The plumber package for R makes it easy to expose existing R code as a webservice via an API (Trestle Technology, LLC 2017).

You take an existing R script and make it accessible with plumber by simply adding a few lines of comments. If you have worked with Roxygen before, e.g. when building a package, you will already be familiar with the core concepts. If not, here are the most important things to know:

* you define the output or endpoint
* you can add additional annotation to customize your input, output and other functionalities of your API
* you can define every input parameter that will go into your function
* every such annotation will begin with either #' or #\*

With this setup, we can take a trained machine learning model and make it available as an API. With this API, other programs can access it and use it to make predictions.

**What are APIs and webservices?**

With plumber, we can build so called **HTTP APIs**. HTTP stands for Hypertext Transfer Protocol and is used to transmit information on the web; API stands for Application Programming Interface and governs the connection between some software and underlying applications. Software can then communicate via HTTP APIs. This way, our R script can be called from other software, even if the other program is not written in R and we have built a tool for machine-to-machine communication, i.e. a webservice.

**How to convert your R script into an API with plumber**

**Training and saving a model**

Let’s say we have trained a machine learning model as in below.

**Example Code**

* the following libraries were loaded:

library(tidyverse) # for tidy data analysis

library(farff) # for reading arff file

library(missForest) # for imputing missing values

library(dummies) # for creating dummy variables

library(caret) # for modeling

library(lime) # for explaining predictions

**Data**

The Chronic Kidney Disease dataset was downloaded from UC Irvine’s Machine Learning repository: <http://archive.ics.uci.edu/ml/datasets/Chronic_Kidney_Disease>

data\_file <- file.path("path/to/chronic\_kidney\_disease\_full.arff")

* load data with the farff package

data <- readARFF(data\_file)

**Features**

* age - age
* bp - blood pressure
* sg - specific gravity
* al - albumin
* su - sugar
* rbc - red blood cells
* pc - pus cell
* pcc - pus cell clumps
* ba - bacteria
* bgr - blood glucose random
* bu - blood urea
* sc - serum creatinine
* sod - sodium
* pot - potassium
* hemo - hemoglobin
* pcv - packed cell volume
* wc - white blood cell count
* rc - red blood cell count
* htn - hypertension
* dm - diabetes mellitus
* cad - coronary artery disease
* appet - appetite
* pe - pedal edema
* ane - anemia
* class - class

**Missing data**

* impute missing data with Nonparametric Missing Value Imputation using Random Forest (missForest package)

data\_imp <- missForest(data)

**One-hot encoding**

* create dummy variables (caret::dummy.data.frame())
* scale and center

data\_imp\_final <- data\_imp$ximp

data\_dummy <- dummy.data.frame(dplyr::select(data\_imp\_final, -class), sep = "\_")

data <- cbind(dplyr::select(data\_imp\_final, class), scale(data\_dummy,

center = apply(data\_dummy, 2, min),

scale = apply(data\_dummy, 2, max)))

**Modeling**

# training and test set

set.seed(42)

index <- createDataPartition(data$class, p = 0.9, list = FALSE)

train\_data <- data[index, ]

test\_data <- data[-index, ]

# modeling

model\_rf <- caret::train(class ~ .,

data = train\_data,

method = "rf", # random forest

trControl = trainControl(method = "repeatedcv",

number = 10,

repeats = 5,

verboseIter = FALSE))

model\_rf

## Random Forest

##

## 360 samples

## 48 predictor

## 2 classes: 'ckd', 'notckd'

##

## No pre-processing

## Resampling: Cross-Validated (10 fold, repeated 5 times)

## Summary of sample sizes: 324, 324, 324, 324, 325, 324, ...

## Resampling results across tuning parameters:

##

## mtry Accuracy Kappa

## 2 0.9922647 0.9838466

## 25 0.9917392 0.9826070

## 48 0.9872930 0.9729881

##

## Accuracy was used to select the optimal model using the largest value.

## The final value used for the model was mtry = 2.

# predictions

pred <- data.frame(sample\_id = 1:nrow(test\_data), predict(model\_rf, test\_data, type = "prob"), actual = test\_data$class) %>%

mutate(prediction = colnames(.)[2:3][apply(.[, 2:3], 1, which.max)], correct = ifelse(actual == prediction, "correct", "wrong"))

confusionMatrix(pred$actual, pred$prediction)

## Confusion Matrix and Statistics

##

## Reference

## Prediction ckd notckd

## ckd 23 2

## notckd 0 15

##

## Accuracy : 0.95

## 95% CI : (0.8308, 0.9939)

## No Information Rate : 0.575

## P-Value [Acc > NIR] : 1.113e-07

##

## Kappa : 0.8961

## Mcnemar's Test P-Value : 0.4795

##

## Sensitivity : 1.0000

## Specificity : 0.8824

## Pos Pred Value : 0.9200

## Neg Pred Value : 1.0000

## Prevalence : 0.5750

## Detection Rate : 0.5750

## Detection Prevalence : 0.6250

## Balanced Accuracy : 0.9412

##

## 'Positive' Class : ckd

##

**LIME**

* LIME needs data without response variable

train\_x <- dplyr::select(train\_data, -class)

test\_x <- dplyr::select(test\_data, -class)

train\_y <- dplyr::select(train\_data, class)

test\_y <- dplyr::select(test\_data, class)

* build explainer

explainer <- lime(train\_x, model\_rf, n\_bins = 5, quantile\_bins = TRUE)

