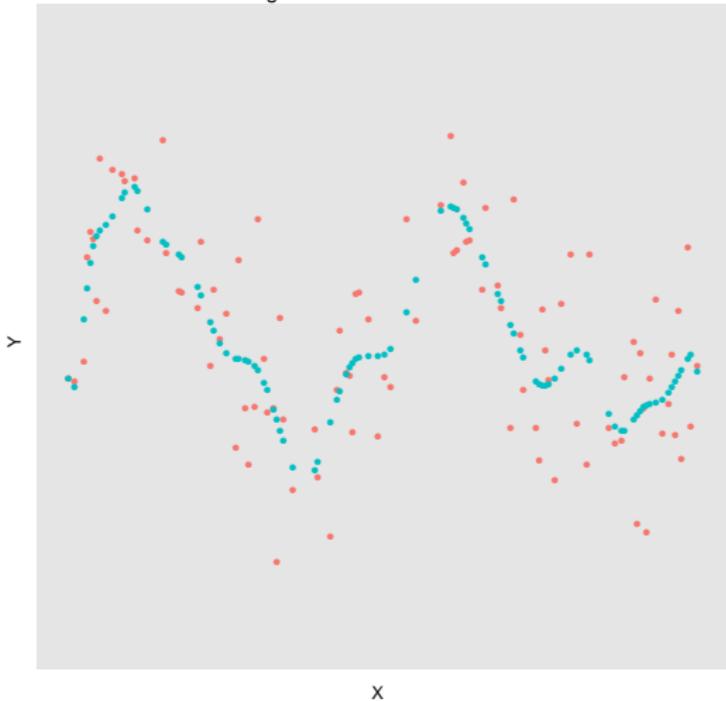
Lasso Regression with Lambda = 0.00286



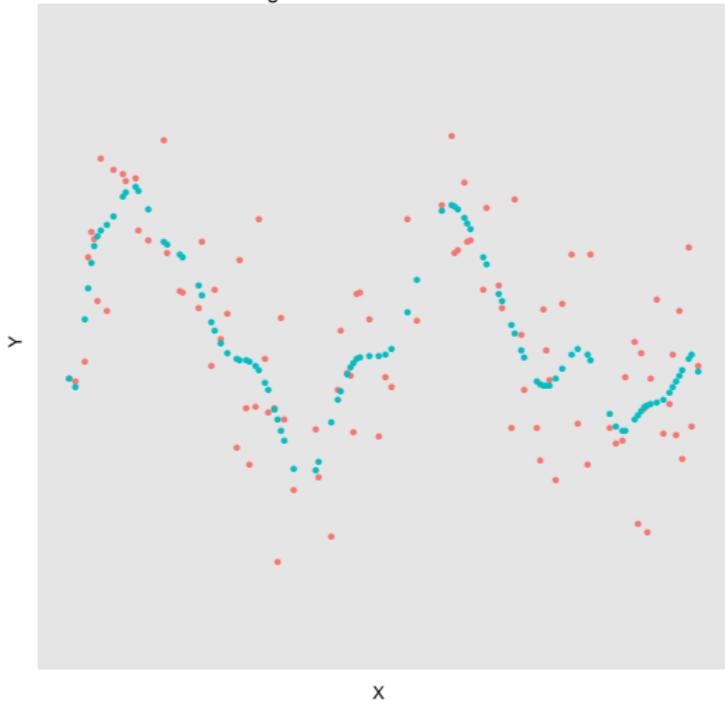
Lasso Regression with Lambda = 0.00261



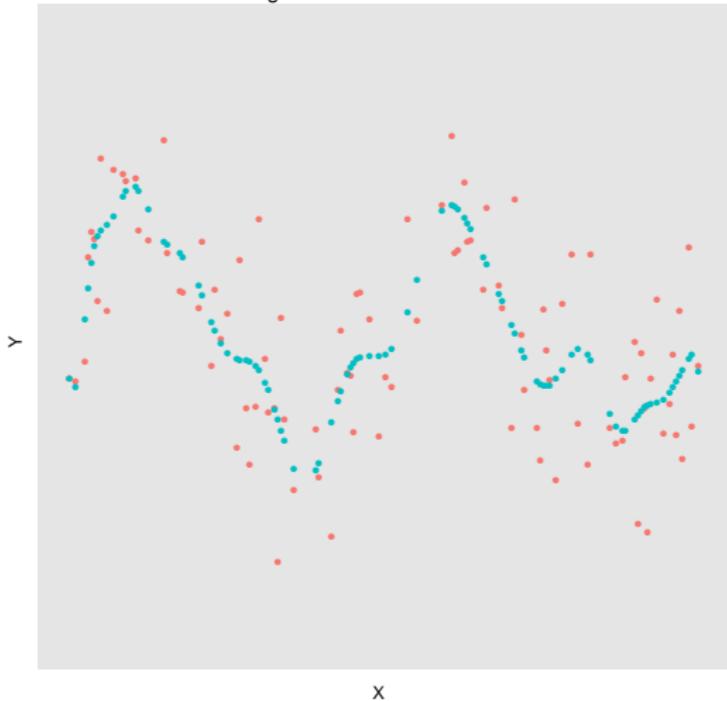
Lasso Regression with Lambda = 0.00238



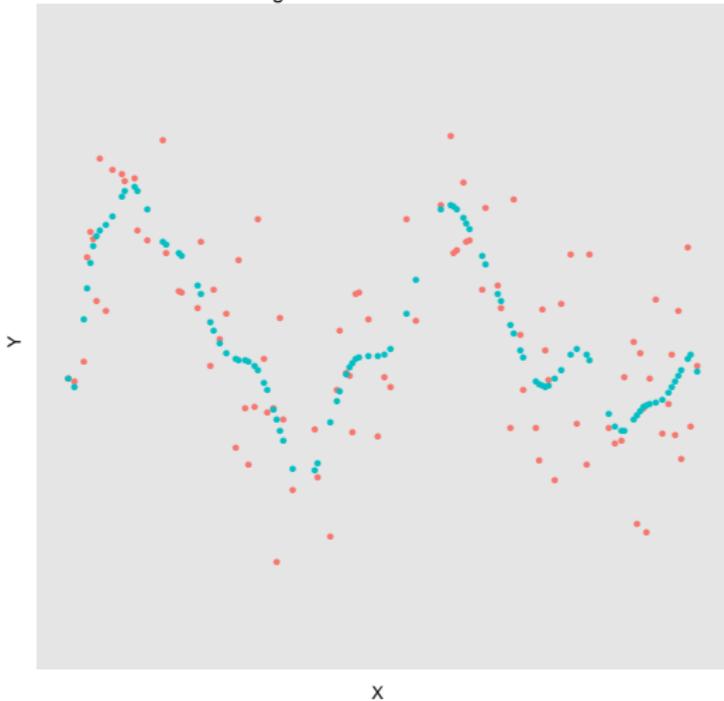
Lasso Regression with Lambda = 0.00217



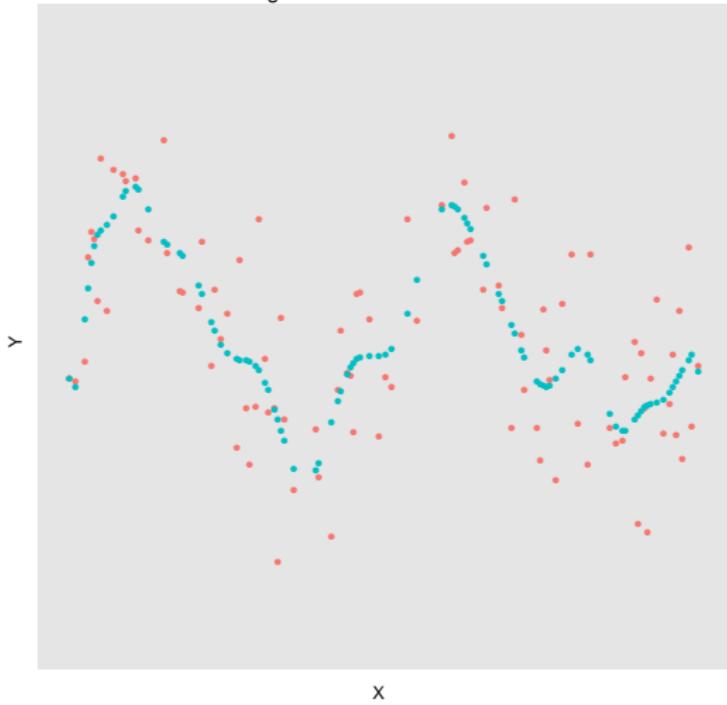
Lasso Regression with Lambda = 0.00197



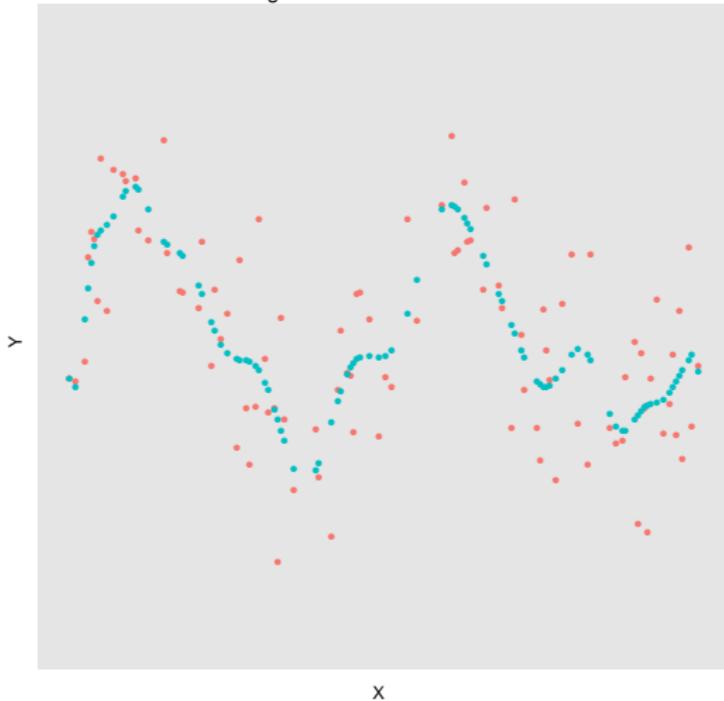
Lasso Regression with Lambda = 0.0018



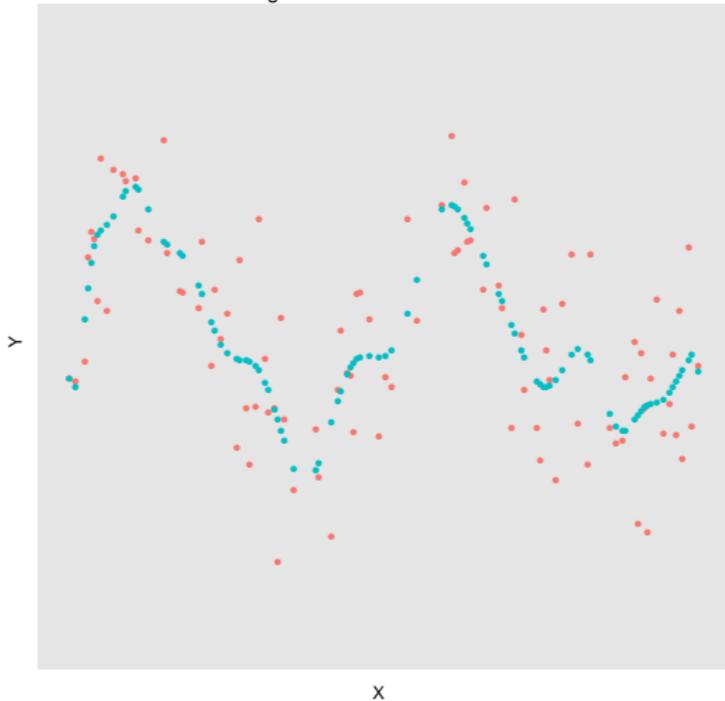
Lasso Regression with Lambda = 0.00164



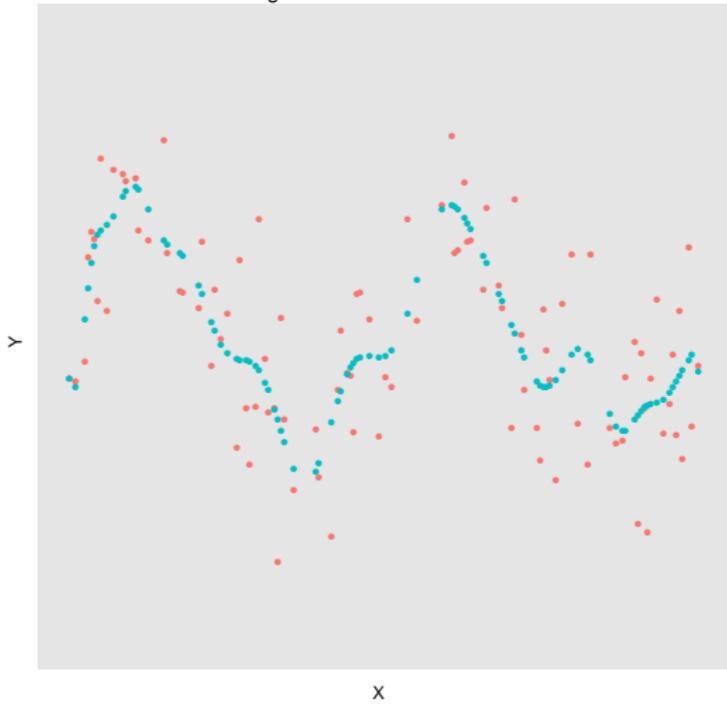
Lasso Regression with Lambda = 0.00149



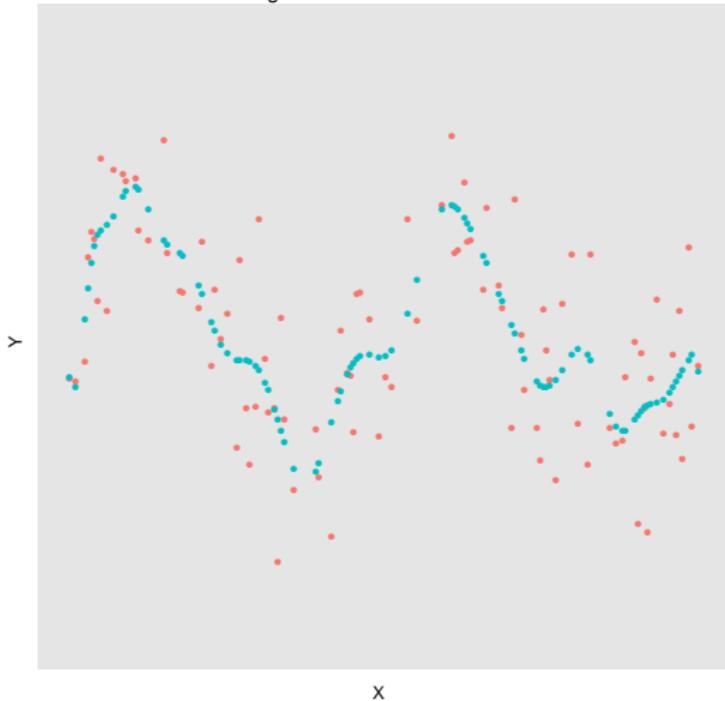
Lasso Regression with Lambda = 0.00136



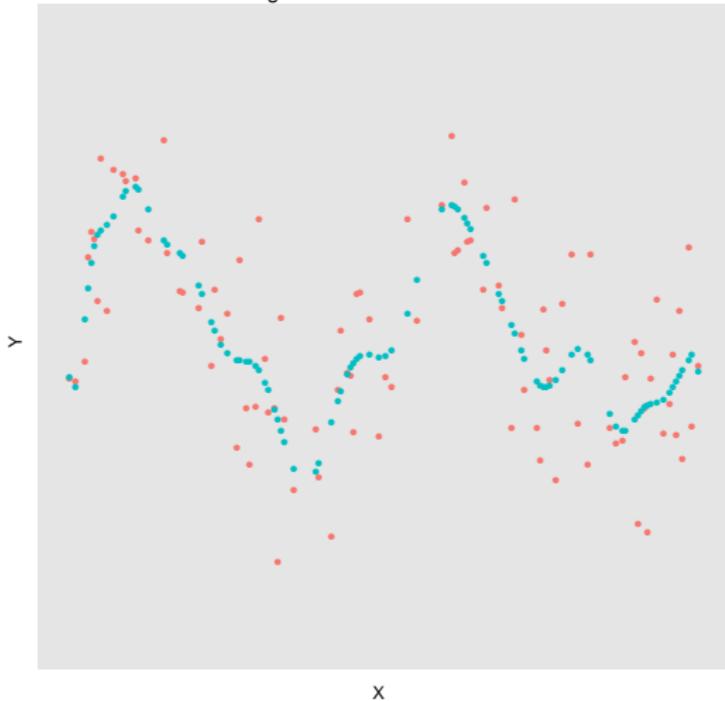
Lasso Regression with Lambda = 0.00124



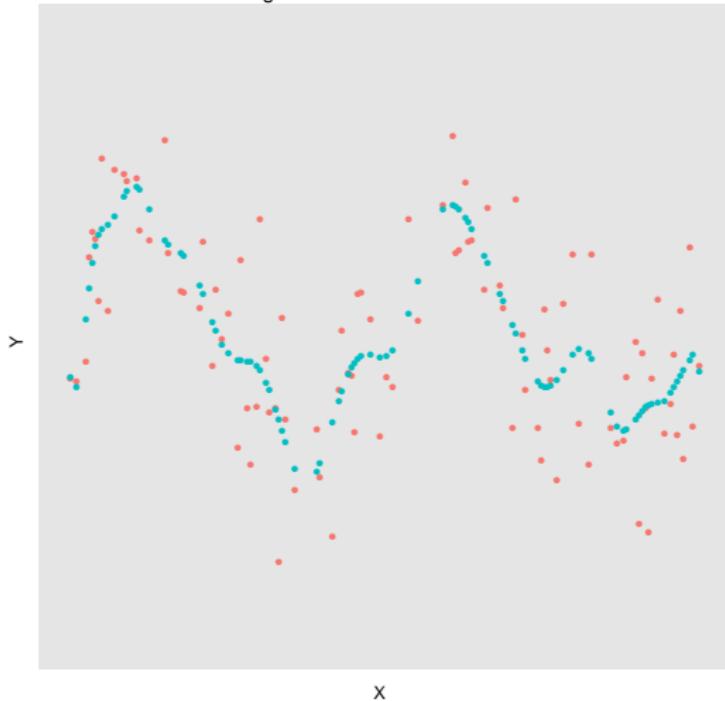
Lasso Regression with Lambda = 0.00113



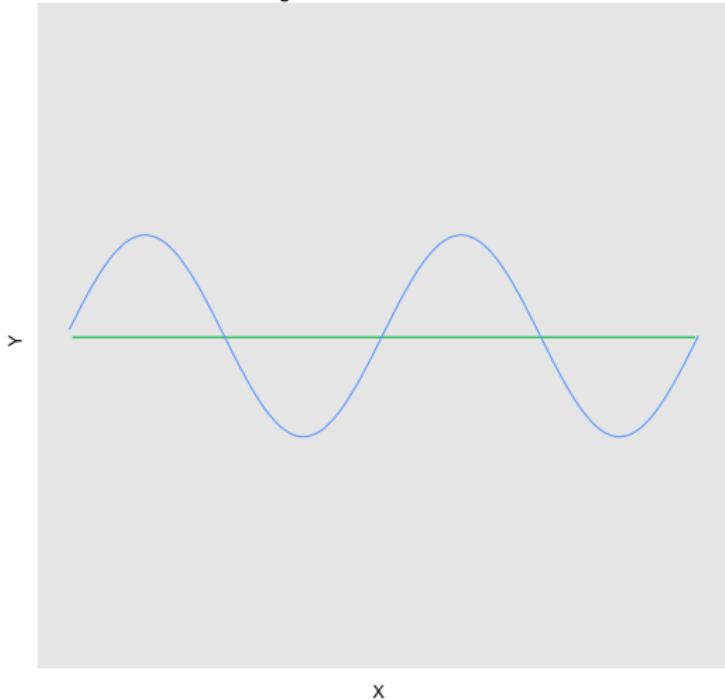
Lasso Regression with Lambda = 0.00103



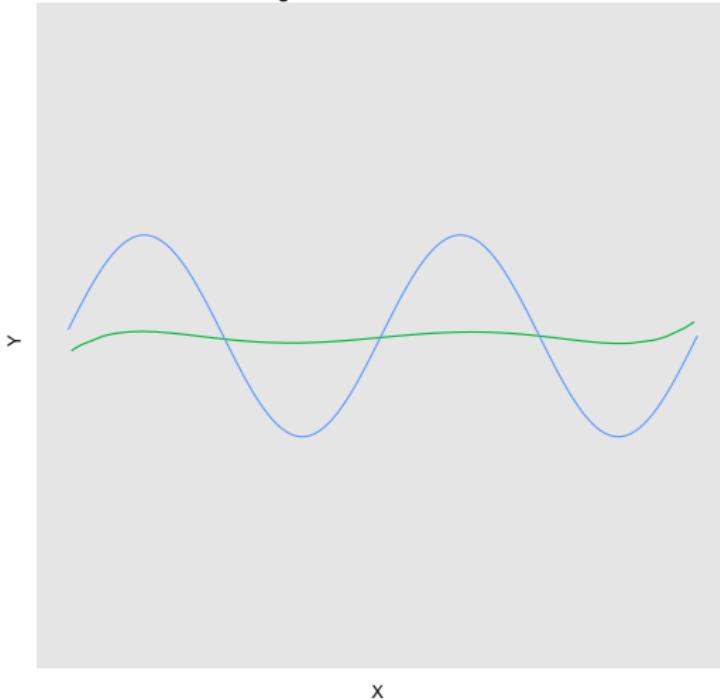
Lasso Regression with Lambda = 0.000937



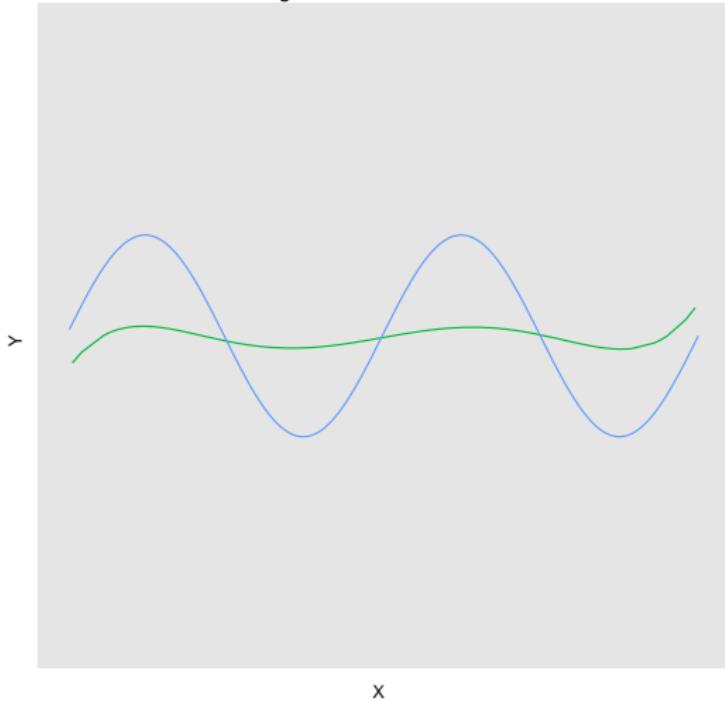
Lasso Regression with Lambda = 0.524



