1. Read X\_train file using read.table to create data frame
2. Read features file using read.table to create data frame
3. Read Activity\_labels using read.table to create data frame
4. Assign column names “activity\_id” & “activity\_desc” to the columns in Activity\_label file
5. Assign the features names to columns (observations) for training file
6. Read subject\_train file using read.table to create data frame
7. Read y\_train file using read.table to create data frame
8. Assign “subject” column name to subjects in subject\_Trai data frame
9. Assign “activity” column name to activity column to activities in training file
10. Column merge (X\_train,y\_train,subject\_train) data frames using cbind
11. Repeat steps #1 to # 10 for test files (X\_test,y\_test,subject\_test)
12. Combine the resultant Test and Train dataframes using rbind to combine the rows from Test and Train data frames
13. Extracts only the measurements on the mean and standard deviation for each measurement by Subsetting columns mean and standard deviation from the combined dataframe in step # 12
14. Uses descriptive activity names to name the activities in the data set by merging the acivity dataframe with dataframe in step # 14 to get the activity names
15. Creates a tidy dataset with the average of each variable for each activity and each subject the average of each variable for each activity and each subject
16. Write the tidy datset in “TidyDataset.txt”

1

ICPSR 9721 Page

CODEBOOK FOR RUNAWAY SLAVE HEIGHTS

FROM COLONIAL NEWSPAPER ADVERTISEMENTS

SPSS VARIABLE NAME VARIABLE VALUES OR

AND POSITION EXPLANATION

STATE 1-2 STATE OF NEWSPAPER IN USES STANDARD POST

WHICH ADVERTISEMENT WAS OFFICE STATE

LOCATED ABBREVIATIONS

YEAR 3-6 YEAR OF ADVERTISEMENT

LNAME 7-15 LAST NAME OF RUNAWAY

FNAME 22-30 FIRST NAME OF RUNAWAY

RACE 42 RACE OF RUNAWAY W = WHITE (NA)

B = BLACK

I = INDIAN

M = MULATTO

SEX 43 SEX OF RUNAWAY M = MALE

F = FEMALE

W = FEMALE

XVAR 44 EXTRA VARIABLE NA

AGE 45-46 AGE OF RUNAWAY AGE IN YEARS

FEET 49 HEIGHT OF RUNAWAY USED IN CALCULATION

IN FEET OF HEIGHT (IN

INCHES)

INCHES 50-53 HEIGHT OF RUNAWAY IN USED IN CALCULATION

INCHES OF HEIGHT (IN

INCHES)

BPLACE 56-61 PLACE OF BIRTH AS RECORDED, SEE

RECODE STATEMENTS

FOR SUMMARIES

IJ 64-65 LEGAL STATUS CO = CONVICT SERVANT

(NA)

IJ = IN JAIL AT TIME

OF

ADVERTISEMENT

1

Page 2 ICPSR

CODEBOOK FOR RUNAWAY SLAVE HEIGHTS

FROM COLONIAL NEWSPAPER ADVERTISEMENTS (CONTINUED)

[1]PROF [2]66 [3]PROFESSION [4]0 = MISSING DATA

[5]1 =[6] MENTION

2 = APPRENTICE

3 = SOLDIER

4 = SAILOR

5 = FARMER

6 = RUNAWAY FROM

SHIP

7 = THIEF

9 = MISSING PERSON

LANGUAGE 67 LANGUAGE (KNOWLEDGE 0 = MISSING DATA

OF ENGLISH) 1 = SPEAKS

2 = BROKEN/WITH

ACCENT

3 = LITTLE

4 = NONE

\*NA = NOT APPLICABLE

1

ICPSR 9721 Page

CODEBOOK FOR RUNAWAY SLAVE HEIGHT DESCRIPTORS

FROM COLONIAL NEWSPAPER ADVERTISEMENTS

SPSS VARIABLE NAME VARIABLE VALUES OR

AND POSITIONS EXPLANATION

STATE 1-2 STATE OF NEWSPAPER IN USES STANDARD POST

WHICH ADVERTISEMENT WAS OFFICE STATE

LOCATED ABBREVIATIONS

YEAR 3-6 YEAR OF ADVERTISEMENT

LNAME 7-21 LAST NAME OF RUNAWAY

FNAME 22-40 FIRST NAME OF RUNAWAY

RACE 42 RACE OF RUNAWAY W = WHITE (NA)

B = BLACK

I = INDIAN

M = MULATTO

SEX 43 SEX OF RUNAWAY M = MALE

F = FEMALE

W = FEMALE

AGE 45-46 AGE OF RUNAWAY AGE IN YEARS

FEET 49 HEIGHT OF RUNAWAY USED IN CALCULATION

IN FEET OF HEIGHT (IN

INCHES)

INCHES 50-53 HEIGHT OF RUNAWAY IN USED IN CALCULATION

INCHES OF HEIGHT (IN

INCHES)

