concordance=TRUE

IT497 OSEMN Assignment

Sonali Changkakoti
School of Information Technology
Illinois State University
schangk@ilstu.edu

October 20, 2014

United States School Statistics (100 Largest Cities): Chicago, Illinois

Abstract

There are around 100 schools in Chicago.-Rough-Write more.

1 Introduction

2 Data

We have to collect data for plotting the graph consisting of total numbers of schools, teachers and schools in Chicago, Illinois. We will get the required data from the Quandl API. We will be using various libraries available in R like RCurl, ggplot2, reshape2 and Quandl.

1. Obtaing the data.

   We will obtain data by using ggplot2.

```
# We will be using RCurl, ggplot2, reshape2 and Quandl.
library(RCurl)

## Loading required package:  bitops

library(ggplot2)
library(reshape2)
library(Quandl)
library(knitr)


# To get all tables on the page
```

```
myCsv <- getURL("https://www.quandl.com/api/v1/datasets/
                NCES/SCHOOLS_CITIES_CHICAGOILLINOIS.csv",ssl.verifypeer = FALSE)
school_data<-read.csv(textConnection(myCsv))

# Using Quandl to get all data

myData <- Quandl("NCES/SCHOOLS_CITIES_CHICAGOILLINOIS",
                 authcode="sUMBj-Lb7MRzwnUWXvxe")
```

2. Cleaning data

   We will scrub and clean our data to get the relevant data needed to obtain the results.

```
# Filtering the data needed to plot a graph showing total
#students, total teachers and total schools in Chicago, Illinois
cleanData<-myData[,1:4]
# Changing the columns' name
colnames(cleanData) <- c("year", "schools", "students", "teachers")
cleanData
```

```
##          year schools students teachers
## 1  2011-12-31     620   400383 21847.46
## 2  2010-12-31     614   402951 22588.93
## 3  2009-12-31     610   420193 21062.10
## 4  2008-12-31     600   399013 19674.00
## 5  2007-12-31     597   408311 18715.00
## 6  2006-12-31     600   415293 24659.00
## 7  2005-12-31     588   420787 23417.50
## 8  2004-12-31     588   428221 21261.90
## 9  2003-12-31     581   432478 22876.80
## 10 2002-12-31     574   432027 22419.10
## 11 2001-12-31     573   429684 23012.00
```

3. Explore data

   We will need to Explore (the E) the data by showing the following commands and their results - class(), str(), summary() and xtabs().

```
# Class
class(cleanData)
```

```
## [1] "data.frame"
```

```
# Str
str(cleanData)
```

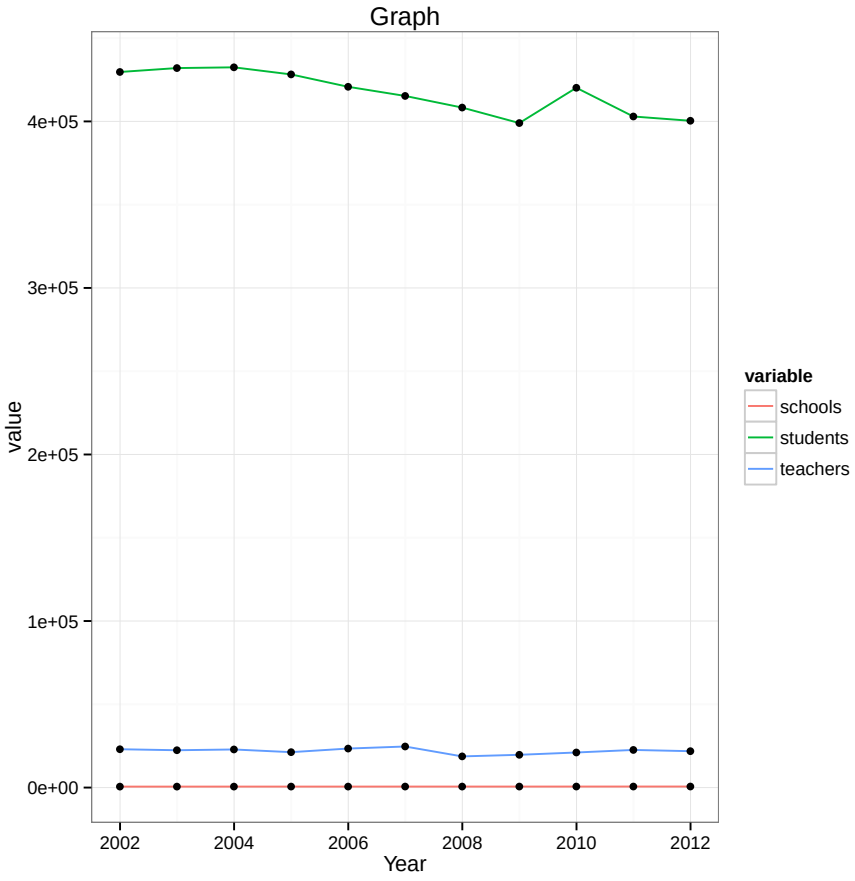
```
## 'data.frame': 11 obs. of  4 variables:
##  $ year    : Date, format: "2011-12-31" "2010-12-31" ...
##  $ schools : num  620 614 610 600 597 600 588 588 581 574 ...
##  $ students: num  400383 402951 420193 399013 408311 ...
##  $ teachers: num  21847 22589 21062 19674 18715 ...

# Summary
summary(cleanData)

##       year               schools         students         teachers
##  Min.   :2001-12-31   Min.   :573.0   Min.   :399013   Min.   :18715
##  1st Qu.:2004-07-01   1st Qu.:584.5   1st Qu.:405631   1st Qu.:21162
##  Median :2006-12-31   Median :597.0   Median :420193   Median :22419
##  Mean   :2006-12-31   Mean   :595.0   Mean   :417213   Mean   :21958
##  3rd Qu.:2009-07-01   3rd Qu.:605.0   3rd Qu.:428953   3rd Qu.:22944
##  Max.   :2011-12-31   Max.   :620.0   Max.   :432478   Max.   :24659
```

3 Result

```
# Melting the data
moltenData <- melt(cleanData,id.vars="year")
ggplot(moltenData, aes(as.Date(year,"%e %b %Y"),value))+
  geom_line(aes(color = variable))+
   geom_point()  + xlab("Year") +
  labs(title = "Graph")+ theme_bw()
```


Graph
4e+05
3e+05
2e+05
1e+05
0e+00
value
2002
2004
2006
2008
2010
2012
Year
variable
schools
students
teachers