* run explain() function

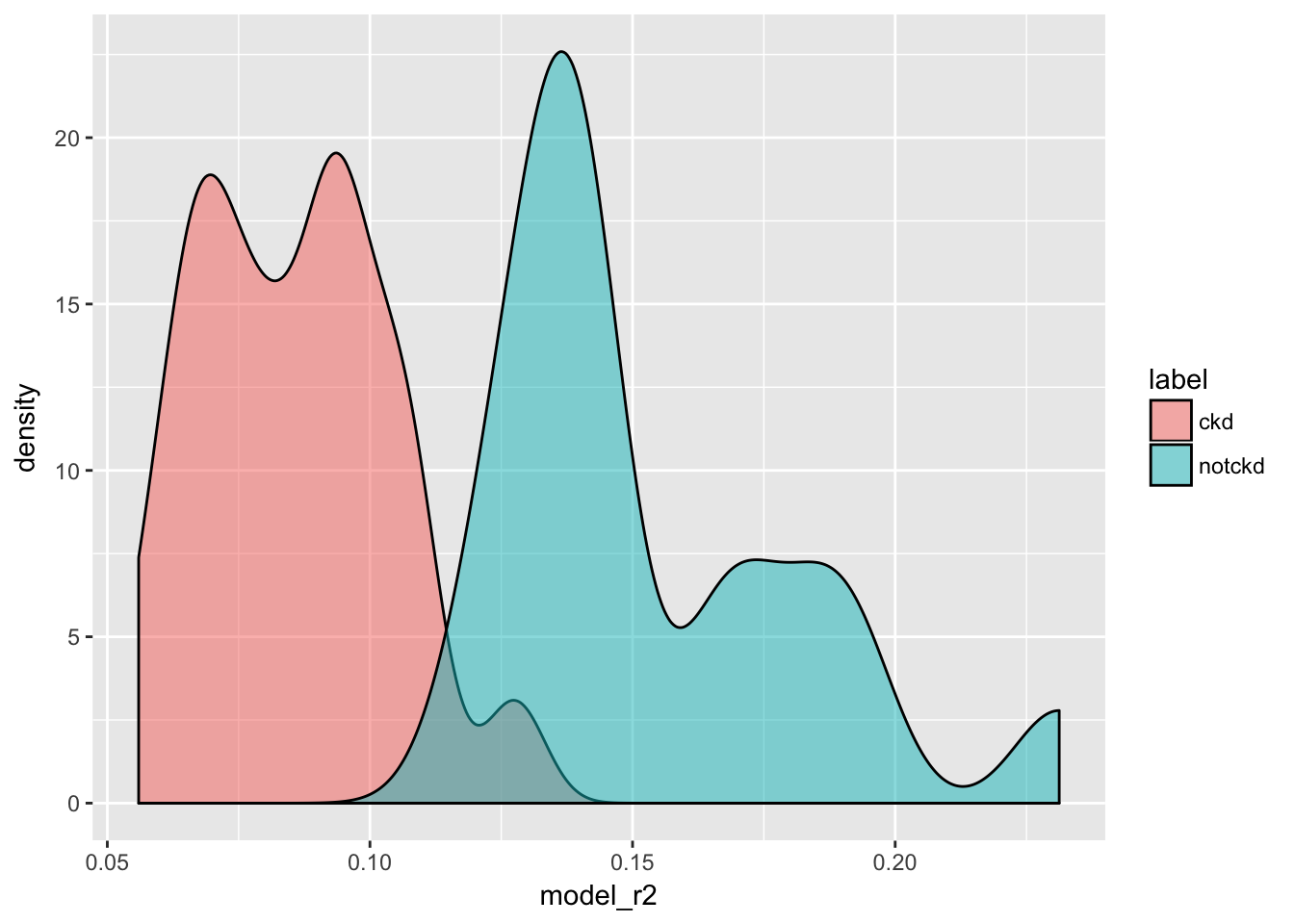
explanation\_df <- lime::explain(test\_x, explainer, n\_labels = 1, n\_features = 8, n\_permutations = 1000, feature\_select = "forward\_selection")

* model reliability

explanation\_df %>%

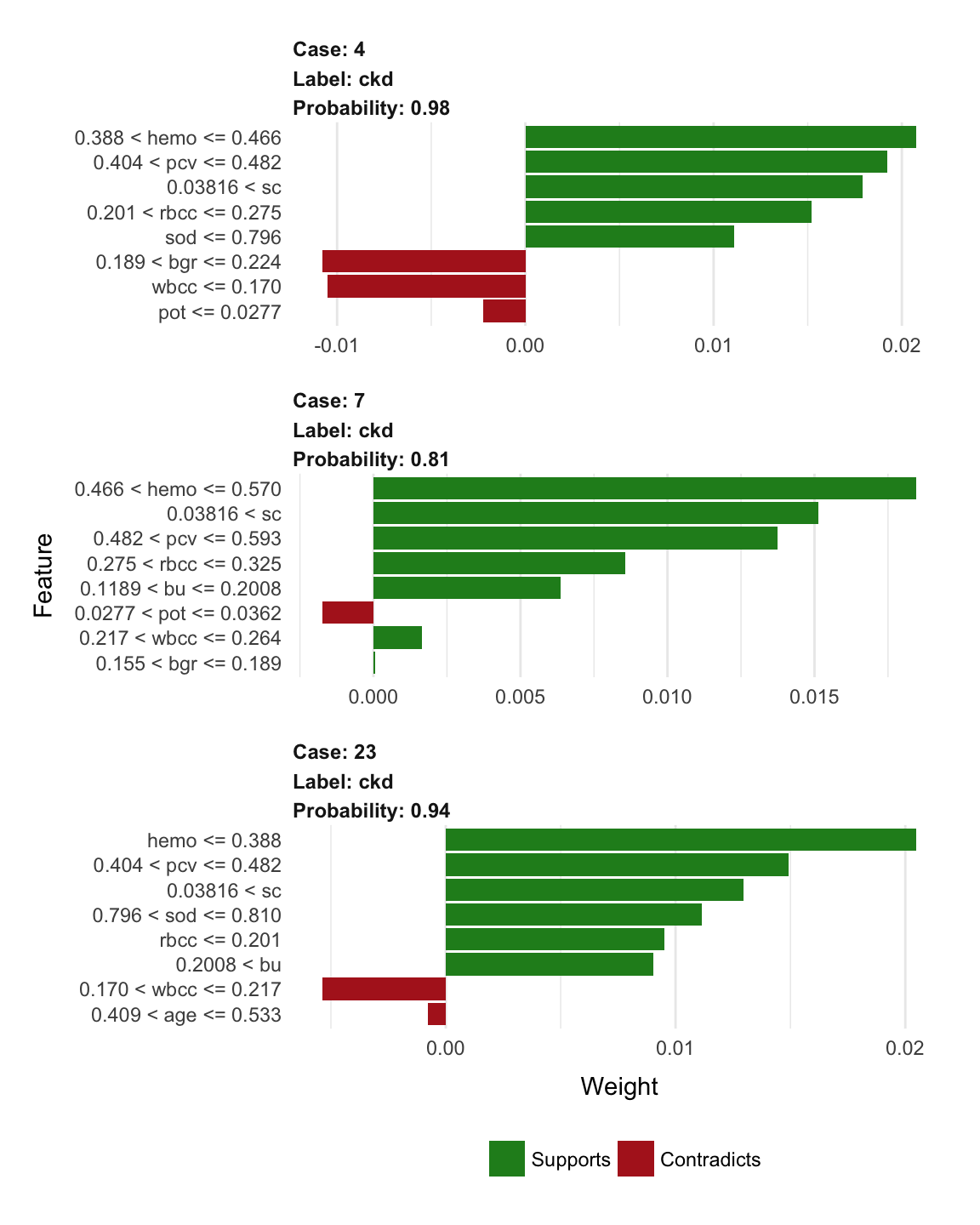
ggplot(aes(x = model\_r2, fill = label)) +

geom\_density(alpha = 0.5)



