### No Child Left Behind

J.A.Hartigan Yale University September 2013

### **1** Connectictut Schools

Primary and secondary education in Connecticut is carried out by 166 school districts funded by local property taxes. There are big variations in per capita income and in fractions of poor minorities over the various districts; a few big city districts have almost all the minority students. There is some equalisation provided by the state, so the amount of money spent per student, about \$17000 per year, is roughly the same in all districts. Nevertheless, there are big differences in academic performance between the various districts, with the poor minority students in urban areas one or two grades behind by the time they reach high school, and with 42% of Hispanics and 33% of African Americans failing to complete high school after 4 years.

In 2001 the Bush Administration, with bipartisan support, passed the No Child Left Behind Act, at

http://en.wikipedia.org/wiki/No Child Left Behind Act

requiring states to assess the academic performance of every child annually with standardised tests, and to identify and correct "failing" schools when the test scores for poor performing students fail to improve.

years	<-	2006:201	.2					
pBlack	<-	c(53, 57,	58,	64,	69,	68, '	70)	
pHispanic	<-	c(54, 57,	59,	63,	68,	68, '	70)	
pWhite	<-	c(89, 90,	91,	93,	94,	94, 9	95)	
pAsian	<-	c(92, 92,	93,	94,	94,	96, 9	95)	
sBlack	<-	c(218, 22	2, 22	28, 2	28,	235,	235,	235)
sHispanic	<-	c(219, 22	23, 22	24, 2	29,	235,	235,	236)
sWhite	<-	c(265, 26	59, 20	68, 2	273,	276,	276,	277)
sAsian	<-	c(279, 28	32, 28	82, 2	284,	288,	292,	291)

Proficiency fractions, 8th grade math: rbind(years, pBlack, pHispanic, pWhite, pAsian)

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]
years	2006	2007	2008	2009	2010	2011	2012
pBlack	53	57	58	64	69	68	70
pHispanic	54	57	59	63	68	68	70
pWhite	89	90	91	93	94	94	95
pAsian	92	92	93	94	94	96	95

Average Scale Scores, 8th grade math:

from <u>http://solutions1.emetric.net/CMTPublic/Index.aspx.</u>								
rbind(ye	ears,	sBla	ick,	sHis	panio	c, sW	hite,	sAsian)
	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	
years	2006	2007	2008	2009	2010	2011	2012	
sBlack	218	222	228	228	235	235	235	
sHispanic	219	223	224	229	235	235	236	
sWhite	265	269	268	273	276	276	277	
sAsian	279	282	282	284	288	292	291	

In both tables, we see substantial improvement over the first 4 years, and more improvement in minority groups than in whites and Asians. The improvement levels off in 2010. Already in 2006, 90% of the white and Asian students exceed the Proficient level, so there is not much possibility of improvement there. The better measure is the Scale Score, which approximates more accurately the proportion of questions answered by all students in a particular group. The standard deviation of scale score is about 40 for each group. For example, Asians improved their score by 17 points over the 7 years, nearly half a standard deviation, a substantial improvement. The "achievement gap" between Asians and Whites on the one hand, and Asians and Hispanics on the other is barely changed over the period..it begins at 47 in 2006 and drops to 42 in 2012. It may be that the effect of requiring standardised testing for all students is to improve the test scores initially, but the students and teachers have now adjusted to the testing process, and new gains will be small, as they have been for the last 3 years.

### 2 Adequate Yearly Progress

Since 2003, Connecticut has been following the federal guidelines, testing the children every year and keeping track of the results by race and ethnic group. The goal is to have 100% of the students achieving proficiency in mathematics by 2012. If a school is not making adequate yearly progress (AYP) towards that goal then various actions are required by the NCLB law. The data for AYP is retained at <u>http://www.ctreports.com</u>.

The Connecticut Department of education lists five levels of achievement: Under Basic, Basic, Proficient, Goal, Advanced. The description of proficiency suggests adequacy rather than proficiency:

Generally, fourth-grade students who perform at this level demonstrate adequate knowledge of grade-level content. These students demonstrate adequate conceptual understanding, computational skills and problem-solving skills, as well as an ability to solve complex and abstract mathematical problems. Typically, the solutions these students provide to math problems are adequate and include sufficient explanations.

The 8th grade math test has a raw score range of 0-146. Items in the test are worth 1 or 2 points. The lower limit of the "proficient group", 213, corresponds to a raw score of 62. In the test a person who answered only the arithmetic items correctly would get a score of 64. The raw score is converted into a "scale" score with range 100-400 according to the transformation in:

http://www.csde.state.ct.us/public/cedar/assessment/cmt/resources/misc\_cmt/ 2010%20CMT%20Score%20Conversion%20Tables%20%28Technical%20Bulleti n%29%208-17-10%20Final.pdf.

This transformation is not linear, but compresses the outlying raw scores. For classification purposes, the scale score is roughly the raw score plus 150. Then the scale score is converted into the five levels using break points 190, 213, 244, 286 on the raw scores, available at:

http://www.sde.ct.gov/sde/lib/sde/pdf/curriculum/cali/2010cmtinterpretivegui definal.pdf.

You are proficient if you score more than 63 on the raw score, that is,

43%. The various transformations obscure how mathematical ability is being measured. The arbitrary break points on an arbitrary scale can be manipulated to make sure that nearly everyone is proficient.( Counting the number of students who exceed a certain threshold level rather than comparing average scores reduces efficiency in comparisons between groups . For example, if the threshold level is chosen optimally, and if there is actually a one standard deviation difference between the groups, you need 30% more students to detect a difference between the groups in a standard hypothesis testing framework.)

The scale scores are approximately normal with mean 264 and standard deviation 50 in 2011 from:

#### http://solutions1.emetric.net/CMTPublic/CMTCode/Report.aspx?data=3494B57 834941FEAD3C67FFFA3A83B70

The expected scale score within each of the 5 performance intervals is 162,198, 227, 265, 319. From the distribution over the five categories available in the published data, we compute an expected scale score in which the expectations within each category are weighted by the reported number in each category. These scale scores provide a more direct indicator of student performance.

In 2012, minorities are 70% proficient, up from 53% in 2006, apparently good progress. On the other hand, some investigators have noted that a Proficient level on the Connecticut definition, (scale score between 213 and 244) can be Below Basic level in the NAEP( National Assessment of Educational Progress.) Roughly, Connecticut Proficiency matches NAEP Basic, Connecticut Goal matches NAEP Proficient. The Connecticut Proficient range is entirely below the Connecticut average score.

The NCLB goal of 100% must be met by all groups by 2014. As a result most Connecticut schools with any substantial minority enrollment have been declared to fail AYP; some have achieved "SafeHarbor", which means they are not on course to achieve the 100% goal but did have a modest improvement over the previous year's result. (The goal of 100% is a fantasy, and will never be met. One way people try to escape punishment for failing absurd goals is to

cheat; in April 2013, 35 teachers and staff in the Atlanta school district were arrested for improving the students' test scores by erasing the wrong answers and replacing them with correct ones.) We aren't going to close all the schools with high minority enrollments in 2014. Still, impracticable as the 100% goal is, there has been substantial improvement in all subgroups meeting the "proficiency" standard and in meeting the "goal" standard. We will look at the performance of students in the eighth grade math test as a function of race, ethnicity, and wealth.

