

Kenya Data - First Look and Cleaning

Loading Data

Load in the data with new column headings:

```
# Load the mapping table of new to old column names and make into a vector
ColumnNameData <- read.csv("NewColumnNameMapping.csv")
ColumnNameMap <- ColumnNameData$NewColName
names(ColumnNameMap) <- ColumnNameData$OldColName
#Make proposed names syntactically valid for R
ColumnNameMap <- make.names(ColumnNameMap, unique=TRUE)
#Load in data using new column names
Kenya <- read.csv("Addressing the -1 for DataKind - KENYA SURVEY - KMM UPDATES 22 JAN cleaned data - Sh
                header = FALSE, stringsAsFactors = FALSE, skip = 2,
                col.names = ColumnNameMap)

#Also add a function to make proper case
properCase <- function(x) {
  lower <- tolower(x)
  s <- strsplit(lower, " ")[[1]]
  paste(toupper(substring(s, 1,1)), substring(s, 2),
        sep="", collapse=" ")
}
```

Show the dimensions of the loaded data to make sure it was loaded ok (rows, columns)

```
dim(Kenya)
```

```
## [1] 205 397
```

Analysing breakdown of gender and age

For Q10 on sex, 1: Male, 2: Female

For Q11 on age, 1: <=30, 2: 31-41, 3: 41-50, 4: 57-70, 5: >70

```
table(Kenya$Q10.Sex)
```

```
##
##  1  2
## 156 49
```

```
table(Kenya$Q11.Age)
```

```
##
##  1  2  3  4  5
##  9 49 57 73 17
```

```
table(Kenya$Q10.Sex, Kenya$Q11.Age)
```

```
##
##      1  2  3  4  5
##    1  6 36 35 62 17
##    2  3 13 22 11  0
```

Standardising text fields (Q3, 4 and 7)

Convert the 'Cooperative Society' field, Question 3, to proper case then make a table of all values.

(Question: Are any of these similar spellings actually the same?)

```
Kenya$Q3.CoopSociety <- sapply(Kenya$Q3.CoopSociety, properCase)
table(Kenya$Q3.CoopSociety)
```

```
##
##      Barasendu      Barazendu      Barezendu
##          19          6          1
##      Barsendu      Chaiyat      Chaiyat/kolenge
##          1          33          2
##      Cheptabach      Chuiyat Does Not Belong To Any
##          10          1          1
##      Kapsiliboi      Kechem      Ketchem
##          9          6          5
##      Keteng      Kipsebwo      Kolenge
##          2          4          19
##      Kosiywo      Kosoiywo      Luguitany
##          1          31          6
##      Malot      Mosine      N/a
##          1          2          2
##      Nandi Cooperative      Nandi Farmers      Sarma
##          1          5          21
##      Sireet  Oep      Siret      Sokot
##          1          3          1
##      Temso      Ye
##          10          1
```

Convert the 'Village' field, Question 4, to proper case then make a table of all values.

(Question: Are any of these similarly-spelled villages actually the same?)

```
Kenya$Q4.Village <- sapply(Kenya$Q4.Village, properCase)
table(Kenya$Q4.Village)
```

```
##
##      Chemoru  Chepkaitit  Chepkiwen  Chepsire  Cheptabach  Cheptililik
##          8          1          7          2          12          2
##      Choimim  Choimin  Kabirer  Kaboch  Kabokwa  Kabooch
##          7          1          2          3          6          4
##      Kaborer  Kabote  Kamagei  Kamanang  Kamaram  Kamung'ei
##          1          3          1          4          1          1
```

```
##      Kamungei      Kapboch Kapchouriai Kapchuria Kapchuriai Kapkagaron
##          1          1          1          7          5          1
##      Kapkembur      Kapkilel      Kapkioo      Kapkorio      Kapkoros      Kaplamai
##          4          1          1          1          2          1
##      Kaplamaywo      Kapsila      Kapsile      Kapsilei      Kapsiliboi      Kaptien
##          1          1          1          1          2          5
##      Kaptoroi      Keburo      Keteng      Kiminda      Kipkoiyo      Kipsebwo
##          2          1          15          2          2          7
##      Kipsigak      Kipsikak      Kiptemuria      Koilel      Kokamei      Kong'asis
##          2          2          1          1          1          3
##      Lelwak      Luguityany      Meswo      Mosine      Mosobejo      Mosobesho
##          2          1          4          12          2          1
##      Mosombor      Nduroto      Nduroton      Nduruto      Olotwo      Samoei
##          3          4          1          1          1          2
##      Sarma      Siksik      Singorwa      Singorwo      Sosio      Sosion
##          1          7          1          1          3          1
##      Soson      Taboiyat      Tachasis      Taito      Taretmoi      Temso
##          1          1          1          1          1          6
##      Tendwet      Tenwet      Tururo      Urhf
##          1          1          1          1
```

Convert the 'Producer organisation' field, Question 7, to proper case then make a table of all values.