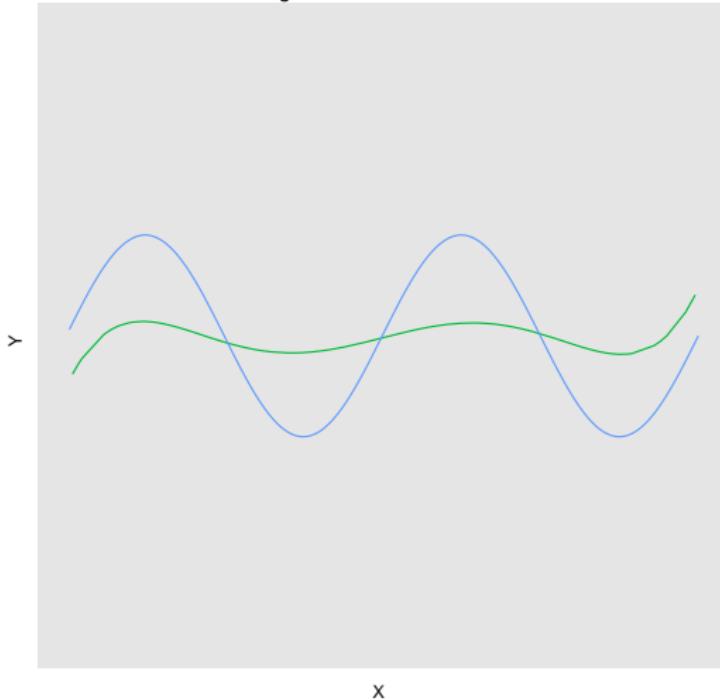
Lasso Regression with Lambda = 0.478



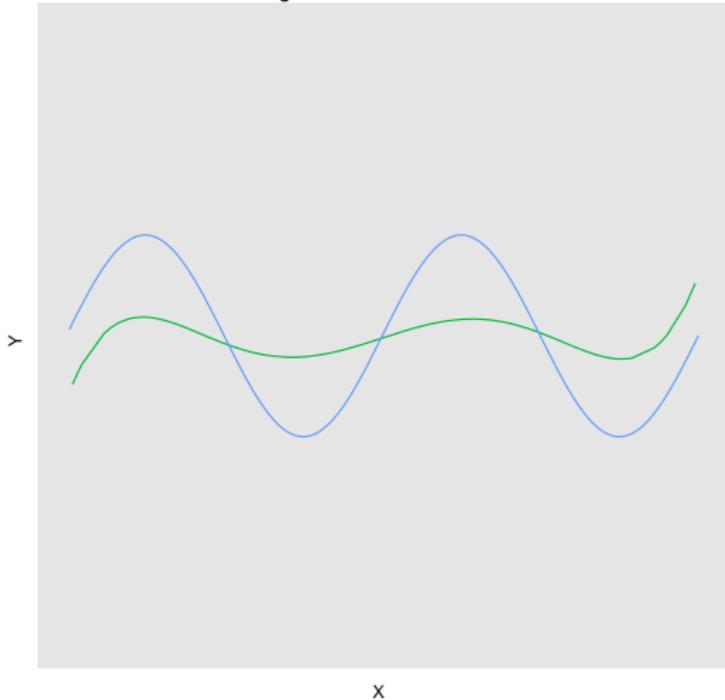
Lasso Regression with Lambda = 0.435



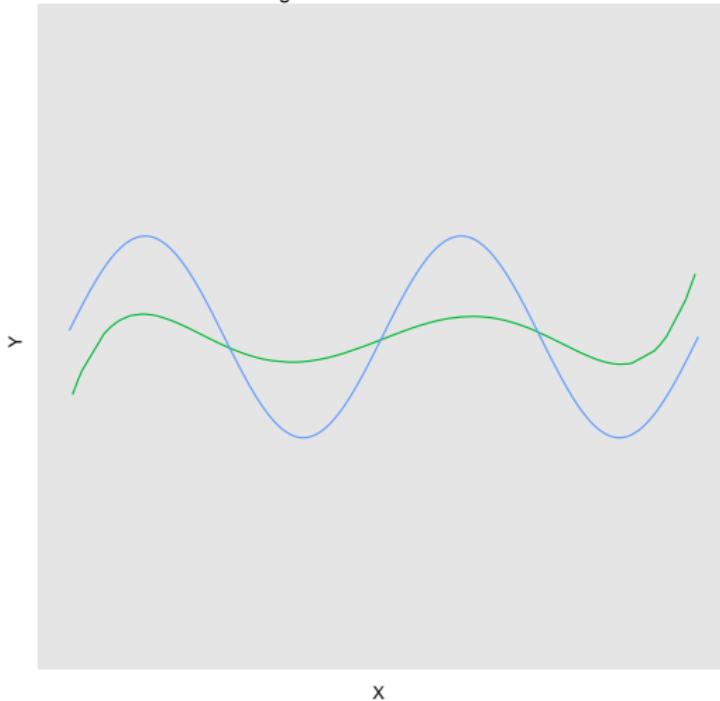
Lasso Regression with Lambda = 0.396



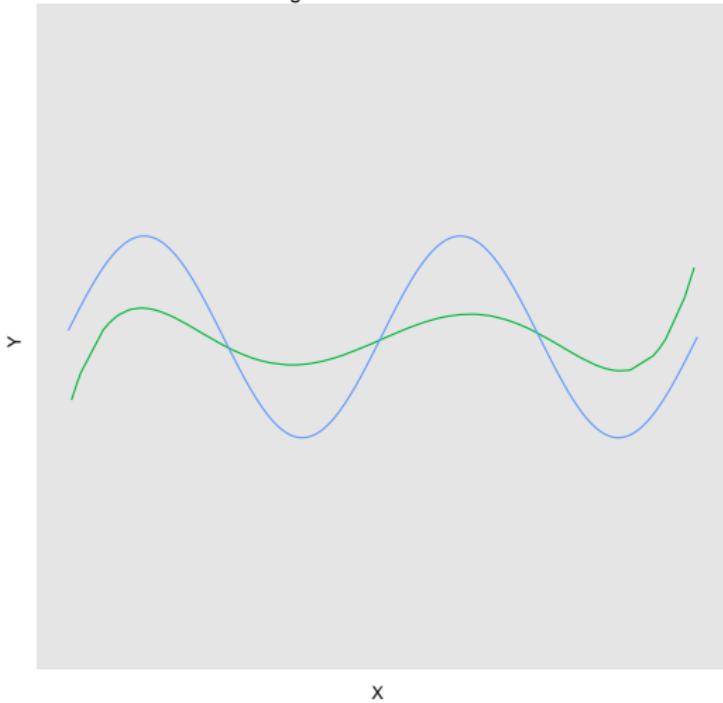
Lasso Regression with Lambda = 0.361



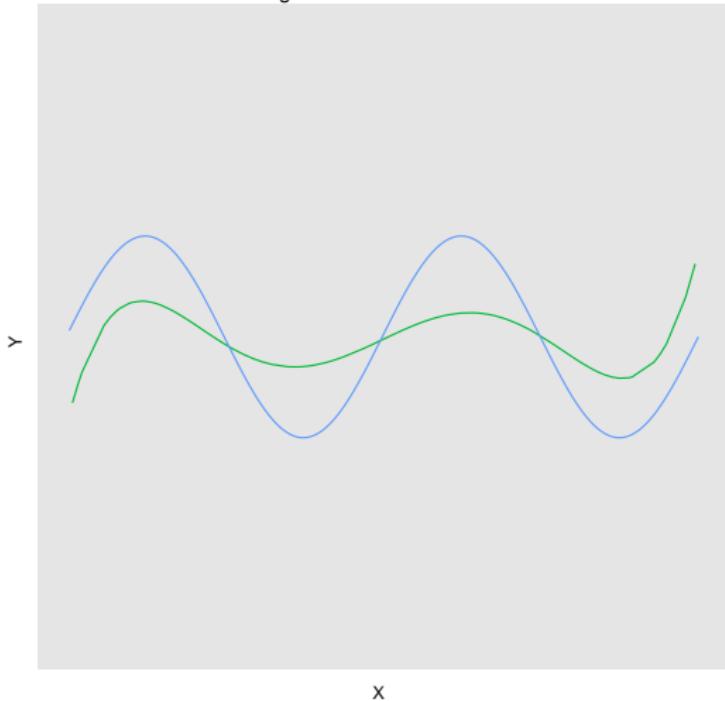
Lasso Regression with Lambda = 0.329



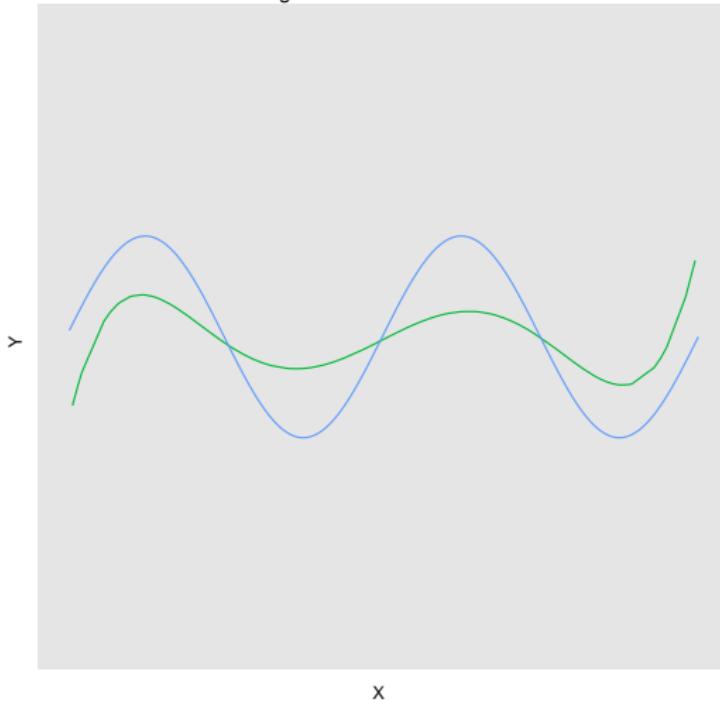
Lasso Regression with Lambda = 0.3



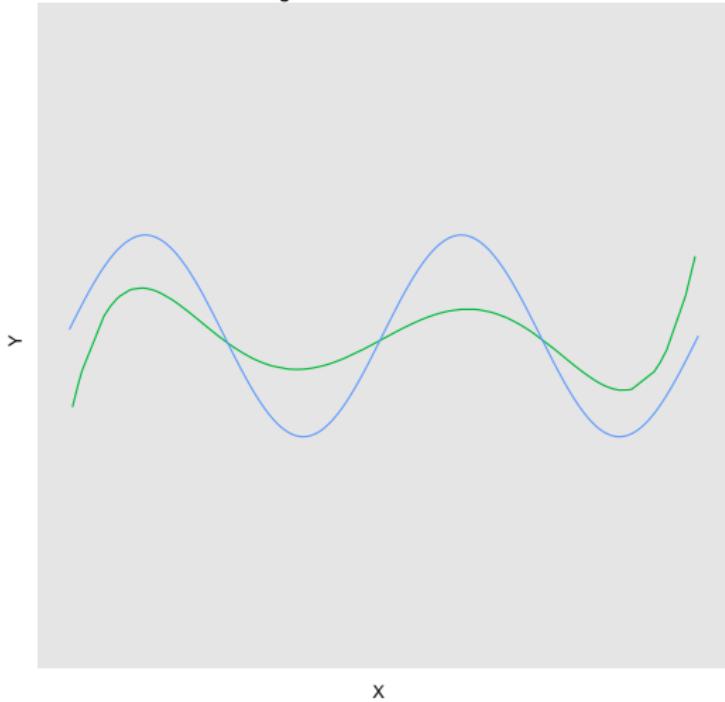
Lasso Regression with Lambda = 0.273



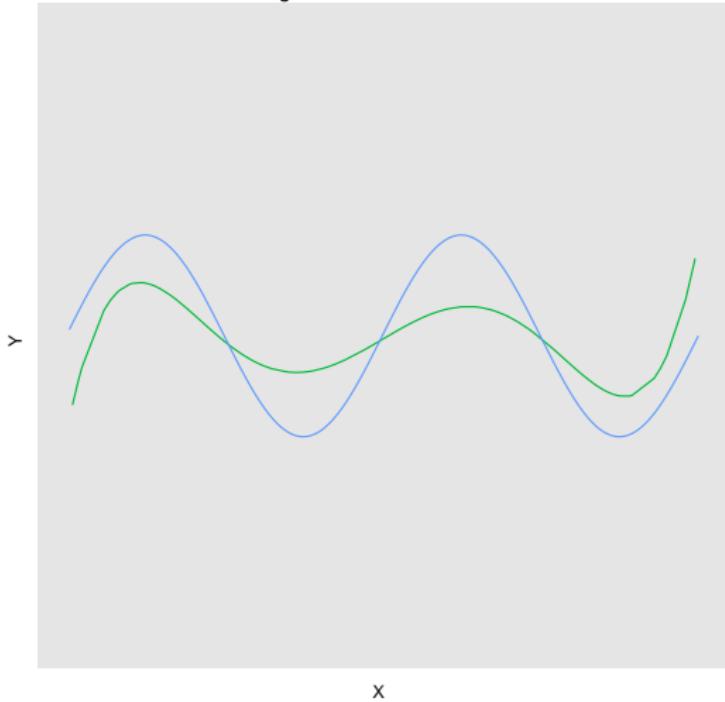
Lasso Regression with Lambda = 0.249



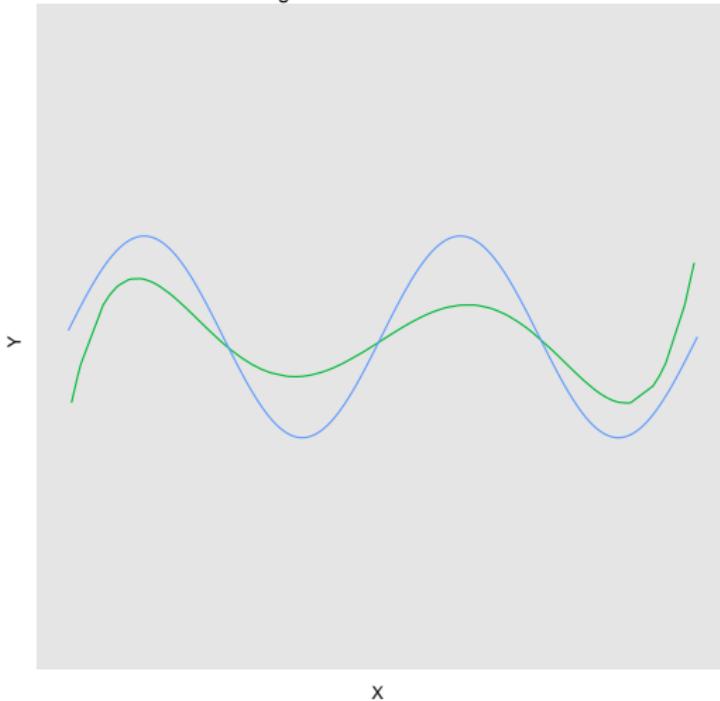
Lasso Regression with Lambda = 0.227



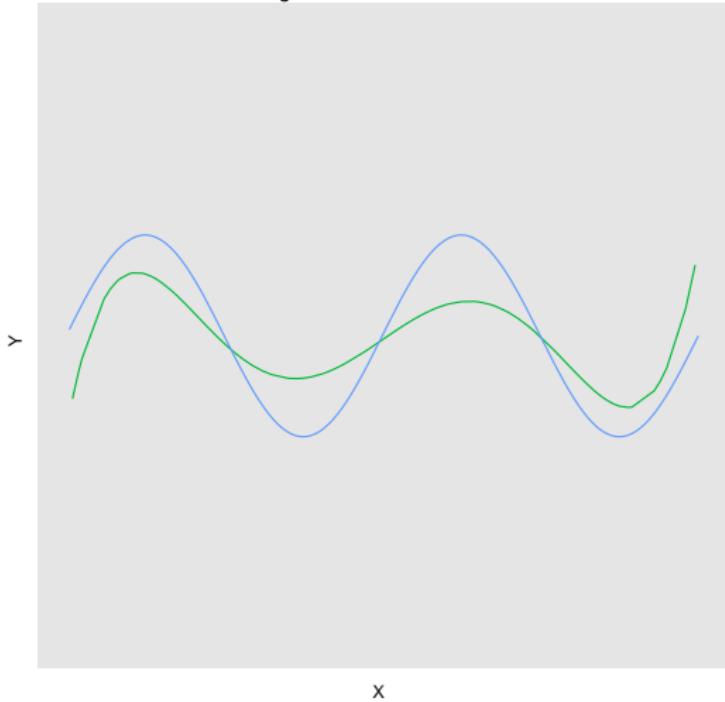
Lasso Regression with Lambda = 0.207



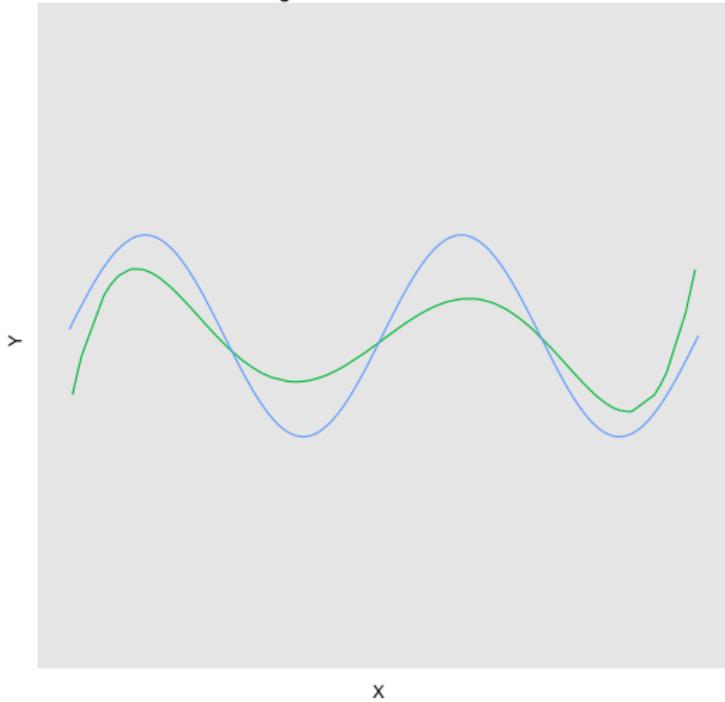
Lasso Regression with Lambda = 0.188



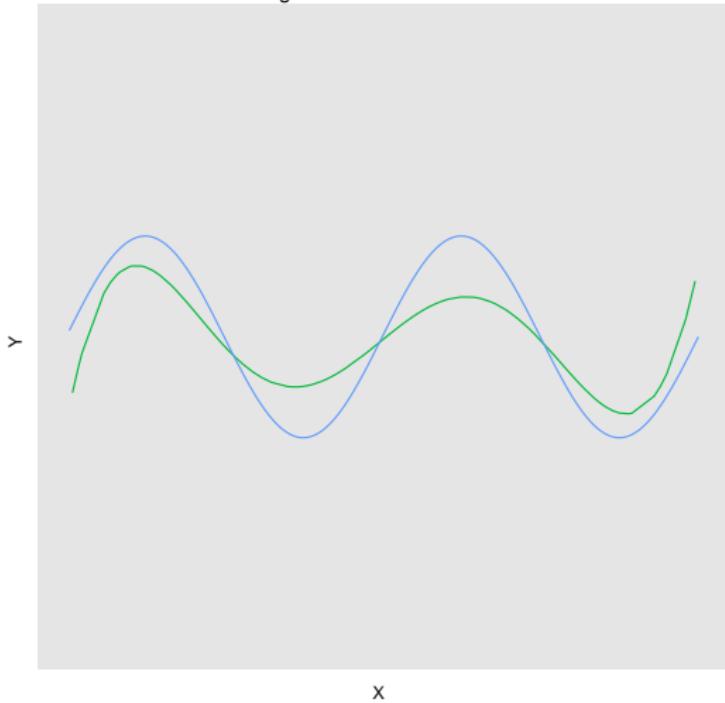
Lasso Regression with Lambda = 0.172



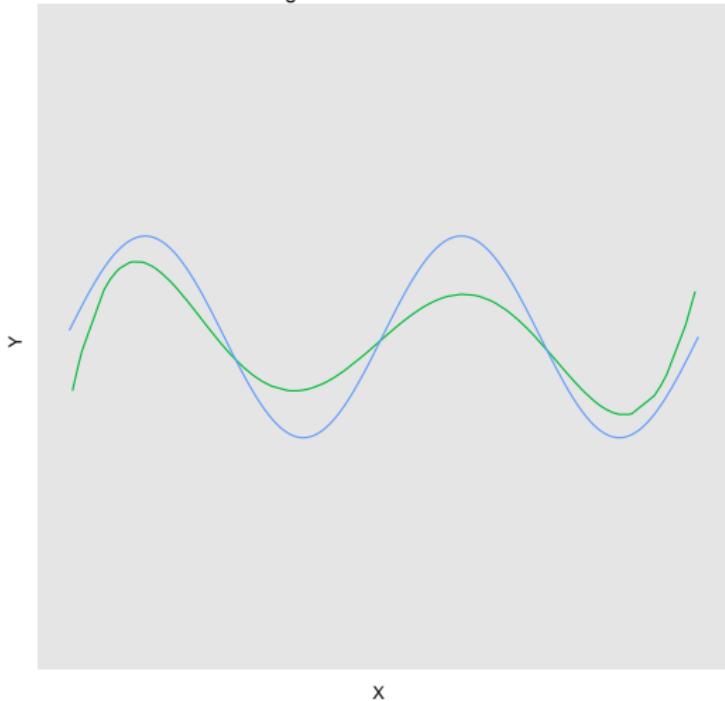
Lasso Regression with Lambda = 0.156



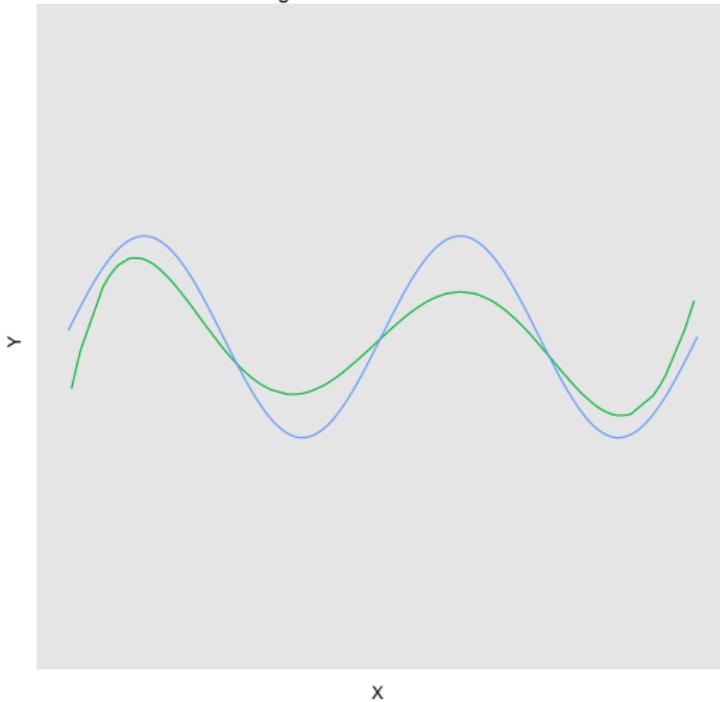
Lasso Regression with Lambda = 0.142



