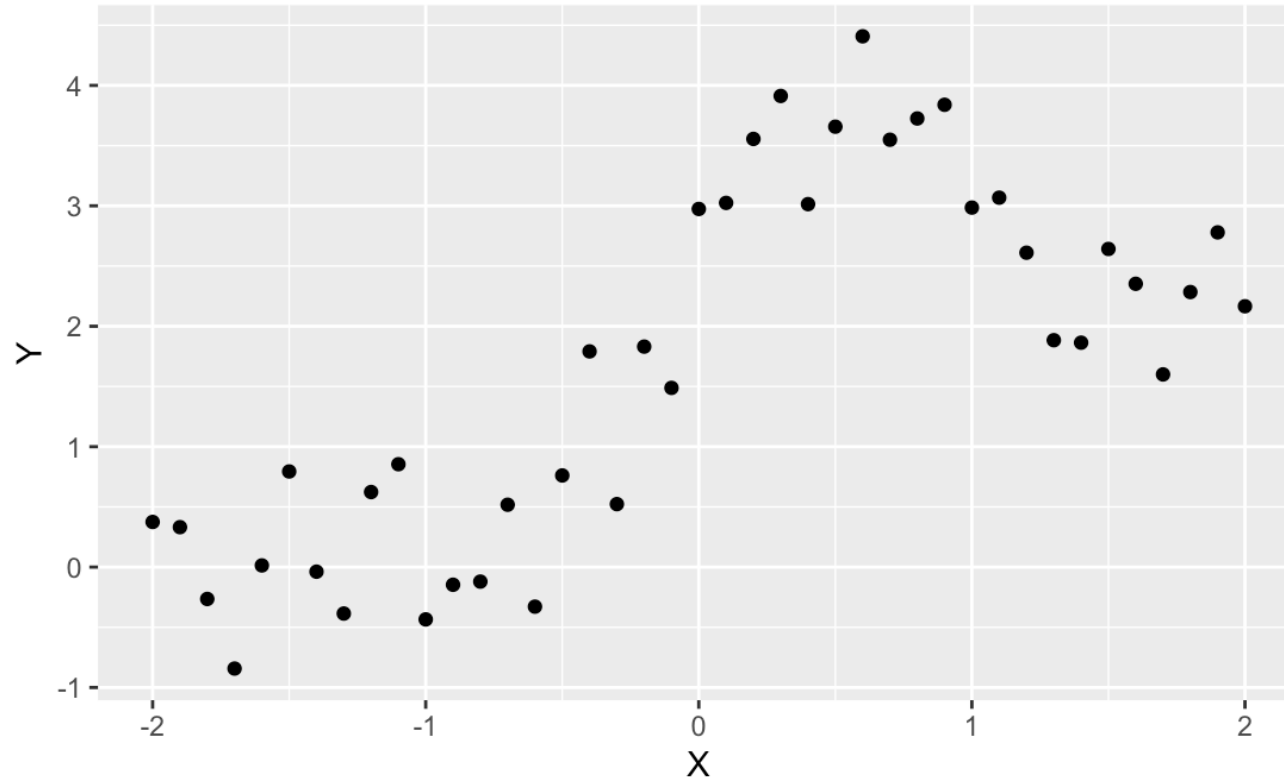


Reproducing Kernel Hilbert Spaces

David Sharkansky

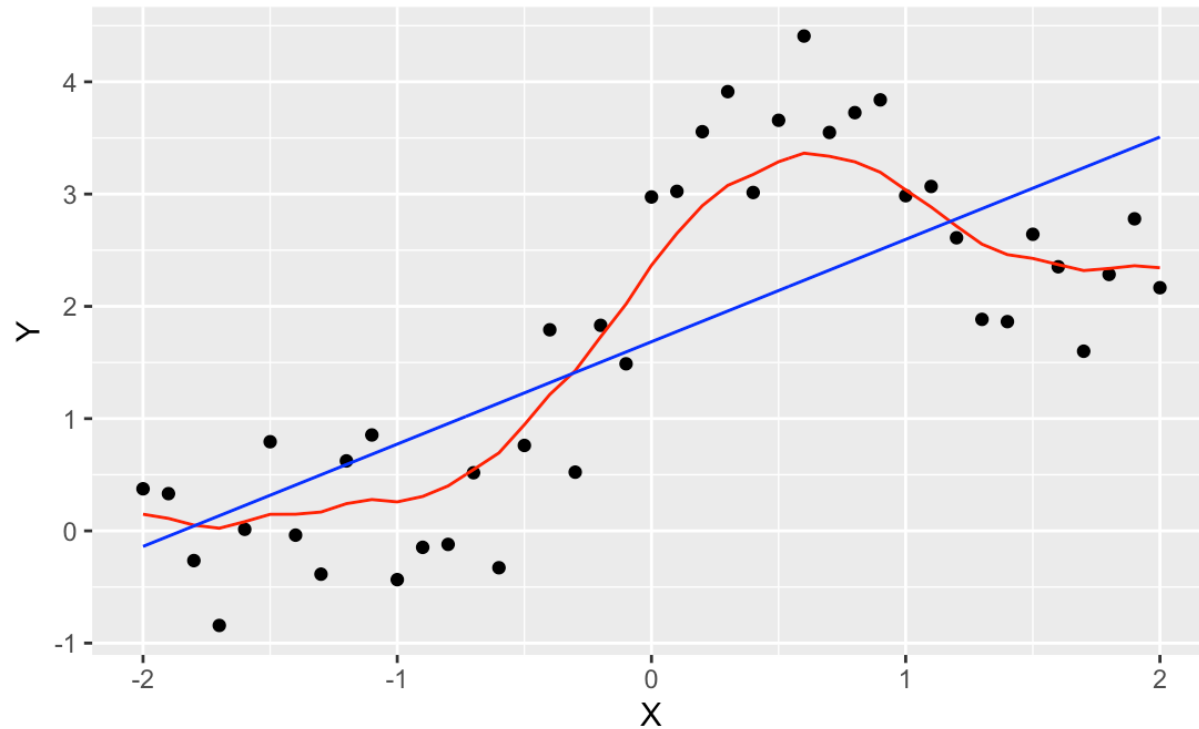
Motivation

- Fitting a function to data
- We want a function that is close to the data points but not overfitted
- We can make up any function but how do we make it make sense?