# 3 Comparisons between ethnic and poverty categories, by minority fraction

### 3.1 Scale score vs minority fraction

The data set dma is extracted from the Connecticut Department of Education web site in the Data Preparation section. It consists of District level data giving average scale scores within each ethnic and poverty level group, averaged over the years 2006-2011.

```
dma <-read.csv("data/DistrictConn8Math.csv",</pre>
header=T, as.is=T)
tiff("pictures/Race and wealth.tif", w=900, h=950)
s <- 500
Grid( c(-5, seq(0,100,20), 110),
   c(195, seq(200, 320, 20)),
ylab=" Conn. 8th Grade Math/Scale Score/by minority
fraction", at=c(50, 0, 50))
# insert circles for each district and race by
poverty class
circle(minority,Black.pS,Black.pN,s,"black",3,dma)
circle(minority,Black.rS,Black.rN,s,"black",1,dma)
circle (minority, Asian.pS, Asian.pN, s, "green", 3, dma)
circle(minority,Asian.rS, Asian.rN,s,"green",1,dma)
circle(minority, Hispanic.pS,
    Hispanic.pN,s,"red",3,dma)
circle(minority, Hispanic.rS,
    Hispanic.rN,s,"red",1,dma)
circle(minority,White.pS,White.pN,s,"blue",3,dma)
circle(minority,White.rS,White.rN,s,"blue",1,dma)
# put in district names for poor districts
dd <- dma$District
```

```
dd[dma$minority < 30 | dma$minority > 97] <- ""</pre>
# sawtooth pattern for vertical placement of
district names
saw <- 300 + 2 * (rank(dma$minority) %% 10)</pre>
text(dma$minority, saw, dd, cex=1)
# identify race and ethnicity
text(10, 200, "White", col="blue", pos=4, cex=2)
text(10, 205, "Asian", col="green", pos=4, cex=2)
text(10, 215, "Hispanic", col="red", pos=4, cex=2)
text(10, 210, "Black", col="black", pos=4, cex = 2)
# identify rich and poor
text(30, 210, "rich", cex=2)
text(30, 206, "poor", cex=2)
points(37, 206, cex = 2.2)
points(37, 206, cex = 1.8)
points(37, 210, cex= 2)
```

dev.off()



The center of each circle is the District minority fraction on the x-axis, and the District Average Scale Score for a particular ethnic and poverty group. The areas of the circles of indicate the sizes of the various student groups. The races are indicated by color, the rich and poor students by the line thickness of the circles.

The Asians do best. The main effect visible is the segregation of the schools, with most minorities in very high minority districts, and most majority students in very low minority districts. There is a 60 point difference in the average scale scores between the minority and majority students.

### 3.2 Scale score by minority quintiles

```
tiff("pictures/Race and wealth quintiles.tif",
   w=1000, h=1000)
Grid(c(-5, seq(0,100,20), 110),
  c(195, seq(200, 300, 20), 305),
  ylab=" Quintile Averages/Conn.8th Grade
Math/Scale Score/By Minority Fraction",
   at=c(50, 50, -5, 50), cex=1.8)
# identify race and ethnicity
text(10, 200, "White", col="blue", pos=4, cex=2)
text(10, 205, "Asian", col="green", pos=4, cex=2)
text(10, 215, "Hispanic", col="red", pos=4, cex=2)
text(10, 210, "Black", col="black", pos=4, cex = 2)
text(30,206,"poor",cex=2);rect(36,205,38,207,1wd=2)
text(30,210,"rich",cex=2);rect(36,209,38,211)
# draw a average lines for the different regions
meanline(dma$minority, dma$Black.pS, dma$Black.pN,
 cutpoints=seq(0,100,20), lwd=3)
meanline(dma$minority, dma$Black.rS, dma$Black.rN,
 cutpoints = seq(0, 100, 20))
meanline(dma$minority, dma$Asian.pS, dma$Asian.pN,
 cutpoints=seq(0,100,20), col="green", lwd=3)
meanline(dma$minority, dma$Asian.rS, dma$Asian.rN,
          cutpoints=seq(0,100,20), col="green")
meanline (dma$minority, dma$White.pS, dma$White.pN,
 cutpoints=seq(0,100,20), col="blue", lwd=3)
meanline(dma$minority, dma$White.rS, dma$White.rN,
 cutpoints=seq(0,100,20), col="blue")
meanline(dma$minority,dma$Hispanic.pS,dma$Hispanic.
pN, cutpoints=seq(0,100,20), col="red", lwd=3)
meanline(dma$minority,dma$Hispanic.rS,dma$Hispanic.
rN, cutpoints=seq(0,100,20), col="red")
```

dev.off()



Within each quintile of the minority score, compute the average scale score in each race and poverty class, weighted by the number of persons in each such class. The average scores are connected by lines with color indicating ethnic group and thickness indicating poverty. The areas of the points are proportional to the numbers of students in each quintile average. Note the many whites in the rich low minority quintile, and many minorities in the poor high minority quintile.

## **3.2 Scale Score for poverty, ethnicity, and minority fraction quintiles**

11

```
meanscore <- function(x) {</pre>
# get means and sums 8 poverty-ethnic categories
x <- na.omit(x)
mv < - rep(0, 16)
ms < - mv
for (i in c(5:8, 13:16) ) {
ms[i] <- sum(x[ ,i-4])</pre>
if(ms[i]>0) mv[i]<-sum(x[,i]*x[,i-4])/ms[i]
}
mv <- as.character(round(mv))</pre>
mvv <- paste(mv[c(6:8,5)],mv[c(14:16,13)],sep=":")</pre>
ms <- as.character(round(ms))</pre>
mss <- paste(ms[c(6:8,5)],ms[c(14:16,13)],sep=":")</pre>
mvv < - c(mvv, mss)
return(mvv)
}
means <- data.frame(matrix("", nrow=5, ncol=8),</pre>
                   stringsAsFactors=F)
for( row in 1:5) means[row, ] <-</pre>
meanscore(dma[floor(dma$minority/20) == row-1,
2:17])
names(means) <- c( "Black", "Hispanic", "White",</pre>
"Asian", "nBlack", "nHispanic", "nWhite", "nAsian")
row.names(means) <-</pre>
c("0-20%", "20-40%", "40-60%", "60-80%", "80-100%")
print(means[, 5:8])
          nBlack nHispanic
                              nWhite
                                       nAsian
        1173:2127 1900:3665 7931:82978 644:4298
0-20%
       1233:1088 1566:902 2286:10146 317:676
20-40%
```

40-60% 3849:2287 4563:1979 2744:9762 508:947

80-100% 10599:2180 10966:981 1488:820 388:83

4694:2984 8625:2422 2342:5164 328:460

60-80%

The table gives the total number of students during 2006-2011 in the various ethnic poor and ethnic rich classes. About 50% of the Blacks and 30% of the Hispanics are in the highest minority fraction group. About 75% of the Whites, and 60% of the Asians are in the lowest minority fraction group.

```
print(means[, 1:4])
```

```
Black Hispanic
                          White
                                  Asian
0-20%
                 238:256 251:277 275:300
        232:251
        226:240
                 229:246 243:268 266:285
20-40%
40-60%
        221:234
                 227:241 242:266 263:285
60-80%
        218:232
                 214:230 236:263 252:279
80-100% 217:227
                 216:224 245:267 259:277
```

A 20 point rule sums up many of the differences: the average difference in scale scores between Blacks & Hispanics, Whites, and Asians at the same poverty level is about 20 points. The average difference between poor and rich students of the same ethnicity is about 20 points. The average difference between high majority districts and high minority districts is about 20 points for the same poverty levels and each ethnic group. For example, the average difference between poor Black students in high minority districts and rich White students in high majority districts is about 60 points.

