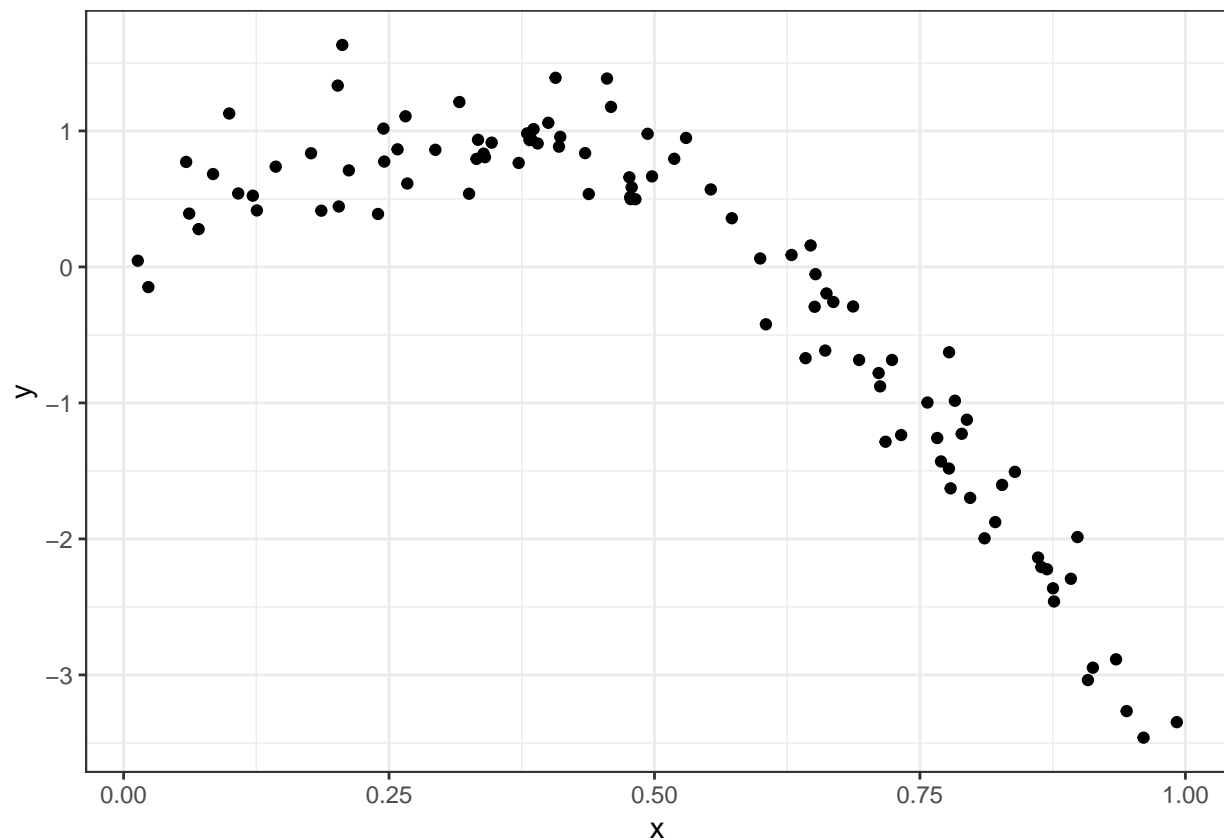


Cross validation

Annie Yu (xy2404)

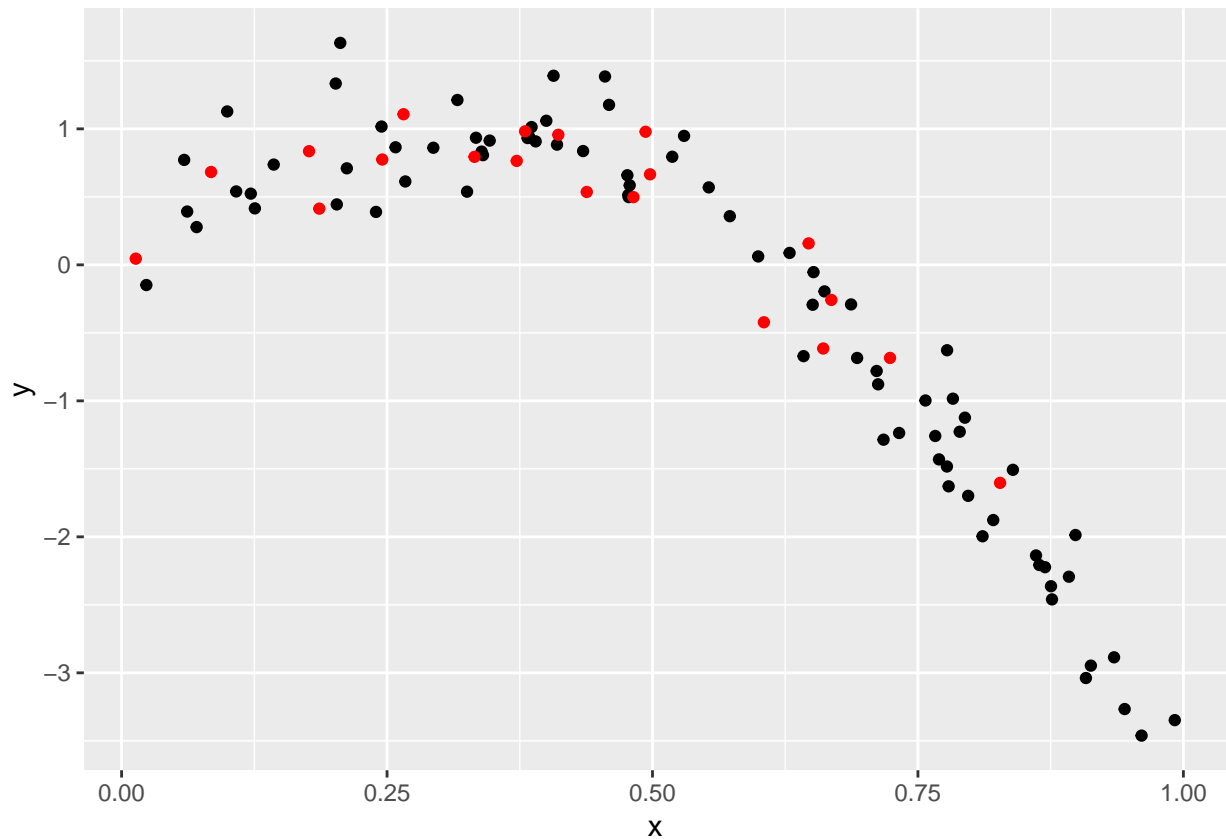
11/13/2018

```
nonlin_df = tibble(  
  id = 1:100,  
  x = runif(100, 0, 1),  
  y = 1 - 10 * (x - .3) ^ 2 + rnorm(100, 0, .3)  
)  
  
ggplot(nonlin_df, aes(x = x, y = y)) + geom_point() + theme_bw()
```