* plot explanations

plot\_features(explanation\_df[1:24, ], ncol = 1)



**Session Info**

## Session info -------------------------------------------------------------

## setting value

## version R version 3.4.3 (2017-11-30)

## system x86\_64, darwin15.6.0

## ui X11

## language (EN)

## collate de\_DE.UTF-8

## tz Europe/Berlin

## date 2018-04-22

## Packages -----------------------------------------------------------------

## package \* version date source

## assertthat 0.2.0 2017-04-11 CRAN (R 3.4.0)

## backports 1.1.2 2017-12-13 CRAN (R 3.4.3)

## base \* 3.4.3 2017-12-07 local

## BBmisc 1.11 2017-03-10 CRAN (R 3.4.0)

## bindr 0.1 2016-11-13 CRAN (R 3.4.0)

## bindrcpp \* 0.2 2017-06-17 CRAN (R 3.4.0)

## blogdown 0.5 2018-01-24 CRAN (R 3.4.3)

## bookdown 0.7 2018-02-18 CRAN (R 3.4.3)

## broom 0.4.3 2017-11-20 CRAN (R 3.4.2)

## caret \* 6.0-78 2017-12-10 CRAN (R 3.4.3)

## cellranger 1.1.0 2016-07-27 CRAN (R 3.4.0)

## checkmate 1.8.5 2017-10-24 CRAN (R 3.4.2)

## class 7.3-14 2015-08-30 CRAN (R 3.4.3)

## cli 1.0.0 2017-11-05 CRAN (R 3.4.2)

## codetools 0.2-15 2016-10-05 CRAN (R 3.4.3)

## colorspace 1.3-2 2016-12-14 CRAN (R 3.4.0)

## compiler 3.4.3 2017-12-07 local

## crayon 1.3.4 2017-09-16 CRAN (R 3.4.1)

## CVST 0.2-1 2013-12-10 CRAN (R 3.4.0)

## datasets \* 3.4.3 2017-12-07 local

## ddalpha 1.3.1.1 2018-02-02 CRAN (R 3.4.3)

## DEoptimR 1.0-8 2016-11-19 CRAN (R 3.4.0)

## devtools 1.13.5 2018-02-18 CRAN (R 3.4.3)

## digest 0.6.15 2018-01-28 CRAN (R 3.4.3)

## dimRed 0.1.0 2017-05-04 CRAN (R 3.4.0)

## dplyr \* 0.7.4 2017-09-28 CRAN (R 3.4.2)

## DRR 0.0.3 2018-01-06 CRAN (R 3.4.3)

## dummies \* 1.5.6 2012-06-14 CRAN (R 3.4.0)

## e1071 1.6-8 2017-02-02 CRAN (R 3.4.0)

## evaluate 0.10.1 2017-06-24 CRAN (R 3.4.1)

## farff \* 1.0 2016-09-11 CRAN (R 3.4.0)

## forcats \* 0.3.0 2018-02-19 CRAN (R 3.4.3)

## foreach \* 1.4.4 2017-12-12 CRAN (R 3.4.3)

## foreign 0.8-69 2017-06-22 CRAN (R 3.4.3)

## ggplot2 \* 2.2.1.9000 2018-02-28 Github (thomasp85/ggplot2@7859a29)

## glmnet 2.0-13 2017-09-22 CRAN (R 3.4.2)

## glue 1.2.0 2017-10-29 CRAN (R 3.4.2)

## gower 0.1.2 2017-02-23 CRAN (R 3.4.0)

## graphics \* 3.4.3 2017-12-07 local

## grDevices \* 3.4.3 2017-12-07 local

## grid 3.4.3 2017-12-07 local

## gtable 0.2.0 2016-02-26 CRAN (R 3.4.0)

## haven 1.1.1 2018-01-18 CRAN (R 3.4.3)

## highr 0.6 2016-05-09 CRAN (R 3.4.0)

## hms 0.4.1 2018-01-24 CRAN (R 3.4.3)

## htmltools 0.3.6 2017-04-28 CRAN (R 3.4.0)

## htmlwidgets 1.0 2018-01-20 CRAN (R 3.4.3)

## httpuv 1.3.6.1 2018-02-28 CRAN (R 3.4.3)

## httr 1.3.1 2017-08-20 CRAN (R 3.4.1)

## ipred 0.9-6 2017-03-01 CRAN (R 3.4.0)

## iterators \* 1.0.9 2017-12-12 CRAN (R 3.4.3)

## itertools \* 0.1-3 2014-03-12 CRAN (R 3.4.0)

## jsonlite 1.5 2017-06-01 CRAN (R 3.4.0)

## kernlab 0.9-25 2016-10-03 CRAN (R 3.4.0)

## knitr 1.20 2018-02-20 CRAN (R 3.4.3)

## labeling 0.3 2014-08-23 CRAN (R 3.4.0)

## lattice \* 0.20-35 2017-03-25 CRAN (R 3.4.3)

## lava 1.6 2018-01-13 CRAN (R 3.4.3)

## lazyeval 0.2.1 2017-10-29 CRAN (R 3.4.2)

## lime \* 0.3.1 2017-11-24 CRAN (R 3.4.3)

## lubridate 1.7.3 2018-02-27 CRAN (R 3.4.3)

## magrittr 1.5 2014-11-22 CRAN (R 3.4.0)

## MASS 7.3-49 2018-02-23 CRAN (R 3.4.3)

## Matrix 1.2-12 2017-11-20 CRAN (R 3.4.3)

## memoise 1.1.0 2017-04-21 CRAN (R 3.4.0)

## methods \* 3.4.3 2017-12-07 local

## mime 0.5 2016-07-07 CRAN (R 3.4.0)

## missForest \* 1.4 2013-12-31 CRAN (R 3.4.0)

## mnormt 1.5-5 2016-10-15 CRAN (R 3.4.0)

## ModelMetrics 1.1.0 2016-08-26 CRAN (R 3.4.0)

## modelr 0.1.1 2017-07-24 CRAN (R 3.4.1)

## munsell 0.4.3 2016-02-13 CRAN (R 3.4.0)

## nlme 3.1-131.1 2018-02-16 CRAN (R 3.4.3)

## nnet 7.3-12 2016-02-02 CRAN (R 3.4.3)

## parallel 3.4.3 2017-12-07 local

## pillar 1.2.1 2018-02-27 CRAN (R 3.4.3)

## pkgconfig 2.0.1 2017-03-21 CRAN (R 3.4.0)

## plyr 1.8.4 2016-06-08 CRAN (R 3.4.0)

## prodlim 1.6.1 2017-03-06 CRAN (R 3.4.0)

## psych 1.7.8 2017-09-09 CRAN (R 3.4.1)

## purrr \* 0.2.4 2017-10-18 CRAN (R 3.4.2)

## R6 2.2.2 2017-06-17 