

**Prediction of oxygen uptake dynamics by machine learning analysis of
wearable sensors during activities of daily living**

NINA ZHANG

215319003

Load Packages

Load some packages we will use.

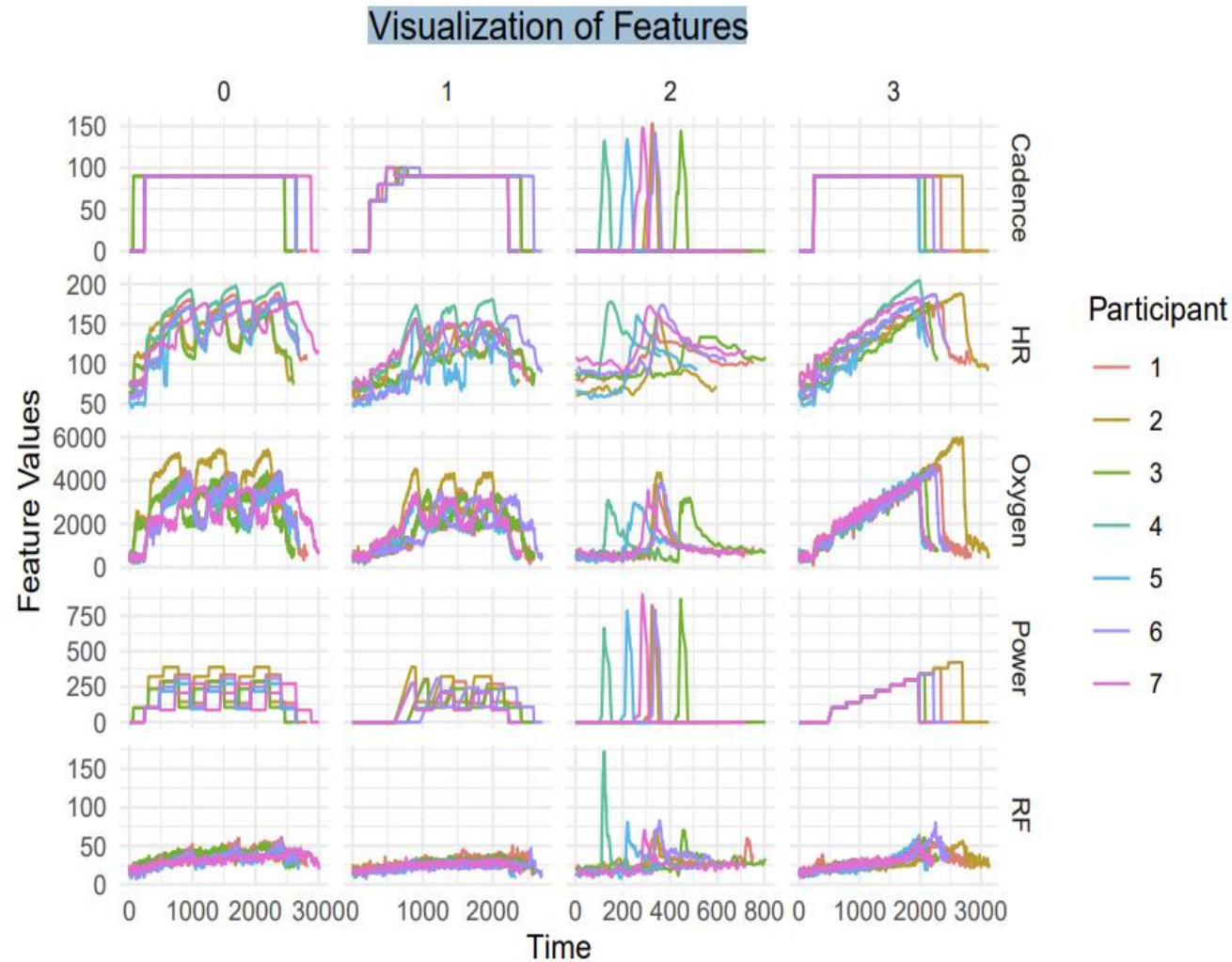
- `library(tidyverse)`
- `library(tidytext)`
- `library(corrplot)`
- `library(doMC)`
- `library(caret)`