(Question: Are any of these similarly-spelled orgs actually the same?)

```
Kenya$Q7.ProducerOrg <- sapply(Kenya$Q7.ProducerOrg, properCase)
table(Kenya$Q7.ProducerOrg)
```

```
##
##      Chemomi      Epk      Hfhf      None      Sireet Sireet Eop
##          1          2          1          1          47          1
## Sireet Oep      Siret
##      149          3
```

Columns with only one value

Identify which columns only have one unique value

```
names(Filter(function(x)(length(unique(x))<2), Kenya))
```

```
## [1] "Filter"
## [2] "SbjNam"
## [3] "UsrUnq"
## [4] "SubjData"
## [5] "RvwTime"
## [6] "RvwComment"
## [7] "SrvyrComment"
## [8] "Complete"
## [9] "StopQ"
## [10] "Test"
## [11] "ParentID"
## [12] "Status"
```

```
## [13] "QAScore"
## [14] "FrScName"
## [15] "ExReNum"
## [16] "RvwName"
## [17] "Q8.Consent"
## [18] "Q13.Role.04"
## [19] "Q30.Other.Text"
## [20] "Q43.MainFarmingChallenges.07"
## [21] "Q48.HarvestRice"
## [22] "Q50.HarvestPlantains"
## [23] "Q52.HarvestMango"
## [24] "Q73.AmountEarnedGoats"
## [25] "Q74.NumberGoatsSold"
## [26] "Q75.PxPerGoat"
## [27] "Q76.TotalEarningGoats"
## [28] "Q89.MainSourcesIncome.HydroponicsIncome"
## [29] "Q89.MainSourcesIncome.HandicraftIncome"
## [30] "Q89.MainSourcesIncome.RemitIncome"
## [31] "Q89.MainSourcesIncome.PigIncome"
## [32] "Q92.BankAccountHowOften"
## [33] "Q93.BankAccountChallenges"
## [34] "Q100.OtherFinancingFacilities"
## [35] "Q101.NameOtherFinancingFacilities"
## [36] "Q102.OtherHowOften"
## [37] "Q103.OtherChallenges"
## [38] "Q105.GoodAgPracticeTrained.014"
## [39] "Q117.FromWhereLearned.07.Text"
## [40] "Q130.AgExperimentsFrequency.Text"
## [41] "Q132.FarmPromotorGuidanceFrequency.Text"
## [42] "Q133.CoopExtensionSupportFrequency.Text"
## [43] "Q135.RadioStationFrequency.Text"
## [44] "Q136.WeFarmFrequency.Text"
## [45] "Q137.MobileFrequency.Text"
## [46] "Q138.NewspaperFrequency.Text"
## [47] "Q139.TVFrequency.Text"
## [48] "Q142.LocalGroupFrequency.Text"
## [49] "Q143.TrainingSessionFrequency.Text"
## [50] "Q144.TrainingMaterialsFrequency.Text"
## [51] "Q145.MostValuableSources.014"
## [52] "Q165.ShareNewFarmingPractice.Other"
```

Duplicate columns

Check if Q155 values are identical to Q158 values

```
table(Kenya$Q155.HowLongUsingIdea == Kenya$Q158.HowLongUsingIdea)
```

```
##
## TRUE
## 99
```

Eliminate one field from the dataset since they are identical

```
Kenya$Q158.HowLongUsingIdea <- NULL
```

Cleaning free-text number fields (Q38-40, 61-3, 69-72, 77-9, 113, 118)