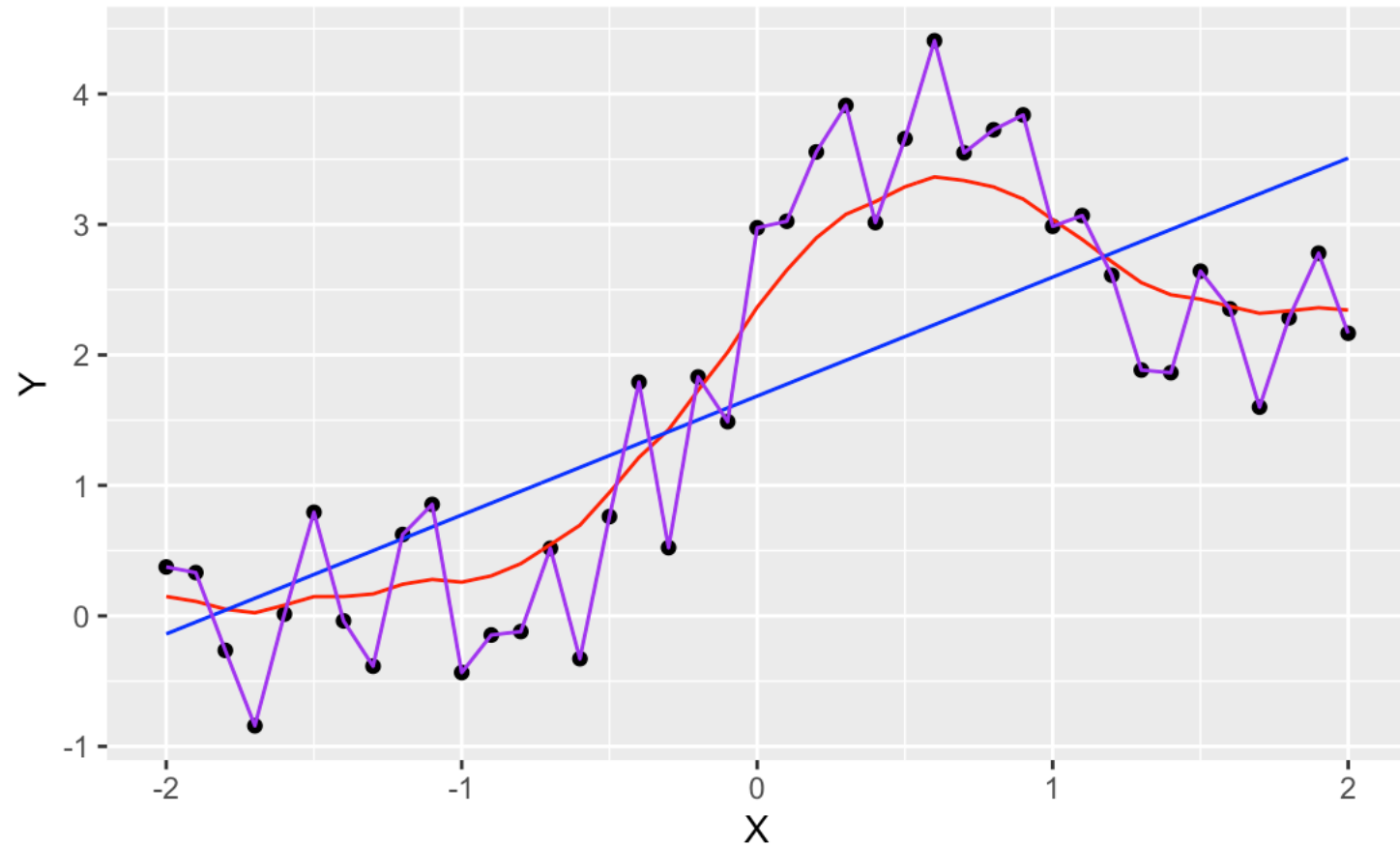
Reproducing Kernel Hilbert Spaces

- Find a function within a family of functions called a Reproducing Kernel Hilbert Space
- Start with a function of two points $K(x, y)$
- Build functions using this kernel function fixing one point