CRAN (R 3.4.0)

## randomForest \* 4.6-12 2015-10-07 CRAN (R 3.4.0)

## Rcpp 0.12.15 2018-01-20 CRAN (R 3.4.3)

## RcppRoll 0.2.2 2015-04-05 CRAN (R 3.4.0)

## readr \* 1.1.1 2017-05-16 CRAN (R 3.4.0)

## readxl 1.0.0 2017-04-18 CRAN (R 3.4.0)

## recipes 0.1.2 2018-01-11 CRAN (R 3.4.3)

## reshape2 1.4.3 2017-12-11 CRAN (R 3.4.3)

## rlang 0.2.0.9000 2018-02-28 Github (tidyverse/rlang@9ea33dd)

## rmarkdown 1.8 2017-11-17 CRAN (R 3.4.2)

## robustbase 0.92-8 2017-11-01 CRAN (R 3.4.2)

## rpart 4.1-13 2018-02-23 CRAN (R 3.4.3)

## rprojroot 1.3-2 2018-01-03 CRAN (R 3.4.3)

## rstudioapi 0.7 2017-09-07 CRAN (R 3.4.1)

## rvest 0.3.2 2016-06-17 CRAN (R 3.4.0)

## scales 0.5.0.9000 2018-02-28 Github (hadley/scales@d767915)

## sfsmisc 1.1-1 2017-06-08 CRAN (R 3.4.0)

## shiny 1.0.5 2017-08-23 CRAN (R 3.4.1)

## shinythemes 1.1.1 2016-10-12 CRAN (R 3.4.0)

## splines 3.4.3 2017-12-07 local

## stats \* 3.4.3 2017-12-07 local

## stats4 3.4.3 2017-12-07 local

## stringdist 0.9.4.6 2017-07-31 CRAN (R 3.4.1)

## stringi 1.1.6 2017-11-17 CRAN (R 3.4.2)

## stringr \* 1.3.0 2018-02-19 CRAN (R 3.4.3)

## survival 2.41-3 2017-04-04 CRAN (R 3.4.3)

## tibble \* 1.4.2 2018-01-22 CRAN (R 3.4.3)

## tidyr \* 0.8.0 2018-01-29 CRAN (R 3.4.3)

## tidyselect 0.2.4 2018-02-26 CRAN (R 3.4.3)

## tidyverse \* 1.2.1 2017-11-14 CRAN (R 3.4.2)

## timeDate 3043.102 2018-02-21 CRAN (R 3.4.3)

## tools 3.4.3 2017-12-07 local

## utils \* 3.4.3 2017-12-07 local

## withr 2.1.1.9000 2018-02-28 Github (jimhester/withr@5d05571)

## xfun 0.1 2018-01-22 CRAN (R 3.4.3)

## xml2 1.2.0 2018-01-24 CRAN (R 3.4.3)

## xtable 1.8-2 2016-02-05 CRAN (R 3.4.0)

## yaml 2.1.17 2018-02-27 CRAN (R 3.4.3)

I loaded a data set on chronic kidney disease, did some preprocessing (converting categorical features into dummy variables, scaling and centering), split it into training and test data and trained a Random Forest model with caret. We can use this trained model to make predictions for one test case with the following code:

library(tidyverse)

# load test and train data

load("../../data/test\_data.RData")

load("../../data/train\_data.RData")

# load model

load("../../data/model\_rf.RData")

# take first test case for prediction

input\_data <- test\_data[1, ] %>%

select(-class)

# predict test case using model

pred <- predict(model\_rf, input\_data)

cat("----------------\nTest case predicted to be", as.character(pred), "\n----------------")

## ----------------

## Test case predicted to be ckd

## ----------------

**The input**

For our API to work, we need to define the input, in our case the features of the test data. When we look at the model object, we see that it expects the following parameters:

var\_names <- model\_rf$finalModel$xNames

var\_names

## [1] "age" "bp" "sg\_1.005" "sg\_1.010"

## [5] "sg\_1.015" "sg\_1.020" "sg\_1.025" "al\_0"

## [9] "al\_1" "al\_2" "al\_3" "al\_4"

## [13] "al\_5" "su\_0" "su\_1" "su\_2"

## [17] "su\_3" "su\_4" "su\_5" "rbc\_normal"

## [21] "rbc\_abnormal" "pc\_normal" "pc\_abnormal" "pcc\_present"

## [25] "pcc\_notpresent" "ba\_present" "ba\_notpresent" "bgr"

## [29] "bu" "sc" "sod" "pot"

## [33] "hemo" "pcv" "wbcc" "rbcc"

## [37] "htn\_yes" "htn\_no" "dm\_yes" "dm\_no"

## [41] "cad\_yes" "cad\_no" "appet\_good" "appet\_poor"

## [45] "pe\_yes" "pe\_no" "ane\_yes" "ane\_no"

Good practice is to write the input parameter definition into you [API Swagger UI](https://swagger.io/swagger-ui/), but the code would work without these annotations. We define the parameters by annotating them with name and description in our R-script using @parameter. For this purpose, I want to know the type and min/max values for each of my variables in the training data. Because categorical data has been converted to dummy variables and then scaled and centered, these values will all be numeric and between 0 and 1 in this example. If I would build this script for a real case, I’d use the raw data as input and add a preprocessing function to my script, though!