Note: Look into Q73-76 since column descriptions appear duplicative to 69-71

Show unique values in Q38:

```
unique(Kenya$Q38.CashCropEarnings)
```

```
## [1] "360000"      "60000"      "50000"      "117000"
## [5] "300000"      "75000"      "180000"     "15000"
## [9] "Ksh 70000"   "Ksh160,000" "Ksh48000"   "150000"
## [13] "Ksh 500,000" "Ksh 40,000" "144000"     "336,000"
## [17] "20100"       "240,000"    "160,000"    "160000"
## [21] "54000"       "124 000"    "350 000"    "87500"
## [25] "146000"      "124000"     "750000"     "350000"
## [29] "ff"          "125,000"    "200,000"    "211,200"
## [33] "ksh.250,000" "ksh. 138,000" "ksh.500,000" "ksh. 14000"
## [37] "ksh.11,250"  "ksh. 30,000"  "ksh.150,000" "ksh. 200,000"
## [41] "ksh.5000"    "ksh. 625,000" "ksh.60,000"  "ksh. 115,000"
## [45] "ksh.7000"    "240000"      "3000"        "ksh.75000"
## [49] "ksh. 300,000" "ksh.825,000"  "ksh.300,000" "ksh.175,000"
## [53] "ksh. 100,000" "ksh.75,000"   "ksh.176,000" "ksh. 50,000"
## [57] "ksh. 125,000" "ksh. 150,000" "ksh. 75,000"  "ksh.50,000"
## [61] "ksh.37,500"  "ksh.200,000"  "ksh.100,000"  "ksh. 37500"
## [65] "ksh.125,000" "ksh.120,000"  "ksh. 25000"   "ksh.450,000"
## [69] "90000"       "900000"      "120000"      "12000"
## [73] "270000"      "450000"      "600000"      "420000"
## [77] "30000"       "280,000"     "200000"      "84000"
## [81] "120,000"     "210000"      "34800"       "18000"
## [85] "240 000"     "ksh.1000,000" "863000"      "573000"
## [89] "108000"      "37000"       "9000"        "330000"
## [93] "14400"       "220000"      "100,000"     "360,000"
## [97] "56,000"      "215080"      "10,000"      "14,000"
## [101] "180,000"     "32400"       "12960000 KSHS" "11400 KSHS"
## [105] "97200 KSHS"  "108000 KSHS" "257,143 kshs"  "180000 kshs"
## [109] "250000 kshs" "150000 kshs"  "161000 kshs"   "25000 kshs"
## [113] "260000 kshs" "50000 kshs"   "10000 kshs"    "127000 kshs"
## [117] "122000 kshs" "209000 kshs"  "230000 kshs"   "42000 kshs"
## [121] "103000 kshs" "142000 kshs"  "152000 kshs"   "15000 kshs"
## [125] "299000 kshs" "79000 kshs"   "14500 kshs"    "87000 kshs"
## [129] "60000 kshs"  "52800 kshs"   "157500 kshs"
```

Replace all non-digit characters with blank space, convert field to integer, then show unique values in Q38 and also plot histogram:

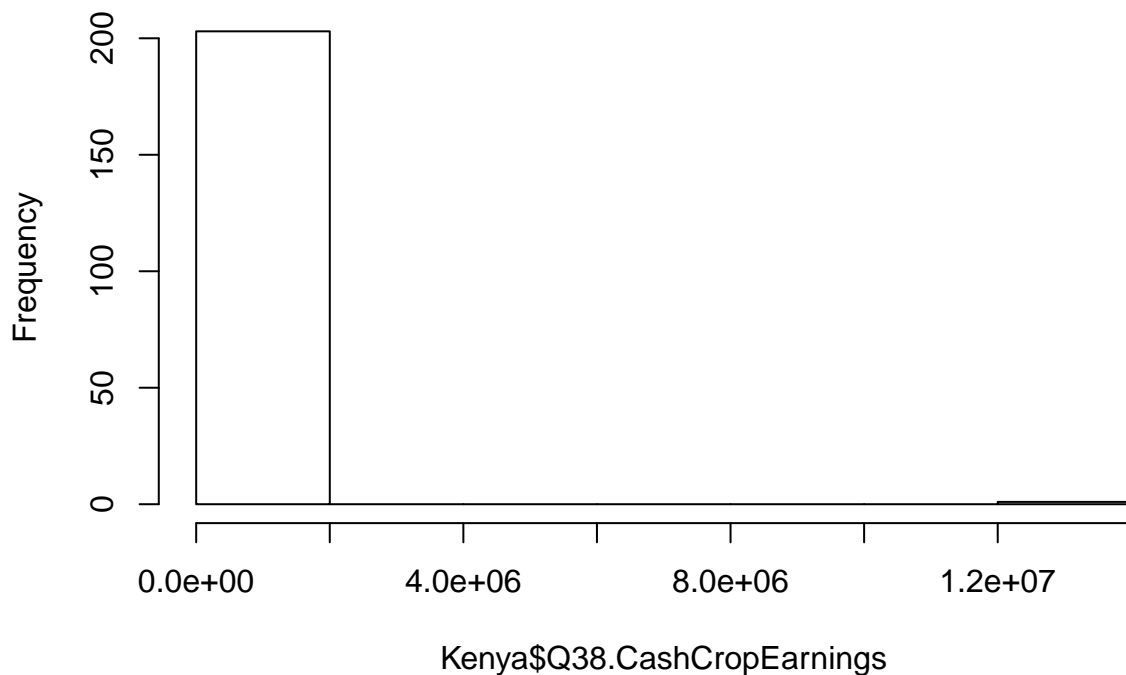
```
Kenya$Q38.CashCropEarnings <- as.integer(gsub("[^0-9]", "", Kenya$Q38.CashCropEarnings))
unique(Kenya$Q38.CashCropEarnings)
```

```
## [1] 360000 60000 50000 117000 300000 75000 180000
```

```
## [8] 15000 70000 160000 48000 150000 500000 40000
## [15] 144000 336000 20100 240000 54000 124000 350000
## [22] 87500 146000 750000 NA 125000 200000 211200
## [29] 250000 138000 14000 11250 30000 5000 625000
## [36] 115000 7000 3000 825000 175000 100000 176000
## [43] 37500 120000 25000 450000 90000 900000 12000
## [50] 270000 600000 420000 280000 84000 210000 34800
## [57] 18000 1000000 863000 573000 108000 37000 9000
## [64] 330000 14400 220000 56000 215080 10000 32400
## [71] 12960000 11400 97200 257143 161000 260000 127000
## [78] 122000 209000 230000 42000 103000 142000 152000
## [85] 299000 79000 14500 87000 52800 157500
```

```
hist(Kenya$Q38.CashCropEarnings)
```