The differences between rich and poor are compressed in the minority districts; the average difference there is only 14. The Hispanics benefit more than Blacks by being in the near majority districts; for example, the poor Hispanics increase by 23 points going from minority to majority districts, and the rich Hispanics by 28; while the poor Blacks increase by 16 and the rich Blacks by 22.

If all districts were equally integrated, all districts would be similar to the present 20-40% minority districts. The Blacks and Hispanics would gain about 9 points, the Hispanics would gain 13 points, the Whites would lose 9 points, and the Asians would lose 15 points.

# 4 Effect of per capita income and minority fraction on scale scores

### 4.1 Rich Whites

(Rich: students who are not eligible for free lunch.)

To discount the effects of very large per capita income in a few districts near New York, we use logpci: dma\$logpci <- log(dma\$pci)

The regression to predict rich white scores is weighted by the number of rich white students. This ensures that the same coefficients would be obtained if the regression were done on the scale scores for each individual student:

```
rwcs<-lm(
White.rS~logpci+minority,weights=White.rN,data=dma)
options(digits=3)
summary(rwcs)$coef[,c(1,3)]</pre>
```

	Estimate t	value
(Intercept)	50.166	2.53
logpci	21.989	11.60
minority	-0.151	-3.70

Some pci values are missing; we identify which values are used in the regression by names(rwcs\$fit). We use line width 2 (lwd=2) in the plot so that the very small circles corresponding to very few students remain visible.

```
tiff("pictures/Rich Whites.tif", w=900, h=800)
Grid(c(238, seq(240,300, 10)),
    c(238, seq(240, 300, 10)),
    ylab=" Prediction from income and minority
fraction/Rich White/ Scale Score",
    at=c(270,240,240 ))
points(rwcs$fit, rwcs$fit+rwcs$res, lwd=2,
    cex=sqrt(dma[names(rwcs$fit), "White.rN"])/10)
```

```
# The fitted line:
abline(reg=lm(rwcs$fit+rwcs$res~rwcs$fit),col="blue
")
arrows(295, 270, 295, 280)
text(pos=4, 283, 265, "Number of Students", cex=2)
dev.off()
```

Prediction from income and minority fraction Rich White Scale Score



Examine the high residual and the two influential values in the left of the plot: dma[names(rwcs\$res)[rwcs\$fit < 255], ][, c(1, 13, 17, 18, 19, 21)]

	District	White.rN	White.rS	Total	minority	logpci
10	Bridgeport	48	241	8550	89.3	9.70
<b>4</b> 8	Hartford	. 313	271	9008	91.2	9.51
114	Waterbury	746	249	7352	2 75.8	9.78
126	Windham	270	254	1378	67.2	9.74

Hartford has a rather high white.rS value for such a large minority population. Two of the Hartford middle schools are about 70% minority, whereas the rest are near 100% minority; about 100 of the white students were at those 70% minority schools. We only have district level data, so we can't adjust for big differences in ethnicity between schools within districts. The surprisingly high Hartford scores are due to those schools.

```
Make indicator variables for outliers and influential values:
dma$bridgeport <- dma$District == "Bridgeport"
dma$hartford <- dma$District == "Hartford"
dma$waterbury <- dma$District == "Waterbury"
dma$windham <- dma$District == "Windham"
```

Also include a squared logpci variable to handle the apparent nonlinearity of the plot; for higher values of logpci, the same increases in logpci cause smaller increases in scale score than for lower values of logpci. Subtract the mean log pci from logpci before squaring to get near orthogonality between the square term and the linear term:

```
mlogpci <- mean(dma$logpci, na.rm=T)
print(mlogpci)</pre>
```

[1] 10.3

```
dma$logpci2 <- (dma$logpci - mlogpci)^2</pre>
```

Estimate 1	t value
-27.89	-0.9049
32.60	11.0324
-17.84	-4.3531
-0.16	-3.8661
L -1.13	-0.0844
-21.74	-3.7794
-1.94	-0.5239
-4.81	-0.8309
	Estimate ( -27.89 32.60 -17.84 -0.16 L -1.13 -21.74 -1.94 -4.81

We identify observations as outliers in a sample of size 163 when the absolute t-value for the indicator variable exceeds sqrt(2 log(163))=3.19, (the approximate size of the largest of 163 independent unit normals). Following that standard, we exclude only the Hartford case from the model (paradoxically, by including the Hartford outlier indicator in the model.) We accept the square term also.

```
Estimate t value(Intercept)-29.880-1.07logpci31.92011.34logpci2-16.859-4.33minority-0.149-3.87hartford1-20.787-3.73
```

tiff("pictures/Rich white curved prediction.tif", w=900, h=800) Grid(c(238, seq(240,300, 10)), c(238, seq(240, 300, 10)), ylab=" Curved Prediction from income and minority fraction/Rich White/ Scale Score", at=c(270, 240, 240 )) points(shrwcs\$fit, shrwcs\$fit+shrwcs\$res, lwd=2, cex=sqrt(dma[names(shrwcs\$fit),"White.rN"])/10)

# The fitted line: abline(reg=lm(shrwcs\$fit+shrwcs\$res~shrwcs\$fit),col ="blue") dev.off()

Curved Prediction from income and minority fraction Rich White Scale Score



We see that the curvature is removed, that most of the scale scores are over 270, and there are some interesting deviations above and below the fitted curve.

To understand the effect of minority enrollment on rich white student scale score, we also need to know the relation between minority and pci:

```
summary(lm(logpci~minority, data=dma,
    weights=Total))$coef[,c(1,3)]
        Estimate t value
(Intercept) 10.46698 250.72
minority -0.00813 -8.14
```

Thus a 10% increase in minority causes -8% change in pci.

### **Effects on Rich White Students:**

pci: +10% -> +3.2 change in scale score (at median pci) direct minority: +10% -> -1.5 change in scale score overall minority: +10% -> -1.5 - 0.8\*3.2 = -4.1 change in scale score

So, for example, if a Rich White Student moved from a school with no minorities, to a school with all minorities, (100% change in minority), that school would be in a low pci district, and that student's scale score is predicted to decline by 41.