- Examples: linear, Gaussian, Sobolev



Penalization

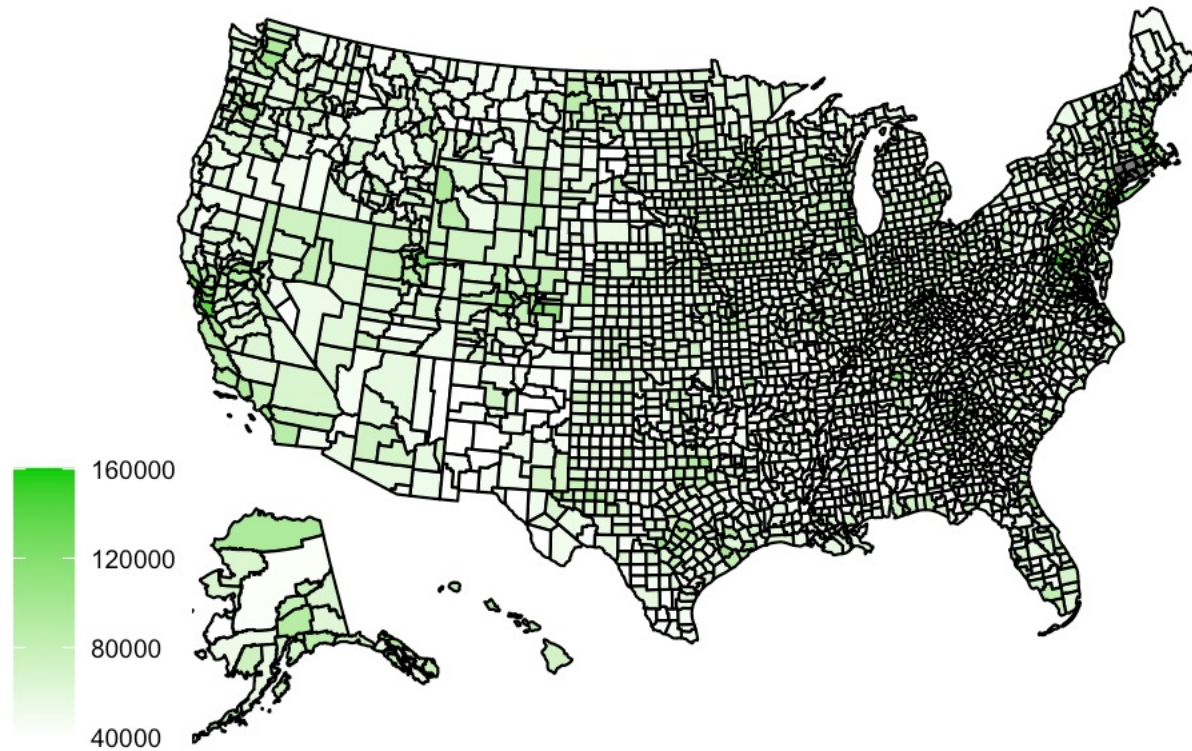
- Add a penalization parameter for functions that are not smooth
- Need the function to fit unknown data points as well, so we need to account for error and smooth the functions