Histogram of Kenya\$Q38.CashCropEarnings



Show unique values in Q39

```
unique(Kenya$Q39.CashCropPx)
```

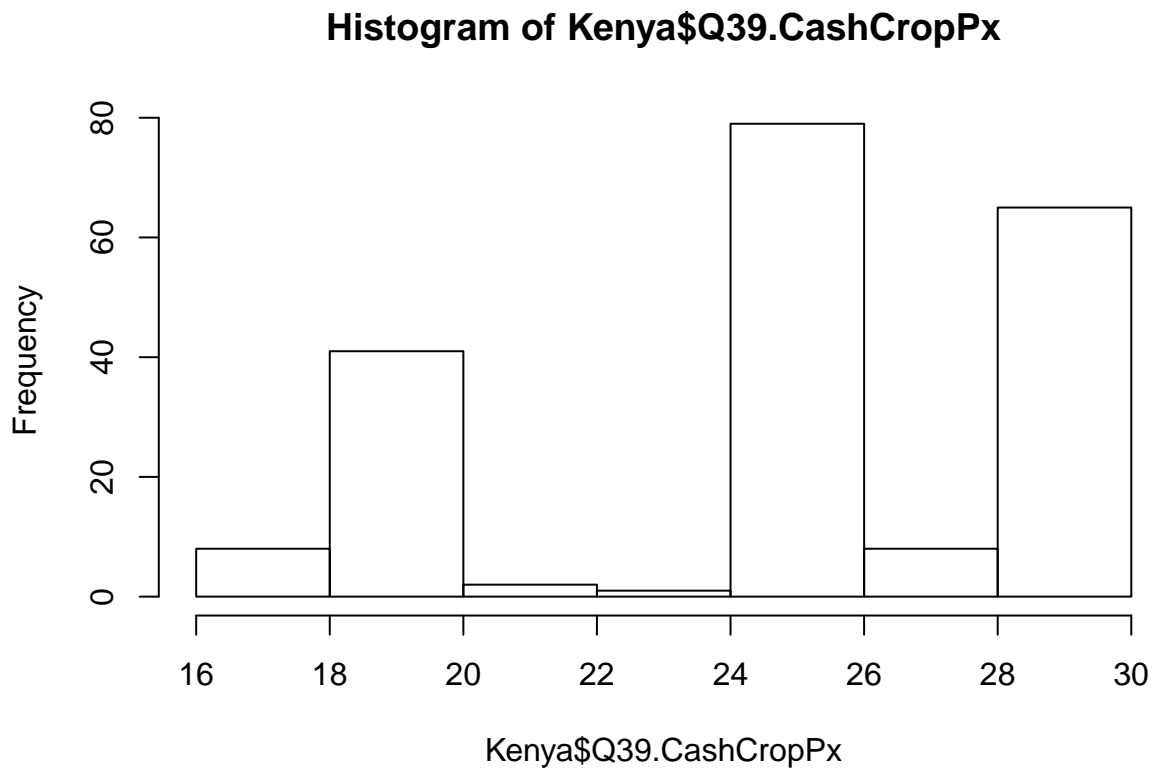
```
## [1] "30" "Ksh 20" "20" "27 ksh"
## [5] "27" "25 shilings" "25 per kg" "25 sh per kg"
## [9] "29 sh per kg" "ff" "18" "16"
## [13] "25" "29" "22" "26 sh"
## [17] "29 per kg" "ksh. 25" "ksh 25" "29 sh perkg"
## [21] "26 per kg" "25 perkg" "18 KSHS" "27 KSHS"
## [25] "29 kshs" "20 kshs" "18 kshs" "28 kshs"
## [29] "25 kshs" "27 kshs" "22 kshs" "22.50 kshs"
```

