

Annual Arts in School Report 2013-2014

CUNY MSDA - DATA607 - Project 2_a

Completed by: Duubar Villalobos Jimenez mydvtech@gmail.com

March 12, 2017



Figure 1:

The goal of this assignment is to give you practice in preparing different datasets for downstream analysis work.

Your task is to:

- (1) Choose any **three** of the “**wide**” **datasets** identified in the Week 5 Discussion items. (You may use your own dataset; please don’t use my Sample Post dataset, since that was used in your Week 5 assignment!)

For each of the three chosen datasets:

- Create a **.CSV** file (or optionally, a **MySQL** database!) that includes all of the information included in the dataset. You’re encouraged to use a “wide” structure similar to how the information appears in the discussion item, so that you can practice tidying and transformations as described below.
- Read the information from your **.CSV** file into **R**, and use **tidyr** and **dplyr** as needed to tidy and transform your data. [Most of your grade will be based on this step!]
- Perform the analysis requested in the discussion item.
- Your code should be in an R Markdown file, posted to **rpubs.com**, and should include narrative descriptions of your data cleanup work, analysis, and conclusions.

- (2) Please include in your homework submission, for each of the three chosen datasets:

The **URL** to the **.Rmd** file in your **GitHub** repository, and The URL for your **rpubs.com** web page.

PROCEDURE

Library definitions

```
# Need to employ kable table reporting functionality
library(knitr)
# Need to employ stringr for Regular Expressions
library(stringr)
# Need to employ to use tidy data functions
library(tidyr)
library(dplyr)
```

Annual Arts in School Report Data 2013 - 2014.

Dataset url location: <https://data.cityofnewyork.us/Education/Annual-Arts-in-School-Reports-Raw-Data/vdgp-ddvg>

The Annual Arts in Schools Report includes data about arts teachers, arts budgeting, space for the arts, partnerships with arts and cultural organizations, and parent involvement for elementary, middle, and high schools. These reports help school administrators, parents, and students understand how their schools are progressing towards offering universal arts education to all students.

Last Updated:

September 1, 2015

Data Provided by:

Department of Education (DOE)

Dataset Owner:

NYC OpenData

Dictionary

This dataset does not seem to have a dictionary. The download link is for a **.zip** file containing three **.csv** files as follows:

- *Arts Survey Data_ 2011-2012_ ODP.csv*
- *Arts Survey Data_ 2012-2013_ ODP.csv*
- *Arts Survey Data_ 2013-2014_ ODP.csv*

For simplicity reasons, I will read the raw data directly from the source.

URL Raw data location:

```
url <- "https://data.cityofnewyork.us/api/file_data/bXGEFYYP6F8SYozNljEvaG4BmTKdDtilNHRWeDEyEY4?filename="
```

(1) Read information from .CSV file into R.

From the above file, I will choose “**Arts Survey Data_ 2013-2014_ ODP.csv**” for the latest information contained on that .zip file, I am just keeping in mind that the other two files can be worked and compared in a yearly fashion since each file represents the information for a single school year.

Read .csv from url by employing read.csv()

For this project I will try something new. I will load the data directly from the original location from a zip file. For this, I will be reading such url for the **.zip** file into **R** without unzipping the file from the web in advance; then I will unzip the file that I want to work with and read it into a data frame.

```
# Procedure to download .zip file containing various files including the one I want to work with.
temp <- tempfile()
download.file(url, temp)
my.file <- unz(temp, "Arts Survey Data_ 2013-2014_ ODP.csv")
my.data <- read.csv(my.file, header=TRUE, sep=",", stringsAsFactors=FALSE)

# Deleting downloaded file
unlink(temp)
```