Problem

- Election data by county based on several demographic factors
- Can we accurately predict how counties vote based on demographics?

Median Household Income by County



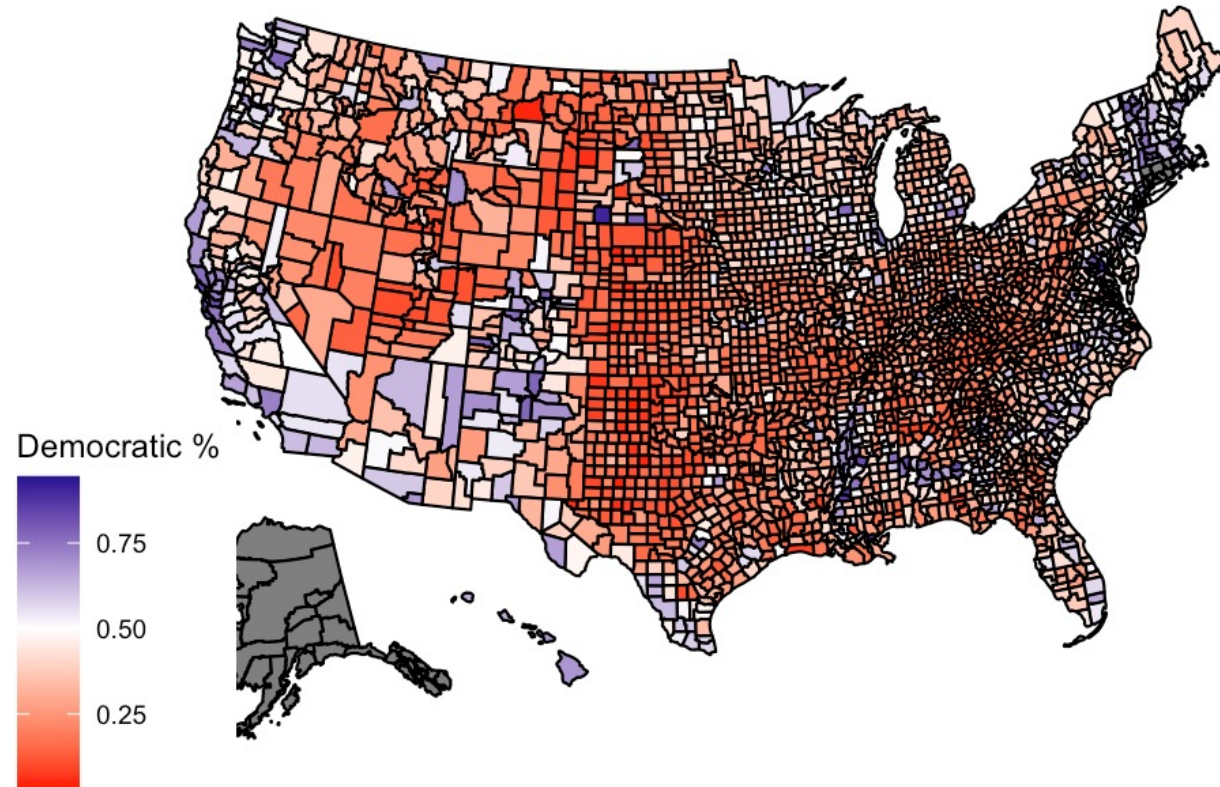
Applying RKHS

- Can we estimate vote proportions as a function of demographic factors?
- Which kernel does the best job of predicting vote share?

