Using the Illinois Report Card Data to Teach Statistics

MMC Conference of Workshops

Frank Briody
Prospect High School
frankbriody@gmail.com

2/1/2020

Contents

1 Variables

The ISBE raw data file *rx17.txt* contains 1,471 variables. The variable definitions are in the Excel file *RC17_layout.xlsx* and have been categorized into the groups shown below. The first number represents available variables in each group while the second is the number actually imported into the processed data file. The import script produces 316 variables from 20 of the 21 categories for all 3,796 Illinois public schools. (None of the NAEP variables were imported.) Usable files will be discussed in section 3.

School information (13 variables;12 imported)
Student demographics (396;71)
ACT (44;11)
Instructional setting (92;2)
Teacher and admin statistics (78;26)
District financial (67;40)
Region and legislative (3;2)
National Assmnt. of Educ. Progress (NAEP) (184;0)
College and Career readiness (16;3) CTE (4;1)
Advanced coursework (12;3)
AP courses (168;42)
IB courses (168;42)
Dual credit (168;42)
AP exams (36;12)
Post secondary remediation (4;1)
Response rate (5E survey) (4;2)
Health and wellness (3;1)
Teacher Attendance (4;1)
Teacher Evaluation (2;1)
School District Count (3;1)

---

2 Descriptive Statistics via State Demographics

2.1 Categorical Count (Raw)

```
school_type <- rc17 %>%
  count(SCHOOL_TYPE_NAME, sort = TRUE) %>%
  mutate(rel_freq = n/sum(n))
school_type
```

```
## # A tibble: 4 x 3
##   SCHOOL_TYPE_NAME     n rel_freq
##   <chr>            <int>    <dbl>
## 1 ELEMENTARY        2406   0.634
## 2 HIGH SCHOOL        644   0.170
## 3 MIDDLE SCHL        604   0.159
## 4 CHARTER SCH        142   0.0374
```

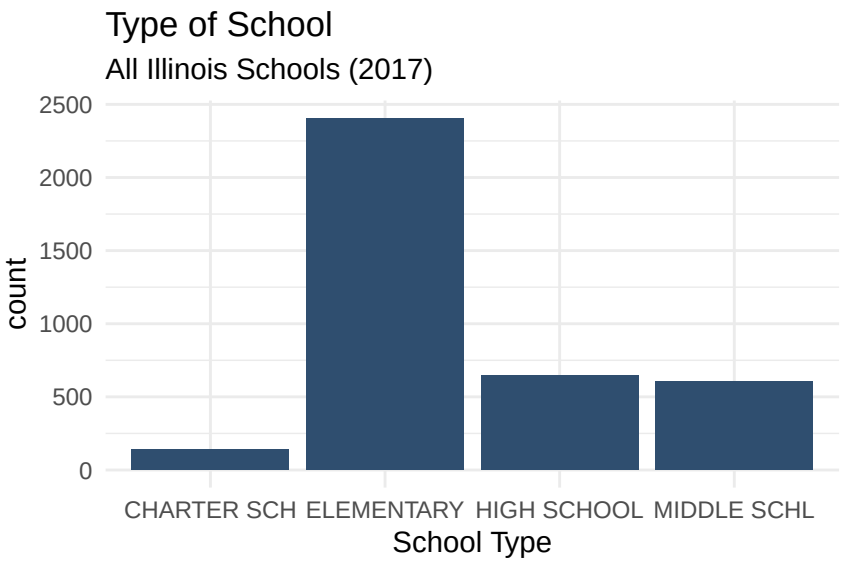
2.2 Categorical Count (Formatted)

```
kable(school_type) %>%
  kable_styling(bootstrap_options = "striped", full_width = F)
```

SCHOOL_TYPE_NAME	n	rel_freq
ELEMENTARY	2406	0.6338251
HIGH SCHOOL	644	0.1696523
MIDDLE SCHL	604	0.1591149
CHARTER SCH	142	0.0374078