Imported file structure display

```
## 'data.frame':   1415 obs. of  1659 variables:
## $ School_Year      : chr  "2013-14" "2013-14" "2013-14" "2013-14" ...
## $ DBN              : chr  "01M015" "01M019" "01M020" "01M034" ...
## $ BN              : chr  "M015" "M019" "M020" "M034" ...
## $ District         : int   1 1 1 1 1 1 1 1 1 ...
## $ School_Name      : chr  "P.S. 015 Roberto Clemente" "P.S. 019 Asher Levy" "P.S. 020 Anna ...
## $ School_Level     : chr  "Elementary" "Elementary" "Elementary" "K-8" ...
```

```

## $ PreK           : int 26 36 54 18 18 36 33 18 12 31 ...
## $ K              : int 39 39 114 33 40 63 41 24 30 69 ...
## $ FirstGr        : int 39 38 111 32 33 85 38 28 31 48 ...
## $ SecondGr       : int 21 36 98 35 26 56 53 30 31 61 ...
## $ ThirdGr        : int 16 45 109 34 23 74 46 43 34 53 ...
## $ FourthGr       : int 26 47 71 45 20 56 49 33 34 63 ...
## $ FifthGr        : int 23 44 74 43 19 46 58 26 28 59 ...
## $ SixthGr        : int NA NA NA 43 NA NA NA NA 61 NA ...
## $ SeventhGr      : int NA NA NA 57 NA NA NA NA 63 NA ...
## $ EighthGr       : int NA NA NA 53 NA NA NA NA 69 NA ...
## $ NinthGr        : int NA NA NA NA NA NA NA NA NA NA ...
## $ TenthGr        : int NA NA NA NA NA NA NA NA NA NA ...
## $ EleventhGr     : int NA NA NA NA NA NA NA NA NA NA ...
## $ TwelfthGr      : int NA NA NA NA NA NA NA NA NA NA ...
## $ G01_CLS_01     : chr "1-012" "1-101" "1-015" "1-101" ...
## $ G01_CLS_02     : chr "1-102" "1-102" "1-101" "1-102" ...
## $ G01_CLS_03     : chr "" "1-103" "1-102" "" ...
## $ G01_CLS_04     : chr "" "1-901" "1-103" "" ...
## $ G01_CLS_05     : chr "" "" "1-104" "" ...
## $ G01_CLS_06     : chr "" "" "1-105" "" ...
## $ G01_CLS_07     : chr "" "" "" "" ...
## $ G01_CLS_08     : chr "" "" "" "" ...
## $ G01_CLS_09     : chr "" "" "" "" ...
## $ G01_CLS_10     : chr "" "" "" "" ...
## $ G01_CLS_11     : chr "" "" "" "" ...
## $ G01_CLS_12     : chr "" "" "" "" ...
## $ G01_CLS_13     : chr "" "" "" "" ...
## $ G01_CLS_14     : chr "" "" "" "" ...
## $ G01_CLS_15     : chr "" "" "" "" ...
## $ G01_CLS_16     : chr "" "" "" "" ...
## $ G02_CLS_01     : chr "2-202" "2-201" "2-103" "2-201" ...
## $ G02_CLS_02     : chr "" "2-202" "2-201" "2-202" ...
## $ G02_CLS_03     : chr "" "2-901" "2-202" "" ...
## $ G02_CLS_04     : chr "" "" "2-203" "" ...
## $ G02_CLS_05     : chr "" "" "2-204" "" ...
## $ G02_CLS_06     : chr "" "" "2-205" "" ...
## $ G02_CLS_07     : chr "" "" "" "" ...
## $ G02_CLS_08     : chr "" "" "" "" ...
## $ G02_CLS_09     : chr "" "" "" "" ...
## $ G02_CLS_10     : chr "" "" "" "" ...
## $ G02_CLS_11     : chr "" "" "" "" ...
## $ G02_CLS_12     : chr "" "" "" "" ...
## $ G02_CLS_13     : chr "" "" "" "" ...
## $ G03_CLS_01     : chr "3-302" "3-301" "3-203" "3-301" ...
## $ G03_CLS_02     : chr "" "3-302" "3-301" "3-302" ...
## $ G03_CLS_03     : chr "" "3-303" "3-302" "3-951" ...
## $ G03_CLS_04     : chr "" "3-902" "3-303" "" ...
## $ G03_CLS_05     : chr "" "" "3-304" "" ...
## $ G03_CLS_06     : chr "" "" "" "" ...
## $ G03_CLS_07     : chr "" "" "" "" ...
## $ G03_CLS_08     : chr "" "" "" "" ...
## $ G03_CLS_09     : chr "" "" "" "" ...
## $ G03_CLS_10     : chr "" "" "" "" ...
## $ G03_CLS_11     : chr "" "" "" "" ...

