Lab 3

Stencila SchoolCal

Jimmy(Jinye) Lu
Nelson Wong
Jacob(Xiangchen) Liu

School performances and Poverty in California

This example implements the California School SAT Performance and Poverty Data which offers information about schools in California, their SAT performance, and economic information (Free-or-Reduced-Price Meal eligibility).

Usual intro:

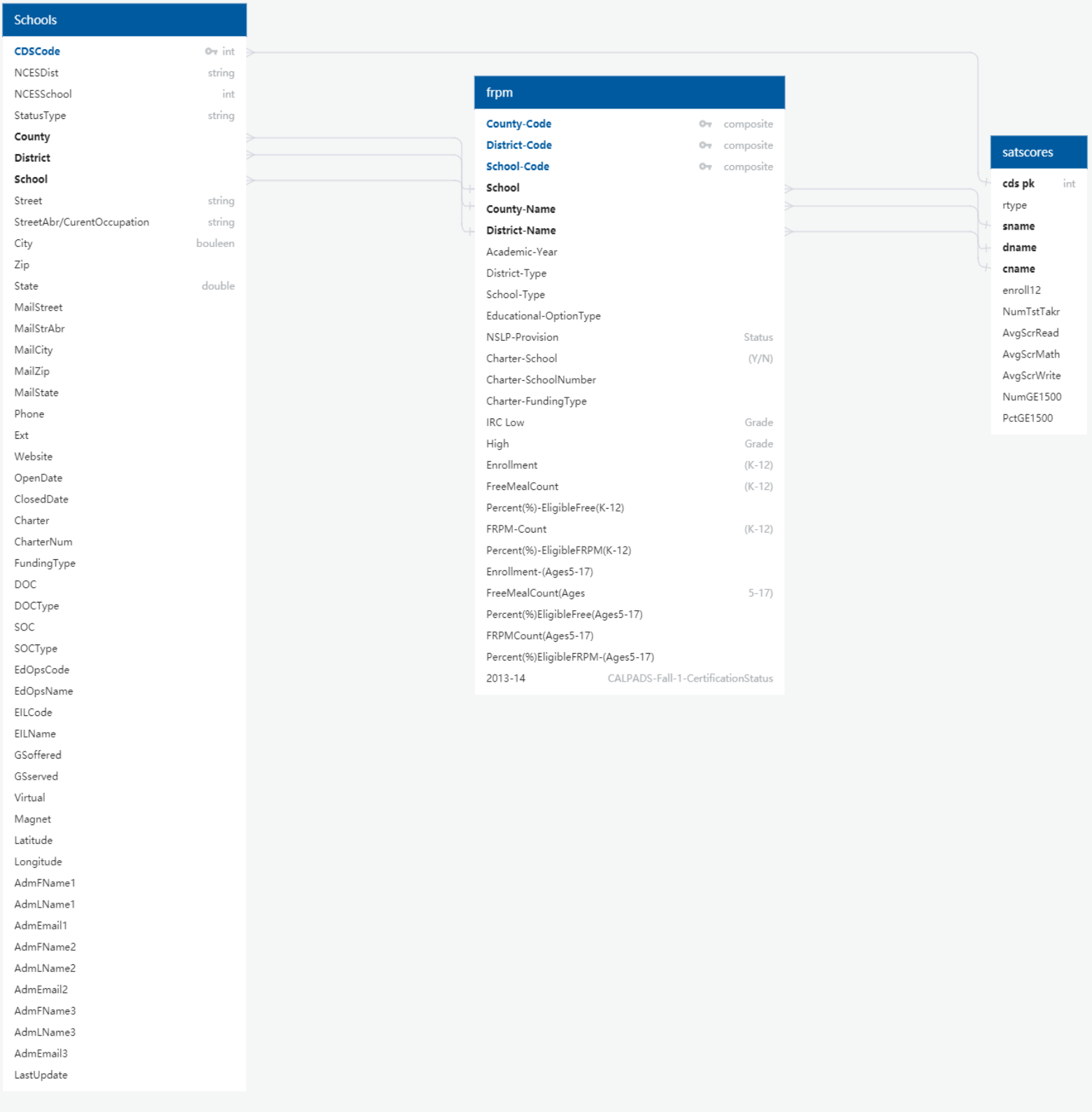
As usual, we start by querying the `sqlite_master` table to get a list of the tables in the database

```sql
SELECT name, sql FROM sqlite_master WHERE type='table'
```