### 4.2 Rich Black students

```
rbcs <- lm(Black.rS~logpci + minority,</pre>
             weights=Black.rN, data=dma)
summary(rbcs) $coef[, c(1,3)]
         Estimate t value
(Intercept) 150.97
                    4.60
logpci
             9.63
                    3.06
           -0.24 -7.14
minority
tiff("pictures/Rich black students.tif", w=900,
h=900)
Grid(c(208, seq(210, 260, 10)),
c(198, seq(200, 300, 10)),
ylab=" Prediction from income and minority
fraction/Rich Black/ Scale Score",
at=c(230, 210, 210))
points(rbcs$fit, rbcs$fit+rbcs$res, lwd=2,
cex = sqrt(dma[names(rbcs$fit), "Black.rN"])/10 )
# The fitted line:
abline(reg=lm(rbcs$fit+rbcs$res~rbcs$fit),col="blue
")
dev.off()
```



We see that these values are nearly all below 240. The dots in the high majority districts correspond to very small counts of rich black students in those districts. The 5 or 6 large circles correspond to the city districts that contain nearly all the black students, rich or poor.

**Effects on Rich Black Students:** 

pci: +10% -> +1.0 change in scale score direct minority: +10% -> -2.4 change in scale score overall minority:  $+10\% \rightarrow -2.4 - 0.8*1.0 = -3.2$  change in scale score Thus the effect of minority fraction on Rich Black Students is somewhat less than for Rich White Students.

#### **4.3 Poor White Students**

```
pwcs <- lm(White.pS ~ logpci + minority ,</pre>
               weights=White.pN, data=dma)
summary(pwcs) $coef[, c(1, 3)]
         Estimate t value
(Intercept) 106.8771 3.23
loqpci
                    4.38
          14.1247
minority -0.0917 -2.87
tiff("pictures/Poor White Students.tif", w=900,
h=500)
Grid(c(228, seq(230,270, 10)),
c(218, seq(220, 290, 10)),
ylab=" Prediction from income and minority
fraction/Poor White/ Scale Score",
at=c(250, 230, 230))
points(pwcs$fit, pwcs$fit+pwcs$res, lwd=2,
cex = sqrt(dma[names(pwcs$fit), "White.pN"])/10 )
# The fitted line:
abline(reg=lm(pwcs$fit+pwcs$res~pwcs$fit),col="blue
")
dev.off()
```



Poor white Effects:

pci: +10% -> +1.4 change in scale score direct minority: +10% -> -0.9 change in scale score overall minority: +10% -> -1.4 - 0.8\*0.9 = -2.1% change in scale score

The overall negative effect of minority percent is less for poor white students than for rich white or black students.

### 4.4 Poor Black students

	Estimate t	value
(Intercept)	190.051	7.02
logpci	3.798	1.48
minority	-0.135	-4.18

```
tiff("pictures/Poor Black Students.tif", w=900,
h=900)
Grid(c(208, seq(210,230, 10), 235),
c(158, seq(160, 300, 20)),
ylab=" Prediction from income and minority
fraction/Poor Black/ Scale Score",
at=c(225, 210, 210 ))
```

```
points(pbcs$fit, pbcs$fit+pbcs$res, lwd=2,
cex = sqrt(dma[names(pbcs$fit), "Black.pN"])/10)
```

```
# The fitted line:
abline(reg=lm(pbcs$fit+pbcs$res~pbcs$fit),col="blue
")
dev.off()
```



Effects of minority and pci on poor blacks: pci: +10% -> +0.3 change in scale score direct minority: +10% -> -1.3 change in scale score overall minority: +10% -> -1.3 -0.8\*0.3 = -1.5 change in scale score

The average scale score is affected by minority fraction about equally for rich white and rich black students, and quite a bit less for poor white and poor black students. In these data, almost all the poor blacks are concentrated in just three poor high minority districts, so there is not much data to show what happens for poor blacks in low minority rich districts. In particular, the coefficient for logpci is not significantly different from zero; there are not enough blacks living in high pci districts to determine what the effect due to pci would be.

### 5 Change

In order to examine the progress of "no child left behind" we use the data containing results by years:

```
dm <- read.csv("data/DistrictYearConnMath8.csv",
header=T, as.is=T)
```

Some districts lacking either 2006 or 2011 must be eliminated. Find the table for district and year, select districts without 6 years, and then eliminate those districts from dm.

```
tt <- table(dm$District, dm$Year)</pre>
dy <- apply(tt, 1, sum)</pre>
names (dy) [dy < 6]
[1] "Achievement First" "Cornwall"
[3] "Elm City Col Prep" "Highville"
[5] "Jumoke Academy"
                       "LEARN"
[7] "New Beginnings"
                       "Park City Prep"
[9] "The Bridge Academy"
dm <- dm[!dm$District %in% names(dy)[dy < 6],]</pre>
tiff("pictures/change.tif", w=1000, h=700)
Grid(c(-5, seq(0, 100, 20), 110),
   c(175, seq(180, 340, 20)),
   ylab ="Change in 8th grade math 2006-2011 /by
minority fraction/Scale Score", at=c(50, 50, -5))
first <- dm$Year == 2006
last <- dm$Year == 2011
x <- (dm$minority[first] + dm$minority[last])/2</pre>
quad(x, diff=5,
dm$Black.pN[first]/25, dm$Black.pS[first],
dm$Black.pN[last]/25, dm$Black.pS[last],
minwt=1, lwd=2)
quad(x, diff=5,
dm$Black.rN[first]/25, dm$Black.rS[first],
dm$Black.rN[last]/25, dm$Black.rS[last],
```

```
minwt=1, lwd=1)
quad(x, diff=5,
dm$Hispanic.pN[first]/25, dm$Hispanic.pS[first],
dm$Hispanic.pN[last]/25, dm$Hispanic.pS[last],
minwt=1, lwd=2, col="red")
quad(x, diff=5,
dm$Hispanic.rN[first]/25, dm$Hispanic.rS[first],
dm$Hispanic.rN[last]/25, dm$Hispanic.rS[last],
minwt=1, lwd=1, col="red")
quad(x, diff=5,
dm$Asian.pN[first]/25, dm$Asian.pS[first],
dm$Asian.pN[last]/25, dm$Asian.pS[last],
minwt=1, lwd=2, col="green")
quad(x, diff=5,
dm$Asian.rN[first]/25, dm$Asian.rS[first],
dm$Asian.rN[last]/25, dm$Asian.rS[last],
minwt=1, lwd=1, col="green")
quad(x, diff=5,
dm$White.pN[first]/25, dm$White.pS[first],
dm$White.pN[last]/25, dm$White.pS[last],
minwt=1, lwd=2, col="blue")
quad(x, diff=5,
dm$White.rN[first]/25, dm$White.rS[first],
dm$White.rN[last]/25, dm$White.rS[last],
minwt=1, lwd=1, col="blue")
text(0, 190, "White" , col="blue", pos=4, cex=2)
text(0, 195, "Asian", col="green", pos=4, cex=2)
text(0, 185, "Hispanic", col="red",pos=4, cex=2)
text(0, 180, "Black", col="black", pos=4, cex=2)
text(30, 180, "poor", pos=4, cex=2)
rect(40, 178, 44, 182, lwd=2)
text(30, 186, "rich",pos=4,cex=2)
rect(40, 184, 44, 188)
# put in district names for poor districts
dd <- dm$District
                                                 11 11
dd[dm$minority < 30 | dm$minority > 97 ] <-
```