```

```
## $ G03_CLS_12      : chr  "" "" "" "" ...
## $ G03_CLS_13      : chr  "" "" "" "" ...
## $ G03_CLS_14      : chr  "" "" "" "" ...
## $ G04_CLS_01      : chr  "4-402" "4-401" "4-203" "4-401" ...
## $ G04_CLS_02      : chr  "" "4-402" "4-401" "4-402" ...
## $ G04_CLS_03      : chr  "" "4-403" "4-402" "4-951" ...
## $ G04_CLS_04      : chr  "" "4-902" "4-403" "" ...
## $ G04_CLS_05      : chr  "" "" "4-404" "" ...
## $ G04_CLS_06      : chr  "" "" "" "" ...
## $ G04_CLS_07      : chr  "" "" "" "" ...
## $ G04_CLS_08      : chr  "" "" "" "" ...
## $ G04_CLS_09      : chr  "" "" "" "" ...
## $ G04_CLS_10      : chr  "" "" "" "" ...
## $ G04_CLS_11      : chr  "" "" "" "" ...
## $ G04_CLS_12      : chr  "" "" "" "" ...
## $ G04_CLS_13      : chr  "" "" "" "" ...
## $ G04_CLS_14      : chr  "" "" "" "" ...
## $ G04_CLS_15      : chr  "" "" "" "" ...
## $ G05_CLS_01      : chr  "5-502" "5-501" "5-402" "5-501" ...
## $ G05_CLS_02      : chr  "" "5-502" "5-403" "5-502" ...
## $ G05_CLS_03      : chr  "" "5-902" "5-501" "5-951" ...
## $ G05_CLS_04      : chr  "" "" "5-502" "" ...
## $ G05_CLS_05      : chr  "" "" "" "" ...
## $ G05_CLS_06      : chr  "" "" "" "" ...
## $ G05_CLS_07      : chr  "" "" "" "" ...
## $ G05_CLS_08      : chr  "" "" "" "" ...
## $ G05_CLS_09      : chr  "" "" "" "" ...
## $ G05_CLS_10      : chr  "" "" "" "" ...
## $ G05_CLS_11      : chr  "" "" "" "" ...
## $ G05_CLS_12      : chr  "" "" "" "" ...
## $ G05_CLS_13      : chr  "" "" "" "" ...
## $ G05_CLS_14      : chr  "" "" "" "" ...
## $ G05_CLS_15      : chr  "" "" "" "" ...
## $ G05_CLS_16      : logi  NA NA NA NA NA NA ...
## $ REG01_CLS_01     : int   9 10 7 14 15 22 22 3 26 3 ...
## $ REG01_CLS_02     : int   30 10 24 18 18 24 16 25 5 21 ...
## $ REG01_CLS_03     : int   NA 14 27 NA NA 16 NA NA NA 24 ...
## $ REG01_CLS_04     : int   NA 4 14 NA NA 18 NA NA NA NA ...
## $ REG01_CLS_05     : int   NA NA 19 NA NA 5 NA NA NA NA ...
## [list output truncated]
```

In summary, this data frame contains 1416 independent observations of 1659 recognizable possible values.

Data transformation

Now that I have the data frame I will **subset** a small portion (first 20 columns) in order to create some possible outcomes from the given information.

```
# New subset of the main data (first 20 columns) from School.Year to TwelfthGr
my.new.data <- my.data %>% subset(select=(School.Year:TwelfthGr))
```

New subset table.

School.Year	DBN	BN	District	School_Name	School_Level	PreK	K	FirstGr	SecondGr
2013-14	01M015	M015	1	P.S. 015 Roberto Clemente	Elementary	26	39	39	
2013-14	01M019	M019	1	P.S. 019 Asher Levy	Elementary	36	39	38	
2013-14	01M020	M020	1	P.S. 020 Anna Silver	Elementary	54	114	111	
2013-14	01M034	M034	1	P.S. 034 Franklin D. Roosevelt	K-8	18	33	32	
2013-14	01M063	M063	1	The STAR Academy - P.S.63	Elementary	18	40	33	
2013-14	01M110	M110	1	P.S. 110 Florence Nightingale	Elementary	36	63	85	

Creating a Tidy Table

From the given table, we can modify it by taking the grade as one variable, that is from **PreK** to **TwelfthGr** will be put into a column named **Grade**. This table transformation will be done by employing the **gather()** function from the **tidyr** library.

```
my.tidy.data <- my.new.data %>% gather("Grade", "n.Students", PreK:TwelfthGr)
```

Short tidy table example after transformation.