name	sql
schools	CREATE TABLE schools ("CDSCode" VARCHAR(14) NOT NULL, "NCESDist" VARCHAR(7), "NCESSchool" VARCHAR(5), "StatusType" VARCHAR(7) NOT NULL, "County" VARCHAR(15) NOT NULL, "District" VARCHAR(74) NOT NULL, "School" VARCHAR(89), "Street" VARCHAR(62), "StreetAbr" VARCHAR(59), "City" VARCHAR(25), "Zip" VARCHAR(10), "State" VARCHAR(4), "MailStreet" VARCHAR(70), "MailStrAbr" VARCHAR(70), "MailCity" VARCHAR(22), "MailZip" VARCHAR(10), "MailState" VARCHAR(4), "Phone" VARCHAR(14), "Ext" VARCHAR(6), "Website" VARCHAR(96), "OpenDate" DATE, "ClosedDate" DATE, "Charter" BOOLEAN, "CharterNum" VARCHAR(4), "FundingType" VARCHAR(23), "DOC" VARCHAR(2) NOT NULL, "DOCType" VARCHAR(42) NOT NULL, "SOC" VARCHAR(4), "SOCType" VARCHAR(41), "EdOpsCode" VARCHAR(7), "EdOpsName" VARCHAR(43), "EILCode" VARCHAR(8), "EILName" VARCHAR(31), "GSoffered" VARCHAR(16), "GSserved" VARCHAR(5), "Virtual" VARCHAR(4), "Magnet" BOOLEAN, "Latitude" FLOAT, "Longitude" FLOAT, "AdmFName1" VARCHAR(17), "AdmLName1" VARCHAR(22), "AdmEmail1" VARCHAR(49), "AdmFName2" VARCHAR(15), "AdmLName2" VARCHAR(18), "AdmEmail2" VARCHAR(40), "AdmFName3" VARCHAR(8), "AdmLName3" VARCHAR(15), "AdmEmail3" VARCHAR(39), "LastUpdate" DATE NOT NULL, CHECK ("Charter" IN (0, 1)), CHECK ("Magnet" IN (0, 1)))
frpm	CREATE TABLE frpm ("Academic Year" VARCHAR(9), "County Code" VARCHAR(4), "District Code" INTEGER, "School Code" VARCHAR(7), "County Name" VARCHAR(15), "District Name" VARCHAR(75), "School Name" VARCHAR(85), "District Type" VARCHAR(32), "School Type" VARCHAR(41), "Educational Option Type" VARCHAR(43), "NSLP Provision Status" VARCHAR(24), "Charter School (Y/N)" BOOLEAN, "Charter School Number" VARCHAR(4), "Charter Funding Type" VARCHAR(23), "IRC" BOOLEAN, "Low Grade" VARCHAR(5), "High Grade" VARCHAR(14), "Enrollment (K-12)" FLOAT, "Free Meal Count (K-12)" FLOAT, "Percent (%) Eligible Free (K-12)" FLOAT, "FRPM Count (K-12)" FLOAT, "Percent (%) Eligible FRPM (K-12)" FLOAT, "Enrollment (Ages 5-17)" FLOAT, "Free Meal Count (Ages 5-17)" FLOAT, "Percent (%) Eligible Free (Ages 5-17)" FLOAT, "FRPM Count (Ages 5-17)" FLOAT, "Percent (%) Eligible FRPM (Ages 5-17)" FLOAT, "2013-14 CALPADS Fall 1 Certification Status" BOOLEAN, CHECK ("Charter School (Y/N)" IN (0, 1)), CHECK ("IRC" IN (0, 1)), CHECK ("2013-14 CALPADS Fall 1 Certification Status" IN (0, 1)))
satscores	CREATE TABLE satscores (cds VARCHAR(14) NOT NULL, rtype VARCHAR(1) NOT NULL, sname VARCHAR(50), dname VARCHAR(50), cname VARCHAR(15), enroll12 INTEGER NOT NULL, "NumTstTakr" INTEGER NOT NULL, "AvgScrRead" INTEGER, "AvgScrMath" INTEGER, "AvgScrWrite" INTEGER, "NumGE1500" INTEGER, "PctGE1500" FLOAT)

Then we take a look at each tables in the database which contains different fields, to get a taste of what they contain.
Firstly, we tried to create a diagram of the tables and their possible relationships on quickdatabase.

www.quickdatabasediagrams.com



```
sql
select *
from schools
```

CDSCode	NCESDist	NCESSchool	StatusType	County	District	School	Street	StreetAbr	City	Zip
01100170000000	0691051		Active	Alameda	Alameda County Office of Education		313 West Winton Avenue	313 West Winton Ave.	Hayward	94544-1136
01100170109835	0691051	10546	Closed	Alameda	Alameda County Office of Education	FAME Public Charter	39899 Balentine Drive, Suite 335	39899 Balentine Dr., Ste. 335	Newark	94560-5359
01100170112607	0691051	10947	Active	Alameda	Alameda County Office of Education	Envision Academy for Arts & Technology	1515 Webster Street	1515 Webster St.	Oakland	94612-3355
01100170118489	0691051	12283	Closed	Alameda	Alameda County Office of Education	Aspire California College Preparatory Academy	2125 Jefferson Avenue	2125 Jefferson Ave.	Berkeley	94703-1414
01100170123968	0691051	12844	Active	Alameda	Alameda County Office of Education	Community School for Creative Education	2111 International Boulevard	2111 International Blvd.	Oakland	94606-4903
01100170124172	0691051	12901	Active	Alameda	Alameda County Office of Education	Yu Ming Charter	1086 Alcatraz Avenue	1086 Alcatraz Ave.	Oakland	94608-1265
01100170125567	0691051	13008	Active	Alameda	Alameda County Office of Education	Urban Montessori Charter	5328 Brann Street	5328 Brann St.	Oakland	94619-3312
01100170130302	0691051		Closed	Alameda	Alameda County Office of Education	Technical, Agricultural & Nat.	597 C Street	597 C St.	Hayward	94541
01100170130401	0691051	09264	Active	Alameda	Alameda County Office of Education	Alameda County Juvenile Hall/Court	2500 Fairmont Avenue	2500 Fairmont Ave.	San Leandro	94578-1005
01100170130419	0691051	06830	Active	Alameda	Alameda County Office of Education	Alameda County Community	313 West Winton Avenue	313 West Winton Ave.	Hayward	94544-1136

Showing 10 out of 17686 rows

Through we review the schools table, we found that "CDSCode" is the only field that every observation in the table is not blank. So we can determine that "CDSCode" is the key of the schools tables.