# sawtooth pattern for placement of district names saw <- 310 + 3 \* (rank(dm\$minority[first]) %% 10) text(dm\$minority[first], saw, dd[first], cex=1) dev.off()



Only groups of students with at least 25 students in 2006 and in 2011 are included. The left vertical edge of each quadrilateral has length proportional to the number of students in 2006, and average position equal to the average scale score in 2006; and the right vertical edge similarly shows 2011 counts and scores. The horizontal shift in each quadrilateral is the same. Thus we can see both whether or not scores are improving, and whether or not the school system is losing count in some groups. Most of the quads point upwards showing remarkable improvement over the six years, for all groups. Overall, the minorities are improving a little more than the majority groups, perhaps because the No Child Left Behind program emphasizes improving performance of the lowest scorers. The Asians are the one

#### 29

group whose scores are declining. There are bigger improvements for the minorities in the integrated districts than in the minority districts. The Hispanics and poor whites seem to be improving faster than the blacks in the integrated districts. Rich white students are declining in number in all the integrated districts.

### **6 Conclusions: Failing Schools**

The state has successfully improved scale scores in almost all groups, and especially in the minority groups. The only "failing" schools are those few where ethnic-poverty groups, of size of at least 24 in both 2006 and 2011, have declines in scores between 2006 and 2011. (A list based on percentage meeting goal rather than scale scores is similar). Here they are:

```
changeScore <- function(dm, eth, mult=0) {</pre>
# identify districts where ethS is decreasing
significantly
# sd of 40 in scale scores is based on expected
score computation # mult is the standard error
multiplier in comparisons
# fix up ethnic poverty names
names(dm) <- gsub("pN", "poor:", names(dm))</pre>
names(dm) <- gsub("rN", "rich:", names(dm))</pre>
year1 <- dm$Year == 2006
year2 <- dm$Year == 2011</pre>
ethN <- dm[, eth]
ethS <- dm[, eth + 4]
se <- 40 * sqrt( 1/(ethN[year1]+.00001) +</pre>
1/(ethN[year2] + .00001))
sem <- mult * se</pre>
bad <- (ethS[year2] + mult * se < ethS[year1]) &</pre>
(ethN[year1] > 24) \& (ethN[year2] > 24)
return(c(names(dm)[eth], dm$District[year1][bad]))
}
```

```
Run through ethnic poverty combinations:
for( i in c(3:6, 11:14) ) print(changeScore(dm, i))
```

```
[1] "Asian.poor:" "Bridgeport"
[1] "Black.poor:" "New Britain"
[1] "Hispanic.poor:" "East Hartford" "New London"
[4] "West Haven" "CREC"
[1] "White.poor:" "Derby" "New Milford"
```

```
[1] "Asian.rich:" "Greenwich" "Stamford"
[1] "Black.rich:"
[1] "Hispanic.rich:" "New Britain"
[1] "White.rich:" "Derby" "East Haddam"
[4] "East Hartford" "East Windsor" "Old Saybrook"
[7] "Windham"
```

The usual statistical test of significance for the comparison between groups (standard error multiplier = 2) exhibits no ethnic-poverty group with significant declines. There are a few groups with declining average scores. We see no rich black groups with declining scores, but that is because there are few rich black groups to begin with. There are 6 rich white groups with declining scores. This is just accidental variation. The "No child Left Behind Program" intends to improve groups with very low average scores, with program success being 100% of the groups reaching the "proficient" level, which still remains below average for the state as a whole. It does not have much effect on the students who are already scoring above average, but overall, their scores have improved too. The Black and Hispanic districts with declining scores do deserve scrutiny within the program.

### 7 Data Preparation

Scraping 8th grade Math scores for Connecticut Students:

The detailed data was obtained from the Connecticut state department of education website:

http://solutions1.emetric.net/cmtpublic/Index.aspx.

The state does not usually provide data about racial groups of size less than 20 within districts, which makes it impossible to evaluate how well minorities are doing in the different school districts. The state did not follow the size 20 rule in reporting the percent meeting the state goal( one level up from proficient) for the different minority groups, separately for students eligible for free lunch (poor) and students not eligible for free lunch(rich), for all school districts.

In addition we need data on per capita income in the various districts available at:

http://www.cslib.org/pathfinders/elemed.htm.

The web site produces a graphical display for performance of poor students in the different racial/ethnic categories; the page source for the HTML making that display is saved as "poorConn8MathSource.txt". Similarly, the page source for the the rich students is saved as "richConn8MathSource.txt".

These files are in the directory ~No Child Left Behind/data: list.files("data/")

- [1] "conn8math.csv"
- [2] "Conn8MathData.csv"
- [3] "DistrictConn8Math.csv"
- [4] "DistrictPerCapitaIncome2011.csv"
- [5] "DistrictYearConn8MathR.txt"
- [6] "DistrictYearConnMath8.csv"
- [7] "poorConn8MathSource.txt"
- [8] "richConn8MathSource.txt"

### 7.1 Scrape source file "poorConn8MathSource.txt"

First scan in all data as character strings, then pick out those strings
containing "showGraphs" that contain data:
poor <scan("data/poorConn8MathSource.txt",what="",sep="\n
")</pre>

```
uselines <- grep("showGraphs", poor)
poor[uselines] [1]</pre>
```

#### [1] "\'5ct\'5ct<td

```
class=\"barTestContent\">   
Mathematics8<td
class=\"barYearContent\">2006<a
href=\"javascript:showGraphs('State','2006','8','Mathematic
s','',11935,23.8,21.4,28.3,21.7,4.9);\"><img
class=\"barNContent\">11935<td
class=\"barGraphContent\"><img</pre>
onMouseMove=\"ChartshowTip('23.8',event)\"
onMouseOut=\"CharthideTip();\" src=\"../images/qlbar.gif\"
style=\"border-width:0px;height:17px;width:96px;\"/><img</pre>
onMouseMove=\"ChartshowTip('21.4',event)\"
onMouseOut=\"CharthideTip();\" src=\"../images/q2bar.gif\"
style=\"border-width:0px;height:17px;width:86px;\"/><img</pre>
onMouseMove=\"ChartshowTip('28.3',event)\"
onMouseOut=\"CharthideTip();\" src=\"../images/q3bar.gif\"
style=\"border-width:0px;height:17px;width:114px;\"/><img</pre>
onMouseMove=\"ChartshowTip('21.7',event)\"
onMouseOut=\"CharthideTip();\" src=\"../images/q4bar.gif\"
style=\"border-width:0px;height:17px;width:87px;\"/><img</pre>
onMouseMove=\"ChartshowTip('4.9',event)\"
onMouseOut=\"CharthideTip();\" src=\"../images/q5bar.gif\"
style=\"border-
width:0px;height:17px;width:17px;\"/>
class=\"barTotalContent\">100.0"
```

The first line of data is for the whole state, 11935 students, with the percentages in the five different performance levels. For example the percentage meeting "goal", the two highest levels, is 21.7 + 4.9 = 26.6.