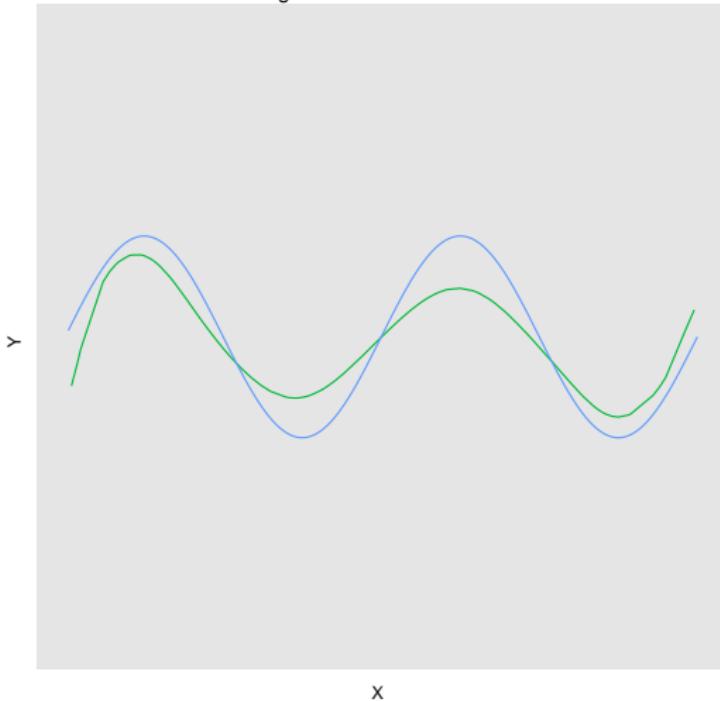
Lasso Regression with Lambda = 0.13



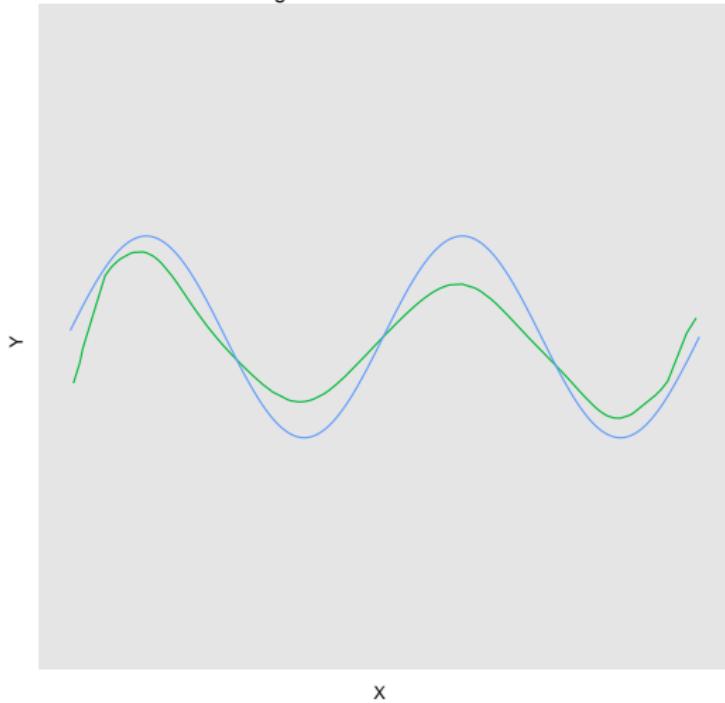
Lasso Regression with Lambda = 0.118



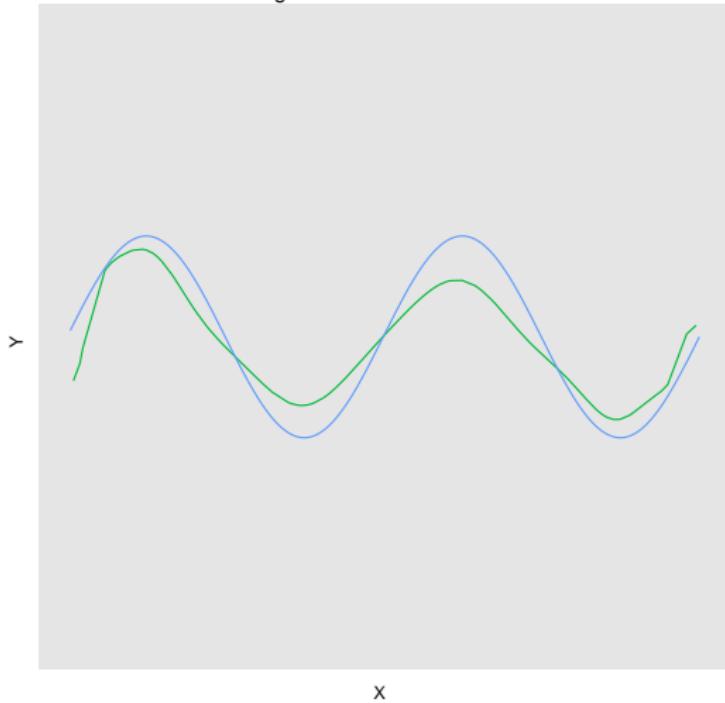
Lasso Regression with Lambda = 0.108



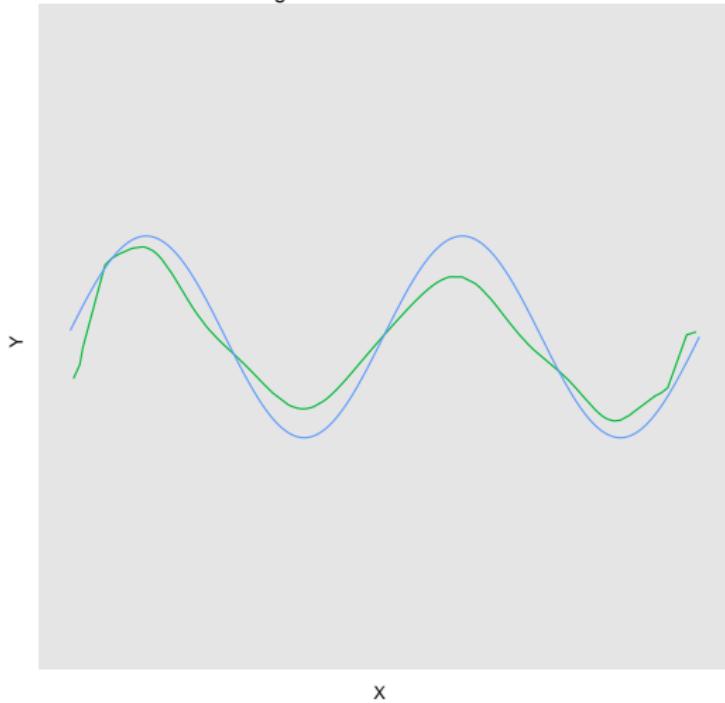
Lasso Regression with Lambda = 0.0982



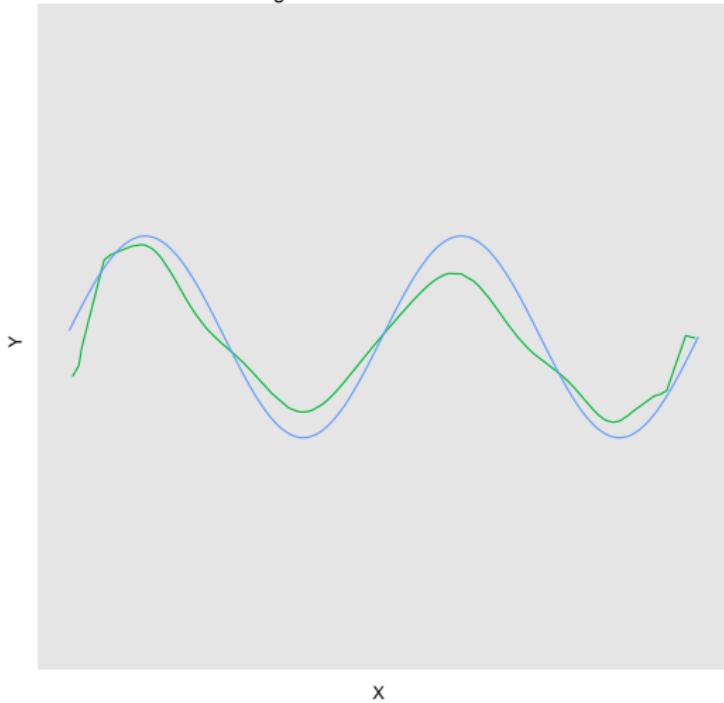
Lasso Regression with Lambda = 0.0895



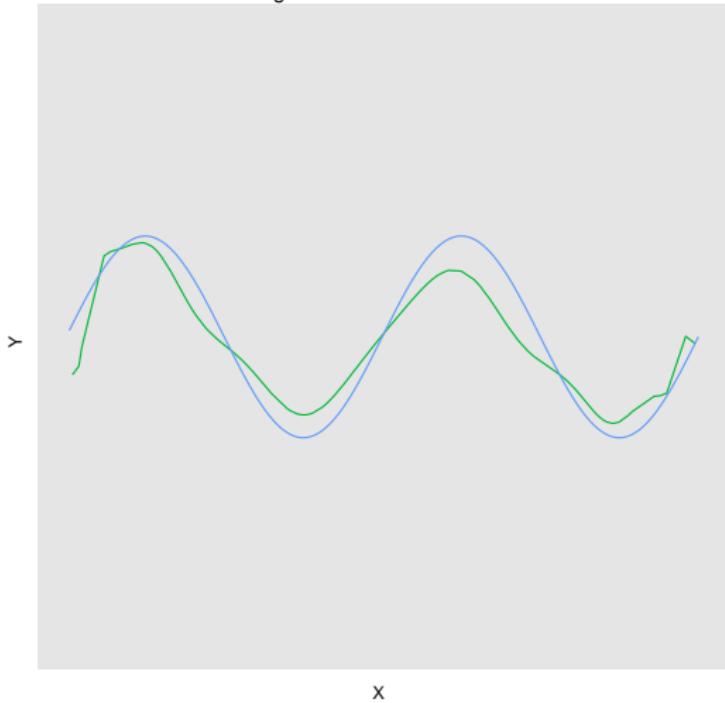
Lasso Regression with Lambda = 0.0815



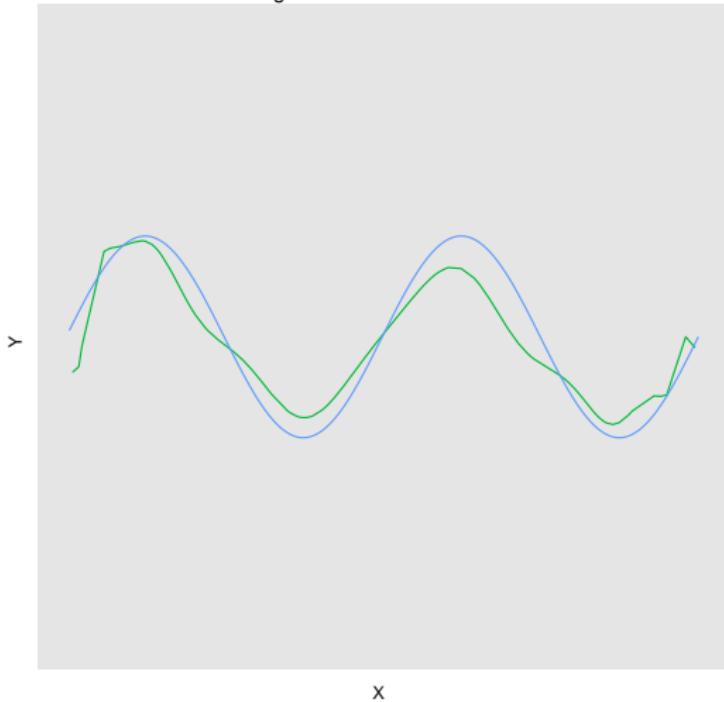
Lasso Regression with Lambda = 0.0743



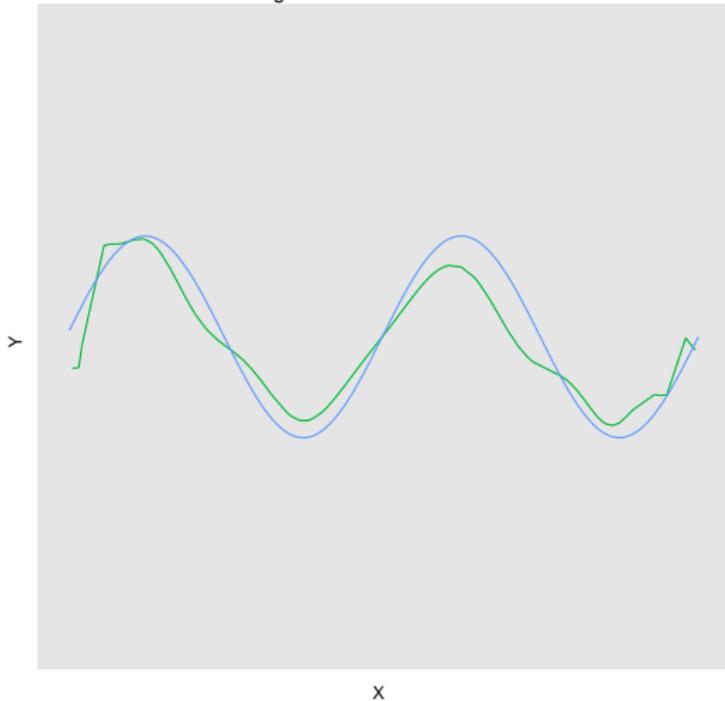
Lasso Regression with Lambda = 0.0677



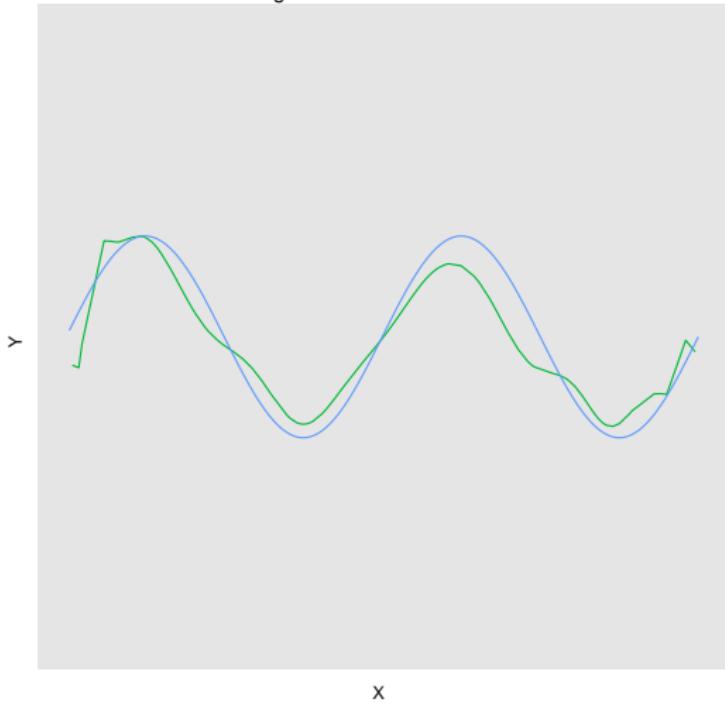
Lasso Regression with Lambda = 0.0617



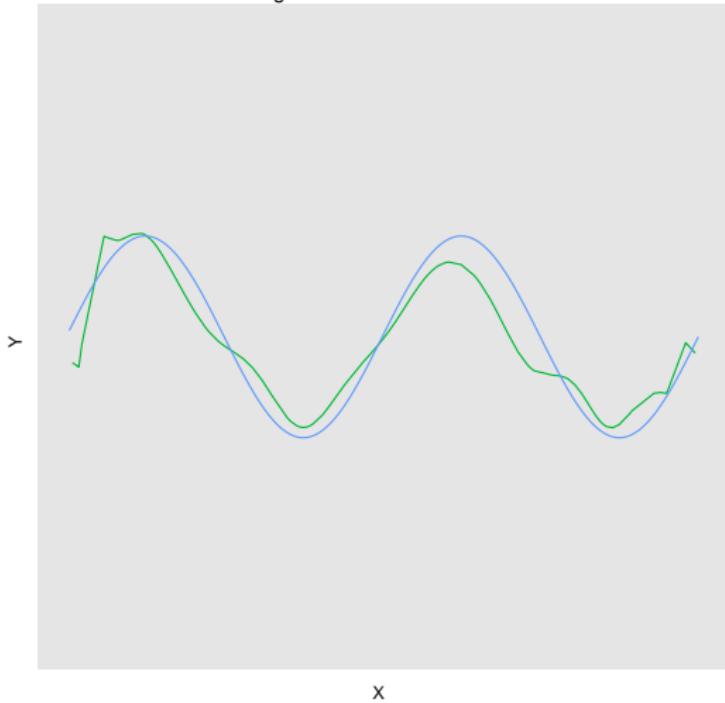
Lasso Regression with Lambda = 0.0562



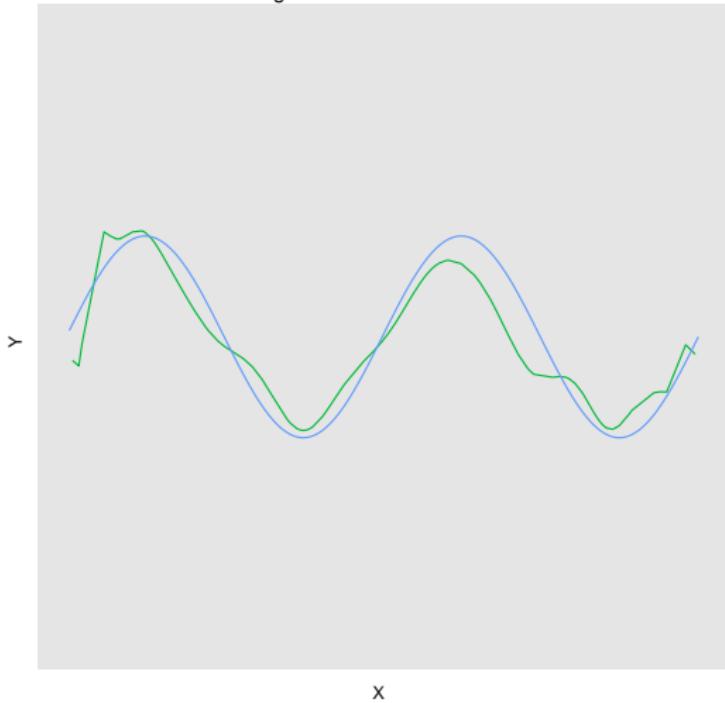
Lasso Regression with Lambda = 0.0512



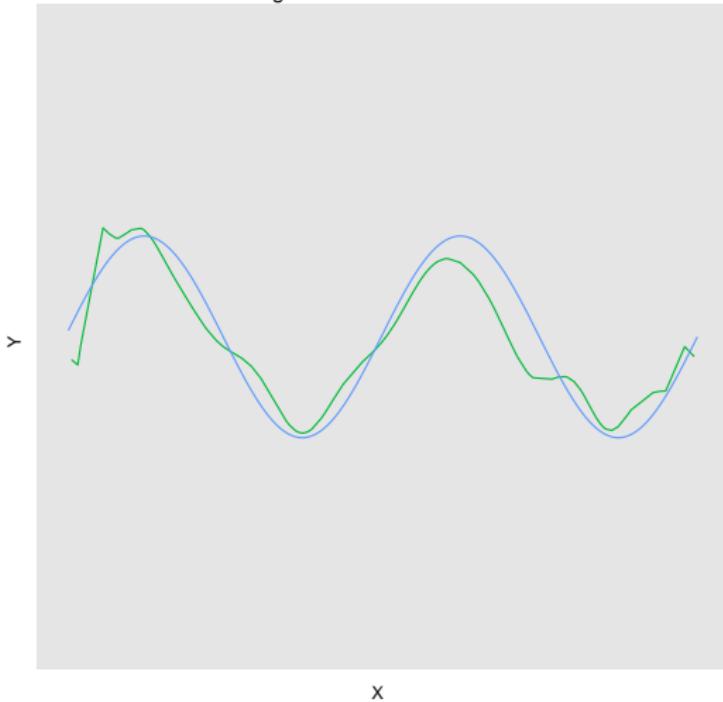
Lasso Regression with Lambda = 0.0467



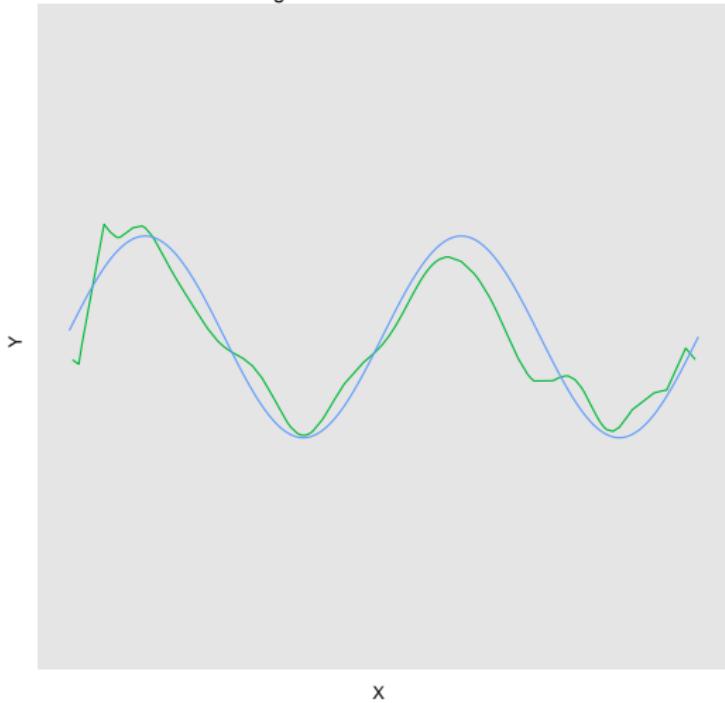
Lasso Regression with Lambda = 0.0425



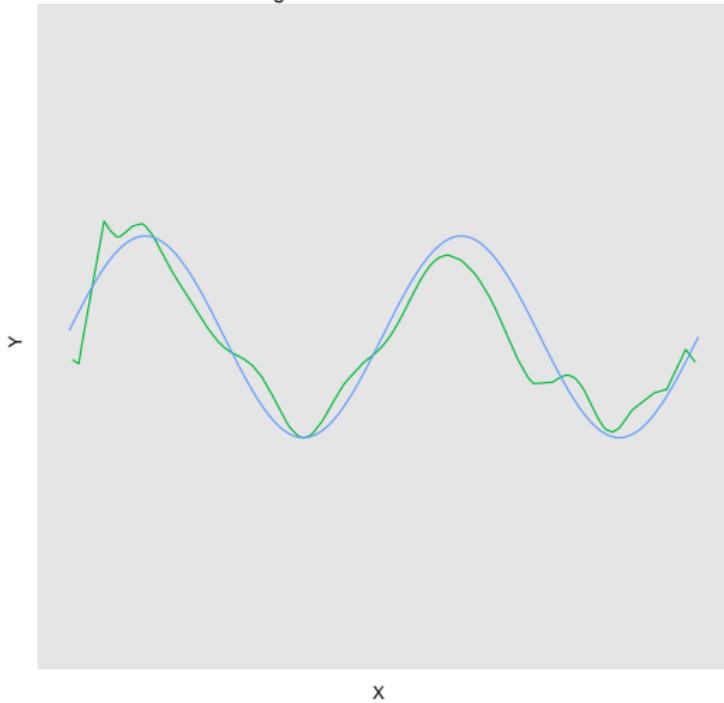
Lasso Regression with Lambda = 0.0387



