

MIT-Analytics Edge Course - May 2016

May 26, 2016

1 This is the record of the coding work done.

This is the first part of the course in learning terms and shows how we can create a dataset

```
sd(c(5,8,12))

## [1] 3.511885

which.min(c(4,1,6))

## [1] 2

seq(0,100,2)

## [1] 0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32
## [18] 34 36 38 40 42 44 46 48 50 52 54 56 58 60 62 64 66
## [35] 68 70 72 74 76 78 80 82 84 86 88 90 92 94 96 98 100

Country=c("Brazil","China","India","Switzerland","USZ")
LifeExpec=c(74,76,65,83,79)
CountryData=data.frame(Country,LifeExpec)
CountryData

##      Country LifeExpec
## 1      Brazil      74
## 2       China      76
## 3       India      65
## 4 Switzerland      83
## 5         USZ      79

CountryData$Population=c(199,130,124,7.9,318)
CountryData

##      Country LifeExpec Population
## 1      Brazil      74      199.0
## 2       China      76      130.0
## 3       India      65      124.0
## 4 Switzerland      83       7.9
## 5         USZ      79      318.0

Country=c("Australia","Greece")
LifeExpec=c(82,81)
Population=c(23,11)
NewCountryData=data.frame(Country,LifeExpec,Population)
NewCountryData

##      Country LifeExpec Population
## 1 Australia      82         23
## 2   Greece      81         11

AllCountryData=rbind(CountryData,NewCountryData)
AllCountryData
```

```
##      Country LifeExpec Population
## 1      Brazil      74      199.0
## 2       China      76      130.0
## 3       India      65      124.0
## 4 Switzerland      83       7.9
## 5         USZ      79      318.0
## 6   Australia      82       23.0
## 7     Greece      81       11.0
```