# show parameter definition for the first three features

for (i in 1:3) {

# if you wanted to see it for all features, use

#for (i in 1:length(var\_names)) {

var <- var\_names[i]

train\_data\_subs <- train\_data[, which(colnames(train\_data) == var)]

type <- class(train\_data\_subs)

if (type == "numeric") {

min <- min(train\_data\_subs)

max <- max(train\_data\_subs)

}

cat("Variable:", var, "is of type:", type, "\n",

"Min value in training data =", min, "\n",

"Max value in training data =", max, "\n----------\n")

}

## Variable: age is of type: numeric

## Min value in training data = 0

## Max value in training data = 0.9777778

## ----------

## Variable: bp is of type: numeric

## Min value in training data = 0

## Max value in training data = 0.7222222

## ----------

## Variable: sg\_1.005 is of type: numeric

## Min value in training data = 0

## Max value in training data = 1

## ----------

Unless otherwise instructed, all parameters passed into plumber endpoints from query strings or dynamic paths will be character strings.

This means that we need to convert numeric values before we process them further. Or we can define the parameter type explicitly, e.g. by writing variable\_1:numeric if we want to specifiy that *variable\_1* is supposed to be numeric.

To make sure that the model will perform as expected, it is also advisable to add a few validation functions. Here, I will validate

* whether every parameter is numeric/integer by checking for NAs (which would have resulted from as.numeric()/as.integer() applied to data of character type)
* whether every parameter is between 0 and 1

In order for plumber to work with our input, it needs to be part of the HTTP request, which can then be routed to our R function. The [plumber documentation](https://www.rplumber.io/docs/routing-and-input.html#query-strings) describes how to use query strings as inputs. But in our case, manually writing query strings is not practical because we have so many parameters. Of course, there are programs that let us generate query strings but the easiest way to format the input from a line of data I found is to use JSON.

The toJSON() function from the rjson package converts our input line to JSON format:

library(rjson)

test\_case\_json <- toJSON(input\_data)

cat(test\_case\_json)

## {"age":0.511111111111111,"bp":0.111111111111111,"sg\_1.005":1,"sg\_1.010":0,"sg\_1.015":0,"sg\_1.020":0,"sg\_1.025":0,"al\_0":0,"al\_1":0,"al\_2":0,"al\_3":0,"al\_4":1,"al\_5":0,"su\_0":1,"su\_1":0,"su\_2":0,"su\_3":0,"su\_4":0,"su\_5":0,"rbc\_normal":1,"rbc\_abnormal":0,"pc\_normal":0,"pc\_abnormal":1,"pcc\_present":1,"pcc\_notpresent":0,"ba\_present":0,"ba\_notpresent":1,"bgr":0.193877551020408,"bu":0.139386189258312,"sc":0.0447368421052632,"sod":0.653374233128834,"pot":0,"hemo":0.455056179775281,"pcv":0.425925925925926,"wbcc":0.170454545454545,"rbcc":0.225,"htn\_yes":1,"htn\_no":0,"dm\_yes":0,"dm\_no":1,"cad\_yes":0,"cad\_no":1,"appet\_good":0,"appet\_poor":1,"pe\_yes":1,"pe\_no":0,"ane\_yes":1,"ane\_no":0}

**Defining the endpoint and output**

In order to convert this very simple script into an API, we need to define the endpoint(s). Endpoints will return an output, in our case it will return the output of the predict() function pasted into a line of text (e.g. “Test case predicted to be ckd”). Here, we want to have the predictions returned, so we annotate the entire function with @get. This endpoint in the API will get a custom name, so that we can call it later; here we call it predict and therefore write #' @get /predict.

According to the design of the HTTP specification, GET (along with HEAD) requests are used only to read data and not change it. Therefore, when used this way, they are considered safe. That is, they can be called without risk of data modification or corruption — calling it once has the same effect as calling it 10 times, or none at all. Additionally, GET (and HEAD) is idempotent, which means that making multiple identical requests ends up having the same result as a single request.

In this case, we could also consider using @post to avoid caching issues, but for this example I’ll leave it as @get.

The POST verb is most-often utilized to **create** new resources. In particular, it’s used to create subordinate resources. That is, subordinate to some other (e.g. parent) resource. In other words, when creating a new resource, POST to the parent and the service takes care of associating the new resource with the parent, assigning an ID (new resource URI), etc. On successful creation, return HTTP status 201, returning a Location header with a link to the newly-created resource with the 201 HTTP status. POST is neither safe nor idempotent. It is therefore recommended for non-idempotent resource requests. Making two identical POST requests will most-likely result in two resources containing the same information.

We can also customize the output. Keep in mind though, that the output should be “serialized”. By default, the output will be in JSON format. Here, I want to have a text output, so I’ll specify @html without html formatting specifications, although I could add them if I wanted to display the text on a website. If we were to [store the data in a database](https://www.rplumber.io/docs/runtime.html#external-data-store), however, this would not be a good idea. In that case, it would be better to output the result as a JSON object.

**Logging with filters**

It is also useful to provide some sort of logging for your API. Here, I am using the simple example from the [plumber documentation](https://www.rplumber.io/docs/routing-and-input.html#filters) that uses filters and output the logs to the console or your API server logs. In production, it would be better to use a real logging setup that stores information about each request, e.g. the time stamp, whether any errors or warnings occurred, etc. The forward() part of the logging function passes control on to the next handler in the pipeline, here our predict function.

**Running the plumber script**

We need to save the entire script with annotations as an *.R* file as seen below. The regular comments # describe what each section does.

# script name:

# plumber.R

# set API title and description to show up in http://localhost:8000/\_\_swagger\_\_/

#' @apiTitle Run predictions for Chronic Kidney Disease with Random Forest Model

#' @apiDescription This API takes as patient data on Chronic Kidney Disease and returns a prediction whether the lab values

#' indicate Chronic Kidney Disease (ckd) or not (notckd).