PBIRTH 54-61 PLACE OF BIRTH AS RECORDED, SEE

RECODE STATEMENTS

FOR SUMMARIES

CO 64-65 LEGAL STATUS CO = CONVICT SERVANT

(NA)

IJ = IN JAIL AT TIME

OF

ADVERTISEMENT

1

Page 4 ICPSR

CODEBOOK FOR RUNAWAY SLAVE HEIGHT DESCRIPTORS

FROM COLONIAL NEWSPAPER ADVERTISEMENTS (CONTINUED)

P 66 PROFESSION 0 = MISSING DATA

1 = MENTION

2 = APPRENTICE

3 = SOLDIER

4 = SAILOR

5 = FARMER

6 = RUNAWAY FROM

SHIP

7 = THIEF

9 = MISSING PERSON

L 67 LANGUAGE (KNOWLEDGE 0 = MISSING DATA

OF ENGLISH) 1 = SPEAKS

2 = BROKEN/WITH

ACCENT

3 = LITTLE

4 = NONE

HEIGHT 68-73 DESCRIPTOR OF HEIGHT SEE RECODES FOR

OF RUNAWAY SLAVE SUMMARIES

WEIGHT 74-81 DESCRIPTOR OF WEIGHT SEE RECODES FOR

OF RUNAWAY SLAVE SUMMARIES

\*NA = NOT APPLICABLE

1

ICPSR 9721 Page

CODEBOOK FOR RUNAWAY INDENTURED SERVANT HEIGHTS

FROM COLONIAL NEWSPAPER ADVERTISEMENTS

SPSS VARIABLE NAME VARIABLE VALUES OR AND

POSITION EXPLANATION

STATE 1-2 STATE OF NEWSPAPER IN USES STANDARD POST

WHICH ADVERTISEMENT WAS OFFICE STATE

LOCATED ABBREVIATIONS

YEAR 3-6 YEAR OF ADVERTISEMENT

LNAME 7-15 LAST NAME OF RUNAWAY

FNAME 22-30 FIRST NAME OF RUNAWAY

RACE 42 RACE OF RUNAWAY W = WHITE

B = BLACK

I = INDIAN

M = MULATTO

SEX 43 SEX OF RUNAWAY M = MALE

F = FEMALE

W = FEMALE

XVAR 44 EXTRA VARIABLE NA

AGE 45-46 AGE OF RUNAWAY AGE IN YEARS

FEET 49 HEIGHT OF RUNAWAY USED IN CALCULATION

IN FEET OF HEIGHT (IN

INCHES)

INCHES 50-53 HEIGHT OF RUNAWAY IN USED IN CALCULATION

INCHES OF HEIGHT (IN

INCHES)

BPLACE 54-61 PLACE OF BIRTH AS RECORDED, SEE

RECODE STATEMENTS

FOR SUMMARIES

IJ 64-65 LEGAL STATUS CO = CONVICT SERVANT

IJ = IN JAIL AT TIME

OF

ADVERTISEMENT

1

Page 6 ICPSR

CODEBOOK FOR RUNAWAY INDENTURED SERVANT HEIGHTS

FROM COLONIAL NEWSPAPER ADVERTISEMENTS (CONTINUED)

PROF 66 PROFESSION 0 = MISSING DATA

1 = MENTION

2 = APPRENTICE

3 = SOLDIER

4 = SAILOR

5 = FARMER

6 = RUNAWAY FROM

SHIP

7 = THIEF

9= MISSING PERSON

LANGUAGE 67 LANGUAGE (KNOWLEDGE 0 = MISSING DATA

OF ENGLISH) 1 = SPEAKS

2 = BROKEN/WITH

ACCENT

3 = LITTLE

4 = NONE

1

ICPSR 9721 Page

CODEBOOK FOR RUNAWAY INDENTURED SERVANT HEIGHT DESCRIPTORS

FROM COLONIAL NEWSPAPER ADVERTISEMENTS

SPSS VARIABLE NAME VARIABLE VALUES OR AND

POSITION EXPLANATION

STATE 1-2 STATE OF NEWSPAPER IN USES STANDARD POST

WHICH ADVERTISEMENT WAS OFFICE STATE

LOCATED ABBREVIATIONS

YEAR 3-6 YEAR OF ADVERTISEMENT

LNAME 7-21 LAST NAME OF RUNAWAY

FNAME 22-40 FIRST NAME OF RUNAWAY

RACE 42 RACE OF RUNAWAY W = WHITE

B = BLACK

I = INDIAN

M = MULATTO

SEX 43 SEX OF RUNAWAY M = MALE

F = FEMALE

W = FEMALE

AGE 45-46 AGE OF RUNAWAY AGE IN YEARS

FEET 49 HEIGHT OF RUNAWAY USED IN CALCULATION

IN FEET OF HEIGHT (IN

INCHES)