2.3 Categorical Plot

```
ggplot(rc17, aes(x=factor(SCHOOL_TYPE_NAME)))+
  geom_bar(fill="#2F4E6F")+
  labs(title = "Type of School", x = "School Type", subtitle = "All Illinois Schools (2017)") +
  theme_minimal()
```

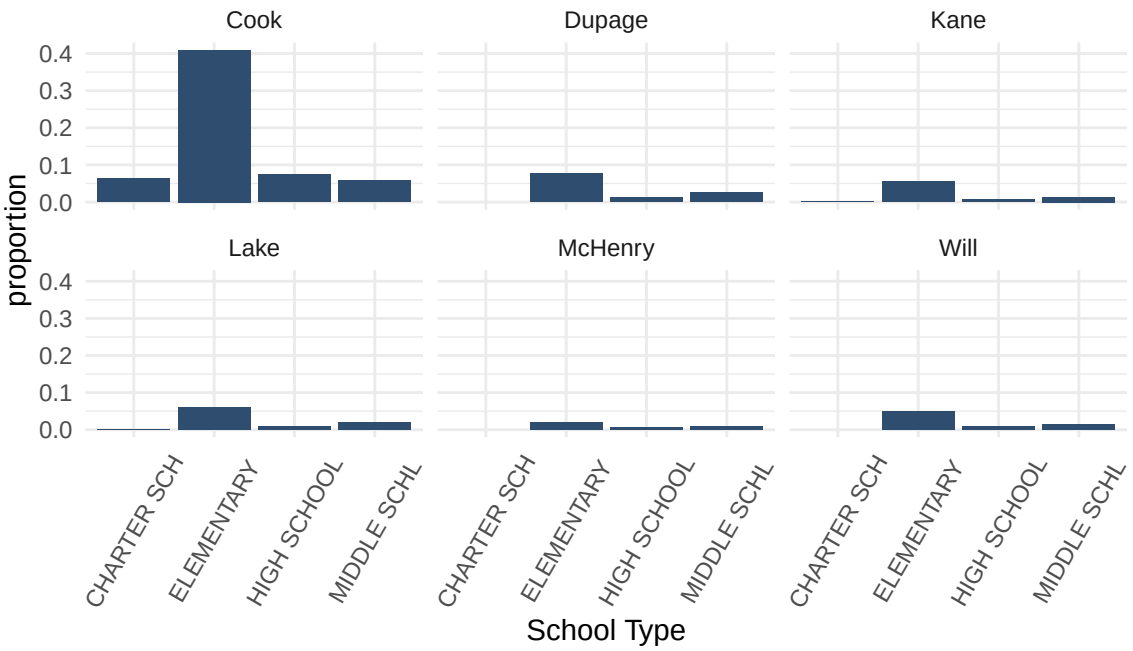


2.4 Categorical Analysis I

Write a short analysis for the types of schools in the state of Illinois.

2.5 Categorical Analysis II

```
rc17 %>%
filter(COUNTY == "Dupage" | COUNTY == "Will" | COUNTY == "Kane" |
         COUNTY == "Lake" | COUNTY == "Cook"| COUNTY == "McHenry") %>%
ggplot(aes(x=factor(SCHOOL_TYPE_NAME), y = (..count..)/sum(..count..))) +
  geom_bar(fill="#2F4E6F")+
  facet_wrap(~COUNTY, nrow = 2) +
  labs(title = "Type of School by County",
       x = "School Type",
       y = "proportion",
       subtitle = "Six Counties in the Chicago Metropolitan Region (2017)") +
  theme_minimal()  +
  theme(axis.text.x = element_text(angle = 60, vjust = 0.5))
```



Write a short analysis for the types of schools in the six county region.