#' For details on how the model is built, see https://shirinsplayground.netlify.com/2017/12/lime\_sketchnotes/

#' For further explanations of this plumber function, see https://shirinsplayground.netlify.com/2018/01/plumber/

# load model

# this path would have to be adapted if you would deploy this

load("/Users/shiringlander/Documents/Github/shirinsplayground/data/model\_rf.RData")

#' Log system time, request method and HTTP user agent of the incoming request

#' @filter logger

function(req){

cat("System time:", as.character(Sys.time()), "\n",

"Request method:", req$REQUEST\_METHOD, req$PATH\_INFO, "\n",

"HTTP user agent:", req$HTTP\_USER\_AGENT, "@", req$REMOTE\_ADDR, "\n")

plumber::forward()

}

# core function follows below:

# define parameters with type and description

# name endpoint

# return output as html/text

# specify 200 (okay) return

#' predict Chronic Kidney Disease of test case with Random Forest model

#' @param age:numeric The age of the patient, numeric (scaled and centered to be btw 0 and 1)

#' @param bp:numeric The blood pressure of the patient, numeric (scaled and centered to be btw 0 and 1)

#' @param sg\_1.005:int The urinary specific gravity of the patient, integer (1: sg = 1.005, otherwise 0)

#' @param sg\_1.010:int The urinary specific gravity of the patient, integer (1: sg = 1.010, otherwise 0)

#' @param sg\_1.015:int The urinary specific gravity of the patient, integer (1: sg = 1.015, otherwise 0)

#' @param sg\_1.020:int The urinary specific gravity of the patient, integer (1: sg = 1.020, otherwise 0)

#' @param sg\_1.025:int The urinary specific gravity of the patient, integer (1: sg = 1.025, otherwise 0)

#' @param al\_0:int The urine albumin level of the patient, integer (1: al = 0, otherwise 0)

#' @param al\_1:int The urine albumin level of the patient, integer (1: al = 1, otherwise 0)

#' @param al\_2:int The urine albumin level of the patient, integer (1: al = 2, otherwise 0)

#' @param al\_3:int The urine albumin level of the patient, integer (1: al = 3, otherwise 0)

#' @param al\_4:int The urine albumin level of the patient, integer (1: al = 4, otherwise 0)

#' @param al\_5:int The urine albumin level of the patient, integer (1: al = 5, otherwise 0)

#' @param su\_0:int The sugar level of the patient, integer (1: su = 0, otherwise 0)

#' @param su\_1:int The sugar level of the patient, integer (1: su = 1, otherwise 0)

#' @param su\_2:int The sugar level of the patient, integer (1: su = 2, otherwise 0)

#' @param su\_3:int The sugar level of the patient, integer (1: su = 3, otherwise 0)

#' @param su\_4:int The sugar level of the patient, integer (1: su = 4, otherwise 0)

#' @param su\_5:int The sugar level of the patient, integer (1: su = 5, otherwise 0)

#' @param rbc\_normal:int The red blood cell count of the patient, integer (1: rbc = normal, otherwise 0)

#' @param rbc\_abnormal:int The red blood cell count of the patient, integer (1: rbc = abnormal, otherwise 0)

#' @param pc\_normal:int The pus cell level of the patient, integer (1: pc = normal, otherwise 0)

#' @param pc\_abnormal:int The pus cell level of the patient, integer (1: pc = abnormal, otherwise 0)

#' @param pcc\_present:int The puc cell clumps status of the patient, integer (1: pcc = present, otherwise 0)

#' @param pcc\_notpresent:int The puc cell clumps status of the patient, integer (1: pcc = notpresent, otherwise 0)

#' @param ba\_present:int The bacteria status of the patient, integer (1: ba = present, otherwise 0)

#' @param ba\_notpresent:int The bacteria status of the patient, integer (1: ba = notpresent, otherwise 0)

#' @param bgr:numeric The blood glucose random level of the patient, numeric (scaled and centered to be btw 0 and 1)

#' @param bu:numeric The blood urea level of the patient, numeric (scaled and centered to be btw 0 and 1)

#' @param sc:numeric The serum creatinine level of the patient, numeric (scaled and centered to be btw 0 and 1)

#' @param sod:numeric The sodium level of the patient, numeric (scaled and centered to be btw 0 and 1)

#' @param pot:numeric The potassium level of the patient, numeric (scaled and centered to be btw 0 and 1)

#' @param hemo:numeric The hemoglobin level of the patient, numeric (scaled and centered to be btw 0 and 1)

#' @param pcv:numeric The packed cell volume of the patient, numeric (scaled and centered to be btw 0 and 1)

#' @param wbcc:numeric The white blood cell count of the patient, numeric (scaled and centered to be btw 0 and 1)

#' @param rbcc:numeric The red blood cell count of the patient, numeric (scaled and centered to be btw 0 and 1)

#' @param htn\_yes:int The hypertension status of the patient, integer (1: htn = yes, otherwise 0)

#' @param htn\_no:int The hypertension status of the patient, integer (1: htn = no, otherwise 0)

#' @param dm\_yes:int The diabetes mellitus status of the patient, integer (1: dm = yes, otherwise 0)

#' @param dm\_no:int The diabetes mellitus status of the patient, integer (1: dm = no, otherwise 0)

#' @param cad\_yes:int The coronary artery disease status of the patient, integer (1: cad = yes, otherwise 0)

#' @param cad\_no:int The coronary artery disease status of the patient, integer (1: cad = no, otherwise 0)

#' @param appet\_good:int The appetite of the patient, integer (1: appet = good, otherwise 0)

#' @param appet\_poor:int The appetite of the patient, integer (1: appet = poor, otherwise 0)

#' @param pe\_yes:int The pedal edema status of the patient, integer (1: pe = yes, otherwise 0)

#' @param pe\_no:int The pedal edema status of the patient, integer (1: pe = no, otherwise 0)

#' @param ane\_yes:int The anemia status of the patient, integer (1: ane = yes, otherwise 0)

#' @param ane\_no:int The anemia status of the patient, integer (1: ane = no, otherwise 0)

#' @get /predict

#' @html

#' @response 200 Returns the class (ckd or notckd) prediction from the Random Forest model; ckd = Chronic Kidney Disease

calculate\_prediction <- function(age, bp, sg\_1.005, sg\_1.010, sg\_1.015, sg\_1.020, sg\_1.025, al\_0, al\_1, al\_2,

al\_3, al\_4, al\_5, su\_0, su\_1, su\_2, su\_3, su\_4, su\_5, rbc\_normal, rbc\_abnormal, pc\_normal, pc\_abnormal,

pcc\_present, pcc\_notpresent, ba\_present, ba\_notpresent, bgr, bu, sc, sod, pot, hemo, pcv,

wbcc, rbcc, htn\_yes, htn\_no, dm\_yes, dm\_no, cad\_yes, cad\_no, appet\_good, appet\_poor, pe\_yes, pe\_no,

ane\_yes, ane\_no) {

# make data frame from numeric parameters

input\_data\_num <<- data.frame(age, bp, bgr, bu, sc, sod, pot, hemo, pcv, wbcc, rbcc,

stringsAsFactors = FALSE)

