Final Case Study 1

Rob Rivens

3/22/2017

Case Study1 Code - Rob Rivens Intro to Data Science Section 402

This document combines GDP Rankings for 190 countries around the ##globe with economic data for each country, as compiled by the ##Worldbank website.

The following is a set of instructions describing the process to #extract, cleanse and analyze the data in an effort to determine the #relationship between “Lower Middle Income” countries which fall #into the 1st quantile in terms of global GDP Rank.

```
options(repos = c(CRAN = "http://cran.rstudio.com"))
```

Download GDP Data from website

```
setwd("/Users/robertrivens/SMU Data Science/Intro to Data Science/CaseStudy1")
url <- "https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2FGDP.csv"
```

Set working directory and specify path for file download

```
f <- file.path(getwd(), "GDP.csv")
download.file(url, f)
install.packages("data.table")
```

```
##
##   There is a binary version available but the source version is
##   later:
##            binary source needs_compilation
## data.table 1.10.0 1.10.4              TRUE
```

```
## installing the source package 'data.table'
```

```
library(data.table)
```

Populate GDP Data Table; set variable names

```
dtGDP <- data.table(read.csv(f, skip = 4, nrows = 215))
dtGDP <- dtGDP[X != ""]
dtGDP <- dtGDP[, list(X, X.1, X.3, X.4)]
setnames(dtGDP, c("X", "X.1", "X.3", "X.4"), c("CountryCode","Ranking","Long.Name","gdp"))

#Download Education Data from website
url <- "https://d396qusza40orc.cloudfront.net/getdata%2Fdata%2FEDSTATS_Country.csv"

#Set working directory and specify path for file download
f <- file.path(getwd(), "FEDSTATS_Country.csv")
download.file(url, f)

#Populate Education Data Table
dtEd <- data.table(read.csv(f))
```

Merge GDP file with Education file, by Country Code

```
dt <- merge(dtGDP, dtEd, all = TRUE, by = c("CountryCode"))

#Count number of matches
sum(!is.na(unique(dt$Ranking)))
```

```
## [1] 189
```

```
##There were 189 matches from the GDP file to the Income file.  The countries which did not match were very small
 countries; there were also summary-level codes in the Income file which did not match a country by name.

#Sort merged file by Country Code in ascending order; display 13th Country with GDP ranking
dt[order(Ranking, decreasing = TRUE), list(CountryCode, Long.Name.x, Long.Name.y, Ranking, gdp)][13]
```

```
##    CountryCode         Long.Name.x         Long.Name.y Ranking   gdp
## 1:         KNA St. Kitts and Nevis St. Kitts and Nevis     178  767
```

```
##The 13th Country in terms of GDP rank is St. Kitts.

#Display average GDP Rankings by Income Group
dt[, mean(Ranking, na.rm = TRUE), by = Income.Group]
```

```
##            Income.Group        V1
## 1: High income: nonOECD  91.91304
## 2:           Low income 133.72973
## 3:  Lower middle income 107.70370
## 4:  Upper middle income  92.13333
## 5:    High income: OECD  32.96667
## 6:                   NA 131.00000
## 7:                            NaN
```

High Income: OECD GDP rank was 32.9 and the High Income nonOECD rank was 91.9

install package ggplot2 and themes (for analysis)

```
install.packages("ggplot2")
```