And after the review of frpm table, we found that Academic Year and County Code and Street Code with District Code

are all the compulsory field that are not blank in every row except two rows. We found two observations that almost all the data in the table is missing or unknown. So we just filter these two rows data to get a precise insight about the data in frpm.

```sql
select *
from frpm
order by "Academic Year"
```

Academic Year	County Code	District Code	School Code	County Name	District Name	School Name	District Type	School Type	Educational Option Type	NSLP Provision Status	Charter School (Y/N)	Charte Schoo Numbe
		NA									NA	
		NA									NA	
2014-2015	01	10017	0109835	Alameda	Alameda County Office of Education	FAME Public Charter	County Office of Education (COE)	K-12 Schools (Public)	Traditional		1	072
2014-2015	01	10017	0112607	Alameda	Alameda County Office of Education	Envision Academy for Arts & Technology	County Office of Education (COE)	High Schools (Public)	Traditional		1	081
2014-2015	01	10017	0118489	Alameda	Alameda County Office of Education	Aspire California College Preparatory Academy	County Office of Education (COE)	High Schools (Public)	Traditional		1	104
2014-2015	01	10017	0123968	Alameda	Alameda County Office of Education	Community School for Creative Education	County Office of Education (COE)	Elementary Schools (Public)	Traditional	Breakfast Provision 2	1	128
2014-2015	01	10017	0124172	Alameda	Alameda County Office of Education	Yu Ming Charter	County Office of Education (COE)	Elementary Schools (Public)	Traditional		1	129
2014-2015	01	10017	0125567	Alameda	Alameda County Office of Education	Urban Montessori Charter	County Office of Education (COE)	Elementary Schools (Public)	Traditional		1	138
2014-2015	01	10017	0130401	Alameda	Alameda County Office of Education	Alameda County Juvenile Hall/Court	County Office of Education (COE)	Juvenile Court Schools	Juvenile Court School			
2014-2015	01	10017	0130419	Alameda	Alameda County Office of Education	Alameda County Community	County Office of Education (COE)	County Community	County Community School			

Showing 10 out of 10395 rows

Also, in the satscores table, we found that the "cds" field is the only field that is not blank in every row. And due to its uniqueness in the table, we think it is the key of the table. And we also found that there are some observation that, for example, their "cds" fields are "00000000000000" or "01000000000000" which are different from other observation in the table. The first observation represents all the schools' satscore data. And the second one "01000000000000" observation represents all the schools' data in Alameda County. And when the field "rtype" is "D", it means that the observation is the whole data in that district.

```sql
select *
from satscores
where cds
```

cds	rtype	sname	dname	cname	enroll12	NumTstTakr	AvgScrRead	AvgScrMath	AvgScrWrite	NumGE1500
01000000000000	C			Alameda	16978	8855	516	536	517	4900
01100170000000	D		Alameda County Office of Education	Alameda	398	88	418	418	417	14
01100170109835	S	FAME Public Charter	Alameda County Office of Education	Alameda	62	17	503	546	505	9
01100170112607	S	Envision Academy for Arts & Technology	Alameda County Office of Education	Alameda	75	71	397	387	395	5
01100170118489	S	Aspire California College Preparatory Academy	Alameda County Office of Education	Alameda	61		NA	NA	NA	NA
01611190000000	D		Alameda Unified	Alameda	922	544	521	546	519	333
01611190106401	S	Alameda Science and Technology Institute	Alameda Unified	Alameda	36	36	562	590	555	29
01611190119222	S	Nea Community Learning Center	Alameda Unified	Alameda	12	6	NA	NA	NA	NA
01611190130229	S	Alameda High	Alameda Unified	Alameda	465	325	543	573	543	229
01611190130609	S	Alameda Community Learning Center	Alameda Unified	Alameda	26	20	575	593	581	16

Showing 10 out of 2330 rows

After a initial review of the tables, we found that the three tables are related to each other with a unique field "CDS Code". And these tables are all one to one relationships. It means that one observation in

schools table can only refer to one observation in frpm table or satscores table. And one observation in frpm table can only refer to one observation in satscores table.

When we delve into the schools table, we want to gain some insight about California School information. So we find out which county has the most schools that is active now.

```sql
select County, count() as NumberofSchool
from schools
where StatusType=="Active"
group by County
order by NumberofSchool DESC
limit 10
```

County	NumberofSchool
Los Angeles	2431
San Diego	827
Orange	660
San Bernardino	608
Riverside	547
Santa Clara	465
Alameda	442
Sacramento	414
Fresno	394
Kern	324

And we also want to find out that which county's schools have the best average SAT performance in active schools.

```sql
select County, avg(PctGE1500) as SATPerformance
from schools join satscores on satscores.cds==schools.CDSCode
where StatusType=="Active"
group by County
order by SATPerformance DESC
limit 10
```

County	SATPerformance
Mariposa	71.79
Nevada	67.7286
El Dorado	67.0827
Marin	64.7638
Plumas	63.0275
Calaveras	62.885
Modoc	61.54
Amador	60.57
Santa Cruz	60.2265
San Luis Obispo	59.5493

Question:Does a correlation exist between SAT scores with a school's percentage share of students eligible for free or

reduced price school food?

Here is an example for schools' student percentage of eligible FRPM(K-12) and its corresponding Pencentage of students who have SAT scores above 1500.

```
sql()
```

```
select "County Code" ||"District Code"||"School Code" as CDSCode,"Percent (%) Eligible FRPM (K-12)",
Â Â Â Â PctGE1500
from frpm join satscores on satscores.cds==frpm."County Code" ||"District Code"||"School Code"
WHERE PctGE1500!="NA" & "Percent (%) Eligible Free (Ages 5-17)"!="NA"
```

CDSCode	Percent (%) Eligible FRPM (K-12)	PctGE1500
01100170109835	0.6578	52.94
01100170112607	0.4709	7.04
01611190106401	0.3059	80.56
01611190130229	0.2268	70.46
01611190130609	0.125	80
01611190130625	0.9193	4.35
01611190132878	0.4249	42.42
01611270130450	0.1941	76.4
01611430122697	0.7396	5.56
01611430131177	0.2935	74.44