Replace all non-digit and non “.” characters with blank space, convert field to numeric, then show unique values in Q39 and also plot histogram

```
library(stringr)
Kenya$Q39.CashCropPx <- as.numeric(str_extract(Kenya$Q39.CashCropPx, "\\d{1,}\\.|\\.?(\\d{1,})?"))
unique(Kenya$Q39.CashCropPx)
```

```
## [1] 30.0 20.0 27.0 25.0 29.0 NA 18.0 16.0 22.0 26.0 28.0 22.5
```

```
hist(Kenya$Q39.CashCropPx)
```



Show unique values in Q40

```
unique(Kenya$Q40.CostOfProduction)
```

```
## [1] "50000" "20000" "23000" "40000"
## [5] "30000" "100000" "70000" "44000"
## [9] "Ksh 35000" "Ksh 30,000" "25000" "Ksh 280,000"
## [13] "15,000" "35,000" "3000" "10000"
## [17] "Ksh 90,000" "50 000 ksh" "80,000" "200,000"
## [21] "60 000" "75000" "78000" "72000"
## [25] "200000" "240000" "190000" "gg"
## [29] "59000" "113,200" "104,400" "70,000"
## [33] "ksh. 70,000" "ksh 56000" "ksh. 200,000" "ksh.200,000"
## [37] "ksh.3920" "ksh.4000" "ksh.15000" "ksh. 15000"
## [41] "ksh.60,000" "ksh. 120,000" "ksh. 2000" "ksh.300,000"
## [45] "ksh3000" "ksh,23,000" "ksh.44,200" "ksh.28,800"
## [49] "ksh.70,000" "ksh.3500" "ksh.30,000" "ksh.100,000"
## [53] "ksh. 175,000" "ksh.350,000" "ksh.175,000" "ksh,67,000"
## [57] "ksh80,000" "ksh.2950" "ksh.45000" "ksh. 35,000"
```

```
## [61] "ksh. 94,000" "ksh.26,000" "ksh.75,000" "ksh.45,000"
## [65] "ksh. 20,000" "ksh.20,000" "ksh.90,000" "ksh. 80,000"
## [69] "ksh.18,000" "ksh.35,000" "ksh. 40,000" "ksh.120,000"
## [73] "ksh.124,000" "ksh.190,000" "ksh.80,000" "250000"
## [77] "ksh. 246,000" "60000" "41000" "2800"
## [81] "90000" "27000" "280000" "148000"
## [85] "65000" "158000" "15000" "ksh.50,000"
## [89] "153,400" "135,000" "50,000" "102,000"
## [93] "85400" "127000" "120000" "400000"
## [97] "40800" "1000" "ksh,600000" "ksh. 150,000"
## [101] "500076" "300000" "203000" "21000"
## [105] "86000" "5000" "80000" "55000"
## [109] "44,000" "4800" "112000" "25,000"
## [113] "138,000" "40,000" "14,000" "130,000"
## [117] "9500" "6000" "Ksh 90000" "8000"
## [121] "8000000 KSHS" "6500 KSHS" "20000 KSHS" "10000 KSHS"
## [125] "116,069 kshs" "25000 kshs" "150000 kshs" "24000 kshs"
## [129] "30000 kshs" "5000 kshs" "50000 kahs" "15000 kshs"
## [133] "10000 kshs" "82000 kshs" "200000 kshs" "26000 kshs"
## [137] "79000 kshs" "45500 kshs" "17000 kshs" "32500 kshs"
## [141] "70000 kshs" "46000 kshs" "13600 kshs" "147000 kshs"
## [145] "55000 kshs" "5600 kshs" "129000 kshs"
```