2 Loading Data Files

This allows us to load a csv file and create a subset and write that to a file

```
setwd("/home/samar/Desktop/NewProjectsDone/MOOC/MIT_AnalyticsEdge")
getwd()
```

```
## [1] "/home/samar/Desktop/NewProjectsDone/MOOC/MIT_AnalyticsEdge"
```

```
WHO=read.csv("WHO.csv")
str(WHO)
```

```
## 'data.frame': 194 obs. of 13 variables:
```

```
## $ Country      : Factor w/ 194 levels "Afghanistan",...: 1 2 3 4 5 6 7 8 9 10 ...
## $ Region       : Factor w/ 6 levels "Africa","Americas",...: 3 4 1 4 1 2 2 4 6 4 ...
## $ Population   : int  29825 3162 38482 78 20821 89 41087 2969 23050 8464 ...
## $ Under15      : num  47.4 21.3 27.4 15.2 47.6 ...
## $ Over60       : num  3.82 14.93 7.17 22.86 3.84 ...
## $ FertilityRate : num  5.4 1.75 2.83 NA 6.1 2.12 2.2 1.74 1.89 1.44 ...
## $ LifeExpectancy : int  60 74 73 82 51 75 76 71 82 81 ...
## $ ChildMortality : num  98.5 16.7 20 3.2 163.5 ...
## $ CellularSubscribers : num  54.3 96.4 99 75.5 48.4 ...
## $ LiteracyRate   : num  NA NA NA NA 70.1 99 97.8 99.6 NA NA ...
## $ GNI            : num  1140 8820 8310 NA 5230 ...
## $ PrimarySchoolEnrollmentMale : num  NA NA 98.2 78.4 93.1 91.1 NA NA 96.9 NA ...
## $ PrimarySchoolEnrollmentFemale: num  NA NA 96.4 79.4 78.2 84.5 NA NA 97.5 NA ...
```

```
summary(WHO)
```

```
##      Country      Region      Population
## Afghanistan      : 1  Africa      :46  Min.      : 1
## Albania          : 1  Americas     :35  1st Qu.: 1696
## Algeria          : 1  Eastern Mediterranean:22  Median : 7790
## Andorra          : 1  Europe      :53  Mean   : 36360
## Angola           : 1  South-East Asia :11  3rd Qu.: 24535
## Antigua and Barbuda: 1  Western Pacific :27  Max.   :1390000
## (Other)          :188
##      Under15      Over60      FertilityRate      LifeExpectancy
## Min.      :13.12  Min.      : 0.81  Min.      :1.260  Min.      :47.00
## 1st Qu.:18.72  1st Qu.: 5.20  1st Qu.:1.835  1st Qu.:64.00
## Median :28.65  Median : 8.53  Median :2.400  Median :72.50
## Mean      :28.73  Mean      :11.16  Mean      :2.941  Mean      :70.01
## 3rd Qu.:37.75  3rd Qu.:16.69  3rd Qu.:3.905  3rd Qu.:76.00
## Max.      :49.99  Max.      :31.92  Max.      :7.580  Max.      :83.00
##      NA's      :11
##      ChildMortality      CellularSubscribers      LiteracyRate      GNI
## Min.      : 2.200  Min.      : 2.57  Min.      :31.10  Min.      : 340
## 1st Qu.: 8.425  1st Qu.: 63.57  1st Qu.:71.60  1st Qu.: 2335
## Median :18.600  Median : 97.75  Median :91.80  Median : 7870
## Mean      :36.149  Mean      : 93.64  Mean      :83.71  Mean      :13321
## 3rd Qu.:55.975  3rd Qu.:120.81  3rd Qu.:97.85  3rd Qu.:17558
## Max.      :181.600  Max.      :196.41  Max.      :99.80  Max.      :86440
##      NA's      :10      NA's      :91      NA's      :32
##      PrimarySchoolEnrollmentMale      PrimarySchoolEnrollmentFemale
## Min.      : 37.20      Min.      : 32.50
```

```
## 1st Qu.: 87.70      1st Qu.: 87.30
## Median : 94.70      Median : 95.10
## Mean   : 90.85      Mean    : 89.63
## 3rd Qu.: 98.10      3rd Qu.: 97.90
## Max.   :100.00      Max.    :100.00
## NA's   :93          NA's     :93

WHO_Europe = subset(WHO, Region == "Europe")
str(WHO_Europe)

## 'data.frame': 53 obs. of 13 variables:
## $ Country          : Factor w/ 194 levels "Afghanistan",...: 2 4 8 10 11 16 17 22 26 42
## $ Region           : Factor w/ 6 levels "Africa","Americas",...: 4 4 4 4 4 4 4 4 4 4 ...
## $ Population       : int  3162 78 2969 8464 9309 9405 11060 3834 7278 4307 ...
## $ Under15          : num  21.3 15.2 20.3 14.5 22.2 ...
## $ Over60           : num  14.93 22.86 14.06 23.52 8.24 ...
## $ FertilityRate     : num  1.75 NA 1.74 1.44 1.96 1.47 1.85 1.26 1.51 1.48 ...
## $ LifeExpectancy    : int  74 82 71 81 71 71 80 76 74 77 ...
## $ ChildMortality    : num  16.7 3.2 16.4 4 35.2 5.2 4.2 6.7 12.1 4.7 ...
## $ CellularSubscribers : num  96.4 75.5 103.6 154.8 108.8 ...
## $ LiteracyRate      : num  NA NA 99.6 NA NA NA NA 97.9 NA 98.8 ...
## $ GNI               : num  8820 NA 6100 42050 8960 ...
## $ PrimarySchoolEnrollmentMale : num  NA 78.4 NA NA 85.3 NA 98.9 86.5 99.3 94.8 ...
## $ PrimarySchoolEnrollmentFemale: num  NA 79.4 NA NA 84.1 NA 99.2 88.4 99.7 97 ...

write.csv(WHO_Europe, "WHO_Europe.csv")
ls()

## [1] "AllCountryData" "Country"         "CountryData"     "LifeExpec"
## [5] "NewCountryData" "Population"      "WHO"             "WHO_Europe"

rm(WHO_Europe)
ls()

## [1] "AllCountryData" "Country"         "CountryData"     "LifeExpec"
## [5] "NewCountryData" "Population"      "WHO"

sd(WHO$Over60)

## [1] 7.149331
```