Showing 10 out of 1280 rows

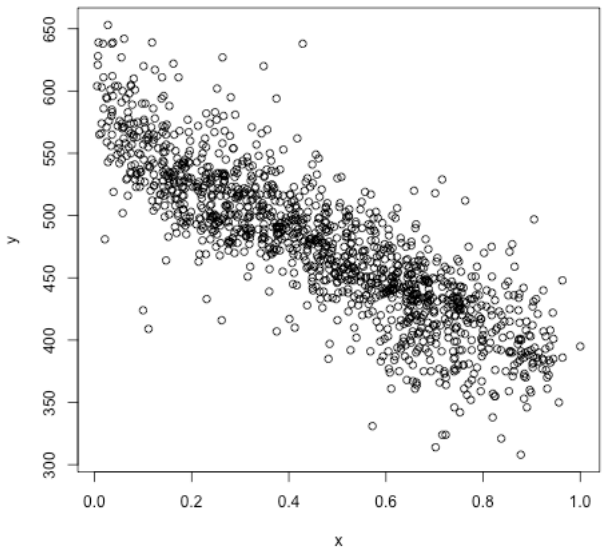
There is a strong negative correlation between percentage shares of students, between from K-12, who are elligible for FRPM with average total SAT scores.

```
table100 = sql()
```

```
select "Percent (%) Eligible FREE (Ages 5-17)" as pe5, AvgScrRead
from frpm join satscores on satscores.cds==frpm."County Code" ||"District Code"||"School Code"
WHERE AvgScrRead!="NA" & "Percent (%) Eligible FRPM (Ages 5-17)"!="NA"
```

```
r(table100)
```

```
x = table100 $ pe5
y = table100 $ AvgScrRead
plot(x,y)
```



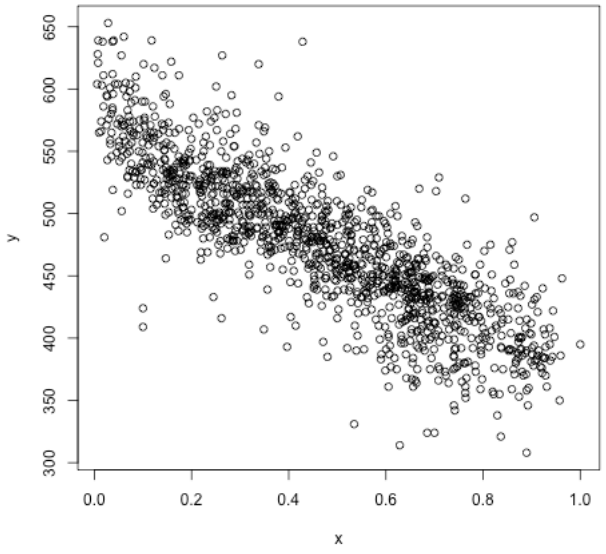
There is a strong negative correlation between percentage shares of students, between ages 5-17, who are elligible for free meals with average reading scores in the SAT.

```
table199 = sql()
```

```
select "Percent (%) Eligible FREE (K-12)" as pek, AvgScrRead
from frpm join satscores on satscores.cds==frpm."County Code" ||"District Code"||"School Code"
WHERE AvgScrRead!="NA" & "Percent (%) Eligible FREE (K-12)"!="NA"
```

```
r(table199)
```

```
x = table199 $ pek
y = table199 $ AvgScrRead
plot(x,y)
```



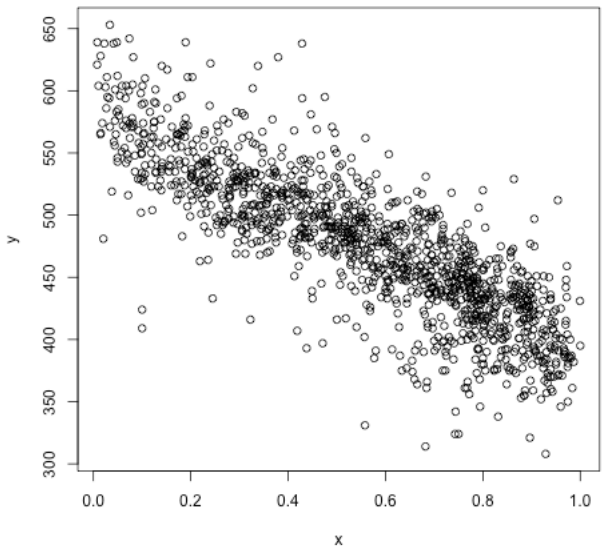
There is a strong negative correlation between percentage shares of students, between grades K-12, who are elligible for free meals with average reading scores in the SAT.

```
table6 = sql()
```

```
select "Percent (%) Eligible FRPM (K-12)" as pefk, AvgScrRead
from frpm join satscores on satscores.cds==frpm."County Code" ||"District Code"||"School Code"
WHERE AvgScrRead!="NA" & "Percent (%) Eligible FRPM (K-12)"!="NA"
```

```
r(table6)
```

```
x = table6 $ pefk
y = table6 $ AvgScrRead
plot(x,y)
```