# and make sure they really are numeric

input\_data\_num <<- as.data.frame(t(sapply(input\_data\_num, as.numeric)))

# make data frame from (binary) integer parameters

input\_data\_int <<- data.frame(sg\_1.005, sg\_1.010, sg\_1.015, sg\_1.020, sg\_1.025, al\_0, al\_1, al\_2,

al\_3, al\_4, al\_5, su\_0, su\_1, su\_2, su\_3, su\_4, su\_5, rbc\_normal, rbc\_abnormal, pc\_normal, pc\_abnormal,

pcc\_present, pcc\_notpresent, ba\_present, ba\_notpresent, htn\_yes, htn\_no, dm\_yes, dm\_no,

cad\_yes, cad\_no, appet\_good, appet\_poor, pe\_yes, pe\_no, ane\_yes, ane\_no,

stringsAsFactors = FALSE)

# and make sure they really are numeric

input\_data\_int <<- as.data.frame(t(sapply(input\_data\_int, as.integer)))

# combine into one data frame

input\_data <<- as.data.frame(cbind(input\_data\_num, input\_data\_int))

# validation for parameter

if (any(is.na(input\_data))) {

res$status <- 400

res$body <- "Parameters have to be numeric or integers"

}

if (any(input\_data < 0) || any(input\_data > 1)) {

res$status <- 400

res$body <- "Parameters have to be between 0 and 1"

}

# predict and return result

pred\_rf <<- predict(model\_rf, input\_data)

paste("----------------\nTest case predicted to be", as.character(pred\_rf), "\n----------------\n")

}

Note that I am using the “double-assignment” operator <<- in my function, because I want to make sure that objects are overwritten at the top level (i.e. globally). This would have been relevant had I set a global parameter, but to show it the example, I decided to use it here as well.

We can now call our script with the plumb() function, run it with run() and open it on port 800. Calling plumb() creates an environment in which all our functions are evaluated.

library(plumber)

r <- plumb("/Users/shiringlander/Documents/Github/shirinsplayground/static/scripts/plumber.R")

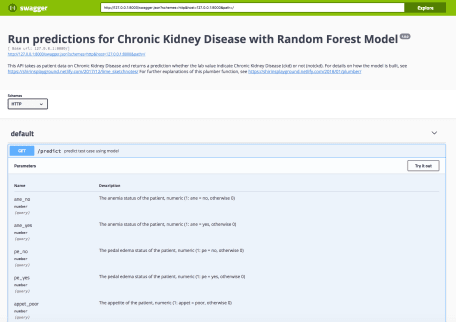
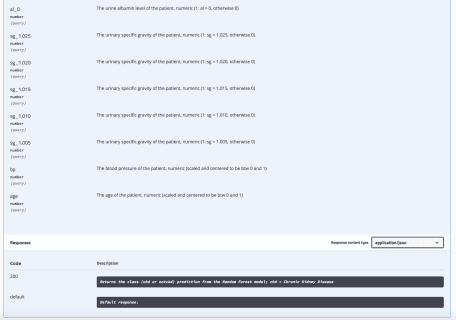
r$run(port = 8000)

We will now see the following message in our R console:

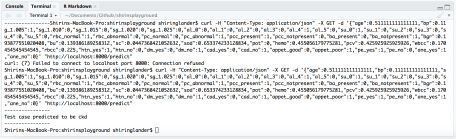
Starting server to listen on port 8000

Running the swagger UI at http://127.0.0.1:8000/\_\_swagger\_\_/

If you go to \*[http://localhost:8000/\_\_swagger\_\_/\*](http://localhost:8000/__swagger__/*), you could now try out the function by manually choosing values for all the parameters we defined in the script.

… 

Because we annotated the calculate\_prediction() function in our script with #' @get /predict we can access it via \*[http://localhost:8000/predict\*](http://localhost:8000/predict*). But because we have no input specified as of yet, we will only see an error on this site. So, we still need to put our JSON formatted input into the function. To do this, we can use [*curl*](https://en.wikipedia.org/wiki/CURL) from the terminal and feed in the JSON string from above. If you are using RStudio in the latest version, you have a handy terminal window open in your working directory. You find it right next to the Console.



Terminal in RStudio

curl -H "Content-Type: application/json" -X GET -d '{"age":0.511111111111111,"bp":0.111111111111111,"sg\_1.005":1,"sg\_1.010":0,"sg\_1.015":0,"sg\_1.020":0,"sg\_1.025":0,"al\_0":0,"al\_1":0,"al\_2":0,"al\_3":0,"al\_4":1,"al\_5":0,"su\_0":1,"su\_1":0,"su\_2":0,"su\_3":0,"su\_4":0,"su\_5":0,"rbc\_normal":1,"rbc\_abnormal":0,"pc\_normal":0,"pc\_abnormal":1,"pcc\_present":1,"pcc\_notpresent":0,"ba\_present":0,"ba\_notpresent":1,"bgr":0.193877551020408,"bu":0.139386189258312,"sc":0.0447368421052632,"sod":0.653374233128834,"pot":0,"hemo":0.455056179775281,"pcv":0.425925925925926,"wbcc":0.170454545454545,"rbcc":0.225,"htn\_yes":1,"htn\_no":0,"dm\_yes":0,"dm\_no":1,"cad\_yes":0,"cad\_no":1,"appet\_good":0,"appet\_poor":1,"pe\_yes":1,"pe\_no":0,"ane\_yes":1,"ane\_no":0}' "http://localhost:8000/predict"

**-H** defines an extra header to include in the request when sending HTTP to a server (<https://curl.haxx.se/docs/manpage.html#-H>).

**-X** pecifies a custom request method to use when communicating with the HTTP server (<https://curl.haxx.se/docs/manpage.html#-X>).

**-d** sends the specified data in a request to the HTTP server, in the same way that a browser does when a user has filled in an HTML form and presses the submit button. This will cause curl to pass the data to the server using the content-type application/x-www-form-urlencoded (<https://curl.haxx.se/docs/manpage.html#-d>).