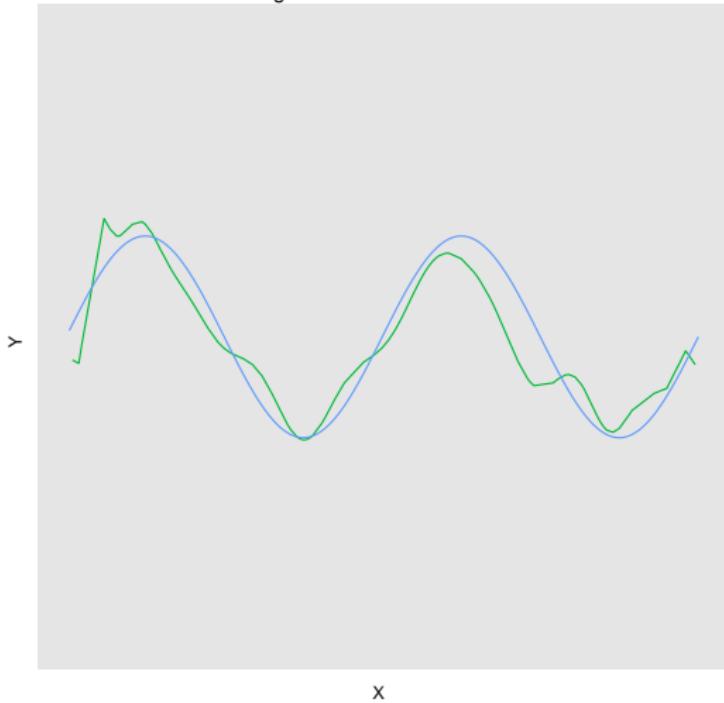
Lasso Regression with Lambda = 0.0353



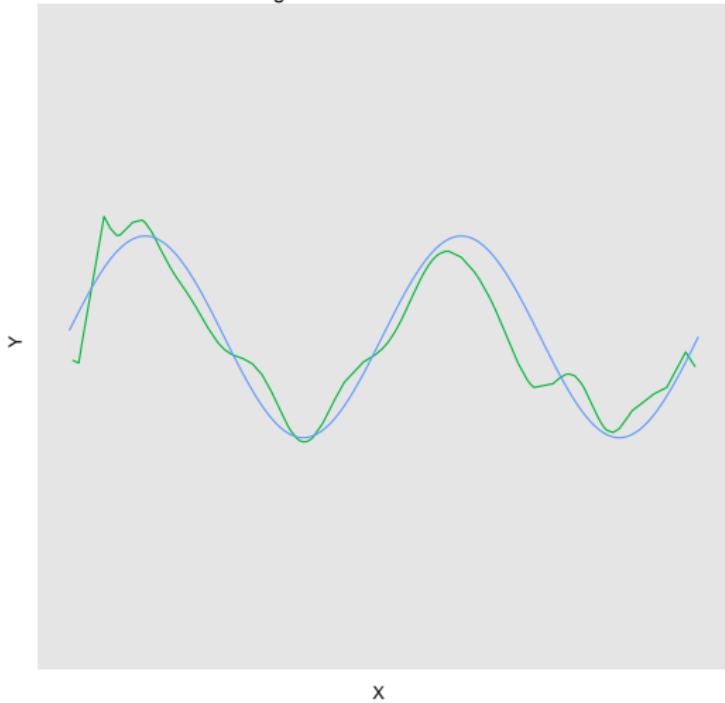
Lasso Regression with Lambda = 0.0322



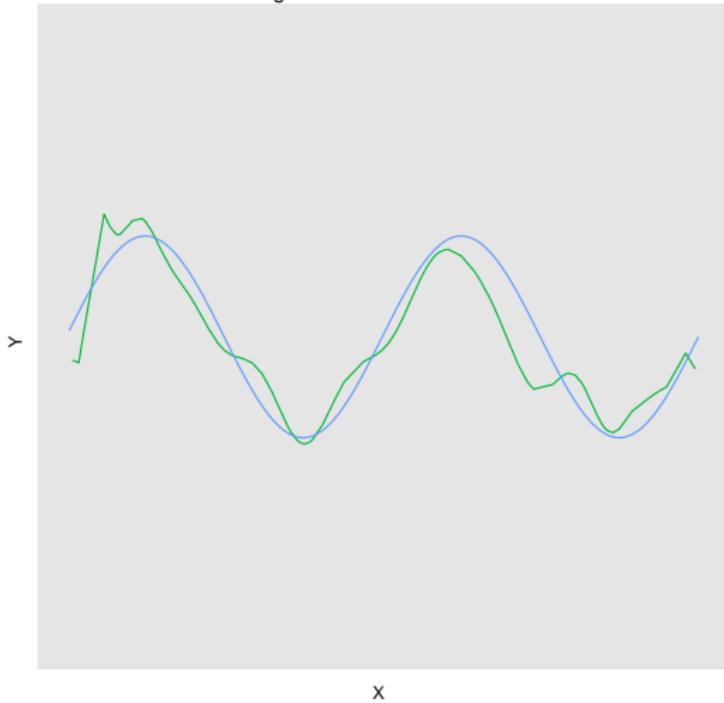
Lasso Regression with Lambda = 0.0293



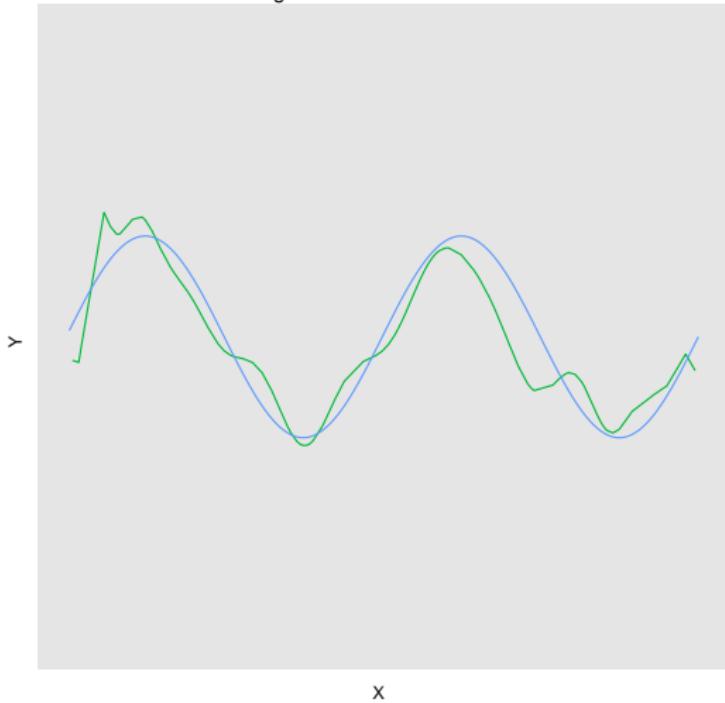
Lasso Regression with Lambda = 0.0267



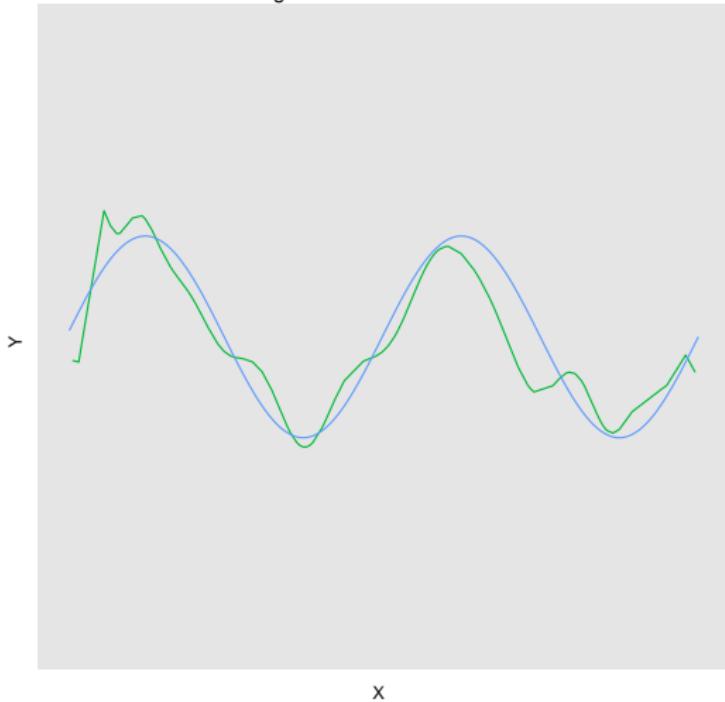
Lasso Regression with Lambda = 0.0243



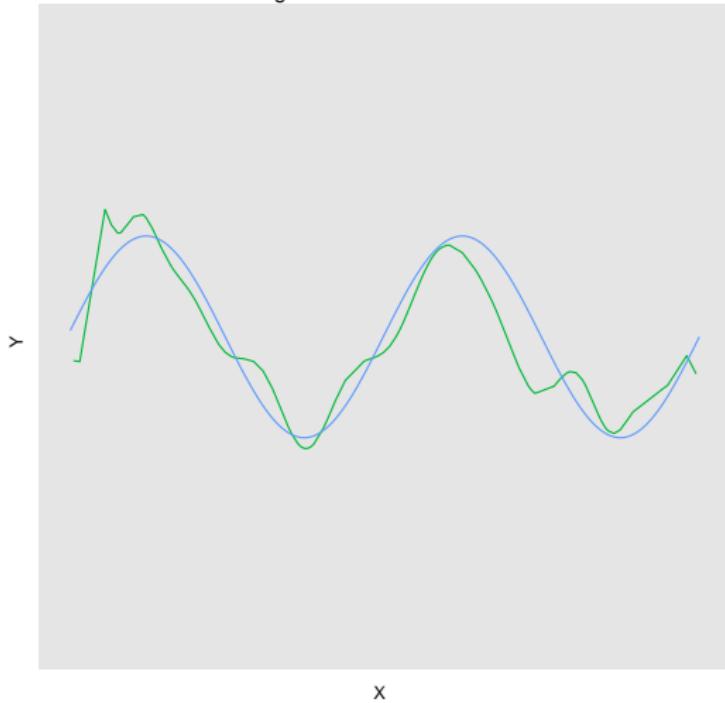
Lasso Regression with Lambda = 0.0222



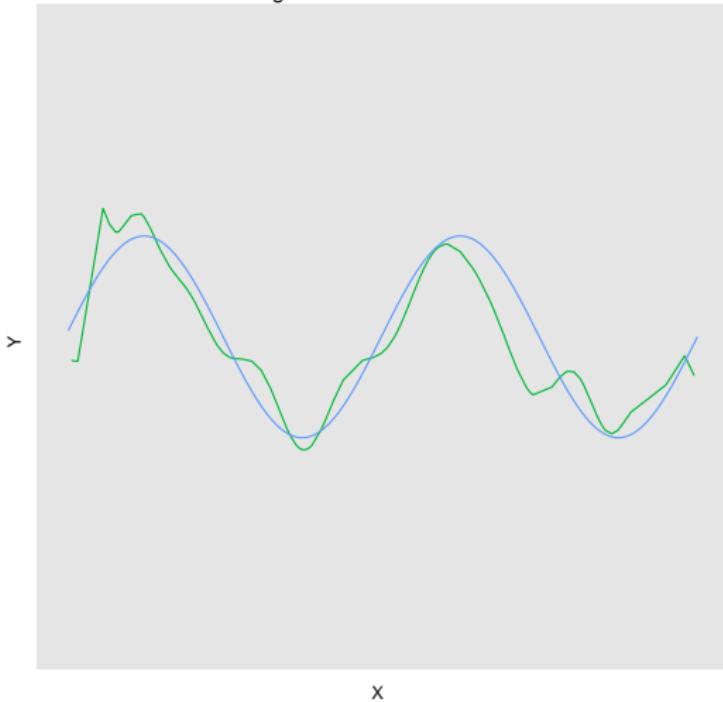
Lasso Regression with Lambda = 0.0202



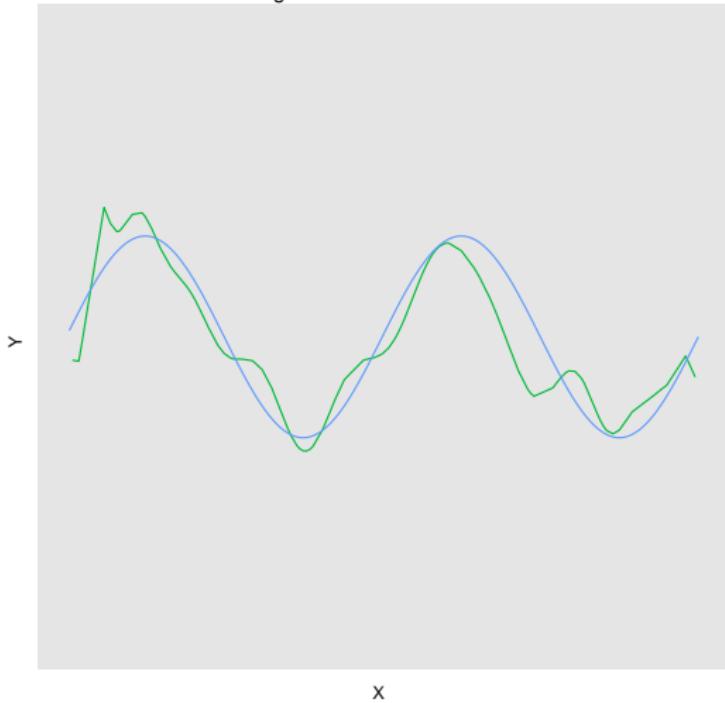
Lasso Regression with Lambda = 0.0184



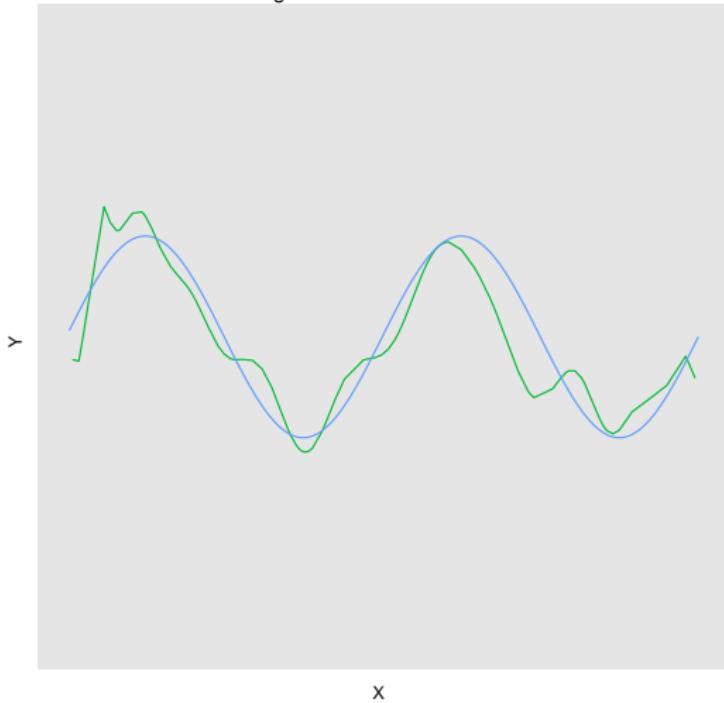
Lasso Regression with Lambda = 0.0168



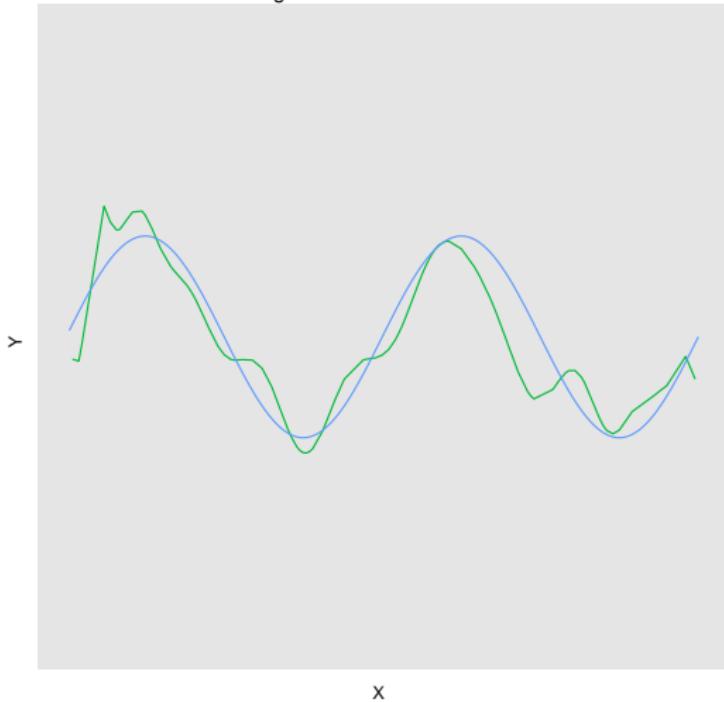
Lasso Regression with Lambda = 0.0153



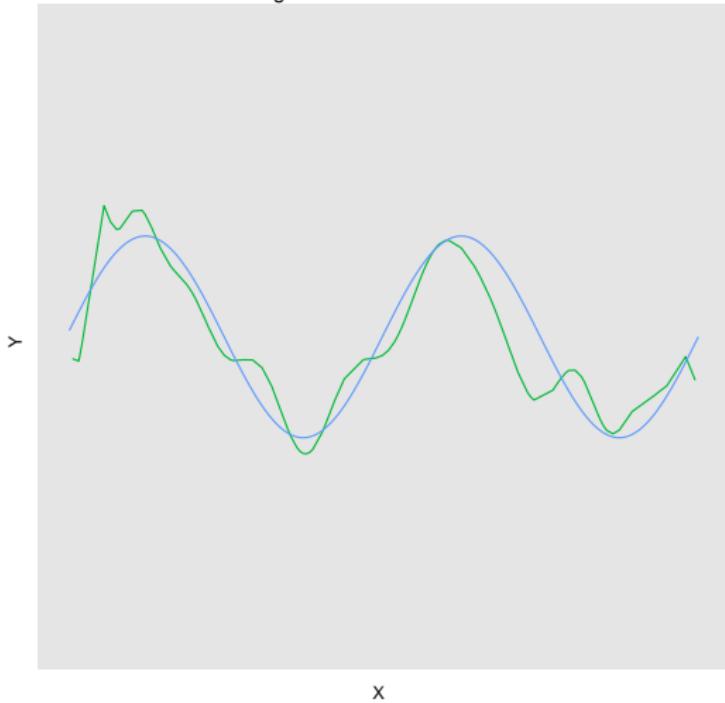
Lasso Regression with Lambda = 0.0139



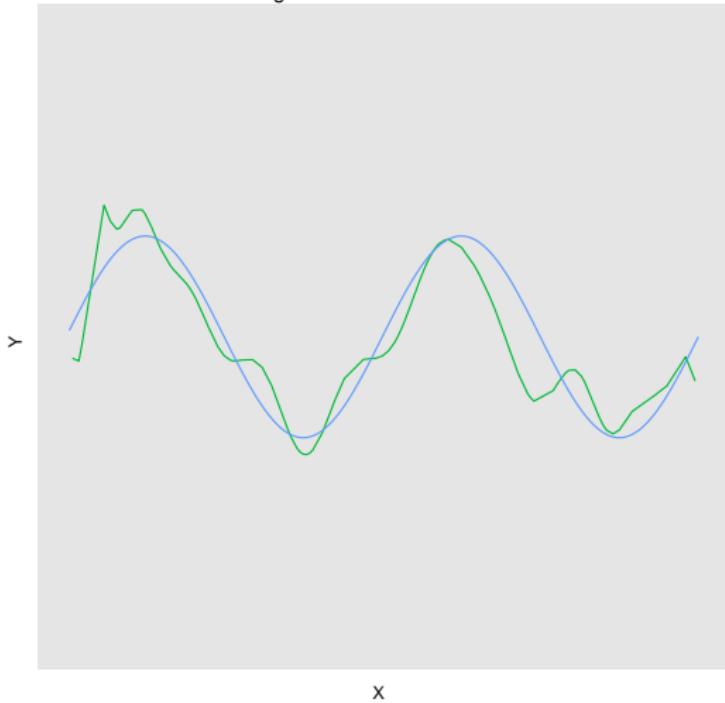
Lasso Regression with Lambda = 0.0127



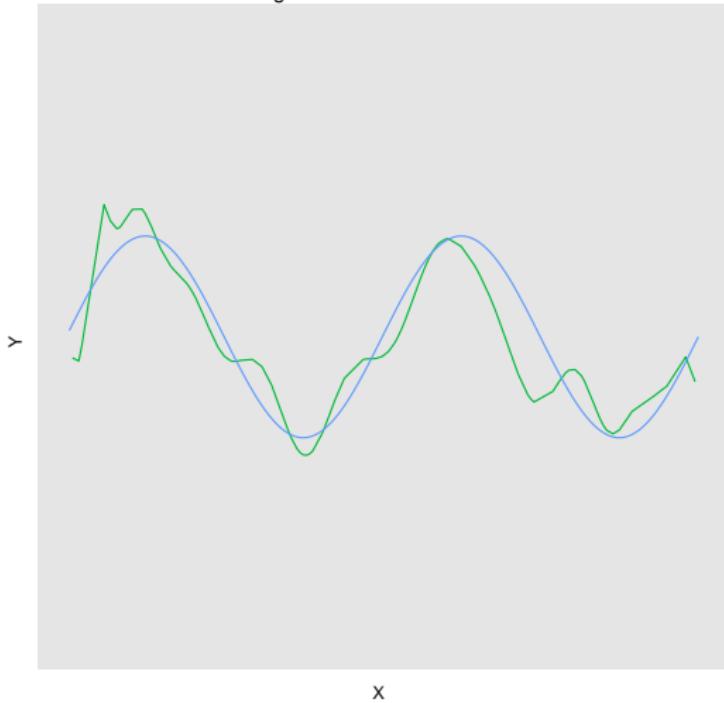
Lasso Regression with Lambda = 0.0116



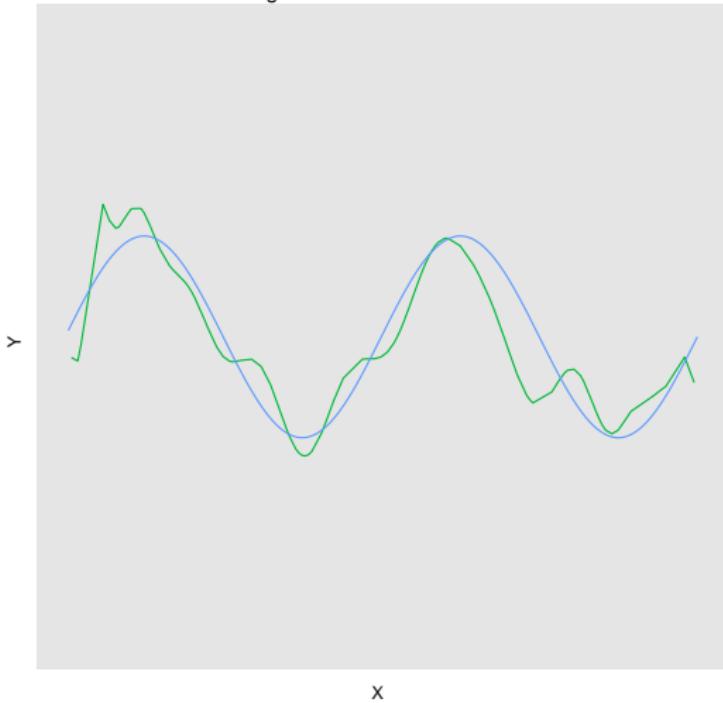
Lasso Regression with Lambda = 0.0105