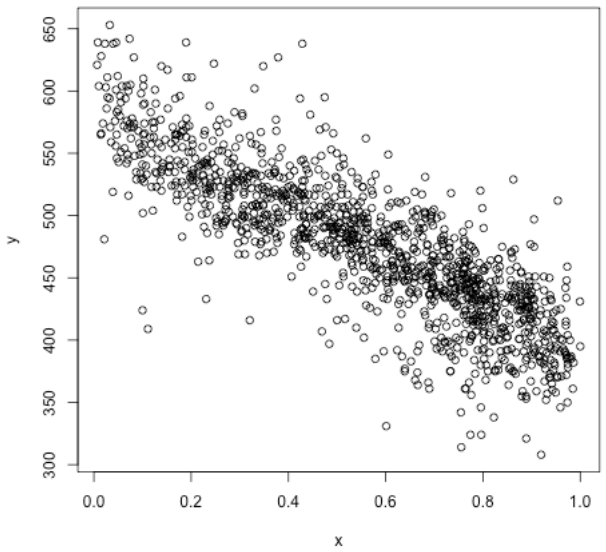
There is a strong negative correlation between percentage shares of students, between grades K-12, who are elligible for FRPM with average reading scores in the SAT.

```
table5 = sql()
```

```
select "Percent (%) Eligible FRPM (Ages 5-17)" as pe5, AvgScrRead
from frpm join satscores on satscores.cds==frpm."County Code" ||"District Code"||"School Code"
WHERE AvgScrRead!="NA" & "Percent (%) Eligible FRPM (Ages 5-17)"!="NA"
```

```
r(table5)
```

```
x = table5 $ pe5
y = table5 $ AvgScrRead
plot(x,y)
```



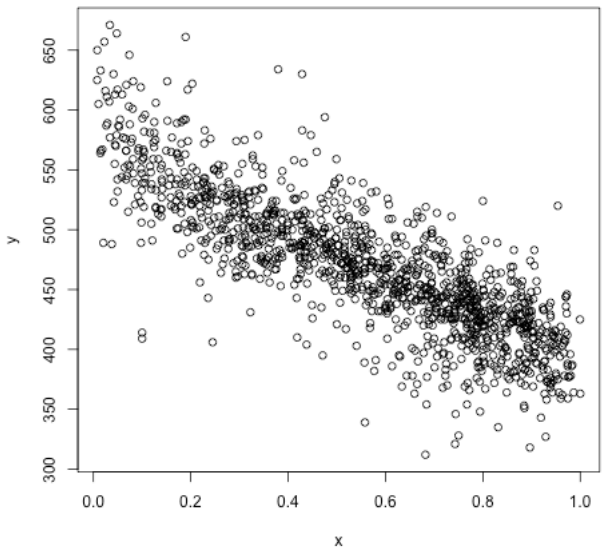
There is a strong negative correlation between percentage shares of students, between ages 5-17, who are elligible for FRPM with average reading scores in the SAT.

```
table13 = sql()
```

```
select "Percent (%) Eligible FRPM (K-12)" as pek, AvgScrWrite
from frpm join satscores on satscores.cds==frpm."County Code" ||"District Code"||"School Code"
WHERE AvgScrWrite!="NA" & "Percent (%) Eligible FRPM (K-12)"!="NA"
```

```
r(table13)
```

```
x = table13 $ pek
y = table13 $ AvgScrWrite
plot(x,y)
```



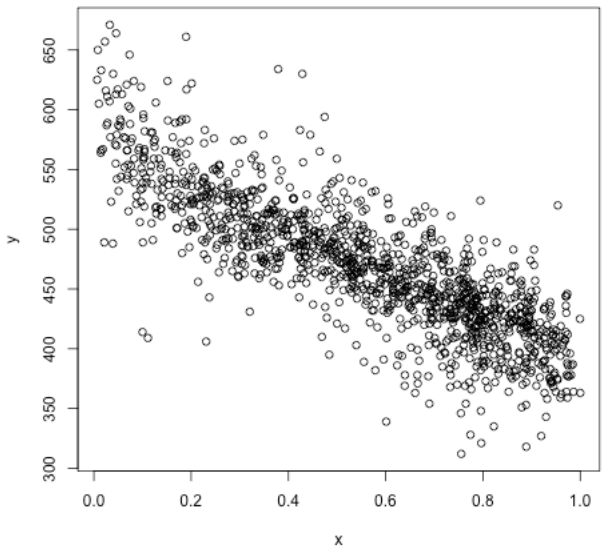
There is a strong negative correlation between percentage shares of students, between grades K-12, who are elligible for FRPM with average writing scores in the SAT.

```
table14 = sql()
```

```
select "Percent (%) Eligible FRPM (Ages 5-17)" as pe5, AvgScrWrite
from frpm join satscores on satscores.cds==frpm."County Code" ||"District Code"||"School Code"
WHERE AvgScrWrite!="NA" & "Percent (%) Eligible FRPM (Ages 5-17)"!="NA"
```

```
r(table14)
```

```
x = table14 $ pe5
y = table14 $ AvgScrWrite
plot(x,y)
```



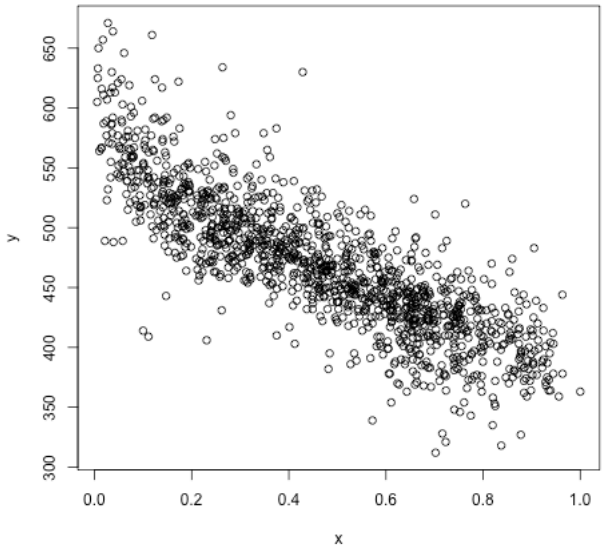
There is a strong negative correlation between percentage shares of students, between ages 5-17, who are elligible for FRPM with average writing scores in the SAT.

```
table15 = sql()
```

```
select "Percent (%) Eligible FREE (Ages 5-17)" as pe5, AvgScrWrite
from frpm join satscores on satscores.cds==frpm."County Code" ||"District Code"||"School Code"
WHERE AvgScrWrite!="NA" & "Percent (%) Eligible FREE (Ages 5-17)"!="NA"
```

```
r(table15)
```

```
x = table15 $ pe5
y = table15 $ AvgScrWrite
plot(x,y)
```



