**To Do Paper:**

* Overview table which models have which preprocessing steps
* Flow chart of the modelling approach, including tree structure for the directory

**Feature Engineering**

* Which problems remain that have to be solved by data cleaning?
  + Outliers, measurement errors?

**Models**

* Partitioning the data:
  + Splits and CV
* Hyperparameter tuning approach
  + Other ways that were tried for partitioning or for hyperparameter tuning? (maybe use different evaluation metrics, see which one works best)
  + Tune the number of folds between 5 and 10
* Baseline models
  + Always predicting the mean
  + Linear Models
* Model Specifications
  + GB
  + nonlinear RBF SVM
    - Normalisation changes nothing
  + OLS
    - Interactions, natural spline, Polynomial terms, Log Transforms?
  + step\_interact( ~ Gr\_Liv\_Area:starts\_with("Bldg\_Type\_") ) %>%
  + step\_ns(Latitude, deg\_free = 20)