3 Data Import

3.1 Data Files

- ISBE Report Card Data Library [https://www.isbe.net/Pages/Illinois-State-Report-Card-Data.aspx]
  - rc17.txt
  - six_county
- Import script
  - define variables
  - fix issues i.e. "$" and ","
  - load libraries
  - available here

Creating the six county subset:

```
sixco <- rc17 %>%
  filter(COUNTY %in% c("Cook", "Lake", "Will", "Kane", "McHenry", "Dupage"))
```

4 Distributions

4.1 Lists

4.1.1 Number of High Schools in the Six County Region

```
sixco %>%
  filter(SCHOOL_TYPE_NAME == "HIGH SCHOOL") %>%
  group_by(COUNTY) %>%
  summarise(count = n())
```

```
## # A tibble: 6 x 2
##   COUNTY  count
##   <chr>   <int>
## 1 Cook      151
## 2 Dupage     23
## 3 Kane       16
## 4 Lake       21
## 5 McHenry    14
## 6 Will       17
```

4.1.2 Single List of Scores

```
mchenry_act <- rc17%>% filter(COUNTY == "McHenry", is.na(ACT_COMP_SCHOOL) == FALSE )
mchenry_act$ACT_COMP_SCHOOL
```

```
##  [1] 22.4 19.7 18.1 23.1 22.6 23.8 22.7 24.0 21.1 19.9 22.9 22.9 21.4 21.2
```

4.1.3 Lake County ACT Scores (Ordered and Formatted)

```
lake_ACT <- rc17 %>%
  filter(SCHOOL_TYPE_NAME == "HIGH SCHOOL", COUNTY == "Lake") %>%
  arrange(desc(ACT_COMP_SCHOOL)) %>%
  select(COUNTY, SCHOOL_NAME, ACT = ACT_COMP_SCHOOL)
kable(lake_ACT)
```

COUNTY	SCHOOL_NAME	ACT
Lake	Adlai E Stevenson High School	26.9
Lake	Deerfield High School	26.4
Lake	Lake Forest High School	26.3
Lake	Libertyville High School	25.9
Lake	Highland Park High School	25.2
Lake	Vernon Hills High School	25.1
Lake	Lake Zurich High School	24.9
Lake	Barrington High School	24.8
Lake	Grayslake Central High School	23.3
Lake	Lakes Community High School	22.6
Lake	Grayslake North High School	22.4
Lake	Warren Township High School	22.1
Lake	Wauconda High School	21.8
Lake	Antioch Comm High School	21.7
Lake	Mundelein Cons High School	21.4
Lake	Grant Community High School	21.3
Lake	New Tech High - Zion-Benton East	20.1
Lake	Zion-Benton Twnshp Hi Sch	18.6
Lake	Waukegan High School	17.9
Lake	Round Lake Senior High School	17.8
Lake	North Chicago Community High Sch	17.5

Table 1: Types of School Districts

	Dupage	Lake
LARGE	182	150
MEDIUM	52	38
SMALL	0	4

Create a boxplot for Lake County ACT scores. How could you compare to DuPage county?

4.2 Two-Way Tables

```
district_type <- rc17 %>%
  filter(COUNTY == "Lake" | COUNTY == "Dupage") %>%
  group_by(COUNTY)

two_way <- with(district_type, table(DISTRICT_SIZE_NAME, COUNTY))

kable(two_way, caption = "Types of School Districts") %>%
  kable_styling(bootstrap_options = "striped", full_width = F)
```