Extract the percentage data between the parentheses, and add a "poor" to the string:

```
betweenpar<-regexpr("\(.*[0-9]\);",poor[uselines])
spoor <- regmatches(poor[uselines], betweenpar)
spoor <- gsub("\(|\)|;", "", spoor)
pspoor <- paste("poor", spoor, sep=",")
pspoor[1]
[1]
"poor,'State','2006','8','Mathematics','',11935,23.8,21.4,2
8.3,21.7,4.9"</pre>
```

```
Repeat the same operations on the rich students source file "rich.txt":
rich <- scan("data/richConn8MathSource.txt",</pre>
what="", sep="n")
uselines <- grep("showGraphs", rich)</pre>
betweenpar<-regexpr("\(.*[0-9]\);",rich[uselines])</pre>
srich <- regmatches(rich[uselines], betweenpar)</pre>
srich <- gsub("\setminus(|\setminus)|;", "", srich)
rsrich <- paste("rich", srich, sep=",")</pre>
rsrich[1]
[1]
"rich, 'State', '2006', '8', 'Mathematics', '', 32009, 5.0, 7.2, 17.
7,39.4,30.8"
Combine the two files and remove the inconvenient quotes :
pr <- c(pspoor,rsrich)</pre>
pr <- gsub("'", "", pr)</pre>
c(head(pr, 1), tail(pr, 1))
```

```
[1]
"poor,State,2006,8,Mathematics,,11935,23.8,21.4,28.3,21.7,4
.9"
[2] "rich,Elm City Col Prep,2011,8,Mathematics,Two or more
races,1,0.0,0.0,0.0,0.0,100.0"
```

The quotes are gone, and the poor and rich are combined.

Concatenate the data out to a csv file, convenient for comma delimited data: cat(pr, file = "data/conn8math.csv", sep="\n")

### 7.2 Edit data frame

```
Read from here after first scrape, getting a data frame:
dpr <- read.csv("data/conn8math.csv", header=F,
as.is=T)
dpr[1:5, ]
```

```
V1
         V2
             V3 V4
                             V5
                                            V6
                                                  V7
                                                       V8
1 poor State 2006 8 Mathematics
                                                11935 23.8
2 poor State 2006 8 Mathematics Asian American
                                                  327
                                                       7.0
3 poor State 2006
                  8 Mathematics
                                                 3830 29.1
                                          Black
4 poor State 2006 8 Mathematics
                                       Hispanic
                                                 4473 28.4
5 poor State 2006 8 Mathematics
                                    Am. Indian
                                                   42 16.7
   V9 V10 V11 V12
1 21.4 28.3 21.7
                 4.9
2 10.7 23.5 36.1 22.6
3 24.6 28.0 15.8 2.5
4 23.9 28.3 17.0
                 2.4
5 11.9 42.9 21.4
                 7.1
```

```
Eliminate state entries:
dpr <- dpr[dpr[,2] != "State", ]
dpr[1:5, ]
```

```
V1
           V2
                V3 V4
                                V5
                                               V6 V7
                                                        V8
39 poor Ansonia 2006 8 Mathematics
                                                   112 19.6
40 poor Ansonia 2006 8 Mathematics Asian American
                                                     3
                                                       0.0
41 poor Ansonia 2006 8 Mathematics
                                                   30 23.3
                                            Black
42 poor Ansonia 2006 8 Mathematics
                                         Hispanic
                                                   35 25.7
43 poor Ansonia 2006 8 Mathematics
                                       Am. Indian
                                                     1
                                                        0.0
    V9 V10 V11
                   V12
39 21.4 35.7 17.0
                    6.3
40 0.0 33.3 66.7
                   0.0
41 40.0 23.3 6.7
                    6.7
42 20.0 40.0 14.3
                    0.0
43 0.0 0.0 0.0 100.0
```

Replace recent expanded Race/Ethnicity names with original names, and elide sparse data with American Indian or Pacific Islander ethnicity:

```
eth <- dpr[, 6]
eth[eth == "Am Ind or AK Native"] <- "Am. Indian"
eth[eth == "Black or African Am"] <- "Black"
eth[eth == "Hisp/Lat or any race"] <- "Hispanic"
eth[eth == "Two or more races"] <- "White"
eth[eth == "Nat of HI or Pac Isl"] <- NA
eth[eth == "Asian American"] <- "Asian"
eth[eth == "Am. Indian"] <- NA
eth[eth == ""] <- NA
dpr[, 6] <- eth
table(eth)
```

```
eth
```

Asian	Black Hi	<b>spanic</b>	White
1197	1427	1540	1936

The differences in counts of ethnic names occurring in the different districts is caused by school districts having zero count for some minorities. Remove NAs: dpr <- na.omit(dpr) head(dpr)

```
V2
                V3 V4
                               V5
                                        V6 V7
                                                V8
                                                     V9
    V1
40 poor Ansonia 2006 8 Mathematics
                                      Asian 3 0.0 0.0
41 poor Ansonia 2006 8 Mathematics
                                      Black 30 23.3 40.0
42 poor Ansonia 2006 8 Mathematics Hispanic 35 25.7 20.0
44 poor Ansonia 2006 8 Mathematics
                                      White 43 14.0 11.6
                                      Asian 3 33.3 33.3
46 poor Ansonia 2007
                     8 Mathematics
                                      Black 36 16.7 33.3
47 poor Ansonia 2007
                     8 Mathematics
   V10
       V11 V12
40 33.3 66.7 0.0
41 23.3 6.7 6.7
42 40.0 14.3 0.0
44 41.9 23.3 9.3
46 0.0 33.3 0.0
47 36.1 13.9 0.0
```

## 7.3 Compute expected scale values for ethnic groups:

From

http://solutions1.emetric.net/CMTPublic/CMTCode/Report.aspx?data=6581DA E1BB8B81461ADB4AAC5B5808D