Import Data

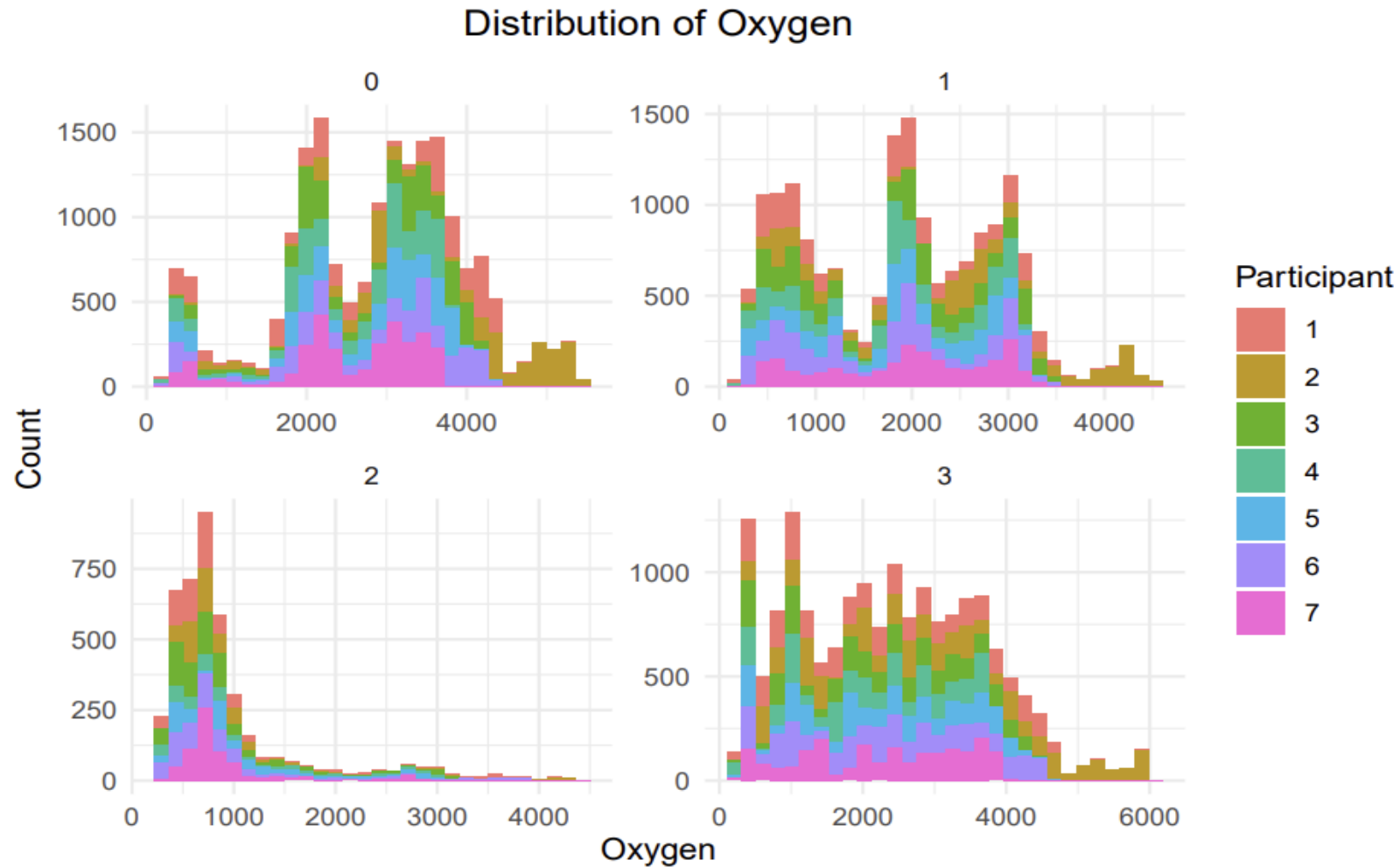
- Import “Kaggle_Data.csv” into R.
- `data = read_csv('Kaggle_Data.csv')`