5 Correlation and Regression

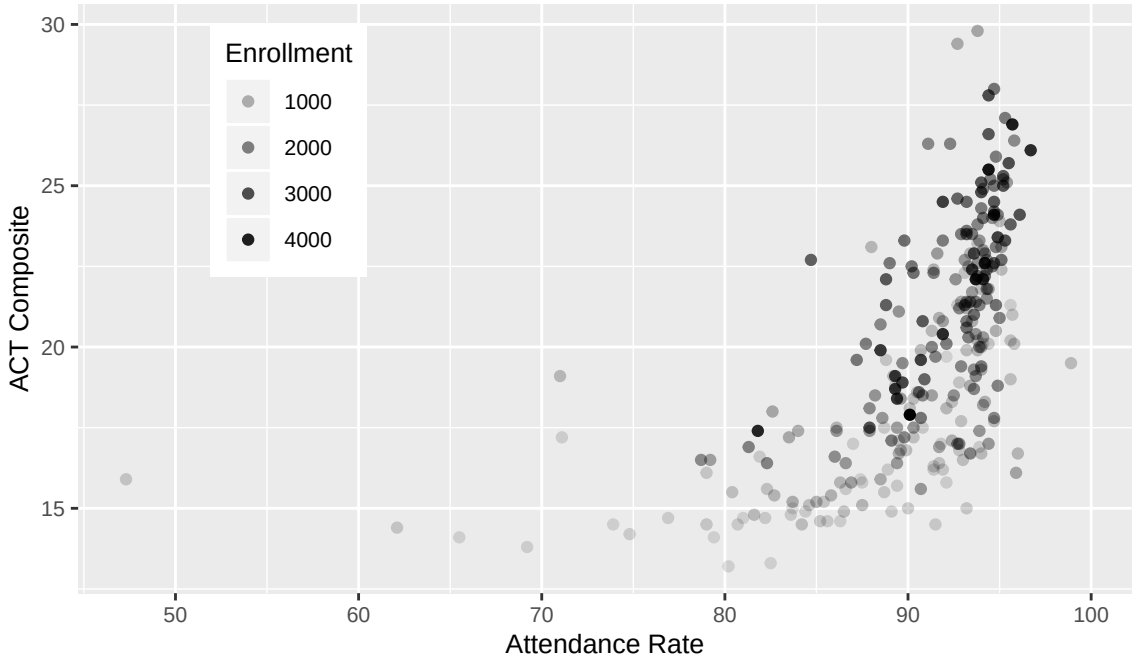
5.1 Guess the Correlation

- ACT Composite vs Chronically Truant (#) Guess:_____ Actual:_____
- ACT Composite vs Chronically Truant (%) Guess:_____ Actual:_____
- ACT Composite vs Student Mobility Guess:_____ Actual:_____
- ACT Composite vs Attendance rate (%) Guess:_____ Actual:_____

5.2 Predicting ACT Scores

Suppose you choose 3 variables (Composite ACT Score, Enrollment and Attendance Rate) for all schools in the Six County region. What question(s) and display(s) would you explore?