School.Year	DBN	BN	District	School_Name	School_Level	Grade	n.Students
2013-14	01M015	M015	1	P.S. 015 Roberto Clemente	Elementary	PreK	26
2013-14	01M019	M019	1	P.S. 019 Asher Levy	Elementary	PreK	36
2013-14	01M020	M020	1	P.S. 020 Anna Silver	Elementary	PreK	54
2013-14	01M034	M034	1	P.S. 034 Franklin D. Roosevelt	K-8	PreK	18
2013-14	01M063	M063	1	The STAR Academy - P.S.63	Elementary	PreK	18
2013-14	01M110	M110	1	P.S. 110 Florence Nightingale	Elementary	PreK	36

From the above table we can explore a few things as follows:

Total number of art students:

The grand total of art students for all districts, school levels and grades is 883813 art students.

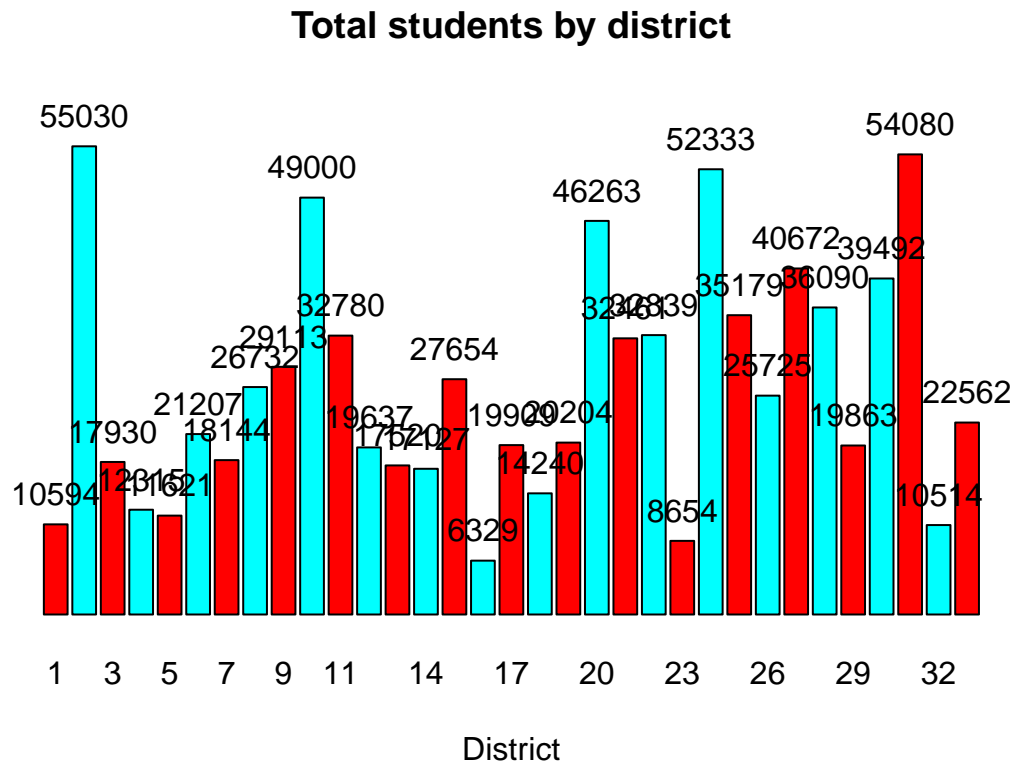
Top 5 districts with highest art student population and percentages.

District	Total Students	Percentage
2	55030	6.23 %
31	54080	6.12 %
24	52333	5.92 %
10	49000	5.54 %
20	46263	5.23 %

Bottom 5 districts with lowest art student population and percentages.