Exploratory Data Analysis

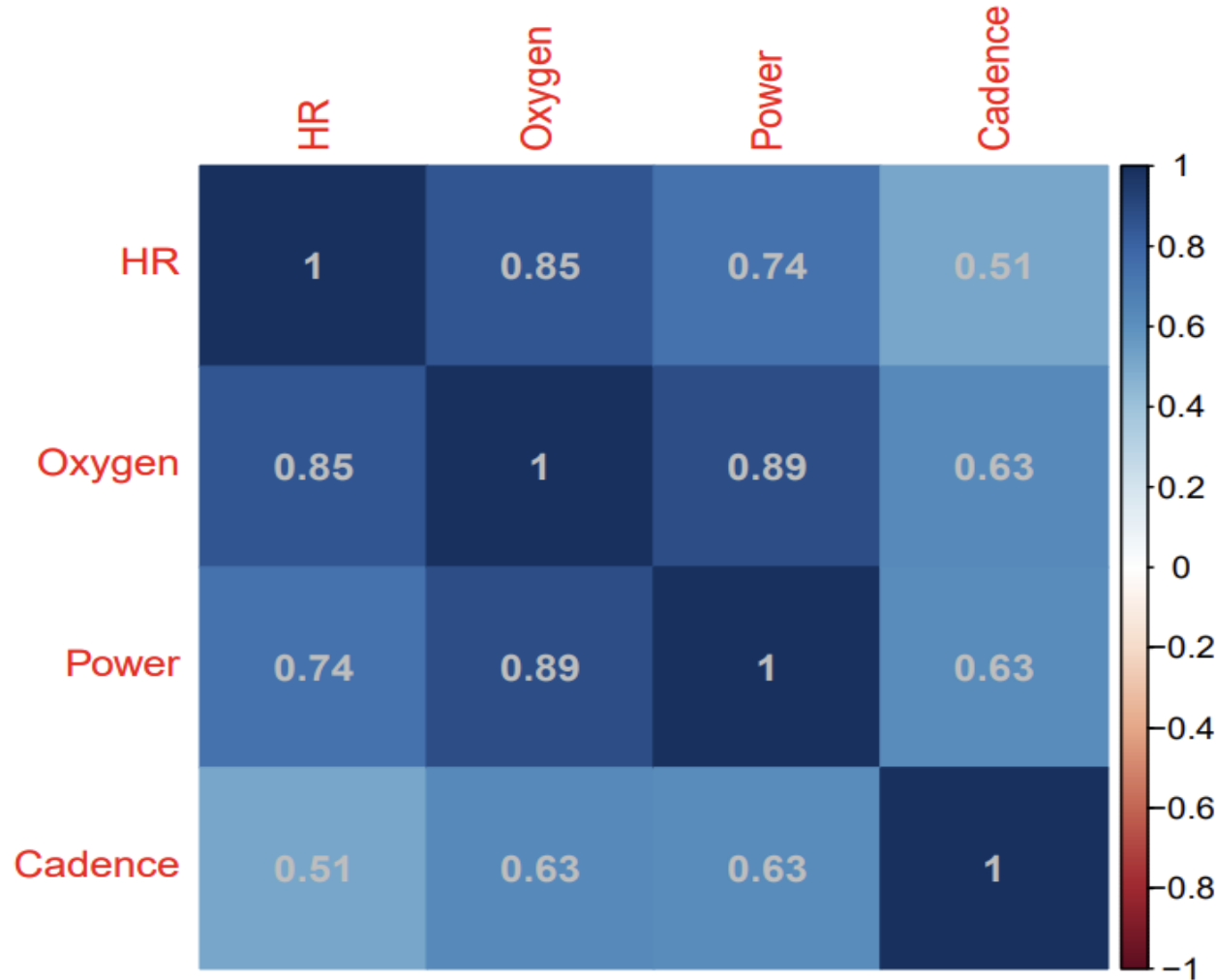
Visualize the features in the data.