ACT-INCHES HEIGHT OF RUNAWAY IN USED IN CALCULATION

50-53 INCHES OF HEIGHT (IN

INCHES)

PBIRTH 54-61 PLACE OF BIRTH AS RECORDED, SEE

RECODE STATEMENTS

FOR SUMMARIES

CO 64 -6 5 LEGAL STATUS CO = CONVICT SERVANT

IJ = IN JAIL AT TIME

OF

ADVERTISEMENT

1

Page 8 ICPSR

CODEBOOK FOR RUNAWAY INDENTURED SERVANT HEIGHT DESCRIPTORS

FROM COLONIAL NEWSPAPER ADVERTISEMENTS (CONTINUED)

P 66 PROFESSION 0 = MISSING DATA

1 = MENTION

2 = APPRENTICE

3 = SOLDIER

4 = SAILOR

5 = FARMER

6 = RUNAWAY FROM

SHIP

7 = THIEF

9 = MISSING PERSON

L 67 LANGUAGE (KNOWLEDGE 0 = MISSING DATA

OF ENGLISH) 1 = SPEAKS

2 = BROKEN/WITH

ACCENT

3 = LITTLE

4 = NONE

HEIGHT 68-73 DESCRIPTOR OF HEIGHT SEE RECODES FOR

OF RUNAWAY SERVANT SUMMARIES

WEIGHT 74-81 DESCRIPTOR OF WEIGHT SEE RECODES FOR

OF RUNAWAY SERVANT SUMMARIES

Code

run\_analysis{

# Read X\_train

train.file <- read.table("C:/Users/srinivas/Documents/R/getdata\_projectfiles\_UCI HAR Dataset/UCI HAR Dataset/train/X\_train.txt", header = FALSE, sep = "")

# Read features file

feat.file <- read.table("C:/Users/srinivas/Documents/R/getdata\_projectfiles\_UCI HAR Dataset/UCI HAR Dataset/features.txt", header = FALSE, sep = "")

# Read Activity file

act.file <- read.table("C:/Users/srinivas/Documents/R/getdata\_projectfiles\_UCI HAR Dataset/UCI HAR Dataset/activity\_labels.txt", header = FALSE, sep = "")

# Question 43:Appropriately labels the data set with descriptive variable namesAssign names from features file to training file

names(train.file) <- feat.file[,2]

# Read subject train file

subjtrain.file <- read.table("C:/Users/srinivas/Documents/R/getdata\_projectfiles\_UCI HAR Dataset/UCI HAR Dataset/train/subject\_train.txt", header = FALSE, sep = "")

#Read Y\_train

ytrain.file <- read.table("C:/Users/srinivas/Documents/R/getdata\_projectfiles\_UCI HAR Dataset/UCI HAR Dataset/train/y\_train.txt", header = FALSE, sep = "")

# Assign col namnes to subject file and train activity file

names(subjtrain.file) <- "subject"

names(ytrain.file) <- "activity"

# Add columns Subject and Activity to the training file

train.file1 <- cbind(train.file,subjtrain.file,ytrain.file)

#Read X\_test file

test.file <- read.table("C:/Users/srinivas/Documents/R/getdata\_projectfiles\_UCI HAR Dataset/UCI HAR Dataset/test/X\_test.txt", header = FALSE, sep = "")

#Read subject test file

subjtest.file <- read.table("C:/Users/srinivas/Documents/R/getdata\_projectfiles\_UCI HAR Dataset/UCI HAR Dataset/test/subject\_test.txt", header = FALSE, sep = "")

#Read Activity test file

ytest.file <- read.table("C:/Users/srinivas/Documents/R/getdata\_projectfiles\_UCI HAR Dataset/UCI HAR Dataset/test/y\_test.txt", header = FALSE, sep = "")

# Add columns Subject and Activity to the testing file

test.file1 <- cbind(test.file,subjtest.file,ytest.file)

# Question 1:Merges the training and the test sets to create one data setrbind to combine the rows of the training file and Test file

comb.df <- rbind(test.file1,train.file1)

# Question 2:Extracts only the measurements on the mean and standard deviation for each measurement.Subsetting columns mean and standard deviaiton

mscomb.df <- comb.df[,c(1,2,3,4,5,6,41,42,43,44,45,46,81,82,83,84,85,86,121,122,123,124,125,126,161,162,163,164,165,166,201,202,214,215,227,228,240,241,253,254,266,267,268,269,270,271,373,374,375,424,425,426,427,428,452,453,454,529,530,542,543,552,562,563

)]

# Question 3:Uses descriptive activity names to name the activities in the data set.merge to get the activity names

mergedata <- merge(mscomb.df, act.file, by.x = "activity", by.y = "activity\_id")

# Question 5:Creates a second, independent tidy data set with the average of each variable for each activity and each subjectthe average of each variable for each activity and each subject

aggrDF <- aggregate(. ~ subject+activity\_desc, mergedata,mean)

}