The expected scale score values corresponding to each performance level for each ethnic group are based on the distributions over the performance levels for each ethnic group in 2009. These distributions change only a little from year to year. It is assumed that the underlying scale score values for each ethnic group are normally distributed, and the mean and standard deviation for the normal is determined by the given performance level distributions.

```
#Black 2009
bes <- expectScale(c(16, 20, 32, 25))
#hispanic 2009
hes <- expectScale(c(17, 20, 30, 26))
#white 2009
wes <- expectScale(c(2, 5, 16, 39))
#asian 2009
aes <- expectScale(c(2, 4, 11, 33))
rbind(bes, hes, wes, aes)
     [,1] [,2] [,3] [,4] [,5]</pre>
```

```
bes 168 202 230 263 303
hes 168 202 229 262 304
wes 175 204 231 266 316
aes 173 204 230 267 326
```

```
eth <- dpr[, 6]
black <- eth == "Black"
hispanic <- eth == "Hispanic"
white <- eth == "White"
asian <- eth == "Asian"
dpr$scale <- NA</pre>
```

```
dpr$scale[black] <- round(as.matrix(dpr[black,
8:12]) %*% bes /100 )
dpr$scale[hispanic] <-
round(as.matrix(dpr[hispanic, 8:12]) %*% hes/100)
dpr$scale[white] <- round( as.matrix(dpr[white,
8:12]) %*% wes/100 )
dpr$scale[asian] <- round( as.matrix(dpr[asian,
8:12]) %*% aes/100 )
dpr$goal <- dpr[, 11] + dpr[, 12]
dpr <- dpr[dpr[, 4] == 8, c(1, 2, 3, 6, 7, 13, 14)]
names(dpr) <- c("Wealth", "District", "Year",
        "Ethnicity","Number", "Scale", "Goal")
head(dpr)
Wealth District Year Ethnicity Number Scale Goal
```

40	poor	Ansonia	2006	Asian	3	255	66.7
41	poor	Ansonia	2006	Black	30	211	13.4
42	poor	Ansonia	2006	Hispanic	35	213	14.3
44	poor	Ansonia	2006	White	43	236	32.6
46	poor	Ansonia	2007	Asian	3	214	33.3
47	poor	Ansonia	2007	Black	36	215	13.9

```
Save data assembled so far
write.csv(dpr, file="data/Conn8MathData.csv",
row.names=F)
```

40

### 7.4 Reframe data by district and year:

The reason for doing this is to compare different groups within districts.

```
dpr <- read.csv("data/Conn8MathData.csv", as.is=T,
header=T)
```

```
head(dpr, 4)
```

Wealth District Year Ethnicity Number Scale Goal								
1	poor	Ansonia	2006	Asian	3	255 66.7		
2	poor	Ansonia	2006	Black	30	211 13.4		
3	poor	Ansonia	2006	Hispanic	35	213 14.3		
4	poor	Ansonia	2006	White	43	236 32.6		

```
Initialise new data frame, one row for each district and year:
dname=unique(dpr$District)
years <- unique(dpr$Year)
eth <- unique(dpr$Ethnicity)
print(eth)
```

[1] "Asian" "Black" "Hispanic" "White"

```
dm < -
data.frame(matrix(0,length(dname)*length(years),18)
)
names(dm) <- c("District", "Year",</pre>
paste(eth, "pN", sep="."), paste(eth, "pS",
sep="."), paste(eth, "rN", sep="."), paste(eth,
"rS", sep="."))
dm$Total <- dm$Year
for (i in 1:length(dname)) {
for ( j in 1:length(years)) {
ij <- 6 * (i - 1) + j
dm[ij, 1] <- dname[i]</pre>
dm[ij, 2] <- years[j]</pre>
uij<-dpr$District==dname[i] & dpr$Year == years[j]</pre>
for (k in 1:length(eth)) {
u <- uij&dpr$Ethnicity==eth[k] &</pre>
```

```
dpr$Wealth=="poor"
if (sum(u) == 1) dm[ij, k + 2] <- dpr[u, 5]
if (sum(u) == 1) dm[ij, k + 6] <- dpr[u, 6]
u <- uij & dpr$Ethnicity == eth[k] & dpr$Wealth ==
"rich"
if (sum(u) == 1) dm[ij, k + 10] <- dpr[u, 5]
if (sum(u) == 1) dm[ij, k + 14] <- dpr[u, 6]
}
dm$Total[ij] <- sum(dm[ij, c(3:6, 11:14)])</pre>
}
}
dm$minority <-
dm$Black.pN + dm$Hispanic.pN + dm$Black.rN +
dm$Hispanic.rN
dm$minority <- round(100*dm$minority/dm$Total, 2)</pre>
dm <- na.omit(dm)
tt <- table(dm$District, dm$Year)</pre>
dy <- apply(tt, 1, sum)</pre>
names (dy) [dy < 6]
[1] "Achievement First" "Cornwall"
[3] "Elm City Col Prep" "Highville"
[5] "Jumoke Academy"
                        "LEARN"
[7] "New Beginnings"
                        "Park City Prep"
[9] "The Bridge Academy"
write.csv(dm,"data/DistrictYearConnMath8.csv",
  row.names=F)
dm[1, ]
District Year Asian.pN Black.pN Hispanic.pN White.pN
1 Ansonia 2006
                      3
                             30
                                         35
                                                  43
 Asian.pS Black.pS Hispanic.pS White.pS Asian.rN Black.rN
1
      255
               211
                          213
                                   236
                                              4
                                                      13
 Hispanic.rN White.rN Asian.rS Black.rS Hispanic.rS
                          296
                                   230
          18
                   98
                                               223
1
 White.rS Total minority
1
      249
            244
                    39.3
```

### 7.5 Sum data over years

```
Many districts have small minority counts; sum over years to make
comparisons possible:
dm <- read.csv("data/DistrictYearConnMath8.csv",</pre>
                 header=T, as.is=T)
udistrict <- unique(dm$District)</pre>
nrows <- length(udistrict)</pre>
dma <- dm[1:nrows, ]</pre>
dma$District <- udistrict
for (d in 1:nrows) {
 use <- dm$District == udistrict[d]</pre>
 dma$Total[d] <- sum(dm$Total[use])</pre>
     (var in c(3:6, 11:14) ){
 for
   dma[d, var] <- sum(dm[use, var])</pre>
 }
 for (var in c(7:10, 15:18) ) {
   vartotal <- dma[d, var-4]</pre>
   dma[d, var] <- NA
    if
         (vartotal > 0) dma[d, var] <-
       sum(dm[use, var-4]*dm[use, var])/vartotal
 }
ł
dma[, -1] <- round(dma[, -1])
dma <- dma[, -2]
dma[1, ]
District Asian.pN Black.pN Hispanic.pN White.pN Asian.pS
                 11
                         177
                                     186
                                              248
                                                       251
1 Ansonia
 Black.pS Hispanic.pS White.pS Asian.rN Black.rN
1
       222
                   231
                            245
                                      13
                                               78
 Hispanic.rN White.rN Asian.rS Black.rS Hispanic.rS
                           274
1
           81
                   433
                                     233
                                                 235
 White.rS Total minority
                       39
1
       257 1227
```

Minority has to be recomputed to reflect sums over years: dma\$minority <- dma\$Black.pN + dma\$Hispanic.pN + dma\$Black.rN + dma\$Hispanic.rN dma\$minority<-round(100\*dma\$minority/dma\$Total, 2)</pre>

### 7.6 Add per capita income:

We get per capita income from <u>http://www.sde.ct.gov/sde/cwp</u>

head(pci)