Data

- Chose 5 demographic variables as predictors
- Democratic vote share as a proportion of total Democratic and Republican votes as the response

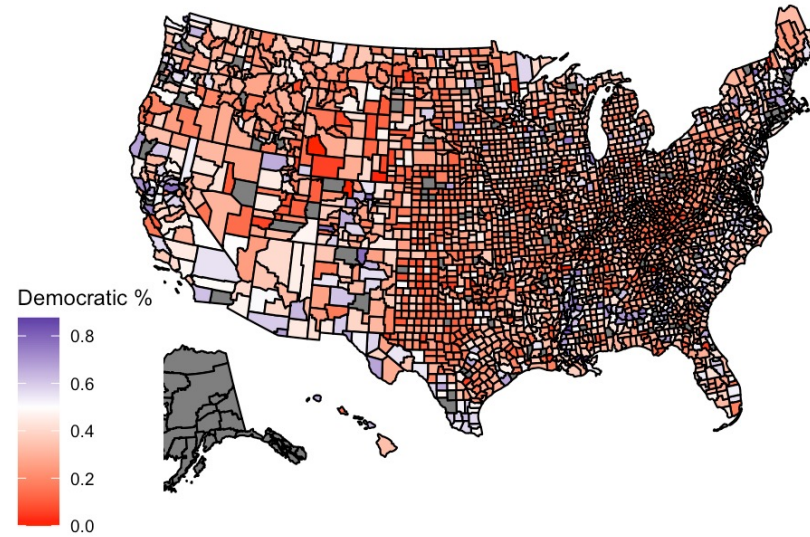
Democratic Vote Share in 2020 by County



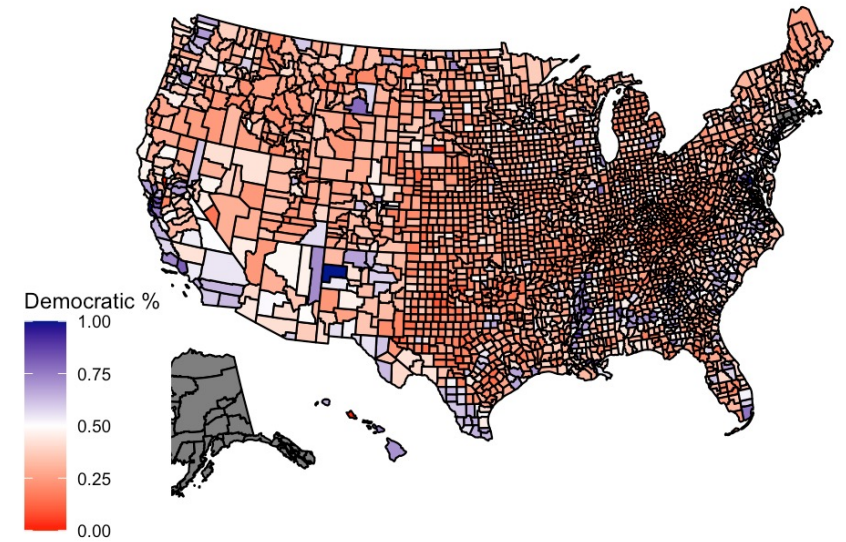
Results

- Sobolev kernel mean squared error: 0.01726362
- Gaussian kernel mean squared error: 0.01184931

Sobolev Kernel Estimate of 2020 Vote



Gaussian Kernel Estimate of 2020 Vote



What I Learned

- Reproducing kernel Hilbert spaces can be helpful when we do not know what kind of function we want
- Finding the right function is difficult
- Tuning functions is very important

Thank you!