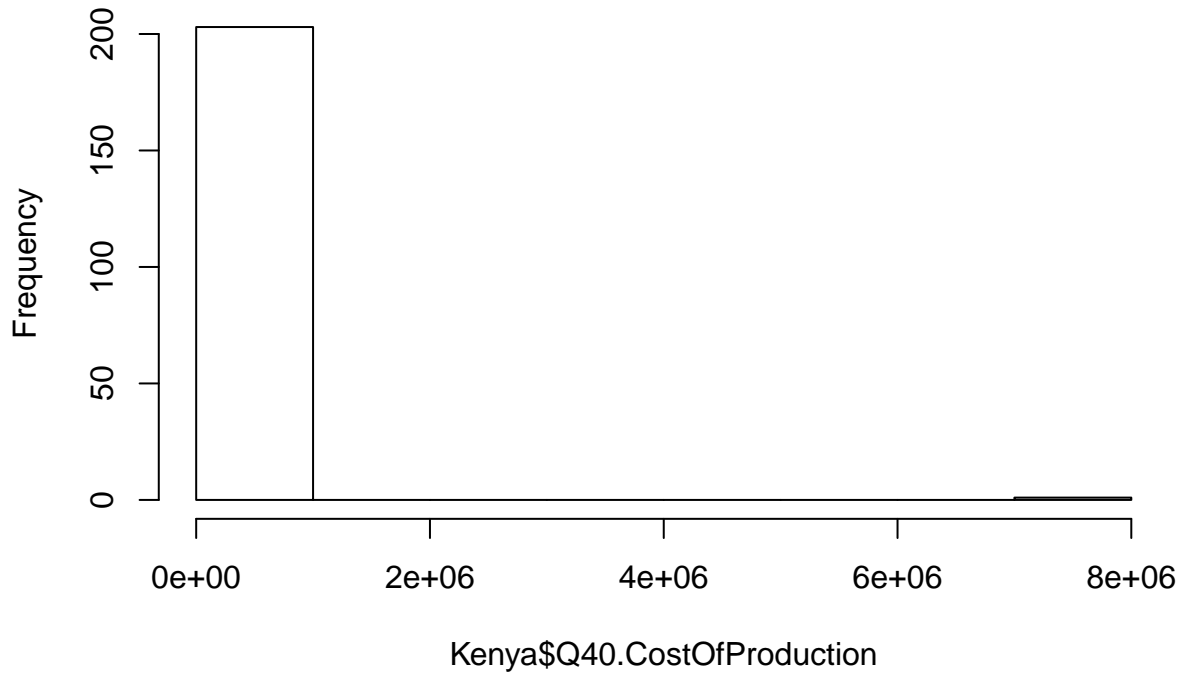
Replace all non-digit characters with blank space, convert field to integer, then show unique values in Q40 and also plot histogram:

```
Kenya$Q40.CostOfProduction <- as.integer(gsub("[^0-9]", "", Kenya$Q40.CostOfProduction))
unique(Kenya$Q40.CostOfProduction)
```

```
## [1] 50000 20000 23000 40000 30000 100000 70000 44000
## [9] 35000 25000 280000 15000 3000 10000 90000 80000
## [17] 200000 60000 75000 78000 72000 240000 190000 NA
## [25] 59000 113200 104400 56000 3920 4000 120000 2000
## [33] 300000 44200 28800 3500 175000 350000 67000 2950
## [41] 45000 94000 26000 18000 124000 250000 246000 41000
## [49] 2800 27000 148000 65000 158000 153400 135000 102000
## [57] 85400 127000 400000 40800 1000 600000 150000 500076
## [65] 203000 21000 86000 5000 55000 4800 112000 138000
## [73] 14000 130000 9500 6000 8000 8000000 6500 116069
## [81] 24000 82000 79000 45500 17000 32500 46000 13600
## [89] 147000 5600 129000
```

```
hist(Kenya$Q40.CostOfProduction)
```


Histogram of Kenya\$Q40.CostOfProduction



Show unique values in Q61, replace all non-digit characters with blank space, convert field to integer, then show resulting unique values

```
unique(Kenya$Q61.CropEarning)
```

```
## [1] "1000"      "3000"      NA
## [4] "2100"      "9000"      "14400"
## [7] "1800"      "0"         "8800"
## [10] "Ksh 150,000" "30000"     "60,000"
## [13] "100000"     "56000"     "not recorded"
## [16] "dd"         "3100"      "9700"
## [19] "12000"      "5000"      "ksh.10,125"
## [22] "None"       "ksh.2000"  "ksh. 3000"
## [25] "Ksh. 0"     "ksh. 0"    "ksh.0"
## [28] "ksh.12,000" "ksh.1000"  "ksh. 0"
## [31] "ksh.25,000" "ksh.5000"  "ksh.13,000"
## [34] "ksh.100,000" "ksh.7500"  "ksh.10,000"
## [37] "ksh.144,000" "family use" "0(use at home only)"
## [40] "family use only" "7000"      "84,000"
## [43] "for family use only" "5600"      "no sold as was food"
## [46] "1500"       "30000 Kshs" "13800 KSHS"
## [49] "57500 KSHS" "50000 KSHS" "11000 KSHS"
## [52] "1000 kshs"  "18000 kshs" "13000 kshs"
## [55] "36000 kshs" "47000 kshs" "3000 kshs"
## [58] "500 kshs"   "20000 kshs" "5500 kshs"
## [61] "2000 kshs"  "8000 kshs"  "700 kshs"
## [64] "3500 kshs"  "ksh.3000"  "31000 kshs"
```

```
Kenya$Q61.CropEarning <- as.integer(gsub("[^0-9]", "", Kenya$Q61.CropEarning))
unique(Kenya$Q61.CropEarning)
```

```
## [1] 1000 3000 NA 2100 9000 14400 1800 0 8800 150000
## [11] 30000 60000 100000 56000 3100 9700 12000 5000 10125 2000
## [21] 25000 13000 7500 10000 144000 7000 84000 5600 1500 13800
## [31] 57500 50000 11000 18000 36000 47000 500 20000 5500 8000
## [41] 700 3500 31000
```