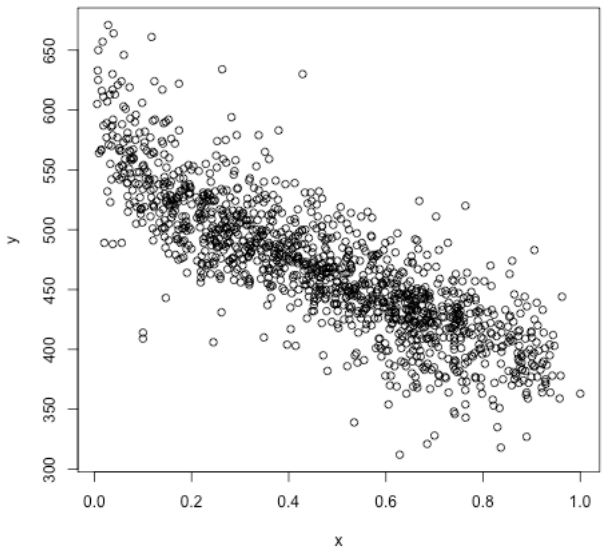
There is a strong negative correlation between percentage shares of students, between ages 5-17, who are elligible for free meals with average writing scores in the SAT.

```
table16 = sql()
```

```
select "Percent (%) Eligible FREE (K-12)" as pek, AvgScrWrite
from frpm join satscores on satscores.cds==frpm."County Code" ||"District Code"||"School Code"
WHERE AvgScrWrite!="NA" & "Percent (%) Eligible FREE (K-12)"!="NA"
```

```
r(table16)
```

```
x = table16 $ pek
y = table16 $ AvgScrWrite
plot(x,y)
```



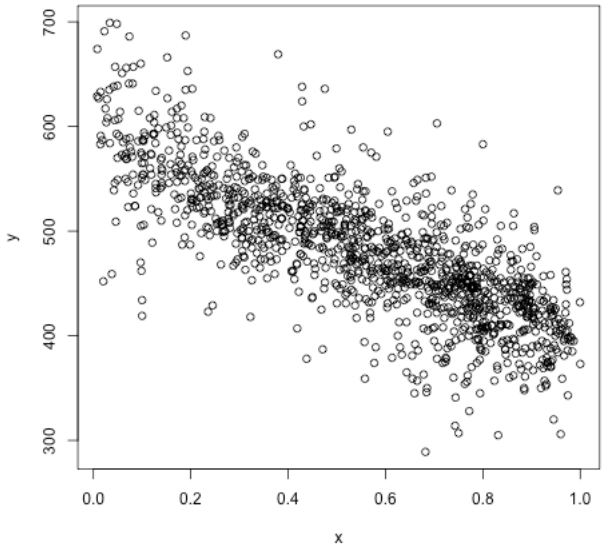
There is a strong negative correlation between percentage shares of students, between K-12, who are elligible for free meals with average writing scores in the SAT.

```
table27 = sql()
```

```
select "Percent (%) Eligible FRPM (K-12)" as pek, AvgScrMath
from frpm join satscores on satscores.cds==frpm."County Code" ||"District Code"||"School Code"
WHERE AvgScrMath!="NA" & "Percent (%) Eligible FRPM (K-12)"!="NA"
```

```
r(table27)
```

```
x = table27 $ pek
y = table27 $ AvgScrMath
plot(x,y)
```



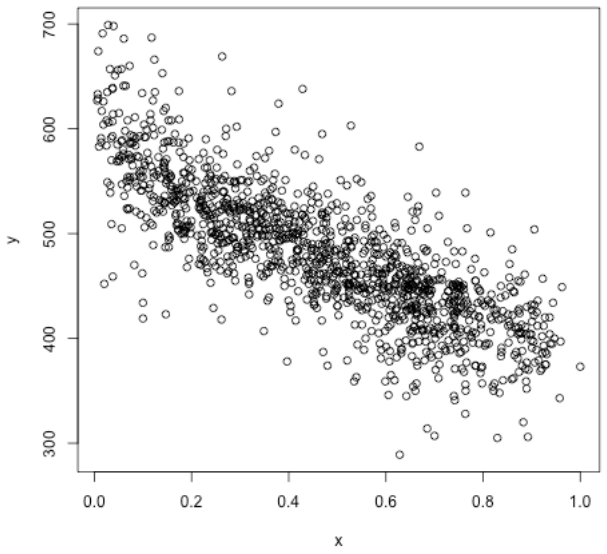
There is a strong negative correlation between percentage shares of students, between K-12, who are elligible for FRPM with average math scores in the SAT.

```
table9 = sql()
```

```
select "Percent (%) Eligible FREE (K-12)" as pek, AvgScrMath
from frpm join satscores on satscores.cds==frpm."County Code" ||"District Code"||"School Code"
WHERE AvgScrMath!="NA" & "Percent (%) Eligible FREE (K-12)"!="NA"
```

```
r(table9)
```

```
x = table9 $ pek
y = table9 $ AvgScrMath
plot(x,y)
```