		<b>V1</b>	V2	<b>V</b> 3	V4			
1	ANDOVER		376,368,494	3,210	30,273			
2	ANSONIA		1,533,969,464	18,514	20,504			
3	ASHFORD		462,339,581	4,470	26,104			
4	AVON		3,744,303,900	17,357	51,706			
5	BARKHAMSTED		527,705,389	3,692	28,961			
6	BEACON FALLS		685,384,414	5,866	25,285			
	<b>V</b> 5	<b>V6</b>						
1	43,260.39	102						
2	20,705.31	161						
3	32,906.93	130						
4	135,945.26	21						
5	50,451.04	79						
6	36,006.58	122						
p p t	<pre>pci\$V4 &lt;- as.numeric(sub(",", "", pci\$V4)) pci\$V1 &lt;- paste(substr(pci\$V1, 1, 1), tolower(substr(pci\$V1, 2, 20)), sep="")</pre>							

Match the districts in the two data sets with brute force, after eliminating blanks:

```
dma[, 1] <- gsub(" ", "", dma[, 1])
# initialise pci value in dma array
dma$pci <- NA
pci$V1 <- gsub(" ", "", pci$V1)</pre>
```

```
for (i in 1:dim(dma)[1]){
  for (ii in 1:length(pci$V1)){
    if (pci$V1[ii] == dma$District[i])
      dma$pci[i] <- pci$V4[ii]
  }
}
write.csv(dma, "data/DistrictConn8Math.csv",
row.names=F)
head(dma, 2)</pre>
```

```
District Asian.pN Black.pN Hispanic.pN White.pN Asian.pS
1 Ansonia
                 11
                          177
                                      186
                                                248
                                                         251
2 Ashford
                  3
                            1
                                         6
                                                 39
                                                         256
 Black.pS Hispanic.pS White.pS Asian.rN Black.rN
       222
                   231
                             245
                                        13
                                                 78
1
                                         2
2
       263
                   214
                             271
                                                  3
  Hispanic.rN White.rN Asian.rS Black.rS Hispanic.rS
1
           81
                             274
                                      233
                                                   235
                    433
            5
2
                   247
                             248
                                      303
                                                   296
 White.rS Total minority
                             pci
       257 1227
                       522 20504
1
2
       274
                        15 26104
             306
```

### **8** Functions

```
Grid <- function(xticks, yticks, ylab="",</pre>
at=(min(xticks)+ mean(xticks))/2, cex=2.5){
# background for plot using grid of light grey
lines
par(mar=c(3,3,6,2))
plot(1, 1, xlim=range(xticks), ylim =
range(yticks),
   xlab="", ylab="", axes=F, pch="")
# use only interior values of tick ranges in plots
usey <- rep( T, length(yticks) )</pre>
usey[c( 1, length(yticks) )] <- F</pre>
usex <- rep( T, length(xticks) )</pre>
usex[c( 1, length(xticks) )] <- F</pre>
# grey lines in both directions
for ( row in yticks[usey] )
 lines(range(xticks),c(row, row),col="light grey")
for ( col in xticks[usex] )
 lines(c(col, col),range(yticks),col="light grey")
# put ylab on left top, using / to split long
expressions
ylabs <- unlist(strsplit(ylab,"/"))</pre>
# identify tick marks on both axes
if (length(yticks) > 2)
 text(pos=2, rep(min(xticks), length(yticks)-2),
      yticks[usey], yticks[usey], cex=2, xpd=T)
if (length(xticks)>2)
 text(pos=1, xticks[usex], rep(min(yticks),
      length(xticks)-2), xticks[usex], cex=2, xpd=T)
lylabs <- min(5, length(ylabs))</pre>
```

```
if(lylabs > 0)
 mtext(ylabs,side=3,line =(5/lylabs)*(lylabs-1):0,
       at = at, cex=cex)
par(mar=c(5, 4, 4, 2))
invisible()
}
quad <- function(x, diff, w1, y1, w2, y2,
col=1, lwd=1, minwt=0)
{
# draws a polygon representing a trend between two
weighted points; points must have weight greater
than minwt to be included; shiver the x differences
a bit, randomly, to avoid overwritten lines
if (length(x) != length(w1) |
   length(x) != length(y1))
 stop(" x w1 y1 not equal length")
if (length(x) != length(w2) |
    length(x) !=
                     length(y2))
 stop(" x w2 y2 not equal length")
x1 < - x
x^2 < -x
diff <- diff*(0.9 + 0.2 * runif(length(x)))
for (i in 1:length(w1)) {
 if
     (w1[i] >= minwt & w2[i] >= minwt) {
   x1[i] <- x[i] - diff[i]</pre>
   x2[i] <- x[i] + diff[i]</pre>
   lines(c(x1[i], x1[i], x2[i], x2[i], x1[i]),
         c(y1[i] - w1[i]/2, y1[i]+w1[i]/2,
   y2[i]+w2[i]/2, y2[i] - w2[i]/2, y1[i]-w1[i]/2),
          col=col, lwd=lwd)
 }
}
invisible()
}
```

```
meanline <-
function(x, y, w, cutpoints=range(x),col=1, lwd=1) {
# draw meanlines in intervals set by cutpoints on x
xb <- rep(0, length(cutpoints)-1)</pre>
yb <- xb
wt <-xb
for (i in 1: (length(cutpoints) - 1)) {
 use <- x >= cutpoints[i] & x < cutpoints[i + 1]</pre>
 xb[i] <- sum((w*x)[use & !is.na(x)])/sum(w[use])</pre>
 yb[i] <- sum((w*y)[use & !is.na(y)])/sum(w[use])</pre>
 wt[i] <- sum(w[use])</pre>
}
lines(xb, yb, col=col, lwd=lwd)
points(xb, yb, col=col, cex=sqrt(wt)/20)
}
df <- function(xx,data) {</pre>
colx <- which(names(data) %in% xx)</pre>
if(length(colx)==0) return(NULL)
return(data[, colx])
}
```

```
circle <- function(x, y, d, s, border, lwd, data) {
# circle of radius d/s located at x,y
x <- df(deparse(substitute(x)),data)</pre>
y <- df(deparse(substitute(y)),data)</pre>
d <- df(deparse(substitute(d)),data)</pre>
d < - d/s
lx <- length(x)
xcircle <- cos( (1:100) * 2*pi/100 )</pre>
ycircle <- sin( (1:100) * 2*pi/100 )</pre>
for( case in 1:1x)
polygon(x[case] + d[case] * xcircle,
  y[case] + d[case] * ycircle,
  border=border, lwd=lwd)
invisible()
}
expectScale <- function(p) {</pre>
# expected scale values within each performance
level for each ethnic group, assuming normal
distribution of scale values over group.
# cuts are chosen by connecticut education dept
# p specifies percentiles within first 4 groups
perc <- cumsum(p)/100
z <- qnorm(perc)</pre>
# compute expectations within each performance
level on normal scale
zz < - c(-5, z, 5)
phi <- dnorm(zz)</pre>
cump <-pnorm(zz)</pre>
```

```
expz <- -diff(phi)/diff(cump)
# coef gives the mean and sd of the corresponding
normal
cuts <- c(190, 213, 244, 287)
coef <- summary(lm(cuts ~ z))$coef[, 1]
# rescale back to scale values according to coef
return( round(coef[1] + coef[2]*expz) )
}</pre>
```