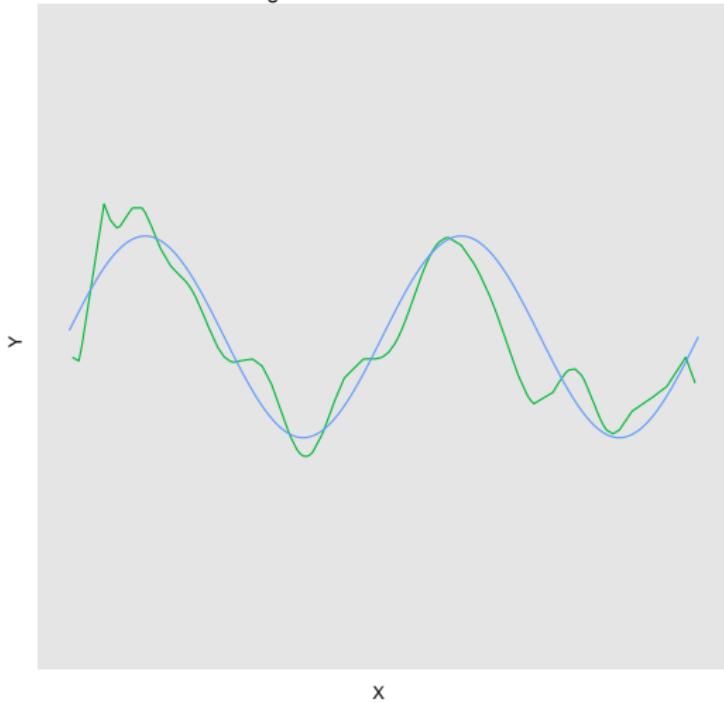
Lasso Regression with Lambda = 0.0096



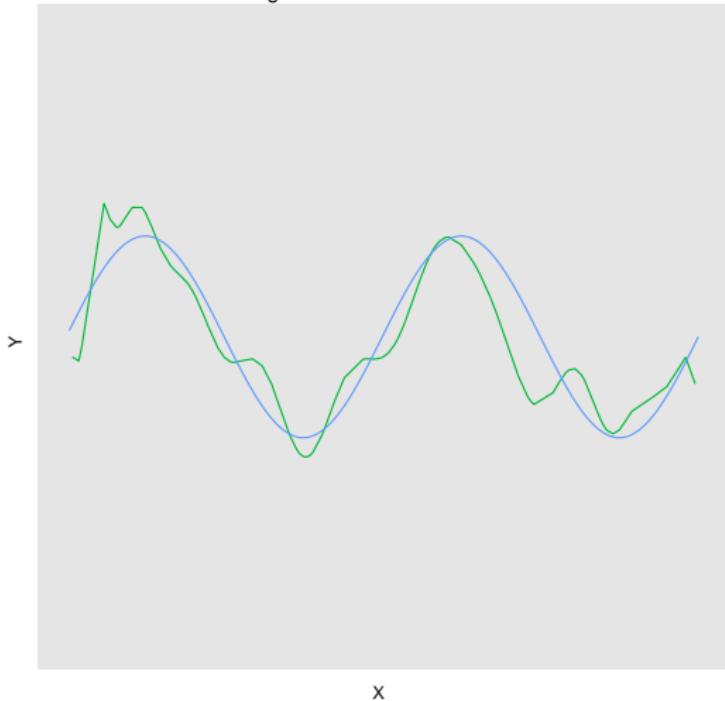
Lasso Regression with Lambda = 0.00874



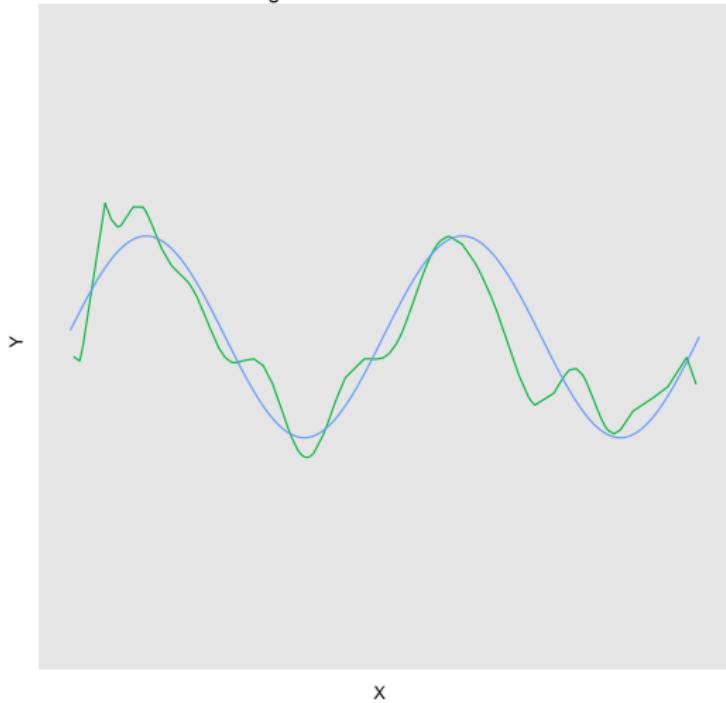
Lasso Regression with Lambda = 0.00797



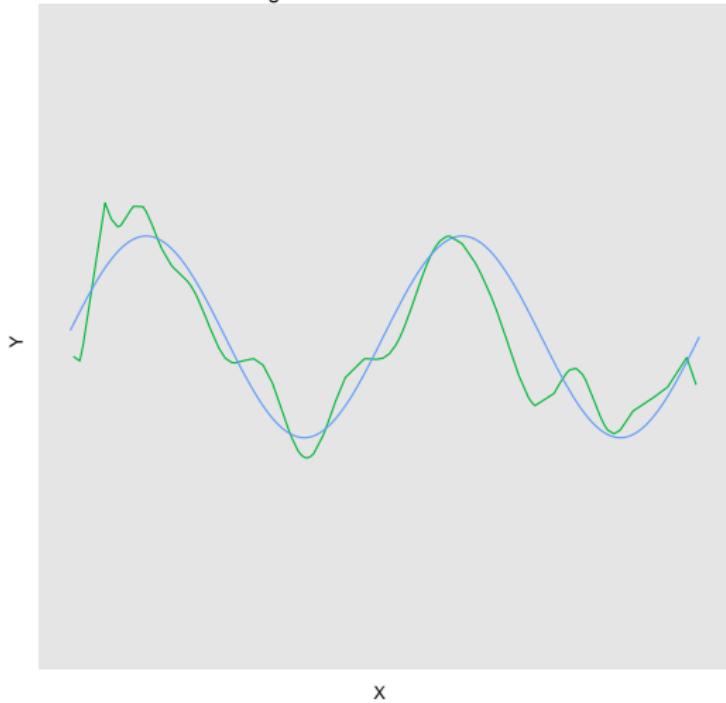
Lasso Regression with Lambda = 0.00726



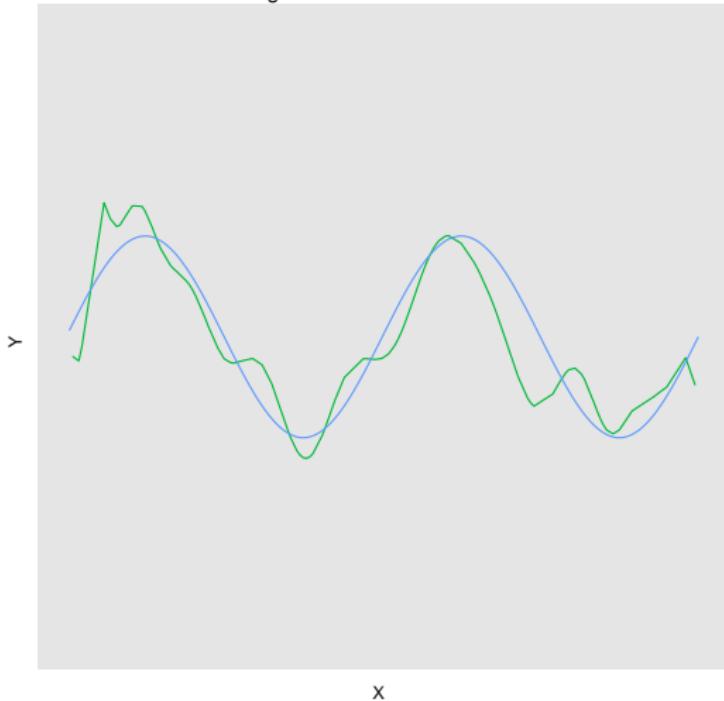
Lasso Regression with Lambda = 0.00661



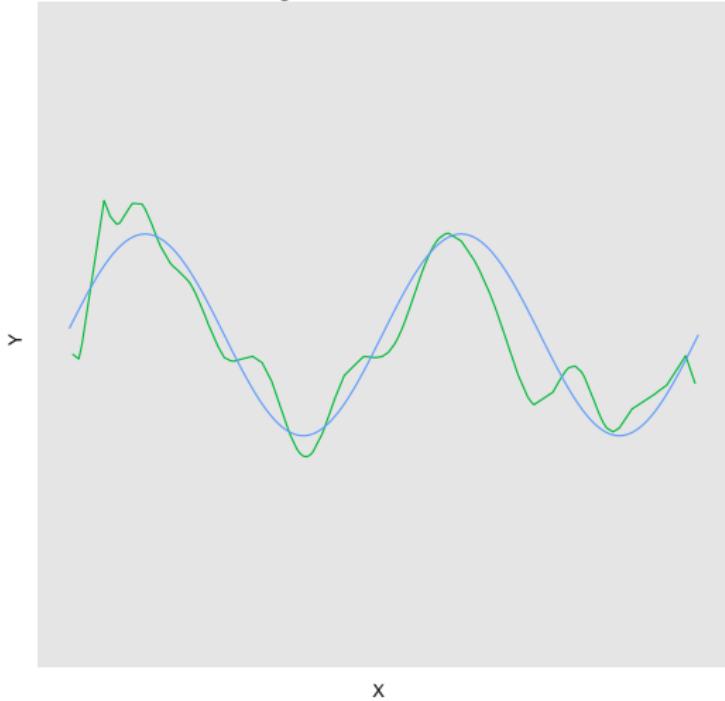
Lasso Regression with Lambda = 0.00603



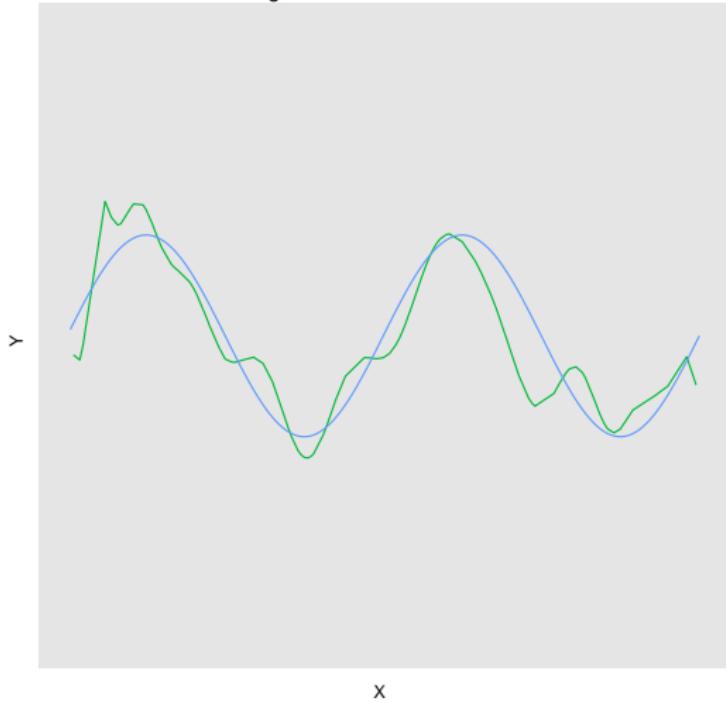
Lasso Regression with Lambda = 0.00549



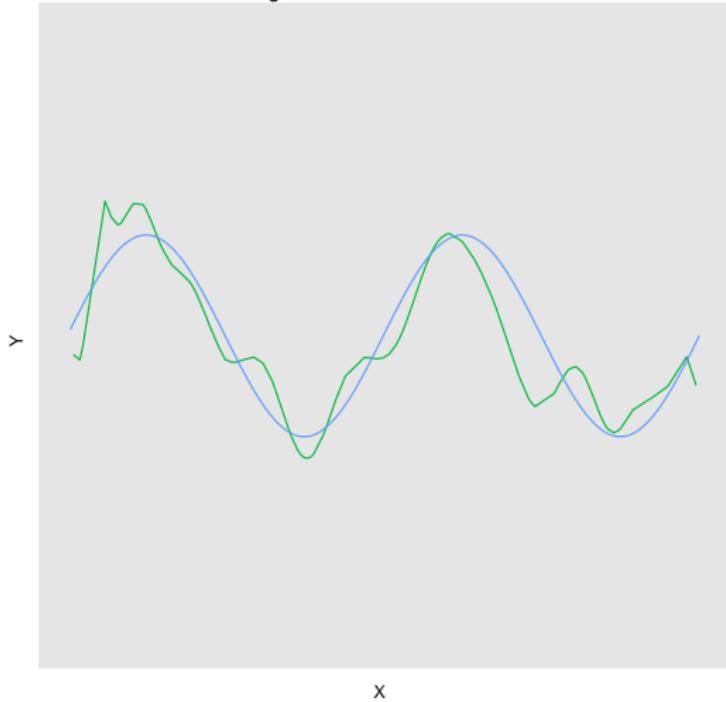
Lasso Regression with Lambda = 0.005



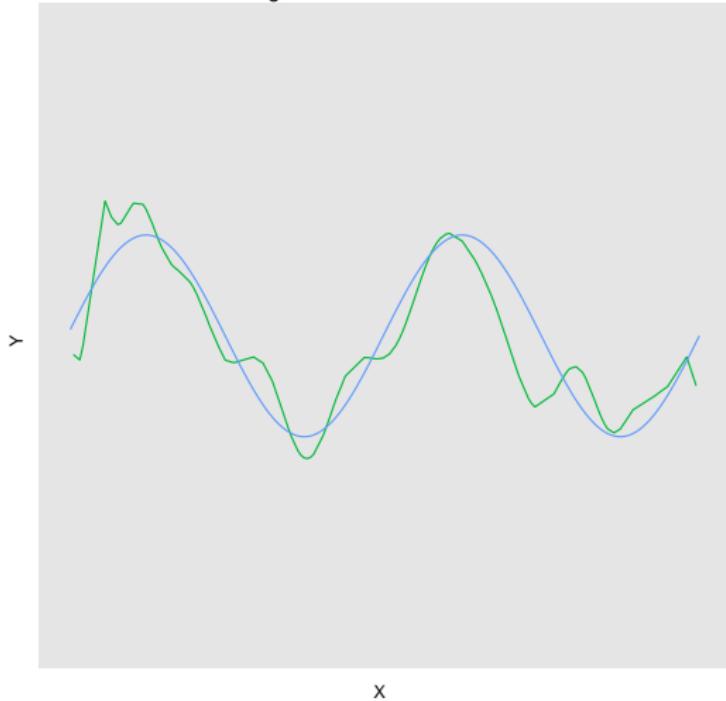
Lasso Regression with Lambda = 0.00456



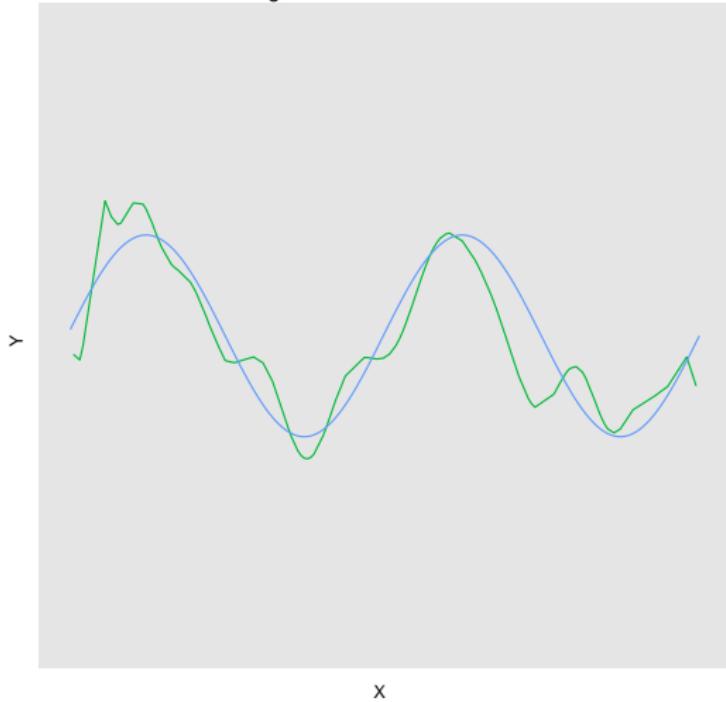
Lasso Regression with Lambda = 0.00415



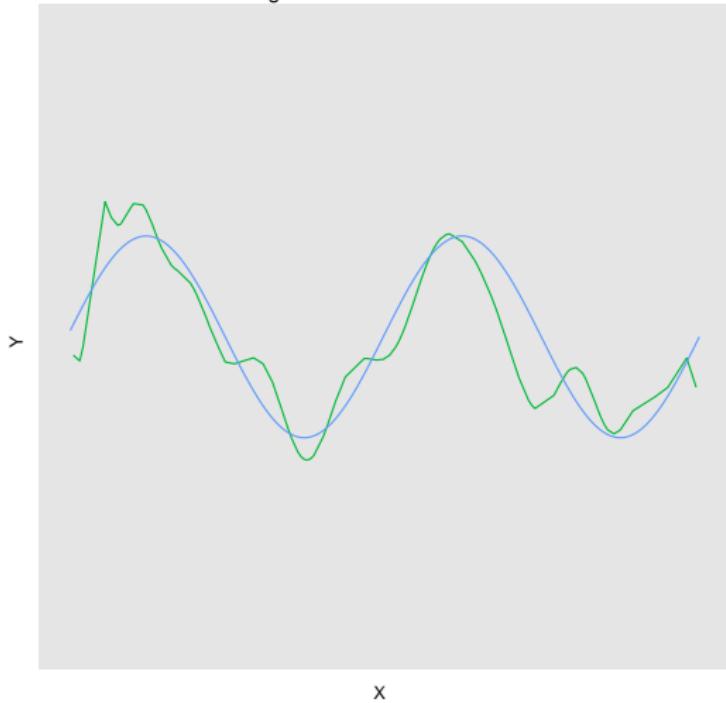
Lasso Regression with Lambda = 0.00378



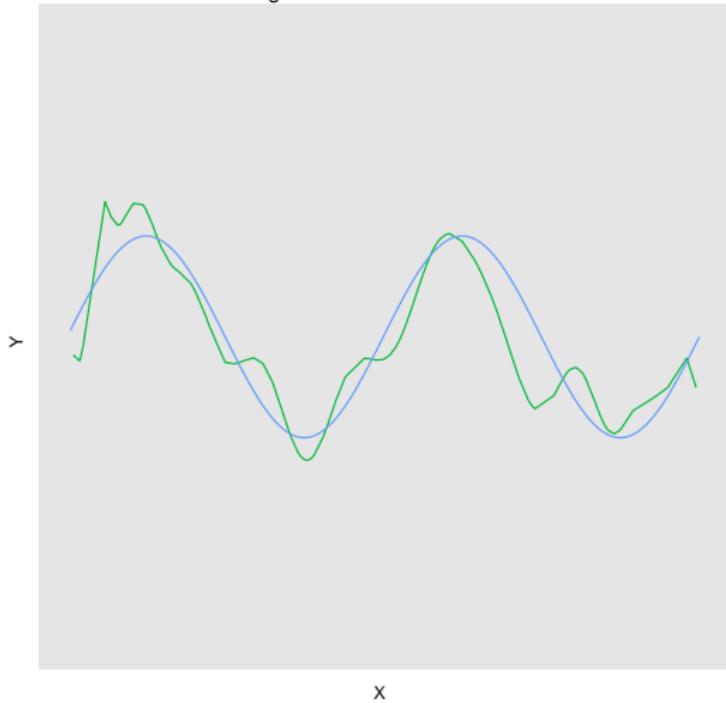
Lasso Regression with Lambda = 0.00345



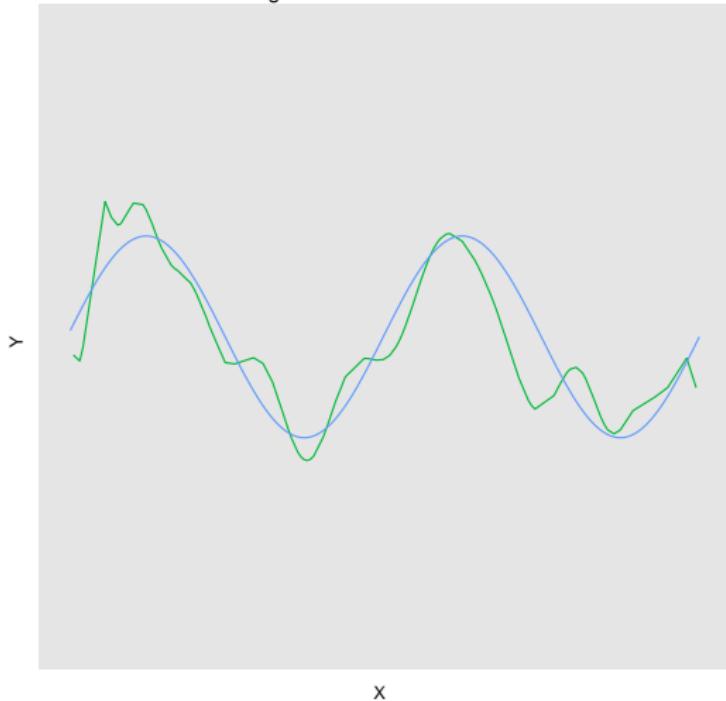
Lasso Regression with Lambda = 0.00314



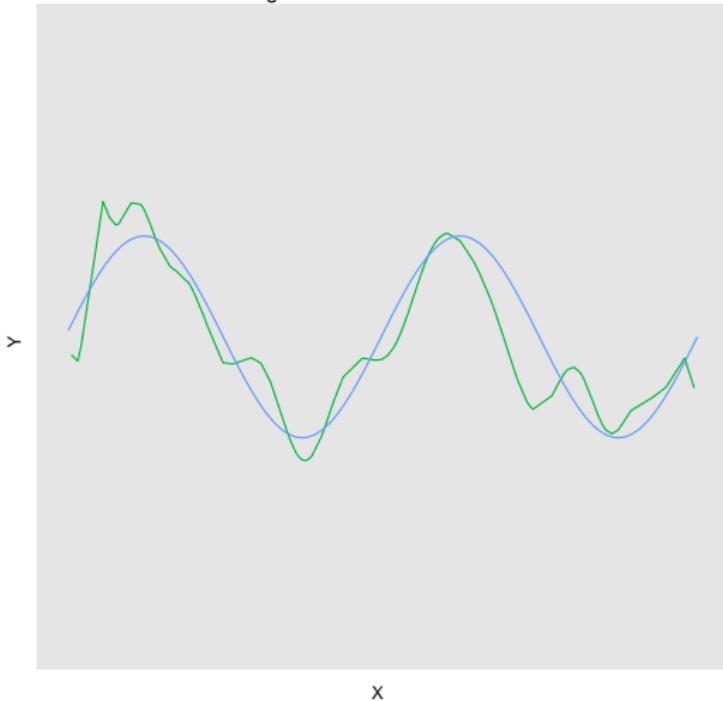
Lasso Regression with Lambda = 0.00286