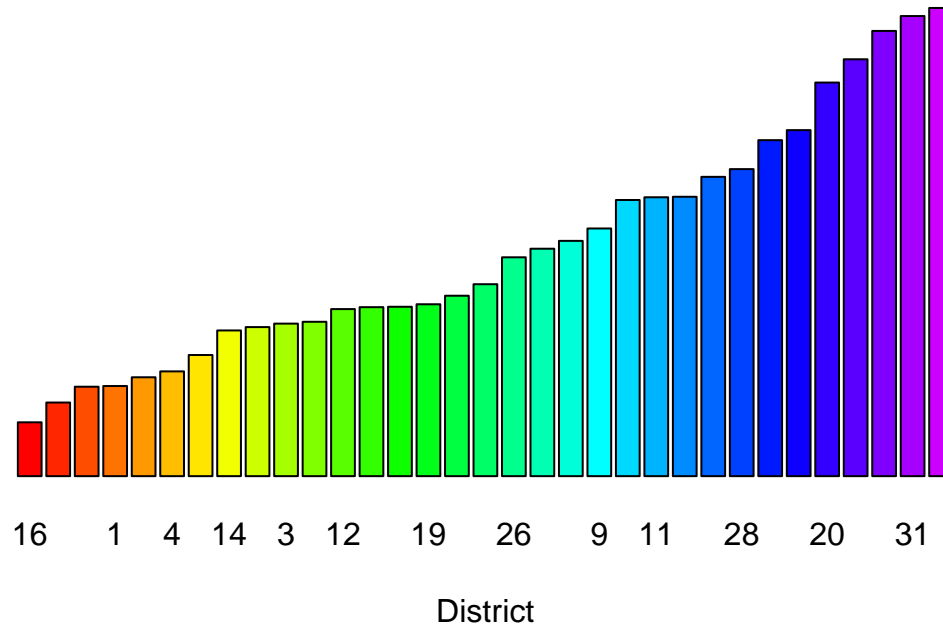
District	Total Students	Percentage
5	11621	1.31 %
1	10594	1.2 %
32	10514	1.19 %
23	8654	0.98 %
16	6329	0.72 %

Bar plot: Total of art students in a district sorted by district.



Bar plot: Total of art students in a district sorted by the number of art students in a district.

Total students by district

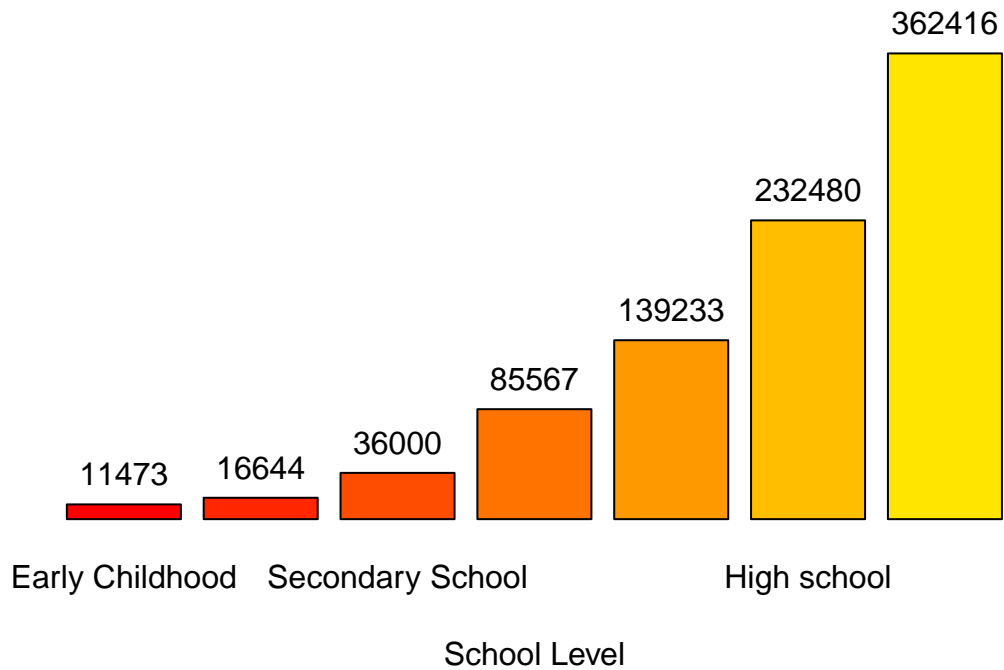


School level: art student population and percentages.