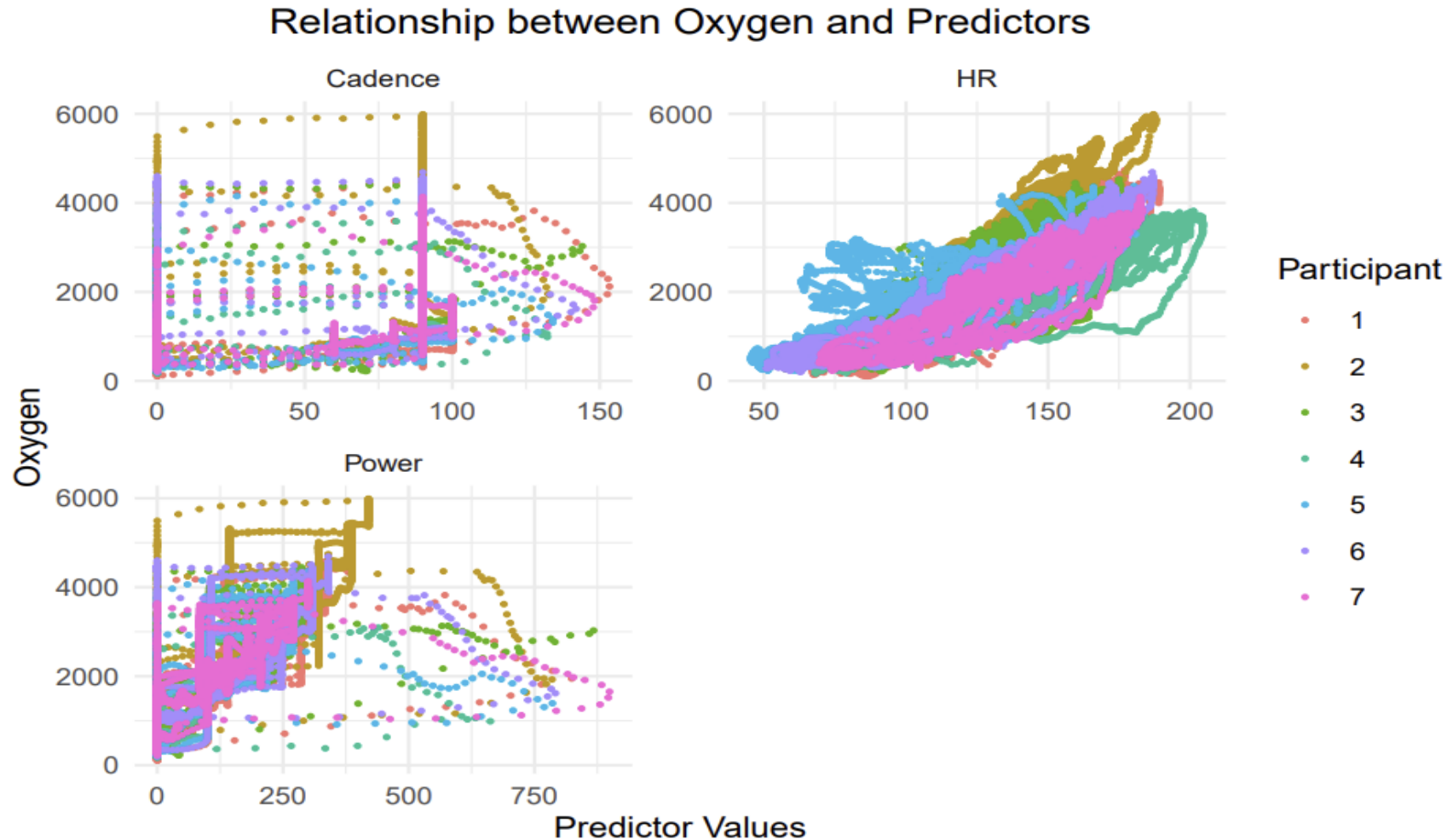
Visualize the distribution of Oxygen



Visualize the correlations between these features.



Visualize the relationship between predictors and Oxygen.



Modeling

- Data Split
- Split the data into training set(70%) and test set(30%).
- We will fit random forest, xgboost and neural network using 10-fold cross-validation.

Random Forest

- Tune hyper-parameters for a random forest on the training set.
- The final parameters for random forest are as follows.
- `RF$bestTune`
- Fit random forest using best hyper-parameters on training set.

XGBoost

- Tune hyper-parameters for xgboost on training set.
- The final parameters for xgboost are as follows.
- `XGBoost$bestTune`
- Fit xgboost using best hyper-parameters on training set.

Neural Network

- Tune hyper-parameters for neural network on training set.
- The final parameters for the neural network are as follows.
NNET\$bestTune
- Fit neural network using best hyper-parameters on training set.

Model Evaluation

- Predict Oxygen using three models on test set.
- Calculate RMSE, MAE and R-squared for three models on test set.

Visualize these metric results.