Show unique values in Q62 Note: Q62 appears to be all in different units (sack, kg, bundle, punch, crate). Best to avoid this for analysis at the moment.

```
unique(Kenya$Q62.CropSalePx)
```

```
## [1] "200" NA
## [3] "700" "3000"
## [5] "180per tin" "300"
## [7] "0" "400"
## [9] "Ksh 1500 per Crate" "50"
## [11] "59" "40 per kg"
## [13] "not sold" "tt"
## [15] "ksh.200per kg" "ksh 45 per kg"
## [17] "ksh 2000 per sack" "ksh 1000 per bunch"
## [19] "ksh 75 per kg" "ksh 0"
## [21] "ksh.540" "ksh. 40 per kg"
## [23] "ksh.0" "N/A"
## [25] "ksh. 0" "ksh. 100"
## [27] "ksh.500 per sack" "ksh.12500 per sack"
## [29] "ksh.1000 per sack" "ksh.3200 per sack"
## [31] "ksh0" "ksh.300 per punch"
## [33] "ksh.75 per kg" "ksh.200 per punch"
## [35] "ksh.100 per kg" "ksh.20 per bundle of vegetables"
## [37] "20 per kilogram" "4000"
## [39] "60per tin" "100"
## [41] "70 Kshs" "100 KSHS PER KG"
## [43] "575 KSHS" "120 KSHS"
## [45] "3000 PER KG" "70 Per kg"
## [47] "180 kshs" "150 kshs"
## [49] "100 kshs" "5 kshs"
## [51] "750 kshs" "10 kshs"
## [53] "5kshs" "1300 kshs"
## [55] "535 kshs" "15 kshs"
## [57] "ksh.70" "400kshs"
## [59] "800 kshs"
```

Show unique values in Q63, replace all non-digit characters with blank space, convert field to integer, then show resulting unique values

```
unique(Kenya$Q63.ProductionCost)
```

```
## [1] "2400" "30000" NA "3000"
```

```
## [5] "7000"      "1000"      "5000"      "Ksh 300,000"
## [9] "4000"      "Ksh 15,000" "60000"     "10000"
## [13] "not recorded" "rr"        "1500"      "10,000"
## [17] "ksh 2500"   "ksh. 5000" "ksh. 2000" "ksh. 30,000"
## [21] "ksh. 1500"  "ksh. 3000" "ksh. 500"   "ksh.20,000"
## [25] "ksh. 10,000" "ksh. 4000" "ksh.1000"   "ksh. 0"
## [29] "ksh.5000"   "ksh. 30,000" "ksh.50,000" "ksh.25,000"
## [33] "ksh. 2500"  "ksh.2000"   "ksh.10,000" "ksh. 20,000"
## [37] "ksh.15,000" "ksh. 50,000" "ksh.4000"   "ksh. 5,000"
## [41] "ksh.30,000" "ksh.3000"   "ksh.70,000" "ks.30,000"
## [45] "ksh.40,000" "15000"      "1200"       "12000"
## [49] "1300"       "16000"      "500"        "0"
## [53] "ksh. 35,000" "5,000"      "2000"       "6000"
## [57] "20000"      "67000"      "Ksh 10,000" "13,000"
## [61] "5000 KSHS"  "4000 KSHS"  "2000 KSHS"  "10000 kshs"
## [65] "2000 kshs"  "5000 kshs"  "7500 kshs"  "10500 kshs"
## [69] "35000 kshs" "6500 kshs"  "27000 kshs" "7000 kshs"
## [73] "4000 kshs"  "3500 kss"   "8500 kshs"  "1200 kshs"
## [77] "1300 kshs"  "250 kshs"   "3000 kshs"  "4500 kshs"
## [81] "3300 kshs"  "3750 kshs"
```

```
Kenya$Q63.ProductionCost <- as.integer(gsub("[^0-9]", "", Kenya$Q63.ProductionCost))
unique(Kenya$Q63.ProductionCost)
```

```
## [1] 2400 30000 NA 3000 7000 1000 5000 300000 4000 15000
## [11] 60000 10000 1500 2500 2000 500 20000 0 50000 25000
## [21] 70000 40000 1200 12000 1300 16000 35000 6000 67000 13000
## [31] 7500 10500 6500 27000 3500 8500 250 4500 3300 3750
```

NOTE: Similar analyses and modifications are made for individual crop results in Questions 69-79