This will return the following output:

* cat() outputs to the R console if you use R interactively; if you use R on a server, it will be included in the server logs.

System time: 2018-01-15 13:34:32

Request method: GET /predict

HTTP user agent: curl/7.54.0 @ 127.0.0.1

* paste outputs to the terminal

----------------

Test case predicted to be ckd

----------------

**Security**

This example shows a pretty simply R-script API. But if you plan on deploying your API to production, you should consider the [security section of the plumber documentation](https://www.rplumber.io/docs/security.html). It give additional information about how you can make your code (more) secure.

**Finalize**

If you wanted to deploy this API you would need to host it, i.e. provide the model and run an R environment with plumber, ideally on a server. A good way to do this, would be to package everything in a Docker container and run this. Docker will ensure that you have a working snapshot of the system settings, R and package versions that won’t change. For more information on dockerizing your API, check out below:

Plumber allows you to create a web API by merely decorating your existing R source code with special comments. Take a look at an example.

# plumber.R

#\* Echo back the input

#\* @param msg The message to echo

#\* @get /echo

function(msg=""){

list(msg = paste0("The message is: '", msg, "'"))

}

#\* Plot a histogram

#\* @png

#\* @get /plot

function(){

rand <- rnorm(100)

hist(rand)

}

#\* Return the sum of two numbers

#\* @param a The first number to add

#\* @param b The second number to add

#\* @post /sum

function(a, b){

as.numeric(a) + as.numeric(b)

}

These comments allow plumber to make your R functions available as API endpoints. You can use either #\* as the prefix or #', but we recommend the former since #' will collide with Roxygen.

> library(plumber)

> r <- plumb("plumber.R") # Where 'plumber.R' is the location of the file shown above

> r$run(port=8000)

You can visit this URL using a browser or a terminal to run your R function and get the results. For instance <http://localhost:8000/plot> will show you a histogram, and <http://localhost:8000/echo?msg=hello> will echo back the 'hello' message you provided.

Here we're using curl via a Mac/Linux terminal.

$ curl "http://localhost:8000/echo"

{"msg":["The message is: ''"]}

$ curl "http://localhost:8000/echo?msg=hello"

{"msg":["The message is: 'hello'"]}

As you might have guessed, the request's query string parameters are forwarded to the R function as arguments (as character strings).

$ curl --data "a=4&b=3" "http://localhost:8000/sum"

[7]

You can also send your data as JSON:

$ curl --data '{"a":4, "b":5}' http://localhost:8000/sum

[9]

## Installation

You can install the latest stable version from CRAN using the following command:

install.packages("plumber")

If you want to try out the latest development version, you can install it from GitHub. The easiest way to do that is by using devtools.

library(plumber)

sessionInfo()

## R version 3.4.2 (2017-09-28)

## Platform: x86\_64-apple-darwin15.6.0 (64-bit)

## Running under: macOS High Sierra 10.13.2

##

## Matrix products: default

## BLAS: /Library/Frameworks/R.framework/Versions/3.4/Resources/lib/libRblas.0.dylib

## LAPACK: /Library/Frameworks/R.framework/Versions/3.4/Resources/lib/libRlapack.dylib

##

## locale:

## [1] de\_DE.UTF-8/de\_DE.UTF-8/de\_DE.UTF-8/C/de\_DE.UTF-8/de\_DE.UTF-8

##

## attached base packages:

## [1] methods stats graphics grDevices utils datasets base

##

## other attached packages:

## [1] plumber\_0.4.4 rjson\_0.2.15 forcats\_0.2.0 stringr\_1.2.0

## [5] dplyr\_0.7.4 purrr\_0.2.4 readr\_1.1.1 tidyr\_0.7.2

## [9] tibble\_1.3.4 ggplot2\_2.2.1 tidyverse\_1.2.1

##

## loaded via a namespace (and not attached):

## [1] httr\_1.3.1 ddalpha\_1.3.1 sfsmisc\_1.1-1

## [4] jsonlite\_1.5 splines\_3.4.2 foreach\_1.4.3

## [7] prodlim\_1.6.1 modelr\_0.1.1 assertthat\_0.2.0

## [10] stats4\_3.4.2 DRR\_0.0.2 cellranger\_1.1.0

## [13] yaml\_2.1.15 robustbase\_0.92-8 ipred\_0.9-6

## [16] backports\_1.1.1 lattice\_0.20-35 glue\_1.2.0

## [19] digest\_0.6.12 randomForest\_4.6-12 rvest\_0.3.2

## [22] colorspace\_1.3-2 recipes\_0.1.1 httpuv\_1.3.5

## [25] htmltools\_0.3.6 Matrix\_1.2-12 plyr\_1.8.4

## [28] psych\_1.7.8 timeDate\_3042.101 pkgconfig\_2.0.1

## [31] CVST\_0.2-1 broom\_0.4.3 haven\_1.1.0

## [34] caret\_6.0-77 bookdown\_0.5 scales\_0.5.0

## [37] gower\_0.1.2 lava\_1.5.1 withr\_2.1.0

## [40] nnet\_7.3-12 lazyeval\_0.2.1 cli\_1.0.0

## [43] mnormt\_1.5-5 survival\_2.41-3 magrittr\_1.5

## [46] crayon\_1.3.4 readxl\_1.0.0 evaluate\_0.10.1

## [49] nlme\_3.1-131 MASS\_7.3-47 xml2\_1.1.1

## [52] dimRed\_0.1.0 foreign\_0.8-69 class\_7.3-14

## [55] blogdown\_0.3 tools\_3.4.2 hms\_0.4.0

## [58] kernlab\_0.9-25 munsell\_0.4.3 bindrcpp\_0.2

## [61] compiler\_3.4.2 RcppRoll\_0.2.2 rlang\_0.1.4

## [64] grid\_3.4.2 iterators\_1.0.8 rstudioapi\_0.7

## [67] rmarkdown\_1.8 gtable\_0.2.0 ModelMetrics\_1.1.0

## [70] codetools\_0.2-15 reshape2\_1.4.2 R6\_2.2.2

## [73] lubridate\_1.7.1 knitr\_1.17 bindr\_0.1

## [76] rprojroot\_1.2 stringi\_1.1.6 parallel\_3.4.2

## [79] Rcpp\_0.12.14 rpart\_4.1-11 tidyselect\_0.2.3

## [82] DEoptimR\_1.0-8