School_Level	Total Students	Percentage
Elementary	362416	41.01 %
High school	232480	26.3 %
Junior High-Intermediate-Middle	139233	15.75 %
K-8	85567	9.68 %
Secondary School	36000	4.07 %
K-12 all grades	16644	1.88 %
Early Childhood	11473	1.3 %

Bar plot: Total of art students by school level sorted by the number art of students.

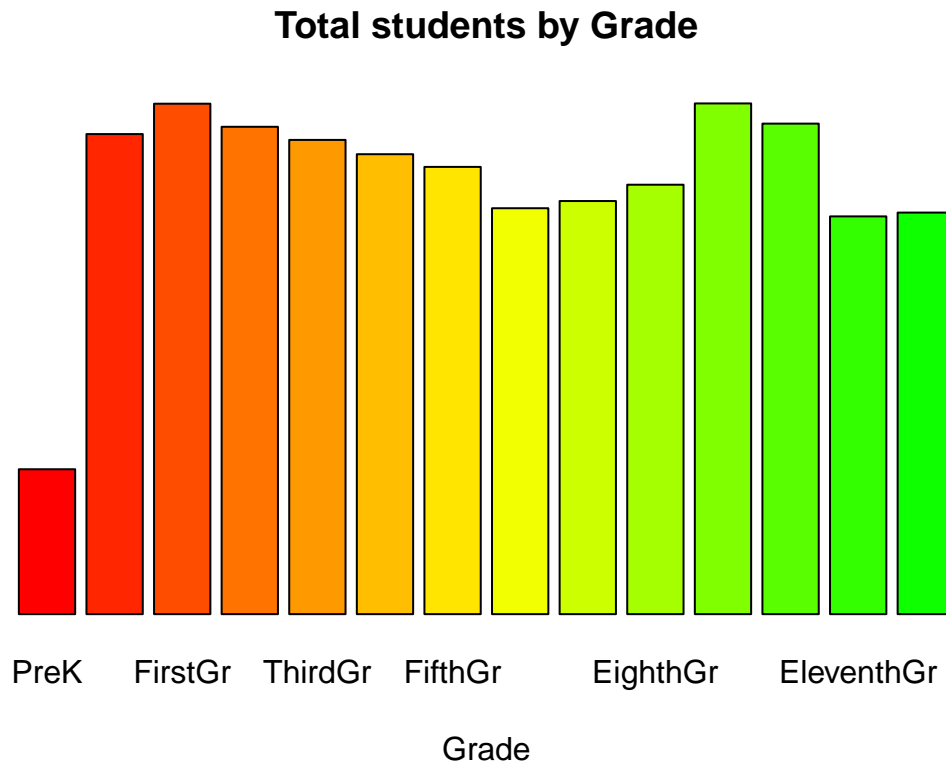
Total students by School Level



School Grades: art student population and percentages.

Grade	Total Students	Percentage
PreK	21157	2.39 %
K	70095	7.93 %
FirstGr	74529	8.43 %
SecondGr	71157	8.05 %
ThirdGr	69241	7.83 %
FourthGr	67154	7.6 %
FifthGr	65302	7.39 %
SixthGr	59266	6.71 %
SeventhGr	60318	6.82 %
EighthGr	62700	7.09 %
NinthGr	74568	8.44 %
TenthGr	71619	8.1 %
EleventhGr	58074	6.57 %
TwelfthGr	58633	6.63 %

Bar plot: Total of art students in a grade sorted by grade.



Conclusions

Districts:

Based on data analysis, we can appreciate major art student population differences in between the top 5 districts and bottom 5 districts with as much as 48701 art students. This will represent a good starting point to find out the multiple reasons that could affect as much; could it be population density playing a role on this?

School level:

Based on simple observation, we can observe how higher school levels tend to have more students enrolled in art classes while basic school levels tend to have less students enrolled. This, I think is an excellent topic to discuss in regards of early child development and communication trough art expression.

Grade:

If we look at the chart and results, we noticed that over all the NYC department of education in the 2013-2014 school year, presented very similar populations of all grade levels taking art classes, once again on this comparison comes to show how early aged students tend to have less populations taking art classes.

Final conclusion:

Since there's more data available, it will be interesting to perform school year to year comparisons in regards of the enrolled art students; this, to find out if some programs have been put in place in order to increase or decrease student participation in art classes. Also, it will be interesting to see how the populations transform or behave from year to year.