Show unique values in Q113

```
unique(Kenya$Q113.CostOfChange)
```

```
## [1] NA "0" "4000"
## [4] "1000" "ksh.45000" "ksh,50,000"
## [7] "ksh. 10,000" "ksh. 5000" "ksh. 3000"
## [10] "ksh. 200,000" "ksh.1000" "ksh.10,000"
## [13] "ksh.60,000" "ksh.30,000" "ksh.40,000"
## [16] "ksh.20,000" "ksh. 100,000" "ksh.62,000"
## [19] "50000" "13000" "ksh.2000"
## [22] "ksh.55,000" "32,000" "40000"
## [25] "0(i was given free)" "ksh.5000" "2000"
## [28] "50,000" "500" "150 kshs"
## [31] "5000 KSHS" "500 kshs" "58000 kshs"
```

Show unique values in Q118

```
unique(Kenya$Q118.CostStarting)
```

```
## [1] "100(1972)" NA "Ksh 120,000" "500"
```

```
## [5] "85,000"      "10000 ksh"    "ksh,84,000"   "ksh.3000"
## [9] "ksh. 2,800,000" "ksh. 5000"    "ksh.20,000"   "ksh.1000"
## [13] "15,000"      "20,000"      "5000"         "2000"
## [17] "2500"        "3000"        "3000 kshs"    "50000 kshs"
```

Longitude and Latitude values

First check how many values are present (not NA), then plot histogram showing distribution

```
length(which(!is.na(Kenya$Latitude)))
```

```
## [1] 142
```

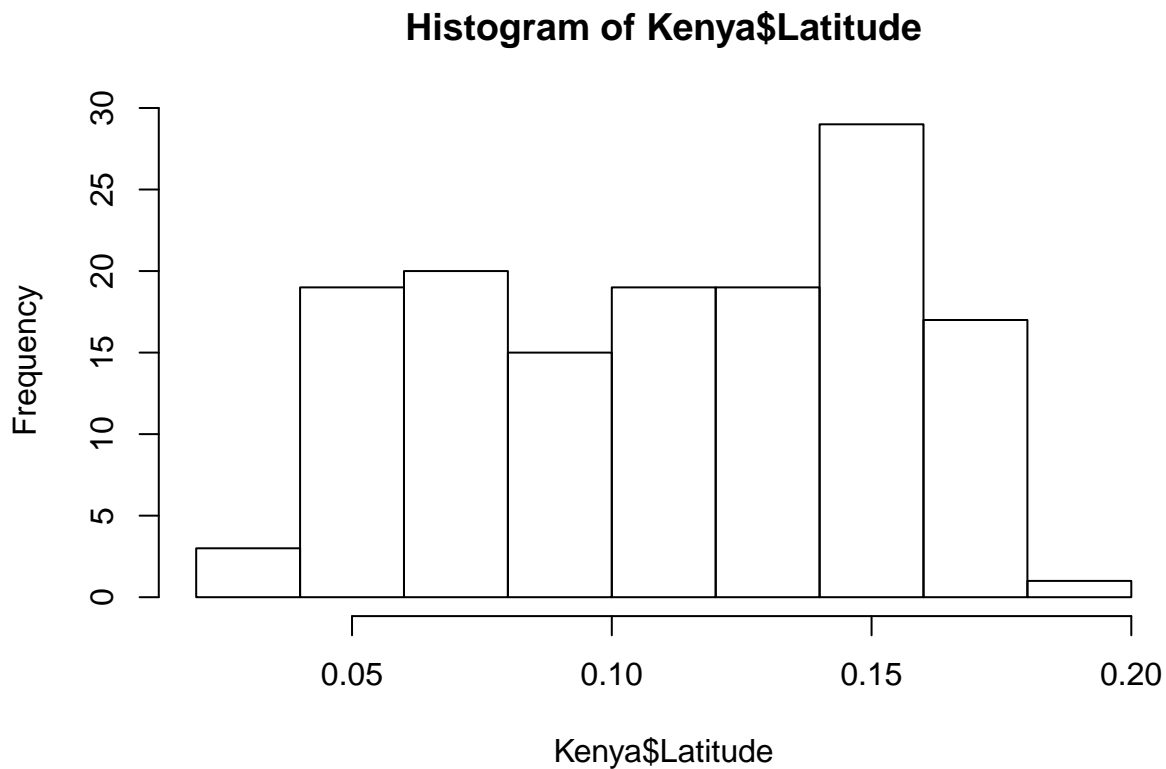
```
length(which(!is.na(Kenya$Longitude)))
```

```
## [1] 142
```

```
summary(Kenya$Latitude)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
## 0.03699 0.07817 0.10870 0.11100 0.14980 0.18490     63
```

```
hist(Kenya$Latitude)
```



```
summary(Kenya$Longitude)
```

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
##	35.08	35.11	35.17	35.17	35.21	35.27	63

```
hist(Kenya$Longitude)
```