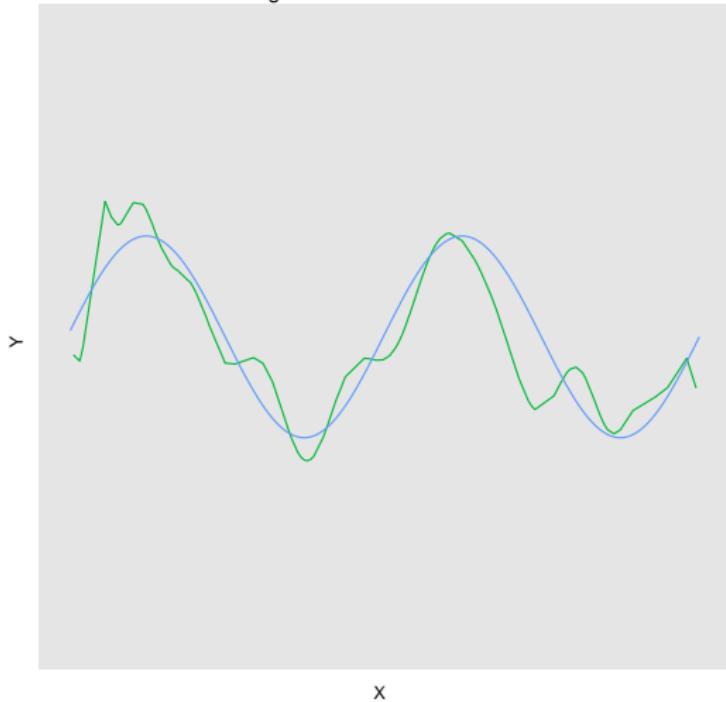
Lasso Regression with Lambda = 0.00261



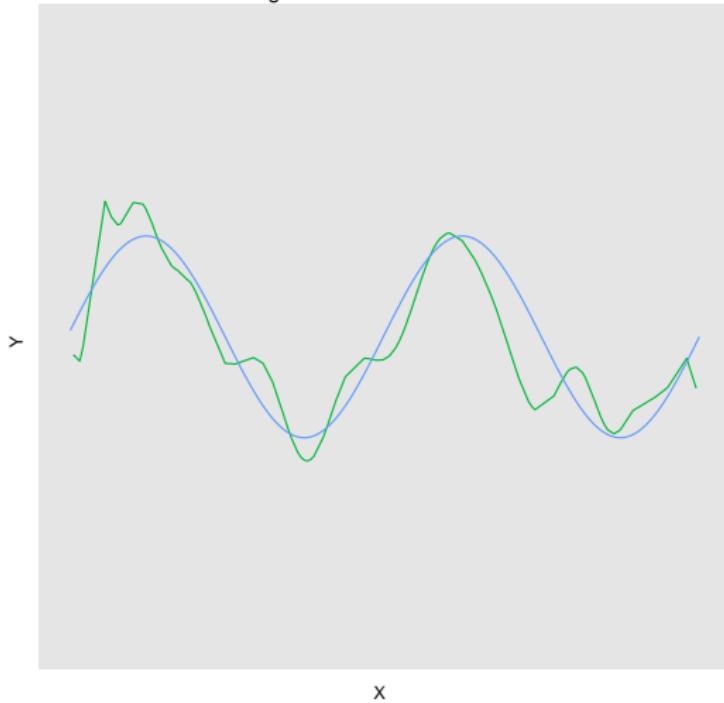
Lasso Regression with Lambda = 0.00238



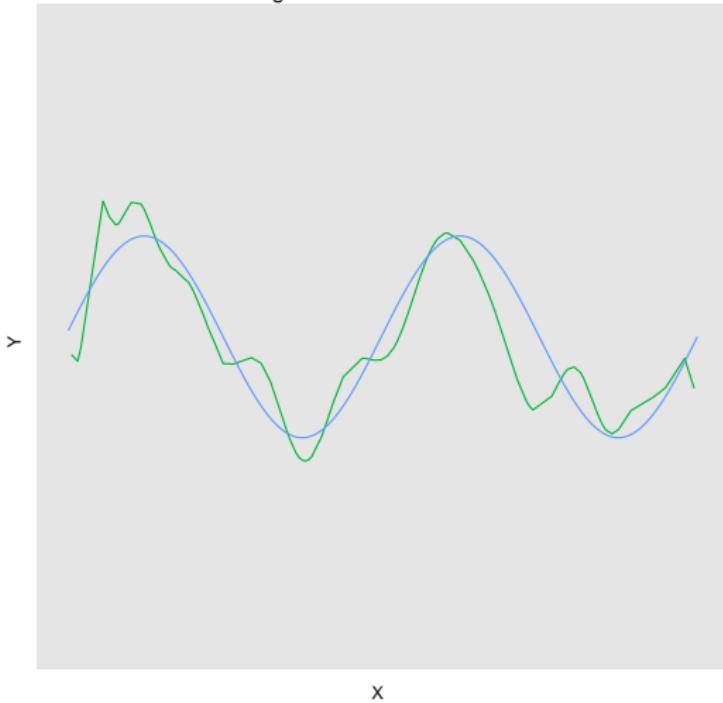
Lasso Regression with Lambda = 0.00217



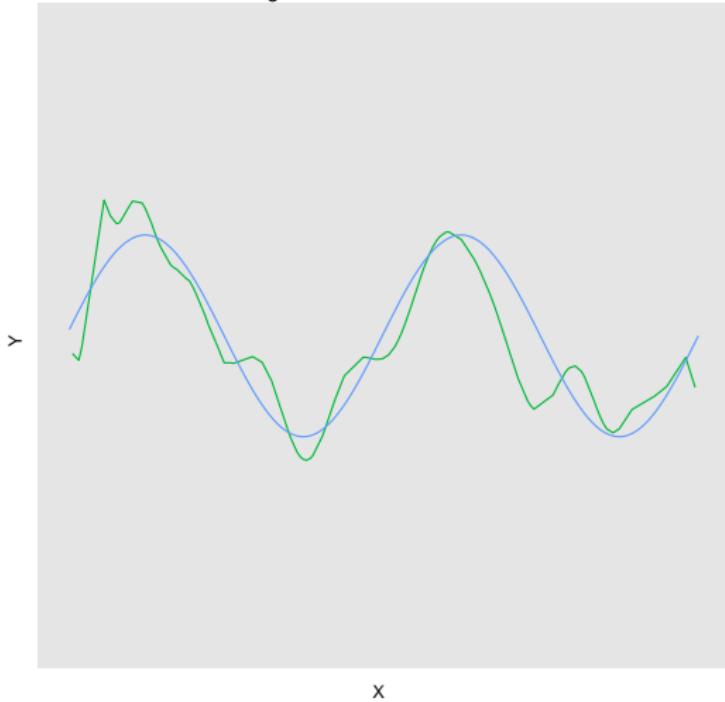
Lasso Regression with Lambda = 0.00197



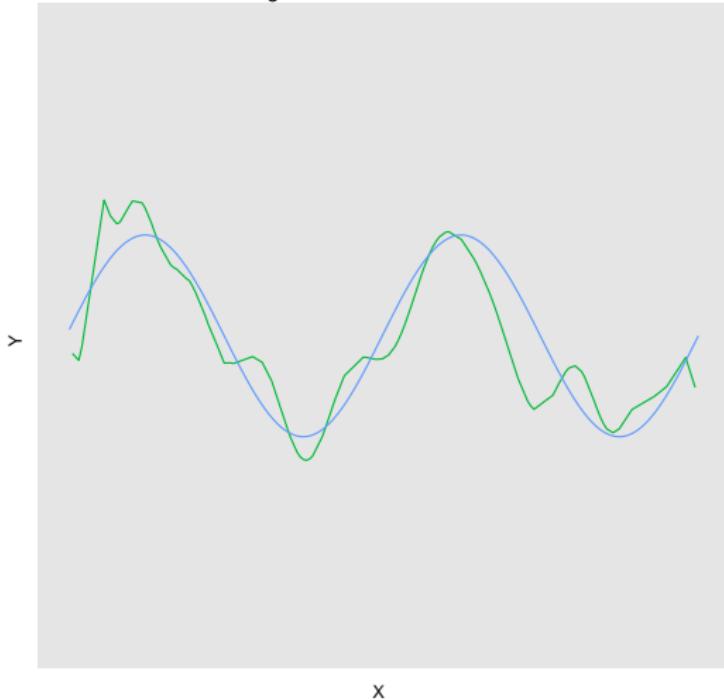
Lasso Regression with Lambda = 0.0018



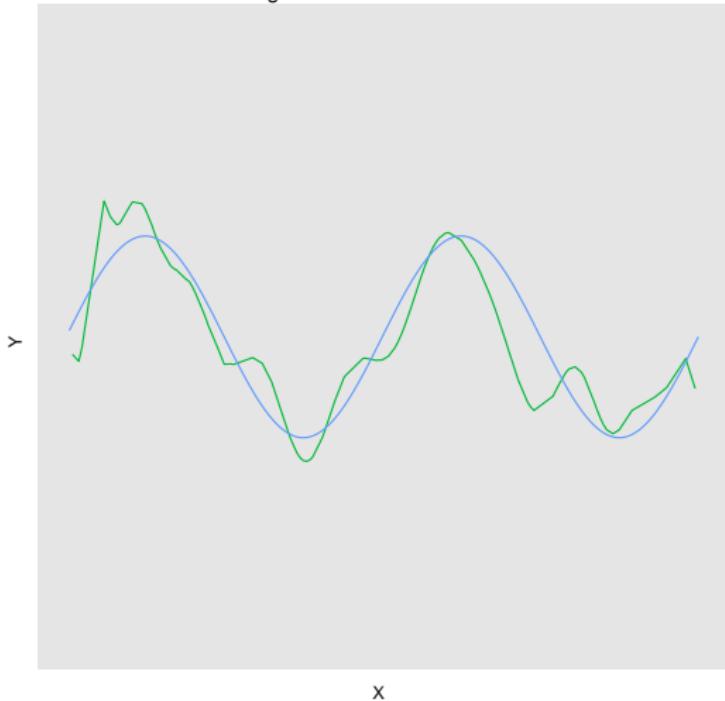
Lasso Regression with Lambda = 0.00164



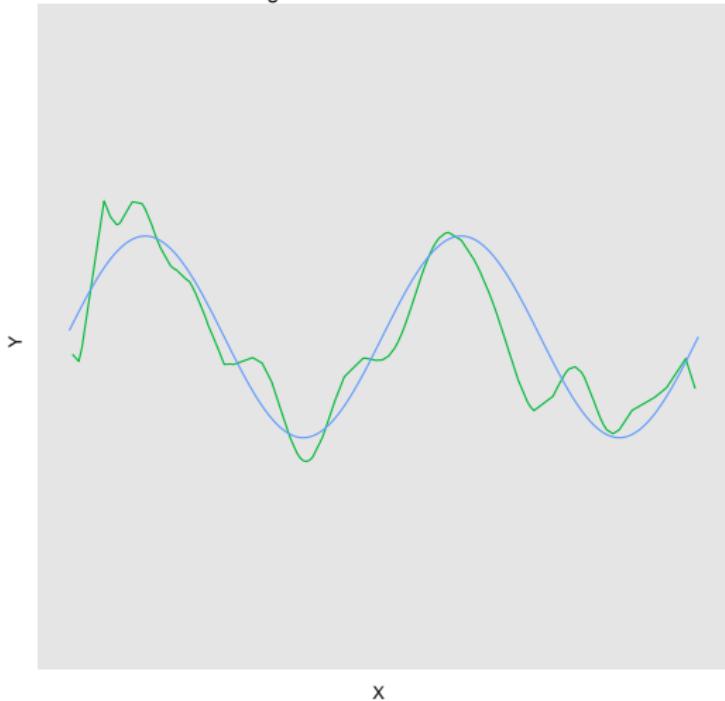
Lasso Regression with Lambda = 0.00149



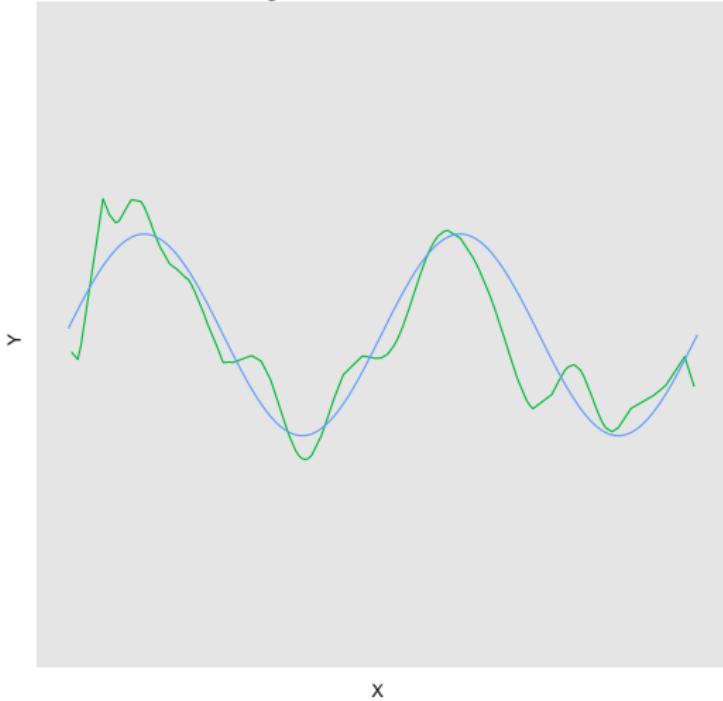
Lasso Regression with Lambda = 0.00136



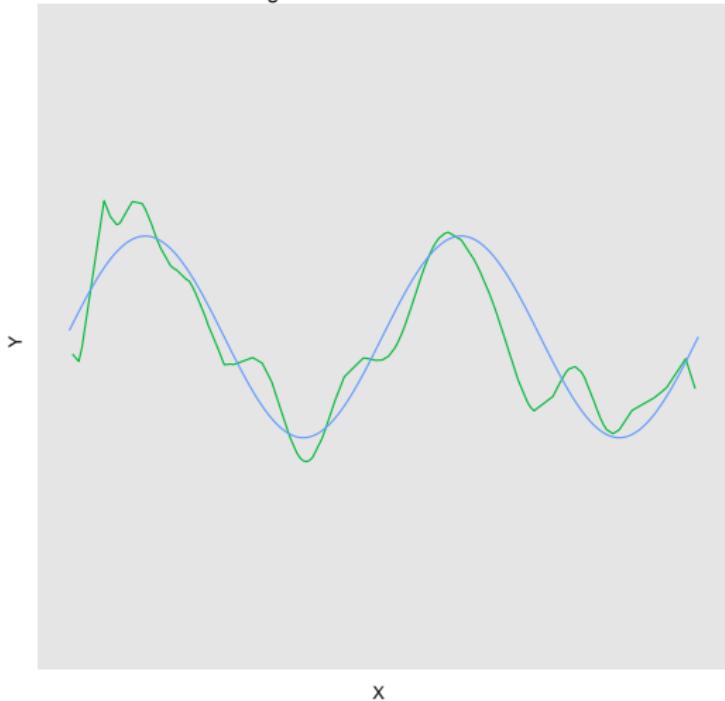
Lasso Regression with Lambda = 0.00124



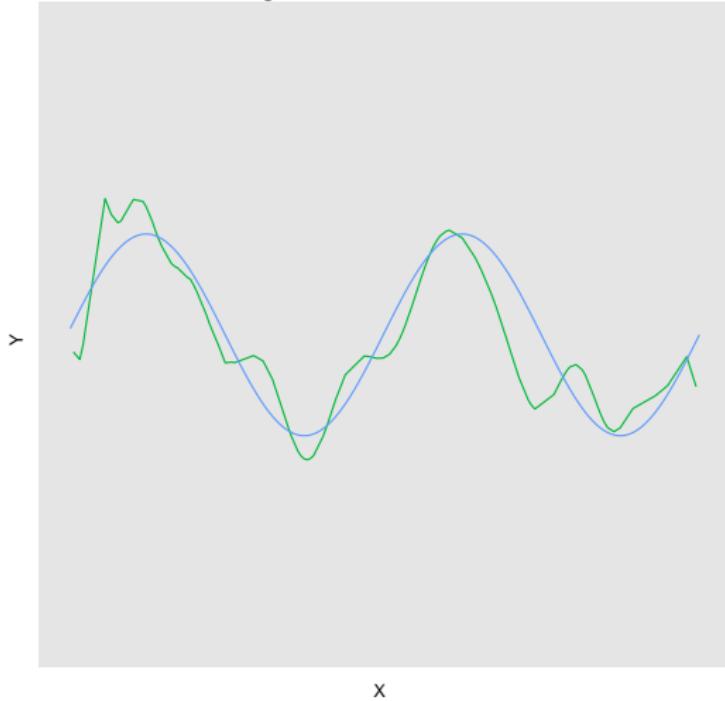
Lasso Regression with Lambda = 0.00113



Lasso Regression with Lambda = 0.00103



Lasso Regression with Lambda = 0.000937



How can we use regularization to solve real world problems?

The text regression problem:

- ▶ N documents
- ▶ D words
- ▶ Predict continuous value for each document from word counts

Examples:

- ▶ From IPO notices, predict stock volatility
- ▶ From press releases, predict Congress member's politics
- ▶ ...

We need regularization because we:

- ▶ Observe more words than documents
- ▶ Want sparse solutions, e.g. a few words that matter a lot

From press releases, how can we find words that signal politics?

*Hey, does anybody notice this crazy thing that we're on
the road to socialism? I'm just saying. Wow. We got —
we got the SCHIPs thing going for us. That's great.*

How about that McDonalds two blocks from Ground Zero? That's killed more people than the nineteen hijackers.

Who thinks:

1. Text A was pro-Democrat and Text B was pro-Republican?
2. Text A was pro-Republican and Text A was pro-Democrat?

Corpus Statistics:

- ▶ $N = 1,408$ unique documents
- ▶ $D = 20,521$ unique words

Document A:

i want to talk about jobs lately it seems that everyone says they want to talk about jobs and that we'll get around to tackling jobs next week or the week after

Document B:

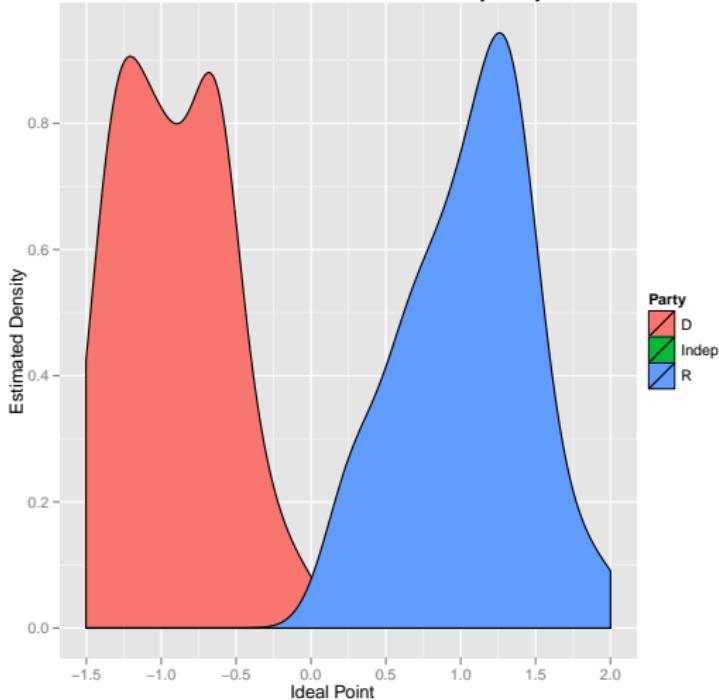
*there was a major legislative accomplishment in
washington last week and it's getting less attention than