Partition into training and testing

```
train_df = sample_n(nonlin_df, 80)  
test_df = anti_join(nonlin_df, train_df, by = "id")  
  
ggplot(train_df, aes(x = x, y = y)) +  
  geom_point() +  
  geom_point(data = test_df, color = "red")
```

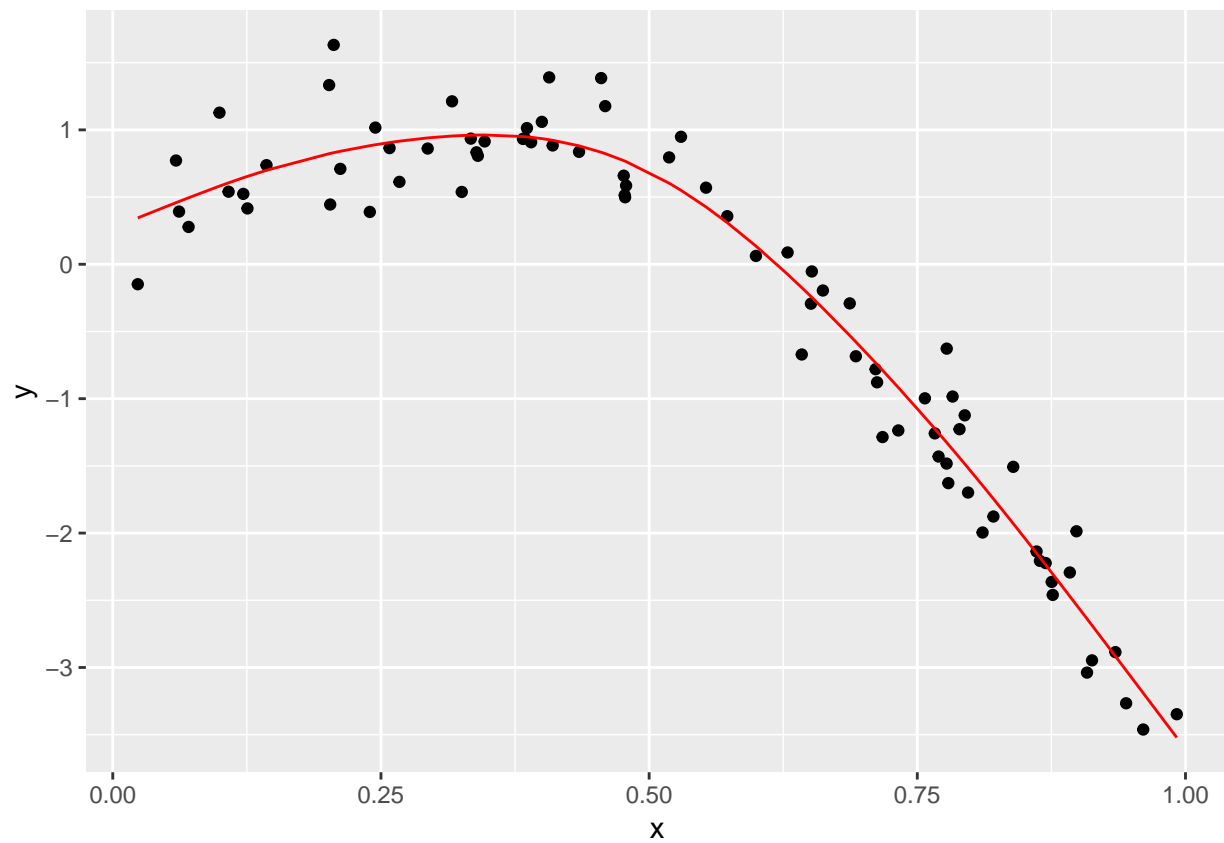


fit a few model

```
lin_mod = lm(y ~ x, data = train_df)
nonlin_mod = mgcv::gam(y ~ s(x), data = train_df)
wiggly_mod = mgcv::gam(y ~ s(x, k = 30), sp = 10e-6, data = train_df)
```

Show what these look like:

```
train_df %>%
  add_predictions(nonlin_mod) %>%
  ggplot(aes(x = x, y = y)) + geom_point() +
  geom_line(aes(y = pred), color = "red")
```



```
train_df %>%  
  add_predictions(wiggly_mod) %>%  
  ggplot(aes(x = x, y = y)) + geom_point() +  
  geom_line(aes(y = pred), color = "red")
```