5.3 Scatterplot Analysis



5.4 Regression Output

5.4.1 All Schools

```
summary(lm(sixco$ACT_COMP_SCHOOL~sixco$ATTENDANCE_RATE_SCHOOL_perALL))
```

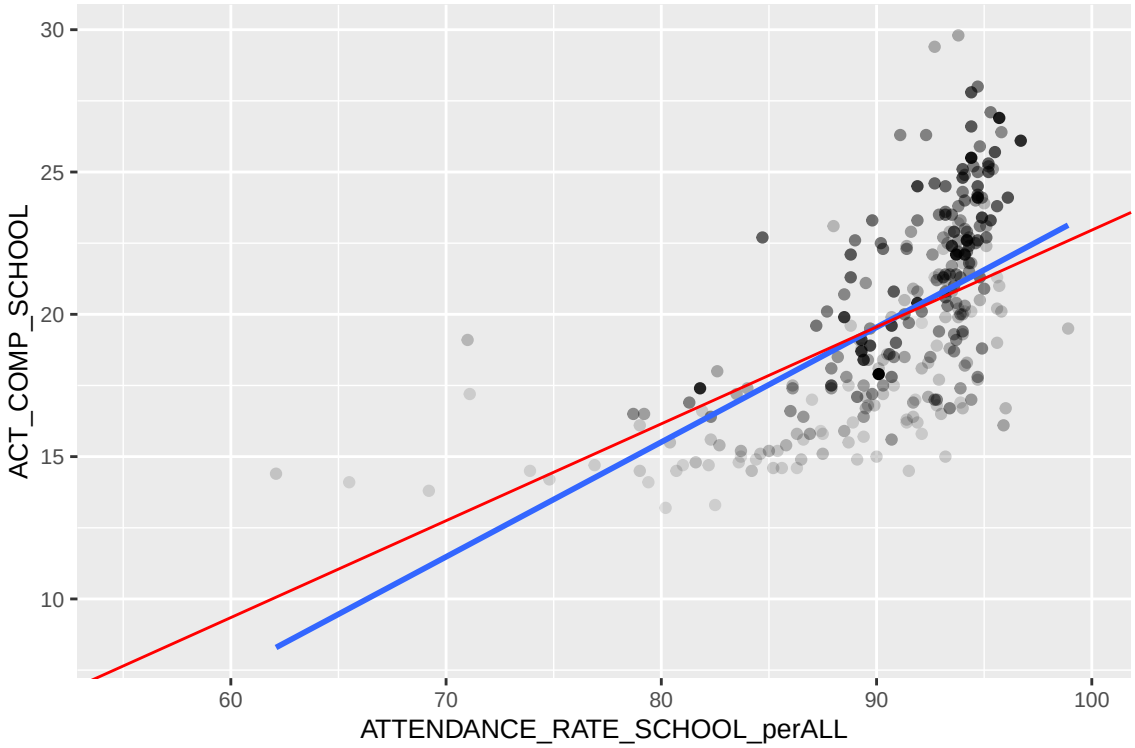
```
##
## Call:
## lm(formula = sixco$ACT_COMP_SCHOOL ~ sixco$ATTENDANCE_RATE_SCHOOL_perALL)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.7088 -2.2108 -0.2343  1.8209 10.9398
##
## Coefficients:
##                                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)                       -11.26875    2.61192  -4.314 2.25e-05
## sixco$ATTENDANCE_RATE_SCHOOL_perALL   0.34311    0.02887  11.886  < 2e-16
##
## (Intercept)                       ***
## sixco$ATTENDANCE_RATE_SCHOOL_perALL ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.844 on 268 degrees of freedom
##   (1779 observations deleted due to missingness)
## Multiple R-squared:  0.3452, Adjusted R-squared:  0.3427
## F-statistic: 141.3 on 1 and 268 DF,  p-value: < 2.2e-16
```

