MWAS code structure:

mwas_analysis.r —— command line function, includes "train", "predict", "plot", "statistics" options.

Dependencies (required packages):
1. optparse   – option parsing
2. biom   – biom format file operation
3. e1071   – classifier package (SVM, cross-validation…)
4. randomForest – random forest
5. glmnet   – for multi-normial regression
6. pROC   – ROC analysis
7. caret   – variable importance analysis for feature selection
8. xlsx   – for Excel workbook output
9. beeswarm   – beeswarm plot
10. kernlab – future version will add this function

Input Option list
Preprocessing: (universal options for all modes)
      -t   transform_type [default: none]
      -r   suppress_relative_abundance_conversion[default: FALSE]
      -v   min_prevalence [default: 0]
      -e   collapse_table [default: FALSE]

Training:
Rscript mwas_analysis.r –w train
      -i   otu_fp [required]
      -m   map_fp [required]
      -c   category [required]
      -o   out_dir [default: "."]
      -M   method (classifier type) [default: RF]
      -C   classifier_parameter [optional]
      -a   validation_type [default: cv]
      -n   number_of_fold_cv [default: 10]
      -f   is_feature_selection [default: FALSE]
      -s   feature_selection_paramter [default: 0.0]

Rscript bin/mwas_analysis.R -w train –i
test/data/GG_100nt_even10k-adults-s20.biom –m test/data/gg-
map-adults.txt –c COUNTRY -o example/train_out/ –M svm

Predicting:
Rscript mwas_analysis.r –w predict
      -i   otu_fp [required]
      -m   map_fp [optional]
      -c   category [required only if map_fp is provided]
      -o   out_dir [default: "."]
      -M   trained_model_fp (method) [required]

```
Rscript bin/mwas_analysis.R -w predict -i
test/data/GG_100nt_even10k-adults-s20.biom -m test/data/gg-
map-adults.txt -c COUNTRY -o example/predict_out/ -M
example/train_out/trained_model.rds
```

Plot: (all capital for plot-only parameters)
```
Rscript mwas_analysis.r -w plot
    -i      otu_fp [required]
    -o      out_dir [default: "."]
    -c      category [required]
    -m      map_fp [required]
    -F      feat_stats_fp (feature_parameter) [optional]
    -M      plot_type (method) [required]
    -D      distance_fp [optional]
    -P      pcoa_fp [optional]
    -T      which_taxa [optional]
    -S      shorten_taxa [default: FALSE]
    -X      multiple_axis [default: FALSE]
    -A      Maximum false discovery rate (alpha) [optional]
    -K      is.filter_kegg (for heatmap) [optional]
    -N      number of plots (beeswarm,scatterplot) [default: NULL]
```

Gradients plot
```
Rscript bin/mwas_analysis.R -w plot -M gradients -i
test/data/taxa/merged-taxa.txt -T Bacteroides,Prevotella -o
example/plot_out_gradient -S
```

Beeswarm with feature stats table
```
Rscript bin/mwas_analysis.R -w plot -M beeswarm -i
test/data/taxa/merged-taxa.txt -o example/beeswarm -m
test/data/gg-map-adults.txt -c COUNTRY -F
test/data/stats/taxon-stats-table.txt -A 0.01 -N 20
```

Beeswarm without feature stats table
```
Rscript bin/mwas_analysis.R -w plot -M beeswarm -i
test/data/taxa/merged-taxa.txt -o example/beeswarm -m
test/data/gg-map-adults.txt -c COUNTRY -A 0.1 -N 20 -S
```

Scatterplot is similar to the Beeswarm plots, except
-M scatterplot

Statistics:
```
Rscript mwas_analysis.r -w statistics
    -i      otu_fp
    -m      map_fp
    -c      category
    -o      out_dir
```

Function list

File name	Function names	Notes
import_data.R	import.train.params	
	import.predict.params	
	import.plot.params	
	import.stats.params	
util.load.R	read.qiime.table.mwas	
	read.qiime.classic.table	
	load.qiime.mapping.file	
	load.qiime.otu.table	
	otu.table.has.metadata	
	get.header.index	
	load.qiime.taxon.table	
	load.qiime.distance.matrix	
	load.qiime.pcoa.file	
	load.qiime.feature.stats	
	remove.nonoverlapping.samples	
	cluster.by.correlation	
	collapse.by.correlation	
	inspect.env	unused
	shorten.taxonomy	
	balanced.folds	unused
	is.outlier	unused
	factor.to.numeric	unused
	parse.params	updated for heatmap.mwas
model.train.R	train.mwas	
	persist.model.mwas	
	cross.validation.mwas	
model.evaluation.R	model.evaluation.mwas	
	predict.knn	S3 method
preprocess_mwas.R	preprocess.mwas	
	filter.pathways	
	get.next.kegg	
visualization.R	plot.mwas	
	diff.plot.parameters	
	run.beeswarm	
	run.2d.scatterplot	
	plot.gradients	
export_data.R	export.mwas	
	save.xlsx	
	write.differentiation.test.results	

stats.r	differentiation.test	
	linear.test	not used
	t.test.wrapper	not used
roc_mwas.R	roc.mwas	
gradients.R	show.gradients	
	show.metadata	
	get.gradient.ixs	
	best.legend.location	
features.selection.R	feature.scores.rf.mwas	
heatmap.3.R	heatmap.3	
heatmap.mwas.R	cluster.columns	
	cluster.rows	
	shorten.taxonomy	Similar to the one in util.load.R; commented out
	create.color.bars	
	heatmap.mwas	