it deserves because it isn't national health care reform*

Document Term Matrix:

Document	I	Want	Talk	Jobs	Week
A	1	2	2	3	2
B	0	0	0	0	1

- ▶ Fit Lasso regression to word counts
- ▶ Predict ideal points for senators

Estimated Distribution of Ideal Points by Party



Top 10 Most Republican Terms:

Term	Value
okla	1.23
bailey	0.647
johnny	0.588
administering	0.561
neb	0.556
sam	0.542
986	0.532
texans	0.493
patriotism	0.466
demint	0.417

Top 10 Most Democratic Terms:

Term	Value
sherrod	-0.367
sheldon	-0.249
dec	-0.196
possess	-0.168
salaries	-0.158
tom	-0.152
debbie	-0.151
dark	-0.148
lautenberg	-0.133
fought	-0.106

Debugging:

- ▶ Too many names of senators in our list
- ▶ Strip out all the names from corpus
- ▶ Run analysis from scratch on clean corpus

Top 10 Most Republican Terms excluding Names:

Term	Value
okla	1.13
neb	0.726
bailey	0.674
2415	0.638
986	0.578
kansans	0.543
administering	0.516
texans	0.467
profoundly	0.459
patriotism	0.430

Top 10 Most Democratic Terms excluding Names:

Term	Value
cedar	-0.224
chaired	-0.197
dec	-0.158
dark	-0.146
blocked	-0.138
reverses	-0.134
1960s	-0.125
insurers	-0.0958
fought	-0.0926
possess	-0.0923