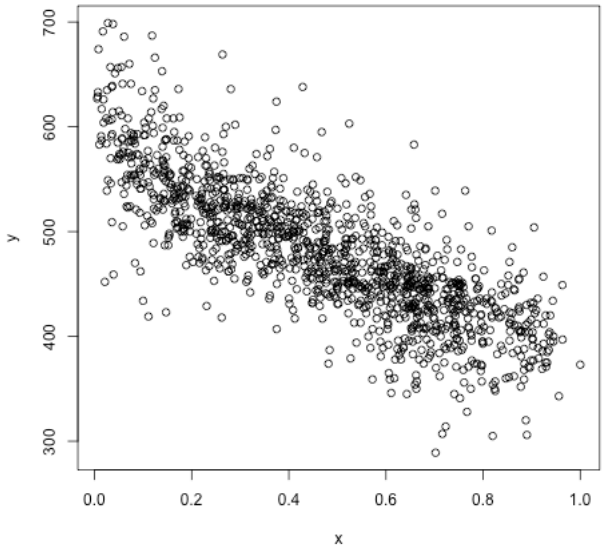
There is a strong negative correlation between percentage shares of students, between K-12, who are elligible for free meals with average math scores in the SAT.

```
table10 = sql()
```

```
select "Percent (%) Eligible FREE (Ages 5-17)" as pe5, AvgScrMath
from frpm join satscores on satscores.cds==frpm."County Code" ||"District Code"||"School Code"
WHERE AvgScrMath!="NA" & "Percent (%) Eligible FREE (Ages 5-17)"!="NA"
```

```
r(table10)
```

```
x = table10 $ pe5
y = table10 $ AvgScrMath
plot(x,y)
```



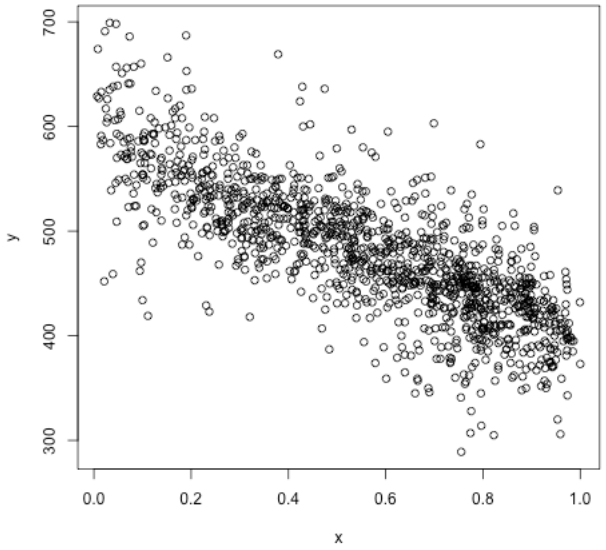
There is a strong negative correlation between percentage shares of students, between ages 5-17, who are elligible for free meals with average math scores in the SAT.

```
table11 = sql()
```

```
select "Percent (%) Eligible FRPM (Ages 5-17)" as pe5, AvgScrMath
from frpm join satscores on satscores.cds==frpm."County Code" ||"District Code"||"School Code"
WHERE AvgScrMath!="NA" & "Percent (%) Eligible FRPM (Ages 5-17)"!="NA"
```

```
r(table11)
```

```
x = table11 $ pe5
y = table11 $ AvgScrMath
plot(x,y)
```



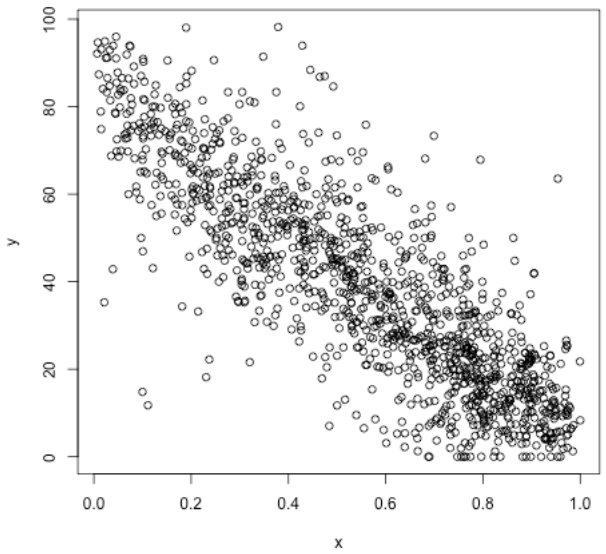
There is a strong negative correlation between percentage shares of students, between ages 5-17, who are elligible for FRPM with average math scores in the SAT.

```
table1 = sql()
```

```
select "Percent (%) Eligible FRPM (Ages 5-17)" as pe5, PctGE1500
from frpm join satscores on satscores.cds==frpm."County Code" ||"District Code"||"School Code"
WHERE PctGE1500!="NA" & "Percent (%) Eligible Free (Ages 5-17)"!="NA"
```

```
r(table1)
```

```
x = table1 $ pe5
y = table1 $ PctGE1500
plot(x,y)
```