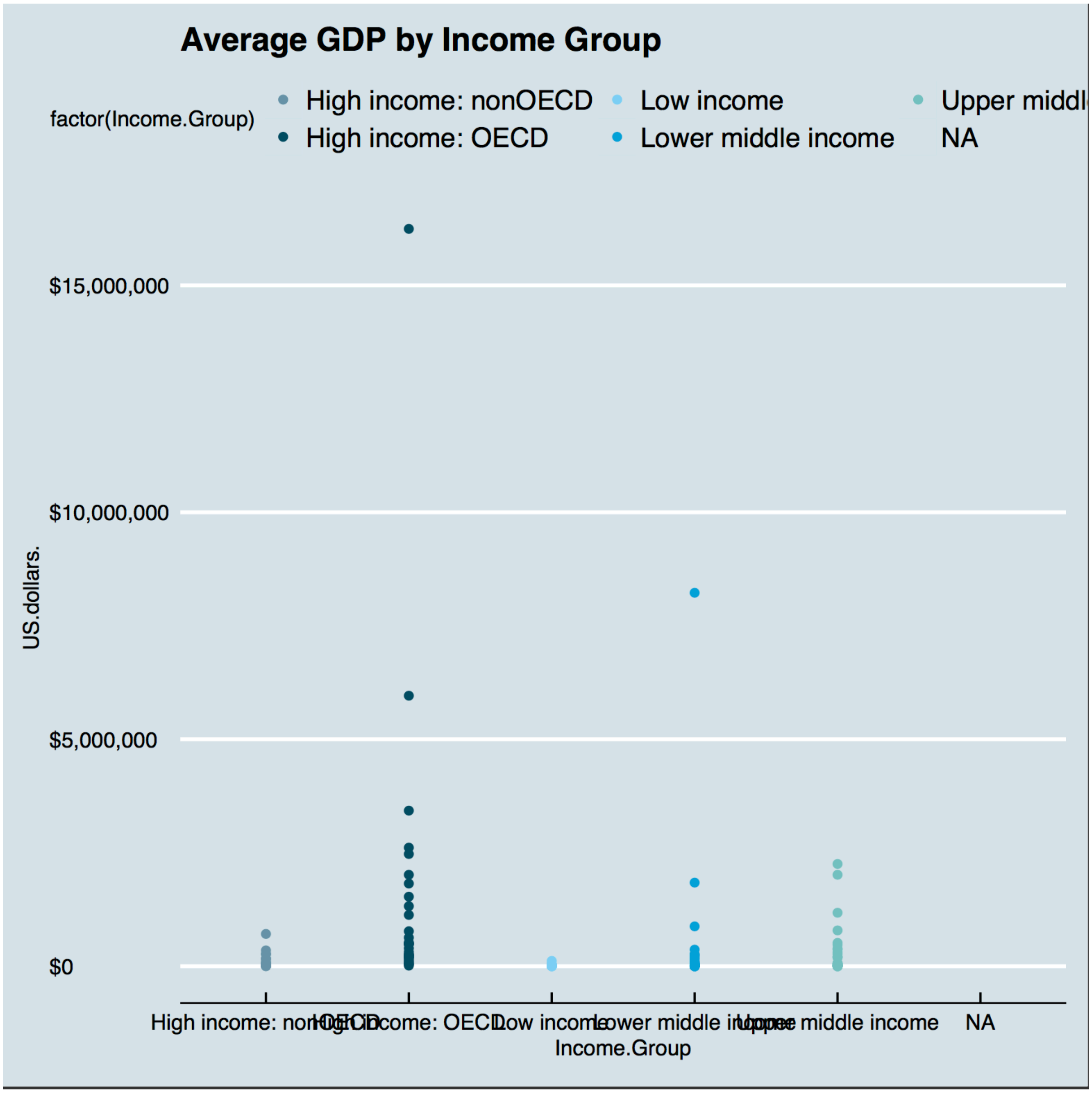
```
##
## The downloaded binary packages are in
##  /var/folders/5g/7gjy3t3s10s44xxfsd3g0_pm0000gn/T//RtmpLWRpeT/downloaded_packages
```

```
install.packages("ggthemes")
```

```
##
## The downloaded binary packages are in
##  /var/folders/5g/7gjy3t3s10s44xxfsd3g0_pm0000gn/T//RtmpLWRpeT/downloaded_packages
```

```
library(ggplot2)
library(ggthemes)

#Create color plot of GDP value by Income Group
ggplot(dt, aes(Income.Group, gdp, color=factor(Income.Group)))
```



Cut GDP Ranking into 5 quantile groups; make table Income Group by #GDP

```
breaks <- quantile(dt$Ranking, probs = seq(0, 1, 0.2), na.rm = TRUE)
dt$quantileGDP <- cut(dt$Ranking, breaks = breaks)
dt[Income.Group == "Lower middle income", .N, by = c("Income.Group", "quantileGDP")]
```

```
##           Income.Group quantileGDP  N
## 1: Lower middle income (38.8,76.6] 13
## 2: Lower middle income   (114,152]  8
## 3: Lower middle income   (152,190] 16
## 4: Lower middle income  (76.6,114] 12
## 5: Lower middle income    (1,38.8]  5
## 6: Lower middle income          NA  2
```

There were eight (8) countries in the Lower middle income bracket that fell into the top 38 GDP rank (1st quantile). These countries apparently have high exports since their average populations are Lower middle income. These countries are also high in terms of total population counts – which would further saturate the income spread per capita. The countries in order of highest GDP were China, Egypt, India, Indonesia and Thailand.