5.4.2 Influential Removed

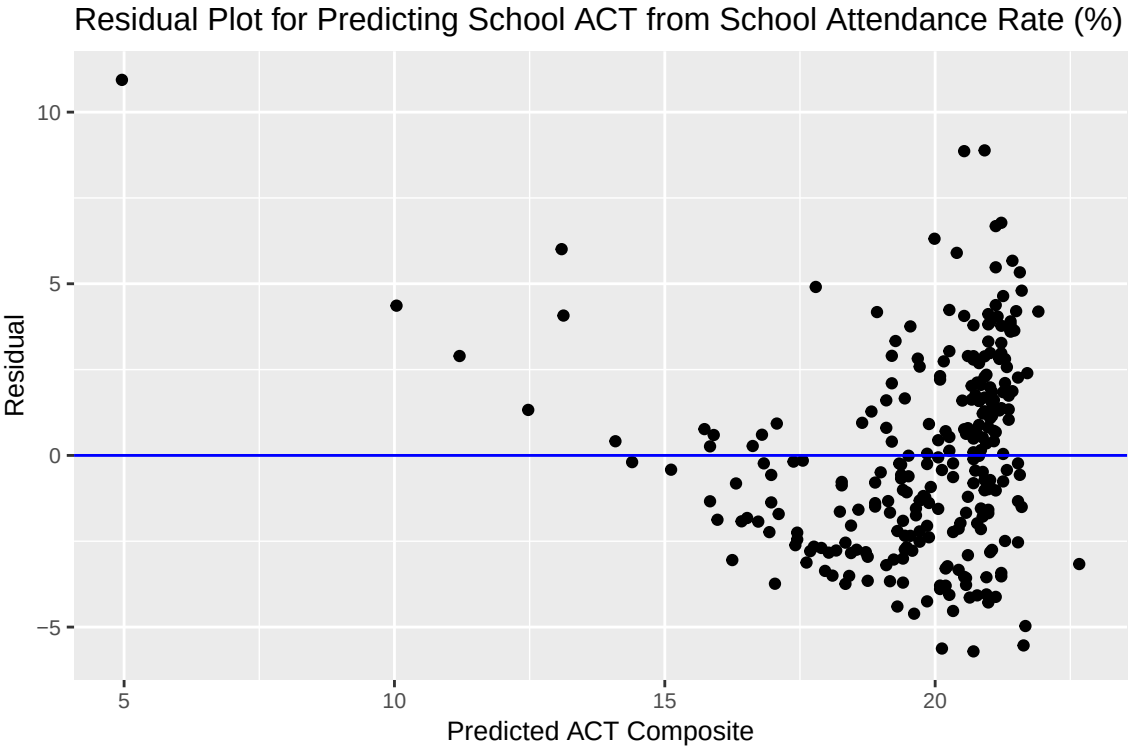
```
sixco_removed <- sixco %>%
  filter(ATTENDANCE_RATE_SCHOOL_perALL>50)

  summary(lm(sixco_removed$ACT_COMP_SCHOOL~sixco_removed$ATTENDANCE_RATE_SCHOOL_perALL))
```

```
##
## Call:
## lm(formula = sixco_removed$ACT_COMP_SCHOOL ~ sixco_removed$ATTENDANCE_RATE_SCHOOL_perALL)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.8340 -2.1147 -0.1728  1.6627  8.7676
##
## Coefficients:
##                                            Estimate Std. Error t value
## (Intercept)                                -16.74644    2.81182  -5.956
## sixco_removed$ATTENDANCE_RATE_SCHOOL_perALL   0.40322    0.03104  12.993
##                                           Pr(>|t|)
## (Intercept)                               8.13e-09 ***
## sixco_removed$ATTENDANCE_RATE_SCHOOL_perALL  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.75 on 267 degrees of freedom
##   (1778 observations deleted due to missingness)
## Multiple R-squared:  0.3873, Adjusted R-squared:  0.385
## F-statistic: 168.8 on 1 and 267 DF,  p-value: < 2.2e-16
```

5.4.3 Both Models



5.4.4 Residual Plot



6 Original Material

6.1 R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

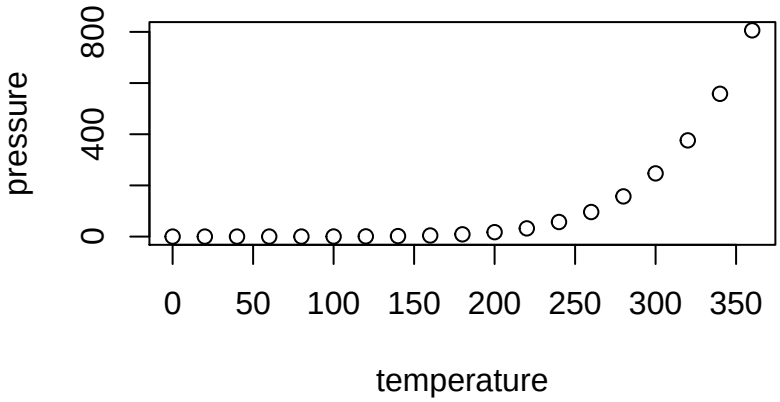
When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

```
##      speed           dist
##  Min.   : 4.0   Min.   :  2.00
##  1st Qu.:12.0   1st Qu.: 26.00
##  Median :15.0   Median : 36.00
##  Mean   :15.4   Mean   : 42.98
##  3rd Qu.:19.0   3rd Qu.: 56.00
##  Max.   :25.0   Max.   :120.00
```

6.2 Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.