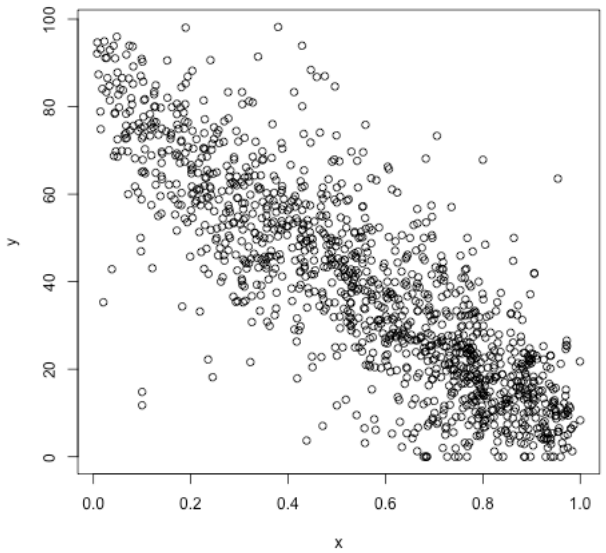
There is a strong negative correlation between percentage shares of students, between ages 5-17, who are elligible for FRPM with average total SAT scores.

```
table2 = sql()
```

```
select "Percent (%) Eligible FRPM (K-12)" as pefk, PctGE1500
from frpm join satscores on satscores.cds==frpm."County Code" ||"District Code"||"School Code"
WHERE PctGE1500!="NA" & "Percent (%) Eligible FRPM (Ages K-12)"!="NA"
```

```
r(table2)
```

```
x = table2 $ pefk
y = table2 $ PctGE1500
plot(x,y)
```



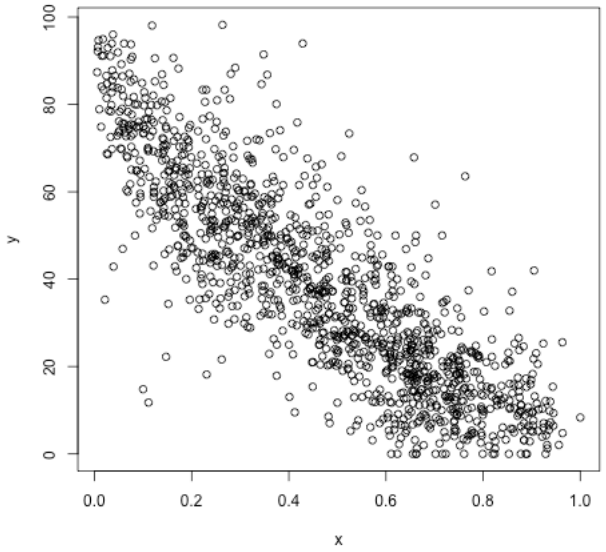
There is a strong negative correlation between percentage shares of students, between K-12, who are elligible for FRPM with average total SAT scores.

```
table3 = sql()
```

```
select "Percent (%) Eligible Free (Ages 5-17)" as pef5, PctGE1500
from frpm join satscores on satscores.cds==frpm."County Code" ||"District Code"||"School Code"
WHERE PctGE1500!="NA" & "Percent (%) Eligible Free (Ages 5-17)"!="NA"
```

```
r(table3)
```

```
x = table3 $ pef5
y = table3 $ PctGE1500
plot(x,y)
```



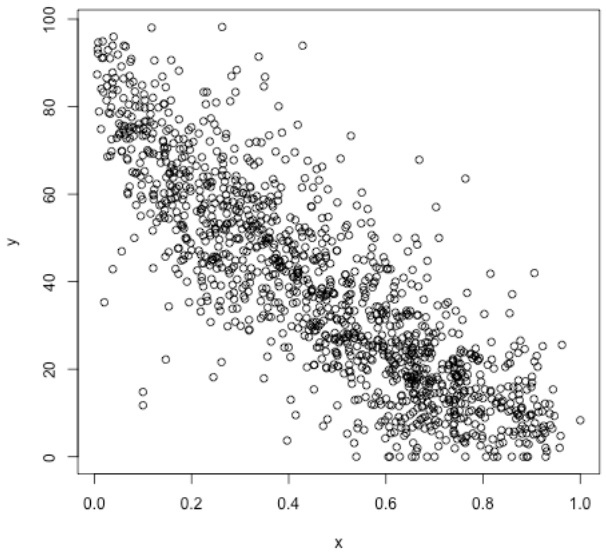
There is a strong negative correlation between percentage shares of students, between ages 5-17, who are elligible for free meals with average total SAT scores.

```
table4 = sql()
```

```
select "Percent (%) Eligible Free (K-12)" as pefk, PctGE1500
from frpm join satscores on satscores.cds==frpm."County Code" ||"District Code"||"School Code"
WHERE PctGE1500!="NA" & "Percent (%) Eligible Free (K-12)"!="NA"
```

```r
r(table4)
```

```r
x = table4 $ pefk
y = table4 $ PctGE1500
plot(x,y)
```



There is a strong negative correlation between percentage shares of students, between K-12, who are elligible for free meals with average total SAT scores.

Since a student must be sufficiently poor to be eligible for FRPM, schools with larger shares of free or FRPM eligble students imply that the students are relatively poor. Children living in poverty have lower qualities of life and are subject to burdens that may inhibit their ability to perform well at school. Moreover, they have less access to educationanl tools, such as tutoring services, that wealthier families can afford. Consequently, poorer students are more likely to underperform at school. Although the FRPM policy is meant to alleviate burdens of poverty, there is no evidence that FPRM makes poor students perform as well as their wealthier counterparts. However, this is not to say that FRPM has no effect on a student's productivity at school. It could be the case that FRPM has increased or even decreased the average student's ability to perform well in SATs, but this cannot be ascertained without a treatment and control group, and a before and after treatment effect sample. Finally, the relationship between FRPM and SAT scores may not exist beyond a correlation. This is because shares of students eligible for FRPM is a reflection of the shares of students in poverty, therefore, the relationship between FRPM and SAT scores may be spuriously driven by the poverty levels itself.

The correlation between shares of students eliglbe for FRPM and SAT scores is robust across an array of measurements. Students eligible for FRPM are calculated by age and by grade, while SAT scores are compared as a whole and by subject. This implies that the relationship is robust to measurement error.